Bioestatística Básica

RCA 5804

a. Comparando Proporções
b. Qual o real Significado de não-significante
c. Intervalos de confiança da diferença de medias

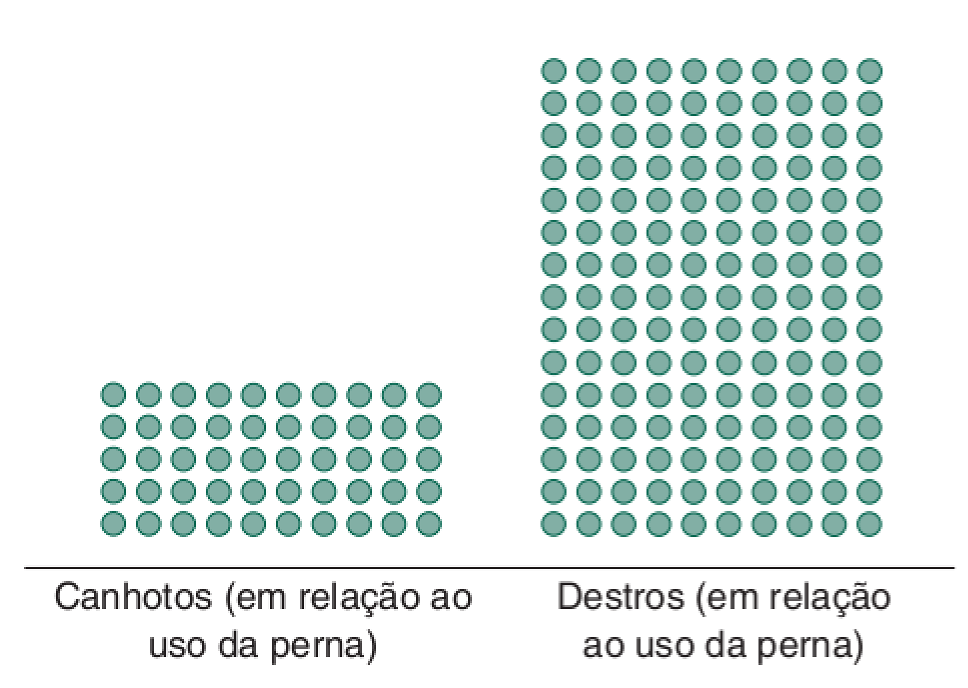
Prof. Dr. Alfredo J Rodrigues

Departamento de Cirurgia e Anatomia
Faculdade de Medicina de Ribeirão Preto
Universidade de São Paulo

