

# מבוא לסטטיסטיקה

פרק 26 - מבוא ללרגרסיה

תוכן העניינים

1. כללי ..... 1

## מבוא לקורס:

### רקע:

#### הגדרות וסימונים:

**משתנה אמפירי** – תוצאותיו ידועות מראש (למשל: רמת הכנסה, גיל, מס' שנות לימוד במדגם מסוים).

**משתנה מקרי** – תוצאותיו לא ידועות מראש (כגון תוצאה בהטלת קובייה או בהטלת מטבע). באקונומטריקה נעסוק בעיקר במשתנים מקריים.

שני סוגי המשתנים יסומנו באות לועזית עם אינדקס (למשל:  $X_t$  או  $Y_t$ ).  
**קבוע** – מקבל ערך אחד בלבד (מסומן באות לועזית ללא אינדקס – למשל  $a$  או  $b$ ).  
 לכל משתנה מקרי  $X_t$  יש תוחלת המייצגת את מרכז ההתפלגות ( $\mu_X$  או  $E(X)$ ).  
**השונות** – מייצגת את מידת הפיזור של ההתפלגות ( $\sigma_X^2$  או  $V(X)$ ).

**סטית התקן** – היא השורש של השונות ( $\sigma_X$ ).

**שונות משותפת (covariance)** – מדד להתפלגות המשותפת של שני משתנים מקריים ומייצגת את הכיוון של הקשר ביניהם ( $\text{Cov}(X, Y)$ ):

$$\text{Cov}(X, Y) = 0 \Leftrightarrow X, Y \text{ בלתי מתואמים.}$$

$$\text{Cov}(X, Y) > 0 \Leftrightarrow \text{מתאם חיובי בין המשתנים.}$$

$$\text{Cov}(X, Y) < 0 \Leftrightarrow \text{מתאם שלילי בין המשתנים.}$$

$$X, Y \text{ בלתי תלויים} \Leftarrow X, Y \text{ בלתי מתואמים.}$$

**מקדם המתאם של פירסון** – מדד לכיוון ולעוצמת הקשר הליניארי בין שני

$$\text{משתנים: } \eta_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

$$-1 \leq \eta \leq 1$$

$$\eta = 1 \text{ מתאם ליניארי חיובי מלא בין שני המשתנים.}$$

$$\eta = -1 \text{ מתאם ליניארי שלילי מלא בין שני המשתנים.}$$

$$\eta = 0 \text{ לא קיים מתאם ליניארי בין שני המשתנים.}$$

## אמידה:

פרמטר – ערך המשתנה הנחקר המתאר את כל האוכלוסייה.  
סטטיסטי/אומד – ערך המשתנה הנחקר המתאר את המדגם.

| מדגם                                                                       | אוכלוסייה                                                      |
|----------------------------------------------------------------------------|----------------------------------------------------------------|
| $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$                                     | $E(X) = \mu$                                                   |
| $S_X^2 = \frac{S_{XX}}{n} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$      | $V(X) = \sigma^2 = E(X - E(X))^2$                              |
| $\frac{S_{XY}}{n} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}$ | $\text{cov}(X, Y) = E(X - E(X))(Y - E(Y))$                     |
| $r_{XY} = \frac{S_{XY}}{\sqrt{S_{XX}} \sqrt{S_{YY}}}$                      | $\eta_{XY} = \frac{\text{cov}(X, Y)}{\sqrt{V(X)} \sqrt{V(Y)}}$ |

## נוסחאות וחוקים בסטטיסטיקה:

יהיו  $X$  ו- $Y$  משתנים מקריים, ו- $a, b$  קבועים:

חוקי הסיגמה:

$$1. \sum_{t=1}^T X_t = X_1 + X_2 + X_3 + \dots + X_T$$

$$2. \sum_{t=1}^T a = Ta \quad \text{סכום של קבוע:}$$

$$3. \sum_{t=1}^T aX_t = a \sum_{t=1}^T X_t \quad \text{סכום של קבוע כפול משתנה = לקבוע כפול הסכום:}$$

$$4. \sum_{t=1}^T (X_t \pm Y_t) = \sum_{t=1}^T X_t \pm \sum_{t=1}^T Y_t \quad \text{סכום של סכום/הפרש = לסכום/הפרש הסכומים:}$$

$$5. \sum_{t=1}^T X_t^2 \neq \left( \sum_{t=1}^T X_t \right)^2 \quad \text{יש לשים לב כי:}$$

$$\sum_{t=1}^T X_t Y_t \neq \sum_{t=1}^T X_t \sum_{t=1}^T Y_t$$

הגדרות ופיתוחים:

