**Exemplo de teste de hipóteses com a distribuição normal (gaussiana)**

**Teste de observações individuais**

Suponha uma distribuição normal de valores de frequência fundamental em enunciados de um falante do português brasileiro com média μ = 130 Hz e desvio-padrão σ = 30 Hz em situação de leitura. Suponha que se faça outro experimento com o mesmo falante e se encontre, num determinado ponto de sua fala um valor de frequência fundamental de 200 Hz. Deseja-se fazer um teste de hipóteses para avaliar as chances de o novo valor ser associado à primeira distribuição.

Em virtude de termos uma ideia de valor extremo à direita, montamos o seguinte conjunto de hipóteses:

H0: μ = 130

Ha: μ > 130

α = 0,05

Para tomar a decisão por uma das hipóteses acima faça:

prob <- pnorm (200, 130, 30,lower.tail=F)

Esse comando calcula a probabilidade de se terem valores maiores do que 200 Hz (não há sentido o valor igual na distribuição normal) obtidos a partir da distribuição normal de média 130 Hz. Observe que o comando ‘lower.tail=F’ toma a probabilidade à direita de 200 Hz, e não à esquerda!

O valor de **0.009815329**, menor que o nível de significância de 5% informa que a hipótese nula precisa ser rejeitada Assim, a probabilidade de cometer um erro (do tipo I) é de aproximadamente 1% (~0,01).

O mesmo se obtém usando a função qnorm, que informa o intervalo de confiança dos valores individuais:

limiar <- qnorm(0.05, 130, 30, lower.tail=F)

O resultado de **179.3456 Hz** revela que a partir desse valor a probabilidade de ocorrerem observações é menor que 5%. Assim, rejeita-se a hipótese nula, pois 200 Hz está à direita daquele valor (200 > 179.3456).

Se não se há certeza sobre a direção da variação, o melhor é montar o seguinte esquema de hipóteses (observe o sinal de diferente, indicando que pode ser maior ou menor que o valor da hipótese nula):

H0: μ = 130

Ha: μ ≠ 130

α = 0,05

É preciso calcular as probabilidades à esquerda e à direita da média. Façamos pelo cálculo do intervalo de confiança a 95%, isto é deixando os 5% de probabilidade restante para as pontas da distribuição:

limiaresq <- qnorm(0.025, 130, 30)

com resultado: 71.2 Hz

e

limiardir <- qnorm(0.025, 130, 30, lower.tail=F)

com resultado: 188.8 Hz

Isto é, o intervalo entre 71 e 189 Hz contem 95% dos valores da distribuição inicial. Isto é, fora do mesmo a chance é igual ou menor que 5%. Como 200 Hz está fora do intervalo, rejeita-se a hipótese nula também nesse caso: 200 Hz não pertence à mesma população que a distribuição inicial.