

**FUNDAÇÃO GETULIO VARGAS - FGV**

**TRABALHO II – Casos Práticos**

Ana Paula Pudo  
Davi Khoury  
Fábio Monteiro  
Lucas Sena Alves  
Marcos Soares

**MBA Business Analytics e Big Data | Inferência  
Estatística**

Prof. João Rafael Dias Pinto

Brasília  
2021

## SUMÁRIO

Relação de Ilustrações.....	3
Tabelas .....	3
Imagens .....	3
Gráficos .....	3
1. Introdução.....	4
2. Metodologia .....	4
3. Primeiro Case: Fator de Risco para parto de bebês abaixo do peso .....	6
3.1 Visão Geral das informações - Análise Univariada.....	6
3.2 Diagrama sobre a relação da cor da pele das mães e o nascimento de bebês abaixo do peso .....	8
3.3 Diagrama sobre o consumo de tabaco das mães em relação ao nascimento de bebês abaixo do peso .....	9
3.4 Diagrama sobre o histórico de partos prematuros das mães em relação ao nascimento de bebês abaixo do peso.....	9
3.5 Diagrama sobre o histórico de hipertensão das mães em relação ao nascimento de bebês abaixo do peso .....	10
3.6 Diagrama sobre o histórico de Irritabilidade Uterina das mães em relação ao nascimento de bebês abaixo do peso.....	11
4. Segundo case: IMC e gênero em ataques cardíacos .....	12
4.1 Diagrama sobre a relação do Índice da Massa Corporal e o Gênero .....	13
5. Terceiro case: Interferência do álcool na habilidade de direção .....	13
5.1 Diagrama sobre a relação sobre o consumo de álcool e o tempo de reação de um motorista .....	13
6. Quarto case: Ganho de peso em ratos.....	14
6.1 Diagrama sobre a quantidade e tipo de alimento e o ganho de peso em ratos .....	15
7. Quinto case: Gordura corporal em pacientes do sexo masculino .....	16
7.1 Regressão linear simples .....	16
7.1.1 <i>Ganho de Peso x Idade</i> .....	16
7.1.2 <i>Gordura Corporal x Peso</i> .....	17
7.1.3 <i>Gordura Corporal x Altura</i> .....	17
7.1.4 <i>Gordura Corporal x Abdômen</i> .....	18
7.1.5 <i>Gordura Corporal x Quadril</i> .....	18
7.1.6 <i>Gordura Corporal x Pescoço</i> .....	19
7.1.7 <i>Gordura Corporal x Coxa</i> .....	20
7.1.8 <i>Gordura Corporal x Punho</i> .....	20
7.1.9 <i>Gordura Corporal x Bíceps</i> .....	21
7.1.10 <i>Gordura Corporal x Antebraço</i> .....	21

7.2 Regressão Múltipla .....	22
7.3 Teste de Multicolinearidade.....	22
7.4 Conclusões .....	22

## **Relação de Ilustrações**

### **Tabelas**

Tabela 1: Técnicas estatísticas adotadas nas análises dos Estudos de Caso

### **Imagens**

Imagem 1: Sumário da Regressão Linear Múltipla

### **Gráficos**

Gráfico 1: Distribuição de nascidos abaixo do peso  
 Gráfico 2: Distribuição de mães - Cor da pele  
 Gráfico 3: Distribuição de Mães - Consumo de Tabaco  
 Gráfico 4: Distribuição de Histórico de Parto Prematuro  
 Gráfico 5: Distribuição de Histórico de Hipertensão  
 Gráfico 6: Mães com histórico de Irritabilidade Uterina  
 Gráfico 7: Idade das mães  
 Gráfico 8: Análise bivariada: Nascidos abaixo do peso x Cor da pele da mãe  
 Gráfico 9: Análise bivariada: Nascidos abaixo do peso x Consumo de Tabaco  
 Gráfico 10: Análise bivariada: Histórico de Nascimento abaixo do Peso x Parto Prematuro  
 Gráfico 11: Análise bivariada: Histórico de Nascimento abaixo do Peso X Histórico de Hipertensão  
 Gráfico 12: Análise bivariada: Histórico de Nascimento abaixo do peso e Irritabilidade Uterina  
 Gráfico 13: Distribuição por gênero  
 Gráfico 14: Análise bivariada IMC X Gênero  
 Gráfico 15: Análise bivariada Consumo Alcool x Tempo de Reação  
 Gráfico 16: Análise bivariada: Ganho de peso x Quantidade de alimentos  
 Gráfico 17: Análise bivariada Tipo de alimento x Ganho de peso  
 Gráfico 18: Regressão linear simples: Gordura corporal x Idade  
 Gráfico 19: Regressão linear simples: Gordura Corporal x Peso  
 Gráfico 20: Regressão linear simples: Gordura corporal x Altura  
 Gráfico 21: Regressão Linear Simples: Gordura Corporal x Abdômen  
 Gráfico 22: Regressão Linear Simples: Gordura Corporal x Quadril  
 Gráfico 23: Regressão Linear Simples: Gordura Corporal x Pescoço  
 Gráfico 24: Regressão Linear Simples: Gordura Corporal x Coxa  
 Gráfico 25: Regressão Linear Simples: Gordura Corporal x Punho

Gráfico 26: Regressão Linear Simples: Gordura Corporal x Bíceps

Gráfico 27: Regressão Linear Simples: Gordura Corporal x Antebraço

## 1. Introdução

Este é um estudo realizado com o objetivo de analisar os dados e trazer conclusões para hipóteses em cinco casos práticos: análise dos riscos de partos com bebês abaixo do peso; ataques cardíacos e a relação entre Índice de Massa Corporal e Gênero nos casos; relação entre a ingestão de cervejas e o impacto na habilidade de direção de veículos; ganho de peso e diferentes dietas em ratos; e circunferência a possíveis relações lineares de gordura corporal.

O estudo buscará formular e testar hipóteses, com o objetivo de gerar conclusões estatisticamente relevantes, necessárias para a elaboração de insights.

