Testes de Hipóteses

- Objetivo: Validar afirmações
  - sobre parâmetros
    - … de uma população
    - … comparação entre grupos
  - sobre a forma
    - … de uma população
    - … comparação da forma entre grupos

Testes de Hipóteses – Introdução

A fábrica A produz componentes cuja duração média, por questões de controlo de qualidade, não deverá ser inferior a 35 u.t. Admita-se que:

i) o tempo de vida das componentes se distribui de forma Normal e

ii) a variabilidade do tempo de vida é conhecida e igual a 2,2 u.t.

Seis componentes foram testadas, tendo-se registado os valores abaixo.

Estará o processo fora de controlo?

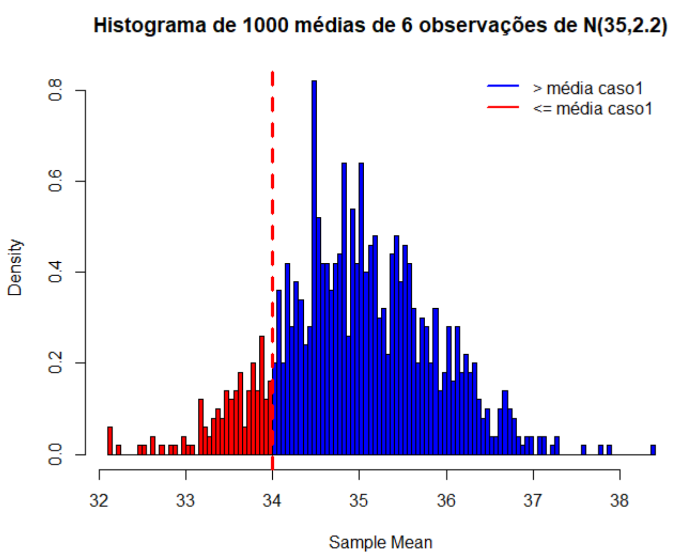
A
32
33
34
37
34
34

Média=34

Será "muito" ou "pouco" provável obter este tipo de amostras (ou outras com médias ainda mais baixas) quando efetivamente μ é 35, sendo $\sigma = 2{,}2$?

HIPÓTESES EM TESTE

H0: $\mu = 35$
H1: $\mu < 35$ (processo fora de controlo)



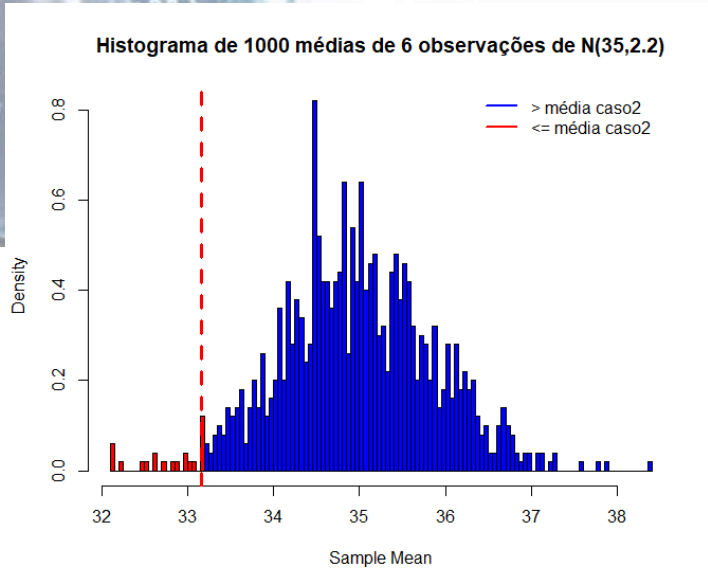

```
> caso1 <- mean(xx <= m_obs1)
> round(caso1,4)
[1] 0.122
```

E nestes outros casos?

Será "muito" ou "pouco" provável obter este tipo de amostras (ou outras com médias ainda mais baixas) quando efetivamente μ é 35, sendo σ = 2,2?

