Regressão Linear Múltipla

Frases

"Por serem mais precisos que as palavras, os números são particularmente adequados para transmitir conclusões científicas"

Pagano e Gauvre, 2004

Roteiro

Especificação do Modelo

Exemplo 5 – Rede de Lanchonetes

- Modelo para explicar receita total em cadeia de lanchonete
- Semanalmente decide-se:
  - √ quanto gastar com propaganda e
  - √ quais promoções (preços mais baixos) serão implementadas
- Dados: *lanchonete*
- *Fonte: Hill*

Exemplo 5 – Rede de Lanchonetes

- Despesas de propaganda:
  - √ Aumento despesas gera aumento na receita?
  - √ Aumento de receita é suficiente para justificar despesas?
  - √ Redução de preços leva a aumento ou diminuição de receita total?

Exemplo 5 – Rede de Lanchonetes

- Promoções:
  - √ Redução de preços leva a aumento ou diminuição de receita total?
  - √ Situações a ser verificadas:

Redução de preços	➔	**Pequeno aumento das vendas**	**Demanda <u>inelástica</u> em relação ao preço**
Redução de preços	➔	**Grande aumento das vendas**	**Demanda <u>elástica</u> em relação ao preço**

Modelo Econômico

$$E(RT) = b_0 + b_1 p + b_2 a$$

- *RT*: receita total ($ mil)
- *p*: preço médio de todos os produtos
- *a*: nível de despesas com propaganda ($ mil)
- Comportamento esperado (ou médio) de muitos pontos de venda distintos
- Componente sistemático

Modelo Econométrico

$$RT_i = b_0 + b_1 p_i + b_2 a_i + e_i$$

- Incorpora o termo de erro aleatório
- *e*: todos os fatores que fazem a receita total da semana diferir de seu valor esperado

  (condições meteorológicas, comportamento dos concorrentes, etc.)

$$RT_i = \underbrace{E(RT)}_{\text{componente sistemático}} + \underbrace{e_i}_{\text{componente aleatório}}$$

Interpretação dos Parâmetros

- $ß_1$: Variação em RT (\$ 1000) quando *p* sofre um aumento de 1 unidade e *a* é mantido constante:

$$b_1 = \frac{\Delta RT}{\Delta p}, a\ cte. \qquad\qquad b_1 = \frac{\partial RT}{\partial p}$$

√ $ß_1 > 0$: demanda <u>inelástica</u> em relação ao preço aumento de preço ocasiona aumento receita

√ $ß_1 < 0$: demanda <u>elástica</u> em relação ao preço aumento de preço ocasiona decréscimo receita

Interpretação dos Parâmetros (2)

- $ß_2$: Variação em RT (\$ 1000) quando *a* é aumentado de \$1 mil e *p* é mantido constante. Espera-se que seja positivo:

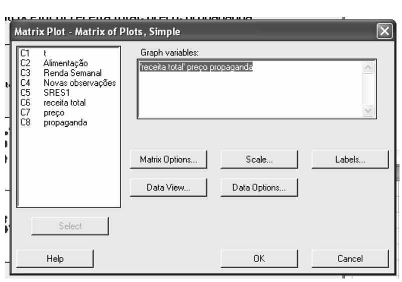
$$b_2 = \frac{\Delta RT}{\Delta a}, p\ cte. \qquad\qquad b_2 = \frac{\partial RT}{\partial a}$$

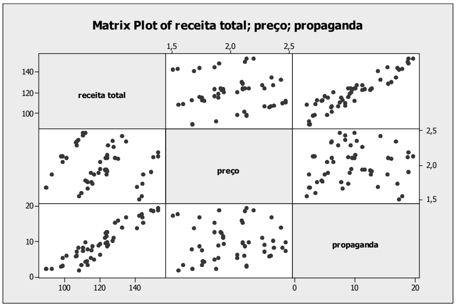
√ $ß_2 < 1$: aumento de \$1.000 na propaganda ocasionará aumento inferior a \$1.000 em RT

√ $ß_2 > 1$: aumento na propaganda acarretará aumento maior na receita

Análise Exploratória

Graph > Matrix Plot > Matrix of plots: Simple ➔





Correlations: receita total; preço; propaganda

```
                  receita tota         preço
preço                   -0,014
                         0,921

propaganda               0,925         0,101
                         0,000         0,474
```

Comentários

- Embora a correlação entre receita Total e preço seja não significante, o diagrama de dispersão mostra padrões nos pontos;
- A correlação entre Receita Total e propaganda é bastante forte e aparenta haver uma relação linear entre elas
- Preço e propaganda são fracamente correlacionados

Modelo Geral

Regressão Múltipla

- Predizer valores de uma variável dependente (Y) em função de variáveis independentes ($X_1, X_2, ..., X_k$).
- Conhecer o quanto as variações de X_j ($j = 1,...,k$) podem afetar Y.

