Estatística teórica 28/10/2019 18:10

— // —

Revisão e conceitos de Estatística Descritiva

Análise de dados estatísticos.

Dados Estatísticos

- Qualitativos
  - nominais (não se consegue ordenar)
  - ordinais
- Quantitativas
  - Discretas
  - contínuas

n_i → Freq Abs.

$f_i = \frac{n_i}{N}$ → Freq Relativa

n → Número total de observações.

N_i → Freq acumulada

F_i → Freq Relativa acumulada

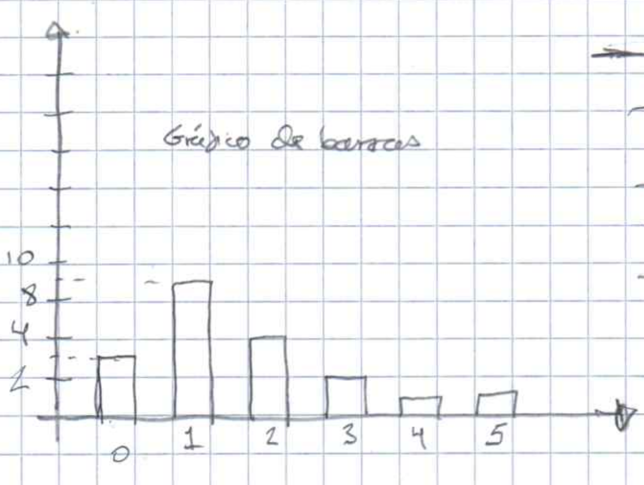
Exemplo:

X - "Nº de filhos por Pessoa"

1, 0, 1, 2, 1, 3, 1, 2
1, 4, 3, 1, 2, 1, 2, 1, 0, 0, 1, 5

tabela de Frequência

x_i	contagem	n_i	f_i	N_i
0	III	3	0,15	3
1	IIIIIIIII	9	0,45	12
2	IIII	4	0,2	16
3	II	2	0,1	18
4	I	1	0,05	19
5	I	1	0,05	20



exemplo tempo que demora

5,1 2,31
3,4
10,5
20,1 → falta de se criar classes

Regra para definir classes.

[0, 5 [	3
[5, 10 [	
[10, 15 [	1
[15, 120 [	
[20, 25 [	1

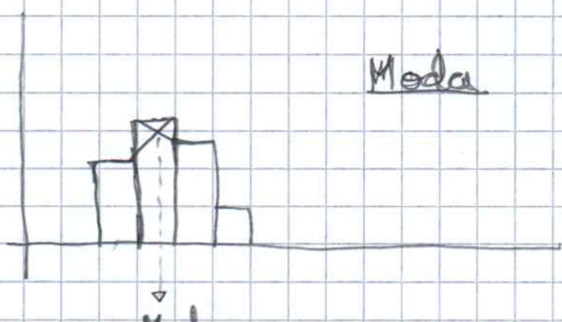
Marca da classe é sua média → $x_i = \frac{a+b}{2}$

Para valores contínuos vamos usar o histograma, na qual a área de cada barra é seu valor ou frequência absoluta. (Pg 11)

Medidas descritivas

(Pg 13)

Moda



Moda

Mediana

500
500 1500
1800
15000 → $\bar{x} = 3860$

500, 500, 1500, 1800, 15000,

x_1 x_2 x_3 x_4 x_5

↓

Mediana

$M_e = x_{\left(\frac{5+1}{2}\right)}$

$= x_3$

Estatística
teórica

28/10/2019
19:00

— // —

000 1111111111 2222 45

ordenar por ordem depois encontrar item central.

Estatística I — 4/11/2019 18:15

— // —

tabela de frequências acumuladas. Pg 9.

