האוניברסיטה העברית בירושלים
המחלקה לסטטיסטיקה

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

תאריך: ה' בניסן תשע"ו, 13.4.16
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!!

פיתרון

שאלה 1 (25 נקודות)

נתייחס לנתונים דלהלן. נניח כי הנתונים מתנהגים לפי המודל

$$Y_{ij} = \mu_i + \epsilon_{ij}\ ;\quad \epsilon_{ij} \underset{iid}{\sim} N(0, \sigma^2)$$

חשבו סטטיסטי F לבדוק את השערת האפס הבאה:

$$H_0:\ \ \mu_3 - \mu_1 = 2(\mu_2 - \mu_1)\ \text{ and }\ \mu_4 - \mu_1 = 4(\mu_2 - \mu_1)$$

יש להגיע לתשובה **מספרית**.

בנוסף, ציינו (בצורה מלאה) את התפלגותו של הסטטיסטי תחת H_0.

רמז: תחת H_0 ניתן לבטא כל אחד מה- μ_i במונחים של שני פרמטרים $\beta_1^{(0)}, \beta_2^{(0)}$. למשל, אפשר לקחת $\beta_1^{(0)} = \mu_1$. אז, ניתן לרשום את המודל בצורה $Y = X^{(0)}\beta^{(0)} + \epsilon$, עבור מטריצה $X^{(0)}$ מתאימה. עליכם להגדיר בצורה נכונה את $X^{(0)}$ ו- $\beta^{(0)}$.

הנתונים:

GROUP	N	mean	SD
1	49	4.04	8.79
2	48	-0.46	7.83
3	48	-3.67	7.44
4	53	-8.89	9.86

פיתרון

סטטיסטי F מתאים לבדיקת H_0 יהיה,

$$F_0 = \frac{\left\|\hat{Y} - \hat{Y}^{(0)}\right\|^2/(d - d_0)}{\left\|Y - \hat{Y}\right\|^2/(N - d)} \sim \mathcal{F}_{(d-d_0, N-d)}$$

כאשר,

$$d = I = 4$$

$$N = \sum_{i=1}^{4} n_i = 198$$

$$\frac{\left\|Y - \hat{Y}\right\|^2}{N-d} = \frac{1}{N-d} \sum_{i=1}^{I} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_{i.}\right)^2 = s^2$$

$$\left\|\hat{Y} - \hat{Y}^{(0)}\right\|^2 = \sum_{i=1}^{I} n_i \left(\bar{Y}_{i.} - \hat{\mu}_i^{(0)}\right)^2$$

לפנינו מודל ניתוח שונות חד-כיווני. נזכור כי ניתן להציגו בצורה הבאה.

$$Y = X\beta + \epsilon$$

כאשר, $Y = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{In_I} \end{pmatrix}$, $\beta = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_I \end{pmatrix}$ ו- X הינה מטריצת תכנון מסדר גודל $(N \times I)$, אשר בכל אחת מעמודותיה תופיע הספרה 1 - לו התצפית משתייכת לקבוצה אותה מייצגת העמודה ו- 0 - אחרת.

במצב הנ״ל, אומדי הווקטור β יהיו ממוצעי הקבוצות. כלומר,

$$\hat{\beta} = \begin{pmatrix} \bar{Y}_{1.} \\ \vdots \\ \bar{Y}_{I.} \end{pmatrix}$$

נשים לב כי במודל תחת H_0, ניתן להביע את $\mu_1, \dots, \mu_4$ באמצעות μ_1 ו- μ_2 בלבד. על כן, נגדיר צמד המקדמים הבא.

$$\beta_1^{(0)} = \mu_1$$
$$\beta_2^{(0)} = \mu_2 - \mu_1$$

(זו אינה הדרך היחידה בה ניתן להגדיר את $\beta_1^{(0)}$ ו- $\beta_2^{(0)}$).

מהגדרה זו, ניווכח כי

$$\mu_1^{(0)} = \beta_1^{(0)}$$
$$\mu_2^{(0)} = \beta_1^{(0)} + \beta_2^{(0)}$$
$$\mu_3^{(0)} = \beta_1^{(0)} + 2\beta_2^{(0)}$$
$$\mu_4^{(0)} = \beta_1^{(0)} + 4\beta_2^{(0)}$$

בנוסף, אנו למדים כי $d_0 = 2$, היות וקיימים שני מקדמי רגרסיה.

כעת, נותר לנו לבנות את מטריצת התכנון תחת H_0, $X^{(0)}$.

$$X^{(0)}_{(2\times 198)} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & 2 \\ 1 & 4 \\ \vdots & \vdots \\ 1 & 4 \end{pmatrix}$$

(49 השורות הראשונות תקבלנה הערכים $(0 \quad 1)$, 48 השורות הבאות תקבלנה הווקטור $(1 \quad 1)$ וכן הלאה).

