

# **AULA 12**

# **Inferência a Partir de Duas Amostras**

**Ernesto F. L. Amaral**

**15 de setembro de 2011**  
**Metodologia de Pesquisa (DCP 854B)**

**Fonte:**

**Triola, Mario F. 2008. “Introdução à estatística”. 10<sup>a</sup> ed. Rio de Janeiro: LTC. Capítulo 9 (pp.360-407).**

## ESQUEMA DA AULA

- Inferências sobre duas proporções.
- Inferências sobre duas médias: amostras independentes.
- Inferências a partir de amostras emparelhadas.
- Comparação da variância em duas amostras.

## VISÃO GERAL

- Os capítulos anteriores (estimação de valores de parâmetros populacionais e teste de hipóteses) envolveram métodos para uma única amostra, usada para se fazer inferência sobre um único parâmetro populacional.
- Na prática, há muitas situações em que desejamos comparar dois conjuntos de dados amostrais.
- Portanto, este capítulo estende os métodos abordados anteriormente para situações que envolvem comparações de duas amostras em vez de apenas uma.

# **INFERÊNCIAS SOBRE DUAS PROPORÇÕES**

# INFERÊNCIAS SOBRE DUAS PROPORÇÕES

- Objetivo é de usar duas proporções amostrais:
  - Para **teste de afirmativa** sobre duas proporções populacionais.

ou

- Para construção de estimativa de **intervalo de confiança** da diferença entre proporções populacionais correspondentes.

## REQUISITOS

- No **teste de hipótese** sobre duas proporções populacionais ou na construção de um **intervalo de confiança** para diferença entre duas proporções populacionais, temos estes requisitos:
  - Temos proporções de duas **amostras aleatórias simples independentes** (valores amostrais selecionados de uma população não estão relacionados ou emparelhados com valores amostrais selecionados da outra população).
  - Para cada uma das duas amostras, o número de sucessos é, pelo menos, cinco e o número de fracassos também.

# NOTAÇÃO PARA DUAS PROPORÇÕES

- Para a população 1, fazemos:
  - $p_1$  = proporção populacional
  - $n_1$  = tamanho da amostra
  - $x_1$  = número de sucessos na amostra
- Proporção amostral:  $\hat{p}_1 = \frac{x_1}{n_1}$
- $\hat{q}_1 = 1 - \hat{p}_1$
- A população 2 possui o mesmo tipo de notação.

## PROPORÇÃO AMOSTRAL COMBINADA

- A proporção amostral combinada é simbolizada por  $p$ -barra e é dada por:

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

- O complementar de  $p$ -barra é dado por:

$$\bar{q} = 1 - \bar{p}$$

# ESTATÍSTICA DE TESTE PARA DUAS PROPORÇÕES

– Hipótese nula ( $H_0$ ):  $p_1 = p_2$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

– Onde:  $p_1 - p_2 = 0$  (pressuposto na hipótese nula)

$$\hat{p}_1 = \frac{x_1}{n_1} \quad \text{e} \quad \hat{p}_2 = \frac{x_2}{n_2}$$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} \quad \text{e} \quad \bar{q} = 1 - \bar{p}$$

– Os valores de  $p$  e valores críticos são encontrados com base no valor calculado do escore  $z$  (Tabela A-2).

## ESTIMATIVA DE INTERVALO DE CONFIANÇA

– A estimativa de intervalo de confiança para  $p_1 - p_2$  é:

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

– Onde a margem de erro  $E$  é dada por:

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$$

## DETERMINAÇÃO DO NÚMERO DE SUCESSOS $x_1$ e $x_2$

- Para calcular os testes de hipótese e intervalos de confiança, é preciso especificar os valores de  $x_1$ ,  $n_1$ ,  $x_2$  e  $n_2$ .
- Por exemplo, em uma pesquisa com 1.125 pessoas, 47% delas disseram que nunca ou raramente viajaram de avião.
  - $n_1 = 1125$
  - $\hat{p}_1 = 0,47$
  - Sendo:  $x_1 = n_1 * \hat{p}_1$
  - Temos:  $x_1 = 1125 * 0,47 = 528,75 \approx 529$
- Usamos os valores de  $n_1$  e  $x_1$ , além dos valores da população 2 (não exibidos), nos cálculos de estatística de teste para duas proporções.