alfredo@fmrp.usp.br

Dados nominais

Duas classes mutuamente excludentes

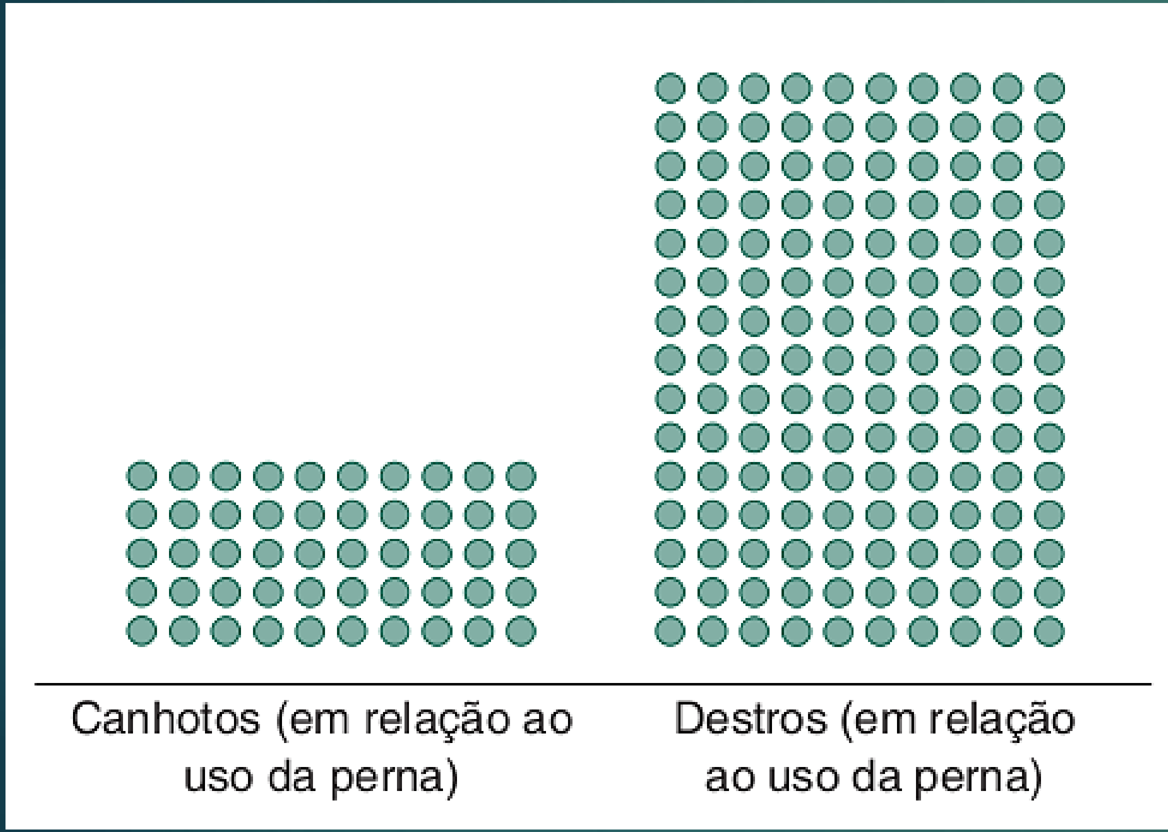


$P_{destro} = 1 - P_{canhoto}$

$$\sigma = \sqrt{p(1-p)}$$

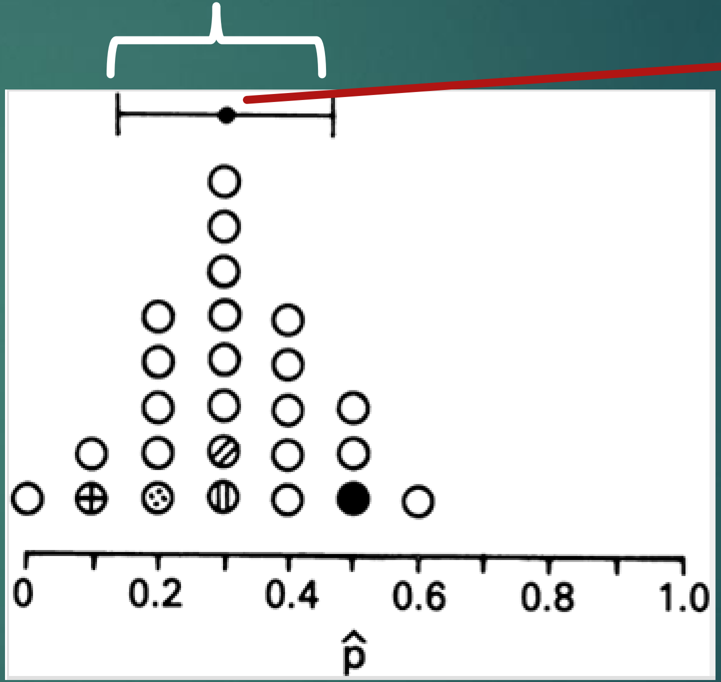
(Equivalente à média")

Qual a proporção de canhotos ?
4 amostras de 10 indivíduos



Erro-padrão da proporção média (P)

P média



- Erro-padrão $\Rightarrow \sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$

Teste de hipóteses para Proporções

O análogo de "t" para proporções é a **Estatística Z**

$$z = \frac{\text{Diferença das proporções amostrais}}{\text{Erro-padrão das diferenças das proporções amostrais}}$$

$$z = \frac{|\hat{p}_1 - \hat{p}_2| - \frac{1}{2}(1/n_1 + 1/n_2)}{\sqrt{\hat{p}(1 - \hat{p})(1/n_1 + 1/n_2)}}$$

- Quando **“Z”** é “GRANDE” concluímos que é **pouco provável que** ambas as amostras provêm da mesma população.

