Estudo de caso 3: Emparelhamento de dados

25 de abril de 2016

1. Introdução

Uma percepção comum é a de que pessoas tendem a sistematicamente declarar um peso corporal inferior ao valor real. Com base nessa afirmação, investigaremos o viés de relato de peso dos alunos do curso de graduação em Engenharia de Sistemas da UFMG, mediante comparação dos valores estimados pelos estudantes da disciplina com valores aferidos por uma balança digital de uso doméstico.

O experimento realizado utiliza o conceito de emparelhamento de dados, no qual a hipótese a ser testada passa a ser relacionada com a diferença média entre as observações de cada amostra.

2. Coleta e análise exploratória dos dados

A Tabela 1 contém a amostra de dados coletada, ou seja, os pesos informados pelos alunos da disciplina e os aferidos pela balança, juntamente com um identificador aleatório e único para cada um deles:

ID	Peso Estimado(kg)	Peso Medido(kg)
EngSis.A	78.4	79.6
EngSis.B	69.5	69.3
EngSis.C	66.2	66.4
EngSis.D	61.0	60.4
EngSis.E	59.0	57.7
EngSis.F	70.5	71.3
EngSis.G	63.5	63.3
EngSis.H	81.0	79.3
EngSis.I	57.7	53.5
EngSis.J	49.0	52.0
EngSis.K	68	71.7

Tabela 1: Tabela de Amostras

Os valores dos pesos declarados e medidos foram exibidos no gráfico da Figura 1 para possibilitar uma melhor visualização do comportamento da amostra. O diagrama da Figura 2 retrata a diferença entre esses valores.

```r
#Ler dados de entrada
dados <- read.table("dados.csv", header=TRUE, sep=",");
aggdata <- aggregate(dados$Peso~dados$Fonte,data=dados,FUN=mean);
summary(aggdata);
```

```
##    dados$Fonte   dados$Peso
##  Estimado:1    Min.   :65.80
##  Medido  :1    1st Qu.:65.84
##                Median :65.88
##                Mean   :65.88
##                3rd Qu.:65.92
##                Max.   :65.95
```

```r
#Separando
estimados <- c(0);
ei <- 1;
medidos <- c(0);
mi <- 1;
for (i in seq(1:22)) {
  if (dados$Fonte[i] == 'Estimado') {
    estimados[ei] <- dados$Peso[i];
    ei <- ei + 1;
  } else {
    medidos[mi] <- dados$Peso[i];
    mi <- mi + 1;
  }
}

#Boxplot Pesos Estimados e Medidos em Balança
boxplot(estimados, medidos, ylab="Peso (kg)",
        names=c("Estimados","Medidos"), col = "lightgray",
        main="Pesos Estimados e Medidos em Balança");
```

