## DEPARTAMENTO DE MATEMÁTICA



# Probabilidades e Estatística

## LEGM, LEIC-A, LEIC-T, LEMat, MEBiom, MEFT, MEQ

 $2^{o}$  semestre – 2011/2012

2º Teste A

(3.0)

08/06/2012 - 9:00

Duração: 1 hora e 30 minutos

Justifique convenientemente todas as respostas!

### Grupo I

3.0 + 2.0 + 2.5 + 2.5 = 10 valores

- 1. O número de carros que passam em cada período de um minuto num determinado ponto de uma autoestrada é uma variável aleatória com distribuição de Poisson de valor esperado  $\lambda$ . Tendo por base a contabilização do número de automóveis que passam nesse ponto da autoestrada em cada um de 100 intervalos de um minuto, seleccionados ao acaso:
  - (a) Deduza o estimador de máxima verosimilhança de  $\lambda$ . Será que o estimador é centrado?
    - V.a. de interesse

X = número de carros que passam em cada período de um minuto...

• Distribuição

 $X \sim \text{Poisson}(\lambda)$ 

• Parâmetro DESCONHECIDO

$$\lambda = E(X) = V(X), \ \lambda > 0$$

$$P(X = x) \stackrel{form}{=} \frac{e^{-\lambda} \lambda^x}{x!}, x = 0, 1, 2, \dots$$

• Amostra

 $\underline{x} = (x_1, \dots, x_n)$  amostra de dimensão n proveniente da população X

• Obtenção do estimador de MV de  $\lambda$ 

So I — Função de Verosiminança 
$$L(\lambda|\underline{x}) \stackrel{X_i \, indep}{=} \prod_{i=1}^n P(X_i = x_i)$$
 
$$\stackrel{X_i \simeq X}{=} \prod_{i=1}^n \left(\frac{e^{-\lambda} \, \lambda^{x_i}}{x_i!}\right)$$
 
$$= \frac{e^{-n\lambda} \, \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}, \, \lambda > 0$$

Passo 2 — Função de log-verosimilhança

$$\ln L(\lambda | \underline{x}) = -n\lambda + \ln(\lambda) \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!)$$

Passo 3 — Maximização<sup>1</sup>

A estimativa de MV de  $\lambda$  é aqui representada por  $\lambda$  e

timativa de MV de 
$$\lambda$$
 é aqui representada por  $\hat{\lambda}$  e 
$$\hat{\lambda} : \begin{cases} \frac{d \ln L(\lambda | \underline{x})}{d \lambda} \Big|_{\lambda = \hat{\lambda}} = 0 & \text{(ponto de estacionaridade)} \\ \frac{d^2 \ln L(\lambda | \underline{x})}{d \lambda^2} \Big|_{\lambda = \hat{\lambda}} < 0 & \text{(ponto de máximo)} \end{cases}$$
$$\begin{cases} -n + \frac{\sum_{i=1}^n x_i}{\hat{\lambda}} = 0 \\ -\frac{\sum_{i=1}^n x_i}{\hat{\lambda}^2} < 0 \end{cases}$$
$$\begin{cases} \hat{\lambda} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}, \text{ média da amostra} \end{cases}$$
Proposição verdadeira

<sup>&</sup>lt;sup>1</sup>Este procedimento só deve ser aplicado se  $\sum_{i=1}^{n} x_i \neq 0$ , sendo o resultado obtido,  $\hat{\lambda} = \bar{x}$ , também válido se  $\sum_{i=1}^{n} x_i = 0$ .

## Passo 4 — Estimador de MV de $\lambda$

Será representado pela v.a.  $\text{EMV}(\lambda) = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}$ , média da amostra aleatória.

