Bioestatística Básica
RCA 5804

COMPARANDO 2 ou + GRUPOS INDEPENDENTES

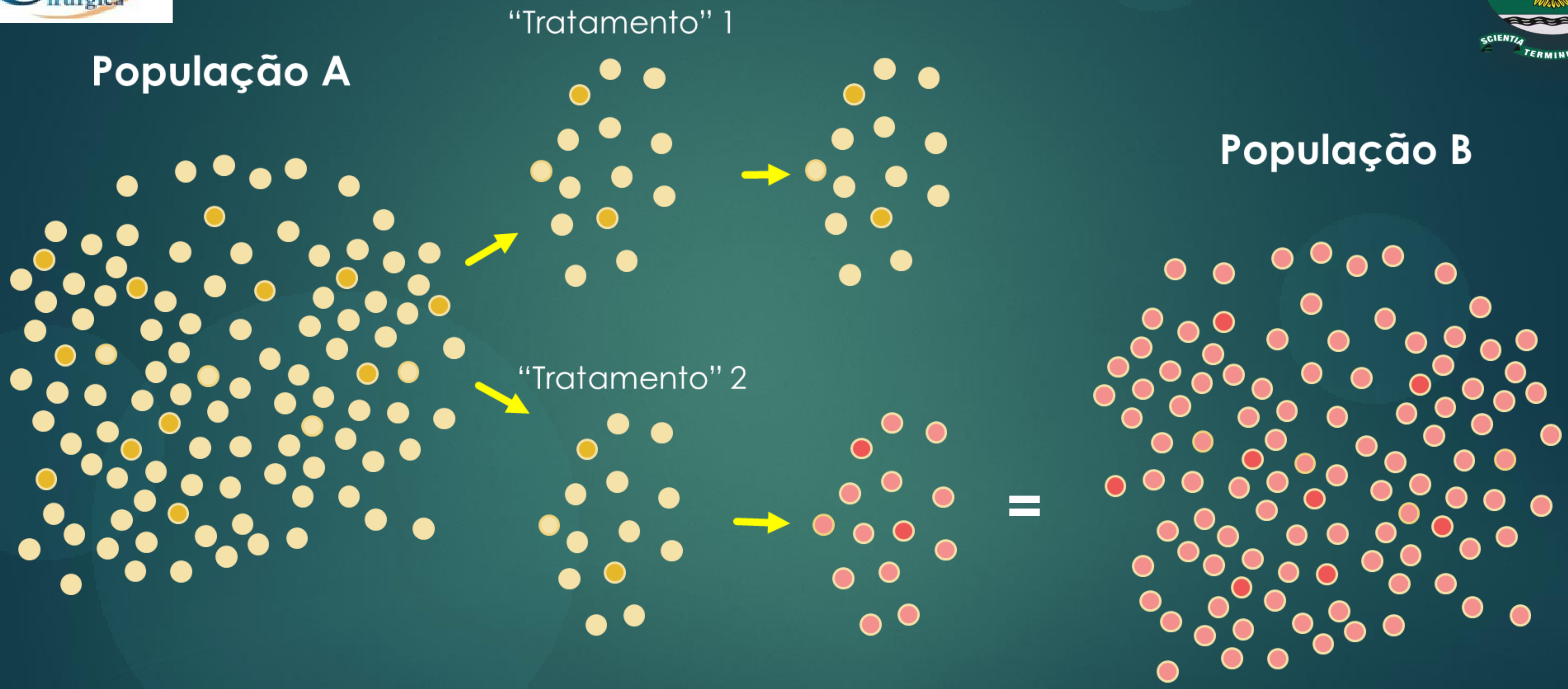
Prof. Dr. Alfredo J Rodrigues

Departamento de Cirurgia e Anatomia

Faculdade de Medicina de Ribeirão Preto

Universidade de São Paulo

alfredo@fmrp.usp.br



Indivíduos do "Tratamento" 2 ≠ dos Indivíduos do "Tratamento" 1, portanto provêm de populações dierentes

Inferência estatística (teste de hipóteses)

Inferência Estatística consiste em ”testar hipóteses” acerca de uma população através de evidências fornecidas por amostra(s).

Hipótese H0 ou Nula

PROVÊM DA MESMA POPULAÇÃO?

