Bioestatística Básica

RCA 5804

1. Experimentos no qual o sujeito recebe + de 1 tratamento
2. Alternativas para teste "T" e Análise de Variância
3. Intervalo de confiança da ≠ de medias
4. Alternativas para distribuições não-Gaussianas

Prof. Dr. Alfredo J Rodrigues

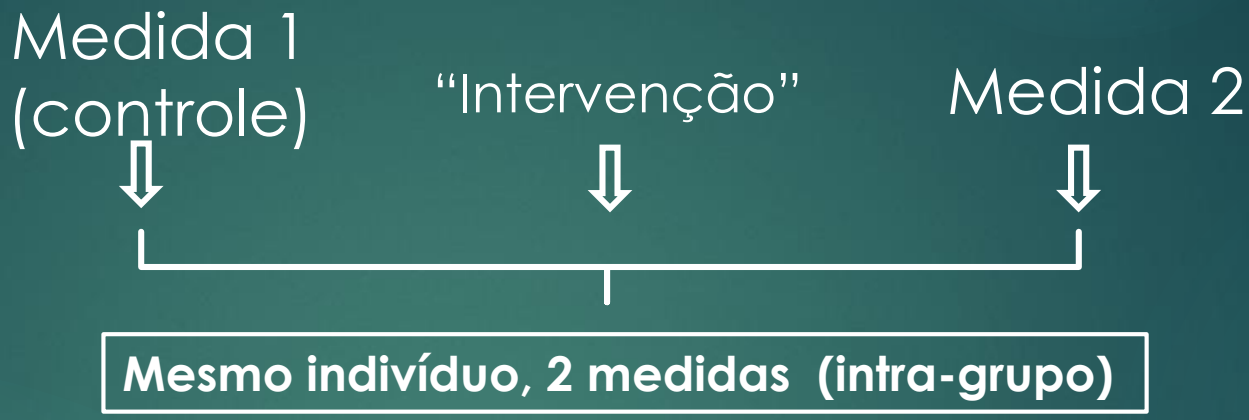
Departamento de Cirurgia e Anatomia

Faculdade de Medicina de Ribeirão Preto

Universidade de São Paulo

alfredo@fmrp.usp.br

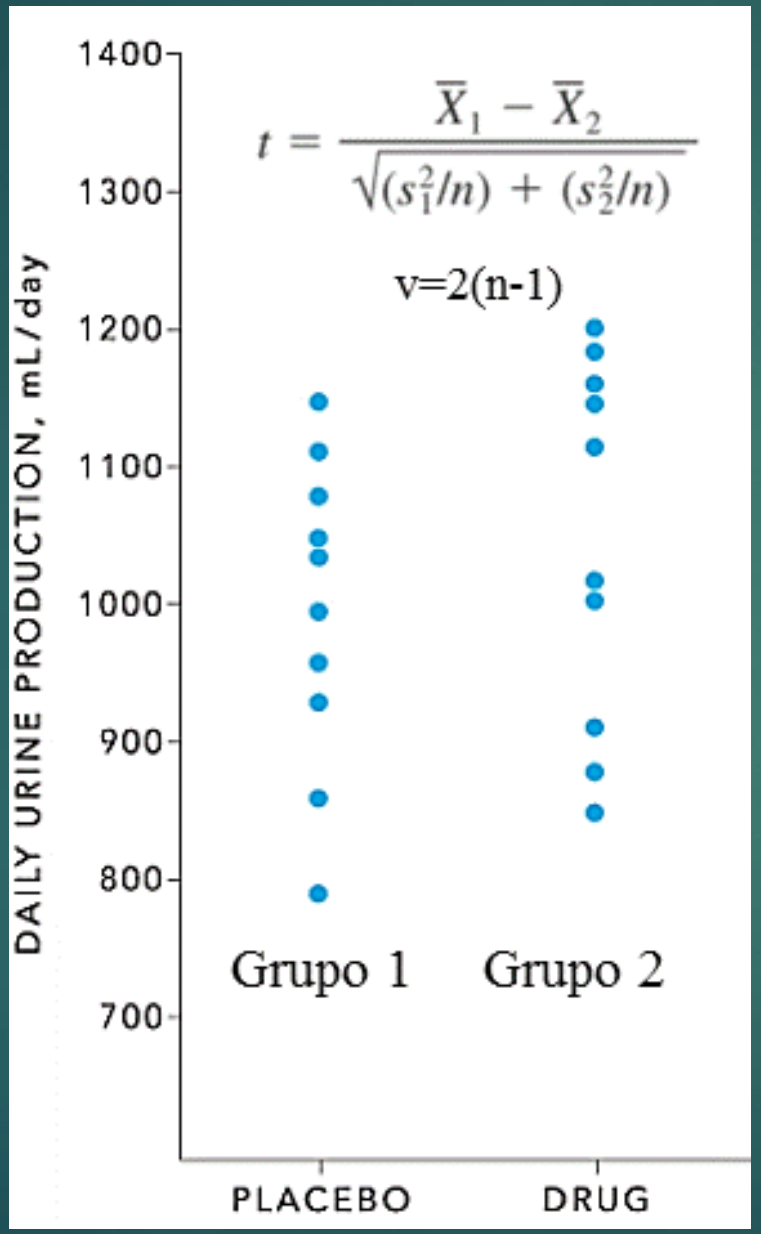
Antes e Após tratamento



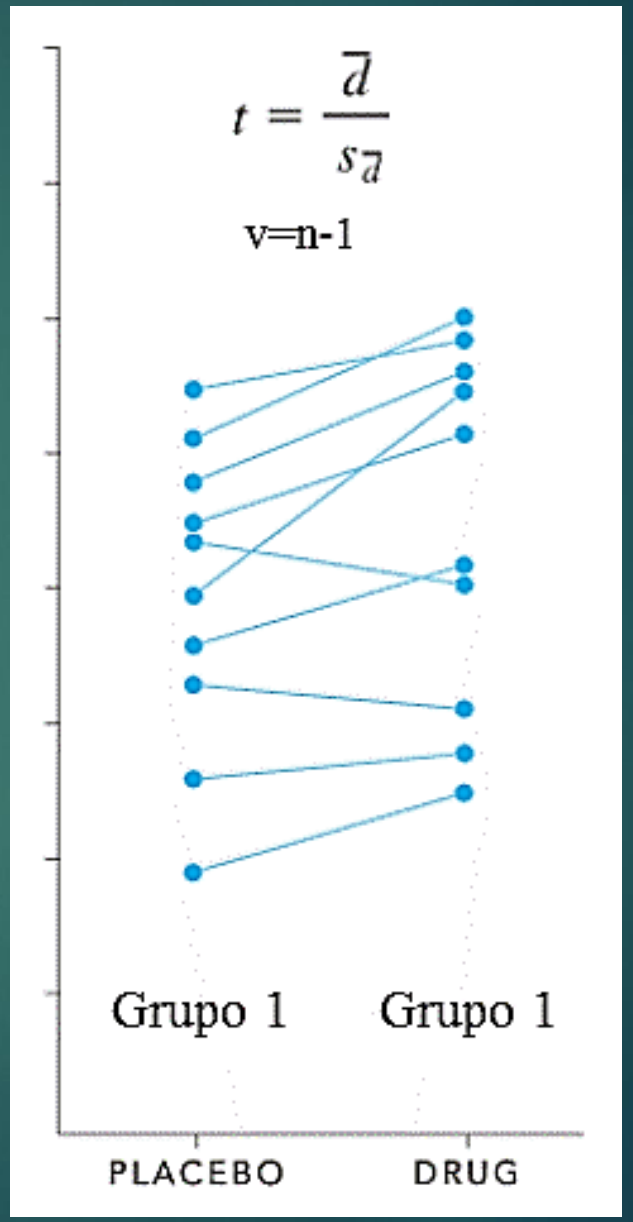
Teste "T" Pareadado:

- Característica do tipo de estudo:
  - O indivíduo é controle dele mesmo.
  - Evita influência da variabilidade biológica entre sujeitos no tratamento
  - O teste se interessa pela mudança e não pela ≠ provocada pelo tratamento.

Teste “T”



Teste “T” pareado



TESTE “t” (amostras independents)
vs
“t” Pareado (amostras dependentes)

EXPERIMENTO

1	10,00	12,00
2	11,00	13,00
3	12,00	13,00
4	10,00	11,00
5	9,00	11,00
6	6,00	7,00
7	8,00	9,00
8	9,00	11,00
9	12,00	14,00
10	10,00	12,00
Total N	10	10

TESTE " t" vs "t" Pareado

Teste "T" (amostras Independentes)

Case Summaries

	grupo 1	grupo 2
1	10,00	12,00
2	11,00	13,00
3	12,00	13,00
4	10,00	11,00
5	9,00	11,00
6	6,00	7,00
7	8,00	9,00
8	9,00	11,00
9	12,00	14,00
10	10,00	12,00
Total N	10	10

Teste "T" pareado (amostras Independentes)

Case Summaries

	antes	depois
1	10,00	12,00
2	11,00	13,00
3	12,00	13,00
4	10,00	11,00
5	9,00	11,00
6	6,00	7,00
7	8,00	9,00
8	9,00	11,00
9	12,00	14,00
10	10,00	12,00
Total N	10	10

Utilizando teste "T" independente

Group Statistics

	grupo	N	Mean	Std. Deviation	Std. Error Mean
grupo 1	1	10	9,7000	1,82878	,57831
	2	10	11,3000	2,05751	,65064

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
grupo 1	Equal variances assumed	,065	,802	-1,838	18	,083	-1,60000	,87050	-3,42886	,22886
	Equal variances not assumed			-1,838	17,756	,083	-1,60000	,87050	-3,43067	,23067

Utilizando teste"T" pareado

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	antes	9,7000	10	1,82878	,57831
	depois	11,3000	10	2,05751	,65064

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	antes - depois	-1,60000	,51640	,16330	-1,96941	-1,23059	-9,798	9	,000

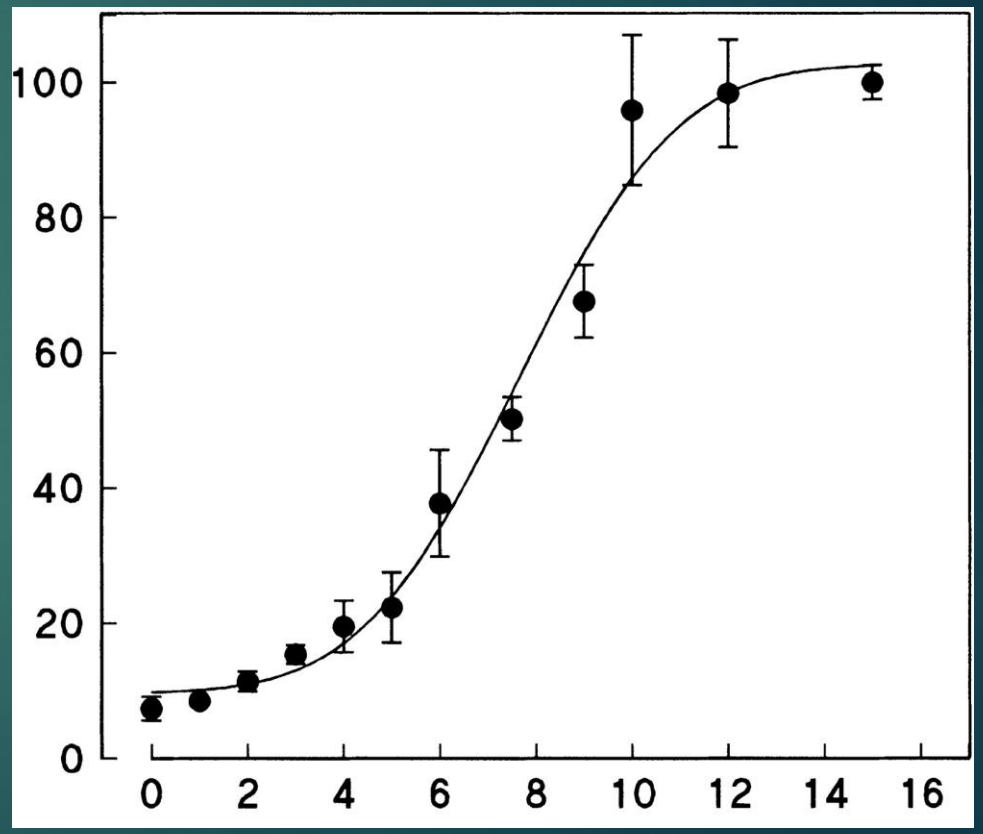
Análise de variância com medidas repetidas com apenas 1 grupo

