רגרסיה- תרגול 5- תשפ"ה

תזכורת

פונקציית סיכון MSE:

יהי θ פרמטר לא ידוע ו-$\hat{\theta}$ אומד ל-$\theta \in \Theta$. נגדיר את פונקציית ה-$MSE(\theta,\hat{\theta})$:

$$E(\hat{\theta}-\theta)^2 = Var(\hat{\theta}) + E^2(\hat{\theta}-\theta)$$

ובאופן כללי עבור $\theta \in \Theta \subseteq R^p$:

$$MSE(\hat{\theta},\theta) = E\left|\left|\hat{\theta}-\theta\right|\right|^2$$

נאמר כי אומד $\hat{\theta}$ טוב יותר מאומד $\tilde{\theta}$ אם מתקיים שלכל $\theta \in \Theta$:

$$MSE(\theta,\tilde{\theta}) \geq MSE(\theta,\hat{\theta})$$

וקיים לפחות θ' אחד עבורו אי השיוויון חזק.

שאלה:

הראו כי במקרה הרב מימדי:

$$MSE(\hat{\theta},\theta) = E\left|\left|\hat{\theta}-\theta\right|\right|^2 = tr[Var(\hat{\theta})] + \left|\left|E(\hat{\theta})-\theta\right|\right|^2$$

פתרון:

$$E\left|\left|\hat{\theta}-\theta\right|\right|^2 = E\left|\left|\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta\right|\right|^2$$
$$= E\left|\left|\hat{\theta}-E(\theta)\right|\right|^2 + 2E\left[\left(\hat{\theta}-E(\hat{\theta})\right)^T\left(E(\hat{\theta})-\theta\right)\right] + E\left|\left|E(\hat{\theta})-\theta\right|\right|^2$$

הביטוי הימני ביותר דטרמיניסטי ולכן שווה לתוחלתו. באשר לביטוי האמצעי- $\left(E(\hat{\theta})-\theta\right)$ גם כן דטרמיניסטי ולכן ניתן להוציאו מהתוחלת. נקבל:

$$E\left[\left(\hat{\theta}-E(\hat{\theta})\right)^T\left(E(\hat{\theta})-\theta\right)\right] = \left(E[\hat{\theta}-E(\hat{\theta})]\right)^T\left(E(\hat{\theta})-\theta\right) = 0^T\left(E(\hat{\theta})-\theta\right) = 0$$

כעת, בעבור הביטוי השמאלי, נשים לב שהמשתנה $\hat{\theta}-E(\hat{\theta})$ הוא מ"מ עם תוחלת 0. נשתמש בטענה מתרגול 3:

הסיקו מכך כי אם $\mathbb{E}[Z]=0$ אזי מתקיים כי

$$\mathbb{E}\left(||Z||^2\right) = tr(cov[Z])$$

ונסיק כי:

$$E\left|\left|\hat{\theta}-E(\theta)\right|\right|^2 = tr\left(Var\left(\hat{\theta}-E(\theta)\right)\right) = tr\left(Var(\hat{\theta})\right)$$

שמשלים את ההוכחה.

משפט גאוס-מרקוב (אומד OLS הוא $BLUE$)

תהי $\theta := a^T\beta$. נסמן את אומד OLS ל-θ: $\hat{\theta} := a^T\hat{\beta} := c^TY$ עבור $c = X(X^TX)^{-1}a$. זהו אומד לינארי ב-Y חסר הטיה ל-θ. נסמן ב-d^TY אומד חסר הטיה אחר שגם הוא לינארי ב-Y. אז מתקיים שלכל θ:

$$Var(d^TY,\theta) \geq Var(c^TY,\theta) \Rightarrow MSE(d^TY,\theta) \geq MSE(c^TY,\theta)$$

שאלה- לא צריכים לדעת להוכיח בעצמכם כרגע

הראו כי אומד ה-$BLUE$ הינו יחיד :