## • Estimador de MV de $\lambda$ é centrado?

 $\mathrm{EMV}(\lambda) = \bar{X}$  é um estimador centrado de  $\lambda$  sse  $E(\bar{X}) = \lambda, \forall \lambda > 0$ . Ora,

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n}X_{i}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}E(X_{i})$$

$$X_{i} \stackrel{\sim}{=} X \quad \frac{1}{n}\sum_{i=1}^{n}E(X)$$

$$= E(X)$$

$$= \lambda,$$

 $\forall \lambda > 0$ . Assim, conclui-se que  $\bar{X}$  é um estimador centrado de  $\lambda$ .

- (b) Sabendo que foi contabilizada a passagem de um total de 950 carros no referido ponto da (2.0) autoestrada no conjunto dos 100 intervalos de um minuto seleccionados, construa um intervalo, com nível de confiança de aproximadamente 95%, para o parâmetro λ.
  - V.a.

 $X_i=$ número de carros que passam no i- ésimo período de um minuto,  $i=1,\dots,100$   $X_i\stackrel{i.i.d.}{\sim} X$ 

### • Situação

 $X \sim \text{Poisson}(\lambda)$ 

 $\lambda$  desconhecido

n = 100 >> 30 (suficientemente grande)

### • Obtenção de IC para $\lambda$

– Passo 1 — Selecção da v.a. fulcral para  $\lambda$ 

Utilizaremos a v.a. fulcral para  $\lambda$ 

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{\mathrm{EMV}[V(\bar{X})]}} = \frac{\bar{X} - \lambda}{\sqrt{\mathrm{EMV}(\lambda/n)}} = \frac{\bar{X} - \lambda}{\sqrt{\frac{\bar{X}}{n}}} \stackrel{a}{\sim} \mathrm{normal}(0, 1),$$

uma vez que nos foi solicitada a determinação de um IC aproximado para o parâmetro do modelo de Poisson e a dimensão da amostra justifica o recurso a uma aproximação distribucional.

### - Passo 2 — Obtenção dos quantis de probabilidade

Dado que  $(1 - \alpha) \times 100\% = 95\% \Leftrightarrow \alpha = 0.05$ , os quantis a utilizar são

$$\begin{cases} a_{\alpha} = -\Phi^{-1}(1 - \alpha/2) = -\Phi^{-1}(0.975) \stackrel{tabela}{=} -1.9600 \\ b_{\alpha} = \Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.975) = 1.9600. \end{cases}$$

Estes enquadram a v.a. fulcral para  $\lambda$  com probabilidade aproximadamente igual a  $(1-\alpha)$ .

## – Passo 3 — Inversão da desigualdade $a_{\alpha} \leq Z \leq b_{\alpha}$

$$\begin{split} &P(a_{\alpha} \leq Z \leq b_{\alpha}) \simeq 1 - \alpha \\ &P\left(a_{\alpha} \leq \frac{\bar{X} - \lambda}{\sqrt{\frac{\bar{X}}{n}}} \leq b_{\alpha}\right) \simeq 1 - \alpha \\ &P\left[\bar{X} - b_{\alpha} \times \sqrt{\frac{\bar{X}}{n}} \leq \lambda \leq \bar{X} - a_{\alpha} \times \sqrt{\frac{\bar{X}}{n}}\right] \simeq 1 - \alpha \\ &P\left[\bar{X} - \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{\bar{X}}{n}} \leq \lambda \leq \bar{X} + \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{\bar{X}}{n}}\right] \simeq 1 - \alpha. \end{split}$$

## - Passo 4 — Concretização

Ao ter-se em consideração que

$$* n = 100$$

\* 
$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{950}{100} = 9.5$$

\* 
$$\Phi^{-1}(1 - \alpha/2) = 1.9600$$
,

conclui-se que o IC aproximado a 95% para  $\lambda$  é dado por

$$IC(\lambda) = \left[ \bar{x} \pm \Phi^{-1} (1 - \alpha/2) \times \sqrt{\frac{\bar{x}}{n}} \right]$$
  
=  $\left[ 9.5 \pm 1.9600 \times \sqrt{\frac{9.5}{100}} \right]$   
\(\simeq \text{[8.895887, 10.104113].}