עתה, עלינו לאמוד הווקטור $\beta^{(0)}$. נעשה זאת בעזרת שיטת הריבועים הפחותים.

$$\hat{\beta}^{(0)} = \begin{pmatrix} \hat{\beta}_1^{(0)} \\ \hat{\beta}_2^{(0)} \end{pmatrix} = \left(X^{(0)^T} X^{(0)}\right)^{-1} X^{(0)^T} Y$$

לשם נוחות, נבצע החישוב בשלבים.

$$\left(X^{(0)^T} X^{(0)}\right)^{-1} = \begin{pmatrix} N & \sum_{i=2}^{4} (2)^{i-2} n_i \\ \sum_{i=2}^{4} (2)^{i-2} n_i & \sum_{i=2}^{4} (4)^{i-2} n_i \end{pmatrix}^{-1} = \begin{pmatrix} 198 & 356 \\ 356 & 1{,}088 \end{pmatrix}^{-1} = \frac{1}{88{,}688} \begin{pmatrix} 1{,}088 & -356 \\ -356 & 198 \end{pmatrix}$$

$$\approx \begin{pmatrix} 0.0123 & -0.004 \\ -0.004 & 0.0022 \end{pmatrix}$$

$$X^{(0)^T} Y = \begin{pmatrix} \sum_{i=1}^{4} \sum_{j=1}^{n_i} Y_{ij} \\ \sum_{i=2}^{4} (2)^{i-2} \sum_{j=1}^{n_i} Y_{ij} \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{4} n_i \bar{Y}_{i.} \\ \sum_{i=2}^{4} (2)^{i-2} n_i \bar{Y}_{i.} \end{pmatrix} = \begin{pmatrix} -471.45 \\ -2{,}259.08 \end{pmatrix}$$

ומכאן נקבל,

$$\hat{\beta}^{(0)} = \begin{pmatrix} 0.0123 & -0.004 \\ -0.004 & 0.0022 \end{pmatrix} \begin{pmatrix} -471.45 \\ -2{,}259.08 \end{pmatrix} = \begin{pmatrix} 3.2375 \\ -3.0842 \end{pmatrix}$$

כעת, נחשב הערכים החזויים תחת H_0, $\hat{Y}^{(0)}$.

$$\hat{Y}_{1,1}^{(0)} = \cdots = \hat{Y}_{1,49}^{(0)} = \hat{\mu}_1^{(0)} = 3.2375$$
$$\hat{Y}_{2,1}^{(0)} = \cdots = \hat{Y}_{2,48}^{(0)} = \hat{\mu}_2^{(0)} = 0.1533$$
$$\hat{Y}_{3,1}^{(0)} = \cdots = \hat{Y}_{3,48}^{(0)} = \hat{\mu}_3^{(0)} = -2.9309$$
$$\hat{Y}_{4,1}^{(0)} = \cdots = \hat{Y}_{4,53}^{(0)} = \hat{\mu}_4^{(0)} = -9.0993$$

לכן, מונה הסטטיסטי F_0 יהיה

$$\sum_{i=1}^{I} n_i \left(\bar{Y}_{i.} - \mu_i^{(0)}\right)^2 = 49(4.04 - 3.2375)^2 + \cdots + 53(-8.89 + 9.0993)^2 = 78.1535$$

הרכיב האחרון אשר נותר למצוא, בטרם נחשב הסטטיסטי המבוקש, הינו s^2.

$$s^2 = \frac{1}{N-d} \sum_{i=1}^{I} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_{i.}\right)^2 = \frac{1}{194} \sum_{i=1}^{I} (n_i - 1) SD_i^2 = 73.4394$$

כעת, נחשב ערכו של הסטטיסטי $F_0 \sim \mathcal{F}_{(2,194)}$.

$$F_0 = \frac{78.1535/2}{73.4394} = \boxed{0.5321}$$

שאלה 2 (10 נקודות)

להלן תצפיות ממדגם מקרי פשוט :

$$1.81, \quad -0.55, \quad 2.50, \quad 0.92, \quad 1.19, \quad -1.12, \quad -1.03, \quad 0.44$$

חשבו ושרטטו את פונקציית ההתפלגות האמפירית $\hat{F}_n$.

פיתרון

ראשית, נסדר המדגם בסדר עולה.

$-1.12, -1.03, -0.55, 0.44, 0.92, 1.19, 1.81, 2.5$

עתה, נחשב את פונקצית ההתפלגות האמפירית בעזרת הנוסחה הבאה.