Medida 1 controle

"Tratamento" (ou não)

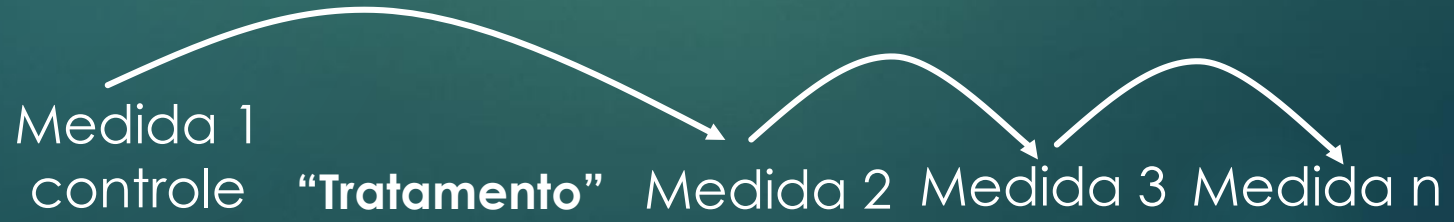
Medida 2 Medida 3 Medida n

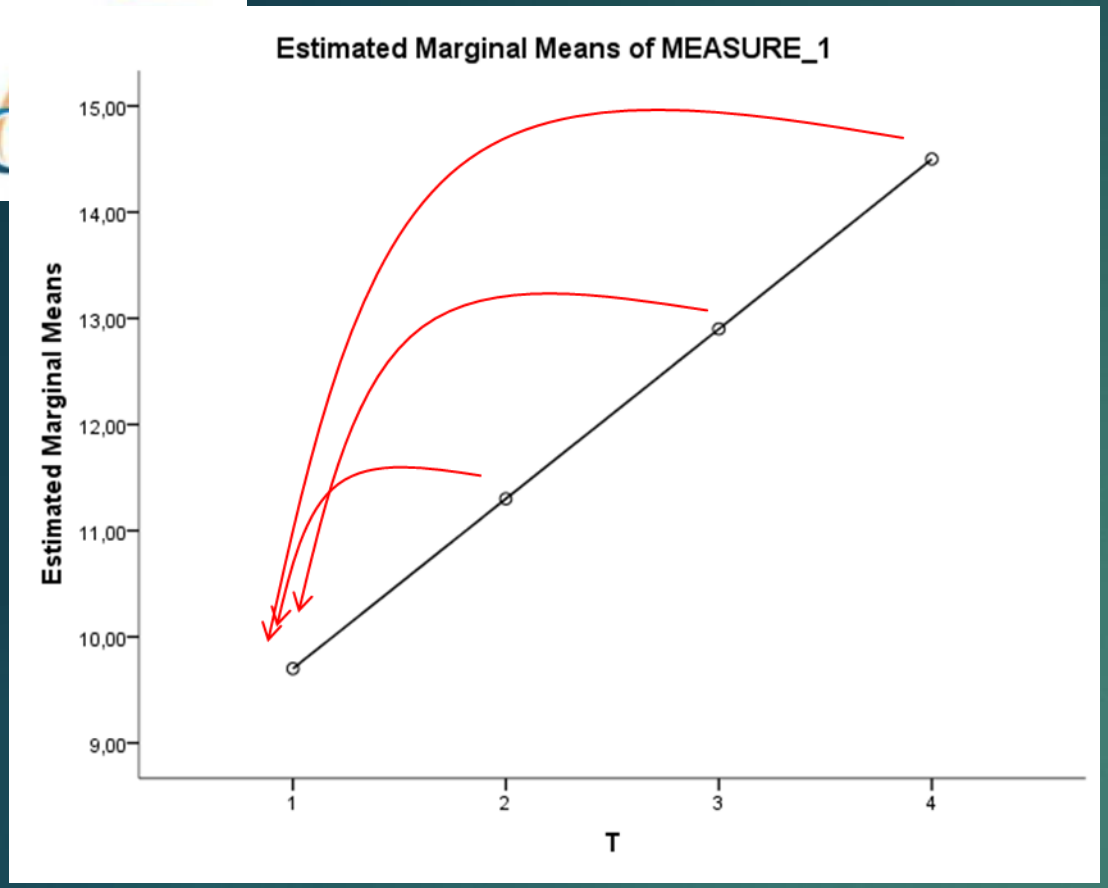
Mesmo indivíduo

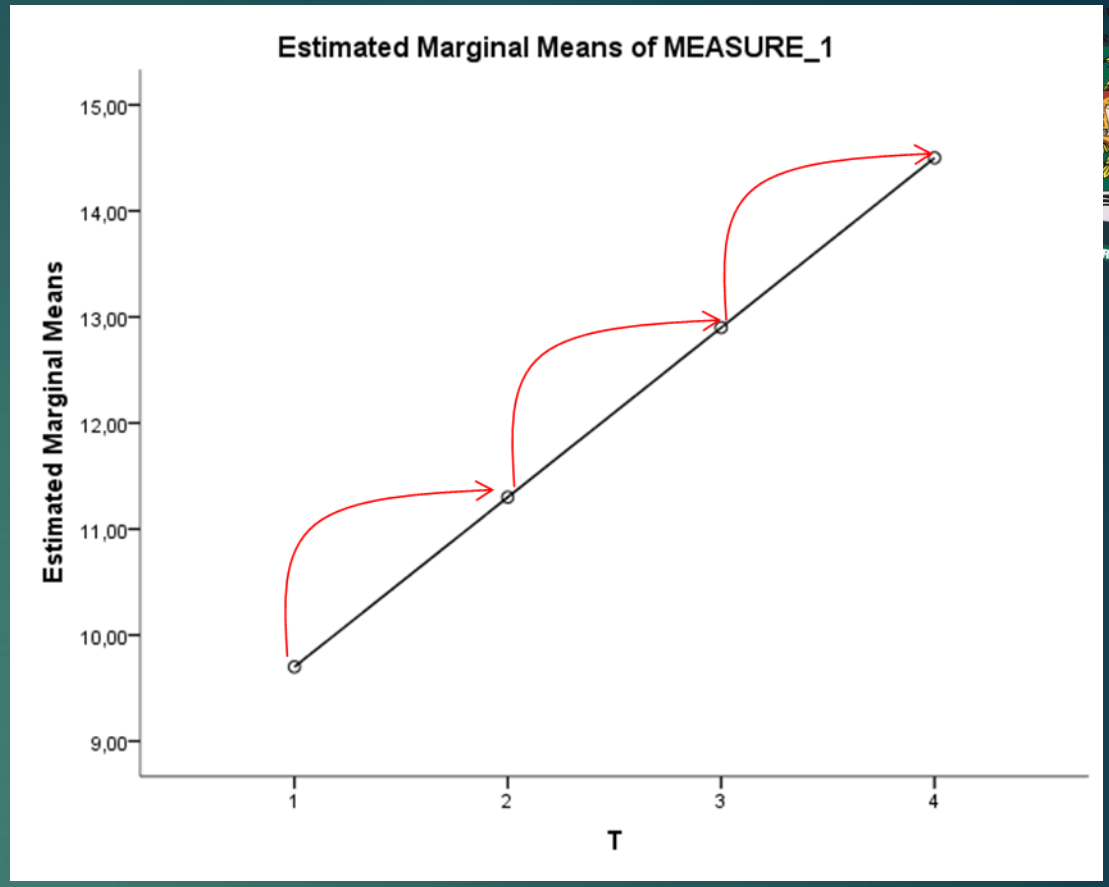


Análise de variância (ANOVA) com medidas repetidas ou múltiplas (one-way repeated measures - ANOVA)

- Há apenas um nível de comparação (intra-grupo)
- Vários CONTRASTES (modos de comparação)







Tests of Within-Subjects Contrasts

Measure:MEASURE_1

Source	T	Type III Sum of Squares	df	Mean Square	F	Sig.
T	Level 2 vs. Level 1	25,600	1	25,600	96,000	,000
	Level 3 vs. Level 1	102,400	1	102,400	164,571	,000
	Level 4 vs. Level 1	230,400	1	230,400	370,286	,000
Error(T)	Level 2 vs. Level 1	2,400	9	,267		
	Level 3 vs. Level 1	5,600	9	,622		
	Level 4 vs. Level 1	5,600	9	,622		

Tests of Within-Subjects Contrasts

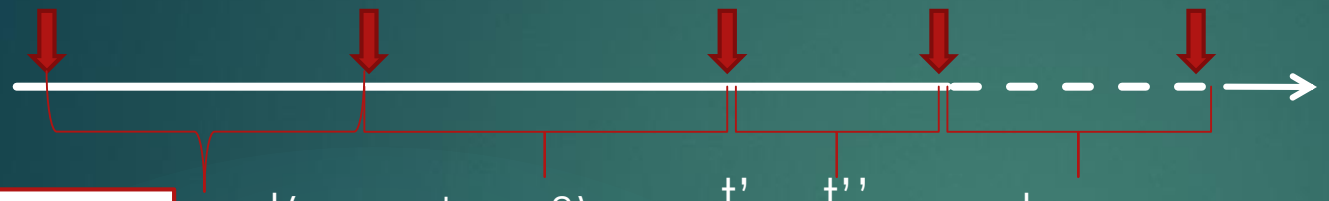
Measure:MEASURE_1

Source	T	Type III Sum of Squares	df	Mean Square	F	Sig.
T	Level 1 vs. Level 2	25,600	1	25,600	96,000	,000
	Level 2 vs. Level 3	25,600	1	25,600	96,000	,000
	Level 3 vs. Level 4	25,600	1	25,600	96,000	,000
Error(T)	Level 1 vs. Level 2	2,400	9	,267		
	Level 2 vs. Level 3	2,400	9	,267		
	Level 3 vs. Level 4	2,400	9	,267		