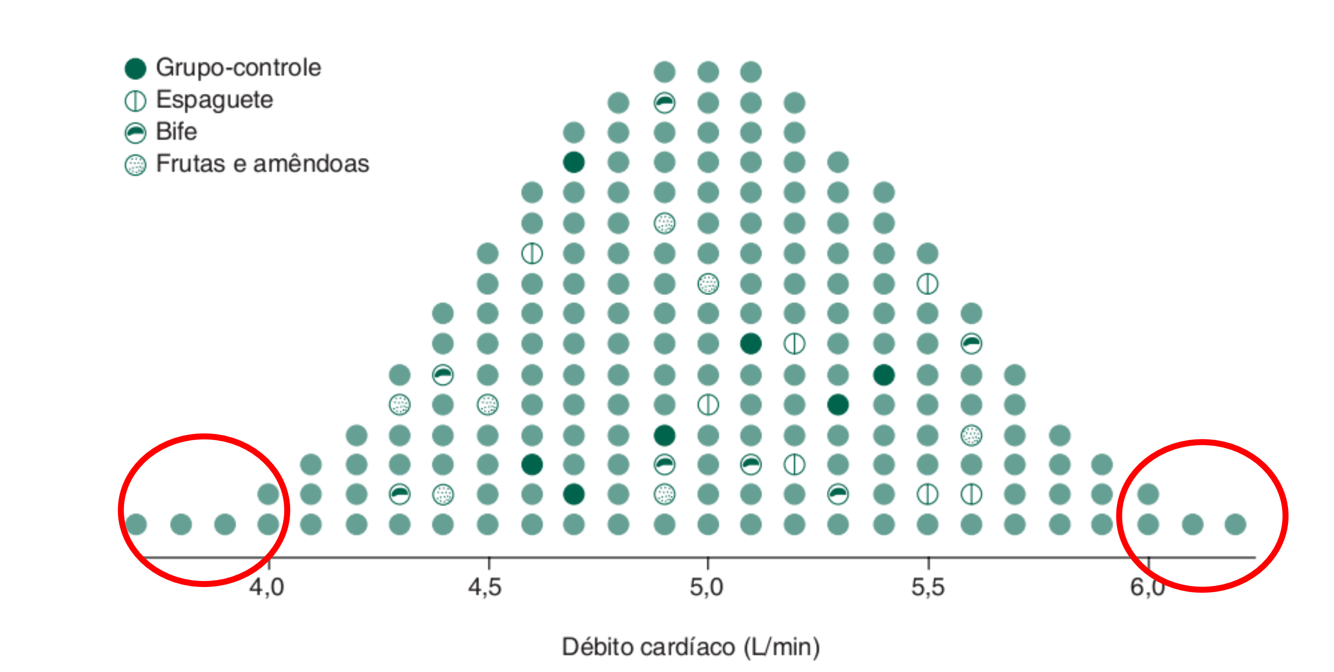
AMOSTRAS

Hipótese H1 ou alternativa

PROVÊM DE POPULAÇÕES DIFERENTES?

Hipótese nula ou H0:

- Os indivíduos da população tem débito cardíaco (DC) diferentes
- Alguns indivíduos tem DC maiores ou menores que a maioria. Eles estão nas extremidades da curva

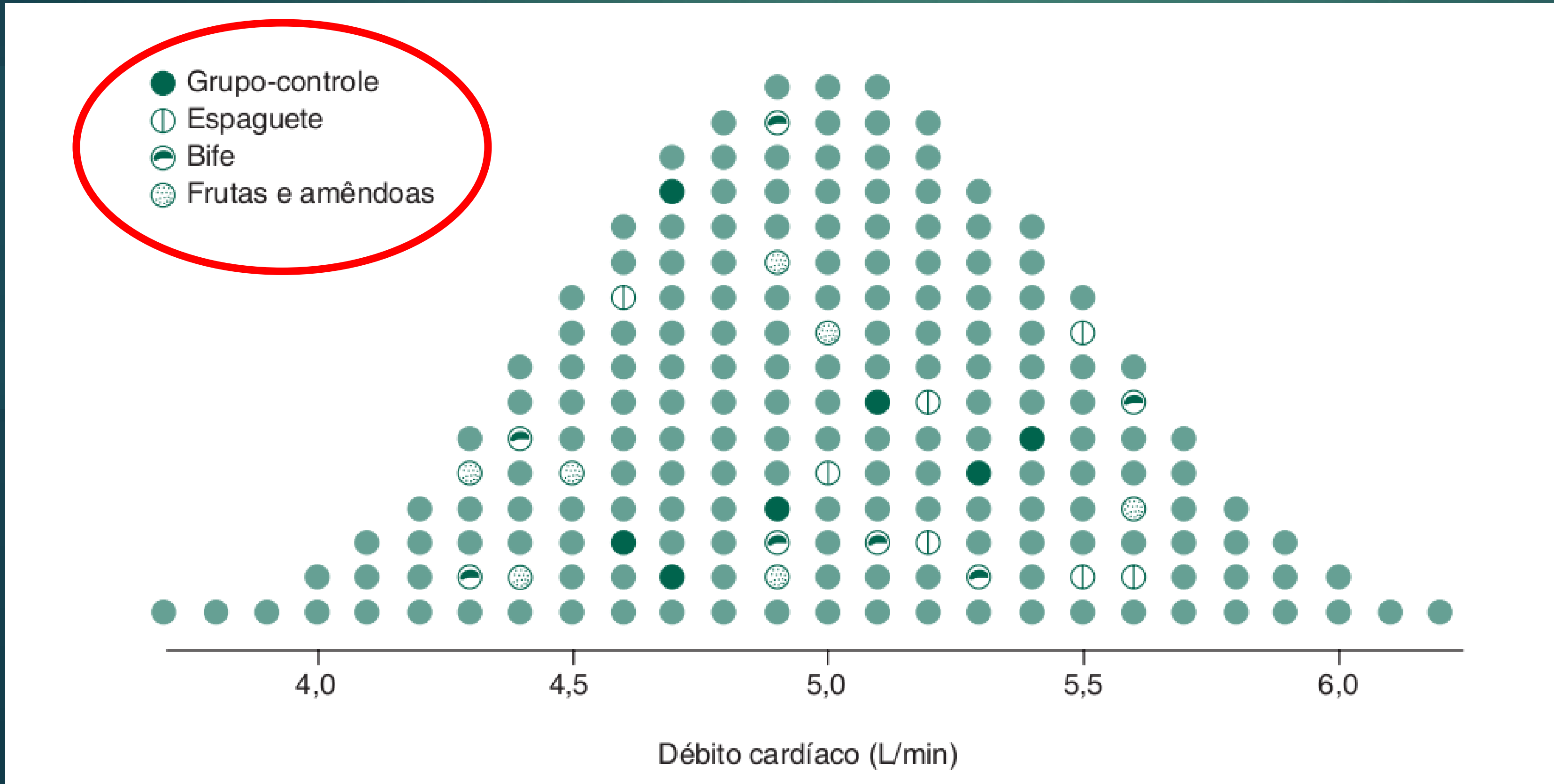


Pesquisa

A dieta influencia no débito cardíaco?

