- Tutorial
- Exercícios
- Apostila

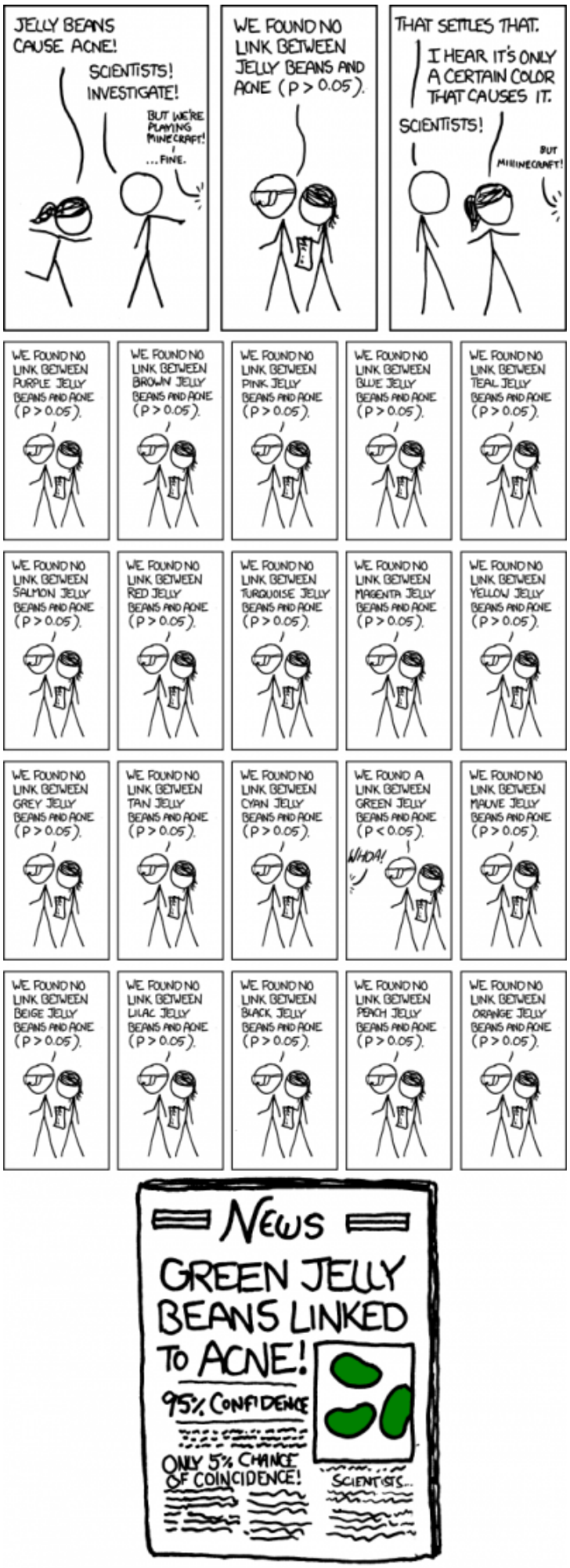
6a. Teste de Hipótese

Video de aula síncrona gravada no google meet em 28/09/2020, não editado.





Video


JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
... FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
News
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