Análise de variância com medidas repetidas com 2 ou + grupos

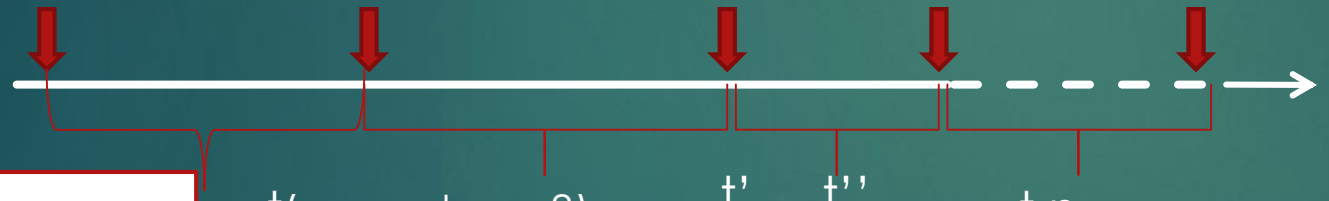
Grupo 1

Medida 1 "Tratamento" Medida 2 Medida 3 Medida n



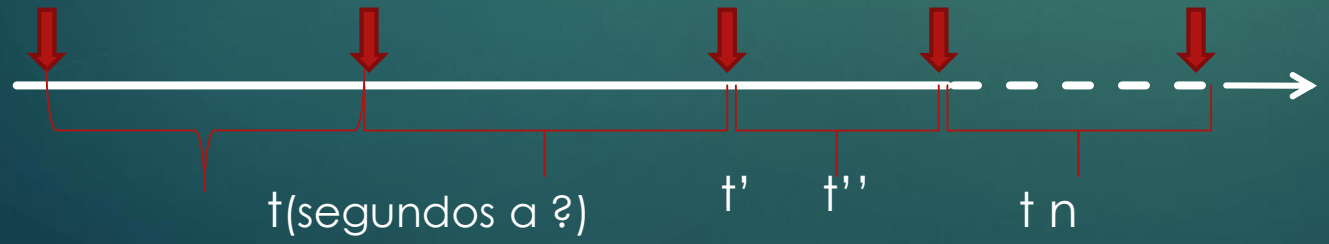
Grupo 2

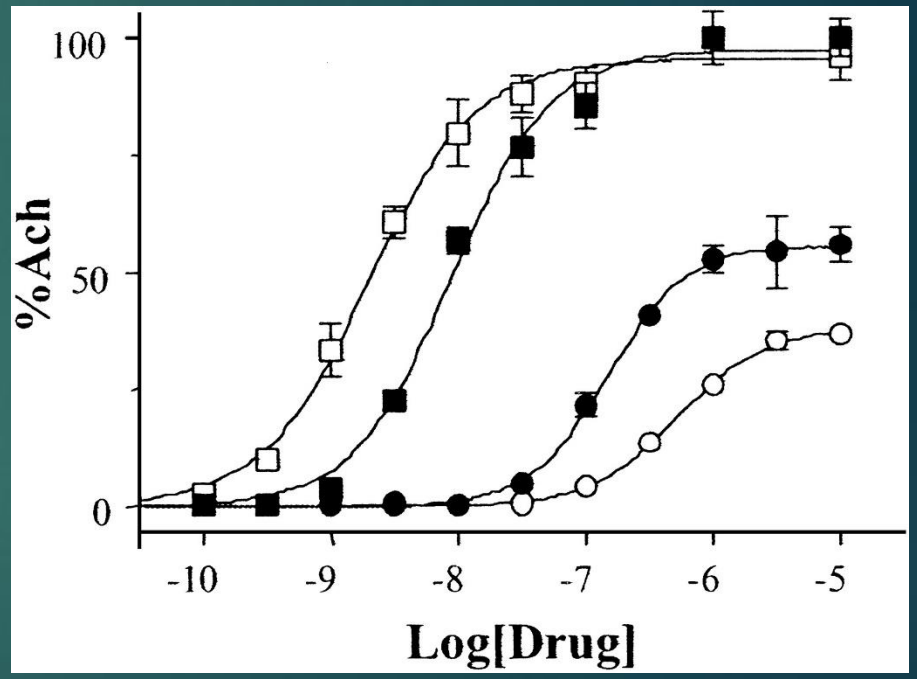
Medida 1 "Tratamento" Medida 2 Medida 3 Medida n



Grupo n

Medida 1 "Tratamento" Medida 2 Medida 3 Medida n

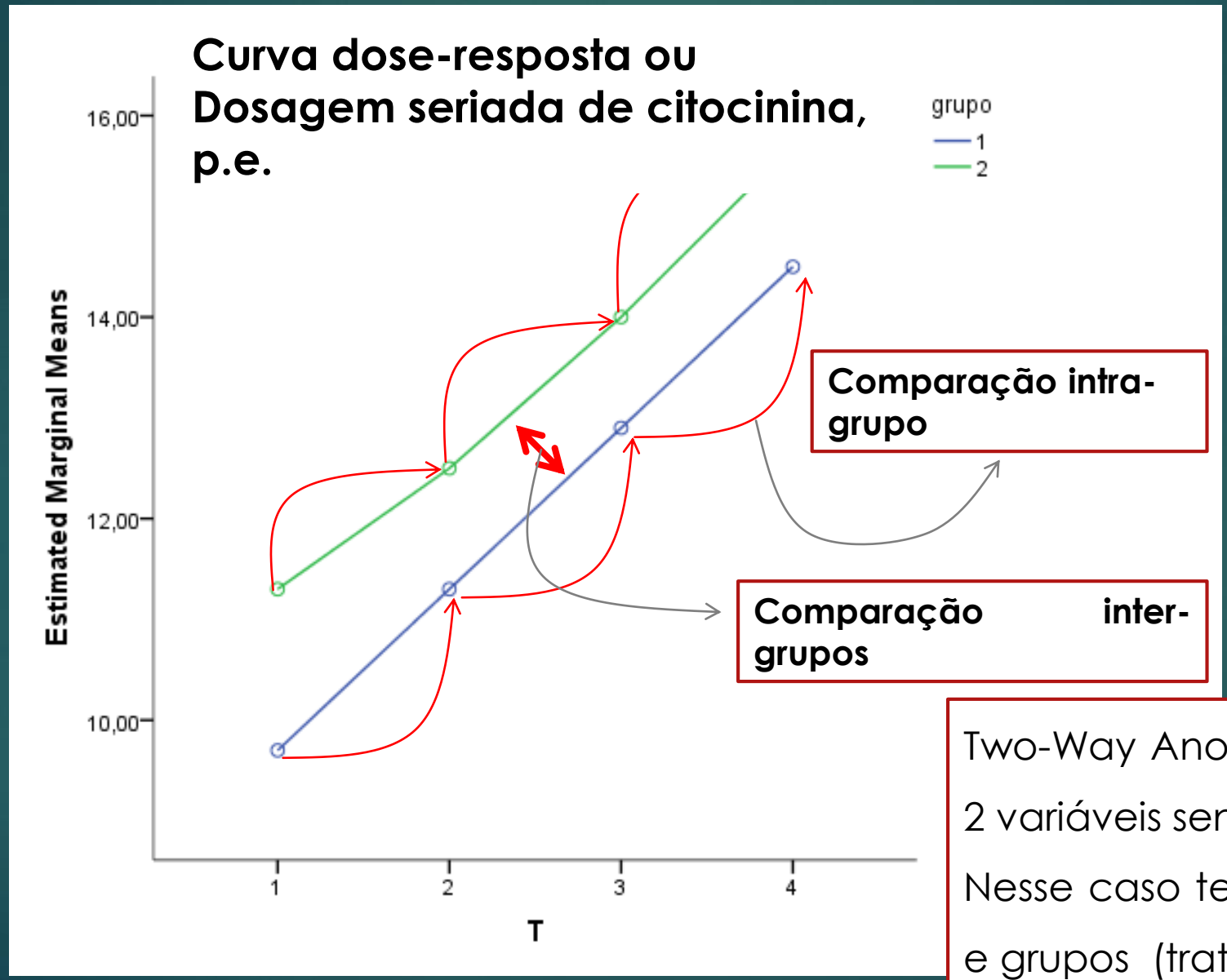




Análise de variância com medidas repetidas com 2 ou + grupos

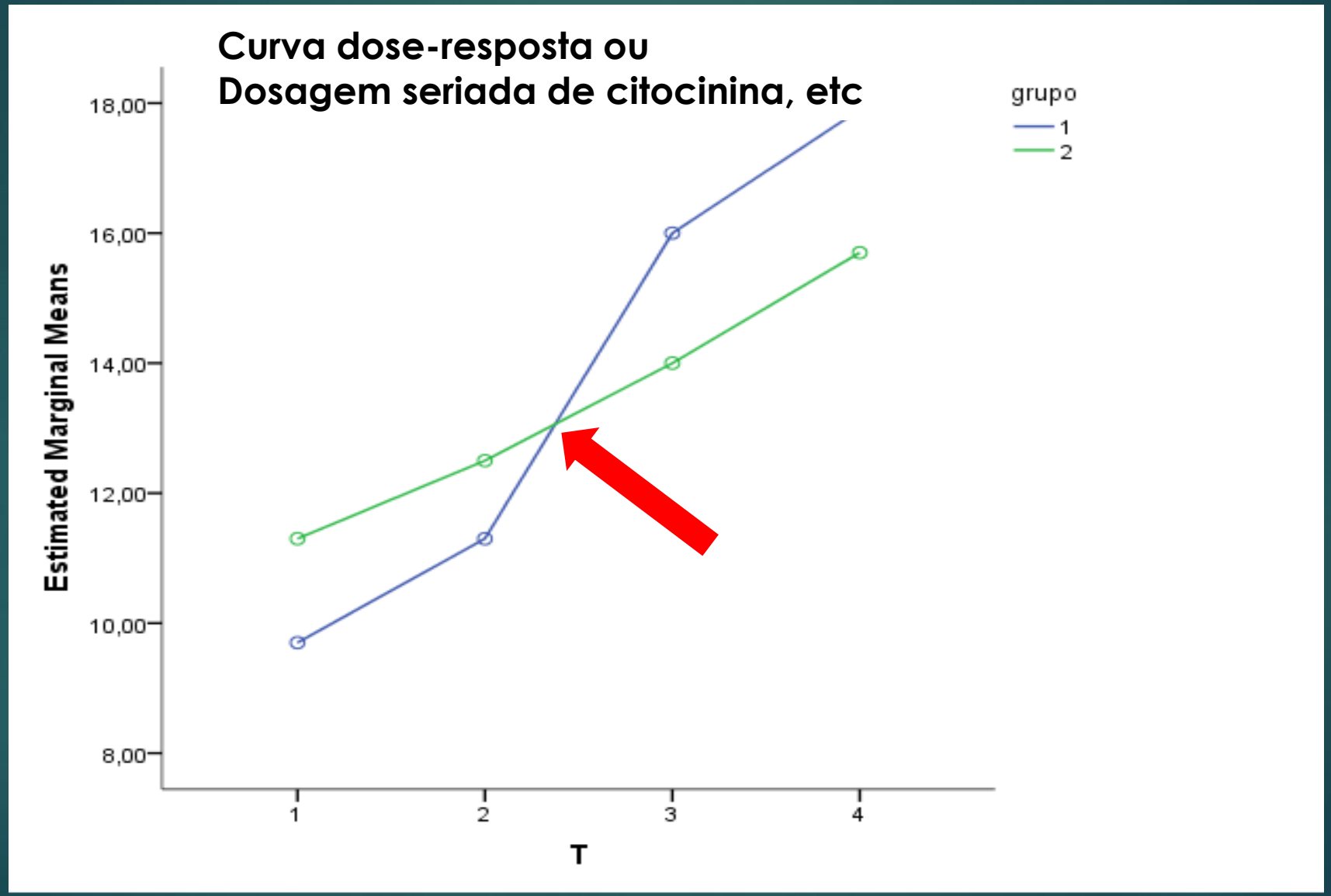
- **Três níveis de comparação**
  - **Intra-grupos**
  - **Inter-grupos**
  - **Interação entre grupos**

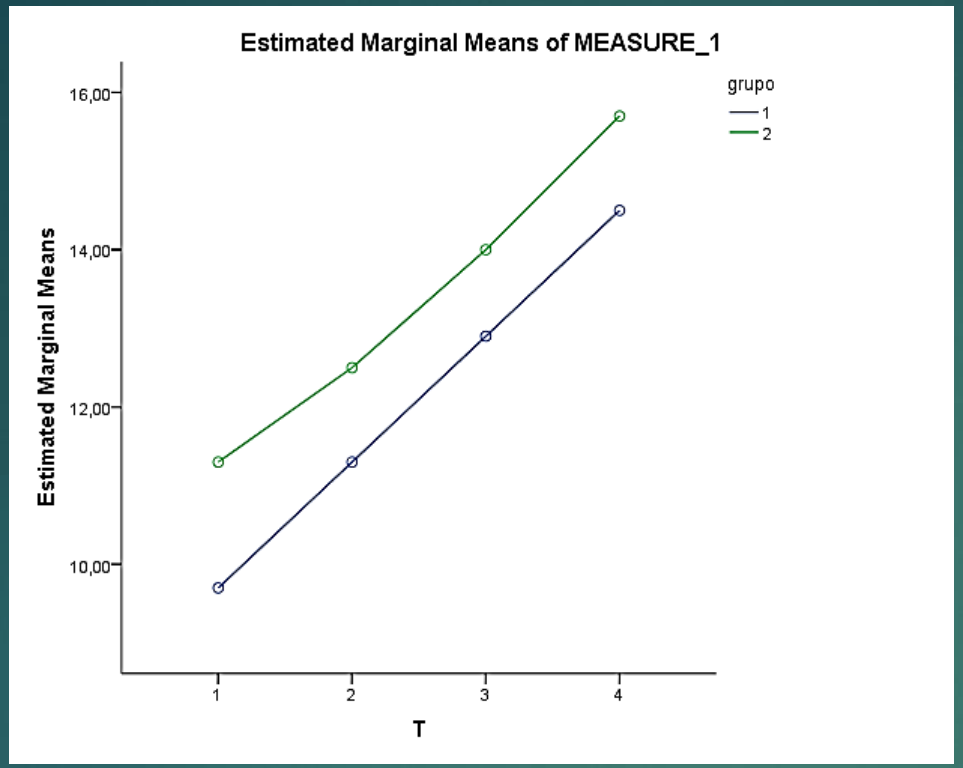
Comparação intra-grupos e inter-grupos



Two-Way Anova porque são 2 variáveis sendo avaliadas. Nesse caso temos T (tempo) e grupos (tratamentos ≠s)

Interação


Curva dose-resposta ou
Dosagem seriada de citocinina, etc
grupo
1
2
Estimated Marginal Means
18,00
16,00
14,00
12,00
10,00
8,00
1
2
3
4
T



Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
factor1	,001	77,189	20	,000	,436	,597	,167

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept + grupo
Within Subjects Design: factor1

Esfericidade é a condição em que as variâncias das diferenças entre todas as combinações de grupos relacionados (níveis) são iguais.