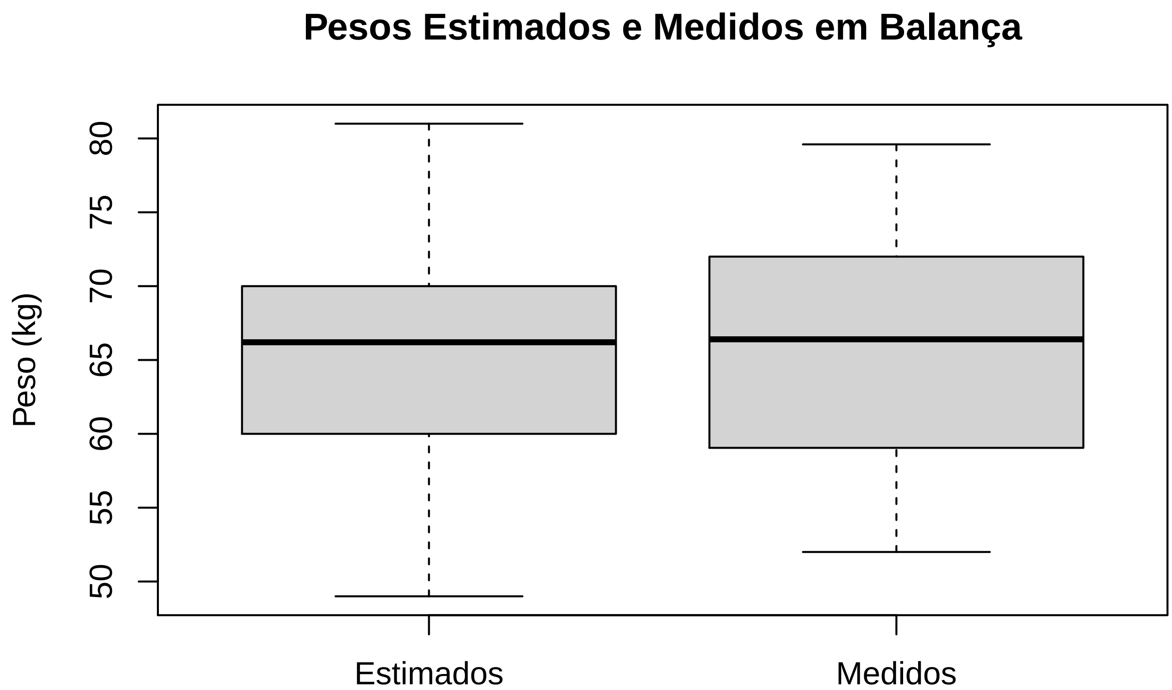

Figura 1:Pesos estimados e medidos

```r
#Boxplot Diferença dos Pesos Estimados e Medidos
diffPesos <- estimados - medidos;
boxplot(diffPesos, ylab="Peso (kg)",
        names=c("Diferença"), col = "lightgray",
        main="Diferença dos Pesos Estimados e Medidos")
```

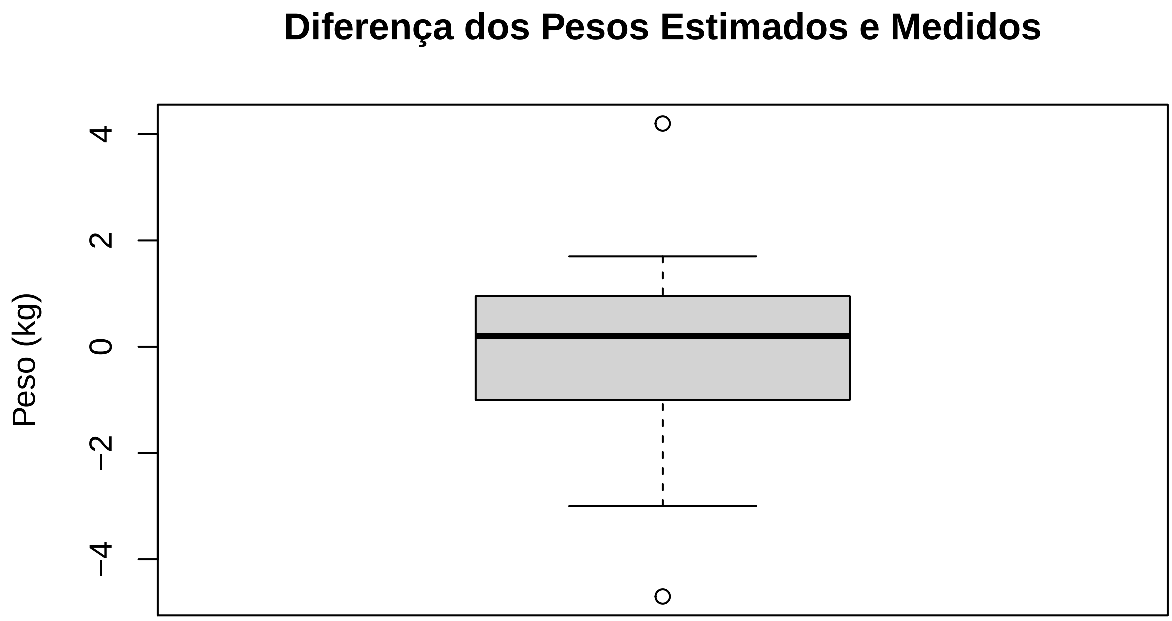

Figura 2: Diferença dos pesos estimados e medidos

De acordo com o Teorema do Limite Central, se a amostra tiver tamanho n suficiente, a distribuição amostral de $\bar{x}$ é aproximadamente Normal. Nesse caso, iremos assumir que n é suficiente, como será apresentado na seção de "Premissas do teste" deste relatório, verificamos esta premissa através do gráfico de normalidade presente na figura 3.

```
# qqPlot para testar a normalidade das diferenças entre
# o peso estimado e o peso medido.
require("car");
```

```
## Loading required package: car
```

```
## Warning: package 'car' was built under R version 3.1.3
```

```
qqPlot(diffPesos, pch=16, cex=1.5, las=1, main="Normalidade da diferença dos dados");
```

