10.2 Distribuição Amostral da Média

Se amostras aleatórias de tamanho n forem tomadas de uma população com média μ e variância σ^2, então a distribuição amostral de $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ tem as seguintes propriedades:

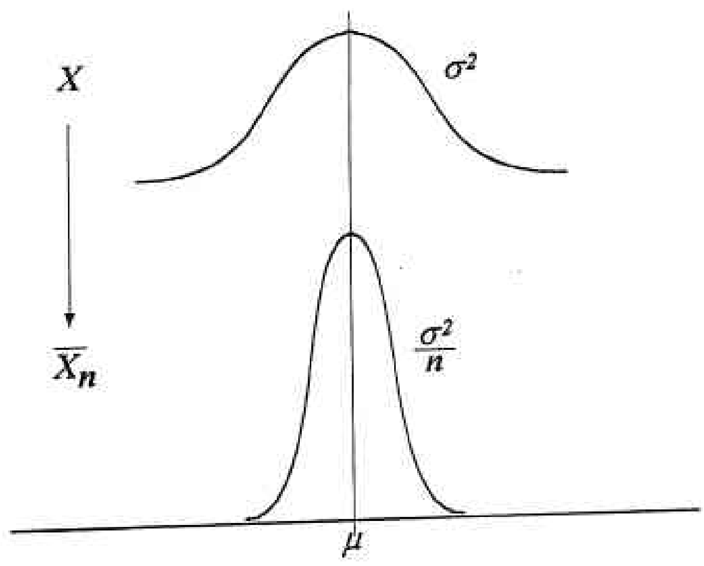
1) $\mu_{\overline{X}} = E(\overline{X}) = \mu$

2) $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} \Rightarrow \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$

Se o tamanho da amostra cresce, o desvio padrão da média amostral decresce.

3) Se a população original tem distribuição Normal, então $\overline{X}$ também tem distribuição Normal:

$$X \cap N(\mu,\sigma^2) \Rightarrow \overline{X} \cap N(\mu,\frac{\sigma^2}{n})$$

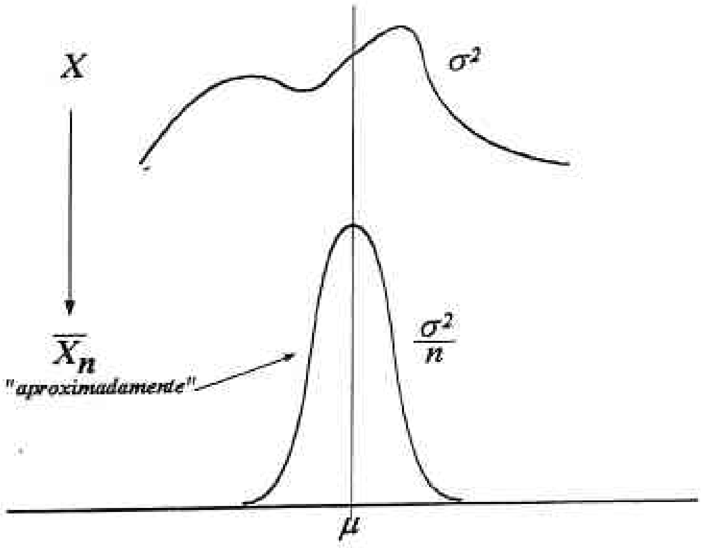


4) ***Teorema Central do Limite:***

Se a população original tem uma distribuição qualquer com média μ e variância σ^2, para n "suficientemente grande" (na prática, quando <u>n>30</u>), então $\overline{X}$ tem distribuição <u>aproximadamente</u> Normal:

$$\left.\begin{array}{r} E(X) = \mu \\ Var(X) = \sigma^2 \end{array}\right\} \Rightarrow \overline{X} \overset{a}{\cap} N(\mu; \frac{\sigma^2}{n})$$

onde $\overset{a}{\cap}$ significa aproximadamente distribuído



Exercício. Faça um esboço das distribuições de X e de $\overline{X}$, assumindo que X tem distribuição normal com média 40, variância 36 e que o tamanho de amostra foi igual a 9.

