

# סטטיסטיקה 3

מנהל עסקים – המכללה למנהל

## סטודנטים יקרים

לפניכם ספר תרגילים בקורס סטטיסטיקה 3 לחוג מנהל עסקים במכללה למנהל.

הספר הוא חלק מפרויקט חדשני וראשון מסוגו בארץ במקצוע זה, המועבר ברשת האינטרנט On-line. הקורס באתר כולל פתרונות מלאים לספר התרגילים. **הפתרונות מוגשים במוגשים בסרטוני וידאו המלווים בהסבר קולי, כך שאתם רואים את התהליכים בצורה מובנית, שיטתית ופשוטה, ממש כפי שנעשה בשיעור פרטי.**

הקורס נבנה ע"י קרן ברוסרד, המנוסה בלימוד בקורס יותר מ 10 שנים עם הצלחות מוכחות של מאות סטודנטים.

אז אם אתם עסוקים מידי בעבודה, סובלים מלקויות למידה, רוצים להצטיין או פשוט אוהבים ללמוד בשקט בבית, אנחנו מזמינים אתכם לחוויית לימודים יוצאת דופן וחדשה לחלוטין, היכנסו עכשיו לאתר



אנו מאחלים לכם הצלחה מלאה בבחינות

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

## תוכן עניינים

5.....	הקדמה	
6.....	<a href="#">חלק א'-</a> ניתוח שונות חד גורמי	
6.....	1. מתי משתמשים בניתוח שונות חד גורמי?	
6.....	1.1 זיהוי המודל המחקרי	
6.....	2. השערות	
7.....	3. אמידת גודל האפקט	
9.....	4. ביצוע המבחן הסטטיסטי	
9.....	4.1 חישוב סטטיסטי המבחן F	
9.....	4.2 כלל הכרעה	
9.....	4.3 מסקנה	
11.....	5. הנחות המודל	
11.....	5.1 רישום ההנחות במונחי השאלה	
11.....	5.2 הנחת שוויון שונות	
13.....	6. ניתוח פלטים	
13.....	6.1 ביצוע המבחן הסטטיסטי על סמך פלט ה-Descriptives	
13.....	6.2 בדיקת ההנחות על סמך הפלטים	
15.....	7. פרופורציית השונות המוסברת- $eta^2$	
17.....	8. השוואות מרובות (POST-HOC TESTS)	
17.....	8.1 השערות	
17.....	8.2 השיטות להשוואות מרובות: TUKY ו-L.S.D	
18.....	8.3 ניתוח פלטים	
21.....	<a href="#">חלק ב'</a> גרסיה ליניארית חד משתנית	
22.....	1. מבוא- מקדם המתאם הליניארי	
22.....	1.1 הגדרה	
22.....	1.2 דיאגרמת פיזור	

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

23.....	תכונות מקדם המתאם הליניארי.....	1.3
25.....	קיצוץ תחום והשפעתו על המתאם.....	1.4
26.....	השפעת ערכים קיצוניים/חריגים על המתאם.....	1.5
28.....	השפעת טרנספורמציה ליניארית על המתאם.....	1.6
28.....	נוסחת המתאם הליניארי.....	1.7
29.....	מובהקות המתאם.....	1.8
31.....	קו הרגרסיה במדגם.....	2.
31.....	בניית קו הרגרסיה במדגם.....	2.1
34.....	תכונות קו הרגרסיה במדגם.....	2.2
34.....	אומדנים חשובים ברגרסיה פשוטה.....	2.3
36.....	מתאמים בין משתני הרגרסיה.....	2.4
37.....	משוואת הרגרסיה המתוקנת.....	2.5
38.....	2.6 השוואה בין מודל רגרסיה חד משתנית לבין מודל ניתוח שונות חד כיוונית.....	2.6
39.....	מובהקות הרגרסיה באוכלוסיה.....	3.
39.....	ההנחות הנדרשות למודל הרגרסיה הפשוטה.....	3.1
43.....	בדיקת השערות למובהקות משוואת הרגרסיה באוכלוסיה.....	3.2
46.....	בדיקת השערות למובהקות מקדם השיפוע באוכלוסיה.....	3.3
48.....	קשר בין התפלגויות F-I T.....	3.4
49.....	רב"ס לתוחלת ערכי $Y$ ( $\mu_0$ ) עבור ערך מסוים של $X$ .....	3.5
51.....	רב"ס לערכי $Y$ עבור ערך מסוים של $X$ .....	3.6
53.....	חלק ג'-רגרסיה מרובה ומשתנה דמי.....	
52.....	רגרסיה מרובה.....	
52.....	1. מאפייני קו הרגרסיה במדגם.....	1.
52.....	<b>1.1 מקדמי הרגרסיה</b> .....	1.1
55.....	<b>1.2 אומדנים חשובים ברגרסיה מרובה</b> .....	1.2
59.....	2. מובהקות קו הרגרסיה ומקדמיו באוכלוסיה.....	2.
59.....	<b>2.1 הנחות מודל הרגרסיה המרובה</b> .....	2.1

59.....	מובהקות משוואת הרגרסיה המרובה	2.2
60.....	מובהקות השיפועים	2.3
63.....	שיטות להרצת רגרסיה רבת משתנים	3
63.....	3.1 שיטת ENTER	
65.....	3.2 רגרסיה בצעדים חכמים (stepwise)	
67.....	3.3 רגרסיה היררכית	
68.....	3.4 השוואה בין שיטת stepwise והיררכית	
71.....	3.5 התוספת לניבוי ומובהקותה	
76.....	4. בדיקת קו ליניאריות	
77.....	5. שאלות מסכמות	
81.....	משתנה דמי	
81.....	1. יצירת משתני הדמי	
81.....	2. רגרסיה עם משתנה ב"ת איכותי אחד בלבד	
81.....	2.1 בניית משוואת הרגרסיה	
82.....	2.2 השוואה בין מודל הרגרסיה עם משתני דמי לניתוח שונות חד גורמי	
86.....	2.3 השוואה בין מודל רגרסיה עם משתני דמי לרגרסיה פשוטה	
88.....	3. רגרסיה המשלבת משתנים ב"ת איכותיים וכמותיים	
89.....	3.1 בניית קו הרגרסיה	
89.....	3.2 ניתוח הרגרסיה	
	מבחנים לדוגמה 1+2	4

# הקדמה

מבנה הקורס:

- 1) ניתוח שונות חד גורמי
- 2) רגרסיה חד משתנית ("פשוטה")
- 3) רגרסיה מרובה ומשתני דמי
- 4) פיתרון מבחנים לדוגמא

# ניתוח שונות חד גורמי

## 1. מתי משתמשים בניתוח שונות חד גורמי?

בניתוח שונות חד גורמי משתמשים כאשר רוצים לבדוק האם קיים הבדל בין מספר קבוצות במשתנה מסוים.

דוגמא מס' 1: חוקר רצה לבחון השפעתן של שלוש שיטות טיפול פיזיותרפי שונות על רמת ההתפתחות המוטורית של ילדים קטנים (ציוני ההתפתחות המוטורית נעים בין 1 "נמוכה מאד" ל- 10 "גבוהה מאד").

### 1.1 זיהוי המודל המחקרי

זיהוי המודל המחקרי יתבצע על ידי הגדרת המשתנים: התלוי והבלתי-תלוי

שאלה מס' 1 (מבוססת על דוגמא מס' 1)

הגדר את המשתנים וזהה את המודל המחקרי

תשובה:

## 2. השערות

המשך דוגמא מס' 1: לצורך הבדיקה נבדקו 14 ילדים בעלי רמת התפתחות זהה והם חולקו באקראי ל- 3 קבוצות. לאחר 8 חודשי טיפול עבר כל אחד מהילדים הערכה של רמת ההתפתחות המוטורית שלו.

להלן התוצאות

שהתקבלו:

	שיטה ג'	שיטה ב'	שיטה א'	
	2	5	8	
	4	4	5	
	3	7	4	
	2	7	7	
	4		6	
$n = 14$	$n_3 = 5$	$n_2 = 4$	$n_1 = 5$	$n_j$
$\bar{x} = 4.86$	$\bar{x}_3 = 3$	$\bar{x}_2 = 5.75$	$\bar{x}_1 = 6$	$\bar{x}_j$
$\hat{s}_T = 1.92$	$\hat{s}_3 = 1$	$\hat{s}_2 = 1.5$	$\hat{s}_1 = 1.58$	$\hat{s}_j$

**השערות:**

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_K$$

$$H_1: otherwise$$

אפקט הטיפול =  $\Delta$  = הבדל בין התוחלות באוכלוסיה

**3. אמידת גודל האפקט**

כיצד נסיק על פי ההבדל בין הממוצעים במדגם ( $\bar{x}_j$ ) שקיים הבדל בין התוחלות באוכל' ( $\mu_j$ )?

התשובה: באמצעות ניתוח השונות (סטטיסטי המבחן F):

$$F = \frac{MSB}{MSW}$$

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

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

בכדי להבין את הרציונאל עליו נשען סטטיסטי המבחן F נעזר בדוגמא ההיפותטית הבאה:

נתונות שתי אוכלוסיות שבשתייהן  $\Delta = 0$

אוכל' מס' 2:  $\sigma_j^2 \neq 0$

אוכל' מס' 1:  $\sigma_j^2 = 0$

שיטה א'	שיטה ב'	שיטה ג'		שיטה א'	שיטה ב'	שיטה ג'
1	2	2		6	6	6
3	5	4		6	6	6
6	6	5		6	6	6
6	6	7		6	6	6
10	7	8		6	6	6
8	10	10		6	6	6
$\mu_1 = 6$	$\mu_2 = 6$	$\mu_3 = 6$		$\mu_1 = 6$	$\mu_2 = 6$	$\mu_3 = 6$

$$\bar{x}_j \neq \mu_j$$

$$\bar{x}_j = \mu_j$$

## שאלה מס' 2

חוה/י דעתך על הטענות הבאות:

א. MSB הוא אומדן חסר הטיה ל:

1. שונות הטעויות.

תשובה:

2. לגודל האפקט

תשובה:

ב. במבחן F הבדוק הבדלים בין תוחלות, ה-MSB וה-MSW הינם משתנים מקריים בעלי תוחלת  $E(MSB)$  ו-  $E(MSW)$  בהתאמה. במידה ולא קיים הבדל בין התוחלות של המשתנה התלוי, ציין מה מבין שלהלן מתקיים ונמק תשובתך:

a.  $E(MSW) = E(MSB)$

b. MSB בהכרח שווה או קטן מ-MSW

c. MSB בהכרח שווה ל-MSW

d.  $E(MSB) < E(MSW)$

e. אף אחת מהתשובות הנ"ל

תשובה:



#### 4. ביצוע המבחן הסטטיסטי

##### 4.1 חישוב סטטיסטי המבחן F

את סטטיסטי המבחן  $F = \frac{MSB}{MSW}$  נחשב באמצעות טבלת מקור שונות (ANOVA):

מקור השונות	df (degrees of freedom) דרגות חופש	SS (sum of squares) סכומי ריבועים	MS (Mean of squares) שנויות	F
B	C-1	$SSB = \sum_{j=1}^c nj(\bar{x}_j - \bar{x})^2$	MSB=SSB/C-1	$F = \frac{MSB}{MSW}$
w	n-C	$SSW = \sum_{j=1}^c \hat{S}_j^2(n_j - 1)$	MSW=SSW/n-C	
T	n-1	$SST = (N - 1)\hat{S}_T^2$ $= SSB + SSW$	MST=SST/n-1 $= \hat{S}_T^2$ <u>הערה חשובה:</u> MST $\neq$ MSB+MSW	

##### 4.2 כלל הכרעה

דוחים את H0 אם:  $F > F_c \alpha(c-1, n-c)$

##### 4.3 מסקנה

יש/אין עדות להבדל במשתנה התלוי בין קבוצות המשתנה הב'ת באוכלוסייה, ברמת מובהקות של אלפא

**שאלה מס' 3 (מבוססת על דוגמא מס' 1)**

האם שיטת הטיפול משפיעה על רמת ההתפתחות המוטורית של הילדים? בדוק ברמת מובהקות של 0.05.

**תשובה:**

## 5. הנחות המודל

### 5.1 רישום ההנחות במונחי השאלה

דגימה מקרית של האובייקט שמדגם.

מדגמים בלתי תלויים.

התפלגות נורמאלית של המשתנה התלוי בתוך כל אוכלוסיית המשתנה ה"ב"ת.

שוונות המשתנה התלוי שווה בכל אוכלוסיות המשתנה ה"ב"ת.

מודל קבוע.

### 5.2 הנחת שוויון שוונות

הנחת שוויון שוונות גורסת כי כל השוונות של האוכלוסיות שוות אחת לשנייה ולשוונות הטעויות

$$(\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \dots = \sigma_k^2 = \sigma_e^2)$$

תחת הנחת שוויון שוונות :

1) כל אחת מן האומדים לשוונות מתוך נתוני המדגם  $(\hat{s}_j^2)$  וה-MSW (המהווה ממוצע משוקלל של

$$MSW = \frac{\sum \hat{s}_j^2 \cdot (n_j - 1)}{n - c}$$
 האומדים לשוונות של כל קבוצה

האוכלוסייה  $(\sigma_j^2)$  וגם את השוונות הטעויות  $(\sigma_e^2)$ .

2) בכדי לאמוד כל אחת מן השוונות באוכלוסייה  $(\sigma_j^2)$  כמו גם את שוונות הטעויות  $(\sigma_e^2)$ , נעדיף

להשתמש באומד לשוונות MSW במקום באומדים לשוונות  $\hat{s}_j^2$ .

הסיבה להעדפה זו היא שחישוב ה-MSW מבוסס על כל תצפיות המדגם ואילו חישוב ה- $\hat{s}_j^2$

מבוסס רק על מס' התצפיות של הקבוצה הספציפית. ככל שהאומד לשוונות מבוסס על יותר

תצפיות כך הוא יהיה יעיל ומדויק יותר כאומדן לפרמטר.

3) חישוב ה-MSW (האומד לשוונות הטעויות  $(\sigma_e^2)$  נשען על הנחת שוויון שוונות.

#### שאלה מס' 4

חוה/י דעתך על הטענות הבאות:

א. תחת הנחות המודל כל אחת משוניות המדגמים היא אומדן חסר הטיה ל-  $\sigma_e^2$ .

תשובה:

ב. הנחת שוויון שוניות משמשת להערכת גודל האפקט.

תשובה:

ג. תחת הנחות המודל, בכדי לאמוד את השונות של אחת האוכלוסיות נעדיף להשתמש ב-MSW במקום באומד לשונות הספציפי של אותה אוכלוסייה ( $\hat{\sigma}_j^2$ ).

