רגרסיה מרובה
וישומי הרגרסיה

ב. אם ידוע שהנבדק קיבל את הציון $\bar{x}$ וציון זה הוא 70, האם עובדה זאת תעזור לך לנבא את ציונו ב-y? הסבר.

ג. אם ידוע שהמתאם בין x ל-y הוא 1 והנבדק קיבל את הציון $\bar{x} = 70$. האם עובדה זאת תעזור לך לנבא את ציונו ב-y? הסבר.

שאלה 4

חוקר א' טוען כי יש קשר ליניארי חיובי בין רמת ההכנסה של ההורים (x_1) לבין אוצר המלים שיש לילד (x_2).

ברגרסיה שביצע (x_2 - מנובא, x_1 - מנבא) נתקבל מתאם של 0.40 בין שני משתנים אלה.

חוקר ב' טוען כי המתאם בין המשתנים הללו הוא מזויף ולמעשה מושפע ממשתנה שלישי: מידת החשיפה של הילד לגרויים (x_3).

בבדיקה שערך לצורך זה מצא כי:

$$Sx_1 = 10 \qquad Cov_{1,3} = 0.81$$
$$Sx_2 = 8 \qquad Cov_{2,3} = 3.60$$
$$Sx_3 = 5$$

א. מה תצטרך לחשב על מנת לבחון האם חוקר ב' צדק בטענתו? הסבר.

ב. חשב את מה שציינת ב- א'.

ג. האם חוקר ב' צדק? הסבר בקצרה.

שאלה 5

האסיסטנט של חוקר א' משאלה 4 טוען כי המתאם האמיתי בין x_1 לבין x_2 גדול יותר למעשה מזה שנתקבל במחקרו (0.40). לטענתו חוקר א' קיבל מתאם זה משום שבדק זאת על מדגם שכלל ילדים מכל שכבות האוכלוסיה.

כדי לבדוק טענה זאת הציע האסיסטנט לבדוק שוב את המתאם אך הפעם רק לגבי הילדים הגרים בשכונות מצוקה.

א. האם צדק האסיסטנט בטענתו? הסבר בקצרה.

ב. אילו היתה נערכת הבדיקה שהציעה האסיסטנט, מה היה גובה המתאם שהיית מצפה לו, נמוך או גבוה מזה שחוקר א' קיבל? הסבר.

שאלה 3

התשובות בסעיף זה, רלוונטיות בהנחה שקיים קשר ליניארי כלשהו, שעליו יכולה להתבסס רגרסיה

חלק א

נבדק שקיבל ב X את הממוצע, ציון ב Y יהיה Y ממוצע. ניתן להוכחי זאת באמצעות נוסחת הניבוי (ראה/י בחוברת).

חלק ב

העובדה שידוע כי הציון הממוצע והמנבא המתקבל של X הוא 70, לא מסייעת בניבוי הציון של Y.

באופן עקרוני התשובה כאן זהה לתשובה הקודמת. הציון שנתקבל יתאפס מול הציון הממוצע (ביניהם קיימת פעולת חיסור), ולכן הערך המנובא יהיה שווה לערכו של Y ממוצע.

ניתן לומר, כי אם הציון המתקבל שווה לממוצע, הניבוי של Y יהיה לפי הציון הממוצע של Y.

חלק ג

חלק ג

ההבדל בין הסעיף הזה לקודמיו הוא שכאשר יש מתאם של 1, אין רגרסיה לממוצע ובעצם כל X בתחום הבעייה, יביא לתוצאה מדוייקת (ולא ניבוי) של Y - אם ידועה משוואת ההתאמה ("רגרסיה" עם מתאם 1).

המקרה הפרטי שבו ידוע מהו ממוצע X, לא מוסיף יכולת לנבא את Y. מכאן שעדיין המתאם לא ממש עוזר. אין לנו נתונים על התפלגות Y.

שאלה 4

חלק א

כדי לבחון את טענתו של חוקר ב', יש לחשב את $\Gamma_{12.3}$. המתאם בין 1 לבין 2 בניטרול 3.

על מנת לחשב זאת ניעזר גם בחישוב Γ_{13} , Γ_{23}. מתוך הנתונים הקיימים.

חלק ב

$$\Gamma_{13} = \frac{COV_{13}}{S_1 S_3} = \frac{0.81}{10*5} = 0.0162$$

$$\Gamma_{23} = \frac{COV_{23}}{S_2 S_3} = \frac{3.60}{8*5} = 0.09$$

$$\Gamma_{12.3} = \frac{0.4 - 0.0162*0.09}{\sqrt{(1-0.0162^2)(1-0.09^2)}} = \frac{0.3985}{\sqrt{(0.9997)(0.9919)}} = 0.4002 \cong 0.4$$

חלק ג

החוקר לא צדק. בניטרול הגורם הנוסף (מידת החשיפה של הילד לגרויים) התקבל מתאם דומה מאד למקורי.

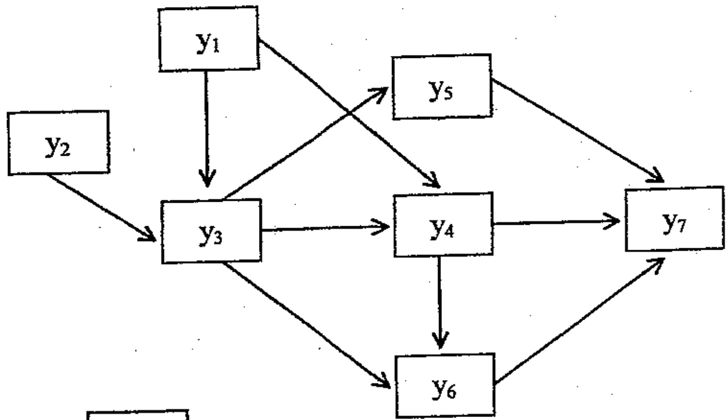
שאלה 5

האסיסטנט לא צודק. קיצוץ תחום הרגרסיה היה מקטין את המתאם ולא מגדיל אותו.

