INSTITUTO POLITÉCNICO DE BRAGANÇA — Escola Superior de Tecnologia e de Gestão	Estatística Engenharia Informática 23 junho 2023 Prova Época Normal

Consulta de Tabelas e Formulário autorizada. É obrigatória a devolução do enunciado, devidamente identificado. Utilização de telemóveis e internet não autorizada. Duração: Parte 2 (50%): 60 min; Global (100%): 120 min.

Nome:____ Número:________________

Se pretender fazer a <u>Prova 2 (avaliação distribuída) responda apenas aos grupos 3 e 4.</u>

1. Uma empresa de segurança aeroportuária está a testar um dispositivo para detetar materiais explosivos através do odor do passageiro quando este passa por uma porta de embarque. Todos os passageiros têm de passar por esta porta, devendo o dispositivo fazer atuar um alarme sonoro sempre que detetar algum material explosivo. Como o sistema ainda está em testes, não é perfeitamente fiável, fazendo soar falsamente o alarme em 2% das situações e não o fazendo em 8% dos casos em que o passageiro transporta material explosivo.
Assuma que 1 em cada 10000 passageiros transporta consigo algum material explosivo. Se um passageiro qualquer, escolhido aleatoriamente, passar pela porta de embarque referida:
1.1 Qual a probabilidade de o alarme tocar?
1.2 Se o alarme for acionado, qual é a probabilidade de que o passageiro esteja a transportar material explosivo?
1.3 Os acontecimentos "o alarme ser acionado" e "o passageiro transportar explosivos" são independentes? Justifique.

2. Um vendedor de automóveis adquire, no início de cada mês, 4 carros usados por 25000 € por unidade a um stand da sua confiança. Durante um mês, o vendedor em questão contacta potenciais clientes para vender os referidos automóveis a 27500 € cada. O número de automóveis que consegue vender durante um mês segue a função de probabilidade que se indica na tabela seguinte. No fim de cada mês o vendedor revende ao stand os automóveis que não conseguiu vender a 22500 € cada.

y	0	1	2	3	4
p(y)	0.10	0.15	0.45	0.20	0.10

2.1 Verifique se a atividade do vendedor, nas condições descritas anteriormente, é rentável.
2.2 O referido vendedor considera que a sua atividade só será viável se conseguir um lucro médio mensal de 2500 €. Dado que não pretende aumentar o seu preço de venda com receio de uma quebra na procura, o vendedor quer propor um novo preço de revenda a praticar pelo stand, de forma a atingir o seu objetivo. Qual deverá ser esse valor de revenda mínimo dos automóveis que não conseguiu vender?

3. Uma pesquisa científica concluiu que a quantidade de riboflavina (vitamina B2) em ovos caseiros, segue uma distribuição normal com um valor médio de 0.25 mg e um desvio padrão de 0.02 mg. Considerando este tipo de ovos:

3.1 Qual é a proporção de ovos que possuem menos do que 0.20 mg de riboflavina?

3.2 Caracterize a distribuição da quantidade média de riboflavina numa embalagem de 30 ovos (determine os seus parâmetros principais).

3.3 Qual é a probabilidade de que a quantidade média de riboflavina na embalagem de 30 ovos estar compreendida entre 0.250 e 0.255 mg?

4. Responsáveis governamentais garantem que as exportações nacionais têm tido um crescimento significativo. A conclusão foi baseada num estudo conduzido em 12 empresas exportadoras selecionadas ao acaso. No estudo em questão foi registado, para cada empresa, a variação percentual das exportações. Os dados recolhidos permitiram determinar que a amostra possui uma variação média percentual de 0.3500 com um desvio padrão de 0.8062.

4.1 Recorrendo à metodologia dos testes de hipóteses, verifique utilizando um nível de significância de 5%, se os dados suportam a afirmação dos responsáveis governamentais. Apresente o valor de prova do teste.

4.2 Faça novamente a verificação da alínea anterior, mas recorrendo a um intervalo de confiança.

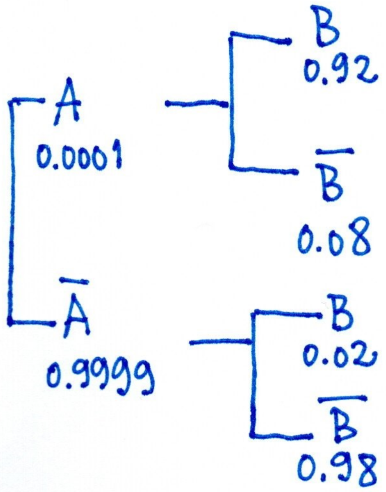
4.3 Se, de facto, as exportações crescerem 0.5% e pretendendo-se detetar tal facto com uma probabilidade não inferior a 90%, qual seria o tamanho mínimo necessário da amostra? (Admita, que nesta situação, é válida a aproximação $\sigma \approx s$)

Questão	1.1	1.2	1.3	2.1	2.2	3.1	3.2	3.3	4.1	4.2	4.3
Cotação(%)	10	10	10	10	10	8	8	8	10	8	8

① Utilizando um diagrama em árvore:

A: transporta explosivos

B: O alarme é acionado



1.1 $P(B) = P(B \cap A) + P(B \cap \bar{A})$

$$= P(A) \times P(B|A) + P(\bar{A}) \times P(B|\bar{A})$$

$$= 0.92 \times 0.0001 + 0.02 \times 0.9999$$

$$= 0.02009$$

1.2 $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

$$= \frac{P(A) \times P(B|A)}{P(B)}$$

$$= \frac{0.0001 \times 0.92}{0.02009}$$

$$= 0.00458$$

1.3 Os acontecimentos A e B serão independentes se $P(A \cap B) = P(A) \times P(B)$ ou por outras palavras $P(B|A) = P(B)$

Como $P(B|A) = 0.92$ e $P(B) = 0.02009$, os acontecimentos não são independentes.