תשובה:

## 6. ניתוח פלטים

### 6.1 ביצוע המבחן הסטטיסטי על סמך פלט ה-Descriptives

דוגמא מס' 2:

במחקר מסוים רצו לבדוק האם קיימים הבדלים בהישגים בבחינה בסטטיסטיקה 3 של הסטודנטים במכללה כפונקציה של שיטות למידה שונות. לשם כך נדגמו באופן מקרי מהמכללה מס' תלמידים מכל אחת משיטות הלימוד ולהלן הנתונים שהתקבלו:

GRADES = ציונים בבחינה

TRAINING METHOD = שיטת הלימוד

### שאלה מס' 5

בדוק את שאלת המחקר ברמת מובהקות של 0.05, לפי השלבים הבאים:

- הצב השערות
- רשום את ההנחות הדרושות למודל במונחי השאלה המחקרית
- הצג את הטבלה המקובלת לניתוח הנתונים וסכם בה את הממצאים
- נסח מסקנה מילולית

#### Descriptives

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
					1 alone	20.00		
2 class	20.00	73.50	10.51	2.35	68.58	78.42	48.00	90.00
3 small groups	20.00	78.45	6.35	1.42	75.48	81.42	65.00	90.00
Total	60.00	71.88	12.06	1.56	68.77	75.00	33.00	90.00

תשובה:

### 6.2 בדיקת ההנחות על סמך הפלטים

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

שתי ההנחות הניתנות לבדיקה באמצעות הפלטים הן הנחת שוויון שוניות והנחת הנורמאליות.

נדגים את בדיקת ההנחות ברמת מובהקות של 0.05 על נתוני דוגמא מס' 2.

### 6.2.1 בדיקת הנחת שוויון שוניות

הפלט המתאים לבדיקת הנחת שוויון שוניות הוא הפלט הבא:

#### Test of Homogeneity of Variances

GRADES Score on training exam

Levene Statistic	df1	df2	Sig.
2.604	2	57	.083

#### השערות:

$$H0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

$H1: otherwise$

כלל הכרעה (על סמך ה- sig שבפלט):

$sig = 0.083 > \alpha = 0.05$  לכן אין סיבה מספקת לדחות את  $H0$  ברמת מובהקות של 0.05

#### מסקנה:

ניתן להניח שוויון שוניות של הציונים בבחינה בכל אוכלוסיית שיטות הלימוד ברמת מובהקות של 0.05.

### 6.2.2 בדיקת הנחת הנורמאליות

הפלט המתאים לבדיקת הנחת הנורמאליות הוא הפלט הבא:

#### Tests of Normality

GROUP Sales training method	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
GRADES Score 1 alone	.147	20	.200*	.956	20	.470
on training exam 2 class	.131	20	.200*	.954	20	.433
3 small groups	.156	20	.200*	.944	20	.281

\*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

**\*\*יש להתייחס רק למבחן Kolmogorov-Smirnov**

#### השערות:

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

עבור כל אחת משלושת אוכלוסיות שיטות הלימוד יש להניח:

H0: התפלגות נורמאלית

H1: התפלגות אחרת

כלל הכרעה (על סמך ה-sig שבפלט):

עבור כל אחת משלושת אוכלוסיות שיטות הלימוד:

$$sig = 0.2 > \alpha = 0.05 \text{ ולכן אין סיבה מספקת לדחות את } H_0 \text{ ברמת מובהקות של } 0.05$$

מסקנה:

ניתן להניח התפלגות נורמאלית של הציונים בבחינה עבור כל אחת מאוכלוסיות שיטות הלימוד ברמת מובהקות של 0.05.

\*\*הערה: מספיק שנדחה את H0 עבור אחת האוכלוסיות בכדי שהנחת הנורמאליות תופר.

שאלה נוספת (מבוססת על דוגמא מס' 2)

סטודנט שעיין בפלטים הציע 2 אומדנים חסרי הטיה לשונות ציוני הבחינה בקבוצת הלומדים לבד. הערך האחד הוא:  $13.5^2$  והערך השני הוא: 111.02. רשום מתי היית מעדיף את אומדן א', מתי את אומדן ב' ומדוע?

תשובה:

7. פרופורציית השונות המוסברת -  $eta^2$

מדד זה מהווה פרופורציה של השונות המוסברת מתוך כלל השונות:

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

$$0 \leq \eta^2 = \frac{SSB}{SST} \leq 1$$

פרופורציית השונות הלא מוסברת:

$$0 \leq 1 - \eta^2 = \frac{SSW}{SST} \leq 1$$

שאלה מס' 6 (מתייחסת לדוגמא מס' 2)

א. כמה אחוזים מן השונות של ציוני הבחינה מוסברים ע"י שיטות הלימוד?

תשובה:

ב. כמה אחוזים מן השונות של ציוני הבחינה אינם מוסברים ע"י שיטות הלימוד?

תשובה:

ג. מהי עוצמת הקשר בין הציונים בבחינה לבין שיטות הלימוד?

תשובה:



## 8. השוואות מרובות (POST-HOC TESTS)

תוצאה מובהקת בניתוח השונות משמעה כאמור- לפחות אחד מן הממוצעים באוכלוסייה שונה מהאחרים.

כדי לדעת איזה מן הממוצעים שונה מהאחרים יש לבצע מבחני המשך (post-hoc tests).  
אנו נשווה כל זוג ממוצעים זה לזה.

$$\frac{c \cdot (c-1)}{2} \text{ מספר ההשוואות הזוגיות האפשריות עבור } C \text{ קבוצות:}$$

בדוגמא מס' 2:

$$\frac{3 \cdot 2}{2} = 3 \text{ מס' ההשוואות המרובות:}$$

### 8.1 השערות

\*\*מתייחסות לדוגמא מס' 2

$$\begin{array}{lll} H_0: \mu_2 - \mu_3 = 0 & H_0: \mu_1 - \mu_3 = 0 & H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_2 - \mu_3 \neq 0 & H_1: \mu_1 - \mu_3 \neq 0 & H_1: \mu_1 - \mu_2 \neq 0 \end{array}$$

### 8.2 השיטות להשוואות מרובות: TUKY ו-L.S.D

נלמדו שתי שיטות להשוואות מרובות:

TUKY (1)

L.S.D (2)

#### 8.2.1 ההבדל בין השיטות

TUKY מחמירה יותר בדחיית  $H_0$  מ-L.S.D.

למה?

בשיטת L.S.D קיימת בעיית תיסוף ה- $\alpha$ : ככל שנעשה יותר השוואות כך הטעות מסוג ראשון ( $\alpha$ ) של כל מערכת ההשוואות תתנפח.

$$\alpha_{PE} = 1 - (1 - \alpha_{PC})^N$$

שיטת TUKY מתגברת על בעיית תיסוף ה- $\alpha$  על ידי כך שקובעת את ה- $\alpha$  הכללית או הסופית של כל מערכת ההשוואות מראש. כתוצאה מכך ה- $\alpha$  עבור כל השוואה קטנה וקשה יותר לדחות את  $H_0$  לעומת שיטת L.S.D.

משמעות ההבדל בין השיטות:

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

אם דחינו H0 ב-TUKY  $\Leftarrow$  נדחה את H0 גם ב-L.S.D.

אם קיבלנו את H0 ב-L.S.D  $\Leftarrow$  נקבל את H0 גם ב-TUKY.

אם קיבלנו את H0 ב-TUKY  $\Leftarrow$  לא ניתן לדעת האם נקבל או נדחה את H0 ב-L.S.D.

אם דחינו את H0 ב-L.S.D  $\Leftarrow$  לא ניתן לדעת האם נקבל או נדחה את H0 ב-TUKY.

### 8.3 ניתוח פלטים

ישנם שני פלטים העוסקים בהשוואות מרובות:

(1) פלט ה-Multiple Comparisons

(2) פלט ה-Homogeneous Subsets

המשך דוגמא מס' 2...

## 1) Post Hoc Tests

### Multiple Comparisons

Dependent Variable: GRADES Score on training exam

	(I) GROUP Sales training method	(J) GROUP Sales training method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	1 alone	2 class	-9.80*	3.33	.01	-17.82	-1.78
		3 small groups	-14.75*	3.33	.00	-22.77	-6.73
	2 class	1 alone	9.80*	3.33	.01	1.78	17.82
		3 small groups	-4.95	3.33	.31	-12.97	3.07
	3 small groups	1 alone	14.75*	3.33	.00	6.73	22.77
		2 class	4.95	3.33	.31	-3.07	12.97
LSD	1 alone	2 class	-9.80*	3.33	.00	-16.47	-3.13
		3 small groups	-14.75*	3.33	.00	-21.42	-8.08
	2 class	1 alone	9.80*	3.33	.00	3.13	16.47
		3 small groups	-4.95	3.33	.14	-11.62	1.72
	3 small groups	1 alone	14.75*	3.33	.00	8.08	21.42
		2 class	4.95	3.33	.14	-1.72	11.62

\*. The mean difference is significant at the .05 level.

### שאלה מס' 7

א. בהסתמך על הפלט Multiple Comparisons - מהן שיטות הלימוד אשר קיימים ביניהן הבדלים מובהקים בהישגים בבחינה?

תשובה:

ב. מדוע קיים הבדל ב- sig בהשוואה בין הקבוצה של הלומדים בקבוצה קטנה לבין קבוצת הלומדים בכיתה בין השיטות TUKY ו-L.S.D?

תשובה:

ג. באיזה שיטה-TUKY או L.S.D הרב"ס לאותו ההפרש בין התוחלות יהיה צר יותר? נמק.

תשובה:

## 2) Homogeneous Subsets

GRADES Score on training exam

GROUP Sales training method	N	Subset for alpha = .05	
		1.00	2.00
Tukey HSD <sup>a</sup> 1 alone	20.00	63.70	
2 class	20.00		73.50
3 small groups	20.00		78.45
Sig.		1.00	.31

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 20.000.

ד. בהסתמך על הפלט Homogeneous Subsets - מהן שיטות הלימוד אשר קיימים ביניהן הבדלים מובהקים על פי שיטת TUKY?

תשובה:

ה. מה צפויות להיות תוצאות הבדיקה לפי שיטת L.S.D? לגבי כל השוואה ציין מה תהיה מסקנה. במידה ואינך יכול לתת תשובה חד משמעית נמק מדוע.

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

**תשובה:**

1. מכון מחקר בדק את השאלה האם המקצוע הנלמד משפיע על המשכורת ההתחלתית (באלפי שקלים). להלן הנתונים:

**Descriptives**

SALARY starting salary

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00 Agriculture	55	6.5911	1.75608	.23679	6.1164	7.0659	1.90	10.15
2.00 Business	61	7.5319	1.91268	.24489	7.0420	8.0217	3.48	12.59
3.00 Engineering	53	9.7779	2.33067	.32014	9.1355	10.4203	4.60	14.34
4.00 Communication	81	8.1221	2.55649	.28405	7.5568	8.6873	.74	15.02
Total	250	7.9923	2.44013	.15433	7.6883	8.2962	.74	15.02

בוצעו השוואות מרובות והתקבלו התוצאות הבאות:

**Multiple Comparisons**

Dependent Variable: SALARY starting salary

	(I) COLLEGE	(J) COLLEGE	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	1.00 Agriculture	2.00 Business	-.9408	.40920	.101	-1.9992	.1177
		3.00 Engineering	-3.1868*	.42359	.000	-4.2825	-2.0911
		4.00 Communication	-1.5309*	.38450	.001	-2.5255	-.5364
	2.00 Business	1.00 Agriculture	.9408	.40920	.101	-.1177	1.9992
		3.00 Engineering	-2.2460*	.41324	.000	-3.3150	-1.1771
		4.00 Communication	-.5902	.37307	.391	-1.5552	.3748
	3.00 Engineering	1.00 Agriculture	3.1868*	.42359	.000	2.0911	4.2825
		2.00 Business	2.2460*	.41324	.000	1.1771	3.3150
		4.00 Communication	1.6559*	.38880	.000	.6502	2.6616
	4.00 Communication	1.00 Agriculture	1.5309*	.38450	.001	.5364	2.5255
		2.00 Business	-.5902	.37307	.391	-.3748	1.5552
		3.00 Engineering	-1.6559*	.38880	.000	-2.6616	-.6502
LSD	1.00 Agriculture	2.00 Business	-.9408*	.40920	.022	-1.7467	-.1348
		3.00 Engineering	-3.1868*	.42359	.000	-4.0211	-2.3525
		4.00 Communication	-1.5309*	.38450	.000	-2.2883	-.7736
	2.00 Business	1.00 Agriculture	.9408*	.40920	.022	.1348	1.7467
		3.00 Engineering	-2.2460*	.41324	.000	-3.0600	-1.4321
		4.00 Communication	-.5902	.37307	.115	-1.3250	.1446
	3.00 Engineering	1.00 Agriculture	3.1868*	.42359	.000	2.3525	4.0211
		2.00 Business	2.2460*	.41324	.000	1.4321	3.0600
		4.00 Communication	1.6559*	.38880	.000	.8901	2.4217
	4.00 Communication	1.00 Agriculture	1.5309*	.38450	.000	.7736	2.2883
		2.00 Business	-.5902	.37307	.115	-.1446	1.3250
		3.00 Engineering	-1.6559*	.38880	.000	-2.4217	-.8901

\*. The mean difference is significant at the .05 level.

שחזר את פלט ה- Homogeneous Subsets לפי שיטת TUKY ברמת מובהקות של 0.05.

תשובה:

ז. בהיסתמך על פלט ה- Homogeneous Subsets שבנית, אילו היית עורך השוואות מרובות עבור רמת מובהקות של 0.1, בין אילו קבוצות היית מוצא הבדלים מובהקים?

תשובה:

## רגרסיה ליניארית חד משתנית

מטרת הרגרסיה הליניארית היא ניבוי משתנה מסוים (משתנה תלוי, המכונה "משתנה מנבא") באמצעות משתנים אחרים (משתנים ב"ת, המכונים "משתנים מנבאים"). כאשר יש לי משתנה מנבא/ב"ת אחד בלבד מדובר ברגרסיה חד משתנית ("פשוטה").

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

ניתוח מודל הרגרסיה מתחלק ל-3 שלבים:

- 1) זיהוי צורת הקשר בין המשתנים כקשר ליניארי.
- 2) בניית קו הרגרסיה במדגם.
- 3) בדיקת מובהקות קו הרגרסיה ומקדמיו באוכלוסיה.

