תרגול 11- רגרסיה ליניארית פשוטה

<u>שאלה 1</u>

מוצר הדגל בקפיטריה של הפקולטה להנדסה הינו כריך עם גבינה בולגרית. מנהל הקפיטריה מאמין שקיים קשר ליניארי בין הפידיון מיחידת כריך בש"ח (Y) לאורך הלחמנייה ממנה עשוי הכריך, בס"מ (X). הוא הורה למהנדס התעשייה של הקפיטריה לבחון את הקשר באמצעות רגרסיה ליניארית פשוטה. לשם כך, אסף המהנדס 30 תצפיות והתקבל הממצא הבא: $\hat{y} = 10 + 0.1x$.

רק אז הבחין המהנדס כי בתצפיות שאסף, אורך הכריכים נמדד במ"מ במקום בס"מ, ולכן התקבל קו ניבוי שגוי.

מהי משוואת הרגרסיה שהייתה מתקבלת אילו המדידות היו מתבצעות בס"מ? הסבירו את תשובתכם.

<u>שאלה 1</u>

נסמן את המשתנים (חלקם הוגדרו בשאלה):

Y – המשתנה המוסבר: הפידיון מכריך

X – המשתנה המסביר במודל המקורי "הנכון": אורך הכריך בס"מ

Z – המשתנה המסביר במודל "השגוי": אורך הכריך במ"מ

במדגם שאסף המהנדס, כל תצפית מורכבת מנתון אחד של y_i ונתון אחד של z_i. משוואת הרגרסיה שהוא בנה אומדת את **y כפונקציה של של z**, בעוד שהמנהל מעוניין במודל שמסביר את y כפונקציה של x.

כדי לחשב את האומדים במודל שנבנה, חושבו SS_{zy} ו-SS_z.

כדי לחשב את האומדים במודל הרצוי, נדרשים SS_{xy} ו-SS_x.

נשים לב לקשרים הבאים בין המשתנים:

לכל תצפית i, מתקיים $x_i = \frac{1}{10} z_i$. לכן $\bar{x} = \frac{\sum_i x_i}{n} = \frac{\sum_i \frac{1}{10} z_i}{n} = \frac{1}{10}\bar{z}$.

אם כך:

$$SS_z = \sum_i^n (z_i - \bar{z})^2 = \sum_i^n (10x_i - 10\bar{x})^2 = 100\sum_i^n (x_i - \bar{x})^2 = 100 \cdot SS_x$$

$$SS_{zy} = \sum_i^n (z_i - \bar{z})(y_i - \bar{y}) = \sum_i^n (10x_i - 10\bar{x})(y_i - \bar{y}) = 10\sum_i^n (x_i - \bar{x})(y_i - \bar{y}) = 10 \cdot SS_{xy}$$

נסמן:

- b_1^x, b_0^x – האומדים במודל שבו המשתנה המסביר הוא x (אותם מחפשים)
- b_1^z, b_0^z – האומדים במודל שבו המשתנה המסביר הוא z (הם נתונים בשאלה)

$$b_1^x = \frac{SS_{xy}}{SS_x} = \frac{\frac{1}{10} SS_{zy}}{\frac{1}{100} SS_z} = \frac{100}{10}\frac{SS_{zy}}{SS_z} = 10 b_1^z = 10 \cdot 0.1 = 1$$

$$b_0^x = \bar{y} - b_1^x \bar{x} = \bar{y} - 10 b_1^z \cdot \frac{1}{10}\bar{z} = \bar{y} - b_1^z \bar{z} = b_0^z = 10$$

כלומר האומד לשיפוע גדל פי 10, והאומד לחותך לא משתנה.

לכן קו הניבוי החדש הינו $\hat{y} = 10 + x$.

<u>שאלה 2</u>

החליטו לבדוק את הקשר בין כמות המטען שמובילה חברת תעופה לבין ההכנסות שלה מכך. לשם כך בדקו את כמות המטען וההכנסה ב-10 חברות תעופה גדולות בארצות הברית. נסמן ב- X את כמות המטען (במיליוני טון) וב- Y את ההכנסה מהובלת מטענים (במיליוני דולרים). התקבלו הממצאים הבאים:

$$\bar{x} = 432.2, \qquad \sum_{i=1}^{10} x_i^2 = 2{,}408{,}810 \quad , \quad \sum_{i=1}^{10} y_i^2 = 103{,}195$$

האומד לקו הרגרסיה חושב והתקבל: $\hat{y} = 5.821 + 0.195x$.

