

נוסחאון במבוא לסטטיסטיקה

סטטיסטיקה תיאורית

$$P = \frac{f}{n} = \frac{\text{שכיחות}}{\text{גודל המדגם}}$$

א. שכיחות יחסית (%):

$$f' = \frac{\text{שכיחות}}{\text{רוחב הקבוצה}}$$

ב. צפיפות:

$$\bar{X} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n}{n} = \frac{\sum_{i=1}^n x_i f_i}{n}$$

ג. מדדי מרכז:

I. \bar{X} - ממוצע:

$$\bar{X} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_k f_k}{n} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

הממוצע \bar{X} בטבלת השכיחויות:

(k = מספר הקבוצות בטבלת השכיחויות)

II. Me - חציון:

חציון למשתנה בדיד:

$$Me = X_{\left(\frac{n+1}{2}\right)}$$

אם n אי-זוגי

$$Me = \frac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2}$$

אם n זוגי

$$Me = L + \frac{\ell}{f} \left(\frac{n}{2} - F \right)$$

חציון למשתנה רציף:

L - הגבול התחתון של הקבוצה החציונית

ℓ - רוחב הקבוצה החציונית

- f - שכיחות הקבוצה החציונית
- F - השכיחות המצטברת עד הקבוצה החציונית (לא כולל אותה)

\tilde{X} - שכיח

I. השכיח למשתנה בדיד הוא הערך בעל השכיחות הגבוהה ביותר.

II. השכיח למשתנה רציף:

$$\tilde{X} = L + \frac{f'_i - f'_{i-1}}{(f'_i - f'_{i-1}) + (f'_i - f'_{i+1})} \cdot \ell$$

כאשר:

- L - הגבול התחתון של קבוצת השכיח
- ℓ - הרוחב של קבוצת השכיח
- f'_i - הצפיפות בקבוצת השכיח
- f'_{i-1} - הצפיפות בקבוצה שלפני השכיח
- f'_{i+1} - הצפיפות בקבוצה שאחרי קבוצת השכיח

ד. מדדי פיזור:

$$V = \frac{(x_1 - \bar{x})^2 f_1 + (x_2 - \bar{x})^2 f_2 + \dots + (x_n - \bar{x})^2 \cdot f_n}{n} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 f_i}{n}$$

V - השונות:

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

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

השונות V בטבלת השכיחויות:

$$Q_3 - Q_1 = \left[L + \frac{\ell}{f} \cdot \left(\frac{3}{4} N - F \right) \right] - \left[L + \frac{\ell}{f} \cdot \left(\frac{N}{4} - F \right) \right]$$

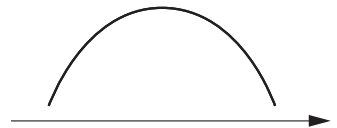
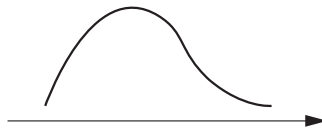
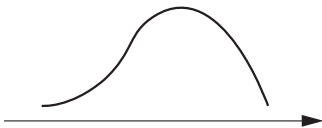
התחום הביך-רבעוני:

ה. צורות התפלגות פעמוניות:

התפלגות א-סימטרית שמאלית

התפלגות א-סימטרית ימנית

התפלגות סימטרית



$$\tilde{X} > Me > \bar{X}$$

$$\tilde{X} < Me < \bar{X}$$

$$\tilde{X} = Me = \bar{X}$$

ממוצע > חציון > שכיח

ממוצע < חציון < שכיח

ממוצע = חציון = שכיח

רגרסיה לינארית

א. r_{XY} – מקדם מתאם (פירסון) בין המשתנים X ו-Y

$$S_X^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2 \quad S_Y^2 = \frac{\sum_{i=1}^n Y_i^2}{n} - \bar{Y}^2$$

