

Coleta e Modelagem dos Dados de Entrada

Prof. Marcelo Xavier Guterres

1 Exercício: Questionário sobre Testes de Aderência à Normalidade

Leia atentamente os seguintes artigos:

1. **"Utilização dos testes estatísticos de Kolmogorov-Smirnov e Shapiro-Wilk para verificação da normalidade para materiais de pavimentação"** de Lopes et al.
2. **"Normalidade de variáveis: métodos de verificação e comparação de alguns testes não-paramétricos por simulação"** de Torman et al.

Em seguida, responda às questões abaixo:

1. Por que é importante verificar a normalidade dos dados antes de realizar análises estatísticas?
2. Quais são os principais métodos gráficos para avaliar visualmente a normalidade de uma variável? Quais são as vantagens e desvantagens desses métodos?
3. Cite pelo menos 5 testes estatísticos de aderência à normalidade mencionados nos artigos.
4. De acordo com os resultados da simulação realizada no segundo artigo, quais testes de normalidade tiveram o melhor desempenho geral? E quais tiveram o pior?
5. Como o tamanho da amostra afeta o desempenho dos testes de normalidade, segundo os resultados apresentados?
6. Para amostras pequenas ($n \leq 10$), qual é a recomendação dos autores em relação aos testes de normalidade?

7. Por que a distribuição t-student foi a mais difícil de ser identificada como não-normal pelos testes?
8. Quais são as implicações de usar técnicas estatísticas paramétricas (como o teste t) quando os dados não seguem uma distribuição normal?
9. Como os resultados desses estudos podem influenciar a escolha de métodos estatísticos em pesquisas na área de pavimentação e transportes?
10. Baseado nos artigos, qual seria sua recomendação para verificar a normalidade de dados em um estudo prático? Justifique sua resposta.

2 Exercício: Análise de Sistema de Filas de Atendimento

Você recebeu um conjunto de dados simulados de um sistema de filas de atendimento para um período de 4 horas (das 09:00 às 13:00). Os dados estão no arquivo *dados_fila_atendimento.csv* que contém informações sobre o horário de chegada dos clientes, o tempo entre chegadas sucessivas e o tempo de serviço para cada cliente.

Instruções

1. Carregue os dados do arquivo CSV no ambiente de análise de sua preferência.
2. Para as variáveis "tempo entre chegadas" e "tempo de serviço":
 - (a) Calcule as medidas de posição: média, mediana, moda, mínimo e máximo.
 - (b) Calcule as medidas de dispersão: amplitude, variância, desvio padrão e coeficiente de variação.
 - (c) Crie histogramas para visualizar a distribuição.
 - (d) Crie box plots para identificar outliers.
3. Analise a aderência a distribuições teóricas:
 - (a) Para o "tempo entre chegadas", teste a distribuição exponencial.

- (b) Para o "tempo de serviço", teste as distribuições normal, lognormal e gamma.
 - (c) Utilize testes estatísticos e gráficos (Q-Q plots) para avaliar a aderência.
4. Responda:
- (a) Qual a melhor distribuição para o tempo entre chegadas? Justifique.
 - (b) Qual a melhor distribuição para o tempo de serviço? Justifique.
 - (c) Existem outliers? Como tratá-los?
 - (d) Qual é a taxa média de chegada (λ) e de atendimento (μ)?
 - (e) O sistema está em estado estacionário? (Compare λ e μ .)
5. Crie um gráfico de dispersão dos tempos de chegada e de serviço. Há correlação?
6. Quais recomendações para melhorar a eficiência do sistema de atendimento?

Bônus

- 7. Implemente uma simulação do sistema de filas e compare com os dados originais.
- 8. Calcule o tempo médio no sistema e o número médio de clientes na fila.