נניח שרוצים לאמוד את הקומבינציה הלינארית $\theta := a^T\beta$, תוך שימוש באומדים לינאריים בוקטור Y. נגדיר $a^T\hat{\beta}_{OLS} = c^TY$.
נניח שקיים אומד לינארי ב-Y אחר וח"ה עם שונות זהה לשל אומד OLS. נסמנו ב-d^TY, אז הצירוף:

$$T = 0.5c^TY + 0.5d^TY = 0.5(c+d)^TY$$

גם הוא לינארי וחסר הטיה. נחשב את השונות שלו:
תזכורת: לכל זוג משתנים X_1, X_2 מתקיים:

$$Cov(X_1, X_2) \leq \sqrt{Var(X_1)Var(X_2)}$$

לכן:

$$\begin{aligned} Var(T) &= 0.25Var(c^TY) + 0.25Var(d^TY) + 0.5 \cdot cov(c^TY, d^TY) \\ &\leq 0.25Var(c^TY) + 0.25Var(d^TY) + 0.5 \cdot var(c^TY) = var(c^TY) \end{aligned}$$

כיוון שהראינו ש-c^TY הוא האומד הלינארי חסר ההטיה בעל השונות הנמוכה ביותר, מתחייב שאי השיוויון מתקיים בשיוויון. ובפרט קיבלנו כי $cov(c^TY, d^TY) = var(c^TY)$

כעת נסתכל על ההפרש $c^TY - d^TY$. זהו משתנה מקרי (כי Y מקרי) בעל תוחלת 0 (כי שני האומדים חסרי הטייה). נקבל:

$$var(c^TY - d^TY) = (c-d)^TVar(Y)(c-d) = \sigma^2 \left\| c-d \right\|^2$$

מצד שני,

$$var(c^TY - d^TY) = var(c^TY) + var(d^TY) - 2cov(c^TY, d^TY) = 2var(c^TY) - 2var(c^TY) = 0$$

לכן מתכונות של נורמה וחיוביות השונות של Y, בהכרח מתקיים $c = d$.

שאלה

3 רגולריזציות לרגרסיית ריבועים פחותים (Ridge & Lasso)

Ridge

$$\hat{\beta}_{\text{Ridge}} = \arg\min_{\beta} \|X\beta - y\|_2^2 + \lambda_{\text{Ridge}}\|\beta\|_2^2$$

הניחו מודל לינארי $Y = X\beta + \varepsilon$, עבור $E[\varepsilon] = 0$, $Cov(\varepsilon) = \sigma^2 I$.

אומד רכס (ridge) עבור $c>0$ כלשהו מוגדר בצורה הבאה:

$$\hat{\beta}_{ridge} = (X'X + cI)^{-1}X'Y$$

נסמן ב $X'X = UDU'$ את הפירוק הספקטרלי של $X'X$, כאשר $D = diag(d_1, ..., d_p)$ הם הערכים העצמיים.

א) מצאו ביטוי עבור התוחלת של $\hat{\beta}_{ridge}$.

ב) בטאו את השונות המשותפת של $\hat{\beta}_{ridge}$ בעזרת $U,\ d_1, ..., d_p, c, \sigma^2$.

ג) הוכיחו שלכל וקטור קבוע $a \in R^p$, $a \neq 0$ מתקיים $Var(a'\hat{\beta}_{ridge}) < Var(a'\hat{\beta}_{OLS})$,
והסבירו כיצד התכונה מתיישבת עם משפט גאוס-מרקוב. (כלומר משפט ה"אומד BLUE").

התפלגות רב נורמלית

נגיד שוקטור מקרי $W \in R^m$ מפולג רב-נורמלי ממימד m ניתן לכתוב את W:

$$W = \mu + AZ$$

עבור $Z \in R^d, \mu \in R^m, A \in R^{m\times d}$ כאשר $Z = (Z_1, \dots, Z_d), Z_i \sim N(0,1), iid's.$

ונסמן $W \sim N(\mu, AA^T)$.