Exercício: Produção de Resina em <u>Pinus elliottii</u>

Seja X a produção anual de resina de árvores de Pinus elliottii. Suponha que X tem distribuição Normal com média 2,3 kg e desvio padrão 0,7kg.

a) Faça um esboço da distribuição de X.

b) Calcule a proporção esperada de árvores com produção maior do que 2,8 kg.

c) Foi tomada uma amostra aleatória de 16 árvores. Qual é a probabilidade de que a produção média das 16 árvores amostradas seja maior que 2,8 kg?

d) Uma amostra aleatória de 49 árvores foi tomada. Qual é a probabilidade de que a produção média das 49 árvores amostradas seja maior que 2,8 kg?

e) Uma amostra aleatória de 25 árvores foi tomada. Obter $\bar{x}$ tal que:

- $P(\overline{X} < \bar{x}) = 0{,}985$
- $P(\overline{X} < \bar{x}) = 0{,}975$

Exercício. Visitantes num Parque

Uma proporção de 37% dos visitantes de um parque favorecem a cobrança de taxas de entrada. Uma amostra aleatória de 200 visitantes foi tomada.

a) Qual é o parâmetro? Qual é a estatística?

b) Qual a probabilidade que na amostra de 200 visitantes pelo menos 40% favoreçam a cobrança de taxas?

c) Qual a probabilidade que na amostra de 200 visitantes, a proporção dos que favorecem a cobrança de taxas fique entre 35 e 39% ?

d) Uma nova amostra de 10 visitantes foi tomada. Qual a probabilidade de que pelo menos 50% dos visitantes na amostra favoreçam a cobrança de taxas ? É válido utilizar o mesmo método utilizado em (b) e (c) ? Por que?

Qual método deveria ser utilizado neste caso para encontrar a probabilidade?

Exemplo 13.A Mortalidade de Árvores numa Floresta

Após vários anos de acompanhamento de parcelas permanentes, uma Eng. Florestal concluiu que, das árvores que morrem num fragmento florestal, 75% são devido ao abafamento da copa por cipós. No ano seguinte, ocorreu um período prolongado de intensa seca e 30 árvores morreram durante o ano, das quais 24 mortes podem ser atribuídas ao efeito dos cipós. A Eng. Florestal afirma que a propòrção de árvores que morreram devido aos cipós (p) foi maior neste ano de seca intensa do que nos anos anteriores.

a. Estabeleça H_0 e H_α.

   H_0: A proporção de árvores mortas por cipós no ano de seca intensa é igual a 75%.

   H_α: A proporção de árvores mortas por cipós no ano de seca intensa é maior que 75%.

b. Estabeleça H_0 e H_α de forma quantitativa.

   $H_0 : p = 0.75$

   $H_\alpha : p > 0.75$

c. De acordo com a afirmação da Eng. Florestal, a hipótese nula foi rejeitada ou aceita? Explique

d. Se o teste de hipótese utilizou o nível de 5% de significância, o valor-p foi maior ou menor que 0.05? Explique.

e. Qual o tamanho de amostra utilizado para os resultados no ano de seca intensa?

Testando Hipóteses sobre a Proporção de uma População

Como a hipótese acima pode ser testada?

Como a taxa de mortalidade do último ano (ano de seca intensa) foi baseada numa amostra, o valor

$$\frac{24}{30} \times 100 = 80\%$$

é uma estatística da amostra e deve ser representado por $\hat{p}$. Nós sabemos que $\hat{p}$ varia de amostra para amostra. Se a amostragem fosse repedita e outras 30 árvores mortas observadas, nós provavelmente obteríamos um outro valor para $\hat{p}$.

Como já foi visto, a distribuição amostral de $\hat{p}$ tem as seguintes propriedades:

1. a média da distribuição de $\widehat{p}$ será p;
2. a variância da distribuição de $\widehat{p}$ será

$$\frac{p(1-p)}{n};$$

3. a distribuição de $\widehat{p}$ será aproximadamente Normal.