- 2. Uma máquina produz peças cujo comprimento, X, é uma variável aleatória com distribuição normal, de parâmetros  $\mu=20$  cm e  $\sigma=1$  cm se a máquina está afinada. De modo a controlar a produção, um operador da máquina seleccionou ao acaso 10 peças que conduziram aos seguintes resultados:  $\sum_{i=1}^{10} x_i = 211.2 \text{ e} \sum_{i=1}^{10} (x_i \bar{x})^2 = 5.95.$ 
  - (a) Considerando  $\sigma$  desconhecido, diga o que pode concluir sobre a hipótese de  $\mu=20$  cm, ao nível (2.5) de significância de 2%.

### • V.a. de interesse

X = comprimento de uma peça

### • Situação

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

## • Hipóteses

 $H_0: \mu = \mu_0 = 20$  $H_1: \mu \neq \mu_0 = 20$ 

## • Nível de significância

$$\alpha_0 = 0.02$$

### • Estatística de teste

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

pois pretendemos efectuar um teste sobre o valor esperado de uma população normal com variância desconhecida.

## $\bullet$ Região de rejeição de $H_0$ (para valores da estatística de teste)

Tratando-se de um teste bilateral  $(H_1: \mu \neq \mu_0)$ , a região de rejeição de  $H_0$  é uma reunião de intervalos do tipo  $W = (-\infty, -c) \cup (c, +\infty)$ , onde  $c: P(\text{Rejeitar } H_0 | \mu = \mu_0) = \alpha_0$ , i.e.,

$$c = F_{t_{(n-1)}}^{-1} (1 - \alpha_0/2)$$

$$= F_{t_{(10-1)}}^{-1} (1 - 0.02/2)$$

$$= F_{t_{(9)}}^{-1} (0.99)$$

$$\stackrel{tabela}{=} 2.821.$$

### • Decisão

Uma vez que

$$\begin{split} n &= 10 \\ \bar{x} &= \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{211.2}{10} = 21.12 \\ s^2 &= \frac{1}{n-1} \sum_{i=1}^{n} (xi - \bar{x})^2 = \frac{5.95}{10-1} \simeq 0.66, \end{split}$$

o valor observado da estatística é igual a

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$\simeq \frac{21.12 - 20}{\frac{\sqrt{0.66}}{\sqrt{10}}}$$

$$\simeq 4.36.$$

Como  $t \simeq 4.36 \in W = (-\infty, -2.821) \cup (2.821, +\infty)$ , devemos rejeitar  $H_0$  (hipótese do fabricante) a qualquer n.s. maior ou igual a 2%.

- (b) Teste, ao nível de significância de 5%, a hipótese da variabilidade do comprimento das peças (2.5) produzidas pela máquina ser igual ao esperado ( $\sigma = 1$  cm) contra a alternativa de ser superior ao esperado ( $\sigma > 1$  cm).
  - Situação

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

• Hipóteses

$$H_0: \sigma = \sigma_0 = 1$$
  
$$H_1: \sigma > \sigma_0 = 1$$

• Nível de significância

$$\alpha_0 = 0.05$$

• Estatística de teste

$$T = \frac{(n-1)S^2}{\sigma_0^2} \sim_{H_0} \chi_{(n-1)}^2$$

pois pretendemos efectuar um teste sobre a variância de uma população normal com valor esperado desconhecido.

• Região de rejeição de  $H_0$  (para valores da estatística de teste)

Tratando-se de mais um teste unilateral superior  $(H_1: \sigma > \sigma_0)$ , a região de rejeição de  $H_0$  é um intervalo à direita do tipo  $W = (c, +\infty)$ , onde  $c: P(\text{Rejeitar } H_0 | \sigma = \sigma_0) = \alpha_0$ , i.e.,

$$c = F_{\chi^{2}_{(n-1)}}^{-1}(1 - \alpha_{0})$$

$$= F_{\chi^{2}_{(9)}}^{-1}(0.95)$$

$$\stackrel{tabela}{=} 16.92.$$

• Decisão

Dado que  $n=10,\,\sum_{i=1}^n(x_i-\bar{x})^2=5.85$  e  $\sigma_0=1,$  o valor observado da estatística é igual a

