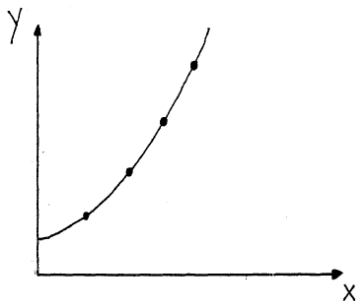


Econometria e Análise de Regressão

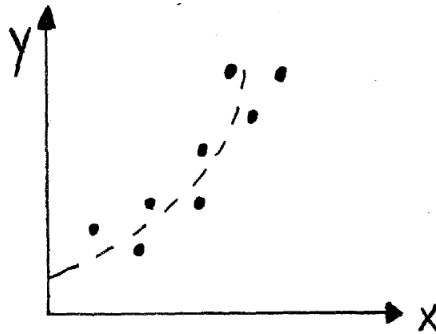
- O que é Econometria?
- Análise de Regressão
- Representação geral de uma relação funcional: $Y = f(X_1, X_2, \dots, X_k)$
- Exemplos de relações funcionais entre variáveis:
 - Variação do preço (Y) de um produto no mercado em função da quantidade produzida (X): $Y = f(X)$
 - Variação da produção (Y) obtida numa cultura em função da área plantada (X_1) e da quantidade de adubo utilizada (X_2): $Y = f(X_1, X_2)$

Modelo matemático e modelo estatístico

- Sejam duas variáveis, X e Y , relacionadas por uma função matemática:
 $Y = f(X)$
- Dado um conjunto de valores de X_j e Y_j ($j = 1, 2, \dots, n$), a representação dos pontos (X_j, Y_j) no eixo cartesiano é dada por:



- É comum, no entanto, que a variável dependente Y seja afetada por outros fatores, além dos considerados no modelo:
 - fatores desconhecidos;
 - fatores conhecidos, mas não mensuráveis (gostos, preferências, risco, incerteza).
- Subsiste, assim, um erro ou resíduo ao modelo (μ): $Y = f(X) + \mu$
- Nesse caso, ao representar o conjunto de pares (X_j, Y_j) no eixo cartesiano, obteremos:



- Outra justificativa para a existência do erro em um modelo estatístico é dada pela ocorrência de erros de mensuração na variável dependente Y

O modelo estatístico de uma regressão linear simples

- Se admitirmos que Y é função **linear** de X , podemos estabelecer uma regressão linear simples, cujo modelo estatístico é

$$Y = \alpha + \beta X + \mu$$

em que

Variáveis
aleatórias

- Y = variável dependente/explicada
- X = variável independente/explanatória *ou* regressor
- μ = termo de erro ou termo estocástico

Parâmetros

- α = termo constante ou intercepto
- β = coeficiente de regressão = *interpretação*

- Exemplo 1:

$$Y = \alpha + \beta X + \mu$$

Y = custo total de produção

X = quantidade produzida

α = *qual é o sentido econômico?*

β = *qual é o sentido econômico?*

- Exemplo 2:

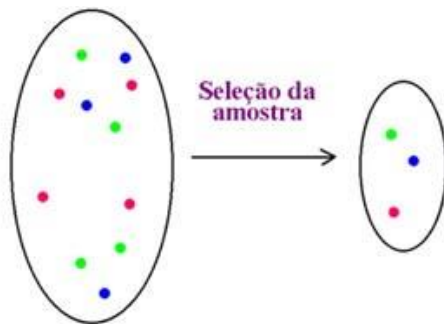
$$Y = \alpha + \beta X + \mu$$

Y = gasto ou consumo (em R\$)

X = renda disponível (em R\$)

Quanto valem α e β ???

- α e β são **parâmetros populacionais**. Se tivéssemos à disposição toda a população estudada, seria possível obtê-los.
- Posso **estimá-los** por meio de uma amostra aleatória extraída da população:



- As *estimativas* de α e β são indicadas por “a” e “b”

- Assim, tenho o modelo “teórico”:

$$Y = \alpha + \beta X + \mu$$

- E o modelo estimado (FRA):

$$\hat{Y} = a + bX$$

- Y é o valor *observado*, ao passo que \hat{Y} é o valor *previsto*

- Para cada observação da amostra pode-se calcular o desvio ou resíduo (e) entre o valor observado/real Y e o valor previsto \hat{Y} : $e = Y - \hat{Y}$

Estimativas dos parâmetros α e β pelo método dos mínimos quadrados ordinários (MQO)