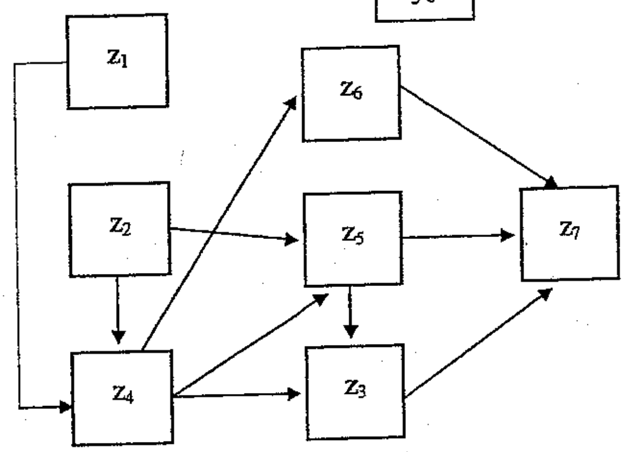
שאלה 2

נתונים תרשימים של שתי תיאוריות.

תאוריה א'



תאוריה ב'



לגבי **כל אחד מהתרשימים** שלעיל ענה בנפרד על השאלות הבאות:

- (2 נק') . מהם המשתנים החיצוניים והפנימיים?
- (6 נק') . כמה משוואות רגרסיה יש להציב על מנת לקבל את כל מקדמי הנתיב שבמודל.
- (8 נק') . הכן טבלה במבנה הבא:

משוואה	משתנה תלוי	משתנה/ים בלתי תלוי/ים	מקדמי הנתיב שתקבל מן המשוואה
1			
2			
3			

(16 נק')   ד.   פרק את המתאם בין x_2 ל-x_7 למרכיביו השונים.

פתרון שאלה 2

תיאוריה א'

א. משתנים חיצוניים : y1,y2

משתנים פנימיים : y3,y4,y5,y6,y7

ב. 5 משוואות רגרסיה, כמספר המשתנים הפנימיים.

ג. משוואות הרגרסיה לבניית המודל

משוואה	משתנה תלוי	משתנה/ים בלתי תלוי/ים	מקדמי הנתיב שתקבל מן המשוואה
1	y3	y1,y2	p_{31}, p_{32}
2	y4	y3,y2	p_{43}, p_{41}
3	y5	y3	p_{53}
4	y6	y3,y4	p_{63}, p_{64}
5	y7	y4,y5,y6	p_{74}, p_{75}, p_{76}

ד. פירוק מתאם

$$r_{27} = \underbrace{P_{32}*P_{43}*P_{74}}_{\text{עקיף}} + \underbrace{P_{32}*P_{63}*P_{76}}_{\text{עקיף}} + \underbrace{P_{32}*P_{53}*P_{75}}_{\text{עקיף}} + \underbrace{P_{32}*P_{43}*P_{64}*P_{76}}_{\text{עקיף}}$$

תיאוריה ב'

א. משתנים חיצוניים : z1,z2

משתנים פנימיים : z3,y4,z5,y6,z7

ב. 5 משוואות רגרסיה, כמספר המשתנים הפנימיים.

ג. משוואות הרגרסיה לבניית המודל

משוואה	משתנה תלוי	משתנה/ים בלתי תלוי/ים	מקדמי הנתיב שתקבל מן המשוואה
1	z3	Z5,z4	p_{35}, p_{34}
2	z4	Z1,z2	p_{42}, p_{41}
3	z5	Z2,z4	p_{52}, P_{54}
4	z6	z4	p_{64}
5	z7	Z3,z5,z6	p_{73}, p_{75}, p_{76}

ד. פירוק מתאם – גם כאן כולם עקיפים.

$$r_{27} = \underbrace{P_{52}*P_{75}}_{\text{עקיף}} + \underbrace{P_{42}*P_{34}*P_{73}}_{\text{עקיף}} + \underbrace{P_{42}*P_{54}*P_{35}*P_{73}}_{\text{עקיף}} + \underbrace{P_{42}*P_{64}*P_{76}}_{\text{עקיף}} + P_{52}*P_{35}*P_{73} + P_{42}*P_{54}*P_{75}$$

שאלה 3

נתונים ציוניהם של 18 תלמידים בשני מקצועות: x ו-y.

x	y	f(x,y)
70	75	2
75	70	3
85	85	1
80	65	2
90	80	1
50	45	1
60	65	2
55	50	1
65	60	3
70	55	2

(4 נק') א. חשב את המתאם בין x ל-y.

(3 נק') ב. מהי שונות טעויות הניבוי?

(3 נק') ג. מהי שונות הניבויים?

(5 נק') ד. האם ניתן היה לנבא מתוך התוצאה בסעיף א' את היחס בין שונות הטעויות לשונות הניבויים.

(5 נק') ה. חשב את קווי הרגרסיה לניבוי y מתוך x ולניבוי x מתוך y.

(4 נק') ו. צייר דיאגרמת פיזור של הנתונים והעבר בה את שני קווי הרגרסיה.

פתרון שאלה 3

טבלת חישובי עזר

מספר סידורי	X1	Y	$x1^2$	y2	x1*y
1	70	75	5250	5625	5250
2	70	75	4900	5625	5250
3	75	70	5625	4900	5250
4	75	70	5625	4900	5250
5	75	70	5625	4900	5250
6	85	85	7225	7225	7225
7	80	65	6400	4225	5200
8	80	65	6400	4225	5200
9	90	80	8100	6400	7200
10	50	45	2500	2025	2250
11	60	65	3600	4225	3900
12	60	65	3600	4225	3900
13	55	50	3025	2500	2750
14	65	60	4225	3600	3900
15	65	60	4225	3600	3900
16	65	60	4225	3600	3900
17	70	55	4900	3025	3850
18	70	55	4900	3025	3850
סכומים	1260	1170	90350	77850	83275

$$\bar{x} = 1260 / 18 = 70 \quad \bar{y} = 1170/18 = 65$$

א. חישוב המתאם בין X ל- Y

$$Sx = \frac{\sqrt{18 * 90{,}000 - 1260^2}}{18} = 10 \quad Sy = \frac{\sqrt{18 * 77{,}850 - 1170^2}}{18} = 10$$

$$rxy = \frac{\frac{83275}{18} - 70 * 65}{10 * 10} = 0.763$$

ב. שונות טעויות הניבוי $S^2e = \left[1 - 0.763^2\right] * 10^2 = 0.418 * 100 = 41.8$

