סטטיסטיקה למדעי המחשב – פתרון תרגיל 10

1. $f_x(x,\theta)=2\theta x+1-\theta;\ x\in[0,1]$. x תצפית בודדת מההתפלגות הנתונה.

א. **קיים מבחן UMP** להשארת ה-0 והאלטרנטיבה הללו. זה אכן **אותו מבחן** שמצאנו בתרגיל 7.

ב. **קיים מבחן UMP** להשארת ה-0 והאלטרנטיבה הללו. זה אכן **אינו אותו מבחן** שמצאנו בתרגיל 7.

ג. **לא קיים מבחן UMP** להשארת ה-0 והאלטרנטיבה הללו כיוון שאזור הדחייה אינו באותו הכיוון.

2.

א. ההנחה **הראשונה** אינה סבירה (הנחת הליניאריות) שכן באופן כללי לא ניתן להניח ליניאריות אל מחוץ לתחום הנתונים ובפרט, הנחה זו צופה שלאחר מספיק שנים זמן הריצה יהיה שלילי.

ב.

$$\hat{b}\sim N\left(b,\frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2}\right)=N\left(b,\frac{0.15^2}{24573.33}\right)=N(b,0.00000092)$$

ג.

i. **לא ניתן לטעון זאת** כיוון ש-$\hat{a}$ הוא גם משתנה מקרי ולכן לא ניתן להתייחס אליו כאל קבוע.

ii. **לא ניתן לטעון זאת** כיוון ש-$\hat{a}$ ו-$\hat{b}$ תלויים ולכן יש להוסיף את ערך השונות המשותפת (Co-Var) שלהם.

ד. נגדיר H_0: b=0 ו- H_A: b≠0. נבחר את סטטיסטי המבחן להיות $\hat{b}$ בצורה מנורמלת (נסמן $\widehat{b_n}$). כלומר סטטיסטי המבחן יהיה:

$$\widehat{b_n}=\frac{\hat{b}-0}{\sqrt{\frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2}}}=\frac{\hat{b}\sqrt{\sum(x_i-\bar{x})^2}}{\sigma_\varepsilon}\sim_{H_0}N(0,1)$$

נמצא אזור דחייה ברמת מובהקות α=0.05. אזור הדחייה יהיה מהצורה $R=(-\infty,C_1]\cup[C_2,\infty)$. נחשב ערכי C:

$$C_1=b_0+Z_{1-\frac{\alpha}{2}}\sqrt{\sigma_{H0}^2}=0-1.96\cdot1=-1.96$$

$$C_2=b_0+Z_{1-\frac{\alpha}{2}}\sqrt{\sigma_\varepsilon^2}=0+1.96\cdot1=1.96$$

סה"כ נקבל **$R=(-\infty,-1.96]\cup[1.96,\infty)$**.

נמצא אומדן ל-$\widehat{b_n}$:

$$\widehat{b_n}=\frac{\hat{b}-0}{\sqrt{\frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2}}}=\frac{-0.011005562}{\sqrt{0.00000092}}=\mathbf{-11.50146031}$$

לכן **נדחה את השערת ה-0** כיוון שהאומדן נמצא באזור הדחייה.

נמצא p-value:

$$P_{H_0}\left(\widehat{b_n} \le \mathbf{-11.50146031}\right) = \phi(\mathbf{-11.50146031}) = 6.485111e-31 = \frac{\alpha^*}{2}$$

קיבלנו ערך הקטן מאוד מ-0.05 ולכן נדחה את השערת ה-0 עבור רמת מובהקות α=0.05.

ה. נמצא רווח סמך ל-b:

$$\hat{b} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2}} = -0.011005562 \pm 1.96 \cdot \sqrt{0.00000092} \Longrightarrow$$
$$\Longrightarrow [\mathbf{-0.0128}, \mathbf{-0.00913}]$$

ו. כיוון שהשונות אינה ידועה, נשתמש באומד לשנות:

$$\widehat{\sigma^2} = \frac{\sum(y_i-\bar{y})^2}{n-2} = \frac{\sum(y_i-\bar{y})^2}{22} = \frac{0.39921}{22} = \mathbf{0.0181}$$

עתה נחשב מחדש את הערכים:

נמצא אזור דחייה ברמת מובהקות α=0.05. אזור הדחייה יהיה מהצורה $R=(-\infty,C_1]\cup[C_2,\infty)$. נחשב ערכי C:

$$C_1 = b_0 + t_{n-2,1-\frac{\alpha}{2}}\sqrt{\sigma_{H0}^2} = 0 - 2.074 \cdot 1 = -2.074$$
$$C_1 = b_0 + t_{n-2,1-\frac{\alpha}{2}}\sqrt{\sigma_{H0}^2} = 0 + 2.074 \cdot 1 = 2.074$$

סה"כ נקבל **$R=(-\infty,-2.074]\cup[2.074,\infty)$**.

