CAPÍTULO 8. REGRESSÃO MÚLTIPLA – TESTES DE HIPÓTESES E INTERVALOS DE CONFIANÇA

PARTE II: TESTES DE HIPÓTESES E INTERVALOS DE CONFIANÇA PARA DIVERSOS β_j's OU DIVERSAS COMBINAÇÕES $\boldsymbol{a}_i'\boldsymbol{\beta}$

8.5 TESTES SOBRE β_j E PARA UMA COMBINAÇÃO $\boldsymbol{a}'\boldsymbol{\beta}$

Muitas vezes queremos realizar **diversos** testes separados, do tipo $H_0\colon \beta_j = 0$, ao invés de um **único** teste $H_0\colon \boldsymbol{\beta}_1 = \mathbf{0}$. Ou então, realizar diversos testes separados, do tipo $H_0\colon \boldsymbol{a}_i'\boldsymbol{\beta} = 0$, ao invés de um único teste $H_0\colon \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$.

Problema: Quando realizamos diversos testes separados do tipo $H_0\colon \beta_j = 0$ ou do tipo $H_0\colon \boldsymbol{a}_i'\boldsymbol{\beta} = 0$, usando um nível de significância $\alpha = 0{,}05$ para cada teste, pode ocorrer um aumento no <u>nível de significância global</u>, o que pode alterar os resultados das inferências.

8.5.1 Testando um β_j ou uma combinação $\boldsymbol{a}'\boldsymbol{\beta}$

Usando a abordagem **modelo completo** *versus* **modelo reduzido.**, para testar $H_0: \beta_j = 0$ utilizamos a estatística:

$$F = \frac{\left(\hat{\boldsymbol{\beta}}'\mathbf{X}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}_1^{*\prime}\mathbf{X}_1'\boldsymbol{y}\right)}{\left(\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\boldsymbol{y}\right)/(n-k-1)} \qquad \textbf{(8.37)}$$

que tem distribuição $F(1, n-k-1)$ se H_0 é verdadeira.

Nesta abordagem, β_j é o último $\boldsymbol{\beta}$ e é feito um arranjo nas colunas $\mathbf{X}$ de tal forma que $\mathbf{X} = \left[\mathbf{X}_1 \ \vdots \ \boldsymbol{x}_j\right]$, onde $\boldsymbol{x}_j$ é a última coluna de $\mathbf{X}$.

Para testar se **uma combinação dos $\boldsymbol{\beta}$'s** é nula, ou seja, $H_0: \boldsymbol{a}'\boldsymbol{\beta} = 0$ *versus* $H_\mathrm{a}: \boldsymbol{a}'\boldsymbol{\beta} \neq 0$, usamos o **teste da hipótese linear geral**, com o vetor $\boldsymbol{a}'$ no lugar da matriz $\mathbf{C}$ em $H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, para $q = 1$.

A estatística utilizada é:

$$F = \frac{(\boldsymbol{a}'\hat{\boldsymbol{\beta}})' \left[\boldsymbol{a}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{a}\right]^{-1}(\boldsymbol{a}'\hat{\boldsymbol{\beta}})}{SQResiduo/(n-k-1)} = \frac{(\boldsymbol{a}'\hat{\boldsymbol{\beta}})^2}{s^2\boldsymbol{a}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{a}} \qquad (8.38)$$

onde $s^2 = QMResiduo$. Sob H_0 a estatística F em (8.38) tem distribuição $F(1, n-k-1)$.

Nota: Em modelos de regressão múltipla é <u>pouco comum</u> testar hipóteses sobre <u>combinações lineares</u> dos β's

Para testar $H_0: \beta_j = 0$ nesta abordagem usamos $\boldsymbol{a}' = [0, \ldots, 0, 1, 0, \ldots, 0]$, onde o número 1 está na posição $(j+1)$ e a estatística:

$$F = \frac{\left(\hat{\beta}_j\right)^2}{s^2 g_{j+1,j+1}} \tag{8.39}$$

onde $g_{j+1,j+1}$ é o $(j+1)$-ésimo elemento da diagonal de $(\mathbf{X}'\mathbf{X})^{-1}$.

