

## מתאם פירסון – תזכורת

$$r = \frac{\text{cov}(x, y)}{s_x \cdot s_y}$$

$$s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2, \quad \text{Cov}(X, Y) = \frac{\sum_{i=1}^n X_i Y_i}{n} - \bar{X} \cdot \bar{Y}$$

### תרגיל

בסמסטר מסוים נבדק הקשר בין המשתנים הבאים:

$X$  – מספר היעדרויות של סטודנטים בקורס סטטיסטיקה (במאה)

$Y$  – ציון סופי בקורס (במאה)

במדגם של 5 סטודנטים התקבלו התוצאות הבאות:  $n=5$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{10}{5} = 2$$

$$\bar{y} = \frac{375}{5} = 75$$

$$\sum_{i=1}^5 x_i y_i = 665 \quad \sum_{i=1}^5 x_i = 10 \quad \sum_{i=1}^5 y_i = \underline{375}$$

$$\sum_{i=1}^5 x_i^2 = 30 \quad \sum_{i=1}^5 y_i^2 = 28,970$$

א. חשבו את מקדם המתאם הלינארי של פירסון  $r_{xy}$ .

$$s_x^2 = \frac{\sum x_i^2}{n} - (\bar{x})^2 = \frac{30}{5} - 2^2 = 2$$

$$s_y^2 = \frac{28,970}{5} - (75)^2 = 169 \quad s_x = \sqrt{2}$$

$$s_y = \sqrt{169} = 13$$

$$\text{cov}(x, y) = \frac{\sum x_i y_i}{n} - \bar{x} \cdot \bar{y}$$

$$= \frac{665}{5} - 2 \cdot 75 = -17$$

$$r = \frac{-17}{\sqrt{2} \cdot 13} \approx -0.92$$

קשר אילן

$$x=3$$

ב. מה יהיה הניבוי לציון הסופי לתלמיד שנעדר 3 שיעורים ?

כדי לנבא את  $y$  באמצעות  $x$  :

$$\tilde{y}_i = bx_i + a \quad \text{קו הרגרסיה :}$$

$$b = \frac{rs_y}{s_x} = \frac{\text{cov}(x, y)}{s^2 x}$$

$$a = \bar{y} - b\bar{x} \quad , \quad \text{כאשר :}$$

$b$  - מבטא את שיפוע קו הרגרסיה (כמה זזים במשתנה  $y$  כאשר עולים ב-  $x$  ביחידה 1)  
 $a$  - מבטא את נקודת החיתוך עם ציר ה-  $y$

$$\tilde{y} = bx + a$$

פתרון: ציון 113

$$b = r \cdot \frac{s_y}{s_x} = -0.92 \cdot \frac{13}{\sqrt{2}} = -8.45$$

$$a = \bar{y} - b\bar{x} = 75 - (-8.45) \cdot 2 \approx 92$$

$$\tilde{y} = -8.45 \cdot x + 92$$

שאלת הנזכר

$$\tilde{y} = -8.45(3) + 92 = 66.65$$

3' ב  $x=3$

$$r = -0.92$$

$$r^2 = (-0.92)^2 = 0.846 \approx 84.6\%$$

84% מהשונות של y נבונה על ידי x

**שונות הניבויים והטעויות של y כפונקציה של x**

מאסה של 169 גרם, מסת המים 169 גרם

$$r^2 = \frac{s_{\hat{y}}^2}{s_y^2} \quad \leftarrow \text{שאר הנבויים}$$

(מאסה של מים)

ל-  $r^2$  קוראים אחוז הניבוי ול-  $s_{\hat{y}}^2$  קוראים **שונות הניבויים** של y כפונקציה של x

ולכן הנוסחה **לשונות הניבויים של y כפונקציה של x**:

$$s_{\hat{y}}^2 = r^2 \cdot s_y^2$$

הנבויים של y

והנוסחה **לשונות הטעויות של y כפונקציה של x**:

$$s_{error}^2 = (1 - r^2) \cdot s_y^2$$

$$= (1 - (-0.92)^2) \cdot 169 = 0.1536(169)$$

נוסחה נוספת לחישוב שונות הניבויים והטעויות:  $= 25.9584$

$$s_y^2 = s_{\hat{y}}^2 + s_{error}^2$$

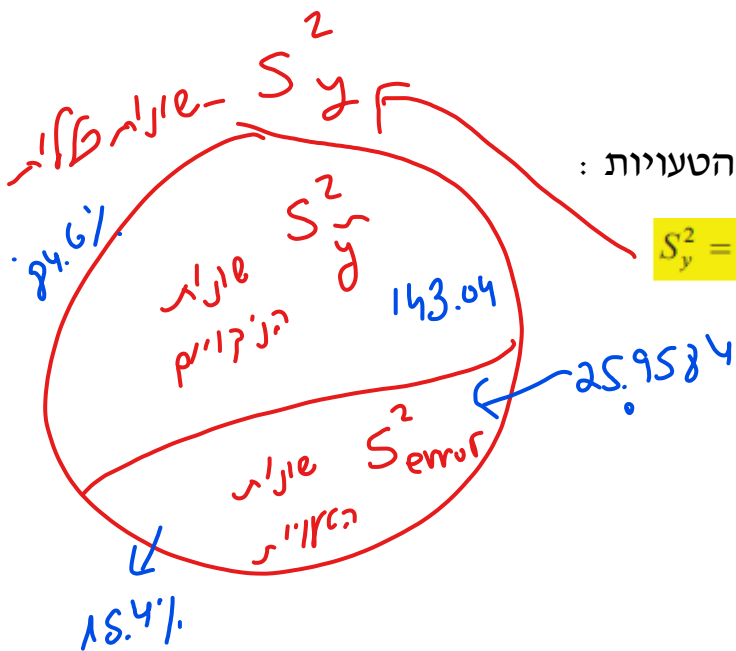
$s_y^2$  - השונות הכללית של y

$s_{\hat{y}}^2$  - שונות הניבויים של y

$s_{y-\hat{y}}^2$  - שונות הטעויות בניבוי y לפי x

$$1 - r^2$$

השונות של y



## אחוז שונות מוסברת (אחוז הניבוי)

$200 \cdot R^2$  - אחוז שונות מוסברת. (אחוז נקיון)

נתון: 
$$\begin{cases} X \text{ - מספר שנת אימון} \\ Y \text{ - משלוח} \end{cases}$$

אכן  $r^2 = R^2 = 0.9^2 = 0.81 = 81\%$

כלומר גודל מההבדל'ם ומהשונות בלכד אנשי להסדיר  
ע"י מספר שנת אימון (א).

אזורים שלא אקראי בהשקון (ווייט / מיקור)

$19\% = 100\% - 81\% \Rightarrow$  אחוז שונות בלתי מוסברת (19%)  
↓ (אחוז הטעויות)

קואדנטה נוספת:  
 $X$  - צפייה בטלויזיה. (בג')  
 $Y$  - הוצאה בלתי מיוזים (יוליו)

יחס שקדנות  $r = -0.6$   
 $R^2 = r^2 = (-0.6)^2 = 36\%$

כלומר 36% מההוצאה בלתי מיוזים מוסברת  
ע"י מספר שעות צפייה בטלויזיה.

$100\% - 36\% = 64\%$

↓ אחוז שונות בלתי מוסברת (גולומ'ם האחרים)

אם  $R^2 = 1 \Rightarrow R = 1$  כלומר אחוז השונות המוסברת הוא 100% (אחוז נקיון)  
ואחוז הטעויות (אחוז שונות בלתי מוסברת) 0%.

$y = bx + a$

