فصل هشتم

تحلیل رگرسیون چند متغیره (مرکب):

مسأله استنتاج آماری

فهرست

چکیده:

در این فصل به بررسی مفاهیم تخمین فاصله‌ای و آزمون فرضیه‌ای مدلهای رگرسیون می‌پردازیم و به ویژگیهایی که مختص چنین مدلهایی است اشاره می‌نمائیم.

1-8- بررسی مجدد مفهوم فرض نرمال بودن (داشتن توزیع نرمال)

$u_i \sim N(0,\sigma^2)$

فروض

$$\frac{(N-3)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-3}$$

$\hat{\beta}_3, \hat{\beta}_2, \hat{\beta}_1$ دارای توزیع نرمال می‌باشند.

آماره t مربوط به پارامترها

$$t = \frac{\hat{\beta}_1 - \beta_1}{Se\,(\hat{\beta}_1)}$$

$$t = \frac{\hat{\beta}_2 - \beta_2}{Se\,(\hat{\beta}_2)}$$

$$t = \frac{\hat{\beta}_3 - \beta_3}{Se\,(\hat{\beta}_3)}$$

8-2- مثال 1-8: رابطه مصرف شخصی و درآمد قابل تصرف شخصی در ایالات متحده طی دوره 70-1956

$$E(Y \mid X_2, X_3) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i}$$

Y: هزینه مصرف شخصی (PCE)

X_2: درآمد شخصی قابل تصرف (PDI)

X_3: زمان

$$\hat{Y}_i = 53.1603 + 0.7266 X_{2i} + 2.7363 X_{3i}$$
$$(13.0261) \quad (0.0487) \quad (0.8486)$$
$$t: (4.0811) \quad (14.9060) \quad (3.2246)$$

$$R^2 = 0.9988$$
$$\bar{R}^2 = 0.9988$$
$$df = 12$$
$$F_{2,12} = 5128.88$$

_ دلایل وارد کردن متغیر زمان (روند):

1) بررسی رفتار متغیر وابسته در طول زمان

2) به عنوان شاخص متغیرهایی که بر Y تأثیر می‌گذارند اما مستقیماً قابل مشاهده نیست. (تکنولوژی)

3-8- آزمون فرضیه درباره ضرائب جزئی رگرسیون:

$$\begin{cases} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0 \end{cases}$$

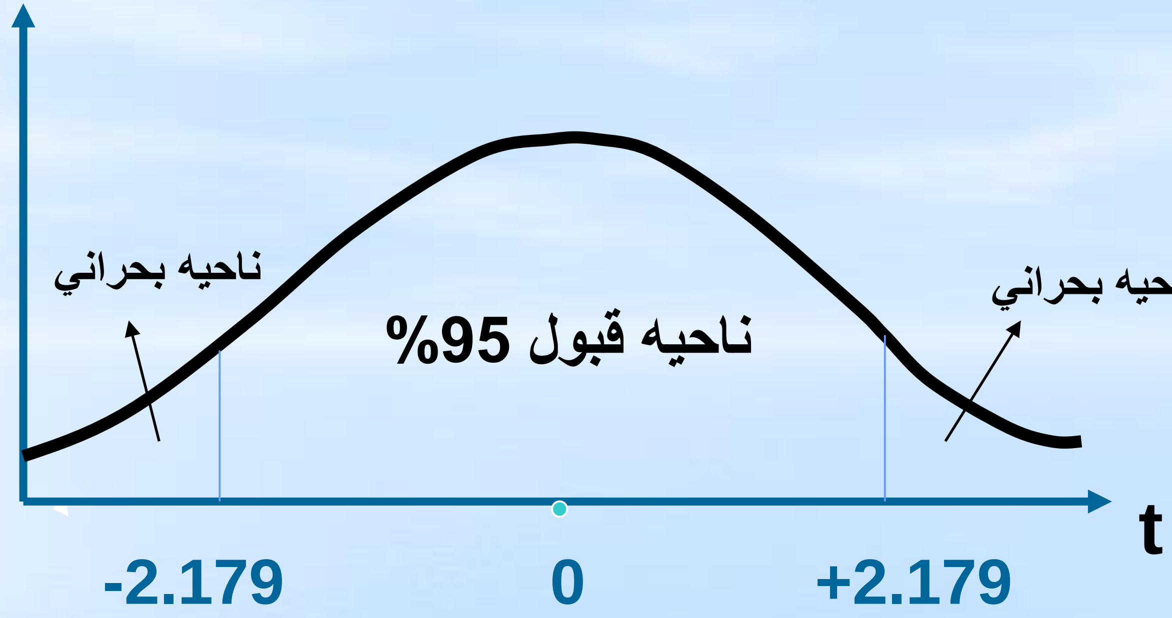
$$t = \frac{\hat{\beta}_2}{Se(\hat{\beta}_2)}$$