B
32
33
32
31
34
37

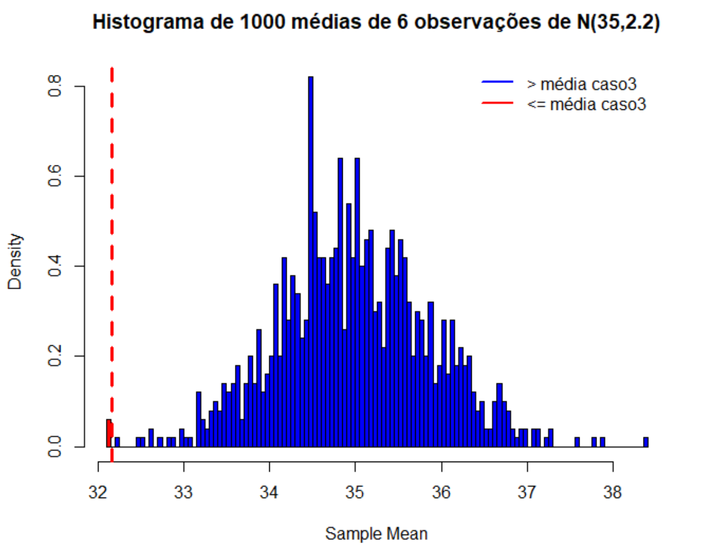
Média=33,167




```
> caso2 <- mean(xx <= m_obs2)
> round(caso2,4)
[1] 0.018
```

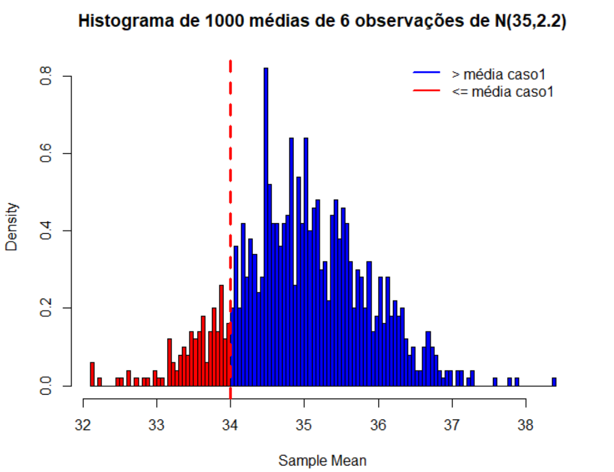
C
32
33
32
31
34
31

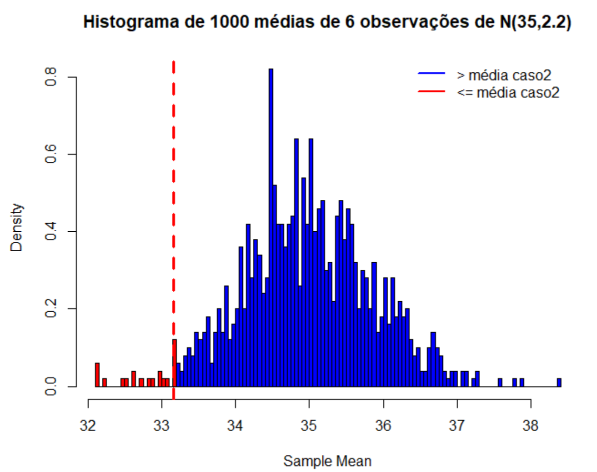
Média=32,167

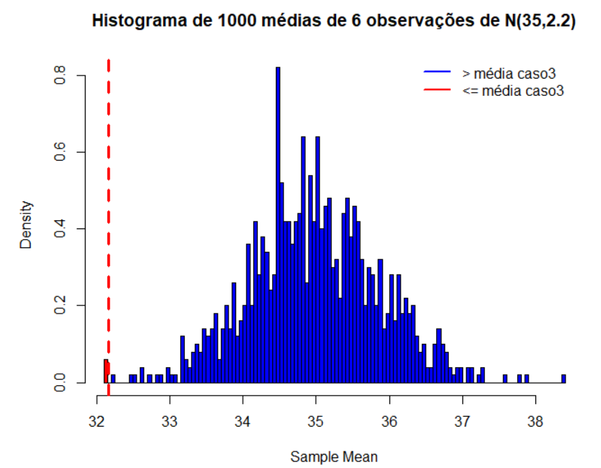



```
> caso3 <- mean(xx <= m_obs3)
> round(caso3,4)
[1] 0.003
```

Conclusões?







