<u>סטטיסטיקה / תרגיל 7#</u>

אריאל סטולרמן
קבוצה 03

(1)

x תצפית בודדת מהתפלגות: $f_X(x,\theta)=\begin{cases}2\theta x+1-\theta, & x\in[0,1]\\ 0, & o/w\end{cases}$

(a) נפתר בתרגול. תשובה סופית: זו פונקצית צפיפות לגיטימית.

(b) נפתר בתרגול. תשובה סופית: סטטיסטי המבחן יהיה התצפית הבודדת x, וכיוון הדחיה יהיה ערכים קרובים ל-1.

(c) נפתר בתרגול. תשובה סופית: עבור ההשערות $H_0:\theta=0; H_1:\theta=1$: $P_{H_0}(x\in[1-\alpha,1])=\alpha$.

(d) השינויים באיזורי הדחייה עבור שינוי H_0, H_1:

- $H_0:\theta=0; H_1:\theta=2$: איזור הדחייה ישאר זהה (כיוון הדחייה נשאר אותו דבר, ומסתכלים על הקטע $[c,1]$ כיוון שאין משמעות ל-x גדול מ-1, שכן $\forall x\notin[0,1].f(x)=0$).
- $H_0:\theta=0.5; H_1:\theta=1$: איזור הדחייה יקטן.

(2)

זמן הגעה של טכנאי i – X_i, $X_i\sim exp(\lambda)$. זמן תיקון בעיה j: $t_j=\sum_{i=1}^{30}X_i$

(a) נפתר בתרגול. תשובה סופית: $P(t_j<200)\cong\Phi\left(\frac{200-30/8}{30/64}\right)=\Phi(-0.0208)=\mathbf{0.49}$.

(b) נפתר בתרגול. תשובה סופית: לא סביר להניח ש-iid ריאלית.

(c) נפתר בתרגול. תשובה סופית: $\left[\frac{1}{30}\cdot\left(\bar{t}-\sqrt{\frac{\bar{t}^2}{30m}}\cdot Z_{1-\frac{\alpha}{2}}\right),\frac{1}{30}\cdot\left(\bar{t}+\sqrt{\frac{\bar{t}^2}{30m}}\cdot Z_{1-\frac{\alpha}{2}}\right)\right]$.

(d) נפתר בתרגול. תשובה סופית: $[6.84, 7.83]$.

(e) נפתר בתרגול. תשובות סופיות:

(i) סטטיסטי המבחן: $\bar{t}$ - זמן התיקון הממוצע.

(ii) כיוון איזור הדחיה יהיה ערכים נמוכים.

(iii) נסמן את סף הדחיה ב-c: $c=240+\sqrt{\frac{30}{m}}\cdot 8Z_\alpha$

(3)

מספר השאילתות בשניה ה-i: X_i. $X_i\sim Pois(\lambda)$.

(a) מרכז החישובים: $\boldsymbol{H_0:\lambda=2}$, חשב האוניברסיטה: $\boldsymbol{H_1:\lambda=1}$

(b) עבור ערכים **נמוכים** של מספר השאילתות הממוצע בשניה נדחה את השערת האפס.

(c) נסמן $T_k=\sum_{i=1}^{4}X_{j_i}$, לפי תכונות התפלגות פואסונית: $T_k\sim Pois(4\lambda)\Leftarrow$

- אם מרכז החישובים צודק: $\boldsymbol{T_k\sim Pois(8)}$
- אם חשב האונ' צודק: $\boldsymbol{T_k\sim Pois(4)}$

(d) להלן מבחן בו הסיכוי לטעות קטן מ-5%:

$$P_{H_o}(T_k \in [0,c]) \leq \alpha = 0.05$$

T_k מתפלג פואסונית עם פרמטר $\lambda = 8$ תחת H_0, ולכן $P_{H_0}(T_k = k) = \frac{8^k}{k!} \cdot e^{-8}$, ומתקיים:

$$P_{H_0}(T_k \in [0,c]) \leq \alpha \Leftrightarrow \frac{8^c}{8!} \cdot e^{-8} \leq 0.05$$

נמצא את c המקסימלי המקיים את אי השוויון, והוא $c = 3$, שכן:

$$c = 3: \frac{8^3}{3!} \cdot e^{-8} = 0.028 < 0.05, c = 4: \frac{8^4}{4!} \cdot e^{-8} = 0.057 > 0.05$$

