מבוא לסטטיסטיקה ואקונומטריקה (סטטיסטיקה ב) קורס 66155

פרק 30 - מבחן t

תוכן העניינים

מבחן t:

רקע:

המבחן הסטטיסטי למובהקות מקדמי הרגרסיה.

המבחן הסטטיסטי	ניסוח	השערות	סטטיסטי המבחן	כלל הכרעה לדחיית H_0	מסקנה
מובהקות השיפוע	האם משתנה מסביר מסוים רלוונטי / למודל משפיע על התלוי?	$H_0:\beta=0$ $H_1:\beta\neq 0$	$t_{\beta=0}=\frac{\hat{\beta}-\beta_0}{S_{\hat{\beta}}}$	שימוש בטבלת T: $\lvert t_{\beta=0}\rvert > t_{\left(n-K,\frac{\alpha}{2}\right)}$ **K=מספר מקדמים (כולל חותך)	יש/אין עדות לכך שהמשתנה הב"ת מובהק באוכ'
מבחן חד צדדי לשיפוע	האם מקדם השיפוע חיובי/שלילי באוכ'?	$H_0:\beta=0$ $H_1:\beta>/<0$		שימוש בטבלת T: $t_{\beta=0}>t_{(n-K,\alpha)}$ $t_{\beta=0}<-t_{(n-K,\alpha)}$	יש/אין עדות לכך שהשיפוע חיובי/שלילי באוכ'
השיפוע = ערך מסוים באוכ'	האם מקדם השיפוע = לערך מסוים (למשל ל-2)?	$H_0:\beta=2$ $H_1:\beta\neq 2$	$t_{\beta=2}=\frac{\hat{\beta}-2}{S_{\hat{\beta}}}$	שימוש בטבלת t	יש/אין עדות לכך שהשיפוע = ל-2 באוכ'.
מבחן למובהקות החותך **ניתן לבצע גם מבחן חד צדדי ושהחותך = לערך מסוים באוכ'	האם קו הרגרסיה יוצא מראשית הצירים?	$H_0:\alpha=0$ $H_1:\alpha\neq 0$	$t_{\alpha=0}=\frac{\hat{\alpha}-\alpha_0}{S_{\hat{\alpha}}}$	נדחה את H_0 אם: שימוש בטבלת T: $\lvert t_{\alpha=0}\rvert > t_{\left(n-K,\frac{\alpha}{2}\right)}$	יש/אין עדות לכך שקו הרגרסיה עובר דרך ראשית הצירים

רב"ס ל- β : $P\left(\hat{\beta}-t_{\left(n-K,\frac{\alpha}{2}\right)}\cdot S_{\hat{\beta}}\le\beta\le\hat{\beta}+t_{\left(n-K,\frac{\alpha}{2}\right)}\cdot S_{\hat{\beta}}\right)=1-\alpha$.

רב"ס ל- α : $P\left(\hat{\alpha}-t_{\left(n-K,\frac{\alpha}{2}\right)}\cdot S_{\hat{\alpha}}\le\alpha\le\hat{\alpha}+t_{\left(n-K,\frac{\alpha}{2}\right)}\cdot S_{\hat{\alpha}}\right)=1-\alpha$.

- ניתן לבדוק השערות באמצעות הרב"ס.
  צריך לבדוק האם הרב"ס מכיל את הערך המבוקש לפי השערת האפס.
  אם כן – נקבל את H_0 ואם לא – נדחה אותה.

תחזית:

המטרה של קו הרגרסיה הוא ביצוע תחזיות:
<u>תחזית נקודתית</u> מחושבת על פי קו הרגרסיה שאמדנו.
נציב במקום ה- X ים ערכים נתונים ונקבל למה שווה ה- Y המנובא.

