

Estatística Computacional

Licenciatura em Ciência de Dados, 2º ano

Preâmbulo

O objetivo destas fichas é dar-vos a oportunidade de estudarem a matéria e aplicarem os conceitos, consolidando assim os conhecimentos.

É da vossa responsabilidade e da vossa consciência trabalharem efetivamente nos problemas propostos.

As fichas podem/devem ser realizadas por pares de alunos. Aproveitem para estudar em conjunto e discutir abordagens.

Ficha 3

- No topo do documento devem incluir a vossa identificação (Nome e número de aluno).
- O próprio nome do documento a entregar deve seguir o formato Ficha3_Id1_Id2, onde Id1 e Id2 correspondem à identificação dos alunos que realizaram a ficha.
- A resposta a esta ficha deve constar de um único script em R, ou, se preferirem e souberem usar, um documento integrado em R Markdown (neste caso devem enviar quer o ficheiro fonte quer o resultado em pdf).
- Usem comentários!
 - Cada pergunta deve estar bem identificada.
 - O código utilizado deve estar comentado.
 - Os resultados obtidos devem ser comentados.
- Os ficheiros devem ser enviados por e-mail (alunos dia: teresa.calapez@iscte-iul.pt; alunos PL: Andre_Fernandes@iscte-iul.pt)

A.

Numa empresa de sondagens existem duas equipas, com dois chefes alocados. Cada chefe foi incumbido de calcular um determinado indicador de produtividade composto (ou seja, que inclui várias variáveis, com vários pesos), que varia entre 0 e 10 pontos, com possibilidade de existência de números não inteiros. Quanto mais próximo de 10, mais produtiva é a equipa. Foi recolhida uma amostra de 20 trabalhos para cada equipa onde esse indicador foi medido, estando os resultados desta amostra presente neste ficheiro (dados_equipas.xlsx). Sabe-se que o valor deste indicador segue uma distribuição normal, onde também não se conhece o verdadeiro valor para a variância. Com base na amostra recolhida, o diretor da empresa solicitou aos chefes que fornecessem uma estimativa por intervalos, com uma confiança de 95%, para o verdadeiro valor médio da produtividade de cada equipa.

1. (2.0) Determine os Intervalos de Confiança concretos para o parâmetro em causa (a média), tendo como base a amostra existente no ficheiro dados_equipas.xlsx. Num comentário, interprete os intervalos de confiança obtidos (um para cada equipa, separadamente).
2. (2.0) Indique o valor da Margem de Erro (Erro Amostral) para os intervalos em causa.
3. (2.0) A Direção exige uma confiança maior na previsão efetuada. Por isso, decidiu solicitar que os Intervalos de Confiança em causa fossem recalculados para uma confiança de 99%. Calcule os novos intervalos de confiança, usando a amostra existente no ficheiro (dados_equipas.xlsx). Num breve comentário, comente qual o efeito desse aumento do nível de confiança nos intervalos encontrados.
4. (2.0) Exercício Teórico. Admitindo que a variância populacional do **Indicador de Produtividade da Equipa 1** é 0.36, e que a distribuição desse Indicador pode ser considerada Normal, qual a dimensão da amostra a considerar para garantir um erro máximo de 0.2 na estimação por intervalos da respetiva média, a 95% de confiança?

B.

Numa reunião com a Direção, o chefe da equipa 2 afirmou o seguinte:

Chefe equipa 2: “A produtividade média da minha equipa é superior a 7 pontos”

- a) (1.0) Tendo em conta a opinião do chefe da equipa 2, se se pretender fazer um teste de hipóteses sobre o valor médio do indicador em causa, formule/indique as hipóteses a serem consideradas no respetivo Teste de Hipóteses.
- b) (2.0) Indique se o chefe da equipa 2 tem razão. Para o efeito, efetue um teste de hipóteses para o parâmetro em causa, utilizando o nível de significância de $\alpha = 0.05$ e utilizando os dados amostrais recolhidos no ficheiro dados_equipas.xlsx. Interprete os resultados obtidos.

O chefe da equipa 2 prosseguiu a reunião dizendo:

Chefe equipa 2: “Posso afirmar que não existe diferença entre a produtividade média da minha equipa e a produtividade média da equipa 1.”

- c) (1.0) Tendo em conta a opinião do chefe da equipa 2, se se pretender fazer um teste de hipóteses sobre o valor médio do indicador em causa, formule/indique as hipóteses a serem consideradas no respetivo Teste de Hipóteses.
- d) (3.0) Indique se o chefe da equipa 2 tem razão. Para o efeito, efetue um teste de hipóteses para o parâmetro em causa, utilizando o nível de significância de $\alpha = 0.05$ e utilizando os dados amostrais recolhidos no ficheiro dados_equipas.xlsx. Interprete os resultados obtidos.
- e) (1.0) Num breve comentário, indique qual o tipo de erro em que poderá estar a incorrer, com as conclusões que retirou no ponto d).
- f) (4.0) O Chefe da equipa 1 afirma que a produtividade média da sua excede em mais de 1 ponto em mais de 1 ponto (na escala de 0 a 10) a produtividade média da equipa 2. Terá razão? i) Formule as hipóteses, ii) realize o teste e iii) interprete os resultados obtidos.