PROBABILITY AND STATISTICS

הסתברות וסטטיסטיקה

מאת **יוג'ין קנציפר**

■ דף פתרונות לשאלות בית 9

9.1 בעיות אמידה קשורות להתפלגות נורמלית

■

שאלה H9.1

א. נתון כי $\sigma=22$, $n=16$, $\bar{X}=863$. רווח סמך ברמה של 90%: $1-\alpha=0.9$ $\Leftarrow$ $\alpha=0.1$; צריך למצוא $z_{\frac{\alpha}{2}}=z_{0.05}$ כזה שעבורו מתקיים $\Phi(z_{0.05})=1-\frac{\alpha}{2}=0.95$.

מחפשים בטבלה של פונקצית Φ את המספר $0.95-0.5=0.45$ ורואים שהוא מתייחס ל—$z_{0.05}\approx1.645$. אזי, באפשרותנו לחשב רווח סמך על פי הנוסחא $\bar{X}-z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}<\mu<\bar{X}+z_{\frac{\alpha}{2}}\cdot\frac{\sigma}{\sqrt{n}}$. אנחנו מקבלים:

$$863-1.645\cdot\frac{22}{\sqrt{16}}<\mu<863+1.645\cdot\frac{22}{\sqrt{16}}$$

כך ש— $854<\mu<872$. התשובה היא: רווח סמך ברמה של 90% עבור μ הוא $854<\mu<872$.

ב. נתון כי $\sigma=22$, $n=100$, $\bar{X}=868$. גם כאן $z_{0.05}\approx1.645$. אזי, חישוב פשוט מביא: $868-1.645\cdot\frac{22}{\sqrt{100}}<\mu<868+1.645\cdot\frac{22}{\sqrt{100}}$. התשובה היא: רווח סמך ברמה של 90% עבור μ הוא $864.4<\mu<871.6$.

ג. משתמשים בנוסחא $n\ge\left(\frac{z_{\frac{\alpha}{2}}\cdot\sigma}{\varepsilon}\right)^2$ עם $z_{0.05}\approx1.645$, $\sigma=22$ ו—$\varepsilon=1$. מקבלים:

$n\ge\left(\frac{1.645\cdot22}{1}\right)^2\approx1309.7$ כך ש— $n=1310$.

ד. $\sigma=22$, $\varepsilon=2$, $\alpha=0.05$, $\Phi(z_{0.025})=1-\frac{0.05}{2}=0.975$ $\Leftarrow$ $z_{0.025}\approx1.96$. אזי,

$n\ge\left(\frac{1.96\cdot22}{2}\right)^2\approx464.8$; התשובה היא: $n=465$.

ה. $\sigma = 22$, $\varepsilon = 2$, $\alpha = 0.01$, $\Phi(z_{0.005}) = 1 - \frac{0.01}{2} = 0.995 \quad \Leftarrow \quad z_{0.005} \approx 2.575$. אזי,

$n \geq \left(\frac{2.575 \cdot 22}{2}\right)^2 \approx 802.3$; התשובה היא: $n = 803$.

שאלה H9.2

נתחיל מהנוסחאות

$$\overline{Y}_{(k)} = \frac{1}{k}\sum_{j=1}^{k} Y_j \;,\; \overline{X}_{(n)} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

ו—

$$\hat{S}^2_{(Y,k)} = \frac{1}{k-1}\left[\sum_{j=1}^{k} Y_j^2 - k\overline{Y}^2_{(k)}\right] ,\; \hat{S}^2_{(X,n)} = \frac{1}{n-1}\left[\sum_{j=1}^{n} X_j^2 - n\overline{X}^2_{(n)}\right].$$

על פי הגדרה, הממוצע של המדגם המאוחד $Z = \{X_1,\ X_2,\ \ldots,\ X_n, Y_1, Y_2, \ldots, Y_k\}$ הוא

$$\overline{Z}_{(n+k)} = \frac{1}{n+k}\left(\sum_{j=1}^{n} X_j + \sum_{j=1}^{k} Y_j\right) = \frac{1}{n+k}\left(n\overline{X}_{(n)} + k\overline{Y}_{(k)}\right).$$

על פי הגדרה, אומדן לשונות של המדגם המאוחד הוא

$$\hat{S}^2_{(Z,n+k)} = \frac{1}{n+k-1}\left[\sum_{j=1}^{n} X_j^2 + \sum_{j=1}^{k} Y_j^2 - (n+k)\overline{Z}^2_{(n+k)}\right].$$

כיוון ש—

$$\sum_{j=1}^{n} X_j^2 = (n-1)\hat{S}^2_{(X,n)} + n\overline{X}^2_{(n)}$$

ו—

$$\sum_{j=1}^{k} Y_j^2 = (k-1)\hat{S}^2_{(Y,k)} + k\overline{Y}^2_{(k)},$$

אנחנו מקבלים:

$$\hat{S}^2_{(Z,n+k)} = \frac{1}{n+k-1}\left[(n-1)\hat{S}^2_{(X,n)} + (k-1)\hat{S}^2_{(Y,k)} + n\overline{X}^2_{(n)} + k\overline{Y}^2_{(k)} - (n+k)\overline{Z}^2_{(n+k)}\right].$$