- Seja o modelo $Y_j = \alpha + \beta X_j + \mu_j$ ①
- Se a e b forem, respectivamente, as estimativas de α e β , a reta de regressão estimada é $\hat{Y}_j = a + bX_j$ ②
- Para cada par de valores X_j e Y_j podemos calcular o desvio ou erro (e_j) entre o valor observado Y_j e o valor esperado \hat{Y}_j : $e_j = Y_j - \hat{Y}_j$ ③

- Substituindo \hat{Y}_j por ②: $e_j = Y_j - (a + bX_j)$

$$e_j = Y_j - a - bX_j$$

- Medida agregada dos desvios: $\sum_{j=1}^n e_j^2$ (soma dos quadrados dos desvios/resíduos)

- O método dos mínimos quadrados consiste em adotar como estimativa dos parâmetros os valores de α e β que minimizam:

$$Z = \sum_{j=1}^n e_j^2 = \sum_{j=1}^n (Y_j - a - bX_j)^2$$

- **Condição de 1ª ordem:** a função Z terá mínimo quando suas derivadas parciais em relação a a e b forem nulas.
- **Condição de 2ª ordem:** derivadas segundas em relação a a e b devem ser positivas.
- Desenvolvendo as condições de 1ª ordem obtemos:
 - i.
$$\frac{\partial Z}{\partial a} = -2 \sum_{j=1}^n (Y_j - a - bX_j) = 0$$
 - ii.
$$\frac{\partial Z}{\partial b} = -2 \sum_{j=1}^n (Y_j - a - bX_j)X_j = 0$$

- Do desenvolvimento algébrico dessas duas expressões obtemos:

$$a = \bar{Y} - b\bar{X}$$

$$b = \frac{\sum x_j y_j}{\sum x_j^2}$$

em que $x_j = (X_j - \bar{X})$ e $y_j = (Y_j - \bar{Y})$

Exemplo (ex. 1a da Lista 1) - É dada uma amostra de 10 pares de valores

X	Y
-2	0
-2	0
-1	2
-1	3
0	4
0	4
1	5
1	6
2	8
2	8

Lista 1: 1a, 3a, 4a,
9a, 9e, 10a

Admite-se que as variáveis X e Y estão relacionadas de acordo com o modelo $Y_i = \alpha + \beta X_i + \mu_i$, onde os μ_i são variáveis aleatórias independentes com distribuição normal de média zero e variância σ^2 . Determine as estimativas dos parâmetros da regressão linear.

Como sei se as estimativas que obtive são boas?

Critérios para saber se a estimativa é “boa”:

A estimativa será boa se o estimador for:

1. Linear

2. Não tendencioso ou não viesado

3. Eficiente

Melhor estimador Linear
Não Tendencioso (MELNT)
ou
Best Linear Unbiased
Estimator (BLUE)

1. Linear

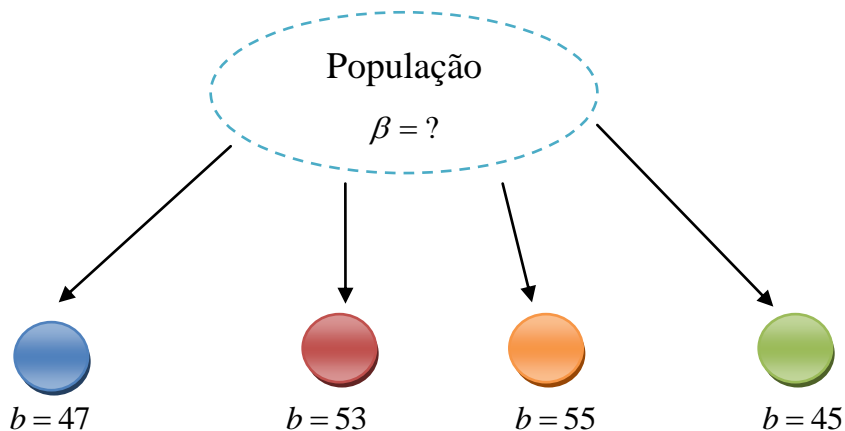
Parâmetro populacional: β = coeficiente de regressão

