Introdução ao teste de hipóteses

Teste de hipóteses no R

<https://jmsallan.netlify.app/blog/introducing-hypothesis-testing/>

O **teste de hipótese** é uma **técnica de inferência estatística** para adquirir informações sobre um parâmetro populacional a partir de observações de uma amostra, um subconjunto da população.

O teste de hipóteses é amplamente utilizado em muitos campos da pesquisa científica. Quando examinamos se há uma diferença em um parâmetro entre os grupos de tratamento e controle no projeto experimental, ou se existe uma correlação entre duas variáveis, estamos usando o teste de hipóteses.

Os elementos de um fluxo de trabalho de teste de hipótese são:

1. A **hipótese nula** e a **hipótese alternativa**.
2. A **distribuição de probabilidade subjacente** de uma variável aleatória para a qual a hipótese nula é verdadeira.
3. O **nível de significância** para rejeitar a hipótese nula.

Aqui, vou apresentar esses elementos para um teste de hipótese simples, junto com algumas possibilidades do ggplot.

**Hipótese nula e alternativa**

Suponhamos que queremos saber se a média de uma população é diferente de zero. Na hipótese nula, o efeito que estamos procurando não existe, então a média da população será zero:

H0:μ=0

A hipótese alternativa é válida se a hipótese nula não for:

H1:μ≠0

Na maioria dos casos, a hipótese nula é a ausência de efeito, enquanto a hipótese alternativa está relacionada com a presença de efeito. Podemos saber se o efeito existe se pudermos descartar a hipótese nula com um nível de significância.

**Distribuição de probabilidade subjacente**

Não sabemos realmente a média da população μ, pois precisaríamos de acesso a toda a população, algo geralmente impossível ou não prático. Tudo o que podemos obter é uma média da amostra X\_barra de uma amostra de n elementos da população. Como obteremos um valor diferente de X\_barra para cada amostra, X\_barra é uma variável aleatória com uma distribuição de probabilidade. Precisamos saber a distribuição de X\_barra se a hipótese nula for verdadeira. Podemos usar o teorema do limite central para saber que:

X\_barra−μ/ s/√n ∼ N(0,1)

O gráfico abaixo mostra a função de densidade de N(0,1).

library(ggplot2)

ggplot(data.frame(x = c(-4, 4)), aes(x)) +

stat\_function(fun = dnorm) +

annotate(geom = "text", x = 2, y = 0.3, label = "normal N(0,1)", hjust = "left") +

#annotate(geom = "text", x = 2, y = 0.27, label = "t-Student n=8", color = "red", hjust = "left") +

#annotate(geom = "text", x = 2, y = 0.24, label = "t-Student n=5", color = "blue", hjust = "left") +

theme\_bw() +

labs(title = "Distribuição Normal", x = "", y = "")

Nível de significância

Para fazer o teste de hipótese a partir de uma amostra, precisamos calcular o valor:

Quais valores provavelmente assumirão este valor se H0 é verdade? O gráfico abaixo apresenta uma distribuição N (0,1). Os valores da cauda apresentados em azul são improváveis ​​de ocorrer nesta distribuição: na verdade, cada cauda tem uma probabilidade de p = 0,025, então a probabilidade de cair em qualquer uma das duas caudas é p = 0,05.

p <- 0.025

tail\_low <- seq(-4, qnorm(p), 0.01)

df\_tl <- data.frame(x=c(tail\_low,qnorm(p)), y =c(dnorm(tail\_low),0))

tail\_high <- seq(qnorm(1-p), 4, 0.01)

df\_th <- data.frame(x=c(qnorm(1-p),tail\_high), y=c(0,dnorm(tail\_high)))

ggplot(data.frame(x = c(-4, 4)), aes(x)) +

stat\_function(fun = dnorm) +

geom\_polygon(data = df\_tl, aes(x,y), fill = "#99CCFF") +

geom\_polygon(data = df\_th, aes(x,y), fill = "#99CCFF") +

geom\_vline(xintercept = qnorm(p), lty = "dashed", lwd = 0.3) +

geom\_vline(xintercept = qnorm(1-p), lty = "dashed", lwd = 0.3) +

annotate(geom = "text", -1.97, 0.1, label = "-1.96", hjust = "right") +

annotate(geom = "text", 1.97, 0.1, label = "1.96", hjust = "left") +

theme\_classic() +

labs(x="", y="")

Podemos obter os valores finais fazendo:

qnorm(p)

qnorm(1-p)

Temos duas explicações alternativas para uma observação acima de 1,96 ou abaixo de -1,96:

1. É uma amostra pouco frequente de uma população para a qual a hipótese nula é verdadeira. Valores ​​fora do intervalo [-1,96,1,96] pode ocorrer com p = 0,05

2. A hipótese nula é falsa.

Este é o significado do valor p:

*O valor p é a probabilidade de observar o valor obtido ou mais um extremo se a hipótese do casco for verdadeira.*

O limite para o valor p rejeitar uma hipótese nula é uma decisão a ser tomada pelo investigador. Um valor limite comum em muitos campos é p = 0,05 , embora em algumas circunstâncias seja aconselhável escolher um valor ainda mais baixo. Portanto, se observarmos um valor de p abaixo de 0,05, rejeitamos a hipótese nula e consideramos que a hipótese alternativa é verdadeira.