Aplicações

X_1 = renda X_2 = taxa de juros X_3 = poupança

$\Rightarrow$ Y = consumo

X_1 = área construída X_2 = custo do m^2 X_3 = localização

$\Rightarrow$ Y = preço imóvel

Modelo de Regressão Múltipla

- $E\{Y\} = f(X_1, X_2, ..., X_k)$
- Linear: $E\{Y\} = ß_0 + ß_1X_1 + ß_2X_2 + ... + ß_kX_k$

  Admite-se que $X_1, ..., X_k$ são variáveis não estocásticas e Y é uma variável aleatória.

Modelo de Regressão Múltipla (2)

- $E\{Y\} = ß_0 + ß_1X_1 + ß_2X_2 + ... + ß_kX_k$
- O coeficiente $ß_j$ representa a variação esperada de Y para cada unidade de variação em X_j ($j = 1, 2, ..., k$), considerando as outras variáveis independentes fixas.
- O primeiro objetivo é estimar os coeficientes: $ß_0, ß_1, ß_2, ... , ß_k$

Modelo de Regressão

AMOSTRA:

	variáveis				
obs.	**Y**	$\mathbf{X_1}$	$\mathbf{X_2}$	**...**	$\mathbf{X_k}$
1	y_1	x_{11}	x_{12}	...	x_{1k}
2	y_2	x_{21}	x_{22}	...	x_{2k}
...	...	...	...	...	...
n	y_k	x_{n1}	x_{n2}	...	x_{nk}

- $E\{Y_i\} = ß_0 + ß_1X_{i1} + ß_2X_{i2} + ... + ß_kX_{ik}$
- $Y_i = ß_0 + ß_1X_{i1} + ß_2X_{i2} + ... + ß_kX_{ik} + e_i$

termo aleatório

Modelo de Regressão Múltiplo

$$Y_i = b_0 + b_1X_{i1} + b_2X_{i2} + ... + b_kX_{ik} + e_i$$

- $ß_j$: efeito da variação da variável x_{ij} sobre o valor esperado de Y_i, mantidas as outras variáveis constantes
  √ $i=1, ..., n$ e $j=1, ..., k$
- $ß_0$: intercepto

Supostos do Modelo

- Os erros (e_i) são independentes e variam aleatoriamente segundo uma distribuição (normal) com média zero e variância constante (homocedástico).

  A média de todas a variáveis omitidas do modelo é zero (o modelo em média é correto).

  Para nenhuma observação a incerteza do modelo será maior ou menor (homocedástico)

Supostos do Modelo

- *cov*(e_i, e_j): o tamanho do erro de uma observação não tem qualquer influência sobre o tamanho provável do erro de outra observação;
- Nenhuma variável explanatória é uma função linear exata de qualquer outra (multicolinearidade exata)

  Nenhuma variável é redundante

Estimativa de Mínimos Quadrados Ordinários para 2 Variáveis

- Minimização da soma dos quadrados das diferenças entre os valores observados e seu valor esperado

$$S(b_0, b_1, b_2) = \sum_{i=1}^{n} \left(y_i - E(Y_i)\right)^2 = \sum_{i=1}^{n} \left(y_i - b_0 - b_1 x_{i1} - b_2 x_{i2}\right)^2$$

- Os estimadores de mínimos quadrados são obtidos através das derivadas de S em relação a $ß_0$, $ß_1$ e $ß_2$.
  - √ Os cálculos não são triviais

Modelos Lineares – Regressão Simples

i	x_i	y_i	$y_i = b_0 + b_1 x_i + e_i$
1	20	98	$98 = \beta_0 + \beta_1.20 + e_1$
2	25	110	$110 = \beta_0 + \beta_1.25 + e_2$
3	30	112	$112 = \beta_0 + \beta_1.30 + e_3$
4	35	115	$115 = \beta_0 + \beta_1.35 + e_4$
5	40	122	$122 = \beta_0 + \beta_1.40 + e_5$

Notação Matricial – Regressão Simples

$$\begin{pmatrix} 98 \\ 110 \\ 112 \\ 115 \\ 122 \end{pmatrix} = \begin{pmatrix} 1 & 20 \\ 1 & 25 \\ 1 & 30 \\ 1 & 35 \\ 1 & 40 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{pmatrix}$$

$$\boldsymbol{Y} = \boldsymbol{X}\, b + e$$

Modelos Lineares – Regressão Múltipla

i	x_{1i}	x_{2i}	y_i
1	20	70	98
2	25	68	110
3	30	83	112
4	35	77	115
5	40	65	122

Notação Matricial – Regressão Múltipla

$$\begin{pmatrix} 98 \\ 110 \\ 112 \\ 115 \\ 122 \end{pmatrix} = \begin{pmatrix} 1 & 20 & 70 \\ 1 & 25 & 68 \\ 1 & 30 & 83 \\ 1 & 35 & 77 \\ 1 & 40 & 65 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{pmatrix}$$

$$\boldsymbol{Y} = \boldsymbol{X}\, b + e$$

Estimador de Mínimos Quadrados

$$Y = Xb + e$$

Estimador de mínimos quadrados de β, isto é, o vetor que minimiza a função $L(\beta) = \varepsilon'\varepsilon = (Y - X\beta)'(Y - X\beta)$:

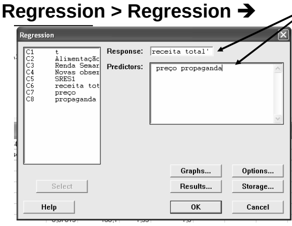
$$\Rightarrow \hat{b} = (X'X)^{-1}(X'Y)$$

$$\hat{b} = (\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_k)$$

Cálculo das *EMQO*

- Na prática, usam-se pacotes computacionais para calcular as estimativas de mínimos quadrados (*EMQO*)
- No Minitab, o caminho é o mesmo:

Stat > Regression > Regression ➔



Lanchonetes – Regressão Ajustada

Regression Analysis: receita total versus preço; propaganda

```
The regression equation is
receita total = 105 - 6,64 preço + 2,98 propaganda

Predictor      Coef  SE Coef      T      P
Constant    104,786    6,483  16,16  0,000
preço        -6,642    3,191  -2,08  0,043
propaganda   2,9843   0,1669  17,88  0,000

S = 6,06961   R-Sq = 86,7%   R-Sq(adj) = 86,2%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       2  11776,2  5888,1  159,83  0,000
Residual Error  49   1805,2    36,8
Total           51  13581,4
```

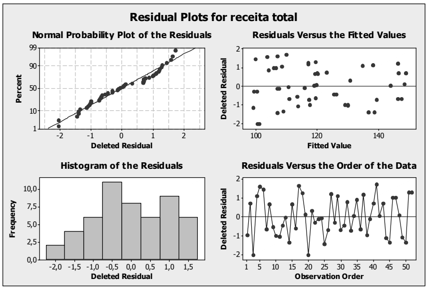
Interpretação

- A demanda aparenta ser elástica em relação ao preço
  - √ Aumento de $1 no preço leva a queda de $6.642 na receita
- O coeficiente de propaganda é positivo:
  - √ Aumento de $1.000 na propaganda resulta aumento de $2.984 na receita

Interpretação (2)

- Intercepto: Se tanto o preço como as despesas com propaganda fossem zero, a receita seria $104.790
  - √ Interpretação enganosa

Lanchonetes – Gráfico de Resíduos



Equação para predição

- Prever a receita total para um preço de \$2 e despesa com propaganda de \$1.000:

$$\hat{R}_{Ti} = 104{,}785 - 6{,}642(2) + 2{,}9843(10) = \$121{,}34$$

- Os modelos de regressão descrevem as variáveis econômicas para valores análogos aos dos dados amostrais. Extrapolar resultados para valores extremos não é recomendado

Estimação da Variância

- A variância é estimada por:

$$\hat{s}^2 = \frac{\sum_i \hat{e}_i^2}{n-(k+1)}$$

onde,

k+1: quantidade de parâmetros a estimar

n: tamanho na amostra

- No exemplo da receita total das lanchonetes:

$$\hat{s}^2 = \frac{1.805}{52-3} = 36{,}8$$

- No exemplo da receita total das lanchonetes:

$$\hat{s}^2 = \frac{1.805}{52-3} = 36,8$$

Propriedades Amostrais dos *EMQO*

Teorema de Gauss – Markov

- Atendidas as hipóteses do modelo de regressão, os EMQO são os melhores estimadores lineares não viciados dos parâmetros (*BLUE – best linear unbiased estimators).*
  - √ Não é necessária a hipótese de normalidade

Variâncias e Covariâncias dos *EMQO*

- Para o caso de duas variáveis explicativas, temos que:

$$Var(\hat{b}_1) = \frac{s^2}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2(1 - r_{12}^2)}$$

r_{12}: coeficiente de correlação entre X_1 e X_2

- Há fórmulas análogas para as outras variáveis

Conseqüências

- Quanto maior a variância do erro (s^2), maior a variância dos *EMQO*;
  - √ A incerteza global na especificação do modelo é transmitida a seus estimadores
- Quanto maior o tamanho *n* da amostra, menor é a variância dos *EMQO*
  - √ Mais observações resultam em estimação mais precisa

Conseqüências (2)

- Quanto maior a variação da variável explicativa em torno de sua média, menor a variância do *EMQO*
  - √ Se a variação em x_1 é pequena, é difícil medir seu efeito
- Quanto maior a correlação entre X_1 e X_2, maior a variância do EMQO.
  - √ Se a variação em uma explicativa está relacionada com a variação de outras variáveis, é difícil separar seus efeitos

Matriz de Variâncias e Covariâncias

- Para o caso de duas variáveis explicativas:

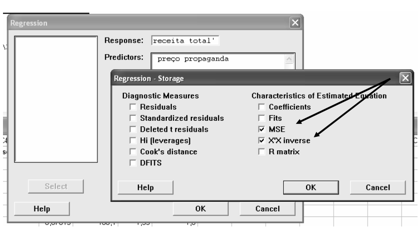
$$\Sigma = \begin{bmatrix} Var(\hat{b}_0) & Cov(\hat{b}_0,\hat{b}_1) & Cov(\hat{b}_0,\hat{b}_2) \\ Cov(\hat{b}_0,\hat{b}_1) & Var(\hat{b}_1) & Cov(\hat{b}_1,\hat{b}_2) \\ Cov(\hat{b}_0,\hat{b}_2) & Cov(\hat{b}_1,\hat{b}_2) & Var(\hat{b}_2) \end{bmatrix}$$

- A maioria dos pacotes apresenta esta matriz como um de seus resultados de saída

Cálculo no Minitab

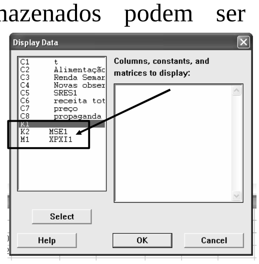
- O Minitab, não apresenta a matriz de variância e covariância em sua saída
- Para seu cálculo:

Stat > Regression > Regression ➔



- Serão armazenados uma matriz de resultados *(X'X)*$^{-1}$ e uma constante (*MSE*)
- No Minitab, matriz é identificada com a letra *M* e constante com a letra *K*.
- Os resultados armazenados podem ser verificados em:

Data > Display Data ➔



Variâncias e Covariâncias – Lanchonetes

- Para se obter a matriz de variâncias e covariâncias, basta se multiplicar o *MSE* (constante) por $(X'X)^{-1}$ (matriz)

```
MTB > Multiply K2 M1 M2
MTB > Print M2
```

Data Display

```
 Matrix M2

 42,0256  -19,8631  -0,161109
-19,8631   10,1837  -0,054021
 -0,1611   -0,0540   0,027868
```

Variâncias dos *EMQO* – Lanchonetes

	42,026
	10,184
	0,02787

Estimativas Intervalares e Testes de Hipóteses

Normalidade

- Se acrescentarmos a suposição de normalidade dos erros às hipóteses do modelo, então os *EMQO* são os melhores estimadores não viciados dos parâmetros da regressão;
- Além disso, podemos construir intervalos de confiança e realizar testes de hipóteses

Inferência para $ß_j$

- Distribuição do estimador $ß_j$, $j = 0, 1, \dots , k$:

$$\hat{b}_j \sim N\left(b_j, ep\left(\hat{b}_j\right)\right)$$

- Intervalo de confiança

$$\hat{b}_j - t_{a/2;(n-(k+1))} ep\left(\hat{b}_j\right) \leq b_j \leq \hat{b}_j + t_{a/2;(n-(k+1))} ep\left(\hat{b}_j\right)$$

Exemplo – Lanchonetes

Regression Analysis: receita total versus preço; propaganda

```
The regression equation is
receita total = 105 - 6,64 preço + 2,98 propaganda

Predictor      Coef  SE Coef      T      P
Constant    104,786    6,483  16,16  0,000
preço        -6,642    3,191  -2,08  0,043
propaganda   2,9843   0,1669  17,88  0,000

S = 6,06961   R-Sq = 86,7%   R-Sq(adj) = 86,2%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       2  11776,2  5888,1  159,83  0,000
Residual Error  49   1805,2    36,8
Total           51  13581,4
```

- Tamanho da amostra: *52*

Intervalos com *95%* de Confiança

- $t_{a/2,(52-(2+1))} = 2{,}010$
- Do Parâmetro $ß_0$:

$$104{,}79 - (2{,}010)(6{,}483) \le b_0 \le 104{,}79 + (2{,}010)(6{,}483)$$
$$91{,}759 \le b_0 \le 117{,}821$$

Intervalos com *95%* de Confiança

- Do Parâmetro $ß_1$:

$$-6{,}642 - (2{,}010)(3{,}191) \le b_1 \le -6{,}642 + (2{,}010)(3{,}191)$$
$$-13{,}056 \le b_1 \le -0{,}228$$

- Intervalo <u>pouco informativo</u>
- Redução de *$1* no preço acarreta um aumento da receita de *$ 228* a *$ 13.056*.

Intervalos com *95%* de Confiança

- Do Parâmetro $ß_2$:

$$2{,}984 - (2{,}010)(0{,}1669) \le b_2 \le 2{,}984 + (2{,}010)(0{,}1669)$$
$$2{,}649 \le b_1 \le 3{,}319$$

- Aumento de *$1.000* nas despesas com propaganda levam a um aumento da receita de *$ 2.649* a *$ 3.319*.