Entre os principais objetivos desta análise estão:

- Realização de análise exploratória para entendimento da natureza dos dados e seus comportamentos.
- Utilização de técnicas estatísticas na formulação e resolução de hipóteses para subsidiar conclusões obtidas acerca de estudos baseados nos dados.
- Utilização de técnicas estatísticas ao estabelecer relações lineares (simples e múltiplas), testes de comparação de médias e de proporção entre as variáveis dos dataframes.
- Plotagem de gráficos e tabelas para melhor visualização dos produtos extraídos nos estudos realizados

## 2. Metodologia

Este estudo consiste na análise realizada em 5 conjuntos de dados distintos, segmentados em casos analisados individualmente (low\_birth\_weight.csv; heart\_attack.csv; beers.csv; ratfeed.txt e bodyfat\_men.csv). A metodologia aplicada a todos os conjuntos foram similares, levando-se em conta principalmente a classificação quanto à natureza quantitativa ou qualitativa dos dados contidos nas amostras, fundamental para a melhor aplicação das técnicas estatísticas mais adequadas na formulação das conclusões descritas nos cenários de análises.

Foram realizados procedimentos de análise exploratórias para melhor entendimento dos dados e a relação entre eles. Os devidos tratamentos e rearranjos adotados foram feitos a partir de técnicas aplicadas no melhor refinamento das análises e na remoção de elementos que demonstram possíveis erros ou inconsistências, mitigando a ocorrência de ruídos aos resultados finais. Entre eles:

- Remoção de outliers e missing values (quando constatado que esses não faziam parte e/ou impactassem o universo amostral);
- Rearranjo dos conjuntos analisados: técnicas aplicadas para melhor aplicar as ferramentas de estatísticas identificadas como adequadas às análises mapeadas;

- Transformações das unidades de medida para padronização nas análises feitas.

Os resultados obtidos a partir das investigações e entendimento dos dataframes estudados possibilitaram a aplicação de técnicas de inferência estatística, na formulação dos testes de hipótese em cenários em que os conjuntos de dados fossem capazes de responder. Foi adotado o nível de significância estatística de 5% e os testes de hipóteses.

Tivemos como objeto principal dos estudos as variáveis target definidas para cada um dos conjuntos de dados através de associações as variáveis explicativas. Os datasets e suas targets foram descritas nas análises subsequentes como:

- **Fator de risco para baixo peso ao nascer (low\_birth\_weight.csv):**
  - Target: *Crianças que nasceram ou não abaixo do peso;*
- **IMC vs Gênero em pacientes com ataque cardíaco (heart\_attack.csv):**
  - Target: *IMC de mulheres;*
- **Ingestão de álcool vs Direção (beers.csv):**
  - Target: *Média de tempo de reação;*
- **Ganho de peso em ratos (ratfeed.txt):**
  - Target: *Média de ganho de peso;*
- **Gordura corporal em pacientes do sexo masculino (bodyfat\_men.csv):**
  - Target: *Gordura corporal.*

<i>Técnicas Estatísticas adotadas nas análises dos Casos de Estudos</i>	Análise Tópico 3	Análise Tópico 4	Análise Tópico 5	Análise Tópico 6	Análise Tópico 7
	Crianças do abaixo peso	IMC vs Gênero	Álcool vs Direção	Ganho de peso (ratos)	Gordura corporal
<i>Conversão das unidades de medidas para padronização do projeto;</i>	OK	-	-	-	OK
<i>Remoção de colunas identificadas como variáveis não relevantes ao trabalho;</i>	OK	-	-	-	-
<i>Remoção de registros identificados na amostra como missing values e outliers;</i>	OK	-	-	-	OK
<i>Criação de colunas a partir do rearranjo de dados para adoção de técnicas estatísticas;</i>	-	OK	OK	-	-
<i>Transformação dos tipos de variáveis e padronização dos domínios (labels) utilizados no dataframe, para possibilitar adoção de técnicas estatísticas além de melhor entendimento;</i>	OK	-	OK	OK	OK
<i>Plotagem de gráficos e/ou tabelas;</i>	OK	OK	OK	OK	OK
<i>Teste Chi-quadrado;</i>	OK	-	-	-	-
<i>Teste de Variância;</i>	-	OK	-	OK	-

<i>Teste T;</i>	-	OK	OK	OK	-
<i>Teste ANOVA;</i>	-	-	-	OK	-
<i>Regressão Linear Simples;</i>	-	-	-	-	OK
<i>Regressão Linear Múltipla.</i>	-	-	-	-	OK

*Tabela: Técnicas Estatísticas adotadas nas análises dos Casos de Estudos nas investigações realizadas.*

Todos os dados foram trabalhados no RStudio, onde também realizamos a representação de histograma, scatterplot, boxplot, análise bivariada e univariada ao postarmos os gráficos e tabelas. Para calcular todas as variáveis e realizar as análises necessárias, utilizamos códigos de programação da linguagem R.

### 3. Primeiro Case: Fator de Risco para parto de bebês abaixo do peso

O nascimento de bebês com baixo peso é preocupante, uma vez que este é um dos indicadores mais relacionados às taxas de mortalidade infantil e deficiências. Existem vários padrões comportamentais da mãe durante a gestação que podem influenciar diretamente no peso da criança. Para verificar quais fatores podem exercer influência no peso do bebê, foi realizado um levantamento com 189 mulheres em um hospital nos EUA. Os dados foram obtidos em 1986.

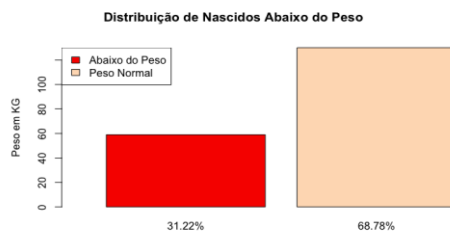
A partir destes dados, analisamos quais são os fatores de risco associados ao parto de um bebê com peso aquém do esperado.

#### 3.1 Visão Geral das informações - Análise Univariada

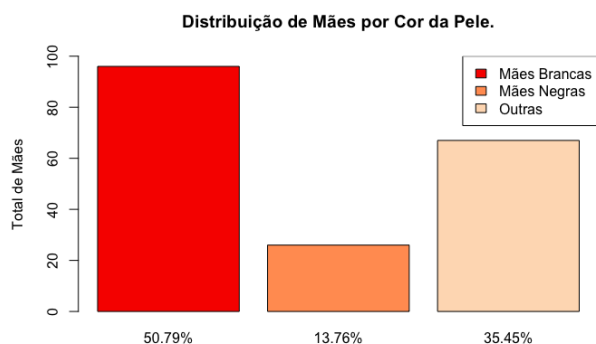
Foram analisadas individualmente cada variável categórica, para que fosse possível traçar um panorama geral.