O teste de hipótese é um instrumento poderoso para a tomada de decisão e parte fundamental do procedimento científico de experimentos. Os testes estão baseados no conceito de variável aleatória, que são aquelas em que o resultado de um evento pode variar. Ou seja, quase tudo o que nos rodeia. Por exemplo, *Eucalyptus saligna* em talhão de cultivo terão uma taxas de crescimento similar, mas não exatamente a mesma. O diâmetro do tronco, após sete anos de plantio, não será o mesmo para todas as árvores. Essa variabilidade tem várias fontes, genética, ambiental ou acidental, e é inerente aos dados biológicos. O esforço no cultivo é justamente no sentido de buscar as melhores taxas de crescimento e menor variação possível, para que o resultado seja eficiente e previsível. Por isso se usa mudas provenientes de clones a partir de cultura de tecido, para controlar pelo menos essa fonte de variabilidade. No teste de hipótese partimos do fato que os dados podem variar e avaliamos se o resultado encontrado pode ter sido gerado pelo acaso e não pelo tratamento que estamos testando. No caso do *Eucalyptus* poderíamos estar interessados no efeito, por exemplo, de um tipo específico de adubo. Comparando mudas que foram colocadas em tratamentos com e sem adubo iremos, quase certamente, encontrar diferenças nos tamanhos das árvores dos dois grupos. A pergunta subjacente é: será que essa diferença encontrada poderia ter sido gerada apenas por outros fatores ou o acaso? Por exemplo, por sorte, poderíamos ter amostrado uma proporção de árvores que cresceram mais em um dos tratamentos e uma proporção menor no outro. Isso simplesmente por acaso! Considerando que há variação no crescimento dos indivíduos, há uma probabilidade desse padrão emergir, nesse caso, simplesmente porque fizemos uma amostra das árvores nas duas condições. O teste de hipótese é o instrumento para nos guiar nessa interpretação. Vamos visitar estes e outros conceitos associados, utilizando as ferramentas disponíveis no R.

Chacal Dourado



Vamos utilizar o exemplo de tamanho de mandíbulas de chacal dourado que se encontra no livro do Bryan Manly[1] sobre randomização e outros métodos que iremos ver mais a frente. Os dados são provenientes da coleção do Museu Britânico de História Natural em Londres e foram publicados em um artigo de 1980. Aqui estamos interessados apenas nos tamanhos das mandíbulas e se há diferença entre fêmeas e machos.

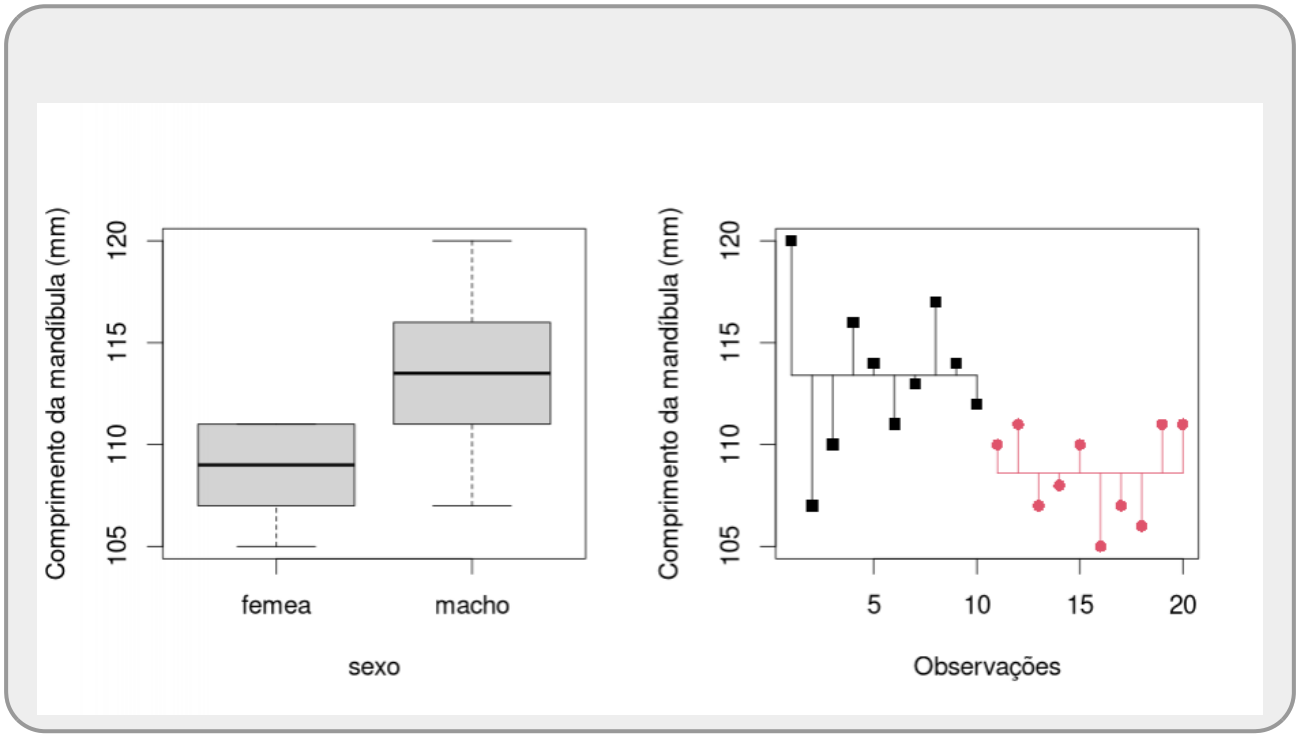
Como foram avaliados apenas 10 machos e 10 fêmeas, vamos entrar esses dados diretamente no R:

```
macho <- c(120, 107, 110, 116, 114, 111, 113, 117, 114, 112)
femea <- c(110, 111, 107, 108, 110, 105, 107, 106, 111, 111)
chacal <- c(macho, femea)
sexo <- factor(rep(c("macho", "femea"), each=10))
```