Teste de Significância para Coeficiente Único

- Teste de significância do parâmetro $ß_j$, $j = 0, 1, \ldots, k$:

H_0: $ß_j = 0$ vs *regressão não é significativa*
H_1: $ß_j$? 0 *regressão significativa*

- Estatística de teste:

$$T = \frac{\hat{b}_j}{ep\left(\hat{b}_j\right)}$$

- Distribuição da estatística de teste: $T \sim t_{(n-(k+1))}$

Inferência para Parâmetros – Caso Geral

- Teste de Hipóteses de parâmetro:

H_0: $ß_j = ß^0$ vs
H_1: $ß_j$? $ß^0$, $i = 0, \ldots, k$

- Estatística de teste:

$$T = \frac{\hat{b}_j - b_j^0}{ep\left(\hat{b}_j\right)}, i = 0, \cdots, k$$

- Distribuição da estatística de teste: $T \sim t_{(n-(k+1))}$

Regression Analysis: receita total versus preço; propaganda

Estatística t

```
The regression equation is
receita total = 105 - 6,64 preço + 2,98 propaganda

Predictor       Coef  SE Coef      T      P
Constant     104,786    6,483  16,16  0,000
preço         -6,642    3,191  -2,08  0,043
propaganda    2,9843   0,1669  17,88  0,000

S = 6,06961   R-Sq = 86,7%   R-Sq(adj) = 86,2%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       2  11776,2  5888,1  159,83  0,000
Residual Error  49   1805,2    36,8
Total           51  13581,4
```

$$T_{b_0} = \frac{104{,}786}{6{,}483} = 16{,}163 \quad T_{b_1} = \frac{-6{,}642}{3{,}191} = -2{,}081 \quad T_{b_2} = \frac{2{,}9843}{0{,}1669} = 17{,}881$$

Rejeitamos H_0 se $|T| < 2{,}010$

Teste de Elasticidade da Demanda

- Quer-se saber se:
  - √ $ß_1$ = 0: demanda <u>inelástica</u> em relação ao preço
    redução no preço ocasiona redução da receita
  - √ $ß_1$ < 0: demanda <u>elástica</u> em relação ao preço
    redução no preço ocasiona aumento da receita

Teste de Elasticidade da Demanda (2)

- Teste de Hipóteses de parâmetro:

H_0: $ß_1 = 0$ *vs* *elasticidade unitária ou inelástica*
H_1: $ß_1 < 0$ *demanda elástica*

- Estatística de teste: $T_{b_1} = \frac{-6{,}642}{3{,}191} = -2{,}081$

- Limite da região crítica: $t_{0,05;\ (52-(2+1))} = -1{,}68$

$T = -2{,}081 < t_c = -1{,}68$

Há evidências para concluir que demanda elástica é mais compatível com os dados

Teste da Eficácia da Propaganda

- Quer-se saber se:
  - √ $ß_2$ = 1: aumento de $1.000 na propaganda ocasionará aumento inferior a $1.000 na receita total
  - √ $ß_2$ >1: aumento na propaganda acarretará aumento maior na receita total

Teste de Elasticidade da Demanda (2)

- Teste de Hipóteses de parâmetro:

H_0: $ß_2 = 1$ *vs* *propaganda ineficaz*
H_1: $ß_1 > 1$ *propaganda eficaz*

- Estatística de teste: $T_{b_2} = \frac{2{,}984 - 1}{0{,}1669} = 11{,}89$

- Limite da região crítica: $t_{0,05;\,(52-(2+1))} = 1{,}68$

$T = 11{,}89 > t_c = 1{,}68$

Há evidências para concluir que a despesa com propaganda é justificada com aumento de receita

Qualidade do Ajuste e Teste de Significância

Coeficiente de Determinação (1)

$$R^2 = \frac{SQReg}{SQT} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$
$$= 1 - \frac{SQRes}{SQT} = 1 - \frac{\sum \hat{e}_i^2}{\sum(y_i - \bar{y})^2}$$

- Medida da proporção da variável dependente que é explicada por <u>todas</u> as variáveis explicativas

Exemplo – Lei de Consumo

Regression Analysis: receita total versus preço; propaganda

```
The regression equation is
receita total = 105 - 6,64 preço + 2,98 propaganda

Predictor      Coef  SE Coef      T      P
Constant    104,786    6,483  16,16  0,000
preço        -6,642    3,191  -2,08  0,043
propaganda   2,9843   0,1669  17,88  0,000

S = 6,06961   R-Sq = 86,7%   R-Sq(adj) = 86,2%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       2  11776,2  5888,1  159,83  0,000
Residual Error  49   1805,2    36,8
Total           51  13581,4
```

R^2

Explicada

Não explicada

$$R^2 = 1 - \frac{1.805{,}2}{13.581{,}4} = 0{,}867$$

R^2

- É o quadrado do coeficiente de correlação amostral entre o valor ajustado e o valor observado
- É um recurso descritivo para informação sobre o ajuste do modelo
- Dificuldade: É sempre possível aumentar o R^2 através da adição de novas variáveis no modelo

