אלונה רביב

סטטיסטיקה אפרמטרית

מהדורה שניה (אלקטרונית)

על הספר

בספר **סטטיסטיקה אפרמטרית** מוצגות השיטות האפרמטריות המקובלות ביותר, באופן שיתאים כספר לימוד לקורס סמסטריאלי של שלוש שעות שבועיות. הספר נכתב לתלמידים הלומדים סטטיסטיקה כחוג ראשי או משני, והוא פרי ניסיון רב של המחברת בהוראת קורס כזה בחוג לסטטיסטיקה וחקר ביצועים באוניברסיטת תל אביב. הספר מתאים גם לתלמידים בחוגים אחרים, אשר למדו קורס אלמנטרי במבוא לסטטיסטיקה בלבד, ואשר לומדים שיטות אפרמטריות כקורס נפרד או כחלק מקורס בסטטיסטיקה.

הניתוח הסטטיסטי והסקת המסקנות מובאים בצורה פשוטה וברורה המתאימה לקהל רחב של קוראים. על כן הספר יכול להיות לעזר רב לחוקרים במדעי החברה, בביולוגיה או ברפואה, שזקוקים בעבודתם לשימוש בשיטות אפרמטריות לצורך הניתוח הסטטיסטי.

הספר מכיל מבחר של תרגילים עם תשובות לחלקם הגדול וכן טבלאות התפלגות של רוב המבחנים המובאים כאן.

הספר יצא לראשונה בדפוס בהוצאת עמיחי (2006).
המהדורה השניה, האלקטרונית הזאת (בהוצאת המחברת) של הספר "סטטיסטיקה אפרמטרית" הוכנה לשימוש **חפשי** של מרצים, סטודנטים וחוקרים המתעניינים בנושא.

על המחַברת

ד"ר אלונה רביב שימשה שנים רבות כמרצה בחוג לסטטיסטיקה וחקר ביצועים באוניברסיטת תל אביב ולימדה קורסים רבים בסטטיסטיקה לתואר ראשון ולתואר שני. היא חיברה, יחד עם פרופ' תלמה לויתן, את שני הספרים המצליחים במבוא להסתברות וסטטיסטיקה: "הסתברות" ו"הסקה סטטיסטית", בהוצאת "עמיחי".

מבוא

בבעיות רבות השיטות האפרמטריות לניתוח נתוני מחקר עדיפות על השיטות הפרמטריות המקובלות. ראשית, לעתים קרובות ההנחות המתחייבות עבור שימוש בשיטות הפרמטריות (המתאימות למודלים נורמלים) אינן מתקיימות. נוסף על כך, במקרים רבים אנו מוצאים שהעוצמה של המבחנים המקובלים עבור מודלים נורמלים היא יחסית נמוכה, בהשוואה לשיטה האפרמטרית המתאימה. חשוב להכיר את השיטות האפרמטריות, שהן מאוד מקובלות ונמצאות בשימוש רב בניתוחים סטטיסטיים של נתוני מחקר. רוב המבחנים האפרמטריים מאופיינים בפשטותם, בהעדר צורך בחישובים מסובכים, וכמו כן, הם מהווים מקור רב השראה להיכרות של הקורא עם החשיבה המקורית והייחודית של סטטיסטיקאים ועם רעיונות סטטיסטיים מעניינים ויפים לפתרון בעיות.

בספר זה מוצגות רק השיטות האפרמטריות המקובלות ביותר, באופן שיתאים כספר לימוד (text book) לקורס של סמסטר אחד בן שלוש שעות שבועיות, לפי ניסיוני בהוראת קורס כזה במשך שנים רבות בחוג לסטטיסטיקה וחקר ביצועים באוניברסיטת תל אביב. הספר מתאים לתלמידים הלומדים סטטיסטיקה כחוג ראשי או משני, אולם גם לתלמידים בחוגים אחרים, אשר למדו רק קורס אלמנטרי במבוא לסטטיסטיקה ולומדים שיטות אפרמטריות כקורס נפרד או כחלק מקורס בסטטיסטיקה.

הספר יכול להועיל לחוקרים שמשתמשים בעבודתם בשיטות אפרמטריות. ניסיתי להביא כאן את צורת השימוש והסקת המסקנות בצורה פשוטה וברורה לקוראים.

נושאים מסוימים וכן הנמקות או הוכחות ברמה גבוהה יותר, שעליהן ניתן לדלג, מסומנות בכוכבית. בנוסף, חלק מן ההוכחות מובאות בנספח, כדי לא להעמיס על הקוראים במהלך השוטף של הלימוד.

מבחינה טכנית של הבאת החומר, הנוסחאות מסומנות במספרים סידוריים בכל פרק בנפרד. בעת התייחסות בטקסט לנוסחה מפרק שונה מזה שבו אנו נמצאים, אנו מצביעים על הפרק המתאים. כשאין אזכור של פרק אחר, הכוונה היא לנוסחה באותו פרק.

הספר אינו מתיימר לכלול את כל השיטות האפרמטריות שפותחו ופורסמו עד עתה, והן רבות מאוד, ולכן גם לא הבאנו רשימה גדולה של מקורות. רשימת המקורות כוללת את העבודות הראשונות שפורסמו בנושאים שעליהם אנו מדווחים בספר. פרסומים אלה

הם, למעשה, העבודות הראשונות שפורסמו בשיטות אפרמטריות, והם מציגים את ראשוני המדענים שעבדו בנושאים אלה.

הפרק הראשון בספר הוא הקדמה ובו אנו מציגים את המושגים הקשורים בהסקה סטטיסטית, וכן את ההגדרה של שיטה אפרמטרית. בפרקים הבאים מובאות השיטות האפרמטריות המתאימות לבעיות מחקר שונות. במידת האפשר (כפונקציה של הרמה המתמטית הנדרשת לפתרון) השתדלנו להביא גם חישובי עוצמה של המבחנים, לפחות בקירוב. זאת, כדי לאפשר לחוקר לקבוע סדר גודל של מספר התצפיות שתידרשנה כדי שהמחקר יהיה יעיל. לא הבאנו כאן מבחני חי־בריבוע עבור לוחות שכיחות. נושא זה מוצג כמעט בכל ספר של מבוא לסטטיסטיקה והוא מובא באופן מאוד מפורט ובהיר בספר "מבוא להסתברות וסטטיסטיקה: הסקה סטטיסטית" שכתבתי עם עמיתתי תלמה לויתן, המיועד לתלמידי כלכלה (רביב ולויתן, 2000).

בסוף כל פרק הבאנו אוסף של תרגילים, ובסוף הספר נמצאות תשובות לחלקם הגדול.

בספר הבאנו טבלאות של התפלגויות הסטטיסטים היותר מקובלים ושימושיים המוצגים בַּספר. טבלאות נוספות ניתן למצוא בספרים רבים המוקדשים לשיטות אפרמטריות כמו, למשל, Lehmann, 1975; Conover, 1980; או Hollander and Wolfe, 1975.

לגבי כל השיטות המוצגות בספר זה ישנם קירובים טובים מאוד של ההסתברויות הדרושות, שבהם ניתן להסתפק ברוב המקרים.

אני מבקשת להודות לחוג לסטטיסטיקה וחקר ביצועים בבית הספר למדעי המתמטיקה באוניברסיטת תל אביב על העזרה הרבה בהכנת הספר. כמו כן אני מודה לעמיתַי, לד"ר תלמה לויתן, שעברה על חלק מן הפרקים והעירה הערות חשובות ולפרופ' יואב בנימיני על ההערות והעידוד. תודתי נתונה גם לחברַי החוקרים ד"ר יפה זינגר, ד"ר אריקה עמיר, ד"ר דורית ארם, פרופ' אבי שדה ופרופ' דניאל בר־טל, שהעמידו לרשותי נתונים שאספו במחקריהם, ובהם השתמשתי כדוגמאות וכתרגילי בית. כמו כן השתמשתי בנתונים מעבודות מחקר של הסטודנטיות ברכה בירן, עירית דר, שירן רזנבלט-שטיין, ועדנה שדה ואני מודה גם להן. בספר זה יש שימוש גם בנתונים ממחקרים שהגיעו לידי במשך זמן רב ולצערי לא הצלחתי לשחזר את מקורם. אני מתנצלת בזאת בפני כל אלה שלא הזכרתי את שמם ומודה להם מאוד. תודתי נתונה גם לתלמידי הקורס "שיטות אפרמטריות" בחוג לסטטיסטיקה באוניברסיטת תל אביב, בשנת תשס"ה, שעברו על הטיוטה הראשונה של הספר והעירו הערות מועילות. אחרון אחרון חביב, אני מודה מקרב לב לבעלי, עמירם, שללא תמיכתו ועידודו ספר זה לא היה רואה אור.

אלונה רביב

תוכן העניינים

פרק 1

מבוא להסקה סטטיסטית: מבחני תמורות

בפרק זה נביא דוגמה לבעיה סטטיסטית ובאמצעותה נסקור את הרעיונות המרכזיים של הסקה סטטיסטית. לא נציג פה בצורה פורמלית את העקרונות של בדיקת השערות סטטיסטיות, אלא נביא אותם כתזכורת בהקשר לדוגמאות שונות, וכך גם מי שלא שמע קודם שום קורס מתקדם בסטטיסטיקה יוכל להפיק תועלת מן הדברים.

דוגמה 1.1. מקבוצה של 7 חולים במחלה מסוימת נבחרה באופן מקרי קבוצה של 3 חולים ואלה קיבלו תרופה חדשה. הקבוצה השנייה שימשה כביקורת. לכל אחד מ-7 החולים נרשם אורך הזמן (חודשים) עד שקיבלו שוב התקף של המחלה.

נסמן: X_i – משך הזמן עד ההתקף אצל חולה i בקבוצת הביקורת $(i=1,2,3,4)$

Y_j – משך הזמן עד ההתקף אצל חולה j בקבוצת הטיפול $(j=1,2,3)$

התוצאות שהתקבלו מסודרות לפי גודלן (פיקטיביות):

התצפיות:	1.3	2.1	3.3	3.7	5.2	8.2	11.2
הדרגות:	1	2	3	4	5	6	7
הקבוצה:	x	x	y	x	x	y	y

בקבוצת הטיפול (ה-y-ים) נמצא החולה השלישי מבחינת הזמן עד התקף נוסף, וכן נמצאים שני החולים שאצלם הזמן ארוך מכל השאר.

מטרת המחקר הייתה לברר אם התרופה החדשה מועילה להארכת "זמן ההמתנה" להתקף נוסף של המחלה. ראשית עלינו למצוא מדד לפער בין שתי הקבוצות (קבוצת הטיפול בהשוואה לקבוצת הביקורת) מבחינת משתנה המחקר, ובהמשך נרצה לבחון אם הפער גדול דיו כדי להכריע שהתרופה אמנם יעילה.

נציע כאן ארבעה מדדים, הנקראים **סטטיסטים** (פונקציות של התצפיות):

א. $\bar{Y}-\bar{X}$ – הפרש ממוצעי התצפיות

ב. M_y-M_x – הפרש חציוני התצפיות

ג. W – סכום הדרגות של שלושת ה-Y-ים

ד. S – מספר ה-y־ים הגדולים או שווים לחציון של המדגם המצורף – 3.7

בלוח 1.1 רשומות כל 35 הבחירות האפשריות של שלוש דרגות ה-y-ים מבין הדרגות של כל 7 התצפיות, וכן ערכי ארבעת הסטטיסטים (המדדים) המתאימים לכל אחת מהאפשרויות הללו.

לוח 1.1. *רשימת כל אפשרויות הבחירה של 3 מתוך 7 דרגות אפשריות עבור קבוצת הטיפול, וערכי ארבעה סטטיסטים*

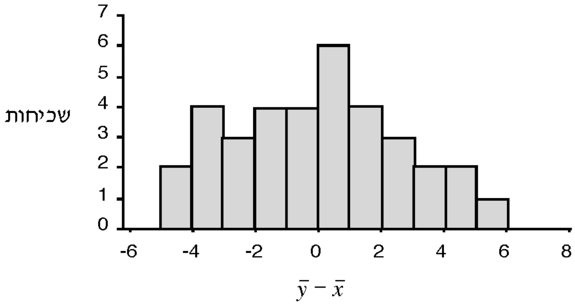
	דרגות ה-y-ים	$\bar{y}-\bar{x}$	m_y-m_x	w	s
1)	1 2 3	-4.84	-4.60	6	0
2)	1 2 4	-4.61	-4.60	7	1
3)	1 2 5	-3.73	-3.85	8	1
4)	1 2 6	-1.98	-2.35	9	1
5)	1 2 7	-0.23	-2.35	10	1
6)	1 3 4	-3.91	-3.40	8	1
7)	1 3 5	-3.03	-2.65	9	1
8)	1 3 6	-1.28	-1.15	10	1
9)	1 3 7	0.47	-1.15	11	1
10)	1 4 5	-2.80	-2.05	10	2
11)	1 4 6	-1.05	-0.55	11	2
12)	1 4 7	0.70	-0.55	12	2
13)	1 5 6	-0.18	1.70	12	2
14)	1 5 7	1.58	1.70	13	2
15)	1 6 7	3.33	4.70	14	2
16)	2 3 4	-3.44	-3.40	9	1
17)	2 3 5	-2.57	-2.65	10	1
18)	2 3 6	-0.82	-1.15	11	1
19)	2 3 7	0.93	-1.15	12	1
20)	2 4 5	-2.33	-2.05	11	2
21)	2 4 6	-0.58	-0.55	12	2
22)	2 4 7	1.17	-0.55	13	2
23)	2 5 6	0.29	1.70	13	2
24)	2 5 7	2.04	1.70	14	2
25)	2 6 7	3.79	4.70	15	2
26)	3 4 5	-1.63	-1.45	12	2
27)	3 4 6	0.12	0.05	13	2
28)	3 4 7	1.87	0.05	14	2
29)	3 5 6	0.99	2.30	14	2
30)	3 5 7	2.74	2.30	15	2
31)	**3 6 7**	**4.49**	**5.30**	**16**	**2**
32)	4 5 6	1.23	2.50	15	3
33)	4 5 7	2.98	2.50	16	3
34)	4 6 7	4.73	5.50	17	3
35)	5 6 7	5.60	5.50	18	3

הערה: התוצאה שהתקבלה בפועל מצוינת על ידי אות עבה (מס' 31).

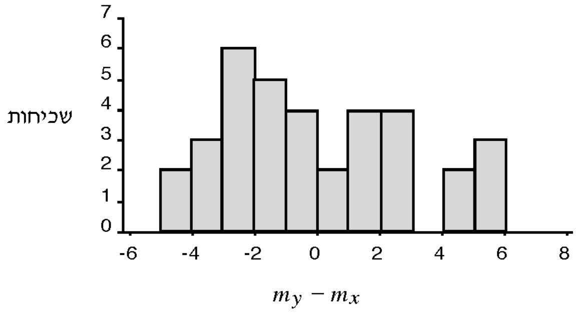
נשים לב שאם למעשה הטיפול החדש אינו עוזר כלל, הרי כל 7 החולים, למעשה, זהים ביניהם מבחינת מצבם הבריאותי ולכן לכל סדר שלהם מבחינת הזמן עד ההתקף היה אותו סיכוי להתרחש. במקרה זה, מראש ההסתברות לקבל כל אחת מ-35 התוצאות בלוח 1.1 הייתה שווה (ושווה, כמובן, ל-$1/35$). נדגיש שבניסוי שנערך התקבלה, כמובן, רק **אחת** מהתוצאות הרשומות בלוח (מסומנת באות עבה, מס' 31).

בציורים 1.1 עד 1.4 מוצגות ההיסטוגרמות של התפלגות כל אחד מהסטטיסטים.

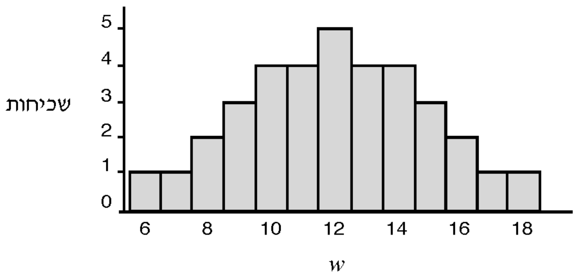
התפלגויות אלה התקבלו על ידי ספירת האפשרויות (שכיחות) לקבלת כל אחד מהערכים האפשריים של הסטטיסטי. (הקוראים מוזמנים לבדוק את הספירה.) למשל, יש 3 אפשרויות שעבורן סכום הדרגות W שווה ל-9 ורק אפשרות אחת שעבורה $W=6$. בשני הציורים הראשונים, 1.1 ו-1.2 ערכי הסטטיסטי נתונים בקבוצות ונרשם מספר התוצאות הנמצא בטווח הערכים בקבוצה. ההיסטוגרמות הללו אמנם מתארות את התפלגות הסטטיסטים, בהנחה שהתרופה אינה מועילה, באופן שלכל אחת מ-35 התוצאות האפשריות אותו סיכוי להתקבל כתוצאה של הניסוי.



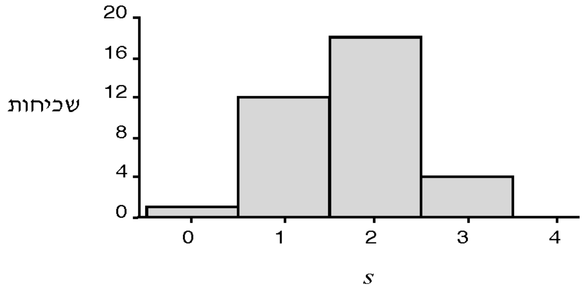
ציור 1.1. *התפלגות הפרש הממוצעים*



ציור 1.2. *התפלגות הפרש החציונים*



ציור 1.3. *התפלגות סכום הדרגות בקבוצת הטיפול*



ציור 1.4. *התפלגות מספר ערכי קבוצת הטיפול הגדולים או שווים לחציון 3.7*

בדיקת השערת המחקר

נסתכל, למשל, על הסטטיסטי הראשון – הפרש הממוצעים. במדגם של החולים שנחקרו, הערך שנמצא עבור הסטטיסטי הזה הוא: $\bar{y}-\bar{x}=7.567-3.075=4.492$ (זהו הערך הרשום עבור אפשרות מס׳ 31 בלוח 1.1). כלומר, ההפרש בין שני הממוצעים הוא די גבוה. האם המסקנה המתחייבת היא שגם באוכלוסיית החולים כולה, אילו היינו נותנים אותו טיפול **לכל החולים**, היינו מקבלים הפרש דומה בין אורך הזמן הממוצע שלהם לבין אורך הזמן הממוצע עד ההתקף הבא אילולא קיבלו את הטיפול? מובן שלא!

ננסח את הבעיה העומדת בפנינו כבעיה סטטיסטית.

כדי לבדוק אם אמנם הטיפול שניתן לחולים מועיל, נעמיד למבחן את התיאוריה הזאת כנגד **השערת האפס** שהטיפול כלל אינו מועיל. נרשום את שתי ההשערות הנבדקות זו מול זו:

השערת האפס (מסומנת H_0) : הטיפול אינו מועיל.

השערה נגדית או אלטרנטיבית (מסומנת H_1): הטיפול מועיל.

(את ההשערות הללו ניתן לרשום גם במונחים הסתברותיים, לגבי התפלגויות המשתנים X ו-Y. נעשה זאת יותר מאוחר, בפרק 2.)

מבחן סטטיסטי לבדיקת השערת האפס

השערת האפס היא ההשערה אשר עומדת למבחן באמצעות הניסוי שנערך, ועלינו להחליט, על סמך הנתונים שהתקבלו, אם מתקבל על הדעת שהשערה זו נכונה, או שהיא כנראה איננה נכונה ויש **לדחות** אותה.

מבחן לבדיקת השערת האפס הוא כלל החלטה שבו משתמשים כדי לקבוע אם יש לדחות את השערת האפס על סמך תוצאות הניסוי.

הטעויות שעלולות לקרות כתוצאה משימוש בכלל החלטה כלשהו נקראות:

טעות מסוג ראשון – הטעות הנגרמת כאשר דוחים בטעות את השערת האפס (מחליטים שהיא איננה נכונה כשהיא, למעשה, נכונה);

טעות מסוג שני – הטעות הנגרמת כאשר בטעות לא דוחים את השערת האפס (מחליטים שהיא נכונה, כשהיא, למעשה, איננה נכונה).

בבעיה המוצגת בדוגמה 1.1 טעות מסוג ראשון נגרמת אם מחליטים לאמץ את הטיפול החדש, בעוד שלמעשה אין בו כל תועלת. טעות מסוג שני נגרמת אם מחליטים שלא לאמץ את הטיפול החדש, בעוד שלמעשה יש בו תועלת.

לשם הבהרת רעיון המבחן הסטטיסטי נחזור לסטטיסטי של הפרש הממוצעים. אם הטיפול מועיל, אזי אורך הזמן עד ההתקף אצל חולים שמקבלים את הטיפול נוטה להיות "ארוך" יותר משל החולים שאינם מקבלים טיפול. מובן שבגלל ההבדלים האינדיבידואליים בין אנשים שונים, לא הגיוני לצַפות שכל אחד ואחד מהחולים שקיבלו טיפול ימתין יותר זמן מכל אחד ואחד מאלה שלא קיבלו טיפול. אולם ניתן להניח שאם הטיפול מועיל, אזי המשתנה Y (אורך הזמן של אדם מקרי שקיבל טיפול) **נוטה** לקבל ערכים גבוהים יותר מן המשתנה X (אורך הזמן של אדם מקרי מקבוצת הביקורת). למשל, התוחלת של Y גבוהה מזו של X, או החציון של Y גבוה מזה של X. בהנחה זו, נצַפה שאם אכן הטיפול מועיל, אזי הערך של הפרש הממוצעים $\bar{Y}-\bar{X}$ ייטה להיות גדול יחסית, לעומת המצב שבו אין הבדל בין התפלגויות המשתנים X ו-Y (כלומר, במקרה שהטיפול אינו מועיל). בהתאם לכך נקבע שככל שהערך של ההפרש $\bar{Y}-\bar{X}$ גדול, כך יש בידינו ראָיָה יותר מוצקה לכך שהטיפול אמנם מועיל.

מובהקות התוצאה (*P*-value)

המובהקות של התוצאה (מסומנת P) היא ההסתברות לקבלת תוצאה **קיצונית** לפחות כמו זו שהתקבלה בניסוי, בהנחה שלמעשה השערת האפס נכונה. הקיצוניות של התוצאה נקבעת בכיוון הנדרש לשם אישוש התיאוריה הרשומה כַּהשערה האלטרנטיבית.

בדוגמה שהצגנו, מובהקות התוצאה מתקבלת על ידי ההסתברות לכך שהפרש הממוצעים $\bar{Y}-\bar{X}$ יהיה **גדול** לפחות כמו הערך 4.49 (הערך שהתקבל בניסוי), אם, למעשה, הטיפול איננו מועיל.

אנו רושמים את מובהקות התוצאה: $P=P_{H_0}(\bar{Y}-\bar{X}\geq 4.49)$.

לפני שנחשב את המובהקות הזאת, נעיר שערך זה מבטא את הסיכוי לכך שבַּניסוי תתקבל באופן מקרי תוצאה כל כך קיצונית (או חריגה), אם למעשה אין כל הבדל בין התפלגויות המשתנים X ו-Y. בדוגמה שלנו, ערך זה הוא הסיכוי לכך שלקבוצת הניסוי ייבחרו במקרה דווקא החולים שעבורם זמן ההמתנה עד ההתקף הוא יחסית גבוה, באופן שהפרש הממוצעים בין קבוצת הטיפול לקבוצת הביקורת יהיה גבוה מ-4.49 חודשים.

חישוב מובהקות התוצאה במקרה זה קל מאוד. בהנחה שהשערת האפס נכונה, הרי לכל

אחת מ-35 התוצאות האפשריות הרשומות בלוח 1.1 אותה הסתברות. בלוח 1.1 רואים שישנן בסך הכל שלוש אפשרויות שבהן התוצאה של $\bar{Y}-\bar{X}$ גבוהה לפחות כמו זו שהתקבלה (מס' 31, 34 ו-35). מכאן מובהקות התוצאה היא

$$P=P_{H_0}(\bar{Y}-\bar{X}\geq 4.49)=\frac{3}{35}\cong 0.086$$

קיבלנו, אפוא, מובהקות של כ-8.6%.

במִלים אחרות, אם למעשה הטיפול איננו מועיל כלל, יש סיכוי של כ-8.6% לכך שבמדגם מקרי של 3 מתוך 7 החולים שנבחרו להשתתף בניסוי, ממוצע אורך הזמן של קבוצת הטיפול יעלה על זה של קבוצת הביקורת ב-4.49 חודשים לפחות.

מסקנת בדיקת ההשערות – הסקה סטטיסטית

כיצד קובעים אם נמצאה ראָיָה לכך שהטיפול מועיל? כאשר מובהקות התוצאה קטנה, ניתן להסיק שהתוצאה שהתקבלה היא "חריגה", באופן שהסיכוי לקבלת תוצאה כה קיצונית הוא קטן. כלומר, אם השערת האפס הייתה נכונה, לא סביר היה לקבל תוצאה כזאת מלכתחילה. במקרה כזה יש בידינו ראָיָה מוצקה דיה לכך שהשערת האפס כנראה איננה נכונה ולכן ההשערה האלטרנטיבית שהעמדנו כנגדה מתארת באופן מתאים יותר את המצב האמיתי.

אימתי נחליט שהראיה מוצקה דיה? או, לחלופין, מתי מובהקות התוצאה היא קטנה מספיק? מובן שעניין זה הוא סובייקטיבי לחלוטין ותלוי בסוג המחקר ובנזקים שעלולים להיגרם מטעויות בהחלטה.

רמת המובהקות של מבחן

רמת המובהקות (המסומנת ב-α), היא הערך הגבוה ביותר של מובהקות התוצאה P, שעבורו מחליטים שיש בתוצאת הניסוי ראָיָה כנגד השערת האפס. כלומר, כל ערך של P הקטן מהגודל α ייקרא "קטן מספיק" כדי להוות ראָיָה **כנגד** נכונותה של השערת האפס (ולפיכך תמיכה בהשערה האלטרנטיבית).

רמת המובהקות היא גם ההסתברות לדחיית השערת האפס בטעות, או ההסתברות ל**טעות מסוג ראשון**.

הערך המקובל של רמת המובהקות של מבחנים הוא $\alpha=0.05$, אך ייתכנו מקרים שבהם ניתן להסתפק ברמת מובהקות $\alpha=0.10$, ומקרים אחרים שבהם נדרוש רמת מובהקות קטנה יותר, למשל $\alpha=0.01$.

נעיר כאן כי מקובל לסמן את ההסתברות לטעות מסוג שני ב-β. נדון בחישוב הסתברויות לטעות מסוג שני יותר מאוחר, בפרק 2.7.

מבחן לבדיקת השערות

מבחן לבדיקת השערת האפס ניתן, אם כך, על ידי הכלל הבא.

יש לחשב את מובהקות התוצאה שהתקבלה בניסוי – P. בהינתן רמת מובהקות α:

אם התקבלה תוצאה שעבורה $P \leq \alpha$, יש **לדחות** את השערת האפס;

במקרה שהתקבלה תוצאה שעבורה $P > \alpha$, אין לדחות את השערת האפס.

בַדוגמה לעיל, כאשר השתמשנו בהפרש הממוצעים, קיבלנו $P = 0.086$. לפיכך, אם החלטנו להשתמש ברמת מובהקות של 5% ($\alpha = 0.05$), לא נוכל לדחות את השערת האפס, כיוון ש- $P > \alpha$. לכן נסיק כי על פי הנתונים שהתקבלו בניסוי אין רְאָיָה מוצקה דיה לכך שהטיפול אכן מועיל. במלים אחרות, יש סיכוי סביר, "לא קטן" (כ-8.6%), לקבלת הפרש גדול לפחות כמו זה שהתקבל, גם אם אין כלל הבדל בין התפלגות אורך הזמן עם טיפול לבין אורך הזמן ללא כל טיפול. אנו אומרים שהסיכוי הוא "לא קטן", כיוון שהוא עולה על רמת המובהקות α שנקבעה מראש.

מובן שאם היינו מחליטים להשתמש ברמת מובהקות של 10% ($\alpha = 0.10$), אזי $P < \alpha$ ולכן במקרה זה ניתן היה להסיק שיש בתוצאות הניסוי רְאָיָה לכך שהטיפול מועיל (דוחים את השערת האפס לטובת ההשערה האלטרנטיבית).

אפשרות השימוש בסטטיסטים אלטרנטיביים

בדוגמה לעיל השתמשנו בסטטיסטי של הפרש הממוצעים כדי לבדוק את ההשערות. אנו אומרים במקרה זה שהפרש הממוצעים הוא **סטטיסטי המבחן**.

נסתכל עתה על **הסטטיסטי השני** – הפרש החציונים של שתי הקבוצות. בלוח 1.1 רואים שהתוצאה שהתקבלה בניסוי (מס' 31) היא $M_y - M_x = 5.30$ וכמו לגבי הפרש הממוצעים, גם פה ישנן בסך הכל שתי אפשרויות נוספות שבהן התוצאה עולה על זו (מס' 34 ו-35). לכן גם במקרה זה מתקבלת מובהקות התוצאה כמו קודם:

$$P_M = P_{H_0}(M_y - M_x \geq 5.30) = \frac{3}{35} \cong 0.086$$

לגבי **הסטטיסטי השלישי** – סכום הדרגות, ניתן לראות שהתקבלה התוצאה $W = 16$ וישנן בסך הכל 4 אפשרויות לתוצאה גבוהה לפחות כמו זו (מס' 31, 33, 34 ו-35). מכאן שמובהקות התוצאה היא

$$P_W = P_{H_0}(W \geq 16) = \frac{4}{35} \cong 0.114$$

בניגוד למסקנה המבוססת על שני הסטטיסטים הראשונים, במקרה זה גם אם היינו מסתפקים ברמת מובהקות של 10% לא היינו יכולים לדחות את השערת האפס.

הסטטיסטי הרביעי נותן תוצאה בעלת מובהקות גדולה עוד יותר. מספר הקומבינציות

שעבורן $S \geq 2$ (מס' 31) הוא גדול מאוד. ספירה קפדנית מראה שמובהקות התוצאה במקרה זה היא

$$P_U = P_{H_0}(S \geq 2) = \frac{22}{35} \cong 0.629$$

כלומר, יש סיכוי העולה על 60% לקבלת ערך של S הגדול לפחות כמו זה שהתקבל בניסוי.

מן הדוגמה הזאת ברור שבחירת הסטטיסטי שישמש לבדיקת ההשערות העומדות למבחן, מהווה כלי חשוב מאוד בתהליך בדיקת השערות.

הערה: העובדה שעבור תוצאות הניסוי הנוכחי המובהקות של הסטטיסטי הרביעי הייתה כל כך גבוהה ביחס לסטטיסטים האחרים, אינה מחייבת שכך יקרה תמיד. אם, למשל, תוצאת הניסוי הייתה האפשרות מס' 32 בלוח 1.1, היינו מקבלים את המובהקויות הבאות (בדקו!):

א. הפרש הממוצעים: $P_{\bar{Y}-\bar{X}} = P_{H_0}(\bar{Y}-\bar{X} \geq 1.23) = \frac{11}{35}$

ב. הפרש החציונים: $P_M = P_{H_0}(M_y - M_x \geq 2.50) = \frac{7}{35}$

ג. סכום הדרגות: $P_W = P_{H_0}(W \geq 15) = \frac{7}{35}$

ד. מספר ה-y-ים הגדולים או שווים ל-3.7: $P_S = P_{H_0}(S \geq 3) = \frac{4}{35}$

במקרה של תוצאה כזאת המובהקות הקטנה ביותר מתקבלת דווקא כאשר משתמשים בסטטיסטי הרביעי.

הכללה: מבחן תמורות (פרמוטציות)

המבחן שבו השתמשנו בדוגמה 1.1 נקרא **מבחן תמורות**. ברגע שקבענו סטטיסטי מתאים כסטטיסטי המבחן, ניתן לחשב את **כל** התמורות השונות של אותם נתונים ספציפיים שהתקבלו בניסוי (למשל, על ידי קביעה אילו מן התצפיות היו יכולות להיות שייכות לקבוצת הניסוי ואילו לביקורת). אם השערת האפס נכונה, לכל אחת מן הקומבינציות הללו יש אותה הסתברות. לכן, לשם חישוב מובהקות התוצאה יש צורך לספור את הקומבינציות שעבורן ערך הסטטיסטי קיצוני לפחות כמו הערך שהתקבל בניסוי. מובהקות התוצאה היא המנה בין המספר הזה לבין מספר התמורות הכולל.

עבור בעיית שתי קבוצות – ניסוי וביקורת, כמו בדוגמה 1.1, אם בקבוצת הטיפול n תצפיות ובקבוצת הביקורת m תצפיות, מספר כל הקומבינציות האפשריות הוא כמספר

האפשרויות לבחור את n התצפיות לטיפול, מתוך $m + n$ הנבדקים העומדים לרשותנו, כלומר, $\binom{m+n}{n}=\frac{(m+n)!}{m!n!}$.

הבעיה במבחן זה היא שרישום כל התמורות האפשריות, חישוב סטטיסטי המבחן לכל אחת מהן וספירתן אינו קשה במיוחד עבור מדגמים קטנים (וגם אז זוהי טרחה בלתי מבוטלת...). לעומת זאת, במדגמים גדולים יותר הטרחה רבה מאוד, ולמעשה החישוב של מובהקות התוצאה הוא כמעט בלתי אפשרי ללא שימוש בתכנת מחשב מתאימה. למשל, אם בַניסוי שני מדגמים בגודל 10 כל אחד (וזה נחשב ניסוי קטן ביותר!), מספר כל הקומבינציות האפשריות הוא גדול מאוד – $\binom{20}{10}=\frac{(20)!}{10!10!}=184{,}756$.

מטעם זה לא נהוג בדרך כלל להשתמש במבחני תמורות.

שמירה על רמת מובהקות – מבחן אפרמטרי

היתרון של מבחן תמורות (לכל סטטיסטי מבחן שבו משתמשים) הוא, שניתן לחשב את מובהקות התוצאה **בדיוק**, והחישוב איננו תלוי כלל בהתפלגות של האוכלוסייה שממנה נבחר המדגם. למשל, בדוגמה 1.1 שהצגנו לעיל, לא חשוב כלל מהי התפלגות זמן ההמתנה של החולים; אם החלוקה לקבוצות ניסוי וביקורת הייתה חלוקה אקראית, ובתנאי שלגבי כל אחד מהחולים אין הבדל אם הוא קיבל טיפול או שלא זכה לקבל טיפול (זמן ההמתנה שלו "נקבע מראש" ואינו תלוי בטיפול), אזי מובהקות התוצאה, כפי שחושבה לעיל עבור כל אחד מהסטטיסטים, היא מדויקת.

נזכיר שכלל ההחלטה לגבי בדיקת ההשערות נקבע אך ורק כפונקציה של מובהקות התוצאה – P:

אם $P\leq\alpha$, יש לדחות את השערת האפס, ואם $P>\alpha$ אין דוחים אותה.

בגלל העובדה שניתן לחשב את המובהקות בלי כל ידיעה על התפלגות התצפיות, מבחן כזה נקרא **מבחן אפרמטרי** (באנגלית Nonparametric או Distribution free). המונח "אפרמטרי" בא מן העובדה שאין צורך להניח כל הנחה לגבי מודל פרמטרי של התצפיות (כמו, למשל, התפלגות נורמלית, מעריכית, וכד'), אלא רק את נכונותה של השערת האפס.

תרגילים

1. הסתכלו על נתוני דוגמה 1.1 להשוואת שני מדגמים. נגדיר את סטטיסטי המבחן על ידי הסטטיסטי $T = \max\{Y_1, Y_2, Y_3\} - \max\{X_1, X_2, X_3, X_4\}$.

חשבו את הערך של הסטטיסטי במדגם, עִרכו מבחן פרמוטציות ותנו את מובהקות התוצאה שהתקבלה לגבי הסטטיסטי T.

2. להשוואת שלושה טיפולים לשם הורדה במשקל, חולקו אקראית 6 גברים בין הטיפולים. התוצאות הרשומות בלוח הן אחוזי הירידה במשקל של הגברים שנבדקו.

הטיפול		
א	ב	ג
19.6	21.7	16.1
17.8	19.3	17.3

א) נסמן ב-$\overline{X}_i$ את ממוצע המדגם ה-i, $i = 1,2,3$, וב-$\overline{X}.$ את הממוצע של כל 6 התצפיות. נגדיר סטטיסטי לבדיקת ההבדל בין שלושת הטיפולים:

$$S = (\overline{X}_1 - \overline{X}.)^2 + (\overline{X}_2 - \overline{X}.)^2 + (\overline{X}_3 - \overline{X}.)^2$$

השתמשו במבחן תמורות המבוסס על הסטטיסטי S, כדי לבדוק אם יש הבדל בין שלושת הטיפולים. קבעו את מובהקות התוצאה.

(רמז: בסך הכל ישנן רק 15 אפשרויות **שונות** של תמורות!)

ב) מה המסקנה עבור רמת מובהקות $\alpha = .10$? ועבור $\alpha = .05$?

פרק 2

מבחן ווילקוקסון לשני מדגמים

2.1 ההשערות בבעיית שני מדגמים

בפרק 1 ראינו דוגמה לבעיה סטטיסטית שבה יש לבחון את ההבדל בין קבוצת טיפול לקבוצת ביקורת. נחזור שנית לדוגמה 1.1 ונשים לב שישנו הבדל עקרוני חשוב מאוד בין ארבעת הסטטיסטים שהוצעו עבור המבחן. לגבי הסטטיסטי הראשון (הפרש הממוצעים) והשני (הפרש החציונים) ברור שהתפלגות כל הערכים האפשריים של הסטטיסטי, הנתונה בציורים 1.1 ו-1.2, תלויה באופן מאוד חזק בתוצאות הספציפיות שהתקבלו בניסוי. לכל אוסף של 7 נתוני הזמן עד ההתקף, התפלגות הפרשי הממוצעים או הפרשי החציונים תהיה שונה (ומיוחדת למדגם זה).

לעומת זאת, לגבי הסטטיסטי השלישי (סכום הדרגות), קל לראות שהתפלגותו אינה תלויה כלל בתוצאות הספציפיות של הניסוי. אותו דבר נכון גם לגבי הסטטיסטי הרביעי, אך ייתכן שהדבר אינו כה מובן מאליו, ונתייחס לכך מאוחר יותר (פרק 3).

בכל מקרה שהתפלגות סטטיסטי המבחן איננה תלויה בתוצאות הספציפיות שהתקבלו, ניתן להכין מראש את התפלגות הסטטיסטי בהנחה שהשערת האפס נכונה, ולהשתמש בה לשם ניתוח תוצאות הניסוי.

אנשים שונים הציעו סטטיסטים כאלה לבעיות שונות, ובכך אִפשרו ניתוח של נתונים בלי הזדקקות לרישום כל התמורות. אחד המבחנים הללו הוא המבחן של ווילקוקסון לשני מדגמים (Wilcoxon two-sample test). ווילקוקסון (Wilcoxon, 1945) וכן מאן וּוויטני (Mann & Whitney, 1947) הציעו את המבחן שאותו נביא כאן.

נציג את בעיית שני המדגמים באופן כללי.

נתון מדגם של N נבדקים, שמהם n נבחרו אקראית לקבלת טיפול ו-m הושארו כביקורת ($N=m+n$).

נסמן: X_i – התצפית ה-i בקבוצת הביקורת $(i=1,2,\dots,m)$ בעלת התפלגות F;

Y_j – התצפית ה-j בקבוצת הטיפול $(j=1,2,\dots,n)$ בעלת התפלגות G.

נניח ששתי ההתפלגויות F ו-G רציפות.

השערת האפס שאין הבדל בין טיפול לביקורת ניתנת להירשם:

$$H_0: F(t)=G(t) \quad \text{לכל } t,$$

בכל הדיון כאן נניח שההשערה האלטרנטיבית טוענת שערכי המשתנה Y, שהתפלגותו G, נוטים להיות גבוהים מערכי המשתנה X, שהתפלגותו F (כלומר, שהטיפול מעלה את תוצאת הבדיקה). נשים לב שההשערה האלטרנטיבית אינה כה פשוטה להצגה כמו השערת האפס. ניתן לרשום, למשל, שהתוחלת של Y גבוהה מן התוחלת של X– $H_1': EY>EX$, או שיחס דומה קיים לגבי שני החציונים – $H_1'': M_Y>M_X$.

צורה אחרת לרשום את ההנחה שערכי Y נוטים להיות גבוהים מערכי X היא, ש**כל** ערכי החלוקה של Y גדולים מאלה של X, בהתאמה, כלומר: $y_p \geq x_p$ לכל $0<p<1$.

[ערך החלוקה x_p, נקרא גם האחוזון ה-p של X, הוא הערך המקיים: $P(X\leq x_p)=p$.]

2.2 המבחן של ווילקוקסון

המבחן מיועד לבדיקת ההשערה שהתפלגות המשתנה X שווה להתפלגות המשתנה Y, כנגד האלטרנטיבה ש-Y נוטה להיות גדול מ-X.

נסדר את כל $N=m+n$ התצפיות הנתונות לפי הגודל, מהקטנה לגדולה. לכל תצפית נרשום את הדרגה שלה בסדר שהתקבל: התצפית הקטנה ביותר מקבלת את הדרגה 1, השנייה בגודלה את הדרגה 2, וכך הלאה, כאשר התצפית הגדולה ביותר מקבלת את הדרגה N.

נסמן את הדרגות המתאימות:

$R_1,R_2,\dots,R_m$ הן הדרגות של $X_1,X_2,\dots,X_m$ בהתאמה.

$S_1,S_2,\dots,S_n$ הן הדרגות של $Y_1,Y_2,\dots,Y_n$ בהתאמה.

הדרגות הללו הן כולן משתנים מקריים, התלויים בסדר שהתקבל בין N התצפיות של הניסוי.

סטטיסטי המבחן של ווילקוקסון הוא סכום הדרגות בקבוצת הטיפול, כלומר,

$$W_s=S_1+S_2+\cdots+S_n=\sum_{i=1}^{n}S_i \qquad (1)$$

(נזכיר שזה היה הסטטיסטי השלישי שהוצע בדוגמה 1.1.)

היות שההשערה האלטרנטיבית טוענת שהמשתנה Y נוטה לקבל ערכים גבוהים יותר מאשר המשתנה X, נצַפה לקבל ערך גבוה של W_s אם למעשה האלטרנטיבה נכונה. לכן

אנו דוחים את השערת האפס עבור ערכים גבוהים מספיק של הסטטיסטי W_s. כלומר, מובהקות התוצאה תחושב באופן הבא: $P=P(W_s \ge w)$, כאשר w הוא הערך של סכום הדרגות שהתקבל בניסוי.

דוגמה 2.1 (המשך דוגמה 1.1). בדוגמה 1.1 נתון מדגם של $N=7$ חולים, שמהם $n=3$ הוקצו לטיפול ו-$m=4$ לביקורת. הדרגות של קבוצת הטיפול הן 3, 6, ו-7 ולפיכך סכום הדרגות שהתקבל הוא $W_s=3+6+7=16$.
מובהקות התוצאה חושבה שם בעזרת לוח 1.1, שבו רשומות כל התוצאות האפשריות של סדר זמן ההמתנה להתקף של 7 החולים. היות שתחת המודל של השערת האפס (אין הבדל בין טיפול לביקורת) לכל סידור כזה אותה הסתברות, מצאנו כי מובהקות התוצאה היא $P=P_{H_0}(W_s \ge 16)=4/35 \cong 0.114$.
את רישום כל התוצאות האפשריות של הדרגות ניתן להכין מראש, ללא ידיעה על תוצאות הניסוי, ולכן ניתן להכין את טבלת ההתפלגות של W_s תחת השערת האפס, עבור ערכי m ו-n שונים. לצערנו, אין נוסחה סגורה לחישוב התפלגות זו.

משפט 2.1. בהנחה שהשערת האפס נכונה (התפלגויות הטיפול והביקורת שוות), פונקציית ההסתברות של סכום הדרגות W_s ניתנת על ידי:

$$P_{H_0}(W_s=w)=\frac{\#\{W_s=w\}}{\binom{N}{n}} \qquad (2)$$

המכנה $\binom{N}{n}$ הוא המספר הכולל של אפשרויות הבחירה של n דרגות לקבוצת הטיפול מתוך N הדרגות האפשריות, והמונה $\#\{W_s=w\}$ הוא מספר האפשרויות מתוכן, שבהן קיים: $W_s=w$.
הוכחה: כבר הראינו בפרק 1 את צורת החישוב הזאת במבחני תמורות. תחת השערת האפס כל אחד מהסידורים השונים של דרגות התצפיות הוא בעל אותה הסתברות. ההנחה במקרה זה היא שערכי התצפיות עצמן כבר נקבעו מראש, ואקראיות הבחירה של קבוצת הטיפול היא שקובעת מה יהיה ערכו של סטטיסטי המבחן כתוצאה מהניסוי.

♣

יש לשים לב שחישוב מספר האפשרויות $\#\{W_s=w\}$ בנוסחה (2) עבור ערכי w שונים מהווה טרחה רבה כבר עבור מדגמים קטנים יחסית, כפי שאפשר לראות בדוגמה 2.2.

דוגמה 2.2. נביא לדוגמה את החישוב של חלק מהתפלגות W_s עבור $m=5,\ n=4$, כאשר נראה כאן כיצד ניתן לספור את האפשרויות לקבלת הערכים השונים של W_s.

בלוח 2.1 הדרגות של ה-y-ים מסומנות ב-*. התחלנו את הספירה מן הדרגות הנמוכות ביותר. בכל שלב נרשמו האפשרויות (אחת או יותר) להעלות את סכום הדרגות ביחידה אחת, על ידי העלאת אחת הדרגות מהשלב הקודם ב-1 (הזזת אחת הכוכביות צעד אחד ימינה).

לוח 2.1. *רשימת הקומבינציות של דרגות ה-y-ים עבור $m=5,\ n=4$, לחישוב התפלגות W_s (ערכים 10-13).*

אפשרות	הדרגות 1	2	3	4	5	6	7	8	9	W_s
1	*	*	*	*						10
2	*	*	*		*					11
3	*	*	*			*				12
4	*	*		*	*					12
5	*	*	*				*			13
6	*	*		*		*				13
7	*		*	*	*					13

* דרגות ה-y-ים

באופן זה ניתן להמשיך ולקבל את כל צירופי הדרגות האפשריים. מספר כל הצירופים האפשריים הוא $\binom{9}{4}=126$. מן הטבלה ניתן למצוא עבור המקרה הנידון את ההסתברויות:

$$P(W_s=11)=\frac{\#\{W_s=11\}}{126}=\frac{1}{126} \qquad P(W_s=10)=\frac{\#\{W_s=10\}}{126}=\frac{1}{126}$$

$$P(W_s=13)=\frac{\#\{W_s=13\}}{126}=\frac{3}{126} \qquad P(W_s=12)=\frac{\#\{W_s=12\}}{126}=\frac{2}{126}$$

היות שאין נוסחה סגורה לחישוב מספר האפשרויות של ערכי W_s השונים, הוכנו טבלאות מתאימות והן קיימות בספרים הדנים בשיטות אפרמטריות (למשל Lehmann, 1975; Hollander & Wolfe, 1973). (את הטבלאות שהבאנו בספר זה הכנו למעשה על סמך שיטת החישוב המודגמת בתרגיל 5, שבה אפשר למצוא את ההסתברות של כל ערך על סמך הסתברויות של ערכים נמוכים יותר.) במקומות שונים הטבלאות ניתנות בצורה שונה. בספר זה מובאת טבלת ההתפלגות המצטברת של המשתנה W_s (טבלה 2 בנספח) עבור m ו-n הקטנים מ-10.

תחום ערכי W_s נע בין הערך המינימלי $1+2+\cdots+n=n(n+1)/2$

לבין הערך המקסימלי שהוא $N(N+1)/2-m(m-1)/2$ (תרגיל 1).
כדי לחסוך מקום, בטבלה נתונות ההסתברויות רק עבור המקרים שבהם $n\le m$. למשל, את מובהקות התוצאה $W_s=16$ שהתקבלה בדוגמה 1.1 $(m=4,\ n=3)$, ניתן למצוא בטבלה באופן הבא:

$$P=P\ (W_s\ge 16)=1-P(W_s\le 15)=1-.8857=.1143$$

ההסתברות $P(W_s\le 15)=.8857$ רשומה בטבלה 2, בעמודה שכותרתה $n=3,\ m=4$.
כל ההסתברויות כאן מחושבות תחת השערת האפס.
מציאת ההסתברות המתאימה כאשר $n>m$ איננה מסובכת. נסתכל על דוגמה 2.3.

דוגמה 2.3. הנתונים להלן הם חלק ממחקר שנערך בישראל (עירית דר) כדי לבדוק את הנושא של אלימות בסביבות שונות, כפי שמדווחים ילדים צעירים. מובאים כאן ציוני מדד האלימות בבית הספר כפי שדיווחו 9 בנות ו-10 בנים, הלומדים בכיתה ב. נרצה לבדוק אם בבית הספר בנים עדים למקרי אלימות יותר מאשר בנות.

לוח 2.2. *ציוני מדד אלימות בבית הספר אצל 9 בנים ו-10 בנות.*

	בנות		בנים	
	דרגה	אלימות	דרגה	אלימות
	1	0.82	12	2.09
	13	2.18	8	1.82
	2	1.00	7	1.73
	6	1.64	18	2.55
	5	1.55	15	2.36
	17	2.50	10	1.91
	11	2.00	4	1.36
	9	1.83	16	2.36
	3	1.17	19	2.73
			14	2.33
סך הכול	67		123	

נסמן ב-X את ערכי מדד האלימות אצל הבנות וב-Y את ערכי הבנים. גודלי המדגמים הם: $m=9,\ n=10$.
מלוח 2.2 רואים שהסטטיסטי של ווילקוקסון המתקבל כאן (סכום הדרגות אצל הבנים) הוא $W_s=123$. מובהקות התוצאה היא אפוא: $P=P(W_s\ge 123)$. בטבלת התפלגות ווילקוקסון (טבלה 2 בנספח) לא נמצא את ההסתברות הזאת, כיוון שכאן $n>m$.

נסתכל, אם כך, על סכום הדרגות אצל הבנות: $W_r=67$. ברור שהסיכוי לכך שסכום הדרגות של הבנים גדול מדי שווה לסיכוי שסכום הדרגות של הבנות קטן מדי.

מובהקות התוצאה היא, אפוא $P=P(W_s\geq 123)=P(W_r\leq 67)=.0326$. הערך $P=.0326$ נמצא בטבלה 2 בנספח, בעמודה שכותרתה $n=9,\ m=10$. עבור רמת מובהקות של $.05$ יש לדחות את השערת האפס ולהסיק שבנים אמנם עדים לאלימות בבית הספר יותר מאשר בנות.

כדי למצוא את מובהקות התוצאה במקרה שהערך הדרוש אינו מופיע בטבלה, לא עבור דרגות ה-x-ים וגם לא עבור דרגות ה-y-ים, ניתן להיעזר בסימטרייה של W_s. תכונה זו מובאת במשפט הבא.

משפט 2.2. תחת השערת האפס התפלגות המשתנה W_s היא סימטרית סביב $\frac{n(N+1)}{2}$.

הוכחה: כדי להוכיח את המשפט, יש להראות שעבור כל ערך t, קיים:

$$P\left\{W_s=\frac{n(N+1)}{2}+t\right\}=P\left\{W_s=\frac{n(N+1)}{2}-t\right\}$$

כאשר ההסתברויות מחושבות תחת השערת האפס (שוויון התפלגויות).

נדרג את N התצפיות בסדר הפוך – מהגדולה ביותר (דרגה 1) עד לקטנה ביותר (דרגה N). נסמן את דרגות ה-y-ים המתאימות על ידי S_j^*, כאשר $S_j^*=N+1-S_j$. (בִדקו שזה אמנם הקשר בין הדרגות המקוריות לדרגות ה"הפוכות".)

סכום הדרגות "ההפוכות" של ה-y-ים הוא

$$W_s^*=\sum_{j=1}^{n}S_j^*=\sum_{j=1}^{n}(N+1-S_j)=n(N+1)-\sum_{j=1}^{n}S_j=n(N+1)-W_s$$

מצד שני, תחת השערת האפס, הסתברות כל תמורה של N הדרגות היא שווה, ולכן לסכום הדרגות "ההפוכות" של ה-y-ים $W_s^*=\sum_{j=1}^{n}S_j^*$ יש אותה התפלגות כמו לסכום הדרגות המקוריות $W_s=\sum_{j=1}^{n}S_j$. שימו לב שסכום הדרגות המקוריות הוא, למשל, $n(n+1)/2$ כאשר ה-y-ים מקבלים את n הדרגות הנמוכות ביותר. אולם באותו אופן סכום הדרגות ה"הפוכות" הוא $n(n+1)/2$ כאשר ה-y-ים מקבלים את n הדרגות הגבוהות ביותר. הסיכוי לכל אחת מהתוצאות הללו הוא זהה (אפשרות אחת בלבד לבחירת n הדרגות הללו). מכאן מקבלים:

$$P\left\{W_s=\frac{n(N+1)}{2}+t\right\}=P\left\{W_s^*=\frac{n(N+1)}{2}+t\right\}$$

$$=P\left\{n(N+1)-W_s=\frac{n(N+1)}{2}+t\right\}$$

$$=P\left\{W_s=\frac{n(N+1)}{2}-t\right\}$$

זו בדיוק דרישת הסימטרייה. ♣

מסקנה 2.1. את הסתברות הזנב ניתן למצוא בעזרת הסימטרייה לפי

$$P\left\{W_s\geq\frac{n(N+1)}{2}+t\right\}=P\left\{W_s\leq\frac{n(N+1)}{2}-t\right\}$$

ניצול הסימטרייה של W_s מאפשר לחסוך מקום בטבלת ההתפלגות של W_s. נמצא, למשל, את ההסתברות $P(W_s\geq 62)=1-P(W_s\leq 61)$, כאשר $m=10,\ n=6$. הערך 61 אינו מופיע בטבלה. נשתמש, אפוא, בתכונת הסימטרייה של W_s. מרכז הסימטרייה של ההתפלגות הוא $n(N+1)/2=6\cdot 17/2=51$. לפיכך הערך הסימטרי של $62=51+11$ הוא $51-11=40$.

ההסתברות הדרושה היא, אפוא: $P=P(W_s\geq 62)=P(W_s\leq 40)=.1317$.

ההסתברות $.1317$ נמצאת בטבלה 2 בנספח, תחת הכותרת $n=6,\ m=10$.

הערה: בכל הפרק הזה אנו מניחים שמשמעות העובדה שהטיפול יעיל מהביקורת היא שערכי התצפיות תחת הטיפול נוטים להיות גבוהים יותר מהערכים של נבדקים שלא טופלו. כלומר, אנו מניחים שיש לדחות את השערת האפס עבור ערכים **גבוהים** של סטטיסטי המבחן W_s (שהוא סכום הדרגות של ה-Y-ים, או, דרגות הנבדקים בקבוצת הטיפול). אם יעילות הטיפול מתבטאת **בירידה** של הערכים הנמדדים, אזי ניתן לתקן זאת באחת משתי צורות:

א. ניתן להחליף את התפקיד של X ו-Y ולהגדיר את ה-X-ים כתצפיות הטיפול ואת ה-Y-ים כביקורת. בצורה כזאת כל שאר ההגדרות והסימונים לגבי סטטיסטי המבחן, אזור הדחייה וחישוב מובהקות התוצאה נשארים כמו קודם.

ב. אם משאירים את אותם הסימונים של שני המדגמים כרגיל, יש לדחות את השערת האפס עבור ערכים נמוכים של W_s, ובהתאם לכך לחשב את מובהקות התוצאה שהתקבלה על ידי הזנב השמאלי.

התיקון א עדיף על ב, מכיוון שבמקרה זה ניתן להשתמש בכל התוצאות הקודמות ללא כל שינוי, ובכך נמנע עצמנו מבלבול. בהתאם לכך, הדבר הפשוט ביותר הוא לסמן ב-X את המשתנה הצפוי (תחת השערת המחקר האלטרנטיבית) לקבל את הערכים הנמוכים

וב-Y את המשתנה הצפוי לקבל את הערכים הגבוהים יותר.

אנו נמשיך ונקרא למדגם ה-X-ים בשם "קבוצת הביקורת" ולמדגם ה-Y-ים נקרא "קבוצת הטיפול", ללא קשר לבעיה המחקרית הספציפית.

השערה דו־צדדית

עד עתה דנו רק במקרים שבהם ההשערה האלטרנטיבית הייתה חד־כיוונית, דהיינו, השערת המחקר טוענת לכיוון מסוים של הערכים הצפויים של התפלגות מדגם ה"טיפול" בהשוואה להתפלגות ה"ביקורת". נסתכל עתה על דוגמה לבעיה שבה אין השערה כיוונית.

דוגמה 2.4. במחקר שמטרתו לברר את מידת הסמכות של מורים בעיני תלמידיהם (ברכה בירן), נשאלו תלמידים על מידת סמכות הידע שהם מייחסים לכל אחד מ-4 מורים. השאלון כלל 9 פריטים, ביניהם "יש לו ידע רב", או "הוא מקפיד לדייק בעובדות". התשובות לכל פריט ניתנו על סולם בין 1 ל-6. כמו כן, השאלון התייחס לשני תחומים: התחום המקצועי שהמורה מלמד, ותחומים אחרים. השאלונים הועברו בכיתות ז ובכיתות י. הציון של תלמיד חושב על ידי ממוצע כל 9 הפריטים בשאלון.

לוח 2.3. *ממוצעים כיתתיים של סמכות ידע של מורים למתמטיקה בתחומים האחרים*

כיתות י		כיתות ז	
דרגה	סמכות (Y)	דרגה	סמכות (X)
1	2.89	2	3.26
3	3.28	5	3.32
4	3.30	8	3.74
6	3.57	9	3.86
7	3.65	10	3.89
11	3.94	12	4.02
13	4.09	14	4.17
17	4.85	15	4.23
		16	4.81
62		91	סך הכול

בלוח 2.3 נתונים חלקיים של המחקר, רק לגבי המורה למתמטיקה בתחומים האחרים (באופן כללי), עבור 9 כיתות ז ו-8 כיתות י. הנתונים בלוח הם הממוצעים הכיתתיים של ציוני התלמידים. נרצה לבדוק אם יש הבדל בין מידת הסמכות שמייחסים תלמידים

צעירים ומבוגרים למורים למתמטיקה.

בלוח 2.3 רשומות גם הדרגות של הכיתות לפי מידת הסמכות שייחסו תלמידי הכיתה למורה למתמטיקה.

אנו מסמנים ב-X את מידת הסמכות של מורה בכיתות ז וב-Y את אלה של כיתות י. גודלי המדגם המתאימים הם: $m=9,\ n=8$. מלוח 2.3 רואים שהסטטיסטי של ווילקוקסון המתקבל כאן הוא $W_s=62$. כדי להחליט אם יש לדחות את השערת האפס על סמך התוצאה שהתקבלה, נרצה לקבוע, לדוגמה, את אזור הדחייה המתאים עבור רמת מובהקות $\alpha=.10$ ונברר אם התוצאה שהתקבלה כלולה באזור הדחייה.

אזור הדחייה עבור אלטרנטיבה דו־צדדית

במקרה של הדוגמה לעיל אין לנו השערה כיוונית, או תאוריה שתנבא איזו קבוצת גיל מייחסת יותר סמכות ידע למורה למתמטיקה. במִלים אחרות, השערת האפס טוענת ששתי ההתפלגויות שוות, והאלטרנטיבה טוענת שהן שונות (ייתכן ש-Y נוטה לקבל ערכים גדולים מ-X, או להיפך – X נוטה לקבל ערכים גדולים מ-Y.)

אם משתמשים בסטטיסטי של ווילקוקסון, יש לדחות את השערת האפס כאשר הסטטיסטי W_s גדול מדי או קטן מדי. במקרה כזה, כל ערך קיצוני של W_s מהווה ראָיה לכך שהשערת שוויון ההתפלגויות איננה נכונה. לכן אזור הדחייה של השערת האפס הוא אזור דחייה **דו־צדדי**. כיוון שהתפלגות W_s היא סימטרית, אזור הדחייה גם הוא סימטרי. כדי שרמת המובהקות (ההסתברות לדחייה תחת H_0) תהיה α, נדחה את השערת האפס בשני הזנבות הקיצוניים של התפלגות W_s, שהסתברות כל אחד מהם היא $\alpha/2$.

ניתן לראות דוגמה לאזור דחייה כזה בציור 2.1, שם מוצגת התפלגות W_s עבור הדוגמה לעיל, שבה $m=9,\ n=8$. עבור רמת מובהקות $\alpha=.10$ נדחה את השערת האפס כאשר $W_s\le 54$ או $W_s\ge 90$. הסתברות הדחייה תחת השערת האפס מתוארת על ידי שני התחומים הכהים בזנבות.

בדוגמה שלנו התוצאה שהתקבלה, $W_s=62$, איננה נופלת באזור הדחייה, ולכן מהנתונים לעיל אין לנו ראָיה לכך שהתפלגויות שתי קבוצות הגיל הן שונות.

מובהקות התוצאה בבעיה דו־צדדית

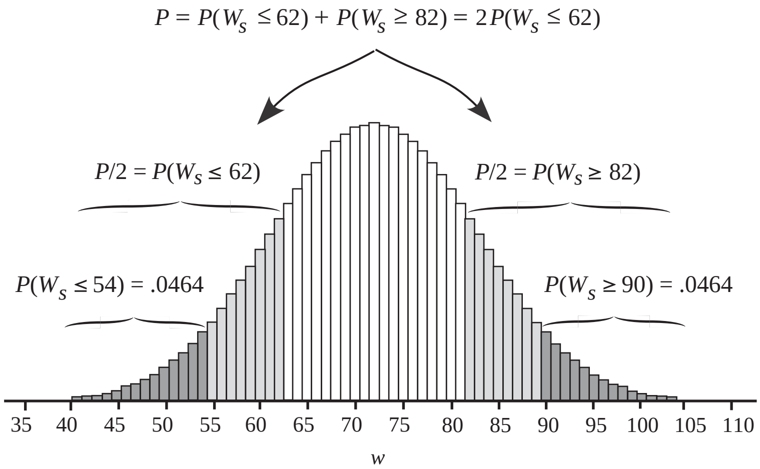
כיצד נחשב את מובהקות התוצאה, שהיא ההסתברות לתוצאה **קיצונית** לפחות כמו זו שהתקבלה בניסוי? ראשית יש לברר באיזה צד של ההתפלגות נפלה התוצאה הזאת (אם היא נמצאת בזנב הימני או בזנב השמאלי של ההתפלגות). לשם כך נבדוק ראשית אם התוצאה גדולה או קטנה ממרכז הסימטרייה – $n(N+1)/2$. בדוגמה זו מרכז הסימטרייה הוא $8(18)/2=72$. התוצאה 62 נמצאת, אפוא, בזנב השמאלי.

ההסתברות לתוצאה **נמוכה** כל כך היא: $P(W_s \le 62) = .1852$. הערך הזה נמצא בטבלה 2 בנספח, בעמודה שכותרתה $n=8,\ m=9$.

היות שהאלטרנטיבה היא דו־צדדית, תוצאה **קיצונית** לפחות כמו זו שהתקבלה היא גם כאשר W_s גדול או שווה לערך הסימטרי בזנב הימני – 82. לכן מובהקות התוצאה היא סכום שתי ההסתברויות הללו. בגלל הסימטרייה של התפלגות W_s, ההסתברות בזנב הימני שווה לזו של הזנב השמאלי, ולכן מובהקות התוצאה היא למעשה כפולה מהסתברות כל אחד מהזנבות (ראו ציור 2.1):

$$P = P(W_s \le 62) + P(W_s \ge 82) = 2P(W_s \le 62) = 2(.1852) = .3704$$

מובהקות זו גבוהה מאוד, ולכן אי אפשר לומר שיש הבדל בין צעירים ומבוגרים יותר ביחס להערכת סמכות ידע של מורים למתמטיקה.



ציור 2.1 *מובהקות התוצאה* $W_s = 62$ *במבחן ווילקוקסון עבור בעיה דו־צדדית* ($m=9, n=8$)

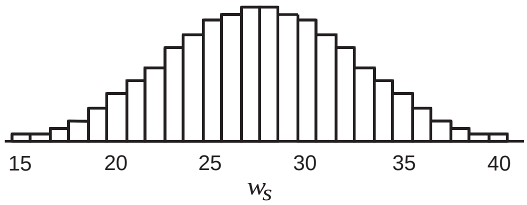
באופן כללי, נניח שאין לחוקר השערה כיוונית ספציפית והוא מעוניין רק להחליט אם ניתן להניח ששתי ההתפלגויות שוות, כנגד האלטרנטיבה שההתפלגויות שונות. (ייתכן ש-Y נוטה לקבל ערכים גדולים מ-X, או להיפך – X נוטה לקבל ערכים גדולים מ-Y.) בגלל הסימטרייה של התפלגות W_s, כדי לקבל את מובהקות התוצאה (ההסתברות לקבלת תוצאה **כה קיצונית**) יש לכפול פי 2 את הסתברות הזנב שמעבר לתוצאה שהתקבלה בניסוי, כאשר התחום של הזנב נקבע על פי המיקום של התוצאה ביחס למרכז הסימטרייה $n(N+1)/2$. ציור 2.1 מתאר את התפלגות הסטטיסטי של ווילקוקסון עבור $m = 9,\ n = 8$, ואת מובהקות התוצאה כאשר $W_s = 62$.

הכלל של השוואת מובהקות התוצאה – P לרמת המובהקות – α, כדי לקבוע אם התוצאה שהתקבלה היא מובהקת (ואז יש לדחות את השערת האפס) תופס, כמובן, גם

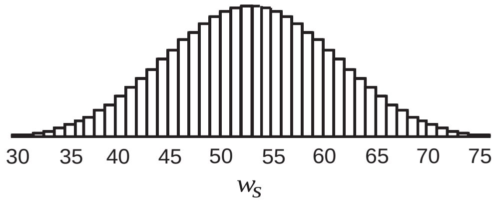
במקרה הזה. אם $P \leq \alpha$, אזי שטח הזנב בכיוון שבו התקבלה התוצאה אינו עולה על $\alpha/2$, ולכן התוצאה כלולה באזור הדחייה.

2.3 הקירוב הנורמלי להתפלגות W_s

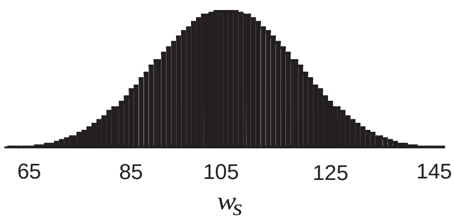
בציור 1.3 ניתן לראות את הסימטרייה של התפלגות הסטטיסטי של ווילקוקסון (משפט 2.2) וכן נוכל לראות שצורת ההתפלגות דומה במקצת לעקום הנורמלי. עבור מדגמים גדולים יותר, ההתפלגות קרובה יותר לעקום הנורמלי. ציור 2.2 מציג את התפלגות W_s עבור גודלי מדגמים שונים $(m=n=5,7,10)$.



ציור 2.2א. *התפלגות W_s עבור מדגמים בגודל* $m=n=5$



ציור 2.2ב. *התפלגות W_s עבור מדגמים בגודל* $m=n=7$



ציור 2.2ג. *התפלגות W_s עבור מדגמים בגודל* $m=n=10$

ניתן לראות מציור 2.2ב שכבר עבור מדגמים בגודל 7 ההתפלגות קרובה מאוד להתפלגות הנורמלית. באופן כללי, ניתן לרשום זאת על ידי המשפט הבא (שאותו איננו מוכיחים פה, כיוון שההוכחה היא מעבר לידע המתמטי הנדרש מקוראי הספר הזה).

משפט 2.3. תחת השערת האפס (שוויון התפלגות X ו-Y), התפלגות סכום הדרגות W_s היא אסימפטוטית נורמלית. כלומר, המשתנה המתוקנן $\frac{W_s-EW_s}{\sqrt{Var(W_s)}}$ הוא בעל התפלגות קרובה להתפלגות נורמלית סטנדרטית, כאשר m ו-n גדולים מאוד. המומנטים EW_s ו-$Var(W_s)$ הרשומים לעיל מחושבים תחת המודל של H_0.

כדי להשתמש במשפט 2.3 לחישוב מקורב של התפלגות הסטטיסטי של ווילקוקסון, עלינו לחשב את המומנטים (התוחלת והשונות) של W_s תחת השערת האפס.

טענה 2.1. אם השערת האפס נכונה, אזי כל אחת מדרגות ה-Y-ים היא בעלת התפלגות אחידה, כלומר $S_j \sim U(1,N)$, $j=1,\ldots,n$.

הוכחה: תחת השערת האפס כל התמורות (פרמוטציות) על N הדרגות הן שוות הסתברות. נסתכל, לדוגמה, על S_1 – הדרגה של Y_1. נחשב את ההסתברות ש-Y_1 יקבל את הדרגה 1 מבין כל N התצפיות (כלומר, שזו תהיה התצפית המינימלית). מספר כל התמורות של N התצפיות שבהן Y_1 היא התצפית המינימלית הוא $(N-1)!$, אלה כל הסידורים של שאר $N-1$ התצפיות בסדר כלשהו. מכיוון שכל הסידורים הם שווי הסתברות, ההסתברות המבוקשת היא

$$P(S_1=1)=\frac{(N-1)!}{N!}=\frac{1}{N}$$

הסתברות זו שווה גם להסתברות שדרגת Y_1 היא 2, 3, או N. כלומר, ישנה הסתברות שווה לקבלת כל אחת מהדרגות האפשריות. אותו דבר נכון, כמובן, לגבי כל אחת מהתצפיות האחרות. מכאן קיבלנו:

$$P(S_j=k)=1/N \quad k=1,\ldots,N \quad , j=1,\ldots,n \text{ לכל}$$

כלומר, התפלגות אחידה כפי שנטען. ♣

על סמך טענה 2.1 קל לחשב את תוחלת הסטטיסטי של ווילקוקסון, כפי שהיא מובאת במסקנה הבאה.

מסקנה 2.2. התוחלת של W_s (סכום הדרגות של n התצפיות בקבוצת הטיפול) היא

$$EW_s=\frac{n(N+1)}{2} \qquad (3)$$

הוכחה: נרשום את W_s כסכום הדרגות: $W_s=\sum_{j=1}^{n}S_j$. לפי טענה 2.1, אם השערת האפס נכונה, כל אחת מהדרגות היא בעלת התפלגות אחידה $U(1,N)$. לפיכך התוחלת שלה

היא

$$ES_j=\frac{N+1}{2}\ ,\qquad j=1,\ldots,n$$

התוחלת של W_s היא סכום התוחלות הללו:

$$EW_s=E\sum_{j=1}^{n}S_j=\sum_{j=1}^{n}ES_j=n\frac{N+1}{2}=\frac{n(N+1)}{2}$$

♣

הערה: את ערכה של התוחלת ניתן להסיק ישירות מהסימטרייה של התפלגות W_s סביב הערך $n(N+1)/2$ (משפט 2.2).

כדי למצוא את השונות, יש לחשב את שונות סכום הדרגות. היות שהדרגות אינן בלתי תלויות זו בזו, יש צורך לחשב זאת בעזרת השונות המשותפת בין כל שתי דרגות. ניתן לעשות זאת בצורות שונות. אנו נראה את החישוב באופן כללי בלמה 2.1, ונשתמש לאחר מכן בנוסחאות שהתקבלו עבור הסטטיסטי של ווילקוקסון.

למה 2.1. תהיה נתונה אוכלוסייה בת N איברים בעלת ממוצע μ ושונות σ^2. בוחרים מתוך האוכלוסיה n איברים **בלי החזרה** ($n\le N$) ומקבלים מדגם $U_1,U_2,\ldots,U_n$. אזי קיים:

$$E\sum_{j=1}^{n}U_j=n\mu \tag{4}$$

$$Var(\sum_{j=1}^{n}U_j)=n\sigma^2\frac{N-n}{N-1} \tag{5}$$

זאת אומרת, שונות הסכום שווה לסכום השונויות, הכופל בביטוי $\frac{N-n}{N-1}$, הנקרא כופל הסופיות. כאשר N שואף לאינסוף, כופל הסופיות שואף ל-1 (ואז שונות הסכום שווה לסכום השונויות, כמו בדגימה עם החזרה).

*הוכחה**: לגבי התוחלת הדבר ברור, כיוון ש-$EU_j=\mu$, לכל $j=1,\ldots,n$.

לגבי השונות, שונות הסכום היא:

$$Var(\sum_{j=1}^{n}U_j)=\sum_{j=1}^{n}Var(U_j)+\sum_{j=1}^{n}\sum_{\substack{l=1\\ l\ne j}}^{n}Cov(U_j,U_l)$$

$$=n\sigma^2+n(n-1)Cov(U_1,U_2) \tag{6}$$

בגלל הסימטרייה בין התצפיות, השונויות המשותפות של כל הזוגות הן שוות. רשמנו בנוסחה (6) את השונות המשותפת בין שתי התצפיות הראשונות.

את השונות המשותפת הבודדת ניתן לחשב בצורות שונות.

א. נסתכל על המקרה שבו $n = N$, כלומר, בוחרים את **כל** N איברי האוכלוסייה ללא החזרה. (בחירה כזאת היא למעשה סידור של איברי האוכלוסייה בסדר אקראי.) במקרה זה סכום כל התצפיות שווה תמיד לסכום כל איברי האוכלוסייה, כלומר, $\sum_{i=1}^{N} U_i = N\mu$. לפיכך שונות הסכום הזה היא אפס: $Var\left(\sum_{i=1}^{N} U_i\right) = 0$.

מצד שני, ניתן לרשום את השונות הזאת על פי נוסחה (6) לעיל, כאשר מציבים $n = N$. מקבלים:

$$0 = Var(\sum_{j=1}^{N} U_j) = N\sigma^2 + N(N-1)Cov(U_1,U_2)$$

מכאן:

$$Cov(U_1,U_2) = -\frac{N\sigma^2}{N(N-1)} = -\frac{\sigma^2}{N-1} \qquad (7)$$

נציב בנוסחה (6) את הערך שהתקבל בנוסחה (7) עבור השונות המשותפת ונקבל את שונות הסכום של n תצפיות:

$$Var(\sum_{j=1}^{n} U_j) = n\sigma^2 + n(n-1)Cov(U_1,U_2)$$

$$= n\sigma^2 - n(n-1)\frac{\sigma^2}{N-1} = n\sigma^2\left[1 - \frac{n-1}{N-1}\right]$$

$$= n\sigma^2\left[\frac{N-n}{N-1}\right]$$

ב. את השונות המשותפת (7) ניתן למצוא על ידי חישוב ישיר, בעזרת ההתפלגות המשותפת של שתי תצפיות. החישוב נמצא בנספח 2א. ♣

הערה: מנוסחת השונות שקיבלנו רואים שאם בוחרים מדגם מקרי בלי החזרה מאוכלוסייה סופית, שונות סכום התצפיות (ולכן גם שונות ממוצע התצפיות) תמיד קטנה יותר מהשונות במקרה שהדגימה נערכת עם החזרה.

משפט 2.4. אם השערת האפס נכונה, אזי התוחלת והשונות של W_s הן:

$$EW_s = \frac{n(N+1)}{2} \qquad (8)$$

$$Var(W_s) = \frac{nm(N+1)}{12} \qquad (9)$$

הוכחה: את נוסחת התוחלת כבר ראינו קודם (מסקנה 2.2). את נוסחת השונות קל להראות על ידי שימוש בלמה 2.1. הדרגות $S_1, S_2, \ldots, S_n$ הן מדגם בלי החזרה מאוכלוסיית המספרים הטבעיים: $\{1,2,\ldots,N\}$. השונות של כל דרגה בודדת (משתנה אחיד, לפי

טענה 2.1 היא $\sigma^2 = Var(S_j) = \frac{N^2-1}{12}$, לכל $j=1,\ldots,n$, ולפי נוסחה (5) שונות סכום הדרגות היא:

$$Var\left(\sum_{j=1}^{n} S_j\right) = n\sigma^2\left[\frac{N-n}{N-1}\right] = n\frac{N^2-1}{12}\left[\frac{m}{N-1}\right] = \frac{nm(N+1)}{12}$$

♣

על פי משפט 2.4 ניתן לחשב בקירוב את התפלגות הסטטיסטי של ווילקוקסון עבור מדגמים די גדולים. כדי לברר את טיב הקירוב, נסתכל על התוצאה בדוגמה 2.4 ונחשב את ההסתברויות המתאימות על סמך הקירוב הנורמלי.

דוגמה 2.5 (המשך דוגמה 2.4). נחשב את מובהקות התוצאה $W_s = 62$ שהתקבלה בדוגמה 2.4 (סמכות מורה למתמטיקה), עבור $m=9,\ n=8$. המומנטים המתאימים כאן הם (לפי משפט 2.4):

$$EW_s = \frac{n(N+1)}{2} = \frac{8(18)}{2} = 72$$

$$Var(W_s) = \frac{mn(N+1)}{12} = \frac{8(9)(18)}{12} = 108$$

הסתברות הזנב, על פי הקירוב הנורמלי:

$$P(W_s \le 62) \cong \Phi\left(\frac{62.5-72}{\sqrt{108}}\right) = \Phi(-0.91) = .1814$$

השתמשנו בתיקון רציפות עבור המשתנה הבדיד W_s. ההסתברות המדויקת, על פי טבלה 2 בנספח, נמצאה $P(W_s \le 62) = .1852$. אנו רואים, אפוא, שהקירוב הנורמלי נותן תוצאה די קרובה לערך הנכון, וזאת כבר עבור מדגמים לא גדולים.

2.4 התיאוריה של המושג "גדול סטוכסטית"*

נסתכל על שני משתנים X ו-Y, בעלי פונקציות התפלגות F ו-G, בהתאמה. כלומר:

$$F(t) = P(X \le t), \quad G(t) = P(Y \le t), \qquad -\infty < t < \infty \text{ עבור}$$

נרשום מודל עבור ההתפלגויות, שיבטא את העובדה שהמשתנה Y נוטה להיות גדול מהמשתנה X. למשל, נרצה שכל ערך חלוקה של Y יהיה גבוה מערך החלוקה המתאים של X.

הגדרה 2.1. אנו אומרים ש-Y **גדול סטוכסטית** מ-X, ומסמנים $Y \succ X$, אם פונקציות ההתפלגות שלהם מקיימות:

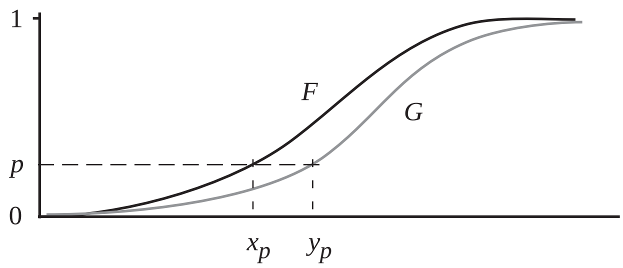
(10) $F(t) \geq G(t)$ לכל t, ויש ערך t_0 שעבורו $F(t_0) > G(t_0)$

משמעות ההגדרה היא שפונקציית ההתפלגות (המצטברת) של X אינה קטנה מזו של Y בשום נקודה על הישר הממשי, והיא ממש גדולה ממנה לפחות בנקודה אחת (ואז היא גדולה ממנה בקטע שלם כלשהו).

לפי הגדרת התפלגות מצטברת, כאשר Y גדול סטוכסטית מ-X, ההתפלגויות מקיימות: $F(t) = P(X \leq t) \geq P(Y \leq t) = G(t)$ לכל ערך של t, ומשמעות אי-שוויון זה היא שההסתברות של X לקבל ערכים נמוכים גדולה יותר מההסתברות של Y לקבל את הערכים הנמוכים הללו.

טענה 2.2. אם Y גדול סטוכסטית מ-X, אזי כל אחד מערכי החלוקה (אחוזונים) של Y אינו קטן מזה של X. כלומר, לכל $0 < p < 1$ מתקיים $x_p \leq y_p$.

הוכחה: ערכי החלוקה הם הערכים המקיימים: (i) $F(x_p) = p$, (ii) $G(y_p) = p$.
מהנחה (10) נובע $G(x_p) \leq F(x_p) = p$, כלומר, $G(x_p) \leq p$. אבל לפי (ii) מתקיים $G(y_p) = p$. קיבלנו, אפוא, שחייב להתקיים: $G(x_p) \leq G(y_p)$. כיוון שההתפלגות G היא מונוטוניות לא יורדת, חייב להתקיים היחס $x_p \leq y_p$. ראו ציור 2.3. ♣



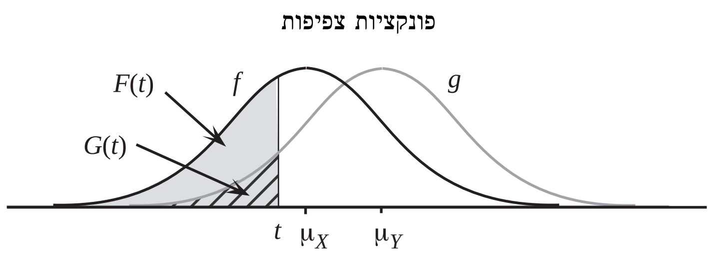
ציור 2.3. *שתי פונקציות התפלגות שעבורן Y גדול סטוכסטית מ-X.*

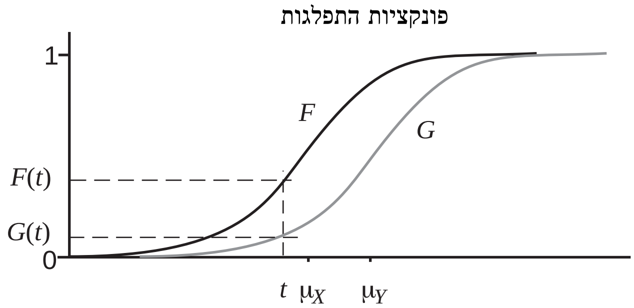
העובדה הכלולה בטענה לעיל מסבירה את המונח "גדול סטוכסטית" עבור המודל המוגדר ב-(10).
ניתן, אפוא, לרשום את ההשערות עבור בעיית שני המדגמים שבה אנו דנים כאן בצורה הבאה:

(11) $H_0 : F(t) \equiv G(t)$ $\quad H_1 : Y \succ X$

נסתכל על דוגמאות אחדות.

דוגמה 2.6 נניח ש-F ו-G הן התפלגויות נורמליות עם אותה שונות σ^2. כלומר, $X \sim N(\mu_X, \sigma^2)$ ו-$Y \sim N(\mu_Y, \sigma^2)$. כדי ש-$Y$ יהיה גדול סטוכסטית מ-X דרוש $\mu_Y > \mu_X$. שתי פונקציות הצפיפות ושתי פונקציות ההתפלגות ניתנות בציור 2.4. הציור מבהיר את הסיבה לכך שההתפלגות F נמצאת **מעל** להתפלגות G בכל נקודה על הישר.





ציור 2.4. *פונקציות הצפיפות ופונקציות ההתפלגות הנורמליות של X ושל Y בדוגמה 2.6 (שונויות שוות)*

חישוב ישיר מראה את הדבר:

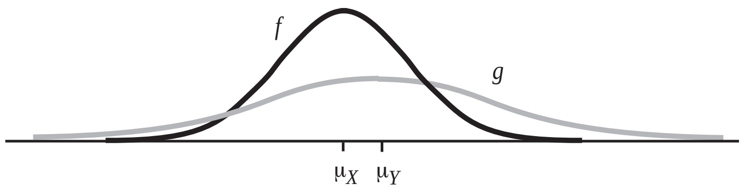
$F(t) = P(X \le t) = \Phi\left(\dfrac{t-\mu_X}{\sigma}\right)$ ובאופן דומה $G(t) = P(Y \le t) = \Phi\left(\dfrac{t-\mu_Y}{\sigma}\right)$.

מהעובדה ש-$\mu_Y > \mu_X$ ו-Φ מונוטונית עולה, נובע

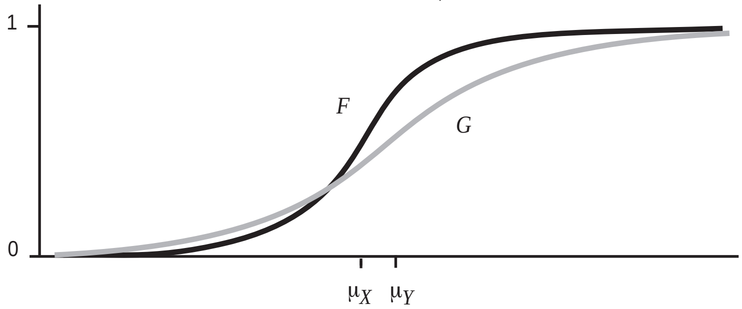
$$G(t) = \Phi\left(\frac{t-\mu_Y}{\sigma}\right) < \Phi\left(\frac{t-\mu_X}{\sigma}\right) = F(t)$$

הערה: אם השונויות אינן שוות בשתי ההתפלגויות הנורמליות, אזי אף אחד משני המשתנים איננו גדול סטוכסטית מהשני. במקרה כזה פונקציות ההתפלגות חותכות זו את זו בנקודה מסוימת. ניתן לראות זאת בציור 2.5. נוכל לרשום זאת כטענה כללית.

פונקציות צפיפות



פונקציות התפלגות



ציור 2.5. *פונקציות הצפיפות ופונקציות ההתפלגויות הנורמליות של X ושל Y (שונויות שונות)*

טענה 2.3. יהיו $X \sim N(\mu_X, \sigma_X^2)$ ו- $Y \sim N(\mu_Y, \sigma_Y^2)$, כאשר $\mu_X < \mu_Y$. אזי Y גדול סטוכסטית מ-X אם ורק אם $\sigma_X^2 = \sigma_Y^2$.

הוכחה: בדוגמה 2.6 הראינו שאם השונויות שוות, אזי קיים היחס הגדול סטוכסטית המתאים בין המשתנים. נרצה להוכיח את ההיפך, כלומר שלא ייתכן יחס סטוכסטי כזה, אלא אם כן השונויות שוות.

נניח, אפוא, שקיים היחס הסטוכסטי הדרוש. כלומר, מתקיים:

$$\Phi\left(\frac{t-\mu_Y}{\sigma_Y}\right) \le \Phi\left(\frac{t-\mu_X}{\sigma_X}\right) \quad \text{לכל } t.$$

בגלל המונוטוניות של Φ חייב להתקיים: $\frac{t-\mu_Y}{\sigma_Y} \le \frac{t-\mu_X}{\sigma_X}$ לכל t.

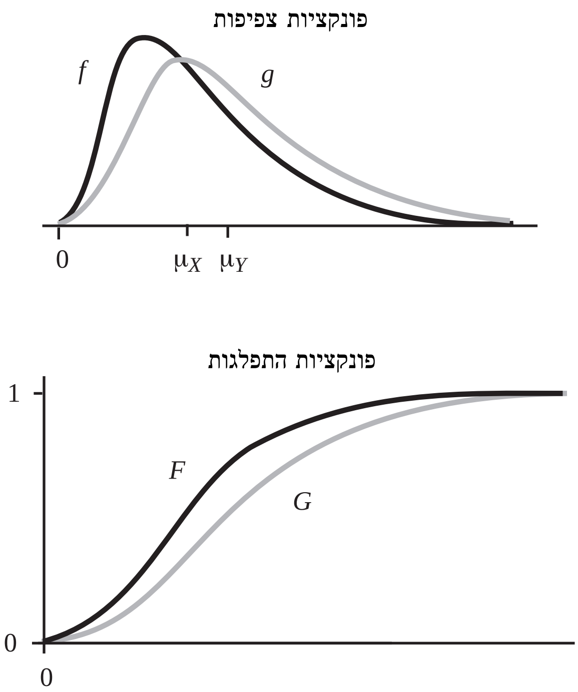
נשים לב שאם שתי סטיות התקן שוות, אזי האי־שוויון הזה מתקיים **לכל** t, אולם אם הן שונות, זה יתקיים רק בתחום מסוים, התלוי בכל 4 הפרמטרים. קל לראות שהאי־שוויון מתקיים רק אם $t(\sigma_X - \sigma_Y) \le \mu_Y\sigma_X - \mu_X\sigma_Y$.

אם $\sigma_X > \sigma_Y$, אזי האי־שוויון לעיל יתקיים כאשר $t \le \frac{\mu_Y\sigma_X - \mu_X\sigma_Y}{\sigma_X - \sigma_Y}$,

ואם $\sigma_X < \sigma_Y$, אזי הוא יתקיים כאשר $t \ge \frac{\mu_X\sigma_Y - \mu_Y\sigma_X}{\sigma_Y - \sigma_X}$.

בכל מקרה, ישנו תחום שעבורו האי־שוויון הדרוש איננו מתקיים. מובן שבמקרה זה אף אחד משני המשתנים אינו גדול סטוכסטית מן המשתנה האחר. מכאן אין יחס גדול סטוכסטית בין שני משתנים נורמלים אם השונויות שלהם שונות. ♣

דוגמה 2.7. בציור 2.6 אנו רואים התפלגות הכנסות באוכלוסייה בעלת רמת הכנסות נמוכה (משתנה X בעל הצפיפות f) וברמת הכנסות גבוהה יותר (משתנה Y בעל הצפיפות g). שלא כמו בדוגמה 2.6, בה שתי ההתפלגויות היו בעלות אותה צורה בדיוק, אך האחת הייתה "מוזזת" ביחס לשנייה, כאן הצפיפויות שונות בצורתן. עם זאת, ניתן לראות שהמשתנה Y גדול סטוכסטית מהמשתנה X. (ההתפלגויות בציור 2.6 הן ממשפחת התפלגויות הנקראות התפלגויות גאמא.)



ציור 2.6. *פונקציות הצפיפות ופונקציות ההתפלגות של ההכנסות בשתי אוכלוסיות*

2.5 הסטטיסטי של מאן-וויטני

בשנת 1947 הציעו מאן וּוויטני (Mann & Whitney, 1947) מבחן לבעיית שני המדגמים, בעיה (11). סטטיסטי המבחן מוגדר בעזרת המשתנים המציינים (אינדיקטורים)

הבאים:

$$U_{ij} = \begin{cases} 1 & X_i < Y_j \\ 0 & X_i > Y_j \end{cases} \qquad i = 1,\ldots,m; j = 1,\ldots,n \tag{12}$$

(היות שאנו מדברים על משתנים רציפים, ההסתברות לשוויון היא 0.)

סטטיסטי המבחן של מאן-וויטני מוגדר על ידי

$$W_{xy} = \sum_{i=1}^{m}\sum_{j=1}^{n} U_{ij} \tag{13}$$

המשתנה W_{xy} מבטא את מספר כל הזוגות של ההשוואות בין ערך של X במדגם הביקורת לערך של Y במדגם הטיפול, שבהן $X < Y$. בסך הכול יש לערוך mn השוואות כאלה, כשכל אחת מהתצפיות בקבוצת הביקורת מושווית לכל אחת מהתצפיות בקבוצת הטיפול.

דוגמה 2.8 (המשך דוגמה 2.3). נרשום שוב, בלוח 2.4, את תוצאות הניסוי לגבי רמת האלימות של בנים ובנות. בכל אחד מהמדגמים הנתונים כבר מסודרים בסדר עולה.

לוח 2.4. *חישוב הסטטיסטי של מאן-וויטני עבור נתוני אלימות מדוגמה 2.3*
($m = 9,\ n = 10$)

דרגה (בנות)	אלימות (x)	דרגה (בנים)	אלימות (y)	מספר ה-x-ים הקטנים מ-y	
1	0.82	4	1.36	3	
2	1.00	7	1.73	5	
3	1.17	8	1.82	5	
5	1.55	10	1.91	6	
6	1.64	12	2.09	7	
9	1.83	14	2.33	8	
11	2.00	15	2.36	8	
13	2.18	16	2.36	8	
17	2.50	18	2.55	9	
		19	2.73	9	
67		123		68	סך הכול

החישוב במקרה כזה הוא קל. לכל אחד מערכי y יש לספור את מספר ה-x-ים הקטנים ממנו ולסכם ערכים אלה עבור כל ה-y-ים במדגם. נסתכל על הבן הראשון, שציונו $y = 1.36$ (דרגתו 4). ישנן 3 בנות בעלות ציון x נמוך יותר. הערך הזה רשום בעמודה

הראשונה משמאל בלוח 2.4. באופן דומה, עבור כל אחד מהבנים רשום בלוח מספר הבנות, שעבורן ציון רמת האלימות x נמוך מהציון של אותו הבן y. בשורה השנייה אנו רושמים את מספר הבנות שעבורן ציון רמת האלימות קטן מ-1.73 (הערך y של הבן השני). יש חמש בנות כאלה. וכך הלאה. סיכום כל הערכים הללו נותן את הסטטיסטי של מאן-וויטני $W_{xy} = 68$.

למרות הצורה השונה שבה מוגדרים הסטטיסטים של מאן-וויטני ושל ווילקוקסון, מסתבר שהם למעשה אקוויוולנטיים. הקשר ביניהם מתבטא במשפט הבא.

משפט 2.5. יש קשר ליניארי פשוט בין הסטטיסטי של מאן-וויטני לבין הסטטיסטי של ווילקוקסון. עבור כל שני מדגמים בגודל m ו-n קיים:

$$W_{xy} = W_s - \frac{n(n+1)}{2} \qquad (14)$$

כאשר W_s הוא סכום n הדרגות של מדגם ה-y-ים.

הוכחה: כדי להוכיח את המשפט נרשום את דרגות ה-Y-ים במונחים של המשתנים המציינים U_{ij} מנוסחה (12). ראשית נסדר את תצפיות קבוצת הטיפול לפי הסדר, כלומר, נניח שקיים: $Y_1 < Y_2 < \cdots < Y_n$. הדרגות של התצפיות הללו הן $S_1, S_2, \cdots, S_n$. הדרגה של ה-Y המינימלי – S_1, היא מספר כל התצפיות הקטנות או השוות ל-Y_1, מבין **כל** N התצפיות של X ושל Y יחד. לכן דרגה זו שווה למספר ה-X-ים הקטנים מ-Y_1, נוסף על Y_1 עצמו. לפיכך ניתן לרשום:

$$S_1 = 1 + \#\{i : X_i < Y_1\} = 1 + \sum_{i=1}^{m} U_{i1}$$

כאשר המשתנים U_{ij} מוגדרים בנוסחה (12).

באותו אופן, הדרגה של Y_2 היא

$$S_2 = 2 + \#\{i : X_i < Y_2\} = 2 + \sum_{i=1}^{m} U_{i2}$$

(שני ה-Y-ים הקטנים או השווים ל-Y_2, נוסף על ה-X-ים הקטנים ממנו). באופן כללי ניתן לרשום עבור כל אחת מהדרגות של ה-Y-ים:

$$S_j = j + \#\{i : X_i < Y_j\} = j + \sum_{i=1}^{m} U_{ij} \qquad j = 1, \ldots, n \qquad (15)$$

הסטטיסטי של ווילקוקסון הוא סכום כל n הדרגות:

$$W_s = \sum_{j=1}^{n} S_j = \sum_{j=1}^{n}\left(j + \sum_{i=1}^{m} U_{ij}\right) = \sum_{j=1}^{n} j + \sum_{j=1}^{n}\sum_{i=1}^{m} U_{ij} = \frac{n(n+1)}{2} + W_{xy}$$

אגף ימין מתקבל לפי הגדרת W_{xy} בנוסחה (13), כסכום כל המשתנים המציינים.

♣

נבדוק את הקשר הזה עבור מדד האלימות. הערך של הסטטיסטי של ווילקוקסון (סכום הדרגות) שהתקבל הוא $W_s = 123$ (ראו סכום הדרגות בלוח 2.4). הערך של הסטטיסטי של מאן-וויטני (לוח 2.4) הוא $W_{xy} = 68$. ההפרש בין שני הערכים הוא $W_s - W_{xy} = 123 - 68 = 55$.

