DISTRIBUIÇÃO AMOSTRAL

- A população define um conjunto vasto, em geral, impossível de conhecer.
- A amostra constitui um subconjunto da população.
- Uma amostra aleatória é uma amostra em que a probabilidade de cada elemento ser seleccionado é conhecida.
- O objectivo é, a partir da amostra, estabelecer conclusões para o todo representado pela população.

DEFINIÇÕES

- Amostra aleatória $x_1, x_2, \ldots, x_n$

- Uma **estatística** é uma medida numérica calculada a partir dos dados amostrais.

- Um parâmetro é uma medida numérica de uma população.

DEFINIÇÕES

	População	Amostra
Média	μ	$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Desvio Padrão	σ	$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}$
Variância	σ^2	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$

ESTATÍSTICA INFERENCIAL

- **Estatística Inferencial** é o conjunto de procedimentos que permitem, a partir de uma amostra, fazer inferências para a população.

AMOSTRA ALEATÓRIA

- Se $x_1, x_2, \ldots, x_n$ são variáveis aleatórias independentes e identicamente distribuídas, então constituem uma amostra aleatória de uma população infinita caracterizada pela sua distribuição comum.

DEFINIÇÃO

- Se $x_1, x_2, \ldots, x_n$ constituem uma amostra aleatória, então

  a média amostral é dada por, $$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

  e a variância amostral por $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$$

- As estatísticas são funções de variáveis aleatórias

DISTRIBUIÇÃO DA MÉDIA

- Se $x_1, x_2, \ldots, x_n$ constituem uma amostra aleatória de uma população infinita com média μ e variância σ^2 então

$$E\left[\overline{x}\right] = \mu$$

$$V\left[\overline{x}\right] = \frac{\sigma^2}{n}$$

TEOREMA DO LIMITE CENTRAL

- Se $x_1, x_2, \ldots, x_n$ constituem uma amostra aleatória de uma população com média μ e variância σ^2 finita, a distribuição limite de

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

à medida que $n \to \infty$ é a distribuição normal padrão.

Exemplo:

Suponha que as classificações de uma grande turma têm média de 72 e um desvio padrão de 9.

a) Encontre a probabilidade de que numa a.a. de 10 estudantes tenha uma média acima de 80.
b) Se a população é Normal, encontre a probabilidade de que um estudante seleccionado aleatoriamente, tenha uma classificação acima de 80.

Solução

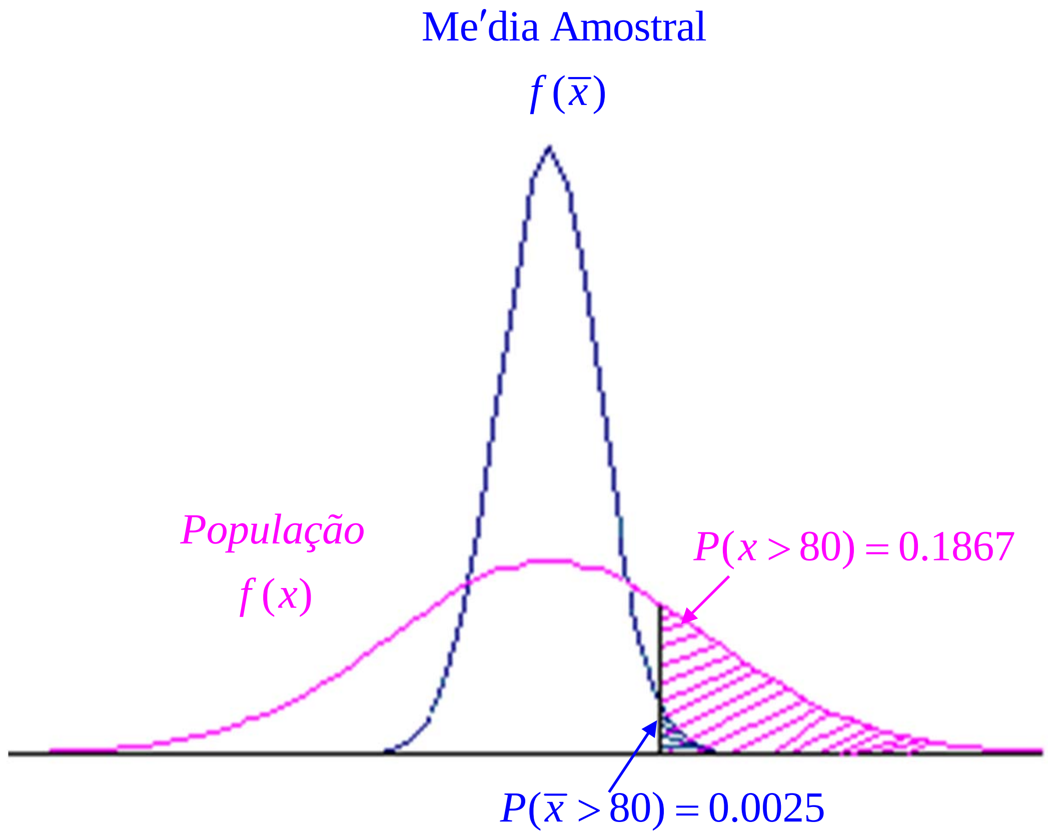
a)

$$P(\overline{x} > 80) = P\left(\frac{\overline{x} - \mu_X}{\sigma_X \big/ \sqrt{n}} > \frac{80 - 72}{9 \big/ \sqrt{10}} \right) = P(z > 2.81) = 0.0025$$

b)

$$P(x > 80) = P\left(\frac{x - \mu_X}{\sigma_X} > \frac{80 - 72}{9} \right) = P(z > 0.89) = 0.1867$$

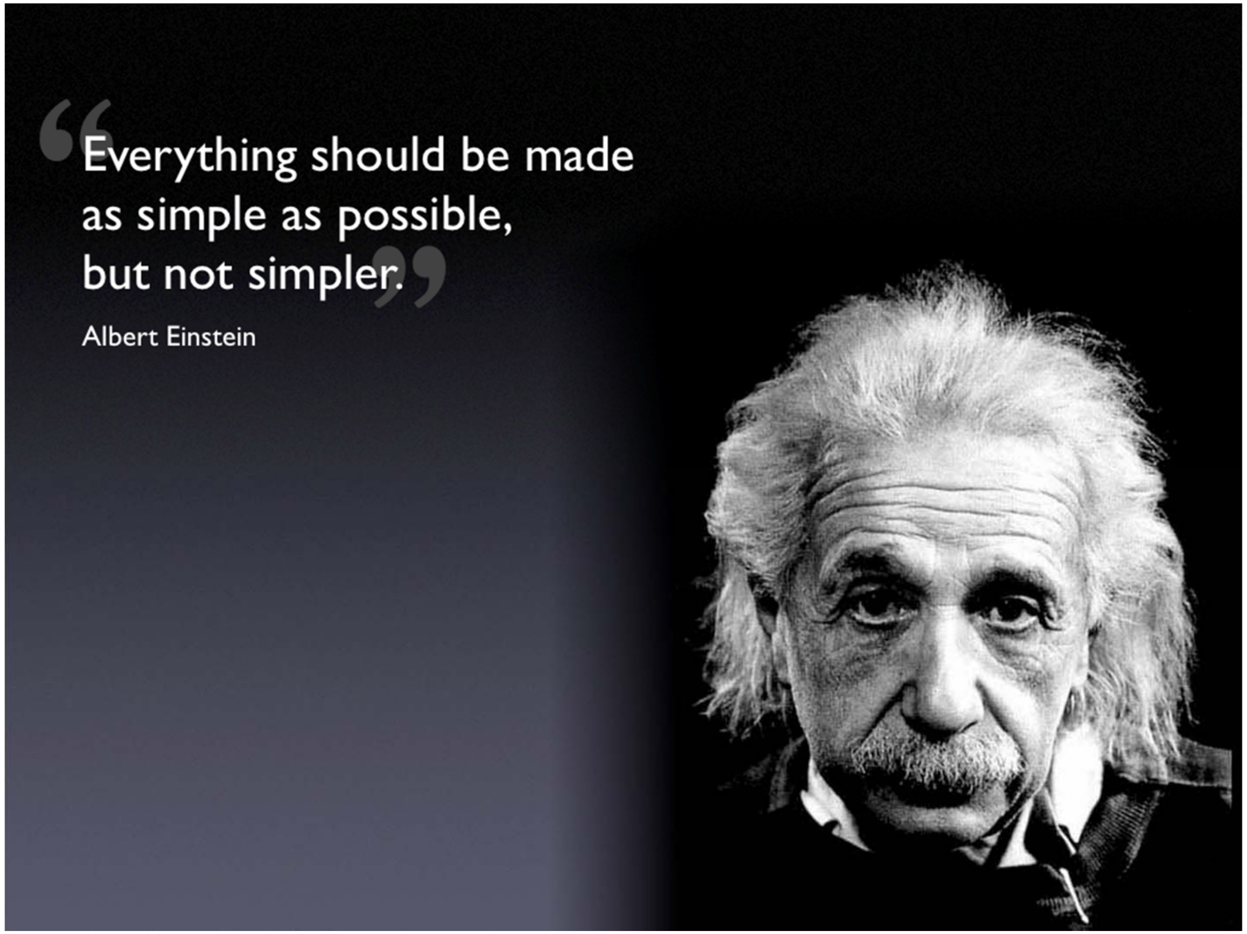
Solução



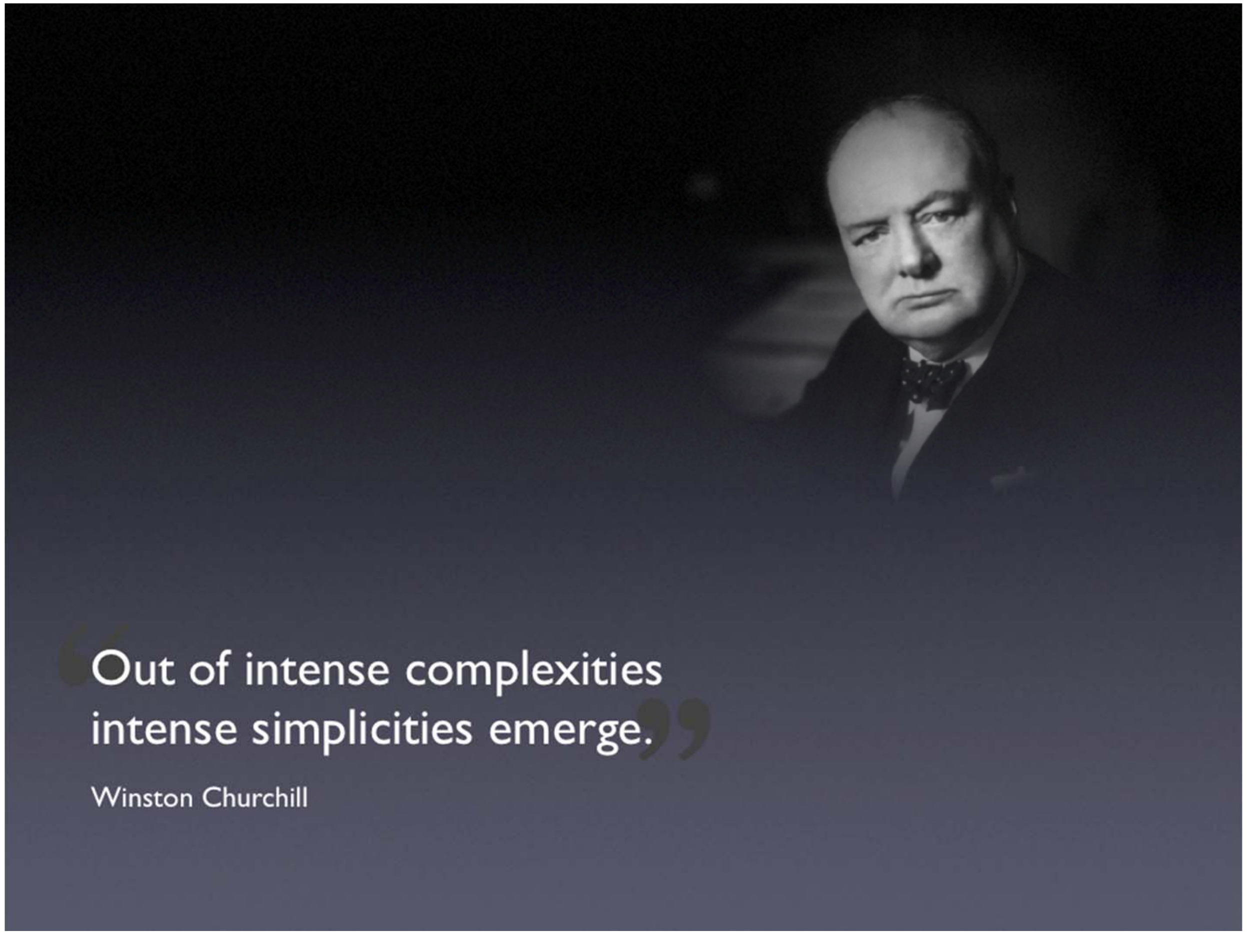
Me′dia Amostral
$f(\overline{x})$
População
$f(x)$
$P(x > 80) = 0.1867$
$P(\overline{x} > 80) = 0.0025$