Representação Gráfica

Histograma

Área = frequência absoluta. Pg 11

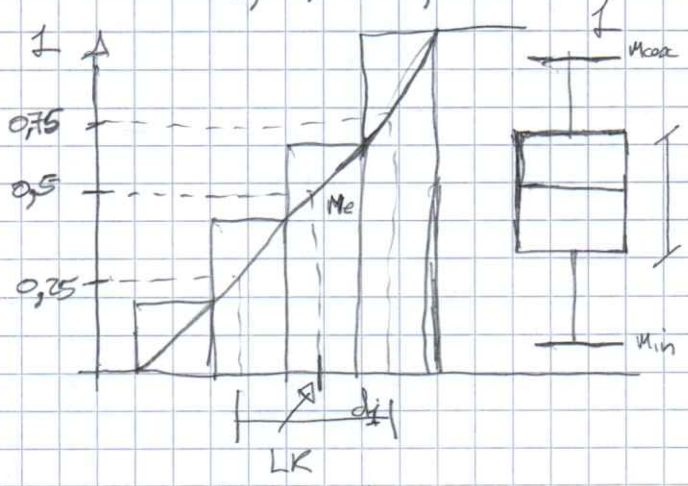
Medidas de Localização Pg 12

Média $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i n_i = \frac{\sum_{i=1}^{n} x_i f_i}{n}$$

Moda

Mediana ⇒ $F(\cancel{0},5) = 0,5$



1 Quartil = 0,25

2 Quartil = 0,5

Medida Dispersão "Range" r

↳ Variância

$\boxed{IQR}$ 3 Quartil − 1 Quartil.

Outliers — when a very encounter a lie.

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

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{c} (x_i - \bar{x}_i)^2$$ Pg 21

Uma é para amostras ⇒ $\frac{1}{n-1}$

Media
Moda
Mediana
Dispersão $\rightarrow$ variância

Medidas de forma pag 23.

coeficiente de assimetria amostral
- $\rightarrow a_3 > 0$
- $\rightarrow a_3 < 0$
- $\rightarrow a_3 = 0$

coeficiente de curtose amostral (importante) pag 25

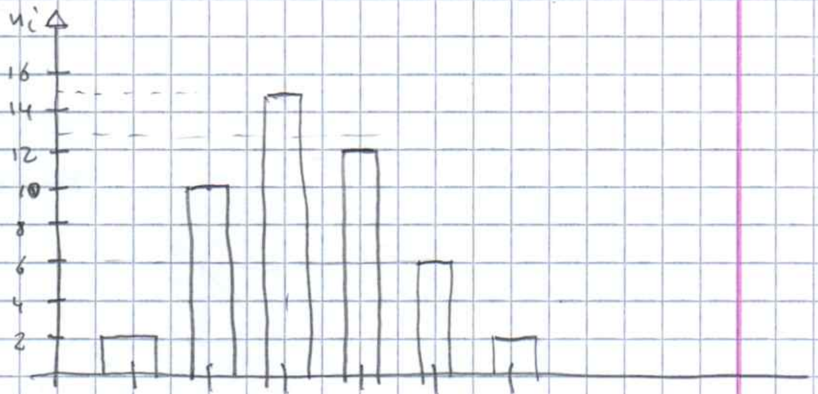
Exercício 1 pag 26

X - "Nº de elementos do agregado familiar"

frequencia absoluta

x_i	n_i	f_i	F_i	
1	4	0,08	0,08	4
2	10	0,2	0,28	14
3	15	0,3	0,58	29
4	13	0,26	0,84	42
5	6	0,12	0,96	48
6	2	0,04	1	50

N=50



c) Moda = 3

$$\text{Mediana} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{3+3}{2} = 3$$

$$\text{Média} = \frac{\sum_{i=1}^{6} x_i n_i}{50} = 3,26$$

Fazer em Excel
usando Dados obtidos em extenso. 1,1,1,1 2,2,2 etc
- Variancia
- Moda
- Media
-

