Modelos de Regressão Múltipla - Parte III

Erica Castilho Rodrigues

10 de Dezembro de 2015

A Soma de Quadrados Extra e o Teste F Parcial

Variáveis Indicadoras

A Soma de Quadrados Extra e o Teste F Parcial

- A quantidade

$$SQT = \sum_i (Y_i - \overline{Y})^2$$

mede a

- A quantidade

$$SQT = \sum_i (Y_i - \overline{Y})^2$$

  mede a variância de Y.
- Essa quantidade muda de um modelo para o outro?

- A quantidade

$$SQT = \sum_i (Y_i - \overline{Y})^2$$

mede a variância de Y.

- Essa quantidade muda de um modelo para o outro? Não.
- Considere o modelo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

vamos denotar sua soma de quadrados totais por $SQT(X_1, X_2)$.

- Considere o modelo

$$Y = \beta_0^* + \beta_1^* X_1 + \epsilon^*$$

vamos denotar sua soma de quadrados totais por $SQT(X_1)$.

- Considere o modelo

$$Y = \beta_0^{**} + \beta_1^{**} X_2 + \epsilon^{**}$$

vamos denotar sua soma de quadrados totais por $SQT(X_2)$.

- Temos então que

$$SQT(X_1, X_2) =$$

- Considere o modelo

$$Y = \beta_0^* + \beta_1^* X_1 + \epsilon^*$$

vamos denotar sua soma de quadrados totais por $SQT(X_1)$.

- Considere o modelo

$$Y = \beta_0^{**} + \beta_1^{**} X_2 + \epsilon^{**}$$

vamos denotar sua soma de quadrados totais por $SQT(X_2)$.

- Temos então que

$$SQT(X_1, X_2) = SQT(X_1) =$$

- Considere o modelo

$$Y = \beta_0^* + \beta_1^* X_1 + \epsilon^*$$

vamos denotar sua soma de quadrados totais por $SQT(X_1)$.

- Considere o modelo

$$Y = \beta_0^{**} + \beta_1^{**} X_2 + \epsilon^{**}$$

vamos denotar sua soma de quadrados totais por $SQT(X_2)$.

- Temos então que

$$SQT(X_1, X_2) = SQT(X_1) = SQT(X_2)\,.$$

- Sabemos ainda que

$$SQT =$$

- Sabemos ainda que

$$SQT = SQR + SQE$$

logo

$$SQT(X_1, X_2) = SQT(X_1) \Rightarrow$$

$$SQR(X_1, X_2) + SQE(X_1, X_2) =$$

- Sabemos ainda que

$$SQT = SQR + SQE$$

logo

$$SQT(X_1, X_2) = SQT(X_1) \Rightarrow$$

$$SQR(X_1, X_2) + SQE(X_1, X_2) = SQR(X_1) + SQE(X_1)\,.$$

- Portanto

$$SQR(X_1, X_2) > SQR(X_1) \Rightarrow$$

- Sabemos ainda que

$$SQT = SQR + SQE$$

logo

$$SQT(X_1, X_2) = SQT(X_1) \Rightarrow$$

$$SQR(X_1, X_2) + SQE(X_1, X_2) = SQR(X_1) + SQE(X_1)\,.$$

- Portanto

$$SQR(X_1, X_2) > SQR(X_1) \Rightarrow SQE(X_1, X_2) < SQE(X_1)\,.$$

- Ao incluirmos uma mariável:
  - melhoramos o ajuste do modelo -

- Sabemos ainda que

$$SQT = SQR + SQE$$

logo

$$SQT(X_1, X_2) = SQT(X_1) \Rightarrow$$

$$SQR(X_1, X_2) + SQE(X_1, X_2) = SQR(X_1) + SQE(X_1)\,.$$

- Portanto

$$SQR(X_1, X_2) > SQR(X_1) \Rightarrow SQE(X_1, X_2) < SQE(X_1)\,.$$

- Ao incluirmos uma mariável:
  - melhoramos o ajuste do modelo -$SQR \Uparrow$ $SQE \Downarrow$

- Queremos medir qual o ganho de se acrescentar uma variável ao modelo.
- Esse ganho é medido pela **Soma de Quadrados Extra**.

Soma de Quadrados Extras

- Mantemos fixas todas as preditoras que já estão no modelo.
- Acrescentamos uma variável a mais.
- A Soma de Quadrados Extras é a redução da Soma de Quadrados Residuais (SQE).

- A Soma de Quadrados Extras ao acrescentarmos X_2 no modelo é dada por

$$SQR(X_2|X_1) = SQE(X_1) - SQE(X_1, X_2)\,.$$

- Observe que

$$SQE =$$

- A Soma de Quadrados Extras ao acrescentarmos X_2 no modelo é dada por

$$SQR(X_2|X_1) = SQE(X_1) - SQE(X_1, X_2)\,.$$

- Observe que

$$SQE = SQT - SQR \Rightarrow$$

- A Soma de Quadrados Extras ao acrescentarmos X_2 no modelo é dada por

$$SQR(X_2|X_1) = SQE(X_1) - SQE(X_1, X_2)\,.$$

- Observe que

$$SQE = SQT - SQR \Rightarrow$$

$SQR(X_2|X_1) =$

- A Soma de Quadrados Extras ao acrescentarmos X_2 no modelo é dada por

$$SQR(X_2|X_1) = SQE(X_1) - SQE(X_1, X_2)\,.$$

- Observe que

$$SQE = SQT - SQR \Rightarrow$$

$$SQR(X_2|X_1) = SQT(X_1) - SQR(X_1) - [SQT(X_1, X_2) - SQR(X_1, X_2)]$$

$$=$$

- A Soma de Quadrados Extras ao acrescentarmos X_2 no modelo é dada por

$$SQR(X_2|X_1) = SQE(X_1) - SQE(X_1, X_2)\,.$$

- Observe que

$$SQE = SQT - SQR \Rightarrow$$

$$SQR(X_2|X_1) = SQT(X_1) - SQR(X_1) - [SQT(X_1, X_2) - SQR(X_1, X_2)]$$

$$= SQT(X_1) - SQR(X_1) - SQT(X_1, X_2) + SQR(X_1, X_2) =$$

- A Soma de Quadrados Extras ao acrescentarmos X_2 no modelo é dada por

$$SQR(X_2|X_1) = SQE(X_1) - SQE(X_1, X_2)\,.$$

- Observe que

$$SQE = SQT - SQR \Rightarrow$$

$$SQR(X_2|X_1) = SQT(X_1) - SQR(X_1) - [SQT(X_1, X_2) - SQR(X_1, X_2)]$$

$$= SQT(X_1) - SQR(X_1) - SQT(X_1, X_2) + SQR(X_1, X_2) =$$

$$SQR(X_1, X_2) - SQR(X_1)$$

pois $SQT(X_1) = SQT(X_1, X_2)$.

- A Soma de Quadrados Extras também pode ser vista como:
  - aumento na SQR quando acrescentamos uma variável no modelo.
- Ou seja

$$SQR(X_2|X_1) = SQR(X_1, X_2) - SQR(X_1)$$

$$SQR(X_1|X_2) = SQR(X_1, X_2) - SQR(X_2)\ .$$

- Devemos verificar se esse acréscimo é grande o suficiente antes de incluirmos a variável no modelo.
- Essas somas aparecem em vários testes sobre os coeficientes de regressão.
- Queremos saber se certas variáveis devem ser retiradas ou não do modelo, dado que outras variáveis já estão nele.

Exemplo:

- Foi feito um estudo para avaliar a quantidade de gordura corporal (Y).
- Foram selecionadas as seguintes variáveis explicativas:
  - X_1 - medida de expessura da dobra do tríceps;
  - X_2 - circunferência da coxa;
  - X_3 - circunferência do antebraço.
- Os dados de 20 mulheres de 25 a 24 anos foram coletados.

