

Especificação
do projeto

2011.2

Sumário

Sobre o projeto	2
Observações Gerais.....	3
Scripts	4
Relatório.....	5
Formato	5
Introdução.....	5
Desenvolvimento	5
Conclusão	6
Apresentação	7

Sobre o projeto

O projeto consiste em uma análise estatística de um conjunto de dados com um tema específico. Cada grupo receberá um tema com um conjunto de dados e deverá, utilizando os conhecimentos adquiridos ao longo da disciplina, fazer um estudo das variáveis do problema. Alguns itens pedidos são: cálculo da média, mediana, definição do intervalo de confiança, etc. Para tal atividade, os alunos deverão utilizar a linguagem R, que aprenderam nas aulas no laboratório.

Basicamente, o projeto se baseia em: cálculo das medidas em R, relatório e apresentação. É importante que os grupos sigam as instruções fornecidas na especificação, aonde serão dados detalhes sobre cada etapa e o que será cobrado.

Observações Gerais

Organização dos grupos: grupos de no máximo cinco pessoas.

Nota e avaliação: a apresentação e o relatório cada um corresponde à metade da nota do projeto. O projeto vale 50% da terceira nota.

Entrega e apresentação: Todos os códigos utilizados em R deverão ser salvos e entregues no dia 24/11/2011 até as 23:59:59 horas, juntamente com relatório e a apresentação para o e-mail da monitoria (monitoriaestatistica@googlegroups.com). Todos esses arquivos devem estar comprimidos seguindo o padrão abaixo.

- Nome: [Tema].zip
- Conteúdo: Três pastas - Scripts, Apresentação e Relatório. Cada uma com seus respectivos conteúdos no formato final de entrega.

Projetos não entregues até o momento da apresentação receberão nota zero. O relatório também deverá ser entregue no dia da apresentação impresso e encadernado.

É importante que todos os grupos estejam presentes durante as apresentações. A ausência não justificada de algum integrante poderá implicar na perda de pontuação sobre a nota final do projeto do integrante em questão.

Scripts

Todas as tarefas envolvendo cálculos e processamento de dados deverão ser feitos utilizando a linguagem R – incluindo geração de gráficos e leitura de tabelas. Não haverá nenhuma restrição quanto à utilização da linguagem – qualquer função pode ser utilizada.

Os scripts feitos em R deverão ser enviados juntamente com o relatório e apresentação. Os grupos poderão ficar à vontade para organizar os scripts e as funções como quiserem, desde que fiquem legíveis para análise e correção posterior.

Caso haja alguma dificuldade em relação ao processamento dos dados – em especial, quanto à leitura das tabelas – um e-mail deverá ser enviado para a monitoria, para que nós possamos analisar o caso e sugerir alguma solução.

Relatório

Formato

O relatório deverá ser composto por:

- Capa
- Identificação do grupo
- Índice
- Introdução
- Desenvolvimento
- Conclusão
- Referências bibliográficas

O relatório pode ser redigido utilizando o estilo padrão do Word, com as fontes Calibri, Times New Roman ou similares, usando tamanhos 11 ou 12. O relatório deve ser entregue encadernado e deve ter em média de 10 a 15 páginas, podendo ter um pouco mais do que isto.

Em caso de omissão em algum item requisitado, pontos serão descontados, então é preferível ultrapassar 15 páginas a não adicionar tudo o que foi pedido.

Introdução

Será uma breve descrição do trabalho desenvolvido pelo grupo. Além disso, deverá conter os objetivos do projeto, além de uma breve apresentação do tema. O tamanho da amostra utilizada deve ser especificado.

Desenvolvimento

O desenvolvimento é composto pelos itens que serão explicados abaixo:

Descrição das variáveis: serão escolhidas três variáveis pelo grupo, duas devem ser contínuas e a outra, discreta. O grupo deve justificar a escolha de tais variáveis, além de apresentar para cada uma delas:

- As medidas estatísticas estudadas ao longo do período (média, mediana, moda, variância, desvio padrão, coeficiente de assimetria, coeficiente de curtose, primeiro quartil, terceiro quartil).
 - Deverá ser feita uma análise estatística de cada variável utilizando-se de tais medidas.
- Elaborar histogramas, polígono de frequência (deve estar presente no histograma) e tabelas de frequência para cada variável.
 - Para cada variável devem ser gerados dois histogramas: um utilizando um método de separação de classes dado em sala e outro utilizando um método que não foi dado em sala. Os alunos devem estudar e explicar este método, assim como comparar os histogramas e explicar as diferenças entre eles.

- A distribuição de cada variável deverá ser comentada (as medidas calculadas anteriormente deverão ser utilizadas).

Intervalos de confiança: deverão ser construídos intervalos de confiança para cada variável, uma mesma variável terá dois intervalos de confiança com graus de significância diferentes. É importante que seja exposto no relatório uma comparação entre esses intervalos.

- Observação: Os gráficos de intervalos de confiança deverão estar no relatório.

Teste de Hipótese: o relatório deverá conter testes de hipótese de uma amostra e duas amostras. Deverá ser explicado o que significa o *p-value* e através dele e do intervalo de confiança descrever o resultado do teste de hipótese. O grupo deverá escolher as variáveis para fazer os testes de hipótese e justificar o porquê de tê-las utilizado. Além disso, é importante que seja explicado como foi feita a amostragem e qual foi o critério ao se escolher o tamanho da amostra.

- **Observação:** É muito importante que os testes de hipótese sejam contextualizados e que seja apresentado um gráfico ilustrando o teste. Projetos com testes de hipótese bem contextualizados, por exemplo, que tenham um acontecimento histórico associado, receberão bonificações.

Conclusão

Deverá conter os resultados do estudo estatístico a respeito do tema abordado.

Apresentação

A apresentação deverá ser realizada em 15 minutos. Elas ocorrerão nos dias 25/11/2011 e 29/11/2011. É importante que todos os grupos estejam presentes em ambas as apresentações.

O conteúdo apresentado deverá ser:

- Uma descrição do tema abordado
- A descrição das variáveis escolhidas pelo grupo
- Gráficos, tabelas e medidas relativas às variáveis devem ser explicadas (mais uma vez é importante relacionar a distribuição de cada variável com suas respectivas medidas).
- Aplicações de estimação intervalar e testes de hipótese sobre a amostra
- Conclusão do trabalho.
- Referências bibliográficas.

Observação: É de extrema importância que todos os dados contidos na apresentação sejam analisados estatisticamente, ou seja, apenas citar os dados não é suficiente, é necessário explicar o que cada um representa no âmbito do tema.