Estatística	Parâmetro
• Característica da amostra	• Característica da população
• Valor variável	• Valor fixo
• Valor conhecido	• Valor desconhecido

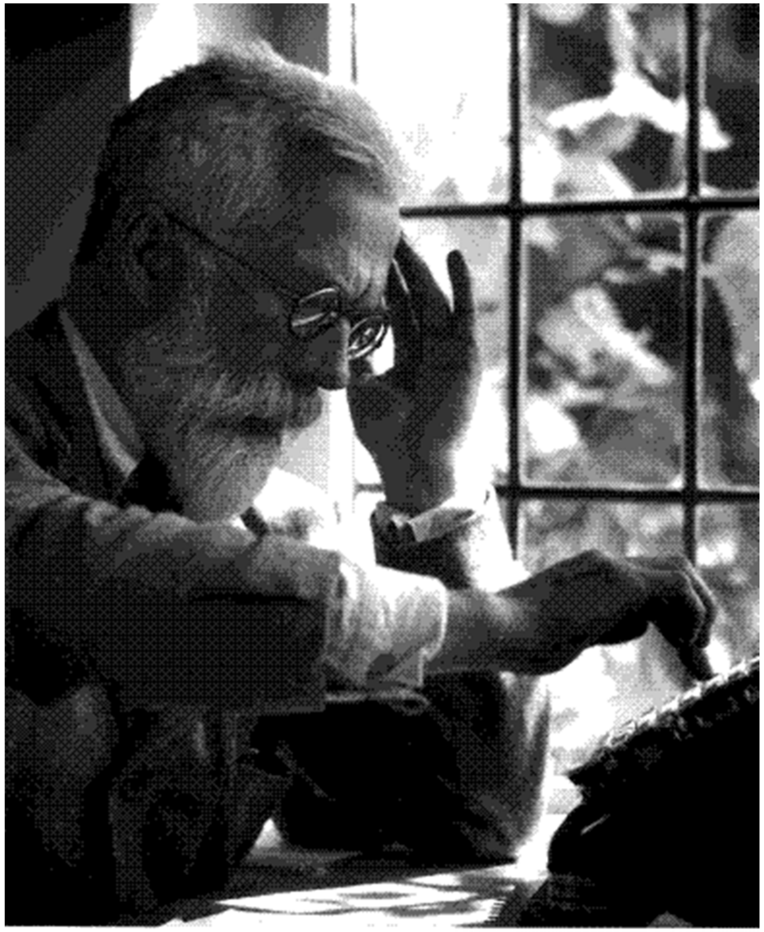
ESTATÍSTICA E LITERACIA



ESTATÍSTICA E LITERACIA



ESTATÍSTICA E LITERACIA



Quando consultam um estatístico pedindo a análise de dados recolhidos sem o seu aconselhamento prévio, pretendem um diagnóstico, mas em geral só já é possível fazer uma AUTÓPSIA.

Ronald Aylmer Fisher

INTERVALOS DE CONFIANÇA

INTERVALOS DE CONFIANÇA

- Estabelecer um intervalo de confiança para o parâmetro θ .

$$P(\hat{\theta}_I < \theta < \hat{\theta}_S) = 1 - \alpha$$

- Determinar os dois limites que definem o intervalo,

$$\hat{\theta}_I < \theta < \hat{\theta}_S$$

  limites que dependem da distribuição amostral de θ e são, respectivamente, os limites inferior e superior do intervalo.

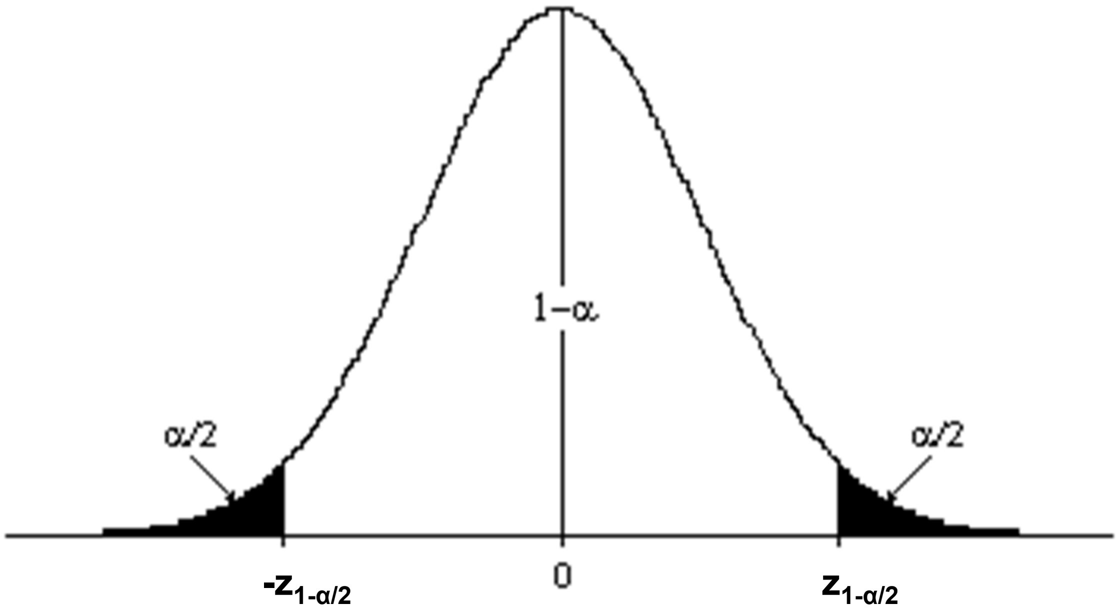
INTERVALO DE CONFIANÇA

- A média de uma amostra possui uma distribuição, σ² conhecido.

$$\mu_{\overline{x}} = \mu \qquad \sigma_{\overline{x}}^2 = \sigma^2 / n \qquad z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}}$$

$$P\left(-z_{1-\alpha/2} < z < z_{1-\alpha/2}\right) = 1 - \alpha$$

INTERVALO DE CONFIANÇA



INTERVALO DE CONFIANÇA

$$P\left(-z_{1-\alpha/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2} \right) = 1-\alpha$$

$$P\left(\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \right) = 1-\alpha$$

INTERVALO DE CONFIANÇA: MÉDIA

- σ^2 conhecido

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\bar{x} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} \approx \bar{x} \pm z_{1-\alpha/2}\frac{s}{\sqrt{n}}$$

EXEMPLO 1

- Suponha que era conhecido que a média e o desvio padrão das alturas dos rapazes com 20 anos era

$$\mu = 170\ cm, \sigma = 10\ cm$$

- Considere que foram recolhidas 5 amostras de 25 rapazes, tendo sido observadas as seguintes médias

Amostra	**1**	**2**	**3**	**4**	**5**
Média (*cm*)	172	168	171	165	172

SOLUÇÃO 1

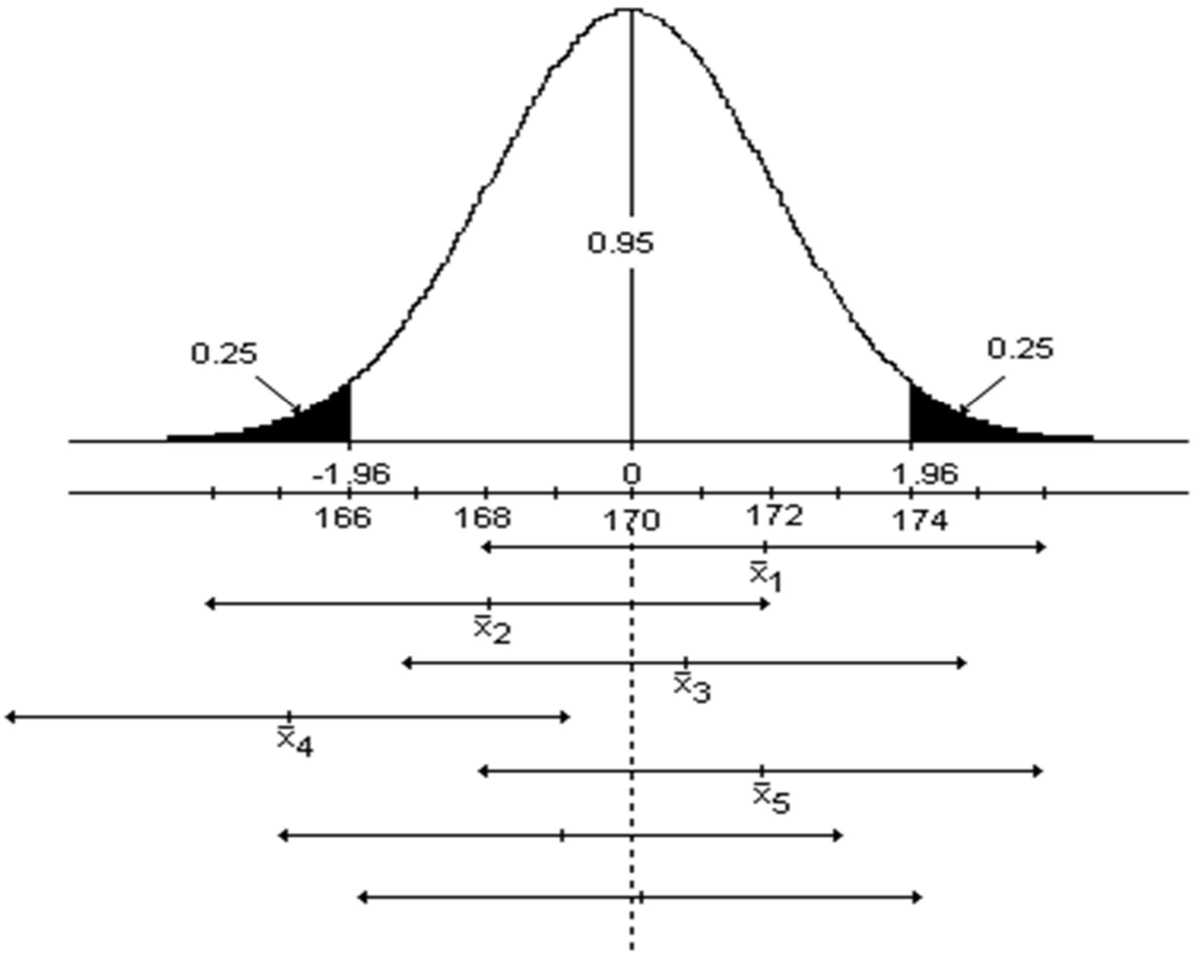
- Intervalo de Confiança de 95%

$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\bar{x} - 1.96\frac{10}{\sqrt{25}} < \mu < \bar{x} + 1.96\frac{10}{\sqrt{25}}$$

$$\bar{x} \pm 4cm$$

SOLUÇÃO 1



EXEMPLO 2

- O peso ao nascer é uma das variáveis mais importantes na avaliação do bem-estar de um recém nascido.

- Suponha que o valor do desvio padrão para os bebés de sexo masculino é 562 gramas. Num determinado centro de saúde, uma amostra de 19 recém nascidos apresentou uma média 3222 gramas.

- Construa um intervalo de confiança de 95% para média do peso dos bebés.

