

רגרסיה ליניארית

תוכן עניינים

4.....	1.קו הרגרסיה במדגם
4.....	1.1 בניית קו הרגרסיה במדגם
7.....	1.2 תכונות קו הרגרסיה במדגם
7.....	1.3 אומדנים חשובים ברגרסיה פשוטה
9.....	1.4 מתאמים בין משתני הרגרסיה
10.....	1.5 משוואת הרגרסיה המתוקנת
11.....	2.6 השוואה בין מודל רגרסיה חד משתנית לבין מודל ניתוח שונות חד כיוונית
13.....	2. מובהקות הרגרסיה באוכלוסיה
13.....	2.1 ההנחות הנדרשות למודל הרגרסיה הפשוטה
17.....	2.2 בדיקת השערות למובהקות משוואת הרגרסיה באוכלוסיה
19.....	2.3 בדיקת השערות למובהקות מקדם השיפוע באוכלוסיה
21.....	2.4 קשר בין התפלגויות F ו T
23.....	2.5 רב"ס לתוחלת ערכי Y (μ_0) עבור ערך מסוים של X
24.....	2.6 רב"ס לערכי Y עבור ערך מסוים של X
25.....	2.7 שאלות מסכמות
28.....	רגרסיה מרובה
28.....	רגרסיה מרובה
28.....	1. מאפייני קו הרגרסיה במדגם
28.....	1.1 מקדמי הרגרסיה
31.....	1.2 אומדנים חשובים ברגרסיה מרובה
34.....	2. מובהקות קו הרגרסיה ומקדמיו באוכלוסיה
34.....	2.1 הנחות מודל הרגרסיה המרובה
34.....	2.2 מובהקות משוואת הרגרסיה המרובה
35.....	2.3 מובהקות השיפועים
38.....	3. שיטות להרצת רגרסיה רבת משתנים
39.....	3.1 שיטת ENTER

40.....	3.2 רגרסיה בצעדים חכמים (stepwise)
43.....	3.3 רגרסיה היררכית
44.....	3.4 השוואה בין שיטת stepwise והיררכית
47.....	3.5 התוספת לניבוי ומובהקותה
51.....	4. בדיקת קו ליניאריות
53.....	5. שאלות מסכמות

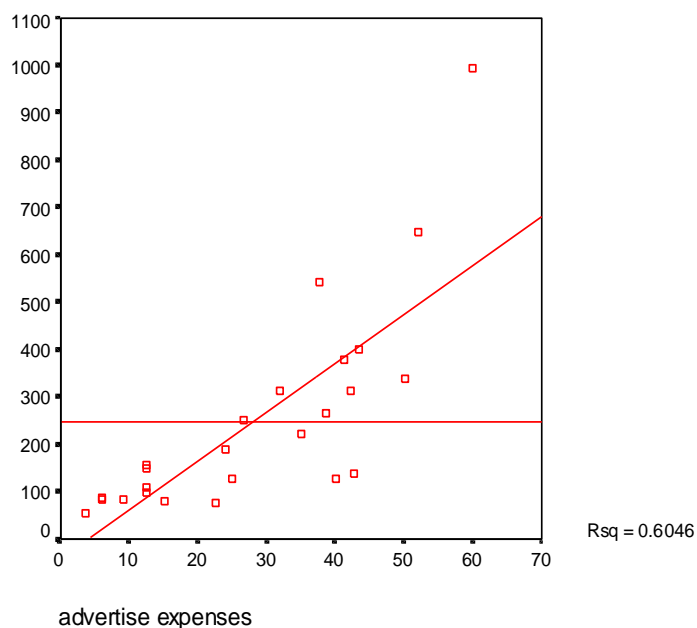
1. קו הרגרסיה במדגם

1.1 בניית קו הרגרסיה במדגם

1.1.1 עיקרון הריבועים הפחותים (L.S.E)

לאחר שנכחנו שיש מתאם ליניארי בין שני המשתנים אנו מעוניינים לבנות את קו הניבוי שיאפשר לנו לנבא משתנה אחד על סמך השני.

נשאלת השאלה: מהו הקו (מבין קווים רבים אפשריים) שיהיה המתאים ביותר לנתוני המדגם?



תשובה: קו הרגרסיה אשר יביא למינימום את הטעויות בניבוי ממנו.

את הניבוי למשתנה y במדגם נסמן: \hat{y}_i

את הערך האמיתי של y במדגם נסמן: y_i

הטעות בניבוי במדגם מהווה את ההפרש בין שני הערכים: $e_i = y_i - \hat{y}_i$

הקו הטוב ביותר בנתוני המדגם הוא הקו שמבטיח כי סכום כל הטעויות הריבועיות הוא המינימאלי ביותר, או במילים אחרות קו רגרסיה אשר נבנה על בסיס "עיקרון הריבועים הפחותים" (Least Square Estimation):

$$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum [y_i - (\hat{\alpha}_i + \hat{\beta}_i x_i)]^2 = MIN$$

באמצעות פונקציה זו (L.S.E) נוכל למצוא את מקדמי הקו הישר $(\hat{\alpha}, \hat{\beta})$ אשר יביאו למינימום את הטעויות מהקו.

לאחר פיתרון המשוואה (אין צורך לדעת לפתור) מתקבלות הנוסחאות הבאות עבור $\hat{\alpha}$ ו- $\hat{\beta}$:

$$\hat{\beta} = r \cdot \frac{\hat{s}_y}{\hat{s}_x} = \frac{\cos(x, y)}{\hat{s}_x^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

1.1.2 מקדמי הקו

א. משמעות המקדמים:

(1) $\hat{\beta}$ -מקדם השיפוע של הקו.

קצב ההשתנות של \hat{y}_i כפונקציה של עליה ביחידה אחת של x_i .

עבור עליה ביחידה אחת של x_i , \hat{y}_i עולה פי $\hat{\beta}$.

יש קשר ישיר בין $\hat{\beta}$ ל- r : ה- $\hat{\beta}$ מקבלת את הסימן של ה- r .

כאשר אין קשר בין X ל-Y ($r=0$), $\hat{\beta} = 0$.

(2) $\hat{\alpha}$ -מקדם החיתוך של הקו.

נקודת החיתוך של קו הרגרסיה עם ציר ה-Y.

כאשר $x_i=0$, $\hat{y}_i = \hat{\alpha}$.

כאשר $r=0$, $\hat{y}_i = \hat{\alpha} = \bar{y}$.

ב. חישוב המקדמים:

? דוגמא מס' 2: ניבוי המכירות באמצעות ההוצאות על פרסום (אלפי ש"ח).

להלן הנתונים המתארים את הקשר בין המכירות לבין ההוצאות על הפרסום:

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
X advertise expenses	25	28.1040	16.36932	3.70	60.00
Y sales	25	249.2400	217.85283	55.00	994.00

Correlations^a

		X advertise expenses	Y sales
X advertise expenses	Pearson Correlation	1	.778**
	Sig. (2-tailed)	.	.000
	Sum of Squares and Cross-products	6430.910	66548.276
	Covariance	267.955	2772.845
Y sales	Pearson Correlation	.778**	1
	Sig. (2-tailed)	.000	.
	Sum of Squares and Cross-products	66548.276	1139037
	Covariance	2772.845	47459.857

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=25

א. חשב את משוואת הרגרסיה לניבוי המכירות על סמך הוצאות הפרסום.

ב. מה יהיה ניבוי המכירות עבור הוצאות פרסום של 50 אלף ₪?

ג. מה יהיה ניבוי הוצאות הפרסום עבור מכירות של 267 אלף ₪?

1.2 תכונות קו הרגרסיה במדגם

(1) סכום הטעויות בניבוי ($\sum e_i$) וממוצע הטעויות בניבוי (\bar{e}_i) שווים ל-0:

$$\bar{e}_i = \sum e_i = 0$$

(2) נקודת הממוצעים (\bar{x}, \bar{y}) נמצאת תמיד על קו הרגרסיה.

(3) ממוצע הניבויים/האומדנים (\hat{y}) = ממוצע ציונים אמיתיים/מנובאים (\bar{y}):

$$\hat{y} = \bar{y}$$

(4) ככל שהציון המנבא (X_i) קרוב יותר לממוצע שלו (\bar{X}) כך הניבוי באמצעותו (\hat{Y}_i) יהיה מדויק יותר.

(5) ניתן להשתמש בקו הרגרסיה רק עבור תחום ציוני המנבא- X עליו הוא נבנה.

(6) עבור ערך x_i מסוים מתקיים כי:

טעות בניבוי + אומדן לניבוי = ציון אמיתי

$$y_i = \hat{y}_i + e_i$$

עבור כלל התצפיות מתקיים כי:

סכום ריבועי הטעויות + סכום ריבועי הניבויים = סכום ריבועי הציונים האמיתיים

$$ssy = ssreg + ssres$$

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

1.3 אומדנים חשובים ברגרסיה פשוטה

1.3.1 פרופורציית השונות המוסברת

$$r^2 = \frac{ssreg}{ssy}$$

פרופורציית השונות הלא-מוסברת:

$$1 - r^2 = \frac{ssres}{ssy}$$

1.3.2 שונות הטעויות

$$MSRES = \frac{SSRES}{n-2} = \frac{(1-r^2) \cdot (n-1) \cdot \hat{s}_y^2}{n-2}$$

MSRES מהווה אומדן חסר הטיה לשונות הלא מוסברת: $E(MSRES) \rightarrow \sigma_e^2$

ניסוחים נוספים לשונות הלא מוסברת:

- אומדן לשונות של Y לאחר ניכוי הקשר הקווי בין X ל-Y.
- אומדן לשונות של Y עבור ערך מסוים של X.

? על סמך נתוני דוגמה מס' 2 :

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
X advertise expenses	25	28.1040	16.36932	3.70	60.00
Y sales	25	249.2400	217.85283	55.00	994.00

כאשר נתון בנוסף כי $r_{xy} = 0.778$

חשב את האומדנים הבאים:

- א. סכום הסטיות הריבועיות של המכירות
- ב. אומדן לשונות הכללית של המכירות
- ג. אומדן לשונות הכללית של המכירות לאחר ניכוי הקשר הקווי עם הוצאות הפרסום
- ד. אומדן לשונות של המכירות עבור הוצאות פרסום של 52 אלף ₪ אומדן לשונות הלא מוסברת של המכירות
- ה. אומדן לסטיית התקן של הטעויות בניבוי המכירות
- ו. אומדן לפרופורציית השונות הלא מוסברת של המכירות על ידי הוצאות הפרסום
- ז. אומדן לפרופורציית שונות הטעויות של הוצאות הפרסום על ידי המכירות סכום ריבועי הסטיות של ניבוי המכירות
- ח. סכום ריבועי הסטיות של הטעויות בניבוי המכירות
- ט. סכום הסטיות הריבועיות של הוצאות הפרסום

1.4 מתאמים בין משתני הרגרסיה

	y ערכי המשתנה התלוי/מנובא	\hat{y} ניבויים/אומדנים/ PRED	E טעויות בניבוי/ שאריות/ RES
X ערכי המשתנה הב"ת/מנבא	r_{xy}	$r_{x\hat{y}} = 1$ כי המתאם בין משתנה (x) לטרנספורמציה הליניארית של אותו משתנה (\hat{y}) שווה תמיד ל-1.	$r_{xe} = 0$ כי מתאם בין קבוע (x) למשתנה (e) שווה תמיד ל-0.
y		$r_{y\hat{y}} = r_{xy}$ כי טרנספורמציה ליניארית (מ-x ל- \hat{y}) לא משנה את ערך המתאם.	$r_{ye} = \sqrt{1 - r_{xy}^2}$ כי r_{ye}^2 מהווה את פרופורציית שונות הטעויות ולכן שווה ל- $1 - r_{xy}^2$ **יש יחס הפוך בין r_{ye} לבין r_{xy} או $r_{y\hat{y}}$
\hat{y}			$r_{\hat{y}e} = 0$ כי מתאם בין קבוע (\hat{y}) למשתנה (e) שווה תמיד ל-0.

? על סמך נתוני דוגמא מס' 2, כאשר נתון כי: $r_{xy} = 0.778$
נא השלם את המתאמים בין משתני הרגרסיה:

	X	Y	PRED	RES
X				
Y				
PRED				
RES				

1.5 משוואת הרגרסיה המתוקנת

עד עתה בנינו את משוואת הרגרסיה בציוני גלם. כעת נבנה את משוואת הרגרסיה בציוני תקן.

להזכירכם- ציון תקן הוא מדד למיקום יחסי; מדד זה נותן אינפורמציה על מיקומה של תצפית מסוימת ביחס להתפלגות כלל התצפיות:

$$Z_i = \frac{x_i - \bar{x}}{\hat{s}_x}$$

בדוגמא א' נתון קשר בין שעות לימוד לפני המבחן לציון במבחן ערכו טרנספורמציה ליניארית לשני המשתנים והעבירו אותם לציוני תקן. דוגמא לטרנספורמציה על 5 תצפיות ראשונות בקובץ:

Case Summaries^a

	HOURS No. of learning hours before exam	MARK Test mark	ZHOURS Zscore: No. of learning hours before ex	ZMARK Zscore: Test mark
1	21.12	69.24	.18838	-.05646
2	15.75	51.00	-.87911	-1.18054
3	14.01	50.21	-1.22270	-1.22902
4	20.13	70.11	-.00817	-.00283
5	23.51	77.07	.66221	.42633
Total N	5	5	5	5

a. Limited to first 5 cases.