Hipótese nula ou H0: a dieta não afeta o DC

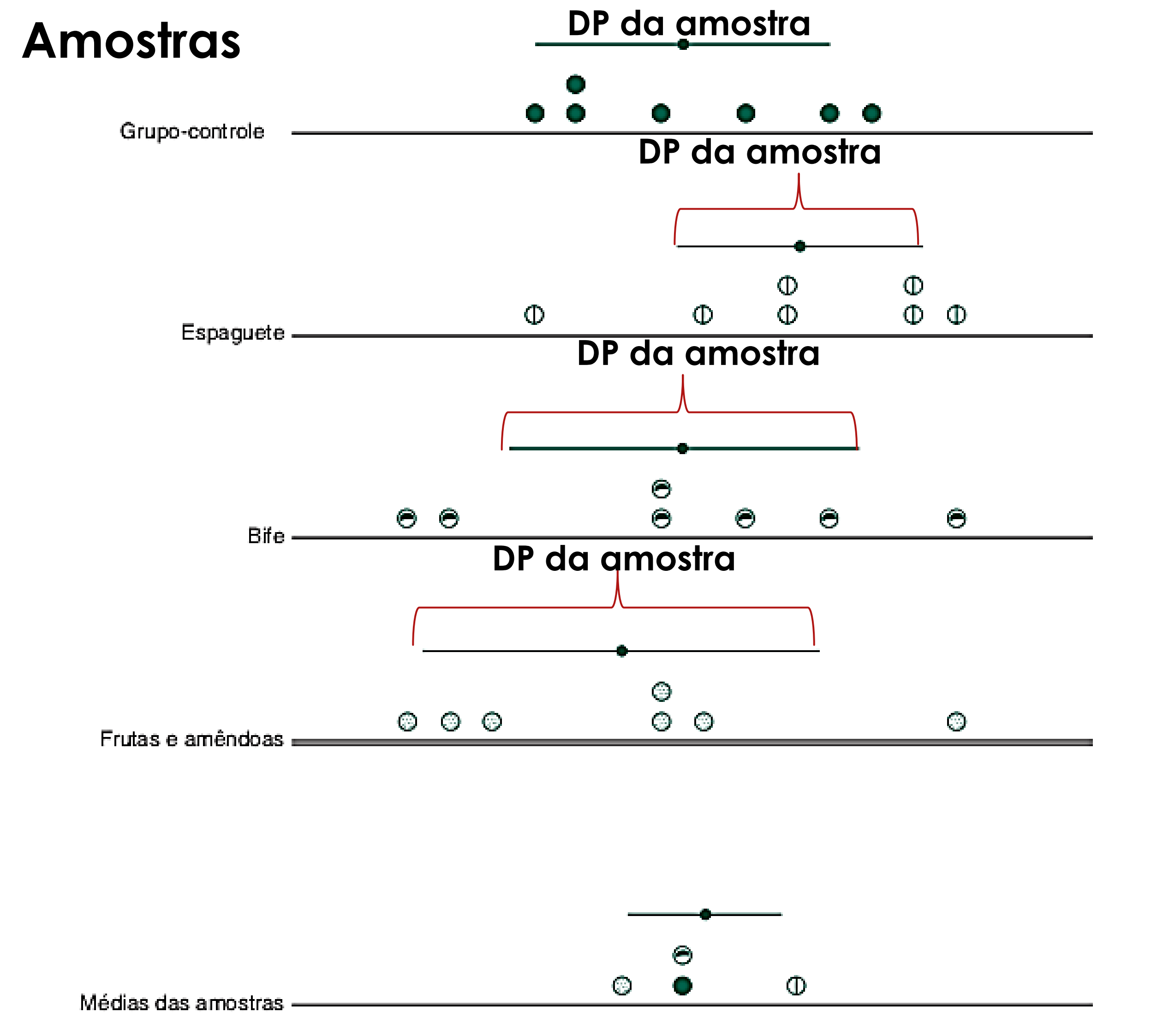
Hipótese alternativa ou H1: a dieta afeta o DC



Após a intervenção os indivíduos de cada grupo são semelhantes (mesma população) ou dessemelhantes (populações ≠) ?