1. מבוא - מקדם המתאם הליניארי

1.1 הגדרה

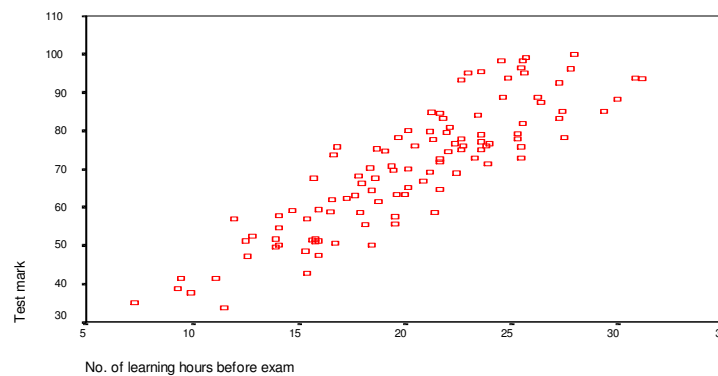
קשר בין שני משתנים כמותיים אשר ניתן להגדירו על פי נוסחת הקו הישר ( $Y=a+bX$ ).

1.2 דיאגרמת פיזור

דוגמא מס' 1: קשר בין שעות למידה בבחינה והציונים בבחינה:

ציונים	שעות למידה	
76	19	1
60	17	2
85	30	3
...	...	...
70	26	109

בכדי לאמוד את טיב הקשר בין המשתנים יש להציגו בדיאגרמת פיזור:



מדיאגרמת הפיזור הנ"ל נוכל להתרשם כי הקשר הוא ליניארי, חיובי וחזק.

$$-1 \leq r_p \leq 1$$

מקדם המתאם של פירסון מוגדר על ידי שני פרמטרים בלתי תלויים אחד בשני-כיוון ועוצמה:

(1) כיוון הקשר

א. חיובי/עולה/ישר- שני המשתנים נעים באותו הכיוון. אם האחד עולה או יורד אז גם השני. מקדם המתאם (r) יקבל סימן +.

ב. שלילי/יורד/הפוך-שני המשתנים נעים בכיוונים הפוכים. אם האחד עולה השני יורד ולהפך. מקדם המתאם יקבל סימן (-).

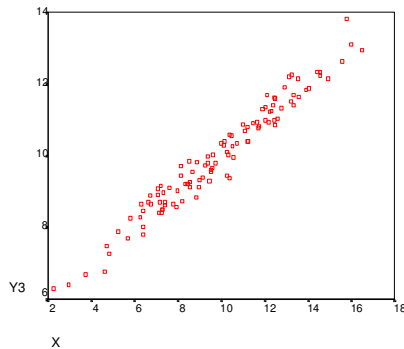
(2) עוצמת הקשר

ככל שעוצמת הקשר גבוהה יותר כך יהיה ערך מקדם המתאם קרוב יותר ל-1 בערך מוחלט. ככל שהעוצמה נמוכה יותר כל יהיה קרוב יותר ל-0 בערך מוחלט.

א. דוגמאות לקשר ליניארי חלקי:

ככל שהנקודות קרובות יותר לקו הישר כך עוצמת הקשר חזקה יותר ולהיפך:

Graph

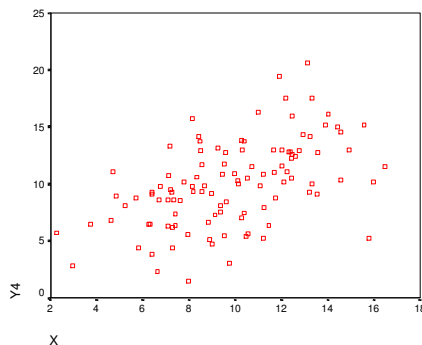


Correlations<sup>a</sup>

		X	Y3
X	Pearson Correlation	1	.979**
	Sig. (2-tailed)	.	.000
Y3	Pearson Correlation	.979**	1
	Sig. (2-tailed)	.000	.

\*\* .Correlation is significant at the 0.01 level

a. Listwise N=109

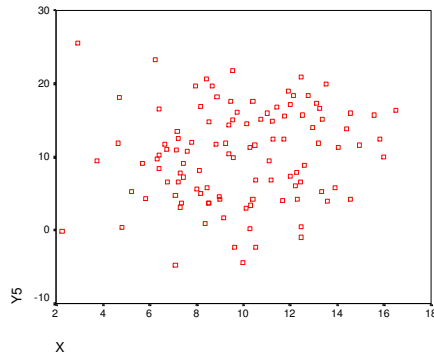


Correlations<sup>a</sup>

		X	Y4
X	Pearson Correlation	1	.518**
	Sig. (2-tailed)	.	.000
Y4	Pearson Correlation	.518**	1
	Sig. (2-tailed)	.000	.

\*\* .Correlation is significant at the 0.01 level

a. Listwise N=109



**Correlations<sup>a</sup>**

		X	Y5
X	Pearson Correlation	1	.122
	Sig. (2-tailed)	.	.207
Y5	Pearson Correlation	.122	1
	Sig. (2-tailed)	.207	.

a. Listwise N=109

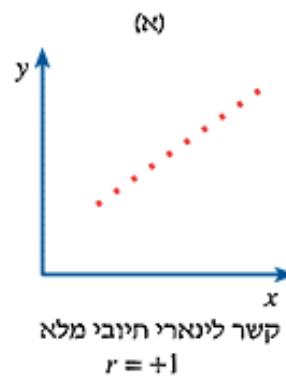
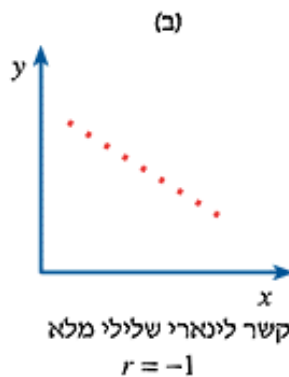
**הערכת עוצמת הקשר:**

$0 < |r| < 0.3$  מתאם נמוך

$0.3 < |r| < 0.7$  מתאם בינוני

$0.7 < |r| < 1$  מתאם חזק

**ב. דוגמאות לקשר ליניארי מושלם/מלא:**

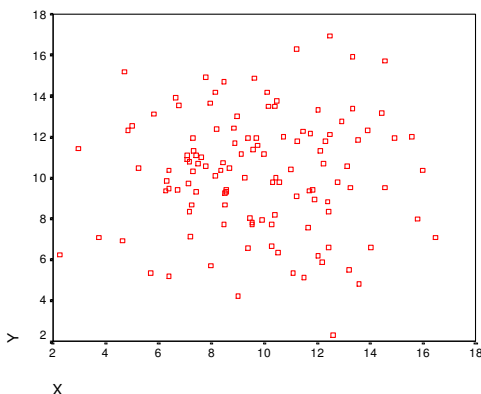


**תכונה מס' 1:** קשר ליניארי מושלם מהווה קשר בין משתנה א' לבין משתנה ב' המהווה טרנספורמציה ליניארית של משתנה א'.

**\*\*תזכורת:** טרנספורמציה ליניארית של משתנה-הוספה/החסרה של קבוע ו/או הכפלה/חילוק בקבוע של כל ערכי המשתנה

לדוגמא: כאשר  $Y = X + 2$ ,  $r_{xy} = 1$  (כי Y הוא טרנספורמציה ליניארית של X).

**ג. דוגמאות לחוסר קשר (ליניארי):**

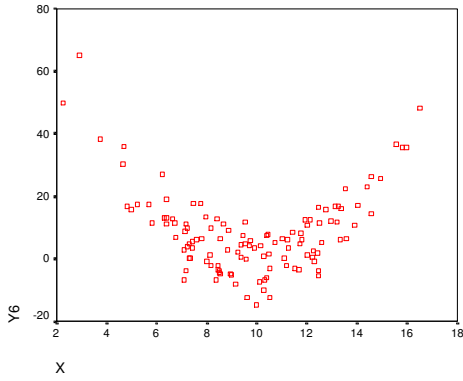


**Correlations<sup>a</sup>**

		X	Y
X	Pearson Correlation	1	.000
	Sig. (2-tailed)	.	.998
Y	Pearson Correlation	.000	1
	Sig. (2-tailed)	.998	.

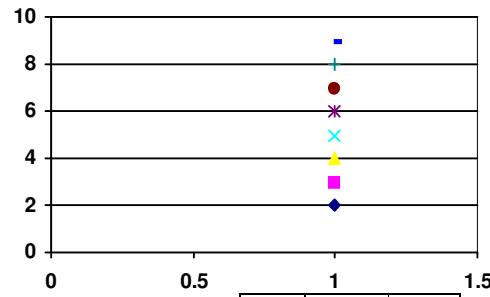
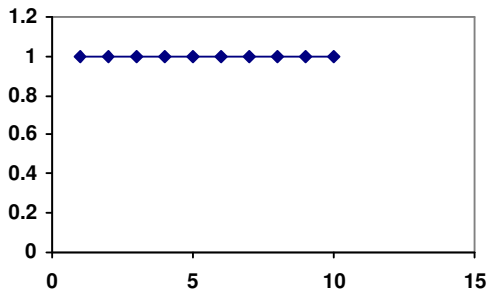
a. Listwise N=120





**Correlations**

		X	Y6
X	Pearson Correlation	1	-.043
	Sig. (2-tailed)	.	.639
	N	120	120
Y6	Pearson Correlation	-.043	1
	Sig. (2-tailed)	.639	.
	N	120	120



Y	X	
1	4	1
1	7	2
1	1	3
...	...	...
1	8	10

$$\hat{S}_Y^2 = 0$$

Y	X	
6	1	1
2	1	2
9	1	3
...	...	...
7	1	8

$$\hat{S}_X^2 = 0$$

**תכונה מס' 2:** קשר בין משתנה לקבוע יהיה שווה תמיד 0.

1.4 קיצוץ תחום והשפעתו על המתאם

קיצוץ תחום משמעו התייחסות לתחום צר יותר של אחד המשתנים (בדר"כ המשתנה ה"ב"ת) באופן המקטין את השונות של המשתנים. הקטנת השונות של המשתנים גוררת הקטנת המתאם ביניהם. המתאם בין המשתנים אחרי הקיצוץ יהיה נמוך יותר מן המתאם לפני הקיצוץ.

בדוגמא 1- מתואר קשר בין שעות לימוד לפני המבחן לציון במבחן.

הקשר נבדק כעת על קבוצה בעלת תחום צר יותר בשעות לימוד (מעל - 25).

אחרי הקיצוץ:

לפני הקיצוץ:

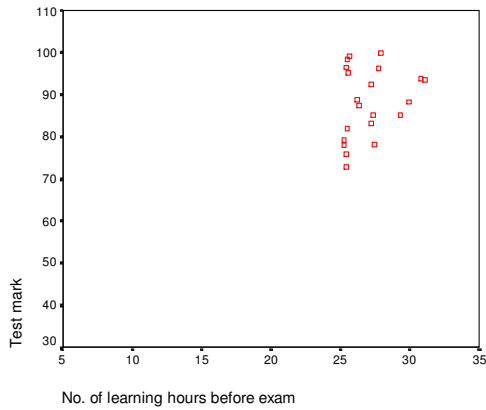
Descriptive Statistics

	Mean	Std. Deviation	N
No. of learning hours before exam	27.0251	1.87348	21
Test mark	88.0571	8.28331	21

Descriptive Statistics

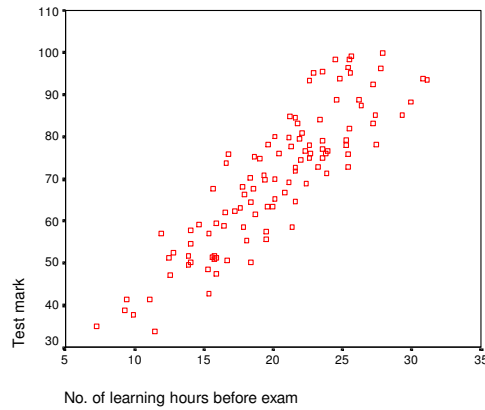
	Mean	Std. Deviation	N
HOURS No. of learn hours before exam	20.1754	5.03946	109
MARK Test mark	70.1530	16.22455	109

דיאגרמת הפיזור לאחר הקיצוץ:



$r=0.264$

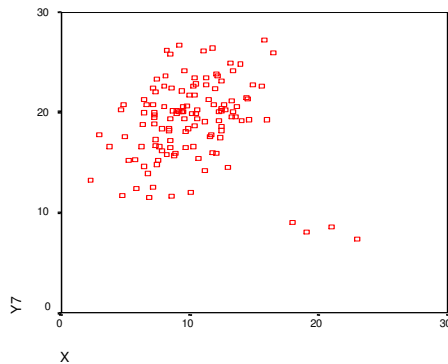
דיאגרמת הפיזור המקורית:



$r=0.88$

1.5 השפעת ערכים קיצוניים/חריגים על המתאם

• מקרה ראשון שבו הקיצוניים מקטינים קשר:



Correlations<sup>a</sup>

		X	Y7
X	Pearson Correlation	1	.023
	Sig. (2-tailed)	.	.799
Y7	Pearson Correlation	.023	1
	Sig. (2-tailed)	.799	.

a. Listwise N=124

**אותם נתונים לאחר השמטת 4 תצפיות שבניגוד למגמה:**



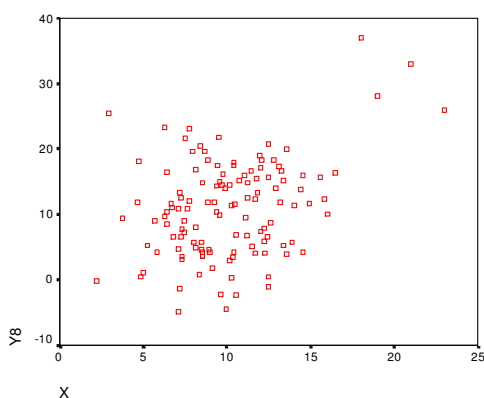
**Correlations<sup>a</sup>**

		X	Y7
X	Pearson Correlation	1	.413**
	Sig. (2-tailed)	.	.000
Y7	Pearson Correlation	.413**	1
	Sig. (2-tailed)	.000	.

\*\* . Correlation is significant at the 0.01 level

a. Listwise N=120

**מקרה אחר שבו הקיצוניים מחזקים קשר:**



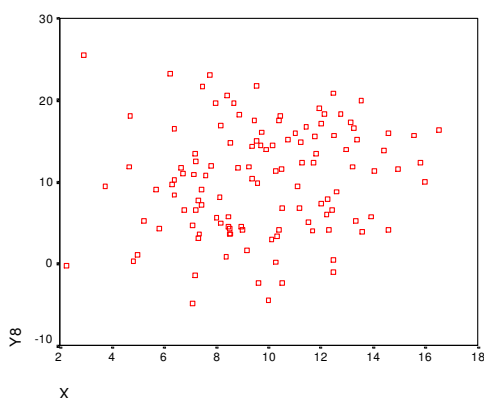
**Correlations<sup>a</sup>**

		X	Y8
X	Pearson Correlation	1	.350**
	Sig. (2-tailed)	.	.000
Y8	Pearson Correlation	.350**	1
	Sig. (2-tailed)	.000	.