R^2 Ajustado

- Medida alternativa para a qualidade do ajustamento:

$$\bar{R}^2 = 1 - \frac{\dfrac{SQRes}{n-(k+1)}}{\dfrac{SQT}{n-1}}$$

- O R^2 ajustado penaliza a quantidade de variáveis

Regression Analysis: receita total versus preço; propaganda

```
The regression equation is
receita total = 105 - 6,64 preço + 2,98 propaganda


Predictor      Coef  SE Coef      T      P
Constant    104,786    6,483  16,16  0,000
preço        -6,642    3,191  -2,08  0,043
propaganda   2,9843   0,1669  17,88  0,000

S = 6,06961   R-Sq = 86,7%   R-Sq(adj) = 86,2%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       2  11776,2  5888,1  159,83  0,000
Residual Error  49   1805,2    36,8
Total           51  13581,4
```

R^2 Ajustado

$$\bar{R}^2 = 1 - \frac{\dfrac{1.805,2}{52-(2+1)}}{\dfrac{13.581,4}{52-1}} = 0,862$$

Considerações sobre R^2

- Não mede a adequação do modelo linear
- Para comparação entre modelos, é importante observar a variação do <u>erro quadrático médio</u>
- Sua grandeza depende também do intervalo de variação da variável regressora

  valor grande de R^2 pode ser resultado de variação irrealista de x

Abusos Comuns

- Deve-se tomar cuidado na <u>forma do modelo</u> e na <u>seleção das variáveis</u> que serão usadas.
  - √ Forte associação não implica relação causal entre variáveis
- Relações de regressão são válidas somente dentro da faixa dos dados originais de x.
  - √ Modelos de regressão não são necessariamente válidos para fins de <u>extrapolação</u>

Teste de Significância Global do Modelo

- Hipóteses:

$H_0: \quad b_1 = b_2 = \cdots = b_k = 0 \quad H_1: \quad pelo\ menos\ um \neq 0$

- Estatística de teste:

$$F = \frac{\dfrac{SQReg}{k}}{\dfrac{SQRes}{n-(k+1)}}$$

Distribuição amostral de F:

$$F \sim F_{k,[n-(k+1)];a}$$

Graus de Liberdade

- SQReg: k
- SQRes: n – (k+1)
- SQT: n – 1

Análise de Variância do Modelo

Fonte de variação	*gl*	*SQ*	*QM*	*Razão*
Regressão	k	*SQReg*	$\frac{SQReg}{k}$	$F = \frac{QMReg}{QMRes}$
Erro	$n-(k+1)$	*SQRes*	$\frac{SQRes}{n-(k+1)}$	
Total	$n-1$	*SQT*		

Exemplo – Lanchonetes

Regression Analysis: receita total versus preço; propaganda

```
The regression equation is
receita total = 105 - 6,64 preço + 2,98 propaganda


Predictor      Coef  SE Coef      T      P
Constant    104,786    6,483  16,16  0,000
preço        -6,642    3,191  -2,08  0,043
propaganda   2,9843   0,1669  17,88  0,000


S = 6,06961   R-Sq = 86,7%   R-Sq(adj) = 86,2%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       2  11776,2  5888,1  159,83  0,000
Residual Error  49   1805,2    36,8
Total           51  13581,4
```

Estatística F

$$F = \frac{\frac{11.776,2}{2}}{\frac{1.805,2}{52-(2+1)}} = 159,825$$

Teste de Significância Global

- Nível de significância de 5%
- Limite da região crítica:

$$F_{0,05;\, 2,\, [52-(2+1)]} = f_c = 3,18658$$

$$F = 159,83 > f_c = 3,187$$

Há evidências para concluir que a relação estimada é significante, ou seja, ou o preço, ou a despesa com propaganda, ou ambos têm influência na receita total

Relação entre Teste Conjunto e Testes Individuais

- Qual diferença entre testar os coeficientes conjuntamente ($ß_2=ß_3=0$) ou testá-los separadamente ($ß_2=0$ e $ß_3=0$)?
- O teste F, além de abranger as duas hipóteses conjuntamente, considera a correlação entre os *EMQO*.
- Cada teste *t* não leva isto em conta, logo, eles não se equivalem.

Conflito entre Teste *t* e Teste *F*

- É possível os testes *t* não indicarem coeficiente significante, enquanto o teste *F* implica os coeficientes conjuntamente significativos
- Esta situação ocorre com freqüência quando os dados são multicolineares

Modelo Restrito

- O teste F indicou que ou o preço ou a propaganda ou ambas explicam a Receita Total;
- Como saber se a mudança no preço não tem qualquer efeito sobre a receita, contornando o efeito dos testes t individuais?

- Vamos testar:

  H_0: *preço não tem efeito na receita*

  H_1: *preço tem algum efeito na receita*

  Ou, de outra maneira:

  H_0: $RT_i = b_0 + b_2 a_i + e_i$

  H_1: $RT_i = b_0 + b_1 p_i + b_2 a_i + e_i$

- Se H_0 for verdadeira, esperamos uma pequena variação na soma dos quadrados dos resíduos dos dois modelos.