Média e DP das amostras

Se H0 é verdadeira então:

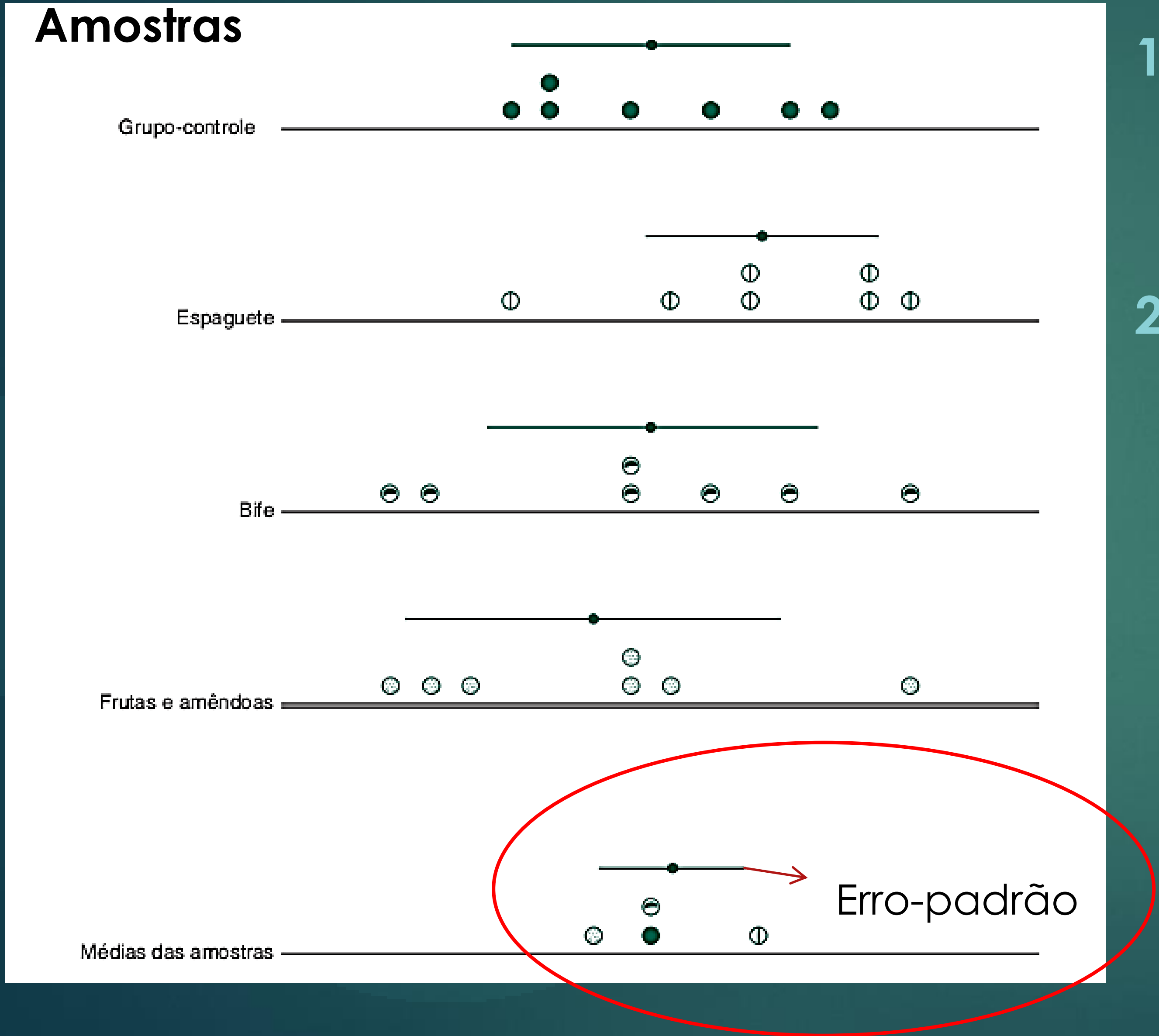


1. A variância, ou DP de cada amostra é uma estimativa da variância na população (se elas provêm da mesma população)
2. A média da variâncias (DP) das amostras será uma estimativa da variância populacional

$$s_{\text{dentro}}^2 = \frac{1}{4}\left(s_{\text{controle}}^2 + s_{\text{espaguete}}^2 + s_{\text{bife}}^2 + s_{\text{frutas}}^2\right)$$

Média e DP das amostras

Se H0 é verdadeira então:



1. A variância populacional pode ser estimada a partir das médias amostrais
2. A estimativa da variância populacional calculada entra as médias amostrais é o DP das médias amostrais (erro-padrã da média)