הפרש זה הוא בדיוק הערך הרשום במשפט 2.5:

$$\frac{n(n+1)}{2} = \frac{10(11)}{2} = 55$$

מסקנה 2.3. לפי משפט 2.5 שני הסטטיסטים – של ווילקוקסון ושל מאן-וויטני – הם אקוויוולנטיים; כל ההבדל ביניהם הוא רק בהזזה בגודל קבוע. את התפלגות הסטטיסטי של מאן-וויטני ניתן למצוא בעזרת ההתפלגות של W_s. כמו כן, על סמך משפט 2.4 המומנטים של הסטטיסטי של מאן־וויטני, תחת השערת האפס, ניתנים על ידי

$$EW_{xy} = EW_s - \frac{n(n+1)}{2} = \frac{n(N+1)}{2} - \frac{n(n+1)}{2} = \frac{nm}{2} \tag{16}$$

$$Var(W_{xy}) = Var(W_s) = \frac{nm(N+1)}{12} \tag{17}$$

את התוחלת של W_{xy} קל לחשב גם באופן ישיר מן ההגדרה (13). המשתנים X_i ו- Y_j הם בלתי תלויים ובעלי התפלגויות רציפות לכל i ו-j. תחת השערת האפס (שוויון התפלגויות) כל N המשתנים הם בעלי אותה התפלגות (רציפה) ובלתי תלויים, ולכן מתקיים: $P(X_i < Y_j) = P(X_i > Y_j) = 1/2$. לפיכך התוחלת של כל אחד מהמשתנים המציינים היא $EU_{ij} = P(X_i < Y_j) = 1/2$.

עבור סכום המשתנים המציינים מקבלים את סכום התוחלות:

$$EW_{xy} = E\sum_{i=1}^{m}\sum_{j=1}^{n} U_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} EU_{ij} = mn\frac{1}{2} = \frac{mn}{2}$$

2.6 בעיות של ערכי תיקו (תוצאות שוות בניסוי)

אם ההתפלגויות F ו-G אינן רציפות, יכולים להתקבל ערכים זהים של תצפיות שונות. אנו קוראים לערכים כאלה **ערכי תיקו**. (באנגלית ערכים זהים נקראים ties ולכן יש המתרגמים זאת לעברית כ"קשרים".)

כדי להשתמש במבחן ווילקוקסון יש לתת דרגה לכל תצפית. שיטת הדירוג נעשית באופן הבא: ערכים זהים מקבלים אותה דרגה; הדרגה שתינתן לקבוצה של ערכים שווים היא הדרגה הממוצעת המגיעה להם, אילו היו כולם שונים.

דוגמה 2.9. נסתכל על תצפיות פיקטיביות והדרגות הניתנות להן.

תצפיות:	10	9	9	9	7	7	5	2
דרגות:	8	6	6	6	3.5	3.5	2	1

לשתי התצפיות שערכן 7 מגיעות הדרגות 3 ו-4, ולכן שתיהן מקבלות את הדרגה הממוצעת 3.5; לשלוש התצפיות שערכן 9 מגיעות הדרגות 5, 6 ו-7, שממוצען הוא 6.

הערה: שימו לב שסכום כל 8 הדרגות הממוצעות כאן הוא 36, השווה לסכום המספרים הטבעיים בין 1 ל-8: $36 = 8\cdot 9/2$. זהו בדיוק סכום הדרגות שהיה מתקבל כאן אילו כל התצפיות היו שונות (ללא ערכי תיקו). אין זה מקרה ונראה זאת באופן כללי יותר מאוחר (משפט 2.6).

הגדרה 2.2. קבוצת תיקו בגודל t היא קבוצה של t תצפיות שכולן בעלות אותו ערך, השונה מכל שאר התצפיות שאינן כלולות בקבוצה.

בדוגמה 2.9 יש שתי קבוצות תיקו, האחת בגודל $t_1 = 2$ והשנייה בגודל $t_2 = 3$. (ניתן לראות גם תצפית בודדת, השונה מכל השאר, כקבוצת תיקו בגודל 1.) הסטטיסטי של ווילקוקסון במקרה כזה הוא סכום הדרגות "הממוצעות" שניתנו לתצפיות בקבוצת הטיפול. כלומר, נסמן ב-$\tilde{S}_1,\tilde{S}_2,\ldots,\tilde{S}_n$ את הדרגות הממוצעות של ה-Y-ים. אזי הסטטיסטי של ווילקוקסון מוגדר על ידי:

$$\tilde{W}_s = \sum_{j=1}^{n} \tilde{S}_j \qquad (18)$$

הבעיה במקרה של ערכי תיקו היא בכך שגם כאשר השערת האפס נכונה (ישנה אקראיות גמורה בחלוקה של אותן N דרגות "ממוצעות" בין m ערכי X ל-n ערכי Y), אי אפשר לקבוע את התפלגות סכום הדרגות $\tilde{W}_s$ על פי הטבלה הנתונה ל-W_s (טבלה 2 בנספח). לשם הבהרה נסתכל שוב על דוגמה 2.9.

דוגמה 2.10 (המשך דוגמה 2.9). נניח שבדוגמה זו $m = n = 4$. כדי לקבוע את התפלגות הסטטיסטי $\tilde{W}_s$ יש לרשום את כל הקומבינציות האפשריות של בחירת ארבע מתוך 8 הדרגות הרשומות. (חלק מאפשרויות הבחירה הן זהות, היות שישנן דרגות זהות.) כל הקומבינציות האפשריות הללו רשומות בלוח 2.5.

לוח 2.5. *רשימת כל הקומבינציות האפשריות של בחירת 4 מתוך 8 הדרגות:*
1, 2, 3.5, 3.5, 6, 6, 6, 8

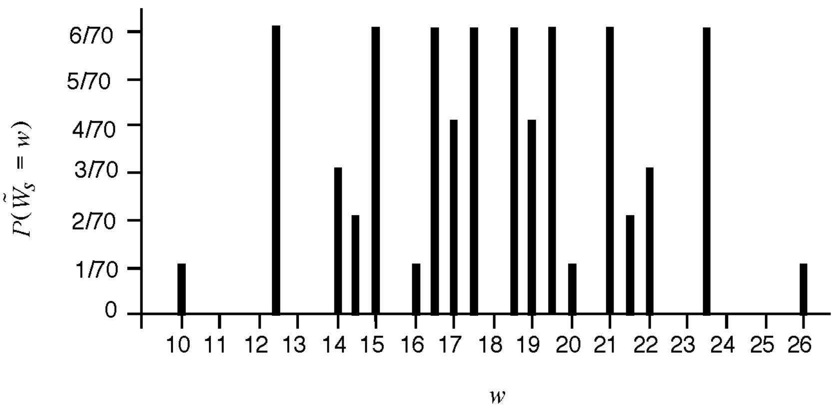
	הדרגות									
אפשרות	1	2	3.5	3.5	6	6	6	8	$\tilde{W}_s$	#
1	*	*	*	*					10	1
2	*	*	*		*				12.5	6
3	*	*	*				*		14.5	2
4	*	*			*	*			15	3
5	*	*			*			*	17	3
6	*		*	*	*				14	3
7	*		*	*				*	16	1
8	*		*		*	*			16.5	6
9	*		*		*			*	18.5	6
10	*				*	*	*		19	1
11	*				*	*		*	21	3
12		*	*	*	*				15	3
13		*	*	*				*	17	1
14		*	*		*	*			17.5	6
15		*	*		*			*	19.5	6
16		*			*	*	*		20	1
17		*			*	*		*	22	3
18			*	*	*	*			19	3
19			*	*	*			*	21	3
20			*		*	*	*		21.5	2
21			*		*	*		*	23.5	6
22					*	*	*	*	26	1
סך הכול										70

$\tilde{W}_s$ הוא סכום הדרגות המסומנות ב-*. # הוא מספר האפשרויות לכל צירוף.

המקומות המסומנים ב-* הן הדרגות שניתנות ל-Y-ים. סכום הדרגות הללו הוא $\tilde{W}_s$ והסימון # מציין את מספר הפעמים שהצירוף המסוים של הדרגות יכול להתקבל. למשל, אם נסתכל על אפשרות מס' 2, כל אחד משני האיברים שדרגתם 3.5 וכן כל אחד מהאיברים שדרגתם 6 יכלו להיבחר לקבוצת הטיפול (Y) ולכן יש בסך הכול $2\times3=6$ אפשרויות לבחירת ארבעה איברים בעלי הדרגות 1, 2, 3.5, 6 מתוך שמונה הדרגות הנתונות. באפשרות מס' 3 יש לבחור אחד משני האיברים שדרגתם 3.5 כדי לקבל את הדרגות 1, 2, 3.5, 8.

המספר הכולל של הקומבינציות האפשריות הוא $70=\binom{8}{4}=\binom{N}{n}$.

ההתפלגות של $\tilde{W}_S$ תחת השערת האפס מתקבלת מלוח 2.5 על ידי ספירת האפשרויות לקבלת כל אחד מהערכים, ודיאגרמת מקלות מתאימה מוצגת בציור 2.7.



ציור 2.7. *פונקציית ההסתברות של $\tilde{W}_S$, תחת השערת האפס, בדוגמה 2.10*

בציור 2.7 אנו רואים שהתפלגות הסטטיסטי של ווילקוקסון במקרה שבמדגמים יש ערכי תיקו אינה דומה למקרה הרציף, שבו כל התצפיות הן שונות. התפלגות זו היא אמנם סימטרית, אבל אינה "רגולרית" כמו בציור 1.3, למשל, וכך גם ההפרש בין שני ערכים סמוכים אינו זהה; אנו רואים פערים של 0.5, 1, וגם של 2.5 בין שני ערכים סמוכים של סכום הדרגות.

על פי צורת חישוב ההתפלגות ברור שההתפלגות בציור 2.7 מתאימה אך ורק לבעיה הספציפית בדוגמה 2.10, עם הדרגות המסוימות שהתקבלו שם. בכל מקרה של ערכי תיקו שונים מאלה, ההתפלגות תהיה שונה. ההתפלגות בציור 2.7 היא, אם כן, **ההתפלגות המותנית** של הסטטיסטי של ווילקוקסון, בהינתן רשימת שמונה הדרגות של התצפיות שהתקבלו בניסוי.

מסקנה 2.4. במקרה שיש ערכי תיקו בין N התצפיות שהתקבלו בניסוי, ניתן לחשב את **ההתפלגות המותנית** של $\tilde{W}_S$, בהינתן ערכי הדרגות הממוצעות. ההתפלגויות המותנות הללו שונות זו מזו ויש לחשב אותן בכל מקרה בנפרד, על ידי ספירת האפשרויות לקבלת כל אחד מהערכים של סכום הדרגות.

ברור שאין בידינו טבלאות של התפלגות $\tilde{W}_S$ במקרים של תיקו בין התצפיות, ולכן אם תרצו לקבל את המובהקות המדויקת של תוצאת הניסוי, תצטרכו לערוך את החישובים בעצמכם. מהתבוננות בדוגמה 2.10 אנו ערים לבעייתיות של חישוב כזה, אפילו עבור

מדגמים כה קטנים ($m = n = 4$). מתברר שבמקרים שההתפלגויות אינן רציפות, רצוי לקחת מדגמים גדולים יותר, שעבורם כבר ניתן להשתמש בקירוב הנורמלי לשם חישוב ההתפלגות. כדי להשתמש בקירוב נורמלי עלינו למצוא את התוחלת ואת השונות של $\tilde{W}_s$. התוצאות מובאות במשפט 2.6.

משפט 2.6. בהינתן N תצפיות מהתפלגות F שאיננה רציפה, התוחלת והשונות של $\tilde{W}_s$, תחת השערת האפס, הן

$$E\tilde{W}_s = EW_s = \frac{n(N+1)}{2} \tag{19}$$

$$Var(\tilde{W}_s) = Var(W_s) - \frac{nm}{12N(N-1)}\sum_r t_r(t_r^2 - 1) \tag{20}$$

כאשר r עובר על כל קבוצות התיקו.

*הוכחה**: נרשום את $\tilde{W}_s$ כסכום הדרגות הממוצעות: $\tilde{W}_s = \sum_{j=1}^{n} \tilde{S}_j$.

המשתנים $\tilde{S}_1, \tilde{S}_2, \ldots, \tilde{S}_n$ הם מדגם מקרי מתוך אוכלוסיית כל N הדרגות (הממוצעות) שהתקבלו בניסוי. לפי למה 2.1, נוסחאות (4) ו-(5), התוחלת והשונות של סכום הדרגות הממוצעות הללו ניתנות לחישוב באמצעות התוחלת והשונות של דרגה בודדת כזאת (כלומר התוחלת והשונות של "אוכלוסיית" הדרגות הממוצעות שהתקבלו בניסוי). נוכיח את המשפט ראשית עבור המקרה הפרטי שבו רק t התצפיות הנמוכות ביותר הן שוות, וכל השאר שונות זו מזו. כלומר, ישנה רק קבוצת תיקו אחת בגודל t, והיא קבוצת t התצפיות הנמוכות ביותר. רשימת כל N הדרגות של התצפיות שהתקבלו היא, אפוא,

$$\underbrace{\frac{t+1}{2}, \frac{t+1}{2}, \ldots, \frac{t+1}{2}}_{t},\ t+1,\ t+2,\ \ldots,\ N-1,\ N \tag{21}$$

[הדרגה של כל אחת מ-t התצפיות הראשונות היא ממוצע הערכים $1,2,\ldots,t$ והיא שווה ל-$(t+1)/2$. כל שאר הדרגות הן המספרים הטבעיים העוקבים מ- $t+1$ עד N.]

התוחלת של $\tilde{S}_1$ היא ממוצע אוכלוסיית N הדרגות (הממוצעות).

סכום t הדרגות הראשונות ברשימה (21), הוא

$$t\left(\frac{t+1}{2}\right) = 1+2+\cdots+t$$

וסכום כל הדרגות ברשימה הוא, אם כך,

$$[1+\ldots+t]+(t+1)+\ldots+N = \frac{N(N+1)}{2}$$

היות שהדרגות של t ערכי התיקו שוות לדרגה הממוצעת המגיעה להן, סכום הדרגות שלהן נשאר שווה לסכום t המספרים הטבעיים הראשונים. לכן גם סכום כל N הדרגות שווה לסכום N המספרים הטבעיים הראשונים, כלומר לסכום של N הדרגות שהיו מתקבלות אילו כל N התצפיות היו שונות זו מזו (ללא כל ערכי תיקו).

התוחלת, לפיכך, זהה לתוחלת של S_1 (כאשר אין ערכי תיקו כלל):

$$E\tilde{S}_1 = \frac{1}{N}\frac{N(N+1)}{2} = \frac{N+1}{2} = ES_1$$

תוחלת סכום הדרגות הממוצעות שווה, אפוא, לתוחלת סכום הדרגות במקרה שאין תיקו. דהיינו, $E\tilde{W}_S = EW_S$.

נשים לב שחישוב התוחלת לעיל נכון גם במקרה שיש יותר מקבוצת תיקו אחת, כיוון שבתוך כל קבוצת תיקו בנפרד, סכום הדרגות הממוצעות שווה לסכום המספרים הטבעיים המתאימים.

נחזור לרשימת הדרגות (21). את השונות של $\tilde{S}_1$ נמצא על ידי מציאת ההפרש בין השונות של $\tilde{S}_1$ לבין השונות של S_1 (הידועה לנו). $\tilde{S}_1$ מקבל כל אחד מהערכים ברשימה (21) בהסתברות $1/N$, ולכן שונותו היא

$$Var(\tilde{S}_1) = \frac{1}{N}\left[t\left(\frac{t+1}{2} - \frac{N+1}{2}\right)^2 + \sum_{k=t+1}^{N}\left(k - \frac{N+1}{2}\right)^2\right]$$

S_1 מקבל כל אחד מהערכים $1,2,\ldots,N$ בהסתברות $1/N$, ולכן שונותו היא

$$Var(S_1) = \frac{1}{N}\sum_{k=1}^{N}\left(k - \frac{N+1}{2}\right)^2$$

כל המחוברים מ- $k=t+1$ עד $k=N$ זהים עבור שתי השונויות הללו ולכן ההפרש בין שתי השונויות הוא ההפרש עבור t האיברים הראשונים בלבד. כלומר:

$$Var(S_1) - Var(\tilde{S}_1) = \frac{1}{N}\left[\sum_{k=1}^{t}\left(k - \frac{N+1}{2}\right)^2 - t\left(\frac{t+1}{2} - \frac{N+1}{2}\right)^2\right] \tag{22}$$

נסתכל עתה על הביטוי (22) ללא חלוקה ב-N. נשתמש בכלל הידוע לגבי סכום ריבועי הסטיות של רשימת מספרים מערך כלשהו:

$$\sum_{i=1}^{m}(a_i - c)^2 = \sum_{i=1}^{m}(a_i - \bar{a})^2 + m(\bar{a} - c)^2 \tag{23}$$

כאשר $\bar{a}$ הוא ממוצע m הערכים $a_1,\ldots,a_M$. (קל מאוד לבדוק את נכונות הפירוק הזה. נסו את כוחכם!)

נפרק את המחובר הראשון באגף ימין של (22) לפי נוסחה (23), כאשר אנו מציבים $m=t$ (מספר המחוברים) ו- $\bar{a} = (t+1)/2$ (ממוצע t האיברים הללו). נקבל:

(24) $$\sum_{k=1}^{t}\left(k-\frac{N+1}{2}\right)^2=\sum_{k=1}^{t}\left(k-\frac{t+1}{2}\right)^2+t\left(\frac{t+1}{2}-\frac{N+1}{2}\right)^2$$

הצבה של (24) בביטוי (22) נותן את הפרש השונויות:

(25) $$Var(S_1)-Var(\tilde{S}_1)=\frac{1}{N}\left[\sum_{k=1}^{t}\left(k-\frac{t+1}{2}\right)^2\right]=\frac{t}{N}Var(U)$$

$$=\frac{t}{N}\cdot\frac{t^2-1}{12}=\frac{t(t^2-1)}{12N}$$

כאשר U הוא משתנה אחיד $U\sim U(1,t)$, בעל תוחלת $\frac{t+1}{2}$ ושונות $\frac{t^2-1}{12}$.

מכאן קיבלנו את הקשר בין השונות של דרגה בודדת במקרה של קבוצת תיקו בגודל t, לעומת המקרה שבו אין ערכי תיקו:

(26) $$Var(\tilde{S}_1)=Var(S_1)-\frac{t(t^2-1)}{12N}$$

כדי לקבל את שונות סכום הדרגות, נציב את (26) בנוסחת השונות של סכום הדרגות, לפי למה 2.1, נוסחה (5), ונקבל את השונות של $\tilde{W}_s$:

$$Var(\tilde{W}_s)=Var\left(\sum_{i=1}^{n}\tilde{S}_1\right)=nVar(\tilde{S}_1)\left[\frac{N-n}{N-1}\right]$$

$$=n\left[Var(S_1)-\frac{t(t^2-1)}{12N}\right]\cdot\left[\frac{N-n}{N-1}\right]$$

מכאן:

(27) $$Var(\tilde{W}_s)=nVar(S_1)\left[\frac{N-n}{N-1}\right]-\frac{nt(t^2-1)}{12N}\left[\frac{N-n}{N-1}\right]$$

נשים לב שהמחובר הראשון בנוסחה (27) הוא בדיוק השונות של סכום הדרגות W_s (ללא תיקו). מכאן נוכל לרשום את (27) בצורה יותר פשוטה, כשנציב $N-n=m$:

(28) $$Var(\tilde{W}_s)=Var(W_s)-\frac{mnt(t^2-1)}{12N(N-1)}$$

בכך הוכחנו את המשפט עבור המקרה הפרטי של קבוצת תיקו אחת בלבד, שהיא קבוצת t הערכים הנמוכים ביותר. [זו בדיוק הנוסחה (20) הרשומה במשפט, עבור המקרה הנידון.]

מתוך חישוב השונות של $\tilde{S}_1$ שערכנו כאן ברור שאותה תוצאה מתקבלת גם אם ישנה בדיוק קבוצה אחת של t ערכים שווים, אשר אינם דווקא הערכים הנמוכים ביותר. ההפרש בין סכום ריבועי הסטיות של הדרגות הרגילות לבין זה של הדרגות הממוצעות תלוי רק בשונות של הדרגות הרגילות, ללא ערכי תיקו, בקבוצה שבה למעשה כל

הערכים שווים. השונות הזאת שווה לשונות של משתנה אחיד, שהוא הזזה פשוטה של $U\sim U(1,t)$, ולכן השונות שלו זהה לשונות הקודמת.

הכללה נוספת של התוצאה לגבי יותר מקבוצת תיקו אחת גם היא ברורה. את ההפרש (22) ניתן לרשום כסכום של ההפרשים לגבי כל קבוצות התיקו. שימוש בפירוק (24) עבור כל קבוצת תיקו בנפרד, נותן בסך הכול את הקשר הבא בין שונות של דרגה בודדת, כאשר אין ערכי תיקו, לבין שונות של דרגה בודדת עם נוכחות של ערכי תיקו. כל קבוצת תיקו בגודל t_r מורידה את השונות של דרגה בודדת בְּגודל המתאים לשונות משתנה אחיד $U(1,t_r)$. יש לסכם, אם כך, את כל השונויות הללו כדי לקבל את ההפרש בין השונות של S_1 לבין השונות של $\tilde{S}_1$. לכן מקבלים באופן כללי:

$$Var(S_1)-Var(\tilde{S}_1)=\sum_r \frac{t_r(t_r^2-1)}{12N} \qquad (29)$$

שונות סכום הדרגות $\tilde{W}_s$ מתקבלת בהתאם. ניתן לרשום אותה, בדומה לנוסחה (28):

$$Var(\tilde{W}_s)=Var(W_s)-\frac{mn}{12N(N-1)}\sum_r t_r(t_r^2-1) \qquad (30)$$

וזו הצורה הרשומה במשפט. ♣

על ידי הצבת השונות של W_s מנוסחה (9) בנוסחה (30), מקבלים נוסחה סגורה עבור השונות של $\tilde{W}_s$:

$$\begin{aligned}Var(\tilde{W}_s)&=\frac{nm(N+1)}{12}-\frac{nm}{12N(N-1)}\sum_r t_r(t_r^2-1)\\&=\frac{mn}{12N(N-1)}\left[N(N^2-1)-\sum_r t_r(t_r^2-1)\right] \qquad (31)\end{aligned}$$

מנוסחה (31) ברור שבכל מקרה של נוכחות ערכי תיקו, השונות של הסטטיסטי של ווילקוקסון קטנה מזו שהייתה מתקבלת ללא ערכי תיקו. עם זאת, כאשר N גדול, אם קבוצות התיקו הן די קטנות אזי אין להן השפעה רבה על השונות.

נחשב, לדוגמה, את השונות עבור הנתונים בדוגמה 2.9, באופן ישיר וגם על סמך הנוסחה (31).

דוגמה 2.11 (*המשך דוגמה 2.9*). השונות ללא ערכי תיקו עבור גודלי המדגם $m = n = 4$ היא:

$$Var(W_s)=\frac{mn(N+1)}{12}=\frac{4\cdot4\cdot9}{12}=12$$

לשם חישוב השונות עם התחשבות בערכי התיקו, נסתכל על ערכי $\tilde{W}_s$ האפשריים, כפי

שהם רשומים בלוח 2.5, או בציור 2.7.

$$E\tilde{W}_s = \frac{1}{70}[10+6\cdot 12.5+\cdots+6\cdot 23.5+26] = \frac{1260}{70} = 18$$

$$E\tilde{W}_s^2 = \frac{1}{70}[10^2+6\cdot 12.5^2+\cdots+6\cdot 23.5^2+26^2]$$

$$= \frac{23{,}470}{70} = 335.286$$

מכאן השונות:

$$Var(\tilde{W}_s) = E\tilde{W}_s^2 - [E\tilde{W}_s]^2 = 335.286 - 18^2 = 11.286$$

נחשב עתה את השונות על פי נוסחה (31).

בין הנתונים יש שתי קבוצות תיקו בגדלים $t_1 = 2,\ t_2 = 3$. לפי זה:

$$\sum_r t_r(t_r^2-1) = 2(2^2-1)+3(3^2-1) = 6+24 = 30$$

$$Var(\tilde{W}_s) = \frac{mn}{12N(N-1)}\left[N(N^2-1) - \sum_r t_r(t_r^2-1)\right]$$

$$= \frac{4\cdot 4}{12\cdot 8\cdot 7}\left[8(8^2-1)-30\right] = 11.286$$

התוצאה זהה, כמובן, לשונות שחושבה באופן ישיר.

הפער היחסי בין השונות ללא נוכחות של ערכי תיקו לבין השונות כאשר ישנם ערכי תיקו בדוגמה זו, הוא של

$$\frac{Var(W_s) - Var(\tilde{W}_s)}{Var(W_s)} = \frac{12-11.286}{12} = \frac{0.714}{12} = 0.0595$$

כלומר, השונות קטנה בכ-6%.

באופן כללי, הפער היחסי בין השונות ללא נוכחות של ערכי תיקו לבין השונות כאשר ישנם ערכי תיקו, הוא של

$$\frac{Var(W_s) - Var(\tilde{W}_s)}{Var(W_s)} = \frac{\frac{nm}{12N(N-1)}\sum_r t_r(t_r^2-1)}{\frac{nm(N+1)}{12}}$$

$$= \frac{1}{N(N^2-1)}\sum_r t_r(t_r^2-1)$$

הפער זניח אם המספר הכולל של תצפיות הוא גדול מאוד ואין הרבה ערכי תיקו.

נראה עתה דוגמה למקרה שקבוצות התיקו הן יחסית גדולות.

דוגמה 2.12. להלן נתונים בטבלת שכיחויות, לגבי הערכה שנתנו תלמידי קורס בסטטיסטיקה לתואר שני לגבי מידת הקושי של הקורס. (השאלה הייתה: "עד כמה הקורס קשה לך באופן כללי?") התוצאות רשומות בלוח 2.6 בנפרד עבור תלמידים שסיימו תואר ראשון לפני 5 שנים לכל היותר, ועבור אלה שסיימו לפני יותר מ-5 שנים.

נרצה לבחון אם לתלמידים שסיימו את לימודי התואר הראשון לפני זמן רב קשה יותר ללמוד סטטיסטיקה מאשר לאלה שסיימו לא מזמן (השוואת השכיחויות בשורה הראשונה לשורה השנייה בלוח 2.6).

לוח 2.6. *שכיחויות של תלמידים לפי הערכת הקושי של הקורס ומשך הזמן מסיום תואר ראשון*

סיום תואר ראשון	הערכת הקושי 1 (בכלל לא קשה)	2	3	4	5 (קשה מאוד)	סך הכול
קבוצה א (לא יותר מ-5 שנים)	2	6	7	4	0	19
קבוצה ב (יותר מ-5 שנים)	2	9	5	11	5	32
סך הכול	4	15	12	15	5	51
דרגה ממוצעת	2.5	12	25.5	39	49	

נסמן ב-X את הערכת הקושי של תלמידים שסיימו תואר ראשון לא מזמן (קבוצה א) וב-Y את אלה שסיימו מזמן (קבוצה ב).

אנו מעוניינים לבדוק את הבעיה החד־צדדית, שלאחר יותר זמן הקושי גדול יותר. כלומר, נבדוק את ההשערה:

לכל t, $H_0: F(t)=G(t)$ כנגד האלטרנטיבה $H_1: Y \succ X$.

יש כאן הרבה ערכי תיקו – כל הערכים הכלולים באותה עמודה בלוח 2.6 מהווים ערכי תיקו (אלה התלמידים שהעריכו את הקושי באותו אופן בדיוק). כלומר, גודלי קבוצות התיקו הם: $t_1=4$, $t_2=15$, $t_3=12$, $t_4=15$ ו- $t_5=5$. הדרגה של כל הערכים באותה עמודה היא הדרגה הממוצעת המגיעה להם. נִראה כיצד חושבו הדרגות הממוצעות.

עבור העמודה הראשונה (הערך 1), הדרגות המגיעות ל-$t_1=4$ הערכים הן 1, 2, 3 ו-4 והדרגה הממוצעת היא $\tilde{s}_1=(1+4)/2=2.5$;

עבור העמודה השנייה (הערך 2), הדרגות המגיעות ל-$t_2=15$ הערכים הן $5,6,\ldots,19$ והדרגה הממוצעת היא $\tilde{s}_2=(5+19)/2=12$;

עבור העמודה השלישית (הערך 3), הדרגה הממוצעת היא $\tilde{s}_3=(20+31)/2=25.5$

(בְדקו!);

עבור העמודה הרביעית (הערך 4) הדרגה הממוצעת היא $\tilde{s}_4=(32+46)/2=39$;

ועבור העמודה החמישית (הערך 5), הדרגה הממוצעת היא $\tilde{s}_5=(47+51)/2=49$.

הסטטיסטי של ווילקוקסון הוא סכום הדרגות הממוצעות בקבוצה השנייה (סיימו מזמן), כלומר:

$$\begin{aligned}\tilde{W}_s &= \tilde{s}_1\cdot 2+\tilde{s}_2\cdot 9+\tilde{s}_3\cdot 5+\tilde{s}_4\cdot 11+\tilde{s}_5\cdot 5\\ &=2.5\cdot 2+12\cdot 9+25.5\cdot 5+39\cdot 11+49\cdot 5=914.5\end{aligned}$$

לשם חישוב מובהקות התוצאה יש למצוא את התוחלת ואת השונות של $\tilde{W}_s$. גודלי המדגם כאן הם $n=32$, $m=19$, וכן $N=51$.

התוחלת:
$$E\tilde{W}_s=EW_s=\frac{n(N+1)}{2}=\frac{32(52)}{2}=832$$

לפי נוסחה (31) השונות היא

$$Var(\tilde{W}_s)=\frac{mn}{12N(N-1)}\left[N(N^2-1)-\sum_r t_r(t_r^2-1)\right]$$

הביטוי שהוא פונקציה של קבוצות התיקו בלבד:

$$\begin{aligned}\sum_r t_r(t_r^2-1) &= 4(4^2-1)+15(15^2-1)+12(12^2-1)\\ &+15(15^2-1)+5(5^2-1)=8{,}616\end{aligned}$$

השונות המתקבלת היא

$$Var(\tilde{W}_s)=\frac{19\cdot 32}{12\cdot 51\cdot 50}\left[51(51^2-1)-8{,}616\right]=2{,}463.47$$

השונות שהייתה מתקבלת ללא ערכי תיקו עבור גודלי המדגם הללו היא

$$Var(W_s)=\frac{mn(N+1)}{12}=\frac{19\cdot 32\cdot 52}{12}=2{,}634.67$$

הפער היחסי לעומת השונות של W_s הוא

$$\frac{Var(W_s)-Var(\tilde{W}_s)}{Var(W_s)}=\frac{\sum_r t_r(t_r^2-1)}{N(N^2-1)}=\frac{8{,}616}{51(51^2-1)}=0.065$$

הפער היחסי קצת גבוה מזה שהתקבל בדוגמה 2.11. זוהי תוצאה של קבוצות תיקו יחסית גדולות.

מובהקות התוצאה מתקבלת על סמך קירוב נורמלי.

$$P=P_{H_0}(\tilde{W}_s\ge 914.5)\cong 1-\Phi\left(\frac{914-832}{\sqrt{2{,}463.47}}\right)=1-\Phi(1.65)=0.0495$$

(עשינו תיקון רציפות של $1/2$, למרות שכלל לא ברור מהו הערך האפשרי הקודם

לתוצאה 914.5 שהתקבלה בניסוי.)
עבור רמת מובהקות של 5% יש לדחות את השערת האפס ולהסיק שכנראה התלמידים שסיימו את לימודי התואר הראשון לפני יותר מ-5 שנים נוטים לדווח על קושי גדול יותר לעומת התלמידים שסיימו לפני 5 שנים או פחות.

נראה עתה מה קורה בבעיה שבה הרבה מאוד ערכי תיקו (כחצי מהתצפיות זהות). נניח שבניסוי N תצפיות דיכוטומיות, כל אחת מקבלת את הערך 0 או 1, ונניח ששתי קבוצות התיקו הן בגודל זהה: $t_1 = t_2 = N/2$. הפער היחסי של שונות סכום הדרגות הממוצעות לעומת המקרה ללא תיקו, הוא

$$\frac{Var(W_s)-Var(\tilde{W}_s)}{Var(W_s)}=\frac{1}{N(N^2-1)}\sum_r t_r(t_r^2-1)=\frac{N^2-4}{4(N^2-1)}$$

הפער היחסי הזה שואף ל-$1/4$ כאשר N שואף לאינסוף, אבל גם עבור מדגמים קטנים הוא בסביבות הערך הזה.
אם, למשל, $N=10$, הפער היחסי הוא

$$\frac{N^2-4}{4(N^2-1)}=\frac{10^2-4}{4(10^2-1)}=\frac{96}{4(99)}=0.2424$$

ואם $N=100$, הפער היחסי הוא

$$\frac{100^2-4}{4(100^2-1)}=\frac{9996}{4(9999)}=0.2499$$

מכאן רואים שהפער היחסי בין השונות בניסוי עם כל כך הרבה ערכי תיקו לבין ניסוי עם מדידות רציפות הוא של כ-25%.

מבחן מאן-וויטני במקרה של ערכי תיקו

את הסטטיסטי של מאן-וויטני עבור מקרי תיקו מגדירים באופן הבא.
הגדרה 2.3. נגדיר את המשתנים ה"מציינים" להשוואת שני ערכים על ידי

$$\tilde{U}_{ij}=\begin{cases}1 & X_i<Y_j\\ 1/2 & X_i=Y_j\\ 0 & X_i>Y_j\end{cases}\qquad i=1,\ldots,m;\ j=1,\ldots,n \tag{32}$$

ואת הסטטיסטי של מאן-וויטני כסכום המשתנים המציינים הללו $\tilde{W}_{xy}=\sum_{i=1}^{m}\sum_{j=1}^{n}\tilde{U}_{ij}$.

$\tilde{W}_{xy}$ הוא למעשה מספר הזוגות (x_i,y_j) שעבורם $x_i<y_j$ ועוד חצי ממספר הזוגות שעבורם $x_i=y_j$.

משפט 2.7. בין $\tilde{W}_{xy}$ לבין $\tilde{W}_s$ קיים אותו הקשר כמו בין W_{xy} לבין W_s (משפט

2.5). דהיינו, $\tilde{W}_{xy}=\tilde{W}_S-n(n+1)/2$, כאשר $\tilde{W}_S$ הוא סכום הדרגות הממוצעות במדגם בגודל n.
ההוכחה נמצאת בנספח 2ב.

2.7 עוצמת מבחן ווילקוקסון (מאן-וויטני)*

אחת התכונות החשובות של מבחנים היא העוצמה של המבחן, המבטאת את הסיכוי להוכחה של תיאוריה חדשה, או לגילוי חדש ומעניין.

הגדרה 2.4. העוצמה של מבחן לבדיקת השערות היא ההסתברות לדחיית השערת האפס, תחת המודל האלטרנטיבי. נסמן את העוצמה: $\pi=P_{H_1}\{H_0$ דחיית$\}$.

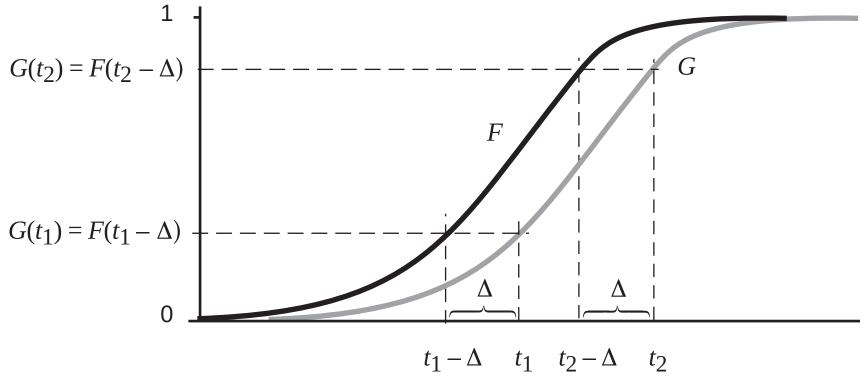
ניתן לרשום את העוצמה על ידי $\pi=1-\beta$, כאשר β היא ההסתברות לטעות מסוג שני: $\beta=P_{H_1}\{H_0$ קבלת$\}$.
העוצמה היא ההסתברות לעשות החלטה נכונה בעת דחיית השערת האפס.
נראה כאן כיצד ניתן למצוא את העוצמה של מבחן ווילקוקסון עבור מודל הזזה, שאותו נגדיר להלן.

מודל הזזה
מקרה פרטי של המודל שבו Y גדול סטוכסטית מ-X הוא מודל הזזה, המניח שהמשתנים X ו-Y הם משתנים בעלי התפלגויות דומות, שכל ההבדל ביניהן הוא בהזזה בלבד.
מודל הזזה: נניח שניתן לרשום את G כפונקציה של F באופן הבא:

$$G(t)=F(t-\Delta) \quad \text{לכל } t, \qquad (33)$$

הגודל Δ נקרא **פרמטר ההזזה.**

יש לשים לב שלפי מודל ההזזה (33) אין הבדל קבוע של Δ יחידות בין שתי ההתפלגויות, אלא הגודל Δ הוא המרחק שבין שני ערכים שווים של F ו-G. כלומר, ההתפלגות G בנקודה מסוימת t^* שווה לערכה של ההתפלגות F בנקודה $t^*-\Delta$. ניתן לראות דוגמה לכך בציור 2.8.



ציור 2.8. *הדגמה של מודל ההזזה עם פרמטר Δ*

טענה 2.3. תחת מודל הזזה (33) קיימים היחסים הבאים:

(1) $Y-\Delta \sim F$; (2) $EY = EX+\Delta$; (3) $Med(Y)=Med(X)+\Delta$.

*הוכחה**:

(1) לפי הנחת המודל, $P(Y-\Delta \le t)=P(Y\le t+\Delta)=G(t+\Delta)=F(t)$.

(2) בהנחה ש-F ו-G רציפות בעלות צפיפויות f ו-g, בהתאמה, אזי על סמך מודל ההזזה ניתן לרשום:

$$g(t) = \frac{dG(t)}{dt} = \frac{dF(t-\Delta)}{dt} = f(t-\Delta)$$

כלומר, גם פונקציית הצפיפות g מהווה הזזה פשוטה של פונקציית הצפיפות f.

חישוב התוחלת של Y הוא, לפיכך

$$EY = \int_{-\infty}^{\infty} tg(t)dt = \int_{-\infty}^{\infty} tf(t-\Delta)dt = \int_{-\infty}^{\infty}(x+\Delta) f(x)dx$$

$$= \int_{-\infty}^{\infty} (x) f(x)dx + \Delta \int_{-\infty}^{\infty} f(x)dx = EX + \Delta$$

(למי שלא מצוי באינטגרציה, התוחלת היא "מרכז הכובד" של פונקציית הצפיפות, והיות שהצפיפות g היא בעלת צורה זהה לצפיפות f, אולם מוזזת ב-Δ יחידות, גם התוחלת שלה מוזזת באותו גודל.)

(3) לגבי החציון, הטענה נובעת ישירות מהגדרת המודל. נסמן ב-$x_{1/2}$ את החציון של המשתנה X וב-$y_{1/2}$ את החציון של המשתנה Y. אזי קיים:

$1/2=G(y_{1/2})=F(y_{1/2}-\Delta)$ ולכן $y_{1/2}-\Delta$ הוא החציון של ההתפלגות F, כלומר, חציון המשתנה X: $y_{1/2}-\Delta=x_{1/2}$ ומכאן נובעת הטענה. ♣

הערה: כמו לגבי החציון, גם כל ערכי החלוקה האחרים של המשתנה Y שווים לערכי החלוקה המתאימים של X לאחר הוספת פרמטר ההזזה Δ. כלומר, $y_p = x_p + \Delta$ לכל $0 < p < 1$. ההוכחה זהה למקרה ש- $p = 1/2$ (הוכיחו!).

בהתאם למודל ההזזה (33), נסתכל על ההשערות:

$$H_0 : \Delta = 0 \qquad H_1 : \Delta > 0 \tag{34}$$

תחת השערת האפס שתי ההתפלגויות שוות, ותחת האלטרנטיבה המשתנה Y גדול סטוכסטית מ-X, היות שעבור כל t מתקיים: $G(t) = F(t - \Delta) \le F(t)$.

נרצה עתה לחשב את העוצמה של מבחן ווילקוקסון. לבדיקת הבעיה (34) יש לדחות את השערת האפס עבור ערכים גבוהים של W_s (הסטטיסטי של ווילקוקסון), או ערכים גבוהים של W_{xy} (הסטטיסטי של מאן-וויטני). עבור רמת מובהקות דרושה α, נניח שדוחים כאשר $W_{xy} \ge c$, כלומר, c הוא הערך הקריטי לפי מבחן מאן-וויטני.

עלינו לחשב, אפוא, את הביטוי הבא:

$$\pi(\Delta) = P_\Delta(W_{xy} \ge c) \tag{35}$$

מובן שהעוצמה (הסיכוי לדחות את השערת האפס תחת המודל האלטרנטיבי) תלויה בגודל Δ. לכן זו למעשה פונקציה של Δ והיא נקראת **פונקציית העוצמה**. הפונקציה מוגדרת היטב גם עבור ערכים שליליים של הפרמטר, ולכן היא מוגדרת היטב לכל ערך של $-\infty < \Delta < \infty$ (נִראה יותר מאוחר שהעוצמה תלויה גם בהתפלגות הבסיסית F).

חישוב העוצמה המדויקת מסובך מאוד. אנו נביא כאן את העוצמה המקורבת, עבור מדגמים גדולים (נקראת עוצמה אסימפטוטית).

טענה 2.4. ההתפלגות של הסטטיסטי של ווילקוקסון (או מאן-וויטני) היא אסימפטוטית נורמלית גם כאשר השערת האפס איננה נכונה. כלומר, עבור m ו-n די גדולים ההתפלגות של המשתנה המתוקנן $\dfrac{W_{xy} - EW_{xy}}{\sqrt{Var(W_{xy})}}$ קרובה להתפלגות נורמלית סטנדרטית, גם כאשר שתי ההתפלגויות F ו-G אינן זהות, בתנאי שמתקיים $0 < P(X < Y) < 1$.

לא נוכיח כאן את הטענה.

הדרישה שההסתברות לכך ש-Y גדול מ-X לא תהיה שווה ל-0 וגם לא תהיה שווה ל-1 ברורה. אם, למשל, $P(X < Y) = 1$, אזי נקבל בוודאות $W_{xy} = mn$ (בכל ההשוואות של הזוגות אותה תוצאה). ולכן הסטטיסטי של מאן-וויטני מקבל את הערך mn בהסתברות 1, כלומר, המשתנה מנוון. התפלגותו במקרה זה איננה נורמלית ולא חשוב מהם גודלי

המדגמים. אותו שיקול נכון גם כאשר $P(X<Y)=0$, אז $W_{xy}=0$ בוודאות.

כדי להשתמש בטענה 2.4 יש לדעת כיצד ניתן לחשב את התוחלת ואת השונות של W_{xy} כאשר ההתפלגויות אינן שוות. בנספח 2ג נראה באופן כללי את צורת המומנטים הללו. הבעיה בשימוש בנוסחה (17) מנספח 2ג היא, שחישוב הפרמטרים המעורבים הוא בדרך כלל מאוד מסובך וקשור בהתפלגות משותפת של שלושה משתנים בלתי תלויים, שאחד מהם בעל התפלגות השונה משני האחרים.

לפיכך נמצא כאן את השונות רק למקרה של מודל הזזה (33), וכן נחשב אותה רק בקירוב, עבור מדגמים מאוד גדולים, ועבור פרמטר Δ מאוד קטן.

עוצמה מקורבת של מבחן המבוסס על סטטיסטי בעל התפלגות אסימפטוטית נורמלית

נסתכל באופן כללי על מבחן בעל אזור דחייה מהצורה $\{T\ge c\}$, כאשר T מתפלג אסימפטוטית נורמלית. כדי שהמבחן יהיה בעל רמת מובהקות מקורבת α, הערך הקריטי c צריך לקיים את השוויון

$$\alpha=P_{H_0}(T\ge c)\cong 1-\Phi\left(\frac{c-E_0T}{\sqrt{Var_0(T)}}\right)=1-\Phi\left(\frac{c-E_0T}{\sigma_0(T)}\right) \tag{36}$$

כאשר התוחלת והשונות לעיל מחושבות תחת השערת האפס. לפי זה הערך של c הוא

$$c=E_0T+z_{1-\alpha}\sigma_0(T) \tag{37}$$

נרשום את העוצמה בקירוב.

$$\pi=P_{H_1}(T\ge c)\cong 1-\Phi\left(\frac{c-E_1T}{\sqrt{Var_1(T)}}\right)=1-\Phi\left(\frac{c-E_1T}{\sigma_1(T)}\right) \tag{38}$$

כאן התוחלת והשונות מחושבות תחת המודל האלטרנטיבי.

נציב את הערך של c מנוסחה (37) ונקבל את העוצמה המקורבת

$$\begin{aligned}\pi&\cong 1-\Phi\left(\frac{[E_0T+z_{1-\alpha}\sigma_0(T)]-E_1T}{\sigma_1(T)}\right)\\&=1-\Phi\left(\frac{E_0T-E_1T}{\sigma_1(T)}+z_{1-\alpha}\frac{\sigma_0(T)}{\sigma_1(T)}\right)\\&=\Phi\left(\frac{E_1T-E_0T}{\sigma_1(T)}-z_{1-\alpha}\frac{\sigma_0(T)}{\sigma_1(T)}\right)\end{aligned} \tag{39}$$

את השוויון האחרון קיבלנו על סמך הסימטרייה של ההתפלגות הנורמלית הסטנדרטית סביב אפס.

עצמת מבחן מאן-וויטני (ווילקוקסון)

נשתמש בנוסחה (39) לחישוב העוצמה של W_{xy}, כיוון שהתפלגותו האסימפטוטית היא נורמלית. במודל הזזה האלטרנטיבה מוגדרת על ידי פרמטר ההזזה Δ. נוכל, אפוא, לרשום את הנוסחה (39) כפונקציה של Δ:

$$\pi(\Delta) \cong \Phi\left(\frac{E_\Delta T - E_0 T}{\sigma_\Delta(T)} - z_{1-\alpha}\frac{\sigma_0(T)}{\sigma_\Delta(T)}\right) \qquad (40)$$

עבור $T = W_{xy}$, התוחלת והשונות של W_{xy} תחת H_0 הן, לפי נוסחאות (16) ו-(17):

$$E_0 T = E_0 W_{xy} = \frac{mn}{2} \qquad \sigma_0^2(T) = Var_0(W_{xy}) = \frac{mn(N+1)}{12}$$

נסמן את הפרמטר

$$p_\Delta = P_\Delta(X < Y) \qquad (41)$$

את התוחלת תחת האלטרנטיבה ניתן לרשום באמצעות p_Δ באופן הבא:

$$E_\Delta W_{xy} = E_\Delta \sum_{i=1}^{m}\sum_{j=1}^{n} U_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} E_\Delta U_{ij} = mnP_\Delta(X<Y) = mnp_\Delta \qquad (42)$$

מכאן העוצמה (40) המתקבלת עבור הסטטיסטי של מאן-וויטני היא

$$\pi(\Delta) \cong \Phi\left(\frac{mn\cdot(p_\Delta - 1/2)}{\sigma_\Delta(W_{xy})} - z_{1-\alpha}\frac{\sigma_0(W_{xy})}{\sigma_\Delta(W_{xy})}\right) \qquad (43)$$

בכדי להשתמש בנוסחה זו נצטרך לחשב את הפרמטר p_Δ ואת השונות במודל האלטרנטיבי, עבודה הדורשת טרחה רבה. לפיכך נעשה כאן קירוב גס יותר, שאותו קל יחסית לחשב.

הנחה 2.1 נניח שלכל m ו-n, השונות $Var_\Delta(W_{xy})$ רציפה ב-Δ.

נמצא את העוצמה המקורבת של מבחן ווילקוקסון בהנחה 2.1, עבור ערכי Δ הקרובים לאפס.

את הפרמטר p_Δ ניתן לרשום באופן הבא:

$$p_\Delta = P_\Delta(X<Y) = P_\Delta\{X-(Y-\Delta)<\Delta\} = P(X-Y'<\Delta) = F^*(\Delta)$$

כבר ראינו (טענה 2.3) שלמשתנה $Y' = Y - \Delta$ יש התפלגות זהה להתפלגות של X. לכן F^* היא התפלגות ההפרש בין שני משתנים בלתי תלויים, שניהם בעלי התפלגות F.

טענה 2.5 F^* היא התפלגות סימטרית סביב אפס.

הוכחה: יהיו X_1 ו-X_2 משתנים בלתי תלויים שווי התפלגות. מטעמי סימטרייה התפלגות המשתנה $X_1 - X_2$ זהה להתפלגות של $X_2 - X_1$. לכן מתקיים:

$$F^*(t) = P(X_1 - X_2 \le t) = P(X_2 - X_1 \le t)$$

$$= P(X_1 - X_2 \geq -t) = 1 - F^*(t)$$

ומכאן הסימטרייה. ♣

נפַתח את F^* לטור טיילור סביב 0, בהנחה ש-F^* בעלת פונקציית צפיפות f^*, וניקח רק את שני האיברים הראשונים של הטור. אנו מניחים ש-Δ קרוב לאפס, ולכן האיברים הנוספים בטור הם זניחים. נקבל לפי זה קירוב עבור p_Δ:

$$p_\Delta = F^*(\Delta) \cong F^*(0) + \Delta f^*(0)$$

בגלל הסימטרייה של F^* סביב אפס, מתקיים $F^*(0) = 1/2$ ומכאן:

$$p_\Delta \cong 1/2 + \Delta f^*(0)$$

או

$$p_\Delta - 1/2 \cong \Delta f^*(0) \qquad (44)$$

לגבי השונות תחת האלטרנטיבה, בגלל רציפות השונות (הנחה 2.1) מתקיים:

$$\frac{\sigma_0(W_{xy})}{\sigma_\Delta(W_{xy})} \xrightarrow[\Delta \to 0]{} 1 \qquad (45)$$

נשתמש בגבול הזה ונציב את התוצאה (44) בנוסחה (43). מכאן נקבל את העוצמה האסימפטוטית של מבחן מאן-וויטני:

$$\pi(\Delta) \cong \Phi\left(\frac{mn \cdot (p_\Delta - 1/2)}{\sigma_\Delta(W_{xy})} - z_{1-\alpha} \frac{\sigma_0(W_{xy})}{\sigma_\Delta(W_{xy})} \right)$$

$$\cong \Phi\left(\frac{mn\Delta f^*(0)}{\sigma_0(W_{xy})} - z_{1-\alpha} \right) = \Phi\left(\frac{mn\Delta f^*(0)}{\sqrt{mn(N+1)/12}} - z_{1-\alpha} \right)$$

קיבלנו, אפוא, את נוסחת העוצמה המקורבת של מבחן מאן-וויטני (או ווילקוקסון)

$$\pi(\Delta) \cong \Phi\left(\frac{\sqrt{12mn}}{\sqrt{(N+1)}} \cdot f^*(0) \cdot \Delta - z_{1-\alpha} \right) \qquad (46)$$

החישוב המקורב של העוצמה נערך בשני שלבים. ראשית, הסתכלנו על מדגמים גדולים מספיק, באופן שניתן לעשות קירוב נורמלי להתפלגות סטטיסטי המבחן. שנית, הסתכלנו על ערכי Δ קטנים מספיק, באופן שהשונות תחת המודל האלטרנטיבי קרובה מאוד לשונות תחת השערת האפס, וכן בטור טיילור האיברים שבהם החזקה של Δ היא 2 או יותר הם זניחים.

הנוסחה (46) מתאימה, אפוא, במקרים של מדגמים גדולים, כאשר יש לחשב עוצמה עבור פער קטן בין הטיפול לביקורת. העוצמה לעיל תלויה, כמובן, בגודל הפער Δ בין התפלגות הטיפול להתפלגות הביקורת, אולם היא תלויה גם בהתפלגות F של תצפיות קבוצת הביקורת, כלומר, במודל ההסתברותי של הבעיה.

הערה: בדרך כלל חישוב מדויק של הגודל הספציפי של העוצמה אינו כה חשוב. זאת מכיוון ששימוש בנוסחת העוצמה נעשה לרוב כדי לנסות לקבוע מראש גודלי מדגמים שעבורם נוכל לגלות את התיאוריה שלנו (דחיית השערת האפס) בסיכוי גדול מספיק. כך, אם הקירוב אינו טוב מספיק, תהיה לנו טעות מסוימת בגודל המדגם הדרוש, אולם בדרך כלל סדר הגודל של המדגם שייבחר יהיה מתאים. למשל, אם נקבל דרישה לגודל מדגם של כ-100 תצפיות במקום 98 או 105, הטעות אינה משמעותית ביותר.

דוגמה 2.13 נחשב את העוצמה המקורבת של מבחן ווילקוקסון במודל הנורמלי. נניח ש-$X \sim N(\mu,\sigma^2)$, כלומר F היא התפלגות נורמלית. התפלגות ההפרש בין שני משתנים נורמלים בלתי־תלויים כאלה היא נורמלית $N(0,2\sigma^2)$ (תוחלת ההפרש היא 0 והשונות היא סכום השונויות). פונקציית הצפיפות של ההפרש היא, אפוא,

$$f^*(t)=\frac{1}{\sqrt{2\pi\cdot 2\sigma^2}}e^{-t^2/4\sigma^2}=\frac{1}{\sqrt{4\pi\sigma^2}}e^{-t^2/4\sigma^2}$$

ולכן

$$f^*(0)=\frac{1}{\sqrt{4\pi\sigma^2}} \tag{47}$$

נוסחת העוצמה המקורבת של מבחן ווילקוקסון במודל הנורמלי מתקבלת על ידי הצבת (47) בנוסחה (46).

$$\begin{aligned}\pi(\Delta)&\cong\Phi\left(\frac{\sqrt{12mn}}{\sqrt{(N+1)}}\cdot\frac{1}{\sqrt{4\pi\sigma^2}}\cdot\Delta-z_{1-\alpha}\right)\\&=\Phi\left(\frac{\sqrt{3mn}}{\sqrt{(N+1)\pi\sigma^2}}\cdot\Delta-z_{1-\alpha}\right)\end{aligned} \tag{48}$$

זו נוסחה כללית לחישוב עוצמת מבחן ווילקוקסון במודל הנורמלי.
עוצמה זו עולה כפונקציה של Δ ויורדת כפונקציה של σ^2. ניתן לראות תאור גרפי של פונקציית העוצמה הזאת בציור 2.9 בהמשך. מבחינת גודלי המדגמים, אם שני המדגמים הם באותו סדר גודל, אזי העוצמה עולה כפונקציה של גודלי המדגמים והיא מקסימלית כאשר שני המדגמים באותו גודל (ראו תרגיל 18).

נסתכל עתה על דוגמה מעשית.

דוגמה 2.14 לחברה גדולה מאוד יש סניפים בשתי ערים שונות. באחד הסניפים (סניף א) מינו עובדת סוציאלית שתפקידה לטפל בבעיות אישיות של העובדים. לאחר זמן רצו

לבדוק אם שביעות הרצון בעבודה של העובדים בסניף א גבוהה יותר מאשר בסניף ב. למחקר נבחרו 10 עובדים באופן מקרי בכל אחד משני הסניפים. נניח שציוני העובדים בשאלון "שביעות רצון בעבודה" הוא משתנה נורמלי, עם סטיית תקן של כ-1.5.

נחשב את העוצמה המקורבת של מבחן ווילקוקסון ברמת מובהקות $\alpha=.05$, עבור פער של $\Delta=1$ בין הסניפים. הצבה של הערכים $m=n=10$, $\sigma=1.5$, $z_{.95}=1.645$ בנוסחה (48) נותנת את העוצמה המקורבת:

$$\pi(1)\cong\Phi\left(\frac{\sqrt{3\cdot100}}{\sqrt{21\pi(1.5^2)}}\cdot1-1.645\right)=\Phi(-0.22)=.4129$$

העוצמה היא בערך 41%. זהו הסיכוי לגלות פער של 1 בציון שביעות הרצון של שני הסניפים. זו, כמובן, עוצמה לא גבוהה ביותר.

הערה: בחישוב העוצמה על סמך חישוב מדויק של התוחלת והשונות של סטטיסטי המבחן תחת האלטרנטיבה (ראו את החישוב המדויק בנספח 2ג), מקבלים:

$$p_\Delta=p_{\Delta=1}(X<Y)=p_{\Delta=1}(Y-X>0)$$

$$=1-\Phi\left(\frac{0-1}{\sqrt{2\cdot1.5^2}}\right)=\Phi(0.47)=.6808$$

מכאן התוחלת של הסטטיסטי של מאן-וויטני היא, לפי נוסחה (42):

$$E_{\Delta=1}W_{xy}=mnp_\Delta=10(10)(.6808)=68.08$$

השונות המדויקת של הסטטיסטי של ווילקוקסון (או מאן-וויטני) המתקבלת (החישוב מסובך יותר ולא נביא אותו כאן) היא $Var_{\Delta=1}(W_{xy})=177.45$.

נשתמש בנוסחה (40) לחישוב העוצמה. התוחלת והשונות תחת השערת האפס הן

$$E_0W_{xy}=mn(0.5)=100(.5)=50$$

$$Var_0(W_{xy})=\frac{mn(N+1)}{12}=\frac{100(21)}{12}=175$$

היחס בין שונות זו לשונות תחת האלטרנטיבה, $175/177.45=0.986$, אכן קרוב מאוד ל-1. עצמת המבחן היא בקירוב

$$\pi(1)\cong\Phi\left(\frac{68.08-50.0}{\sqrt{177.45}}-1.645\frac{\sqrt{175}}{\sqrt{177.45}}\right)=\Phi(-0.28)=.3897$$

העוצמה לפי החישוב המדויק יותר היא קצת קטנה מזו שהתקבלה בקירוב הגס.

גודל המדגם

כדי לקבוע את גודלי המדגם הדרושים לקבלת עוצמה מסוימת, נדון רק במדגמים שווים

בגודלם, כלומר, $m=n=N/2$. אין טעם לחפש מדגמים בגדלים שונים זה מזה, וכך החישוב פשוט יחסית. נציב $m=n$ בנוסחה הכללית של העוצמה (46) ונקבל:

$$\pi(\Delta)\cong\Phi\left(\frac{\sqrt{12n^2}}{\sqrt{(2n+1)}}\cdot f^*(0)\cdot\Delta-z_{1-\alpha}\right)$$

$$\cong\Phi\left(\sqrt{6n}\cdot f^*(0)\cdot\Delta-z_{1-\alpha}\right)$$

אנו מניחים ש-n שיתקבל יהיה די גדול וכך יתקיים $\sqrt{\frac{n^2}{2n+1}}\approx\sqrt{\frac{n}{2}}$.

כדי שעוצמה זו תהיה לפחות π^*, דרוש ש-n יקיים (בדקו!):

$$\sqrt{n}\geq\frac{z_{\pi^*}+z_{1-\alpha}}{\sqrt{6}f^*(0)\cdot\Delta}$$

ולכן גודל המדגם הדרוש לכל קבוצה הוא

$$n\geq\frac{(z_{\pi^*}+z_{1-\alpha})^2}{6\left[f^*(0)\right]^2\cdot\Delta^2}\tag{49}$$

במקרה הפרטי של מודל נורמלי, עם $f^*(0)$ הנתונה בנוסחה (47), מקבלים:

$$n\geq\frac{(z_{\pi^*}+z_{1-\alpha})^2}{6\left[\frac{1}{4\pi\sigma^2}\right]\cdot\Delta^2}=\frac{2\pi\sigma^2(z_{\pi^*}+z_{1-\alpha})^2}{3\Delta^2}\tag{50}$$

גודל המדגם הדרוש הוא פונקציה יורדת של Δ, כלומר, ככל שהפער בין האוכלוסיות קטן, כך נזדקק למדגמים גדולים יותר כדי לגלות זאת.

דוגמה 2.15 (המשך דוגמה 2.14). בבעיית שביעות הרצון בעבודה קיבלנו עוצמה קטנה עבור מדגמים בגודל 10. נחפש את גודלי המדגם הדרושים כדי שעצמת המבחן, עבור $\Delta=1$, תהיה לפחות 80%. אנו מניחים מודל נורמלי של הציונים. נציב בנוסחה (50) את הערכים: $\alpha=.05$, $\pi^*=.80$, $\sigma=1.5$, $\Delta=1$ ונקבל:

$$n\geq\frac{2\pi(1.5^2)(z_{.80}+z_{.95})^2}{3(1^2)}=\frac{2\pi(2.25)(0.84+1.645)^2}{3}=29.1$$

לפיכך יש לבחור לפחות 30 איש בכל אחד מהסניפים כדי לקבל את העוצמה הדרושה. ברור שכדי לקבל הערכה לגודל המדגם הדרוש אין כל צורך לחשב את העוצמה בצורה המדויקת יותר, כמו בנספח 2ג. בחישוב מדוקדק שעשינו במקרה זה, העוצמה המתקבלת עבור 30 תצפיות בכל מדגם היא קצת קטנה מן הנדרש – $\pi=.773$ וגודלי המדגמים הדרושים לקבלת עוצמה של 80% הם $m=n=33$, שם $\pi=.807$. ההבדל שמתקבל אינו מצדיק את המאמץ.

עוצמה אסימפטוטית של מבחן t

המבחן הפרמטרי המקובל והמתאים ביותר (בעל עוצמה מקסימלית במשפחה גדולה של מבחנים) להשוואת שני מדגמים בלתי תלויים במודל נורמלי הוא המבחן המתבסס על הפרש ממוצעי שני המדגמים. המבחן נקרא **מבחן t לשני מדגמים**

הנחות המודל: $X_i \sim N(\mu_X, \sigma^2)$, $i=1,\ldots,m$; $Y_j \sim N(\mu_Y, \sigma^2)$, $j=1,\ldots,n$.
כמו כן, כל $N=m+n$ התצפיות הן בלתי תלויות. כלומר, שני המדגמים הם בלתי תלויים זה בזה ולקוחים מהתפלגויות נורמליות בעלות אותה שונות.
מודל זה שקול למודל הזזה עם פרמטר $\Delta=\mu_Y-\mu_X$.
במודל זה המשתנה Y גדול סטוכסטית מ-X, אם $\mu_Y>\mu_X$ (ראו דוגמה 2.6).
השערת האפס הנבדקת היא $H_0: \mu_X=\mu_Y$, או $H_0: \Delta=0$, והאלטרנטיבה יכולה להיות חד־כיוונית או דו־כיוונית.
לא נציג פה את התיאוריה של בניית הסטטיסטי של מבחן t ואת התפלגותו, אלא נציג את המבחן בקיצור. ניתן למצוא זאת בכל הספרים העוסקים בהסקה סטטיסטית.
נסתכל על המשתנה

$$Z=\frac{(\bar{Y}-\bar{X})-(\mu_Y-\mu_X)}{\sqrt{\sigma^2\left[\frac{1}{m}+\frac{1}{n}\right]}} \tag{51}$$

הפרש הממוצעים $\bar{Y}-\bar{X}$ הוא משתנה נורמלי, ובמכנה רשומה סטיית התקן של הפרש הממוצעים (הוכיחו!). לכן המשתנה Z הוא נורמלי סטנדרטי. היות שהשונות σ^2 איננה ידועה, המשתנה Z אינו סטטיסטי ואינו יכול לשמש כסטטיסטי מבחן. נסתכל, אפוא, על המשתנה

$$T=\frac{(\bar{Y}-\bar{X})-(\mu_Y-\mu_X)}{\sqrt{S^2\left[\frac{1}{m}+\frac{1}{n}\right]}} \tag{52}$$

כאשר S^2 הוא האומד עבור השונות σ^2 (אותה שונות ל-X ול-Y), וניתן על ידי הממוצע המשוקלל של אומדי השונות בכל אחד מהמדגמים בנפרד:

$$S^2=\frac{\sum_{i=1}^{m}(X_i-\bar{X})^2+\sum_{j=1}^{n}(Y_j-\bar{Y})^2}{m+n-2}=\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}$$

המשתנה T הוא משתנה מקרי שהתפלגותו איננה נורמלית, מכיוון שהמכנה שלו הוא משתנה מקרי ולא גודל קבוע. התפלגות המשתנה T נקראת **התפלגות t של "סטודנט"**, או בקיצור, **התפלגות t**, עם פרמטר $m+n-2$, הנקרא **דרגות החופש** מסמנים $T \sim t_{m+n-2}$.

קיימות טבלאות של התפלגות t בכל ספר סטטיסטיקה. איננו מביאים טבלה כזאת כאן, מכיוון שאיננו משתמשים במבחן זה, אלא לצורך הדגמת העוצמה.

עבור בדיקת ההשערות (34) ניתן להשתמש **במבחן t** המבוסס על הסטטיסטי:

$$t = \frac{\bar{Y} - \bar{X}}{\sqrt{S^2\left[\frac{1}{m} + \frac{1}{n}\right]}}$$

תחת H_0 שתי התוחלות שוות ולכן התפלגות הסטטיסטי t היא בדיוק התפלגות t עם $m+n-2$ דרגות חופש.

מבחן t לשני מדגמים ברמת מובהקות α עבור אלטרנטיבה חד־צדדית $H_1: \Delta > 0$, הוא המבחן הדוחה את השערת האפס כאשר $t > t_{1-\alpha}$. ($t_{1-\alpha}$ הוא ערך החלוקה ה-$1-\alpha$ של התפלגות t המתאימה.)

הערה: התפלגות הסטטיסטי t, כפי שהיא מובאת בטבלאות t, תלויה בהנחות המודל (נורמליות ושוויון שונויות). עם זאת, כאשר המדגמים מאוד גדולים, אומדן השונות יהיה קרוב ביותר לשונות האמיתית. כמו כן, ממוצעי המדגמים הם בעלי התפלגות אסימפטוטית נורמלית ולכן גם הפרש הממוצעים הוא בעל התפלגות אסימפטוטית נורמלית. מכאן, כאשר המדגמים גדולים מאוד, הסטטיסטי t הוא בעל התפלגות הקרובה להתפלגות נורמלית סטנדרטית.

קיבלנו בנוסחה (39) באופן כללי את העוצמה עבור מבחן המבוסס על סטטיסטי בעל התפלגות אסימפטוטית נורמלית. ניתן להשתמש בנוסחה זו גם עבור מבחן t. כדי לקבל את התוחלות והשונויות בנוסחה, נרשום את הערך של t בקירוב, כאשר המדגמים גדולים (החלפנו את אומד השונות בשונות האמיתית).

$$t \cong \frac{\bar{Y} - \bar{X}}{\sqrt{\sigma^2\left[\frac{1}{m} + \frac{1}{n}\right]}}$$

כפי שראינו קודם, משתנה זה הוא למעשה המשתנה Z מנוסחה (51), עם $\mu_Y - \mu_X = 0$. תחת השערת האפס זהו משתנה נורמלי סטנדרטי. את המודל האלטרנטיבי ניתן לרשום במונחים של פרמטר ההזזה $\Delta = \mu_Y - \mu_X$. התפלגות הסטטיסטי t היא, אפוא, נורמלית בקירוב. התוחלת היא

$$E_\Delta t \cong \frac{E_\Delta(\bar{Y} - \bar{X})}{\sqrt{\sigma^2\left[\frac{1}{m} + \frac{1}{n}\right]}} = \frac{\mu_Y - \mu_X}{\sqrt{\sigma^2\left[\frac{m+n}{mn}\right]}} = \frac{\sqrt{mn}\cdot\Delta}{\sigma\sqrt{m+n}} \qquad (53)$$

השונות של t שווה בערך ל-1 גם תחת האלטרנטיבה. נציב ערכים אלה בנוסחה (39) ונקבל:

$$\pi_t(\Delta) \cong \Phi\left(\sqrt{\frac{mn}{m+n}} \cdot \frac{\Delta}{\sigma} - z_{1-\alpha}\right) \qquad (54)$$

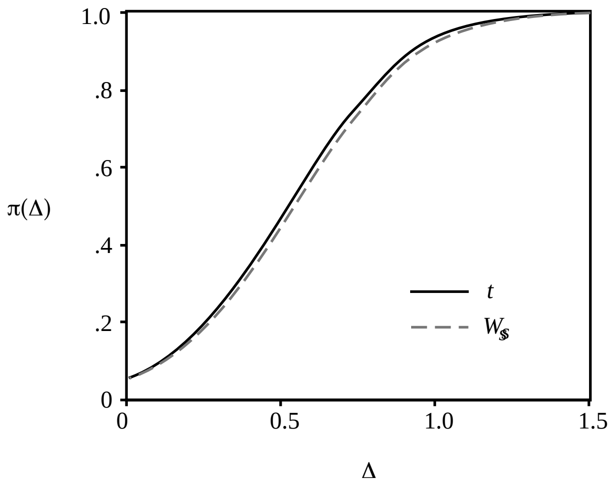
הערה: כדאי לשים לב שהעוצמה האסימפטוטית של מבחן t שמצאנו כאן **אינה תלויה** בהתפלגות F של התצפיות. זאת מכיוון שעבור מדגמים גדולים מאוד, התפלגות הסטטיסטי t היא נורמלית עם שונות השווה ל-1, באופן בלתי תלוי בהתפלגות המדויקת של האוכלוסייה שממנה הוצאו התצפיות.

דוגמה 2.16 (המשך דוגמה 2.14). נחשב את העוצמה האסימפטוטית של מבחן t בתנאי הבעיה של שביעות הרצון בעבודה. הערכים הנתונים הם $\alpha = .05$, $\sigma = 1.5$, $\Delta = 1$. נציב בנוסחה (54) ונקבל:

$$\pi_t(1) \cong \Phi\left(\frac{\sqrt{100}}{\sqrt{20}} \cdot \frac{1}{1.5} - 1.645\right) = \Phi(-0.15) = .4404$$

אם משתמשים במבחן ווילקוקסון, מקבלים (דוגמה 2.14) עוצמה מקורבת של 0.4129, שהיא קצת יותר נמוכה, אולם ההבדל אינו גדול.

בציור 2.9 מתוארת פונקציות העוצמה של מבחן ווילקוקסון בהשוואה לזו של מבחן t עבור מודל נורמלי, עם $m = n = 20$, $\alpha = .05$, $\sigma = 1$. אנו רואים שפונקציית העוצמה של מבחן t גבוהה מזו של מבחן ווילקוקסון לכל ערך של Δ, אולם הפער בין שתי הפונקציות קטן ביותר.



ציור 2.9. *פונקציית העוצמה של מבחן ווילקוקסון ושל מבחן t במודל נורמלי*

גודל המדגם במבחן t

נסתכל שוב על שני מדגמים באותו גודל $m=n$. להשגת עוצמה π^* יש לרשום את נוסחה(54) עבור המקרה הזה ולחלץ את n. מקבלים (קל לבדוק):

$$n \geq \frac{2\sigma^2(z_{\pi^*}+z_{1-\alpha})^2}{\Delta^2} \qquad (55)$$

דוגמה 2.17 (המשך דוגמה 2.14). גודל המדגם הדרוש להשגת עוצמה של 80% בתנאי הבעיה הוא

$$n \geq \frac{2(1.5^2)(z_{.80}+z_{.95})^2}{1^2} = 2(2.25)(0.84+1.645)^2 = 27.8$$

כלומר, יש לקחת לפחות 28 תצפיות.

שימו לב שזה נכון בכל מודל, ובפרט במודל הנורמלי, שעבורו מצאנו שלמבחן ווילקוקסון דרושות 29.1 תצפיות (ולמעשה נזדקק ל-30 תצפיות לפחות). ההבדל בין שני המבחנים מבחינת גודלי המדגמים הדרושים אינו גדול.

יעילות יחסית אסימפטוטית

בספר זה לא נוכל לדון באופן כללי ובצורה מדויקת במושג המורכב של היעילות האסימפטוטית של פיטמן. ננסה לתת פה על קצה המזלג את האינטואיציה העומדת מאחורי ההגדרה של פיטמן.

נסתכל על סטטיסטי מבחן T המתאים לבדיקת השערות בבעיה מסוימת. נסמן ב- n_T את גודל המדגם (האסימפטוטי) הדרוש להשגת עוצמה קבועה π^* במבחן ברמת מובהקות קבועה α, עבור מודל אלטרנטיבי מסוים. המושג "אסימפטוטי" כאן הוא במובן שהשתמשנו בו בפרק זה, כלומר, גודל המדגם המקורב, כאשר הפער בין המודל בהשערת האפס לבין המודל האלטרנטיבי הוא קטן והמדגם המתקבל הוא גדול יחסית.

הגדרה 2.5. היעילות היחסית האסימפטוטית (Asymptotic relative efficiency, או בקיצור ARE) בין שני מבחנים T_1 ו-T_2 מוגדרת על ידי היחס בין גודלי המדגם האסימפטוטיים הדרושים להשגת עוצמה קבועה π^* במבחן ברמת מובהקות קבועה α, עבור מודל אלטרנטיבי מסוים:

$$ARE(T_1,T_2) = \frac{n_{T_2}}{n_{T_1}} \qquad (56)$$

במִלים אחרות, היעילות היחסית היא המנה בין גודלי המדגם הדרושים בשני המבחנים. שימו לב שגודל המדגם של הסטטיסטי הראשון מופיע במכנה ושל השני – במונה. כך, אם המנה גדולה מ-1, למבחן T_2 דרושות יותר תצפיות, ולכן T_1 יעיל יותר. אם המנה

קטנה מ-1 אזי המבחן T_2 יעיל יותר.

לדוגמה, עבור הבעיה בדוגמה 2.15, מצאנו לגבי ערכים ספציפיים במודל הנורמלי, את גודלי המדגם עבור מבחן ווילקוקסון – $n_{W_s}=29.1$ (דוגמה 2.15), ועבור מבחן t – $n_t=27.3$ (דוגמה 2.17). (שימו לב שחישוב היעילות האסימפטוטית נעשה על סמך גודלי המדגם כפי שהתקבלו בנוסחה, לפני העיגול לערך שלם.) היעילות היחסית האסימפטוטית של מבחן ווילקוקסון ביחס למבחן t היא:

$$ARE(W_s,t)=\frac{n_t}{n_{W_s}}=\frac{27.8}{29.1}=0.955$$

היעילות שהתקבלה אמנם קטנה מ-1, אבל למעשה היא קרובה מאוד ל-1.

מסקנה: גם במודל הנורמלי, שעבורו מבחן t הוא המבחן המתאים ביותר (בעל העוצמה המקסימלית מבין קבוצה גדולה של מבחנים), מבחן ווילקוקסון הוא בעל עוצמה גבוהה יחסית, ולכן אין צורך בהרבה יותר תצפיות להשגת אותה עוצמה בהשוואה למבחן t.

הערה: אם היינו מחשבים את המנה של גודלי המדגם השלמים שיש לבחור, היינו מקבלים תוצאה מעט קטנה יותר: $28/30=0.933$.

הערה נוספת: ניתן לחשוב שהיעילות הגבוהה לעיל התקבלה עבור הערכים הספציפיים של α, π^*, σ ו-Δ. מיד נראה שלא כך הדבר, ולמעשה היעילות היחסית בין מבחן ווילקוקסון למבחן t במודל הנורמלי היא קבועה ואינה תלויה בערכים הספציפיים של הפרמטרים הללו.

נסתכל על המקרה הכללי ונרשום את היעילות היחסית של מבחן ווילקוקסון ביחס למבחן t, לפי הנוסחאות (49) ו- (55) של גודלי המדגם שקיבלנו קודם.

$$ARE(W_s,t)=\frac{n_t}{n_{W_s}}=\frac{\dfrac{2\sigma^2(z_{\pi^*}+z_{1-\alpha})^2}{\Delta^2}}{\dfrac{(z_{\pi^*}+z_{1-\alpha})^2}{6\left[f^*(0)\right]^2\cdot\Delta^2}}=12\sigma^2\left[f^*(0)\right]^2$$

קיבלנו שבאופן כללי, במודל כלשהו, היעילות היחסית האסימפטוטית של מבחן ווילקוקסון ביחס למבחן t היא

$$ARE(W_s,t)=12\sigma^2\left[f^*(0)\right]^2 \qquad (57)$$

אנו רואים, אפוא, שהיעילות היחסית איננה תלויה ברמת המובהקות α או בעוצמה

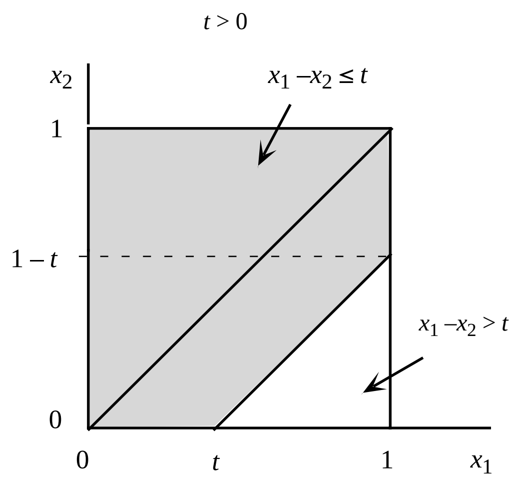
הדרושה π^*, וגם לא בפרמטר ההזזה Δ. ישנם מודלים הסתברותיים שעבורם מבחן t יעיל יותר ואחרים שעבורם דווקא מבחן ווילקוקסון (המבוסס על הדרגות בלבד) יעיל יותר. נסתכל עתה על כמה מודלים.

דוגמה 2.18 מודל נורמלי: כבר ערכנו את כל החישובים לגבי המודל הנורמלי. נציב את $f^*(0)$ לפי נוסחה (47) ונקבל:

$$ARE(W_s,t)=12\sigma^2\frac{1}{4\pi\sigma^2}=\frac{3}{\pi}\cong 0.955$$

זוהי בדיוק היעילות היחסית שהתקבלה עבור הערכים הספציפיים שבהם בחרנו בדוגמה 2.14. יעילות זו אינה תלויה אפילו בשונות של המשתנים והיא קבועה לכל בעיה של השוואת שני מדגמים במודל הנורמלי.

דוגמה 2.19 מודל אחיד: נניח ש-F היא ההתפלגות האחידה על $[0,1]$.
השונות של משתנה בעל התפלגות F היא $\sigma^2=Var(X)=1/12$. עלינו למצוא בנוסף את ההתפלגות F^*, שהיא התפלגות ההפרש בין שני משתנים אחידים כאלה, בלתי תלויים. לא קשה למצוא את ההתפלגות המבוקשת, על סמך ההתפלגות המשותפת של שני המשתנים.
יהיו $X_1\sim U(0,1)$ ו-$X_2\sim U(0,1)$ משתנים בלתי תלויים. ההתפלגות המשותפת שלהם היא אחידה על ריבוע היחידה. בציור 2.10 מתואר התחום שבו ההפרש קטן מערך t, עבור $t>0$.
על פי הציור קל לחשב את התפלגות ההפרש.



ציור 2.10 *ההתפלגות המשותפת של* (X_1,X_2)

אם $0<t\le 1$,

$$F^*(t)=P(X_1-X_2\le t)=1-P(X_1-X_2>t)=1-\frac{(1-t)^2}{2}$$

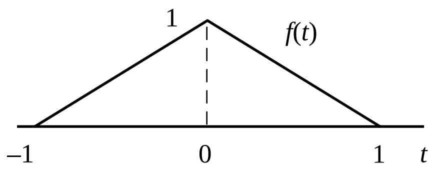
והצפיפות היא הנגזרת $\quad f^*(t)=1-t \qquad 0\le t\le 1$

אם $t\le 0$, ניתן לרשום את הצפיפות על סמך הסימטרייה (טענה 2.5)

$$f^*(t)=1+t \qquad -1\le t\le 0$$

או לחשב במפורש על סמך ציור מתאים.

פונקציית הצפיפות נראית כמו בציור 2.11. הצפיפות בנקודה 0 היא $f^*(0)=1$.



ציור 2.11. *פונקציית הצפיפות של הפרש משתנים אחידים בלתי תלויים*

נציב את הערך של $f^*(0)$ ואת הערך של השונות σ^2 בנוסחה (57) ונקבל את היעילות היחסית

$$ARE(W_s,t)=12\sigma^2\left[f^*(0)\right]^2=12\cdot\frac{1}{12}\cdot 1=1 \qquad (58)$$

היעילות האסימפטוטית היא 1, כלומר מבחן ווילקוקסון ומבחן t יעילים בערך באותה מידה.

דוגמה 2.20. מודל מעריכי: נניח ש-X הוא משתנה מעריכי עם תוחלת 1, כלומר $X\sim\exp(1)$.

ההתפלגות המצטברת היא $F(t)=1-e^{-t},\ t>0$ והצפיפות $f(t)=e^{-t},\ t>0$.

השונות של משתנה מעריכי כזה היא $\sigma^2=1$.

חישוב התפלגות ההפרש F^* יותר מסובך במקרה זה. נסתכל ראשית על $t>0$.

$$F^*(t)=P(X_1-X_2\le t)=\int_0^\infty\int_0^{t+x_2} e^{-x_1}e^{-x_2}\,dx_1dx_2$$

$$=\int_0^\infty e^{-x_2}F(t+x_2)dx_2=\int_0^\infty e^{-x_2}\left[1-e^{-(t+x_2)}\right]dx_2$$

$$=\int_0^\infty e^{-x_2}\,dx_2-\int_0^\infty e^{-t}e^{-2x_2}\,dx_2=1-e^{-t}\cdot\frac{1}{2}$$

הצפיפות היא

$$f^*(t)=\frac{dF^*(t)}{dt}=\frac{1}{2}e^{-t}\qquad t>0$$

ומטעמי סימטרייה (טענה 2.5),

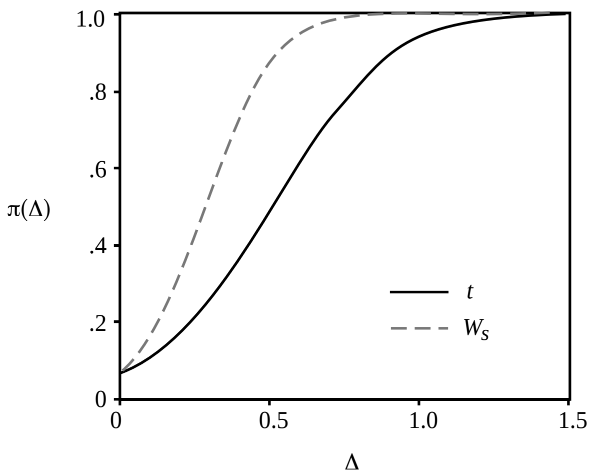
$$f^*(t)=\frac{1}{2}e^{t}\qquad t<0$$

הצפיפות בנקודה 0 היא $f^*(0)=1/2$. היעילות היחסית, לפי נוסחה (57) היא

$$ARE(W_s,t)=12\sigma^2\left[f^*(0)\right]^2=12\cdot1\cdot\left[\frac{1}{2}\right]^2=3\qquad(59)$$

כלומר, מבחן ווילקוקסון הרבה יותר יעיל ממבחן t כשההתפלגות היא מעריכית. שימו לב שההתפלגות המעריכית היא מאוד אסימטרית, עם זנב ימני, ורחוקה ביותר מהתפלגות נורמלית.

ציור 2.12 מדגים את פונקציות העוצמה של שני המבחנים ברמת מובהקות $\alpha=.05$, במודל מעריכי עם פרמטר $\theta=1$ ומדגמים בגודל $m=n=20$. היעילות היחסית הגבוהה נובעת מהפער בין העוצמות עבור ערכי Δ הקרובים לאפס. יחד עם זאת ברור שעוצמת מבחן ווילקוקסון גבוהה בהרבה מזו של מבחן t בכל מקום, פרט לקצוות.



ציור 2.12. *ונקציית העוצמה של מבחן ווילקוקסון ושל מבחן t במודל מעריכי*

דוגמה 2.21. מודל מעריכי סימטרי. נסתכל על מודל שבו הצפיפות היא מעריכית בצורתה, אך מתחלקת באופן סימטרי סביב 0. זו נקראת צפיפות מעריכית כפולה

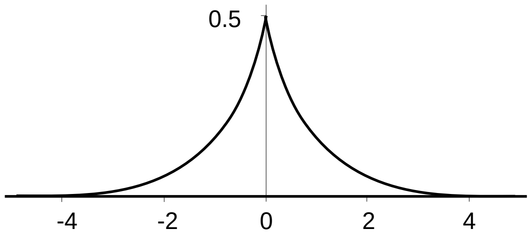
:(double exponential)

$$f(t)=\frac{1}{2}e^{-|t|} \qquad -\infty<t<\infty$$

עקום הצפיפות הזאת מובא בציור 2.13.

(ניתן לראות מהדוגמה הקודמת, דוגמה 2.20, שההתפלגות המעריכית הכפולה היא למעשה התפלגות הפרש בין שני משתנים מעריכיים בלתי תלויים.)

השונות של משתנה מעריכי כזה היא $\sigma^2=2$ וכמו כן $f^*(0)=\frac{1}{4}$ (לא נוכיח זאת כאן).



ציור 2.13. *צפיפות מעריכית כפולה*

היעילות היחסית, לפי נוסחה (57) מתקבלת

$$ARE(W_S,t)=12\sigma^2\left[f^*(0)\right]^2=12\cdot 2\cdot\left[\frac{1}{4}\right]^2=\frac{3}{2}$$

לסיכום, אי אפשר לומר שמבחן t עדיף על מבחן ווילקוקסון. גם במודל הנורמלי, שעבורו מבחן t הוא בעל העוצמה הגבוהה ביותר מבין קבוצה רחבה של מבחנים, אין עוצמתו גבוהה בהרבה מזו של מבחן ווילקוקסון. לעומת זאת, ישנם מודלים שבהם מבחן ווילקוקסון יעיל בהרבה ממבחן t. המסקנה מכך היא שאין סיבה לחשוש משימוש במבחן המבוסס על הדרגות בלבד. התועלת בשימוש במבחן דרגות היא בכך שרמת המובהקות של המבחן אינה תלויה במודל ההסתברותי, גם עבור מדגמים קטנים, ואנו יכולים לסמוך על כך שההסתברות לטעות מסוג I נשמרת, גם ללא כל ידיעה על התפלגות האוכלוסייה.

2.8 רווח בר־סמך לפרמטר הזזה בבעיית שני מדגמים

כדי להעריך את הפער בין שתי האוכלוסיות שמהן הוצאו המדגמים, יש לרשום מודל המתאר את הבעיה במונחים של פרמטר המודד את ה"הבדל" בין ההתפלגויות המתאימות.

אנחנו הסתכלנו על מודל ההזזה (33), ובו פרמטר ההזזה Δ הוא מדד להבדל בין המיקום של שני המשתנים.

נראה כאן כיצד ניתן לאמוד את הערך של Δ על סמך שני המדגמים, וכן כיצד ניתן לבנות רווח בר־סמך עבורו.

אמידת פרמטר ההזזה

נסתכל על משתנה ההפרש $D=Y-X$. ההסתברות ש-D אינו עולה על הערך Δ (פרמטר ההזזה) היא

$$P(D\le\Delta)=P(Y-X\le\Delta)=P(Y-\Delta\le X)$$

לפי טענה 2.3 (1) , התפלגות המשתנה $Y'=Y-\Delta$ זהה להתפלגות המשתנה X. כמו כן X ו- Y' הם בלתי תלויים. מכאן נובע $P(Y'\le X)=1/2$ ולכן מקבלים

$$P(D\le\Delta)=P(Y-\Delta\le X)=P(Y'\le X)=1/2 \qquad (60)$$

כלומר, הפרמטר Δ הוא החציון של התפלגות משתנה ההפרש $D=Y-X$.

ניתן, אם כך, לאמוד את הערך של Δ על סמך נתוני המדגם על ידי חציון כל ההפרשים שהתקבלו בניסוי.

נסמן את ההפרשים הללו, לפי הסדר, על ידי $D_{(1)}<D_{(2)}<\cdots<D_{(M)}$, כאשר $D_{(1)}$ הוא ההפרש הקטן ביותר, $D_{(M)}$ ההפרש הגדול ביותר, ו-M הוא מספר כל ההפרשים בין הזוגות Y_j-X_i, כלומר, $M=mn$.

(רשימת הערכים המסודרים הללו נקראת **סטטיסטי הסדר** של ההפרשים.)

לפי זה חציון כל ההפרשים במדגם ניתן להירשם על ידי:

$$\hat{\Delta}=\begin{cases} D_{((M+1)/2)} & M \text{ אי זוגי} \\ \frac{1}{2}\left[D_{(M/2)}+D_{(M/2+1)}\right] & M \text{ זוגי} \end{cases} \qquad (61)$$

למשל, אם נתונים שני מדגמים בגודל 8 כל אחד, אזי בידינו $8\times8=64$ הפרשים בין כל הזוגות, וחציון ההפרשים הללו יהיה הממוצע בין ההפרש ה-32 להפרש ה-33 בגודלו. אומד זה נקרא אומד הודג'ס-להמן.

דוגמה 2.22. להלן נתונים חלקיים ממחקר (פרופ' אבי שדה וחברים) שנעשה כדי לברר השפעה של משך השינה על התפקוד הקוגניטיבי של תלמידי בית ספר יסודי. הניסוי נערך במשך 5 ימים, כאשר לאחר הלילה השני חלק מן הילדים התבקשו להאריך את שעות השינה בשעה אחת, חלקם התבקשו לקצר אותה בשעה אחת, והאחרים שימשו

כביקורת. משך השינה של כל הילדים נרשם במשך שלושת הלילות הבאים. הנתונים כאן הם לגבי 12 ילדים בגיל 10. שישה הצליחו להאריך את ממוצע שעות השינה שלהם בחצי שעה לפחות, בעקבות בקשה ספציפית של עורכי המחקר ושישה הצליחו לקצר את ממוצע שעות השינה בחצי שעה לפחות. הציונים בלוח 2.7 הם נתונים לגבי הפער בזמן התגובה למטלת קשב (אלפיות השניה) בין המדידה בתחילת הניסוי לבין המדידה בסוף תקופת הניסוי. (זמן תגובה גבוה מעיד על תוצאה גרועה, ולכן פער גבוה בין שני מועדי המדידה מעיד על שיפור הביצוע.)

לוח 2.7 *ציוני הפער בזמן התגובה למטלת קשב*

קיצרו (X):	-77	-26	-14	-8	15	16
האריכו (Y):	-30	13	42	57	58	98

כל $M=36$ ההפרשים בין התלמידים שהאריכו את שעות השינה לבין אלה שקיצרו רשומים בלוח 2.8 בהמשך. אם מסדרים את ההפרשים הללו לפי גודלם, ניתן לראות כי $D_{(18)}=50$ ו- $D_{(19)}=56$ והם רשומים בלוח באות עם קו תחתי. החציון של כל 36 הערכים בלוח הוא הממוצע של שני ערכים אלה:

$$\hat{\Delta}=\frac{1}{2}\left[D_{(18)}+D_{(19)}\right]=\frac{1}{2}[50+56]=53$$

האומדן שקיבלנו עבור הפער בין מידת השיפור של אלה שהאריכו את שעות השינה לבין אלה שקיצרו אותן, הוא 53 אלפיות השניה.

רווח בר־סמך לפרמטר ההזזה

נזכיר (או נחדש, למי שעדיין לא מכיר) את הרעיון של רווח בר־סמך עבור פרמטר. רווח בר־סמך ברמת סמך $1-\alpha$ הוא תחום אקראי המבוסס על תוצאות הניסוי, הבנוי כך שהסיכוי לכך שהרווח אמנם יכלול את הערך הנכון של הפרמטר הוא $1-\alpha$ (והסיכוי שלא יכלול אותו הוא, לפיכך, α). ההגדרה הפורמלית ניתנת להלן.

הגדרה 2.6 רווח בר־סמך ברמת סמך $1-\alpha$ עבור פרמטר θ ניתן על ידי התחום **האקראי** $[T_1,T_2]$, אם קיים: $P_\theta(T_1\le\theta\le T_2)=1-\alpha$ לכל ערך של θ.
רמת הסמך $1-\alpha$ היא ההסתברות לכך שהרווח האקראי $[T_1,T_2]$ יכלול את הערך האמיתי של הפרמטר.

נראה עתה כיצד ניתן לבנות רווח בר־סמך עבור פרמטר ההזזה Δ על סמך $M=mn$

ההפרשים (המסודרים) $D_{(1)} < D_{(2)} < \cdots < D_{(M)}$. המשפט הבא קושר את התפלגות סטטיסטי הסדר של ההפרשים להתפלגות הסטטיסטי של מאן-וויטני.

משפט 2.8. ההסתברות שסטטיסטי סדר מסוים של ההפרשים קטן מהפרמטר Δ ניתנת על ידי:

$$P\{D_{(r)} < \Delta\} = P_{H_0}\{W_{xy} \geq r\} \tag{62}$$

הוכחה: עבור הפרש בודד $D_{ij} = Y_j - X_i$, קיימת זהות בין המאורעות הבאים, כמו בפיתוח נוסחה (60),

$$\{D_{ij} < \Delta\} = \{Y'_j < X_i\} \tag{63}$$

כאשר $Y'_j = Y_j - \Delta$.

נגדיר משתנה מציין

$$T_{ij} = \begin{cases} 1 & Y'_j < X_i \\ 0 & Y'_j > X_i \end{cases}$$

סכום המציינים הללו הוא המשתנה של מאן-וויטני, הסופר את הזוגות (i, j) שעבורם $Y'_j < X_i$. כלומר,

$$\sum_{i=1}^{m}\sum_{j=1}^{n} T_{ij} = \#\left\{(i, j) : Y'_j < X_i\right\} = W_{y'x}$$

אבל לפי נוסחה (63) מספר הזוגות הללו הוא בדיוק מספר ההפרשים D הקטנים מ-Δ. מקבלים אפוא:

$$\#\left\{1 \leq k \leq M : D_k < \Delta\right\} = \#\left\{(i, j) : Y'_j < X_i\right\} = W_{y'x} \tag{64}$$

זאת אומרת, מספר ההפרשים D הקטנים מ-Δ הוא למעשה הסטטיסטי של מאן-וויטני, שהתפלגותו (תחת השערת שוויון התפלגויות) נתונה בטבלה 2 בנספח, במונחים של הסטטיסטי של ווילקוקסון. שימו לב שהמשתנים X ו-Y' בעלי אותה התפלגות, לכל ערך של Δ.

המאורע $\{D_{(r)} < \Delta\}$ הוא למעשה המאורע: {לפחות r מן ההפרשים $D_1, \ldots, D_M$ קטנים מ-Δ}.

היות שמספר ההפרשים הללו זהה למשתנה של מאן-וויטני, לפי נוסחה (64), ההסתברות הדרושה מתקבלת על ידי

$$P\{D_{(r)} < \Delta\} = P\{W_{y'x} \geq r\} = P_{H_0}\{W_{xy} \geq r\}$$

המעבר האחרון נובע מהעובדה שתחת H_0 התפלגות Y שווה להתפלגות X (ושווה להתפלגות Y').

קיבלנו כאן את נוסחה (62). ♣

רווח בר־סמך עבור פרמטר ההזזה Δ מתקבל על סמך משפט 2.8 באופן הבא.

משפט 2.9. עבור $r<s$ קיים:

$$P\{D_{(r)}<\Delta<D_{(s)}\}=P_{H_0}\{r\le W_{xy}<s\} \qquad (65)$$

הוכחה: עבור $r<s$ נשתמש בפירוק

$$\begin{aligned}P\{D_{(r)}<\Delta\}&=P\{(D_{(r)}<\Delta)\cap(D_{(s)}<\Delta)\}\\&\quad+P\{(D_{(r)}<\Delta)\cap(D_{(s)}\ge\Delta)\}\\&=P\{D_{(s)}<\Delta\}+P\{D_{(r)}<\Delta\le(D_{(s)}\}\end{aligned}$$

השוויון האחרון נובע מן העובדה שאם $r<s$ אזי $D_{(r)}<D_{(s)}$.