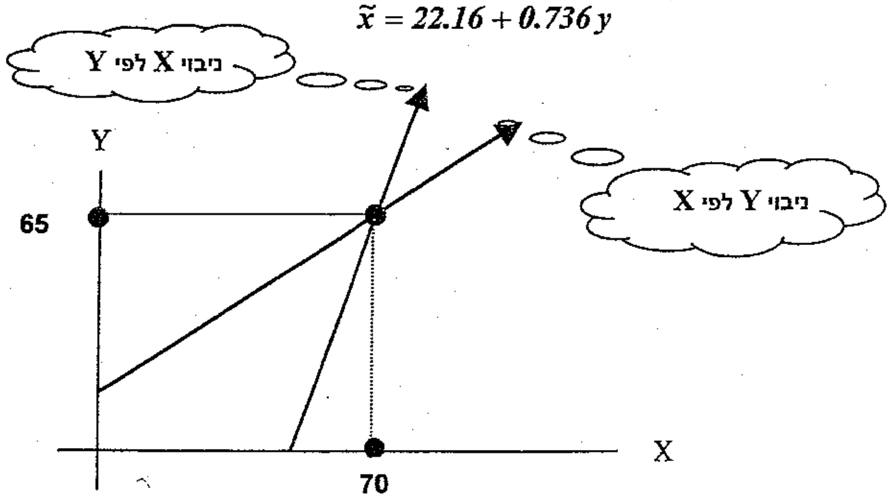
ג. שונות הניבויים $S^2\tilde{y} = 0.763^2 * 10^2 = 0.582 * 100 = 58.2$

ד. היחס בין שונות הטעויות לשונות הניבויים הוא יחס בין פרופורציה למשלים שלה. יחס זה נובע מהמתאם בין המשתנים. ככל שהמתאם גבוה יותר טיב הניבוי גבוה יותר, כלומר החלק של שונות הניבויים מסך כל השונות של הקריטריון גבוה יותר. העלאת המתאם בריבוע והכפלתו ב- 100 נותנת את אחוז השונות המוסברת. החלק המשלים ל- 100% הוא אחוז הטעויות.

ה. קו הרגרסיה.

$$b = 0.736 \frac{10}{10} = 0.736$$
$$a = 65 - 0.736 * 70 = 13.48$$
$$\tilde{y} = 13.48 + 0.736X$$
$$b = 0.736 \frac{10}{10} = 0.736$$
$$a = 70 - 0.736 * 65 = 22.16$$
$$\tilde{x} = 22.16 + 0.736y$$

ו.



ניבויים

אם x=60 אז הניבוי ל- Y 13.48+0.736*60=57.64

אם x=80 אז הניבוי ל- Y 13.48+0.736*80=72.36

אם y=60 אז הניבוי ל- X 22.16+0.736*60=66.32

אם y=80 אז הניבוי ל- X 22.16+0.736*80=81.04

שאלה 4

חוקר טוען כי רמת העצמאות של הילד האחרון במשפחה תלויה במספר הילדים שיש לאמו. ככל שלאם יש יותר ילדים כך גדלה מידת העצמאות של הילד האחרון במשפחה.

במחקר שערך לקחו חלק משפחות, שהלידות הסתיימו בהן, אך שלהן עד שלושה ילדים – 50 אמהות ו50- ילדים אחרונים. נתקבל מתאם של 0.30 בין מספר הילדים שיש להם ((x_1) לבין מידת העצמאות של הילד האחרון ((x_2).

חוקר ב' טוען, כי למעשה שני המשתנים הללו מושפעים מרמת העצמאות של האם עצמה (x_3.)

בבדיקה שערך על אותם נתונים מצא כי:

$$\sum(x_{i1}-\overline{x}_1)^2=1500$$
$$n\sum x_{i2}^2=5500 \quad ; \quad \left(\sum x_{i2}\right)^2=1200$$
$$Sx_3=0.8$$
$$Cov(1,3)=0.59$$
$$Cov(2,3)=0.20$$

(5 נק') א. מה תצטרך לחשב על מנת לבחון האם חוקר ב' צודק בטענתו? הסבר בקצרה.

(6 נק') ב. חשב את מה שציינת ב-א'.

(3 נק') ג. האם חוקר ב' צדק?

פתרון שאלה 4

על מנת לבדוק את הטענה צריך לחשב מתאם חלקי. נחשב את המתאם בין מספר הילדים לבין מידת העצמאות של היליד האחרון בניכוי ההשפעה של רמת העצמאות של האם (שונות משתנה זה מוסרת מן החישוב).

$$Sx1=\sqrt{\frac{1500}{50}}=5.477 \quad Sx2=\frac{\sqrt{5500-1200}}{50}=1.311$$

$$r1,3=\frac{0.59}{5.477*0.8}=0.134 \quad r2,3=\frac{0.2}{1.311*0.8}=0.1906$$

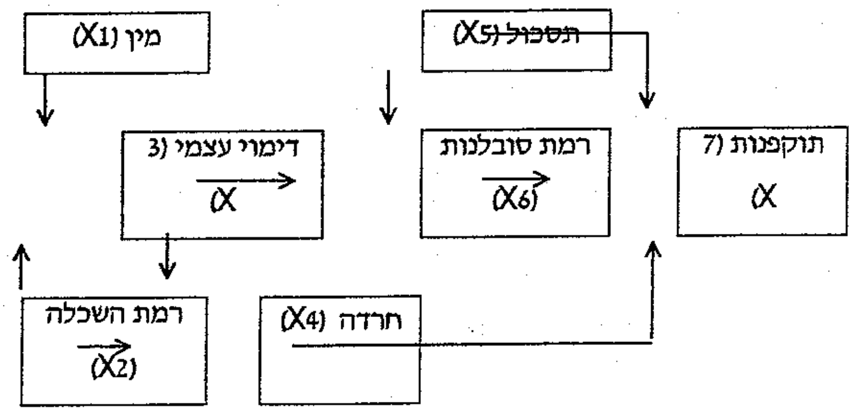
$$r12.3=\frac{0.3-0.134*0.1906}{\sqrt{(1-0.134^2)*(1-0.1906^2)}}=\frac{0.27446}{\sqrt{0.982*0.963}}=\frac{0.27446}{0.97245}\approx 0.2822$$

המתאם החלקי מעט יותר קטן. ההשפעה של X3נמוכה ואפשר להגיד יחסית זניחה. לכן ניתן לומר כי חוקר ב' אינו צודק. ניתן לומר שהוא צודק כי בכל זאת חל שינוי. כמובן שלא ניתן לדעת בוודאות כי לא עשינו מבחן מובהקות לתוצאות.

