

1.2 - סטטיסטיקה תיאורית

מציאות: אוכלוסיה (מסובך) ← תאור: מדגם/מדגם ניתן להבנה.

הגדרה - משתנה

ערך (מספרי) של חבר באוכלוסיה.

- איכותי (צבע, מוצא, ...) (אין משמעות לשוני)
- אורדינלי (השכלה, ...) (יש השוואת גודל בין המשתנים – גדול/קטן)
- אינטרוולי (מידת נעליים, גובה, ...) (יש משמעות להפרש)
- מנתי (משקל, שכר, ...) (יש משמעות ליחס)

מדדים סטטיסטיים

מדדי מרכז

נניח שנתוני מדגם הם x_1, \dots, x_n .

תכונות של מדדים $f: \mathbb{R}^n \rightarrow \mathbb{R}$:

• הומוגניות:

$$\text{אם } f(c \cdot x_1, \dots, c \cdot x_n) = c \cdot f(x_1, \dots, x_n)$$

• שקיפות להזזות:

$$\text{אם } f(\alpha + x_1, \dots, \alpha + x_n) = \alpha + f(x_1, \dots, x_n)$$

• סימטריות

$$\text{אם } \forall \sigma \in S_n: f(x_{\sigma 1}, \dots, x_{\sigma n}) = f(x_1, \dots, x_n)$$

ממוצע

$$\frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} (x_1 + \dots + x_n)$$

הממוצע הוא מדד הומוגני, שקיף להזזות וסימטרי.

חציון

המספר שבדיוק מחצית הנתונים מעליו ומחצית מתחתיו.

אמצע הטווח

$$\frac{\max x_i + \min x_i}{2}$$

שכיח

הערך השכיח ביותר (זה שמופיע מספר גדול ביותר של פעמים).

הכללה של החציון

לכל קבוע $0 < \alpha < 1$ אפשר להגדיר:

המספר שבדיוק α נמצא מתחתיו ו- $(1 - \alpha)$ מעליו $P_\alpha :=$

מדדי פיזור**טווח**

$$\max - \min$$

טווח בין רבעוני

$$Q_3 - Q_1 = P_{0.75} - P_{0.25}$$

ממוצע המדגם

ממוצע המדגם הוא $\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$,

השונות של המדגם

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

סטיית התקן

$$\sigma = \sqrt{\text{השונות}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

הערה

יהיו x_1, \dots, x_n נתוני מדגם. מהו t שעבורו סכום ריבועי השגיאות $(g(t) = \sum (x_i - t)^2)$ מינימלי?

$$g'(t) = \sum 2(t - x_i) = 0 \Rightarrow t = \bar{x}$$