Modelos de Regressão Linear Simples - parte III

Erica Castilho Rodrigues

20 de Setembro de 2016

Análise de Variância

Intervalos de Confiança e Testes para β_0 e β_1

Análise de Variância

- A variável X é um bom preditor da resposta Y?
- Quanto da variação da variável resposta é explicada pela reta de regressão?
- A distância de cada indivíduo em relação à média do grupo é dada por

- A variável X é um bom preditor da resposta Y?
- Quanto da variação da variável resposta é explicada pela reta de regressão?
- A distância de cada indivíduo em relação à média do grupo é dada por
$$(Y_i - \overline{Y})^2 .$$
- A variândia da variável resposta é dada por

- A variável X é um bom preditor da resposta Y?
- Quanto da variação da variável resposta é explicada pela reta de regressão?
- A distância de cada indivíduo em relação à média do grupo é dada por
$$(Y_i - \overline{Y})^2 .$$
- A variândia da variável resposta é dada por
$$\sum_{i=1}^{n} (Y_i - \overline{Y})^2 .$$
- Se não tivéssemos o modelo de regressão, o melhor preditor para Y_i é $\overline{Y}$.

- A variação dos dados pode ser decomposta em uma soma de parcelas.
- Vamos ver como isso é feito.
- Somando e subtraindo $\overline{Y}$ temos que

$$Y_i - \hat{Y}_i =$$

- A variação dos dados pode ser decomposta em uma soma de parcelas.
- Vamos ver como isso é feito.
- Somando e subtraindo $\overline{Y}$ temos que

$$Y_i - \hat{Y}_i = (Y_i - \overline{Y}) + (\hat{Y}_i - \overline{Y})\,.$$

- Somando e subraindo $\hat{Y}_i$ temos que

$$(Y_i - \overline{Y}) =$$

- A variação dos dados pode ser decomposta em uma soma de parcelas.
- Vamos ver como isso é feito.
- Somando e subtraindo $\overline{Y}$ temos que

$$Y_i - \hat{Y}_i = (Y_i - \overline{Y}) + (\hat{Y}_i - \overline{Y})\,.$$

- Somando e subraindo $\hat{Y}_i$ temos que

$$(Y_i - \overline{Y}) = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \overline{Y})\,.$$

- Elevando a segunda igualdade ao quadrado

$$(Y_i - \overline{Y})^2 = (Y_i - \hat{Y}_i)^2 + 2(Y_i - \hat{Y}_i)(\hat{Y}_i - \overline{Y}) + (\hat{Y}_i - \overline{Y})^2\,.$$

- Somando a expressão

$$(Y_i - \overline{Y})^2 = (Y_i - \hat{Y}_i)^2 + 2(Y_i - \hat{Y}_i)(\hat{Y}_i - \overline{Y}) + (\hat{Y}_i - \overline{Y})^2$$

em *i* temos que

- Somando a expressão

$$(Y_i - \overline{Y})^2 = (Y_i - \hat{Y}_i)^2 + 2(Y_i - \hat{Y}_i)(\hat{Y}_i - \overline{Y}) + (\hat{Y}_i - \overline{Y})^2$$

em i temos que

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + 2\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \overline{Y}) + \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2.$$

- Vamos mostrar agora que

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \overline{Y}) = 0\,.$$

- Temos que

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \overline{Y}) =$$

- Temos que

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \overline{Y}) = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)\hat{Y}_i - \overline{Y}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)\,.$$

- Vamos olhar cada parcela separadamente.
- Para o segundo termo temos que

$$\overline{Y}\sum_{i=1}^{n}(Y_i - \hat{Y}_i) = \overline{Y}\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

- A segunda igualdade é válida pois é uma das equações normais.
- Lembre que

$$\left.\frac{\partial L}{\partial \beta_0}\right|_{\hat{\beta}_0, \hat{\beta}_1} =$$

- Temos que

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \overline{Y}) = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)\hat{Y}_i - \overline{Y}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)\,.$$

- Vamos olhar cada parcela separadamente.
- Para o segundo termo temos que

$$\overline{Y}\sum_{i=1}^{n}(Y_i - \hat{Y}_i) = \overline{Y}\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

- A segunda igualdade é válida pois é uma das equações normais.
- Lembre que

$$\left.\frac{\partial L}{\partial \beta_0}\right|_{\hat{\beta}_0, \hat{\beta}_1} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0\,.$$

- Para a segunda parcela temos que

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)\hat{Y}_i =$$

- Para a segunda parcela temos que

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)\hat{Y}_i = \sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 X_i)e_i = \hat{\beta}_0 \sum_{i=1}^{n} e_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i e_i$$

- Observe que

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) =$$

- Para a segunda parcela temos que

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)\hat{Y}_i = \sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1 X_i)e_i = \hat{\beta}_0 \sum_{i=1}^{n} e_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i e_i$$

- Observe que

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

pois é a primeira equação normal.

- Além disso

$$\sum_{i=1}^{n} X_i e_i = \sum_{i=1}^{n} X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) =$$

- Além disso

$$\sum_{i=1}^{n} X_i e_i = \sum_{i=1}^{n} X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

pois essa é a segunda equação normal.

- Lembre-se que

$$\left.\frac{\partial L}{\partial \beta_1}\right|_{\hat{\beta}_0, \hat{\beta}_1} =$$

- Além disso

$$\sum_{i=1}^{n} X_i e_i = \sum_{i=1}^{n} X_i (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0$$

pois essa é a segunda equação normal.

- Lembre-se que

$$\left.\frac{\partial L}{\partial \beta_1}\right|_{\hat{\beta}_0, \hat{\beta}_1} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0 \, .$$

- Dessa maneira temos que

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i) \hat{Y}_i = \sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 X_i) e_i = \hat{\beta}_0 \sum_{i=1}^{n} e_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i e_i$$

$$\hat{\beta}_0 0 + \hat{\beta}_1 0 = 0 \, .$$

- Mostramos assim que

$$\sum_{i=1}^{n}(Y_i - \hat{Y}_i)(\hat{Y}_i - \overline{Y}) = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)\hat{Y}_i - \overline{Y}\sum_{i=1}^{n}(Y_i - \hat{Y}_i) = 0\,.$$

- Portanto

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 + \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2\,.$$

- Então a soma de quadrados total pode ser decomposta em

$$\underbrace{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}_{\text{Soma de Quadrados Totais}} = \underbrace{\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2}_{\text{Soma de Quadrados da Regressão}} + \underbrace{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}_{\text{Soma de Quadrados dos Resíduos}} .$$

$$SQT = SQR + SQE .$$

- SQT $\Rightarrow$ mede a variação dos valores de Y na amostra.
- SQR $\Rightarrow$ mede o quanto da SQT é explicada pelo modelo de regressão ajustado.
- SQE $\Rightarrow$ mede o quanto da SQT **não** é explicada pelo modelo de regressão ajustado.