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

- **Estatística "Z"** testa hipótese **no qual existe apenas 2 desfechos mutuamente excludentes e a probabilidade permanece constante** (ensaio independente de Bernoulli)
- Nas situações nas quais existe **mais de dois grupos** ou **mais de dois eventos possíveis** há necessidade de um **ANÁLOGO AO TESTE DE VARIÂNCIA (ANOVA)**.
- O "análogo" ao teste Anova é o **teste X^2 (Chi-quadrado** ou **Chi-quadrado de Pearson com** correção de Yates para tabelas 2x2)
- **Para comparações múltiplas em tables maiores 2x2 (como ANOVA com >2 grupos) há a necessidade de testes múltiplos 2x2 com correção de erro grupal (Bonferroni, Holm, Holm-Sidak, etc)**

Tratamento	Sobreviveu	Morreu	total
Tratado	459	26	485
Controle	141	27	168
	600	53	653

600 (459+141) sobreviventes/653 indivíduos (total) x 100 = 91,8 % sobreviventes

1. Se H0 é verdadeira (tratamento não tem efeito) **o ESPERADO** é que 91,8 % do grupo Tratado (445 indivíduos) e que 91,8% do grupo controle (145 indivíduos) sobrevirão

2. O OBSERVADO foi 459 sobreviventes no grupo tratado e 141 no controle.

ESTA DIFERENÇA ENTRE O OBSERVADO E O ESPERADO É DEVIDA APENAS AO ACASO?

teste X^2 - Chi-quadrado

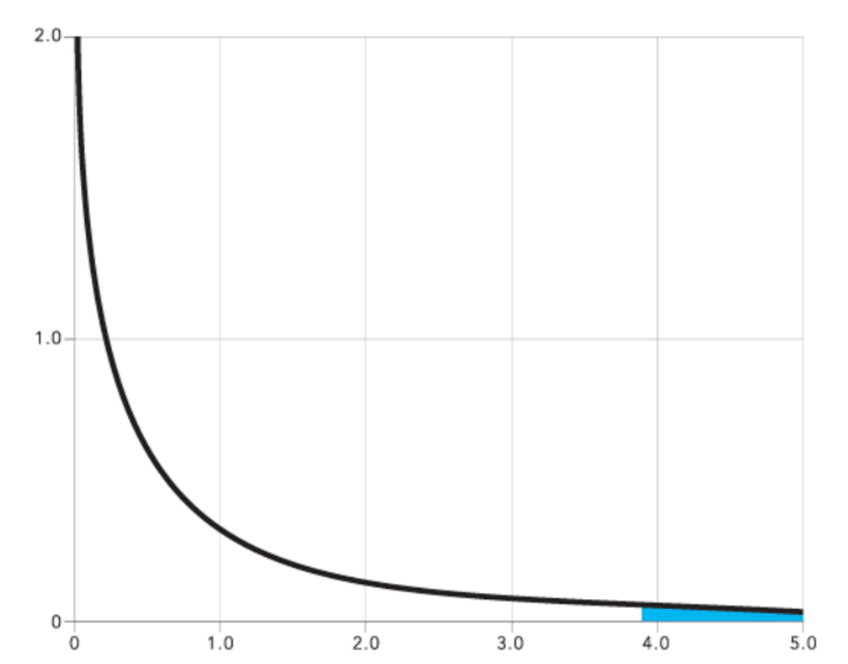
$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Correção de

Yates para tabelas 2x2

$$\chi^2 = \sum \frac{(|O-E| - \frac{1}{2})^2}{E}$$

Teste X^2 (Chi-quadrado)



A distribuiça de X^2 depende do "n" de indivíduos e número de possíveis "desfecho" (graus de liberdade)

Quando o valor **de X^2 é "GRANDE"** (tabela) concluímos que é **pouco provável que** ambas as amostras provêm da mesma população, rejeita-se H0 (diferença é significativa).