אמידת התחזית באוכלוסייה עבור ערך מסוים של X (ברגרסיה פשוטה):

$$\hat{Y}\pm t_{\left(n-2,\frac{\alpha}{2}\right)}S_u\sqrt{1+\frac{1}{n}+\frac{\left(X_f-\bar{X}\right)^2}{\sum\left(X_i-\bar{X}\right)^2}}$$

$$S_u^2=MSE=\frac{SSE}{n-2}\qquad\sum\left(X_i-\bar{X}\right)^2=S_{XX}=(n-1)S_X^2$$

רישום הרב"ס: $p(\_\_\_\le Y\le\_\_\_\_)=1-\alpha$.

התחזית מדויקת יותר (שונות התחזית קטנה יותר) כאשר:

1. n (גודל המדגם) גדול יותר.
2. שונות המשתנה המסביר X גדולה יותר.
3. X_f קרוב יותר ל- $\bar{X}$.
4. האומד לשונות הטעויות – S_u , קטן יותר.

מבחן t מורכב (בחינת קשרים ליניאריים בין הפרמטרים):

משמש לבדיקת השערות העוסקות בקשרים בין הפרמטרים.
כמו למשל: $H_0:\alpha=5\beta$ או $H_0:\beta_1=2\cdot\beta_2$.
במקרים אלו נרשום את השערות האפס כך: $H_0:\alpha-5\beta=0$ ו- $H_0:\beta_1-2\cdot\beta_2=0$
ונחשב את סטטיסטי המבחן t: $t_{\hat\alpha-5\hat\beta}=\dfrac{\left(\hat\alpha-5\hat\beta\right)-0}{S_{\hat\alpha-5\hat\beta}}$ או $t_{\hat\beta_1-2\hat\beta_2}=\dfrac{\left(\hat\beta_1-2\hat\beta_2\right)-0}{S_{\hat\beta_1-2\hat\beta_2}}$.
כאשר את טעות התקן של המבחן מחשבים תוך שימוש בנוסחאות:

$$V\left(X\pm Y\right)=V\left(X\right)+V\left(Y\right)\pm2\operatorname{cov}\left(X,Y\right)$$

$$V\left(aX\right)=a^2V\left(X\right)$$

$$\operatorname{cov}\left(aX,bY\right)=a\cdot b\cdot\operatorname{cov}\left(X,Y\right)$$

ואחר כך מוציאים לשונות שורש כדי לקבל את סטית התקן.
לשם כך יש לקבל נתונים על השונויות המשותפות של הפרמטרים (cov).

שאלות:

מובהקות מקדמי הרגרסיה:

1) חוקר רצה לבחון את השפעת ההכנסה ($INCOME$) על גובה המס (TAX) (במיליארדי $) שגובה מדינה במערב לפי המודל: $TAX_i = \alpha + \beta \cdot INCOME_i + u_i$.
לשם כך אסף נתונים מ-51 מדינות. להלן התוצאות:
$TAX_i = -0.086912 + 0.152232 \cdot INCOME_i$.
(0.08953) (0.01622).
סטיות התקן של האומדים נתונות בסוגריים.

א. מהי המשמעות הכלכלית של β ושל α?
ב. האם ההכנסה משפיעה על גודל המס? בדקו ברמת מובהקות של 5%.
ג. בדקו את ההשערה כי כאשר ההכנסה אפסית, גודל המס שונה מ-0 באוכלוסייה.
ד. בדקו את ההשערה כי ככל שההכנסה עולה כך עולה גם המס ברמת מובהקות של 5% וברמת מובהקות של 1%.
ה. בנו רווח-סמך לשיפוע הרגרסיה ברמת ביטחון של 95%.
ו. בדקו את ההשערה שתוספת של מיליארד $ להכנסה תגדיל את המס ב-0.2 מיליארד $, ברמת מובהקות של 0.05.

2) חוקר רצה לבדוק את השפעת הותק בעבודה (EXP) על השכר ($SALARY$) לפי המודל: $\ln(SALARY_t) = \alpha + \beta \cdot EXP_t + u_t$. הוא אסף 403 תצפיות, ואמד את הפרמטרים. להלן תוצאות האמידה:
$\ln(SALARY)_i = 7.334 - 0.0087 \cdot EXP_i$
(0.068) (0.0026).