Tabela de Análise de Variância (Tabela ANOVA)

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão	1	$SQR = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2$	$QMR = SQR/1$
Residual	$n-2$	$SQE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$	$QME = SQE/(n-2)$
Total	$n-1$	$SQT = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$	

Tabela: Tabela ANOVA

Graus de Liberdade

- Número associado a uma soma de quadrados.
- Quantos pedaços independentes de informação envolvendo as n quantidades independentes $Y_1; \ldots; Y_n$ são necessários para calcular a soma de quadrados.

- **Soma de Quadrados Totais**

$$SQT = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

envolve $n - 1$ quantidades independentes $(Y_i - \overline{Y})$ pois $\sum_{i=1}^{n} (Y_i - \overline{Y}) = 0$.

- **Soma de Quadrados da Regressão**

$$SQR = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2$$

pode-se mostrar que $\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = \hat{\beta}_1^2 S_{XX}$.

- A única informação que vem do vetor $Y_1, \ldots, Y_n$ está em $\hat{\beta}_1$. Portanto temos apenas um grau de liberdade.

- **Soma de Quadrado dos Resíduos**

$$SQE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

por subtração, tem $(n-1)-1 = n-2$ graus de liberdade.

$$\text{Quadrado médio} = \frac{\text{Soma de Quadrados}}{\text{Graus de Liberdade}}$$

Coeficiente de Determinação (R^2)

Mede a proporção da variabilidade total dos dados explicada pelo modelo de regressão.

$$R^2 = \frac{\text{Soma de Quadrados da Regressão}}{\text{Soma de Quadrados totais}} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$$

Observações

- $R^2 \leq 1$
- $R^2 = 1$ somente se o ajuste do modelo é perfeito.
- Toda a variabilidade dos dados é explicada pelo modelo

$$Y_i = \hat{Y}_i \quad \text{para todo } i.$$

- É possível mostrar que

$$R^2 = (\text{coeficiente de correlação entre } X \text{ e } Y)^2$$

ou seja, $R =$ coeficiente de correlação entre X e Y.

- Vejamos porque isso é verdade.
- Vimos que
$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} .$$
- Pode-se mostrar que (lista de exercício)
$$\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = \hat{\beta}_1^2 S_{XX} .$$
- Substituindo $\hat{\beta}_1$ temos
$$\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 =$$

- Vejamos porque isso é verdade.
- Vimos que

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} .$$

- Pode-se mostrar que (lista de exercício)

$$\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = \hat{\beta}_1^2 S_{XX} .$$

- Substituindo $\hat{\beta}_1$ temos

$$\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = \left(\frac{S_{XY}}{S_{XX}}\right)^2 S_{XX} =$$

- Vejamos porque isso é verdade.
- Vimos que
$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} \, .$$
- Pode-se mostrar que (lista de exercício)
$$\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = \hat{\beta}_1^2 S_{XX} \, .$$
- Substituindo $\hat{\beta}_1$ temos
$$\sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 = \left(\frac{S_{XY}}{S_{XX}} \right)^2 S_{XX} = \frac{S_{XY}^2}{S_{XX}} \, .$$

- Então
$$R^2 = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} = \frac{\frac{S_{XY}^2}{S_{XX}}}{S_{YY}}$$
pois $\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = S_{YY}$.
- Ou seja
$$R^2 = \frac{S_{XY}^2}{S_{XX}S_{YY}}$$
- Mas sabemos que o coeficiente de correlação de Pearson é dado por
$$r = \frac{n\sum_i x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n(\sum_i x_i^2) - (\sum_i x_i)^2}\sqrt{n(\sum_i y_i^2) - (\sum_i y_i)^2}} = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$$
- Mostramos assim que
$$R^2 = (\text{coeficiente de correlação entre } X \text{ e } Y)^2.$$

Exemplo

- Vamos analisar a relação entre duas variáveis:
  - temperatura atmosférica média do mês;
  - consumo mensal de gás residencial.
- Qual é a variável resposta?

Exemplo

- Vamos analisar a relação entre duas variáveis:
  - temperatura atmosférica média do mês;
  - consumo mensal de gás residencial.
- Qual é a variável resposta? Consumo.
- Qual é a variável explicativa?

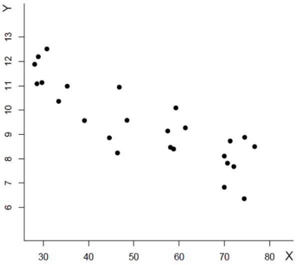
Exemplo

- Vamos analisar a relação entre duas variáveis:
  - temperatura atmosférica média do mês;
  - consumo mensal de gás residencial.
- Qual é a variável resposta? Consumo.
- Qual é a variável explicativa? Temperatura.

Exemplo (continuação)

- Abaixo encontram-se os dados coletados e o gráfico de dispersão das duas variáveis.

X_i	Y_i
35,3	10,98
29,7	11,13
30,8	12,51
58,8	8,40
61,4	9,27
71,3	8,73
74,4	6,36
76,7	8,50
70,7	7,82
57,5	9,14
46,4	8,24
28,9	12,19
28,1	11,88
39,1	9,57
46,8	10,94
48,5	9,58
59,3	10,09
70,0	8,11
70,0	6,83
74,5	8,88
72,1	7,68
58,1	8,47
44,6	8,86
33,4	10,36
28,6	11,08



Exemplo (continuação)

- A partir dos dados coletados temos que

$$n = 25 \quad \sum Y_i == 235,60 \quad \overline{Y} = 9,424 \quad \sum X_i = 1315$$

$$\overline{X} = 52,60 \quad \sum X_i Y_i = 11821,4320 \quad \sum X_i^2 = 76323,42$$

- A partir desses dados podemos calcular os coeficientes da reta

$$\hat{\beta}_1 = -0,079829 \quad \hat{\beta}_0 = 13,623005\,.$$

- O modelo ajustado é

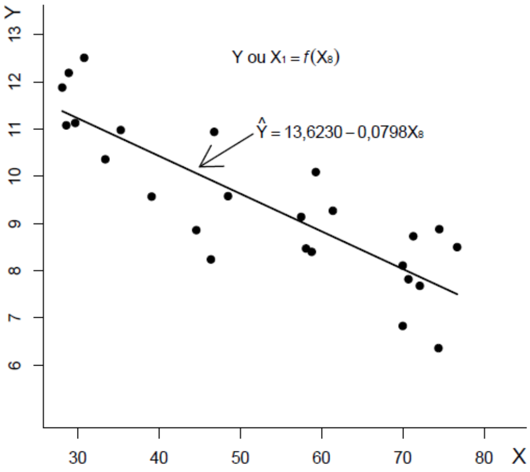
$$Y = 13,623005 - 0,079829X + \epsilon$$

onde $\epsilon \sim N(0, \sigma^2)$.

