

שיטות מתקדמות בביוסטטיסטיקה פרופ' דוד שטיינברג

משך הבחינה: 3 שעות.
מותר להשתמש בכל חומר עזר כתוב ובמחשבון.
לענות על כל השאלות.

הרצאה

תמצית של טבלת ערכים מן ההתפלגות הנורמלית סטנדרטית.

0.99	0.975	0.95	0.90	שטח משמאל
2.326	1.96	1.645	1.28	אחוזון

שאלה 1 (25)

נערך מחקר חתך לבדוק את הקשר בין BMI (מדד סטנדרטי להשמנה, המוגדר כמשקל (בק"ג) מחולק בריבוע הגובה (במטרים)) לבין שעות צפייה בטלוויזיה בקרב ילדים בגילאים 10 עד 18. היקף הצפייה בטלוויזיה נרשם על פי ראיון עם הילד במונחים של ממוצע שעות לשבוע, במשתנה בשם TV.

- 1.1 תחילה התאימו החוקרים מודל רגרסיה $BMI_i = \beta_0 + \beta_1 TV_i + \varepsilon_i$. איזה מקדם במודל הרגרסיה מבטא את הקשר (במידה ויש) בין שעות הצפייה בטלוויזיה והשמנה?
- 1.2 החוקרים מעוניינים לבדוק את ההשערה שאין קשר בין שעות צפייה לבין BMI. נסח השערות סטטיסטיות מתאימות במסגרת המודל הסטטיסטי של סעיף 1.1.
- 1.3 מהן ההנחות הסטטיסטיות הסטנדרטיות למודל הרגרסיה של סעיף 1.1? ציין בקיצור איך אפשר לבדוק אם ההנחות סבירות.
- 1.4 ידוע ש-BMI משתנה עם הגיל, דבר שלא נלקח בחשבון בסעיף 1.1. החוקרים רוצים "לנטרל" את ההשפעה של גיל על BMI בבדיקה של הקשר שלו עם שעות צפייה. (לצורך השאלה, ניתן להניח ש-BMI טיפוסי עולה באופן לניארי עם גיל). הצע מודל סטטיסטי מתאים לכך. הסבר בכמה מילים מה מייצג כל מקדם במודל.

שאלה 2 (25)

מחקר מקרה-ביקורת בדק האם קיים קשר בין צריכת אלכוהול לבין סרטן הוושט. המחקר כלל 200 גברים מברטניה (בצרפת) עם אבחון לסרטן הוושט ו- קבוצת ביקורת בת 778 גברים מאותו אזור שנבחרו באקראי מרשימות בוחרים. אספו מגוון של נתונים על כל הגברים, כולל מידע על הרגלים בצריכת אלכוהול, דיאטה, עישון, גיל ועוד. נתמקד בשאלה זו על תת-קבוצה של הנתונים לגברים בגילאים 35 עד 74 (מתחת לגיל 35 היה רק חולה 1 ומעל גיל 74 היו מעט נבדקים בשתי הקבוצות).

הטבלאות הבאות מציגות את מספר החולים ומספר הגברים בקבוצת הביקורת, לפי 4 קבוצות גיל ולפי 2 קבוצות של צריכת אלכוהול: עד 79 גרם ליום מול 80 גרם ליום או יותר.

גברים בגילאים 35 עד 44:

צריכת אלכוהול		
80 או יותר	עד 79	
4	5	מקרה
26	164	ביקורת

גברים בגילאים 45 עד 54:

צריכת אלכוהול		
80 או יותר	עד 79	
25	21	מקרה
29	138	ביקורת

גברים בגילאים 55 עד 64:

צריכת אלכוהול		
80 או יותר	עד 79	
42	34	מקרה
27	139	ביקורת

גברים בגילאים 65 עד 74:

צריכת אלכוהול		
80 או יותר	עד 79	
19	36	מקרה
18	88	ביקורת

2.1 הגדר את יחס הסיכויים (Odds Ratio) לחלות בסרטן הוושט, לגברים השותים הרבה אלכוהול (80 גרם ליום או יותר) לעומת גברים ששותים מעט אלכוהול (עד 79 גרם ליום), לקבוצת הגיל 35 עד 44.

2.2 חשב אמדנים ליחס הסיכויים בכל קבוצת גיל בנפרד.

