Second exam
January 5, 2008
Statistics II

1. (5 points) A uma amostra aleatória de 344 grossistas no sector da distribuição na região Norte do país foi colocada a seguinte questão: "Qual é a política da sua empresa relativamente à aceitação por parte do pessoal do departamento de compras de ofertas dos vendedores?". Para 83 destes grossistas a política consistia em deixar que o pessoal do departamento de compras tomasse a sua decisão.

   (a) Encontre um intervalo de confiança a 90% para a verdadeira proporção de grossistas que permite que o pessoal do departamento de compras tome a decisão de aceitar ou não ofertas dos vendedores.

   Solution: X = número de grossistas na região Norte que permitem que os seus compradores decidam se aceitam ou não ofertas dos vendedores, em 344
   $\Longrightarrow X \sim \mathrm{b}\,(N = 344, p)$

   $$[p]_{90\%} = \left[f_N \pm z_{\alpha/2} \sqrt{\frac{f_N\,(1 - f_N)}{N}} \right]$$

   onde $f_N = 83/344 = 0.24128$ e $z_{\alpha/2} = z_{0.05} = 1.645$.
   Então, $[p]_{90\%} = [0.2033; 0.2792]$

   (b) Determine a dimensão mínima da amostra para que o intervalo de confiança a 90% para a verdadeira proporção de grossistas que permite que o pessoal do departamento de compras tome a decisão de aceitar ou não ofertas dos vendedores tenha uma margem máxima de erro de 3%.

   Solution:
   $$0.03 = z_{\alpha/2} \sqrt{\frac{p\,(1-p)}{N}}$$
   onde $p = 0.5$ para obter o valor máximo. Então,
   $$N = 1.645^2 \frac{0.25}{0.03^2} = 751.69 \approx 752$$

   Na região Sul do país foi colocada a mesma questão acima mencionada a uma amostra de 250 grossistas, verificando-se que 17% permitiam que o seu pessoal do departamento de compras decidisse aceitar ou não ofertas dos vendedores.

   (c) Verifique se é razoável admitir que a diferença entre a proporção de grossistas que são permissivos quanto à política de aceitação de ofertas nas regiões Norte e Sul está aproximadamente entre 3.8% e 10.4% (assuma um nível de confiança de 95%).

Solution: X = número de grossistas na região Norte que permitem que os seus compradores decidam se aceitam ou não ofertas dos vendedores, em 344
$\Longrightarrow X \sim \mathrm{b}\left(N = 344, p_X\right)$
Y = número de grossistas na região Sul que permitem que os seus compradores decidam se aceitam ou não ofertas dos vendedores, em 250
$\Longrightarrow X \sim \mathrm{b}\left(N = 250, p_Y\right)$

$$[p_X - p_Y]_{95\%} = \left[\left(f_{N_X} - f_{N_Y}\right) \pm z_{\alpha/2}\sqrt{\frac{f_{N_X}\left(1 - f_{N_X}\right)}{N_X} + \frac{f_{N_Y}\left(1 - f_{N_Y}\right)}{N_Y}}\right]$$

onde $f_{N_X} = 83/344 = 0.24128$, $f_{N_Y} = 0.17$ e $z_{\alpha/2} = z_{0.025} = 1.96$.
Então, $[p_X - p_Y]_{95\%} = [0.0382; 0.1044]$

2. (5 points) Numa determinada fábrica, quando o processo de produção de rolamentos para esferas está a funcionar correctamente, o peso destes rolamentos segue uma distribuição normal com média 141.75 gramas e desvio-padrão de 2.835 gramas. Foi feito um ajustamento no processo de produção e o gerente da fábrica suspeita que isso levou a um aumento do peso médio dos rolamentos para esferas, deixando o desvio-padrão inalterado. Foi recolhida uma amostra de 16 peças, verificando-se que, em média, o peso de cada rolamento foi de 142.2 gramas.

   (a) Pensa que a suspeita do gerente da fábrica é verdadeira e que, portanto, o peso dos rolamentos para esferas é agora maior? Use um nível de significância de 5%.