- Qual a interpretação dos parâmetros?

Exemplo (continuação)

- O gráfico asseguir apresenta a reta ajustada.



Exemplo (continuação)

- Precisamos dos seguintes dados para calcular as Somas de Quadrados.

X_i	Y_i	$\hat{Y}_i$	$Y_i - \hat{Y}_i$
35,3	10,98	10,81	0,17
29,7	11,13	11,25	-0,12
30,8	12,51	11,17	1,34
58,8	8,40	8,93	-0,53
61,4	9,27	8,72	0,55
71,3	8,73	7,93	0,80
74,4	6,36	7,68	-1,32
76,7	8,50	7,50	1,00
70,7	7,82	7,98	-0,16
57,5	9,14	9,03	0,11
46,4	8,24	9,92	-1,68
28,9	12,19	11,32	0,87
28,1	11,88	11,38	0,50
39,1	9,57	10,50	-0.93
46,8	10,94	9,89	1,05
48,5	9,58	9,75	-0,17
59,3	10,09	8,89	1,20
70,0	8,11	8,03	0,08
70,0	6,83	8,03	-1,20
74,5	8,88	7,68	1,20
72,1	7,68	7,87	-0,19
58,1	8,47	8,98	-0,51
44,6	8,86	10,06	-1,20
33,4	10,36	10,96	-0,60
28,6	11,08	11,34	-0,26

Exemplo (continuação)

- Vamos agora calcular as Somas de Quadrados

$$SQR = \sum_{i=1}^{n}(\hat{Y}_i - \overline{Y})^2 = 45,5924$$

$$SQE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = 18,2234$$

$$SQT = \sum_{i=1}^{n}(Y_i - \overline{Y})^2 = 63,8158$$

Exemplo (continuação)

- A Tabela ANOVA fica da seguinte maneira

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão	1	45,5924	$QMR = 45,5924/1$
Residual	23	18,2234	$QME = 18,2234/23 = 0,7923$
Total	24	63,8158	

Tabela: Tabela ANOVA

- O Coeficiente de Determinação é dado por

$$R^2 = \frac{SQR}{SQT} = \frac{45,5924}{63,8158} = 0,7144\,.$$

- Isso significa que

- O Coeficiente de Determinação é dado por

$$R^2 = \frac{SQR}{SQT} = \frac{45,5924}{63,8158} = 0,7144\,.$$

- Isso significa que
  - 71,44% da varibilidade dos dados pode ser explicada pelo modelo de regressão.

Intervalos de Confiança e Testes para β_0 e β_1

- Lembre-se que o modelo é definido como

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

onde $\epsilon_i \sim^{iid} N(0, \sigma^2)$.

Suposições do Modelo de Regressão

- Os erros são variáveis aleatórias com média zero e variância constante desconhecida

$$E(\epsilon_i) = 0 \quad Var(\epsilon_i) = \sigma^2 .$$

- Os erros são independentes entre si.
- Isso implica que $Cov(\epsilon_i, \epsilon_j) = 0$?

- Lembre-se que o modelo é definido como

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

onde $\epsilon_i \sim^{iid} N(0, \sigma^2)$.

Suposições do Modelo de Regressão

- Os erros são variáveis aleatórias com média zero e variância constante desconhecida

$$E(\epsilon_i) = 0 \quad Var(\epsilon_i) = \sigma^2 \; .$$

- Os erros são independentes entre si.
- Isso implica que $Cov(\epsilon_i, \epsilon_j) = 0$? Sim.
- O contrário é verdade?

- Lembre-se que o modelo é definido como

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

onde $\epsilon_i \sim^{iid} N(0, \sigma^2)$.

Suposições do Modelo de Regressão

- Os erros são variáveis aleatórias com média zero e variância constante desconhecida

$$E(\epsilon_i) = 0 \quad Var(\epsilon_i) = \sigma^2 \, .$$

- Os erros são independentes entre si.
- Isso implica que $Cov(\epsilon_i, \epsilon_j) = 0$? Sim.
- O contrário é verdade? Não.
- Os erros tem distribuição normal

$$\epsilon_i \sim N(0, \sigma^2) \, .$$

- Dessas supósições decorre que os Y_i são indendentes com a seguinte distribuição

- Dessas supósições decorre que os Y_i são indendentes com a seguinte distribuição

$$Y_i | x_i \sim N\left(\beta_0 + \beta_1 x_i, \sigma^2\right)$$

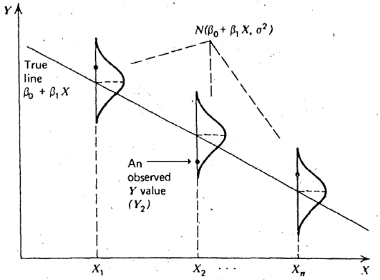
para $i = 1, \ldots, n$.

- Os Y_i são iid?

- Dessas supósições decorre que os Y_i são indendentes com a seguinte distribuição

$$Y_i|x_i \sim N\left(\beta_0 + \beta_1 x_i, \sigma^2\right)$$

para $i = 1, \ldots, n$.
- Os Y_i são iid? Não, pois não são identicamente distribuídos.



Distribuição F

- É uma distribuição contínua que só coloca massa de probabilidade em valores positivos.
- Essa função tem dois parâmetros, dois graus de liberdade

  $$F_{d_1, d_2}$$

  d_1 são os graus de liberdade do numerador e d_2 os graus de liberdade do denominador.
- Na tabela devemos olhar α e os graus de liberdade.
- Vamos olhar de um lado só pois rejeitamos para valores altos de D apenas.

Tabela F

Tabela 5. Limites unilaterais da distribuição F de Fisher-Snedecor ao nível de **5%** de probabilidade.

