Teste de Hipótese Continuação – Aula 05

Statistics for Business and Economics 7 edição, by Paul Newbold , William Carlson , Betty Thorne (cap. Hypothesis Testing: Single Population)

Bussab e Morettin (Cap 13 Teste de Hipótese)

Statistics for Economics, Accounting and Business Studies, capítulo 5, Barrow (Cap. 5 Hypothesis testing)

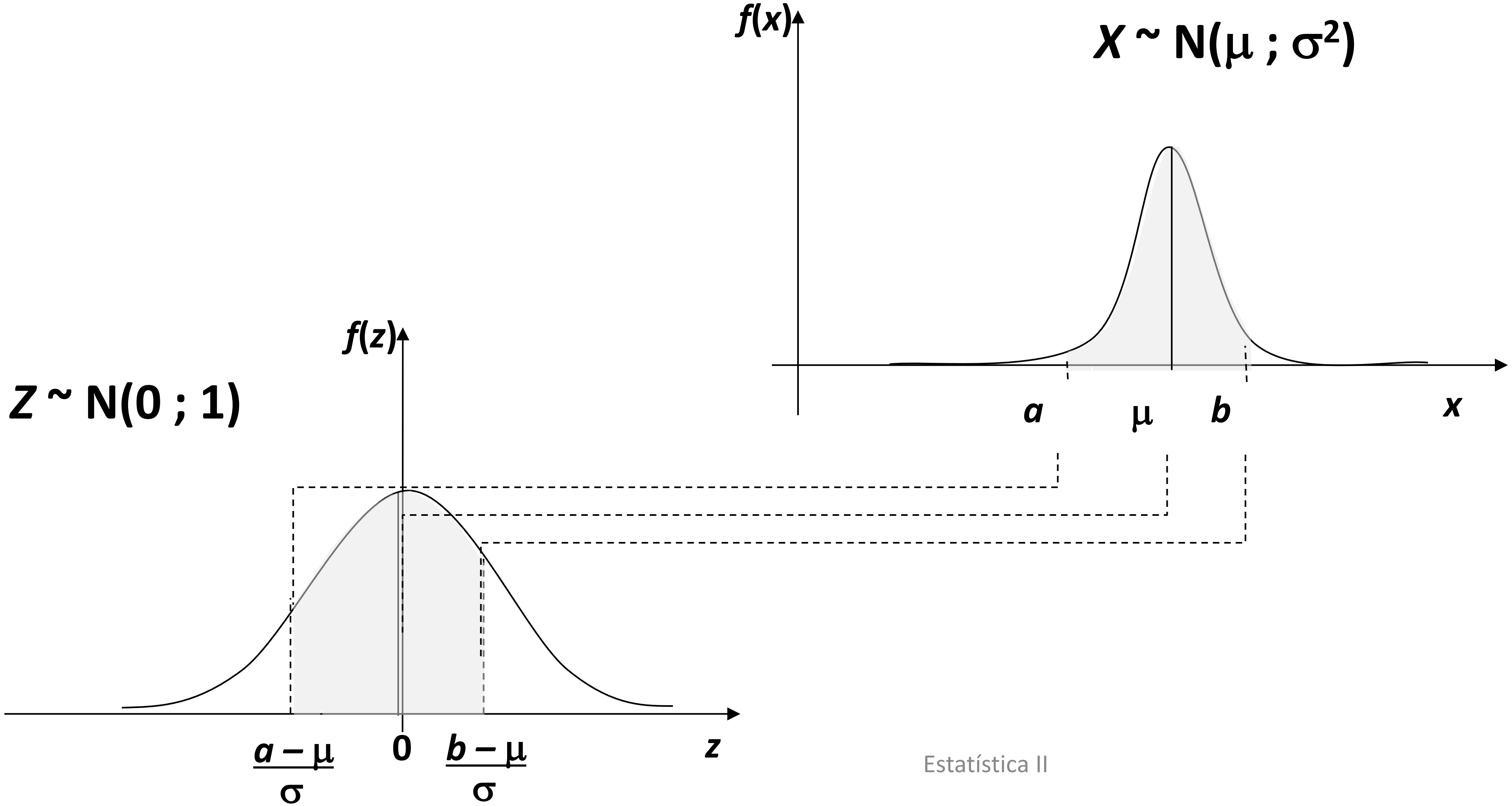
Marislei Nishijima

Se $X \sim N(\mu\,;\sigma^2)$, definimos $\boxed{Z = \dfrac{X - \mu}{\sigma}}$

Lembrando!!!