# TESTES DE HIPÓTESES

- Consideraremos testes de hipóteses sobre duas proporções populacionais:

$$H_0: p_1 = p_2$$

- Sob a suposição de proporções iguais, a melhor estimativa da proporção comum é obtida pela combinação de ambas amostras em uma amostra grande, de modo que  $p$ -barra se torna uma estimativa mais óbvia da proporção populacional comum.

## EXEMPLO

- Pensar que a política é importante na vida é maior entre homens do que entre mulheres?

política	homem		Total
	0	1	
0	22,394	19,634	42,028
1	15,262	17,822	33,084
Total	37,656	37,456	75,112

- $n_0 = 37.656$

- $n_1 = 37.456$

- $H_0: p_0 = p_1$

- $H_1: p_1 > p_0$

- $\alpha = 0,05$

$$\bar{p} = \frac{x_0 + x_1}{n_0 + n_1} = \frac{15.262 + 17.822}{37.656 + 37.456} = 0,44046224$$

$$z = \frac{(\hat{p}_1 - \hat{p}_0) - (p_1 - p_0)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_0}}}$$

$$z = \frac{\left(\frac{17.822}{37.456} - \frac{15.262}{37.656}\right) - (0)}{\sqrt{\frac{(0,44046224)(0,55953776)}{37.456} + \frac{(0,44046224)(0,55953776)}{37.656}}} = 19,46$$

- $P \approx 0$  é menor do que  $\alpha = 0,05$ . Rejeitamos hipótese nula. Há evidência de que política é mais importante dentre homens.

## DESVIO PADRÃO EXATO $\neq$ ESTIMADO

- Podemos construir uma estimativa de intervalo de confiança da diferença entre proporções populacionais ( $p_1 - p_2$ ).
- Se um intervalo de confiança não inclui o zero, temos evidência que sugere que  $p_1$  e  $p_2$  tenham valores diferentes.
- O desvio padrão usado para intervalos de confiança é diferente do desvio padrão usado para o teste de hipótese.
  - O **teste de hipótese** usa desvio padrão **exato**, baseado na suposição de que não há diferença entre proporções.
  - O **intervalo de confiança** usa um desvio padrão baseado em valores **estimados** das proporções populacionais.

## INTERVALOS DE CONFIANÇA

- Se desejo é de estimar diferença entre duas proporções, utilize o **intervalo de confiança**.
- Se desejo é de testar alguma afirmativa sobre duas proporções, use um método de **teste de hipótese**.
- **NÃO** teste a igualdade de duas proporções populacionais pela determinação da existência de sobreposição de dois intervalos de confiança individuais.
- A análise da sobreposição de dois intervalos de confiança individuais é mais conservadora (menos rejeição de  $H_0$ ) do que estimativa de um intervalo de confiança  $p_1 - p_2$ .

## EXEMPLO DE INTERVALO DE CONFIANÇA

– Use os dados do exemplo anterior para construir intervalo de 95% de confiança para a diferença entre as proporções.

–  $\alpha = 0,05$

–  $\hat{p}_0 = 15.262/37.656 = 0,4053$

–  $z_{\alpha/2} = 1,96$

–  $\hat{p}_1 = 17.822/37.456 = 0,4758$

– Margem de erro:

–  $\hat{p}_1 - \hat{p}_0 = 0,0705$

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}_0 \hat{q}_0}{n_0} + \frac{\hat{p}_1 \hat{q}_1}{n_1}}$$

$$E = 1,96 \sqrt{\frac{\left(\frac{15.262}{37.656}\right) \left(\frac{22.394}{37.656}\right)}{37.656} + \frac{\left(\frac{17.822}{37.456}\right) \left(\frac{19.634}{37.456}\right)}{37.456}} = 1,96 * 0,0036 = 0,0071$$

– Intervalo de confiança:

$$(\hat{p}_1 - \hat{p}_0) - E < (p_1 - p_0) < (\hat{p}_1 - \hat{p}_0) + E$$

$$(0,4758 - 0,4053) - 0,0071 < (p_1 - p_0) < (0,4758 - 0,4053) + 0,0071$$

$$0,0634 < (p_1 - p_0) < 0,0776$$

## INTERPRETAÇÃO DO RESULTADO

- Limites do intervalo de confiança não contêm zero, sugerindo que há diferença significativa entre as duas proporções populacionais.
- Temos 95% de confiança que porcentagem de homens que pensam que política é importante é maior do que porcentagem de mulheres que pensam que política é importante por uma quantidade entre 6,34% e 7,76%.

