מודלים סטטיסטיים ויישומיהם (52518)

תשע"ח, סמסטר א'

בוחן בית 1

שאלה 1 (35 נקודות)

נתייחס לניתוח שונות חד-כיווני עם $I = 4$. נגדיר

$$\psi(\mathbf{c}) = \mathbf{c}^T \boldsymbol{\mu} = \sum_{i=1}^{4} c_i \mu_i$$

נניח שנרצה לבדוק את ההשערה $H_0 : \psi(\mathbf{c}) = 0$. ניתן לעשות זאת בשתי דרכים. הדרך הפשוטה הינה להגדיר

$$\hat{\psi}(\mathbf{c}) = \sum_{i=1}^{4} c_i \bar{Y}_{i\bullet}$$

לחשב את

$$T = \frac{\hat{\psi}(\mathbf{c})}{s\left[\sum_{i=1}^{4} c_i^2/n_i\right]^{1/2}}$$

ולדחות כאשר $|T| \geq t_{N-I}(1-\alpha)$. דרך אחרת הינה כך: נכתוב את המודל בצורה

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

כמו שעשינו בכיתה, עם $\boldsymbol{\beta} = [\mu_1, \mu_2, \mu_3, \mu_4]^T$. נמצא 3 וקטורים $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ בגודל 4 שמהווים בסיס לתת מרחב הניצב ל- $\mathbf{c}$. נבנה מטריצה $\mathbf{A}$ בגודל 4×3 שעמודיה הינן $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ ונגדיר $\tilde{\mathbf{X}} = \mathbf{X}\mathbf{A}$. אז נשתמש במבחן F להשוות את המודל

$$\mathbf{Y} = \tilde{\mathbf{X}}\boldsymbol{\omega} + \boldsymbol{\epsilon}, \quad \boldsymbol{\omega} \in \mathbb{R}^3 \tag{1}$$

מול המודל

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}.$$

ניתן לראות כי שתי הדרכים הנ"ל שקולות. כאן נדגים זאת עבור מקרה פרטי. נגדיר

$$\mathbf{c}^T = [0.45, -0.22, 0.87, -1.10]$$

בצעו את הפעולות הבאות:

א. עבור נתונים בקובץ qn1.txt וה- $\mathbf{c}$ הנ"ל, חשבו את הערך של T.

ב. מצאו את $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$.

רמז: ניתן לעשות זאת על ידי ביצוע תהליך גרם-שמיט בווקטורים בנאים:

$$\mathbf{v}_1 = \mathbf{c} = [0.45, -0.22, 0.87, -1.10]^T$$
$$\mathbf{v}_2 = [0, 1, 0, 0]^T$$
$$\mathbf{v}_3 = [0, 0, 1, 0]^T$$
$$\mathbf{v}_4 = [0, 0, 0, 1]^T$$

ג. חשבו את $\hat{\mathbf{Y}}^{(0)}$ עבור הנתונים שבקובץ. ד. חשבו את הסטטיסטי

$$F = \frac{\|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}^{(0)}\|^2}{s^2}$$

(באשר $\hat{\mathbf{Y}}^{(0)}$ הינו הווקטור של ערכי חיזוי תחת המודל של H_0) וודאו כי $F = T^2$.

שאלה 2 (25 נקודות)

נניח שני מודלי רגרסיה:

$$(1) \quad \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$
$$(2) \quad \mathbf{Y} = \mathbf{Z}\boldsymbol{\xi} + \boldsymbol{\epsilon}$$

עם $\boldsymbol{\xi} = \mathbf{H}\boldsymbol{\beta}$ ו- $\mathbf{Z} = \mathbf{X}\mathbf{H}^{-1}$. האומדים של $\boldsymbol{\beta}$ ו- $\boldsymbol{\xi}$ הינם

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$
$$\hat{\boldsymbol{\xi}} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{Y}$$

הוכיחו כי $\hat{\boldsymbol{\xi}} = \mathbf{H}\hat{\boldsymbol{\beta}}$.

<u>שאלה 4 (40 נקודות)</u>

נתייחס למודל הקלאסי לניתוח שונות חד כיווני

$Y_{ij} = \mu_i + \epsilon_{ij}; i = 1, \ldots, I, J = 1, \ldots, n_i$, כאשר $\epsilon_{ij} \sim N(0, \sigma^2)$.

נניח כי הקבוצות i=2,…,I מייצגות טיפולים שונים ו-i=1 מייצגת קבוצת בקרה. נגדיר את הקונטרסטים הבאים:

$$\psi^{(k)} = \mu_{k+1} - \mu_1, k = 1, \ldots, I - 1$$

a. פתחו שיטה לרו"ס בו-זמניים עם כיסוי בו זמני של $1 - \alpha$ במדוייק עבור המקרה בו σ^2 ידוע ובמקרה בו הוא אינו ידוע. יש לרשום את הביטויים הרלבנטיים בצורה מפורשת ככל שאפשר.

b. מצורף קובץ trts.txt עם סטטיסטיקות סיכום ממחקר חקלאי. חשבו את הרו"ס שפיתחתם בסעיף הקודם עם $\alpha = 0.05$ על הנתונים עבור המקרים הבאים:
   i. $\sigma^2 = 9.5$
   ii. σ^2 אינו ידוע

c. עבור נתונים אלה חשבו רו"ס בו-זמניים עבור הקונטרסטים המוגדרים בתחילת השאלה לפי שיטת בונפרוני, עבור שונות לא ידועה. השווה עם סעיף b.

d. כעת נניח את המצב המתואר בשאלה כאשר I=3 ו-$n_1 = m, n_2 = n_3 = dm$. d,m מספרים שלמים ידועים, וכן השונות ידועה. נתייחס לקונטרסטים $\psi^{(1)} = \mu_2 - \mu_1, \psi^{(2)} = \mu_3 - \mu_1$ ונרצה לבדוק באופן בו-זמני את ההשערות $H_0^{(k)}: \psi^{(k)} = 0, k = 1,2$ עם מבחן מהצורה $\left|\hat{\psi}^{(k)}\right| > \frac{c}{\sqrt{m}}$ עבור ערך קריטי מתאים שיתן רמת מובהקות בו זמנית של 0.1 במדוייק. מצאו את הערך המתאים של c עבור d=1,2,3,4,5. מה קורה ל-c כאשר d גדל? מדוע? איך ערכים אלו משתווים לערך c שיוצא משיטת בונפרוני? הסבירו.