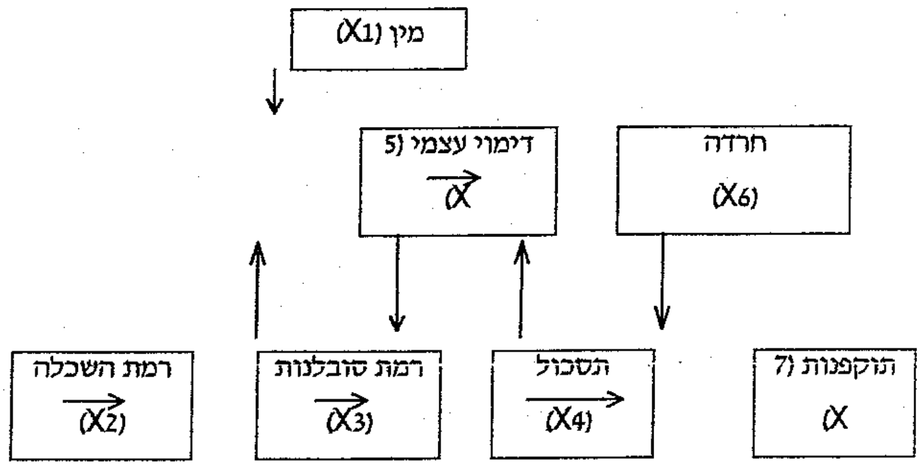
שאלה

לפניך שני תרשימי נתיבים המתארים את הקשר בין מין, רמת השכלה, תסכול, חרדה, דימוי עצמי, רמת סובלנות לתוקפנות.

<u>תאוריה א'</u>



<u>תאוריה ב'</u>



לגבי כל אחד מהתרשימים ענה על השאלות הבאות :

א.	מה הם המשתנים החיצוניים והפנימיים?
ב.	כמה משוואות רגרסיה יש להציב על מנת לקבל את כל מקדמי הנתיב שבמודל?
ג.	הכן טבלה במבנה הבא :

משוואה	משתנה תלוי	משתנה/ים בלתי תלוי/ים	מקדמי הנתיב שתקבל מן המשוואה
1 2 3 . .			

ד.	פרק את המתאם בין X3 לבין X7 למרכיביו השונים. עבור אלו מהחיצים מקדמי הנתיב זהים למתאם הפשוט בין המשתנים המעורבים?
ה.	מר בקורת בחן את תאוריה א' וטען כי אם X5 (תסכול) משפיע הן על תוקפנות (X7) והן על רמת סובלנות (X6) הרי מדובר על קשר מזויף בין רמת סובלנות (X6) לתוקפנות (X7) והקשר הישיר בין שני משתנים אלו שגוי. מחבר התאוריה השיב לו כי הקשר בין רמת סובלנות ותוקפנות יכול להיות גם מזויף וגם ישיר.
(1)	מי מבין החוקרים צודק? נמק.
(2)	ניתן למצוא מצבים דומים גם בתאוריה ב'? אם כן, תן דוגמה אחת.

שאלה 2

חברת פרסום שוקלת שיווק מוצר חדש לשוק. במסגרת זו נערך מחקר בו נבדקה יכולת הניבוי של שני משתנים : גיל (x_2) וסטטוס חברתי-כלכלי (x_3), על הערכת הצורך במוצר מסוג כזה (x_1).

המחקר הועבר ל-10 נבדקים. לכל נבדק תואר המוצר והועבר שאלון בו היה עליו להעריך את חיוניותו של מוצר מסוג כזה.

להלן התוצאות :

x_1	8	5	8	9	10	10	9	9	8	3

x_2	20	30	40	40	60	70	30	30	30	25
x_3	1	1	2	3	5	5	4	3	3	2

(המשך השאלה בעמוד הבא)

א. מהם המנבאים ומהו הקריטוריון?

ב. האם זו רגרסיה פשוטה או מרובה? הסבר בקצרה.

ג. מהם ערכי ה-b? האם הם מובהקים ב- $\alpha=0.05$?

ד. מהו ערך ה-a?

ה. מהו ערך המתאם המרובה? האם הוא מובהק ב- $\alpha=0.05$?

ו. מהו אחוז השונות המוסברת באמצעות שני המנבאים גם יחד?

ז. מהי התוספת בהסבר של משתנה הסטטוס החברתי-כלכלי מעבר להסבר של משתנה הגיל?

שאלה 3

חוקר א' טוען כי בקרב חולי לב, מידת ההסתגלות למחלה תלויה במידת החרדה של החולה. על מנת להוכיח טענתו, ערך מחקר בו בחן אצל 200 חולי לב את מידת החרדה ומידת ההסתגלות למחלה. הוא ביצע רגרסיה חד-משתנית בה המנובא היה מידת ההסתגלות של החולה (X2) והמנבא מידת החרדה (X1). נתקבל מתאם 0.50 בין שני המשתנים.

חוקר ב' טוען כי המתאם בין שני המשתנים האלה הוא מזויף ולמעשה מושפע ממשתנה שלישי: חומרת המחלה (X3). בבדיקה שערך לצורך זה מצא כי:

$$\sum(x_{i1}-\bar{x}_1)^2 = 1000$$
$$n\sum x_{i2}^2 = 5000$$
$$\left(\sum x_{i2}\right)^2 = 1000$$
$$Sx_3 = 0.5$$
$$Cov(1,3) = 0.45$$
$$Cov(2,3) = 0.112$$

א. מה תצטרך לחשב על-מנת לבחון האם חוקר ב' צדק בטענתו? הסבר.

ב. חשב את מה שציינת בסעיף א'.

ג. האם חוקר ב' צדק? הסבר.