Dois Gráficos para ver os mesmos dados

Vamos avaliar esses dados graficamente. O código abaixo produz um gráfico de caixa (boxplot) e também um gráfico, pouco usual, mas que nos permite visualizar a variação que existe nos dados. Vamos usar esse tipo de representação gráfica ao longo desse tutorial. Tenha certeza que entendeu o que está representado nessa figura!

```
x11(width = 12, heigh = 6)
par(mfrow = c(1, 2), cex = 1.5)
boxplot(chacal ~ sexo, ylab = "Comprimento da mandíbula (mm)")
plot(1:20, chacal, pch = rep(c(15, 16), each = 10), col = rep(1:2, each =
10), xlab = "Observações", ylab = "Comprimento da mandíbula (mm)")
medsex <- c(mean(macho), mean(femea))
segments(x0 = 1:20, y0 =  chacal, y1 =  rep(medsex, each = 10), col=
rep(1:2, each = 10))
lines(c(1,10), c(medsex[1], medsex[1]),col=1)
lines(c(11,20),c(medsex[2], medsex[2]),col=2)
```



Agora que já visualizou o gráfico, vamos retornar os parâmetros globais para o padrão e fechar o dispositivo de tela:

```
par(mfrow = c(1,1), cex = 1)
dev.off()
```

Pergunta



Por que estamos olhando para esses dados? Por que estou fazendo esse curso? Podemos nos perguntar tantas coisas... O importante é saber o que estamos fazendo e para quê. Qual a nossa hipótese?! Neste tutorial, a pergunta é usada para ilustrar o teste de hipótese. Mesmo assim, é a condutora de toda a lógica do teste. Podemos fazer uma pergunta simples, por exemplo: se o tamanho da mandíbula difere entre machos e fêmeas. Uma outra pergunta poderia ser se os **machos tem mandíbulas maiores que as fêmeas**. A princípio vamos nos ater à segunda. Nesse caso, uma medida que nos fornece informações sobre nossa pergunta é a diferença entre as mandíbulas de machos e fêmeas. Mas qual valor iremos usar para calcular esse diferença? Há várias possibilidades, vamos usar a mais intuitiva, a média! E nossa pergunta passa a ser um pouco diferente e mais precisa: será que **a mandíbula de machos são em média maiores que as das fêmeas**?

