سعید مرادی - 9613028

< سری سوم تمرین >

7.2) (a) با استفاده از رابطه‌ی بازه‌ی اطمینان، معنادار بودن ضریب High School Earnings را مشخص می‌کنیم

(*) $\text{Confidence interval} = \hat{\beta}_i \pm 1.96 \times SE(\hat{\beta}_i) \rightarrow CI = 0.352 \pm 1.96 \times (0.021)$

$CI_+ = 0.39316 \;,\; CI_- = 0.31084$ → پس این بازه به صورت $(0.31084, 0.39316)$ درمی‌آید

بنابراین چون صفر در این بازه قرار ندارد، فرض صفر (H_0) رد شده و در نتیجه ضریب معنادار است

(البته فرمول * برای بازه‌ی اطمینان 95% برقرار است)

(b) بازه‌ی اطمینان را با استفاده از رابطه‌ی نوشته شده در قسمت a، محاسبه می‌کنیم

$CI = \hat{\beta}_i \pm 1.96 \times SE(\hat{\beta}_i) \rightarrow CI = 0.458 \pm 1.96 \times 0.021 \rightarrow CI_+ = \cancel{0.49896}\ 0.49916 \;,\; CI_- = \cancel{0.41684}\ 0.41684$

پس یعنی با اطمینان 95% می‌توان گفت که ضریب در این بازه است که [illegible]

از آن جایی که صفر در بازه‌ی $CI = (0.41684, 0.49916)$ قرار نمی‌گیرد، پس فرض صفر (H_0) رد شده و ضریب Male معنادار است.

7.3) (a) با استفاده از Exclusion Restrictions یا همان مدل F_Statistics، معنادار بودن ضریب age و موثر بودن آن را بررسی می‌کنیم.

(نکته: ممکن است 2 متغیر به تنهایی معنادار نباشند اما در کنار هم قدرت توضیح‌دهندگی مدل را افزایش دهند) می‌دهد

که بنابراین با قرار گرفتن صفر در بازه‌ی اطمینان نمی‌توان نتیجه گرفت که وجود آن قدرت توضیح‌دهندگی مدل را افزایش

در روش Exclusion Restrictions، نیاز داریم که 2 مدل Restricted (مدلی که متغیر تشریحی age را ندارد) و Unrestricted (مدلی که متغیر تشریحی age را دارد) را بررسی کنیم

$$F = \frac{(SSR_r - SSR_{ur})}{SSR_{ur}} \times \frac{n-k-1}{q}$$

n: تعداد مشاهدات — k: تعداد کل متغیرهای تشریحی — q: تعداد متغیرهای حذف شده در مدل restricted

چون SSR , SST در جدول داده نشده، از تبدیل $SSR = SST(1-R^2)$ استفاده می‌کنیم در نتیجه رابطه به صورت زیر خواهد شد:

حال با استفاده از جایگذاری داریم:

$$F = \frac{(R^2_{ur} - R^2_r)}{1 - R^2_{ur}} \times \frac{(n-k-1)}{q} = \frac{(0.0761 - 0.0710)}{1 - 0.0761} \times \frac{(10973-3-1)}{1}$$

$$= 60.549734$$

از t_test برای تشخیص چه مهم بودن ضریب استفاده می‌کنیم. آماره به صورت زیر خواهد بود:

$t\_statistics = \dfrac{\hat{\beta}_1}{SE(\hat{\beta}_i)} \rightarrow t\_statistics = \dfrac{0.011}{0.001} = 11 > \underbrace{1.96}_{\text{critical value with 95\% confidence}}$ → ضریب Age برای AWE معنادار است

با استفاده از Confidence Interval نیز می توان موضوع را بررسی کرد: $CI = [\beta_1 - SE \times C, \beta_1 + SE \times C]$

با اطمینان 95 درصد ⇐ $[0.011 - 1.96 \times 0.001, 0.011 + 1.96 \times 0.001] \simeq [0.009, 0.013]$

ضریب در این بازه است و چون صفر در این بازه نیست پس فرض H_0 رد شده و ضریب معنادار است.

b) می دانیم که تفاوت در حقوق با ضریب Age مشخص شده و ضریب Age نیز عددی بین $[0.009, 0.013]$ است. در نتیجه تفاوت بین سن Alvo و Kal ، تفاوت در AWE را مشخص می کند (با توجه به اینکه بقیه ی متغیرها در هر دو مشاهده مقدار یکسانی دارند.) $\Delta AWE = \Delta Age \times CI$ ⇐ $\Delta AWE = 10 \times [0.009, 0.013] = [0.09, 0.13]$ ⇐ در نتیجه تفاوت در میانگین حقوق هفتگی دو فرد مشاهده شده عددی بین 0.09 و 0.13 دلار است.