Aparentemente, se a média da população for mesmo 35, obter amostras de dimensão 6 com média amostral até 34 (caso 1), não é muito raro… (neste exemplo, aconteceu em cerca de 12% dos casos)

Mas obter amostras de dimensão 6 com média amostral até 33.2 (caso 2) acontece muito menos vezes… (neste exemplo, aconteceu em cerca de 1.8% dos casos)

E obter amostras de dimensão 6 com média amostral até 32.2 (caso 3) é muito pouco frequente… (neste exemplo, aconteceu em cerca de 0.3% dos casos)

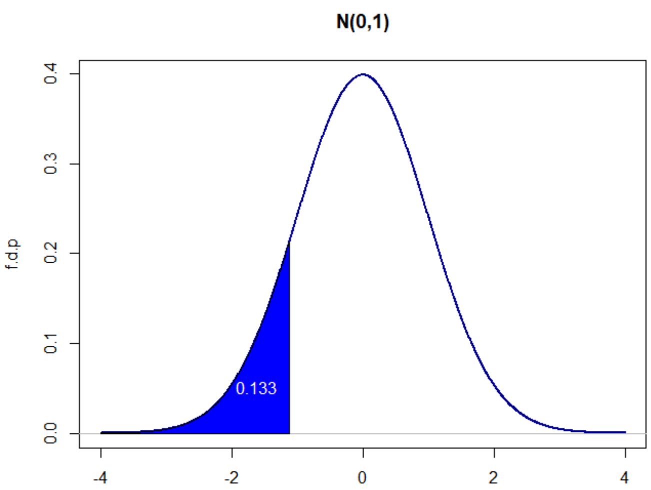
Não deve ser de rejeitar que μ seja 35

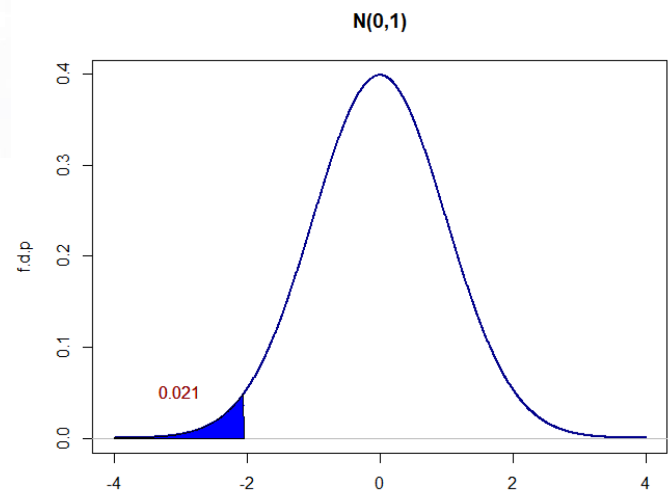
Talvez seja de rejeitar que μ seja 35… ??

Parece que μ não deve ser 35…

Testes de Hipóteses – Introdução

Podemos calcular estes valores teoricamente, já que sabemos que $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