Amostra (n): $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$

Estimador linear: pode ser expresso como uma função linear dos dados da variável dependente: $b = w_1 Y_1 + w_2 Y_2 + \dots + w_n Y_n$

em que $w_i = f(X_i)$

2. Não tendencioso ou não viesado

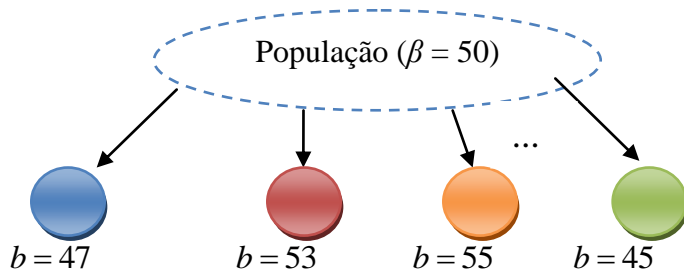


Não tendenciosidade ou ausência de viés: $E(b) = \beta$

3. Eficiente

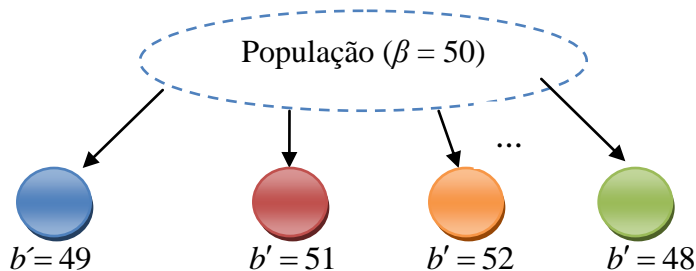
Eficiência: comparando todos os estimadores não tendenciosos de um mesmo parâmetro, “***o***” ***eficiente*** é o que tem ***menor variância***.

Estimador 1 (b) produz as seguintes estimativas:



Ambos são não tendenciosos, então, qual é o melhor???

Estimador 2 (b') produz as seguintes estimativas:



Melhor estimador Linear Não Tendencioso

Se um estimador for linear, não tendencioso e “o” eficiente, será o *melhor*: Melhor Estimador Linear Não Tendencioso (MELNT) ou Best Unbiased Linear Estimator (BLUE).

Sob quais condições o estimador de mínimos quadrados ordinários

$\left(b_{MQO} = \frac{\sum x_j y_j}{\sum x_j^2} \right)$ será o *melhor* estimador de β ?

Hipóteses do modelo de regressão linear simples

H.1 – A relação entre X e Y é linear, dada pela equação $Y = \alpha + \beta X + \mu$ (*hip. de linearidade nos parâmetros*).

H.2 – X não guarda correlação com o termo de erro do modelo (μ): $\text{cov}(X_j, \mu_j) = 0$, para todo $j = 1, 2, \dots, n$ (*hip. de exogeneidade do regressor*).

Isso significa X e μ exercem influência separada (independente) e cumulativa em Y .

Essa hip. é violada quando μ contém fatores que afetam Y e que também estão correlacionados com X . Isso geralmente acontece quando há variáveis omitidas no modelo.

$$Y = \alpha + \beta X + \mu$$

% de alunos
aprovados
em
Matemática

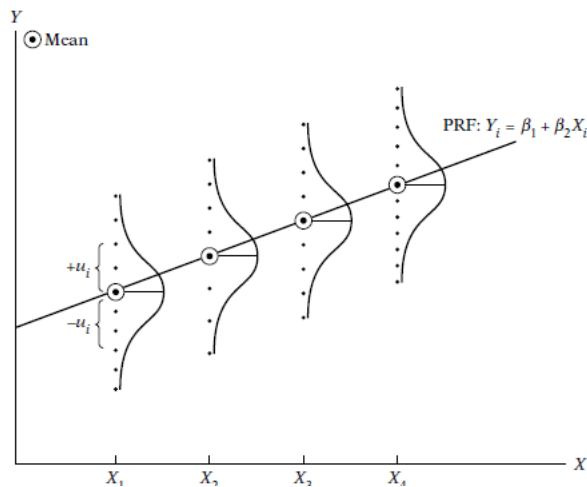
% de alunos aptos a
participar do programa de
merenda escolar

Contém
características da
escola e do
estudante que
afetam o
desempenho
escolar

$$\hat{Y} = 32,14 + -0,319X$$

(Woodridge meap93.raw)

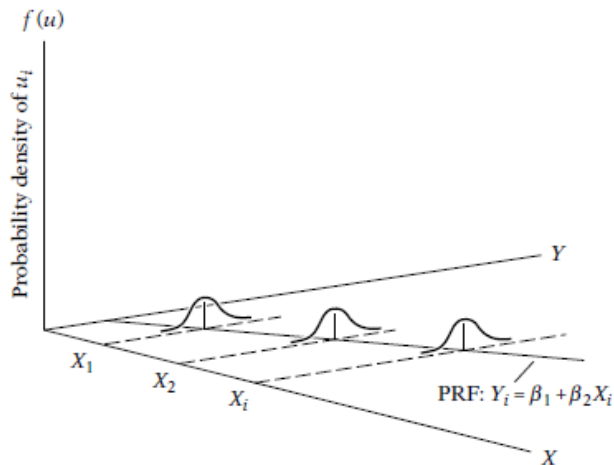
H.3 – Para um dado valor de X , o valor médio ou esperado do termo de erro (μ) é zero: $E(\mu_j|X_j)=0$, para todo $j=1,2,\dots,n$ (*hip. de média condicional zero*).



