

## שאלות ממבחנים-רגרסיה

מבחן מס' 1

### שאלה 1

להלן תוצאות הרצת רגרסיה של  $Y$  בתלות ב-  $X$  עבור 10 תצפיות (חלק מהנתונים הושמטו בכוונה מהפלט, אך ניתנים לחישוב על ידך).

$$\text{נתון כי: } \sum (X_i - \bar{X})^2 = 1475.6$$

מקור	סכום ריבועים SS
רגרסיה	$SSR = 2148.6$
שאריות	$SSE = ?$
סה"כ	$SST = ?$

משתנה	מקדם	טעות תקן	ערך סטטיסטי(מתוקנן) t	p-value מובהקות
קבוע (חותך)	- 24.7	11.3	?	?
X	1.20	?	10.5	

א. מהו SST?

א. לא ניתן לקבוע.

ב. 1994.42

ג. 2304.1

ד. 1629.78

ב. האם הרגרסיה מובהקת? בדוק לפי p value

א. הרגרסיה מובהקת

ב. הרגרסיה אינה מובהקת.

### שאלה מס' 2

לפניך פלט רגרסיה פשוטה (ממנו הושמטו נתונים שבאפשרותך להשלים), המתאר את ציון המבחן כפונקציה של מספר התרגילים שהגיש הסטודנט במהלך הסמסטר, ידוע כי כל הנחות המודל תקפות.

#### SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.842105598
R Square	0.709141839
Adjusted R Square	0.688366256
Standard Error	5.315523758
Observations	16

## ANOVA

	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	964.4329004	964.4329004	34.13343	4.28E-05
Residual	14	395.5670996	28.25479283		
Total	15	1360			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	50.99134199	4.318918368	11.80650747	1.15E-08
מספר התרגילים	4.086580087	0.699471573	5.842381939	4.28E-05

1. ע"פ הנתונים, אחוז השונות של ציוני המבחן המוסברת ע"י מספר התרגילים שהגיש הסטודנט, היא \_\_\_\_\_ . אם נוסף משתנים נוספים, אחוז השונות המוסברת \_\_\_\_\_, ו-  $R_{\text{adjusted}}$  \_\_\_\_\_ .

א. 84% יגדל, לא ניתן לקבוע ללא נתונים נוספים.

ב. 84% , יקטן, יגדל.

ג. 70.9% , יגדל, לא ניתן לקבוע ללא נתונים נוספים.

ד. 70.9% , יגדל, יקטן.

ה. 68.8% יגדל, יגדל.

ו. 68.8% יקטן, יקטן.

2. מהו הרב"ס של שיפוע הרגרסיה  $\beta_1$ ? (בר"מ של 1%)

א.  $2.58 < \beta < 5.58$

ב.  $2 < \beta < 6.17$

ג.  $1.74 < \beta < 6.88$

ד.  $2.86 < \beta < 5.3$

במטרה לנבא בצורה טובה יותר את הצלחת הסטודנטים בבחינה, החליט החוקר להוסיף 2 משתנים נוספים לניתוח הרגרסיה.

מספר השיעורים בהם נכח הסטודנט, ומספר השעות שלמד לבחינה.

לפניכם הפלט החסר:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.880163577
R Square	0.774687922
Adjusted R Square	0.718359902
Standard Error	5.053253294
Observations	16

ANOVA

	df	SS	MS	F
Regression	3	?	351.1918579	13.75315
Residual	12	306.4244263	25.53536886	
Total	15	?		

	Coefficients	Standard Error	t Stat	P-value
Intercept	37.08959571	8.726136945	4.250402663	0.001127
מספר השיעורים	2.610000755	1.429904588	1.82529714	0.092932
מספר התרגילים	3.198068014	1.239162836	2.58082951	0.024061
שעות לימוד לבחינה	-0.108373802	1.021202927	-0.106123669	0.917238

1. בדוק את ההשערה כי הרגרסיה מובהקת לכל אחד מהמשתנים המסבירים בר"מ של 1%.