Vamos olhar para esses valores e armazenar essa diferença entre médias:

```
tapply(chacal, sexo, summary)
tapply(chacal, sexo, mean)
tapply(chacal, sexo, sd)
diff(tapply(chacal, sexo, mean))
difsex <- diff(tapply(chacal, sexo, mean))
```

Estatísticas de interesse

A estatística de interesse é um valor que resume nossa hipótese, nesse caso, a diferença do tamanho médio entre sexo. Nossa resposta está dada: as mandíbulas de machos são em média maiores que as das fêmeas! Além disso, temos uma estimativa do efeito associado: em média os machos tem mandíbulas `4.8 mm` maiores que as fêmeas. Algumas indagações emergem a partir desta afirmação:

* **Essa diferença pode ser gerada pelo acaso?**

Ou ainda mais precisamente:

- **Qual a possibilidade desta diferença ser gerada pelo acaso?**

Variável Aleatória

A pergunta acima pode ser respondida se tomarmos algumas pressupostos sobre a variabilidade dos dados. Por exemplo, que a variável aleatória, tamanho de mandíbula, se ajusta a uma distribuição normal. Vamos olhar os valores e compará-lo com a distribuição normal com mesma média e desvio padrão dos dados.

```
hist(chacal, freq=FALSE,xlim=c(95,125))
curve(exp=dnorm(x, mean=mean(chacal),sd=sd(chacal)),from=95,to=125,
col="red", add=TRUE)
```

Como só temos 20 valores, vamos ser generosos com os dados e considerar que estão bem acoplados a uma distribuição probabilística normal, para fins didáticos[2)].

Simulando Dados

Tendo a premissa acima podemos utilizar a função `rnorm` para simular amostras com as mesmas características dos nossos dados. Mais que isso, podemos simular o cenário associado à nossa hipótese e calcular a estatística de interesse:

```
## 10 valores de um distribuicao normal
rnorm(10,mean = mean(chacal),sd = sd(chacal))
## diferenca entre as medias de 10 valores
mean(rnorm(10, mean = mean(chacal), sd = sd(chacal))) - mean(rnorm(10, mean
= mean(chacal), sd = sd(chacal)))
```

Para deixar o código mais compacto, vamos atribuir a média e o desvio padrão dos dados ao objeto `mch` e `sdch`.

```
(mch <- mean(chacal))
(sdch <- sd(chacal))
(mean(rnorm(10, mean = mch,sd = sdch)) - mean(rnorm(10, mean = mch), sd =
sdch))
```

Cenário Nulo



O cenário acima simula a situação em que os valores de machos e fêmeas são provenientes da mesma distribuição normal. Portanto, qualquer diferença encontrada é devida apenas ao acaso. A fonte nesse caso é o procedimento de amostra. Justamente o caso descrito na abertura desse tutorial. Percebam como os valores resultantes variam:

```
mean(rnorm(10, mean = mch, sd = sdch)) - mean(rnorm(10, mean = mch, sd =
```

```
sdch))
mean(rnorm(10, mean = mch, sd = sdch)) - mean(rnorm(10, mean = mch, sd =
sdch))
mean(rnorm(10, mean = mch, sd = sdch)) - mean(rnorm(10, mean = mch, sd =
sdch))
mean(rnorm(10, mean = mch, sd = sdch)) - mean(rnorm(10, mean = mch, sd =
sdch))
mean(rnorm(10, mean = mch, sd = sdch)) - mean(rnorm(10, mean = mch, sd =
sdch))
mean(rnorm(10, mean = mch, sd = sdch)) - mean(rnorm(10, mean = mch, sd =
sdch))
mean(rnorm(10, mean = mch, sd = sdch)) - mean(rnorm(10, mean = mch, sd =
sdch))
mean(rnorm(10, mean = mch, sd = sdch)) - mean(rnorm(10, mean = mch, sd =
sdch))
mean(rnorm(10, mean = mch, sd = sdch)) - mean(rnorm(10, mean = mch, sd =
sdch))
```

Fazendo isso muitas vezes podemos comparar o valor observado com a distribuição esperada devido ao acaso. Com isso podemos responder nossa pergunta: será que o acaso poderia gerar a diferença observada? Copiar esse linha de comando 1000 vezes seria tedioso! Para resolver isso, vamos aprender uma ferramenta poderosa e um clássico na linguagem de programação: a **iteração** ou ciclos.