Tests of Within-Subjects Effects

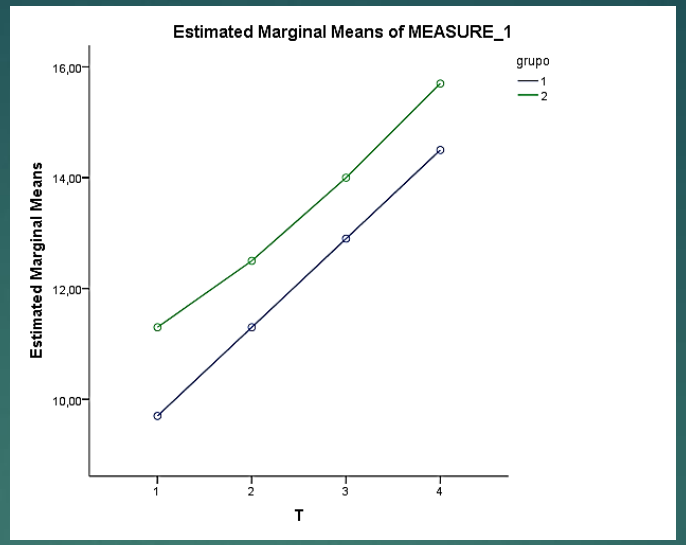
Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
T	Sphericity Assumed	235,937	3	78,646	234,959	,000
	Greenhouse-Geisser	235,937	1,851	127,482	234,959	,000
	Huynh-Feldt	235,937	2,168	108,816	234,959	,000
	Lower-bound	235,937	1,000	235,937	234,959	,000
interação	Sphericity Assumed	,738	3	,246	,734	,536
	Greenhouse-Geisser	,738	1,851	,398	,734	,477
	Huynh-Feldt	,738	2,168	,340	,734	,497
	Lower-bound	,738	1,000	,738	,734	,403
Error(T)	Sphericity Assumed	18,075	54	,335		
	Greenhouse-Geisser	18,075	33,314	,543		
	Huynh-Feldt	18,075	39,028	,463		
	Lower-bound	18,075	18,000	1,004		

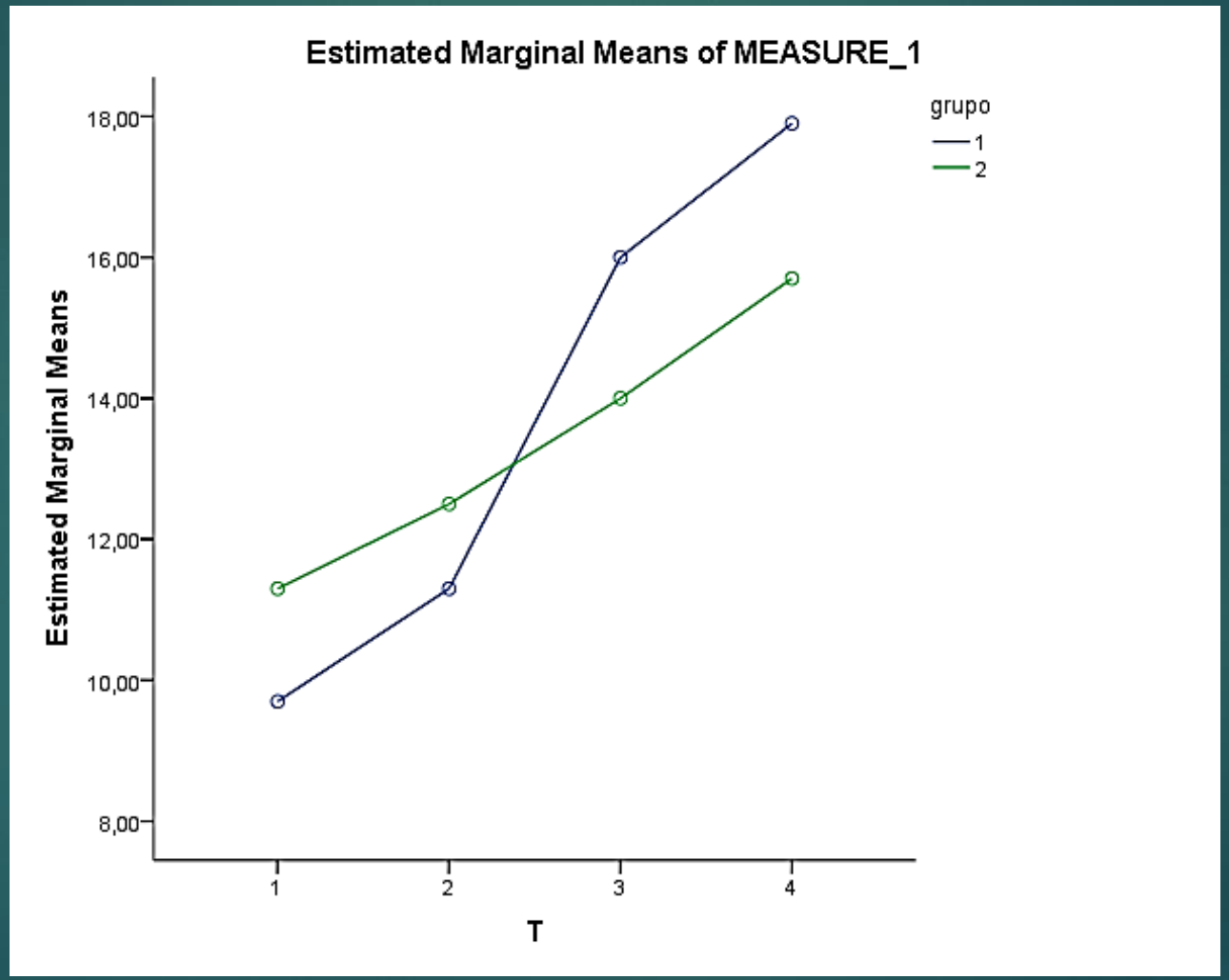
Tests of Between-Subjects Effects

Measure:MEASURE_1
Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	3244,878	1	3244,878	729,585	,000
grupos	8,128	1	8,128	1,828	,193
Error	80,056	18	4,448		



- Conclusão:
- Há diferença significativa antre as doses dentro de cada grupo (intra-grupos, p<0,001), ou seja, as curvas não são horizontais
- Como o teste intergrupos não foi signficativo (p=0,678) não há diferença entre os grupos, ou seja, a "distância entre as curvas não é significative"
- Como o valor de "p" na interação foi >0,05 (p=0,477) as curvas são paralelas (não há interação)

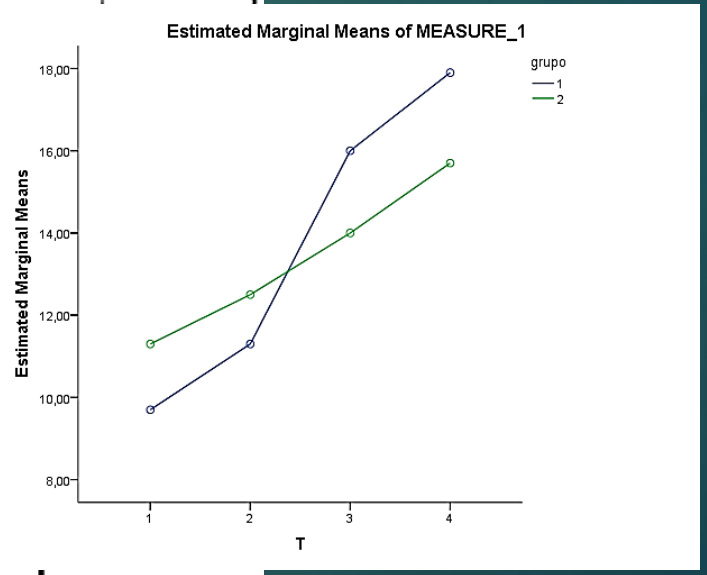

Estimated Marginal Means of MEASURE_1
grupo
1
2
Estimated Marginal Means
18,00
16,00
14,00
12,00
10,00
8,00
1
2
3
4
T

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
T	Sphericity Assumed	493,800	3	164,600	108,461	,000
	Greenhouse-Geisser	493,800	2,076	237,902	108,461	,000
	Huynh-Feldt	493,800	2,481	199,009	108,461	,000
	Lower-bound	493,800	1,000	493,800	108,461	,000
T *	Sphericity Assumed	61,750	3	20,583	13,563	,000
	Greenhouse-Geisser	61,750	2,076	29,750	13,563	,000
	Huynh-Feldt	61,750	2,481	24,886	13,563	,000
	Lower-bound	61,750	1,000	61,750	13,563	,002
Error(T)	Sphericity Assumed	81,950	54	1,518		
	Greenhouse-Geisser	81,950	37,362	2,193		
	Huynh-Feldt	81,950	44,663	1,835		
	Lower-bound	81,950	18,000	4,553		

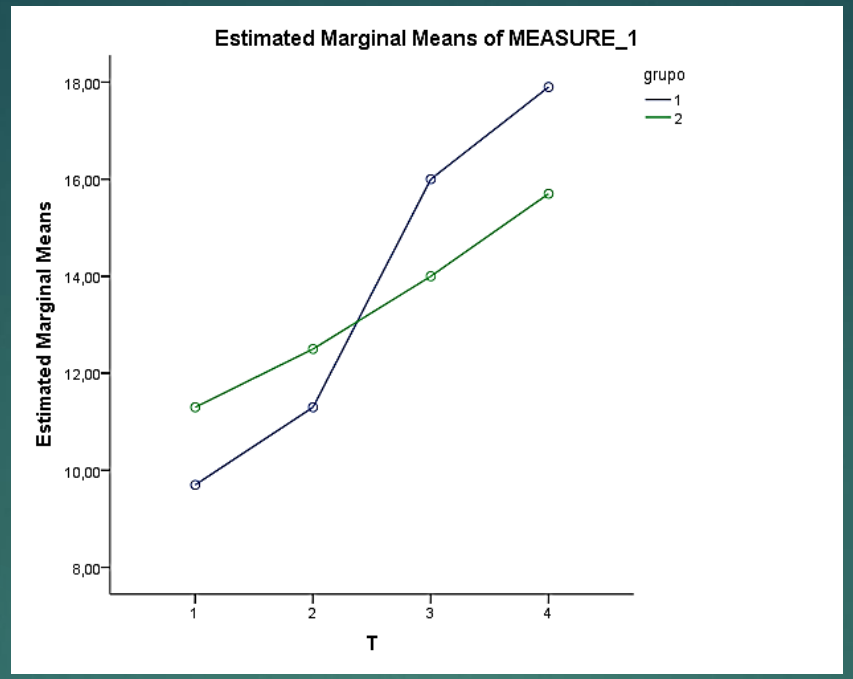
interação



Tests of Between-Subjects Effects

Measure:MEASURE_1
Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
	3672,050	1	3672,050	1066,724	,000
grupos	,613	1	,613	,178	,678
Error	61,963	18	3,442		