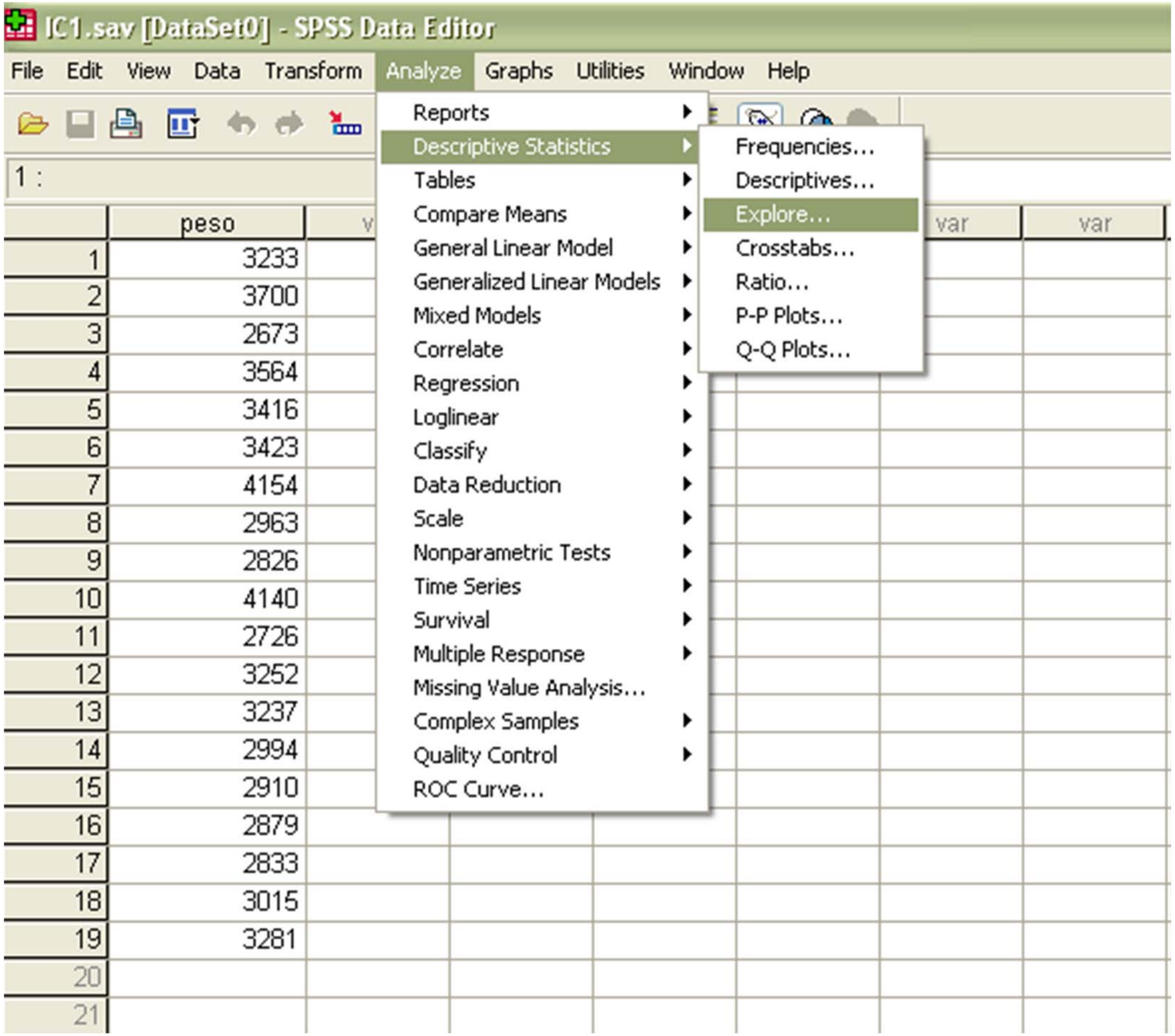
SOLUÇÃO 2

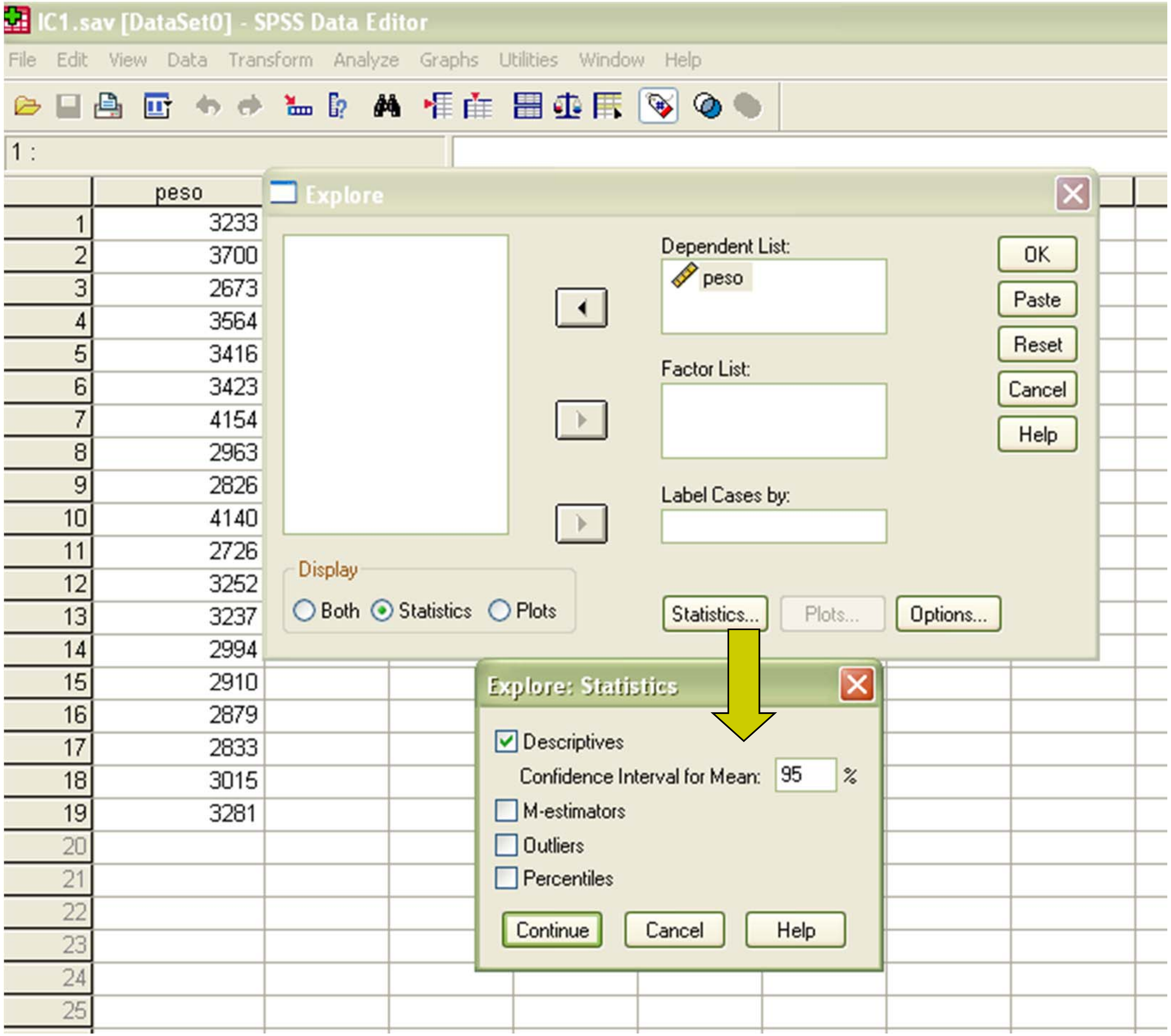
$$\bar{x} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

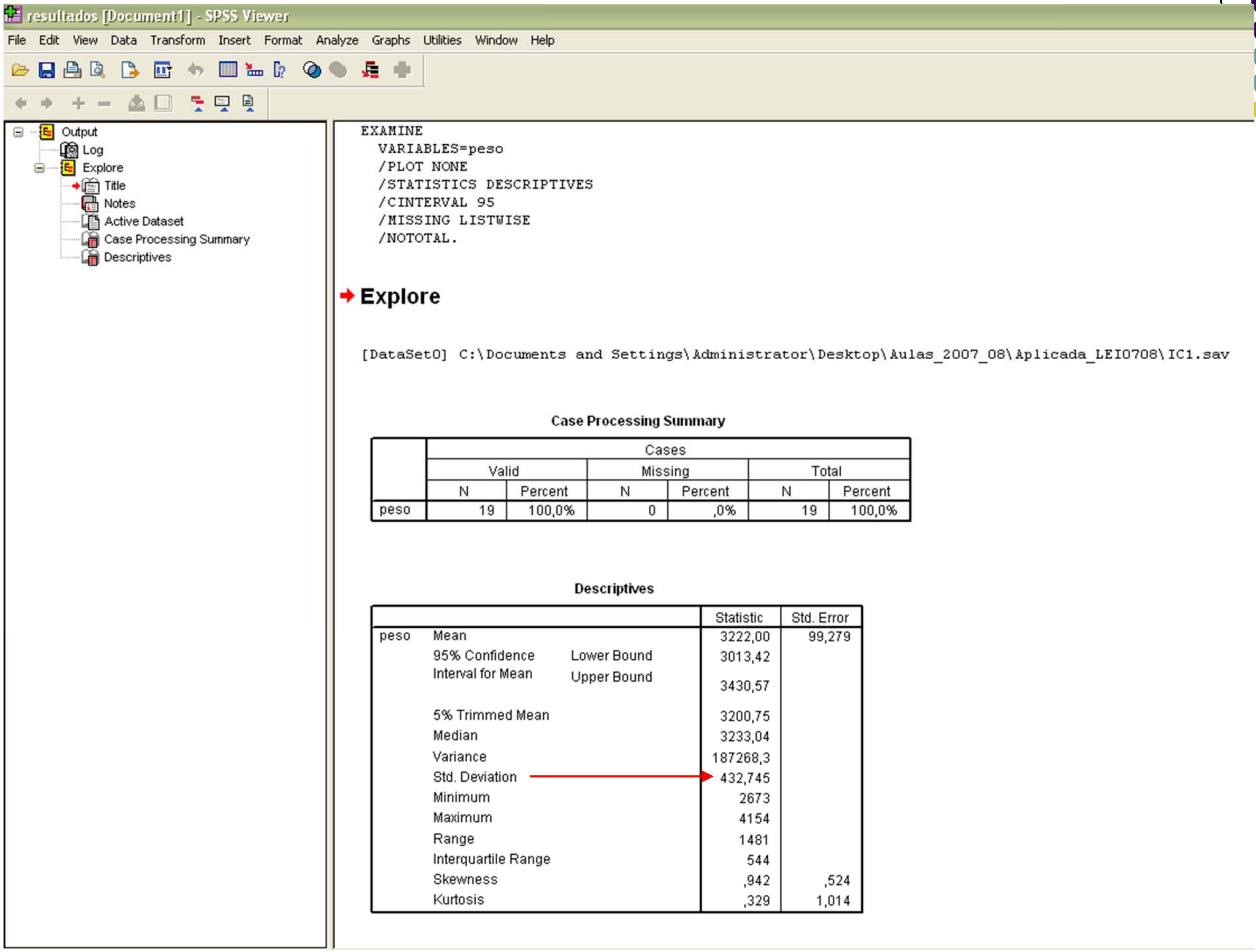
$$3222 - 1.96\frac{562}{\sqrt{19}} < \mu < 3222 + 1.96\frac{562}{\sqrt{19}}$$

$$3222 \pm 253g$$

$$2969 < \mu < 3475$$


IC1.sav [DataSet0] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples
Quality Control
ROC Curve...
Frequencies...
Descriptives...
Explore...
Crosstabs...
Ratio...
P-P Plots...
Q-Q Plots...
peso
var
var
1 3233
2 3700
3 2673
4 3564
5 3416
6 3423
7 4154
8 2963
9 2826
10 4140
11 2726
12 3252
13 3237
14 2994
15 2910
16 2879
17 2833
18 3015
19 3281
20
21