GL	V1																			
V2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	40	60	120	240
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.0	243.9	244.7	245.4	245.9	248.0	251.1	252.2	253.3	253.8
2	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396	19.405	19.412	19.419	19.424	19.429	19.446	19.471	19.479	19.487	19.492
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785	8.763	8.745	8.729	8.715	8.703	8.660	8.594	8.572	8.549	8.538
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.936	5.912	5.891	5.873	5.858	5.803	5.717	5.688	5.658	5.643
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.704	4.678	4.655	4.636	4.619	4.558	4.464	4.431	4.398	4.382
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027	4.000	3.976	3.956	3.938	3.874	3.774	3.740	3.705	3.687
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603	3.575	3.550	3.529	3.511	3.445	3.340	3.304	3.267	3.249
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347	3.313	3.284	3.259	3.237	3.218	3.150	3.043	3.005	2.967	2.947
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102	3.073	3.048	3.025	3.006	2.936	2.826	2.787	2.748	2.727
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943	2.913	2.887	2.865	2.845	2.774	2.661	2.621	2.580	2.559
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.818	2.788	2.761	2.739	2.719	2.646	2.531	2.490	2.448	2.426
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.717	2.687	2.660	2.637	2.617	2.544	2.426	2.384	2.341	2.319
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.635	2.604	2.577	2.554	2.533	2.459	2.339	2.297	2.252	2.230
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.565	2.534	2.507	2.484	2.463	2.388	2.266	2.223	2.178	2.155
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.507	2.475	2.448	2.424	2.403	2.328	2.204	2.160	2.114	2.090
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.456	2.425	2.397	2.373	2.352	2.276	2.151	2.106	2.059	2.035
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.413	2.381	2.353	2.329	2.308	2.230	2.104	2.058	2.011	1.986
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.374	2.342	2.314	2.290	2.269	2.191	2.063	2.017	1.968	1.943
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.340	2.308	2.280	2.256	2.234	2.155	2.026	1.980	1.930	1.905
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.310	2.278	2.250	2.225	2.203	2.124	1.994	1.946	1.896	1.870
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321	2.283	2.250	2.222	2.197	2.176	2.096	1.965	1.916	1.866	1.839
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.259	2.226	2.198	2.173	2.151	2.071	1.938	1.889	1.838	1.811
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.236	2.204	2.175	2.150	2.128	2.048	1.914	1.865	1.813	1.785

Teste F para significância da Regressão

- Queremos verificar se a variável explicativa é significativa para explicar a resposta.
- As hipóteses a serem testadas são:

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0 \, .$$

- Sob H_0 ($\beta_1 = 0$) temos que

$$\frac{SQR}{\sigma^2} \sim \chi^2_1 \quad \frac{SQE}{\sigma^2} \sim \chi^2_{n-2}$$

  e essas duas quantidades são independentes.
- Dessa maneira, temos que

$$F = \frac{SQR/1}{SQE/(n-2)} = \frac{QMR}{QME} \sim F_{(1,n-2)} \, .$$

- Rejeitamos H_0 so o modelo de regressão explica muito a variabilidade dos dados.
- Devemos rejeitar para valores altos ou baixos de F?

- Rejeitamos H_0 so o modelo de regressão explica muito a variabilidade dos dados.
- Devemos rejeitar para valores altos ou baixos de F? Altos.
- O teste é unilateral.
- Rejeitamos H_0 apenas para valores grandes de F.
- Olhamos na tabela α e não $\alpha/2$.
- A região crítica vai ser do tipo

$$F > F_c \, .$$

Exemplo

- Considere novamente o exemplo do consumo de gás e temperatura.
- Vimos que a Tabela ANOVA nesse caso é dada por

Fonte de Variação	Graus de Liberdade	Soma de Quadrados	Quadrado Médio
Regressão	1	45,5924	$QMR = 45,5924/1$
Residual	23	18,2234	$QME = 18,2234/23 = 0,7923$
Total	24	63,8158	

Tabela: Tabela ANOVA

Exemplo (continuação)

- As hipóteses a serem testadas são

Exemplo (continuação)

- As hipóteses a serem testadas são

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0 \, .$$

- A estatística F será dada por

$$F = \frac{QMR}{QME} =$$

Exemplo (continuação)

- As hipóteses a serem testadas são

$$H_0 : \beta_1 = 0 \quad \text{vs} \quad H_1 : \beta_1 \neq 0\,.$$

- A estatística F será dada por

$$F = \frac{QMR}{QME} = \frac{45,5924}{0,7923} = 57,54\,.$$

- Vamos usar $\alpha = 0,05$.

Tabela 5. Limites unilaterais da distribuição F de Fisher-Snedecor ao nível de 5% de probabilidade.

GL												V1								
V2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	40	60	120	240
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.0	243.9	244.7	245.4	245.9	248.0	251.1	252.2	253.3	253.8
2	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396	19.405	19.412	19.419	19.424	19.429	19.446	19.471	19.479	19.487	19.492
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785	8.763	8.745	8.729	8.715	8.703	8.660	8.594	8.572	8.549	8.538
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.936	5.912	5.891	5.873	5.858	5.803	5.717	5.688	5.658	5.643
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735	4.704	4.678	4.655	4.636	4.619	4.558	4.464	4.431	4.398	4.382
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027	4.000	3.976	3.956	3.938	3.874	3.774	3.740	3.705	3.687
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603	3.575	3.550	3.529	3.511	3.445	3.340	3.304	3.267	3.249
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347	3.313	3.284	3.259	3.237	3.218	3.150	3.043	3.005	2.967	2.947
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102	3.073	3.048	3.025	3.006	2.936	2.826	2.787	2.748	2.727
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943	2.913	2.887	2.865	2.845	2.774	2.661	2.621	2.580	2.559
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.818	2.788	2.761	2.739	2.719	2.646	2.531	2.490	2.448	2.426
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.717	2.687	2.660	2.637	2.617	2.544	2.426	2.384	2.341	2.319
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.635	2.604	2.577	2.554	2.533	2.459	2.339	2.297	2.252	2.230
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.565	2.534	2.507	2.484	2.463	2.388	2.266	2.223	2.178	2.155
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544	2.507	2.475	2.448	2.424	2.403	2.328	2.204	2.160	2.114	2.090
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.456	2.425	2.397	2.373	2.352	2.276	2.151	2.106	2.059	2.035
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.413	2.381	2.353	2.329	2.308	2.230	2.104	2.058	2.011	1.986
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.374	2.342	2.314	2.290	2.269	2.191	2.063	2.017	1.968	1.943
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.340	2.308	2.280	2.256	2.234	2.155	2.026	1.980	1.930	1.905
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.310	2.278	2.250	2.225	2.203	2.124	1.994	1.946	1.896	1.870
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321	2.283	2.250	2.222	2.197	2.176	2.096	1.965	1.916	1.866	1.839
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.259	2.226	2.198	2.173	2.151	2.071	1.938	1.889	1.838	1.811
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.236	2.204	2.175	2.150	2.128	2.048	1.914	1.865	1.813	1.785

