52518

מודלים סטטיסטיים ויישומיהם

תשע"ו

<u>מועד א'</u>

<u>פיתרון</u>

<u>שאלה 1</u>

תזכורת,

$$\beta^T X_i = X_i^T \beta = \sum_{j=0}^{p} \beta_j X_{ij}$$

המשוואות אשר נצטרך לפתור לשם קבלת אומד הנראות המרבית לווקטור β הינן,

$$\frac{\partial l(Y,X;\ \beta)}{\partial \beta_0} = 0$$

$$\vdots$$

$$\frac{\partial l(Y,X;\ \beta)}{\partial \beta_p} = 0$$

כאשר, $l(Y,X;\ \beta) = \log\big(L(Y,X,;\beta)\big)$.

נציג כיצד תראה משוואה שכזו, עבור $\beta_r, r \in [0,\dots,p]$ כלשהו.

$$L(Y,X,\beta) = \prod_{i=1}^{n}\left[1 - e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}\right]^{Y_i}\left[1 - \left(1 - e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}\right)\right]^{1-Y_i}$$

$$= \prod_{i=1}^{n}\left[1 - e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}\right]^{Y_i}\left[e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}\right]^{1-Y_i}$$

$$l(Y,X,\beta) = \sum_{i=1}^{n}\left[Y_i \log\left(1 - e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}\right) + (1-Y_i)\log\left(e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}\right)\right]$$

$$= \sum_{i=1}^{n}\left[Y_i \log\left(1 - e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}\right) - (1-Y_i)e^{\sum_{j=0}^{p}\beta_j X_{ij}}\right]$$

$$\frac{\partial l(Y,X,\beta)}{\partial \beta_r} = \sum_{i=1}^{n}\left[Y_i \frac{X_{ir} e^{\left(\sum_{j=0}^{p}\beta_j X_{ij} - e^{\sum_{j=0}^{p}\beta_j X_{ij}}\right)}}{1-e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}} - (1-Y_i) X_{ir} e^{\sum_{j=0}^{p}\beta_j X_{ij}}\right]$$

$$= \sum_{i=1}^{n} X_{ir} e^{\sum_{j=0}^{p}\beta_j X_{ij}} \left[\frac{Y_i e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}}{1-e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}} - (1-Y_i)\right]$$

$$= \sum_{i=1}^{n} X_{ir} e^{\sum_{j=0}^{p}\beta_j X_{ij}} \left[\frac{Y_i e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}} - \left(1-e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}\right) + Y_i - Y_i e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}}{1-e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}}\right]$$

$$= \sum_{i=1}^{n} X_{ir} e^{\sum_{j=0}^{p}\beta_j X_{ij}} \left[\frac{Y_i - \left(1-e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}\right)}{1-e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}}\right]$$

$$= \sum_{i=1}^{n} X_{ir} e^{\sum_{j=0}^{p}\beta_j X_{ij}} \left[\frac{Y_i}{1-e^{-e^{\sum_{j=0}^{p}\beta_j X_{ij}}}} - 1\right] = 0$$

שאלה 2

משמעות המודל (AB, AC, AD) הינה, B, C ו- D בלתי תלויים בהינתן A.

להלן, המודל הלוג ליניארי המתאים.

$$\log \pi_{ijkl} = \bar{\theta}_{....} + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{il}^{AD}$$

כעת, נציג ההוכחה.

כיוון I:

$$\Pr(B=j, C=k, D=l | A=i) = \Pr(B=j|A=i)\Pr(C=k|A=i)\Pr(D=l|A=i)$$

$$\Rightarrow \frac{\Pr(A=i, B=j, C=k, D=l)}{\Pr(A=i)}$$

$$= \frac{\Pr(A=i, B=j)}{\Pr(A=i)} \frac{\Pr(A=i, C=k)}{\Pr(A=i)} \frac{\Pr(A=i, D=l)}{\Pr(A=i)}$$

נזכור כי, $\pi_{ijkl} = \Pr(A=i, B=j, C=k, D=l)$. ומכאן נסיק כי עלינו להוכיח את קיום השוויון הבא.

$$\pi_{ijkl} = \frac{\pi_{ij..}\pi_{i.k.}\pi_{i..l}}{\pi_{i...}\pi_{i...}}$$

בטרם נמשיך בהוכחה, נגדיר הפרמטרים הבאים.