\*\* . Correlation is significant at the 0.01 level

a. Listwise N=124

**אותם נתונים לאחר השמטת 4 תצפיות שגורמות למגמה:**



**Correlations<sup>a</sup>**

		X	Y8
X	Pearson Correlation	1	.128
	Sig. (2-tailed)	.	.165
Y8	Pearson Correlation	.128	1
	Sig. (2-tailed)	.165	.

a. Listwise N=120

**לסיכום: מקדם המתאם מושפע מאוד מערכים חריגים.**

כאשר הערכים החריגים הם בניגוד למגמה הליניארית של מרבית הנתונים-הם יפגעו בעוצמת המתאם

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

כאשר הערכים החריגים מצויים בהתאם למגמה הליניארית של מרבית הנתונים- הם "ינפחו" את המתאם.

1.6 השפעת טרנספורמציה ליניארית על המתאם

טרנספורמציה ליניארית איננה משנה את ערכו של המתאם.

הדגמה על דוגמא מס' 1 (הקשר בין שעות לימוד לבחינה לציון בבחינה):

הוחלט לתת פקטור לציון בבחינה-להוסיף לכל ציון 10 נקודות. המתאם בין הציון לאחר הטרנספורמציה לבין שעות הלימוד יהיה שווה למתאם לפני הטרנספורמציה:

$$r' = r = 0.88$$

1.7 נוסחת המתאם הליניארי

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

הנוסחה מבוססת על ההשתנות המשותפת של  $x$  ושל  $y$ :

$$\text{cov}(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

ההשתנות המשותפת של משתנה עם עצמו מהווה את האומדן לשונות של המשתנה:

$$\text{cov}(x, x) = \frac{\sum (x_i - \bar{x})^2}{n - 1} = \hat{s}_x^2$$
$$\text{cov}(y, y) = \frac{\sum (y_i - \bar{y})^2}{n - 1} = \hat{s}_y^2$$

כך שניתן לרשום את הנוסחה של המתאם גם כך:

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

פלט ה-correlation מתאר את המתאם ואת אומדני ה-cov השונים:

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

Correlations<sup>a</sup>

		HOURS No. of learning hours before exam	MARK Test mark
HOURS No. of learning hours before exam	Pearson Correlation	1	.880**
	Sig. (2-tailed)	.	.000
	Sum of Squares and Cross-products	2742.785	7766.854
	Covariance	25.396	71.915
MARK Test mark	Pearson Correlation	.880**	1
	Sig. (2-tailed)	.000	.
	Sum of Squares and Cross-products	7766.854	28429.479
	Covariance	71.915	263.236

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

a. Listwise N=109

### 1.8 מובהקות המתאם

את מובהקות המתאם נבדוק על סמך הפלט בלבד (ללא חישוב). הפלט הרלוונטי למובהקות המתאם הוא פלט ה-correlations .

נדגים את בדיקת מובהקות המתאם (ברמת מובהקות של 0.05) על נתוני דוגמא מס' 1 (הקשר בין שעות לימוד לבחינה לציון בבחינה):

Correlations<sup>a</sup>

		HOURS No. of learning hours before exam	MARK Test mark
HOURS No. of learning hours before exam	Pearson Correlation	1	.880**
	Sig. (2-tailed)	.	.000
	Sum of Squares and Cross-products	2742.785	7766.854
	Covariance	25.396	71.915
MARK Test mark	Pearson Correlation	.880**	1
	Sig. (2-tailed)	.000	.
	Sum of Squares and Cross-products	7766.854	28429.479
	Covariance	71.915	263.236

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

a. Listwise N=109

### השערות:

$$H_0: \eta = 0$$

$$H_1: \eta \neq 0$$

כלל הכרעה (על סמך ה-sig שבפלט ה-correlations):

$$\text{Sig} = 0.00 < \alpha = 0.05$$

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)



יש סיבה מספקת לדחות את  $H_0$  ברמת מובהקות של 0.05

**מסקנה:**

יש עדות למתאם מובהק באוכלוסייה בין שעות לימוד לבחינה לבין הציונים בבחינה ברמת מובהקות של 0.05.

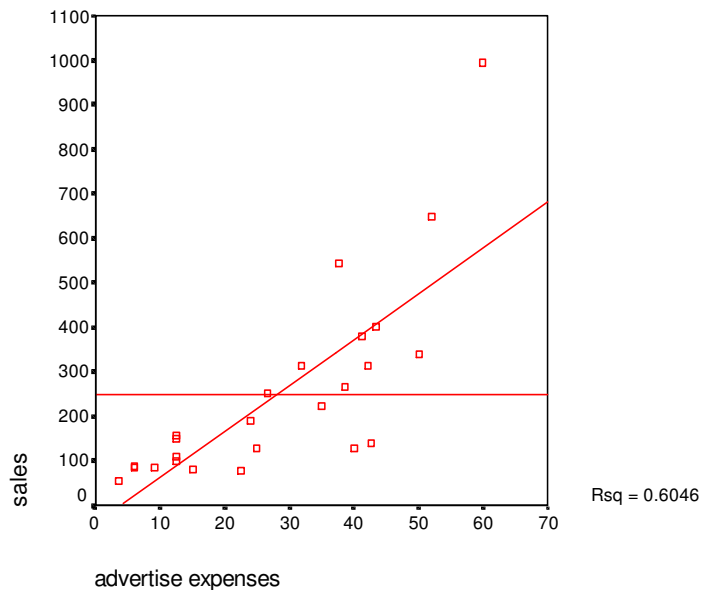
## 2. קו הרגרסיה במדגם

### 2.1 בניית קו הרגרסיה במדגם

#### 2.1.1 עיקרון הריבועים הפחותים (L.S.E)

לאחר שנוכחנו שיש מתאם ליניארי בין שני המשתנים אנו מעוניינים לבנות את קו הניבוי שיאפשר לנו לבנא 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):

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

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

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-  $\hat{\beta}$ :

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

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

2.1.2 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:

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)



**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<sup>a</sup>**

		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 ₪?

2.2 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 ( $\bar{\hat{y}}$ ) = ממוצע ציונים אמיתיים/מנובאים ( $\bar{y}$ ):

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

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

$$ssy = ssreg + ssres$$

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

2.3 אומדנים חשובים ברגרסיה פשוטה

2.3.1 פרופורציית השונות המוסברת

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

פרופורציית השונות הלא-מוסברת:

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

2.3.2 שונות הטעויות

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

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

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 ניבויי המכירות
- ח. סכום ריבועי הסטיות של הטעויות בניבוי המכירות
- ט. סכום הסטיות הריבועיות של הוצאות הפרסום

2.4 מתאמים בין משתני הרגרסיה

	y ערכי המשתנה התלוי/מנובא	$\hat{y}$ ניבויים/אומדנים/ PRED	E טעויות בניבוי/ RES
X ערכי המשתנה הב"ת/מנובא	$r_{xy}$	$r_{x\hat{y}} = 1$ 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				

## 2.5 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<sup>a</sup>

	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

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

### 2.5.1 מקדמי המשוואה המתוקנת

א. מקדם השיפוע המתוקן:

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

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

מקדם השיפוע במשוואה המתוקנת שווה למתאם בין X ל-Y

ב. מקדם החיתוך המתוקן:

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

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

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

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 רגרסיה 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  $\Leftarrow eta^2 = r^2$ ,  
מודל הרגרסיה מתאים יותר לניתוח הנתונים מניתוח השונות כי הוא ממוקד ויעיל יותר.

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

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

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

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 ינותחו 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-  $\eta^2$  (שווים או שאחד גדול יותר מהשני)? נמק.

### 3. מובהקות הרגרסיה באוכלוסיה

3.1 ההנחות הנדרשות למודל הרגרסיה הפשוטה

3.1.1 ההנחות ומשמעותן

למודל הרגרסיה החד משתנית נדרשות שתי ההנחות הבאות:

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

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)  $\varepsilon_i \approx N(0, \sigma_\varepsilon^2)$  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 באוכלוסיה הוא קו התחלות-הציונים המנובאים שווים לממוצע ערכי  $Y$  עבור כל ערך קבוע של  $x_i$  ( $\mu_{y_i|x=x_i} = \alpha + \beta x_i$ ). 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 הטיות למקדמי הקו באוכלוסיה-  $\alpha$  ו- $\beta$ .

ב. שונות הטעויות ( $\sigma_\varepsilon^2$ ) 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  $x_i$  ( $MSRES \rightarrow \sigma_{y_1}^2 = \sigma_{y_2}^2 = \dots = \sigma_\varepsilon^2$ ).

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

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.

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

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

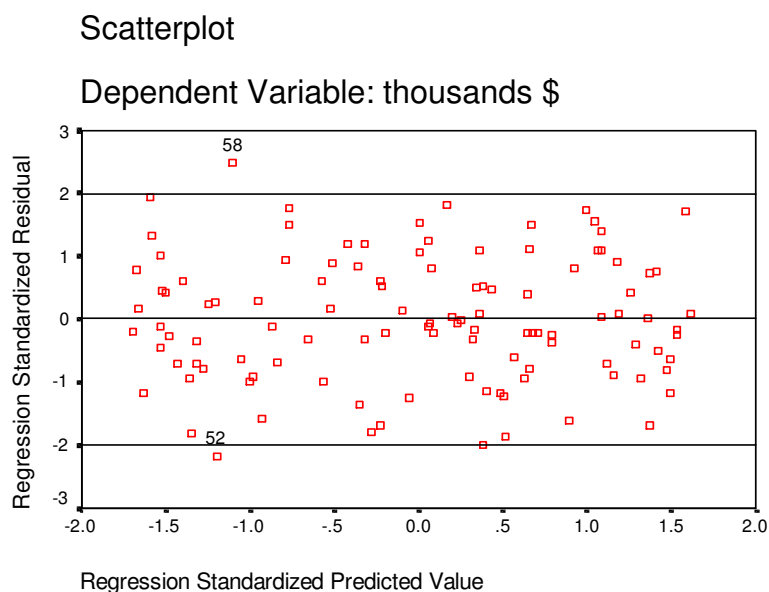
לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)



שונות קבועה ומינימאלית  $\hat{\alpha}, \hat{\beta}$  יעילים לפרמטרים  $\alpha, \beta$  באוכל' תוחלת התפלגות הטעויות = 0 האומדנים  $\hat{\alpha}, \hat{\beta}$  חסרי הטיה לפרמטרים  $\alpha, \beta$  באוכל'

### 3.1.2 בדיקת ההנחות

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



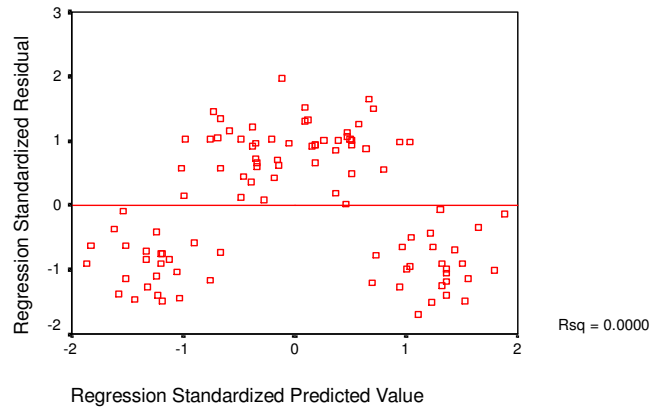
על דיאגרמה מעין זו ניתן לבצע הערכה (שאיננה מבחן סטטיסטי) האם מתקיימות כל אחת משלושת ההנחות. דיאגרמה זו מאפשרת בנוסף איתור תצפיות חריגות (הסוטות יותר משתי סטיות תקן מעל או מתחת לממוצע).

### דוגמאות לאי עמידה בהנחות המודל:

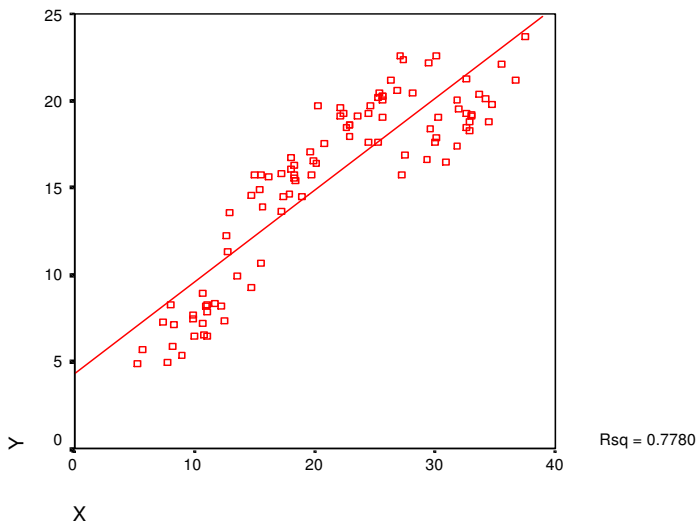
א. אי עמידה בהנחת תוחלת הטעויות = 0  $(E(\varepsilon_i) = 0)$  עבור כל ערך של X :

### Scatterplot

Dependent Variable: Y



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



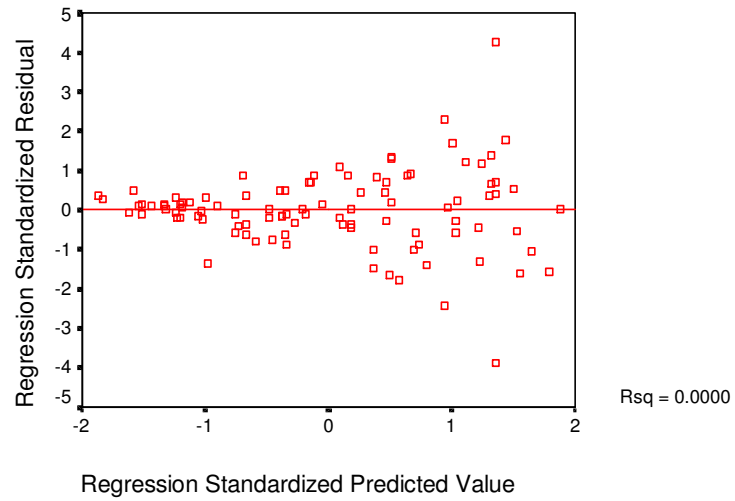
במקרה זה האומדנים  $\hat{\alpha}$  ו- $\hat{\beta}$  יהיו אומדנים מוטים לפרמטרים  $\alpha$  ו- $\beta$  באוכלוסיה.