Exemplo: (continuação)

- A base de dados é apresentada a seguir.

i	X_{1i}	X_{2i}	X_{3i}	Y_i
1	19.5	43.1	29.1	11.9
2	24.7	49.8	28.2	22.8
3	30.7	51.9	37.0	18.7
...	...	...	...	...
18	30.2	58.6	24.6	25.4
19	22.7	48.2	27.1	14.8
20	25.2	51.0	27.5	21.1

Exemplo: (continuação)

- A medição da quantidade de gordura corporal é um procedimento complicado e caro.
- Seria muito útil um modelo que estimasse a gorudra usando poucas medidas simples.
- As academias e consultórios fazem dessa maneira.
- Devemos decidir entre as variáveis

$$X_1, X_2, X_3$$

  quais devem entrar no modelo.
- Algumas possibilidades

Exemplo: (continuação)

- A medição da quantidade de gordura corporal é um procedimento complicado e caro.
- Seria muito útil um modelo que estimasse a gorudra usando poucas medidas simples.
- As academias e consultórios fazem dessa maneira.
- Devemos decidir entre as variáveis

$$X_1, X_2, X_3$$

  quais devem entrar no modelo.
- Algumas possibilidades

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

$$Y = \beta_0 + \beta_2 X_2 + \epsilon$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \epsilon$$

- Qual modelo é o melhor?

Exemplo: (continuação)

- Vejamos como ficam as Tabelas ANOVA para os modelos.

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão (X_1)	1	$SQR = 352.27$	QMR=352.27
Resíduo	18	$SQE = 143.12$	QME=7.95
Total	19	$SQT = 495.39$	

Tabela: Tabela ANOVA para o modelo $Y = \beta_0 + \beta_1 X_1 + \epsilon$

$$F_{obs} = 44.31 \quad F_{1,18} = 4,41$$

- Conclusão:

Exemplo: (continuação)

- Vejamos como ficam as Tabelas ANOVA para os modelos.

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão (X_1)	1	$SQR = 352.27$	QMR=352.27
Resíduo	18	$SQE = 143.12$	QME=7.95
Total	19	$SQT = 495.39$	

Tabela: Tabela ANOVA para o modelo $Y = \beta_0 + \beta_1 X_1 + \epsilon$

$$F_{obs} = 44.31 \quad F_{1,18} = 4,41$$

- Conclusão: com 5% de singificância temos evidência de que X_1 é significativa para explicar Y.

Exemplo: (continuação)

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão (X_2)	1	$SQR = 381.97$	QMR=381.97
Resíduo	18	$SQE = 113.42$	QME=6.30
Total	19	$SQT = 495.39$	

Tabela: Tabela ANOVA para o modelo $Y = \beta_0 + \beta_2 X_2 + \epsilon$

$$F_{obs} = 60.63 \quad F_{1,18} = 4,41$$

- Conclusão:

Exemplo: (continuação)

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão (X_2)	1	$SQR = 381.97$	QMR=381.97
Resíduo	18	$SQE = 113.42$	QME=6.30
Total	19	$SQT = 495.39$	

Tabela: Tabela ANOVA para o modelo $Y = \beta_0 + \beta_2 X_2 + \epsilon$

$$F_{obs} = 60.63 \quad F_{1,18} = 4,41$$

- Conclusão:com 5% de singificância temos evidência de que X_2 é significativa para explicar Y.

Exemplo: (continuação)

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão (X_1, X_2)	2	$SQR = 385.44$	QMR=192.72
Resíduo	17	$SQE = 109.95$	QME=6.47
Total	19	$SQT = 495.39$	

Tabela: Tabela ANOVA para o modelo $Y = \beta_0 + \beta_1 X_2 + \beta_2 X_2 + \epsilon$

$$F_{obs} = 60.63 \quad F_{2,17} = 3.59$$

- Conclusão:

Exemplo: (continuação)

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão (X_1, X_2)	2	$SQR = 385.44$	QMR=192.72
Resíduo	17	$SQE = 109.95$	QME=6.47
Total	19	$SQT = 495.39$	

Tabela: Tabela ANOVA para o modelo $Y = \beta_0 + \beta_1 X_2 + \beta_2 X_2 + \epsilon$

$$F_{obs} = 60.63 \quad F_{2,17} = 3.59$$

- Conclusão:com 5% de singificância temos evidência de que pelo menos X_1 ou X_2 são signficativas para explicar Y.

Exemplo: (continuação)

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão (X_1, X_2, X_3)	3	$SQR = 396.98$	QMR=132.33
Resíduo	16	$SQE = 98.41$	QME=6.15
Total	19	$SQT = 495.39$	

Tabela: Tabela ANOVA para o modelo
$Y = \beta_0 + \beta_1 X_2 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$

$$F_{obs} = 21.52 \quad F_{3,16} = 3.24$$

- Conclusão:

Exemplo: (continuação)

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão (X_1, X_2, X_3)	3	$SQR = 396.98$	QMR=132.33
Resíduo	16	$SQE = 98.41$	QME=6.15
Total	19	$SQT = 495.39$	

Tabela: Tabela ANOVA para o modelo
$Y = \beta_0 + \beta_1 X_2 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$

$$F_{obs} = 21.52 \quad F_{3,16} = 3.24$$

- Conclusão:com 5% de singificância temos evidência de que pelo menos X_1, X_2 ou X_3 são signficativas para explicar Y.

Exemplo: (continuação)

- Considere o modelo que já tem X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_2?

Exemplo: (continuação)

- Considere o modelo que já tem X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_2?

$SQR(X_2|X_1) =$

Exemplo: (continuação)

- Considere o modelo que já tem X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_2?

$$SQR(X_2|X_1) = SQR(X_2, X_1) - SQR(X_1) =$$

Exemplo: (continuação)

- Considere o modelo que já tem X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_2?

$SQR(X_2|X_1) = SQR(X_2, X_1) - SQR(X_1) = 385.44 - 352.27 = 33.17$

ou

$SQR(X_2|X_1) =$

Exemplo: (continuação)

- Considere o modelo que já tem X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_2?

$SQR(X_2|X_1) = SQR(X_2, X_1) - SQR(X_1) = 385.44 - 352.27 = 33.17$

ou

$SQR(X_2|X_1) = SQE(X_1) - SQE(X_1, X_2) =$

Exemplo: (continuação)

- Considere o modelo que já tem X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon\,.$$

- Qual a contribuição de acrescentarmos X_2?

$$SQR(X_2|X_1) = SQR(X_2, X_1) - SQR(X_1) = 385.44 - 352.27 = 33.17$$

ou

$$SQR(X_2|X_1) = SQE(X_1) - SQE(X_1, X_2) = 143.12 - 109.95 = 33.17.$$

Exemplo: (continuação)

- Considere o modelo que já tem X_1 e X_2

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_3?

Exemplo: (continuação)

- Considere o modelo que já tem X_1 e X_2

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_3?

$SQR(X_3|X_1, X_2) =$

Exemplo: (continuação)

- Considere o modelo que já tem X_1 e X_2

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_3?

$$SQR(X_3|X_1, X_2) = SQR(X_3, X_2, X_1) - SQR(X_2, X_1) =$$

Exemplo: (continuação)

- Considere o modelo que já tem X_1 e X_2

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_3?

$$SQR(X_3|X_1, X_2) = SQR(X_3, X_2, X_1) - SQR(X_2, X_1) = 396.98 - 385.44$$

$$= 11.54$$

ou

$$SQR(X_3|X_1.X_2) =$$

Exemplo: (continuação)

- Considere o modelo que já tem X_1 e X_2

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_3?

$$SQR(X_3|X_1, X_2) = SQR(X_3, X_2, X_1) - SQR(X_2, X_1) = 396.98 - 385.44$$

$$= 11.54$$

ou

$$SQR(X_3|X_1.X_2) = SQE(X_1, X_2) - SQE(X_1, X_2, X_3) =$$

Exemplo: (continuação)

- Considere o modelo que já tem X_1 e X_2

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_3?

$$SQR(X_3|X_1, X_2) = SQR(X_3, X_2, X_1) - SQR(X_2, X_1) = 396.98 - 385.44$$

$$= 11.54$$

ou

$$SQR(X_3|X_1.X_2) = SQE(X_1, X_2) - SQE(X_1, X_2, X_3) = 109.95 - 98.41$$

$$= 11.54 \, .$$

Exemplo: (continuação)

- Considere o modelo que já tem X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_2 e X_3?