מכאן ניתן לרשום את ההסתברות המבוקשת

$$\begin{aligned}P\{D_{(r)}<\Delta<D_{(s)}\}&=P\{D_{(r)}<\Delta\}-P\{D_{(s)}<\Delta\}\\&=P_{H_0}\{W_{xy}\ge r\}-P_{H_0}\{W_{xy}\ge s\}\\&=P_{H_0}\{r\le W_{xy}<s\}\end{aligned}$$

השוויון במעבר בין השורה הראשונה לשורה השנייה נובע מנוסחה (62).
שימו לב שההפרשים $D_{(k)}$ הם משתנים רציפים, ולכן אפשר להחליף בין אי־שוויון חלש לאי־שוויון חזק עבורם, אולם W_{xy} הוא משתנה בדיד, ולכן יש חשיבות גדולה לסוג האי־שוויון שרושמים עבורו – אם הוא חלש או חזק. ♣

בהמשך נוותר על רישום השערת האפס לגבי ההסתברות, כשאנו מבינים שאת ההסתברויות המבוקשות אנו מוצאים בטבלה של מאן-וויטני (ווילקוקסון), טבלה 2 בנספח.

מסקנה 2.4. ניתן לבנות רווח בר־סמך עבור Δ באופן הבא: נבחר $r<s$ כלשהם המקיימים:

$$1-\alpha\ge P\{r\le W_{xy}<s\}=\sum_{k=r}^{s-1}P(W_{xy}=k)$$

מכאן, לפי הנוסחה (65) $\quad P\{D_{(r)}\le\Delta\le D_{(s)}\}\ge 1-\alpha$

ורווח בר-סמך ברמת סמך $1-\alpha$ עבור פרמטר ההזזה Δ יינתן על ידי: $[D_{(r)},D_{(s)}]$.

במקרים רבים ניתן לבחור את זוג הדרגות r ו-s בצורות שונות. נוכל לראות זאת בדוגמה 2.23 בהמשך. בגלל צורת ההתפלגות של W_{xy}, שהיא סימטרית עם שכיח במרכז בדומה להתפלגות נורמלית, בדרך כלל הרווח הקצר ביותר יתקבל אם נבחר את r ו-s באופן סימטרי. נוציא את הזנב השמאלי של ההתפלגות, שהסתברותו לא עולה על $\alpha/2$ ואת הזנב הימני הסימטרי. בין שני ערכים אלה ההסתברות היא לפחות $1-\alpha$.

נקבל את הרווח באופן הבא:

מסקנה 2.5. נבחר את הערך c המקסימלי המקיים

$$P\{W_{xy} \le c\} \le \frac{\alpha}{2} \qquad (66)$$

ובהתאם לכך יהיה $r = c+1$.

הערך של s נקבע על סמך הסימטרייה: $s = M - c$ והוא מקיים $P\{W_{xy} \ge s\} \le \frac{\alpha}{2}$.

לפי נוסחה (65) מקבלים מכאן את הרווח הדרוש ברמת סמך $1-\alpha$:

$$\begin{aligned} P\{D_{(r)} \le \Delta \le D_{(s)}\} &= P\{r \le W_{xy} < s\} \\ &= 1 - P\{W_{xy} < r\} - P\{W_{xy} \ge s\} \ge 1-\alpha \end{aligned}$$

דוגמה 2.23. בהינתן גודלי המדגמים m ו-n, וללא כל נתונים לגבי תוצאות הניסוי, ניתן למצוא מראש את צורת הרווח עבור Δ במונחים של הדרגות r ו-s.

למשל, אם $m = n = 6$, כמו בדוגמה 22.2, מספר כל ההפרשים $Y_j - X_i$ הוא $M = 36$.

עבור רמת סמך של $1-\alpha = .90$, מוצאים בטבלת התפלגות ווילקוקסון עם $m = n = 6$, את הערך המקסימלי שעבורו ההסתברות קטנה מ-0.05: $P(W_s \le 28) = .0465$. הערך המתאים עבור מאן-וויטני הוא 7 (בדקו!), כלומר $P(W_{xy} \le 7) = .0465$. לכן $c = 7$, כאשר c מוגדר בנוסחה (66). לפי זה הדרגות המתאימות עבור הרווח הן $r = c+1 = 8$ ו- $s = M - c = 36 - 7 = 29$. הרווח המתקבל עבור Δ הוא $D_{(8)} \le \Delta \le D_{(29)}$.

שימו לב שהרווח הוא סימטרי – מצד שמאל "מסלקים" את 7 ההפרשים הקטנים ביותר, וכך הקצה התחתון של הרווח הוא ההפרש ה-8 בגודלו, וגם בצד ימין "מסלקים" את 7 ההפרשים הגדולים ביותר, וכך הקצה העליון של הרווח הוא ההפרש ה-29 בגודלו, מבין כלל 36 ההפרשים.

נסתכל עתה על הנתונים של דוגמה 2.22. כדי לרשום את כל ההפרשים, הכנו את הנתונים המסודרים של שני המדגמים בטבלה (לוח 2.8), ובכל תא רשמנו את ההפרש המתאים $y_j - x_i$.

בכל שורה בלוח 2.8 הערכים עולים משמאל לימין ובכל עמודה הם יורדים מלמעלה למטה, ולכן ההפרש הנמוך ביותר רשום בעמודה השמאלית בשורה התחתונה (הפינה ה"דרומית-מערבית") וההפרש הגדול ביותר – בעמודה הימנית בשורה העליונה (הפינה ה"צפונית-מזרחית"). מהלוח מוצאים בנקל את ההפרש ה-8 ואת ההפרש ה-29 בגודלו. שתי הקבוצות הכוללות את 7 הערכים הקיצוניים בשני הצדדים מוגבלות במסגרת. קצות הרווח הם ההפרשים הבאים לפי הסדר, והם מסומנים באות עבה.

לוח 2.8 *ההפרשים $y_j - x_i$ בין כל הזוגות של ציוני הילדים שהאריכו לעומת אלה שקיצרו את שעות השינה*

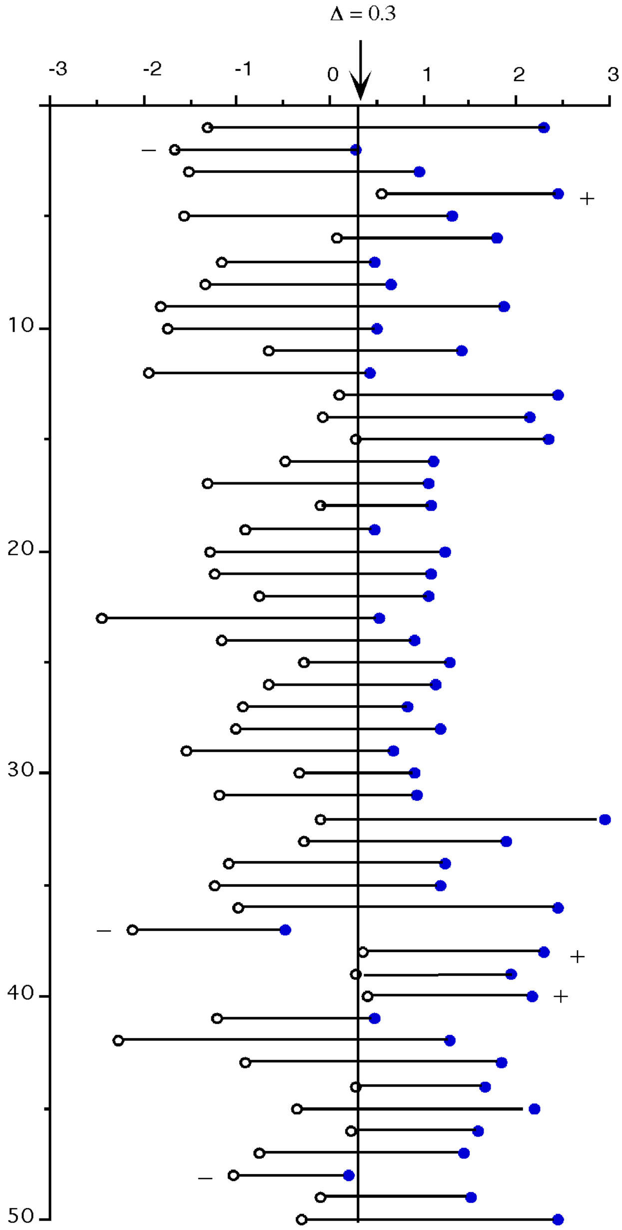
	y_j (האריכו)					
x_j (קיצרו)	-30	13	42	57	58	98
-77	47	**90**	119	134	135	175
-26	-4	39	68	83	84	124
-14	-16	27	<u>56</u>	71	72	112
-8	-22	**21**	<u>50</u>	65	66	106
15	-45	-2	27	42	43	83
16	-46	-3	26	41	42	82

הרווח המתקבל עבור Δ, ברמת סמך של 0.90, הוא $[D_{(8)}, D_{(29)}]=[21,\ 90]$. משמעות הרווח היא שערכי Δ הכלולים בו הם אלה שעבורם התוצאות כפי שהתקבלו הן "סבירות" (ברמת סמך $1-\alpha=.90$). אם הפער בין ההתפלגויות גדול מ-90, כמו 100 למשל, התוצאות שקיבלנו הן בלתי סבירות. ההערכה שלנו היא, אפוא, שהפער בין מידת השיפור במהירות התגובה אצל ילדים שהאריכו את שעות השינה לבין אלה שקיצרו את שעות השינה הוא בערך בין 21 לבין 90 אלפיות השניה.

כדי להסביר את המשמעות הסטטיסטית של רווח בר־סמך אנו מביאים להלן דוגמה של סימולציה שערכנו.

דוגמה 2.24. דגמנו אקראית שני מדגמים בגודל $m=n=6$ מהתפלגויות נורמליות עם שונות $\sigma^2=1$ והפרש תוחלות $\Delta=0.3$. על סמך שני מדגמים אלה בנינו רווח בר־סמך 90% עבור Δ על ידי $D_{(8)} \le \Delta \le D_{(29)}$. חזרנו על דגימה זו 50 פעם. תוצאות הסימולציה מוצגות בציור 2.14.

ליד רווחי סמך שלא כללו את הערך הנכון של Δ רשמנו + או –, בהתאם לכיוון של הטעות. בסך הכל התקבלו בסימולציה 3 רווחים גבוהים מדי ו-3 רווחים נמוכים מדי. שאר 44 הרווחים כללו את הערך הנכון של Δ. רמת הסמך שבחרנו היא 90% ולכן אפשר היה לצַפות שבערך 90% מן הרווחים אמנם יכסו את הערך הנכון של Δ. בסימולציה זו 88% מן הרווחים כללו את הערך $\Delta=.3$.



ציור 2.14 *סימולציה של 50 רווחי סמך 90% עבור פרמטר ההזזה Δ*

משמעות רווח בר־סמך ברמת סמך $1-\alpha$ היא, אפוא, שהסיכוי לכך שהרווח המקרי אמנם יכלול את הערך הנכון של הפרמטר היא $1-\alpha$.

שימוש בקירוב נורמלי

אם m ו-n די גדולים, באופן שניתן להשתמש בקירוב נורמלי לחישוב ההתפלגות של W_{xy}, אפשר למצוא את ערך החלוקה ה- $\alpha/2$ של ההתפלגות על פי הקירוב הנורמלי:

$$P\left(W_{xy} \le c\right) \cong \Phi\left(\frac{c+1/2-mn/2}{\sqrt{mn(N+1)/12}}\right) \le \frac{\alpha}{2}$$

ומכאן ערך המשתנה המתוקנן שווה לערך החלוקה המתאים:

$$\frac{c+1/2-mn/2}{\sqrt{mn(N+1)/12}} \le z_{\alpha/2}$$

הערך c הוא המספר הטבעי המקסימלי המקיים את האי־שוויון

$$c \le \frac{mn-1}{2} + z_{\alpha/2}\sqrt{\frac{mn(N+1)}{12}} \qquad (67)$$

דוגמה 2.24 נמצא, לדוגמה, את הערך c על פי הקירוב הנורמלי, עבור גודלי מדגם $m=n=10$, ורמת־סמך $1-\alpha=.90$. ערך החלוקה המתאים של ההתפלגות הנורמלית סטנדרטית הוא $z_{\alpha/2}=z_{.05}=-1.645$. לפי נוסחה (67) נקבל:

$$c \le \frac{99}{2} - 1.645\sqrt{\frac{100(21)}{12}} = 27.7$$

יש לבחור, אפוא, $c=27$. לפי זה $r=28$ ו- $s=100-27=73$.

נבדוק עתה אם הקירוב טוב. באופן מדויק, בטבלה 2 בנספח מוצאים

$$P(W_{xy} \le 27) = P(W_s \le 82) = .0446$$

והסתברות הערך הבא אחריו היא $P(W_{xy} \le 28) = P(W_s \le 83) = .0526$.

בעזרת הקירוב אמנם קיבלנו את הערך המדויק.

תרגילים

1. N נבדקים מחולקים אקראית לשתי קבוצות: n מהם בקבוצת הטיפול ו-m בקבוצת הביקורת.
   א) תנו את הערך המינימלי ואת הערך המקסימלי שהסטטיסטי של ווילקוקסון W_s יכול לקבל.
   ב) מהי ההסתברות לכך ש-W_s יקבל את הערך המינימלי שחישבתם בחלק א? ומהי ההסתברות לגבי הערך המקסימלי?

2. חשבו את התפלגות W_s כאשר $n=1$, עבור m כלשהו. תארו אותה באופן גרפי עבור $m=10$.

3. עבור המקרה $m=n=4$, חשבו את ההסתברויות $P(W_s \geq 12)$ ו- $P(W_s \leq 14)$ על ידי ספירת כל האפשרויות (רִשמו אותן בצורה מסודרת). השוו את התוצאות שקיבלתם לטבלה 2 בנספח.

4. כדי לבדוק אם הכנה מוקדמת יכולה לעזור להעלאת ציון מנת המשכל, כפי שנמדד על ידי מבחן אינטליגנציה, נבחרו אקראית חמישה מתוך 11 נבדקים וניתן להם מבחן דומה שבו ראו את השאלות והתשובות. שאר השישה שִמשו כקבוצת ביקורת ולא קיבלו כל הכנה מוקדמת. תוצאות מנת המשכל שנמדדו לאחר מכן היו:

עם הכנה:	108	111	114	120	126	
ללא הכנה:	98	101	102	105	110	112

מהי מובהקות התוצאה ומהי המסקנה עבור $\alpha=.01$? ועבור $\alpha=.10$?

5*. א) הוכיחו את נוסחת הרקורסיה הבאה:

$$\#\left\{W_S(n,m)=w\right\}=\#\left\{W_S(n-1,m)=w-N\right\}$$
$$+\#\left\{W_S(n,m-1)=w\right\}$$

כאשר $W_S(k,r)$ הוא הערך של סכום הדרגות של k תצפיות קבוצת הטיפול, כאשר בקבוצת הביקורת r תצפיות.

[רמז: הפרידו לשני מקרים – (1) אם הדרגה המקסימלית N נמצאת בקבוצת הטיפול , או (2) אם היא בקבוצת הביקורת.]

ב) השתמשו בנוסחת הרקורסיה מחלק א, כדי להוכיח את נוסחת הרקורסיה לגבי ההסתברויות:

$$P\left\{W_S(n,m)=w\right\}=\frac{n}{N}P\left\{W_S(n-1,m)=w-N\right\}$$
$$+\frac{m}{N}P\left\{W_S(n,m-1)=w\right\}$$

כאשר ההסתברויות מחושבות תחת השערת האפס.

ג) רִשמו את צורת החישוב של ההסתברות $P\{W_S(4,6)=17\}$ בעזרת נוסחת הרקורסיה מחלק ב. השתמשו בערכים מהטבלה שבידכם וודאו שהתוצאה שקיבלתם אמנם זהה לערך שבטבלה.

ד) כמו בחלק ג, עבור $P\{W_S(4,6)=12\}$.

6*. הוכיחו באופן ישיר, על סמך פונקציית ההסתברות המשותפת של הדרגות (S_i,S_j), את הנוסחה $Cov(S_i,S_j)=-\frac{N+1}{12}$ עבור $i \neq j$

7. לבדיקת ההשפעה האפשרית של נטילת ויטמין C על הצטננויות נבחרו מבין עובדי

מפעל 14 מתנדבים. שמונה מהם נבחרו מקרית וקיבלו טיפול שוטף בויטמין C, ושאר המתנדבים קיבלו גלולות סוכר (פלציבו). במשך שלוש שנות המחקר היו מספר ימי ההיעדרויות בגלל הצטננות:

טיפול:	16	9	8	7	4	3	1	0
ביקורת:	28	23	19	15	13	2		

כדי לבדוק אם ויטמין C מוריד את כמות ההצטננויות,

א) השתמשו בטבלה כדי לקבוע את מובהקות התוצאה לפי הסטטיסטי של ווילקוקסון. מה המסקנה עבור $\alpha=.05$?

ב) חשבו את הערך של הסטטיסטי של מאן-וויטני עבור הנתונים הללו. ודאו שאמנם הקשר בין שני הסטטיסטים מתקיים פה.

ג) השתמשו בקירוב הנורמלי כדי לחשב את המובהקות בקירוב. השוו לתוצאה של חלק א.

.8 טיפול חדש בחולים לאחר ניתוח משווה n חולים בקבוצת הניסוי ל-m חולים בקבוצת הביקורת (טיפול סטנדרטי).

א) מִצאו עבור איזה ערכים של W_s יש לדחות את ההשערה של חוסר הבדל בין שני הטיפולים, כנגד אלטרנטיבה חד־צדדית (הטיפול החדש יעיל יותר מן הסטנדרטי), עבור רמת מובהקות קרובה ככל האפשר ל-0.05, אם

(1) $m=n=8$; (2) $m=8,\ n=9$; (3) $m=9,\ n=8$.

ב) הנתונים להלן הם הזמן (בימים) עד להחלמה, כאשר $m=n=9$.

טיפול סטנדרטי:	21	24	30	32	36	37	40	48	54
טיפול חדש:	19	20	22	25	26	28	29	34	38

מהי מובהקות התוצאה? מה המסקנה עבור רמת מובהקות $\alpha=.05$?

ג) בִּדקו לגבי הנתונים לעיל אם יש הבדל בין שני הטיפולים (מבחן **דו־צדדי**), ברמת מובהקות $\alpha=.05$. תנו את מובהקות התוצאה.

.9 מִצאו את המובהקות **המדויקת** (על ידי מְנִייה) של תוצאת שני המדגמים להלן (מבחן חד־צדדי):

קבוצת הטיפול: $2\ ,\ 2\ ,\ 1\ ,\ -3$

קבוצת הביקורת: $1\ ,\ 0\ ,\ -3\ ,\ -3\ ,\ -1$

.10 מִצאו את התפלגות $\tilde{W}_s$ ותארו אותה על ידי היסטוגרמה, עבור המקרה $m=n=3$ במקרים הבאים:

א) $t_1=2,\ t_2=2,\ t_3=1,\ t_4=1$;

ב) $t_1=1,\ t_2=2,\ t_3=3$;

ג) $t_1=3,\ t_2=3$.

.11 במהלך מחקר שנערך לשם ניסוי של תכניות התערבות בגני ילדים (ד"ר דורית

ארם) נאספו נתונים שונים הקשורים בקריאה וכתיבה. הנתונים כאן הם הציונים בכתיבת שם של 14 ילדי גן טרום טרום חובה בתחילת הניסוי, מתוכם 6 בנות.

1 5 3 1 3 2 3 3 3 5 3 3 3 3

א) דרגו את הנתונים הללו וחשבו ישירות על סמך הדרגות שקיבלתם את שונות "אוכלוסיית" הדרגות הללו.

ב) חשבו לפי התוצאה בחלק א את שונות סכום הדרגות הממוצעות של מדגם בגודל 6 של בנות מאוכלוסיה זו.

ג) מִצאו את שונות סכום הדרגות לפי נוסחת השונות של $\tilde{W}_S$, נוסחה (20), למקרה הנתון: $m=8,\ n=6$.

12. פסיכולוג ביקש לברר אם תרופה מסוימת יכולה לעזור בשליטה על התנהגות תוקפנית אצל חולים. 14 חולים קיבלו את התרופה ו-14 אחרים קיבלו פלציבו. כל הנבדקים עברו מבחן לגבי שליטה על התנהגות תוקפנית וקיבלו את הציונים הבאים (ציון גבוה מעיד על התנהגות תוקפנית רבה).

נתחו את הנתונים כדי לקבוע אם יש ממש בהשערתו של הפסיכולוג ($\alpha=.05$).

פלציבו				תרופה			
9	12	13	11	15	14	10	8
25	20	10	14	16	18	13	10
	10	16	18		12	12	16
	21	14	16		17	13	7

13. עשרה חולים מטופלים מושווים לעשרה חולים המשמשים כביקורת, כאשר החולים ממוינים לפי מידת ההתקדמות במצב בריאותם, בהתפלגות השכיחויות הבאה:

	טוב מאוד	טוב	בינוני	חלש	חלש מאוד
ביקורת	0	1	5	2	2
טיפול	1	3	4	2	0

א) מִצאו את הערך הקריטי c עבור הסטטיסטי של ווילקוקסון, הנותן רמת מובהקות קרובה ל-0.01. השתמשו בקירוב הנורמלי המתאים.

ב) האם תוצאת הניסוי מובהקת עבור רמת מובהקות $\alpha=.01$? מה המסקנה?

ג) מִצאו את המובהקות המדויקת של התוצאה, אם, למעשה, התפלגות החולים שקיבלו טיפול הייתה שונה:

9 הרגישו טוב מאוד ו-1 הרגיש טוב. (נתוני קבוצת הביקורת אינם משתנים.)

*14. במבחן שנערך על השפעת רקע תרבותי על מבחני אינטליגנציה הוצע ניסוי של חונכות לילדים בעלי רקע תרבותי שונה. החונכות הייתה אמורה להיערך על ידי

סטודנטים על בסיס שבועי, במשך שנה. הייתה חשיבות רבה להחליט מראש על גודל קבוצת הטיפול. מובן שהוחלט גם לעקוב אחר קבוצת ביקורת, שגודלה שווה לקבוצת הטיפול.

מה צריך להיות גודלן של שתי הקבוצות, כדי להבטיח סיכוי של 95% לפחות לגילוי שיפור של 10 נקודות IQ, אם משתמשים במבחן ווילקוקסון ברמת מובהקות של 0.01? (ממחקרים קודמים ניתן להניח לצורך זה שבאוכלוסייה כזו התפלגות ציוני ה-IQ היא בערך נורמלית, עם סטיית תקן $\sigma=15$.)

15*. בתנאי שאלה 13, מה צריך להיות גודל קבוצות הילדים, כדי להבטיח באותן נסיבות אותה עוצמה, אם משתמשים במבחן t במקום במבחן ווילקוקסון?

16. חשבו את היעילות היחסית של מבחן ווילקוקסון ביחס למבחן t בתנאים לעיל.

17*. בתנאי שאלה 8, נניח שזמני ההחלמה הם נורמלים עם שונות $\sigma^2=5$, ונשתמש במדגמים בגודל $m=n=25$. מִצאו את העוצמה המקורבת של מבחן ווילקוקסון ברמת מובהקות $\alpha=.02$, אם הטיפול החדש מוריד את זמן ההחלמה

(א) ביום אחד; (ב) ביומיים.

18*. כמו שאלה 16, עבור מודל הזזה, כאשר אתם מניחים שהתפלגות זמן ההחלמה תחת השערת האפס היא מעריכית עם שונות $\sigma^2=5$.

19*. הוכיחו כי עבור מדגם בגודל קבוע $m+n=N$, העוצמה המקורבת הנתונה על ידי

$$\pi(\Delta)=\Phi\left(\sqrt{\frac{12mn}{N+1}}f^*(0)\Delta-z_{1-\alpha}\right)$$

היא מקסימלית, אם בוחרים:

א) $m=n=k$ כאשר $N=2k$ (זוגי);

ב) $n=k,\ m=k+1$ או $n=k+1,\ m=k$ כאשר $N=2k+1$ (אי זוגי).

כלומר, העוצמה גבוהה ככל שגודלי המדגמים קרובים זה לזה.

20. נסמן ב-$\hat{\Delta}(\mathbf{x},\mathbf{y})=med(y_j-x_i)$ את חציון כל $m\cdot n$ ההפרשים, $i=1,\dots,m$, $j=1,\dots,n$.

הוכיחו כי לכל $a,\ b$ קבועים קיים: $\hat{\Delta}(a+b\mathbf{x},a+b\mathbf{y})=b\hat{\Delta}(\mathbf{x},\mathbf{y})$.

21. נגדיר $\overline{\Delta}=\overline{Y}-\overline{X}$ (ההפרש בין ממוצעי המדגמים).

הוכיחו כי במקרה ש-$m=n=2$, שני האומדנים $\overline{\Delta}$ ו-$\hat{\Delta}$ הם תמיד זהים.

22. שני תהליכי ייצור שונים של אותו מוצר הושוו מבחינת אחוז המוצרים הפגומים המיוצרים בתהליך, על ידי בדיקת מנות של מוצרים. האחוזים שהתקבלו:

שיטה א:	7.0	9.4	7.2	8.0	7.1	6.2	8.5	6.9
שיטה ב:	8.8	6.8	5.7	6.5	6.1	7.3	5.0	

א) נסחו את מודל ההזזה המתאים לבעיה זו.

ב) אִמדו את פרמטר ההזזה Δ. מה פירושו?

ג) תנו רווח בר־סמך 95% עבור Δ.

ד) עִרכו מבחן דו־צדדי ברמת מובהקות $\alpha=.05$ לבדיקת ההבדל בין אחוז הפגומים בשתי השיטות. השוו לתוצאות שקיבלתם בחלק ג והסבירו.

23. השתמשו בקירוב הנורמלי כדי למצוא רווח בר־סמך מקורב עבור Δ.

רִשמו את הרווח בכל מקרה במונחים של סטטיסטי הסדר של ההפרשים $D_{(1)}<\ldots<D_{(mn)}$, כאשר אתם מניחים שכל ההפרשים הם שונים:

א) $m=n=25,\ 1-\alpha=.90$;

ב) $m=n=25,\ 1-\alpha=.95$;

ג) $m=20,\ n=30,\ 1-\alpha=.95$.

פרק 3

מבחנים נוספים להשוואת שתי התפלגויות

בפרק זה נביא מבחנים שונים להשוואת התפלגויות, כאשר הבעיה היא לאו דווקא בדיקה לגבי המיקום של ההתפלגויות הללו.

3.1 השוואת פיזורים

לעתים המחקר מתמקד בבעיה של ניסוי חדש, שלגביו לא ברור אם שיטת המדידה היא אמינה ומדויקת דיה. נשים לב שאם המדידות של הניסוי והביקורת מתרכזות סביב אותו ערך מרכזי פחות או יותר, ועם זאת תוצאות הניסוי נוטות, למשל, להיות יותר מפוזרות (עם שונות גדולה), הסטטיסטי של ווילקוקסון לא ייתן לנו אינדיקציה לכך. נביא כאן שני מבחנים שונים לבעיה של השוואת הפיזור של שתי התפלגויות. לשם שימוש בשני המבחנים הללו נצטרך להניח שהחציונים של שתי ההתפלגויות שווים. שימו לב שבמקרה שהחציונים אינם שווים, בעיית הפיזור בדרך כלל אינה משמעותית, מכיוון שממילא ההתפלגויות שונות זו מזו. לכן בדרך כלל השוואת ההתפלגויות מבחינת הפיזור משמעותית רק כאשר החציונים שווים.

המודל של הבעיה:

נתונים שני מדגמים בלתי תלויים $X_1,X_2,\ldots,X_m$ ו-$Y_1,Y_2,\ldots,Y_n$, בעלי התפלגויות F ו-G, בהתאמה. אנו מסמנים $m+n=N$.

הנחות: F ו-G רציפות, בעלות אותו חציון.

ההשערות הנבדקות: $H_0: Var(X)=Var(Y)$ כנגד $H_1: Var(X)<Var(Y)$.

ההשערה האלטרנטיבית היא השערה חד־צדדית, הטוענת שהתפלגות תוצאות הניסוי מפוזרת יותר.

מבחן זיגל-טוקי

מסדרים את כל N התצפיות לפי הגודל ונותנים לתצפית i "משקל" b_i באופן הבא:

א) מסתכלים על זוג התצפיות הקיצוניות ביותר (המינימלית והמקסימלית). לתצפית שדרגתה 1 נותנים משקל $b_1 = 1$ ולתצפית שדרגתה N נותנים משקל $b_N = 2$.

ב) מסתכלים על זוג התצפיות הקיצוניות ביותר, פרט לזוג שעבורו כבר נקבעו המשקלים. לתצפית שדרגתה $N-1$ נותנים את המשקל הבא בתור, כלומר, $b_{N-1} = 3$ ולתצפית שדרגתה 2 נותנים משקל $b_2 = 4$.

וכך הלאה. המשקלים הם הערכים $1,2,\ldots,N$, והם ניתנים באופן שבכל שלב שני המשקלים העוקבים הבאים בתור ניתנים לשתי התצפיות הקיצוניות שעדיין לא קיבלו משקל, ובכל שלב מחליפים את הסדר בין משקל התצפית הגדולה לבין משקל התצפית הקטנה בין בני הזוג.

לדוגמה, אם נתונות $N = 7$ תצפיות, המשקלים ניתנים באופן הבא:

הדרגה: i	7	6	5	4	3	2	1
המשקל: b_i	2	3	6	7	5	4	1

הזוגות של תצפיות מקבילות מסומנים בציור.

בשיטת המשקלים הללו, המדגם מן האוכלוסייה בעלת השונות הגדולה נוטה לקבל את המשקלים הקטנים יותר (אלה השייכים לתצפיות היותר קיצוניות).

סטטיסטי המבחן של זיגל-טוקי (Siegel & Tukey, 1960) הוא סכום המשקלים הללו בקבוצת הניסוי, וניתן לרשום אותו על ידי

$$S = \sum_{i=1}^{N} b_i Z_i \qquad (1)$$

כאשר Z_i הוא המשתנה המציין תצפית בקבוצת הניסוי (המשתנה Y).

תחת האלטרנטיבה (שהתפלגות ה-Y-ים יותר מפוזרת) S נוטה לקבל ערכים קטנים, ולכן דוחים את השערת האפס עבור ערכים נמוכים של S.

התפלגות סטטיסטי המבחן תחת השערת האפס זהה להתפלגות הסטטיסטי של ווילקוקסון לשני מדגמים בלתי תלויים. נשים לב שהמשקלים $b_1, b_2, \ldots, b_N$ הם למעשה המספרים $1,2,\ldots,N$ המסודרים לפי השיטה שנקבעה. כמו כן, תחת השערת האפס לשני המדגמים בדיוק אותה התפלגות, ולכן כל התמורות של המשקלים הללו הן שוות הסתברות. מכאן

ניתן לרשום: $P_{H_0}(S=k)=P_{H_0}(W_s=k)$.
התפלגות הסטטיסטי של ווילקוקסון נתונה בטבלה 2 בנספח, וכן ניתן להשתמש בקירוב נורמלי מתאים עבור מדגמים גדולים יותר, עם התוחלת והשונות (8) ו-(9) בפרק 2.

דוגמה 3.1. הושוו 12 מדידות בשיטה סטנדרטית (X) ל-10 מדידות בשיטה חדשה, מהירה יותר (Y). רוצים לבדוק אם ייתכן שהשיטה המהירה פחות מדויקת. הנתונים מובאים בלוח 3.1, שבו כבר מחושבים המשקלים b_i. (המשקלים הנוספים המובאים בטבלה משמשים למבחן אחר, שיובא יותר מאוחר.)

לוח 3.1. *מדידות בשתי שיטות. התצפיות מסודרות לפי גודלן.*

משקל a_i	משקל b_i	דרגה	תצפית	קבוצה
1	**1**	1	1.87	y
2	4	2	3.19	x
3	**5**	3	3.35	y
4	8	4	3.39	x
5	**9**	5	4.00	y
6	**12**	6	4.10	y
7	13	7	4.13	x
8	**16**	8	4.37	y
9	17	9	4.41	x
10	20	10	4.46	x
11	21	11	4.53	x
11	22	12	4.55	x
10	19	13	4.81	x
9	18	14	4.90	x
8	**15**	15	5.16	y
7	**14**	16	5.21	y
6	**11**	17	5.68	y
5	10	18	5.92	x
4	7	19	6.55	x
3	6	20	6.58	x
2	**3**	21	7.34	y
1	**2**	22	8.15	y

ההשערות הנבדקות הן $H_0:Var(X)=Var(Y)$ כנגד $H_1:Var(X)<Var(Y)$.
הסטטיסטי של זיגל-טוקי ניתן על ידי סכום המשקלים של 10 תצפיות ה-y-ים. משקלים אלה רשומים באות עבה בלוח 3.1. התוצאה המתקבלת היא

$$S=1+5+9+12+16+15+14+11+3+2=88$$

מובהקות התוצאה היא ההסתברות (תחת H_0) לקבלת תוצאה קטנה או שווה לזו שהתקבלה. היות שהמדגמים גדולים מאלה שעבורם ניתן להשתמש בטבלה 2 ($m=12,\ n=10$), נשתמש בקירוב הנורמלי. המומנטים תחת השערת האפס, לפי משפט 2.4, הם

$$ES=EW_s=\frac{n(N+1)}{2}=\frac{10(23)}{2}=115$$

$$Var(S)=Var(W_s)=\frac{mn(N+1)}{12}=\frac{12(10)(23)}{12}=230$$

ומכאן מובהקות התוצאה היא

$$P=P_{H_0}(S\le 88)\cong\Phi\left(\frac{88.5-115}{\sqrt{230}}\right)=\Phi(-1.747)=.040$$

ברמת מובהקות של 5% התוצאה מובהקת. כלומר, אכן ניתן להסיק שבשיטה החדשה השונות גדולה מאשר בשיטה הסטנדרטית. השיטה המהירה היא פחות מדויקת.

ערכי תיקו

אם ההתפלגויות אינן רציפות וקיימים ערכי תיקו בין התצפיות, ניתן להשתמש בקירוב הנורמלי, כפי שנעשה עבור המבחן של ווילקוקסון [תיקון לתיקו עבור השונות – ראו פרק 2, נוסחה (31)]. עם זאת, יש לשים לב שקבוצות התיקו חייבות לקבל כולן אותו משקל (ממוצע המשקל בקבוצה), כשהמשקלים הם מספרים טבעיים עוקבים. זאת אומרת, אם שתי התצפיות הנמוכות ביותר שוות, והמשקלים המגיעים להן אילו היו שונות הם 1 ו-4, אין לתת לכל אחת מהן משקל ממוצע של ערכים אלה – $(1+4)/2=2.5$. במקום זאת יש לתת להן את המשקל $(1+2)/2=1.5$, המתקבל כאילו המשקלים המגיעים להן הם 1,2. הצורה הכללית לתת את המשקלים השווים בקבוצת תיקו בגודל t חייבת להיות ממוצע של t ערכים עוקבים. לכן הופכים את המשקלים בין הקצה הימני לקצה השמאלי לפי ה"גושים" של תצפיות התיקו. נסתכל על דוגמה פיקטיבית.

דוגמה 3.2. נניח שהתצפיות שהתקבלו נתונות בלוח הבא (ללא קשר אם הן שייכות לניסוי או לביקורת). הקפנו כל "גוש" של תצפיות שוות. יש כאן בסך הכול חמישה "גושים", שאותם מסדרים לפי הסדר שנקבע בשיטת זיגל-טוקי, כפי שרואים בלוח, כאילו כל גוש כולל תצפית אחת בלבד. ה"גוש" הנמוך ביותר מקבל את המשקלים הנמוכים. לפיכך שתי התצפיות הראשונות (5) קיבלו את המשקל הממוצע של הדרגות 1 ו-2. התצפית המקסימלית (17) קיבלה את המשקל הבא – 3, אחריה, שתי התצפיות

הגדולות השוות ביניהן (12) קיבלו את המשקל הממוצע של 4 ו-5 וכך הלאה.

התצפית	17	12	12	9	9	6	5	5
סדר הגושים	(2)	(3)		(5)		(4)	(1)	
סדר המשקלים בגוש	3	4	5	8	7	6	2	1
משקל b_i	3	4.5	4.5	7.5	7.5	6	1.5	1.5

בצורה כזאת אנו דואגים לכך שהמשקלים שניתנו לתצפיות הם רשימה של ערכים שחלקם שווים ביניהם, ושווים לממוצע של מספר **ערכים עוקבים**. זאת בדומה למה שאנו עושים בעת דירוג כל התצפיות בנוכחות ערכי תיקו עבור מבחן ווילקוקסון. אם לא יינתנו משקלים בשיטה הזאת, התפלגות הסטטיסטי של זיגל-טוקי לא תהיה זהה להתפלגות הסטטיסטי של ווילקוקסון כאשר ישנם ערכי תיקו.

הערה: למבחן זיגל-טוקי יש חיסרון, והוא שהמבחן איננו סימטרי. ברור שאת המשקלים ניתן להתחיל בערך הגדול ביותר במקום בערך הקטן ביותר ולהפוך את הסדר בהתאם לכך. סכום המשקלים בקבוצת הניסוי יכול, כמובן, להשתנות. היתרון הגדול של המבחן הוא בכך שאין צורך בטבלאות חדשות כדי לחשב את מובהקות התוצאה.

מבחן אנסרי-ברדלי

מבחן אנסרי-ברדלי (Ansari & Bradley, 1960) מתקן את הבעיה של חוסר סימטרייה בשיטה של זיגל וטוקי. המשקלים הניתנים לתצפיות המסודרות (a_i) הם סימטריים, כאשר המשקלים הנמוכים ניתנים לתצפיות הקיצוניות. ראו, למשל, את הציור הבא, עם 7 תצפיות.

הדרגה: i	7	6	5	4	3	2	1
המשקל: a_i	1	2	3	4	3	2	1

באופן כללי, ניתן לרשום את המשקלים של התצפיות באופן הבא:

א) אם N זוגי, יתקבלו המשקלים לפי

הדרגה: i	N	$N-1$	$N-2$	$\cdots$	$N/2+1$	$N/2$	$\cdots$	3	2	1
המשקל: a_i	1	2	3	$\cdots$	$N/2$	$N/2$	$\cdots$	3	2	1

ב) אם N אי־זוגי, יתקבלו המשקלים לפי

הדרגה: i	N	$N-1$	$N-2$	$\cdots$	$(N+1)/2$	$\cdots$	3	2	1
המשקל: a_i	1	2	3	$\cdots$	$(N+1)/2$	$\cdots$	3	2	1

סטטיסטי המבחן של אנסרי-ברדלי הוא סכום המשקלים הללו בקבוצת הניסוי:

$$A=\sum_{i=1}^{n} a_i Z_i \qquad (2)$$

כאשר Z_i הוא המשתנה המציין של תצפית בקבוצת הניסוי (המשתנה Y).
את ההתפלגות המדויקת של A ניתן לחשב על ידי

$$P_{H_0}(A=t)=\frac{\#\{A=t\}}{\binom{N}{n}} \qquad (3)$$

טבלאות מתאימות נבנו על ידי אנסרי וברדלי (Ansari and Bradley, 1960).
עבור מדגמים גדולים ניתן להשתמש בקירוב נורמלי. לשם כך יש לחשב את התוחלת והשונות של A. ברור שההתפלגויות, וכן התוחלת והשונות, שונות בהתאמה אם N זוגי או אי־זוגי.
נשים לב שבכל מקרה, A הוא סכום של תת־קבוצה בגודל n מבין N המשקלים הרשומים ב-א (עבור N זוגי) או ב-ב (אם N אי־זוגי). נחשב, אפוא, את המומנטים על סמך המומנטים של סכום תצפיות בדגימה אקראית ללא החזרה מתוך אוכלוסייה סופית.
א) נניח N זוגי. כל אחד מהערכים $1,2,\ldots,N/2$ מתקבל בהסתברות שווה – $2/N$.
כלומר, זו אוכלוסייה בעלת התפלגות אחידה $U(1,N/2)$.

תוחלת האוכלוסייה היא $$\mu=\frac{1+N/2}{2}=\frac{N+2}{4}$$

והשונות $$\sigma^2=\frac{(N/2)^2-1}{12}=\frac{N^2-4}{48}$$

לפי נוסחאות התוחלת והשונות של סכום תצפיות במדגם ללא החזרה מאוכלוסייה סופית, נוסחאות (4) ו-(5) בפרק 2, מקבלים עבור הסטטיסטי A:

$$EA=n\mu=\frac{n(N+2)}{4} \qquad (4)$$

$$Var(A)=n\sigma^2\frac{N-n}{N-1}=\frac{nm(N^2-4)}{48(N-1)} \qquad (5)$$

ב) נניח N אי־זוגי. כל אחד מהערכים ברשימה ב מתקבל בהסתברות שווה. התוחלת והשונות של האוכלוסייה הן במקרה זה (פרטי החישובים מובאים בנספח 3):

$$\mu=\frac{(N+1)^2}{4N}$$

$$\sigma^2=\frac{(N+1)(N-1)(N^2+3)}{48N^2}$$

ומכאן התוחלת והשונות של A הן

$$EA=n\mu=\frac{n(N+1)^2}{4N} \tag{6}$$

$$Var(A)=n\sigma^2\cdot\frac{N-n}{N-1}=\frac{nm(N+1)(N^2+3)}{48N^2} \tag{7}$$

דוגמה 3.3 (המשך דוגמה 3.1). המשקלים a_i רשומים בלוח 3.1. סכום המשקלים הללו עבור תצפיות מדגם הניסוי (מסומנים באות עבה) מתקבל:

$$A=1+3+5+6+8+8+7+6+2+1=47$$

בדוגמה זו $N=22$ הוא זוגי, ולכן נערוך את התִקנון לפי הנוסחאות (4) ו-(5), עם $m=12,\ n=10$.

התוחלת היא
$$EA=\frac{n(N+2)}{4}=\frac{10(24)}{4}=60$$

והשונות
$$Var(A)=\frac{nm(N^2-4)}{48(N-1)}=\frac{10(12)(22^2-4)}{48(21)}=57.143$$

המובהקות המקורבת של התוצאה היא, אפוא,

$$P=P_{H_0}(A\le 47)\cong\Phi\left(\frac{47.5-60}{\sqrt{57.143}}\right)=\Phi(-1.65)=.0495$$

התוצאה מובהקת באופן גבולי. המובהקות קצת יותר חלשה מזו שהתקבלה על פי מבחן זיגל-טוקי בדוגמה 3.1, אולם ההבדל אינו גדול.

ערכי תיקו

אין בידינו נוסחה קצרה לחישוב השונות של הסטטיסטי של אנסרי-ברדלי ענור ערכי תיקו. את השונות יש לחשב באופן ישיר על סמך הדרגות הממוצעות שהתקבלו במדגם והיא תלויה ברשימת הדרגות הממוצעות שהתקבלו, ולא רק בגודל קבוצות התיקו.

השונות של A מתקבלת לפי נוסחת השונות של מדגם ללא החזרות:

$$Var(A)=n\sigma^2\left[\frac{N-n}{N-1}\right]=\frac{nm}{N-1}\sigma^2$$

כאשר σ^2 היא שונות "אוכלוסיית" הדרגות שהתקבלו במדגם.

הערה: שני המבחנים שהצענו לעיל, מבחן זיגל-טוקי ומבחן אנסרי-ברדלי, מבוססים על ההנחה שהחציונים של שתי ההתפלגויות שווים והבעיה היא רק הבדלי השונויות. אם הנחה זו אינה נכונה, ניתן לעבור למשתנים חדשים, למשל, הסטייה של כל תצפית

מחציון המדגם שלה. התפלגות הסטטיסטים במקרה זה אינה מדויקת.

3.2 מבחן קולמוגורוב-סמירנוב*

המודל של הבעיה: נתונים שני מדגמים בלתי תלויים $X_1,X_2,\dots,X_m$ ו- $Y_1,Y_2,\dots,Y_n$, בעלי התפלגויות F ו-G, בהתאמה. אנו מסמנים $m+n=N$.

ההשערות הנבדקות:

א. השערה חד־צדדית: $H_0: F(t)=G(t)$ לכל t

עבור ערך אחד של t לפחות $F(t)>G(t)$, לכל t $H_1: F(t)\geq G(t)$ (8)

השערה זו טוענת שהמשתנה Y גדול סטוכסטית מהמשתנה X.

ב. השערה דו־צדדית: $H_0 : F(t) = G(t)$ לכל t

עבור ערך אחד של t לפחות $H_1 : F(t) \neq G(t)$ (9)

הנחה: F ו-G רציפות.

סטטיסטי המבחן מבוסס על אמידת ההתפלגויות F ו-G, על סמך שני המדגמים. האמידה נעשית באמצעות ההתפלגויות האמפיריות.

הגדרה 3.1. יהי נתון מדגם $Z_1,Z_2,\dots,Z_n$ של n תצפיות בלתי תלויות מהתפלגות F. ההתפלגות האמפירית של המדגם מוגדרת על ידי

$$F_n(t)=\frac{1}{n}\sum_{i=1}^{n}U_i(t)=\frac{1}{n}\#\{Z_i\leq t\} \qquad -\infty<t<\infty \tag{10}$$

כאשר
$$U_i(t)=\begin{cases}1 & Z_i\leq t\\ 0 & Z_i>t\end{cases}$$

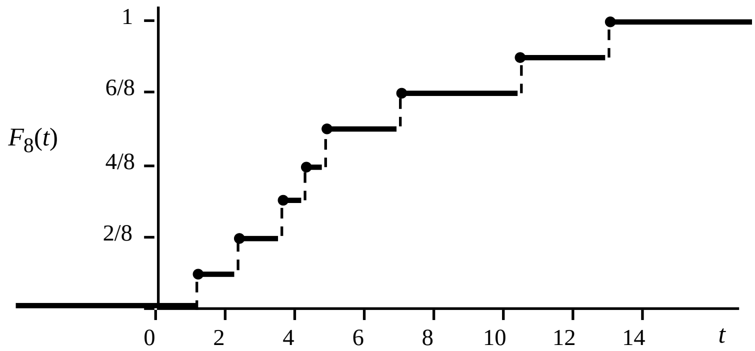
כלומר, עבור כל נקודה t, $F_n(t)$ שווה לפרופורציית התצפיות בְמדגם בגודל n, שערכן אינו עולה על t.

ההתפלגות האמפירית של המדגם היא למעשה ההתפלגות המצטברת המתאימה לפונקציית ההסתברות הנותנת משקל שווה (של $1/n$) לכל אחת מ-n התצפיות שהתקבלו במדגם. זוהי, למעשה, פונקציית "מדרגות", כאשר הקפיצות הן בנקודות של ערכי המדגם וגובה כל אחת מהקפיצות הוא $1/n$. דוגמה לכך נראה בדוגמה 3.4.

דוגמה 3.4. נסתכל על מדגם פיקטיבי של 8 תצפיות:

1.2, 3.4, 3.6, 4.3, 4.9, 7.0, 10.5, 12.5 (התצפיות כבר מסודרות לפי גודלן).

ההתפלגות האמפירית של המדגם ניתנת בציור 3.1.



ציור 3.1. *פונקציית ההתפלגות האמפירית של המדגם בדוגמה 3.4*

משפט 3.1. התוחלת של ההתפלגות האמפירית היא $EF_n(t) = F(t)$ עבור כל t.

הוכחה: את ההתפלגות המצטברת בנקודה t ניתן לרשום על ידי

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} U_i(t) = \frac{S_n(t)}{n}$$

$S_n(t)$ הוא מספר התצפיות שערכן אינו עולה על t. כיוון שהתצפיות בלתי תלויות וכולן בעלות אותה התפלגות F, ברור שהתפלגות המשתנה הזה היא בינומית:

$S_n(t) \sim B[n, F(t)]$. לכן $ES_n(t) = nF(t)$ ומכאן נובעת הטענה. ♣

מן המשפט עולה כי עבור כל נקודה על הישר, ההתפלגות האמפירית בנקודה זו מהווה אומד בלתי מוטה עבור התפלגות האוכלוסייה באותה נקודה.

תכונה נוספת, החשובה ביותר בבעיות של אמידה, היא מידת הקִרבה של האומד לפרמטר הנאמד [כאן – הערך של ההתפלגות $F(t)$]. מתברר שאם המדגם די גדול, אזי הסטייה בין האומד לפרמטר אינה יכולה להיות גדולה מאוד. השונות של ההתפלגות האמפירית נותנת אינדיקציה על הקִרבה בין האומד לפרמטר:

$$Var(F_n(t)) = \frac{Var(S_n(t))}{n^2} = \frac{nF(t)[1-F(t)]}{n^2} = \frac{F(t)[1-F(t)]}{n}$$

שונות זו שואפת לאפס כאשר n שואף לאינסוף. נוכל להסיק, אפוא, שהסטייה בין ההתפלגות האמפירית לבין ההתפלגות באוכלוסייה קטֵנה ככל שהמדגם גדֵל.
ניתן לרשום קִרבה זו בצורה מתמטית מדויקת כמסקנה פשוטה מהחוק החלש של המספרים הגדולים.

משפט 3.2. עבור כל נקודה t, $\lim_{n\to\infty} P\{|F_n(t) - F(t)| > \varepsilon\} = 0$ לכל $\varepsilon > 0$.

התכונה הרשומה כאן נקראת גם **שאיפה בהסתברות** של $F_n(t)$ ל-$F(t)$ ומשמעה שההסתברות לקבלת סטייה מוחלטת חיובית כלשהי (אפילו קטנה מאוד) בין האומד לפרמטר שואפת לאפס כאשר המדגם גדל לאינסוף. באופן מעשי, עבור מדגמים גדולים האומד יהיה די קרוב לפרמטר ולכן שיטת האמידה הזאת היא טובה.

הסטטיסטי של קולמוגורוב-סמירנוב להשוואת שני מדגמים מבוסס על שתי ההתפלגויות האמפיריות של המדגמים:

$$F_m(t)=\frac{1}{m}\#\{X_i\le t\} \tag{11}$$

$$G_n(t)=\frac{1}{n}\#\{Y_j\le t\} \tag{12}$$

סטטיסטי המבחן הוא מדד של המרחק בין שתי ההתפלגויות האמפיריות. המדד שונה בהתאם להשערת המחקר (Smirnov, 1939).

א) לבדיקת השערה חד־צדדית (8) הסטטיסטי הוא

$$D^{+}_{m,n}=\max_{-\infty<t<\infty}[F_n(t)-G_n(t)] \tag{13}$$

הסטטיסטי (13) הוא ההפרש המקסימלי בין שתי ההתפלגויות האמפיריות.
השערת שוויון ההתפלגויות נדחית (לטובת האלטרנטיבה שהמשתנה Y גדול סטוכסטית מהמשתנה X) עבור ערכים גבוהים של $D^{+}_{m,n}$.

ב) לבדיקת השערה דו־צדדית (9) הסטטיסטי הוא

$$D_{m,n}=\max_{-\infty<t<\infty}|F_n(t)-G_n(t)| \tag{14}$$

הסטטיסטי (14) הוא ההפרש **המוחלט** המקסימלי בין שתי ההתפלגויות האמפיריות.
השערת שוויון ההתפלגויות נדחית (לטובת האלטרנטיבה שההתפלגויות שונות) עבור ערכים גבוהים של $D_{m,n}$.

הפונקציה F_m מקבלת $m+1$ ערכים שונים (כולל 0 ו-1) והפונקציה G_n מקבלת $n+1$ ערכים שונים (כולל 0 ו-1). לכן ההפרש בין שתי ההתפלגויות האמפיריות יכול לקבל לכל היותר $m+n$ ערכים שונים.

הערה: במבחנים קודמים השתמשנו באותו סטטיסטי לבדיקת השערה חד־צדדית או דו־צדדית, כאשר ההבדל בין הפתרונות לשתי הבעיות הוא באזור הדחייה השונה. במבחנים של קולמוגורוב-סמירנוב, לכל בעיה משתמשים בסטטיסטי שונה ולכל אחד מהסטטיסטים הללו התפלגות משלו.

טענה 3.1. הסטטיסטים $D_{m,n}$ ו-$D^+_{m,n}$ מבוססים אך ורק על **הדרגות** של התצפיות במדגם המצורף: $R_1,\ldots,R_m$ – דרגות ה-X-ים ו-$S_1,\ldots,S_n$ – דרגות ה-Y-ים בין כל $N=m+n$ התצפיות.

הוכחה: ערכו של ההפרש המקסימלי תלוי רק בערכי ההתפלגויות האמפיריות בנקודות שהתקבלו במדגמים. נסמן ב-$X_{(1)}<\ldots<X_{(m)}$ את סטטיסטי הסדר של מדגם ה-X-ים וב-$Y_{(1)}<\ldots<Y_{(n)}$ את סטטיסטי הסדר של מדגם ה-Y-ים (אלה ערכי המדגמים שהתקבלו, כאשר הם מסודרים לפי גודלם).

ערכי ההתפלגויות האמפיריות בנקודות אלה מתקבלים באופן פשוט על ידי

$$F_m(X_{(l)})=\frac{1}{m}\#\{X_i\le X_{(l)}\}=\frac{l}{m}\qquad l=1,\ldots,m$$

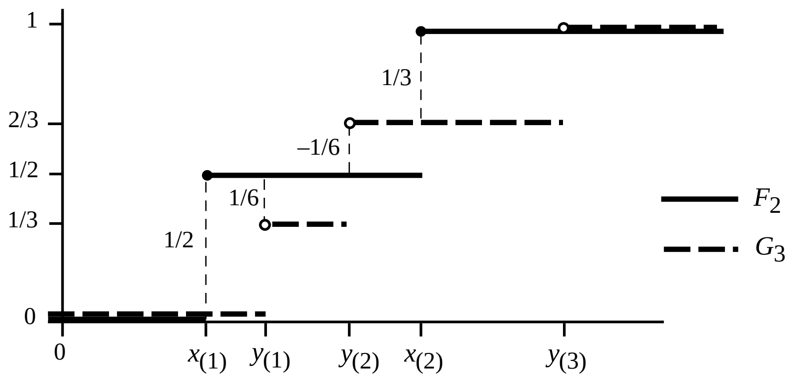
$$F_m(Y_{(r)})=\frac{1}{m}\#\{X_i\le Y_{(r)}\}=\frac{1}{m}(S_r-r)\qquad r=1,\ldots,n$$

$$G_n(Y_{(r)})=\frac{1}{n}\#\{Y_j\le Y_{(r)}\}=\frac{r}{n}\qquad r=1,\ldots,n$$

$$G_n(X_{(l)})=\frac{1}{n}\#\{Y_j\le X_{(l)}\}=\frac{1}{n}(R_l-l)\qquad l=1,\ldots,m$$

מכיוון שערכי ההתפלגויות האמפיריות תלויים רק בדרגות, גם ההפרש ביניהן תלוי רק בדרגות. ♣

דוגמה 3.5. נניח שבניסוי עם $m=2,\ n=3$ תצפיות התקבלה סדרת הדירוגים הבאה: $x_{(1)}<y_{(1)}<y_{(2)}<x_{(2)}<y_{(3)}$. בציור 3.2 מתוארות שתי ההתפלגויות האמפיריות וכל ההפרשים $F_2(t)-G_3(t)$. בסך הכול ישנם 4 ערכים שונים מאפס של ההפרשים. המקסימום שלהם הוא $D^+_{2,3}=1/2$ וכך גם מקסימום הערכים המוחלטים $D_{2,3}=1/2$.



ציור 3.2. *שתי התפלגויות אמפיריות וההפרשים ביניהן*

במקום להשתמש בציור כדי למצוא את ההפרש המקסימלי בין ההתפלגויות האמפיריות, ניתן לרשום את ההתפלגויות בטבלה וכך למצוא את כל ההפרשים. הדרגות של נתוני הדוגמה רשומות בלוח 3.2, ועבור כל דרגה רשום מספר ה-x-ם הקטנים או השווים לה וכן מספר ה-y-ם הקטנים או השווים לה [אלה הם הערכים $2F_2(t)$ ו-$3G_3(t)$ בהתאמה].

לוח 3.2. *חישוב* $D_{2,3}$ ו- $D^+_{2,3}$ בדוגמה 3.5

הדרגות	1	2	3	4	5
התצפיות	x	y	y	x	y
$2F_2$	1	1	1	2	2
$3G_3$	0	1	2	2	3
F_2	1/2	1/2	1/2	1	1
G_3	0	1/3	2/3	2/3	1
F_2-G_3	1/2	1/6	–1/6	1/3	0

התפלגות הסטטיסטים תחת השערת האפס

כמו במקרים הקודמים שבהם הסתכלנו על סטטיסטי המבוסס על הדרגות, גם כאן נסתמך על העובדה שתחת H_0 לכל אחת מהבחירות של n מתוך N דרגות ה-Y-ים יש אותה הסתברות. מכאן מקבלים:

$$P_{H_0}(D_{m,n}=d)=\frac{\#\{D_{m,n}=d\}}{\binom{N}{n}}$$

וכן
$$P_{H_0}(D^+_{m,n}=d)=\frac{\#\{D^+_{m,n}=d\}}{\binom{N}{n}}$$

קיימות טבלאות של ההתפלגויות הללו עבור מדגמים קטנים. כאן לא הבאנו טבלאות אלה.

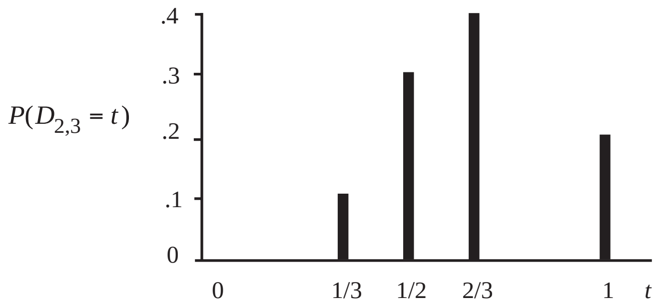
דוגמה 3.6. נציג כאן את התפלגות הסטטיסטים של קולמוגורוב-סמירנוב עבור שני מדגמים קטנים, $m=2,\ n=3$, שנמצאה על ידי מְנִייה. במקרה זה בסך הכול $\binom{5}{2}=10$ אפשרויות לבחירת דרגות שונות עבור ה-y-ים.

ההתפלגויות המתקבלות הן (בדקו! תרגיל 6)

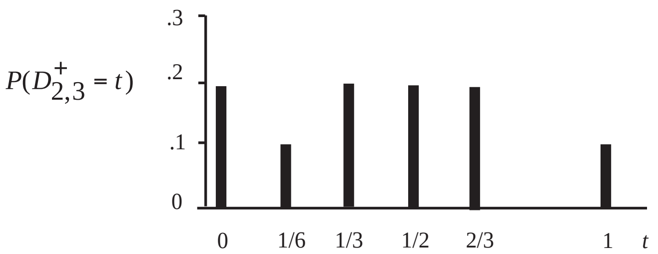
t	1/3	1/2	2/3	1
$P(D_{2,3} = t)$	.1	.3	.4	.2

t	0	1/6	1/3	1/2	2/3	1
$P(D^+_{2,3} = t)$	.2	.1	.2	.2	.2	.1

התיאור הגרפי של ההתפלגויות לעיל מובא בציורים 3.3א ו-3.3ב. הציורים מדגימים את העובדה שהתפלגויות שני הסטטיסטים של קולמוגורוב-סמירנוב שונות מאוד זו מזו. כמו כן, ניתן לראות ששתי ההתפלגויות אינן סימטריות.



ציור 3.3א. *התפלגות* $D_{2,3}$



ציור 3.3ב. *התפלגות* $D^+_{2,3}$

התפלגויות אסימפטוטיות

ההתפלגויות המקורבות של הסטטיסטים של קולמוגורוב-סמירנוב אינן נורמליות. נשים לב שסטטיסטֵי המבחן אינם סכום או קומבינציה ליניארית אחרת של הדרגות, כמו

במקרים קודמים, שבהם ההתפלגות האסימפטוטית הייתה נורמלית.

משפט 3.3.

א) ההתפלגות האסימפטוטית של הסטטיסטי של קולמוגורוב-סמירנוב לבדיקת השערה **חד־צדדית** ניתנת על ידי

$$\lim_{m,n\to\infty} P\left(\sqrt{\frac{mn}{m+n}}\cdot D^{+}_{m,n} \ge t\right) = e^{-2t^2} \qquad t>0 \text{ לכל} \tag{15}$$

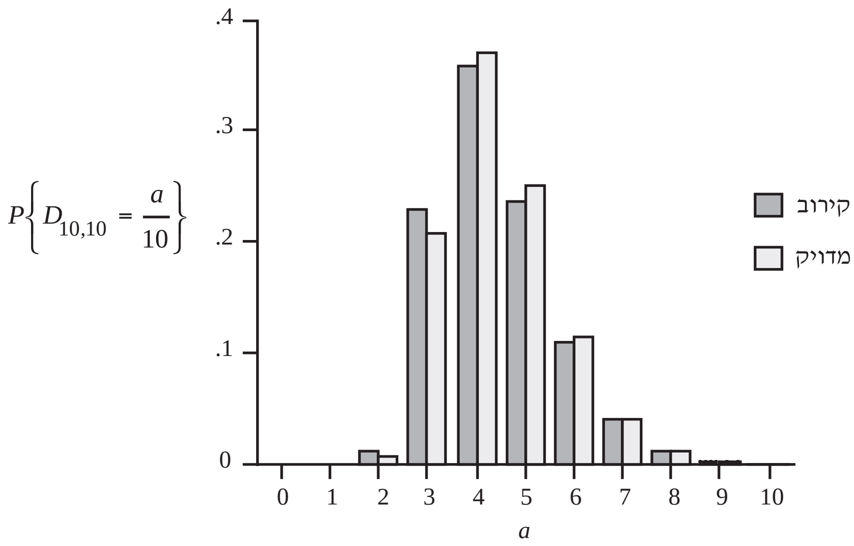
ב) ההתפלגות האסימפטוטית של הסטטיסטי של קולמוגורוב-סמירנוב לבדיקת השערה **דו־צדדית** ניתנת על ידי

$$\lim_{m,n\to\infty} P\left(\sqrt{\frac{mn}{m+n}}\cdot D_{m,n} \ge t\right) = L(t) \qquad t>0 \text{ לכל} \tag{16}$$

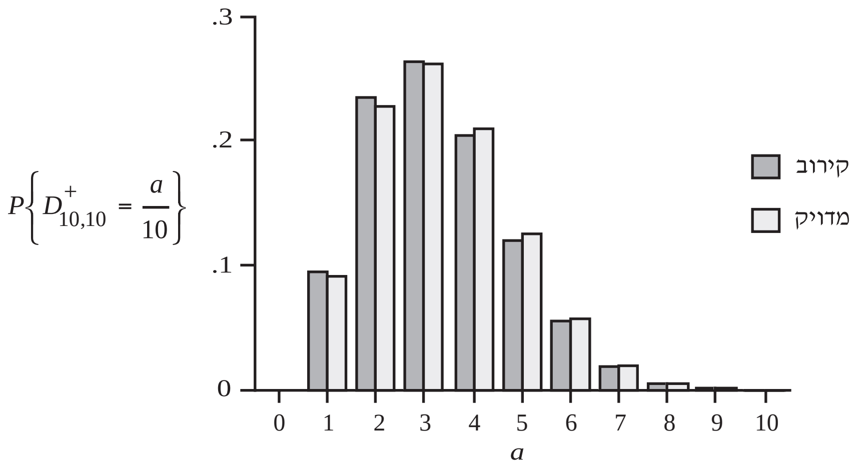
כאשר $L(t)=2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2t^2}$.

טבלה של הפונקציה L מובאת בטבלה 6 בנספח.

לשם הדגמה נביא כאן את ההתפלגות המדויקת ואת ההתפלגות המקורבת של שני הסטטיסטים של קולמוגורוב-סמירנוב עבור $m=n=10$. בציורים 3.4א ו-3.4ב מוצגים תיאורים גרפיים של ההתפלגויות המתאימות של $D_{10,10}$ ושל $D^{+}_{10,10}$. הקירוב די טוב כבר עבור מדגמים בגודל 10, במיוחד במקרה החד־צדדי.



ציור 3.4א. *ההתפלגות המקורבת וההתפלגות המדויקת של הסטטיסטי של קולמוגורב-סמירנוב לבעיה דו־צדדית* ($m = n = 10$)



ציור 3.4ב. *ההתפלגות המקורבת וההתפלגות המדויקת של הסטטיסטי של קולמוגורב-סמירנוב לבעיה חד־צדדית* ($m=n=10$)

נסתכל עתה על שימוש בסטטיסטי של קולמוגורוב-סמירנוב עבור דוגמה של ניסוי אמיתי בשטח.

דוגמה 3.7 (המשך דוגמה 2.3). הנתונים בדוגמה 2.3 הם חלק ממחקר שנערך בישראל (עירית דר) כדי לבדוק את הנושא של אלימות בסביבות שונות, כפי שמדווחים ילדים צעירים. הובאו שם ציוני מדד האלימות כפי שדיווחו 9 בנות ו-10 בנים, הלומדים בכיתה ב. נרצה לבדוק אם בבית הספר בנים עדים למקרי אלימות יותר מאשר בנות. בדוגמה 2.3 השתמשנו לניתוח הנתונים במבחן ווילקוקסון לשני מדגמים וקיבלנו תוצאה מובהקת ($P=.0326$). נערוך עתה מבחן קולמוגורוב-סמירנוב חד־צדדי עבור אותם הנתונים.

כדי להקל על חישוב הפרש ההתפלגויות האמפיריות במקרה זה (כאשר $m \neq n$), נרשום את ההפרש באופן הבא:

$$F_m(t)-G_n(t)=\frac{\#\{X_i \le t\}}{m}-\frac{\#\{Y_j \le t\}}{n}$$

$$=\frac{1}{mn}\left[n\cdot\#\{X_i \le t\}-m\cdot\#\{Y_j \le t\}\right]$$

ובדוגמה זו

$$F_9(t)-G_{10}(t)=\frac{1}{9\cdot 10}\left[10\cdot\#\{X_i \le t\}-9\cdot\#\{Y_j \le t\}\right]$$

בלוח 3.3 נתונות שוב דרגות התצפיות בשני המדגמים.

לוח 3.3. *חישוב הסטטיסטי של קולמוגורוב-סמירנוב עבור נתוני דוגמה 2.3 ו-3.7*

הדרגה	התצפית	$\#\{x_i \le t)$	$\#\{y_j \le t)$	$10 \cdot \#\{x_i \le t\} - 9 \cdot \#\{y_j \le t\}$
1	x	1	0	10
2	x	2	0	20
3	x	3	0	30
4	y	3	1	21
5	x	4	1	31
6	x	5	1	**41**
7	y	5	2	32
8	y	5	3	23
9	x	6	3	33
10	y	6	4	24
11	x	7	4	34
12	y	7	5	25
13	x	8	5	35
14	y	8	6	26
15	y	8	7	17
16	y	8	8	8
17	x	9	8	18
18	y	9	9	9
19	y	9	10	0

מקסימום ההפרש – 41 (מסומן באות עבה) מתקבל עבור הדרגה 6. הבעיה היא חד־צדדית ולכן הערך של סטטיסטי המבחן הוא $D^+_{9,10} = \frac{1}{9 \cdot 10}(41) = .4556$.
כדי להשתמש בהתפלגות המקורבת נעבור למשתנה המתוקנן:

$$\sqrt{\frac{mn}{m+n}} D^+_{m,n} = \sqrt{\frac{9 \cdot 10}{9+10}} D^+_{9,10} = \sqrt{\frac{90}{19}} \cdot 0.4556 = 0.99$$

המובהקות המקורבת של התוצאה היא, לפי נוסחה (15),

$$P = P_{H_0}(D^+_{9,10} \ge .4556) \cong e^{-2(.4556^2)} = .140$$

קיבלנו תוצאה לא מובהקת. לפיכך, אם משתמשים במבחן קולמוגורוב-סמירנוב לבעיה זו לא מצליחים לאשש את הטענה שהבנים עדים לאלימות בבית הספר יותר מבנות.

נביא עתה דוגמה לבעיה דו־צדדית.

דוגמה 3.8. נערוך מבחן קולמוגורוב-סמירנוב לנתוני דוגמה 3.1 (מדידות בשיטה סטנדרטית

לעומת שיטה חדשה). התוצאה המתקבלת (בִדקו את החישוב!) היא

$$D_{12,10}=\frac{1}{12\cdot 10}(30)=\frac{3}{12}=.25$$

כדי להשתמש בהתפלגות המקורבת נעבור למשתנה המתוקנן:

$$\sqrt{\frac{mn}{m+n}}D_{m,n}=\sqrt{\frac{12\cdot 10}{12+10}}D_{12,10}=\sqrt{\frac{120}{22}}\cdot 0.25=0.584$$

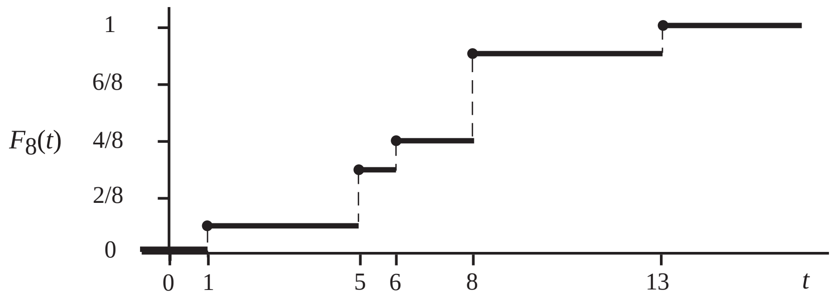
על פי הטבלה של ההתפלגות המקורבת, טבלה 6 בנספח, מקבלים את מובהקות התוצאה: $P(D_{12,10}\geq .25)\cong L(.584)=.8896$. זוהי, כמובן, מובהקות גדולה ביותר והמסקנה היא שלא נמצא הבדל מובהק בין שתי ההתפלגויות.

הערה: התוצאות של מבחן ווילקוקסון להשוואת בנים ובנות בדוגמה 2.3 והמבחנים להשוואת פיזורים בדוגמה 3.1 ובדוגמה 3.3 היו מובהקות, עם מובהקויות הנמוכות קצת מ-0.05. תוצאות אלה שונות מהתוצאות שקיבלנו בדוגמה 3.7 ובדוגמה 3.8, שבהן השתמשנו במבחני קולמוגורוב-סמירנוב. למעשה אין כאן סתירה. לגבי הבעיה החד־צדדית, הנתונים מצביעים על פער בין שתי ההתפלגויות בכיוון של מיקום גבוה יותר של התפלגות הבנים בהשוואה להתפלגות הבנות, שמבחן ווילקוקסון רגיש במיוחד לסוג כזה של פער. המבחן של קולמוגורוב-סמירנוב חייב לגלות הבדלים שונים שעלולים להיות בין שתי ההתפלגויות ולכן במקרה שההבדל הוא דווקא במיקום, סביר שמבחן המיועד לבדיקה ספציפית זו ייתן מובהקות קטנה יותר. באופן דומה, לגבי הבעיה הדו־צדדית, הנתונים מצביעים על פער בין ההתפלגויות בכיוון של פיזורים שונים, שהמבחנים להשוואת פיזורים רגישים במיוחד לסוג כזה של פער. באופן כללי, אם אנו מתעניינים בפער מסוג **ספציפי**, כמו הבדל במיקום, או הבדל בפיזור, כדאי להשתמש במבחנים המתאימים לבעיות הספציפיות, שהם בעלי עוצמה גבוהה יותר לבעיה זו.

ערכי תיקו במבחן קולמוגורוב-סמירנוב

אין כל בעיה לחשב את סטטיסטי המבחן כאשר ישנם ערכי תיקו. מספר התצפיות שהן קטנות או שוות לערך t כלשהו מוגדר היטב. לדוגמה, בציור 3.5 מתוארת פונקציית ההתפלגות האמפירית של המדגם הבא: .1, 5, 5, 6, 8, 8, 8, 14

עבור מדגמים קטנים מאוד ניתן גם לחשב את ההתפלגות המדויקת של סטטיסטי המבחן, תחת השערת האפס (אם כי החישוב מסורבל, כמובן). הבעיה היא שאין בידינו התפלגות מקורבת של סטטיסטי המבחן במקרה של תיקו. תוצאה שעוזרת בפתרון הבעיה מובאת בטענה 3.2.



ציור 3.5. *התפלגות אמפירית של מדגם עם ערכי תיקו*

טענה 3.2. נסמן ב-$\tilde{D}_{m,n}$ את הסטטיסטי של קולמוגורוב-סמירנוב לבעיה דו־צדדית במקרה של ערכי תיקו. אזי קיים

$$P_{H_0}(\tilde{D}_{m,n} \geq a) \leq P_{H_0}(D_{m,n} \geq a) \qquad \text{לכל } a > 0 \qquad (17)$$

לא נוכיח טענה זו כאן. רעיון ההוכחה הוא פשוט – ניתן להראות בקלות כי אם "נשבור" את ערכי התיקו, באופן שכל התצפיות תהיינה שונות, יהיה ההפרש המקסימלי בין ההתפלגויות האמפיריות גדול לפחות כמו זה שהתקבל קודם (תוספת של ערכים אפשריים להפרש יכולה רק להגדיל את המקסימום).
טענה 3.2 תקפה גם לגבי בעיה חד־צדדית.

לפי טענה 3.2, לבדיקת השערות במקרה שקיימים ערכי תיקו נוהגים להשתמש בערכים הקריטיים המתאימים לבעיה שבה אין ערכי תיקו. כך מתקבל מבחן "שמרני", בעל רמת מובהקות שאינה עולה על α. מובהקות התוצאה היא למעשה נמוכה יותר מזו המתקבלת בהנחה שהמדגמים אינם כוללים ערכי תיקו. נראה דוגמה לכך.

דוגמה 3.9. להלן נתונים מתוך מחקר שנערך לשם השוואת שני חומרי אִלחוש שונים ששימשו בעת טיפול שיניים אצל ילדים (ד"ר אריקה עמיר). התוצאות בלוח 3.4 הן מדד לתחושת כאב מיד עם סיום הטיפול, על סקלה בין 0 = אין כאב, עד 5 = כואב מאוד. (נראה מהלוח שאף ילד לא הרגיש כאב חזק.)
פונקציות ההתפלגות האמפיריות של שני המדגמים מוצגות בלוח 3.5. כדי לבדוק אם החומר החדש גורם לכאבים פחותים מאלה של החומר הסטנדרטי, נשתמש במבחן חד־צדדי.
ההפרש המקסימלי מתקבל עבור $t = 0$: $\tilde{D}^{+}_{39,23} = F_{39}(0) - G_{23}(0) = 0.21$.

לוח 3.4. *ילדים שהיו בטיפול שיניים, לפי החומר ומדד הכאב*

הטיפול	0	1	2	סך הכול
	מדד הכאב			
סטנדרטי (y)	4	13	6	23
חדש (x)	15	14	10	39
סך הכול	19	27	16	62

לוח 3.5. *פונקציות ההתפלגות האמפיריות של נתוני דוגמה 3.9*

	0	1	2
F_{39}	15/39	29/39	39/39
G_{23}	4/23	17/23	23/23
$F_{39}-G_{23}$	.211	.004	0

כדי לקבוע אם התוצאה מובהקת נעבור למשתנה המתוקנן:

$$\sqrt{\frac{39\cdot 23}{39+23}}\cdot 0.21=0.80$$

לפי נוסחת הקירוב (15) מקבלים

$$P(D^+\geq 21)=P\left(\sqrt{\frac{39\cdot 23}{39+23}}\cdot D^+\geq 0.80\right)\cong e^{-2(.80^2)}=.278$$

ולכן מובהקות התוצאה שהתקבלה עם ערכי תיקו מקיימת

$$P=P_{H_0}(\tilde{D}^+\geq .21)\leq .278$$

החסם כאן גבוה מאוד והתוצאה אינה מובהקת. ניתן להסיק, אפוא, שאין הבדל בין שני החומרים מבחינת תחושת הכאב.

הערה: איננו יכולים לומר מהי בדיוק מובהקות התוצאה שהתקבלה, אלא רק שהיא אינה עולה על 0.278. המסקנה שלנו מעובדה זו (ההחלטה לא לדחות את השערת האפס) מתבססת על המבחן השמרני, שבו אנו משווים את החסם שהתקבל לרמת המובהקות שנקבעה.

3.3 מבחן החציון

הסטטיסטי של קולמוגורוב-סמירנוב הוא פונקציה של שתי ההתפלגויות האמפיריות, המוגדרות כמספר התצפיות (x או y) הקטנות או שוות ל-t עבור **כל** ערך t על הישר. לעומת זאת, במבחן החציון אנו מסתכלים על נקודה אחת בלבד על הישר – חציון המדגם המצורף.

הבעיה הנחקרת היא בעיית שני המדגמים, כפי שהוגדרה בסעיף 3.2. כלומר, נתונים שני מדגמים בלתי תלויים $X_1, X_2, \ldots, X_m$ ו-$Y_1, Y_2, \ldots, Y_n$, בעלי התפלגויות רציפות F ו-G, בהתאמה. אנו מסמנים $N = m + n$.

ההשערות הנבדקות הן $H_0 : F(t) = G(t)$ לכל t,

כנגד האלטרנטיבה ש-Y נוטה לקבל ערכים גבוהים מ-X: $H_1 : Y \succ X$

נסמן ב-M את החציון של N התצפיות בשני המדגמים ונסמן את הסטטיסטים:

U – מספר ה-X_i-ים הקטנים מ-M, $i = 1, \ldots, m$

V – מספר ה-Y_j-ים הקטנים מ-M, $j = 1, \ldots, n$

וכן נסמן $t = U + V$.

(נזכיר ש-$S = n - V$ היה הסטטיסטי הרביעי שהוצע בדוגמה 1.1.)

נשים לב ש-t הוא מספר קבוע (לא משתנה מקרי) והוא תלוי רק בגודל N באופן הבא:

א) אם N זוגי, אזי החציון של N התצפיות הוא הממוצע בין שתי התצפיות האמצעיות, ולכן מספר התצפיות הקטנות ממנו הוא $t = U + V = \frac{N}{2}$.

ב) אם N אי־זוגי, אזי M הוא בדיוק הערך האמצעי של N התצפיות ולכן מספר התצפיות הקטנות ממנו הוא $t = U + V = \frac{N-1}{2}$.

התפלגות U תחת השערת האפס

תחת H_0 בידינו N תצפיות בלתי תלויות, כולן מאותה התפלגות. בסך הכול ישנן t תצפיות הקטנות מ-M מתוך N התצפיות בשני המדגמים. יש הסתברות שווה לכל החלוקות של N התצפיות לשתי קבוצות (אלו שמתחת לחציון M, שמספרן t, ואלו שמעליו), והיא $\frac{1}{\binom{N}{t}}$.

לפי זה מקבלים

$$P(U = u) = \frac{\#\{U = u\}}{\binom{N}{t}} = \frac{\binom{m}{u}\binom{n}{t-u}}{\binom{N}{t}} \qquad u = 0, 1, \ldots, t \tag{18}$$

במונה רשום מספר האפשרויות לכך שבדיוק u מבין m ה-X-ים יהיו בקבוצת התצפיות הקטנות מ-M (ובדיוק $t-u$ בקבוצה זו הם Y-ים).

הנוסחה (18) היא נוסחת התפלגות היפרגאומטרית. כלומר, $U \sim H(N,m,t)$.

[נזכיר לאלה ששכחו, או לא הכירו התפלגות זו קודם, כי אנו אומרים ש-T מתפלג היפרגאומטרית, ומסמנים $T \sim H(N,D,r)$, אם T הוא מספר האיברים ה"מיוחדים" המתקבלים בדגימה **ללא החזרה** של r איברים מתוך אוכלוסייה בת N איברים, שביניהם D הם "מיוחדים". במקרה הנידון כאן, N הוא המספר הכולל של תצפיות, $D=m$ הוא מספר ה-x-ים ביניהן, ו-$r=t$ הוא מספר התצפיות הקטנות מהחציון. אנו דוגמים t תצפיות מתוך אוכלוסייה בת N איברים, וסופרים כמה מבין אלה שבחרנו הם x-ים.]

הערה: ייתכן שערך מסוים u הוא גדול מדי או קטן מדי ואינו יכול להתקבל כלל כתוצאה מהניסוי, בגלל המגבלה שחייבים להתקיים שני האי־שוויונים: $u \le m$ וגם $t-u \le n$. אולם הנוסחה (18) נשארת נכונה אם מגדירים באופן כללי $\binom{r}{k}=0$ כאשר $r<k$. נראה דוגמה לכך מיד בדוגמה 3.10.

מבחן החציון

ההשערה האלטרנטיבית היא ש-Y גדול סטוכסטית מ-X. על כן נצַפה שתחת האלטרנטיבה מספר ה-X-ים הקטנים מהחציון M יהיה יחסית גדול. לכן דוחים את השערת האפס עבור ערכים גבוהים של U, או, לחלופין, עבור ערכים נמוכים של V. מובהקות התוצאה היא, אפוא, $P=P_{H_0}(U \ge u)$, כאשר u היא התוצאה שהתקבלה בניסוי.

דוגמה 3.10. נסתכל, לדוגמה, על ניסוי שבו $m=8,\ n=10$, כאשר 7 מן ה-x-ים קטנים מהחציון המשותף M ורק 2 מן ה-y-ים קטנים מ-M. ניתן לרשום את התוצאות בטבלת שכיחויות:

	קטן מ-M	גדול או שווה M	סך הכול
x	7	1	8
y	2	8	10
סך הכול	9	9	18

כאן $N=18$ ולכן $t=9$ (זהו סכום איברי העמודה הראשונה משמאל). סטטיסטי המבחן שהתקבל הוא $U=7$ והוא רשום בתא השמאלי העליון. מובהקות התוצאה היא $P=P_{H_0}(U \ge 7)=P(V \le 2)$. כדי למצוא אותה יש לרשום את כל לוחות השכיחות

שבהם ניתן לקבל את הערכים הקיצוניים הללו ולחשב את ההסתברות של כל אחד מהם.
התוצאה $U=8$ מתאימה ללוח השכיחות הבא:

	קטן מ-M	גדול או שווה M	סך הכול
x	8	0	8
y	1	9	10
סך הכול	9	9	18

שולי הטבלאות: $m,\ n,\ t,\ N-t$ הם **קבועים**, ובמקרה שלנו הם 9 ,9 ,10 ,8. התוצאות השונות האפשריות ניתנות על ידי אפשרויות סידור השכיחויות **בתוך** הטבלה. במקרה שלנו, התוצאה $U=9$ איננה אפשרית, בגלל המגבלה של השוליים בשורה הראשונה. (אי אפשר לקבל 9 תצפיות של X הקטנות מ-M, כיוון שבידינו רק 8 תצפיות של X בסך הכול.) מובהקות התוצאה היא, אפוא,

$$P(U\geq 7)=P(U=7)+P(U=8)=\frac{\binom{8}{7}\binom{10}{2}}{\binom{18}{9}}+\frac{\binom{8}{8}\binom{10}{1}}{\binom{18}{9}}=.0076$$

זו מובהקות נמוכה מאוד, ולכן נדחה את השערת שוויון ההתפלגויות ונסיק ש-Y אמנם נוטה להיות גדול מ-X.

התפלגות אסימפטוטית של U

עבור מדגמים קצת יותר גדולים, וכשהתוצאה לא מאוד קיצונית, החישוב המדויק על פי ההתפלגות ההיפרגאומטרית הוא מסורבל. במקרים כאלה ניתן לעבור להתפלגות מקורבת. ההתפלגות ההיפרגאומטרית קרובה להתפלגות נורמלית כאשר N גדול (בתנאי שגם D וגם $N-D$ גדולים מספיק). לגבי מבחן החציון, התפלגות הסטטיסטי U תחת השערת האפס קרובה להתפלגות נורמלית כאשר m ו-n די גדולים.

התוחלת והשונות של ההתפלגות (18) של U, תחת השערת האפס, מתקבלות לפי הנוסחאות המתאימות עבור ההתפלגות ההיפרגאומטרית:

$$EU=m\frac{t}{N} \qquad (19)$$

$$Var(U)=m\frac{t}{N}\cdot\frac{N-t}{N}\cdot\frac{N-m}{N-1}=\frac{mnt(N-t)}{N^2(N-1)} \qquad (20)$$

ניתן לרשום את הנוסחאות במפורש עבור N זוגי ועבור N אי־זוגי.

א) אם N זוגי, $t=N/2$ ומקבלים

$$EU=\frac{m}{2} \qquad Var(U)=\frac{mn}{4(N-1)} \tag{21}$$

ב) אם N אי־זוגי, $t=(N-1)/2$ ומקבלים

$$EU=\frac{m(N-1)}{2N} \qquad Var(U)=\frac{mn(N+1)}{4N^2} \tag{22}$$

דוגמה 3.11 (המשך דוגמה 3.10). נחשב בקירוב את מובהקות התוצאה בדוגמה 3.10. היות ש- $N=18$ הוא זוגי, התוחלת והשונות מחושבות מהנוסחאות (21).

$$EU=\frac{8}{2}=4 \qquad Var(U)=\frac{8\cdot 10}{4\cdot 17}=\frac{20}{17}$$

על פי הקירוב הנורמלי (עם תיקון רציפות, כמובן)

$$P=P_{H_0}(U\geq 7)\cong 1-\Phi\left(\frac{6.5-4}{\sqrt{20/17}}\right)=1-\Phi(2.30)=.0107$$

ההסתברות המקורבת קצת גדולה מההסתברות המדויקת בחישוב בדוגמה 3.10.

לסיכום, מבחן החציון להשוואת שני מדגמים אינו לוקח בחשבון את ערכי התצפיות, אלא משתמש רק בידיעה אם כל תצפית גדולה או קטנה מחציון שני המדגמים. לפיכך מבחן זה הוא בעל עוצמה יחסית נמוכה לעומת המבחנים של ווילקוקסון או של קולמוגורוב־סמירנוב. אנו לא מביאים כאן את עוצמת מבחן החציון.

תרגילים

1* הוכיחו כי עבור N אי־זוגי שונות הסטטיסטי של אנסרי-ברדלי היא:

$$Var(A)=\frac{mn(N+1)(N^2+3)}{48N^2}$$

(ראו נספח 3.)

2. פותחה שיטה מהירה (אך כנראה פחות מדויקת) לקביעת הריכוז של חומר כימי מסוים בתמיסה. נבדקו שמונה מדגמֵי חומר בשיטה זו וכן ארבעה מדגמֵי חומר בשיטה הסטנדרטית. התוצאות הן:

שיטה סטנדרטית:	25.1	24.2	25.3	26.1				
שיטה מהירה:	23.0	18.7	22.5	28.4	17.0	25.4	19.3	16.8

א) עִרכו מבחן זיגל־טוקי לבדיקת ההבדל בדיוק קביעת הריכוז בשתי השיטות,

כאשר אתם נותנים דרגה 1 לתצפית המינימלית. תנו את מובהקות התוצאה והסיקו עבור $\alpha=.05$.

ב) ערכו מבחן כנ"ל, כאשר אתם נותנים דרגה 1 לתצפית המקסימלית. השוו לחלק א.

ג) תנו את מובהקות התוצאה (על־סמך הקירוב הנורמלי) לפי מבחן אנסרי־ברדלי.

3. סרטטו את פונקציית ההתפלגות האמפירית של כל אחד מן המדגמים להלן:

א) 2.8 3.3 2.4 2.1 1.6 3.1

ב) 1.4 1.5 1.4 2.1 1.3 1.5 1.5 1.1

ג) 0.1 -0.7 -0.3 0.1 0.0 -0.7 -0.4 -0.3 -0.7

4. יצרן טקסטיל מעוניין לברר את מידת השונוּת בהתכווצות הבד שהוא מייצר, אחרי כביסה. הוא מאמין כי אחוז ההתכווצות של הבד הוא פחות או יותר דומה בטמפרטורות מים שונות, אך טמפרטורת מים גבוהה עלולה להגדיל את השונוּת. נערכו בדיקות על ידי כביסה של דוגמאות בד זהות במים פושרים ובמים חמים. להלן התוצאות של אחוז ההתכווצות:

מים פושרים:	19.3	18.2	16.8	21.0	15.1	23.6	20.0	24.4		
מים חמים:	21.5	26.6	27.0	25.1	17.0	19.5	30.5	14.1	15.7	15.3

א) נתחו את הנתונים בעזרת מבחן זיגל־טוקי והסיקו עבור $\alpha=.05$.

ב) תארו את פונקציות ההתפלגות האמפיריות של שני המדגמים באופן גרפי. הניתן להבחין בהבדל ביניהן? באיזה כיוון?

ג) נתחו את הנתונים בעזרת מבחן קולמוגורוב-סמירנוב. השוו את מובהקות התוצאה בשני המבחנים. מה המסקנה? הסבירו!

5. חמישה אנשים חולקו אקראית לשתי קבוצות. שלושה מהם קיבלו תרופה מסוימת והשניים הנותרים קיבלו תרופת דמה (פלצבו). לאחר זמן הם דיווחו על הרגשתם:

תרופה: טוב, בינוני, טוב מאוד

דמה: בינוני, רע

א) חשבו את הערך של הסטטיסטי של קולמוגורוב-סמירנוב לבדיקת הבעיה הנידונה.

ב) חשבו את ההתפלגות **המדויקת** של הסטטיסטי תחת השערת האפס (על ידי מְנִייה).

ג) לפי חלק ב, מהי מובהקות התוצאה שהתקבלה במדגם? מה המסקנה מכאן עבור $\alpha=.10$?

ד) חשבו את מובהקות התוצאה, כשאתם משתמשים במבחן ווילקוקסון. השוו למבחן קולמוגורוב-סמירנוב.

6. חשבו על ידי מנִייה את התפלגות הסטטיסטים של קולמוגורוב-סמירנוב עבור שני מדגמים בגודל $m=2,\ n=3$. בדקו את נכונות ההתפלגויות הרשומות בדוגמה 3.6.

7. ערכו מבחן קולמוגורוב-סמירנוב לגבי נתוני תרגיל 10 בפרק 2 (טיפול בתוקפנות).

א) מבחן דו־צדדי עם $\alpha = .05$. השתמשו בנוסחת הקירוב. מה המסקנה?

ב) מבחן חד־צדדי כנגד האלטרנטיבה שהתרופה מועילה למניעת תוקפנות. מִצאו את מובהקות התוצאה בקירוב. מהי המסקנה עבור $\alpha = .05$?

ג) ערכו מבחן ווילקוקסון מתאים (חד־צדדי) והשוו את התוצאה לזו שהתקבלה בחלק ב. איזה מבחן (ווילקוקסון או קולמוגורוב-סמירנוב) נראה לכם יעיל יותר במקרה הזה? הסבירו!

8. ערכו מבחן החציון לגבי נתוני תרגיל 10 בפרק 2 (טיפול בתוקפנות).

א) מצאו את הערך הקריטי למבחן ברמת מובהקות $\alpha = .05$.

ב) מִצאו את המובהקות המדויקת של התוצאה. מה המסקנה?

ג) מצאו את המובהקות המקורבת של התוצאה על סמך קירוב נורמלי.

ד) השוו לתוצאות המבחנים האחרים שהשתמשתם לבדיקת אותה בעיה.

9. בלוח להלן התפלגות נשים העובדות במפעל גדול, לפי מספר ימי ההיעדרות מן העבודה בשנה ולפי המצב המשפחתי (סך הכול 100 נשואות ו-200 רווקות).

בְּדקו בעזרת מבחן קולמוגורוב-סמירנוב אם יש הבדל במידת ההיעדרות בין נשים נשואות לרווקות ($\alpha = .05$).

מס' ימים	נשואות	רווקות
0	15	35
1	18	40
2	15	30
3	12	25
4	9	20
5	7	18
6	3	6
7	2	6
8	3	4
9	3	3
10	3	2
11	2	1
12	1	2
13+	7	8
סך הכול	100	200

10. לגבי הנתונים בתרגיל 9 עִרכו מבחן החציון ברמת מובהקות $\alpha = .05$, כדי לבדוק אם נשים נשואות נוטות להיעדר מהעבודה יותר מרווקות.
(רמז: בגלל ערכי התיקו, המספר הכולל של תצפיות הנמוכות מהחציון המשותף קטן מ- $N/2$.)

פרק 4

מדגם מזווג

השוואה בין טיפול לביקורת (או בין שני טיפולים) ניתן לעשות גם באמצעות מדגם מזווג לכל אחד מהנבדקים המקבלים את הטיפול ניתן לפעמים להתאים נבדק אחר המתאים לו מבחינות היכולות להשפיע על תוצאת הטיפול (זה יכול להיות אפילו אותו הנבדק עצמו), והוא משמש כביקורת. לדוגמה, נניח שבודקים השפעת תרופה חדשה נגד מיגרנה על ידי נתינתה לחולה בזמן התקף והשוואת השפעתה (הזמן עד חלוף הכאב) לתרופה סטנדרטית. את התרופה הסטנדרטית ניתן לתת לאותו החולה בעת התקף מיגרנה אחר (אז הוא משמש כביקורת של עצמו), או לחולה הזהה לו מבחינת המין, הגיל, המשקל ומצב הבריאות.

את תוצאות הניסוי אנו מסמנים על ידי התצפיות הזוגיות (X_i, Y_i), $i=1,2,\ldots,n$, כאשר X_i היא תוצאת הביקורת ו- Y_i היא תוצאת הטיפול.

נתונות לנו, אפוא, n תצפיות בלתי תלויות מהתפלגות דו־ממדית כלשהי. בדרך כלל נהיה מעוניינים לבדוק את השערת האפס שאין הבדל בין הטיפול לביקורת, כנגד האלטרנטיבה שהטיפול יעיל יותר. לא נגדיר כרגע את ההשערות באופן מדויק. נעשה זאת יותר מאוחר בעזרת מודל מתאים.

המבחנים הפשוטים ביותר להשוואת שני טיפולים כאשר הניסוי נערך על מדגם מזווג מבוססים על ההפרשים בין שני הערכים עבור כל אחד מן הזוגות.

נגדיר $D_i = Y_i - X_i$ – ההפרש בין המדידה לאחר טיפול לבין המדידה של הביקורת, עבור הזוג ה-i, $i=1,\ldots,n$.

בפרק זה נביא שני מבחנים שונים המבוססים על ההפרשים הללו. יש לשים לב שמבחנים אלה מבוססים על המדגם של n ההפרשים $D_1, D_2, \ldots, D_n$ בלבד (ולא על $2n$ המדידות המקוריות).

4.1 מבחן הסימן

מבחן הסימן הוא המבחן הפשוט והקל ביותר לשימוש עבור מדגם מזווג. סטטיסטי המבחן הוא "מספר ה-D-ים החיוביים". ניתן להגדירו באופן הבא.
נסמן:

$$Z_i = \begin{cases} 1 & X_i < Y_i \\ 0 & X_i > Y_i \end{cases} = \begin{cases} 1 & D_i > 0 \\ 0 & D_i < 0 \end{cases} \qquad (1)$$

Z_i הוא המשתנה המציין (אינדיקטור) של המאורע $\{D_i > 0\}$.
הסטטיסטי של מבחן הסימן הוא

$$S_n = \#\{\ i : X_i < Y_i\ \} = \sum_{i=1}^{n} Z_i \qquad (2)$$

הנחה 4.1. נניח שהמשתנים $D_1, D_2, \ldots, D_n$ הם בלתי תלויים, בעלי התפלגות F_D רציפה.

תחת הנחה 4.1 המשתנים $Z_1, Z_2, \ldots, Z_n$ גם הם בלתי תלויים ושווי התפלגות. התפלגות כל אחד מהם היא התפלגות ברנולי $Z_i \sim B(1,p)$, כאשר

$$p = P(Z_i = 1) = P(D_i > 0) = P(X_i < Y_i) \qquad (3)$$

בגלל רציפות ההתפלגות F_D, מתקיים: $P(D_i = 0) = 0$. לכן המשתנים Z_i מוגדרים היטב. סכומם הוא, כמובן, משתנה בינומי: $S_n = \sum_{i=1}^{n} Z_i \sim B(n,p)$. באמצעות S_n ניתן לאמוד את הפרמטר p על ידי $\hat{p} = \frac{S_n}{n}$, כלומר, פרופורציית ההפרשים החיוביים.

