מבוא לאקונומטריקה ב'

החוג לכלכלה

תוכן עניינים

משתני דמי

הנושא של משתני דמי מטפל בהכנסת משתנים ב"ת איכותיים למודל הרגרסיה.

עד כה כל המשתנים הב"ת שהכנסנו למודל היו כמותיים, כלומר קיבלו ערכים מספריים.

למשל, נניח שאנו סבורים שמס' שנות הלימוד של אדם משפיעות על שכרו:

W_t=**השכר (המשתנה התלוי)**

S_t = **שנות לימוד (המשתנה הב"ת)**

משוואת הרגרסיה: $W_t = \alpha + \beta \cdot S_t$

במקרה זה המשתנה המסביר (כמו גם המוסבר) הוא כמותי.

נניח שאנו סבורים שגם משתנה המגדר משפיע על השכר. משתנה זה איננו כמותי כמו שנות לימוד אלא איכותי שכן הוא לא מקבל ערכים מספריים אלא ערכים קטגוריאליים כ"גבר" או "אישה".

נשאלת השאלה כיצד נכניס אותו לתוך משוואת הרגרסיה?

נגדיר משתנה D שיקבל את הערך 0 אם מדובר ב"אישה" ואת הערך 1 אם מדובר ב"גבר".

משתנה כזה נקרא משתנה דמי (dummy variable).

לעומת משתנה רגיל ש"פועל" תמיד, משתנה זה "יפעל" רק אם מדובר בגבר.

ניתן להכניס את משתנה הדמי למודל בשלושה אופנים שונים:

(1) משתנה דמי לחותך- המין משפיע על השכר ההתחלתי בלבד

(2) משתנה דמי לשיפוע- המין משפיע על התוספת לשכר בגין שנות הלימוד

(3) משתנה דמי לכל הפונקציה- המין משפיע גם על החותך וגם על השיפוע

(1) משתנה דמי לחותך

המין משפיע על השכר ההתחלתי בלבד.

המודל: $W_t = \alpha_0 + \alpha_1 D + \beta \cdot S_t + u_t$

החותך מייצג כאן את השכר ההתחלתי.

שכר ההתחלתי של אישה: α_0

שכר התחלתי של גבר: $\alpha_0 + \alpha_1$

הבדל בשכר בין נשים וגברים: α_1 **(הפרש בין החותכים)**

בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש החותכים: $H0: \alpha_1 = 0$

**** השיפוע מייצג את התוספת בשכר כפונקציה של מס' שנות הלימוד והוא זהה עבור נשים וגברים.**

? על בסיס מדגם של 50 איש העובדים בחברה מסוימת התקבלו התוצאות הבאות:

$$W_t = 5500 + 1043 \cdot D + 119 \cdot S_t$$

(S.E) (134) (56) (24)

המספרים בסוגריים הם טעויות התקן של מבחני המובהקות לפרמטרים.

א. **מהו השכר ההתחלתי של גבר בעל 12 שנות לימוד?**
ב. **מה ההבדל בשכר ההתחלתי בין גברים לנשים?**
ג. **האם הבדל זה מובהק באוכלוסיה?**
ד. **בדקו את הטענה כי השכר ההתחלתי של גברים גבוה ביותר מ- 500 ₪ מזה של נשים.**
ה. **בדקו את הטענה שהשכר ההתחלתי של נשים נמוך ב-600 ₪ מזה של גברים.**

<u>פונקצית רגרסיה המכילה משתנים איכותיים בלבד</u>

המגדר הוא המשתנה היחיד במשוואה:

$$W_t = \alpha_0 + \alpha_1 D + u_t$$

החותך מייצג כאן את השכר ה<u>ממוצע</u> עבור כל קטגוריה.

שכר הממוצע של אישה: α_0

שכר הממוצע של גבר: $\alpha_0 + \alpha_1$

הבדל בשכר הממוצע בין נשים וגברים: α_1 **(הפרש בין החותכים)**

בדיקת השערות על משתנה הדמי: מבחן t: $H0: \alpha_1 = 0$ **(מבחן זהה למבחן t להבדל בין ממוצעים).**

? על אותו המדגם של 50 איש העובדים בחברה מסוימת ביקש החוקר לבדוק האם יש הבדל בשכר הממוצע בין גברים לנשים. תוצאות האמידה:

$$W_t = 5200 + 1120 \cdot D$$

נתון: $S_{\hat{\alpha}_1} = 63$

בדקו האם קיים הבדל מובהק בשכר <u>הממוצע</u> בין נשים וגברים?

(2) משתנה דמי לשיפוע

המגדר משפיע על התוספת לשכר בגין שנות הלימוד.

$$W_t = \alpha + \beta_0 S_t + \beta_1 D S_t + u_t$$

השיפוע מייצג כאן את התוספת לשכר בגין שנות לימוד.

אצל אישה: התוספת לשכר בגין שנות לימוד- β_0

אצל גבר: התוספת לשכר בגין שנות לימוד- $\beta_0 + \beta_1$

הבדל בין גברים לנשים בתוספת לשכר בגין שנות הלימוד: β_1 **(הפרש השיפועים)**

בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש השיפועים: $H0:\beta_1 = 0$

**** החותך, המייצג את השכר ההתחלתי, יהיה זהה עבור גברים ונשים.**

? על בסיס אותו מדגם, ביקש החוקר לדעת האם קיים הבדל מובהק בין גברים לנשים בתוספת לשכר בגין שנות הלימוד. תוצאות האמידה נתונות להלן:

$$W_t = 5000 + 110 \cdot S_t + 120 \cdot D \cdot S_t + u_t$$

(68) (23) (25)

בדוק את ההשערה.

(3) משתנה דמי לכל הפונקציה

המין משפיע גם על החותך וגם על השיפוע. הווה אומר, גם על השכר ההתחלתי וגם על התוספת לשכר ההתחלתי בגין שנות הלימוד.

