Probabilidades e Estatística

Primeiro exame/segundo teste — 2º semestre – 2009/2010

Duração: 180/90 minutos — 17/06/2010 – 9:00 horas

Justifique convenientemente todas as respostas.

Grupo I — 5 valores

1. Com base no número de acidentes que cada condutor tem num ano, uma companhia de seguros classifica os condutores em três tipos: X, Y e Z. A probabilidade de ter pelo menos um acidente num ano é de 0.02, 0.04 e 0.01 para condutores dos tipos X, Y e Z, respectivamente. Nesta companhia as proporções de segurados de cada tipo são 30% (X), 60% (Y) e 10% (Z).

(a) Calcule a probabilidade de um segurado escolhido ao acaso não ter nenhum acidente num ano. (1.0)

Sejam $X(Y,Z)=$"um segurado é do tipo $X(Y,Z)$" e $A=$"um segurado tem pelo menos um acidente num ano". Tem-se $P(X)=0.3$, $P(Y)=0.6$, $P(Z)=0.1$, $P(A|X)=0.02$, $P(A|Y)=0.04$ e $P(A|Z)=0.01$.

$P(\bar{A})=1-P(A)=1-(P(A|X)P(X)+P(A|Y)P(Y)+P(A|Z)P(Z))=$ (teo. da probabilidade total)

$=1-(0.02\times 0.3+0.04\times 0.6+0.01\times 0.1)=1-0.031=0.969$.

(b) Se um novo segurado não tiver acidentes num ano, qual é a probabilidade de este ser um condutor do tipo X? (1.0)

$$P(X|\bar{A})=\frac{P(\bar{A}\cap X)}{P(\bar{A})}=\frac{P(\bar{A}|X)P(X)}{P(\bar{A})}=\frac{0.98\times 0.3}{0.969}\approx 0.303$$

2. O salário mensal (em euros) auferido por um funcionário de uma empresa privada segue uma distribuição Normal, com desvio padrão igual a 350 euros.

(a) Determine o salário médio, sabendo que 10% dos funcionários desta empresa ganham um salário mensal superior a 3500 euros. (1.5)

Seja $X=$"salário mensal auferido por um funcionário, em euros" com $X\sim N(\mu,350^2)$.

$P(X>3500)=0.1 \Leftrightarrow P(X\leq 3500)=0.9 \Leftrightarrow P\left(\frac{X-\mu}{350}\leq\frac{3500-\mu}{350}\right)=0.9 \Leftrightarrow \Phi\left(\frac{3500-\mu}{350}\right)=0.9 \Leftrightarrow \frac{3500-\mu}{350}=\Phi^{-1}(0.9)=1.2816 \Leftrightarrow \mu\approx 3051.5$ euros.

(b) Considere um grupo de 10 indivíduos escolhidos ao acaso de entre o grande número de funcionários desta empresa que ganham mais que o salário médio. Deduza o número mais provável de funcionários com um salário mensal superior a 3500 euros nesse grupo. (1.5)

Seja Y ="nº de funcionários com salário superior a 3500 num grupo de 10 funcionários que ganham mais do que o salário médio". Então, Y é o nº de sucessos em 10 indivíduos (ensaios de Bernoulli) em que 10 é muito menor do que o nº de funcionários que ganham mais do que o salário médio, pelo que $Y \sim Bin(10, p)$, aproximadamente, com $p = P(X > 3500|X > \mu) = \frac{P(X>3500 \wedge X>\mu)}{P(X>\mu)} = \frac{P(X>3500)}{P(X>\mu)} = \frac{0.1}{0.5} = 0.2$.
Seja $g(y) = \frac{f_Y(y+1)}{f_Y(y)} = \frac{10-y}{4(y+1)}$, para $y = 0, \ldots, 9$. Como $g(y) > 1$ para $y < 2$ e $g(y) < 1$ para $y \geq 2$ tem-se que a moda de Y é igual a 2.

Grupo II	5 valores

1. Seja X uma variável aleatória representando o desvio (relativamente a uma dada constante) do diâmetro de uma peça metálica, com a seguinte função de distribuição

$$F_X(x) = \begin{cases} 0.5e^{0.5x}, & x < 0 \\ 1 - 0.5e^{-0.5x}, & x \geq 0 \end{cases}$$

(a) Qual a probabilidade de ser necessário inspeccionar mais de 8 peças, seleccionadas de forma independente, até se encontrar uma peça com um desvio superior a 2? (1.0)

Seja Z ="nº de peças inspeccionadas até se encontrar uma com um desvio superior a 2". Assim, Z representa o nº de realizações independentes de uma prova de Bernoulli até ser observado o 1º sucesso, pelo que $Z \sim Geo(p)$ em que $p = P(X > 2) = 1 - F_X(2) = 0.5e^{-1} \approx 0.184$.
Então, $P(Z > 8) = 1 - F_Z(8) \approx 1 - 0.803 = 0.197$ (ou $= (1 - 0.184)^8$).

