52571 - רגרסיה ומודלים סטטיסטיים - תשפ״ה סמסטר ב׳

בוחן אמצע

הנחיות כלליות

1. זמן הבחינה: שעתיים אקדמיות
2. מותר להשתמש בארבעה דפי נוסחאות דו-צדדיים (או 8 חד-צדדיים), אסור כל אמצעי דיגיטלי
3. מותר לצטט כל תוצאה שראינו בכיתה, אלא אם השאלה מבקשת במפורש לפתח או להוכיח את התוצאה

שאלה 1

הניחו שהמטריצה $X \in R^{n\times(p+1)}$ בעלת דרגת עמודות מלאה. עבור כל אחת מהטענות הבאות, סמנו אם היא: נכונה רק אם מתקיימות הנחות המודל הלינארי (1), נכונה רק אם מתקיימות הנחות המודל הלינארי הנורמלי (2), או נכונה ללא קשר להנחות המודל הסטטיסטי (3). **נמקו במשפט אחד**.

א. $\hat{\sigma}^2 = \dfrac{SSE}{n-p-1}$ אומד חסר-הטיה ל-σ^2

ב. $P\left(\hat{\beta}_j - t_{1-\frac{\alpha}{2},n-p-1}\sqrt{\hat{\sigma}^2\left(X^TX\right)^{-1}_{j+1,j+1}} \le \beta_j \le \hat{\beta}_j + t_{1-\frac{\alpha}{2},n-p-1}\sqrt{\hat{\sigma}^2\left(X^TX\right)^{-1}_{j+1,j+1}}\right) = 1-\alpha$.

ג. תהי $M \in \mathbb{R}^{(p+1)\times n}$ מטריצה קבועה, ונניח $\tilde{\beta} = MY$ אומד חסר-הטיה כלשהו ל-β. אז כל וקטור קבוע $v \in \mathbb{R}^{p+1}$ מתקיים $MSE\left(v^T\hat{\beta}\right) \le MSE\left(v^T\tilde{\beta}\right)$

ד. וקטור השאריות e אורתוגונלי לוקטור הערכים החזויים $\hat{Y}$ (כלומר המכפלה הפנימית ביניהם היא 0)

ה. הוקטורים המקריים $e, \hat{\beta}$ הם ב״ת סטטיסטית

שאלה 2

נסמן

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \qquad 1_n = (1,...,1)^\top \in \mathbb{R}^n, \qquad M = \mathrm{Im}(X), \qquad L = \mathrm{Im}(1_n)$$

הערה: כרגיל, העמודה הראשונה של $X \in R^{n\times(p+1)}$ היא 1_n

א. הוכיחו שמתקיים $\bar{Y}1_n = P_LY$, כאשר P_L זאת מטריצת ההיטל על L (**הערה**: $\bar{Y}1_n = (\bar{Y},\ldots,\bar{Y})^\top$).

ב. הוכיחו $(Y-\hat{Y}) \perp (\hat{Y}-\bar{Y}1_n)$, כלומר שהוקטורים האלה אורתוגונלים.

רמז: כתבו את $\hat{Y}$ בתור היטל של Y על המרחב המתאים, ואז היעזרו בסעיף א׳ ובתכונות שראינו לגבי מטריצות הטלה.

ג. הוכיחו שמתקיים $\|Y-\bar{Y}1_n\|^2 = \|Y-\hat{Y}\|^2 + \|\hat{Y}-\bar{Y}1_n\|^2$

רמז: תחילה החסירו והוסיפו $\hat{Y}$ בתוך הנורמה באגף שמאל כדי לקבל

$$\|Y-\bar{Y}1_n\|^2 = \|(Y-\hat{Y}) + (\hat{Y}-\bar{Y}1_n)\|^2$$

ואז היעזרו בסעיפים קודמים + פיתגורס.

הערה: שימו לב שאת הטענה המקורית אפשר לכתוב באופן שקול

$$\underbrace{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}_{SST} = \underbrace{\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{Y}\right)^2}_{SSR} + \underbrace{\sum_{i=1}^{n}\left(Y_i-\hat{Y}_i\right)^2}_{SSE}$$

לכן הפירוק של סכומי הריבועים, שראינו בכיתה שמתקיים ברגרסיה פשוטה ($p=1$), למעשה מתקיים באופן כללי יותר גם ברגרסיה מרובה (p כללי).

ד. תחת הנחות המודל הליניארי הראו שמתקיים $\mathbb{E}\|\hat{Y}-\bar{Y}1_n\|^2 = \sigma^2 p + \|QX\beta\|^2$ עבור מטריצת הטלה Q מתאימה.

רמז: בעזרת סעיפים קודמים, כתבו $\hat{Y}-\bar{Y}1_n = QY$ עבור מטריצת הטלה Q מתאימה, ואז חשבו את $\mathbb{E}\|QY\|^2$ בעזרת הזהות שראינו בכיתה, שאומרת שעבור וקטור מקרי Z כלשהו, מתקיים $\mathbb{E}(\|Z\|^2) = \text{cov}(Z) + \mu_Z^\top \mu_Z$ כאשר $\mu_Z := \mathbb{E}[Z]$. לבסוף השתמשו בעובדה $\text{tr}(Q) = \text{dim}(Q)$ עבור מטריצת הטלה Q.

ה. תחת הנחות המודל הליניארי הנורמלי, מצאו את ההתפלגות של $SSE = \|Y-\hat{Y}\|^2 = \|e\|^2$.

בהצלחה!