to que 
$$n = 10$$
,  $\sum_{i=1}^{n} (x_i - x_i)^2$   

$$t = \frac{(n-1)s^2}{\sigma_0^2}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sigma_0^2}$$

$$= \frac{5.95}{1}$$

$$= 5.95.$$

Dado que  $t=5.95 \notin W=(16.92,+\infty)$ , não devemos rejeitar  $H_0$  a qualquer n.s. menor ou igual a 5%.

1. As medições de uma enzima, referentes a uma amostra casual de 100 pacientes que sofrem de hepatite viral aguda, encontram-se agrupadas em classes na tabela seguinte:

Classe	]2.4, 2.5]	]2.5, 2.6]	]2.6, 2.7]	]2.7, 2.8]
Número de pacientes	14	38	36	12

Será que uma distribuição normal, com valor esperado 2.6 e desvio padrão 0.1, se ajusta bem a estes dados? Teste esta hipótese com base no valor-p.

### • V.a. de interesse

X = medição de uma enzima em paciente com hepatite viral aguda

### • Hipóteses

 $H_0: X \sim \text{Normal}(2.6, 0.1^2)$  $H_1: X \nsim \text{Normal}(2.6, 0.1^2)$ 

### • Estatística de Teste

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

onde:

k = No. de classes:

 $O_i$  = Frequência absoluta observável da classe i;

 $E_i$  = Frequência absoluta esperada, sob  $H_0$ , da classe i;

 $\beta = \text{No. de parâmetros a estimar} = 0.$ 

### • Região de rejeição de $H_0$ (para valores de T)

Por estar a efectuar-se um teste de ajustamento, a região de rejeição de  $H_0$  escrita para valores de T é um intervalo à direita  $W=(c,+\infty)$ .

## $\bullet$ Frequências absolutas esperadas sob $H_0$

Para já, note-se que o conjunto de valores possíveis da distribuição Normal(2.6, 0.1²) é  $I\!\!R$ ; daí que se passe a considerar duas novas classes: os intervalos ]  $-\infty$ , 2.4] e ]2.8,  $+\infty$ [, respectivamente, com frequência observada 0. Se para além disso atendermos a que  $F_{X|H_0}(x) = \Phi\left(\frac{x-2.6}{0.1}\right)$ , as frequências absolutas esperadas sob  $H_0$ ,  $E_i = n\,p_i^0 = n\times P(X\in {\rm classe}\;i\mid H_0)$ , são, para i=1,2, iguais a

$$E_{1} = n \times P \left[ X \in ]-\infty, 2.4 \right] \mid X \sim \text{Normal}(2.6, 0.1^{2}) \right]$$

$$= 100 \times \Phi \left( \frac{2.4 - 2.6}{0.1} \right)$$

$$= 100 \times \Phi(-2)$$

$$= 100 \times [1 - \Phi(2)]$$

$$\stackrel{tabela}{=} 100 \times (1 - 0.9772)$$

$$= 100 \times 0.0228$$

$$= 2.28$$

$$E_{2} = n \times P \left[ X \in ]2.4, 2.5 \right] \mid X \sim \text{Normal}(2.6, 0.1^{2}) \right]$$

$$= 100 \times \left[ \Phi \left( \frac{2.5 - 2.6}{0.1} \right) - \Phi \left( \frac{2.4 - 2.6}{0.1} \right) \right]$$

$$= 100 \times [\Phi(-1) - \Phi(-2)]$$

$$= 100 \times (0.1587 - 0.0228)$$

$$= 100 \times 0.1359$$

$$= 13.59.$$

$$E_{3} = n \times P \left[ X \in ]2.5, 2.6 \right] \mid X \sim \text{Normal}(2.6, 0.1^{2}) \right]$$

