

## מודלים סטטיסטיים ויישומיהם 52518 תשע"ח – פתרון תרגיל 1

1. כתבו קוד R לביצוע מבחן F של רגסיה. עבור הנתונים בקובץ המצורף `targil1.xls`, השתמשו בתוכנית שלכם כדי לבדוק את השערת האפס שהמתקדמים של המשתנים  $x_1, \dots, x_8$  שווים כולם לאפס.

קוד R:

```

FTest <- function(y,X,Hcols) {
  X0 <- X[,-Hcols]
  reg.full <- summary(lm(y ~ X))
  reg.0 <- summary(lm(y ~ X0))
  SSE <- reg.full$df[2] * (reg.full$sigma)^2
  SSE0 <- reg.0$df[2] * (reg.0$sigma)^2
  Fstat <- (((SSE0 - SSE) / (length(Hcols))) / ((SSE) / (reg.full$df[2])))
  return(pf(Fstat, length(Hcols), reg.full$df[2], lower.tail = F))
}

library(readxl)
D <- read_excel('targil1.xlsx')
X <- as.matrix(D[,2:14])
y <- D[, "Y"]

FTest(y,as.matrix(X),c(8:13))

```

הפלט המתכבר הוא  $3.679 \cdot 10^{-46}$ , שכן בהינתן השערת אלטרנטיבית פשוטה נדחה את  $H_0$  (ב翦 עט) בכל רמת מובהקות.

2. בקובץ המצורף `fertilizers.csv` מצויים נתונים תנובות של שדות (בעל אוthon גודל) שקיבלו דשן שונים. על מנת להשוות בין סוג הדשן, מבקשים לערוך מבחן F.
- הגדירו מהי השערת האפס ומה השערת האלטרנטיבית במקרה זה.
  - חשבו את ממוצעי הקבוצות השונות, הגדירו את המטריצה W וחשבו את סטטיסטי המבחן F.
  - בהתאם לשערות אותן הגדרתם בסעיף א' ולסטטיסטי המבחן אותו קיבלתם בסעיף ב', הכריעו האם לדוחות את השערת האפס או לא. הסבירו את צעדיכם.
- א. השערת האפס היא כי לסוג הדשן אין השפעה על התנובה הממוצעת, כלומר  $\bar{y}_1 = \bar{y}_2 = \dots = \bar{y}_I$ . בהתאם, השערת האלטרנטיבית תהיה כי לסוג הדשן ישנה השפעה על התנובה הממוצעת.

ב.

$$I = 4, N = 24, n_1 = n_2 = n_3 = n_4 = 6, \\ \bar{y}_1 = 21, \bar{y}_2 = 28.6, \bar{y}_3 = 25.867, \bar{y}_4 = 29.2, \bar{y} = 26.167$$

$$W = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \vdots & \vdots & & \\ 1 & 0 & & \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 1 & 0 & & \\ 0 & 1 & & \\ \vdots & & 1 & 0 \\ & & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\|\hat{y} - \hat{y}^{(0)}\|^2 = \sum_{i=1}^I n_i (\bar{y}_i - \bar{y})^2 = 6(26.694 + 5.921 + 0.09 + 9.201) = 251.44$$

$$\|y - \hat{y}\|^2 = \sum_{i=1}^I \sum_{j=1}^{n_i} (\bar{y}_{ij} - \bar{y}_{i.})^2 = 61.033$$

$$F = \frac{\frac{\|\hat{y} - \hat{y}^{(0)}\|^2}{I-1}}{\frac{\|y - \hat{y}\|^2}{N-I}} = \frac{\frac{251.44}{3}}{\frac{61.033}{20}} = 27.465$$

ג. המבנים של סטטיסטי המבחן הם  $3-20$  בהתאם לבן לבדוק מול התפלגות  $\mathcal{F}_{3,20}$ . ה-value-k המתקבל עבור סטטיסטי זה הוא  $2.712 \cdot 10^{-7}$  לבן בהגנת אלטרנטיבת פשוטה, נדחה את השערת האפס כמעט בכל רמת מובהקות.

3. הרואו כי בניתוח שונות חד-כונני עם  $2 = I$ , הסטטיסטי  $F$  שקול לרכיבי הסטטיסטי  $t$  (באשר מדובר במבחן  $t$  להפרש תוחלות עבור שתי קבוצות בלתי-תלוויות).

$$\text{ראינו בכחיה כי בניתוח שונות חד-כונני מתקיים } F = \frac{\frac{\|\hat{y} - \hat{y}^{(0)}\|^2}{I-1}}{\frac{\|y - \hat{y}\|^2}{N-I}}. \text{ עבור } 2 = I \text{ קיבל:}$$

$$F = \frac{\frac{\|\hat{y} - \hat{y}^{(0)}\|^2}{I-1}}{\frac{\|y - \hat{y}\|^2}{N-I}} = \frac{\frac{\|\hat{y} - \hat{y}^{(0)}\|^2}{1}}{\frac{\|y - \hat{y}\|^2}{N-2}} = \frac{\frac{\|\hat{y} - \hat{y}^{(0)}\|^2}{\sigma^2}}{\frac{\|y - \hat{y}\|^2}{\sigma^2(N-2)}} = \frac{\frac{\|\hat{y} - \hat{y}^{(0)}\|^2}{\sigma^2}}{\frac{\|y - \hat{y}\|^2(N-2)}{\sigma^2(N-2)^2}} =$$