(2)

2.1) | compra : 10 000 €/4 unid.
venda : 27500 €/unid.
revenda : 22500 € /unid. |

A atividade será rentável se o lucro for positivo.

lucro = proveitos − custos = venda + revenda − compra

$L = 27500Y + 22500 \times (4-Y) - 100000$

Y: nº de unidades vendidas/mês ($y = \{0,1,2,3,4\}$)

y	p(y)	L	L×p(y)
0	0.10	−10 000	−1 000
1	0.15	−5 000	−750
2	0.45	0	0
3	0.20	5 000	1 000
4	0.10	10 000	1000
		Σ	250

$\rightarrow \sum L \cdot p(y)$ = "lucro médio" é positivo

A atividade é rentável

2.2) Neste caso o valor (R) da revenda é desconhido.

$27500 \times 0 + R \times (4-0) - 100000$

y	p(y)	L	L×p(y)
0	0.10	4R − 100 000	0.4R − 10000
1	0.15	3R − 72 500	0.45R − 10 875
2	0.45	2R − 45 000	0.9R − 20 250
3	0.20	R − 17 500	0.2R − 3500
4	0.10	10 000	1000
		Σ	1.95R − 43625

Para o lucro médio ser positivo: $1.9R - 43625 > 0 \rightarrow R = 22961$ €

Para o lucro médio ser 2500 €: $1.9R - 43625 = 2500$

$\Leftrightarrow R = 24276$ €

③

3.1) X : Quantidade de riboflavina nos ovos (em mg)

$$X \sim N\left(\mu_X = 0.25 \; ; \; \sigma_X^2 = 0.02^2\right)$$

$$P(X < 0.20) = P\left(Z < \frac{0.20 - 0.25}{0.02}\right) = P(Z < -2.5) =$$

$$= P(Z > 2.5) = 0.00621$$

3.2) $\bar{X}$: Quantidade média de riboflavina numa embalagem de 30 ovos. (em mg)

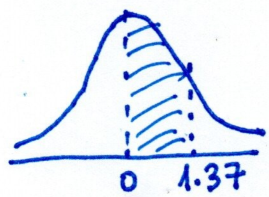
$$\bar{X} = \sum_{i=1}^{30} X_i / 30$$

Pelo T.L.C. $\bar{X} \sim N\left(\mu_{\bar{X}} = 0.25 \; ; \; \sigma^2_{\bar{X}} = \frac{0.02^2}{30}\right)$

$$\sigma^2_{\bar{X}} = \frac{0.02^2}{30} = 0.000013 \text{ mg}^2$$

3.3) $$P(0.250 < \bar{X} < 0.255) = P\left(\frac{0.250 - 0.25}{0.02/\sqrt{30}} < Z < \frac{0.255 - 0.25}{0.02/\sqrt{30}}\right)$$

$$= P(0 < Z < 1.3693) =$$



$$\cong P(Z > 0) - P(Z > 1.37)$$

$$= 0.5000 - 0.0853 = 0.4147$$

④

4.1) X: Variação percentual das vendas (%)

$X \sim N(\mu, \sigma^2)$

<u>amostra</u>: $n = 12$, $\bar{x} = 0.3500$, $s = 0.8062$

<u>T.H.</u>: $\begin{cases} H_0: \mu = 0 \\ H_1: \mu > 0 \end{cases}$ (teste à afirmação)

Se H_0 for verdadeira, $ET = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$

H_0 será rejeitada se $ET > ET(\alpha)$ → teste unilateral à direita

$ET(\alpha) = t_{11}(0.05) = 1.7959$

$ET = \dfrac{0.3500 - 0}{0.8062/\sqrt{12}} = 1.5038$

Como $ET < ET(\alpha)$ não se deve rejeitar H_0 (não existem evidências estatísticas, a um nível de significância de 5%, de que a afirmação seja verdadeira)

Valor prova $= P(t_{11} > 1.5038) \simeq \dfrac{0.10 + 0.05}{2} = 0.075$ $(> \alpha)$

4.2) Neste caso, o I.C. é dado por: $\left[\bar{x} - t_{n-1}(\alpha) \times \dfrac{s}{\sqrt{n}} ; +\infty\right[$

$\left[0.3500 - 1.7959 \times \dfrac{0.8062}{\sqrt{12}} ; +\infty\right[$

$[-0.0680 ; +\infty[$

Como o valor definido na hipótese nula $(\mu = 0)$ está dentro do I.C., então a hipótese nula não deve ser rejeitada. (Conclusão idêntica à da alínea 4.1)

4.3) Dimensionamento da amostra de forma a que a P.T. do teste, quando o crescimento é de 0.5%, seja superior a 90%.