$E(Z) = 0$

$Var(Z) = 1$



Testes de Duas Amostras

Testes de Duas Amostras

- Média da população, amostras dependentes
- Média da população, amostras independentes
- Proporção da população
- Variâncias Populacionais

Exemplos:

- Mesmo grupo antes e depois do tratamento
- Grupo 1 vs. Grupo 2 independente
- Proporção 1 vs. Proporção 2
- Variância 1 vs. Variância 2

Amostras dependentes

amostras dependentes

Teste de Média para 2 Populações Relacionadas

- Amostras Pareadas ou combinadas
- Medidas repetidas (antes/depois)
- Usar diferença entre valores pareados:

$$\boxed{d_i = x_i - y_i}$$

- Hipóteses:
  - Ambas Populações são Normalmente Distribuídas

Estatística de teste: Dependent Samples

A Estatística de teste para adiferença de média é uma t valor, com

n – 1 graus de liberdade:

amostras dependentes

sendo

$$t = \frac{\bar{d} - D_0}{\frac{s_d}{\sqrt{n}}}$$

$$\bar{d} = \frac{\sum d_i}{n} = \bar{x} - \bar{y}$$

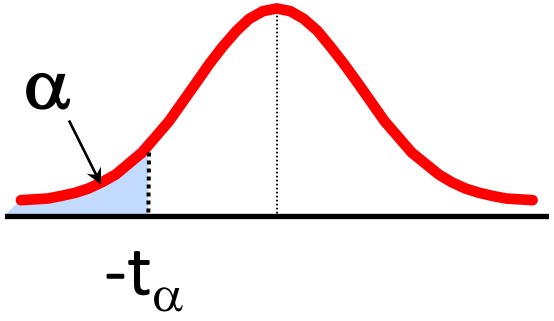
D_0 = Diferença de média suposta
s_d = desvio padrão amostral das diferenças
n = tamanho da amostra (número de pares)

Regra de decisão: Pares combinados

Amostras Pareadas ou combinadas

Teste cauda inferior:

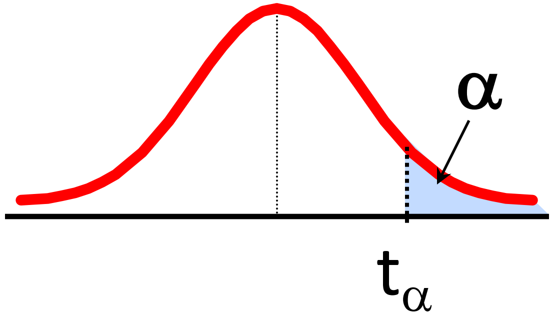
$H_0: \mu_x - \mu_y \geq 0$
$H_1: \mu_x - \mu_y < 0$



Rejeita H_0 se $t < -t_{n-1,\,\alpha}$

Teste cauda superior :

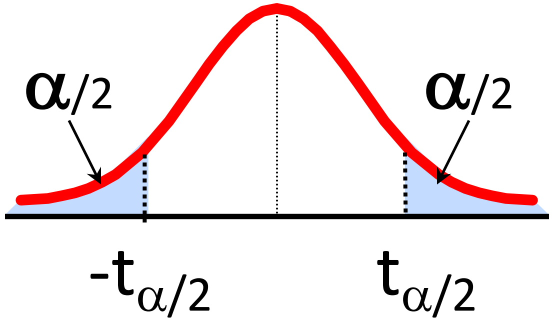
$H_0: \mu_x - \mu_y \leq 0$
$H_1: \mu_x - \mu_y > 0$



Rejeita H_0 se $t > t_{n-1,\,\alpha}$

Teste duas caudas:

$H_0: \mu_x - \mu_y = 0$
$H_1: \mu_x - \mu_y \neq 0$



Rejeita H_0 se $t < -t_{n-1\,,\,\alpha/2}$

ou $t > t_{n-1\,,\,\alpha/2}$

Sendo que $t = \dfrac{\bar{d} - D_0}{\dfrac{s_d}{\sqrt{n}}}$ tem n - 1 g.l.

Exemplo de Pares combinados

Suponha que você envie seus vendedores para um workshop de treinamento de “atendimento ao cliente”. O treinamento fez diferença no número de reclamações? Você coleta os seguintes dados:

Vendedor	Número de reclamações: Antes (1)	Depois (2)	(2) - (1) Diferença, d_i
C.B.	6	4	- 2
T.F.	20	6	-14
M.H.	3	2	- 1
R.K.	0	0	0
M.O.	4	0	- 4
			-21

$$\bar{d} = \frac{\sum d_i}{n}$$

$$= -4.2$$

$$S_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}}$$

$$= 5.67$$

Pares combinados: Solução

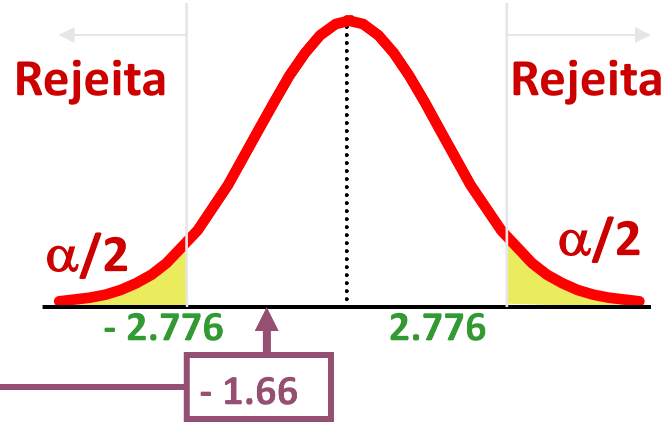
- O treinamento fez diferença no número de reclamações? (ao nível de $\alpha = 0.05$)?

