



**Universidade Estadual de Campinas
Instituto de Matemática, Estatística e Computação Científica
Departamento de Estatística**

Relatório Trabalho de ME610

**Eliane Ramos de Siqueira RA:155233
Guilherme Pazian RA:160323**

Professor: Maurício Zevallos

Campinas-SP, 28 de Abril de 2017

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

1. Introdução

O objetivo deste trabalho é verificar esta opinião através da análise do peso da criança ao nascer

Os dados utilizados foram coletados através de entrevistas com as mães durante a gravidez e correspondem a um ano de estudo. Eles incluem 1236 nascimentos onde a criança sobreviveu pelo menos 28 dias. Além do peso das crianças, foram coletadas informações de outras variáveis que influenciam no peso das crianças.

Para este estudo estamos considerando como variável resposta, o peso da criança ao nascer.

Limpendo os dados

Excluímos as variáveis: Sexo (apenas masculino), Qtd_feto (apenas 5) e Vivo (apenas 1).

Na variável Cor_mae, temos uma observação com valor desconhecido (99)

Na variável Idade_mae temos duas observações desconhecidas (99)

Na variável Educacao_mae temos uma observação desconhecida (9)

Na variável Altura_mae temos 22 observações desconhecidas (99)

Na variável Peso_mae temos 36 observações desconhecidas (999)

Na variável Cor_pai, temos 5 observações com valor desconhecido (99)

Na variável Idade_pai, temos 7 observações com valor desconhecido (99)

Na variável Educacao_pai, temos 13 observações com valor desconhecido (9)

Na variável Altura_pai temos 492 observações desconhecidas (99)

Na variável Peso_pai temos 499 observações desconhecidas (999)

Na variável Rendimento_anual temos 124 observações desconhecidas (98)

Na variável Fuma temos 10 observações desconhecidas (9)

Na variável Tempo_sem_fumar temos 9 observações desconhecidas (98) e 1 não perguntado(99)

Na variável Numero_cigarros temos 10 observações desconhecidas (98)

Tiramos as variáveis Peso_pai e Altura_pai devido a quantidade elevada de observações desconhecidas.

Variáveis disponíveis ainda:

Data_nasc
Tempo_gestacao
Peso
Cor_mae
Educacao_mae
Peso_mae
Peso_pai
Altura_pai
Estado_civil 2 desconhecidas(0)
Fuma
Numero_cigarros Numero_gestacoes Idade_mae Idade_pai Altura_mae Cor_pai Educacao_pai
Rendimento_anual Tempo_sem_fumar

Convertendo unidade de medidas

Frequencias

```
#verificando a frequencia  
freq_fuma=table(dados$Fuma)  
names(freq_fuma)<-c("Nunca", "Fuma agora", "Até a gravidez atual", "Uma vez, não agora")  
freq_fuma=as.data.frame(freq_fuma)  
names(freq_fuma)<-c("Variável Fuma", "Frequência")  
kable(freq_fuma)
```

Variável Fuma	Frequência
Nunca	469
Fuma agora	404
Até a gravidez atual	79
Uma vez, não agora	94

```
#library(xtable)  
#print(xtable(freq_fuma)) #tabela em latex
```

Análise Descritiva

Boxplot para as variáveis categóricas e Dispersão para continua