2.3 חשב רווח בר סמך ליחס הסיכויים בקבוצת הגיל 35 עד 44.

2.4 בהמשך פלט מניתוח של Mantel-Haenszel. מה ההשערה הסטטיסטית הנבדקת על ידי שיטה זו?

2.5 על פי הפלט,

- (א) האם יש עדות לכך שיחס הסיכויים הוא תלוי בגיל.
 (ב) האם ניתן לדחות את ההשערה העיקרית הנבדקת על ידי השיטה?
 (ג) ציין רווח בר סמך (ברמת סמך 95%) ליחס הסיכויים לחלות בסרטן הוושט, לגברים השותים הרבה אלכוהול (80 גרם ליום או יותר) לעומת גברים ששותים מעט אלכוהול (עד 79 גרם ליום).

2.6 אחד החוקרים טוען שאין כל חשש מן הפרדוקס של Simpson כאן, כיוון שיחס הסיכויים די דומה בכל אחת מקבוצות הגיל. האם הוא צודק? נמק את תשובתך בשניים או שלושה משפטים.

Tests for Homogeneity of the Odds Ratio

Statistics		Chi-Squared	df	Asymp. Sig. (2-sided)
Conditional Independence	Cochran's	72.950	1	.000
	Mantel-Haenszel	70.905	1	.000
Homogeneity	Breslow-Day	3.682	3	.298
	Tarone's	3.675	3	.299

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0.

Mantel-Haenszel Common Odds Ratio Estimate

Estimate			4.766
ln(Estimate)			1.562
Std. Error of ln(Estimate)			.192
Asymp. Sig. (2-sided)			.000
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	3.270
		Upper Bound	6.947
	ln(Common Odds Ratio)	Lower Bound	1.185
		Upper Bound	1.938

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1.000 assumption. So is the natural log of the estimate.

שאלה 3 (25)

גם שאלה זו מתייחסת למחקר על הקשר בין צריכת אלכוהול לבין סרטן הוושט.

3.1 הצע מודל של רגרסיה לוגיסטית שיבצע ניתוח מקביל לזה של הניתוח של Mantel-Haenszel. הקפד להגדיר את כל המשתנים שימשו אותך במודל.

(א) ציין איך תסיק מן המודל האם יחס הסיכויים הומוגני.

(ב) ציין איך תסיקו האם יחס הסיכויים שווה ל-1.

3.2 בניתוח של שאלה 2 ושל הסעיף הקודם, התייחסת לגיל כמשתנה קבוצתי, בעל 4 רמות. למעשה, גיל הוא משתנה רציף ולא קבוצתי. בעזרת רגרסיה לוגיסטית, ניתן לבדוק את מידת הקשר בין צריכת אלכוהול לבין סרטן הוושט, תוך ניטרול ההשפעה של גיל, כאשר גיל הוא משתנה רציף. הגדר מודל מתאים. הקפד להגדיר את כל המשתנים שישתתפו במודל.

(א) הסבר איך תסיק האם יחס הסיכויים נשאר קבוע לאורך כל הגילאים.

(ב) הסבר איך תסיק האם יחס הסיכויים שווה ל-1.

שאלה 4 (25)

נערך מחקר המשווה בין שתי שיטות להכנסת קטטר אצל חולי סוכרת – הכנסה כירורגית מול הכנסה דרך העור (percutaneous). המטרה העיקרית של המחקר הייתה לבדוק האם יש הבדל בין השיטות מבחינת התפתחות של זיהומים אצל החולים. לכן, לכל חולה, נרשם משך הזמן (בחודשים) עד להתפתחות זיהום. במקרה שלא התפתח זיהום, החולה נחשב כ"מצונזר" ונרשם משך הזמן מהכנסת הקטטר עד ליום סגירת המחקר.

להלן רשימת הזמנים שנרשמו, מחולקים לארבע קבוצות, לפי סוג הטיפול והאם התפתח זיהום. היו 76 חולים שקיבלו את הקטטר דרך העור ו- 43 חולים בקבוצה הכירורגית. שימו לב שלקבוצה שקיבלה את הקטטר דרך העור, יש שתי עמודות בטבלה לאלה שלא פיתחו זיהום.

