

Pergunta 1

Respondida

Nota: 1,80 em 2,00

No ficheiro de dados [econ.xlsx](#) encontra informação relativa a dados económicos dos Estados Unidos providenciados pela empresa [FRED](#). Este conjunto de dados possui as seguintes variáveis: tempo (Data do registo); gcp (gastos de consumo pessoal, em biliões de dólares); pop (população total); tpp (taxa de poupança pessoal); ddesemp (duração mediana do desemprego, em semanas); ndesemp (número de desempregados, em milhares).

Considere as variáveis  $x_1 = \text{gcp}$  e  $x_2 = \text{ndesemp}$  para os anos superiores ou iguais a 1998. Com recurso ao pacote `ggplot` produza um único gráfico que lhe permita fazer uma análise da evolução dessas duas variáveis para esses anos.

Uma vez que as variáveis podem não ter a mesma escala, antes de construir o gráfico proceda do seguinte modo:

- a. Selecione os dados a usar.
- b. Faça a seguinte transformação aos dados associados a cada variável

$$X_k : z_{ik} = \frac{x_{ik} - \bar{x}_k}{sx_k}, \quad i = 1, 2, \dots, n,$$

onde  $n$  é a dimensão da amostra,  $\bar{x}_k$  e  $sx_k$  correspondem, respectivamente, à média e desvio-padrão da amostra associada à variável  $x_k$ .

Submeta um ficheiro em formato PDF, com uma única página A4, que inclua:

- 1. O código em `R`.  
**Nota:** no código devem também constar os comandos para leitura e seleção dos dados do ficheiro.
- 2. O gráfico que achar mais adequado para analisar a evolução dessas variáveis nesse período de tempo.

**Pergunta 2**

Respondida

Nota: 2,00 em 2,00

O ficheiro [TIME\\_USE\\_24092022.csv](#) contém uma compilação de dados enviados por diversos países para a OCDE ([Organização para a Cooperação e Desenvolvimento Económico](#)) sobre o tempo médio diário (em minutos) despendido pelas pessoas entre os 15 e os 64 anos em diferentes tipos de ocupações.

1. Leia o ficheiro de dados no **R** e elimine todos os registos referentes à África do Sul (dados incompletos).
2. Submeta um ficheiro em formato PDF com uma única página A4, que inclua, num único gráfico, dois diagramas de extremos e quantis que permitam comparar os tempos médios diários registados para **Homens** em duas ocupações distintas: **Cuidados pessoais** e **Trabalho não remunerado**.

**Nota:** o código apresentado deve incluir os comandos para leitura e seleção dos dados do ficheiro.

**Pergunta 3**

Respondida

Nota: 2,00 em 2,00

O ficheiro [GENDER\\_EMP\\_19032023152556091.txt](#) contém uma compilação de dados sobre emprego enviados por diversos países para a OCDE ([Organização para a Cooperação e Desenvolvimento Económico](#)).

Com recurso ao pacote **ggplot** produza um único gráfico de barras que permita comparar os valores da variável **EMP3** (*Unemployment rate, by sex and age group*) entre homens e mulheres nos grupos etários 15–24, 25–54 e 55–64, registados em **2010** no seguinte país: **Sweden**.

Por simplicidade, mantenha todo o texto no gráfico em Inglês.

Submeta um ficheiro em formato PDF com uma única página A4, que inclua:

1. O código em **R**, que deve incluir os comandos para leitura e seleção dos dados do ficheiro.
2. O gráfico produzido.

Pergunta 4

Correta

Nota: 2,00 em 2,00

- 1. Fixando a semente em 2315, gere uma amostra de dimensão  $k = 1396$  proveniente de uma distribuição Exponencial de parâmetro  $\lambda = 10.5$ . Os valores gerados correspondem aos tempos entre acontecimentos sucessivos.
- 2. Considere agora a soma sucessiva destas observações, i.e., se  $x_i$  designar o  $i$ -ésimo valor gerado, então  $s_j = \sum_{i=1}^j x_i$  é o instante de ocorrência do  $j$ -ésimo acontecimento. Seja  $T = \lceil s_{1396} \rceil$  o menor número inteiro maior ou igual ao instante de ocorrência do último acontecimento.
- 3. Divida o intervalo  $]0, T]$  em intervalos de amplitude unitária e contabilize o número de acontecimentos que ocorreram em cada um desses subintervalos.
- 4. Calcule a média do número de acontecimentos por subintervalo e de seguida calcule o desvio absoluto entre este valor e o valor esperado (teórico) do número de acontecimentos num subintervalo. Indique este desvio arredondado a 4 casas decimais.

Pergunta 5

Correta

Nota: 2,00 em 2,00

Ensaio de Bernoulli independentes, cada um dos quais com probabilidade de sucesso 0.3, são sucessivamente realizados. Seja  $x$  o número de insucessos até ao primeiro ensaio que resulta em sucesso. A distribuição da variável aleatória  $x$  é conhecida por distribuição geométrica de parâmetro  $p = 0.3$ , cuja função (massa) de probabilidade é dada por:

$$f_X(x) = \begin{cases} (1 - p)^x p, & x = 0, 1, 2, \dots \\ 0, & \text{caso contrário.} \end{cases}$$

Podemos gerar valores de uma distribuição geométrica a partir de uma distribuição uniforme contínua usando o método de transformação inversa. Nesse sentido, requer-se a execução dos seguintes passos:

- i. Simula-se um valor,  $u$ , proveniente de uma distribuição uniforme no intervalo  $]0, 1[$ .
- ii. Se  $F_X(x - 1) < u \leq F_X(x)$ , aceita-se  $x$  como um valor simulado de  $X$ , onde  $F_X(x)$  é a função de distribuição de  $X$ .