ולכן נקבל כי המבחן המבוקש הוא:

$$\boldsymbol{P_{H_0}(T_k \in [0,3]) \leq 0.05}$$

(e) אם נשתמש ב-$\bar{X}_k$, אזי $\bar{X}_k \sim Pois(\lambda)$, כלומר תחת H_0 מתקיים כי $\overline{X_k} \sim Pois(2)$ ולכן צריך למצוא את $c \in \mathbb{N}$ המקיים $P_{H_0}(\overline{X_k} \in [0,c]) \leq 0.05$. ניתן לראות כי כבר עבור $c = 0$ זה לא מתקיים:

$$\frac{2^0}{0!} \cdot e^{-2} = 0.135 > 0.05$$

ולכן לא קיימות 4 תצפיות כלשהן כך שעבור הממוצע שלהן נדחה את H_0 בסיכוי טעות קטן מ-5%. לכן, לא קיים אזור דחייה כמבוקש (סיכוי הטעות הקטן ביותר עבורו ניתן יהיה לבנות איזור דחיה יהיה 13.5%, עבורו איזור הדחיה הוא $\{0\}$).

(f) כיוון איזור הדחיה:

(i) ערכים **נמוכים**.

(ii) ערכים **גבוהים**.

(iii) ערכים **נמוכים**.

(iv) ערכים **גבוהים או נמוכים**.

(v) ערכים **נמוכים**.

(vi) ערכים **נמוכים**.

(g) נסמן את גודל המדגם ב-n. אם משפט הגבול המרכזי תקף עבור n, ונסמן $y = \sum_{j=1}^{n} X_{i_j}$ אזי הקירוב יהיה:

$$\boldsymbol{y \sim N(\lambda n, \lambda n)}$$

(h) עבור $n = 100$:

תחת H_0 מתקיים כי $y \sim Pois(200{,}200)$. מכאן:

$$P_{H_0}(y \in [0,c]) = \underbrace{P_{H_0}(y < 0)}_{=0} + P_{H_0}(y \in [0,c]) = P_{H_0}(y \leq c) \leq \alpha \Leftrightarrow \Phi\left(\frac{c-200}{\sqrt{200}}\right) \leq 0.05$$

$$\Leftrightarrow \frac{c-200}{\sqrt{200}} \leq Z_{0.05} \Leftrightarrow c \leq \sqrt{200} \cdot (-1.645) + 200 = 176.73 \Rightarrow$$

נשתמש במבחן:

$$\boldsymbol{P_{H_0}(y \in [0,176]) \leq 0.05}$$

(*i*) רב״ס מקורב למספר השאילתות בשניה טיפוסית (λ):

$$\sigma^2 = \mu = \lambda \Rightarrow \left[\boldsymbol{\lambda - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\lambda}{n}}, \lambda + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\lambda}{n}}}\right]$$

(*j*) נציב $\alpha = 0.05$, $n = 100$, $\hat{\lambda} = 1.54$:

$$1.54 \pm 1.96 \cdot \sqrt{\frac{1.54}{100}} \Rightarrow [\mathbf{1.296, 1.783}]$$

(4)

(*a*)

$\alpha = 0.05$. נחשב את אורך הרב״ס לפי האומד השמרני, ונחסום אותו ע״י אחוז הסטייה המקסימלי הנדרש (1% לכל צד):

$$\hat{P} + Z_{1-\frac{\alpha}{2}} \cdot \frac{0.5}{\sqrt{n}} - \left(\hat{P} - Z_{1-\frac{\alpha}{2}} \cdot \frac{0.5}{\sqrt{n}}\right) = Z_{0.975} \cdot \frac{1}{\sqrt{n}} \leq 0.02 \Leftrightarrow \sqrt{n} \geq 98 \Leftrightarrow \boldsymbol{n \geq 9604}$$

(*b*)

במקרה זה המקסימום האפשרי ל-$\hat{P}$ יהיה 0.2, ולכן n החדש יהיה:

$$\hat{P} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{0.2(1-0.2)}{n}} - \left(\hat{P} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{0.2(1-0.2)}{n}}\right) = Z_{0.975} \cdot \frac{4}{5 \cdot \sqrt{n}} \leq 0.02 \Leftrightarrow n \geq 6146.56$$

$$\Rightarrow \boldsymbol{n \geq 6147}$$

המדגם קטן ואין זה מפתיע כיוון שנוספה לנו אינפורמציה על פרופורציית המובטלים מקרב האוכלוסיה.