המודל: $W_t = \alpha_0 + \alpha_1 D + \beta_0 S_t + \beta_1 D S_t + u_t$

השכר ההתחלתי של אישה: α_0

השכר ההתחלתי של גבר: $\alpha_0 + \alpha_1$

הבדל בשכר ההתחלתי בין המינים: α_1 **(הבדל בחותכים)**

אצל אישה-התוספת לשכר בגין שנות הלימוד: β_0

אצל גבר-התוספת לשכר בגין שנות הלימוד: $\beta_0 + \beta_1$

הבדל בין המינים בתוספת לשכר בגין שנות הלימוד: β_1 **(הבדל בשיפועים)**

<u>בדיקת השערות למשתני הדמי:</u>

באמצעות מבחן WALD יש לבדוק: $H0:\alpha_1=\beta_1=0$

$H1$: לפחות אחד הפרמטרים שונה מ-0

אם דוחים את השערת האפס, יש לבצע מבחני t עבור כל אחד מהפרמטרים בנפרד:

$HO:\alpha_1=0$ **ו-** $H0:\beta_1=0$

<u>מבחן CHOW</u>

דרך נוספת לבדיקת ההבדל בין הקטגוריות, <u>בלא יצירת משתני דמי</u>:

חלוקת המדגם לפי הקטגוריות של המשתנה האיכותי. מדגם של גברים (T_m) ושל נשים (T_f).

עבור כל קבוצה לאמוד משוואות רגרסיה לניבוי שכר על ידי שנות לימוד .

נשים: $W_t=\alpha_f+\beta_f X_t+u_t$

גברים: $W_t=\alpha_m+\beta_m X_t+u_t$

השערות: $H_0:\alpha_f=\alpha_m;\beta_f=\beta_m$

לבדיקת ההשערה נשתמש במבחן CHOW (הזהה למבחן WALD שהשתמשנו בו מקודם):

המודל המוגבל (R) לא לוקח בחשבון את השפעת המגדר ולכן לא יכלול את המדגם המאוחד כי אין צורך בשתי רגרסיות נפרדות:

$W_t=\alpha+\beta X_t+u_t$

המודל הלא מוגבל (U) כולל את שני חלקי המדגם:

$$ESS_U=ESS_f+ESS_m$$
$$DF_U=DF_f+DF_m$$

$$CHOW_{stat}=\frac{\dfrac{ESS_R-(ESS_f+ESS_m)}{DF_R-(DF_f+DF_m)}}{\dfrac{ESS_f+ESS_m}{DF_f+DF_m}}=WALD_{stat}$$

למרות התוצאות הזהות בשתי הדרכים, שיטת משתני הדמי עדיפה:

1. **אם דחינו את HO במבחן CHOW נתקשה לברר את מקור ההבדל שנמצא.**
2. **בהרצת שתי רגרסיות נפרדות אנו בודקים הבדל בכל הפונקציה ואילו שיטת משתני הדמי מאפשרת לבדוק הבדל רק בחותך או רק בשיפוע.**

? חוקר רצה לבדוק את הטענה שסוג הכביש משפיע על מס' תאונות הדרכים בקטעי כביש בינעירוניים, בהינתן נפח התנועה.

החוקר בדק האם <u>הפונקציה</u> של מס' התאונות בהינתן נפח התנועה, שונה בין כבישים מהירים לבין כבישים שאינם מהירים. לשם כך אמד החוקר את ארבע המשוואות הבאות:

(1) $NUM_t=\gamma_1+\delta_1\cdot AVGD_t+\varepsilon_{1t}$ **כבישים מהירים בלבד**

(2) $NUM_t=\gamma_2+\delta_2\cdot AVGD_t+\varepsilon_{2t}$ **כבישים לא - מהירים בלבד**

(3) $NUM_t=\gamma_3+\delta_3\cdot AVGD_t+\varepsilon_{3t}$ **שני סוגי הכביש (כל המדגם)**

(4) $NUM_t=\alpha+\beta_1\cdot TYPE_t+\beta_2\cdot AVGD_t+\beta_3\cdot(AVGD\cdot TYPE)_t+U_t$

כאשר : NUM_t **= מס' תאונות הדרכים הקטלניות בקטע כביש t בשנה**

$AVGD_t$ **= נפח התנועה בקטע כביש t ליום באלפים**

$TYPE_t$ **= משתנה דמי המקבל את הערך 1 כאשר הכביש מהיר , ו- 0 כאשר הכביש לא מהיר.**

תוצאות אמידת המשוואות מופיעות בהמשך השאלה.

1) בדקו את טענת החוקר בשתי דרכים שונות. ציינו איזה מן המשוואות רלוונטיות עבור כל דרך.

2) חשבו את הערכים המספריים עבור אומדני משוואה (4).

3) מהו האומדן הנקודתי למס' התאונות בכביש מהיר כאשר נפח התנועה עומד על ארבעת מכוניות ליום בקטע הכביש האמור?

הועלתה הטענה כי המקדם להשפעה של נפח התנועה בדרכים מהירות הינו כפול מזה שבדרכים לא -מהירות.

4) מהי השערת האפס לבדיקת הטענה (במונחי משוואה (4))?

5) מהי הרגרסיה "תחת H_0" למבחן WALD ?