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le t\}}$$

נתחיל בערכו המינימאלי של המדגם, $t = \min(X_i) = -1.12$.

$$\hat{F}_8(-1.12) = \frac{1}{8}\sum_{i=1}^{8} I_{\{X_i \le -1.12\}} = \frac{1}{8}$$

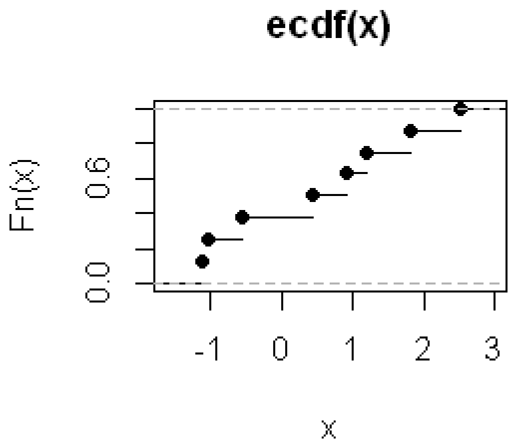
הערך הבא עבורו נחשב את פונקצית ההפלגות האמפירית יהיה -1.03.

$$\hat{F}_8(-1.03) = \frac{1}{8}\sum_{i=1}^{8} I_{\{X_i \le -1.03\}} = \frac{2}{8} = \frac{1}{4}$$

נמשיך באופן דומה, עד אשר נגיע לאיבר המכסימאלי במדגם, $t = \max(X_i) = 2.5$.

$$\hat{F}_8(2.5) = \frac{1}{8}\sum_{i=1}^{8} I_{\{X_i \le 2.5\}} = \frac{8}{8} = 1$$

פונקצית ההתפלגות תיראה כך.



שאלה 3 (25 נקודות)

נתייחס לנתונים הבאים (המספרים הינם תצפיות Y_{ijk}).

	Age ≤ 45	Age > 45
0 mg	12,13	11,12
50 mg	9,10	16,18
100 mg	15,16	17,16
150 mg	7,18	12,18

להלן לוח הממוצעים $\bar{Y}_{ij.}$:

	Age ≤ 45	Age > 45

0 mg	12.5	11.5
50 mg	9.5	17
100 mg	15.5	16.5
150 mg	12.5	15

בנוסף,

$$\sum_{i=1}^{4}\sum_{j=1}^{2}\sum_{k=1}^{2} Y_{ijk}^2 = 3206$$

ננתח את הנתונים האלה באמצעות ניתוח שונות דו-כיווני **ללא אינטראקציה**.

א. נסחו המודל בצורה $Y = X\eta + \epsilon$, עם ההגדרות המתאימות של X ו- η. וחשבו המטריצה $(X^TX)^{-1}$. השתמשו במשקולות $\pi_i = 1/I \quad \forall i$ ו- $\tau_j = 1/J \quad \forall j$. שימו לב, $(X^TX)^{-1}$ מטריצה סימטרית. [6]

<u>תזכורת</u> : נוסחת הפיכת מטריצה.

$$(A^{-1})_{ij} = (-1)^{i+j}\det(A_{ji})/\det(A)$$

כאשר, A_{ji} הינה המטריצה המתקבלת ממחיקת השורה j והעמודה i במטריצה A.

<u>פיתרון</u>

לפנינו המודל הבא,

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}\ ;\quad \epsilon_{ijk} \sim N(0,\sigma^2)$$

כאשר,

$$i = 1,\dots,4\ ;\ j = 1,2\ ;\ k = 1,2$$

אי לכך, הצגתו בצורה $Y = X\eta + \epsilon$ תעשה כך. $Y = \begin{pmatrix} Y_{111} \\ \vdots \\ Y_{IJn_{IJ}} \end{pmatrix}$, $\epsilon = \begin{pmatrix} \epsilon_{111} \\ \vdots \\ \epsilon_{IJn_{IJ}} \end{pmatrix}$ ו- $\eta =$ $(\mu \quad \alpha_1 \quad \alpha_2 \quad \alpha_3 \quad \beta_1)^T$. המטריצה X תהיה מסדר גודל $(N \times (I+J-1)) = (16 \times 5)$.

כל שורה במטריצה תכיל את מקדמי הווקטור η, אשר יסתכמו ל- μ_{ij} הרלוונטי ($\mu_{ij} = \mu +$ $\alpha_i + \beta_j$). לשם טיפול בפרמטרים α_4 ו- β_2, אשר אינם מופיעים בווקטור η, נעזר באילוצים הבאים.

$$\sum_{i=1}^{I}\pi_i\alpha_i = 0 \Rightarrow \frac{1}{I}\sum_{i=1}^{I}\alpha_i = 0 \Rightarrow \sum_{i=1}^{I}\alpha_i = 0$$
$$\sum_{j=1}^{J}\tau_j\beta_j = 0 \Rightarrow \frac{1}{J}\sum_{j=1}^{J}\beta_j = 0 \Rightarrow \sum_{j=1}^{J}\beta_j = 0$$

(היות ואנו מצויים במצב מאוזן, האילוצים לעיל ברורים מאליהם).