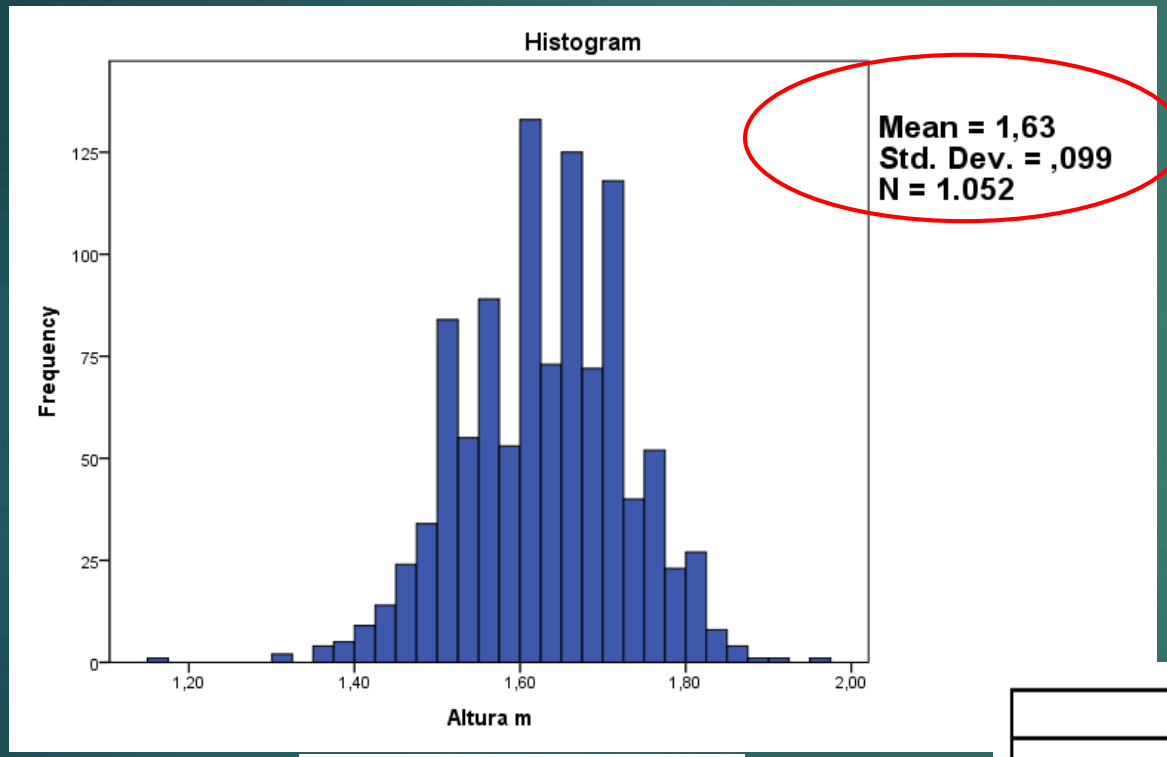


- Conclusão:
- Há diferença significativa antre as dosegens dentro de cada grupo (intra-grupos, p<0,001), ou seja, as curvas não são horizontais
- Como o teste inter-grupos foi não signficativo (p=0,678) não há diferença entre os grupos, ou seja, a distância entre as curvas não é significativa
- Mas como o valor de "p" na interação foi <0,05 (p<0,001) as curvas NÃO são paralelas

Intervalo de Confiança da Diferença das Médias

Intervalo de confiança da média (IC)

- **O intervalo de confiança de uma média nos da o "grau" de certeza (90%, 95%, 99%), de que o intervalo CONTÉM a verdadeira média POPULACIONAL**



➡ POPULAÇÃO

Amostra de 50 (~5% da população)
média = 1,64
95% IC: 1,60 – 1,67

Amostra de 200 (~20% da população)
média = 1,63
95% IC: 1,61 – 1,64

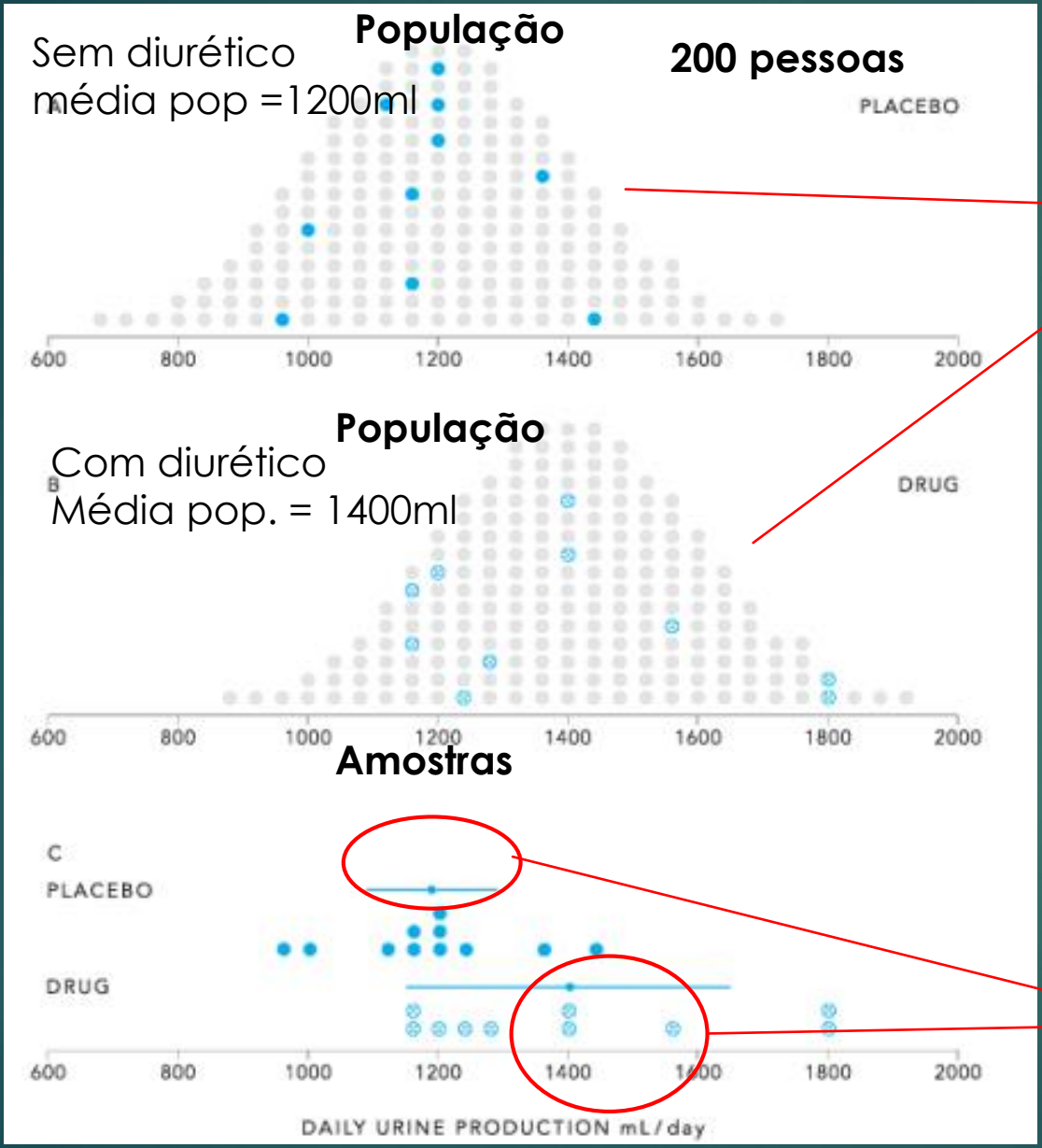
$$\overline{x} \pm z \frac{s}{\sqrt{n}}$$

Z Scores for Commonly Used Confidence Intervals	
Desired Confidence Interval	Z Score
90%	1.645
95%	1.96
99%	2.576

Intervalo de confiança da ≠ de médias(IC)

- O valor de "p" pode nos indicar com x% de "certeza de que uma diferença de médias não é devido ao acaso e portanto as amostras não provem da mesma população (o tratamento teve efeito)
- O valor de "p" nos diz apenas se uma diferença foi estatisticamente significativa, mas não nos permite avaliar a magnitude desta diferença de modo a ajuizar sua relevância.
- O tamanho da diferença entre as médias nos permite estimar a magnitude do "efeito de um tratamento"
- O intervalo de confiança da diferença de médias nos permite estimar com x% (95% ou 99%) de certeza que o intervalo contem a real diferença das médias.

Intervalo de confiança das diferenças de média



≠ na média **populacional com e sem** diurético = 200 ml **(TAMANHO REAL DO EFEITO, DESCONHECIDO PELO PESQUISADOR)**

Há uma **incerteza** quando relatamos o tamanho do efeito baseado na ≠ da média das amostras

≠ das médias amostrais = 250 ml (**é apenas uma estimativa**)

Podemos utilizar o IC (95%, 99%, etc) para descrever o tamaho do efeito, em média, causado pelo "tratamento", e demonstrar o quão incerto é esse tamanho (≠ de efeito), calculando o intervalo de confiaça da diferença entre as médias das amostras

$$(\overline{X}_1 - \overline{X}_2) - t_\alpha s_{\overline{X}_1 - \overline{X}_2} < \mu_1 - \mu_2 < (\overline{X}_1 - \overline{X}_2) + t_\alpha s_{\overline{X}_1 - \overline{X}_2}$$

- Para nosso exemplo, utilizando IC da ≠ de médias observadas de 95%:

  31ml/dia < ≠ de médias < 409ml/dia

  (inclui os 250ml da ≠ observada)

- Consequentemente, podemos estar 95% confiantes de que o diurético aumenta a diurese entre 31 e 409 ml/dia.
- Ademais, como não inclui valor "0", a ≠ significativa

Vantagens de utilizar IC da ≠ das médias

- **Determinar o Tamanho do efeito do "tratamento (relevância)**
  - **Teste de droga antihipertensiva**
  - Grupo controle: PA média 85mmHg
  - Grupo tratado: PA média 81mmHg
  - DP=9mmHg, n=100 cada
  - Valor "t" -2,82 (cutoof para 0,01 = -2,61), portanto significativo

$$-6.9\ \text{mmHg} < \mu_{dr} - \mu_{pla} < -1.2\ \text{mmHg}$$

95% confiança de que a medicação diminuia PA entre 1,2 a 6,9mmHg

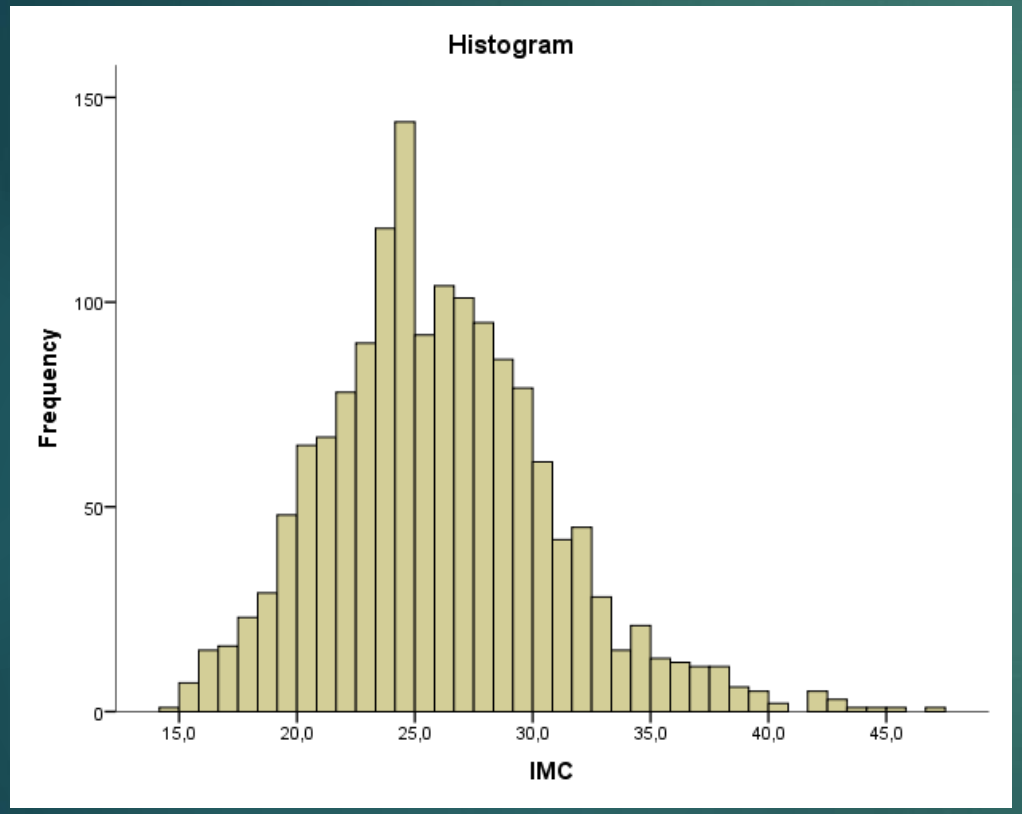
- **Permite testar hipótese**
  - Se o intervalo inclui o valor 0 significa que há probabilidade (dentro dos 95%) de não haver diferença 9diferença = 0), portanto a ≠ é não-significative para p<0,05

