

**MIEIC — 2017/2018**  
**ESTATÍSTICA**  
**ÉPOCA NORMAL (2018.06.08)**  
**Duração: 2h 30m**  
**(Consulta limitada a Formulários e Tabelas Estatísticas)**

- Em todas as suas respostas defina e apresente, com clareza, a **terminologia** (acontecimentos e variáveis aleatórias), a **metodologia** e as **hipóteses subjacentes aos diferentes passos da sua análise**.
- Nos testes de hipóteses que realizar deve apresentar também o **valor de prova** e uma **conclusão** no contexto do problema.
- Todos os problemas devem ser respondidos em **folhas diferentes**.

1. (4 VALORES) *Redundant Array of Independent Disks* (RAID) é uma tecnologia que usa vários discos para aumentar a velocidade de transferência de dados, possibilitar *backup* de dados instantâneo e fornecer proteção contra falhas. Esta tecnologia tem vários níveis *standard* (RAID 0 a RAID 6) que oferecerem diferentes níveis de fiabilidade, disponibilidade, desempenho e capacidade. Os vários níveis RAID distinguem-se pela utilização, ou não, de *striping* (informação distribuída por vários discos), *mirroring* (replicação da informação por vários discos) e *parity* (verificação de erros). Suponha que a probabilidade de um disco rígido falhar num dia seja 2.5% (nota: valor anormalmente alto para simplificar os cálculos, valores típicos seriam inferiores a 0.01%) e que as falhas dos discos são independentes entre si.

- a) Numa configuração RAID 1 (*mirroring*) com dois discos a informação é duplicada nos dois discos, *i.e.*, um disco é uma cópia integral do outro. Qual é a probabilidade de ocorrer perda de informação num dia? Considere que há perda de informação se os dois discos avariarem no mesmo dia.

**Solução: 0.062%**

- b) Numa configuração RAID 0 (*striping*) com dois discos a informação é distribuída pelos dois discos, permitindo ganhos de desempenho no acesso aos discos. Qual é a probabilidade de ocorrer perda de informação num dia? Considere que há perda de informação se pelo menos um disco avariar num dia.

**Solução: 4.9375%**

- c) Numa configuração RAID 0 0 (*striping*) com dois discos e admitindo que, no final do dia, um disco avariado é substituído por um novo, calcule a probabilidade de num ano (365 dias) haver no máximo 10 dias sem perda de informação.

**Solução:  $\approx 0\%$  (distribuição Binomial com  $N = 365$  e  $p = 1 - 0.049375$ )**

2. (4 valores) Um produto é classificado de acordo com o número de defeitos que tem e com o tamanho do lote onde foi produzido. Sejam  $X$  e  $Y$  duas variáveis aleatórias que representam, respetivamente, o número de defeitos por unidade (podendo tomar os valores 0, 1, 2 e 3) e o tamanho do lote (podendo tomar os valores 500 e 1000). A tabela seguinte representa a função de probabilidade conjunta  $p_{XY}(x, y)$ .

$p_{XY}(x, y)$		$Y$	
		500	1000
$X$	0	0.12	0.10
	1	0.10	0.10
	2	0.14	0.14
	3	0.10	0.20

- a) Calcule a probabilidade de um produto, selecionado aleatoriamente, ter pelo menos 2 defeitos

**Solução: 58%**

- b) Calcule o valor esperado e o desvio padrão do número de defeitos por produto.

**Solução:**  $\mu_X = 1.66$  e  $\sigma_X = 1.12$

- c) Pretende efetuar uma grande encomenda deste produto e tem a possibilidade de escolher o tamanho do lote onde os produtos serão produzidos. Escolheria lotes de tamanho 500 ou 1000, ou será que tanto faz? Justifique convenientemente a sua resposta.

**Resposta:** Lote de 500 unidades (lotes de 500 apresentam menor valor esperado condicionado:  $\mu_{X|Y=500} = 1.48 < \mu_{X|Y=1000} = 1.81$ ; necessário também calcular e comentar a influência dos desvios padrões condicionados)

3. (4 valores). Um topógrafo pretende determinar a altura de um monte que se julga inferior a 300 metros de altura. Para tal recorre a um instrumento de medida adequado, o teodolito, que, como qualquer instrumento de medida, tem um erro de medição associado. Para ultrapassar/controlar o erro de medição realiza 10 medições independentes com o mesmo teodolito, tendo obtido os valores apresentados na tabela seguinte.

Medições em metros (X)	301.7	297.2	300.2	296.9	299.2	300.8	296.3	301.6	294.2	296.2
	$\bar{X} = 298.4$					$s = 2.61$				

- a) Teste, ao nível de significância de 5%, se a altura do monte é realmente inferior a 300 metros.

