REGRESSÃO LINEAR MÚLTIPLA

renda = $\beta_1 + \beta_2$ educação + ε $\quad\rightarrow\quad$ regressão linear simples

renda = $\beta_1 + \beta_2$ educação + β_3 idade + β_4 idade^2 + ε $\quad\rightarrow\quad$ regressão linear múltipla

Forma geral:

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_K x_{iK} + \varepsilon_i \qquad i = 1, \ldots, n$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = y \qquad \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix} = x_1 \qquad \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \varepsilon$$

$$y = x_1\beta_1 + x_2\beta_2 + \ldots + x_K\beta_K + \varepsilon$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & & x_{2K} \\ x_{31} & x_{32} & & x_{3K} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & & x_{nK} \end{bmatrix}_{n \times K} \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}_{K \times 1}$$

pode ser uma coluna de 1's

$$y = X\beta + \varepsilon$$

<u>Pressuposições do Modelo Linear:</u>

<u>Pressuposição 1</u> – *y* é uma função linear (nos parâmetros) de um conjunto específico de variáveis *X* mais um erro.

Violação: omissão de variável relevante, inclusão de variável irrelevante e não linearidade.

<u>Pressuposição 2</u> – a esperança do erro é zero

Violação: tendenciosidade da constante

$$y = \alpha + \beta X + \varepsilon \qquad E(\varepsilon) = c$$

$$y = \alpha + \beta X + \varepsilon + c - c$$

$$y = \alpha + c + \beta X + \varepsilon - c$$

$$y = \alpha^* + \beta X + \varepsilon^* \qquad E(\varepsilon^*) = 0$$

$$E(\varepsilon) = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \vdots \\ E(\varepsilon_n) \end{bmatrix} = 0$$

o que implica em,

$$E(y) = X\beta$$

<u>Pressuposição 3</u> – a variância do erro é constante,isto é, não depende de X (homoscedasticia), e não há correlação entre os erros (não autocorrelação).

$$\varepsilon\varepsilon' = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n \end{bmatrix} = \begin{bmatrix} \varepsilon_1\varepsilon_1 & \varepsilon_1\varepsilon_2 & \cdots & \varepsilon_1\varepsilon_n \\ \varepsilon_2\varepsilon_1 & \varepsilon_2\varepsilon_2 & & \varepsilon_2\varepsilon_n \\ \vdots & & & \\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \cdots & \varepsilon_n\varepsilon_n \end{bmatrix}$$

$$E(\varepsilon\varepsilon') = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 I$$

Pressuposição 4 – não há correlação entre o erro e as variáveis *x*.

Violação: equações simultâneas, erros em variáveis.

$$E(\varepsilon | X) = 0 \Rightarrow \operatorname{cov}(X, \varepsilon) = 0$$

$$C = \alpha + \beta Y + \varepsilon$$

$$Y = C + I$$

<u>Pressuposição 5</u> – X é uma matriz ($n \times K$) não estocástica com rank K.

Isto implica que devemos ter pelo menos K observações, e que as colunas de X devem ser linearmente independentes.

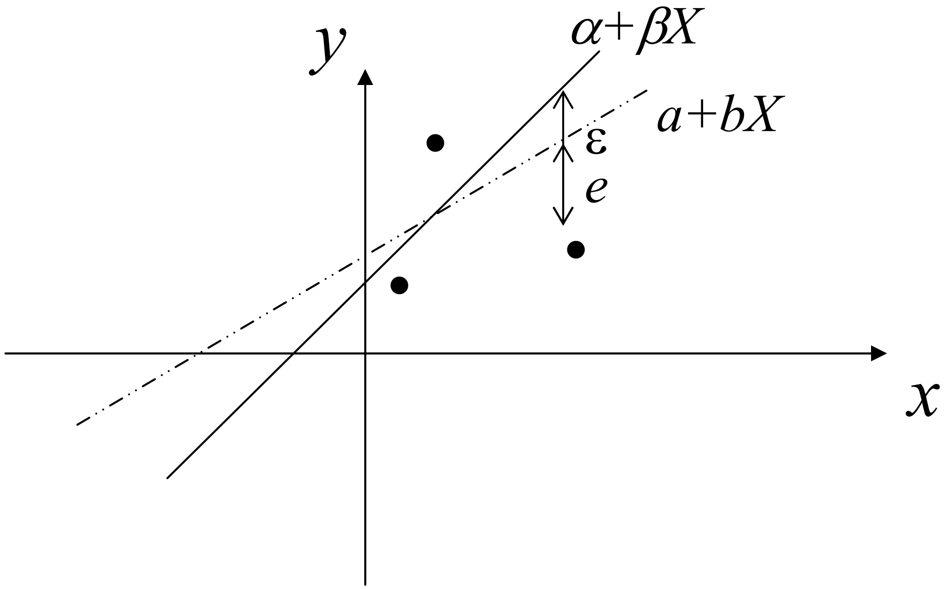
Violação: multicolinearidade.