Alternativas para distribuições não-Gaussianas

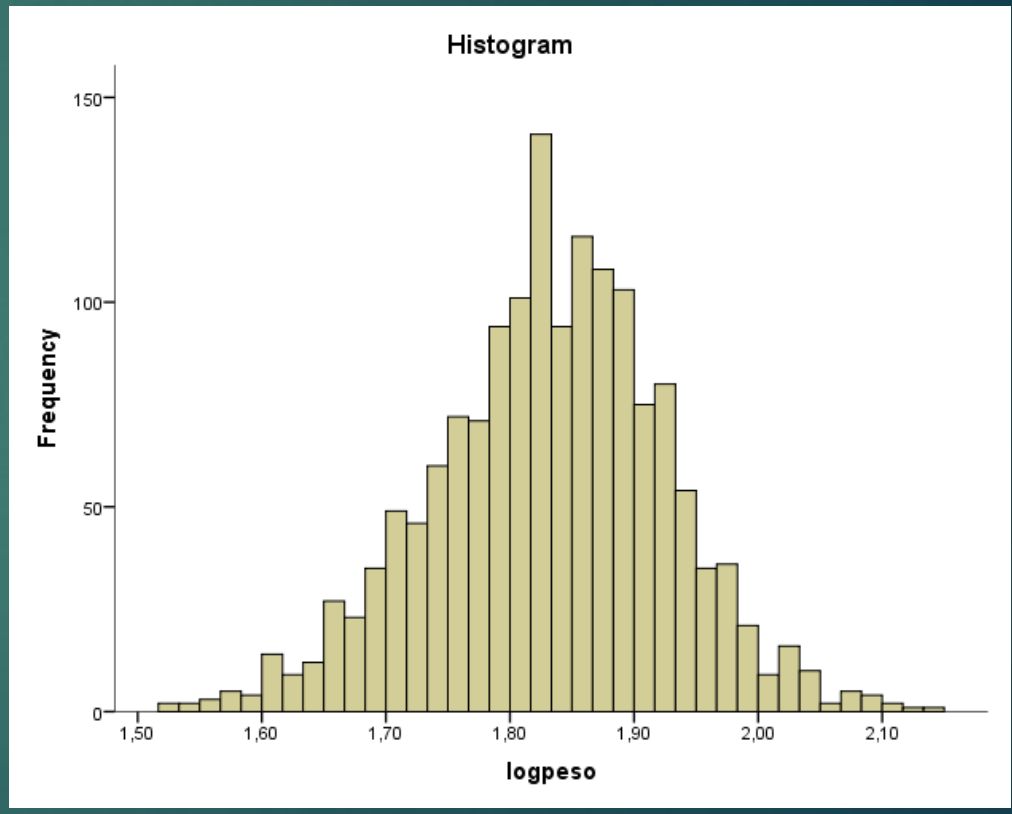
- **Condições para uso testes paramétricos**
  - Distribuição Normal
  - Variâncias semelhantes entre grupos (testes específicos, Levene p.e)
- E quando essas condições não são satisfeitas?
  - Alternativas para distribuição não-simétrica
    - **Utilizar teste não-paramétrico**
    - **Retirar valores extremos (justificável apenas se forem "erro")***
    - **Transformação matemática das variáveis**:
      - $\log x$
      - $\sqrt{x}$
      - $1/x$
      - escore reverso

Data Transformation	Can Correct For
Log transformation ($\log(X_i)$): Taking the logarithm of a set of numbers squashes the right tail of the distribution. As such it's a good way to reduce positive skew. However, you can't get a log value of zero or negative numbers, so if your data tend to zero or produce negative numbers you need to add a constant to all of the data before you do the transformation. For example, if you have zeros in the data then do $\log(X_i + 1)$, or if you have negative numbers add whatever value makes the smallest number in the data set positive.	Positive skew, unequal variances
Square root transformation ($\sqrt{X_i}$): Taking the square root of large values has more of an effect than taking the square root of small values. Consequently, taking the square root of each of your scores will bring any large scores closer to the centre – rather like the log transformation. As such, this can be a useful way to reduce positive skew; however, you still have the same problem with negative numbers (negative numbers don't have a square root).	Positive skew, unequal variances
Reciprocal transformation ($1/X_i$): Dividing 1 by each score also reduces the impact of large scores. The transformed variable will have a lower limit of 0 (very large numbers will become close to 0). One thing to bear in mind with this transformation is that it reverses the scores: scores that were originally large in the data set become small (close to zero) after the transformation, but scores that were originally small become big after the transformation. For example, imagine two scores of 1 and 10; after the transformation they become 1/1 = 1, and 1/10 = 0.1: the small score becomes bigger than the large score after the transformation. However, you can avoid this by reversing the scores before the transformation, by finding the highest score and changing each score to the highest score minus the score you're looking at. So, you do a transformation $1/(X_{Highest} - X_i)$.	Positive skew, unequal variances
Reverse score transformations: Any one of the above transformations can be used to correct negatively skewed data, but first you have to reverse the scores. To do this, subtract each score from the highest score obtained, or the highest score + 1 (depending on whether you want your lowest score to be 0 or 1). If you do this, don't forget to reverse the scores back afterwards, or to remember that the interpretation of the variable is reversed: big scores have become small and small scores have become big!	Negative skew

IMC
Skewness=0,609



Log(IMC)
Skewness=-0,009



Alternativas para distriuições não-normais (Gaussianas)

Usar testes que não utilizam parâmetros da população, e portanto são não-paramétricos.

Estes teste utilizam dos valores ordenados por grandeza ou sinais (ranks, - ou +) e focam nas diferenças de mediana ao invés de médias

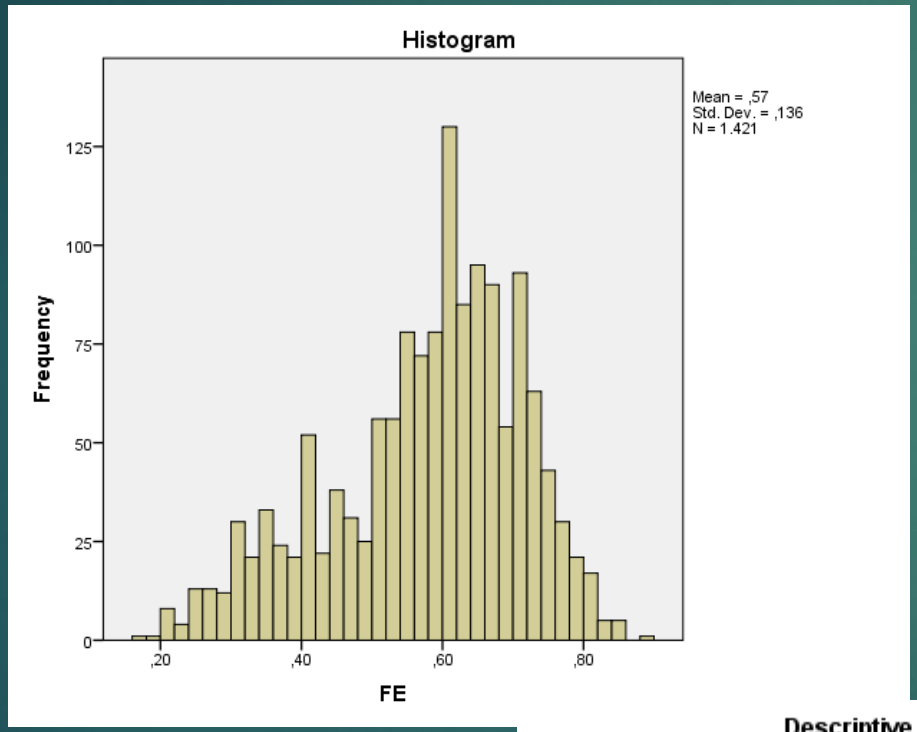
PARAMÉTRICO	EQUIVALENTE NÃO-PARAMÉTRICO
Teste “t”	Mann-Whitney
Teste “t” pareado	Wilcoxon ou McNemar (proporções antes e após trat.)
Anova	Kruskal-Wallis
Análise de medidas repetidas	Friedman (não compara contrastes/entre grupos)
Correlação de Pearson	Spearman
	X^2 qui-quadrado

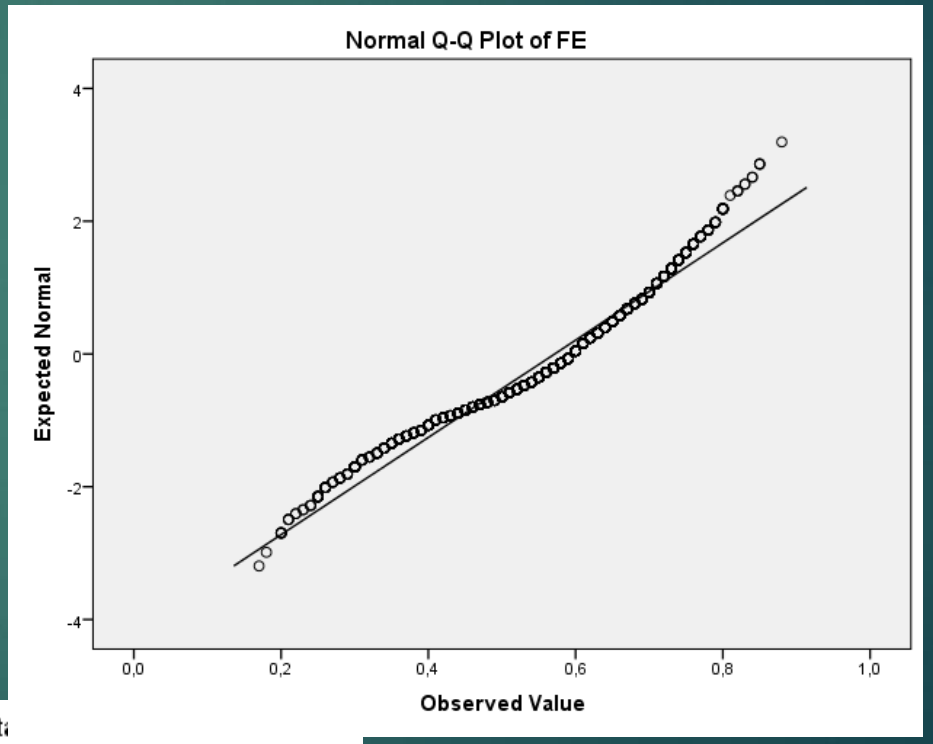
Assim como para os paramétricos é possível proceder testes Post Hoc para não paramétricos mediante procedimentos matemáticos para os ajustes de valor de “p” (Bonferroni, Dunn-Bonferroni, Dunn-Sidak , etc.) que às vezes são opções fornecidos pelos softwares de estatística.

Exemplo

- As frações de ejeção (FE) do ventrículos esquerdo são semelhantes entre as classes I a IV da NYHA?

4 grupos: ANOVA?





Descriptive Sta

	N	Mean	Skewness	
	Statistic	Statistic	Statistic	Std. Error
FE	1421	,5714	-,604	,065
Valid N (listwise)	1421			

Não-paramétrico para mais de 2 grupos

Report

FE

Classe Func NYHA	Mean	N	Std. Deviation
Classe I	,6036	583	,11264
Classe II	,5552	434	,14047
Classe III	,5454	324	,15230
Classe IV	,5252	64	,15775
Total	,5716	1405	,13624

Rank

Ranks

	Classe Func NYHA	N	Mean Rank
FE	Classe I	583	783,83
	Classe II	434	655,78
	Classe III	324	645,13
	Classe IV	64	579,88
	Total	1405	

Test Statistics[a,b]

	FE
Chi-Square	41,536
df	3
Asymp. Sig.	,000

a. Kruskal Wallis Test

b. Grouping Variable: Classe Func NYHA

Rejeito H0, há diferenças da FE entre os grupos, mas entre quais grupos?