אם המטריצה $V \coloneqq AA^T$ הפיכה אז קיימת צפיפות והיא מוגדרת על ידי:

$$f_W(\boldsymbol{w}) = (2\pi)^{-m/2}|\boldsymbol{V}|^{-1/2}\exp\left[-(\boldsymbol{w}-\boldsymbol{\mu})^\top \boldsymbol{V}^{-1}(\boldsymbol{w}-\boldsymbol{\mu})/2\right], \quad \boldsymbol{w} \in \mathbb{R}^m$$

א. הראו כי אם ל-W קיימת צפיפות, אז:

$$\mathbf{c}^\top \boldsymbol{W} \sim \mathcal{N}\left(\mathbf{c}^\top \boldsymbol{\mu}, \mathbf{c}^\top \boldsymbol{V}\mathbf{c}\right) \quad \forall \mathbf{c} \in \mathbb{R}^m \quad \Longleftrightarrow \quad \boldsymbol{W} \sim \mathcal{N}_m(\boldsymbol{\mu}, \boldsymbol{V}).$$

ב. מצאו את ההתפלגות השולית של הכניסה ה-j של W_j.

ג. הכלילו את התוצאה למציאת ההתפלגות של:

$$\text{If } \boldsymbol{C} \in \mathbb{R}^{m\times k} \text{ constant matrix then } \boldsymbol{CW} \sim \mathcal{N}_m\left(\boldsymbol{C\mu}, \boldsymbol{CVC}^T\right).$$

א. נוכיח בעבור המקרה הכללי של סעיף ג'. נתבונן בוקטור המקרי:

$$CW = C\mu + CAZ$$

נסמן ב-$\tilde{A} \coloneqq CA \in R^{m\times k}$ וכן $\tilde{\mu} \coloneqq C\mu \in R^m$ ונקבל:

$$CW = \widetilde{W} = \tilde{\mu} + \tilde{A}Z$$

בעבור אותו הוקטור Z שכניסותיו מ"מ נורמליים סטנדרטיים ב"ת ולכן רב נורמלי m מימדי. בפרט זה קורה כאשר $m = 1$ במקרה של סעיף א. כלומר- טרנספורמציה לינארית של מ"מ נורמליים, מפולגת גם היא נורמלית. התוחלת והשונות מיידיים.

מצד שני: אם לכל $c \in R$ מתקיים ש-$c^T W$ מפולג נורמלי עם הפרמטרים לעיל, אז הפונקציה יוצרת המומנטים של $c^T W$:

$$M_{CW}(s) = E\left(e^{(Cs)^T W}\right) = e^{s^T C^T \mu} e^{s^T C^T V C s}, \forall c \in R^{m\times k}$$

נסמן $Cs =: \tilde{s}$ ונקבל:

$$e^{s^T C^T \mu} e^{s^T C^T V C s} = e^{\tilde{s}^T \mu} e^{\tilde{s}^T V \tilde{s}} = M_W(\tilde{s})$$

בעבור W המפולג כמו בטענה. כיוון שפיי"מ מגדירה באופן יחיד את ההתפלגות, מתקבלת הטענה.

ב. כדי למצוא את ההתפלגות השולית של W_j נוכל לקחת $c = e_j \in R^m$. שימו לב שנקבל מכך את המשתנה W_j. לפי הטענה מהסעיף הקודם, נקבל שההתפלגות של $c^T W$ היא נורמלית חד מימדית. התוחלת של המ"מ היא $e_j^T \mu = \mu_j$ והשונות היא $e_j^T V e_j = V_{jj}$.

ג. וכן במקרה הרב מימדי:

כאשר ההוכחות מבוססות על כך שניתן להסתכל על המשתנה החדש :

$$CW = C\mu + CAZ = \tilde{\mu} + \tilde{A}Z$$

מצד שני, אם לכל $c \in R$ מתקיים ש-c^TW מפולג נורמלי עם הפרמטרים לעיל, אז הפונקציה יוצרת המומנטים של c^TW :

$$M_{c^TW}(s) = E\left(e^{Sc^TW}\right) = e^{sc^T\mu}e^{s^2c^TVc}, \forall c \in R^m$$

נסמן $sc =: \tilde{s}$ ונקבל:

$$e^{sc^T\mu}e^{s^2c^TVc} = e^{\tilde{s}^T\mu}e^{\tilde{s}^TV\tilde{s}} = M_W(\tilde{s})$$

בעבור W המפולג כמו בטענה. כיוון שפיי"מ מגדירה באופן יחיד את ההתפלגות, מתקבלת הטענה.

הערה: לא צריכים לדעת להוכיח את הטענות שמשתמשות בפונקציה יוצרת מומנטים- אלא רק צד אחד. ראו הקלטה.