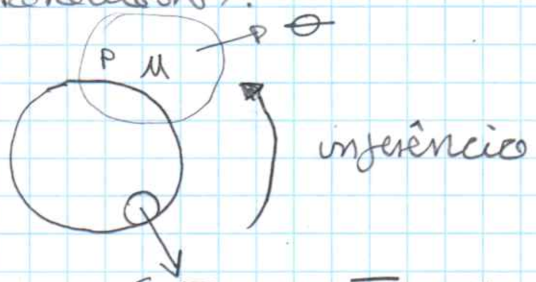
d)

e) Skew (table) 0,1234 (extensos)
Kurt (table) -0,42349

Estatística 18/11/2019 18:15

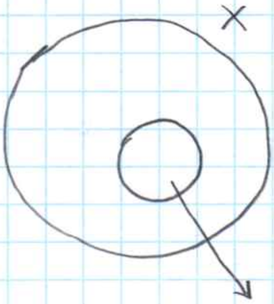
— // — T

Estimação de Parâmetros.



$$\left\{\begin{matrix} \bar{x}_0 \\ n \end{matrix}\right. \rightarrow \bar{X} = F(X_1, X_2, \ldots, X_n) = \frac{\sum X_i}{n}$$

$$\bar{x}_0 = F(x_1, x_2, \ldots, x_n) = \frac{\sum x_i}{n}$$



$X_1, X_2, \ldots, X_n$

$1 - \alpha \leftarrow$ Grande confiança

$$\Delta = z \times \boxed{\sigma}$$

$$\frac{\sigma}{\sqrt{n}}$$

$1 - \alpha = 0.95$ (por defeito)

$$IC_\theta = [\hat{\theta} - \Delta, \hat{\theta} + \Delta] \quad c/\ (1-\alpha) \times 100\%$$

AMPLITUDE DO IC = 2Δ

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \underset{A,\mu}{\sim} N\left(\mu ; \frac{\sigma^2}{n}\right)$$

$$\mu = \bar{X} - Z \frac{\sigma}{\sqrt{n}}$$

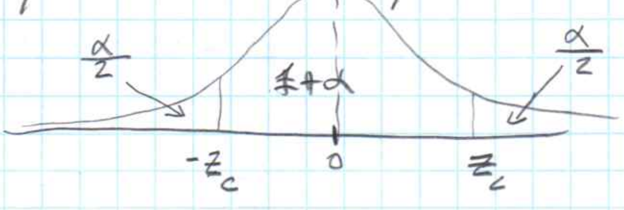
Pg 5

$$P(A \leq \mu \leq B) = 1 - \alpha$$

$$P\left(A \leq \bar{X} - Z \frac{\sigma}{\sqrt{n}} \leq B\right) = 1 - \alpha$$

$$P\left(\frac{A - \bar{X}}{\sigma/\sqrt{n}} \leq -Z \leq \frac{B - \bar{X}}{\sigma/\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(\frac{\bar{X} - B}{\sigma/\sqrt{n}} \leq Z \leq \frac{\bar{X} - A}{\sigma/\sqrt{n}}\right) = 1 - \alpha$$



$$\left\{\begin{aligned} -z_c &= \frac{\bar{X} - B}{\sigma/\sqrt{n}} \\ z_c &= \frac{\bar{X} - A}{\sigma/\sqrt{n}} \end{aligned}\right. \qquad \left\{\begin{aligned} B &= \bar{X} + z_c \frac{\sigma}{\sqrt{n}} \\ A &= \bar{X} - z_c \frac{\sigma}{\sqrt{n}} \end{aligned}\right.$$

$$z_c = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$$

Pg 7

~ 1.96

$$IC_\mu = [\, \bar{X} - \Delta \,;\, \bar{X} + \Delta \,]$$

$$\Delta = Z_c \frac{\sigma}{\sqrt{n}} \leftarrow \text{erro de estimativa.}$$

intervalo de confiança

$$Z_c = \Phi^{-1}(1 - \alpha/2)$$

intervalo de confiança

$1 - \alpha = 0.95$

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

$Z_{1-\frac{\alpha}{2}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)$

$$\Delta = Z_c \frac{\sigma}{\sqrt{n}} \leqslant \varepsilon$$

Exemplo 1

Pg 12

X – " Diametro de uma peças [mm] "

