

Métodos Computacionais para Estatística II

Prof: Jony Arrais Pinto Junior

Lista 04

1. Crie a seguinte função no R e faça o que se pede abaixo.

$$f(x, k) = \frac{\Gamma(k)e^{-kx-3}}{k}, x, k > 0.$$

- (a) Plote a função $f(x, 4)$, $-2 < x < 6$.
 - (b) Plote a função $f(2, k)$, $-2 < k < 10$.
 - (c) Maximize a função acima, considerando que $k = 2$.
 - (d) Plote a função que foi maximizada no item anterior e uma linha vertical no valor que maximiza a função.
2. Crie uma lista com amostras de tamanho 10, 20, 32, 42, 78 da variável Poisson(4), Gama(1,1), $U\{1,2,3\}$, Normal(10,100) e Binomial(3,0.9), respectivamente. Usando a função `sapply` obtenha
- (a) as médias amostrais
 - (b) as variâncias amostrais
 - (c) os quantis
 - (d) a soma de cada valor ao quadrado
 - (e) o produto de todos os valores da amostra
3. Encontre uma estimativa de máxima verossimilhança do parâmetro desconhecido das distribuições abaixo se foram observadas as amostras fornecidas em cada caso (Faça cada item maximizando a função de verossimilhança e a função de log-verossimilhança).
- (a) $X \sim \text{Gama}(3, \beta)$, $x = (1, 2, 2.4, 2.8, 5, 1, 3, 4, 6.3, 2.9)$
 - (b) $X \sim \text{Normal}(\mu, 10)$, $x = (10, 12, 11, 12.8, 13, 14.9, 12, 16.2)$
 - (c) $X \sim \text{Geométrica}(p)$, $x = c(2, 3, 2, 2, 3, 3, 3, 4, 4, 5, 7, 2, 2, 2)$
 - (d) $X \sim \text{Normal}(10, \sigma^2)$, $x = c(25, 23, 22, 21, 27, 39, 35, 33, 32)$
 - (e) $X \sim \text{Exponencial}(\eta)$, $x = (4, 5, 6.2, 4, 3, 5, 6.9, 7, 9.3)$
4. Para checar o resultado do teorema do limite central faça o seguinte: (1) gere 1000 amostras de tamanho n , (2) Calcule a média de cada amostra (3) Faça um histogramas com as médias das amostras e compare com a curva da respectiva normal apontado pelo teorema. Siga os passos para todos os seguintes valores de $n = 2, 5, 10, 50, 100$. O que acontece a medida que n aumenta em cada um dos casos abaixo?
- (a) $X \sim \text{Gama}(2, 8)$
 - (b) $X \sim \text{Geométrica}(0.8)$
 - (c) $X \sim \text{Uniforme}(2, 20)$

x	$P(X = x)$
1	0,05
2	0,35
3	0,15
4	0,05
5	0,20
6	0,15
7	0,05

(d) $X \sim \text{Normal}(2,8)$

(e) $X \sim \text{Binomial}(20,0.8)$

(f) $X \sim \chi^2_{(3)}$

5. Sejam X_1, X_2, \dots, X_n , dado μ e σ^2 , uma a.a.s de uma distribuição $N(\mu, \sigma^2)$. Um resultado importante diz que a distribuição de

$$V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)},$$

em que $S^2 = \sum \frac{(X_i - \bar{X})^2}{n-1}$. Ilustre esse resultado usando uma simulação. **Dica:** Você pode seguir os seguintes passos para ilustrar o resultado:

- escolha os parâmetros de uma distribuição normal;
 - escolha o tamanho da amostra, dado por n , e o número de amostras de tamanho n , dado por N ;
 - Gere N amostras de tamanho n ;
 - Para cada amostra calcule S^2 e $V = (n-1)S^2/\sigma^2$;
 - Faça um histograma com os valores de V e compare com a curva de uma distribuição $\chi^2_{(n-1)}$.
6. Existe um resultado que diz que soma de distribuições normais tem distribuição normal. Como você checaria esse resultado computacionalmente? Suponha 5 normais quaisquer, todas com parâmetros diferentes.
7. Leia o banco de dados *Banco de Métodos II - I* do site da disciplina no SPSS e faça o que se pede:
- (a) Um gráfico apropriado para cada uma das variáveis *Situação Atual no Trabalho*, *Idade*, *Escolaridade*, *Ansiedade*, *Depressão* e *Frequenta algum Grupo Religioso*.
 - (b) Refaça os gráficos do item anterior considerando os grupos, isto é, o gráfico de cada variável tem que ser feito para cada grupo e devem ser apresentados numa mesma figura.
8. Ainda considerando as variáveis do banco anterior faça o que se pede:
- (a) Calcule a frequência para as variáveis *Situação Atual no Trabalho* e *Frequenta algum Grupo Religioso*.
 - (b) Calcule a média, o desvio padrão, o mínimo e o máximo para as variáveis *Idade*, *Escolaridade*, *Ansiedade* e *Depressão*.