$$= 100 \times \left[ \Phi \left( \frac{2.6 - 2.6}{0.1} \right) - \Phi \left( \frac{2.5 - 2.6}{0.1} \right) \right]$$

$$= 100 \times \left[ \Phi(0) - \Phi(-1) \right]$$

$$= 100 \times (0.5 - 0.1587)$$

$$= 100 \times 0.3413$$

$$= 34.13.$$

E, por simetria da f.d.p. da Normal $(\mu, \sigma^2)$  em torno de  $\mu = 2.6$ , obtêm-se as restantes frequências esperadas sob  $H_0$ :

$$E_4 = E_3 = 34.13,$$
  $E_5 = E_2 = 13.59,$   $E_6 = E_1 = 2.28.$ 

É necessário agrupar classes uma vez que se verifica  $E_i \geq 5$  em menos de 80% das classes (em todas elas tem-se  $E_i \geq 1$ ). Agrupem-se as duas classes que violam essa condição, a  $1^{\underline{a}}$  e a  $6^{\underline{a}}$ : ficando a nova classe igual a  $]-\infty,2.4]\cup]2.8,+\infty[$  com frequência observada 0 e frequência esperada sob  $H_0$  igual a 4.56.

#### • Decisão

No cálculo do valor observado da estatística de teste convém adiantar a seguinte tabela auxiliar.

i	Classe $i$	Freq. abs. obs. $o_i$	Freq. abs. esper. sob $H_0$ $E_i = n \times p_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i-E_i)^2}{E_i}$
$1_{nova} = 1 + 6$	$]-\infty, 2.4] \cup ]2.7, +\infty[$	0	4.56	$\frac{\frac{(0-4.56)^2}{4.56}}{\frac{(14-13.59)^2}{13.59}} = 4.56$
2	]2.4, 2.5]	14	13.59	$\frac{(14-13.59)^2}{13.59} = 0.012369$
3	]2.5, 2.6]	38	34.13	0.438819
4	]2.6, 2.7]	36	34.13	0.102458
5	]2.7, 2.8]	12	13.59	0.186026
		$\sum_{i=1}^{k} o_i = n = 100$	$\sum_{i=1}^{k} E_i = n = 100$	$t = \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i} = 5.2997$

## • Decisão (com base em intervalo para o valor-p)

Uma vez que este teste está associado a uma região de rejeição que é um intervalo à direita temos:

$$valor - p = P(T > t \mid H_0)$$
  
=  $P(T > 5.2997 \mid H_0)$   
 $\simeq 1 - F_{\chi^2_{(5-0-1)}}(5.2997).$ 

Recorrendo às tabelas de quantis da distribuição do qui-quadrado podemos adiantar um intervalo para o valor-p deste teste. Com efeito, ao enquadrarmos convenientemente t=5.2997, obtemos sucessivamente

$$\begin{split} F_{\chi^2_{(4)}}^{-1}(0.70) &= 4.045 &< 5.2997 < 5.989 = F_{\chi^2_{(4)}}^{-1}(0.80) \\ &0.70 &< F_{\chi^2_{(4)}}(5.2997) < 0.80 \\ &0.20 = 1 - 0.80 &< valor - p < 1 - 0.70 = 0.30. \end{split}$$

Logo:

- não devemos rejeitar  $H_0$  a qualquer n.s.  $\alpha_0 \le 20\%$ , por exemplo, a qualquer dos níveis usuais de significância de 1%, 5% e 10%;
- devemos rejeitar  $H_0$  a qualquer n.s.  $\alpha_0 \geq 30\%$ .
- Alternativa Decisão (com base no valor-p determinado usando máquina de calcular)

Uma vez que este teste está associado a uma região de rejeição que é um intervalo à direita temos:

$$valor - p = P(T > t \mid H_0)$$

$$= P(T > 5.2997 \mid H_0)$$

$$\simeq 1 - F_{\chi^2_{(5-0-1)}}(5.2997)$$

$$= 0.257905.$$

## Consequentemente:

- não devemos rejeitar  $H_0$  a qualquer n.s.  $\alpha_0 \le 25.7905\%$ , por exemplo, a qualquer dos níveis usuais de significância de 1%, 5% e 10%;
- devemos rejeitar  $H_0$  a qualquer n.s.  $\alpha_0 > 25.7905\%$ .
- 2. Num estudo sobre segurança rodoviária, pretende-se analisar a influência da velocidade a que um veículo pesado se desloca (x, em metros por segundo) sobre a distância percorrida pelo veículo após o início da travagem (Y, em metros), denominada distância de travagem.