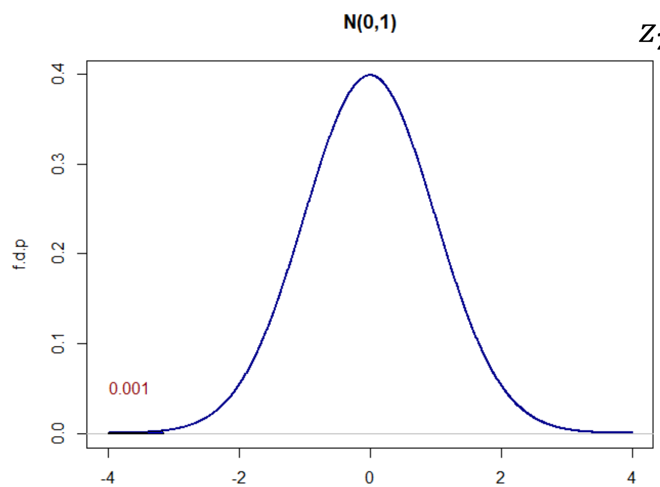




$$z_1 = \frac{34-35}{2.2/\sqrt{6}} = -1.113$$

$$z_2 = \frac{33.17-35}{2.2/\sqrt{6}} = -2.041$$

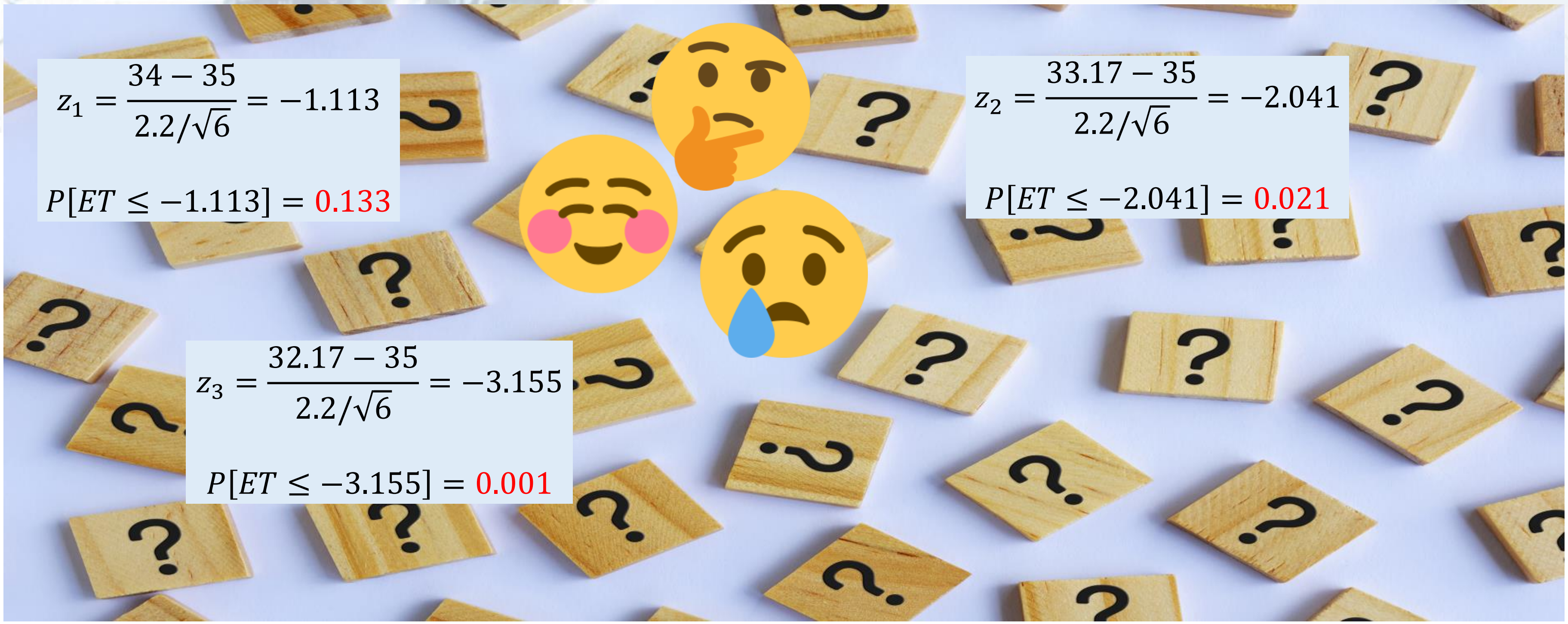
$$z_3 = \frac{32.17-35}{2.2/\sqrt{6}} = -3.155$$



Conclusão em cada uma das situações??

Testes de Hipóteses – Introdução

Como definir em que condições dizemos que os valores obtidos são “grandes” ou são “demasiadamente” pequenos?



$$z_1 = \frac{34 - 35}{2.2/\sqrt{6}} = -1.113$$

$$P[ET \leq -1.113] = 0.133$$

$$z_2 = \frac{33.17 - 35}{2.2/\sqrt{6}} = -2.041$$

$$P[ET \leq -2.041] = 0.021$$

$$z_3 = \frac{32.17 - 35}{2.2/\sqrt{6}} = -3.155$$

$$P[ET \leq -3.155] = 0.001$$

Para respondermos a esta questão vamos ter de introduzir mais alguns conceitos

Testes de Hipóteses

- Para construir um ensaio de hipóteses, é preciso, em 1º lugar, **estabelecer as hipóteses em teste**.
- Existe sempre:
  - Uma hipótese, dita **hipótese nula, H0**, que estabelece pelo menos um valor específico para o parâmetro (ou para a forma)
    - (ou seja, **tem** de conter uma igualdade, podendo ser =, <= ou >=)
  - Uma hipótese, dita **hipótese alternativa, H1**, que contradiz a anterior

No exemplo, recorde-se o ponto de partida:

"A fábrica A produz componentes cuja duração média, por questões de controlo de qualidade, não deverá ser inferior a 35 u.t"

