

שיטות מתקדמות בביוסטטיסטיקה פרופ' דוד שטיינברג

משך הבחינה: 3 שעות.
מותר להשתמש בכל חומר עזר כתוב ובמחשבון.
נא לענות על כל השאלות.
בבקשה לכתוב בכתב יד גדול וברור.

הצהרה

בפניכם 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 (20)

שרייר, אסטציו, אסלר ומהלר דיווחו על ניסוי קליני שבדק השפעות של שליטה אגרסיבית של לחץ דם בקרב 480 חולי סוכרת. החולים הוקצו באקראי לשתי קבוצות טיפול: הורדה משמעותית של לחץ דם או הורדה מתונה של לחץ דם. כל חולה קיבל טיפול תרופתי הולם לקבוצה.

המדד העיקרי להצלחת הטיפול היה creatinine clearance (CC) והוא נמדד כל 6 חודשים במשך תקופת מעקב, שנמשכה בממוצע מעט יותר מ-5 שנים. הניתוח הסטטיסטי של מדד זה התבסס על שימוש של המודל הלינארי המעורב. המודל כלל רגרסיה כנגד זמן, השוואה בין 2 קבוצות הטיפול והשוואה בין 3 קבוצות של חולים שנוצרו בהתאם לרמת ה-albumin בשתן. המחברים לא מציינים אילו גורמים אקראיים נכללו במודל שלהם.

1.1 עליכם לרשום את המודל ב- Level 1 המתאר את רמת ה-CC בדם לנבדק בודד, תוך ציפיה לקשר לינארי בין CC לבין הזמן. להציע מודל לגורמי השגיאה במודל שלך ולנמק במשפט או שניים למה מודל זה סביר. יש להקפיד להגדיר כל סימון שאתם כותבים.

1.2 עבור אחד הנבדקים, ביקור המעקב הראשון היה 8 חודשים לאחר תחילת הטיפול ולא 6 חודשים כפי שתוכנן. האם המודל שרשמת בסעיף הקודם ייקח בחשבון סטייה זו מלוח הזמנים המתוכנן? האם ניתוח לפי "תצפיות חוזרות" (repeated measures) יטפל באותו אופן בסטייה מלוח הזמנים?

1.3 לרשום את המודל ב- Level 2 הקושר את התוצאות של הנבדק למשתנים ברמת הנבדק – קבוצת טיפול וקבוצת albumin. שוב להקפיד להגדיר כל סימון. להניח כאן שהמודל כולל חותך אקראי ושיפוע אקראי לכל נבדק.

1.4 החוקרים מצאו שלא היה שוני בין קבוצות הטיפול מבחינת השוני ב-CC. הם כן מצאו שוני בין קבוצות ה-albumin. להסביר, במונחים של המודל שרשמת, מה זה אומר לגבי המקדמים. (למשל, הם מדווחים על התוצאות הבאות לחולים עם רמה גבוהה של albumin: רמת ה-CC של מטופלים בקבוצה האגרסיבית ירדה במוצע מ-81.8 בתחילת המחקר ל-51.5 כעבור 5 שנים, ורמת ה-CC של מטופלים בטיפול המתון ירדה במוצע מ-76.8 ל-56.6.)

שאלה 2 (32)

מחקר MONICA בדק גורמי סיכון פוטנציאליים למחלות לב (CVD). אחד הגורמים שנבדק היה פקטור IX, שהוא אחד הגורמים המעורבים בקרישת הדם. במחקר מקרה-ביקורת, נבדקים הסובלים מ-CVD הושוו למדגם מן האוכלוסייה של אנשים ללא CVD. מדגם הבריאים הותאם לחולים על פי שכבות של גיל ומין, ולכן כולל מספרים דומים של נבדקים בכל קבוצה לפי גיל ומין. רשמו לכל נבדק האם הוא/היא מתחת לחציון (לפי מין) של פקטור IX (מקודד 0) או מעל החציון (מקודד 1). באופן דומה, $1=CVD$ מתאים לנבדקים הסובלים מ-CVD. קבוצות 1 עד 5 מתאימות לגברים, לפי 5 קבוצות גיל, מ-25 עד 34 (קבוצה 1) עד ל-65 עד 74 (קבוצה 5). קבוצות 6 עד 10 הן קבוצות הגיל של הנשים, בהתאמה. טבלה של התוצאות מופיעה בעמוד הבא. להלן פלט מהפעלת המבחן של מנטל והנסזל.

Tests of Homogeneity of the Odds Ratio

	Chi-Squared	df	Asymp. Sig. (2-sided)
Breslow-Day	8.736	9	.462
Tarone's	8.736	9	.462

Tests of Conditional Independence

	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	.270	1	.603
Mantel-Haenszel	.211	1	.646

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			1.062
ln(Estimate)			.060
Std. Error of ln(Estimate)			.117
Asymp. Sig. (2-sided)			.606
Asymp. 95% Confidence Interval	Common Odds Ratio	Lower Bound	.844
		Upper Bound	1.337
	ln(Common Odds Ratio)	Lower Bound	-.169
		Upper Bound	.290

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.