ב. אי עמידה בהנחת שוויון שוניות:

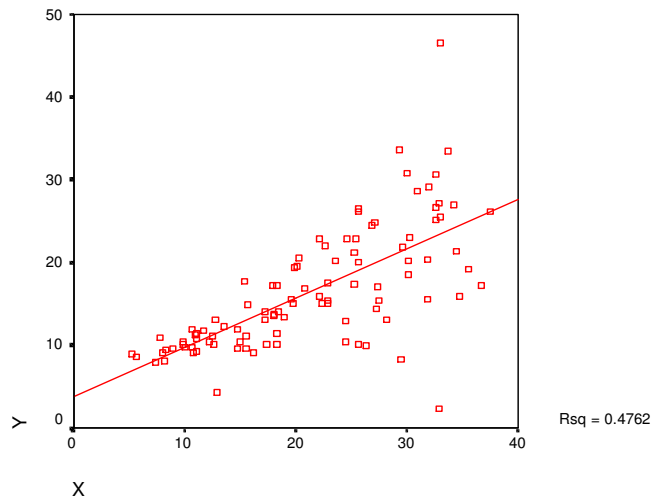
לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

## 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  $\alpha$  ו- $\beta$  באוכלוסיה.

3.2 בדיקת השערות למובהקות משוואת הרגרסיה באוכלוסיה

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

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

### 3.2.1 השערות

$$H_0; \beta = 0$$

$$H_1; \beta \neq 0$$

ההשערות מנוסחות במונחי הפרמטר  $\beta$ , 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	$F_{\alpha; (1, n-2)}$ 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  $F <$  קריטי

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

**מסקנה: יש/אין** עדות לכך שמשוואת הרגרסיה לניבוי  $Y$  על ידי  $X$  מובהקת באוכלוסיה ברמת מובהקות של  $\alpha$ .

**? 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<sup>a</sup>**

		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<sup>b</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.823 <sup>a</sup>	.677	.674	25.39155

a. Predictors: (Constant), SIZE square meter

b. Dependent Variable: PRICE thousands \$

כאשר נתון בנוסף ש:  $n=112$

- עם חישוב טבלת מקור שונות.
- 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

3.3 בדיקת השערות למובהקות מקדם השיפוע באוכלוסיה  
מבחן סטטיסטי שמטרתו לבדוק האם מקדם השיפוע מובהק באוכלוסיה.

#### 3.3.1 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.

#### 3.3.2 סטטיסטי המבחן

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

#### 3.3.3 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  $\alpha$  שלמה.

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

#### 3.3.4 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-האומד לשונות הטעויות-MSRES. קיים קשר ישיר בינו לבין  $\hat{s}_{\hat{\beta}}$ .

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

קיים קשר הפוך בין ה-MSRES לבין עוצמת הקשר הליניארי בין המשתנים  
 $(|r_{xy}|)$ .

ב. במכנה-סכום הסטיות הריבועיות של X-SSX. קיים קשר הפוך בינו לבין  $\hat{s}_{\beta}$ .

קשור לתופעת קיצוץ תחום; כאשר מקוצץ תחום ערכי X (הב"ת), ה-SSX קטן.

3.3.5 רב"ס ל- $\beta$

כאשר אנו מתבקשים לאמוד את שיפוע המשוואה, נבנה רווח בר סמך ל- $\beta$ :

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

??

(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<sup>a</sup>

		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 באוכלוסייה 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.

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

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

### 3.3.6 פלט המקדמים (coefficients)

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

		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:**

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

3.4 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 ( $\beta$ )

(3) מבחן למובהקות מקדם המתאם ( $\eta$ )

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):

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)



ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	148427.7	1	148427.680	230.216	.000 <sup>a</sup>
	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<sup>a</sup>

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<sup>a</sup>

		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

3.5 רב"ס לתוחלת ערכי Y ( $\mu_0$ ) עבור ערך מסוים של X

רב"ס שמטרתו לאמוד את תוחלת ערכי המשתנה התלוי ( $\mu_0$ ) עבור ערך מסוים של המשתנה ה"ב" ( $x_0$ ). 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:

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

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

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

- א. MSRES -האומדן לשונות הטעויות. ככל שגדל, טעות התקן/ הרב"ס גדלים ולהפך.
- ב. נ- גודל המדגם. ככל שגדל, טעות התקן/ הרב"ס קטנים ולהפך.
- ג. SSX-מונה השונות של X (קשור לתופעת קיצוץ תחום). ככל שגדל, טעות התקן/הרב"ס קטנים ולהפך.
- ד.  $(x_0 - \bar{x})$ -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) והפליטים 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

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?

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

- 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?

3.6 רב"ס לערכי Y עבור ערך מסוים של X

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

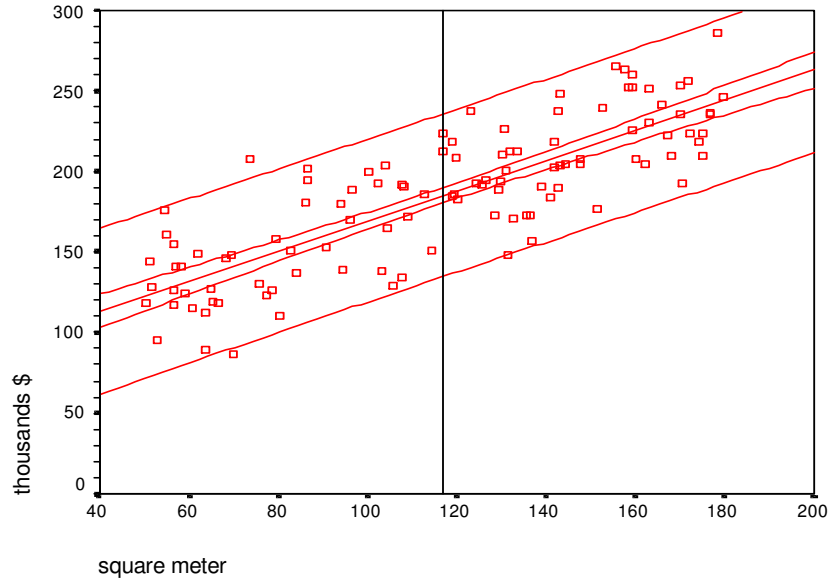
נוסחת הרב"ס:

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

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?**

# חלק ג' - רגרסיה מרובה ומשתנה דמי

## 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 הב"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"	$\hat{\beta}_1 = \frac{(r_{y1} - r_{y2} * r_{12})}{1 - r_{12}^2} * \frac{\hat{s}_y}{\hat{s}_{x1}}$ $\hat{\beta}_2 = \frac{(r_{y2} - r_{y1} * r_{12})}{1 - r_{12}^2} * \frac{\hat{s}_y}{\hat{s}_{x2}}$ <b>הערה:</b> ניתן לחישוב רק ברגרסיה דו משתנית	$z\hat{\beta}_1 = \frac{(r_{y1} - r_{y2} * r_{12})}{1 - r_{12}^2}$ $z\hat{\beta}_2 = \frac{(r_{y2} - r_{y1} * r_{12})}{1 - r_{12}^2}$

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

<p>התרומה היחסית של המשתנה הב"ת לניבוי המשתנה התלוי.</p> <p>השוואת הבטות המתוקנות בערך מוחלט- <math> z\hat{\beta} </math> מאפשרת להעריך את התרומה היחסית של כל אחד מהמשתנים הב"ת לניבוי המשתנה התלוי.</p> <p>הסיבה שאנו משווים בין השיפועים המתוקנים ולא הגולמיים היא כי המתוקנים 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><math display="block">z\hat{\beta}_1 = r_{y1}</math></p> <p><math display="block">z\hat{\beta}_2 = r_{y2}</math></p> <p>הבטה המתוקנת שווה למתאם הפשוט.</p>	<p><math display="block">\hat{\beta}_1 = r_{y1} \cdot \frac{\hat{s}_y}{\hat{s}_{x1}}</math></p> <p><math display="block">\hat{\beta}_2 = r_{y2} \cdot \frac{\hat{s}_y}{\hat{s}_{x2}}</math></p> <p>שיפועי הרגרסיה המרובה שווים לשיפועי הפשוטה</p>	<p>כאשר המתאם בין המשתנים הב"ת=0 <math>r_{12} = 0</math></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<sup>a</sup>

			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:

$$r_{12} = 0$$

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

$$\hat{s}_{x1}^2 = 9, \hat{s}_{x2}^2 = 25$$

$$\hat{y}_i = 1.8zx_{1i} - 1.2zx_{2i} \text{ : המשוואה המתוקנת שהתקבלה}$$

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, תהייה 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.  $AdjR_{y12}^2 = 0.611$ .

ב.  $R_{y12}^2 = 0.611$ .

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**

	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

**חשב את האומדנים הבאים:**

- א. האומדן לפרופורציית השונות הלא מוסברת של השכר הנוכחי ע"י רמת ההשכלה והוותק
- ב. א.ח.ה לשונות של השכר 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 באוכלוסייה

### 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  $X$ ,  $\varepsilon_i \approx N(0, \sigma_\varepsilon^2)$ .

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 שבטה 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	$F_{\alpha}(k, n-k-1)$
אמיתית y	n-1	SSy		קריטי

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  $\alpha$

? **בדוק האם המשוואה לניבוי השכר הנוכחי של העובד ע"י ההשכלה והוותק שלו מובהקת באוכל' 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<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.857 <sup>a</sup>	.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 ה"ב"ת מובהק באוכלוסייה.

#### השערות:

$H_0; \beta_j = 0$   
כאשר המשתנים ה"ב"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  $\alpha$ .

**?** 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<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.857 <sup>a</sup>	.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

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

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<sup>a</sup>**

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- $\hat{\beta}_j$

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

**?** 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  $H_0$  במבחן למובהקות השיפוע של משתנה ב"ת מסוים ברגרסיה פשוטה, האם בהכרח נצליח לדחות את  $H_0$  למובהקות אותו השיפוע גם במרובה (כאשר ישנם משתנים ב"ת נוספים בתוך המשוואה)?

**תשובה:**

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

מרובה

פשוטה

$$r_{12} = 0 \quad (1) \quad S.E.(\hat{\beta}) = \sqrt{\frac{MSRES}{SSX}}$$

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

ולכן אם דחינו H0 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 H0 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

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

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

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

ניתוח הנתונים לפי שיטת Enter:

Variables Entered/Removed<sup>b</sup>

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

a. All requested variables entered.

b. Dependent Variable: Y Export

Model Summary<sup>b</sup>

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.506 <sup>a</sup>	.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<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	155428.3	5	31085.659	7.414	.000 <sup>a</sup>
	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<sup>a</sup>

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<sup>a</sup>

		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<sup>d</sup>

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

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 השערתו צריך החוקר לבנות משוואת רגרסיה בשני צעדים/שלב. בשלב הראשון להכניס את המשתנים -המחאות נוסעים ויבוא ואילו בשלב השני להכניס את שלושת המשתנים הנותרים:

Variables Entered/Removed<sup>b</sup>

Model	Variables Entered	Variables Removed	Method
1	X5 Traveler Checks, X2 Import <sup>a</sup>	.	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<sup>a</sup>

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.

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

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 <sup>a</sup>	.156	.149	67.68720	.156	20.760	1	112	.000
2	.461 <sup>b</sup>	.213	.198	65.68529	.056	7.931	1	111	.006
3	.505 <sup>c</sup>	.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<sup>d</sup>

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

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<sup>d</sup>**

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

		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 באוכלוסיה.

#### 3.5.1 חישוב התוספת לניבוי

יתבצע ע"י הנוסחה:

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

כל המשתנים הב"ת      משתנים שהיו כבר במשוואה

לדוגמא:

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

**? 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 <sup>a</sup>	.156	.149	67.68720	.156	20.760	1	112	.000
2	.461 <sup>b</sup>	.213	.198	65.68529	.056	7.931	1	111	.006
3	.505 <sup>c</sup>	.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 בייצוא 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 שחושבו 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):

$$H_0: \beta_2 = 0$$

כאשר המשתנים חיסכון והמחאות נוסעים מצויים במשוואה.

$$H_1: \beta_2 \neq 0$$

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



כאשר המשתנה המחאות נוסעים מצוי במשוואה.  
 $H_0: \beta_2 = \beta_3 = 0$   
 $H_1: otherwise$

### הנחות:

אופן שלושת הנחות של מודל רגרסיה מרובה. ההנחות מתייחסות לכל המשתנים ה"ב"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.  $\varepsilon_i \approx N(0, \sigma_\varepsilon^2)$  עבור כל קומבינציה ליניארית קבועה של המחאות נוסעים, יבוא ו-פעילות בחשבונות חיסכון.

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  $F$  מחושב  $< F_{k_{all} - k_{part}, n - k_{all} - 1}$ , 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<sup>a</sup>

		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 <sup>a</sup>	.156	.149	67.68720	.156	20.760	1	112	.000
2	.461 <sup>b</sup>	.213	.198	65.68529	.056	7.931	1	111	.006
3	.505 <sup>c</sup>	.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.

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

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 <sup>a</sup>	.156	.149	67.68720	.156	20.760	1	112	.000
2	.461 <sup>b</sup>	.213	.198	65.68529	.056	7.931	1	111	.006
3	.505 <sup>c</sup>	.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<sup>a</sup>

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 הב"ת- $r_{12} = 0$  מתקיים:

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

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

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

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

(5) Tolerance=1 (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

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

קיים מתאם נמוך יותר בין אותו משתנה ב"ת למשתנים ב"ת האחרים המצויים במשוואה.

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

נדגים באמצעות נתוני דוגמא מס' 2 :

Excluded Variables<sup>d</sup>

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

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

### Descriptive statistic

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

### Correlations

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

\*. 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
<b>1</b>	<b>a</b>			
<b>2</b>	<b>b</b>			<b>653.959</b>
<b>3</b>	<b>c</b>			<b>645.613</b>

a. Predictors: (Constant), EDLEVEL.