משוואה (1) - כבישים מהירים בלבד

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	344
Number of Observations Used	344

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4700.81174	4700.81174	89.12	<.0001
Error	342	18039	52.74684		
Corrected Total	343	22740			

Root MSE	7.26270	R-Square	0.2067
Dependent Mean	5.10465	Adj R-Sq	0.2044
Coeff Var	142.27617		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.55289	0.54303	2.86	0.0045
avgd	1	0.02098	0.00222	9.44	<.0001

משוואה (2) - כבישים לא מהירים בלבד

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	410
Number of Observations Used	410

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	971.99073	971.99073	145.83	<.0001
Error	408	2719.34830	6.66507		
Corrected Total	409	3691.33902			

Root MSE	2.58168	R-Square	0.2633
Dependent Mean	1.38780	Adj R-Sq	0.2615
Coeff Var	186.02612		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.14978	0.16360	0.92	0.3605
avgd	1	0.02877	0.00238	12.08	<.0001

משוואה (3) - שני סוגי הכביש (כל המדגם)

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	754
Number of Observations Used	754

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	8052.00804	8052.00804	288.84	<.0001
Error	752	20964	27.87730		
Corrected Total	753	29016			

Root MSE	5.27990	R-Square	0.2775
Dependent Mean	3.08355	Adj R-Sq	0.2765
Coeff Var	171.22758		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.73903	0.23665	3.12	0.0019
avgd	1	0.02330	0.00137	17.00	<.0001

משוואה (4)

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	754
Number of Observations Used	754

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	8256.966	2752.322	99.44	<.0001
Error	750	20759	27.678		
Corrected Total	753	29016			

Root MSE	5.26102	R-Square	0.2846
Dependent Mean	3.08355	Adj R-Sq	0.2817
Coeff Var	170.61553		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.14978	0.33340	0.45	0.6534
type	1				0.0067
avgd	1				<.0001
avgdtype	1				0.1283

סיכום ביניים:

	משתנה דמי לחותך	**משתנה דמי לשיפוע**	**משתנה דמי לכל הפונקציה**
המודל	$Y_t = \alpha_0 + \alpha_1 D + \beta \cdot X_t + u_t$	$Y_t = \alpha + \beta_0 X_t + \beta_1 DX_t + u_t$	$Y_t = \alpha_0 + \alpha_1 D + \beta_0 X_t + \beta_1 DX_t + u_t$
ההשערה במילים	**קיים הבדל בין הקטגוריות ב-Y ההתחלתי (בחותך).**	**קיים הבדל בין הקטגוריות בתוספת ל- Y בגין X (בשיפוע).**	**קיים הבדל בין הקטגוריות במשוואת הרגרסיה כולה (בחותך ובשיפוע).**
בדיקת ההשערה	**מבחן t להפרש החותכים:** $H0: \alpha_1 = 0$	**מבחן t להפרש השיפועים:** $H0: \beta_1 = 0$	**מבחן WALD להפרש בין הפונקציות (החותכים והשיפועים):** $H0: \alpha_1 = \beta_1 = 0$ ****ניתן לבדוק את ההשערה בדבר הבדל בין הפונקציות גם במבחן CHOW.** **אם דוחים את HO יש לברר את מקור ההבדל באמצעות מבחני t (אפשרי רק ב-WALD):** $H0: \alpha_1 = 0$ $H0: \beta_1 = 0$

משתני דמי אם המשתנה האיכותי יכול לקבל יותר משני ערכים

כאשר המשתנה האיכותי כלל שני ערכים בלבד (למשל, מגדר: גבר, אישה) הסתפקנו במשתנה דמי אחד.

במקרים רבים המשתנה האיכותי כולל יותר משני ערכים/קטגוריות. במקרה כזה נגדיר מס' משתני דמי כמספר הקטגוריות פחות אחד.

למשל, את המשתנה האיכותי של עונות השנה הכולל 4 ערכים: אביב, קיץ, סתיו, חורף נייצג באמצעות 3 משתני דמי:

D_1 **יקבל את הערך 1 אם מדובר באביב ו-0 אחרת.**

D_2 **יקבל את הערך 1 אם מדובר בקיץ ו-0 אחרת.**

D_3 **יקבל את הערך 1 אם מדובר בסתיו ו-0 אחרת.**

אם מדובר בחורף אז כל משתני הדמי יקבלו את הערך 0 ולכן החורף היא קבוצת הייחוס.

נניח שאנו רוצים לבדוק עונתיות במחירי הירקות:

$V_t=$ **מדד מחירי הירקות**

$p_t=$**מדד המחירים לצרכן**

(1) <u>משתני דמי לחותך</u>

<u>הטענה:</u> יש הבדל בין עונות השנה במחיר ההתחלתי של הירקות

<u>המודל:</u> $V_t=\alpha_0+\alpha_1 D_{1t}+\alpha_2 D_{2t}+\alpha_3 D_{3t}+\beta\cdot P_t+u_t$

כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב- β.

למחיר זה יתווסף α_0 **בחורף,** $\alpha_0+\alpha_1$ **באביב,** $\alpha_0+\alpha_2$ **בקיץ ו-** $\alpha_0+\alpha_3$ **בסתיו.**

ניתן לראות כי: α_0**: החותך בקטגוריה שהושמטה**

$\alpha_0+\alpha_i$**: החותך בקטגוריה i.**

<u>בדיקת השערות:</u>

השערות:

$$H0:\alpha_1=\alpha_2=\alpha_3=0$$
$$H1: OTHERWISE$$

המבחן הסטטיסטי : מבחן WALD:

(U) $V_t=\alpha_0+\alpha_1 D_{1t}+\alpha_2 D_{2t}+\alpha_3 D_{3t}+\beta\cdot P_t+u_t$

(R) $V_t=\alpha\ \ +\beta\cdot P_t+u_t$

****שימו לב שהחותך במשוואה המוגבלת איננו α_0 שכן המשתנה המסביר של עונות השנה ירד.**

אם נדחה את HO במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין החותכים על ידי מבחני t:

1. **האם יש הבדל במחיר ההתחלתי של הירקות בין האביב לחורף:** $H0:\alpha_1=0$
2. **האם יש הבדל במחיר ההתחלתי של הירקות בין הקיץ לחורף:** $H0:\alpha_2=0$
3. **האם יש הבדל במחיר ההתחלתי של הירקות בין הסתיו לחורף:** $H0:\alpha_3=0$

? א. הועלתה הטענה כי יש הבדל במחיר ההתחלתי בין האביב לקיץ.

