First exam October 23, 2006
Statistics II

1. (7 points) Numa determinada empresa de recursos humanos um teste de aptidão é realizado por um elevado número de candidatos a um emprego. A pontuação obtida nesse teste (a qual varia entre 0 e 100) segue uma distribução normal, com mediana igual a 65. Sabe-se ainda que 10% dos candidatos têm uma nota superior a 75.

   (a) Calcule a probabilidade da pontuação obtida no teste estar entre 60 e 65.

   Solution: Seja X uma variável aleatória que representa a pontuação obtida no teste de aptidão. Então, X segue uma distribuição normal com média μ e variância σ^2. Como a distribuição normal é simétrica, a média é igual à mediana, logo $\mu = 65$. Sabe-se ainda que

   $$\mathbb{P}(X > 75) = 0.1,$$

   ou

   $$\Phi\left(\frac{75-65}{\sigma}\right) = 0.9.$$

   Então, usando as tabelas da distribuição normal, temos que $10/\sigma = 1.282$, logo $\sigma = 7.8$. Então, a probabilidade é

   $$\begin{aligned}\mathbb{P}(60 < X < 65) &= \Phi\left(\frac{65-65}{7.8}\right) - \Phi\left(\frac{60-65}{7.8}\right) = \Phi(0) - \Phi(-0.64) = \\ &= 0.5 - [1 - \Phi(0.64)] = 0.2389.\end{aligned}$$

   (b) Indique qual a pontuação correspondente ao 3° quartil. Interprete o resultado obtido.

   Solution: Represente-se o 3° quartil da variável por Q_{III}. Então, temos de ter $\mathbb{P}(X < Q_{III}) = 0.75$. Assim,

   $$\Phi\left(\frac{Q_{III}-65}{7.8}\right) = 0.75,$$

   pelo que temos que $Q_{III} = 70.2611$. Cerca de 75% dos candidatos a um emprego obtêm uma pontuação no teste igual ou inferior a 70.26.

   (c) Num total de 18 testes seleccionados aleatoriamente, qual a probabilidade de mais de metade terem uma pontuação inferior a 77.9 valores?

   Solution: Seja Y = número de testes com pontuação inferior a 77.9 num total de 18 testes.
   Então, Y segue uma distribuição binomial com $N = 18$ e $p = \mathbb{P}(X < 77.9)$ que

é

$$p = \mathbb{P}(X < 77.9) = \Phi\left(\frac{77.9 - 65}{7.8}\right) \approx 0.95.$$

Portanto,

$$\mathbb{P}(Y > 9) = \mathbb{P}(W < 9) = \mathbb{P}(W \leq 8) \approx 1,$$

onde $W \sim \mathrm{b}(N = 18, p = 1 - 0.95 = 0.05)$.

(d) Qual a probabilidade de ser necessário seleccionar menos de 4 testes até encontrar o 2° com pontuação inferior à referida na alínea anterior?

Solution: Seja L = número de testes até encontrar o 2° com nota inferior a 77.9.
Então, L segue uma distribuição binomial negativa com $K = 2$ e $p = 0.95$.

$$\begin{aligned}
\mathbb{P}(L < 4) &= \mathbb{P}(L = 2) + \mathbb{P}(L = 3) = \mathrm{bn}(2; 2, 0.95) + \mathrm{bn}(3; 2, 0.95) = \\
&= \mathrm{b}(2; 2, 0.95) + \frac{2}{3}\,\mathrm{b}(2; 3, 0.95) = \mathrm{b}(0; 2, 0.05) + \frac{2}{3}\,\mathrm{b}(1; 3, 0.05) = \\
&= 0.9025 + \frac{2}{3}\,0.1354 = 0.9928.
\end{aligned}$$

Sabe-se que em 55% dos casos o tempo de correcção de um teste de aptidão é inferior a 10 minutos.

(e) Qual a probabilidade de serem corrigidos no máximo 3 testes em 45 minutos?

Solution: Seja J = número de testes corrigidos em 45 minutos.
Então, J segue uma distribuição de Poisson com λ_J desconhecido. Para calcular λ, seja T = tempo entre a correcção de dois testes em minutos.
Então, $T \sim \exp(\lambda_T)$ com λ_T desconhecido. Sabe-se que $\mathbb{P}(T < 10) = 0.55$, portanto,

$$\mathbb{P}(T < 10) = 1 - e^{-10\lambda_T} = 0.55,$$

donde temos que $\lambda_T = 0.07985$.
Assim, $\lambda_J = 45 \cdot 0.07985 \approx 3.6$ pelo que $J \sim \mathcal{P}(3.6)$. Então, $\mathbb{P}(J \leq 3) = 0.5152$.