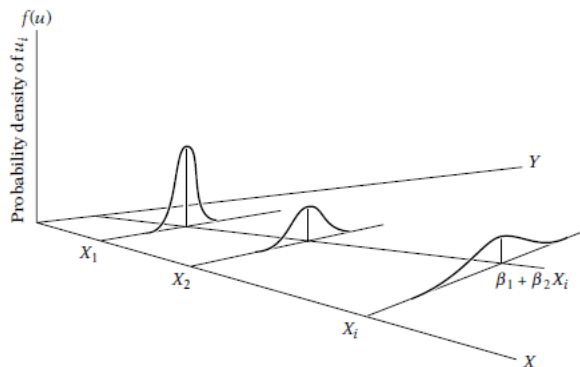
Isso significa que os fatores não incluídos explicitamente no modelo (incluídos, portanto, em μ), não afetam sistematicamente o valor médio de Y , pois os valores positivos de μ se anulam com os valores negativos. Desse modo, o efeito médio de μ sobre Y é nulo.

H.1 a H.3 são necessárias para demonstrar que os estimadores de mínimos quadrados ordinários de α e β são *não tendenciosos* e *consistentes*.

H.4 – Para um dado valor de X , a variância do termo de erro (μ) é constante: $V(\mu_j|X_j)=\sigma^2$ ou $V(Y_j|X_j)=\sigma^2$ (*hip. de homocedasticidade*).



De outra forma, a população dos Y dado X têm igual dispersão ou variância. Se a dispersão variar para os diversos valores fixados de X temos uma situação de heterocedasticidade:



É realista supor que a variabilidade do salário em torno de sua média é a mesma para todos os níveis educacionais?

H.5 – Dado dois valores de X quaisquer, X_i e X_j , a correlação entre quaisquer dois valores μ_i e μ_j ($i \neq j$) é zero: $\text{cov}(\mu_i, \mu_j | X_i, X_j) = 0$.

Em outras palavras, o erro associado a uma observação (μ_j) é não correlacionado com o erro em qualquer outra observação (μ_i).

Por exemplo, se houvesse uma correlação entre μ_2 e μ_1 , Y_2 não dependeria apenas de X_2 , mas de μ_1 , pois este também explica μ_2 .

Teorema de Gauss-Markov: sob as hipóteses H.1 a H.5, o estimador de MQO de α e β é o MELNT/BLUE.

- Se qualquer uma das hipóteses de Gauss-Markov for violada, o teorema não é mais válido.
- A heterocedasticidade não faz com que o MQO seja tendencioso. Entretanto, ele deixa de ser o MELNT, pois não tem mais a menor variância entre os estimadores lineares não-tendenciosos.

Lista 1: 8a

Decomposição da soma de quadrados total

Causas de Variação	Soma de Quadrados (SQ)
Regressão	$\sum \hat{y}_j^2$
Resíduo	$\sum e_j^2$
Total	$\sum y_j^2$

Essa tabela mostra que a variação dos valores de Y em torno de sua média (SQ Total = $\sum y_j^2$) pode ser dividida em duas partes: uma (SQ Regressão = $\sum \hat{y}_j^2$) que é “explicada” pela regressão e outra (SQ Residual = $\sum e_j^2$) devida ao fato de que nem todos os pontos estão sobre a reta de regressão, que é a parte não “explicada” pela regressão.