בשאלה זו נראה שעבור התפלגות דו-נורמלית, אי-תלות שקולה לחוסר קורלציה. עבור

$$X = (X_1, X_2)^T \sim N(\mu, \Sigma)$$

א. הראו שאם X_1 ו- X_2 בלתי תלויים אזי $Cov(X_1, X_2) = 0$, כלומר Σ מטריצה אלכסונית.

ב. הראו שאם Σ מטריצה אלכסונית אזי X_1 ו- X_2 בלתי תלויים.

פתרון:

א. זה נכון לכל זוג מ"ם ב"ת:

$$E(X_1X_2) = \int\int x_1x_2f_{X_1,X_2}(x_1,x_2)dx_1dx_2 = \int\int x_1x_2f_{X_1}(x1)\cdot f_{X_2}(x_2)dx_1dx_2$$
$$= \int x_2f_{X_2}(x_2)dx_2\int x_1f_{X_1}(x_1)dx_1 = E(X_2)E(X_1)$$

ב. אם Σ אלכסונית (ונניח כי בעלי צפיפות):

$$f_{\vec{X}}(\vec{x}) = (2\pi)^{-\frac{n}{2}}\cdot|\Sigma|^{-\frac{1}{2}}\cdot e^{-\frac{1}{2}\overrightarrow{(x-\mu)}^T\Sigma^{-1}\overrightarrow{(x-\mu)}} = (2\pi)^{\frac{n}{2}}\cdot\left(\Pi_{i=1}^n\Sigma_{ii}\right)^{-\frac{1}{2}}\cdot e^{-\frac{1}{2}\Sigma_{i=1}^n(x_i-\mu_i)^T\Sigma_{ii}^{-1}(x_i-\mu_i)}$$
$$= \Pi_{i=1}^n(2\pi\Sigma_{ii})^{-\frac{1}{2}}e^{\frac{(x_i-\mu_i)^2}{\Sigma_{ii}}}$$

כלומר הצפיפות המשותפת מתפרקת למכפלת צפיפויות רב נורמליות חד מימדיות ולכן ב"ת.

שאלה- דוגמה מספרית

נתונים 4 משתנים מקריים Y_1, Y_2, Y_3, Y_4 בעלי תוחלת 0 ושונות משותפת המקיימת $Cov(Y_i, Y_j) = min(i,j)$ לכל $1 \le i, j \le 4$. ההתפלגות המשותפת של 4 המשתנים היא רב נורמלית. נגדיר $S = Y_1 + Y_2 + Y_3 + Y_4$.

(א) זהו באופן מפורש את ההתפלגות של S.

פתרון:

נמצא את ההתפלגות של $S = \sum_{i=1}^{4} Y_i$

$$S = \sum_{i=1}^{4} Y_i = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix}$$

ונסמן $1_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}$. נתון כי $\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{pmatrix}\right)$ ולכן

$$S \sim N\left(\begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}\right)$$

כלומר,

$$S \sim N(0, 30)$$

התפלגויות קשורות להתפלגות הנורמלית ומבוא להסקה סטטיסטית

Definition (Chi-square distribution). If $Z_1, Z_2, \ldots, Z_k \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, then the distribution of

$$Q = \sum_{j=1}^{k} Z_j^2$$

is called the Chi-square distribution with k degrees of freedom, and we denote $Q \sim \chi_k^2$ (in R: `pchisq()`, `qchisg()`, `rchisq()`).

Definition 5 (t-distribution)**.** . If $Z \sim \mathcal{N}(0,1), V \sim \chi_k^2$, are independent random variables, then the distribution of

$$T = \frac{Z}{\sqrt{V/k}}$$

is called the t-distribution with k degrees of freedom, and we denote $T \sim t_k$ (in R: `pt ()`, `qt( )`, `rt( )`).

Definition 6 (F distribution)**.** If $V_1 \sim \chi_{k_1}^2, V_2 \sim \chi_{k_2}^2$, are independent random variables, the distribution of

$$F = \frac{V_1/k_1}{V_2/k_2}$$

is called the F-*distribution with* k_1 *and* k_2 (numerator and denominator, respectively) *degrees of freedom*, and we denote $F \sim \mathrm{F}_{k_1,k_2}$.