**Erros quando a hipótese nula é verdadeira**

Vamos ver o que acontece se realizarmos muitos testes de hipóteses quando a hipótese nula for verdadeira. Usaremos este código para examinar que:

set.seed(1313)

p\_values\_mu0 <- sapply(1:10000, function(x){

sample <- rnorm(n = 100, mean = 0, sd = 1)

f <- t.test(sample, mu = 0)

return(f$p.value)

Este código obtém 10.000 vezes uma amostra de tamanho 100 de uma população de distribuição normal com média zero e desvio padrão um. Em seguida, ele testa se a média da população é zero usando t.test e armazena o valor p obtido no teste. Portanto, temos 10.000 valores p diferentes.

Vejamos uma representação da distribuição dos valores p obtidos:

ggplot(data.frame(p = p\_values\_mu0), aes(p)) +

geom\_histogram(binwidth = 0.05, center = 0.025, fill = c("#FF9999", rep("#C0C0C0", 19)), color = "#606060") +

theme\_bw() +

scale\_x\_continuous(name="p-value", breaks = seq(0, 1, 0.1))

A distribuição de p é uniforme: se a hipótese nula for verdadeira, qualquer valor de p tem a mesma probabilidade. A barra em vermelho apresenta as situações em que recusamos a hipótese nula com um nível de significância de 0,05. Nessas situações, estamos cometendo um erro Tipo I.

Erros quando a hipótese nula é falsa

Vamos fazer uma experiência um pouco diferente:

p\_values\_mu1 <- sapply(1:10000, function(x){

sample <- rnorm(n = 100, mean = 0.2, sd = 1)

f <- t.test(sample, mu = 0)

return(f$p.value)

})

Agora, a média da população não é zero, então sabemos que a hipótese nula de que a média é zero é falsa. Vamos ver quais valores p obtemos:

ggplot(data.frame(p = p\_values\_mu1), aes(p)) +

geom\_histogram(binwidth = 0.05, center = 0.025, fill = c("#C0C0C0", rep("#FF9999", 19)), color = "#606060") +

theme\_bw() +

scale\_x\_continuous(name="p-value", breaks = seq(0, 1, 0.1), limits = c(0,1))

As barras vermelhas são de valores de p acima do nível de significância de 0,05. Lá estamos aceitando a hipótese nula quando ela é falsa. Este é um erro do Tipo II. Observamos que a probabilidade desse erro é bastante alta, em torno de 0,5.

Podemos melhorar o poder estatístico deste teste aumentando o tamanho da amostra de 100 para 1000:

p\_values\_mu12 <- sapply(1:10000, function(x){

sample <- rnorm(n = 1000, mean = 0.2, sd = 1)

f <- t.test(sample, mu = 0)

return(f$p.value)

})

Agora que o tamanho da amostra é 1000, a probabilidade de erro Tipo II foi reduzida para quase zero:

ggplot(data.frame(p = p\_values\_mu12), aes(p)) +

geom\_histogram(binwidth = 0.05, center = 0.025, fill = c("#C0C0C0", rep("#FF9999", 19)), color = "#606060") +

theme\_bw() +

scale\_x\_continuous(name="p value", breaks = seq(0, 1, 0.1), limits = c(0,1))

É importante ter em mente que, quando não rejeitamos a hipótese nula, não podemos ter certeza se ela é verdadeira ou não. Pode ser que a hipótese nula seja verdadeira e não haja nenhum efeito, ou que o efeito seja muito pequeno para ser observado com o poder estatístico de nossa amostra.

**Erros tipo I e tipo II em testes de hipótese**

O teste de hipóteses é uma ferramenta importante na pesquisa científica quantitativa. O mau uso do teste de hipóteses, aliado aos incentivos do mercado de trabalho de pesquisa, está no cerne da crise de reprodutibilidade da pesquisa científica baseada em inferência estatística. É de grande importância que qualquer cientista entenda o teste de hipóteses corretamente.

**Existem dois tipos diferentes de erro no teste de hipótese:**

Cometemos um erro Tipo I quando rejeitamos a hipótese nula quando ela é verdadeira. Isso significa que estamos detectando um efeito quando ele não está presente.

Cometemos um erro do Tipo II quando não rejeitamos a hipótese nula quando ela é falsa. Isso significa que não estamos detectando um efeito quando ele está presente. Podemos reduzir a probabilidade de erro Tipo II aumentando o poder estatístico do experimento, o que geralmente significa aumentar o tamanho da amostra.