א. האם קיים קשר חיובי מובהק בין ותק ללוג השכר?
ב. בדוק את ההשערה כי שיעור התשואה בשכר לשנת ותק קטנה מ: -0.9.
ג. מהי תחזית השכר עבור אדם בעל 10 שנות ותק?

3) נאמד המודל: $Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + \beta_3 W_i + \beta_4 S_i + u_i$ והתקבלו התוצאות הבאות:
$\hat{y}_i = 5.06 + 0.97x_i + 3z_i - 5.02w_i + 8.97s_i$.
(0.456) (0.42) (0.08) (0.7) (0.29).

א. האם משתנה W רלוונטי למודל? בדקו ברמת מובהקות של 0.01.
ב. בנו רווח בר סמך להשפעת X על Y.

תחזית:

4) נתונה משוואת הרגרסיה הבאה: $\hat{y}_i = 13 + 8x_{1i} + 7x_{2i} + 2x_{3i} + 9x_{4i}$.
כאשר y_i הינו סה"כ הוצאות משק בית i לחינוך לשבוע, x_{ji} הינו גילו של הילד j.
מה יהיה סה"כ הוצאות משק הבית אם גיל הילד הראשון הוא 2 שנים, של השני 4.5 שנים, השלישי הוא בן 5 ואילו הרביעי בן 8?

5) במדגם של 30 דירות המושכרות לסטודנטים ברדיוס של עד 2 ק"מ מסביב למכללה נחקר הקשר בין שכר דירה למספר הסטודנטים הגרים בדירה.
להלן תוצאות האמידה: $\hat{Y}_i = 686.207 + 233.52 \cdot X_i$.
נתון בנוסף כי:
$S_x^2 = 1.313^2$
$S_u^2 = 414.055^2$
$\bar{x} = 3$
א. חשבו אומדן נקודתי לשכר הדירה אותו ישלמו סטודנטים החולקים את הדירה עם שותף אחד בלבד.
ב. אמוד את שכר הדירה שישלמו סטודנטים החולקים את הדירה עם שותף אחד בלבד, ברמת בטחון של 95%.

t מורכב:

6) נתון המודל: $Y_t = \alpha + \beta X_t + u_t$ שלצורך אמידתו נאספו 240 תצפיות ונתקבל ש:
$\hat{Y}_i = 5.25 + 0.96X_i$.
(0.12) (0.25).
נתון בנוסף כי: $\text{cov}\left(\hat{\alpha}, \hat{\beta}\right) = -0.003$.
יש לבדוק את ההשערה: $H_0 : \alpha = 5\beta$.

7) על מנת לאמוד את פונקציית התצרוכת נאספו נתונים על 42 משקי בית בשנת 2007 ונאמדה המשוואה הבאה: $C_i = \alpha + \beta_1 \cdot W_i + \beta_2 \cdot P_i + u_i$.
להלן תוצאות האמידה של המשוואה הנ"ל:
$C_i = -107.226 + 0.743W_i + 0.561P_i$.
(0.4) (0.0678).
נתון גם ש: $\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -0.009$.
יש לבדוק את ההשערה שהנטייה השולית לצרוך (נש"צ) מתוך ההכנסה זהה לנטייה השולית לצרוך מתוך ההון.

תרגול מסכם:

8) כלכלן בנה עבור מכבי ת"א מודל החוזה את השכר שיש לשלם לשחקן כדורסל לחוזה של שנה: $\hat{Y}_i=\hat{\alpha}+\hat{\beta}_1X_{1i}+\hat{\beta}_2X_{2i}+\hat{\beta}_3X_{3i}+u_i$.

כאשר:

Y: שכר השחקן באלפי $.

X_1: מס' נקודות שקולע השחקן בממוצע למשחק.

