

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:

$$W_t = \text{השכר (המשתנה התלוי)}$$

$$S_t = \text{שנות לימוד (המשתנה הב"ת)}$$

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.

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

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: $H_0: \alpha_1 = 0$

** השיפוע מייצג את התוספת לשכר 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.

(2) 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- 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: $H_0: \beta_1 = 0$

** 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.

? חברה מסוימת מתמרצת עובדים ונותנת להם תגמול לפי מספר השעות הנוספות שעבדו בחודש.

נגדיר $Y = \text{סה"כ התגמול שמקבל עובד שעבד } X \text{ שעות נוספות בחודש}$.

$D = 1$ אם $10 < X$ ו $D = 0$ אחרת.

איזו משוואת רגרסיה יש להריץ בכדי לאמוד את סה"כ התגמול בתנאים הבאים:

1. מי שעבד עד וכולל 10 שעות נוספות בחודש תגמולו יהיה קבוע וזהה לכל השעות הנוספות ואילו מי שעבד מעבר ל-10 שעות נוספות מקבל תגמול שונה לכל שעה נוספת מעבר ל-10 הראשונות.

2. עובדים שעבדו יותר מ-10 שעות נוספות בחודש מקבלים סכום קבוע ללא תלות במספר השעות הנוספות שעבדו.

(3) מישתנה דמי לכל הפונקציה

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

$$\text{המודל: } W_t = \alpha_0 + \alpha_1 D + \beta_0 S_t + \beta_1 DS_t + u_t$$

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)

בדיקת השערות למשתני הדמי:

$$H0: \alpha_1 = \beta_1 = 0$$

באמצעות מבחן WALS יש לבדוק:

H1: לפחות אחד הפרמטרים שונה מ-0

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

$$H0: \beta_1 = 0 \quad \text{vs} \quad H0: \alpha_1 = 0$$

? 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 המשוואות הבאות:

$$NUM_t = \gamma_3 + \delta_3 \cdot AVGD_t + \varepsilon_{3t} \quad (1)$$

$$NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_2 \cdot AVGD_t + \beta_3 \cdot (AVGD \cdot TYPE)_t + U_t \quad (2)$$

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 מוצגות להלן:

$$NUM_t = 0.739 + 0.0233 \cdot AVGD_t \quad (1)$$

$$NUM_t = 0.14978 + 1.40311 \cdot TYPE_t + 0.002877 \cdot AVGD_t - 0.008 \cdot (AVGD \cdot TYPE)_t \quad (2)$$

$$ESS = 20963 \quad Pt_{\hat{\alpha}} = 0.0019; Pt_{\hat{\beta}} = 0.0001 \quad (1)$$

$$Pt_{\hat{\alpha}} = 0.6534; Pt_{\hat{\beta}_1} = 0.0067; Pt_{\hat{\beta}_2} = 0.0001; Pt_{\hat{\beta}_3} = 0.1283 \quad ESS = 20759 \quad (2)$$

1) בדקו את טענת החוקר.

2) מהו האומדן הנקודתי למס' התאונות בכביש מהיר כאשר נפח התנועה עומד על 4 אלפי מכוניות ליום בקטע הכביש האמור?

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

3) מהי השערת האפס לבדיקת הטענה?

4) מהי הרגרסיה "תחת" H_0 למבחן WALT ?

סיכום ביניים:

משתנה דמי לכול 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 (בחותרך ובשיפוע).	קיים הבדל בין הקטגוריות בתוספת ל- Y בגין X (בשיפוע).	קיים הבדל בין הקטגוריות ב- Y ההתחלתי (בחותרך).	ההשערה במילים
מבחן WALT להפרש בין הפונקציות (החותכים והשיפועים): $H_0: \alpha_1 = \beta_1 = 0$ אם דוחים את H_0 יש לברר את מקור ההבדל באמצעות מבחני t (אפשרי רק ב-WALT): $H_0: \alpha_1 = 0$ $H_0: \beta_1 = 0$	מבחן t להפרש השיפועים: $H_0: \beta_1 = 0$	מבחן t להפרש החותכים: $H_0: \alpha_1 = 0$	בדיקת ההשערה