<u>Pressuposição 6</u> – o erro tem distribuição normal

$$\varepsilon \sim N[0, \sigma^2 I]$$

Vetor de Coeficientes por Mínimos Quadrados.

Minimizar a soma de quadrados dos erros, isto é, minimizar



$$S = \varepsilon'\varepsilon$$

$$S = (y - X\beta)'(y - X\beta)$$

$$S = y'y - y'X\beta - \beta'X'y + \beta'X'X\beta$$

$$S = y'y - 2\beta'X'y + \beta'X'X\beta$$

Condição necessária para mínimo:

$$\frac{\partial S}{\partial \beta} = -2X'y + 2X'X\beta = 0$$

Deixe *b* ser a solução, então *b* satisfaz a equação acima.

$$X'Xb = X'y$$

Da pressuposição de rank total, $(X'X)^{-1}$ existe. Então,

$$\boxed{b = (X'X)^{-1}X'y}$$

Aspectos Algébricos:

As equações normais são:

$$X'Xb - X'y = -X'(y - Xb) = -X'e = 0$$

Isto significa que para cada coluna x_k de X,

$$x_k' e = 0$$

Em particular, se a primeira coluna de X é uma coluna de 1's, a soma dos resíduos será zero.

Implicações:

1) O hiperplano de regressão passa pelas médias dos dados.

$$\bar{y} = b'\overline{X}$$

$$y = \underbrace{a + bx}_{\hat{y}} + e$$

$$\sum y = na + b\sum x + \sum e$$

$$\bar{y} = a + b\bar{x}$$

2) A média do valor ajustado da regressão iguala a média do valor atual, isto é,

$$\bar{\hat{y}} = \bar{y}$$

$$\hat{y} = a + bx$$

$$\sum \hat{y} = na + b\sum x$$

$$\bar{\hat{y}} = a + b\bar{x} = \bar{y}$$

Isto segue de 1, pois os valores ajustados são:

$$\hat{y} = Xb$$

Estes resultados não precisam necessariamente ocorrer em modelos sem constante.

Por definição, o vetor de resíduos de mínimos quadrados é:

$$e = y - Xb = y - X(X'X)^{-1}X'y$$

$$= (I - X(X'X)^{-1}X')y = My$$

M é uma matriz ($n \times n$), simétrica ($M = M'$) e idempotente ($M = MM$).

M pode ser interpretada como uma matriz que quando prémultiplica qualquer vetor *y*, produz um vetor de resíduos da regressão de *y* em função de *X*. Segue imediatamente que:

$$MX = (I - X(X'X)^{-1}X')X = X - X\underbrace{(X'X)^{-1}X'X}_{I} = 0$$

uma forma de interpretar este resultado é que a regressão de X em função de X fornece um ajustamento perfeito e, portanto, os resíduos serão zero.

Formas úteis de soma de quadrado dos resíduos:

$$e'e = y'MMy = y'My = y'e$$

$$e'e = y'e = y'(y - Xb) = y'y - y'Xb = y'y - b'X'y$$

$$e'e = y'y - b'X'Xb$$

$$e'e = (y - Xb)'(y - Xb) = y'y - 2b'X'\underbrace{(Xb + e)}_{y} + b'X'Xb = y'y - 2b'X'Xb - 2b'\cancelto{0}{X'e} + b'X'Xb$$

Regressão Parcial

Que cálculos estão envolvidos para obter, isoladamente, os coeficientes de um conjunto de variáveis na regressão múltipla? Por exemplo, o coeficiente de educação, na regressão de renda em função de educação e idade.