X	Y	$\hat{Y} = 4 + 1,9X$	$y = Y - \bar{Y}$	$\hat{y} = \hat{Y} - \bar{Y}$	y^2	\hat{y}^2
-2	0					
-2	0					
-1	2					
-1	3					
0	4					
0	4					
1	5					
1	6					
2	8					
2	8					
Somatório	40					

SQ da Regressão	$\sum \hat{y}_j^2 =$
SQ Residual	
SQ Total	$\sum y_j^2 =$

Maneira mais fácil de calcular a SQ Regressão: $\sum \hat{y}_j^2 = b \sum x_j y_j$

Assim:

Causas de Variação	Soma de Quadrados (SQ)	
Regressão	$b \sum x_j y_j$	=
Resíduo		=
Total	$\sum y_j^2$	=

$$\hat{Y} = 4 + 1,9X$$

$$\sum x^2 = 20$$

$$\sum y^2 = 74$$

$$\sum xy = 38$$

Coeficiente de determinação (R^2) do modelo

- $R^2 = \frac{\text{SQ Regressão}}{\text{SQ Total}}$
- Indica a proporção da variação de Y que é “explicada” pela regressão, isto é, pela variável X .
- $0 \leq R^2 \leq 1$ (quanto mais próximo de 1, melhor).
- Exemplo: $R^2 = \frac{\text{SQ Regressão}}{\text{SQ Total}} = \frac{72,2}{74,0} = 97,57\%$

Lista 1: 1c, 2b,
4b, 8c, 9d, 10b.

Análise de variância da regressão (ANOVA)

- Associamos às somas de quadrados (SQ) da regressão, residual e total k , $(n - k - 1)$ e $(n - 1)$ graus de liberdade, respectivamente.
- Os quadrados médios (QM) são obtidos dividindo-se as somas de quadrados pelos respectivos graus de liberdade.
- O QM Resíduos é uma estimativa não tendenciosa da variância do erro (σ^2).

Causas de Variação	Graus de liberdade (gl)	Soma de Quadrados (SQ)	Quadrados Médios (QM)
Regressão	$k =$	$b \sum xy =$	
Resíduos			$s^2 =$
Total	$n - 1 =$	$\sum y^2 =$	

k = número de variáveis explanatórias

H.6 – Os erros têm distribuição normal: $\mu_j \sim N$.

H.6 é necessária para que se possam utilizar as distribuições t e F para realizar testes de hipóteses a respeito dos valores dos parâmetros e para construir IC para os parâmetros.

Teste F

Pode-se demonstrar que, se os erros têm distribuição normal e se $\beta=0$, o quociente $\frac{\text{QM Regressão}}{\text{QM Resíduos}}$ tem distribuição F com 1 e $(n-2)$ gl. Então, para

testar $\begin{cases} H_0 : \beta = 0 \\ H_A : \beta \neq 0 \end{cases}$ podemos utilizar a estatística F .

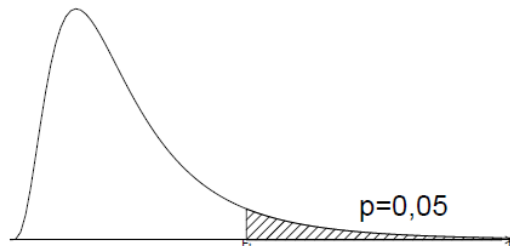
Exemplo

$$\begin{cases} H_0 : \beta = 0 \\ H_A : \beta \neq 0 \end{cases} \quad \text{ao} \quad ns = 5\%$$

$$F_{calc} = \frac{\text{QM Regressão}}{\text{QM Resíduos}} = \frac{72,2}{0,225} = 320,89$$

$$F_0 = ?$$

Regra de decisão: se $F_{calc} \geq F_0$,
rejeita-se H_0 .



Lista 1: 1b, 3b, 4c, 9b.

TABELA IV – Distribuição de F. Valor crítico F_0 tal que $P(F > F_0) = 0,05$

Nº de graus de liberdade do denominador	Número de graus de liberdade do numerador																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,46	19,47	19,48	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,66	8,64	8,62	8,59	8,57	8,55	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

Interpolações devem ser feitas com base nos recíprocos dos graus de liberdade (interpolação harmônica).

Fonte: Christ (1966, p. 670) e Pimentel Gomes (1966, p. 406-407).

	1	2	3	4	5	6	7	8	9
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49

Nesse caso, $F_{calc} = 320,89 > 5,32$, logo, rejeita-se H_0 . **O que isso significa?**

Testes de hipóteses sobre β

Sob H.1 a H.6, o quociente $t_b = \frac{b - \beta}{s(b)}$ tem distribuição t de Student com $(n - 2)$ gl. Assim, t_b pode ser utilizado para testar hipóteses sobre β .

Exemplo: $\begin{cases} H_0 : \beta = 2 \\ H_A : \beta \neq 2 \end{cases}$ ao $ns = 5\%$

$$t_b = \frac{b - E(b)}{s(b)} = \frac{b - \beta}{s(b)}$$

$$s(b) = \sqrt{V(b)} \quad \text{e}$$

$$V(b) = \frac{\sigma^2}{\sum x^2}$$

Quanto maior σ^2 , maior $V(b)$, pois uma variação maior nos fatores não-observáveis que afetam Y faz com que seja mais difícil estimar com precisão β .

