Lisbon School of Economics & Management
Universidade de Lisboa

Estatística II

Licenciatura em Gestão
2.º Ano/1.º Semestre
2023/2024

Aulas Teóricas N.ºs 24 e 25 (Semana 13)

Docente: Elisabete Fernandes
E-mail: efernandes@iseg.ulisboa.pt

https://doity.com.br/estatistica-aplicada-a-nutricao

https://basiccode.com.br/produto/informatica-basica/

Conteúdos Programáticos

Aulas Teóricas (Semanas 1 a 5)	Aulas Teóricas (Semanas 5 a 7)	Aulas Teóricas (Semanas 7 a 9)	Aulas Teóricas (Semanas 10 a 13)
• **Capítulo 1**: Estimação	• **Capítulo 2**: Testes de Hipóteses	• **Capítulo 3**: Modelo de Regressão Linear	• **Capítulo 4**: Complementos ao Modelo de Regressão Linear

Material didático: Exercícios do Livro Murteira et al (2015), Formulário e Tabelas Estatísticas

Bibliografia: B. Murteira, C. Silva Ribeiro, J. Andrade e Silva, C. Pimenta e F. Pimenta; *Introdução à Estatística*, 2ª ed., Escolar Editora, 2015.

8ª semana (07/11 e 09/11)

T14 - Modelo de Regressão Linea (MRL)r

Interpretação dos parâmetros da regressão; exemplos; Resíduos MQ e regressão ajustada; Propriedades dos estimadores MQ dos coeficientes da regressão; Estimador não enviesado da variância da variável residual; Exemplo.

T15 - Modelo de regressão Linear

Coeficiente de determinação e sua interpretação. Hipótese adicional (H6) e inferência estatística sobre o modelo; Inferência sobre um parâmetro beta. Exemplos

9ª semana (14/11 e 16/11)

T16 - Modelo de Regressão Linear

Mais exemplos de inferência sobre um parâmetro beta; Inferência sobre uma combinação linear de betas; exemplos.

T17 - Modelo de Regressão Linear

Teste de nulidade conjunta de vários coeficientes; exemplo; Teste F à significância global da regressão; Teste de um conjunto de restrições lineares; exemplo.

10ª semana (21/11 e 23/11)

T18 - Complementos ao MRL

Variáveis artificiais: Introdução à modelação de fatores qualitativos, conceito de variável artificial, estimação e interpretação do modelo com variáveis artificiais; exemplos.

T19 - Complementos ao MRL

Inferência na Regressão Linear Múltipla

1

Teste t para Combinações Lineares dos Coeficientes do MRLM

Combinação Linear de Parâmetros

Combinação linear dos parâmetros:

Dado o modelo de RLM:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

Sob as premissas do Teorema de Gauss-Markov, temos que:

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \quad \text{e} \quad Var(\hat{\boldsymbol{\beta}}) = E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T] = (\mathbf{X}^T\mathbf{X})^{-1}\sigma^2$$

Uma combinação linear dos parâmetros, onde cada coeficiente β_j seja multiplicado por uma constante c_j , seria dado por:

$$c_0\alpha + c_1\beta_1 + \ldots + c_k\beta_k = \begin{pmatrix} c_0 & c_1 & \ldots & c_k \end{pmatrix} \begin{pmatrix} \alpha \\ \beta_1 \\ \ldots \\ \beta_k \end{pmatrix} = \mathbf{c}^T\boldsymbol{\beta}$$

Não é difícil demonstrar que:

$$E(\mathbf{c}^T\hat{\boldsymbol{\beta}}) = \mathbf{c}^T E(\hat{\boldsymbol{\beta}}) = \mathbf{c}^T\boldsymbol{\beta}$$

ou seja, uma função linear dos estimadores de MQO é um estimador não viesado da função linear dos parâmetros

E:

$$Var(\mathbf{c}^T\hat{\boldsymbol{\beta}}) = E[(\mathbf{c}^T\hat{\boldsymbol{\beta}} - \mathbf{c}^T\boldsymbol{\beta})(\mathbf{c}^T\hat{\boldsymbol{\beta}} - \mathbf{c}^T\boldsymbol{\beta})^T] = \mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}\sigma^2$$

a variância da função linear será uma grandeza escalar.

A combinação linear dos parâmetros pode ser utilizada em teste de hipóteses para combinações dos parâmetros ou intervalos de confiança para os valores previstos.

Teste t para uma Combinação Linear

Seja o ajuste de MQ:

$$Y_i = \hat{\alpha} + \hat{\beta}_1 X_{1_i} + \hat{\beta}_2 X_{2_i} + \ldots + \hat{\beta}_k X_{k_i} + \hat{e}_i$$

Suponha as seguintes hipóteses para os parâmetros β_1 e β_2 do modelo:

$$\begin{cases} H_0: \beta_1 = \beta_2 \\ H_1: \beta_1 \neq \beta_2 \end{cases}$$

Testar a hipótese nula que $\beta_1 = \beta_2$ é o mesmo que testar a combinação:

$$(0)\alpha + (1)\beta_1 + (-1)\beta_2 \ldots + (0)\beta_k = 0$$

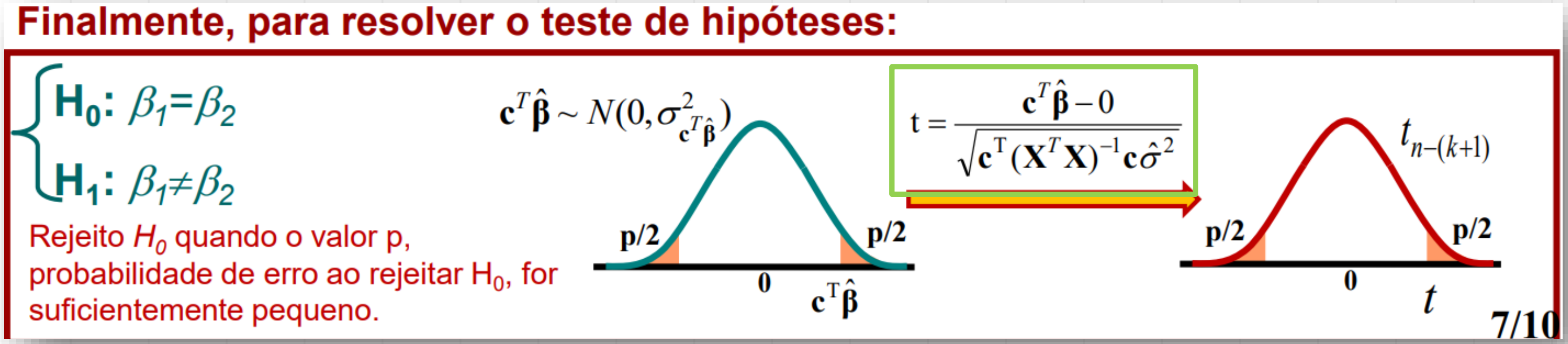
ou, matricialmente: $\begin{pmatrix} 0 & 1 & -1 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta_1 \\ \ldots \\ \beta_k \end{pmatrix} = \mathbf{c}^T \boldsymbol{\beta} = 0$

A estatística $\begin{pmatrix} 0 & 1 & -1 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \ldots \\ \hat{\beta}_k \end{pmatrix} = \mathbf{c}^T \hat{\boldsymbol{\beta}}$

Como combinação linear de v.a.s normais é também uma normal, teremos: $\mathbf{c}^T \hat{\boldsymbol{\beta}} \sim N(\mathbf{c}^T \boldsymbol{\beta}, \sigma^2_{\mathbf{c}^T \hat{\boldsymbol{\beta}}})$

