

Introdução aos modelos de regressão com erros nas variáveis independentes

Prof. Caio Azevedo

Introdução

- Vamos nos concentrar no modelo de regressão linear simples:

$$Y_i = \beta_0 + \beta_1 x_i + \xi_i, \xi_i \overset{i.i.d.}{\sim} N(0, \sigma^2).$$

- A variável independente/explicativa (x_i) é assumida :
 - Ser não aleatória: os valores a serem observados são fixados antes do experimento ser realizado.
 - Medida sem erros: os valores observados são medidos de modo preciso.

Introdução

- Apesar de assumirmos um comportamento aleatório para a variável resposta Y_i ; ela é considerada ser medida sem erros.
- Objetivo: estudar situações em que a variável explicativa possui um comportamento aleatório ou é medida com erros (nesta última situação quase sempre considera-se uma estrutura aleatória para ela).

Exemplos

- Variável explicativa não aleatória: Exemplo 1 (consumo e carga) ; Exemplo 3 (produção de milho em função do fósforo); Exemplo 5 (tipos de solvente usados na extração de pigmentos); Exemplo 10 (resistência de materiais).
- Variável explicativa aleatória: Exemplo 4 (consumo de combustível); Exemplo 7 (censo IBGE 2000); Exemplo 8 (perfil dos clientes de uma loja), Exemplo 9 (risco de assegurar automóveis usando as covariáveis originais).

Exemplos

- Variável explicativa medida com erros: Exemplo 2 (comparação dos tipos de escova - convencional e hugger, cada criança medida duas vezes); Exemplo 6 (estudo longitudinal da eficácia das escovas, convencional e monobloco), Exemplo 9 (risco de assegurar automóveis usando as componente principais).

Voltando ao Exemplo 9

- Vamos considerar o Exemplo 9 supondo que a primeira componente principal é suficiente para os nossos objetivos.
- Consideremos, por enquanto, que essa componente principal é uma variável aleatória medida sem erros.
- Que mudanças ocorrem no modelo de regressão normal linear homocedástico e nas inferências construídas a partir dos estimadores de mínimos quadrados ordinários?

Exemplo 9: continuação

$$Y_i = \beta_0 + \beta_1 X_i + \xi_i, i = 1, \dots, 199$$

- X_i : representa o valor da componente principal para o i -ésimo carro.
- $\xi_i \stackrel{i.i.d}{\sim} N(0, \sigma^2)$, $X_i \stackrel{i.i.d}{\sim} N(\mu_x, \sigma_x^2)$ e $\xi_i \perp X_i, \forall i$. Contudo note que Y_i e X_i são dependentes para o mesmo i .
- $\mathcal{E}(Y_i | X_i = 0) = \beta_0$: risco esperado condicional de assegurar carro com valor nulo para a componentes.
- $\mathcal{E}(Y_i | X_i = x_i + 1) - \mathcal{E}(Y_i | X_i = x_i) = \beta_1$: incremento (positivo ou negativo) no risco de assegurar esperado e condicional, para o aumento em uma unidade na componente.

Exemplo 9: continuação

- Estimadores de mínimos quadrados:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}; \hat{\beta}_1 = \sum_{i=1}^n C_i Y_i; C_i = \frac{X_i - \bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

- Note que agora os estimadores dependem de dois conjuntos de variáveis aleatórias $(\mathbf{Y}', \mathbf{X}')'$, $\mathbf{Y} = (Y_1, \dots, Y_n)'$ e $\mathbf{X} = (X_1, \dots, X_n)'$.
- Resultados importantes : $\mathcal{E}_X(X) = \mathcal{E}_Y(\mathcal{E}_{X|Y}(X|Y))$ e $\mathcal{V}(X) = \mathcal{E}_X(\mathcal{V}_{X|Y}(X|Y)) + \mathcal{V}_X(\mathcal{E}_{X|Y}(X|Y))$.

Exemplo 9: continuação

- Temos que :

- $\mathcal{E}(\hat{\beta}_1|\mathbf{X} = \mathbf{x}) = \beta_1$ e $\mathcal{V}(\hat{\beta}_1|\mathbf{X} = \mathbf{x}) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$.

