فصل سوم

مدل‌های چند متغیره

مقدمه

در دنیای واقعی رابطه بین متغیرها به صورت ساده و دو متغیره نمی‌باشد اصولاً متغیرهای بسیاری وجود دارند که در هر معادله رگرسیونی اثرگذار بوده و می‌بایستی در مدل ظاهر شوند. مدل رگرسیونی دو متغیره توانایی تشریح و نشان دادن اثر سایر متغیرهای توضیحی بر متغیر وابسته در مدل را ندارد برای همین منظور از مدل‌های رگرسیونی چند متغیره استفاده می‌شود. در مدل رگرسیون چند متغیره امکان وارد کردن چندین عامل مهم و اثرگذار بر متغیر وابسته وجود دارد.

به طوری که در معادله رگرسیونی $Y = \beta_۰ + \beta_۱ X_۱ + \beta_۲ X_۲ + \mathrm{U}$ ، $\beta_۱$ معیاری برای اندازه گیری تغییرات Y با توجه به $X_۱$ با ثابت در نظر گرفتن سایر عوامل اثرگذار در مدل می‌باشد. همچنین فرض می‌شود که در این مدل $E\,(U|\, X_۱ \ , X_۲\,) = ۰$ می‌باشد و یا به عبارت دیگر برای هر مقداری از $X_۱$ و $X_۲$ در جامعه میانگین عوامل غیرقابل مشاهده صفر می‌باشد. در هر مدل رگرسیونی چند متغیره مدل می‌تواند شامل k+۱ پارامتر نامشخص باشد که شامل عرض از مبداء و ضرایب شیب‌ها می‌باشد و رابطه تبعی بین متغیرهای مستقل و وابسته به صورت درست تصریح شده باشد.

بسط روش OLS به مدل‌های چند متغیره

روش OLS را می‌توان به آسانی به مدل‌هایی که شامل چند متغیر مستقل باشند بسط داد. مدل چند متغیره زیر را در نظر بگیرید (به طوری که K=۲ متغیر مستقل (X) و (Y) متغیر وابسته باشد):

$$Y_t = \beta_۰ + \beta_{t۱} X_{t۱} + \beta_۲ X_{t۲} + u_t$$

با مجذور نمودن طرفین و گرفتن مجموع، داریم:

$$S = \sum \left(Y_t - \hat{\beta}_۰ - \hat{\beta}_۱ X_{t۱} - \hat{\beta}_۲ X_{t۲}\right)^۲ = \sum \hat{u}_t$$

$$\frac{\partial S}{\partial \hat{\beta}_۰} = -۲ \sum \left(Y_t - \hat{\beta}_۰ - \hat{\beta}_۱ X_{t۱} - \hat{\beta}_۲ X_{t۲}\right)^۲ = ۰$$

$$\frac{\partial S}{\partial \hat{\beta}_۱} = -۲ \sum X_{t_۱} \left(Y_t - \hat{\beta}_۰ - \hat{\beta}_۱ X_{t۱} - \hat{\beta}_۲ X_{t۲}\right)^۲ = ۰$$

$$\frac{\partial S}{\partial \hat{\beta}_۲} = -۲ \sum X_{t_۲} \left(Y_t - \hat{\beta}_۰ - \hat{\beta}_۱ X_{t۱} - \hat{\beta}_۲ X_{t۲}\right)^۲ = ۰$$

با استفاده از معادلات فوق داریم:

$$\sum \left(Y_t - \hat{\beta}_۰ - \hat{\beta}_۱ X_{t۱} - \hat{\beta}_۲ X_{t۲}\right) = ۰$$

$$\sum X_{t_۱} \left(Y_t - \hat{\beta}_۰ - \hat{\beta}_۱ X_{t۱} - \hat{\beta}_۲ X_{t۲}\right) = ۰$$

$$\sum X_{t_۲} \left(Y_t - \hat{\beta}_۰ - \hat{\beta}_۱ X_{t۱} - \hat{\beta}_۲ X_{t۲}\right) = ۰$$

$$\sum Y_t = T\,\hat{\beta}_۰ + \hat{\beta}_۱ \sum X_{t۱} + \hat{\beta}_۲ \sum X_{t۲}$$

$$\sum X_{t۱} Y_t = \hat{\beta}_۰ \sum X_{t۱} + \hat{\beta}_۱ \sum X_{t۱}^۲ + \hat{\beta}_۲ \sum X_{t۱} X_{t۲}$$

$$\sum X_{t۲} Y_t = \hat{\beta}_۰ \sum X_{t۲} + \hat{\beta}_۱ \sum X_{t۱} X_{t۲} + \hat{\beta}_۲ \sum X_{t۲}^۲$$

$$\begin{bmatrix} \sum Y_t \\ \sum X_{t۱} Y_t \\ \sum X_{t۲} Y_t \end{bmatrix} \begin{bmatrix} T & \sum X_{t۱} & \sum X_{t۲} \\ \sum X_{t۱} & \sum X_{t۱}^۲ & \sum X_{t۱} X_{t۲} \\ \sum X_{t۲} & \sum X_{t۱} X_{t۲} & \sum X_{t۲}^۲ \end{bmatrix} \begin{bmatrix} \hat{\beta}_۰ \\ \hat{\beta}_۱ \\ \hat{\beta}_۲ \end{bmatrix}$$

سه معادله سه مجهولی فوق را می‌توان مجدداً با استفاده از قاعده کرامر حل کرد. (مجهولات $\hat{\beta}_۰$ تا $\hat{\beta}_۲$ می‌باشند).

اجازه دهید مشاهده t ام از تعداد T مشاهده را برای K متغیر به صورت زیر در نظر بگیریم:

$$Y_t = \beta_۱ X_{t۱} + \beta_۲ X_{t۲} + \cdots + \beta_k X_{kt} + u_t$$

$$Y = X\beta + u$$

شکل بسته ماتریسی آن خواهد بود:

$$\begin{bmatrix} Y_{۱۱} \\ Y_{۲۱} \\ \vdots \\ Y_{T۱} \end{bmatrix} \begin{bmatrix} X_{۱۱} & X_{۱۲} & \cdots & X_{۱\mathrm{k}} \\ X_{۲۱} & X_{۲۲} & \cdots & X_{۲K} \\ \vdots & \vdots & & \vdots \\ X_{T۱} & X_{T۲} & \ldots & X_{T\mathrm{k}} \end{bmatrix} \begin{bmatrix} \hat{\beta}_۱ \\ \hat{\beta}_۲ \\ \vdots \\ \hat{\beta}_k \end{bmatrix} + \begin{bmatrix} u_{۱۱} \\ u_{۲۱} \\ \vdots \\ u_{T۱} \end{bmatrix}$$

S، مجموع مجذورات پسماندها است:

$$S = \sum \left(Y_t - \hat{\beta}_۱ X_{t۱} - \hat{\beta}_۲ X_{t۲} - \cdots - \hat{\beta}_k X_{tk} \right)^۲ = \sum \hat{u}_t^۲$$

$$\frac{\partial S}{\partial \hat{\beta}_۱} = -۲ \sum X_{t۱} \left(Y_t - \hat{\beta}_۱ X_{t۱} - \hat{\beta}_۲ X_{t۲} - \cdots - \hat{\beta}_k X_{tk} \right) = ۰$$

$$\hat{\beta}_۱ \sum X_۱^۲ + \hat{\beta}_۲ \sum X_۱ X_۲ + \cdots + \hat{\beta}_k \sum X_۱ X_k = \sum X_۱ Y$$

$$\hat{\beta}_۱ \sum X_۱ X_۲ + \hat{\beta}_۲ \sum X_۲^۲ + \cdots + \hat{\beta}_k \sum X_۲ X_k = \sum X_۲ Y$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\hat{\beta}_۱ \sum X_۱ X_k + \hat{\beta}_۲ \sum X_۲ X_k + \cdots + \hat{\beta}_k \sum X_\mathrm{k}^۲ = \sum X_k Y$$

$$\begin{bmatrix} \sum X_۱^۲ & \sum X_۱ X_۲ & \cdots & \sum X_۱ X_k \\ \sum X_۱ X_۲ & \sum X_۲^۲ & \ldots & \sum X_۲ X_k \\ \vdots & & & \\ \sum X_k X_۱ & \sum X_k X_۲ & \ldots & \sum X_\mathrm{k}^۲ \end{bmatrix} \begin{bmatrix} \hat{\beta}_۱ \\ \hat{\beta}_۲ \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \sum X_۱ Y \\ \sum X_۲ Y \\ \vdots \\ \sum X_k Y \end{bmatrix}$$

که به شکل ماتریس می‌توان نوشت:

$$(\mathrm{X'X})\,\hat{\beta} \;=\; \mathrm{X'Y}$$

دو طرف را در معکوس X′X پیش ضرب می‌کنیم:

$$(\mathrm{X'X})^{-۱}\,\left(\mathrm{X'}\,\hat{X}\right)\hat{\beta} \;= (\mathrm{X'X})^{-۱}\;\mathrm{X'Y}$$

داریم:

$$\hat{\beta} \;= (\mathrm{X'X})^{-۱}\;\mathrm{X'Y}$$

مثال:

$$Y_t \;=\; \beta_۱ \;+\; \beta_۲\; X_t \;+ u_t$$

$$X = \begin{bmatrix} ۱ & X_۱ \\ ۱ & X_۲ \\ \vdots & \vdots \\ ۱ & X_t \end{bmatrix}, X' = \begin{bmatrix} ۱ & ۱ & ۱ & \cdots & ۱ \\ X_۱ & X_۲ & \cdots & & X_۱ \end{bmatrix}\; X'X = \begin{bmatrix} \mathrm{T} & \sum X_t \\ \sum X_t & \sum x_t^۲ \end{bmatrix}$$

$$X'Y = \begin{bmatrix} \sum Y_t \\ \sum X_t\; Y_t \end{bmatrix}, \quad Y = \begin{bmatrix} Y_۱ \\ \vdots \\ Y_T \end{bmatrix}$$

$$\hat{\beta} \;= \begin{bmatrix} \hat{\beta}_۱ \\ \hat{\beta}_۲ \end{bmatrix} = \begin{bmatrix} \mathrm{T} & \sum X_t \\ \sum X_t & \sum x_t^۲ \end{bmatrix}^{-۱} \begin{bmatrix} \sum Y_t \\ \sum X_t\; Y_t \end{bmatrix} = \begin{bmatrix} \dfrac{\sum Y_t \sum X_t^۲ - \sum Y_t X_t \;\sum X_t}{T\;\sum x_t^۲ - (\sum X_t)^۲} \\ \dfrac{T \sum X_t\; Y_t - \sum X_t \sum Y_t}{T\;\sum x_t^۲ - \;(\sum X_t)^۲} \end{bmatrix}$$

فرض جدید

فرض جدیدی را باید به فروض مدل رگرسیون اضافه کنیم و آن این است که اگر ماتریس مشاهدات بر روی متغیرهای مستقل را با X نشان دهیم، ماتریس T.K می‌باشد، این ماتریس می‌باید مرتبه‌ای برابر K≤T داشته باشد. تنها در این صورت است که می‌توانیم K پارامتر مدل را به دست آوریم. اما چنانچه مرتبه ماتریس X کوچک‌تر از K باشد، یعنی ماتریس مشاهدات، بر روی متغیرهای مستقل دارای K ستون مستقل

از هم باشد آنگاه ضرایب متغیرهایی که با یکدیگر رابطه خطی دارند، قابل تفکیک از یکدیگر نمی‌باشند و اصطلاحاً به این مسئله مشکل هم خطی[1] گفته می‌شود که بعداً به آن خواهیم پرداخت.

یادآوری:در $\hat{\beta} = (X'X)^{-۱} X'Y$ نیازمند به معکوس نمودن ماتریس X'X می‌باشیم، چنان چه ماتریس X دارای مرتبه کمتر از K باشد، در آن صورت ماتریس X'X معکوس پذیر نبوده و اصطلاحاً آن را منفرد[2] گویند.

در مورد تفسیر معادله رگرسیون برآورد شده به صورت روش حداقل مربعات معمولی می‌توان نشان داد که اگر معادله زیر را داشته باشیم:

$$\hat{Y} = \hat{\beta}_{۰} + \hat{\beta}_{۱} X_{۱} + \hat{\beta}_{۲} X_{۲}$$

با استفاده از تفاضل معادله فوق داریم:

$$\Delta\hat{Y} = \hat{\beta}_{۱} \Delta X_{۱} + \hat{\beta}_{۲} \Delta X_{۲}$$

حال اگر $X_{۲}$ ثابت باشد داریم:

$$\Delta X_{۲} = ۰ \rightarrow \Delta\hat{Y} = \hat{\beta}_{۱} \Delta X_{۱}$$

بنابراین در یک معادله گرسیون چند متغیره ضریب برآورد شده با فرض ثابت بودن سایر متغیرها تحلیل می‌شود.