IC1.sav [DataSet0] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
peso
1 3233
2 3700
3 2673
4 3564
5 3416
6 3423
7 4154
8 2963
9 2826
10 4140
11 2726
12 3252
13 3237
14 2994
15 2910
16 2879
17 2833
18 3015
19 3281
Explore
Dependent List:
peso
Factor List:
Label Cases by:
OK
Paste
Reset
Cancel
Help
Display
Both Statistics Plots
Statistics... Plots... Options...
Explore: Statistics
Descriptives
Confidence Interval for Mean: 95 %
M-estimators
Outliers
Percentiles
Continue Cancel Help




```
EXAMINE
  VARIABLES=peso
  /PLOT NONE
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

Explore

[DataSet0] C:\Documents and Settings\Administrator\Desktop\Aulas_2007_08\Aplicada_LEI0708\IC1.sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
peso	19	100,0%	0	,0%	19	100,0%

Descriptives

			Statistic	Std. Error
peso	Mean		3222,00	99,279
	95% Confidence Interval for Mean	Lower Bound	3013,42	
		Upper Bound	3430,57	
	5% Trimmed Mean		3200,75	
	Median		3233,04	
	Variance		187268,3	
	Std. Deviation		432,745	
	Minimum		2673	
	Maximum		4154	
	Range		1481	
	Interquartile Range		544	
	Skewness		,942	,524
	Kurtosis		,329	1,014

INTERVALO DE CONFIANÇA: MÉDIA

- σ^2 desconhecido, n<30

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} \qquad P\left(-t_{\alpha/2,n-1} < t < t_{\alpha/2,n-1}\right) = 1 - \alpha$$

$$P\left(-t_{\alpha/2,n-1} < \frac{\overline{x} - \mu}{s/n} < t_{\alpha/2,n-1}\right) = 1 - \alpha$$

$$P\left(\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

INTERVALO DE CONFIANÇA: MÉDIA

- σ^2 desconhecido, n<30

$$\bar{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

$$\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

EXEMPLO 3

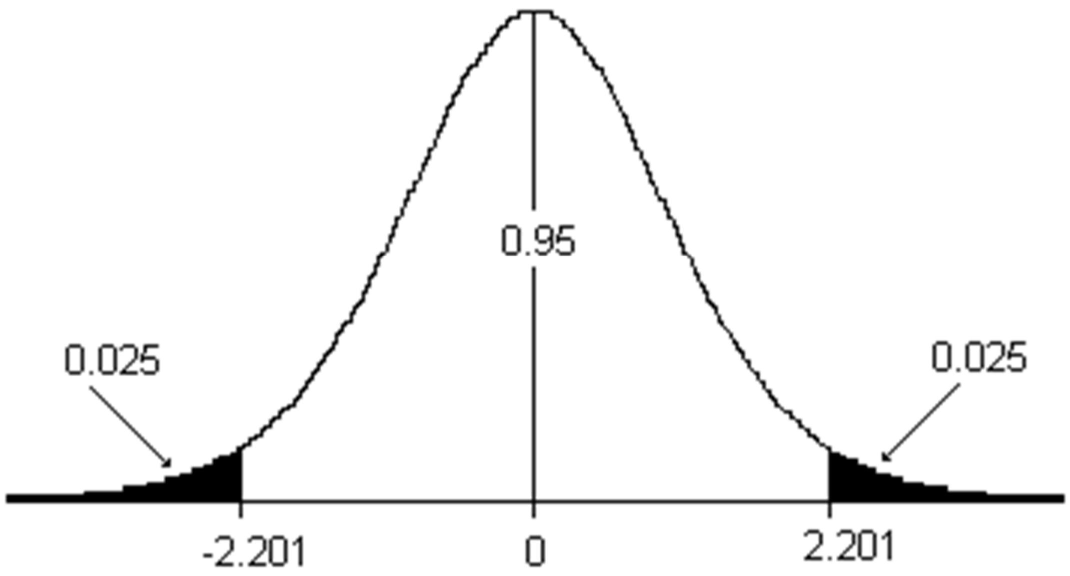
- Numa universidade, uma amostra de 12 estudantes foi seleccionada.

- O comprimento médio da mão encontrado foi de 19.92 cm com um desvio padrão de 0.17cm.

- Construa um intervalo de confiança de 95% para o verdadeiro valor do comprimento médio.

SOLUÇÃO 3

- *t*-Student, com 11 graus de liberdade



SOLUÇÃO 3

$$19.92 - 2.201\frac{0.17}{\sqrt{12}} < \mu < 19.92 + 2.201\frac{0.17}{\sqrt{12}}$$

$$19.92 \pm 0.108$$

$$19.812 < \mu < 20.028$$

INTERVALO DE CONFIANÇA DIFERENÇA DE MÉDIAS

- $\overline{x}_1$, $\overline{x}_2$ médias de amostras aleatórias independentes, de dimensão n_1, n_2
- Populações normais com médias μ_1 e μ_2 e variância comum desconhecida σ^2

$$t = \frac{\left(\overline{x}_1 - \overline{x}_2\right) - \left(\mu_1 - \mu_2\right)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

INTERVALO DE CONFIANÇA DIFERENÇA DE MÉDIAS

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2, n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2$$

$$< \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2, n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{\alpha/2, n_1+n_2-2}\, s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

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

EXEMPLO 5

- Pretende-se testar duas formulações alimentares no crescimento de frangos de aviário. Os frangos, distribuídos por dois pavilhões A e B, foram alimentados durante cinco semanas com a respectiva ração. No fim do período de crescimento, foram seleccionadas duas amostras.

Grupo	n	Média (g)	Desvio Padrão (g)
Pav. A	16	1623,7500	192,7131
Pav. B	10	1588,0000	167,1194

SOLUÇÃO 5

- *t*-Student

$$s_p^2 = \frac{(16-1)(192.7131)^2 + (10-1)(167.1194)^2}{16+10-2}$$

$$s_p^2 = 33684.7970 \qquad t_{0.025,24} = 2.06$$

$$(1623.75 - 1588.00) \pm (2.06)(183.5342)\sqrt{\frac{1}{16} + \frac{1}{10}}$$

$$35.75 \pm 152.4091$$

$$-116.6591 < \mu_1 - \mu_2 < 188.1591$$

INTERVALO DE CONFIANÇA DIFERENÇA DE MÉDIAS

- $\overline{x}_1$, $\overline{x}_2$ médias de amostras aleatórias independentes
- Populações normais com médias μ_1 e μ_2 e variâncias desconhecidas e diferentes

INTERVALO DE CONFIANÇA DIFERENÇA DE MÉDIAS

$$n_1 = n_2 = n$$

$$n_1 + n_2 - 2 = 2(n-1)$$

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{\alpha/2,2(n-1)} \sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}$$

$$n_1 \neq n_2$$

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t_{\alpha/2,\nu} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\nu = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$$

INTERVALO DE CONFIANÇA DIFERENÇA DE MÉDIAS

- Amostras emparelhadas

$$\mu_d = (\mu_1 - \mu_2)$$

$$\bar{d} - t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right) < \mu_d < \bar{d} + t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right)$$

$$\bar{d} \pm t_{\alpha/2,n-1}\left(\frac{s_d}{\sqrt{n}}\right)$$

EXEMPLO 6

- Uma amostra de dez trabalhadores de uma fábrica onde existe a manipulação de dioxinas foi seleccionada aleatoriamente.

- Nestes trabalhadores foi determinada a concentração (em ppm, partes por milhão) de dioxinas no plasma e no tecido gordo.

- Construa um intervalo de confiança para a diferença entre as concentrações de dioxina no plasma e no tecido gordo.

EXEMPLO 6

Trabalhador	Plasma	Tecido Gordo
1	2.5	4.9
2	3.5	6.9
3	1.8	4.2
4	4.7	4.4
5	7.2	7.7
6	4.1	2.5
7	3.0	5.5
8	3.3	2.9
9	3.1	5.9
10	2.5	2.3

SOLUÇÃO 5

- *t*-Student

$$\bar{d} = -1.1500$$

$$s_d = 1.7335$$

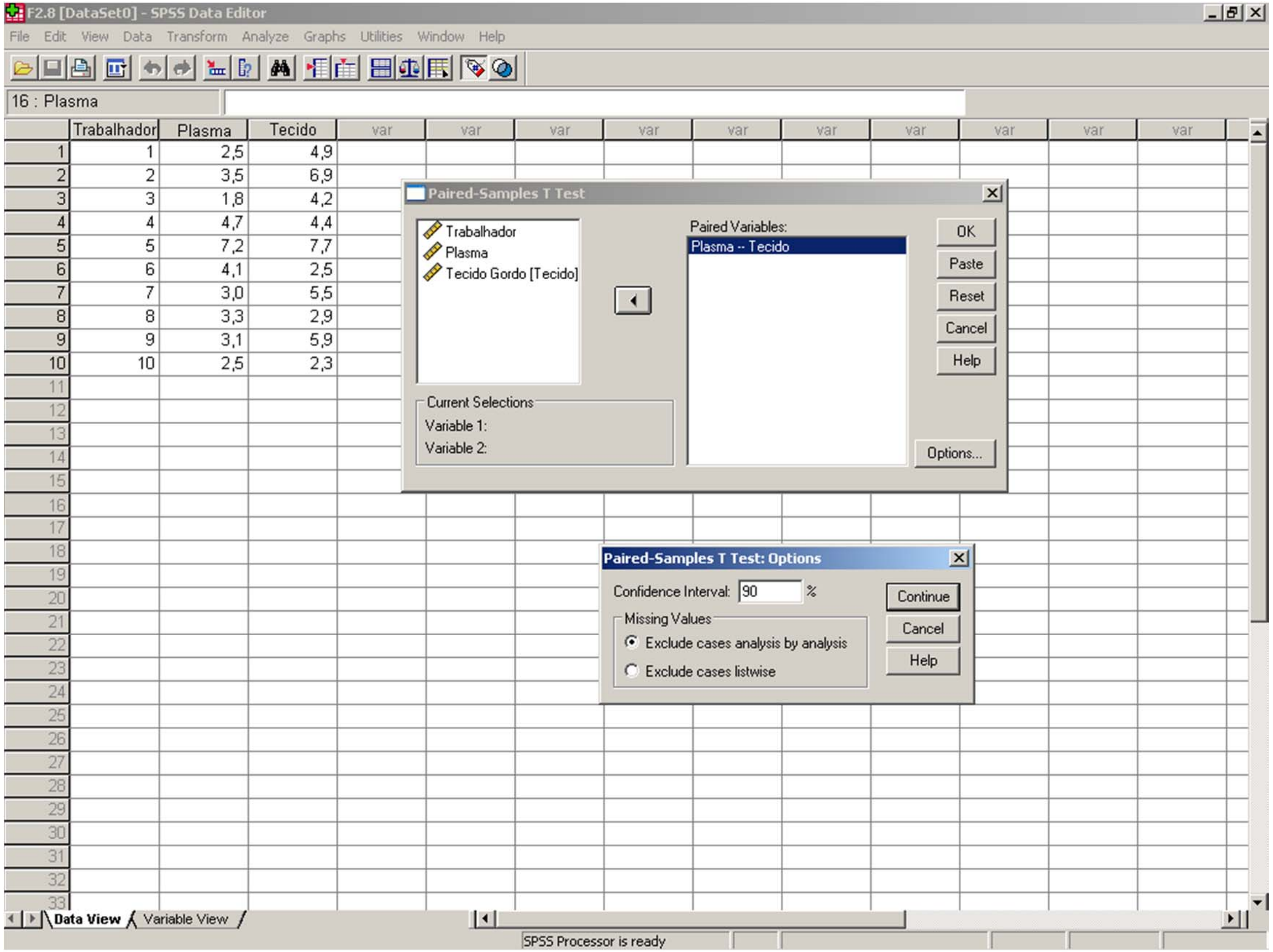
$$t_{0.025,9} = 2.262$$

$$-1.1500 - 2.262\frac{1.7335}{\sqrt{10}} < \mu_d < -1.1500 + 2.262\frac{1.7335}{\sqrt{10}}$$

$$-1.1500 \pm 2.262\frac{1.7335}{\sqrt{10}}$$

$$-1.1500 \pm 1.2400$$

$$-2.3900 < \mu_d < 0.0900$$


F2.8 [DataSet0] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
16 : Plasma
Trabalhador Plasma Tecido var var var var var var var var var var
1 1 2,5 4,9
2 2 3,5 6,9
3 3 1,8 4,2
4 4 4,7 4,4
5 5 7,2 7,7
6 6 4,1 2,5
7 7 3,0 5,5
8 8 3,3 2,9
9 9 3,1 5,9
10 10 2,5 2,3
Paired-Samples T Test
Trabalhador
Plasma
Tecido Gordo [Tecido]
Paired Variables:
Plasma -- Tecido
OK
Paste
Reset
Cancel
Help
Current Selections
Variable 1:
Variable 2:
Options...
Paired-Samples T Test: Options
Confidence Interval: 90 %
Continue
Cancel
Help
Missing Values
Exclude cases analysis by analysis
Exclude cases listwise
Data View
Variable View
SPSS Processor is ready

Output1 - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