Exemplo: (continuação)

- Considere o modelo que já tem X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_2 e X_3?

$SQR(X_2, X_3 | X_1) =$

Exemplo: (continuação)

- Considere o modelo que já tem X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_2 e X_3?

$$SQR(X_2, X_3 | X_1) = SQR(X_3, X_2, X_1) - SQR(X_1) =$$

Exemplo: (continuação)

- Considere o modelo que já tem X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_2 e X_3?

$$SQR(X_2, X_3|X_1) = SQR(X_3, X_2, X_1) - SQR(X_1) = 396.98 - 352.27$$

$$= 44.71$$

ou

$$SQR(X_3|X_1.X_2) =$$

Exemplo: (continuação)

- Considere o modelo que já tem X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_2 e X_3?

$$SQR(X_2, X_3|X_1) = SQR(X_3, X_2, X_1) - SQR(X_1) = 396.98 - 352.27$$

$$= 44.71$$

ou

$$SQR(X_3|X_1.X_2) = SQE(X_1) - SQE(X_1, X_2, X_3) =$$

Exemplo: (continuação)

- Considere o modelo que já tem X_1

$$Y = \beta_0 + \beta_1 X_1 + \epsilon \, .$$

- Qual a contribuição de acrescentarmos X_2 e X_3?

$$SQR(X_2, X_3|X_1) = SQR(X_3, X_2, X_1) - SQR(X_1) = 396.98 - 352.27$$

$$= 44.71$$

ou

$$SQR(X_3|X_1.X_2) = SQE(X_1) - SQE(X_1, X_2, X_3) = 143.12 - 98.41$$

$$= 44.71 \, .$$

Soma de Quadrados Extras

- Envolve a diferença entre:
  - Soma de Quadrados Residual do modelo contendo as variáveis X já presentes.
  - Soma de Quadrados Residual do modelo contendo estas X e as novas X.
- Ou ainda a diferença entre:
  - Soma de Quadrados da Regressão do modelo contendo as variáveis X já presentes.
  - Soma de Quadrados da Regressão do modelo contendo estas X e as novas X.

Tabela ANOVA com Decomposição da Soma de Quadrados da Regressão

- A Soma de Quadrados da Regressão pode ser decomposta na:
  - Soma dos Quadrados Extras individuais relativas à entrada no modelo de uma variável por vez.
- Cada uma dessas somas individuais tem um grau de liberdade.
- Pois elas equivalem a acrescentar uma variável ao modelo.
- Podemos usar testes F a fim de avaliar quais variáveis incluir no modelo.

Exemplo de Tabela ANOVA com Decomposição da Soma de Quadrados da Regressão

Fonte de Variação	Soma de Quadrados	G.L.	Quadrado Médio
Regressão	$SQR(X_1, X_2, X_3)$	3	$QMR(X_1, X_2, X_3)$
X_1	$SQR(X_1)$	1	$QMR(X_1)$
$X_2 \mid X_1$	$SQR(X_2 \mid X_1)$	1	$QMR(X_2, \mid X_1)$
$X_3 \mid X_1, X_2$	$SQR(X_3 \mid X_1, X_2)$	1	$QMR(X_3 \mid X_2, X_1)$
Residual	$SQE(X_1, X_2, X_3)$	$n-4$	$QME(X_1, X_2, X_3)$
Total	SQT	$n-1$	

- A Soma de Quadrados Extras ao se acrescentar duas variáveis é dada pelas somas individuais

$$SQR(X_2, X_3|X_1) = SQR(X_2|X_1) + SQR(X_3|X_1, X_2)\,.$$

- Essa soma tem dois graus de liberdade.
- Pois estamos acrescentando duas variávies ao modelo.

Exemplo

- Vamos retomar o exemplo da gordura corporal.
- Vimos que a SQR do modelo com as três variáveis é dada por

$$SQR = 396.98 \text{ com}$$

Exemplo

- Vamos retomar o exemplo da gordura corporal.
- Vimos que a SQR do modelo com as três variáveis é dada por

$$SQR = 396.98 \text{ com } 3 \text{ graus de liberdade.}$$

- A SQR do modelo apenas com a variável X_1 é dada por

$$SQR = 352.27 \text{ com}$$

Exemplo

- Vamos retomar o exemplo da gordura corporal.
- Vimos que a SQR do modelo com as três variáveis é dada por

$$SQR = 396.98 \text{ com 3 graus de liberdade.}$$

- A SQR do modelo apenas com a variável X_1 é dada por

$$SQR = 352.27 \text{ com 1 grau de liberdade.}$$

- A Soma de Quadrado Extra ao acresentarmos X_2 no modelo que só tinha X_1 é dada por

$$SQR(X_2|X_1) =$$

Exemplo

- Vamos retomar o exemplo da gordura corporal.
- Vimos que a SQR do modelo com as três variáveis é dada por

$$SQR = 396.98 \text{ com 3 graus de liberdade.}$$

- A SQR do modelo apenas com a variável X_1 é dada por

$$SQR = 352.27 \text{ com 1 grau de liberdade.}$$

- A Soma de Quadrado Extra ao acresentarmos X_2 no modelo que só tinha X_1 é dada por

$$SQR(X_2|X_1) = SQR(X_2, X_1) - SQR(X_1) = 33.17$$

com

Exemplo

- Vamos retomar o exemplo da gordura corporal.
- Vimos que a SQR do modelo com as três variáveis é dada por

$$SQR = 396.98 \text{ com 3 graus de liberdade.}$$

- A SQR do modelo apenas com a variável X_1 é dada por

$$SQR = 352.27 \text{ com 1 grau de liberdade.}$$

- A Soma de Quadrado Extra ao acresentarmos X_2 no modelo que só tinha X_1 é dada por

$$SQR(X_2|X_1) = SQR(X_2, X_1) - SQR(X_1) = 33.17$$

com 1 grau de liberdade.

Exemplo (continuação)

- A Soma de Quadrado Extra ao acresentarmos X_3 no modelo que tinha X_1 e X_2 é dada por

$$SQR(X_3|X_1, X_2) =$$

Exemplo (continuação)

- A Soma de Quadrado Extra ao acresentarmos X_3 no modelo que tinha X_1 e X_2 é dada por

$$SQR(X_3|X_1, X_2) = SQR(X_3, X_2, X_1) - SQR(X_1, X_2) = 11.54$$

com

Exemplo (continuação)

- A Soma de Quadrado Extra ao acresentarmos X_3 no modelo que tinha X_1 e X_2 é dada por

$$SQR(X_3|X_1, X_2) = SQR(X_3, X_2, X_1) - SQR(X_1, X_2) = 11.54$$

com 1 grau de liberdade.

