

# Tarefa 2 Aula de Laboratório 2

24/10/2021

**SME 0820** “Modelos de Regressão e Aprendizado Supervisionado I”

Universidade de São Paulo ICMC

**Estagiário:** Daniel Camilo Fuentes Guzmán

**Supervisora:** Profa. Dra. Reiko Aoki

Baseado na aula 2 de laboratório, resolva o problema dado e escreva um relatório apresentando a análise feita.

## Instruções:

- A entrega será feita de um relatório por grupo até o dia 02/11/2021. (Trabalhos enviados após esse horário serão penalizados por hora de atraso e após 10h do horário do limite de entrega eletrônica serão desconsiderados. Casos excepcionais devem ser discutidos com a docente antes do limite de entrega eletrônica);
- Especificar Nome completo e número USP dos autores;
- No título do trabalho deve ser especificado o número do grupo;
- As ilustrações(Gráficos, tabelas, etc.) devem ser identificadas (e numeradas e com legenda) e citadas ao longo do texto;
- Envio do código utilizado completo especificando a linguagem utilizada.
- O trabalho deverá ser entregue no moodle em formato PDF, com número do trabalho, do grupo e código da disciplina especificado (**Exemplo:** Tarefa 2 SME0820 - Grupo X).
- O nível de significância para o teste de hipóteses e para os intervalos de confiança de cada grupo será o número do grupo, ou seja, o grupo 1 utiliza  $\alpha=1\%$ . O conjunto de dados de cada grupo será disponibilizado pelo estagiário via e-mail para todos os integrantes do grupo.
- Mesmo que o resultado do código esteja certo e o PDF esteja escrito errado não significa que a questão será aceita como certa. Se for, terá uma penalização.

## Tarefa:

Baseado na aula 2 de laboratório, resolva o problema dado e escreva um relatório apresentando a análise feita.

Com o conjunto de dados disponibilizados pelo estagiário para o seu grupo. Utilizando R ou Python:

## 1 Documento de Apresentação

- Apresente um relatório em pdf da solução da tarefa 2 correspondente à aula de laboratório 2 com estrutura de trabalho, tendo uma capa, um índice e conclusões e resumo na parte final do código utilizado bem organizado e detalhado já seja do R ou da linguagem Python, levando em consideração as instruções descritas no início deste documento.

## 2 Análise Descritiva dos Dados

- Faça uma análise descritiva completa dos seus dados e intérprete, faça gráficos de dispersão, box plots e histogramas de cada uma das variáveis de seu conjunto de dados e interprete.

### 3 Ajuste um Modelo de Regressão Linear Múltipla

- Ajuste um modelo de regressão linear múltipla considerando todas as variáveis do seu conjunto de dados. Apresente os valores numéricos de da matriz  $X^T X$  e sua inversa. Assim como da matriz  $X^T Y$  e estimativas do vetor dos  $\beta$ 's e intérprete. Explique se existe diferença entre o **EMV** e o estimador de mínimos quadrados de  $\beta$ ? justifique.

### 4 Estimação de $\sigma^2$

- Calcule o  $SQRes$  e estime  $\sigma^2$ . Descreva e justifique as características deste estimador, é viciado?, é consistente?. Calcule o valor dos  $\hat{Y}$ 's e o valor dos resíduos para o seu modelo.

### 5 ANOVA

- Faça o teste de significância da regressão utilizando a ANOVA. Determine se existe relação linear entre a variável resposta  $Y$  e qualquer uma das covariáveis. Para a ANOVA ilustre o passo a passo da construção ou cálculo da mesma (Calcule os valores de:  $SQTotal$ ,  $SQreg$ ,  $SQres$ ,  $QMReg$ ,  $QMres$ , g.l. da regressão, g.l. do resíduo, F.). Apresente a tabela ANOVA e sua respectiva análise.

### 6 $R^2$ e $R^2_{ajustado}$

- Quanto da variabilidade total esta sendo explicada pela regressão? Explique acerca da diferença entre o  $R^2$  e  $R^2_{ajustado}$ . Qual seria mais adequado de usar no contexto de adicionar covariáveis a seu modelo? Justifique. Faça comparativo de  $R^2$  e  $R^2_{ajustado}$  para pelo menos 3 cenários e descreva o que acontece com ditas medidas. (uma possibilidade pode ser ir tirando de uma em uma ou um conjunto de covariáveis de seu modelo e verificar o  $R^2$  e  $R^2_{ajustado}$ ). Faça um quadro ilustrando os resultados.

### 7 Teste nos Coeficientes de Regressão Individualmente

- No contexto de seleção de Covariáveis e Modelo com Covariáveis Significativas: Ao nível de significância  $\alpha$  definido para seu grupo, o modelo deveria conter quais co-variáveis?. Apresente um segundo modelo para as covariáveis significativas com um nível de  $\alpha$  de significância. O que foi observado de diferente? (Neste ponto obtenha os teste parcial ou marginal nos coeficientes de regressão individualmente que faz uso da estatística t ilustrado na aula 13 - testes de hipóteses de significância dos betas individualmente).

### 8 Teste um subconjunto de coeficientes

- Faça análise e Escolha do Melhor Modelo, Qual modelo seu grupo acredita ser o melhor? Apresente o modelo escolhido pelo grupo (Equação do modelo ajustado, com as estimativas dos  $\beta$ 's e  $\sigma^2$  e interpretação do modelo). Justifique a escolha das co-variáveis. Faça uso do teste de um subconjunto de coeficientes nesta questão ilustrado na aula 14 e 15.

### 9 IC Para Coeficientes de Regressão

Faça o intervalo de confiança para os coeficientes de regressão de seu modelo. Apresente conclusões ou interpretação para cada um deles.

# 10 IC para $E(Y)$

Faça o intervalo de confiança para  $E(Y)$  no esquema especificado a seu grupo pelo monitor. Apresente conclusões ou interpretação para cada um deles.

## **Cr terios de Avalia  o**

- Execu  o dos itens do problema enunciado.
- Organiza  o e coer ncia do texto.
- Organiza  o do c digo.
- Interpreta  o dos resultados, (an lise descritiva, ajuste do modelo, teste de hip tese, intervalo de confian a).
- Pontualidade na entrega.
- A nota de cada grupo poder  variar de 0 a 10.
- Dentro dos grupos que entregarem o trabalho at  o hor rio limite, o que conseguir maior pontua  o ter  um pr mio de um ponto para a pr xima tarefa de laborat rio. (Quer dizer que trabalhos atrasados s  ter o direito a penaliza  o.)