- א. הרגרסיה אינה מובהקת לכל המשתנים שנבדקו.
- ב. לא ניתן לקבוע מהנתונים האם הרגרסיה מובהקת.
- ג. הרגרסיה מובהקת למשתנה מספר התרגילים, אך אינה מובהקת למשתנים מספר השיעורים ומספר שעות הלימוד לבחינה.
- ד. הרגרסיה מובהקת למשתנה מספר התרגילים ומספר השיעורים, אך אינה מובהקת למשתנה שעות הלימוד לבחינה.

2. מהו SSR של הרגרסיה המרובה?

- א.  $SSR=964.4$ .
- ב.  $SSR=1053.57$ .
- ג.  $SSR=694.57$ .
- ד.  $SSR=853.57$ .
- ה. לא ניתן לחשב את SSR מהנתונים שהתקבלו.

## מבחן מס' 2

### שאלה מס' 1

הועלתה השערה שהוצאות האחזקה של מערכת לעיבוד נתונים קשורות למספר שעות השימוש השבועיות במערכת. להלן תוצאות חלקיות של פלט EXCEL של ניתוח רגרסיה בין- Y הוצאות אחזקה שנתיות (במאות \$) ו- X מספר שעות השימוש השבועיות.

#### SUMMARY OUTPUT

<i>Regression Statistics</i>					
Multiple R					
R Square					
Adjusted R Square					
Standard Error					
Observations 10					

  

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression		860.051			0.00012
Residual					
Total		1004.525			

  

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	10.528	3.745		0.023
שעות שימוש	0.953		6.901	

א. לאור התוצאות ה- p-value בבדיקת ההשערה:

$$H_0: \beta_1 \leq 0.5$$

$$H_1: \beta_1 > 0.5$$

1.  $Pv < 0.005$

2.  $0.005 < pv < 0.01$

3.  $0.01 < pv < 0.025$

4.  $0.025 < pv < 0.05$

5.  $Pv > 0.05$

ב. אם היינו מריצים רגרסיה בה- Y מספר שעות השימוש השבועיות ואילו המשתנה המסביר- X, הוצאות אחזקה שנתיות (במאות \$), אזי השיפוע של קו הרגרסיה יהיה:

1. 0.953

2. 1.049

3. 10.528

4. 0.095

5. 0.898

## שאלה מס' 2

הועלתה השערה שמספר התקלות ברכב Y קשורה לגיל הנהג X. לשם כך נלקח מדגם של 10 נהגים.

כמו כן חושבו הסכומים הבאים:

$$\sum X_i^2 = 14,227 \quad \sum X_i = 363 \quad \sum Y_i = 13 \quad \sum Y_i^2 = 29 \quad \sum X_i Y_i = 366$$

משוואת קו הרגרסיה נתונה על ידי:  $\hat{Y} = 4.96 - 0.1X_i$

א. ערכו של מקדם המתאם הלינארי בין מספר התקלות לבין גיל הנהג הוא:

1. -10.59

2. -0.9395

3. 0.8826

4. -0.1

החוקר לא היה מרוצה מעוצמת הקשר ולכן החליט להוסיף לרגרסיה את המשתנים הבאים :

מספר הק"מ שהמכונית נסעה (באלפי ק"מ) וסוג הרכב. במדגם נכללו 2 סוגי רכבים: A ו B

כאשר סוג רכב B קודד כערך 0 וסוג רכב A קודד בערך 1.

להלן פלט הרגרסיה המרובה. שימו לב כי חלק מהערכים בפלט חסרים:

#### SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	
R Square	
Adjusted R Square	
Standard Error	0.204047
Observations	10



## ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	3			0.00002
Residual	6	0.249811		
Total	9			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	1.407943	0.71094		
X1 גיל הנהג	-0.03094	0.014611		
X2 סוג הרכב	0.239574	0.152994		
X3 ק"מ באלפים	0.027666	0.005329		

ב. להלן מספר טענות לגבי מובהקות הרגרסיה ומובהקות המשתנה סוג הרכב ברמת מובהקות 0.1:

1. הרגרסיה מובהקת והמשתנה סוג הרכב מובהק ברמת מובהקות 0.1.
2. הרגרסיה מובהקת, אך המשתנה סוג הרכב אינו מובהק ברמת מובהקות 0.1.
3. הרגרסיה אינה מובהקת, אך המשתנה סוג הרכב מובהק ברמת מובהקות 0.1.
4. הרגרסיה והמשתנה סוג הרכב אינם מובהקים ברמת מובהקות 0.1.

