

# מבוא לאקונומטריקה

פרק 17 - משתנה דמי

תוכן העניינים

1. כללי..... 1

## משתנה דמי:

### רקע:

הכנסת משתנים ב"ת איכותיים למודל הרגרסיה.

למשל, נתונה משוואת הרגרסיה:  $W_t = \alpha + \beta \cdot S_t$ .

$W_t$  = השכר (התלוי).

$S_t$  = שנות לימוד (הב"ת) שניהם כמותיים.

נניח שאנו סבורים שגם משתנה המגדר (משתנה איכותי) משפיע על השכר.

כדי להכניסו למשוואת הרגרסיה יש להגדיר משתני דמי (dummy variable):

נגדיר משתנה  $D$  שיקבל את הערך 0 אם מדובר ב"אישה" ואת הערך 1 אם מדובר ב"גבר". ניתן להכניס את משתנה הדמי למודל בשלושה אופנים שונים:

1. משתנה דמי לחותך – המגדר משפיע על השכר ההתחלתי בלבד.
2. משתנה דמי לשיפוע – המגדר משפיע על התוספת לשכר בגין שנות הלימוד.
3. משתנה דמי לכל הפונקציה – המגדר משפיע גם על החותך וגם על השיפוע.

### 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:  $\alpha_0$ .

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

הבדל בשכר בין נשים וגברים:  $\alpha_1$  (הפרש בין החותכים).

בדיקת השערות על משתנה הדמי: מבחן  $t$  למובהקות הפרש החותכים:  $H_0: \alpha_1 = 0$ . השיפוע מייצג את התוספת בשכר כפונקציה של מסי שנות הלימוד והוא זהה עבור נשים וגברים.

### פונקציית רגרסיה המכילה משתנים איכותיים בלבד:

המגדר הוא המשתנה היחיד במשוואה:  $W_t = \alpha_0 + \alpha_1 D + u_t$ .

החותך מייצג כאן את השכר הממוצע עבור כל קטגוריה:

שכר הממוצע של אישה:  $\alpha_0$ .

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

הבדל בשכר הממוצע בין נשים וגברים:  $\alpha_1$  (הפרש בין החותכים).

בדיקת השערות על משתנה הדמי: מבחן  $t$ :  $H_0: \alpha_1 = 0$  (מבחן זהה למבחן  $t$  להבדל בין ממוצעים).

**משתנה דמי לשיפוע:**

המגדר משפיע על התוספת לשכר בגין שנות הלימוד:  $W_t = \alpha + \beta_0 S_t + \beta_1 DS_t + u_t$ .  
 השיפוע מייצג כאן את התוספת לשכר בגין שנות לימוד.  
 אצל אישה: התוספת לשכר בגין שנות לימוד:  $\beta_0$ .  
 אצל גבר: התוספת לשכר בגין שנות לימוד:  $\beta_0 + \beta_1$ .  
 הבדל בין גברים לנשים בתוספת לשכר בגין שנות הלימוד:  $\beta_1$  (הפרש השיפועים).  
 בדיקת השערות על משתנה הדמי: מבחן  $t$  למובהקות הפרש השיפועים:  $H_0: \beta_1 = 0$ .  
 החותך, המייצג את השכר ההתחלתי, יהיה זהה עבור גברים ונשים.

**משתנה דמי לכל הפונקציה:**

המין משפיע גם על החותך וגם על השיפוע – גם על השכר ההתחלתי וגם על התוספת לשכר ההתחלתי בגין שנות הלימוד.  
 המודל:  $W_t = \alpha_0 + \alpha_1 D + \beta_0 S_t + \beta_1 DS_t + u_t$ .  
 השכר ההתחלתי של אישה:  $\alpha_0$ .  
 השכר ההתחלתי של גבר:  $\alpha_0 + \alpha_1$ .  
 הבדל בשכר ההתחלתי בין המינים:  $\alpha_1$  (הבדל בחותכים).  
 אצל אישה: התוספת לשכר בגין שנות הלימוד:  $\beta_0$ .  
 אצל גבר: התוספת לשכר בגין שנות הלימוד:  $\beta_0 + \beta_1$ .  
 הבדל בין המינים בתוספת לשכר בגין שנות הלימוד:  $\beta_1$  (הבדל בשיפועים).

