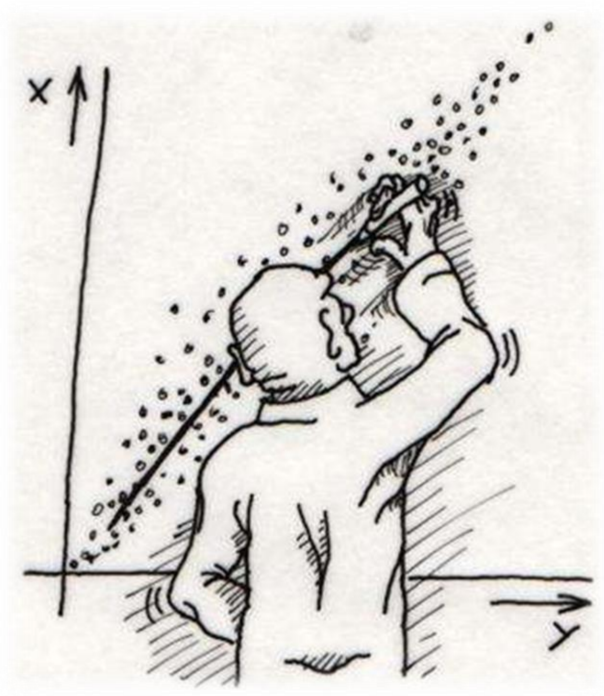
پاسخ‌نامه امتحان میان‌ترم اقتصادسنجی دوره فرعی

محمد مهدی بناساز



بهار ۱۳۹۳

دانشکده مدیریت و اقتصاد
دانشگاه صنعتی شریف

سوال اول

برای حل این سوال معادله اصلی رگرسیون را $y = \beta_0 + \beta_1 x + u \quad (1)$ و معادله‌ای که متغیرهای آن تغییر کرده‌اند را $y' = \beta_0' + \beta_1' x' + u' \quad (2)$ درنظر می‌گیریم. سپس با وارد کردن تغییرات در معادله 2 سعی می‌کنیم معادلات 1 و 2 را شبیه کنیم. سپس با متحد قرار دادن آن‌ها، ارتباط ضرایب قبل و بعد از تغییرات را استخراج می‌نماییم. راه‌حل دیگر آن است که فرمول‌های مربوط به هر تخمین زننده و انحراف معیار آن را بنویسیم و تغییرات مورد نظر را در آن اعمال کنیم.

الف)

$$(1) \quad y = \widehat{\beta_0} + \widehat{\beta_1} x + \hat{u}$$

$$(2) \quad y' = \widehat{\beta_0'} + \widehat{\beta_1'} x' + \widehat{u'}; \{y' = y + c, x' = x, \widehat{u'} = \hat{u}\} \rightarrow y + c = \widehat{\beta_0}' + \widehat{\beta_1}' x + u$$

$$\rightarrow \; y = (\widehat{\beta_0'} - c) + \widehat{\beta_1}' x + u \, .$$

$$(1) \cong (2) \rightarrow y = \widehat{\beta_0} + \widehat{\beta_1} x + u \; \cong \; y = (\widehat{\beta_0'} - c) + \widehat{\beta_1}' x + u$$

$$\widehat{\beta_0} = (\widehat{\beta_0'} - c) \;\rightarrow\; \boldsymbol{\beta_0'} = \boldsymbol{\beta_0} + \boldsymbol{c}. \quad std(\widehat{\beta_0'}) = std(\widehat{\beta_0} + c) \rightarrow \boldsymbol{std(\widehat{\beta_0'})} = \boldsymbol{std(\widehat{\beta_0})}.$$

$$\boldsymbol{\beta_1} = \boldsymbol{\beta_1'}. \quad \boldsymbol{std\left(\widehat{\beta_1'}\right)} = \boldsymbol{std\left(\widehat{\beta_1}\right)}.$$