- $\mathcal{E}(\hat{\beta}_0|\mathbf{X} = \mathbf{x}) = \beta_0$ e $\mathcal{V}(\hat{\beta}_0|\mathbf{X} = \mathbf{x}) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$.

- Portanto, facilmente prova-se que $\mathcal{E}(\hat{\beta}_1) = \mathcal{E}(\mathcal{E}(\hat{\beta}_1|\mathbf{X} = \mathbf{x})) = \beta_1$ e $\mathcal{E}(\hat{\beta}_0) = \mathcal{E}(\mathcal{E}(\hat{\beta}_0|\mathbf{X} = \mathbf{x})) = \beta_0$

Exemplo 9: continuação

- Contudo, temos que:

$$\begin{aligned}\mathcal{V}(\hat{\beta}_1) &= \mathcal{E}(\mathcal{V}(\hat{\beta}_1|\mathbf{X})) + \mathcal{V}(\mathcal{E}(\hat{\beta}_1|\mathbf{X})) \\ &= \sigma^2 \mathcal{E}\left(\frac{1}{\sum_{i=1}^n (X_i - \bar{X})^2}\right) + \mathcal{V}(\beta_1) \\ &= \sigma^2 \mathcal{E}\left(\frac{1}{S_X}\right)\end{aligned}$$

- Mas, $W_X = \frac{S_X}{\sigma_X^2} \sim \chi_{(n-1)}^2$. Assim prova-se que $\mathcal{E}(W_X^{-1}) = 1/(n-3)$.
Logo $\mathcal{E}(S_X^{-1}) = 1/(\sigma_X^2(n-3))$.
- Portanto, $\mathcal{V}(\hat{\beta}_1) = \frac{\sigma^2}{\sigma_X^2(n-3)} \neq \mathcal{V}(\hat{\beta}_1|\mathbf{X} = \mathbf{x})$.

Exemplo 9: continuação

- Analogamente, temos que:

$$\begin{aligned} \mathcal{V}(\hat{\beta}_0) &= \mathcal{E}(\mathcal{V}(\hat{\beta}_0|\mathbf{X})) + \mathcal{V}(\mathcal{E}(\hat{\beta}_0|\mathbf{X})) \\ &= \sigma^2 \left[\frac{1}{n} + \mathcal{E} \left(\frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right) \right] + \mathcal{V}(\beta_0) \\ &= \sigma^2 \left[\frac{1}{n} + \mathcal{E} \left(\frac{\bar{X}^2}{S_X} \right) \right] \end{aligned}$$

- Mas, nesse caso, é possível provar que $\bar{X} \perp (X_i - \bar{X}), \forall i$. Assim, \bar{X}^2 e $\sum_{i=1}^n (X_i - \bar{X})^2$ também o são.

Exemplo 9: continuação

- Portanto, $\mathcal{E}\left(\frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2}\right) = \mathcal{E}(\bar{X}^2)\mathcal{E}\left(\frac{1}{S_X}\right)$.
- Mas, $\mathcal{E}(\bar{X}^2) = \mathcal{V}(\bar{X}) + \mathcal{E}(\bar{X})^2 = \sigma_x^2/n + \mu_x^2$.
- Logo, $\mathcal{V}(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\sigma_x^2 + \mu_x^2 n}{n\sigma_x^2(n-3)}\right)$.
- A obtenção das distribuições exatas de $\hat{\beta}_0$ e $\hat{\beta}_1$ tornam-se complicadas.

Exemplo 9: continuação

- Contudo, podemos usar resultados assintóticos. Com efeito, para n suficientemente grande, temos que:

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{\hat{\sigma}_X^2 + \hat{\mu}_X^2 n}{n \hat{\sigma}_X^2 (n-3)} \right]}} \approx N(0, 1); \quad \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\hat{\sigma}_X^2 (n-3)}}} \approx N(0, 1)$$

em que $\hat{\mu}_X = \bar{X}$; $\sigma_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$;
 $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$.

- Logo, inferências para cada parâmetro podem ser feitas utilizando-se os resultados acima.