$1-\beta \geq 0.90 \Leftrightarrow P(\text{rejeitar } H_0 \mid H_0 \text{ falsa}) \geq 0.90$

T.H.:

$\begin{cases} H_0: \mu = 0 \\ H_1: \mu = 0.50 \end{cases}$

H_0 será rejeitada quando $ET > t_{n-1}(\alpha)$

Como $\sigma \approx s$ então podemos simplificar e o novo critério de rejeição será: $ET > z(\alpha)$

$$\frac{\bar{x}-\mu_0}{s/\sqrt{n}} > z(\alpha) \Leftrightarrow \bar{x} > \mu_0 + z(\alpha) \times \frac{s}{\sqrt{n}}$$

então, $1-\beta \geq 0.90 \Leftrightarrow P\left(\bar{x} > \mu_0 + z(\alpha) \times \frac{s}{\sqrt{n}} \;\middle|\; H_0 \text{ falsa}\right) \geq 0.90$

$$(\ldots) \Leftrightarrow P\left(Z > \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z(\alpha)\right) \geq 0.90$$

$$\Leftrightarrow \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z(\alpha) \leq z(0.90) \Leftrightarrow n \geq \left(\frac{\sigma \times (z(0.90) - z(\alpha))}{\mu_0 - \mu_1}\right)^2$$

$$\Leftrightarrow n \geq \left(\frac{0.8062 \times (-1.2816 - 1.6449)}{0 - 0.50}\right)^2$$

$$\Leftrightarrow n \geq 22.3 \Leftrightarrow n \geq 23$$

INSTITUTO POLITÉCNICO DE BRAGANÇA — ipb — Escola Superior de Tecnologia e de Gestão	Estatística Engenharia Informática 7 julho 2023 Prova Época Recurso

Consulta de Tabelas e Formulário autorizada. É obrigatória a devolução do enunciado, devidamente identificado. Utilização de telemóveis e internet não autorizada. Duração: 120 min. Justifique convenientemente todas as suas respostas.

Nome:____ Número:_______________

1. As empresas Caracol Expressos e a Sempressas Bus monopolizam o transporte de passageiros em autocarro entre o Porto e Bragança. As duas empresas competem relativamente à pontualidade que anunciam aos seus clientes. Recolheram-se alguns dados referentes à pontualidade das viagens realizadas pelas duas empresas entre estas duas cidades (apenas sentido Bragança), tendo-se verificado que a Caracol Expressos representa 65% do total das viagens. Os dados recolhidos mostram que as viagens da Caracol Expressos possuem atrasos de tempo em 66% das ocasiões enquanto esse valor é de 40% para a Sempressas Bus.

1.1 Apresente um diagrama em árvore que traduza a informação do enunciado, definindo os acontecimentos correspondentes.

1.2 Se selecionar aleatoriamente um autocarro que realizou a referida rota e pertence a uma destas duas empresas, qual a probabilidade de estar atrasado?

1.3 Se um autocarro proveniente de uma qualquer destas empresas, chegar atrasado a Bragança, qual é a probabilidade de pertencer à Caracol Expressos?

1.4 Se um autocarro chegar a Bragança e não estiver atrasado, qual é a probabilidade de pertencer à Sempressas Bus?

2. Três cartas são retiradas aleatoriamente e sem reposição de um baralho normal contendo 52 cartas. Considere que Y representa o número de copas obtido. Nestas condições, responda às seguintes questões.

2.1 A variável Y é discreta ou contínua?

2.2 Determine a função de probabilidade e a função distribuição de probabilidade da variável Y. Apresente um esboço gráfico de ambas as funções.

2.3 Determine os parâmetros de localização (média, mediana e moda) e de dispersão (variância), para a variável Y.

2.4 Qual a probabilidade de se obterem pelo menos 2 copas?

3. Admita que a quantidade de proteína no hambúrguer “Diabólico” comercializado por uma cadeia internacional de fast-food segue aproximadamente uma distribuição normal com um valor médio de 21.0 g e um desvio padrão de 0.8 g.

3.1 Se um hambúrguer for escolhido aleatoriamente de entre os “diabólicos”, qual é a probabilidade de conter menos do que 20.0 g?

3.2 Qual é a distribuição da quantidade total de proteína em 2 hambúrgueres escolhidos aleatoriamente de entre os “diabólicos”?

3.3 Qual é a distribuição da quantidade média de proteína em 2 hambúrgueres escolhidos aleatoriamente de entre o referido tipo de hambúrgueres? Defina o seu valor médio e variância)

3.4 Se forem escolhidos aleatoriamente 2 hambúrgueres de entre os “diabólicos”, qual é a probabilidade de a quantidade média de proteína ser pelo menos 20.5 g?

4. Pretende-se calcular um intervalo de confiança a 99% para a proporção de economistas que discorda com a subida das taxas de juros bancários como forma de conter a subida da inflação.

4.1 Numa sondagem de opinião envolvendo 400 economistas, 150 manifestaram concordância com a medida económica referida. Nestas condições, determine o intervalo de confiança para a proporção de economistas que concorda com esta medida.

4.2 Qual o tamanho mínimo da amostra de economistas a inquirir de modo que o intervalo a calcular tenha uma amplitude inferior a 5 pontos percentuais.

Questão	1.1	1.2	1.3	1.4	2.1	2.2	2.3	2.4	3.1	3.2	3.3	3.4	4.1	4.2
Cotação(%)	8	8	8	6	6	8	6	6	8	8	6	6	8	8

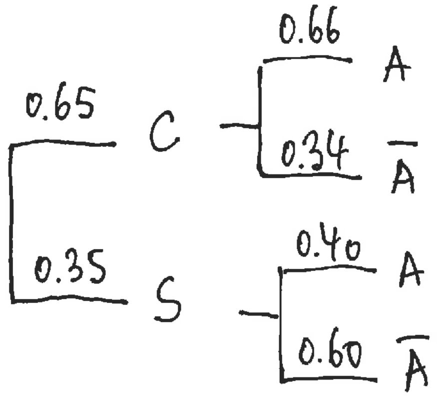
①

1.1) Acontecimentos:

C : O transporte é efetuado pela Caracol Expresso

S : O transporte é efetuado pela Sempressas Bus

A : A viagem é efetuada com atraso na chegada



1.2) $P(A) = P(A \cap C) + P(A \cap S) = P(A|C) \times P(C) + P(A|S) \times P(S)$