Exemplo (continuação)

- A Tabela ANOVA fica da seguinte forma

Fonte de Variação	Soma de Quadrados	G.L.	Quadrado Médio
Regressão	$SQR(X_1, X_2, X_3) =$		

Exemplo (continuação)

- A Tabela ANOVA fica da seguinte forma

Fonte de Variação	Soma de Quadrados	G.L.	Quadrado Médio
Regressão	$SQR(X_1, X_2, X_3) = 396.98$	3	132.33
X_1	$SQR(X_1) =$		

Exemplo (continuação)

- A Tabela ANOVA fica da seguinte forma

Fonte de Variação	Soma de Quadrados	G.L.	Quadrado Médio
Regressão	$SQR(X_1, X_2, X_3) = 396.98$	3	132.33
X_1	$SQR(X_1) = 352.27$	1	352.27
$X_2 \mid X_1$	$SQR(X_2 \mid X_1) =$		

Exemplo (continuação)

- A Tabela ANOVA fica da seguinte forma

Fonte de Variação	Soma de Quadrados	G.L.	Quadrado Médio
Regressão	$SQR(X_1, X_2, X_3) = 396.98$	3	132.33
X_1	$SQR(X_1) = 352.27$	1	352.27
$X_2 \mid X_1$	$SQR(X_2 \mid X_1) = 33.17$	1	33.17
$X_3 \mid X_1, X_2$	$SQR(X_3 \mid X_1, X_2) = 11.54$	1	11.54
Residual	$SQE(X_1, X_2, X_3) = 98.41$	16	6.15
Total	495.39	19	

Exemplo (continuação)

- Se quisermos saber o ganho de acrescentar duas variáveis, basta somar os quadrados individuais.
- Por exemplo, se quisermos saber o ganho de se acrescentar X_2 e X_3.
- A Soma de Quadrados Extras é dada por

$$SQR(X_2, X_3|X_1) =$$

Exemplo (continuação)

- Se quisermos saber o ganho de acrescentar duas variáveis, basta somar os quadrados individuais.
- Por exemplo, se quisermos saber o ganho de se acrescentar X_2 e X_3.
- A Soma de Quadrados Extras é dada por

$$SQR(X_2, X_3|X_1) = SQR(X_2|X_1) + SQR(X_3|X_2, X_1) =$$

Exemplo (continuação)

- Se quisermos saber o ganho de acrescentar duas variáveis, basta somar os quadrados individuais.
- Por exemplo, se quisermos saber o ganho de se acrescentar X_2 e X_3.
- A Soma de Quadrados Extras é dada por

$$SQR(X_2, X_3|X_1) = SQR(X_2|X_1) + SQR(X_3|X_2, X_1) =$$

$$33.17 + 11.54 = 44.71 \, .$$

- Essa soma tem quantos graus de liberdade?

Exemplo (continuação)

- Se quisermos saber o ganho de acrescentar duas variáveis, basta somar os quadrados individuais.
- Por exemplo, se quisermos saber o ganho de se acrescentar X_2 e X_3.
- A Soma de Quadrados Extras é dada por

$$SQR(X_2, X_3|X_1) = SQR(X_2|X_1) + SQR(X_3|X_2, X_1) =$$

$$33.17 + 11.54 = 44.71 \, .$$

- Essa soma tem quantos graus de liberdade? 2.

Uso da Soma de Quadrados Extra nos Testes dos Coeficientes

- Vejamos agora como usar Soma de Quadrados Extra para testar se devemos ou não incluir variáveis no modelo.
- Considere o modelo

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \, .$$

- Podemos testar se cada variável é significativa individualmente.
- Queremos testar as hipóteses

$$H_0 : \beta_j = 0 \quad \text{vs} \quad H_1 : \beta_j \neq 0 \, .$$

- Já vimos como fazer isso usando o teste-t

$$t = \frac{\hat{\beta}_j}{\sqrt{Var(\hat{\beta}_j)}} \sim t_{n-4} \text{ sob } H_0.$$

- Nesse caso, testamos se a variável é significativa, dado que as demais estão no modelo.
- Por exemplo, se queremos testar

$$H_0 : \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_3 \neq 0 \, .$$

- Usamos

$$t = \frac{\hat{\beta}_3}{\sqrt{Var(\hat{\beta}_3)}} \sim t_{n-4} \text{ sob } H_0.$$

- Podemos fazer esse teste usando a estatística **F parcial**.
- A Soma de Quadrado Extra devido à inclusão de X_3 é dada por $SQR(X_3|X_1, X_2)$.
- Pode-se motrar que se $\beta_3 = 0$, então

$$E\left(\frac{SQR(X_3|X_1, X_2)}{1}\right) = E(QMR(X_3|X_1, X_2)) = \sigma^2$$

- Portanto se a razão

$$\frac{QMR(X_3|X_1, X_2)}{\sigma^2} >> 1$$

indica que

- Podemos fazer esse teste usando a estatística **F parcial**.
- A Soma de Quadrado Extra devido à inclusão de X_3 é dada por $SQR(X_3|X_1, X_2)$.
- Pode-se motrar que se $\beta_3 = 0$, então

$$E\left(\frac{SQR(X_3|X_1, X_2)}{1}\right) = E(QMR(X_3|X_1, X_2)) = \sigma^2$$

- Portanto se a razão

$$\frac{QMR(X_3|X_1, X_2)}{\sigma^2} >> 1$$

  indica que $\beta_3 \neq 0$.
- Não sabemos o valor de σ^2.
- Podemos estimar pelo

- Podemos fazer esse teste usando a estatística **F parcial**.
- A Soma de Quadrado Extra devido à inclusão de X_3 é dada por $SQR(X_3|X_1, X_2)$.
- Pode-se motrar que se $\beta_3 = 0$, então

$$E\left(\frac{SQR(X_3|X_1, X_2)}{1}\right) = E(QMR(X_3|X_1, X_2)) = \sigma^2$$

- Portanto se a razão

$$\frac{QMR(X_3|X_1, X_2)}{\sigma^2} >> 1$$

  indica que $\beta_3 \neq 0$.
- Não sabemos o valor de σ^2.
- Podemos estimar pelo

$$S^2 =$$

- Podemos fazer esse teste usando a estatística **F parcial**.
- A Soma de Quadrado Extra devido à inclusão de X_3 é dada por $SQR(X_3|X_1, X_2)$.
- Pode-se motrar que se $\beta_3 = 0$, então

$$E\left(\frac{SQR(X_3|X_1, X_2)}{1}\right) = E(QMR(X_3|X_1, X_2)) = \sigma^2$$

- Portanto se a razão

$$\frac{QMR(X_3|X_1, X_2)}{\sigma^2} >> 1$$

  indica que $\beta_3 \neq 0$.
- Não sabemos o valor de σ^2.
- Podemos estimar pelo

$$S^2 = \ = QME$$

- Podemos fazer esse teste usando a estatística **F parcial**.
- A Soma de Quadrado Extra devido à inclusão de X_3 é dada por $SQR(X_3|X_1, X_2)$.
- Pode-se motrar que se $\beta_3 = 0$, então

$$E\left(\frac{SQR(X_3|X_1, X_2)}{1}\right) = E(QMR(X_3|X_1, X_2)) = \sigma^2$$

- Portanto se a razão

$$\frac{QMR(X_3|X_1, X_2)}{\sigma^2} >> 1$$

  indica que $\beta_3 \neq 0$.
- Não sabemos o valor de σ^2.
- Podemos estimar pelo

$$S^2 = \ = QME = \frac{SQE}{n-p}.$$

- Usamos a SQE estimado do modelo com todas as variáveis.
- A estatística F é dada por

$$F_0 = \frac{SQR(X_3|X_1, X_2)/(1)}{SQE(X_1, X_2, X_3)/(n-4)} = \frac{QMR(X_3|X_1, X_2)}{QME(X_1, X_2, X_3)}$$

- Sob H_0 temos que

$$F_0 \sim F_{1,n-4} \, .$$

- Podemos também estar interessados em testar vários coeficientes simultanemente.
- Queremos verificar, por exemplo, se X_2 e X_3 podem ser retiradas do modelo.
- As hipóteses a serem testadas são