Suponha que a regressão envolve dois conjuntos de variáveis, X_1 e X_2. Então,

$$y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

Qual é a solução algébrica de b_2? As equações normais são:

$$\begin{bmatrix} X_1' \\ X_2' \end{bmatrix} \begin{bmatrix} X_1 & X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}$$

$$(X_1'X_1)b_1 + (X_1'X_2)b_2 = X_1'y \qquad \rightarrow$$

$$b_1 = (X_1'X_1)^{-1}X_1'y - (X_1'X_1)^{-1}X_1'X_2b_2$$

$$b_1 = (X_1'X_1)^{-1}X_1'(y - X_2b_2)$$

$$(X_2'X_1)b_1 + (X_2'X_2)b_2 = X_2'y$$

$$(X_2'X_1)(X_1'X_1)^{-1}X_1'y - (X_2'X_1)(X_1'X_1)^{-1}X_1'X_2b_2 + (X_2'X_2)b_2 = X_2'y$$

$$(X_2'X_2)b_2 - (X_2'X_1)(X_1'X_1)^{-1}(X_1'X_2)b_2 = X_2'y - (X_2'X_1)(X_1'X_1)^{-1}X_1'y$$

$$X_2'(I - X_1(X_1'X_1)^{-1}X_1')X_2b_2 = X_2'(I - X_1(X_1'X_1)^{-1}X_1')y$$

$$b_2 = \left[X_2'(I - X_1(X_1'X_1)^{-1}X_1')X_2\right]^{-1} X_2'(I - X_1(X_1'X_1)^{-1}X_1')y$$

$$b_2 = \left[X_2'M_1X_2\right]^{-1} X_2'M_1y$$

Neste caso, M_1X_2 é uma matriz de resíduos. Cada coluna de M_1X_2 é um vetor de resíduos da regressão da correspondente coluna de X_2 em função das variáveis em X_1

Como M_1 é uma matriz simétrica e idempotente podemos escrever.

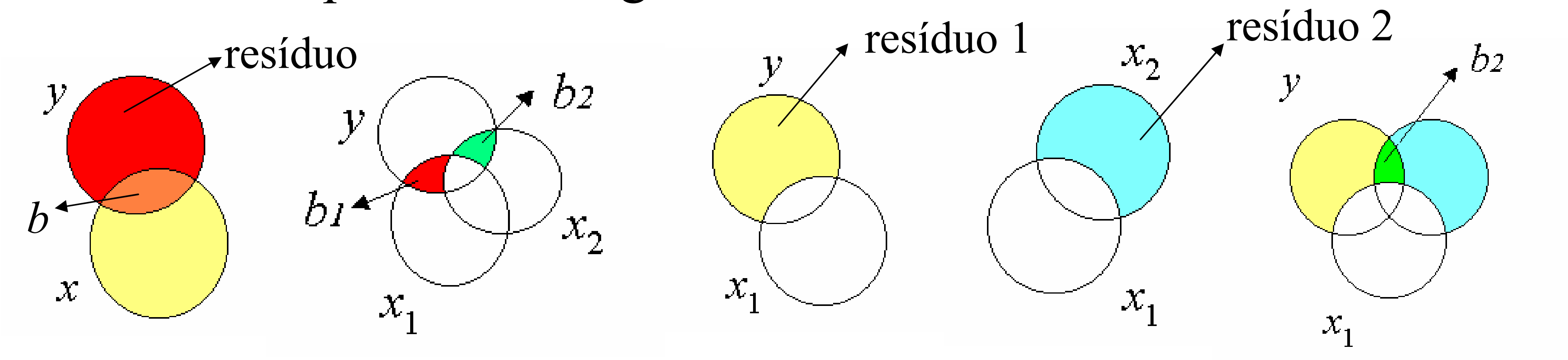
$$b_2 = \left[X_2'M_1'M_1X_2\right]^{-1}\left[X_2'M_1'M_1y\right]$$

$$b_2 = (X_2^{*}{}'X_2^{*})^{-1}(X_2^{*}{}'y^{*})$$

Portanto, b_2 é o conjunto de coeficientes obtido quando é feita a regressão dos resíduos da regressão de y em função de X_1 com os resíduos da regressão de X_2 em função de X_1

No exemplo, poderíamos obter o coeficiente de educação fazendo a regressão de renda em função da idade (resíduo 1) e de educação em função da idade (resíduo 2) e, então, fazendo a regressão do resíduo 1 em função do resíduo 2.