Fixando a semente em 1126, implemente este método de simulação estocástica repetindo os passos anteriores até obter uma amostra de dimensão  $n = 1041$ .

Indique a proporção de valores simulados que são superiores à soma da média com o desvio padrão amostrais, de entre os que são superiores à respetiva média amostral. Apresente o resultado com 4 casas decimais.

Pergunta 6

Correta

Nota: 2,00 em 2,00

Considere a variável aleatória  $x$  que representa o primeiro algarismo de um número inteiro escrito em base decimal. Admita que  $x$  possui distribuição de Benford, com função de probabilidade dada por:

$$P(X = x) = \log_{10}\left(1 + \frac{1}{x}\right), \quad x \in \{1, 2, \dots, 9\}.$$

- 1. Calcule a probabilidade de  $x$  ser igual a 5 ou 7.
- 2. Obtenha a fração de potências de dois no intervalo  $[2^3, 2^{30}]$  cujo primeiro algarismo é igual a 5 ou 7.
- 3. Calcule o desvio absoluto entre os valores calculados em 1. e 2.
- 4. Indique este desvio arredondado a 4 casas decimais.

Pergunta 7

Correta

Nota: 2,00 em 2,00

Fixando a semente em 1473, simule  $m = 1621$  amostras de dimensão  $n = 19$  de uma população normal de média nula e variância unitária. Para cada uma das amostras, calcule a soma dos quadrados dos valores observados.

Indique a diferença em valor absoluto (arredondado a 4 casas decimais), entre o quantil de probabilidade 0.3 da amostra das somas dos quadrados dos valores observados e o quantil correspondente à distribuição teórica da soma de quadrados de variáveis normais reduzidas independentes.

**Nota:** Use a função `quantile` com a opção `type=2`.

Pergunta 8

Respondida

Nota: 2,00 em 2,00

Considere uma variável aleatória com distribuição de Cauchy, com parâmetros de localização e escala iguais a 1.2 e 1.6, respectivamente.

Usando o R e fixando a semente em 1637, gere uma amostra de dimensão  $n = 141$  desta população.

Represente num único gráfico:

- 1. Os valores gerados ordenados por ordem crescente versus os quantis de probabilidade  $i/(141 + 1)$ ,  $i = 1, \dots, 141$  desta população.
- 2. Os valores gerados ordenados por ordem crescente versus os quantis de probabilidade  $i/(141 + 1)$ ,  $i = 1, \dots, 141$  de uma população normal com valor esperado  $\mu = 2.2$  e variância  $\sigma^2 = 3.6$ .
- 3. A recta bissectriz dos quadrantes ímpares.

Submeta um ficheiro em formato PDF, com uma única página A4, que inclua:

- 1. O código em R.
- 2. O gráfico produzido.

Pergunta 9

Respondida

Nota: 1,90 em 2,00

Para a construção de intervalos de confiança para o parâmetro  $p$  de uma distribuição de Bernoulli podemos recorrer à variável fulcral

$$Z_1 = \frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \stackrel{a}{\sim} N(0,1)$$

obtida pela aplicação do teorema do limite central a uma amostra aleatória de tamanho  $n$  suficientemente grande da referida população. Duas variantes são possíveis:

Método 1

Usando  $z_1$ , não é difícl mostrar que os limites do intervalo de confiança são as soluções da seguinte equação do segundo grau em  $p$ :

$$\bar{X}^2 - 2p\bar{X} + p^2 - z^2 \frac{p(1-p)}{n} = 0,$$

em que  $\bar{x}$  representa a média amostral e  $z = \phi^{-1}\left(\frac{1+\gamma}{2}\right)$ , para um nível de confiança aproximado  $\gamma \in ]0,1[$ .

Método 2

Uma segunda aproximação conduz à variável fulcral

$$Z_2 = \frac{\bar{X} - p}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}} \stackrel{a}{\sim} N(0,1)$$

que permite a construção de intervalos de confiança de uma forma mais simples e habitual.

Com o objetivo de comparar os dois métodos e, em particular, avaliar a adequação da segunda aproximação, implemente os seguintes passos no **R**:

1. Fixe a semente em 1296 e para cada valor de  $n \in \{30, 50, 100, 200, 300, 500, 1000\}$ :
  - a. gere  $k = 3000$  amostras de tamanho  $n$  de uma distribuição de Bernoulli com parâmetro  $p = 0.3$ ;
  - b. para cada amostra gerada, calcule a diferença entre os comprimentos dos intervalos de confiança construídos pelo **Método 2** e pelo **Método 1**, com um nível de confiança aproximado  $\gamma = 0.97$ .
  - c. calcule a média das  $k = 3000$  diferenças anteriores.
2. Construa um gráfico que ilustre a variação das diferenças médias em função do tamanho da amostra.

Submeta um ficheiro em formato PDF, com uma única página A4, que inclua:

1. O código em **R**.
2. O gráfico pedido.
3. Comentários sobre os resultados obtidos.

Pergunta 10

Correta

Nota: 2,00 em 2,00

Considere uma variável aleatória  $X$  com distribuição Normal de valor esperado  $\mu$  desconhecido e variância  $\sigma^2 = 4$ . Construa um teste de hipóteses  $H_0: \mu = 23.7$  contra  $H_1: \mu \neq 23.7$ , ao nível de significância de  $\alpha = 0.04$ .

Com recurso ao R e fixando a semente em 443, gere  $m = 300$  amostras de dimensão  $n = 44$  dessa variável, admitindo que  $\mu = 25.1$ . Aplique o teste de hipóteses que construiu para cada amostra gerada, e use o conjunto de resultados para obter uma estimativa da probabilidade do teste conduzir à não rejeição de  $H_0$ . Indique o resultado com 3 casas decimais.