מהאילוצים לעיל נסיק,

$$\alpha_4 = -\alpha_1 - \alpha_2 - \alpha_3$$
$$\beta_2 = -\beta_1$$

להלן, המטריצה X.

$$X=\begin{pmatrix}1&1&0&0&1\\1&1&0&0&1\\1&1&0&0&-1\\1&1&0&0&-1\\1&0&1&0&1\\1&0&1&0&1\\1&0&1&0&-1\\1&0&1&0&-1\\1&0&0&1&1\\1&0&0&1&1\\1&0&0&1&-1\\1&0&0&1&-1\\1&-1&-1&-1&1\\1&-1&-1&-1&1\\1&-1&-1&-1&-1\\1&-1&-1&-1&-1\end{pmatrix}$$

כעת, נחשב המטריצה $(X^TX)^{-1}$.

$$(X^TX)^{-1}=\begin{pmatrix}16&0&0&0&0\\0&8&4&4&0\\0&4&8&4&0\\0&4&4&8&0\\0&0&0&0&16\end{pmatrix}^{-1}$$

היות ולפנינו מטריצת בלוקים אלכסונית, נגדיר האיברים הבאים.

$$A_1=16$$
$$A_2=\begin{pmatrix}8&4&4\\4&8&4\\4&4&8\end{pmatrix}$$
$$A_3=16$$

כלומר,

$$(X^TX)^{-1}=\begin{pmatrix}A_1&\cdots&0\\\vdots&A_2&\vdots\\0&\cdots&A_3\end{pmatrix}^{-1}=\begin{pmatrix}A_1^{-1}&\cdots&0\\\vdots&A_2^{-1}&\vdots\\0&\cdots&A_3^{-1}\end{pmatrix}$$

ניתן לראות בנקל כי $A_1^{-1}=A_3^{-1}=0.0625$. נותר לחשב את הבלוק האמצעי.

$$A_2^{-1}=\frac{\operatorname{adj}(A)}{\det(A)}=\frac{1}{256}\begin{pmatrix}\begin{vmatrix}8&4\\4&8\end{vmatrix}&-\begin{vmatrix}4&4\\4&8\end{vmatrix}&\begin{vmatrix}4&8\\4&4\end{vmatrix}\\-\begin{vmatrix}4&4\\4&8\end{vmatrix}&\begin{vmatrix}8&4\\4&8\end{vmatrix}&-\begin{vmatrix}8&4\\4&4\end{vmatrix}\\\begin{vmatrix}4&8\\4&4\end{vmatrix}&-\begin{vmatrix}8&4\\4&4\end{vmatrix}&\begin{vmatrix}8&4\\4&8\end{vmatrix}\end{pmatrix}^T$$
$$=\begin{pmatrix}0.1875&-0.0625&-0.0625\\-0.0625&0.1875&-0.0625\\-0.0625&-0.0625&0.1875\end{pmatrix}$$

ב. מצאו את האומדים של α_i, עבור $i=1,\dots,4$. [3]

<u>פיתרון</u>

היות ואנו במצב המאוזן, נשתמש בנוסחה הבאה.

$$\hat{\alpha}_i=\hat{\mu}_{i.}-\hat{\mu}=\bar{Y}_{i..}-\bar{Y}_{...}$$

ולהלן האומדים.

$$\hat{\alpha}_1=12-13.75=-1.75$$

$$\hat{\alpha}_2 = 13.25 - 13.75 = -0.5$$
$$\hat{\alpha}_3 = 16 - 13.75 = 2.25$$
$$\hat{\alpha}_4 = 13.75 - 13.75 = 0$$

ג. חשבו רווח סמך בו-זמני (ברמת ביטחון של 95%) לפי שיטת שאפה, עבור הקונטראסט הבא.

$$\psi(\alpha) = \frac{\alpha_3 + \alpha_4}{2} - \alpha_1$$

(הניחו כי הקונטראסט אינו נקבע מראש). [16]

<u>פיתרון</u>

ראשית, נזכור כי וקטור האומדים מתפלג כך,

$$\hat{\alpha} \sim N(\alpha, \sigma^2 C_{\hat{\alpha}})$$

כאשר, $C_{\hat{\alpha}}$ הינה תת-מטריצה של $(X^T X)^{-1}$, בה מצויות השורות והעמודות המתייחסות ל- $\hat{\alpha}$ (במקרה שלנו $C_{\hat{\alpha}} = A_2^{-1}$).

רווח סמך בו-זמני לפי שיטת שאפה יחושב באופן הבא.

$$\hat{\psi}(d) \pm \left[(I-1)F_{(I-1,N-(I+J-1))}^{(1-\omega)}\right]^{\frac{1}{2}} \cdot s\sqrt{d^T C_{\hat{\alpha}} d}$$

(בכדי למנוע בלבול, רמת המובהקות מוצגת באות ω).