با توجه به مثال قبل:

$$t = \frac{0.7266}{0.0487} = 14.9060$$

t جدول : $t_{\frac{\alpha}{2}, 12} = 2.179 \Rightarrow$ t محاسباتی > t جدول

فرضیه عدم رد می‌شود و $\hat{\beta}_2$ از نظر آماری معنادار است.



جدول 3–8

(فاصله اطمینان 95٪ برای t با درجه آزادی 12)

$$\hat{\beta}_2 - t_{\frac{\alpha}{2}}\ Se\,(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{\frac{\alpha}{2}}\ Se\,(\hat{\beta}_2)$$

$$.7266 - 2.179 \times .0487 \le \beta_2 \le .7266 + 2.179 \times .0487$$

$$.6205 \le \beta_2 \le .8327$$

8-4- آزمون معنی‌دار بودن کلی رگرسیون نمونه

$H_0: \quad \boldsymbol{\beta_2 = \beta_3 = 0}$

H_1: حداقل یکی مخالف صفر

آزمون پیاپی از فرضیه‌های انفرادی (تکی) با آزمون مرکب همان فرضیه‌ها یکسان نیست. نخستین استدلال برای این امر آن است که در یک آزمون مرکب فرضیه های مختلف ؛ هر فرضیه تکی، از اطلاعات بکار رفته در سایر فرضیه‌ها متأثر می‌شود.

8-5- روش آنالیز واریانس برای آزمون معنادار بودن کلی رگرسیون چندمتغیره (مرکب) مشاهده شده: آزمون F

$$\sum y_i^2 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \sum e_i^2$$

$$TSS = ESS + RSS$$

منبع تغییرات	SS	df	MSS
بعلت رگرسیون (ESS)	$\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$	2	$\frac{SS}{df} = \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{k-1=2}$
بعلت باقیمانده‌ها (RSS)	$\sum e_i^2$	N-3	$\hat{\sigma}^2 = \frac{\sum e_i^2}{N-k=N-3}$
جمع	$\sum y_i^2$	N-1	

آزمون F

$$F = \frac{(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i})/2}{\sum e_i^2/(N-3)} = \frac{ESS/df}{RSS/df}$$

if $F_{\text{محاسباتی}} > F_{\text{جدول}} \Rightarrow$ فرضیه عدم رد شده و معادله رگرسیونی معنادار خواهد بود

برای مثال ذکر شده: جدول 3-8

$$F = \frac{32982.5502}{6.4308} = 5128.8781 > F_{0.05\ \text{و}\ (2\ \text{و}12)} \Rightarrow$$ رگرسیون معنادار است

نکته

$$if \quad \beta_2 = \beta_3 = 0 \;\Rightarrow E\left[\frac{\sum e_i^2}{N-3}\right] = E(\hat{\sigma}^2) = \frac{E(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i})}{2} = \sigma^2$$

به عبارتی تنها منبع تغییرات در Y بعلت عوامل تصادفی نشان داده شده با u_i می‌باشد.