$= 0.66 \times 0.65 + 0.40 \times 0.35 = 0.569$

1.3) $P(C|A) = \dfrac{P(C \cap A)}{P(A)} = \dfrac{P(A|C) \times P(C)}{P(A)} = \dfrac{0.66 \times 0.65}{0.569} = 0.754$

1.4) $P(S|\bar{A}) = \dfrac{P(S \cap \bar{A})}{P(\bar{A})} = \dfrac{P(\bar{A}|S) \times P(S)}{1 - P(A)} = \dfrac{0.60 \times 0.35}{1 - 0.569} = 0.487$

(2)

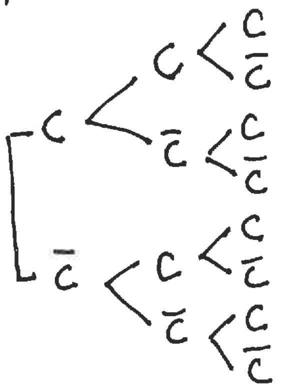
2.1) A variável Y é uma variável discreta já que toma apenas valores inteiros: $y = \{0, 1, 2, 3\}$

2.2) Função de probabilidade: $p(y) = P(Y = y)$

Função distribuição de probabilidade: $F(y) = P(Y \leq y)$

52 cartas: 13 cartas/naipe (39 não copas + 13 copas)

$$p(0) = \frac{39}{52} \times \frac{38}{51} \times \frac{37}{50} = 0.4135$$



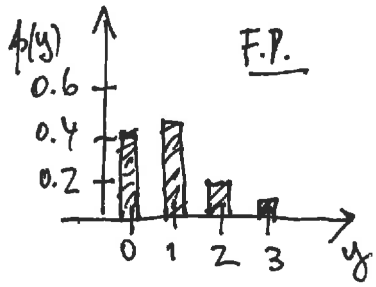
(sugestão)

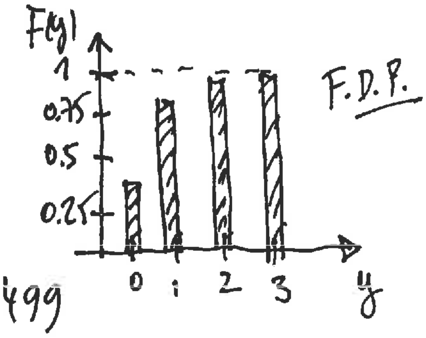
$$p(1) = \frac{13}{52} \times \frac{39}{51} \times \frac{38}{50} \times 3 = 0.4360$$

$$p(2) = \frac{13 \times 12}{52 \quad 51} \times \frac{39}{50} \times 3 = 0.1376$$

$$p(3) = \frac{13}{52} \times \frac{12}{51} \times \frac{11}{50} = 0.0129$$

y	$p(y)$	$F(y)$	$y \times p(y)$	$y^2 \times p(y)$
0	0.4135	0.4135	0	0
1	0.4360	0.8495	0.4360	0.4360
2	0.1376	0.9871	0.2752	0.5504
3	0.0129	1.000	0.0387	0.1161
Σ	1.000		0.7499	1.1025





2.3) Moda = 1 ; Mediana = 1 ; Média $= \mu_Y = \sum y \times p(y) = 0.7499$

Variância $= \sum y^2 \times p(y) - \mu_Y^2 = 1.1025 - 0.7499^2 = 0.5401$

2.4) $P(Y \geq 2) = p(2) + p(3) = 0.1376 + 0.0129 = 0.1506$

3

3.1) X: Quantidade de proteína no hamburguer "diabólico" (em g)

$X \sim N(\mu_X = 21.0 \,;\, \sigma_X^2 = 0.8^2)$

$P(X < 20.0) = P\left(Z < \frac{20.0 - 21.0}{0.8}\right) = P(Z < -1.25) = P(Z > 1.25) = 0.1056$

3.2) S: Soma da quantidade de proteína em 2 hamburgueres "diabólicos" (g)

$S = X_1 + X_2$ (2 variáveis Normais)

$$\left| \begin{array}{l} \mu_S = 2 \times \mu_X = 42.0\ g \\ \sigma_S^2 = 2 \times \sigma_X^2 = 2 \times 0.8^2 = 1.28\ g^2 \end{array} \right. \rightarrow S \sim N(\mu_S = 42.0 \,;\, \sigma_S^2 = 1.28)$$

3.3) $\overline{X}$: Quantidade média de proteína em 2 hamburgueres "diabólicos" (g)

$\overline{X} = \frac{X_1 + X_2}{2}$ pelo T.L.C. $\mu_{\overline{X}} = \mu_X = 21.0\ g$

$\sigma_{\overline{X}}^2 = \frac{\sigma_X^2}{n} = \frac{0.8^2}{2} = 0.32\ g^2$

$\overline{X} \sim N(\mu_{\overline{X}} = 21.0 \,;\, \sigma_{\overline{X}}^2 = 0.32)$

3.4) $P(\overline{X} \geq 20.5) = P\left(Z \geq \frac{20.5 - 21.0}{\sqrt{0.32}}\right) = P(Z \geq -0.88)$

$= 1 - P(Z < -0.88) = 1 - P(Z > 0.88) =$

$= 1 - 0.1894 = 0.8106$

4

4.1) Trata-se de um intervalo de confiança para a proporção Binomial. Como a amostra é de grande dimensão ($n=400$), o intervalo é dado por: $\hat{p} \pm z(\alpha/2) \times \sqrt{\frac{\hat{p} \times (1-\hat{p})}{n}}$; $\hat{p} = \frac{y}{n} = \frac{150}{400} = 0.375$

