INSTITUTO SUPERIOR DE AGRONOMIA

Exame de Matemática II (2^a. Chamada)

Uma possível resolução

Duração: 2 horas 30 de Junho de 2015

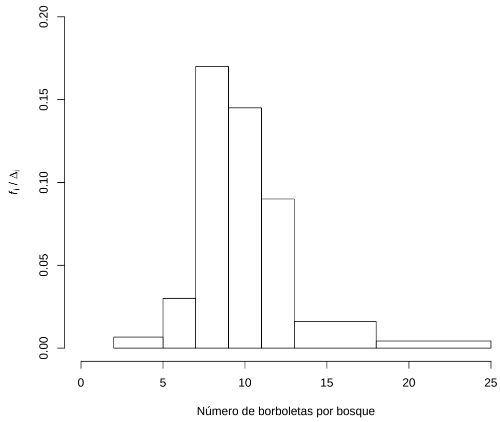
1. a) Número de espécies de borboletas por bosque - variável quantitativa discreta pois é uma variável numérica que só pode tomar valores em $\mathbf{N}_0$

b)

Classe i	Freq. absoluta n_i	Freq. relativa f_i	Freq. rel. acumulada, F_i	Amplitude Δ_i	Altura rectângulo f_i/Δ_i
$]2,5]$	2	0.02	0.02	3	0.007
$]5,7]$	6	0.06	0.08	2	0.030
$]7,9]$	34	0.34	0.42	2	0.170
$]9,11]$	29	0.29	0.71	2	0.145
$]11,13]$	18	0.18	0.89	2	0.090
$]13,18]$	8	0.08	0.97	5	0.016
$]18,25]$	3	0.03	1.00	7	0.004

Nota: as duas últimas colunas não fazem parte da resposta à alínea b) mas são necessárias para responder à questão seguinte.

c)



d) A grande diferença entre os dois diagramas é a existência de observações atípicas (*outliers*). Vamos calcular valores aproximados do primeiro e do terceiro quartis ($Q'_{0.25}$ e $Q'_{0.75}$) e das barreiras inferior e superior (BI e BS) para analisar a possível existência deste tipo de observações. Utilizando as fórmulas disponíveis no formulário e como a primeira classe com frequência relativa acumulada superior a 0.25 é a terceira ($F_3 = 0.42 \geq 0.25$) e superior a 0.75 é a quinta ($F_5 = 0.89 \geq 0.75$),

$Q'_{0.25} = 7 + 2 \times \frac{0.25-0.08}{0.34} = 8.0$

$Q'_{0.75} = 11 + 2 \times \frac{0.75-0.71}{0.18} = 11.44$

$BI \simeq 8.0 - 1.5 \times (11.44 - 8.0) = 2.84$

$BS \simeq 11.44 + 1.5 \times (11.44 - 8.0) = 16.6$

Como o valor aproximado de BI é superior a 2, extremo inferior da primeira classe, e o de BS é inferior a 25, extremo superior da última classe, a existência de observações atípicas neste conjunto de dados é admissível. Deste modo, o **diagrama A** deverá ser o correspondente a este conjunto de dados.

2. a) Vamos calcular o coeficiente de correlação entre as duas variáveis (r_{xy}) utilizando as fórmulas disponíveis no formulário

$$r_{xy} = \frac{cov(x,y)}{s_x\, s_y} = \frac{\frac{n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n(n-1)}}{\sqrt{\frac{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}}\sqrt{\frac{n\sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}{n(n-1)}}}$$

$$\Leftrightarrow \quad r_{xy} = \frac{16 \times 29641 - 592 \times 691}{\sqrt{(16 \times 25968 - 592^2)(16 \times 34719 - 691^2)}} = 0.915.$$

Como o valor de r_{xy} está perto de um dos extremos do intervalo de variação deste parâmetro, $[-1, 1]$, podemos dizer que a intensidade de relacionamento linear entre x e y é forte. Esta análise quantitativa devia ser, no entanto, complementada pela análise de um diagrama de dispersão entre x e y.