Regra: Rejeitamos H_0 se $F > F(\alpha, 1, n-k-1)$ ou se p-valor $\leq \alpha$.

Importante: Como em (8.39) $F \sim F(1; n{-}k{-}1)$ graus de liberdade, para testar $H_0\colon \beta_j = 0$, podemos usar de **forma equivalente**, a estatística $t = \sqrt{F}$ definida como:

$$t_j = \frac{\hat{\beta}_j}{s\sqrt{g_{j+1,j+1}}} \tag{8.40}$$

Neste caso, rejeitamos $H_0\colon \beta_j = 0$ se $|t_j| > t(\alpha/2, n{-}\,k{-}\,1)$ ou, equivalentemente, se p-valor $\leq \alpha$.

Vale notar que para um teste bilateral $H_0\colon \beta_j = 0 \; versus \; H_a\colon \beta_j \neq 0$,

$$p\text{-valor} = 2P\left(t > |t_j|\right)$$

onde t_j é o valor da estatística de teste calculado em (8.40).

8.5.2 Testar diversos β_j's ou diversas combinações $\boldsymbol{a}_i'\boldsymbol{\beta}$

Quando testamos diversas hipóteses $H_0: \beta_j = 0$ ou $H_0: \boldsymbol{a}_i'\boldsymbol{\beta} = 0$, existem dois níveis de significância (α) diferentes envolvidos nas comparações:

- nível de significância geral ou $familywise\ \alpha\ level$ (α_f)
- nível de significância para cada comparação (α_c) ou $comparison\text{-}wise\ \alpha\ level$.

Alguns testes controlam o nível de significância geral (α_f) e outros, o nível de significância para cada comparação (α_c)

1) O teste de regressão global (ou geral) $H_0: \boldsymbol{\beta}_1 = \mathbf{0}$ em que $\boldsymbol{\beta}_1 = [\beta_1, \beta_2, \dots, \beta_k]'$ é feito utilizando um nível α de <u>significância geral</u> ou $familywise\ \alpha\ level$ (α_f).

2) Vamos testar $H_0: \beta_j = 0$, para $j = 1, 2, \dots, k$.

Seja $\alpha_c =$ "nível de significância por comparação ou para cada teste" ou $comparison$-$wise\ \alpha\ level.$

Problema: Se o teste $H_0: \beta_j = 0$ for executado k vezes ao nível α_c de significância em cada teste ocorrerá um **aumento** do nível α-geral (α_f). Probabilisticamente tem-se:

$$\alpha_f = 1 - (1 - \alpha_c)^k$$

Exemplo: Realizar diversos testes $H_0: \beta_j = 0$, fixando um nível de significância $\alpha_c = 0{,}05$ para cada teste:

# testes	1	2	4	6	8
α_f	0,05	0,0975	0,1855	0,2649	0,3366

Note que o nível de significância geral (α_f) aumenta bastante com o aumento do número de testes

- O método de Bonferroni e o de Scheffé, foram desenvolvidos para proteger contra essa inflação do nível α-global quando diversos testes são realizados.

Vamos assumir que os testes para $H_0: \beta_j = 0$ serão executados **sem considerar** se a hipótese global $H_0: \boldsymbol{\beta}_1 = \mathbf{0}$ foi rejeitada.

MÉTODO DE BONFERRONI

Suponhamos que sejam executados k testes de hipóteses do tipo $H_0: \beta_j = 0$, para $j = 1, 2, \ldots, k$.

- Para assegurar que o nível α_f seja menor ou igual a um valor desejado, $\alpha_f \leq \alpha^*$, basta assumir em cada um dos k testes, um nível de significância

$$\alpha_c = \frac{\alpha^*}{k}$$

- É comum fixar $\alpha^* = 0{,}05$.
- A tabela de Bonferroni de valores críticos $t(\alpha^*/2k)$ pode ser encontrada em Rencher (1995, pág. 499-500), dentre outros.