Como $1-\alpha = 0.99 \Leftrightarrow \alpha = 0.01 \rightarrow z(\alpha/2) = z(0.005) = 2.5758$

o intervalo de confiança será: $0.375 \pm 2.5758 \times \sqrt{\frac{0.375 \times (1-0.375)}{400}}$

$0.375 \pm 0.062 \rightarrow [0.313 ; 0.437]$

4.2) A amplitude do intervalo de confiança é dada por:

$A = 2 \times z(\alpha/2) \times \sqrt{\frac{\hat{p} \times (1-\hat{p})}{n}}$ e resolvendo em ordem a $\underline{n}$:

$$n = \frac{(2 \times z(\alpha/2))^2 \times \hat{p} \times (1-\hat{p})}{A^2}$$

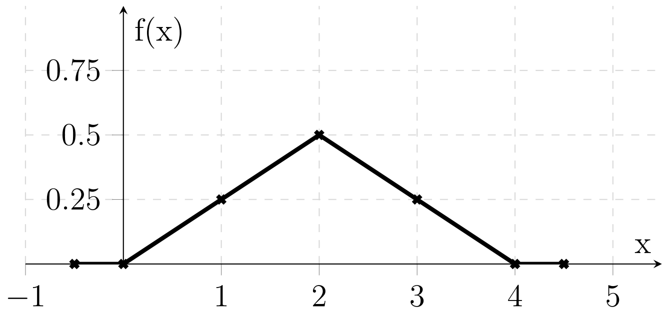
Para garantir que $A < 0.05$, deve-se utilizar $\hat{p} = 0.5 = \hat{q}$

$$n = \frac{(2 \times 2.5758)^2 \times 0.5 \times (1-0.5)}{0.05^2} \Leftrightarrow n = 2653.9 \rightarrow \underline{n \geq 2654}$$

Se utilizarmos como estimativa $\hat{p} = 0.375$,
a dimensão mínima será ligeiramente inferior: $\underline{n \geq 2488}$

Avaliação contínua de Estatística	
2019/2020	Cursos: EC, IG, EI, C
Ficha sobre variáveis aleatórias	**Notas**: Consulta bibliográfica autorizada apenas ao Formulário e Tabelas. Indique com clareza quais as hipóteses subjacentes às análises que efetuar.
novembro de 2019	
Duração: 15 minutos	
Número:	Nome:

1. A variável aleatória X representa as vendas semanais (em UM) de determinado acessório eletrónico da marca Applex Lda. À variável X está associada a função densidade de probabilidade, f(x), que a seguir se representa.



Determine as seguintes alíneas:

a) Comprove que $f(x)$ é uma função densidade de probabilidade.

$$\int_0^4 f(x)\mathrm{d}x = 1 \Leftrightarrow 4 \times \frac{0.5}{2} = 1$$

b) Qual a probabilidade da marca vender mais do que 3 UM?

$$P(x > 3) = \int_3^4 f(x)\mathrm{d}x = 1 \times \frac{0.25}{2} = 0.125$$

c) Qual a probabilidade da marca vender exatamente 3 UM?

$$P(x = 3) = 0$$

d) Qual a média das vendas por semana do referido acessório?

$\mu_X = 2$

2. Considere que Y representa a variável aleatória lucro anual (em UM) de uma empresa que iniciou atividade há 9 anos. Suponha que esta empresa obteve os seguintes lucros anuais:

0.5	0.5	0.2
0.8	0.2	-0.1
0.2	0.5	-0.1

Calcule as seguintes alíneas:

a) Defina a função de probabilidade da variável Y.

y	$P(y)$
-0.1	2/9
0.2	3/9
0.5	3/9
0.8	1/9

b) Qual a probabilidade da empresa perder dinheiro num ano?

$P(y < 0) = 2/9$

c) Qual a probabilidade do lucro anual ser superior a 0.2 UM ?

$P(y > 0.2) = 4/9$

d) Qual o lucro médio anual da empresa?

$$\mu_Y = \sum_y x \cdot P(x) = -0.1 \cdot \frac{2}{9} + 0.2 \cdot \frac{3}{9} + 0.5 \cdot \frac{3}{9} + 0.8 \cdot \frac{1}{9} = 0.3$$

INSTITUTO POLITÉCNICO DE BRAGANÇA
Escola Superior de Tecnologia e Gestão

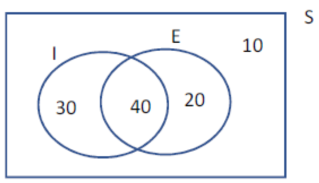
Avaliação contínua de Estatística	
2019/2020	Cursos: EC, IG, EI, C
Ficha sobre Teoria da Probabilidade	**Notas**: Consulta bibliográfica autorizada apenas ao Formulário e Tabelas. Indique com clareza quais as hipóteses subjacentes às análises que efetuar.
novembro de 2019	
Duração: 15 minutos	
Número:	Nome:

1. Uma escola tem 100 alunos dos quais 60 estudam Estatística, 70 estudam Inglês e 40 estudam simultaneamente Estatística e Inglês. O diretor da escola resolveu escolher um aluno, ao acaso, para representar os alunos numa comissão. Determine:

   a) Qual a probabilidade do aluno estudar Estatística?