می‌توان نشان داد که ضریب برآورد شده در یک معادله رگرسیون چند متغیره تحت چه شرایطی با ضریب برآورد شده در معادله رگرسیون دو متغیره با یکدیگر برابر می‌باشد. برای این منظور کافی است نشان دهیم که چنانچه:

$$\hat{Y} = \hat{\beta}_{۰} + \hat{\beta}_{۱} X_{۱} + \hat{\beta}_{۲} X_{۲}$$

---

[1] Multicollinearity

[2] Singular

حال اگر X_1 را بر روی X_2 برازش کنیم و جملات اخلال ناشی از این معادله رگرسیون را به عنوان مثال تحت عنوان r_{i1} داشته باشیم تحت این شرایط از برازش این جمله اخلال بر روی Y خواهیم داشت:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1}\, Y_i}{\sum_{i=1}^{n} \hat{r}_{i1}}$$

به طوری که $\hat{r}_{i1}$ بخشی از X_1 می‌باشد که دارای همبستگی با X_2 نمی‌باشد. بنابراین با توجه به این رابطه می‌توان بیان کرد اگر دو معادله رگرسیونی زیر را داشته باشیم تحت چه شرایطی $\hat{\beta}_1 = \tilde{\beta}_1$ برقرار می‌باشد.

$$\tilde{Y} = \tilde{\beta}_0 + \tilde{\beta}_1 X_1$$

$$\tilde{Y} = \tilde{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

رابطه بین $\hat{\beta}_1$ و $\tilde{\beta}_1$ را می‌توان به صورت زیر نمایش داد:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$$

که در آن $\tilde{\delta}_1$ ضریب مربوط به معادله رگرسیون برازش X_{i2} بر روی X_{i1} می‌باشد:

$$X_{i2} = \tilde{\delta}_0 + \tilde{\delta}_1 X_{i1} + r_i$$

بنابراین زمانی $\hat{\beta}_1 = \tilde{\beta}_1$ می‌باشد که دو حالت زیر وجود داشته باشد:

۱) $\hat{\beta}_2 = 0$؛ به عبارتی متغیر X_2 که وارد معادله رگرسیون شده هیچ گونه قدرت توضیح دهندگی Y را نداشته باشد و یا X_2 یک متغیر بی ربط باشد که همبستگی با Y ندارد.

۲) $\tilde{\delta}_1 = 0$؛ به عبارت دیگر هیچ گونه همبستگی بین X_1 و X_2 وجود نداشته باشد.

مثال ۱) چنانچه معادله رگرسیون زیر به صورت انحراف از میانگین در نظر گرفته شده باشد، برآوردهای حداقل مربعات معمولی از ضرایب معادله رگرسیون را به دست آورید:

$$y_t = \beta_۱ X_{۱t} + \beta_۲ X_{۲t} + u_t$$

$$\sum X_۱^۲ = ۴۲, \quad \sum X_۲^۲ = ۲۶, \quad \sum y^۲ = ۵۲۵, \quad \sum X_۱ y = ۳۰, \quad \sum X_۲ y = ۲۰,$$

$$\sum X_۱ X_۲ = ۰, \qquad n = ۲۴$$

با توجه به استخراج ضرایب معادله رگرسیونی با استفاده از روش ماتریس داریم:

$$\hat{\beta} = \left(X_۱' X_۱\right)^{-۱} X_۱' Y$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_۱ \\ \hat{\beta}_۲ \end{bmatrix} = \begin{bmatrix} \sum X_۱^۲ & \sum X_۱ X_۲ \\ \sum X_۱ X_۲ & \sum X_۲^۲ \end{bmatrix}^{-۱} \begin{bmatrix} \sum X_۱ y \\ \sum X_۲ y \end{bmatrix} = \begin{bmatrix} ۴۲ & ۰ \\ ۰ & ۲۶ \end{bmatrix}^{-۱} \begin{bmatrix} ۳۰ \\ ۲۰ \end{bmatrix}$$

$$= \frac{۱}{۲۶ * ۴۲} \begin{bmatrix} ۲۶ & ۰ \\ ۰ & ۴۲ \end{bmatrix} \begin{bmatrix} ۳۰ \\ ۲۰ \end{bmatrix} = \frac{۱}{۱۰۹۲} \begin{bmatrix} ۲۶ & ۰ \\ ۰ & ۴۲ \end{bmatrix} \begin{bmatrix} ۳۰ \\ ۲۰ \end{bmatrix} = \begin{bmatrix} \frac{۲۶}{۱۰۹۲} & ۰ \\ ۰ & \frac{۴۲}{۱۰۹۲} \end{bmatrix} \begin{bmatrix} ۳۰ \\ ۲۰ \end{bmatrix}$$

$$= \begin{bmatrix} \frac{۲۶}{۱۰۹۲} * ۳۰ \\ \frac{۴۲}{۱۰۹۲} * ۲۰ \end{bmatrix} = \begin{bmatrix} ۰.۷۱۴۲ \\ ۰.۷۶۹۲ \end{bmatrix}$$

مثال ۲) با توجه به اطلاعات جدول ذیل معادله رگرسیون چند متغیره مورد نظر را برازش کرده و ضرایب آن را محاسبه کنید:

جدول

مشاهده	Y	X۱	X۲
۱	۴۶	۱۴	۱
۲	۵۱	۱۵	۱,۲۵
۳	۶۹	۱۶	۳
۴	۷۴	۱۷	۳,۲۵
۵	۸۰	۱۸	۴
۶	۸۲	۱۹	۵,۲۵
۷	۹۷	۲۰	۵,۵

با توجه به فرم ماتریسی بیان شده در این فصل می‌توان ساختار زیر را بیان کرد:

$$N\hat{\beta}_0 + \hat{\beta}_1 \sum X_{1i} + \hat{\beta}_2 \sum X_{2i} = \sum Y_i$$

$$\hat{\beta}_0 \sum X_{1i} + \hat{\beta}_1 \sum X_{1i}^2 + \hat{\beta}_2 \sum X_{1i} X_{2i} = \sum X_{1i} Y_i$$

$$\hat{\beta}_0 \sum X_{2i} + \hat{\beta}_1 \sum X_{1i} X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 = \sum X_{2i} Y_i$$

$$7\hat{\beta}_0 \qquad 119\hat{\beta}_1 \qquad 23.25\,\hat{\beta}_2 = 499 \qquad \hat{\beta}_0 = -29.231$$

$$119\hat{\beta}_0 \qquad 2051\hat{\beta}_1 \qquad 417.75\,\hat{\beta}_2 = 8709 \qquad \hat{\beta}_1 = -5.219$$

$$23.25\hat{\beta}_0 \qquad 417.75\hat{\beta}_1 \qquad 95.94\,\hat{\beta}_2 = 1841.25 \qquad \hat{\beta}_2 = -3.549$$

مثال ۳) چنانچه به دنبال بررسی عملکرد افراد در محیط کار باشیم با توجه به ویژگی‌های بیان شده افراد تأثیر شاخص‌های تجربه کاری و ارزش تحصیلات افراد را بر عملکرد افراد مورد بررسی قرار دهید:

جدول

Y	X۱	X۲	X۱❋Y	X۲❋Y	X۱ ❋ X۲	
۱	۴۰	۲۵	۴۰	۲۵	۱۰۰۰	
۲	۴۵	۲۰	۹۰	۴۰	۹۰۰	
۱	۳۸	۳۰	۳۸	۳۰	۱۱۴۰	
۳	۵۰	۳۰	۱۵۰	۹۰	۱۵۰۰	
۲	۴۸	۲۸	۹۶	۵۶	۱۳۴۴	
۳	۵۵	۳۰	۱۶۵	۹۰	۱۶۵۰	
۳	۵۳	۳۴	۱۵۹	۱۰۲	۱۸۰۲	
۴	۵۵	۳۶	۲۲۰	۱۴۴	۱۹۸۰	
۴	۵۸	۳۲	۲۳۲	۱۲۸	۱۸۵۶	
۳	۴۰	۳۴	۱۲۰	۱۰۲	۱۳۶۰	
۵	۵۵	۳۸	۲۷۵	۱۹۰	۲۰۹۰	
۳	۴۸	۲۸	۱۴۴	۸۴	۱۳۴۴	
۳	۴۵	۳۰	۱۳۵	۹۰	۱۳۵۰	
۲	۵۵	۳۶	۱۱۰	۷۲	۱۹۸۰	
۴	۶۰	۳۴	۲۴۰	۱۳۶	۲۰۴۰	
۵	۶۰	۳۸	۳۰۰	۱۹۰	۲۲۸۰	
۵	۶۰	۴۲	۳۰۰	۲۱۰	۲۵۲۰	
۵	۶۵	۳۸	۳۲۵	۱۹۰	۲۴۷۰	
۴	۵۰	۳۴	۲۰۰	۱۳۶	۱۷۰۰	
۳	۵۸	۳۸	۱۷۴	۱۱۴	۲۲۰۴	
Y	X۱	X۲	X۱❋Y	X۲❋Y	X۱ ❋ X۲	
۶۵	۱۰۳۸	۶۵۵	۳۵۱۳	۲۲۱۹	۳۴۵۱۰	جمع
۲۰	۲۰	۲۰	۲۰	۲۰	۲۰	تعداد

بر اساس روش‌های خطی بیان شده می‌توان برآوردگرهای روش حداقل مربعات را به صورت ذیل استخراج کرد:

$$\hat{\beta}_۱ = \frac{\left(\sum X_۲^۲\right)\left(\sum X_۱ y\right)-\left(\sum X_۱ X_۲\right)\left(\sum X_۲ Y\right)}{\left(\sum X_۱^۲\right)\left(\sum X_۲^۲\right)-\left(\sum X_۱ X_۲\right)^۲} = \frac{(۵۲۱.۷۵)(۱۳۹.۵)-(۵۱۵.۵)(۹۰.۲۵)}{(۱۰۹۱.۸)(۵۲۱.۷۵)-(۵۱۵.۵)^۲}$$

$$= \frac{(۷۲۷۸۴.۱۳)-(۴۶۵۲۳.۸۸)}{(۵۶۹۶۴۶.۷)-(۲۶۵۷۴۰.۳)} = \frac{۲۶۲۶۰.۲۵}{۳۰۳۹۰۶.۴} = ۰.۰۸۶۴$$

$$\hat{\beta}_۲ = \frac{\left(\sum X_۱^۲\right)\left(\sum X_۲ y\right)-\left(\sum X_۱ X_۲\right)\left(\sum X_۱ Y\right)}{\left(\sum X_۱^۲\right)\left(\sum X_۲^۲\right)-\left(\sum X_۱ X_۲\right)^۲} = \frac{(۱۰۹۱.۸)(۹۰.۲۵)-(۵۱۵.۵)(۱۳۹.۵)}{(۱۰۹۱.۸)(۵۲۱.۷۵)-(۵۱۵.۵)^۲}$$

$$= \frac{۲۶۶۲۲.۷}{۳۰۳۹۰۶.۴} = ۰.۰۸۶۴$$

$$\hat{\beta}_. = \bar{Y} - \hat{\beta}_۱ \bar{X}_۱ - \hat{\beta}_۲ \bar{X}_۲ = ۳.۲۵ - ۰.۰۸۶۴\left(۵۱.۹\right) - ۰.۰۸۷۶\left(۳۲.۷۵\right) = -۴.۱۰$$

$$\hat{Y} = -۴.۱۰ + ۰.۰۸۶X_۱ + ۰.۰۸۷X_۲$$

فصل چهارم

خواص آماری

برآورد کننده OLS

مقدمه

همان‌گونه که در بخش رگرسیون دو متغیره و روش حداقل مربعات معمولی بیان شد، رابطه بین Y_i متغیر وابسته) و X_i (متغیر مستقل) را در جامعه آماری به صورت مدل رگرسیون زیر در نظر می‌گیریم:

$$Y_i = \beta_. + \beta_۱ X_i + U_i \qquad i = ۱, \cdots, n$$

فروض کلاسیک رگرسیون در مورد جمله اختلال عبارتند از:

$E(u_i \,|X_i) = ۰$ (۱)

$var(u_i \,|X_i) = \sigma^{۲}$ (۲)

(۳) جملات اختلال مستقل از یکدیگر بوده و $\text{cov}\,(u_i,\ u_j) = ۰$ برای $i\ \neq j$

(۴) جمله اختلال u_i از متغیر مستقل X_i مستقل بوده و $\text{cov}\,(u_i\,,\,X_i\,) = ۰$

۵) به منظور استنباط‌های آماری عموماً جمله اخلال u_i دارای توزیع نرمال فرض می‌شود. به علاوه برای تسهیل در محاسبات و استخراج نتایج، فرض غیر تصادفی بودن X_i به فروض فوق اضافه می‌گردد.

چنانچه برآوردگرهای حداقل مربعات معمولی از $\beta_{۰}$ و $\beta_{۱}$ را به ترتیب $\hat{\beta}_{۰}$ و $\hat{\beta}_{۱}$ بنامیم خواهیم داشت:

$$Y_i = \hat{\beta}_{۰} + \hat{\beta}_{۱} X_i + e_i \qquad (۱)$$

و یا:

$$\hat{Y}_i = \hat{\beta}_{۰} + bX_i$$

که در آن e_i برآورد u_i و $\hat{Y}_i$ برآورد Y_i می‌باشد. با در نظر گرفتن متغیرها به صورت انحراف از میانگین ضریب و از روابط فوق حذف شده و داریم:

$$y_i = \hat{\beta}_{۱} x_i + e_i$$

$$\hat{y}_i = bx_i$$

که در آن:

$$y_i = Y_i - \bar{Y}, \qquad \hat{y}_i = \hat{y}_i - \bar{Y}, \qquad x_i = X_i - \bar{X}$$

برآوردگرهای حداقل مربعات از $\beta_{۱}$ و $\beta_{۰}$ به ترتیب از روابط زیر به دست می‌آیند.

$$\hat{\beta}_{۱} = \sum \omega_i Y_i, \qquad \omega_i = \frac{x_i}{\sum x_i^{۲}} \qquad (۲)$$

$$\hat{\beta}_{۰} = \bar{Y} - \hat{\beta}_{۱}\bar{X} = \sum h_i Y_i, \qquad h_i = \frac{۱}{n} - \bar{X}\omega_i \qquad (۳)$$

همان طور که ملاحظه می‌گردد برآوردگرهای مذکور خطی می‌باشند بدین مفهوم که از ترکیب خطی Xها با وزن‌های ω_i و h_i حاصل می‌گردند. این وزن‌ها به گونه ای محاسبه شده اند که $\hat{\beta}_۱$ و $\hat{\beta}_۰$ به ترتیب برآوردهای نااریبی از $\beta_۱$ و $\beta_۰$ بوده و در عین حال کمترین واریانس را در میان برآوردگرهای خطی نااریب داشته باشند. لذا به این برآوردها به اصطلاح blue[۳] می‌گویند. به علاوه ثابت شده است که چنانچه جمله اختلال u_i دارای توزیع نرمال باشد برآوردگر حداقل مربعات معمولی کمترین واریانس را در میان کلیه برآوردگرهای خطی و غیرخطی دارد.