נחשב כל חלק מהנוסחה בנפרד.

$$\psi(d) = \frac{\alpha_3 + \alpha_4}{2} - \alpha_1 = \frac{1}{2}(\alpha_3 - \alpha_1 - \alpha_2 - \alpha_3) - \alpha_1 = \begin{pmatrix} -\frac{3}{2} & -\frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = d^T \alpha$$

ולכן,

$$\hat{\psi}(d) = d^T \hat{\alpha} = \begin{pmatrix} -\frac{3}{2} & -\frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} -1.75 \\ -0.5 \\ 2.25 \end{pmatrix} = 2.875$$

היות ומדובר במודל נטול אינטראקציות, נחשב s^2 באופן הבא.

$$s^2 = \frac{1}{N-(I+J-1)} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{n} (Y_{ijk} - \hat{Y}_{ijk})^2$$
$$= \frac{1}{N-(I+J-1)} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{n} \left(Y_{ijk} - (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)\right)^2$$

כאשר,

$$\hat{\beta}_1 = \bar{Y}_{.1.} - \bar{Y}_{...} = 12.5 - 13.75 = -1.25$$
$$\hat{\beta}_2 = \bar{Y}_{.2.} - \bar{Y}_{...} = 15 - 13.75 = 1.25$$

($\hat{\mu}$ ו- $\hat{\alpha}_1, \dots, \hat{\alpha}_4$ חושבו בסעיף הקודם)

$$s^2 = \frac{1}{16-(4+2-1)} \sum_{i=1}^{4} \sum_{j=1}^{2} \sum_{k=1}^{2} \left(Y_{ijk} - (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)\right)^2$$
$$= \frac{1}{11}\left(\left(12 - (13.75 - 1.75 - 1.25)\right)^2 + \cdots + (18-15)^2\right) = 11.1364$$

לסיום, נחשב ערכו של $d^T C_{\hat{\alpha}} d$.

$$d^T C_{\hat{\alpha}} d = \begin{pmatrix} -\frac{3}{2} & -\frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} 0.1875 & -0.0625 & -0.0625 \\ -0.0625 & 0.1875 & -0.0625 \\ -0.0625 & -0.0625 & 0.1875 \end{pmatrix} \begin{pmatrix} -\frac{3}{2} \\ -\frac{1}{2} \\ 0 \end{pmatrix} = 0.375$$

רווח הסמך המתקבל הינו,

$$CI\big(\psi(d)\big) = 2.875 \pm 3.2806 \cdot \sqrt{11.1364} \cdot \sqrt{0.375} = \boxed{[-3.8291 \quad 9.5791]}$$

<u>שאלה 4 (25 נקודות)</u>

במסגרת של מודלים לוג-ליניאריים, נגדיר

$$\Psi^{k}_{ii'jj'} = \left(\frac{\Pr(B=j'|A=i',C=k)}{\Pr(B=j|A=i',C=k)}\right) \Big/ \left(\frac{\Pr(B=j'|A=i,C=k)}{\Pr(B=j|A=i,C=k)}\right)$$

הוכיחו כי המודל (AB,AC,BC) שקול למצב בו $\Psi^{k}_{ii'jj'}$ לא תלוי ב-k, ז״א מנת יחס הסיכויים של כל שני משתנים אינה תלויה בערכו של המשתנה השלישי.

<u>פיתרון</u>

לפנינו המודל הלוג-ליניארי הבא,

$$\log \pi_{ijk} = \theta_{ijk} = \bar{\theta}_{\dots} + \lambda^{A}_{i} + \lambda^{B}_{j} + \lambda^{C}_{k} + \lambda^{AB}_{ij} + \lambda^{AC}_{ik} + \lambda^{BC}_{jk}$$

נציג את π_{ijk} כך.

$$\pi_{ijk} = \Pr(A=i,B=j,C=k) = \omega\alpha_i\beta_j\gamma_k\delta_{ij}\epsilon_{ik}\phi_{jk}$$

כאשר,

$$\begin{aligned}
\omega &= \exp(\bar{\theta}_{\dots})\\
\alpha_i &= \exp\left(\lambda^{A}_{i}\right)\\
\beta_j &= \exp\left(\lambda^{B}_{j}\right)\\
\gamma_k &= \exp\left(\lambda^{C}_{k}\right)\\
\delta_{ij} &= \exp\left(\lambda^{AB}_{ij}\right)\\
\epsilon_{ik} &= \exp\left(\lambda^{AC}_{ik}\right)\\
\phi_{jk} &= \exp\left(\lambda^{BC}_{jk}\right)
\end{aligned}$$