ההשערות הנבדקות

הפרמטר היחיד שבו תלויה התפלגות סטטיסטי המבחן S_n הוא $p = P(D_i > 0)$, וההשערות תירשמנה כהשערות על p. תחת השערת האפס, שאין הבדל בין טיפול לביקורת, אנו מניחים שקיים:

$$P(X_i < Y_i) = P(X_i > Y_i) = 1/2 \qquad (4)$$

ולכן תחת השערת האפס $P(D_i > 0) = P(D_i < 0) = 1/2$

תחת אלטרנטיבה חד־צדדית, הטוענת שלאחר הטיפול תוצאת המדידה נוטה להיות גבוהה יותר, יתקיים $P(D_i > 0) > P(D_i < 0)$ כלומר, $P(D_i > 0) > 1/2$.

הערות:

א) השוויון (4) מגדיר סוג מסוים של סימטרייה בין X ל-Y (אנו אומרים שהמשתנים

הללו "ניתנים להחלפה" – exchangeable). אין זה מחייב, כמובן, שההתפלגות הדו־ממדית של (X,Y) תהיה בהכרח סימטרית. הרעיון העומד מאחורי מודל זה הוא שאם למעשה אין הבדל בין טיפול לביקורת, אזי הסיכוי לכך שאצל נבדק שקיבל טיפול יימצא ערך גבוה מאשר אצל בן־זוגו (ביקורת), שווה לסיכוי שהתוצאה תהיה הפוכה.

ב) השוויון (4) אינו אקוויוולנטי לשוויון ההתפלגויות השוליות של X ו־Y. ניתן למצוא התפלגויות דו־ממדיות שבהן ההתפלגויות השוליות שוות, אולם $P(X<Y)\neq 1/2$, וגם להיפך, ישנן התפלגויות דו־ממדיות שבהן ההתפלגויות השוליות הן שונות, ועם זאת $P(X<Y)=1/2$ (ראו דוגמאות בנספח 4).

נרשום, אפוא, את ההשערות במונחים של הפרמטר p המוגדר בנוסחה (3):

$$H_0: p=1/2 \qquad H_1: p>1/2 \tag{5}$$

אם אמנם ההשערה האלטרנטיבית נכונה, כלומר $p>1/2$, אזי נצַפה ליותר זוגות שעבורם $D>0$ ולכן אנו דוחים את השערת האפס עבור ערכים גבוהים של סטטיסטי המבחן S_n. תחת השערת האפס S_n הוא בעל התפלגות בינומית, $S_n \sim B(n,1/2)$, ולכן מובהקות התוצאה ניתנת לחישוב פשוט באמצעות ההתפלגות הבינומית עם פרמטר $p=1/2$.

מבחן הסימן

נתונות n תצפיות בלתי תלויות של זוגות (X_i,Y_i), $i=1,\ldots,n$, ויהיו $D_i=Y_i-X_i$, $i=1,\ldots,n$, ההפרשים המתאימים. סטטיסטי המבחן S_n, הוא מספר הזוגות שעבורם $Y_i>X_i$, והוא נתון בנוסחה (2). לבדיקת ההשערות (5) דוחים את H_0 עבור ערכים גבוהים של S_n, שהתפלגותו תחת השערת האפס היא בינומית – $S_n \sim B(n,1/2)$.

מובהקות התוצאה. אם במדגם של n זוגות התקבלה התוצאה $S_n=a$, אזי מובהקות התוצאה היא

$$P=P_{H_0}(S_n\geq a)=\sum_{k=a}^{n}\binom{n}{k}\left[\frac{1}{2}\right]^n=\left[\frac{1}{2}\right]^n\sum_{k=a}^{n}\binom{n}{k} \tag{6}$$

את ההסתברויות הללו קל לחשב כאשר n לא גדול. טבלאות של ההסתברויות הבינומיות נמצאות בספרים רבים. בפרט, בטבלה 3 בנספח מוצגות טבלאות בינומיות $B(n,1/2)$, עבור $2\leq n\leq 30$. לשימוש בטבלה, שימו לב שהתפלגות זו היא סימטרית סביב $n/2$.

דוגמה 4.1. הנתונים בלוח 4.1 הם חלק ממחקר שנערך לגבי אנשים שנפגעו בהלם קרב (ד"ר יפה זינגר). לגבי כל אחד מ-15 נפגעים נרשם ציון החומרה של מצבו, כפי שדווח על ידי המטפל שלו וכפי שדווח על ידי החולה הנפגע בעצמו. הציון נקבע על ידי ממוצע הערכות החומרה שניתנו עבור 17 פריטים. כל פריט מקבל ערכים בין 0 = בכלל לא חמור, עד 4 = חמור. אנו מסמנים כאן ב-X_i את ציון החומרה כפי שדווח על ידי החולה ה-i וב-Y_i את ציון החומרה שנתן המטפל שלו. נרצה לבדוק אם הנפגע עצמו רואה את מצבו כחמור יותר מאשר רואה זאת המטפל שלו.

לוח 4.1. *ציוני הערכה של חומרת מצבם של חולים לפי דיווח עצמי ולפי דיווח המטפל*

החולה	הערכת המטפל (X)	הערכת החולה (Y)	הפרש $(Y-X)$	Z
1	2.94	3.41	0.47	1
2	2.65	3.53	0.88	1
3	2.06	3.47	1.41	1
4	2.29	3.35	1.06	1
5	2.24	2.12	–0.12	0
6	1.47	1.24	–0.24	0
7	2.59	2.00	–0.59	0
8	1.48	2.06	0.58	1
9	1.76	1.94	0.18	1
10	3.65	3.24	–0.41	0
11	2.88	3.88	1.00	1
12	3.12	2.76	–0.35	0
13	0.65	1.41	0.76	1
14	2.94	3.65	0.71	1
15	1.47	0.29	–1.18	0

אם אין הבדל בין ההערכה העצמית של החולה להערכת המטפל, נצַפה שציון החולה לא יהיה בהכרח גבוה מציון המטפל שלו. בהתאם לכך נוכל לרשום את ההשערות הנבדקות: $H_0: p=1/2$ כנגד $H_1: p>1/2$, כאשר p, כפי שמוגדר בנוסחה (3), הוא הסיכוי לכך שהערכת החולה לגבי חומרת מצבו גבוהה מהערכת המטפל. (ההשערה האלטרנטיבית שרשמנו היא בהתאמה לתֵאוריה, הטוענת שהחולים נוטים להעריך את מצבם כחמור יותר מאשר המטפלים שלהם.) הסטטיסטי של מבחן הסימן הוא מספר החולים שההערכה העצמית שלהם גבוהה מהערכת המטפל. ערכי המשתנה המציין Z, נוסחה (1), רשומים גם הם בלוח וניתן לראות כי התקבלה התוצאה $S_{15}=9$.

נשים לב שכדי לערוך את מבחן הסימן, למעשה אין צורך לחשב את ההפרשים D_i,

מכיוון שהסטטיסטי אינו תלוי בערכם של ההפרשים, אלא רק בסימנם, שקל לזהות ישירות מן הנתונים.

מובהקות התוצאה היא ההסתברות לקבלת ערך כה גבוה של הסטטיסטי, בהנחה שהשערת האפס נכונה. לפי נוסחה (6):

$$P = P_{H_0}(S_{15} \ge 9)$$

$$= \left[\frac{1}{2}\right]^{14}\left[\binom{15}{9}+\binom{15}{10}+\binom{15}{11}+\binom{15}{12}+\binom{15}{13}+\binom{15}{14}+\binom{15}{15}\right]$$

$$= \frac{9{,}949}{2^{15}} = .3036$$

את המובהקות הזאת אפשר גם למצוא בטבלת ההתפלגות של מבחן הסימן, טבלה 3. על סמך הסימטרייה של התפלגות S_n (הערך $S_{15}=9$ סימטרי לערך $S_{15}=15-9=6$), נמצא בטבלה: $P = P_{H_0}(S_{15} \ge 9) = P_{H_0}(S_{15} \le 6) = .3036$.

מובהקות התוצאה שהתקבלה היא גדולה, ולכן עבור כל רמת מובהקות סבירה התוצאה אינה מובהקת. המסקנה היא שאין לדחות את השערת האפס, כלומר, אי אפשר לומר שהערכות החולים עצמם גבוהות מהערכות המטפלים שלהם. נראה שההערכות די דומות.

קירוב נורמלי למבחן הסימן

כאשר n די גדול ניתן להשתמש בקירוב הנורמלי עבור ההתפלגות הבינומית. לגבי S_n מדובר בהתפלגות בינומית עם פרמטר $p=1/2$. באופן כללי, הקירוב הנורמלי עבור התפלגות בינומית $B(n,p)$ הוא די טוב כאשר $np \ge 5$ וגם $n(1-p) \ge 5$. עבור הסטטיסטי של מבחן הסימן, דרוש $n/2 \ge 5$, או $n \ge 10$. (טבלת ההתפלגות של מבחן הסימן נותנת לנו את ההסתברויות המדויקות גם עבור ערכים גדולים בהרבה.)

התוחלת והשונות של S_n תחת השערת האפס הן

$$(7) \qquad ES_n = n \cdot \frac{1}{2} = \frac{n}{2} \qquad\qquad Var(S_n) = n \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{n}{4}$$

ולפיכך מקבלים את ההתפלגות המקורבת על ידי

$$(8) \qquad P_{H_0}(S_n \le k) \cong \Phi\left(\frac{k + 1/2 - n/2}{\sqrt{n/4}}\right) \qquad k = 0,1,\ldots,n$$

השתמשנו כאן בתיקון רציפות.

דוגמה 4.2 (המשך דוגמה 4.1). נחשב את מובהקות התוצאה בעזרת הקירוב הנורמלי. התוצאה שהתקבלה היא $S_{15}=9$.

התוחלת היא $ES_{15}=7.5$ והשונות $Var(S_{15})=3.75$.
המובהקות המקורבת של התוצאה היא

$$P=P_{H_0}(S_{15}\geq 9)\cong 1-\Phi\left(\frac{8.5-7.5}{\sqrt{3.75}}\right)=1-\Phi(0.52)=.3015$$

בהשוואה למובהקות המדויקת – $P=.3036$, הקירוב טוב מאוד.

4.2 בעיות של ערכי תיקו במבחן הסימן

כשהתפלגות ההפרש F_D היא רציפה, אין בעיות של תיקו, מכיוון ש-$P(D_i=0)=0$ לכל $i=1,\ldots,n$, וכך המשתנה המציין Z_i בנוסחה (1) מוגדר היטב. במקרה שההתפלגות F_D אינה רציפה, ייתכן שההסתברות של הערך 0 היא חיובית. למעשה, זוהי הבעיה היחידה בשימוש במבחן הסימן במקרה ש-F_D אינה רציפה. אין כל בעיה לחשב את הסטטיסטי S_n במקרים אחרים של תיקו, למשל, אם חלק מה-X-ים או חלק מה-Y-ים שווים ביניהם, או גם אם חלק מההפרשים D שווים ביניהם.
נסמן את ההסתברויות הבאות:

$$p^+=P(D>0)\qquad p^-=P(D<0)\qquad p^0=P(D=0)\tag{9}$$

מובן שסכום שלוש ההסתברויות הללו הוא $p^++p^-+p^0=1$.
השערת האפס, שאין הבדל בין טיפול לביקורת, ניתנת להירשם על ידי

$$H_0: p^+=p^-\tag{10}$$

השערה זו טוענת שהסיכוי לכך שתוצאת הטיפול **עדיפה** על הביקורת שווה לסיכוי שתוצאת הביקורת **עדיפה** על הטיפול. אין כאן כל התייחסות לסיכוי שהתוצאות תהיינה שוות.
השערת האפס טוענת, למעשה, שקיים השוויון $P(X_i<Y_i)=P(X_i>Y_i)$, לכל i.
נניח עתה שההסתברות לשוויון בין X ל-Y היא חיובית, כלומר: $p^0=P(D=0)>0$.
במקרה זה סכום שתי ההסתברויות האחרות p^+ ו-p^- קטן מ-1. כך, גם אם השערת האפס נכונה, באופן ששתי הסתברויות אלה שוות ביניהן, כל אחת בנפרד קטנה ממש מ-1/2 ולמעשה אינה ידועה לנו.
התפלגות הסטטיסטי של מבחן הסימן S_n – מספר ההפרשים **החיוביים** – היא בינומית $S_n\sim B(n,p^+)$.
היות שגם תחת השערת האפס הערך של הפרמטר p^+ אינו ידוע, הסטטיסטי S_n איננו אפרמטרי במקרה הזה.
כדי לקבל סטטיסטי אפרמטרי, שהתפלגותו אינה תלויה בפרמטר לא ידוע, נערוך את

השינוי הבא. נסתכל על ההסתברות **המותנית** לקבלת הפרש חיובי, בתנאי שההפרש איננו אפס:

$$p_1^+ = P(D>0/D\neq 0) = \frac{P(D>0)}{P(D\neq 0)} = \frac{p^+}{p^+ + p^-}$$

באופן דומה, ההסתברות **המותנית** לקבלת הפרש שלילי, אם ההפרש איננו אפס היא

$$p_1^- = P(D<0/D\neq 0) = \frac{P(D<0)}{P(D\neq 0)} = \frac{p^-}{p^+ + p^-}$$

תחת השערת האפס (10) שתי ההסתברויות המותנות הללו שוות: $p_1^+ = p_1^-$. כיוון שסכומן הוא 1, כל אחת מהן שווה ל-1/2. לכן ניתן לרשום את ההשערות הנבדקות כהשערות על ההסתברות המותנית:

$$H_0: p_1^+ = 1/2 \qquad H_1: p_1^+ > 1/2 \tag{11}$$

כדי לבדוק את השערת שוויון ההסתברויות המותנות, נסתכל רק על התצפיות שעבורן $D_i \neq 0$, כלומר רק על אלה שבהן $X_i \neq Y_i$. את סטטיסטי המבחן נגדיר בדיוק כמו קודם, על ידי מספר התצפיות שעבורן $X_i < Y_i$.

$$S_n = \#\left\{i: X_i < Y_i\right\}$$

נסמן ב-N את מספר ההפרשים השונים מאפס: $N = \#\{i: D_i \neq 0\}$. יש לשים לב ש-N הוא משתנה מקרי התלוי בתוצאות הניסוי, והוא תמיד קטן או שווה ל-n.

טענה 4.1. תחת השערת האפס (10) פונקציית ההתפלגות **המותנית** של S_n, בתנאי ש-$N = k$, היא בינומית:

$$S_n | N = k \sim B(k, 1/2) \tag{12}$$

ההוכחה ברורה. תחת השערת האפס, מתוך k ההפרשים השונים מאפס שכולם בלתי תלויים, ההסתברות המותנית לכך שהפרש מסוים הוא חיובי, היא $p_1^+ = 1/2$. ♣

בהתאם לטענה 4.1 לעיל, משתמשים במקרה של ערכי תיקו בשיטה הבאה:

א. מסלקים מן המדגם את כל הזוגות של תצפיות שעבורן התקבל תיקו ($x_i = y_i$).

ב. מפעילים את מבחן הסימן הרגיל על N התצפיות שנותרו ($N \leq n$).

דוגמה 4.3. בלוח 4.2 נתונים חלקיים מתוך מחקר לגבי שימוש במִלות שאלה. שיחות עם ילדים בני שנתיים הוקלטו במשך זמן קצוב ונרשם מספר הפעמים שהילדים השתמשו בכל אחת משתי המִלים "מה?" ו-"איפה?". אצל שניים מבין 17 הפעוטות שהשתתפו בניסוי, מספר הפעמים שאמרו "מה" היה זהה למספר הפעמים שאמרו "איפה". אחרי הוצאתם מהמדגם נותרו $N = 15$ תצפיות. מספר ההפרשים החיובים הוא

$S_{15}=4$. תחת השערת האפס התפלגות S_{15} היא בינומית $B(15,1/2)$. הבעיה כאן היא דו־צדדית (אין תֵאוריה לגבי המילה היותר שכיחה אצל פעוטות).

לוח 4.2. *מספר הפעמים שפעוטות השתמשו במִלות שאלה*

הפעוט	מה (Y)	איפה (X)	$D=Y-X$	Z
1	0	1	–1	0
2	1	4	–3	0
3	1	4	–3	0
4	3	4	–1	0
5	1	3	–2	0
6	1	0	1	1
7	0	5	–5	0
8	2	7	–5	0
9	3	6	–3	0
10	1	8	–7	0
11	0	0	0	–
12	9	7	2	1
13	2	1	1	1
14	0	3	–3	0
15	0	0	0	–
16	6	2	4	1
17	14	18	–4	0

התוצאה שהתקבלה נמצאת בזנב השמאלי של ההתפלגות. הסתברות הזנב היא

$$P_{H_0}(S_{15}\le 4)=\frac{1}{2^{15}}\left[1+15+\binom{15}{2}+\binom{15}{3}+\binom{15}{4}\right]=0.0592$$

ערך זה נמצא גם בטבלת מבחן הסימן, טבלה 3, עבור $n=15$. ההסתברות לעיל היא, למעשה, ההסתברות **המותנית** לקבלת 4 הפרשים חיוביים לכל היותר, מתוך 15 הזוגות שבהם ההפרש שונה מאפס.

למציאת מובהקות התוצאה יש לכפול את הסתברות הזנב לעיל ב-2 (ההתפלגות הבינומית עם $p=1/2$ היא סימטרית), ולכן המובהקות שהתקבלה כאן היא גבוהה – $P=2(.0592)=.1184$. המסקנה היא שאין הבדל מובהק בין השימוש בשתי מִלות השאלה הללו.

4.3 שימוש במבחן הסימן לבדיקה לגבי ערכי חלוקה שונים

עד עתה התייחסנו למבחן הסימן כמבחן לגבי החציון של התפלגות המשתנה D (הפרש בין זוג תצפיות בבעיית מדגם מזווג). נסתכל עתה על מדגם בודד $X_1, X_2,, X_n$. מובן שניתן להשתמש במבחן הסימן כדי לבדוק השערה לגבי חציון התפלגות המשתנה X. נראה זאת בהמשך בפירוט. אולם באופן דומה ניתן לבדוק השערות לגבי ערכי חלוקה אחרים, כמו הרבעון התחתון, העשירון העליון וכד׳.

אנו מסמנים ב-x_p את ערך החלוקה ה-p של התפלגות המשתנה X, כלומר: $P(X \le x_p) = p$.

לדוגמה, לפי הסימון הזה $x_{.5}$ הוא החציון, $x_{.25}$ הוא הרבעון התחתון ו-$x_{.90}$ הוא העשירון העליון.

נניח שיש לבדוק את ההשערות: $H_0 : x_p = a$ כנגד $H_1 : x_p > a$, כאשר p היא הסתברות מסוימת ו-a הוא ערך נתון כלשהו. אנו בודקים כאן אם ערך החלוקה ה-p של התפלגות X שווה ל-a. למשל, השערת אפס מהצורה $H_0 : x_{.10} = 7$ משמעה שהעשירון התחתון של התפלגות X הוא 7, או במִלים אחרות, שהתפלגות המשתנה X מקיימת: $P(X \le 7) = 0.10$.

לבדיקת השערה כזאת נסתכל על סטטיסטי דומה לזה המוגדר בנוסחה (2) – מספר התצפיות הגדולות מהערך המשוער a: $S_n(a) = \#\{i : X_i > a\}$.

התצפיות הן בלתי תלויות, ולכן תחת H_0, $S_n(a)$ מתפלג בינומית:

$S_n(a) \sim B(n,\, 1-p)$. ($1-p$ היא ההסתברות שתצפית בודדת **גדולה** מהערך a.)

בהתאם להשערה האלטרנטיבית, במקרה ש-H_0 איננה נכונה נצפה לקבל ערכים גבוהים של סטטיסטי המבחן (יותר תצפיות יהיו גדולות מן הערך המשוער a). לכן נדחה את השערת האפס עבור ערכים גבוהים של $S_n(a)$. מובהקות התוצאה $S_n(a) = s$ היא, אפוא:

$$P = P_{H_0}\{S_n(a) \ge s\} = \sum_{k=s}^{n} \binom{n}{k} (1-p)^k p^{n-k}$$

אם n קטן, ניתן לחשב הסתברות זו בעזרת מחשבון כיס. עבור n גדול יותר אפשר להיעזר בטבלאות של התפלגות בינומית שנמצאות בספרי הסתברות (למשל Devore, 1991; Walpole & Myers, 1972 ורבים אחרים). אם ההשערה היא לגבי החציון $x_{.5}$, אזי $S_n(a) \sim B(n, 1/2)$ והטבלה המתאימה היא טבלת מבחן הסימן, טבלה 3. כפי שהערנו קודם, קירוב נורמלי יהיה מתאים במקרה ש-n די גדול ומקיים את שני התנאים: $np \ge 5$ וגם $n(1-p) \ge 5$.

דוגמה 4.4. בייצור השוטף במפעל לייצור פרופילים מאלומיניום, כ-25% מהפרופילים

מיוצרים בעובי שאינו עולה על 2 מ"מ. כדי לערוך ביקורת איכות בודקים בכל יום מדגם של 20 פרופילים ורושמים את העובי X של כל פרופיל. ההשערות הנבדקות הן $H_0 : X_{.25} = 2$ כנגד $H_1 : X_{.25} > 2$. משמעות ההשערה האלטרנטיבית היא שהרבעון התחתון גבוה מ-2 מ"מ, או במִלים אחרות, שהפרופילים נוטים להיות עבים מדי.

כדי להשתמש במבחן הסימן, די לספור כמה פרופילים עבים מדי (יותר מ-2 מ"מ). המספר הזה הוא $S_{20}(2) = \#\{i : X_i > 2\}$. יש לדחות את השערת האפס עבור ערכים גבוהים של הסטטיסטי $S_{20}(2)$, כאשר תחת השערת האפס $S_{20}(2) \sim B(20,0.75)$.

שיטת ההחלטה לגבי הבדיקה יכולה להיקבע בשתי צורות.

1) בכל יום ניתן לחשב את מובהקות תוצאת המדגם של אותו יום, על סמך ההתפלגות הבינומית $B(20,0.75)$, כדי להחליט אם המדגם עומד בדרישות. אם, למשל, התקבלו ביום אחד 8 פרופילים שעוביים יותר מ-2 מ"מ (פגומים), המובהקות היא ההסתברות לקבלת 8 פגומים לפחות, תחת השערת האפס:

$$P = P_{H_0}\{S_{20}(2) \geq 8\} = \sum_{k=8}^{20} \binom{20}{k} (0.75)^k (0.25)^{20-k} = 0.1018$$

ההסתברות לעיל התקבלה לפי טבלה בינומית. עבור רמות מובהקות מקובלות התוצאה אינה מובהקת.

2) ניתן **לחשב מראש** את הערך הקריטי של הסטטיסטי, שהָחֵל ממנו יש לדחות את השערת האפס עבור רמת המובהקות שנקבעה מראש. נניח שדרושה רמת מובהקות $\alpha = 0.05$. כדי למצוא את הערך הקריטי נצטרך למצוא את האחוזון המתאים של ההתפלגות הבינומית $B(20,0.75)$. שוב, לפי טבלה בינומית ניתן לראות שההסתברות לקבלת לפחות 9 מוצרים עבים מדי היא

$$P_{H_0}\{S_{20}(2) \geq 9\} = \sum_{k=9}^{20} \binom{20}{k} (0.75)^k (0.25)^{20-k} = 0.0409$$

קודם ראינו שההסתברות לקבלת 8 מוצרים כאלה לפחות, כבר גדולה מ-0.10. לכן 9 הוא הערך הקריטי כאן. אם כן, ניתן לקבוע מראש, שבכל יום שבו עורכים את הביקורת, אם $S_{20}(2) \geq 9$ (ימצאו לפחות תשעה מוצרים פגומים) יש לדחות את השערת האפס ולבדוק אם קרתה תקלה בעת הייצור.

4.4 מבחן ווילקוקסון למדגם אחד

המבחן של ווילקוקסון למדגם אחד, או למדגם מזווג (Wilcoxon, 1945) מבוסס גם על הפערים בין זוגות התצפיות, נוסף על הסימן שלהם.

נחזור שוב על הגדרת הבעיה.

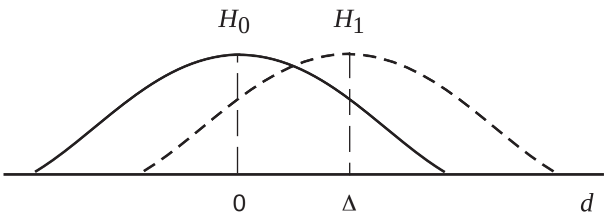
נתונים (X_i,Y_i), $i=1,\ldots,n$, n זוגות בלתי תלויים ונגדיר את $D_i=Y_i-X_i$ ההפרשים המתאימים.

הנחה 4.2. נניח שהמשתנים $D_1,D_2,\ldots,D_n$ הם בלתי תלויים, בעלי התפלגות **רציפה וסימטרית**, שנסמנה F_D.

ההבדל בין הנחה 4.2 להנחה 4.1 הוא שכאן אנו מניחים גם סימטרייה של ההתפלגות. נראה יותר מאוחר בפרק זה מדוע הנחת הסימטרייה חשובה.
היות שהתפלגות ההפרשים – F_D סימטרית, ההשערות לגבי הפער בין X ל-Y ניתנות להירשם כהשערות על חציון התפלגות ההפרשים D, כלומר על חציון ההתפלגות F_D.
את הבעיה החד־צדדית נרשום, אפוא:

$$H_0: Med(F_D)=0 \qquad H_1: Med(F_D)>0 \tag{13}$$

או בצורה אחרת
$$H_0: d_{.5}=0 \qquad H_1: d_{.5}>0$$
כאשר $d_{.5}$ הוא חציון התפלגות המשתנה D.
משמעות ההשערה האלטרנטיבית היא שההפרש $D=Y-X$ נוטה להיות חיובי, כלומר, Y נוטה להיות גדול מ-X. בציור 4.1 ניתן לראות שתי התפלגויות כאלה, האחת – סימטרית בעלת חציון 0 והשנייה – סימטרית בעלת חציון חיובי Δ.



ציור 4.1. *דוגמה לצפיפויות סימטריות תחת המודלים המוגדרים ב-H_0 וב-H_1*

סטטיסטי המבחן של ווילקוקסון

כאן אנו לוקחים בחשבון לא רק את הסימן של ההפרש D, אלא גם את גודל המרחק שלו מן האפס.
מחשבים את הערך המוחלט של כל אחד מההפרשים – $|D_1|,|D_2|,\ldots,|D_n|$, ומדרגים את הערכים המוחלטים הללו מ-1 (הערך המוחלט הקטן ביותר) עד n (הערך המוחלט הגדול ביותר).
נסמן ב- $r(|D_i|)$ את הדרגה של $|D_i|$ וכן נסמן ב-Z_i את המשתנה המציין של הפרש חיובי, כמו בנוסחה (1).

סטטיסטי המבחן של ווילקוקסון מוגדר על ידי

$$V^+ = \sum_{i=1}^{n} r(|D_i|) \cdot Z_i = \sum_{i:D_i>0} r(|D_i|) \qquad (14)$$

הסטטיסטי V^+ הוא סכום הדרגות שקיבלו הערכים המוחלטים של ההפרשים החיוביים.

דוגמה 4.5 (המשך דוגמה 4.1). נסתכל שוב על ההערכה העצמית והערכת המטפל לגבי מצבם של 15 נפגעי הלם קרב, ונוסיף את דרגות הערכים המוחלטים (לוח 4.3).

לוח 4.3. *הערכות חומרת מצב החולים לפי דיווח עצמי ולפי דיווח המטפל*

i	המטפל (X_i)	החולה (Y_i)	$D_i = Y_i - X_i$	Z_i	$r(\|D_i\|)$
1	2.94	3.41	0.47	1	6
2	2.65	3.53	0.88	1	11
3	2.06	3.47	1.41	1	15
4	2.29	3.35	1.06	1	13
5	2.24	2.12	–0.12	0	1
6	1.47	1.24	–0.24	0	3
7	2.59	2.00	–0.59	0	8
8	1.48	2.06	0.58	1	7
9	1.76	1.94	0.18	1	2
10	3.65	3.24	–0.41	0	5
11	2.88	3.88	1.00	1	12
12	3.12	2.76	–0.35	0	4
13	0.65	1.41	0.76	1	10
14	2.94	3.65	0.71	1	9
15	1.47	0.29	–1.18	0	14

הסטטיסטי V^+ הוא סכום הדרגות השייכות לערכי $d>0$, או אלה שעבורם $z=1$:

$$V^+ = \sum_{i=1}^{n} r(|d_i|) \cdot z_i = \sum_{i:D_i>0} r(|d_i|)$$
$$= 6+11+15+13+7+2+12+10+9 = 85$$

ניתן, כמובן, לחשב את V^+ בעזרת הדרגות של הערכים השליליים. נרשום

$$V^- = 1+3+8+5+4+14 = 35$$

היות שסכום כל הדרגות הוא קבוע:

$$V^+ + V^- = 1+\cdots+n = n(n+1)/2 = 15(16)/2 = 120$$

מקבלים $V^+ = 120 - V^- = 120 - 35 = 85$

הגדרת הסטטיסטי V^+ בצורה שונה

כדי להקל על חישובים שונים הקשורים בסטטיסטי V^+, ניתן לרשום אותו בצורה שונה מזו שרשמנו בנוסחה (14). נסדר את רשימת הנבדקים, כפי שרשמנו בלוח 4.3, לפי גודל הערך המוחלט של ההפרש D. תתקבל הרשימה החדשה בלוח 4.4. הנתונים בלוח 4.4 זהים לאלה הרשומים בלוח 4.3, פרט לכך שהנבדקים מסודרים בסדר שונה.

לוח 4.4. *הערכות חומרת מצב החולים, המסודרים לפי דרגת הערך המוחלט של ההפרש*

j	המטפל (X_j)	החולה (Y_j)	$D_j = Y_j - X_j$	T_j	$r(\|D_j\|) = j$
1	2.24	2.12	–0.12	0	1
2	1.76	1.94	0.18	1	2
3	1.47	1.24	–0.24	0	3
4	3.12	2.76	–0.35	0	4
5	3.65	3.24	–0.41	0	5
6	2.94	3.41	0.47	1	6
7	1.48	2.06	0.58	1	7
8	2.59	2.00	–0.59	0	8
9	2.94	3.65	0.71	1	9
10	0.65	1.41	0.76	1	10
11	2.65	3.53	0.88	1	11
12	2.88	3.88	1.00	1	12
13	2.29	3.35	1.06	1	13
14	1.47	0.29	–1.18	0	14
15	2.06	3.47	1.41	1	15

אנו רואים, כמו בלוח 4.3, שהדרגות השייכות להפרשים שליליים הן הדרגות 1, 3, 4, 5, 8 ו-14. סכום הדרגות של ההפרשים החיוביים הוא, כמובן, כמו קודם, $V^+ = 85$. המספר הסידורי j נקבע לפי הדרגה, במקום לפי הסדר שבו נרשמו הנבדקים במקור (סדר אלפביתי).

עקרונית אין הבדל בין צורות הרישום הללו, אלא שלפי לוח 4.4 קל יותר לרשום את הנוסחה עבור הסטטיסטי V^+, וחשוב יותר, קל מאוד לחשב עבורו תוחלת ושונות.

נגדיר, אפוא, את הסטטיסטי V^+ לפי צורת הרישום לעיל. נסמן את ערכי ההפרשים המסודרים לפי ערכם המוחלט על ידי $D^{(1)}, D^{(2)}, \ldots, D^{(n)}$, כאשר $D^{(j)}$ הוא ההפרש D שדרגת הערך המוחלט שלו היא j. כך, $D^{(1)}$ הוא ההפרש הקטן ביותר בערכו המוחלט ו-$D^{(n)}$ הוא ההפרש הגדול ביותר בערכו המוחלט. המשתנים המציינים מוגדרים

באופן הבא:

$$T_j = \begin{cases} 1 & D^{(j)} > 0 \\ 0 & D^{(j)} < 0 \end{cases} \qquad (15)$$

במִלים אחרות, T_j מקבל את הערך 1 אם הדרגה j שייכת להפרש חיובי, ואת הערך 0 אם היא שייכת להפרש שלילי.

בַדוגמה בלוח 4.4, $T_1 = 0$, כיוון שההפרש המינימלי בערכו המוחלט שייך להפרש שלילי, אבל $T_2 = 1$, כיוון שהדרגה 2 שייכת להפרש חיובי.

הסטטיסטי V^+ ניתן להירשם, אפוא, לפי נוסחה (14) על ידי

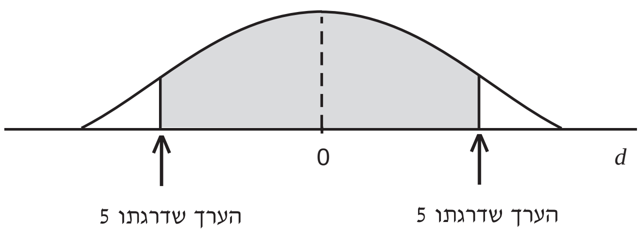
$$V^+ = \sum_{j=1}^{n} j \cdot T_j \qquad (16)$$

התפלגות הסטטיסטי V^+ של ווילקוקסון תחת השערת האפס

נסתכל על הצורה (16) שבה רשמנו את V^+. המשפט הבא נותן את התפלגות המשתנים $T_1, T_2, \ldots, T_n$.

משפט 4.1. אם השערת האפס נכונה, אזי תחת הנחה 2 המשתנים $T_1, T_2, \ldots, T_n$ הם בלתי תלויים וכל אחד מהם הוא משתנה ברנולי, עם פרמטר $p = 1/2$. כלומר $T_j \sim B(1,1/2)$ עבור $j = 1, \ldots, n$.

*הוכחה**: תחת השערת האפס החציון של התפלגות ההפרשים F_D הוא 0, וכמו כן ההתפלגות F_D היא סימטרית (ולכן סימטרית סביב אפס). בהנחות אלה, נסתכל על הפרש D שדרגתו, למשל, 1 (זו התצפית **הקרובה ביותר** ל-0). אינטואיטיבית ניתן לראות שבגלל הסימטרייה של ההתפלגות, הסיכוי שתצפית זו היא חיובית שווה לסיכוי שתצפית זו היא שלילית. אותו דבר נכון לכל אחת מן הדרגות – הסיכוי שדרגה מסוימת (של הערך המוחלט) שייכת לערך חיובי שווה לסיכוי שהיא שייכת לערך שלילי. ציור 4.2 מדגים זאת עבור הדרגה $j = 5$.



ציור 4.2. *התפלגות סימטרית סביב 0*

כדי להוכיח זאת באופן מתמטי, נַראה ראשית שהערך המוחלט $|D_i|$ בלתי תלוי בסימן

של D_i, על ידי כך שנוכיח שהסתברות החיתוך של שני מאורעות מתאימים שווה למכפלת ההסתברויות.

בגלל הסימטרייה של F_D,

$$P(|D_i| \le t) = P(-t \le D_i \le t) = 2[F_D(t) - 1/2]$$

כמו כן, לכל i המשתנה המציין Z_i הוא משתנה ברנולי $Z_i \sim B(1,1/2)$ ולכן $P(Z_i = 1) = 1/2$.

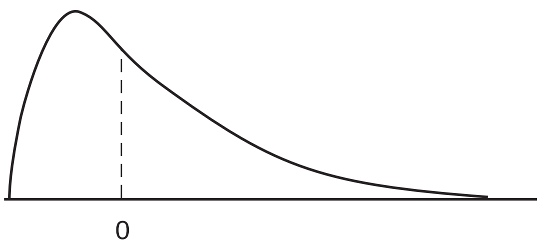
הסתברות החיתוך של שני המאורעות לעיל היא

$$\begin{aligned} P\left\{(|D_i| \le t) \cap (Z_i = 1)\right\} &= P\left\{(|D_i| \le t) \cap (D_i > 0)\right\} \\ &= P(0 < D \le t) = F_D(t) - 1/2 \\ &= P(|D_i| \le t) \times P(Z_i = 1) \end{aligned}$$

קיבלנו שהסתברות החיתוך שווה למכפלת ההסתברויות, ולכן המשתנים Z_i ו- $|D_i|$ הם בלתי תלויים. במִלים אחרות, הסימן של הפרש מסוים D אינו תלוי בערך המוחלט שלו. מכיוון שכל n המשתנים $Z_1, Z_2, \ldots, Z_n$ הם בלתי תלויים, וכן המשתנים $|D_1|, |D_2|, \ldots, |D_n|$ גם הם בלתי תלויים, קיבלנו שכל המשתנים המציינים את הסימן הם בלתי תלויים בכל הערכים המוחלטים. מכאן, כמובן, נובע שהדרגה של $|D_i|$ (הקשורה לרשימת כל הערכים המוחלטים) אינה תלויה בסימן שלו. המשתנה T_j מציין את הסימן של ההפרש שדרגתו j, והיות שהסימן לא תלוי בדרגה, נובע שהמשתנה מתפלג בינומית, כמו Z_j.

♣

שימו לב שאם ההתפלגות F_D אינה סימטרית, משפט 4.1 אינו נכון. נסתכל, לדוגמה, על ציור 4.3. ההתפלגות בציור היא בעלת חציון אפס, אולם קל לראות שהסיכוי שדרגה גבוהה (של ערך רחוק מהחציון) תהיה שייכת להפרש חיובי הוא גדול יותר בהשוואה לערך שלילי, שצפוי שיהיה קרוב יותר ל-0. משפט 4.1 נכון, אפוא, רק בהנחה 4.2 (אי-תלות בין התצפיות **וסימטרייה** של התפלגות ההפרש).



ציור 4.3. *צפיפות אסימטרית בעלת חציון 0 וזנב ימני*

נשתמש עתה במשפט 4.1 לחישוב התפלגות הסטטיסטי V^+, תחת השערת האפס.

מסקנה 4.1. ממשפט 4.1 נובע שכל אחת מהדרגות $1,2,\ldots,n$ יכולה להיות שייכת להפרש חיובי או שלילי בהסתברות $1/2$. בגלל האי-תלות בין המשתנים המציינים $T_1,T_2,\ldots,T_n$, הסתברות של כל קומבינציה של רשימת סימנים (+ או –) ליד כל אחת מ-n הדרגות היא בעלת הסתברות שווה, השווה ל- $\frac{1}{2^n}=\left[\frac{1}{2}\right]^n$. לפיכך פונקציית ההסתברות של V^+ היא

$$P_{H_0}(V^+=v)=\frac{\#\{V^+=v\}}{2^n} \qquad (17)$$

ערכי V^+ האפשריים הם בין 0 (כל ההפרשים שליליים) לבין $n(n+1)/2$ (כולם חיוביים).

דוגמה 4.6. נראה כאן כיצד ניתן לחשב את התפלגות V^+ עבור מדגם של 4 תצפיות (של זוגות). עבור $n=4$ ישנם בסך הכול $2^4=16$ צירופים אפשריים של סימנים. רשמנו את כולם בלוח 4.5.

לוח 4.5. *רשימת כל הצירופים האפשריים של סימני דרגות ההפרשים, עבור $n = 4$*

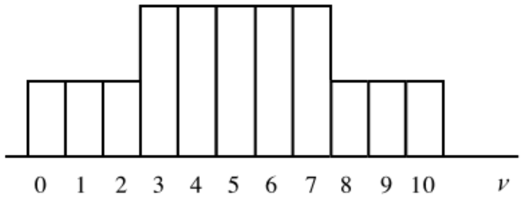
V^+	4	3	2	1	V^+	4	3	2	1
5	–	+	+	–	0	–	–	–	–
6	+	–	+	–	1	–	–	–	+
7	+	+	–	–	2	–	–	+	–
6	–	+	+	+	3	–	+	–	–
7	+	–	+	+	4	+	–	–	–
8	+	+	–	+	3	–	–	+	+
9	+	+	+	–	4	–	+	–	+
10	+	+	+	+	5	+	–	–	+

(בשתי קבוצות העמודות 1–4 מופיעה הכותרת: הדרגות)

לפי הרשימה בַלוח, התפלגות V^+ תחת השערת האפס היא

$$P_{H_0}(V^+=v)=\begin{cases}1/16 & v=0,1,2,8,9,10\\ 2/16 & v=3,4,5,6,7\end{cases}$$

ההיסטוגרמה של התפלגות זו ניתנת בציור 4.4. זו התפלגות סימטרית סביב הערך $v=5$. נראה מיד שאין זה מקרה.



ציור 4.4. התפלגות V^+ תחת H_0, $n=4$

על פי מסקנה 4.1 ניתן לחשב את התפלגות הסטטיסטי V^+ תחת השערת האפס עבור כל גודל מדגם n על ידי הספירה המתאימה של האפשרויות, על פי הנוסחה (17). טבלת ההתפלגות הזאת, עבור גודלי מדגם בין 2 ל-15, מוצגת בטבלה 4 בנספח.

דוגמה 4.7 (המשך דוגמה 4.5). מטבלה 4 בנספח ניתן למצוא את מובהקות התוצאה $V^+=85$ שהתקבלה במבחן ווילקוקסון, עבור מדגם של $n=15$ תצפיות:

$$P=P_{H_0}(V^+\geq 85)=1-P_{H_0}(V^+\leq 84)=.0844$$

(על פי מבחן הסימן לגבי אותם נתונים, קיבלנו בדוגמה 4.1 את המובהקות 0.3036.)

הערה: על פי מבחן ווילקוקסון קיבלנו כאן מובהקות קטנה יותר מזו שהתקבלה על סמך מבחן הסימן. פעמים רבות אמנם זו תהיה התוצאה, אולם אין זה מהווה כלל ודאי. ייתכנו נתונים שעבורם התוצאה על פי מבחן הסימן תהיה קיצונית יותר מזו המבוססת על מבחן ווילקוקסון. (נסו למצוא דוגמה פיקטיבית למקרה כזה!)

משפט 4.2. עבור מדגם בגודל n התפלגות הסטטיסטי V^+, תחת השערת האפס, היא סימטרית סביב $n(n+1)/4$.

הוכחה: תרגיל 6.

נסתכל, לדוגמה, כיצד ניתן למצוא מטבלה 4 את ההסתברות $P(V\geq 95)$, עבור $n=14$ תצפיות מזווגות. הערך $v=95$ אינו מופיע בטבלה. ניתן למצוא את ההסתברות הדרושה בשני אופנים:

א) נשתמש בסימטרייה של V^+. מרכז הסימטרייה הוא

$$n(n+1)/4=14(15)/4=52.5$$

הערך הסימטרי לתוצאה $95=52.5+42.5$ הוא $52.5-42.5=10$ ולכן מקבלים:

$$P(V^+\geq 95)=P(V^+\leq 10)=.0026$$

הערך 0.0026 מצוי בטבלה 4.

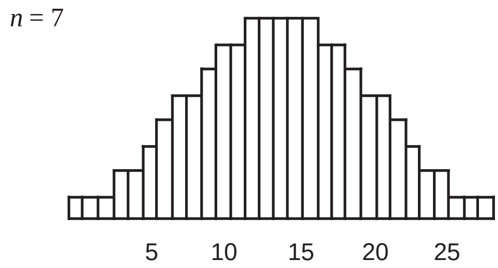
ב) על סמך סכום הדרגות השייכות להפרשים שליליים V^-. אם סכום הדרגות של ההפרשים החיוביים הוא $V^+ = 95$, סכום הדרגות של ההפרשים השליליים הוא $V^- = 10$. [עבור $n = 14$, $\frac{n(n+1)}{2} = 105 = 95 + 10$.] היות שתחת השערת האפס לסטטיסטי V^- אותה התפלגות כמו לסטטיסטי V^+, ניתן למצוא את המובהקות מתוך טבלה 4 על ידי

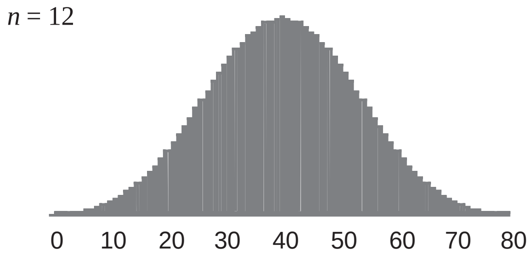
$$P(V^- \le 10) = P(V^+ \le 10) = .0026$$

באופן כללי, דחיית השערת האפס עבור ערכים גבוהים של V^+ אקוויוולנטית לדחייתה עבור ערכים נמוכים של V^-.

קירוב נורמלי עבור התפלגות הסטטיסטי V^+

גם הסטטיסטי V^+ של ווילקוקסון הוא בעל התפלגות הקרובה לנורמלית עבור n די גדול. עבור $n = 4$ רואים בציור 4.4 שההתפלגות כלל איננה קרובה לנורמלית. נביא כאן, בציור 4.5, את ההיסטוגרמות של התפלגות V^+ תחת השערת האפס, עבור 7 ועבור 12 תצפיות (מזווגות). אנו רואים שההתפלגות עבור 7 תצפיות עדיין אינה קרובה לנורמלית, אולם כבר עבור 12 תצפיות הקְרבה להתפלגות נורמלית ברורה. נרשום זאת במשפט הבא (שאותו לא נוכיח כאן).





ציור 4.5. *ההיסטוגרמות של התפלגויות הסטטיסטי V^+ עבור $n = 7, 12$*

משפט 4.3. תחת השערת האפס (חציון התפלגות ההפרש היא 0), התפלגות הסטטיסטי V^+ היא אסימפטוטית נורמלית. כלומר, המשתנה המתוקנן $\frac{V^+ - EV^+}{\sqrt{Var(V^+)}}$ הוא בעל התפלגות קרובה להתפלגות נורמלית סטנדרטית, כאשר n גדול. התוחלת EV^+ והשונות $Var(V^+)$ המצוינים לעיל מחושבים תחת המודל של H_0.

חישוב התוחלת והשונות של V^+ תחת השערת האפס

קל מאוד לחשב את התוחלת ואת השונות של V^+ תחת המודל של השערת האפס, בהתבסס על מסקנה 4.1.

משפט 4.4. תחת השערת האפס המומנטים של הסטטיסטי V^+ הם

$$EV^+ = \frac{n(n+1)}{4} \tag{18}$$

$$Var(V^+) = \frac{n(n+1)(2n+1)}{24} \tag{19}$$

הוכחה: לפי משפט 4.1 המשתנים המציינים מתפלגים לפי התפלגות ברנולי –

$T_j \sim B(1,1/2)$, ולכן קיים: $ET_j = 1/2$ וכן $Var(T_j) = 1/4$ לכל $j = 1,\ldots,n$.

מכאן מקבלים את התוחלת של הסכום כסכום התוחלות:

$$EV = E\sum_{j=1}^{n} j\cdot T_j = \sum_{j=1}^{n} j\cdot ET_j = \frac{1}{2}\sum_{j=1}^{n} j = \frac{1}{2}\left[\frac{n(n+1)}{2}\right] = \frac{n(n+1)}{4}$$

מובן שהתוחלת שהתקבלה שווה למרכז הסימטרייה של ההתפלגות (משפט 4.2).

נוסף על כך, n המשתנים המציינים הם בלתי-תלויים ולכן שונות הסכום שלהם שווה לסכום השונויות. מכאן מקבלים

$$Var(V^+) = Var\left(\sum_{j=1}^{n} j\cdot T_j\right) = \sum_{j=1}^{n} j^2\cdot Var(T_j) = \frac{1}{4}\sum_{j=1}^{n} j^2$$

$$= \frac{1}{4}\left[\frac{n(n+1)(2n+1)}{6}\right] = \frac{n(n+1)(2n+1)}{24}$$

(הערה: הנוסחה עבור סכום הריבועים של המספרים הטבעיים העוקבים היא ידועה וקל להוכיח אותה בעזרת אינדוקציה על מספר המחוברים.) ♣

דוגמה 4.8 (המשך דוגמה 4.7). נחשב את מובהקות התוצאה שהתקבלה בדוגמת נפגעי הלם קרב על סמך הקירוב הנורמלי. עבור $n = 15$ התוחלת והשונות של V^+ תחת

השערת האפס, מתקבלות מנוסחאות (18) ו-(19):

$$EV^+=\frac{15(16)}{4}=60.0 \qquad Var(V^+)=\frac{15(16)(31)}{24}=310.0$$

על ידי תקנון הערך של V^+ שהתקבל במדגם ($v=85$), נקבל את המובהקות בקירוב (עם שימוש בתיקון רציפות):

$$P=P_{H_0}(V^+\geq 85)\cong 1-\Phi\left(\frac{84.5-60}{\sqrt{310}}\right)=1-\Phi(1.39)=.0823$$

ההסתברות המדויקת שהתקבלה על סמך טבלה 4 היא $P=.0844$. הערכים קרובים מאוד.

4.5 בעיות של תיקו במבחן ווילקוקסון למדגם אחד

בעת שימוש במבחן ווילקוקסון, ערכי תיקו יכולים להיות משני סוגים: זוגות שבהם $X=Y$, כלומר, $D=0$ (ואז אי אפשר לקבוע אם זו תצפית עם הפרש חיובי או שלילי), וכן זוגות שונים שעבורם ההפרשים D זהים (ואז יש קושי בהגדרת הדרגות).
ראשית, כמו בבעיות של תיקו במבחן הסימן, פרק 4.2, מוציאים מן המדגם את כל התצפיות שעבורן התקבל $D=0$. גודל המדגם קטֵן בהתאמה. לשם הנוחיות נמשיך ונסמן אותו ב-n.
שנית, אם יש ערכי תיקו בין הערכים המוחלטים $|D_1|,\ldots,|D_n|$, נותנים דרגה שווה לערכים שווים – הדרגה הממוצעת המגיעה להם (כפי שנעשה במבחן ווילקוקסון לשני מדגמים, פרק 2.6). סטטיסטי המבחן הוא סכום הדרגות (הממוצעות) של הערכים המוחלטים השייכים לערכי D חיוביים. כלומר:

$$\tilde{V}=\sum_{i=1}^{n}\tilde{r}(|D_i|)$$

כאשר $\tilde{r}$ היא דרגה ממוצעת.
ברור שהתפלגות $\tilde{V}$ אינה זהה להתפלגות V^+, הניתנת בטבלה 4 בנספח. עם זאת, עבור n די גדול ניתן להשתמש בקירוב הנורמלי. קל להראות (בדומה להוכחת משפט 2.6) שהתוחלת של $\tilde{V}$ ניתנת על ידי

$$E\tilde{V}=EV^+=\frac{n(n+1)}{4} \qquad (20)$$

והשונות היא

$$Var(\tilde{V})=Var(V^+)-\frac{1}{48}\sum_j t_j(t_j^2-1) \qquad (21)$$

$$=\frac{n(n+1)(2n+1)}{24}-\frac{1}{48}\sum_j t_j(t_j^2-1)$$

כאשר t_j הוא גודל קבוצת התיקו ה-j.

דוגמה 4.9 (המשך דוגמה 4.3). אנו חוזרים על הנתונים שהבאנו בלוח 4.2 לגבי השימוש במִלות שאלה אצל פעוטות, כאשר הוספנו גם את דרגות הערכים המוחלטים של ההפרשים. אצל שניים מבין 17 הפעוטות שהשתתפו בניסוי, מספר הפעמים שאמרו "מה" היה זהה למספר הפעמים שהם אמרו "איפה". אחרי הוצאתם מהמדגם אנו נותרים עם $N=15$ תצפיות. הדירוג ניתן רק לערכים המוחלטים שאינם אפס (בִּדקו את נכונות הדירוג!).

לוח 4.6. *מספר הפעמים שפעוטות השתמשו במִלות שאלה*

הפעוט	מה (Y)	איפה (X)	$D=Y-X$	Z	$\tilde{r}(\lvert D\rvert)$
1	0	1	–1	0	2.5
2	1	4	–3	0	8.5
3	1	4	–3	0	8.5
4	3	4	–1	0	2.5
5	1	3	–2	0	5.5
6	1	0	1	1	2.5
7	0	5	–5	0	13.5
8	2	7	–5	0	13.5
9	3	6	–3	0	8.5
10	1	8	–7	–	–
11	0	0	0	–	–
12	9	7	2	1	5.5
13	2	1	1	1	2.5
14	0	3	–3	0	8.5
15	0	0	0	–	–
16	6	2	4	1	11.5
17	14	18	–4	0	11.5

סכום הדרגות הממוצעות של ההפרשים החיוביים נמצא:

$$\tilde{V}=2.5+5.5+2.5+11.5=22$$

נחשב את התוחלת ואת השונות של $\tilde{V}$, כאשר מספר התצפיות הרלבנטיות (שונות מ-0) הוא $n=15$.

$$E\tilde{V}=\frac{n(n+1)}{4}=\frac{15(16)}{4}=60$$

עבור השונות יש לחשב ראשית את הביטוי השני באגף ימין של (21). בין הדרגות בדוגמה, הדרגה 2.5 מופיעה ארבע פעמים, הדרגה 5.5 מופיעה פעמיים, הדרגה 8.5 מופיעה 4 פעמים, הדרגה 11.5 מופיעה פעמיים, והדרגה 13.5 מופיעה פעמיים. כלומר,

$$t_1=4,\ t_2=2,\ t_3=4,\ t_4=t_5=2$$

הביטוי הנדרש לתיקון השונות הוא, אפוא,

$$\frac{1}{48}\sum_j t_j(t_j^2-1)=\frac{1}{48}\Big[2\cdot 4(4^2-1)+3\cdot 2(2^2-1)\Big]=\frac{138}{48}=2.875$$

והשונות לפי (21) היא

$$Var(\tilde{V})=\frac{15(16)(31)}{24}-2.875=307.125$$

התוצאה שהתקבלה, $\tilde{V}=22$, קטנה מהתוחלת 60. לכן נחשב את הסתברות הזנב השמאלי. על פי הקירוב הנורמלי מקבלים

$$P(\tilde{V}\le 22)\cong\Phi\left(\frac{22.5-60}{\sqrt{307.125}}\right)=\Phi(-2.14)=.0162$$

היות שהמבחן הוא דו־צדדי, מובהקות התוצאה היא כפולה — $P=2(.0162)=.0324$.
עבור רמת מובהקות של 5% ניתן להסיק שיש הבדל מובהק בין מידת השימוש בין שתי מִלות השאלה הללו: במִלה "איפה" נוטים הפעוטות להשתמש יותר מאשר במִלה "מה".
נזכיר שעל פי מבחן הסימן, בדוגמה 4.3, התקבלה מובהקות של 0.1184, שהיא גדולה יותר מן המובהקות לפי מבחן ווילקוקסון. הסיבה היא שמבחן ווילקוקסון רגיש להפרשים גדולים בין שתי התצפיות, בעוד שהסטטיסטי של מבחן הסימן מתחשב רק בסימנם של ההפרשים ולא בגודלם.

4.6 רווח בר־סמך לפרמטר מיקום במדגם מזווג על סמך הסטטיסטי של מבחן הסימן

נניח שבידינו מדגם מזווג בגודל n ומדגם ההפרשים המתאימים $D_1,\ldots,D_n$. אנו מניחים $D_i\sim F_D$, משתנים בלתי תלויים, כאשר F_D התפלגות רציפה.
נגדיר את הפרמטר Δ כחציון התפלגות ההפרשים:

$$\Delta=med(F_D)=d_{.5} \qquad (22)$$

הערה: במונחים של המודל הזה, ההשערות בנוסחה (5), ניתנות להירשם כך:

$$H_0: \Delta = 0 \qquad\qquad H_1: \Delta > 0$$

אמידת הפרמטר Δ

כיוון ש-Δ הוא חציון התפלגות המשתנה D, ניתן לאמוד את Δ על ידי חציון המדגם

$$\hat{\Delta} = med(D_1,\ldots,D_n) \qquad (23)$$

רווח בר־סמך עבור Δ

יש למצוא רווח בר־סמך עבור Δ ברמת ביטחון $1-\alpha$, כלומר, רווח המבוסס על תוצאות המדגם, באופן שההסתברות שהוא יכלול את הפרמטר Δ תהיה $1-\alpha$ (הגדרה 2.6). נעשה זאת בעזרת סטטיסטי הסדר של ההפרשים $D_{(1)} < D_{(2)} < \cdots < D_{(n)}$, כלומר רשימת הערכים שהתקבלו, כשהם מסודרים לפי גודלם. (שימו לב שהרשימה כאן היא של ההפרשים עצמם ולא של הערכים המוחלטים שלהם.)

נרצה למצוא שתי דרגות, r ו-s, באופן ש-$r < s$, שיקיימו:

$$P\{D_{(r)} \le \Delta \le D_{(s)}\} \ge 1-\alpha \qquad (24)$$

כלומר, ההסתברות שהרווח בין ההפרש ה-r בגודלו לבין ההפרש ה-s בגודלו אמנם יכלול את הפרמטר Δ, תהיה שווה לרמת הביטחון (או רמת הסמך) $1-\alpha$ הנדרשת. זהו רווח בר־סמך עבור Δ ברמת ביטחון (או רמת סמך) $1-\alpha$.

ההסתברות הרשומה ב-(24) מחושבת תחת ההנחה ש-Δ הוא חציון ההתפלגות F_D.

את ההסתברות (24) ניתן לרשום:

$$P\{D_{(r)} \le \Delta \le D_{(s)}\} = P\{D_{(r)} \le \Delta\} - P\{D_{(s)} \le \Delta\} \qquad (25)$$

(ניתן לראות את הפירוט של הפירוק הזה בהוכחה של משפט 2.8.)

נרשום עתה נוסחה כללית להסתברות $P\{D_{(k)} \le \Delta\}$ עבור k כלשהו ($1 \le k \le n$).

$$P\{D_{(k)} \le \Delta\} = P\{\text{לפחות } k \text{ מבין } D_1,\ldots,D_n \text{ קטנים או שווים ל- } \Delta\} \qquad (26)$$

$$= \sum_{i=k}^{n} \binom{n}{i} \left[\frac{1}{2}\right]^n$$

השוויון האחרון נובע מהעובדה ש-n ההפרשים $D_1,\ldots,D_n$ הם בלתי תלויים וכן, היות ש-Δ הוא חציון ההתפלגות של ההפרשים הללו, עבור כל i, קיים: $P(D_i \le \Delta) = 1/2$.

סכום ההסתברויות ב-(26) הוא סכום הסתברויות בינומיות $B(n,1/2)$ וזהו בדיוק סכום ההסתברויות המצטבר של הזנב הימני של התפלגות המשתנה S_n, תחת השערת האפס.

קיבלנו, אפוא, את ההסתברות הדרושה:

$$P\{D_{(k)} \le \Delta\} = P(S_n \ge k) \qquad (27)$$

במִלים אחרות, ההסתברות לכך שההפרש ה-k בגודלו, מבין n הפרשים, לא יעלה על

חציון ההתפלגות – Δ, שווה להסתברות שמשתנה בינומי $B(n,1/2)$ יקבל ערך גדול או שווה ל- k.

נציב את התוצאה (27) ב-(25) ונקבל את נוסחת הסתברות הרווח:

$$P(D_{(r)} \le \Delta \le D_{(s)}) = P(D_{(r)} \le \Delta) - P(D_{(s)} \le \Delta)$$

$$= P(S_n \ge r) - P(S_n \ge s) = P(r \le S_n < s)$$

$$= \sum_{i=r}^{s-1} \binom{n}{i} \left[\frac{1}{2}\right]^n$$

קיבלנו, אפוא, את הנוסחה עבור הסתברות הרווח:

$$P(D_{(r)} \le \Delta \le D_{(s)}) = \sum_{i=r}^{s-1} \binom{n}{i} \left[\frac{1}{2}\right]^n \qquad (28)$$

ההתפלגות הבינומית לעיל, שהיא התפלגות הסטטיסטי של מבחן הסימן S_n, נתונה בטבלה 3 בנספח.

לפי נוסחה (28), כדי לקבל רווח בר־סמך ברמת ביטחון $1-\alpha$, יש לבחור r ו-s שיקיימו את הדרישה:

$$\sum_{i=r}^{s-1} \binom{n}{i} \left[\frac{1}{2}\right]^n = P(r \le S_n \le s-1) \ge 1-\alpha \qquad (29)$$

בהתאם לדרגות r ו-s שנבחרו, הרווח ברמת סמך $1-\alpha$ עבור Δ הוא $[D_{(r)}, D_{(s)}]$.

הערה: כדי שהרווח יהיה מוגדר לכל ערכי r ו-s האפשריים (כולל הקצוות – 0 או n), נגדיר את הערכים הקיצוניים של סטטיסטי הסדר על ידי $D_{(0)} = -\infty$ ו- $D_{(n+1)} = \infty$.

ברור שניתן לבחור ערכים שונים של r ו-s שיקיימו את הדרישה (29). כדי שהרווח יהיה קצר ככל האפשר, רצוי לבחור ערכים סביב מרכז ההתפלגות הבינומית.

דוגמה 4.10. נניח שבידינו מדגם של 10 זוגות של תצפיות ואנו רוצים לתת רווח בר־סמך ל-Δ ברמת סמך של $1-\alpha = 0.90$. להלן (לוח 4.7) חלק מהטבלה של התפלגות S_n עבור $n = 10$ (מתוך טבלה 3 בנספח).

לוח 4.7. *ההתפלגות המצטברת של S_n עבור $n = 10$*

k	0	1	2	3	4	5	6	7	8	9	10
$P(S_n \le k)$	.0010	.0107	.0547	.1719	.3770	.6230	.8281	.9453	.9893	.9990	1.0000

ישנן אפשרויות אחדות לבחור ערכים $r < s$ שיקיימו את הדרישה (29) עבור $1-\alpha = .90$.

א) $P(0\le S_n\le 7)=.9453$. ולכן $r=0,\ s=8$
הרווח המתקבל הוא $[D_{(r)},D_{(s)}]=[-\infty,D_{(8)}]$

ב) $P(1\le S_n\le 7)=.9453-.0010=.9443$. כאן $r=1,\ s=8$.
הרווח המתקבל הוא $[D_{(r)},D_{(s)}]=[D_{(1)},D_{(8)}]$

ג) $P(2\le S_n\le 7)=.9453-.0107=.9346$. כאן $r=2,\ s=8$.
הרווח המתקבל הוא $[D_{(r)},D_{(s)}]=[D_{(2)},D_{(8)}]$

ד) $P(3\le S_n\le 8)=.9893-.0547=.9346$. כאן $r=3,\ s=9$.
הרווח המתקבל הוא $[D_{(r)},D_{(s)}]=[D_{(3)},D_{(9)}]$

באופן דומה ניתן לקבל רווחי־סמך המבוססים על הדרגות הגבוהות ביותר.
ברור שהרווח ג עדיף על הרווחים א ו-ב, אולם אי אפשר להשוות את הרווחים שיתקבלו לפי ג או ד בלי ידיעת תוצאות הניסוי.

קביעת רווח בר־סמך כללי

כדי לא להזדקק לכל החישובים כפי שעשינו בדוגמה **4.10**, ניתן לקבוע כלל די טוב לבחירת הדרגות r ו-s, כפי שהבאנו בפרק 2.8. ההתפלגות הבינומית $B(n,1/2)$ היא סימטרית סביב $n/2$, וצורתה דומה לצורת פעמון, באופן שההסתברויות במרכז הן הגבוהות ביותר. לכן נבחר את התחום המרכזי של ההתפלגות, שהסתברותו $1-\alpha$ לפחות. במִלים אחרות, על פי טבלת S_n נבחר את הערך המקסימלי c שעבורו

$$P(S_n\le c)\le\frac{\alpha}{2}\qquad(30)$$

מטעמי סימטרייה, $P(S_n\ge n-c)\le\frac{\alpha}{2}$ ולכן $P(c+1\le S_n<n-c)\ge 1-\alpha$.
בהתאם לכך הדרגות הדרושות לפי נוסחה (29) הן

$$r=c+1;\ s=n-c\qquad(31)$$

דוגמה 4.11 (*המשך דוגמה 4.10*). רמת הסמך הדרושה היא $1-\alpha=0.90$. לפי הכלל (30), נראה בלוח 4.7 שהערך הגבוה ביותר שעבורו ההתפלגות המצטברת אינה עולה על $\alpha/2=.05$ הוא $c=1$, שעבורו מתקיים: $P(S_{10}\le 1)=.010$. לפי (31) יש לקבוע, אפוא, $r=2,\ s=10-1=9$. הרווח המתקבל הוא: $[D_{(2)},D_{(9)}]$. זה רווח סימטרי מבחינת הדרגות – "מסלקים" את הערך הנמוך ביותר ואת הערך הגבוה ביותר, ומקבלים את הרווח בין הערך השני לתשיעי. רמת הסמך המדויקת המתקבלת היא

$$P(D_{(2)}\le\Delta\le D_{(9)})=P(2\le S_n<9)=1-P(S_n\le 1)-P(S_n\ge 9)$$

$$=1-2(0.0107)=0.9786$$

הערה: לפי החישוב בדוגמה **4.10** ברור שלפעמים ניתן לקבל רווח יותר טוב מהרווח הסימטרי, בגלל אי־רציפות ההתפלגות הבינומית. גם הרווח ב-ג וגם זה שב-ד יותר טובים מהרווח הסימטרי שקיבלנו פה. אלה רווחים צרים יותר, אשר מקיימים שניהם את הדרישה (29). עם זאת, רמת הסמך של הרווח הסימטרי כאן גבוהה, כמובן, יותר מרמות הסמך ב-ג וב-ד. במקרים של מדגמים גדולים יש קושי לחשב את הרווחים המדויקים ולכן נהוג להשתמש ברווח סימטרי.

שימוש בקירוב נורמלי

אם המדגם גדול מספיק, באופן שניתן להשתמש בקירוב הנורמלי עבור התפלגות S_n, ניתן לקבל את הדרגות r ו-s בעזרת הקירוב הנורמלי. לפי נוסחה (30), דרוש שהערך c יקיים:

$$P(S_n \le c) \cong \Phi\left(\frac{c+1/2-n/2}{\sqrt{n/4}}\right) \le \frac{\alpha}{2} \qquad (32)$$

מכאן דרוש שיתקיים $\frac{c+1/2-n/2}{\sqrt{n/4}} \le z_{\alpha/2} = -z_{1-\alpha/2}$ והערך c צריך לקיים:

$$c \le \frac{n-1}{2} - z_{1-\alpha/2}\sqrt{\frac{n}{4}} \qquad (33)$$

דוגמה 4.12. נתונים כאן ההפרשים בין ציונים פסיכומטריים לציוני בגרות (מתואמים) של 14 סטודנטים למתמטיקה:

50, 128, 10, 96, 136, –22, 46, –16, 42, 74, 12, –4, 54, 162

נרצה למצוא רווח בר־סמך לחציון Δ של התפלגות הפער בין הציון הפסיכומטרי לציון בגרות.

נשתמש בקירוב הנורמלי כדי למצוא את הערך c עבור רמת סמך $1-\alpha=.90$.

נציב את ערך החלוקה $z_{1-\alpha/2}=z_{.95}=1.645$ וגודל המדגם $n=14$ בנוסחה (33) ונקבל:

$$c \le \frac{14-1}{2} - 1.645\sqrt{\frac{14}{4}} = 3.4$$

לכן יש לבחור $c=3$.

לפי נוסחה (31), הדרגות המתאימות הן: $r=c+1=4;\ s=n-c=14-3=11$.

רווח בר־סמך ברמת סמך 90% עבור Δ (הפער בין התפלגות הציון הפסיכומטרי לציון בגרות) הוא $[D_{(r)},D_{(s)}]=[D_{(4)},D_{(11)}]$.

כדי למצוא את הערכים המתאימים עבור נתוני המדגם, יש לסדר אותם לפי הגודל:

$$-22, -16, -4, 10, 12, 42, 46, 50, 54, 74, 96, 128, 136, 162$$

שימו לב שתצפיות ההפרשים מסודרות כפי שהן במקור (ולא לפי הערך המוחלט!).

לפי הרשימה הזאת, $D_{(4)}=10$ ו-$D_{(11)}=96$ ולפיכך הרווח עבור Δ הוא $[10, 96]$.

("מסלקים" 3 תצפיות בכל אחד מהזנבות, וכך הרווח המתקבל הוא בין התצפית הרביעית מצד שמאל של הרשימה לבין התצפית הרביעית מצד ימין.)

משמעות הרווח שקיבלנו היא שכנראה הפער בין שני הציונים הוא בין 10 נקודות ל-96 נקודות. נשים לב שהערך 0 אינו כלול ברווח, ולכן ניתן לומר (ברמת ביטחון של 90%) שאצל תלמידי מתמטיקה הציונים הפסיכומטריים נוטים להיות גבוהים מציוני הבגרות.

שימוש בטבלת ההתפלגות המדויקת של S_n עבור $n=14$ נותן את הערך המקסימלי שעבורו ההתפלגות המצטברת אינה עולה על 0.05: $P(S_n\le 3)=.0287$. [הערך הבא אחריו כבר גדול מדי: $P(S_n\le 4)=.0898$.] כלומר, גם לפי ההתפלגות המדויקת יש לבחור $c=3$, אותה תוצאה כמו זו שהתקבלה על ידי שימוש בקירוב הנורמלי.

האומדן המתקבל מנתונים אלה לגבי הפער בין הציון הפסיכומטרי לבין ציון הבגרות של סטודנטים למתמטיקה הוא חציון מדגם ההפרשים, לפי נוסחה (23), כלומר:

$$\hat{\Delta}=\frac{D_{(7)}+D_{(8)}}{2}=\frac{46+50}{2}=48$$

4.7 רווח בר־סמך לפרמטר מיקום במדגם מזווג על סמך הסטטיסטי של ווילקוקסון

נרשום מודל הזזה המתאים לשימוש בהתפלגות של הסטטיסטי V^+.

נניח שהמשתנים $D_1, D_2, \ldots, D_n$ הם בלתי תלויים בעלי התפלגות F_D **רציפה וסימטרית**.

מודל ההזזה מוגדר על ידי

$$F_D(t)=F_0(t-\Delta) \qquad (34)$$

תחת המודל (34) התפלגות ההפרשים היא הזזה מסוימת של התפלגות סימטרית סביב אפס.

לפי מודל ההזזה (34), F_0 סימטרית סביב 0, ולכן Δ הוא בדיוק חציון ההתפלגות F_D:

$$F_D(\Delta)=F_0(0)=\frac{1}{2}$$

לפיכך פרמטר ההזזה Δ הוא גם הפרמטר שהוגדר ב-(22).

נגדיר עתה משתנים חדשים:

$$W_{ij}=\frac{1}{2}(D_i+D_j) \qquad i\le j \qquad (35)$$

ערכי W הם כל הממוצעים של שתי תצפיות שונות $(i<j)$, וכן הממוצעים של כל תצפית עם עצמה $(i=j)$, שהם למעשה התצפיות הבודדות. נסמן את מספר המשתנים הללו על ידי M:

$$M=\binom{n}{2}+n=\frac{n(n+1)}{2} \tag{36}$$

אמידת Δ

ברור שחציון כל אחד מהמשתנים $W_{ii}=D_i$ הוא בדיוק Δ. כמו כן מטעמי סימטרייה, עבור $i<j$, למשתנה W_{ij} – שהוא ממוצע שתי תצפיות בלתי תלויות מאותה התפלגות, כל אחת עם חציון Δ – גם כן התפלגות עם חציון Δ. לכן ניתן לאמוד את Δ ידי החציון של כל ערכי W במדגם:

$$\hat{\Delta}_H=med(W_1,\ldots,W_M) \tag{37}$$

אומד זה נקרא אומד הודג'ס-להמן.

הערה: בסעיף 4.6 הצענו, נוסחה (23), כאומד ל-Δ את $\hat{\Delta}$ – חציון התצפיות הבודדות D. האומד המוצע כאן הוא חציון **הממוצעים** בין כל שתי תצפיות כאלה.

הקשר בין הסטטיסטי V^+ לבין ערכי הממוצעים W

הסטטיסטי של ווילקוקסון למדגם אחד למעשה מבוסס על המשתנים W_{ij} המוגדרים בנוסחה (35), לפי המשפט הבא.

משפט 4.5. הסטטיסטי של ווילקוקסון למדגם אחד שווה למספר ה-W_{ij}-החיובים. כלומר:

$$V^+=\#\{i\le j: W_{ij}>0\} \tag{38}$$

הוכחה: נזכיר ש-V^+ הוא סכום הדרגות של הערכים המוחלטים השייכים להפרשים חיוביים. כפי שהוגדר בנוסחה (16), $V^+=\sum_{j=1}^{n} jT_j$, כאשר T_j הוא המשתנה המציין של המאורע {הדרגה j שייכת ל-D חיובי}.

לשם נוחיות הכתיבה, נניח שהתצפיות D מסודרות לפי דרגת הערך המוחלט שלהן. זאת אומרת, דרגת $|D_j|$ היא j, עבור $j=1,\ldots,n$.

את הספירה של הממוצעים W_{ij} החיוביים נעשה בצורה הבאה. נסמן את המשתנים המציינים:

$$h_{ij}=\begin{cases}1 & W_{ij}>0\\ 0 & W_{ij}<0\end{cases}$$

וכך ניתן לרשום

$$\#\{i \le j : W_{ij} > 0\} = \sum_{j=1}^{n} \sum_{i=1}^{j} h_{ij} = \sum_{j=1}^{n} H_j \qquad (39)$$

כאשר H_j הוא מספר ה- W_{ij} החיוביים עבור j קבוע: $H_j = \sum_{i=1}^{j} h_{ij}$.

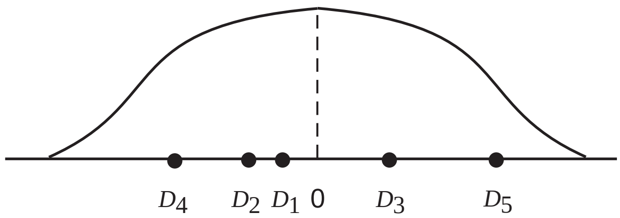
לכן נספור ראשית את מספר ה- W_{ij} החיוביים עבור כל j בנפרד, כאשר $i \le j$, ולאחר מכן נסכם ערכים אלה עבור כל ערכי j. נשים לב שהממוצע $W_{ij} = (D_i + D_j)/2$ הוא חיובי רק אם הסכום $2W_{ij} = D_i + D_j$ הוא חיובי.

נסתכל ראשית על $j = 1$. עלינו לברר מהו מספר הממוצעים W_{i1} שהם חיוביים. יש רק ערך אחד i המקיים $i \le 1$ והוא $i = 1$. הערך המתאים של הממוצע הוא D_1 ולכן מספר ה- W_{i1} החיוביים הוא 1 אם $D_1 > 0$ והוא 0 אם $D_1 < 0$. קיבלנו, אפוא:

$$H_1 = h_{11} = \begin{cases} 1 & D_1 > 0 \\ 0 & D_1 < 0 \end{cases}$$

ניקח עתה $j = 2$. עלינו לברר מהו מספר הממוצעים W_{i2} שהם חיוביים, כאשר $i \le 2$. במקרה זה יש שני ערכי i מתאימים והממוצעים המתאימים הם $W_{12} = \frac{1}{2}(D_1 + D_2)$ ו- $W_{22} = D_2$. ברור ש- $W_{22} > 0$ רק אם $D_2 > 0$. כמו כן, היות שהנחנו שערכי D מסודרים כבר לפי ערכם המוחלט, קיים $|D_1| < |D_2|$ ולכן סכום שני המשתנים הללו הוא חיובי רק אם הגדול יותר בערכו המוחלט הוא חיובי, כלומר, אם $D_2 > 0$ (ראו ציור **4.6**). בסך הכול מספר הממוצעים W_{i2} שהם חיוביים הוא 2 אם $D_2 > 0$ והוא 0 אם $D_2 < 0$. כלומר,

$$H_2 = \sum_{i=1}^{2} h_{i2} = \begin{cases} 2 & D_2 > 0 \\ 0 & D_2 < 0 \end{cases}$$



ציור 4.6. *מדגם של 5 תצפיות*

נסתכל עתה באופן כללי על דרגה ספציפית כלשהי j ונראה מהו מספר ה- W_{ij} החיוביים עבור $i \le j$. בגלל הסדר שקבענו בין התצפיות, אם $i < j$ אזי $|D_i| < |D_j|$. נפריד את הדיון לשתי אפשרויות.

א) אם $D_j > 0$, אזי לכל i הקטן או שווה ל-j מתקיים $2W_{ij} = D_i + D_j > 0$ (כל המשתנים D_i קטנים בערכם המוחלט מ- D_j), ולכן בדיוק j מבין הממוצעים W_{ij}

הם חיוביים. ניתן לראות זאת בציור 4.6.

ב) אם $D_j<0$ אזי כל אחד מבין הסכומים $2W_{ij}=D_i+D_j$ הוא שלילי עבור $i\le j$.

נסתכל על הנתונים בציור 4.6 כדוגמה. עבור $j=2$, למשל, $D_2<0$, ומכאן $W_{22}<0$, כמובן, וכן גם הסכום $2W_{12}=D_1+D_2<0$. לכן במקרה זה אין שום ערך חיובי מבין W_{i2} עבור $i\le 2$.

עבור $j=3$, $D_3>0$, ולכן $W_{33}>0$. כמו כן D_3 גדול בערכו המוחלט גם מ-D_1 וגם מ-D_2 (שניהם שליליים), ולכן הסכומים $2W_{13}=D_1+D_3$ וגם $2W_{23}=D_2+D_3$ שניהם חיוביים. בסך הכול ישנם 3 ערכי W_{i3} חיוביים, עבור $i\le 3$.

מאפשרויות א ו-ב לעיל נובע בסך הכול, לפי נוסחה (39), שעבור כל j ספציפי מתקיים:

$$H_j=\sum_{i=1}^{j}h_{ij}=\begin{cases} j & D_j>0 \\ 0 & D_j<0 \end{cases}=jT_j$$

זה נכון לכל j ולכן הסכום עבור כל ערכי j מתקבל:

$$\#\{i\le j: W_{ij}>0\}=\sum_{j=1}^{n}H_j=\sum_{j=1}^{n}jT_j=V^+$$

בכך התקבלה טענת המשפט. ♣

רווח בר־סמך עבור Δ

כדי לקבל רווח בר־סמך ל-Δ נזדקק למשפט הבא, הקושר את התפלגות סטטיסטי הסדר של W_{ij} להתפלגות ווילקוקסון.

משפט 4.6. יהיו $D_1,D_2,\ldots,D_n$ משתנים בלתי-תלויים בעלי התפלגות רציפה וסימטרית F_D, כאשר F_D מקיימת את תנאי (34).

אזי קיים:

$$P(V^+\le k)=P_\Delta\{W_{(M-k)}\le\Delta\} \qquad \text{עבור } k=0,\ldots,M \tag{40}$$

או בצורה אחרת:

$$P_\Delta\{W_{(r)}\le\Delta\}=P(V^+\ge r) \qquad \text{עבור } r=0,\ldots,M \tag{41}$$

כאשר הסתברויות המשתנה V^+ הן אלה המתקבלות תחת השערת האפס (טבלה 4 בנספח).

הוכחה: נגדיר משתנים חדשים $D_i'=D_i+\Delta$, $i=1,\ldots,n$. ההתפלגות של D_i' היא

$$P(D_i'\le t)=P(D_i-\Delta\le t)=P(D_i\le t+\Delta)=F_D(t+\Delta)=F_0(t)$$

כלומר, המשתנים החדשים $D_1',D_2',\ldots,D_n'$ הם בלתי-תלויים בעלי התפלגות F_0.

הראינו במשפט 4.5, נוסחה (38), ש-V^+ הוא למעשה מספר ה-W_{ij} החיוביים. את

התפלגות הסטטיסטי V^+ (תחת המודל של השערת האפס שבו $\Delta = 0$) ניתן לרשום, אפוא, על ידי מספר הממוצעים החיוביים של זוגות D'_i, D'_j, עבור $i \le j$. כלומר, נרשום את המשתנה V^+, שהתפלגותו נתונה בנספח 4, על ידי

$$V^+ = \#\left\{ i \le j : \frac{D'_i + D'_j}{2} > 0 \right\}$$

$$= \#\left\{ i \le j : \frac{(D_i - \Delta) + (D_j - \Delta)}{2} > 0 \right\} = \#\left\{ i \le j : W_{ij} > \Delta \right\}$$

מכאן השקילות בין המאורעות: $V^+ \le k$ אם ורק אם $\#\left\{ i \le j : W_{ij} > \Delta \right\} \le k$.

המאורע האחרון שקול למאורע $\#\left\{ i \le j : W_{ij} \le \Delta \right\} \ge M - k$ [יש בסך הכול M ערכים ל- W_{ij}, עבור $i \ge j$, כאשר M נתון בנוסחה (36)]. משמעות המאורע האחרון היא שיש לפחות $M-k$ ערכי W_{ij} שאינם עולים על Δ. לכן קיים: $W_{(M-k)} \le \Delta$. קיבלנו, אפוא, את השקילות של שני המאורעות: $\{V^+ \le k\} = \{W_{(M-k)} \le \Delta\}$.

לכן הסתברויותיהם שווה: $P_{H_0}(V^+ \le k) = P_\Delta(W_{(M-k)} \le \Delta)$

בכך הוכחנו את המשפט, נוסחה (40).

נוסחה זו ניתן לרשום בצורה אחרת על ידי הצבת $r = M - k$:

$$P_\Delta\{W_{(r)} \le \Delta\} = P_{H_0}(V^+ \le M - r) = P_{H_0}(V^+ \ge r)$$

השוויון האחרון נובע מהסימטרייה של התפלגות V^+.

בכך הוכחנו את הנוסחה (41). ♣

בניית רווח בר־סמך עבור Δ

בדומה למקרים הקודמים, נמצא רווח בר־סמך עבור פרמטר ההזזה Δ בהתבסס על סטטיסטי הסדר של ערכי W.

נרצה לבחור שתי דרגות $r < s$, באופן שיתקיים:

$$P_\Delta\{W_{(r)} \le \Delta \le W_{(s)}\} \ge 1 - \alpha \qquad (42)$$

נרשום את ההסתברות (42) כהפרש בין שתי הסתברויות, כמו בנוסחה (25):

$$P_\Delta\{W_{(r)} \le \Delta \le W_{(s)}\} = P_\Delta\{W_{(r)} \le \Delta\} - P_\Delta\{W_{(s)} \le \Delta\}$$

לפי נוסחה (41) נקבל:

$$P_\Delta\{W_{(r)} \le \Delta \le W_{(s)}\} = P(V^+ \ge r) - P(V^+ \ge s) = P(r \le V^+ \le s - 1) \qquad (43)$$

בדומה לפתרונות הקודמים של רווחי-סמך, נבחר רווח סימטרי (התפלגות V^+ היא סימטרית) באופן הבא.

יהי c הערך המקסימלי שעבורו קיים:

$$(44)\qquad P(V^+ \le c) \le \frac{\alpha}{2}$$

לפי (43) הדרגות המתאימות הן

$$(45)\qquad r = c+1;\ s = M-c$$

רווח בר־סמך ברמת סמך $1-\alpha$ עבור Δ מתקבל על ידי

$$(46)\qquad [D_{(r)}, D_{(s)}] = [D_{(c+1)}, D_{(M-c)}]$$

הערה: כמו במקרים הקודמים, הרווח (46) לעיל הוא סימטרי מבחינת הדרגות. "מסלקים" את c הדרגות הנמוכות ביותר ואת c הדרגות הגבוהות ביותר מבין M ערכי W.

דוגמה 4.13 (המשך דוגמה 4.12). נתוני ההפרשים בין ציונים פסיכומטריים וציוני בגרות מדוגמה 4.13 רשומים בלוח 4.8 לפי הסדר, בשורה העליונה ובעמודה הראשונה משמאל.

לוח 4.8. *ערכי W_{ij} של נתוני ההפרשים בין ציונים פסיכומטריים לציוני בגרות ($n = 14$)*

	d_j													
d_i	–22	–16	–4	10	12	42	46	50	54	74	96	128	136	162
–22	–22	–19	–13	–6	–5	10	12	14	16	26	37	53	57	70
–16		–16	–10	–3	–2	13	15	17	19	29	40	56	60	73
–4			–4	3	4	19	21	**23**	25	35	46	62	66	79
10				10	11	26	28	30	32	42	53	69	73	**86**
12					12	27	29	31	33	43	54	70	74	87
42						42	44	46	48	58	69	85	89	102
46							46	48	50	60	71	87	91	104
50								50	52	62	73	89	93	106
54									54	64	75	91	95	108
74										74	85	101	105	118
96											96	112	116	129
128												128	132	145
136													136	149
162														162

בַּלוח נתונים כל הממוצעים של שני הפרשים במדגם, $w_{ij} = (d_i + d_j)/2$, עבור $i \le j$.
מספר כל ממוצעי זוגות ההפרשים הוא $105 = \frac{14(15)}{2} = \frac{n(n+1)}{2} = M$. באלכסון מופיעים ההפרשים עצמם. נרצה לקבל רווח סמך 90% עבור Δ.
לפי טבלת התפלגות V^+ (טבלה 4 בנספח), $P(V^+ \le 25) = .0453$. לכן יש לבחור את הערך הנמוך $c = 25$. בלוח 4.8, הגבולות של 25 הערכים הנמוכים ביותר (בפינה השמאלית העליונה) ו־25 הערכים הגבוהים ביותר (בפינה הימנית התחתונה) מסומנים

במסגרת, והערכים הבאים מיד אחריהם (ה-26 וה-80) מודגשים באות עבה. הרווח עבור החציון Δ הוא

$$[W_{(c+1)},W_{(105-c)}]=[W_{(26)},W_{(80)}]=[23,86]$$

כשהשתמשנו ברווח המבוסס על ההתפלגות של S_n קיבלנו את הרווח

$$[D_{(4)},D_{(11)}]=[10,96]$$

ההשוואה מראה שהשימוש ב- V^+ נתן לנו רווח קצר יותר.

מלוח 4.8 ניתן גם למצוא את אומדן הודג׳ס ולהמן עבור Δ על ידי החציון של ערכי W במדגם – $\hat{\Delta}_H=W_{(53)}=53$ (זו איננה טעות. סִפרו ובִדקו את התוצאה!). ערך זה מסומן בקו תחתי. נזכיר שהאומדן שהתקבל על סמך חחתציון ההפרררשים עצמם היה $\hat{\Delta}=48$ (דוגמה 4.12).

4.8 עוצמת מבחן הסימן

חישוב העוצמה של מבחן הסימן הוא פשוט, מכיוון שסטטיסטי המבחן הוא בעל התפלגות בינומית: $S_n=\sum_{i=1}^{n}Z_i$, כאשר $Z_i=\begin{cases}1 & X_i<Y_i\\0 & X_i>Y_i\end{cases}=\begin{cases}1 & D_i>0\\0 & D_i<0\end{cases}$.

המשתנים $Z_1,Z_2,\ldots,Z_n$ הם בלתי תלויים ושווי התפלגות, ולכן סכומם הוא משתנה בינומי – $S_n\sim B(n,p)$, כאשר p מוגדר על ידי $p=P(D_i>0)$.

לחישוב העוצמה נניח שהתפלגות ההפרש D היא רציפה.

נזכיר שבמקרה זה ההשערות הנבדקות הן אלה המובאות ב-(5):

$$H_0: p=1/2 \qquad\qquad H_1: p>1/2$$

עוצמת המבחן

העוצמה היא ההסתברות לדחיית השערת האפס תחת הערך האלטרנטיבי. עבור מבחן הסימן לבעיה החד־צדדית, אזור הדחייה עבור רמת מובהקות α כולל את הערכים הגבוהים של S_n, שהסתברותם כאשר $p=1/2$ היא α. ראינו קודם כיצד מוצאים אזור דחייה מתאים עבור כל גודל מדגם נתון. בהנחה שהערך הקריטי הוא c, כלומר, דוחים את השערת האפס כאשר $S_n\geq c$, עוצמת המבחן עבור ערך ספציפי p היא

$$\pi(p)=P(S_n\geq c)=\sum_{k=c}^{n}\binom{n}{k}p^k(1-p)^{n-k} \tag{47}$$

כאשר המדגם קטן ניתן לחשב את העוצמה בקלות.

דוגמה 4.14. מוציאים מדגם של 30 תצפיות מזווגות לבדיקת השערה חד־צדדית (5). לפי

טבלה 3, הערך הקריטי עבור רמת מובהקות של 5% הוא $c=20$:

$$P_{H_0}(S_{18}\ge 20)=P_{H_0}(S_{18}\le 10)=.0494$$

נחשב את עוצמת המבחן אם למעשה $p=0.6$.

$$\pi(0.6)=P_{p=.6}(S_{30}\ge 20)=\sum_{k=20}^{30}\binom{30}{k}0.6^k\cdot 0.4^{30-k}=.2915$$

עבור מדגמים גדולים החישוב המדויק קשה. אולם אז ניתן להשתמש בקירוב הנורמלי עבור ההתפלגות הבינומית, בתנאי שמתקיים $np\ge 5$ וגם $n(1-p)\ge 5$.

נחשב את העוצמה המקורבת בַדוגמה. באופן כללי, המומנטים של S_n הם

$$ES_n=np \qquad Var(S_n)=np(1-p)$$

כאשר $n=30$ ו- $p=0.6$ מקבלים:

$$ES_{18}=30(0.6)=18 \qquad Var(S_{18})=30(0.6)(0.4)=7.2$$

העוצמה המקורבת:

$$\pi(0.6)=P_{p=.6}(S_{30}\ge 20)\cong 1-\Phi\left(\frac{19.5-18}{\sqrt{7.2}}\right)$$
$$=1-\Phi(0.56)=.2877$$

הקירוב טוב.

עוצמה מקורבת של מבחן הסימן

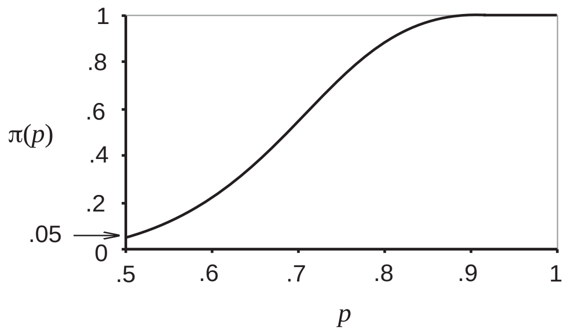
מדוגמה 4.14 ברור שהעוצמה של מבחן הסימן היא פונקציה של הערך p. נרשום צורה כללית של העוצמה המקורבת, ובעזרתה נוכל גם לחשב את גודל המדגם הדרוש להשגת עוצמה גדולה כרצוננו. נעיר כאן, כפי שהערנו בפרק 2, שלמעשה הערך של העוצמה המדויקת אינו כה חשוב, כיוון שבדרך כלל הבעיה שלשמה דרוש חישוב העוצמה היא למציאת גודל מדגם מתאים. כך, גם אם החישוב הוא מקורב, נקבל סדר גודל נכון של המדגם הדרוש.

על פי הקירוב הנורמלי, מקבלים את פונקציית העוצמה של מבחן הסימן:

$$\pi(p)\cong\Phi\left(\frac{\sqrt{n}(p-1/2)}{\sqrt{p(1-p)}}-z_{1-\alpha}\frac{1/2}{\sqrt{p(1-p)}}\right) \qquad (49)$$

(הוכיחו בעצמכם, תרגיל 5 א.)

הערה: נסתכל על ערכי הפרמטר הרלוונטיים, בתחום $1/2 \le p \le 1$. בתחום זה סטיית התקן $\sqrt{p(1-p)}$ במכנה בנוסחה (49) היא יורדת ב-p. המונה עולה ב-p, ולכן פונקציית העוצמה מונוטונית עולה ב-p. ציור 4.8 מדגים את פונקציית העוצמה $\pi(p)$, עבור $n=18$, כאשר $p \ge 1/2$. אנו רואים שהעוצמה היא פונקציה עולה של p, והיא שווה ל-1 עבור ערכי p גדולים מאוד. זה נכון לכל גודל מדגם n, למרות שהצורה המדויקת של הפונקציה שונה, כמובן, בגוודלי מדגם שונים.



ציור 4.8. *פונקציית העוצמה של מבחן הסימן,* $n=18$

קביעת גודל המדגם

קל לראות מנוסחה (49) שעבור כל p קבוע, העוצמה עולה ב-n, ומקיימת: $\lim_{n\to\infty} \pi(p)=1$. לכן ניתן למצוא גודל מדגם n שעבורו העוצמה תהיה גדולה כרצוננו.

אם נדרוש שעבור ערך ספציפי p העוצמה תהיה לפחות π, נקבל את הדרישה:

$$\pi(p)=\Phi\left(\frac{\sqrt{n}(p-1/2)}{\sqrt{p(1-p)}}-z_{1-\alpha}\frac{1/2}{\sqrt{p(1-p)}}\right)\ge\pi$$

חילוץ הערך של n מהאי־שוויון לעיל נותן את גודל המדגם (בדקו! תרגיל 5 ב):

$$n\ge\frac{\left[z_{1-\alpha}\cdot 1/2+z_{\pi}\sqrt{p(1-p)}\right]^2}{(p-1/2)^2} \qquad (50)$$

דוגמה 4.15 (המשך דוגמה 4.14). עבור מדגם בגודל 18 מצאנו בדוגמה 4.14 שעוצמת מבחן הסימן ברמת מובהקות $\alpha=.05$, אם $p=.6$, היא $\pi(0.6)=.2088$, וזו עוצמה די קטנה.

נחשב את גודל המדגם הדרוש כדי לקבל עוצמה של 80% לפחות, כאשר $p=0.6$.

האחוזונים המתאימים של ההתפלגות הנורמלית הם: $z_{1-\alpha}=z_{.95}=1.645$, $z_{\pi}=z_{.80}=0.84$.

לפי נוסחה (50) מקבלים

$$n \ge \frac{\left[z_{1-\alpha} \cdot 1/2 + z_{\pi}\sqrt{p(1-p)}\right]^2}{(p-1/2)^2}$$

$$= \frac{\left[1.645 \cdot 1/2 + 0.84\sqrt{0.6(0.4)}\right]^2}{(0.6-1/2)^2} = 152.3$$

יש לדגום לפחות 153 תצפיות כדי לקיים את הדרישה.

עוצמה מקורבת של מבחן הסימן במודל הזזה

כדי להשתמש בנוסחת העוצמה (49) יש למצוא את הערך של p התלוי, כמובן, בהתפלגות F_D. במקרים רבים החישוב די מסבך. ניתן לחשב ערך זה של p בקירוב, בהנחות מסוימות על המודל של הבעיה.

נסתכל על **מודל הזזה**, שבו התפלגות ההפרשים F_D מקיימת: $F_D(t) = F_0(t-\Delta)$, כאשר F_0 היא התפלגות עם חציון השווה לאפס. במודל זה נוכל לרשום את ההשערות הרשומות ב-(5) בצורה הבאה:

$$H_0 : \Delta = 0 \qquad H_1 : \Delta > 0 \qquad (51)$$

בדומה לצורת החישוב של עוצמת מבחן ווילקוקסון לשני מדגמים בפרק 2, נסתכל על המקרה של ערכי Δ הקרובים ל-0. את הפרמטר p ניתן לרשום באופן הבא:

$$p = P(D>0) = 1 - F_D(0) = 1 - F_0(-\Delta)$$

נפַתח את F_0 לטור טיילור סביב 0, בהנחה ש-F_0 בעלת פונקציית צפיפות f_0, וניקח רק את שני האיברים הראשונים של הטור. אנו מניחים ש-Δ קרוב לאפס, ולכן האיברים הנוספים בטור הם זניחים. נקבל לפי זה את הקירוב:

$$F_0(-\Delta) \cong F_0(0) - \Delta f_0(0) = \frac{1}{2} - \Delta f_0(0)$$

הצבנו $F_0(0) = \frac{1}{2}$, מכיוון שהחציון של F_0 הוא 0.