- Podemos também estar interessados em testar vários coeficientes simultanemente.
- Queremos verificar, por exemplo, se X_2 e X_3 podem ser retiradas do modelo.
- As hipóteses a serem testadas são

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_2 \neq 0 \text{ e/ou } \beta_3 \neq 0 \, .$$

- Devemos comparar o acréscimo na SQR ao acrescentarmos as duas variáveis com a SQE do modelo todo.
- A estatística F é dada por

$$F_0 = \frac{SQR(X_2, X_3|X_1)/2}{SQE(X_1, X_2, X_3)/(n-4)} = \frac{QMR(X_3|X_1, X_2)}{QME(X_1, X_2, X_3)} \, .$$

- Sob H_0

$$F_0 \sim F_{2,n-4} \, .$$

Exemplo:

- Considere novamente o exemplo da gordura corporal.
- Queremos verificar se devemos de fato incluir a variável X_3 no modelo.
- As hipóteses a serem testadas são

$$H_0 : \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_3 \neq 0 \, .$$

- A estatística F é dada por

$$F_0 =$$

Exemplo:

- Considere novamente o exemplo da gordura corporal.
- Queremos verificar se devemos de fato incluir a variável X_3 no modelo.
- As hipóteses a serem testadas são

$$H_0 : \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_3 \neq 0 \, .$$

- A estatística F é dada por

$$F_0 = \frac{SQR(X_3|X_1, X_2)/(1)}{SQE(X_1, X_2, X_3)/(n-4)} = \frac{QMR(X_3|X_1, X_2)}{QME(X_1, X_2, X_3)}$$

$$=$$

Exemplo:

- Considere novamente o exemplo da gordura corporal.
- Queremos verificar se devemos de fato incluir a variável X_3 no modelo.
- As hipóteses a serem testadas são

$$H_0 : \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_3 \neq 0 \, .$$

- A estatística F é dada por

$$F_0 = \frac{SQR(X_3|X_1, X_2)/(1)}{SQE(X_1, X_2, X_3)/(n-4)} = \frac{QMR(X_3|X_1, X_2)}{QME(X_1, X_2, X_3)}$$

$$= \frac{11.54}{6.15} = 1.88 \, .$$

- Fixando $\alpha = 5\%$ temos que $F_{1,16} = 8.53$.
- Conclusão:

Exemplo:

- Considere novamente o exemplo da gordura corporal.
- Queremos verificar se devemos de fato incluir a variável X_3 no modelo.
- As hipóteses a serem testadas são

$$H_0 : \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_3 \neq 0 \, .$$

- A estatística F é dada por

$$F_0 = \frac{SQR(X_3|X_1, X_2)/(1)}{SQE(X_1, X_2, X_3)/(n-4)} = \frac{QMR(X_3|X_1, X_2)}{QME(X_1, X_2, X_3)}$$

$$= \frac{11.54}{6.15} = 1.88 \, .$$

- Fixando $\alpha = 5\%$ temos que $F_{1,16} = 8.53$.
- Conclusão: Não rejeitamos H_0.
- Com 5% de significância não podemos afirmar que a variável circunferência do antebraço é significativa para explicar a gordura corporal.

Exemplo: (continuação)

- Suponha que as hipóteses a serem testadas são

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_2 \neq 0 \text{ e/ou } \beta_3 \neq 0 \, .$$

- A estatística de teste é dada por

$$F_0 =$$

Exemplo: (continuação)

- Suponha que as hipóteses a serem testadas são

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_2 \neq 0 \text{ e/ou } \beta_3 \neq 0 \, .$$

- A estatística de teste é dada por

$$F_0 = \frac{SQR(X_2, X_3|X_1)/2}{SQE(X_1, X_2, X_3)/(n-4)} = \frac{QMR(X_3|X_1, X_2)}{QME(X_1, X_2, X_3)}$$

$$= \frac{44.71/2}{6.15} = 3.635 \, .$$

- Para $\alpha = 0.05$, $F_{2,16} = 3.63$.
- Conclusão:

Exemplo: (continuação)

- Suponha que as hipóteses a serem testadas são

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0 \quad \text{vs} \quad H_1 : \beta_2 \neq 0 \text{ e/ou } \beta_3 \neq 0 \, .$$

- A estatística de teste é dada por

$$F_0 = \frac{SQR(X_2, X_3|X_1)/2}{SQE(X_1, X_2, X_3)/(n-4)} = \frac{QMR(X_3|X_1, X_2)}{QME(X_1, X_2, X_3)}$$

$$= \frac{44.71/2}{6.15} = 3.635 \, .$$

- Para $\alpha = 0.05$, $F_{2,16} = 3.63$.
- Conclusão: Rejeitamos H_0.
- Com 5% de significância, pelo menos uma das duas variáveis é significativa.

Variáveis Indicadoras

- As variáveis indicadoras também são conhecidas como variáveis Dummy.
- Elas vão representar uma variável categórica nominal.
- Recebem valores zero ou um de acordo com a presença ou não da categoria.
- Em séries temporais:
  - indicam o acontecimento de um evento atípico, como guerra ou crise financeira.
- Em regressão linear:
  - pode representar o sexo dos indivíduos.

- A variáveis categóricas podem também ser vistas como subgrupos dos dados.
- Nesse caso as variáveis indicadoras recebem 1 se o indivíduo pertence ao grupo e 0 caso contrário.
- Elas permitem representar vários grupos em uma única equação.
- Não precisamos escrever uma equação para cada grupo.
- As variáveis indicadoras podem ser tratadas como qualquer outra no modelo de regressão.

- Suponha que queremos comparar dois grupos:
  - grupo controle e tratamento.
- Vamos definir a variável Z_i tal que

$$Z_i = \begin{cases} 1 & \text{se o i-ésimo indivíduo percente ao grupo tratamento} \\ 0 & \text{se o i-ésimo indivíduo percente ao grupo controle.} \end{cases}$$

- O modelo fica

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi} + \alpha_1 Z_i + \epsilon_i$$

onde

- Y_i é a variável resposta,
- X_i's são variáveis explicativas quantitativas.

- Qual parâmetro estima a difrenreça entre os grupos?

- Qual parâmetro estima a difrenreça entre os grupos? α_1.
- Vamos ver como fica o modelo para cada um dos grupos.
- Se o indivíduo pertence ao grupo controle a equação fica

- Qual parâmetro estima a difrenreça entre os grupos? α_1.
- Vamos ver como fica o modelo para cada um dos grupos.
- Se o indivíduo pertence ao grupo controle a equação fica

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi} + \epsilon_i, \text{ pois } Z_i = 0 \, .$$

- Se ele pertence ao grupo tratamento

- Qual parâmetro estima a difrenreça entre os grupos? α_1.
- Vamos ver como fica o modelo para cada um dos grupos.
- Se o indivíduo pertence ao grupo controle a equação fica

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi} + \epsilon_i, \text{ pois } Z_i = 0 \, .$$

- Se ele pertence ao grupo tratamento

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi} + \alpha_1 + \epsilon_i, \text{ pois } Z_i = 1 \, .$$

- Ou seja

$$E(Y_i|\mathbf{X}, Z_i = 0) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi}$$

$$E(Y_i|\mathbf{X}, Z_i = 1) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_p X_{pi} + \alpha_1 \, .$$

- A diferença das médias dos dois grupos é dada por

$$E(Y_i|\mathbf{X}, Z_i = 1) - E(Y_i|\mathbf{X}, Z_i = 0) =$$

- A diferença das médias dos dois grupos é dada por

$$E(Y_i|\mathbf{X}, Z_i = 1) - E(Y_i|\mathbf{X}, Z_i = 0) = \alpha_1 \ .$$

- Evitamos então ter que criar uma equação para cada grupo.