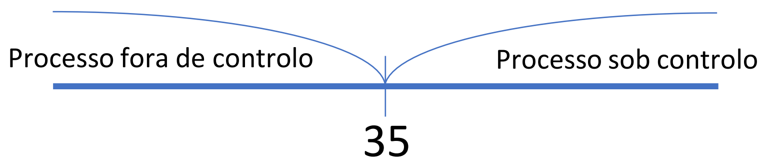
Então o processo está

i) sob controlo, se a duração média for superior ou igual a 35 ($\mu \geq 35$); e

ii) fora de controlo **se a duração média for inferior a 35 u.t,** ($\mu < 35$)

E as hipóteses em teste serão:

H0: $\mu \geq 35$	hipótese nula
H1: $\mu < 35$	hipótese alternativa



Testes de hipóteses

H0: $\mu \geq 35$ hipótese nula
H1: $\mu < 35$ hipótese alternativa

- Quando recolhemos dados, vamos confrontá-los com as hipóteses colocadas
- Sabemos que, numa amostra, a média $\bar{x}$ não vai ser igual à “verdadeira média”, ou seja, a média populacional, μ.
  - É preciso tentar ver se a diferença observada é ou não “credível” se a hipótese nula for verdadeira
    - Parece algo credível observarmos a amostra 1, mas o mesmo não se passa com as amostras 2 ou 3…

No fundo, tentamos avaliar em que medida os dados são ou não contraditórios com a afirmação colocada na hipótese nula.

Se decidirmos que os dados contradizem a H0, então REJEITAMOS H0

Se decidirmos que os dados NÃO contradizem a H0, então NÃO REJEITAMOS H0

Em qualquer dos casos, corremos sempre o risco de **errar**…

Erros nos ensaios de hipóteses

- Em qualquer dos casos, corremos sempre o risco de errar…

Podemos, por exemplo, decidir que uma diferença é “demasiadamente grande” para ser devida ao acaso, e decidir rejeitar H0… mas acontecer que esta até é verdadeira…

Rejeitar H0 | H0 verdadeira

Erro tipo I

No exemplo, concluir que a média populacional é inferior a 35, ou seja, que o processo está fora de controlo, quando na realidade isso não acontece

Podemos, por outro lado, decidir que uma diferença é “pequena” e pode ser devida ao acaso e consequentemente não rejeitar a H0, mas acontecer que ela é efetivamente falsa…

Não Rejeitar H0 | H0 falsa

Erro tipo II

No exemplo, concluir que a média populacional não é inferior a 35, ou seja, que o processo parece estar sob controlo, quando na realidade isso não acontece

Erros nos ensaios de hipóteses

- H0: O réu é inocente
  H1: O réu é culpado

Decisão baseada nas provas	**Situação real** **H0 é verdadeira** (o réu é mesmo inocente)	**H0 é falsa** (o réu é realmente culpado)
Não rejeitar H0 (réu não é considerado culpado, sendo absolvido)	O juiz tomou uma **<u>decisão correta</u>**	O juiz tomou uma decisão incorreta: considerou inocente um réu que, na verdade, era culpado **<u>Erro tipo II</u>** <u>(Não rejeitar H0 | H0 falsa)</u>
Rejeitar H0 (réu é considerado culpado, sendo condenado)	O juiz tomou uma decisão incorreta: considerou culpado um réu que, na verdade, era inocente **<u>Erro tipo I</u>** (Rejeitar H0 | H0 verdadeira)	O juiz tomou uma **<u>decisão correta</u>**

Erros nos ensaios de hipóteses

- Estes erros variam inversamente, ou seja, quando aumenta o erro tipo I diminui o erro tipo II, e vice-versa
- Nos procedimentos estatísticos a utilizar, a probabilidade máxima de ocorrência do erro tipo I é fixada a priori
  (ou seja, controla-se a probabilidade de rejeitar indevidamente a hipótese nula)
- Esta probabilidade,
  P[Rejeitar H0 | H0 verdadeira],
  chama-se significância e designa-se por α.
- Uma vez escolhido o ensaio, a probabilidade de ocorrência do erro tipo II, β, é a menor **possível**, para o nível de significância escolhido, α (e para esse ensaio)

Testes de Hipóteses

- Para construir um ensaio de hipóteses, é preciso, em 2º lugar, **fixar α, nível de significância de referência**, ou seja, a probabilidade de erro tipo I.