Foi percebido nesta amostra que a idade das mães varia entre 15 a 45 anos, mas a maior concentração de ocorrências está nas mães mais jovens entre 15 a 25 anos, e majoritariamente mães brancas e não fumantes. A maioria delas não apresenta histórico de parto prematuro, hipertensão e irritabilidade uterina.

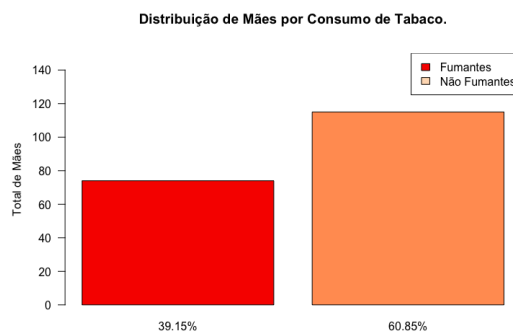
O percentual de nascimento de bebês abaixo do peso é de 31,22% .



**Gráfico 1: Distribuição de nascidos abaixo do peso**

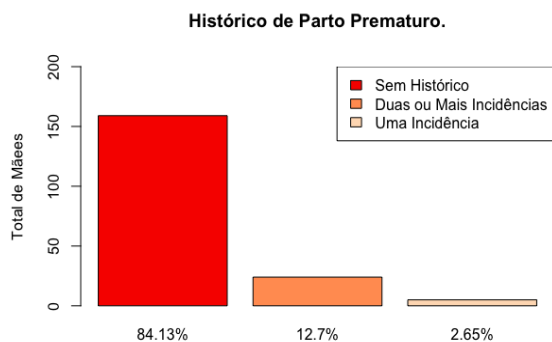


**Gráfico 2: Distribuição de mães - Cor da pele**

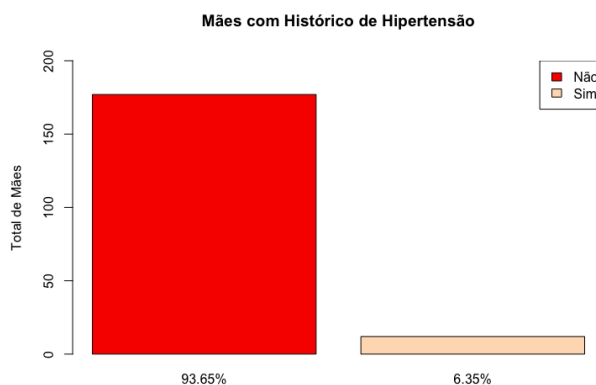


**Gráfico 3: Distribuição de Mães - Consumo de**

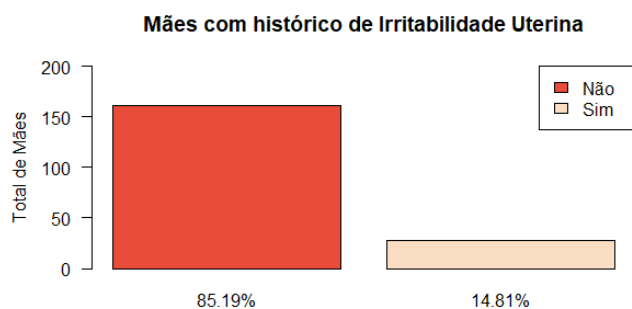
**Tabaco**



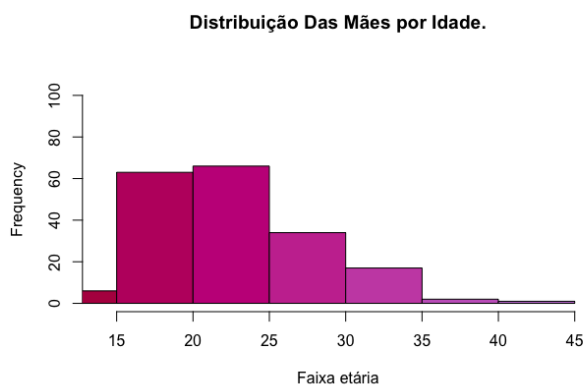
**Gráfico 4: Distribuição de Histórico de Parto Prematuro**



**Gráfico 5: Distribuição de Histórico de Hipertensão**



**Gráfico 6: Mães com histórico de Irritabilidade Uterina**



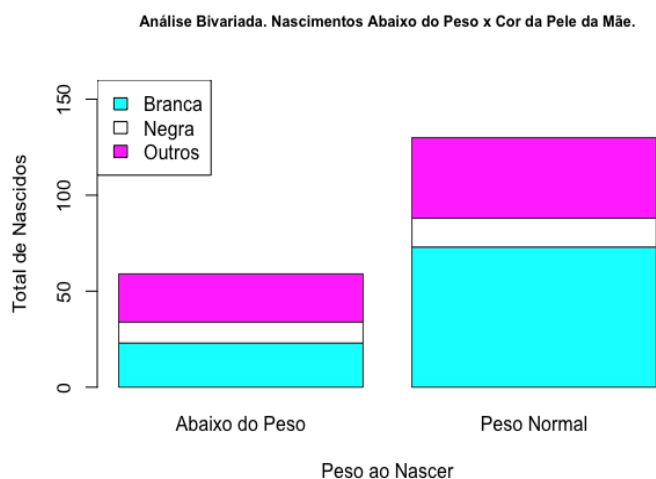
**Gráfico 7: Idade das mães**

### 3.2 Diagrama sobre a relação da cor da pele das mães e o nascimento de bebês abaixo do peso

Na análise bivariada entre a cor da pele e a variável alvo, foram formuladas as hipóteses nula ( $H_0$ ) e alternativa ( $H_a$ ):

$H_0$  - A cor da pele das mães não tem relação com o baixo peso ao nascer.

$H_a$  - A cor da pele das mães tem relação com o baixo peso ao nascer.



**Gráfico 8: Análise bivariada: Nascidos abaixo do peso x Cor da pele da mãe**

Chisq.test / P-valor = 0,081



**CONCLUSÃO:** Como o p-valor = 0.081 % é maior que 5%, não foi possível rejeitar a H0. Portanto, é possível concluir que não existe dependência entre a cor da pele das mães e o baixo peso da criança ao nascer, ao nível de significância de 5%.