ب)

$$(1) \quad y = \widehat{\beta_0} + \widehat{\beta_1} x + \hat{u}$$

$$(2) \quad y' = \widehat{\beta_0'} + \widehat{\beta_1'} x' + \widehat{u'}; \{y' = y, x' = x + c, \widehat{u'} = \hat{u}\} \rightarrow y = \widehat{\beta_0}' + \widehat{\beta_1}'(x + c) + u$$

$$\rightarrow \; y = (\widehat{\beta_0'} + \widehat{\beta_1}' c) + \widehat{\beta_1}' x + u \, .$$

$$(1) \cong (2) \rightarrow y = \widehat{\beta_0} + \widehat{\beta_1} x + u \; \cong \; y = (\widehat{\beta_0'} + \widehat{\beta_1}' c) + \widehat{\beta_1}' x + u$$

$$\boldsymbol{\beta_1} = \boldsymbol{\beta_1'}. \quad \boldsymbol{std\left(\widehat{\beta_1'}\right)} = \boldsymbol{std\left(\widehat{\beta_1}\right)}.$$

$$\widehat{\beta_0} = (\widehat{\beta_0'} + \widehat{\beta_1}' c) \;\rightarrow\; \widehat{\beta_0'} = \widehat{\beta_0} - \widehat{\beta_1}' c \;\rightarrow\; \boldsymbol{\widehat{\beta_0'}} = \boldsymbol{\widehat{\beta_0}} - \boldsymbol{\widehat{\beta_1} c}.$$

$$std\left(\widehat{\beta_0'}\right) = std\left(\widehat{\beta_0} - \widehat{\beta_1} c\right) \rightarrow \boldsymbol{std\left(\widehat{\beta_0'}\right)} = \sqrt{\boldsymbol{var\left(\widehat{\beta_0}\right) + c^2 var\left(\widehat{\beta_1}\right) + 2c.cov\left(\widehat{\beta_0}, \widehat{\beta_1}\right)}}\,.$$

ت)

$$(1) \quad y = \widehat{\beta_0} + \widehat{\beta_1}x + \hat{u}$$

$$(2) \quad y' = \widehat{\beta_0'} + \widehat{\beta_1'}x' + \widehat{u'}; \{y' = y, x' = x \times c, \widehat{u'} = \hat{u}\} \to y = \widehat{\beta_0}' + \widehat{\beta_1}'(x \times c) + u$$

$$\to \; y = \widehat{\beta_0'} + c\widehat{\beta_1}'x + u\,.$$

$$(1) \cong (2) \to y = \widehat{\beta_0} + \widehat{\beta_1}x + u \;\cong\; y = \widehat{\beta_0'} + c\widehat{\beta_1}'x + u$$

$$\beta_1 = c\beta_1' \;\to \boldsymbol{\beta_1'} \;= \frac{\boldsymbol{\beta_1}}{\boldsymbol{c}}. \;\; std\left(\widehat{\beta_1'}\right) = std\left(\widehat{\beta_1}/c\right) \;\to \boldsymbol{std}\left(\widehat{\boldsymbol{\beta_1'}}\right) = \frac{\mathbf{1}}{\boldsymbol{c}}\boldsymbol{std}\left(\widehat{\boldsymbol{\beta_1}}\right).$$

$$\widehat{\boldsymbol{\beta_0}} = \widehat{\boldsymbol{\beta_0'}}. \;\; \boldsymbol{std}\left(\widehat{\boldsymbol{\beta_0'}}\right) = \boldsymbol{std}\left(\widehat{\boldsymbol{\beta_0}}\right).$$

سوال دوم

از جدول رگرسیون ۱ ما می‌توانیم ضرایب تخمین زده‌شده را به صورت معادله زیر خلاصه کنیم:

$$\widehat{colgpa} = 1.241365 \; - 0.0568543\, hsize + 0.0046754\, hsize^2 - 0.0132126\, hsperc + 0.0016464 sat + 0.1548814 female + 0.1693064 athlete$$