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:

$$V_t = \text{מדד מחירי הירקות}$$

$$p_t = \text{מדד המחירים לצרכן}$$

(1) משתני דמי לחותך

הטענה: יש הבדל בין עונות השנה במחיר ההתחלתי של הירקות

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

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

בדיקת השערות:

השערות:

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

$$H1: OTHERWISE$$

המבחן הסטטיסטי :

מבחן WALT:

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

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

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

אם נדחה את $H0$ 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.

? כלכלן הציע לאמוד את המודל הבא: $ta25_t = \alpha_0 + \alpha_1 Q1_t + \alpha_2 Q2_t + \alpha_3 Q3_t + \alpha_4 Q4_t + u_t$

כאשר:

Ta25 - תשואת מדד המניות ת"א 25.

Q_i - משתנה דמי המקבל ערך 1 עבור רבעון i ו-0 אחרת.

מי מהטענות הבאות נכונה (יש רק אחת):

1. לא ניתן לאמוד את המודל כיוון שאין שום משתנה מסביר אלא רק משתני דמי.

2. לא ניתן לאמוד את המודל יש בעיית מולטיקוליניאריות מלאה.

3. אין שום בעיה לאמוד את המודל.

4. כל התשובות האחרות שגויות.

(2) משתני דמי לשיפוע

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

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

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 .

בדיקת השערות:

השערות:

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

$H_1: OTHERWISE$

המבחן הסטטיסטי:

מבחן WALT:

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

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

**שימו לב שהשיפוע במשוואה המוגבלת איננו β_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 בפונקצית הרגרסיה לניבוי מחיר הירקות באמצעות המחיר לצרכן.

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

בדיקת השערות:

השערות:

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

המבחן הסטטיסטי:

מבחן WALT:

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

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

אם דוחים את H_0 , יש לבדוק במבחני WALT האם ההבדל הוא בין החותכים או בין השיפועים:

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

באם דוחים את H_0 יש להמשיך לבדוק באמצעות מבחני t:

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

? ידוע כי התוצר בישראל תלוי בתוצר בארה"ב, על כן הוצע לבחון את המודל:

$$gdp_i = \alpha + \beta \cdot gdp_usa_i + u_i$$

כאשר gdp הוא אחוז הצמיחה בתוצר בישראל.

ו- gdp_usa הוא אחוז הצמיחה בתוצר בארה"ב.

הועלתה הטענה שיש עונתיות לפי רבעונים לשם כך הוגדר משתנה הדמי $q = \text{רבעון}$.

רשמו את השערות האפס עבור הטענות הבאות:

א. הרבעונים משפיעים על אחוז הצמיחה בתוצר בישראל באופן כללי וכתלות בהשפעת

התוצר האמריקאי?

ב. השפעת הרבעון השלישי על התוצר בישראל זהה להשפעת הרבעון הרביעי?

ג. השפעת התוצר האמריקאי על התוצר בישראל ברבעון השני זהה לרבעון הראשון?

ד. הרבעונים השלישי והראשון משפיעים באופן זהה על התוצר בישראל באופן ישיר וגם

כתלות בתוצר האמריקאי.

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) הבדל בחותך ללא אינטראקציה

המודל:

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

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. הבדל בשכר ההתחלתי בין גברים לנשים: $H_0: \alpha_1 = 0$

2. הבדל בשכר ההתחלתי בין שחורים ללבנים: $H_0: \alpha_2 = 0$

(2) הבדל בחותך עם אינטראקציה

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

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:

$$H_0: \alpha_3 = 0 \quad .3$$

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

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

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

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

בטרם ניגשים לפיתרון השאלה יש להכין את טבלת העזר הבאה:

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

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.