### 3.3 Diagrama sobre o consumo de tabaco das mães em relação ao nascimento de bebês abaixo do peso

Já considerando o consumo de tabaco e o peso, foram formuladas as seguintes hipóteses:

H0 - Não existe dependência entre o consumo de tabaco e o nascimento de bebês abaixo do peso.

Há - Existe dependência entre o consumo de tabaco e o nascimento de bebês abaixo do peso.

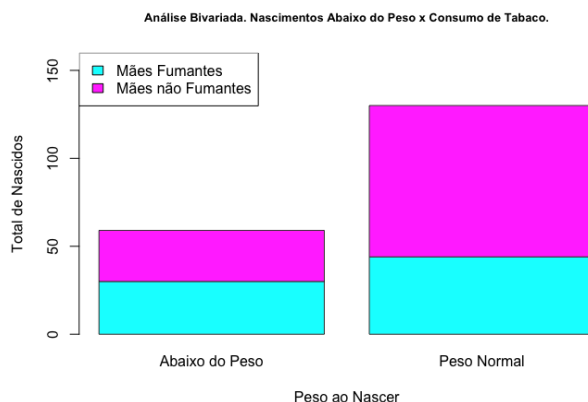


Gráfico 9: Análise bivariada: Nascidos abaixo do peso x Consumo de Tabaco

Chisq.test / P-valor = 0,0395

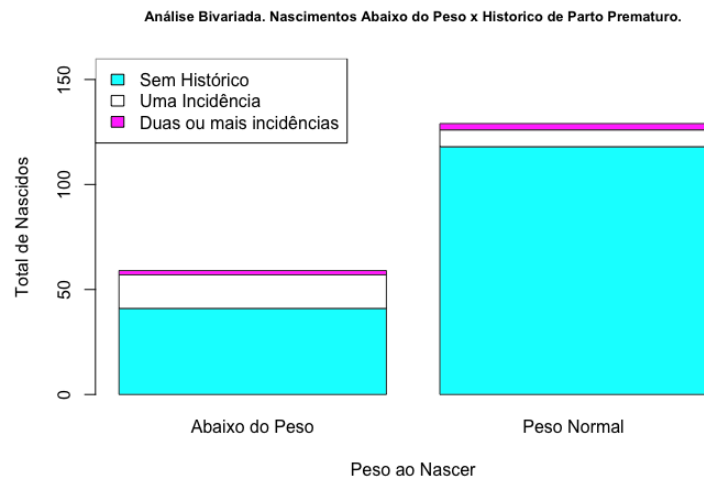
**CONCLUSÃO:** Como o p-valor = 0,0395 é menor que 5%, rejeitamos a H0. Portanto, concluímos que o consumo de tabaco das mães tem relação ao nascimento de bebês abaixo do peso, ao nível de significância de 5%.

### 3.4 Diagrama sobre o histórico de partos prematuros das mães em relação ao nascimento de bebês abaixo do peso

Na análise bivariada entre partos prematuros e o peso, as hipóteses formuladas foram:

H0: Não existe dependência entre mães que tenham histórico de partos prematuros e o nascimento de bebês abaixo do peso.

Ha: Existe dependência entre mães que tenham histórico de partos prematuros e o nascimento de bebês abaixo do peso.



**Gráfico 10: Análise bivariada. Histórico de Nascimentos abaixo do Peso x Parto Prematuro**

Chisq.test / P-valor = 0,00049

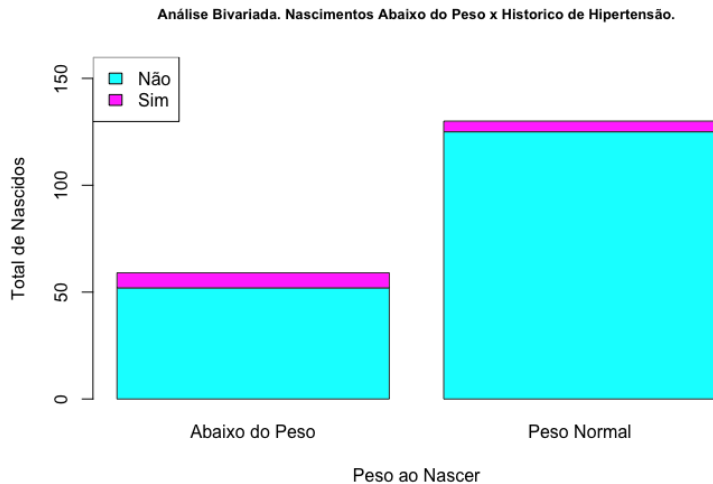
**CONCLUSÃO:** Como o p-valor = 0,00049 é menor que 5% rejeitamos a H0, portanto, concluímos que existe dependência entre mães que tem um histórico de partos prematuros em relação ao nascimento de bebês abaixo do peso, ao nível de significância de 5%.

### 3.5 Diagrama sobre o histórico de hipertensão das mães em relação ao nascimento de bebês abaixo do peso

Na análise bivariada que considerou o histórico de hipertensão e o peso, as hipóteses formuladas foram:

H0: Não existe dependência entre a existência de histórico de hipertensão e o nascimento de bebês abaixo do peso.

Ha: Existe dependência entre ter histórico de hipertensão e o nascimento de bebês abaixo do peso.



**Gráfico 11: Análise bivariada: Histórico de Nascimentos abaixo do Peso X Histórico de Hipertensão**

Chisq.test / P-valor = 0,076

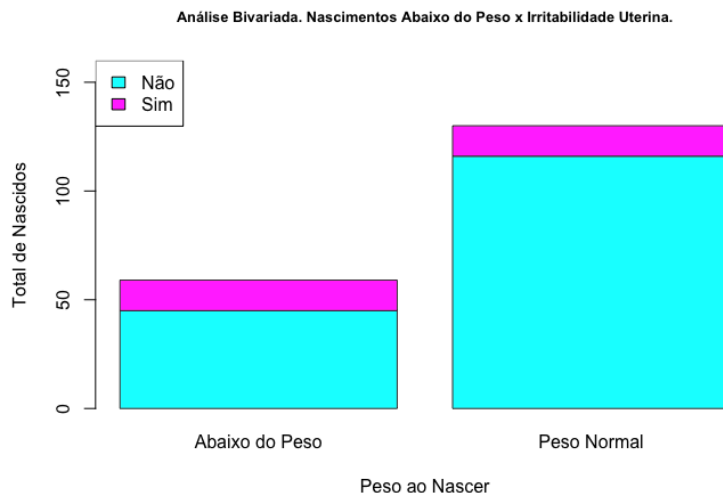
**CONCLUSÃO:** Como o p-valor = 0,076 é maior que 5% não podemos rejeitar a H0. Ou seja, é possível concluir que não existe dependência entre mães que têm um histórico de hipertensão e o nascimento de bebês abaixo do peso, ao nível de significância de 5%.