No exemplo, admita-se $\alpha = 0{,}05 = P[Rejeitar\ H_0 | H_0\ verdadeira]$

Recordando as hipóteses

H0: $\mu \geq 35$ hipótese nula

H1: $\mu < 35$ hipótese alternativa → **Teste unilateral esquerdo**

- E definir a regra de decisão, tradicionalmente definindo duas regiões:
- A **região CRÍTICA**, RC: caso a Estatística de teste pertença à RC rejeitar-se-á a H0
- A **região NÃO CRÍTICA**, RNC: caso a Estatística de teste pertença à RNC, H0 não será rejeitada

- A regra de decisão pode ser construída tendo por base a probabilidade de Estatística de teste assumir um valor tão ou mais extremo que o observado na amostra concreta recolhida, e que se chama valor-p (***p-value***):
- Se p-value ≤ α, rejeita-se H0
- Se p-value > α, não se rejeita H0

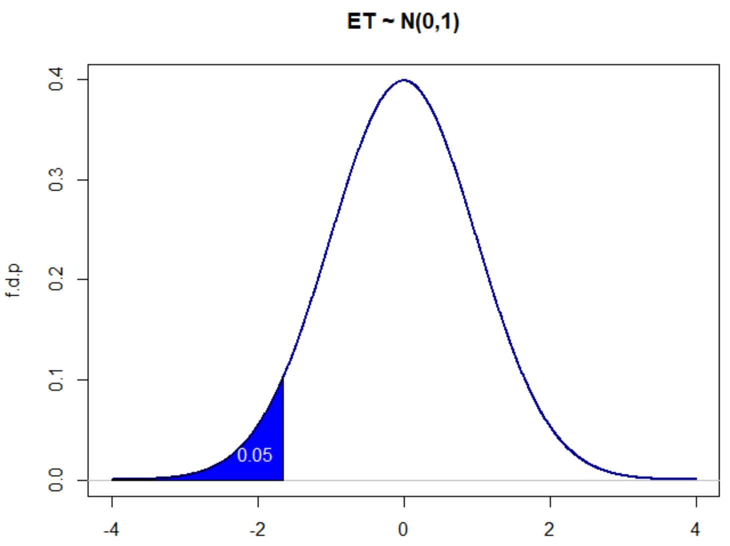
Testes de Hipóteses

No exemplo, admita-se $\alpha = 0{,}05 = P[Rejeitar\ H_0|H_0\ verdadeira]$

Recordando as hipóteses

H0: $\mu \geq 35$ hipótese nula

H1: $\mu < 35$ hipótese alternativa



Como o teste é unilateral esquerdo, o ponto fronteira entre RC e RNC é o quantil de probabilidade $\alpha = 0{,}05$ de uma normal standard, $z_{crit} = -1.645$

Temos então $RC =]-\infty, -1.645]$ e $RNC =]-1.645\ +\infty[$

Se tivermos observado a amostra 1, $z_1 = -1.113 \in RNC$ e não rejeitamos H0, para $\alpha = 0{,}05$

Se tivermos observado a amostra 2, $z_2 = -2{,}041 \in RC$ e rejeitamos H0, para $\alpha = 0{,}05$

Se tivermos observado a amostra 3, $z_3 = -3{,}155 \in RC$ e rejeitamos H0, para $\alpha = 0{,}05$

OU (regra alternativa)

$pvalue_1 = 0.133 > \alpha = 0{,}05$, logo não rejeitamos H0, para $\alpha = 0{,}05$

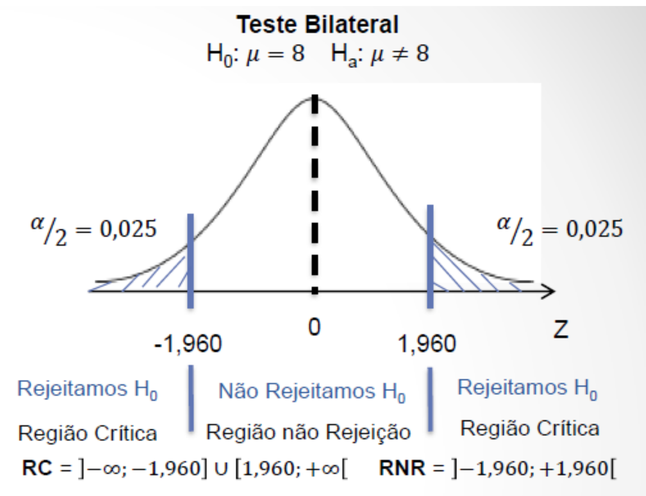
$pvalue_2 = 0.021 \leq \alpha = 0{,}05$, logo rejeitamos H0, para $\alpha = 0{,}05$