מכאן נקבל קירוב עבור p:

$$p = 1 - F_0(-\Delta) \cong 1 - \frac{1}{2} + \Delta f_0(0) = \frac{1}{2} + \Delta f_0(0)$$

מכאן מקבלים:

$$p - \frac{1}{2} \cong \Delta f_0(0) \qquad (52)$$

נוסף על הקירוב שעשינו עבור הערך של p, נסתכל גם על סטיית התקן במכנה – $\sqrt{p(1-p)}$. זוהי פונקציה שטוחה מאוד, במיוחד בסביבה קרובה של $p=1/2$, שם מתקבל המקסימום שלה. לפיכך סטיית התקן היא כמעט קבועה בסביבה שאינה מאוד רחוקה מ-$1/2$. לפיכך במקרה ש-Δ קרוב לאפס p קרוב ל-$1/2$, ולכן ניתן לרשום

בקירוב $1/2 = \sqrt{1/2 \cdot 1/2} \cong \sqrt{p(1-p)}$.
העוצמה המקורבת של מבחן הסימן במודל של הזזה מתקבלת על ידי הצבת הערכים המקורבים הללו בנוסחה (49):

$$\pi(p) \cong \Phi\left(\frac{\sqrt{n}\Delta f_0(0)}{1/2} - z_{1-\alpha}\right) = \Phi\left(2\sqrt{n}\Delta f_0(0) - z_{1-\alpha}\right) \qquad (53)$$

גודל המדגם של מבחן הסימן במודל הזזה

אם משתמשים בנוסחה המקורבת (53) לערך של π במודל הזזה, ניתן למצוא את גודל המדגם על ידי חילוץ n מהמשוואה

$$\pi(p) \cong \Phi\left(2\sqrt{n}\Delta f_0(0) - z_{1-\alpha}\right) \geq \pi$$

מתקבלת נוסחה מקורבת עבור גודל המדגם הדרוש כדי שמבחן הסימן יהיה בעל עוצמה π לפחות, עבור ערכי Δ קטנים:

$$n \cong \frac{\left[z_{1-\alpha} + z_{\pi}\right]^2}{4\Delta^2[f_0(0)]^2} \qquad (54)$$

נשים לב שגודל המדגם הדרוש תלוי במודל ההסתברותי של הבעיה, כלומר בפונקציית הצפיפות f_0 של ההפרשים תחת השערת האפס.

דוגמה 4.16. מודל נורמלי. נניח שהתפלגות ההפרשים F_D היא נורמלית, $D \sim N(\mu_D, \sigma_D^2)$. כלומר, F_0 היא התפלגות נורמלית עם תוחלת $\mu = 0$.

מכאן $f_0(t) = \frac{1}{\sqrt{2\pi\sigma_D^2}} \cdot e^{-t^2/2\sigma_D^2}$ ולכן $f_0(0) = \frac{1}{\sqrt{2\pi\sigma_D^2}}$.

גודל המדגם הדרוש לקבלת עוצמה π^* לפחות, ניתן לפי נוסחה (54):

$$n \cong \frac{\left[z_{1-\alpha} + z_{\pi^*}\right]^2}{4\Delta^2[f_0(0)]^2} = \frac{2\pi\sigma_D^2\left[z_{1-\alpha} + z_{\pi^*}\right]^2}{4\Delta^2} = \frac{\pi\sigma_D^2\left[z_{1-\alpha} + z_{\pi^*}\right]^2}{2\Delta^2}$$

(כדי למנוע בלבול, רשמנו כאן את העוצמה הדרושה כ-π^*.)

גודל המדגם עולה כפונקציה של שונות ההפרשים והוא יורד כפונקציה של פרמטר ההזזה Δ.

נחשב את הערך המקורב של n לבעיה דומה לזו שהוצגה בדוגמה 4.15, כלומר, למבחן הסימן ברמת מובהקות $\alpha = .05$ כשהעוצמה הדרושה היא $\pi^* = .80$, בהנחה של מודל נורמלי, עם $\Delta = \sigma_D/2$. מקבלים:

$$n \cong \frac{\pi\sigma_D^2\left[z_{1-\alpha}+z_{\pi^*}\right]^2}{2\Delta^2} = \frac{\pi\sigma_D^2\left[1.645+0.84\right]^2}{2\cdot\sigma_D^2/4}$$
$$= 2\pi\left[1.645+0.84\right]^2 = 38.8$$

דרושות לפחות 39 תצפיות.

כדי לברר את מידת הדיוק של הקירוב שעשינו לגבי גודל המדגם, כשהערך של p נרשם בקירוב במודל של הזזה בנוסחה (52), נחשב עתה את גודל המדגם הנדרש בבעיה שלפנינו לפי נוסחה (50) המדויקת יותר. במודל הנורמלי קל לחשב את p.

$D \sim N(\Delta, \sigma_D^2)$ ולכן כאשר $\Delta = \sigma_D/2$, מקבלים

$$p = P_\Delta(D>0) = 1-\Phi\left(\frac{-\Delta}{\sigma_D}\right) = \Phi\left(\frac{\Delta}{\sigma_D}\right) = \Phi\left(\frac{1}{2}\right) = .6915$$

מכאן, לפי נוסחה (50),

$$n \geq \frac{\left[1.645\cdot 1/2 + 0.84\sqrt{(.6915)(.3185)}\right]^2}{(.6915-.5)^2} = 40.36$$

לפי זה דרושות לפחות 41 תצפיות. קיבלנו שהתוצאה המקורבת די דומה לתוצאה המדויקת.

בהמשך נחשב את העוצמה המקורבת ואת גודל המדגם הדרוש לגבי מבחן ווילקוקסון למדגם מזווג ומבחן t למדגם מזווג ונשווה אותם לתוצאות שקיבלנו עבור מבחן הסימן.

4.9 עוצמת מבחן ווילקוקסון למדגם מזווג*

את עוצמת מבחן ווילקוקסון נחשב רק עבור מודל ההזזה (34), ובו יש להניח כי תחת השערת האפס התפלגות ההפרשים F_0 היא **סימטרית** סביב 0. הנחה זו מאפשרת להשתמש במבחן ווילקוקסון.

המודל הוא, אפוא: $F_D(t) = P(D \leq t) = F_0(t-\Delta)$, כאשר $F_0(0) = 1/2$.

ההשערות הנבדקות הן $H_0: \Delta = 0$ כנגד $H_1: \Delta > 0$.

נזכיר שבסעיף קודם ראינו, משפט 4.5, נוסחה (38) שאת הסטטיסטי של ווילקוקסון ניתן לרשום על ידי $V^+ = \#\{i \leq j : W_{ij} > 0\}$, כאשר $W_{ij} = \frac{D_i + D_j}{2}$.

נסמן עתה

$$p_1 = P(D_i + D_j > 0) \qquad i \neq j \tag{55}$$

נשתמש בקירוב הנורמלי של הסטטיסטי V^+. בפרק 2 קיבלנו נוסחה (39) לחישוב

עוצמה אסימפטוטית של מבחן המבוסס על סטטיסטי T בעל התפלגות אסימפטוטית נורמלית. נרשום נוסחה זו שוב כאן:

$$\pi(\Delta) \cong \Phi\left(\frac{E_\Delta T - E_0 T}{\sigma_\Delta(T)} - z_{1-\alpha}\frac{\sigma_0(T)}{\sigma_\Delta(T)}\right) \qquad (56)$$

היות שהמשתנה V^+ הוא בעל התפלגות אסימפטוטית נורמלית, נוכל להשתמש בנוסחה זו. נחזור ונרשום את V^+ כסכום משתנים מציינים, כמו בנוסחה (39) כאן:

$$V^+ = \sum_{j=1}^{n}\sum_{i\le j} h_{ij}$$

כאשר $h_{ij} = \begin{cases} 1 & W_{ij} > 0 \\ 0 & W_{ij} < 0 \end{cases}$.

התוחלת היא של V^+ היא

$$E_\Delta V^+ = \sum_{j=1}^{n}\sum_{i\le j} E_\Delta h_{ij} = \sum_{j=1}^{n} E_\Delta h_{jj} + \sum_{j}\sum_{i<j} E_\Delta h_{ij}$$

$$= nP_\Delta(W_{ii} > 0) + \binom{n}{2} P_\Delta(W_{ij} > 0)$$

$$= nP(D_i > 0) + \binom{n}{2} P(D_i + D_j > 0) = np + \binom{n}{2} p_1$$

תחת H_0, המומנטים של V^+ מנוסחאות (18) ו- (19) הם

$$E_0 V^+ = \frac{n(n+1)}{4} \qquad Var_0(V^+) = \frac{n(n+1)(2n+1)}{24}$$

עבור ערכי Δ קטנים, קירוב די טוב עבור העוצמה נקבל אם נציב בנוסחה (56) במקום השונות במודל האלטרנטיבי את השונות תחת H_0. נקבל, אפוא, את הקירוב הבא:

$$\pi(\Delta) \cong \Phi\left(\frac{E_\Delta V^+ - E_0 V^+}{\sigma_0(V^+)} - z_{1-\alpha}\right) \qquad (57)$$

ההפרש בין שתי התוחלות הוא

$$E_\Delta V^+ - E_0 V^+ = np + \binom{n}{2} p_1 - \frac{n(n+1)}{4}$$

$$= np + \frac{n(n-1)}{2} p_1 - \frac{n(n+1)}{4}$$

ניתן לרשום: $n(n+1) = n(n-1) + 2n$ ומכאן

$$E_\Delta V^+ - E_0 V^+ = \frac{n(n-1)}{2}\left(p_1 - \frac{1}{2}\right) - n\left(p - \frac{1}{2}\right) \qquad (58)$$

נציב בנוסחת הקירוב (56) את השונות תחת H_0 ואת הפרש התוחלות (58):

$$\pi(\Delta)\cong\Phi\left(\frac{\frac{n(n-1)}{2}\left(p_1-\frac{1}{2}\right)-n\left(p-\frac{1}{2}\right)}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}-z_{1-\alpha}\right) \tag{59}$$

כדי למצוא את העוצמה המקורבת, יש לחשב את ערכי שני הפרמטרים p ו-p_1, התלויים, כמובן, בהתפלגות הספציפית F_D של ההפרשים.

עוצמה אסימפטוטית של מבחן ווילקוקסון במודל הזזה

לפי מודל ההזזה קיים $F_D(t)=F_0(t-\Delta)$, כאשר F_0 היא התפלגות סימטרית סביב 0. את הערך של p במודל ההזזה (ללא הנחה על סימטריה) קיבלנו כשדנו במבחן הסימן, נוסחה (52):

$$p-\frac{1}{2}\cong\Delta f_0(0) \tag{60}$$

לגבי חישוב p_1, נזכור שאם $D\sim F_\Delta$, אזי $D-\Delta\sim F_0$, כאשר F_0 סימטרית סביב 0.

$$p_1=P_\Delta(D_1+D_2>0)=P_\Delta\{(D_1-\Delta)+(D_2-\Delta)>-2\Delta\}$$
$$=1-\tilde{F}(-2\Delta)=\tilde{F}(2\Delta)$$

כאשר $\tilde{F}$ היא ההתפלגות של סכום שני משתנים בלתי תלויים מהתפלגות F_0. נניח שלהתפלגות $\tilde{F}$ יש פונקציית צפיפות $\tilde{f}$. אזי ניתן לפתח את $\tilde{F}$ לטור טיילור סביב 0, בדומה לפעולה שביצענו לגבי הפרמטר p. שוב, אנו מניחים ש-Δ קרוב לאפס. הקירוב המתקבל:

$$p_1=\tilde{F}(2\Delta)\cong\tilde{F}(0)+2\Delta\tilde{f}(0)=\frac{1}{2}+2\Delta\tilde{f}(0)$$

ולכן

$$p_1-\frac{1}{2}\cong 2\Delta\tilde{f}(0) \tag{61}$$

עבור ערכים קטנים של Δ מתקבלת העוצמה האסימפטוטית, על ידי הצבת (60) ו-(61) בנוסחת העוצמה (59):

$$\pi(\Delta)\cong\Phi\left(\frac{n(n-1)\Delta\tilde{f}(0)-n\Delta f_0(0)}{\sqrt{n(n+1)(2n+1)/24}}-z_{1-\alpha}\right)$$
$$=\Phi\left(\frac{\sqrt{n}\left[(n-1)\,\tilde{f}(0)-f_0(0)\right]}{\sqrt{(n+1)(2n+1)/24}}\cdot\Delta-z_{1-\alpha}\right) \tag{62}$$

דוגמה 4.17. מודל נורמלי. כמו בדוגמה 4.16, נניח שהתפלגות ההפרשים F_D היא

נורמלית $N(\mu_D,\sigma_D^2)$. כלומר, F_0 היא התפלגות נורמלית עם תוחלת 0 $\mu=0$.

במודל זה ראינו ש- $f_0(0)=\dfrac{1}{\sqrt{2\pi\sigma_D^2}}$.

כמו כן, כאשר $D_1,D_2\sim N(0,\sigma_D^2)$ בלתי תלויים, אזי $D_1+D_2\sim N(0,2\sigma_D^2)$. כלומר, $\tilde{f}$ – צפיפות סכום שני משתנים נורמלים בלתי תלויים, בעלי תוחלת 0, היא צפיפות נורמלית עם תוחלת 0 ושונות $2\sigma_D^2$.

מכאן
$$\tilde{f}(0)=\frac{1}{\sqrt{2\pi\cdot 2\sigma_D^2}}=\frac{1}{2\sigma_D\sqrt{\pi}}$$

נציב את שתי הצפיפויות המתאימות בנוסחת העוצמה (62) ונקבל שבמודל הנורמלי העוצמה האסימפטוטית ניתנת על ידי

$$\pi(\Delta)\cong\Phi\left(\frac{\sqrt{n}\left[\dfrac{(n-1)}{2\sigma_D\sqrt{\pi}}-\dfrac{1}{\sqrt{2\pi\sigma_D^2}}\right]}{\sqrt{\dfrac{(n+1)(2n+1)}{24}}}\cdot\Delta-z_{1-\alpha}\right)$$

$$=\Phi\left(\frac{\sqrt{n}\left[(n-1)/2-1/\sqrt{2}\right]}{\sqrt{(n+1)(2n+1)/24}}\cdot\frac{\Delta}{\sigma_D\sqrt{\pi}}-z_{1-\alpha}\right)\qquad(63)$$

זוהי נוסחה כללית עבור העוצמה המקורבת של מבחן ווילקוקסון למדגם מזווג במודל הנורמלי.

נחשב עוצמה זו עבור מבחן ווילקוקסון ברמת מובהקות $\alpha=.05$, המבוסס על $n=18$ זוגות של תצפיות, בהנחה של מודל נורמלי, עם $\Delta=\sigma_D/2$.

$$\pi(\Delta)\cong\Phi\left(\frac{\sqrt{18}\left[17/2-1/\sqrt{2}\right]}{\sqrt{(19)(37)/24}}\cdot\frac{1}{2\sqrt{\pi}}-1.645\right)=\Phi(0.08)=.5319$$

גודל המדגם של מבחן ווילקוקסון למדגם מזווג במודל הזזה

נסתכל על הנוסחה המקורבת (62) לעוצמה. הדרישה שהעוצמה תהיה לפחות π, אקוויוולנטית לדרישה:

$$\frac{\sqrt{n}[(n-1)\tilde{f}(0)-f_0(0)]}{\sqrt{(n+1)(2n+1)/24}}\cdot\Delta-z_{1-\alpha}\geq z_\pi\qquad(64)$$

את הביטוי הכופל את Δ ניתן לרשום בקירוב, כאשר n גדול

$$A_n=\frac{\sqrt{n}[(n-1)\tilde{f}(0)-f_0(0)]}{\sqrt{(n+1)(2n+1)/24}}\approx\frac{\sqrt{n}\,\tilde{f}(0)}{\sqrt{1/12}}=\sqrt{12}\sqrt{n}\,\tilde{f}(0)$$

לפי זה, אם המדגם יהיה גדול מספיק, הדרישה (68) תתקיים אם יתקיים

$$\sqrt{12}\sqrt{n}\,\tilde{f}(0)\cdot\Delta-z_{1-\alpha}\geq z_\pi \tag{65}$$

חילוץ הערך של n מהאי־שוויון (65) נותן את גודל המדגם הדרוש:

$$n\geq\frac{[z_{1-\alpha}+z_\pi]^2}{12[\tilde{f}(0)]^2\Delta^2} \tag{66}$$

דוגמה 4.18. נחשב את הערך המקורב של n לבעיה שהוצגה בדוגמה 4.15, כלומר, למבחן ווילקוקסון למדגם מזווג ברמת מובהקות $\alpha=.05$, כשהעוצמה הדרושה היא $\pi^*=.80$, בהנחה של מודל נורמלי עם $\Delta=\frac{\sigma_D}{2}$. במודל נורמלי מצאנו בדוגמה 4.17: $\tilde{f}(0)=\frac{1}{2\sigma_D\sqrt{\pi}}$. נציב בנוסחה (66) ונקבל את גודל המדגם הדרוש:

$$n\geq\frac{[z_{1-\alpha}+z_{\pi^*}]^2}{12[\tilde{f}(0)]^2\Delta^2}=\frac{[1.645+0.84]^2}{12\left[\frac{1}{4\pi\sigma_D^2}\right]\frac{\sigma_D^2}{4}}=\frac{4\pi[1.645+0.84]^2}{3}=25.9$$

דרושות לפחות 26 תצפיות.

גודל המדגם הדרוש במודל נורמלי

ניתן לקבל נוסחה כללית לגודל המדגם הדרוש במבחן ווילקוקסון עבור מודל נורמלי. על פי הביטוי שקיבלנו בדוגמה 4.18 עבור $\tilde{f}(0)$ ושימוש בנוסחה (66), אם משתמשים במבחן ווילקוקסון ברמת מובהקות α ודרושה עוצמה של π^* לפחות, גודל המדגם הדרוש הוא

$$n\geq\frac{[z_{1-\alpha}+z_{\pi^*}]^2}{12\cdot\frac{1}{4\pi\sigma_D^2}\cdot\Delta^2}=\frac{\pi\sigma_D^2[z_{1-\alpha}+z_{\pi^*}]^2}{3\Delta^2} \tag{67}$$

4.10 עוצמה מקורבת של מבחן t מזווג

אם מניחים שהפרשי התצפיות D_i מתפלגים נורמלית, כלומר $D_i\sim N(\Delta,\sigma_D^2)$, $i=1,\ldots,n$, את הבעיה הסטטיסטית ניתן לרשום על ידי $H_0:\Delta=0$, $H_1:\Delta>0$. המבחן המתאים הוא מבחן t מזווג. סטטיסטי המבחן הוא

$$t=\frac{\overline{D}}{\sqrt{S_D^2/n}}=\frac{\sqrt{n}\overline{D}}{S_D} \tag{68}$$

כאשר $\overline{D}=\frac{1}{n}\sum_{i=1}^{n}D_i$ ו- $S_D^2=\frac{1}{n-1}\sum_{i=1}^{n}(D_i-\overline{D})^2$ הוא אומד חסר הטיה לשונות σ_D^2. בהנחת נורמליות של התצפיות, תחת השערת האפס הסטטיסטי t הוא בעל התפלגות t של סטודנט, עם $n-1$ דרגות חופש. דוחים את השערת האפס עבור ערכים גדולים של t, או, כאשר $t>t_{n-1,1-\alpha}$.

התפלגות t עם k דרגות חופש שואפת להתפלגות נורמלית סטנדרטית כאשר k שואף לאינסוף, ולכן תחת השערת האפס, הסטטיסטי t בנוסחה (68) הוא בעל התפלגות אסימפטוטית נורמלית. לפיכך את עוצמת המבחן ניתן לרשום באופן הבא:

$$\pi(\Delta)=P_\Delta(t\geq t_{1-\alpha})=P_\Delta\left(\frac{\sqrt{n}\overline{D}}{S_D}\geq t_{1-\alpha}\right)$$

$$=P_\Delta\left(\frac{\sqrt{n}(\overline{D}-\Delta)}{S_D}\geq t_{1-\alpha}-\frac{\sqrt{n}\Delta}{S_D}\right)$$

כאשר n גדול $S_D\cong\sigma_D$ ולכן $\frac{\sqrt{n}(\overline{D}-\Delta)}{s_D}\approx N(0,1)$ וכמו כן $t_{1-\alpha}\cong z_{1-\alpha}$. מכאן נמשיך את השוויון

$$\cong P\left(Z\geq t_{1-\alpha}-\frac{\sqrt{n}\Delta}{S_D}\right)\cong 1-\Phi\left(z_{1-\alpha}-\frac{\sqrt{n}\Delta}{\sigma_D}\right)$$

$$=\Phi\left(\frac{\sqrt{n}\Delta}{\sigma_D}-z_{1-\alpha}\right)$$

קיבלנו, אפוא, נוסחת קירוב עבור העוצמה של מבחן t מזווג כאשר מדגם הזוגות גדול:

$$\pi(\Delta)\cong\Phi\left(\frac{\sqrt{n}\Delta}{\sigma_D}-z_{1-\alpha}\right) \tag{69}$$

חישוב נוסחת הקירוב התבסס על שני קירובים:

(1) על כך שסטיית התקן במדגם של n תצפיות בלתי תלויות מאותה התפלגות שואפת לסטיית התקן של האוכלוסייה, כאשר גודל המדגם שואף לאינסוף;

(2) על כך שהתפלגות משתנה t שואפת להתפלגות נורמלית סטנדרטית, כאשר מספר דרגות החופש שואף לאינסוף.

הערה: העוצמה האסימפטוטית של מבחן t תלויה במודל ההסתברותי של הבעיה רק דרך שונות התפלגות ההפרשים.

דוגמה 4.19 (*המשך דוגמה* 4.17). נשווה את עוצמת מבחן t לעוצמת מבחן ווילקוקסון,

כפי שחושבה בדוגמה 4.17, עם $n=18$ תצפיות, כאשר $\Delta=\frac{\sigma_D}{2}$. נציב את Δ בנוסחה (69) ונקבל את העוצמה המקורבת של מבחן t:

$$\pi(\Delta)\cong\Phi\left(\frac{\sqrt{n}\Delta}{\sigma_D}-z_{1-\alpha}\right)=\Phi\left(\frac{\sqrt{18}}{2}-1.645\right)=\Phi(0.48)=.6844$$

עוצמה זו גבוהה במקצת מהעוצמה שהתקבלה עבור מבחן ווילקוקסון, שם קיבלנו $\pi(\Delta)\cong.5319$.

מציאת גודל המדגם

כדי למצוא את גודל המדגם הדרוש לקבלת עוצמה π כשמשתמשים במבחן t מזווג, יש לחלץ את n מהאי־שוויון המתקבל לפי נוסחת העוצמה (69). מקבלים:

$$n\geq\frac{[z_{1-\alpha}+z_\pi]^2\sigma_D^2}{\Delta^2} \qquad (70)$$

לדוגמה, עבור הבעיה שהובאה בדוגמה 4.18, שם $\alpha=.05$, $\Delta=\sigma_D/2$ ודרושה עוצמה של 0.80, מספר התצפיות הדרושות כשמשתמשים במבחן t הוא

$$n\geq\frac{[1.645+0.84]^2\sigma_D^2}{\sigma_D^2/4}=4(6.175)=24.7$$

דרושות לפחות 25 תצפיות.

בהשוואה למבחן הדרגות של ווילקוקסון, דוגמה 4.18, שבו נדרשו לפחות 26 תצפיות, אנו רואים שהפער הוא קטן ביותר.

יעילות יחסית אסימפטוטית

בפרק 2 נוסחה (56) הגדרנו יעילות אסימפטוטית יחסית בין שני מבחנים על ידי המנה בין גודלי המדגם הדרושים להשגת עוצמה שווה.

בפרק זה דנו בשלושה מבחנים המתאימים לבעיה של מדגם מזווג – מבחן הסימן, מבחן ווילקוקסון ומבחן t. על סמך גודלי המדגם המקורבים שמצאנו עבור שלושת המבחנים הללו, בנוסחאות (54), (66) ו-(70) ניתן למצוא את היעילויות היחסיות ביניהם, במודל של הזזה, כאשר ההתפלגות סימטרית (רק אז לגיטימי להשתמש במבחן ווילקוקסון!).

א) היעילות היחסית האסימפטוטית בין מבחן ווילקוקסון למדגם מזווג לבין מבחן t למדגם מזווג ניתנת על ידי

$$ARE(V^+,t)=\frac{n_t}{n_V}=12\sigma_D^2[\tilde{f}(0)]^2 \qquad (71)$$

ב) היעילות היחסית האסימפטוטית בין מבחן הסימן לבין מבחן t מזווג ניתנת על ידי

$$ARE(S_n,t)=\frac{n_t}{n_S}=4\sigma_D^2[f_0(0)]^2 \qquad (72)$$

ג) את היעילות היחסית האסימפטוטית בין מבחן הסימן לבין מבחן ווילקוקסון למדגם מזווג ניתן לקבל בעזרת שתי היעילויות הקודמות, לפי:

$$ARE(S_n,V^+)=\frac{n_V}{n_S}=\frac{n_t/n_S}{n_t/n_V}=\frac{ARE(S_n,t)}{ARE(V^+,t)}=\frac{[f_0(0)]^2}{3[\tilde{f}(0)]^2}$$

נביא כאן את היעילויות האסימפטוטיות המתקבלות לגבי שני מודלים, שבהם ההתפלגות היא סימטרית.

מודל נורמלי

במודל נורמלי מצאנו את הביטויים הנחוצים:

$$f_0(0)=\frac{1}{\sqrt{2\pi\sigma_D^2}} \quad \text{וכן} \quad \tilde{f}(0)=\frac{1}{2\sigma_D\sqrt{\pi}}\,.$$

מכאן היעילויות היחסיות הן

$$ARE(V^+,t)=12\sigma_D^2[\tilde{f}(0)]^2=12\sigma_D^2\frac{1}{4\sigma_D^2\pi}=\frac{3}{\pi}\cong 0.955$$

יעילות זו זהה ליעילות בין מבחן ווילקוקסון למבחן t בבעיית שני מדגמים (דוגמה 2.19).

$$ARE(S_n,t)=4\sigma_D^2[f_0(0)]^2=4\sigma_D^2\frac{1}{2\pi\sigma_D^2}=\frac{2}{\pi}\cong 0.637$$

$$ARE(S_n,V^+)=\frac{[f_0(0)]^2}{3[\tilde{f}(0)]^2}=\frac{1/2\pi\sigma_D^2}{3\left[1/4\sigma_D^2\pi\right]}=\frac{2}{3}$$

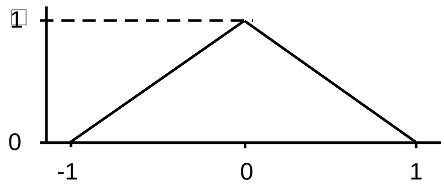
מסקנה: במודל הנורמלי מבחן ווילקוקסון יעיל כמעט כמו מבחן t, בעוד שמבחן הסימן יעיל הרבה פחות. להשגת עוצמה באותו גודל, במבחן ווילקוקסון דרושים כ-$2/3$ של מספר התצפיות בהשוואה למספר התצפיות הדרוש במבחן הסימן.

מודל אחיד

נסתכל על המודל שבו תחת השערת האפס D משתנה אחיד $D\sim U(-0.5,\ 0.5)$.

הצפיפות f_0 היא הצפיפות האחידה על $[-0.5,\ 0.5]$ $(-0.5,\ 0.5)$ ולכן $f_0(0)=1$.

$\tilde{f}$ היא צפיפות הסכום של שני משתנים בלתי תלויים בעלי התפלגות אחידה כנ"ל. קל לראות שהצפיפות המתאימה ניתנת על ידי התיאור הגרפי בציור 4.9. לפי זה $\tilde{f}(0)=1$.



ציור 4.9. *פונקציית הצפיפות של סכום שני משתנים אחידים בלתי תלויים*

השונות σ_D^2 היא שונות המשתנה האחיד המתאים:

$$\sigma_D^2 = Var(D) = 1/12$$

מהצבת הביטויים שמצאנו בנוסחאות היעילות היחסית מקבלים:

$$ARE(V^+, t) = 12\sigma_D^2[\tilde{f}(0)]^2 = 12 \cdot \frac{1}{12} \cdot 1 = 1$$

גם יעילות זו זהה ליעילות בין מבחן ווילקוקסון למבחן בבעיית שני המדגמים (דוגמה 2.20)

$$ARE(S_n, t) = 4\sigma_D^2[f_0(0)]^2 = 4 \cdot \frac{1}{12} \cdot 1 = \frac{1}{3}$$

$$ARE(S_n, V^+) = \frac{[f_0(0)]^2}{3[\tilde{f}(0)]^2} = \frac{1}{3}$$

מסקנה: במודל אחיד מבחן הסימן הוא גרוע מאוד, יחסית לשני המבחנים האחרים.

הערה: ניתן למצוא מודלים שעבורם מבחן הסימן עדיף על מבחן ווילקוקסון וגם על מבחן t. למשל, כאשר התפלגות ההפרשים היא מעריכית כפולה (ראו הגדרת התפלגות זו בפרק 2, דוגמה 2.21), היעילות של מבחן הסימן ביחס למבחן ווילקוקסון היא $4/3$ וביחס למבחן t היעילות היא 2. עובדה זו מפתיעה, כיוון שאיננו מצפים לקבל עוצמה גבוהה במבחן הסימן.

תרגילים

1. בפני 20 ילדים הוצגו משקאות זהים, השונים זה מזה רק בצבע. 15 מתוכם בחרו את המשקה מצבע א והשאר את המשקה מצבע ב.
   א) תנו את המובהקות המדויקת של התוצאה על פי הטבלה המתאימה.
   ב) חשבו את המובהקות בעזרת הקירוב הנורמלי. השוו את שתי התוצאות והסיקו עבור $\alpha = .05$ אם הצבע משפיע על הבחירה.

2. ספָּק של צבעים טוען שקיימת תוספת חדשה לצבע אשר מקצרת את זמן הייבוש של החלק הצבוע. לבדיקת הטענה נצבעו 10 קרשים, כאשר חצי של כל קרש נצבע בצבע הרגיל והחצי השני – בצבע החדש. זמני הייבוש, בשעות, רשומים להלן:

הקרש:	1	2	3	4	5	6	7	8	9	10
עם תוספת:	6.4	5.8	7.4	5.5	6.3	7.8	8.6	8.2	7.0	4.9
ללא תוספת:	6.6	5.8	7.8	5.7	6.0	8.4	8.8	8.4	7.3	5.8

תנו את המובהקות המדויקת של התוצאה על פי מבחן הסימן והסיקו, עבור $\alpha=.01$, אם הספק צודק.

3. מהנדס מזון בחן 15 צנצנות ריבה כדי לקבוע את אחוז החומר הסינתטי שהוסף לריבה. האחוזים שנרשמו הם

2.4, 2.3, 1.7, 1.7, 2.3, 1.2, 1.1, 3.6, 3.1, 1.0, 4.2, 2.3, 1.6, 2.5, 2.4

התקן הקבוע הוא של התפלגות עם חציון 2% חומר סינתטי.

א) האם התוצרת הזאת עומדת בתקן? בדקו בר"מ $\alpha=.01$.

ב) בכל מקרה נדרש שלא יותר מעשירית הצנצנות יכילו יותר מ-3% חומר סינתטי. בדקו ($\alpha=.05$) אם התוצרת עומדת בדרישה זו.

4. על סמך 12 תצפיות של זוגות (X_i,Y_i), $i=1,\ldots,12$ רוצים לערוך מבחן סימן (חד־צדדי) ברמת מובהקות של $\alpha=.04$.

א) מהו אזור הדחייה?

ב) מה עוצמת המבחן שקבעתם, אם למעשה $P(Y>X)=.6$?

ג) מהו גודל המדגם המינימלי שתצטרכו לבחור כדי שעבור רמת מובהקות $\alpha=.04$ תתקבל עוצמה של .90 לפחות?

5. יהי S_n הסטטיסטי של מבחן הסימן ברמת מובהקות α המבוסס על n תצפיות, לבדיקת ההשערה $H_0:p=1/2$ כנגד $H_1:p>1/2$, כאשר $p=P(X<Y)$.

א) רִשמו נוסחה כללית של פונקציית העוצמה של המבחן כפונקציה של p, כאשר $p>1/2$. הציבו והשוו לתוצאה שקיבלתם בשאלה 4 חלק ב'.

ב) רִשמו נוסחה כללית לגודל המדגם הדרוש כדי שמבחן הסימן ברמת מובהקות α יהיה בעל עוצמה π לפחות, אם למעשה $p=P(X<Y)$, כאשר $p>1/2$. הציבו והשוו לתוצאה שקיבלתם בשאלה 4 חלק ג'.

6. בודקים את ההשערה $H_0:P(Y>X)=1/2$ על סמך n זוגות של תצפיות (X_i,Y_i), $i=1,\ldots,n$. הוכיחו כי תחת H_0 התפלגות הסטטיסטי V^+ של ווילקוקסון היא סימטרית סביב $n(n+1)/4$.

(רמז: הסתכלו על המשתנה $1-T_j$. כיצד הוא מתפלג?)

7. כדי לבחון את ההבדל בין מבחן הסימן לבין מבחן ווילקוקסון למדגם מזווג, נתונים

בזה שני מדגמים פיקטיביים של הפרשים. עִרכו לגבי כל אחד מן המדגמים הללו את שני המבחנים לבדיקת ההשערה שהחציון שלהם הוא אפס כנגד האלטרנטיבה שהוא חיובי. חשבו את מובהקות התוצאה לגבי כל מבחן והסבירו ממה נובעים ההבדלים שהתקבלו.

מדגם I :	10	7	6	4	4	3	2	1	–1
מדגם II :	–10	7	6	4	4	3	2	1	1

8. עִרכו מבחן ווילקוקסון לגבי נתוני שאלה 2, בעיית הצבע. חשבו את המובהקות **המדויקת** וכן את המובהקות שמתקבלת בעזרת הקירוב הנורמלי. השוו את התוצאות והסיקו לגבי טיב הצבע החדש, בר"מ של 0.01.

9. בבדיקה שנערכה להשוואת השפעת גלולות מסוג חדש להרגעת כאבי ראש לעומת תרופה סטנדרטית, נרשמו תשובותיהם של 9 חולים לגבי מידת ההעדפה של הסוג החדש על הסוג הסטנדרטי:

הרבה פחות	קצת פחות	אין העדפה	קצת יותר	הרבה יותר
1	1	2	4	1

א) נתחו את הנתונים בעזרת מבחן הסימן ותנו את מובהקות התוצאה. הסיקו אם ניתן לומר שהתרופה החדשה עדיפה ($\alpha = .05$).

ב) השתמשו במבחן ווילקוקסון וחשבו (על ידי מְנִייה) את המובהקות המדויקת של התוצאה שהתקבלה. מה המסקנה?

10. השתמשו בקירוב הנורמלי (עבור מבחן הסימן ועבור מבחן ווילקוקסון) כדי לענות על שאלה 8, בהנחה שהניסוי נערך לגבי 30 חולים שהעדפותיהם היו כדלהלן:

הרבה פחות	קצת פחות	אין העדפה	קצת יותר	הרבה יותר
1	6	8	12	3

11. תנו רווח בר־סמך 95% עבור החציון של התפלגות אחוז החומר הסינתטי בריבה (שאלה 3), על סמך ההתפלגות הבינומית.

12. במחקר לבדיקת האפשרות לאבחון טוב של ילדים לקויי למידה (עדנה שדה) נבחר מדגם של ילדים שאובחנו כלקויי למידה והועבר להם מבחן אינטליגנציה וקסלר-R. לכל ילד נרשם הציון המילולי והציון הביצועי שקיבל במבחן. לשם השוואה נערכו מבחנים אלה גם לקבוצה מקבילה של ילדים רגילים. אנו מביאים כאן נתונים חלקיים, עבור 12 ילדים מכל קבוצה.

ילדים רגילים		לקויי למידה	
ציון ביצועי	ציון מילולי	ציון ביצועי	ציון מילולי
68	68	92	73
80	85	88	102
73	92	110	88
98	75	85	72
63	85	72	77
122	133	97	58
85	67	115	107
95	97	93	92
117	102	115	83
100	88	58	72
90	107	115	55
100	100	102	88

א) בדקו ($\alpha=.05$) אם אצל ילדים לקויי למידה יש הבדל בין הציון הביצועי והציון המילולי.

ב) חזרו על חלק א עבור ילדים רגילים.

ג) בדקו ($\alpha=.05$) אם הציון הביצועי אצל ילדים לקויי למידה נמוך מזה של ילדים רגילים וכן אם הציון המילולי אצל ילדים לקויי למידה נמוך מזה של ילדים רגילים.

ד) האם הפער בין הציון הביצועי לציון המילולי יכול לתת אבחון טוב להיותו של ילד לקוי למידה? ערכו מבחן מתאים ($\alpha=.05$) והסבירו.

ה) סכמו את הממצאים לגבי בעיית המחקר.

13. תנו הערכה לגודל המדגם הדרוש, כדי שעבור רמת מובהקות $\alpha=.01$ תתקבל עוצמה של 0.95 בערך, תחת מודל של הזזה, אם F_0 היא נורמלית ו- $\frac{\Delta}{\sigma_D}=\frac{1}{2}$:

א) אם משתמשים במבחן הסימן (הערכה יותר מדויקת וקצת פחות מדויקת);

ב) אם משתמשים במבחן ווילקוקסון מזווג;

ג) אם משתמשים במבחן t מזווג.

14. חִזרו על שאלה 13 אם F_0 היא אחידה $U(-1/2,1/2)$ ו- $\Delta=0.1$.

15. חשבו את היעילויות היחסיות עבור כל אחד משני המודלים לעיל (הנורמלי והאחיד):

א) בין מבחן ווילקוקסון מזווג לבין מבחן t מזווג;

ב) בין מבחן הסימן לבין מבחן t מזווג;

ג) בין מבחן הסימן למבחן ווילקוקסון מזווג.

מהי המסקנה?

16. הנתונים בטבלה להלן הם המשקל (בק"ג) ש-12 תלמידים הצליחו להרים לפני תכנית אימונים של 8 שבועות ואחריה.

התלמיד	לפני	אחרי
1	28.8	40.8
2	31.8	45.8
3	28.8	38.8
4	27.8	48.8
5	33.2	30.2
6	34.8	41.8
7	37.2	49.2
8	40.4	48.8
9	40.4	49.8
10	30.8	39.8
11	30.8	42.8
12	28.2	42.8

יהי Δ – חציון התפלגות ההפרש (מידת השיפור בעזרת האימונים).

א) תנו אומדן ל-Δ בשתי שיטות – $\hat{\Delta}$ ו- $\hat{\Delta}_H$ (שיטת הודג'ס-להמן).

ב) תנו עבור Δ רווח בר־סמך 90% בעזרת ההתפלגות הבינומית.

ג) תנו רווח בר־סמך 90% עבור Δ בעזרת ההתפלגות של הסטטיסטי V^+ של ווילקוקסון. השוו לחלק ב.

פרק 5

השוואת יותר משני מדגמים בלתי תלויים

בפרק 2 הבאנו את מבחן ווילקוקסון להשוואת המיקום של שתי אוכלוסיות. כאן נביא הכללה של מבחן זה, שהוצעה על ידי קרוסקל וואליס (Kruskal & Wallis, 1952), כאשר הניסוי כולל יותר משני מדגמים, או טיפולים. (ההכללה מתבררת מתרגיל 5.)
נניח שאנו רוצים לבחון אם יש הבדל בין k טיפולים שונים (או בין k אוכלוסיות). בידינו k מדגמים בלתי-תלויים, כאשר כל מדגם הוא אוסף התצפיות שנמדדו מהנבדקים שעברו את הטיפול המתאים, או מדגם של תצפיות מהאוכלוסייה המתאימה.
נסמן את התצפיות על ידי
X_{ij} – התצפית ה-j במדגם ה-i, $(i=1,2,\ldots,k;\ j=1,2,\ldots n_i)$, והיא בעלת התפלגות שנסמנה F_i.

המבנה של הנתונים שבידינו הוא לפי המתואר בלוח 5.1.

לוח 5.1. *נתונים של k מדגמים וסימונם*

הטיפול			
1	2	…	k
X_{11}	X_{21}	…	X_{k1}
X_{12}	X_{22}	…	X_{k2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_{1n_1}	X_{2n_1}	…	X_{kn_k}

מספר התצפיות במדגם ה-i מסומן n_i ונסמן שוב, כמו בפרק 2, את מספר התצפיות הכולל שבידינו ב-N. כלומר, $N=\sum_{i=1}^{k} n_i$.

5.1 מבחן קרוסקל-וואליס

נניח שכל k ההתפלגויות הן רציפות. כלומר, אין סיכוי לקבל שתי תוצאות זהות (או יותר).

השערת האפס, שאין הבדל בין הטיפולים, ניתנת להירשם:

$$H_0 : F_1(t) = F_2(t) = \cdots = F_k(t) \qquad \text{לכל } t, \qquad (1)$$

כלומר, כל ההתפלגויות שוות.

ההשערה האלטרנטיבית היא שישנו איזשהו הבדל בין ההתפלגויות (או האוכלוסיות). כרגע לא נדון כלל באלטרנטיבה כיוונית. השאלה הנשאלת כאן היא אם על סמך נתוני המדגמים ניתן לומר שכל האוכלוסיות הן זהות, או שיש לדחות השערה זו ולהחליט שהן אינן זהות.

המבחן של קרוסקל-וואליס

הסטטיסטי של קרוסקל וּוואליס הוא הכללה פשוטה של הסטטיסטי של ווילקוקסון לשני מדגמים.

מדרגים את כל N התצפיות מהערך הנמוך ביותר (דרגה 1) ועד הגבוה ביותר (דרגה N). הדרגה של תצפית X_{ij} מסומנת R_{ij}. מסכמים את הדרגות בכל אחד מהמדגמים:

$$R_i = \sum_{j=1}^{n_i} R_{ij} \qquad (2)$$

סטטיסטי המבחן של קרוסקל-וואליס הוא מדד להבדל בין סכומי הדרגות הללו (או בין הדרגות הממוצעות במדגמים השונים).

הסטטיסטי של קרוסקל-וואליס מוגדר על ידי

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{1}{n_i} \left[R_i - \frac{n_i(N+1)}{2} \right]^2 \qquad (3)$$

דוחים את השערת האפס עבור ערכים גבוהים של H.

נביא כאן הצדקה לשימוש בנוסחה (3).

כמו במקרה של שני מדגמים, תחת המודל של השערת האפס כל N התצפיות הן בעלות אותה התפלגות (לקוחות מאוכלוסיות זהות), ולכן כל הסידורים של הדרגות בין התצפיות הם שווי הסתברות, ולכל אחת מהתמורות של הדרגות אותה הסתברות – $1/N!$. מכאן נובע שכל אחת מהדרגות בנפרד היא בעלת התפלגות אחידה $U(1,N)$, זאת אומרת, ההסתברות שדרגת תצפית מסוימת תהיה שווה ל-1, ל-2, או לכל ערך אחר בין 1 ל-N היא $1/N$.

ניתן לרשום, אפוא,

(4) $$R_{ij} \sim U(1,N)$$ לכל i ולכל j

בהנחה זו, המומנטים של כל אחת מהדרגות הם

(5) $$ER_{ij} = (N+1)/2 \quad i=1,\ldots,k;\ j=1,\ldots n_i$$

(6) $$Var(R_{ij}) = (N^2-1)/12 \quad i=1,\ldots,k;\ j=1,\ldots n_i$$

עבור סכום הדרגות במדגם ה-j נקבל, אפוא, את התוחלת:

(7) $$ER_i = E\sum_{i=1}^{n_i} R_{ij} = \sum_{i=1}^{n_i} ER_{ij} = n_i \frac{N+1}{2}$$

השונות של R_i מתקבלת בקלות על ידי שימוש בנוסחה שקיבלנו עבור השונות של הסטטיסטי של ווילקוקסון, נוסחה (9) בפרק 2. נסתכל על המדגם ה-i כקבוצת הטיפול וכל שאר התצפיות כביקורת ונציב $m = N - n_i$. נקבל את השונות:

(8) $$Var(R_i) = \frac{n_i(N-n_i)(N+1)}{12} \qquad i=1,\ldots,k$$

נִראה שהסטטיסטי H די קרוב לסכום הריבועים של המשתנים R_i, כשהם מתוקננים תחת השערת האפס. ריבוע המשתנה המתוקנן על פי (7) ו-(8) הוא

$$Z_i^2 = \frac{\left[R_i - ER_i\right]^2}{Var(R_i)} = \frac{\left[R_i - n_i(N+1)/2\right]^2}{n_i(N-n_i)(N+1)/12}$$

$$= \frac{12}{(N-n_i)(N+1)} \cdot \frac{1}{n_i}\left[R_i - n_i(N+1)/2\right]^2$$

מצד שני, את H מנוסחה (3) ניתן לרשום כסכום: $$H = \sum_{i=1}^{k} W_i^2$$

כאשר $$W_i^2 = \frac{12}{N(N+1)} \cdot \frac{1}{n_i}\left[R_i - n_i(N+1)/2\right]^2$$

W_i^2 אינו זהה ל- Z_i^2, אולם ההבדל ביניהם הוא רק במכנה, שבו מוחלף $N-n_i$ ב-N. ראינו, אפוא, שצורת הסטטיסטי H היא קרובה לסכום ריבועים של משתנים מתוקננים. למעשה הסטטיסטי H דומה לסטטיסטי F שבו משתמשים להשוואת כמה אוכלוסיות במודל הנורמלי. יותר מאוחר נביא בפירוט את הקשר הזה. משפט 5.1 מביא את ההצגה המתאימה של הסטטיסטי של קרוסקל-וואליס.

משפט 5.1. את הסטטיסטי H ניתן לרשום בצורה הבאה: $$H = \frac{SSB}{MST}$$

כאשר $$SSB = \sum_{i=1}^{k} n_i \left[\overline{R}_i - \frac{(N+1)}{2}\right]^2$$

הוא סכום ריבועי הסטיות בין ממוצעי הדרגות של המדגמים לבין ממוצע הדרגות של כל N התצפיות (נהוג לקרוא לו סכום הריבועים **בין** המדגמים – between-groups sum of squares), וכן

$$MST = \frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[R_{ij} - \frac{N+1}{2}\right]^2 = \frac{SST}{N-1}$$

כאשר SST הוא סכום ריבועי הסטיות של כל הדרגות מהממוצע שלהן (נהוג לקרוא לביטוי זה סכום הריבועים הכולל – total sum of squares).

הוכחה: ראשית נעיר כי הממוצע של כל N הדרגות R_{ij} שווה לתוחלת של כל אחת מהדרגות:

$$\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} R_{ij} = \frac{1}{N}(1+2+\cdots N) = \frac{N+1}{2} = ER_{ij}$$

חישוב SST הוא קל מאוד.

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[R_{ij} - \frac{N+1}{2}\right]^2 = \sum_{r=1}^{N}\left[r - \frac{N+1}{2}\right]^2 = NVar(U)$$

כאשר המשתנה U הוא משתנה אחיד $U(1,N)$. שימו לב שהשונות של U ניתנת על ידי סכום הריבועים לעיל, כאשר כל ריבוע כזה יש לכפול בהסתברות המתאימה – $1/N$. נציב את השונות (הידועה!) של U ונחלק ב- $N-1$. נקבל:

$$MST = \frac{SST}{N-1} = \frac{1}{N-1}\cdot\frac{N(N^2-1)}{12} = \frac{N(N+1)}{12} \qquad (9)$$

זהו בדיוק המכנה של הסטטיסטי H.

מצד שני, סכום הריבועים הרשום ב-(3) הוא למעשה סכום הסטיות הריבועיות של ממוצעי הדרגות. נסמן ב- $\bar{R}_i$ את הדרגה הממוצעת במדגם ה-i:

$$\bar{R}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} R_{ij} = \frac{R_i}{n_i}$$

נרשום את סכום הריבועים בין ממוצעי המדגמים:

$$SSB = \sum_{i=1}^{k} n_i\left[\bar{R}_i - \frac{(N+1)}{2}\right]^2 = \sum_{i=1}^{k}\frac{1}{n_i}\left[R_i - \frac{n_i(N+1)}{2}\right]^2 \qquad (10)$$

וביטוי זה הוא בדיוק המונה של H. הצבת (9) ו-(10) בנוסחה (3) נותנת את הסטטיסטי H:

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{1}{n_i}\left[R_i - \frac{n_i(N+1)}{2}\right]^2 = \frac{SSB}{MST} \qquad (11)$$

♣

הצורה הזאת מסבירה את הרעיון של בניית הסטטיסטי. SSB הוא מדד להבדלים בין הדרגות הממוצעות שהתקבלו ב-k המדגמים. כאשר יש הבדלים בין האוכלוסיות, מדד זה נוטה לקבל ערכים גבוהים, לכן דוחים את השערת האפס של שוויון ההתפלגויות עבור ערכים גבוהים של הסטטיסטי H.

הערה: במודל נורמלי נהוג להשתמש לבעיה שבה אנו דנים כאן בסטטיסטי המבחן F, המוגדר על ידי $F=\frac{SSB/(k-1)}{(SST-SSB)/(N-k)}$, אולם סכומי הריבועים לעיל מחושבים על סמך התצפיות המקוריות X_{ij}, בעוד שהסטטיסטי של ווילקוקסון מחושב על סמך הדרגות R_{ij}. המבחן המתאים נקרא "ניתוח שונות חד־כיווני". קל לראות שאם מבחן ניתוח שונות כזה ייערך עבור הדרגות, אזי הערך של F הוא פונקציה של H (תרגיל 6). לפיכך, המבחן של קרוסקל-וואליס אקוויוולנטי למבחן ניתוח שונות חד-כיווני המבוצע על הדרגות.

הערה נוספת: למעשה, המכנה MST בנוסחה (11) הוא קבוע, לפי נוסחה (9), ואינו משנה דבר לגבי כלל ההחלטה. הסיבה היחידה להימצאותו בנוסחת הסטטיסטי H היא האפשרות לקבל נוסחת קירוב עבור התפלגות H למדגמים גדולים, תחת השערת האפס. לאחר דוגמה 5.1 נביא את צורת החישוב של התפלגות H תחת השערת האפס.
את הסטטיסטי H ניתן לחשב גם באופן אחר, כפי שמובא בטענה 5.1.

טענה 5.1. את H ניתן לחשב גם על ידי

$$H=\frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i}-3(N+1) \qquad (12)$$

הוכחה: תרגיל 1.

דוגמה 5.1. להלן נתונים חלקיים ממחקר (פרופ' אבי שדה וחברים) שנעשה כדי לברר השפעה של משך השינה על התפקוד הקוגניטיבי של תלמידי בית ספר יסודי. הנתונים כאן הם לגבי ילדים בכיתה ד. הניסוי נערך במשך 5 ימים, כאשר לאחר הלילה השני חלקם התבקשו להאריך את שעות השינה בשעה אחת, חלקם התבקשו לקצר אותה בשעה אחת והאחרים שימשו כביקורת. משך השינה של כל הילדים נרשם במשך שלושת הלילות הבאים. קבוצה א הם התלמידים שהאריכו את ממוצע שעות השינה שלהם בחצי שעה לפחות, בעקבות בקשה ספציפית של עורכי המחקר; קבוצה ב הם אלה שקיצרו את שעות השינה שלהם בחצי שעה לפחות; וקבוצה ג הם אלה שלא קיצרו ולא האריכו את שעות השינה. הציונים בלוח 5.2 הם נתונים לגבי הפער בזמן התגובה

למטלת קשב (אלפיות השניה) בין המדידה בתחילת הניסוי לבין המדידה בסוף תקופת הניסוי. (זמן תגובה גבוה מעיד על תוצאה גרועה, ולכן פער גבוה בין שני מועדי המדידה מעיד על שיפור הביצוע.)

לוח 5.2. *תוצאות ניסוי להשוואת מדד קשב לפי קבוצת הניסוי (דוגמה 5.1)*

	קבוצת הניסוי					
	א (האריכו)		ב (קיצרו)		ג (ביקורת)	
	דרגה	x	דרגה	x	דרגה	x
	1	-146	3	-77	8	-30
	2	-116	6	-52	17	-9
	4	-69	11	-26	21	8
	5	-58	12	-20	22	13
	7	-50	13	-16	25	42
	9	-28	15	-14	26	57
	10	-27	18	-8	27	58
	14	-15	20	6	28	82
	16	-11	23	15	29	98
	19	5	24	16	30	99
R_i	87		145		233	
n_i	10		10		10	
$\bar{R}_i$	8.7		14.5		23.3	

בדוגמה זו $k=3$, $n_1=n_2=n_3=10$, ו- $N=30$.

סכום כל הדרגות הוא $465=\frac{N(N+1)}{2}$

התוחלת בכל קבוצה היא $155=\frac{10(31)}{2}=\frac{n_i(N+1)}{2}$.

חישוב הסטטיסטי H לפי ההגדרה (3) נותן

$$H=\frac{12}{30(31)}\cdot\frac{1}{10}\left[(233-155)^2+(145-155)^2+(87-155)^2\right]$$

$$=13.95$$

חישוב לפי נוסחה (12) נותן אותה תוצאה, כמובן:

$$H=\frac{12}{30\cdot 33}\cdot\frac{1}{10}\left(233^2+145^2+87^2\right)-3(31)=13.95$$

בינתיים לא ברור לנו אם ערך זה גדול מספיק כדי לדחות את השערת האפס. כדי לברר זאת יש למצוא את מובהקות התוצאה $P_{H_0}(H\geq 13.95)$.

התפלגות הסטטיסטי של קרוסקל-וואליס תחת השערת האפס

כבר הזכרנו שבהנחה שהשערת האפס נכונה, כל N המשתנים X_{ij} הם בלתי תלויים ושווי התפלגות, ולכן לכל אחת מהפרמוטציות של N הדרגות אותה הסתברות – $1/N!$. מצד שני, היות שאנו מסתכלים על סכום הדרגות בכל מדגם, כל הפרמוטציות של דרגות **בתוך** אותה קבוצה אינן משנות את הערך של H. במִלים אחרות, היות שהסטטיסטי H תלוי רק ברשימת **סכומי הדרגות** $R_1,\ldots,R_k$, אין צורך לספור את כל הפרמוטציות של הדרגות שנכללות באותו מדגם. לפיכך ההסתברות לקבלת ערך מסוים של H היא

$$P_{H_0}(H=h)=\frac{\#\{H=h\}}{\binom{N}{n_1\ n_2\ldots n_k}}=\frac{\#\{H=h\}}{\frac{N!}{n_1!n_2!\ldots n_k!}} \qquad (13)$$

המכנה של (13) נקרא **המקדם המולטינומי**, והוא מבטא את מספר האפשרויות לחלק N איברים ל-k קבוצות בנות $n_1,n_2,\ldots,n_k$ איברים, בהתאמה.
דוגמה לבניית התפלגות זו נראה בדוגמה 5.2.

דוגמה 5.2. נמצא את התפלגות הסטטיסטי של קרוסקל-וואליס עבור שלושה מדגמים בגדלים $n_1=1,\ n_2=2,\ n_3=1$. בלוח 5.3 רשומות כל האפשרויות של חלוקת $N=4$ הדרגות $1,\ 2,\ 3,\ 4$ בין שלוש הקבוצות. עבור כל חלוקה כזאת חושב הערך של הסטטיסטי H. מספר האפשרויות השונות של חלוקת ארבע הדרגות ל-3 הקבוצות, לפי גודלי המדגם הנתונים, הוא $\frac{N!}{n_1!n_2!\ldots n_k!}=\frac{4!}{1!\ 2!\ 1!}=12$ וזהו בדיוק מספר האפשרויות הרשומות בלוח 5.3.

פונקציית ההסתברות של H מתקבלת על ידי ספירת האפשרויות לכל אחד מן הערכים של H מן הלוח:

סך הכול	2.7	1.8	0.3	h
1	$\frac{6}{12}$	$\frac{4}{12}$	$\frac{2}{12}$	$P(H=h)$

לוח 5.3. *חלוקת הדרגות וערכי הסטטיסטי H עבור שלושה מדגמים בגדלים*
$n_1=1, n_2=2, n_3=1$

	דרגות בקבוצות				סכום הדרגות R_i				
	א	ב		ג	א	ב	ג	SSB	H
1.	1	2	3	4	1	5	4	4.5	2.7
2.	1	2	4	3	1	6	3	3	1.8
3.	1	3	4	2	1	7	2	4.5	2.7
4.	2	1	3	4	2	4	4	3	1.8
5.	2	1	4	3	2	5	3	0.5	0.3
6.	2	3	4	1	2	7	1	4.5	2.7
7.	3	1	2	4	3	3	4	4.5	2.7
8.	3	1	4	2	3	5	2	0.5	0.3
9.	3	2	4	1	3	6	1	3	1.8
10.	4	1	2	3	4	3	3	4.5	2.7
11.	4	1	3	2	4	4	2	3	1.8
12.	4	2	3	1	4	5	1	4.5	2.7

על פי שיטת הספירה לעיל ניתן לחשב את התפלגות הסטטיסטי H עבור גודלי מדגם שונים. קיימות טבלאות של ההתפלגויות הללו, בדרך כלל עבור שלושה מדגמים בלבד, ועבור מדגמים לא גדולים (עד 5 תצפיות בכל מדגם), כאשר ברוב המקרים נתונים רק ערכי חלוקה ספציפיים, כמו האחוזון ה-90, ה-95 וה-99 של ההתפלגות. איננו מביאים כאן טבלת הסתברויות כזאת, כיוון שכבר עבור מדגמים בסדר גודל בינוני ניתן להשתמש בקירוב די טוב.

התפלגות מקורבת של H

ההתפלגות המקורבת של H היא התפלגות **חי-בריבוע**. ניקח, אפוא, סטייה קלה ונסביר מהי ההתפלגות הזאת.

התפלגות חי־בריבוע

התפלגות הריבוע של משתנה נורמלי סטנדרטי נקראת התפלגות חי־בריבוע עם דרגת חופש אחת. אנו מסמנים $Z^2 \sim \chi_1^2$, כאשר $Z \sim N(0,1)$.
התפלגות חי־בריבוע עם m דרגות חופש, מסומנת χ_m^2, היא התפלגות סכום ריבועים של m משתנים נורמלים סטנדרטיים בלתי־תלויים. כלומר, אם $Z_1,\ldots,Z_m$ הם משתנים

בלתי-תלויים נורמלים סטנדרטיים, אז $T=\sum_{i=1}^{m} Z_i^2 \sim \chi_m^2$. ההתפלגות של T תלויה רק בפרמטר m.

המשתנה T המתפלג חי-בריבוע הוא חיובי תמיד, וכמו כן ההתפלגות איננה סימטרית. דוגמה להתפלגות זאת מוצגת בציור 5.1. טבלת התפלגות חי־בריבוע נמצאת בטבלה 5 בנספח. בטבלה רשומים ערכי חלוקה מסוימים של ההתפלגות, עבור כל ערך של דרגות החופש m.

שימו לב שבמקרה של דרגת חופש אחת בלבד, ניתן למצוא את האחוזונים של התפלגות חי־בריבוע על סמך ההתפלגות הנורמלית, טבלה 1, באופן הבא: אם $T \sim \chi_1$ אזי $T=Z^2$. בגלל הסימטרייה של Z, הסתברות הזנב של המשתנה T היא

$$P(T \geq t) = P(Z^2 \geq t) = P(|Z| \geq \sqrt{t}) = 2P(Z \geq \sqrt{t})$$

מכאן נובע הקשר בין ערכי החלוקה של משתנה חי־בריבוע עם דרגת חופש אחת לבין משתנה נורמלי סטנדרטי: $\chi_{1-\alpha}^2 = (Z_{1-\alpha/2})^2$.

למשל, $\chi_{.95}^2 = (Z_{.975})^2 = 1.96^2 = 3.8416$, וזה אמנם הערך המופיע בטבלה 5.

נחזור עתה לסטטיסטי של קרוסקל-וואליס.

עבור המדגם ה-i, אם גודל המדגם n_i הוא די גדול, אזי המשתנה המתוקנן $Z_i = \frac{R_i - ER_i}{\sqrt{Var(R_i)}}$ הוא בקירוב נורמלי סטנדרטי (כמו סכום הדרגות במבחן ווילקוקסון לשני מדגמים). אם כל k המדגמים הם די גדולים, אזי כל k המשתנים המתוקננים הללו הם בקירוב נורמלים סטנדרטיים. אם משתנים אלה היו בלתי תלויים, סכום הריבועים שלהם היה בעל התפלגות חי־בריבוע עם k דרגות חופש. אולם במקרה זה k המשתנים הללו אינם בלתי תלויים, כיוון שסכום כל הדרגות הוא קבוע:

$$R_1 + \ldots + R_k = 1 + 2 + \cdots + N = \frac{N(N+1)}{2}$$

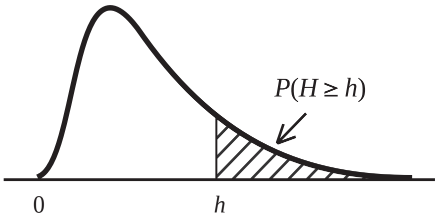
קרוסקל הוכיח שתחת השערת האפס המשתנה H, המוגדר בנוסחה (3), מתפלג בקירוב לפי התפלגות חי־בריבוע עם $k-1$ דרגות חופש. כלומר, יש להוריד דרגת חופש אחת בגלל האילוץ לגבי סכום הדרגות ב-k המדגמים. ננסח זאת כמשפט.

משפט 5.2. אם k המדגמים הם די גדולים, הסטטיסטי H מתפלג בקירוב לפי התפלגות חי־בריבוע עם $k-1$ דרגות חופש.

לא נוכיח כאן את המשפט.

בהסתמך על ההתפלגות המקורבת, מבחן ברמת מובהקות α דוחה את השערת האפס

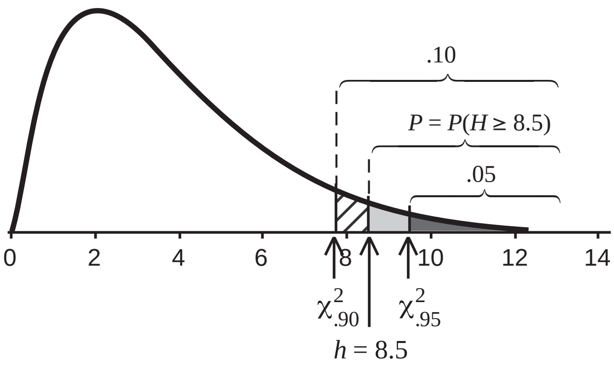
של שוויון התפלגויות כאשר הערך של H גדול מערך החלוקה ה-$1-\alpha$ של התפלגות χ^2_{k-1}. כלומר, אנו דוחים את ההשערה שכל ההתפלגויות שוות כאשר $H>\chi^2_{k-1,1-\alpha}$.
המובהקות המקורבת של תוצאה h היא השטח תחת עקום של הצפיפות מימין לערך h:
$P=P_{H_0}(H\geq h)\cong P(T\geq h)$, כאשר $T\sim\chi^2_{k-1}$. ניתן לראות זאת בציור 5.1.



ציור 5.1. *התפלגות חי־בריבוע והסתברות הזנב הימני*

הבעיה במציאת המובהקות לעיל היא בכך שטבלה 5 בנספח אינה מציגה את ההסתברויות עבור כל ערך h, אלא מביאה רק ערכי חלוקה מסוימים. לדוגמה, נניח שבדקנו את ההבדל בין חמש אוכלוסיות וקיבלנו את הערך של הסטטיסטי של קרוסקל-וואליס $H=8.5$. אנו מביאים כאן רק את השורה המתאימה (4 דרגות חופש) מתוך טבלת חי־בריבוע. הערך $h=8.5$ נמצא בין שני ערכי החלוקה $\chi^2_{.90}=7.78$ ו-$\chi^2_{.95}=9.49$. לפיכך הסתברות הזנב מעל לתוצאה זו היא בין שתי הסתברויות הזנב לעיל, כלומר, בין 0.05 לבין 0.10.

	p												
ד"ח	.005	.010	.025	.050	.100	.250	.500	.750	.900	.950	.975	.990	.995
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.14	13.28	14.86



ציור 5.2. **התפלגות** χ^2_4 *ומובהקות התוצאה במבחן קרוסקל-וואליס*

לשם הבהרה ראו ציור 5.2. המובהקות P מקיימת, אפוא, $0.05 < P < 0.10$. על פי הטבלה המצויה בידינו אין באפשרותנו לדעת את הערך P בדיוק רב יותר.
ניתן למצוא את השטח תחת עקום הצפיפות של חי־בריבוע בעזרת תכנה כמו Excel, למשל. עם זאת, כדי להגיע למסקנה לגבי בדיקת השערות, בדרך כלל אפשר להסתפק באחוזונים הרשומים בטבלה. אם נרצה לדייק לגבי מובהקות התוצאה המקורבת בדוגמה זו, ניתן למצוא את השטח המדויק: $P(T \geq 8.5) = .0749$, כאשר $T \sim \chi^2_4$.

דוגמה 5.3 (המשך דוגמה 5.1). נמצא את מובהקות התוצאה $H = 13.95$ שהתקבלה בדוגמה 5.1 (ניסוי בשינוי משך השינה). נשתמש בהתפלגות המקורבת חי־בריבוע. הניסוי נערך לגבי $k = 3$ מדגמים ולכן ההתפלגות המקורבת של H היא חי־בריבוע עם 2 דרגות חופש. לפי טבלת חי־בריבוע, טבלה 5 בנספח, עבור 2 דרגות חופש, ניתן לראות שערך החלוקה הגדול ביותר המצוי בטבלה הוא $\chi^2_{.995} = 10.6$. תוצאת הניסוי שהתקבלה – 13.95 – גדולה אף מערך זה. לפיכך מובהקות התוצאה קטנה מהסתברות הזנב המתאים, כלומר: $P = P_{H_0}(H \geq 13.95) < .005$. היות שהמובהקות קטנה מאוד, עבור כל רמת מובהקות סבירה ניתן להסיק שיש הבדל בין שלוש הקבוצות שנבדקו. מן הערכים של ממוצעי הדרגות בלוח 5.2 נראה שקבוצה א (אלה שהאריכו את שעות השינה) קיבלה את הדרגות הגבוהות ביותר, קבוצה ב (אלה שקיצרו) – את הדרגות הבינוניות, וקבוצה ג (ביקורת) קיבלה את הדרגות הנמוכות. כלומר, ייתכן שהארכת השינה אמנם השפיעה על שיפור מהירות התגובה. עם זאת, לא ברור לנו אם ההבדלים בין כל שתיים מהקבוצות הם מובהקים.
נדון בבעיה זו בחלק הבא.

5.2 השוואות מרובות

אם חוקר דחה את השערת האפס, והסיק ש-k האוכלוסיות כנראה אינן זהות, הוא בדרך כלל מעוניין להשוות בין זוגות של אוכלוסיות כדי לברר אם חלקן נוטות להיות שונות זו מזו. למשל, אם הושוו 4 קבוצות והוחלט שהן שונות, יש עניין להחליט מהי הקבוצה שבה הציון הוא הגבוה ביותר, אם ניתן למצוא קבוצה כזאת העדיפה באופן מובהק על כל השאר.

נניח, למשל, שחוקר שהשווה את ארבע הקבוצות רוצה להשוות בין כל $6 = \binom{4}{2}$ הזוגות של המדגמים. כלומר, הוא רוצה לבדוק את ההשערות:

$$H_0^1 : F_1 \equiv F_2 \ , \ H_0^2 : F_1 \equiv F_3 \ , \ \ldots \ , \ H_0^6 : F_3 \equiv F_4 \qquad (14)$$

הוא יכול, כמובן, לבדוק כל אחת מההשערות הרשומות ב-(14) בעזרת המבחן של ווילקוקסון לשני מדגמים (או מאן-וויטני). הבעיה היא שבצורה כזאת רמת המובהקות של המבחן איננה פשוטה לחישוב. רמת המובהקות (ההסתברות לעשות טעות מסוג I) במקרה של השוואות מרובות היא ההסתברות לדחות לפחות אחת מן ההשערות הרשומות ב-(14), אם, למעשה, השערת האפס הרשומה ב-(1) נכונה (כל ההתפלגויות שוות). נניח שהחוקר בודק את ההבדל בין כל שני מדגמים ברמת מובהקות $\alpha = .05$. ההסתברות שהוא יעשה טעות מסוג I בבדיקה בשיטה זו היא, למעשה, ההסתברות שהוא ידחה לפחות אחת מבין שש ההשערות, אם למעשה כולן נכונות. ברור שהסתברות זו **גבוהה** מרמת המובהקות .05 שנקבעה לכל אחת מהבדיקות הזוגיות בנפרד. למשל, אם 6 המבחנים היו בלתי תלויים, היינו מקבלים את רמת המובהקות של המבחן על ידי

$$\alpha = P_{H_0}(\text{דחייה לפחות באחת מ-6 הבדיקות})$$
$$= 1 - P_{H_0}(\text{שום דחיה ב-6 הבדיקות}) = 1-(1-\alpha)^6$$
$$= 1-(.95)^6 = .2649$$

זאת אומרת, בהנחה של אי־תלות בין הבדיקות הנערכות, ההסתברות לטעות מסוג I גדולה בהרבה מההסתברות הנדרשת של 0.05. מובן שבמקרה שבו אנו דנים כאן הבדיקות אינן בלתי תלויות, ולכן ההסתברות לטעות איננה ניתנת לחישוב פשוט. ישנן שיטות שונות לעריכת השוואות זוגיות, כדי להבטיח שרמת המובהקות תִּשָּׁמר ולא תעלה על הגודל α שנקבע מראש. נביא כאן שיטה אחת בלבד, והיא מתאימה לבעיה של השוואות מרובות בכל מודל הסתברותי.

שיטת בונפרוני להשוואות מרובות

נניח שאנו רוצים לערוך m מבחנים. נסתכל על משפחת m ההשערות הנבדקות. – H_0^i, $i=1,\ldots,m$. רמת המובהקות במשפחה היא ההסתברות לעשות לפחות טעות אחת מסוג I בין כל m הבדיקות. הסתברות זו מסומנת FWE (family-wise error). נרצה לדאוג לכך שרמת המובהקות הזאת לא תעלה על הגודל הנדרש α. במִלים אחרות, אנו דורשים שההסתברות לדחות בטעות לפחות אחת מ-m ההשערות לא תעלה על α, במקרה שכל m ההשערות נכונות. השיטה הנקראת "שיטת בונפרוני" מסתמכת על אי־שוויון בונפרוני, שבעזרתו ניתן לרשום

$$P\{A_1 \cup A_2 \cup \ldots \cup A_m\} \leq P(A_1)+P(A_2)+\cdots+P(A_m) \qquad (15)$$

עבור כל m מאורעות.

משמעות האי־שוויון (15) היא שההסתברות לכך שלפחות אחד מבין m מאורעות יקרה, לעולם אינה עולה על סכום ההסתברויות של m המאורעות הבודדים.

לא נוכיח אי-שוויון זה כאן.

לגבי בעיית בדיקת ההשערות המרובות, נסמן את המאורעות: A_i – ההשערה H_0^i נדחית, $i=1,\ldots,m$, ונניח שאת המבחן לבדיקת H_0^i אנו עורכים ברמת מובהקות α_i.

רמת המובהקות במשפחה כולה, FWE, היא ההסתברות שיקרה לפחות אחד מבין המאורעות $A_1,\ldots,A_m$, בהנחה שכל m ההשערות הנבדקות הן נכונות. לפי אי-שוויון (15) ישנו חסם פשוט על הסתברות זו:

$$\text{FWE}=P_{H_0^1\cap\ldots\cap H_0^m}\left(A_1\cup\ldots\cup A_m\right) \qquad (16)$$
$$\leq P_{H_0^1}(A_1)+\ldots+P_{H_0^m}(A_m)=\alpha_1+\cdots+\alpha_m$$

מכאן שרמת המובהקות במשפחה (FWE) אינה עולה על סכום רמות המובהקות הספציפיות ב-m המבחנים הבודדים. על סמך הנוסחה (16) מקבלים את השיטה של בונפרוני להשוואות מרובות. בשיטה זו איננו יכולים להבטיח שהסתברות הטעות תהיה בדיוק α, אלא רק שהיא לא תעלה על α. במִלים אחרות, מקבלים חסם עליון עבור רמת המובהקות במשפחת m המבחנים.

שיטת בונפרוני. כדי שרמת המובהקות במשפחת m המבחנים לא תעלה על α, נבחר את רמות המובהקות α_i של ההשוואות הספציפיות, באופן שסכומן יהיה שווה ל-α. כלומר, $\alpha_1+\cdots+\alpha_m=\alpha$.

בדרך כלל מקובל לבחור רמות מובהקות שוות לכל המבחנים, זאת אומרת

$$\alpha_i=\frac{\alpha}{m} \qquad i=1,\ldots,m \qquad (17)$$

לפי אי־שוויון (16) מובטח בכך שיתקיים $\text{FWE}\leq\alpha$.

השימוש בשיטת בונפרוני להשוואות זוגיות

בבעיית השוואות זוגיות של k אוכלוסיות, מספר ההשערות הנבדקות הוא $m=\binom{k}{2}$ ולפי שיטת בונפרוני יש לערוך כל אחת מההשוואות הזוגיות ברמת מובהקות

$$\alpha_i=\frac{\alpha}{\binom{k}{2}} \qquad (18)$$

למשל, אם רוצים לערוך השוואות זוגיות בין ארבע אוכלוסיות, עם הסתברות לטעות מסוג I שלא תעלה על 0.05, אזי $k=4$, $\binom{k}{2}=6$ וכל השוואה תיערך באמצעות מבחן ווילקוקסון (או מאן-וויטני) **דו־צדדי** ברמת מובהקות $\alpha_i=0.05/6=.0083$. במקרה כזה מחליטים על השוואה בודדת כמובהקת, רק אם מובהקות התוצאה של השוואה זו אינה עולה על 0.0083.

שימו לב שרמת המובהקות הספציפית שבה אנו משתמשים במבחן בודד להשוואת שתי

אוכלוסיות, היא קטנה מאוד. ככל שתרצו לערוך יותר השוואות, תצטרכו להשתמש ברמת מובהקות קטנה יותר עבור כל השוואה בודדת.

הערה: כל אחד מהמבחנים הספציפיים להשוואה זוגית הוא מבחן דו־צדדי, כיוון שבמקור לא הייתה השערה אלטרנטיבית כיוונית לגבי היחס בין k האוכלוסיות.

דוגמה 5.4 (המשך דוגמה 5.1). נסתכל שוב על הנתונים בדוגמה 5.1 (לוח 5.2) ונרשום בלוח 5.4 רק את הדרגות המתאימות. בלוח 5.4 הדרגות בכל קבוצה כבר מסודרות לפי גודלן. מאחר שבמבחן קרוסקל-וואליס שערכנו קיבלנו הבדל מובהק בין שלוש קבוצות הניסוי, נרצה לנסות לברר איזה מהקבוצות השתפרה באופן מובהק יותר מהאחרות. נערוך את ההשוואות הזוגיות באופן שנשמור על רמת מובהקות כללית FWE של $\alpha = .10$.

לוח 5.4. *הדרגות של נתוני דוגמה 5.1 (הארכה וקיצור של שעות שינה)*

א	ב	ג	
1	3	8	
2	6	17	
4	11	21	
5	12	22	
7	13	25	
9	15	26	
10	18	27	
14	20	28	
16	23	29	
19	24	30	
8.7	14.5	23.3	$\overline{R}_i$

את ההשוואות הזוגיות נוח יותר לעשות בעזרת הסטטיסטי של מאן-וויטני מאשר בעזרת הסטטיסטי של ווילקוקסון, כיוון שהשימוש בסטטיסטי של מאן-וויטני אינו מחייב לדרג מחדש את התצפיות הכלולות בכל זוג של מדגמים. נערוך, אם כך, את 3 ההשוואות הזוגיות (להשוואת כל זוג של קבוצות ניסוי) בעזרת מבחני מאן-וויטני, ברמת מובהקות $\alpha_i = 0.10/3 = 0.0333$ כל אחד. היות שהמבחנים הם דו־צדדיים, כדי שתוצאת מבחן כלשהו תהיה מובהקת, דרוש שהסתברות הזנב שהתקבל לא תעלה על $\alpha_i/2 = 0.0333/2 = .0167$.

את הערכים של W_{xy} להשוואת שלושת הזוגות של הקבוצות נוח לרשום בְּטבלה, לוח 5.5 (בדקו!). טבלה כזאת מאוד יעילה כאשר יש להשוות מספר גדול יחסית של מדגמים. לדוגמה, בהשוואה בין קבוצה א (x) לקבוצה ב (y) על ידי הסטטיסטי של מאן-וויטני התקבלה התוצאה $W_{xy}=\#\{x_i<y_j\}=16$. תמצאו ערך זה בלוח 5.5 במקום המתאים.

לוח 5.5. *ערכי W_{xy} להשוואת כל שתי קבוצות ניסוי*

x	y: א	ב	ג
א (האריכו)	–	16*	6*
ב (קיצרו)		–	26
ג (ביקורת)			–

* הבדל מובהק

הערך $W_{xy}=16$ נמצא בזנב השמאלי של ההתפלגות ולכן הסתברות הזנב המתאימה היא $P(W_{xy}\le 16)$.

בחישוב הסתברות ה"זנב" לגבי כל אחת מהתוצאות צריך לקחת בחשבון את הכיוון שבו נפלה התוצאה (אם נפלה בזנב הימני או השמאלי של התפלגות W_{xy}). קל לבדוק זאת על פי התוחלת. באופן כללי $EW_{xy}=mn/2$. ניתן לחשב את מובהקות התוצאה עבור כל אחת מההשוואות כדי לבדוק היכן הסתברות הזנב קטנה מ-0.0167. מצד שני, היות שכל הקבוצות בדוגמה זו הן באותו גודל ($n_i=10$), יותר קל למצוא את הערכים הקריטיים שעבורם הסתברות הזנב אינה עולה על הערך הנדרש 0.0167. טבלה 2 בנספח נותנת את התפלגות הסטטיסטי של ווילקוקסון W_s ולכן יש לעבור לערכים המתאימים של W_{xy}. לגבי הזנב השמאלי: על פי הטבלה ($m=n=10$), $P(W_s\le 76)=.0144$ בעוד ש-$P(W_s\le 77)=.0177$. לפיכך הערך 77 גדול מדי ולכן יש לבחור את הערך הקריטי $W_s=76$. הערך המתאים של הסטטיסטי של מאן-וויטני הוא

$$W_{xy}=W_s-\frac{n(n+1)}{2}=76-\frac{10(11)}{2}=21$$

את הזנב הימני ניתן למצוא בעזרת סימטרייה. הערך הסימטרי של $W_{xy}=21$ הוא $W_{xy}=10\cdot 10-21=79$. שיטת ההחלטה לגבי ההשוואות המרובות: השוואה שבה יתקבל $W_{xy}\le 21$ או $W_{xy}\ge 79$ היא מובהקת.

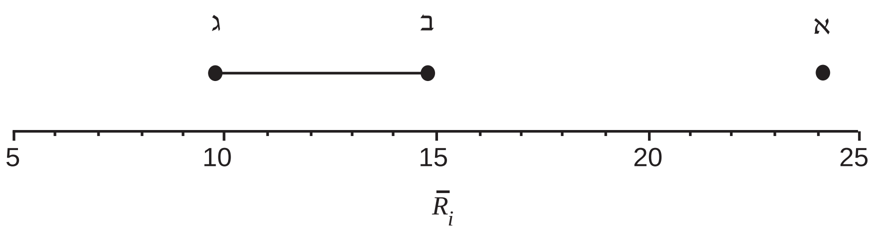
מסקנה: בלוח 5.5 ישנן שתי תוצאות מובהקות (שתיהן בזנב השמאלי). אלה ההשוואות בין קבוצה א לקבוצה ב ובין קבוצה א לקבוצה ג. תוצאות אלה מסומנות בכוכבית.

מסקנה מתוצאות ניסוי זה היא שהתלמידים שהאריכו את שעות השינה שיפרו את הביצועים לעומת שתי הקבוצות האחרות (אלה שקיצרו ואלה שלא שינו את משך השינה). בין אלה שקיצרו את שעות השינה לבין קבוצת הביקורת אין הבדל מובהק. ניתן, אפוא, להסיק שכדאי לנסות להאריך את שעות השינה של תלמידים כדי לשפר את התפקוד הקוגניטיבי שלהם.

תיאור תוצאות ההשוואות הזוגיות אופן גרפי

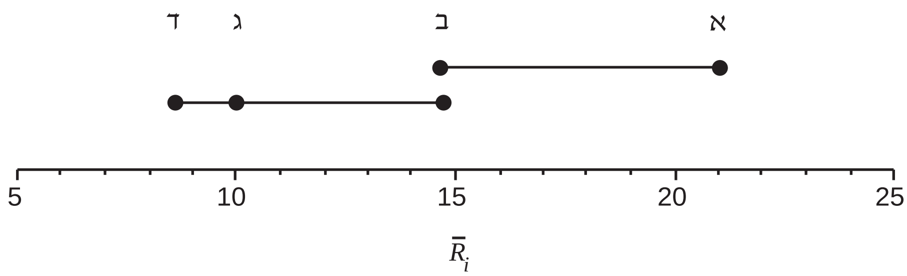
כאשר יש הרבה השוואות, נוח להבין את הממצאים באמצעות תיאור גרפי מתאים. נתאר כאן את התוצאות שקיבלנו בניסוי השינה. אנו רושמים על ציר ה-x סקלה של הדרגות. מעל לציר ה-x אנו מציינים בסימן מיוחד (כאן – עיגול שחור) את מיקום הדרגה הממוצעת של כל אחד מהמדגמים. מחברים על ידי קו מאוזן את ציוני הדרגות הממוצעות שביניהן **לא נמצא** הבדל מובהק.

ציור 5.3 מדגים את התיאור הזה עבור דוגמה 5.4. חיברנו בקו את ממוצעי קבוצה ב וקבוצה ג (שביניהם אין הבדל מובהק). לא חיברנו את הממוצעים של א ו-ב ולא את אלה של א ו-ג (שביניהם התקבל הבדל מובהק).



ציור 5.3. *תוצאות ההשוואות הזוגיות בדוגמה 5.4*

לעתים התוצאות של השוואות מרובות אינן כה ברורות כמו אלה שהתקבלו בדוגמת השינה. ציור 5.4 הוא דוגמה פיקטיבית לתיאור גרפי של ממצאי השוואות זוגיות עבור $k=4$ מדגמים.



ציור 5.4. *תוצאות ההשוואות הזוגיות בדוגמה פיקטיבית*

לפי הציור, קבוצה א הייתה גבוהה באופן מובהק מקבוצות ג ו-ד, אך לא מקבוצה ב. כמו כן, בין שלוש הקבוצות ב, ג ו-ד לא נמצאו הבדלים מובהקים.

הערה 1. שימו לב שבציור 5.4 נתקבלה לכאורה סתירה בין הממצאים. מצד אחד אין הבדל בין קבוצה א ל-ב, כמו כן אין הבדל בין קבוצה ב ל-ג, אבל יש הבדל בין א ל-ג (תוצאות דומות התקבלו גם עבור קבוצה ד). הסיבה לכך היא שהמסקנה של בדיקת ההשערה הסטטיסטית על שוויון התפלגויות איננה מוחלטת. כשאנו לא דוחים את השערת האפס, אין זה אומר ששתי ההתפלגויות בדיוק שוות, אלא רק שאי אפשר לדחות את השערת השוויון על בסיס הנתונים הקיימים. במקרה הנידון, ההבדל בין הדרגה הממוצעת של קבוצה א לזו של קבוצה ב לא היה די גדול, אבל בהשוואה של קבוצה א לקבוצה ג ההבדל היה גדול מספיק כדי להוות ראָיה לכך שכנראה קיים הבדל בין שתי האוכלוסיות הללו.

הערה 2. בגלל הדרישה החמורה מאוד של שמירה על רמת מובהקות נמוכה במשפחת ההשערות הנבדקות, רמת המובהקות של כל אחד מהמבחנים הבודדים חייבת להיות קטנה מאוד. מסיבה זו קורה לעתים שההשערה הכללית של שוויון כל k ההתפלגויות נדחית בהסתמך על המבחן של קרוסקל-וואליס, אבל אף אחת מההשוואות הזוגיות איננה מובהקת. כלומר, קורה שלמרות שיש, כנראה, הבדלים בין ההתפלגויות, איננו יכולים לאתר אותם, אם משתמשים בשיטה של השוואות מרובות עם שליטה על רמת המובהקות "המשפחתית" FWE.

5.3 בעיות של ערכי תיקו במבחן קרוסקל-וואליס

במקרה שההתפלגויות אינן רציפות, ומתקבלים ערכי תיקו, מדרגים את התצפיות כמו במקרים הקודמים – ערכים שווים מקבלים דרגות שוות, השוות לדרגה הממוצעת המגיעה להם.

נסמן: $\tilde{R}_{ij}$ – הדרגה (הממוצעת) של התצפית X_{ij} , $i=1,\ldots,k;\ j=1,\ldots,n_i$

$\tilde{R}_i=\sum_{j=1}^{n_i}\tilde{R}_{ij}$ – סכום הדרגות הממוצעות במדגם ה-i, $i=1,\ldots,k$.

כפי שהראינו בנוסחה (11), עבור המקרה שבו אין תיקו, הסטטיסטי H ניתן להירשם על ידי $H=SSB/MST$. באותו אופן נוכל לחשב אותו גם כאשר יש ערכי תיקו בניסוי. סכומי הריבועים המתאימים הם

$$\tilde{SSB}=\sum_{i=1}^{k}\frac{1}{n_i}\left[\tilde{R}_i-\frac{n_i(N+1)}{2}\right]^2 \qquad (19)$$

$$\tilde{MST}=\frac{1}{N-1}\sum_{i=1}^{k}\sum_{j=1}^{n}\left[\tilde{R}_{ij}-\frac{N+1}{2}\right]^2 \qquad (20)$$

את שני סכומי הריבועים לעיל ניתן לחשב על סמך הדרגות הממוצעות שהתקבלו. עם זאת, אפשר להראות שהערך של $\tilde{MST}$ תלוי אך ורק בגודלן של קבוצות התיקו.

טענה 5.2. נסמן ב- $\tilde{H}$ את הסטטיסטי המתקבל על ידי המנה $\tilde{H}=\frac{\tilde{SSB}}{\tilde{MST}}$. אזי

$$\tilde{H}=\frac{H}{\left[1-\frac{1}{N(N^2-1)}\sum_l t_l(t_l^2-1)\right]} \qquad (21)$$

כאשר $H=\frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{1}{n_i}\left[\tilde{R}_i-\frac{n_i(N+1)}{2}\right]^2$

הוכחה: נסתכל על קבוצת כל N הדרגות הממוצעות שהתקבלו בניסוי. נסמן ב- $\tilde{U}$ את המשתנה המקבל כל אחת מ-N הדרגות בקבוצה זו (חלקן שוות) בהסתברות שווה (השווה ל-$1/N$). בנוסחה (20) רשום סכום ריבועי הסטיות של הערכים בקבוצת הדרגות לעיל מהממוצע שלהם (הממוצע נשאר כפי שהיה ללא ערכי תיקו). מכאן הביטוי MST הוא למעשה השונות של המשתנה $\tilde{U}$, הכופלת בקבוע $N/(N-1)$. כלומר, במקרה של תיקו מקבלים

$$\tilde{MST}=\frac{N}{N-1}Var(\tilde{U}) \qquad (22)$$

את השונות של המשתנה $\tilde{U}$ כבר חישבנו בפרק 2, שם קיבלנו את נוסחה (29):

$$Var(\tilde{U})=Var(U)-\frac{1}{12N}\sum_l t_l(t_l^2-1)=\frac{N^2-1}{12}-\frac{1}{12N}\sum_l t_l(t_l^2-1) \qquad (23)$$

כאשר $U\sim U(1,N)$ ו-t_l הוא מספר האיברים בקבוצת התיקו ה-l.

נציב את הנוסחה (23) ב-(22) ונקבל את MST במקרה תיקו:

$$\tilde{MST}=\frac{N}{N-1}Var(\tilde{U})=\frac{N(N+1)}{12}\left[1-\frac{1}{N(N^2-1)}\sum_l t_l(t_l^2-1)\right]$$

הסטטיסטי $\tilde{H}$ הוא

$$\tilde{H}=\frac{\tilde{SSB}}{\tilde{MST}}=\frac{\sum_{i=1}^{k}\frac{1}{n_i}\left[\tilde{R}_i-\frac{n_i(N+1)}{2}\right]^2}{\frac{N(N+1)}{12}\left[1-\frac{1}{N(N^2-1)}\sum_l t_l(t_l^2-1)\right]}$$

$$= \frac{H}{\left[1 - \frac{1}{N(N^2-1)} \sum_l t_l (t_l^2 - 1)\right]}$$

ובכך הוכחנו את הטענה. ♣

דוגמה 5.5. מחקר נערך כדי להעריך את האפקט של אינפורמציה לגבי הביצוע הנדרש במשימה של ייצור חוזר. המשימה הייתה ללטש פיסת מתכת לגודל וצורה מסוימים. עשרים ושניים פועלים במפעל השתתפו בניסוי. הנבדקים בקבוצת הביקורת A לא קיבלו כל אינפורמציה לגבי התוצרת שלהם, בקבוצה B הנבדקים קיבלו אומדן גס לגבי התוצרת, והנבדקים בקבוצה C קיבלו אינפורמציה מדויקת לגבי התוצרת ויכלו לבדוק את עבודתם על ידי השוואה לסרטוט מתאים. הממצאים בלוח 5.6 הם מספר הפיסות שיוצרו על ידי הפועלים בשלוש הקבוצות בתקופת המחקר. הלוח כולל גם את הדרגות של כל 22 התצפיות. גודלי המדגם הם $n_1 = 8$, $n_2 = n_3 = 7$.

לוח 5.6 *מספר הפיסות שיוצרו על ידי הפועלים בשלוש קבוצות הניסוי (דוגמה 5.5)*

	C		B		A	
	x	דרגה	x	דרגה	x	דרגה
	41	8.5	38	4.5	35	1
	43	12.5	39	6	36	2
	44	15.5	42	10	37	3
	45	17.5	43	12.5	38	4.5
	46	19	43	12.5	40	7
	47	20.5	45	17.5	41	8.5
	47	20.5	48	22	43	12.5
					44	15.5
R_i		114		85		54
n_i		7		7		8
$\bar{R}_i$		16.29		12.14		6.75

חישוב הסטטיסטי H נעשה לפי הנוסחה (12), הנכונה גם עבור דרגות ממוצעות:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{\tilde{R}_i^2}{n_i} - 3(N+1)$$

$$= \frac{12}{22 \cdot 23} \cdot \left[\frac{1}{8} \cdot 54^2 + \frac{1}{7} \cdot 85^2 + \frac{1}{7} \cdot 114^2\right] - 3(23) = 8.151$$

יש כאן חמש קבוצות תיקו בגודל $t=2$ וקבוצה אחת בגודל $t=4$. נחשב את הביטוי הקשור בקבוצות התיקו הללו בנוסחה (21).

$$\sum_l t_l(t_l^2-1)=5[2(2^2-1)]+4(4^2-1)=90$$

המכנה של $\tilde{H}$ הוא

$$1-\frac{1}{N(N^2-1)}\sum_l t_l(t_l^2-1)=1-\frac{90}{22(22^2-1)}=0.9915$$

התיקון הוא, אפוא, לא כל כך משמעותי. הסטטיסטי המתוקן המתקבל הוא

$$\tilde{H}=\frac{H}{\left[1-\frac{1}{N(N^2-1)}\sum_l t_l(t_l^2-1)\right]}=\frac{8.151}{0.9915}=8.221$$

לפי טבלת χ_2^2 מובהקות התוצאה מקיימת $.01<P<.025$ (בדקו!). [לפי ההתפלגות המדויקת של χ_2^2, המובהקות היא $P\cong P\{\chi_2^2\geq 8.221\}=.0164$.] נוכל להסיק שעבור רמת מובהקות $\alpha=.05$ התוצאה מובהקת. על סמך הניסוי הזה ניתן להניח שיש הבדל בין שיטות העבודה.

כדי לערוך השוואות זוגיות בין השיטות הללו, נשתמש בשיטת בונפרוני, עם $\alpha=.05$. בסך הכול יש לערוך כאן 3 השוואות ולכן רמת המובהקות לכל השוואה היא $\alpha_i=.05/3=.0167$, $i=1,2,3$. את ההשוואות נעשה בעזרת מבחני מאן-וויטני ונשתמש בקירוב הנורמלי לקביעת מובהקות התוצאה. מובהקויות מדויקות אינן ניתנות לחישוב על פי טבלה 2 בגלל ערכי התיקו.

(1) השוואת הקבוצות A ו-B: כאן $m=8,\ n=7$ והסטטיסטי $W_{xy}=42.5$.

התוחלת והשונות הן

$$EW_{xy}=mn/2=28$$

$$Var(W_{xy})=\frac{mn(N+1)}{12}-\frac{nm}{12N(N-1)}\sum_r t_r(t_r^2-1)$$

$$=\frac{8(7)(16)}{12}-\frac{8(7)}{12(15)(14)}(30)=74.667-0.667=74.0$$

ערכי התיקו חושבו רק בשתי הקבוצות A ו-B.

המובהקות המקורבת של התוצאה היא

$$P=P(W_{xy}\geq 42.5)\cong 1-\Phi\left(\frac{42-28}{\sqrt{74}}\right)=1-\Phi(1.63)=.0516$$

התוצאה אינה מובהקת.

(2) השוואת הקבוצות A ו-C: גם כאן $m=8,\ n=7$ והסטטיסטי $W_{xy}=51.5$.

התוחלת זהה למקרה הקודם והשונות היא

$$Var(W_{xy})=\frac{8(7)(16)}{12}-\frac{8(7)}{12(15)(14)}(24)=74.134$$

ערכי התיקו חושבו רק בשתי הקבוצות A ו-C.

המובהקות המקורבת של התוצאה היא

$$P=P(W_{xy}\geq 51.5)\cong 1-\Phi\left(\frac{51-28}{\sqrt{74.134}}\right)=1-\Phi(2.67)=.0038$$

מובהקות זו קטנה מ-0.0167 ולכן התוצאה מובהקת. הציונים בקבוצה C גבוהים מאלה בקבוצה A.

(3) השוואת הקבוצות B ו-C: כאן $m=7,\ n=7$ והסטטיסטי $W_{xy}=34.5$.

התוחלת היא $EW_{xy}=mn/2=24.5$ והשונות היא

$$Var(W_{xy})=\frac{7(7)(15)}{12}-\frac{7(6)}{12(14)(13)}(36)=60.558$$

ערכי התיקו חושבו רק בשתי הקבוצות B ו-C.

המובהקות המקורבת של התוצאה היא

$$P=P(W_{xy}\geq 34.5)\cong 1-\Phi\left(\frac{34-24.5}{\sqrt{60.558}}\right)=1-\Phi(1.22)=.1112$$

תוצאה זו אינה מובהקת.

לסיכום, שיטה C (אינפורמציה מדויקת) עדיפה על שיטה A (ללא כל אינפורמציה). לעומת זאת אין הבדל בין אינפורמציה חלקית (B) לבין חוסר אינפורמציה (A) או אינפורמציה מלאה (C).

5.4 מבחן יונקירי לאלטרנטיבה סדורה

הסטטיסטי של קרוסקל-וואליס איננו רגיש לצורה מסוימת של אלטרנטיבה, אלא הוא מיועד לבחון הבדלים כלשהם בין k ההתפלגויות. אם, לעומת זאת, ההשערה האלטרנטיבית היא בכיוון מסוים, אזי ניתן להציע סטטיסטי שונה שיהיה רגיש בדיוק לפערים בכיוון המשוער. לדוגמה, אם יש צורך לבדוק כיצד משפיעה העלאה בכמות ההשקיה על היבול של עגבניות, ייתכן שההשערה האלטרנטיבית תהיה שככל שכמות המים גבוהה, כך היבול נוטה לעלות. השערה כזאת נקראת **אלטרנטיבה סדורה**. באופן תֵאורטי ניתן

לרשום השערה כזאת על ידי

$$H_1: F_1(t) \geq F_2(t) \geq \ldots \geq F_k(t) \qquad \text{לכל } t,$$

כאשר F_i היא ההתפלגות של X_{ij} – התצפית ה-j מהטיפול i.

