האוניברסיטה העברית בירושלים
המחלקה לסטטיסטיקה

מודלים סטטיסטיים ויישומיהם (52518)
פרופ' דוד צוקר
תשע"ז סמסטר א', מועד ב'

תאריך: י"ב בשבט תשע"ז, 26.3.17
משך הבחינה: שעתיים וחצי
חומר עזר: מחשב כיס ודפי רשימות (שני דפים בגודל A4, שני צדדים)

חל איסור מוחלט להעתיק. תלמיד שייתפס יורחק לשנה מלימודיו.

בהצלחה!!

שאלה 1 (10 נקודות)

נתייחס למודל

$$Y_{ij} = \mu_i + \varepsilon_{ij},\ \ \varepsilon_{ij} \sim N(0,\sigma^2) \text{ independent}$$

נגדיר

$$\psi(\mathbf{c}) = \sum_{i=1}^{N} c_i \mu_i$$

מצאו את ההתפלגות של $\psi(\mathbf{c})$. נמקו.

שאלה 2 (10 נקודות)

הסביר ב- 3-4 משפטים את המושגים הבאים. השתמשו בסימונים לפי הצורך.

[5] א. רווחי סמך בו-זמניים

[5] ב. שיטת Welch

שאלה 3 (20 נקודות)

נתייחס לנתונים עם שלושה משתנים אכותיים A,B,C, כל אחד עם שתי רמות. מה הפירוש של המודל (AB,AC) ? הוכיחו בצורה מדוקדקת.

שאלה 4 (20 נקודות)

להלן מובא פלט מהרצה של רגרסיה לוגיסטית. המשתנה המוסבר הינו משקל לידה נמוך (0=לא, 1=כן). המשתנים המסבירים הינם AGE=גיל האם, WEIGHT=משקל האם לפני ההיריון (קילו), ו-SMOKE=סטטוס עישון של האם (0=לא, 1=כן). חישבו אומד נקודתי ורווח סמך ברמה 95% ל-odds ratio למשקל לידה נמוך בין אם בגיל 24 ששקלה 50 קילו לפני ההיריון ולא מעשנת לבין אם בגיל 30 ששקלה 60 קילו לפני ההיריון ולא מעשנת.

הפלט:

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.4118	1.0151	1.9342	0.1643
age	1	-0.0408	0.0328	1.5442	0.2140
weight	1	-0.0268	0.0135	3.9456	0.0470
smoke	1	0.6981	0.3269	4.5596	0.0327

Estimated Covariance Matrix of Estimated Coefficient Vector

Parameter	Intercept	age	weight	smoke
Intercept	1.030523	-0.02084	-0.00896	-0.05179
age	-0.02084	0.001077	-0.00006	-0.00009
weight	-0.00896	-0.00006	0.000182	0.00011
smoke	-0.05179	-0.00009	0.00011	0.106895

שאלה 5 (30 נקודות)

נתייחס למודל של ניתוח שונות דו-כיווני :

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

$\varepsilon_{ijk} \sim N(0,\sigma^2)$ ב"ת, σ^2 לא ידוע

$$i=1,\dots,I;\ j=1,\dots,J;\ k=1,\dots,n_{ij}$$

נניח כי $I = J = 2$ ו- $n_{11}=2, n_{12}=4, n_{21}=4, n_{22}=8$. נגדיר את $\mu,\alpha_i,\beta_j,\gamma_{ij}$ לפי המשקולות

$$\pi_1 = 1/3,\ \pi_2 = 2/3,\ \tau_1 = 1/3,\ \tau_2 = 2/3$$

[15] א. רשמו את המודל בצורה $\mathbf{Y} = \mathbf{X}\boldsymbol{\eta} + \boldsymbol{\varepsilon}$ עם הגדרות מתאימות.

[15] ב. הראו כי

$$SSAB = \sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij}(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

באשר

$$SSAB = SSE(\mu,\alpha,\beta) - SSE(\mu,\alpha,\beta,\gamma)$$

$$\bar{Y}_{...} = \frac{1}{N}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n_{ij}} Y_{ijk}$$

$$\bar{Y}_{i..} = \frac{1}{n_{i.}}\sum_{j=1}^{2}\sum_{k=1}^{n_{ij}} Y_{ijk}$$

$$\bar{Y}_{.j.} = \frac{1}{n_{.j}}\sum_{i=1}^{2}\sum_{k=1}^{n_{ij}} Y_{ijk}$$

$$n_{i.} = n_{i1} + n_{i2},\ \ n_{.j} = n_{1j} + n_{2j},\ \ N = n_{11} + n_{12} + n_{21} + n_{22}$$

שאלה 6 (10 נקודות)

נתייחס למחקר באשר דוגמים מספר צמחים, דגומים 3 עלים מכל צמח, לוקחים 2 דגימות של 100 מ"ג מכל עלה ומודדים ריכוז הסידן בכל דגימה. נניח שנרצה לנתח את הנתונים באמצעות המודל

$$Y_{ijk} = \mu + a_i + b_{ij} + \epsilon_{ijk}$$
$$a_i \sim N(0,\sigma_a^2)$$
$$b_{ij} \sim N(0,\sigma_b^2)$$
$$\epsilon_{ijk} \sim N(0,\sigma_\epsilon^2)$$

עם אי-תלות בין כל המ"מ $a_i, b_{ij}, \epsilon_{ijk}$, באשר i מסמן את צמח, j מסמן את עלה בתוך הצמח, ו- k מסמן את הדגימה . נסמן את הווקטור של כל התצפיות מצמח i ב-

$$\mathbf{Y}_i = [Y_{i11}, Y_{i12}, Y_{i21}, Y_{i22}, Y_{i31}, Y_{i32}]^T$$

רשמו את ההתפלגות של $\mathbf{Y}_i$.