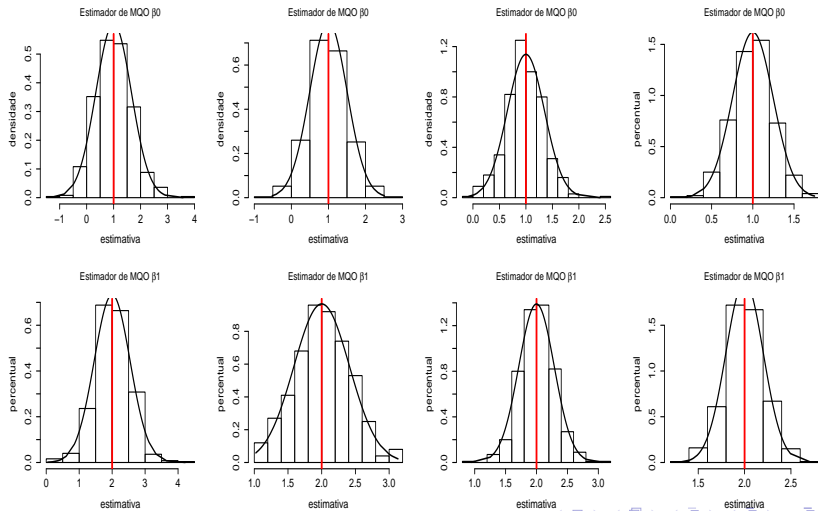
Estudo de simulação

- Modelo : $Y_i = 1 + 2x_i + \xi_i, i = 1, 2, \dots, n, x_i \overset{i.i.d.}{\sim} N(1, 0.5)$ e $\xi_i \overset{i.i.d.}{\sim} N(0, 4)$.
- $n \in \{30, 50, 100, 200\}$.
- OBS: a diferença entre o que está sendo feito aqui e o que é feito comumente em simulação de modelos de regressão é que a variável explicativa é simulada em cada réplica (em geral ela é simulada somente uma vez).

n	Para.	M.Est.	Var.	M. Var.	Var. Verd.
30	β_0	1,001	0,436	0,438	0,435
	β_1	2,022	0,285	0,302	0,296
50	β_0	0,982	0,259	0,252	0,252
	β_1	2,020	0,168	0,169	0,170
100	β_0	0,980	0,122	0,124	0,123
	β_1	2,010	0,076	0,084	0,082
200	β_0	0,997	0,062	0,061	0,061
	β_1	2,001	0,041	0,041	0,041

M. Est.: média das estimativas entre as réplicas, Var.: variância das estimativas entre as réplicas, M. Var.: média das variâncias do estimador de MQO entre as réplicas, Var. Verd.: variância verdadeira do estimador de MQO.

Resultados (n=30,50,100,200)



Resultados para dos dados do risco de assegurar carros

MQO usual

Param.	Estim.	EP	IC(95%)	Estat. t	p-valor
β_0	0,789	0,084	[0,623 ; 0,955]	9,349	<0,0001
β_1	0,109	0,032	[0,045 ; 0,173]	3,357	0,0009

MQO com as variâncias verdadeiras

Parm.	Estim.	Ep	IC(95%)	Estat. Z	p-valor
β_0	0,789	0,084	[0,623 ; 0,955]	9,325	<0,0001
β_1	0,109	0,032	[0,045 ; 0,173]	3,340	0,0008

Modelos com erros nas variáveis explicativas

$$Y_i = \beta_0 + \beta_1 w_i + \xi_i, i = 1, \dots, n$$

- X_i : representa o valor da covariável para a i -ésima observação.
- $\xi_i \stackrel{i.i.d}{\sim} N(0, \sigma^2)$, $X_i = w_i + \epsilon_i$, $\epsilon_i \stackrel{i.i.d}{\sim} N(0, \sigma_\epsilon^2)$ e $\xi_i \perp \epsilon_j, \forall i, j$, em que w_i é o valor verdadeiro (não observável) e X_i é o valor observável (medido com erros) da covariável.
- Modelo funcional: admite que $w_i, i = 1, \dots, n$ são desconhecidos mas não aleatórios.
- Modelo estrutural: admite que W_i são desconhecidos e aleatórios.
- Focaremos no modelo estrutural.