7.7) a) از t_test استفاده می کنیم. $t\_statistics = \frac{\hat{\beta}_1}{SE(\beta_1)} = \frac{0.567}{1.23} \simeq 0.461 < 1.96$

پس فرض صفر (صفر بودن ضریب) رد نمی شود پس ضریب ها BDR به شکل معناداری از صفر متفاوت نیست.

b) با توجه به اینکه در بررسی ضریب BDR (تعداد اتاق های خانه) مساحت خانه ثابت در نظر گرفته شده است ولی در حقیقت با افزایش تعداد اتاق ها، مساحت نیز افزایش می یابد. در نتیجه افزایش تعداد اتاق ها خود تابع مساحت خانه است در نتیجه BDR قدرت توضیح دهندگی کمی در مدل دارد. به عنوان مثال یک منزل 1000 فوت مربعی می تواند بیش از 2 اتاق خواب داشته باشد و با این حال داشته باشد تاثیر چندانی در قیمت خانه ندارد زیرا این مساحت خانه است که تعداد اتاق ها و قیمت خانه را تعیین می کند. در واقع BDR، اثر جزئی تعداد اتاق های خواب را با در نظر گرفتن مساحت خانه (Hsize) در نظر می گیرد چون در یک مساحت مشخص تعداد اتاق ها چندان متغیر نیست، پس اثر ثابت معناداری روی قیمت ندارد.

c) بازه ی اطمینان 95% برای خانه ای با مساحت 2500 فوت مربع برابر است با: $2500 \times$ Confidence Interval:

$= 2500 \times [0.005 - 0.00072 \times 1.96, 0.005 + 0.00072 \times 1.96] = [8.972, 16.028]$

d) ضرایب مدل رگرسیون اگر به یکدیگر نزدیک باشند، تفسیر و مقایسه ی آنها با یکدیگر راحت تر است، بنابراین با توجه به

اینجا ضریب size بسیار کوچکتر از ضرایب دیگر است. بهتر است که مقدار آن با مقیاس بزرگتری در مدل گنجانده شود مثلاً برحسب 100 متر مربع یا 1000 فوت مربع در مدل قرار گیرد.

e) مقدار بحرانی در سطح 10% برابر با 2.30 است. از آن جایی که $F > 2.30$ است پس Age, BDR به طور همزمان با اطمینان 90% صفر نخواهند بود

7: a) اگر $H_0: \beta_1 = \beta_2$ باشد با تعریف متغیر $\theta = \beta_1 - \beta_2$ خواهیم داشت: $\beta_1 = \theta + \beta_2$

معادله‌ی اولیه داده شده به شکل رو‌به‌رو است ← $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$

حال باید صفر بودن فرض $H_0: \theta = 0$ را بررسی کنیم. برای بدست آوردن معادله باید تغییر متغیر انجام دهیم به طوری که تبدیل به متغیری با ضریب θ بشود (متغیر اولیه). در نتیجه $\beta_2 X_{1i}$ به آن اضافه و از آن کم می‌کنیم

$$Y_i = \beta_0 + \beta_1 x_{1i} - \beta_2 x_{1i} + \beta_2 x_{1i} + \beta_2 x_{2i} + u_i \Rightarrow$$

$$Y_i = \beta_0 + \underbrace{(\beta_1 - \beta_2)}_{\theta} x_{1i} + \beta_2 \underbrace{(x_{1i} + x_{2i})}_{x'_i} + u_i = \beta_0 + \theta x_{1i} + \beta_2 x'_i$$

x'_i ← نامگذاری می‌کنیم.

حال صحت فرض $H_0: \theta = 0$ را روی این رگرسیون بررسی کنیم به طوری که $\theta = \beta_1 - \beta_2$ ، $x'_i = x_{1i} + x_{2i}$

b) اگر $H_0: \beta_1 + 2\beta_2 = 0$ باشد آنگاه با تعریف متغیر $\theta = \beta_1 + 2\beta_2$ خواهیم داشت: $H_0: \theta = 0$

حال مانند مورد a ، هم تبدیل متغیر را انجام می‌دهیم: عبارت $2\beta_2 x_{1i}$ را اضافه و کم می‌کنیم:

$$Y_i = \beta_0 + \beta_1 x_{1i} + 2\beta_2 x_{1i} - 2\beta_2 x_{1i} + \beta_2 x_{2i} + u_i = \beta_0 + u_i + \underbrace{(\beta_1 + 2\beta_2)}_{\theta} x_{1i} + \beta_2 \underbrace{(-2x_{1i} + x_{2i})}_{x'_i} \Rightarrow Y_i = \beta_0 + \theta x_{1i} + \beta_2 x'_i + u_i$$

حال فرض $H_0: \theta = 0$ را بررسی می‌کنیم.

c) در این حالت اگر $H_0: \beta_1 + \beta_2 = 1$ باشد، با تعریف متغیر $\theta = \beta_1 + \beta_2 - 1$ خواهیم داشت $H_0: \theta = 0$

مانند 2 مورد قبلی تبدیل را انجام می‌دهیم. عبارت $\beta_2 x_{1i} - x_{1i}$ را اضافه و کم می‌کنیم:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i} - x_{1i} - \beta_2 x_{1i} + \beta_2 x_{2i} + x_{1i} + u_i = Y_i = \underbrace{(\beta_1 + \beta_2 - 1)} x_{1i} + \beta_2 \underbrace{(x_{2i} - x_{1i})}_{x'_i} + x_{1i} + u_i \Rightarrow \underbrace{Y_i - x_{1i}}_{Y'_i} = \theta x_{1i} + \beta_2 x'_i + \beta_0 + u_i$$

$$Y'_i = \theta x_{1i} + \beta_2 x'_i + \beta_0 + u_i$$