Este processo é freqüentemente chamado de partição e por esta razão os coeficientes da regressão múltipla são chamados de coeficientes parciais de regressão.



Propriedades Estatísticas dos Estimadores

$$b = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon$$

$$b = \beta + (X'X)^{-1}X'\varepsilon = \beta + A\varepsilon = Ay$$

onde $\qquad A = (X'X)^{-1}X'$

Se X é não estocástico ou se E($X'\varepsilon$) = 0 então E(b) = β

Portanto b é um estimador não tendencioso de β.

$$Var(b) = E\left[(b-\beta)(b-\beta)'\right] = E\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right]$$

$$= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1} = (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

$$Var(b) = E(A\varepsilon)(A\varepsilon)' = E(A\varepsilon\varepsilon'A') = AE(\varepsilon\varepsilon')A' = \sigma^2 AA'$$

Pelo teorema de Gauss-Markov b é o estimador linear não tendencioso mais eficiente de β.

Prova:

Deixe $\tilde{b} = Cy$ ser qualquer outro estimador linear não

tendencioso de β, onde C é uma matriz $(K \times n)$. Se é não tendencioso, então,

$$E(Cy) = E(CX\beta + C\varepsilon) = \beta$$

o que implica $CX = I$

$$Var(\tilde{b}) = E\left[(\tilde{b} - \beta)(\tilde{b} - \beta)'\right] = E\left[C\varepsilon\varepsilon'C'\right] = \sigma^2 CC'$$

$$C = A + (C - A)$$

$$\sigma^2 CC' = \sigma^2 (A + (C - A))(A + (C - A))'$$

$$\sigma^2 CC' = \sigma^2 AA' + \sigma^2 A(C - A)' + \sigma^2 (C - A)A' + \sigma^2 (C - A)(C - A)'$$

$$(C-A)A' = [C-(X'X)^{-1}X']X(X'X)^{-1} = CX(X'X)^{-1} - (X'X)^{-1} = (\cancelto{I}{CX} - I)(X'X)^{-1} = 0$$

Portanto,

$$\sigma^2 CC' = \sigma^2 AA' + \sigma^2 (C-A)(C-A)'$$

Var($\widetilde{b}$) = Var(b) + matriz semi-definida positiva

Portanto,

$$\mathrm{Var}(\widetilde{b}) \geq \mathrm{Var}(b)$$

Testando hipóteses sobre os coeficientes.

Como b é uma função linear de ε. Se pressupormos que ε tem distribuição normal multivariada, então:

$$b \sim N\left[\beta, \sigma^2 (X'X)^{-1}\right]$$

Assim, cada elemento de b é normalmente distribuído.

$$b_k \sim N\left[\beta_k, \sigma^2 \left(X'X\right)^{-1}_{kk}\right]$$

Seja s^{kk} o k-ésimo elemento da diagonal de $(X'X)^{-1}$, então:

$$Z_k = \frac{b_k - \beta_k}{\sqrt{\sigma^2 s^{kk}}} \sim N(0,1)$$

tem distribuição normal padronizada. Se σ^2 fosse conhecido, a inferência estatística sobre β_K poderia ser baseada em Z_K. Entretanto, σ^2 terá que ser estimado e Z_K não será usado.

Voltando aos resíduos

$$e = y - Xb = y - X(X'X)^{-1}X'y = My = M\left(X\beta + \varepsilon\right) = M\varepsilon$$

já que $\qquad MX = 0$

Um estimador de σ^2 será baseado na soma de quadrados dos resíduos.

$$e'e = \varepsilon' M' M\varepsilon = \varepsilon' M\varepsilon$$

$$e'e = \varepsilon' M \varepsilon$$

$$E(e'e) = E(\varepsilon' M \varepsilon)$$

$\varepsilon' M\ \varepsilon$ é um escalar (1x1) e assim, é igual ao seu traço.