כלומר, האוכלוסיות גדולות סטוכסטית זו מזו לפי הסדר הרשום. אם המודל הוא של הזזה, ההשערה האלטרנטיבית ניתנת להירשם, למשל, בעזרת החציונים $M_1, M_2, \ldots, M_k$ של k ההתפלגויות:

$$H_1: M_1 \leq M_2 \leq \ldots \leq M_k$$

הסטטיסטי של יונקירי

יהי W_{il} הסטטיסטי של מאן-וויטני להשוואת אוכלוסייה i לאוכלוסייה l, $i < l$, כלומר:

$$W_{il} = \#\{(r,s) : X_{ir} < X_{ls}\}$$

סטטיסטי המבחן של יונקירי (Jonckheere, 1954) מוגדר על ידי סכום הסטטיסטים לעיל:

$$J = \sum_{i=1}^{k} \sum_{l=1}^{i-1} W_{il} \tag{24}$$

בהתאם לאלטרנטיבה הסדורה, דוחים את השערת האפס עבור ערכים גבוהים של J.

קיימות טבלאות של ההתפלגות, אולם ניתָן להשתמש בקירוב נורמלי כאשר המדגמים גדולים. לשם כך יש למצוא את התוחלת והשונות של הסטטיסטי תחת השערת האפס (שוויון כל k ההתפלגויות).

חישוב התוחלת והשונות של J תחת השערת האפס

התוחלת והשונות של הסטטיסטי של יונקירי מובאות במשפט הבא.

משפט 5.3. תחת השערת האפס התוחלת של הסטטיסטי של יונקירי היא

$$EJ = \frac{1}{4}\left[N^2 - \sum_{i=1}^{k} n_i^2\right] \tag{25}$$

והשונות היא

$$Var(J) = \frac{1}{72}\left[N^2(2N+3) - \sum_{i=1}^{k} n_i^2(2n_i+3)\right] \tag{26}$$

הוכחה: התוחלת של כל אחד מהמחוברים בנוסחה (24) היא תוחלת סטטיסטי של מאן-וויטני, השווה $EW_{il} = \frac{n_i n_l}{2}$. תוחלת הסכום שווה לסכום התוחלות:

$$EJ = E\sum_{i=1}^{k}\sum_{l=1}^{i-1} W_{il} = \sum_{i=1}^{k}\sum_{l<i} \frac{n_i n_l}{2} = \frac{1}{4}\sum_{i=1}^{k}\sum_{l \neq i} n_i n_l$$

$$= \frac{1}{4}\left[\sum_{i=1}^{k}\sum_{l=1}^{k} n_i n_l - \sum_{i=1}^{k} n_i^2\right] = \frac{1}{4}\left[N^2 - \sum_{i=1}^{k} n_i^2\right]$$

חישוב השונות קצת יותר מורכב והוא מובא בנספח 5. ♣

דוגמה 5.6. כדי לבחון את השפעתו האפשרית של חומר מסוג של סם קל על התנועתיות, ניתן חומר זה במינונים שונים ל-4 קבוצות של חולדות. ההשערה הייתה שמינון גבוה מוריד את התנועתיות. בלוח 5.7 נתוני השינויים בתנועתיות של החולדות, בכל אחד מהמינונים השונים. בלוח מוצגות גם הדרגות של כל התצפיות.

לפי התֵאוריה של החוקרים, התנועתיות צפויה לרדת ככל שהמינון גבוה. לכן את הסטטיסטים המתאימים של מאן-וויטני יש לחשב בכיוון המתאים, וזו הסיבה שרשמנו את הקבוצות בלוח 5.7 משמאל לימין, בכיוון המשוער.

לוח 5.7. *השינוי בתנועתיות של חולדות בארבעה מינונים של סם*

0.1 מ"ג		0.5 מ"ג		1 מ"ג		2 מ"ג	
דרגה	ציון	דרגה	ציון	דרגה	ציון	דרגה	ציון
3	5.6	11	7.6	5	6.4	1	3.4
12	8.4	16	9.9	6	6.5	2	4.8
13	8.8	21	11.8	8	7	4	5.7
15	9.6	24	12.2	9	7.1	7	6.8
18	10.5	26	13	14	9.1	10	7.2
23	12	27	13.5	20	11.5	17	10.1
25	12.6	28	14.2	29	14.7	19	10.6
32	18.6	30	14.8	31	15.3	22	11.9

ערכי הסטטיסטים של מאן-וויטני רשומים בלוח 5.8 וסכומם הוא $J = 250$.

לוח 5.8. ערכי W_{xy} *להשוואת כל 6 הזוגות של מינונים*

	y		
0.1 מ"ג	0.5 מ"ג	1 מ"ג	x
47	57	42	2 מ"ג
38	46		1 מ"ג
20			0.5 מ"ג

את המובהקות נמצא בעזרת הקירוב הנורמלי. גודלי המדגם כאן הם $n_i=8$, $i=1,\ldots,4$ וכך $N=32$. התוחלת היא, לפי נוסחה (25)

$$EJ=\frac{1}{4}\left[N^2-\sum_{i=1}^{k}n_i^2\right]=\frac{1}{4}\left[32^2-4(8^2)\right]=192$$

השונות על פי נוסחה (26):

$$Var(J)=\frac{1}{72}\left[N^2(2N+3)-\sum_{i=1}^{k}n_i^2(2n_i+3)\right]$$
$$=\frac{1}{72}\left[32^2(67)-4(64\cdot 19)\right]=885.33$$

מכאן מובהקות התוצאה (דחייה עבור ערכים גבוהים):

$$P=P_{H_0}(J\geq 250)\cong 1-\Phi\left(\frac{249.5-192}{\sqrt{885.33}}\right)=1-\Phi(1.93)=.0268$$

עבור רמת מובהקות של 5% יש לדחות את השערת האפס ולהסיק כי אמנם ישנה ירידה בתנועתיות כשהמינון עולה.

הערה: עיון מעמיק יותר בתוצאות בלוח 5.8 מצביע על כך שייתכן שהתֵאוריה אינה מדויקת. התוחלת של הסטטיסטי של מאן-וויטני להשוואת שתי קבוצות בנות $n=8$ כל אחת היא $EW_{xy}=8\cdot 8/2=32$. נשים לב שההשוואה בין מינון של 0.5 מ"ג לבין 0.1 מ"ג נתנה תוצאה הפוכה מן הצפוי (ערך נמוך מן התוחלת). מכאן נובע שייתכן כי למעשה עבור מינונים נמוכים התנועתיות עולה עם העלייה במינון ועבור מינונים גבוהים יותר היא מתחילה לרדת. התוצאה הכוללת של מבחן יונקירי התקבלה מובהקת בכיוון המשוער למרות התוצאה ההפוכה לעיל, מכיוון ששאר ההשוואות היו כולן בכיוון המתאים לתֵאוריה, ובסך הכול התקבל ערך יחסית גבוה של J .

תרגילים

1. הוכיחו את נוסחה (12) לחישוב H.
2. חשבו על ידי מנִייה את ההתפלגות המדויקת של הסטטיסטי H של קרוסקל־וואליס, תחת H_0, עבור המקרים הבאים ($k=3$):
   א) $n_1=1,\ n_2=1,\ n_3=4$; ב) $n_1=n_2=n_3=2$.
3. להשוואת ארבעה טיפולים לשם הורדה במשקל, חולקו אקראית 24 גברים בין

הטיפולים. התוצאות הרשומות הן אחוזי הירידה במשקל של הגברים שנבדקו.

הטיפול			
א	ב	ג	ד
19.3	21.7	18.7	20.2
18.0	19.6	17.3	18.5
17.8	19.4	16.1	17.7
16.5	18.8	14.6	17.2
15.0	18.3	13.9	17.0
12.5	17.5	12.2	15.7

א) בְּדקו אם יש הבדלים מובהקים בין הטיפולים ($\alpha = .05$).

ב) עִרכו השוואות זוגיות עם $\alpha = .06$ לקביעת הסדר בין הטיפולים מבחינת יעילותם להורדת המשקל.

4. בטבלה נתונות היחידות, בספרה השניה והשלישית אחרי הנקודה העשרונית, של קביעת קבוע הגרביטציה G על ידי Heyl בשנת 1930 (כלומר, נתון של 83 מתאים לתוצאת ניסוי של 6.683). החוקר השתמש בכדורים משלושה חומרים שונים.

זהב	72	79	78	76	81	83
פלטינה	64	67	67	61	61	
זכוכית	74	72	75	71	78	

קִבעו את המובהקות המקורבת של התוצאה לבדיקת ההבדל בקבוע הגרביטציה והסיקו אם יש הבדל בין החומרים ($\alpha = .05$). במידת הצורך עִרכו השוואות זוגיות.

5. הוכיחו כי עבור שני מדגמים ($k = 2$) מבחן קרוסקל־וואליס אקוויוולנטי למבחן ווילקוקסון לשני מדגמים הבודק בעיה דו־צדדית.

כלומר: ראשית הראו את הקשר בין הסטטיסטי H לבין הסטטיסטי W_s לאחר התקנון המתאים עבור קירוב נורמלי. שנית, בדקו כי הערכים הקריטים הם זהים.

עִרכו את הבדיקה עבור רמות המובהקות $\alpha = .01, .05, .10$. (הסבירו!)

הסיקו מכאן על השקילות בין שני המבחנים גם לגבי חישוב המובהקות המקורבת למדגמים גדולים.

6. הוכיחו את הקשר בין הסטטיסטי H לסטטיסטי F של ניתוח שונות חד־כיווני:

$$\frac{1}{F} = \frac{k-1}{N-k}\left[\frac{N-1}{H} - 1\right]$$

כאשר $F = \dfrac{SSB/(k-1)}{(SST - SSB)/(N-k)}$.

7. במחקר להשוואת האפקט של הצגות שונות של עדויות, נעשו שלושה סרטים שהציגו משפט פלילי. הסרטים הראו את אותן העובדות, אלא שהם נערכו בצורות שונות. כל סרט הוצג בפני קבוצה שונה של 30 איש. הצופים התבקשו להחליט לגבי אשמתו של הנאשם: "אשם", "לא בטוח", או "לא אשם". תוצאות הניסוי מובאות בלוח השכיחויות להלן. האם יש הבדל בין הצורות השונות של עריכת הסרט לגבי ההחלטה על האשמה? מה מובהקות התוצאה? אם יש צורך, ערכו השוואות זוגיות.

	לא אשם	לא בטוח	אשם
סרט א'	10	10	10
סרט ב'	10	8	12
סרט ג'	22	6	2

8. הסתכלו על נתוני דוגמה 5.5. השתמשו במבחן יונקירי כדי לבדוק אם תוספת אינפורמציה משפרת את יעילות העבודה. תנו את מובהקות התוצאה והסיקו עבור $\alpha=.05$.

פרק 6

מדדי קשר בין שני משתנים, מקדמי מתאם

עד עתה הדיון בכל מקום היה לגבי משתנים חד־ממדיים בלבד, כאשר הבעיות שבהן דנו היו לגבי מיקום ההתפלגויות, השוואתן מבחינת המיקום והפיזור ורווחי־סמך עבור המיקום. אולם במקרים רבים בעיות מחקריות עוסקות בייתכנות של קשר בין משתנים שונים. קיום של קשר בין משתנים מסביר תֵאוריות בנושא הנידון ועוזר לנבא או להעריך ערכים של משתנה מסוים בעזרת משתנים אחרים. למשל, רוצים להעריך עד כמה (אם בכלל) יש קשר בין עישון לבין סרטן הריאות; בין ציון בבחינה הפסיכומטרית לבין מידת הצלחה בלימודים באוניברסיטה; או בין הטמפרטורה של התנור במפעל לבין מידת הקשיחות של המוצר. בדומה לכך, רוצים לדעת מהם המשתנים המנבאים בהצלחה את מזג האוויר.

כאן נדון רק בקשר בין שני משתנים בלבד.

לבדיקת הקשר בין שני משתנים מסתכלים על מדגם של n תצפיות של זוגות (X_i,Y_i), $i=1,\ldots,n$. מניחים שהזוגות במדגם הם בלתי תלויים זה בזה, בעלי אותה התפלגות משותפת.

השערת האפס בבעיות כאלה היא שאין קשר בין X ל-Y (אי־תלות), בעוד שהאלטרנטיבה היא בדרך כלל חד־כיוונית, הגורסת ש-Y נוטה לעלות כאשר X עולה ("קשר חיובי"), או השערה הפוכה, ש-Y נוטה לרדת כאשר X עולה ("קשר שלילי"). האלטרנטיבות של "קשר חיובי" או של "קשר שלילי" ניתנות להירשם במונחים הסתברותיים באופנים שונים, וכרגע לא נדון בכך. בפרק זה נתאר שני מדדים שונים לקשר בין המשתנים – מְתאם הדרגות של ספירמן, והמדד של קנדל.

6.1 מתאם הדרגות של ספירמן

המתאם של ספירמן (Spearman, 1904) הוא מקרה פרטי של מקדם המתאם המקובל של פירסון, המודד את הקשר הליניארי בין שני משתנים.
נזכיר את הגדרת המתאם של פירסון.

$$r=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2\sum_{i=1}^{n}(Y_i-\overline{Y})^2}} \qquad (1)$$

המקדם r הוא בעל התכונות הבאות:

א) לכל מדגם של n זוגות $-1\le r\le 1$;

ב) הערכים הקיצוניים של r מתקבלים כאשר n הזוגות (X_i,Y_i) נמצאים על קו ישר, $r=1$ אם שיפוע הקו חיובי (Y עולה כפונקציה של X) ו-$r=-1$ אם השיפוע שלילי (Y יורד כפונקציה של X).

ההתפלגות של r ידועה כאשר ההתפלגות המשותפת של הזוגות היא דו־נורמלית, ואת המודל הזה תמצאו בספרות הסטטיסטית הקלאסית בנושאים של מקדם מתאם ורגרסיה ליניארית. בניגוד לכך, המדד של ספירמן הוא אפרמטרי, מבחינה זאת שהתפלגותו, תחת השערת האפס, איננה תלויה כלל בהתפלגות המשותפת של הזוגות.

בניית המדד של ספירמן

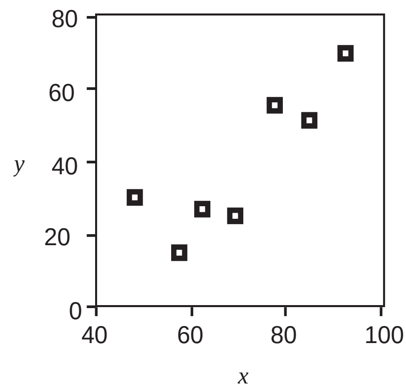
בחלק זה נניח שההתפלגות המשותפת היא רציפה.
מדרגים את הערכים $X_1,\ldots,X_n$ **ובנפרד** את הערכים $Y_1,\ldots,Y_n$.
מסמנים ב-$R_1,\ldots,R_n$ את דרגות ה-X-ים וב-$S_1,\ldots,S_n$ את דרגות ה-Y-ים.
מתאם הדרגות של ספירמן, מסומן r_s, הוא המקדם r של פירסון, נוסחה (1), כשהוא מחושב על הדרגות. כלומר,

$$r_s=\frac{\sum_{i=1}^{n}(R_i-\overline{R})(S_i-\overline{S})}{\sqrt{\sum_{i=1}^{n}(R_i-\overline{R})^2\sum_{i=1}^{n}(S_i-\overline{S})^2}} \qquad (2)$$

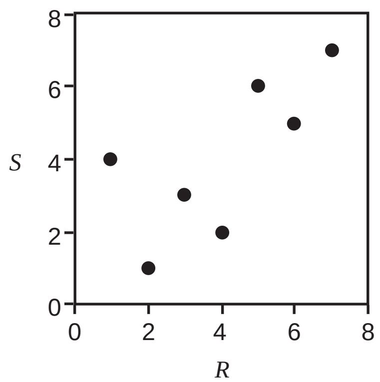
לפני שנַראה כיצד ניתן לרשום את מתאם הדרגות r_s בצורה פשוטה יותר, נוכל להיעזר בציורים 6.1 ו-6.2 כדי לבחון את משמעות ההבדל בין שני המתאמים – r ו-r_s.

דוגמה 6.1. בציור 6.1 מוצגת דיאגרמת הפיזור של ציוני 7 תלמידים בשתי בחינות. ציור 6.2 מציג את דיאגרמת הפיזור של הדרגות המתאימות. לא רשמנו כאן את הציונים עצמם. ניתן לזהות אותם מציור 6.1.

שתי דיאגרמות הפיזור שונות זו מזו. מקדם המתאם של פירסון, המודד את הקשר הליניארי בין שני המשתנים, כפי שהם מתוארים בציור 6.1, הוא $r=0.73$. לעומת זאת, מתאם הדרגות של ספירמן הוא אותו מקדם בדיוק, כשהוא מחושב לפי דיאגרמת הפיזור בציור 6.2, והוא כאן $r_s=0.51$. הקשר הליניארי בין הציונים עצמם נראה הדוק יותר מאשר בין הדרגות שלהם. הסיבה היא שפערים קטנים מאוד בין שני ציונים מהווים הבדל של דרגה שלמה, ולהיפך, פער מאוד גדול בין שני ציונים עוקבים מהווה גם הוא הבדל של דרגה אחת בלבד. ייתכנו, כמובן, מקרים שבהם מתקיים $r_s>r$. ראו, למשל, דוגמה 6.2 בהמשך.



ציור 6.1. *דיאגרמת פיזור של ציונים בשתי בחינות*



ציור 6.2. *דיאגרמת פיזור של דרגות הציונים בשתי בחינות*

חישוב מתאם הדרגות של ספירמן

הדרגות של ה-x-ים או של ה-y-ים כוללות את רשימת המספרים $1,2,\ldots,n$ בסדר המתאים לגודל התצפיות. לפיכך הממוצע של הדרגות הללו הוא קבוע:

$$\overline{R}=\overline{S}=\frac{1}{n}\left[1+2+\cdots+n\right]=\frac{n+1}{2} \qquad (3)$$

באותו אופן גם סכום ריבועי הסטיות מן הממוצע הוא קבוע:

$$\sum_{i=1}^{n}(R_i-\overline{R})^2=\sum_{k=1}^{n}\left(k-\frac{n+1}{2}\right)^2=nVar(U)=\frac{n(n^2-1)}{12} \qquad (4)$$

כאשר המשתנה U הוא משתנה אחיד $U(1,n)$ כפי שכבר ראינו בפרקים קודמים.

סכום הסטיות הריבועיות עבור הדרגות S_i זהה, כמובן, לזה של הדרגות R_i. נציב את

התוצאות של (3) ו-(4) בנוסחה (2) ונקבל את מתאם ספירמן:

$$r_s = \frac{12\sum_{i=1}^{n}\left(R_i - \frac{n+1}{2}\right)\left(S_i - \frac{n+1}{2}\right)}{n(n^2-1)} \qquad (5)$$

ניתן לחשב את r_s לפי הנוסחה (5), אולם החישוב קל יותר אם מפרקים את המונה באופן הבא:

$$\sum_{i=1}^{n}\left(R_i - \frac{n+1}{2}\right)\left(S_i - \frac{n+1}{2}\right) = \sum_{i=1}^{n} R_i S_i - n\left[\frac{n+1}{2}\right]^2 \qquad (6)$$

לפי נוסחה זו ניתן לרשום את r_s גם בצורה הבאה:

$$r_s = \frac{12\sum_{i=1}^{n} R_i S_i - n\left[\frac{n+1}{2}\right]^2}{n(n^2-1)} = \frac{12\sum_{i=1}^{n} R_i S_i}{n(n^2-1)} - \frac{3(n+1)}{n-1} \qquad (7)$$

דוגמה 6.2. הנתונים בדוגמה (לוח 6.1) הם מדגם מקרי של עשר ערים בארה"ב שלגביהם נרשמה הטמפרטורה המקסימלית בחודש ינואר בשנים 1931-1960 במעלות פרנהייט (מה לעשות, כך עדיין מקובל למדוד טמפרטורות ב"עולם החדש") והגובה מעל פני הים של העיר (רגל). נרצה למדוד את הקשר בין שני המשתנים הללו.

לוח 6.1. *נתוני טמפרטורה וגובה מעל פני הים בעשר ערים בארה"ב*

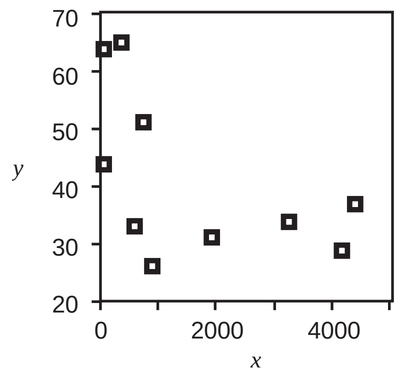
העיר	גובה x_i	דרגה R_i	טמפ' מקסימלית y_i	דרגה S_i	R_iS_i
Baltimore	20	1	44	7	7
Houston	40	2	64	9	18
Los Angeles	340	3	65	10	30
Detroit	585	4	33	4	16
Charlotte	720	5	51	8	40
Madison	860	6	26	1	6
Spokane	1,890	7	31	3	21
Rapid City	3,230	8	34	5	40
Helena	4,155	9	29	2	18
Salt Lake City	4,390	10	37	6	60
סך הכול		55		55	256

רשמנו בלוח 6.1 את הערים לפי סדר גודל עולה של גובהן מעל פני הים (המשתנה X).

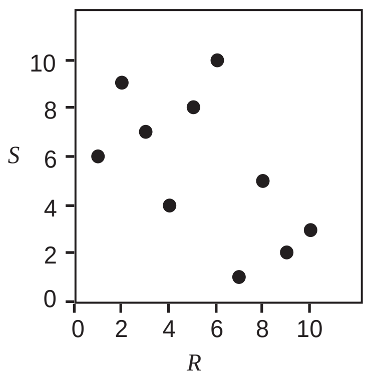
סיכום העמודה האחרונה (הימנית) נותן את סכום המכפלות של הדרגות. נציב ערך זה בנוסחה (7) ונקבל:

$$r_s = \frac{12\cdot 256}{10(10^2-1)} - \frac{3\cdot 11}{9} = -\frac{558}{990} = -.564$$

קיבלנו, אפוא, מתאם דרגות שלילי שמשמעו כי ככל שהגובה מעל פני הים עולה, כך הטמפרטורה המקסימלית נוטה לרדת. להדגמת המשמעות הזאת נראה כאן (ציורים 6.3 ו-6.4) את שתי דיאגרמות הפיזור – של הנתונים המקוריים ושל הדרגות המתאימות. בציור 6.3 בולטת העובדה שבין ערים הנמצאות בגובה נמוך מאוד (פחות מ-1,000 רגל מעל פני הים) יש הבדלי טמפרטורות גדולים מאוד, ולעומת זאת הערים הנמצאות בגבהים גדולים יותר במדגם זה הן בעלות טמפרטורות דומות, ללא נטייה של ירידת הטמפרטורה עם הגובה. אולם כאשר מסתכלים על הדרגות בלבד (ציור 6.4) נטיית הירידה של הטמפרטורה עם הגובה בולטת ביותר, כלומר, עיר הנמצאת בגובה רב יותר מעיר אחרת, נוטה להיות קרה ממנה.



ציור 6.3. *דיאגרמת פיזור של הגובה (x) והטמפרטורה (y)*



ציור 6.4. דיאגרמת פיזור של הדרגות

הערה: ערכו של מקדם המתאם של פירסון עבור נתוני דוגמה 6.2 קצת קטן מזה של מתאם הדרגות של ספירמן: $r = -.539$.

צורה נוספת לחישוב מתאם ספירמן ניתנת בעזרת הטענה הבאה.

טענה 6.1. נסמן את הפרשי הדרגות $D_i = R_i - S_i$, $i = 1,\ldots,n$. אזי

$$r_s = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)} \qquad (8)$$

הוכחה: נרשום את ריבוע ההפרש בצורה הבאה (הוספה והחסרה של הדרגה הממוצעת):

$$D_i^2=(R_i-S_i)^2=\left[\left(R_i-\frac{n+1}{2}\right)-\left(S_i-\frac{n+1}{2}\right)\right]^2$$
$$=\left[R_i-\frac{n+1}{2}\right]^2+\left[S_i-\frac{n+1}{2}\right]^2-2\left[R_i-\frac{n+1}{2}\right]\cdot\left[S_i-\frac{n+1}{2}\right]$$

סכום כל הביטויים הללו הוא

$$\sum_{i=1}^{n}D_i^2=2\frac{n(n^2-1)}{12}-2r_s\frac{n(n^2-1)}{12}=\frac{1}{6}n(n^2-1)(1-r_s)$$

העברת אגפים נותנת בדיוק את הנוסחה (8) הרשומה בטענה. ♣

מסקנה: מתאם הדרגות תלוי בהפרשי הדרגות המתאימות לכל תוצאה. קל לראות מנוסחה (8) שמתאם הדרגות של ספירמן מקבל את הערכים הבאים:

1) הערך $r_s=1$ מתקבל כאשר $\sum_{i=1}^{n}D_i^2=0$, כלומר, כאשר ה-x-ים וה-y-ים מסודרים בדיוק באותו סדר. ברור שזה גם הערך המקסימלי שהמתאם r_s יכול לקבל, כיוון ש-$\sum_{i=1}^{n}D_i^2\geq 0$ תמיד.

2) הערך המינימלי $r_s=-1$ מתקבל כאשר הדרגות של ה-y-ים בדיוק בסדר הפוך לאלה של ה-x-ים: $S_i=n+1-R_i$. את העובדה שהערך המינימלי הוא -1 נסחנו כטענה 6.2 בהמשך ואנו מוכיחים אותה בנספח 6א.

3) בכל מקרה $-1\leq r_s\leq 1$.

הערה: כיוון ש-r_s הוא, למעשה, מקדם מתאם של פירסון בין שתי סדרות של מספרים, ולגבי המדד של פירסון תכונה 3 ידועה, ניתן לוותר על ההוכחה. אנו מביאים אותה כאן לטובת הקוראים שאין להם ידע נוסף בסטטיסטיקה או בהסתברות. מעניין, עם זאת, שניתן להוכיח את העובדה שמתאם דרגות מקבל רק ערכים בין -1 ל-1, באמצעים מתמטיים פשוטים ביותר.

טענה 6.2. לכל סדרת n זוגות, מתקיים $r_s\geq -1$.
ההוכחה בנספח 6א.

התפלגות מתאם ספירמן תחת השערת האפס

ביטוי סטטיסטי להשערת האפס של חוסר קשר בין שני המשתנים X ו-Y מוגדר כאי־תלות בין המשתנים. כלומר,

H_0: X ו-Y בלתי תלויים

בינתיים לא נרשום את ההשערה האלטרנטיבית על ידי מודל סטטיסטי.

לבדיקת ההשערה נשתמש במקדם המתאם של ספירמן. זה סטטיסטי אפרמטרי, כיוון שהתפלגותו תחת H_0 אינה תלויה בהתפלגות המשותפת של (X,Y).

לוח 6.2. *התמורות של הדרגות* S_1,S_2,S_3,S_4 *ומתאם ספירמן כאשר* $R_1=1,\ R_2=2,\ R_3=3,\ R_4=4$

אפשרות	S_1	S_2	S_3	S_4	$\sum_{i=1}^{n} D_i^2$	r_s
1	1	2	3	4	0	1.0
2	1	2	4	3	2	0.8
3	1	3	2	4	2	0.8
4	1	3	4	2	6	0.4
5	1	4	2	3	6	0.4
6	1	4	3	2	8	0.2
7	2	1	3	4	2	0.8
8	2	1	4	3	4	0.6
9	2	3	1	4	6	0.4
10	2	3	4	1	12	-0.2
11	2	4	1	3	10	0.0
12	2	4	3	1	14	–0.4
13	3	1	2	4	6	0.4
14	3	1	4	2	10	0.0
15	3	2	1	4	8	0.2
16	3	2	4	1	14	–0.4
17	3	4	1	2	16	–0.6
18	3	4	2	1	18	–0.8
19	4	1	2	3	12	–0.2
20	4	1	3	2	14	–0.4
21	4	2	1	3	14	–0.4
22	4	2	3	1	18	–0.8
23	4	3	1	2	18	–0.8
24	4	3	2	1	20	–1.0

בהנחה שהמשתנים X ו-Y הם בלתי תלויים, אין כל קשר בין הסדר של התצפיות $X_1,\ldots,X_n$ לבין הסדר של התצפיות $Y_1,\ldots,Y_n$. מכאן נובע שכל $n!$ התמורות של $S_1,\ldots,S_n$ ביחס ל-$R_1,\ldots,R_n$ הן שוות הסתברות. לפיכך, התפלגות r_s תחת השערת האפס ניתנת על ידי

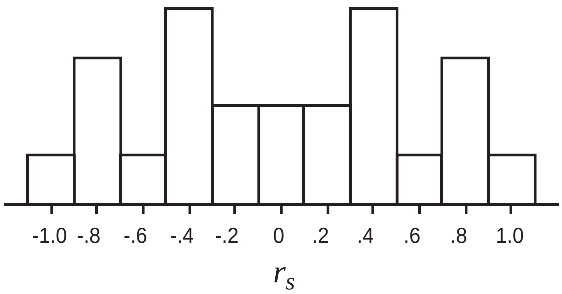
$$P_{H_0}(r_s = t) = \frac{\#\{r_s = t\}}{n!} \qquad (9)$$

בלוח 6.2 מוצגת צורת הספירה של כל האפשרויות עבור המקרה $n=4$. במקרה זה $n!=24$. בַלוח מצוינים גם סכומי הריבועים של ההפרשים והמתאם r_s, עבור כל אחת מהתמורות.

מלוח 6.2 ניתן למצוא את התפלגות הסטטיסטי r_s תחת השערת האפס, על ידי ספירת האפשרויות. בהתאם לכך מקבלים את ההתפלגות בלוח 6.3, וההיסטוגרמה של התפלגות זו מוצגת בציור 6.5.

לוח 6.3. *התפלגות הסטטיסטי של ספירמן תחת H_0, עבור $n=4$*

t	–1.0	–0.8	–0.6	–0.4	–0.2	0.0	0.2	0.4	0.6	0.8	1.0
$P(r_s=t)$	$\frac{1}{24}$	$\frac{3}{24}$	$\frac{1}{24}$	$\frac{4}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{2}{24}$	$\frac{4}{24}$	$\frac{1}{24}$	$\frac{3}{24}$	$\frac{1}{24}$



ציור 6.5. התפלגות r_s תחת השערת האפס, $n=4$

התפלגויות מתאם הדרגות של ספירמן לערכי n שונים ניתנות בטבלאות רבות, בדרך כלל עבור מדגמים שאינם עולים על 30. איננו מביאים טבלה זו. נראה מיד כיצד ניתן לחשב את ההתפלגות בקירוב.

בציור 6.5 רואים שעבור $n=4$ ההתפלגות של r_s תחת השערת האפס היא סימטרית סביב אפס. ניתן להראות שהסימטרייה הזאת קיימת עבור כל גודל מדגם n.

טענה 6.3. תחת השערת האפס למתאם הדרגות r_s התפלגות סימטרית סביב אפס.

הוכחה: לפי הכלל (9) הסתברות כל ערך של r_s פרופורציונלית למספר התמורות שעבורן ערך זה מתקבל. לכן, כדי להראות שהסתברויות שני ערכים הן שוות, די להראות שמספר האפשרויות לקבלת כל אחד מהם הוא שווה. נסתכל עתה על תמורה מסוימת של דרגות $R_1,\ldots,R_n$ ו-$S_1,\ldots,S_n$ שעבורה $r_s=t$. לאותה תמורה ישנה בדיוק

תמורה אחת שהיא "ההפוכה" שלה, הנותנת דרגה 1 לערך הגבוה ביותר של y ודרגה n לערך הנמוך ביותר של y. את הדרגות ה"הפוכות" של ה-y-ים ניתן לרשום $S_i^* = n+1-S_i$, $i=1,\dots,n$. נסמן ב-t^* את מתאם הדרגות המתקבל בין הדרגות $R_1,\dots,R_n$ לבין הדרגות ה"הפוכות" $S_1^*,\dots,S_n^*$. הסטייה בין הדרגה "ההפוכה" לדרגה הממוצעת היא

$$S_i^* - (n+1)/2 = n+1 - S_i - (n+1)/2 = -\left[S_i - (n+1)/2\right]$$

לכן נקבל את המתאם, לפי נוסחה (5)

$$t^* = r_s^* = \frac{12\sum_{i=1}^{n}\left(R_i - \frac{n+1}{2}\right)\cdot\left(S_i^* - \frac{n+1}{2}\right)}{n(n^2-1)}$$

$$= \frac{-12\sum_{i=1}^{n}\left(R_i - \frac{n+1}{2}\right)\cdot\left(S_i - \frac{n+1}{2}\right)}{n(n^2-1)} = -r_s = -t$$

קיבלנו, אפוא, שלכל אחת מהאפשרויות שעבורן $r_s = t$ יש בדיוק אפשרות אחת מתאימה שעבורה $r_s = -t$. לכן מספר האפשרויות הכולל שעבורן $r_s = t$ שווה בדיוק למספר האפשרויות הכולל שעבורן $r_s = -t$, ומכאן שההסתברויות לקבלת שני ערכים אלה הן שוות. זה נכון לכל ערך של t ומכאן הסימטרייה. ♣

קירוב נורמלי עבור התפלגות r_s

גם עבור התפלגות הסטטיסטי r_s ניתן להשתמש בקירוב נורמלי. שימו לב שנוסחה (7) מציגה את r_s כפונקציה ליניארית של סכום המכפלות $\sum_{i=1}^{n} R_i S_i$ וכסכום של משתנים מקריים, אין פלא שעבור n די גדול ההתפלגות היא בקירוב נורמלית.

משפט 6.1. תחת השערת האפס (אי־תלות) הסטטיסטי r_s הוא בעל התפלגות אסימפטוטית נורמלית. כלומר, עבור מדגם די גדול, התפלגות המשתנה המתוקנן $\frac{r_s - Er_s}{\sqrt{Var(r_s)}}$ היא בקירוב נורמלית סטנדרטית.

לא נוכיח את המשפט כאן.

יש למצוא את התוחלת ואת השונות של r_s תחת H_0. אלה מובאים במשפט 6.2.

משפט 6.2.. תחת השערת האפס (אי־תלות)

$$Er_s = 0 \qquad (10)$$

(11) $$Var(r_s)=\frac{1}{n-1}$$

הוכחה: מטענה 6.3 (סימטרייה סביב 0) נובעת התאפסות התוחלת.

את השונות נחשב בהתבסס על נוסחה (7). לשם הנוחיות נניח שה-x-ים כבר מסודרים לפי גודלם, באופן ש-$R_i=i$ עבור $i=1,\dots,n$. נסתכל ראשית על השונות של סכום המכפלות.

(12) $$Var(\sum_{i=1}^{n}R_iS_i)=Var(\sum_{i=1}^{n}iS_i)=\sum_{i=1}^{n}i^2Var(S_i)+\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}ijCov(S_iS_j)$$

תחת השערת האפס כל אחת מהדרגות S_i היא בעלת התפלגות אחידה $U(1,n)$ ולכן השונות של דרגה בודדת היא

(13) $$Var(S_i)=\frac{n^2-1}{12}$$

תחת השערת האפס אין כל קשר בין דרגות ה-x-ים לדרגות ה-y-ים, ולכן ההתפלגות המשותפת של שתי דרגות S_i ו-S_j מתקבלת כמו בבחירה מקרית של שני ערכים מבין $n,\dots,1$ ללא החזרה. השונות המשותפת של שתי דרגות שונות ניתנת, אפוא, לפי נוסחה (7) מפרק 2, לגבי שונות משותפת בין שתי תצפיות בדגימה בלי החזרה.

(14) $$Cov(S_i,S_j)=-\frac{Var(S_i)}{n-1}=-\frac{n^2-1}{12(n-1)}=-\frac{n+1}{12}$$

נציב את (13) ואת (14) בנוסחה (12) ונקבל את שונות הסכום:

(15) $$Var(\sum_{i=1}^{n}R_iS_i)=\sum_{i=1}^{n}i^2\frac{n^2-1}{12}+\sum_{i=1}^{n}\sum_{j\neq i}ij\left[-\frac{n+1}{12}\right]$$

נחשב כל אחד משני הסכומים באגף ימין בנפרד. סכום ריבועי המספרים הטבעיים העוקבים הוא על פי נוסחה ידועה (שכבר השתמשנו בה לא אחת):

$$\sum_{i=1}^{n}i^2=\frac{n(n+1)(2n+1)}{6}$$

סכום המכפלות מתקבל באופן הבא:

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}ij=\sum_{i=1}^{n}\sum_{j=1}^{n}ij-\sum_{i=1}^{n}i^2=\left[\sum_{i=1}^{n}i\right]^2-\sum_{i=1}^{n}i^2$$

הצבת שני הביטויים הללו בנוסחה (15) נותנת את שונות הסכום:

$$Var(\sum_{i=1}^{n}R_iS_i)=\frac{n^2-1}{12}\cdot\sum_{i=1}^{n}i^2-\frac{n+1}{12}\left\{\left[\sum_{i=1}^{n}i\right]^2-\sum_{i=1}^{n}i^2\right\}$$

$$=\left(\frac{n^2-1}{12}+\frac{n+1}{12}\right)\sum_{i=1}^{n} i^2-\frac{n+1}{12}\left[\sum_{i=1}^{n} i\right]^2$$

$$=\frac{(n+1)^2 n^2 (n-1)}{12^2}$$

(הציבו וודאו את התשובה!)

בסך הכול קיבלנו:

$$Var\left(\sum_{i=1}^{n} R_i S_i\right)=\frac{(n+1)^2 n^2 (n-1)}{12^2} \qquad (16)$$

לפי זה מתקבלת שונות מתאם הדרגות:

$$Var(r_s)=Var\left(\frac{12\sum_{i=1}^{n} R_i S_i}{n(n^2-1)}-\frac{3(n+1)}{n-1}\right)=\frac{12^2 Var\left(\sum_{i=1}^{n} R_i S_i\right)}{\left[n(n^2-1)\right]^2}$$

$$=\frac{12^2}{n^2(n+1)^2(n-1)^2}\cdot\frac{(n+1)^2 n^2 (n-1)}{12^2}=\frac{1}{n-1}$$

בכך הוכחנו את המשפט. ♣

מסקנה 6.1. תחת השערת האפס (אי־תלות) התפלגות המשתנה $\sqrt{n-1}\cdot r_s$ היא בקירוב נורמלית סטנדרטית.

דוגמה 6.3 (*המשך דוגמה 6.2*). הבעיה הנבדקת כאן היא אם יש קשר בין הגובה מעל פני הים (X) לבין הטמפרטורה המקסימלית (Y). ההשערות הנבדקות הן

H_0: X ו-Y בלתי תלויים   כנגד האלטרנטיבה – יש קשר שלילי בין X ו- Y :H_1

מטבע הדברים הבעיה לגבי הקשר בין הטמפרטורה והגובה היא חד־צדדית, ולכן הטענה שאותה נרצה לבדוק היא אם נכון הדבר שבמקומות גבוהים הטמפרטורה נוטה להיות נמוכה יותר מאשר במקומות נמוכים.

נחשב את מובהקות התוצאה שהתקבלה בדוגמה: $r_s=-.564$, כאשר $n=10$, בעזרת הקירוב הנורמלי. המשתנה המתוקנן הוא $z=\sqrt{9}\cdot(-.564)=-1.692$. המובהקות המקורבת של התוצאה היא, אפוא,

$$P=P_{H_0}(r_s\le -.564)\cong\Phi(-1.692)=.0455$$

עבור רמת מובהקות $\alpha=.05$ התוצאה אמנם מובהקת, אבל מניסיונכם ציפיתם בוודאי לתוצאה מובהקת הרבה יותר.

הסיבה שהקשר אינו חזק כל־כך בדוגמה זו היא שהערים שנבחרו למדגם הן מאזורי

רוחב גאוגרפי שונים מאוד, החל מהעיר הדרומית יוסטון (טקסס) ועד הצפונית ספוקן (מדינת וושינגטון). מובן שמתקבל על הדעת שיש קשר שלילי גם בין הרוחב הגאוגרפי לבין הטמפרטורה המקסימלית (בצפון קר יותר בחלק הצפוני של כדור הארץ) ולכן הגובה אינו מסביר את כל ההבדל בין הטמפרטורות של הערים השונות. לא נביא כאן את נתוני קו הרוחב הגאוגרפי, אולם רק נציין כי עבור אותן ערים שנכללו במדגם של דוגמה 6.2, המתאם של ספירמן בין הרוחב הגאוגרפי והטמפרטורה המקסימלית נמצא $r_s = -.879$. כדי לבודד את הקשר "הנקי" בין הגובה לבין הטמפרטורה יש לחשב את המתאם בין שני המשתנים הללו רק עבור ערים שכולן בערך באותו קו רוחב גאוגרפי. צורת חישוב כזאת נקראת **מתאם חלקי**. במודל רב-נורמלי התפלגות המתאם החלקי ידועה, ולכן ניתן להסיק לגבי הקשר בין המשתנים על סמך מתאם זה. אולם כשמחשבים מתאם חלקי על סמך הדרגות, התפלגותו תלויה בהתפלגות המשתנים הנחקרים ולכן הוא איננו סטטיסטי אפרמטרי.

קשר וסיבתיות

אין לבלבל את הקשר הסטטיסטי בין שני משתנים עם הסיבתיות. הסיפור המפורסם בהקשר זה הוא על חייזר מן המאדים שחזר מסיור על כדור הארץ וסיפר בהתפעלות לאנשי המאדים שלבני האדם יש תכשיר מצוין להשמנה, שנקרא סוכרזית. מתברר, הוא אמר, ששתיית משקה עם סוכרזית מעלה את המשקל של בני אדם. ובכן, הסיפור הוא רק סיפור, אולם לעתים מזומנות אנו נתקלים בטעויות דומות לאלה אפילו אצל חוקרים מכובדים.

לדוגמה, מצאו שאצל הרופאים המנתחים הידועים כיותר מומחים בתחומם ישנה נטייה לאחוזים גבוהים יותר של ניתוחים לא מוצלחים. האם זה אומר שהמומחים הללו אינם, למעשה, כל כך מומחים וכדאי לחולה להיות מנותח על ידי הרופא הפחות מומחה בתחומו? נראה שזה לא בהכרח כך. סביר שהקשר השלילי שנמצא נובע למעשה מכך שהרופאים המומחים מנתחים בדרך כלל את החולים היותר קשים, ולכן גם אחוז ההצלחה בניתוחים שלהם נמוך יותר.

אם כך, נִזָּהֵר ולא נמהר להסיק מקשר בין שני משתנים על הסיבתיות.

6.2 המתאם של ספירמן כאשר ישנם ערכי תיקו

ערכי תיקו המהווים בעיה בחישוב המתאם של ספירמן הם ערכים שווים בין ה-x-ים או ערכים שווים בין ה-y-ים. במקרה זה הדירוג נעשה כמו בכל המקרים הקודמים, על ידי דרגה ממוצעת. המתאם בין הדרגות הממוצעות מחושב בדיוק כמו עבור המקרה שאין

תיקו, נוסחה (2). כלומר,

$$\tilde{r}_s = \frac{\sum_{i=1}^{n}(\tilde{R}_i - \bar{\tilde{R}})(\tilde{S}_i - \bar{\tilde{S}})}{\sqrt{\sum_{i=1}^{n}(\tilde{R}_i - \bar{\tilde{R}})^2 \sum_{i=1}^{n}(\tilde{S}_i - \bar{\tilde{S}})^2}} \qquad (17)$$

כאשר $\tilde{R}_i, \tilde{S}_i$, $i=1,\ldots,n$, הן הדרגות הממוצעות.

כדי לחסוך בחישובים, ניתן לרשום את המתאם הזה בצורה יותר פשוטה. ראשית, ברור, כמו בכל המקרים הקודמים של ערכי תיקו, שגם כאן הממוצע של הדרגות אינו משתנה: $\bar{\tilde{R}} = \bar{\tilde{S}} = \bar{R} = \bar{S} = \frac{n+1}{2}$. בהתאם לכך, את המונה של $\tilde{r}_s$ ניתן לפרק, כמו במקרה שאין תיקו, נוסחה (6):

$$\sum_{i=1}^{n}\left(\tilde{R}_i - \frac{n+1}{2}\right)\left(\tilde{S}_i - \frac{n+1}{2}\right) = \sum_{i=1}^{n}\tilde{R}_i\tilde{S}_i - n\left[\frac{n+1}{2}\right]^2$$

גם את המכנה ניתן לרשום בצורה פשוטה יותר:

$$\sum_{i=1}^{n}(\tilde{R}_i - \bar{R})^2 = \sum_{i=1}^{n}\left(\tilde{R}_i - \frac{n+1}{2}\right)^2 = nVar(\tilde{U})$$

כאשר $\tilde{U}$ הוא משתנה המקבל את הערכים $\tilde{R}_1, \tilde{R}_2, \ldots, \tilde{R}_n$ בהסתברויות שוות. כבר חישבנו את השונות של $\tilde{U}$ והיא נתונה בנוסחה (26) בפרק 2. שונות זו שווה לשונות משתנה אחיד $U(1,n)$ שממנה יש לחסר את סכום השונויות של משתנים אחידים בכל קבוצות התיקו:

$$\sum_{i=1}^{n}(\tilde{R}_i - \bar{R})^2 = \frac{n(n^2-1)}{12} - \frac{1}{12}\sum_j t_j(t_j^2 - 1) \qquad (18)$$

לשם קיצור הכתיבה, נסמן $\sum_j t_j(t_j^2-1) = t*$, $\sum_k u_k(u_k^2-1) = u*$, כאשר t_j הוא גודל קבוצת התיקו ה-j בין ה-x-ים, ו-u_k הוא גודל קבוצת התיקו ה-k בין ה-y-ים. לפי נוסחה (18), כשהיא מופעלת בדומה גם לגבי דרגות ה-y-ים, נקבל את צורת החישוב של מתאם הדרגות מנוסחה (17).

$$\tilde{r}_s = \frac{\sum_{i=1}^{n}\tilde{R}_i\tilde{S}_i - n\left[(n+1)/2\right]^2}{\sqrt{\frac{1}{12}\left[n(n^2-1) - t*\right]\cdot\frac{1}{12}\left[n(n^2-1) - u*\right]}}$$

$$= \frac{12\sum_{i=1}^{n}\tilde{R}_i\tilde{S}_i - 3n(n+1)^2}{\sqrt{\left[n(n^2-1) - t*\right]\cdot\left[n(n^2-1) - u*\right]}} \qquad (19)$$

טענה 6.4. התוחלת והשונות של $\tilde{r}_s$ תחת השערת האפס זהות לאלה של r_s, כלומר

$$E\tilde{r}_s = 0 \qquad Var(\tilde{r}_s) = \frac{1}{n-1}$$

התאפסות התוחלת נובעת מהסימטרייה סביב אפס, גם במקרה של תיקו.
את נוסחת השונות לא נוכיח כאן.

דוגמה 6.4. להלן נתונים חלקיים (שירן שטיין) של דיווחי אבות לילדים צעירים לגבי מידת המעורבות שלהם בגידול הילדים (הציונים על סולם בין 1 ל-5) וכן ציוני "נשיות" שלהם, כפי שהתקבלו משאלון המודד משתנים הקשורים במגדר. נרצה לבדוק אם יש קשר בין ציון "נשיות" של האב לבין מידת מעורבותו בגידול הילדים.

לוח 6.4. *ציוני נשיות ומעורבות בגידול הילדים אצל אבות*

האב	מעורבות		נשיות	
	ציון (X)	דרגה $(\tilde{R})$	ציון (Y)	דרגה $(\tilde{S})$
1	2	1	5.47	11
2	3	2.5	3.87	3.5
3	3	2.5	3.64	2
4	4	6.5	3.87	3.5
5	4	6.5	3.27	1
6	4	6.5	5.80	14
7	4	6.5	5.40	9.5
8	4	6.5	5.40	9.5
9	4	6.5	4.47	5
10	5	13	4.87	7
11	5	13	4.80	6
12	5	13	5.53	12
13	5	13	6.07	16
14	5	13	5.87	15
15	5	13	5.33	8
16	5	13	5.73	13

התוצאות: $\sum_{i=1}^{16} \tilde{R}_i \tilde{S}_i = 1{,}302$

קבוצות התיקו הן בגודל: $t_1 = 2,\ t_2 = 6,\ t_3 = 7\ ; u_1 = 2,\ u_2 = 2$
המונה של מקדם המתאם בנוסחה (19) הוא

$$12(1{,}302)-3(16)(17^2)=1{,}752$$

עבור המכנה יש לחשב:

$$n(n^2-1)=16(16^2-1)=4{,}080$$

$$t^*=2(3)+6(35)+7(48)=552 \qquad u^*=2[2(3)]=12$$

הביטוי מתחת לשורש במכנה של $\tilde{r}_s$ הוא

$$\left[n(n^2-1)-t^*\right]\cdot\left[n(n^2-1)-u^*\right]=14{,}351{,}904$$

נציב את כל התוצאות הללו בנוסחה (19) ונקבל את מתאם הדרגות

$$\tilde{r}_s=\frac{12\sum_{i=1}^{n}\tilde{R}_i\tilde{S}_i-3n(n+1)^2}{\sqrt{\left[n(n^2-1)-t^*\right]\cdot\left[n(n^2-1)-u^*\right]}}$$

$$=\frac{1{,}752}{\sqrt{14{,}351{,}904}}=0.462$$

מובהקות התוצאה מתקבלת על פי הקירוב הנורמלי, לפי טענה **6.4**:

$$P=P(\tilde{r}_s\geq .462)\cong 1-\Phi(.462\sqrt{15})=1-\Phi(1.79)=.0366$$

בהנחה שההשערה שעמדה לבדיקה הייתה חד־צדדית, מובהקות זו קטנה מרמת המובהקות המקובלת של .05. ולכן נוכל להסיק על קשר חיובי בין מעורבות של אב בגידול ילדיו לבין מידת "נשיותו". כלומר, אב בעל תכונות "נשיות" נוטה להיות מעורב יותר בגידול ילדיו.

נעיר שכדי למצוא מובהקות מדויקת של מתאם דרגות כשהנתונים כוללים ערכי תיקו אי אפשר להיעזר בטבלאות הקיימות. אין טעם לחשב את המובהקות עבור המדגם הספציפי, היות שזו בדרך כלל טרחה רבה – יש למצוא את מספר כל התמורות שעבורן המתאם קיצוני לפחות כמו זה שהתקבל במדגם.

הערה: עבור מקרי תיקו שימוש בהפרשי הדרגות נותן נוסחה מקבילה לנוסחה (8), אך לא זהה לה. ריבוע הפרש הדרגות הממוצעות המתקבל:

$$\tilde{D}_i^2=(\tilde{R}_i-\tilde{S}_i)^2=\left[\left(\tilde{R}_i-\frac{n+1}{2}\right)-\left(\tilde{S}_i-\frac{n+1}{2}\right)\right]^2$$

$$=\left[\left(\tilde{R}_i-\frac{n+1}{2}\right)\right]^2+\left[\left(\tilde{S}_i-\frac{n+1}{2}\right)\right]^2$$

$$-2\left[\left(\tilde{R}_i-\frac{n+1}{2}\right)\cdot\left(\tilde{S}_i-\frac{n+1}{2}\right)\right]$$

נשתמש בנוסחה (18) לסכום האיברים עבור שני המחוברים הראשונים, בהתאמה, וכן בנוסחת מתאם הדרגות עבור המחובר האחרון, ונקבל את סכום ריבועי ההפרשים:

$$\sum_{i=1}^{n}\tilde{D}_i^2=\frac{1}{12}\left[n(n^2-1)-t^*\right]+\frac{1}{12}\left[n(n^2-1)-u^*\right]$$

$$-2\tilde{r}_s\frac{1}{12}\sqrt{\left[n(n^2-1)-t^*\right]\cdot\left[n(n^2-1)-u^*\right]}$$

$$=\frac{1}{6}\left[n(n^2-1)-\frac{t^*+u^*}{2}\right]$$

$$-\tilde{r}_s\frac{1}{6}\sqrt{\left[n(n^2-1)-t^*\right]\cdot\left[n(n^2-1)-u^*\right]}$$

לפיכך מקבלים את המתאם של ספירמן כפונקציה של סכום ריבועי ההפרשים בין הדרגות הממוצעות:

$$\tilde{r}_s=\frac{\left[n(n^2-1)-(t^*+u^*)/2\right]-6\sum_{i=1}^{n}\tilde{D}_i^2}{\sqrt{\left[n(n^2-1)-t^*\right]\cdot\left[n(n^2-1)-u^*\right]}} \qquad (20)$$

[חשבו שוב את המתאם בדוגמה 6.4 לפי נוסחה (20) וודאו שקיבלתם אותה תוצאה כמו קודם.]

נביא עתה דוגמה נוספת לחישוב מקדם המתאם של ספירמן עבור נתונים שבהם הרבה מאד ערכי תיקו, והם מוצגים בלוח שכיחות.

דוגמה 6.5. בסקר שנערך בבית חולים נשאלו נשים בהיריון אם הן מתכוונות להיניק את התינוק שיוולד. בלוח 6.5 נתונים חלקיים לגבי תשובות על ההנקה ומידת הדתיות של האישה. נרצה לבדוק אם בכלל יש קשר בין שני המשתנים הללו.

הבעיה כאן היא דו־צדדית, איננו יודעים באיזה כיוון יהיה הקשר הזה, אם בכלל הוא קיים.

נסמן ב-X את כוונת ההנקה (בשורות) וב-Y את מידת הדתיות (בעמודות). בין ערכי x יש שתי קבוצות תיקו, שכלולות בהן כל הנשים הנמצאות באותה שורה.

גודל קבוצות התיקו הן השכיחויות בשוליים: $t_1=37,\ t_2=139$;

באותו אופן עבור ערכי y קבוצות התיקו הן בגודל: $u_1=73,\ u_2=47,\ u_3=56$.

בלוח 6.5 רשמנו גם את השכיחויות השוליות המצטברות (המהוות כלי עזר לחישוב

הדרגות והן מצוינות באות עבה בלוח). כמו כן, רשמנו את הדרגה עבור כל רמה בכל אחד מהמשתנים (גם אלה מצוינות באות עבה). נסביר להלן כיצד חושבו הדרגות.

לוח 6.5. *נשים בהיריון לפי דתיות והנקה*

הנקה (X)	דתיות (Y)			סך הכול	מצטבר	דרגה $\tilde{R}_i$
	חילונית	מסורתית	דתית			
לא	20	11	6	37	**37**	**19**
כן	53	36	50	139	**176**	**107**
סך הכול	73	47	56	176		
מצטבר	**73**	**120**	**176**			
דרגה $\tilde{S}_i$	**37**	**97**	**148.5**			

דרגת ה-x-ים בכל אחת מהשורות היא הדרגה הממוצעת המגיעה לערך של השורה. בשורה הראשונה, הדרגות המגיעות ל-37 הנשים שאינן מתכוונות להיניק הן הדרגות $1,2,\ldots,37$ שהממוצע שלהן הוא $19=(1+37)/2$. בשורה השנייה, הדרגות המגיעות ל-139 הנשים שמתכוונות להיניק הן הדרגות $38,39,\ldots,176$ (הערך העליון הוא השכיחות המצטברת של שתי השורות הראשונות, השווה למספר הנשים הכולל) שהממוצע שלהן הוא $107=(38+176)/2$. הדרגות הללו רשומות בלוח 6.5 בעמודה הימנית.

בדיוק באותו אופן מקבלים את הדרגות של ה-y-ים בעמודות והן רשומות בשורה התחתונה של הלוח. כדי לחשב את המתאם של ספירמן יש לחשב את סכום המכפלות של הדרגות עבור כל אחת מהנשים.

נסתכל על התא הראשון (האישה אינה מתכוונת להיניק והיא חילונית). כל אחת מ-6 הנשים הכלולות בתא הזה הן בעלות הדרגות $\tilde{R}_i=19,\ \tilde{S}_i=37$, והמכפלה המתאימה לכל אחת מהן היא $\tilde{R}_i\tilde{S}_i=19\times 37$. באופן דומה, כל אחת מהנשים הכלולות בתא מסוים הן בעלות אותה דרגה גם לפי X וגם לפי Y. לפיכך סכום כל המכפלות של הדרגות ניתן על ידי

$$\sum_{i=1}^{n}\tilde{R}_i\tilde{S}_i=20(19)(37)+11(19)(97)+6(19)(148.5)$$
$$+53(107)(37)+36(107)(97)+50(107)(148.5)$$
$$=1,429,208$$

המונה של מקדם המתאם בנוסחה (19) הוא

$$12\sum_{i=1}^{n}\tilde{R}_i\tilde{S}_i-3n(n+1)^2=12(1{,}673{,}569.5)-3(176)(177^2)$$
$$=608{,}784$$

הביטויים עבור המכנה המתקבלים:

$$n(n^2-1)=176(176^2-10=5{,}451{,}600$$
$$t^*=37(37^2-1)+139(139^2-1)=2{,}736{,}096$$
ולכן $$n(n^2-1)-t^*=5{,}451{,}600-2{,}736{,}096=2{,}715{,}504$$
$$u^*=73(73^2-1)+47(47^2-1)+56(56^2-1)=668{,}280$$
ולכן $$n(n^2-1)-u^*=5{,}451{,}600-668{,}280=4{,}783{,}320$$

הצבה בנוסחה (19) נותנת את מתאם ספירמן

$$\tilde{r}_s=\frac{12\sum_{i=1}^{n}\tilde{R}_i\tilde{S}_i-3n(n+1)^2}{\sqrt{\left[n(n^2-1)-t^*\right]\cdot\left[n(n^2-1)-u^*\right]}}$$
$$=\frac{608{,}784}{\sqrt{(2{,}715{,}504)(4{,}783{,}320)}}=.169$$

המתאם הזה הוא קטן. על סמך הקירוב הנורמלי נקבל את מובהקות התוצאה (מבחן דו־צדדי):

$$P=2P_{H_0}(\tilde{r}_s\geq.169)\cong2\left[1-\Phi\left(.169\sqrt{175}\right)\right]=2(.0125)=.025$$

למרות שהמתאם קטן, הוא גדול מאפס באופן מובהק, עבור רמת מובהקות של 5%. מצאנו שיש קשר בין הכוונה להיניק לבין דתיות, כאשר הכוונה להיניק נוטה לעלות עם העלייה בדתיות. את כיוון הקשר קבענו לפי קביעת הסדר בטבלת השכיחויות, מה שקבע את הדירוג של כל אחד מן המשתנים.

בִּדקו והיווכחו שאם הייתם משנים את הסדר של ערכי ההנקה, ומסמנים את הערך "כן" כערך הנמוך ואת הערך "לא" כערך הגבוה, המתאם היה מתקבל באופן סימטרי $\tilde{r}_s=-.169$.

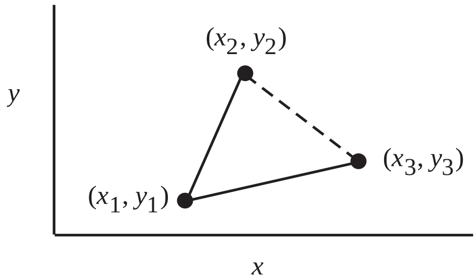
יעילות יחסית אסימפטוטית

היעילות היחסית האסימפטוטית של המבחן המבוסס על מתאם הדרגות של ספירמן ביחס למבחן המבוסס על המתאם של פירסון היא $9/\pi^2\cong.912$, במודל הדו-נורמלי (שבו השימוש במתאם של פירסון הוא מתאים). אנו רואים, אפוא, שאין הפסד גדול בשימוש במתאם דרגות במקום במתאם הרגיל, המבוסס על התצפיות המקוריות.

6.3 המתאם של קנדל

נתון מדגם של n תצפיות של זוגות בלתי תלויים (X_i,Y_i), $i=1,\ldots,n$. נניח, ראשית, שאין שום ערכי תיקו במדגם (ההתפלגות המשותפת רציפה).

נסתכל על שני זוגות מסוימים, למשל (x_1,y_1) ו-(x_2,y_2). אנו אומרים ששני הזוגות הם **מתאימים** (באנגלית concordant), אם הסדר בין שני ערכי x זהה לסדר בין שני ערכי y. כלומר, אם $x_1<x_2$ וגם $y_1<y_2$, או $x_1>x_2$ וגם $y_1>y_2$. משמעות *ההתאמה* היא שהנבדק שעבורו המשתנה X גדול יותר, הוא גם זה שעבורו המשתנה Y גדול יותר. למשל: שני הזוגות $(x_1,y_1)=(3,10)$ ו-$(x_2,y_2)=(5,45)$ הם מתאימים מכיוון שלגבי שני המשתנים הערכים של הזוג הראשון נמוכים מאלה של הזוג השני. אולם שני הזוגות $(x_2,y_2)=(5,45)$ ו-$(x_3,y_3)=(8,20)$ אינם מתאימים (באנגלית discordant), כי הכיוונים שלהם הפוכים – הערך של X גדול יותר בתצפית השנייה ואילו הערך של Y גדול יותר בתצפית הראשונה. הדגמה לכך תמצאו בציור 6.6. הזוג (x_1,y_1) הוא בהתאמה גם עם (x_2,y_2) וגם עם (x_3,y_3). לעומת זאת שני הזוגות (x_2,y_2) ו-(x_3,y_3) אינם מתאימים. נקל לראות זאת על פי הקו המחבר את שתי הנקודות במישור (x,y). כאשר קו זה הוא בעל שיפוע חיובי (קו "עולה"), שני הזוגות הם מתאימים, וכאשר קו זה הוא בעל שיפוע שלילי (קו "יורד"), שני הזוגות אינם מתאימים.



ציור 6.6. *הזוגות המחוברים בקו שלם הם מתאימים והמחוברים בקו מקווקו אינם מתאימים*

אם יש קשר חיובי בין שני המשתנים, הערך של Y נוטה להיות גבוה ככל שהערך של X גבוה, ואז נצַפה שבסיכוי גבוה שני זוגות מקריים יהיו בהתאמה.

בהתאם לכך נגדיר את המדד של קנדל (Kendall, 1938). בהינתן מדגם של n תצפיות (X_i,Y_i), נסתכל על כל $\binom{n}{2}$ הזוגות של תצפיות ונספור כמה מהזוגות הללו מתאימים. נסמן מספר זה ב-N_c. מספר הזוגות שאינם מתאימים הוא, כמובן, כל השאר, ונסמנו: $N_d=\binom{n}{2}-N_c$. ניתן לרשום את N_c ואת N_d בצורה מתמטית, בעזרת המשתנים

המציינים הבאים:

$$H_{ij}=\begin{cases}1 & (X_i-X_j)(Y_i-Y_j)>0 \\ 0 & (X_i-X_j)(Y_i-Y_j)<0\end{cases} \qquad i\neq j \tag{21}$$

מכפלת ההפרשים הללו היא חיובית כאשר לשני ההפרשים אותו סימן, כלומר, כאשר שתי התצפיות מתאימות. המכפלה שלילית כאשר הסימנים שונים, ואז אין התאמה.
לפי זה מספר הזוגות המתאימים מתקבל על ידי סכום המשתנים המציינים ב-(21)

$$N_c=\sum_{i=1}^{n}\sum_{j>i}H_{ij}=\#\left\{i<j:(X_i-X_j)(Y_i-Y_j)>0\right\} \tag{22}$$

מקדם המתאם של קנדל

מקדם המתאם של קנדל מוגדר על ידי

$$T=\frac{N_c-N_d}{\binom{n}{2}}=\frac{N_c}{\binom{n}{2}}-\frac{N_d}{\binom{n}{2}}=\frac{2N_c}{n(n-1)}-\frac{2N_d}{n(n-1} \tag{23}$$

כלומר, זהו ההפרש בין פרופורציית הזוגות המתאימים לאלה שאינם מתאימים, מתוך כלל ההשוואות הנערכות.

נסתכל שוב על נתוני דוגמה 6.1 ונחשב עבורם את המתאם של קנדל.

דוגמה 6.6 (המשך דוגמה 6.1). בלוח 6.6 רשומות רק הדרגות של נתוני הגובה מעל פני הים (R_i) והטמפרטורה המקסימלית (S_i) של עשר ערים בארה"ב. הערים מסודרות בלוח לפי הגובה שלהן. בלוח רשמנו עבור כל עיר (x_i,y_i) את $N_{c(i)}$ – מספר הערים (x_j,y_j) המתאימות לה, כאשר $j>i$. (שתי ערים הן מתאימות, אם העיר הגבוהה יותר מעל פני הים היא גם בעלת הטמפרטורה המקסימלית הגבוהה יותר.) כמו כן, רשמנו עבור כל עיר את מספר הערים שאינן מתאימות לה – $N_{d(i)}$. הסכום של שני הערכים הללו רשום בעמודה האחרונה בצד ימין, והוא, למעשה, מספר כל הערים שאליהן השווינו את העיר ה-i (x_i,y_i).
נסתכל על השורה הראשונה. השווינו את העיר בולטימור, שהדרגות שלה הן $(1,7)$, לכל אחת מ-9 הערים האחרות. כיוון שדרגות הגובה R_i כבר מסודרות בסדר עולה, די להסתכל על עמודת S_i ולספור כמה דרגות בעמודה זו, מתוך כל 9 הדרגות האפשריות, גבוהות מהדרגה 7. הדרגות 8, 9 ו-10 מקיימות את התנאי ולכן יש 3 ערים המתאימות לבולטימור – יוסטון, לוס אנג'לס ושרלוט. מכאן $N_{c(1)}=3$. הערך $N_{d(1)}=6$ הוא מספר הדרגות הנמוכות מ-7. הסכום 9 הוא מספר כל ההשוואות שערכנו עם בולטימור.

לוח 6.6. *הדרגות של נתוני טמפרטורה וגובה מעל פני הים בעשר ערים בארה"ב, וחישוב המתאם של קנדל*

העיר i	R_i	S_i	$N_{c(I}$	$N_{d(i}$	$N_{c(i)}+N_{d(i)}$
1. Baltimore	1	7	3	6	9
2. Houston	2	9	1	7	8
3. Los Angeles	3	10	0	7	7
4. Detroit	4	4	3	3	6
5. Charlotte	5	8	0	5	5
6. Madison	6	1	4	0	4
7. Spokane	7	3	2	1	3
8. Rapid City	8	5	1	1	2
9. Helena	9	2	1	0	1
10. Salt Lake City	10	6	–	–	–
סך הכול	55	55	15	30	45

את העיר השנייה (יוסטון), שדרגותיה $(2,9)$, אנו משווים רק לשאר 8 הערים שנמצאות בשורות הנמוכות יותר, מתחת לקו המקווקו. כל הדרגות R_i גבוהות שם יותר מהדרגה 2 וביניהן מצאנו רק אחת שבה הדרגה S_i גם היא גבוהה יותר מהדרגה 9 (הדרגה 10 – לוס אנג'לס). כל שאר 7 הדרגות נמוכות יותר. כך רשמנו $N_{c(2)}=1$ ו-$N_{d(2)}=7$. הסכום $1+7=8$ הוא מספר כל התצפיות מתחת לקו המקווקו.

באותו אופן מסתכלים על כל אחת מהערים ומשווים אותה לכל אלה הנמצאות בשורה נמוכה יותר. בעמודה האחרונה מימין רשומות מספר ההשוואות שנעשו עבור כל עיר (ערך זה יורד ב-1 בכל שורה). סכום הערכים בעמודה זו הוא $45=\frac{10\cdot 9}{2}=\binom{10}{2}$, שזה מספר כל ההשוואות בין שתי ערים שונות. סכום הערכים בעמודה שכותרתה $N_{c(i)}$ הוא בדיוק N_c, וסכום הערכים בעמודה שכותרתה $N_{d(i)}$ הוא בדיוק N_d. מכאן נקבל את המתאם של קנדל:

$$T=\frac{N_c-N_d}{n(n-1)/2}=\frac{15-30}{45}=-\frac{15}{45}=-\frac{1}{3}$$

מתאם זה הוא שלילי, כפי שהיה גם המתאם של ספירמן עבור אותם הנתונים (בדוגמה **6.1** מצאנו $r_s=-.564$).

הרעיון העומד מאחורי המתאם של קנדל הוא שאם המשתנה Y נוטה לעלות כאשר X עולה, יותר זוגות יִטו להיות מתאימים זה לזה (הקווים המחברים ביניהם יהיו בכיוון

חיובי, כמו הקווים השלמים בציור 6.6). במקרה זה יתקיים $N_c > N_d$, ואז יהיה $T>0$. אם, לעומת זאת, המשתנה Y נוטה לרדת כאשר X עולה, יותר זוגות יִטו להיות לא מתאימים (הקווים המחברים ביניהם יהיו בכיוון שלילי, כמו הקווים המקווקווים בציור 6.6) ואז יהיה $T<0$. אם אין קשר בין המשתנים, נצַפה שבערך חצי מן ההשוואות תהיינה של זוגות מתאימים, ולכן T יהיה בסביבות 0.

הערכים האפשריים של T מתקבלים באופן הבא:

אם כל הזוגות הם מתאימים, אזי $N_c = n(n-1)/2$, $N_d = 0$ ו-$T=1$.

אם כל הזוגות אינם מתאימים, אזי $N_c = 0$, $N_d = n(n-1)/2$ ו-$T=-1$.

כל שאר המקרים נופלים בין שני הערכים הקיצוניים הללו ולכן תמיד $-1 \le T \le 1$.

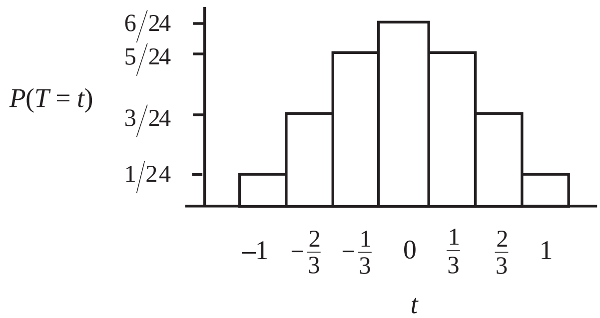
התפלגות T תחת השערת האפס

כפי שראינו בדוגמה 6.6, הערך של T תלוי אך ורק בסדר בין דרגות ה-y-ים ביחס לדרגות ה-x-ים. לכן, כמו עבור המתאם של ספירמן, תחת השערת האפס של אי־תלות, לכל אחת מהתמורות אותה הסתברות, וההתפלגות של T ניתנת על ידי

$$P_{H_0}(T=t) = \frac{\#\{T=t\}}{n!} \qquad (24)$$

גם עבור הסטטיסטי הזה קיימות טבלאות של ההתפלגות תחת השערת האפס, שאותן איננו מביאים כאן.

את ההתפלגות המתקבלת עבור $n=4$ תצפיות, ניתן למצוא על סמך לוח 6.2, שם רשמנו את כל התמורות וחישבנו את התפלגות המתאם של ספירמן. ההתפלגות המתקבלת נתונה על ידי ההיסטוגרמה בציור 6.7 (בדקו!).



ציור 6.7. *התפלגות הסטטיסטי של קנדל תחת השערת האפס,* $n=4$

השוואת ציור 6.7 לציור 6.5 מראה את ההבדל בין התפלגויות שני המתאמים המוצגים כאן. הסטטיסטי של קנדל הוא בעל התפלגות מאוד "רגולרית" – פונקציית ההסתברות

היא סימטרית, עם שכיח במרכז. בטבלאות מתאימות ניתן למצוא את מובהקות התוצאה של דוגמה 6.6. התוצאה שהתקבלה היא $T=-1/3$ והמובהקות $P=.108$.

את הסימטרייה של התפלגות המתאם של קנדל לא קשה להראות, והיא מובאת בטענה הבאה.

טענה 6.5. לכל n, התפלגות המתאם של קנדל T תחת השערת האפס היא סימטרית סביב אפס.

הוכחה: לפי הכלל (24) הסתברות כל ערך של T פרופורציונלית למספר התמורות שעבורן ערך זה מתקבל. לכן, כדי להראות שהסתברויות שני ערכים הן שוות, די להראות שלשניהם אותו מספר אפשרויות (בדיוק כמו שעשינו לגבי הסטטיסטי של ספירמן, טענה 6.3).

נסתכל על תמורה מסוימת של דרגות: $R_1,\ldots,R_n$ ו-$S_1,\ldots,S_n$ שעבורה $T=\dfrac{N_c-N_d}{n(n-1)/2}=t$. לאותה תמורה ישנה בדיוק תמורה אחת שהיא "ההפוכה" שלה, כלומר $S_i^*=n+1-S_i$, $i=1,\ldots,n$. (כבר ראינו את צורת ההפיכה הזאת בהוכחה לטענה 6.3.) נסמן ב-t^* את הערך של מתאם קנדל T המתקבל בין $R_1,\ldots,R_n$ לבין $S_1^*,\ldots,S_n^*$. היות שערכי $R_1,\ldots,R_n$ זהים בשני המקרים, וערכי $S_1^*,\ldots,S_n^*$ הפוכים לאלה של $S_1,\ldots,S_n$, כל זוג שנמצא מתאים קודם הוא עתה אינו מתאים, ולהיפך. מספר הזוגות המתאימים עתה הוא $N_c^*=N_d$ ומספר הזוגות שאינם מתאימים הוא $N_d^*=N_c$.

לפיכך הערך החדש של מתאם קנדל הוא

$$t^*=\frac{N_c^*-N_d^*}{n(n-1)/2}=\frac{N_d-N_c}{n(n-1)/2}=-t$$

ומכאן הסימטרייה סביב אפס. ♣

קירוב נורמלי עבור התפלגות T

משפט 6.3. תחת השערת האפס (אי־תלות) הסטטיסטי T הוא בעל התפלגות אסימפטוטית נורמלית. כלומר, התפלגות המשתנה המתוקנן $\dfrac{T-ET}{\sqrt{Var(T)}}$ היא בקירוב נורמלית סטנדרטית עבור מדגם די גדול.

משפט זה לא נוכיח כאן.

התוחלת והשונות של T מתקבלות לפי המשפט הבא.

משפט 6.4. תחת השערת האפס (אי־תלות) $ET=0$, וכן

$$Var(N_c - N_d) = \frac{n(n-1)(2n+5)}{18} \qquad (25)$$

ושונות הסטטיסטי T היא

$$Var(T) = \frac{Var(N_c - N_d)}{\left[n(n-1)/2\right]^2} = \frac{2(2n+5)}{9n(n-1)} \qquad (26)$$

הוכחה: מן הסימטרייה סביב 0 של התפלגות הסטטיסטי T ברור שהתוחלת תחת השערת האפס היא 0. חישוב השונות הרבה יותר מסובך והמתעניינים יוכלו למצוא אותו בנספח 6ב.

דוגמה 6.7 (המשך דוגמה 6.6). השונות עבור מדגם של $n=10$ תצפיות, כפי שהבאנו בדוגמה 6.6, היא

$$Var(T) = \frac{2(20+5)}{9(10)(9)} = 0.0617$$

המובהקות המקורבת של תוצאת המדגם $T=-1/3$ היא, אפוא:

$$P = P_{H_0}(T \le -1/3) \cong \Phi\left(\frac{-1/3}{\sqrt{0.0617}}\right) = \Phi(-1.34) = .0901$$

לפי ההתפלגות המדויקת, אותה תוצאה מראה מובהקות $P=.108$, שהיא קצת גבוהה מהמובהקות המקורבת.

תיקון רציפות

אפשר לקבל קירוב טוב יותר אם עושים תיקון רציפות. המונה של הסטטיסטי T, $N_c - N_d$, הוא משתנה בדיד. הפער בין שני ערכים סמוכים של משתנה זה הוא 2. זאת מכיוון שסכומם קבוע $N_c + N_d = n(n+1)/2$, ולכן כאשר N_c עולה ביחידה אחת, חייב N_d לרדת ביחידה אחת. בהתאם לכך את ההתפלגות המקורבת ניתן למצוא על ידי תקנון $N_c - N_d$, עם השונות מנוסחה (25) באופן הבא:

$$P(N_c - N_d \le t) \cong \Phi\left(\frac{t+1}{\sqrt{n(n-1)(2n+5)/18}}\right)$$

נחשב שוב את המובהקות המקורבת של התוצאה מדוגמה 6.6, שם $n=10$ והתוצאה $N_c - N_d = 15$, עם תיקון רציפות.

$$P = P(N_c - N_d \ge 15) \cong \Phi\left(\frac{14}{\sqrt{(10)(9)(25)/18}}\right) = \Phi(1.25) = .1052$$

תוצאה זו מאוד קרובה למובהקות המדויקת.

התאוריה של קנדל לגבי המדד לקשר בין שני משתנים*

נסתכל על ההתפלגות הדו־ממדית של המשתנה (X,Y). מגדירים את ההסתברות להתאמה ואת ההסתברות לאי התאמה בין זוג של שני משתנים בלתי תלויים כאלה – (X_1,Y_1) ו- (X_2,Y_2) על ידי

$$\pi_c = P\left\{(X_1-X_2)(Y_1-Y_2)>0\right\} \qquad (27)$$

$$\pi_d = P\left\{(X_1-X_2)(Y_1-Y_2)<0\right\} \qquad (28)$$

π_c היא ההסתברות לכך שההפרש בין שתי התצפיות על המשתנה X וההפרש בין שתי התצפיות על המשתנה Y הם באותו סימן (הזוגות מתאימים). π_d היא ההסתברות לכך ששני ההפרשים הם בסימנים שונים (הזוגות אינם מתאימים).

הגדרה: המדד של קנדל לקשר בין X ו-Y מוגדר על ידי

$$\tau = \pi_c - \pi_d \qquad (29)$$

כאשר π_c ו-π_d מוגדרות בנוסחאות (27) ו-(28).

כלומר, הפרמטר τ הוא הפער בין ההסתברות להתאמה לבין ההסתברות לחוסר התאמה בין שתי תצפיות דו־ממדיות.

טענה 6.6. הפרמטר τ מקבל ערכים בתחום $-1\le\tau\le1$.

הוכחה: הפרמטר τ הוא הפרש בין שתי הסתברויות, המקבלות ערכים בין 0 ל-1, שסכומן 1: $\pi_c+\pi_d=1$. ההפרש ביניהן אינו יכול, אפוא, לעלות על 1 וכן אינו יכול להיות נמוך מ--1. ♣

טענה 6.7. הסטטיסטי T של קנדל לבדיקת ההשערה שהמשתנים X ו-Y הם בלתי תלויים, כפי שהוגדר בנוסחה (23), הוא אומד בלתי מוטה של הפרמטר τ.

הוכחה: נסתכל על ההגדרה (22) למספר הזוגות המתאימים. כל המשתנים H_{ij}, $(i<j)$, הם בעלי אותה התפלגות, ולכן, כמובן, בעלי אותה תוחלת.

$$EH_{ij} = P\left\{(X_i-X_j)(Y_i-Y_j)>0\right\} = \pi_c$$

לפי ההגדרה (22)

$$EN_c = \binom{n}{2}\pi_c$$

ותוחלת הפרופורציה של הזוגות המתאימים במדגם של n תצפיות היא

$$E\frac{N_c}{\binom{n}{2}} = \pi_c$$

באופן דומה,
$$E\frac{N_d}{\binom{n}{2}} = P\left\{(X_i - X_j)(Y_i - Y_j) < 0\right\} = \pi_d$$

ומכאן תוחלת ההפרש היא
$$ET = E\frac{N_c - N_d}{\binom{n}{2}} = \pi_c - \pi_d = \tau$$

♣

טענה 6.8. אם המשתנים X ו-Y הם בלתי תלויים אזי $\tau = 0$.

הוכחה: אם X ו-Y הם בלתי תלויים, אז
$$\begin{aligned}\pi_c &= P(X_1 > X_2, Y_1 > Y_2) + P(X_1 < X_2, Y_1 < Y_2)\\ &= P(X_1 > X_2)P(Y_1 > Y_2) + P(X_1 < X_2)P(Y_1 < Y_2)\\ &= \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{2}\end{aligned}$$

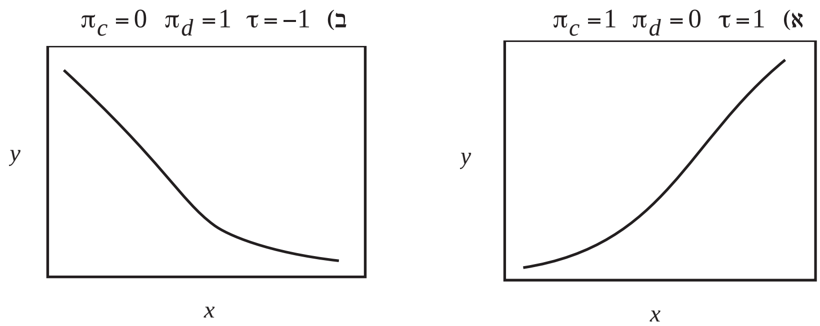
באופן דומה ברור שגם $\pi_d = 1/2$ ומכאן $\tau = 0$. ♣

הערה 1. ההיפך מטענה 6.8 אינו נכון. כלומר, ייתכן שהזוג (X, Y) הוא בעל התפלגות משותפת שעבורה $\tau = 0$, ועם זאת המשתנים הם תלויים. (אותו דבר נכון גם למקדם המתאם הרגיל של פירסון, שיכול להתאפס מבלי שהמשתנים יהיו בלתי תלויים.) דוגמה לכך ניתן יהיה לראות בדוגמה 6.8, שהיא דוגמה פשוטה ביותר, עבור מקרה של תיקו.

הערה 2. נוסיף ונעיר, כי ניתן למצוא קשר פונקציונלי בין הפרמטרים τ ו-ρ עבור ההתפלגות המשותפת הדו־נורמלית, שם התאפסות המתאם ρ אקוויוולנטית לאי־תלות. אם המשתנה הדו־ממדי (X,Y) הוא בעל התפלגות דו־נורמלית עם מקדם מתאם ρ, אזי קיים:
$$\tau = \frac{2}{\pi}\arcsin(\rho)$$
לא נביא כאן את ההוכחה, כיוון שהיא מעבר לרמת הידע בהסתברות הנדרשת בספר זה. עם זאת, ניתן לראות מכאן בקלות, שבמקרה הדו־נורמלי הפרמטרים ρ ו-τ מתאפסים יחד. כלומר, $\rho = 0 \Leftrightarrow \tau = 0$.

היות שבמודל הדו־נורמלי התאפסות המתאם ρ אקוויוולנטית לאי־תלות בין המשתנים, גם התאפסות τ אקוויוולנטית לאי־תלות במודל זה. כלומר, במקרה של התפלגות דו־נורמלית לא תיתכן תלות בין המשתנים אם הערך של τ הוא 0.

באופן כללי קל לראות שאם יש קשר "חיובי" בין X ל-Y אזי $\tau > 0$, ואם הקשר הוא "שלילי", אזי $\tau < 0$. במקרים הקיצוניים: (א) כשכל הנקודות (x,y) שעבורן הצפיפות המשותפת היא חיובית נמצאות על עקום עולה (כלומר, Y הוא פונקציה עולה של X), הערך של הפרמטר הוא $\tau = 1$, או (ב) כאשר הן נמצאות על עקום יורד (Y הוא פונקציה יורדת של X), הפרמטר הוא $\tau = -1$. תוכלו לראות זאת בציור 6.8. במקרים הקיצוניים הללו ברור שכל שתי תצפיות שיתקבלו במדגם יהיו בוודאות מתאימות (במקרה א) או בוודאות לא מתאימות (במקרה ב) ולכן הסטטיסטי T הוא משתנה קבוע, המקבל בוודאות את הערך 1 במקרה א, או את הערך 1– במקרה ב.



ציור 6.8. *המשתנה Y הוא פונקציה מונוטונית של המשתנה X.*

הערה: לגבי מתאם הדרגות של ספירמן, מדובר על סטטיסטי המתקבל במדגם, ולא הגדרנו פרמטר הקשור בהתפלגות המשותפת, שסטטיסטי זה הוא אומד שלו. לא נעשה זאת במסגרת הספר הזה, כיוון שחישוב התוחלת של מתאם ספירמן דורש רמת ידע מעל לנדרש בספר זה.

6.4 המתאם של קנדל כאשר ישנם ערכי תיקו

הטיפול בערכי תיקו כשמשתמשים בסטטיסטי של קנדל אינו ברור מאליו. נסתכל על הפרמטר τ במקרה שלפחות אחד משני המשתנים, X או Y, אינו רציף. ברור שההגדרה (29) במונחים של ההגדרות (27) ו-(28) זקוקה לתיקון, כיוון שבמקרה של אי־רציפות יש הסתברות חיובית שלפחות אחד ההפרשים: $X_1 - X_2$ או $Y_1 - Y_2$ הוא 0, ולכן במקרה זה סכום ההסתברויות להתאמה ולאי התאמה קטן מ-1.

נסתכל, אפוא, על ההסתברויות המותנות:

$$\tilde{\pi}_c = P\left\{(X_1 - X_2)(Y_1 - Y_2) > 0 \mid (X_1 - X_2)(Y_1 - Y_2) \neq 0\right\}$$

$$\tilde{\pi}_d = P\left\{(X_1-X_2)(Y_1-Y_2)<0 \mid (X_1-X_2)(Y_1-Y_2)\neq 0\right\}$$

אלה הן ההסתברויות ששתי התצפיות הן בהתאמה, או באי התאמה, **כשנתון** שאין ביניהן ערכי תיקו. ברור שסכום שתי ההסתברויות המותנות הללו הוא 1.

שימו לב שאת הסתברויות המותנות לעיל ניתן לחשב באופן הבא:

$$\tilde{\pi}_c = \frac{P\left\{(X_1-X_2)(Y_1-Y_2)>0\right\}}{P\{(X_1-X_2)(Y_1-Y_2)\neq 0\}} = \frac{\pi_c}{\pi_c+\pi_d}$$

$$\tilde{\pi}_d = \frac{P\left\{(X_1-X_2)(Y_1-Y_2)<0\right\}}{P\{(X_1-X_2)(Y_1-Y_2)\neq 0\}} = \frac{\pi_d}{\pi_c+\pi_d}$$

את מקדם המתאם של קנדל במקרה זה נגדיר על ידי ההפרש בין שתי ההסתברויות המותנות הללו:

$$\tilde{\tau} = \tilde{\pi}_c - \tilde{\pi}_d = \frac{\pi_c-\pi_d}{\pi_c+\pi_d} \qquad (30)$$

נסתכל עתה על דוגמה 6.8 שבה נדגים חישוב של $\tilde{\tau}$ במקרה פשוט.

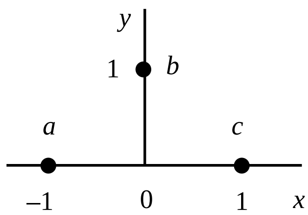
דוגמה 6.8. זו דוגמה לזוג משתנים מקריים שהם תלויים, ולמרות זאת $\tilde{\tau}=0$ (וגם $\rho=0$).

נניח שלמשתנה (X,Y) פונקציית ההסתברות (הבדידה) הבאה:

$$P(X=-1,Y=0)=P(X=1,Y=0)=P(X=1,Y=1)=1/3$$

במישור הדו־ממדי פונקציית ההסתברות נראית כמו בציור **6.9**.

בסך הכול ישנן שלוש תוצאות אפשריות לקבלת הזוג (X,Y), שהסתברות כל אחת מהן היא $1/3$. סימנו את הנקודות הללו ב-a, b ו-c בציור **6.9**.



ציור 6.9. *התפלגות משותפת לדוגמה 6.8*

שימו לב שאם בהוצאת שתי תצפיות בלתי תלויות באופן מקרי מקבלים שתי תוצאות זהות $(X_1,Y_1)=(X_2,Y_2)$, אזי אין התאמה ואין גם אי התאמה ביניהן. כמו כן, אם מתקבלות שתי התוצאות $(X_1,Y_1)=a$ ו-$(X_2,Y_2)=c$, הרי שעבורן גם כן אין התאמה

ואין אי התאמה, כיוון שערכו של X זהה בשתי הנקודות. ההסתברות להתאמה בהתפלגות זו היא, אפוא, רק ההסתברות לקבלת שתי התצפיות a ו-b (לא חשוב באיזה סדר). נרשום זאת במונחים של המשתנים (X_1,Y_1) ו-(X_2,Y_2):

$$\pi_c = P\left\{(X_1=-1,Y_1=0)\cap(X_2=0,Y_2=1)\right\}$$

$$+P\left\{(X_1=0,Y_1=1)\cap(X_2=-1,Y_2=0)\right\}=2\cdot\frac{1}{3}\cdot\frac{1}{3}=\frac{2}{9}$$

ההסתברות לחוסר התאמה היא ההסתברות לקבלת שתי הנקודות b ו-c בציור. מטעמי סימטרייה הסתברות זו שווה, כמובן, להסתברות להתאמה. כלומר, $\pi_d=\pi_c=2/9$. סכום שתי ההסתברויות הוא $\pi_c+\pi_d=4/9$.

ההסתברות **המותנית** להתאמה, כשנתון שבין שני הזוגות אין תיקו, היא ההסתברות להתאמה מחולקת בהסתברות שאין תיקו, כלומר,

$$\tilde{\pi}_c=\frac{\pi_c}{\pi_c+\pi_d}=\frac{2/9}{4/9}=\frac{1}{2} \quad \text{וכך גם} \quad \tilde{\pi}_d=\frac{\pi_d}{\pi_c+\pi_d}=\frac{2/9}{4/9}=\frac{1}{2}.$$

(שימו לב שוב, שסכום שתי ההסתברויות המותנות הללו הוא 1.)

מכאן מקדם המתאם של קנדל במקרה זה הוא 0 $\tilde{\tau}=\tilde{\pi}_c-\tilde{\pi}_d=0$.

קיבלנו התאפסות של המתאם, אם כי המשתנים X ו-Y הם, כמובן, תלויים.

הערה: בדוגמה זו מקדם המתאם ρ גם הוא 0. החישוב קל מאוד: פונקציית ההסתברות השולית של X היא אחידה: $P(X=K)=1/3$, עבור $k=-1,\ 0,\ 1$, והתוחלת שלו $EX=0$.

פונקציית ההסתברות השולית של Y היא $P(Y=0)=2/3$, $P(Y=1)=1/3$, והתוחלת שלו $EY=1/3$. המכפלה XY מקבלת רק את הערך 0 (בכל שלוש הנקודות a, b ו-c אחד הערכים הוא 0). מכאן מקבלים את השונות המשותפת:

$$Cov(X,Y)=EXY-EXEY=0$$

ולכן גם המתאם מתאפס:

$$\rho=\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}=0$$

סטטיסטי המבחן של קנדל במקרה של ערכי תיקו

אמידת המתאם של קנדל $\tilde{\tau}=\tilde{\pi}_c-\tilde{\pi}_d$ אינה מובנת מאליה. האומד הטבעי ביותר הוא, על פי הנוסחה (30),

$$\tilde{T}=\frac{N_c-N_d}{N_c+N_d} \qquad (31)$$

כאשר N_c הוא מספר כל הזוגות המתאימים, N_d הוא מספר כל הזוגות הלא מתאימים, וכן N_c+N_d הוא מספר כל הזוגות שאין ביניהם ערכי תיקו (לא עבור ה-x-ים ולא עבור ה-y-ים). האומד $\tilde{T}$ מקבל ערכים בין -1 לבין 1.
(בספרות מקובל לסמן את המקדם $\tilde{T}$ ב-γ.)
יש לשים לב שהמכנה של $\tilde{T}$ אינו קבוע, והוא תלוי בתוצאות המדגם.

אומד אחר, המקובל מאוד למקרי תיקו, הוא האומד הבא:

$$\tau_b=\frac{N_c-N_d}{\sqrt{[n(n-1)/2-t^{**}]\cdot[n(n-1)/2-u^{**}]}} \qquad (32)$$

כאשר $t^{**}=\frac{1}{2}\sum_j t_j(t_j-1)$, $u^{**}=\frac{1}{2}\sum_l u_l(u_l-1)$.

שימו לב שאם במדגם אין ערכי תיקו, האומד τ_b זהה לאומד המקורי T, שהוגדר בנוסחה (23).

ישנן הצעות נוספות לאמידת המתאם $\tilde{\tau}$ על ידי תקנונים שונים מאלה שהבאנו לעיל. לא נדון בהם כאן. נאמר רק שכדי לבחון את השערת האי־תלות, די להסתכל על המונה של האומדים – N_c-N_d כסטטיסטי המבחן ולהשתמש עבורו בקירוב נורמלי.
חישוב הסטטיסטי N_c-N_d עבור מדגם של תצפיות עם ערכי תיקו בין חלק מן ה-x-ים או חלק מן ה-y-ים (או בשניהם) קצת פחות פשוט מהמקרה הרציף. עלינו לקחת בחשבון בתהליך הספירה רק את הזוגות שבהם גם שני ערכי X שונים זה מזה וגם שני ערכי Y שונים זה מזה. השוואות בין זוגות שבהם אחד המשתנים לפחות מקבל ערך זהה, אינן נלקחות בחשבון. נסתכל על הדוגמה הבאה.

דוגמה 6.9. להלן דוגמה פיקטיבית של נתונים, עם דרגות הרשומות בלוח 6.7.
נסתכל, לדוגמה, על התצפית השנייה $(i=2)$. לחישוב $N_{c(2)}$ עלינו לספור את כל הדרגות $\tilde{S}_i$ מתחת לקו המקווקו, שהן גדולות (ממש) מהדרגה $\tilde{S}_2=4.5$. ישנה דרגה אחת כזאת ולכן $N_{c(2)}=1$. לחישוב $N_{d(2)}$ עלינו לספור את כל הדרגות $\tilde{S}_i$ מתחת לקו המקווקו, שהן קטנות (ממש) מהדרגה 4.5 וישנן שתיים כאלה. אולם אחת מהן שייכת להשוואה בין $i=2$ לבין $i=3$, ששם דרגות ערכי x הן שוות (ל-2.5), ולכן רק אחת מההשוואות היא ללא ערכי תיקו. מכאן מקבלים $N_{d(2)}=1$.
בסך הכול קיבלנו $N_c+N_d=9$ זוגות שביניהם אין שום ערך של תיקו, מתוכם $N_c=6$ זוגות מתאימים ו-$N_d=3$ אינם מתאימים. הסטטיסטי של קנדל (31) המתקבל הוא $\tilde{T}=\frac{6-3}{9}=\frac{1}{3}$.

לוח 6.7. *חישוב הסטטיסטי של קנדל כשיש ערכי תיקו*

i	$\tilde{R}_i$	$\tilde{S}_i$	$N_{c(i)}$	$N_{d(i)}$	$N_{c(i)}+N_{d(i)}$
1	1	2	3	0	3
2	2.5	4.5	1	1	2
3	2.5	2	2	0	2
4	4	6	0	2	2
5	5.5	4.5	0	0	0
6	5.5	2	–	–	–
סך הכול	21	21	6	3	9

נחשב גם את הסטטיסטי τ_b מנוסחה (32). הביטויים הדרושים:

$$n(n-1)/2=6(5)/2=15,\quad t^{**}=2\cdot 2(1)/2=2,\quad u^{**}=[2(1)+3(2)]/2=4$$

והסטטיסטי המתקבל הוא $\tau_b=\dfrac{6-3}{\sqrt{(15-2)(15-4)}}=0.251$.

יש הבדל די גדול בין שני אומדני המתאם של קנדל על פי נתוני הדוגמה. היחס בין שני האומדנים אינו מפתיע. מתברר שהמכנה של τ_b תמיד גדול או שווה ל-N_c+N_d ולכן תמיד מתקיים $\tau_b\le\tilde{T}$.
לא נוכיח זאת כאן.

נביא עתה דוגמה נוספת לחישוב הסטטיסטי של קנדל, למקרה שישנם ערכי תיקו רבים. לשם כך נחזור לדוגמה 6.5.