   E: Estudar estatística

   I: Estudar inglês



   P(I)=0.7

   P($E \cap I$)=0.4

   P(E)=60/100=0.6

   b) Qual a probabilidade do aluno estudar Estatística ou Inglês?
   P($E \cup I$)=P(E)+P(I)-P($E \cap I$)=0.6+0.7-0.4=0.9

   c) Qual a probabilidade do aluno não estudar nenhuma destas disciplinas?
   P($\bar{E} \cap \bar{I}$)=1-P($E \cup I$)=1-0.9=0.1

   d) Qual a probabilidade do aluno estudar Estatística, sabendo que estuda Inglês?

   P(E|I)=P($E \cap I$)/P(I)=0.4/0.7=0.5714

Estatística - 15 de janeiro de 2015

Proposta de Resolução

Enunciado A

1. Representem X_A e X_B os consumos diários (em kWh) nos edifícios A e B, respetivamente. Do enunciado sabe-se que $X_A \sim \mathrm{N}\left(\mu_A = 620,\, \sigma_A = 15\right)$ e sabe-se que $X_B \sim \mathrm{N}\left(\mu_B = 580,\, \sigma_B = 10\right)$.

   a) Pretende-se calcular $P\left(X_B > X_A\right)$. Então

   $$P\left(X_B > X_A\right) = P\left(X_B - X_A > 0\right)$$

   Fazendo $X_B - X_A = X_{B-A}$, a nova variável X_{B-A} resulta de uma combinação linear de v.a. Normais, sendo também uma v.a. Normal com

   $$\mu = \mu_A - \mu_B = 580 - 620 = -40 \text{ e com}$$

   $$\sigma^2 = \sigma_B^2 + \sigma_A^2 = 10^2 + 15^2 = 325$$

   Então,

   $$P\left(X_{B-A} > 0\right) = P\left(Z > \frac{0 - -40}{\sqrt{325}}\right) = P\left(Z > 2.2188\right) = 0.0133$$

   b) Pretende-se calcular $P\left(X_B > 600\right)$. Então,

   $$P\left(X_B > 600\right) = P\left(Z > \frac{600 - 580}{10}\right) = P\left(Z > 2\right) = 0.0228$$

   c) Represente-se por Y o número de dias, em 30, em que o consumo de 600 kWh é ultrapassado. Admitindo que os consumos em dias consecutivos são independentes, então $Y \sim \mathrm{B}\left(n = 30,\, p\right)$ com $p = P\left(X_B > 600\right) = 0.0228$, conforme calculado na alínea anterior. Pretende-se calcular $P\left(Y \geq 1\right)$. Utilizando a distribuição Binomial,

   $$P\left(Y \geq 1\right) = 1 - P\left(Y \leq 0\right) = 1 - F(0) = 1 - 0.5014 = 0.4986$$

2. Represente Y o número de acidentes num dia e represente X o intervalo de tempo (em dias) entre acidentes. Do enunciado sabe-se que $Y \sim \mathrm{Poisson}\left(\lambda = 0.01\right)$ e que $X \sim \mathrm{EN}\left(\lambda = 0.01\right)$.

   a) Represente Y_m o número de acidentes mensais. Então,

   $$Y_m \sim \mathrm{Poisson}\left(\lambda = 0.01 \times 30 = 0.3\right)$$

Pretende-se calcular $P(Y_m \geq 1)$. Utilizando a distribuição de Poisson,

$$P(Y_m \geq 1) = 1 - P(Y_m \leq 0) = 1 - F(0) = 1 - 0.7408 = 0.2592$$

b) Como o processo de Poisson não tem memória, o facto de não terem ocorrido acidentes até maio é irrelevante para o cálculo da probabilidade de acidente em junho. Logo, a resposta é $P(Y_m = 0) = 0.7408$. Alternativamente, utilizando a distribuição Exponencial Negativa, poder-se-ia calcular

$$\begin{aligned} P(X \geq 180|X \geq 150) &= \frac{P(X \geq 180 \cap X \geq 150)}{P(X \geq 150)} = \frac{P(X \geq 180)}{P(X \geq 150)} = \\ &= \frac{1 - F(180)}{1 - F(150)} = \frac{1 - 0.8347}{1 - 0.7769} = 0.7408 \end{aligned}$$

c) Esta probabilidade pode ser calculada pela distribuição Exponencial Negativa:

$$P(X \geq 180) = 1 - F(180) = 1 - 0.8347 = 0.1653$$

d) O número médio de acidentes por ano (λ_a) é proporcional ao número médio de acidentes diários, ou seja,

$$\lambda_a = \lambda \cdot 30 \cdot 12 = 0.01 \times 360 = 3.6$$

O tempo médio entre acidentes é o inverso do número médio de acidentes, ou seja, $1/\lambda = 1/0.01 = 100$ dias.

3. Definam-se os seguintes acontecimentos:

GI: ocorrência de gémeos idênticos, com $P(GI) = 0.3$;

$GF = \overline{GI}$: ocorrência de gémeos fraternos, com $P(GF) = 1 - P(GI) = 0.7$;

MM: gémeos do sexo masculino, com $P(MM|GI) = 0.5$ e com $P(MM|GF) = 0.25$;

FF: gémeos do sexo feminino, com $P(FF|GI) = 0.5$ e com $P(FF|GF) = 0.25$;

MF: um gémeo masculino e outro feminino, com $P(MF|GF) = 0.5$.

