סטטיסטיקה / תרגיל 8#

אריאל סטולרמן
קבוצה 03

(1)

(a)

נסמן את מוצאו של אדם ב-θ, והיא יכולה לקבל את הערכים 0 עבור אנגלי, ו-1 עבור צרפתי. כמו כן נסמן את המשקה אותו מזמין אדם ב-$d \in \{beer, brendy, whiskey, wine\}$. תחת סימונים אלו:

- השערת האפס היא $H_0: \theta = 0$
- האלטרנטיבה: $H_1: \theta = 1$
- סטטיסטי המבחן הוא d.

(b) להלן הסיכוי לטעות בשני הפאבים – נחשב את הסיכוי לזיהוי כצרפתי בהינתן שהאדם אנגלי:

$i.$ בפאב A:

$$P_{H_0}(d \in \{beer, whiskey\}) = 0.1 + 0.2 = 0.3$$

$ii.$ בפאב B:

$$P_{H_0}(d = brendy) = 0.2$$

(c) להלן הסיכוי שהפאבים לא יהיו רגישים מספיק:

$i.$ בפאב A:

$$P_{H_1}(d \notin \{beer, whiskey\}) = P_{H_1}(d \in \{brendy, wine\}) = 0.2 + 0.6 = 0.8$$

$ii.$ בפאב B:

$$P_{H_1}(d \neq brendy) = P_{H_1}(d \in \{beer, whiskey, wine\}) = 0.1 + 0.1 + 0.6 = 0.8$$

(d) פאב B מזהה יותר טוב את מוצאו של אדם, כיוון שהסיכוי לטעות בזיהוי המוצא בו נמוך מאשר בפאב A.

(e) מספר המבחנים שאפשר לבנות הוא כמספר הצירופים האפשריים ללא סדר של המשקאות הנתונים, כאשר צירוף הוא בין משקה אחד לארבעה משקאות, מספר זה הוא: $\sum_{i=1}^{4}\binom{4}{i} = 2^4 - 1 = 15$.

(2)

(a)

$i.$ תחת נתונים אלו, לא ידוע האם חוקר ב׳ ידחה את השערת האפס, שכן יתכן וכיוון הדחייה שלו (המושפע מהאלט׳ שלו) שונה מכיוון הדחייה של חוקר א׳.

$ii.$ במקרה זה חוקר ב׳ גם כן ידחה את השערת האפס, כיוון שאזור הדחייה שלו גדול יותר ומכיל את אזור הדחייה של חוקר א׳.

(b)

$i.$ תשובה זהה לסעיף (a),i.

$ii.$ יתכן (אך לא הכרחי) שחוקר ב׳ יחליט כן לדחות את השערת האפס למרות בחירתו של חוקר א׳, כיוון שכאמור אזור הדחייה שלו גדול יותר, ותתכן תוצאה שתיפול באזור הדחייה שלו, שלא נפלה באזור הדחייה של חוקר א׳.

(3)

נסמן X_i - כמות הגשמים שירדה בשנה i בירושלים, כאשר $X_i \sim N(\mu = 800, \sigma^2)$.

(a)

.i ניסוח ההשערות: $H_0: \mu = 800$, $H_1: \mu > 800$.

.ii סטטיסטי המבחן יהיה $\bar{X}$.

.iii כיוון הדחייה הוא ערכים גבוהים, נחשב את הערך הקריטי c עבור רמת מובהקות $\alpha = 0.05$:

במקרה זה המדגם הנתון:

$$\bar{x} = \frac{801 + \cdots + 810}{10} = 793.3$$

תחילה נחשב אומד לשונות מהמדגם הנתון:

$$\widehat{\sigma^2} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{(801 - 793.3)^2 + \cdots + (810 - 793.3)^2}{10 - 1} = 268\frac{1}{90} \Rightarrow \hat{\sigma} = 16.37$$

כעת נחשב את הערך הקריטי ברמת מובהקות $\alpha = 0.05$:

$$c = \mu_0 + t_{n-1,1-\alpha} \cdot \frac{\hat{\sigma}}{\sqrt{n}} = 800 + \underbrace{t_{9,0.95}}_{=1.833} \cdot \frac{16.37}{\sqrt{10}} = \mathbf{809.48}$$

.iv המסקנה תהיה שברמת ביטחון של 95% ניתן להגיד כי השערת האפס **נכונה**.

(b)

.i ניסוח ההשערות: $H_0: \mu = 800$, $H_1: \mu \neq 800$

.ii סטטיסטי המבחן יהיה גם כן $\bar{X}$.

.iii כיוון הדחייה הוא או ערכים נמוכים או ערכים גבוהים. נחשב את הערך הקריטי c עבור רמת מובהקות $\alpha = 0.05$:

במקרה זה אזור הדחיה הוא דו כיווני, ונשתמש בנתונים שחישבנו קודם לאומד לשונות. נסמן ב-c_{low} את הערך הקריטי התחתון וב-c_{high} את הערך הקריטי העליון, כך ש: $P_{H_0}\left(\bar{X} \in (-\infty, c_{low}] \uplus [c_{high}, \infty)\right) = 0.05$

$$c_{low} = \mu_0 - t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{n}} = 800 - \underbrace{t_{9,0.975}}_{=2.262} \cdot \frac{16.37}{\sqrt{10}} = \mathbf{788.29}$$

$$c_{high} = \mu_0 + t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{n}} = \mathbf{811.70}$$

.iv המסקנה תהיה שברמת ביטחון של 95% ניתן להגיד כי השערת האפס גם כן **נכונה**.

