

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 2

- 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 Ficha1_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.
- Nesta primeira ficha, os ficheiros devem ser enviados por e-mail (alunos dia: teresa.calapez@iscte-iul.pt; alunos PL: Andre_Fernandes@iscte-iul.pt)

A.

Uma equipa de uma determinada empresa de sondagens necessita de contactar potenciais eleitores, para elaborar as sondagens requisitadas. Por dia, esta equipa consegue contactar em média, cerca de 20 pessoas. De referir que as pessoas seleccionadas são contactadas de forma independente e aleatória. Desta forma, o número de pessoas contactadas num dia, não afeta o número de pessoas contactado no dia seguinte.

O chefe de equipa definiu que para a equipa cumprir o planeamento a equipa devia contactar por dia pelo menos 15 pessoas. O chefe fica satisfeito se o planeamento for cumprido.

1. Num comentário em R, defina a variável em causa, indicando o respetivo modelo de distribuição de probabilidades (nome e parâmetros).
2. Tendo em conta a distribuição em causa, crie o respetivo gráfico da função de probabilidades e da função de distribuição.
3. Qual a probabilidade de em determinado dia, o chefe ficar satisfeito?
4. Qual deverá ser o número de contactos mínimo por dia, para que seja, pelo menos 0.95 a probabilidade do chefe ficar satisfeito?
5. Qual a probabilidade de num determinado mês (30 dias), serem inquiridas no máximo 550 pessoas?
6. Crie uma amostra aleatória de 365 observações (1 ano) com base na distribuição do número de contacto diários. Represente graficamente a informação recolhida. Calcule e interprete a média amostral, no contexto dado.
7. Qual a probabilidade de em 15 dias, existirem 10, onde o chefe fica satisfeito?
8. Qual a probabilidade de em 15 dias, existirem 13, onde o chefe fica satisfeito? Compare o resultado com a alínea anterior e num breve comentário em R, comente o porquê da diferença.

B.

A variável aleatória X , que designa os custos para manter uma determinada equipa em funcionamento (em milhares de euros) por mês, tem a seguinte função densidade de probabilidade:

$$f(x) = \frac{1}{496}x^2 + \frac{1}{24}, \quad 2 < x < 10$$

9. Calcule as seguintes probabilidades:

1. $P(X = 5)$;
2. A probabilidade do custo de manutenção da equipa ser superior a 6 mil euros.

10. Represente o respetivo gráfico da função densidade de probabilidade.

11. Calcule o valor esperado de X .

C.

O número de horas necessário para que uma equipa efetue um determinado trabalho é uma variável aleatória com distribuição normal de média 180 horas e desvio-padrão de 36 horas.

12. Sabendo que a partir de 200 horas, a equipa começa a receber horas extra, qual a probabilidade de de existirem horas extra pagas num determinado trabalho?

Para executar um projeto de grande envergadura, a empresa necessitou de utilizar uma equipa interna (onde o número de horas afetas ao projeto segue uma distribuição normal de média 180 horas e desvio-padrão 36 horas) e uma equipa externa (onde o número de horas afetas ao projeto segue uma distribuição normal de média 100 horas e desvio-padrão 49 horas).

13. Qual a probabilidade de serem necessárias menos de 350 horas para terminar o projeto?

14. Qual a duração máxima para esse projeto que garante ser a capacidade de trabalho do conjunto das duas equipas suficiente para o término do projeto em 90% dos casos?

15. Num breve comentário em R, indique se concorda com o facto da empresa projetar um orçamento para o seu cliente onde são orçamentadas 230 horas para concluir este trabalho de grande envergadura.

16. Efetue uma simulação para 1000 trabalhos onde é utilizada uma equipa interna e uma equipa externa. Simule separadamente as horas afetas à equipa interna e à equipa externa, calculando o número de horas totais afetas ao projeto através da soma. Crie o respetivo gráfico que permite observar os resultados da simulação.

17. Verifique em quantos desses 1000 trabalhos não se ultrapassaram as 350 horas totais de trabalho. Compare com o valor que esperava obter a partir dos resultados de 13.