الف)

$$\frac{\partial \widehat{colgpa}}{\partial hsize} = -\,0.0568543 \; + 2 \;\times 0.0046754\, hsize$$

$$\frac{\partial \widehat{colgpa}}{\partial hsize} |hsize = 5| = -\,0.0568543 \; + 10 \;\times 0.0046754 = \; \mathbf{-0.0101003}$$

$$\frac{\partial \widehat{colgpa}}{\partial hsize} = \; \widehat{\beta_1} + 2\,\widehat{\beta_2} \;\; hsize$$

$$var\left(\frac{\partial \widehat{colgpa}}{\partial hsize}\right) = var(\widehat{\beta_1} + 2\,\widehat{\beta_2}\ hsize)$$
$$= var(\widehat{\beta_1}) + 4hsize^2 . var(\widehat{\beta_2}) + 4hsize. cov(\widehat{\beta_1},\widehat{\beta_2})$$
$$= 0.0163513^2 + 4 \times 25 \times 0.0022494^2 - 4 \times 5 \times 0.00003503$$
$$= 0.00026737 + 0.0005980 - 0.0007006 = 0.00016477$$

$$std\left(\frac{\partial \widehat{colgpa}}{\partial hsize}\right) = \sqrt{0.00016477} = \mathbf{0.0128}$$

ب)

$$\frac{\partial \widehat{colgpa}}{\partial hsize} = 0 \ \rightarrow -0.0568543 \ + 2 \ \times 0.0046754\, hsize = 0 \ \rightarrow hsize$$
$$= \frac{0.0568543}{2 \times 0.0046754} = \mathbf{6.0801}$$

ت) برای فهم بهتر ضرایب متغیرهای مجازی، ضابطه مربوطه را در معادله رگرسیون جاگذاری می‌کنیم:

$$\widehat{colgpa} = \widehat{\beta_0} + \widehat{\beta_1}\ hsize + \widehat{\beta_2}\ hsize^2 + \widehat{\beta_3}\ hsperc \ + \widehat{\beta_4} sat + \widehat{\beta_5} female + \widehat{\beta_6} athlete$$

$$\begin{cases} athlete = 1 \ \rightarrow \ \widehat{colgpa}_1 = \widehat{\beta_0} + \widehat{\beta_1}\ hsize + \ \widehat{\beta_2}\ hsize^2 + \widehat{\beta_3}\ hsperc \ + \widehat{\beta_4} sat + \widehat{\beta_5} female + \widehat{\beta_6} \\ \\ athlete = 0 \ \rightarrow \ \widehat{colgpa}_0 = \widehat{\beta_0} + \widehat{\beta_1}\ hsize + \ \widehat{\beta_2}\ hsize^2 + \widehat{\beta_3}\ hsperc \ + \widehat{\beta_4} sat + \widehat{\beta_5} female \end{cases} \rightarrow$$

$$\widehat{\beta_6} = \widehat{colgpa}_1 - \widehat{colgpa}_0 = 0.1693064$$

ضریب موردنظر بیان می‌کند که معدل انتظاری دانش‌آموزی که عضو تیم ورزشی است از دانش‌آموزی که عضو تیم ورزشی نیست، ۰,۱۷ نمره بالاتر است. با توجه به آماره t ضریب تخمین زده شده برای athlete که ۴ است، این ضریب اختلاف معنی‌داری از صفر دارد.

آزمون فرضیه زیر را درنظر می‌گیریم:

$$\begin{cases} H_0: \beta_6 = 0.2 \\ H_1: \beta_6 \neq 0.2 \end{cases} .$$

آماره آزمون آزمون بالا به شکل زیر است:

$$t-stat=\frac{0.17-0.2}{0.0423}=-0.71$$

عدد آماره آزمون با توجه به مقدار ۱٫۹۶ برای مقدار بحرانی توزیع t، در محدوده قبول قرار می‌گیرد و فرضیه $H_0: \beta_6 = 0.2$ **رد نمی‌شود**.