Observação

- Suponha que a nossa variável categórica tem k categorias ou k grupos.
- Só precisamos de $k - 1$ variáveis indicadoras.
- Uma delas (categoria de referência) recebe valor 0 para todas indicadoras.
- Assim as demais categorias são comparadas com relação à categoria de referência.

Exemplo:

- Queremos modelar duas variáveis:
  - peso do peru (em libras);
  - idade do peru (em semanas).
- Qual a variável resposta?

Exemplo:

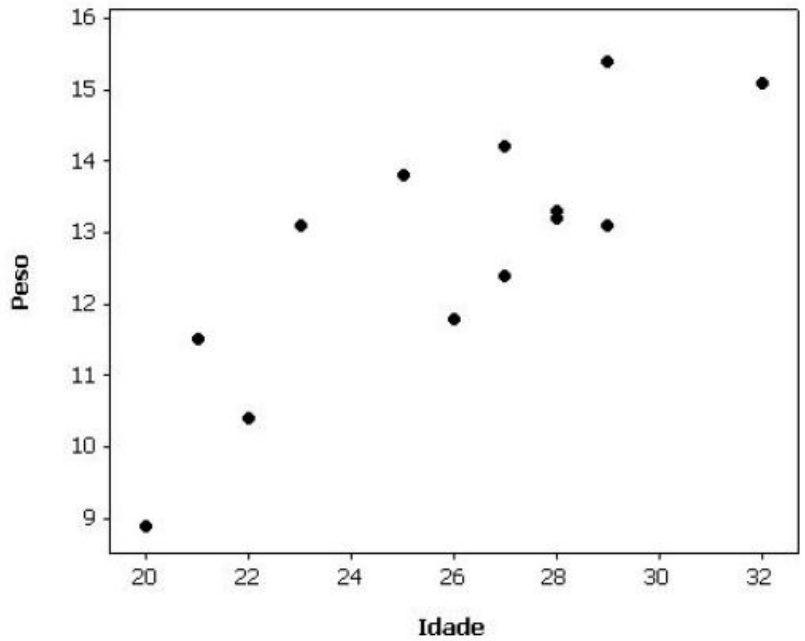
- Queremos modelar duas variáveis:
  - peso do peru (em libras);
  - idade do peru (em semanas).
- Qual a variável resposta? Peso.
- Espera-se uma relação positiva ou negativa?

Exemplo:

- Queremos modelar duas variáveis:
  - peso do peru (em libras);
  - idade do peru (em semanas).
- Qual a variável resposta? Peso.
- Espera-se uma relação positiva ou negativa? Positiva.
- Quanto maior a idade, maior o peso.

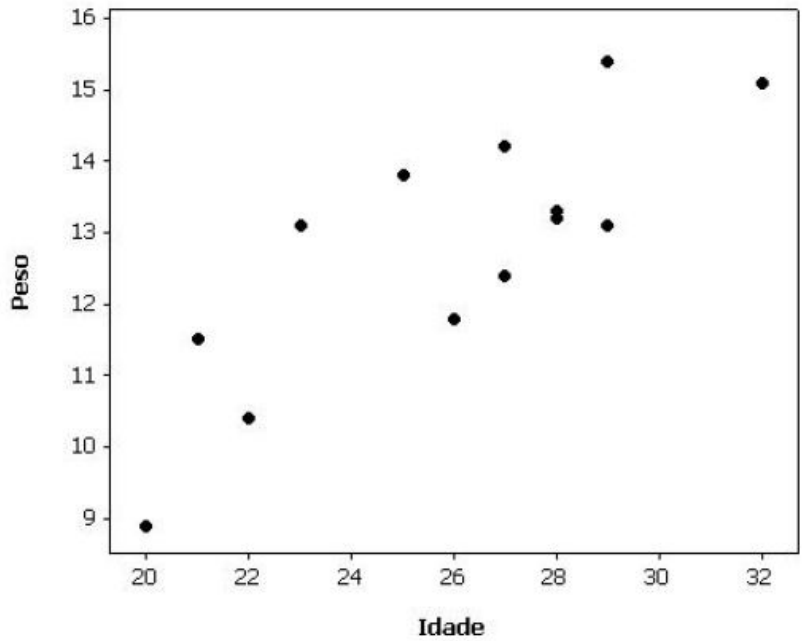
Exemplo: (continuação)

- A figura a seguir mostra o gráfico de dispersão entre as variáveis.



Exemplo: (continuação)

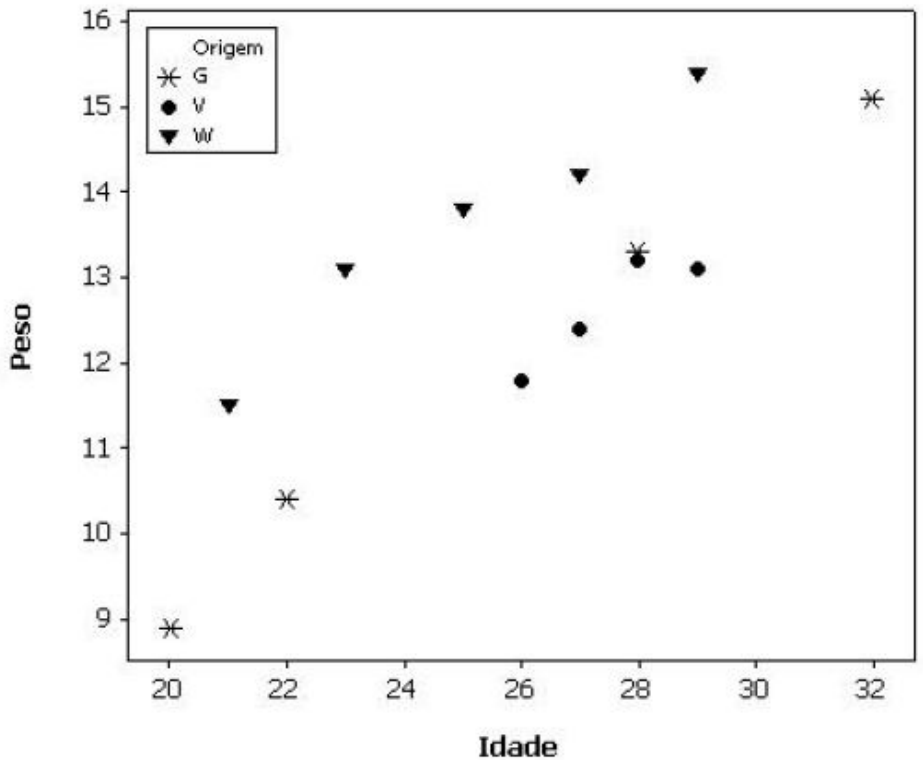
- A figura a seguir mostra o gráfico de dispersão entre as variáveis.



- Parece existir uma relação linear positiva entre as variáveis.

Exemplo: (continuação)

- Talvez o local de onde vem o peru pode estar influenciando.
- Os perus podem vir de um dos seguintes estados:
  - Geórgia, Virgínia, Winsconsin.



Exemplo: (continuação)

- Criamos duas variáveis indicadoras para representar o estado de onde os perus foram originados.

Origem	Z_1	Z_2
Geórgia	1	0
Virgínia	0	1
Wisconsin	0	0

- Se o peru pertence à Geórgia,
  - sua média fica somada de

Exemplo: (continuação)

- Criamos duas variáveis indicadoras para representar o estado de onde os perus foram originados.

Origem	Z_1	Z_2
Geórgia	1	0
Virgínia	0	1
Wisconsin	0	0

- Se o peru pertence à Geórgia,
  - sua média fica somada de Z_1.
- Se pertence à Virgínia,
  - sua média é somada de

Exemplo: (continuação)

- Criamos duas variáveis indicadoras para representar o estado de onde os perus foram originados.