- Notamos então que $F_{0,05;1,23} = 4,28$.
- A região crítica do teste é dada por

- Notamos então que $F_{0,05;1,23} = 4,28$.
- A região crítica do teste é dada por

$$F > 4,28$$

- Conclusão:

- Notamos então que $F_{0,05;1,23} = 4,28$.
- A região crítica do teste é dada por

$$F > 4,28$$

- Conclusão: Regeitamos H_0.
- Com 5% de significância há evidência de que a temperatura é significativa para explicar o consumo de gás.

- Veremos agora como encontrar intervalos para os parâmetros.
- Para isso precisamos encontrar variância dos estimadores.
- Vamos verificar primeiro se eles são não viesados.
- O que isso significa?

- Veremos agora como encontrar intervalos para os parâmetros.
- Para isso precisamos encontrar variância dos estimadores.
- Vamos verificar primeiro se eles são não viesados.
- O que isso significa?

$$E(\hat{\beta}_0) = \beta_0 \quad E[\hat{\beta}_1] = \beta_1$$

Intervalos de Confiança

- Vamos mostrar primeiro que $E(\hat{\beta}_0) = \beta_0$.

$E[\hat{\beta}_0] =$

- Vamos mostrar primeiro que $E(\hat{\beta}_0) = \beta_0$.

$$E[\hat{\beta}_0] = E[\overline{Y} - \hat{\beta}_1\overline{X}] = E[\overline{Y}] - E[\beta_1\overline{X}] =$$

- Vamos mostrar primeiro que $E(\hat{\beta}_0) = \beta_0$.

$$E[\hat{\beta}_0] = E[\overline{Y} - \hat{\beta}_1 \overline{X}] = E[\overline{Y}] - E[\beta_1 \overline{X}] = \frac{1}{n} E[\sum_i Y_i] - \overline{X} E[\hat{\beta}_1] =$$

- Vamos mostrar primeiro que $E(\hat{\beta}_0) = \beta_0$.

$$E[\hat{\beta}_0] = E[\overline{Y} - \hat{\beta}_1 \overline{X}] = E[\overline{Y}] - E[\beta_1 \overline{X}] = \frac{1}{n} E[\sum_i Y_i] - \overline{X} E[\hat{\beta}_1] =$$

$$\frac{1}{n} \sum_i E(Y_i) - \beta_1 \overline{X} =$$

- Vamos mostrar primeiro que $E(\hat{\beta}_0) = \beta_0$.

$$E[\hat{\beta}_0] = E[\overline{Y} - \hat{\beta}_1 \overline{X}] = E[\overline{Y}] - E[\beta_1 \overline{X}] = \frac{1}{n} E[\sum_i Y_i] - \overline{X} E[\hat{\beta}_1] =$$

$$\frac{1}{n} \sum_i E(Y_i) - \beta_1 \overline{X} = \frac{1}{n} \sum_i (\beta_0 + \beta_1 X_i) - \beta_1 \overline{X} =$$

- Vamos mostrar primeiro que $E(\hat{\beta}_0) = \beta_0$.

$$E[\hat{\beta}_0] = E[\overline{Y} - \hat{\beta}_1\overline{X}] = E[\overline{Y}] - E[\beta_1\overline{X}] = \frac{1}{n}E[\sum_i Y_i] - \overline{X}E[\hat{\beta}_1] =$$

$$\frac{1}{n}\sum_i E(Y_i) - \beta_1\overline{X} = \frac{1}{n}\sum_i(\beta_0 + \beta_1 X_i) - \beta_1\overline{X} =$$

$$\beta_0 + \beta_1\overline{X} - \beta_1\overline{X} = \beta_0 \ .$$

- Vejamos agora que $E(\hat{\beta}_1) = \beta_1$.
- Sabemos que

$$\sum_i (X_i - \overline{X})(Y_i - \overline{Y}) = \sum_i (X_i - \overline{X}) Y_i - \overline{Y} \sum_i (Y_i - \overline{Y}) = \sum_i (X_i - \overline{X}) Y_i$$

pois $\sum_i (Y_i - \overline{Y}) = 0$.

- Temos então que

$$E[\hat{\beta}_1] = E\left[\frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_i (X_i - \overline{X})^2}\right] = \frac{E[\sum_i (X_i - \overline{X}) Y_i]}{\sum_i (X_i - \overline{X})^2}$$

$$= \frac{[\sum_i (X_i - \overline{X}) E(Y_i)]}{\sum_i (X_i - \overline{X})^2} = \frac{[\sum_i (X_i - \overline{X})(\beta_0 + \beta_1 X_i)]}{\sum_i (X_i - \overline{X})^2}$$

$$E[\hat{\beta}_1] = \frac{\sum_i (X_i - \overline{X})\beta_0 + \beta_1 \sum_i (X_i - \overline{X})X_i}{\sum_i (X_i - \overline{X})^2}$$

mas

$$\sum_i (X_i - \overline{X}) = 0$$

$$\sum_i (X_i - \overline{X})X_i - \overline{X}\sum_i (X_i - \overline{X}) = \sum_i (X_i - \overline{X})^2$$

Logo

$$E[\hat{\beta}_1] = \frac{0\beta_0}{\sum_i (X_i - \overline{X})^2} + \beta_1 \frac{\sum_i (X_i - \overline{X})^2}{\sum_i (X_i - \overline{X})^2} = \beta_1$$