از طرف دیگر روابط (۲) و (۳) برای محاسبه $\hat{\beta}_۱$ و $\hat{\beta}_۰$ با حداقل کردن $\sum e_i^2$ یعنی مجموع مربعات جملات اخلال نیز به دست می‌آیند به همین دلیل آنها را برآوردگر حداقل مربعات معمولی یا OLS می‌نامند. در فرایند حداقل کردن $\sum e_i^2$ دو معادله موسوم به معادلات نرمال به صورت زیر تشکیل می‌گردد:

$$\sum e_i = ۰ \qquad\qquad \sum X_i e_i = ۰$$

که از حل آنها روابط (۲) و (۳) برای محاسبه $\hat{\beta}_۱$ و $\hat{\beta}_۰$ هم به دست می‌آید.

با جایگذاری از مدل اصلی در معادلات (۲) و (۳) داریم:

$$\hat{\beta}_۱ = \beta_۱ + \sum \omega_i u_i,$$

$$\hat{\beta}_۰ = \beta_۰ + \sum h_i u_i \qquad (۵)$$

دو رابطه مذکور در حل بسیاری از مسائل مورد استفاده قرار می‌گیرند. در واقع می‌توان امید ریاضی عبارتهای مشتمل بر متغیرهای تصادفی Y_i، $\hat{\beta}_۱$ و $\hat{\beta}_۰$ را با جایگذاری از روابط (۴) و (۵) در معادله اصلی محاسبه کرد.

---

[3] The Best Linear Unbiased Estimator

به عنوان مثال امید ریاضی، واریانس و کوواریانس برآوردگرهای $\hat{\beta}_۱$ و $\hat{\beta}_۰$ با استفاده از روابط فوق و در نظر گرفتن فروض کلاسیک به صورت زیر به دست می‌آیند:

$$E\left(\hat{\beta}_۱\right) = \beta_۱ + \sum \omega_i E(u_i) = \hat{\beta} \qquad (۶)$$

$$E\left(\hat{\beta}_۰\right) = \beta_۰ + \sum h_i E(u_i) = \hat{\beta}_۰ \qquad (۷)$$

$$var\left(\hat{\beta}_۱\right) = E\left(\hat{\beta}_۱ - \beta_۱\right)^۲ - E(\sum \omega_i u_i)^۲ = \sigma^2 \sum \omega_i^۲ - \sigma^۲/\sum x_i^۲ \qquad (۸)$$

$$var\left(\hat{\beta}_۰\right) = E\left(\hat{\beta}_۰ - \beta_۰\right)^۲ - E(\sum h_i u_i)^۲ = \sigma^۲ \sum h_i^۲ - \sigma^۲ \Big/ \frac{۱}{n} + \frac{\bar{X}^۲}{\sum x_i^۲} \qquad (۹)$$

$$covE\,(\hat{\beta}_۰ - \hat{\beta}_۱) = E(\hat{\beta}_۱ - \hat{\beta}_۱)\,(\hat{\beta}_۰ - \beta_۰) = E(\sum \omega_i u_i)\ (\sum h_i u_i)$$

$$= \sigma^۲ \sum h_i u_i = \sigma^۲ \left(\frac{-\bar{X}}{\sum x_i^۲}\right)$$

که در آن $\sigma^۲$ یا واریانس خطا از رابطه زیر برآورد می‌گردد:

$$S^۲ = \frac{\sum e_t^۲}{n-۲} = \frac{\sum\left(y_i - \hat{\beta}_۱ x_i\right)^2}{n-۲} = \frac{\sum e_i^۲ - \hat{\beta}_۱^۲ \sum x_i^۲}{n-۲}$$

خواص آماری برآوردکننده $\hat{\beta}$ در حالت ماتریسی

این خواص بدون تورش بودن تخمین زننده OLS در مدل چند متغیره است:

$$Y_{T.۱} = X_{T.K}\beta_{T.۱} + u_{T.۱}$$

$$\hat{\beta} = (X'X)^{-۱} X'Y$$

$$\hat{\beta} = (X'X)^{-۱} X'[X\beta + U] = (X'X)^{-۱} X'X\beta = (X'X)^{-۱} X'u$$

$$\hat{\beta} = \beta + (X'X)^{-۱} X'u$$

$$E(\hat{\beta}) = \beta + (X'X)^{-۱} X'E(u) = \beta$$

ماتریس واریانس کوواریانس $\hat{\beta}$

$$\hat{\beta} = \beta + (X'X)^{-۱} X'u$$

$$\hat{\beta} - \beta = (X'X)^{-۱} X'u$$

$$VC(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta') = E(X'X)^{-۱} X'E(uu') X(X'X)^{-۱}$$

$$= (X'X)^{-۱} X'E(uu') X(X'X)^{-۱}$$

$$= \sigma^{۲}(X'X)^{-۱}$$

چون $E(uu') = \sigma^{۲} I_T$ می‌باشد:

$$E\begin{bmatrix} \hat{\beta}_۱ - \hat{\beta}_۱ \\ \hat{\beta}_۲ - \hat{\beta}_۲ \\ \vdots \\ \hat{\beta}_k - \hat{\beta}_k \end{bmatrix}\left[\left(\hat{\beta}_۱ - \hat{\beta}_۱\right)\left(\hat{\beta}_۲ - \hat{\beta}_۲\right) \dots (\hat{\beta}_k - \hat{\beta}_k)\right]$$

$$= E\begin{bmatrix} \left(\hat{\beta}_۱ - \hat{\beta}_۱\right)^۲ & \left(\hat{\beta}_۱ - \hat{\beta}_۱\right)\left(\hat{\beta}_۲ - \hat{\beta}_۲\right) & \cdots & \left(\hat{\beta}_۱ - \hat{\beta}_۱\right)(\hat{\beta}_k - \hat{\beta}_k) \\ \left(\hat{\beta}_۱ - \hat{\beta}_۱\right)\left(\hat{\beta}_۲ - \hat{\beta}_۲\right) & \left(\hat{\beta}_۱ - \hat{\beta}_۱\right)^۲ & \cdots & \left(\hat{\beta}_۲ - \hat{\beta}_۲\right)(\hat{\beta}_k - \hat{\beta}_k) \\ \vdots & & & \\ (\hat{\beta}_k - \hat{\beta}_k)\left(\hat{\beta}_۱ - \hat{\beta}_۱\right) & & \cdots & (\hat{\beta}_k - \hat{\beta}_k)^۲ \end{bmatrix}$$

$$= \sigma^{۲}(X'X)^{-۱}$$

$$\begin{array}{c} VC(\hat{\beta}) \\ \text{ماتریس} \\ \text{واریانس} \\ \text{کوراینس} \end{array} = \begin{bmatrix} \sigma^۲_{\hat{\beta}_۱} & cov\left(\hat{\beta}_۲ - \hat{\beta}_۱\right) & \cdots & \left(\hat{\beta}_۱ - \hat{\beta}_k\right) \\ & & & cov\left(\hat{\beta}_۲ - \hat{\beta}_k\right) \\ cov\left(\hat{\beta}_۲ - \hat{\beta}_۱\right) & \sigma^۲_{\hat{\beta}_۲} & \cdots & cov\left(\hat{\beta}_{k-۱} - \hat{\beta}_k\right) \\ \vdots & & & \\ cov\left(\hat{\beta}_k - \hat{\beta}_۱\right) & & \cdots & \sigma^۲_{\hat{\beta}_۲} \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} \sum X_{t1}^2 & \sum X_{t1} X_{t2} & \cdots & \sum X_{t1} X_{tk} \\ \sum X_{t1} X_{t1} & \sum X_{t2}^2 & \cdots & \sum X_{t2} X_{tk} \\ & & & \vdots \\ \sum X_{tk} X_{t1} & & \cdots & \sum X_{tk}^2 \end{bmatrix}$$

مثال: در مدل دو متغیره

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \sigma^2 \begin{bmatrix} T & \sum x_t \\ \sum x_t & \sum X_t^2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} \sum X_t^2 & -\sum x_t \\ -\sum x_t & T \end{bmatrix} . \frac{1}{\sum X_t^2 - (\sum x_t)^2}$$

$$var\left(\hat{\beta}_0\right) = \frac{\sigma^2 \sum X_t^2}{T \sum X_t^2 - (\sum x_t)^2}, \ \ var\left(\hat{\beta}_1\right) = \frac{T\sigma^2}{T \sum X_t^2 - (\sum x_t)^2}$$

جدول ۴ ادامه جدول ۱ می‌باشد که عامل دیگری را برای تولید محصول ذرت در نظر گرفته که عبارت است از سموم آفات نباتی (X_2):

$$Y_i = \alpha + \beta X_{1i} + \gamma X_{2i} + u_i$$

$$\hat{\alpha} = \bar{Y} - \hat{\beta} X_1 - \hat{\gamma} \bar{X}_2$$

$$\hat{\beta} = \frac{\left(\sum X_1\right)\left(\sum X_2^2\right) - \left(\sum X_2 Y\right)\left(\sum X_1 X_2\right)}{\left(\sum X_1^2\right)\left(\sum X_1^2\right) - \left(\sum X_1 X_2\right)^2}$$

$$\hat{\lambda} = \frac{\left(\sum X_۲ Y\right)\left(\sum X_۱^۲\right) - \left(\sum X_۱ Y\right)\left(\sum X_۱ X_۲\right)}{\left(\sum X_۱^۲\right)\left(\sum X_۲^۲\right) - \left(\sum X_۱ X_۲\right)^۲}$$

با محاسبه فرمول‌های فوق از جدول ۳ خواهیم داشت:

$$\hat{\beta} = \frac{(۹۵۶)(۵۰۴) - (۹۰۰)(۵۲۴)}{(۵۷۶)(۵۰۴) - (۵۲۴)^۲} \approx ۰/۶۵$$

$$\hat{\gamma} = \frac{(۹۰۰)(۵۷۶) - (۹۵۶)(۵۲۴)}{(۵۷۶)(۵۰۴) - (۵۲۴)^۲} \approx ۱/۱۱$$

$$\hat{\alpha} = ۵۷ - (۰/۶۵)(۱۸) - (۱/۱۱)(۱۲) = ۳۱/۹۸$$

جدول

i	Y	X_1	X_2	y	x_1	x_2	$x_1 y$	$x_2 y$	$x_1 x^2$	X_1^2	X_2^2
۱	۴۰	۶	۴	-۱۷	-۱۲	-۸	۲۰۴	۱۳۶	۹۶	۱۴۴	۶۴
۲	۴۴	۱۰	۴	-۱۳	-۸	-۸	۱۰۴	۱۰۴	۶۴	۶۴	۶۴
۳	۴۶	۱۲	۵	-۱۱	-۶	-۷	۶۶	۷۷	۴۲	۳۶	۴۶
۴	۴۸	۱۴	۷	-۹	-۴	-۵	۳۶	۴۵	۲۰	۱۶	۲۵
۵	۵۲	۱۶	۹	-۵	-۲	-۳	۱۰	۱۵	۶	۴	۹
۶	۵۸	۱۸	۱۲	۱	۰	۰	۰	۰	۰	۰	۰
۷	۶۰	۲۲	۱۴	۳	۴	۲	۱۲	۶	۸	۱۶	۴
۸	۶۸	۲۴	۲۰	۱۱	۶	۸	۶۶	۸۸	۴۸	۳۶	۶۴
۹	۷۴	۲۶	۲۱	۱۷	۸	۹	۱۳۶	۱۵۳	۷۲	۶۴	۸۱
۱۰	۸۰	۳۲	۲۴	۲۳	۱۴	۱۲	۳۲۲	۲۷۶	۱۶۸	۱۹۶	۱۴۴
N۱۰	$\sum Y =$ ۵۷۰ Y =۵۷	$\sum x_1 =$ ۱۸۰ $x_1 =$ ۱۸	$\sum x_2 =$ ۱۲۰ $x_2 =$ ۱۲	$\sum y =$ ۰	$\sum x_1 =$ ۰	$\sum x_2 =$ ۰	$\sum x_1 y =$ ۹۵۶	$\sum x_2 y =$ ۹۰۰	$\sum x_1 x_2 =$ ۵۲۴	$\sum X_1^2 =$ ۵۷۶	$\sum X_2^2 =$ ۵۰۴

مثال ۱) مشاهدات بر روی $X_{t۲}$ و $X_{t۳}$ در جدول زیر ارائه شده است.

$$Y_t = \beta_۱ + \beta_۲ X_{t۳} + \beta_۳ X_{t۳} = u_t$$

ضرایب $\beta_۱$ ، $\beta_۲$ ، $\beta_۳$ را با استفاده از روش OLS تخمین بزنید.

t	y_t	$x_{t۲}$	$x_{t۳}$	$x_{t۲}^{۲}$	$x_{t۳}^{۳}$	$x_{t۲}\, y_{t۳}$	$x_۲\, y_t$	$x_۳\, y_t$
۱	-۲	۱	-۶	۱	۳۶	-۶	-۲	-۱۲
۲	۰	-۲	۰	۴	۰	۰	۰	۰
۳	۲	-۱	۲	۱	۴	-۲	-۲	۴
۴	۳	۲	۴	۴	۱۶	۸	۶	۱۲
	۳	**۰**	**۰**	**۱۰**	**۵۶**	**۰**	**۲**	**۲۸**