Descriptive Statistics

	Mean	Std. Deviation	N
ZHOURS Zscore: No. of learning hours before ex	.000	1.000	109
ZMARK Zscore: Test mark	.000	1.000	109

$$r_{xy} = r_{zxy} = 0.88$$

1.5.1 מקדמי המשוואה המתוקנת

א. מקדם השיפוע המתוקן:

$$z\hat{\beta} = r$$

$$z\hat{\beta} = r_{zxy} \cdot \frac{\hat{s}_{zy}}{\hat{s}_{zx}} = r_{xy} \cdot \frac{1}{1} = r_{xy} \quad \text{הסבר:}$$

מקדם השיפוע במשוואה המתוקנת שווה למתאם בין X ל-Y

ב. מקדם החיתוך המתוקנן:

$$z\hat{\alpha} = 0$$

$$\text{הסבר: } z\hat{\alpha} = z\bar{y} - z\hat{\beta} \cdot z\bar{x} = 0 - r \cdot 0 = 0$$

משוואת הרגרסיה המתוקנת יוצאת תמיד מראשית הצירים.

ג. משוואת הרגרסיה המתוקנת:

$$z\hat{y}_i = r \cdot zx_i$$

? חוקר מנבא את Y ע"י X דרך משוואת הרגרסיה. הנתונים שהיו בפניו הינם:

$$\text{א.ח.ה לשונות של } X = 16$$

$$\text{א.ח.ה לשונות של } Y = 100$$

$$\text{הממוצע של } X = 20$$

$$\text{ממוצע ערכי הניבויים} = 45$$

$$\text{המתאם בין } X \text{ ל- } Y = 0$$

א. חשב את משוואת הניבוי של Y ע"י X בציוני תקן.

ב. מהו הציון הגולמי שנבא עבור $X = 25$?

ג. חווה דעתך על הטענה הבאה: ברגרסיה בציוני תקן, ציון התקן המנובא יהיה

קטן יותר בערכו המוחלט מציון התקן המנבא.

2.6 השוואה בין מודל הרגרסיה חד משתנית לבין מודל ניתוח שונות חד כיוונית

רגרסיה חד משתנית היא מקרה פרטי של ניתוח שונות חד כיוונית: $eta^2 \geq r^2$

א. כאשר בין ממוצעי הקבוצות קיים פער קבוע ($\hat{\beta}$)-הנתונים ליניאריים $\Leftrightarrow eta^2 = r^2$, מודל הרגרסיה מתאים יותר לניתוח הנתונים מניתוח השונות כי הוא ממוקד ויעיל יותר.

ב. כאשר הפער בין ממוצעי הקבוצות אינו קבוע- הנתונים אינם ליניאריים $\Leftrightarrow eta^2 > r^2$, מודל הרגרסיה איננו מתאים לנתונים (אינם ליניאריים) ויש לבצע ניתוח שונות.

? חברה יצרה מוצר חדש ומתלבטת ביחס למחיר המכירה של מוצר זה.

מנהל השיווק יודע שיש לקבוע מחיר שהינו בסביבות 10 ₪ ליחידה אולם הוא איננו בטוח אם קביעת המחיר כ- 9 ₪ ליחידה או 11 ₪ ליחידה ישנה באופן משמעותי את רמת המכירות.

לצורך מתן מענה לשאלה זו בוצע הניסוי הבא: המוצר החדש שווק למדגם מקרי של 60 חנויות הנמצאות באזורים בעלי מאפיינים דומים. ב-20 חנויות שנבחרו באופן מקרי מבין ה-60 נמכר המוצר במחיר של 9 ₪ ליחידה, ב-20 חנויות מקריות אחרות הוא נמכר ב-10 ₪ ליחידה וביתר במחיר של 11 ₪ ליחידה.

בתום תקופת הניסוי נרשם סך היחידות שנמכרו בכל אחת מהחנויות.

לפניך פלט ובו סכום חלק מהממצאים:

Level	Number	Mean
9 ₪	20	153.6
10 ₪	20	151.5
11 ₪	20	123.2

שני סטודנטים שהסתכלו על הנתונים, הציעו לנתחם באמצעות 2 מודלים שונים:

סטודנט א' הציע לנתח את הנתונים באמצעות מודל ניתוח שונות חד כיווני.

סטודנט ב' הציע לנתח את הנתונים באמצעות מודל הרגרסיה הפשוטה כאשר X הינו המחיר בו נמכר המוצר בחנות ואילו ה-Y הוא מספר האריזות הנמכרות בחנות זו (במחיר זה).

סטודנט ג' שהתבקש לחוות את דעתו טען שהיות ובשני המודלים נבדקת השפעה של המחיר על המכירות לא חשוב באמצעות איזה מודל ינתחו הנתונים. התוצאות תהינה זהות בשני המודלים.

א. חווה דעתך על טענתו של הסטודנט השלישי. במידה ואתה סבור שיש הבדל בין שני המודלים, באיזה מודל היית בוחר לצורך ניתוח הנתונים? נמק.

ב. במידה והיית מחשב את הקשר בין המחיר למס' האריזות הנמכר בשני המודלים- מה להערכתך היה היחס בין שני מדדי הקשר- r ו- η^2 (שווים או שאחד גדול יותר מהשני)? נמק.

2. מובהקות הרגרסיה באוכלוסייה

2.1 ההנחות הנדרשות למודל הרגרסיה הפשוטה

3.1.1 ההנחות ומשמעותן

למודל הרגרסיה החד משתנית נדרשות שתי ההנחות הבאות:

(1) דגימה מקרית של התצפיות; אי תלות בין הטעויות

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad : i \neq j$$

הסבר: ההנחה הראשונה טוענת שאם התצפיות נדגמו באופן בלתי תלוי אחת בשנייה, הדבר מבטיח שהטעויות בניבוי שלהן יהיו אף הן בלתי תלויות.

(2) $\varepsilon_i \approx N(0, \sigma_e^2)$ עבור כל ערך קבוע של X_i .

הסבר: ההנחה השנייה מגלמת בתוכה 3 הנחות שונות המתייחסות להתפלגות הטעויות

א. תוחלת התפלגות הטעויות שווה לאפס ($E(\varepsilon_i) = 0$) עבור כל ערך של x_i .

הנחה זו מבטיחה שקו הרגרסיה באוכלוסייה הוא קו התוחלות – הציונים המנובאים שווים לממוצע ערכי Y עבור כל ערך קבוע של x_i ($\mu_{y_i/x_i} = \alpha + \beta x_i$). במילים אחרות הנחה זו מבטיחה כי מדובר במודל ליניארי ולא אחר.

רק כאשר הנחה זו מתקיימת – מקדמי קו הרגרסיה במדגם – $\hat{\alpha}$ ו- $\hat{\beta}$ הם אומדנים חסרי הטיה למקדמי הקו באוכלוסייה – α ו- β .

ב. שונות הטעויות (σ_e^2) שווה וקבועה עבור כל ערך של x_i .

הנחה זו מבטיחה כי שונות הטעויות תהיה שווה סביב קו הרגרסיה ומינימאלית. תחת הנחה זו, האומד לשונות הטעויות MSRES מהווה אומד לשונות ערכי Y עבור כל אחד מערכי x_i ($MSRES \rightarrow \sigma_{y_1}^2 = \sigma_{y_2}^2 = \dots = \sigma_e^2$).

רק כאשר הנחה זו מתקיימת האומדנים $\hat{\alpha}$ ו- $\hat{\beta}$ יהיו אומדנים יעילים לפרמטרים α ו- β . באוכלוסייה.

ג. הטעויות מתפלגות נורמאלית עבור כל ערך של x_i .

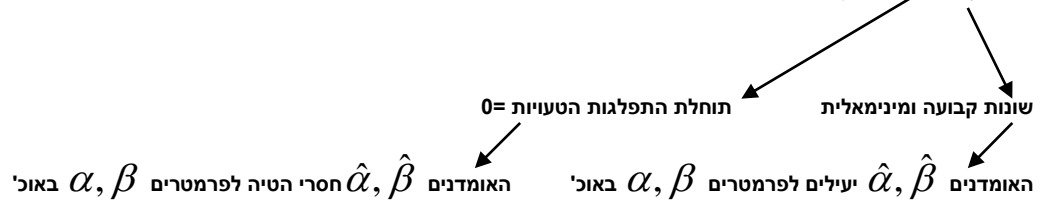
הנחה זו מבטיחה התפלגות נורמאלית של הטעויות (או ערכי Y) עבור כל ערך קבוע של x_i .

לסיכום ההנחות של מודל הרגרסיה:

1. דגימה מקרית של תצפיות / טעויות ב"ת.

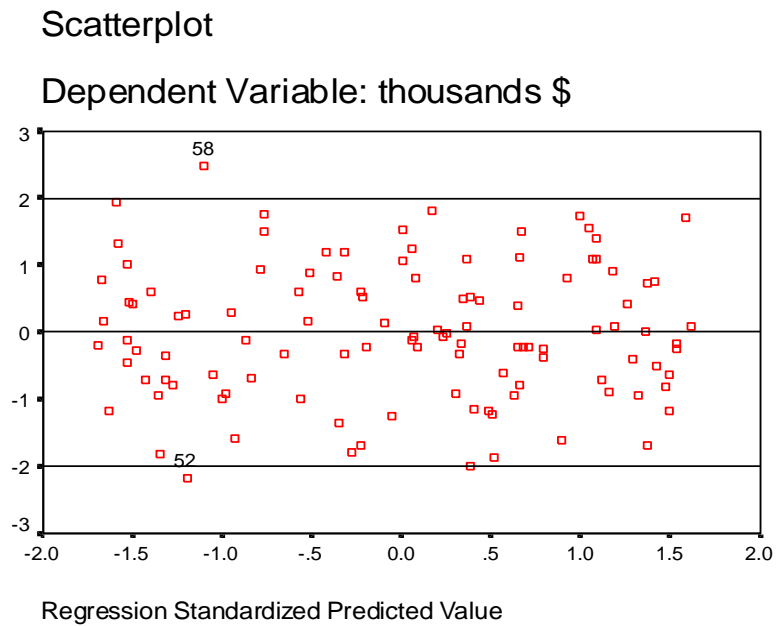
$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j \quad \text{עבור כל } j$$

2. $\varepsilon_i \approx N(0, \sigma_e^2)$ עבור כל ערך קבוע של X_i .



3.1.2 בדיקת ההנחות

את ההנחות נהוג לבדוק על ידי דיאגרמת פיזור של ניבויים (\hat{y}_i) * שאריות (z_{e_i}) בציוני תקן כמו לדוגמא:



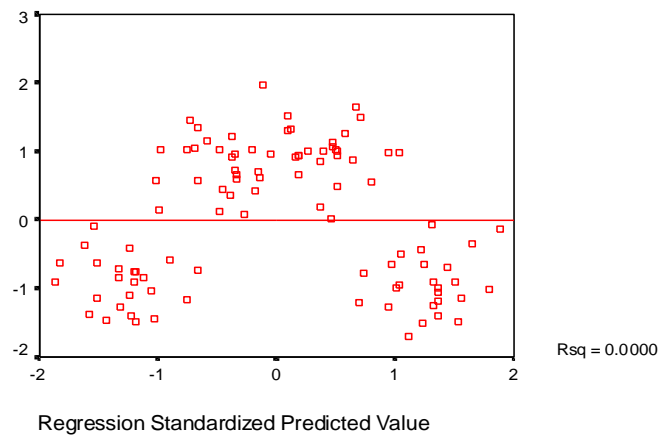
על דיאגרמה מעין זו ניתן לבצע הערכה (שאיננה מבחן סטטיסטי) האם מתקיימות כל אחת משלושת ההנחות. דיאגרמה זו מאפשרת בנוסף איתור תצפיות חריגות (הסוטות יותר משתי סטיות תקן מעל או מתחת לממוצע).