### 3.6 Diagrama sobre o histórico de Irritabilidade Uterina das mães em relação ao nascimento de bebês abaixo do peso

Na análise bivariada que considerou o histórico de irritabilidade uterina e o peso dos bebês, as hipóteses formuladas foram:

H0: Não existe dependência entre o fato de mães terem irritabilidade uterina e o nascimento de bebês abaixo do peso.

Ha: Existe dependência entre o fato de mães terem irritabilidade uterina e o nascimento de bebês abaixo do peso.



**Gráfico 12: Análise bivariada: Histórico de Nascimentos abaixo do peso e Irritabilidade Uterina**

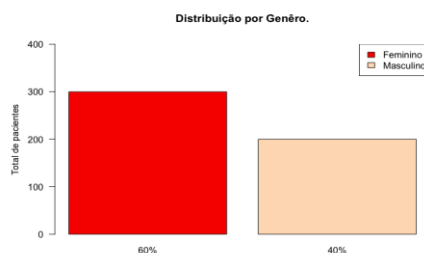
Chisq.test / P-valor = 0,03

**CONCLUSÃO:** Como o p-valor = 0,03 é maior que 5%, não podemos rejeitar a  $H_0$ , e concluímos que não existe dependência entre mães que tem um histórico de irritabilidade uterina em relação ao nascimento de bebês abaixo do peso, ao nível de significância de 5%.

#### 4. Segundo case: IMC e gênero em ataques cardíacos

Um estudo foi realizado com pacientes que haviam sofrido ataque cardíaco. Eles são habitantes de uma certa região metropolitana. A partir dos dados levantados, a análise a seguir vai buscar hipóteses sobre para a relação entre o Índice de Massa Corporal e o Gênero entre tais pacientes. Os indivíduos que participaram deste estudo se apresentaram ao hospital e foram selecionados aleatoriamente para participar da análise.

A distribuição por gênero dos participantes do estudo foi a seguinte:



**Gráfico 13: Distribuição por gênero**

#### 4.1 Diagrama sobre a relação do Índice da Massa Corporal e o Gênero

Considerando a análise bivariada entre o Índice de Massa Corporal e o gênero dos pacientes, foram estabelecidas as seguintes hipóteses:

H<sub>0</sub> - A média do IMC de mulheres que sofreram infarto é igual ou inferior à média de IMC de homens na mesma condição.

H<sub>a</sub> - A média do IMC de mulheres que sofreram infarto é superior à média de IMC de homens na mesma condição.

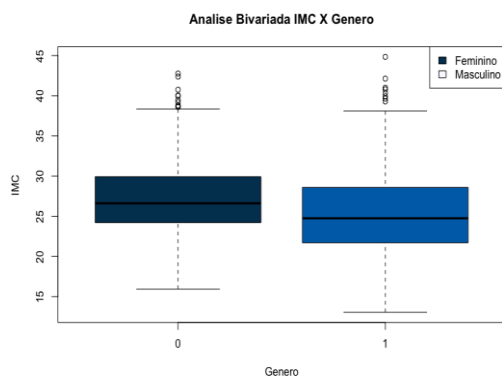


Gráfico 14: Análise bivariada IMC X Gênero

Var.test - P- Valor 0,0004069% - As Variâncias são diferentes.

T.test - P-Valor 0,0007

**CONCLUSÃO:** Como o p-valor = 0,0007 é menor que 5, rejeitamos a H<sub>0</sub> em favor a H<sub>a</sub>. Desta forma, é possível concluir que a média de IMC das mulheres que sofreram infarto é superior à média de IMC de homens na mesma condição, apontando uma relação entre IMC e gênero.

#### 5. Terceiro case: Interferência do álcool na habilidade de direção

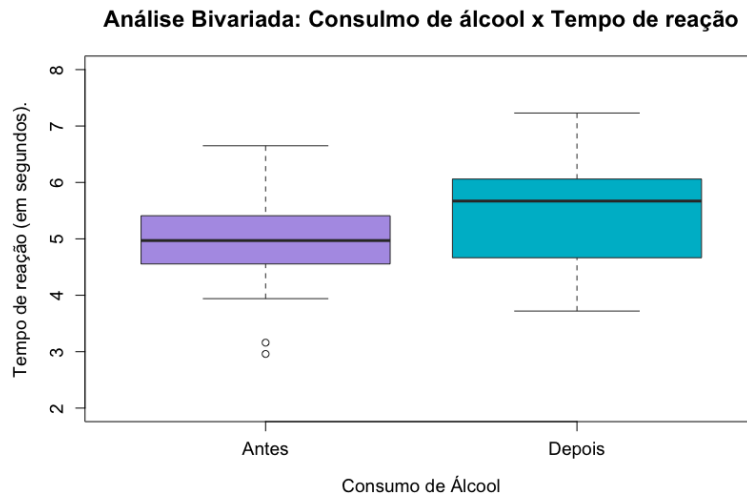
A ingestão de bebidas alcoólicas é uma das causas principais de acidentes com veículos. Uma amostra de 20 motoristas foi escolhida e seu tempo de reação em uma pista de obstáculos foi medido antes e depois da ingestão de duas cervejas. O objetivo do estudo foi verificar se os motoristas ficam prejudicados após beber duas cervejas. As amostras são pareadas.

##### 5.1 Diagrama sobre a relação sobre o consumo de álcool e o tempo de reação de um motorista

Para realizar esta análise bivariada, foram consideradas como hipóteses:

H0: A média do tempo de reação dos motoristas após o consumo de álcool é igual ou inferior do que a média do tempo de reação antes do consumo de álcool.

Ha: média do tempo de reação dos motoristas após o consumo de álcool é maior do que a média do tempo de reação antes do consumo de álcool



**Gráfico 15: Análise bivariada Consumo Álcool x Tempo de Reação**

t.test - P-Valor 0,0182

**CONCLUSÃO:** Os dados apresentam evidências de que o consumo de álcool torna o tempo de reação dos motoristas maior do que antes do consumo de álcool. Desta forma, rejeitamos a H0 em favor a Ha.

