

**Disciplina:** Introdução à Estatística, Sistemas de Informação, EACH, USP, 2011.

**Professor:** Fernando Fagundes Ferreira.

**Monitor:** Eder Lucio da Fonseca.

## REGRAS DO TRABALHO FINAL

O trabalho é baseado nas monitorias sobre Regressão Linear e Teste de Kruskal-Wallis. Para resolver as questões, poderão ser consultadas as seguintes fontes: apresentação utilizada nas monitorias (disponível no COL USP), apostila disponibilizada no xérox e quaisquer outros materiais didáticos adicionais. Além disso, o trabalho é baseado em algumas regras simples descritas abaixo. A infração destas regras pode trazer prejuízos à nota final do trabalho.

- 1) O trabalho deve ser enviado para o e-mail [eder.luciofonseca@usp.br](mailto:eder.luciofonseca@usp.br), com o assunto "TRABALHO\_SI\_2011". O corpo do e-mail deve conter o nome completo dos integrantes do grupo e o número USP. Serão aceitos somente trabalhos em **pdf**.
- 2) O trabalho deve ser enviado impreterivelmente até a meia-noite do dia **30/06/11**. Trabalhos enviados após esta data/horário serão desconsiderados.
- 3) Serão atribuídas notas de 0 a 10 aos trabalhos. O peso será aplicado posteriormente pelo professor Fernando.
- 4) Cada grupo deverá ter no máximo **três** alunos. Trabalhos enviados com mais de três alunos serão desconsiderados imediatamente, sendo atribuída nota zero para todos os integrantes. Por outro lado, podem existir grupos com menos de três integrantes.
- 5) O conjunto de dados a ser utilizado para a resolução das questões está na planilha que segue em anexo.
- 6) **Quanto à cópia parcial de questões:** Os trabalhos que possuírem questões exatamente iguais terão as mesmas desconsideradas na correção.
- 7) **Quanto à cópia integral de trabalhos:** Dois trabalhos com todas as questões exatamente iguais terão suas notas divididas por três.
- 8) Os casos não descritos aqui serão julgados posteriormente, se necessário, pelo professor Fernando.

**QUESTÃO 01:** (2,5 pontos)

- a) Comente um pouco sobre o coeficiente de correlação ( $\rho$ ). Como ele pode ser interpretado? Existe algum caso em que ele não pode trazer informação sobre a relação entre as variáveis? Comente.
- b) Calcule o coeficiente de correlação amostral ( $r$ ) entre as variáveis  $X_1$  e  $Y_1$  pertinentes ao seu grupo.
- c) Teste se a correlação é igual a zero apresentando todos os passos necessários para a tomada de decisão (teste bilateral com  $\alpha=0,04$ ). O que aconteceria se utilizássemos  $\alpha=0,01$ ?

**QUESTÃO 02:** (2,5 pontos)

Para o conjunto de dados pertinente ao seu grupo:

- a) Construa um gráfico de dispersão para  $X_2$  e  $Y_2$ .
- b) Justifique a necessidade de utilizarmos uma regressão linear.
- c) Efetue um teste t unilateral para a nulidade do coeficiente de correlação, com  $\alpha = 0,05$ . Formule as hipóteses claramente.
- d) Encontre a equação da reta ajustada ao conjunto de dados.
- e) Interprete o coeficiente angular e o intercepto. O intercepto faz sentido?
- f) Construa um intervalo de confiança para o coeficiente angular com  $\alpha = 0,01$ . Qual é a sua conclusão a respeito do coeficiente angular? Interprete o intervalo de confiança.
- g) Escolha dois valores de  $X_2$ : um dentro do conjunto de dados e outro fora do conjunto de dados. Em seguida gere duas previsões com os valores escolhidos. Interprete os resultados.

**QUESTÃO 03:** (2,5 pontos)

Para o conjunto de dados pertinente ao seu grupo e utilizando a ferramenta Action:

- a) Construa um gráfico de dispersão para  $X_3$  e  $Y_3$ .
- b) Escreva a equação da reta ajustada.

- c) Interprete a estatística F obtida e o p-valor obtido.
- d) Qual a relação entre a estatística F, a Soma dos Quadrados da Regressão e Soma dos Quadrados dos Resíduos? *Dica: Pense em como foram calculados os Quadrados Médios dos Resíduos e Quadrados Médios da Regressão (utilizando os graus de liberdade).*
- e) O ajuste obtido foi bom? Justifique utilizando a estatística  $R^2$ .
- f) Teste as suposições adotadas ao ajustar uma regressão linear:
  - Os erros são normalmente distribuídos?
  - Os erros têm variância constante?
  - Os erros são independentes (não autocorrelacionados)? Utilize o teste de Durbin-Watson.

**QUESTÃO 04:** (2,5 pontos)

Um grupo de 36 pacientes portadores de anemia foi dividido em três grupos (Homens, Mulheres e Crianças), e todos foram submetidos ao mesmo tratamento. A doença é caracterizada por uma baixa taxa de hemoglobina, ou seja, valores inferiores a 13 g/dl para homens, inferiores a 12 g/dl para mulheres e inferiores a 11g/dl para crianças.

Para o conjunto de dados pertinente ao seu grupo, verifique, aos níveis de 5% e 1% de significância, se a melhoria do tratamento foi igual para os três grupos.