1. **מהי השערת האפס לבדיקת הטענה?**
2. **פרטו שני מבחנים סטטיסטיים בעזרתם ניתן לבדוק את הטענה.**

ב. הועלתה הטענה כי יש רק שתי עונות המשפיעות על מחיר הירקות ההתחלתי: קיץ+אביב, חורף+סתיו.

1. **מהי השערת האפס לבדיקת הטענה?**

2. מהו המבחן הסטטיסטי המתאים? פרטו.

(2) <u>משתני דמי לשיפוע</u>

<u>הטענה:</u> יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן

<u>המודל:</u> $V_t = \alpha + \beta_0 P_t + \beta_1(D_{1i}P_t) + \beta_2(D_{2i}P_t) + \beta_3(D_{3i}P_t) + u_t$

המחיר ההתחלתי של הירקות שווה בין עונות השנה (α) אולם כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב:

β_0 **בחורף,** $\beta_0 + \beta_1$ **באביב,** $\beta_0 + \beta_2$ **בקיץ ו-** $\beta_0 + \beta_3$ **בסתיו.**

ניתן לראות כי: β_0 **: השיפוע בקטגוריה שהושמטה**

$\beta_0 + \beta_i$ **: השיפוע בקטגוריה i.**

<u>בדיקת השערות:</u>

השערות:

$$H0: \beta_1 = \beta_2 = \beta_3 = 0$$
$$H1: OTHERWISE$$

המבחן הסטטיסטי : מבחן WALD:

(U) $V_t = \alpha + \beta_0 P_t + \beta_1(D_{1i}P_t) + \beta_2(D_{2i}P_t) + \beta_3(D_{3i}P_t) + u_t$

(R) $V_t = \alpha + \beta \cdot P_t + u_t$

****שימו לב שהשיפוע במשוואה המוגבלת איננו** β_0 **שכן המשתנה המסביר של עונות השנה ירד.**

אם נדחה את HO במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין השיפועים על ידי מבחני t.

(3) <u>משתני דמי לכל הפונקציה</u>

<u>הטענה</u>:יש הבדל בין עונות השנה בפונקצית הרגרסיה לניבוי מחיר הירקות באמצעות המחיר לצרכן.

<u>המודל</u>: $V_t = \alpha_0 + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta_0 P_t + \beta_1 (D_{1i} P_t) + \beta_2 (D_{2i} P_t) + \beta_3 (D_{3i} P_t) + u_t$

<u>בדיקת השערות:</u>

השערות:

$$HO: \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$$

המבחן הסטטיסטי: מבחן WALD:

(U) $V_t = \alpha_0 + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta_0 P_t + \beta_1 (D_{1i} P_t) + \beta_2 (D_{2i} P_t) + \beta_3 (D_{3i} P_t) + u_t$

(R) $V_t = \alpha + \beta \cdot P_t + u_t$

אם דוחים את HO, יש לבדוק במבחני WALD האם ההבדל הוא בין החותכים או בין השיפועים:

$$HO: \beta_1 = \beta_2 = \beta_3 = 0 \quad HO: \alpha_1 = \alpha_2 = \alpha_3 = 0$$

באם דוחים את H0 יש להמשיך לבדוק באמצעות מבחני t:

$$HO: \beta_j = 0 \qquad H0: \alpha_j = 0$$

משתני דמי עבור שני משתנים איכותיים

נתבונן בדוגמא שבה יש שני משתנים איכותיים המשפיעים על פונקצית השכר-

מגדר (אישה, גבר) וגזע (לבן, שחור).

נגדיר משתנה דמי G שיקבל 1 אם מדובר בגבר ו-0 אחרת (אישה).

נגדיר משתנה דמי R שיקבל 1 אם מדובר בלבן ו-0 אחרת (שחור).

נבדוק כיצד מגדר וגזע משפיעים על השכר ההתחלתי (החותך), כאשר השכר תלוי גם בשנות לימוד (S_t).

(1) <u>הבדל בחותך ללא אינטראקציה</u>

<u>המודל:</u> $W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \beta \cdot S_t + u_t$

במודל זה- אין השפעה משולבת של מגדר וגזע על השכר ההתחלתי.

במילים אחרות, ההבדל בשכר ההתחלתי בין גברים ונשים לא תלוי בגזע (זהה עבור שחורים ועבור לבנים) ולהיפך-ההבדל בשכר ההתחלתי בין לבנים לשחורים לא תלוי במגדר (זהה עבור נשים וגברים).

ניתן לבדוק השערות על כל אחד מהמשתנים הב"ת האיכותיים בנפרד:

1. **ההבדל בשכר ההתחלתי בין גברים לנשים :** $HO: \alpha_1 = 0$
2. **ההבדל בשכר ההתחלתי בין שחורים ללבנים:** $H0: \alpha_2 = 0$

(2) <u>הבדל בחותך עם אינטראקציה</u>

<u>המודל:</u> $W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \alpha_3 G \cdot R + \beta \cdot S_t + u_t$

במודל זה הטענה היא כי קיימת, בנוסף להשפעה של מגדר וגזע בנפרד על השכר, גם השפעה משולבת (אינטראקציה) של מגדר וגזע על השכר ההתחלתי.

במילים אחרות, ההבדל בשכר ההתחלתי בין גברים ונשים תלוי בגזע (שונה אם מדובר בשחורים או בלבנים) ולהיפך.

במודל זה, לעומת הקודם, נוספת ההשערה לבדיקת השפעת האינטראקציה בין מגדר לגזע על השכר ההתחלתי:

3. $HO: \alpha_3 = 0$

<u>דרך נוספת ליצירת מודל עם אינטראקציה:</u>

הגדרת משתני דמי המייצגים שילוב בין המשתנים האיכותיים גזע ומגדר באופן הבא:

D_1 **יקבל 1 אם מדובר בגבר לבן ו-0 אחרת**

D_2 **יקבל 1 אם מדובר בגבר שחור ו-0 אחרת**

D_3 **יקבל 1 אם מדובר באשה לבנה ו-0 אחרת**

הנשים השחורות מהוות כאן את קבוצת הייחוס.