$$S_{XY} = \text{cov}(X, Y) = \frac{\sum_{i=1}^n X_i Y_i}{n} - \bar{X} \cdot \bar{Y}$$

$$r_{XY} = \frac{\text{cov}(X, Y)}{S_X \cdot S_Y}$$

$$-1 \leq r_{XY} \leq 1$$

ב. קו הרגרסיה של Y על סמך X

$$\hat{Y} = a_{Y/X} X + b_{Y/X}$$

$$a_{Y/X} = \frac{S_{YX}}{S_X^2} \quad b_{Y/X} = \bar{Y} - a_{Y/X} \bar{X}$$

ג. קו הרגרסיה של X על סמך Y

$$\hat{X} = a_{X/Y} Y + b_{X/Y}$$

$$a_{X/Y} = \frac{S_{XY}}{S_Y^2} \quad b_{X/Y} = \bar{X} - a_{X/Y} \cdot \bar{Y}$$

הסתברות

$$P(A) = \frac{\text{מספר האפשרויות של מאורע A}}{\text{מספר האפשרויות של מרחב המדגם}} = P(A)$$

$$0 \leq P(A) \leq 1$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cap B) = P(A) \cdot P(B)$$

כאשר A ו-B הם מאורעות בלתי תלויים

התפלגויות

$$X \sim B(n, p)$$

א. התפלגות בינומית

$$E(X) = np$$

תוחלת

$$V(X) = npq = np(1-p)$$

שונות

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$k = 0, 1, 2, \dots, n$$

$$0 \leq p \leq 1$$

ב. התפלגות נורמלית

$$X \sim N(\mu, \sigma)$$

μ - ממוצע האוכלוסייה

σ - סטיית התקן של האוכלוסייה

$$z = \frac{X - \mu}{\sigma} \sim N(0, 1) \quad (\text{ציון תקן})$$

התפלגות נורמלית סטנדרטית I.

$$P(X \leq a) = \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(X > a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(a < X < b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$\Phi(-z) = 1 - \Phi(z)$$

II. משפט הגבול המרכזי

יהיו x_1, x_2, \dots, x_n משתנים מקריים בלתי תלויים ושווי-התפלגות בעלי תוחלת μ וסטיית תקן σ , המוגדרים על אותו מרחב מדגם.

עבור n מספיק גדול מתקיים בקירוב:

$$\sum_{i=1}^n X_i = S_n \sim N(n\mu, \sqrt{n} \sigma) \quad \text{התפלגות הסכום:}$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad \text{התפלגות הממוצע:}$$

פונקציות שימושיות באקסל

מניית מספר התאים המכילים מספרים	Count
החזרת הערך הגדול ביותר בקבוצת ערכים	Max
החזרת הערך הקטן ביותר בקבוצת ערכים	Min
סכום כל המספרים בטווח תאים	Sum
החזרת ממוצע חשבוני של ארגומנטים	Average/mean
החזרת הערך השכיח במערך	Mode
החזרת הערך החציוני במערך	Median
סטיית התקן בהתבסס על מדגם	Stdev
שונות המדגם	Var
החזרת מידת האסימטריה של ההתפלגות	Skew
החזרת ההתפלגות המצטברת הנורמלית עבור ממוצע וסטיית תקן	Normdist
החזרת ההתפלגות המצטברת הנורמלית הסטנדרטית	Normsdist
החזרת ערך מנורמל מתוך התפלגות המאופיינת על-ידי ממוצע וסטיית תקן	Standardize
החזרת מקדם המתאם בין שתי קבוצות נתונים	Correl
החזרת השיפוע של קו הרגרסיה הליניארית	Slope

החזרת הקבוע של קו הרגרסיה הלינארית	Intercept
ניכוי ערך עתידי לאורך מגמה לינארית	Forecast
החזרת המגמה הלינארית המתאימה לנתונים	Linest
החזרת ערך לוגי המבוסס על בחינת ערך תא	if

בהצלחה!