1. סכום הסטיות מהממוצע  $= 0$  :  $\sum_{t=1}^T (X_t - \bar{X}) = 0$
2. סכום הסטיות הריבועיות מהממוצע (מונה השונות):  

$$S_{XX} = \sum_{t=1}^T (X_t - \bar{X})^2 = \sum_{t=1}^T X_t^2 - T\bar{X}^2 = \sum_{t=1}^T (X_t - \bar{X})X_t$$
3. מונה של השונות המשותפת:  

$$S_{XY} = \sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y}) = \sum_{t=1}^T X_t Y_t - T\bar{X}\bar{Y} = \sum_{t=1}^T (X_t - \bar{X})Y_t = \sum_{t=1}^T (Y_t - \bar{Y})X_t$$

חוקי התוחלת:

1. תוחלת של קבוע = קבוע:  $E(a) = a$
2. תוחלת של סכום/הפרש = לסכום/הפרש התוחלות:  

$$E(X \pm Y) = E(X) \pm E(Y)$$

$$E\left(\sum (X_i)\right) = \sum E(X_i)$$
3. תוחלת של כפל/חילוק  $\neq$  לכפל/חילוק התוחלות:  

$$E(X \cdot Y) \neq E(X) \cdot E(Y)$$

$$E\left(\frac{X}{Y}\right) \neq \frac{E(X)}{E(Y)}$$

$$E(X^2) \neq [E(X)]^2$$
4. השפעת טרנספורמציה ליניארית על התוחלת:  

$$E\left(a + \frac{1}{a}X \pm b\right) = a + \frac{1}{a} \cdot E(X) \pm b$$

חוקי השונות:

1. עבור  $X$  ו- $Y$  בלתי תלויים/בלתי מתואמים מתקיים:  
שונות של סכום/הפרש = סכום השונות:  

$$V(X \pm Y) = V(X) + V(Y)$$

$$V\sum(X_i) = \sum V(X_i)$$
2. עבור  $X$  ו- $Y$  תלויים/מתואמים מתקיים:  
שונות של סכום/הפרש  $\neq$  סכום השונות:  

$$V(X \pm Y) = V(X) + V(Y) \pm 2 \cdot Cov(X, Y)$$

$$V(a) = 0$$

3. שונות של קבוע = 0 :  $V(a \pm x) = V(X)$

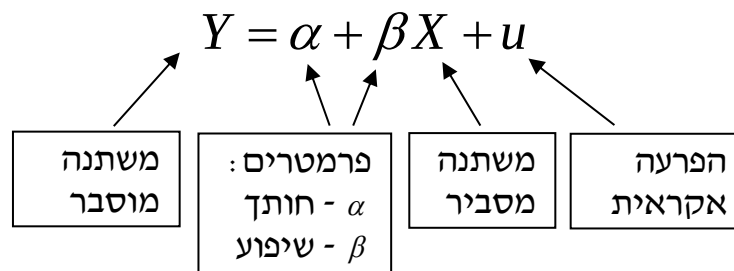
4. השפעת טרנספורמציה ליניארית על השונות :  $V(aX + b) = a^2 V(X)$

- חוקי התוחלת והשונות מתייחסים למשתנים אמפיריים כאל קבועים (יוצאים מחוץ לתוחלת או לשונות).  
חוקי הסכימה מתייחסים למשתנים אמפיריים כמשתנים הנשארים בתוך הסיגמא (רק הקבועים ייצאו מחוץ לסיגמא).

#### חוקי השונות המשותפת :

1. שונות משותפת בין משתנה לקבוע = 0 :  $\text{cov}(X, a) = 0$
2. שונות משותפת של משתנים המוכפלים בקבוע :  $\text{cov}(aX, bY) = ab \cdot \text{cov}(X, Y)$
3. שונות משותפת של משתנה עם עצמו = שונות המשתנה :  $\text{cov}(X, X) = V(X)$   
 $\text{cov}(Y, Y) = V(Y)$