- Soma do quadrado dos resíduos:
  - √ $SQRes_R$: soma restrita de quadrados dos resíduos $SQRes$ de H_0
  - √ $SQRes_U$: soma não restrita quadrados resíduos $SQRes$ de H_1
- Tem-se sempre que: $SQRes_R - SQRes_U \geq 0$
- A estatística F para testar esta diferença é:

$$F = \frac{\dfrac{SQRes_R - SQRes_U}{J}}{\dfrac{SQRes_U}{n-(k+1)}} \qquad F \sim F_{J,[n-(k+1)];a}$$

J: quantidade de restrições sob H_0

Teste *F*

- H_0 consiste em uma ou mais hipóteses (J) de igualdades

  (não pode incluir hipóteses do tipo: = ou =)
- H_1: ao menos uma das igualdade não é verdadeira
- Estatística F: $F = \dfrac{(SQRes_R - SQRes_U)/J}{SQRes_U/[n-(k+1)]}$
- Distribuição amostral da estatística $F \sim F_{J,[n-(k+1)];a}$

Modelo Restrito

$$RT_i = b_0 + b_2 a_i + e_i$$

$SQRes_R = 1.965$
$J = 1$

Regression Analysis: receita total versus propaganda

```
The regression equation is
receita total = 91,8 + 2,95 propaganda


Predictor     Coef  SE Coef      T      P
Constant    91,831    1,871  49,07  0,000
propaganda  2,9491   0,1715  17,19  0,000


S = 6,26858   R-Sq = 85,5%   R-Sq(adj) = 85,2%


Analysis of Variance

Source          DF     SS     MS       F      P
Regression       1  11617  11617  295,62  0,000
Residual Error  50   1965     39
Total           51  13581
```

Modelo Irrestrito

$$RT_i = b_0 + b_1 p_i + b_2 a_i + e_i$$

$SQRes_U = 1.805,2$

Regression Analysis: receita total versus preço; propaganda

```
The regression equation is
receita total = 105 - 6,64 preço + 2,98 propaganda


Predictor      Coef  SE Coef      T      P
Constant    104,786    6,483  16,16  0,000
preço        -6,642    3,191  -2,08  0,043
propaganda   2,9843   0,1669  17,88  0,000


S = 6,06961   R-Sq = 86,7%   R-Sq(adj) = 86,2%


Analysis of Variance

Source          DF       SS      MS       F      P
Regression       2  11776,2  5888,1  159,83  0,000
Residual Error  49   1805,2    36,8
Total           51  13581,4
```

Teste *F*

- a = 5%
- Estatística de teste:

$$F = \frac{\frac{1.965 - 1.805,168}{1}}{\frac{1805,168}{52-3}} = 4,332$$

- Limite da região crítica:

$F_{0,05;\ 1,\ [52-(2+1)]} = f_c = 4,03839$

$$F = 4,332 > f_c = 4,03839$$

Há evidências para concluir que o preço tem efeito significativo sobre a receita total

Modelo Ampliado

Questionamento

- A relação linear entre receita, preço e gastos com propaganda é uma boa aproximação da realidade?
- Modelo linear implica que crescimento de gastos com propaganda leva a crescimento proporcional da receita, independente do nível
- Amento de gasto com propaganda não pode levar a retornos decrescentes?

Modelo Alternativo

$$RT_i = b_0 + b_1 p_i + b_2 a_i + b_3 a_i^2 + e_i$$

- O termo do quadrado dos gastos leva em conta a diminuição dos resultados em função dos gastos com propaganda
- A resposta de $E(RT_i)$ é:

$$\frac{\Delta E(RT)}{\Delta a_i}, \ p \ cte. \qquad \frac{\partial E(RT_i)}{\partial a_i} = b_2 + 2b^{3} a_i$$

Interpretação

$$\frac{\partial E(RT_i)}{\partial a_i} = b_2 + 2b^{3} a_i$$

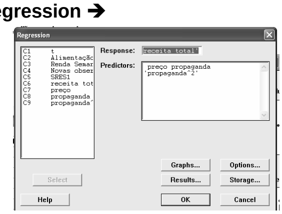
- Quando a_i aumenta de um unidade ($1.000) e p_i é mantido constante, $E(RT_i)$ aumenta de $ß_2 + 2ß_3 a_i$
- Sinais esperados:
  - √ $ß_2 > 0$: Positiva quando $a_i = 0$
  - √ $?_3 < 0$: Para o retorno cair quando a_i aumenta

Estimação do Modelo Ampliado

- No Minitab, criamos uma coluna com os gastos da propaganda ao quadrado.
  - √ Em *Session*, *Editor* > *Enable Command*.

```
MTB > let Let 'propaganda^2' = 'propaganda'**2
```