H_0: $\mu_x - \mu_y = 0$
H_1: $\mu_x - \mu_y \neq 0$

$\alpha = .05$ $\bar{d} = -4.2$

Valor crítico = ± 2.776

g.l. = n − 1 = 4



Estatística de Teste:

$$t = \frac{\bar{d} - D_0}{s_d/\sqrt{n}} = \frac{-4.2 - 0}{5.67/\sqrt{5}} = \boxed{-1.66}$$

Decisão: Não rejeita *H_0*
(a estatística de teste não cai na região de rejeição)

Conclusão: Não há mudança significative no número de reclamações.

Diferença Entre Duas Médias

Médias das populações, amostras independentes

Objetivo: Testar a diferença entre duas médias populacionais, $\mu_x - \mu_y$

- Diferentes populações
  - Não relacionadas
  - Independentes
    - A Amostra selecionada de uma população não tem efeito sobre a amostra selecionada na outra população
  - Normalmente distribuída

Diferença Entre Duas Médias

(cont.)

Médias das populações, amostras independentes

- σ_x^2 e σ_y^2 conhecidas → A Estatística de Teste é um **z** valor
- σ_x^2 e σ_y^2 desconhecidas
  - σ_x^2 e σ_y^2 supostas iguais
  - σ_x^2 e σ_y^2 supostas diferentes

  → A Estística de teste é um valor da distribuição **t** de Student

σ_x^2 e σ_y^2 conhecidos

Médias das populações, amostras independentes

- σ_x^2 e σ_y^2 conhecidas *
- σ_x^2 e σ_y^2 desconhecidas

Hipóteses:

As amostras são aleatórias e retiradas de forma independente

Ambas as distribuições de população são normais

As variâncias populacionais são conhecidas

σ_x^2 e σ_y^2 Conhecidos

(cont.)

Médias das populações, amostras independentes

- σ_x^2 e σ_y^2 conhecidas *
- σ_x^2 e σ_y^2 desconhecidas

Quando σ_x^2 e σ_y^2 são conhecidas e ambas normalmente distribuídas, a variância de $\bar{X} - \bar{Y}$ é

$$\sigma^2_{\bar{X}-\bar{Y}} = \frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}$$

…e a variável aleatória Z tem uma distribuição normal padrão

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_x^2}{n_X} + \frac{\sigma_y^2}{n_Y}}}$$

Estatística de Teste, σ_x^2 e σ_y^2 Conhecidas

Médias das populações, amostras independentes

- σ_x^2 e σ_y^2 conhecidas *
- σ_x^2 e σ_y^2 desconhecidas

$$H_0 : \mu_x - \mu_y = D_0$$

A Estatística de Teste pata

$\mu_x - \mu_y$ é:

$$z = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\dfrac{\sigma_x^2}{n_x} + \dfrac{\sigma_y^2}{n_y}}}$$

Teste de Hipótese para Duas Médias Populacionais

Duas Médias Populacionais, amostras independentes

Teste de cauda inferior:

H_0: $\mu_x \geq \mu_y$
H_1: $\mu_x < \mu_y$

i.e.,

H_0: $\mu_x - \mu_y \geq 0$
H_1: $\mu_x - \mu_y < 0$

Teste de cauda superior :

H_0: $\mu_x \leq \mu_y$
H_1: $\mu_x > \mu_y$

i.e.,

H_0: $\mu_x - \mu_y \leq 0$
H_1: $\mu_x - \mu_y > 0$

Teste de duas caudas:

H_0: $\mu_x = \mu_y$
H_1: $\mu_x \neq \mu_y$

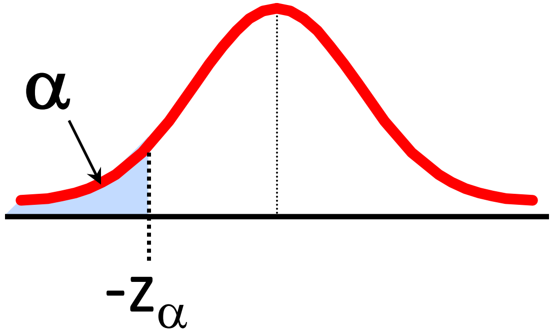
i.e.,

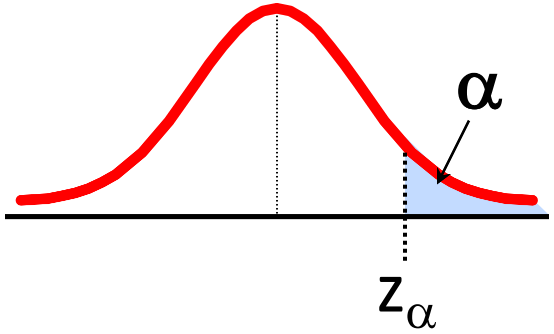
H_0: $\mu_x - \mu_y = 0$
H_1: $\mu_x - \mu_y \neq 0$