נמצא אומדן ל-$\widehat{b_n}$:

$$\widehat{b_n} = \frac{\hat{b}-0}{\sqrt{\frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2}}} = \frac{-0.011005562}{\sqrt{0.0000007384}} = \mathbf{-12.80713889}$$

נדחה את השערת ה-0 כיוון שהאומדן נמצא באזור הדחייה.

נמצא p-value:

$$P_{H_0}\left(\widehat{b_n} \le -12.80713889\right) = \chi_{22}(-12.80713889) \approx 0 = \frac{\alpha^*}{2}$$

קיבלנו ערך הקטן מאוד מ-0.05 (כמעט ממש 0) ולכן נדחה את השערת ה-0 עבור רמת מובהקות α=0.05.

נמצא רווח סמך ל-b:

$$\hat{b} \pm t_{n-1,1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_\varepsilon^2}{\sum(x_i-\bar{x})^2}} = -0.011005562 \pm 2.069 \cdot \sqrt{0.0000007384} \Longrightarrow$$
$$\Longrightarrow [\mathbf{-0.01278}, \mathbf{-0.009227}]$$

ז. כיוון שהאומד מנורמל, אזור הדחייה יהיה גם **$R=(-\infty,-1.96]\cup[1.96,\infty)$**.

נמצא אומדן ל-$\widehat{b_n}$:

$$\widehat{b_n} = \frac{\hat{b} - 0}{\sqrt{\frac{\sigma_\varepsilon^2}{\sum(x_i - \bar{x})^2}}} = \frac{-0.01682207}{\sqrt{0.00000428}} = \mathbf{-8.133953644}$$

לכן **נדחה את השערת ה-0** כיוון שהאומדן נמצא באזור הדחייה.

נמצא p-value:

$$P_{H_0}\left(\widehat{b_n} \le -8.133953644\right) = \phi(-8.133953644) = 2.077557e - 16 = \frac{\alpha^*}{2}$$

קיבלנו ערך הקטן מאוד מ-0.05 ולכן נדחה את השערת ה-0 עבור רמת מובהקות α=0.05.

נמצא רווח סמך ל-b:

$$\hat{b} \pm Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_\varepsilon^2}{\sum(x_i - \bar{x})^2}} = -0.01682207 \pm 1.96 \cdot \sqrt{0.00000428} \Rightarrow$$
$$\Rightarrow [\mathbf{-0.02087}, \mathbf{-0.01276}]$$

כיוון שהשונות אינה ידועה, נשתמש באומד לשנות:

$$\widehat{\sigma^2} = \frac{\sum(y_i - \bar{y})^2}{n-2} = \frac{\sum(y_i - \bar{y})^2}{16} = \frac{0.70855}{16} = \mathbf{0.04428}$$

עתה נחשב מחדש את הערכים:

נמצא אזור דחייה ברמת מובהקות α=0.05. אזור הדחייה יהיה מהצורה $R=(-\infty,C_1]\cup[C_2,\infty)$. נחשב ערכי C:

$$C_1 = b_0 + t_{n-2,1-\frac{\alpha}{2}}\sqrt{\sigma_{H0}^2} = 0 - 2.12 \cdot 1 = -2.12$$
$$C_1 = b_0 + t_{n-2,1-\frac{\alpha}{2}}\sqrt{\sigma_{H0}^2} = 0 + 2.12 \cdot 1 = 2.12$$

סה"כ נקבל **$R=(-\infty,-2.12]\cup[2.12,\infty)$**.

נמצא אומדן ל-$\widehat{b_n}$:

$$\widehat{b_n} = \frac{\hat{b} - 0}{\sqrt{\frac{\sigma_\varepsilon^2}{\sum(x_i - \bar{x})^2}}} = \frac{-0.01682207}{\sqrt{0.0000047353}} = \mathbf{-7.730468929}$$

נדחה את השערת ה-0 כיוון שהאומדן נמצא באזור הדחייה.

נמצא p-value:

$$P_{H_0}\left(\widehat{b_n} \le -7.730468929\right) = \chi_{16}(-7.730468929) \approx 0 = \frac{\alpha^*}{2}$$

קיבלנו ערך הקטן מאוד מ-0.05 (כמעט ממש 0) ולכן נדחה את השערת ה-0 עבור רמת מובהקות α=0.05.

נמצא רווח סמך ל-b:

$$\hat{b} \pm t_{n-1,1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_\varepsilon^2}{\sum(x_i - \bar{x})^2}} = -0.01682207 \pm 2.11 \cdot \sqrt{0.00000092} \Rightarrow$$
$$\Rightarrow [\mathbf{-0.0214}, \mathbf{-0.0122}]$$