FacIX * CVD * Group Crosstabulation

Count

Group	FacIX	0	CVD		Total
			0	1	
1	FacIX	0	75	6	81
		1	20	4	24
	Total		95	10	105
2	FacIX	0	56	12	68
		1	56	15	71
	Total		112	27	139
3	FacIX	0	40	24	64
		1	52	29	81
	Total		92	53	145
4	FacIX	0	39	35	74
		1	54	36	90
	Total		93	71	164
5	FacIX	0	33	34	67
		1	52	38	90
	Total		85	72	157
6	FacIX	0	69	18	87
		1	32	11	43
	Total		101	29	130
7	FacIX	0	87	21	108
		1	34	11	45
	Total		121	32	153
8	FacIX	0	62	21	83
		1	58	36	94
	Total		120	57	177
9	FacIX	0	37	19	56
		1	59	36	95
	Total		96	55	151
10	FacIX	0	24	27	51
		1	62	57	119
	Total		86	84	170

סיכום לפי CVD ופקטור IX בלבד נותן את הטבלה הבאה:

FacIX * CVD Crosstabulation

Count

		CVD		Total
		0	1	
FacIX	0	522	217	739
	1	479	273	752
Total		1001	490	1491

- 2.1 אם מתעלמים מגיל ומין, נא חשבו את יחס הסיכויים (odds ratio) בין פקטור IX מעל החציון לבין CVD, ולחשב עבורו רווח בר סמך ברמת סמך 95%.
- 2.2 לחשב יחס סיכויים (odds ratio) בין פקטור IX מעל החציון לבין CVD בקרב גברים בקבוצת הגיל 65 עד 74. (אין צורך בחישוב רווח סמך.)
- 2.3 עמית למחקר לא למד הרבה סטטיסטיקה. להסביר לו את התוצאות של מבחן מנטל-הנסזל, כולל הסבר מה בודק כל חלק של המבחן, מה חשיבותו, ומה המסקנות לגבי השאלה שעל הפרק כאן (האם יש קשר בין הרמה של פקטור IX לבין CVD). במהלך התשובה, לתת רווח בר סמך ברמת סמך 95% ליחס הסיכויים (odds ratio) בין פקטור IX מעל החציון לבין CVD מתוקן לגיל ומין.
- 2.4 האם נתונים אלה הם דוגמה לפרדוקס של סימפסון? נמק בשניים-שלושה משפטים.

שאלה 3 (18)

כתבה מה- 16.2.2010 הודיעה על ממצאים של שכחות חריגה של מחלת פרקינסון בקרב תושבי באקה אל גרבייה. הממצאים באו ממחקר על השכיחות של פרקינסון ביישובים ערביים בוואדי ערה. על פי המאמר, מצאו חוקרים שהשכיחות של פרקינסון ב באקה אל גרבייה הייתה 73.39 למאה אלף תושבים, לעומת 34.84 באום אל פאחם, 25.45 בכפר קרע ו- 18.45 בערה-ערעה.

- 3.1 להסביר בכמה מילים למה השוואה ישירה של השכיחות עלולה להוביל למסקנות לא נכונות.
- 3.2 להסביר איך להפעיל כאן את השיטה של "תקנון לא ישיר" כדי להשוות בין הישובים באמצעות SMR לפרקינסון.

שאלה 4 (30)

שאלה זו מציגה תוצאות ממחקר על גנים העשויים להיות קשורים לפרוגנוזה של חולים בסרטן ריאות. אחד הסימנים הברורים להידרדרות בקרב חולי סרטן ריאות הוא הופעה של גרורות במוח. לכן "משך הזמן" במחקר זה הוגדר לפי הזמן (בחודשים) מאבחון ועד מציאת גרורות במוח. חולים עם גרורות במוח בעת האבחון לא נכללו כלל במחקר. לכל חולה, החוקרים מדדו רמות חומר בדם הקשורות לאוסף של גנים. הרמות בדם משקפות את ההרכב של אותם גנים. היו ניתוחים שהתייחסו לרמות החומרים הגנטיים כמשתנים רציפים, והיו ניתוחים אחרים בהם החוקרים ייצרו קבוצות תוך חלוקת החולים על הציר של כל גן (למשל לחולים עם רמת חומר נמוכה, בינונית או גבוהה).