<u>המודל:</u> $W_t = \gamma_0 + \gamma_1 D_1 + \gamma_2 D_2 + \gamma_3 D_3 + \delta \cdot S_t + u_t$

נעזר בטבלה בכדי לנסח את ההשערות לבדיקת האינטראקציה:

	גבר	**אישה**	**הפרש**
לבן	$\gamma_0 + \gamma_1$	$\gamma_0 + \gamma_3$	$\gamma_1 - \gamma_3$
שחור	$\gamma_0 + \gamma_2$	γ_0	γ_2
הפרש	$\gamma_1 - \gamma_2$	γ_3	

ההשערות לבדיקת קיום האינטראקציה:

$HO: \gamma_1 - \gamma_3 = \gamma_2$ **או** $HO: \gamma_1 - \gamma_2 = \gamma_3$

התוצאות שיתקבלו כאן יהיו כמובן זהות לחלוטין לתוצאות שהתקבלו בדרך הקודמת:

$$WALD = t^2$$
$$PF = Pt$$

שאלה מס' 1

חוקר בדק השפעות של השכלה, גזע (שחור, לבן) וניסיון (EXP) על לוג השכר (ln(Y)) במדגם בן 306 תצפיות:

$$\ln(Y)_t = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \beta_1 EXP_t + \beta_2 EXP_t^2 + u_t$$

ln(Y)-לוג השכר

EXP- שנות ניסיון

D_1 **מקבל את הערך 1 עבור שחורים בעלי השכלה גבוהה (ו-0 אחרת)**

D_2 **מקבל את הערך 1 עבור שחורים בעלי השכלה נמוכה (ו-0 אחרת)**

D_3 **מקבל את הערך 1 עבור לבנים בעלי השכלה גבוהה (ו-0 אחרת)**

תוצאות אמידת משוואת הרגרסיה מוצגות בפלט להלן:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	--------	------	------	------
Error	300	140	------		
Corrected Total	305	210			

Root MSE	-------	R-Square	-------
Dependent Mean	-------	Adj R-Sq	-------
Coeff Var	-------		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-------	--------	60.84	0.00
D1	1	-------	--------	-3.20	0.00
D2	1	-------	--------	-5.56	0.00
D3	1	-------	--------	7.23	0.00
EXP	1	-------	--------	8.11	0.00
EXP^2	1	--------	--------	-7.45	0.00

בטרם ניגשים לפיתרון השאלה יש להכין את טבלת העזר הבאה:

	השכלה נמוכה	השכלה גבוהה	הפרש
לבנים	α_0	$\alpha_0+\alpha_3$	α_3
שחורים	$\alpha_0+\alpha_2$	$\alpha_0+\alpha_1$	$\alpha_1-\alpha_2$
הפרש	α_2	$\alpha_1-\alpha_3$	

(א) לפי המשוואה הניסיון זהה עבור שחורים ולבנים: נכון/לא נכון/ לא ניתן לדעת

(ב) בדוק את הטענה כי בקרב אנשים בעלי השכלה נמוכה אין השפעה לגזע.

(ג) בדוק את הטענה כי אין השפעות השכלה בקרב לבנים.

(ד) הי השערת האפס לבדיקת הטענה כי אין אינטראקציה בין גזע להשכלה?

(ה) לבדיקת ההשערה של הסעיף הקודם בוצע מבחן WALD.

הרגרסיה המוגבלת תחת השערת האפס הינה:

$$Z_0 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \gamma_4 Z_4 + \nu$$ **מהם ה-Z-ים?**

(ו) בדוק את ההשערה אם ידוע שבמודל המוגבל $R^2 = 0.33$

(ז) החוקר החליט לאמוד במקום את המשוואה המקורית את המשוואה:

$$\ln(Y)_t = \lambda_0 + \lambda_1 S + \lambda_2 E + \lambda_3 (S \cdot E) + \delta_1 EXP + \delta_2 EXP^2 + \omega_t$$

כאשר: S מקבל את הערך 1 עבור שחורים ו-0 אחרת (לבנים)

E מקבל את הערך 1 עבור השכלה גבוהה ו-0 אחרת (השכלה נמוכה).

מה הקשר בין המקדמים של שני המודלים?

(ח) אם יאמוד החוקר את המשוואה:

$$\ln(Y)_t = \lambda_0 + \lambda_1 S + \lambda_2 E + \delta_1 EXP + \delta_2 EXP^2 + \omega_t$$

האם תהיה טעות ספציפיקציה של השמטת משתנה רלוונטי (היעזר בסעיפים ד',ו' ו-ז').

<u>שאלה מס' 2</u>

חוקרת בדקה השפעות השכלה, מגדר וניסיון על הכנסה מעבודה לפי המשוואה הבאה:

$$\begin{aligned}\ln(MWAGE) = {} & \alpha_0 + \alpha_1 \cdot S + \alpha_2 \cdot E + \alpha_3 \cdot (S \cdot E) + \\ & \beta_0 \cdot EXP + \beta_1 \cdot (EXP \cdot S) + \beta_2 \cdot (EXP \cdot E) + \beta_3 \cdot (EXP \cdot S \cdot E) + U\end{aligned}$$

כאשר: S משתנה דמי: 1= עבור נשים, 0=גברים

E משתנה דמי: 1= עבור השכלה גבוהה (scl >12) , 0=השכלה נמוכה

בטרם ניגשים לפיתרון השאלה יש להכין את טבלת העזר הבאה:

	השכלה גבוהה (E=1)	**השכלה נמוכה (E=0)**	**הפרש**
נשים (S=1)	$\alpha_0+\alpha_1+\alpha_2+\alpha_3+(\beta_0+\beta_1+\beta_2+\beta_3)EXP_t$	$\alpha_0+\alpha_1+(\beta_0+\beta_1)EXP_t$	**חותך:** $\alpha_2+\alpha_3$ **שיפוע:** $\beta_2+\beta_3$
גברים (S=0)	$\alpha_0+\alpha_2+(\beta_0+\beta_2)EXP_t$	$\alpha_0+\beta_0 EXP_t$	**חותך:** α_2 **שיפוע:** β_2
הפרש	**חותך:** $\alpha_1+\alpha_3$ **שיפוע:** $\beta_1+\beta_3$	**חותך:** α_1 **שיפוע:** β_1	