Criando Ciclos

A função `for()` cria ciclos de eventos que se repetem, sua sintaxe é simples, basta eleger uma variável, no exemplo abaixo `i`[3] e indicar em um vetor, quais os valores que essa variável vai assumir em cada ciclo. No caso do código abaixo, os valores de 1 a 10. Entre `{ }` colocamos o procedimento que queremos repetir. A cada novo ciclo, `i` assume o valor seguinte do vetor que foi definido (`1:10`).

Veja o efeito do código abaixo. A função `cat()` mostra na tela (no console do R) o valor atribuído ao objeto que fornecemos. Os símbolos "\t" e "\n" são os caracteres no R que definem tabulação e quebra de linha em uma cadeia de caracteres, respectivamente.

```
for(i in 1:10)
{
    cat("\n\t", i)
}
```

Pronto! **Acabaram de aprender uma das ferramentas mais poderosa em programação!**

Vamos utiliza-la então para simular, muitas e muitas vezes, o nosso cenário nulo. Aqui, o pulo do gato é preparar um objeto, antes de iniciar o ciclo, para guardar os resultados de cada iteração. A variável `i`, neste caso, serve tanto como contador dos ciclos, quanto para posicionar o resultado no objeto onde iremos guardar a informação. Uma boa prática é preencher o objeto que irá guardar o resultado com NAs ao criá-lo. Assim, se algo sair errado na iteração, iremos perceber logo que tentarmos operar o resultado!

```
nsim <- 1000
## criando o objeto para guardar o resultado
cenaNula <- rep(NA, nsim)
## ciclo de iteração
cenaNula[1] <- difsex
for(i in 2:nsim)
{
    cenaNula[i] <- mean(rnorm(10, mean = mch, sd = sdch)) - mean(rnorm(10,
mean = mch, sd = sdch))

}
str(cenaNula)
```

Visualizando a simulação

Vamos agora usar a **animada** função chamada simula.r que produzimos para ilustrar a simulação:



Atenção usuários do RStudio: a função `simula()` não é tão animada assim na janela do RStudio. Para que ela funcione, antes de usá-la abra uma janela gráfica com o comando `x11()`

Primeiro precisamos carregá-la no espaço de trabalho, com o comando `source`. Depois, é só simular!

```
x11(width = 10, height = 10)
source("simula.r")
simula(chacal[sexo == "macho"], chacal[sexo == "femea"])
```

Probabilidade do Acaso

Vamos usar os valores que produzimos para gerar um histograma da nossa simulação.

```
par(las = 1, cex = 1.2)
hist(cenaNula, col = rgb(0, 0, 1, 0.3))
```

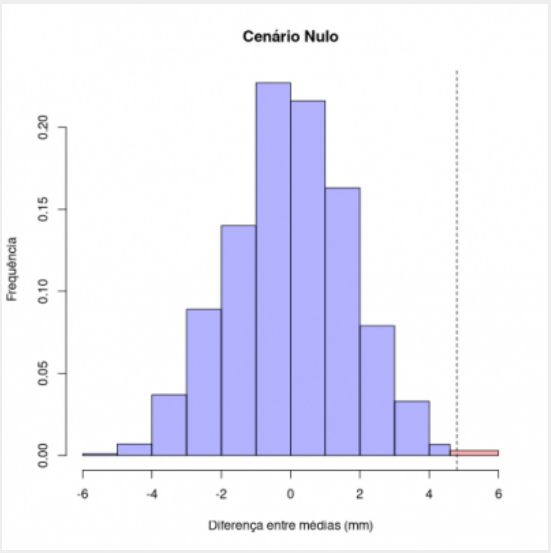
Agora, podemos calcular agora quantas vezes o cenário nulo, ou o acaso, gerou diferenças iguais ou maiores que a observada nos dados, assim como, a probabilidade associada a este valor:

```
sum(cenaNula >= difsex)
```