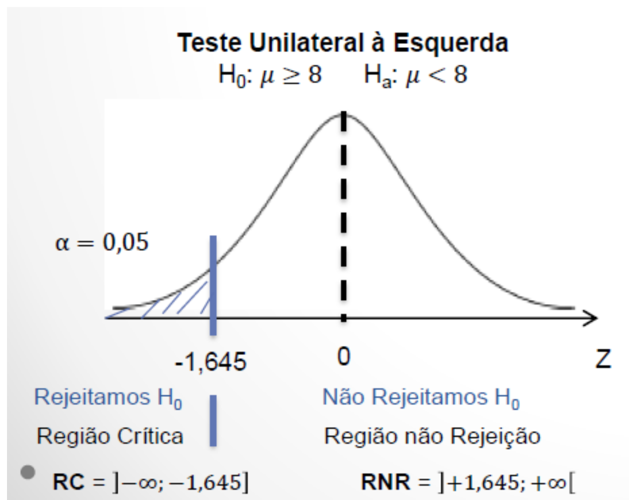
$pvalue_3 = 0.001 \leq \alpha = 0{,}05$, logo rejeitamos H0, para $\alpha = 0{,}05$

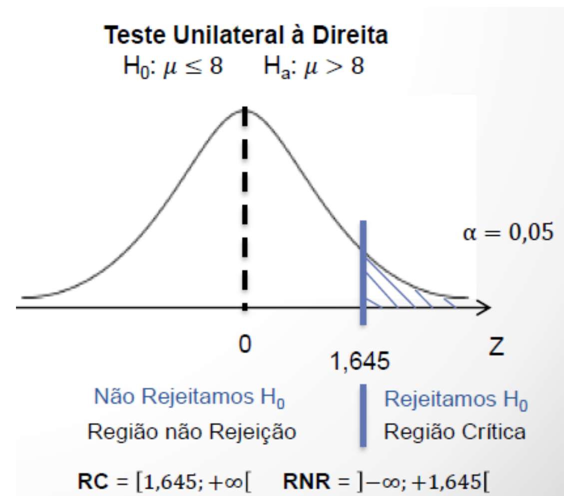
Tipos de testes de hipóteses

Genericamente, classificamos os testes de acordo com o posicionamento dos valores indicados na hipótese alternativa.

Tendo como exemplo uma população Normal e uma significância (normalmente representado por um α) de 0.05:







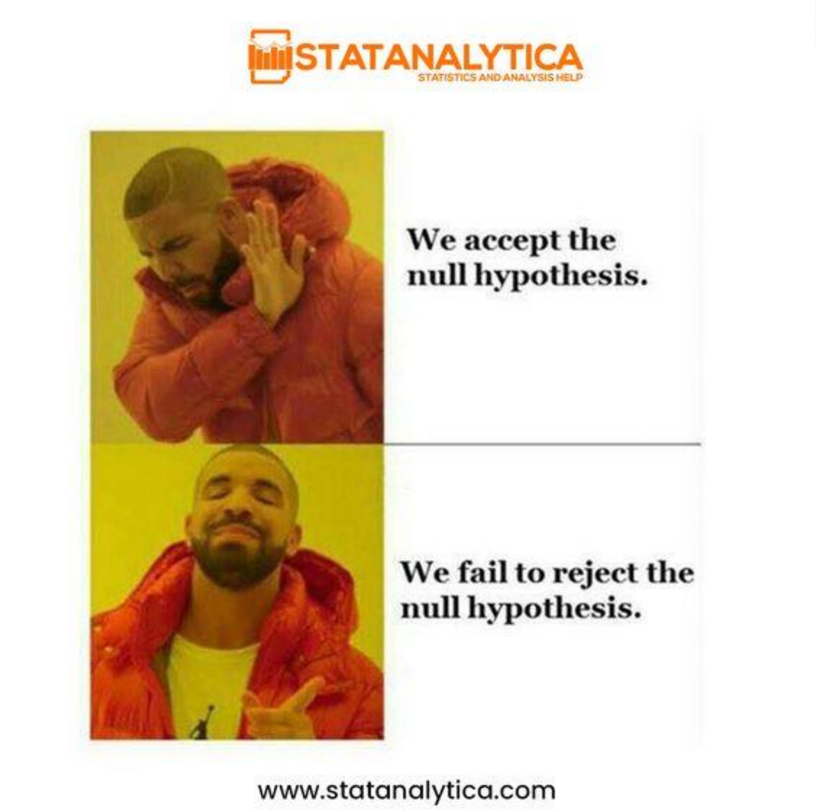
- A igualdade está sempre na hipótese nula!!