Stat > Regression > Regression ➔



Resultados

Regression Analysis: receita tota versus preço; propaganda; propaganda^2

```
The regression equation is
receita total = 105 - 6,58 preço + 2,95 propaganda + 0,0017 propaganda^2

Predictor        Coef  SE Coef      T      P
Constant      104,815    6,578  15,93  0,000
preço          -6,582    3,459  -1,90  0,063
propaganda     2,9475   0,7855   3,75  0,000
propaganda^2  0,00173  0,03609   0,05  0,962

S = 6,13236   R-Sq = 86,7%   R-Sq(adj) = 85,9%

Analysis of Variance

Source          DF       SS      MS       F      P
Regression       3  11776,3  3925,4  104,38  0,000
Residual Error  48   1805,1    37,6
Total           51  13581,4
```

Comentários

- O coeficiente de a_i^2 não é significativo, ou seja, não há diminuição do retorno em relação à propaganda ou o coeficiente não foi estimado com precisão
- O preço se tornou não significativo
- Diferenças nas estimativas:

$$\underset{}{RT_i} = \underset{(6{,}483)}{104{,}786} - \underset{(3{,}191)}{6{,}642\,p_i} + \underset{(0{,}1669)}{2{,}9843\,a_i} \qquad R^2 = 86{,}7\% \quad \hat{s}^2 = 36{,}8$$

$$RT_i = \underset{(6{,}578)}{104{,}815} - \underset{(3{,}459)}{6{,}582\,p_i} + \underset{(0{,}7855)}{2{,}9843\,a_i} + 0{,}00173\,a_i^2 \qquad R^2 = 86{,}7\% \quad \hat{s}^2 = 37{,}6$$

Novos Dados

- Uma das soluções quando não há estimativa precisa de parâmetros econômicos é obter mais e melhores dados
- Neste caso, foram coletados dados de outras 26 semanas, disponíveis na planilha: *lanchonete_new*
- O nome das colunas têm o sufixo _new.
- Cria-se a coluna do quadrado dos gastos com propaganda.

Regression Analysis: receita_new versus preço_new; propaganda_new; ...

```
The regression equation is
receita_new = 110 - 10,2 preço_new + 3,36 propaganda_new
              - 0,0268 propaganda_new^2
```

Predictor	Coef	SE Coef	T	P
Constant	110,464	3,741	29,52	0,000
preço_new	-10,198	1,582	-6,45	0,000
propaganda_new	3,3610	0,4217	7,97	0,000
propaganda_new^2	-0,02675	0,01589	-1,68	0,096

S = 5,91871   R-Sq = 87,9%   R-Sq(adj) = 87,4%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	18752,0	6250,7	178,43	0,000
Residual Error	74	2592,3	35,0		
Total	77	21344,3			

Teste de Significância Unilateral

- H_0: $ß_3 = 0$ vs H_1: $ß_3 < 0$
- Nível de significância de 5%
- Limite crítico: $t_{0,05;\,[78-(3+1)]} = f_c = -1,66571$

$$t = -1{,}68 < t_c = -1{,}67$$

Os dados de que dispomos são compatíveis com a hipótese da diminuição do retorno em função da despesa

Comentários

- Comparação com o ajuste anterior:

$$\underset{}{RT_i} = \underset{(6,578)}{104{,}815} - \underset{(3,459)}{6{,}582 p_i} + \underset{(0,7855)}{2{,}9843 a_i} + \underset{(0,03609)}{0{,}00173 a_i^2} \qquad \overline{R}^2 = 85{,}9\% \quad \hat{s}^2 = 37{,}6$$

$$RT^*_i = \underset{(3,741)}{110{,}464} - \underset{(1,582)}{10{,}198 p^*_i} + \underset{(0,4217)}{3{,}3610 a^*_i} - \underset{(0,01589)}{0{,}02675 a^{*2}_i} \qquad \overline{R}^2 = 87{,}4\% \quad \hat{s}^2 = 35{,}0$$

- Houve aumento da precisão (menor erro-padrão)
- O termo a_i^2 tem o sinal esperado

Modelo Restrito

Regression Analysis: receita_new versus preço_new

```
The regression equation is
receita_new = 112 + 5,06 preço_new


Predictor     Coef  SE Coef      T      P
Constant   111,713    8,854  12,62  0,000
preço_new    5,057    4,012   1,26  0,211


S = 16,5860   R-Sq = 2,0%   R-Sq(adj) = 0,8%


Analysis of Variance

Source          DF       SS     MS     F      P
Regression       1    436,9  436,9  1,59  0,211
Residual Error  76  20907,3  275,1
Total           77  21344,3
```

Referências

Bibliografia Recomendada

- Hill, R. C., Griffiths, W. E. e Judge, G. (Saraiva)
  Econometria
- Gujarati, D. N. (Pearson)
  Econometria Básica
- Maddala, G. S. (LTC)
  Introdução à econometria