(c) להלן רווח בר סמך ברמת סמך 99% (כלומר $\alpha = 0.01$) עבור כמות הגשם השנתית הממוצעת בירושלים:

$$\left[\bar{X} - t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{n}}, \bar{X} + t_{n-1,1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{n}}\right] = \left[793.3 - \underbrace{t_{9,0.995}}_{=3.25} \cdot \frac{16.37}{\sqrt{10}}, \dots\right] = [\mathbf{776.47}, \mathbf{810.12}]$$

(4)

(*a*) ניסוח ההשערות (פרופורציה בינומית): $H_0: p = \frac{1}{2}$, $H_1: p < \frac{1}{2}$

(*b*) מבחן לבדיקת השערת רופאי השיניים עבור $\alpha = 0.01$ (נשתמש באומד שמרני ל-p):

כיוון הדחייה הוא ערכים נמוכים. נחשב ערך קריטי c:

$$c = p_{H_0} - Z_{1-\alpha} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{1}{2} - \underbrace{Z_{0.99}}_{=2.326} \cdot \sqrt{\frac{0.5^2}{144}} = \frac{1}{2} - 2.326 \cdot \frac{1}{24} = \mathbf{0.403}$$

ומכאן קיבלנו את המבחן: עבור $p \leq 0.403$ נדחה את השערת האפס. טענת רופאי השיניים היא ש-$p = \frac{60}{144} = 0.416 > 0.403$ ולכן ניתן לקבוע (ב-99%) כי טענת הרופאים **שגויה**.

(*c*) לפי תוצאות המדגם כאמור $\hat{p} = 0.416$. כמו כן רמת הסמך הדרושה היא 80% כלומר $\alpha = 0.2$. לפיכך:

רווח סמך כללי:

$$\left[\hat{p} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = \left[0.416 - Z_{0.9} \cdot \frac{\sqrt{\hat{p}(1-\hat{p})}}{12}, 0.416 + Z_{0.9} \cdot \frac{\sqrt{\hat{p}(1-\hat{p})}}{12}\right]$$

רווח סמך רגיל:

$$= \left[0.416 - 1.282 \cdot \frac{\sqrt{0.416 \cdot 0.584}}{12}, 0.416 + 1.282 \cdot \frac{\sqrt{0.416 \cdot 0.584}}{12}\right] = [\mathbf{0.363}, \mathbf{0.468}]$$

רווח סמך שמרני (נציב $\frac{1}{2}$ תחת השורש):

$$= \left[0.416 - 1.282 \cdot \frac{0.5}{12}, 0.416 + 1.282 \cdot \frac{0.5}{12}\right] = [\mathbf{0.362}, \mathbf{0.469}]$$

(5)

נסמן X_i = משקל ביצה, $X_i \sim N(\mu = 40, \sigma = 14)$. נסמן את תוחלת משקלי הביצים לאחר טיפול הורמונלי ב-μ_{new}.

מדגם נתון: $n = 15$, $\bar{x} = 43.5$

(*a*) מסקנת המחקר ברמת מובהקות $\alpha = 0.04$:

תחילה ננסח השערות: $H_0: \mu_{new} = 40$, $H_1: \mu_{new} > 40$ – כלומר השערת האפס אותה רוצים לסתור היא שלמרות ההורמון, משקל הביצה לא יגדל. נבחר כסטטיסטי המבחן את המשקל הממוצע של הביצים במדגם.

מניסוח ההשערות עולה כי נדחה את השערת האפס עבור ערכים גבוהים, על כן נמצא ערך קריטי c תחת הנתונים:

$$c = \mu_{new_0} + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}} = 40 + \underbrace{Z_{0.96}}_{=1.75} \cdot \frac{14}{\sqrt{15}} = 46.32$$

כלומר עבור משקל ממוצע ≤ 46.32 נדחה את השערת האפס. לאור המדגם הנתון, **לא נדחה** את השערת האפס. לכן, מסקנת המחקר תהיה שבסיכוי של 96%, ההורמון אינו מגדיל את משקלי הביצים.

(*b*) רווח סמך ברמת סמך 97%, כלומר $a = 0.03$:

$$\left[\bar{x} - Z_{1-\frac{0.03}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{1-\frac{0.03}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = \left[43.5 - \underbrace{Z_{0.985}}_{=2.17} \cdot \frac{14}{\sqrt{15}}, 43.5 + Z_{0.985} \cdot \frac{14}{\sqrt{15}}\right] = [\mathbf{35.65}, \mathbf{51.34}]$$

(6) $\mu, \sigma = 0.5$

נסמן קוטר כל בורג כ-X_i, ובבוקר הדגימה מתקיים: $X_i \sim N(\mu = 5, \sigma = 0.5)$, מדגם $n = 100$.

עבור רמת מובהקות של $\alpha = 0.05$, סטטיסטי המבחן יהיה הממוצע $\bar{X}$, ונבצע בדיקת השערות כאשר: $H_0: \mu = 5$, $H_1: \mu \neq 5$. במקרה זה הערכים הקריטיים c יהיו:

$$c_{1,2} = \mu_0 \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 5 \pm Z_{0.975} \cdot \frac{0.5}{10} = 4.902, 5.098$$

כלומר, אם ממוצע הקוטרים יהיה גבוה מ-5.098 מ״מ או נמוך מ-4.902 מ״מ, בסיכוי של 95% שהמכונה לא כוילה כראוי.

עבור $a = 0.10$: נחליף בחישוב את $Z_{0.975}$ ב-$Z_{0.95}$ ונקבל את הערכים $c_{1,2} = 4.917, 5.082$, עבור הסיכוי 90% כמובן.