Onde: $\sigma^2_{\mathbf{c}^T \hat{\boldsymbol{\beta}}} = \mathbf{c}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{c} \sigma^2$

Teste t para uma Combinação Linear



[Slide 1 (unicamp.br)]

Teste t para uma Combinação Linear: Exemplo

Sejam os dados da relação entre renda familiar (*Y*), anos de estudo (*X_1*) e idade (*X_2*) do responsável pela família:

$$Y_i = 1{,}9 + 1X_{1_i} + 0{,}06X_{2_i} + \hat{e}_i$$

Há evidências de que o efeito dos anos de estudo seja maior que o da idade?

$$\begin{cases} \mathbf{H_0}: \beta_1 = \beta_2 \\ \mathbf{H_1}: \beta_1 > \beta_2 \end{cases}$$

Testar a hipótese nula que $\beta_1 = \beta_2$ é o mesmo que testar a combinação:

$$(0)\alpha + (1)\beta_1 + (-1)\beta_2 = 0$$

ou, matricialmente: $\begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \end{pmatrix} = \mathbf{c}^T\boldsymbol{\beta} = 0$

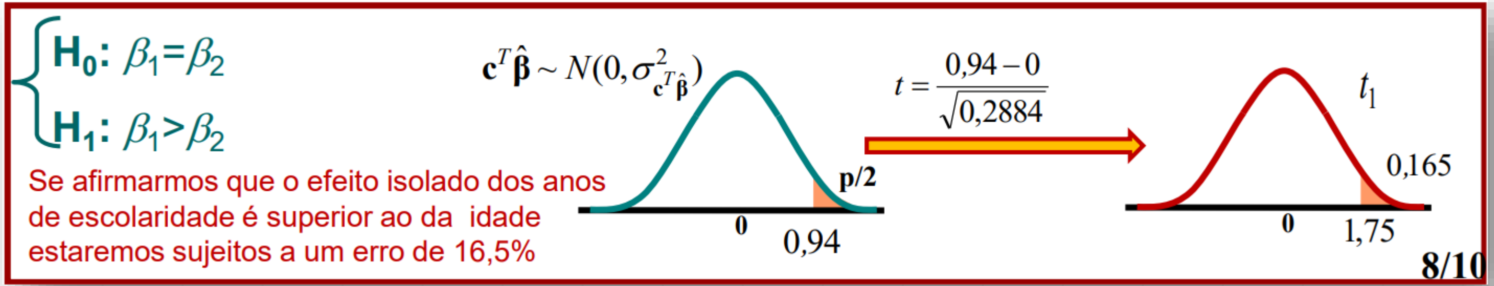
E a estatística : $\mathbf{c}^T\hat{\boldsymbol{\beta}} = \begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1{,}9 \\ 1 \\ 0{,}06 \end{pmatrix} = 0{,}94$

Que estará normalmente distribuída e a estimativa de sua variância será:

$$S^2_{\mathbf{c}^T\hat{\boldsymbol{\beta}}} = \mathbf{c}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{c}\hat{\sigma}^2 = \begin{pmatrix} 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 8{,}95 & 2{,}5 & -0{,}57 \\ 2{,}5 & 1 & -0{,}2 \\ -0{,}57 & -0{,}2 & 0{,}042 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} 0{,}2 = 0{,}2884$$

Slide 1 (unicamp.br)

Teste t para uma Combinação Linear: Exemplo



Slide 1 (unicamp.br)

Seja a equação de regressão múltipla,

$$salário = \beta_0 + \beta_1 educ + \beta_2 produtividade + \varepsilon$$

Verifique, a partir da formulação e construção de um teste de hipóteses, se a variável *educ* apresenta um impacto superior ao da variável *produtividade* na variável resposta. Nesse exemplo, utilize o banco de dados *TEMCOPROD.wf1*.

Index of /wp-content/uploads/2014/02 (hedibert.org)

Resolução do Exercício

Modelo

$$salário = \beta_0 + \beta_1 educ + \beta_2 produtividade + \varepsilon$$

Hipóteses

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

$$H_A : \beta_1 > \beta_2$$

[Index of /wp-content/uploads/2014/02 (hedibert.org)]

Testes de Hipóteses para uma Combinação Linear de Parâmetros

Sejam as hipóteses

$$H_0 : \beta_i = \beta_j \Leftrightarrow \beta_i - \beta_j = 0$$

$$H_A : \beta_i \neq \beta_j \; (\beta_i < \beta_j \quad \text{ou} \quad \beta_i > \beta_j)$$

Estatística do teste

$$t = \frac{(\hat{\beta}_i - \hat{\beta}_j) - 0}{se(\hat{\beta}_i - \hat{\beta}_j)}$$

$$se(\hat{\beta}_i - \hat{\beta}_j) = \sqrt{\left[se(\hat{\beta}_i)\right]^2 + \left[se(\hat{\beta}_j)\right]^2 - 2C\hat{o}v(\hat{\beta}_i, \hat{\beta}_j)}$$

[Index of /wp-content/uploads/2014/02 (hedibert.org)]

Testes de Hipóteses para uma Combinação Linear de Parâmetros

ALTERNATIVAS DE SOLUÇÃO

A) Calcular todos os componentes do erro padrão (o *software* *Eviews* gera a matriz de variâncias e covariâncias para os estimadores dos parâmetros do modelo de regressão. Lembra onde está?).

$$se(\hat{\beta}_i - \hat{\beta}_j) = \sqrt{\left[se(\hat{\beta}_i)\right]^2 + \left[se(\hat{\beta}_j)\right]^2 - 2C\hat{o}v(\hat{\beta}_i, \hat{\beta}_j)}$$

Index of /wp-content/uploads/2014/02 (hedibert.org)

Testes de Hipóteses para uma Combinação Linear de Parâmetros

B) Trabalhar com um modelo transformado para obter o resultado diretamente

Seja

$$\theta = \beta_i - \beta_j,$$

por exemplo, β_j pode ser escrito como $\beta_j = \beta_i - \theta$ e, substituindo este resultado na equação de regressão múltipla, podemos testar H_0: $\theta = 0$, contra uma alternativa apropriada.

Resolução do Exercício

Modelo **(solução *B*)**

$$salário = \beta_0 + \beta_1 educ + \beta_2 produtividade + \varepsilon$$

Hipóteses

$$H_0 : \beta_1 = \beta_2 \Leftrightarrow \beta_1 - \beta_2 = 0$$
$$H_A : \beta_1 > \beta_2$$

Escrevendo $\theta = \beta_1 - \beta_2$, vem que

$$H_0 : \theta = 0$$
$$H_A : \theta > 0$$

Index of /wp-content/uploads/2014/02 (hedibert.org)

Resolução do Exercício

Mas, $\theta = \beta_1 - \beta_2 \Leftrightarrow \beta_1 = \theta + \beta_2$, e substituindo este resultado no modelo proposto, vem que

$$salário = \beta_0 + \beta_1 educ + \beta_2 produtividade + \varepsilon$$

$$\Updownarrow$$

$$salário = \beta_0 + (\theta + \beta_2) educ + \beta_2 produtividade + \varepsilon$$

$$\Updownarrow$$

$$salário = \beta_0 + \theta\, educ + \beta_2 educ + \beta_2 produtividade + \varepsilon$$

$$\Updownarrow$$

$$salário = \beta_0 + \theta\, educ + \beta_2 (educ + produtividade) + \varepsilon$$

Index of /wp-content/uploads/2014/02 (hedibert.org)