Vamos indicar esses valores no gráfico:

```
histNull <- hist(cenaNula, plot = FALSE, breaks = c(min(floor(cenaNula)),
-4.6:4.6, max(ceiling(cenaNula))))
str(histNull)
```

```
histNull$breaks
cols <- rep(c(rgb(0, 0, 1, 0.3), rgb(1, 0, 0, 0.3)),   c(10, 2))
plot(histNull,  main = "Cenário Nulo", xlab = "Diferença entre médias (mm)",
ylab = "Frequência", col = cols)
abline(v = difsex, lty = 2)
```



Dividindo esse valor pelo número de simulações que foram feitas temos uma estimativa da **probabilidade do acaso**.

```
sum(cenaNula >= difsex)/length(cenaNula)
```

Interpretado o resultado

O cenário nulo, onde não há diferenças entre os sexos, gerou valores maiores ou iguais à diferença observada em 3 casos[4] em 1000. Isso significa uma proporção de **0.003** ou **0.3%** dos casos.

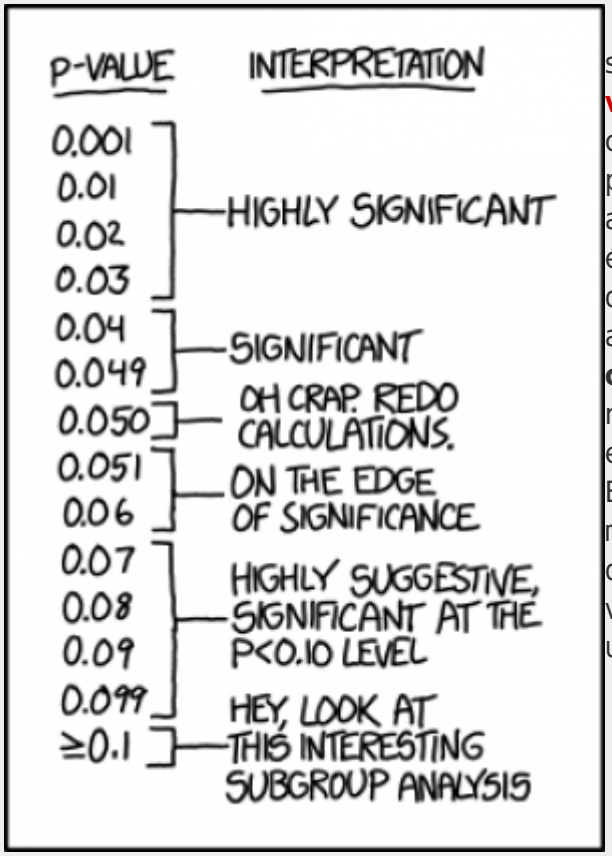
Como essa probabilidade é muito baixa, eu particularmente, me sinto confortável em afirma que:

Os machos de chacal dourado apresentam mandíbulas maiores que as fêmeas.

E posso dizer ainda, que posso estar errado, mas a probabilidade de incorrer em erro ao fazer essa afirmação é pequena, cerca de `0.3%`.

O valor de probabilidade que acabamos de calcular é, sem dúvida, a estatística mais famosa e polêmica do teste de hipótese: o **p-valor**.

I CAN SEE CLEAR NOW!



O p-valor está em crise! Muitos artigos tem sido publicados recentemente discutindo o **p-valor**. Um dos pontos mais atacados é o já consagrado limite de 5% de probabilidade para a significância do resultado. Outro ponto atacado é a própria palavra **significância** que é mal interpretada ou pior, não define bem o que o resultado do p-valor significa! Alguns autores sugerem que usemos algo como **clareza** do resultado. Por exemplo, o nosso resultado é claro, as mandíbulas de machos, em média, são maiores que as das fêmeas. Entretanto, será que essa diferença de `4.8 mm` é relevante ou significante? Será que essa diferença, tão pequena, condiciona alguma variação biologicamente significativa? Aqui uma seleção de artigos sobre o tema:

- I can see clearly now: reinterpreting statistical significance. Dushoff J.; Klain M.P. & Bolker B.M. 2019.Methods in Ecology and Evolution 10(6): 756-759.
- The ASA Statement on p Values Context Process and Purpose. Wasserstein R.L. & Lazar, N.A. 2016. The ASA Statement on p-Values: context, process, and purpose. The American Statistician 70(2): 129-133.
- Retire statistical significance. Amrhein.pdf |Amrhein V.; Greenland S.; McShane B. et al. 2019. Nature 567:305-307.
- Field effects studies in the Chernobyl exclusion zone:lessons to be learnt. Beresford N.A.; Scott E.M. & Copplestone D.2020. Journal of Environmental Radioactivity 211 105893.

Bicaudal e Unicaudal

A outra pergunta simples que poderíamos fazer sobre esses dados é:

- **Existe diferença entre machos e fêmeas?**

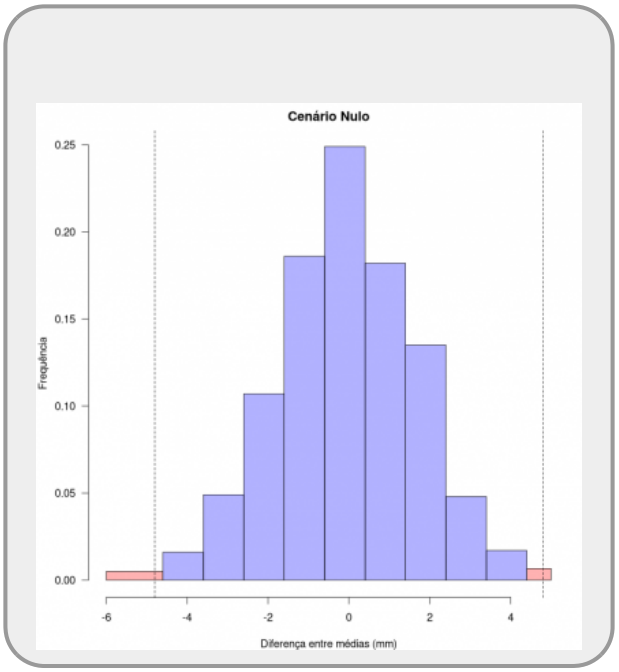
Neste caso, não estamos preocupados se a diferença é macho maior que fêmea ou fêmea maior que macho. Apenas nos perguntamos se há diferença entre os sexos. Nesse caso, o calculo do `p-valor` é diferente, pois podemos encontrar diferenças para qualquer dos dois lados da nossa distribuição de

diferenças. Vamos calcular o `p-valor` para esse caso, para tanto, precisamos apenas fazer a mesma operação com os valores absolutos.

```
sum(Mod(cenaNula) >= abs(difsex))
sum(Mod(cenaNula) >= abs(difsex))/length(cenaNula)
```

O resultado deve ser por volta de 6 simulações com valores iguais ou maiores que `4.8`. O que significa um $p\text{-value} \approx 0.006$. Apesar de pequeno (`0.6%`), esse valor é o dobro do que encontramos com o teste uni-caudal. Vamos representar esses valores no nosso histograma:

```
cols <- rep(c(rgb(1, 0, 0, 0.3), rgb(0, 0, 1, 0.3), rgb(1, 0, 0, 0.3)),
c(1, 9, 2))
plot(histNull,  main = "Cenário Nulo", xlab = "Diferença entre médias (mm)",
ylab = "Frequência", col = cols)
abline(v = c(difsex, -1* difsex), lty = 2)
savePlot("biNula.png", type = "png")
```