Para testar $H_0: \beta_j = 0$ *versus* $H_a: \beta_j \neq 0$, para $j = 1, 2, \ldots, k$ a um nível de significância $\alpha_j = \alpha^*/k$ usamos a estatística:

$$t_j = \frac{\widehat{\beta}_1}{s\sqrt{g_{j+1,j+1}}} \qquad (8.42)$$

Rejeitamos H_0 se $\left|t_j\right| > t(\alpha^*/2k, n - k - 1)$ ou se p-valor $\leq \alpha^*/2k$.

Para testar $H_{0i}: \boldsymbol{a_i}'\boldsymbol{\beta} = 0$, para $i = 1, 2, \ldots, d$, usamos a estatística:

$$F_i = \frac{(\boldsymbol{a}_i'\widehat{\boldsymbol{\beta}})'\left[\boldsymbol{a}_i'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{a}_i\right]^{-1}(\boldsymbol{a}_i'\widehat{\boldsymbol{\beta}})}{s^2} = \frac{(\boldsymbol{a}_i'\widehat{\boldsymbol{\beta}})^2}{s^2\boldsymbol{a}_i'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{a}_i} \qquad (8.43)$$

e rejeitamos H_{0i} se $F_i > F(\alpha^*/d, 1, n - k - 1$ ou se p-valor $\leq \alpha^*/d$.

Observações importantes:

- O α-global do procedimento para testar d combinações dos β's, $H_{0i}\colon \boldsymbol{a}_i'\boldsymbol{\beta} = 0$, $i = 1, 2, \ldots, d$, é válido somente se os coeficientes dos vetores $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d$ forem **especificados $\boldsymbol{a\ priori}$**, ou seja, se forem escolhidos <u>antes de olhar as estimativas dos β's</u>.

- Se desejarmos fixar os coeficientes $\boldsymbol{a}_i$ das combinações dos β's, $a\ posteriori$, depois de olhar os dados ou as estimativas dos β's, devemos usar o **método de Scheffé**, que será apresentado a seguir.

MÉTODO DE SCHEFFÉ

O procedimento de Scheffé para controlar α_f produz testes simultâneos de H_0: $\boldsymbol{a}'\boldsymbol{\beta} = 0$ (ou H_0: $\boldsymbol{a}'\boldsymbol{\beta} = t$) para **todo vetor $\boldsymbol{a}$** definido a *priori* ou a *posteriori.*

Para encontrar um valor crítico grande o suficiente que possa ser usado para comparar <u>todos</u> os possíveis vetores $\boldsymbol{a}$, a distribuição do $max_{\boldsymbol{a}}(F)$ é utilizada.

Para testar s hipótese H_0: $\boldsymbol{a}'\boldsymbol{\beta} = 0$ é usada a estatística

$$F = \frac{\left(\boldsymbol{a}'\hat{\boldsymbol{\beta}}\right)^2}{s^2\boldsymbol{a}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{a}} \tag{8.44}$$

e rejeitamos H_0 se $F > (k+1)\, F(\alpha^*, k+1, n-k-1)$.

Para testar $H_{0j}: \beta_j = 0$ individuais usando Scheffé, tomamos $\boldsymbol{a}' =$ [0,…, 0, 1, 0,…, 0], com o número 1 na (j+1)-ésima posição.

A raiz quadrada da estatística F é $t_j = \hat{\beta}_j / \sqrt{s^2\, g_{j+1,j+1}} \sim t_{(n-k-1)}$ como em (8.42).

$\Rightarrow$ Rejeitamos $H_0: \boldsymbol{a}'\boldsymbol{\beta} = \beta_j = 0$ se

$$\left|t_j\right| \geq \sqrt{(k+1)\, F(\alpha^*, k+1, n-k-1)}$$

Observações:

- Os testes de Bonferroni para β_j's individuais em (8.42) são **conservadores** e **mais poderosos** que os testes de Scheffé.
- Teste conservativo ou conservador: controla bem a ocorrência do erro tipo I, que consiste em rejeitar erroneamente H_0.
- Teste mais poderoso: maior probabilidade de rejeitar corretamente H_0.
- Para um **grande número de combinações lineares** $\boldsymbol{a}'\boldsymbol{\beta}$ o teste de Scheffé é melhor, porque para qualquer número de escolhas de $\boldsymbol{a}'\boldsymbol{\beta}$ pode ser testado contra um **único valor crítico**:

$$(k+1)\,F(\alpha^*, k+1, n-k-1)$$

enquanto o valor crítico para o teste de Bonferroni em (8.43) $F(\alpha^*/d, 1; n-k-1)$ aumenta com o número de testes, d.