## 6. Quarto case: Ganho de peso em ratos

Um estudo foi conduzido com ratos para avaliar o impacto da dieta no ganho de peso nos animais. Para isso, eles foram induzidos a diferentes dietas com diferentes quantidades e foi avaliado o ganho de peso em cada um deles. No banco de dados temos as seguintes informações:

- Gain: ganho de peso (em gramas)
- Amount: classificação binária se recebiam pouca ou muita alimentação
- Type: qual tipo de comida recebeu: 1 = beef; 2 = pork; 3 = cereal.

## 6.1 Diagrama sobre a quantidade e tipo de alimento e o ganho de peso em ratos

H0 - A média de ganho de peso de ratos que receberam muita quantidade de alimentos é igual ou inferior a média de ganho de peso de ratos que receberam pouca quantidade.

Ha - A média de ganho de peso de ratos que receberam muita quantidade de alimentos é superior a média de ganho de peso de ratos que receberam pouca quantidade.

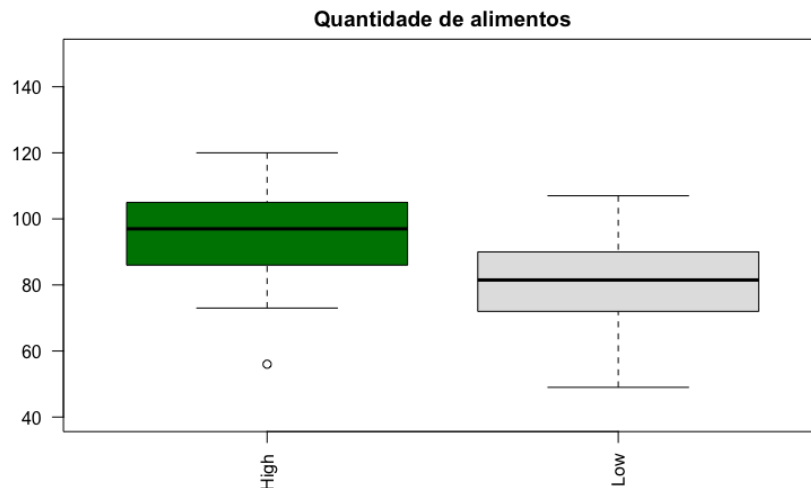


Gráfico 16: Análise bivariada: Ganho de peso x Quantidade de alimentos

var.test - P-Valor 0,95

t.test - <<<<<5%

**CONCLUSÃO:** Como P-VALUE é menor que 5%, podemos rejeitar a H0 em favor de Ha. Desta forma, concluímos que a quantidade de alimento dada aos ratos tem relação com o ganho de peso.

A seguir, testamos a relação entre o ganho de peso e o tipo de alimentação. As hipóteses elaboradas foram:

H0 - As médias de ganho de peso não se diferem com relação ao tipo de alimento consumido pelos ratos.

Ha - Ao menos uma das médias de ganho de peso se difere com relação ao tipo de alimento consumido pelos ratos.

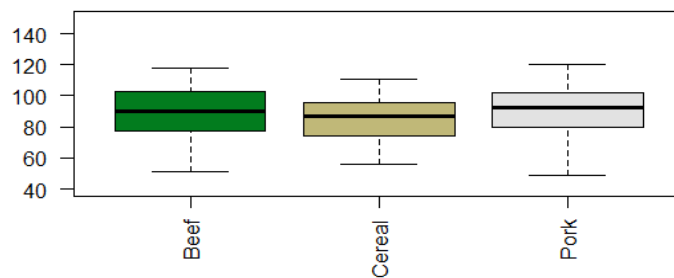


Gráfico 17: Análise bivariada Tipo de alimento x Ganho de peso

**CONCLUSÃO:** Os resultados mostram que ao menos uma das médias de ganho de peso se difere com relação a quantidade e tipo de alimento consumido pelos ratos. Desta forma, concluímos que a quantidade e o tipo de alimento dado aos ratos têm relação com o ganho de peso.

## 7. Quinto case: Gordura corporal em pacientes do sexo masculino

O estudo de caso a seguir busca entender se existem relações lineares entre medições de circunferência de várias partes do corpo de homens (além de outras informações) com o percentual de gordura corporal desses indivíduos.

### 7.1 Regressão linear simples

#### 7.1.1 Ganho de Peso x Idade

No teste de regressão linear simples por idade, temos o seguinte cenário:

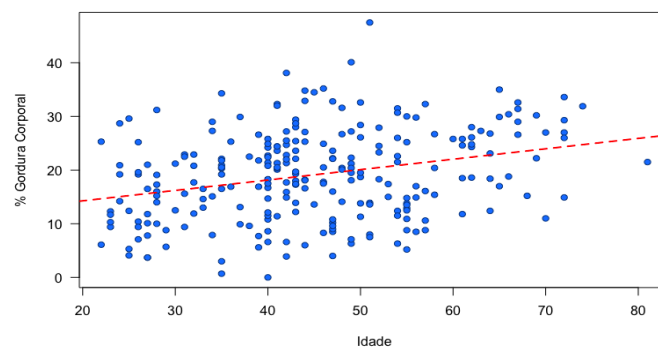


Gráfico 18: Regressão linear simples: Gordura corporal x Idade

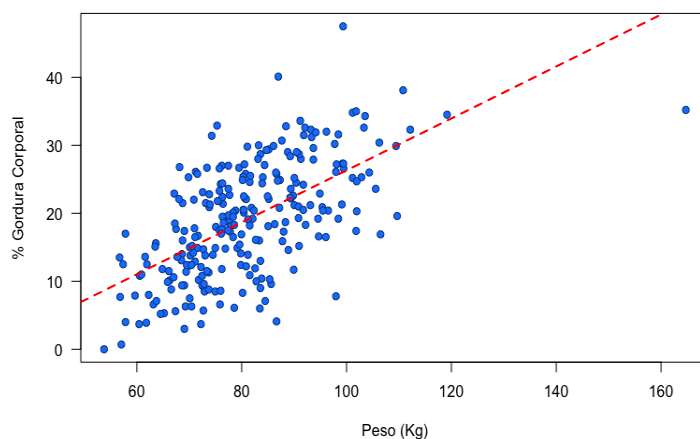


P – Valor <<<<<5%

R<sup>2</sup> = 0,086

Como o P-value é menor que 5%, concluímos que há relação linear entre idade e percentual de gordura corporal, de modo que a cada ano de idade, há um acréscimo de 0,19% de gordura corporal.