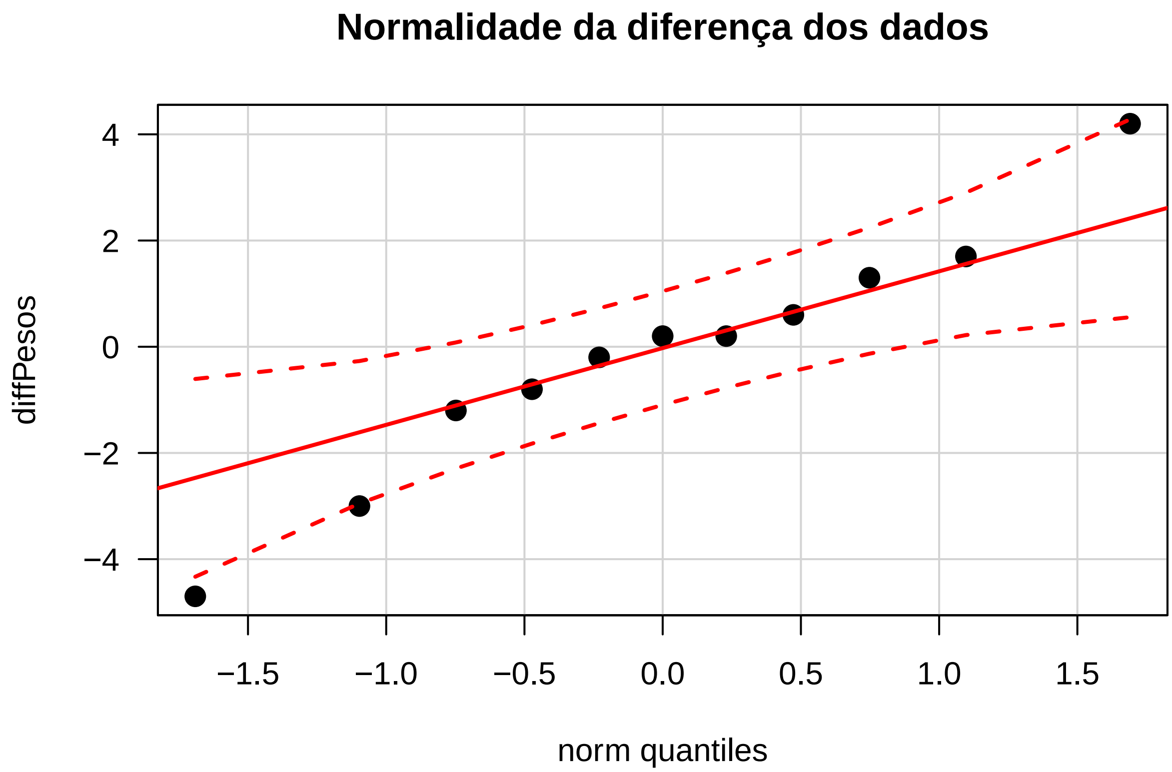

Figura 3: Normalidade da diferença dos dados

3. Estratégia de Inferência

O processo de inferência estatística consiste em tirar conclusões sobre uma população com base em informações extraídas de amostras da mesma. No presente estudo de caso, o parâmetro sobre o qual temos interesse é a diferença média entre os pesos medidos e estimados μ_D dos alunos de graduação em Engenharia de Sistemas.

O método se baseia em calcular a diferença entre cada par de pesos estimado e medido de cada aluno, determinar a média dessas diferenças e informar se essa média é estatisticamente significativa. Podemos utilizar o método de inferência para uma única amostra, visto que temos apenas duas observações de uma mesma amostra, as quais serão pareadas.

Parâmetros necessários

- O nível de significância (α), representa a probabilidade de erro tipo I, ou seja, a probabilidade de rejeitarmos a hipótese nula quando ela é efetivamente verdadeira. Pensando em uma taxa de erro aceitável para o domínio do problema, fixamos $\alpha = 0.05$.
- β representa a probabilidade de erro tipo II, ou seja, aceitarmos a hipótese nula quando ela é efetivamente falsa. Resolvemos permitir um erro tipo II de 20% ($\beta = 0.2$), considerando que esse tipo de erro tem menor impacto negativo do que o erro tipo I.
- O Nível de confiança ($1 - \alpha$) tem como objetivo conhecer o quanto o teste de hipóteses controla um erro do tipo I, ou qual a probabilidade de aceitar a hipótese nula se realmente for verdadeira.