2 דרכים לבדיקה האם יש השפעה למשתנה האיכותי:

1. בדיקת השערות למשתני הדמי:  
 באמצעות מבחן WALS יש לבדוק:  $H_0: \alpha_1 = \beta_1 = 0$ .  
 לפחות אחד הפרמטרים שונה מ-0:  $H_1$ .  
 אם דוחים את השערת האפס, יש לבצע מבחני  $t$  עבור כל אחד מהפרמטרים  
 בנפרד:  $H_0: \alpha_1 = 0$  ו-  $H_0: \beta_1 = 0$ .
2. מבחן CHOW:  
 דרך נוספת לבדיקת ההבדל בין הקטגוריות בלא יצירת משתני דמי:  
 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 WALS):  
 המודל המוגבל (R) לא לוקח בחשבון את השפעת המגדר ולכן יכול את  
 המדגם המאוחד:  $W_t = \alpha + \beta X_t + u_t$ .

המודל הלא מוגבל (U) כולל את שני חלקי המדגם:  
 $ESS_U = ESS_f + ESS_m$   
 $DF_U = DF_f + DF_m$

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

למרות התוצאות הזהות בשתי הדרכים, שיטת משתני הדמי עדיפה:

1. אם דחינו את  $H_0$  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.

### סיכום ביניים:

| משתנה דמי לכל הפונקציה  | משתנה דמי לשיפוע                                    | משתנה דמי לחותך                                       |               |
|---|---|---|---------------|
| $Y_t = \alpha_0 + \alpha_1 D + \beta_0 X_t + \beta_1 DX_t + u_t$  | $Y_t = \alpha + \beta_0 X_t + \beta_1 DX_t + u_t$   | $Y_t = \alpha_0 + \alpha_1 D + \beta \cdot X_t + u_t$ | המודל         |
| קיים הבדל בין הקטגוריות במשוואת הרגרסיה כולה (בחותך ובשיפוע).   | קיים הבדל בין הקטגוריות בתוספת ל-Y בגין X (בשיפוע). | קיים הבדל בין הקטגוריות ב-Y ההתחלתי (בחותך).          | ההשערה במילים |
| מבחן WALS להפרש בין הפונקציות (החותכים והשיפועים):<br>$H_0: \alpha_1 = \beta_1 = 0$<br>**ניתן לבדוק את ההשערה בדבר הבדל בין הפונקציות גם במבחן CHOW.<br>אם דוחים את $H_0$ יש לברר את מקור ההבדל באמצעות מבחני t (אפשרי רק ב-WALS):<br>$H_0: \alpha_1 = 0$<br>$H_0: \beta_1 = 0$ | מבחן t להפרש השיפועים:<br>$H_0: \beta_1 = 0$        | מבחן t להפרש החותכים:<br>$H_0: \alpha_1 = 0$          | בדיקת ההשערה  |

**משתני דמי אם המשתנה האיכותי יכול לקבל יותר משני ערכים:**

כאשר המשתנה האיכותי כולל יותר משני ערכים/קטגוריות 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. משתני דמי לחותך:**

הטענה: יש הבדל בין עונות השנה במחיר ההתחלתי של הירקות.

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

כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב- $\beta$ . למחיר זה יתווסף  $\alpha_0$  בחורף,  $\alpha_0 + \alpha_1$  באביב,  $\alpha_0 + \alpha_2$  בקיץ ו- $\alpha_0 + \alpha_3$  בסתיו.

ניתן לראות כי:  $\alpha_0$  - החותך בקטגוריה שהושמטה,  $\alpha_0 + \alpha_1$  - החותך בקטגוריה i.

בדיקת השערות:

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$

השערות:  $H_1: OTHERWISE$

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  $\alpha_0$  שכן המשתנה המסביר של עונות השנה ירד.

אם נדחה את  $H_0$  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:  
 $H_0: \alpha_1 = 0$

2. האם יש הבדל במחיר ההתחלתי של הירקות בין הקיץ לחורף:  
 $H_0: \alpha_2 = 0$

3. האם יש הבדל במחיר ההתחלתי של הירקות בין הסתיו לחורף:  
 $H_0: \alpha_3 = 0$

2. משתני דמי לשיפוע:

הטענה: יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן.

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

המחיר ההתחלתי של הירקות שווה בין עונות השנה ( $\alpha$ ) אולם כל עליה

של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות

ב:  $\beta_0$  בחורף,  $\beta_0 + \beta_1$  באביב,  $\beta_0 + \beta_2$  בקיץ ו- $\beta_0 + \beta_3$  בסתיו.