Modelos com erros nas variáveis explicativas

- Sob qualquer modelo (funcional/estrutural) temos que:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1(X_i - \epsilon_i) + \xi_i = \beta_0 + \beta_1 X_i + (\xi_i - \beta_1 \epsilon_i) \\ &= \beta_0 + \beta_1 X_i + v_i \end{aligned}$$

em que $v_i = (\xi_i - \beta_1 \epsilon_i)$. Portanto, a estrutura acima não corresponde mais ao modelo de regressão linear homocedástico estudado até o momento.

- Assim, consoante à modelagem adotada (funcional ou estrutural), novas abordagens inferenciais precisam ser utilizadas.

Modelo funcional

$$Y_i = \beta_0 + \beta_1 w_i + \xi_i, i = 1, \dots, n$$

- $\xi_i \stackrel{i.i.d}{\sim} N(0, \sigma^2)$, $X_i = w_i + \epsilon_i$, $\epsilon_i \stackrel{i.i.d}{\sim} N(0, \sigma_\epsilon^2)$ e $\xi_i \perp \epsilon_j, \forall i, j$, em que w_i é o valor verdadeiro (não observável) e X_i é o valor observável (medido com erros).
- $w_i : i = 1, \dots, n$ são desconhecidos mas não aleatórios. Logo, $X_i \stackrel{i.i.d}{\sim} N(x_i, \sigma_\epsilon^2)$.
- Nesse caso $\theta = (\beta_0, \beta_1, \sigma^2, \sigma_\epsilon^2, \mathbf{w}')'$, em que $\mathbf{w} = (w_1, \dots, w_n)'$ são os parâmetros de interesse. Note que o aumento do tamanho da amostra implica no aumento do número de parâmetros a serem estimados.

Modelos funcional

- Verossimilhança (note que, nesse caso, $Y_i \perp X_i$):

$$L(\theta) =$$

$$-\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln \sigma_\epsilon^2 - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^n (x_i - w_i)^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

- A obtenção das estimativas de MV tornam-se inviáveis a não ser que algumas suposições adicionais acerca dos parâmetros sejam feitas.
- Pode-se também utilizar adaptações de outros métodos de estimação como mínimos quadrados ponderados.

Modelo estrutural

$$Y_i = \beta_0 + \beta_1 W_i + \xi_i, i = 1, \dots, n$$

- $\xi_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$, $X_i = W_i + \epsilon_i$, $\epsilon_i \stackrel{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$, em que W_i é o valor verdadeiro (não observável) e X_i é o valor observável (medido com erros).
- Modelo funcional, admite que $W_i, i = 1, \dots, n$ são desconhecidos e aleatórios. Assumiremos que $W_i \stackrel{i.i.d.}{\sim} N(\mu_W, \sigma_W^2)$ e $W_i \perp \epsilon_j \perp \xi_k, \forall i, j, k$. Assim, $X_i \stackrel{i.i.d.}{\sim} N(\mu_W, \sigma_X^2)$, em que $\sigma_X^2 = \sigma_W^2 + \sigma_\epsilon^2$.
- Nesse caso, $\theta = (\mu_W, \beta_0, \beta_1, \sigma^2, \sigma_\epsilon^2, \sigma_W^2)'$ são os parâmetros de interesse. Note que o aumento do tamanho da amostra não implica

Modelo estrutural

- Nesse caso, temos que:

$$\begin{pmatrix} Y_i \\ X_i \end{pmatrix} \sim N_2 \left[\begin{pmatrix} \beta_0 + \beta_1 \mu_W \\ \mu_W \end{pmatrix}; \begin{pmatrix} \beta_1^2 \sigma_W^2 + \sigma^2 & \beta_1 \sigma_W^2 \\ \beta_1 \sigma_W^2 & \sigma_W^2 + \sigma_\epsilon^2 \end{pmatrix} \right]$$

- Primeiramente, podemos provar que $\hat{\beta}_{1(MQO)} \xrightarrow[n \rightarrow \infty]{P} \beta_1 \frac{\sigma_W^2}{\sigma_X^2}$.
- Como a verossimilhança é a função conjunta de $(Y_1, X_1)', \dots, (Y_n, X_n)'$, temos um problema de falta de identificabilidade.
- Assim, a menos que se coloquem restrições nos parâmetros, não é possível obter estimativas de máxima verossimilhança.