ث) از رگرسیون شماره ۳ با توجه به ضریب شیب (۱۲۲٫۰۳۳۱-)و معنی‌داری شدید آن مشخص است که متغیرهای sat و athlete، همبستگی منفی با هم دارند. حال فرض کنید که ما متغیر sat را از مدل حذف کرده‌ایم. با ثابت فرض کردن سایر متغیرهایی که در معادله رگرسیون ۲ حضور دارند، متغیر athlete اگر به جای ۰، ۱ باشد - یک واحد افزایش یابد، با ضریب مربوط به خود باعث افزایش colgpa می‌شود. نکته این‌جا است که متغیر sat در معادله رگرسیون ۲ نبوده‌است و کنترل نشده‌است. پس با افزایش athlete، sat کاهش می‌یابد و باعث کاهش colgpa می‌شود. در واقع تغییر sat، و اثر آن، اثر متغیر athlete را بر colgpa خنثی می‌کند. و از دید ما که رگرسیون ۲ را انجام داده‌ایم، ضریب athlete از نظر آماری بی‌معنی است.

ج)

$$colgpa=\widehat{\alpha_0}+\widehat{\alpha_1}\,hsize+\widehat{\alpha_2}\,hsize^2+\widehat{\alpha_3}hsperc+\widehat{\alpha_4}sat+\boldsymbol{\widehat{\alpha_5}female.athlete}+\hat{u}.$$

ضریب $\widehat{\alpha_5}$ پاسخ سوال ماست.

ح)

$$colgpa=\widehat{\alpha_0}+\widehat{\alpha_1}\,hsize+\widehat{\alpha_2}\,hsize^2+\widehat{\alpha_3}hsperc+\widehat{\alpha_4}sat+\widehat{\alpha_5}female+\widehat{\alpha_6}athlete+\widehat{\alpha_7}female.athlete+\hat{u}.$$

خ)

$$\widehat{lcolgpa}=0.4787139-0.0248034\,hsize+0.001954\,hsize^2-0.0056961\,hsperc+0.0005757sat+0.0639424female+0.0937825athlete-0.1428296black$$

$$\frac{\%\Delta colgpa}{\Delta hsize}=\left(\frac{\partial \widehat{lcolgpa}}{\partial hsize}\right)\times 100=100\times(-0.0248034+2\times 0.001954\,hsize)$$

$$\frac{\%\Delta colgpa}{\Delta hsize}[hsize=5]=100\times(-0.0248034+2\times 0.001954\times 5)=\boldsymbol{-0.53}\%$$

$$\frac{\partial \widehat{lcolgpa}}{\partial hsize}=\widehat{\beta_1}+2\widehat{\beta_2}\,hsize\rightarrow var\left(\frac{\partial \widehat{lcolgpa}}{\partial hsize}\right)=var\left(\widehat{\beta_1}+2\widehat{\beta_2}\,hsize\right)$$
$$=var\left(\widehat{\beta_1}\right)+4hsize^2.var\left(\widehat{\beta_2}\right)=0.0074566^2+4\times 25\times 0.001026^2$$
$$=0.0001609$$

$$std\left(\frac{\partial \widehat{lcolgpa}}{\partial hsize}\right) = \sqrt{0.0001609} = \mathbf{0.0126}$$

د) معدل مردان به طور میانگین ، ۶,۴ درصد کم‌تر از زنان بوده‌است. که انحراف معیار این اثر طبق رگرسیون ۴ ، ۰,۰۰۸۲ است.

ذ)

$$F \equiv \frac{\frac{RSS_r - RSS_{ur}}{k-1}}{\frac{RSS_{ur}}{n-(k)}} = \frac{\frac{350.49105 - 263.535913}{7}}{\frac{263.535913}{4135-8}} = \frac{12.42}{0.06} = 207$$

با توجه به عدد ۲۰۷ ضرایب کمکی به توضیح نمره دانش‌آموزان کمک می‌کنند. و در کل **رگرسیون معنی دار** است.