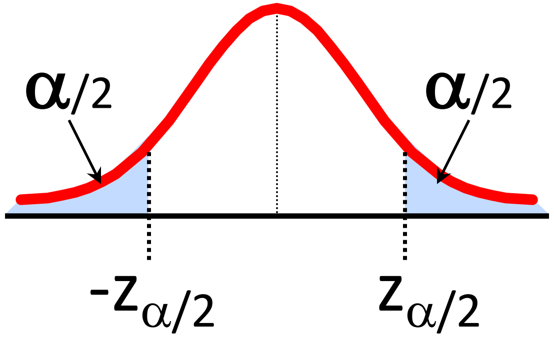
Regras de Decisão

Duas Médias Populacionais, amostras independents, Variâncias conhecidas

Teste cauda inferior:	Teste cauda superior:	Teste duas caudas:
$H_0: \mu_x - \mu_y \geq 0$	$H_0: \mu_x - \mu_y \leq 0$	$H_0: \mu_x - \mu_y = 0$
$H_1: \mu_x - \mu_y < 0$	$H_1: \mu_x - \mu_y > 0$	$H_1: \mu_x - \mu_y \neq 0$







Rejeita H_0 if $z < -z_\alpha$

Rejeita H_0 if $z > z_\alpha$

Rejeita H_0 if $z < -z_{\alpha/2}$

or $z > z_{\alpha/2}$

σ_x^2 e σ_y^2 desconhecidas, Supostas Iguais

Médias das populações, amostras independentes

- σ_x^2 e σ_y^2 conhecidas
- σ_x^2 e σ_y^2 desconhecidas
  - σ_x^2 e σ_y^2 desconhecidas, Supostas Iguais *
  - σ_x^2 e σ_y^2 desconhecidas, Supostas Diferentes

Hipóteses:

As amostras são aleatórias e retiradas de forma independente

Ambas as distribuições de população são normais

As variâncias populacionais são desconhecidas e supostas iguais

σ_x^2 e σ_y^2 desconhecidas, Supostas Iguais

(cont.)

- Médias das populações, amostras independentes
  - σ_x^2 e σ_y^2 conhecidas
  - σ_x^2 e σ_y^2 desconhecidas
    - σ_x^2 e σ_y^2 desconhecidas, Supostas Iguais *
    - σ_x^2 e σ_y^2 desconhecidas, Supostas Diferentes

- As variâncias populacionais são supostas iguais, então use os dois desvio-padrões amostrais agrupados para estimar σ
- usar um t value com $(n_x + n_y - 2)$ graus de liberdade

σ_x^2 e σ_y^2 desconhecidas, Supostas Iguais

σ_x^2 e σ_y^2 desconhecidas

- σ_x^2 e σ_y^2 desconhecidas, Supostas Iguais *
- σ_x^2 e σ_y^2 desconhecidas, Supostas Diferentes

A Estatística de Teste para $\mu_x - \mu_y$ é:

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{s_p^2}{n_x} + \frac{s_p^2}{n_y}}}$$

Sendo que t tem $(n_1 + n_2 - 2)$ g.l., e

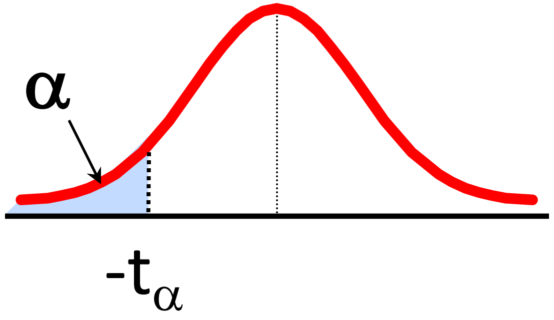
$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

Regras de decisão

Duas Médias Populacionais, amostras independents, Variâncias desconhecidas

Teste cauda inferior:

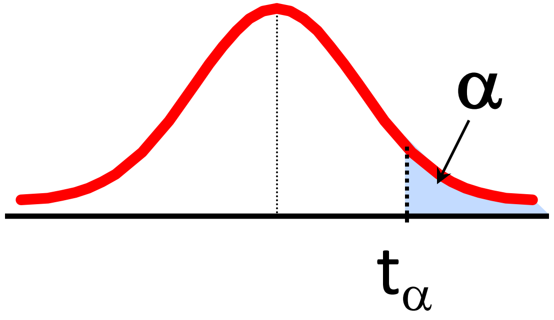
H_0: $\mu_x - \mu_y \geq 0$
H_1: $\mu_x - \mu_y < 0$



Rejeita H_0 se
$t < -t_{(n1+n2-2), \alpha}$

Teste cauda superior:

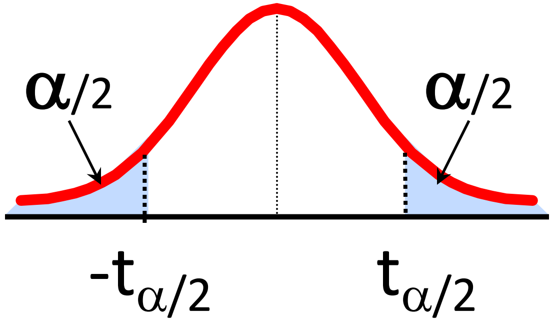
H_0: $\mu_x - \mu_y \leq 0$
H_1: $\mu_x - \mu_y > 0$