Pretende-se calcular $P(GI|MM)$. Pela lei da probabilidade condicionada (ou desenvolvendo diretamente com o teorema de Bayes),

$$\begin{aligned} P(GI|MM) &= \frac{P(GI \cap MM)}{P(MM)} = \frac{P(GI \cap MM)}{P(GI \cap MM) + P(GF \cap MM)} \\ &= \frac{P(MM|GI) \cdot P(GI)}{P(MM|GI) \cdot P(GI) + P(MM|GF) \cdot P(GF)} \\ &= \frac{0.5 \times 0.3}{0.5 \times 0.3 + 0.25 \times 0.7} = 0.4615 \end{aligned}$$

4. Represente X o número de funcionários de uma loja. Sabe-se que $X = Y + 1$ e que $Y \sim \mathrm{B}(n = 5,\, p = 0.6)$.

   a) Como Y varia entre 0 e $n = 5$, então, $X = Y+1$ vai variar entre 1 e $n+1 = 6$. Logo, o número mínimo de funcionários por loja é 1 e o número máximo de funcionários por loja é 6.

   b) Pretende-se calcular $P(X = 4)$. Utilizando a distribuição Binomial,

   $$P(X = 4) = P(Y + 1 = 4) = P(Y = 3) = f(3) = 0.3456$$

   c) Represente-se por S o número total de funcionários a trabalhar no centro comercial. S resulta da soma dos funcionários das 240 lojas, representados por X_1, X_2, ..., X_{240}. Admitindo que o número de funcionários das várias lojas é independente entre si, temos uma soma de variáveis independentes e identicamente distribuídas. Logo, $S = X_1 + X_2 + \cdots + X_{240}$. Pelo teorema do limite central, $S \approx$ Normal com $\mu_S = 240 \cdot \mu_X$ e com $\sigma_S^2 = 240 \cdot \sigma_X^2$. Os parâmetros de X podem ser obtidos com:

   $$\mu_X = E(Y + 1) = E(Y) + 1 = np + 1 = 5 \times 0.6 + 1 = 4$$

   $$\sigma_X^2 = Var(Y + 1) = Var(Y) = np(1 - p) = 5 \times 0.6 \times (1 - 0.6) = 1.2$$

   Logo, $\mu_S = 240 \times 4 = 960$ e $\sigma_S^2 = 240 \times 1.2 = 288$
   Pretende-se calcular[1] $P(S > 1000 + 0.5)$:

   $$P\,(S > 1000.5) = P\left(Z > \frac{1000.5 - 960}{\sqrt{288}}\right) = P\,(Z > 2.3865) = 0.0085$$

5. Utilizando a variável X para representar o diâmetro das peças produzidas, sabe-se que $X \approx$ Normal, com valor esperado, μ, desconhecido, e desvio padrão $\sigma = 0.05$ mm.

   a) Dado o enunciado, pretende-se testar, com $\alpha = 0.5\%$, as seguintes hipóteses:

   $$H_0 : \mu = 10\,\text{mm}$$
   $$H_1 : \mu \neq 10\,\text{mm}$$

   Como se trata de uma população Normal com variância conhecida, a estatística do teste ET é dada por

   $$ET = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

   Se H_0 for verdadeira, $ET \sim \mathrm{N}(0,\, 1)$. Nestas condições, H_0 deve ser rejeitada se $ET < -z(\alpha/2)$ ou se $ET > z(\alpha/2)$, sendo $\alpha/2 = 0.0025$ e $z(\alpha/2) = 2.807$.

[1] Utilizando a correção de continuidade

No caso da amostra fornecida, $\bar{x} = 9.9764$. Logo,

$$ET = \frac{9.9764 - 10}{0.05/\sqrt{5}} = -1.0554$$

De acordo com o critério definido, a hipótese nula não deve ser rejeitada, não havendo evidência de que a média dos diâmetros se tenha desviado significativamente de 10 mm. Procedendo ao cálculo do valor de prova ($v.p.$):

$$v.p. = 2 \cdot P(Z \geq |ET|) = 2 \cdot P(Z \geq |-1.0554|) = 2 \times 0.1456 = 0.2912$$

b) Há evidência de desvios significativos quando H_0 é rejeitada, ou seja, quando $ET < -z(\alpha/2)$ ou quando $ET > z(\alpha/2)$. Desenvolvendo a primeira condição,

$$ET < -z(\alpha/2) \Leftrightarrow \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < -z(\alpha/2) \Leftrightarrow \bar{x} < \mu - z(\alpha/2)\frac{\sigma}{\sqrt{n}} \Leftrightarrow$$
$$\Leftrightarrow \bar{x} < 10 - 2.807\frac{0.05}{\sqrt{5}} \Leftrightarrow \bar{x} < 9.9372$$

ou seja, $\bar{x}_i = 9.9372$. Desenvolvendo a segunda expressão,

$$ET > z(\alpha/2) \Leftrightarrow \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > z(\alpha/2) \Leftrightarrow \bar{x} > \mu + z(\alpha/2)\frac{\sigma}{\sqrt{n}} \Leftrightarrow$$
$$\Leftrightarrow \bar{x} > 10 + 2.807\frac{0.05}{\sqrt{5}} \Leftrightarrow \bar{x} > 10.0628$$

conclui-se que $\bar{x}_s = 10.0628$.

6. a) Trata-se de um intervalo de confiança para a proporção Binomial. Como se trata de uma amostra de grande dimensão, pode utilizar-se no cálculo a expressão

$$\hat{p} \pm z(\alpha/2)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Na amostra obteve-se $y = 150$ e $n = 400$. Logo, $\hat{p} = y/n = 150/400 = 0.375$. Sabe-se também que $1 - \alpha = 0.95$ e, logo, $z(\alpha/2) = z(0.025) = 1.96$. O intervalo de confiança será

$$0.375 \pm 1.96\sqrt{\frac{0.375(1-0.375)}{400}} \Leftrightarrow [0.3276,\ 0.4224]$$

O resultado exato, caso tivesse sido utilizada a distribuição Binomial diretamente, seria [0.3274, 0.4245].

b) A amplitude do intervalo calculado na alínea anterior é dada por

$$A = 2 \cdot z(\alpha/2)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Resolvendo em ordem a n, fica

$$n = \frac{\left[2 \cdot z(\alpha/2)\right]^2 \hat{p}(1-\hat{p})}{A^2}$$

Para garantir uma amplitude menor que o especificado ($A = 0.05$), em qualquer circunstâncias, deve utilizar-se $\hat{p} = 0.5$. Então,

$$n = \frac{\left[2 \times 1.96\right]^2 \times 0.5 \times (1-0.5)}{0.05^2} = 1536.5835$$

devendo o resultado ser arredondado por excesso, ou seja, $n = 1537$.