Resolução do Exercício

Dependent Variable: SALARIO
Method: Least Squares
Date: 09/06/10 Time: 17:24
Sample: 1 46
Included observations: 46
SALARIO=C(1)+C(2)*EDUC+C(3)*(EDUC+PRODUTIVIDADE)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	12281.96	3944.548	3.113654	0.0033
C(2)	1255.717	591.9782	2.121222	0.0397
C(3)	470.0954	132.2545	3.554476	0.0009

R-squared	0.693710	Mean dependent var	39827.39
Adjusted R-squared	0.679464	S.D. dependent var	10999.24
S.E. of regression	6227.324	Akaike info criterion	20.37427
Sum squared resid	1.67E+09	Schwarz criterion	20.49353
Log likelihood	-465.6083	Hannan-Quinn criter.	20.41895
F-statistic	48.69484	Durbin-Watson stat	1.198179
Prob(F-statistic)	0.000000		

Index of /wp-content/uploads/2014/02 (hedibert.org)

Resolução do Exercício: Decisão

(solução *B*)

Hipóteses

$$H_0 : \theta = 0$$
$$H_A : \theta > 0$$

Sob H_0

$$t_{obs} = \frac{1255{,}717 - 0}{591{,}978} = 2{,}121222$$

$$t_{crítico} = t_{n\text{-}k}^{\alpha} = t_{43}^{0{,}05} = @qtdist(0.95{,}43) = 1{,}681$$

$$p - valor = 1 - @ctdist(2.121222{,}43) = 0{,}01985$$

Index of /wp-content/uploads/2014/02 (hedibert.org)

Inferência na Regressão Linear Múltipla

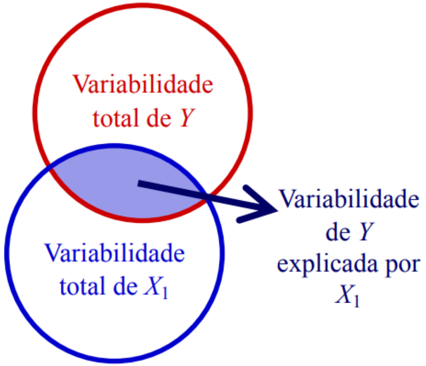
2

Teste para Restrições dos Coeficientes do MRLM

Contribuição Marginal

Seja o modelo de RLS:

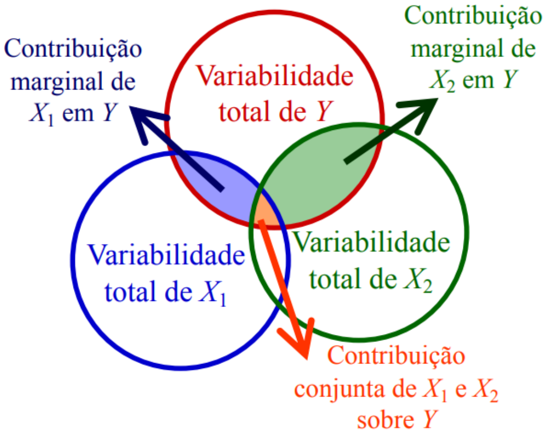
$$Y = \alpha + \beta_1 X_1 + e$$



O grau de relacionamento linear entre X_1 e Y determinará a parcela da variabilidade de Y explicada unicamente por X_1.

Seja o modelo de RLM:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$



Qual a parcela da variabilidade de Y explicada isoladamente por X_2?
Em outras palavras, qual a contribuição marginal de X_2, depois de considerada a contribuição de X_1?

Econometria_Cap9_ContribuicaoMarginal (unicamp.br)

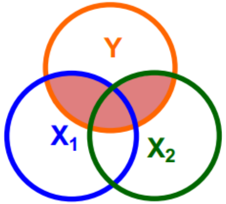
Contribuição Marginal (q=1)

SQReg devido a X_1 e X_2 (Irrestrito):

$SQReg(Y / X_1 \ e \ X_2)$ ou $SQReg_{ir}$

Variabilidade de Y explicada pelo conjunto das variáveis X_1 e X_2.

Graus de liberdade: 2 *coeficientes angulares do modelo* $Y=\alpha+\beta_1 X_1+\beta_2 X_2+e$.

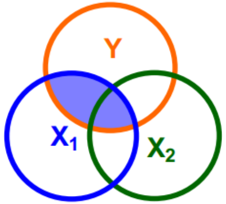


SQReg devido exclusivamente a X_1 (Restrito):

$SQReg(Y / X_1)$ ou $SQReg_r$

Variabilidade de Y explicada exclusivamente por X_1.

Graus de liberdade: 1 *coeficiente angular do modelo* $Y=\alpha+\beta_1 X_1+e$.

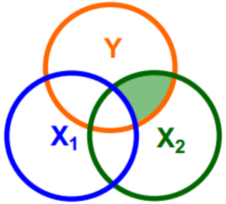


SQReg devido ao acréscimo de X_2:

$$Contribuição \ X_2 = SQReg_{ir} - SQReg_r$$

Variabilidade de Y explicada por X_2 após considerada a variabilidade explicada por X_1.

Graus de liberdade: 1 *novo coeficiente angular incorporado no modelo* (β_2).



Econometria_Cap9_ContribuicaoMarginal (unicamp.br)

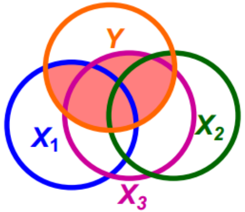
Contribuição Marginal (q=2)

SQReg devido a X_1, X_2 e X_3 (Irrestrito):

$SQReg(Y / X_1, X_2, X_3)$ ou $SQReg_{ir}$

Variabilidade de Y explicada pelo conjunto das variáveis X_1, X_2 e X_3:

Graus de liberdade: 3 *coeficientes angulares do modelo $Y=\alpha+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+e$*

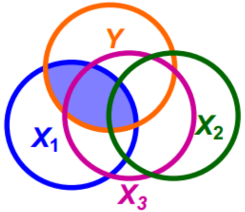


SQReg devido exclusivamente a X_1 (Restrito):

$SQReg(Y / X_1)$ ou $SQReg_r$

Variabilidade de Y explicada exclusivamente por X_1

Graus de liberdade: 1 *coeficiente angular do modelo $Y=\alpha+\beta_1 X_1+e$.*

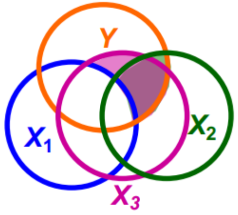


SQReg devido ao acréscimo de X_2 e X_3:

$$Contribuição\ de\ X_2\ e\ X_3 = SQReg_{ir} - SQReg_r$$

Variabilidade de Y explicada por X_2 e X_3 após considerada a variabilidade explicada por X_1

Graus de liberdade: 2 *novos coeficientes angulares incorporados no modelo (β_2 e β_3).*



Econometria_Cap9_ContribuicaoMarginal (unicamp.br)

Teste de Restrição aos Parâmetros

Seja o modelo *irrestrito* de RLM:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + e$$

Para verificarmos se a contribuição de um grupo de q variáveis independentes é significativa no modelo irrestrito devemos, primeiro, elaborar um modelo com restrição aos parâmetros. Suponha que, por simplicidade, as q variáveis que desejamos testar são as últimas das k variáveis do modelo irrestrito (a ordem, obviamente, não faz importância). Nosso modelo *restrito* seria:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{k-q} X_{k-q} + e$$

Em outras palavras, estaríamos interesados em testar a hipótese nula de que os q coeficientes do modelo irrestrito são nulos:

$$H_0 : \beta_{k-q+1} = 0, \ \ldots, \ \beta_k = 0 \quad \text{e} \quad H_1 : pelo\ menos\ um\ \beta_{k-q+j} \neq 0$$

O teste estatístico para restrição aos parâmetros consiste em verificar se a contribuição marginal dessas q variáveis é significativa . A estatística F será dada por:

$$F = \frac{(SQ\,Reg_{ir} - SQ\,Reg_r)/q}{SQ\,Res_{ir}/(n-k-1)} \quad \text{ou} \quad F = \frac{(SQ\,Res_r - SQ\,Res_{ir})/q}{SQ\,Res_{ir}/(n-k-1)}$$

Onde $SQReg_{ir}$ e $SQReg_r$ são, respectivamente, a soma dos quadrados da regressão sem e com restrição nos parâmetros, $SQRes_{ir}$ é a soma dos quadrados dos resíduos da regressão sem restrição.