- O poder do teste $(1 - \beta)$ tem como objetivo conhecer o quanto o teste de hipóteses controla um erro do tipo II ou, qual a probabilidade de rejeitar a hipótese nula se a mesma realmente for falsa.
- O menor tamanho de efeito de importância prática (δ^*) como 0.5, considerando o peso médio das roupas e um erro de precisão aceitável para a balança.

Hipóteses de Teste

O teste estatístico é planejado para avaliar a força da evidência **contra** a hipótese nula H_0. Usualmente, a hipótese nula é uma afirmativa de "nenhum efeito". A afirmativa sobre a população a **favor** da qual estamos tentando achar evidência é a hipótese alternativa H_1. Logo, as hipóteses são:

- $H_0 : \mu_D = 0$
- $H_1 : \mu_D \neq 0$

Com a formulação acima, vemos que nossa população de interesse corresponde à diferença entre os pesos estimados e medidos dos alunos de Engenharia de Sistemas, sendo μ_D a diferença média e o nosso parâmetro de interesse.

Premissas do teste

Antes da realização do teste temos que apresentar as seguintes premissas com relação ao mesmo.

1. Os dados formam uma distribuição normal com média μ e variância σ^2;
2. O desvio padrão de cada população é desconhecido;

4. Projeto experimental e Análise dos Resultados

Com as hipóteses definidas, assim como a população e seu parâmetro de interesse, devemos apresentar o tipo de teste de hipótese a ser utilizado para que a análise dos resultados possa ser feita corretamente.

O problema consiste na análise das estimações e das medidas dos pesos, ambas feitas no mesmo indivíduo, portanto, como se trata de um experimento pareado, utilizaremos o teste t pareado para concluir a investigação citada na introdução, o que é o objetivo deste trabalho.

Outra motivação para o uso desse teste é que, neste caso, ao contrário do teste t para duas amostras, o teste t pareado pode obter maior poder estatístico, pois não está sujeito a variação adicional causada pela independência das observações, afinal as observações pareadas são dependentes.

Ao rodar o teste t pareado no R, obtemos o seguinte resultado:

```
#Teste t pareado
t.test(dados$Peso~dados$Fonte, paired=T, data=aggdata)
```

```
##
##  Paired t-test
##
## data:  dados$Peso by dados$Fonte
## t = -0.2174, df = 10, p-value = 0.8323
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
## -1.738752  1.429661
## sample estimates:
## mean of the differences
##             -0.1545455
```

Analisando o resultado a cima, podemos notar que o valor de t obtido se encontra dentro do intervalo de confiança e o valor p é maior do que α. Portanto, não podemos rejeitar nossa hipótese nula.

O teste de potência considerando o desvio padrão da população como o desvio da amostra apresenta o seguinte resultado:

```
#Teste de potência
s <- sd(diffPesos);
power.t.test(n=11,
             delta=0.5,
             sd=s,
             sig.level=0.05,
             type="paired");
```

```
##
##      Paired t test power calculation
##
##               n = 11
##           delta = 0.5
##              sd = 2.358119
##       sig.level = 0.05
##           power = 0.09313085
##     alternative = two.sided
##
## NOTE: n is number of *pairs*, sd is std.dev. of *differences* within pairs
```

Dessa forma, o teste apresentou uma potência, significativamente baixa, de 9%. Esse valor é consequência da escolha do tamanho de efeito e do tamanho da amostra.

5. Conclusão

Com base nos resultados estamos 95% confiantes que a diferença média de peso está entre -1.739 e 1.43. Sendo assim, não temos evidências suficientes para afirmar que os alunos do curso de Engenharia de Sistemas tendem a informar peso menor que o real.

Como forma de descobrir quais seriam os parâmetros ideais para a obtenção de uma potência de teste de 80%, fizemos dois gráficos nos quais variamos o tamanho de efeito, ou o tamanho da amostra, versus a potência em cada um dos casos. As figuras 4 e 5 exibem esses resultados.

```
#Parâmetros de plots.
par(bg = "#dddddd");

#Potência variando tamanho de efeito.
delta <- 0.5;
powerBySe <- c(0);
es <- c(0);
for (i in seq(1:300)) {
```

```
  powerBySe[i] <- power.t.test(n=11,
                               delta=delta,
                               sd=s,
                               sig.level=0.05,
                               type="paired")$power;
  es[i] <- delta;
  delta <- delta + 0.01;
}
plot(es, powerBySe, type="l", lwd=2, las=1,
     main="Potência por Tamanho de Efeito", xlab="Tamanho de Efeito", ylab="Potência");
grid(NA,NULL,"white",lwd=2,lty=1);
```