$\mu = 20 \qquad \sigma = 3.1$ (

$$\begin{cases} \bar{x}_0 = 21.4 \\ n = 25 \end{cases}$$

Determinar $IC_\mu = [A, B]$

$\bar{X}$ – " Diametro médio, 1E 2I peças [mm] "

$$\bar{X} = \frac{\sum_{i=1}^{25} X_i}{25} \sim N\left(\mu, \frac{3.1^2}{25}\right)$$

Após Amostragem:

$1 - \alpha = 0.95$

$\alpha = 0.05$

$$Z_{(1-\frac{\alpha}{2})} = \Phi^{-1}\left(1 - \frac{0.05}{2}\right)$$

$$= \Phi^{-1}(0.975) = 1.96$$

$$\Delta = Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{25}} = 1.96 \times \frac{3.1}{\sqrt{25}}$$

$$= 1.22$$

$$IC_\mu = [\, \bar{x}_0 - \Delta \,;\, \bar{x}_0 + \Delta \,]$$

$$= [\, 21.4 - 1.22 \,;\, 21.4 + 1.22 \,]$$

R: $IC_{\mu\,(95\%)} = [20.18 \,,\, 22.62]$

$= c \quad 1 - \alpha = 0.95$

Estatística 18/11/2019

18:51

— || —

$$IC_{\mu_1-\mu_2} = [A, B]$$

formula pagina $\boxed{15}$

exemplo 2.

$$IC_{\mu_1-\mu_2} = [\bar{x}_1 - \bar{x}_2 - \Delta ; \bar{x}_1 - \bar{x}_2 + \Delta]$$

$$\Delta = Z_{(1-\frac{\alpha}{2})} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\begin{cases} \bar{x}_A = 425.3 \\ s_A = 51.3 \\ n_A = 30 \end{cases} \quad \begin{cases} \bar{x}_B = 441,2 \\ s_B = 55,1 \\ n_B = 30 \end{cases}$$

"intervalo de confiança" - parametro

$$\Delta = 1.96 \times \sqrt{\frac{s_A^2}{30} + \frac{s_B^2}{30}}$$
$$= 1.96 \times \sqrt{\frac{51.3^2}{30} + \frac{55.1^2}{30}}$$
$$= 26.9$$

$s_A \approx S_A$ $n_1 \geq 30$

$s_B \approx S_B$ $n_2 \geq 30$

$$IC_{\mu_1-\mu_2} = (\bar{x}_1 - \bar{x}_2 - \Delta ; \bar{x}_1 - \bar{x}_2 + \Delta)$$
$$= [(425.3 - 441.2) - 26.9, (426.5 - 446.2) + 26.9)$$

R: $IC_{\mu_1-\mu_2} = (-19.2 - 26.9, -19.2 + 26.9]$

$= [-41.6, 7.7]$ km

$\leftarrow 1-\alpha = 0.95$

Estatistica. 25/11/2019.

teoria 18:10

— 11 —

capitulo 5   C5

test de hipoteses.

Duas hipoteses possiveis $\begin{cases} H0 \\ H1 \end{cases}$

Pg 4 $\Theta = \{\mu, p\}$

θ_0 valor fixo.

$H0 : \mu = \mu_0$

$H1 : \mu \neq \mu_0$ $\begin{cases} \bar{x}_0 \\ n \end{cases}$

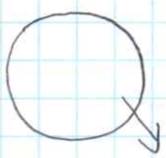
RC - região critico ou de rejeição.