$$\omega = \exp\{\bar{\theta}_{....}\}$$
$$\alpha_i = \exp\{\lambda_i^A\}$$
$$\beta_j = \exp\{\lambda_j^B\}$$
$$\gamma_k = \exp\{\lambda_k^C\}$$
$$\delta_l = \exp\{\lambda_l^D\}$$
$$\epsilon_{ij} = \exp\{\lambda_{ij}^{AB}\}$$
$$\phi_{ik} = \exp\{\lambda_{ik}^{AC}\}$$
$$\eta_{il} = \exp\{\lambda_{il}^{AD}\}$$

נתחיל בפיתוח צד שמאל.

$$\pi_{ijkl} = \omega\alpha_i\beta_j\gamma_k\delta_l\epsilon_{ij}\phi_{ik}\eta_{il}$$

כעת, נוודא כי צידו הימני של השוויון שקול לביטוי לעיל.

$$\frac{\pi_{ij..}\pi_{i.k.}\pi_{i..l}}{\pi_{i...}\pi_{i...}} = \frac{\omega\alpha_i\beta_j\gamma_.\delta_.\epsilon_{ij}\phi_{i.}\eta_{i.}\omega\alpha_i\beta_.\gamma_k\delta_.\epsilon_{i.}\phi_{ik}\eta_{i.}\omega\alpha_i\beta_.\gamma_.\delta_l\epsilon_{i.}\phi_{i.}\eta_{il}}{\omega\alpha_i\beta_.\gamma_.\delta_.\epsilon_{i.}\phi_{i.}\eta_{i.}\omega\alpha_i\beta_.\gamma_.\delta_.\epsilon_{i.}\phi_{i.}\eta_{i.}}$$
$$= \omega\alpha_i\beta_j\gamma_k\delta_l\epsilon_{ij}\phi_{ik}\eta_{il}$$

כיוון II:

נתייחס אל השוויון הבא כנתון, ונמצא מהו המודל הלוג ליניארי המתאים לו.

$$\pi_{ijkl} = \frac{\pi_{ij..}\pi_{i.k.}\pi_{i..l}}{\pi_{i...}\pi_{i...}}$$

כמובן,

$$\theta_{ijkl} = \log\pi_{ijkl} = \log\pi_{ij..} + \log\pi_{i.k.} + \log\pi_{i..l} - 2\log\pi_{i...}$$

עתה, נתבונן במודל המלא של ארבעת הגורמים.

$$\theta^\star_{ijkl} = \bar{\theta}_{....} + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{il}^{AD} + \lambda_{jk}^{BC} + \lambda_{jl}^{BD} + \lambda_{kl}^{CD} + \lambda_{ijk}^{ABC}$$
$$+ \lambda_{ijl}^{ABD} + \lambda_{jkl}^{BCD} + \lambda_{ijkl}^{ABCD}$$

בכדי לקבל המודל הרצוי, עלינו להוכיח כי,

$$\lambda_{jk}^{BC} = 0 \ \ \forall(j,k)$$
$$\lambda_{jl}^{BD} = 0 \ \ \forall(j,l)$$
$$\lambda_{kl}^{CD} = 0 \ \ \forall(k,l)$$

כמו כן, עלינו להראות כי כל ה- λ-ות מסדר שלישי ורביעי הן אפס.

על פי הגדרה, ידוע כי,

$$\lambda_{jk}^{BC} = \bar{\theta}_{.jk.} - \bar{\theta}_{.j..} - \bar{\theta}_{..k.} + \bar{\theta}_{....}$$

נפתח כל איבר בשוויון לעיל.

$$\begin{aligned}
\bar{\theta}_{.jk.} &= \frac{1}{IL}\sum_{i'}\sum_{l'}\theta_{i'jkl'} \\
&= \frac{1}{I}\sum_{i'}\log\pi_{i'j..} + \frac{1}{I}\sum_{i'}\log\pi_{i'.k.} + \frac{1}{IL}\sum_{i'}\sum_{l'}\log\pi_{i'..l'} \\
&\quad - 2\frac{1}{I}\sum_{i'}\log\pi_{i'...}
\end{aligned}$$

$$\bar{\theta}_{.j..} = \frac{1}{I}\sum_{i'}\log\pi_{i'j..} + \frac{1}{IK}\sum_{i'}\sum_{k'}\log\pi_{i'.k'.} + \frac{1}{IL}\sum_{i'}\sum_{l'}\log\pi_{i'..l'} - 2\frac{1}{I}\sum_{i'}\log\pi_{i'...}$$

$$\bar{\theta}_{..k.} = \frac{1}{IJ}\sum_{i'}\sum_{j'}\log\pi_{i'j'..} + \frac{1}{I}\sum_{i'}\log\pi_{i'.k.} + \frac{1}{IL}\sum_{i'}\sum_{l'}\log\pi_{i'..l'} - 2\frac{1}{I}\sum_{i'}\log\pi_{i'...}$$