Modelo estrutural

- Estimadores de mínimos quadrados $\hat{\beta}_{1(MQO)} = \frac{\sum_{i=1}^n (X_i - \bar{X}) Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2}$ e $\hat{\beta}_{0(MQO)} = \bar{Y} - \hat{\beta}_{1(MQO)} \bar{X}$.
- Estimadores de máxima verossimilhança. Maximizar

$$l(\theta) = -\frac{1}{2} \sum_{i=1}^n (\mathbf{z}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{z}_i - \boldsymbol{\mu}) - \frac{n}{2} \ln |\boldsymbol{\Sigma}|.$$

em que $\mathbf{z}_i = (y_i, x_i)'$, $\boldsymbol{\mu} = (\beta_0 + \beta_1 \mu_X, \mu_X)'$ e

$$\boldsymbol{\Sigma} = \begin{pmatrix} \beta_1^2 \sigma_W^2 + \sigma^2 & \beta_1 \sigma_W^2 \\ \beta_1 \sigma_W^2 & \sigma_W^2 + \sigma_\epsilon^2 \end{pmatrix}.$$

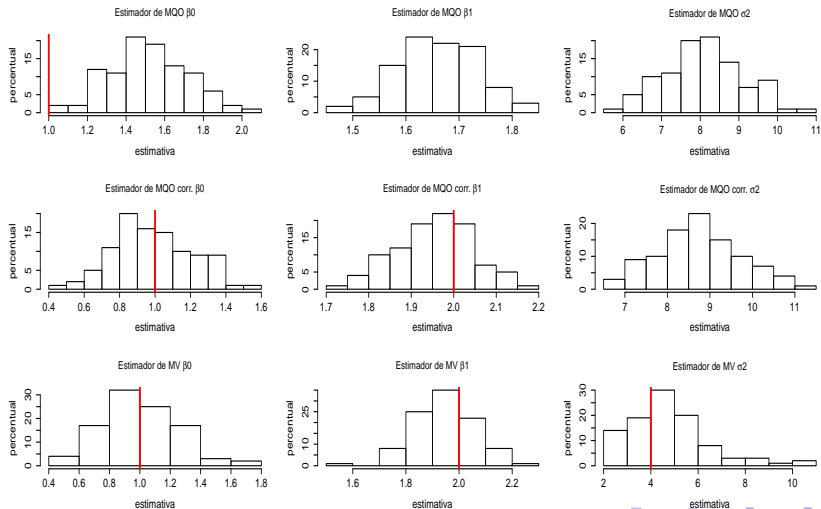
Inferência com $\frac{\sigma_W^2}{\sigma_X^2}$ conhecido

- Vamos discutir o processo de inferência considerando o fator de atenuação $\frac{\sigma_W^2}{\sigma_X^2}$ conhecido, ou seja $\frac{\sigma_W^2}{\sigma_X^2} = k, k \in (0, 1)$, pois $\sigma_X^2 = \sigma_W^2 + \sigma_\epsilon^2$.
- Assim, prova-se que $\sigma_W^2 = \frac{k}{1-k} \sigma_\epsilon^2$.
- Podemos provar também que $\mathcal{E}(\hat{\beta}_{1(MQO)} | \mathbf{X} = \mathbf{x}) = \beta_1 k$ e $\mathcal{V}(\hat{\beta}_{1(MQO)} | \mathbf{X} = \mathbf{x}) = \frac{\sigma_Y^2 - \beta_1 \sigma_{XY}}{\sum_{i=1}^n (X_i - \bar{X})^2}$, em que $\sigma_Y^2 = \mathcal{V}(Y)$ e $\sigma_{XY} = \text{Cov}(X, Y)$.
- Logo, $\hat{\beta}_{1(C)} = k^{-1} \hat{\beta}_{1(MQO)}$ é um estimador não viciado de β_1 . Analogamente, $\hat{\beta}_{0(C)} = \bar{Y} - \hat{\beta}_{1(C)} \bar{X}$ é um estimador não viciado de β_0 .

