

Exercícios – Estatística e Delineamento 2016-17

0 Conceitos introdutórios de Estatística e do programa R

1. Um agricultor instalou um pluviómetro para medir a precipitação num dado terreno. Durante um ano, obteve os seguintes totais mensais (em mm):

| | | | | | |
|-----------|-------|--------|------|----------|-------|
| Janeiro | 101.0 | Maio | 26.7 | Setembro | 5.7 |
| Fevereiro | 60.7 | Junho | 10.5 | Outubro | 51.7 |
| Março | 75.1 | Julho | 2.5 | Novembro | 50.1 |
| Abril | 19.9 | Agosto | 39.8 | Dezembro | 170.6 |

Numa sessão de trabalho no programa R, responda às seguintes alíneas:

- (a) Crie um vector com os 12 totais mensais indicados. Designe o objecto criado por `precip`.
 - (b) Crie o vector `meses` com o nome dos 12 meses do ano.
 - (c) Associe a cada medição o nome do respectivo mês, utilizando o comando `names` do R.
 - (d) Calcule, com a ajuda dos comandos estatísticos elementares de que o R dispõe, as seguintes quantidades:
 - i. A precipitação total anual;
 - ii. A precipitação mensal média;
 - iii. A precipitação mensal mediana;
 - iv. A variância das precipitações mensais;
 - v. O desvio padrão das precipitações mensais;
 - vi. A precipitação mensal mínima;
 - vii. A precipitação mensal máxima.
 - (e) Seleccione o subvector
 - i. da precipitação no mês de Outubro;
 - ii. das precipitações nos meses de Junho a Setembro (inclusive);
 - (f) Seleccione o subvector dos meses com precipitação
 - i. superior a 50 mm;
 - ii. acima da média.
 - (g) Identifique, com auxílio de comandos do R:
 - i. qual o mês onde se verificou a precipitação mínima;
 - ii. qual o mês onde se verificou a precipitação máxima.
 - (h) Aplique o comando `plot` ao vector `precip` que criou na alínea 1a. Comente o resultado.
 - (i) Execute os comandos


```
> plot(precip, type="l")
> plot(precip, type="h")
```

 Comente o resultado.
2. O programa R disponibiliza alguns conjuntos de dados. Os seus nomes e breves descrições podem ser consultados através do comando

```
> data()
```

Entre estes dados encontra-se o vector `sunspots`, onde se registam o número médio de manchas solares observadas nos dias de cada mês, entre 1749 e 1983¹. Os valores podem ser vistos escrevendo apenas o nome do objecto.

- (a) Determine o comprimento do vector `sunspots`, utilizando o comando `length`.
 - (b) Crie um histograma dos valores registados, utilizando o comando `hist`:
 - i. deixando que o comando defina as classes de valores utilizadas;
 - ii. pedindo a criação de classes de comprimento 10, começando em zero e acabando em 260.
 - (c) Calcule, com a ajuda do comando `quantile`:
 - i. os três quartis (primeiro quartil, mediana e terceiro quartil) dos dados;
 - ii. o nono decil dos dados.
 - (d) Aplique o comando `summary` ao objecto `sunspots` e inspeccione o resultado.
 - (e) Construa um diagrama de extremos e quartis dos dados, utilizando o comando `boxplot`.
3. Uma experiência sobre o enraizamento de estacas semi-lenhosas de oliveiras visa saber se quatro diferentes tratamentos afectam a forma como as estacas se distribuem por três possíveis resultados.

Os resultados possíveis, para cada estaca, são:

- morte da estaca;
- estaca com calo;
- enraizamento da estaca.

Os quatro diferentes tratamentos utilizados foram:

- sem incisão/sem utilização de boro;
- com incisão/sem utilização de boro;
- sem incisão/com boro;
- com incisão/com boro.

Em cada tratamento, foram ensaiadas 60 estacas. Os resultados obtidos foram os seguintes:

| Tratamento | Resultado | | |
|----------------------|-----------|----------|--------------|
| | Morte | Com calo | Enraizamento |
| Sem incisão/sem boro | 26 | 18 | 16 |
| Com incisão/sem boro | 32 | 22 | 6 |
| Sem incisão/com boro | 24 | 24 | 12 |
| Com incisão/com boro | 39 | 19 | 2 |

- (a) Crie, numa sessão de trabalho do R, uma matriz de nome `estacas`, contendo os resultados da experiência (utilize o comando `matrix` do R). Tenha em atenção que, por omissão, o R introduz os dados na matriz *por columnas*.
- (b) Utilize os comandos `rownames` e `colnames` para dar nomes às linhas e colunas da matriz criada na alínea anterior.