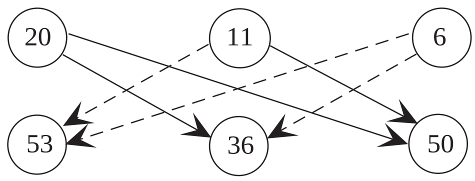
דוגמה 6.10 (*המשך דוגמה 6.5*). נסתכל שוב על לוח השכיחות של נשים בהיריון לפי מידת הדתיות וכוונת ההנקה שלהן. כל שתי נשים הכלולות בתא מסוים הן בעלות אותו ערך של X ואותו ערך של Y. כמו כן, כל אלה הכלולות באותה שורה הן בעלות ערך זהה של X, ואלה הכלולות באותה עמודה הן בעלות ערך זהה של Y. לכן, כדי להשוות שתי נשים שאין ביניהן שום תיקו, עלינו להשוות רק כאלה שאינן כלולות באותה שורה וגם לא באותה עמודה.
נוכל להשוות, למשל נשים הכלולות בשני התאים המסומנים בעיגול. ניקח אישה בתא המסומן בעיגול בהיר (אינה מתכוונת להיניק וחילונית) ונשווה אותה עם אישה בתא המסומן בעיגול כהה (מתכוונת להיניק ומסורתית). בתא הכהה כל אחד מהמשתנים גדול

מזה שבתא הבהיר (רמת ההנקה גבוהה יותר והאישה יותר דתית). לפיכך שתי נשים בתאים הללו הן מתאימות. לעומת זאת אישה בתא הכהה ואישה בתא המסומן במרובע (אינה מתכוונת להיניק ודתית) אינן מתאימיות (רמת ההנקה של האישה בתא המרובע נמוכה יותר ואילו רמת הדתיות שלה גבוהה יותר).

לוח 6.8. *שכיחויות נשים בהריון לפי דתיות והנקה*

הנקה (X)	חילונית	מסורתית	דתית	סך הכול
	דתיות (Y)			
לא	20	11	$\boxed{6}$	37
כן	53	36	50	139
סך הכול	73	47	56	176

באופן כללי השוואת נשים כאן ניתנת להיעשות רק לגבי שני תאים "אלכסוניים" – יש התאמה כאשר האלכסון יורד בכיוון ימין ואין התאמה כאשר האלכסון יורד בכיוון שמאל. נביא כאן שוב (ציור 6.10) את השכיחויות מלוח 6.8, עם חצים המסמנים כיווני ההתאמה ואי התאמה.



ציור 6.10. *השוואות בין תאים.* קו רצוף – *יש התאמה, קו מקווקו – אין התאמה*

כדי למצוא את הסטטיסטי של קנדל, יש לספור את כל הזוגות המתאימים של נשים ואת כל הזוגות הלא המתאימים של נשים. מספר הזוגות **המתאימים** הוא סכום המכפלות של השכיחויות בתאים המקושרים בחץ עם קו רצוף בציור 6.10:

$$N_c = 20 \cdot 36 + 20 \cdot 50 + 11 \cdot 50 = 2{,}270$$

מספר הזוגות **הלא מתאימים** הוא סכום המכפלות של השכיחויות בתאים המקושרים בחץ עם קו מקווקו:

$$N_d = 11 \cdot 53 + 6 \cdot 53 + 6 \cdot 36 = 1{,}117$$

מכאן מקבלים את מקדם המתאם של קנדל, לפי נוסחה (31):

$$\tilde{T} = \frac{2{,}270 - 1{,}117}{2{,}270 + 1{,}117} = \frac{1{,}153}{3{,}387} = 0.3404$$

שימו לב שעבור מדגם של 176 נשים, מספר ההשוואות הכולל בין שתי נשים הוא $(176)(175)/2 = 15,400$. אולם בגלל קבוצות התיקו הגדולות בדוגמה זו, מספר כל ההשוואות ללא תיקו הוא רק 3,387.

לחישוב הערך של τ_b בדוגמה זו, גודלי קבוצות התיקו הן השכיחויות בשולי הטבלה בלוח 6.8: $t_1 = 37,\ t_2 = 139\ ;\quad u_1 = 73,\ u_2 = 47,\ u_3 = 56$

מכאן מקבלים את הביטויים הנחוצים:

$$t^{**} = \frac{1}{2}[37(36) + 139(138)] = 10{,}257$$

$$u^{**} = \frac{1}{2}[73(72) + 47(46) + 56(55)] = 5{,}249$$

$$n(n-1)/2 = (176)(175)/2 = 15{,}400$$

מכאן נקבל את המדד, לפי נוסחה (32)

$$\tau_b = \frac{N_c - N_d}{\sqrt{[n(n-1)/2 - t^{**}]\cdot[n(n-1)/2 - u^{**}]}}$$

$$= \frac{1{,}153}{\sqrt{[15{,}400 - 10{,}257]\cdot[15{,}400 - 5249]}} = .1596$$

גם בדוגמה זו $\tau_b \le \tilde{T}$.

הקירוב הנורמלי במקרה של תיקו

כדי להשתמש בקירוב הנורמלי עבור $N_c - N_d$ עלינו לדעת את התוחלת והשונות של הסטטיסטי בהינתן ערכי התיקו. תחת השערת האפס התוחלת אינה משתנה (ההתפלגות עדין סימטרית סביב אפס). השונות משתנה כפונקציה של גודל קבוצות התיקו וניתן לחשב אותה באופן הבא:

$$Var(N_c - N_d) = \frac{n(n-1)(2n+5)}{18} - \frac{1}{18}\left[\sum_j t_j(t_j - 1)(2t_j + 5) + \sum_k u_k(u_k - 1)(2u_k + 5)\right] + \frac{1}{9n(n-1(n-2)}\left[\sum_j t_j(t_j - 1)(t_j - 2)\right] \times \left[\sum_j u_j(u_j - 1)(u_j - 2)\right] \qquad (33)$$

$$+\frac{1}{2n(n-1)}\left[\sum_j t_j(t_j-1)\right]\times\left[\sum_j u_j(u_j-1)\right]$$

לא נוכיח את הנוסחה הזאת כאן.

שימו לב שהמחובר הראשון בנוסחה (33) הוא בדיוק השונות $Var(N_c-N_d)$ במקרה הרציף, ללא ערכי תיקו, כפי שהיא מוצגת בנוסחה (25).

נשתמש בנוסחה (33) כדי לחשב את מובהקות התוצאה מדוגמה **6.10**.

דוגמה 6.11 (המשך דוגמה 6.10). הסטטיסטי שהתקבל בדוגמה **6.10** הוא $N_c-N_d=1{,}153$, וכן גודלי קבוצות התיקו הן השכיחויות בשולי הטבלה בלוח **6.6**:

$$u_1=73,\ u_2=47,\ u_3=56\ \ ;t_1=37,\ t_2=139$$

נחשב בנפרד כל אחד מהמחוברים בנוסחת השונות (33).

א) $$\frac{n(n-1)(2n+5)}{18}=\frac{176(175)(357)}{18}=617{,}848$$

ב) $$\frac{1}{18}\left[\sum_j t_j(t_j-1)(2t_j+5)+\sum_l u_l(u_l-1)(2u_l+5)\right]=390{,}598$$

ג) $$\frac{1}{9n(n-1(n-2)}\left[\sum_j t_j(t_j-1)(t_j-2)\right]\times\left[\sum_l u_l(u_l-1)(u_l-2)\right]$$
$$=36{,}591.35$$

ד) $$\frac{1}{2n(n-1)}\left[\sum_j t_j(t_j-1)\right]\times\left[\sum_l u_l(u_l-1)\right]=3{,}496.04$$

נצרף את התוצאות הללו לפי הנוסחה (33) ונקבל את השונות:

$$Var(N_c-N_d)=617{,}848-390{,}598+36{,}591.35+3{,}496.04$$
$$=267{,}337.39$$

המובהקות המקורבת של התוצאה היא (מבחן דו־צדדי)

$$P=2P_{H_0}(N_c-N_d\geq 1{,}153)\cong 2\left[1-\Phi\left(\frac{1{,}152}{\sqrt{267{,}337.39}}\right)\right]=.0258$$

התוצאה מובהקת. נזכיר שמבחן ספירמן בדוגמה **6.5** נתן תוצאה כמעט זהה. יש אפוא קשר בין רמת הדתיות לבין כוונת ההנקה, כאשר אישה דתייה יותר נוטה להיניק יותר מאישה פחות דתייה.

יעילות יחסית אסימפטוטית

היעילות היחסית האסימפטוטית של המבחן המבוסס על המתאם של קנדל ביחס למבחן המבוסס על המתאם של פירסון היא $9/\pi^2 \cong .912$, במודל הדו-נורמלי (שבו השימוש במתאם של פירסון הוא מתאים). יעילות זו זהה ליעילות היחסית המתאימה של מתאם ספירמן.

תרגילים

1. כל אחד משבעה נבדקים קיבל שתי משימות. להלן הציונים שהתקבלו:

הנבדק:	7	6	5	4	3	2	1
משימה א:	20	19	31	6	9	18	27
משימה ב:	31	12	32	10	11	13	30

חשבו את המתאם של ספירמן ואת המתאם של קנדל בין ציוני שתי המשימות. מה מובהקות התוצאה שהתקבלה בכל אחת מן השיטות? הסיקו אם ניתן לומר שיש קשר בין ביצועי שתי המשימות, או שהן בלתי תלויות ($\alpha = .05$).

2. הוכיחו שאם דרגות ה-y-ים בדיוק הפוכות מדרגות ה-x-ים, אזי $r_s = -1$.

3. חשבו את ההתפלגות המדויקת של מתאם ספירמן ומתאם קנדל, תחת השערת אי-תלות, עבור מדגם של 5 תצפיות, שבו דרגות ה-x-ים היו $1, 2.5, 2.5, 4, 5$ ודרגות ה-y-ים $1, 3, 3, 3, 5$.

4. כל אחד מ-12 סטודנטים עבר בחינה באנגלית. הציונים לפי שנת הלימודים באוניברסיטה היו

שנה ראשונה	70	78	88	71
שנה שניה:	91	72	69	79
שנה שלישית:	89	84	81	93

השערת החוקר הייתה שידיעת אנגלית **משתפרת** עם השנים באוניברסיטה. בדקו את ההשערה בר"מ של 5%.

5. חוקר אסף נתונים כדי להצביע על קשר שקיים בין גודל של יישוב לבין הופעת מגיפות של שפעת. הוא טוען כי ככל שהיישוב קטן, כך עולה שכיחות ההופעות של מגיפות שפעת. הנתונים להלן הינם לגבי ערים באנגליה, בשנים 1956-1940.

זמן ממוצע בין מגיפות (שבועות)	אוכלוסייה (אלפים)	עיר	זמן ממוצע בין מגיפות (שבועות)	אוכלוסייה (אלפים)	עיר
79	12	K	73	1046	A
98	11	L	106	658	B
199	7.1	M	92	415	C
149	5.6	N	93	269	D
105	4.0	O	94	180	E
>284	3.5	P	80	113	F
191	3.1	Q	74	66	G
>284	2.6	R	75	65	H
175	1.7	S	86	22	I
			92	18	J

נסחו השערות מתאימות ובדקו בר"מ $\alpha = .01$.

6. במהלך מחקר שנערך לשם ניסוי של תכניות התערבות בגני ילדים (ד"ר דורית ארם) נאספו נתונים שונים הקשורים בקריאה וכתיבה. לוח השכיחות כאן מציג את הציונים של ילדי גן טרום טרום חובה בתחילת הניסוי לגבי המשתנים: X – ציון בכתיבת שם; Y – ציון בזיהוי אותיות.
בְּדקו אם יש קשר בין יכולת בכתיבת השם לבין יכולת זיהוי אותיות. עִרכו את הבדיקה בעזרת מבחן ספירמן ובעזרת מבחן קנדל. מה המסקנה?

X (כתיבת שם): 6	4	3	2	1	Y (זיהוי אותיות)
0	1	29	10	7	1
2	1	12	5	2	2
2	0	6	2	0	3

פרק 7

ניתוח דו־כיווני – בלוקים אקראיים

כשחוקר מעוניין לבדוק אם יש הבדל בין כמה טיפולים, הוא יכול לתכנן את הדגימה בצורות שונות. אפשר לבחור מדגמים מקריים בלתי תלויים ולתת לנבדקים בכל מדגם טיפול שונה. את ניתוח התוצאות בתכנון כזה ניתן לעשות בעזרת המבחן של קרוסקל-וואליס (פרק 5). ברור שבתכנון כזה הבדלים גדולים בין נבדקים שונים שקיבלו אותו טיפול (שונויות גדולות **בתוך** המדגמים) יכולים לגרום לכך שהבדלים בין המדגמים השונים עלולים להיטשטש, כלומר, להתקבל קטנים יחסית, למרות שלמעשה יש הבדל בין הטיפולים. פתרון לבעיה זו ניתן למצוא, בין השאר, על ידי "זיווג" של הטיפולים, כפי שראינו בתכנון של מדגמים מזווגים לבדיקת הבדל בין שני טיפולים (פרק 4).
בדיוק באותה צורה שבה הסתכלנו על מדגם מזווג בהשוואה לשני מדגמים בלתי תלויים, כך נסתכל עתה על מדגם של בלוקים אקראיים בהשוואה למספר כלשהו של מדגמים בלתי תלויים.

דוגמה 7.1. מובאים כאן נתונים חלקיים מתוך מחקר שנערך בבתי ספר בישראל (ברכה בירן) בנושא הערכת סמכות ידע של מורים בארבעה מקצועות שונים (ספרות, היסטוריה, מתמטיקה וביולוגיה). לגבי כל ארבעת המורים, ציין כל תלמיד עד כמה הוא סומך עליהם, כלומר, מאמין ובוטח בידע הכללי שלהם במקצוע שהם מלמדים. הציון שניתן לכל מורה הוא ממוצע ההערכות של תלמיד לגבי תשעה פריטים שונים (למשל: יש לו ידע רב; הוא מקפיד לדייק בעבודות; אפשר לסמוך על הידע שלו בלב שלם), כאשר בכל פריט ההערכות האפשריות היו בתחום 1-6.
התוצאות הרשומות בלוח 7.1 הן הציונים הכיתתיים (ממוצעי ההערכה של תלמידי הכיתה) לגבי ארבע מורות, בכל אחת מעשר כיתות י. (אף לא אחת מהמורות אינה מלמדת בשתי כיתות שונות שנכללו במדגם.)

לוח 7.1. *ממוצעים כיתתיים של הערכת מורות במדגם של עשר כיתות י*

ביולוגיה	מתמטיקה	היסטוריה	ספרות	הכיתה
מקצוע ההוראה				
4.18	4.30	4.61	4.52	1
4.33	4.76	4.57	4.35	2
4.22	4.97	5.01	4.60	3
5.41	3.05	5.19	4.33	4
3.87	4.73	4.72	4.35	5
5.09	4.34	4.68	4.30	6
5.26	5.00	4.56	4.45	7
2.47	4.56	5.13	4.10	8
4.01	5.53	4.54	3.94	9
4.08	5.08	4.96	4.25	10

נשים לב שעשר הכיתות שנבדקו הן בלתי תלויות זו בזו, אבל בכל כיתה הציונים של המורות השונות הם, כמובן, תלויים, כיוון שתלמידי כיתה מסוימת רשמו את הערכותיהם בדיוק לאותן ארבע מורות המלמדות בכיתה את המקצועות הללו. כדי להשוות בין המקצועות, יש לקחת בחשבון שהציונים הכיתתיים הם למעשה בלוקים של 4 תצפיות (המקצועות) לגבי כל אחת מהכיתות בנפרד. ציוני כל כיתה לגבי ארבע המורות מהווים בלוק. במונחים של תכנון הניסוי מדובר כאן ב-10 בלוקים אקראיים (הכיתות) של 4 "טיפולים" (המקצועות). בעיית המחקר הייתה לבחון אם יש הבדל בין מורות במקצועות שונים לגבי הערכת הידע שלהן.

הצגת הבעיה כבעיה סטטיסטית

נתונים n בלוקים ("נבדקים") של k איברים ("טיפולים"), כאשר n הבלוקים הם בלתי תלויים זה בזה.

נסמן את התצפיות על ידי

X_{ij} – התצפית של הטיפול j עבור הנבדק ה-i, $i=1,\dots,n;\ j=1,\dots,k$.

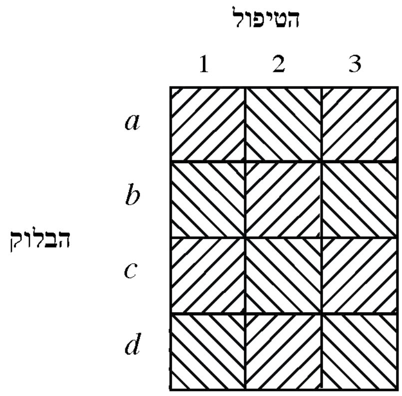
נניח שהתפלגות X_{ij} היא F_j עבור כל i, $i=1,\dots,n$.

את השערת האפס, שאין כל הבדל בין הטיפולים נרשום:

$$H_0: F_1(t)=F_2(t)=\cdots=F_k(t) \qquad \text{לכל } t \qquad (1)$$

תכנונים כאלה של ניסוי אנו מוצאים לעתים קרובות בחקלאות, שם בוחנים הבדל בין שיטות טיפול שונות (דישון, השקיה וכד') על ידי כך שעורכים את הניסוי ב-n חלקות, כאשר כל חלקה מהווה בלוק, שבו מנסים את כל k הטיפולים. ראו לדוגמה ציור 7.1

שבו יש 4 בלוקים, כל אחד של 3 טיפולים.



ציור 7.1. *ארבעה בלוקים של שלושה טיפולים*

7.1 מבחן פרידמן

סטטיסטי המבחן של פרידמן (Friedman, 1937) מבוסס על הדרגות של התצפיות בכל אחד מהבלוקים בנפרד. אנו מדרגים מ-1 עד k את התצפיות בכל בלוק.

נסמן: R_{ij} – הדרגה של X_{ij} מבין $X_{i1},X_{i2},\ldots,X_{ik}$. כלומר, R_{ij} היא הדרגה של הטיפול ה-j מבין כל k הטיפולים של הנבדק (הבלוק) ה-i.

למעשה, הדרגות נקבעות על ידי הסדר בין הטיפולים אצל כל אחד מהנבדקים.

נסמן את סכום הדרגות שקיבל הטיפול ה-j אצל כל הנבדקים ב-T_j: $T_j = \sum_{i=1}^{n} R_{ij}$.

ברור שאם יש הבדל גדול בין הטיפולים, הסדר בין התצפיות יהיה דומה אצל כל הנבדקים ואז יהיה הבדל גדול בין סכומי הדרגות של הטיפולים השונים. נראה, לדוגמה, שני סידורים שונים של דרגות ארבעה בלוקים עבור שלושה טיפולים בלוחות 7.2א ו-7.2ב.

בלוח 7.2א הדרגות של כל הבלוקים הן זהות, באופן שטיפול 3 עדיף על טיפול 2 וזה עדיף על טיפול 1 אצל כל הנבדקים. סכומי הדרגות של כל הבלוקים (הסכומים של שלוש העמודות) מאוד שונים זה מזה. בלוח 7.2ב הדירוגים בַבלוקים שונים זה מזה, וכך סכומי הדרגות של שלוש העמודות הם זהים. שימו לב שסכום הדרגות בכל בלוק (שורה) הוא קבוע, כיוון שזה בדיוק סכום הערכים $1+2+3=6$, כאשר הסדר אינו חשוב.

באופן כללי, בבעיה עם n בלוקים ו-k טיפולים, סכום הדרגות בכל בלוק הוא קבוע –

$$\sum_{j=1}^{k} R_{ij} = 1+2+\cdots+k = \frac{k(k+1)}{2}$$

וכך סך כל הדרגות בטבלה הוא תמיד

$$\sum_{j=1}^{k} T_j = \sum_{i=1}^{n}\sum_{j=1}^{k} R_{ij} = \frac{nk(k+1)}{2} \qquad (2)$$

לוח 7.2א

הבלוק	הטיפול 1	2	3	סך הכול
a	1	2	3	6
b	1	2	3	6
c	1	2	3	6
d	1	2	3	6
T_j	4	8	12	24

לוח 7.2ב

הבלוק	הטיפול 1	2	3	סך הכול
a	1	2	3	6
b	2	3	1	6
c	3	2	1	6
d	2	1	3	6
T_j	8	8	8	24

סכום הדרגות שקיבלו הטיפולים השונים תלוי, כמובן, בתוצאות הניסוי.
אם השערת האפס נכונה, אין הבדל בין הטיפולים ולכן נצַפה לקבל טבלת דרגות הדומה ללוח 7.2ב, כלומר, סכומי הדרגות של הטיפולים השונים יהיו דומים. הסטטיסטי של פרידמן לבדיקת השערת האפס הוא מדד להבדל בין סכומי הדרגות הללו והוא ניתן על ידי

$$S = \sum_{j=1}^{k} (T_j - \overline{T})^2 = \sum_{j=1}^{k}\left[T_j - \frac{n(k+1)}{2}\right]^2 \qquad (3)$$

הביטוי (3) הוא סכום ריבועי הסטיות של $T_1, T_2, \ldots, T_k$ מהממוצע שלהם, $\overline{T}$, שערכו, לפי (2) הוא קבוע:

$$\overline{T} = \frac{1}{k}\sum_{j=1}^{k} T_j = \frac{n(k+1)}{2} \qquad (4)$$

ערכים גבוהים של הסטטיסטי S מנוסחה (3) מעידים על פער גדול בין הדרגות שניתנו לטיפולים השונים, ולכן אנו דוחים את השערת האפס עבור ערכים גבוהים של S.

טענה 7.1. את S ניתן לחשב גם על ידי הנוסחה

$$S = \sum_{j=1}^{k} T_j^2 - \frac{n^2 k(k+1)^2}{4} \qquad (5)$$

ההוכחה קלה ואנו משאירים זאת לבדיקתכם.

בדוגמה 7.2 נחשב את הערך של הסטטיסטי של פרידמן עבור נתוני דוגמה 7.1.

דוגמה 7.2 (המשך דוגמה 7.1). בלוח 7.3 מוצגות הדרגות של ציוני המורות בארבעת המקצועות בכל אחת מהכיתות במדגם. עבור כל כיתה בנפרד מדרגים את ציוני 4 המורות במקצועות השונים מ-1 עד 4. לפי הסימונים לעיל, בדוגמה זו $k=4$ ו-$n=10$.

לוח 7.3. *דרגות הערכת מורות במקצועות השונים עבור עשר כיתות*

הכיתה	מקצוע ההוראה: ספרות	היסטוריה	מתמטיקה	ביולוגיה	סך הכול
1	3	4	2	1	10
2	2	3	4	1	10
3	2	4	3	1	10
4	2	3	1	4	10
5	2	3	4	1	10
6	1	3	2	4	10
7	1	2	3	4	10
8	2	4	3	1	10
9	1	3	4	2	10
10	2	3	4	1	10
סך הכול T_j	18	32	30	20	100

ממוצע סכומי הדרגות הוא, לפי נוסחה (4): $\bar{T}=\frac{n(k+1)}{2}=\frac{10(5)}{2}=25$, או, פשוט, $\bar{T}=\frac{100}{4}=25$. מכאן מקבלים את הסטטיסטי S לפי נוסחה (3)

$$S=\sum_{j=1}^{k}(T_j-\bar{T})^2$$

$$=(18-25)^2+(32-25)^2+(30-25)^2+(20-25)^2=148$$

ניתן לחשב את הסטטיסטי גם לפי נוסחה (5). מקבלים אותו ערך, כמובן:

$$S=18^2+32^2+30^2+20^2-\frac{10^2(4)(5^2)}{4}=148$$

התפלגות S תחת השערת האפס

בהנחה שהשערת האפס (1) נכונה, הדירוג של הטיפולים בכל אחד מהבלוקים הוא מקרי ולכן בכל בלוק כל $k!$ התמורות של הדרגות הן בעלות אותה הסתברות. היות ש-n הבלוקים השונים הם בלתי תלויים זה בזה, ההסתברות לקבלת טבלת דרגות מסוימת היא מכפלת ההסתברויות של n הבלוקים המרכיבים אותה. בסך הכול קיימות $(k!)^n$ קומבינציות שונות של דירוגי k הטיפולים ב-n בלוקים ולכן התפלגות S היא

$$P_{H_0}(S=s)=\frac{\#\{S=s\}}{(k!)^n}$$

ברוב הספרים הדנים בשיטות אפרמטריות אין בנמצא טבלאות של התפלגות S עבור יותר מ-3 טיפולים ויותר מ-8 בלוקים. לכן, בדרך כלל, נזדקק לקירוב כדי לחשב את מובהקות התוצאה.

ההתפלגות המקורבת של S

בדומה למקרה של מדגמים בלתי תלויים, גם בניסוי עם בלוקים אקראיים, תחת השערת האפס המשתנה S, כשהוא מתוקנן באופן מתאים, מתפלג בקירוב לפי התפלגות חי־בריבוע עם $k-1$ דרגות חופש. התִקנון המתאים מוגדר במשפט הבא.

משפט 7.1. תחת השערת האפס (1) של שוויון התפלגויות, הסטטיסטי של פרידמן מתפלג אסימפטוטית לפי התפלגות חי־בריבוע. במִלים אחרות, תחת השערת האפס המשתנה המתוקנן

$$F_r=\frac{12S}{nk(k+1)} \tag{6}$$

מתפלג בקירוב χ^2_{k-1} (חי־בריבוע עם $k-1$ דרגות חופש), בתנאי שמספר הבלוקים n די גדול.

לא נוכיח כאן את המשפט, אולם נסביר בהמשך את התִקנון שנעשָׂה.

מסקנה: קל להראות שאם משתמשים בנוסחה (5) לחישוב S, הסטטיסטי המתוקנן בנוסחה (6) יכול להירשם בצורה הבאה:

$$F_r=\frac{12\sum_{j=1}^{k}T_j^2}{nk(k+1)}-3n(k+1) \tag{7}$$

הנמקה לתִקנון

נסתכל על התפלגות הדרגות של התצפיות הבודדות – R_{ij} בבלוק מסוים i. תחת השערת האפס שאין הבדל בין הטיפולים, כל $k!$ התמורות של הדרגות בבלוק הן שוות הסתברות, ולכן לכל j הדרגה R_{ij} שקיבל הטיפול ה-j היא בעלת התפלגות אחידה $R_{ij} \sim U(1,k)$. מכאן לכל $i=1,\ldots,n$ ולכל $j=1,\ldots,k$ התוחלת של הדרגה R_{ij} היא $ER_{ij}=(k+1)/2$, והשונות היא $Var(R_{ij})=(k^2-1)/12$.

[שימו לב שהתוחלת לעיל היא גם הדרגה הממוצעת בכל הטבלה:

$$\bar{R}=\frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n}R_{ij}=\frac{1}{nk}\cdot\frac{nk(k+1)}{2}=\frac{k+1}{2}\text{ .]}$$

התוחלת של סכום דרגות כאלה עבור אותו טיפול היא סכום התוחלות:

$$ET_j = E\sum_{i=1}^{n}R_{ij} = \sum_{i=1}^{n}ER_{ij} = \frac{n(k+1)}{2}$$

[תוחלת זו היא גם הממוצע $\bar{T}$ של k סכומי הדרגות $T_1,\ldots,T_k$, נוסחה (4).]

כמו כן, הבלוקים (או הנבדקים) הם בלתי תלויים ולכן גם שונות סכום הדרגות שקיבל אותו טיפול בכל הבלוקים שווה לסכום השונויות:

$$Var(T_j)=Var\left(\sum_{i=1}^{n}R_{ij}\right)=\sum_{i=1}^{n}Var(R_{ij})=\frac{n(k^2-1)}{12}$$

לפיכך התקנון המתאים של T_j הוא

$$Z_j=\frac{T_j-ET_j}{\sqrt{Var(T_j)}}=\frac{T_j-\frac{n(k+1)}{2}}{\sqrt{\frac{n(k^2-1)}{12}}}$$

והריבוע שלו הוא

$$Z_j^2=\frac{12\left[T_j-\frac{n(k+1)}{2}\right]^2}{n(k^2-1)}$$

סכום k הריבועים הללו דומה לַמשתנה F_r, המוגדר בנוסחה (6), כאשר ההבדל הוא רק בַמכנה:

$$\sum_{j=1}^{k}Z_j^2=\frac{12\sum_{j=1}^{k}\left[T_j-\frac{n(k+1)}{2}\right]^2}{n(k^2-1)}=\frac{12S}{n(k^2-1)}=F_r\frac{k-1}{k}$$

כלומר, F_r הוא סכום ריבועים של משתנים המתפלגים בקירוב נורמלית סטנדרטית, הכופל בקבוע $k/(k-1)$. ההכפלה בקבוע נדרשת מכיוון שסכומי הדרגות של הטיפולים השונים אינם בלתי תלויים.

הערה: את הסטטיסטי S ניתן לרשום גם בעזרת סכום הריבועים בין ממוצעי הדרגות, SSB, (Between-groups Sum of Squares) כפי שהוגדר בפרק 5, משפט 5.1.

נסמן את הדרגה הממוצעת שקיבל טיפול j על ידי $\bar{R}_j = \frac{1}{k}\sum_{i=1}^{k} R_{ij} = \frac{T_j}{k}$. כאן נוכל לרשום את SSB באופן הבא:

$$SSB = \sum_{j=1}^{k} n_j (\bar{R}_j - \bar{R})^2 = n \sum_{j=1}^{k} \left[\frac{T_j}{n} - \frac{(k+1)}{2} \right]^2 = \frac{1}{n} S \qquad (8)$$

כאשר $\bar{R} = \frac{k+1}{2}$ היא הדרגה הממוצעת בכל הטבלה.

ולכן הסטטיסטי של פרידמן הוא, למעשה, מהצורה $S = nSSB$.

הגדרה 7.1. בבעיית בלוקים אקראיים ממוצע ריבועי הסטיות של כל הדרגות מהממוצע שלהן (Total Mean Squares), המסומן MST, מוגדר על ידי

$$MST = \frac{SST}{n(k-1)} = \frac{1}{n(k-1)} \sum_{i=1}^{n} \sum_{j=1}^{k} \left(R_{ij} - \bar{R} \right)^2 \qquad (9)$$

המכנה של MST הוא מספר "דרגות החופש" של סכום הריבועים, כאשר בכל בלוק בנפרד יש $k-1$ "דרגות חופש" לסך הסטיות הריבועיות מהממוצע הכולל (שהוא גם ממוצע הבלוק).

טענה 7.2. את המשתנה המתוקנן F_r ניתן לרשום על ידי $F_r = \frac{SSB}{MST}$, כאשר SSB ו-MST מוגדרים בנוסחאות (8) ו-(9).

[זוהי גם הצורה שבה רשמנו את המשתנה H של קרוסקל-וואליס, נוסחה (11) בפרק 5, משפט 5.1, עבור הבעיה של השוואת k מדגמים בלתי תלויים, אלא ששם ל- MST אין אותן דרגות חופש כמו פה.]

הוכחה: נראה זאת על ידי חישוב סכום הריבועים MST.

כדי לחשב את סך הסטיות הריבועיות SST, נסתכל על בלוק מסוים i.

סכום ריבועי הסטיות של כל אחת מהדרגות בבלוק מהממוצע הכללי $\bar{R}$ הוא

$$\sum_{j=1}^{k} \left(R_{ij} - \frac{k+1}{2} \right)^2$$

ממוצע הדרגות בבלוק גם הוא $(k+1)/2$, כמובן, ולכן סכום הסטיות הריבועיות הזה הוא השונות של משתנה אחיד $U(1,k)$ הכופלת ב-k. מכאן מקבלים

$$\sum_{j=1}^{k} \left(R_{ij} - \frac{k+1}{2} \right)^2 = \frac{k(k^2-1)}{12}$$

סכום ריבועי הסטיות הללו בכל הטבלה הוא:

$$SST=\sum_{i=1}^{n}\sum_{j=1}^{k}\left(R_{ij}-\overline{R}\right)^2=\sum_{i=1}^{n}\frac{k(k^2-1)}{12}=\frac{nk(k^2-1)}{12} \qquad (10)$$

מכאן ממוצע הריבועים הכולל הוא

$$MST=\frac{SST}{n(k-1)}=\frac{k(k+1)}{12} \qquad (11)$$

הביטוי הזה הוא **קבוע** ואינו תלוי בתוצאות הניסוי.

את הסטטיסטי המתוקנן של פרידמן מקבלים על ידי המנה של סכומי הריבועים:

$$\frac{SSB}{MST}=\frac{S/n}{k(k+1)/12}=\frac{12S}{nk(k+1)}=F_r$$

קיבלנו, אפוא, את הטענה. ♣

ההתפלגות המקורבת של F_r היא, אפוא, לפי משפט 7.1, התפלגות חי־בריבוע עם $k-1$ דרגות חופש.

כדי לחשב את המובהקות המקורבת של התוצאה בדוגמה 7.2 יש למצוא את F_r. קיבלנו בדוגמה 7.2 את התוצאה $S=148$, כאשר $n=10,\ k=4$. מכאן הסטטיסטי המתוקנן הוא

$$F_r=\frac{12(148)}{10(4)(5)}=8.88$$

מובהקות התוצאה היא בקירוב $P=P_{H_0}(F_r\geq 8.88)\cong P(\chi^2_3\geq 8.88)$.

על פי טבלת חי־בריבוע עם 3 דרגות חופש נמצא $.025<P<.05$, ולפי ההתפלגות המדויקת של חי־בריבוע, המובהקות המקורבת היא $P(\chi^2_3\geq 10.95)=.0309$.

המסקנה היא שיש הבדל בין מורות במקצועות שונים מבחינת הערכת התלמידים לגבי סמכות הידע שלהן במקצוע שאותו הן מלמדות.

השוואות מרובות

כמו במקרה של מדגמים בלתי תלויים, אם תוצאת המבחן להשוואת k הטיפולים היא מובהקת, נהיה מעוניינים בדרך כלל לערוך השוואות בין כל הזוגות של הטיפולים (או חלקם), כדי לנסות ולזהות את הטיפולים העדיפים.

להשוואה בין שני טיפולים ניתן להשתמש במבחן המתאים למדגמים מזווגים – למשל, מבחן הסימן, או מבחן ווילקוקסון. נשתמש בשיטת בונפרוני, כפי שהוגדרה בפרק 5.2, כדי לתקן את רמת המובהקות של השוואה בודדת, בגלל ריבוי המבחנים. אם עורכים את כל $\frac{k(k+1)}{2}=\binom{k}{2}$ ההשוואות הזוגיות, ונדרשת רמת מובהקות כללית $\mathrm{FWE}=\alpha$, אזי רמת המובהקות הספציפית של השוואה בודדת צריכה להיות $\alpha_i=\frac{\alpha}{k(k+1)/2}$.

ההשוואות הפשוטות ביותר במקרה זה הן באמצעות מבחן הסימן.
נפעיל עתה את שיטת בונפרוני לגבי דוגמת ההערכות של המורות במקצועות השונים.

דוגמה 7.3 (הנמשך דוגמה 7.2). באמצעות מבחן פרידמן ההבדל בין המורות בארבעת המקצועות נמצא מובהק. נשווה עתה בין כל הזוגות של המקצועות, ברמת מובהקות כללית $\text{FWE} = .10$. מספר ההשוואות שאנו עורכים הוא $4\cdot3/2=6$ ולכן רמת המובהקות הדרושה לכל השוואה בודדת, לפי כלל התיקון של בונפרוני, היא $\alpha_i=.10/6=.0167$. השוואה בודדת היא מובהקת אם הסתברות הזנב (מבחן דו־צדדי) קטנה מ-$0.0167/2=.0083$.
לפי טבלת מבחן הסימן, טבלה 3 בנספח, עבור $n=10$, $P(S_{10}\le1)=.0107$ אולם $P(S_{10}\le2)=.0547$ ולכן השוואה בודדת היא מובהקת רק עבור $S_{10}\le1$ או $S_{10}\ge9$. (כלומר, לפחות ב-9 מתוך 10 הכיתות אחד המקצועות צריך לקבל ציון גבוה מזה של המקצוע השני.)
מלוח 7.3 אנו מוצאים שרק ההבדל בין מורות לספרות ומורות להיסטוריה הוא מובהק, $S_{10}=10$, באופן שהערכת סמכות הידע של מורות להיסטוריה גבוהה מזו של מורות לספרות. כל שאר ההשוואות אינן מובהקות. (בִּדקו את כל ההשוואות וּודאו שאין תוצאה מובהקת נוספת.)

הערה: את ההשוואות הזוגיות ניתן לערוך באמצעות מבחן ווילקוקסון למדגם מזווג במקום מבחן הסימן. נראה יותר מאוחר בפרק זה (פרק 7.5) את הסיבה לכך שערכנו דווקא את מבחן הסימן.

7.2 מקדם ההסכמה בין שופטים

בניסוי של בלוקים אקראיים, כדי להשוות בין הטיפולים אנו מדרגים את הטיפולים בכל בלוק בנפרד. נניח שהדירוג הזה נעשה על ידי n שופטים, המדרגים k "טיפולים" (מכוניות שונות, מתחרים בתחרות "מלך השרירים", או תרופות שונות להקלת כאבי ראש). נרצה לשאול עד כמה יש הסכמה בין כל השופטים לגבי דירוג הטיפולים.
המדד למידת ההסכמה בין השופטים נקרא **מקדם ההסכמה** והוא מוגדר באופן הבא.

הגדרה 7.2: מקדם ההסכמה מסומן W והוא מוגדר על ידי

$$W=\frac{12S}{n^2k(k^2-1)}=\frac{F_r}{n(k-1)} \qquad (12)$$

המשפט הבא מסביר את משמעות הנוסחה (12).

משפט 7.2. מקדם ההסכמה הוא המנה בין הערך של S שהתקבל בניסוי לבין הערך המקסימלי שהסטטיסטי S יכול לקבל בניסוי כזה. כלומר, אם בניסוי התקבל $S=s$, אזי $W=\frac{s}{\max(S)}$.

*הוכחה**: נוכיח ראשית שהסטטיסטי S מקבל את ערכו המקסימלי כאשר כל הדירוגים שווים, כמו, למשל, כאשר בכל הבלוקים $R_{i1}=1,\ R_{i2}=2,\ \dots,\ R_{ik}=k$.

נזכור ש- SST הוא למעשה קבוע, לפי נוסחה (10). מצד שני, ניתן לרשום אותו בצורה אחרת, על ידי פירוק סכום הריבועים באופן הבא.

נסמן: $\bar{R}_j=\frac{T_j}{n}$ – הדרגה הממוצעת שקיבל טיפול j, $j=1,\dots,k$.

את SST נפרק על ידי חיסור והוספה של הדרגה הממוצעת הזאת בכל אחד מהמחוברים. [החלפנו את סדר הסיכום של השורות והעמודות בנוסחה (10)]:

$$SST=\sum_{j=1}^{k}\sum_{i=1}^{n}\left(R_{ij}-\bar{R}\right)^2=\sum_{j=1}^{k}\sum_{i=1}^{n}\left(R_{ij}-\bar{R}_j+\bar{R}_j-\bar{R}\right)^2$$

$$=\sum_{j=1}^{k}\sum_{i=1}^{n}\left(R_{ij}-\bar{R}_j\right)^2+\sum_{j=1}^{k}n\left(\bar{R}_j-\bar{R}\right)^2$$

$$+2\sum_{j=1}^{k}\left(\bar{R}_j-\bar{R}\right)\sum_{i=1}^{n}\left(R_{ij}-\bar{R}_j\right)$$

המחובר האחרון מתאפס, כיוון שמתקיים

$$\sum_{i=1}^{n}\left(R_{ij}-\bar{R}_j\right)=T_j-n\bar{R}_j=T_j-n\frac{T_j}{n}=0 \qquad \text{לכל } j,$$

בסך הכול קיבלנו את SST, שהוא קבוע ואינו תלוי בתוצאות הניסוי, כסכום של שני סכומי ריבועים, שאותם ניתן לרשום באופן הבא:

$$SST=\sum_{j=1}^{k}\sum_{i=1}^{n}\left(R_{ij}-\bar{R}_j\right)^2+n\sum_{j=1}^{k}\left(\bar{R}_j-\bar{R}\right)^2$$

$$=SSW+SSB=\frac{nk(k^2-1)}{12}$$

המחובר הראשון, שמסומן SSW, הוא סכום עבור כל הטיפולים של ריבועי הסטיות בין דרגות הבלוקים בטיפול זה מהדרגה הממוצעת של אותו הטיפול. במִלים אחרות, זה סכום ריבועי הסטיות בין הבלוקים **בתוך** הטיפולים. SSB הוא סכום ריבועי הסטיות **בין** הטיפולים, כפי שראינו קודם, נוסחה (8). הביטוי האחרון באגף ימין הוא הערך (הקבוע) של SST לפי נוסחה (10). הראינו קודם בנוסחה (8) ש- $SSB=S/n$ ולכן קיבלנו כאן

$$\frac{nk(k^2-1)}{12}=SST=SSW+\frac{S}{n}$$

מכיוון ש- $SSW\geq 0$ תמיד, כסכום של ריבועים, והסכום של שני המחוברים לעיל הוא קבוע, S יהיה מקסימלי כאשר SSW יהיה מינימלי, או, כאשר $SSW = 0$. במקרה זה מקבלים את הערך $S/n=SST$, וזה הערך המקסימלי האפשרי של S/n. מכאן מתקבל המקסימום של S

$$\max S=nSST=\frac{n^2k(k^2-1)}{12} \qquad (13)$$

המנה בין S לבין הערך המקסימלי של S היא

$$\frac{S}{\max S}=\frac{12S}{n^2k(k^2-1)}=W$$

כאשר W הוגדר בנוסחה (12). ♣

הערה: שימו לב שהערך המקסימלי של S מתקבל כאשר $SSW=0$ וזה קורה רק כאשר $R_{ij}=\overline{R}_j=T_j/n$ לכל $i=1,\ldots,n$ ולכל $j=1,\ldots,k$, כלומר, כאשר כל אחד מהטיפולים מקבל אותה דרגה בכל הבלוקים.

מסקנה 7.1. ממשפט 7.2 ברור שמקדם ההסכמה מקבל ערכים $0\leq W\leq 1$. הערך המינימלי $W=0$ מתקבל רק כאשר $S=0$, כלומר, כאשר סכומי הדרגות בכל הטיפולים הם שווים. הערך המקסימלי $W=1$ מתקבל כאשר כל n הדירוגים בבלוקים הם זהים, ואז סכומי הדרגות הם, למשל, $R_1=n,R_2=2n,\ldots,R_k=kn$ וסך ריבועי הסטיות שלהם מהממוצע הוא הגדול ביותר האפשרי.

נחשב, לדוגמה, את מקדם ההסכמה בין הכיתות לגבי הערכת סמכות המורות (דוגמה 7.2). שם $k=4,\ n=10$ והתוצאה שהתקבלה היא $S=148$.

$$W=\frac{12S}{n^2k(k^2-1)}=\frac{12(148)}{10^2(4)(4^2-1)}=0.296$$

ההסכמה כאן די נמוכה. כלומר, הכיתות לא כל כך דומות זו לזו מבחינת דירוג ההערכות של המורות במקצועות השונים. (כנראה שמידת ההערכה תלויה גם במורה עצמה...)

משמעות מקדם ההסכמה. לפי המתואר לעיל, מקדם ההסכמה גבוה כאשר S גבוה, כלומר, כאשר יש הבדלים גדולים בין סכומי הדרגות של הטיפולים השונים. אם אין הסכמה בין השופטים, יתקבלו דירוגים שונים בבלוקים השונים, סכומי הדרגות יהיו די

דומים, ואז S יהיה קטן. כאשר יש הסכמה מלאה בין השופטים לגבי דירוג הטיפולים, S יקבל את הערך המקסימלי האפשרי, ואז נקבל $W=1$. בדיקת השערה על מידת ההסכמה בין שופטים לגבי טיפולים מסוימים אקוויוולנטית, אפוא, לבדיקת השערה על הבדל בין הטיפולים. סטטיסטי המבחן F_r משמש גם לבדיקת השערת האפס שאין כל הסכמה בין השופטים (הדירוג בבלוקים השונים אקראי) כנגד האלטרנטיבה שישנה מידה של הסכמה בין השופטים (או – הטיפולים אינם שווים).

ניקח, לדוגמה, את תחרות "מלך השרירים", שבה הגיעו לגמר 5 גברים והם דורגו על ידי שבעה שופטים באופן בלתי תלוי זה בזה. הגבר שזכה במקום ראשון בתחרות הוא זה שסכום הדרגות שניתנו לו היה הגבוה ביותר. עם זאת, נניח שמקדם ההסכמה בין השופטים היה $W=.1$. הסטטיסטי של פרידמן לבדיקת ההבדל בין המתחרים הוא $F_r=Wn(k-1)=2.8$. המובהקות של תוצאה זו היא מאוד גבוהה, $P>.50$, כך שתוצאה זו כלל אינה מובהקת.

במקרה כזה נוכל להסיק שבין השופטים כלל לא הייתה הסכמה לגבי הדירוג של המתחרים, ולמעשה, הזוכה במקום הראשון זכה בתואר רק במקרה (היה לו מזל גדול!).

7.3 בעיות של ערכי תיקו במבחן פרידמן

אם בחלק מדירוגי הבלוקים יש ערכי תיקו, משתמשים, כמו בכל המקרים הקודמים, בדרגה הממוצעת.

מסמנים $\tilde{R}_{ij}$ – הדרגה (הממוצעת) של התצפית X_{ij}, $i=1,\ldots,n;\ j=1,\ldots,k.$

סכום הדרגות (הממוצעות) של הטיפולים מסומן $\tilde{T}_j=\sum_{i=1}^{n}\tilde{R}_{ij}$.

היות שסכום הדרגות בכל בלוק נשמר כמו במקרה שאין תיקו, ממוצע סכומי הדרגות אינו משתנה והסטטיסטי S המחושב על סמך סכומי הדרגות של הטיפולים הוא

$$\tilde{S}=\sum_{j=1}^{k}\left[\tilde{T}_j-\frac{n(k+1)}{2}\right]^2 \qquad (14)$$

ניתן לחשב אותו, בדומה לנוסחה (5) על ידי

$$\tilde{S}=\sum_{j=1}^{k}(\tilde{T}_j)^2-\frac{n^2k(k+1)^2}{4} \qquad (15)$$

כדי לקבל משתנה בעל התפלגות מקורבת חי־בריבוע, יש לתקנן את $\tilde{S}$ על ידי ממוצע הריבועים MST, כמו במקרה שאין תיקו, טענה 7.2. מצורת החישוב של MST בנוסחה (11) ברור שסכום הריבועים הכולל SST תלוי בגודל קבוצות התיקו בכל הבלוקים, באופן שכבר ראינו בבעיות קודמות:

$$SST = \frac{nk(k^2-1)}{12} - \frac{1}{12k}\sum_j \sum_l t_{jl}(t_{jl}^2-1) \qquad (16)$$

כאשר t_{jl} הוא גודל קבוצת התיקו ה-l בבלוק j.

מכאן מקבלים את סטטיסטי המבחן (בדקו!):

$$\tilde{F}_r = \frac{SSB}{MST} = \frac{12\tilde{S}}{nk(k+1) - \frac{1}{k-1}\sum_j \sum_l t_{jl}(t_{jl}^2-1)}$$

$$= \frac{Fr}{1 - \frac{1}{nk(k^2-1)}\sum_j \sum_l t_{jl}(t_{jl}^2-1)} \qquad (17)$$

דוגמה 7.4. נסתכל שוב על הניסוי שהצגנו בפרק 4, דוגמה 4.5, לגבי מידת השימוש בִמלות שאלה אצל פעוטות. בניסוי המקורי נרשם, פרט לִמלים "מה?" ו"איפה?", גם השימוש בִמלות השאלה "מה זה?" ו"מי זה?".

מספר הפעמים שהפעוטות השתמשו בכל אחת מ-4 המִלים הללו נתון בלוח 7.4.

לוח 7.4. *מספר הפעמים שפעוטות השתמשו ב-4 מִלות שאלה*

מי זה	מה זה	איפה	מה	הפעוט
0	1	1	0	1
2	0	4	1	2
5	17	4	1	3
2	15	4	3	4
0	7	3	1	5
2	0	0	1	6
1	0	5	0	7
1	2	7	2	8
0	2	6	3	9
0	4	8	1	10
0	0	0	0	11
0	8	7	9	12
0	4	1	2	13
0	12	3	0	14
0	1	0	0	15
0	11	2	6	16
15	37	18	14	17

נרצה להשוות בין ארבע מִלות השאלה, כדי לברר מהן המִלים שבהן הפעוטות משתמשים בתדירות גבוהה יותר. דירוג המִלים אצל כל אחד מהפעוטות בנפרד נתון בלוח 7.5.
מספר ה"טיפולים" כאן הוא $k=4$ מִלים ומספר הבלוקים, או ה"נבדקים", הוא $n=17$ פעוטות.

לוח 7.5. *דרגות מידת השימוש ב-4 מִלות שאלה*

הפעוט	מה	איפה	מה זה	מי זה	סך הכול
1	1.5	3.5	3.5	1.5	10
2	2	4	1	3	10
3	1	2	4	3	10
4	2	3	4	1	10
5	2	3	4	1	10
6	3	1.5	1.5	4	10
7	1.5	4	1.5	3	10
8	2.5	4	2.5	1	10
9	3	4	2	1	10
10	2	4	3	1	10
11	2.5	2.5	2.5	2.5	10
12	4	2	3	1	10
13	3	2	4	1	10
14	1.5	3	4	1.5	10
15	2	2	4	2	10
16	3	2	4	1	10
17	1	3	4	2	10
סך הכול T_j	37.5	49.5	52.5	30.5	170

ממוצע סכומי הדרגות הוא $\bar{T}=\frac{1}{4}\sum_{j=1}^{4}T_j=\frac{170}{4}=42.5$.
הסטטיסטי של פרידמן, לפי נוסחה (14):

$$\tilde{S}=\sum_{j=1}^{k}(\tilde{T}_j-\bar{T})^2=(37.5-42.5)^2+\cdots+(30.5-42.5)^2=318$$

קבוצות התיקו t_{il} ב-17 הבלוקים הן בגדלים הבאים: בבלוק 1 – $2, 2$; בבלוק 6 – 2 ; בבלוק 7 – 2 ; בבלוק 8 – 2 ; בבלוק 11 – 4 ; בבלוק 14 – 2.

התיקון הדרוש: $\sum\sum t_{ij}(t_{ij}^2-1)=6[2\cdot 3]+[4\cdot 15]=96$

מכאן סטטיסטי המבחן מתקבל, לפי נוסחה (17):

$$\tilde{F}_r = \frac{12\tilde{S}}{nk(k+1) - \frac{1}{k-1}\sum_j \sum_l t_{jl}(t_{jl}^2 - 1)} = \frac{12(318)}{17(4)(5) - \frac{96}{3}} = 12.39$$

המובהקות המקורבת של תוצאה זו מתקבלת על סמך טבלת התפלגות חי־בריבוע עם 3 דרגות חופש: $0.005 < P < 0.01$. המובהקות קטנה מאוד, ולכן התוצאה מאוד מובהקת. ניתן להסיק שיש הבדל בין תדירות השימוש של פעוטות במִלים השונות.

לפי שיטת בונפרוני, אם רוצים לערוך השוואות בין כל הזוגות של 4 המִלים, ברמת מובהקות כללית של $\alpha = .10$, יש לערוך 6 מבחנים, שכל אחד מהם בעל רמת מובהקות $\alpha_i = 0.10/6 = 0.0167$. היות שהמבחנים הם דו־צדדיים, כדי שהשוואה זוגית ספציפית תהיה מובהקת, הסתברות הזנב המתאים, $P/2$, צריכה להיות נמוכה מ- $0.0167/2 = .0083$.

נשתמש שוב במבחן הסימן להשוואת כל שתי מִלים. בלוח 7.6 מוצגים ערכי S_n, הערך של n הרלוונטי (ללא הפרשים שערכם אפס) והסתברות הזנב של התוצאה, לפי טבלה 3 בנספח. מכל ההשוואות שערכנו, רק ההשוואה בין "איפה" לבין "מי זה" היא מובהקת. פעוטות נוטים לשאול "איפה" לעתים קרובות יותר מאשר "מי זה". בין כל שאר המִלים אין הבדל מובהק.

לוח 7.6. *תוצאות ההשוואות בין כל שתי מִלים באמצעות מבחן הסימן*

	מי זה			מה זה			איפה		
	$P/2$	n	S_n	$P/2$	n	S_n	$P/2$	n	S_n
מה	.2905	13	8	.0898	14	4	.0592	15	4
איפה	.0037*	15	13	.2120	14	5			
מה זה	.0106	16	13						

* תוצאה מובהקת

יש כאן, לכאורה, סתירה לנתונים, מכיוון שסך הדרגות של המִלה "מה זה" גבוה מזה של המִלה "איפה", ולמרות זאת לא התקבל הבדל מובהק בין "מה זה" לבין "מי זה" (המִלה שדורגה נמוך ביותר). הסיבה היא, כמובן, הבדלים בין מספר ערכי התיקו בהשוואות בין המִלים.

מקדם ההסכמה במקרי תיקו

נדון בצורת החישוב של מקדם ההסכמה במקרים של תיקו בניסוי, בהתייחס לדוגמה 7.4, שם התקבל $\tilde{S}=318$. ניתן לחשב את הערך של W בשלושה אופנים שונים.

א) חישוב לפי הנוסחה (12) שהתקבלה עבור בעיה ללא תיקו. מקבלים ערך מקורב

$$\tilde{W}_1=\frac{\tilde{S}}{n^2k(k^2-1)/12}=\frac{12(318)}{17^2(4)(15)}=\frac{3816}{17{,}340}=0.220$$

ב) לתקן את SST לפי נוסחה (16), ולהפעיל את הקשר הרשום בנוסחה (13) עבור המקסימום. הנוסחה המתקבלת עבור מקדם ההסכמה היא

$$\tilde{W}_2=\frac{\tilde{S}}{nSST}=\frac{12\tilde{S}}{n^2k(k^2-1)-n\sum t(t^2-1)} \qquad (18)$$

ערך זה תמיד גדול מהערך שמתקבל עבור מקדם ההסכמה ללא התיקון. התוצאה בדוגמה 7.4 היא

$$\tilde{W}_2=\frac{3{,}816}{17{,}340-17(96)}=\frac{3{,}816}{15{,}708}=0.2429$$

נזכיר כי בתהליך הוכחת משפט 7.2 הערך המקסימלי של S התקבל על סמך הפירוק $SST=SSW+SSB$, כאשר $SSB=S/n$, ומכאן שהנוסחה (13) עבור הערך המקסימלי הזה, התקבלה למקרה שבו $SSW=0$. עם זאת, יש להעיר שבמקרים שבהם קיימים ערכי תיקו, לרוב אי אפשר להשיג את המינימום $SSW=0$, מכיוון שבדרך כלל אי אפשר לסדר את הדרגות (הממוצעות) שהתקבלו כך, שיהיו **בדיוק זהות** אצל כל השופטים. ראו, למשל, את הדרגות בלוח 7.5, שם לא בכל שורה רשומות אותן 4 דרגות. בהתאם לכך, כשנתונה רשימת הדירוגים שניתנו על ידי n השופטים, הערך המינימלי של SSW יכול להיות חיובי, ולכן הערך המקסימלי של SSB יהיה נמוך מ-SST. במקרים כאלה $nSST$ אינו בדיוק המקסימום האפשרי של $\tilde{S}$, כפי שהיה כאשר לא היו מקרי תיקו. הערך שבו חילקנו את $\tilde{S}$ כאן הוא

$$nSST=\frac{n^2k(k^2-1)-n\sum t(t^2-1)}{12}=\frac{15{,}708}{12}=1{,}309$$

ונראה מיד שערך זה אינו ערך אפשרי עבור הדרגות הממוצעות שהתקבלו בניסוי.

ג) ניתן לחשב את הערך המדויק של $\tilde{W}$ לפי הערך המקסימלי האפשרי של $\tilde{S}$, בהתחשב בערכי התיקו הספציפיים שהתקבלו למעשה במדגם. את הערך המקסימלי האפשרי של $\tilde{S}$ בדוגמה 7.4 אפשר למצוא על ידי רישום הדירוגים שהתקבלו אצל כל אחד מהפעוטות בסדר דומה ככל האפשר. באופן כזה מתקבלים ההבדלים הגדולים ביותר בין סכומי הדרגות. עשינו זאת בלוח 7.7 עבור הדרגות המוצגות

בלוח 7.5. בכל שורה סודרו הדירוגים הספציפיים שהתקבלו אצל אותו הפעוט מהדירוג הנמוך ביותר, עבור המִלה "מה", ועד הגבוה ביותר, עבור המִלה "מה זה". למשל, אצל פעוט מספר 1 התקבלו הדרגות: 1.5, 3, 3, 1.5 והן סודרו בלוח 7.7 בסדר עולה: 1.5, 1.5, 3, 3.

לוח 7.7. *הדרגות הרשומות בלוח 7.5 כשהן מסודרות בסדר עולה אצל כל אחד מהפעוטות*

הפעוט	מה	איפה	מה זה	מי זה	סך הכול
1	1.5	1.5	3.5	3.5	10
2	1	2	3	4	10
3	1	2	3	4	10
4	1	2	3	4	10
5	1	2	3	4	10
6	1.5	1.5	3	4	10
7	1.5	1.5	3	4	10
8	1	2.5	2.5	4	10
9	1	2	3	4	10
10	1	2	3	4	10
11	2.5	2.5	2.5	2.5	10
12	1	2	3	4	10
13	1	2	3	4	10
14	1.5	1.5	3	4	10
15	2	2	2	4	10
16	1	2	3	4	10
17	1	2	3	4	10
סך הכול T_j	21.5	33.0	49.5	66.0	170

הערך המקסימלי האפשרי של $\tilde{S}$ בניסוי זה הוא הערך המחושב מטבלת הדירוגים כפי שהם מסודרים בלוח 7.7. כאן קיים: $\sum_{j=1}^{4}\tilde{T}_j^2 = 8{,}357.5$ ולכן

$$\max\tilde{S} = \sum_{j=1}^{k}(\tilde{T}_j - 42.5)^2 = 1{,}132.5$$

הערך המקסימלי של $\tilde{S}$ הוא, אפוא, נמוך מהערך 1,309 שהיה מתקבל אילו היה ניתן להשיג $SSW = 0$. מקדם ההסכמה המדויק, עם תיקון לתיקו, הוא, אפוא,

$$\tilde{W}_3 = \frac{\tilde{S}}{\max\tilde{S}} = \frac{318}{1{,}132.5} = 0.2808$$

התוצאה קצת גבוהה מזו שהתקבלה על פי הנוסחה (18).

החישוב כפי שערכנו כאן דורש בדרך כלל עבודה רבה ואין זה מקובל להשתמש בו.

הנוסחאות המקובלות לחישוב מקדם ההסכמה הן, אפוא:

במקרה שאין ערכי תיקו
$$W=\frac{S}{nSST}=\frac{12S}{n^2k(k^2-1)}$$

כשישנם ערכי תיקו, נוסחה (18) לעיל
$$\tilde{W}=\frac{\tilde{S}}{nSST}=\frac{12\tilde{S}}{n^2k(k^2-1)-n\sum t(t^2-1)}$$

7.4 הקשר בין מקדם ההסכמה לבין מתאם הדרגות של ספירמן

נסתכל על מקרה פרטי של שני בלוקים בלבד. את לוח הדרגות של k הטיפולים ניתן לרשום כמו בלוח 7.9.

טענה 7.3. עבור המקרה של שני בלוקים בלבד, $n=2$, הסטטיסטי של פרידמן מתקבל על ידי

$$S=\frac{k(k^2-1)}{6}(1+r_s) \qquad (19)$$

הוכחה: נחשב את הסטטיסטי של פרידמן מן הטבלה של שני הבלוקים, בלוח 87.

לוח 7.8. *הדרגות בשני בלוקים*

הבלוק	1	2	…	k	סך הכול
1	U_1	U_2	…	U_k	$k(k+1)/2$
2	V_1	V_2	…	V_k	$k(k+1)/2$
סך הכול R_j	U_1+V_1	U_2+V_2	…	U_k+V_k	$k(k+1)$

סך הדרגות של הטיפול ה-j הוא $T_j=U_j+V_j$, $j=1,\dots,k$, וממוצע k סכומי הדרגות

הללו הוא $k(k+1)/k=(k+1)$.

סך ריבועי הסטיות מהממוצע:

$$S=\sum_{j=1}^{k}[U_j+V_j-(k+1)]^2=\sum_{j=1}^{k}\left[U_j-\frac{k+1}{2}+V_j-\frac{k+1}{2}\right]^2$$
$$=\sum_{j=1}^{k}\left[U_j-\frac{k+1}{2}\right]^2+\sum_{j=1}^{k}\left[V_j-\frac{k+1}{2}\right]^2$$
$$+2\sum_{j=1}^{k}\left[U_j-\frac{k+1}{2}\right]\cdot\left[V_j-\frac{k+1}{2}\right]$$

שני המחוברים הראשונים הם סך ריבועי הסטיות של איברי הבלוק מממוצע הבלוק, כלומר, $k(k^2-1)\big/12$, והמחובר השלישי הוא המונה של מתאם הדרגות של ספירמן, r_s, נוסחה (5) בפרק 6. מכאן מקבלים

$$S=2\frac{k(k^2-1)}{12}+2r_s\frac{k(k^2-1)}{12}=\frac{k(k^2-1)}{6}(1+r_s)$$

אנו רואים, אפוא, שבמקרה של שני בלוקים בלבד, הסטטיסטי של פרידמן הוא פונקציה ליניארית של מתאם הדרגות של ספירמן. ♣

את מקדם ההסכמה עבור שני בלוקים מקבלים על ידי הצבת הערך S כפונקציה של מקדם ההסכמה W כפי שמוגדר בנוסחה (12):

$$W=\frac{12S}{n^2k(k^2-1)}=\frac{1+r_s}{2}$$

הערך המקסימלי של S מתקבל, כמובן, כאשר שני הדירוגים זהים, כלומר, כאשר $r_s=1$.

במקרה זה הקשר בין מבחן פרידמן למתאם ספירמן ברור. ככל ששני הדירוגים דומים יותר, כך S גדול יותר ויחד עם זאת r_s גדול יותר. מצד שני, אם הדירוגים אינם דומים, או שהם לגמרי הפוכים, הערך של S הוא נמוך וכן גם r_s יכול להיות קרוב לאפס, או אפילו שלילי.

לפי הנוסחה (19), עבור $n=2$ בלוקים מוצאים את הקשרים הבאים:

א) אם $r_s=1$ אזי $S=\dfrac{k(k^2-1)}{3}$ ולכן $W=\dfrac{12S}{n^2k(k^2-1)}=1$.

ב) אם $r_s=0$ אזי $S=\dfrac{k(k^2-1)}{6}$ לכן $W=\dfrac{1}{2}$.

ג) אם $r_s=-1$ אזי $S=0$ וכך גם $W=0$.

הערך המינימלי של הסטטיסטי של פרידמן מתקבל כאשר המתאם הוא -1, $r_s=-1$, ואז

מקדם ההסכמה הוא אפס.

קשר דומה קיים בין הסטטיסטי של פרידמן לבין מתאמי ספירמן גם כשמדובר במספר גדול יותר של בלוקים, $n>2$. נסתכל על ניסוי של n בלוקים ו-k טיפולים, כאשר $k \ge 2$ ו-$n \ge 2$.

נסמן: r_{il} – מתאם הדרגות של ספירמן בין הבלוק ה-i לבלוק ה-l, $(i \ne l)$, ונסמן את הממוצע של המתאמים הללו בין כל הזוגות של דירוגי בלוקים שונים על ידי

$$\bar{r} = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{\substack{l=1 \\ l>i}}^{n} r_{il}$$

הקשר בין ממוצע המתאמים לבין הסטטיסטי של פרידמן מוצג במשפט 7.3.

משפט 7.3. הקשר בין ממוצע מתאמי ספירמן לבין מקדם ההסכמה ניתן על ידי

$$\bar{r} = \frac{nW-1}{n-1} \qquad (20)$$

או

$$W = \frac{(n-1)\bar{r}+1}{n} \qquad (21)$$

ההוכחה קצת ארוכה ולכן אנו מביאים אותה בנספח 7א.

הצבת $n=2$ בנוסחה (21) אמנם נותנת את הערך של W כפונקציה של r_s כפי שהתקבלה עבור שני בלוקים (במקרה כזה יש רק מתאם אחד ולכן $\bar{r} = r_s$).

הערות:

1) בדומה למה שמצאנו עבור שני בלוקים (טענה 7.3), מקדם ההסכמה הוא פונקציה ליניארית של ממוצע מתאמי הדרגות של ספירמן.

2) אם $n>2$ ממוצע המתאמים אינו יכול לקבל את הערך -1. נסתכל, לדוגמה, על ניסוי עם שלושה בלוקים. במקרה כזה ישנם שלושה מתאמים זוגיים – r_{12}, r_{13} ו-r_{23}. הממוצע שלהם יכול לקבל את הערך $\bar{r}=1$ רק אם כל אחד מהמתאמים הבודדים שווה ל-1, כלומר, אם כל הדירוגים זהים. אולם, כדי שהממוצע יהיה שווה ל--1 חייבים כל שלושת המתאמים להיות שווים ל--1, מה שאיננו אפשרי. נשים לב ש-$r_{12}=-1$ רק אם הדרגות בבלוק הראשון הפוכות מאלה שבבלוק השני. באותו אופן $r_{13}=-1$ רק אם הדרגות בבלוק הראשון הפוכות מאלה שבבלוק השלישי. אבל מכאן נובע, כמובן, שדרגות הבלוק השני והשלישי זהות, כלומר $r_{23}=1$. הממוצע של שלושת המתאמים במקרה זה הוא $\bar{r}=-1/3$.

3) ניתן למצוא חסם עבור ממוצע המתאמים על סמך הקשר של ממוצע זה עם מקדם ההסכמה W. היות ש-$0 \le W \le 1$, נקבל מנוסחה (20) עבור $\bar{r}$:

$$-\frac{1}{n-1} \le \bar{r} \le 1$$

4) הנוסחה (21) נותנת הסבר אינטואיטיבי להגדרת מקדם ההסכמה W. ככל שהדירוגים בבלוקים השונים דומים יותר זה לזה, כך מקדם ההסכמה גבוה יותר.

7.5 מבחן פרידמן עבור שני טיפולים, הקשר למבחן הסימן

המקרה $k=2$ הוא מקרה פרטי של ניסוי עם n בלוקים אקראיים, שבו רק שני טיפולים. למעשה זה מודל של מדגם מזווג, שבניתוח המתאים לבעיה דנו בפרק 4. מצד שני, היות שמבחן פרידמן מתאים לכל בעיה של בלוקים אקראיים עם $k \ge 2$, נראה כאן מהו הסטטיסטי של פרידמן המתקבל בבעיה ספציפית זו.

הדרגות בכל אחד מהבלוקים הן 1 ו-2, כאשר התצפית הגדולה יותר מקבלת דרגה 2.

נסמן את המשתנה U – מספר הבלוקים שעבורם התצפית השנייה גדולה מן הראשונה, כלומר:

$$U=\#\{1 \le i \le n : R_{i1}=1, R_{i2}=2\}$$

סכום הדרגות בכל אחד משני הטיפולים, מתקבל:

$$T_1=\sum_{i=1}^{n} R_{i1}=\{i:R_{i1}=1\}+2\{i:R_{i1}=2\}=U+2(n-U)=2n-U$$

$$T_2=\sum_{i=1}^{n} R_{i2}=\{i:R_{i2}=1\}+2\{i:R_{i2}=2\}=n-U+2U=n+T$$

ואמנם הסכום הכולל של הדרגות הוא $T_1+T_2=3n$ (בכל בלוק סכום הדרגות הוא 3).

הסטיות של סכומי הדרגות מהממוצע הן

$$T_1-\frac{3n}{2}=2n-U-\frac{3n}{2}=\frac{1}{2}n-U$$

$$T_2-\frac{3n}{2}=n+U-\frac{3n}{2}=U-\frac{1}{2}n$$

(את הסטייה השנייה ניתן היה לחשב ישירות מהסטייה הראשונה, כיוון שסכום הסטיות מהממוצע שווה תמיד לאפס.)

אנו רואים ששתי הסטיות זהות בערכן המוחלט. הסטטיסטי S מתקבל על ידי סכום ריבועי הסטיות הללו:

$$S=\left(T_1-\frac{3n}{2}\right)^2+\left(T_2-\frac{3n}{2}\right)^2=2\left(U-\frac{1}{2}n\right)^2$$

קיבלנו את הסטטיסטי של פרידמן כפונקציה של מספר הבלוקים שעבורם התצפית

השניה גדולה מהראשונה. נשים לב כי המשתנה U הוא בדיוק הסטטיסטי S_n של מבחן הסימן בבעיה של מדגם מזווג.

דחייה עבור ערך גבוה של הסטטיסטי S של פרידמן אקוויוולנטית לדחייה עבור **סטייה מוחלטת** גבוהה של S_n מתוחלתו $n/2$. כלומר, זהו אזור דחייה דו־צדדי של מבחן הסימן.

נסתכל עתה על המשתנה המתוקנן F_r מנוסחה (6), כאשר נציב $U=S_n$.

$$F_r=\frac{12S}{nk(k+1)}=\frac{12S}{n\cdot 6}=\frac{2S}{n}=\frac{4\left(S_n-n/2\right)^2}{n}=\frac{\left(S_n-n/2\right)^2}{n/4} \qquad (22)$$

אם משתמשים במבחן הסימן לבעיה, סטטיסטי המבחן המתוקנן עבור הקירוב הנורמלי הוא $Z=\frac{S_n-n/2}{\sqrt{n/4}}$. [נזכיר שתחת השערת האפס של שוויון בין שני הטיפולים $ES_n=n/2$, $Var(S_n)=n/4$.]

לפיכך הסטטיסטי F_r הרשום בנוסחה (22) הוא בדיוק Z^2, כלומר, ריבוע הערך המתוקנן של S_n.

הוכחנו בכך את הטענה הבאה:

טענה 7.4. במקרה של שני טיפולים בלבד ($k=2$), הסטטיסטי F_r של פרידמן שווה לריבוע הסטטיסטי של מבחן הסימן, המתוקנן תחת השערת האפס:

$$F_r=\frac{\left(S_n-n/2\right)^2}{n/4}.$$

עבור n די גדול $F_r\sim\chi^2_{k-1}$ וכאשר $k=2$, $F_r\sim\chi^2_1$. עבור מבחן ברמת מובהקות α יש לדחות את השערת האפס כאשר $F_r>\chi^2_{1,1-\alpha}$. אבל ערך גדול של F_r אקוויוולנטי, לפי טענה 7.4, לערך גדול של Z^2. לכן אזור הדחייה לעיל הוא למעשה אזור דחייה דו־צדדי של מבחן הסימן. יש לדחות את השערת האפס כאשר $\frac{\left(S_n-n/2\right)^2}{n/4}>\chi^2_{1,1-\alpha}$, או כאשר $\frac{\left|S_n-n/2\right|}{\sqrt{n/4}}>z_{1-\alpha/2}$. שתי הצורות אקוויוולנטיות.

הערך הקריטי של מבחן הסימן – $z_{1-\alpha/2}$ מתקבל מהקשר בין התפלגות חי־בריבוע להתפלגות הנורמלית סטנדרטית, כאשר $\chi^2_1=Z^2$. (ראו הסבר בפרק 5.1.)

מסקנה 7.2. במקרה של שני טיפולים, המבחן של פרידמן אקוויוולנטי למבחן הסימן הדו־צדדי.

שימו לב שמבחן פרידמן למספר כלשהו של טיפולים הוא הכללה של מבחן הסימן למדגם מזווג, ואיננו הכללה של מבחן ווילקוקסון, למשל. זוהי גם הסיבה שערכנו השוואות מרובות בין כל שני זוגות של טיפולים (דוגמאות 7.3 ו-7.4) בעזרת מבחן הסימן.

7.6 מבחן קוקרן

נניח שהתצפיות שנמדדו הן משתנים דיכוטומיים, המקבלים את הערכים 0 או 1 בלבד. (נעיר שכל משתנה דיכוטומי אחר ניתן להציג גם כמשתנה מציין כזה ואין זה משפיע על התוצאות.)

זאת אומרת, התצפיות הן מהצורה $X_{ij}=0,1$ לכל $i=1,\ldots,n;\ j=1,\ldots,k$.

פונקציות ההסתברות של התצפיות תלויות בפרמטרים $p_j=P(X_{ij}=1)$.

את ההשערה של שוויון התפלגויות בין כל k הטיפולים ניתן לרשום על ידי

$$H_0: p_1=p_2=\cdots=p_k$$

במקרה כזה כל הנתונים בטבלה של n בלוקים ו-k טיפולים הם כולם 0 או 1 בלבד, כמובן.

נסמן: L_i – סכום הערכים בבלוק ה-i (מספר ה-1-ים בשורה i),

B_j – סכום הערכים בטיפול ה-j (מספר ה-1-ים בעמודה j).

הסטטיסטי של קוקרן (Cochran, 1950) מוגדר על ידי

$$Q=\frac{k(k-1)\sum_{i=1}^{k}(B_i-\overline{B})^2}{k\sum_{i=1}^{n}L_i-\sum_{i=1}^{n}L_i^2} \qquad (23)$$

תחת השערת האפס הסטטיסטי Q מתפלג בקירוב χ^2_{k-1}.

נביא דוגמה לשימוש בסטטיסטי של קוקרן.

דוגמה 7.5. אנו מביאים נתונים חלקיים ממחקר שנערך לשם בדיקת יעילות של חומר אִלחוש חדש לטיפול שיניים אצל ילדים (ד"ר אריקה עמיר). בין השאר נרשם כיצד הגיבו הילדים לזריקה: א) אם בכו; ב) אם מצמצו בעיניים; ג) אם הזיזו ידיים. בלוח 7.9 רשומות התגובות של 23 ילדים שקיבלו זריקת אִלחוש של החומר החדש (0 = לא הייתה תגובה, 1 = הייתה תגובה).

לוח 7.9. *תגובות ילדים לזריקת אַלחוש*

סך הכול (L_i)	ידיים	עיניים	בכי	הילד
3	1	1	1	1
3	1	1	1	2
0	0	0	0	3
0	0	0	0	4
1	0	0	1	5
0	0	0	0	6
0	0	0	0	7
0	0	0	0	8
1	0	0	1	9
0	0	0	0	10
1	0	1	0	11
1	0	1	0	12
0	0	0	0	13
0	0	0	0	14
0	0	0	0	15
1	0	0	1	16
1	0	0	1	17
0	0	0	0	18
0	0	0	0	19
1	0	0	1	20
0	0	0	0	21
1	0	0	1	22
1	0	0	1	23
15	2	4	9	סך הכול (B_j)

מספר הטיפולים הוא $k=3$, ומספר הנבדקים $n=23$. התוצאות המתקבלות:

$$\bar{B}=\frac{15}{3}=5 \qquad \sum_{i=1}^{16} L_i=15 \qquad \sum_{i=1}^{16} L_i^2=27$$

$$\sum_{j=1}^{4}(B_j-\bar{B})^2=4^2+1^2+3^2=26$$

את סך הסטיות הריבועיות ניתן, כמובן, לחשב גם על ידי

$$\sum_{j=1}^{k}(B_j-\bar{B})^2=\sum_{j=1}^{k}B_j^2-k\bar{B}^2$$

ועבור הדוגמה מקבלים

$$\sum_{j=1}^{3}(B_j-\overline{B})^2=\sum_{j=1}^{3}B_j^2-3\overline{B}^2=9^2+4^2+2^2-3\cdot5^2=26$$

הסטטיסטי של קוקרן

$$Q=\frac{k(k-1)\sum_j(B_j-\overline{B})^2}{k\sum_i L_i-\sum_i L_i^2}=\frac{3(2)(26)}{3(15)-27}=8.67$$

לפי טבלת התפלגות χ^2 עם 2 דרגות חופש (טבלה 5 בנספח), $.01<P<.025$, כאשר לפי ההתפלגות המדויקת של משתנה χ_3^2 מקבלים $P=.013$. התוצאה מובהקת ולכן נסיק שיש הבדל בין התגובות השונות.

הערה: היות שלכל i, $L_i\le k$, המכנה של Q אינו יכול להיות שלילי, כלומר, תמיד מתקיים $\sum_{i=1}^{n}L_i^2\le k\sum_{i=1}^{n}L_i$. מכאן נובע שתמיד מתקיים $Q\ge0$. נבדוק מתי המכנה מתאפס. $L_i^2=kL_i$ כאשר $L_i=0$, או כאשר $L_i=k$. לכן כדי שהמכנה יתאפס דרוש שבכל בלוק כל הערכים יהיו זהים (כל הערכים הם 0 או כל הערכים הם 1). הדבר קורה רק כאשר, למעשה, אין כל הבדל בין k הטיפולים, וכל אחד מהנבדקים מגיב באופן זהה בדיוק לכל אחד מהם.

במקרה הקיצוני הזה סכומי הדרגות B_j בכל הטיפולים הם שווים ואז מתקבל $\sum(B_j-\overline{B})^2=0$. מובן שערך זה, שהוא הנמוך ביותר האפשרי, אינו בשום אופן ערך שעבורו יש לדחות את השערת האפס, ולכן גם אין צורך בתקנון שלו ל- Q.

מתברר שיש קשר קשר בין הסטטיסטי של קוקרן לבין הסטטיסטי של פרידמן, והוא מוצג במשפט הבא.

משפט 7.4. הסטטיסטי של קוקרן שווה לסטטיסטי של פרידמן עם ערכי תיקו. כלומר, $Q=\tilde{F}_r$, כאשר $\tilde{F}_r$ מוגדר בנוסחה (17).

ההוכחה בנספח 7ב.

השוואות זוגיות

כמו לגבי מבחן פרידמן באופן כללי, גם לאחר קבלת תוצאה מובהקת במבחן קוקרן ניתן לערוך השוואות מרובות בין כל הזוגות של הטיפולים, על ידי שימוש במבחן הסימן לגבי כל זוג, עם תיקון בונפרוני. בכל השוואה יש, כמובן, לקחת בחשבון רק את הזוגות שבהם ערכים שונים לשני הטיפולים – (0,1) או (1,0).

לגבי דוגמה 7.5, כדי שרמת המובהקות הכללית לא תעלה על $\alpha=.10$, דרוש שבכל

אחת מ-3 ההשוואות הזוגיות רמת המובהקות תהיה $0.1/3 = .0333$, ולכן הסתברות הזנב החד־צדדי צריכה להיות קטנה מ- $0.0333/2 = .0167$. בכל השוואה של זוגות התגובות יש מספר אחר של ערכי הפרשים השונים מאפס. בדקו וודאו את הממצאים הבאים:

א) השוואת בכי למצמוץ בעיניים: $S_9 = 7$, $P/2 = .0898$ (זו תוצאה לא מובהקת, כמובן).

ב) השוואת בכי לתזוזה של הידיים: $S_7 = 7$, $P/2 = .0078$. זו תוצאה מובהקת, והמסקנה שילדים בוכים יותר מאשר מזיזים ידיים.

ג) השוואת מצמוץ בעיניים לתזוזה של הידיים: $S_2 = 2$. רק אצל שני ילדים התגובות בעיניים ובידיים היו שונות. מובן שאין הבדל בין שתי התגובות הללו.

7.7 מבחן מקנמר

מקרה פרטי של הבעיה שבה מטפל המבחן של קוקרן הוא המקרה של משתנה דיכוטומי, עם שני טיפולים בלבד, $k = 2$. מצד שני, זה גם מקרה פרטי של מדגם מזווג, כאשר התוצאות האפשריות הן דיכוטומיות (למשל, 0 או 1).

נסמן את זוגות המשתנים המתאימים (X_{i1}, X_{i2}) על ידי

$X_{i1} = X_i$ – ציון הביקורת, $X_{i2} = Y_i$ – ציון הטיפול, בבלוק i (נבדק i), $(i = 1, \ldots, n)$.

נגדיר את הפרמטרים $p_1 = P(X_i = 1),\ p_2 = P(Y_i = 1)$.

השערת האפס הנבדקת היא $H_0: p_1 = p_2$, כלומר, הסיכוי ל"הצלחה" שווה בשני הטיפולים.

את התוצאות שהתקבלו במדגם ניתן להציג בטבלת שכיחות דו־ממדית 2×2, כמו לוח **7.10**.

לוח 7.10. *טבלת שכיחויות של n בלוקים בגודל 2*

ביקורת	טיפול 0	טיפול 1	סך הכול
0	a	b	
1	c	d	B_1
סך הכול		B_2	n

הסימונים, כפי שרואים בטבלה:

a – מספר הבלוקים מהצורה $(0,0)$ – בטיפול ובביקורת תוצאה נמוכה;

b – מספר הבלוקים מהצורה $(0,1)$ – בטיפול תוצאה גבוהה מאשר בביקורת;

c – מספר הבלוקים מהצורה $(1,0)$ – בביקורת תוצאה גבוהה מאשר בטיפול;

d – מספר הבלוקים מהצורה $(1,1)$ – בטיפול ובביקורת תוצאה גבוהה.

B_1 – מספר ה-1-ים בביקורת (מספר ה"הצלחות" בקבוצת הביקורת), $B_1 = c+d$

B_2 – מספר ה-1-ים בטיפול (מספר ה"הצלחות" בקבוצת הטיפול), $B_2 = b+d$.

המבחן של מקנמר (לבחינת ההבדל בין שני הטיפולים מבוסס על המשתנה $B_2 - B_1 = b-c$.

את הפרמטרים p_1 ו-p_2 לעיל ניתן לרשום:

$$p_1 = P(X_i = 1) = P(X_i = 1, Y_i = 1) + P(X_i = 1, Y_i = 0)$$

$$p_2 = P(Y_i = 1) = P(X_i = 1, Y_i = 1) + P(X_i = 0, Y_i = 1)$$

לכן ההפרש בין שתי ההסתברויות הוא למעשה ההפרש בין ההסתברויות באלכסון הטבלה:

$$p_1 - p_2 = P(X_i = 1, Y_i = 0) - P(X_i = 0, Y_i = 1)$$
$$= P(1,0) - P(0,1)$$

את השערת האפס ניתן לרשום, לפיכך, על ידי $H_0 : P(1,0) = P(0,1)$, כלומר, שוויון ההסתברויות של שני התאים באלכסון הטבלה בלוח **7.10**.

סטטיסטי המבחן של מקנמר (McNemar, 1947) לבדיקת השערת האפס הוא

$$Z = \frac{b-c}{\sqrt{b+c}} \qquad (24)$$

כשהשערת האפס נכונה, הסטטיסטי Z מתפלג בקירוב נורמלית סטנדרטית, בתנאי ש-$b+c$ די גדול.

נסתכל על בעיה דו־צדדית, כלומר, על האלטרנטיבה $H_1 : p_1 \neq p_2$, או, בניסוח אחר, $H_1 : P(1,0) = P(0,1)$. במקרה זה יש לדחות את השערת האפס עבור ערכים גבוהים של הערך המחלט $|Z|$, או, באופן אקוויוולנטי, עבור ערכים גבוהים של Z^2.

תחת השערת האפס $Z^2 = \dfrac{(b-c)^2}{b+c}$ מתפלג בקירוב חי־בריבוע עם דרגת חופש אחת.

מכאן, עבור רמת מובהקות α יש לדחות את השערת האפס כאשר $|Z| \geq z_{1-\alpha/2}$, או אלטרנטיבית, כאשר $Z^2 > \chi^2_{1,1-\alpha}$.

טענה 7.5. המבחן של מקנמר אקוויוולנטי למבחן הסימן עבור בעיה דו־צדדית.

הוכחה: ניתן להסתכל על n הבלוקים (הנבדקים) כ-n זוגות, שעבורם אפשר להפעיל את

מבחן הסימן. את כל הבלוקים שעבורם שתי התוצאות שוות מוציאים מן הניתוח. אלה הן $a+d$ התצפיות באלכסון הראשי בלוח 7.10. אנו נותרים, אפוא, עם $b+c$ תצפיות שעבורן התוצאות שונות: $X_i \neq Y_i$. הסטטיסטי של מבחן הסימן הוא מספר הבלוקים שעבורם $X_i < Y_i$, כלומר מספר הבלוקים מהצורה (0,1). מקבלים, אפוא, $S_N = b$, כאשר $N = b+c$.

המשתנה המתוקנן של מבחן הסימן הוא

$$Z_1 = \frac{S_N - N/2}{\sqrt{N/4}} = \frac{b-(b+c)/2}{\sqrt{(b+c)/4}} = \frac{(b-c)/2}{\sqrt{(b+c)/4}} = \frac{b-c}{\sqrt{b+c}}$$

קיבלנו $Z_1 = Z$, כאשר Z הוא הסטטיסטי של מקנמר, נוסחה (24). עבור אלטרנטיבה דו־צדדית, מבחן הסימן דוחה את השערת האפס כאשר Z_1 גדול בערכו המוחלט. קיבלנו, אפוא, ששני המבחנים זהים. ♣

דוגמה 7.6. כדי להתקבל לעבודה מסוימת יש לעבור שני ראיונות נפרדים (אצל שני מראיינים שונים). בלוח 7.11 מוצגות תוצאות הראיונות של 68 מועמדים. נרצה לבדוק אם הסיכוי להצלחה בראיון הראשון שונה מאשר בראיון השני (ייתכן שהמועמדים לומדים מן הניסיון, או אולי הם דווקא יותר לחוצים).

לוח 7.11. *לוח שכיחות של 68 מועמדים לפי התוצאות בשני ראיונות*

ריאיון ראשון	ריאיון שני: כישלון	ריאיון שני: הצלחה	סך הכול
כישלון	26	14	40
הצלחה	10	18	28
סך הכול	36	32	68

מנתוני הלוח מקבלים $b=14,\ c=10$ ומכאן הסטטיסטי (24) של מקנמר הוא

$$Z = \frac{b-c}{\sqrt{b+c}} = \frac{14-10}{\sqrt{24}} \cong 0.816$$

זוהי, כמובן, תוצאה לא מובהקת לפי ההתפלגות הנורמלית סטנדרטית. מובהקות התוצאה (מבחן דו־צדדי) היא $P = 2[1-\Phi(.816)] = .4142$

חישוב המובהקות על פי התפלגות חי־בריבוע נותן, כמובן, אותה תוצאה. הסטטיסטי הוא

$$Z^2 = \frac{(14-10)^2}{24} = \frac{16}{24} = \frac{2}{3} = 0.667$$

השוואה לטבלת חי־בריבוע עם דרגת חופש אחת נותנת את המובהקות

$$P = P(Z^2 \geq 0.667) = P(\chi_1^2 \geq 0.667) = .4142$$

את ההסתברות לעיל חישבנו באופן מדויק ולא על פי טבלה 5 בנספח, שבה ניתן למצוא רק חסמים עבור המובהקות. הסתכלות בטבלה נותנת את המובהקות בתחום $0.25 < P < 0.50$.

המסקנה מן הניתוח היא שאין הבדל במידת ההצלחה בין שני הראיונות.

תיקון רציפות עבור הסטטיסטי של מקנמר

עבור מדגמים לא כל כך גדולים, רצוי לערוך תיקון רציפות כדי להשתמש בקירוב הנורמלי במבחן מקנמר. נראה כאן כיצד נערך התיקון, כאשר אנו מסתכלים על הסטטיסטי האקוויוולנטי של מבחן הסימן.

באופן עקרוני אין צורך להשתמש בסטטיסטי של מקנמר, כיוון שהוא למעשה זהה למבחן הסימן, שאותו אנו כבר מכירים. עם זאת, לעתים נוח יותר להסתכל על טבלה מהצורה של לוח 7.10 וממנה על הסטטיסטי (24). לכן נראה כיצד משתנה הנוסחה (24) אם מעוניינים לעשות תיקון רציפות.

נחזור למבחן הסימן. מצאנו שניתן לרשום $N = b + c$, $S_N = b$, כאשר N הוא מספר התצפיות שבהן שתי התוצאות שונות. אנו דנים במקרה שבודקים אלטרנטיבה דו־צדדית.

א) אם $b > c$ מתקבל ערך חיובי של Z ומובהקות התוצאה מתקבלת על ידי הזנב הימני של התפלגות S_N, כלומר:

$$P/2 = P(S_N \geq b) \cong 1 - \Phi\left(\frac{b - 1/2 - N/2}{\sqrt{N/4}}\right)$$

זאת אומרת שהערך המתוקנן של Z הוא

$$Z = \frac{b - 1/2 - N/2}{\sqrt{N/4}} = \frac{b - 1/2 - (b+c)/2}{\sqrt{(b+c)/4}} = \frac{b - c - 1}{\sqrt{b+c}} \geq 0$$

והערך המוחלט זהה לערך המקורי: $|Z| = \dfrac{b - c - 1}{\sqrt{b+c}}$

ב) אם $b < c$ מתקבל ערך שלילי של Z ומובהקות התוצאה מתקבלת על ידי הזנב השמאלי של התפלגות S_N, כלומר:

$$P/2 = P(S_N \leq b) \cong \Phi\left(\frac{b + 1/2 - N/2}{\sqrt{N/4}}\right)$$

זאת אומרת שהערך המתוקנן של Z הוא

$$Z = \frac{b + 1/2 - N/2}{\sqrt{N/4}} = \frac{b - c + 1}{\sqrt{b+c}} \leq 0$$

והערך המוחלט שלו מקיים:
$$|Z|=-Z=\frac{c-b-1}{\sqrt{b+c}}$$
אפשר לרשום את הערך המתוקנן של Z עבור שתי האפשרויות לעיל בנוסחה אחת:
$$|Z|=\frac{|b-c|-1}{\sqrt{b+c}} \qquad (25)$$
או את הריבוע שלו:
$$Z^2=\frac{(|b-c|-1)^2}{b+c} \qquad (26)$$
נוסחה (25) היא הצורה המתאימה לתיקון רציפות של סטטיסטי המבחן של מקנמר, המושווה להתפלגות הנורמלית, ונוסחה (26) היא הצורה המתאימה לתיקון רציפות של סטטיסטי המבחן של מקנמר, המושווה להתפלגות חי־בריבוע.

דוגמה 7.7 (המשך דוגמה 7.6). נחשב ראשית את מובהקות התוצאה עם תיקון רציפות לפי מבחן הסימן. בניסוי התקבלו התוצאות $N=14+10=24$, ו-$S_{24}=14$.

א) המובהקות המדויקת על סמך טבלת מבחן הסימן, טבלה 3 בנספח, המתקבלת:
$$P=2P(S_{24}\geq 14)=2P(S_{24}\leq 10)=2(.2706)=.5412$$
ב) על סמך קירוב נורמלי, עם תיקון רציפות מקבלים:
$$P=2P(S_{24}\geq 14)\cong 2\left[1-\Phi\left(\frac{13.5-24/2}{\sqrt{24/4}}\right)\right]$$
$$=2[1-\Phi(0.61)]=2(.2709)=.5418$$
ההבדל בין המובהקות המדויקת למובהקות המקורבת הוא זניח.

ג) הערך המתוקנן של התוצאה לפי הסטטיסטי של מקנמר עם תיקון רציפות, נוסחה (25), הוא
$$|Z|=\frac{|b-c|-1}{\sqrt{b+c}}=\frac{14-10-1}{\sqrt{24}}=0.61$$
וזה, כמובן, זהה לערך המתוקנן שהתקבל באופן ישיר לפי מבחן הסימן.

ד) אם אין עורכים תיקון רציפות, מקבלים תוצאה שונה. לפי נוסחה (24), הסטטיסטי של מקנמר כפי שחישבנו בדוגמה 7.6, שהתקבל הוא $|Z|=0.816$. ערך זה גבוה יותר מזה שהתקבל עם תיקון רציפות (0.61) והוא נותן מובהקות קטנה בהרבה מהמובהקות המדויקת או המקורבת, כפי שהתקבלו ב-א או ב-ב לעיל.

הקשר בין הסטטיסטי של מקנמר לסטטיסטי של פרידמן

נסתכל על n הבלוקים של הזוגות בלוח $n\times 2$, שבו כל תצפית היא 0 או 1. זהו מקרה פרטי של המודל של קוקרן, עם $k=2$. נראה כיצד ניתן לרשום את הסטטיסטי של

קוקרן במקרה הספציפי הזה. כדי להשתמש בסטטיסטי של קוקרן, נשתמש בסימון המתאים: L_i – מספר ה-1-ים אצל נבדק i ($L_i = 0,1,2$).

מטבלת השכיחויות, לוח 7.10, מוצאים את הערכים הדרושים עבור הסטטיסטי של קוקרן:

$$\sum_{i=1}^{n} L_i = (b+c)+2d \qquad \sum_{i=1}^{n} L_i^2 = (b+c)+4d$$

ערכי B_j, $j=1,2$ (מספר ה-1-ים בביקורת ומספר ה-1-ים בטיפול) ניתנים גם הם מלוח השכיחויות 7.10: $B_1 = c+d$, $B_2 = b+d$ והממוצע שלהם הוא

$$\overline{B} = \frac{1}{2}(B_1+B_2) = \frac{1}{2}(b+c+2d) = \frac{1}{2}(b+c)+d$$

סכום ריבועי הסטיות המתקבל (בִּדקו!):

$$(B_1-\overline{B})^2 + (B_2-\overline{B})^2 = 2\left[\frac{1}{2}(b-c)\right]^2 = \frac{(b-c)^2}{2}$$

מכאן ניתן לרשום את הסטטיסטי (23) של קוקרן במונחים של b ו-c:

$$Q = \frac{k(k-1)\sum_{i=1}^{k}(B_i-\overline{B})^2}{k\sum_{i=1}^{n}L_i - \sum_{i=1}^{n}L_i^2} = \frac{2(1)\cdot\frac{(b-c)^2}{2}}{2(b+c+2d)-(b+c+4d)}$$

$$= \frac{(b-c)^2}{b+c} = Z^2$$

כאשר Z הוא הסטטיסטי (24) של מקנמר. קיבלנו, אפוא, שהסטטיסטי של מקנמר שווה לסטטיסטי של קוקרן. השוויון מתקבל לגבי הסטטיסטי של מקנמר **ללא תיקון רציפות**.

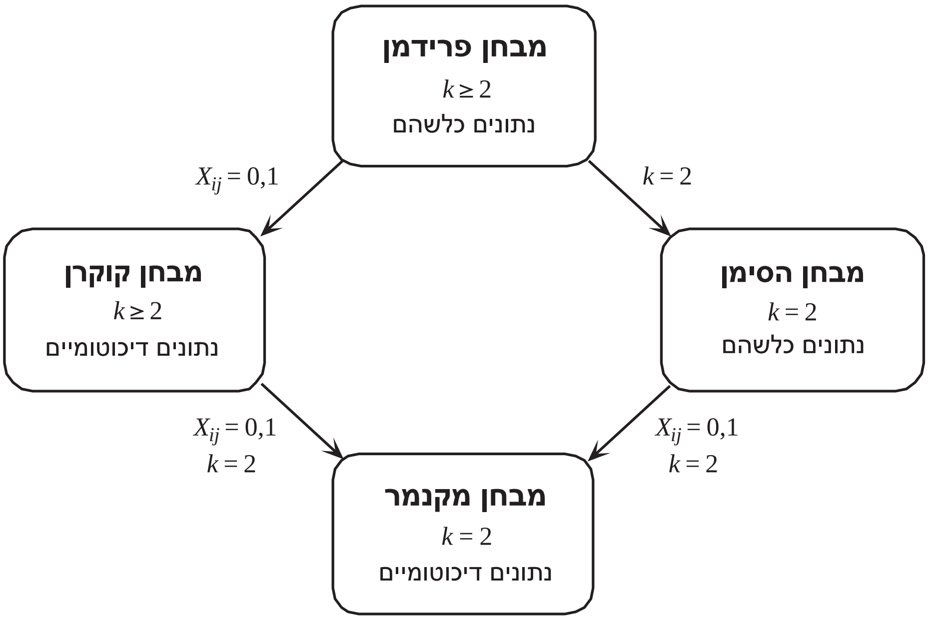
מסקנה 7.3. המבחן של מקנמר הוא מקרה פרטי של מבחן קוקרן לבעיית בלוקים אקראיים, כאשר מספר הטיפולים הוא $k=2$.

לפי משפט 7.4 הסטטיסטי של קוקרן אקוויוולנטי לסטטיסטי של פרידמן, עם תיקון לערכי תיקו. כמו כן, הראינו (טענה 7.4) שכאשר $k=2$, המבחן של פרידמן אקוויוולנטי למבחן הסימן (דו־צדדי). מכאן נובע שהמבחן של מקנמר הוא למעשה מקרה פרטי של מבחן פרידמן, כאשר הנתונים הם דיכוטומיים ומספר הטיפולים הוא $k=2$ בלבד.