b) A equação da recta de regressão linear ($y = b_0 + b_1 x$) é, neste caso, $\boxed{y = 6.1135 + 1.002\,x}$ pois,

$$b_1 = \frac{cov(x,y)}{s_x^2} = \frac{\frac{n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n(n-1)}}{\frac{n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n-1)}}$$

$$\Leftrightarrow \quad b_1 = \frac{16 \times 29641 - 592 \times 691}{16 \times 25968 - 592^2} = 1.002$$

e

$$b_0 = \bar{y} - b_1 \bar{x} \quad \Leftrightarrow \quad b_0 = \frac{691}{16} - 1.002 \times \frac{592}{16} = 6.1135$$

c) A precisão da recta de regressão é dada pelo coeficiente de determinação $R^2 = (r_{xy})^2 = 0.915^2 = 0.837$, o que significa que 83.7% da variabilidade de y é explicada pela regressão linear de y sobre x.

d) Valor ajustado: $\hat{y}_1 = 6.1135 + 1.002\,x_1 = 6.1135 + 1.002 \times 12 = 18.1375$
Resíduo: $e_1 = y_1 - \hat{y}_1 = 16 - 18.1375 = -2.1375$

3. X - v.a. que representa a procura semanal (em t) de terra vegetal num viveiro de plantas.

a) Seja $f(x)$ a função densidade de probabilidade de X. Sabemos que $\int_{-\infty}^{+\infty} f(x)\,dx = 1$ o que corresponde, neste caso, à área de um triângulo de base 2 e altura a. Logo,

$$\frac{2 \times \mathrm{a}}{2} = 1 \Leftrightarrow \mathrm{a} = 1.$$

b) De acordo com a figura do enunciado, o gráfico de $f(x)$ é:
- o segmento de recta contido na bissectriz dos quadrantes ímpares, quando $x \in [0, 1]$,
- o segmento de recta que une os pontos $(1, 1)$ e $(2, 0)$, quando $x \in [1, 2]$,
- o eixo dos xx, para os restantes valores de x.
Assim,

$$f(x) = \begin{cases} x & \text{se } 0 \le x \le 1 \\ -x + 2 & \text{se } 1 < x \le 2 \\ 0 & \text{restantes valores de } x \end{cases}$$

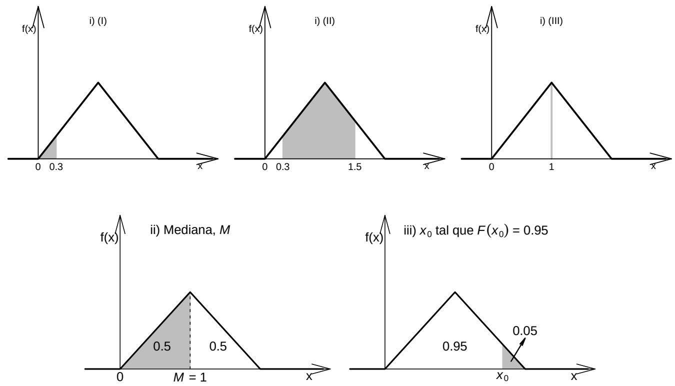
Seja $F(x)$ a função distribuição cumulativa de X.
Por definição, $F(x) = P[X \le x] = \int_{-\infty}^{x} f(t)\,dt, \quad x \in \mathbf{R}$. Assim,

$$\begin{array}{ll} x \le 0, & \int_{-\infty}^{x} 0\,dt = 0 \\ 0 < x \le 1, & \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{0} 0\,dt + \int_0^x t\,dt = \left[\frac{t^2}{2}\right]_0^x = \frac{x^2}{2} \\ 1 < x \le 2, & \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{0} 0\,dt + \int_0^1 t\,dt + \int_1^x (-t+2)\,dt = \frac{1}{2} + \left[-\frac{t^2}{2} + 2t\right]_1^x = -\frac{x^2}{2} + 2x - 1 \\ x > 2, & \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{0} 0\,dt + \int_0^1 t\,dt + \int_1^2 (-t+2)\,dt + \int_2^x 0\,dt = 1 \end{array}$$