Erro tipo I - nivel de significância.
α para verdadeiro

β para falso

Pg 7
Quadro de casos possiveis

Pg 9   teste de média para uma população



$\bar{x}_0$

$\bar{X} \sim N(\mu ; \frac{\sigma^2}{n})$

normalizado 5%

Pg 10

$H_0 : \mu = \mu_0$

$H_1 : \begin{cases} \mu \neq \mu_0 \\ \mu > \mu_0 \\ \mu < \mu_0 \end{cases}$

* teste de média unilateral a esquerda.

exemplo $\mu = 5000$ horas ?

$\bar{x}_0 = 4850\ h$

hipóteses $\begin{cases} H_0: & \mu = 5000 \\ H_1: & \mu < 5000 \end{cases}$

α

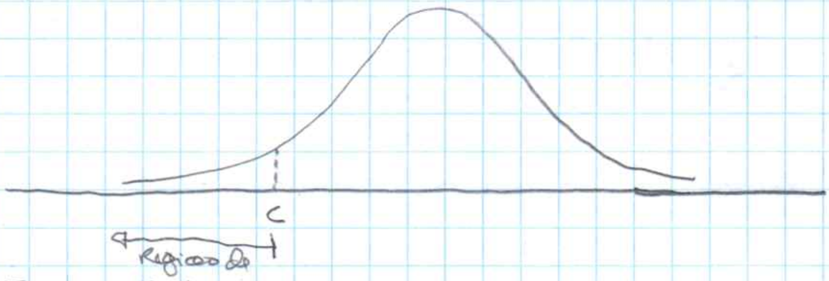
$H_0: \mu = \mu_0$

$H_1: \mu < \mu_0$

$\bar{X} \sim N(\mu ; \frac{s^2}{n})$

Determinar RC_x

sob $H_0: \bar{X}_{H_0} \sim N(\mu_0 ; \frac{s^2}{n})$



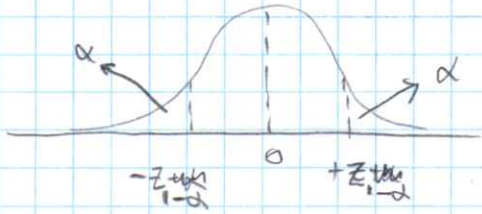
$\begin{cases} \bar{x}_0 = 4850 \\ n = 30 \end{cases}$

$RC_x =]-\infty, c]$

$P(\bar{X}_{H_0} < c) = \alpha$

$P(Z < \frac{c - \mu_0}{s/\sqrt{n}}) = \alpha \Leftrightarrow \Phi\left(\frac{c-\mu_0}{s/\sqrt{n}}\right) = \alpha$

$\Leftrightarrow \frac{c - \mu_0}{s/\sqrt{n}} = \Phi^{-1}(\alpha) = -Z_{1-\alpha}$



$c = \mu_0 - Z_{1-\alpha} \frac{\alpha}{\sqrt{n}}$

$RC_x =]-\infty, \mu_0 - Z_{1-\alpha} \cdot \frac{\alpha}{\sqrt{n}}]$

Decisão se $\bar{x}_0 \in RC_x$

H_0 Deve ser Rejeitado senão H_0 Não deve ser rejeitado

Estatística 25/11/2019

teórica 18h30

— 11 —

$$\begin{cases} \bar{x}_0 = 4850h \\ n = 30 \\ s = 500h \\ \alpha = 0{,}05 \end{cases}$$

apart exemplo

provar que 4850 está dentro 5% aos 5000h

② $\bar{X} \sim N(\mu ; \frac{\sigma^2}{30})$

sob H_0: $\bar{X}_{H_0} \sim N(5000, \frac{500^2}{30})$

$\sigma = s$

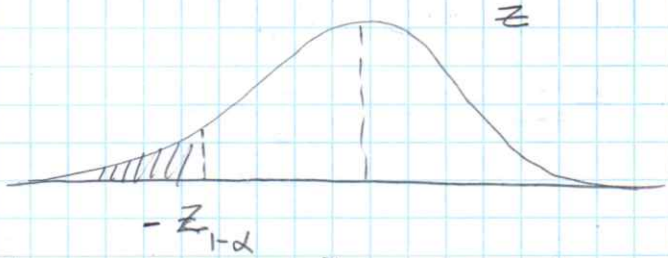
$$RC =]-\infty, \mu_0 - z_{1-\alpha} \frac{\sigma}{\sqrt{n}}]$$

$$=]-\infty, 5000 - 1.645 \times \frac{500}{\sqrt{30}}$$

$$=]-\infty, 4849{,}83]$$

como $\bar{x}_0 = 4850h \notin RC_x$ Não se rejeita H_0 ao nível de significância de 5%.