$$E[tr(\varepsilon' M \varepsilon)] = E[tr(M \varepsilon \varepsilon')]$$

Como M é fixo

$$tr[ME(\varepsilon\varepsilon')] = tr(M\sigma^2 I) = \sigma^2 tr(M)$$

O traço de M é

$$\text{tr}\left(I_n - X(X'X)^{-1}X'\right) = tr(I_n) - tr\left[(X'X)^{-1}X'X\right] = tr(I_n) - tr(I_K) = n - K$$

Portanto,

$$E(e'e) = (n-K)\sigma^2$$

e um estimador não tendencioso de σ^2 é

$$s^2 = \frac{e'e}{n-K} \qquad Va\hat{r}(b) = s^2 (X'X)^{-1}$$

Lembre-se que:

O traço de uma matriz quadrada é igual a soma dos elementos da diagonal principal. $\Rightarrow$ $$\mathrm{tr}(A) = \sum_i a_{ii}$$

Propriedades:

$$\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$$

$$\mathrm{tr}(AB) = \mathrm{tr}(BA)$$

$$\mathrm{tr}(ABCD) = \mathrm{tr}(BCDA) = \mathrm{tr}(CDAB) = \mathrm{tr}(DABC)$$

Lembre-se que:

$$X_i \sim N(\mu, \sigma^2) \qquad Z_i = \frac{X_i - \mu}{\sigma} \sim N(0,1) \qquad \sum_{i=1}^{n} Z_i^2 = \chi^2(n)$$

$$\begin{cases} X \sim N(0,1) \\ Y \sim \chi^2(K) \end{cases} \qquad \frac{X}{\sqrt{Y/K}} \sim t(K)$$

com X e Y independentes

$$\begin{cases} Y_1 \sim \chi^2(K_1) \\ Y_2 \sim \chi^2(K_2) \end{cases} \qquad \frac{Y_1/K_1}{Y_2/K_2} \sim F(K_1, K_2)$$

com Y_1 e Y_2 independentes

$$\varepsilon \sim Normal \Rightarrow b_k \sim N(\beta_k, \sigma^2 (X'X)^{-1}_{kk})$$

$$\frac{b_k - \beta_k}{\sqrt{\sigma^2 s^{kk}}} \sim N(0,1)$$

$$e'e = \varepsilon' M \varepsilon$$

$$\frac{e'e}{\sigma^2} = \left(\frac{\varepsilon'}{\sigma}\right) M \left(\frac{\varepsilon}{\sigma}\right) \sim \chi^2 (n-K)$$

com número de graus de liberdade igual ao rank de *M*.

$$\frac{\varepsilon}{\sigma} \sim N(0, I)$$

$$\mathrm{Var}\left(\frac{\varepsilon_i}{\sigma}\right) = \frac{1}{\sigma^2} Var(\varepsilon_i) = 1$$

M é uma matriz idempotente $\Rightarrow$ o rank de *M* = traço de *M*

Se cov(X,Y) = 0 $\Rightarrow$ X e Y são independentes se X e Y ~ Normal

$$\operatorname{cov}(b,e) = E\left[\left(b-\beta\right)\left(e-0\right)'\right] = E\left[\left(X'X\right)^{-1}X'\varepsilon\varepsilon'M\right] = \sigma^2\left(X'X\right)^{-1}X'M =$$

$$\sigma^2(X'X)^{-1}X'(I - X(X'X)^{-1}X') = \sigma^2(X'X)^{-1}X' - \sigma^2(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I}X' = 0$$

Então b e e são independentes e,

$$\frac{\dfrac{b_k-\beta_k}{\sqrt{\sigma^2 s^{kk}}}}{\sqrt{\dfrac{e'e/\sigma^2}{n-K}}} \sim t\left(n-K\right)$$

ou

$$\frac{\dfrac{b_k-\beta_k}{\sqrt{\cancel{\sigma}^2 s^{kk}}}}{\sqrt{\dfrac{e'e\cancel{/\sigma^2}}{n-K}}} = \frac{b_k-\beta_k}{\sqrt{\dfrac{\cancel{(n-K)}s^2 s^{kk}}{\cancel{n-K}}}} \sim t\left(n-K\right)$$

substituindo $e'e = (n-K)s^2$

$$t_k = \frac{b_k - \beta_k}{\sqrt{s^2 s^{kk}}} = \frac{b_k - \beta_k}{s(b_k)}$$

Intervalo de confiança:

$$b_k - t_0 s(b_k) < \beta_k < b_k + t_0 s(b_k)$$

<u>Testando uma Restrição Linear</u>

$$H_0 : r_1\beta_1 + r_2\beta_2 + ... + r_K\beta_K = r'\beta = q$$

Geralmente, alguns *r*'s serão zero. A estimativa amostral será:

$$r_1 b_1 + r_2 b_2 + ... + r_K b_K = r'b = \hat{q}$$

Exemplo:

$$H_0 : \beta_1 - \beta_2 = 0$$

Se $\hat{q}$ difere significativamente de q, concluímos que os dados da amostra não são consistentes com a hipótese.