א. רשמו את הפונקציה לחישוב:

1. **תחזית לוג השכר עבור גבר בעל השכלה נמוכה ו- 10 שנות ניסיון.**
2. **תחזית לוג השכר ההתחלתי עבור נשים משכילות.**
3. **לאחר כמה שנות ניסיון ישתווה השכר של נשים משכילות לזה של גברים משכילים?**

ב. רשמו את השערות האפס המתאימות לבדיקת הטענות הבאות:

1. **אין השפעה של מגדר והשכלה על השכר.**
2. **השפעת ההשכלה אינה תלויה במגדר.**
3. **אין השפעות השכלה אצל גברים.**
4. **אין הבדל בשיעורי התשואה לניסיון, בקרב הנשים.**

מולטיקוליניאריות

מולטיקוליניאריות היא תופעה סטטיסטית בעייתית המתייחסת למתאם בין המשתנים המסבירים במודל.

נבחין בין מולטיקוליניאריות מלאה לחלקית.

1) מולטיקוליניאריות מלאה

מתאם מלא בין המשתנים המסבירים במודל.

הדבר קורה כאשר משתנה מסביר אחד הוא קומבינציה ליניארית מלאה של המשתנה המסביר השני:

$x_1 = a + bx_2$ (x_1 **הוא קומבינציה ליניארית מלאה של** x_2) **מכאן ש:** $r_{12} = 1$.

****שימו לב כי מדובר בטרנספורמציה ליניארית ולא בטרנספורמציה אחרת (למשל** $x_1 = x_2^2$**), אז בהכרח** $r_{12} \neq 1$.

במצב של מולטיקוליניאריות מלאה אין כל השפעה של המשתנה האחד מעבר לשני.

לדוגמא: נניח שאנו רוצים לאמוד את הביקוש לדירות בתל אביב כפונקציה של מחירן בשקלים (x_1 **) ובדולרים (** x_2 **):**

$$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

בהנחה ששער הדולר נותר קבוע, המחיר בדולרים מהווה קומבינציה ליניארית מלאה של המחיר בשקלים (המחיר בשקלים * שער הדולר). במקרה כזה לא ניתן להפריד את ההשפעה של שני המשתנים הב"ת זה מזה ומדובר בעצם באותו המשתנה.

מדוע זה בעייתי?

כיוון שלא ניתן לאמוד את המודל שכן אר"פ אינם מוגדרים.

הסבר: בגזירת אר"פ, נוצר מצב של תלות ליניארית בין המשוואות הנורמאליות וחלקן יתבטלו. ניוותר עם יותר נעלמים ממשוואות ועם אינסוף פיתרונות, כך שלא נוכל להגדיר את האומדים.

בדוגמא שלנו:

$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$

גזירת $\hat{\alpha}$ **:** $\sum \hat{u}_t = 0$

גזירת $\hat{\beta}_1$ **:** $\sum \hat{u}_t x_1 = 0$

גזירת $\hat{\beta}_2$ **:** $\sum \hat{u}_t x_2 = 0$

מכיוון ש: $x_1 = bx_2$ **(b = שער הדולר) אז גזירת** $\hat{\beta}_1$ **:** $\sum \hat{u}_t (bx_2) = 0$, $b\sum \hat{u}_t x_2 = 0$, $0 = 0$ **והמשוואה מתבטלת.**

<u>פיתרון</u>: הורדת אחד המשתנים ואמידת המשוואה מחדש בלעדיו.

2) מולטיקוליניאריות חלקית

כאשר יש מתאם גבוה מאוד בין משתנים מסבירים במודל (אך לא מושלם) עלולה להיווצר בעיה של מולטיקוליניאריות חלקית.

לדוגמא: נניח שאנו רוצים לאמוד את הציונים בתואר ראשון ע"י ציוני הפסיכומטרי (x_1 **) וציוני הבגרות (** x_2 **):**

$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$

מכיוון שיש מתאם גבוה בין ציוני הפסיכומטרי וציוני הבגרות לא נוכל לבודד באופן מלא את ההשפעה המדויקת של כל אחד מהם על ציוני ה-.B.A

כל אחד מהמשתנים הב"ת "יגזול" מן ההשפעה הייחודית שיש למשתנה הב"ת השני על המשתנה התלוי, כך שבסופו של דבר, למרות שהמודל עם שני המשתנים הב"ת יהיה מובהק, התרומה הייחודית של כל משתנה ב"ת לניבוי התלוי לא תהיה מובהקת.

<u>זיהוי מולטיקוליניאריות חלקית:</u>

1) **כאשר קיימת <u>סתירה</u> בין התוצאה <u>במבחן F</u> למובהקות המודל (המודל מובהק) לבין <u>מבחני t</u> למובהקות השיפועים (אף אחד מן השיפועים איננו מובהק).**

הסתירה נוצרת כתוצאה מהגדלת השונות של כל אחד מהשיפועים בשל המתאם הגבוה בין הב"ת, באופן שלא מאפשר לדחות את השערת האפס למובהקות השיפועים:

$$s_{\hat{\beta}_1}^2 = \frac{MSE}{SSX_1(1-r_{12})}$$

$$t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$$

2) **<u>רגישות לספציפיקציה</u>- הורדת משתנה ב"ת שאיננו מובהק תהפוך משתנים ב"ת אחרים במודל למובהקים. אם אין בעיה של מולטיקוליניאריות, הורדת משתנים ב"ת שאינם רלוונטיים מהמודל, לא אמורה להשפיע על מובהקותם של המשתנים הב"ת האחרים.**