- Podemos usar o teste de Scheffé para combinações lineares $\boldsymbol{a}'\boldsymbol{\beta}$ escolhidas **depois de olhar os dados** ($post\ hoc$), o que não é possível com o método de Bonferroni.

Observação:

Se os k testes H_0: $\beta_j = 0, j = 1, 2, \ldots, k$ forem realizados usando, por exemplo, a estatística

$$t_j = \hat{\beta}_1 / s\sqrt{g_{j+1,j+1}}, \quad \text{em (8.42)}$$

somente se H_0: $\boldsymbol{\beta}_1 = \mathbf{0}$ foi rejeitada usando F em (8.5), **o que é bastante comum**, a **taxa de erro global** é reduzida e os valores críticos

$$t(\alpha^*/2k, n-k-1) \quad \text{e} \quad \sqrt{(k+1)\,F(\alpha^*, k+1, n-k-1)}$$

serão ainda **mais conservativos**.

Conclusão: Se a hipótese H_0: $\boldsymbol{\beta}_1 = \mathbf{0}$ foi rejeitada, podemos usar α^* em cada teste H_0: $\beta_j = 0$ ou H_0: $\boldsymbol{a}'\boldsymbol{\beta} = 0$ que o $\alpha_f \cong \alpha^*$

Exemplo 8.5.2. Testar H_{01}: $\beta_1 = 0$ e H_{02}: $\beta_2 = 0$ para os dados da Tabela 7.1. Já sabemos que:

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 5.3754 \\ 3.0118 \\ -1.2855 \end{bmatrix} \qquad s^2 = 2.8288$$

$$\widehat{cov}(\widehat{\boldsymbol{\beta}}) = s^2 \begin{bmatrix} 0.97476 & 0.24290 & -0.22871 \\ 0.24290 & 0.16207 & -0.11120 \\ -0.22871 & -0.11120 & 0.08360 \end{bmatrix}$$

Usando a estatística t_j em (8.42) temos:

$$t_1 = \frac{\hat{\beta}_1}{s\sqrt{g_{2,2}}} = \frac{3.0118}{\sqrt{2.8288}\sqrt{0.16207}} = 4.4482$$

$$t_2 = \frac{\hat{\beta}_2}{s\sqrt{g_{3,3}}} = \frac{-1.2855}{\sqrt{2.8288}\sqrt{0.08360}} = -2.6435$$

a) Usando um nível de significância $\alpha = 5\%$ para cada teste, **rejeitaremos** H_{01} e H_{02} porque $t(0.025, 9) = 2.262$.

Os p-valores (**bilaterais**)para t_1 e t_2 são:

$$p\text{-}valor_1 = 2P(t > |t_1|) = 2P(t > |4.4482|) = 0.0016$$

$$p\text{-}valor_2 = 2P(t > |t_2|) = 2P(t > |-2.6435|) = 0.0268$$

b) Usando o método de Bonferroni: 2 testes $\Rightarrow \alpha_c = 0{,}05/2 = 0{,}025$ para cada teste $\Rightarrow$ **não rejeitaríamos** $H_{02}: \beta_2 = 0$ porque

$$p\text{-}valor_2 = 0{,}0268 > 0{,}025$$

Note que:

Usual:	$t(0{,}025; 9\ gl) = 2{,}2622$
Bonferroni:	$t(0{,}0125; 9\ gl) = 2{,}6850$
Scheffé:	$t(0{,}025; 9\ gl) = 3{,}4041$

8.6 INTERVALOS DE CONFIANÇA E INTERVALOS DE PREDIÇÃO

Vamos considerar intervalos de confiança para β_j, $\boldsymbol{a}'\boldsymbol{\beta}$, $E(y)$ e σ^2, além de intervalos de predição para futuras observações.