Rejeita H_0 se
$t > t_{(n1+n2-2), \alpha}$

Two-tail test:

H_0: $\mu_x - \mu_y = 0$
H_1: $\mu_x - \mu_y \neq 0$



Rejeita H_0 se
$t < -t_{(n1+n2-2), \alpha/2}$ or
$t > t_{(n1+n2-2), \alpha/2}$

Exemplo: Teste t com variância agrupada:

Você é um analista financeiro de uma corretora. Existe uma diferença no rendimento de dividendos entre as ações listadas na NYSE e NASDAQ? Você coleta os seguintes dados:

	NYSE	**NASDAQ**
Número	**21**	**25**
Média amostral	**3.27**	**2.53**
Des.Padr amostral	**1.30**	**1.16**

Supondo que ambas as populações sejam aproximadamente normal com variâncias iguais, há diferença no rendimento médio?
(use α = 0.05)

Calculando a Estatística de Teste

A a Estatística de Teste é:

$$t = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\left(3.27 - 2.53\right) - 0}{\sqrt{1.5021\left(\frac{1}{21} + \frac{1}{25}\right)}} = \boxed{2.040}$$

$$S_p^2 = \frac{\left(n_1 - 1\right)S_1^{\ 2} + \left(n_2 - 1\right)S_2^{\ 2}}{(n_1 - 1) + (n_2 - 1)} = \frac{\left(21 - 1\right)1.30^2 + \left(25 - 1\right)1.16^2}{(21 - 1) + (25 - 1)} = 1.5021$$

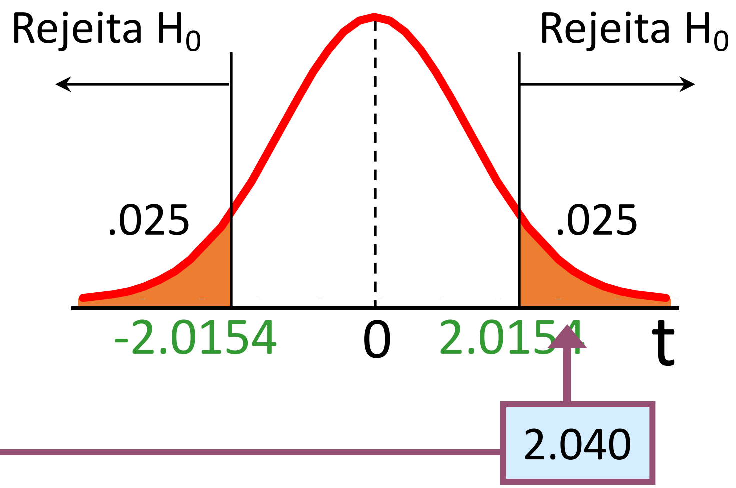
Solução

H_0: $\mu_1 - \mu_2 = 0$ i.e. $(\mu_1 = \mu_2)$
H_1: $\mu_1 - \mu_2 \neq 0$ i.e. $(\mu_1 \neq \mu_2)$

$\alpha = 0.05$

df = 21 + 25 − 2 = 44

Valores Críticos: t = ± 2.0154



Estatística de Teste:

$$t = \frac{3.27 - 2.53}{\sqrt{1.5021\left(\frac{1}{21} + \frac{1}{25}\right)}} = \boxed{2.040}$$

Decisão:

Rejeita H_0 at $\alpha = 0.05$

Conclusão:

Existe evidência de diferença nas médias.

σ_x^2 e σ_y^2 desconhecidas, Supostas Diferentes

- Médias das populações, amostras independentes
  - σ_x^2 e σ_y^2 conhecidas
  - σ_x^2 e σ_y^2 desconhecidas
    - σ_x^2 e σ_y^2 desconhecidas, Supostas Iguais
    - σ_x^2 e σ_y^2 desconhecidas, Supostas Diferentes *

Hipóteses:

As amostras são aleatórias e retiradas de forma independente

Ambas as distribuições de população são normais

As variâncias populacionais são desconhecidas e supostas diferentes

σ_x^2 e σ_y^2 desconhecidas, Supostas Diferentes

(cont.)

Médias das populações, amostras independentes

σ_x^2 e σ_y^2 conhecidas

σ_x^2 e σ_y^2 desconhecidas

σ_x^2 e σ_y^2 supostas iguais

σ_x^2 e σ_y^2 supostas diferentes *

Estimando intervalo:

- As variâncias da população são supostas serem diferentes, portanto, uma variância combinada não é apropriada
- use um t valor com ν Graus de Liberdade, sendo

$$\nu = \frac{\left[\left(\frac{s_x^2}{n_x}\right) + \left(\frac{s_y^2}{n_y}\right)\right]^2}{\left(\frac{s_x^2}{n_x}\right)^2 / (n_x - 1) + \left(\frac{s_y^2}{n_y}\right)^2 / (n_y - 1)}$$

Estatística de Teste, σ_x^2 e σ_y^2 Desconhecidas e Diferentes

σ_x^2 e σ_y^2 desconhecidas

- σ_x^2 e σ_y^2 desconhecidas, Supostas Iguais
- σ_x^2 e σ_y^2 desconhecidas, Supostas Diferentes *