Surgical Insertion Infection		Percutaneous Insertion Infection		
Yes	No	Yes	No	No
1.5	2.5	0.5	0.5	6.5
3.5	2.5	0.5	0.5	7.5
4.5	3.5	0.5	0.5	7.5
4.5	3.5	0.5	0.5	7.5
5.5	3.5	0.5	0.5	8.5
8.5	4.5	0.5	0.5	8.5
8.5	5.5	2.5	0.5	8.5
9.5	6.5	2.5	0.5	9.5
10.5	6.5	3.5	0.5	9.5
11.5	7.5	6.5	0.5	10.5
15.5	7.5	15.5	1.5	10.5
16.5	7.5		1.5	10.5
18.5	7.5		1.5	11.5
23.5	8.5		1.5	11.5
26.5	9.5		2.5	12.5
	10.5		2.5	12.5
	11.5		2.5	12.5
	12.5		2.5	12.5
	12.5		2.5	14.5
	13.5		3.5	14.5
	14.5		3.5	16.5
	14.5		3.5	16.5
	21.5		3.5	18.5
	21.5		3.5	19.5
	22.5		4.5	19.5
	22.5		4.5	19.5
	25.5		4.5	20.5
	27.5		5.5	22.5
			5.5	24.5
			5.5	25.5
			5.5	26.5
			5.5	26.5
				28.5

להלן השורות הראשונות של טבלת האירועים המתאימה לאומד של Kaplan-Meier לקבוצה שקיבלה את הקטטר דרך העור.

Time	Status	Cumulative Survival	Standard Error	Cumulative Events	Number Remaining
.50	1			1	75
.50	1			2	74
.50	1			3	73
.50	1			4	72
.50	1			5	71
.50	1	.9211	.0309	6	70
.50	0			6	69
.50	0			6	68
.50	0			6	67
.50	0			6	66
.50	0			6	65
.50	0			6	64
.50	0			6	63
.50	0			6	62
.50	0			6	61
.50	0			6	60
1.50	0			6	59
1.50	0			6	58
1.50	0			6	57
1.50	0			6	56
2.50	1			7	55
2.50	1	.8882	.0376	8	54
2.50	0			8	53
2.50	0			8	52
2.50	0			8	51
2.50	0			8	50
2.50	0			8	49
3.50	1	.8700	.0409	9	48
3.50	0			9	47
3.50	0			9	46
3.50	0			9	45
3.50	0			9	44
3.50	0			9	43
4.50	0			9	42
4.50	0			9	41
4.50	0			9	40
5.50	0			9	39
5.50	0			9	38
5.50	0			9	37
5.50	0			9	36
5.50	0			9	35
6.50	1	.8452	.0467	10	34
6.50	0			10	33
7.50	0			10	32
7.50	0			10	31
7.50	0			10	30

- 4.1 הסבר איך חושב המספר 0.888 בתור האמדן לסיכוי לא לפתח זיהום במשך חודשיים וחצי לאחר הכנסת קטטר דרך העור.
- 4.2 בקבוצה של "דרך העור", חשב רווח בר סמך ברמת סמך 95% לסיכוי להיות בלי זיהום 7 חודשים לאחר הכנסת הקטטר.
- 4.3 בתחית העמוד מופיע גרף עם האמדנים של Kaplan-Meier לשתי הקבוצות. פרש בכמה מילים מה את(ה) רואה בגרף.
- 4.4 החוקרים ביצעו את מבחן log rank כדי להשוות בין שני הטיפולים. מהן ההשערות הסטטיסטיות הנבדקות במבחן זה?
- 4.5 מבחן log rank מתבסס על סדרה של טבלאות 2 על 2. בנה בצורה מפורשת את הטבלה הראשונה בסדרה זו והסבר בכמה מילים מה היה צפוי בטבלה אילו השערת האפס (מן הסעיף הקודם) נכונה.
- 4.6 להלן פלט מהשוואת השיטות להכנסת קטטר. האם יש הבדל מובהק בין השיטות מבחינת משך הזמן עד להתפתחות זיהום?

Survival Analysis for TIME

		Total	Number Events	Number Censored	Percent Censored
METHOD	1	43	15	28	65.12
METHOD	2	76	11	65	85.53
Overall		119	26	93	78.15

Test Statistics for Equality of Survival Distributions for METHOD

	Statistic	df	Significance
Log Rank	2.53	1	.1117