נשאר לפשט את החלק שמתייחס לממוצעים:

$$\begin{aligned}
&n\overline{X}^2_{(n)} + k\overline{Y}^2_{(k)} - (n+k)\overline{Z}^2_{(n+k)} \\
&\quad = n\overline{X}^2_{(n)} + k\overline{Y}^2_{(k)} - (n+k)\frac{1}{(n+k)^2}\left(n\overline{X}_{(n)} + k\overline{Y}_{(k)}\right)^2 \\
&\quad = \frac{1}{n+k}\left[n(n+k)\overline{X}^2_{(n)} + k(n+k)\overline{Y}^2_{(k)} - \left(n\overline{X}_{(n)} + k\overline{Y}_{(k)}\right)^2\right] \\
&\quad = \frac{nk}{n+k}\left(\overline{X}_{(n)} - \overline{Y}_{(k)}\right)^2.
\end{aligned}$$

כתוצאה,

$$\hat{S}^2_{(Z,n+k)} = \frac{1}{n+k-1}\left[(n-1)\hat{S}^2_{(X,n)} + (k-1)\hat{S}^2_{(Y,k)} + \frac{nk}{n+k}\left(\overline{X}_{(n)} - \overline{Y}_{(k)}\right)^2\right].$$

שאלה H9.3 (אין חובה לפתור)

זו בעיית אמידה קלאסית. פותרים אותה בשלבים.

שלב ראשון. יש לבנות אומד נקודתי $F_\theta(X_1,\ldots,X_n;Y_1,\ldots,Y_m)$ כזה שתוחלתו היא פרמטר עצמו: $E[F_\theta(X_1,\ldots,X_n;Y_1,\ldots,Y_m)] = \theta$.

במקרה הספציפי של $\theta = \mu_X \mp \mu_Y$ קל לראות כי הבחירה הנכונה היא

$$F_\theta(X_1,\ldots,X_n;Y_1,\ldots,Y_m) = \overline{X}_{(n)} \mp \overline{Y}_{(m)}.$$

כאן,

$$\overline{X}_{(n)} = \frac{1}{n}\sum_{j=1}^{n} X_j \quad \text{ו—} \quad \overline{Y}_{(m)} = \frac{1}{m}\sum_{j=1}^{m} Y_j$$

הם ממוצעים.

שלב שני. יש לענות על השאלה "איך האומד $F_\theta(X_1,\ldots,X_n;Y_1,\ldots,Y_m)$ מפולג?" כיוון שהתצפיות מפולגות נורמלית,

$$F_\theta(X_1,\ldots,X_n;Y_1,\ldots,Y_m) = \overline{X}_{(n)} \mp \overline{Y}_{(m)} \sim \mathrm{N}\left(\mu_X \mp \mu_Y, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right).$$

הפרמטרים של ההתפלגות חושבו בדרך רגילה. החישוב התבסס על הנוסחאות:

$$E\left[\overline{X}_{(n)}\right] = E\left[\frac{1}{n}\sum_{j=1}^{n} X_j\right] = \mu_X,$$

$$\mathrm{var}\left[\overline{X}_{(n)}\right] = \mathrm{var}\left[\frac{1}{n}\sum_{j=1}^{n} X_j\right] = \frac{\sigma_X^2}{n}$$

וכד'.

שלב שלישי. חישוב ההסתברות $P\left(\left|F_\theta(X_1,\ldots,X_n;Y_1,\ldots,Y_m) - \theta\right| \le \varepsilon_\alpha\right) = 1-\alpha$ ומציאת ε_α.

במקרה הספציפי של $\theta = \mu_X \mp \mu_Y$, יש לטפל בהסתברות

$$P\left(\left|\overline{X}_{(n)} \mp \overline{Y}_{(m)} - (\mu_X \mp \mu_Y)\right| \le \varepsilon_\alpha\right) = 1-\alpha.$$

במילים אחרות, יש לטפל ב—

$$\underbrace{P\left((\mu_X \mp \mu_Y) - \varepsilon_\alpha \le \overline{X}_{(n)} \mp \overline{Y}_{(m)} \le (\mu_X \mp \mu_Y) + \varepsilon_\alpha\right)}_{\Phi\left(\frac{(\mu_X \mp \mu_Y)+\varepsilon_\alpha-(\mu_X \mp \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\right)-\Phi\left(\frac{(\mu_X \mp \mu_Y)-\varepsilon_\alpha-(\mu_X \mp \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\right)} = 1-\alpha.$$

זה מביא את המשוואה:

$$\Phi\left(\frac{\varepsilon_\alpha}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\right) - \Phi\left(-\frac{\varepsilon_\alpha}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\right) = 2\Phi\left(\frac{\varepsilon_\alpha}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\right) - 1 = 1-\alpha$$

כך ש—

$$\Phi\left(\frac{\varepsilon_\alpha}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}\right) = 1-\frac{\alpha}{2}.$$

מכאן אפשר למצוא את ε_α:

$$\frac{\varepsilon_\alpha}{\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} = z_{\frac{\alpha}{2}},$$

או

$$\varepsilon_\alpha = z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}.$$

שלב רביעי. קביעת רווח סמך:

$$F_\theta(X_1,\dots,X_n;Y_1,\dots,Y_m) - \varepsilon_\alpha \le \theta \le F_\theta(X_1,\dots,X_n;Y_1,\dots,Y_m) + \varepsilon_\alpha.$$

במקרה הספציפי שלנו, רווח סמך ל— $\theta = \mu_X \mp \mu_Y$ הוא:

$$\overline{X}_{(n)} \mp \overline{Y}_{(m)} - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} \le \mu_X \mp \mu_Y \le \overline{X}_{(n)} \mp \overline{Y}_{(m)} + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}.$$