<u>חוקר</u> בדק את הקשר בין ציוניהם של 60 סטודנטים לרפואה בפסיכומטרי לבין ציוניהם בתום שנה א' וקיבל מתאם של 0.28 (מובהק).

האסיסטנט שלו, <u>מר למדני</u>, אומר שלמעשה המתאם בין ציוני הפסיכומטרי לבין ציוני סוף שנה א' באוניברסיטה גבוה יותר מזה שנתקבל. לטענתו, החוקר קיבל מתאם נמוך, משום שחשב אותו רק עבור הסטודנטים לרפואה ולא עבור כלל הסטודנטים באוניברסיטה.

אסיסטנט אחר, <u>מר שקדני</u>, גורס כי המתאם אכן גבוה יותר אך הסיבה לקבלת מתאם נמוך היא אחרת. הסיבה לכך היא שהחוקר בדק את הציון של הסטודנטים שהתקבלו לרפואה ולא של כל אלו שניסו להתקבל לפקולטה לרפואה.

<u>החוקר</u> השיב להם כי טענותיהם רק מחזקות את העובדה כי הפסיכומטרי אכן מנבא ציונים של סטודנטים לרפואה בתום שנה א'.

א.	הסבר כל אחת משלוש הטענות.
ב.	לו היתה נערכת כל אחת מן הבדיקות שהציעו האסיסטנטים מה היה גובה המתאם שהיית מצפה לו – נמוך, שווה או גבוה מזה שקיבל החוקר עצמו?
ג.	באילו מההצעות – זו של מר למדני או זו של מר שקדני – היה מתקבל מתאם גבוה יותר? הסבר מהם הנתונים להם אתה זקוק כדי להחליט.

<u>שאלה 1</u>

<u>תאוריה א'</u>

א. משתנים חיצוניים: X1, X2, X5

משתנים פנימיים: X3, X4, X6, X7

ב. מספר משוואות הרגרסיה הוא כמספר המשתנים הפנימיים במודל – 4.

ג.

מס' משוואה	משתנה תלוי	משתנים בלתי תלויים	מקדמי הנתיב
1	X3	X1, X2	P31, P32
2	X4	X2, X3	P42, P43
3	X6	X3, X5	P63, P65
4	X7	X4, X5, X6	P74, P75, P76

ד. בין X1 ל-X2 אין קשר, מאחר ושנייהם משתנים חיצוניים, ולכן P31=r31 ו-P32=r32. בין X3 ל-X5 אין קשר, ולכן P63=r63 ו-P65=r65.

ה. $r37 = P_{43}P_{74} + P_{63}P_{76} + P_{32}P_{42}P_{74}$.

1. השפעה ישירה ← אין.

2. השפעה עקיפה ← $X_3 \rightarrow X_4 \rightarrow X_7 \quad P_{43}P_{74}$

$P_{63}P_{76} \quad X_3 \rightarrow X_6 \rightarrow X_7$

3. השפעה מזוייפת ←

$$\begin{array}{ccc} X_2 & \rightarrow & X_3 \\ \downarrow & & \\ X_4 & \rightarrow & X_7 \end{array} \qquad P_{32}P_{42}P_{74}$$

ו. 1. מחבר התיאוריה טוען שההשפעות השונות עליהן מר ביקורת מדבר אינן מבטלות זו את זו (כפי שטוען מר ביקורת), אלא מתווספות אחת על השנייה. מחבר התיאוריה צודק, מאחר ומודל הנתיבים הינו מודל אדיטיבי. הקשר בין המשתנים יכול לנבוע בחלקו ממתאם מזויף ובחלקו מקשר ישיר.

2. בתיאוריה ב' ניתן לראות שתסכול (X4) משפיע על חרדה (X6) ועל תוקפנות (X7) ובין X6 ל-X7 קיים קשר ישיר.

<u>תאוריה ב'</u>

א. משתנים חיצוניים : X1, X2.

משתנים פנימיים : X3-X7.

ב. מספר משוואות הרגרסייה הם כמספר המשתנים הפנימיים – 5.

ג.

מס' משוואה	משתנה תלוי	משתנים בלתי תלויים	מקדמי הנתיב
1	X3	X2	P32
2	X4	X3, X5	P43, P45
3	X5	X1, X3	P51, P53
4	X6	X4, X5	P64, P65
5	X7	X4, X6	P74, P76

ד. P32=r32. בין X1 ל-X3 לא קיים קשר, ולכן גם P51=r51 ו-P53=r53.

$r37 = P_{43}P_{74} + P_{53}P_{65}P_{76} + P_{43}P_{64}P_{76} + P_{53}P_{45}P_{64}P_{76} + p53p45p74$.

אין השפעה ישירה ואין השפעה מזוייפת, אלא רק השפעות עקיפות.

שאלה 2

א. המנבאים הם גיל (X2) וסטטוס חברתי כלכלי (X3=X1).

הקריטריון הינו הערכת הצורך במוצר (X1=Y).

ב. רגרסייה מרובה, שכן ישנם 2 משתנים מנבאים לניבוי הקריטריון.

ג.

$$
\begin{aligned}
& n=10; \\
& \sum y = 8+5+8+9+10+10+9+9+8+3=79 \quad \bar{y} \quad 7.9 \\
& \sum x_1 = 1+1+2+3+5+5+4+3+3+2=29 \quad \bar{x}_1 \quad 2.9 \\
& \sum x_2 = 20+30+40+40+60+70+30+30+30+25=375 \quad \bar{x}_2 \quad 37.5 \\
& \sum y^2 = 8^2+5^2+8^2+9^2+10^2+10^2+9^2+9^2+8^2+3^2=669 \\
& \sum x_1^2 = 1^2+1^2+2^2+3^2+5^2+5^2+4^2+3^2+3^2+2^2=103 \\
& \sum x_2^2 = 20^2+30^2+40^2+40^2+60^2+70^2+30^2+30^2+30^2+25^2=16,325 \\
& \sum yx_1 = 8\cdot 1 \;\; 5\cdot 1 \;\; 8\cdot 2 \;\; 9\cdot 3 \;\; 10\cdot 5 \;\; 10\cdot 5 \;\; 9\cdot 4 \;\; 9\cdot 3 \;\; 8\cdot 3 \;\; 3\cdot 2 \quad 249 \\
& \sum yx_2 = 8\cdot 20 \;\; 5\cdot 30 \;\; 8\cdot 40 \;\; 9\cdot 40 \;\; 10\cdot 60 \;\; 10\cdot 70 \;\; 9\cdot 30 \;\; 9\cdot 30 \;\; 8\cdot 30 \;\; 3\cdot 25 \quad 3,145 \\
& \sum x_1x_2 = 1\cdot 20 \;\; 1\cdot 30 \;\; 2\cdot 40 \;\; 3\cdot 40 \;\; 5\cdot 60 \;\; 5\cdot 70 \;\; 4\cdot 30 \;\; 3\cdot 30+3\cdot 30 \;\; 2\cdot 25 \quad 1,250
\end{aligned}
$$