Quanto maior a variabilidade em X , menor $V(b)$, pois será mais fácil estimar com precisão a relação entre $E(Y | X)$ e X .

Como não disponho de σ^2 , utilizo em seu lugar $s^2 = \text{QM Resíduos}$: $\hat{V}(b) = \frac{s^2}{\sum x^2}$

$$\text{Logo, } \hat{V}(b) = \frac{0,225}{20} = 0,01125$$

$$t_b = \frac{b - \beta}{s(b)} = \frac{1,9 - 2}{\sqrt{0,01125}} = -0,94$$

$$RC = \{(-\infty; -t_0) \cup (t_0; +\infty)\}$$

Regra de decisão: se $t_b \in RC$, rejeita-se H_0 .

Apêndice

TABELA I. Distribuição de t de Student. Valor crítico t_0 tal que

$$P(t > t_0) = P(t < t_0) = \alpha/2$$

Número de Graus de Liberdade	Nível de significância para o teste bilateral (α)					
	0,20	0,10	0,05	0,02	0,01	0,005
1	3,078	6,314	12,706	31,821	63,657	127,32
2	1,886	2,920	4,303	6,965	9,925	14,089
3	1,638	2,353	3,182	4,541	5,841	7,453
4	1,533	2,132	2,776	3,747	4,604	5,598
5	1,476	2,015	2,571	3,365	4,032	4,773
6	1,440	1,943	2,447	3,143	3,707	4,317
7	1,415	1,895	2,365	2,998	3,499	4,029
8	1,397	1,860	2,306	2,896	3,355	3,832
9	1,383	1,833	2,262	2,821	3,250	3,690
10	1,372	1,812	2,228	2,764	3,169	3,581
11	1,363	1,796	2,201	2,718	3,106	3,497
12	1,356	1,782	2,179	2,681	3,055	3,428
13	1,350	1,771	2,160	2,650	3,012	3,372
14	1,345	1,761	2,145	2,624	2,977	3,326
15	1,341	1,753	2,131	2,602	2,947	3,286

Equivalência entre os testes t e F

Quando testamos $\begin{cases} H_0 : \beta = 0 \\ H_A : \beta \neq 0 \end{cases}$, a estatística $F_{calc} = t_b^2$ e $F_0 = t_0^2$

Equivalência entre os valores calculados:

$$F_{calc} = \frac{\text{QMRegressão}}{\text{QMResíduos}} = \frac{72,2}{0,225} = 320,89$$

$$t_b = \frac{b - \beta}{s(b)} = \frac{1,9 - 0}{\sqrt{0,01125}} = 17,9134$$

Observa-se que $17,9134^2 = 320,89$

Equivalência entre os valores críticos:

$$F_0[5\%, 1 \text{ e } 8 \text{ gl}] = 5,32$$

$$t_0[5\%, 8 \text{ gl}] = 2,306$$

Observa-se que $2,306^2 = 5,32$

Atenção:

Teste F : - exclusivo para $\begin{cases} H_0 : \beta = 0 \\ H_A : \beta \neq 0 \end{cases}$

Teste t : - admite outros valores no lado direito da igualdade: $H_0 : \beta = \theta$
- admite outros formatos da H_A : $H_A : \beta > \theta$ ou $H_A : \beta < \theta$

(Anpec 2009, Questão 10) Com relação aos testes de hipótese, é correto afirmar:

- Considere o seguinte modelo de regressão linear: $y = \beta_0 + \beta_1 X + \mu$, em que μ é o erro da regressão, y é a variável dependente e X é a variável explicativa. Para testarmos a hipótese $H_0 : \beta_1 = 0$ contra a alternativa $H_1 : \beta_1 > 0$, devemos utilizar um teste t unilateral.

Teste de hipóteses sobre α

Sob H.1 a H.6, o quociente $t_a = \frac{a - \alpha}{s(a)}$ têm distribuição t de Student com $(n - 2)$

gl. Assim, t_a pode ser utilizado para testar hipóteses sobre α .