הערה: המבחן של קוקרן מטפל באופן כללי בהשוואת k טיפולים, ולגביו אין אפשרות לעשות תיקון רציפות, מכיוון שקשה לקבוע מהו המרחק בין שני ערכים סמוכים של

המשתנה. המרחק הזה תלוי בפערים בין הסכומים $B_1,\ldots,B_k$. לעומת זאת, המבחן המקביל עבור המקרה של שני טיפולים בלבד מאפשר לעשות תיקון רציפות ולשפר את הקירוב בחישוב מובהקות התוצאה.

נציג עתה באופן כללי את הקשרים שמצאנו בין המבחנים השונים לבעיית n בלוקים אקראיים עם k טיפולים. הקשרים מתוארים בציור 7.2. לכל בעיה של k בלוקים אקראיים ניתן להשתמש בסטטיסטי של פרידמן, עם תיקונים מתאימים לערכי תיקו. כאשר $k=2$ מבחן פרידמן הוא למעשה מבחן הסימן (דו־צדדי); כאשר הנתונים הם דיכוטומיים, מבחן פרידמן הוא למעשה מבחן קוקרן; וכאשר $k=2$ ויחד עם זאת הנתונים דיכוטומיים, מבחן פרידמן הוא למעשה מבחן מקנמר (אך בלי תיקון רציפות).



ציור 7.2. *הקשרים בין ארבעת המבחנים לבעיית בלוקים אקראיים*

מסקנה: מבחן הסימן, מבחן קוקרן ומבחן מקנמר הם כולם מקרים פרטיים של מבחן פרידמן, עם תיקון מתאים לערכי תיקו.

תרגילים

1. שמונה עשר נבדקים התבקשו להגיב לטונים חלשים מאוד על ידי הקשה על מקש מיוחד עבור כל טון. הניסוי נערך תחת חמישה תנאי רעש שונים. בכל אחד מהתנאים הושמעו 100 טונים ונרשם מספר הזיהויים הנכונים. התוצאות נתונות בטבלה להלן.

א) האם קיים הבדל בין התנאים $(\alpha = .05)$? אם כן – נסו לקבוע, בעזרת השוואות זוגיות, מהו התנאי העדיף, שבו קל יותר לזהות את הטון.

ב) חשבו את מקדם ההסכמה בין הנבדקים. הסבירו מה ניתן לומר על סמך הערך שהתקבל. מה הקשר לתוצאה שהתקבלה בחלק א לגבי בדיקת ההבדל בין התנאים?

	התנאי				
הנבדק	1	2	3	4	5
1	73	75	71	74	64
2	91	89	83	80	82
3	92	94	90	93	83
4	84	82	76	73	75
5	56	58	54	57	47
6	60	58	52	49	51
7	73	75	69	72	62
8	70	68	64	61	63
9	87	89	83	86	76
10	75	73	69	66	68
11	77	75	69	72	62
12	68	70	64	61	63
13	73	75	69	72	62
14	75	73	68	65	67
15	93	95	89	92	82
16	90	89	85	82	84
17	84	86	80	83	73
18	69	67	63	60	62

2. בניסוי לגבי האפקט של אסוציאציות ריגושיות על הזיכרון, כל אחד מ-23 נבדקים נתבקש להיזכר בשמותיהם של 18 סיפורים שנכתבו קודם לכן. מבין הסיפורים

הללו לשישה ניתנה הערכה חיובית (+), לשישה ניתנה הערכה שלילית (–) ושישה לא הוערכו כלל (0). הנבדקים האמינו שלהערכות ישנה משמעות, אך למעשה הן הוקצו באופן אקראי. הנתונים בטבלה הם מספר הסיפורים ששמותיהם נשכחו על ידי כל נבדק, בכל אחת מקטגוריות ההערכה:

הנבדק	–	0	+	הנבדק	–	0	+
1	2	3	1	13	5	3	2
2	2	1	2	14	2	0	2
3	0	1	2	15	3	2	3
4	4	3	2	16	0	1	0
5	5	3	4	17	2	2	1
6	1	1	1	18	4	2	2
7	4	3	3	19	4	3	3
8	0	0	1	20	1	1	0
9	1	0	0	21	3	3	1
10	0	0	1	22	4	4	3
11	0	1	2	23	2	0	1
12	1	1	0				

חשבו את מובהקות התוצאה והסיקו אם הזכירה תלויה בהערכה ($\alpha = .05$). אם יש צורך – עִרכו השוואות זוגיות. הסבירו את התוצאה.

3. דגימות קרקע נלקחו מחמישה מקומות באזור מסוים ונבדקו לגבי תכולה של חומר מסוים (1 – החומר נמצא, 0 – החומר לא נמצא). הדגימות נלקחו מאותם מקומות 12 פעם בזמנים שונים. להלן התוצאות:

	הזמן											
מקום	1	2	3	4	5	6	7	8	9	10	11	12
א	0	0	0	1	1	0	0	1	1	1	0	0
ב	0	1	0	1	1	0	0	1	0	0	0	0
ג	1	1	1	1	1	1	1	1	1	1	0	0
ד	1	1	1	1	0	1	1	0	1	0	0	0
ה	0	1	0	0	1	0	0	0	0	0	0	0

בְּדקו אם ההסתברות למציאת החומר שווה בכל המקומות ($\alpha = .05$). אם יש צורך – עִרכו השוואות מרובות. הסבירו את התוצאה.

4. לגבי הנתונים בתרגיל 3, ללא קשר לתוצאה שהתקבלה שם, עִרכו טבלת שכיחויות מתאימה להשוואת מקום ג למקום ה. רִשמו את הסטטיסטי של מקנמר ובדקו

בעזרתו את ההשערה שבשני מקומות אלה ישנו אותו סיכוי למציאת החומר, כנגד האלטרנטיבה שבמקום ג הסיכוי גדול יותר (השערה חד־צדדית). עִרכו גם מבחן סימן מתאים (עם תיקון רציפות). השוו את מובהקות התוצאה בשני המבחנים.

5. להלן חלק מטבלת דירוג המבקרים בעיתון "עכבר העיר" לגבי שבעה סרטי קולנוע בשבוע מסוים (שבעה ה"טובים" ביותר). הטבלה רשומה כפי שהופיעה בעיתון.

הסרט	שמוליק דובדבני	יהודה סתיו	ניסים דיין	אורי קליין	יהודית אוריין	גידי אורשר
א	*****	*****	****	***	*****	****
ב	*****	****	*****	***	***	****
ג	****	*****	****	****	****	***
ד	****	****	***	***	****	****
ה	***	****	***	****	***	****
ו	****	****	***	**	***	****
ז	****	***	***	***	****	***

מקרא: ♦ איום   * גרוע   ** סתמי   *** שווה לראות   **** טוב מאוד   ***** מצוין

א) חשבו את מקדם ההסכמה בין המבקרים לגבי הערכת הסרטים הללו.

ב) בִּדקו אם יש הבדל באיכותם של הסרטים, או שאפשר לומר ש"כולם טובים". מהי מובהקות התוצאה? מה המסקנה עבור $\alpha=.10$.

מקורות

רביב, א׳ ולויתן, ת׳ (2000). מבוא להסתברות וסטטיסטיקה: הסקה סטטיסטית. תל אביב: הוצאת עמיחי.

Ansari, A. R. and Bradely, R. A. (1960). Rank sum tests for dispersion. *Ann. Math. Stat*., *31*, 1174-1189.

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, *37*, 256-266.

Conover, W. J. (1980). *Practical Nonparametric Statistics*. (2nd ed). New York: John Wiley & Sons.

Devore, J. L. (1991). *Probability and statistics for engineering and the sciences* (3rd ed). Pacific Grove, California: Brooks/Cole.

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. Statist. Ass.*, *32*, 675-701.

Hollander, M. and Wolfe, D. A. (1973). *Nonparametric Statistical Methods*. New York: John Wiley & Sons.

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika*, *41*, 133-145.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*, 81-93.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion analysis of variance. *J. Amer. Statist. Ass.*, *47*, 583-621.

Lehmann, E. L. (1975). *Nonparametrics, Statistical Methods Based on Ranks*. San Francisco: Holden-Day.

Mann, H. B. and Whitney, D. R. (1947). On a test whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.*, *18*, 50-60.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*, 153-157

Siegel, S. and Tukey, J. W. (1960). A nonparametric sum of ranks procedure for relative spread in unpaired samples. *J. Amer. Statist. Ass.*, *55*, 429-445.

Smirov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples (in Russian). *Bull. Moscow University*, *2*, 3-16.

Spearman, C. (1904). The proof of measurement of association between two things. *Am. J. Psychol.*, *15*, 72-101.

Walpole, R. E. and Myers, R. H. (1993). *Probability and statistics for engineers and scientists* (5th ed). New York: Macmillan.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics*, *1*, 80-85.

תשובות לתרגילים נבחרים

פרק 1

1. $T = 6$; $P = \frac{5}{35} = .143$

2. א) $S = 7.62$. ; $P = \frac{2}{15} = .1333$

פרק 2

2. $W_s \sim U(1,11)$

4. $W_s = 42$; $P = P(W_s \geq 42) = .0152$

5. ג) $P\left\{W_s(4,6) = 17\right\} = \frac{4}{10}P\left\{W_s(3,6) = 7\right\} + \frac{6}{10}P\left\{W_s(4,5) = 17\right\} = .0477$

ד) $P\left\{W_s(4,6) = 12\right\} = \frac{4}{10}P\left\{W_s(3,6) = 2\right\} + \frac{6}{10}P\left\{W_s(4,5) = 12\right\} = .0095$

7. א) $W_s = 61$; $P = P(W_s \geq 61) = .0213$. התוצאה מובהקת.

ב) $W_{xy} = 40$

ג) $P(W_s \geq 61) \cong .0227$

8. א) (1) דוחים עבור $W_s \geq 84$ (רמת המובהקות היא .0524);

(2) דוחים עבור $W_s \geq 99$ (רמת המובהקות היא .0464);

(3) דוחים עבור $W_s \geq 90$ (רמת המובהקות היא .0464).

ב) $W_s = 107$. $P = .0313$. תוצאה מובהקת.

ג) $P = 2(.0313) = .0626$. תוצאה לא מובהקת.

9. $\tilde{W}_s = 25.5$. $P = \frac{16}{\binom{9}{4}} = 0.127$

12. $\tilde{W}_s = 226.5$; $P = P_{H_0}(\tilde{W}_s \geq 226.5) \cong .1446$

13. א) $c = 135$

ג) $\tilde{W}_s = 9(16) + 10.5 = 154.5$; $P = P_{H_0}(\tilde{W}_s \geq 154.5) = \frac{2}{\binom{20}{10}}$

14. $n \geq 75$

15. $n \geq 71$

17. א) $\pi(1) = .3015$; ב) $\pi(2) = .8437$

18. א) $\pi(1) = .7459$; ב) $\pi(2) = .9996$

22. ב) $M = 8(7) = 56$;

$$\hat{\Delta} = \frac{1}{2}\left[D_{(28)} + D_{(29)}\right] = \frac{1}{2}(-0.9 - 1.0) = -0.95$$

ג) הרווח הוא $[-2.3,0.3] = \left[D_{(11)}, D_{(46)}\right]$.

23. א) $r = 228$, $s = 398$.

ג) $r = 201$, $s = 400$

פרק 3

2. א) סכום המשקלים במדגם הקטן $S = 31$; $P = P(S \geq 31) = .2303$

ג) $A = 25$; $P = P(A \leq 25) \cong .1949$

4. א) $S = 75$; $P = P(S \leq 75) = P\{W_s \leq 75\} = .0416$

ג) $D = 0.4$; $P = P\left\{D_{8,10} \geq .4\right\} \cong L(0.84) = .4806$

5. א) $D^+_{2,3} = 2/3$

ג) $P = P_{H_0}(D^+_{2,3} \geq 2/3) = 3/10$

ד) $W_s = 11.5$; $P = \frac{2}{10}$.

7. א) $D_{14,14} = 2/7 = .286$; $P = P\left\{D_{14,14} \geq \frac{2}{7}\right\} \cong L(0.756) = .6104$.

ב) $D^+_{14,14} = 2/7$; $P = P\left\{D^+_{14,14} \geq \frac{2}{7}\right\} \cong e^{-2(0.756)^2} = .319$.

ג) $\tilde{W}_s = 226.5$. $P = P(\tilde{W}_s \geq 226.5) \cong .1446$.

9. בעיה דו־צדדית: $D_{100,200} = .09$; $P = P\left\{D_{100,200} \geq .09\right\} \cong L(0.735) \cong .66$.

פרק 4

1. א) $P = 2P(S_{20} \geq 15) = 2(.0207) = .0414$

ב) $P(S_{20} \geq 15) \cong .0222$; $P \cong 2(.0222) = .0444$

2. $P = P(S_9 \geq 8) = .0195$

3. א) הבעיה כאן: $H_0\colon x_{.5} = 2\%$ $H_1\colon x_{.5} > 2\%$

S_n – מספר הצנצנות שעבורן יותר מ-2% חומר סינתטי.

$S_{15} = 9$; $P = P(S_{15} \geq 9) = .3036$.

ב) הבעיה: $H_0\colon x_{.9} \leq 3\%$ $H_1\colon x_{.9} > 3\%$

S_n – מספר הצנצנות שעבורן יותר מ-3% חומר סינתטי.

לפי התפלגות $B(15,0.1)$, $P = P(S_{15} \geq 3) = .1841$.

4. א) אזור הדחייה: $S_{12} \geq 10$

ב) העוצמה: $\pi = P_{p=.6}(S_{12} \geq 10) = .3032$

ג) $n \geq 226$

8. $V^+ = 39.5$; $P = \frac{13}{2^9} = .0254$

9. א) $S_7 = 5$; $P = P(S_7 \geq 5) = .2266$

ב) $V^+ = 18.5$; $P = \frac{38}{2^7} = .2969$

.10 א) $S_{22} = 15$; $P = P(S_{22} \geq 15) \cong .0681$

ב) $V^+ = 175.5$; $P \cong .0455$

.11 א) $r = 4, s = 12$. הרווח המתקבל: $[1.6,\ 2.5]$.

.13 א) לפי נוסחה (54) בפרק 4, $n \geq 101$;

לפי נוסחה (58) בפרק 4, במודל הזזה, $n \geq 100$.

ב) $n \geq 66$.

ג) $n \geq 63$

.14 א) לפי נוסחה (54) בפרק 4, $n \geq 388$;

לפי נוסחה (58) במודל הזזה $n \geq 345$.

ב) $n \geq 132$.

ג) $n \geq 132$.

.16 א) $\tilde{\Delta} = \frac{1}{2}[d_{(6)} + d_{(7)}] = 11$.

ב) $r = 3$, $s = 10$. הרווח $[d_{(3)}, d_{(10)}] = [8.4, 14.0]$.

ג) $r = 18$, $s = 61$. הרווח $[d_{(18)}, d_{(61)}] = [8.7, 13.0]$

פרק 5

.3 א) $H = 8.707$; $.025 < P < .05$.

.4 $H = 11.32$; $\tilde{H} = 11.39$; $P < .005$. השוואות מרובות $(\alpha = .05)$:

קבוע הגרביטציה של פלטינה נמוך מזה של זהב ושל זכוכית.

.7 $H = 12.25$; $\tilde{H} = 14.24$; $P < .005$. השוואות מרובות $(\alpha = .05)$:

בסרט ג מאשימים פחות מאשר בשני הסרטים האחרים.

פרק 6

.1 ספירמן: $r_s = .9286$; מובהקות מדויקת – $P = \frac{17}{7!} = .0034$;

מובהקות מקורבת (עם תיקון רציפות) – $P \cong 1 - \Phi(2.23) = .0129$.

קנדל: $T = .8095$;

מובהקות מקורבת (עם תיקון רציפות) – $P = 1 - \Phi(2.40) = .0082$.

.4 $\tilde{r}_s = .532$; $P \cong 1 - \Phi(1.765) = .0388$.

.5 $r_s = -.6868$; $P = P(r_s \leq -.6874) \cong \Phi(-2.92) = .0018$.

פרק 7

.1 א) $F_r = 52.0$. לפי התפלגות חי־בריבוע עם 4 דרגות חופש $P < .005$.

ב) מקדם ההסכמה: $W = 0.722$.

.2 $F_r = 4.76$; $\hat{F}_r = 6.00$. לפי התפלגות חי־בריבוע עם 2 דרגות חופש $P \cong .05$.
השוואות מרובות ($\alpha = .05$): אף לא אחת מהשוואות אינה מובהקת.

.3 $Q = 13.778$; לפי התפלגות חי־בריבוע עם 4 דרגות חופש $.005 < P < .01$.

.5 **א)** לפי נוסחה (18) מקדם ההסכמה הוא $\hat{W} = 0.3004$.
ב) $\hat{F}_r = 10.814$. טבלת התפלגות חי־בריבוע עם 8 ד"ח, $.05 < P < .10$.
יש הבדל בין הסרטים.

נספחים

השלמות והוכחות

נספח 2א. חישוב השונות המשותפת של שתי תצפיות בדגימה בלי החזרות

משפט. תהי נתונה אוכלוסייה בת N איברים בעלת ממוצע μ ושונות σ^2. בוחרים מתוך האוכלוסייה n איברים **בלי החזרה** ($n \le N$) ומקבלים מדגם $U_1, U_2, \ldots, U_n$. אזי קיים:

$$Cov(U_i, U_j) = -\frac{\sigma^2}{N-1} \qquad \text{לכל } i \ne j$$

הוכחה: נניח שאיברי האוכלוסייה הם $\{a_1, a_2, \ldots, a_N\}$ (N הערכים אינם חייבים להיות כולם שונים זה מזה). בבחירה מקרית של איבר מהאוכלוסייה, ההסתברות של כל אחד מאיברי האוכלוסייה היא שווה. לכן פונקציית ההסתברות של U_1 היא אחידה על האוכלוסייה. כלומר

$$P(U_1 = a_k) = \frac{1}{N} \qquad k = 1, \ldots, N$$

מטעמי סימטרייה, זו גם פונקציית ההסתברות של כל אחת משאר התצפיות. זאת אומרת, לכל $i = 1, \ldots, n$, מתקיים

$$P(U_i = a_k) = \frac{1}{N} \qquad k = 1, \ldots, N$$

היות שלכל n התצפיות אותה התפלגות, הן גם בעלות אותה תוחלת ואותה שונות. תוחלת כל אחת מהתצפיות שווה לממוצע האוכלוסייה:

$$EU_i = \frac{1}{N}\sum_{k=1}^{N} a_k = \bar{a} = \mu \tag{1}$$

כמו כן השונות של כל תצפית שווה לשונות האוכלוסייה:

$$Var(U_i) = \frac{1}{N}\sum_{k=1}^{N}(a_k - \mu)^2 = \sigma^2 \tag{2}$$

שוב, מטעמי סימטרייה, ההתפלגות המשותפת של כל זוג תצפיות זהה. הדגימה היא ללא החזרה, ולכן פונקציית ההסתברות המשותפת של שתי תצפיות (למשל, שתי הראשונות) היא

$$P(U_1 = a_k, U_2 = a_l) = \frac{1}{N(N-1)} \qquad k \ne l, \;\; k, l = 1, \ldots, N$$

חישוב השונות המשותפת:

$$Cov(U_1, U_2) = E(U_1 - \mu)(U_2 - \mu)$$

$$= \frac{1}{N(N-1)} \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l \neq k}}^{n} (a_k - \mu)(a_l - \mu)$$

$$= \frac{1}{N(N-1)} \left[\sum_{k=1}^{n} \sum_{l=1}^{n} (a_k - \mu)(a_l - \mu) - \sum_{k=1}^{n} (a_k - \mu)^2 \right]$$

$$= \frac{1}{N(N-1)} \left[\sum_{k=1}^{n} (a_k - \mu) \sum_{l=1}^{n} (a_l - \mu) - \sum_{k=1}^{n} (a_k - \mu)^2 \right]$$

התוחלת μ היא למעשה ממוצע איברי האוכלוסייה $\bar{a}$, לפי נוסחה (1). סך הסטיות של איברי קבוצה מממוצע הקבוצה הוא אפס ולכן המחובר הראשון בביטוי לעיל מתאפס. המחובר השני הוא שונות האוכלוסייה, נוסחה (2), מחולקת ב-$(N-1)$. מכאן מקבלים בהמשך לשוויון האחרון

$$= \frac{1}{N(N-1)} \left[-\sum_{k=1}^{n} (a_k - \mu)^2 \right] = -\frac{\sigma^2}{N-1}$$

♣

נספח 2ב. הקשר בין הסטטיסטי של ווילקוקסון והסטטיסטי של מאן-וויטני בבעיית שני מדגמים, כאשר ישנם מקרי תיקו

משפט 2.6. הקשר בין $\tilde{W}_{xy}$ לבין $\tilde{W}_s$ זהה לקשר בין W_{xy} לבין W_s. כלומר,

$$\tilde{W}_{xy} = \tilde{W}_s - \frac{n(n+1)}{2}$$

כאשר $\tilde{W}_s$ הוא סכום הדרגות הממוצעות במדגם בגודל n.

הוכחה: נניח, ראשית, שישנה קבוצת תיקו אחת בלבד, בגודל t, כאשר ערכי התיקו הם t התוצאות הנמוכות ביותר (הדרגות המגיעות להם הן הדרגות $1,2,\ldots,t$), ושכל שאר התצפיות שונות זו מזו.

נסמן ב-k את מספר ה-x-ים בקבוצת התיקו וב-l את מספר ה-y-ים בקבוצת התיקו ($k+l=t$). כלומר, $x_1 = x_2 = \cdots = x_k = y_1 = y_2 = \cdots = y_l$.

סכום הדרגות (הממוצעות) של כל ה-y-ים הוא

$$\tilde{W}_s = \sum_{j=1}^{n} \tilde{S}_j = l\left(\frac{t+1}{2}\right) + \sum_{j=l+1}^{n} S_j \qquad (3)$$

הסטטיסטי של מאן-וויטני הוא, לפי ההגדרה (32) בפרק 2

$$(4) \qquad \tilde{W}_{xy} = \sum_{i=1}^{m} \sum_{j=1}^{n} \tilde{U}_{ij} = \#\{i,j : X_i < Y_j\} + \frac{1}{2}\#\{i,j : X_i = Y_j\}$$

המחובר השני באגף ימין של (4) הוא מספר כל ההשוואות בין k ה-x-ים בקבוצת התיקו לבין l ה-y-ים בקבוצת התיקו, שכולם שווים, כלומר $\#\{i,j : X_i = Y_j\} = kl$.

המחובר הראשון של (4) מתקבל על ידי כל ההשוואות בין תצפיות y שאינן בקבוצת התיקו, לבין כל תצפיות x. (תצפית y הנמצאת בקבוצת התיקו אינה גדולה אף לא מאחד מה-x-ים.) נרשום זאת באמצעות המשתנים ה"מציינים":

$$\#\{i,j : x_i < y_j\} = \sum_{j=l+1}^{n} \sum_{i=1}^{m} \tilde{U}_{ij}$$

מכאן, סכום שני הביטויים בנוסחה (4) הוא

$$(5) \qquad \tilde{W}_{xy} = \#\{i,j : X_i < Y_j\} + \frac{1}{2}\#\{i,j : X_i = Y_j\} = \sum_{j=l+1}^{n} \sum_{i=1}^{m} \tilde{U}_{ij} + \frac{1}{2}kl$$

לפי מה שכבר הראינו בשוויון (15) בפרק 2,

$$\sum_{i=1}^{m} U_{ij} = S_j - j \quad , \ j > l \ \text{עבור}$$

ניתן להשתמש בצורה זו גם עבור ערכי התיקו, כיוון שבין התצפיות המושוות כאשר $j > l$ אין שום ערכים שווים.

לכן, לפי (5), נוכל לרשום את הסטטיסטי של מאן-וויטני

$$(6) \qquad \tilde{W}_{xy} = \sum_{j=l+1}^{n} (S_j - j) + \frac{1}{2}kl$$

הסטטיסטי של ווילקוקסון רשום בנוסחה (3) והסטטיסטי של מאן-וויטני בנוסחה (6). ההפרש בין שני הסטטיסטים הוא, אפוא,

$$\begin{aligned}
\tilde{W}_s - \tilde{W}_{xy} &= \left[l\left(\frac{t+1}{2}\right) + \sum_{j=l+1}^{n} S_j \right] - \left[\sum_{j=l+1}^{n} (S_j - j) + \frac{1}{2}kl \right] \\
&= l\left(\frac{t+1}{2}\right) - \left[-\frac{n(n+1)}{2} + \frac{l(l+1)}{2} \right] - \frac{kl}{2} \\
&= \frac{1}{2}\left[n(n+1) - l(l+1) + l(t+1) - kl \right] \\
&= \frac{1}{2}\left[n(n+1) + l(t-l-k) \right] = \frac{1}{2}n(n+1)
\end{aligned}$$

השוויון האחרון נובע מהעובדה ש- $t = k + l$. ♣

נספח 2ג. חישוב התוחלת והשונות של הסטטיסטי של מאן-וויטני תחת מודל אלטרנטיבי

נתונים שני מדגמים בלתי תלויים: $X_1,X_2,\ldots,X_m$; $Y_1,Y_2,\ldots,Y_n$ כאשר $X_i \sim F$ ו-$Y_j \sim G$.

הסטטיסטי של מאן-וויטני, נוסחה (13) בפרק 2, מוגדר על ידי $W_{xy}=\sum_{i=1}^{m}\sum_{j=1}^{n}U_{ij}$

כאשר
$$U_{ij}=\begin{cases}1 & X_i<Y_j\\ 0 & X_i>Y_j\end{cases}$$

לשם חישוב התוחלת נזדקק לפרמטר הבא:

$$p_1=P(X<Y) \qquad (7)$$

כאשר X ו-Y משתנים בלתי תלויים מהתפלגות F ו-G, בהתאמה.

התוחלת של כל אחד מהמציינים היא

$$EU_{ij}=P(X_i<Y_j)=p_1 \qquad (8)$$

התוחלת של W_{xy} מתקבלת כסכום התוחלות של המשתנים המציינים

$$EW_{xy}=E\sum_{i=1}^{m}\sum_{j=1}^{n}U_{ij}=\sum_{i=1}^{m}\sum_{j=1}^{n}EU_{ij}=\sum_{i=1}^{m}\sum_{j=1}^{n}P(U_{ij}=1)=mnp_1$$

כלומר, קיבלנו את הנוסחה עבור התוחלת

$$EW_{xy}=mnp_1$$

כאשר p_1 מוגדר בנוסחה (7).

לגבי השונות נזדקק לשני פרמטרים נוספים:

$$p_2=P\{(X<Y_1)\cap(X<Y_2)\} \qquad (9)$$

כאשר X, Y_1 ו-Y_2 הם שלושה משתנים בלתי תלויים, $X\sim F$, $Y_1,Y_2\sim G$.

$$p_3=P\{(X_1<Y)\cap(X_2<Y)\} \qquad (10)$$

כאשר X_1, X_2 ו-Y הם שלושה משתנים בלתי תלויים, $X_1,X_2\sim F$; $Y\sim G$.

חישוב השונות מתבסס על הנוסחה הידועה לחישוב שונות של סכום משתנים מקריים.

$$Var\left(\sum_{i=1}^{m}\sum_{j=1}^{n}U_{ij}\right)=\sum_{i=1}^{m}\sum_{j=1}^{n}Var(U_{ij})+\sum_{i}\sum_{j}\sum_{k}\sum_{l}Cov(U_{ij},U_{kl}) \qquad (11)$$

סכום השונויות המשותפות עובר על כל האינדקסים $1\le i,k\le m$, $1\le j,l\le n$, באופן ש-$U_{ij}\neq U_{kl}$ (כלומר, $k\neq i$ או $l\neq j$).

בביטוי הראשון השונויות כולן שוות. נרשום, למשל, עבור $i=1,j=1$ (האינדקסים של X ושל Y יכולים להיות שווים!).

$$Var(U_{11})=p_1(1-p_1) \qquad (12)$$

לגבי הביטוי השני, אם שני האינדקסים של המציינים הם שונים ($i \neq k$ וגם $j \neq l$), אזי המשתנים U_{ij} ו- U_{kl} הם בלתי תלויים, כי הם מבוססים על 4 משתנים בלתי תלויים X_i, Y_j, X_k, Y_l . לכן בכל המקרים הללו השונות המשותפת מתאפסת. נותר, אפוא, לחשב את השונויות המשותפות רק במקרה שאחד משני זוגות האינדקסים שווים, $i=k$ או $j=l$. נבדוק את המקרים הללו.

א) נניח $i=k$ אולם $j \neq l$. ניקח לדוגמה $i=k=1, j=1, l=2$.

$$Cov(U_{11}, U_{12}) = EU_{11}U_{12} - EU_{11}EU_{12} \tag{13}$$

התוחלות של המציינים הבודדים נתונות בנוסחה (8). תוחלת המכפלה היא

$$\begin{aligned} EU_{11}U_{12} &= P(\{(U_{11}=1) \cap (U_{12}=1)\} \\ &= P\{(X_1 < Y_1) \cap (X_1 < Y_2)\} = p_2 \end{aligned} \tag{14}$$

כאשר p_2 הוגדר בנוסחה (9).

מכאן השונות המשותפת במקרה זה מתקבלת על ידי הצבת הנוסחאות (8) ו-(14) בנוסחה (13):

$$Cov(U_{11}, U_{12}) = p_2 - p_1^2 \tag{15}$$

מספר השונויות המשותפות מהצורה הזאת הוא $mn(n-1)$, כיוון שאת האינדקס של X ניתן לבחור ב-m אפשרויות ואת שני ה-Y-ים השונים ניתן לבחור ב-$n(n-1)$ אפשרויות.

ב) נניח $j=l$ אולם $i \neq k$. ניקח לדוגמה $i=1, k=2, j=l=1$. בדומה למקרה א,

$$Cov(U_{11}, U_{21}) = EU_{11}U_{21} - EU_{11}EU_{21}$$

תוחלת המכפלה היא

$$\begin{aligned} EU_{11}U_{21} &= P(\{(U_{11}=1) \cap (U_{21}=1)\} \\ &= P\{(X_1 < Y_1) \cap (X_2 < Y_1)\} = p_3 \end{aligned}$$

כאשר p_3 הוגדר בנוסחה (10). מכאן השונות המשותפת במקרה זה המתקבלת

$$Cov(U_{11}, U_{21}) = p_3 - p_1^2 \tag{16}$$

מספר השונויות המשותפות הללו הוא $mn(m-1)$.

נציב את התוצאות (12), (15) ו-(16) בנוסחת השונות (11) ונקבל

$$\begin{aligned} Var(W_{xy}) &= mnp_1(1-p_1) + mn(n-1)(p_2 - p_1^2) \\ &\quad + mn(m-1)(p_3 - p_1^2) \end{aligned} \tag{17}$$

♣

נבדוק את נכונות הנוסחה למקרה ששתי ההתפלגויות שוות (השערת האפס נכונה). יש לחשב את ערכם של שלושת הפרמטרים שהגדרנו פה.

(1) X ו-Y בלתי תלויים שווי התפלגות ולכן

$$p_1 = P(X < Y) = \frac{1}{2}$$

(2) X, Y_1 ו-Y_2 הם שלושה משתנים בלתי תלויים שווי התפלגות ולכן, מטעמי סימטרייה, ההסתברות ש-X הוא המינימלי מביניהם היא

$$p_2 = P\left\{(X < Y_1) \cap (X < Y_2)\right\} = \frac{1}{3}$$

(3) X_1, X_2 ו-Y הם שלושה משתנים בלתי תלויים, שווי התפלגות ולכן ההסתברות ש-Y הוא המקסימלי ביניהם היא

$$p_3 = P\left\{(X_1 < Y) \cap (X_2 < Y)\right\} = \frac{1}{3}$$

הצבת תוצאות אלה בנוסחת השונות (17) נותנת את השונות עבור המודל בהשערת האפס:

$$Var_{H_0}(W_{xy}) = mn\frac{1}{4} + mn(n-1)\left(\frac{1}{3} - \frac{1}{4}\right) + mn(m-1)\left(\frac{1}{3} - \frac{1}{4}\right)$$

$$= \frac{mn(m+n+1)}{12} = \frac{mn(N+1)}{12}$$

זו הנוסחה שהתקבלה עבור השונות של W_s (או W_{xy}) ישירות, נוסחה (9) בפרק 2.

♣

נספח 3. חישוב התוחלת והשונות של הסטטיסטי של אנסרי-ברדלי, במקרה ש-N אי־זוגי

משפט. התוחלת והשונות של האוכלוסייה $\left\{1\ \ 2\ \ 3\ \ \cdots\ \ \frac{N+1}{2}\ \ \cdots\ \ 3\ \ 2\ \ 1\right\}$ הן

$$\mu = \frac{(N+1)^2}{4N}$$

$$\sigma^2 = \frac{(N+1)(N-1)(N^2+3)}{48N^2}$$

הוכחה: את הפרטים הוכיחו בעצמכם (תרגיל 1 בפרק 3). כל אחד מ-N הערכים ברשימה $\left\{1\ \ 2\ \ 3\ \ \cdots\ \ \frac{N+1}{2}\ \ \cdots\ \ 3\ \ 2\ \ 1\right\}$ מתקבל בהסתברות שווה. יהי U תוצאה של בחירת איבר אחד מהאוכלוסייה באופן אקראי. התפלגות U זהה להתפלגות האוכלוסייה. התוחלת של האוכלוסייה היא, אפוא:

$$\mu = EU = \frac{2}{N}\left[1+2+\cdots+\frac{N-1}{2}\right]+\frac{1}{N}\cdot\frac{N+1}{2}=\frac{(N+1)^2}{4N}$$

המומנט השני באוכלוסייה (ממוצע הריבועים):

$$EU^2 = \frac{2}{N}\left[1^2+2^2+\cdots+\left(\frac{N-1}{2}\right)^2\right]+\frac{1}{N}\cdot\left(\frac{N+1}{2}\right)^2$$
$$=\frac{(N+1)}{12N}(N^2+2N+3)$$

מכאן ששונות האוכלוסייה היא

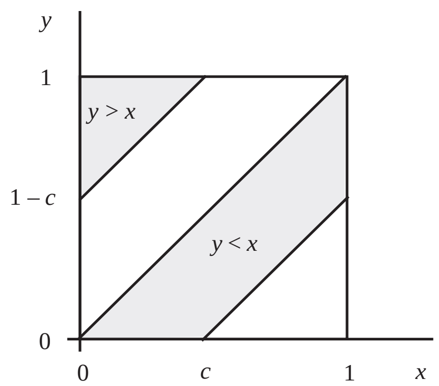
$$\sigma^2 = EU^2-\mu^2=\frac{(N+1)}{12N}(N^2+2N+3)-\left[\frac{(N+1)^2}{4N}\right]^2$$
$$=\frac{(N+1)(N-1)(N^2+3)}{48N^2}$$

♣

נספח 4. דוגמאות להתפלגויות דו־ממדיות

דוגמה 4א. אנו מביאים כאן דוגמה להתפלגות דו־ממדית שעבורה קיים שוויון ההתפלגויות השוליות של X ו-Y, אבל חציון התפלגות ההפרש $D=Y-X$ אינו אפס, כלומר $P(X<Y)\neq P(X>Y)$.

יהי (X,Y) זוג משתנים מקריים בעלי פונקציית צפיפות משותפת אחידה על התחום הכהה בציור.



השטח הכהה כולו שווה $c=\frac{c^2}{2}+\left[\frac{1}{2}-\frac{(1-c)^2}{2}\right]$ ולכן ניתן לרשום את הצפיפות:

$f(x,y)=\frac{1}{c}$ כאשר $0\le x-y\le c$ או $y-x\ge 1-c$ $(0<c<1)$.

הצפיפויות השוליות של X ושל Y הן אחידות, כלומר, $X\sim U(0,1)$ וגם $Y\sim U(0,1)$ (קל לבדוק). לכן, כמובן, גם החציונים של שני המשתנים הם שווים: $x_{.5}=y_{.5}=1/2$.

נסתכל על ההפרש $D=Y-X$. לא נחשב את התפלגות המשתנה D, אולם נברר היכן החציון של התפלגות זו. קל לחשב על פי הציור, על ידי חישוב שטח המשולש כפול הצפיפות בתחום, את ההסתברות:

$$P(D>0)=P(Y>X)=\frac{c^2}{2}\cdot\frac{1}{c}=\frac{c}{2}<\frac{1}{2}$$

ומכאן
$$P(D<0)=1-P(D>0)=1-\frac{c}{2}>\frac{1}{2}$$

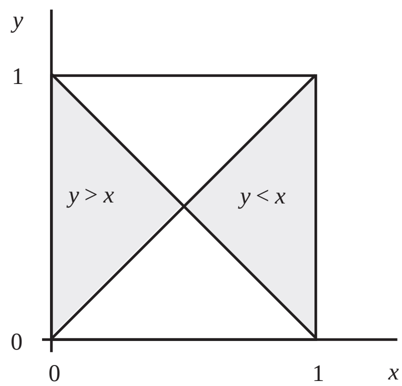
הראינו, אפוא, שהערך 0 איננו החציון של המשתנה D, ולמעשה, החציון של D חייב להיות שלילי.

שימו לב שההסתברות $P(D>0)=c/2$ יכולה להיות קטנה כרצוננו (ורחוקה מאוד מ-$1/2$), למרות שחציוני שני המשתנים X ו-Y הם שווים!

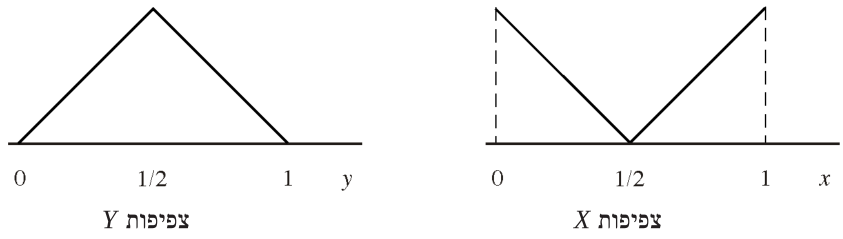
מטעמי סימטרייה ברור שבמקרה ההפוך, שבו הצפיפות המשותפת היא חיובית על התחומים **הלבנים**, חציון המשתנה D יהיה חיובי.

דוגמה 4ב. כאן אנו מביאים דוגמה למקרה הפוך מזה שבדוגמה 4א, ובו התפלגות דו־ממדית שעבורה חציון התפלגות ההפרש D הוא אפס, אבל למשתנים X ו-Y התפלגויות שוליות שונות.

יהיו (X,Y) זוג משתנים מקריים בעלי פונקציית צפיפות משותפת אחידה על התחום הכהה בציור.



ברור מן הציור שבגלל הסימטרייה $P(Y>X)=P(Y<X)=1/2$, ולכן חציון התפלגות ההפרש D הוא אפס. מצד שני, התפלגויות המשתנים X ו-Y אינן זהות. הצפיפויות המתאימות הן מהצורה (בְּדקו!)



נספח 5. חישוב שונות הסטטיסטי של יונקירי

יהיו נתונות התצפיות X_{ij}, $i=1,\ldots,k$, $j=1,\ldots,n_i$ ו- $N=\sum_{i=1}^{k} n_i$. הסטטיסטי של יונקירי מוגדר, נוסחה (24) בפרק 5: $J=\sum_{i=1}^{k}\sum_{l=1}^{i-1} W_{il}$, כאשר W_{il} הוא הסטטיסטי של מאן-וויטני להשוואת אוכלוסייה i לאוכלוסייה l.

שונות הסכום היא סכום השונויות והשונויות המשותפות של כל המשתנים. השונות של כל אחד מהמחוברים היא השונות של הסטטיסטי של ווילקוקסון, נוסחה (17) בפרק 2:

$$Var(W_{il})=\frac{n_i n_l(n_i+n_l+1)}{12}$$

נסמן את המשתנים המציינים המגדירים את הסטטיסטים של מאן-וויטני להשוואת שני המדגמים i ו-l:באופן הבא:

$$U_{rs}^{(il)}=\begin{cases}1 & X_{ir}<X_{ls}\\ 0 & X_{ir}>X_{ls}\end{cases}$$

ואז נרשום את הסטטיסטי להשוואת המדגם ה-i למדגם ה-l: $W_{il}=\sum_{r=1}^{n_i}\sum_{s=1}^{n_l} U_{rs}^{(il)}$.

תחת השערת האפס, לכל אחד מהמציינים התפלגות $B(1,1/2)$ ולכן תוחלת $1/2$ ושונות $1/4$. השונות המשותפת של שני מציינים כאלה היא, אפוא,

$$Cov\left(U_{rs}^{(i_1 l_1)},U_{uv}^{(i_2 l_2)}\right)=E\left(U_{rs}^{(i_1 l_1)}U_{uv}^{(i_2 l_2)}\right)-\frac{1}{4}$$

את השונויות המשותפות יש לחשב רק עבור המקרים שבהם משווים אחד המדגמים לשני מדגמים אחרים. נבדוק את כל האפשרויות.

א) השוואת מדגם ראשון למדגם שני ולמדגם שלישי.

$$Cov(W_{12},W_{13})=Cov\left(\sum_{r=1}^{n_1}\sum_{s=1}^{n_2}U_{rs}^{(12)},\sum_{u=1}^{n_1}\sum_{v=1}^{n_3}U_{uv}^{(13)}\right)$$
$$=\sum_{r=1}^{n_1}\sum_{s=1}^{n_2}\sum_{u=1}^{n_1}\sum_{v=1}^{n_3}Cov\left(U_{rs}^{(12)},U_{uv}^{(13)}\right)$$

השונויות המשותפות שאינן מתאפסות הן אלה שבהן האינדקס של התצפית מהמדגם הראשון זהה בשתי ההשוואות.

$$Cov\left(U_{rs}^{(12)},U_{rv}^{(13)}\right)=P(X_{1r}<X_{2s},X_{1r}<X_{3v})-\frac{1}{4}$$
$$=\frac{1}{3}-\frac{1}{4}=\frac{1}{12}$$

הסתברות החיתוך לעיל היא $1/3$ מכיוון ששלוש התצפיות בסוגריים הן בלתי תלויות ובעלות אותה התפלגות (תחת השערת האפס). שתי תמורות מתוך 6 מקיימות את הדרוש. כבר חישבנו ביטויים כאלה בנספח 2ג.

ישנם $n_1n_2n_3$ מחוברים מהצורה הזאת בהשוואה המסוימת שרשמנו. לכן השונות המשותפת כאן היא

$$Cov(W_{12},W_{13})=\sum_{r=1}^{n_1}\sum_{s=1}^{n_2}\sum_{v=1}^{n_3}Cov\left(U_{rs}^{(12)},U_{rv}^{(13)}\right)=n_1n_2n_3\cdot\frac{1}{12}$$

ב) באופן דומה נקבל את השונות המשותפת של השוואת מדגם ראשון לשני ומדגם שני לשלישי:

$$Cov\left(U_{rs}^{(12)},U_{sv}^{(23)}\right)=P(X_{1r}<X_{2s},X_{2s}<X_{3v})-\frac{1}{4}$$
$$=\frac{1}{6}-\frac{1}{4}=-\frac{1}{12}$$

רק תמורה אחת מתוך 6 מקיימת את הדרוש ולכן הסתברות החיתוך לעיל היא $1/6$. בסך הכול מקבלים כאן:

$$Cov(W_{12},W_{23})=\sum_{r=1}^{n_1}\sum_{s=1}^{n_2}\sum_{v=1}^{n_3}Cov\left(U_{rs}^{(12)},U_{sv}^{(23)}\right)=n_1n_2n_3\cdot\left(-\frac{1}{12}\right)$$

ג) נסתכל עתה על השוואת מדגם ראשון ושני למדגם שלישי. בדומה ל-א,

$$Cov\left(U_{rs}^{(13)},U_{us}^{(23)}\right)=P(X_{1r}<X_{3s},X_{2u}<X_{3s})-\frac{1}{4}$$
$$=\frac{1}{3}-\frac{1}{4}=\frac{1}{12}$$
$$Cov(W_{13},W_{23})=\sum_{r=1}^{n_1}\sum_{s=1}^{n_2}\sum_{v=1}^{n_3}Cov\left(U_{rs}^{(13)},U_{us}^{(23)}\right)=n_1n_2n_3\cdot\frac{1}{12}$$

נצרף את כל ההשוואות האפשריות מצורה א, ב ו-ג וכן את השונויות הבודדות ונקבל את השונות של הסטטיסטי של יונקירי:

$$Var(J)=\frac{1}{12}\left[\sum_{i<j} n_i n_j (n_i+n_j+1)+\sum_{\substack{i<j,l\\ l\neq j}} n_i n_j n_l\right.$$

$$\left.+\sum_{\substack{i,j<l\\ i\neq j}} n_i n_j n_l-\sum_{i<j<l} n_i n_j n_l-\sum_{l<j<i} n_i n_j n_l\right]$$

את ארבעת המחוברים האחרונים ניתן לצרף בצורה פשוטה.

היות שקיים $\sum_{\substack{i<j,l\\ l\neq j}} n_i n_j n_l=2\sum_{i<j<l} n_i n_j n_l$ וכך גם $\sum_{\substack{i,j<l\\ i\neq j}} n_i n_j n_l=2\sum_{i<j<l} n_i n_j n_l$,

לכן בסך הכול

$$\sum_{\substack{i<j,l\\ l\neq j}} n_i n_j n_l+\sum_{\substack{i,j<l\\ i\neq j}} n_i n_j n_l-\sum_{i<j<l} n_i n_j n_l-\sum_{l<j<i} n_i n_j n_l=2\sum_{i<j<l} n_i n_j n_l$$

השונות היא ,אפוא,

$$Var(J)=\frac{1}{12}\left[\sum_{i<j} n_i n_j (n_i+n_j+1)+2\sum_{i<j<l} n_i n_j n_l\right]$$

נחשב כל מחובר בנפרד.

$$\sum_{i<j} n_i n_j (n_i+n_j+1)=\frac{1}{2}\sum_{i\neq j} n_i n_j (n_i+n_j+1)$$

$$=\frac{1}{2}\left[\sum_{i=1}^{k}\sum_{j=1}^{k} n_i n_j (n_i+n_j+1)-\sum_{i=1}^{k} n_i^2(2n_i+1)\right]$$

$$=\frac{1}{2}\left[\sum_{i=1}^{k}\sum_{j=1}^{k} n_i^2 n_j+\sum_{i=1}^{k}\sum_{j=1}^{k} n_i n_j^2+\sum_{i=1}^{k}\sum_{j=1}^{k} n_i n_j-2\sum_{i=1}^{k} n_i^3-\sum_{i=1}^{k} n_i^2\right]$$

$$=\frac{1}{2}\left[2N\sum_{i=1}^{k} n_i^2+N^2-2\sum_{i=1}^{k} n_i^3-\sum_{i=1}^{k} n_i^2\right]$$

$$\sum_{i<j<l} n_i n_j n_l=\frac{1}{6}\sum_{i\neq j\neq l} n_i n_j n_l$$

$$=\frac{1}{6}\left[\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{l=1}^{k} n_i n_j n_l-3\sum_{i=1}^{k}\sum_{j\neq i}^{k} n_i^2 n_j-\sum_{i=i}^{k} n_i^3\right]$$

$$=\frac{1}{6}\left[N^3-3N\sum_{i=1}^{k}n_i^2+2\sum_{i=1}^{k}n_i^3\right]$$

סיכום כל הביטויים:

$$Var(J)=\frac{1}{12}\left[\frac{1}{2}\left(2N\sum_{i=1}^{k}n_i^2+N^2-2\sum_{i=1}^{k}n_i^3-\sum_{i=1}^{k}n_i^2\right)\right]$$
$$+\frac{1}{12}\cdot\frac{2}{6}\left[N^3-3N\sum_{i=1}^{k}n_i^2+2\sum_{i=1}^{k}n_i^3\right]$$
$$=\frac{1}{72}\left[\left(2N^3+3N^2-2\sum_{i=1}^{k}n_i^3-3\sum_{i=1}^{k}n_i^2\right)\right]$$
$$=\frac{1}{72}\left[\left(N^2(2N+3)-\sum_{i=1}^{k}n_i^2(2n_i+3)\right)\right]$$

נספח 6א. הערך המינימלי של מתאם הדרגות של ספירמן

טענה 6.2. לכל סדרת n זוגות, מתקיים $r_s \geq -1$.

הוכחה: ראה שהמתאם המינימלי מתקבל כאשר דרגות ה-y-ים בדיוק הפוכות מדרגות ה-x-ים. במקרה זה המתאם הוא $r_s \geq -1$ (תרגיל 2 בפרק 7).

נניח שה-x-ים מסודרים לפי גודלם, כלומר, $R_i = i$, $i=1,\ldots,n$ ויהיו $S_1,\ldots,S_n$ דרגות ה-y-ים המתאימות. נוכיח שאם שתי דרגות סמוכות של ה-y-ים נמצאות בסדר המתאים ל-x-ים, אזי החלפת הסדר ביניהן מקטינה את המתאם r_s.

ניעזר בנוסחה (8) בפרק 6 עבור המתאם, הרשום כפונקציה יורדת של סכום ההפרשים בין הדרגות: $r_s=1-\frac{6\sum_{i=1}^{n}D_i^2}{n(n^2-1)}$, כאשר $D_i=R_i-S_i$, $i=1,\ldots,n$.

עבור סדר מסוים נסמן $A=\sum_{i=1}^{n}D_i^2$. נסתכל עתה, למשל, על S_1 ו-S_2. נניח ש-$S_1<S_2$ ונסמן ב-B את סכום ריבועי ההפרשים, כאשר מחליפים את הסדר בין שתי הדרגות הללו. סכום ריבועי ההפרשים של כל $n-2$ הדרגות האחרות אינו משתנה. לכן ההפרש בין שני סכומי הריבועים הוא

$$B-A=\left[(S_2-1)^2+(S_1-2)^2\right]-\left[(S_1-1)^2+(S_2-2)^2\right]$$
$$=2S_2-2S_1=2(S_2-S_1)>0$$

בדיוק אותו יחס בין A ל-B מתקבל כשמחליפים כל שתי דרגות סמוכות (לאו דווקא שתי הראשונות). עבור החלפת S_k ו-S_{k+1}, בהנחה שהם היו מראש בסדר "הנכון", כלומר, $S_k < S_{k+1}$, נקבל (בדקו את החישוב):

$$B-A=\left[(S_{k+1}-k)^2+(S_k-(k+1))^2\right]$$
$$-\left[(S_k-k)^2+(S_{k+1}-(k+1))^2\right]=2(S_{k+1}-S_k)>0$$

זאת אומרת, בכל מקרה שדרגות סמוכות של y הן בסדר המתאים לסדר דרגות x, אם מחליפים את הסדר ביניהן סכום הריבועים $\sum_{i=1}^{n} D_i^2$ עולה (ולכן, כמובן, מקדם המתאם יורד).

המסקנה מכך היא שכל עוד ישנן דרגות של y הנמצאות בסדר "הנכון" ($S_k < S_{k+1}$), ניתן להקטין את המתאם על ידי החלפת הסדר ביניהן. במקרה הקיצוני, שבו כל הדרגות של y מסודרות בסדר הפוך – אין אפשרות להחלפה שתקטין את המתאם ולכן זהו המקרה שבו המתאם הוא מינימלי. ♣

נ□ראה כאן דוגמה לסדרת החלפות של דרגות סמוכות כדי להגיע למתאם המינימלי. בכל שלב החלפנו שתי דרגות סמוכות. בשלב האחרון הגענו לרשימת הדרגות ההפוכות. מסומנות באות עבה הדרגות הסמוכות שהוחלפו בכל שלב.

דרגות ה-x-ים הן: 1 2 3 4

	דרגות y				$\sum_{i=1}^{4} D_i^2$	r_s
הדרגות המקוריות	3	1	2	4	6	.4
החלפה 1	3	**2**	**1**	4	8	.2
החלפה 2	3	2	**4**	**1**	14	–.4
החלפה 3	3	**4**	**2**	1	18	–.8
החלפה 4	**4**	**3**	2	1	20	–1.0

ברור כי מכל תמורה של דרגות y ניתן להגיע לתמורה ההפוכה עם מתאם מינימלי על ידי החלפה סדרתית של שתי דרגות סמוכות.

נספח 26ב. השונות של מתאם קנדל תחת אי־תלות

נסתכל על הסטטיסטי N_c. ראינו שניתן לרשום אותו על ידי סכום של משתנים מציינים, לפי נוסחה (22) בפרק 6:

$$N_c = \sum_{i=1}^{n}\sum_{j>i} H_{ij}$$

כאשר המשתנים המציינים מוגדרים ב-(21) שם.

שונות הסכום שווה לסכום כל השונויות והשונויות המשותפות:

$$Var(N_c) = Var\left(\sum_{i=1}^{n}\sum_{j>i} H_{ij}\right)$$

$$= \sum_{i<j}\sum Var(H_{ij}) + \sum_{i<j}\sum \sum_{k<l}\sum Cov(H_{ij}, H_{kl}) \qquad (18)$$

נסתכל, אפוא, על כל ביטוי בנפרד. נזכור כי כל החישוב נערך בהנחה של אי־תלות בין X ל-Y. כמו כן, נניח שה-X-ים כבר מסודרים לפי גודלם, ולכן המשתנים המציינים מתקבלים על ידי

$$H_{ij} = \begin{cases} 1 & y_i > y_j \\ 0 & y_i < y_j \end{cases}$$

המשתנה H_{ij} הוא משתנה ברנולי $H_{ij} \sim B(1,1/2)$. לפיכך $EH_{ij} = 1/2$ ו-$Var(H_{ij}) = 1/4$.

נסתכל עתה על השונויות המשותפות $Cov(H_{ij}, H_{kl})$. בכל מקרה,

$$Cov(H_{ij}, H_{kl}) = EH_{ij}H_{kl} - EH_{ij}EH_{kl} = EH_{ij}H_{kl} - 1/4 \qquad (19)$$

נבדוק את המקרים השונים האפשריים.

א) כאשר זוג האינדקסים $i<j$ שונה מהזוג $k<l$ אזי $Cov(H_{ij}, H_{kl}) = 0$.

ב) נניח ש-$i=k$, ו-$j \neq l$. ניקח לדוגמה $i=k=1$, $j=2$, ו-$l=3$.

אזי $EH_{12}H_{13} = P(Y_1 > Y_2, Y_1 > Y_3) = 1/3$, כיוון ש Y_1, Y_2, Y_3 הם שלושה משתנים בלתי תלויים שווי התפלגות, וכל אחת מ-$3! = 6$ התמורות של הסדר שלהן היא בעלת אותה הסתברות. שתיים מן התמורות מקיימות את התנאי ש-Y_1 הוא הגבוה ביותר: $Y_2 < Y_3 < Y_1$ ו- $Y_3 < Y_2 < Y_1$. השונות המשותפת היא, אפוא, לפי (19),

$$Cov(H_{12}, H_{13}) = EH_{12}H_{13} - 1/4 = 1/3 - 1/4 = 1/12 \qquad (20)$$

בַּסכום (18) ישנם $2\binom{n}{3}$ איברים כאלה. אלה הם כל הבחירות של שלושה אינדקסים i,j,l השונים זה מזה, כאשר i הוא הקטן ביותר, ו-j ו-l יכולים להתחלף ביניהם.

ג) נניח ש-$j=l$, ו-$i \neq k$. ניקח לדוגמה $i=1$, $j=l=2$ ו-$k=3$.

אזי $EH_{12}H_{32}=P(Y_1>Y_2,Y_3>Y_2)=1/3$, כיון שאלה שוב שלושה משתנים בלתי תלויים שווי התפלגות, ושתיים מן התמורות של הסדר ביניהן מקיימות את התנאי ש-Y_2 הוא הנמוך ביותר. השונות המשותפת היא, אפוא, כמו במקרה ב

$$Cov(H_{12},H_{32})=EH_{12}H_{32}-1/4=1/3-1/4=1/12 \qquad (21)$$

בַסכום (18) ישנם, כמו במקרה ב, $2\binom{n}{3}$ איברים כאלה.

ד) נניח ש-$i=l$, ו-$j\neq k$. ניקח לדוגמה $i=l=1$, $j=2$, ו-$k=3$.

אזי $EH_{12}H_{31}=P(Y_1>Y_2,Y_3>Y_1)=1/6$, כיוון שיש רק תמורה אחת המקיימת את התנאי $Y_2<Y_1<Y_3$. לפיכך השונות המשותפת היא

$$Cov(H_{i12},H_{31})=EH_{12}H_{31}-1/4=1/6-1/4=-1/12 \qquad (22)$$

מספר האיברים הללו בַסכום (18) הוא $\binom{n}{3}$.

ה) נניח ש-$k=j$, ו-$i\neq l$. ניקח לדוגמה $i=1$, $k=j=2$, ו-$l=3$

כמו במקרה ד, $EH_{12}H_{23}=P(Y_1>Y_2,Y_2>Y_3)=1/6$ (רק תמורה אחת מתאימה) ולכן גם במקרה זה

$$Cov(H_{12},H_{23})=EH_{12}H_{23}-1/4=1/6-1/4=-1/12 \qquad (23)$$

זו בדיוק אפשרות סימטרית לאפשרות ד. מספר האיברים הללו בַסכום (18) גם הוא $\binom{n}{3}$.

נסכם עתה את הביטויים שקיבלנו בנוסחאות (20) עד (23) ונקבל את שונות הסכום (18):

$$\begin{aligned}Var(N_c)&=\binom{n}{2}Var(H_{12})+2\binom{n}{3}\left[Cov(H_{12},H_{13})+Cov(H_{12},H_{32})\right]\\&\quad+\binom{n}{3}\left[Cov(H_{12},H_{31})+Cov(H_{12},H_{23})\right]\\&=\binom{n}{2}\frac{1}{4}+2\binom{n}{3}\left[\frac{1}{12}+\frac{1}{12}\right]+\binom{n}{3}\left[-\frac{1}{12}-\frac{1}{12}\right]\\&=\binom{n}{2}\frac{1}{4}+\binom{n}{3}\frac{2}{12}=\frac{n(n-1)}{2}\cdot\frac{1}{4}+\frac{n(n-1)(n-2)}{6}\cdot\frac{1}{6}\\&=\frac{n(n-1)(5+2n)}{72}\end{aligned}$$

כלומר, קיבלנו

$$Var(N_c)=\frac{n(n-1)(5+2n)}{72} \qquad (24)$$

נזכיר את הקשר בין N_c ל-N_d: $N_c + N_d = \binom{n}{2}$. שונות ההפרש היא, לפיכך,

$$Var(N_c - N_d) = Var\left[2N_c - \binom{n}{2}\right] = 4Var(N_c)$$
$$= \frac{4n(n-1)(5+2n)}{72} = \frac{2(5+2n)}{9n(n-1)}$$

זוהי הנוסחה (25) הרשומה במשפט 6.4 בפרק 6. ♣

נספח 7א. הקשר בין הסטטיסטי של פרידמן לבין מתאמי דרגות בין הבלוקים

משפט 7.3. נסתכל על ניסוי של בלוקים אקראיים עם n שופטים (בלוקים) המדרגים k טיפולים.

יהי r_{il} מתאם הדרגות של ספירמן בין נבדק i לנבדק l ויהי $\bar{r}$ ממוצע ערכי r_{il}, כלומר,

$$\bar{r} = \frac{1}{\binom{n}{2}} \sum_{i=1}^{n} \sum_{l>i} r_{il} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{l \neq i} r_{il}$$

אזי מתקיים: $\bar{r} = \frac{nW-1}{n-1}$, כאשר W הוא מקדם ההסכמה בין הנבדקים.

הוכחה: נזכיר את הגדרת מקדם ההסכמה, נוסחה (12) בפרק 7:

$$W = \frac{12S}{n^2 k(k^2-1)}$$

R_{ij} היא דרגת הטיפול j בבלוק i (בין כל k נתוני הבלוק). לשם נוחיות הכתיבה, נסמן את הסטיות $e_{ij} = R_{ij} - \frac{1}{2}(k+1)$.

לדוגמה, מתאם הדרגות בין שני הבלוקים הראשונים ניתן להירשם במונחים של הסטיות e_{ij}:

$$r_{12} = \frac{12\sum_{j=1}^{k}\left(R_{1j} - \frac{k+1}{2}\right)\left(R_{2j} - \frac{k+1}{2}\right)}{k(k^2-1)} = \frac{12\sum_{j=1}^{k} e_{1j}e_{2j}}{k(k^2-1)}$$

את מתאם הדרגות בין שני בלוקים כלשהם ניתן לרשום באופן דומה

$$r_{il} = \frac{12\sum_{j=1}^{k} e_{ij}e_{lj}}{k(k^2-1)}$$

הממוצע של המתאמים:

$$\bar{r} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{l\neq i}\frac{12\sum_{j=1}^{k} e_{ij}e_{lj}}{k(k^2-1)} = \frac{12}{n(n-1)k(k^2-1)}\sum_{i=1}^{n}\sum_{\substack{l=1\\ l\neq i}}^{n}\sum_{j=1}^{k} e_{ij}e_{lj} \tag{25}$$

כדי לחשב את הסכום לעיל נחליף את סדר הסיכום:

$$\sum_{i=1}^{n}\sum_{\substack{l=1\\ l\neq i}}^{n}\sum_{j=1}^{k} e_{ij}e_{lj} = \sum_{j=1}^{k}\left[\sum_{i=1}^{n}\sum_{\substack{l=1\\ l\neq i}}^{n} e_{ij}e_{lj}\right] = \sum_{j=1}^{k}\left[\left(\sum_{i=1}^{n} e_{ij}\right)^2 - \sum_{i=1}^{n} e_{ij}^2\right]$$

נציב את ההגדרה של e_{ij}:

$$= \sum_{j=1}^{k}\left[\sum_{i=1}^{n}\left(R_{ij}-\frac{k+1}{2}\right)\right]^2 - \sum_{j=1}^{k}\left[\sum_{i=1}^{n}\left(R_{ij}-\frac{k+1}{2}\right)^2\right]$$

$$= \sum_{j=1}^{k}\left[T_j-\frac{n(k+1)}{2}\right]^2 - \sum_{i=1}^{n}\left[\sum_{j=1}^{k}\left(R_{ij}-\frac{k+1}{2}\right)^2\right]$$

$$= S - n\frac{k(k^2-1)}{12}$$

הכפלה במקדם בנוסחה (25) נותנת את ממוצע המתאמים

$$\bar{r} = \frac{12}{n(n-1)k(k^2-1)}\left[S-\frac{nk(k^2-1)}{12}\right]$$

$$= \frac{12S}{n(n-1)k(k^2-1)} - \frac{1}{n-1} = \frac{nW-1}{n-1}$$

♣

נספח 7ב. הקשר בין הסטטיסטי של קוקרן לסטטיסטי של פרידמן

משפט 7.4. הסטטיסטי של קוקרן שווה לסטטיסטי של פרידמן עם תיקון לערכי תיקו. כלומר, $Q=\tilde{F}_r$ כאשר $\tilde{F}_r$ ו-Q מוגדרים בנוסחאות (17) ו- (23) בפרק 7.

הוכחה: יש לרשום את דרגות התצפיות בכל בלוק. היות שהתצפיות כולן הן 0 או 1, בכל בלוק יש לכל היותר שתי דרגות ממוצעות שונות. דרגת ה-0-ים היא הנמוכה ודרגת ה-1-ים היא הגבוהה.

בבלוק i יש $k-L_i$ ערכים של 0, שהדרגות המגיעות להם הן $1,\ldots,k-L_i$ וכן L_i ערכים של 1, שהדרגות המגיעות להם הן $k-L_i+1,\ldots,k$.

הדרגה הממוצעת של ערכי 0 היא $\quad \dfrac{1+(k-L_i)}{2}=\dfrac{1}{2}(k+1)-\dfrac{1}{2}L_i$

הדרגה הממוצעת של ערכי 1 היא $\quad \dfrac{k-L_i+1+k}{2}=k+\dfrac{1}{2}-\dfrac{1}{2}L_i$

סך הדרגות שקיבל טיפול j תלוי במספר ערכי 0 ובמספר ערכי 1 שהתקבלו בטיפול זה:

$$\tilde{T}_j=\sum_{i:x_{ij}=0}\left[\frac{k+1}{2}-\frac{1}{2}L_i\right]+\sum_{i:x_{ij}=1}\left[k+\frac{1}{2}-\frac{1}{2}L_i\right]$$

המחובר השני כולל B_j מחוברים (מספר ה-1-ים בעמודה j) והמחובר השני כולל $n-B_j$ מחוברים. לפיכך הסיכום לעיל המתקבל הוא

$$=(n-B_j)\frac{k+1}{2}+B_j\left(k+\frac{1}{2}\right)-\frac{1}{2}\sum_{i=1}^{n}L_i$$

$$=\frac{n(k+1)}{2}+\frac{k}{2}B_j-\frac{1}{2}\sum_{i=1}^{n}L_i$$

הסכום האחרון באגף ימין מקיים $\sum_{i=1}^{n}L_i=\sum_{j=1}^{k}B_j$ שהוא סכום כל הערכים בטבלה.

מכאן קיבלנו את סכום הדרגות (הממוצעות) בטיפול ה-j

$$\tilde{T}_j=\frac{n(k+1)}{2}+\frac{k}{2}B_j-\frac{1}{2}\sum_{j=1}^{k}B_i=\frac{n(k+1)}{2}+\frac{k}{2}(B_j-\overline{B})$$

הסטיות של סכומי הדרגות הללו מהממוצע הן

$$\tilde{T}_j-\frac{n(k+1)}{2}=\frac{k}{2}\left(B_j-\overline{B}\right)$$

סך ריבועי הסטיות המתקבל הוא, אפוא:

$$\tilde{S}=\sum_{j=1}^{k}\left(\tilde{T}_j-\frac{n(k+1)}{2}\right)^2=\frac{k^2}{4}\sum_{j=1}^{k}\left(B_j-\overline{B}\right)^2$$

לפיכך הסטטיסטי המתוקנן של פרידמן הוא:

$$F_r=\frac{12\tilde{S}}{nk(k+1)}=\frac{12k^2}{4nk(k+1)}\sum_{j=1}^{k}\left(B_j-\overline{B}\right)^2=\frac{3k\sum_{j=1}^{k}\left(B_j-\overline{B}\right)^2}{n(k+1)} \tag{26}$$

יש לחלק את F_r בביטוי המתאים כדי לתקן אותו עבור ערכי התיקו בבלוקים, לפי נוסחה (17) בפרק 7.

$$\tilde{F}_r=\frac{F_r}{1-\dfrac{1}{nk(k^2-1)}\sum_i\sum_l t_{il}(t_{il}^2-1)} \tag{27}$$

נחשב את הביטוי הקשור בתיקון בגלל קבוצות התיקו. בבלוק i גודלי קבוצות התיקו הם: $t_{i1}=k-L_i$, $t_{i2}=L_i$. מכאן מקבלים:

$$\begin{aligned}\sum_{i=1}^{n}\sum_{l}t_{il}(t_{il}^2-1)&=\sum_{i=1}^{n}\left\{t_{i1}(t_{i1}^2-1)+t_{i2}(t_{i2}^2-1)\right\}\\&=\sum_{i=1}^{n}\left\{[(k-L_i)[(k-L_i)^2-1]+L_i(L_i^2-1)\right\}\\&=nk(k^2-1)+3k\sum_{i=1}^{n}L_i^2-3k^2\sum_{i=1}^{n}L_i\end{aligned}$$

(חשבו וו□דאו את נכונות השוויון האחרון.)

המכנה של $\tilde{F}_r$, נוסחה (27), הוא, לפיכך:

$$\begin{aligned}1-\frac{1}{nk(k^2-1)}\sum_{i}\sum_{l}t_{il}(t_{il}^2-1)&=1-\frac{nk(k^2-1)+3k\sum_{i=1}^{n}L_i^2-3k^2\sum_{i=1}^{n}L_i}{nk(k^2-1)}\\&=\frac{-3k\sum_{i=1}^{n}L_i^2+3k^2\sum_{i=1}^{n}L_i}{nk(k^2-1)}\\&=\frac{3}{n(k^2-1)}\left[k\sum_{i=1}^{n}L_i-\sum_{i=1}^{n}L_i^2\right]\end{aligned}$$

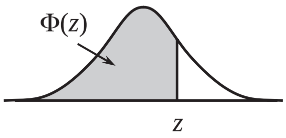
נציב את התוצאה הזאת במכנה של נוסחה (27) ונקבל, לפי (26):

$$\tilde{F}_r=\frac{3k\sum_{j=1}^{k}\left(B_j-\overline{B}\right)^2\Big/n(k+1)}{\frac{3}{n(k^2-1)}\left(k\sum_{i=1}^{n}L_i-\sum_{i=1}^{n}L_i^2\right)}=\frac{k(k-1)\sum_{j=1}^{k}\left(B_j-\overline{B}\right)^2}{k\sum_{i=1}^{n}L_i-\sum_{i=1}^{n}L_i^2}=Q$$

זוהי בדיוק הנוסחה עבור הסטטיסטי Q של קוקרן, נוסחה (23)בפרק 7. בכך הוכחנו את המשפט. ♣

טבלאות

ההתפלגות המצטברת הנורמלית סטנדרטית: $P(Z \le z) = \Phi(z)$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.00	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.10	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.20	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.30	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.40	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.50	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.60	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.70	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.80	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.90	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.00	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.10	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.20	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.30	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.40	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.50	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.60	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.70	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.80	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.90	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.00	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.10	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.20	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.30	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.40	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.50	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.60	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.70	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.80	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.90	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.00	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.10	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.20	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.30	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.40	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.50	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

z	1.282	1.645	1.960	2.326	2.576	3.090	3.291	3.719	3.891	4.265
$\Phi(z)$	.90	.95	.975	.99	.995	.999	.9995	.9999	.99995	.99999

טבלה 2. התפלגות הסטטיסטי של ווילקוקסון לשני מדגמים: $P(W_s \le k)$
(W_s – סכום הדרגות של n מתוך $N = n + m$ תצפיות)

$n=3$	m							
k	3	4	5	6	7	8	9	10
6	.0500	.0286	.0179	.0119	.0083	.0061	.0045	.0035
7	.1000	.0571	.0357	.0238	.0167	.0121	.0091	.0070
8	.2000	.1143	.0714	.0476	.0333	.0242	.0182	.0140
9	.3500	.2000	.1250	.0833	.0583	.0424	.0318	.0245
10	.5000	.3143	.1964	.1310	.0917	.0667	.0500	.0385
11	.6500	.4286	.2857	.1905	.1333	.0970	.0727	.0559
12	.8000	.5714	.3929	.2738	.1917	.1394	.1045	.0804
13	.9000	.6857	.5000	.3571	.2583	.1879	.1409	.1084
14	.9500	.8000	.6071	.4524	.3333	.2485	.1864	.1434
15	1.0000	.8857	.7143	.5476	.4167	.3152	.2409	.1853
16		.9429	.8036	.6429	.5000	.3879	.3000	.2343
17		.9714	.8750	.7262	.5833	.4606	.3636	.2867
18		1.0000	.9286	.8095	.6667	.5394	.4318	.3462
19			.9643	.8690	.7417	.6121	.5000	.4056
20			.9821	.9167	.8083	.6848	.5682	.4685
21			1.0000	.9524	.8667	.7515	.6364	.5315
22				.9762	.9083	.8121	.7000	.5944

$n=4$	m						
k	4	5	6	7	8	9	10
10	.0143	.0079	.0048	.0030	.0020	.0014	.0010
11	.0286	.0159	.0095	.0061	.0040	.0028	.0020
12	.0571	.0317	.0190	.0121	.0081	.0056	.0040
13	.1000	.0556	.0333	.0212	.0141	.0098	.0070
14	.1714	.0952	.0571	.0364	.0242	.0168	.0120
15	.2429	.1429	.0857	.0545	.0364	.0252	.0180
16	.3429	.2063	.1286	.0818	.0545	.0378	.0270
17	.4429	.2778	.1762	.1152	.0768	.0531	.0380
18	.5571	.3651	.2381	.1576	.1071	.0741	.0529
19	.6571	.4524	.3048	.2061	.1414	.0993	.0709
20	.7571	.5476	.3810	.2636	.1838	.1301	.0939
21	.8286	.6349	.4571	.3242	.2303	.1650	.1199
22	.9000	.7222	.5429	.3939	.2848	.2070	.1518
23	.9429	.7937	.6190	.4636	.3414	.2517	.1868
24	.9714	.8571	.6952	.5364	.4040	.3021	.2268
25	.9857	.9048	.7619	.6061	.4667	.3552	.2697
26	1.0000	.9444	.8238	.6758	.5333	.4126	.3177
27		.9683	.8714	.7364	.5960	.4699	.3666
28		.9841	.9143	.7939	.6586	.5301	.4196
29		.9921	.9429	.8424	.7152	.5874	.4725
30		1.0000	.9667	.8848	.7697	.6448	.5275

טבלה 2. התפלגות הסטטיסטי של ווילקוקסון לשני מדגמים: $P(W_s \le w)$ (המשך)

$n=5$	m						$n=6$, m				
k	5	6	7	8	9	10	6	7	8	9	10
15	.0040	.0022	.0013	.0008	.0005	.0003					
16	.0079	.0043	.0025	.0016	.0010	.0007					
17	.0159	.0087	.0051	.0031	.0020	.0013					
18	.0278	.0152	.0088	.0054	.0035	.0023					
19	.0476	.0260	.0152	.0093	.0060	.0040					
20	.0754	.0411	.0240	.0148	.0095	.0063					
21	.1111	.0628	.0366	.0225	.0145	.0097	.0011	.0006	.0003	.0002	.0001
22	.1548	.0887	.0530	.0326	.0210	.0140	.0022	.0012	.0007	.0004	.0002
23	.2103	.1234	.0745	.0466	.0300	.0200	.0043	.0023	.0013	.0008	.0005
24	.2738	.1645	.1010	.0637	.0415	.0276	.0076	.0041	.0023	.0014	.0009
25	.3452	.2143	.1338	.0855	.0559	.0376	.0130	.0070	.0040	.0024	.0015
26	.4206	.2684	.1717	.1111	.0734	.0496	.0206	.0111	.0063	.0038	.0024
27	.5000	.3312	.2159	.1422	.0949	.0646	.0325	.0175	.0100	.0060	.0037
28	.5794	.3961	.2652	.1772	.1199	.0823	.0465	.0256	.0147	.0088	.0055
29	.6548	.4654	.3194	.2176	.1489	.1032	.0660	.0367	.0213	.0128	.0080
30	.7262	.5346	.3775	.2618	.1818	.1272	.0898	.0507	.0296	.0180	.0112
31	.7897	.6039	.4381	.3108	.2188	.1548	.1201	.0688	.0406	.0248	.0156
32	.8452	.6688	.5000	.3621	.2592	.1855	.1548	.0903	.0539	.0332	.0210
33	.8889	.7316	.5619	.4165	.3032	.2198	.1970	.1171	.0709	.0440	.0280
34	.9246	.7857	.6225	.4716	.3497	.2567	.2424	.1474	.0906	.0567	.0363
35	.9524	.8355	.6806	.5284	.3986	.2970	.2944	.1830	.1142	.0723	.0467
36	.9722	.8766	.7348	.5835	.4491	.3393	.3496	.2226	.1412	.0905	.0589
37	.9841	.9113	.7841	.6379	.5000	.3839	.4091	.2669	.1725	.1119	.0736
38	.9921	.9372	.8283	.6892	.5509	.4296	.4686	.3141	.2068	.1361	.0903
39	.9960	.9589	.8662	.7382	.6014	.4765	.5314	.3654	.2454	.1638	.1099
40	1.0000	.9740	.8990	.7824	.6503	.5235	.5909	.4178	.2864	.1942	.1317
41		.9848	.9255	.8228	.6968	.5704	.6504	.4726	.3310	.2280	.1566
42		.9913	.9470	.8578	.7408	.6161	.7056	.5274	.3773	.2643	.1838
43		.9957	.9634	.8889	.7812	.6607	.7576	.5822	.4259	.3035	.2139
44		.9978	.9760	.9145	.8182	.7030	.8030	.6346	.4749	.3445	.2461
45		1.0000	.9848	.9363	.8511	.7433	.8452	.6859	.5251	.3878	.2811
46			.9912	.9534	.8801	.7802	.8799	.7331	.5741	.4320	.3177
47			.9949	.9674	.9051	.8145	.9102	.7774	.6227	.4773	.3564
48			.9975	.9775	.9266	.8452	.9340	.8170	.6690	.5227	.3962
49			.9987	.9852	.9441	.8728	.9535	.8526	.7136	.5680	.4374
50			1.0000	.9907	.9585	.8968	.9675	.8829	.7546	.6122	.4789
51				.9946	.9700	.9177	.9794	.9097	.7932	.6555	.5211
52				.9969	.9790	.9354	.9870	.9312	.8275	.6965	.5626
53				.9984	.9855	.9504	.9924	.9493	.8588	.7357	.6038
54				.9992	.9905	.9624	.9957	.9633	.8858	.7720	.6436
55				1.0000	.9940	.9724	.9978	.9744	.9094	.8058	.6823
56					.9965	.9800	.9989	.9825	.9291	.8362	.7189
57					.9980	.9860	1.0000	.9889	.9461	.8639	.7539

טבלה 2. התפלגות הסטטיסטי של ווילקוקסון לשני מדגמים: $P(W_s \le w)$ (המשך)

$n=7$	m			
k	7	8	9	10
28	.0003	.0002	.0001	.0001
29	.0006	.0003	.0002	.0001
30	.0012	.0006	.0003	.0002
31	.0020	.0011	.0006	.0004
32	.0035	.0019	.0010	.0006
33	.0055	.0030	.0017	.0010
34	.0087	.0047	.0026	.0015
35	.0131	.0070	.0039	.0023
36	.0189	.0103	.0058	.0034
37	.0265	.0145	.0082	.0048
38	.0364	.0200	.0115	.0068
39	.0487	.0270	.0156	.0093
40	.0641	.0361	.0209	.0125
41	.0825	.0469	.0274	.0165
42	.1043	.0603	.0356	.0215
43	.1297	.0760	.0454	.0277
44	.1588	.0946	.0571	.0351
45	.1914	.1159	.0708	.0439
46	.2279	.1405	.0869	.0544
47	.2675	.1678	.1052	.0665
48	.3100	.1984	.1261	.0806
49	.3552	.2317	.1496	.0966
50	.4024	.2679	.1755	.1148
51	.4508	.3063	.2039	.1349
52	.5000	.3472	.2349	.1574
53	.5492	.3894	.2680	.1819
54	.5976	.4333	.3032	.2087
55	.6448	.4775	.3403	.2374
56	.6900	.5225	.3788	.2681
57	.7325	.5667	.4185	.3004
58	.7721	.6106	.4591	.3345
59	.8086	.6528	.5000	.3698
60	.8412	.6937	.5409	.4063
61	.8703	.7321	.5815	.4434
62	.8957	.7683	.6212	.4811
63	.9175	.8016	.6597	.5189
64	.9359	.8322	.6968	.5566
65	.9513	.8595	.7320	.5937
66	.9636	.8841	.7651	.6302
67	.9735	.9054	.7961	.6655
68	.9811	.9240	.8245	.6996

$n=8$	m		
k	8	9	10
36	.0001	.0000	.0000
37	.0002	.0001	.0000
38	.0003	.0002	.0001
39	.0005	.0003	.0002
40	.0009	.0005	.0003
41	.0015	.0008	.0004
42	.0023	.0012	.0007
43	.0035	.0019	.0010
44	.0052	.0028	.0015
45	.0074	.0039	.0022
46	.0103	.0056	.0031
47	.0141	.0076	.0043
48	.0190	.0103	.0058
49	.0249	.0137	.0078
50	.0325	.0180	.0103
51	.0415	.0232	.0133
52	.0524	.0296	.0171
53	.0652	.0372	.0217
54	.0803	.0464	.0273
55	.0974	.0570	.0338
56	.1172	.0694	.0416
57	.1393	.0836	.0506
58	.1641	.0998	.0610
59	.1911	.1179	.0729
60	.2209	.1383	.0864
61	.2527	.1606	.1015
62	.2869	.1852	.1185
63	.3227	.2117	.1371
64	.3605	.2404	.1577
65	.3992	.2707	.1800
66	.4392	.3029	.2041
67	.4796	.3365	.2299
68	.5204	.3715	.2574
69	.5608	.4074	.2863
70	.6008	.4442	.3167
71	.6395	.4813	.3482
72	.6773	.5187	.3809
73	.7131	.5558	.4143
74	.7473	.5926	.4484
75	.7791	.6285	.4827
76	.8089	.6635	.5173

טבלה 2. התפלגות הסטטיסטי של ווילקוקסון לשני מדגמים: $P(W_s \leq w)$ (המשך)

n=9	m	
k	9	10
45	.0000	.0000
46	.0000	.0000
47	.0001	.0000
48	.0001	.0001
49	.0002	.0001
50	.0004	.0002
51	.0006	.0003
52	.0009	.0005
53	.0014	.0007
54	.0020	.0011
55	.0028	.0015
56	.0039	.0021
57	.0053	.0028
58	.0071	.0038
59	.0094	.0051
60	.0122	.0066
61	.0157	.0086
62	.0200	.0110
63	.0252	.0140
64	.0313	.0175
65	.0385	.0217
66	.0470	.0267
67	.0567	.0326
68	.0680	.0394
69	.0807	.0474
70	.0951	.0564
71	.1112	.0667
72	.1290	.0782
73	.1487	.0912
74	.1701	.1055
75	.1933	.1214
76	.2181	.1388
77	.2447	.1577
78	.2729	.1781
79	.3024	.2001
80	.3332	.2235
81	.3652	.2483
82	.3981	.2745
83	.4317	.3019
84	.4657	.3304
85	.5000	.3598
86	.5343	.3901
87	.5683	.4211
88	.6019	.4524
89	.6348	.4841
90	.6668	.5159

$n=10$, $m=10$			
k		k	
55	.0000	81	.0376
56	.0000	82	.0446
57	.0000	83	.0526
58	.0000	84	.0615
59	.0001	85	.0716
60	.0001	86	.0827
61	.0002	87	.0952
62	.0002	88	.1088
63	.0004	89	.1237
64	.0005	90	.1399
65	.0008	91	.1575
66	.0010	92	.1763
67	.0014	93	.1965
68	.0019	94	.2179
69	.0026	95	.2406
70	.0034	96	.2644
71	.0045	97	.2894
72	.0057	98	.3153
73	.0073	99	.3421
74	.0093	100	.3697
75	.0116	101	.3980
76	.0144	102	.4267
77	.0177	103	.4559
78	.0216	104	.4853
79	.0262	105	.5147
80	.0315	106	.5441

טבלה 3. התפלגות הסטטיסטי של מבחן הסימן: $P(S_n \le k)$

k	n = 2	3	4	5	6	7	8	9	10
0	.2500	.1250	.0625	.0313	.0156	.0078	.0039	.0020	.0010
1	.7500	.5000	.3125	.1875	.1094	.0625	.0352	.0195	.0107
2	1.0000	.8750	.6875	.5000	.3438	.2266	.1445	.0898	.0547
3		1.0000	.9375	.8125	.6563	.5000	.3633	.2539	.1719
4			1.0000	.9688	.8906	.7734	.6367	.5000	.3770
5				1.0000	.9844	.9375	.8555	.7461	.6230
6					1.0000	.9922	.9648	.9102	.8281
7						1.0000	.9961	.9805	.9453
8							1.0000	.9980	.9893
9								1.0000	.9990
10									1.0000

k	n = 11	12	13	14	15	16	17	18	19	20
0	.0005	.0002	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000
1	.0059	.0032	.0017	.0009	.0005	.0003	.0001	.0001	.0000	.0000
2	.0327	.0193	.0112	.0065	.0037	.0021	.0012	.0007	.0004	.0002
3	.1133	.0730	.0461	.0287	.0176	.0106	.0064	.0038	.0022	.0013
4	.2744	.1938	.1334	.0898	.0592	.0384	.0245	.0154	.0096	.0059
5	.5000	.3872	.2905	.2120	.1509	.1051	.0717	.0481	.0318	.0207
6	.7256	.6128	.5000	.3953	.3036	.2272	.1662	.1189	.0835	.0577
7	.8867	.8062	.7095	.6047	.5000	.4018	.3145	.2403	.1796	.1316
8	.9673	.9270	.8666	.7880	.6964	.5982	.5000	.4073	.3238	.2517
9	.9941	.9807	.9539	.9102	.8491	.7728	.6855	.5927	.5000	.4119
10	.9995	.9968	.9888	.9713	.9408	.8949	.8338	.7597	.6762	.5881

k	n = 21	22	23	24	25	26	27	28	29	30
0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
1	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
3	.0007	.0004	.0002	.0001	.0001	.0000	.0000	.0000	.0000	.0000
4	.0036	.0022	.0013	.0008	.0005	.0003	.0002	.0001	.0001	.0000
5	.0133	.0085	.0053	.0033	.0020	.0012	.0008	.0005	.0003	.0002
6	.0392	.0262	.0173	.0113	.0073	.0047	.0030	.0019	.0012	.0007
7	.0946	.0669	.0466	.0320	.0216	.0145	.0096	.0063	.0041	.0026
8	.1917	.1431	.1050	.0758	.0539	.0378	.0261	.0178	.0121	.0081
9	.3318	.2617	.2024	.1537	.1148	.0843	.0610	.0436	.0307	.0214
10	.5000	.4159	.3388	.2706	.2122	.1635	.1239	.0925	.0680	.0494
11	.6682	.5841	.5000	.4194	.3450	.2786	.2210	.1725	.1325	.1002
12	.8083	.7383	.6612	.5806	.5000	.4225	.3506	.2858	.2291	.1808
13	.9054	.8569	.7976	.7294	.6550	.5775	.5000	.4253	.3555	.2923
14	.9608	.9331	.8950	.8463	.7878	.7214	.6494	.5747	.5000	.4278
15	.9867	.9738	.9534	.9242	.8852	.8365	.7790	.7142	.6445	.5722

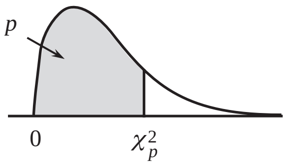
טבלה 4. התפלגות הסטטיסטי של ווילקוקסון למדגם מזווג: $P(V^+ \le k)$

	n									
k	1	2	3	4	5	6	7	8	9	10
0	.5000	.2500	.1250	.0625	.0313	.0156	.0078	.0039	.0020	.0010
1	1.0000	.5000	.2500	.1250	.0625	.0313	.0156	.0078	.0039	.0020
2		.7500	.3750	.1875	.0938	.0469	.0234	.0117	.0059	.0029
3		1.0000	.6250	.3125	.1563	.0781	.0391	.0195	.0098	.0049
4			.7500	.4375	.2188	.1094	.0547	.0273	.0137	.0068
5			.8750	.5625	.3125	.1563	.0781	.0391	.0195	.0098
6			1.0000	.6875	.4063	.2188	.1094	.0547	.0273	.0137
7				.8125	.5000	.2813	.1484	.0742	.0371	.0186
8				.8750	.5938	.3438	.1875	.0977	.0488	.0244
9				.9375	.6875	.4219	.2344	.1250	.0645	.0322
10				1.0000	.7813	.5000	.2891	.1563	.0820	.0420
11					.8438	.5781	.3438	.1914	.1016	.0527
12					.9063	.6563	.4063	.2305	.1250	.0654
13					.9375	.7188	.4688	.2734	.1504	.0801
14					.9688	.7813	.5313	.3203	.1797	.0967
15					1.0000	.8438	.5938	.3711	.2129	.1162
16						.8906	.6563	.4219	.2480	.1377
17						.9219	.7109	.4727	.2852	.1611
18						.9531	.7656	.5273	.3262	.1875
19						.9688	.8125	.5781	.3672	.2158
20						.9844	.8516	.6289	.4102	.2461
21						1.0000	.8906	.6797	.4551	.2783
22							.9219	.7266	.5000	.3125
23							.9453	.7695	.5449	.3477
24							.9609	.8086	.5898	.3848
25							.9766	.8438	.6328	.4229
26							.9844	.8750	.6738	.4609
27							.9922	.9023	.7148	.5000
28							1.0000	.9258	.7520	.5391
29								.9453	.7871	.5771
30								.9609	.8203	.6152
31								.9727	.8496	.6523
32								.9805	.8750	.6875
33								.9883	.8984	.7217
34								.9922	.9180	.7539
35								.9961	.9355	.7842
36								1.0000	.9512	.8125
37									.9629	.8389
38									.9727	.8623
39									.9805	.8838
40									.9863	.9033
41									.9902	.9199
42									.9941	.9346
43									.9961	.9473
44									.9980	.9580

טבלה 4. התפלגות הסטטיסטי של ווילקוקסון למדגם מזווג: $P(V^+ \le k)$ (המשך)

	n				
k	11	12	13	14	15
0	.0005	.0002	.0001	.0001	.0000
1	.0010	.0005	.0002	.0001	.0001
2	.0015	.0007	.0004	.0002	.0001
3	.0024	.0012	.0006	.0003	.0002
4	.0034	.0017	.0009	.0004	.0002
5	.0049	.0024	.0012	.0006	.0003
6	.0068	.0034	.0017	.0009	.0004
7	.0093	.0046	.0023	.0012	.0006
8	.0122	.0061	.0031	.0015	.0008
9	.0161	.0081	.0040	.0020	.0010
10	.0210	.0105	.0052	.0026	.0013
11	.0269	.0134	.0067	.0034	.0017
12	.0337	.0171	.0085	.0043	.0021
13	.0415	.0212	.0107	.0054	.0027
14	.0508	.0261	.0133	.0067	.0034
15	.0615	.0320	.0164	.0083	.0042
16	.0737	.0386	.0199	.0101	.0051
17	.0874	.0461	.0239	.0123	.0062
18	.1030	.0549	.0287	.0148	.0075
19	.1201	.0647	.0341	.0176	.0090
20	.1392	.0757	.0402	.0209	.0108
21	.1602	.0881	.0471	.0247	.0128
22	.1826	.1018	.0549	.0290	.0151
23	.2065	.1167	.0636	.0338	.0177
24	.2324	.1331	.0732	.0392	.0206
25	.2598	.1506	.0839	.0453	.0240
26	.2886	.1697	.0955	.0520	.0277
27	.3188	.1902	.1082	.0594	.0319
28	.3501	.2119	.1219	.0676	.0365
29	.3823	.2349	.1367	.0765	.0416
30	.4155	.2593	.1527	.0863	.0473
31	.4492	.2847	.1698	.0969	.0535
32	.4829	.3110	.1879	.1083	.0603
33	.5171	.3386	.2072	.1206	.0677
34	.5508	.3667	.2274	.1338	.0757
35	.5845	.3955	.2487	.1479	.0844
36	.6177	.4250	.2709	.1629	.0938
37	.6499	.4548	.2939	.1788	.1039
38	.6812	.4849	.3177	.1955	.1147
39	.7114	.5151	.3424	.2131	.1262
40	.7402	.5452	.3677	.2316	.1384
41	.7676	.5750	.3934	.2508	.1514
42	.7935	.6045	.4197	.2708	.1651
43	.8174	.6333	.4463	.2915	.1796
44	.8398	.6614	.4730	.3129	.1947
45	.8608	.6890	.5000	.3349	.2106
46	.8799	.7153	.5270	.3574	.2271
47	.8970	.7407	.5537	.3804	.2444
48	.9126	.7651	.5803	.4039	.2622
49	.9263	.7881	.6066	.4276	.2807
50	.9385	.8098	.6323	.4516	.2997
51	.9492	.8303	.6576	.4758	.3193
52	.9585	.8494	.6823	.5000	.3394
53	.9663	.8669	.7061	.5242	.3599
54	.9731	.8833	.7291	.5484	.3808
55	.9790	.8982	.7513	.5724	.4020
56	.9839	.9119	.7726	.5961	.4235
57	.9878	.9243	.7928	.6196	.4452
58	.9907	.9353	.8121	.6426	.4670
59	.9932	.9451	.8302	.6651	.4890
60	.9951	.9539	.8473	.6871	.5110
61	.9966	.9614	.8633	.7085	.5330
62	.9976	.9680	.8781	.7292	.5548
63	.9985	.9739	.8918	.7492	.5765
64	.9990	.9788	.9045	.7684	.5980
65	.9995	.9829	.9161	.7869	.6192
66	1.0000	.9866	.9268	.8045	.6401
67		.9895	.9364	.8212	.6606
68		.9919	.9451	.8371	.6807
69		.9939	.9529	.8521	.7003
70		.9954	.9598	.8662	.7193
71		.9966	.9659	.8794	.7378
72		.9976	.9713	.8917	.7556
73		.9983	.9761	.9031	.7729
74		.9988	.9801	.9137	.7894
75		.9993	.9836	.9235	.8053
76		.9995	.9867	.9324	.8204
77		.9998	.9893	.9406	.8349
78		1.0000	.9915	.9480	.8486
79			.9933	.9547	.8616
80			.9948	.9608	.8738
81			.9960	.9662	.8853
82			.9969	.9710	.8961
83			.9977	.9753	.9062
84			.9983	.9791	.9156
85			.9988	.9824	.9243
86			.9991	.9852	.9323
87			.9994	.9877	.9397

טבלה 5. התפלגות חי־בריבוע: ערכי החלוקה χ^2_p



	p												
ד"ח	.005	.01	.025	.05	.10	.25	.50	.75	.90	.95	.975	.99	.995
1	0.0^439	0.0^316	0.0^398	0.0^239	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7
35	17.2	18.5	20.6	22.5	24.8	29.1	34.3	40.2	46.1	49.8	53.2	57.3	60.3
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	79.5
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	92.0
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	107.6	113.1	118.1	124.1	128.3
120	83.9	86.9	91.6	95.7	100.6	109.2	119.3	130.1	140.2	146.6	152.2	159.0	163.6

טבלה 6. ההתפלגות המקורבת של הסטטיסטי של קולמוגורוב-סמירנוב:

$$L(t) = \lim_{m,n\to\infty} P\left\{\sqrt{\frac{mn}{m+n}} \cdot D_{m,n} \geq t\right\}$$

t	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.3	1.0000	1.0000	1.0000	.9999	.9998	.9997	.9995	.9992	.9987	.9981
0.4	.9972	.9960	.9945	.9926	.9903	.9874	.9840	.9800	.9753	.9700
0.5	.9639	.9572	.9497	.9415	.9325	.9228	.9124	.9013	.8896	.8772
0.6	.8643	.8508	.8367	.8222	.8073	.7920	.7764	.7604	.7442	.7278
0.7	.7112	.6945	.6777	.6609	.6440	.6272	.6104	.5936	.5770	.5605
0.8	.5441	.5280	.5120	.4962	.4806	.4653	.4503	.4355	.4209	.4067
0.9	.3927	.3791	.3657	.3527	.3399	.3275	.3154	.3036	.2921	.2809
1.0	.2700	.2594	.2492	.2392	.2296	.2202	.2111	.2024	.1939	.1857
1.1	.1777	.1700	.1626	.1555	.1486	.1420	.1356	.1294	.1235	.1177
1.2	.1122	.1070	.1019	.0970	.0924	.0879	.0836	.0794	.0755	.0717
1.3	.0681	.0646	.0613	.0582	.0551	.0522	.0495	.0469	.0443	.0420
1.4	.0397	.0375	.0354	.0335	.0316	.0298	.0282	.0266	.0250	.0236
1.5	.0222	.0209	.0197	.0185	.0174	.0164	.0154	.0145	.0136	.0127
1.6	.0120	.0112	.0105	.0098	.0092	.0086	.0081	.0076	.0071	.0066
1.7	.0062	.0058	.0054	.0050	.0047	.0044	.0041	.0038	.0035	.0033
1.8	.0031	.0029	.0027	.0025	.0023	.0021	.0020	.0018	.0017	.0016
1.9	.0015	.0014	.0013	.0012	.0011	.0010	.0009	.0009	.0008	.0007
2.0	.0007	.0006	.0006	.0005	.0005	.0004	.0004	.0004	.0003	.0003
2.1	.0003	.0003	.0002	.0002	.0002	.0002	.0002	.0002	.0001	.0001
2.2	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
2.3	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000