$$\begin{aligned}
\bar{\theta}_{....} &= \frac{1}{J}\sum_{j'}\bar{\theta}_{.j'..} \\
&= \frac{1}{IJ}\sum_{i'}\sum_{j'}\log\pi_{i'j'..} + \frac{1}{IK}\sum_{i'}\sum_{k'}\log\pi_{i'.k'.} + \frac{1}{IL}\sum_{i'}\sum_{l'}\log\pi_{i'..l'} \\
&\quad - 2\frac{1}{I}\sum_{i'}\log\pi_{i'...}
\end{aligned}$$

ניתן לראות בנקל כי,

$$\lambda_{jk}^{BC} = \bar{\theta}_{.jk.} - \bar{\theta}_{.j..} - \bar{\theta}_{..k.} + \bar{\theta}_{....} = 0$$

בצורה זהה, נוכל להוכיח כי λ_{jl}^{BD} ו- λ_{kl}^{CD} מתאפסות.

כעת, נתבונן ב- λ_{ijk}^{ABC}.

$$\lambda_{ijk}^{ABC} = \bar{\theta}_{ijk.} - \bar{\theta}_{ij..} - \bar{\theta}_{i.k.} - \bar{\theta}_{.jk.} + \bar{\theta}_{i...} + \bar{\theta}_{.j..} + \bar{\theta}_{..k.} - \bar{\theta}_{....}$$

בעזרת אשר הוכח כבר נראה כי ניתן לפתח השוויון באופן הבא.

$$\lambda_{ijk}^{ABC} = \bar{\theta}_{ijk} - \bar{\theta}_{ij..} - \bar{\theta}_{i.k.} + \bar{\theta}_{i...} - \lambda_{jk}^{BC}$$

היות והוכחנו כי λ_{jk}^{BC} שווה לאפס, נותר לפתח שאר ה- $\bar{\theta}$-ות כפי שהוצג לעיל. וכך להשלים ההוכחה.

את λ_{ijkl}^{ABCD} ניתן להציג כך.

$$\lambda_{ijkl}^{ABCD} = \theta_{ijkl} - \lambda_i^A - \lambda_j^B - \lambda_k^C - \lambda_l^D - \lambda_{ij}^{AB} - \lambda_{ik}^{AC} - \lambda_{il}^{AD} - \bar{\theta}_{....}$$

שימו לב, ה- λ-ות אשר הוכחו כאפסים אינן מופיעות. נמשיך,

$$\lambda_{ijkl}^{ABCD} = \theta_{ijkl} - \bar{\theta}_{ij..} - \bar{\theta}_{i.k.} - \bar{\theta}_{i..l} + 2\bar{\theta}_{i...}$$

נפתח הביטוי בטכניקה המקובלת, ונראה כי גם הוא מתאפס.

כנדרש!

שאלה 3

ראשית נציג את מודל ניתוח השונות החד-כיווני כמודל רגרסיה ליניארית.

$$Y = X\beta + \epsilon\,; \quad \epsilon \sim N(0, \sigma^2 I)$$

כאשר, $Y = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{In_I} \end{pmatrix}$, $\beta = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_I \end{pmatrix}$, $\epsilon = \begin{pmatrix} \epsilon_{11} \\ \vdots \\ \epsilon_{In_I} \end{pmatrix}$ ומטריצת התכנון $X_{(N\times I)}$ תקבל בכל אחת מ- I עמודותיה 1 לו התצפית שייכת לקבוצה ה- i, ו- 0 אחרת (למשל, כל התצפיות השייכות לקבוצה הראשונה תקבלנה 1 בעמודה הראשונה ואפס בשאר העמודות).

נזכור כי נוסחה כללית עבור סטטיסטי F ברגרסיה ליניארית הינו,

$$F = \frac{\left\|\hat{Y} - \hat{Y}^{(H_0)}\right\|^2 / (d - d_0)}{\left\|Y - \hat{Y}\right\|^2 / (N - d)}$$

כאשר, d - מספר מקדמי הרגרסיה במודל המלא ו- d_0 - מספר מקדמי הרגרסיה במודל השערת האפס.

במקרה של ניתוח שונות חד-כיווני נקבל,

$$d = I$$

$$\hat{Y}_{ij} = \bar{Y}_{i.}$$

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

לשם השלמת חישוב הסטטיסטי לעיל, נותר עלינו לחשב את המונה שלו. בפרט, את $\hat{Y}^{(H_0)}$.