Origem	Z_1	Z_2
Geórgia	1	0
Virgínia	0	1
Wisconsin	0	0

- Se o peru pertence à Geórgia,
  - sua média fica somada de Z_1.
- Se pertence à Virgínia,
  - sua média é somada de Z_2.
- Se pertence a Wisconsin,
  - sua média não se altera.

Exemplo: (continuação)

- A tabela a seguir mostra os dados do problema com as variáveis indicadoras.

Y=Peso	X=Idade	Origem	Z_1	Z_2
13.3	28	G	1	0
8.9	20	G	1	0
15.1	32	G	1	0
10.4	22	G	1	0
13.1	29	V	0	1
12.4	27	V	0	1
13.2	28	V	0	1
11.8	26	V	0	1
11.5	21	W	0	0
14.2	27	W	0	0
15.4	29	W	0	0
13.1	23	W	0	0
13.8	25	W	0	0

Figura:

Exemplo: (continuação)

- O modelo é dado por

$$\underbrace{Y_i}_{peso} = \beta_0 + \beta_1 \underbrace{X_i}_{idade} + \underbrace{\alpha_1 Z_{1i} + \alpha_2 Z_{2i}}_{origem} + \epsilon_i$$

- Ajustamos o modelo usando mínimos quadrados, da maneira usual.
- A reta ajustada é dada por

$$\underbrace{Y_i}_{peso} = 1.43 + 0.487 \underbrace{X_i}_{idade} \underbrace{-1.92 Z_{1i} - 2.19 Z_{2i}}_{origem} .$$

Exemplo: (continuação)

- A Tabela ANOVA é apresentada a seguir.

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão	$p = 3$	$SQR = 38.606$	QMR=12.869
Resíduo	$n - p - 1 = 9$	$SQE = 0.811$	QME=0.09
Total	$n - 1 = 12$	$SQT = 39.417$	

Tabela: Tabela ANOVA

Exemplo: (continuação)

- Vamos fazer o teste F para esses dados.
- Queremos testar as seguintes hipóteses:

Exemplo: (continuação)

- Vamos fazer o teste F para esses dados.
- Queremos testar as seguintes hipóteses:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0, \text{ ou seja, nenhum dos } \beta_j \text{ é significante}$$

$$H_1 : \beta_j \neq 0 \text{ para pelo menos um } j \geq 1, \text{ ou seja, } \beta_j \text{ é significante}$$

- A estatística de teste é dada por

Exemplo: (continuação)

- Vamos fazer o teste F para esses dados.
- Queremos testar as seguintes hipóteses:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0, \text{ ou seja, nenhum dos } \beta_j \text{ é significante}$$

$$H_1 : \beta_j \neq 0 \text{ para pelo menos um } j \geq 1, \text{ ou seja, } \beta_j \text{ é significante}$$

- A estatística de teste é dada por

$$F = \frac{QMR}{QME} \sim F_{p,n-p-1} \text{ sob } H_0 \, .$$

Exemplo: (continuação)

- A estatística F é dada por

$$F = \frac{QMR}{QME} = \frac{12.869}{0.09} = 142.72\,.$$

- Sob H_0 temos que

$$F \sim F_{3,9}$$

- Fixando $\alpha = 0.05$ temos que $F_{3,9;0.05} \approx F_{3,9;0.05} = 3.863$.
- A região crítica é dada por

$$F > 3.863\,.$$

- Qual a conclusão?

Exemplo: (continuação)

- A estatística F é dada por

$$F = \frac{QMR}{QME} = \frac{12.869}{0.09} = 142.72\,.$$

- Sob H_0 temos que

$$F \sim F_{3,9}$$

- Fixando $\alpha = 0.05$ temos que $F_{3,9;0.05} \approx F_{3,9;0.05} = 3.863$.
- A região crítica é dada por

$$F > 3.863\,.$$

- Qual a conclusão? Como $F_{obs} = 142.72 > 3.863$, rejeitamos H_0.
- O que isso significa?

Exemplo: (continuação)

- A estatística F é dada por

$$F = \frac{QMR}{QME} = \frac{12.869}{0.09} = 142.72\,.$$

- Sob H_0 temos que

$$F \sim F_{3,9}$$

- Fixando $\alpha = 0.05$ temos que $F_{3,9;0.05} \approx F_{3,9;0.05} = 3.863$.
- A região crítica é dada por

$$F > 3.863\,.$$

- Qual a conclusão? Como $F_{obs} = 142.72 > 3.863$, rejeitamos H_0.
- O que isso significa? Com 5% de significância, podemos dizer que pelo uma das variáveis (X_1, Z_1 ou Z_2) é significativa para explicar Y.

Exemplo: (continuação)

- Vejamos agora os resultados dos testes-t individuais.

Preditor	Estimativa Pontual	Erro Padrão	Estatística T	p-valor
Intercepto	1.4309	0.6574	2.18	0.058
Idade	0.4868	0.0257	18.91	0.000
Z_1	-1.0184	0.2018	-0.51	0.000
Z_2	-2.1919	0.2114	-10.37	0.000

- Quais variáveis são significativas ?

Exemplo: (continuação)

- Vejamos agora os resultados dos testes-t individuais.

Preditor	Estimativa Pontual	Erro Padrão	Estatística T	p-valor
Intercepto	1.4309	0.6574	2.18	0.058
Idade	0.4868	0.0257	18.91	0.000
Z_1	-1.0184	0.2018	-0.51	0.000
Z_2	-2.1919	0.2114	-10.37	0.000

- Quais variáveis são significativas ? Todas.
- **Obs.:** uma variável indicadora não ser significativa indica

Exemplo: (continuação)

- Vejamos agora os resultados dos testes-t individuais.

Preditor	Estimativa Pontual	Erro Padrão	Estatística T	p-valor
Intercepto	1.4309	0.6574	2.18	0.058
Idade	0.4868	0.0257	18.91	0.000
Z_1	-1.0184	0.2018	-0.51	0.000
Z_2	-2.1919	0.2114	-10.37	0.000

- Quais variáveis são significativas ? Todas.
- **Obs.:** uma variável indicadora não ser significativa indica não diferença entre grupos.

Exemplo: (continuação)

- Como vimos que existe diferença entre os grupos, vamos escrever as equações de cada um deles.
- Para os perus da Geógia

$$Y_i =$$

Exemplo: (continuação)

- Como vimos que existe diferença entre os grupos, vamos escrever as equações de cada um deles.
- Para os perus da Geógia

$$Y_i = \beta_0 + \beta_1 X_i + \alpha_1 + \epsilon_i$$

a reta ajustada é dada por

$$Y_i = (\hat{\beta}_0 + \hat{\alpha}_1) + \hat{\beta}_1 X_i + \epsilon_i$$

$$Y_i = (1.43 - 1.92) + 0.487 X_i \, .$$

- Interpretação:

Exemplo: (continuação)

- Como vimos que existe diferença entre os grupos, vamos escrever as equações de cada um deles.
- Para os perus da Geógia

$$Y_i = \beta_0 + \beta_1 X_i + \alpha_1 + \epsilon_i$$

a reta ajustada é dada por

$$Y_i = (\hat{\beta}_0 + \hat{\alpha}_1) + \hat{\beta}_1 X_i + \epsilon_i$$

$$Y_i = (1.43 - 1.92) + 0.487 X_i \, .$$

- Interpretação: o peso esperado de um pero da Geórgia é 1.92 libras menor do que o peso esperado de um peso de Winsconsin. (pois Winsconsin é a categoria de referência)

Exemplo: (continuação)

- Para os perus da Virgínia

$$Y_i =$$

Exemplo: (continuação)

- Para os perus da Virgínia

$$Y_i = \beta_0 + \beta_1 X_i + \alpha_2 + \epsilon_i$$

a reta ajustada é dada por

$$Y_i = (\hat{\beta}_0 + \hat{\alpha}_2) + \hat{\beta}_1 X_i + \epsilon_i$$

$$Y_i = (1.43 - 2.19) + 0.487 X_i \, .$$

- Interpretação:

Exemplo: (continuação)

- Para os perus da Virgínia

$$Y_i = \beta_0 + \beta_1 X_i + \alpha_2 + \epsilon_i$$

a reta ajustada é dada por

$$Y_i = (\hat{\beta}_0 + \hat{\alpha}_2) + \hat{\beta}_1 X_i + \epsilon_i$$

$$Y_i = (1.43 - 2.19) + 0.487 X_i \, .$$

- Interpretação: o peso esperado de um pero da Virgínia é 2.19 libras menor do que o peso esperado de um peso de Winsconsin.