$$S_y = \frac{\sqrt{n\cdot\sum y^2-\left(\sum y\right)^2}}{n} = \frac{\sqrt{10\cdot 669-(79)^2}}{10} = \frac{21.18}{10} = 2.12$$

$$S_{x_1} = \frac{\sqrt{n\cdot\sum x_1^2-\left(\sum x_1\right)^2}}{n} = \frac{\sqrt{10\cdot 103-(29)^2}}{10} = \frac{13.74}{10} = 1.37$$

$$S_{x_2} = \frac{\sqrt{n\cdot\sum x_2^2-\left(\sum x_2\right)^2}}{n} = \frac{\sqrt{10\cdot 16,325-(375)^2}}{10} = \frac{150.41}{10} = 15.04$$

$$\mathrm{cov}(1,y) = \frac{\sum x_1 y}{n} - \bar{x}_1 \bar{y} \quad \frac{249}{10} = 2.9 \;\; 7.9 \quad 1.99$$

$$\mathrm{cov}(2,y) = \frac{\sum x_2 y}{n} - \bar{x}_2 \;\; \bar{y} \quad \frac{3,145}{10} - 37.5 = 7.9 \quad 18.25$$

$$\mathrm{cov}(1,2) = \frac{\sum x_1 x_2}{n} - \bar{x}_1 \;\; \bar{x}_2 \quad \frac{1,250}{10} = 2.9 \;\; 37.5 \quad 16.25$$

$$r(1,y) = \frac{\mathrm{cov}(1,y)}{S_{x_1}\cdot S_y} = \frac{1.99}{1.37\cdot 2.12} = 0.68$$

$$r(2,y) = \frac{\mathrm{cov}(2,y)}{S_{x_2}\cdot S_y} = \frac{18.25}{15.04\cdot 2.12} = 0.57$$

$$r(1,2) = \frac{\mathrm{cov}(1,2)}{S_{x_1}\cdot S_{x_2}} = \frac{16.25}{1.37\cdot 15.04} = 0.79$$

$$b_1 = \frac{S_y}{S_{x_1}}\left(\frac{r_{1y}-r_{2y}\;\; r_{12}}{1-r_{12}^2}\right) = \frac{2.12}{1.37}\left(\frac{0.68-0.57\;\; 0.79}{1-0.79^2}\right) = 0.95$$

$$b_2 = \frac{S_y}{S_{x_2}}\left(\frac{r_{2y}-r_{1y}\;\; r_{12}}{1-r_{12}^2}\right) = \frac{2.12}{15.04}\left(\frac{0.57-0.68\;\; 0.79}{1-0.79^2}\right) = 0.01$$

מבחן מובהקות למקדמי הרגרסיה (b_j)

$$R^2_{y\cdot 12}=\frac{r_{1y}^2+r_{2y}^2-2\cdot r_{1y}r_{2y}r_{12}}{1-r_{12}^2}=\frac{0.68^2+0.57^2-2\cdot 0.68\cdot 0.57\cdot 0.79}{1-0.79^2}=0.47$$

$$Sb_1=\sqrt{\frac{(1-R^2)\cdot S_y\cdot n/(n-k-1)}{n\cdot S_{x_1}^2\left(1-r_{12}^2\right)}}=\sqrt{\frac{(1-0.47)\cdot 2.12\cdot 10/(7)}{10\cdot 1.37^2\left(1-0.79^2\right)}}=0.69$$

$$Sb_2=\sqrt{\frac{(1-R^2)\cdot S_y\cdot n/(n-k-1)}{n\cdot S_{x_2}^2\left(1-r_{12}^2\right)}}=\sqrt{\frac{(1-0.47)\cdot 2.12\cdot 10/(7)}{10\cdot 15.04^2\left(1-0.79^2\right)}}=0.06$$

$$t_1=\frac{b_1}{Sb_1}=\frac{0.95}{0.69}=1.38 \qquad t_{c_{n-k-1}}^{\frac{\alpha}{2}}=t_{c_7}^{0.025}=2.365$$

$$t_2=\frac{b_2}{Sb_2}=\frac{0.01}{0.06}=0.17$$

מקדמי הרגרסייה של שני המשתנים המנבאים אינם מובהקים סטטיסטית, מאחר ו- $t_1,t_2<t$, כלומר תרומתם הייחודית של כל אחד מהם בניבוי קריטריון Y אינה מובהקת.

ד. $a=\bar{y}-b_1\bar{x}_1-b_2\bar{x}_2=7.9-0.95\cdot 2.9-0.01\cdot 37.5=4.77$.

ה. מבחן מובהקות לשונות המוסברת:

$$F=\frac{R^2/k}{(1-R^2)/(n-k-1)}=\frac{0.47/2}{(1-0.47)/(7)}=3.1$$

$$F^{\alpha}_{c_{k,n-k-1}}=F^{0.05}_{(2,7)}=4.74$$

התוצאה אינה מובהקת סטטיסטית ($F<F_c$), כלומר שני המנבאים יחד אינם מצליחים לנבא את הקריטריון, כלומר השונות המוסברת ע״י שניהם אינה מובהקת.

ו. השונות המוסברת ב-Y ע״י שני המנבאים היא $R^2_{y\cdot 12}=0.47$.

ז. התוספת בהסבר של משתנה הסטטוס החברתי-כלכלי (X1) מעבר להסבר של משתנה הגיל (X2) היא: $R^2_{y\cdot 12}-r^2_{2y}=0.47-0.57^2=0.145$, כלומר X1 הוסיף 14.5% על השונות.