Com base nestas propriedades sabemos que a variável aleatória:

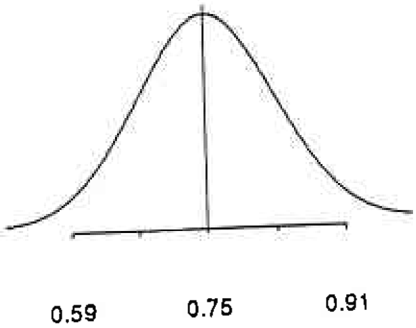
$$Z = \frac{\widehat{p} - p}{\sqrt{p(1-p)/n}}$$

terá distribuição aproximadamente Normal Padronizada, isto é, distribuição Normal com média 0 e desvio padrão 1.

Assumindo que a hipótese nula é verdadeira ($H_0 : p = 0.75$), $\widehat{p}$ seria aproximadamente Normal com média 0.75 e desvio padrão

$$\sqrt{\frac{0.75(1-0.75)}{30}}$$

(distribuição ao lado).



Podemos agora quantificar, sob as condições de H_0, a probabilidade de observarmos um valor de $\widehat{p} = 0.80$ ou mais extremo, .

O que quer dizer um valor mais extremo ?

A hipótese alternativa (H_α) estabelece que $p > 0.75$, portanto a Eng. Florestal acredita que a mortalidade no ano de seca intensa é **maior** do que nos anos normais. Se isso de fato ocorrer, o valor amostral $\widehat{p}$ poderia ser ainda maior do que 0.80.

$$\begin{aligned} H_\alpha : p > 0.75 \quad &\Rightarrow \quad P[\widehat{p} \geq 0.80] \\ H_0 : p = 0.75 \quad &\Rightarrow \quad P[\widehat{p} \geq 0.80] = P\left[Z \geq \frac{0.80 - 0.75}{\sqrt{0.75(1-0.75)/30}}\right] = P[Z \geq 0.63] = 0.2643 \end{aligned}$$

Portanto, sob H_0, a probabilidade de se observar um valor de $\widehat{p}$ ou maior é 0.2643. Se rejeitarmos H_0 com base no valor observado de $\widehat{p} = 0.80$, teremos a probabilidade de 0.2643 de cometermos um erro tipo I. Logo, 0.2643 é o valor-p, mas a margem de erro tipo I aceitável nas Ciências Florestais é de 0.05 (α - nível de significância). Como o valor-p (0.2643) é bem maior do que o nível de significância ($\alpha = 0.05$) nós não rejeitamos H_0 e podemos concluir que, do ponto de vista estatístico, a afirmação da Eng. Florestal não procede.

Testando Hipóteses sobre a Média de uma População

Até agora foi visto o teste Z para se testar a proporção de uma população. Qual seria o teste mais apropriado para testarmos hipóteses sobre *a média de uma população*?

Exemplo 13.D Testando o pH do Solo

Um cientista deseja saber se o pH de um solo é acido. Ele uma amostra com cinco unidades e obteve os valores de ph:

$$5.8;\ 6.3;\ 6.9;\ 6.2;\ 5.5$$

Considere os seguintes aspectos:

- O cientista considera o solo ácido se o seu pH for menor que 7. Portanto as hipótese a serem testadas são:

  H_0: ________________________________

  H_α: ________________________________

- Utilizando a teoria da distribuição da média amostral podemos utilizar o seguinte raciocínio: Seja X o pH do solo e possui distribuição Normal com média μ e desvio padrão σ, então:

  $\bar{x}_5$ terá média igual μ;

  $\bar{x}_5$ terá desvio padrão igual a $\sigma/\sqrt{5}$;

  $\bar{x}_5$ terá distribuição Normal.

- Logo, pela teoria da distribuição amostral, a estatística:

  $$Z = \frac{\bar{x}_5 - \mu}{\sigma/\sqrt{5}}$$

  terá distribuição Normal com média 0 e desvio padrão 1. Poderíamos utilizá-la para testar H_0.

Estas considerações apresentam um problema: a hipótese nula (H_0) nos fornece uma indicação para a média da população (μ), mas não nos dá qualquer dica sobre o desvio padrão da população (σ). O desvio padrão da população (σ) permanece desconhecido e, não há modo de se obter a estatística Z com base na amostra tomada pelo cientísta.