8.6.1 Região de confiança para $\boldsymbol{\beta}$

(*Detalhes no livro do Rencher*)

Uma região de $100(1-\alpha)\%$ confiança para $\beta_0, \beta_1, \dots, \beta_k$ consiste de todos os valores de $\boldsymbol{\beta}$ que satisfazem:

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}\,(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq (k+1)F(\alpha, k{+}1, n{-}k{-}1) \qquad (8.46)$$

Se $k = 1$ variável regressora: a região de confiança é uma elipse em duas dimensões. Para mais de uma variável regressora a região é um elipsoide, difícil de visualizar e interpretar.

8.6.2. Intervalos de confiança para β_j

Se $\beta_j \neq 0$ em (8.40) nós podemos escrever que:

$$P\left[-t_{\alpha/2,n-k-1} \leq \frac{\hat{\beta}_j - \beta_j}{s\sqrt{g_{j+1,j+1}}} \leq t_{\alpha/2,n-k-1}\right] = 1 - \alpha$$

Resolvendo para β_j, temos:

$$P\left[\hat{\beta}_j - t_{\alpha/2}(s)\sqrt{g_{j+1,j+1}} \leq \beta_j \leq \hat{\beta}_j + t_{\alpha/2}(s)\sqrt{g_{j+1,j+1}}\right] = 1 - \alpha$$

Podemos escrever que:

$$\hat{\beta}_j \pm t_{\alpha/2,n-k-1}(s)\sqrt{g_{j+1,j+1}} \tag{8.47}$$

são os limites do **intervalo de confiança** 100(1 – α)% para β_j.

Dizemos que "o intervalo obtido em (8.47) contém o verdadeiro valor de β_j com 100(1–α)% de confiança".

Vale observar que o coeficiente de confiança (1− α) vale para um **único intervalo de confiança** para um dos β_j's.

Para calcular intervalos de confiança para todos os β_j's, com coeficiente de confiança global de (1–α), veja a Seção 8.6.7.

Exemplo 8.6.2. Vamos calcular o $IC(95\%)$ para cada β_j usando y_2 no conjunto de dados apresentados na Tabela 7.4.

São dados: $s = 4{,}0781$, $t_{0{,}025;15} = 2{,}1314$.

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 65.3755 & -0.3388 & -0.3125 & -0.0204 \\ -0.3388 & 0.0018 & 0.0013 & -0.0004 \\ -0.3125 & 0.0013 & 0.0041 & -0.0018 \\ -0.0204 & -0.0004 & -0.0018 & 0.0216 \end{bmatrix}$$

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} -26.0353 \\ 0.4046 \\ 0.2930 \\ 1.0338 \end{bmatrix}$$

Para β_1, obtemos por (8.47):

$$\hat{\beta}_1 \pm t_{0.025;15}(s)\sqrt{g_{2,2}} = 0.4046 \pm (2.1314)(4.0781)\sqrt{0.0018}$$

$$= 0{,}4046 \pm 0.3723$$

$$\Rightarrow IC(\beta_1, 95\%) = (0.0322;\ 0.7769)$$

Para os outros β_j's temos:

β_0: -26.0353 ± 70.2812 $\Rightarrow IC(\beta_0, 95\%) = (-96.3165; 44.2459)$

β_2: 0.2930 ± 0.5551 $\Rightarrow IC(\beta_2, 95\%) = (-0.2621; 0.8481)$

β_3: 1.0338 ± 1.2777 $\Rightarrow IC(\beta_3, 95\%) = (-0.2439; 2.3115)$

Vale observar que:

- O coeficiente de confiança de 95% vale somente para um dos quatro intervalos de confiança.
- Para mais de um I.C. ver Exemplo 8.6.7.