<u>שאלה 3</u>

א. חוקר ב׳ טוען שהקשר בין 2 המשתנים X1 (מידת החרדה) ל-X2 (מידת ההסתגלות של החולה) מושפע ממשתנה שלישי (חומרת המחלה – X3). ע״מ לבחון זאת יש לחשב את המתאם החלקי של X1 ו-X2 בנטרול X3 ולהשוותו למתאם המלא.

ב. נתון: $r(1,2)=0.5; n=200$

$$S_{x_2}=\frac{\sqrt{n\cdot\sum x_2^2-\left(\sum x_2\right)^2}}{n}=\frac{\sqrt{5,000-1,000}}{200}=0.32$$

$$r(2,3)=\frac{\text{cov}(2,3)}{S_{x_2}\cdot S_{x_3}}=\frac{0.112}{0.32\cdot 0.5}=0.7$$

$$S_{x_1}=\sqrt{\frac{\sum\left(x_{i_1}-\bar{x}\right)^2}{n}}=\sqrt{\frac{1,000}{200}}=\sqrt{5}=2.23$$

$$r(1,3)=\frac{\text{cov}(1,3)}{S_{x_1}\cdot S_{x_3}}=\frac{0.45}{2.23\cdot 0.5}=0.4$$

$$r_{12\cdot 3}=\frac{r_{12}-r_{23}\ \ r_{13}}{\sqrt{\left(1-r_{23}^2\right)\left(1-r_{13}^2\right)}}=\frac{0.5-0.7\ \ 0.4}{\sqrt{\left(1-0.7^2\right)\left(1-0.4^2\right)}}=0.36$$

ג. קיים הבדל בין המתאם המתאם החלקי (0.36) למתאם המלא (0.5), כלומר למשתנה הנוסף X3 יש השפעה מסויימת על המתאם המלא.

שאלה 4

א. הן מר למדני והן מר שקדני טוענים שהחוקר התייחס במחקרו רק לאותם סטודנטים לרפואה, כאשר מר למדני טוען שאינו התייחס לכלל הסטודנטים באוניברסיטה ומר שקדני טוען שהחוקר לא התייחס לכל אותם סטודנטים שניסו להתקבל לפקולטה לרפואה (אלו שהצליחו ואלו שלא הצליחו). אם כן שניהם טוענים שהחוקר צימצם את טווח הערכים של שני המשתנים – גם של ציוני הפסיכומטרי וגם של ציונייהם בתום שנה א'. בהנחה שהקשר הוא לינארי (לא נאמר אחרת), קיצוץ התחום הקטין את השונות של המשתנים וכתוצאה מכך הקטין גם את הסיכוי שהם ימצאו בקשר אחד עם השני. ולכן בשני המקרים המתאם שהתקבל בפועל נמוך ביחס למתאם שהיה מתקבל.

ב. לו הייתה נערכת כל אחת מהבדיקות שהציעו האסיסטנטים שונות המשתנים היתה גדלה ומחזקת את הסיכוי של המשתנים להיות בקשר זה עם זה, ולכן גובה המתאם שהייתי מצפה לו היה גבוה מזה שקיבל החוקר עצמו (בהנחה שקיים קשר לינארי בין השניים).

ג. כלל הסטודנטים באוניברסיטה כולל בתוכו את הסטודנטים שניסו להתקבל לפקולטה לרפואה ואלו כוללים בתוכם את הסטודנטים לרפואה (אותם אלה שהצליחו להתקבל). כלומר, מר למדני שמתייחס לכלל הסטודנטים באוניברסיטה, בעצם טוען שקיצוץ התחום שביצע החוקר רחב יותר מזה של מר שקדני (המתייחס לסטודנטים שניסו להתקבל לפקולטה לרפואה), משמע שהשונות אצלו קטנה יותר מזו של מר שקדני.
במילים אחרות, במידה והיו פועלים בהתאם להצעותיהם של האסיסטנטים היה מתקבל מתאם גבוה יותר אצל מר למדני, שכן השונות אצלו גדולה יותר מזו של מר שקדני.

שאלה 5 (39 נקי)

נבדקה יכולת ניבוי של רמת המוטיבציה (x_1) והכושר הגופני (x_2) על מהירות ריצה (y) של 100 ילדים בני 8 . ציוני המוטיבציה נעים בין :

(0) מוטיבציה נמוכה ועד (20) מוטיבציה גבוהה.

וציוני הכושר הגופני נעים בין :

(0) כושר לקוי ועד (100) כושר גופני מעולה.

רמת מוטיבציה (x_1)	15	5	16	15	8	18	17	12	14	15
כושר גופני (x_2)	90	30	90	75	30	95	90	40	70	80
מהירות ריצה (y)	12	20	14	15	18	12	14	17	16	15

(2 נקי) א. מהם המנבאים ומהו הקריטריון?

(10 נקי) ב. חשב את קו הניבוי לניבוי מהירות ריצה מתוך רמת המוטיבציה והכושר הגופני.

(5 נקי) ג. חשב את ערך המתאם המרובה ואת אחוז השונות המוסברת בניבוי מהירות ריצה מתוך כושר גופני ורמת המוטיבציה.

(8 נקי) ד. האם ערכי b מובהקים ב- $\alpha = 0.05$?

(5 נקי) ה. האם ערך המתאם המרובה מובהק ב- $\alpha = 0.05$?

(5 נקי) ו. מהי התוספת בהסבר של משתנה כושר הגופני (x_1) על פני רמת מוטיבציה (x_2) ?

(5 נקי) ז. האם אפשר היה להקיש על תוצאות אלו מן הערכים של b? הסבר.

תשובה 4

א. המנבאים: רמת מוטיבציה (X_1) ועושר לשוני (X_2)
הקריטריון: מהירות קריאה (y).

ב.