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

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

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

Excluded Variables <sup>d</sup>

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	TIME Job seniority (months)	.102 <sup>a</sup>	1.752	.082	.162	.974
	AGE Age of manager	.104 <sup>a</sup>	1.751	.083	.162	.927
	EXP Work experience (years)	.042 <sup>a</sup>	.716	.475	.067	.971
	WORKERS No. of workers	-.220 <sup>a</sup>	-2.188	.031	-.201	.321
2	TIME Job seniority (months)	.114 <sup>b</sup>	1.992	.049	.184	.966
	AGE Age of manager	.103 <sup>b</sup>	1.750	.083	.162	.927
	EXP Work experience (years)	.041 <sup>b</sup>	.707	.481	.066	.971
3	AGE Age of manager	.093 <sup>c</sup>	1.593	.114	.149	.919
	EXP Work experience (years)	.038 <sup>c</sup>	.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 <sup>a</sup>	.624	.618	658.748	.624	94.654	2	114	.000
2	.797 <sup>b</sup>	.636	.623	654.240	.012	1.788	2	112	.172
3	.808 <sup>c</sup>	.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 השכלה 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?**



## משתנה דמי

עד עתה המשתנים הב"ת שנכנסו לתוך מודל הרגרסיה היו כמותיים (סולם מדידה רווח או מנה) אולם ניתן גם להכניס משתנים ב"ת איכותיים (סולם מדידה שמי או סדר) לתוך מודל הרגרסיה. הכנסת משתנים איכותיים לרגרסיה מצריכה קידוד מיוחד של אותם המשתנים כ"משתני דמי".

### 1. יצירת משתני דמי

עבור משתנה ב"ת איכותי עם C קטגוריות ניצור C-1 משתני דמי.

משתני הדמי יכולים לקבל רק שני ערכים: 1=שייך לקטגוריה ו-0=לא שייך לקטגוריה. כל נבדק שייך לקטגוריה אחת בלבד.

לדוגמא:

סוג הסניף (BRANCH): משתנה איכותי שמי – עירוני(1), אזורי(2), עסקי(3)

מהמשתנה סוג הסניף יצרו שני משתני דמי: BRANCH1- עירוני (1), לא עירוני(0)

BRANCH2- אזורי (1), לא אזורי(0)

אדם שלא שייך לסניף העירוני (BRANCH1=0) ולא שייך גם לסניף האזורי (BRANCH2=0) בהכרח יהיה שייך לעסקי ולכן אין צורך לקודד כמשתני דמי את כל הקטגוריות.

### 2. רגרסיה עם משתנה ל"ת איכותי אחד בלבד

#### 2.1 בניית משוואת הרגרסיה

משוואת הרגרסיה עם משתני דמי נראית כך:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 \cdot \text{demi}1 + \hat{\beta}_2 \cdot \text{demi}2 + \dots + \hat{\beta}_{c-1} \cdot \text{demi}(c-1)$$

משתני הדמי (demi) יכולים כאמור לקבל שני ערכים בלבד: 0 או 1. ניבוי באמצעות המשוואה משמעו ניבוי Y עבור שייכות לאחת הקטגוריות של המשתנה הב"ת.

דוגמא מס' 3: בדיקת הקשר בין סוג הסניף (אותו קודדנו בסעיף הקודם כמשתני דמי) להיקף פעילות בכרטיסי אשראי.

משתני הדמי הם אלו שיכנסו אל תוך משוואת הרגרסיה:

**Variables Entered/Removed<sup>d</sup>**

Model	Variables Entered	Variables Removed	Method
1	Branch Type 2, Branch <sup>a</sup> Type 1		Enter

- a. All requested variables entered.  
 b. Dependent Variable: Credit Card

**להלן הממצאים:**

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	611.249	12.392		49.327	.000
	Branch Type 1	-417.047	17.640	-1.036	-23.643	.000
	Branch Type 2	-272.673	17.525	-.682	-15.559	.000

a. Dependent Variable: Credit Card

**בנה את משוואת הרגרסיה על סמך הפלט:**

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 \text{Branch1} + \hat{\beta}_2 \text{Branch2}$$

$$\hat{y}_i = 611.249 - 417.047 \text{Branch1} - 272.673 \text{Branch2}$$

**סוגי הניבויים שניתן לעשות על סמך המשוואה הנ"ל:**

ניבוי הפעילות בכרטיסי אשראי עבור סניף עירוני:  $\hat{y}_i = 611.249 - 417.047 = 194.202$

ניבוי הפעילות בכרטיסי אשראי עבור סניף אזורי:  $\hat{y}_i = 611.249 - 272.673 = 338.576$

ניבוי הפעילות בכרטיסי אשראי עבור סניף עסקי:  $\hat{y}_i = 611.249$

**2.2 השוואה בין מודל הרגרסיה עם משתני דמי לניתוח שונות חד גורמי**

שני המודלים – רגרסיה עם משתנה ב"ת איכותי אחד וניתוח שונות חד גורמי עם אותו משתנה ב"ת זהים לגמרי.

בדיקת השפעה של משתנה ב"ת איכותי על המשתנה התלוי יכולה אם כן להיעשות באמצעות שני המודלים כאשר התוצאות בשניהם תהינה זהות.

נדגים זאת בעזרת ניתוח שונות לנתוני דוגמא מס' 3:

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

## Report

Credit Card			
Branch type	Mean	N	Std. Deviation
city	194.2018	38	43.89064
regional	338.5756	39	58.16948
business	611.2487	39	112.02326
Total	382.9553	116	189.73656

ניתן לראות כי הממוצעים של ניתוח השונות עבור הקטגוריות של המשתנה הב"ת זהים לניבויים על קו הרגרסיה שחישבנו קודם לכן:

$$\hat{y}_i = 611.249 - 417.047 = 194.202 \quad \text{ניבוי הפעילות בכרטיסי אשראי עבור סניף עירוני:}$$

$$\hat{y}_i = 611.249 - 272.673 = 338.576 \quad \text{ניבוי הפעילות בכרטיסי אשראי עבור סניף אזרחי:}$$

$$\hat{y}_i = 611.249 \quad \text{ניבוי הפעילות בכרטיסי אשראי עבור סניף עסקי:}$$

בנוסף, טבלאות ה-ANOVA של ניתוח הרגרסיה ושל ניתוח השונות הן זהות:

ניתוח הרגרסיה:

### ANOVA<sup>b</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3463269	2	1731634.546	289.149	.000 <sup>a</sup>
	Residual	676726.5	113	5988.730		
	Total	4139996	115			

a. Predictors: (Constant), Branch Type 2, Branch Type 1

b. Dependent Variable: Credit Card

ניתוח השונות:

### ANOVA Table

		Sum of Squares	df	Mean Square	F	Sig.
Credit Card * Branch type	Between Groups	3463269	2	1731634.546	289.149	.000
	Within Groups	676726.5	113	5988.730		
	Total	4139996	115			

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

מובהקות משוואת הרגרסיה מעידה על מובהקות ניתוח השונות ולהיפך. ניתן לראות כי כל המדדים של הטבלאות זהים.

השפעה מובהקת באוכלוסיה של סוג הסניף על הפעילות בכרטיסי אשראי משמעה כי קיים הבדל מובהק באוכלוסיה בפעילות כרטיסי אשראי בין סוגי הסניף השונים.

בנוסף לכך גם מדדי הקשר  $r$  ו- $\eta^2$  (ופרופורציית השונות המוסברת  $r^2$  ו- $\eta^2$ ) הם זהים: ניתוח הרגרסיה:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.915 <sup>a</sup>	.837	.834	77.38689

a. Predictors: (Constant), Branch Type 2, Branch Type 1

ניתוח השונות:

Measures of Association

	Eta	Eta Squared
Credit Card * Branch type	.915	.837

ולבסוף משמעות מובהקת המקדמים היא כי ממוצע הקטגוריה המסוימת שונה מהממוצעים האחרים.

במקרה זה כל שיפועי הרגרסיה היו מובהקים:

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	611.249	12.392		49.327	.000
	Branch Type 1	-417.047	17.640	-1.036	-23.643	.000
	Branch Type 2	-272.673	17.525	-.682	-15.559	.000

a. Dependent Variable: Credit Card

כך שכל הממוצעים של פעילות האשראי בסניפים שונים מהאחרים כפי שבא לידי ביטוי בטבלת ה-homogenous subsets של ניתוח השונות:

### ISKA1 Credit Card

Tukey HSD<sup>a,b</sup>

BRANCH Branch type	N	Subset for alpha = .05		
		1	2	3
1 city	38	194.2018		
2 regional	39		338.5756	
3 business	39			611.2487
Sig.		1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

- a. Uses Harmonic Mean Sample Size = 38.661.
- b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

#### 2.2.1 שיחזור מקדמי הרגרסיה על פי הממוצעים

ניתן לשחזר את מקדמי הרגרסיה עם משתני הדמי מתוך ממוצעי ניתוח השונות:

$$\hat{\alpha} = \text{ממוצע הקטגוריה שלא הוגדרה ע"י משתנה דמי}$$

$$\hat{\beta}_1 + \hat{\alpha} = \text{ממוצע קטגורית משתנה דמי 1}$$

$$\hat{\beta}_2 + \hat{\alpha} = \text{ממוצע קטגורית משתנה דמי 2}$$

וכך הלאה.

נדגים על סמך נתוני דוגמא מס' 3:

בניתוח שונות לשאלת המחקר התקבל ש:

#### Report

Credit Card			
Branch type	Mean	N	Std. Deviation
city	194.2018	38	43.89064
regional	338.5756	39	58.16948
business	611.2487	39	112.02326
Total	382.9553	116	189.73656

ידוע כי:

מהמשתנה סוג הסניף יצרו שני משתני דמי : BRANCH1 - עירוני (1), לא עירוני (0)

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

## BRANCH2 - אזורי (1), לא אזורי (0)

על סמך הממוצעים של ניתוח השונות נשחזר את מקדמי הרגרסיה (האלפא והבטות) של ניתוח הרגרסיה :

$$\hat{\alpha} = 611.2487$$

$$\hat{\beta}_1 + \hat{\alpha} = 194.2018$$

$$\hat{\beta}_1 + 611.2487 = 194.2018$$

$$\hat{\beta}_1 = 194.2018 - 611.2487 = -417.04$$

$$\hat{\beta}_2 + \hat{\alpha} = 338.5756$$

$$\hat{\beta}_2 + 611.2487 = 338.5756$$

$$\hat{\beta}_2 = 338.5756 - 611.2487 = -272.67$$

משוואת הרגרסיה עם משתני הדמי:

$$\hat{y}_i = 611.2487 - 417.04 \text{Branch1} - 272.67 \text{Branch2}$$

### 2.3 השוואה בין מודל רגרסיה עם משתני דמי לרגרסיה פשוטה

מדובר במקרים בהם נתבקש להשוות בין מודל רגרסיה עם משתנה ב"ת איכותי (רגרסיה עם משתני דמי) למודל רגרסיה בו מתייחסים לאותו משתנה ב"ת ככמותי (רגרסיה פשוטה).

מאחר וכידוע מודל רגרסיה עם משתנה ב"ת איכותי זהה לניתוח שונות חד גורמי, אותם עקרונות ששימשו אותנו בהשוואה בין ניתוח שונות לרגרסיה פשוטה תקפים גם כאן (ראה סעיף בחלק של רגרסיה פשוטה).

בכדי לדעת איזה מודל עדיף לניתוח הנתונים-מודל רגרסיה עם משתני דמי או מודל רגרסיה פשוטה, יש לבחון האם הניבויים של משוואת הרגרסיה עם משתני הדמי "יושבים" על קו אחד ישר (כלומר כאשר עולים ביחידה אחת ב-X הקפיצה בין הניבויים היא פי קבוע). במידה וכן נעדיף להתייחס ל-X ככמותי ולנתח את הקשר בינו לבין Y במודל רגרסיה פשוטה. אם הניבויים לא "יושבים" על קו ישר נעדיף להתייחס ל-X כאיכותי ולנתח את הנתונים באמצעות רגרסיה עם משתני דמי (או ניתוח שונות).

בדוגמא מס' 3, למשל, הניבויים על קו הרגרסיה הינם:

$$\hat{y}_i = 194.202 \text{ ניבוי הפעילות בכרטיסי אשראי עבור סניף עירוני:}$$

$$\hat{y}_i = 338.576 \text{ ניבוי הפעילות בכרטיסי אשראי עבור סניף אזורי:}$$

$$\hat{y}_i = 611.249 \text{ ניבוי הפעילות בכרטיסי אשראי עבור סניף עסקי:}$$

ניתן לראות כי ניבויים אלו אינם יושבים על קו ישר (הקפיצה בין הניבויים איננה קבועה).

**?** חברה בודקת שיטות שונות לתמרוץ אנשי מכירות על מנת לבחור את גודל התמריץ שיגדיל

באופן משמעותי את המכירות. מנהל המכירות בחברה מעוניין לבדוק האם ככל שעולה גודל התמריץ לאנשי המכירות, עולה כמות המכירות. מנכ"ל החברה הגביל את גודל התמריץ (לכמות מכירות מעל לנורמה)

ל-120 ₪. לצורך מתן מענה לשאלה זו בוצע הניסוי הבא: נדגם מדגם מקרי של 80 אנשי מכירות באזורים בעלי מאפיינים דומים. 80 אנשי המכירות חולקו באופן מקרי במספר שווה ל-4 קבוצות כאשר כל קבוצה תומרצה באופן שונה: קבוצה א' תומרצה ב-90 ₪, קבוצה ב' תומרצה ב-100 ₪, קבוצה ג' תומרצה ב-110 ₪ וקבוצה ד' תומרצה ב-120 ₪. בתום תקופת הניסוי נרשם סך היחידות שנמכרו על ידי כל אחד מאנשי המכירות.

להלן חלק מן הממצאים:

תמריץ	Number	Mean
90 ₪	20	52
100 ₪	20	56
110 ₪	20	60
120 ₪	20	64

שני סטודנטים שהסתכלו על הנתונים הציעו לנתחם באמצעות שני מודלים שונים.

**סטודנט א'** הציע לבנות מודל רגרסיה עם משתני דמי באופן הבא:

$$\text{כאשר: } y_i = \alpha + \beta_1 D_{i1} + \beta_2 D_{i2} + \beta_3 D_{i3} + \epsilon_i$$

$D_{i1}$  שווה ל-1 אם איש המכירות ה-i קיבל תמריץ של 90 ₪ והוא שווה ל-0 אחרת.

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

$D_{i2}$  שווה ל-1 אם איש המכירות ה- $i$  קיבל תמריץ של 100 ₪ והוא שווה ל-0 אחרת.

$D_{i3}$  שווה ל-1 אם איש המכירות ה- $i$  קיבל תמריץ של 110 ₪ והוא שווה ל-0 אחרת.

סטודנט ב' הציע לנתח את הנתונים באמצעות מודל רגרסיה פשוטה כאשר  $x_i$  (משתנה כמותי)

גובה התמריץ של איש המכירות  $i$  -  $y_i$  מס' היחידות שמכר בתמריץ זה.

א. התייחס למודל הרגרסיה עם משתני הדמי: חשב את  $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$  על סמך הנתונים

והצג את משוואת הניבוי לכמות המכירות.

ב. איזה מודל לדעתך מתאים יותר לנתונים ולשאלת המחקר? הסבר.