A soma dos quadrados de regressão SSR dos dois modelos são comparados!!

A soma dos quadrados dos resíduos SSE dos dois modelos são comparados!!

Formulário

Testes de restrições lineares sobre os coeficientes de regressão

Caso geral (m restrições lineares): $F = \dfrac{(VR_0 - VR_1)/m}{VR_1/(n-k)} \sim F(m, n-k)$

$VR_0 =$ Variação residual do modelo com as m restrições lineares;

$VR_1 =$ Variação residual do modelo sem restrições

Econometria_Cap9_ContribuicaoMarginal (unicamp.br)

Testes de Hipóteses para Restrições

Testes de restrições lineares sobre os coeficientes de regressão

Caso geral (m restrições lineares): $\boxed{F = \dfrac{(VR_0 - VR_1)/m}{VR_1/(n-k)} \sim F(m, n-k)}$

VR_0 = Variação residual do modelo com as m restrições lineares;

VR_1 = Variação residual do modelo sem restrições

Casos particulares

- Uma única restrição (m=1): $t_{\hat{\delta}} = \dfrac{\hat{\delta} - \delta}{s_{\hat{\delta}}} \sim t(n-k)$ em que $s_{\hat{\delta}}$ é o erro padrão de $\hat{\delta} = cb$.
- Nulidade conjunta: $F = \dfrac{R^2/(k-1)}{(1-R^2)/(n-k)} = \dfrac{VE/(k-1)}{VR/(n-k)} \sim F(k-1, n-k)$
- Nulidade de um subconjunto (m)

$$F = \frac{(VR_0 - VR_1)/m}{VR_1/(n-k)} \sim F(m, n-k) \text{ ou } F = \frac{(R^2 - R_0^2)/m}{(1-R^2)/(n-k)} \sim F(m, n-k)$$

VR_1 = Variação residual do modelo sem restrições;

VR_0 = Variação residual do modelo com restrições

R_0 = Coeficiente de determinação do modelo com restrições

O Teste F para a Contribuição Marginal

Seja o modelo de RLM:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + e$$

Para testarmos a contribuição marginal de um grupo de q variáveis, teremos as hipóteses:

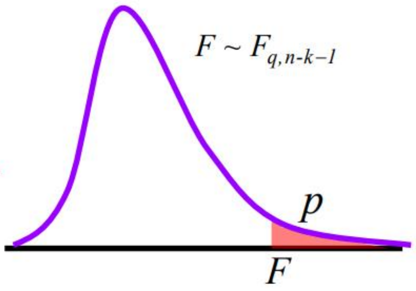
$$\begin{cases} \mathbf{H_0}: \beta_{k-q+1} = \ldots = \beta_k = 0 \text{ (não contribui)} \\ \mathbf{H_1}: \text{Pelo menos um } \beta_j \neq 0 \text{ (contribui)} \end{cases}$$

A estatística de teste será

$$\boxed{F = \frac{(SQReg_{ir} - SQReg_r)/q}{SQRes_{ir}/(n-k-1)}}$$

A soma dos quadrados de regressão SSR dos dois modelos são comparados!!

Considerando H_0 verdadeiro, a distribuição de *F* será...



Rejeitar H_0 significa afirmar que o grupo de *q* variáveis contribui para explicar *Y*, ou seja, há relação significativa entre pelo menos uma destas variáveis explicativas e a variável dependente.

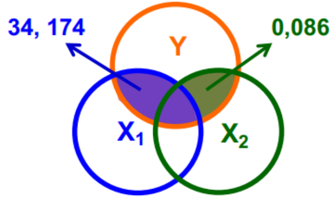
Econometria_Cap9_ContribuicaoMarginal (unicamp.br)

O Teste F para a Contribuição Marginal: Exemplo

Na relação entre renda familiar (*Y*), anos de estudo (X_1) e idade (X_2) do responsável pela família, a contribuição marginal da idade é significativa?

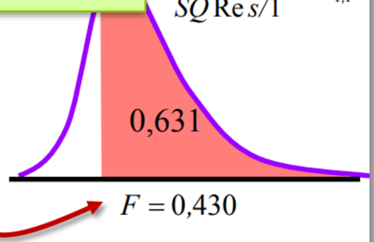
Modelo Irrestrito: $Y_i = 1{,}9 + 1X_{1_i} + 0{,}06X_{2_i} + \hat{e}_i$

Modelo Restrito: $Y_i = 2{,}714 + 1{,}286X_{1_i} + \hat{e}_i$



Fonte	gl	Soma dos Quadrados	Quadrados Médios	F
Regressão Restrita	1	34,714	34,714	F =0,086/0,2 = 0,43
Contribuição Marginal	1	0.086	0.086	0.430
Regressão Irrestrita	2	34.800	17.400	87.000
Resíduos	1	0.200	0.200	
Total	3	35.000		

$$\frac{SQ\,Reg_{cont}/1}{SQ\,Res/1} \sim F_{1,1}$$



Não há evidências para afirmar que a contribuição marginal da idade sobre a variabilidade da renda familiar seja significativa. A probabilidade de erro ao fazermos tal afirmação é muito alta, de aproximadamente 63%.

Econometria_Cap9_ContribuicaoMarginal (unicamp.br)

Anova para a Contribuição Marginal

Fonte	gl	Soma dos Quadrados	Quadrados Médios	F
Regressão Restrita	$k - q$	$SQ\text{Re}\,g_r$	$SQ\text{Re}\,g_r/(k-q)$	
Contribuição	q	$SQ\text{Re}\,g_{ir} - SQ\text{Re}\,g_r$	$\dfrac{(SQ\text{Re}\,g_{ir} - SQ\text{Re}\,g_r)}{q}$	$F = \dfrac{QMReg_{contribuição}}{QMRes_{ir}}$
Regressão Irrestrita	k	$SQ\text{Re}\,g_{ir}$	$SQ\text{Re}\,g_{ir}/k$	$F = \dfrac{QMReg_{ir}}{QMRes_{ir}}$
Resíduos	$n-(k+1)$	$SQ\text{Re}\,s_{ir}$	$SQ\text{Re}\,s_{ir}/(n-k-1)$	
Total	$n-1$	STQ		

[Econometria_Cap9_ContribuicaoMarginal (unicamp.br)]

Curiosidade

Coeficiente de Correlação Parcial

Seja o modelo de RLM:

$$Y_i = \alpha + \beta_1 X_{1i} + \ldots + \beta_k X_{k_i} + e_i$$

Da mesma forma que análise de variância permite considerar a contribuição marginal de uma variável explanatória, podemos estender o conceito de correlação simples para ver em que medida a variável dependente e uma das variáveis independentes estão relacionadas, depois de isolados os efeitos das demais variáveis explanatórias do modelo:

$$r_{Y1.23\ldots k}$$

O $r_{Y1.23\ldots k}$ é a correlação parcial entre Y e X_1, mantendo-se constantes os efeitos das k–1 variáveis independentes restantes. Seu quadrado, $r^2_{Y1.23\ldots k}$ é o respectivo coeficiente de determinação parcial.