**Solução:**  $H_0: \mu = 300$  e  $H_1: \mu < 300$ ;  $\alpha = 5\%$ ;  $VC = -1.83$ ;  $ET = -1.94$ ;  $pvalue = 4.22\% \rightarrow$  rejeitar  $H_0$

- b) Admita agora que o teodolito está devidamente calibrado e que os erros de medição são independentes, com média igual a 0 metros e desvio padrão igual a 3 metros. Qual deverá ser o número de medições a realizar caso o topógrafo pretenda ter uma confiança de 95% de que a média das medições não está a mais de 0.30 metros do verdadeiro valor da altura do monte.

**Solução:**  $N \geq 384.16 \rightarrow N_{min} = 385$  (trata-se de um I.C. para a medição, com desvio padrão conhecido e onde metade da amplitude do I.C. deverá ser inferior a 0.30 metros)

4. (5 VALORES) O grupo PRESS detém um conjunto de jornais quer em suporte físico quer em suporte digital relacionados com diferentes temáticas (e.g., economia, desporto, cultura). O grupo encontra-se a efetuar um investimento forte nos canais digitais e tinha, no final de 2017, uma quota de mercado de 60% dos jornais digitais em Portugal. Para perceber a tendência do negócio de venda de jornais em suporte digital, o grupo promoveu um inquérito a nível nacional cujos resultados acabaram de ser publicados. O inquérito foi realizado a 5 000 pessoas e verificou-se que 3 150 assinam um jornal digital do grupo PRESS.

- a) Teste, ao nível de significância de 5%, se existe uma tendência crescente no número de assinaturas de jornais digitais do grupo PRESS, com base nos resultados do inquérito obtido.

**Solução:**  $H_0: p = 0.6$  e  $H_1: p > 0.6$ ;  $\alpha = 5\%$ ;  $VC = 1.645$ ;  $ET = 4.33$ ;  $pvalue \approx 0\% \rightarrow$  rejeitar  $H_0$

- b) Se a quota de mercado do grupo PRESS for de 61% e usando a mesma dimensão da amostra, qual a probabilidade de não se detetar que é falsa a hipótese de a quota de mercado ser de 60%.

**Solução:** 55.7% (trata-se do cálculo do erro do tipo II,  $\beta$ )

- c) Qual o número de pessoas com assinaturas de jornais digitais do grupo PRESS a partir do qual se tomaria uma decisão diferente da tomada na alínea a).

**Solução:**  $N_{min} = 3057$

- d) Por forma a conhecer a notoriedade dos jornais em suporte físico, o grupo PRESS promoveu um inquérito adicional. O inquérito abrangeu 4 000 pessoas e verificou-se que 2 450 adquiriram recentemente um jornal em suporte físico do grupo. Construa um intervalo de confiança a 90% para a diferença entre as proporções de pessoas que têm uma assinatura de um jornal digital e de pessoas que adquiriram recentemente um jornal em suporte físico.

**Solução:** [0.00056, 0.03443]

5. (3 VALORES) Relativamente à disciplina Métodos Estatísticos a tabela seguinte apresenta a média das horas de estudo semanais e as classificações obtidas em exame, de uma amostra aleatória de 10 alunos inscritos na disciplina no ano letivo 2016/2017.

Horas de estudo semanais (X)	2.5	5.0	6.0	3.5	0.0	1.5	3.0	4.0	6.5	8.0
Classificação no exame (Y)	10.5	15.0	15.2	11.0	5.0	6.1	12.0	13.6	17.0	18.2
$S_{X,X} = 53.0$	$S_{X,Y} = 92.2$		$S_{Y,Y} = 171.0$		$\bar{X} = 4.0$		$\bar{Y} = 12.36$			

- a) Estime e teste, ao nível de significância de 5%, a relação entre a classificação que um aluno tem no exame da disciplina e as horas semanais que, em média, dedicou ao seu estudo. Interprete as estimativas obtidas para os parâmetros do modelo de regressão obtido.

*Solução:*  $Y = 12,36 + 1,74 \cdot (X - 4)$  (pvalue do teste ao  $\beta$ :  $< 0.0005$ )

- b) Admita que a relação estimada, entre a média das horas de estudo semanais e as classificações obtidas em exame, é válida para qualquer ano letivo.
- Com base nas horas que, em média, dedicou por semana ao estudo desta disciplina, estime a classificação que irá obter no presente exame.
  - Quantas horas a mais por semana teria, em média, de ter estudado para obter uma classificação de 20 valores?
  - Considere a situação de um colega que estudou 12 horas por semana. Comente sobre a adequabilidade da utilização do modelo obtido na alínea a) nesta situação.

*Resposta: ...*