- Para encontrármos intervalos e fazer testes precisamos da variância dos estimadores.
- Vamos encontrar primeiro $Var(\hat{\beta}_1)$

$$Var(\hat{\beta}_1) = Var\left[\frac{\sum_i (X_i - \overline{X})Y_i}{\sum_i (X_i - \overline{X})^2}\right] = \frac{Var[\sum_i (X_i - \overline{X})Y_i]}{[\sum_i (X_i - \overline{X})^2]^2}$$

$$= \frac{\sum_i (X_i - \overline{X})^2 Var[Y_i]}{[\sum_i (X_i - \overline{X})^2]^2} = \frac{\sum_i (X_i - \overline{X})^2 \sigma^2}{[\sum_i (X_i - \overline{X})^2]^2}$$

$$= \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2}$$

- Temos que

$$Var(\hat{\beta}_0) = Var[\overline{Y} - \hat{\beta}_1 \overline{X}] = Var[\overline{Y}] + Var[-\hat{\beta}_1 \overline{X}] + 2Cov[\overline{Y}, -\hat{\beta}_1 \overline{X}]$$

$$= \frac{\sigma^2}{n} + \overline{X}^2 Var(\hat{\beta}_1) - 2\overline{X} Cov(\overline{Y}, \hat{\beta}_1)$$

- Mostraremos mais a frente que

$$Cov[\overline{Y}, -\hat{\beta}_1 \overline{X}] = 0 \, .$$

- Temos então que

$$Var(\hat{\beta}_0) = \frac{\sigma^2}{n} + \overline{X}^2 Var(\hat{\beta}_1)$$

- Vimos que

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2}$$

- Logo

$$Var(\hat{\beta}_0) = \frac{\sigma^2}{n} + \overline{X}^2 \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2}$$

$$= \sigma^2 \left[\frac{\sum_i (X_i - \overline{X})^2 + n\overline{X}^2}{\sum_i (X_i - \overline{X})^2} \right]$$

- Observe que

$$\sum_i (X_i - \overline{X})^2 + n\overline{X}^2 = \sum_i X_i^2 - 2\overline{X} \sum_i X_i + n\overline{X}^2 + n\overline{X}^2$$

$$= \sum_i X_i^2 - 2n\overline{X}^2 + n\overline{X}^2 + n\overline{X}^2 = \sum_i X_i^2$$

- Portanto

$$Var(\hat{\beta}_0) = \sigma^2 \frac{\sum_i X_i^2}{n \sum_i (X_i - \overline{X})^2}$$

- Vamos mostrar agora que $Cov[\overline{Y}, \hat{\beta}_1] = 0$.
- Temos que

$$Cov[\overline{Y}, \hat{\beta}_1] = Cov\left(\overline{Y}, \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_i (X_i - \overline{X})^2}\right)$$

$$\frac{Cov\left(\overline{Y}, \sum_i (X_i - \overline{X})(Y_i - \overline{Y})\right)}{\sum_i (X_i - \overline{X})^2} = \frac{\sum_i Cov\left(\overline{Y}, (X_i - \overline{X})(Y_i - \overline{Y})\right)}{\sum_i (X_i - \overline{X})^2}$$

$$= \frac{\sum_i (X_i - \overline{X}) Cov\left(\overline{Y}, (Y_i - \overline{Y})\right)}{\sum_i (X_i - \overline{X})^2}$$

- Observe que

$$Cov\left(\overline{Y}, (Y_i - \overline{Y}\right) = Cov\left(\overline{Y}, Y_i\right) - Cov\left(\overline{Y}, \overline{Y}\right)$$

$$= Cov\left(\frac{\sum_j Y_j}{n}, Y_i\right) - Var(\overline{Y}) = n^{-1}\sum_j Cov(Y_j, Y_i) - n^{-1}\sigma^2$$

$$n^{-1}\left[Cov(Y_i, Y_i) + \sum_{j:j\neq i} Cov(Y_j, Y_i) - \sigma^2\right]$$

$$n^{-1}\left[\sigma^2 + \sum_{j:j\neq i} Cov(Y_j, Y_i) - \sigma^2\right] = 0$$

- Pois Y's são não correlacionados

$$Cov(Y_j, Y_i) = 0 \text{ para } i \neq j\,.$$

Resumindo ...

- Mostramos que

$$E(\hat{\beta}_0) = \beta_0 \quad Var(\hat{\beta}_0) = \sigma^2 \frac{\sum_i X_i^2}{n \sum_i (X_i - \overline{X})^2}$$

$$E(\hat{\beta}_1) = \beta_1 \quad Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_i (X_i - \overline{X})^2}$$

Inferência sobre β_1

- Desvio padrão de $\hat{\beta}_1$ é dado por

$$DP(\hat{\beta}_1) = \sqrt{Var(\beta_1)} = \left[\frac{\sigma^2}{\sum_i (X_i - \overline{X})^2}\right]^{1/2}$$

que é estimado por

$$\hat{DP}(\hat{\beta}_1) = \left[\frac{S^2}{\sum_i (X_i - \overline{X})^2}.\right]^{1/2}$$

- O intervalo de confiança é dado por

$$\hat{\beta}_1 \pm \hat{DP}(\hat{\beta}_1) t_{n-2;\alpha/2} \quad \hat{\beta}_1 \pm \left[\frac{S^2}{\sum_i (X_i - \overline{X})^2}.\right]^{1/2} t_{n-2;\alpha/2}$$

- Se quisermos testar as hipóteses

$$H_0 : \beta_1 = b_1 \quad H_1 : \beta_1 \neq b_1$$

- A estatística de teste é dada por

$$t = \frac{\hat{\beta}_1 - b_1}{\hat{DP}(\hat{\beta}_1)} = \frac{\hat{\beta}_1 - b_1}{\left[\frac{S^2}{\sum_i (X_i - \overline{X})^2}\right]^{1/2}} \sim t_{n-2}$$

Inferência sobre β_0

- Desvio padrão de $\hat{\beta}_0$ é dado por

$$DP(\hat{\beta}_0) = \sqrt{Var(\beta_0)} = \left[\sigma^2 \frac{\sum_i X_i^2}{n\sum_i (X_i - \overline{X})^2}\right]^{1/2}$$

que é estimado por

$$\hat{DP}(\hat{\beta}_0) = \left[S^2 \frac{\sum_i X_i^2}{n\sum_i (X_i - \overline{X})^2}\right]^{1/2}$$

- O intervalo de confiança é dado por

$$\hat{\beta}_0 \pm \hat{DP}(\hat{\beta}_0) t_{n-2;\alpha/2} \quad \hat{\beta}_0 \pm \left[S^2 \frac{\sum_i X_i^2}{n\sum_i (X_i - \overline{X})^2}\right]^{1/2} t_{n-2;\alpha/2}$$