(f) Sabe-se que a pessoa responsável pela correcção dos testes de aptidão passa cerca de 6 horas por dia nesta tarefa. Qual a probabilidade de em 5 dias ele conseguir corrigir 150 testes de aptidão? Indique todas as hipóteses que julgue necessárias. A probabilidade pedida deverá ser calculada com base nas tabelas estatísticas.

Solution: Seja M = número de testes corrigidos por hora.

Então, M segue uma distribuição de Poisson com λ desconhecido. Na alínea anterior vimos que $\lambda = 3.6$ para 45 minutos, assim para uma hora $\lambda = 4.8$.

Seja S = número de testes corrigidos em 30 horas (6 horas $\times 5$ dias).

S também segue uma distribuição de Poisson, $S \sim \mathcal{P}(\lambda = 4.8 \cdot 30 = 144)$. Pelo teorema de límite central, $S \overset{a}{\sim} \mathcal{N}(\mu = 144, \sigma^2 = 144)$. Assim,

$$\begin{aligned}\mathbb{P}(S = 150) &= \mathbb{P}(149.5 < S < 150.5) = \\ &= \mathbb{P}\left(\frac{149.5 - 144}{\sqrt{144}} < Z < \frac{150.5 - 144}{\sqrt{144}}\right) = \\ &= \Phi(0.54) - \Phi(-0.54) = 2\Phi(0.54) - 1 = 0.4108.\end{aligned}$$

2. (6.5 points) O peso das abóboras cultivadas pelo Sr. Isaías no seu terreno de Loures segue uma distribuição normal com média igual a 4 Kg e variância igual a 10 Kg^2. Sabendo que o Sr. Isaías vende as suas abóboras em lotes de 10 unidades,

   (a) Qual a probabilidade de que o peso médio das abóboras de um lote se situe entre os 3 e os 5 Kg?

   (b) E quantas abóboras deverá ter um lote (qual a dimensão do lote), por forma a assegurar que o peso médio das abóboras nesse lote seja superior a 3.5 Kg com 95% de probabilidade?

   O Sr. Isaías está a pensar comprar um outro terreno em Bucelas, também para o cultivo de abóboras, tendo-lhe sido garantido que o peso destas também se distribui normalmente, mas agora com uma média de 5 Kg e um desvio padrão de 3 Kg.

   (c) Qual a probabilidade de que o peso médio das abóboras de um lote produzido em Bucelas seja 1.5 Kg superior ao peso médio das abóboras de um lote produzido em Loures? (Admita que ambos os lotes têm 10 abóboras)

   Mais tarde o Sr. Isaías recebe uma proposta para a compra de um outro terreno perto de Vila Franca. É-lhe assegurado que, num lote de 10, a variância corrigida, do peso das abóboras desse lote é inferior a 6 Kg^2 com uma probabilidade de 90%. Admita que também em Vila Franca o cultivo das abóboras segue uma distribuição normal.

   (d) Qual a variância do peso das abóboras cultivadas neste terreno de Vila Franca?

3. (6.5 points) (a) Explique, **por palavras**, porque a distribuição exponencial não tem memória.

   Diga se as seguintes afirmações são verdadeiras ou falsas. Justifique as suas respostas.

   (b) Se X é uma variável aleatória que segue uma distribuição Poisson com média λ, onde $0 < \lambda < 1$, então $\mathbb{E}(X!) = e^{-\lambda}/(1 - \lambda)$.

Solution:

$$\begin{aligned}\mathbb{E}(X!) &= \sum_{x=0}^{\infty} x!\frac{\lambda^x}{x!}e^{-\lambda} = \sum_{x=0}^{\infty} \lambda^x e^{-\lambda} = \\ &= e^{-\lambda}\sum_{x=0}^{\infty}\lambda^x.\end{aligned}$$

Como $0 < \lambda < 1$, a soma converge e é igual a $1/(1-\lambda)$. Assim, $\mathbb{E}(X!) = e^{-\lambda}/(1-\lambda)$, portanto a afirmação é **VERDADEIRA**.

(c) Se a proporção de filhos varões numa população é 0.5 e são seleccionadas 4 famílias com 3 filhos, então a probabilidade de pelo menos uma destas famílias ter todos os filhos varões é 0.5862.

Solution: Seja X = número de famílias entre as 4 seleccionadas com 3 filhos varões.
Então, $X \sim \mathrm{b}(N = 4, p = ?)$. A probabilidade pode ser calculada como $p = 0.5 \cdot 0.5 \cdot 0.5 = 0.125$. Assim,

$$\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - \binom{4}{0} 0.125^0 0.875^4 = 0.4138 \neq 0.5862.$$

Portanto, a afirmação é **FALSA**.