### 3. רגרסיה המשלבת משתנים $\chi^2$ איכותיים וכמותיים

דוגמא מס' 4:

בדיקת הקשר בין סוג הסניף והיקף הפעילות בהלוואות להיקף פעילות בכרטיסי אשראי.

הרצה בשיטה ההיררכית לבדיקת התוספת של סוג הסניף על הלוואות.

להלן הממצאים:

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	.375 <sup>a</sup>	.140	.133	176.68312	.140	18.620	1	114	.000
2	.889 <sup>b</sup>	.790	.784	88.11468	.650	173.176	2	112	.000

a. Predictors: (Constant), X8 Loans

b. Predictors: (Constant), X8 Loans, BRANCH2 Branch Type 2, BRANCH1 Branch Type 1

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)



Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	131.604	60.515		2.175	.032
	X8 Loans	1.091	.253	.375	4.315	.000
2	(Constant)	400.612	33.718		11.881	.000
	X8 Loans	.806	.127	.277	6.350	.000
	BRANCH1 Branch Type 1	-369.364	20.221	-.918	-18.266	.000
	BRANCH2 Branch Type 2	-245.308	20.032	-.613	-12.246	.000

a. Dependent Variable: Y Credit Card

### 3.1 בניית קו הרגרסיה

הבטות של משתני הדמי יצטרפו לקבוע אלפא בהגדרת הקבוצה אליה שייכת התצפית ולאחריהם נוסיף את הבטות של המשתנים הכמותיים.

לדוגמא, נבנה את קו הרגרסיה הסופית באמצעות פלט המקדמים שלב שני:

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

? איזה פעילות בכרטיסי אשראי נוכל לחזות עבור לקוח השייך לסניף עירוני ואשר היקף הלוואותיו עומד על 252 ?

### 3.2 ניתוח הרגרסיה

העקרונות אותם למדנו לניתוח הרגרסיה תקפים גם לניתוח הרגרסיה בשילוב משתני דמי. יש לשים לב כי בניתוח אנו מתייחסים לכל אחד ממשתני הדמי כמשתנה בפני עצמו.

?

א. האם התוספת של הלוואות על פני סוג הסניף מובהקת? בדוק ברמת מובהקות של 0.05 (רשום הנחות, הצב השערות והסק מסקנה מילולית בלבד, ללא חישוב).

ב. בנה טבלת מקור שונות לבדיקת מובהקות המשוואה הסופית.

ג. האם התוספת לניבוי הפעילות בכרטיסי אשראי של סוג הסניף מעבר להיקף הלוואות מובהקת באוכלוסיה ברמת מובהקות של 0.05? (רשום הנחות, הצב השערות, חשב את סטטיסטי המבחן והסק מסקנה מילולית).

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

ד. כזכור, נמצא הבדל מובהק בהיקף הפעילות באשראי בין סוגי הסניף השונים בניתוח השונות. באם נעשה רגרסיה היררכית שבה משתנה סוג הסניף נכנס בשלב הראשון ואילו בשלב השני נכנס משתנה- הלוואות, האם על סמך תוצאות אלו ניתן לומר כי משתנה סוג הסניף חייב להיות מובהק הן בשלב הראשון של הרגרסיה (כאשר הוא היחיד במשוואה) והן בשלב השני, כאשר הלוואות מצוי גם הוא במשוואה? חווה דעתך.

# מבחן לדוגמה מספר 1

שאלה מס' 1 (31 נק')

מכון מחקר מעוניין לבחון האם רמת ההשכלה של אבות משפיעה על ציוני הפסיכומטרי של בניהם.  
לשם כך נדגמו באופן מקרי סטודנטים ונאספו נתונים לגבי המשתנים הבאים:

Psy score

Father education

להלן הפלטים שהתקבלו:

## Descriptives

psy score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
12 to lowest	348	613.13	63.474	3.403	606.43	619.82	414	767
proffesional	187	624.80	58.132	4.251	616.42	633.19	443	779
B.A	186	634.70	56.047	4.110	626.60	642.81	506	777
secont & third degree	157	651.86	57.823	4.615	642.74	660.98	497	792
Total	878	627.11	61.391	2.072	623.04	631.18	414	792

## Test of Homogeneity of Variances

psy score

Levene Statistic	df1	df2	Sig.
1.742	3	874	.157

psy score

father education	N	Subset for alpha = .05		
		1	2	3
Tukey HSD <sup>a,c</sup> 12 to lowest	348	613.13		
professional	187	624.80	624.80	
B.A	186		634.70	
secont & third degrees	157			651.86
Sig.		.207	.348	1.000

Means for groups in homogeneous subsets are displayed.

- a. Uses Harmonic Mean Sample Size = 200.331.
- b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

א. מתוך הנתונים לעיל, האם ניתן לצפות להבדל בציוני הפסיכומטרי של הבנים בין רמת ההשכלה הנמוכה ביותר לגבוהה ביותר של האבות? נמק. התייחס בתשובתך לרווחי הסמך בפלט הסטטיסטיקה התיאורית. (3 נק')

Descriptives

psy score

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
12 to lowest	348	613.13	63.474	3.403	606.43	619.82	414	767
professional	187	624.80	58.132	4.251	616.42	633.19	443	779
B.A	186	634.70	56.047	4.110	626.60	642.81	506	777
secont & third degree	157	651.86	57.823	4.615	642.74	660.98	497	792
Total	878	627.11	61.391	2.072	623.04	631.18	414	792

ב. בדוק את שאלת המחקר לפי השלבים הבאים: (28 נק')

1. הצב השערות. (1 נק')

2. רשום את ההנחות הדרושות (במונחי השאלה) למבחן הסטטיסטי שהנך בודק. (1 נק')

(נק')

3. בדוק את ההנחות הניתנות לבדיקה על סמך הפלט ברמת מובהקות  $\alpha = 0.05$

(הצב השערות והסק מסקנות במונחי השאלה). (3 נק')

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

4. האם ניתן לדעת ללא חישוב סטטיסטי F מה תהיה מסקנת החוקר? אם לא, מדוע?  
 אם כן, מהי? נמק תשובתך תוך הסתמכות על הפלט המתאים. (4 נק')
5. חשב את ערכו של הסטטיסטי F ונסח מסקנה מילולית במונחי השאלה. (7 נק')
6. חשב את פורפורצית השונות של ציוני הפסיכומטרי אשר איננה מוסברת על ידי  
 ההבדלים ברמות ההשכלה של האב. (4 נק')
7. בין אלו רמות השכלה של האב קיימים הבדלים בציוני הפסיכומטרי ברמת מובהקות  
 של 0.05? (4 נק')
8. האם ניתן להסיק בין אלו רמות השכלה של האב קיימים הבדלים בציוני הפסיכומטרי  
 ברמת מובהקות  $\alpha = 0.05$  גם על פי שיטת L.S.D? נא רשום תשובה מפורטת  
 ומנומקת עבור כל השוואה. (4 נק')

## שאלה 2 (5 נק')

במודל רגרסיה דו- משתנית נמצא מתאם אפס בין שני המשתנים הבלתי תלויים  $(x_1, x_2)$ . על כן הסיק החוקר שערך הסטטיסטי t בבדיקת מובהקות השיפוע  $x_1$  במודל הדו- משתני יהיה קטן יותר מערך הסטטיסטי t בבדיקת מובהקות השיפוע  $x_1$  במודל החד משתני. ציין נכון/לא נכון והסבר מדוע.

## שאלה 3 (64 נק')

מכון מחקר מעוניין לבחון את הגורמים הקשורים לגודל ההכנסה לנפש בקרב סטודנטים הלומדים לתואר שני. לשם כך נבדקו המשתנים הבאים:

Income - הכנסה חודשית בשקלים.

Psy score - ציוני הפסיכומטרי.

B.A - ציוני התואר הראשון.

Age - גיל.

Satisfaction - מידת שביעות הרצון מהעבודה.

להלן הפלטים שהתקבלו:

**Descriptive Statistics**

	Mean	Std. Deviation	N
incom	5013.3846	1269.73740	878
psy score	627.11	61.391	878
B.A	80.0125	4.02521	878
age	28.0194	2.84664	878
satisfaction	3.9943	.81095	878

### Correlations

		incom	psy score	B.A	age	שביעות רצון מהשכר
incom	Pearson Correlation	1	-.024	.487**	.713**	.348**
	Sig. (2-tailed)	.	.474	.000	.000	.000
	N	941	878	941	941	941
psy score	Pearson Correlation	-.024	1	-.014	-.052	-.008
	Sig. (2-tailed)	.474	.	.684	.122	.808
	N	878	878	878	878	878
B.A	Pearson Correlation	.487**	-.014	1	.685**	.192**
	Sig. (2-tailed)	.000	.684	.	.000	.000
	N	941	878	941	941	941
age	Pearson Correlation	.713**	-.052	.685**	1	.247**
	Sig. (2-tailed)	.000	.122	.000	.	.000
	N	941	878	941	941	941
שביעות רצון מהשכר	Pearson Correlation	.348**	-.008	.192**	.247**	1
	Sig. (2-tailed)	.000	.808	.000	.000	.
	N	941	878	941	941	941

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

### 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	.024 <sup>a</sup>	.001	-.001	270.08995	.001	.513	1	876	.474
2	.483 <sup>b</sup>	.233	.232	113.06795	.233	265.590	1	875	.000
3	.731 <sup>c</sup>	.535	.533	867.93362	.302	283.029	2	873	.000

a. Predictors: (Constant), psy score

b. Predictors: (Constant), psy score, B.A

c. Predictors: (Constant), psy score, B.A, satisfaction, age

### ANOVA<sup>d</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	827844.1	1	827844.146	.513	.474 <sup>a</sup>
	Residual	1.41E+09	876	1613128.479		
	Total	1.41E+09	877			
2	Regression	3.30E+08	2	164936581.6	133.129	.000 <sup>b</sup>
	Residual	1.08E+09	875	1238920.261		
	Total	1.41E+09	877			
3	Regression	7.56E+08	4	189072458.7	250.989	.000 <sup>c</sup>
	Residual	6.58E+08	873	753308.771		
	Total	1.41E+09	877			

a. Predictors: (Constant), psy score

b. Predictors: (Constant), psy score, B.A

c. Predictors: (Constant), psy score, B.A, satisfaction, age

d. Dependent Variable: incom

### Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5327.229	440.194		12.102	.000
	psy score	-.500	.699	-.024	-.716	.474
2	(Constant)	-6935.729	845.595		-8.202	.000
	psy score	-.363	.612	-.018	-.593	.553
	B.A	152.188	9.338	.482	16.297	.000
3	(Constant)	-4344.346	675.265		-6.434	.000
	psy score	.226	.478	.011	.473	.637
	B.A	-4.606	10.015	-.015	-.460	.646
	age	300.210	14.354	.673	20.915	.000
	satisfaction	293.621	37.294	.188	7.873	.000

a. Dependent Variable: incom



**Excluded Variables<sup>f</sup>**

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics	
					Tolerance	
1	B.A	.482 <sup>a</sup>	16.297	.000	.483	1.000
	age	.709 <sup>a</sup>	29.672	.000	.708	.997
	satisfaction	.349 <sup>a</sup>	11.026	.000	.349	1.000
2	age	.713 <sup>b</sup>	21.707	.000	.592	.528
	satisfaction	.267 <sup>b</sup>	9.262	.000	.299	.963

- a. Predictors in the Model: (Constant), psy score  
 b. Predictors in the Model: (Constant), psy score, B.A  
 c. Dependent Variable: incom

השאלות הבאות מתייחסות ל**מודל מספר 1** בלבד:

א. נניח שציוני הפסיכומטרי גדלים בסטיית תקן אחת. בכמה סטיות תקן צפויה לעלות ההכנסה?  
 (4 נק')

ב. מהו ממוצע הניבויים להכנסה בשקלים? (2 נק')

ג. בנה רווח בר סמך ברמת ביטחון של 95% לתוחלת ההכנסה עבור ציון פסיכומטרי של 600.  
 (רשום הנחות, חשב והסק מסקנה מילולית במונחי השאלה). (9 נק')

ד. בדוק את מובהקות המשוואה לניבוי ההכנסה על ידי ציוני הפסיכומטרי ברמת מובהקות של  $\alpha = 0.05$  (הצב השערות והסק מסקנה מילולית ללא חישוב על סמך תוצאות הפלט בלבד. ציין מאיזה פלט נלקחה התוצאה). (5 נק')

השאלות הבאות מתייחסות ל**מודלים 1+2** בלבד:

ה. אמוד את תוחלת ההכנסה של סטודנטים שציון הפסיכומטרי שלהם הוא 570 וציון ה-B.A שלהם הוא 85. (5 נק')

ו. מהי התוספת במונחי  $R^2$  של משתנה ה-B.A על המשתנה - ציון הפסיכומטרי? (2 נק')

ז. האם תוספת זו מובהקת? בדוק ברמת מובהקות  $\alpha = 0.05$  (רשום הנחות, הצב השערות והסק מסקנה מילולית ללא חישוב, על סמך תוצאות הפלט בלבד). מה הוא ה-sig לקבלת המסקנה? (6 נק')

ח. מה להערכתך המתאם בין ציוני הפסיכומטרי לבין ציוני ה-B.A (גבוה, נמוך או אפס). ציין שני נימוקים להערכתך. (6 נק')

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

ט. חוקר הסתכל בטבלת ה- Excluded Variables וגילה כי לאחר הכנסת המשתנה הראשון למשוואה, המשתנה בעל המתאם החלקי (Partial Correlation) הגבוה ביותר הוא גיל. למרות זאת המשתנה שנכנס למשוואה בשלב השני היה ה- B.A. כיצד ניתן להסביר זאת? נמק. (5 נק')

השאלות הבאות מתייחסות **למודל 3** בלבד:

י. מה גודל התרומה היחסית של המשתנה גיל. שחזר אותו באמצעות תוצאות אחרות בפלט. (4 נק')

יא. חשב את התוספת במונחי  $R^2$  של המשתנים גיל ומידת הסיפוק מהעבודה על המשתנים- ציוני הפסיכומטרי וה- B.A. (5 נק')

יב. האם תוספת זו מובהקת? בדוק ברמת מובהקות  $\alpha = 0.05$  (רשום הנחות, הצב השערות, חשב והסק מסקנה מילולית במונחי השאלה). (6 נק')

יג. על פי טבלת המקדמים (Coefficients), למשתנה ה- B.A אין תרומה מובהקת לניבוי ההכנסה במודל השלישי למרות שתרומתו לניבוי ההכנסה במודל השני  $\alpha$  כיצד ניתן להסביר זאת? נמק. (5 נק')

## מבחן לדוגמה מספר 2

שאלה 1(25 נק')

חוקרים מתחום הפסיכולוגיה החיובית בחנו את הגורמים המשפיעים על תחושת האופטימיות של האדם. לשם כך נדגמו באופן מקרי 1169 אנשים אשר נשאלו על מידת האופטימיות שהם חווים בחייהם (שנמדדה בסקלה רציפה של 1-4 כאשר 1- כלל לא אופטימי ואילו 4- אופטימי מאוד). המשתנה הראשון שנבדק היה גיל הנבדק בשנים.