- A significância tem relação com o nível de confiança, conceito apresentado no conjunto de slides anteriores (intervalos de confiança).

Testes de Hipóteses - ETAPAS

1. Escrever as hipóteses;
2. Escolher o teste adequado;
3. Estipular o erro máximo que nos permitimos correr ou por outras palavras, através do nível de significância definir a região crítica e não crítica;
4. Calcular e verificar se o valor do teste(t) está na região crítica ou não
5. Tomar a decisão (sem esquecer que nunca se afirma que se aceita a H0: ou a rejeitamos, ou não a rejeitamos)

Em vez de 4, verificar se a probabilidade de obter um valor tão ou mais extremo que o obtido para o valor do teste (t), ou seja, o valor-p ou *p-value*, é **inferior ou igual (H0 rejeitada)** ou **superior (H0 não rejeitada)** ao nível de significância .



Estatísticas de Teste

Parâmetro a testar	Tipo de populações	Conhece-se σ_1^2 e σ_2^2?	Estatística de teste	Distribuição amostral
μ	Normal (qualquer n) ou Qualquer (TLC, n grande)	Sim	$\frac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$	$N(0,1)$
μ	Normal (qualquer n) ou Qualquer (TLC, n grande)	Não	$\frac{\bar{X}-\mu_0}{\frac{S'}{\sqrt{n}}}$	$t_{(n-1)}$ Se n grande, pode usar-se $N(0,1)$
$\mu_1-\mu_2$	Normais	Sim	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	$N(0,1)$
$\mu_1-\mu_2$	Normais (qualquer n_1 e n_2) ou Quaisquer (TLC, n_1 e $n_2>30$)	Não, e assume-se $\sigma_1^2=\sigma_2^2$	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)_0}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{(n_1-1)s_1'^2+(n_2-1)s_2'^2}{n_1+n_2-2}}}$	$t_{(n_1+n_2-2)}$
$\mu_1-\mu_2$	Normais (qualquer n_1 e n_2) ou Quaisquer (TLC, n_1 e $n_2>30$)	Não, e $\sigma_1^2\neq\sigma_2^2$	$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)_0}{\sqrt{\frac{s_1'^2}{n_1}+\frac{s_2'^2}{n_2}}}$	$t_{(\nu)}$ com ν dado por $\frac{\left(\frac{s_1'^2}{n_1}+\frac{s_2'^2}{n_2}\right)^2}{\frac{s_1'^4}{n_1(n_1+1)}+\frac{s_2'^2}{n_2(n_2+1)}}-2$ Se n grande, pode usar-se $N(0,1)$

Estatísticas de teste

Parâmetro a testar	Tipo de população	Dimensão da amostra	Variável fulcral	Distribuição amostral
$\boldsymbol{p}$	Bernoulli	$n > 30$	$\dfrac{\bar{X} - \mathrm{p}}{\sqrt{\mathrm{p}(1-p)/n}}$	$\dot{\sim} N(0,1)$
$\boldsymbol{p_1 - p_2}$	Bernoulli	$n_1 > 30$ e $n_2 > 30$	$\dfrac{(\bar{X}_1 - \bar{X}_2) - (p_1 - p_2)_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$, com $\hat{p} = \dfrac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1 + n_2}$	$\dot{\sim} N(0,1)$
$\boldsymbol{\sigma^2}$	Normal	qualquer	$\dfrac{(n-1)S'^2}{\sigma_0^2}$	$\chi^2_{(n-1)}$
$\boldsymbol{\dfrac{\sigma_1^2}{\sigma_2^2}}$	Normal	qualquer	$\dfrac{S_1^2}{S_2^2}\left(\dfrac{\sigma_2^2}{\sigma_1^2}\right)_0$	$F_{(n_1-1;n_2-1)}$