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_۱ \\ \hat{\beta}_۲ \\ \hat{\beta}_k \end{bmatrix}, \quad X = \begin{bmatrix} X_{۱۱} & X_{۱۲} & X_{۱۳} \\ X_{۲۱} & X_{۲۲} & X_{۲۳} \\ X_{۳۱} & X_{۳۲} & X_{۳۳} \\ X_{۴۱} & X_{۴۲} & X_{۴۳} \end{bmatrix} = \begin{bmatrix} ۱ & ۱ & -۶ \\ ۱ & -۲ & ۰ \\ ۱ & -۱ & ۲ \\ ۱ & ۲ & ۴ \end{bmatrix} \quad X' = \begin{bmatrix} ۱ & ۱ & ۱ & ۱ \\ ۱ & -۲ & -۱ & ۲ \\ -۶ & ۰ & ۲ & ۴ \end{bmatrix}$$

$$X'X = \begin{bmatrix} \sum X_{t_۱}^{۲} & \sum X_{t_۱} \sum X_{t_۲} & \sum X_{t_۱} \sum X_{t_۳} \\ \sum X_{t_۲} X_{t_۱} & \sum X_{t۲}^{۲} & \sum X_{t۲} X_{t۳} \\ \sum X_{t_۳} X_{t_۱} & \sum X_{t_۳} X_{t_۲} & \sum X_{t_۳}^{۲} \end{bmatrix} = \begin{bmatrix} ۴ & ۰ & ۰ \\ ۰ & ۱۰ & ۰ \\ ۰ & ۰ & ۵۶ \end{bmatrix}$$

$$X'Y = \begin{bmatrix} ۳ \\ ۲ \\ ۲۸ \end{bmatrix}, \hat{\beta} \begin{bmatrix} \frac{۳}{۴} \\ \frac{۲}{۱۰} \\ \frac{۲۸}{۵۶} \end{bmatrix} = \begin{bmatrix} ۰/۷۵ \\ ۰/۲۰ \\ ۰/۵۰ \end{bmatrix} \qquad (X'X)^{-۱} = \begin{bmatrix} \frac{۱}{۴} & ۰ & ۰ \\ ۰ & \frac{۱}{۱۰} & ۰ \\ ۰ & ۰ & \frac{۱}{۵۶} \end{bmatrix}$$

تخمین واریانس u $(\hat{\sigma}^۲)$

در مدل بدون عرض از مبدأ یک متغیره داریم، $\hat{\sigma}^۲ = \frac{\sum e_t^۲}{T-۱}$ ، به طوری که e برآورده u باشد.

مدل بدون عرض از مبدأ مقابل را ملاحظه کنید:

$$y_t = x_t\beta + u_t$$

پس از تخمین و جایگزین نمودن آن در مدل فوق داریم:

$$\hat{y}_t = x_t\hat{\beta}$$

برآورد جمله اخلال u_t، یعنی e_t عبارت است از اختلاف بین y_t و $\hat{y}_t$ ، پس داریم:

$$y_t - \hat{y}_t = e_t$$

برای به دست آوردن تخمین واریانس داریم:

$$\sum e_t^۲ = \sum (y_t - \hat{y}_t)^۲$$

$$\sum e_t^۲ = \sum \left(x_t\beta + u_t - x_t\hat{\beta}_t\right)^۲$$

$$\sum e_t^۲ = \sum \{u_t - x_t(\beta + \hat{\beta}_t)\}^۲$$

$$\sum e_t^۲ = \sum \left\{u_t^۲ + x_t^۲(\beta + \hat{\beta}_t)^۲ + ۲x_t u_t (\beta + \hat{\beta}_t)\right\}$$

$$= \sum u_t^۲ + \sum x_t^۲ (\beta - \hat{\beta}_t)^۲ + ۲\sum x_t u_t \ (\beta + \hat{\beta}_t)$$

چنانچه از طرفین، امید ریاضی بگیریم، داریم:

$$E\left(\sum e_t^{۲}\right) = T\sigma^{۲} + \sum x_t^{۲}\left(\beta - \hat{\beta}_t\right)^{۲} - ۲\mathrm{E}\sum x_t u_t \frac{\sum x_t u_t}{\sum x_t^{۲}}$$

چون در مدل یک متغیره بدون عرض از مبدأ $\left(\beta - \hat{\beta}_t\right)^{۲} = \frac{\sigma^{۲}}{\sum x_t^{۲}}$ و جمله آخر نیز برابر است:

$$\mathbf{۲}\,\sigma^{۲}\,\frac{\sum x_t^{۲}}{\sum x_t^{۲}} = \sigma^{۲}$$ پس داریم:

$$T\hat{\sigma}^{۲} = T\sigma^{۲} + \sum x_t^{۲}\ \frac{\sigma^{۲}}{\sum x_t^{۲}} - ۲\sigma^{۲}\frac{\sum x_t^{۲}}{\sum x_t^{۲}}$$

$$T\hat{\sigma}^{۲} = T\sigma^{۲} - \sigma^{۲} = (T-۱)\sigma^{۲}$$

در نتیجه تخمین واریانس u عبارت از:

$$\hat{\sigma}^{۲} = \frac{T-۱}{T}\,\sigma^{۲}$$

برای بررسی بدون توش بودن $\hat{\sigma}^{۲}$ ، مجدداً از آن امید می‌گیریم:

$$E(\hat{\sigma}^{۲}) = \frac{T-۱}{T}\,\sigma^{۲}$$

ملاحظه می‌شود که دارای تورش است، یعنی:

$$E\left(\hat{\sigma}^{۲}\right) \neq \sigma^{۲}$$

پس چنانچه از ابتدا در مخرج کسر به جای T از ۱-T استفاده شود یعنی $\hat{\sigma}^{۲} = \frac{\sum e_t^{۲}}{T-۱}\,\sigma^{۲}$ باشد، یا به عبارت دیگر از درجه آزادی استفاده کنیم، داریم:

$$\hat{\sigma}^{۲} = \frac{T-۱}{T}\,\sigma^{۲}$$

$$\left(T-۱\right)\hat{\sigma}^{۲} = \left(T-۱\right)\,\sigma^{۲}$$

$$E\left(\hat{\sigma}^{۲}\right) = \sigma^{۲}$$

چنانچه مدل دارای دو پارامتر باشد:

$$\hat{\sigma}^{۲} = \frac{\sum e_t^{۲}}{T - ۲}$$

تخمین واریانس u در مدل K متغیره:

$$\hat{\sigma}^{۲} = \frac{\sum \hat{u}_t^{۲}}{T - \mathrm{k}}$$

تخمین واریانس u به زبان ماتریسی عبارت است از:

$$\hat{\sigma}^{۲} = \frac{\hat{U}'\hat{U}}{T - \mathrm{k}} = \frac{(Y - \hat{Y})'(Y - \hat{Y})}{T - \mathrm{k}}$$

$$\left(Y - \hat{Y}\right)'\left(Y - \hat{Y}\right) = Y'Y - \hat{Y}'Y - Y'\hat{Y} + Y'\hat{Y}$$

$$Y'Y - \left(X\hat{\beta}\right)'\hat{Y}$$

$$Y'\hat{Y} = Y'X\hat{\beta} = Y'X(X'X)^{-۱}X'Y = Y'PY\,;P = X(X'X)^{-۱}X'$$

P عبارت از پروجکشن ماتریس یعنی ضرب آن در Y، تصویر Y یا $\hat{Y}$ را میدهد.

$$\hat{Y}'\hat{Y} = \hat{\beta}'X'\hat{\beta} = Y'X(X'X)^{-۱}X'X\hat{\beta} = Y'X\hat{\beta}$$

$$Y'X(X'X)^{-۱}X'Y = Y'PY$$

$$\hat{\sigma}^{۲} = \frac{\hat{Y}'\hat{Y} - \hat{\beta}'X'X\hat{\beta}}{T - \mathrm{k}} = \frac{Y'Y - \mathrm{Y'X(X'X)^{-۱}X'X(X'X)^{-۱}X'Y}}{T - \mathrm{k}}$$

$$\text{صورت} = \mathrm{Y'}Y - \mathrm{Y'X\,(X'X)^{-۱}X'Y} = \mathrm{Y'}\left[\mathrm{I} - \mathrm{X}\left(\mathrm{X'X}\right)^{-۱}\mathrm{X'}\right]\mathrm{Y} = \mathrm{Y'MY}$$

به طوری که:

$$M = I - X\,(X'X)^{-۱}X'$$

M ماتریسی است هم قوه، یعنی:

$$M^{۲} = M.M \quad , \quad M = M'$$

پس:

$$\hat{\sigma}^{۲} = \frac{Y'MY}{T-K}$$

$$\hat{\sigma}^{۲} = \frac{Y'(I-P)Y}{T-K}$$

یا

به طوری که:

$$I - P = M$$

مثال ۲) با توجه به مشاهدات مربوط به X و Y که به صورت جدول زیر جمع آوری شده اند، پارامترهای معادله رگرسیون $Y_t = \beta_{.} + \beta_{۱}X_t + U_t$ را برآورد کرده و سپس خطاهای معیار ضرایب را محاسبه نمایید.

Y_t	۳	۱	۵	۷	۱۲	۸
X_t	۲	۵	۶	۴	۸	۱۷

برای این منظور ابتدا به محاسبه میانگین مشاهدات پرداخته می‌شود و سپس با استفاده از روش اختلاف از میانگین به محاسبه برآوردگرهای مدل رگرسیون پرداخته می‌شود.

$$\bar{X} = \frac{۲ + ۵ + ۶ + ۴ + ۱۷ + ۸}{۶} = ۷$$

$$\bar{Y} = \frac{۳ + ۱ + ۵ + ۷ + ۱۲ + ۸}{۶} = ۶$$

$$\sum x_t Y_t = (۲ - ۷)\ (۳) + (۵ - ۷)\ (۱) + (۶ - ۷)(۵) + (۴ - ۷)\ (۷) + (۸ - ۷)(۱۲)$$
$$+ (۱۷ - ۷)\ (۸) = ۴۹$$

$$\sum x_t^{۲} = (۲ - ۷)^{۲} + (۵ - ۷)^{۲} + (۶ - ۷)^{۲} + (۴ - ۷)^{۲} + (۸ - ۷)^{۲} + (۱۷ - ۷)^{۲}$$
$$= ۱۴۰$$

$$\sum y_t^{۲} = (۳ - ۶)^{۲} + (۱ - ۶)^{۲} + (۵ - ۶)^{۲} + (۷ - ۶)^{۲} + (۱۲ - ۶)^{۲} + (۸ - ۶)^{۲} = ۷۶$$

$$\hat{\beta}_{۱} = \frac{\sum x_t Y_t}{\sum x_t^{۲}} = \frac{۴۹}{۱۴۰} = ۰.۳۵$$

$$\hat{\beta}_{۰} = \bar{Y} - \hat{\beta}_{۱}\bar{X} = ۶ - (۰.۳۵)(۷) = ۳.۳۵$$

$$RSS = \sum e_t^{۲} = \sum y_t^{۲} - \hat{\beta}_{۱} \sum x_t^{۲} = ۷۶ - (۰.۳۵)^{۲}(۱۴۰) = ۵۸.۸۵$$

$$S^{۲} = \frac{Rss}{n - ۲} = \frac{۵۸.۸۵}{۶ - ۲} = ۱۴.۷۱$$

$$var(\hat{\beta}_{۱}) = \frac{S^{۲}}{\sum x_t^{۲}} = \frac{۱۴.۷۱}{۱۴۰} = ۰.۱۰۵۰ \Rightarrow SE(\hat{\beta}_{۱}) = ۰.۳۲۴$$

$$var(\hat{\beta}_{۱}) = S^{۲}\left(\frac{1}{n} + \frac{\bar{X}^{۲}}{\sum x_t^{۲}}\right) = ۱۴.۷۱\left(\frac{۱}{۶} + \frac{۴۹}{۱۴۰}\right) = ۷.۶۰۰۱ \Rightarrow SE(\hat{\beta}_{۰}) = ۲.۷۵۷$$

جدول ذیل نتیجه خروجی محاسبه شده توسط نرم افزار Excel را نمایش می‌دهد که نتایج به دست آمده با محاسبات انجام شده یکسان می‌باشد.

SSUMMARY OUTPUT		
Regression Statistics		
Multiple R	0.4750	

R Square	0.2257				
Adjusted R Square	0.0321				
Standard Error	3.8357				
Observations	6				
ANOVA					
	df	SS	MS	F	Significance F
Regression	1	17.150	17.150	1.166	0.341
Residual	4	58.850	14.713		
Total	5	76			
	Coefficients	Standard	T Stat	P-value	Lower 95%
Intercept	3.55	2.7571	1.287	0.2673	-4.1049
X	0.35	0.3242	1.079	0.3410	-0.5501

مثال ۳) مدل زیر را در نظر بگیرید

$Y_t \ = \ \alpha + \beta X_t \ + u_t$ t = 1, ...,T

چنانچه جمله اخلال u_t فروض کلاسیک را تأمین کند، مشاهدات را با فرض زوج بودن T بدو گروه مساوی با n مشاهده در هر گروه تقسیم می‌کنیم (T = 2n). $\bar{X}_۱$, $\bar{X}_۲$, $\bar{Y}_۱$ و $\bar{Y}_۲$ نشان دهنده به ترتیب میانگین X_t و Y_t در هر گروه می‌باشد. پیشنهاد شده است که پارامتر β از رابطه زیر تخمین زده شود

$$\widehat{\beta} = \frac{\bar{Y}_۱ - \bar{Y}_۲}{\bar{X}_۱ - \bar{X}_۲}$$

برای این منظور $E(\widehat{\beta})$ و $var(\widehat{\beta})$ را محاسبه کنید.