– آزمون معنی‌دار بودن کلی رگرسیون مرکب: آزمون F

مدل رگرسیون k متغیره:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

$H_0:\ \beta_2 = \beta_3 = \ldots = \beta_K = 0$

H_1: حداقل يكي از βها مخالف صفر باشد

$$F = \frac{ESS\ /\ df}{RSS\ /\ df} = \frac{ESS\ /(K-1)}{RSS\ /(N-K)} \quad if\ \ F > F_{\alpha(K-1,N-K)} \Rightarrow$$ H_0 رد می‌شود

8- یک رابطه مهم بین $R2$ و F در مدل k متغیره:

$$F = \frac{ESS/(k-1)}{RSS/(N-k)} = \frac{N-k}{k-1} \cdot \frac{ESS}{RSS}$$

$$= \frac{N-k}{k-1} \cdot \frac{ESS}{TSS-ESS} = \frac{N-k}{k-1} \cdot \frac{ESS/TSS}{1-(ESS/TSS)}$$

$$= \frac{N-k}{k-1} \cdot \frac{R^2}{1-R^2} = \frac{R^2/(k-1)}{(1-R^2)/(N-k)}$$

توجه:

F تابع آزمون معنی‌دار بودن R^2 نیز می‌باشد.

$$if \begin{cases} R^2 = 0 \Rightarrow F = 0 \\ R^2 \uparrow \Rightarrow F \uparrow \\ R^2 \to 1 \Rightarrow F \to \infty \end{cases}$$

جدول 4-8- جدول AOV بر حسب R^2

منبع تغییرات	SS	df	MSS
به علت رگرسیون	$R^2(\sum y_i^2)$	2	$R^2(\sum y_i^2)/2$
به علت باقیمانده‌ها	$(1-R^2)(\sum y_i^2)$	$N-3$	$(1-R^2)(\sum y_i^2)/N-3$
جمع	$\sum y_i^2$	$N-1$	

مزیت بیان آزمون F برحسب R^2: دانستن مقدار R^2 برای محاسبه F کافی است.

آزمون معنيداربودن کلي رگرسيون مرکب برحسب R^2

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_K X_{ki} + u_i$$

$$\begin{cases} H_0 : \beta_2 = \beta_3 = \ldots = \beta_K = 0 \\ H_1 : \end{cases}$$

حداقل يك ضريب زاويه مخالف صفر وجود دارد

or

$$\begin{cases} H_0 : R^2 = 0 \\ H_1 : R^2 \neq 0 \end{cases}$$

$$F = \frac{R^2/(K-1)}{(1-R^2)/(N-K)}$$

$$if \quad F > F_{\alpha,(K-1,N-K)} \quad \Rightarrow$$

فرضيه H_0 رد ميشود

7-8- اثر (سهم) «نموی» یا «نهایی» یک متغیر توضیحی

آیا افزودن متغیر به مدل، ESS (و بدین ترتیب R^2) را «به طور معنی دار» نسبت به RSS افزایش می‌دهد؟

برای پاسخ به مثال درآمد – مصرف رجوع می‌شود.

Y: هزینه مصرف شخصی

X_2 : درآمـد قابـل تصـرف شخصـی

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_{12} \; X_{2i}$$

$$= 12.762 + 0.8812 \; X_{2i}$$

$$(4.6818) \quad (0.0114)$$

$$t: \; (2.7259) \quad (77.2982)$$

$$r^2 = 0.9978$$

$$\bar{r}^2 = 0.9978$$

$$\begin{cases} H_0 : \beta_{12} = 0 \\ \\ H_1 : \beta_{12} \neq 0 \end{cases}$$

$$t > t_{.05} , t_{.01} \Rightarrow$$

فرضیه H_0 رد می‌شود و X_2 به طور معنی‌داری بر Y اثر می‌گذارد

همچنین :

$$F = \frac{65898.235}{11.08} = 5947.494 > F_{\alpha,(1,13)} \Rightarrow$$

H_0 رد می‌شود

حال اثر افزایش متغیر X_3 به مدل، «اثر نموی» را بررسی می‌کنیم.