שאלה

חימום: הראו שאם $T \sim t_k$ אז $T^2 \sim F_{1,k}$.

א. הראו כי $\widehat{\sigma^2} = \frac{||e||^2}{n-p-1}$ הוא אומד חסר הטיה ל-σ^2.

ב. הראו שאם P מטריצת הטלה מדרגה r, ואם $Z \sim N(0, I)$ אז מתקיים:

$$||PZ||^2 \sim \chi_r^2$$

ג. הסיקו כי תחת הנחת הנורמליות מתקיים: $||e||^2 = ||Y - \hat{Y}||^2 \sim \sigma^2\chi_{n-p-1}^2$.

ד. יהי $Z \sim N_q(\mu, \Sigma)$ כלומר $Z \in \mathbb{R}^q$ מתפלג רב-נורמלית עם וקטור תוחלות μ ומטריצת שונויות משותפות Σ. הראו כי

$$(Z-\mu)^T\Sigma^{-1}(Z-\mu) \sim \chi_q^2$$

ה. יהי $Y \sim N_n(X\beta, \sigma^2 I_q)$.

ו. הראו כי המשתנים $||\hat{Y}||^2$ ו-$||e||^2$ ב"ת.

ו. אפיינו את ההתפלגות של

$$\frac{||\hat{Y} - X\beta||^2}{||e||^2}$$

ז. נסמן ב-$s_j = (X^TX)_{jj}^{-1}$. אפיינו את ההתפלגות של $(\hat{\beta}_j - \beta_j)/\sqrt{(\widehat{\sigma^2} s_j)}$.

פתרון:

אם $T \sim t_k$ אז $T = \frac{Z}{\sqrt{V/k}}$ בעבור $Z \sim N(0,1)$ ו-$V \sim \chi_k^2$ ב"ת. מכאן ש- $T^2 = \frac{Z^2}{V/k}$. המונה מפולג χ_1^2 ולכן גם בחלוקה ב-1 והמכנה הוא χ_k^2 מחולק בדרגות החופש שלו. זו בדיוק ההגדרה של התפלגות $F_{1,k}$.

א. ראשית נסתכל על:

$$E(e) = E[(I-P_X)Y] = (I-P_X)X\beta = 0$$

בתרגול 3 ראינו כי בעבור וקטור מקרי עם תוחלת 0 מתקיים:

$$E\left(||e||^2\right) = tr\left(Var(e)\right)$$

צד ימין שווה במקרה הזה:

$$tr(Var(I-P_X)Y) = tr(\sigma^2(I-P_X)(I-P_X)^T) = \sigma^2 tr\left((I-P_X)\right)$$

כאשר השיוון האחרון נכון כי זו מטריצת הטלה ולכן סימטרית ואיידמפוטנטית. כעת נשתמש בתכונה נוספת של מטריצת הטלה מתרגיל 2 שאומרת ש-$tr(I-P_X) = rank(I-P_X)$, וכן בעבודה ש-$I-P_X$ מטילה למרחב ממימד $n-(p+1)$ ונקבל כי $\sigma^2(n-p-1) = E\left(||e||^2\right)$. נחלק בגורם המתאים ונקבל את הטענה.

ב.

$$||PZ||^2 = Z^TP^TPZ = Z^TPZ = Z^TU\Lambda U^TZ$$

נזכור כי הע"ע של מטריצת הטלה הם 0 ו-1 בריבוי $n-r$ ו-r בהתאמה. נסתכל על המ"מ $\tilde{Z} \coloneqq U^TZ \in R^n$:

זו קומבינציה לינארית של וקטור רב נורמלי ולכן גם הוא מפולג רב נורמלית. תוחלתו: $U^TE(Z) = 0$ ומטריצת השונויות שלו:

$$Var(U^TZ) = U^TIU = I$$

כלומר גם הוא וקטור רב נורמלי סטנדרטי. נציב בפירוק הספקטרלי ונקבל:

$$Z^T U \Lambda U^T Z = \sum_{i=1}^{n} \tilde{Z}_i^2 \lambda_i = \sum_{i=1}^{r} \tilde{Z_i}^2$$

וזהו סכום של r מ״מ ב״ת ש״ה נורמליים סטנדרטיים ולכן מפולגים χ_r^2.