(d) Se X_1 e X_2 são variáveis aleatórias independentes com $X_1 \sim \mathcal{N}(1,2)$ e $X_2 \sim \mathcal{N}(2,2)$, então $W = (X_1-1)^2/2 + (X_2-2)^2/2$ segue uma distribuição χ^2 com 2 graus de liberdade, e $U = \frac{(X_1-1)/\sqrt{2}}{\sqrt{W/2}}$ segue uma distribuição Student-t com 2 graus de liberdade.

Solution:

$$\begin{array}{c} X_1 \sim \mathcal{N}(1,2) \\ X_2 \sim \mathcal{N}(2,2) \end{array} \iff \begin{array}{c} \frac{X_1-1}{\sqrt{2}} \sim \mathcal{N}(0,1) \\ \frac{X_2-2}{\sqrt{2}} \sim \mathcal{N}(0,1) \end{array}$$

Além disso,

$$\frac{(X_1-1)^2}{2} \sim \chi_1^2 \quad \text{e} \quad \frac{(X_2-2)^2}{2} \sim \chi_1^2$$

Como X_1 e X_2 são independentes, a soma $(X_1-1)^2/2 + (X_2-2)^2/2$ também segue uma distribuição χ^2 com 2 graus de liberdade. Assim $W \sim \chi_2^2$.
U é o rácio entre uma variável aleatória normal $\mathcal{N}(0,1)$ e a raíz quadrada de uma variável aleatória χ^2 dividida pelos graus de liberdade. Se o numerador

e o denominador fossem independentes, U seria uma variável aleatória com distribuição t-Student com 2 graus de liberdade. No entanto, X_1 e W são claramente dependentes porque W é função de X_1. Assim, a afirmação é **FALSA**.

(e) Se X segue uma distribuição uniforme entre 0 and 1, então $Y = -\ln X/a$ tem uma distribuição exponencial com média $1/a$.

Solution: Como $X \sim U(0,1)$, $F_X(x) = x$, $0 < x < 1$. Assim, $-\infty < \ln X < 0 \iff 0 < -\ln X < \infty \iff 0 < -\frac{\ln X}{a} < \infty$, portanto o domínio de Y é $(0, \infty)$ que é o domínio de uma variável aleatória exponencial. Ao representar a função de distribuição de Y por F_Y, então para $x \in (0, \infty)$,

$$\begin{aligned} F_Y(x) &= \mathbb{P}(Y \leq x) = \mathbb{P}\left(-\frac{\ln X}{a} \leq x\right) = \mathbb{P}\left(X \geq e^{-ax}\right) = \\ &= 1 - \mathbb{P}\left(X \leq e^{-ax}\right) = 1 - F_X\left(e^{-ax}\right) = 1 - e^{-ax}. \end{aligned}$$

Ao derivar, a função densidade de probabilidade de Y é $f_Y(x) = ae^{-ax}$ que é a função densidade da distribuição exponencial com parametro a. Assim a média é dada por $1/a$. Então, a afirmação é **VERDADEIRA**.

(f) Se X é uma variável aleatória normal, $\mathbb{P}(X \leq 16) = 0.8$ e $\mathbb{P}(X \leq 3) = 0.15$, então $\mathbb{P}(2.5 < X < 4) = 0.1$.

Solution:

$$\begin{array}{l} \mathbb{P}(X \leq 16) = 0.8 \\ \mathbb{P}(2.5 < X < 4) = 0.1 \end{array} \iff \begin{array}{l} \mathbb{P}\left(Z \leq \frac{16-\mu}{\sigma}\right) = 0.8 \\ \mathbb{P}\left(Z \leq \frac{3-\mu}{\sigma}\right) = 0.15 \end{array} \iff \begin{array}{l} \frac{16-\mu}{\sigma} = 0.8416 \\ \frac{3-\mu}{\sigma} = -1.036 \end{array}$$

Ao resolver estas equacões, obtém-se $\mu = 10.1729$ e $\sigma = 6.9237$. Assim,

$$\begin{aligned} \mathbb{P}(2.5 < X < 4) &= \mathbb{P}\left(\frac{2.5 - 10.1729}{6.9237} < Z < \frac{4 - 10.1729}{6.9237}\right) = \\ &= \mathbb{P}(-1.1082 < Z < -0.8916) = \\ &= \Phi(-0.8916) - \Phi(-1.1082) = 0.051 \neq 0.1, \end{aligned}$$

portanto, a afirmação é **FALSA**.