جدول AOV برای ارزیابی اثر (سهم) نموی یک متغیر (یا متغیرهای) اضافی

منبع تغییرات	SS	df	MSS
ESS به علت X_2 به تنهایی	$Q_1 = \hat{\beta}^2_{12} \sum x_2^2$	1	$Q_1 \ / 1$
ESS به علت افزودن X_3	$Q_2 = Q_3 - Q_1$	1	$Q_2 \ / 1$
ESS به علت هر دو X_2 و X_3	$Q_3 = \hat{\beta}_2 \sum y_i \ x_{2i} + \hat{\beta}_3 \sum y_i \ x_{3i}$	2	$Q_3 \ / 2$
RSS	$Q_4 = Q_5 - Q_3$	N-3	$Q_4 \ / \ N-3$
جمع	$Q_5 = \sum y_i^2$	N-1	

$$\boldsymbol{F = \frac{Q_2 / df}{Q_4 / df}} = \frac{(\text{ESS جدید} - \text{ESS قدیم}) / \text{تعداد متغیرهای توضیحی جدید}}{\text{RSS جدید} / df\ (= N - \text{تعداد پارامترها در مدل جدید})}$$

ESS جدید مربوط به مدل زیر است:

$$\boldsymbol{\hat{Y}_i = 53.1603 + 0.7266 X_{2i} + 2.7363 X_{3i}}$$

$$(13.0261) \quad (0.0487) \quad (0.8486) \qquad df = 12$$

$$t = \quad (4.0811) \quad (14.9060) \quad (3.2246)$$

$$R^2 = 0.9988$$

X_3: متغیر روند

$$\bar{R}^2 = 0.9988 \qquad F = \frac{66.865 / 1}{77.1693 / 12} = 10.3973 > F_{0.01,(1,12)}$$

$\Leftarrow$ فرضیه H_0 مبنی بر بی‌معنی بودن β_3 رد شده و در نتیجه افزودن X_3 به مدل بطور معنی‌دار ESS و بالطبع R^2 را افزایش می‌دهد.

از سوی دیگر:

$$F = \frac{(R^2_{\text{جدید}} - R^2_{\text{قدیم}})/df}{(1 - R^2_{\text{جدید}})/df}$$

$$= \frac{(R^2_{\text{جدید}} - R^2_{\text{قدیم}})/(\text{تعداد متغیرهای توضیحی جدید})}{(1 - R^2_{\text{جدید}})/(N)\ (N = \text{تعداد پارامترها در مدل جدید})}$$

در مثال مذکور:

$$F = \frac{(0.9988 - 0.9978)/1}{(1 - 0.9988)/12} = 10 > F_{0.01,(1,12)}$$

در نتیجه β_3 باید به مدل افزوده شود.

چه وقت متغیر جدیدی را اضافه کنیم؟

?

محققان مدلی که بالاترین R^2 تعدیل شده را ارائه نماید، انتخاب مـی کننـد .
به عبارتی زمانی که افزودن متغیر در مدل، R^2 را بیفزاید، اما از لحاظ آماری RSS را بطور معنی‌دار کاهش ندهد، به مدل افزوده نمی‌شود.

اگرقدرمطلق t مربوط به متغیـر جدیـد $> 1 \longleftarrow$ $\bar{R}^2$ افـزایش مـی یابـد
پس متغیر جدید وارد می شود .