### 7.1.2 Gordura Corporal x Peso



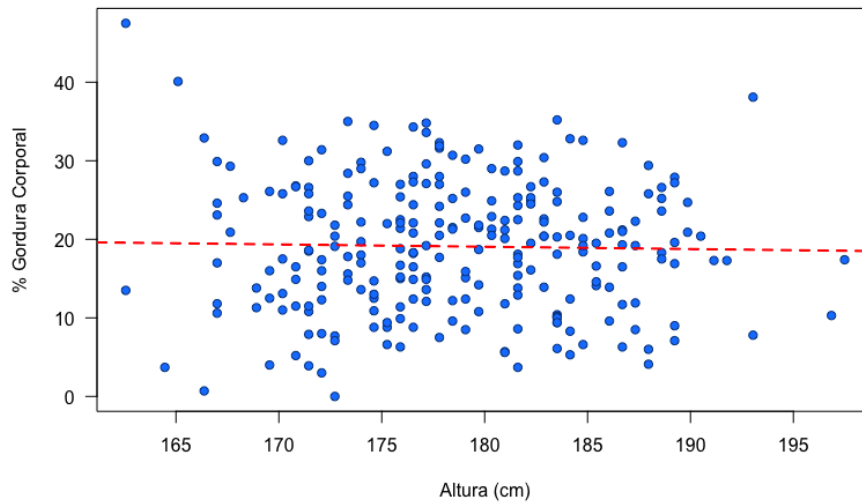
**Gráfico 19: Regressão linear simples: Gordura Corporal x Peso**

P- Valor <<<<<5%

R<sup>2</sup> = 0,3732

Como o P-value é menor que 5%, concluímos que há relação linear entre peso e percentual de gordura corporal, de modo que a cada quilo corporal, há um acréscimo de 1,85% de gordura corporal.

### 7.1.3 Gordura Corporal x Altura



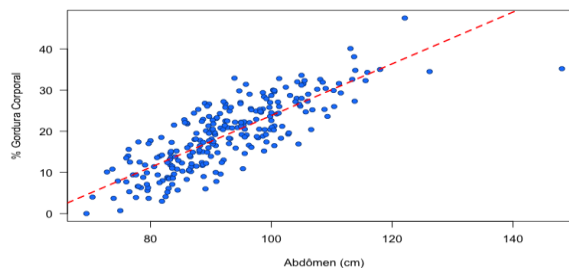
**Gráfico 20: Regressão linear simples: Gordura corporal x Altura**

P- Valor = 0,71

R2 = 0,0005

Como o P-value é maior que 5%, não podemos comprovar estatisticamente que há uma relação linear entre altura e percentual de gordura corporal.

#### 7.1.4 Gordura Corporal x Abdômen



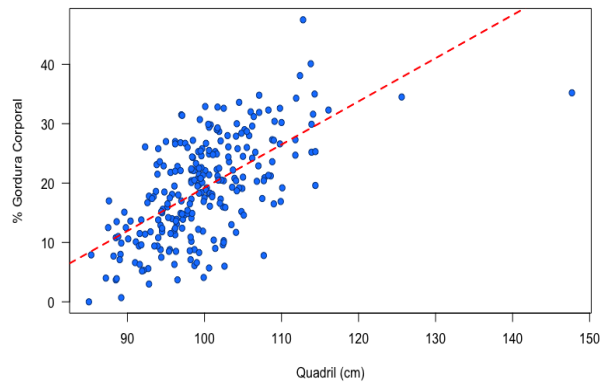
**Gráfico 21: Regressão Linear Simples: Gordura Corporal x Abdômen**

P- Valor <<<<<5%

R2 0,66

Como o P-value é menor que 5%, concluímos que há relação linear entre circunferência abdominal e percentual de gordura corporal, de modo que a cada centímetro da circunferência abdominal, há um acréscimo de 0,63% de gordura corporal.

#### 7.1.5 Gordura Corporal x Quadril



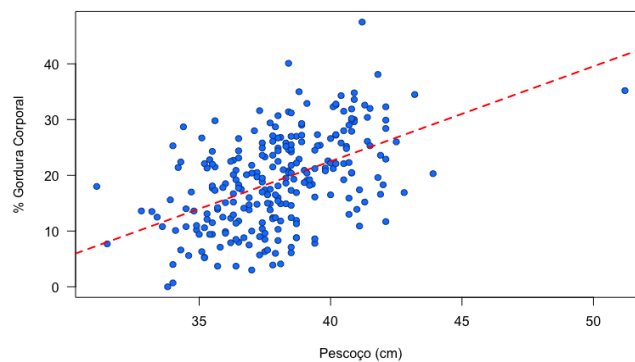
**Gráfico 22: Regressão Linear Simples: Gordura Corporal x Quadril**

P-Valor = <<<<<5%

R2 = 0,3846

Como o P-value é menor que 5%, concluímos que há relação linear entre circunferência do quadril e percentual de gordura corporal, de modo que a cada centímetro da circunferência do quadril, há um acréscimo de 0,73% de gordura corporal.

#### 7.1.6 Gordura Corporal x Pescoço



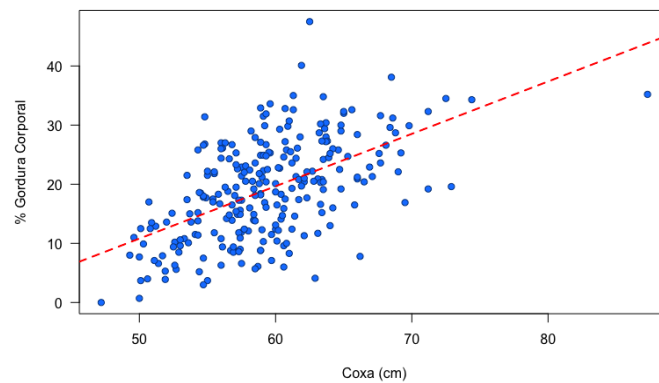
**Gráfico 23: Regressão Linear Simples: Gordura Corporal x Pescoço**

P-Valor <<<<<5%

R2 0,2474

Como o P-value é menor que 5%, concluímos que há relação linear entre circunferência do pescoço e percentual de gordura corporal, de modo que a cada centímetro da circunferência do pescoço, há um acréscimo de 1,70% de gordura corporal.