O respectivo coeficiente de determinação parcial, que mede a parcela da variabilidade de Y não associada aos demais regressores que é explicada exclusivamente por X_1 será representado por

$$r^2_{Y1.23\ldots k}$$

Econometria_Cap9_ContribuicaoMarginal (unicamp.br)

Curiosidade

Coeficiente de Correlação Parcial

Seja o modelo de RLM: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$

Para calcularmos $r_{Y1.2}$, poderíamos seguir os seguintes passos:

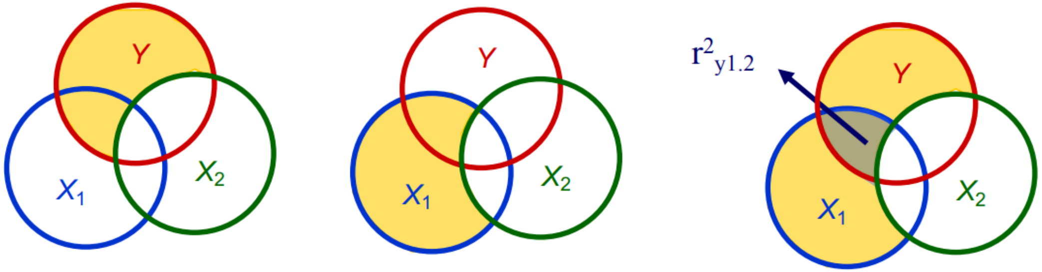
Passo 1) Estimar resíduos do ajuste: $Y = \hat{\delta}_0 + \hat{\delta}_1 X_2 + \hat{e}_{y2}$

$\hat{e}_{y2}$ contém a parcela de Y não associada a X_2.

Passo 2) Estimar resíduos da relação: $X_1 = \hat{\gamma}_0 + \hat{\gamma}_1 X_2 + \hat{e}_{12}$

$\hat{e}_{12}$ contém a parcela de X_1 não associada a X_2.

Passo 3) Estimar correlação entre os resíduos dos dois modelos: $r_{Y1.2} = r_{\hat{e}_{Y2}\hat{e}_{12}}$



Econometria_Cap9_ContribuicaoMarginal (unicamp.br)

Curiosidade

Coeficiente de Correlação Parcial

Coeficiente de Correlação Parcial

No caso de duas variáveis independentes:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e$$

Então a correlação parcial poderá ser diretamente obtida por:

$$r_{Y1.2} = \frac{r_{Y1} - r_{Y2} r_{12}}{\sqrt{(1 - r_{12}^2)(1 - r_{Y2}^2)}} \quad \text{e} \quad r_{Y2.1} = \frac{r_{Y2} - r_{Y1} r_{12}}{\sqrt{(1 - r_{12}^2)(1 - r_{Y1}^2)}}$$

Analogamente, o coeficiente de determinação parcial será dado por:

$$r^2{}_{Y1.2} = \frac{R^2 - r_{Y2}^2}{1 - r_{Y2}^2}$$

Econometria_Cap9_ContribuicaoMarginal (unicamp.br)

Curiosidade

Coeficiente de Correlação Parcial: Exemplo

Vamos considerar a relação para a renda (*Y*) como função dos anos de estudo (X_1) e idade (X_2).

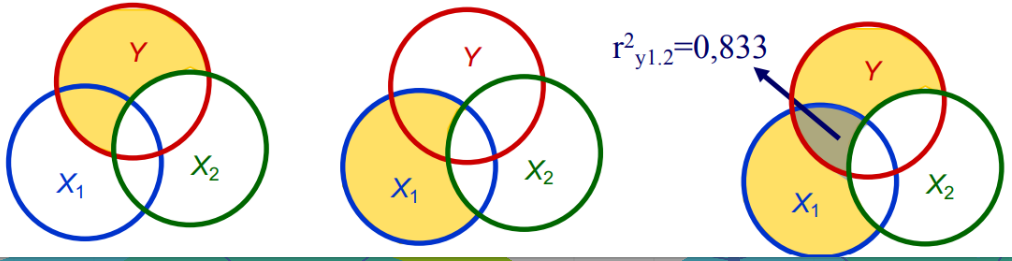
$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e \quad \Rightarrow \quad Y_i = 1{,}9 + 1X_{1_i} + 0{,}06X_{2_i} + \hat{e}_i$$

Para calcularmos a correlação parcial entre renda e anos de estudo ($r_{Y1.2}$), podemos seguir os seguintes passos:

Passo 1) Estimar resíduos do ajuste: $Y = 2{,}714 + 1{,}286X_2 + \hat{e}_{y2}$

Passo 2) Estimar resíduos da relação: $X_1 = -2{,}5 + 0{,}2X_2 + \hat{e}_{12}$

Passo 3) Estimar correlação entre os resíduos dos dois modelos: $r_{Y1.2} = r_{\hat{e}_{Y2}\hat{e}_{12}} = 0{,}913$



Econometria_Cap9_ContribuicaoMarginal (unicamp.br)

Curiosidade

Coeficiente de Correlação Parcial: Exemplo

Vamos considerar a relação para a renda (*Y*) como função dos anos de estudo (*X*$_1$) e idade (*X*$_2$).

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + e \quad \Rightarrow \quad Y_i = 1{,}9 + 1X_{1_i} + 0{,}06X_{2_i} + \hat{e}_i$$

Podemos ainda calcular a correlação parcial entre renda e anos de estudo ($r_{Y1.2}$) diretamente por:

$$r_{Y1.2} = \frac{r_{Y1} - r_{Y2}r_{12}}{\sqrt{(1-r_{12}^2)(1-r_{Y2}^2)}} = \frac{(0{,}9959)-(0{,}9827)(-0{,}9757)}{\sqrt{[1-(0{,}9759)^2][1-(0{,}9827)^2]}} = 0{,}913$$

Onde:

$$r_{Y1} = \frac{\sum_{i=1}^{n} x_{1_i} y_i}{\sqrt{\sum_{i=1}^{n} x_{1_i}^2 \sum_{i=1}^{n} y_i^2}} = \frac{27}{\sqrt{(21)(35)}} = 0{,}9959$$

$$r_{Y2} = \frac{\sum_{i=1}^{n} x_{2_i} y_i}{\sqrt{\sum_{i=1}^{n} x_{2_i}^2 \sum_{i=1}^{n} y_i^2}} = \frac{130}{\sqrt{(500)(35)}} = 0{,}9827$$

$$r_{12} = \frac{\sum_{i=1}^{n} x_{1_i} x_{2_i}}{\sqrt{\sum_{i=1}^{n} x_{1_i}^2 \sum_{i=1}^{n} x_{2_i}^2}} = \frac{100}{\sqrt{(21)(500)}} = 0{,}9759$$

Econometria_Cap9_ContribuicaoMarginal (unicamp.br)

3 Variáveis Artificiais

Complementos ao Modelo de Regressão Linear Múltipla

Variáveis Binárias: Definição

1) Escala Nominal: Valores representam categorias (*nomes*). Não se pode falar que um seja maior que o outro. Exemplo: sexo.

2) Escala Ordinal: Valores representam uma hierarquia de posições. Não se pode, entretanto, falar quão maior é um valor em relação a outro. Exemplo: classe social.

