שלבי הפתרון בשאלות סטטיסטיקה ורגרסיה

מתמטיקה • כיתה י"א • 4 יחידות לימוד (שאלון 471)

מה נתון בשאלה?

משתנה אחד • טבלת שכיחויות

ערך x ושכיחות f • $N = \Sigma f$ = סך הכול

1 ממוצע: $\bar{x} = \frac{\Sigma x \cdot f}{N}$

נעלם בטבלה + ממוצע נתון ← משוואה

2 שונות: $S^2 = \frac{\Sigma f(x-\bar{x})^2}{N}$

סטיית תקן: $S = \sqrt{S^2}$ • בונים עמודות $x - \bar{x}$, $(x-\bar{x})^2$

3 שכיח וחציון

שכיח = הערך עם השכיחות הגבוהה ביותר

חציון: מסדרים מהקטן לגדול • N זוגי ← ממוצע המקומות $\frac{N}{2}$ ו־$\frac{N}{2}+1$

שני משתנים • קשר ורגרסיה

טבלת זוגות (x, y) או דיאגרמת פיזור

1 ממוצעים וסטיות תקן

$\bar{x}$, $\bar{y}$, S_x , S_y • טבלה: $x - \bar{x}$ • $y - \bar{y}$ • מכפלה • $(x-\bar{x})^2$

2 מקדם המתאם: $r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n \cdot S_x \cdot S_y}$

בחירה מאפשרויות: עולה ← $r > 0$ • יורדת ← $r < 0$

רק אם כל הנקודות על ישר אחד ← $r = \pm 1$

3 ישר הרגרסיה: $y - \bar{y} = r \cdot \frac{S_y}{S_x}(x - \bar{x})$

שיפוע = $r \cdot \frac{S_y}{S_x}$ • עובר תמיד דרך $(\bar{x}, \bar{y})$

4 ניבוי ושינוי נתונים

ניבוי: מציבים את x בישר (בתחום הנתונים) • זו **הערכה**, לא ערך מדויק, כי הנקודות לא כולן על הישר

נתון $y = ax + b$ ← מציבים $(\bar{x}, \bar{y})$ ומוצאים b • או $a = r \cdot \frac{S_y}{S_x}$ ← מוצאים r / S_y

שינוי בכל ערכי y	$\bar{y}$	S_y	r	שיפוע הישר
מוסיפים / מחסירים c	$\bar{y} + c$	לא משתנה	לא משתנה	לא משתנה • הישר זז ב־c
כופלים ב־$k > 0$	$k \cdot \bar{y}$	$k \cdot S_y$	לא משתנה	מוכפל ב־k

גדל ב־15% ← $k = 1.15$ • קטן ב־20% ← $k = 0.8$ • שינוי ב־x: באותו אופן משתנים $\bar{x}$ ו־S_x

דוגמה

נתונים הזוגות $(1, 2)$, $(2, 4)$, $(3, 5)$, $(4, 4)$, $(5, 5)$. מצאו את r, את ישר הרגרסיה, ונבאו את y כאשר $x = 3.5$

1. $\bar{x} = 3$, $\bar{y} = 4$, $S_x = \sqrt{2}$, $S_y = \sqrt{1.2}$ • סכום המכפלות: 6
2. $r = \frac{6}{5 \cdot \sqrt{2} \cdot \sqrt{1.2}} \approx 0.775$
3. $r \cdot \frac{S_y}{S_x} = 0.775 \cdot \frac{\sqrt{1.2}}{\sqrt{2}} = 0.6$ ← $y - 4 = 0.6(x - 3)$
4. $y = 0.6x + 2.2$ ← $y = 0.6 \cdot 3.5 + 2.2 = 4.3$

תשובה: $r \approx 0.775$ (קשר חיובי חזק) • $y = 0.6x + 2.2$ • ניבוי: $y \approx 4.3$

מהנוסחאון

הנוסחאות של שלבים 1–3 **מופיעות בנוסחאון**, וגם:

$$r = \frac{1}{N}(z_{x1}z_{y1} + \ldots + z_{xN}z_{yN})$$

לזכור

- $-1 \le r \le 1$
- קרוב ל־± 1 ← קשר חזק
- $r = 0$ ← אין קשר ליניארי
- סימן השיפוע = סימן r

טעות נפוצה

- מחלקים ב־n (מספר הנתונים), לא ב־$n - 1$
- שוכחים להוציא שורש: $S = \sqrt{S^2}$
- מעגלים רק בסוף