### 7.1.7 Gordura Corporal x Coxa



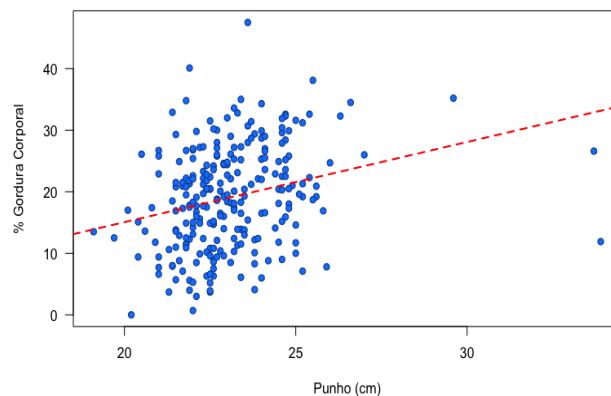
**Gráfico 24: Regressão Linear Simples: Gordura Corporal x Coxa**

P-Valor = <<<<5%

R<sup>2</sup> = 0,3065

Como o P-value é menor que 5%, concluímos que há relação linear entre circunferência da coxa e percentual de gordura corporal, de modo que a cada centímetro da circunferência da coxa, há um acréscimo de 0,88% de gordura corporal.

### 7.1.8 Gordura Corporal x Punho



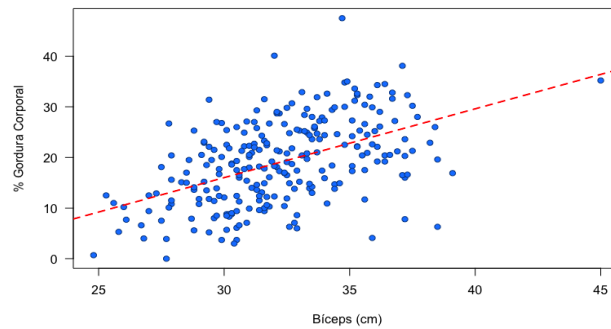
**Gráfico 25: Regressão Linear Simples: Gordura Corporal x Punho**

P- Valor <<<< 5%

R<sup>2</sup> = 0,070

Como o P-value é menor que 5%, concluímos que há relação linear entre circunferência do punho e percentual de gordura corporal, de modo que a cada centímetro da circunferência do punho, há um acréscimo de 1,30% de gordura corporal.

### 7.1.9 Gordura Corporal x Bíceps



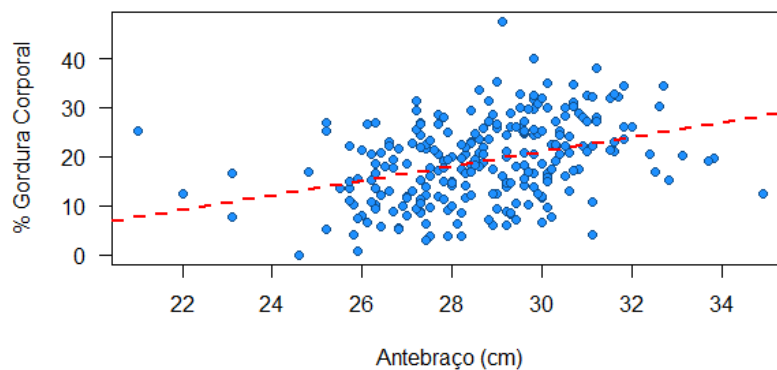
**Gráfico 26: Regressão Linear Simples: Gordura Corporal x Bíceps**

P- Valor= <<<<<5%

R2 = 0,24

Como o P-value é menor que 5%, concluímos que há relação linear entre circunferência do bíceps e percentual de gordura corporal, de modo que a cada centímetro da circunferência do bíceps, há um acréscimo de 1,36% de gordura corporal.

### 7.1.10 Gordura Corporal x Antebraço



**Gráfico 27: Regressão Linear Simples: Gordura Corporal x Antebraço**

P- Valor= <<<<<5%

R2= 0,1319

Como o P-value é menor que 5%, concluímos que há relação linear entre circunferência do antebraço e percentual de gordura corporal, de modo que a cada centímetro da circunferência do antebraço, há um acréscimo de 1,49% de gordura corporal.

## 7.2 Regressão Múltipla

A partir do resultado da regressão múltipla, as variáveis com maior significância estatística são peso, abdômen e pescoço.

```
Call:
lm(formula = Gordura.Corporal ~ ., data = data_bodyfat)

Residuals:
    Min       1Q   Median       3Q      Max
-10.312  -3.158   0.017   3.238  11.044

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -32.248970  19.269868  -1.674  0.09552 .
Idade         0.026999   0.029541   0.914  0.36165
Peso        -1.320902   0.580570  -2.275  0.02378 *
Altura      -0.004006   0.010027  -0.400  0.68987
Abdômen       0.996122   0.084485  11.791 < 2e-16 ***
Quadril     -0.212051   0.145334  -1.459  0.14586
Pescoço     -0.660808   0.228671  -2.890  0.00421 **
Coxa         0.263616   0.138265   1.907  0.05777 .
Punho        0.009571   0.214063   0.045  0.96437
Bíceps       0.167044   0.174491   0.957  0.33937
Antebraço    0.337101   0.196638   1.714  0.08776 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.369 on 240 degrees of freedom
Multiple R-squared:  0.7365,    Adjusted R-squared:  0.7255
F-statistic: 67.08 on 10 and 240 DF, p-value: < 2.2e-16
```

Imagem 1: sumário da regressão linear múltipla

## 7.3 Teste de Multicolinearidade

Foram observadas multicolinearidade entre as variáveis Peso, Abdômen, Quadril e Coxa:

Idade: 1.821803  
Peso: 33.289241  
Altura: 2.418860  
Abdômen: 10.864185  
Quadril: 13.987867  
Pescoço: 4.059452  
Coxa: 6.802001  
Punho: 1.730071  
Bíceps: 3.650845  
Antebraço: 2.075980

## 7.4 Conclusões

Com base nos dados, concluímos que as todas as variáveis, exceto a altura das pessoas da amostragem, mostraram correlação linear com o percentual de gordura corporal.