Estudo de Caso I - Modelos Estatísticos e Inferência

Marcus e Paulo

2 de abril de 2016

O Experimento

O BMI (*Body Mass Index* ou Índice de Massa Corporal) é o padrão pelo qual você pode ver se seus níveis de gordura e peso estão dentro do recomendado pela Organização Mundial de Saúde.

Sendo assim, o professor da disciplina de Modelos Estatísticos e Inferência da UFMG, Felipe Campelo, decidiu fazer um experimento com o intuito de comparar seu BMI com o dos seus alunos do curso de Engenharia de Sistemas. Sendo a pergunta motivadora do experimento "**Os alunos do curso de Engenharia de Sistemas estão, em média, mais "acima do peso" (de acordo com BMI) do que este professor**?".

Tentamos inferir o que seria a verdadeira resposta à pergunta acima a partir de uma amostra com o peso e a altura de cada aluno da disciplina de Modelos Estatísticos e Inferência coletada pelo professor e preenchida pelos mesmos. A partir disto, foi calculado o BMI dos 13 alunos, com isso temos um conjunto de informações que podem ser usadas para a análise do estudo de caso.

Definições das Hipóteses do teste

Seguindo a metodologia estatística definimos uma hipótese nula e uma alternativa, com base no BMI do professor de 26.3 Kg m^{-2}:

- $H_0 : \mu = 26.3$
- $H_1 : \mu > 26.3$

Definimos então que iremos utilizar o **teste-t**. O motivo da escolha do mesmo é devido a simplicidade de sua utilização no R e também porque este teste é utilizado quando não conhecemos a variância real do processo.

Dados Utilizados

Para realização dos testes iremos utilizar dados adquiridos de forma voluntária dos alunos da disciplina que submeteram anonimamente suas massas e alturas, e organizaram os dados via internet em um documento **.csv**.

Estimação do tamanho do efeito e intervalo de confiança

O tamanho de efeito é responsável por quantificar o módulo do desvio observado da hipótese nula. A medida de tamanho de efeito apropriada para o teste-t com uma amostra é a Cohen's d. Indicada pela fórmula abaixo:

$$d = \frac{(X - \mu)}{s}$$

Ao colocar o valor da hipótese nula, o desvio padrão estimado e a média dos BMIs das amostras, notamos um tamanho de efeito de -1.473904, o que indica que os dados estão **desviando** para valores menores do que o da hipótese nula.

Podemos observar que na função t.test utilizamos no campo da hipótese alternativa o valor **two.sided** (bilateral), para a função nos retornar um intervalo de confiança bilateral. Mas mantemos, para comparação, a hipotése alternativa como um valor maior que a hipótese nula. O valor de alpha foi escolhido a priori como sendo 0.01, ou seja, o intervalo de confiança foi de 99% e corresponde a valores menores do que a hipótese nula, de acordo com o que será mostrado no tópico de Simulação deste relatório.

Verificação e discussão das Premissas

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

Dados:

1. Os alunos da turma de Modelos Estatísticos e Inferência são um grupo representativo dos alunos do curso de Engenharia de Sistemas;
2. Os alunos reportaram suas massas e alturas corretamente, sem viés.

Teste:

1. O intervalo usado na função t.test deve ser um intervalo bilateral, uma vez que é pedido um intervalo de confiança bilateral;
2. Na análise consideramos o fato da hipótese alternativa ser direcional, apesar do intervalo de confiança ser bilateral;
3. As observações formam uma distribuição normal com média μ e variância σ^2;
4. Desvio padrão da população é desconhecido;