נפשט הביטוי $\Psi^{k}_{ii',jj'}$.

$$\begin{aligned}
\Psi^{k}_{ii',jj'} &= \left(\frac{\Pr(B=j'|A=i',C=k)}{\Pr(B=j|A=i',C=k)}\right) \Big/ \left(\frac{\Pr(B=j'|A=i,C=k)}{\Pr(B=j|A=i,C=k)}\right)\\
&= \left(\frac{\frac{\Pr(A=i',B=j',C=k)}{\Pr(A=i',C=k)}}{\frac{\Pr(A=i',B=j,C=k)}{\Pr(A=i',C=k)}}\right) \Bigg/ \left(\frac{\frac{\Pr(A=i,B=j',C=k)}{\Pr(A=i,C=k)}}{\frac{\Pr(A=i,B=j,C=k)}{\Pr(A=i,C=k)}}\right)\\
&= \frac{\Pr(A=i',B=j',C=k)}{\Pr(A=i',B=j,C=k)} \Big/ \frac{\Pr(A=i,B=j',C=k)}{\Pr(A=i,B=j,C=k)} = \frac{\pi_{i'j'k}\pi_{ijk}}{\pi_{i'jk}\pi_{ij'k}}\\
&= \frac{(\omega\alpha_{i'}\beta_{j'}\gamma_k\delta_{i'j'}\epsilon_{i'k}\phi_{j'k})(\omega\alpha_i\beta_j\gamma_k\delta_{ij}\epsilon_{ik}\phi_{jk})}{(\omega\alpha_{i'}\beta_j\gamma_k\delta_{i'j}\epsilon_{i'k}\phi_{jk})(\omega\alpha_i\beta_{j'}\gamma_k\delta_{ij'}\epsilon_{ik}\phi_{j'k})} = \frac{\delta_{i'j'}\delta_{ij}}{\delta_{i'j}\delta_{ij'}}\\
&= \exp\left(\lambda^{AB}_{i'j'} - \lambda^{AB}_{i'j} - \lambda^{AB}_{ij'} + \lambda^{AB}_{ij}\right)
\end{aligned}$$

נשים לב כי הביטוי המתקבל חף מ- k. ולכן, הטענה הוכחה.

<u>שאלה 5 (15 נקודות)</u>

נתייחס לחומר בנספח 1 עבור הרצה של רגרסיה לוגיסטית.

נסמן ב- x_1 את המצב $\text{term} = 24, \text{amount} = 10{,}000, \text{age} = 30, \text{land} = 0$.
נסמן ב- x_2 את המצב $\text{term} = 36, \text{amount} = 15{,}000, \text{age} = 40, \text{land} = 1$.
בנוסף, עבור כל x אפשרי נסמן $p(x) = \Pr(Y = 1|X = x)$.

א. חשבו רווח סמך (ברמת ביטחון של 95%) עבור $p(x_1)$.

פיתרון

תחילה, נזכור כי

$$p(x_1) = \Pr(Y = 1|X = x_1) = \frac{e^{x_1^T\beta}}{1 + e^{x_1^T\beta}}$$

היות ו- $p(x_1)$ הינה פונקציה "יפה", נחשב רווח סמך לביטוי $x_1^T\hat{\beta}$ ונציבו בהסתברות המבוקשת.

לשם חישוב רווח סמך ל- $x_1^T\hat{\beta}$, נשתמש בעובדות הבאות.

$$\hat{\beta} \sim N(\beta, (X^TWX)^{-1})$$
$$x^T\hat{\beta} \sim N(x^T\beta, x^T(X^TWX)^{-1}x)$$

ולכן, נוסחת רווח הסמך תיראה כך.

$$x_1^T\hat{\beta} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{x_1^T(X^TWX)^{-1}x_1}$$

נגדיר,

$$x_1^T = (1 \quad \log(24) \quad \log(10^4) \quad \log(30) \quad 0)$$

$$\hat{\beta} = \begin{pmatrix} -1.1179 \\ -0.9297 \\ 0.187 \\ 0.956 \\ -0.7341 \end{pmatrix}$$

והמטריצה $(X^TWX)^{-1}$ הינה המטריצה המצויה בתחתית "נספח 1".