# RESOLUÇÃO NO STATA

```
. prtest politica, by(mulher)
```

Two-sample test of proportion

```
0: Number of obs = 37456
1: Number of obs = 37656
```

Variable	Mean	Std. Err.	z	P> z	[95% Conf. Interval]
0	.4758116	.0025805			.470754 .4808693
1	.4053006	.00253			.4003419 .4102593
diff	.070511	.0036138			.063428 .077594
	under Ho:	.0036228	19.46	0.000	

```
diff = prop(0) - prop(1) z = 19.4631
Ho: diff = 0
```

```
Ha: diff < 0 Pr(Z < z) = 1.0000
Ha: diff != 0 Pr(|Z| < |z|) = 0.0000
Ha: diff > 0 Pr(Z > z) = 0.0000
```

- **Teste de hipótese:** a probabilidade de não rejeitarmos  $H_0$ :  $\text{diff}=0$  é muito pequena:  $P(Z>z)=0$ . Rejeitamos hipótese nula.
- Aceitamos  $H_1$ :  $\text{diff}>0$ . Há evidência de que política é mais importante para homens ( $\text{prop}_0$ ) do que para mulheres ( $\text{prop}_1$ ).
- **Intervalo de confiança:** com NC de 95%, diferença das proporções de homens e mulheres varia de 0,0634 a 0,0776.

# **INFERÊNCIAS SOBRE DUAS MÉDIAS: AMOSTRAS INDEPENDENTES**

## DEFINIÇÕES DE AMOSTRAS

- **Amostras independentes:** valores amostrais de uma população não estão relacionados ou combinados com os valores amostrais selecionados da outra população.
  - Ex.: grupo de tratamento e grupo de controle.
  
- **Amostras dependentes:** membros de uma amostra podem ser usados para determinar os membros da outra amostra.
  - Consistem em dados emparelhados dependentes, tais como dados de marido/mulher.
  - Dependência pode ocorrer com amostras relacionadas por associações como membros de uma família.
  - Ex.: dados coletados antes e depois de política pública.

# INFERÊNCIAS SOBRE DUAS MÉDIAS

- Serão apresentados métodos para uso de dados amostrais provenientes de **duas amostras independentes** para:
  - Teste de hipóteses sobre duas médias populacionais.
  - Construção de estimativas de intervalos de confiança para diferença entre duas médias populacionais.
- Esses métodos podem ser aplicados a situações em que:
  - Desvios padrões das duas populações são **desconhecidos e diferentes**. São métodos mais realistas e têm melhor desempenho.
  - Desvios padrões das duas populações são **conhecidos**.
  - Desvios padrões das duas populações são **desconhecidos**, mas **se supõe que sejam iguais**.

## $\sigma_1$ E $\sigma_2$ DESCONHECIDOS E DIFERENTES

- Ao usar duas amostras independentes para testar afirmativa sobre diferença ( $\mu_1 - \mu_2$ ) ou para construir intervalo de confiança utilize este requisitos:
  - $\sigma_1$  e  $\sigma_2$  são desconhecidos e não se faz suposição sobre igualdade entre eles.
  - Duas amostras são independentes.
  - Amostras aleatórias simples.
  - Uma ou ambas destas condições são satisfeitas:
    - Duas amostras são grandes ( $n_1 > 30$  e  $n_2 > 30$ ).
    - Amostras provêm de populações com distribuições normais:
      - Em amostras pequenas, procedimentos funcionam se não houver *outliers*.

## TESTE DE HIPÓTESE PARA DUAS MÉDIAS

- Para obter estatística do teste de hipótese para duas médias com amostras independentes, utilize:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

- Ao determinar valores críticos ou valores  $P$ , é preciso obter o número de graus de liberdade (gl):
  - No livro, gl é o menor número entre  $n_1 - 1$  e  $n_2 - 1$ .
  - Nos pacotes estatísticos:

$$gl = \frac{(A + B)^2}{\frac{A^2}{n_1 - 1} + \frac{B^2}{n_2 - 1}} \quad \text{onde: } A = \frac{s_1^2}{n_1} \quad \text{e} \quad B = \frac{s_2^2}{n_2}$$

## INTERVALO DE CONFIANÇA PARA $\mu_1 - \mu_2$

– Intervalo de confiança para a diferença  $\mu_1 - \mu_2$  é:

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

– Onde:

$$E = t_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

– Graus de liberdade é o mesmo usado para teste de hipótese.