ניתן לראות כי-  $\beta_0$ : השיפוע בקטגוריה שהושמטה  $\beta_0 + \beta_i$ :

השיפוע בקטגוריה i.

בדיקת השערות:

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

השערות:

$$H_1: \text{OTHERWISE}$$

המבחן הסטטיסטי – מבחן WALD:

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

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

- שימו לב שהשיפוע במשוואה המוגבלת איננו  $\beta_0$  שכן המשתנה המסביר של עונות השנה ירד.

אם נדחה את  $H_0$  במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין השיפועים על ידי מבחני  $t$ .

3. משתני דמי לכל הפונקציה :

הטענה : יש הבדל בין עונות השנה בפונקציית הרגרסיה לניבוי מחיר הירקות באמצעות המחיר לצרכן. המודל :

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

בדיקת השערות :

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$$

המבחן הסטטיסטי - מבחן WALD :

(U)

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

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

אם דוחים את  $H_0$ , יש לבדוק במבחן WALD האם ההבדל הוא בין החותכים

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0, H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$$

אם דוחים את  $H_0$  יש להמשיך לבדוק באמצעות מבחן  $t$  :

$$H_0 : \beta_j = 0, H_0 : \alpha_j = 0$$

### משתני דמי עבור שני משתנים איכותיים :

לדוגמא – שני משתנים איכותיים המשפיעים על פונקציית 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. הבדל בחותך ללא אינטראקציה :

$$W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \beta \cdot S_t + u_t$$

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

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

$$1. H_0 : \alpha_1 = 0$$

$$2. H_0 : \alpha_2 = 0$$

2. הבדל בחותך עם אינטראקציה :

$$W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \alpha_3 G \cdot R + \beta \cdot S_t + u_t$$

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

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

$$H_0 : \alpha_3 = 0$$

3. דרך נוספת ליצירת מודל עם אינטראקציה :

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.

$$W_t = \gamma_0 + \gamma_1 D_1 + \gamma_2 D_2 + \gamma_3 D_3 + \delta \cdot S_t + u_t$$

נעזר בטבלה בכדי לנסח את ההשערות לבדיקת האינטראקציה :

| הפרש                  | אישה                  | גבר                   |      |
|-----------------------|-----------------------|-----------------------|------|
| $\gamma_1 - \gamma_3$ | $\gamma_0 + \gamma_3$ | $\gamma_0 + \gamma_1$ | לבן  |
| $\gamma_2$            | $\gamma_0$            | $\gamma_0 + \gamma_2$ | שחור |
|                       | $\gamma_3$            | $\gamma_1 - \gamma_2$ | הפרש |

ההשערות לבדיקת קיום האינטראקציה :  $H_0 : \gamma_1 - \gamma_3 = \gamma_2$  או  $H_0 : \gamma_1 - \gamma_2 = \gamma_3$  התוצאות שיתקבלו כאן יהיו כמובן זהות לחלוטין לתוצאות שהתקבלו בדרך

$$WALD = t^2$$

$$PF = P_t$$

## שאלות:

## משתנה דמי לחותך:

- (1) 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 באוכלוסייה?
- 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.

## פונקציית רגרסיה המכילה משתנים איכותיים בלבד:

- (2) 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 הממוצע בין נשים וגברים?

## משתנה דמי לשיפוע:

- (3) על בסיס אותו מדגם, ביקש החוקר לדעת האם קיים הבדל מובהק בין גברים לנשים בתוספת לשכר בגין שנות הלימוד.
- תוצאות האמידה נתונות להלן:
- $$W_t = 5000 + 110 \cdot S_t + 120 \cdot D \cdot S_t + u_t$$
- (25) (23) (68)
- בדוק את ההשערה.

## משתנה דמי לכל פונקציה:

(4) חוקר רצה לבדוק את הטענה שסוג הכביש משפיע על מס' תאונות הדרכים בקטעי כביש בינעירוניים, בהינתן נפח התנועה. החוקר בדק האם הפונקציה של מס' התאונות בהינתן נפח התנועה, שונה בין כבישים מהירים לבין כבישים שאינם מהירים. לשם כך אמד החוקר את ארבע המשוואות הבאות:

$$1. \quad NUM_t = \gamma_1 + \delta_1 \cdot AVGD_t + \varepsilon_{1t} \quad \text{כבישים מהירים בלבד.}$$

$$2. \quad NUM_t = \gamma_2 + \delta_2 \cdot AVGD_t + \varepsilon_{2t} \quad \text{כבישים לא מהירים בלבד.}$$

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

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

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.