במונחי ניתוח שונות חד-כיווני, ייראה מודל השערת האפס באופן הבא.

$$Y_{ij} = \mu_i^{(H_0)} + \epsilon_{ij} = \gamma v_i + \epsilon_{ij}$$

ובמונחי רגרסיה ליניארית, המודל יראה כך,

$$Y = X^{(H_0)}\beta^{(H_0)} + \epsilon$$

כאשר,

$$\beta^{(H_0)} = \gamma$$

$$X^{(H_0)} = XV = \begin{pmatrix} v_1 \\ \vdots \\ v_1 \\ v_2 \\ \vdots \\ v_2 \\ \vdots \\ v_I \\ \vdots \\ v_I \end{pmatrix}$$

(כל v_i מופיע n_i פעמים).

היות ובמודל השערת האפס קיים מקדם רגרסיה בודד, $d_0 = 1$.

עתה, נאמוד את γ ($\beta^{(H_0)}$) לפי שיטת הריבועים הפחותים.

$$\hat{\beta}^{(H_0)} = \hat{\gamma} = \left(X^{(H_0)^T} X^{(H_0)}\right)^{-1} X^{(H_0)^T} Y = ((XV)^T XV)^{-1}(XV)^T Y$$

$$= (V^T X^T XV)^{-1} V^T X^T Y = \left(V^T \begin{pmatrix} n_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & n_I \end{pmatrix} V\right)^{-1} V^T \begin{pmatrix} \sum_{j=1}^{n_1} Y_{1j} \\ \vdots \\ \sum_{j=1}^{n_I} Y_{Ij} \end{pmatrix}$$

$$= \frac{\sum_{i=1}^{I} v_i Y_{i.}}{\sum_{i=1}^{I} n_i v_i^2} = \frac{\sum_{i=1}^{I} n_i v_i \bar{Y}_{i.}}{\sum_{i=1}^{I} n_i v_i^2}$$

לכן,

$$\hat{Y}_{ij}^{(H_0)} = \left(\frac{\sum_{i'=1}^{I} n_{i'} v_{i'} \bar{Y}_{i'.}}{\sum_{i'=1}^{I} n_{i'} v_{i'}^2}\right) v_i \quad \forall (i,j)$$

נסיים, חישוב המונה.

$$\left\|\hat{Y} - \hat{Y}^{(H_0)}\right\|^2 = \sum_{i=1}^{I} n_i \left(\bar{Y}_{i.} - \left(\frac{\sum_{i'=1}^{I} n_{i'} v_{i'} \bar{Y}_{i'.}}{\sum_{i'=1}^{I} n_{i'} v_{i'}^2}\right) v_i\right)^2$$

$$= \sum_{i=1}^{I} n_i \bar{Y}_{i.}^2 - 2\left(\frac{\sum_{i'=1}^{I} n_{i'} v_{i'} \bar{Y}_{i'.}}{\sum_{i'=1}^{I} n_{i'} v_{i'}^2}\right) \sum_{i=1}^{I} n_i v_i \bar{Y}_{i.}$$

$$+ \left(\frac{\sum_{i'=1}^{I} n_{i'} v_{i'} \bar{Y}_{i'.}}{\sum_{i'=1}^{I} n_{i'} v_{i'}^2}\right)^2 \sum_{i=1}^{I} n_i v_i^2$$

$$= \sum_{i=1}^{I} n_i \bar{Y}_{i.}^2 - 2\frac{\left(\sum_{i'=1}^{I} n_{i'} v_{i'} \bar{Y}_{i'.}\right)^2}{\sum_{i'=1}^{I} n_{i'} v_{i'}^2} + \frac{\left(\sum_{i'=1}^{I} n_{i'} v_{i'} \bar{Y}_{i'.}\right)^2}{\sum_{i'=1}^{I} n_{i'} v_{i'}^2}$$

$$= \sum_{i=1}^{I} n_i \bar{Y}_{i.}^2 - \frac{\left(\sum_{i'=1}^{I} n_{i'} v_{i'} \bar{Y}_{i'.}\right)^2}{\sum_{i'=1}^{I} n_{i'} v_{i'}^2}$$

אי לכך,

$$F = \frac{\left(\sum_{i=1}^{I} n_i \bar{Y}_{i.}^2 - \frac{\left(\sum_{i'=1}^{I} n_{i'} v_{i'} \bar{Y}_{i'.}\right)^2}{\sum_{i'=1}^{I} n_{i'} v_{i'}^2}\right) \Big/ (I-1)}{s^2} \sim F_{I-1,N-I}$$