Dunn-Bonferroni post hoc

Each node shows the sample average rank of Classe Func NYHA.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Classe IV-Classe III	65,245	55,477	1,176	,240	1,000
Classe IV-Classe II	75,897	54,305	1,398	,162	,973
Classe IV-Classe I	203,947	53,405	3,819	,000	,001
Classe III-Classe II	10,652	29,777	,358	,721	1,000
Classe III-Classe I	138,701	28,103	4,935	,000	,000
Classe II-Classe I	128,049	25,712	4,980	,000	,000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.

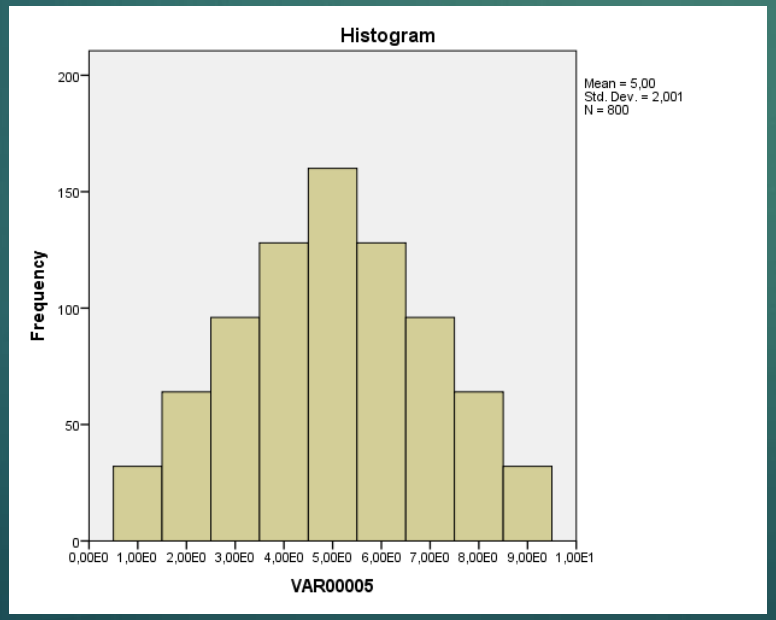
Escolha entre os testes paramétricos e não paramétricos

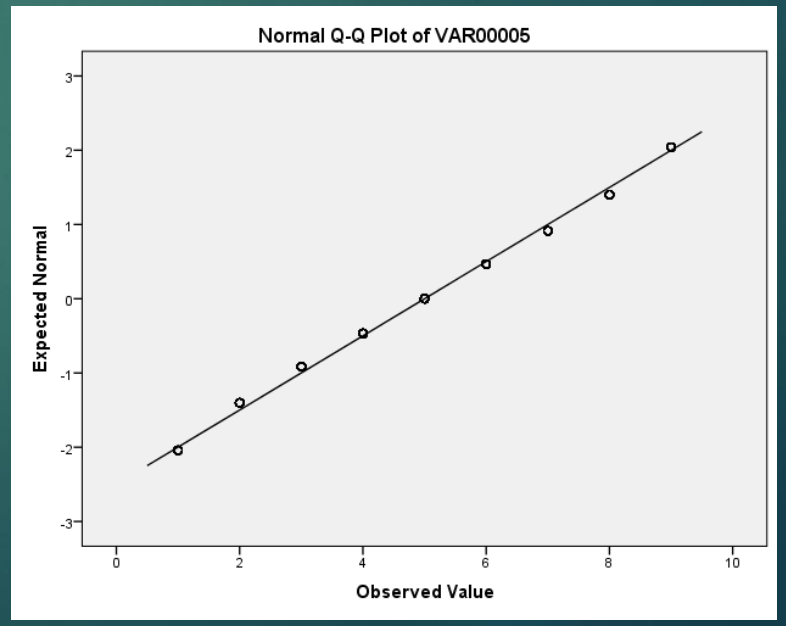
- **Os casos fáceis para não-paramétricos**
  - O resultado é em escores ou pontuação (grau I, grau II, leve,moderado, etc)
  - Alguns valores estão "fora da escala", ou seja, muito altos ou muito baixos. Mesmo se a população for gaussiana, é impossível analisar esses dados com um teste paramétrico.
  - Você tem certeza que **na população a distribuição é não gaussiana** (Salário, escala Apgar p.e; considerar transformação)

Escolha entre os testes paramétricos e não paramétricos

- **A distribuição é normal? Como avaliar?**

- Plotar os dados em histogramas de frequência e examinar
- Construir gráficos "Normais Q-Q" ou de "probabilidade normal" (pontos na reta)





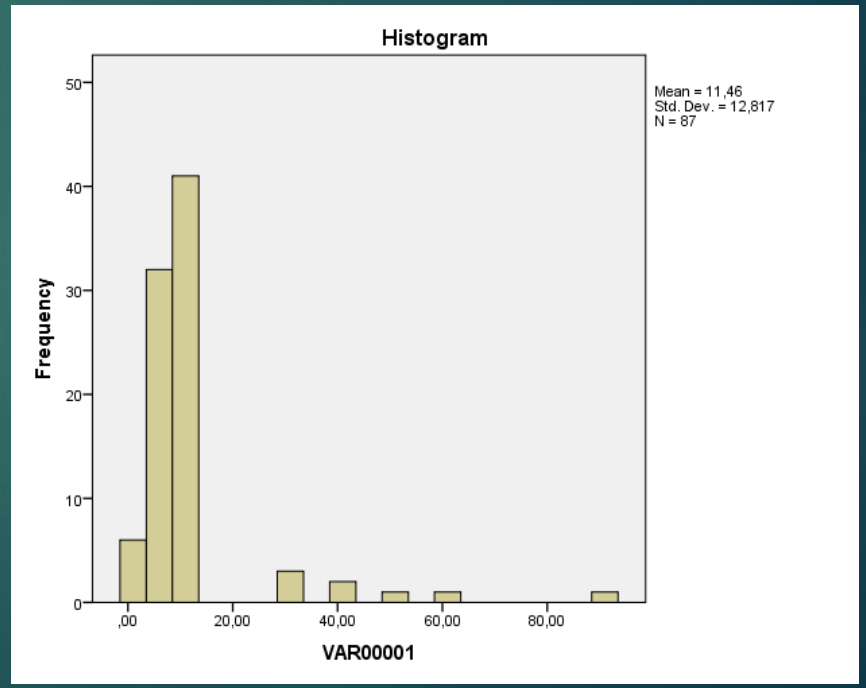
Escolha entre os testes paramétricos e não paramétricos

A distribuição é normal? Como avaliar?

Se o desvio-padrão é semelhante a média, ou maior, e a variável puder ter apenas valores positivos, estes são indicativos de distribuição assimétrica (uma variável normalmente distribuída teria de conter valores negativos)

Descriptive Statistics

	N	Mean	Std. Deviation
VAR00001	87	11,4598	12,81696
Valid N (listwise)	87		



Escolha entre os testes paramétricos e não paramétricos

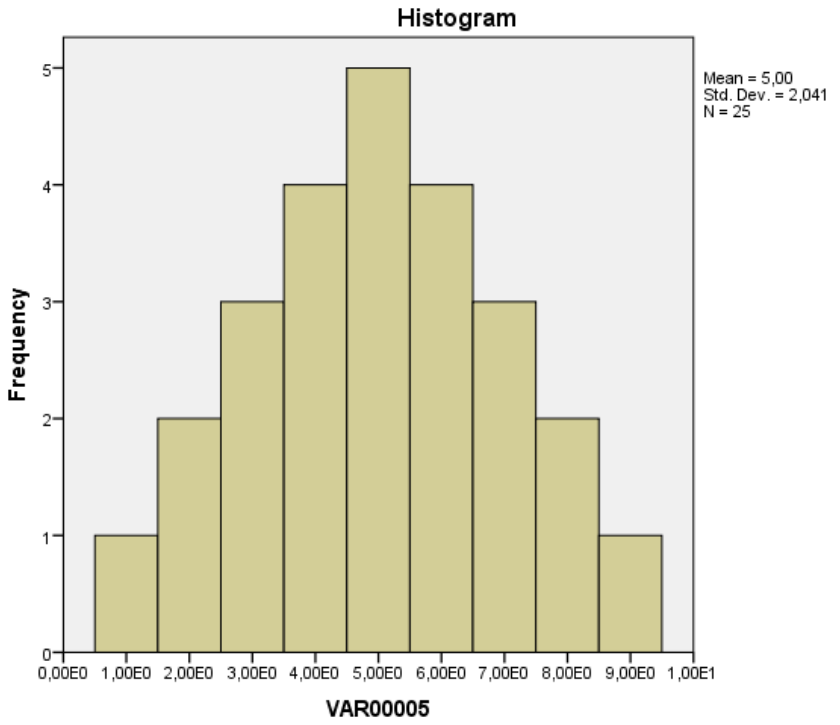
- **Utilizar testes**: Kolmogorov Smirnov (desaconselhado por alguns), Shapiro-Wilk e D'Agostino-Pearson omnibus test *
  - Nestes testes H0 é: a distribuição é normal. Portanto, se p<0,05 a distribuição **não é normal***

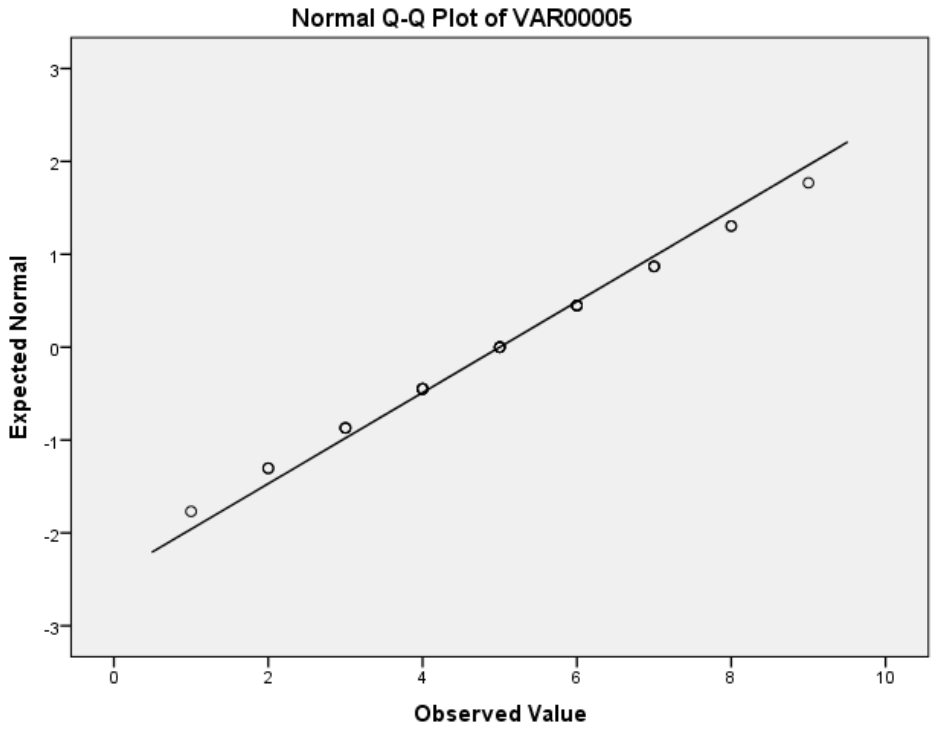
*** Devido ao baixo poder desses testes, AMOSTRAS PEQUENAS quase sempre passam no teste (falsamente como distribuição normal, p>0,05) e as GRANDES quase sempre não passam, pois o "n" grande faz com que o testes detectem mesmo os pequenos desvios na normalidade que na verdade não teriam influência nos testes.**

AMOSTRAS PEQUENAS

Descriptive Statistics

	N	Mean	Skewness	
	Statistic	Statistic	Statistic	Std. Error
VAR00005	25	5,0000	,000	,464
Valid N (listwise)	25			





Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
VAR00005	,100	25	,200*	,978	25	,834