דוגמאות לאי עמידה בהנחות המודל:

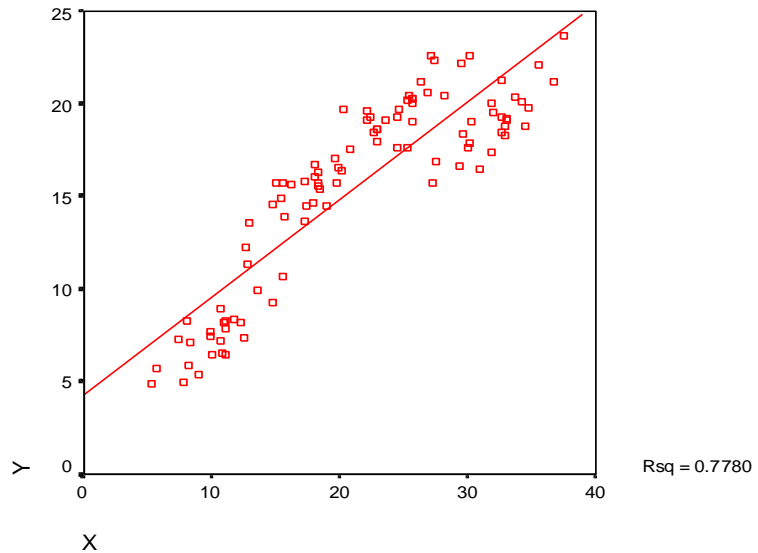
א. אי עמידה בהנחת תוחלת הטעויות $E(\varepsilon_i) = 0$ עבור כל ערך של X :

Scatterplot

Dependent Variable: Y



המשמעות של הפרת הנחה זו היא כי קו הרגרסיה איננו קו התוחלות ולמעשה המודל איננו ליניארי:

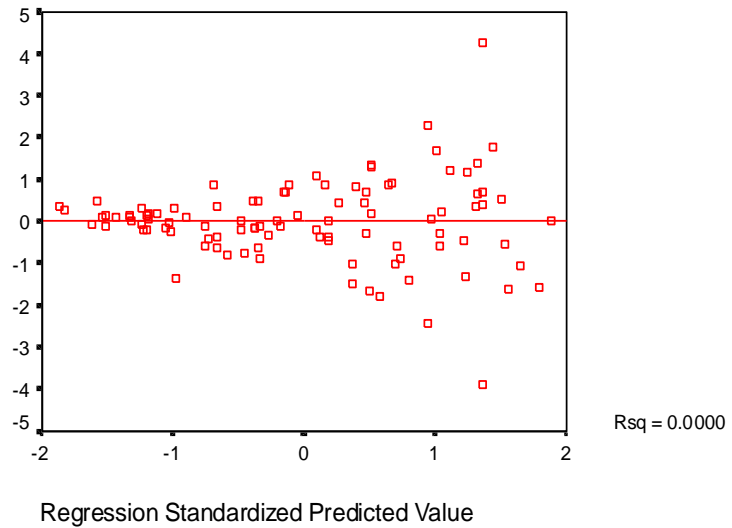


במקרה זה האומדנים $\hat{\alpha}$ ו- $\hat{\beta}$ יהיו אומדנים מוטים לפרמטרים α ו- β באוכלוסיה.

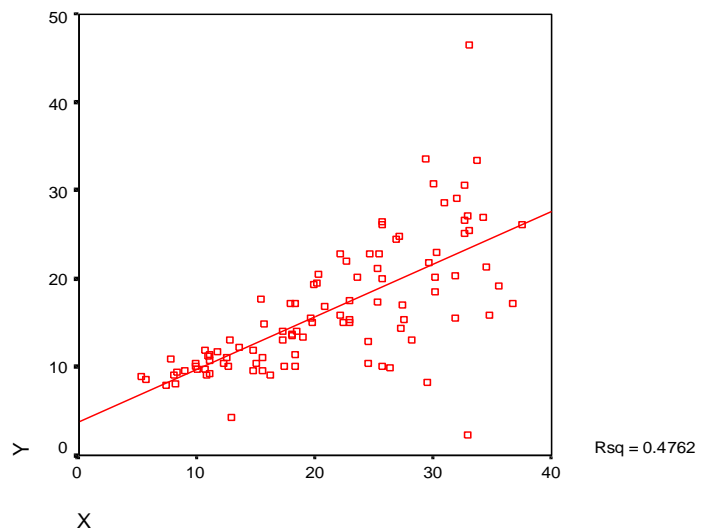
ב. אי עמידה בהנחת שוויון שוניות:

Scatterplot

Dependent Variable: Y



במקרה זה שונות ערכי Y איננה שווה לאורך קו הרגרסיה:



במקרה זה האומדנים $\hat{\alpha}$ ו- $\hat{\beta}$ יהיו אומדנים בלתי יעילים לפרמטרים α ו- β באוכלוסיה.

2.2 בדיקת השערות למובהקות משוואת הרגרסיה באוכלוסיה

מבחן סטטיסטי שמטרתו לבדוק האם קו הרגרסיה שבנינו במדגם מובהק באוכלוסיה.

3.2.1 השערות

$$H_0; \beta = 0$$

$$H_1; \beta \neq 0$$

ההשערות מנוסחות במונחי הפרמטר β , שיפוע הקו. האפקט במודל הרגרסיה תלוי במקדם השיפוע. כאשר הוא שונה מאפס יש אפקט של ניבוי לקו הרגרסיה ואילו כאשר השיפוע = 0 אין אפקט ניבוי לקו הרגרסיה.

3.2.2 סטטיסטי המבחן

$$F = \frac{MSREG}{MSRES}$$

רציונאל המבחן זהה לרציונאל של ניתוח שונות חד כיוונית:

$$E(MSREG) = \Delta + \sigma_e^2$$

$$E(MSRES) = \sigma_e^2$$

תחת $H_0 (\Delta = 0)$:

$$E(MSREG) = \sigma_e^2 = E(MSRES)$$

3.2.3 חישוב הסטטיסטי

את סטטיסטי המבחן F נחשב באמצעות טבלת ה-ANOVA :

מקור השונות	דרגות חופש df	SS	MS=SS/df	$F = \frac{MSREG}{MSRES}$ סטטיסטי
רגרסיה reg	k=1	SSreg	MSreg	
טעויות res	n-k-1=n-2	SSres	MSres	$F_c \alpha; (1, n-2)$ קריטי
אמיתית y	n-1	SSy		

ניתן לחשב את סטטיסטי המבחן F גם באמצעות נוסחה מקוצרת (כאשר אנו לא נדרשים לבנות את טבלת ה-ANOVA):

$$F = \frac{r^2}{(1-r^2)/n-2}$$

3.2.4 כלל הכרעה ומסקנה

נדחה את H_0 אם F מחושב $F < F_{קריטי}$

מסקנה: יש/אין עדות לכך שמשוואת הרגרסיה לניבוי \hat{Y} על ידי X מובהקת באוכלוסיה ברמת מובהקות של α .

? מבוססת על דוגמא מס' 3: ניבוי מחיר הדירה על ידי שטחה.

בבדיקת הקשר בין המשתנים התקבלו הנתונים הבאים:

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
SIZE square meter	112	116.7404	39.13942	50.46	179.76
PRICE thousands \$	112	185.0664	44.45345	86.20	286.56

Correlations^a

		SIZE square meter	PRICE thousands \$
SIZE square meter	Pearson Correlation	1	.823**
	Sig. (2-tailed)	.	.000
PRICE thousands \$	Pearson Correlation	.823**	1
	Sig. (2-tailed)	.000	.

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=112

- א. בדוק האם המשוואה לניבוי מחיר הדירה (באלפי דולרים) על ידי שטח הדירה (במטר מרובע) מובהקת באוכלוסיה ברמת מובהקות של אלפא=0.01 (רשום הנחות, הצב השערות, רשום טבלת מקור שונות והסק מסקנה).
- ב. בדוק את מובהקות המשוואה על סמך הפלט הבא:

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.823 ^a	.677	.674	25.39155

a. Predictors: (Constant), SIZE square meter

b. Dependent Variable: PRICE thousands \$

כאשר נתון בנוסף ש: $n=112$

1. עם חישוב טבלת מקור שונות.

2. ללא צורך בחישוב טבלת מקור שונות.

ג. בנה טבלת מקור שונות על פי הפלט הבא:

Descriptive Statistics

	Mean	Std. Deviation	N
SIZE square meter	116.740	39.139	112
PRICE thousands \$	185.066	44.453	112
PRE_1 Unstandardized Predicted Value	185.066	36.568	112
RES_1 Unstandardized Residual	.000	25.277	112

2.3 בדיקת השערות למובהקות מקדם השיפוע באוכלוסייה

מבחן סטטיסטי שמטרתו לבדוק האם מקדם השיפוע מובהק באוכלוסייה.

2.3.1 השערות

$$H_0; \beta = 0$$

$$H_1; \beta \neq 0$$

הערה: שימו לב כי ניתן לשער גם השערות חד צדדיות וכמו כן השערות המייחסות ערך כלשהו (שאינו 0) לשיפוע.

2.3.2 סטטיסטי המבחן

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta}{\sqrt{\frac{MSRES}{SSX}}}$$

2.3.3 כלל הכרעה ומסקנה

נדחה את H_0 אם:

$$t_{\hat{\beta}} > \pm t\left(\frac{\alpha}{2}\right)(n-2)$$

הערה: במבחן חד צדדי יש לקחת α שלמה.

מסקנה: יש/אין עדות לכך ששיפוע קו הרגרסיה לניבוי \underline{Y} על ידי \underline{X} מובהק באוכלוסייה ברמת מובהקות של $\underline{\alpha}$.

2.3.4 גורמים המשפיעים על טעות התקן של $\hat{\beta}$

$$\hat{s}_{\hat{\beta}} = \sqrt{\frac{MSRES}{SSX}}$$

טעות התקן של $\hat{\beta}$ ($\hat{s}_{\hat{\beta}}$) מושפעת משני גורמים:

- א. במונה-האומדן לשונות הטעות - MSRES. קיים קשר ישיר בינו לבין $\hat{s}_{\hat{\beta}}$.
 - ב. במכנה-סכום הסטיות הריבועיות של X - SSX. קיים קשר הפוך בינו לבין $\hat{s}_{\hat{\beta}}$.
- קיים קשר הפוך בין ה-MSRES לבין עוצמת הקשר הליניארי בין המשתנים $(|r_{xy}|)$.
- קשור לתופעת קיצוץ תחום; כאשר מקוצץ תחום ערכי X (הב"ת), ה-SSX קטן.

2.3.5 רב"ס ל- β

כאשר אנו מתבקשים לאמוד את שיפוע המשוואה, נבנה רווח בר סמך ל- β :

$$\hat{\beta} \pm t_{n-2} \left(\frac{\alpha}{2}\right) \sqrt{\frac{MSRES}{SSX}}$$

??

1) על סמך נתוני דוגמא מס' 3 (ניבוי מחיר הדירה על סמך שטחה):

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
SIZE square meter	112	116.7404	39.13942	50.46	179.76
PRICE thousands \$	112	185.0664	44.45345	86.20	286.56

Correlations^a

		SIZE square meter	PRICE thousands \$
SIZE square meter	Pearson Correlation	1	.823**
	Sig. (2-tailed)	.	.000
PRICE thousands \$	Pearson Correlation	.823**	1
	Sig. (2-tailed)	.000	.

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=112

א. בדוק האם השיפוע לניבוי מחיר הדירה על ידי שטחה מובהק באוכלוסיה ברמת מובהקות של 0.05.

ב. אמוד את השיפוע באוכלוסייה ברמת ביטחון של 0.95.

(2) חווה דעתך על הטענות הבאות:

א. ברגרסיה חד משתנית, ככל שנדגום טווח ערכים גדול יותר במשתנה הב"ת, אזי נגדיל את הפיזור ועקב כך ברוב המקרים שונות הטעויות תגדל ויהיה קשה יותר לדחות את השערת האפס למובהקות מקדם בטא.

ב. השונות של התפלגות הדגימה של בטא גדלה ככל שהקשר בין המשתנה התלוי למשתנה הבלתי תלוי חלש יותר.

2.3.6 פלט המקדמים (coefficients)

להלן פלט המציג את מקדמי הרגרסיה ומובהקותם בציוני גלם ובציוני תקן:

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	75.997	7.578		10.028	.000	60.979	91.015
	SIZE square meter	.934	.062	.823	15.173	.000	.812	1.056

a. Dependent Variable: PRICE thousands \$

**שימו לב לקשר בין סטטיסטי המבחן T לערך המקדם וטעות התקן:

$$t = \frac{B}{Std.Error}$$

2.4 קשר בין התפלגויות T ו-F

ברגרסיה חד משתנית קיימת זהות בין שלושת המבחנים הסטטיסטיים הבאים:

(1) מבחן F למובהקות משוואת הרגרסיה

(2) מבחן T למובהקות מקדם השיפוע (β)

(3) מבחן למובהקות מקדם המתאם (η)

כל אחד מן המבחנים מעיד על האחרים: ה-sig בשלושת שווה.

בנוסף קיים קשר בין סטטיסטי המבחן T ו-F (המבחן הראשון והשני):

$$F(1, n - 2) = t_{n-2}^2 \left(\frac{\alpha}{2} \right)$$

על פי נתוני דוגמא מס' 2 (ניבוי מחיר הדירה על פי שטחה) ניתן לראות כי:

סטטיסטי המבחן F שחישבנו למובהקות משוואת הרגרסיה שווה לריבוע סטטיסטי המבחן T שחישבנו למובהקות מקדם השיפוע:

$$F = 230.9 = t^2 = 15.2^2$$

ניתן לראות בנוסף כי ה-sig של שלושת המבחנים שווה:

ה-sig של מובהקות המשוואה (המופיע בפלט ה-ANOVA):

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	148427.7	1	148427.680	230.216	.000 ^a
	Residual	70920.396	110	644.731		
	Total	219348.1	111			

a. Predictors: (Constant), SIZE square meter

b. Dependent Variable: PRICE thousands \$

שווה ל-sig של מובהקות השיפוע (שורה שנייה בפלט ה-coefficients):

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	75.997	7.578		10.028	.000	60.979	91.015
	SIZE square meter	.934	.062	.823	15.173	.000	.812	1.056

a. Dependent Variable: PRICE thousands \$

ול-sig של מובהקות מקדם המתאם בין שני המשתנים (המופיע בפלט ה-correlations):

Correlations^a

		SIZE square meter	PRICE thousands \$
SIZE square meter	Pearson Correlation	1	.823**
	Sig. (2-tailed)	.	.000
PRICE thousands \$	Pearson Correlation	.823**	1
	Sig. (2-tailed)	.000	.