Solution: X = pesos dos rolamentos para esferas após o ajustamento
$X \sim \mathcal{N}\left(\mu, \sigma = 2.835\right)$
$H_0 : \mu = 141.75$
$H_1 : \mu > 141.75$

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{N}} \sim \mathcal{N}\left(0, 1\right)$$

$$\mathbb{P}\left(\overline{X} > a \mid \mu = 141.75\right) = 0.05 \quad \Longleftrightarrow \quad \mathbb{P}\left(Z > \frac{a - 141.75}{2.835/\sqrt{16}}\right) = 0.05$$

Usando a tabela da distribuição normal padronizada obtém-se que

$$\frac{a - 141.75}{2.835/\sqrt{16}} = 1.645 \quad \Longleftrightarrow \quad a = 142.9159$$

$$RC = \left\{\overline{X} : \overline{X} > 142.9159\right\}$$

Como $142.2 \notin RC$, não se rejeita H_0
A suspeita do gerente da fábrica não parece ser verdadeira.

(b) Assumindo o mesmo nível de significância da alínea anterior, calcule a probabilidade de concluirmos que o gerente não tem razão, sendo que, de facto, o verdadeiro peso médio dos rolamentos para esferas após o ajustamento é de 143 gramas.

Solution: O que se pede é o calculo do erro tipo II

$$\begin{aligned}\beta &= \mathbb{P}\left(\overline{X} < 142.9159 \mid \mu = 143\right) = \mathbb{P}\left(Z < \frac{142.9159 - 143}{2.835/\sqrt{16}}\right) = \\ &= \mathbb{P}\left(Z < -0.12\right) = 1 - 0.5478 = 0.4522\end{aligned}$$

(c) Represente graficamente a função potência do teste para um nível de significância de 5%. Tem de indicar o valor mínimo da função, um valor intermédio e sabe-se que se o verdadeiro peso médio for igual ou superior a 145.7 a função tem o valor 1.

Solution: Função potência do teste

$$\begin{aligned}\pi\left(\mu\right) &= \mathbb{P}\left(\overline{X} > 142.9159 \mid \mu\right) = \mathbb{P}\left(Z > \frac{142.9159 - \mu}{2.835/\sqrt{16}}\right) = \\ &= 1 - \mathbb{P}\left(Z < \frac{142.9159 - \mu}{2.835/\sqrt{16}}\right)\end{aligned}$$

Alguns valores de referência:

$$\begin{aligned}\pi\left(\mu = 141.75\right) &= 0.05 \\ \pi\left(\mu = 143\right) &= 1 - \beta = 0.5478 \\ \pi\left(\mu = 145.7\right) &= 1\end{aligned}$$

3. (5 points) Considerou-se uma amostra aleatória com informação relativa ao peso dos recém-nascidos e à idade das respectivas mães. Esta última variável encontra-se dividida em três escalões da seguinte forma

- Mães com menos de 25 anos
- Mães com idade compreendida entre 25 e 35 anos
- Mães com mais de 35 anos

Utilizando o SPSS obtiveram-se os seguintes resultados

ANOVA

Peso

Source	Sum of sq.	df	Mean sq.	F-ratio	Sig.
Between group	2.560				.000
Within group					
Total	28.536	593			

(a) Identifique o teste efectuado, o seu objectivo e as hipóteses nula e alternativa.

> **Solution:** O teste efectuado é um teste de análise de variância (ANOVA) cujo objectivo é testar se o peso médio dos recém-nascidos varia consoante a idade das respectivas mães.
>
> $H_0 : \mu_1 = \mu_2 = \mu_3 \qquad H_1 : \mu_1 \neq \mu_2 \cup \mu_1 \neq \mu_3 \cup \mu_2 \neq \mu_3$

(b) Quais as hipóteses assumidas na realização deste teste?

> **Solution:** Normalidade e igualdade de variâncias.

(c) Qual a conclusão a que chega com a realização do teste? Fundamente a sua resposta.