Considere o modelo de regressão linear simples,  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$  (i = 1, ..., n), com as hipóteses de trabalho habituais, e que as observações relativas a n = 52 veículos pesados conduziram aos seguintes resultados:

$$\sum_{i=1}^{52} x_i = 1138, \quad \sum_{i=1}^{52} x_i^2 = 28800, \quad \sum_{i=1}^{52} y_i = 249.7, \quad \sum_{i=1}^{52} y_i^2 = 1366.35, \quad \sum_{i=1}^{52} x_i y_i = 6254.9$$

- (a) Obtenha as estimativas de mínimos quadrados de  $\beta_0$  e  $\beta_1$  e interprete a estimativa de  $\beta_1$ . (2.0)
  - [Modelo de RLS

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

 $Y_i = \text{distância de travagem do } i - \text{ésimo veículo pesado}$ 

 $x_i$  = velocidade a que o i – ésimo veículo pesado se desloca

 $\epsilon_i = {
m erro}$  aleatório associado à medição da distância de travagem do i-ésimo veículo]

• Estimativas de  $\beta_0$  e  $\beta_1$ 

[Uma vez que 
$$n = 52$$
 e

$$\sum_{i=1}^{n} x_i = 1138,$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{1138}{52} = 21.885$$

$$\sum_{i=1}^{n} x_i^2 = 28800$$

$$\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2 = 28800 - 52 \times 21.885^2 = 3895.432$$

$$\sum_{i=1}^{n} y_i = 249.7,$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i = \frac{249.7}{52} = 4.802$$

$$\sum_{i=1}^{n} y_i^2 = 1366.35$$

$$\sum_{i=1}^{n} y_i^2 - n(\bar{y})^2 = 1366.35 - 52 \times 4.802^2 = 167.271$$

$$\sum_{i=1}^{n} x_i y_i = 6254.9$$

$$\sum_{i=1}^{n} x_i y_i - n \bar{x} \bar{y} = 6254.9 - 52 \times 21.885 \times 4.802 = 790.128,$$

a estimativa dos mínimos quadrados de  $\beta_1$  e  $\beta_0$  são, para este modelo, iguais a:

$$\hat{\beta}_{1} = \frac{\sum_{i=1}^{n} x_{i} y_{i} - n \bar{x} \bar{y}}{\sum_{i=1}^{n} x_{i}^{2} - n (\bar{x})^{2}}$$

$$= \frac{790.128}{3\,895.432}$$

$$= 0.203$$

$$\hat{\beta}_{0} = \bar{y} - \hat{\beta}_{1} \times \bar{x}$$

$$= 4.802 - 0.203 \times 21.885$$

$$= 0.359$$

 $\bullet$  Interpretação da estimativa de mínimos quadrados de  $\beta_1$ 

$$\hat{\beta}_1 = 0.203$$

Caso a velocidade a que veículo pesado se desloca aumente em um m/s, estima-se que o valor esperado da distância de travagem aumente aproximadamente 0.203 metros.