שם המשתנה	סכום הציונים (ΣX_i)	סכום ריבועי הציונים (ΣX_i^2)	ממוצע
X_1	135	1973	13.5
X_2	690	53650	69
y	153	2399	15.3

$\Sigma X_{1i} X_{2i} = 10205$
$\Sigma X_{1i} y_i = 1980$
$\Sigma X_{2i} y_i = 10005$

$$S_1 = \frac{\sqrt{n \cdot \Sigma X_{1i}^2 - (\Sigma X_{1i})^2}}{n} = \frac{\sqrt{10 \times 1973 - 135^2}}{10} = \frac{\sqrt{1505}}{10} = 3.879$$

$$S_2 = \frac{\sqrt{n \cdot \Sigma X_{2i}^2 - (\Sigma X_{2i})^2}}{n} = \frac{\sqrt{10 \times 53650 - 690^2}}{10} = \frac{\sqrt{60400}}{10} = 24.576$$

$$S_y = \frac{\sqrt{n \cdot \Sigma y_i^2 - (\Sigma y_i)^2}}{n} = \sqrt{10 \times 2399 - 153^2} = \frac{\sqrt{581}}{10} = 2.41$$

$$Cov(1,2)=\frac{\sum X_{1i}X_{2i}}{n}-\bar{X}_1\times\bar{X}_2=\frac{10205}{10}-13.5\times 69=\underline{\underline{89}}$$

$$Cov(1,y)=\frac{\sum X_{1i}Y_i}{n}-\bar{X}_1\bar{Y}=\frac{1980}{10}-13.5\times 15.3=\underline{\underline{-8.55}}$$

$$Cov(2,y)=\frac{\sum X_{2i}Y_i}{n}-\bar{X}_2\bar{Y}=\frac{10005}{10}-69\times 15.3=\underline{\underline{-55.2}}$$

$$r_{(1,2)}=\frac{Cov(1,2)}{S_1\cdot S_2}=\frac{89}{3.879\times 24.576}=\underline{0.934}$$

$$r_{(1,y)}=\frac{Cov(1,y)}{S_1\cdot S_y}=\frac{-8.55}{3.879\times 2.41}=\underline{-0.915}$$

$$r_{(2,y)}=\frac{Cov(2,y)}{S_2\cdot S_y}=\frac{-55.2}{24.576\times 2.41}=\underline{-0.932}$$

$$b_1=\frac{S_y}{S_1}\left[\frac{r_{1y}-r_{2y}\cdot r_{12}}{1-r_{12}^2}\right]=\frac{2.41}{3.879}\times\left[\frac{-0.915-0.934\times(-0.932)}{1-0.934^2}\right]=\underline{\underline{-0.217}}$$

$$b_2=\frac{S_y}{S_2}\left[\frac{r_{2y}-r_{1y}\cdot r_{12}}{1-r_{12}^2}\right]=\frac{2.41}{24.576}\times\left[\frac{-0.932-0.934\times(-0.915)}{1-0.934^2}\right]=\underline{\underline{-0.059}}$$

$$a=\bar{Y}-b_1\bar{X}_1-b_2\bar{X}_2=15.3-(-0.217)\times 13.5-(-0.059)\times 69=\underline{\underline{22.3}}$$

$$\boxed{\hat{Y}_i=22.3-0.217X_{1i}-0.059X_{2i}}$$

$$R^2_{y\cdot 12}=\frac{r_{1y}^2+r_{2y}^2-2r_{12}r_{1y}r_{2y}}{1-r_{12}^2}=\frac{(-0.915)^2+(-0.932)^2-2\times 0.934\times(-0.915)\times(-0.932)}{1-0.934^2}=$$

$$=\underline{\underline{0.884}}$$

ערך המתאם המרובה $R_{y\cdot 12}=\sqrt{0.884}=\underline{\underline{0.94}}$

$\text{שונות מוסברת}=R^2_{y\cdot 12}=0.884 \Rightarrow \%\ \text{שונות מוסברת}=\underline{\underline{88.4\%}}$

3. מבחן מובהקות למקדמי הרגרסיה:

ערך סטטיסטי $t_1=\frac{b_1}{Sb_1}=\frac{-0.217}{0.224}=\underline{\underline{-0.97}}$

$$Sb_1=\sqrt{\frac{SSres/n-k-1}{SSx_1(1-r_{12}^2)}}=\sqrt{\frac{(1-R^2)S_y^2/n-k-1}{S_1^2(1-r_{12}^2)}}=\sqrt{\frac{(1-0.884)\times 2.41^2/7}{3.879^2\times(1-0.934^2)}}=0.224$$

ערך סטטיסטי $t_2=\frac{b_2}{Sb_2}=\frac{-0.059}{0.035}=\underline{\underline{-1.69}}$

$$Sb_2=\sqrt{\frac{1-R^2\cdot S_y^2/n-k-1}{S_2^2\cdot(1-r_{12}^2)}}=\sqrt{\frac{(1-0.884)\times 2.41^2}{24.576^2\times(1-0.934^2)}}=\underline{\underline{0.035}}$$

ערך קריטי מטבלה $t_{(8)}^{\frac{\alpha}{2}=0.025}=\underline{\underline{2.306}}$

התוצאות לא מובהקות סטטיסטית ב- $\alpha=0.05$

<u>מסקנה</u>: התרומה הבודדת של X_1 ושל X_2 אינה מובהקת כלומר אף משתנה לא מוסיף שונות מוסברת מובהקת להסבר הקריטריון מעבר למשתנה האחר!!!

ה. <u>מבחן מובהקות ל-שונות R^2 ניבויים</u>:

ערך סטטיסטי $F=\frac{R^2/k}{(1-R^2)/n-k-1}=\frac{0.884/2}{1-0.884/8}=\underline{\underline{30.48}}$

ערך קריטי מטבלה $F_{(2,8)}^{\alpha=0.05}=\underline{\underline{4.46}}$

$4.46<30.48 \Leftarrow$ התוצאה מובהקת ב-$\alpha=0.05$

<u>מסקנה</u>: אכן שני המנבאים X_1 ו-X_2 ביחד יכולים לנבא באופן מובהק את <u>הקריטריון</u>!!!

ו. <u>התוספת בהסבר של X_1 על פני X_2</u>:

$$\underbrace{R^2_{y\cdot 12}}_{\text{שונות מוסברת של y באמצעות 1 ו-2}}-\underbrace{r^2_{y2}}_{\text{שונות מוסברת רק באמצעות 2}}=0.884-(0.932^2)=\underbrace{\underline{\underline{0.015}}}_{\text{תוספת שונות של } X_1 \text{ מעבר ל-} X_2 \text{ בהסבר של y !!!}}$$