ג.

$$||e||^2 = ||Y - \hat{Y}||^2 = ||(I - P_X)Y||^2 = ||(I - P_X)X\beta + (I - P_X)\epsilon||^2 = ||(I - P_X)\epsilon||^2$$

כעת כיוון ש-$\epsilon \sim N(0, \sigma^2 I)$ אז ניתן לכתוב את $\epsilon = \sigma Z$ בעבור $Z \sim N(0, I)$ ומכאן:

$$||(I - P_X)\epsilon||^2 = \sigma^2 ||(I - P_X)Z||^2 = \sigma^2 \chi_{n-p-1}^2$$

כאשר השיוויון האחרון נובע מסעיף ב׳.

ד.

כיוון ש-Σ סימטרית חיובית, נוכל לכתוב $||\Sigma^{-\frac{1}{2}}(Z - \mu)||^2 = (Z - \mu)^T \Sigma^{-1} (Z - \mu)$. שימו לב שזוהי נורמה בריבוע של וקטור מקרי נורמלי סטנדרטי, ולכן מפולגת χ_q^2.

ה. כפי שראינו בתרגול ובשיעור, תחת הנחת הנורמליות המשתנים:

$$\hat{Y} = X\hat{\beta} = P_X Y \sim N(X\beta, \sigma^2 X(X^T X)^{-1} X^T) = N(X\beta, \sigma^2 P_X)$$
$$e = (I - P_X)Y \sim N\left(0, \sigma^2 (I - P_X)\right)$$

לכן אם המשתנים ב״מ הם גם יהיו ב״ת:

$$Cov\left(\hat{Y}, \hat{e}\right) = Cov(P_X Y, (I - P_X)Y) = P_X \cdot Var(Y)(I - P_X)^T = P_X \sigma^2 (I - P_X) = 0$$

הפונקצייה $f = g: R^n \to R$ היא:

$$f(v) = \sum_{i=1}^{n} v_i^2$$

ו.

$$\hat{Y} - X\beta = P_X Y - X\beta = P_X X\beta + P_X \epsilon - X\beta = X\beta + P_X \epsilon - X\beta = P_X \epsilon$$

אזי

$$||\hat{Y} - X\beta||^2 = ||P_X Y - X\beta||^2 = ||P_X \epsilon||^2$$

תוך שימוש בסעיף ב׳, זה מתפלג $\sigma^2 \chi_{p+1}^2$.

כמו כן, הוכחנו כבר בשיעור (ושוב- תוך שימוש בסעיף ב׳) כי $||e||^2 \sim \sigma^2 \chi_{n-p-1}^2$. לכן הסטטיסטי שבשאלה שקול ל:

$$T := \frac{\frac{||\hat{Y} - X\beta||^2}{\sigma^2}}{\frac{||e||^2}{\sigma^2}} = \frac{V_1}{V_2}$$

עבור $V_1 \sim \sigma^2 \chi_{p+1}^2$ ואילו $V_2 \sim \sigma^2 \chi_{n-p-1}^2$ ושני המשתנים בלתי תלויים (מהסעיף הקודם). לאחר חלוקה של המונה ב- $p + 1$ ושל המכנה ב-$n - p - 1$ נקבל בדיוק את ההגדרה של משתנה מקרי המפולג $F_{p+1, n-p-1}$. כלומר: $\frac{n-p-1}{p+1} T \sim F_{p+1, n-p-1} \Leftrightarrow T \sim \frac{p+1}{n-p-1} F_{p+1, n-p-1}$

ז. תחת המודל הנורמלי מתקיים :

$$\hat{\beta}_j \sim N\left(\beta, \sigma^2 s_j\right)$$
$$(\hat{\beta}_j - \beta_j)/\sqrt{(\widehat{\sigma^2} s_j)} = [(\hat{\beta}_j - \beta_j)/\sigma \cdot \sqrt{s}_j]/\sqrt{(\widehat{\sigma^2}}]/\sigma$$

המונה מפולג נורמלי סטנדרטי בעוד שבמכנה ישנו שורש על פני

$\frac{\chi^2_{n-p-1}}{n-p-1} = \widehat{\sigma^2}\ /\sigma^2 = \frac{||e||^2}{\sigma^2 \cdot n-p-1}$ כלומר חי בריבוע המחולק בדרגות החופש שלו. מכאן שההתפלגות היא התפלגות t_{n-p-1}.