נחשב רכיבי נוסחת רווח הסמך.

$$x_1^T\hat{\beta} = (1 \quad 3.1781 \quad 9.2103 \quad 3.4012 \quad 0)\begin{pmatrix} -1.1179 \\ -0.9297 \\ 0.187 \\ 0.956 \\ -0.7341 \end{pmatrix} = 0.9012$$

$$x_1^T(X^TWX)^{-1}x_1 = (1 \quad 3.1781 \quad 9.2103 \quad 3.4012 \quad 0)(X^TWX)^{-1}\begin{pmatrix} 1 \\ 3.1781 \\ 9.2103 \\ 3.4012 \\ 0 \end{pmatrix} = 0.0304$$

ולכן,

$$CI(x_1^T\beta) = 0.9012 \pm 1.96 \cdot \sqrt{0.0304} = [0.5595 \quad 1.2428]$$

(כידוע, $Z_{0.975} \approx 1.96$).

$$CI\big(p(x_1)\big) = \left[\frac{e^{0.5595}}{1 + e^{0.5595}} \quad \frac{e^{1.2428}}{1 + e^{1.2428}}\right] = \boxed{[0.6363 \quad 0.776]}$$

ב. חשבו רווח סמך (ברמת ביטחון של 95%) עבור $\frac{p(x_2)}{1-p(x_2)} \Big/ \frac{p(x_1)}{1-p(x_1)}$.

פיתרון

ראשית, נפשט הביטוי לעיל.

$$\frac{p(x_2)}{1-p(x_2)}\Big/\frac{p(x_1)}{1-p(x_1)} = \frac{\frac{e^{x_2^T\beta}}{1+e^{x_2^T\beta}}}{1-\frac{e^{x_2^T\beta}}{1+e^{x_2^T\beta}}}\Bigg/\frac{\frac{e^{x_1^T\beta}}{1+e^{x_1^T\beta}}}{1-\frac{e^{x_1^T\beta}}{1+e^{x_1^T\beta}}} = \frac{\frac{e^{x_2^T\beta}}{1+e^{x_2^T\beta}}}{\frac{1}{1+e^{x_2^T\beta}}}\Bigg/\frac{\frac{e^{x_1^T\beta}}{1+e^{x_1^T\beta}}}{\frac{1}{1+e^{x_1^T\beta}}} = \frac{e^{x_2^T\beta}}{e^{x_1^T\beta}}$$
$$= e^{(x_2^T-x_1^T)\beta}$$

נגדיר,

$$u = x_2 - x_1$$

היות ו- $e^{u^T\beta}$ פונקציה "יפה". נוכל לחשב רווח סמך עבור $u^T\beta$, ולהציב התוצאה במנת יחס הסיכויים המבוקשת. התהליך זהה לחלוטין לזה אשר בוצע בסעיף א'.

$$\begin{aligned} u^T &= x_2^T - x_1^T \\ &= (1 \quad \log(36) \quad \log(15\cdot 10^3) \quad \log(40) \quad 1) \\ &- (1 \quad \log(24) \quad \log(10^4) \quad \log(30) \quad 0) \\ &= (0 \quad 0.4055 \quad 0.4055 \quad 0.2877 \quad 1) \end{aligned}$$

$$u^T\hat{\beta} = (0 \quad 0.4055 \quad 0.4055 \quad 0.2877 \quad 1)\begin{pmatrix}-1.1179\\-0.9297\\0.187\\0.956\\-0.7341\end{pmatrix} = -0.7603$$

$$u^T(X^TWX)^{-1}u = (0 \quad 0.4055 \quad 0.4055 \quad 0.2877 \quad 1)(X^TWX)^{-1}\begin{pmatrix}0\\0.4055\\0.4055\\0.2877\\1\end{pmatrix} = 0.0366$$

$$CI(u^T\beta) = -0.7603 \pm 1.96\cdot\sqrt{0.0366} = [-1.1353 \quad -0.3852]$$

$$CI\left(\frac{p(x_2)}{1-p(x_2)}\Big/\frac{p(x_1)}{1-p(x_1)}\right) = [e^{-1.1353} \quad e^{-0.3852}] = \boxed{[0.3213 \quad 0.6803]}$$

נספח 1

Data on 1000 loan customers of a German bank.

The variables are as follows:
gc: 0/1 indicator of whether the person is a good risk for the requested loan (1) or not 0)
term: duration of requested loan
amount: amount requested (in German Marks)
age: age of the customer (years)
land: 0/1 indicator of whether the customer owns land (1) or does not (0)

<u>We apply a log transformation to term, amount, and age before running the regression.</u>

lterm = log(term)
amtlog = log(amount)
l.age = log(age)

Results of Logistic Regression Analysis

	Coef	S.E.	Wald Z	Pr(>|Z|)
Intercept	-1.1179	1.1539	-0.97	0.3327
lterm	-0.9297	0.1714	-5.43	<0.0001
amtlog	0.1870	0.1230	1.52	0.1284
l.age	0.9560	0.2565	3.73	0.0002
land	-0.7341	0.1980	-3.71	0.0002

Covariance Matrix of Estimated Coefficients

	Intercept	lterm	amtlog	l.age	land
Intercept	1.33159839	0.0137349459	-0.0752634941	-0.2258012175	0.0688971954
lterm	0.01373495	0.0293662476	-0.0134884945	0.0013452636	-0.0009395372
amtlog	-0.07526349	-0.0134884945	0.0151238319	-0.0007412548	-0.0036072797
l.age	-0.22580122	0.0013452636	-0.0007412548	0.0658113459	-0.0128218795
land	0.06889720	-0.0009395372	-0.0036072797	-0.0128218795	0.0392093907