Estatística t

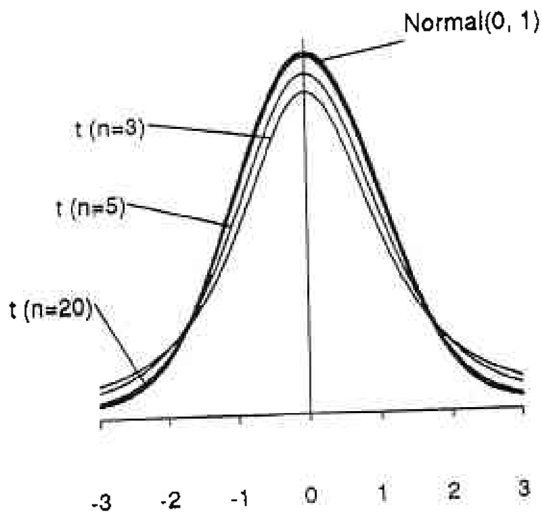
Este tipo de problema já foi estudado no começo do século pelo matemático inglês William Gosset, que usava o pseudônimo "Student" para publicar os seus trabalhos. Student desenvolveu a seguinte estatística:

$$t = \frac{\overline{x}_n - \mu}{s_n/\sqrt{n}}$$

onde s_n é o desvio padrão amostral (amostra de tamanho n).

A estatística t é uma variável aleatória cuja distribuição tem as seguintes propriedades:

1. Tem forma de sino (como a dist. Normal).
2. É simétrica em relação ao zero (como a dist. Normal Padronizada).
3. É mais "achatada" no centro e tem as caudas mais "pesadas".
4. Possui somente um parâmetro: $\nu = n-1$ chamado de graus de liberdade.
5. À medida que $\nu \to \infty$, então $t \to Z$, isto é, para grandes amostras a distribuição de t se aproxima da Normal Padronizada.



É importante notar que para cada valor dos graus de liberdade ($\nu = n - 1$) temos uma distribuição diferente para o t de Student. A distribuição Normal Padronizada é uma única distribuição (o valor dos parâmetros são fixos), mas a distribuição t de Student é na verdade uma *família de distribuições*, pois cada valor particular de ν gera uma distribuição distinta das demais.

Testando Hipóteses sobre a Média de uma População

Até agora foi visto o teste Z para se testar a proporção de uma população. Qual seria o teste mais apropriado para testarmos hipóteses sobre *a média de uma população*?

Exemplo 13.D Testando o pH do Solo

Um cientista deseja saber se o pH de um solo é acido. Ele uma amostra com cinco unidades e obteve os valores de ph:

$$5.8;\ 6.3;\ 6.9;\ 6.2;\ 5.5$$

Considere os seguintes aspectos:

- O cientista considera o solo ácido se o seu pH for menor que 7. Portanto as hipótese a serem testadas são:

  H_0: ______________________________

  H_α: ______________________________

- Utilizando a teoria da distribuição da média amostral podemos utilizar o seguinte raciocínio: Seja X o pH do solo e possui distribuição Normal com média μ e desvio pádrão σ, então:

  $\overline{x}_5$ terá média igual μ;

  $\overline{x}_5$ terá desvio padrão igual a $\sigma/\sqrt{5}$;

  $\overline{x}_5$ terá distribuição Normal.

- Logo, pela teoria da distribuição amostral, a estatística:

$$Z = \frac{\overline{x}_5 - \mu}{\sigma/\sqrt{5}}$$

  terá distribuição Normal com média 0 e desvio padrão 1. Poderíamos utilizá-la para testar H_0.

Estas considerações apresentam um problema: a hipótese nula (H_0) nos fornece uma indicação para a média da população (μ), mas não nos dá qualquer dica sobre o desvio padrão da população (σ). O desvio padrão da população (σ) permanece desconhecido e, não há modo de se obter a estatística Z com base na amostra tomada pelo cientísta.