- (c) Recalcule todas as estatísticas descritivas pedidas no item anterior para as mesmas variáveis dentro de cada grupo, por exemplo, calcule a idade média para o grupo adesão e a idade média para o grupo não adesão.
9. Ainda com relação ao mesmo banco, crie uma nova variável que será obtida em função da variável escolaridade, de modo que
- Se escolaridade for igual a 0 anos, escolaridade2 será igual a analfabeto.
 - Se escolaridade estiver entre 1 e 8 anos, escolaridade2 será igual a ensino fundamental.
 - Se escolaridade estiver entre 9 e 11 anos, escolaridade2 será igual a ensino médio.
 - Se escolaridade for superior a 11 anos, escolaridade2 será igual a ensino superior ou mais.
10. Leia a mesma base utilizadas nas questões anteriores no R e faça o que se pede:
- (a) Acrescente na base uma variável que categoriza a variável depressão: 0 - se depressão < 18 e 1 - caso contrário.
- (b) Acrescente na base uma variável que categoriza a variável escolaridade como na questão anterior.
- (c) Acrescente uma variável que é média entre as pontuações da depressão e ansiedade dos indivíduos.
- (d) Salve essa base com extensões csv e dta.
- (e) Leia as bases salvas e verifique se as mesmas foram criadas corretamente.
11. Crie uma função que calcula o poder do teste para os seguintes testes. **Dica:** Antes de calcular o poder será necessário definir sua região crítica e para isso você precisará especificar um nível de significância, sendo assim sua função deve receber α como uma das entradas.
- (a) Teste para a média unilateral à direita com variância populacional conhecida.
- (b) Teste para a média unilateral à esquerda com variância populacional conhecida.
- (c) Teste para a proporção bilateral.
- (d) Teste para a média unilateral à direita com variância populacional desconhecida.
- (e) Teste para a média bilateral com variância populacional desconhecida.
12. Faça um plot de todas as funções criadas na questão anterior.
13. Crie funções que calcula intervalos de confiança para a média, variância e proporção de uma população, diferença de médias e diferença de proporções de duas populações e razão de variâncias de duas populações.
14. A diretoria de uma firma que fabrica um determinado produto resolveu que retiraria o produto da linha de produção se a média de consumo *per capita* fosse menor do que 8Kg. Caso contrário, continuaria a fabricá-lo. Foi realizada uma pesquisa de mercado e observou-se a seguinte amostra $x = (8.61, 8.72, 8.41, 6.2, 8.73, 8.57, 8.84, 5.91, 7.89, 7.58, 4.95, 8.31, 7.76, 8.57, 7.21, 6.17, 6.45, 7.09, 7.76, 8.86, 6.58, 7.75, 11.01, 5.16, 7.22)'$.
- (a) Construa um teste de hipótese adequado, utilizando $\alpha = 0,05$, e com base na amostra colhida determine a decisão a ser tomada pela diretoria.

- (b) Se $\alpha = 0, 1$, a decisão seria outra?
 - (c) Qual a probabilidade β de se tomar uma decisão errada, se na realidade, a média populacional for $\mu = 7, 8\text{Kg}$?
 - (d) Calcule β com base em $\alpha = 0, 1$. Obs. Crie uma função para calcular o erro tipo II pedido.
15. Gere 700 amostras de tamanho 25 de uma normal com média 10 e variância 3. Armazene essas amostras em uma matriz e faça o que se pede:
- (a) Calcule o intervalo de confiança para a média populacional com nível de confiança de 95% para cada amostra.
 - (b) Calcule o percentual de intervalos que contém a verdadeira média populacional. O resultado obtido foi razoável? Por que?
 - (c) Faça o mesmo para a variância populacional.
16. Leia o banco de dados *Banco Lista 06* e apresente o teste para responder os seguintes questionamentos:
- (a) A proporção de pessoas com idade superior a 80 anos é superior a 10%?
 - (b) A proporção de homens é 50%?
 - (c) A proporção de pessoas com escolaridade inferior a 10 anos é superior a 60%?
 - (d) A média populacional da variável MMSE é igual a 30?
 - (e) A média da variável CAMCOG é menor do que 87?
 - (f) A média da variável memória para pacientes que aceitaram fazer terapia é igual a média dos que recusaram?
 - (g) A média da variável fluência verbal para pacientes que aceitaram fazer terapia é menor do que a média dos que recusaram?
 - (h) A média da variável reconhecimento para pacientes que aceitaram fazer terapia é igual a média dos que recusaram?