Optimism

AGE

להלן הממצאים:

### Statistics

		optimizm	AGE
N	Valid	1201	1169
	Missing	0	32
Mean		2.5793	36.23
Minimum		1.00	18
Maximum		4.00	89

### Correlations

		optimizm	AGE
optimizm	Pearson Correlation	1	.060
	Sig. (2-tailed)	.	.041
	Sum of Squares and Cross-products	290.800	499.553
	Covariance	.242	.428
	N	1201	1169
AGE	Pearson Correlation	.060	1
	Sig. (2-tailed)	.041	.
	Sum of Squares and Cross-products	499.553	246114.2
	Covariance	.428	210.714
	N	1169	1169

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

(4) א. בנה את משוואת הרגרסיה לניבוי תחושת האופטימיות ע"י הגיל.

(6) ב. האם המשוואה לניבוי האופטימיות ע"י הגיל מובהקת? בדוק ברמת מובהקות של  $\alpha = 0.05$

(רשום הנחות, הצב השערות, חשב את סטטיסטי המבחן והסק מסקנה).

(5) ג. חשב אומדנים נקודתיים עבור תוחלת האופטימיות לערכים הבאים:

\*גיל = 17 שנה

\*גיל = 56 שנה

מי מבין שני האומדים הללו יעיל יותר בניבוי תוחלת האופטימיות? הסבר.

(5) ד. חשב רווח בר סמך לתוחלת האופטימיות ברמת ביטחון של 95%, עבור אדם שגילו 56 שנה.

(1) ה. מהי פרופורציית שונות האופטימיות המוסברת ע"י ההבדלים בגיל האדם?

(4) ו. בשלב השני הוחלט להוסיף למשוואה משתנה מנבא נוסף: ההכנסה של האדם בשקלים. באיזה

תנאי לא ישתנה מקדם השיפוע של הגיל כתוצאה מהוספת המשתנה הנוסף? הסבר.

(27 נק')

חוקר נוסף החליט לבדוק את משתני המחקר באופן שונה. הוא קיבץ את הגיל לחמש קטגוריות:

18-24

25-30

31-40

41-50

51-90

החוקר בחן את ההשפעה של חמשת רמות הגיל על האופטימיות.

להלן התוצאות שהתקבלו:

## Report

optimizm

new age	N	Std. Deviation
18-24	293	.46084
25-30	259	.47816
31-40	217	.51997
41-50	210	.50890
51-90	190	.50319
Total	1169	.49342

## Multiple Comparisons

Dependent Variable: optimizm

Tukey HSD

(I) new age	(J) new age	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
18-24	25-30	.0704	.04193	.448	-.0442	.1850
	31-40	.0350	.04403	.932	-.0853	.1553
	41-50	.0225	.04445	.987	-.0990	.1439
	51-90	-.0885	.04580	.301	-.2136	.0367
25-30	18-24	-.0704	.04193	.448	-.1850	.0442
	31-40	-.0354	.04525	.936	-.1590	.0882
	41-50	-.0479	.04566	.832	-.1727	.0768
	51-90	-.1589*	.04696	.007	-.2872	-.0305
31-40	18-24	-.0350	.04403	.932	-.1553	.0853
	25-30	.0354	.04525	.936	-.0882	.1590
	41-50	-.0125	.04759	.999	-.1425	.1175
	51-90	-.1234	.04885	.085	-.2569	.0100
41-50	18-24	-.0225	.04445	.987	-.1439	.0990
	25-30	.0479	.04566	.832	-.0768	.1727
	31-40	.0125	.04759	.999	-.1175	.1425
	51-90	-.1109	.04923	.161	-.2454	.0236
51-90	18-24	.0885	.04580	.301	-.0367	.2136
	25-30	.1589*	.04696	.007	.0305	.2872
	31-40	.1234	.04885	.085	-.0100	.2569
	41-50	.1109	.04923	.161	-.0236	.2454

\*. The mean difference is significant at the .05 level.

12) א. בדוק את שאלת המחקר לפי השלבים הבאים:

1. הצב השערות. מה מייצגים הפרמטרים אליהם מתייחסות השערות?

2. רשום את ההנחות הדרושות למודל במונחי השאלה.

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

3. בנה את טבלת מקור השונות וחשב את סטטיסטי המבחן.

4. רשום מסקנה במונחי השאלה ברמת מובהקות של  $\alpha = 0.05$ .

(4) ב. על סמך הנחת שיוויון השונות, ניתן להסיק כי השונות של קטגוריית הגיל- 31-40 –

$0.519^2$  היא אומדן חסר הטיה לשונות הטעויות. מה דעתך? הסבר.

(4) ג. מה ניתן לדעת על ערכי ה- sig בטבלת ה- multiple comparisons, לו בוצעו ההשוואות

בשיטת L.S.D? האם הם יגדלו, יקטנו או ישארו ללא שינוי? הסבר.

(2) ד. מהי פרופורציית שונות האופטימיות המוסברת ע"י ההבדלים בקטגוריות הגילאים?

(3) ה. ממה נובע ההבדל בין הפרופורציה שחושבה בסעיף ה' של השאלה הקודמת לבין

הפרופורציה שחושבה בסעיף הקודם? הסבר.

(2) ו. בהנחה שהחוקר היה בודק את שאלת המחקר באמצעות מודל רגרסיה עם משתני דמי. מה

תהיה פרופורציית שונות האופטימיות המוסברת ע"י ההבדלים בקטגוריות הגילאים? הסבר

מדוע.

החוקרים המשיכו לבחון את הגורמים המשפיעים על מידת האופטימיות של האדם ובדקו משתנים נוספים למשתנה הגיל.

להלן המשתנים שנבדקו:

optimism מידת האופטימיות (נמדדה בסקלה רציפה של 1- כלל לא אופטימי עד 4- אופטימי מאוד)

AGE גיל הנבדק בשנים

YRS OF SCHOOLING שנות לימוד

ROOMS מספר חדרים בבית

read paper המידה שבה קורא הנבדק עיתונים (בסקלה של 1-לא קורא כלל עד 5- קורא מדי יום)

tv news המידה שבה צופה הנבדק בחדשות בטלוויזיה (בסקלה של 1-לא צופה כלל עד 5- צופה מדי

יום)

להלן הפלטים שהתקבלו:

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

### Descriptive Statistics

	Mean	Std. Deviation	N
optimizm	2.5726	.49491	1108
AGE	35.80	14.316	1108
YRS OF SCHOOLING	13.15	2.358	1108
ROOMS	4.26	1.191	1108
read paper	4.22	1.077	1108
tv news	4.48	1.119	1108

### Correlations

		optimizm	AGE	YRS OF SCHOOLING	ROOMS	read paper	tv news
Pearson Correlation	optimizm	1.000	.055	.107	-.023	.079	.012
	AGE	.055	1.000	-.051	-.045	.051	.241
	YRS OF SCHOOLING	.107	-.051	1.000	.069	.028	-.048
	ROOMS	-.023	-.045	.069	1.000	.044	.030
	read paper	.079	.051	.028	.044	1.000	.371
	tv news	.012	.241	-.048	.030	.371	1.000
Sig. (1-tailed)	optimizm	.	.034	.000	.218	.004	.344
	AGE	.034	.	.044	.069	.045	.000
	YRS OF SCHOOLING	.000	.044	.	.010	.179	.055
	ROOMS	.218	.069	.010	.	.070	.162
	read paper	.004	.045	.179	.070	.	.000
	tv news	.344	.000	.055	.162	.000	.
N	optimizm	1108	1108	1108	1108	1108	1108
	AGE	1108	1108	1108	1108	1108	1108
	YRS OF SCHOOLING	1108	1108	1108	1108	1108	1108
	ROOMS	1108	1108	1108	1108	1108	1108
	read paper	1108	1108	1108	1108	1108	1108
	tv news	1108	1108	1108	1108	1108	1108

### Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.122 <sup>a</sup>	.015	.013	.49163
2	.126 <sup>b</sup>	.016	.013	.49165
3	.148 <sup>c</sup>	.022	.018	.49056

a. Predictors: (Constant), YRS OF SCHOOLING, AGE

b. Predictors: (Constant), YRS OF SCHOOLING, AGE, ROOMS

c. Predictors: (Constant), YRS OF SCHOOLING, AGE, ROOMS, read paper, tv news

ANOVA<sup>d</sup>

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.064	2	2.032	8.406	.000 <sup>a</sup>
	Residual	267.079	1105	.242		
	Total	271.143	1107			
2	Regression	4.283	3	1.428	5.906	.001 <sup>b</sup>
	Residual	266.860	1104	.242		
	Total	271.143	1107			
3	Regression	5.951	5	1.190	4.946	.000 <sup>c</sup>
	Residual	265.191	1102	.241		
	Total	271.143	1107			

a. Predictors: (Constant), YRS OF SCHOOLING, AGE

b. Predictors: (Constant), YRS OF SCHOOLING, AGE, ROOMS

c. Predictors: (Constant), YRS OF SCHOOLING, AGE, ROOMS, read paper, tv news

d. Dependent Variable: optimizm

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.195	.093		23.528	.000
	AGE	2.090E-03	.001	.060	2.022	.043
	YRS OF SCHOOLING	2.300E-02	.006	.110	3.666	.000
2	(Constant)	2.242	.105		21.271	.000
	AGE	2.049E-03	.001	.059	1.981	.048
	YRS OF SCHOOLING	2.341E-02	.006	.112	3.722	.000
	ROOMS	-1.18E-02	.012	-.029	-.952	.341
3	(Constant)	2.145	.122		17.512	.000
	AGE	2.123E-03	.001	.061	1.995	.046
	YRS OF SCHOOLING	2.269E-02	.006	.108	3.610	.000
	ROOMS	-1.29E-02	.012	-.031	-1.039	.299
	read paper	3.887E-02	.015	.085	2.630	.009
	tv news	-1.24E-02	.015	-.028	-.844	.399

a. Dependent Variable: optimizm

עפ"י מודל 1:

(3) א. חשב את פרופורציית השונות של תחושת האופטימיות המוסברת ע"י הגיל ומס' שנות הלימוד בלבד.

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)



- (6) ב. בנה את המשוואה לניבוי תחושת האופטימיות ע"י גיל ומס' שנות הלימוד.
- (8) ג. האם המשוואה לניבוי תחושת האופטימיות מובהקת? בדוק ברמת מובהקות של  $\alpha = 0.05$  (הצב השערות, חשב את הסטטיסטי והסק מסקנה במונחי השאלה).

עפ"י מודל 2:

- (3) ד. מה גודל התוספת במונחי  $R^2$  של מספר החדרים בבית על גיל ומספר שנות לימוד?
- (6) ה. האם תוספת זו מובהקת? בדוק ברמת מובהקות של  $\alpha = 0.05$  (רשום הנחות, הצב השערות, חשב והסק מסקנה במונחי השאלה).
- (2) ו. כיצד מתיישבת התוצאה שקיבלת בסעיף ה' עם תוצאת ה-  $AdjR^2$  שהתקבלה במודל 2?
- (4) ז. השוואה את התוצאה שהתקבלה בסעיף ה' לתוצאת טבלת ה-ANOVA של מודל 2. ממה נובעים ההבדלים בתוצאות?  
הסבר.

עפ"י כל המודלים:

- (3) ח. מהי המשוואה האופטימלית לניבוי תחושת האופטימיות מבין השלושה שקיימים? כיצד הגעת למסקנה זו?

לאחר קבלת הממצאים הנ"ל החוקר שיער שגם המגדר של הנבדק יכול להשפיע על תחושת האופטימיות שהוא חש.

הוא הגדיר את משתנה הדמי הבא:

Male = 1 אם גבר ו- 0 אם אשה.

לפתרונות מלאים בסרטוני וידאו היכנסו ל- [www.Gool.co.il](http://www.Gool.co.il)

להלן הפלט שהתקבל:

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.122 <sup>a</sup>	.015	.013	.49163
2	.126 <sup>b</sup>	.016	.013	.49165
3	.148 <sup>c</sup>	.022	.018	.49056
4	.159 <sup>d</sup>	.025	.020	.48991

- a. Predictors: (Constant), YRS OF SCHOOLING, AGE
- b. Predictors: (Constant), YRS OF SCHOOLING, AGE, ROOMS
- c. Predictors: (Constant), YRS OF SCHOOLING, AGE, ROOMS, read paper, tv news
- d. Predictors: (Constant), YRS OF SCHOOLING, AGE, ROOMS, read paper, tv news, MALE

(4) ט. חשב את גודל התרומה הייחודית, במונחי  $R^2$ , של המגדר כאשר כל יתר המשתנים הב"ת נמצאים במשוואה.

(5) י. האם תרומה זו מובהקת? בדוק ברמת מובהקות של  $\alpha = 0.05$  (הצב השערות, חשב והסק מסקנה במונחי השאלה).

בנוסף התקבל פלט ה- Excluded Variables המוצג להלן:

**Excluded Variables<sup>d</sup>**

Model	Beta In	t	Sig.	Partial Correlation	Collinearity Statistics	
					Tolerance	
1	ROOMS	-.029 <sup>a</sup>	-.952	.341	-.029	.993
	read paper	.073 <sup>a</sup>	2.449	.014	.074	.996
	tv news	.003 <sup>a</sup>	.098	.922	.003	.941
	MALE	.062 <sup>a</sup>	2.088	.037	.063	.999
2	read paper	.075 <sup>b</sup>	2.495	.013	.075	.994
	tv news	.004 <sup>b</sup>	.140	.888	.004	.939
	MALE	.062 <sup>b</sup>	2.074	.038	.062	.999
3	MALE	.059 <sup>c</sup>	1.977	.048	.059	.997

a. Predictors in the Model: (Constant), YRS OF SCHOOLING, AGE

b. Predictors in the Model: (Constant), YRS OF SCHOOLING, AGE, ROOMS

c. Predictors in the Model: (Constant), YRS OF SCHOOLING, AGE, ROOMS, read paper, tv news

d. Dependent Variable: optimizm

(4) יא. אם הנתונים היו מעובדים בשיטת stepwise, איזה משתנה היה נכנס לתוך משוואת הרגרסיה בשלב השני? הסבר.

ציין שני הבדלים עיקריים בין השיטה שבה עובדו הנתונים במחקר המנכחי לבין שיטת stepwise.