3) Escala Intervalar: Valores representam ordem e é possível mensurar intervalo entre eles. Não se pode, entretanto, dizer quantas vezes um é maior que outro. Exemplo: ano.

4) Escala de razão: Valores representam ordem, é possível mensurar intervalo entre eles e quantificar grandezas em uma escala de razão. Exemplo: renda.

Variável Binária

Uma variável binária (variável *dummy*) pode representar dois estados possíveis::

$$X \begin{cases} 0, \text{ ausência da característica de interesse (Fracasso)} \\ 1, \text{ presença da característica de interesse (Sucesso)} \end{cases}$$

Podemos, assim, estimar a influência de variáveis explicativas (independentes) nominais ou ordinais em modelo de regressão, da mesma maneira que fazemos com variáveis quantitativas de escala intervalar ou de razão.

Variáveis Dummy

Uma forma de introduzir características qualitativas em modelos econométricos consiste na utilização de variáveis *dummy* (fictícia, postiça), frequentemente chamadas de variáveis binárias ou dicotômicas, uma vez que assumem apenas um de dois valores – em geral 0 ou 1 – para indicar a presença ou ausência de determinada característica.

regressaolinearmultipla.pdf (hedibert.org)

Variáveis Dummy

Vale lembrar que a variável *dummy* representa estados ou níveis de fatores, ou seja representa algo que não possui valores numéricos ou, caso possua, estes valores não têm realmente um significado numérico.

Assim, uma variável *dummy*, D, pode ser descrita da seguinte maneira:

$$D = \begin{cases} 0, & \text{se a característica não estiver presente} \\ 1, & \text{se a característica estiver presente} \end{cases}$$

regressaolinearmultipla.pdf (hedibert.org)

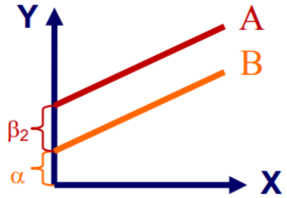
Duas Categorias Nominais

- Para representar duas categorias nominais (A e B) em um modelo de regressão, é necessária apenas uma variável binária (*D*);
- A categoria de referência (ou referência de análise) é aquela em que *D*=0;

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_i + e_i$$

Categoria	D_i
A	1
B	0

O coeficiente β_2 indicaria quanto *Y* seria, em média, maior (ou menor) para a categoria A (*D*=1) que a categoria de referência B (*D*=0), independente do valor de *X*.



Para A: $Y_i = (\alpha + \beta_2) + \beta_1 X_i + e_i$

Para B: $Y_i = \alpha + \beta_1 X_i + e_i$

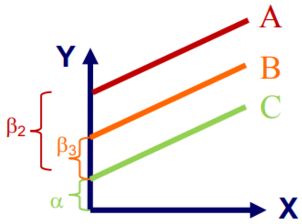
Múltiplas Categorias Nominais

- Para representar *k* categorias nominais, são necessárias *k*–1 variáveis binárias *D's*.
- A referência da análise será a categoria em que todas as variáveis binárias assumem 0 simultaneamente. Exemplo, supondo *k*=3:

Categoria	D_{1i}	D_{2i}
A	1	0
B	0	1
C	0	0

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_{1_i} + \beta_3 D_{2_i} + e_i$$

O coeficiente β_2 indicaria quanto Y seria, em média, maior para a categoria A (D_1=1) que a categoria de referência C (D_1=0 e D_2=0), independente do valor de X. O coeficiente β_3 indicaria quanto Y seria, em média, maior para a categoria B (D_2=1) que a categoria de referência C. Pois teríamos:



Para A: $Y_i = (\alpha + \beta_2) + \beta_1 X_i + e_i$

Para B: $Y_i = (\alpha + \beta_3) + \beta_1 X_i + e_i$

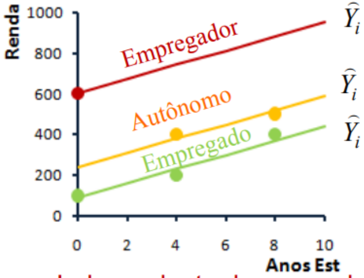
Para C: $Y_i = \alpha + \beta_1 X_i + e_i$

Econometria_Cap11_VariaveisBinarias (unicamp.br)

Múltiplas Categorias: Exemplo

Sejam os dados amostrais para renda (*Y*), anos de estudo (*X*) e posição na ocupação:

Y_i	X_i	*Posição Ocupação*	D_{1i}	D_{2i}
100	0	Empregado	0	0
200	4	Empregado	0	0
400	8	Empregado	0	0
400	4	Autônomo	1	0
500	8	Autônomo	1	0
600	0	Empregador	0	1



$$\hat{Y}_i = (93{,}3 + 506{,}7) + 35X_i$$

$$\hat{Y}_i = (93{,}3 + 146{,}7) + 35X_i$$

$$\hat{Y}_i = 93{,}3 + 35X_i$$

Independente dos anos de escolaridade, o rendimento médio dos autônomos seria 146,7 reais superior ao dos empregados e o dos empregadores 506,7 superior.

Sejam as binárias:

$$D_{1_i} = \begin{cases} 1, & se\ \ Autônomo \\ 0, & c.c. \end{cases}$$

$$D_{2_i} = \begin{cases} 1, & se\ \ Empregador \\ 0, & c.c. \end{cases}$$

Supondo o modelo:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 D_{1_i} + \beta_3 D_{2_i} + e_i$$

O estimador de MQO:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{y})$$

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 6 & 24 & 2 & 1 \\ 24 & 160 & 12 & 0 \\ 2 & 12 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 2200 \\ 9600 \\ 900 \\ 600 \end{pmatrix} = \begin{pmatrix} 93{,}3 \\ 35 \\ 146{,}7 \\ 506{,}7 \end{pmatrix}$$

Econometria_Cap11_VariaveisBinarias (unicamp.br)

A senhorita Rose Jolie, gerente do departamento de RH da empresa TEMCO, gostaria de estimar os parâmetros de um modelo de regressão linear que levasse em consideração as variáveis explicativas *educ* e *dept* na explicação da variável resposta *salário*. Auxilie a senhorita Jolie nesta proposição.

regressaolinearmultipla.pdf (hedibert.org)

Apenas para lembrar, a senhorita Jolie, coletou informações de uma amostra aleatória de 46 funcionários da empresa, sobre as seguintes variáveis:

id – número cadastral do funcionário;

salario – anual, em dólares;

anosemp – tempo (em anos) na empresa;

expprev – experiência anterior (em anos);

educ – anos de estudo após o segundo grau;

sexo – (feminino = 0, masculino = 1);

dept – departamento no qual o funcionário atua

(Compras = 1, Engenharia = 2, Propaganda = 3, Vendas = 4);

super – número de empregados sob responsabilidade do empregado.

regressaolinearmultipla.pdf (hedibert.org)

Resolução do Exercício

À primeira vista, como existem quatro departamentos na empresa *TEMCO*, Rose Jolie poderia optar por usar a variável *dept*, com os valores 1, 2, 3 e 4.