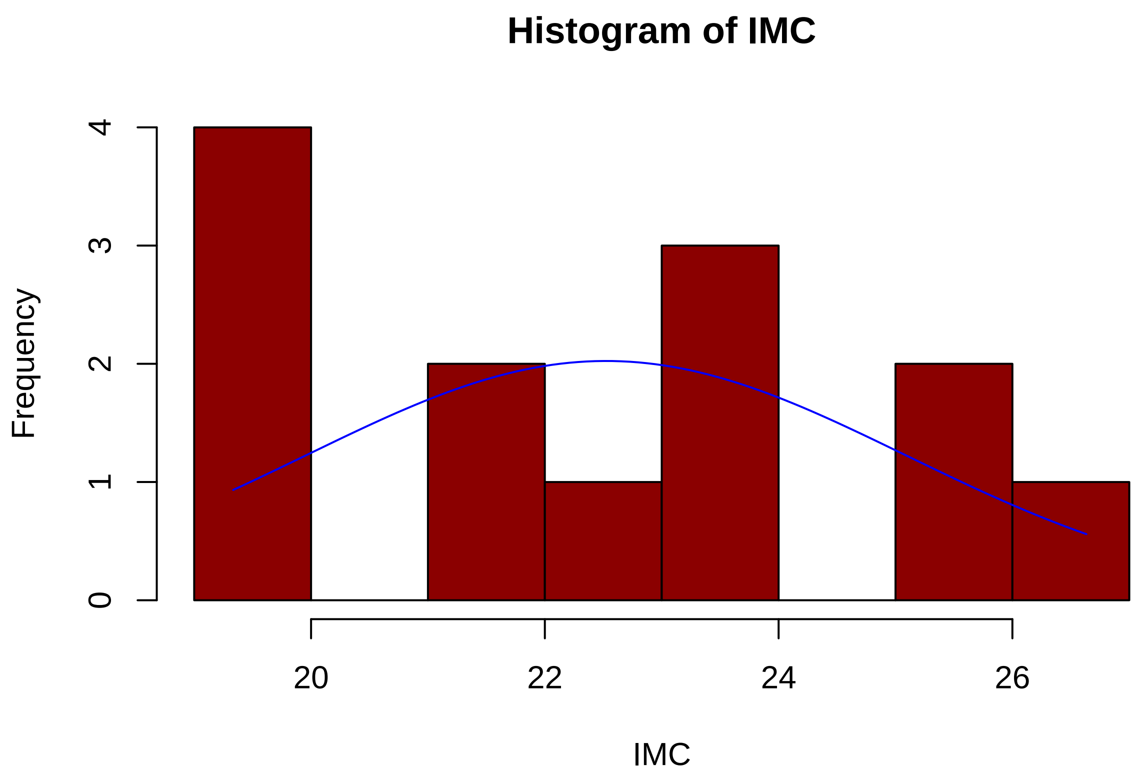
Gráficos Utilizados

Seguem os gráficos que foram utilizados para visualização dos dados e das distribuições:

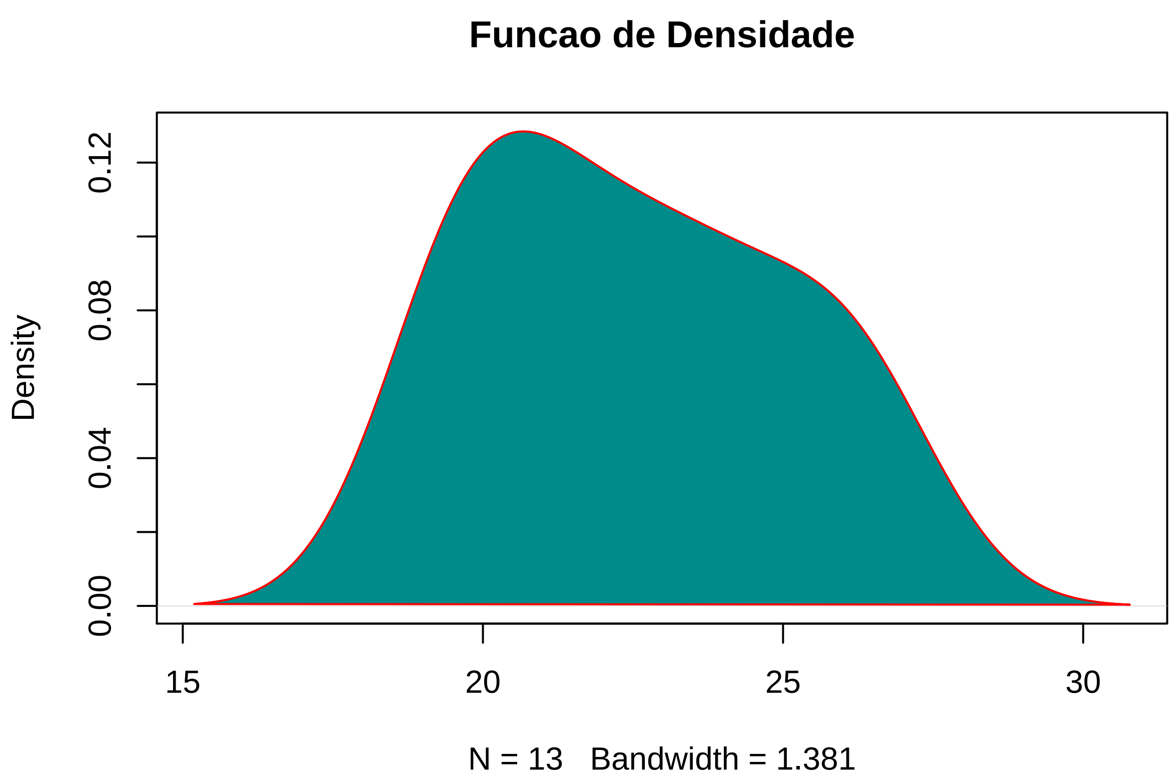
```
##Leitura dos Dados
FileName = "Dados.csv";
FileName = paste(getwd(),FileName,sep="/");
Dados = read.csv(FileName, header = TRUE, sep=",");

## Calculo IMC
Massa = Dados[,1];
Altura = Dados[,2];
IMC = (Massa) / (Altura*Altura);

#Plot do histograma dos dados
h <- hist(IMC,breaks = 7, col = "dark red")
xfit<-seq(min(IMC),max(IMC),length=100)
yfit<-dnorm(xfit,mean=mean(IMC),sd=sd(IMC))
yfit <- yfit*diff(h$mids[1:2])*length(IMC)
lines(xfit, yfit, col="blue")
```




```r
# Prova de que os dados não sao normais
d <- density(x = IMC)
plot(d, main = "Funcao de Densidade", col = "red")
polygon(d, col="darkcyan", border = "red")
```