```
  /COMPRESSED.
T-TEST
  PAIRS = Plasma  WITH Tecido (PAIRED)
  /CRITERIA = CI(.9)
  /MISSING = ANALYSIS.
```

T-Test

[DataSet0]

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Plasma	3,570	10	1,5195	,4805
	Tecido Gordo	4,720	10	1,8299	,5787

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Plasma & Tecido Gordo	10	,477	,163

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	90% Confidence Interval of the Difference: Lower	Upper			
Pair 1	Plasma - Tecido Gordo	-1,1500	1,7335	,5482	-2,1549	-,1451	-2,098	9	,065

SPSS Processor is ready

INTERVALO DE CONFIANÇA PROPORÇÃO

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \qquad P\left(-z_{1-\alpha/2} < \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} < z_{1-\alpha/2} \right) = 1 - \alpha$$

$$P\left(\hat{p} - z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

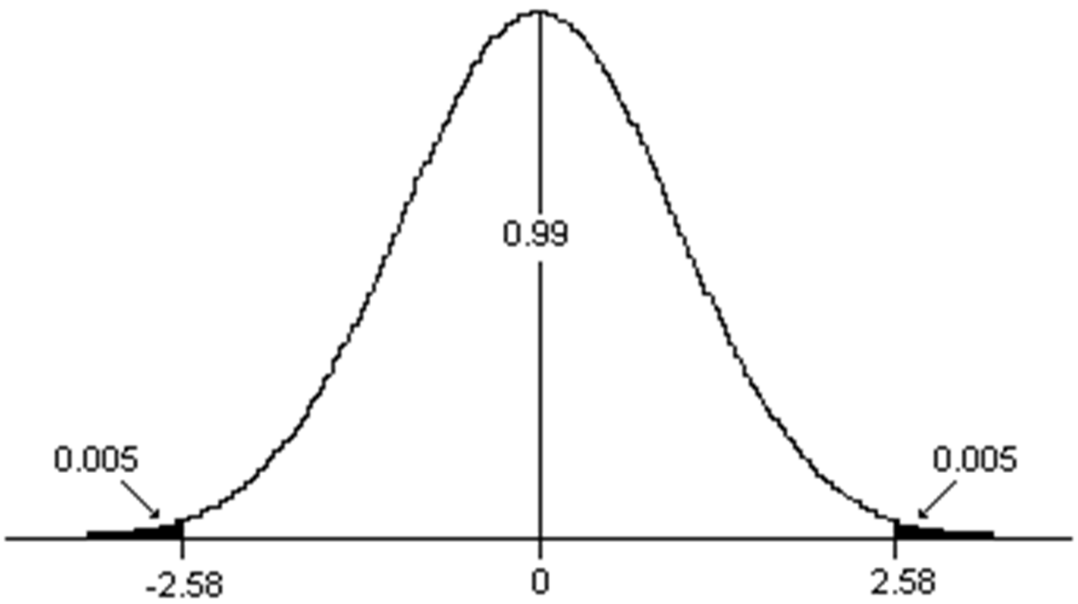
EXEMPLO 7

- Para determinar a incidência de uma determinada doença genética no Norte de Portugal, foi recolhida uma amostra de gotas de sangue de 500 bebés, nascidos no ano de 1994.

- As análises permitiram detectar 37 bebés portadores da doença.

- Estime um intervalo de confiança de 99% para a proporção de portadores da doença.

SOLUÇÃO 7



SOLUÇÃO 7

$$\hat{p} = \frac{37}{500} = 0.074$$

$$0.074 \pm 2.58\sqrt{\frac{0.074(1-0.074)}{500}}$$

$$0.074 \pm 0.030$$

$$0.044 < p < 0.104$$

Estimativa para a diferença de proporções $p_1 - p_2$

$$\boxed{Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sigma_{\hat{p}_1 - \hat{p}_2}} \sim N(0,1)}$$

➡ $\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

➡ $\hat{p}_1 = \dfrac{x_1}{n_1}$ e $\hat{p}_2 = \dfrac{x_2}{n_2}$

$$\boxed{(\hat{p}_1 - \hat{p}_2) - z_{(1-\alpha/2)}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{(1-\alpha/2)}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

Exemplo:

- Quando um sinal de limite de velocidade de 50km/h foi colocado numa estrada, numa amostra de 100 veículos, 49 violaram o limite de velocidade. Quando o limite foi aumentado para 60 km/h, duma amostra de 100 veículos, 19 ultrapassaram o novo limite. Encontre um intervalo de confiança de 99% para $p_1 - p_2$ e interprete o seu resultado.

$$\hat{p}_1 = \frac{49}{100} = 0{,}49 \text{ e } \hat{p}_2 = \frac{19}{100} = 0{,}19$$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{0,49(1-0,49)}{100} + \frac{0{,}19(1-0{,}19)}{100}} = 0{,}0635$$

$$z_{0.995} \overset{tab.5}{=} 2{,}575$$

$$0{,}30 - 0{,}164 < p_1 - p_2 < 0{,}30 + 0{,}164$$

$$0{,}136 < p_1 - p_2 < 0{,}464$$

Estimativa para o desvio padrão, σ

1. **Grandes amostras** $n \geq 100$

Se a variável em estudo apresenta uma distribuição Normal então o intervalo de confiança será:

$$\boxed{s_n - z_{1-\alpha/2} \cdot \frac{0,71.s_n}{\sqrt{n}} < \sigma < s_n + z_{1-\alpha/2} \cdot \frac{0,71.s_n}{\sqrt{n}}}$$

2. **Pequenas amostras** $n < 100$

Se a variável em estudo apresenta uma distribuição Normal então o intervalo de confiança será:

$$\boxed{\frac{(n-1)s_{n-1}^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \frac{(n-1)s_{n-1}^2}{\chi^2_{n-1,1-\alpha/2}}}$$

Para o desvio padrão poder-se-á escrever:

$$\boxed{\sqrt{\frac{(n-1)s_{n-1}^2}{\chi^2_{n-1,\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)s_{n-1}^2}{\chi^2_{n-1,1-\alpha/2}}}}$$

Estimativa para o quociente das variâncias σ_1^2 / σ_2^2

Considerando duas populações, de onde são retiradas as amostras aleatórias, independentes e com distribuições aproximadamente normais, o intervalo de confiança (1-)100% para o quociente das variâncias será dado por.

$$\boxed{\frac{s_1^2}{s_2^2}\frac{1}{F_{\alpha/2,n_1-1,n_2-1}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2}F_{\alpha/2,n_2-1,n_1-1}}$$

onde $F_{\alpha/2,n_1-1,n_2-1}$ é o valor que localiza uma área de $\alpha/2$ na cauda superior da distribuição F com n_1-1 no numerador e n_2-1 graus de liberdade no denominador e $F_{\alpha/2,n_2-1,n_1-1}$ é o valor que localiza um área de $\alpha/2$ na cauda superior da distribuição F com n_2-1 no numerador e n_1-1 graus de liberdade no denominador.

TESTES DE HIPÓTESES

OBJECTIVO

Verificar se os dados amostrais (ou estimativas obtidas a partir deles) são ou não compatíveis com determinadas populações (ou valores previamente fixados dos parâmetros populacionais).

PROCEDIMENTO

- **Definição das hipóteses**
  - **H_0 – hipótese nula**
  - **H_1 – hipótese alternativa**
- **Identificação da estatística de teste (ET)**
- **Definição da regra de decisão**
- **Cálculo da estatística de teste e tomada de decisão**

Critérios para um teste de hipóteses

Decisão / Hipótese	Não rejeitar H_0	Rejeitar H_0
H_0 é verdadeira	Decisão correcta	Erro Tipo I
H_0 é falsa	Erro Tipo II	Decisão correcta

ERROS