ابتدا $\hat{\beta}$ را برحسب جملات اخلال به دست می‌آوریم. مشاهدات مربوط به گروه اول را با $(X_{۱i}, Y_{۱i})$ و مشاهدات مربوط به گروه دوم را با $(X_{۲i}, Y_{۲i})$ نشان می‌دهیم.

$$\hat{\beta} = \frac{\sum Y_{۱i} - \sum Y_{۲i}}{\sum X_{۱i} - \sum X_{۲i}} = \frac{\sum\left(\alpha + \beta X_{۱i} + u_{۱i}\right) - \sum\left(\alpha + \beta X_{۲i} + u_{۲i}\right)}{\sum X_{۱i} - \sum X_{۲i}}$$

$$\frac{n\alpha + \beta \sum X_{۱i} + \sum u_{۱i} - n\alpha + \beta \sum X_{۲i} - \sum u_{۲i}}{\sum X_{۱i} - \sum X_{۲i}} = \beta + \frac{\sum u_{۱i} - \sum u_{۲i}}{\sum X_{۱i} - \sum X_{۲i}}$$

بنابراین امید ریاضی و واریانس $\hat{\beta}$ با توجه به فروض کلاسیک برابر خواهد بود با:

$$E\left(\hat{\beta}\right) = \beta + \frac{\sum E(u_{۱i}) - \sum E(u_{۲i})}{\sum X_{۱i} - \sum X_{۲i}} = \beta$$

$$var\left(\hat{\beta}\right) = E\left(\hat{\beta} - \beta\right)^{۲} = E\left(\frac{\sum u_{۱i} - \sum u_{۲i}}{\sum X_{۱i} - \sum X_{۲i}}\right)^{۲} = \frac{\sum E(u_{۱i}^{۲}) - \sum E(u_{۲i}^{۲})}{\left(\sum X_{۱i} - \sum X_{۲i}\right)^{۲}}$$

$$= \frac{n\sigma^{۲} + n\sigma^{۲}}{\left(\sum X_{۱i} - \sum X_{۲i}\right)^{۲}} = \frac{T\sigma^{۲}}{\left(\sum X_{۱i} - \sum X_{۲i}\right)^{۲}}$$

مثال ۴) معادله رگرسیون $Y_t = \beta X_t + \varepsilon_t$ مفروض است. ε_t کلیه فروض کلاسیک را تأمین می‌کند. سه برآوردگر زیر را از β در نظر بگیرید:

(۱) $$\hat{\beta}_{۱} = \frac{\bar{Y}}{\bar{X}}$$

(۲) $$\hat{\beta}_{۲} = \frac{\sum X_t Y_t}{\sum x_t^{۲}}$$

(۳) $$\hat{\beta}_{۳} = \frac{\sum X_t Y_t}{\sum x_t^{۲}}$$

نشان دهید که آیا هر سه برآورد کننده ناریب اند و کدامیک از سه برآورد کننده فوق بهترین است؟

به طور کلی برای حل اینگونه مسائل بایستی به جای Y_t از معادله رگرسیون مربوطه جایگزین کرده و پس از ساده کردن عبارت حاصله از دو طرف رابطه امید ریاضی گرفته شود. همان طور که در ادامه خواهیم دید سه برآورد کننده مذکور ناریب هستند. برای برآوردگر (۱) داریم:

$$\hat{\beta}_{۱} = \frac{\bar{Y}}{\bar{X}} = \frac{\sum Y_t}{\sum X_t} = \frac{\sum(\beta X_t + \varepsilon_t)}{\sum X_t} = \frac{\beta \sum X_t + \sum \varepsilon_t}{\sum X_t} = \beta + \left(\frac{1}{\sum X_t}\right)\sum \varepsilon_t = \beta + W_{1t}\sum \varepsilon_t \qquad (۴)$$

که در آن $W_1 = \frac{۱}{\sum X_t}$ است از دو طرف رابطه فوق امید ریاضی می‌گیریم. با توجه به آنکه X_t متغیری غیر تصادفی و $E(\varepsilon_t) = ۰$ است خواهیم داشت:

$$E\left(\hat{\beta}_{۱}\right) = \beta + W_{1t}\, E(\varepsilon_t) = \beta$$

به همین ترتیب برای برآوردگر (۲) داریم:

$$\hat{\beta}_{۲} = \frac{\sum X_t\, Y_t}{\sum x_t^{۲}} = \frac{\sum X_t(\beta X_t + \varepsilon_t)}{\sum x_t^{۲}} = \frac{\beta \sum x_t^{۲} + \sum X_t\, \varepsilon_t}{\sum x_t^{۲}} = -\beta + \sum\left(\frac{X_t}{\sum x_t^{۲}}\right)\varepsilon_t = \qquad (۵)$$

$$\beta + W_{۲t}\, \varepsilon_t$$

که در آن $W_{۲t} = \frac{\sum X_t}{\sum x_t^{۲}}$ است توجه کنید که در اینجا وزن‌های $W_{۲t}$ تابعی از مشاهده t ام می‌باشد زیرا X_t در صورت کسر مربوطه ظاهر می‌شود به همین دلیل از اندیس برای وزن‌ها استفاده می‌شود. اما در حالت قبل وزن‌ها برای تمامی جملات اخلال برابر مقدار ثابت $W_{۱}$ به دست آمد. از دو طرف رابطه حاصله در فوق امید ریاضی می‌گیریم:

$$E\left(\hat{\beta}_{۲}\right) = \beta + \sum W_{۲t}\; E(\varepsilon_t) = \beta$$

و بالاخره برای برآوردگر (۳) خواهیم داشت:

$$\hat{\beta}_{۳} = \frac{\sum x_t\, Y_t}{\sum x_t^{۲}} = \frac{\sum X_t(\beta X_t + \varepsilon_t)}{\sum x_t^{۲}} = \frac{\beta \sum X_t\, x_t}{\sum x_t^{۲}} + \sum\left(\frac{X_t}{\sum x_t^{۲}}\right)\varepsilon_t$$

اما با توجه به این نکته که $\sum X_t\, x_t = \sum x_t^{۲}$ است لذا داریم:

$$\hat{\beta} = \beta + \sum(\frac{X_t}{\sum x_t^{۲}})\, \varepsilon_t = \beta + \sum \omega_t\, \varepsilon_t \qquad (۶)$$

همان طور که ملاحظه گردید سه برآوردگر مذکور ناریب می‌باشند. لازم به توضیح است که در این مرحله ما نیازی به تعریف متغیرهای جدید W برای وزن‌ها نداشتیم اما همان طور که خواهیم دید تعریف این متغیرها در قسمت بعد عملیات محاسباتی را ساده تر خواهند نمود.

بهترین برآورد کننده بایستی کمترین واریانس را داشته باشد. واریانس $\hat{\beta}_{۱}$ با استفاده از رابطه (۴) به طریقه زیر محاسبه می‌شود:

$$\hat{\beta}_{۱} - \beta = W_{۱} \sum \varepsilon_t \rightarrow var\left(\hat{\beta}_{۱}\right) = E\left(\hat{\beta}_{۱} - \beta\right)^{۲} = E\left[W_1 \sum \varepsilon_t\right]^{۲}$$

$$= E[W_{۱}^{۲}(\varepsilon_t + \cdots + \varepsilon_n)^{۲} = W_{۱}^{۲} E(\varepsilon_t + \cdots + \varepsilon_n)^{۲}$$

پس از بسط عبارت داخل پرانتز و استفاده از فروض کلاسیک $E(\varepsilon_t \varepsilon_t) = 0$ برای $t \neq s$ و $E(\varepsilon_{t}^{۲}) = ۰$

رابطه تعریفی $W_{۱} = \frac{۱}{\sum X_t}$ داریم:

$$var\left(\hat{\beta}_{۱}\right) = W_{۱}^{۲}(n\sigma^2) = \frac{n\sigma^{۲}}{(\sum X_i)^{۲}} = \frac{n\sigma^{۲}}{(n\bar{X})^{۲}} = \frac{\sigma^{۲}}{n\bar{X}^{۲}}$$

به همین ترتیب واریانس $\hat{\beta}_{۲}$ با استفاده از (۵) به شیوه زیر محاسبه می‌شود.

$$\hat{\beta}_{۲} - \beta = \sum W_{۲t}\varepsilon_t \rightarrow var\left(\hat{\beta}_{۲}\right) = E\left(\hat{\beta}_{۲} - \beta\right)^{۲} = E(\sum W_{۲t}\varepsilon_t)^{۲}$$

پس از بسط عبارت داخل پرانتز و استفاده از فروض کلاسیک $E(\varepsilon_t \varepsilon_t) = ۰$ برای $t \neq s$ و $E(\varepsilon_{t}^{۲}) = \sigma^{۲}$و

رابطه تعریفی $W_{۲t} = \frac{X_t}{\sum x_t^{۲}}$ داریم:

$$var\left(\hat{\beta}_{۲}\right) = \sigma^2 \sum W_{۲t}^{۲} = \sigma^2 \sum \left(\frac{X_t}{\sum x_t^{۲}}\right)^{۲} = \frac{\sigma^{۲}}{\sum x_t^{۲}}$$

و بالاخره برای محاسبه واریانس $\hat{\beta}_۳$ با توجه به رابطه (۶) به ترتیب زیر عمل می‌کنیم:

$$\hat{\beta}_۳ - \beta = \sum \omega_t \varepsilon_t \rightarrow var\left(\hat{\beta}_۳\right) = E\left(\hat{\beta}_۳ - \beta\right)^۲ = E(\sum \omega_t \varepsilon_t)^۲$$

$$\sigma^2 \sum \omega_{\mathrm{t}}^۲ = \sigma^2 \sum \left(\frac{X_t}{\sum x_t^۲}\right) = \frac{\sigma^۲}{\sum x_t^۲}$$

حال نشان می‌دهیم که واریانس $\hat{\beta}_۲$ کمتر از واریانس برآورد کننده‌های دیگر است. بدین منظور ابتدا $var(\hat{\beta}_۳)$ را با $var(\hat{\beta}_۲)$ مقایسه می‌کنیم:

$$var\left(\hat{\beta}_۳\right) = \frac{\sigma^۲}{\sum(X_t - \bar{X})^۲} = \frac{\sigma^۲}{\sum X_t^۲ - n\bar{X}^۲} > \frac{\sigma^۲}{\sum X_t^۲} = var(\hat{\beta}_۲)$$

حال $var(\hat{\beta}_۱)$ را با $var(\hat{\beta}_۳)$ مقایسه می‌کنیم:

$$var\left(\hat{\beta}_۱\right) = \frac{\sigma^۲}{n\bar{X}^۲} = \frac{\sigma^۲}{\sum X_t^۲ - \sum x_t^۲} > \frac{\sigma^۲}{\sum X_t^۲} = var(\hat{\beta}_۲)$$

در عبارت فوق از اتحاد $\sum x_t^۲ = \sum X_t^۲ - n\bar{X}^۲$ و یا به طور معادل $n\bar{X}^۲ = \sum X_t^۲ - \sum x_t^۲$ استفاده نموده‌ایم. همانطور که ملاحظه گردید $\hat{\beta}_۲$ بهترین برآوردگر می‌باشد. در مسئله بعد نشان خواهیم داد که مقدار $\hat{\beta}_۲$از حداقل کردن مجموع مربعات خطا یعنی $\sum e_t^۲$ در معادله رگرسیون $Y_t = \beta X_t + \varepsilon_t$ به دست می‌آید.

مثال ۵) معادله رگرسیون $Y_t = \beta X_t + \varepsilon_t$ را که در آن ضریب مقدار ثابت برابر صفر می‌باشد و $\varepsilon_t \sim IN\,(۰\,, \sigma^۲)$ در نظر بگیرید، برای این منظور برآوردگر حداقل مربعات معمولی را برای β به دست آورده و واریانس آن را محاسبه کنید همچنین برآوردگر نااریبی برای var(b) پیشنهاد نمایید.

برآوردگر حداقل مربعات از β را در رگرسیون مورد نظر با b نشان میدهیم. در اینصورت خواهیم داشت:

$$Y_t = bX_t + e_t$$

$$\min \sum e_t^{۲} \Rightarrow \frac{d(\sum e_t^{۲})}{db} = \frac{d\sum (Y_t - bX_t)^{۲}}{db} = ۰$$

$$-۲\sum X_t(Y_t - bX_t) = ۰ \Rightarrow -\sum X_t Y_t + \mathrm{b}\sum X_t^{۲} = ۰$$

$$\mathrm{b} = \sum W_t Y_t \quad , \quad W_t = \frac{X_t}{\sum x_t^{۲}} \qquad (۱)$$

به علاوه شرایط کافی برای حداقل بودن جواب حاصله را می‌توان به سادگی تحقیق نمود داریم:

$$\frac{d^{۲}\sum (Y_t - bX_t)^{۲}}{db^{۲}} = -۲\sum X_t^{۲} < ۰$$

بنابراین معادله (۱) جواب‌های مورد نظر می‌باشند. در این مورد هیچ یک از متغیرهای X_t و Y_t در به صورت انحراف از میانگین ظاهر نمی‌شوند. این برآوردگر در میان براورد کننده‌های خطی بدون تورش حداقل واریانس را دارد و لذا یک برآوردگر blue است.

حال واریانس b را با توجه به رابطه حاصله (۱) به دست می‌آوریم:

$$b = \frac{\sum X_t Y_t}{\sum X_t^{۲}} = \frac{\sum X_t\,(\beta X_t + \varepsilon_t)}{\sum X_t^{۲}} = \beta + \frac{\sum X_t \varepsilon_t}{\sum X_t^{۲}} = \beta + \sum W_t \varepsilon_t \qquad (۲)$$

که در آن $W_t = \frac{X_t}{\sum X_t^{۲}}$ است. با توجه به $E(b) = \beta$ داریم:

$$var(b) = E(b-\beta)^{۲} = E\left(\sum W_t \varepsilon_t\right)^{۲} = E\left(\sum W_{۱}\varepsilon_{۱} + \cdots + \sum W_n \varepsilon_n\right)^{۲} = \sigma^{۲}\sum W_t^{۲}$$

$$= \sigma^{۲}\sum\left(\frac{X_t}{\sum X_t^{۲}}\right)^{۲} = \frac{\sigma^{۲}\sum X_t^{۲}}{(\sum X_t)^{۲}} = \frac{\sigma^{۲}}{\sum X_t^{۲}}$$

در اینجا نیز تنها تفاوت در محاسبه واریانس b با حالتی که در معادله رگرسیون عرض از مبدأ مخالف صفر می‌باشد آن است که X_t در مخرج کسر به صورت انحراف از میانگین ظاهر نمی‌شود.