$$t(n-K)=\frac{\dfrac{r'b-q}{\sqrt{r'\left[\sigma^2(X'X)^{-1}\right]r}}}{\sqrt{\dfrac{e'e/\sigma^2}{n-k}}}=\frac{r'b-q}{\sqrt{\hat{\mathrm{var}}(\hat{q})}}$$

$\hat{q}$ é uma função linear de b e a $v\hat{a}r(b)=s^2(X'X)^{-1}$

$$E(\hat{q})=E(r'b)=r_1E(b_1)+\ldots+r_KE(b_K)=r_1\beta_1+\ldots+r_K\beta_K=r'\beta$$

$$\mathrm{var}(\hat{q})=\mathrm{var}(r'b)=E\left[r'b-r'\beta\right]^2=E\left[(r'b-r'\beta)(r'b-r'\beta)'\right]$$

$$=E\left[r'(b-\beta)(b-\beta)'r\right]=r'\,\mathrm{var}(b)r$$

$$\mathrm{var}(\hat{q})=r'[\sigma^2(X'X)^{-1}]r$$

Considere J restrições lineares na forma: $H_0 : R_{(J \times K)}\beta = q$

1) $\beta_j = 0$

$$R = [00...10...0] \quad \text{e} \quad q = 0$$

2) $\beta_k = \beta_j$

$$R = [001...-1...0] \quad \text{e} \quad q = 0$$

3) $\beta_2 + \beta_3 + \beta_4 = 1$

$$R = [0 \ 1 \ 1 \ 1 \ 0 \ ...] \quad \text{e} \quad q = 1$$

4) $\beta_1 = 0, \ \beta_2 = 0, \ 2\beta_3 = 1$

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2 & 0 & \cdots & 0 \end{bmatrix} \text{ e } \ q = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

5) $\beta_2 + \beta_3 = 1, \beta_4 + \beta_6 = 0$ e $\beta_5 + \beta_6 = 0$

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Dado o estimador de mínimos quadrados b, nosso interesse se centra na discrepância entre Rb e q. É improvável que seja exatamente 0. Queremos saber então se a diferença entre Rb e q pode ser atribuída a erro de amostragem ou se é significativa. Uma vez que b tem distribuição normal e Rb é uma função linear de b, Rb também tem distribuição normal.

$$X \sim N\left(\mu, \sigma^2\right) \qquad Z = \frac{X - \mu}{\sigma} \sim N(0,1) \qquad \sum_{i=1}^{n} Z_i^2 = \chi^2(n)$$

$$X \sim N\left(\mu, \Sigma\right) \qquad \Sigma^{-1/2}(X - \mu) \sim N(0, I)$$

$$(X - \mu)' \Sigma^{-1} (X - \mu) \sim \chi^2(n)$$

$$H_0 : R\beta = q$$

Se $b \sim$ Normal então $Rb \sim$ Normal

$$E(Rb) = RE(b) = R\beta$$

$$\mathrm{var}(Rb) = E[Rb - q]^2 = E[(Rb - R\beta)(Rb - R\beta)'] =$$

$$E\left[R(b-\beta)(b-\beta)' R'\right] = RE\left[(b-\beta)(b-\beta)'\right]R' = R\,\mathrm{var}(b)R' = \sigma^2 R(X'X)^{-1}R'$$