#### המודל האקונומטרי :

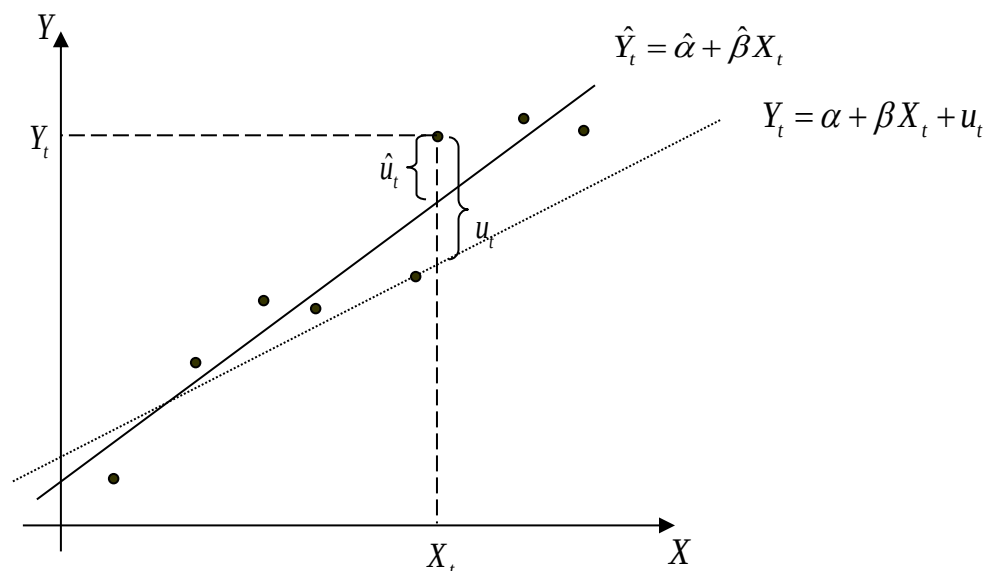


1. במודל :  $Y_t = \alpha + \beta X_t + u_t$ ,  $\alpha$  ו- $\beta$  הם מספרים קבועים אך לא ידועים. אנו יכולים להעריך אותם ולקבל אומדים (תהליך קבלת האומדנים נקרא אמידה).

2.  $\hat{\alpha}$  הוא האומד ל- $\alpha$  ו- $\hat{\beta}$  הוא האומד ל- $\beta$ .

3. אומדי ריבועים פחותים (אר"פ) הם אומדים שחושבו בשיטת הריבועים הפחותים. מסומנים בד"כ ע"י 'כובעי' -  $\hat{\beta}$ .  
אומדים אחרים מסומנים בד"כ ע"י 'תלתלי' -  $\tilde{\beta}$ .

4. בעוד  $\alpha$  ו- $\beta$  הם קבועים,  $\hat{\alpha}$  ו- $\hat{\beta}$  הם משתנים מקריים כיוון שבכל מדגם מתקבלים  $\hat{\alpha}$  ו- $\hat{\beta}$  אחרים.
5. את  $\alpha$  ו- $\beta$  לא ניתן לדעת (אלא רק לאמוד מנתוני המדגם) – הקו האמיתי באוכ' לא ידוע.
6. אפשר לדעת את  $\hat{u}_t$ , שהיא הסטיה מקו הרגרסיה במדגם:
- עבור  $X_t$ , הערך הצפוי של המשתנה המוסבר  $(\hat{Y}_t)$  המתקבל לפי הרגרסיה הוא:  $\hat{Y}_t = \hat{\alpha} + \hat{\beta}X_t$ .
- הסטיה של התצפית  $(Y_t)$  מהערך הצפוי לפי הרגרסיה  $(\hat{Y}_t)$  היא:  $\hat{u}_t = Y_t - \hat{Y}_t$ .



— קו הרגרסיה הנאמד (במדגם)  
 ..... קו הרגרסיה האמיתי באוכלוסייה  
 • תצפית בודדת

## שאלות:

(1) הבא נוכיח את הזהויות הבאות:

$$א. \sum_{t=1}^T (X_t - \bar{X})^2 = \sum_{t=1}^T X_t^2 - T\bar{X}^2$$

$$ב. \sum_{t=1}^T (X_t - \bar{X})^2 = \sum_{t=1}^T (X_t - \bar{X})X_t$$

$$ג. \sum_{t=1}^T (X_t - \bar{X}) = 0$$

$$ד. \sum \frac{(X_t - \bar{X})^2}{\sum (X_t - \bar{X})X_t} = 1$$

$$ה. \sum (\hat{\alpha} + \hat{\beta}x_i)(x_i - \bar{x}) = \hat{\beta}(x_i - \bar{x})^2$$

$$ו. \sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y}) = \sum_{t=1}^T X_t Y_t - T\bar{X}\bar{Y}$$

$$ז. \sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y}) = \sum_{t=1}^T (X_t - \bar{X})Y_t$$

(2) בטא באמצעות:  $\text{cov}(x, y), \text{var}(x), \text{var}(y)$  והקבועים  $a$  ו- $b$  את הביטויים

הבאים:

$$א. \text{Var}(ax)$$

$$ב. \text{Var}(x + y)$$

$$ג. \text{Var}(ax + b)$$

$$ד. \text{Cov}(x, ay)$$

$$ה. \text{Cov}(x + a, y + b)$$

$$ו. \text{מקדם המתאם בין } x \text{ ל-} y$$

## תשובות סופיות:

(1) הוכחה.

(2) א.  $a^2 \text{var}(x)$  ב.  $\text{var}(x) + \text{var}(y) + 2\text{cov}(x, y)$  ג.  $a^2 \text{var}(x)$  ד.  $a \text{cov}(x, y)$ 

$$ה. \text{cov}(x, y) \quad ו. r = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)} \cdot \sqrt{\text{var}(y)}}$$