המונה מתפלג  $\chi_1^2$  והמכנה מתפלג  $\chi_{N-2}^2$  כולם סה"כ נקבעים בי הביטוי מתפלג  $\chi_{N-2}^2 / \chi_{N-2}^2$ , זה ריבוע של  $\sqrt{\frac{\chi_{N-2}^2}{(N-2)^2}}$  שזו בדיקת התפלגות  $t_{N-2}$ .

4. נתיחס לניתוח שונות חד-כונני. בכיתה הוציא המודל בצורה  $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ , באשר:

$$\mu = \frac{1}{I} \sum_{i=1}^I \mu_i, \quad \alpha_i = \mu_i - \mu$$

במספרים מסוימים מוגדר  $\mu$  באופן הבא:

$$\mu = \frac{1}{N} \sum_{i=1}^I n_i \mu_i$$

נניח כי  $I = 3, n_1 = 9, n_2 = 7, n_3 = 5$ . עם ההגדלה חדשה של  $\mu$  נגדיר גם

$$\beta = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \quad \eta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

א. מצאו את המטריצה  $X$  המקיים  $E[Y] = X\eta$

ב. מצאו את המטריצה  $W$  המקיים  $E[Y] = W\beta$

ג. מצאו ביטויים לכל ריבובי הוקטור  $\eta$  במנוחי ריבובי הוקטור  $\beta$  ואת המטריצה  $H$  המקיים  $\eta = H\beta$

ד. מצאו ביטויים לכל ריבובי הוקטור  $\beta$  במנוחי ריבובי הוקטור  $\eta$  ואת המטריצה  $G$  המקיים  $\eta = G\beta$

ה. וודאו ע"י חישוב ב-R כי מתקיים  $X = WG$  וכן  $W = XH$

.  
א.

$$E[y] = X \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \mu + \alpha_1 \\ \vdots (8 \text{ more}) \\ \mu + \alpha_2 \\ \vdots (6 \text{ more}) \\ \mu + \alpha_3 \\ \vdots (4 \text{ more}) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & 0 \\ \vdots (8 \text{ more}) \\ 1 & 0 & 1 \\ \vdots (6 \text{ more}) \\ 1 & -1.8 & -1.4 \\ \vdots (4 \text{ more}) \end{bmatrix}}_X \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \eta \end{bmatrix}$$

.1

$$E[y] = W \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \underbrace{\begin{bmatrix} : & \mu_1 \\ : & (8 \text{ more}) \\ \mu_2 \\ : & (6 \text{ more}) \\ \mu_3 \\ : & (4 \text{ more}) \end{bmatrix}}_{W} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ : & (8 \text{ more}) \\ 0 & 1 & 0 \\ : & (6 \text{ more}) \\ 0 & 0 & 1 \\ : & (4 \text{ more}) \end{bmatrix}}_{W} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \beta \end{bmatrix}$$

.2

$$\eta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \frac{3}{7}\mu_1 + \frac{1}{3}\mu_2 + \frac{5}{21}\mu_3 \\ \frac{4}{7}\mu_1 - \frac{1}{3}\mu_2 - \frac{5}{21}\mu_3 \\ -\frac{3}{7}\mu_1 + \frac{2}{3}\mu_2 - \frac{5}{21}\mu_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 3/7 & 1/3 & 5/21 \\ 4/7 & -1/3 & -5/21 \\ -3/7 & 2/3 & -5/21 \end{bmatrix}}_H \underbrace{\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}}_{\beta}$$

.3

$$\beta = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mu + \alpha_1 \\ \mu + \alpha_2 \\ \mu - 1.8\alpha_1 - 1.4\alpha_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1.8 & -1.4 \end{bmatrix}}_G \underbrace{\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix}}_{\eta}$$

:R טיפן .4

```
X <- matrix(ncol = 3, data = c(rep(c(1,1,0),9),rep(c(1,0,1),7),rep(c(1,-1.8,-1.4),5)), byrow = T)
```

```
W <- matrix(ncol = 3, data = c(rep(c(1,0,0),9),rep(c(0,1,0),7),rep(c(0,0,1),5)), byrow = T)
```

```
H <- matrix(ncol = 3, data = c(3/7,1/3,5/21,4/7,-1/3,-5/21,-3/7,2/3,-5/21), byrow = T)
```

```
G <- matrix(ncol = 3, data = c(1,1,0,1,0,1,1,-1.8,-1.4), byrow = T)
```

```
all.equal(W %*% G, X) && all.equal(X %*% H, W)
```