8.6.3 Intervalo de confiança para $\boldsymbol{a}'\boldsymbol{\beta}$

Se $\boldsymbol{a}'\boldsymbol{\beta} \neq 0$ em (8.44) e pelo Problema 5.12, podemos escrever que os limites do **intervalo de confiança** (1–α) para **um único** $\boldsymbol{a}'\boldsymbol{\beta}$ como:

$$\boldsymbol{a}'\widehat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1}\,(s)\sqrt{\boldsymbol{a}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{a}} \qquad (8.49)$$

8.6.4 Intervalo de confiança para $E(\boldsymbol{y})$

- Seja $\boldsymbol{x}_0 = [1, x_{01}, x_{02}, \cdots, x_{0k}]'$ uma escolha particular de $\boldsymbol{x} = [1, x_1, x_2, \cdots, x_k]'$, que não precisa ser uma linha da matriz $\mathbf{X}$.
- Se $\boldsymbol{x}_0$ estiver muito fora da região coberta pela amostra, a previsão baseada em $\boldsymbol{x}_0$ será pobre.

Um **intervalo de confiança** $100(1–\alpha)\%$ para $E(y_0) = \boldsymbol{x}_0'\boldsymbol{\beta}$ de (8.49), que é a média da distribuição dos valores y correspondente a $\boldsymbol{x}_0$ é dado por:

$$\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1}(s)\sqrt{\boldsymbol{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_0} \qquad (8.52)$$

Garante a confiança de 1– α para uma **única** escolha do vetor $\boldsymbol{x}_0$.

Para I.C.'s cobrindo todas as escolhas de $\boldsymbol{x}_0$'s, veja a Seção 8.6.7.

Para o caso especial de uma **regressão linear simples**, temos que o intervalo de confiança para $E(y_0)$ é dado por:

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2,n-2}(s)\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{8.58}$$

Vale notar que o tamanho do intervalo em (8.58) depende de quão longe x_0 está de $\bar{x}$.

8.6.5. Intervalo de predição para uma observação futura

Um **intervalo de confiança** para uma **observação futura** y_0 correspondente a um $\boldsymbol{x}_0$ é chamado **intervalo de predição**, porque y_0 é uma **observação individual**, uma **variável aleatória** ao invés de um parâmetro.

O **intervalo de predição** para uma observação futura $(\boldsymbol{x}_0, y_0)$ é dado por:

$$\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}} \pm t_{\alpha/2,n-k-1}(s)\sqrt{1+\boldsymbol{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_0} \qquad (8.61)$$

Nota: O intervalo de predição de y_0 deverá ser mais largo que o intervalo de confiança para o parâmetro $E(y_0)$.

Para o caso da regressão linear simples, a expressão (8.61) se reduz a:

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{\alpha/2,n-2}(s)\sqrt{1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \qquad (8.63)$$

Exemplo 8.6.5. Usando os dados da Tabela 6.2 vamos calcular um intervalo de predição 95% para $x_0 = 80$, usando (8.63):

$$\hat{\beta}_0 + \hat{\beta}_1 x_0 \pm t_{0.025;16}(s)\sqrt{1 + \frac{1}{18} + \frac{(80-58.056)^2}{19530.944}}$$

$$80.5386 \pm 2.11998(13.8547)(1.0393)$$

$$80.5386 \pm 30.5258$$

Então, o intervalo [50.0128; 111.0644] contém o verdadeiro valor de y correspondente a $x_0 = 80$, com 95% de confiança.

8.6.6 Intervalo de confiança para σ^2

(ver detalhes no livro do Rencher)

O intervalo de confiança 100(1–α)% para a variância é dado por:

$$\frac{(n-k-1)s^2}{\chi^2_{\alpha/2;n-k-1}} \leq \sigma^2 \leq \frac{(n-k-1)s^2}{\chi^2_{(1-\alpha/2);n-k-1}} \qquad (8.65)$$

em que $\chi^2_{1-\alpha/2;n-k-1}$ é o percentil (superior) de ordem 1–α/2 e $\chi^2_{\alpha/2;n-k-1}$ é o percentil (inferior) de ordem α/2.

O intervalo de confiança (1–α) para o **desvio padrão** é dado por

$$\sqrt{\frac{(n-k-1)s^2}{\chi^2_{\alpha/2;n-k-1}}} \leq \sigma \leq \sqrt{\frac{(n-k-1)s^2}{\chi^2_{1-\alpha/2;n-k-1}}} \qquad (8.66)$$

8.6.7 Intervalos simultâneos

O coeficiente de confiança $(1-\alpha)$ para os intervalos obtidos nas Seções 8.6.1-8.6.6 **é válido para um único intervalo** em cada caso.