Exemplo: (continuação)

- O gráfico a seguir mostra as retas ajustadas para cada um dos grupos.

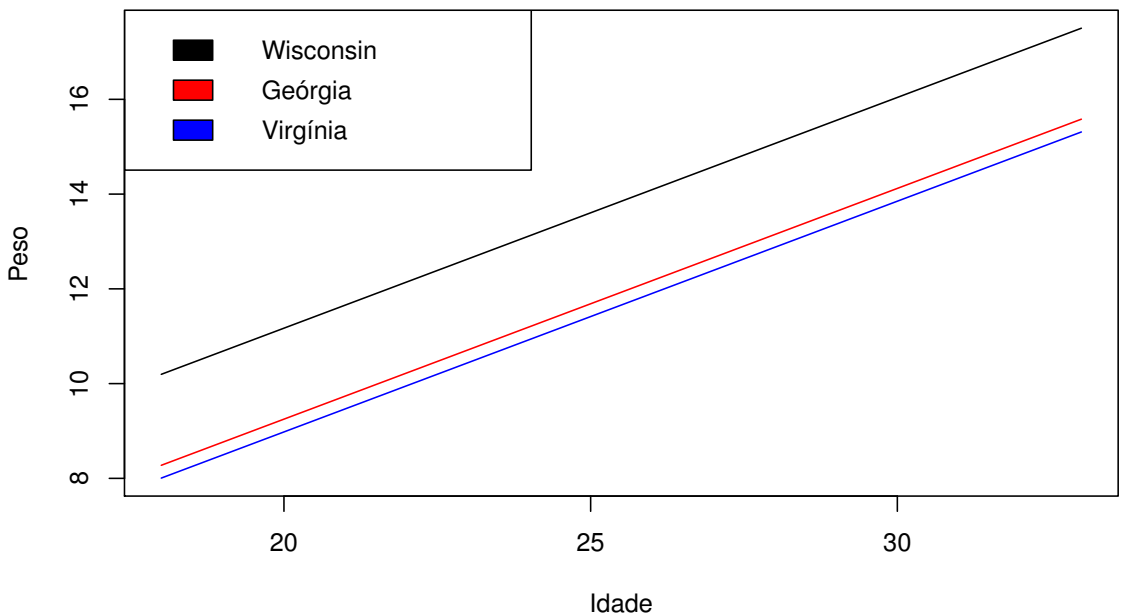


Figura:

Exemplo - Duas variáveis categóricas

- Podemos ter mais de uma variável categórica.
- Um engenheiro está estudando a vida útil de uma ferramenta de corte.
- Esse tempo pode depender das seguintes variáveis:
  - velocidade de rotação do torno (medida em r.p.m.);
  - tipo de ferramenta (A ou B);
  - tipo de óleo lubrificante (baixa ou média viscosidade).
- Variável indicadora para o tipo de ferramenta:

$$Z_1 = \begin{cases} 1 & \text{se o tipo de ferramenta é } A \\ 0 & \text{se o tipo de ferramenta é } B. \end{cases}$$

Exemplo - Duas variáveis categóricas

- Variável indicadora para o tipo de óleo:

$$Z_2 = \begin{cases} 1 & \text{se o óleo é de baixa viscosidade} \\ 0 & \text{se o óleo é de alta viscosidade.} \end{cases}$$

- O modelo geral é dado por

$$\underbrace{Y_i}_{\text{tempo}} = \beta_0 + \beta_1 \underbrace{X_i}_{\text{velocidade}} + \underbrace{\alpha_1 Z_{1i}}_{\text{ferramenta}} + \underbrace{\alpha_2 Z_{2i}}_{\text{óleo}} + \epsilon_i$$

Exemplo - Duas variáveis categóricas

- A tabela a seguir mostra os modelos para cada um dos grupos.

Ferramenta	Óleo	Modelo
A	baixa	$Y_i =$

Exemplo - Duas variáveis categóricas

- A tabela a seguir mostra os modelos para cada um dos grupos.

Ferramenta	Óleo	Modelo
A	baixa	$Y_i = \beta_0 + \beta_1 X_i + \epsilon$
B	baixa	$Y_i =$

Exemplo - Duas variáveis categóricas

- A tabela a seguir mostra os modelos para cada um dos grupos.

Ferramenta	Óleo	Modelo
A	baixa	$Y_i = \beta_0 + \beta_1 X_i + \epsilon$
B	baixa	$Y_i = \beta_0 + \beta_1 X_i + \alpha_1 + \epsilon$
A	média	$Y_i =$

Exemplo - Duas variáveis categóricas

- A tabela a seguir mostra os modelos para cada um dos grupos.

Ferramenta	Óleo	Modelo
A	baixa	$Y_i = \beta_0 + \beta_1 X_i + \epsilon$
B	baixa	$Y_i = \beta_0 + \beta_1 X_i + \alpha_1 + \epsilon$
A	média	$Y_i = \beta_0 + \beta_1 X_i + \alpha_2 + \epsilon$
B	média	$Y_i =$

Exemplo - Duas variáveis categóricas

- A tabela a seguir mostra os modelos para cada um dos grupos.

Ferramenta	Óleo	Modelo
A	baixa	$Y_i = \beta_0 + \beta_1 X_i + \epsilon$
B	baixa	$Y_i = \beta_0 + \beta_1 X_i + \alpha_1 + \epsilon$
A	média	$Y_i = \beta_0 + \beta_1 X_i + \alpha_2 + \epsilon$
B	média	$Y_i = \beta_0 + \beta_1 X_i + \alpha_1 + \alpha_2 + \epsilon$

Exemplo

- Problema: o que a discriminação sexual no emprego significa e como ela pode ser medido?
- Para responder a estas perguntas considere estes dados artificiais referentes ao emprego registros de uma amostra de funcionários de Ace Manufacturing.

i	Sex	Merit Pay	i	Sex	Merit Pay
Bob	M	9.6	Tim	M	6.0
Paul	M	8.3	George	M	1.1
Mary	F	4.2	Alan	M	9.2
John	M	8.8	Lisa	F	3.3
Nancy	F	2.1	Anne	F	2.7

i (c1)	Sex (c2)	Merit Pay (c3)	i (c1)	Sex (c2)	Merit Pay (c3)
Bob	1	9.6	Tim	1	6.0
Paul	1	8.3	George	1	1.1
Mary	0	4.2	Alan	1	9.2
John	1	8.8	Lisa	0	3.3
Nancy	0	2.1	Anne	0	2.7

Exemplo

- Obtemos a seguinte reta estimada:

$$\hat{Y} = 3,08 + 4,09X$$

Observações

- Poderíamos ajustar várias equações de regressão, uma para cada grupo.
- Por que não fazemos isso?
- O modelo assume a mesma variância σ^2 e o mesmo coeficiente β_1 para todos grupos.
- Se fizermos tudo junto teremos uma amostra maior do que se fizermos separados.
- As estimativas ficam melhores.
- Se fizermos separados os tamanhos de amostra são pequenos.