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

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

ג. 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.

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

ה. מהי הרגרסיה "תחת"  $H_0$  למבחן WALT?

### 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  |

## משתנה איכותי עם יותר משתי קטגוריות:

(5) ענה על הסעיפים הבאים:

- א. הועלתה הטענה כי יש הבדל במחיר ההתחלתי בין האביב לקיץ.  
 i. מהי השערת האפס לבדיקת הטענה?  
 ii. 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.  
 i. מהי השערת האפס לבדיקת הטענה?  
 ii. מהו המבחן הסטטיסטי המתאים? פרטו.

## משתנה דמי עבור שני משתנים איכותיים:

(6) 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 בבלט 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 |                | t Value | Pr >  t |
|------------------|----|-----------|----------------|---------|---------|
|                  |    | Estimate  | Standard Error |         |         |
| 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 <sup>2</sup> | 1  | -----     | -----          | -7.45   | 0.00    |

- 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. מהי השערת האפס לבדיקת הטענה כי אין אינטראקציה בין גזע להשכלה?
- ה. לבדיקת ההשערה של הסעיף הקודם בוצע מבחן W.L.D.  
 הרגרסיה המוגבלת תחת השערת האפס הינה :
- $$Z_0 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \gamma_4 Z_4 + v$$
- מהם ה-Zים?
- ו. בדוק את ההשערה אם ידוע שבמודל המוגבל  $R^2 = 0.33$ .
- ז. החוקר החליט לאמוד במקום את המשוואה המקורית את המשוואה :
- $$\ln(Y)_i = \lambda_0 + \lambda_1 S + \lambda_2 E + \lambda_3 (S \cdot E) + \delta_1 EXP + \delta_2 EXP^2 + \omega_i$$
- 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 :
- $$\ln(Y)_i = \lambda_0 + \lambda_1 S + \lambda_2 E + \delta_1 EXP + \delta_2 EXP^2 + \omega_i$$
- ספציפיקציה של השמטת משתנה רלוונטי (היעזר בסעיפים ד', ו' ו-ז').

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

$$\ln(MWAGE) = \alpha_0 + \alpha_1 \cdot S + \alpha_2 \cdot E + \alpha_3 \cdot (S \cdot E) + \beta_0 \cdot EXP + \beta_1 \cdot (EXP \cdot S) + \beta_2 \cdot (EXP \cdot S) + \beta_3 \cdot (EXP \cdot S \cdot E) + U$$

כאשר :

$S$  משתנה דמי : 1 = עבור נשים, 0 = גברים.

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

א. רשמו את הפונקציה לחישוב :

- i. תחזית לוג השכר עבור גבר בעל השכלה נמוכה ו-10 שנות ניסיון.
- ii. תחזית לוג השכר ההתחלתי עבור נשים משכילות.
- iii. לאחר כמה שנות ניסיון ישתווה השכר של נשים משכילות לזה של גברים משכילים?

ב. רשמו את השערות האפס המתאימות לבדיקת הטענות הבאות :

- i. אין השפעה של מגדר והשכלה על השכר.
- ii. השפעת ההשכלה אינה תלויה במגדר.
- iii. אין השפעות השכלה אצל גברים.
- iv. אין הבדל בשיעורי התשואה לניסיון, בקרב הנשים.

### תרגול מסכם :

(8) על מנת לאמוד השפעת מגדר ומצב משפחתי על השכר, נאמדה המשוואה הבאה :

$$WAGE = \alpha_0 + \alpha_1 \cdot GENDER + \alpha_2 \cdot FS + \alpha_3 \cdot (GENDER \cdot FS) + \beta_1 \cdot EDUC + \beta_2 \cdot AGE + U$$

כאשר :

$GENDER$  = מגדר : 1 = גבר, 0 = אישה.

$FS$  = מצב משפחתי : 1 = נשואים, 0 = לא נשואים.

$EDUC$  = מס' שנות לימוד של העובד.

$AGE$  = גיל העובד.

$WAGE$  = שכר העובד.

משוואה (1) נאמדה בפלט מס' 1.

