A

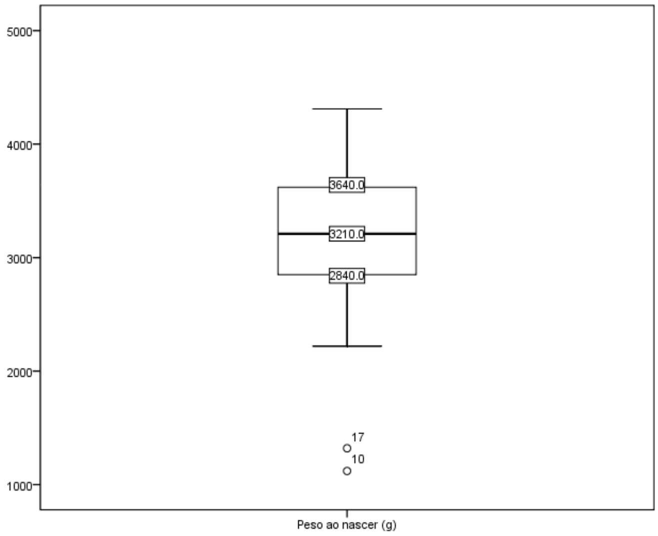
FREQUÊNCIA DE ESTATÍSTICA APLICADA

Universidade do Minho, Escola de Engenharia, Departamento de Produção e Sistemas
Licenciatura em Engenharia Informática
Ano lectivo 2009/2010
11/11/2009 – Duração: 1h 45 minutos

Leia com atenção os enunciados e apresente todos os cálculos que tiver de efectuar.

1. As seguintes representações dizem respeito a dados referentes ao peso ao nascer de 55 bebés, em gramas, registados num Hospital:

Peso ao nascer (g)		Statistic
Mean		*3206.73*
95% Confidence Interval for Mean	Lower Bound	*3033.85*
	Upper Bound	*3379.60*
Median		*3210.00*
Variance		*408940.017*
Std. Deviation		*639.484*
Minimum		*1120*
Maximum		*4310*
Range		*3190*



a. Qual o tipo de variável e qual a designação do gráfico apresentado?

Variável contínua medida numa escala proporcional; o gráfico apresentado é uma caixa de bigodes, ou diagrama de caixa fio.

b. Apenas com base na informação apresentada, indique as medidas centrais e de dispersão dos dados.

Medidas centrais: média= *3206.73;* mediana=*3210*; medidas de dispersão: variância = 408940.017 Desvio padrão=639.484, amplitude = 3190, P25= 2840 e P75=3640.

c. Qual a percentagem de bebés com peso superior ou igual a 3640 g? Qual a percentagem de bebés com peso inferior a 2000 g?

A percentagem de bebés com peso superior ou igual a 3640 g é de 25% (**P75=3640, significa que P(X≤3640)=0.75 ⇒ P(X>3640)=0.25). P(X<2000)=(2/55)=0.0364=3.64%**

d. Que pode dizer relativamente às observações 10 e 17 representadas no gráfico? Justifique.

As observações 10 e 17 são valores extremos moderados, porque estão para além (abaixo) da barreira que delimita Q1 - 1.5*DIQ = 2840-1200 g =1640 g.

2. Num conjunto de 5000 indivíduos observados quanto às características "cor dos olhos" e "cor do cabelo", registaram-se os seguintes resultados:

	Cor do cabelo		
Cor dos olhos	**Loiro**	**Moreno**	**Total**
Claros	1225	3575	4800
Escuros	50	150	200
Total	1275	3725	5000

Na população de onde foi retirada esta amostra, determine:

a. a probabilidade de ocorrência de pessoas com o cabelo loiro.

R: $P(\text{loiros}) = \frac{1275}{5000} = 0.255$

b. a probabilidade de ocorrência de pessoas com cabelo loiro e olhos claros.

R: $P(\text{loiros e olhos claros}) = \frac{1225}{5000} = 0.245$

c. a probabilidade de ocorrência de pessoas com cabelo loiro, caso tenham os olhos claros.