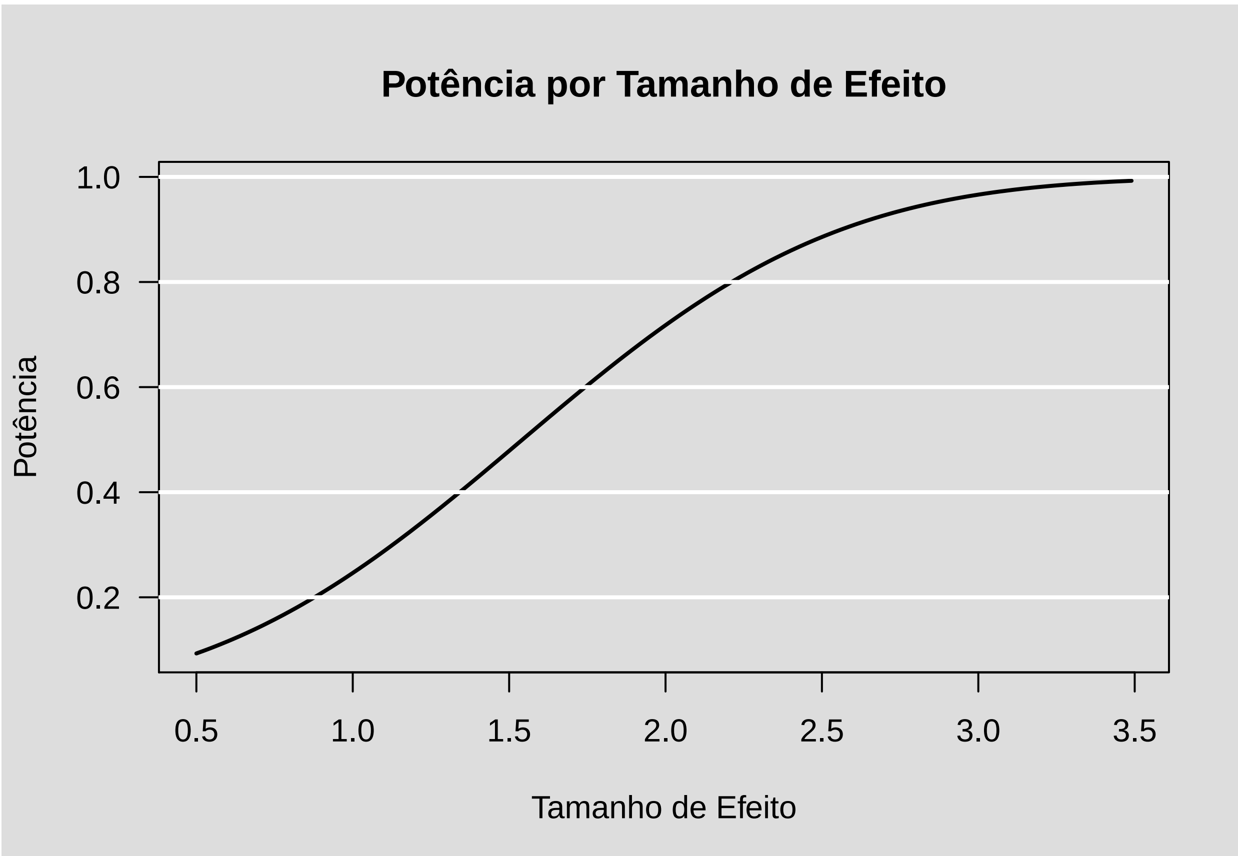

Figura 4: Potência x Tamanho de Efeito

```
#Potência variando quantidade de amostras.
n <- 11;
powerByN <- c(0);
ni <- c(0);
for (i in seq(1:1000)) {
  powerByN[i] <- power.t.test(n=n,
                              delta=0.5,
                              sd=s,
                              sig.level=0.05,
                              type="paired")$power;
  ni[i] <- n;
  n <- n + 0.5;
```

```
}
plot(ni, powerByN, type="l", lwd=2, las=1,
     main="Potência por Tamanho da Amostra", xlab="Tamanho da Amostra", ylab="Potência");
grid(NA,NULL,"white",lwd=2,lty=1);
```