- α = P(erro tipo I) = P(rej. $H_0|H_0$)

- β = P(erro tipo II)=P(não rej $H_0|H_1$)

Função potência de um teste

$$\boldsymbol{k}(\boldsymbol{\theta}) = \begin{cases} \boldsymbol{\alpha} & H_0 \\ \boldsymbol{1} - \boldsymbol{\beta} & H_1 \end{cases}$$

DECISÃO

- Se o valor calculado para a ET > ET(α) então rejeita-se H_0 (valor p < α)

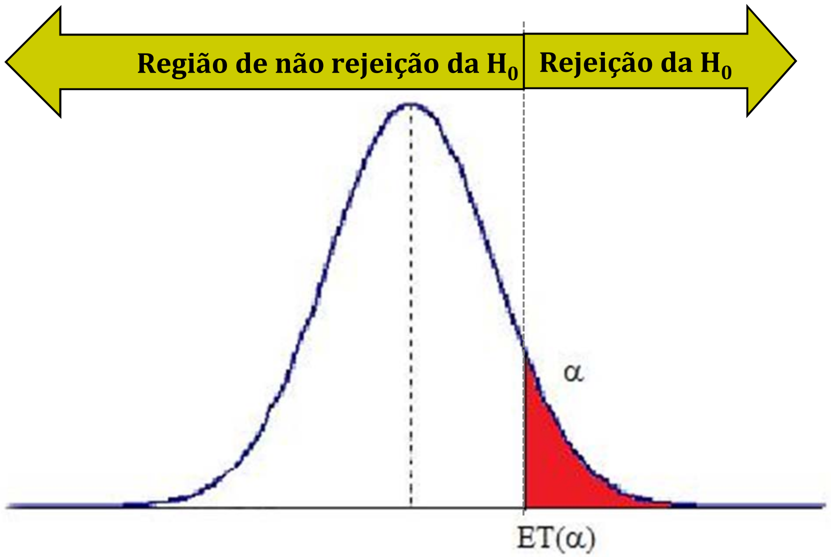
- Se o valor calculado para a ET < ET(α) então não se rejeita H_0; (valor p > α) - resultado inconclusivo.

TESTE DE SIGNIFICÂNCIA

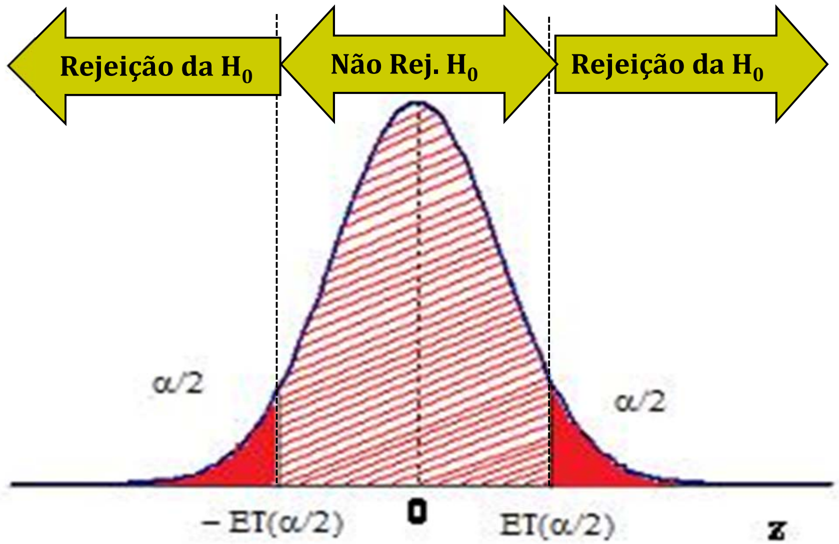
- Se a diferença entre o que esperamos e o que obtemos é tão grande que não pode ser atribuída ao acaso, nós rejeitamos a hipótese na qual baseamos as nossas expectativas.

- Se a diferença entre o que esperamos e o que obtemos é tão pequena que pode ser realmente atribuída ao acaso, nós não rejeitamos a hipótese na qual baseamos as nossas expectativas, dizemos que o resultado do teste não é (estatisticamente) significativo.

Teste de hipóteses unilateral



Testes de hipóteses bilateral



MÉDIA

Teste Unilateral

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0$$

$$\left(H_1 : \mu < \mu_0\right)$$

Teste Bilateral

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Estatística

$$z = \frac{\overline{x} - \mu_0}{\sigma / \sqrt{n}} \approx \frac{\overline{x} - \mu_0}{s / \sqrt{n}}$$

Região de Rejeição

$$z > z_{1-\alpha}$$

$$\left(z < -z_{1-\alpha}\right)$$

Região de Rejeição

$$|z| > z_{1-\alpha/2}$$

EXEMPLO 1

- Suponha que a Inspecção das Actividades Económicas quer verificar se os sacos de cimento de uma determinada fábrica têm um peso médio de 15 Kg. Para tal recolheu uma amostra aleatória de 50 sacos, tendo encontrado uma média de 14.81 Kg com um desvio padrão de 0.62 Kg. Permitem os dados concluir que a fábrica está a fornecer sacos com um peso inferior ao especificado?. Assuma α=0.05.

SOLUÇÃO 1

- 1. Formulação das hipóteses

$$H_0 : \mu = 15$$

$$H_1 : \mu < 15$$

- 2. Região crítica

$$z < -z_{0.95} = -1.65$$

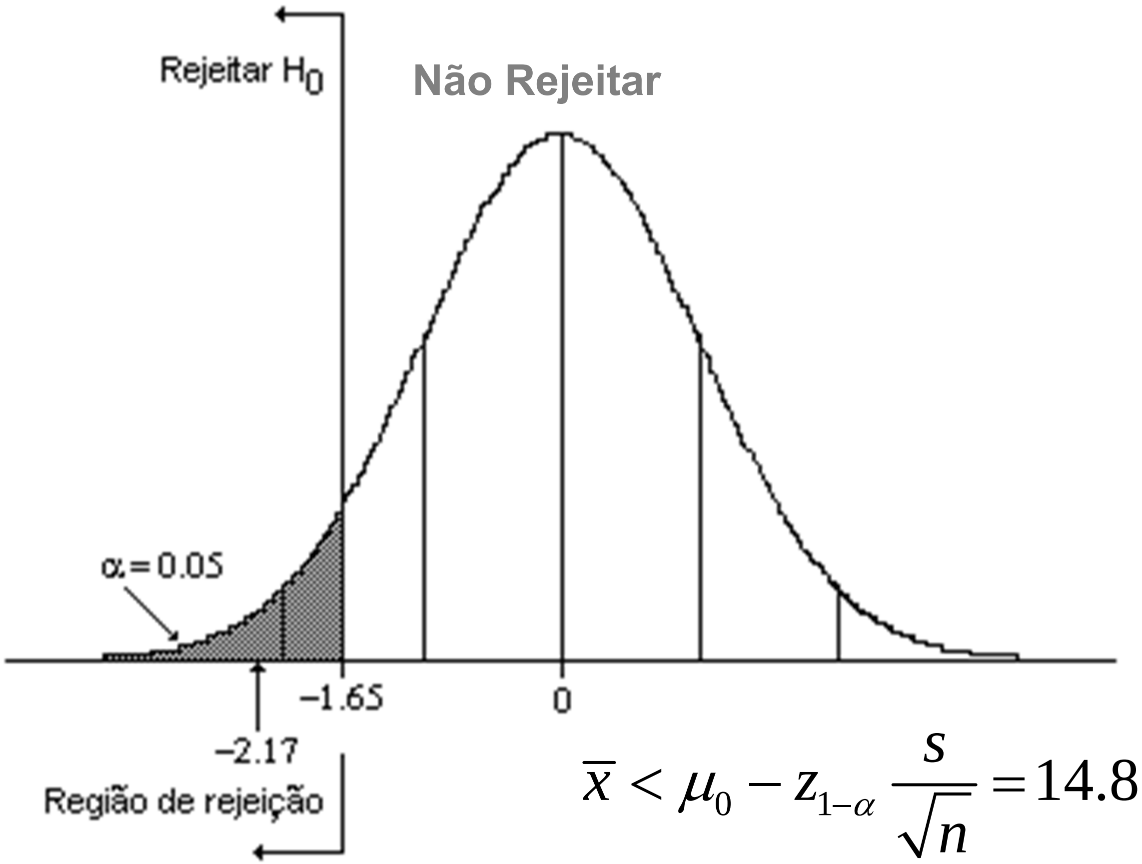
- 3. Teste estatístico

$$z \approx \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{14.81 - 15}{0.62/\sqrt{50}} = -2.17$$

- 4. Decisão

  Rejeitar a Hipótese Nula, ou seja, os sacos têm, em média, um peso inferior a 15 Kg.

SOLUÇÃO 1



$$\bar{x} < \mu_0 - z_{1-\alpha}\frac{s}{\sqrt{n}} = 14.86$$

MÉDIA

Teste Unilateral

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0$$

$$\left(H_1 : \mu < \mu_0\right)$$

Teste Bilateral

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Estatística

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

Região de Rejeição

$$t > t_\alpha$$

$$\left(t < -t_\alpha\right)$$

Região de Rejeição

$$|t| > t_{\alpha/2}$$

EXEMPLO 2

- Uma máquina produz seringas com 2.5 cm de comprimento. No entanto, se as seringas forem demasiado curtas ou longas, serão rejeitadas. Neste caso a máquina necessita de ser ajustada. Para tal, uma amostra de seringas é recolhida, a intervalos regulares, para verificar se as mesmas estão a ser produzidas com o comprimento médio de 2.5 cm.
- Suponha que foi recolhida uma amostra de 16 seringas, com uma média $\overline{x}$=2.52 cm e um desvio padrão s=0.04 cm.
- Há evidência suficiente para assumir que a máquina não está a produzir segundo as especificação, isto é, que a máquina está fora de controlo? Use α=0.01.

SOLUÇÃO 2

- 1. Formulação das hipóteses

$$H_0 : \mu = 2.5$$

$$H_1 : \mu \neq 2.5$$

- 2. Região crítica

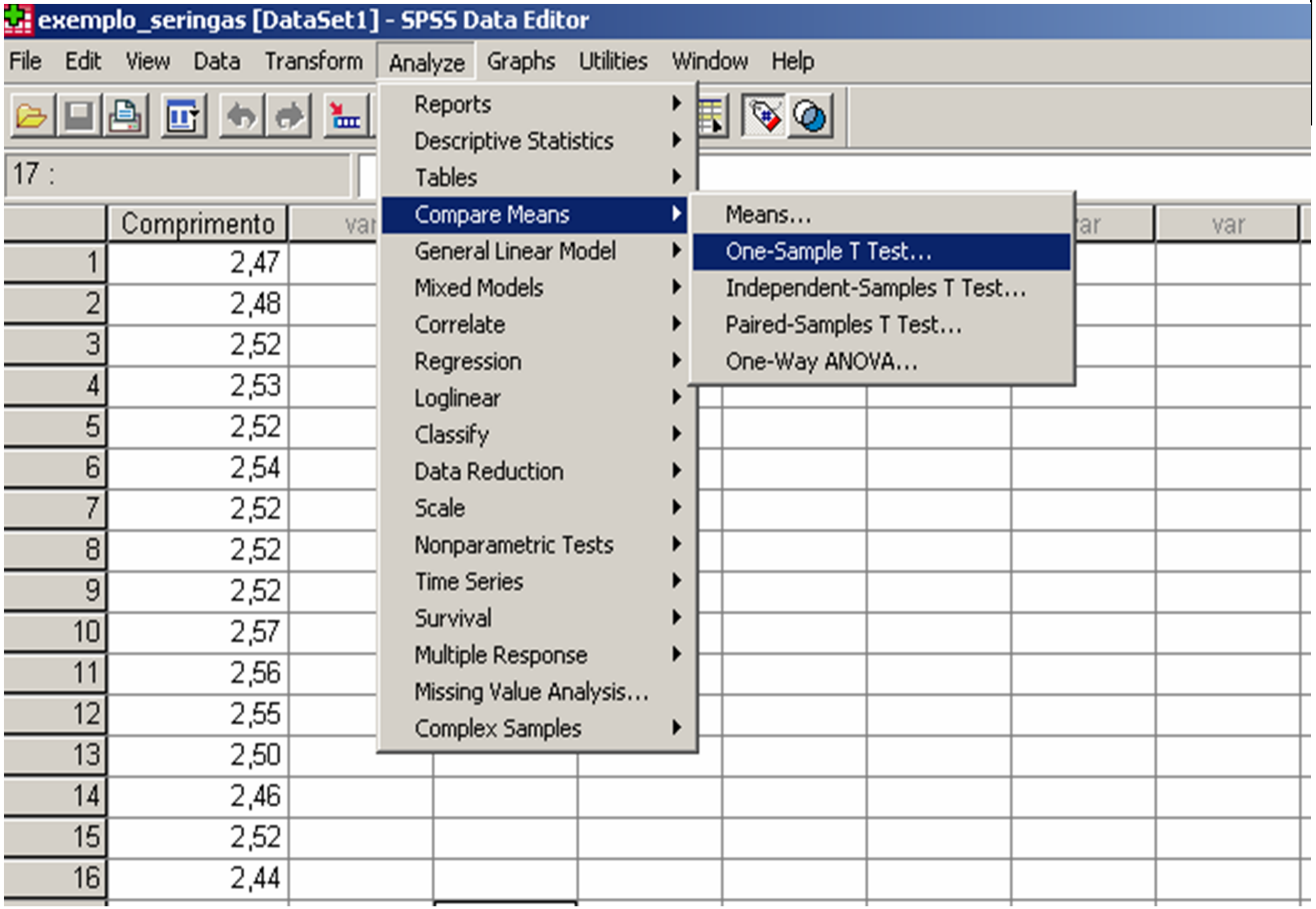
$$|t| \geq t_{0.005} = 2.947$$

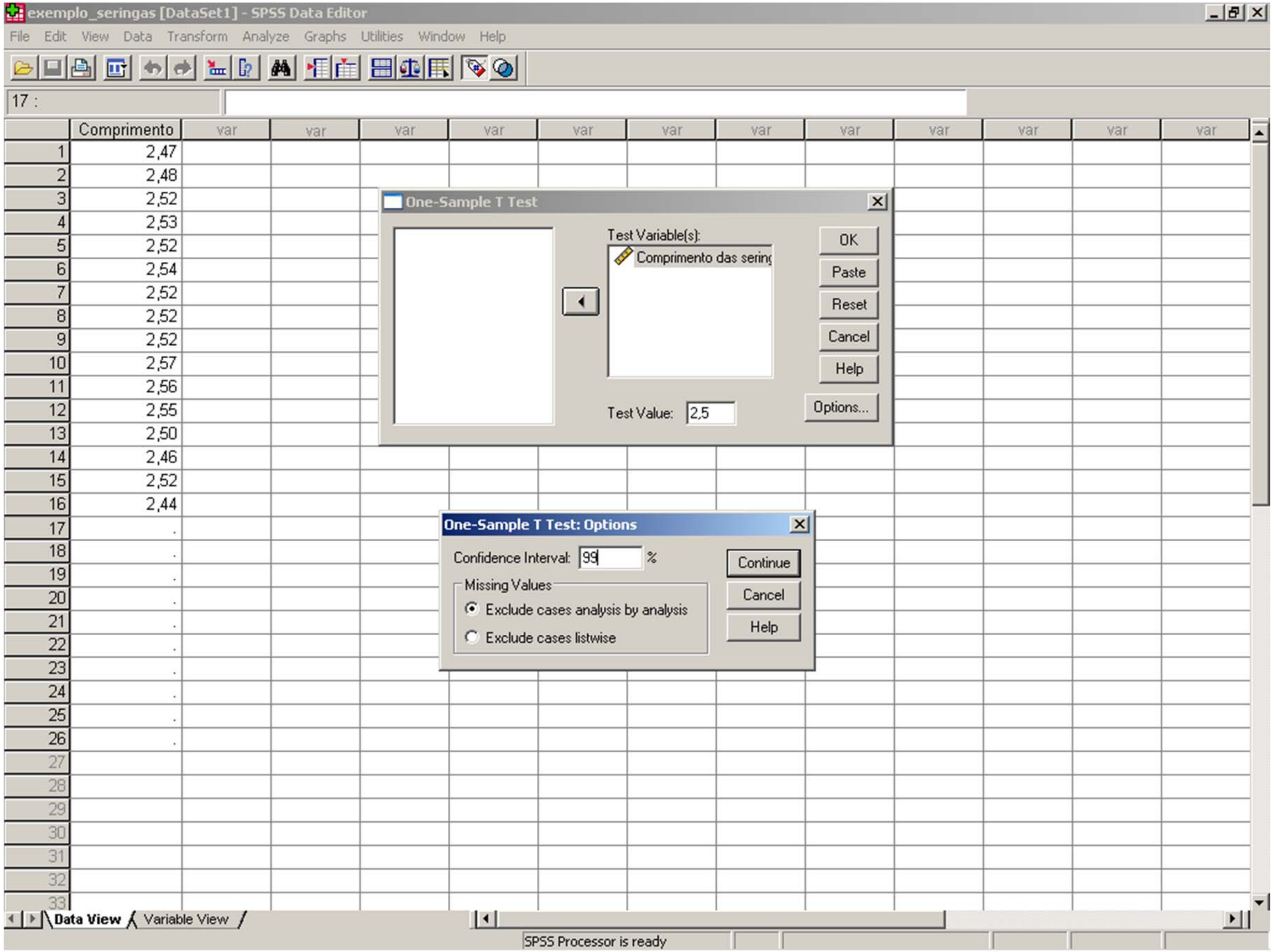
- 3. Teste estatístico

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{2.52 - 2.5}{0.04/\sqrt{16}} = 1.00$$

- 4. Decisão

  Não rejeitar a Hipótese Nula, ou seja, as seringas podem apresentar um comprimento médio de 2.5 cm.


exemplo_seringas [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
17 :
Comprimento
var
var
1 2,47
2 2,48
3 2,52
4 2,53
5 2,52
6 2,54
7 2,52
8 2,52
9 2,52
10 2,57
11 2,56
12 2,55
13 2,50
14 2,46
15 2,52
16 2,44
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples
Means...
One-Sample T Test...
Independent-Samples T Test...
Paired-Samples T Test...
One-Way ANOVA...


exemplo_seringas [DataSet1] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
17 :
Comprimento
var
1 2,47
2 2,48
3 2,52
4 2,53
5 2,52
6 2,54
7 2,52
8 2,52
9 2,52
10 2,57
11 2,56
12 2,55
13 2,50
14 2,46
15 2,52
16 2,44
One-Sample T Test
Test Variable(s):
Comprimento das sering
OK
Paste
Reset
Cancel
Help
Test Value: 2,5
Options...
One-Sample T Test: Options
Confidence Interval: 99 %
Missing Values
Exclude cases analysis by analysis
Exclude cases listwise
Continue
Cancel
Help
Data View
Variable View
SPSS Processor is ready

Output2 - SPSS Viewer

