

## **Lab05 - Estatística Inferencial I**

Aula para sanar dúvidas: dia 13/03/2012 das 18 às 20 horas no LCC2.

Limite de entrega sem penalização: 15/03/2012 às 23:59.

Liberação do gabarito: 19/03/2012 às 00:01.

### **Instruções gerais**

- Em todas as questões, o *script* R deve receber como parâmetro o nome da base de dados que o *script* deverá acessar (em caso de dúvidas pesquise por *commandArgs* no *help* do R). O *script* deve ser executado usando Rscript, por exemplo:

```
Rscript nome-lab05-questaoQ.R basededados-fornecida.txt
```

- O código R de todas as questões deve ser genérico, pois serão utilizadas outras bases de dados com o mesmo formato para testá-lo.

- Se você deseja utilizar no código da questão Y o código (ou parte de código) gerado na questão X, implemente isso por meio de funções e método e NÃO por compartilhamento de arquivo de resultado, por exemplo, não é permitido implementar em Y a leitura de um arquivo gerado por X.

- Todos os arquivos de saída gerados pelo *script* R (\*.png e \*.txt) devem ser salvos no diretório corrente, não coloque de modo *hardcode* um caminho para um diretório.

- Para cada questão deve ser gerado um relatório em formato \*.pdf com a interpretação e a análise dos resultados obtidos com a execução do *script* R. Todos os gráficos gerados devem ser apresentados e discutidos no relatório. O texto deve ser bem estruturado, os gráficos devem estar legíveis, adequadamente identificados e citados no texto. Use um corretor ortográfico!

- Neste lab usaremos uma base de dados de ociosidade de máquinas, que está disponível na página da disciplina, arquivo *atividade-maquinas2-dsc.txt*<sup>1</sup>. A base de dados está estruturada com as seguintes colunas:

---

<sup>1</sup> A base de dados não está completa. Foi utilizado um filtro para que a base de dados apresente cenários relevantes para fins didáticos.

- **intervalo** [*integer*]: tempo em segundos em que a máquina passou em um determinado estado (ociosa ou ocupada).
- **ociosa** [*logical*]: indica o estado da máquina correspondente ao intervalo de tempo indicado, sendo TRUE quando a máquina está ociosa ou FALSE caso contrário.
- **maquina** [*character*]: identificador da máquina.
- **laboratorio** [*character*]: identificador do laboratório ao qual a máquina pertence.

Deste modo, cada linha do arquivo é o *log* de um **intervalo** em que uma **máquina** em um **laboratório** permaneceu **ociosa** ou não.

## Exercício

1) Suponha que você deseja realizar uma análise gráfica para identificar se os tamanhos dos intervalos em que as máquinas permaneceram ociosas e os tamanhos dos intervalos em que elas permaneceram ocupadas seguem uma distribuição normal. Você deseja analisar isso com todos os intervalos juntos e separadamente para cada laboratório. Nesse sentido, siga as instruções abaixo:

- Escreva um *script* R que gera os gráficos FDA, FDP e as qqnorms dos seguintes dados: (i) intervalos em que as máquinas estiveram ocupadas; (ii) intervalos em que as máquinas estiveram ociosas; (iii) intervalos em que as máquinas estiveram ociosas separados por laboratório. Nos gráficos FDPs, plote duas linhas, uma com a densidade dos dados obtidos na base de dados e outra com a densidade de uma distribuição normal gerada com a média e o desvio padrão dos dados obtidos da base de dados. Para cada item solicitado deve ser gerada uma figura no formato \*.png, portanto, ao todo devem ser geradas 3 figuras.
- Gere os gráficos utilizando o seu *script*. Escreva um relatório apresentando os gráficos que você obteve e discutindo: (i) que relações você observa entre os gráficos FDA, FDP e qqnorm, apresente exemplos de alguns comportamentos observados no qqnorm e de como eles se manifestam na FDA e/ou na FDP; (ii) com base no qqnorm, na sua opinião, qual cenário avaliado mais se assemelha ao esperado de uma distribuição normal e qual menos se assemelha.

2) Suponha que após a análise gráfica da questão anterior, você decidiu verificar o  $Z_{skewness}$  e o  $Z_{kurtosis}$  dos dados para os mesmos cenários.

- Escreva um *script* R que calcula o  $Z_{skewness}$  e o  $Z_{kurtosis}$  dos seguintes dados: (i) intervalos em que as máquinas estiveram ocupadas; (ii) intervalos em que as máquinas estiveram ociosas; (iii) intervalos em que as máquinas estiveram ociosas separados por laboratório. Os resultados dos três itens devem ser salvos como uma tabela em um único arquivo no formato \*.txt. A tabela deve ter 3 colunas: cenário,  $Z_{skewness}$  e  $Z_{kurtosis}$ .

- Escreva um relatório apresentando os resultados obtidos com a execução do seu *script*. No relatório você deve responder as seguintes perguntas: (i) que interpretação você faz dos valores de  $Z_{skewness}$  e do  $Z_{kurtosis}$  obtidos? (iii) esses valores permitem tirar alguma conclusão sobre a normalidade dos dados? (iv) o que acontece se você variar o  $\alpha$ ?
- **Dica:** nesse exercício use a biblioteca *moments*. Em caso de dúvidas sobre como instalar um pacote, pesquise no *help ?install.packages* .

3) Use os testes estatísticos Anderson-Darling e Shapiro-Wilk para verificar a normalidade de um conjunto de dados. Para tanto, considere as descrições abaixo:

- Escreva um *script* R que aplica os testes Anderson-Darling e Shapiro-Wilk sobre os seguintes dados: (i) intervalos em que as máquinas estiveram ocupadas; (ii) intervalos em que as máquinas estiveram ociosas; (iii) intervalos em que as máquinas estiveram ociosas separados por laboratório. Você deve salvar os resultados em um único arquivo com uma tabela com quatro colunas: nome do teste, cenário, valor da estatística (A ou W) e o valor do *p-value*.
- Escreva um relatório apresentando os dados obtidos com a execução do seu *script*. No relatório você deve: (i) explicitar qual é a hipótese nula e qual a hipótese alternativa (ii) analisar as estatísticas e os *p-values* gerados pelos testes, explicando o que eles indicam; (iii) concluir se, considerando algum dos teste estatísticos aplicados, você falha ou não em rejeita a hipótese nula; (iv) explicar diferenças que você observa nos resultados gerados pelos dois testes.

4) Considerando os totais de tempo em que cada máquina permaneceu ociosa e os totais de tempo em que cada máquina permaneceu ocupada, implemente uma solução em R que permite utilizar os testes Anderson-Darling e Shapiro-Wilk para verificar se os tempos em que as máquinas permaneceram nesses estados seguem uma distribuição normal, antes e após uma transformação com *sqrt* e *log*. Escreva um relatório explicando sua solução, apresentando os resultados estatísticos que obteve, discutindo as conclusões que os resultados permitem tirar e comparando os resultados gerados pelos testes.