- (b) Construa um intervalo de confiança a 95% para β<sub>1</sub>. Será que existe uma relação linear entre a (4.0) distância média de travagem de um veículo pesado e a velocidade a que o mesmo se desloca?
  - [Hipóteses de trabalho

$$\epsilon_i\sim_{i.i.d.} \text{Normal}(0,\sigma^2),\, i=1,\dots,n$$
 (hipótese de trabalho)  $\beta_0,\beta_1,\sigma^2$  DESCONHECIDOS]

- IC a 95% para  $\beta_1$ 
  - Passo 1 V.a. fulcral para  $\beta_1$

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n \, \bar{x}^2}}} \sim t_{(n-2)}$$

### - Passo 2 — Quantis de probabilidade

Já que  $(1-\alpha)\times 100\%=95\%$  temos  $\alpha=0.05$  e lidaremos com dois quantis simétricos  $a_\alpha=-b_\alpha$  iguais a:

$$\pm F_{t_{(n-2)}}^{-1}(1-\alpha/2) = \pm F_{t_{(52-2)}}^{-1}(1-0.05/2) = F_{t_{(50)}}^{-1}\Phi^{-1}(0.975) \stackrel{tabela}{=} \pm 2.009.$$

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

$$P(a_{\alpha} \le Z \le b_{\alpha}) = 1 - \alpha$$

$$P\left[ -F_{t(n-2)}^{-1} (1 - \alpha/2) \le \frac{\hat{\beta}_1 - \beta_1}{\sqrt{\sum_{i=1}^{n} x_i^2 - n \,\bar{x}^2}} \le F_{t(n-2)}^{-1} (1 - \alpha/2) \right] = 1 - \alpha$$

$$P\left\{\hat{\beta}_{1} - F_{t_{(n-2)}}^{-1}(1 - \alpha/2) \times \sqrt{\frac{\hat{\sigma}^{2}}{\sum_{i=1}^{n} x_{i}^{2} - n\bar{x}^{2}}}\right.$$

$$\leq \beta_{1} \leq \hat{\beta}_{1} + F_{t_{(n-2)}}^{-1}(1 - \alpha/2) \times \sqrt{\frac{\hat{\sigma}^{2}}{\sum_{i=1}^{n} x_{i}^{2} - n\bar{x}^{2}}}\right\} = 1 - \alpha.$$

### - Passo 4 — Concretização

Tendo em conta que a estimativa de  $\sigma^2$  é dada por

$$\hat{\sigma}^2 = \frac{1}{n-2} \left[ \left( \sum_{i=1}^n y_i^2 - n \, \bar{y}^2 \right) - (\hat{\beta}_1)^2 \left( \sum_{i=1}^n x_i^2 - n \, \bar{x}^2 \right) \right]$$

$$= \frac{1}{52-2} \left( 167.310 - 0.203^2 \times 3895.432 \right)$$

$$\simeq 0.139,$$

segue-se

$$IC_{(1-\alpha)\times 100\%}(\beta_1) = \left[ \hat{\beta}_1 \pm F_{t_{(n-2)}}^{-1} (1 - \alpha/2) \times \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n \, \bar{x}^2}} \right]$$

$$IC_{95\%}(\beta_1) = \left[ 0.203 \pm 2.009 \times \sqrt{\frac{0.139}{3\,895.432}} \right]$$

$$= [0.191, 0.215].$$

### • Hipóteses

 $H_0: \beta_1 = \beta_{1,0} = 0$ 

 $H_1: \beta_1 \neq \beta_{1,0} = 0$  (existe relação linear entre o valor esperado de Y e x)

#### Decisão

Invocando a relação entre intervalos de confiança e testes de hipóteses bilaterais, devemos rejeitar a hipótese  $H_0: \beta_1 = \beta_{1,0} = 0^2$  ao n.s. de  $\alpha \times 100\% = 100\% - 95\% = 5\%$  (ou a qualquer outro n.s. maior que 5%) já que

$$0 \notin IC_{95\%}(\beta_1) = [0.191, 0.215].$$

Esta analogia é válida porque:

- a v.a. fulcral para  $\beta_1$  usada na construção do IC é também utilizada para definir a estatística de teste sobre  $\beta_1$ ;
- -o nível de significância do teste, 5%, é igual a (100% nível de confiança do IC).]

<sup>&</sup>lt;sup>2</sup>A favor da hipótese  $H_1: \beta_1 \neq \beta_{1,0} = 0$ .