3) **<u>סימנים הפוכים</u>-כאשר השיפועים של המשתנים הב"ת מקבלים סימנים הפוכים מכיוון ההשפעה שלהם על המשתנה התלוי. אם למשל, x_1 משפיע חיובית על Y ואילו x_2 משפיע שלילית על Y אבל הם יופיעו במשוואת הרגרסיה עם סימנים הפוכים ($\hat{\beta}_1$ שלילית ואילו $\hat{\beta}_2$ חיובית), יש לחשוד שקיימת בעיה של מולטי קולינאריות במודל (מתאם גבוה מאוד בין המשתנים הב"ת ש"מבלבל" את הסימנים שלהם).**

<u>השלכות של מולטיקוליניאריות חלקית:</u>

מולטיקוליניאריות חלקית איננה פוגעת בתכונות של אר"פ (הם נותרים ליניאריים, חסרי הטיה, יעילים ועקיבים) ולא באומד השונות של האומדים (שנותר חסר הטיה) כך שבדיקת השערות תוך שימוש באומדים הללו תהיה תקפה (זאת בניגוד למולטיקוליניאריות מלאה).

במובן הזה, בעיה של מולטיקוליניאריות חלקית דומה לבעיה של הוספת משתנה ב"ת שאיננו רלוונטי.

<u>פיתרונות למולטיקוליניאריות חלקית:</u>

1) **ברוב המקרים נשקול להוריד את אחד המשתנים.**

יחד עם זאת, כאשר המובהקות של המשתנים היא גבולית $1 < t_{\hat{\beta}} < 2$, יתכן ונותיר את שניהם בתוך המודל כיוון שבסך הכול יש עליה ב- $AdjR^2$ (לפי חוק חיטובסקי).

2) **ניתן לעיתים לאחד את שני המשתנים למשתנה אחד. למשל, בדוגמא של ניבוי ציוני ה-.B.A על ידי ציוני הפסיכומטרי והבגרות, יש לשקול להגדיר משתנה חדש שיהווה שקלול ציוני הפסיכומטרי והבגרות ולהשתמש בו לניבוי ה-.B.A**

<u>שלבי בדיקת ההשערות:</u>

1) **מבצעים מבחן F לבדיקת מובהקות המודל.**

2) **במידה והמודל מובהק, מבצעים מבחן t למובהקות כל אחד מהשיפועים.**

3) **ביצוע מבחן WALD לבדיקת כל השיפועים שלא יצאו מובהקים:**

א. אם מקבלים את H0: אין סתירה עם מבחני t- אין בעיה של מולטיקוליניאריות חלקית, נוריד את קבוצת המשתנים הלא רלוונטיים מהמודל.

ב. אם דוחים את H0: יש סתירה עם מבחני t- קיימת בעיה של מולטיקוליניאריות חלקית, יש להוריד מן המודל כל פעם משתנה אחד ולבצע מבחן WALD בלעדיו, עד שמזהים את המשתנה/משתנים שיש להוריד מהמודל.

סיכום ותרגול של טעויות ספציפיקציה ומולטי קולינאריות

סיכום

<table>
<tr><th>הבעיה</th><th>הגדרה</th><th>זיהוי</th><th>השלכות</th><th>פיתרון</th></tr>
<tr>
<td>הוספת משתנה
לא רלוונטי</td>
<td>המודל האמיתי:
$Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$
המודל הנאמד (הטעותי):
$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$</td>
<td>קבלת H0 במבחן t
למובהקות β_2</td>
<td>ניתן לבצע בדיקת השערות
אר"פ ($\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$) חסרי הטיה
אומדי השונות ($S^2_{\hat{\alpha}}, S^2_{\hat{\beta}_1}, S^2_{\hat{\beta}_2}$) חסרי הטיה</td>
<td>*הורדת
המשתנה</td>
</tr>
<tr>
<td>השמטת משתנה
רלוונטי</td>
<td>המודל האמיתי:
$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$
המודל הנאמד (הטעותי):
$Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$</td>
<td>דחית H0 במבחן t
למובהקות β_2</td>
<td>לא ניתן לבצע בדיקת השערות
<table>
<tr><th>בהיעדר x_2 :</th><th>אומד ל- β_1</th><th>אומד ל- α</th><th>אומד לשונות הפרמטרים</th></tr>
<tr><td>$S_{12} = 0$</td><td>חסר הטיה</td><td>מוטה
אלא אם- $\bar{x}_2 = 0$</td><td>מוטה חיובית</td></tr>
<tr><td>$S_{12} \neq 0$</td><td>מוטה חיובית:
S_{12} ו- β_2 שווי סימן
מוטה שלילית:
S_{12} ו β_2 מנוגדי סימן</td><td>מוטה</td><td>מוטה חיובית</td></tr>
</table>
</td>
<td>הוספת
המשתנה</td>
</tr>
<tr>
<td>מולטיקוליניאריות
מלאה</td>
<td>מתאם מלא בין המשתנים המסבירים במודל
$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$
כאשר $r_{12} = \pm 1$</td>
<td>אם: $x_1 = a + bx_2$
אז: $r_{12} = 1$</td>
<td>לא ניתן לבצע בדיקת השערות
אר"פ ($\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$) בלתי מוגדרים.</td>
<td>הורדת
אחד
המשתנים</td>
</tr>
<tr>
<td>מולטיקוליניאריות
חלקית</td>
<td>מתאם חזק בין המשתנים המסבירים במודל
$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$
כאשר $0.7 < \lvert r_{12} \rvert < 1$</td>
<td>א.סתירה בין מבחן F ל- t
ב.רגישות לספציפיקציה
ג.סימנים הפוכים</td>
<td>ניתן לבצע בדיקת השערות
אין פגיעה בתכונות אר"פ ושונותם</td>
<td>**הורדת
אחד
המשתנים
או איחודם</td>
</tr>
</table>

* במידה והמובהקות גבולית ($1 < t_{\hat{\beta}_2} < 2$) נשקול להשאיר משתנה לא רלוונטי כי מעלה את $AdjR^2$ (חוק חיטובסקי).