## EXPLORANDO CONJUNTOS DE DADOS

- Antes de realizar teste de hipótese ou construir intervalo de confiança, devemos explorar as duas amostras:
  - Encontrar estatísticas descritivas para ambos conjuntos de dados ( $n$ , média e desvio padrão).
  - Fazer diagramas de caixa para os dois conjuntos de dados com a mesma escala.
  - Fazer histogramas do dois conjuntos de dados para comparar suas distribuições.
  - Identificar valores extremos (*outliers*).

## $\sigma_1$ E $\sigma_2$ CONHECIDOS

- No caso raro de conhecermos os desvios padrões populacionais, a estatística de teste e o intervalo de confiança se baseiam na distribuição normal em lugar da distribuição  $t$ .
- **Requisitos:**
  - Dois desvios padrões populacionais são conhecidos.
  - Duas amostras são independentes.
  - Amostras aleatórias simples.
  - Uma ou ambas destas condições são satisfeitas:
    - Duas amostras são grandes ( $n_1 > 30$  e  $n_2 > 30$ ).
    - Amostras provêm de populações com distribuições normais. Em amostras pequenas, procedimentos funcionam se não houver *outliers*.

## TESTE DE HIPÓTESE PARA DUAS MÉDIAS

- A estatística ( $z$ ) do teste de hipótese para duas médias de amostras independentes com  $\sigma_1$  e  $\sigma_2$  conhecidos é:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- Procurar valores  $P$  e valores críticos na tabela de distribuição normal padrão (Tabela A-2).

## INTERVALO DE CONFIANÇA

- O intervalo de confiança para  $\mu_1 - \mu_2$  em amostras independentes com  $\sigma_1$  e  $\sigma_2$  conhecidos é:

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

- Onde:

$$E = z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

## $\sigma_1$ E $\sigma_2$ DESCONHECIDOS E IGUAIS

- Se valores de  $\sigma_1$  e  $\sigma_2$  não forem conhecidos, mas se for razoável supor que tenham o mesmo valor, as variâncias amostrais podem ser combinadas para estimar  $\sigma^2$ .
- A estimativa combinada de  $\sigma^2$  é denotada por  $s_p^2$ .

## REQUISITOS PARA $\sigma_1$ E $\sigma_2$ DESCONHECIDOS E IGUAIS

- Dois desvios padrões populacionais não são conhecidos, mas supõe-se que sejam iguais ( $\sigma_1 = \sigma_2$ ).
- Duas amostras são independentes.
- Amostras aleatórias simples.
- Uma ou ambas destas condições são satisfeitas:
  - Duas amostras são grandes ( $n_1 > 30$  e  $n_2 > 30$ ).
  - Amostras provêm de populações com distribuições normais:
    - Em amostras pequenas, procedimentos funcionam se não houver *outliers*.

## TESTE DE HIPÓTESE

- Estatística do teste de hipótese para duas médias com amostras independentes e com  $\sigma_1$  igual a  $\sigma_2$ :

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

- Onde temos a variância combinada:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

- O número de graus de liberdade é dado por  $gl = n_1 + n_2 - 2$ .

## INTERVALO DE CONFIANÇA

- O intervalo de confiança para  $\mu_1 - \mu_2$  com amostras independentes e com  $\sigma_1$  e  $\sigma_2$  iguais é:

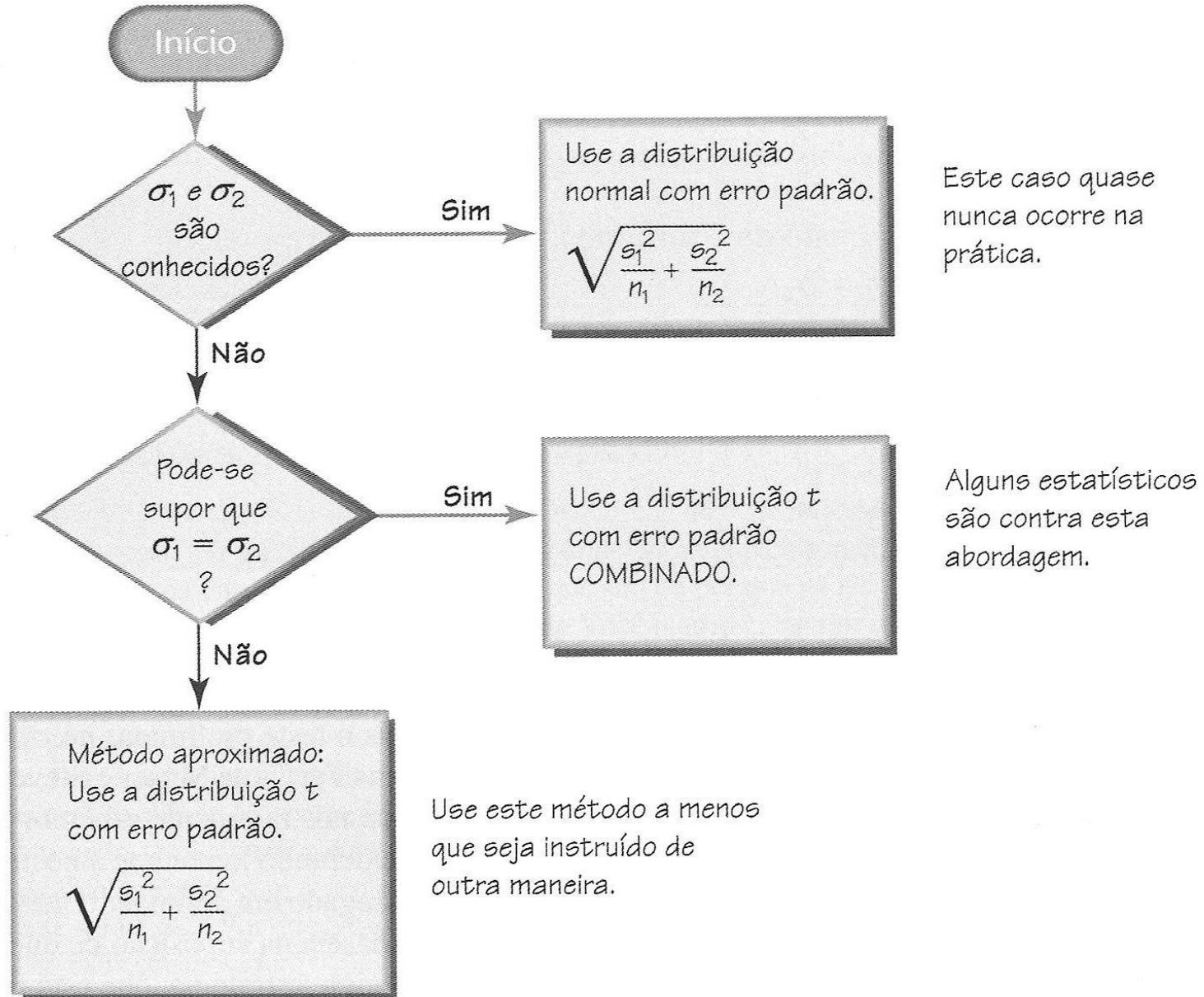
$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

- Onde:

$$E = t_{\alpha/2} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

- A variância combinada ( $s_p^2$ ) e o número de graus de liberdade ( $n_1 + n_2 - 2$ ) é similar ao do teste de hipótese.

# INFERÊNCIA SOBRE DUAS MÉDIAS INDEPENDENTES



# **INFERÊNCIAS A PARTIR DE AMOSTRAS EMPARELHADAS**

# INFERÊNCIAS COM AMOSTRAS EMPARELHADAS

- Duas amostras são dependentes se membros de uma amostra podem ser usados para determinarem os membros da outra amostra.
- Ou seja, com dados emparelhados, há alguma relação, de modo que cada valor em uma amostra está emparelhado com um valor correspondente na outra amostra.

# REQUISITOS PARA AMOSTRAS EMPARELHADAS

- Dados amostrais consistem em dados emparelhados.
- Amostras aleatórias simples.
- Uma ou ambas destas condições são satisfeitas:
  - Número de pares de dados é grande ( $n > 30$ ).
  - Pares têm diferenças que são provenientes de uma população com distribuição aproximadamente normal.
    - Se houver afastamento radical da distribuição normal, não devemos usar os métodos desta seção.