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=112

2.5 רב"ס לתוחלת ערכי Y (μ_0) עבור ערך מסוים של X

רב"ס שמטרתו לאמוד את תוחלת ערכי המשתנה התלוי (μ_0) עבור ערך מסוים של המשתנה ה"ב"ת (x_0). במילים אחרות אנו מתבקשים לאמוד את הניבוי באוכלוסייה עבור ערך מסוים של X.

האומד הנקודתי (הסטטיסטי) סביבו בנוי הרב"ס הוא הניבוי במדגם עבור אותו ה-X:

$$\hat{\mu}_0 = \hat{y}_0 = \hat{\alpha} + \hat{\beta}x_0$$

נוסחת הרב"ס:

$$\hat{\mu}_0 \pm t_{n-2} \left(\frac{\alpha}{2} \right) \cdot \sqrt{MSRES \cdot \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SSX} \right)}$$

טעות התקן/גודל הרב"ס מושפעים מ-4 גורמים:

- MSRES - האומד לשונות הטעויות. ככל שגדל, טעות התקן/ הרב"ס גדלים ולהפך.
- n - גודל המדגם. ככל שגדל, טעות התקן/ הרב"ס קטנים ולהפך.
- SSX - מונה השונות של X (קשור לתופעת קיצוץ תחום). ככל שגדל, טעות התקן/ הרב"ס קטנים ולהפך.
- $(x_0 - \bar{x})$ - הסטייה של ערך X המסוים מהמוצע של X. ככל שגדלה טעות התקן/ הרב"ס גדלים ולהפך.

? השאלה מבוססת על נתוני דוגמא מס' 2 (ניבוי הקשר בין מחיר הדירה על סמך שטחה) והפלטים הבאים:

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
SIZE square meter	112	116.7404	39.13942	50.46	179.76
PRICE thousands \$	112	185.0664	44.45345	86.20	286.56

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	75.997	7.578		10.028	.000	60.979	91.015
	SIZE square meter	.934	.062	.823	15.173	.000	.812	1.056

a. Dependent Variable: PRICE thousands \$

Descriptive Statistics

	Mean	Std. Deviation	N
SIZE square meter	116.740	39.139	112
PRICE thousands \$	185.066	44.453	112
PRE_1 Unstandardized Predicted Value	185.066	36.568	112
RES_1 Unstandardized Residual	.000	25.277	112

- א. חשב רב"ס ברמת סמך של 95% לתוחלת מחיר הדירה כאשר שטח הדירה הוא 100 מ"ר.
- ב. תן אומדן נקודתי לתוחלת מחיר הדירה עבור שטח הדירה 105 מ"ר ועבור שטח דירה של 70 מ"ר. מי משני האומדנים הללו יעיל ומדויק יותר בניבוי התוחלת של מחיר הדירה? נמק.
- ג. האם נוכל לנבא על סמך משוואת הרגרסיה הנ"ל את מחיר הדירה עבור שטח של 40 מ"ר?
- ד. מהו האומדן לשינוי בתוחלת מחיר הדירה עבור עליה ביחידה אחת של שטח הדירה?
- ה. מהו האומדן להפרש הצפוי בתוחלת מחיר הדירה בין שטח דירה ששווה ל-120 מ"ר לבין שטח דירה ששווה ל-90 מ"ר?
- ו. אמוד את שונות מחיר הדירה עבור שטח של 100 מ"ר. על איזה הנחה מהנחות המודל מבוסס אומדן זה?

2.6 רב"ס לערכי Y עבור ערך מסוים של X

רב"ס שמטרתו לאמוד את כל טווח ערכי Y (y_0) עבור ערך X מסוים (x_0).

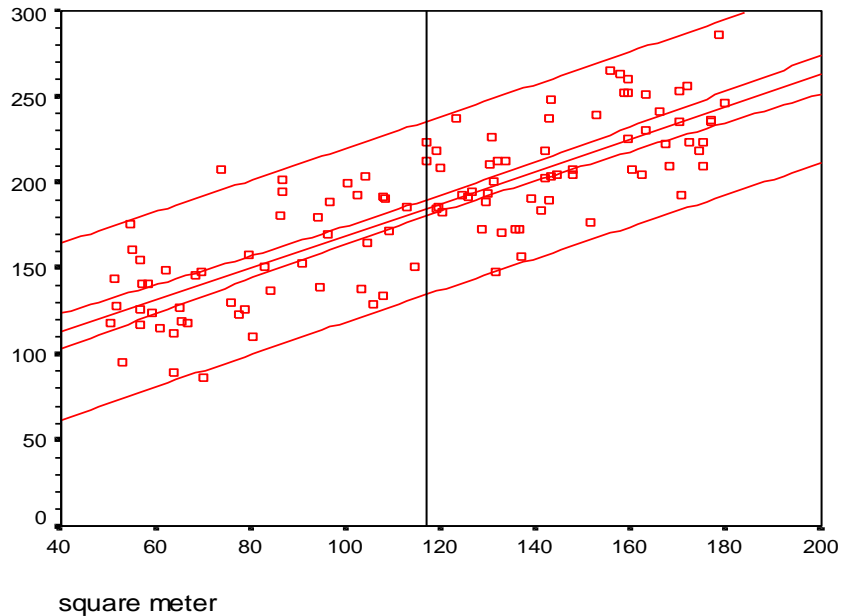
נוסחת הרב"ס:

$$\hat{\mu}_0 \pm t_{n-2} \left(\frac{\alpha}{2} \right) \cdot \sqrt{MSRES \cdot \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{SSX} \right)}$$

ניתן לראות כי גם רב"ס זה בנוי סביב האומדן הנקודתי לתוחלת ערכי Y עבור ערך ה-X המסוים ($\hat{\mu}_0$).

ההבדל בין רב"ס לערכי Y לבין הרב"ס לתוחלת ערכי Y בא לידי ביטוי בטעות התקן. ניתן לראות כי טעות התקן של הרב"ס לערכי Y גדולה יותר מטעות התקן של הרב"ס לתוחלת ערכי Y. כאשר כל יתר הפרמטרים נשארים קבועים רב"ס זה יהיה רחב יותר מן הרב"ס לתוחלת.

התרשים הבא מתאר רב"ס לתוחלת ולערך המשתנה התלוי וממחיש זאת בבירור:



:

? חשב רב"ס ברמת ביטחון של 95% למחיר הדירה עבור שטח דירה של 100 מ"ר. מה ההבדל בין רב"ס זה לרב"ס הקודם?

2.7 שאלות מסכמות

1) בנק הפועלים ערך מחקר לבדיקת השפעת הפעילות בבורסה על הפעילות בחשבונות החיסכון של לקוחותיו. לשם כך נדגמו 115 נבדקים מסניפים שונים ברחבי הארץ.

להלן המשתנים שנחקרו:

ISKA3 - פעילות בבורסה

ISKA6 - פעילות בחשבונות חיסכון

להלן הממצאים:

	N	Mean	Std. Deviation
ISKA3	115	234.7183	73.8843
ISKA6	115	167.5057	
Valid N (listwise)	115		

Model Summary:

Model	R	R Square	Adjusted R Square	Error Std. Of the Estimate
1	0.339	0.115	0.107	48.3545

Predictors: (constant), ISKA3

Dependent Variable: ISKA6

**נתון בנוסף כי במבחן F לבדיקת מובהקות המשוואה התקבל $\text{sig}=0.021$

א. בדוק האם שיפוע הרגרסיה לניבוי פעילות בחשבונות חיסכון ע"י פעילות בבורסה מובהקת ברמת מובהקות של אלפא=0.05, על סמך הנתונים בלבד (רשום הנחות, הצב השערות והסק מסקנה מילולית).

ב. חשב את האומדן חסר ההטיה לשונות של ערכי פעילות בחשבון חיסכון.

ג. בנה את משוואת הרגרסיה לניבוי פעילות בחשבונות חיסכון ע"י פעילות בבורסה.

ד. מהו האומדן הנקודתי לתוחלת פעילות בחשבון חיסכון עבור לקוח שפעילותו בבורסה היא 210?

2) מכון מחקר בחן עבור חברה מסוימת את הקשר בין סך מכירותיה החודשיות (באלפי ש"ח) לבין הוצאותיה על פרסום בחודש (באלפי ש"ח). לשם כך נדגמו 25 חודשים ונבדקו המשתנים הבאים:

Advertise expenses

Sales

להלן הממצאים:

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-41.586	56.465		-.736	.469
	X advertise expenses	10.348	1.745	.778	5.930	.000

a. Dependent Variable: Y sales

	N	Mean	Std .Deviation
Advertise expenses	25	28.1040	
Sales	25	249.2400	217.85283
Valid N (listwise)	25		

- א. האם המשוואה לניבוי המכירות על ידי הוצאות הפרסום מובהקת באוכלוסיה? בדוק ברמת מובהקות של 0.05 על פי נתוני הפלט בלבד (רשום הנחות, הצב השערות והסק מסקנה מילולית).
- ב. מהו ערך סטטיסטי המבחן F למובהקות המשוואה? ענה ללא חישוב.
- ג. מהו האומדן לשונות ערכי המכירות כאשר הוצאות הפרסום מצויות במשוואה?
- ד. בנה רב"ס ברמת ביטחון של 95% לתוחלת המכירות עבור הוצאות פרסום של 35 אלף ₪.
- ה. מהו ערכו של המתאם בין האומדנים של המכירות לבין ערכי הוצאות הפרסום?
- ו. מהו ערכו של המתאם בין האומדנים של המכירות לבין ערכי המכירות?
- ז. מהו ערכו של המתאם בין המכירות לבין הטעות בניבוי המכירות?
- ח. מהו האומדן לשינוי בתוחלת ערכי המכירות עבור עליה ביחידה אחת בהוצאות הפרסום?

רגרסיה מרובה

רגרסיה מרובה

כאשר אנו מנסים לנבא את המשתנה התלוי ביותר ממשתנה ב"ת אחד מדובר ברגרסיה מרובה.

חלק זה יתחלק ל-3 חלקים:

- (1) מאפייני קו הרגרסיה במדגם
- (2) מובהקות קו הרגרסיה ומקדמיו באוכלוסיה
- (3) שיטות להרצת רגרסיה רבת משתנים

הערה: החלק הראשון והשני יעסקו בעיקר ברגרסיה דו משתנית: ניבוי המשתנה התלוי באמצעות שני משתנים ב"ת בלבד.

1. מאפייני קו הרגרסיה במדגם

נוסחת קו הרגרסיה המרובה:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \dots + \hat{\beta}_k x_{ki}$$

האומדן לניבוי (\hat{y}) מהווה קומבינציה ליניארית של ערכי X.

מספר השיפועים (ה- $\hat{\beta}$) הם כמספר המשתנים ה"ת ואילו קבוע הרגרסיה ($\hat{\alpha}$) הוא אחד.

1.1 מקדמי הרגרסיה

נציג את מקדמי קו הרגרסיה המרובה בציוני גלם ובציוני תקן:

המקדם	ציוני גלם	ציוני תקן
אלפא	$\hat{\alpha} = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2 - \dots - \hat{\beta}_k \bar{x}_k$	$z\hat{\alpha} = 0$
המשמעות	קבוע החיתוך של קו הרגרסיה עם ציר ה-Y. כאשר כל ה-Xים שווים ל-0 אז: $\hat{y}_i = \hat{\alpha}$	קו הרגרסיה יוצא מראשית הצירים
בטה "מקדמי רגרסיה חלקיים"	$\hat{\beta}_1 = \frac{(r_{y1} - r_{y2} * r_{12})}{1 - r_{12}^2} * \frac{\hat{s}_y}{\hat{s}_{x1}}$ $\hat{\beta}_2 = \frac{(r_{y2} - r_{y1} * r_{12})}{1 - r_{12}^2} * \frac{\hat{s}_y}{\hat{s}_{x2}}$ <u>הערה:</u> ניתן לחישוב רק ברגרסיה דו משתנית	$z\hat{\beta}_1 = \frac{(r_{y1} - r_{y2} * r_{12})}{1 - r_{12}^2}$ $z\hat{\beta}_2 = \frac{(r_{y2} - r_{y1} * r_{12})}{1 - r_{12}^2}$

<p>התרומה היחסית של המשתנה הב"ת לניבוי המשתנה התלוי.</p> <p>השוואת הבטות המתוקננות בערך מוחלט- $z\hat{\beta}$ מאפשרת להעריך את התרומה היחסית של כל אחד מהמשתנים הב"ת לניבוי המשתנה התלוי.</p> <p>הסיבה שאנו משווים בין השיפועים המתוקננים ולא הגולמיים היא כי המתוקננים בניגוד לגולמיים מנוקים מיחידות המדידה של המשתנים.</p>	<p>התרומה הייחודית של המשתנה הב"ת לניבוי של המשתנה התלוי (בניכוי שאר המשתנים הב"ת המצויים במשוואה).</p>	<p>המשמעות</p>
<p> $z\hat{\beta}_1 = r_{y1}$ $z\hat{\beta}_2 = r_{y2}$ </p> <p>הבטה המתוקננת שווה למתאם הפשוט.</p>	<p> $\hat{\beta}_1 = r_{y1} \cdot \frac{\hat{s}_y}{\hat{s}_{x1}}$ $\hat{\beta}_2 = r_{y2} \cdot \frac{\hat{s}_y}{\hat{s}_{x2}}$ </p> <p>שיפועי הרגרסיה המרובה שווים לשיפועי הפשוטה</p>	<p>כאשר המתאם בין המשתנים הב"ת = $r_{12} = 0$</p>

יציבות השיפועים (הבטות): כאשר נכניס משתנים ב"ת לתוך משוואת הרגרסיה, נשאלת השאלה מה יקרה לבטות של המשתנים הב"ת שכבר מצויים בתוך המשוואה?