Se H_0 é verdadeira,

$$W = (Rb-q)'\left[\sigma^2 R(X'X)^{-1}R'\right]^{-1}(Rb-q) \sim \chi^2(J)$$

onde J é o número de graus de liberdade

Como não conhecemos σ^2 não podemos utilizar este teste, mas utilizamos F com s^2

$$F_{(J,n-K)} = \frac{(Rb-q)'\left[\cancel{\sigma}^2 R(X'X)^{-1}R'\right]^{-1}(Rb-q)/J}{\dfrac{\cancel{(n-K)}s^2/\cancel{\sigma}^2}{\cancel{n-K}}} = \frac{\dfrac{\chi^2(J)}{J}}{\dfrac{\chi^2(n-K)}{n-K}}$$

$$F_{(J,n-K)} = \frac{(Rb-q)'\left[s^2 R(X'X)^{-1} R'\right]^{-1}(Rb-q)}{J}$$

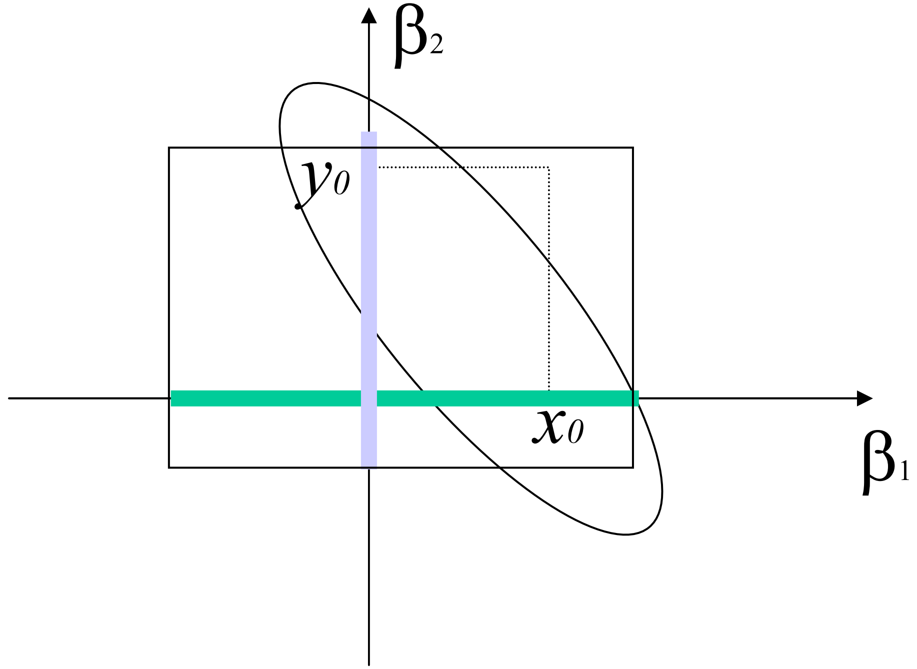
O intervalo de confiança para um único coeficiente é baseado no conjunto de valores para os quais o teste t é menor que o valor crítico especificado. É o conjunto de valores para os quais, a um dado nível de significância, nós não rejeitaríamos a hipótese de que β se iguala a um determinado valor.

No modelo de regressão múltipla, a região conjunta de confidência para um conjunto de coeficientes seria o conjunto de valores para os quais a hipótese de que o conjunto dos verdadeiros coeficientes simultaneamente iguais a determinados valores não poderia ser rejeitada.

O teste F para dois coeficientes seria:

$$F(2, n-K) = \frac{1}{2}(b-\beta)'(est.\mathrm{var}(b))^{-1}(b-\beta)$$

$$\frac{1}{2}\begin{bmatrix} b_1-\beta_1 \\ b_2-\beta_2 \end{bmatrix}' \begin{bmatrix} V(b_1) & \mathrm{cov}(b_1,b_2) \\ \mathrm{cov}(b_2,b_1) & V(b_2) \end{bmatrix}^{-1} \begin{bmatrix} b_1-\beta_1 \\ b_2-\beta_2 \end{bmatrix} \begin{matrix} = F(2,10) \\ < F(2,10) \end{matrix}$$



Se testar $H_0 : \beta_1 = x_0$ não rejeito H_0

Se testar $H_0 : \beta_2 = y_0$ não rejeito H_0

Se testar $H_0 : \beta_1 = x_0$ e $\beta_2 = y_0$ simultaneamente rejeito H_0

Ao testar separadamente não se leva em consideração a covariância entre b_1 e b_2. Se a covariância for negativa, por exemplo, então dificilmente quando b_1 é grande b_2 também será, por isso cai fora da elipse.