¹Na realidade, `sunspots` não é um vector, mas um objecto de outro tipo, designado `ts` (as iniciais de *time series*, ou seja, série cronológica). No entanto, para aquilo que se pede neste exercício o objecto comporta-se como um vector.

- (c) Crie um vector contendo apenas os números de estacas enraizadas, para cada tratamento. Calcule o número total de estacas enraizadas.
- (d) Calcule as frequências absolutas marginais de linhas e de colunas, utilizando o comando `apply`.
- (e) Calcule as frequências relativas marginais de linhas e de colunas.
- (f) Seleccionar a submatriz correspondente aos tratamentos onde não se efectua a incisão na estaca.
4. A distribuição Binomial surge associada a uma variável aleatória X que conta o número de êxitos em m provas de Bernoulli. As distribuições Binomiais caracterizam-se por dois parâmetros: m , o número de provas; e p , a probabilidade de êxito em cada uma dessas provas. Utiliza-se a notação $X \cap B(m, p)$ para indicar que X tem distribuição Binomial com parâmetros m e p , ou seja, que:

$$P[X = x] = \binom{m}{x} p^x (1-p)^{m-x} \quad (x = 0, 1, 2, \dots, m)$$

Admita-se, por exemplo, que a probabilidade de uma ovelha dar à luz um borrego fêmea é 0.5. Pode então admitir-se que o número de borregos fêmeas no lote de 8 crias será dado por uma variável aleatória $X \cap B(8, 0.5)$.

- (a) O comando `dbinom` calcula valores da função de massa probabilística numa distribuição Binomial. Utilize esse comando para determinar a probabilidade de haver 5 crias fêmeas no lote de 8 crias.
- (b) Calcule $P[X = x]$ para todos os possíveis valores de x e construa o diagrama de barras da função de massa probabilística.
- (c) O valor esperado duma variável aleatória discreta X , que toma um número finito de valores x_i com probabilidades p_i ($i = 1, 2, \dots, k$), é dado por $E[X] = \sum_{i=1}^k x_i p_i$. Utilizando o comando `dbinom`, determine o valor esperado duma v.a. $X \cap B(8, 0.5)$. Verifique que o resultado corresponde à expressão conhecida para o valor esperado duma Binomial $B(m, p)$: $E[X] = mp$.
- (d) O comando `pbinom` calcula valores da função distribuição cumulativa numa distribuição Binomial, ou seja, as probabilidades $P[X \leq x]$. Utilize este comando para calcular a probabilidade de que haja menos de seis fêmeas no lote de 8 crias.
- (e) Calcule a probabilidade de pelo menos metade das crias serem fêmeas, utilizando o comando `pbinom`.
5. Diz-se que uma variável aleatória X tem distribuição de Poisson, com parâmetro λ , e escreve-se $X \cap P(\lambda)$, se

$$P[X = x] = \frac{e^{-\lambda} \lambda^x}{x!}, \quad (\forall x \in \mathbb{N}_0)$$

A distribuição de Poisson surge frequentemente associada à contagem de acontecimentos raros. O parâmetro λ é, simultaneamente, o valor esperado e a variância de X , ou seja, $E[X] = V[X] = \lambda$. Considere um estudo sobre a severidade de uma doença causada por um fungo da videira, que é de crescimento micelial rápido. Nesse estudo são dispostos pedaços de planta (madeira com necrose) em várias placas de Petri. Ao fim de 5 dias conta-se o número de colónias por cada placa de Petri. O número médio de colónias por placa, ao fim de 5 dias, é de 2.2. O estudo revela que pode admitir-se que o número de colónias por placa ao quinto dia segue uma lei de Poisson, com parâmetro $\lambda = 2.2$.

- (a) Verifique que o comando `dpois(3,2.2)` devolve a probabilidade de haver 3 colónias numa placa.

- (b) Determine a probabilidade de, numa dada placa, se ter mais de 4 colónias (utilize o comando `ppois`).
- (c) Determine a probabilidade de haver 2, 3 ou 4 colónias numa placa.