```
T-TEST
  /TESTVAL = 2.5
  /MISSING = ANALYSIS
  /VARIABLES = Comprimento
  /CRITERIA = CI(.99) .
```

T-Test

[DataSet1]

One-Sample Statistics

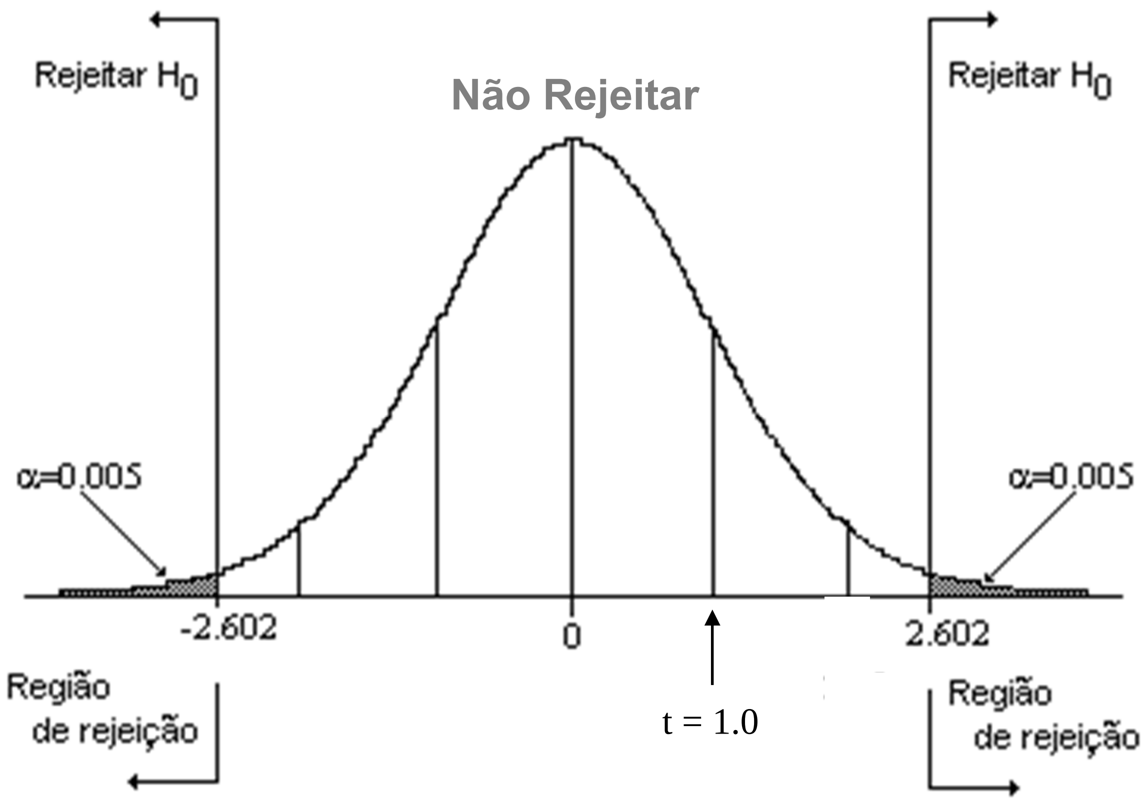
	N	Mean	Std. Deviation	Std. Error Mean
Comprimento das seringas	16	2,5138	,03594	,00898

One-Sample Test

	Test Value = 2.5					
					99% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Comprimento das seringas	1,530	15	,147	,01375	-,0127	,0402

SPSS Processor is ready

SOLUÇÃO 2



Rejeitar H_0
Não Rejeitar
Rejeitar H_0
α=0.005
α=0.005
-2.602
0
2.602
Região
de rejeição
t = 1.0
Região
de rejeição

DIFERENÇA DE MÉDIAS AMOSTRAS INDEPENDENTES

Teste Unilateral

$$H_0 : (\mu_1 - \mu_2) = D_0$$

$$H_1 : (\mu_1 - \mu_2) > D_0$$

$$\left(H_1 : (\mu_1 - \mu_2) < D_0\right)$$

Teste Bilateral

$$H_0 : (\mu_1 - \mu_2) = D_0$$
$$H_1 : (\mu_1 - \mu_2) \neq D_0$$

Estatística

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - D_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

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

Região de Rejeição

$$t > t_\alpha$$

$$(t < -t_\alpha)$$

Região de Rejeição

$$|t| > t_{\alpha/2}$$

EXEMPLO 4

- O desgaste da cabeça do fémur conduz à implantação de uma cabeça de substituição numa liga metálica leve e resistente. Esta implantação é feita com um cimento especial, que alguns médicos suspeitam que possa diminuir a resistência do osso. No entanto, as opiniões dos ortopedistas dividem-se quanto à necessidade de introdução de um tampão que evite que o cimento se espalhe pelo espaço disponível. Para comparar o efeito do uso do tampão na resistência à flexão, foram efectuados vários implantes em animais de laboratório, tendo sido obtidos os seguintes resultados de resistência (Nm):

C/ Tampão	7.0	6.2	7.1	8.1	5.1	5.6
S/ Tampão	8.9	7.7	5.3	8.6	7.1	4.6

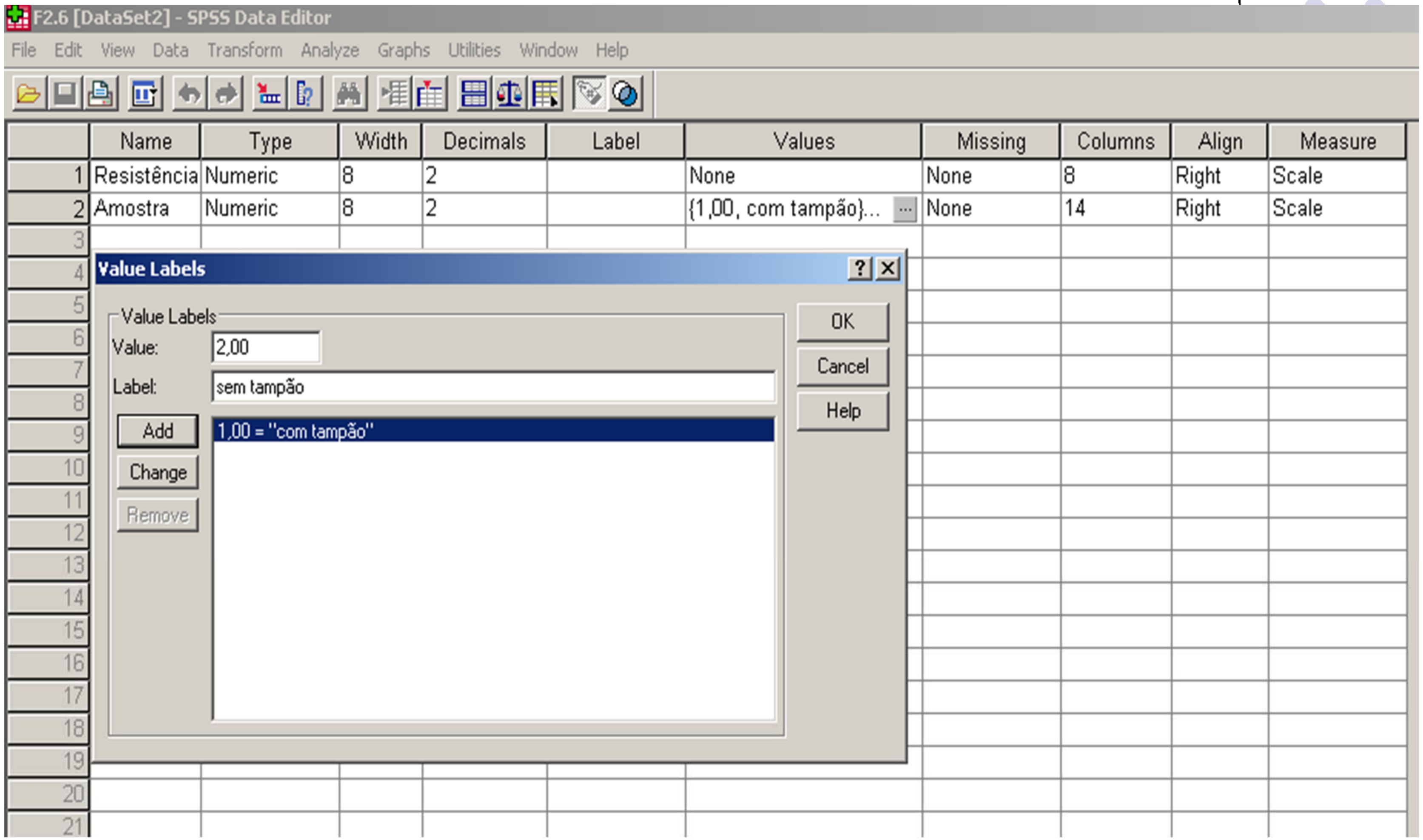
O que pode concluir acerca do efeito do tampão na resistência à flexão? Use $\alpha=0.05$.

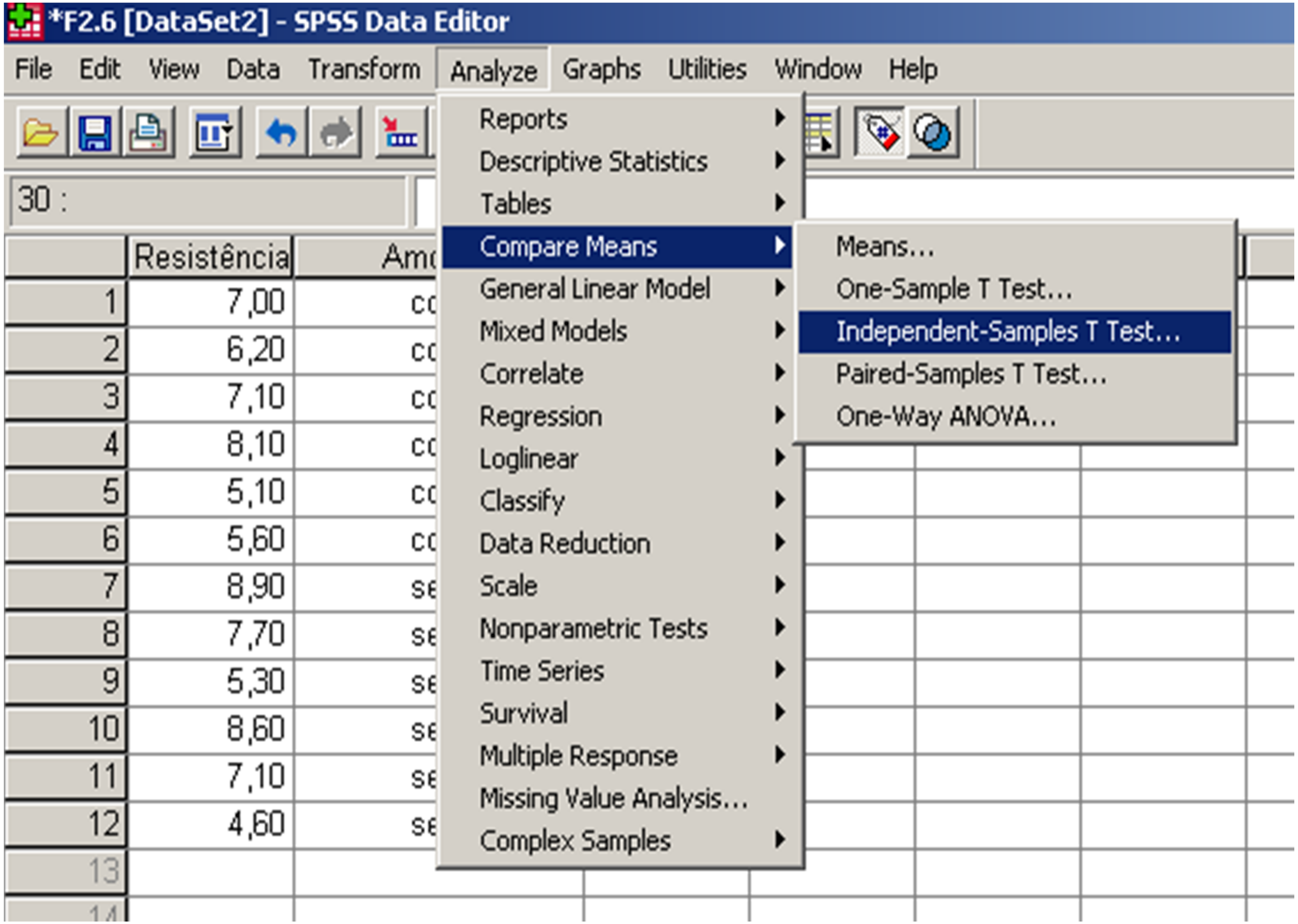
F2.6 [DataSet2] - SPSS Data Editor

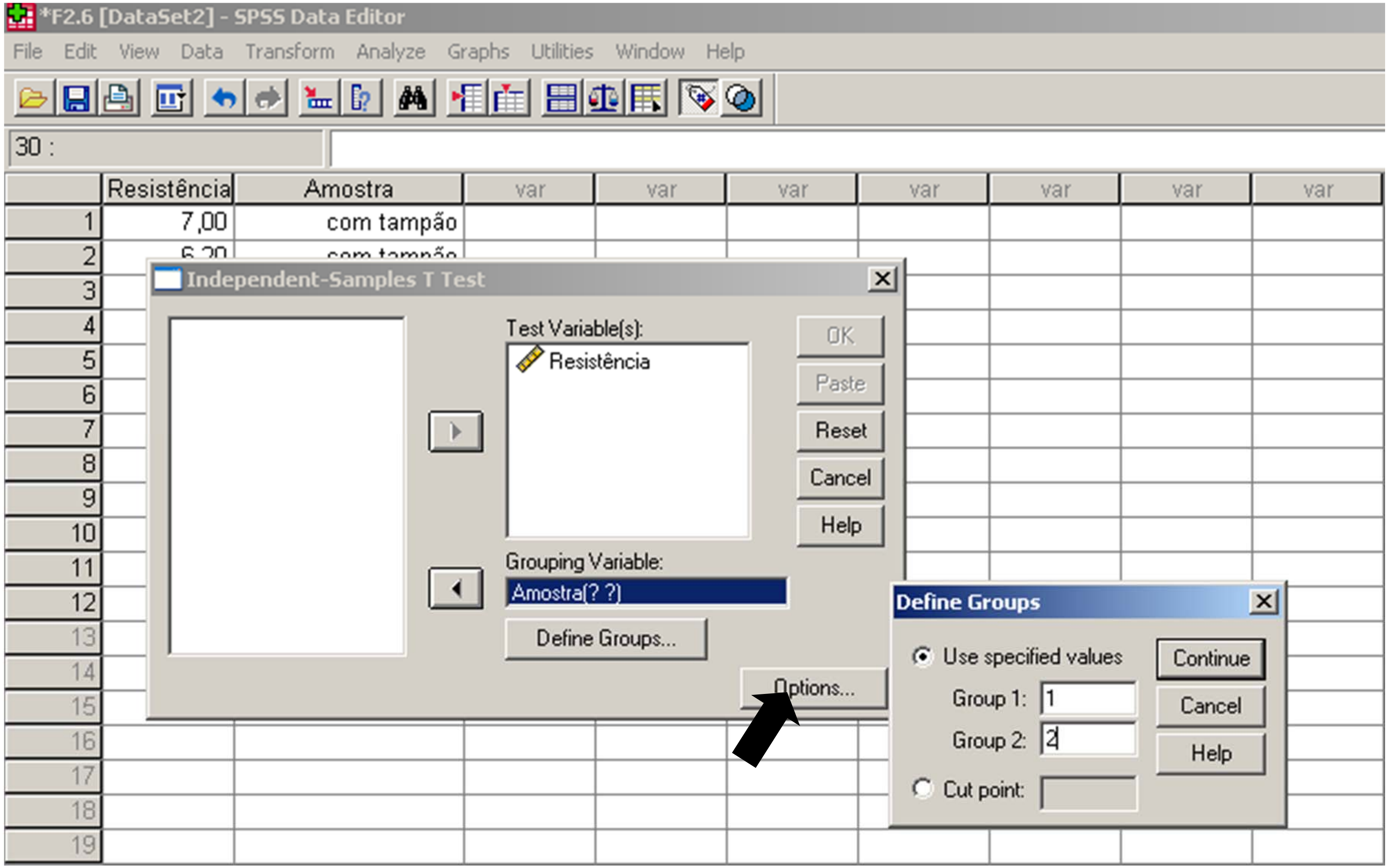
File Edit View Data Transform Analyze Graphs Utilities Window Help

30 :

	Resistência	Amostra	var	var	var
1	7,00	com tampão			
2	6,20	com tampão			
3	7,10	com tampão			
4	8,10	com tampão			
5	5,10	com tampão			
6	5,60	com tampão			
7	8,90	sem tampão			
8	7,70	sem tampão			
9	5,30	sem tampão			
10	8,60	sem tampão			
11	7,10	sem tampão			
12	4,60	sem tampão			
13					


F2.6 [DataSet2] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
Name Type Width Decimals Label Values Missing Columns Align Measure
1 Resistência Numeric 8 2 None None 8 Right Scale
2 Amostra Numeric 8 2 {1,00, com tampão}... None 14 Right Scale
Value Labels
Value Labels
Value: 2,00
Label: sem tampão
Add
Change
Remove
1,00 = "com tampão"
OK
Cancel
Help


*F2.6 [DataSet2] - SPSS Data Editor
File
Edit
View
Data
Transform
Analyze
Graphs
Utilities
Window
Help
30 :
Resistência
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Mixed Models
Correlate
Regression
Loglinear
Classify
Data Reduction
Scale
Nonparametric Tests
Time Series
Survival
Multiple Response
Missing Value Analysis...
Complex Samples
Means...
One-Sample T Test...
Independent-Samples T Test...
Paired-Samples T Test...
One-Way ANOVA...
7,00
6,20
7,10
8,10
5,10
5,60
8,90
7,70
5,30
8,60
7,10
4,60



*F2.6 [DataSet2] - SPSS Data Editor
File Edit View Data Transform Analyze Graphs Utilities Window Help
30 :
Resistência
Amostra
var
7,00
com tampão
Independent-Samples T Test
Test Variable(s):
Resistência
OK
Paste
Reset
Cancel
Help
Grouping Variable:
Amostra(? ?)
Define Groups...
Options...
Define Groups
Use specified values
Group 1: 1
Group 2: 2
Cut point:
Continue
Cancel
Help

SOLUÇÃO 4

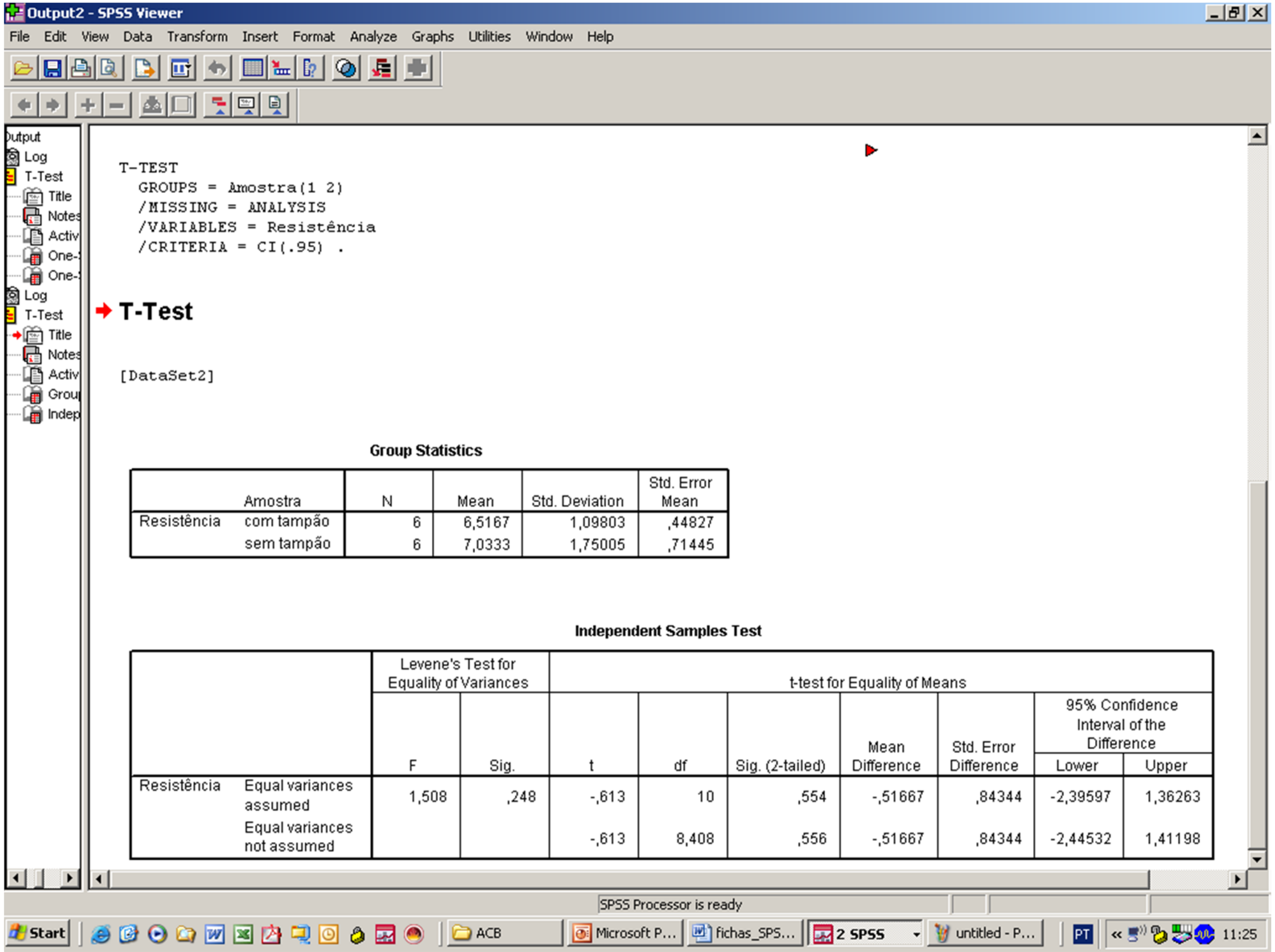
$$\bar{x}_1 = 6.517 \quad s_1 = 1.098$$
$$\bar{x}_2 = 7.033 \quad s_2 = 1.750$$

- 1. Formulação das hipóteses

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

- 2. Região crítica

$$|t| \geq t_{0.025} = 2.228$$