A Estatística de Teste para $\mu_x - \mu_y$ é:

$$t = \frac{(\bar{x} - \bar{y}) - D_0}{\sqrt{\dfrac{s_x^2}{n_X} + \dfrac{s_y^2}{n_Y}}}$$

Sendo que t possui ν graus de liberdade:

$$\nu = \frac{\left[\left(\dfrac{s_x^2}{n_x}\right) + \left(\dfrac{s_y^2}{n_y}\right)\right]^2}{\left(\dfrac{s_x^2}{n_x}\right)^2 / (n_x - 1) + \left(\dfrac{s_y^2}{n_y}\right)^2 / (n_y - 1)}$$

Duas Proporções Populacionais

Proporções Populacionais

Objetivo: Testar hipóteses para a diferença de proporções entre duas populações, $\boxed{P_x - P_y}$

Hipóteses:

Ambas amostras são grandes,

$$nP(1 - P) > 5$$

Duas Proporções Populacionais

(cont.)

Proporções Populacionais

- A variável aleatória

$$Z = \frac{(\hat{p}_x - \hat{p}_y) - (p_x - p_y)}{\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n_x} + \frac{\hat{p}_y(1-\hat{p}_y)}{n_y}}}$$

É aproximadamente normalmente distribuída

Estatística de Teste para Duas Proporções Populacionais

Proporções Populacionais

Estatística de Teste para

$H_0: P_x - P_y = 0$

é um z valor:

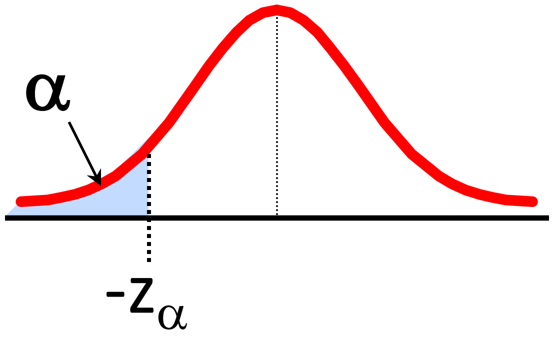
$$z = \frac{(\hat{p}_x - \hat{p}_y)}{\sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_x} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_y}}}$$

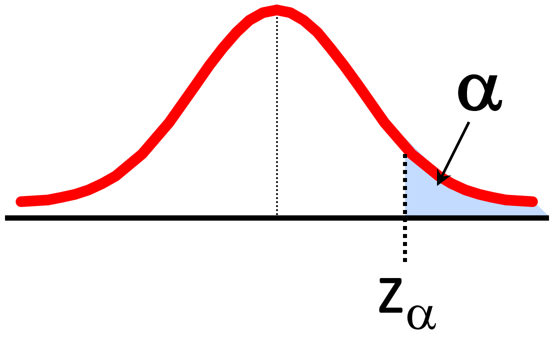
Sendo $\hat{p}_0 = \dfrac{n_x\hat{p}_x + n_y\hat{p}_y}{n_x + n_y}$

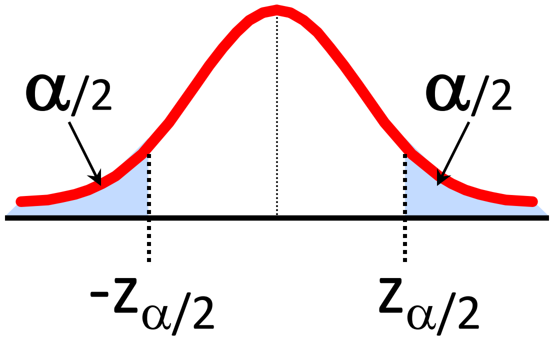
Regras de Decisão: Proporções

Population proportions

Teste cauda inferior:	Teste cauda superior :	Teste de duas caudas:
$H_0: P_x - P_y \geq 0$	$H_0: P_x - P_y \leq 0$	$H_0: P_x - P_y = 0$
$H_1: P_x - P_y < 0$	$H_1: P_x - P_y > 0$	$H_1: P_x - P_y \neq 0$







Rejeita H_0 if $z < -z_{\alpha}$

Rejeita H_0 if $z > z_{\alpha}$

Rejeita H_0 if $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

Exemplo:
Duas Proporções Populacionais

Existe uma diferença significativa entre a proporção de homens e a proporção de mulheres que votarão sim na proposição A?

- Em uma amostra aleatória, 36 de 72 homens e 31 de 50 mulheres indicaram que votariam sim

- Teste usando o nível de significância de 0.05

Exemplo:
Duas Proporções Populacionais

(cont.)