** במידה ומובהקותם גבולית ($1 < t_{\hat{\beta}} < 2$) נשקול להשאיר את שניהם בשל העלייה ב- $AdjR^2$ (חוק חיטובסקי).

תרגול

<u>שאלה מס' 1</u>

להלן מודל של שכר W_t, כפונקציה של שנות לימוד S_t :

$$W_t = \alpha + \beta \cdot S_t + u_t \quad (1)$$

להלן מודל של שכר W_t, כפונקציה של שנות לימוד S_t ושל גיל A_t :

$$W_t = \alpha + \beta_1 \cdot S_t + \beta_2 \cdot A_t + v_t \quad (2)$$

כל האומדים חיוביים ומובהקים וקיים קשר שלילי בין גיל להשכלה.

א. $\hat{\beta}_1$ במשוואה (1) הוא:

1. **אומד חסר הטיה**
2. **אומד מוטה שלילית**
3. **אומד מוטה חיובית**
4. **אומד מוטה, אך לא ניתן לדעת את כיוון ההטיה.**

ב. ניתן להשתמש במבחן t לבדיקת מובהקות השיפוע במשוואה (1). נכון/לא נכון/ לא ניתן לדעת

ג. בנוסף למשתנים במשוואה השנייה, החליט החוקר להוסיף גם את משתנה הוותק, EXP_t. מכיוון שלא היו בידו נתונים על הוותק, החליט החוקר להעריכו עבור כל עובד על ידי הגיל של העובד פחות 24 שנים (מתוך ההנחה שהחיים המקצועיים מתחילים בגיל זה לערך).

להלן משוואה מס' 3:

$$W_t = \alpha + \beta_1 \cdot S_t + \beta_2 \cdot A_t + \beta_3 \cdot EXP_t + w_t \quad (3)$$

חווה דעתך על המשוואה השלישית.

<u>שאלה מס' 2</u>

נתונות ארבע משוואות הרגרסיה הבאות (כאשר הסטיות במודל האמיתי מקיימות את הנחות הרגרסיה הקלאסיות):

(1) $X_{2t} = \lambda + \delta \cdot X_{1t} + V_t$ **כאשר התקבל:** $\sum \hat{V}_t^2 = \sum (X_{2t} - \overline{X}_{2t})^2$

(2)
$$\underset{}{Y_t = \alpha + \underset{(10.3)}{\beta_1} \cdot X_{1t} + \underset{(19.8)}{\beta_2} \cdot X_{2t} + U_t}$$

(3)
$$Y_t = \alpha + \underset{(9.9)}{\beta_1} \cdot X_{1t} + \underset{(17.3)}{\beta_2} \cdot X_{2t} + \underset{(0.37)}{\beta_3} \cdot X_{3t} + W_t$$

(4)
$$Y_t = \alpha + \underset{(6.3)}{\beta_1} \cdot X_{1t} + \Sigma_t$$

(המספרים בסוגריים הם ערכי t של אומדני המקדמים).

לגבי הטענות הבאות, קבעו לגבי כל טענה אם היא נכונה או לא, והסבירו:

א. **האומד של** β_1 **במשוואה (2) הינו חסר הטיה, אך אומד השונות של** β_1 **מוטה.**

ב. **האומד של** β_1 **במשוואה (3) הינו חסר הטיה, אך אומד השונות של** β_1 **מוטה.**

ג. **האומד של** β_1 **במשוואה (4) הינו חסר הטיה, אך אומד השונות של** β_1 **מוטה.**

ד. **האומדן** $\hat{\beta}_1$ **במשוואה (4) זהה ל** $\hat{\beta}_1$ **במשוואה (2).**

ה. **השונות התיאורטית של האומדן** $\hat{\beta}_1$ **במשוואה (4) זהה לשונות התיאורטית של** $\hat{\beta}_1$ **במשוואה (2), אך אומדני השונויות שונים.**

ו. **האומד ל -** α **במשוואה (4) הינו חסר הטיה.**

ז. **האומד ל -** α **במשוואה (3) הינו חסר הטיה.**

ח. R^2 של משוואה (2) גדול מ - R^2 של משוואה (3).

ט. $\overline{R}^2$ של משוואה (2) גדול מ - $\overline{R}^2$ של משוואה (3).

<u>שאלה מס' 3</u>

נתון המודל : $Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + U_t$

חוו דעתכם על הטענות הבאות (כל סעיף עומד בפני עצמו):

א. בהנחה כי מתקיים: $Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + U_t$ $\quad R^2 = 0.92$

(0.3) (0.5)

הערכים בסוגריים הם ערכי t למובהקות הבטות.

יש טעות במודל כי המודל מובהק והמקדמים לא: נכון/לא נכון/ לא ניתן לדעת

ב. בהנחה כי מתקיים : $X_{1t} - 2X_{2t} = 1$

לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים: נכון/לא נכון/ לא ניתן לדעת

ג. בהנחה כי מתקיים: $x_{1t} = x_{2t}^2$.

לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים: נכון/לא נכון/ לא ניתן לדעת.

ד. הוכיחו תשובותיכם לסעיפים א ו- ב.

ה. בהנחה כי מתקיים: $r_{12} = 0.98$

1. לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים:

נכון/לא נכון/ לא ניתן לדעת.

2. איזו בעיה עלולה להיווצר במודל ומהן השלכותיה.

3. בהנחה שהמודל יצא מובהק אולם הבטות אינן מובהקות וערכי t

למובהקות הבטות הן כדלקמן: $t_{\hat{\beta}_1} = 1.31$, $t_{\hat{\beta}_2} = 1.45$ **, מה יהיה הפיתרון הטוב**

ביותר, לדעתכם, לבעיה במודל (אליה התייחסתם בסעיף 2)?

א. להוריד את x_1

ב. להוריד את x_2

ג. להוריד את שני המשתנים.

ד. להותיר את שני המשתנים.