Exemplo 13.D Testando o pH do Solo

Podemos agora testar as hipótese a respeito do pH do solo utilizando a estatística t. Seja μ o pH médio do solo e $\bar{x}$ a média amostral dos pHs na amostra:

- Hipóteses

  $H_0 : \mu = 7$

  $H_\alpha : \mu < 7$

- Dados

  $\bar{x} = (5.8 + 6.3 + 6.9 + 6.2 + 5.5)/5 = 6.14$

  $s = 0.5320$

- Estatística t:

$$t = \frac{6.14 - 7}{0.5320/\sqrt{5}} = -3.61$$

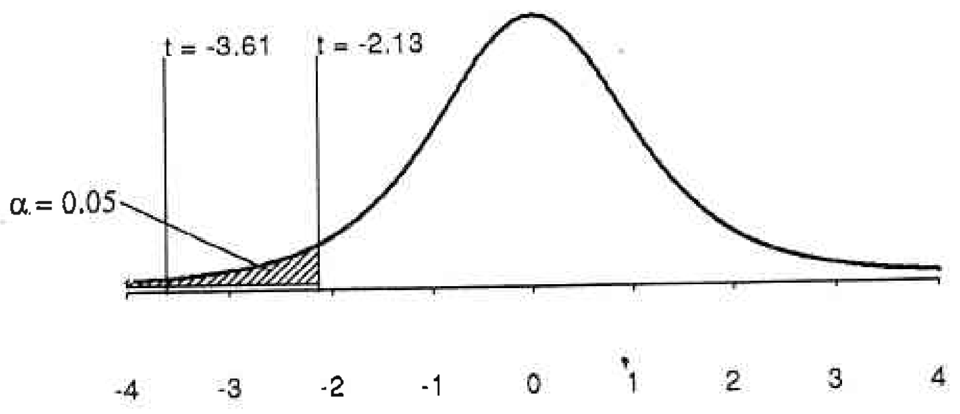
- Valor-p

$H_\alpha : \mu < 7 \quad \Rightarrow \quad P[\bar{x} < 6.14] \Rightarrow P[t < -3.61] \Rightarrow P[t > 3.61]$ (Simétrica em relação ao zero)

$H_0 : \mu = 7 \quad \Rightarrow \quad P[t > 3.61] = \text{valor-p}$

Na tabela do t de Student (Anexo C) encontramos o valor 2.13 para $t_{\alpha=0.05;\nu=4}$.
Logo:

$$P[t > 3.61] < P[t > 2.13] \Rightarrow P[t < -3.61] < P[t < -2.13]$$
$$\Rightarrow \text{valor-p} < \alpha = 0.05$$



- Decisão: Rejeita-se H_0 (ao nível de 5% de probabilidade) e concluí-se que existe evidência estatística para afirmar que o pH médio do solo é menor que 7.

10.2 Distribuição Amostral da Proporção

Seja p a proporção das unidades de uma população que possuem uma certa característica (proporção de "sucessos"), como por exemplo:

- árvores com cancro numa floresta de eucalipto;
- árvores cobertas por cipós numa mata;
- árvores dominadas (árvores vivas com altura inferior a 1,30m) numa floresta plantada de *PinusI*;
- capivaras do sexo masculino.

Se amostras aleatórias de tamanho n forem tomadas de uma população com proporção p, então a distribuição amostral de $\hat{p}_n$ tem as seguintes propriedades:

1. $\mu_{\hat{p}} = p \Rightarrow \hat{p}_n$ é um estimador sem viés de p.

2. $$\sigma^2_{\hat{p}} = \frac{p(1-p)}{n} \Rightarrow \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Se o tamanho da amostra cresce, o desvio padrão da proporção amostral decresce.