Legenda:

MMSE - Mini Estado do Exame Mental

Grupo: 1 - aceitaram fazer terapia e 2 - recusaram a fazer terapia

Gênero: 1 - Masculino e 2 - Feminino

fluenciaverbal: Teste da fluência verbal

Os testes MMSE, reconhecimento, fluência verbal, CAMCOG e memória são testes para diagnosticar a capacidade funcional de pessoas e, quanto menor a pontuação de um indivíduo nesses testes melhor o paciente se encontra.

17. Desenhe o gráfico do poder e do erro tipo II para os testes realizados dos itens (a) ao (e). E interprete os mesmos.
18. Leia o banco *Mortes violentas* e faça o que se pede:
- (a) Crie uma base que contenha somente os dados referentes aos anos de 2008 e 2010 e que a variável suicídio seja igual a 1. Deixe na base somente as variáveis *codmunocor*, *idade*, *sexo*, *racacor*, *homicidio*, *homicidio.paf*.
 - (b) Acrescente as microregiões na base de dados criada no item anterior.

- (c) Com relação a base criada faça o que se pede:
- Sexo: 1 - Masculino e 2 - Feminino
 - Raça: 1 - Branco, 2 - preta, 3 - parda, 4 - amarela e 5 - indígena
 - Homicídio: 0 - Não e 1 - Sim;
 - Homicídio por arma de fogo: 0 - Não e 1 - Sim;
 - Idade: assume valores de 1 até 100.
 - Sendo assim, qualquer outro valor para estas variáveis será considerado um dado faltante, façam esta recodificação.
- (d) Crie uma base que contenha a quantidade homicídio, quantidade de homicídio por arma de fogo e idade média para cada microregião.
- (e) Crie uma base que contenha a quantidade homicídio, quantidade de homicídio por arma de fogo e idade média para cada microregião por gênero.
19. Crie uma função que calcula o p-valor para os seguintes testes.
- (a) Teste para a média unilateral à direita com variância populacional conhecida.
 - (b) Teste para a média unilateral à esquerda com variância populacional conhecida.
 - (c) Teste para a média bilateral com variância populacional conhecida.
 - (d) Teste para a variância bilateral de uma população normal.
20. A seguir são apresentadas algumas situações problemas. Aplique uma das funções criadas no exercício 14 e tire uma conclusão. Suponha que as amostras são normais, isto é, aqui não é preciso fazer o teste de normalidade, só mesmo usar uma das função que vocês criaram e chegar a uma conclusão com base no p-valor obtido.
- (a) Uma empresa está afirmando que o consumo de combustível (em litros) dos modelos *A* e *B* são diferentes. Para checar tal suposição foi observada a seguinte amostra do consumo de gasolina por 100 Km rodados do modelo *A*,

$$12; \quad 13; \quad 12,5; \quad 14; \quad 12,9; \quad 11,7; \quad 12,3; \quad 12; \quad 13,3$$
 e do modelo *B*,

$$10; \quad 11; \quad 9,5; \quad 8; \quad 9; \quad 9,7; \quad 10,3; \quad 10.$$
 Com base no resultado qual sua conclusão?
 - (b) Um psicólogo desconfia que o QI médio dos moradores do Ingá é superior a 70 pontos. Com base na amostra abaixo, qual a sua conclusão?

$$75, \quad 78, \quad 66, \quad 70, \quad 88, \quad 85, \quad 80, \quad 69.$$
 - (c) Um nutricionista acredita que o ganho calórico de pessoas que estão sendo submetidas a uma dieta para ganhar peso não é superior a 2 Kg por mês. Com base na amostra abaixo, qual a sua conclusão?

$$1; \quad 1,1; \quad 1,95; \quad 2; \quad 2,9; \quad 1,97; \quad 1,3; \quad 2,5.$$
21. Leia o banco de dados *Banco Lista 07* e apresente o teste para responder os seguintes questionamentos. Faça todos os testes no R e refaça no SPSS aqueles que forem possíveis:
- (a) O peso depois da dieta diminuiu?
 - (b) A pontuação do CAMCOG antes e depois da terapia são diferentes?
 - (c) E a Pontuação em uma prova de conhecimentos gerais aumentou depois que os alunos foram submetidos à aulas de reforço?