# NOTAÇÃO PARA DADOS EMPARELHADOS

- $d$  = diferença individual entre os dois valores em um único par.
- $\mu_d$  = valor médio das diferenças  $d$  para a população de todos os pares.
- $\bar{d}$  = valor médio das diferenças  $d$  para os dados amostrais emparelhados (igual à média dos valores  $x - y$ ).
- $s_d$  = desvio padrão das diferenças  $d$  para os dados amostrais emparelhados.
- $n$  = número de pares de dados.

## TESTE DE HIPÓTESE

- Estatística de teste de hipótese para dados emparelhados é dada por:

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

- Onde graus de liberdade é igual a  $n - 1$ .

## INTERVALO DE CONFIANÇA

– Intervalo de confiança para dados emparelhados é:

$$\bar{d} - E < \mu_d < \bar{d} + E$$

– Onde:

$$E = t_{\alpha/2} \frac{S_d}{\sqrt{n}}$$

## EXEMPLO

- Índice de valores racionais (tradicional/secular) é maior entre homens do que entre mulheres?

*tab mulher, sum(tradtrat5) mean*

mulher	Mean
0	.2231321
1	.24008968
Total	.23166243

# RESOLUÇÃO DO EXEMPLO

```
. ttest tradrat5, by(mulher)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	39730	.2231321	.0043531	.8676724	.2146	.2316642
1	40216	.2400897	.0045294	.9083252	.2312119	.2489674
combined	79946	.2316624	.003142	.8883898	.2255041	.2378207
diff		-.0169576	.0062839		-.0292739	-.0046413

diff = mean(0) - mean(1) t = -2.6986  
 Ho: diff = 0 degrees of freedom = 79944

Ha: diff < 0  
 Pr(T < t) = 0.0035

Ha: diff != 0  
 Pr(|T| > |t|) = 0.0070

Ha: diff > 0  
 Pr(T > t) = 0.9965

- **Teste de hipótese:** probabilidade de não rejeitar  $H_0$ : diff=0 é muito pequena:  $P(T < t) = 0,0035$ . Rejeitamos hipótese nula.
- Aceitamos  $H_1$ : diff < 0. Evidência de menores valores seculares entre homens ( $mean_0$ ) do que mulheres ( $mean_1$ ).
- **Intervalo de confiança:** com NC de 95%, diferença das médias de homens e mulheres varia de -0,0293 a -0,0046.

# COMPARAÇÃO DA VARIAÇÃO EM DUAS AMOSTRAS

# COMPARAÇÃO DA VARIAÇÃO EM DUAS AMOSTRAS

- Esta seção apresenta o teste  $F$  que usa duas variâncias amostrais (ou desvios padrões) para a comparação de duas variâncias populacionais (ou desvios padrões).
- O teste  $F$  para a comparação de duas variâncias populacionais é muito sensível a afastamentos da distribuição normal.
- Notações de medidas de variação:
  - $s$  = desvio padrão de amostra
  - $s^2$  = variância da amostra (desvio padrão amostral ao quadrado).
  - $\sigma$  = desvio padrão da população.
  - $\sigma^2$  = variância da população (desvio padrão populacional ao quadrado)

## REQUISITOS

- Duas populações são independentes uma da outra:
  - Duas amostras são independentes se amostra selecionada de uma população não se relaciona com amostra selecionada da outra população.