3. Se a população original tem uma distribuição qualquer, para n suficientemente grande ($n \geq 30$), $\hat{p}_n$ terá distribuição aproximadamente normal:

$$\hat{p}_n \overset{a}{\cap} N\left(p, \frac{p(1-p)}{n}\right)$$

Teste de hipóteses para a média de uma população

1° Passo: Hipóteses

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_a : \mu \neq \mu_0 \rightarrow \text{bilateral}$$

$$H_a : \left.\begin{matrix} \mu < \mu_0 \\ \mu > \mu_0 \end{matrix}\right\} \text{unilateral}$$

2° Passo: Estatística do teste

2.1. Variância conhecida $\Rightarrow z_{calc} = \dfrac{\bar{x} - \mu_0}{\sqrt{\dfrac{\sigma^2}{n}}}$

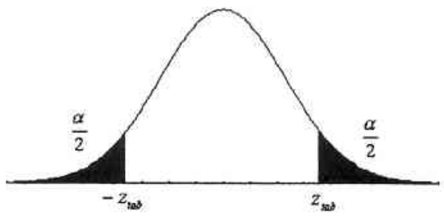
2.2. Variância desconhecida $\Rightarrow t_{calc} = \dfrac{\bar{x} - \mu_0}{\sqrt{\dfrac{s^2}{n}}}$ com (n-1) g.l.

2.3. Variância desconhecida com n > 30 (aproximação normal) $\Rightarrow z_{calc} = \dfrac{\bar{x} - \mu_0}{\sqrt{\dfrac{s^2}{n}}}$

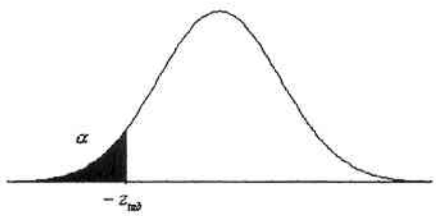
3° Passo: Regiões de Rejeição

3.1. Variância conhecida ou aproximação normal

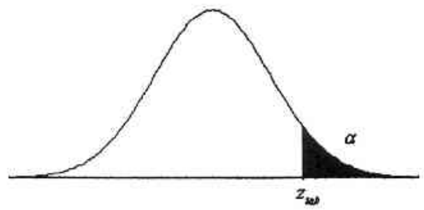
$H_a : \mu \neq \mu_0$



$H_a : \mu < \mu_0$

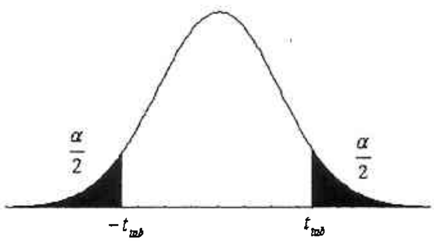


$H_a : \mu > \mu_0$

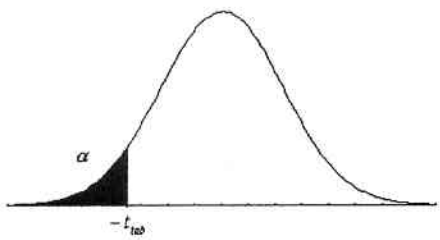


3.2. Variância desconhecida

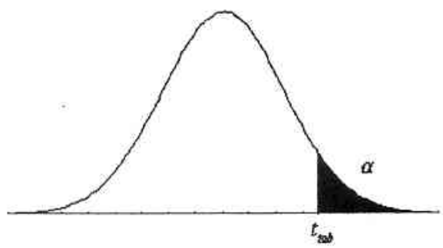
$H_a : \mu \neq \mu_0$



$H_a : \mu < \mu_0$



$H_a : \mu > \mu_0$



4° Passo: Conclusões

Teste de hipóteses para a proporção de uma população

1° Passo: Hipóteses

$$H_a: p \neq p_0 \rightarrow \text{bilateral}$$

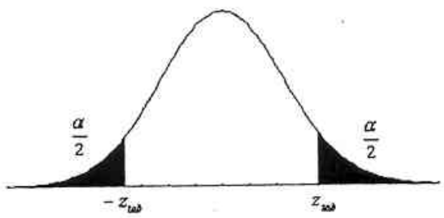
$$H_0: p = p_0 \quad \text{vs} \quad H_a: \left.\begin{matrix} p < p_0 \\ p > p_0 \end{matrix}\right\} \text{unilateral}$$

