

מבוא לסטטיסטיקה וניתוח נתונים

פרק 29 - רגרסיה

תוכן העניינים

1. כללי 1

מדדי קשר – רגרסיה ליניארית:

רקע:

במידה וקיים קשר חזק בין שני המשתנים הכמותיים נהוג לבצע ניבוי. לבנות קו ניבויים הנקרא גם קו רגרסיה המנבא משתנה אחד על סמך האחר. מדובר בקו שמנבא את Y על סמך X . השיטה למציאת הקו הנ"ל נקראת שיטת הריבועים הפחותים והקו המתקבל נקרא קו הרגרסיה או קו הניבויים או קו הריבועים הפחותים. a - נותן את ערך Y כאשר X הנו אפס על גבי קו הניבויים. הוא נקרא החותך של הקו. b - הוא שיפוע הקו נותן בכמה בעצם Y משתנה כאשר X גדל ביחידה אחת על גבי קו הניבויים.

להלן המשוואות למציאת הפרמטרים של קו הרגרסיה: $Y = bX + a$, $b = r \frac{S_y}{S_x}$.

לצורך בניית קו ניבויים לניבוי X על סמך Y נצטרך לעדכן את הנוסחאות בהתאם.

שאלות:

- (1) נסמן ב- X את ההכנסה של משפחה באלפי ₪. נסמן ב- Y את ההוצאות של משפחה באלפי ₪. נלקחו 20 משפחות והתקבלו התוצאות הבאות:

$$\sum_{i=1}^{20} Y_i = 200, \quad \sum_{i=1}^{20} X_i = 240$$

$$\sum_{i=1}^{20} (X_i - \bar{X})^2 = 76, \quad \sum_{i=1}^{20} (Y_i - \bar{Y}) = 76$$

$$\sum_{i=1}^{20} (X_i - \bar{X})(Y_i - \bar{Y}) = 60.8$$

- א. חשבו את מדד הקשר הלינארי בין X ל- Y . מיהו המשתנה התלוי?
 ב. מצאו את קו הרגרסיה לניבוי ההוצאה של משפחה על סמך הכנסה שלה. הסבירו את משמעות הפרמטרים של קו הרגרסיה.
 ג. משפחת כהן הכניסה 15,000₪. מה ההוצאה הצפויה שלה?

- (2) נסמן ב- X את ההשכלה של אדם בשנות לימוד. נסמן ב- Y את הכנסתו באלפי ₪. במחקר התקבלו התוצאות הבאות:

$$S_x = 2, \quad S_y = 5, \quad \bar{X} = 14, \quad \bar{Y} = 8, \quad \text{COV}(X, Y) = 7.5$$

- א. חשבו את מדד הקשר של פירסון בין ההשכלה להכנסה.
 ב. מה ההכנסה הצפויה לאדם שהשכלתו 12 שנים?
 ג. מה ההשכלה הצפויה לאדם שהכנסתו 10,000₪?

- (3) חוקר רצה לחקור את הקשר הקווי שבין הציון המבחן בסטטיסטיקה לבין מספר שעות ההכנה של הסטודנטים למבחן. במדגם של 100 סטודנטים שנבחנו בקורס נרשמו התוצאות הבאות: הציון הממוצע של הסטודנטים היה 65 עם סטיית תקן של 27. מספר שעות ההכנה הממוצע היה 30 עם סטיית תקן של 18. מקדם המתאם בין הציון לשעות ההכנה היה 0.8.

- א. על פי משוואת הרגרסיה, שעת הכנה נוספת משפרת את ציון המבחן ב-?
 ב. על פי משוואת הרגרסיה, תלמיד שייגש למבחן ללא שעות הכנה כלל יקבל ציון?
 ג. מהו קו הרגרסיה לניבוי הציון לפי שעות ההכנה?

- (4) נתונים 2 משתנים X ו- Y . כמו כן נתון: $\bar{X} = 1.5, S_x = S_y = 4$,
 וכן שקו הרגרסיה של Y על בסיס X הינו: $Y = -0.2X + 0.5$.
 חשבו מהו מקדם המתאם בין X ל- Y .

תשובות סופיות:

- | | | |
|----------------------|-----------------------|-------------|
| ג. 12.4 אלפי שח. | ב. $Y = 0.8X + 0.4$. | א. 0.8 (1) |
| ג. 14.6 שנים. | ב. 4.25 אלפי שח. | א. 0.75 (2) |
| ג. $Y = 1.2X + 29$. | ב. 29. | א. 1.2 (3) |
| | | א. -0.2 (4) |