Dessa maneira,

$$salário = \beta_1 + \beta_2 educ + \beta_3 dept + \varepsilon$$

No entanto, ao fazer isto, Rose Jolie estaria introduzindo uma ideia de espaçamento, que ficará mais clara nos resultados descritos nos *slides* a seguir.

regressaolinearmultipla.pdf (hedibert.org)

Resolução do Exercício

Escrevendo a equação de regressão de interesse, para cada um dos departamentos, temos que:

$$E(salário|educ, dept = 1) = (\beta_1 + \beta_3) + \beta_2 educ$$

$$E(salário \mid educ, dept = 2) = (\beta_1 + 2\beta_3) + \beta_2 educ$$

$$E(salário \mid educ, dept = 3) = (\beta_1 + 3\beta_3) + \beta_2 educ$$

$$E(salário \mid educ, dept = 4) = (\beta_1 + 4\beta_3) + \beta_2 educ$$

regressaolinearmultipla.pdf (hedibert.org)

Resolução do Exercício

Dessa forma, admitiríamos, por exemplo, que

$$E(salário \mid educ, dept = 2) - E(salário \mid educ, dept = 1) =$$

$$= E(salário \mid educ, dept = 4) - E(salário \mid educ, dept = 3) =$$

$$= \beta_3$$

ou seja, que a diferença entre os salários esperados dos funcionários dos departamentos de Engenharia e Compras é a mesma que a dos funcionários dos departamentos de Propaganda e Engenharia, mantendo constante o tempo de escolaridade.

regressaolinearmultipla.pdf (hedibert.org)

Resolução do Exercício

Assim, se Rose Jolie utilizasse *dept* da forma como foi construída, então ela estaria impondo uma restrição ao modelo, que não sabemos se é real.

Ainda, se a ordem das categorias da variável departamento fosse alterada, estaríamos propondo um novo conjunto de restrições ao modelo, o que muito provavelmente nos levaria a resultados completamente diferentes do caso anterior.

regressaolinearmultipla.pdf (hedibert.org)

Resolução do Exercício

Portanto, o ideal seria utilizar um grupo de variáveis que representasse os estados de interesse, que no nosso caso não apresentam nenhuma ordenação natural, de tal sorte a nunca alterar o resultado final, qualquer que seja o critério de criação adotado para a construção destas variáveis.

regressaolinearmultipla.pdf (hedibert.org)

Resolução do Exercício

A solução é, portanto, trabalharmos com algumas variáveis *dummy*.

No geral, se temos p estados, devemos trabalhar com $p - 1$ variáveis *dummy*.

regressaolinearmultipla.pdf (hedibert.org)

Resolução do Exercício

Para o nosso exemplo, poderíamos definir as variáveis *dummy* D_C, D_E e D_P da seguinte maneira, para representar os estados da variável departamento:

dept	D_C	D_E	D_P
Compras	1	0	0
Engenharia	0	1	0
Propaganda	0	0	1
Vendas	0	0	0

regressaolinearmultipla.pdf (hedibert.org)

Resolução do Exercício

Assim, partindo do modelo de regressão linear

$$y_i = \beta_1 + \beta_2\, educ_i + \delta_1\, D_{Ci} + \delta_2\, D_{Ei} + \delta_3\, D_{Pi} + \varepsilon_I$$

temos que:

Compras: $y_i = (\beta_1 + \delta_1) + \beta_2 educ_i + \varepsilon_i$

Engenharia: $y_i = (\beta_1 + \delta_2) + \beta_2 educ_i + \varepsilon_i$

Propaganda: $y_i = (\beta_1 + \delta_3) + \beta_2 educ_i + \varepsilon_i$

Vendas: $y_i = \beta_1 + \beta_2\, educ_i + \varepsilon_i$

regressaolinearmultipla.pdf (hedibert.org)

Resolução do Exercício

Do *slide* 14, o parâmetro δ_1, por exemplo, pode ser interpretado como a diferença esperada entre os salários dos profissionais das áreas de Compras e Vendas, que apresentam o mesmo tempo de escolaridade.

Ainda, vale lembrar que, estamos admitindo que o acréscimo médio no salário correspondente ao acréscimo em um ano de escolaridade é o mesmo para os quatro departamentos.

regressaolinearmultipla.pdf (hedibert.org)

4 Exercícios Suplementares

Admita que a variável aleatória X, que representa o número de linhas de código sem erros de sintaxe até à deteção da primeira linha de código com erros de tal tipo, possui função de probabilidade

$$P(X=x)=\begin{cases}(1-p)^x\,p, & x=0,1,2,\ldots\\ 0, & \text{caso contrário,}\end{cases}$$

onde p é um parâmetro desconhecido.

Deduza o estimador de máxima verosimilhança de p, baseado numa amostra aleatória de dimensão n proveniente de X.

Resolução do Exercício

- **V.a. de interesse; f.p.**

  X

  $$P(X=x)=\begin{cases}(1-p)^x\,p, & x=0,1,2,\ldots\\ 0, & \text{caso contrário,}\end{cases}$$

- **Parâmetro desconhecido**

  $p \quad (p\in[0,1])$

- **Amostra**

  $\underline{x}=(x_1,\ldots,x_n)$, amostra de dimensão n proveniente de X.

- **Dedução do estimador de MV de p**

  Passo 1 — Função de verosimilhança

$$\begin{aligned} L(p\,|\,\underline{x}) &= P(\underline{X}=\underline{x}) \stackrel{X_i\,indep}{=} \prod_{i=1}^{n} P(X_i=x_i)\\ &\stackrel{X_i\sim X}{=} \prod_{i=1}^{n} P(X=x_i)\\ &= \prod_{i=1}^{n}\left[(1-p)^{x_i}\,p\right]\\ &= p^n\,(1-p)^{\sum_{i=1}^{n}x_i}, \quad p\in[0,1]\end{aligned}$$

Passo 2 — Função de log-verosimilhança

$$\ln L(p|\underline{x}) = n\ln(p)+\ln(1-p)\times\sum_{i=1}^{n}x_i, \quad p\in[0,1]$$

Resolução do Exercício

A estimativa de MV de p é representada por $\hat{p}$ e

$$\hat{p} \quad : \quad \begin{cases} \left.\dfrac{d \ln L(p|\underline{x})}{dp}\right|_{p=\hat{p}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\dfrac{d^2 \ln L(p|\underline{x})}{dp^2}\right|_{p=\hat{p}} < 0 & \text{(ponto de máximo)} \end{cases}$$

$$\begin{cases} \frac{n}{\hat{p}} - \frac{\sum_{i=1}^n x_i}{1-\hat{p}} = 0 \\ -\frac{n}{\hat{p}^2} - \frac{\sum_{i=1}^n x_i}{(1-\hat{p})^2} < 0 \quad \text{(prop. verdadeira)} \end{cases} \overset{\hat{p}\neq 0,1}{\Leftrightarrow} \begin{cases} n - \hat{p}\left(n + \sum_{i=1}^n x_i\right) = 0 \Leftrightarrow \hat{p} = \frac{n}{n+\sum_{i=1}^n x_i} \\ - \end{cases}$$

Passo 4 — Estimador de MV de p

$$EMV(p) \quad = \quad \frac{n}{n + \sum_{i=1}^n X_i} = (1+\bar{X})^{-1}$$

Admita que o peso máximo suportado por uma marquesa portátil ergonómica é uma variável aleatória X com distribuição normal. Foram selecionados aleatoriamente 28 dessas marquesas, tendo-se observado uma variância amostral corrigida de $s^2 = 3.8$.

Qual é o intervalo de confiança a 90% para a variância de X?