> **Solution:** A tabela completa é

Source	Sum of sq.	df	Mean sq.	F-ratio	Sig.
Between group	2.560	2	1.28	29.12	.000
Within group	25.977	591	0.044		
Total	28.536	593			

> Como a estatística F tem um p-value de 0.000, rejeita-se a hipótese nula. Os pesos médios são diferentes em função da idade das mães.

4. (5 points) Diga se as seguintes afirmações são verdadeiras ou falsas. Justifique as suas respostas.

   (a) Uma empresa de segurança supôs que o número de acidentes que um carro sofre por ano é uma variável aleatória discreta X com domínio $\{0, 1, 2\}$ e com função de probabilidade dada por $f(0) = 3\theta$, $f(1) = \theta$, $f(2) = 1 - 4\theta$. Para avaliar esta hipótese, foi recolhida uma amostra aleatória de 200 carros. Destes 200, 130 não sofreram qualquer acidente num ano, 50 sofreram um acidente e 20 sofreram dois acidentes. Então, se $\theta = 0.2$, a hipótese assumida pelda empresa está correcta para um nível de significância de 5%.

Solution: Como $\theta = 0.2$ temos que

$$\begin{aligned} f(0) &= 3 \cdot 0.2 = 0.6 \implies \hat{E}_0 = 200 \cdot 0.6 = 120 \\ f(1) &= 0.2 \implies \hat{E}_1 = 200 \cdot 0.2 = 40 \\ f(2) &= 1 - 4 \cdot 0.2 = 0.2 \implies \hat{E}_2 = 200 \cdot 0.2 = 40 \end{aligned}$$

A tabela de valores observados e esperados é

	0	1	2
O_i	130	50	20
E_i	120	40	40

A estatística é

$$\chi^2 = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} = 13.33$$

O valor crítico é $\chi^2_{2;0.05} = 5.99 < 13.33 \implies$ rejeita-se H_0
Assim, a afirmação é **FALSA**.

(b) O departamento de recursos humanos quer analisar se há uma relação entre o estado civil dos seus empregados e a continuidade nos seus postos de trabalho. Para realizar este estudo, foi seleccionada uma amostra aleatória de 221 trabalhadores da empresa na última década, conseguindo-se a seguinte informação:

	Solteiros	Casados
Abandonaram a empresa no primeiro ano de contrato	17	12
Abandonaram entre o primeiro e segundo ano	42	34
Não abandonaram nos primeiros dois anos	59	57

Segundo os dados, o estado civil e a continuidade no posto de trabalho são variáveis independentes para um nível de significância de 5%.

Solution: Os valores observados são

	Solteiros	Casados	*Total*
Abandonaram a empresa no 1° ano de contrato	17	12	29
Abandonaram entre o 1° e 2° ano	42	34	76
Não abandonaram nos primeiros dois anos	59	57	116
Total	118	103	221

Os valores esperados são

	Solteiros	Casados	*Total*
Abandonaram a empresa no 1° ano de contrato	15.48	13.52	29
Abandonaram entre o 1° e 2° ano	40.58	35.42	76
Não abandonaram nos primeiros dois anos	61.94	54.06	116
Total	118	103	221

onde $15.48 = \dfrac{118 \cdot 29}{221}$, etc.

O estatístico é

$$\chi^2 = \sum_{i=1}^{3} \sum_{j=1}^{2} \frac{\left(O_{ij} - \hat{E}_{ij}\right)^2}{\hat{E}_{ij}} = 0.72$$

O valor crítico é $\chi_{(3-1)(2-1);0.05} = \chi^2_{2;0.05} = 5.99 > 0.72 \Longrightarrow$ não se rejeita H_0

Assim, a afirmação é **VERDADEIRA**.

(c) Ao testar a hipótese nula que a média duma população normal (μ) é 1, com uma região crítica bilateral, o p-value do teste é 0.038. Podemos, então, assegurar que o intervalo de confiança a 95% para μ não contém o valor 1.

Solution: p-value $= 0.038 > 0.05 \Longrightarrow$ rejeita-se H_0 a 5% $\Longrightarrow$ o intervalo de confiança a 95% para μ não contém o valor 1.

Assim, a afirmação é **VERDADEIRA**.