סטטיסטיקה

תזכורת

בהרצאה הקודמת דנו באמידה נקודתית וברווחי סמך.

$$X \sim \mathcal{F}_\vartheta$$

$$X_1, \cdots, X_n$$

$$\Downarrow$$

$$\vartheta = ?$$

בדיקת השערות

X משתנה מקרי.

מודל: $X \sim \mathcal{F}_\vartheta$ כאשר: ϑ פרמטר קבוע אך לא ידוע, ו - $\mathcal{F}_\vartheta$ התפלגות התלויה ב - ϑ.

השערת האפס

הערה: נטפל בהשערת אפס פשוטה.

טענה על הפרמטר, שאנו מקווים להפריך.

לדוגמה: $\mathcal{H}_0 : \vartheta = 5$, כאשר (נתון) $\vartheta \le 5$.

הערה: לא נטפל בהשערות כמו: $\mathcal{H}_0 : \vartheta \le 5, \mathcal{H}_0 : \vartheta \neq 4, \mathcal{H}_0 : \vartheta \in \{3,5,7,9\}$.

ההשערה האלטרנטיבית (המשלים ביחס לנתונים): $\mathcal{H}_1 : \vartheta < 5$.

תוצאות הניסוי:

1. דוחים את $\mathcal{H}_0$ $\Leftarrow$ מקבלים את $\mathcal{H}_1$.
2. לא דוחים את $\mathcal{H}_0$.

השוואת תוצאות הניסוי לאמת (שאינה ידועה)

התוצאה \ האמת	$\mathcal{H}_0$ נכונה	$\mathcal{H}_0$ לא נכונה
$\mathcal{H}_0$ נדחית	טעות מסוג ראשון $false\ positive$ 5%	הצלחה
$\mathcal{H}_0$ לא נדחית	הצלחה	טעות מסוג שני $false\ negative$

i. **רקע תאורטי**

1. **זיהוי התופעה הנמדדת (X).**

הגדרת המשנה המקרי X.

2. **קביעת המודל ($\mathcal{F}_\vartheta$).**

המודל נקבע על סמך שיקולים תאורטיים, שיקולים מתמטיים וניסויים.

המודל הוא משפחה פרמטרית של התפלגויות: $X \sim \mathcal{F}_\vartheta$, כאשר ϑ לא ידוע.

3. **ניסוח השערת האפס ($\mathcal{H}_0$).**

$\mathcal{H}_0: \vartheta = \vartheta_0$.

המטרה: להפריך את $\mathcal{H}_0$.

4. **ניסוח ההשערה האלטרנטיבית ($\mathcal{H}_1$).**

$\mathcal{H}_1: \vartheta > \vartheta_0$, $\mathcal{H}_1: \vartheta \neq \vartheta_0$, וכו', בהתאם לידע התאורטי.

ההשערה האלטרנטיבית היא המשלים להשערת האפס ביחס למציאות.

5. **קביעת רמת המובהקות (α)**

בדרך כלל $\alpha = 0.05$.

$P_{\mathcal{H}_0}$(טעות מסוג ראשון).

ii. **רקע סטטיסטי**

1. **בחירת סטטיסטי (S)**

הסטטיסטי הוא פונקציה של נתוני המדגם.

בדרך כלל זהו אומד נקודתי לפרמטר.

2. **קביעת אזור דחיה**

אזור הדחיה נקבע לפי ההשערה $\mathcal{H}_0$ ורמת המובהקות α.

זהו אזור שאם S נופל לתוכו, דוחים את $\mathcal{H}_0$.

3. **קביעת גודל המדגם (n)**

גודל המדגם נקבע לפי שיקולים כלכליים ואילוצים על הטעות מסוג שני.

iii. **ביצוע הניסוי**

1. **דגימת $X_1, X_2, \cdots, X_n$**
2. **חישוב הסטטיסטי**
3. **בדיקה האם הסטטיסטי נפל באזור הדחיה**

iv. **החלטות**

אם הסטטיסטי נפל באזור הדחייה – דוחים את $\mathcal{H}_0$, והוכחנו את האלטרנטיבה $\mathcal{H}_1$ (הסיכוי לטעות הוא α).

אחרת, הניסוי נכשל, ולא דוחים את $\mathcal{H}_0$.

■

דוגמה

$$\boxed{X \sim N(\mu, \sigma^2)}$$

רוצים לבדוק את השערות על μ.

נניח ש - σ ידוע.

$$\mathcal{H}_0 : \mu = \mu_0 \,, \mathcal{H}_1 : \mu \neq \mu_0$$

$$X_1, \cdots, X_n \sim N(\mu, \sigma^2)$$

$$\Downarrow$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

אם $\mathcal{H}_0$ נכונה :

$$\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$$

לכן :

$$\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

אזור הדחיה :

$$\left\{ \bar{X} \, : \, \left| \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right| > 1.96 \right\}$$

<u>הערה</u> : 1.96 תלוי ב - α.

אם נופלים לאזור הדחיה, דוחים את $\mathcal{H}_0$.

■