به منظور تخمین var(b) بایستی $\sigma^{۲}$ را برآورد کنیم. می‌توان پیش بینی کرد که $S^{۲} = \frac{\sum e_t^{۲}}{n-۱}$ یک برآوردگر ناااریب از $\sigma^{۲}$ است زیرا در معادله رگرسیون مورد بحث تنها یک پارامتر یعنی β تخمین زده شده و در نتیجه درجه آزادی آن n − ۱ است. در ادامه نااریب بودن $S^{۲}$ را اثبات می‌کنیم:

$$E(S^{۲}) = \frac{\sum \varepsilon_t^{۲}}{n-۱} = \frac{E[\sum (Y_t - bX_t)^{۲}]}{n-۱}$$

به جای b و Y در رابطه فوق به ترتیب از معادله (۲) و $Y_t - bX_t + \varepsilon_t$ جایگزین می‌کنیم:

(۳)

$$E(S^{۲}) = \frac{E\left\{\sum[(\beta X_t - \varepsilon_t) - (\beta X_t - \sum W_t\,\varepsilon_t)X_t]^{۲}\right\}}{n-۱} = \frac{E[\sum(\varepsilon_t - X_t \sum W_t \varepsilon_t)^{۲}}{n-۱}$$

$$= \frac{E\left[\sum \varepsilon_t^{۲} - ۲\sum W_t\,\varepsilon_t\,\sum X_i\,\varepsilon_i + \sum X_t^{۲}\,(\sum W_t\,\varepsilon_t)^{۲}\right]^{۲}}{n-۱}$$

با در نظر گرفتن فروض کلاسیک $\sum \varepsilon_i^{۲} = \sigma^{۲}$ و $E(\varepsilon_i\,\varepsilon_j) = ۰$ برای $i \neq j$ و غیرتصادفی بودن X_t اجزاء صورت کسر را محاسبه می‌کنیم:

$$E\,\sum\left(\varepsilon_i^{۲}\right) = n\,\sigma^{۲},\;\; E(\sum W_t\,\varepsilon_t)(\sum X_t\,\varepsilon_t) = \sigma^{۲}\sum X_t\,W_t = \sigma^{۲}\sum X_t\left(\frac{X_t}{\sum X_t^{۲}}\right) = \sigma^{۲},$$

$$\sum X_t^{۲}\,E\left(\sum W_t\,\varepsilon_t\right)^{۲} = \sigma^{۲}\sum X_t^{۲}\sum W_t^{۲} = \sigma^{۲}\sum X_t^{۲}\left(\frac{X_t}{\sum X_t^{۲}}\right)^{۲} = \sigma^{۲}$$

با جایگذاری عبارات فوق در (۳) داریم:

$$\mathrm{E}\left(S^{۲}\right) = \frac{n\sigma^{۲} - ۲\sigma^{۲} + \sigma^{۲}}{n-۱} = \sigma^{۲}$$

بنابراین بمنظور برآورد واریانس b میتوان از رابطه $\widehat{var}(b) = S^{۲} / \sum X_i^{۲}$ که در آن $S^{۲} = \frac{\sum x_t^{۲}}{n-۱}$ است استفاده نمود.

فصل هفتم

خواص برآورده کننده‌ها در نمونه‌های بزرگ

سازگاری: خواصی که درباره تخمین زننده‌ها تا به حال گفته شد، مانند بدون تورش بودن و کارایی، بیشتر در حجم نمونه‌های کوچک مورد استفاده قرار می‌گیرد، بنابراین لازم است به خواص تخمین زننده‌ها در اندازه‌های بزرگ نمونه توجه کرد. این خواص را خواص مجانبی[۱] گویند. یکی از این خواص ویژگی سازگاری[۲] است، البته بدون تورش بودن به طور مجانبی. کارایی به طور مجانبی نیز از خواص دیگر نمونه‌های حجم زیاد است.

فرض کنید $\hat{\theta}_n$ تخمین زن θ بر اساس حجم نمونه n باشد. آنگاه دنبال تخمین زننده‌های $\hat{\theta}_n$، یک دنباله سازگار خوانده می‌شود. اگر برای هر مقدار اختیاری $\delta > ۰$ و $\varepsilon > ۰$ داشته باشیم:

$$prob\left[\left|\hat{\theta}_n - \theta\right| < \varepsilon\right] > 1 - \delta$$

یعنی با افزایش حجم نمونه n، $\hat{\theta}_n$ با احتمال نزدیک به یک به مقدار واقعی و نزدیک θ می‌شود. این بیان را به صورت زیر هم مینویسند:

$$prob\left(\left|\hat{\theta}_n - \theta\right| < \varepsilon\right) = ۱$$

و به طور خلاصه می‌نویسند:

$$\text{Plim}\ \hat{\theta}_n = \theta$$

$$Plim = Probablity\ Limit$$

می گویند: $\hat{\theta}_n$ به طور احتمالی به سمت θ گرایش پیدا می‌کند.

---

[1] Asymptotic
[2] Consistency

در عمل غالبا مفاهیم فوق را ویژگی، سازگاری برای θ به کار می‌بریم.

شرایط کافی برای سازگاری: شرط کافی برای سازگاری $\hat{\theta}_n$ آن است که وقتی تورش و واریانس اندازه در نمونه، افزایش یابد به سمت صفر تمایل پیدا می‌کند. این شرط، شرط خوبی است که در عمل، سازگاری یک تخمین زننده را بررسی کنیم ولی توجه داریم که شرط لازمی برای سازگاری نیست. یک امین زننده می‌تواند سازگار باشد حتی اگر مقدار تورش آن به سمت صفر میل نکند.

برخی از خواص Plim

اگر $C_۱$ و $C_۲$ مقادیر ثابتی باشند:

$$\text{Plim}\left(C_۱ y_{t۱} + C_۲ y_{t۲}\right) = C_۱ \text{Plim}\, y_{t۱} + C_۲ \text{Plim}\, y_{t۲} \qquad (۱)$$

$$\text{Plim}\left(y_{t۱} y_{t۲}\right) = (\text{Plim}\ y_{t۱}) + (\text{Plim}\, y_{t۲}) \qquad (۲)$$

$$\text{Plim}\left(y_{t۱} / y_{t۲}\right) = \text{Plim}\ y_{t۱} / (\text{Plim}\, y_{t۲} \qquad \text{اگر} \qquad (y_۲ \neq ۰) \qquad (۳)$$

اگر $Plim\ y_t = c$ باشد و $g_t = (y)$ تابعی پیوسته از لا باشد، آنگاه داریم:

$$\text{Plim}\ g(y_t) = g(c) \qquad (۴)$$

تخمین زننده OLS برای ضرایب مدل رگرسیون خطی یک تخمین زننده سازگار می‌باشند؟

مدل دو متغیره را در نظر بگیریم:

$$y_t = \beta_۰ + \beta_۱ X_t + u_t$$

برحسب انحراف از معیار داریم:

$$\hat{\beta}_۱ = \frac{\sum x_t y_t}{\sum x_t^۲}$$

$$\hat{\beta}_۱ = \sum w_t y_t$$

$$w_t = \frac{x_t}{\sum x_t^{۲}}$$

$$\hat{\beta}_{۱} = \beta_{۱} + \sum w_t y_t = \beta_{۱} + \frac{\sum x_t u_t}{\sum x_t^{۲}}$$

برای اثبات سازگاری، از طرفین حد احتمال می‌گیریم:

$$\text{Plim}\ \beta_{۱} = \text{Plim}\ \beta_{۱} + \text{Plim}\ (\frac{\sum x_t u_t}{\sum x_t^{۲}})$$

$$= \beta_{۱} + \frac{\text{Plim}\ (\frac{\sum x_t u_t}{T})}{\text{Plim}\ (\frac{\sum x_t^{۲}}{T})}$$

اما $(\frac{\sum x_t u_t}{T})$ Plim یک برآورد کننده سازگار از $Cov(X_t u_t)$ می‌باشد و $(\frac{\sum x_t^{۲}}{T})$ Plim هم یک براوردکننده ساز گار از $\text{Var}(X_t)$ می‌باشد.

چون X_t یک مقدار ثابت و غیرتصادفی است، $Cov(X_t u_t) = ۰$،

پس

$$\text{Plim}\ (\frac{\sum x_t u_t}{T}) \rightarrow ۰$$

و واریانس X_t هم یک مقدار ثابت است، بنابراین

$$\text{Plim}\hat{\beta}_{۱} = \beta_{۱} + \frac{۰}{\text{مقدار ثابت}}$$

مقدار

پس $\hat{\beta}_{۱}$ یک تخمین زننده سازگار است.

$$\text{Plim}\widehat{\beta}_1 = \beta_1$$

به همین ترتیب برای $\hat{\beta}_0$ می‌توان نشان داد، برآورد کننده‌های سازگار است.

در حالت چند متغیره:

Y = Xβ +U

$$\hat{\beta} = (X'X)^{-1} X'Y = (X'X)^{-1} X'[X\beta + U] = \beta + (X'X)^{-1}X'U$$

$$\hat{\beta} = \beta + (X'X)^{-1}X'U$$

$$\text{Plim}\widehat{\beta} = \text{Plim}\beta + \text{Plim}\left(X'X\right)^{-1} X'U = \beta +$$

$$\text{Plim}\frac{\left(X'X\right)^{-1}}{T}.\text{Plim}\frac{X'U}{T}$$

اگر $\text{Plim}\frac{X'X}{T}^{-1} \to Q$ ، یعنی به Q میل کند، به طوری که Q^{-1} وجود داشته باشد، آنگاه

$\text{Plim}\left[\frac{X'X}{T}^{-1}\right] \to Q^{-1}$ مقداری ثابت است.

برای سازگاری باید ثابت کرد که $\text{Plim}\frac{X'U}{T}$ به سمت صفر میل می‌کند.

$$\text{Plim}\widehat{\beta} = \beta + A^{-1}.\circ = \beta$$

مثال: از نمونه $X_1, X_2, \dots, X_n$ ، $\bar{X}$ یک تخمین زن بدون تورش و سازگار است.

$$E\bar{X} = \mu$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

$$\lim Var(\bar{X}) = \lim_{n\to\infty}\frac{\sigma^2}{n} \to 0$$

بنابراین $\bar{X}$ شرط کافی برای سازگاری را داراست.

فصل هشتم

استفاده از رگرسیون برای پیش بینی

یکی از هدف‌های عمده در مطالعات اقتصادسنجی، پیش بینی مقدار متغیر وابسته به ازای مقدار معینی از متغیر مستقل است. این پیش بینی بر اساس برآورد پارامترها می‌باشد و ممکن است به دو صورت انجام گیرد.

۱- پیش بینی میانگین متغیر وابسته به ازای یک مقدار معین از متغیر مستقل X، مثلا X_{T+1}

۲- پیش بینی متغیر وابسته به ازای یک مقدار معینی از متغیر مستقل X، مانند X_{T+1}.

واضح است که پیش بینی دوم کمتر از حالت اول قابل پیش بینی است و در هر یک از حالات فوق دچار خطا خواهیم شد که متوسط هر دو خطا برابر با صفر است و واریانس آنها را نیز می‌باید محاسبه کرد.

مثال: مدل زیر را در نظر بگیرید، به طوری که β_1 ، β_0 از روش OLS برآورد شده باشند:

$$Y_t = \beta_0 + \beta_1 X_t + U_t$$

$$t = 1, \dots, T$$

حال چنان چه مقدار X برای دوره $(T+1)$ داده شده باشد، برآورد مقدار Y، متناظر با دوره $(T+1)$ را پیش بینی کننده گویند. ما سعی می‌کنیم مقدار واقعی $(Y_T + 1)$ را پیش بینی کنیم.

$$Y_{T+1} = \beta_0 + \beta_1 X_{T+1} + u_{T+1}$$

چون ما فقط بر اساس نمونه برآورد β_1 ، β_0 را داریم، یک منبع خطا، خطای نمونه گیری در برآورد پارامترهاست. اما بدون توجه به دقت تخمین پارامترها، ما هرگز قادر به پیش بینی u_{T+1} به طور کامل نیستیم.

پیش بینی کننده‌ها به صورت زیر خواهند بود:

$$Y_{T+۱}^{p} = \hat{\beta}_{۰} + \hat{\beta}_{۱} X_{T+۱}$$

خطای پیش بینی عبارت است از:

$$e_{T+۱} = Y_{T+۱} - Y_{T+۱}^{p}$$

$$= \beta_{۰} + \beta_{۱} X_{T+۱} + u_{T+۱} - \hat{\beta}_{۰} + \hat{\beta}_{۱} X_{T+۱}$$

$$= (\beta_{۰} - \hat{\beta}_{۰}) + (\beta_{۱} - \hat{\beta}_{۱}) X_{T+۱} + u_{T+۱}$$

با گرفتن امید ریاضی از طرفین، ملاحظه می‌شود که $E(e_{T+1}) = ۰$

است. بنابراین، پیش بینی بر اساس روش OLS بدون تورش است، بدین معنی که خطای پیش بینی دارای میانگین صفر است.