(b) Prove que a variável aleatória $Y = |X|$ tem distribuição exponencial. (1.5)

$\forall y > 0 : F_Y(y) = P(Y \leq y) = P(|X| \leq y) = P(-y \leq X \leq y) = F_X(y) - F_X(-y) = 1 - e^{-0.5y}$ e $F_Y(y) = 0, \forall y \leq 0$. Pela forma da função de distribuição conclui-se que $Y \sim Exp(0.5)$. Ou, como $f_Y(y) = \frac{dF_Y(y)}{dy}$ nos pontos onde a derivada existe, então $f_Y(y) = 0.5e^{-0.5y}$ para $y > 0$.

2. Num dado composto químico, X e Y são as proporções dos elementos químicos A e B, respectivamente. Admite-se que (X, Y) tem função de densidade de probabilidade conjunta

$$f_{X,Y}(x, y) = \begin{cases} 6\,x, & 0 < x < y < 1 \\ 0, & \text{caso contrário.} \end{cases}$$

(a) Mostre que as proporções dos elementos químicos A e B são variáveis dependentes. (1.0)

$f_X(x) = \int_{I\!R} f_{X,Y}(x, y)\, dy = \int_x^1 6x\, dy = 6x(1 - x)$, para $0 < x < 1$
$f_Y(y) = \int_{I\!R} f_{X,Y}(x, y)\, dx = \int_0^y 6x\, dx = 3y^2$, para $0 < y < 1$
Ambas as funções são nulas fora dos intervalos indicados.
Como, por exemplo, $0 = f_{X,Y}(1/2, 1/2) \neq f_X(1/2) \times f_Y(1/2) = 9/8$, as variáveis são dependentes.

(b) Qual o valor esperado condicional da proporção do elemento químico A quando a proporção do elemento químico B é y, para qualquer $y \in]0, 1[$? (1.5)

$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{2x}{y^2}, & 0 < x < y \\ 0, & \text{caso contrário} \end{cases}$, para $0 < y < 1$.
$E[X|Y = y] = \int_{I\!R} x f_{X|Y=y}(x)\, dx = \int_0^y \frac{2x^2}{y^2}\, dx = \frac{2}{y^2} \left[\frac{x^3}{3}\right]_0^y = \frac{2y}{3}$.

1. O *stress* vibratório (em N/mm^2) de uma lâmina de uma turbina de motor (X) tem, sob certas condições, a seguinte função densidade de probabilidade

$$f_X(x) = \frac{x}{\theta}\, e^{-\frac{1}{2\theta}x^2},\ x \geq 0,\ \text{com } \theta > 0.$$

(a) Mostre que $T = \frac{1}{2n}\sum_{i=1}^{n} X_i^2$ é o estimador de máxima verosimilhança de θ para uma amostra aleatória de tamanho n. (1.5)

$\mathcal{L}(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{x_i}{\theta}\, e^{-\frac{1}{2\theta}x_i^2} = \theta^{-n} e^{-\frac{\sum_{i=1}^{n} x_i^2}{2\theta}} \prod_{i=1}^{n} x_i$. Como esta função de θ está definida num aberto de $I\!R$ e é diferenciável, o seu máximo pode ser determinado por derivação.

$\log \mathcal{L} = -n \log \theta - \frac{\sum_{i=1}^{n} x_i^2}{2\theta} + \sum_{i=1}^{n} \log x_i$

$\frac{d \log \mathcal{L}}{d\theta} = 0 \Leftrightarrow -\frac{n}{\theta} + \frac{\sum_{i=1}^{n} x_i^2}{2\theta^2} = 0 \Leftrightarrow -2n\theta + \sum_{i=1}^{n} x_i^2 = 0 \Leftrightarrow \theta = \frac{\sum_{i=1}^{n} x_i^2}{2n}$

$\frac{d^2 \log \mathcal{L}}{d\theta^2} = \frac{n}{\theta^2} - \frac{\sum_{i=1}^{n} x_i^2}{\theta^3} = \frac{n\theta - \sum_{i=1}^{n} x_i^2}{\theta^3}$

$\left.\frac{d^2 \log \mathcal{L}}{d\theta^2}\right|_{\theta = \frac{\sum_{i=1}^{n} x_i^2}{2n}} = -\frac{4n^3}{\left(\sum_{i=1}^{n} x_i^2\right)^2} < 0,\ \forall x_1, \ldots, x_n,\ \forall n \in I\!N.$