X_2: מס' האסיסטים שיש לשחקן בממוצע למשחק.

X_3: מס' הדקות שיושב שחקן על הספסל בממוצע למשחק.

הכלכלן דגם 34 משחקים וקיבל את התוצאות הבאות:

$\hat{Y}_i=120+18X_{1i}+8X_{2i}-22X_{3i}$.

(2.2) (3) (4.4) (-5)

**הערכים שבסוגריים הם ערכי t.

התקבל בנוסף כי: $\text{cov}(\hat{\beta}_1,\hat{\beta}_2)=4, \text{cov}(\hat{\beta}_1,\hat{\beta}_3)=-3, \text{cov}(\hat{\beta}_2,\hat{\beta}_3)=-6$.

א. תנו פירוש למקדמי הרגרסיה.

ב. איזה מהמשתנים הב"ת רלוונטי למודל?

ג. בנו רב"ס למשתנים המובהקים.

ד. מייקל ג'ורדן הצטרף למכבי והוא דורש 2 מיליון $ לעונה.
ידוע כי מייקל קולע 45 נקודות בממוצע למשחק, מוסר 15 אסיסטים בממוצע למשחק ויושב 5 דקות בממוצע על הספסל. כמה צריך לשלם לו?

ה. לטענת שמעון מזרחי מס' הנקודות הממוצע שקולע שחקן למשחק צריך להשפיע פי 4 ממספר האסיסטים הממוצע שלו. האם הוא צודק?

9) כלכלן אמד את המודל הבא: $\ln(Q_i)=\alpha+\beta_1\ln(K_i)+u_i$ שמתאר את הקשר שבין צריכת מוצר מסוים להכנסת הפרט (עקומת אנג'ל):

K - הכנסה חודשית באלפי שקלים.

Q - צריכה שנתית באלפי שקלים.

לשם כך אסף 60 נתונים והריץ רגרסיה.

התוצאות אשר קיבל הן: $\hat{\beta}=-2$, $\hat{\alpha}=4$, $t_{\hat{\beta}}=-7$, $t_{\hat{\alpha}}=3$, $\text{cov}(\hat{\alpha},\hat{\beta})=-0.05$

, $S_Q=0.05$, $S_K=1.5$.

נקודת הממוצעים הינה: (6.7,0.4).

א. הכלכלן ביקש לבדוק את ההשערה כי הגמישות במודל יחידתית ושווה ל-1-.

ב. בדקו את ההשערה כי מקדם החיתוך של קו הרגרסיה הוא כפול ממקדם השיפוע.

ג. חיים משתכר בממוצע לחודש 10,000 ₪, כמה ישקיע בצריכת המוצר בשנה?

ד. בנו רב"ס לתחזית הצריכה של חיים באוכלוסייה.

תשובות סופיות:

1) א. ראה סרטון. ב. כן. ג. אין עדות לכך. ד. יש עדות לכך.
ה. $P\left(0.12\le\beta\le 0.184\right)=0.95$. ו. ניתן לדחות את השערת האפס.

2) א. לא. ב. אין עדות לכך. ג. $\hat{Y}=1404$.

3) א. כן. ב. $P\left(0.13\le\beta\le 1.81\right)=0.95$.

4) 142.5 ₪ לשבוע.

5) א. 1153.247. ב. $P\left(282.94\le Y_{X=2}\le 2023.55\right)=0.95$.

6) אין עדות לכך.

7) אין עדות לכך.

8) א. ראה סרטון. ב. כל שלושת המשתנים.
ג. $P\left(6\le\beta_1\le 30\right)=0.95$, $P\left(4.364\le\beta_2\le 11.636\right)=0.95$, $P\left(-30.8\le\beta_3\le -13.2\right)=0.95$.
ד. 940 אלף $. ה. כן.

9) א. אין עדות לכך. ב. יש עדות לכך. ג. 545 ₪.
ד. $P\left(-6.205\le Q_i\le 7.295\right)=0.95$.