בנוסף נאמד גם פלט מס' 2.

- א. החוקרת הניחה כי פערי השכר, באים לידי ביטוי בשכר ההתחלתי בלבד : נכון/ לא נכון/ לא ניתן לדעת.
- ב. החוקרת הניחה כי הפערים בין נשים לגברים בשכר אינם תלויים בגיל : נכון/ לא נכון/ לא ניתן לדעת.
- ג. השערת האפס לבדיקת הטענה היא : \_\_\_\_\_.
- ד. המשתנה המוסבר ברגרסיה מס' 2 הינו : \_\_\_\_\_ (כתבו את המודל שבו מחושב המשתנה המוסבר).
- ה. הסטטיסטי של LM לבדיקת הטענה :
  - i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.
  - ii. ניתן לחשבו וערכו הוא : \_\_\_\_\_.

ו. המקדם של GENDER בפלט מס' 2 הינו: \_\_\_\_\_.

הועלתה הטענה כי הפערים בין גברים לנשים בקרב העובדים הנשואים גבוהים ביותר מ-1500 ש"ח מאשר הפערים בקרב העובדים שאינם נשואים.

ז. ההשערות לבדיקת הטענה הינן: \_\_\_\_\_.

ח. הסטטיסטי לבדיקת הטענה:

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: \_\_\_\_\_.

ט. ההשערות לבדיקת הטענה הן: \_\_\_\_\_.

י. המודל המוגבל לבדיקת הטענה הוא: \_\_\_\_\_.

יא. הסטטיסטי לבדיקת הטענה:

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחישוב וערכו הוא: \_\_\_\_\_.

### פלט מס' 1 - משוואה 1:

**Dependent variable: WAGE**

**Number of observations used: 17495**

#### Analysis of Variance

| Source  | DF    | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|-------|----------------|-------------|---------|--------|
| Model   | 5     | 1.504815E11    | 30096294654 | 646.42  | <.0001 |
| Error   | 17489 | 8.142567E11    | 46556220    |         |        |
| C Total | 17494 | 9.647382E11    |             |         |        |

|          |            |          |        |
|----------|------------|----------|--------|
| Root MSE | 6823.35843 | R-square | 0.1560 |
| Dep Mean | 7286.58004 | Adj R-sq | 0.1557 |
| C.V.     | 93.64281   |          |        |

#### Parameter Estimates

| Variable  | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob> T |
|-----------|----|--------------------|----------------|-----------------------|---------|
| INTERCEP  | 1  | -3642.10108        | 260.72351      | -13.97                | <.0001  |
| GENDER    | 1  | 2006.13583         | 187.64224      | 10.69                 | <.0001  |
| FS        | 1  | 899.68055          | 159.19316      | 5.65                  | <.0001  |
| GENDER*FS | 1  | 1964.31810         | 227.43348      | 8.64                  | <.0001  |
| EDUC      | 1  | 428.20041          | 12.45434       | 34.38                 | <.0001  |
| AGE       | 1  | 64.72379           | 4.43948        | 14.58                 | <.0001  |

## פלט מס' 2 - מבחן LM:

Dependent variable :

Number of observations used: 17495

## Analysis of Variance

| Source  | DF    | Sum of Squares | Mean Square | F Value | Prob>F |
|---------|-------|----------------|-------------|---------|--------|
| Model   | 5     | 66653745252    | 13330749050 | 286.32  | <.0001 |
| Error   | 17489 | 8.142567E11    | 46558220    |         |        |
| C Total | 17494 | 8.809105E11    |             |         |        |

|          |             |          |        |
|----------|-------------|----------|--------|
| Root MSE | 6823.35843  | R-square | 0.0757 |
| Dep Mean | 2.29222E-12 | Adj R-sq | 0.0754 |
| C.V.     | 2.97675E17  |          |        |

## Parameter Estimates

| Variable  | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob> T |
|-----------|----|--------------------|----------------|-----------------------|---------|
| INTERCEP  | 1  | -1244.40187        | 260.72351      | -4.77                 | <.0001  |
| GENDER    | 1  |                    |                |                       |         |
| FS        | 1  | 899.68055          | 159.19316      | 5.65                  | <.0001  |
| GENDER*FS | 1  | 1964.31810         | 227.43348      | 8.64                  | <.0001  |
| EDUC      | 1  | 23.18457           | 12.45434       | 1.86                  | 0.0627  |
| AGE       | 1  | -38.13257          | 4.43948        | -8.59                 | <.0001  |

9) במטרה לאמוד את העושר הפיננסי של לקוחות הבנק כפונקציה של ההכנסה, הוותק והתנהגות פיננסית נאמדה המשוואה הבאה:

$$OSHER_t = \alpha_0 + \alpha_1 \cdot V + \alpha_2 \cdot G + \alpha_3 \cdot V \cdot G + \beta \cdot INCOME_t + V_t \quad (1)$$

כאשר:

$OSHER_t$  = העושר הפיננסי של הלקוח.