Para intervalos múltiplos adaptamos os métodos da Seção 8.5.2.

Utilizando o método de Bonferroni, intervalos de confiança para β_1, $\beta_2,\ldots,\beta_k$ são dados por:

$$\hat{\beta}_j \pm t_{\alpha^*/(2k),n-k-1}(s)\sqrt{g_{j+1,j+1}} \tag{8.67}$$

Para d funções lineares $\boldsymbol{a}_1'\boldsymbol{\beta}$, $\boldsymbol{a}_2'\boldsymbol{\beta},\ldots,\boldsymbol{a}_d'\boldsymbol{\beta}$ escolhidas $a\ priori$ os intervalos de confiança de Bonferroni são dados por:

$$\boldsymbol{a}_i'\widehat{\boldsymbol{\beta}} \pm t_{\alpha^*/(2d),n-k-1}(s)\sqrt{\boldsymbol{a}_i'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{a}_i} \tag{8.68}$$

para $i = 1, 2\ldots, d$.

Esses IC's asseguram uma **confiança simultânea** de, no mínimo, $(1 - \alpha^*)$; isto é, garantem uma confiança de $100(1-\alpha^*)\%$ que cada um dos d intervalos contenha o verdadeiro valor do parâmetro.

Intervalos de Bonferroni para construir IC´s para $E(y_0) = \boldsymbol{x}_0'\boldsymbol{\beta}$ para **poucos** valores de $\boldsymbol{x}_0$, digamos $\boldsymbol{x}_{01}, \boldsymbol{x}_{02}, \ldots, \boldsymbol{x}_{0d}$:

$$\boldsymbol{x}_{0i}'\widehat{\boldsymbol{\beta}} \pm t_{\alpha^*/2d, n-k-1}(s)\sqrt{\boldsymbol{x}_{0i}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_{0i}} \tag{8.69}$$

para $i = 1, 2\ldots, d$.

Intervalos de predição de Bonferroni para d novas observações y_{01}, $y_{02}, \ldots, y_{0d}$ em d valores de $\boldsymbol{x}_0$, digamos $\boldsymbol{x}_{01}, \boldsymbol{x}_{02}, \ldots, \boldsymbol{x}_{0d}$:

$$\boldsymbol{x}_{0i}'\widehat{\boldsymbol{\beta}} \pm t_{\alpha^*/2d, n-k-1}(s)\sqrt{1 + \boldsymbol{x}_{0i}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_{0i}} \tag{8.70}$$

para $i = 1, 2\ldots, d$.

Os intervalos obtidos por (8.70) terão um coeficiente de confiança global de, no mínimo, (1–α).

Os limites dos intervalos de confiança (conservativos) de Scheffé para **todas** as possíveis funções lineares $\boldsymbol{a}'\boldsymbol{\beta}$ (escolhidas $a\ priori$ ou não) são dados por:

$$\boldsymbol{a}'\widehat{\boldsymbol{\beta}} \pm (s)\sqrt{(k+1)\, F_{\alpha^*;k+1,n-k-1}\ \boldsymbol{a}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{a}} \qquad (8.71)$$

- Intervalos construídos desta forma têm um **coeficiente de confiança global** de, no mínimo, 100(1–α^*)%.
- Para poucas funções lineares, os intervalos em (8.68) serão mais estreitos, mas para um número grande de funções lineares, os intervalos em (8.69) serão mais estreitos.

Limites de confiança para $E(y_0) = \boldsymbol{x}_0'\boldsymbol{\beta}$, para todos os possíveis $\boldsymbol{x}_0$, nós usamos (8.71):

$$\boldsymbol{x}_0'\widehat{\boldsymbol{\beta}} \pm (s)\sqrt{(k+1)\,F_{\alpha^*;k+1,n-k-1}\,\boldsymbol{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_0} \qquad (8.72)$$

Os limites dos intervalos de predição de Scheffé para y_{01}, y_{02},…, y_{0d}, são calculados por:

$$\boldsymbol{x}_{0i}'\widehat{\boldsymbol{\beta}} \pm (s)\sqrt{d\,F_{\alpha^*;k+1,n-k-1}[1+\boldsymbol{x}_{0i}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{x}_{0i}]} \qquad (8.73)$$

para $i = 1, 2\ldots, d$.