התשובה: ככל שהמתאם בין המשתנה הב"ת שנכניס עם משתנים ב"ת אחרים שכבר מצויים במשוואה יהיה חזק יותר כך הבטות יהיו פחות יציבות/קבועות בהכנסתו ולהפך.

? דוגמה 1: ניבוי השכר הנוכחי באמצעות רמת ההשכלה והוותק במקום העבודה:

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
SALNOW Current salary	121	8965.6734	1392.38927	5412.49	12207.19
EDUCATE Level of education in years	121	13.7438	2.76746	8.00	20.00
TIME Job seniority in months	121	76.1653	13.22267	38.00	110.00

Correlations^a

			SALNOW Current salary	EDUCATE Level of education in years	TIME Job seniority in months
SALNOW Current salary	Pearson Correlation		1	.551**	.661**
	Sig. (2-tailed)		.	.000	.000
EDUCATE Level of education in years	Pearson Correlation		.551**	1	.008
	Sig. (2-tailed)		.000	.	.927
TIME Job seniority in months	Pearson Correlation		.661**	.008	1
	Sig. (2-tailed)		.000	.927	.

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=121

א. חשב את משוואת הרגרסיה לניבוי השכר הנוכחי על ידי רמת ההשכלה והוותק בציוני גלם ובציוני תקן.

1. מה יהיה הניבוי, בציוני גלם ובציוני תקן, של השכר הנוכחי לאדם שרמת השכלתו = 14 שנים והוותק שלו בעבודה = 75 חודשים? מה דעתך על מידת הדיוק של ניבוי זה?

2. אמוד את הפרש בשכר הנוכחי של שני עובדים בעלי אותה רמת השכלה אך מס' חודשי הוותק של הראשון גבוה פי 1.5 ממספר חודשי הוותק של השני.

3. למי משני המשתנים הב"ת-רמת השכלה או וותק תרומה ייחודית גבוהה יותר לניבוי המשתנה המנובא?

4. בשלב הראשון הוחלט לנבא את השכר הנוכחי ע"י רמת ההשכלה בלבד. בנה את משוואת הרגרסיה.

5. על סמך הפלט בלבד: האם המשוואה שבנית בסעיף הקודם מובהקת ברמת מובהקות של אלפא = 0.05? (הצב השערות, רשום הנחות והסק מסקנה מילולית בלבד).

6. בשלב השני הוסיף החוקר את המשתנה וותק. האם ישתנה מקדם הרגרסיה החלקי של רמת ההשכלה לאחר הכנסת הוותק לתוך המשוואה? הסבר ללא חישוב.

? חוקר ערך מחקר שבו ניבא את Y על ידי שני משתנים ב"ת. להלן הנתונים:

$$r_{12} = 0$$

$$\bar{x}_1 = 20, \bar{x}_2 = 50$$

$$\hat{s}_{x_1}^2 = 9, \hat{s}_{x_2}^2 = 25$$

המשוואה המתקננת שהתקבלה: $\hat{z}y_i = 1.8zx_{1i} - 1.2zx_{2i}$

האם המשוואה הנ"ל הגיונית? חווה דעתך.

? חווה דעתך על הטענה הבאה: המשתנה בעל התרומה היחסית הנמוכה ביותר במשוואת הרגרסיה המונה 3 משתנים ב"ת לפחות, הוא בהכרח זה בעל הקשר החלש ביותר עם המשתנה התלוי.

1.2 אומדנים חשובים ברגרסיה מרובה

1.2.1 המתאם המרובה R

עד עתה הכרנו את מקדם המתאם r המכונה גם ה"מתאם הפשוט" המתייחס לקשר בין שני משתנים בלבד. המתאם המרובה R מתייחס לקשר בין המשתנה התלוי לכמה משתנים ב"ת, למשל: R_{y123} . (שימו לב כי המשתנים הב"ת (ה-איים) מסומנים במספרים).

בניגוד למתאם הפשוט (r) המתאם המרובה (R) מגדיר רק את עוצמת הקשר ולא את כיוונו ($0 \leq R \leq 1$). בנוסף לכך לעומת המתאם הפשוט המהווה מקדם מתאם סימטרי (לא חשוב מיהו המשתנה התלוי ומיהו הב"ת: $r_{xy} = r_{yx}$), מקדם המתאם המרובה הוא א- סימטרי.

1.2.2 פרופורציית השונות המוסברת R^2

משמעותה (כמו ברגרסיה פשוטה) כמה מתוך השונות של המשתנה התלוי הצליחו המשתנים הב"ת להסביר:

$$0 \leq R^2 = \frac{ssreg}{ssy} \leq 1$$

כאשר ישנם שני משתנים ב"ת בלבד (רגרסיה דו משתנית), ניתן לחשב את פרופורציית השונות המוסברת גם על ידי הנוסחה הבאה:

$$R_{y12}^2 = \frac{r_{y1}^2 + r_{y2}^2 - 2 \cdot r_{y1} \cdot r_{y2} \cdot r_{12}}{1 - r_{12}^2}$$

המתאם המרובה מושפע מהתלות הסטטיסטית בין המשתנים הב"ת:

$$1. \text{ כאשר } r_{12} = 0$$

פרופורציית השונות המוסברת של Y על ידי שני המשתנים הב"ת, תהיה שווה לסכום

$$\text{הפרופורציות המוסברות ה"פשוטות": } R_{y12}^2 = r_{y1}^2 + r_{y2}^2$$

2. כאשר $r_{12} \neq 0$:

פרופורציית השונות המוסברת של Y ע"י שני המשתנים הב"ת, תהיה קטנה מסכום הפרופורציות

$$R_{y12}^2 < r_{y1}^2 + r_{y2}^2 : \text{"פשוטות"}$$

תכונה חשובה של R^2 : בהוספת משתנים ב"ת לתוך משוואת הרגרסיה יכול רק לגדול ("להתנפח") או להישאר אותו דבר אך לא יוכל לעולם לקטון.

1.2.3 האומד המתוקן (חסר ההטיה) לפרופורציית השונות המוסברת- $AdjR^2$

האומד המתוקן לפרופורציית השונות המוסברת- $AdjR^2$ בניגוד לאומד המוטה- R^2 עלול לקטון (ואף לקבל ערכים שליליים) בהוספת משתנים ב"ת לא איכותיים לתוך משוואת הרגרסיה (כאלה שהם בעלי קשר נמוך עם Y וקשרים גבוהים עם X ים אחרים המצויים במשוואה).

לכן $AdjR^2$ משמש ככלי חשוב כאשר נרצה להעריך האם כדאי לנו להכניס משתנה/משתנים ב"ת מסוימים לתוך משוואת הרגרסיה או בכדי לבחור את המשוואה האופטימאלית במדגם.

? במחקר מסוים מנבאים את ציוני ה-B.A באמצעות ציוני הבגרות בלבד כאשר פרופורציית השונות המוסברת שהתקבלה היא: $r_{xy}^2 = 0.64$. בשלב השני הוחלט להוסיף למשוואת הניבוי משתנה מנבא נוסף-פסיכומטרי והתקבל ש:

$$א. \quad AdjR_{y12}^2 = 0.611$$

$$ב. \quad R_{y12}^2 = 0.611$$

חווה דעתך לגבי שני האומדנים הנ"ל לפרופורציית השונות המוסברת של ה-B.A על ידי פסיכומטרי ובגרות- האם הגיוניים? נמק.

תשובה:

האומד הראשון הגיוני ואילו האומד השני איננו הגיוני. שני האומדים מצביעים על ירידה בפרופורציית השונות המוסברת בהוספה של משתנה ב"ת נוסף לתוך המשוואה- הפסיכומטרי. אולם רק האומד המתוקן לפרופורציית השונות המוסברת ($AdjR^2$) יכול לקטון בהוספת משתנים ב"ת לתוך משוואת הרגרסיה. האומד המוטה (R^2) יכול רק לגדול ("להתנפח") ובכל מקרה לא יכול לקטון.

? על סמך נתוני דוגמה 1 (ניבוי השכר הנוכחי באמצעות רמת ההשכלה והוותק במקום העבודה):

Case Summaries

	N	Mean	Std. Deviation	Minimum	Maximum
SALNOW Current salary	121	8965.6734	1392.38927	5412.49	12207.19
EDUCATE Level of education in years	121	13.7438	2.76746	8.00	20.00
TIME Job seniority in months	121	76.1653	13.22267	38.00	110.00

Correlations^a

		SALNOW Current salary	EDUCATE Level of education in years	TIME Job seniority in months
SALNOW Current salary	Pearson Correlation Sig. (2-tailed)	1 .	.551** .000	.661** .000
EDUCATE Level of education in years	Pearson Correlation Sig. (2-tailed)	.551** .000	1 .	.008 .927
TIME Job seniority in months	Pearson Correlation Sig. (2-tailed)	.661** .000	.008 .927	1 .

** . Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=121

חשב את האומדנים הבאים:

- א. האומדן לפרופורציית השונות הלא מוסברת של השכר הנוכחי ע"י רמת ההשכלה והוותק**
- ב. א.ח.ה לשונות של ההשכר הנוכחי**
- ג. א.ח.ה לשונות של השכר הנוכחי כאשר רמת ההשכלה והוותק מצויים במשוואה**
- ד. האומדן לשונות של השכר הנוכחי עבור רמת השכלה=14 שנים וותק=68 חודשים**

2. מובהקות קו הרגרסיה ומקדמיו באוכלוסיה

2.1 הנחות מודל הרגרסיה המרובה

1. דגימה מקרית של תצפיות/ טעויות ב"ת.

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$$

2. עבור כל קומבינציה ליניארית קבועה של ערכי X. $\varepsilon_i \approx N(0, \sigma_e^2)$

3. אין מתאם מלא בין המשתנים הב"ת.

2.2 מובהקות משוואת הרגרסיה המרובה

נבצע מבחן זה כשנרצה לבדוק האם משוואת הרגרסיה שקיבלנו במדגם מובהקת באוכל'.

השערות:

$$H_0; \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1; \text{not } H_0$$

**מספיק שבטח אחת שונה מ-0 כדי שהמשוואה תהיה מובהקת

סטטיסטי המבחן:

$$F = \frac{MS_{REG}}{MS_{RES}}$$

נחשבו באמצעות טבלת ANOVA:

מקור השונות	דרגות חופש df	SS	MS=SS/df	$F = \frac{MS_{REG}}{MS_{RES}}$ סטטיסטי
רגרסיה reg	K	SSreg	MSreg	
טעויות res	n-k-1	SSres	MSres	$F_{\alpha; (k, n-k-1)}$ קריטי
אמיתית y	n-1	SSy		

הערה: ניתן לחשב את סטטיסטי F גם באמצעות הנוסחה המקוצרת:

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

כלל הכרעה ומסקנה:

אם $F > F_c$ יש סיבה מספקת לדחות את H_0 .

מסקנה: יש/אין עדות לכך שמשוואת הרגרסיה מובהקת באוכלוסייה ברמת מובהקות של α

? בדוק האם המשוואה לניבוי השכר הנוכחי של העובד ע"י ההשכלה והוותק שלו מובהקת באוכ' ברמת מובהקות של אלפא=0.05 (הצב השערות, רשום הנחות, בנה טבלת מקור שונות והסק מסקנה). בנה את טבלת מקור השונות על סמך ה- Model Summary בהינתן גם $n=121$:

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.857 ^a	.735	.731	722.66594

a. Predictors: (Constant), TIME Job seniority in months, EDUCATE Level of education in years

b. Dependent Variable: SALNOW Current salary

2.3 מובהקות השיפועים

2.3.1 המבחן הסטטיסטי

נבצע מבחן סטטיסטי זה כאשר נרצה לבדוק האם מקדם השיפוע (הבטה) של אחד המשתנים הב"ת מובהק באוכלוסייה.