O teste de hipótese é:

H_0: $P_M - P_W = 0$ (as duas Proporções são Iguais)
H_1: $P_M - P_W \neq 0$ (existe uma diferença significativa entre as proporções)

- As proporções amostrais são:

Homens: $\hat{p}_M = \frac{36}{72} = 0.50$

Mulheres: $\hat{p}_W = \frac{31}{50} = 0.62$

- A estimativa para a proporção geral comum é:

$$\hat{p}_0 = \frac{n_M \hat{p}_M + n_W \hat{p}_W}{n_M + n_W} = \frac{72(36/72) + 50(31/50)}{72 + 50} = \frac{67}{122} = .549$$

Exemplo:
Duas Proporções Populacionais

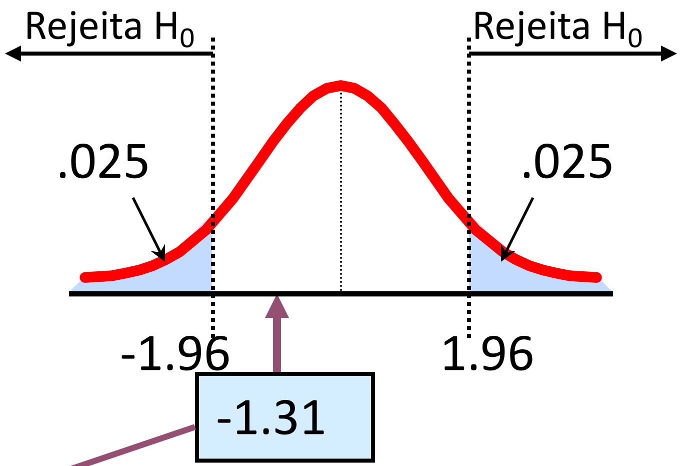
(cont.)

A estatística de teste para $P_M - P_W = 0$ é:

$$z = \frac{(\hat{p}_M - \hat{p}_W)}{\sqrt{\frac{\hat{p}_0(1-\hat{p}_0)}{n_1} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_2}}}$$

$$= \frac{(.50 - .62)}{\sqrt{\left(\frac{.549(1-.549)}{72} + \frac{.549(1-.549)}{50}\right)}}$$

$$= \boxed{-1.31}$$



Valores Críticos= ±1.96
Para α = 0.05

Decisão: Não rejeita H_0

Conclusão: Não há evidência de diferença significativa entre homens e mulheres sobre as proporções em que votarão sim para a proposta.

Teste de Hipótese testando Variâncias Diferentes (Pode-se testar qualquer parâmetro populacional)

Teste de Hipótese com Variâncias Diferentes

- Objetivo: Teste de hipótese sobre variâncias populationais

$H_0: \sigma_x^2 \geq \sigma_y^2$
$H_1: \sigma_x^2 < \sigma_y^2$

Teste de cauda inferior

$H_0: \sigma_x^2 \leq \sigma_y^2$
$H_1: \sigma_x^2 > \sigma_y^2$

Teste de cauda superior

$H_0: \sigma_x^2 = \sigma_y^2$
$H_1: \sigma_x^2 \neq \sigma_y^2$

Teste de duas caudas

As duas populações são consideradas independentes e normalmente distribuídas

Teste de Hipótese testando Variâncias Diferentes

(cont.)

Teste de Hipótese testando Variâncias Diferentes

A variável aleatória tem

$$F = \frac{s_x^2/\sigma_x^2}{s_y^2/\sigma_y^2}$$

Uma distribuição F (Fisher) com $(n_x - 1)$ graus de Liberdade no numerador e $(n_y - 1)$ no denominator.

Denota um valor F com ν_1 Graus de Liberdade no numerador e ν_2 no denominador por F_{ν_1,ν_2}

Estatística de Teste

Teste para
Duas variâncias
populacionais

Estatística de Teste F

A estatística de Teste para um teste de hipótese sobre duas variações populacionais é

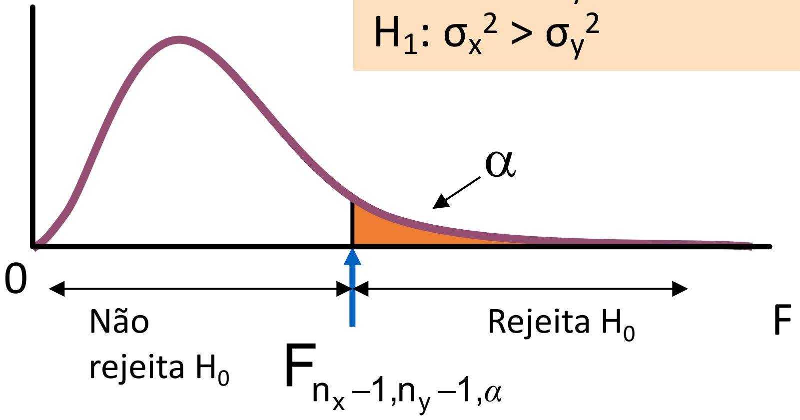
$$F = \frac{s_x^2}{s_y^2}$$

Sendo que F tem $(n_x - 1)$ graus de Liberdade no numerador e $(n_y - 1)$ no denominator.