Inferência com $\frac{\sigma_W^2}{\sigma_\epsilon^2}$ conhecido

- Vamos comparar os estimadores de MQO, MQOC (mqo corrigidos) e MV de $(\beta_0, \beta_1, \sigma^2)'$ através de um estudo de simulação. No caso da estimação por MV, também estimou-se $(\mu_W, \sigma_\epsilon^2)'$.
- Modelo $Y_i = \beta_0 + \beta_1 W_i + \xi_i, i = 1, 2; j = 1, \dots, 100.$,
 $X_i = W_i + \epsilon_i, \xi_i \stackrel{i.i.d}{\sim} N(0, \sigma^2), W_i \stackrel{i.i.d}{\sim} N(0, \sigma_W^2)$ e
 $\epsilon_i \stackrel{i.i.d}{\sim} N(0, \sigma_\epsilon^2), \xi_i \perp W_j \perp \epsilon_k, \forall i, j, k.$
- $\beta_0 = 1, \beta_1 = 2, k = 0,85 \sigma^2 = 4, \sigma_\epsilon^2 = 1$ e $\sigma_W^2 \approx 5,67.$
- Foram geradas 100 réplicas de tamanho 100.

Resultados



Mét. de est.	Par.	Valor verd.	Média	Vício	Var.	RQEQM
MQO	β_0	1,00	1,514	0,514	0,042	0,553
	β_1	2,00	1,660	-0,340	0,006	0,348
	σ^2	4,00	8,099	4,099	1,016	4,221
MQO corr,	β_0	1,00	0,987	-0,013	0,046	0,216
	β_1	2,00	1,953	-0,047	0,008	0,101
	σ^2	4,00	8,753	4,753	1,014	4,858
MV	β_0	1,00	1,006	0,006	0,064	0,253
	β_1	2,00	1,945	-0,055	0,012	0,121
	σ^2	4,00	4,772	0,772	2,805	1,844

Resultados para dos dados do risco de assegurar carros

MQO usual

Param.	Estim.	EP	IC(95%)	Estat. t	p-valor
β_0	0,789	0,084	[0,623 ; 0,955]	9,349	<0,0001
β_1	0,109	0,032	[0,045 ; 0,173]	3,357	0,0009

MV sob o modelo com erro nas variáveis e $k = 0,85$

Parm.	Estim.	Ep	IC(95%)	Estat. Z	p-valor
β_0	0,805	0,083	[0,643 ; 0,968]	9,698	<0,0000
β_1	0,123	0,036	[0,053 ; 0,192]	3,454	0,0006

Comentário sobre ausência de normalidade dos erros

- Voltemos ao MNL:

$$\mathbf{Y}_n = \mathbf{X}_n\boldsymbol{\beta} + \boldsymbol{\xi}_n$$

- Suponha que $\mathcal{E}(\boldsymbol{\xi}_n) = \mathbf{0}_n$ e $\text{Cov}(\boldsymbol{\xi}_n) = \sigma^2\mathbf{I}_n$.

- Suposições adicionais

1 $\max_{1 \leq k \leq n} \mathbf{x}'_{nk} (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{x}_{nk} \rightarrow 0$, quando $n \rightarrow \infty$.

2 $\lim_{n \rightarrow \infty} \mathbf{X}'_n \mathbf{X}_n = \mathbf{V}$.

- Pelo Teorema Central do Limite (TCL) de Linderberg-Feller (se $n \rightarrow \infty$), então

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\mathbf{0}, \sigma^2 \mathbf{V}^{-1})$$

Comentário

- Para n suficientemente grande

$$\hat{\beta} \approx N_p(\beta, \sigma^2(\mathbf{X}_n\mathbf{X}_n)^{-1})$$

- Em relação à primeira suposição adicional, considere que

$$\mathbf{X}_n = \begin{pmatrix} \mathbf{x}'_{n_1} \\ \mathbf{x}'_{n_2} \\ \vdots \\ \mathbf{x}'_{n_n} \end{pmatrix}; \mathbf{X}_n = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} \end{pmatrix}; \mathbf{X}'_n\mathbf{X}_n = \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix}$$

- Assim

$$\mathbf{x}'_{n_1}(\mathbf{X}'_n\mathbf{X}_n)^{-1}\mathbf{x}_{n_1} = n_1^{-1}; \mathbf{x}'_{n_2}(\mathbf{X}'_n\mathbf{X}_n)^{-1}\mathbf{x}_{n_2} = n_2^{-1}$$