השערות:

$$H_0; \beta_j = 0$$

כאשר המשתנים הב"ת האחרים מצויים במשוואה

$$H_1; \beta_j \neq 0$$

**הערה: ניתן לשער גם השערות חד צדדיות וגם על בטה ששווה לערך מסוים באוכלוסייה.

סטטיסטי המבחן:

ניתן לחישוב רק ברגרסיה דו משתנית (עבור יותר משני משתנים ב"ת נשתמש בפלטים בלבד).

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - \beta}{\sqrt{\frac{MSRES}{SSX_1(1 - r_{12}^2)}}}$$

$$t_{\hat{\beta}_2} = \frac{\hat{\beta}_2 - \beta}{\sqrt{\frac{MSRES}{SSX_2(1 - r_{12}^2)}}}$$

כלל הכרעה ומסקנה:

אם $t > t_c\left(\frac{\alpha}{2}, n - k - 1\right)$ או $-t < -t_c\left(\frac{\alpha}{2}, n - k - 1\right)$ נדחה את H_0

מסקנה: יש/אין עדות לכך ששיפוע המשתנה הב"ת המסוים מובהק באוכלוסיה כאשר שאר המשתנים הב"ת מצויים במשוואה, ברמת מובהקות של α .

? האם השיפוע של רמת ההשכלה מובהק כאשר הוותק מצוי במשוואה? (הצב השערות, רשום הנחות, קבע כלל הכרעה והסק מסקנה).

פטור על פי ה- Model summary :

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.857 ^a	.735	.731	722.66594

a. Predictors: (Constant), TIME Job seniority in months, EDUCATE Level of education in years

b. Dependent Variable: SALNOW Current salary

כאשר נתון ש:

$$z\hat{\beta}_2 = 0.652$$

$$r_{12} = 0.008$$

	N	S.Deviation
Y	121	1392.389
X1	121	2.767
X2	121	13.222

? האם התרומה הייחודית של הוותק לניבוי השכר הנוכחי מובהקת? הצב השערות, רשום הנחות והסק מסקנה מילולית על סמך הפלט בלבד וללא חישוב.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-74.959	503.950		-.149	.882
	EDUCATE Level of education in years	274.676	23.839	.546	11.522	.000
	TIME Job seniority in months	69.133	4.989	.657	13.856	.000

a. Dependent Variable: SALNOW Current salary

? לאיזה מבין המשתנים הב"ת במודל יש את התרומה הייחודית הגבוהה ביותר לניבוי המשתנה התלוי?

2.3.2 רב"ס ל- $\hat{\beta}_j$

$$\hat{\beta}_j \pm t\left(\frac{\alpha}{2}, n - k - 1\right) \cdot S.E.(\hat{\beta}_j)$$

? אמוד את השיפוע של הוותק באוכלוסיה ברמת ביטחון של 0.05.

2.3.3 מובהקות השיפוע-השוואה בין רגרסיה פשוטה למרובה

נשאלת השאלה-אם הצלחנו לדחות את H0 במבחן למובהקות השיפוע של משתנה ב"ת מסוים ברגרסיה פשוטה, האם בהכרח נצליח לדחות את H0 למובהקות אותו השיפוע גם במרובה (כאשר ישנם משתנים ב"ת נוספים בתוך המשוואה)?

תשובה:

מרובה

פשוטה

$$r_{12} = 0 \quad (1) \quad S.E.(\hat{\beta}) = \sqrt{\frac{MSRES}{SSX}}$$
$$S.E.(\hat{\beta}) < S.E.(\hat{\beta}) \quad \text{פשוטה} \quad S.E.(\hat{\beta}) = \sqrt{\frac{MSRES}{SSX(1-r_{12}^2)}} = \sqrt{\frac{MSRES}{SSX}}$$

ולכן אם דחינו H_0 בפשוטה בהכרח נדחה גם במרובה.

$$r_{12} \neq 0 \quad (2)$$

$$S.E.(\hat{\beta}) = \sqrt{\frac{MSRES}{SSX(1-r_{12}^2)}}$$

פשוטה $S.E.(\hat{\beta}) > S.E.(\hat{\beta})$ מרובה

ולכן אם דחינו H_0 בפשוטה לא בהכרח שנדחה במרובה.

מסקנה: כאשר שיפוע משתנה ב"ת מסוים מובהק באוכלוסיה ברגרסיה פשוטה מובהקותו של אותו המשתנה ברגרסיה מרובה תלויה במתאם בין המשתנים הב"ת. רק במקרה שאין מתאם בין הב"ת מובהקות המשתנה בפשוטה מעידה בהכרח על מובהקותו במרובה.

? חוקר ניבא את הציונים ב-B.A על סמך המשתנים- פסיכומטרי ובגרות. בשלב הראשון הוא ניבא את הציונים ב-B.A רק על סמך הבגרות ומקדם השיפוע יצא מובהק. בשלב השני החליט להוסיף את משתנה הפסיכומטרי. לטענתו משתנה הבגרות חייב לצאת מובהק גם בשלב זה, כי שונות הטעויות קטנה. חווה דעתך על טענתו.

3. שיטות להרצת רגרסיה רבת משתנים

קיימות 3 שיטות להרצת הנתונים ברגרסיה מרובה, הנבדלות ביניהן בסדר הכנסת המשתנים הב"ת:

שיטת ENTER

רגרסיה בצעדים חכמים (STEPWISE REGRESSION)

רגרסיה היררכית

3.1 שיטת ENTER

מכניסים את כל המשתנים הב"ת למשוואה בעת ובעונה אחת.

דוגמה 2: ניבוי היקף פעולות סניף הבנק ביצוא באמצעות 5 משתנים- היקף פעולות בבורסה, יבוא, חסכונות, עמלות והמחאות נוסעים (כל המשתנים באלפי ש"ח ליום).

ניתוח הנתונים לפי שיטת Enter:

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	X5 Traveler Checks, X1 Stock Market, X4 Commission, X2 Import, X3 Saving ^a	.	Enter

a. All requested variables entered.

b. Dependent Variable: Y Export

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.506 ^a	.256	.221	64.75156

a. Predictors: (Constant), X5 Traveler Checks, X1 Stock Market, X4 Commission, X2 Import, X3 Saving

b. Dependent Variable: Y Export

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	155428.3	5	31085.659	7.414	.000 ^a
	Residual	452818.5	108	4192.764		
	Total	608246.8	113			

a. Predictors: (Constant), X5 Traveler Checks, X1 Stock Market, X4 Commission, X2 Import, X3 Saving

b. Dependent Variable: Y Export

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-259.207	71.279		-3.637	.000
	X1 Stock Market	-2.00E-02	.089	-.020	-.225	.823
	X2 Import	.257	.103	.221	2.490	.014
	X3 Saving	.312	.128	.219	2.447	.016
	X4 Commission	-2.57E-02	.107	-.021	-.240	.811
	X5 Traveler Checks	103.806	29.454	.315	3.524	.001

a. Dependent Variable: Y Export

? על סמך הפלטים הנ"ל, ענה ללא חישוב:

1. האם משוואת הרגרסיה לניבוי פעילות סניף הבנק ביצוא על ידי חמש המשתנים הב"ת מובהקת באוכלוסיה ברמת מובהקות של 0.05?
2. לאיזה מהמשתנים הב"ת תרומה ייחודית מובהקת לניבוי פעילות סניף הבנק ביצוא, ברמת מובהקות של 0.05?
3. מי מבין המשתנים הב"ת תורם את התרומה היחסית הגבוהה ביותר לניבוי פעילות סניף הבנק ביצוא?
4. מהי פרופורציית השונות המוסברת של פעילות סניף הבנק ביצוא ע"י חמשת המשתנים הב"ת?
5. מהו האומדן לשונות הטעויות בניבוי פעילות הבנק ביצוא?

3.2 רגרסיה בצעדים חכמים (stepwise)

מטרתה לבנות את משוואת הרגרסיה האופטימאלית. כלומר לבחור את המשתנים הב"ת אשר תרומתם הייחודית לניבוי המשתנה התלוי היא הגבוהה ביותר ומובהקת.

בכל צעד/שלב יכנס משתנה ב"ת אחד בלבד.

בשלב הראשון: יכנס המשתנה הב"ת שיש לו את המתאם הפשוט הגבוה ביותר עם Y ומובהק.

בשלב השני: יכנס המשתנה הב"ת שיש לו את המתאם החלקי הגבוה ביותר עם Y (בניכוי המשתנה הב"ת שנכנס בשלב הראשון למשוואה) ומובהק.

בשלב השלישי: יכנס המשתנה הב"ת שיש לו את המתאם החלקי הגבוה ביותר עם Y (בניכוי המשתנים שנכנסו בשלב הראשון והשני למשוואת הרגרסיה) ומובהק.

וכך הלאה... במידה ובכל שלב, המשתנה בעל המתאם הגבוה ביותר איננו מובהק עוצרים ולא ממשיכים להכניס משתנים.

הדגמה על פי נתוני דוגמא 2 : בניית משוואת הרגרסיה האופטימאלית לניבוי פעילות הבנק ביצוא בשיטת stepwise :

שלב ראשון -בחירת המשתנה הב"ת תעשה מתוך טבלת המתאמים:

Correlations^a

		Y Export	X1 Stock Market	X2 Import	X3 Saving	X4 Commission	X5 Traveler Checks
Y Export	Pearson Correlation	1	.126	.337**	.283**	.137	.395**
	Sig. (2-tailed)	.	.182	.000	.002	.147	.000
X1 Stock Market	Pearson Correlation	.126	1	.189*	.342**	.019	.094
	Sig. (2-tailed)	.182	.	.044	.000	.841	.318
X2 Import	Pearson Correlation	.337**	.189*	1	.161	.207*	.282**
	Sig. (2-tailed)	.000	.044	.	.086	.027	.002
X3 Saving	Pearson Correlation	.283**	.342**	.161	1	.109	.119
	Sig. (2-tailed)	.002	.000	.086	.	.249	.206
X4 Commission	Pearson Correlation	.137	.019	.207*	.109	1	.281**
	Sig. (2-tailed)	.147	.841	.027	.249	.	.002
X5 Traveler Checks	Pearson Correlation	.395**	.094	.282**	.119	.281**	1
	Sig. (2-tailed)	.000	.318	.002	.206	.002	.

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

a. Listwise N=114

ניתן לראות כי למשתנה "המחאות נוסעים" יש את המתאם הגבוה ביותר עם פעילות הבנק

$$r_{y5} = 0.395 \text{ - (ומובהק)}$$

שימו לב כי טבלת המקדמים, המתארת את המתאמים ה"פשוטים" בין Y לבין ה-Xים, טובה רק לשלב הראשון. בשלבים הבאים נבחרים ה-Xים בעלי המתאם החלקי הגבוה ביותר עם Y. את המתאמים החלקיים ניכל לראות בפלט המכונה-excluded variables:

Excluded Variables^d

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	X1 Stock Market	.089 ^a	1.025	.308	.097	.991
	X2 Import	.245 ^a	2.793	.006	.256	.920
	X3 Saving	.239 ^a	2.816	.006	.258	.986
	X4 Commission	.028 ^a	.305	.761	.029	.921
2	X1 Stock Market	.011 ^b	.118	.906	.011	.880
	X2 Import	.216 ^b	2.496	.014	.232	.904
	X4 Commission	.008 ^b	.093	.926	.009	.915
3	X1 Stock Market	-.019 ^c	-.213	.832	-.020	.864
	X4 Commission	-.020 ^c	-.229	.820	-.022	.900

a. Predictors in the Model: (Constant), X5 Traveler Checks

b. Predictors in the Model: (Constant), X5 Traveler Checks, X3 Saving

c. Predictors in the Model: (Constant), X5 Traveler Checks, X3 Saving, X2 Import

d. Dependent Variable: Y Export

ניתן לראות לפי ה-Partial Correlation כי למשתנה "חיסכון" יש את המתאם החלקי הגבוה ביותר עם פעילות הבנק ביצוא (0.258) ומובהק (sig=0.006) ולכן הוא יכנס בשלב השני לתוך מודל הרגרסיה.

באותו אופן ניתן לראות לפי ה-Partial Correlation כי למשתנה יבוא יש את המתאם החלקי הגבוה ביותר עם פעילות הבנק ביצוא (0.232) ומובהק (sig=0.014) ולכן הוא יכנס בשלב השלישי לתוך משוואת הרגרסיה.

בשלב הבא המשתנה בעל המתאם החלקי הגבוה ביותר עם פעילות הבנק ביצוא-"עמלות" (-0.022) כבר לא מובהק (sig=0.820) ולכן לא ייכנס לתוך משוואת הרגרסיה והפרוצדורה תיעצר.

הפלט הבא מציג את המשתנים שנכנסו למשוואת הרגרסיה בכל שלב בשיטת ה-stepwise:

Variables Entered/Removed^f

Model	Variables Entered	Variables Removed	Method
1	X5 Traveler Checks	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	X3 Saving	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	X2 Import	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: Y Export

? האם ניתן לדעת, ללא חישוב, האם המשוואה לניבוי פעילות הבנק ביצוא המתקבלת בשלב

הסופי על פי שיטת ה-stepwise מובהקת?