$INCOME_t$  = ההכנסה של הלקוח.

$V$  = משתנה דמי. 1 = לקוחות ותיקים. 0 = לקוחות חדשים.

$G$  = משתנה דמי. 1 = לקוחות בעלי התנהגות פיננסית תקינה. 0 = אחרת.

$V_t$  = סטייה מקרית המקיימת את כל ההנחות הקלאסיות.

- משוואה (1) נאמדה בפלט מס' 4. כמו כן נאמד פלט מס' 5. התייחסו לפלטים אלו בלבד.

א. החוקרת הניחה כי הנטייה השולית לחסוך (לצבור עושר) מתוך ההכנסה,

איננה תלויה בוותק או בהתנהגות הפיננסית של הלקוח:

נכון/ לא נכון/ לא ניתן לדעת

ב. החוקרת הניחה כי ההבדל בין לקוחות חדשים ללקוחות וותיקים, תלוי

בהתנהגות הפיננסית של הלקוח:

נכון/ לא נכון/ לא ניתן לדעת

הועלתה הטענה כי אין השפעה של וותק והתנהגות פיננסית על העושר.

ג. השערת האפס לבדיקת הטענה הינה: \_\_\_\_\_

ד. הסטטיסטי של WALT לבדיקת הטענה:

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: \_\_\_\_\_

ה. הסטטיסטי של CHOW לבדיקת הטענה שהוותק וההתנהגות הפיננסית

אינם משפיעים על פונקצית העושר:

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: \_\_\_\_\_

הועלתה הטענה כי אין השפעה של ותק בקרב לקוחות בעלי התנהגות פיננסית לא תקינה.

ו. השערת האפס לבדיקת הטענה הינה: \_\_\_\_\_

ז. הסטטיסטי לבדיקת הטענה:

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: \_\_\_\_\_

ח. אם נתון ש:  $\alpha_2 = 0$  המשמעות הינה:

i. אין השפעה של התנהגות פיננסית בקרב לקוחות חדשים.

ii. אין השפעה של התנהגות פיננסית

iii. אין השפעה של התנהגות פיננסית בקרב לקוחות וותיקים

iv. כל התשובות אינן נכונות

### פלט מס' 4 - משוואה 1:

(4)

Model: MODEL1  
Dependent Variable: OSHER OSHER

Analysis of Variance

| Source          | DF    | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|-------|----------------|-------------|---------|--------|
| Model           | 4     | 2.337947E14    | 5.844868E13 | 441.20  | <.0001 |
| Error           | 58546 | 7.755951E15    | 1.324762E11 |         |        |
| Corrected Total | 58550 | 7.989746E15    |             |         |        |

  

|                |            |          |        |
|----------------|------------|----------|--------|
| Root MSE       | 363973     | R-Square | 0.0293 |
| Dependent Mean | 10546      | Adj R-Sq | 0.0292 |
| Coeff Var      | 3451.27703 |          |        |

Parameter Estimates

| Variable             | Label     | DF | Parameter Estimate | Standard Error | t Value | Pr >  t |
|----------------------|-----------|----|--------------------|----------------|---------|---------|
| $\alpha_0$ Intercept | Intercept | 1  | 1678.45358         | 1975.42706     | 0.85    | 0.3955  |
| $\alpha_1$ v         |           | 1  | 4727.38758         | 3077.65435     | 1.54    | 0.1245  |
| $\alpha_2$ g         |           | 1  | 475649             | 18512          | 25.69   | <.0001  |
| $\alpha_3$ vg        |           | 1  | -290742            | 26068          | -11.15  | <.0001  |
| $\beta$ income       | income    | 1  | 0.25845            | 0.00820        | 31.51   | <.0001  |

## פלט מס' 5:

(R)