להלן פלטים רלוונטים:

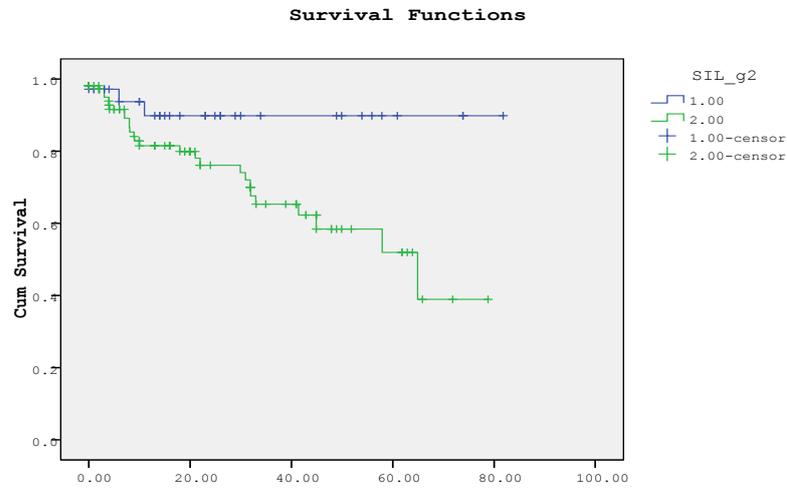
מבחן מובהקות המתאים להשוואה לפי הגן SIL, רמה נמוכה מול רמה גבוהה:

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	4.804	1	.028

Test of equality of survival distributions for the different levels of SIL_g2.

תרשים קפלן-מייר לפי הגן SIL. קבוצה אחת כוללת נבדקים עם רמות גבוהות של הגן וקבוצה שתיים כוללת חולים עם רמות נמוכות.



תחילת טבלת ההישרדות בקבוצה 2 :

Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases	
		Estimate	Std. Error			
1	2	1	0.989	0.01	1	94
2	2	0	.	.	1	93
3	2	0	.	.	1	92
4	2	0	.	.	1	91
5	2	0	.	.	1	90
6	2	0	.	.	1	89
7	3.016	1	.	.	2	88
8	3.016	1	0.967	0.019	3	87
9	3.934	1	0.956	0.021	4	86
10	3.967	0	.	.	4	85
11	4	1	0.945	0.024	5	84
12	4	0	.	.	5	83
13	4	0	.	.	5	82
14	4	0	.	.	5	81
15	4.033	1	0.933	0.026	6	80
16	4.033	0	.	.	6	79
17	4.918	0	.	.	6	78
18	4.984	0	.	.	6	77
19	6.033	0	.	.	6	76
20	6.033	0	.	.	6	75
21	6.033	0	.	.	6	74
22	6.951	0	.	.	6	73
23	7.016	1	0.92	0.029	7	72
24	7.049	1	0.908	0.031	8	71
25	7.967	1	.	.	9	70

26	7.967	1	0.882	0.035	10	69
27	8.033	1	0.869	0.037	11	68
28	8.951	1	0.857	0.039	12	67
29	8.951	0	.	.	12	66
30	9.016	1	0.844	0.04	13	65
31	9.934	0	.	.	13	64
32	9.934	0	.	.	13	63
33	9.967	1	0.83	0.042	14	62
34	9.967	0	.	.	14	61
35	12.984	0	.	.	14	60

תוצאות ממודל קוקס שמוסיף משתנים באופן סדרתי (לפי יחס הנראות). המשתנה NEUR_score הוא ציון למצב נוירולוגי של החולה. שני המשתנים האחרים הם רמות של גנים (משתנים רציפים, לא קבוצתיים), לאחר טרנספורמציה לסקאלה לוגריתמית.

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Step 1 L_FALZ	-.636	.218	8.504	1	.004	.529	.345	.812
Step 2 NEUR_score	.390	.139	7.874	1	.005	1.476	1.125	1.938
L_FALZ	-.831	.234	12.640	1	.000	.436	.275	.689
Step 3 NEUR_score	.362	.136	7.097	1	.008	1.437	1.100	1.875
L_FALZ	-.815	.236	11.948	1	.001	.442	.279	.703
L_KIFC1	1.604	.679	5.577	1	.018	4.971	1.314	18.811

4.1 האם החלוקה של החולים לפי הגן SIL מובילה לקבוצות השונות זו מזו באופן מובהק מבחינת משך הזמן עד לגילוי גרורות במוח?

4.2 לתאר על סמך הגרף של עקומת קפלן-מייר מהם זמני ההישרדות של קבוצה 1 (העקומה העליונה בתרשים).

4.3 לקבוצה 2 לפי SIL, לתת אמדן להסתברות להישאר נקי מגרורות במוח 12 חודשים לאחר האבחון. לציין גם רווח בר סמך ברמת סמך 95% להסתברות זו.

4.4 לפרש את התוצאות של המודל של קוקס. לציין מהם המנבאים של "פרוגנוזה טובה" ושל "פרוגנוזה לא טובה".