$$s^2_{entre} = n s^2_{\bar{X}}$$

Erro-padrão

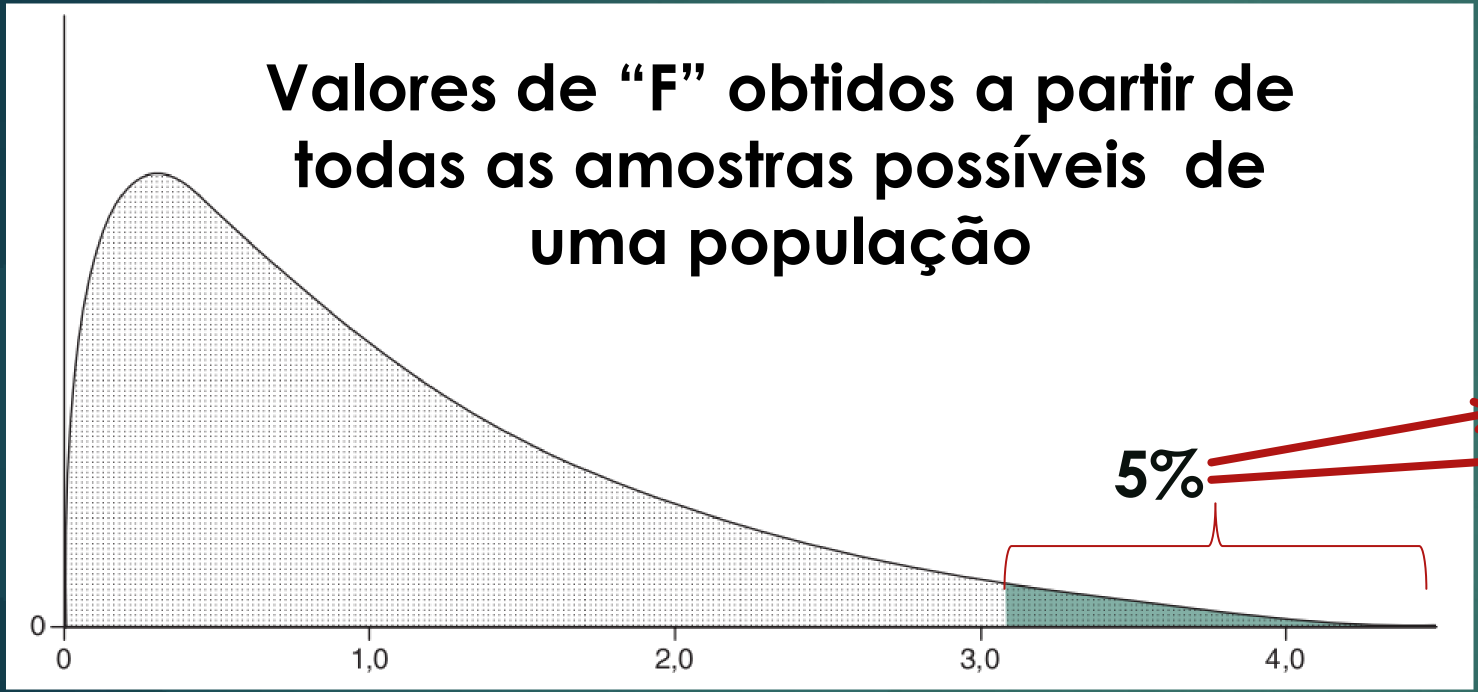
ESTATÍSTICA "F" OU ANÁLISE DE VARIÂNCIA

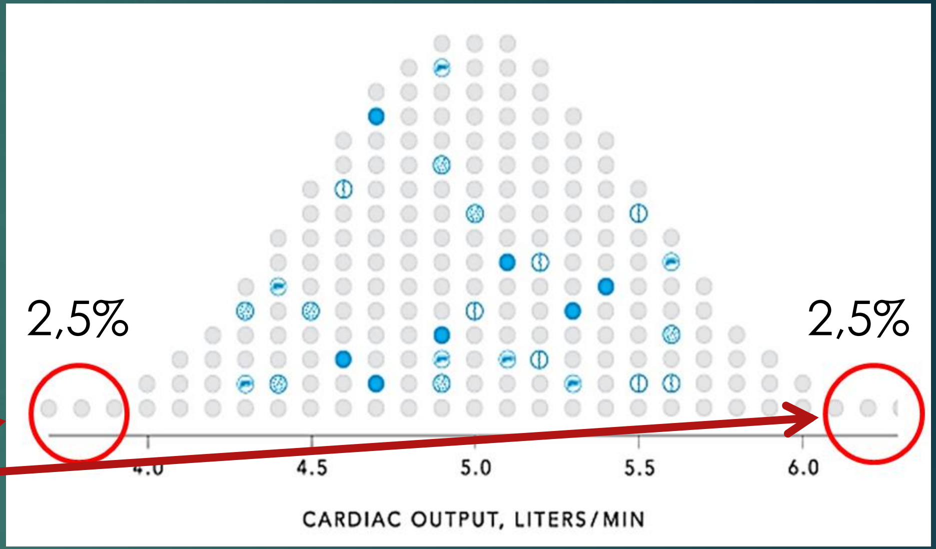
Se as amostras forem da mesma população estas 2 estimativas (S^2_{dentro} e S^2_{entre}) fornecerão resultados semelhantes, portanto:

$$F = \frac{s^2_{entre}}{s^2_{dentro}} = F = \frac{\text{Variância populacional estimada a partir das médias amostrais}}{\text{Variância populacional estimada como média das variâncias amostrais}} \cong 1$$

Se **VALOR de "F" é "grande"** significa que as amostras **não provêm da mesma população**

O que seria um valor de "F" "grande"?





- O ACASO pode levar o pesquisador a obter amostras das extremidades e nesse caso "F" poderá ter valor "grande"

- Portanto temos 5% ou 1% (a escolha é do pesquisador) de que o valor de “F” seja grande por ACASO (azar) e assim rejeita-se erroneamente a hipótese H0. **Este é o chamado erro tipo I ou alfa**

O que é “F” Grande?

- “F” é “grande” a partir do valor (cutoff) que engloba os 5% ($p<0,05$), ou 1% ($P<0,01$), das amostras das extremidade dos valores.