Exemplo: $\begin{cases} H_0 : \alpha = 5 \\ H_A : \alpha < 5 \end{cases}$ ao $ns = 5\%$

$$t_a = \frac{a - \alpha}{s(a)} = \frac{4 - 5}{s(a)}$$

$$s(a) = \sqrt{V(a)}$$

$$V(a) = \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x^2} \right) \sigma^2$$

Como não disponho de σ^2 , utilizo em seu lugar $s^2 = \text{QM Resíduos}$

$$\hat{V}(a) = \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x^2} \right) s^2$$

$$\hat{V}(a) = \left(\frac{1}{10} + \frac{0^2}{20} \right) 0,225 = 0,0225$$

$$t_a = \frac{a - \alpha}{s(a)} = \frac{4 - 5}{s(a)} = \frac{4 - 5}{\sqrt{0,0225}} = -6,67$$

$$RC = \{-\infty; -t_0\}$$

Regra de decisão: se $t_a \in RC$, rejeita-se H_0

Número de Graus de Liberdade	Nível de significância para o teste <u>bilateral</u> (α)					
	0,20	0,10	0,05	0,02	0,01	0,005
1	3,078	6,314	12,706	31,821	63,657	127,32
2	1,886	2,920	4,303	6,965	9,925	14,089
3	1,638	2,353	3,182	4,541	5,841	7,453
4	1,533	2,132	2,776	3,747	4,604	5,598
5	1,476	2,015	2,571	3,365	4,032	4,773
6	1,440	1,943	2,447	3,143	3,707	4,317
7	1,415	1,895	2,365	2,998	3,499	4,029
8	1,397	1,860	2,306	2,896	3,355	3,832
9	1,383	1,833	2,262	2,821	3,250	3,690
10	1,372	1,812	2,228	2,764	3,169	3,581
11	1,363	1,796	2,201	2,718	3,106	3,497
12	1,356	1,782	2,179	2,681	3,055	3,428
13	1,350	1,771	2,160	2,650	3,012	3,372
14	1,345	1,761	2,145	2,624	2,977	3,326

Intervalo de Confiança para α e β

IC para β : $b \pm \underbrace{t_0 \cdot s(b)}_{\text{Erro padrão de } b}$ a C% e $(n-2)$ gl

IC para α : $a \pm \underbrace{t_0 \cdot s(a)}_{\text{Erro padrão de } a}$ a C% e $(n-2)$ gl



Exemplo: construir um IC para β adotando um coeficiente de confiança C de 98%

$$\bar{b} \pm t_0 \cdot \underbrace{s(\bar{b})}_{\sqrt{0,01125}}$$



t_0

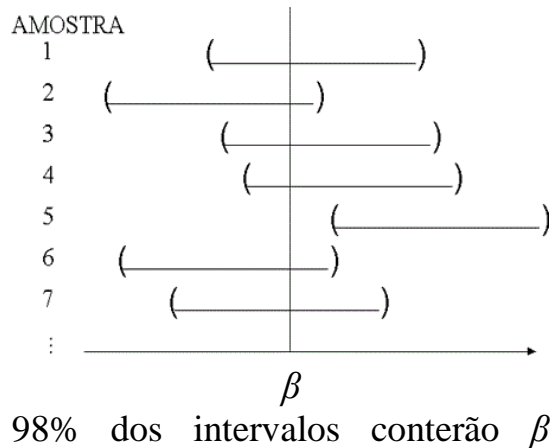
- se $C = 98\%$, então $ns=2\%$.
- $gl = 8$

$$1,9 \pm 2,896\sqrt{0,01125}$$

$$1,9 \pm 0,307$$

$$[1,59; 2,21] \quad \text{ou} \quad 1,59 < \beta < 2,21$$

Interpretação:



Número de Graus de Liberdade	Nível de significância para o teste bilateral (α)					
	0,20	0,10	0,05	0,02	0,01	0,005
1	3,078	6,314	12,706	31,821	63,657	127,32
2	1,886	2,920	4,303	6,965	9,925	14,089
3	1,638	2,353	3,182	4,541	5,841	7,453
4	1,533	2,132	2,776	3,747	4,604	5,598
5	1,476	2,015	2,571	3,365	4,032	4,773
6	1,440	1,943	2,447	3,143	3,707	4,317
7	1,415	1,895	2,365	2,998	3,499	4,029
8	1,397	1,860	2,306	2,896	3,355	3,832
9	1,383	1,833	2,262	2,821	3,250	3,690
10	1,372	1,812	2,228	2,764	3,169	3,581
11	1,363	1,796	2,201	2,718	3,106	3,497
12	1,356	1,782	2,179	2,681	3,055	3,428
13	1,350	1,771	2,160	2,650	3,012	3,372
14	1,345	1,761	2,145	2,624	2,977	3,326

Intervalo de Confiança para uma previsão de \hat{Y}

Exemplo

Determine a estimativa de Y para $X = 3$ e o respectivo intervalo de confiança ao nível de confiança de 95%.