שאלה

א. מצאו את ההתפלגות של $\hat{\beta}_j$ ובנו רווח סמך ל-β_j כאשר σ^2 ידוע. חזרו על כך כאשר σ^2 איננו ידוע.

פתרון :

א. ע״פ השאלה הקודמת, ההתפלגות נורמלית עם הפרמטרים שהוזכרו. מכאן שרווח סמך ל-β_j :

$$\hat{\beta}_j \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{(X^T X)_{jj}}$$

וכאשר σ^2 איננו ידוע:

$$\hat{\beta}_j \pm t_{n-p-1,1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{(X^T X)_{jj}}$$

שאלה

```
Call:
lm(formula = TotalMurderRate ~ Population + Density + Ownership,
    data = guns)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2783 -1.3871 -0.3493  0.9758  5.9019

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.84193    1.02830   1.791   0.0797 .
Population   0.09146       A      2.238   0.0300 *
Density      1.91414    0.20841   9.184 4.63e-12 ***
Ownership       B       2.23611   1.258   0.2146
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.816 on 47 degrees of freedom
Multiple R-squared:  0.6835,    Adjusted R-squared:  C
F-statistic: 33.83 on D and 47 DF,  p-value: 8.432e-12
```

	intercept	Population	Density	Ownership
intercept	0.3205745	-0.0073787	-0.0360011	-0.6548395
Population	-0.0073787	0.0005063	0.0006838	0.0109158
Density	-0.0360011	0.0006838	0.0131685	0.0721368
Ownership	-0.6548395	0.0109158	0.0721368	1.5159148

3. **(10 נק')** נניח, היפותטית, שהיתה מדינה 52 שלא כלולה בקובץ הנתונים המקורי, ועבורה המשתנים המסבירים מקבלים את הערכים: Population = 9.53, density = 0.2, ownership = 0.63. מצאו אומד ליניארי חסר-הטייה בעל שונות מינימלית עבור תוחלת שיעור מקרי הרצח במדינה עם המאפיינים האלה (שהמודל הליניארי שמתאר את הקשר בין המשתנים עבור קובץ הנתונים המקוריים, מתאים גם לתצפית החדשה).
4. **(10 נק')** בנו ר"ס ברמת ביטחון 90% עבור הפרמטר שנאמד בסעיף הקודם. יש לציין אילו הנחות נדרשות על השגיאות ϵ_i כדי שרווח-הסמך אכן יהיה תקף.

4. עבור צירוף לינארי כללי, $\theta = a^T\beta$, האומד הרגיל הוא $\hat{\theta} = a^T\hat{\beta}$, והשונות: $Var(\hat{\theta}) = Var(a^T\hat{\beta}) = \sigma^2 a^T[(X^TX)^{-1}]a$.

במקרה שלנו $a = (1, 9.53, 0.2, 0.63)^T$, והמטריצה $(X^TX)^{-1}$ נתונה.

נחשב:

$$a^T[(X^TX)^{-1}]a = \sum_{i=1}^{4}\sum_{j=1}^{4} [(X^TX)^{-1}]_{ij}\, a_i a_j =$$

$$= 0.32 - .0073\cdot 9.53 - \overset{0.036}{.36}\cdot 0.2 - .654\cdot .63 + \qquad (i=1)$$

$$+\ 9.53\,(-.0073 + .0005\cdot 9.53 + .0006\cdot 0.2 + .011\cdot 0.63) + \qquad (i=2)$$

$$+\ 0.2\,(-\overset{0.036}{.36} + .0006\cdot 9.53 + .0131\cdot 0.2 + 0.72\cdot 0.63) + \qquad (i=3)$$

$$+\ 0.63\,(-.654 + .011\cdot 9.53 + .072\cdot 0.2 + 1.516\cdot 0.63) \qquad (i=4)$$

$$= \boxed{0.14}$$

תזכורת:

	1	9.53	0.2	0.63
1	0.32	-.0073	-.036	-.654
9.53	-.0073	.0005	.0006	$\overset{+}{-}$.011
0.2	-.036	.0006	.0131	.072
0.63	-.654	.011	.072	1.516