? על פי פלט ה-model summary מה יהיה ערכו של הסטטיסטי F לבדיקת מובהקות המשוואה הסופית?

3.3 רגרסיה היררכית

החוקר הוא זה שמחליט אילו משתנים יכנסו לתוך משוואת הרגרסיה בכל שלב במטרה לבדוק את השערותיו. בכל שלב יכולים להיכנס כמה משתנים ב"ת ביחד וממילא המשוואה הסופית או מקדמי השיפוע לא חייבים להיות מובהקים (בניגוד לשיטת ה-stepwise).

נדגים על סמך נתוני דוגמא מס' 2. החוקר משער כי המשתנים חיסכון, עמלות ופעילות בבורסה יתרמו תרומה מובהקת לניבוי פעילות הבנק ביצוא מעבר לתרומת המשתנים המחאות נוסעים ויבוא.

בכדי לבדוק את השערות צריך החוקר לבנות משוואת רגרסיה בשני צעדים/שלבים. בשלב הראשון להכניס את המשתנים - המחאות נוסעים ויבוא ואילו בשלב השני להכניס את שלושת המשתנים הנותרים:

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	X5 Traveler Checks, X2 Import	.	Enter
2	X3 Saving, X4 Commission, X1 Stock Market	.	Enter

- a. All requested variables entered.
b. Dependent Variable: Y Export

לפי טבלת המקדמים ניתן לראות כי בניגוד לשיטת ה-stepwise לא כל שיפועי המשתנים הב"ת בשלב האחרון מובהקים:

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-231.425	70.532		-3.281	.001
	X2 Import	.285	.102	.245	2.793	.006
	X5 Traveler Checks	107.619	28.975	.326	3.714	.000
2	(Constant)	-259.207	71.279		-3.637	.000
	X2 Import	.257	.103	.221	2.490	.014
	X5 Traveler Checks	103.806	29.454	.315	3.524	.001
	X1 Stock Market	-2.00E-02	.089	-.020	-.225	.823
	X3 Saving	.312	.128	.219	2.447	.016
	X4 Commission	-2.57E-02	.107	-.021	-.240	.811

- a. Dependent Variable: Y Export

3.4 השוואה בין שיטת stepwise והיררכית

לעיתים אנו מתבקשים לבחון האם הנתונים עובדו בשיטת stepwise או בשיטה ההיררכית.

קיימים כמה אינדיקציות העוזרות להבחין בין השיטות:

1. ב-stepwise נכנס משתנה ב"ת אחד בכל שלב ואילו בהיררכית לא בהכרח.
2. ב-stepwise המשתנה הראשון שנכנס צריך להיות בעל המתאם הפשוט הגבוה ביותר עם המשתנה התלוי ואילו בהיררכית לא בהכרח.
3. ב-stepwise התוספת של כל משתנה ב"ת שנכנס חייבת להיות מובהקת מעבר למה שכבר מצוי במשוואה ואילו בהיררכית לא בהכרח.
4. ב-stepwise כל מקדמי השיפוע במשוואה הסופית חייבים להיות מובהקים ואילו בהיררכית לא בהכרח.
5. ב-stepwise ה-MSRES צריך לרדת משלב לשלב ואילו ה- $AdjR^2$ צריך לעלות משלב לשלב..

? על סמך הפלטים הבאים הערך באיזו שיטה נותחו הנתונים, stepwise או היררכית? תן לפחות 3 אינדיקציות התומכות בערכתך.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.395 ^a	.156	.149	67.68720	.156	20.760	1	112	.000
2	.461 ^b	.213	.198	65.68529	.056	7.931	1	111	.006
3	.505 ^c	.255	.235	64.19061	.042	6.229	1	110	.014

a. Predictors: (Constant), X5 Traveler Checks

b. Predictors: (Constant), X5 Traveler Checks, X3 Saving

c. Predictors: (Constant), X5 Traveler Checks, X3 Saving, X2 Import

ANOVA^d

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	95112.382	1	95112.382	20.760	.000 ^a
	Residual	513134.4	112	4581.557		
	Total	608246.8	113			
2	Regression	129330.9	2	64665.464	14.988	.000 ^b
	Residual	478915.9	111	4314.557		
	Total	608246.8	113			
3	Regression	154999.0	3	51666.334	12.539	.000 ^c
	Residual	453247.8	110	4120.435		
	Total	608246.8	113			

a. Predictors: (Constant), X5 Traveler Checks

b. Predictors: (Constant), X5 Traveler Checks, X3 Saving

c. Predictors: (Constant), X5 Traveler Checks, X3 Saving, X2 Import

d. Dependent Variable: Y Export

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-221.003	72.540		-3.047	.003
	X5 Traveler Checks	130.447	28.630	.395	4.556	.000
2	(Constant)	-254.426	71.388		-3.564	.001
	X5 Traveler Checks	121.042	27.983	.367	4.326	.000
	X3 Saving	.341	.121	.239	2.816	.006
3	(Constant)	-259.691	69.795		-3.721	.000
	X5 Traveler Checks	102.038	28.387	.309	3.595	.000
	X3 Saving	.301	.119	.211	2.521	.013
	X2 Import	.251	.100	.216	2.496	.014

a. Dependent Variable: Y Export

Excluded Variables^d

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	X1 Stock Market	.089 ^a	1.025	.308	.097	.991
	X2 Import	.245 ^a	2.793	.006	.256	.920
	X3 Saving	.239 ^a	2.816	.006	.258	.986
	X4 Commission	.028 ^a	.305	.761	.029	.921
2	X1 Stock Market	.011 ^b	.118	.906	.011	.880
	X2 Import	.216 ^b	2.496	.014	.232	.904
	X4 Commission	.008 ^b	.093	.926	.009	.915
3	X1 Stock Market	-.019 ^c	-.213	.832	-.020	.864
	X4 Commission	-.020 ^c	-.229	.820	-.022	.900

- a. Predictors in the Model: (Constant), X5 Traveler Checks
- b. Predictors in the Model: (Constant), X5 Traveler Checks, X3 Saving
- c. Predictors in the Model: (Constant), X5 Traveler Checks, X3 Saving, X2 Import
- d. Dependent Variable: Y Export

Correlations^a

		Y Export	X1 Stock Market	X2 Import	X3 Saving	X4 Commission	X5 Traveler Checks
Y Export	Pearson Correlation	1	.126	.337**	.283**	.137	.395**
	Sig. (2-tailed)	.	.182	.000	.002	.147	.000
X1 Stock Market	Pearson Correlation	.126	1	.189*	.342**	.019	.094
	Sig. (2-tailed)	.182	.	.044	.000	.841	.318
X2 Import	Pearson Correlation	.337**	.189*	1	.161	.207*	.282**
	Sig. (2-tailed)	.000	.044	.	.086	.027	.002
X3 Saving	Pearson Correlation	.283**	.342**	.161	1	.109	.119
	Sig. (2-tailed)	.002	.000	.086	.	.249	.206
X4 Commission	Pearson Correlation	.137	.019	.207*	.109	1	.281**
	Sig. (2-tailed)	.147	.841	.027	.249	.	.002
X5 Traveler Checks	Pearson Correlation	.395**	.094	.282**	.119	.281**	1
	Sig. (2-tailed)	.000	.318	.002	.206	.002	.

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

a. Listwise N=114

****הערה:** שימו לב כי ניתן לשלב בין השיטות, כלומר שצעדים מסוימים יתבצעו בשיטת stepwise ואילו אחרים בהיררכית. כך שלמעשה יש לבחון כל שלב ושלב ולראות באיזה שיטה נעשה.

3.5 התוספת לניבוי ומובהקותה

כאשר מתבצעת רגרסיה בצעדים (בין אם בשיטה ההיררכית או ב-stepwise) ניתן לחשב את התוספת לניבוי של משתנה/משתנים ב"ת שנכנסו בשלב מסוים אל תוך משוואת הרגרסיה על פני משתנה/משתנים ב"ת אחרים שכבר היו במשוואה. ניתן בנוסף לבדוק האם תוספת זו לניבוי של המשתנה התלוי מובהקת באוכלוסייה.

3.5.1 חישוב התוספת לניבוי

יתבצע ע"י הנוסחה:

$$R_{all}^2 - R_{part}^2 = \text{התוספת המבוקשת}$$

↙
↘

כל המשתנים ה"ת משתנים שהיו כבר במשוואה

לדוגמא:

? על סמך נתוני דוגמא מס' 2:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.395 ^a	.156	.149	67.68720	.156	20.760	1	112	.000
2	.461 ^b	.213	.198	65.68529	.056	7.931	1	111	.006
3	.505 ^c	.255	.235	64.19061	.042	6.229	1	110	.014

a. Predictors: (Constant), X5 Traveler Checks

b. Predictors: (Constant), X5 Traveler Checks, X3 Saving

c. Predictors: (Constant), X5 Traveler Checks, X3 Saving, X2 Import

א. חשב את התוספת לניבוי פעילות הבנק בייצוא של יבוא על פני פעילות בחשבונות חיסכון והמחאות נוסעים.

ב. חשב את התוספת לניבוי פעילות הבנק ביצוא של פעילות בחשבונות חיסכון ויבוא על פני המחאות נוסעים.

3.5.2 מבחן מובהקות התוספת

מטרתו לבדוק האם התוספת לניבוי של משתנה/משתנים ב"ת מסוימים על פני אחרים מובהקת באוכלוסיה.

נדגים את שלבי המבחן על בדיקת מובהקות התוספות שחשבו בסעיפים א' ו-ב' של השאלה הקודמת.

השערות:

רשומות במונחי הבטה/בטות של התוספת, כאשר בהערה מצוינים המשתנים המצויים במשוואה (part).

למשל, עבור סעיף א' של השאלה הקודמת (תוספת של יבוא על פני חיסכון והמחאות נוסעים):

$$\begin{aligned} H_0: \beta_2 &= 0 \\ H_1: \beta_2 &\neq 0 \end{aligned}$$

כאשר המשתנים חיסכון והמחאות נוסעים מצויים במשוואה.

עבור סעיף ב' (תוספת של חיסכון ויבוא על פני המחאות נוסעים):

$$\begin{aligned} H_0: \beta_2 = \beta_3 &= 0 \\ H_1: &otherwise \end{aligned}$$

כאשר המשתנה המחאות נוסעים מצוי במשוואה.

הנחות:

אותן שלושת הנחות של מודל רגרסיה מרובה. ההנחות מתייחסות לכל המשתנים הב"ת (ל-all).

לדוגמא, ההנחות של בדיקת מובהקות התוספת של סעיף א' וב' של השאלה הקודמת:

1. דגימה מקרית של לקוחות בנק/טעויות ב"ת.

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad i \neq j$$

2. עבור כל קומבינציה ליניארית קבועה של המחאות נוסעים, יבוא ו-פעילות בחשבונות חיסכון. $\varepsilon_i \approx N(0, \sigma_\varepsilon^2)$

3. אין מתאם מלא בין המחאות נוסעים, יבוא ו-פעילות בחשבונות חיסכון.

סטטיסטי המבחן:

$$F = \frac{(R_{all}^2 - R_{part}^2) / (K_{all} - K_{part})}{(1 - R_{all}^2) / (n - K_{all} - 1)} \stackrel{H_0}{\sim} F_{k_{all} - k_{part}, n - k_{all} - 1}$$

כלל הכרעה ומסקנה:

אם $F < F_{k_{all} - k_{part}, n - k_{all} - 1}$, דוחים את H_0 ויש עדות למובהקות התוספת באוכלוסיה.

? צור משוואת רגרסיה אופטימאלית לניבוי פעילות הבנק ביצוא באמצעות המשתנים הבאים:

פעילות בחשבונות חיסכון, יבוא והמחאות נוסעים. איזה מהמשתנים יכנס בשלב הראשון? איזה

מהמשתנים יכנס בשלב השני? פטור על פי טבלת המקדמים בלבד:

Correlations^a

		Y Export	X1 Stock Market	X2 Import	X3 Saving	X4 Commission	X5 Traveler Checks
Y Export	Pearson Correlation	1	.126	.337**	.283**	.137	.395**
	Sig. (2-tailed)	.	.182	.000	.002	.147	.000
X1 Stock Market	Pearson Correlation	.126	1	.189*	.342**	.019	.094
	Sig. (2-tailed)	.182	.	.044	.000	.841	.318
X2 Import	Pearson Correlation	.337**	.189*	1	.161	.207*	.282**
	Sig. (2-tailed)	.000	.044	.	.086	.027	.002
X3 Saving	Pearson Correlation	.283**	.342**	.161	1	.109	.119
	Sig. (2-tailed)	.002	.000	.086	.	.249	.206
X4 Commission	Pearson Correlation	.137	.019	.207*	.109	1	.281**
	Sig. (2-tailed)	.147	.841	.027	.249	.	.002
X5 Traveler Checks	Pearson Correlation	.395**	.094	.282**	.119	.281**	1
	Sig. (2-tailed)	.000	.318	.002	.206	.002	.