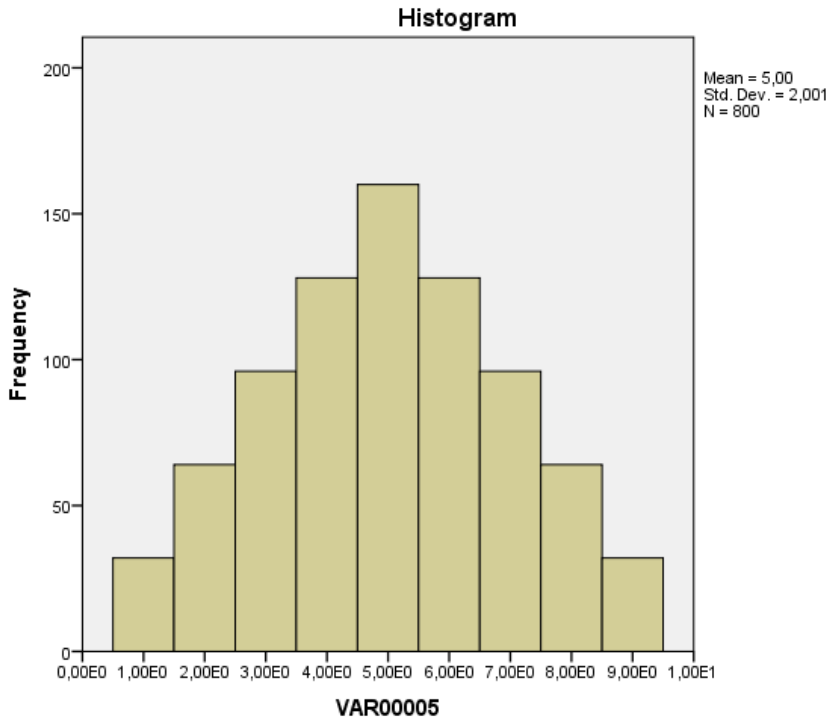
*. This is a lower bound of the true significance.

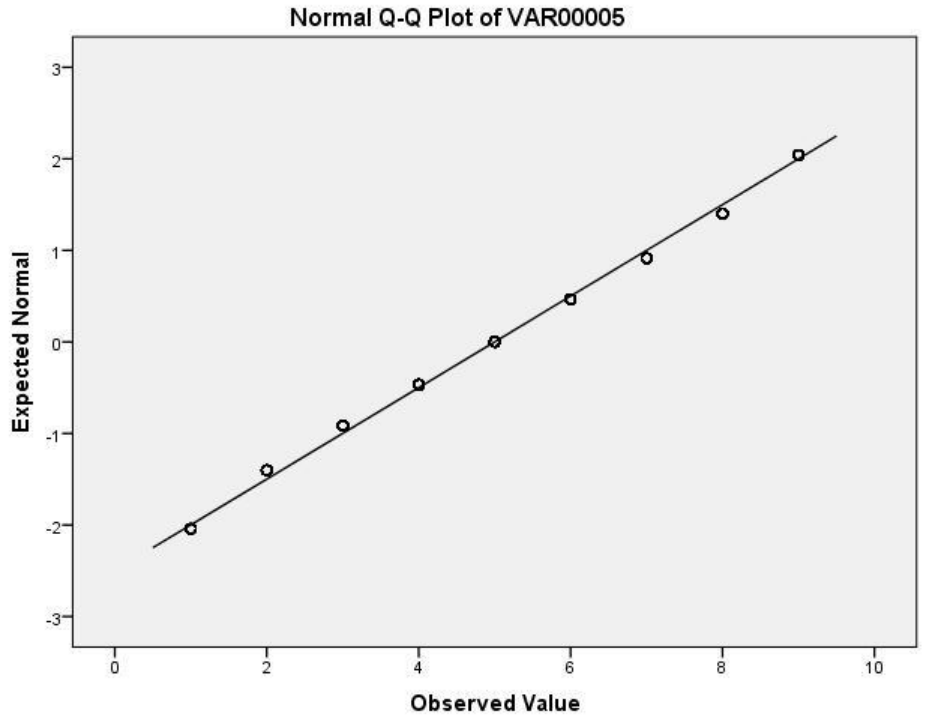
a. Lilliefors Significance Correction

Amostras grandes

Descriptive Statistics

	N	Mean	Skewness	
	Statistic	Statistic	Statistic	Std. Error
VAR00005	800	5,0000	,000	,086
Valid N (listwise)	800			





Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
VAR00005	,100	400	,000	,967	400	,000

a. Lilliefors Significance Correction

Escolha entre os testes paramétricos e não paramétricos

- **Os casos difíceis:**
  - Com poucos dados **(n pequeno)**, é difícil dizer se os dados tem distribuição normal por inspeção dos gráficos e **os testes formais tem pouco poder** de discriminar entre as distribuições normal e não.
  - Alguns argumentam que o importante **é a distribuição da POPULAÇÃO**, e **não a distribuição na amostra**. Para decidir se uma população tem distrbuição normal, olhe todos os dados disponíveis (outras publicações), e não apenas os dados na experiência atual.

Escolhendo entre paramétrico e não-paramétrico

Importa se você escolher um teste paramétrico ou não-paramétrico?

DEPENDE

- **AMOSTRAS GRANDES:**
  - Os testes paramétricos são robustos a **desvios** da distribuição normal (“Teorema limite central”)
  - O quão grande uma amostra é grande o suficiente?
    - Depende da natureza da distribuição não-Gaussiana. Em todo caso, a menos que a distribuição da população seja **realmente estranha**, provavelmente você estará seguro se escolher um teste paramétrico, quando há **pelo menos três dezenas (alguns dizem duas)** de pontos de dados em cada grupo.

Escolhendo entre paramétrico e não-paramétrico

- **Amostra grande:**
  - O que acontece quando você usa um teste **não-paramétrico** com dados de uma **população gaussiana** (com distribuição normal)?
  - **Os testes não-paramétricos serão menos poderoso** do que os testes paramétricos para detectar as diferenças (depende do tamanho da amostra e do tamanho da diferença) o pode resultar em valor de p **um pouco maior**.
  - Vide exemplos a seguir

Paramétrico x não-paramétrico

Group Statistics

	Idoso	N	Mean	Std. Deviation	Std. Error Mean
Peso	Não idoso	861	70,8424	16,52363	,56312
	Idoso	680	68,0260	12,91629	,49532

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Peso	Equal variances assumed	33,054	,000	3,650	1539	,000	2,81636	,77156	1,30295	4,32978
	Equal variances not assumed			3,755	1538,844	,000	2,81636	,74996	1,34530	4,28742

Test Statistics[a]

	Peso
Mann-Whitney U	267428,000
Wilcoxon W	498968,000
Z	-2,918
Asymp. Sig. (2-tailed)	,004

a. Grouping Variable: Idoso

Group Statistics

	Sexo	N	Mean	Std. Deviation	Std. Error Mean
IMC	feminino	664	26,400	5,2085	,2021
	masculino	878	25,800	4,5560	,1538

Independent Samples Test

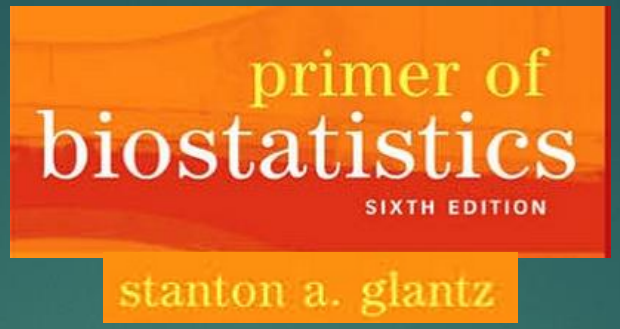
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
IMC	Equal variances assumed	15,172	,000	2,175	1540	,030	,5423	,2493	,0532	1,0313
	Equal variances not assumed			2,135	1318,538	,033	,5423	,2540	,0441	1,0405

Test Statistics[a]

	IMC
Mann-Whitney U	277463,000
Wilcoxon W	663344,000
Z	-1,621
Asymp. Sig. (2-tailed)	,105

a. Grouping Variable: Sexo

Escolhendo entre paramétrico e não-paramétrico



Pode-se resumir tudo basicamente na seguinte diferença de opinião: algumas pessoas pensam que na *ausência* de evidência de que os dados *não* foram tomados a partir de uma população normalmente distribuída, pode-se usar testes paramétricos pois são mais poderosos e mais amplamente usados. Estas pessoas dizem que deveria ser usado um teste não paramétrico somente quando houver uma evidência positiva de que as populações que estão sendo estudadas não são normalmente distribuídas. Outras apontam que os métodos não paramétricos discutidos neste capítulo são 95% tão poderosos quanto métodos paramétricos quando os dados vêm de populações com distribuição normal e são mais confiáveis quando os dados não forem tomados de populações normalmente distribuídas. Estas também acreditam que pesquisadores devem assumir o mínimo possível quando forem analisar os dados. Recomendam, entretanto, que métodos não paramétricos sejam utilizados *exceto* quando há uma *evidência positiva* de que os métodos paramétricos são adequados. Até o momento, não há resposta definitiva determinando qual atitude é preferível. E provavelmente nunca haverá tal resposta.

JANE SUPERBRAIN 15.1

Non-parametric tests and statistical power ②

Ranking the data is a useful way around the distributional assumptions of parametric tests but there is a price to pay: by ranking the data we lose some information about the magnitude of differences between scores. Consequently, non-parametric tests can be less powerful than their parametric counterparts. Statistical power (section 2.6.5) refers to the ability of a test to find an effect that genuinely exists. So, by saying that non-parametric tests are less powerful, we mean that if there is a genuine effect in our data then a parametric test is more likely to detect it than a non-parametric one. However, this statement is true only *if the assumptions of the parametric test are met*. So, if we use a parametric test and a non-parametric test on the same data, and those data meet the appropriate assumptions, then the parametric test will have greater power to detect the effect than the non-parametric test.

The problem is that to define the power of a test we need to be sure that it controls the Type I error rate (the number of times a test will find a significant effect when in reality there is no effect to find – see section 2.6.2). We saw in Chapter 2 that this error rate is normally set at 5%. We know that when the sampling distribution is normally distributed then the Type I error rate of tests based on this distribution is indeed 5%, and so we can work out the power. However, when data are not normal the Type I error rate of tests based on this distribution won't be 5% (in fact we don't know what it is for sure as it will depend on the shape of the distribution) and so we have no way of calculating power (because power is linked to the Type I error rate – see section 2.6.5). So, although you often hear (in the first edition of this book for example!) of non-parametric tests having an increased chance of a Type II error (i.e. more chance of accepting that there is no difference between groups when, in reality, a difference exists), this is true only if the sampling distribution is normally distributed.

- TESTES NÃO-PARAMÉTRICOS SÃO SEMPRE VÁLIDOS, MAS NEM SEMPRE PODEROSOS O SUFICIENTE *(para mostrar que a ≠ é significativa)*

- TESTES PARAMÉTRICOS SÃO OS MAIS PODEROSOS *(para mostrar que a ≠ é significativa)*, MAS NEM SEMPRE VÁLIDOS

Nahm. Korean Journal Anesthesiology. 2016

Tipos de Experimentos

Escala de mensuração	2 grupos de indivíduos ≠s	3 ou + grupos de indivíduos ≠s	Medida antes e depois no mesmo indivíduo	Multiplas medidas no mesmo indivíduo c/ um grupo apenas	Multiplas medidas no mesmo indivíduo c/ 2 ou + grupos ≠s	Associação entre 2 variáveis
Intervalar (Dist. Normal)*	Teste T não-pareado	Análise de variância (ANOVA)	Teste T pareado	ANOVA para medidas repetidas (One-Way ANOVA	Two-Way ANOVA (medidas repetidas c/ 2 ou + grupos)	Regressão Linear Correlação Pearson
Nominal	X^2 **	X^2 **	McNemar	Cochrane Q	-	Point-Biseral correlação p/variável dicotômica
Ordinal	Mann-Whitney	Kruskal-Wallis	Wilcoxon	Friedman	-	Correlação Spearman

- *Se a distribuição NÃO É "NORMAL", utilize os métodos para a escala de mensuração ORDINAL ou Considere transformar
- ** Utilizar teste exato de Fisher se as amostras são pequenas (< 5 em uma das casas)