Regra de decisão: Duas Variâncias

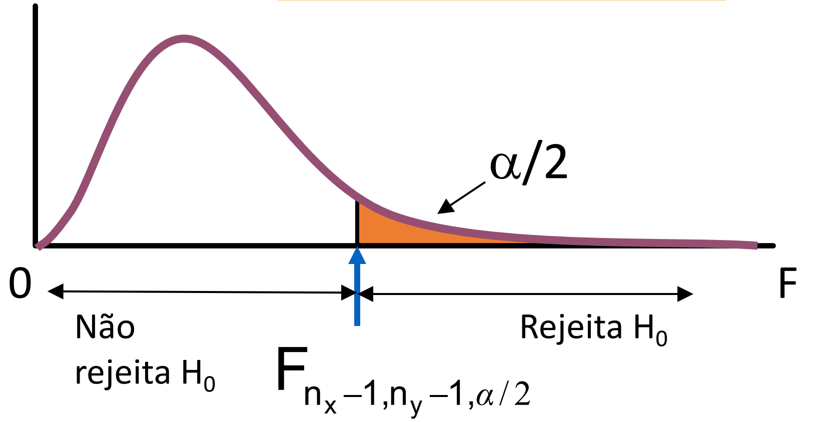
Use s_x^2 para descrever a maior variância.

$H_0: \sigma_x^2 \leq \sigma_y^2$
$H_1: \sigma_x^2 > \sigma_y^2$



$$\boxed{\text{Reject } H_0 \text{ if } F > F_{n_x-1,n_y-1,\alpha}}$$

$H_0: \sigma_x^2 = \sigma_y^2$
$H_1: \sigma_x^2 \neq \sigma_y^2$



- Região de rejeição para um teste de duas caudas é:

$$\boxed{\text{Reject } H_0 \text{ if } F > F_{n_x-1,n_y-1,\alpha/2}}$$

Em que s_x^2 é maior das duas variâncias

Exemplo: Teste F

Você é um analista financeiro de uma corretora. Você deseja comparar rendimentos de dividendos entre ações listadas na NYSE e NASDAQ. Você coleta os seguintes dados:

	NYSE	NASDAQ
Número	**21**	**25**
Média	**3.27**	**2.53**
Des. Pad.	**1.30**	**1.16**

Existe uma diferença nas variâncias entre a NYSE e NASDAQ à $\alpha = 0.10$?

Teste F : Exemplo Solução

- Formule o teste de hipótese:

$H_0: \sigma_x^2 = \sigma_y^2$ (não há diferença entre variâncias)

$H_1: \sigma_x^2 \neq \sigma_y^2$ (há diferença entre variâncias)

- Encontre o valor crítico de F para α = .10/2:

Graus de Liberdade:

- Numerador

  (NYSE tem maior desvio padrão):

  - $n_x - 1 = 21 - 1 = 20$ d.f.

- Denominador:

  - $n_y - 1 = 25 - 1 = 24$ d.f.

$$F_{n_x-1,\, n_y-1,\, \alpha/2}$$

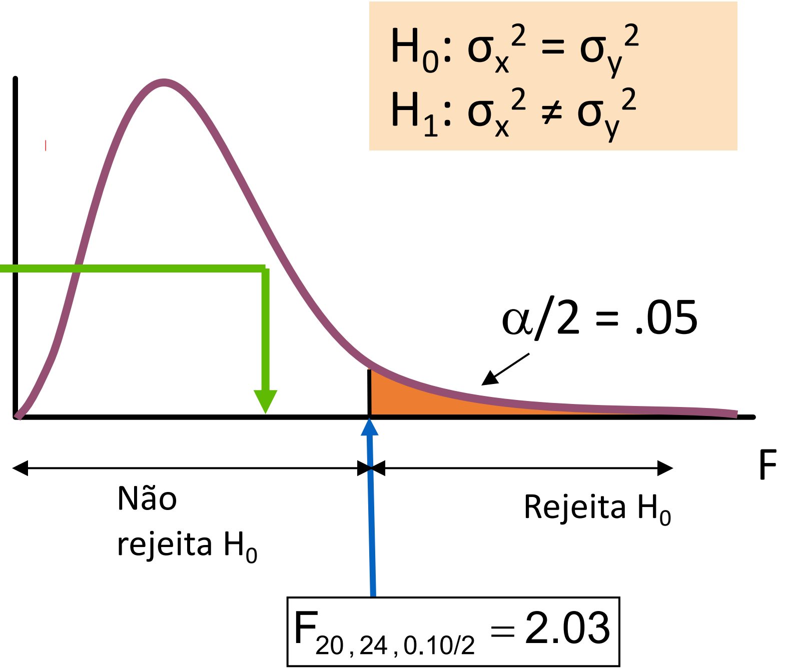
$$= F_{20\,,\,24\,,\,0.10/2} = \boxed{2.03}$$

Teste F : Exemplo Solução

(cont.)

- The test statistic is:

$$F = \frac{s_x^2}{s_y^2} = \frac{1.30^2}{1.16^2} = \boxed{1.256}$$



- F = 1.256 não está na região de rejeição, então não rejeitamos H_0

- Conclusão: Não há evidência suficiente de diferenças entre as variâncias a α = 0.10

Testes de duas amostras em Stata

For paired samples (t test):

https://www.youtube.com/watch?v=GiDSnufmZgI

For independent samples:

https://www.youtube.com/watch?v=by4c3h3WXQc

For variances…

- F test for two variances:
  - https://www.youtube.com/watch?v=by4c3h3WXQc&t=255s