** . Correlation is significant at the 0.01 level (2-tailed).

* . Correlation is significant at the 0.05 level (2-tailed).

a. Listwise N=114

3.5.3 הפלט המתייחס לתוספת ולמובהקותה

חלק ה- change statistic בפלט ה-model summary מתייחס לתוספת ולמובהקותה.

לדוגמא:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.395 ^a	.156	.149	67.68720	.156	20.760	1	112	.000
2	.461 ^b	.213	.198	65.68529	.056	7.931	1	111	.006
3	.505 ^c	.255	.235	64.19061	.042	6.229	1	110	.014

a. Predictors: (Constant), X5 Traveler Checks

b. Predictors: (Constant), X5 Traveler Checks, X3 Saving

c. Predictors: (Constant), X5 Traveler Checks, X3 Saving, X2 Import

? על פי ה- R square change בלבד חשב את פרופורציית השונות המוסברת של פעילות הבנק

ביצוא על ידי המשתנים הב"ת הבאים:

א. יבוא, המחאות נוסעים ופעילות בחשבונות חיסכון.

ב. פרופורציית השונות המוסברת על ידי המשתנים המחאות נוסעים ופעילות בחשבונות חיסכון בלבד.

ג. פרופורציית השונות המוסברת על ידי המשתנים המחאות נוסעים ויבוא בלבד.

3.5.4 הקשר בין מובהקות השיפוע ומובהקות התוספת לניבוי

קיים קשר בין מבחן T למובהקות השיפוע של משתנה ב"ת מסוים כאשר משתנים אחרים מצויים במשוואה (מובהקות התרומה הייחודית שלו) לבין מבחן F למובהקות התוספת של אותו משתנה ב"ת מעבר לאותם משתנים אחרים:

$$t^2\left(\frac{\alpha}{2}\right)(n-k-1) = F(\alpha)(1, n-k-1)$$

$$\text{sig} = \text{sig}$$

נדגים זאת על נתוני דוגמא מס' 2 :

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.395 ^a	.156	.149	67.68720	.156	20.760	1	112	.000
2	.461 ^b	.213	.198	65.68529	.056	7.931	1	111	.006
3	.505 ^c	.255	.235	64.19061	.042	6.229	1	110	.014

a. Predictors: (Constant), X5 Traveler Checks

b. Predictors: (Constant), X5 Traveler Checks, X3 Saving

c. Predictors: (Constant), X5 Traveler Checks, X3 Saving, X2 Import

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-221.003	72.540		-3.047	.003
	X5 Traveler Checks	130.447	28.630	.395	4.556	.000
2	(Constant)	-254.426	71.388		-3.564	.001
	X5 Traveler Checks	121.042	27.983	.367	4.326	.000
	X3 Saving	.341	.121	.239	2.816	.006
3	(Constant)	-259.691	69.795		-3.721	.000
	X5 Traveler Checks	102.038	28.387	.309	3.595	.000
	X3 Saving	.301	.119	.211	2.521	.013
	X2 Import	.251	.100	.216	2.496	.014

a. Dependent Variable: Y Export

ניתן לראות למשל את הקשר בין מובהקות השיפוע של יבוא כאשר חיסכון והמחאות נוסעים במשוואה (פלט המקדמים) לבין מובהקות התוספת של יבוא מעבר לאותם שני משתנים (פלט ה-model summary):

$$t^2 = 2.496^2 = F = 6.229$$

$$sig = 0.014$$

?

1. על סמך פלט ה-model summary בלבד-מה יהיה סטטיסטי T של מבחן מובהקות התרומה הייחודית של חשבונות חיסכון כאשר המחאות נוסעים מצוי במשוואה?
2. על סמך פלט המקדמים בלבד- מה יהיה ערך ה-sig של מובהקות התוספת של חשבונות חיסכון כאשר המחאות נוסעים מצוי במשוואה?

4. בדיקת קו ליניאריות

קו-ליניאריות משמעה מתאם בין המשתנים הב"ת.

אנו שואפים שהקו-ליניאריות במודל (מתאמים בין המשתנים הב"ת) תהיה נמוכה ככל שניתן כי כידוע מתאם גבוה בין הב"ת פוגם בתרומה הייחודית של המשתנים לניבוי המשתנה התלוי.

לעיתים נתבקש להעריך האם קיימת בעיה של קו- ליניאריות במודל הכולל שני משתנים ב"ת. במקרה זה נצטרך למעשה להעריך האם קיים מתאם גבוה בין שני המשתנים הב"ת שמצויים במודל.

אנו יודעים כי כאשר אין מתאם כלל בין ה"ב"ת- $r_{12} = 0$ מתקיים:

$$(1) \quad \frac{\hat{\beta}_1}{z\hat{\beta}_1} = \text{ברגרסיה פשוטה} \quad \text{במרובה (כאשר } x_2 \text{ מצוי גם בתוך המשוואה).}$$

$$(2) \quad \begin{aligned} z\hat{\beta}_1 &= r_{y1} \\ z\hat{\beta}_2 &= r_{y2} \end{aligned} \quad \text{ברגרסיה מרובה (כאשר שני המשתנים ה"ב"ת בתוך המשוואה).}$$

$$(3) \quad \hat{\beta}_1 > \hat{\beta}_1 \quad \text{במרובה (כאשר } x_2 \text{ מצוי גם במשוואה) בפשוטה}$$

$$(4) \quad R_{y12}^2 = r_{y1}^2 + r_{y2}^2$$

$$(5) \quad \text{Tolerance} = 1 \text{ (בפלט ה- excluded variables).}$$

הסבר: מדד ה-Tolerance נותן את פרופורציית השונות הלא מוסברת של משתנה ב"ת מסוים ע"י משתנים ב"ת אחרים המצויים במשוואה. ככל שפרופורציה זו גבוהה יותר כך קיים מתאם נמוך יותר בין אותו משתנה ב"ת למשתנים ב"ת האחרים המצויים במשוואה.

Tolerance=1 מעיד על חוסר מתאם בין ה"ב"ת ואילו Tolerance=0 מעיד על מתאם מושלם בין המשתנה ה"ב"ת למשתנים ה"ב"ת האחרים המצויים במשוואה.

נדגים באמצעות נתוני דוגמא מס' 2 :

Excluded Variables^d

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics	
					Tolerance	
1	X1 Stock Market	.089 ^a	1.025	.308	.097	.991
	X2 Import	.245 ^a	2.793	.006	.256	.920
	X3 Saving	.239 ^a	2.816	.006	.258	.986
	X4 Commission	.028 ^a	.305	.761	.029	.921
2	X1 Stock Market	.011 ^b	.118	.906	.011	.880
	X2 Import	.216 ^b	2.496	.014	.232	.904
	X4 Commission	.008 ^b	.093	.926	.009	.915
3	X1 Stock Market	-.019 ^c	-.213	.832	-.020	.864
	X4 Commission	-.020 ^c	-.229	.820	-.022	.900

a. Predictors in the Model: (Constant), X5 Traveler Checks

b. Predictors in the Model: (Constant), X5 Traveler Checks, X3 Saving

c. Predictors in the Model: (Constant), X5 Traveler Checks, X3 Saving, X2 Import

d. Dependent Variable: Y Export

? על סמך נתוני דוגמא 2 הערך את גודל המתאם בין המשתנים הב"ת: 'המחאות נוסעים' (x_5) ו-'פעילות בחשבונות חיסכון' (x_3).

5. שאלות מסכמות

סקר מטעם הלשכה המרכזית לסטטיסטיקה בחן את הגורמים המשפיעים על רווחי חברות במשק. לשם כך נדגמו 117 חברות במשק ונתקבלו לגביהן הנתונים על המשתנים הבאים:

רווחי החברה

גיל המנהל

רמת השכלתו

הוותק בתפקיד הנוכחי

מס' העובדים בחברה

נתקבלו הממצאים הבאים:

Descriptive statistic

	Mean	Std.deviation	N
Profit	5608.08		117
Time Job seniority	80.02		117
AGE of manager	13.97		117
EDLEVEL	6.888		117
Workers No	3.17		117

Correlations^a

			PROFIT Profit for the firm in \$	TIME Job seniority (months)	AGE Age of manager	EDLEVEL Educational level (years)	EXP Work experience (years)	WORKERS No. of workers
PROFIT Profit for the firm in \$	Pearson Correlation Sig. (2-tailed)	1 .	.227* .014	-.115 .219	.784** .000	-.093 .318	.575** .000	
TIME Job seniority (months)	Pearson Correlation Sig. (2-tailed)	.227* .014	1 .	.041 .661	.162 .080	.001 .993	.183* .048	
AGE Age of manager	Pearson Correlation Sig. (2-tailed)	-.115 .219	.041 .661	1 .	-.270** .003	.798** .000	-.230* .013	
EDLEVEL Educational level (years)	Pearson Correlation Sig. (2-tailed)	.784** .000	.162 .080	-.270** .003	1 .	-.171 .065	.824** .000	
EXP Work experience (years)	Pearson Correlation Sig. (2-tailed)	-.093 .318	.001 .993	.798** .000	-.171 .065	1 .	-.146 .116	
WORKERS No. of workers	Pearson Correlation Sig. (2-tailed)	.575** .000	.183* .048	-.230* .013	.824** .000	-.146 .116	1 .	

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

a. Listwise N=117

Model summary

Model	R	R square	Adjusted R square	Std. Error of the Estimate
1	A			
2	B			653.959
3	C			645.613

a. Predictors: (Constant), EDLEVEL.

b. Predictors: (Constant), EDLEVEL, Workers No.

c. Predictors: (Constant), EDLEVEL, Workers No., Time Job seniority.

Excluded Variables^d

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	TIME Job seniority (months)	.102 ^a	1.752	.082	.162	.974
	AGE Age of manager	.104 ^a	1.751	.083	.162	.927
	EXP Work experience (years)	.042 ^a	.716	.475	.067	.971
	WORKERS No. of workers	-.220 ^a	-2.188	.031	-.201	.321
2	TIME Job seniority (months)	.114 ^b	1.992	.049	.184	.966
	AGE Age of manager	.103 ^b	1.750	.083	.162	.927
	EXP Work experience (years)	.041 ^b	.707	.481	.066	.971
3	AGE Age of manager	.093 ^c	1.593	.114	.149	.919
	EXP Work experience (years)	.038 ^c	.656	.513	.062	.970

- a. Predictors in the Model: (Constant), EDLEVEL Educational level (years)
 b. Predictors in the Model: (Constant), EDLEVEL Educational level (years), WORKERS No. of workers
 c. Predictors in the Model: (Constant), EDLEVEL Educational level (years), WORKERS No. of workers, TIME Job seniority (months)
 d. Dependent Variable: PROFIT Profit for the firm in \$

- א. האם המקדם החלקי של רמת השכלה מובהק כאשר מס' עובדים מצוי במשוואה? בדוק ברמת מובהקות של 0.05 (רשום הנחות, הצב השערות, חשב והסק מסקנה מילולית).
- ב. בנה טבלת מקור שונות לניבוי הרווחים על ידי רמת ההשכלה ומס' העובדים בלבד.
- ג. מהו סכום הסטיות הריבועיות של הרווחים על פי טבלת מקור השונות שבנית בסעיף הקודם? כיצד ישתנה אם נוסיף את המשתנה הב"ת של וותק המנהל? הסבר.
- ד. חשב את גודל התוספת במונחי R^2 של מס' העובדים והוותק על פני רמת ההשכלה. האם תוספת זו מובהקת? בדוק ברמת מובהקות של 0.05 (רשום הנחות, הצב השערות, חשב והסק מסקנה מילולית).
- ה. על פי איזה שיטה נותחו הנתונים (stepwise או היררכית)? נמק תשובתך ותן לה סימוכין מתוך הנתונים.

בהמשך הוחלט לנתח את אותם הנתונים בדרך נוספת והתקבלו הממצאים הבאים:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.790 ^a	.624	.618	658.748	.624	94.654	2	114	.000
2	.797 ^b	.636	.623	654.240	.012	1.788	2	112	.172
3	.808 ^c	.653	.637	641.338	.017	5.552	1	111	.020

a. Predictors: (Constant), EDLEVEL Educational level (years), AGE Age of manager

b. Predictors: (Constant), EDLEVEL Educational level (years), AGE Age of manager, TIME Job seniority (months), EXP Work experience (years)

c. Predictors: (Constant), EDLEVEL Educational level (years), AGE Age of manager, TIME Job seniority (months), EXP Work experience (years), WORKERS No. of workers

$$\hat{y}_i = 400.612 - 369.364Branch1 - 245.308Branch2 + 0.806x_i$$

ו. האם המשוואה לניבוי הרווחים על ידי רמת ההשכלה וגיל המנהל מובהקת באוכלוסייה? בדוק ברמת מובהקות של 0.05 (רשום הנחות, הצב השערות והסק מסקנה מילולית בלבד, ללא חישוב).

ז. מה יהיה סטטיסטי T למובהקות מקדם השיפוע של המשתנה 'מספר עובדים' כאשר יתר המשתנים הב"ת האחרים מצויים בתוך המשוואה?

ח. איזה משתנים ב"ת יופיעו בפלט ה-excluded variables בכל שלב?