- Se quisermos testar as hipóteses

$$H_0 : \beta_0 = b_0 \quad H_0 : \beta_0 \neq b_0$$

- A estatística de teste é dada por

$$t = \frac{\hat{\beta}_0 - b_0}{\hat{DP}(\hat{\beta}_0)} = \frac{\hat{\beta}_0 - b_0}{\left[S^2 \frac{\sum_i X_i^2}{n \sum_i (X_i - \overline{X})^2}\right]^{1/2}} \sim t_{n-2}$$

Exemplo:

- Vamos retormar o exemplo do consumo de gás e temperatura.
- Para esses dados tínhamos que $\hat{\beta}_1 = -0,0798$.
- Vimos ainda que

$$n = 25 \quad \sum_i X_i^2 = 76323.42 \quad \overline{X} = 52.60$$

$$\sum_i (X_i - \overline{X})^2 =$$

Exemplo:

- Vamos retormar o exemplo do consumo de gás e temperatura.
- Para esses dados tínhamos que $\hat{\beta}_1 = -0,0798$.
- Vimos ainda que

$$n = 25 \quad \sum_i X_i^2 = 76323.42 \quad \overline{X} = 52.60$$

$$\sum_i (X_i - \overline{X})^2 = \sum_i X_i^2 - n\overline{X}^2 =$$

Exemplo:

- Vamos retormar o exemplo do consumo de gás e temperatura.
- Para esses dados tínhamos que $\hat{\beta}_1 = -0,0798$.
- Vimos ainda que

$$n = 25 \quad \sum_i X_i^2 = 76323.42 \quad \overline{X} = 52.60$$

$$\sum_i (X_i - \overline{X})^2 = \sum_i X_i^2 - n\overline{X}^2 = 76323.42 - 25(52.60)^2 = 7154.42$$

- Além disso

$$S^2 = 0.7923\ .$$

- Portanto

$$\hat{DP}(\hat{\beta}_1) = \left[\frac{S^2}{\sum_i (X_i - \overline{X})^2}\right]^{1/2} =$$

Exemplo:

- Vamos retormar o exemplo do consumo de gás e temperatura.
- Para esses dados tínhamos que $\hat{\beta}_1 = -0,0798$.
- Vimos ainda que

$$n = 25 \quad \sum_i X_i^2 = 76323.42 \quad \overline{X} = 52.60$$

$$\sum_i (X_i - \overline{X})^2 = \sum_i X_i^2 - n\overline{X}^2 = 76323.42 - 25(52.60)^2 = 7154.42$$

- Além disso

$$S^2 = 0.7923\,.$$

- Portanto

$$\hat{DP}(\hat{\beta}_1) = \left[\frac{S^2}{\sum_i (X_i - \overline{X})^2}\right]^{1/2} = \left[\frac{0.7923}{7154.42}\right]^{1/2} = 0.0105$$

Exemplo: (continuação)

- Queremos calcular intervalo de 95% para β_1.
- Precisamos olhar na tabela

Exemplo: (continuação)

- Queremos calcular intervalo de 95% para β_1.
- Precisamos olhar na tabela $t_{23;0,025} = 2.069$.
- Portanto o intervalo de confiança para β_1 é dado por

$$-0,0798 \pm (2.069)(0.0105)$$

$$IC_{95\%}(\beta_1) = [-0.1015; -0.0581].$$

- Interpretação do intervalo:

Exemplo: (continuação)

- Queremos calcular intervalo de 95% para β_1.
- Precisamos olhar na tabela $t_{23;0,025} = 2.069$.
- Portanto o intervalo de confiança para β_1 é dado por

$$-0,0798 \pm (2.069)(0.0105)$$

$$IC_{95\%}(\beta_1) = [-0.1015; -0.0581].$$

- Interpretação do intervalo: Com 95% de confiança podemos dizer que o verdadeiro valor de β_1 está entre -0,1015 e -0,0581.

Exemplo: (continuação)

- Queremos testar as hipóteses:

$$H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0 \, .$$

- Considere 5% de significância.
- Com base no intervalo de confiança, qual a conclusão do teste?

Exemplo: (continuação)

- Queremos testar as hipóteses:

$$H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0 \,.$$

- Considere 5% de significância.
- Com base no intervalo de confiança, qual a conclusão do teste?
- Como zero não pertence ao intervalo, rejeitamos H_0.
- Conclusão:

Exemplo: (continuação)

- Queremos testar as hipóteses:

$$H_0 : \beta_1 = 0 \quad H_1 : \beta_1 \neq 0 \,.$$

- Considere 5% de significância.
- Com base no intervalo de confiança, qual a conclusão do teste?
- Como zero não pertence ao intervalo, rejeitamos H_0.
- Conclusão: com 5% de significância podemos dizer que a variável temperatura está linearmente correlacionada com a variável consumo de gás.

Exemplo: (continuação)

- Vamos agora encontrar o intervalo de confiança para β_0.
- Vimos que $\hat{\beta}_0 = 13.623$.
- Vimos ainda que

$$n = 25 \quad \sum_i X_i^2 = 76323.42 \quad \sum_i (X_i - \overline{X})^2 = 7154.42$$

$$S^2 = 0.7923$$

- Temos que

$$\hat{DP}(\hat{\beta}_0) = \left[S^2 \frac{\sum_i X_i^2}{n \sum_i (X_i - \overline{X})^2} \right]^{1/2} = \left[0.7923 \frac{76323.42}{25(7154.42)} \right]^{1/2}$$

$$= 0.5814$$

Exemplo: (continuação)

- Queremos calcular intervalo de 95% para β_0.
- Precisamos olhar na tabela

Exemplo: (continuação)

- Queremos calcular intervalo de 95% para β_0.
- Precisamos olhar na tabela $t_{23;0,025} = 2.069$.
- Portanto o intervalo de confiança para β_0 é dado por

$$13.623 \pm (2.069)(0.5814)$$

$$IC_{95\%}(\beta_0) = [12.4201; 14.8259] .$$

- Interpretação do intervalo:

Exemplo: (continuação)

- Queremos calcular intervalo de 95% para β_0.
- Precisamos olhar na tabela $t_{23;0,025} = 2.069$.
- Portanto o intervalo de confiança para β_0 é dado por

$$13.623 \pm (2.069)(0.5814)$$

$$IC_{95\%}(\beta_0) = [12.4201; 14.8259] .$$

- Interpretação do intervalo: Com 95% de confiança podemos dizer que o verdadeiro valor de β_0 está entre 12.4201 e 14.8259.