2° Passo: Estatística do teste

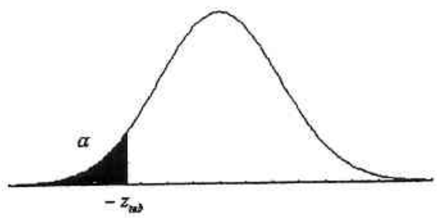
$$z_{calc} = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

3° Passo: Regiões de Rejeição

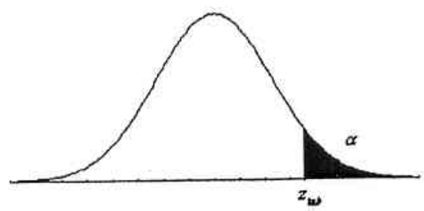
$H_a: p \neq p_0$



$H_a: p < p_0$



$H_a: p > p_0$



4° Passo: Conclusões

Teste para a diferença de médias de 2 populações – amostra pareada

1° Passo: Hipóteses

$$H_0 : \mu_1 - \mu_2 = \theta_0 \quad \text{vs} \quad H_a : \mu_1 - \mu_2 \neq \theta_0 \rightarrow \text{bilateral}$$

$$H_a : \left.\begin{matrix} \mu_1 - \mu_2 < \theta_0 \\ \mu_1 - \mu_2 > \theta_0 \end{matrix}\right\} \text{unilateral}$$

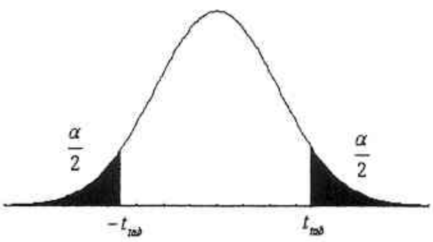
2° Passo: Estatística do teste

					Média	Var.
Amostra 1	x_{11}	x_{12}	...	x_{1n}	$\bar{x}_1$	s_1^2
Amostra 2	x_{21}	x_{22}	...	x_{2n}	$\bar{x}_2$	s_2^2
diferença (d_i)	d_1	d_2	...	d_n	$\bar{d}$	s_d^2

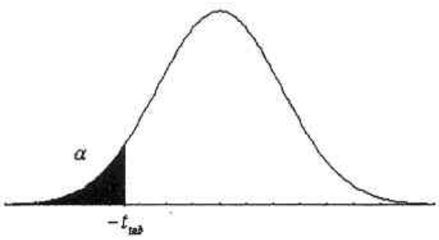
$$t_{calc} = \frac{\bar{d} - \theta_0}{\sqrt{\frac{s_d^2}{n}}}$$ com (n-1) g.l., sendo n = número de pares

3° Passo: Regiões de Rejeição

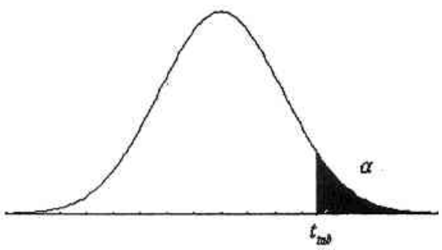
$H_a : \mu_1 - \mu_2 \neq \theta_0$



$H_a : \mu_1 - \mu_2 < \theta_0$



$H_a : \mu_1 - \mu_2 > \theta_0$



4° Passo: Conclusões

Teste para a diferença de médias de 2 populações – amostra não pareada

1° Passo: Hipóteses

$$H_0 : \mu_1 - \mu_2 = \theta_0 \quad \text{vs} \quad H_a : \mu_1 - \mu_2 \neq \theta_0 \rightarrow \text{bilateral}$$

$$H_a : \left.\begin{matrix} \mu_1 - \mu_2 < \theta_0 \\ \mu_1 - \mu_2 > \theta_0 \end{matrix}\right\} \text{unilateral}$$

2° Passo: Estatística do teste

					Média	Var.
Amostra 1	x_{11}	x_{12}	...	x_{1n}	$\bar{x}_1$	s_1^2
Amostra 2	x_{21}	x_{22}	...	x_{2n}	$\bar{x}_2$	s_2^2