Logo,

$$F(x) = \begin{cases} 0 & \text{se } x \le 0 \\ \frac{x^2}{2} & \text{se } 0 < x \le 1 \\ -\frac{x^2}{2} + 2x - 1 & \text{se } 1 < x \le 2 \\ 1 & \text{se } x > 2 \end{cases}$$

c) Interpretação geométrica (a sombreado apresentam-se as áreas que representam as probabilidades referentes a cada uma das alíneas)



Os cálculos pedidos podem ser efectuados com base na interpretação geométrica ou, analiticamente, utilizando a função distribuição cumulativa de X:

i) I) $P[X < 0.3] = F(0.3) = 0.09/2 = 0.045$;
II) $P[0.3 < X < 1.5] = F(1.5) - F(0.3) = -1.5^2/2 + 2 \times 1.5 - 1 - 0.045 = 0.83$;
III) $P[X = 1] = 0$;

ii) a mediana de X é o valor M tal que $F(M) = 0.5$. Facilmente se vê pelo gráfico da função f (e se confirma analiticamente) que $M = 1$;

iii) a quantidade mínima de terra vegetal x_0 tal que $P[X > x_0] = 0.05 \Leftrightarrow P[X \leq x_0] = 0.95 \Leftrightarrow F(x_0) = 0.95$.
Pela análise do gráfico de f, $x_0 \in]1, 2[$ pelo que $F(x_0) = 0.95 \Leftrightarrow -x_0^2/2 + 2x_0 - 1 = 0.95 \Leftrightarrow x_0 = 2 - \sqrt{0.1} \simeq 1.684 \;\vee\; x_0 = 2 + \sqrt{0.1}$ (impossível). Logo $x_0 = 1.684$ t.

4. X - v.a. que representa o nº. de bolbos de túlipa que não vigam num saco com 5 unidades;
Y - v.a. que representa o nº. de bolbos de amarílis que não vigam num saco com 4 unidades.

Como temos uma sucessão de provas de Bernoulli (bolbo não viga/viga), que se podem admitir independentes, com probabilidade de sucesso (bolbo não viga) constante, $X \frown B(5, 0.1)$ e $Y \frown B(4, 0.05)$.

a) i) $P[Y \geq 2] = 1 - P[Y < 2] = 1 - P[Y \leq 1] = 1 - 0.9860 = 0.014$ (pelas tabelas).

ii) $X + Y$ - v.a. que representam o número total de bolbos que não vigam.
Como X e Y são variáveis aleatórias independentes,
$P[X = x_i, Y = y_j] = P[X = x_i] \times P[Y = y_j], \quad \forall i, j.$

$$\begin{aligned} P[X+Y=1] &= P[X=0, Y=1] + P[X=1, Y=0] \\ &= P[X=0] \times P[Y=1] + P[X=1] \times P[Y=0] \\ &= \binom{5}{0} 0.1^0 0.9^5 \times \binom{4}{1} 0.05^1 0.95^3 + \binom{5}{1} 0.1^1 0.9^4 \times \binom{4}{0} 0.05^0 0.95^4 = 0.3684, \end{aligned}$$

b) T - v.a. que representam o número total de bolbos que vigam, $T = 9 - (X + Y)$.

$E[T] = E[9 - (X+Y)] \underset{(1)}{=} 9 - E[X] - E[Y] \underset{(2)}{=} 9 - 5 \times 0.1 - 4 \times 0.05 = 8.3$

$Var[T] = Var[9 - (X+Y)] \underset{(3)}{=} (-1)^2 Var[X+Y] \underset{(4)}{=} Var[X] + Var[Y] \underset{(2)}{=} 5 \times 0.1 \times 0.9 - 4 \times 0.05 \times 0.95 = 0.64$

(1) Propriedades do valor esperado,
(2) Porque X e Y são v.a. com distribuição binomial,
(3) Propriedades da variância,
(4) Porque X e Y são v.a. independentes.