- Duas populações são normalmente distribuídas:
  - Métodos desta seção não são robustos, já que são extremamente sensíveis a afastamentos da normalidade.

# TESTES DE HIPÓTESE

- Notação para testes de hipótese com duas variâncias ou desvios padrões:
  - $s_1^2$  = maior das duas variâncias amostrais.
  - $n_1$  = tamanho da amostra com a maior variância.
  - $\sigma_1^2$  = variância da população da qual se extraiu a amostra com a maior variância.
  - Os símbolos  $s_2^2$ ,  $n_2$  e  $\sigma_2^2$  são usados para a outra amostra e população.

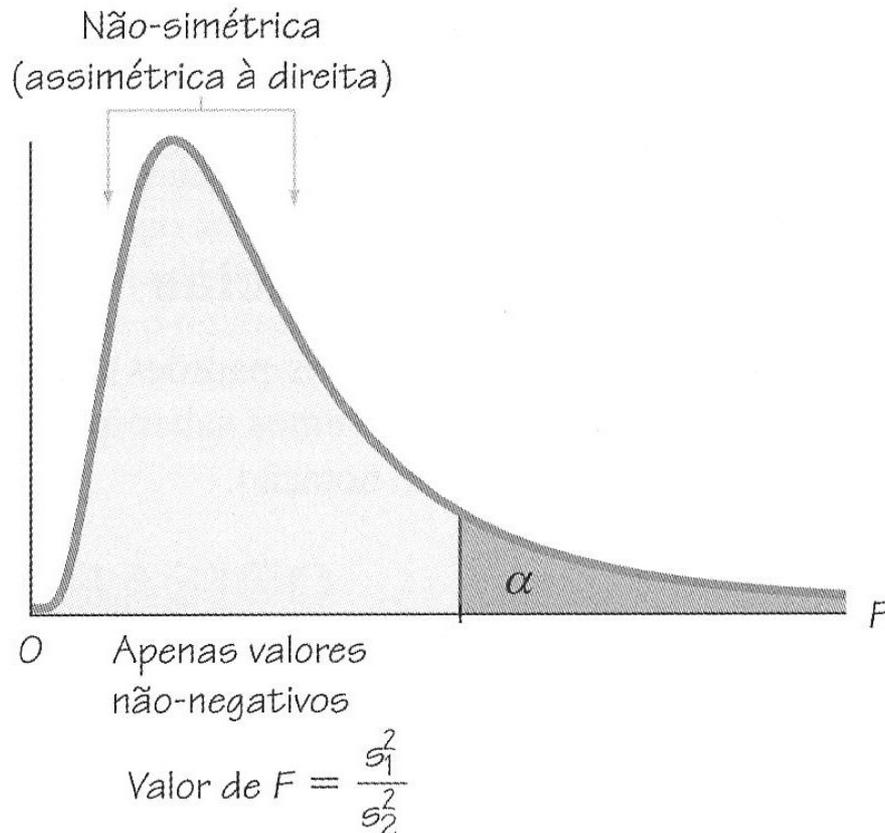
- Estatística de teste de hipótese com duas variâncias:

$$F = \frac{s_1^2}{s_2^2}$$

- Número de graus de liberdade do numerador =  $n_1 - 1$ .
- Número de graus de liberdade do denominador =  $n_2 - 2$ .

# DISTRIBUIÇÃO F

- Se realizarmos vários experimentos para selecionar amostras aleatórias de duas populações normalmente distribuídas com variâncias iguais, a distribuição da razão  $s_1^2/s_2^2$  das variâncias amostrais será a distribuição F.



- Há uma distribuição F diferente para cada par distinto de graus de liberdade para o numerador e o denominador.

## INTERPRETAÇÃO DA ESTATÍSTICA DE TESTE $F$

- Se as duas populações têm variâncias iguais, então a razão  $s_1^2/s_2^2$  tende a se aproximar de 1.
- Como  $s_1^2$  é sempre a maior variância,  $s_1^2$  e  $s_2^2$  terão valores muito distantes um do outro se a razão for um número grande.
- Ou seja, valores grandes de  $F$  são evidência contra  $\sigma_1^2 = \sigma_2^2$ .
- A estatística de teste  $F$  se aplica a uma afirmativa feita sobre duas variâncias, mas pode ser usada para realizar afirmações sobre dois desvios padrões.

## EXEMPLO

- Índice de valores racionais (tradicional/secular) possui desvio padrão maior entre homens do que entre mulheres?

```
. sdtest tradrat5, by(homem)
```

Variance ratio test

Group	obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	40216	.2400897	.0045294	.9083252	.2312119	.2489674
1	39730	.2231321	.0043531	.8676724	.2146	.2316642
combined	79946	.2316624	.003142	.8883898	.2255041	.2378207

```
ratio = sd(0) / sd(1)                                f = 1.0959
Ho: ratio = 1                                         degrees of freedom = 40215, 39729
```

```
Ha: ratio < 1
Pr(F < f) = 1.0000
```

```
Ha: ratio != 1
2*Pr(F > f) = 0.0000
```

```
Ha: ratio > 1
Pr(F > f) = 0.0000
```

- **Teste de hipótese:** probabilidade de não rejeitar  $H_0$ : razão=1, é muito pequena:  $P(F>f)=0,0000$ . Rejeitamos  $H_0$ .
- Aceitamos  $H_1$ : razão>1. Evidência de maior desvio padrão no índice entre mulheres ( $sd_0$ ) do que entre homens ( $sd_1$ ).