$\therefore\ \hat{\theta}_{MV} = \frac{\sum_{i=1}^{n} X_i^2}{2n}$

(b) Determine a estimativa de máxima verosimilhança da probabilidade de o *stress* vibratório exceder 15, com base numa amostra observada em que $\sum_{i=1}^{10} x_i^2 = 1490.11$. (1.0)

$P(X > 15) = 1 - F_X(15) = 1 - \int_0^{15} \frac{x}{\theta}\, e^{-\frac{1}{2\theta}x^2}\, dx = 1 - \left[-e^{-\frac{x^2}{2\theta}}\right]_0^{15} = e^{-\frac{15^2}{2\theta}} = g(\theta).$

Pela invariância dos estimadores de máxima verosimilhança tem-se que

$$\hat{g}_{MV}(\theta) = g(\hat{\theta}_{MV}) = e^{-\frac{225}{2\hat{\theta}_{MV}}} = e^{-\frac{225n}{\sum_{i=1}^{n} X_i^2}}.$$

A estimativa pedida é então $e^{-\frac{2250}{1490.11}} \approx 0.221$.

2. Numa experiência com alunos de uma escola, foram medidas as temperaturas de 6 alunos do turno da manhã (X_1) e de 5 alunos do turno da tarde (X_2), obtendo-se os valores $\bar{x}_1 = 36.25$, $\bar{x}_2 = 36.98$, $\sum_{j=1}^{6} x_{1j}^2 = 7909.20$ e $\sum_{j=1}^{5} x_{2j}^2 = 6854.26$. Considerando que as temperaturas de os alunos de ambos os turnos são variáveis aleatórias independentes com distribuições Normais e variância comum, construa um intervalo de confiança a 99% para a diferença entre as temperaturas médias dos alunos dos dois turnos. Há alguma evidência de que os valores esperados da temperatura dos alunos são distintos nos dois turnos? (2.5)

Variável fulcral: $Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{Sp} \sim t_{(9)}$ com $S_p = \sqrt{\frac{5S_1^2 + 4S_2^2}{9}\left(\frac{1}{6} + \frac{1}{5}\right)}$.

$P(-a < Z < a) = 0.99 \Leftrightarrow a = F^{-1}_{t_{(9)}}(0.995) = 3.25$

$P(-3.25 < Z < 3.25) = 0.99 \Leftrightarrow P(\bar{X}_1 - \bar{X}_2 - 3.25S_p < \mu_1 - \mu_2 < \bar{X}_1 - \bar{X}_2 + 3.25S_p) = 0.99$

$IAC_{0.99}(\mu_1 - \mu_2) =]\bar{X}_1 - \bar{X}_2 - 3.25S_p,\ \bar{X}_1 - \bar{X}_2 + 3.25S_p[$

$\bar{x}_1 - \bar{x}_2 = 36.25 - 36.98 = -0.73$

$s_1^2 = \frac{\sum_{j=1}^{6} x_{1j}^2 - 6\bar{x}_1^2}{5} = 4.965,\ s_2^2 = \frac{\sum_{j=1}^{5} x_{2j}^2 - 5\bar{x}_2^2}{4} = 4.1645$ e $s_p = 1.3$

$IC_{0.99}(\mu_1 - \mu_2) =]-4.955,\ 3.495[$

Uma vez que $0 \in IC_{0.99}(\mu_1 - \mu_2)$, não há evidência suficiente para afirmar que os valores esperados das temperaturas dos alunos são diferentes nos dois turnos para pelo menos níveis de significância ≤ 0.01.

Grupo IV	5 valores

1. A tabela abaixo indica os tempos de montagem, em minutos, de 600 aparelhos de um certo tipo:

Tempo	$[0, 10]$	]10,15]	$]15, 20]$	$]20, 30]$
Nº de aparelhos	190	110	90	210

Para que níveis de significância são estes resultados compatíveis com a hipótese de o tempo de montagem desse tipo de aparelho ter distribuição uniforme no intervalo [0,30]? (2.5)

Seja X =“tempo de montagem de um certo tipo de aparelho, em minutos”.
Pretende-se testar $H_0 : X \sim U[0, 30]$ contra $H_1 : X$ segue qualquer outra distribuição.
Sob H_0, $f_X(x) = f_X^0(x) = 1/30$ para $0 \leq x \leq 30$.

C_i	o_i	p_i^0	$e_i = np_i^0$
$[0, 10]$	190	1/3	200
$]10, 15]$	110	1/6	100
$]15, 20]$	90	1/6	100
$]20, 30]$	210	1/3	200
	$n = 600$	1	600

$p_i^0 = P(X \in C_i|H_0) = \int_{C_i} f_X^0(x)\, dx,\ i = 1, \ldots, 4.$
Estatística de teste: $Q_0 = \sum_{i=1}^4 \frac{(O_i - E_i)^2}{E_i} \overset{a}{\sim} \chi^2_{(3)}$, sob H_0.
$q_0 = 3$ e valor-p=$P(Q_0 \geq 3|H_0) = 1 - F_{\chi^2_{(3)}}(3) = 1 - 0.608 = 0.392$.
Não se rejeita H_0 para n.s. inferiores a 0.392.