Caso se pudesse tomar a amostra da alínea anterior como *amostra piloto* para obtenção de uma estimativa de p, poder-se-ia fazer $\hat{p} = 0.375$. Neste caso o resultado seria $n = 1441$, ligeiramente inferior.

Prova Escrita de Estatística	
2014/2015	Notas
Época Normal	Consulta bibliográfica autorizada. Salvo indicação contrária, trabalhe com um nível de significância de 5%. Grupos de cotações iguais. Embora os cálculos possam ser efetuados no computador, apenas serão avaliadas as respostas na folha de exame. Utilizar uma folha para cada questão. É obrigatória a devolução do enunciado, devidamente identificado.
15 de janeiro de 2015	
Duração: 75 + 75 minutos	
Número: Nome:	P-I P-II

Parte I

1. Uma empresa opera em dois edifícios (A e B) com funcionamento independente. O consumo diário de energia elétrica no edifício A segue uma distribuição Normal com média de 620 kWh e desvio padrão de 15 kWh. No edifício B, o consumo também é normalmente distribuído, mas com média de 580 kWh e desvio padrão de 10 kWh.

   a) Qual a probabilidade de, num dia escolhido ao acaso, o consumo do edifício B ser superior ao consumo do edifício A?

   b) No edifício B, qual a percentagem de dias em que se consome mais do que 600 kWh?

   c) Ainda no edifício B, qual a probabilidade de, durante um mês (30 dias), aquele consumo (600 kWh) ser ultrapassado uma ou mais vezes?

2. Numa implantação fabril, o número médio de acidentes de trabalho por dia é igual a 0.01 e o intervalo de tempo entre acidentes tem uma distribuição Exponencial Negativa. Considere que a empresa labora 30 dias por mês.

   a) Qual a probabilidade de ocorrência de 1 ou mais acidentes de trabalho, num qualquer mês?

   b) No final de maio ainda não tinham ocorrido acidentes desde o início do ano. Qual a probabilidade de terminar o mês de junho sem acidentes?

   c) Qual a probabilidade de não haver acidentes no segundo semestre do ano (julho a dezembro)?

   d) Qual o número médio de acidentes de trabalho por ano e qual o tempo médio entre acidentes.

3. Sabe-se que cerca de 30% dos gémeos são *idênticos*, sendo os restantes, cerca de 70%, *fraternos*. Quando os gémeos são idênticos, são necessariamente do mesmo sexo – metade das vezes do sexo feminino e metade das vezes do sexo masculino. Quando se trata de gémeos fraternos, 1/4 das vezes são ambos do sexo feminino, 1/4 das vezes são ambos do sexo masculino e metade das vezes são um de cada sexo. Um casal acabou de ter gémeos e são ambos do sexo masculino. Sem mais informação, qual a probabilidade de serem gémeos idênticos?

Parte II

4. Num determinado centro comercial vão-se instalar 240 lojas. Cada loja emprega $Y + 1$ funcionários, sendo Y uma v.a. Binomial com $n = 5$ e $p = 0.6$.

   a) Qual o número mínimo e qual o número máximo de funcionários por loja?

   b) Qual a probabilidade de uma loja escolhida ao acaso ter exatamente 4 funcionários?

   c) Determine a probabilidade aproximada de virem a trabalhar neste centro comercial mais do que 1000 funcionários.

5. O diâmetro médio de uma peça a fabricar em série, deve ser igual a 10 mm. Em condições ideais, o desvio padrão dos diâmetros é igual a 0.05 mm e a distribuição dos mesmos é aproximadamente Normal. Para monitorizar a média dos diâmetros fabricados, periodicamente, vai medir-se uma amostra aleatória de 5 peças e proceder-se a um teste de hipóteses com uma significância de 0.5%, com vista a detetar eventuais alterações em relação ao ideal (10 mm).

   a) As medições (em milímetros) de uma das amostras obtidas foram 9.973, 10.012, 10.030, 9.915 e 9.952. Podemos concluir que o diâmetro médio se desviou significativamente de 10 mm? Proceda ao teste de hipóteses nos termos enunciados.

   b) Na afirmação – *há evidência de que o diâmetro se desviou significativamente de 10 mm quando a média amostral* ($\bar{x}$) *for menor que* $\bar{x}_i$ *ou maior que* $\bar{x}_s$ – quais os valores de $\bar{x}_i$ e de $\bar{x}_s$ que tornam a afirmação verdadeira?

6. Pretende-se calcular um intervalo de confiança a 95% para a proporção de contribuintes que concorda com a nova legislação, a chamada *fiscalidade verde*.

   a) Numa sondagem de opinião envolvendo 400 contribuintes, 150 manifestaram concordância com a nova legislação. Nestas condições, determine o intervalo de confiança para a proporção de contribuintes que concorda com a nova legislação.

   b) Qual o tamanho mínimo da amostra de contribuintes a inquirir de modo a que o intervalo a calcular tenha uma amplitude inferior a 5 pontos percentuais.