واریانس خطای پیش بینی به صورت زیر محاسبه می‌شود:

$$Var(e_{T+1}) = var\left(\hat{\beta}_{۰}\right) + \left(X_{T+۱}\right)^{۲} var\left(\hat{\beta}_{۱}\right) + ۲X_{T+۱}\,\mathrm{Cov}\left(\hat{\beta}_{۰} + \hat{\beta}_{۱}\right) + Var(u_{T+1})$$

$$= \sigma^{۲}\left[۱ + \left[\frac{۱}{T} + \frac{\bar{X}^{۱}}{S_{xx}}\right] + \frac{\left(X_{T+۱}\right)^{۲}}{S_{xx}} - \frac{۲\bar{X}X_{T+۱}}{S_{xx}}\right]$$

$$= \sigma^{۲}\left[۱ + \frac{۱}{T} + \frac{\left(X_{T+۱} - \bar{X}\right)^{۲}}{S_{xx}}\right]$$

با قرار دادن تخمین نمونه ای $\sigma^{۲}$که قادر خواهیم بود برای $X_{T+۱}$ فاصله اطمینان بسازیم.

ملاحظه می‌شود که میزان پراکندگی خطای پیش بینی رابطه‌ای غیرمستقیم با حجم نمونه، T دارد یعنی هر چه قدر حجم نمونه بیشتر باشد میزان خطا کمتر است از طرفی میزان پراکندگی رابطه ای مستقیم با فاصله $X_{T+۱}$ از مرکز مشاهدات دارد و همچنین میزان پراکندگی به تغییرات متغیر X_t، رابطه ای مکوس دارد، یعنی وقتی X_t ، دارای تغییرات بیشتری باشند قدرت بیش بینی افزایش می‌یابد.

پیش بینی در مدل‌های چند متغیره

فرض کنیم مشاهدات روی $t = ۱ , \dots , T$ می‌باشد و تابع $Y = X_{t.K}\,\beta + u$ را تخمین زده ایم. حال چنان چه داده‌های بیشتری روی Xها داشته باشیم و آن را $X_{t.K}$ بنامیم، دو مدل داریم:

$Y_* = X_*\,\beta + u_*$

برای آن که دارای یک پیش بینی منطقی باشیم باید در اینجا هم دارای مدلی با ساختار مشابه باشیم، پس فرض زیر ضروری است:

$$E\begin{bmatrix} u \\ u_* \end{bmatrix} (u' u_*') = E\begin{bmatrix} u\,u' & u u_*' \\ u_* u_*' & u_* u_*' \end{bmatrix} = \begin{bmatrix} \sigma^۲ I_T & ۰ \\ ۰ & \sigma^۲ I_S \end{bmatrix}$$

می‌خواهیم دو چیز را پیش بینی کنیم، اول میانگین Y_* و دوم خود Y_*

این دو را بر اساس $\hat{\beta} = (X'X)^{-۱}\,X'Y$ پیش بینی می‌کنیم. Y_*^p عبارتست از پیش مقدار Y در دوره S بر اساس X_s ها.

در اینجا ما دو خطا خواهیم داشت $e_۱$ و $e_۲$:

$$e_۱ = E(Y_*) - Y_*^p$$

$$e_۱ = E(X_*\beta) - X_*\hat{\beta} = X_*(\beta - \hat{\beta})$$

$$e_۱ = Y_* - Y_*^p = X_*(\beta - \hat{\beta}) + u_*$$

میانگین هر دو خطا برابر صفر است. به واریانس‌ها توجه کنید:

$$VC\left(e_۱\right) = VC\left[E\,(Y_*) - Y_*^p\right] = E\,[Y_*^p - E\,(Y_*^p)]\,[\,Y_*^p - E\,(Y_*^p)]'$$

$$VC\left(e_۱\right) = E\,[X_*(\beta - \hat{\beta})]\,[(\beta - \hat{\beta})'\,X_*']$$

$$VC\left(e_۱\right) = X_*\,E(\beta - \hat{\beta})\,(\beta - \hat{\beta})'\,X_*' = \sigma^۲ X_*\left(X'X\right)^{-۱} X_*'$$

$$VC\left(e_{۲}\right) = VC\,(Y_* - Y_*^p) = VC[X_*(\beta - \widehat{\beta})(\beta - \widehat{\beta}) + u_*]$$

چون u_* و $\beta - \hat{\beta}$ با یکدیگر رابطه ندارند:

$$VC\left(e_{۲}\right) = VC[X_*(\beta - \widehat{\beta})] + VC(u_*)$$

$$VC\left(e_{۲}\right) = \sigma^{۲} X_* + (X'X)^{-۱}\, X_*' + \sigma^{۲}$$

$$VC\left(e_{۲}\right) = \sigma^{۲}[X_* + (X'X)^{-۱}\, X_*' + I]$$

مثال ۱)

$$y_t = \alpha + bx_t + u_t$$

یک دوره بعد $\quad X_{۱} = ۱ \qquad X_{.} = \begin{bmatrix} ۱ & X_{T+۱} \end{bmatrix}$

$$Var\left(e_{۲}\right) = \sigma^{۲}\left[۱ + \begin{bmatrix} ۱ & X_{T+۱} \end{bmatrix} \begin{bmatrix} T & \sum x_t \\ \sum x_t & \sum x_t^{۲} \end{bmatrix} \begin{bmatrix} ۱ \\ X_{T+۱} \end{bmatrix}\right]$$

$$Var\left(e_{۲}\right) = \sigma^{۲}\left[۱ + \frac{۱}{T} + \frac{(X_{T+۱} - \bar{X})^{۲}}{\sum(x_t - \bar{X})^{۲}}\right]$$

از این رابطه نتیجه می‌گیریم هر چقدر حجم مشاهدات را زیاد کنیم، واریانس پیش بینی کننده[1] از میزان واقعی آن کمتر می‌شود.

مثال ۲)

مدل رگرسیون دو متغیره زیر را در نظر بگیرید:

$$y_t = \alpha + \beta X_t + u_t$$

به طوری که:

---

[1] Predictor

$$T = ۲۰ \quad , \qquad \sum y_t = ۱۶ \ , \qquad \sum x_t = ۸۰$$

$$\sum y_t^{۲} = ۵۲ \quad , \qquad \sum x_t^{۲} = ۴۰۰ \quad , \qquad \sum x_t y_t = ۴۰$$

چنانچه مشاهده بر روی x در دوره T+۱ برابر با ۱۰ باشد،

۱- واریانس مقدار خطای پیش بینی کننده از مقدار واقعی آن را به دست آورید.

۲- واریانس پیش بینی کننده را محاسبه کنید.

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = (X'X)^{-۱} X' Y = \begin{bmatrix} \mathrm{T} & \sum X_t \\ \sum X_t & \sum x_t^{۲} \end{bmatrix} \begin{bmatrix} \sum Y_t \\ \sum X_t Y_t \end{bmatrix} = \begin{bmatrix} ۲۰ & ۸۰ \\ ۸۰ & ۴۰ \end{bmatrix}^{-1} \begin{bmatrix} ۱۶ \\ ۴۰ \end{bmatrix}$$

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} ۲ \\ -۰/۳ \end{bmatrix},$$

برای $X_{T+۱} = ۱۰$ داریم:

$$\hat{y} = ۲ - ۰/۳ \left(۱۰\right) = -۱$$

$$\hat{y}\left(\hat{y}_f - y_f\right) = \hat{\sigma}^{۲} \left[۱ + \begin{bmatrix} ۱ & ۱۰ \end{bmatrix} \begin{bmatrix} ۰/۵ & -۰/۰۵ \\ -۰/۰۵ & ۰/۰۱۲۵ \end{bmatrix} \begin{bmatrix} ۱ \\ ۱۰ \end{bmatrix} \right] = ۰/۵ \ \hat{\sigma}^{۲}$$

برای پیش بینی کننده:

$$\widehat{var}\left(\hat{y}_f\right) = \hat{\sigma}^{۲}(X_f \left(X'X\right)^{-۱} X_f'$$

$$\hat{\sigma}^{۲} = \frac{\hat{u}'\,\hat{u}}{T-۲} = \frac{y'My}{۱۸} = \frac{۱}{۱۸} \,.\left(y'y - y'X\hat{\beta}\right); M = I - X(X'X)^{-۱} X'$$

$$\hat{\sigma}^{۲} = ۱/۷۷$$

$$\widehat{var}\left(\hat{y}_f\right) = ۱/۷۷ \ (۰/۵)$$

مثال ۳) جدول ۵ (ادامه جدول ۴)، محاسبات اضافی را برای آزمون آماری معنی دار بودن ضرایب $\hat{\beta}$ و $\hat{y}$ نشان میدهد. مقادیر $\hat{Y}_i$ از جدول ۴ با جایگذاری مقادیر X_{1i} و X_{2i} در معادله رگرسیون برآورد شده، به دست آمده است.

جدول ۵- محاسبه آزمون معنی دار بودن آثار پارامترهای مدل

i	Y	$X_۱$	$X_۱$	y	e	$e^۲$	$y^۲$
۱	۴۰	۶	۴	۴۰/۳۲	-۰/۳۲	۰/۱۰۲۴	۲۸۹
۲	۴۴	۱۰	۴	۴۲/۹۲	۱/۰۸	۱/۱۶۶۴	۱۶۹
۳	۴۶	۱۲	۵	۴۵/۳۳	۰/۶۷	۰/۴۴۸۹	۱۲۱
۴	۴۸	۱۴	۷	۴۸/۸۵	-۰/۸۵	۰/۷۲۲۵	۸۱
۵	۵۲	۱۶	۹	۵۲/۳۷	-۰/۳۷	۰/۱۳۶۹	۲۵
۶	۵۸	۱۸	۱۲	۵۷/۰۰	۱/۰۰	۱/۰۰۰۰	۱
۷	۶۰	۲۲	۱۴	۶۱/۸۲	-۱/۸۲	۳/۳۱۲۴	۹
۸	۴۸	۲۴	۲۰	۶۹/۷۸	-۱/۷۸	۳/۱۶۸۴	۱۲۱
۹	۷۴	۲۶	۲۱	۷۲/۱۹	۱/۸۱	۳/۲۷۶۱	۲۸۹
۱۰	۸۰	۳۲	۲۴	۷۹/۴۲	۰/۵۸	۰/۳۳۶۴	۵۲۹
N = ۱۰					$\sum X_۱ = ۰$	$\sum e^۲ = ۱۳/۶۷۰۴$	$\sum y^۲ = ۱/۶۳۴$

مثال ۴) محاسبه $R^۲$ برای مثال فوق که میزان برازش مدل را بررسی می‌کند عبارت است از:

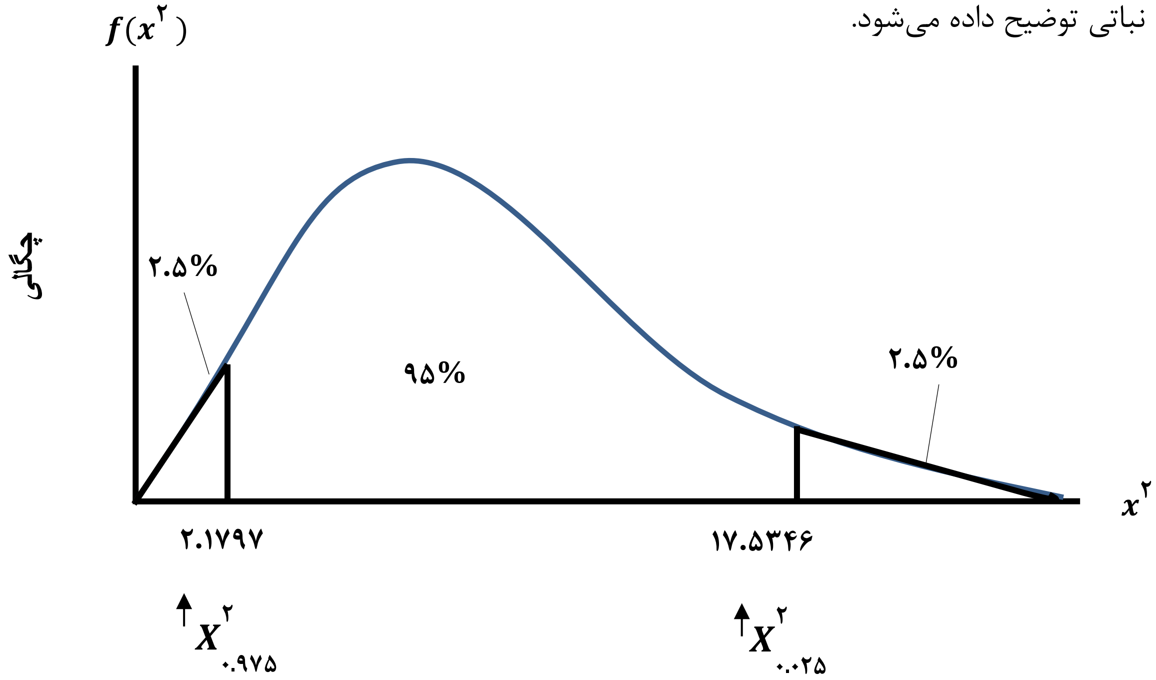
$$R^۲ = ۱ - \frac{\sum e_i^۲}{\sum y_i^۲} = ۱ - \frac{۱۳/۶۷۰۴}{۱۶۳۴} \approx ۱ - ۰/۰۰۸۴$$