Model: MODEL1  
Dependent Variable: OSHER OSHER

Analysis of Variance

| Source          | DF            | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|---------------|----------------|-------------|---------|--------|
| Model           | $k=1$         | 1.328125E14    | 1.328125E14 | 989.70  | <.0001 |
| Error           | $T-k-1=58549$ | 7.856933E15    | 1.341941E11 |         |        |
| Corrected Total | $T-1=58550$   | 7.989746E15    |             |         |        |

  

|                |            |          |        |
|----------------|------------|----------|--------|
| Root MSE       | 366325     | R-Square | 0.0166 |
| Dependent Mean | 10546      | Adj R-Sq | 0.0166 |
| Coeff Var      | 3473.58313 |          |        |

Parameter Estimates

| Variable  | Label     | DF | Parameter Estimate | Standard Error | t Value | Pr >  t |
|-----------|-----------|----|--------------------|----------------|---------|---------|
| Intercept | Intercept | 1  | 8057.00227         | 1515.97451     | 5.31    | <.0001  |
| income    | income    | 1  | 0.25967            | 0.00825        | 31.46   | <.0001  |

## תשובות סופיות:

- (1) א.  $W_t = 7971$  . ב. 1,043 נח. ג. כן. ד. יש עדות לכך.
- ה. יש עדות לכך. (2)  
יש עדות לכך. (3)  
יש עדות לכך. (3)
- (4) א. יש עדות לכך, מבחן CHOW : 1, 2 ו-3, משתנה דמי : 3 ו-4.  
ב.  $\hat{\alpha} = 0.14978$ ,  $\hat{\beta}_1 = 1.40311$ ,  $\hat{\beta}_2 = 0.002877$ ,  $\hat{\beta}_3 = -0.008$ .  
ג.  $NUM_t = 1.532398$ . ד.  $H_0 : \beta_2 + \beta_3 = 2 \cdot \beta_2$   
 $H_0 : \beta_3 = \beta_2$   
ה.  $NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_3 \cdot (AVGD_t + AVGD \cdot TYPE_t) + U_t$ .
- (5) א. i.  $H_0 : \alpha_1 = \alpha_2$  . ii. WALD ו-t. ב. i.  $H_0 : \alpha_1 = \alpha_2, \alpha_3 = 0$  . ii. WALD.  
א. נכון. ב. יש עדות לכך. ג. יש עדות לכך. (6)  
ד.  $H_0 : \alpha_2 = \alpha_1 - \alpha_3$  או  $H_0 : \alpha_3 = \alpha_1 - \alpha_2$   
ה.  $Z_0 = \ln(Y)_t$ ,  $Z_1 = D_1 + D_3$ ,  $Z_2 = D_2 - D_3$ ,  $Z_3 = EXP_t$ ,  $Z_4 = EXP_t^2$   
ו. אין עדות לכך. ז.  $\lambda_0 = \alpha_0$ ,  $\lambda_1 = \alpha_2$ ,  $\lambda_2 = \alpha_3$ ,  $\lambda_3 = \alpha_1 - \alpha_2 - \alpha_3$ . ח. לא.
- (7) א. i.  $\hat{\ln}(MWAGE) = \hat{\alpha}_0 + \hat{\beta}_0 \cdot 10$  . ii.  $\hat{\ln}(MWAGE) = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3$  . iii.  $EXP_t = \frac{-(\alpha_1 + \alpha_3)}{\beta_1 + \beta_3}$   
ב. i.  $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$  . ii.  $H_0 : \alpha_3 = \beta_3 = 0$  . iii.  $H_0 : \alpha_2 = \beta_2 = 0$  . iv.  $H_0 : \beta_2 + \beta_3 = 0$   
א. נכון. ב. נכון. ג.  $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$  . ד. ראו סרטון. ה. ראו סרטון. ו. ראו סרטון. (8)  
ז.  $H_0 : \alpha_3 = 1500$  . ח. ii,  $t_{stat} = 2.04$  . ט.  $H_0 : \alpha_1 = \alpha_2$   
י.  $H_1 : \alpha_3 > 1500$  . יא.  $H_1 : \alpha_1 \neq \alpha_2$   
י.  $WAGE = \alpha_0 + \alpha_2 \cdot (GENDER + FS) + \alpha_3 \cdot (GENDER \cdot FS) + \beta_1 \cdot EDUC + \beta_2 \cdot AGE + U$  . יא. i.