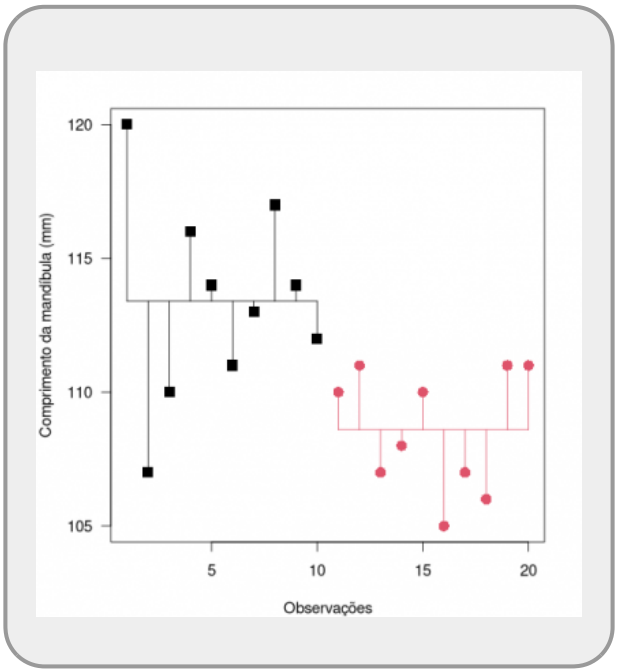
Acabamos de recriar um dos teste mais famosos da estatística frequentista clássica: **o teste t de Student**, apenas usamos a diferença entre médias e não a estatística `t` criada por William Gosset.

O teste t de Gosset

William Gosset trabalhava na cervejaria Guinness em 1907 quando inventou o teste t. Como havia cláusulas no contrato de trabalho que proibiam a publicação de dados da cervejaria, ele usou o pseudônimo de **Student**. Hoje o teste é conhecido como **teste t de Student**. Esse teste está baseada na distribuição da estatística `t`:

A estatística `t` expressa o quanto que a diferença observada é maior que a variação média dos dados. Essa variação média dos dados é similar à média dos seguimentos que apresentamos no gráfico do

início desse tutorial e reproduzido novamente abaixo.



Vamos calcular a estatística `t`.

```
## variancias
(varMacho <- var(macho))
(varFemea <- var(femea))
## residuos medios
(resMF <- sqrt((varMacho/10)+(varFemea/10)))
## statistica t
(tsex  <- difsex/resMF)
```

Vamos agora calcular a probabilidade associada ao teste t, para este valor, usando a função `pt` da mesma forma como usamos a função `pnorm`:

```
pt(tsex, df = 8, lower.tail = FALSE)
```

Podemos também fazer o teste usando a função `t.test` diretamente nos dados:

```
sexoFator <- factor(sexo, levels = c("macho", "femea"))
t.test(chacal ~ sexo, alternative = "greater")
```

As diferenças entre o valor do `t.test` e o `pt` deve-se ao ajuste do teste para amostra pequenas. Perceba como o `df` (graus de liberdade) que o `t.test` retorna é diferente do que colocamos no `pt`. A diferença entre o nosso teste com a diferença entre as médias e o teste t está relacionada a diferença das estatísticas utilizadas, entretanto, os resultados tendem a convergir quando as estatísticas são robustas.



Siga os exercícios 6a e em seguida para o Tutorial de ANOVA.

1)

Manly B. F. J., 2007 Randomization, bootstrap and Monte Carlo methods in biology. 3nd Ed., Chapman and Hall, London

2)

Este é um pressuposto sério e deve ser avaliado para definir o tipo de análise que será usada

3)

pode ser qualquer nome como quando criamos um objeto. Classicamente, em programação utiliza-se `i`, `j`, `ii`, `jj`...

4)

no caso da minha simulação. Esse valor irá variar tendo em vista que o cenário nulo é construido com valores aleatórios, mas ficará próximo de 2 a 5. Além disso, ele tende a convergir para o valor próximo a 3 com o aumento do número de simulações

From:
http://ecor.ib.usp.br/ - **ecoR**

Permanent link:
http://ecor.ib.usp.br/doku.php?id=02_tutoriais:tutorial6:start&rev=1601671321

Last update: **2020/10/02 17:42**