8-8- آزمون برابری دو ضریب یک رگرسیون

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \ldots + \beta_K X_{Ki} + u_i$$

$$\begin{cases} H_0 : \beta_3 = \beta_4 & \text{یا} \quad (\beta_3 - \beta_4) = 0 \\ H_1 : \beta_3 \neq \beta_4 & \text{یا} \quad (\beta_3 - \beta_4) \neq 0 \end{cases}$$

$$t = \frac{(\hat{\beta}_3 - \hat{\beta}_4) - (\beta_3 - \beta_4)}{Se(\hat{\beta}_3 - \hat{\beta}_4)} = \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{\operatorname{var}(\hat{\beta}_3) + \operatorname{var}(\hat{\beta}_4) - 2\operatorname{cov}(\hat{\beta}_3, \hat{\beta}_4)}}$$

مراحل آزمون:

1.تخمین $\hat{\beta}_3$ و $\hat{\beta}_4$

2.محاسبه واریانس و کوواریانس $\hat{\beta}_3$ و $\hat{\beta}_4$

3.محاسبه نسبت t

4.فرضیه عدم رد می‌شود $\quad if \quad t > t_{\alpha\,,\,N-K} \Rightarrow$

8-9- روش حداقل مربعات مقید قیود تساوی خطی

برای بررسی برقراری یک قید دو روش وجود دارد که با در نظر گرفتن تابع تولید کاب- داگلاس توضیح داده می‌شود:

Y : محصول

X_2: نیروی کار

X_3 : نهاده سرمایه

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \longrightarrow \ln Y_i = \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

$$\begin{cases} H_0 : \beta_2 + \beta_3 = 1 \rightarrow \\ H_1 : \beta_2 + \beta_3 \neq 1 \end{cases}$$

قید= بازده ثابت نسبت به مقیاس

الف- روش آزمون t

$$t=\frac{(\hat{\beta}_2+\hat{\beta}_3)-(\hat{\beta}_2+\hat{\beta}_3)}{Se(\hat{\beta}_2+\hat{\beta}_3)}=\frac{(\hat{\beta}_2+\hat{\beta}_3)-1}{\sqrt{\mathrm{var}(\hat{\beta}_2)+\mathrm{var}(\hat{\beta}_3)+2\mathrm{cov}(\hat{\beta}_2,\hat{\beta}_3)}}$$

$if \;\; t > t_{\text{بحرانی}} \Rightarrow$ فرضیه H_0 مبنی بر ثابت بودن بازده نسبت به مقیاس رد می‌شود

ب- روش آزمون F حداقل مربعات

$$\beta_2 = 1 - \beta_3 \qquad (1)$$

$$\beta_3 = 1 - \beta_2 \qquad (2)$$

در ایــن روش از ابتــدا قیــد وارد مــی‌شــود و معنــی‌دار بــودن مدل مقید آزمون می‌شود.

$$(1) \Rightarrow \ln Y_i = \beta_0 + (1-\beta_3)\ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

$$= \beta_0 + \ln X_{2i} + \beta_3(\ln X_{3i} - \ln X_{2i}) + u_i$$

یا $$(\ln Y_i - \ln X_{2i}) = \beta_0 + \beta_3(\ln X_{3i} - \ln X_{2i}) + u_i$$

ویا $$\ln(Y_i / X_{2i}) = \beta_0 + \beta_3 \ln(X_{3i} / X_{2i}) + u_i$$ ➡ رگرسیون مقید

آیا محدودیت اعمال شده برقرار است؟

$$F=\frac{(RSS_R-RSS_{UR})/m}{RSS_{UR}/(N-k)}=\frac{(\sum e_R^2-\sum e_{UR}^2)/m}{\sum e_{UR}^2/(N-k)}$$

رگرسیون غیرمقید $RSS_{UR}=\sum e_{UR}^2$

رگرسیون مقید $RSS_R=\sum e_R^2$

ویا $$F=\frac{(R_{UR}^2-R_R^2)/m}{(1-R_{UR}^2)/(N-k)}$$

$m=$ تعداد محدودیتهای خطی

$k=$ تعداد پارامترها در رگرسیون غیرمقید

$N=$ تعداد مشاهدات

$$\text{if}\quad F>F_{\alpha,(m,N-k)}\ \Rightarrow$$ فرضیه H_0 مبنی بر برقراری قید رد می‌شود.