$$= ۰/۹۹۱۶ \text{ یا } ۹۹/۱۶ \%$$

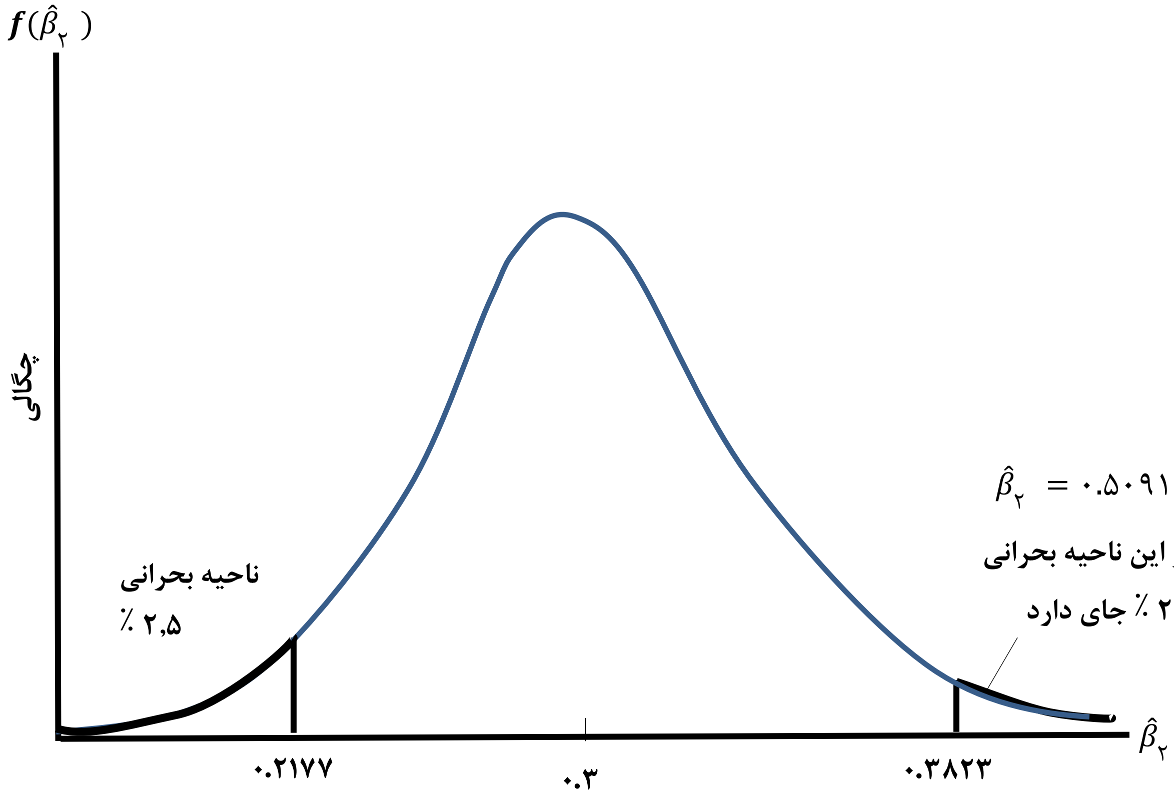
اگر $R^۲$ را با ۹۷/۱۰ درصد در مدل رگرسیون ساده مقایسه کنیم، یعنی زمانی که فقط اثر کودشیمیایی را بررسی می‌کردیم، (به عنوان یک متغیر مستقل) ملاحظه می‌شود که $R^۲$ افزایش یافته است، در حالی که ضریب تعیین تعدیل شده عبارت است از:

$$\bar{R}^{۲} = ۱ - \left(۱ - R^{۲}\right)\frac{n-۱}{n-k} = ۱ - \left(۱ - ۰/۹۹۱۶\right) - \frac{۱۰ - ۱}{۱۰ - ۳} = ۰/۹۹۱۶$$

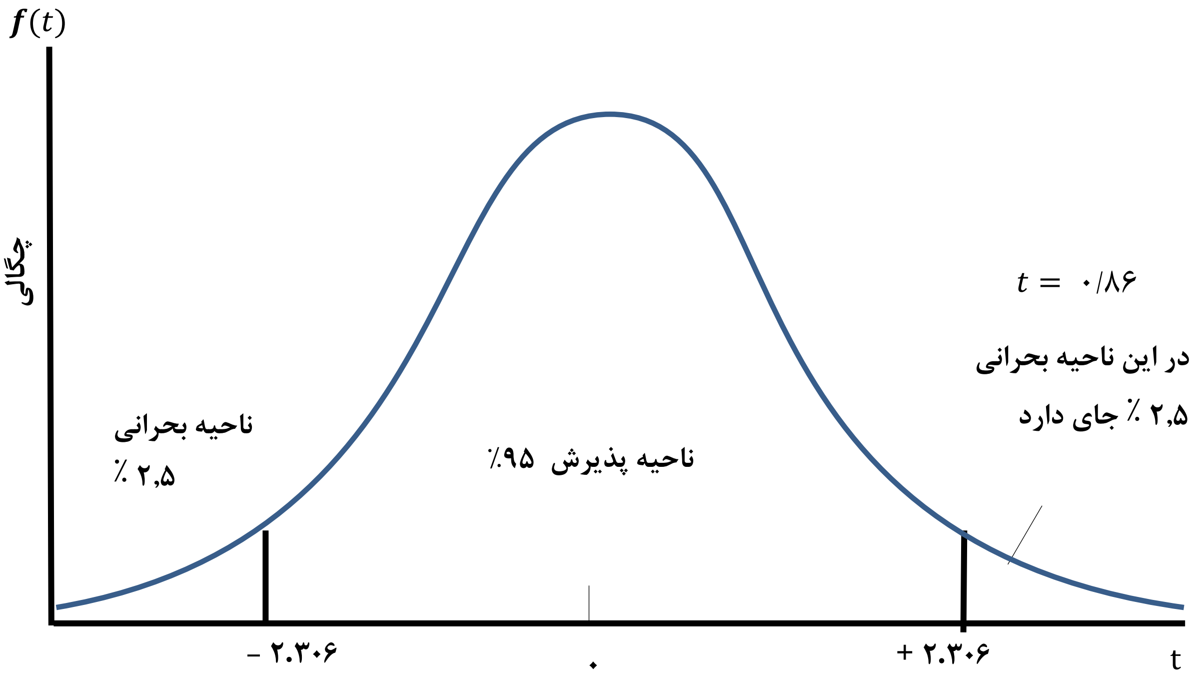
یعنی ۹۸/۹۷ درصد از تغییرات میزان برداشت محصول ذرت، توسط کودشیمیایی و اضافه نمودن سموم آفات نباتی توضیح داده می‌شود.



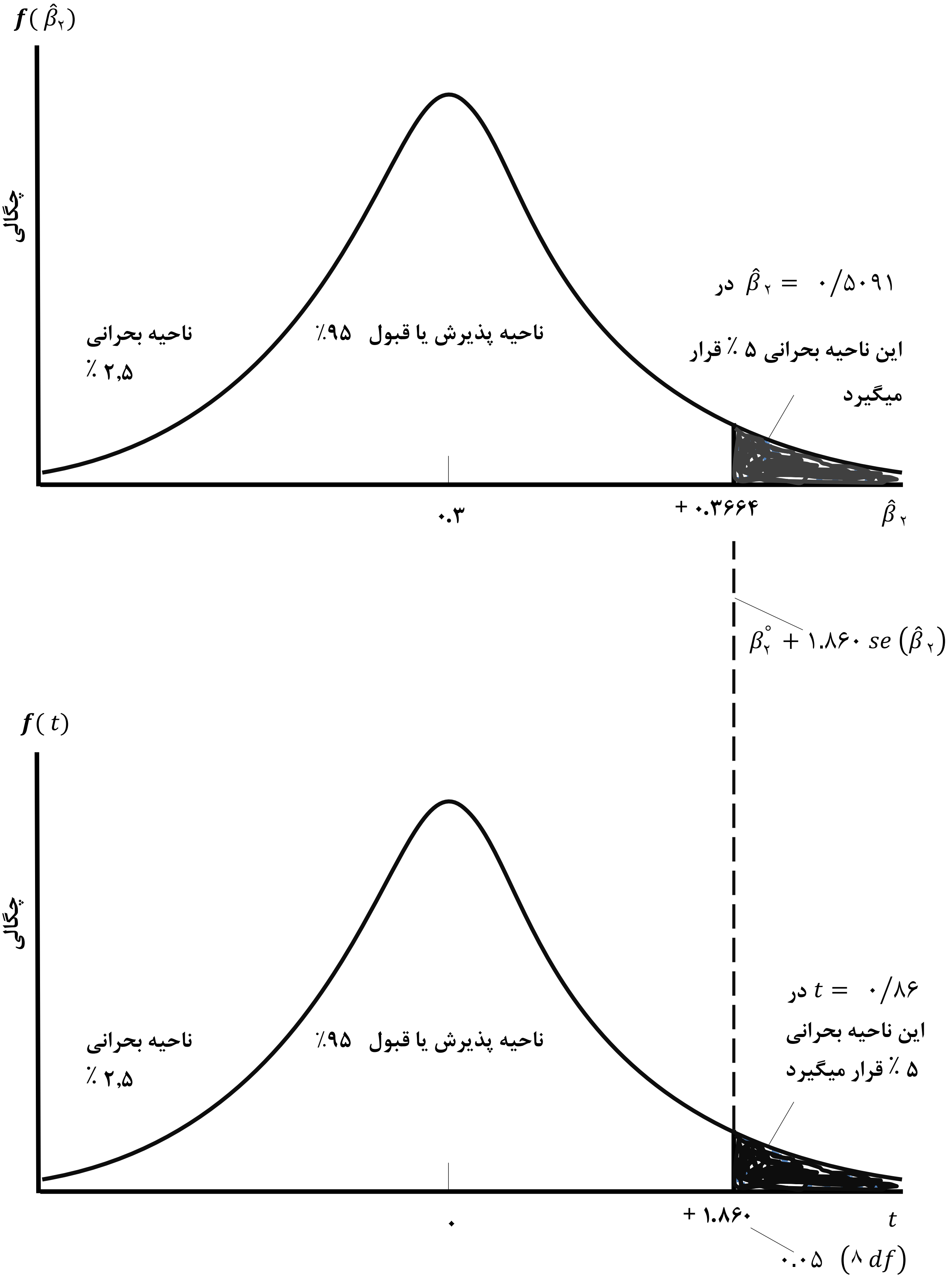
شکل ۱-۵- فاصله اطمینان ۹۵ درصد برای $x^{۲}$ (df = ۸)



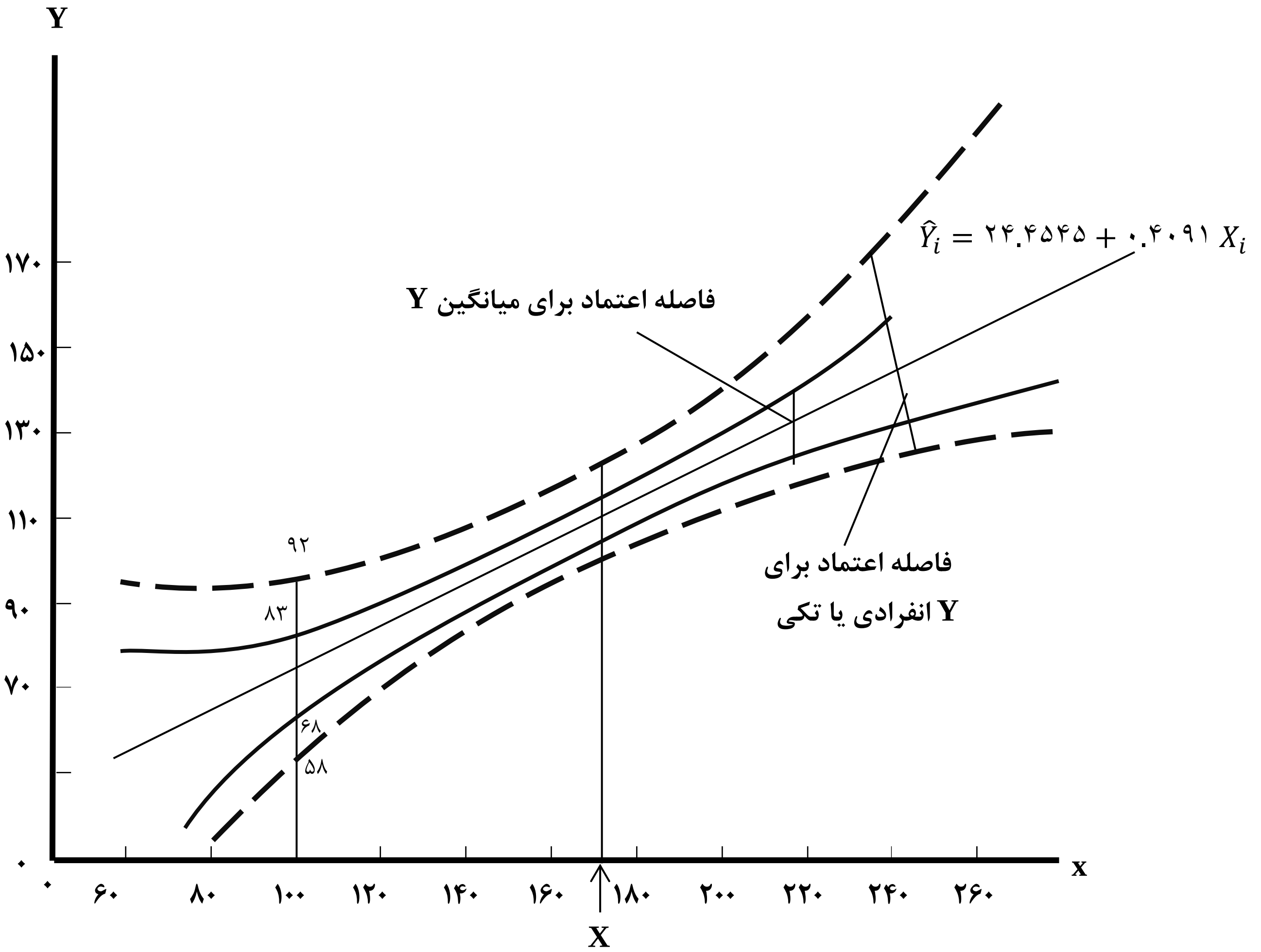
شکل ۳-۵ فاصله اطمینان ۹۵ درصد برای $\hat{\beta}_۲$ بنا بر فرضیه $\hat{\beta}_۲ = ۰.۳$



شکل ۴-۵: فاصله اطمینان ۹۵ درصد برای t با درجه آزادی ۸



شکل ۵-۵ : آزمون معنی دار بودن یک دنباله



شکل ۶-۵- فواصل (محدوده‌های) اطمینان برای میانگین Y و مقادیر تکی Y