- O **valor de F** depende do **tamanho das amostras** e do **número de amostras** e **é verificado em tabelas**
- Em geral os softwares calculam valor de “p” para o valor de “F” obtido

Graus de Liberdade

V1= número de amostras – 1

V2 = número de amostras x (tamanho das amostras -1)

# P< 0,05

v_2 \ v_1	1	2	3	4	5	6	7	8	9	10
1	161.448	199.500	215.707	224.583	230.162	233.986	236.768	238.882	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.687	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
21	4.325	3.467	3.072	2.840	2.685	2.573	2.488	2.420	2.366	2.321
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255
25	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	2.236
26	4.225	3.369	2.975	2.743	2.587	2.474	2.388	2.321	2.265	2.220
27	4.210	3.354	2.960	2.728	2.572	2.459	2.373	2.305	2.250	2.204

ATENÇÃO

- "F" foi obtido a partir de medidas de **tendência central e variância**

- Estas são parâmetros da população somente se a população tiver distribuição simétrica (normal) por esta razão o teste é "F" é **PARAMÉTRICO**

- Como no caso existe apenas **um fator** (a dieta no exemplo) o teste é chamada apenas de Análise de Variância ("One-Way Analysis of Variance") ou "**ANOVA**"
- A "ANOVA" serve **para comparar 2 ou mais grupos**

ANOVA

Condições necessárias para seu uso:

1. A população da qual derivam deve ter **distribuição normal**
2. Cada amostra deve ser **independente** das demais
3. Cada amostra deve ser selecionada **aleatoriamente**
4. **Atenção: As variâncias entre amostras devem ser semelhantes** (Homogeneidade de variâncias ou homocedasticidade)

Teste de Levine: $p<0,05$: sem homogeneidade

ANOVA

- A chance de **erro tipo I** é maior quando as **variâncias das amostras (populações) forem muito diferentes**,
- ANOVA é razoavelmente **robusto** (menor chance de cometer erro tipo I) se as **amostras são do mesmo tamanho**
- O efeito da desigualdade das variâncias **é mais grave quando os tamanhos das amostras são desiguais.**
- **Alternativa para "não-homogeneidade": Testes "Robustos"**
  - **Teste *Welch F***
  - ***Teste de Brown-Forsythe***

ANOVA SPSS

Descriptives

GSHLBA

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
controle	10	366,2000	170,98720	54,07090	243,8831	488,5169	103,00	714,00
6h AO	10	794,5000	609,07950	192,60785	358,7908	1230,2092	172,00	2143,00
Total	20	580,3500	487,69784	109,05255	352,1004	808,5996	103,00	2143,00

Test of Homogeneity of Variances

GSHLBA

Levene Statistic	df1	df2	Sig.
7,527	1	18	,013

ANOVA

GSHLBA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	917204,450	1	917204,450	4,584	,046
Within Groups	3601930,100	18	200107,228		
Total	4519134,550	19			

Robust Tests of Equality of Means

GSHLBA

	Statistic[a]	df1	df2	Sig.
Welch	4,584	1	10,410	,057
Brown-Forsythe	4,584	1	10,410	,057

a. Asymptotically F distributed.

Erro tipo I

Teste “T”

- Teste “T” é uma forma diferente da estatística “F”, pois $F=T^2$, , e que compara média de **apenas 2 grupos**.
- O critério de decisão para se rejeitar a hipótese nula (H0), é o tamanho de “T”, semelhantemente ao “ tamanho de F”.
- Existem tabelas para os valores de “T” assim como para os “F” .

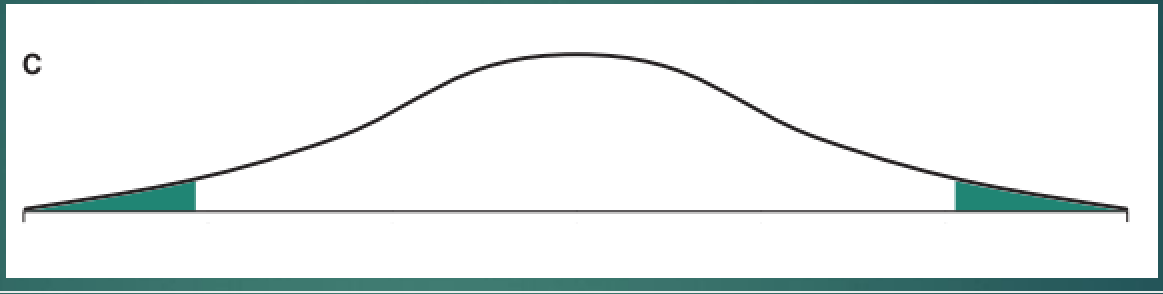
$$t = \frac{\text{Diferença nas médias amostrais}}{\text{Erro-padrão da diferença das médias amostrais}}$$

- Quando **"T" é "pequeno"** concluímos que ambas as amostras provêm da mesma população, (não há diferença significativa entre elas)
- Quanto **"T" é "GRANDE"** concluímos que é **pouco provável** ambas as amostras provêm da mesma população (**há diferença significativa entre elas**)

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{(s^2/n) + (s^2/n)}}$$

- Assim como para F há uma gama de valores possíveis para "T" que dependem de graus de liverdade (v).
- V = 2(n-1) pata o teste T

Distribuição dos Valores de "T" para todas as amostras possíveis de uma população infinita é ilustrada pela figura abaixo.