$$R_{UR}^2\geq R_R^2$$

$$\sum e_{UR}^2\leq\sum e_R^2$$

توجه

10-8- آزمون عمومی F:

$$Y_i = \beta_1 + \beta_2 X_{2i} \ \beta_3 \ X_{3i} + \cdots + \beta_k \ X_{ki} + u_i$$

برای آزمون فرضیه‌هایی مانند:

$$H_0 : \beta_2 = \beta_3$$

$$H_0 : \beta_3 + \beta_4 + \beta_5 = 3$$

$$H_0 : \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$

مراحل زیر را انجام می‌دهیم:

1. برازش مدلهای بدون محدودیت (بزرگتر، غیر مقید UR) و مدلهای مقید (با محدودیت و مدل کوچکتر R)
2. محاسبه R^2_{UR} و R^2_R
3. محاسبه درجات آزادی: $(N-k)$ درجه آزادی مدل غیرمقید و (m) درجه آزادی مدل مقید
4. محاسبه F
5. قاعده تصمیم‌گیری: $\text{if} \ \ F_{\text{محاسباتي}} > F_{\alpha,(m,N-k)} \Rightarrow$

فرضیه عدم رد می‌شود

8-11- پیش‌بینی با استفاده از رگرسیون چند متغیره (مرکب)

مثال مصرف- درآمد را در نظر بگیرید:

$$\hat{Y}_i = 53.1603 + 0.7266\, X_{2i} + 2.7363\, X_{3i}$$
$$(13.0261)\quad (0.0487)\quad (0.4886)\qquad R^2 = 0.9988$$

الف- پیش‌بینی تکی: (پیش‌بینی تکی مقدار Y به شرط مقدار متغیر توضیحی $X=X_0$)

$$if \begin{cases} X_2 = 567 \\ X_3 = 16 \end{cases} \Rightarrow (\hat{Y}_{1971} | X_2 = 567, X_3 = 16) =$$
$$= 53.1603 + 0.7266(567) + 2.7363$$
$$= 508.9297$$

پیش‌بینی تکی:

$$\left[\hat{Y}_{1971} - t_{\alpha/2}\, Se_{(Y_{1971})} \le Y_{1971} \le \hat{Y}_{1971} + t_{\alpha/2}\, Se_{(Y_{1971})}\right]$$

$$508.9297 - 2.179(3.1763) \le Y_{1971} \le 508.9297 + 2.179(3.1763)$$

$$501.9988 \le Y_{1971} \le 515.8412$$

ب- پیش‌بینی میانگین: پیش‌بینی نقطه‌ای روی تابع رگرسیون جامعه اصلی (PRF)

$$\left\lfloor \hat{Y}_{1971} - t_{\alpha/2}\, Se_{(\hat{Y}_{1971})} \leq E(Y_{1971}) \leq \hat{Y}_{1971}\; t_{\alpha/2}\, Se_{(\hat{Y}_{1971})} \right\rfloor$$

$$508.9297 - 2.179(1.9126) \leq E(Y_{1971}) \leq 508.9297 + 2.179(1.9126)$$

$$504.7518 \leq E(Y_{1971}) \leq 513.0808$$

8-12- خلاصه و نتایج :

در این فصل به بررسی موارد زیر پرداختیم:

- آزمون معنی‌دار بودن تکی یک ضریب جزئی رگرسیون و آزمون معنی‌دار بودن کلی یکسان نیست.
- اندازه‌گیری سهمی یک یا چند متغیر توضیحی نسبت به رگرسیون مرکب
- آزمون یکسانی ضرایب در رگرسیون
- وارد کردن محدودیتهای خطی تئوریکی به مدل و آزمون صحت محدودیتها
- آزمون یک یا دو فرضیه درباره رگرسیون مرکب توسط آزمون F


پایان