2.1. Variâncias homogêneas

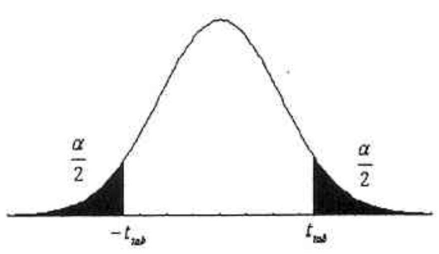
$$t_{calc} = \frac{\bar{x}_1 - \bar{x}_2 - \theta_0}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right)}} \quad \text{com } (n_1+n_2-2) \text{ g.l.}$$

2.1. Variâncias heterogêneas

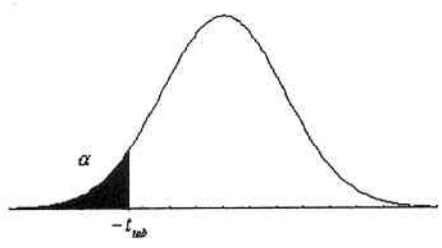
$$t_{calc} = \frac{\bar{x}_1 - \bar{x}_2 - \theta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{com } \nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

3° Passo: Regiões de Rejeição

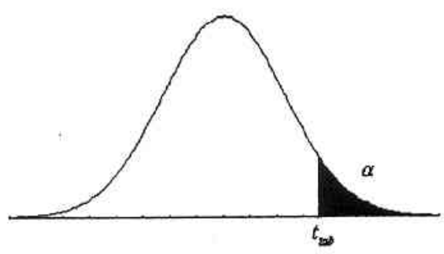
$H_a : \mu_1 - \mu_2 \neq \theta_0$



$H_a : \mu_1 - \mu_2 < \theta_0$



$H_a : \mu_1 - \mu_2 > \theta_0$



4° Passo: Conclusões

Teste de variância de 2 populações

1º Passo: Hipóteses

$H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_a : \sigma_1^2 > \sigma_2^2 \rightarrow$ unilateral

$H_a : \sigma_1^2 \neq \sigma_2^2 \rightarrow$ bilateral

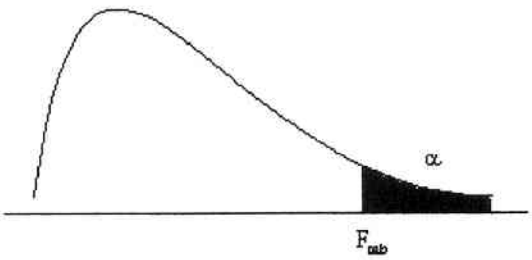
2º Passo: Estatística do teste

					Média	Var.
Amostra 1	x_{11}	x_{12}	...	x_{1n}	$\bar{x}_1$	s_1^2
Amostra 2	x_{21}	x_{22}	...	x_{2n}	$\bar{x}_2$	s_2^2

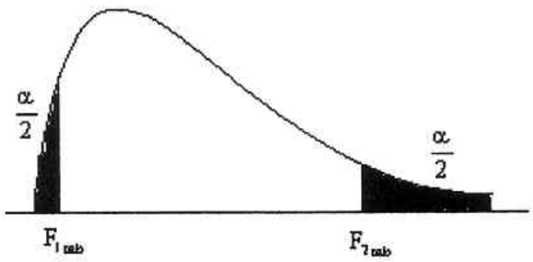
$$F_{calc} = \frac{s_{max}^2}{s_{min}^2} \quad \text{com} \quad \begin{array}{ll} \nu_1 = (n_1 - 1)\ \text{g.l.} & \text{(g.l. do numerador)} \\ \nu_2 = (n_2 - 1)\ \text{g.l.} & \text{(g.l. do denominador)} \end{array}$$

3º Passo: Regiões de Rejeição

$H_a : \sigma_1^2 > \sigma_2^2$



$H_a : \sigma_1^2 \neq \sigma_2^2$



4º Passo: Conclusões

Obs: Se H_0 é rejeitada – variância heterogênea

Se H_0 não é rejeitada – variância homogênea