5. X - v.a. que representa a quantidade (em ppm) de um poluente no solo numa dada região; $X \frown N(24.5, 1.4)$.

a) i) $P[24 < X < 25] = P[X < 25] - P[X \leq 24] = P\left[\frac{X-\mu}{\sigma} < \frac{25-24.5}{1.4}\right] - P\left[\frac{X-\mu}{\sigma} \leq \frac{24-24.5}{1.4}\right] =$
$= \Phi(0.36) - \Phi(-0.36) = \Phi(0.36) - 1 + \Phi(0.36) = 2 \times 0.64058 - 1 = 0.28116$ (pelas tabelas).

ii) $P[X > 27.9] = 1 - P[X \leq 27.9] = 1 - P\left[\frac{X-\mu}{\sigma} \leq \frac{27.9-24.5}{1.4}\right] = 1 - \Phi(2.43) = 1 - 0.99245 = 0.00755$.

b) Seja $(X_1, X_2, \ldots, X_{10})$ a amostra aleatória em questão: $X_1, X_2, \ldots, X_{10}$ são v.a. independentes e $X_i \frown N(24.5, 1.4)$, $\forall i = 1, \ldots, 10$. Logo, pelas propriedades da distribuição normal, a quantidade média de poluente nessa amostra é $\overline{X} = \frac{1}{10}\sum_{i=1}^{10} \frown N\left(24.5, \frac{1.4}{\sqrt{10}}\right)$.

$P[24 < \overline{X} < 25] = P[\overline{X} < 25] - P[\overline{X} \leq 24] = P\left[\frac{\overline{X}-\mu}{\sigma/\sqrt{10}} < \frac{25-24.5}{1.4/\sqrt{10}}\right] - P\left[\frac{X-\mu}{\sigma/\sqrt{10}} \leq \frac{24-24.5}{1.4/\sqrt{10}}\right] =$
$= \Phi(1.13) - \Phi(-1.13) = \Phi(1.13) - 1 + \Phi(1.13) = 2 \times 0.87076 - 1 = 0.74152$ (pelas tabelas).

c) Sendo o fenómeno espacial estudado estacionário e isotrópico, a média e a variância são constantes na região em questão ($\mu = 24.5$ ppm e $\sigma^2 = 1.4^2 = 1.96$ ppm^2), o variograma e a covariância só dependem da distância h (em km) entre os pontos.

i) De acordo com o enunciado,

$$\gamma_{\text{ajustado}}(h) = \begin{cases} 0 & \text{se } h = 0 \\ C_0 + C\left[1.5\frac{h}{\phi} - 0.5\left(\frac{h}{\phi}\right)^3\right] & \text{se } 0 < h \leq \phi \\ C_0 + C & \text{se } h > \phi \end{cases}$$

em que $C_0 = 0.1$ é o efeito pepita, C é o patamar no modelo teórico e ϕ é a amplitude. Como a covariância é nula quando $h \geq 1$, $\phi = 1$. Além disso, $\sigma^2 = C_0 + C \Leftrightarrow 1.96 = C + 0.1 \Leftrightarrow C = 1.86$ ppm^2. Logo,

$$\gamma_{\text{ajustado}}(h) = \begin{cases} 0 & \text{se } h = 0 \\ 0.1 + 1.86(1.5h - 0.5h^3) & \text{se } 0 < h \leq 1 \\ 1.96 & \text{se } h > 1 \end{cases}$$

ii) $\gamma_{\text{ajustado}}(0.5) = 0.1 + 1.86 \times (1.5 \times 0.5 - 0.5 \times 0.5^3) = 0.69125$ ppm^2.