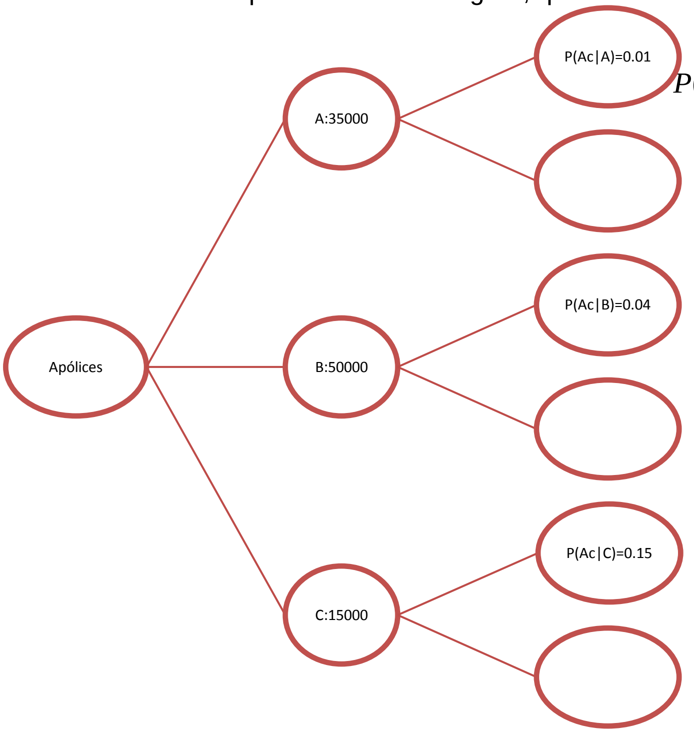
R: $\mathrm{P}(\text{loiros} \mid \text{olhos claros}) = \dfrac{\mathrm{P}(\text{loiros e olhos claros})}{\mathrm{P}(\text{olhos claros})} = \dfrac{0.245}{\frac{4800}{5000}} = 0.2552$

d. a probabilidade de ocorrência de pessoas com olhos claros, caso tenham o cabelo loiro.

R: $\mathrm{P}(\text{olhos claros}|\text{loiros }|) = \dfrac{\mathrm{P}(\text{loiros e olhos claros})}{\mathrm{P}(\text{loiros})} = \dfrac{0.245}{0.255} = 0.9608$

3. Uma companhia de seguros distribui os seus segurados por três classes, A, B e C, consoante o menor ou maior risco que lhes atribui. Em determinado momento, era a seguinte a carteira de apólices: classe A : 35000 segurados; classe B: 50000 segurados; classe C: 15000 segurados. A probabilidade de os segurados de cada classe sofrerem um ou mais acidentes no próximo ano é de 0.01, 0.04 e 0.15, respectivamente.

A companhia de seguros nunca tem a certeza a que classe pertence o subscritor do seguro. Se o segurado tiver um acidente durante o primeiro ano de seguro, que conclusões pode tirar?



$$P(Ac) = \frac{0.01*35000 + 0.04*50000 + 0.15*15000}{100000} = 0.046$$

$$P(A|Ac) = \frac{P(A \cap Ac)}{P(Ac)} = \frac{0.35*0.01}{0.046} = 0.076$$

$$P(B|Ac) = \frac{P(B \cap Ac)}{P(Ac)} = \frac{0.50*0.04}{0.046} = 0.4348$$

$$P(C|Ac) = \frac{P(C \cap Ac)}{P(Ac)} = \frac{0.15*0.15}{0.046} = 0.4891$$

R: Pode concluir que existe maior probabilidade de ser um segurado do tipo C.

4. Seja X uma variável aleatória com a seguinte distribuição de probabilidades:

x_i	1−2k	k−1	k	2k
$f(x_i) = P(X = x_i)$	p	3p	p	p

a. Sabendo que E(X) = 1/3 calcule o valor de p e k.

$$\begin{cases} p + 3p + p + p = 1 \\ (1-2k)*p + (k-1)*p + k*p + 2k*p = \frac{1}{3} \end{cases} \begin{cases} 6p = 1 \\ p - 2kp + kp - p + kp + 2kp = \frac{1}{3} \end{cases} \begin{cases} p = \frac{1}{6} \\ 2kp = \frac{1}{3} \end{cases} \begin{cases} p = \frac{1}{6} \\ k = 1 \end{cases}$$

b. Calcule V[X].

$$V[X] = E\left[X^2\right] - \left(E[X]\right)^2 = E\left[X^2\right] - \frac{1}{9} = 1 - \frac{1}{9} = \frac{8}{9}$$

$$E\left[X^2\right] = (-1)^2 * \frac{1}{6} + 0 * \frac{3}{6} + 1^2 * \frac{1}{6} + 2^2 * \frac{1}{6} = 1$$

c. Determine a função distribuição acumulada de X.

x_i	-1	0	1	2
$f(x_i) = P(X = x_i)$	1/6	3/6	1/6	1/6

$$F(x)=\begin{cases}0 & x<-1\\ 1/6 & -1\le x<0\\ 4/6 & 0\le x<1\\ 5/6 & 1\le x<2\\ 1 & x\ge 2\end{cases}$$

d. Calcule P(X ≥ 0 | X < 2).