Outro método



$$RC_Z =]-\infty, -Z_{1-\alpha}]$$

$$Z_{1-\alpha} = \Phi^{-1}(1-\alpha)$$

$$Z = \frac{\bar{X}_{H_0} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$Z_{observado} = \frac{\bar{x}_0 - x_0}{\frac{s}{\sqrt{n}}}$$

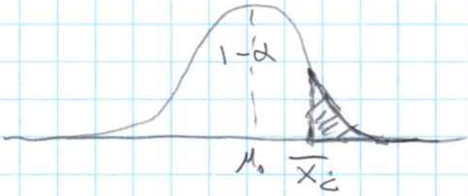
Decisão

se $Z_{obs} \in RC_Z \Rightarrow$ Rejeita-se H_0

senão não se Rejeita H_0

teste unilateral à direita $H0: \mu = \mu_0$

$H1: \mu > \mu_0$



tres testes
- esquerdo
- direito
- bilateral

bilateral

$$RC_{\bar{x}} =]-\infty, \mu_0 - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}] \cup$$

$$]\mu_0 + z_{1-\frac{\alpha}{2}}; +\infty]$$

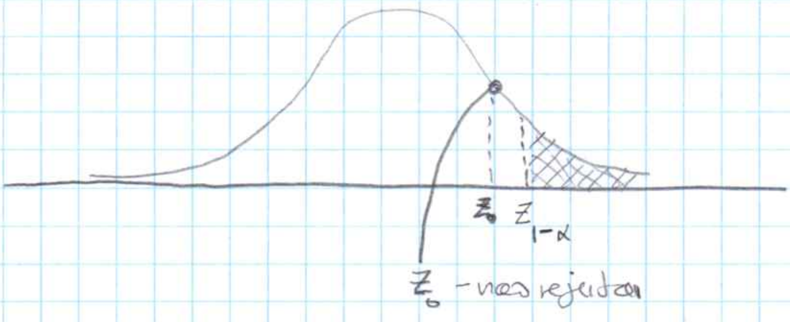
$$z_{1-\frac{\alpha}{2}} = \Phi^{-1}(1-\frac{\alpha}{2})$$

$$RC_z =]-\infty; -z_{1-\frac{\alpha}{2}}] \cup [z_{1-\frac{\alpha}{2}}, +\infty[$$

$$Z_{obs} = \frac{\bar{x}_0 - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

técnica 3

critério de valor de prova (pvalue)



$$RC = [z_{1-\alpha}, +\infty[$$

valor p $= P(Z > z_0) \geq \alpha$

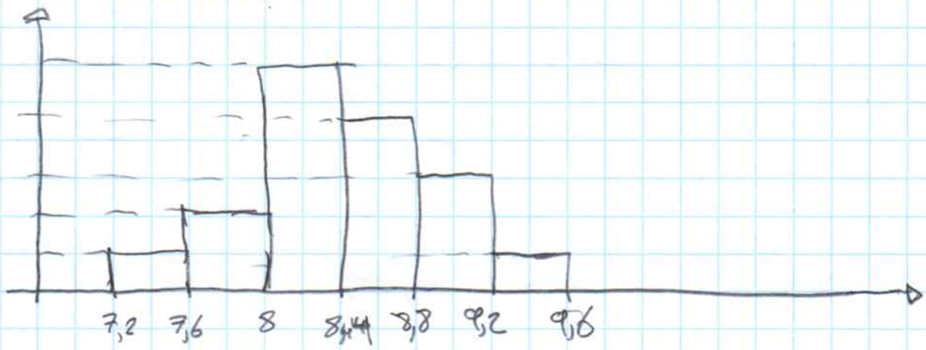
se pvalue $\leq \alpha$ Rejeita-se H_0
se pvalue $\geq \alpha$ não se Rejeita H_0