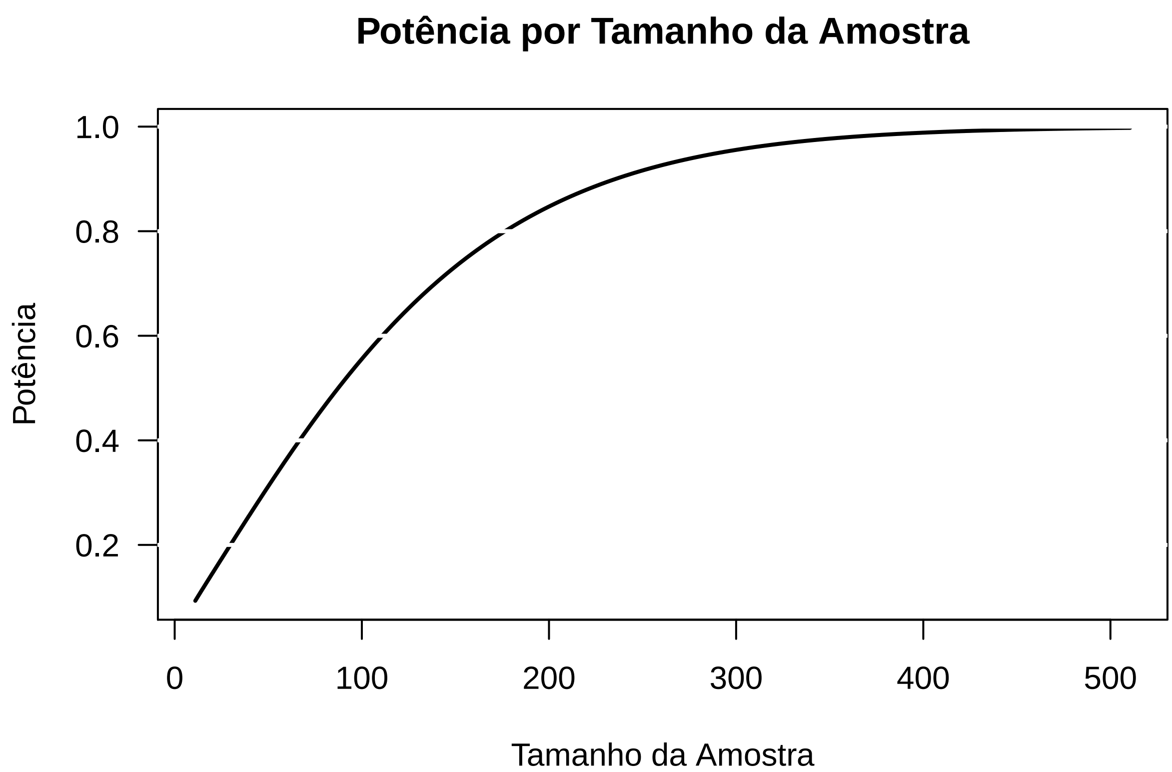

Figura 5: Potência x Tamanho da Amostra

Sendo assim, uma potência de 80% seria garantida se: mantivéssemos o tamanho da amostra fixo (n=11) e mudássemos o tamanho de efeito para, aproximadamente, 2.2; ou mantivéssemos o tamanho de efeito fixo ($\delta^* = 0.5$) e mudássemos o tamanho da amostra para, aproximadamente, 180.

Algumas formas de melhorar a qualidade do experimento seriam:

- Aumentar o número de amostras coletadas;
- Aumentar o tamanho das amostras.

6. Referências

[1] https://github.com/fcampelo/Design-and-Analysis-of-Experiments

[2] Estatística Aplicada e Probabilidade para Engenheiros (4a edição) - Montgomery

[3] A Estatística Básica e Sua Prática (6a edição) - David S. Moore, William I. Nortz, Michael A. Fligner

[4] https://stat.ethz.ch/R-manual/R-devel/library/stats/html/var.test.html

[5] http://ww2.coastal.edu/kingw/statistics/R-tutorials/independent-t.html

[6] http://www.portalaction.com.br/inferencia/56-teste-para-comparacao-de-duas-variancias-teste-f

[7] http://support.minitab.com/pt-br/minitab/17/topic-library/basic-statistics-and-graphs/hypothesis-tests/tests-of-means/why-use-paired-t/

[8] http://www.portalaction.com.br/inferencia/58-teste-t-pareado

[9] http://www.r-bloggers.com/paired-students-t-test/

[10] http://www.statmethods.net/stats/power.html