(*c*)

תשובה זהה לזו בסעיף (*a*), והיא $\boldsymbol{n \geq 9604}$, שכן גם שם לא השתמשנו בשום ידע על שיעור המובטלים במדינה, וגם שם חסמנו את רוחב הרב״ס ב-2% (1% לכל כיוון), ביחס ל-α המביא לרמת סמך של 95%.

(*d*)

השתכנעתי.

(*e*)

בסעיף זה מתקיים: $n = 100, 9\%\ unemployed \Rightarrow \hat{P} = 0.09$. כמו כן לא צויין אך נתייחס לאותו $\alpha = 0.05$.

רב״ס רגיל לפרופורציית המובטלים עם/בלי ההנחה מסעיף (*b*) (לא משפיע על הרב״ס):

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

$$\left[0.09 - 1.96 \cdot \sqrt{\frac{0.09 \cdot 0.91}{100}}, 0.09 + 1.96 \cdot \sqrt{\frac{0.09 \cdot 0.91}{100}}\right] = [\mathbf{0.034, 0.146}]$$

רב״ס שמרני לפרופורציית המובטלים במדינה עם ההנחה מסעיף (b):

$$\left[\hat{P} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{0.2(1-0.2)}{n}}, \hat{P} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{0.2(1-0.2)}{n}}\right] = \left[0.09 - 1.96 \cdot \sqrt{\frac{0.16}{100}}, 0.09 + 1.96 \cdot \sqrt{\frac{0.16}{100}}\right]$$

$$= [\mathbf{0.0116}, \mathbf{0.1684}]$$

רב״ס שמרני לפרופורציית המובטלים במדינה בלי ההנחה מסעיף (b):

$$\left[\hat{P} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{0.5(1-0.5)}{n}}, \hat{P} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{0.5(1-0.5)}{n}}\right] = \left[0.09 - 1.96 \cdot \sqrt{\frac{0.25}{100}}, 0.09 + 1.96 \cdot \sqrt{\frac{0.25}{100}}\right]$$

$$= [-\mathbf{0.008}, \mathbf{0.188}]$$

(5)

להלן הסבר להטיות המודגמות בפרק:

- הטיית הברירה למדגם: הטיה בתוצאות כתוצאה מהאופן בו נאספו התצפיות. למשל, איסוף תוצאות סקר הבחירות בארה״ב מתוך אוכלוסייה עשירה, ומתוכם – אלו שטרחו להיות אקטיבים ולהחזיר את תשובתם למערכת העיתון. דוגמא נוספת: ״משאל עם״ על נסיגה נוספת מהשטחים, שנערך ע״י הצבת קלפיות בתחנות דלק ושליחת ספחי עיתונים למערכת. במקרה זה של הטיה, לא משנה כמה יוגדל המדגם, הטעות תחזור.
- הטייה בדגימת מכסות: בדגימת מכסות אמנם דואגים לבנות את המדגם בפרופורציה לפילוג האוכלוסיה (למשל אחוז גברים לעומת נשים, שחורים לעומת לבנים וכו׳). שיטה זו לכשעצמה אינה מבטיחה ייצוג נאמן של כל האוכלוסיה, אך הגורם העיקרי המשפיע על ההטיה הוא הגורמים המשפיעים על בחירת נדגם אחד על פני אחר, המשפיעה באופן חזק על תוצאות הסקר. למשל, בסקר בחירות 48׳ בארה״ב שנעשה בשיטה זו, הסוקרים פנו בעיקר לדגימות מתוך אוכלוסיית רפובליקנים (שהיו עשירים ומשכילים יותר) מטעמי נוחות, ואכן זה השפיע עד כדי טעות בחיזוי תוצאות הבחירות.
- השפעת צורת הסקר וניסוח השאלות בו על תגובת הנשאלים: אופן ניסוח השאלה יכול להשפיע מאוד על תגובת הנשאל. לכן יש לתת בסיכום תוצאות סקר כלשהו גם מידע על אופן עריכת הסקר והשאלות בו. למשל: ניסוח שאלה באופן פתוח לעומת אופן סגור. ההטייה כאן נובעת מסיבות פסיכולוגיות הקשורות לאופן ההחלטה של בני אדם.
- ניתוח שגוי/חסר של תוצאות: בדוגמת בחירות 96׳ בארץ, לא היתה התייחסות לקולות הצפים, שניתן היה להחשיבם ע״ב תוצאות היסטוריות של התפלגות הקולות הצפים, וכמו כן טווח השגיאה שהוצג לא היה נכון. בוודאי שניתוח כזה יביא להטייה בתוצאות סקר לעומת המציאות.