נספח 2

לוח של התפלגות F

ערכים קריטיים עבור $\alpha = 0.05$ בהתפלגות F עם d_1 ד״ח במונה ו- d_2 ד״ח במכנה

$d_2 \backslash d_1$	1	2	3	4	5	6	7	8	9	10
1	161.45	18.51	10.13	7.71	6.61	5.99	5.59	5.32	5.12	4.96
2	199.50	19.00	9.55	6.94	5.79	5.14	4.74	4.46	4.26	4.10
3	215.71	19.16	9.28	6.59	5.41	4.76	4.35	4.07	3.86	3.71
4	224.58	19.25	9.12	6.39	5.19	4.53	4.12	3.84	3.63	3.48
5	230.16	19.30	9.01	6.26	5.05	4.39	3.97	3.69	3.48	3.33
6	233.99	19.33	8.94	6.16	4.95	4.28	3.87	3.58	3.37	3.22
7	236.77	19.35	8.89	6.09	4.88	4.21	3.79	3.50	3.29	3.14
8	238.88	19.37	8.85	6.04	4.82	4.15	3.73	3.44	3.23	3.07
9	240.54	19.38	8.81	6.00	4.77	4.10	3.68	3.39	3.18	3.02
10	241.88	19.40	8.79	5.96	4.74	4.06	3.64	3.35	3.14	2.98
11	242.98	19.40	8.76	5.94	4.70	4.03	3.60	3.31	3.10	2.94
12	243.91	19.41	8.74	5.91	4.68	4.00	3.57	3.28	3.07	2.91
13	244.69	19.42	8.73	5.89	4.66	3.98	3.55	3.26	3.05	2.89
14	245.36	19.42	8.71	5.87	4.64	3.96	3.53	3.24	3.03	2.86
15	245.95	19.43	8.70	5.86	4.62	3.94	3.51	3.22	3.01	2.85
16	246.46	19.43	8.69	5.84	4.60	3.92	3.49	3.20	2.99	2.83
17	246.92	19.44	8.68	5.83	4.59	3.91	3.48	3.19	2.97	2.81
18	247.32	19.44	8.67	5.82	4.58	3.90	3.47	3.17	2.96	2.80
19	247.69	19.44	8.67	5.81	4.57	3.88	3.46	3.16	2.95	2.79
20	248.01	19.45	8.66	5.80	4.56	3.87	3.44	3.15	2.94	2.77
21	248.31	19.45	8.65	5.79	4.55	3.86	3.43	3.14	2.93	2.76
22	248.58	19.45	8.65	5.79	4.54	3.86	3.43	3.13	2.92	2.75
23	248.83	19.45	8.64	5.78	4.53	3.85	3.42	3.12	2.91	2.75
24	249.05	19.45	8.64	5.77	4.53	3.84	3.41	3.12	2.90	2.74
25	249.26	19.46	8.63	5.77	4.52	3.83	3.40	3.11	2.89	2.73
26	249.45	19.46	8.63	5.76	4.52	3.83	3.40	3.10	2.89	2.72
27	249.63	19.46	8.63	5.76	4.51	3.82	3.39	3.10	2.88	2.72
28	249.80	19.46	8.62	5.75	4.50	3.82	3.39	3.09	2.87	2.71
29	249.95	19.46	8.62	5.75	4.50	3.81	3.38	3.08	2.87	2.70
30	250.10	19.46	8.62	5.75	4.50	3.81	3.38	3.08	2.86	2.70
31	250.23	19.46	8.61	5.74	4.49	3.80	3.37	3.07	2.86	2.69
32	250.36	19.46	8.61	5.74	4.49	3.80	3.37	3.07	2.85	2.69
33	250.48	19.47	8.61	5.74	4.48	3.80	3.36	3.07	2.85	2.69
34	250.59	19.47	8.61	5.73	4.48	3.79	3.36	3.06	2.85	2.68
35	250.69	19.47	8.60	5.73	4.48	3.79	3.36	3.06	2.84	2.68
36	250.79	19.47	8.60	5.73	4.47	3.79	3.35	3.06	2.84	2.67
37	250.89	19.47	8.60	5.72	4.47	3.78	3.35	3.05	2.84	2.67
38	250.98	19.47	8.60	5.72	4.47	3.78	3.35	3.05	2.83	2.67
39	251.06	19.47	8.60	5.72	4.47	3.78	3.34	3.05	2.83	2.66
40	251.14	19.47	8.59	5.72	4.46	3.77	3.34	3.04	2.83	2.66