- Vimos até agora como fazer inferência sobre os parâmetros.
- Podemos estar interessados em fazer inferência para a resposta média.
- Ou seja, queremos inferir qual o valor esperado de Y para determinado valor de X_0

$$\mu_0 = E(Y|X = X_0) \,.$$

- A estimativa pontual é dada por

$$\hat{\mu}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0 \,.$$

- Desvio padrão é dado por

$$DP(\mu_0) = \left\{ \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_i (X_i - \overline{x})^2} \right] \right\}^2$$

que é estimado por

$$\hat{DP}(\mu_0) = \left\{ S^2 \left[\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_i (X_i - \overline{x})^2} \right] \right\}^2$$

- O intervalo de confiança é calculado como

$$\hat{\mu_0} \pm \hat{DP}(\mu_0) t_{n-2;\alpha/2}$$

Inferência sobre observações individuais

- Vimos anteriormente como encontrar o intervalo de confiança para a média (em torno da reta).
- Vamos ver agora como encontrar o intervalo de confiança para a resposta individual.
- O valor individual Y varia em torno da média $E(Y|X)$.
- A variância será maior do que no caso anterior.
- A estimativa pontual é a mesma.
- O intervalo de confiança fica com amplitude maior.

- A estimativa pontual é dada por

$$\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0 \, .$$

- Desvio padrão é dado por

$$DP(Y_0) = \left\{ \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_i (X_i - \overline{x})^2} \right] \right\}^{1/2}$$

que é estimado por

$$\hat{DP}(Y_0) = \left\{ S^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_i (X_i - \overline{x})^2} \right] \right\}^{1/2}$$

- O intervalo de confiança é calculado como

$$\hat{Y}_0 \pm \hat{DP}(Y_0) t_{n-2;\alpha/2}$$

Exemplo:

- Considere o exemplo do consumo de gás e temperatura.
- Suponha que queremos saber qual o consumo de gás para uma temperatura igual a 30.
- Temos que

$$(X_0 - \overline{X})^2 = (30 - 52.6)^2 = 510.76$$

$$\sum_i (X_i - \overline{x})^2 = 7154.42 \quad S^2 = 0.7923$$

- O valor esperado do consumo quando $X = 30$ é dado por

Exemplo:

- Considere o exemplo do consumo de gás e temperatura.
- Suponha que queremos saber qual o consumo de gás para uma temperatura igual a 30.
- Temos que

$$(X_0 - \overline{X})^2 = (30 - 52.6)^2 = 510.76$$

$$\sum_i (X_i - \overline{x})^2 = 7154.42 \quad S^2 = 0.7923$$

- O valor esperado do consumo quando $X = 30$ é dado por

$$Y = 13.623005 - 0.079829(30) \Rightarrow Y = 11.228\,.$$

- O intervalo de confiança para a resposta média é dado por

$$\hat{Y}_0 \pm \hat{DP}(Y_0) t_{n-2;\alpha/2}\,.$$

Exemplo: (continuação)

- Vamos agora calcular $\hat{DP}(\mu_0)$.
- Temos que

$$\hat{DP}(\mu_0) = \left\{ S^2 \left[\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_i (X_i - \overline{x})^2} \right] \right\}^{1/2}$$

$$= \left\{ 0.7923 \left[\frac{1}{25} + \frac{510.76}{7154.42} \right] \right\}^{1/2} = 0.2644$$

- Além disso $t_{13;0,025} = 2.069$.
- O intervalo fica

$$11.228 \pm (2.069)(0.2644)$$

Exemplo: (continuação)

- Portanto
$$IC_{95\%}(\hat{\mu}_0) = [10.6809; 11.7750] \,.$$
- Interpretação:

Exemplo: **(continuação)**

- Portanto

$$IC_{95\%}(\hat{\mu}_0) = [10.6809; 11.7750] \,.$$

- Interpretação: Com 95% de confiança pode-se dizer que a o consumo de gás esperado para uma temperatura igual a 30 está entre 10.6809 e 11.7750.

Exemplo: (continuação)

- Suponha agora que estamos interessados na resposta individual e não na resposta média.
- A estimativa pontual é a mesma obtida anteriormente

Exemplo: (continuação)

- Suponha agora que estamos interessados na resposta individual e não na resposta média.
- A estimativa pontual é a mesma obtida anteriormente

$$Y = 13.623005 - 0.079829(30) \Rightarrow Y = 11.228\,.$$

- O desvio padrão nesse caso é estimado por

$$\hat{DP}(Y_0) = \left\{ S^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum_i (X_i - \overline{x})^2} \right] \right\}^{1/2}$$

$$\left\{ 0.7923 \left[1 + \frac{1}{25} + \frac{510.76}{7154.42} \right] \right\}^{1/2} = 0.8352$$

- Além disso $t_{13;0,025} = 2.069$.
- O intervalo fica

$$11.228 \pm (2.069)(0.8352)$$

Exemplo: (continuação)

- Portanto
$$IC_{95\%}(\hat{Y}_0) = [9.4999; 12.9560] \,.$$
- Interpretação:

Exemplo: (continuação)

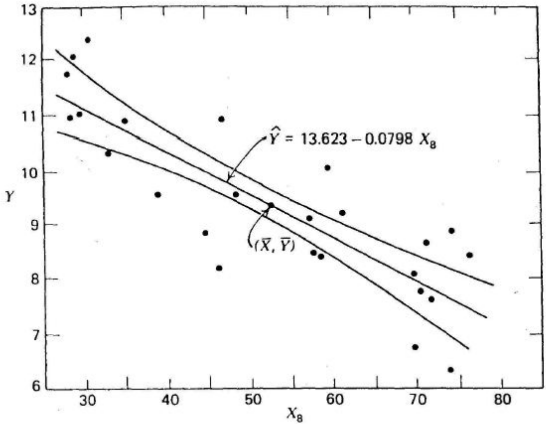
- Portanto

$$IC_{95\%}(\hat{Y}_0) = [9.4999; 12.9560] \,.$$

- Interpretação: Com 95% de confiança pode-se dizer que a o consumo de gás verdadeiro para uma temperatura igual a 30 está entre 10.6809 e 11.7750.

Exemplo: (continuação)

- O gráfico a seguir mostra reta ajusta e o intervalo de confiança de 95% para a resposta média.



Exemplo: (continuação)

- O gráfico a seguir mostra reta ajusta e o intervalo de confiança de 95% **para a resposta individual.**