سوال سوم

معادله‌های بسیار مهم که در این قسمت بسیار از آن‌ها استفاده می‌کنیم. این فرمول‌های در **ضمیمه A** کتاب درسی آمده است!

$$\sum_{i=1}^{n}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2 = \sum_{i=1}^{n}\boldsymbol{x_i^2} - \boldsymbol{n}(\overline{\boldsymbol{x}})^2$$

$$\sum_{i=1}^{n}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})(\boldsymbol{y_i} - \overline{\boldsymbol{y}}) = \sum_{i=1}^{n}\boldsymbol{x_i y_i} - \boldsymbol{n}(\overline{\boldsymbol{x}}.\overline{\boldsymbol{y}})$$

الف)

$$\widehat{\beta_1} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}x_i y_i - n(\bar{x}.\bar{y})}{\sum_{i=1}^{n}x_i^2 - n(\bar{x})^2} = \frac{41140 - 526(5.89 \times 12.56)}{87040 - 526(12.56^2)}$$
$$= \frac{2227.3616}{4061.6064} = 0.55$$

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x} = 5.89 - 0.55 \times 12.56 = -1.018$$

ب)

$$std(\widehat{\beta_1}) = \sqrt{\frac{\frac{1}{n-k}\sum_{i=1}^{n}\widehat{u_i}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} = \sqrt{\frac{\frac{1}{n-k}\left(\sum_{i=1}^{n}(y_i-\bar{y})^2 - \widehat{\beta_1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})\right)}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} =$$
$$\sqrt{\frac{\frac{1}{524}\times(7197.9554-0.55\times2227.3616)}{4061.6064}} = \sqrt{\mathbf{0.0028}} = \mathbf{0.053}$$

$$std(\widehat{\beta_0}) = std(\bar{y} - \widehat{\beta_1}\bar{x}) = \bar{x}std(\widehat{\beta_1}) = 12.56 \times 0.053 = \mathbf{0.66}$$

ت) $\widehat{\beta_1}$ اختلاف معنی‌داری از صفر دارد؛

$$t.stat(\widehat{\beta_1}) = \frac{\widehat{\beta_1}}{std(\widehat{\beta_1})} = \frac{0.55}{0.053} = 10.38$$

ث)

$$E(\widehat{\beta_1}) - \beta_1 = \beta_2\frac{\sum_{i=1}^{n}(x_i-\bar{x})(z_i-\bar{z})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \beta_2\frac{\sum_{i=1}^{n}x_iz_i - n(\bar{x}.\bar{z})}{\sum_{i=1}^{n}x_i^2 - n(\bar{x})^2} = \cdots$$

سوال چهارم

الف) متغیرهای زیر را تعریف می‌کنیم:

M = Mean Contingency Fee, LR = Ln Recovery

معادله رگرسیون زیر را درنظر بگیرید؛ نقطه شکستگی را LR_0 می‌نامیم.

$$(1) \qquad M = a + b.LR + c.D + d.D \times LR + u$$

$$D = \begin{cases} 0, & if\ LR < LR_0 \\ 1, & if\ LR \geq LR_0 \end{cases}$$

معادله رگرسیون قبل از شکستگی:

$$M = a + b.LR + u$$

معادله رگرسیون بعد از شکستگی:

$$M = a + b.LR + c.D + d.D \times LR + u$$

همانطور که در نمودار مشخص است، در نقطه شکستگی مقادیر M برابر است. پس داریم:

$$a + b.LR_0 = (a + c) + (b + d).LR_0 \ \rightarrow c + d.LR_0 = 0$$

عبارت $c + d.LR_0 = 0$ را از سمت راست معادله رگرسیون ۱ کم می‌کنیم تا به معادله زیر برسیم.

$$\boldsymbol{M = a + b.LR + d.D(LR - LR_0) + u}$$

ضریب a مربوط به شیب منحنی قبل از شکستگی و ضریب d شیب منحنی بعد از شکستگی را نمایندگی می‌کنند.

ب) از آزمون‌های cusum, cusumq استفاده می‌کنیم.

ج) بعد از مشخص شدن نقطه شکستگی کافی است در معادله رگرسیونی که برای الف بدست آوردیم معنی‌داری تخمین ضریب d آزمون شود.

پایان