A porção verde mostras os valores extremos de "T" com 5% de chance >2,1 ou < -2,1

- A distribuição de valores de "T" é simétrica ("F" era unicaudal)
- A maioria dos valores esta em torno de "0" e **raramente estão abaixo de -2.1 ou acima de -2.1**
- **O quão raramente?**
  - 2,5% das vezes abaixo de -2,1 e 2,5% das vezes acima de 2.1 **(bicaudal = total 5%)**

Tabela valores críticos de “t”

- Graus de liberdade
- (v)= 2x(n-1): amostras de mesmo tamanho
- V= n1+n2 - 2: amostras de tamanho desigual
  - n= tamanho de CADA AMOSTRA

Grupo 1 e 2 : n=5 cada
v=8
Grupo 1: n=5, Grupo 2 n=7
V= 9

Degrees of Freedom	Probability, p			
	0.1	0.05	0.01	0.001
1	6.31	12.71	63.66	636.62
2	2.92	4.30	9.93	31.60
3	2.35	3.18	5.84	12.92
4	2.13	2.78	4.60	8.61
5	2.02	2.57	4.03	6.87
6	1.94	2.45	3.71	5.96
7	1.89	2.37	3.50	5.41
8	1.86	2.31	3.36	5.04
9	1.83	2.26	3.25	4.78
10	1.81	2.23	3.17	4.59
11	1.80	2.20	3.11	4.44
12	1.78	2.18	3.06	4.32

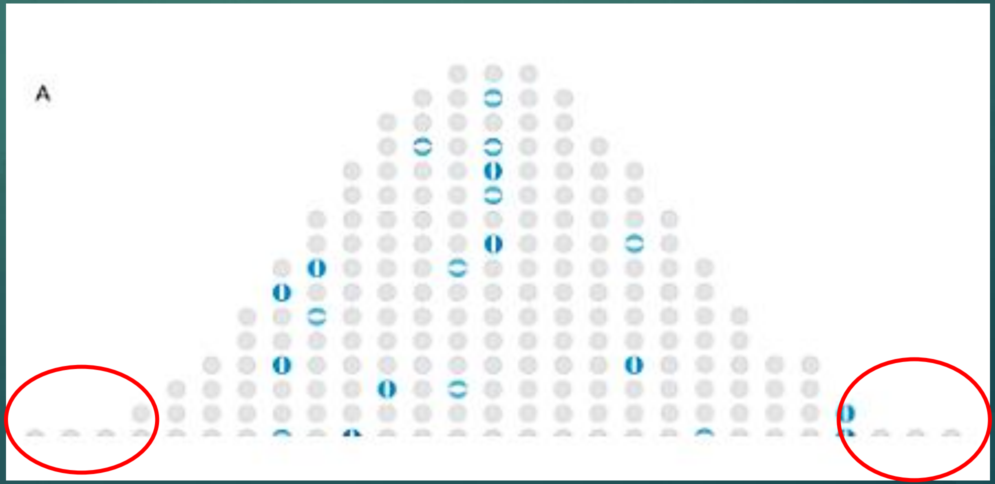
Qual a relação da cerveja com a estatística?

William Sealy Gosset (1876−1937)

Pseudônimo "Student"

Concluindo

- Quando o pesquisador relata um valor de $p<0,05$ ele quer dizer que há menos de 5% de chance de que a diferença observada (**e que forneceu um valor de "T" grande**) é decorrente do acaso ("azar"), ou seja, as amostras, **escolhidos aleatoriamente, c**ontinha predominantemente indivíduos da população com valores extremos**.** (exemplo da figura abaixo).
  - Esse é o chamado **erro tipo I ou alfa**



- Assim, o valor de “p” significa a probabilidade de encontrar uma DIFERENÇA que na realidade NÃO EXISTE (erro tipo I ou erro a)

- Existe o erro tipo II, ou β, quando NÃO SE CONSEGUE OBERVAR uma DIFERENÇA que na realidade EXISTE

Test “T“ de Student

- **Atenção**:
  - Análise de variância compara 2 ou + grupos
  - “ Teste T” APENAS para 2 grupos
- Condições para Uso teste “T”:
  1. Cada amostra deve ser independente
  2. Cada amostra deve ser selecionada aleatoriamente
  3. A POPULAÇÃO DA QUAL DERIVAM DEVE TER DISTRIBUIÇÃO NORMAL
  4. AS VARIÂNCIAS DAS AMOSTRAS DEVEM SER SEMELHANTES
  5. AS AMOSTRAS DEVEM TER TAMANHOS SEMELHANTES

Efeito de tamanho da amostra no teste "T"

Group Statistics

	grupo	N	Mean	Std. Deviation	Std. Error Mean
VAR00002	1,00	13	2,3846	1,02376	,28394
	2,00	12	2,6250	1,04718	,30230

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
VAR00002	Equal variances assumed	,000	,986	-,580	23	,567	-,24038	,41434	-1,09752	,61675
	Equal variances not assumed			-,580	22,744	,568	-,24038	,41473	-1,09886	,61809

Group Statistics

	grupo	N	Mean	Std. Deviation	Std. Error Mean
VAR00002	1,00	156	2,3846	,98676	,07900
	2,00	144	2,6250	1,00610	,08384

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(s_1^2/n) + (s_2^2/n)}}$$

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
VAR00002	Equal variances assumed	,004	,951	-2,088	298	,038	-,24038	,11511	-,46692	-,01385
	Equal variances not assumed			-2,087	295,066	,038	-,24038	,11520	-,46710	-,01367

LEMBRANDO

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
Peso	Equal variances assumed	,700	,403	7,640	1017	,000	6,79758	,88978	5,05157	8,54358
	Equal variances not assumed			7,623	980,294	,000	6,79758	,89171	5,04769	8,54747

Além da distribuição normal, **os testes paramétricos (T e F) exigem que a variância e tamanho das amostras sejam semelhantes.** Muitos softwares fornecem testes para verificar essa semelhança (como o "Levene").