ג. SST בפלט הרגרסיה המרובה שווה ל-

1. 12.1

2. 16.0

3. 20.0

4. אין מספיק נתונים לחשבו.

ד. לאור התוצאות, רב"ס למקדם המשתנה מספר הקילומטרים, בר"מ 5%:

1.  $(0.015, 0.04)$

2.  $(0.017, 0.038)$

3.  $(0.02, 0.035)$

4. אין מספיק נתונים לחשבו.

ה. אם היינו מקודדים את סוג רכב B כערך 1 וסוג רכב A קודד בערך 0 משואת הרגרסיה הייתה:

1. נשאר ללא שינוי

2.  $\hat{Y} = 1.4079 - 0.0309X_{1i} - 0.2395X_{2i} + 0.0276X_{3i}$

3.  $\hat{Y} = 1.6475 - 0.0309X_{1i} - 0.2395X_{2i} + 0.0276X_{3i}$

4. לא ניתן לדעת ללא הרצה מחדש

ו. החוקר רצה להוסיף משתנה מסביר נוסף - X4 מספר השנים שחלפו מאז קבלת רישיון הנהיגה. להלן פלט הרגרסיה (חלק מהנתונים חסרים) עם 4 המשתנים המסבירים:

	Coefficients	Standard Error	t Stat	P-value
Intercept	1.5	0.71		
X1 גיל הנהג	5.1	14.1		
X2 סוג הרכב	0.25	0.2		
X3 ק"מ באלפים	0.02	0.008		
מס' השנים שחלפו מאז קבלת רשיון X4	-5.13094	0.8		

על סמך נתונים אלו:

1. חשש סביר למולטיקוליניאריות
2. התחזית הנקודתית של מספר התקלות תהיה דומה לזו של הרגרסיה הקודמת (עם 3 המשתנים המסבירים).
3. קיימת קורלציה גבוהה בין חלק מהמשתנים המסבירים.
4. כל התשובות נכונות.

### שאלה מס' 3

במפעל מסוים הורץ מודל של רגרסיה ליניארית פשוטה על 8 עובדים כאשר  $Y$  - תפוקת העובד ו-  $X$  - גיל העובד. נמצא שהחלק המוסבר על ידי הרגרסיה הוא 74.

הטעויות שהתקבלו מופיעות בחלקן בטבלה שלהלן:

$e_7$	$e_6$	$e_6$	$e_5$	$e_4$	$e_3$	$e_2$	$e_1$
0	2	-2	2	?	2	-3	0

אחת מהטענות שלהלן נכונה:

1. מקדם ההסבר בין X ל-Y הוא 0.74

2. לא ניתן לחשב את מקדם המתאם המרובה

3.  $SSR=18$

4.  $SSE=74$

5. אף אחת מהטענות איננה נכונה

#### שאלה מס' 4

חברה בדקה את הקשר בין המכירות השבועיות לבין הוצאות הפרסום. מנהל השיווק בדק מדגם של 30 שבועות ומצא כי קו הרגרסיה לניבוי המכירות השבועיות על סמך הוצאות הפרסום הוא:

$$\hat{y} = 1.25x - 0.5$$

איזה מהמשפטים הבאים אינו אפשרי?

- א. קיים קשר מושלם בין המכירות השבועיות לבין הוצאות הפרסום.
- ב. קיים קשר חזק וחיובי בין המכירות השבועיות לבין הוצאות הפרסום.
- ג. קו הרגרסיה לניבוי הוצאות הפרסום על סמך המכירות השבועיות מתלכד עם הקו הנתון.
- ד. קו הרגרסיה לניבוי הוצאות הפרסום על סמך המכירות השבועיות אינו מתלכד עם הקו הנתון.