R:

$$P(X\ge 0\,|\,X<2)=\frac{P(X\ge 0\wedge X<2)}{P(X<2)}=\frac{P(x=0)+P(x=1)}{P(x\le 1)}=\frac{4/6}{5/6}=\frac{4}{5}$$

5. Suponha que o tempo, em minutos, necessário para que um técnico afine cada componente de uma maquina numa linha de produção é descrito por uma variável aleatória X com distribuição U(5,10).

a. Qual a probabilidade de que uma componente escolhida ao acaso:

i. Tenha necessitado de mais de 7 minutos para ser afinada.

$$P(x>7)=\int_7^{10} f(x)dx=\int_7^{10}\frac{1}{10-5}dx=\frac{1}{5}[x]_7^{10}=\frac{1}{5}*3=\frac{3}{5}$$

ii. Tenha exigido ao técnico um tempo de afinação inferior a 9 minutos sabendo que aquele tempo foi superior a 7 minutos.

$$P(x<9\,|\,x>7)=\frac{P(x<9\wedge x>7)}{P(x>7)}=\frac{\int_7^9\frac{1}{5}dx}{\frac{3}{5}}=\frac{\frac{2}{5}}{\frac{3}{5}}=\frac{2}{3}$$

b. Seja Y = 10 + 5X o custo em euros de afinação de cada componente. Qual o custo esperado de afinação de uma componente escolhida ao acaso?

$$E[Y]=E[10+5X]=10+5E[X]=10+5*7.5=47.5€$$

$$E[X]=\frac{5+10}{2}=7.5$$

6. A vida útil de uma calculadora fabricada pela empresa SOMAESEGUE segue uma distribuição normal com μ=54 meses e σ=8 meses. A empresa garante que qualquer calculadora que comece a funcionar mal até ao 36º mês após a compra é substituída por uma nova.

a. Que percentagem de calculadoras feitas por esta empresa espera-se substituir?

$$P(x<36)=P\left(z<\frac{36-54}{8}\right)=P(z<-2.25)\overset{tabela5}{=}0.0122$$

b. Qual período de garantia deve a empresa dar se ela não quiser substituir mais do que 1% das calculadoras avariadas?

$$P(x<k)\le 0.01\Leftrightarrow P(z<k)\le 0.01\overset{tabela5}{\Rightarrow}k=-2.33$$
$$\frac{x-54}{8}=-2.33\Leftrightarrow x\approx 35.36\to x\le 35.36\text{ meses}$$

7. Sejam $x_1, x_2, \ldots, x_n$, os elementos de uma amostra aleatória da distribuição:

$$f(x)=\begin{cases}\frac{1}{\theta}x^{-(1+\theta)/\theta} & x>1,\ \theta>0\\ 0 & x\le 1\end{cases}$$

a. Encontre o Estimador de Máxima Verosimilhança para θ.

$$L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)=\prod_{i=1}^{n}\frac{1}{\theta}x_i^{-\frac{1+\theta}{\theta}}=\left(\frac{1}{\theta}\right)^n\left(\prod_{i=1}^{n}x_i\right)^{-\frac{1+\theta}{\theta}}$$

$$\ln\left[\left(\frac{1}{\theta}\right)^n\left(\prod_{i=1}^{n}x_i\right)^{-\frac{1+\theta}{\theta}}\right]=\ln\left(\frac{1}{\theta}\right)^n+\ln\left(\prod_{i=1}^{n}x_i\right)^{-\frac{1+\theta}{\theta}}=-n\ln\theta-\frac{1+\theta}{\theta}\sum_{i=1}^{n}\ln x_i=-n\ln\theta-\frac{1}{\theta}\sum_{i=1}^{n}\ln x_i-\sum_{i=1}^{n}\ln x_i$$

$$\frac{d\ln L(\theta)}{d\theta}=-\frac{n}{\theta}+\frac{1}{\theta^2}\sum_{i=1}^{n}\ln x_i$$

$$\frac{d\ln L(\theta)}{d\theta}=0\Leftrightarrow-\frac{n}{\theta}+\frac{1}{\theta^2}\sum_{i=1}^{n}\ln x_i=0\Leftrightarrow\frac{1}{\theta}\left(-n+\frac{1}{\theta}\sum_{i=1}^{n}\ln x_i\right)=0\Leftrightarrow-n+\frac{1}{\theta}\sum_{i=1}^{n}\ln x_i=0\Leftrightarrow\frac{1}{\theta}\sum_{i=1}^{n}\ln x_i=n\Leftrightarrow\theta=\frac{\sum_{i=1}^{n}\ln x_i}{n}$$

$$\therefore\quad\hat{\theta}=\frac{\sum_{i=1}^{n}\ln x_i}{n}$$

b. Admita que observou a seguinte amostra: (1.3; 2.1; 1.7; 1.5; 2.7). Determine o valor da estimativa de máxima verosimilhança para θ.