שאלה 4

לפנינו המודל הבא,

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

תהנה $\left(\{\pi_i^\star\}_{i=1}^I, \{\tau_j^\star\}_{j=1}^J\right)$ אוסף משקולות מסוים עבור הפרמטרים μ, α_i ו- β_j.

$$\alpha_r^\star = \bar{\mu}_{r.}^\star - \mu^\star$$

כאשר,

$$\bar{\mu}_{r.}^\star = \sum_{j=1}^J \tau_j^\star \mu_{rj}$$

$$\mu^\star = \sum_{i=1}^I \sum_{j=1}^J \pi_i^\star \tau_j^\star \mu_{ij}$$

ונוכיח הטענה ישירות.

$$\alpha_l^\star - \alpha_m^\star = (\bar{\mu}_{l.}^\star - \mu^\star) - (\bar{\mu}_{m.}^\star - \mu^\star) = \bar{\mu}_{l.}^\star - \bar{\mu}_{m.}^\star = \sum_{j=1}^J \tau_j^\star \mu_{lj} - \sum_{j=1}^J \tau_j^\star \mu_{mj}$$

$$= \sum_{j=1}^J \tau_j^\star (\mu_{lj} - \mu_{mj}) = \sum_{j=1}^J \tau_j^\star \left(\mu + \alpha_l + \beta_j - (\mu + \alpha_m + \beta_j)\right)$$

$$= (\alpha_l - \alpha_m) \sum_{j=1}^J \tau_j^\star = (\alpha_l - \alpha_m)$$

ולהלן ההוכחה כי ההפרש המדובר חף מתלות בבחירת המשקולות.

שאלה 5

נוסחת רווח סמך לפי שיטת בונפרוני נראית כך,

$$\hat{\psi}\left(c^{(k)}\right) \pm t_{\left[N-I\,, 1-\frac{\alpha}{2K}\right]} \cdot s\sqrt{v\left(c^{(k)}\right)}$$

כאשר,

$$\hat{\psi}\left(c^{(k)}\right) = \sum_{i=1}^I c_i^{(k)} \bar{X}_i$$

$$v\left(c^{(k)}\right) = \sum_{i=1}^{I} \frac{{c_i^{(k)}}^2}{n_i}$$

היות ואנו מתבקשים לחשב רווחי סמך בו-זמניים לכלל הפרשי הזוגות במדגם, K הינו 3.

בטרם נחשב את שלושת רווחי הסמך, נאתר הערך $t_{\left[N-I\,,1-\frac{\alpha}{2K}\right]}$.

למעשה, עלינו לחשב ערכו של $t_{\left[21-3,1-\frac{0.05}{2\cdot 3}\right]} = t_{[18,0.9917]}$. היות ואין בטבלת ה- t המצורפת את ההסתברות $1-0.9917=0.0083$ (בטבלת זנב ימני נחפש ערכים עבור $\frac{\alpha}{2K}$), נשתמש בנוסחת האינטרפולציה הליניארית לאתר קירוב לערך המבוקש.

נגדיר,

$$a = 0.0083$$
$$x_1 = 0.005$$
$$x_2 = 0.01$$

(נזכור כי על x_1 להיות גדול מ- x_2, ועל a להימצא בין שני ה- x-ים).

עתה, נשלים החישוב.

$$y_1 = t_{[18,x_1]} = 2.878$$
$$y_2 = t_{[18,x_2]} = 2.552$$
$$t_{[18,a]} \approx y_1 + \left(\frac{y_2-y_1}{x_2-x_1}\right)(a-x_1) = 2.878 + \frac{-0.326}{0.005}\cdot 0.0033 = 2.6628$$

כעת, נחשב את שלושת רווחי הסמך.

$$CI(\psi\left(c^{(1)}\right) = (1\cdot 6 - 1\cdot 4) \pm 2.6628\cdot 2.0494\sqrt{\frac{1}{6}+\frac{1}{7}} = \boxed{[-1.0361 \quad 5.0361]}$$

$$CI(\psi\left(c^{(2)}\right) = (1\cdot 6 - 1\cdot 3.4) \pm 2.6628\cdot 2.0494\sqrt{\frac{1}{6}+\frac{1}{8}} = \boxed{[-0.3472 \quad 5.5472]}$$

$$CI(\psi\left(c^{(1)}\right) = (1\cdot 4 - 1\cdot 3.4) \pm 2.6628\cdot 2.0494\sqrt{\frac{1}{7}+\frac{1}{8}} = \boxed{[-2.2243 \quad 3.4243]}$$