A: [2.482, 6.060] **B**: [1.599, 2.521] $\boxed{\textbf{C}: [2.558, 6.353]}$ **D**: [2.376, 7.042]

Resolução do Exercício

- **V.a. de interesse**

  X = peso máximo suportado por uma marquesa portátil ergonómica

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$, $\mu = E(X)$ desconhecido, $\sigma^2 = V(X)$ DESCONHECIDO

- **Obtenção do IC para σ^2**

  Passo 1 — Variável aleatória fulcral para σ^2

$$Z = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

  Passo 2 — Quantis de probabilidade

  Dado que $n = 28$ e $(1-\alpha) \times 100\% = 90\%$, usaremos os quantis

$$(a_\alpha, b_\alpha) : \begin{cases} P(a_\alpha \le Z \le b_\alpha) = 1-\alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2 \end{cases} \Leftrightarrow \begin{cases} a_\alpha = F^{-1}_{\chi^2_{(n-1)}}(\alpha/2) = F^{-1}_{\chi^2_{(27)}}(0.05) \overset{tabela/calc.}{=} 16.15 \\ b_\alpha = F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2) = F^{-1}_{\chi^2_{(27)}}(0.95) \overset{tabela/calc.}{=} 40.11. \end{cases}$$

Resolução do Exercício

Passo 3 — Inversão da desigualdade $a_\alpha \leq Z \leq b_\alpha$

$$P(a_\alpha \leq Z \leq b_\alpha) = 1-\alpha$$

$$P\left[a_\alpha \leq \frac{(n-1)S^2}{\sigma^2} \leq b_\alpha\right] = 1-\alpha$$

$$P\left[\frac{(n-1)S^2}{b_\alpha} \leq \sigma^2 \leq \frac{(n-1)S^2}{a_\alpha}\right] = 1-\alpha$$

Passo 4 — Concretização

Atendendo à expressão geral do IC para σ^2,

$$IC_{(1-\alpha)\times 100\%}(\sigma^2) = \left[\frac{(n-1)\,s^2}{F^{-1}_{\chi^2_{(n-1)}}(1-\alpha/2)}, \frac{(n-1)\,s^2}{F^{-1}_{\chi^2_{(n-1)}}(\alpha/2)}\right],$$

ao par de quantis acima e ao facto de $s^2 = 3.8$, temos:

$$\begin{aligned} IC_{90\%}(\sigma) &= \left[\frac{(28-1)\times 3.8}{40.11}, \frac{(28-1)\times 3.8}{16.15}\right] \\ &\simeq [2.558, 6.353]. \end{aligned}$$

Suponha que o comprimento (em mm) de uma peça é uma variável aleatória com distribuição normal com valor esperado μ desconhecido e variância σ^2 desconhecida.

Teste a hipótese $H_0 : \mu = 100$ contra $H_1 : \mu < 100$, num dia em que foram recolhidas casualmente 4 peças e a média e a variância corrigida amostrais foram de $\bar{x} = 101.0$ e $s^2 = 2.56$. Decida com base no valor-p.

Resolução do Exercício

- **V.a. de interesse**

  X = comprimento de uma peça metálica

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ DESCONHECIDO

  σ desconhecido

- **Hipóteses**

  $H_0 : \mu = \mu_0 = 100 \quad \text{vs.} \quad H_1 : \mu < \mu_0$

- **Estatística de teste**

$$\boxed{T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim_{H_0} t_{(n-1)}}$$

- **Região de rejeição de H_0 (para valores da estatística de teste)**

  O teste é unilateral inferior ($H_1 : \mu > \mu_0$), logo a região de rejeição de H_0 é tipo $(-\infty, c)$.

  Decisão (com base no valor-p)

  Atendendo a que o valor observado da estatística de teste e o valor-p aproximado são iguais a

$$\boxed{\begin{aligned} t &= \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \\ &= \frac{101.0 - 100}{\frac{\sqrt{2.56}}{\sqrt{4}}} \\ &\simeq 1.25 \end{aligned}}$$

Resolução do Exercício

$$
\begin{aligned}
valor - p &= P(T < t \mid H_0) \\
&= F_{t_{(n-1)}}(t) \\
&= F_{t_{(3)}}(1.25) \\
&\overset{tabelas/calc.}{=} 0.85
\end{aligned}
$$

é suposto:

- não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq valor - p = 85\%$, nomeadamente ao níveis usuais de significância $(1\%, 5\%, 10\%)$[;
- rejeitar H_0 a qualquer n.s. $\alpha_0 > valor - p = 85\%$].

O engenheiro que gere uma linha de apoio *online* pretende averiguar se o tempo de assistência (em minutos) possui função de distribuição dada por $F_0(x) = 1 - e^{-(x/2)^2}$, para $x > 0$ (hipótese H_0).

Recorra ao teste de ajustamento do qui-quadrado e à tabela de frequências seguinte para testar H_0 ao nível de significância de 1%.

Tempo de assistência	$]0,2/3]$	$]2/3,2]$	$]2,3]$	$]3,+\infty[$
Frequência absoluta observada	13	32	13	6
Frequência absoluta esperada sob H_0	6.73	33.73	16.80	6.74

Resolução do Exercício

- **V.a. de interesse**

  $X =$ tempo de assistência$(em minutos)$

- **Hipóteses**

$$\boxed{\begin{array}{l} H_0 : F_X(x) = P(X \leq x) = F_0(x) = 1 - e^{-(x/2)^2}, \quad x > 0 \\ H_1 : F_X(x) \neq F_0(x), \quad \text{para algum } x > 0 \end{array}}$$

- **Nível de significância**

  $\alpha_0 = 1\%$

- **Estatística de teste**

$$\boxed{T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \overset{a}{\sim}_{H_0} \chi^2_{(k-1)},}$$

onde: k = no. de classes = 4; O_i = frequência absoluta observável da classe i; E_i = frequência absoluta esperada, sob H_0, da classe i.

Resolução do Exercício

exp	2,718282
p1	0,105161
n*p1	6,730284
n	64

- [**Frequências absolutas esperadas sob H_0** — De acordo com a tabela facultada, as frequências absolutas esperadas sob H_0 aproximadas às centésimas são: $E_1 \simeq 6.73$; $E_2 \simeq 33.73$; $E_3 \simeq 16.80$; $E_4 \simeq 6.74$. Não é necessário fazer qualquer agrupamento de classes uma vez que se verifica $E_i \geq 5$, em pelo menos 80% das classes, e que $E_i \geq 1$, para todo o i. Caso fosse preciso efectuar agrupamento de classes, os valores de k e $c = F^{-1}_{\chi^2_{(k-1)}}(1-\alpha_0)$ teriam que ser recalculados...]

- **Região de rejeição de H_0** (para valores de T)

  Lidamos com um teste de ajustamento, donde a região de rejeição de H_0 é o intervalo à direita $W = (c, +\infty)$, onde

$$\boxed{c \quad = \quad F^{-1}_{\chi^2_{(k-\beta-1)}}(1-\alpha_0) = F^{-1}_{\chi^2_{(4-1)}}(1-0.01) \overset{tabela/calc.}{=} 11.34.}$$

0.5

- **Decisão**

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esp. sob H_0 E_i	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
1	$]0,2/3]$	13	6.73	$\frac{(13-6.73)^2}{6.73} \simeq 5.841$
2	$]2/3,2]$	32	33.73	0.089
3	$]2,3]$	13	16.80	0.860
4	$]3,+\infty[$	6	6.74	0.081
		$\sum_{i=1}^{k} o_i = n = 64$	$\sum_{i=1}^{k} E_i = n = 64$	$t = \sum_{i=1}^{k} \frac{(o_i-E_i)^2}{E_i} \simeq 6.871$

0.5

Uma vez que $t \simeq 6.871 \not\in W = (11.34, +\infty)$, não devemos rejeitar H_0 ao n.s. de $\alpha_0 = 1\%$ [nem a qualquer outro n.s. inferior a α_0].

Obrigada!

Questões?