X_j	Y_j
-2	0
-2	0
-1	2
-1	3
0	4
0	4
1	5
1	6
2	8
2	8

$$\hat{Y} = 4 + 1,9X$$

$$\text{Para } X_h = 3, \hat{Y}_h = 9,7$$

$$\text{IC para } \hat{Y}_h : \hat{Y}_h \pm t_0 \sqrt{\left(\frac{1}{n} + \frac{x_h^2}{\sum x^2} \right) s^2}$$

$$t_0 = ?$$

- $C = 95\%$, $ns = 5\%$.
- $gl = (n - 2) = 8$

Número de Graus de Liberdade	Nível de significância para o teste bilateral (α)					
	0,20	0,10	0,05	0,02	0,01	0,005
1	3,078	6,314	12,706	31,821	63,657	127,32
2	1,886	2,920	4,303	6,965	9,925	14,089
3	1,638	2,353	3,182	4,541	5,841	7,453
4	1,533	2,132	2,776	3,747	4,604	5,598
5	1,476	2,015	2,571	3,365	4,032	4,773
6	1,440	1,943	2,447	3,143	3,707	4,317
7	1,415	1,895	2,365	2,998	3,499	4,029
8	1,397	1,860	2,306	2,896	3,355	3,832
9	1,383	1,833	2,262	2,821	3,250	3,690
10	1,372	1,812	2,228	2,764	3,169	3,581
11	1,363	1,796	2,201	2,718	3,106	3,497
12	1,356	1,782	2,179	2,681	3,055	3,428
13	1,350	1,771	2,160	2,650	3,012	3,372
14	1,345	1,761	2,145	2,624	2,977	3,326

$$9,7 \pm 2,306 \sqrt{\left(\frac{1}{10} + \frac{3^2}{20}\right) 0,225}$$

$$9,7 \pm 0,811$$

$$8,89 < \hat{Y}_h < 10,51$$

Aplicação

Gretl: <http://gretl.sourceforge.net/>

Use os dados em SLEEP75.RAW para estudar se há um *tradeoff* entre o tempo gasto dormindo por semana e o tempo gasto em um trabalho pago. Estime o modelo

$$sleep = \alpha + \beta totwrk + \mu$$

em que *sleep* corresponde a minutos gastos dormindo à noite por semana, e *totwrk* é o total de minutos trabalhados durante a semana.

- (i) Reporte seus resultados na forma de equação, juntamente com o número de observações e R^2 . O que o intercepto significa nessa equação?
- (ii) Por que razão você acha que foi obtido um valor tão baixo para o R^2 ?
- (iii) Se *totwrk* aumentar em duas horas, em quanto se estima que *sleep* irá cair?

Modelo 1: MQO, usando as observações 1-706
Variável dependente: sleep

Se p-valor > ns,
não se rejeita H_0

	<i>Coefficiente</i>	<i>Erro Padrão</i>	<i>razão-t</i>	<i>p-valor</i>	
const	3586,38	38,9124	92,1653	<0,00001	***
totwrk	-0,150746	0,0167403	-9,0050	<0,00001	***

Média var. dependente	3266,356	D.P. var. dependente	444,4134
Soma resíd. quadrados	1,25e+08	E.P. da regressão	421,1357
R-quadrado	0,103287	R-quadrado ajustado	0,102014
F(1, 704)	81,08987	P-valor(F)	1,99e-18

(ANPEC 2014, Questão 01) Neste exemplo, queremos prever o peso do indivíduo i usando somente sua altura,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

no qual Y é o peso do indivíduo e X a altura. Assumimos que $(Y_i, X_i)_{i=1}^N$ é uma amostra aleatória, $E[\varepsilon_i|X_i] = 0$, $Var[X_i] > 0$, $E[X_i^4] < \infty$, $0 < E[u_i^4] < \infty$ e $Var[\varepsilon_i|X_i] = \sigma_\varepsilon^2$. Após coletar a informação de peso e altura de 100 indivíduos, obtemos a seguinte tabela:

$\sum_{i=1}^N Y_i$	$\sum_{i=1}^N X_i$	$\sum_{i=1}^N (Y_i - \bar{Y})^2$	$\sum_{i=1}^N (X_i - \bar{X})^2$	$\sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X})$
18	8	95	1200	4800

Estimando o modelo por Mínimos Quadrados Ordinários, calcule o valor da estimativa obtida para $\hat{\beta}_1$. Multiplique o resultado por 10.