Quando a variância e/ou tamanho das amostra não são semelhantes

PROBLEMAS

Group Statistics

	Operações resumidas por tipo	N	Mean	Std. Deviation	Std. Error Mean
Idade	Revascularização	655	61,57	10,031	,392
	Outras	28	56,68	17,762	3,357

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
Idade	Equal variances assumed	28,190	,000	2,427	681	,015	4,892	2,016	,934	8,851
	Equal variances not assumed			1,448	27,741	,159	4,892	3,379	-2,033	11,818

Teste “Robustos” : Welch ou Brown-Forsythe

Quando a variância e/ou tamanho das amostra não são semelhantes

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
Idade	Equal variances assumed	28,190	,000	2,427	681	,015	4,892	2,016	,934	8,851
	Equal variances not assumed			1,448	27,741	,159	4,892	3,379	-2,033	11,818

Robust Tests of Equality of Means

Idade

	Statistic[a]	df1	df2	Sig.
Welch	2,096	1	27,741	,159
Brown-Forsythe	2,096	1	27,741	,159

a. Asymptotically F distributed.

ERRO COMUM COM O TESTE “T”

teste “t” 2 (g1 x g3)

teste “t” 1 (g1 x g2)

teste “t” 3 (g1 x g4)

Grupo 1 Grupo 2 Grupo 3 Grupo 4

- O teste de variânça mostra se há ou não ≠ entre as amostras, mas caso haja ≠ o teste não mostra entre quais amostras esta a diferença.
- Então muitos se utilizam do artifício acima (vários testes “T”)

•Ocorre que:

•Se p<0,05 foi escolhido, teremos 5% de (t1) + 5% (t2) + 5%(t3) = 15%, assim a chance de encontrar “t” grande (significativo) por acaso (azar) sem existir ≠ realmente (erro tipo I) é a soma de todos os a (15%).

- **Nos casos de 2 ou + grupos utiliza-se ANOVA o** que mostrará que ao menos um não provêm da mesma população (rejeita-se H0)
- Todavia, quando há mais de 2 grupos, como descobrir entre quais esta a ≠ , se ela existe?
  - Utilizando *procedimentos , ou testes, de comparação múltipla ou **Pós-testes (Post-Hoc)**:*

Bonferroni

Muito conservador : bom controle sobre erro tipo I, mas perde poder estatístico (capacidade de detector diferença quando ela existe), principalmente se usado para comparar **muitos grupos.**

Pós-testes (Post-Hoc)

Dunnet's

Pode ser usado para comparar múltiplos grupos contra o um "controle"

Tukey

Conservador como Bonferroni, mas tem maior poder estatístico que Bonferroni para múltiplos grupos

Pós-testes

Games-Howell

- Grupos com variância diferentes
- Grupos com "n" diferentes
- Poderoso, mas pode ser liberal quando as amostras são pequenas

GT2 Hochbes e Gabriel's pairwise

- Ambos com bom desempenho para mostras com "n" diferentes
- GT2 Hochberg não confiável com amostras de varianças diferentes
- Gabriel: muito liberal quando os tamanhos de amostra são muito diferentes

8.2.11.2. *Post hoc* procedures and violations of test assumptions ②

Most research on *post hoc* tests has been looked at whether the test performs well when the group sizes are different (an unbalanced design), when the population variances are very different, and when data are not normally distributed. The good news is that most multiple comparison procedures perform relatively well under small deviations from normality. The bad news is that they perform badly when group sizes are unequal and when population variances are different.

Hochberg's GT2 and *Gabriel's* pairwise test procedure were designed to cope with situations in which sample sizes are different. Gabriel's procedure is generally more powerful but can become too liberal when the sample sizes are very different. Also, Hochberg's GT2 is very unreliable when the population variances are different and so should be used only when you are sure that this is not the case. There are several multiple comparison procedures that have been specially designed for situations in which population variances differ. SPSS provides four options for this situation: *Tamhane's T2*, *Dunnett's T3*, *Games–Howell* and *Dunnett's C*. Tamhane's T2 is conservative and Dunnett's T3 and C keep very tight Type I error control. The Games–Howell procedure is the most powerful but can be liberal when sample sizes are small. However, Games–Howell is also accurate when sample sizes are unequal.

8.2.11.3. Summary of *post hoc* procedures ②

The choice of comparison procedure will depend on the exact situation you have and whether it is more important for you to keep strict control over the familywise error rate or to have greater statistical power. However, some general guidelines can be drawn (see Toothaker, 1993). When you have equal sample sizes and you are confident that your population variances are similar then use REGWQ or Tukey as both have good power and tight control over the Type I error rate. Bonferroni is generally conservative, but if you want guaranteed control over the Type I error rate then this is the test to use. If sample sizes are slightly different then use Gabriel's procedure because it has greater power, but if sample sizes are very different use Hochberg's GT2. If there is any doubt that the population variances are equal then use the Games–Howell procedure because this generally seems to offer the best performance. I recommend running the Games–Howell procedure in addition to any other tests you might select because of the uncertainty of knowing whether the population variances are equivalent.

Although these general guidelines provide a convention to follow, be aware of the other procedures available and when they might be useful to use (e.g. Dunnett's test is the only multiple comparison that allows you to test means against a control mean).

O que é um teste conservador ?

É aquele que **tem menor probabilidade de cometer o erro tipo I**, ou seja, encontrar ≠ onde ela não existe.

Consequentemente, um teste conservador tem maior probabilidade de não detectar uma ≠ que realmente existe (erro tipo II ou β).

Erro do tipo I: decisão de rejeitar H0 quando de fato H0 é verdadeira.

Erro do tipo II: decisão de não rejeitar H0 quando de fato H0 é falsa