Esses d intervalos de predição asseguram um coeficiente de confiança global de $(1 - \alpha^*)$, no mínimo.

8.7 TESTES DA RAZÃO DE VEROSSIMILHANÇAS

(Para maiores detalhes: Rencher, pág. 204)

Os testes desenvolvidos nas Seções 8.1 e 8.2 foram derivados utilizando métodos **informais** baseados em características de SQ's que têm distribuição de quiquadrado e são independentes.

Esses mesmos testes podem ser obtidos de maneira **mais formal**, através de uma abordagem de razão de verossimilhanças.

Vamos apresentar os principais resultados do **teste da razão de verossimilhanças** (TRV) no contexto simples de testar

$$H_0: \boldsymbol{\beta} = \mathbf{0} \; versus \; H_1: \boldsymbol{\beta} \neq \mathbf{0}.$$

A função de verossimilhança – $L(\boldsymbol{\beta}, \sigma^2)$ – foi definida na Seção 7.6.2 como a densidade conjunta dos y's. Assumindo $\boldsymbol{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}\sigma^2)$

$$L(\boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}}\, e^{-(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta})'(\boldsymbol{y}-\mathbf{X}\boldsymbol{\beta})/2\sigma^2}$$

O **teste da razão de verossimilhanças** consiste em comparar o máximo valor de $L(\boldsymbol{\beta}, \sigma^2)$ restrito por $H_0\!: \boldsymbol{\beta} = \mathbf{0}$, com o máximo valor de $L(\boldsymbol{\beta}, \sigma^2)$ sob a hipótese $H_1\!: \boldsymbol{\beta} \neq \mathbf{0}$, que não é restrita, através da razão de verossimilhanças:

$$LR = \frac{max_{H_0} L(\boldsymbol{\beta}, \sigma^2)}{max_{H_1} L(\boldsymbol{\beta}, \sigma^2)} = \frac{max\, L(\mathbf{0}, \sigma^2)}{max\, L(\boldsymbol{\beta}, \sigma^2)} \qquad (8.75)$$

É claro que $0 \leq LR \leq 1$, porque o máximo de LR restrito a $\boldsymbol{\beta} = \mathbf{0}$ não pode exceder o máximo irrestrito.

Como pequenos valores de LR favorecem a hipótese alternativa, H_1, e grandes valores de LR favorecem H_0, nós rejeitamos H_0 se $LR \leq c$, em que c é escolhido de tal forma que $P(LR \leq c) = \alpha$, se H_0 é verdadeira.

Wald (1943) mostrou que, sob H_0 e para n grande,

$$-2ln(LR) \sim \chi^2(v),$$

onde v é o número de parâmetros estimados sob H_1 menos o número de estimativas sob H_0.

No caso de H_0: $\boldsymbol{\beta} = \mathbf{0}$ *versus* H_1: $\boldsymbol{\beta} \neq \mathbf{0}$ nós temos $v = (k+2)-1 = k+1$, porque os coeficientes de regressão ($\boldsymbol{\beta}$) e a variância (σ^2) são estimados sob H_1, enquanto somente a variância (σ^2) é estimada sob H_0.

Teorema 8.7A Se $\boldsymbol{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}\sigma^2)$ o teste da razão de verossimilhanças para $H_0: \boldsymbol{\beta} = \mathbf{0}$ *versus* $H_1: \boldsymbol{\beta} \neq \mathbf{0}$ pode ser baseado em

$$F = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\boldsymbol{y}/(k+1)}{\left(\boldsymbol{y}'\boldsymbol{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\boldsymbol{y}\right)/(n-k-1)}$$

Rejeitamos H_0 se $F > F_{(\alpha;\, k+1, n-k-1)}$.

Teorema 8.7B Se $\boldsymbol{y} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}\sigma^2)$ então o teste-F para H_0: $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ no Teorema 8.4A é **equivalente** ao Teste da Razão de Verossimilhanças.