2. Uma empresa regista as despesas miúdas em dezenas de euros (que não incluem alojamento e refeições) que os vendedores dos seus produtos fazem durante o tempo (em dias) das viagens de visita aos clientes. Desse conjunto de dados escolheu-se casualmente os valores das variáveis indicadas relativos a 8 viagens que se apresentam resumidamente em seguida:

$$\sum_{i=1}^{8} y_i = 139; \quad \sum_{i=1}^{8} x_i = 37; \quad \sum_{i=1}^{8} y_i^2 = 2669; \quad \sum_{i=1}^{8} x_i^2 = 205.5; \quad \sum_{i=1}^{8} x_i y_i = 734.$$

O núcleo de Estatística da empresa considerou para análise dos dados o seguinte modelo de regressão: Y_i, $i = 1, \ldots, 8$, independentes e identicamente distribuídas com distribuição $N(\beta_0 + \beta_1 x_i, \sigma^2)$.

(a) Indique um estimador de $\eta_i = \beta_0 + \beta_1 x_i$ e averigue se esse estimador é centrado. (1.0)

Recorrendo aos estimadores dos MQ (ou MV) de β_0 e β_1 tem-se $\hat{\eta}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i = \bar{Y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i = \bar{Y} + \hat{\beta}_1 (x_i - \bar{x}) = \bar{Y} + \frac{\sum_j x_j Y_j - n\bar{x}\bar{Y}}{\sum_k x_k^2 - n\bar{x}^2}(x_i - \bar{x}) =$
$= \frac{\sum_j Y_j}{n} + \frac{\sum_j (x_j - \bar{x})(x_i - \bar{x}) Y_j}{\sum_k x_k^2 - n\bar{x}^2} = \sum_j \left(\frac{1}{n} + \frac{(x_j - \bar{x})(x_i - \bar{x})}{\sum_k x_k^2 - n\bar{x}^2}\right) Y_j = \sum_j w_j Y_j.$
$E[\hat{\eta}_i] = \sum_j w_j E[Y_j] = \sum_j w_j (\beta_0 + \beta_1 x_j) = \beta_0 \sum_j w_j + \beta_1 \sum_j w_j x_j$
$\sum_j w_j = 1 + \frac{(x_i - \bar{x})}{\sum_k x_k^2 - n\bar{x}^2} \sum_j (x_j - \bar{x}) = 1$
$\sum_j w_j x_j = \sum_j \frac{x_j}{n} + \frac{(x_i - \bar{x})}{\sum_k x_k^2 - n\bar{x}^2} \sum_j x_j (x_j - \bar{x}) = \bar{x} + \frac{(x_i - \bar{x})}{\sum_k x_k^2 - n\bar{x}^2} (\sum_j x_j^2 - n\bar{x}^2) = x_i$
$\therefore E[\hat{\eta}_i] = \beta_0 + \beta_1 x_i = \eta_i.$

(b) Qual a estimativa do acréscimo da despesa (Y) média correspondente a mais um dia no tempo da visita de um vendedor? (0.5)

Quer-se estimar $E[Y|x + 1] - E[Y|x] = \beta_0 + \beta_1 (x + 1) - (\beta_0 + \beta_1 x) = \beta_1$
$\hat{\beta}_1 = \frac{\sum_j x_j y_j - n\bar{x}\bar{y}}{\sum_k x_k^2 - n\bar{x}^2} \approx 2.651 \times 10$ euros.

(c) Teste a hipótese de a despesa média correspondente a uma viagem de 5 dias não ultrapassar o montante de 180 euros (use como estimativa centrada de σ^2 o valor 2.05). (1.0)

Seja $\eta = E[Y|x=5] = \beta_0 + 5\beta_1$. Pretende-se testar $H_0 : \eta \leq 18$ contra $H_1 : \eta > 18$.
Estatística de teste: $T_0 = \frac{\hat{\eta}-18}{\sqrt{\left(\frac{1}{8}+\frac{(\bar{x}-5)^2}{\sum x_i^2 - 8\bar{x}^2}\right)\hat{\sigma}^2}} \overset{\eta=18}{\sim} t_{(6)}$.
$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 5.115 \Rightarrow \hat{\eta} = 18.369$.
$t_0 = \frac{0.369}{0.514} = 0.717$, valor-p$= P(T_0 \geq 0.717 | \eta = 18) = 1 - F_{t_{(6)}}(0.717) \approx 1 - 0.75 = 0.25$.
$\therefore$ Não se rejeita H_0 aos níveis de significância usuais (≤ 0.1).