א. מהו ממוצע ההכנסות מהובלת המטענים במדגם? **(90.1)**

ב. חשב את ערכו של מקדם המתאם המרובה (r). **(0.966)**

ג. האם קיים קשר ליניארי חיובי בין כמות המטען לבין ההכנסות מהובלתו ברמת מובהקות 0.05? **(כן)**

<u>שאלה 2</u>

א. קו הריבועים הפחותים עובר בנקודת הממוצעים $(\bar{x}, \bar{y})$ ולכן:

$$\bar{y} = b_0 + b_1 \cdot \bar{x} = 5.821 + 0.195 \cdot 432.2 = 90.1$$

ב. $r = \dfrac{SS_{xy}}{\sqrt{SS_x SS_y}}$

$$SS_x = \sum_{i=1}^{10} x_i^2 - n\bar{x}^2 = 2{,}408{,}810 - 10 \cdot 432.2^2 = 540{,}841.6$$

$$SSR = b_1^2 SS_x = 0.195^2 \cdot 540{,}841.6 = 20{,}565.5$$

$$SST = SS_y = \sum_{i=1}^{10} y_i^2 - n\bar{y}^2 = 103{,}195 - 10 \cdot 90.1^2 = 22{,}014.9$$

$$\Rightarrow R^2 = \frac{SSR}{SST} = \frac{20{,}565.5}{22{,}014.9} = 0.934$$

ג. קשר ליניארי חיובי משמעו ש- $\beta_1 > 0$.

מערכת ההשערות: $H_0 : \beta_1 = 0, \quad H_1 : \beta_1 > 0$.

סטטיסטי המבחן: $T_{b_1} = \dfrac{b_1}{s / \sqrt{SS_x}} = \dfrac{0.195}{s / \sqrt{540{,}841.6}}$

$$s = \sqrt{\frac{SS_y - b_1^2 SS_x}{n-2}} = \sqrt{\frac{22{,}014.9 - 0.195^2 \cdot 540{,}841.6}{8}} = 13.46$$

$$\Rightarrow T_{b_1} = 10.65, \quad > \quad t_{1-\alpha}^{n-2} = t_{0.95}^{8} = 1.86$$

ולכן השערת האפס נדחית בר״מ 0.05, כלומר קיים קשר ליניארי חיובי.

שאלה 3

להלן נתונים לגבי עלויות התפעול של דוד חשמלי כפונקציה של דרגות עומס:

עומס	0	1	2	4	5	8
עלות	0.5	2	4.2	6	6.5	9.5

א. אמוד את קו הרגרסיה של העלות כפונקצית ליניארית של העומס. ($\hat{y}=1.15+1.09x$)

ב. האם ע"ס הנתונים ניתן לומר שיש קשר ליניארי בין העלות לעומס ברמת מובהקות 0.05? **(כן)**

שאלה 3

סעיף א'

$$\sum x_i = 20 \qquad \sum y_i = 28.7 \qquad \sum x_i^2 = 110 \qquad \sum x_i y_i = 142.9$$

$$b_1 = \frac{\sum x_i y_i - n\bar{x}\cdot\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{142.9 - 6\cdot\frac{20}{6}\cdot\frac{28.7}{6}}{110 - 6\cdot\left(\frac{20}{6}\right)^2} = 1.09$$

$$b_0 = \bar{y} - b_1\bar{x} = \frac{28.7}{6} - 1.09\cdot\frac{20}{6} = 1.15 \quad \Rightarrow \hat{y} = 1.15 + 1.09x$$

סעיף ב'

$$t_{0.975}^{6-2} = 2.776 \quad . \quad H_0:\beta_1 = 0, \qquad H_1:\beta_1 \neq 0$$

$$\sum y_i^2 = 190.39$$

$$SS_x = \sum x_i^2 - n\bar{x}^2 = 110 - 6\cdot 3.33^2 = 43.33$$

$$SS_y = \sum y_i^2 - n\bar{y}^2 = 190.39 - 6\cdot 4.783^2 = 53.1$$

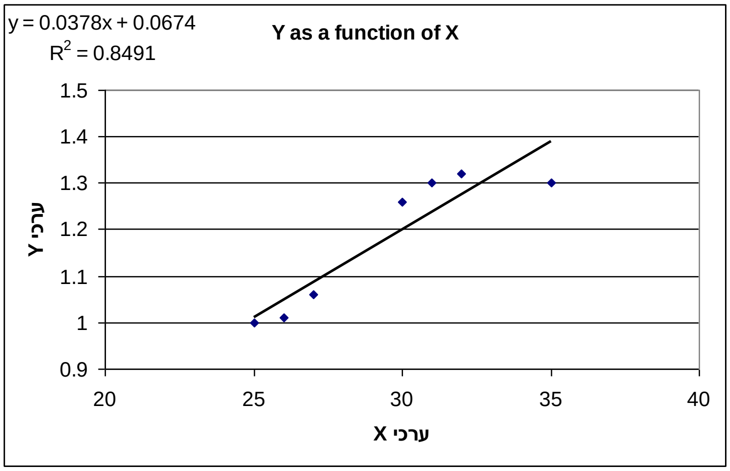
$$s = \sqrt{\frac{SS_y - b_1^2 SS_x}{n-2}} = \sqrt{\frac{53.1 - 1.09^2\cdot 43.33}{4}} = \sqrt{0.406} = 0.636$$

$$T_{b_1} = \frac{b_1}{s/\sqrt{SS_x}} = \frac{1.09*\sqrt{43.33}}{0.636} = 11.28$$