לסיכום:

$$CI = \hat{\theta} \pm \hat{\sigma}\sqrt{a^T(X^TX)^{-1}a}\cdot t_{n-p-1;\,1-\alpha/2}$$

$$= 4.865 \pm 1.816 * \sqrt{0.14}\cdot 1.678$$

$$= 4.865 \pm 1.14 = \boxed{(3.72,\ 6)}$$

הר"ס תקף במובן נורמלי של שגיאות ε_i (בנוסף להנחות המודל הלינארי הכללי: ליניאריות; שגיאות חסרות-מתאם וחסרות הטיה ביניהן).

שאלות נוספות- לקראת בוחן האמצע:

1. נניח כי בידינו נתונים על הגיל (X_1), המשקל (X_2), ועל לחץ הדם (Y) של n מטופלים. כמו כן, הניחו כי ידוע שהמשקל לא משפיע כלל על לחץ הדם בהינתן הגיל. נסמן ב-M את המרחב הנפרש על ידי עמודות המטריצה X. נסמן ב-L את המרחב הנפרש על ידי עמודות המטריצה X כשהשמטנו ממנה את עמודת המשקלים.

האם הטענות הבאות נכונות או לא נכונות? נמקו.

א. $P_M Y$ הוא אומד מוטה ל-$E(Y|X)$.

ב. $E||e_M||^2 > E||e_L||^2$.

א. הטענה לא נכונה. אינטואיטיבית- זה כמו להגיד $\beta_2 = 0$. כיוון שתחת הנחות המודל הלינארי אומד ריבועים פחותים הוא אומד חסר הטיה, נקבל $E(\hat{\beta}_2) = \beta_2 = 0$. באופן פורמלי יותר:

$$E(P_M Y) = P_M E(Y) = P_M X_L \beta = X_L \beta$$

כאשר המעבר האחרון נכון כיוון ש- $L \subseteq M$.

ב. הטענה לא נכונה.

$$\|e_L\|^2 = \|(I - P_L)Y\|^2 = Y^T U \Lambda U^T Y = \tilde{Y}^T \Lambda \tilde{Y} = \sum_{i=1}^{n-p} \tilde{Y}_i^2 \geq \sum_{i=1}^{n-p-1} \tilde{Y}_i^2 = \|(I - P_M)Y\|^2$$
$$= \|e_M\|^2.$$

2. הוכיחו כי

10. If $\boldsymbol{Z}$ is another $n \times m$ matrix s.t. $\mathrm{Im}(\boldsymbol{Z}) = \mathrm{Im}(\boldsymbol{X})$, then $\boldsymbol{P_Z} = \boldsymbol{P_X}$.

3.

Proposition 6. *We have*

1. $\boldsymbol{I} - \boldsymbol{P_X} = P_{Im(\boldsymbol{X})^\perp}$
2. *if L and M are two subspaces of $\mathbb{R}^n$ with $L \subseteq M$, then $\boldsymbol{P}_M - \boldsymbol{P}_L = \boldsymbol{P}_{M \cap L^\perp}$*

ראו פתרון ברשימות השיעור.
עבור שאלה 3 סעיף 2:

$$\forall v \in R^n:$$

$$w := (P_M - P_L)v = P_M v - P_L v \Rightarrow w \in M$$

כי $P_M v \in M$ וכן $P_L v \in L \subseteq M$ ולכן גם ההפרש (תת מרחב).
מצד שני, ניקח $u \in L$ אז:

$$u^T w = u^T P_M v - u^T P_L v = u^T P_M^T v - u^T P_L^T v = (P_M u)^T v - (P_L u)^T v = u^T v - u^T v = 0$$

ולכן $w \in L^\perp$.

נותר להראות כי זו מטריצה סימטרית ואיידמפוטנטית:

$$(P_M - P_L)\text{^}T =$$
$$P_M^T - P_L^T = P_M - P_L$$

וכן:

$$(P_M - P_L)(P_M - P_L) = P_M^2 - P_L P_M - P_M P_L + P_L^2 = P_M - P_L$$

כאשר השתמשנו בכך שאם $L \subseteq M$ אז:

$$P_L P_M = P_M P_L = P_L$$