```
T-TEST
  GROUPS = Amostra(1 2)
  /MISSING = ANALYSIS
  /VARIABLES = Resistência
  /CRITERIA = CI(.95) .
```

T-Test

[DataSet2]

Group Statistics

	Amostra	N	Mean	Std. Deviation	Std. Error Mean
Resistência	com tampão	6	6,5167	1,09803	,44827
	sem tampão	6	7,0333	1,75005	,71445

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Resistência	Equal variances assumed	1,508	,248	-,613	10	,554	-,51667	,84344	-2,39597	1,36263
	Equal variances not assumed			-,613	8,408	,556	-,51667	,84344	-2,44532	1,41198

DIFERENÇA DE MÉDIAS AMOSTRAS EMPARELHADAS

Teste Unilateral

$$H_0 : (\mu_1 - \mu_2) = D_0$$
$$H_1 : (\mu_1 - \mu_2) > D_0$$
$$\left(H_1 : (\mu_1 - \mu_2) < D_0\right)$$

Teste Bilateral

$$H_0 : (\mu_1 - \mu_2) = D_0$$
$$H_1 : (\mu_1 - \mu_2) \neq D_0$$

Estatística

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

Região de Rejeição

$$t > t_\alpha$$
$$\left(t < -t_\alpha\right)$$

Região de Rejeição

$$|t| > t_{\alpha/2}$$

PROPORÇÃO

Teste Unilateral

$$H_0 : p = p_0$$

$$H_1 : p > p_0$$

$$\left(H_1 : p < p_0\right)$$

Teste Bilateral

$$H_0 : p = p_0$$

$$H_1 : p \neq p_0$$

Estatística

$$z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0\left(1 - p_0\right)}{n}}}$$

Região de Rejeição

$$z > z_{1-\alpha}$$

$$\left(z < -z_{1-\alpha}\right)$$

Região de Rejeição

$$|z| > z_{1-\alpha/2}$$

DIFERENÇA DE PROPORÇÕES

Teste Unilateral

$$H_0 : (p_1 - p_2) = D_0$$
$$H_1 : (p_1 - p_2) > D_0$$
$$(H_1 : (p_1 - p_2) < D_0)$$

Teste Bilateral

$$H_0 : (p_1 - p_2) = D_0$$
$$H_1 : (p_1 - p_2) \neq D_0$$

Estatística

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

se $D_0 \neq 0$

Região de Rejeição

$$z > z_{1-\alpha}$$
$$(z < -z_{1-\alpha})$$

Região de Rejeição

$$|z| > z_{1-\alpha/2}$$

DIFERENÇA DE PROPORÇÕES

Teste Unilateral

$$H_0 : (p_1 - p_2) = 0$$
$$H_1 : (p_1 - p_2) > 0$$
$$\left(H_1 : (p_1 - p_2) < 0\right)$$

Teste Bilateral

$$H_0 : (p_1 - p_2) = 0$$
$$H_1 : (p_1 - p_2) \neq 0$$

Estatística

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \text{sendo } \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Região de Rejeição

$$z > z_{1-\alpha}$$
$$\left(z < -z_{1-\alpha}\right)$$

Região de Rejeição

$$|z| > z_{1-\alpha/2}$$

VARIÂNCIA

Teste Unilateral

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_1 : \sigma^2 > \sigma_0^2$$

$$\left(H_1 : \sigma^2 < \sigma_0^2\right)$$

Teste Bilateral

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_1 : \sigma^2 \neq \sigma_0^2$$

Estatística

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Região de Rejeição

$$\chi^2 > \chi_\alpha^2 \left(\chi^2 < \chi_{1-\alpha}^2\right)$$

Região de Rejeição

$$\chi^2 < \chi_{1-\alpha}^2 \quad \text{ou} \quad \chi^2 > \chi_\alpha^2$$

EXEMPLO 5

As garrafas de refrigerantes contêm um volume aproximado de 33 cl. O produtor perderá dinheiro se as garrafas contiverem muito mais do que o volume especificado, e correrá o risco de ser multado, se o volume for bastante inferior. Assim, é necessário controlar a variação do volume de enchimento das garrafas. Se a variância for superior a 0.25 o processo está fora de controlo, e a máquina de enchimento deve ser ajustada Para tal, um controlador de qualidade recolhe uma amostra de 15 garrafas, com um enchimento médio de 33.15 cl e um desvio padrão de 0.71 cl. Face a estes resultados pode-se concluir que o processo está controlado?

SOLUÇÃO 5

1. Formulação das hipóteses

$$H_0 : \sigma^2 = 0.25$$

$$H_1 : \sigma^2 > 0.25$$

2. Região crítica

$$\chi^2 \geq \chi^2_{0.05} = 23.685$$

3. Teste estatístico

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{14(0.71)^2}{0.25} = 28.230$$

4. Decisão

Rejeitar a Hipótese Nula, ou seja, o processo de enchimento está fora de controlo.

RAZÃO DE VARIÂNCIAS

Teste Unilateral

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \Rightarrow \left(\sigma_1^2 = \sigma_2^2\right)$$

$$H_1 : \frac{\sigma_1^2}{\sigma_2^2} > 1 \Rightarrow \left(\sigma_1^2 > \sigma_2^2\right)$$

Estatística

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

$$\left(F = \frac{s_2^2}{s_1^2}\right)$$

Região de Rejeição

$$F > F_\alpha \qquad \left[H_1 : \frac{\sigma_1^2}{\sigma_2^2} < 1 \Rightarrow \left(\sigma_1^2 < \sigma_2^2\right)\right]$$

Teste Bilateral

$$H_0 : H_1 : \frac{\sigma_1^2}{\sigma_2^2} = 1 \Rightarrow \left(\sigma_1^2 = \sigma_2^2\right)$$

$$H_1 : H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1 \Rightarrow \left(\sigma_1^2 \neq \sigma_2^2\right)$$

Estatística

$$F = \begin{cases} \dfrac{s_1^2}{s_2^2} & s_1^2 > s_2^2 \\[2ex] \dfrac{s_2^2}{s_1^2} & s_1^2 < s_2^2 \end{cases}$$

Região de Rejeição

$$F > F_{\alpha/2}$$

EXEMPLO 9

- Uma das características importantes num vinho de qualidade é a constância do gosto no sabor e no aroma. A variabilidade no sabor depende do processo de vinificação, que pode incluir o controlo de variáveis como a temperatura, fermentação, qualidade das leveduras, etc. Um empresa quer ensaiar um novo processo de vinificação com o objectivo de reduzir a variabilidade no gosto, medido por índice. Duas amostras aleatórias, respectivamente de 25 e 15 copos, retiradas da produção dos dois processos de vinificação foram avaliadas por um painel de provadores, produzindo os seguintes resultados:

- Permitem os dados concluir que a variabilidade no segundo processo de vinificação é menor?

SOLUÇÃO 9

1. Formulação das hipóteses

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_1 : \frac{\sigma_1^2}{\sigma_2^2} > 1$$

2. Região crítica

$$F \geq F_{0.01,} = 2.35$$

3. Teste estatístico

$$F = \frac{s_1^2}{s_2^2} = \frac{(1.22)^2}{(0.72)^2} = 2.87$$

4. Decisão

   Rejeitar a Hipótese Nula, ou seja, a variabilidade no segundo processo de vinificação é menor.