נדחה את השערת האפס ונסיק כי בר"מ 0.05 יש קשר ליניארי בין העומס לעלות.

שאלה 4

נתון הגרף הבא:



עוד נתון כי היו 7 תצפיות.

יתר על כן $SS_x = 77.714$.

א. האם הרגרסיה מובהקת בר״מ 1%? **(כן)**

ב. בדוק את ההשערה $H_0 : \beta_1 \ge 0.03$ כנגד $H_1 : \beta_1 < 0.03$ בר״מ 10%. **(מקבלים את H_0)**

שאלה 4

א. <u>אפשרות ראשונה: באמצעות מבחן F</u>

$$SSR = b_1^2 SS_x = 0.111$$

$$SST = \frac{SSR}{R^2} = \frac{0.111}{0.8491} = 0.1307$$

$$\Rightarrow \quad SSE = SST - SSR = 0.0197$$

$$MSR = \frac{SSR}{k} = \frac{SSR}{1} = 0.111$$

$$MSE = \frac{SSE}{n-2} = \frac{0.0197}{7-2} = 0.00394$$

$$\Rightarrow F = \frac{MSR}{MSE} = \frac{0.111}{0.00394} = 28.17$$

כלל ההחלטה ברמת מובהקות 1% הוא דחה אם: $F > f_{1-0.01}^{1,7-2} = f_{0.99}^{1,5} = 16.26$

לכן נדחה את השערת האפס. הרגרסיה מובהקת ברמת מובהקות 1%.

<u>אפשרות שנייה: באמצעות מבחן T</u>

מתחילים עם אותם צעדים:

$$SSR = b_1^2 SS_x = 0.111$$

$$SST = \frac{SSR}{R^2} = \frac{0.111}{0.8491} = 0.1307$$

$$\Rightarrow \quad SSE = SST - SSR = 0.0197$$

$$MSE = \frac{SSE}{n-2} = \frac{0.0197}{7-2} = 0.00394$$

$$\Rightarrow T_{b_1} = \frac{b_1}{s / \sqrt{SS_x}} = \frac{0.0378}{\sqrt{MSE} / \sqrt{77.714}} = \frac{0.0378}{\sqrt{0.00394} / \sqrt{77.714}} = 5.307$$

שימו לב שמתקיים:

$$F=(T_{b_1})^2 \Rightarrow \qquad T_{b_1}=\sqrt{F}=\sqrt{28.17}=5.307$$

$t^{n-2}_{1-\frac{\alpha}{2}}=t^{7-2}_{1-\frac{0.01}{2}}=t^{5}_{0.995}=4.032$. מכיוון ש- $\left|T_{b_1}\right|$ אכן גדול מ- $t^{n-2}_{1-\frac{\alpha}{2}}$ נדחה את השערת האפס ונחליט כי הרגרסיה מובהקת ברמת מובהקות 1%.

ב. שתי דרכים למציאת S_{b1}:

<u>דרך א׳:</u> $S=\sqrt{MSE}=\sqrt{0.00394}=0.0627 \Rightarrow S/\sqrt{SS_x}=0.0627/\sqrt{77.714}=0.00711$

<u>דרך ב׳:</u> $F=(T_{b_1})^2 \Rightarrow T_{b_1}=\sqrt{F}=\sqrt{28.17}=5.307 \Rightarrow \quad S_{b_1}=\frac{b_1}{T_{b_1}}=\frac{0.0378}{5.307}=0.00711$

כאשר T_{b1} הוא סטטיסטי המבחן מסעיף א׳, כלומר הסטטיסטי המתייחס למובהקות הרגרסיה.

נציב בנוסחאות למבחן ההשערות. שימו לב שכעת T_{b1} מתייחס להשערה הנתונה בסעיף ב׳:

$$T_{b_1}=\frac{b_1-\mu_1}{S/\sqrt{SS_x}}=\frac{0.0378-0.03}{0.00711}=1.1$$

$$-t^{n-2}_{1-\alpha}=-t^{7-2}_{0.9}=-1.476$$

מכיוון שערך סטטיסטי המבחן גדול מהערך הקריטי, לא נדחה את השערת האפס.