$$\hat{\theta}_5=\frac{\sum_{i=1}^{5}\ln x_i}{5}=\frac{2.934}{5}=0.587$$

8. O índice de massa corporal é obtido dividindo o peso de uma pessoa pelo quadrado da sua altura e é usado como um indicador de obesidade. Suponha que a distribuição do índice de massa corporal tem um desvio padrão de 3 kg/m^2, e se deseja estimar a média a partir de uma amostra de dimensão 49.

a. Determine a probabilidade de o valor estimado ter um erro máximo de 1 kg/m^2.

$$P(\text{erro estimado}\le 1)=P\left(\left|\bar{x}-\mu\right|<1\right)=P\left(-1<\bar{x}-\mu<1\right)=P\left(-\frac{1}{\sigma/\sqrt{n}}<\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}<\frac{1}{\sigma/\sqrt{n}}\right)$$

$$P\left(-\frac{1}{3/\sqrt{49}}<z<\frac{1}{3/\sqrt{49}}\right)=P\left(-2.33<z<2.33\right)=P\left(z<2.33\right)-P\left(z<-2.33\right)\overset{tabela 5}{=}0.9901-0.0099=0.9802$$

b. Construa um intervalo de confiança a 99% para o verdadeiro valor médio.

$$IC\text{ a 99\% para }\mu\rightarrow\bar{x}-z_{1-\alpha/2}.\frac{\sigma}{\sqrt{n}}<\mu<\bar{x}+z_{1-\alpha/2}.\frac{\sigma}{\sqrt{n}}$$

$$\bar{x}-2.575.\frac{3}{\sqrt{49}}<\mu<\bar{x}+2.575.\frac{3}{\sqrt{49}}$$

$$\bar{x}-1.104<\mu<\bar{x}+1.104$$

$$\bar{x}\pm1.104\ kg/m^2$$

c. Qual deveria ser a dimensão da amostra para que se conseguir um intervalo de confiança a 99% e simultaneamente a probabilidade de o valor estimado ter um erro máximo de 1 kg/m^2?

$$z_{1-\alpha/2}.\frac{\sigma}{\sqrt{n}}\le1\rightarrow2.575.\frac{3}{\sqrt{n}}\le1\Leftrightarrow n\ge59.68\Rightarrow n\ge60$$

R: A amostra deveria ter no mínimo 60 indivíduos

9. Considere a variável aleatória peso de componentes electrónicos produzidos por determinada empresa. Pretendendo-se estudar a variabilidade do peso dos referidos componentes, recolheu-se uma amostra de 11 elementos, cujos valores (em gramas) foram: 98, 97, 102, 100, 98, 101, 102, 105, 95, 102, 100.

a. Apresente uma estimativa para a variância do peso dos componentes.

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = \frac{110080 - 11*100^2}{10} = 8.0$$

b. Construa um intervalo de confiança para a variância do peso, com um grau de confiança de 95%.

$$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}} \qquad \chi^2_{n-1,\alpha/2} = \chi^2_{10,0.025} \overset{tabela7}{=} 20.48320 \quad \text{e} \quad \chi^2_{n-1,1-\alpha/2} = \chi^2_{10,0.975} \overset{tabela7}{=} 3.24696$$

$$\frac{10*8}{20.48320} < \sigma^2 < \frac{10*8}{3.24696}$$

$$3.91 < \sigma^2 < 24.64$$

c. Construa um intervalo de confiança a 99% para o desvio padrão do peso.

$$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}} \qquad \chi^2_{n-1,\alpha/2} = \chi^2_{10,0.005} \overset{tabela7}{=} 25.18805 \quad \text{e} \quad \chi^2_{n-1,1-\alpha/2} = \chi^2_{10,0.995} \overset{tabela7}{=} 2.15585$$

$$\frac{10*8}{25.18805} < \sigma^2 < \frac{10*8}{2.15585}$$

$$3.176 < \sigma^2 < 37.108$$

$$1.78 < \sigma < 6.09$$

d. Qual o pressuposto que se deve verificar para que o intervalo de confiança seja válido?

R: A variável peso tem de ser normalmente distribuída.

Questão	1a	1b	1c	1d	1e	2a	2b	2c	3	4a	4b	4c	4d	5ai	5aii	5b	6a	6b	7a	7b	8a	8b	8c	9a	9b	9c	9d
Cotação	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	2	0,5	1	1	0,5	0,75	0,75	0,5	1	1	1,5	0,5	1	0,75	0,75	0,5	0,75	0,75	0,5