Estatística 9/12/2019

18:12

T

X – "Qtd de corante por bebida (ml)"



$\alpha = 0,05$

X ~ Exp (λ)

(1) $H_0: X \sim N(\quad)$

→ frequência observada

→ frequência esperada sub H_0

	x_i	n_i	e_i
< 7,2	—	0	0,3
7,2 – 7,6	7,4	2	2
7,6 – 8	7,8	7	7,6
8 – 8,4	8,2	18	15,4
8,4 – 8,8	8,6	15	16,7
8,8 – 9,2	9	9	9,7
9,2 – 9,6	9,4	4	3,0
≥ 9,6	—	0	0,5
		55	55,0

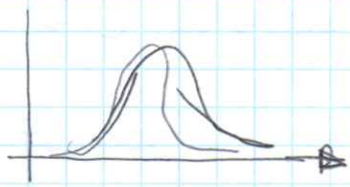
Soma n_i 9

$p_{e_1} = ?$

$P(X < 7,2) \times 55 \simeq 0,3$

$e_i \geq 5$

X ~ Un [a, b]



$$\bar{x}_0 = \frac{\sum x_i n_i}{n} \qquad s^2 = \frac{n\left(\sum x_i^2 n_i\right) - \left(\sum x_i n_i\right)^2}{n(n-1)}$$

$$\frac{464,6}{55} = 8,45 \qquad = \frac{55 \times 3937,56 - 464,6^2}{55 \times 54} = 0,2399$$

c.a) $\sum x_i^2 n_i = 7,4^2 \times 2 + \dots + 9,4^2 \times 4$

$= 3937,56$

$H_1: X \tilde{n} \text{ é } N(8,45;\ 0,2399)$

(2) $Q = \sum_{i=1}^{n} \frac{(N_i - e_i)^2}{e_i} \sim \chi^2_{(k-m-1)}$ sempre desde a direita

(3) Determinar $RC_\alpha = [\,c, +\infty[$

Sub H_0:

$e_i = P(7,2 \leq X < 7,6) \times n$

$= \text{normCDF}(7,2;\ 7,6;\ 8,45;\ \sqrt{0,2399}) \times 55)$

∴ Quadro transformado

n_i	e_i	$\frac{(n_i-e_i)^2}{e_i}$
9	9.9	0,081
18	15.4	0,44
15	16,7	0,17
13	13,2	0,03
55	55,0	0,69

$K=4$

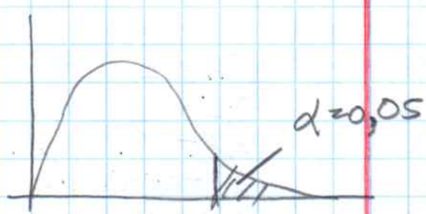
$gl = K - m - 1$

$K \leftarrow N^{o}$ de classes
$m \leftarrow N^{o}$ de parametros

$K=4;\ m=2$

$$Q = \sum_{i=1}^{4} \frac{(N_i - e_i)^2}{e_i} \sim \chi^2_{(gl=4-2-1)}$$

$$P(Q > c) = 0{,}05 = \alpha$$



chi → InvC → Area = 0,05; df = 1   c = Xinv = 3,84145

Decisão

calcular $q_0 = \frac{\sum (n_i - e_i)^2}{e_i}$

$= \frac{(9-9{,}9)^2}{9{,}9} + \cdots + \frac{(13-13{,}2)^2}{13{,}2}$

como $q_0 < 3{,}84$
aceita-se o ajuste ao nível de **5%**