Pelas imagens nota-se que os dados coletados não formam exatamente uma normal, mas podem ser aproximados.

Simulação

```
## Definicao de Variaveis
alpha = 0.01;
expectedMean <- 26.3

#Teste T
t2Sided = t.test(x = IMC,
       alternative= "two.sided",
       mu = expectedMean,
       conf.level = 1-alpha);
print(t2Sided)
print("Hipotese nula é rejeitada.")


#Valores Criticos
a <- qt(1-0.005,12);
b <- qt(0.005,12);

print( paste("Os Valores Criticos sao :",a,"e",b,sep=" ")  )


estimateVar <- function(Data)
{
  Mean <- mean (Data);
  Var <- sapply(Data, function(x) (x-Mean)**2 );
  Var <- sum(Var)/(length(Var)-1);
  Var
}

Var <- estimateVar(IMC);
desvioPadraoEstimado <- sqrt(Var);
effectSize <- ( mean(IMC) - expectedMean) / sqrt(Var)

print( paste("Desvio Padrao estimado = ", desvioPadraoEstimado,sep=" "))
print(paste("Tamanho de efeito", effectSize, sep=" "))
```

```
##
##  One Sample t-test
##
## data:  IMC
## t = -5.3142, df = 12, p-value = 0.000184
## alternative hypothesis: true mean is not equal to 26.3
## 99 percent confidence interval:
##  20.35204 24.69396
## sample estimates:
## mean of x
##    22.523
##
```

```
## [1] "Hipotese nula é rejeitada."
## [1] "Os Valores Criticos sao : 3.0545395893929 e -3.0545395893929"
## [1] "Desvio Padrao estimado =  2.56258202347936"
## [1] "Tamanho de efeito -1.47390393221882"
```

Conclusões e Recomendações

Como estamos considerando a hipótese alternativa sendo direcional (maior que), para que a hipótese nula seja desconsiderada o valor de t_0 tem que ser maior em módulo que o valor de $t_{critico}$, o que neste caso é verdade, como foi visto após a execução do teste-t no arquivo **.R** (5.3142 > 3.05454). Esta conclusão foi tirada simplesmente olhando pro valor de t_0 dado pela função **t.test** do R. Podemos tirar esta mesma conclusão utilizando o valor-p. Quando este for menor que o nível de significância descartamos a hipótese nula. Ao olhar o intervalo de confiança bilateral percebemos que os valores são menores que o da hipótese nula, confirmando o que notamos na seção sobre tamanho de efeito. Com estas informações, observamos que podem ser inferidos valores de BMI médio dos alunos do curso de Engenharia de Sistemas, como sendo valores na faixa de 20.35204 Kg m^{-2} a 24.69396 Kg m^{-2}. Isto sugere que os alunos na verdade estão, em média, menos **acima do peso** (levando em conta o BMI) do que o professor Felipe Campelo. Portanto, podemos parabenizar os alunos que em média estão com seus BMIs não só fora da faixa de pré-obesos (BMI de 25.00 a 29.99 Kg m^{-2}), como dentro da faixa normal (18.50 a 24.99 Kg m^{-2}). É recomendado que esses alunos continuem mantendo estes valores com uma alimentação saudável. Ao professor é recomendado que seja feita uma dieta balanceada juntamente com exercícios para a queima de gordura.

Potência do Teste

A hipótese nula foi descartada depois da execução do teste, o que não torna a potência do mesmo importante na análise deste estudo de caso.

Possíveis formas de melhorar o Experimento

Uma possível forma de melhorar este experimento seria aumentar o espaço amostral (número de observações), pegando informações não só da turma de modelos estatísticos e inferência, mas sim de vários estudantes do curso de Engenharia de Sistemas de vários períodos diferentes. Porque assim a amostra ficaria mais próxima ainda de uma curva normal. A coleta da amostra deve ser feita por alguns alunos selecionados que medirão a altura e a massa de todos os outros, de forma que os dados fiquem mais confiáveis.