## רגרסיה ומודלים סטטיסטיים- פתרון בוחן האמצע- מועד א'

### שאלה 1

א. (1) נכונה רק אם מתקיימות הנחות המודל הלינארי. ראינו כי:

$$E(SSE) = E||e||^2 = E||(I - P_X)Y||^2 = E||(I - P_X)\epsilon||^2 = tr(cov(I - P_X)\epsilon)$$
$$= \sigma^2(n - p - 1)$$

כאשר השוויון השלישי נכון תחת $Y = X\beta + \epsilon$, הרביעי כי $E(\epsilon) = 0$ והחמישי כי $\epsilon \sim (0, \sigma^2 I)$.

ב. (2) נכונה רק אם מתקיימות הנחות המודל הלינארי הנורמלי. תחת הנחות אלו:

$$\frac{\hat{\beta}_j - \beta_j}{\widehat{\sigma^2}\ (X^T X)^{-1})_{(j+1)(j+1)}} \sim t_{n-p-1}$$

ולכן זהו רווח סמך ל- $\beta_j$ ברמת סמך $1 - \alpha$.

ג. (1) נכונה רק אם מתקיימות הנחות המודל הלינארי. זוהי תוצאה ישירה של משפט גאוס מרקוב שקובע שתחת הנחות המודל הלינארי אומד OLS הוא האומד הלינארי חסר ההטיה הטוב ביותר.

ד. (3) נכונה ללא קשר להנחות המודל הסטטיסטי- $\hat{Y} \in IM(X), \hat{e} \in IM(X)^{\perp}$.

ה. (2) נכונה רק אם מתקיימות הנחות המודל הלינארי הנורמלי. הוקטורים בלתי מתואמים תחת הנחות המודל הלינארי- ותחת הנורמליות זה גורר שגם ב"ת (אחרת לא בהכרח).

### שאלה 2

א. אפשר לחשב ישירות או לשים לב שברגרסיה עם חותך בלבד:

$$\widehat{\beta_L} = argmin_{\beta} \sum_{i=1}^{n} (Y_i - \beta)^2$$

ובהגדרה זהו הממוצע $\hat{\beta}_L = \bar{Y}$. לכן $P_L Y = X_L \hat{\beta}_L = 1_n \bar{Y}$.

ב.

$$\left(Y - \hat{Y}\right) = (I - P_M)Y \Rightarrow \left(Y - \hat{Y}\right)^T \left(\hat{Y} - \bar{Y}1_n\right) = Y^T(I - P_M)(P_M - P_L)Y$$
$$= Y^T(I - P_M)P_{M\cap L^{\perp}}Y = 0$$

כאשר השיוויון השלישי נכון כיוון שעבור $L \subseteq M$ תתי מרחבים, $(P_M - P_L) = P_{M\cap L^{\perp}}$ והאחרון נכון כיוון ש- $P_{M\cap L^{\perp}}Y \in M$ ו- $(I - P_M)$ מטילה למשלים האורתוגונלי.

ג.

$$||Y - 1_n\bar{Y}||^2 = \left|\left|Y - \hat{Y} + \hat{Y} - 1_n\bar{Y}\right|\right|^2 = \left|\left|Y - \hat{Y}\right|\right|^2 + \underbrace{2\left(Y - \hat{Y}\right)^T\left(\hat{Y} - \bar{Y}1_n\right)}_{=0} + \left|\left|\hat{Y} - 1_n\bar{Y}\right|\right|^2$$
$$= \left|\left|Y - \hat{Y}\right|\right|^2 + \left|\left|\hat{Y} - 1_n\bar{Y}\right|\right|^2$$

ד.

$$\left|\left|\hat{Y} - 1_n\bar{Y}\right|\right|^2 = \left|\left|P_{M\cap L^\perp}Y\right|\right|^2 = \left|\left|P_{M\cap L^\perp}(X\beta + \epsilon)\right|\right|^2 \Rightarrow$$

$$E\left|\left|\hat{Y} - 1_n\bar{Y}\right|\right|^2 = tr\left(cov(P_{M\cap L^\perp}\epsilon)\right) + E(X\beta + \epsilon)^T P_{M\cap L^\perp} E(X\beta + \epsilon)$$

$$= \sigma^2 tr(P_{M\cap L^\perp}) + ||P_{M\cap L^\perp}X\beta||^2 = \sigma^2 p + ||P_{M\cap L^\perp}X\beta||^2$$

כאשר השיוון לפני האחרון נבע מכך ש-$\epsilon \sim (0, \sigma^2 I)$, וש- $P_{M\cap L^\perp}$ סימטרית ואיידמפוטנטית, והאחרון מכך ש- $tr(P_{M\cap L^\perp}) = rank(P_{M\cap L^\perp}) = \dim(M \cap L^\perp) = p + 1 - 1 = p$.

ה. הוכחנו זאת כמה פעמים בכיתה:

עבור מטריצת הטלה $Q$ שדרגתה $k$ ו-$Z$ וקטור נורמלי סטנדרטי מתקיים $\left|\left|QZ\right|\right|^2 \sim \chi_k^2$. כעת נסמן:

$$Q = (I - P_M)$$

$$Z = \frac{1}{\sigma}\epsilon$$

אז:

$$\left|\left|QZ\right|\right|^2 \sim \chi_{n-p-1}^2 \Rightarrow \left|\left|e\right|\right|^2 = \left|\left|QY\right|\right|^2 = \left|\left|QX\beta + Q\epsilon\right|\right|^2 = \left|\left|Q\epsilon\right|\right|^2 = \left|\left|Q\sigma Z\right|\right|^2$$

$$= \sigma^2\left|\left|QZ\right|\right|^2 \sim \sigma^2 \chi_{n-p-1}^2$$