Tratamento	Sobreviveu	Morreu	total
Tratado	459	26	485
Controle	141	27	168
	600	53	653

$$\chi^2 = \sum \frac{(|O - E| - \frac{1}{2})^2}{E}$$

= 19,194

P<0,001

Probabilidade de um valor maior de *P*								
ν	0,50	0,25	0,10	0,05	0,025	0,01	0,005	0,001
1	0,455	1,323	2,706	3,841	5,024	6,635	7,879	10,828

Valores críticos para X^2

	Probabilidade de um valor maior de *P*							
ν	0,50	0,25	0,10	0,05	0,025	0,01	0,005	0,001
1	0,455	1,323	2,706	3,841	5,024	6,635	7,879	10,828
2	1,386	2,773	4,605	5,991	7,378	9,210	10,597	13,816
3	2,366	4,108	6,251	7,815	9,348	11,345	12,838	16,266
4	3,357	5,385	7,779	9,488	11,143	13,277	14,860	18,467
5	4,351	6,626	9,236	11,070	12,833	15,086	16,750	20,515
6	5,348	7,841	10,645	12,592	14,449	16,812	18,548	22,458
7	6,346	9,037	12,017	14,067	16,013	18,475	20,278	24,322
8	7,344	10,219	13,362	15,507	17,535	20,090	21,955	26,124
9	8,343	11,389	14,684	16,919	19,023	21,666	23,589	27,877
10	9,342	12,549	15,987	18,307	20,483	23,209	25,188	29,588
11	10,341	13,701	17,275	19,675	21,920	24,725	26,757	31,264
12	11,340	14,845	18,549	21,026	23,337	26,217	28,300	32,909
13	12,340	15,984	19,812	22,362	24,736	27,688	29,819	34,528
14	13,339	17,117	21,064	23,685	26,119	29,141	31,319	36,123
15	14,339	18,245	22,307	24,996	27,488	30,578	32,801	37,697
16	15,338	19,369	23,542	26,296	28,845	32,000	34,267	39,252
17	16,338	20,489	24,769	27,587	30,191	33,409	35,718	40,790
18	17,338	21,605	25,989	28,869	31,526	34,805	37,156	42,312

Graus de liberdade= (L-1)x(C-1)

Teste X^2 (Chi-quadrado)

- **Condições para uso teste X^2:**
  - X^2 utilizado para **dados categóricos**, portanto é **não-paramétrico e não necessita distribuição normal.**
  - Nas tabelas 2x2 o número de observações esperadas em **TODAS as células deve ser de no mínimo 5** (o que não ocorre usualmente com amostras pequenas).
    - **Não satisfeitas essa condição, deve-se utilizar o "TESTE EXATO DE FISHER", geralmente oferecido pelos softwares de estatística.**

Teste X^2 (Chi-quadrado)

- **Condições para uso teste X^2:**
  - NAS TABELAS **MAIORES** QUE 2X2 o número de observações esperadas em cada célula não deve ser < 1 e não > 20% das células deve ter < 5 observações. Se for o caso, **aumentar a coleta ou diminuir o numero de categorias**
  - As medidas devem ser independentes.

Teste X^2 (Chi-quadrado)

- **Hipótese H0: mortalidade em homens e mulheres é a mesma em cirurgia cardíaca.**
- **H1: mortalidade é ≠**

Sexo * OBITO Crosstabulation

			OBITO		Total
			não	sim	
Sexo	feminino	Count	513	74	587
		% within Sexo	87,4%	12,6%	100,0%
		% within OBITO	43,8%	46,3%	44,1%
	masculino	Count	658	86	744
		% within Sexo	88,4%	11,6%	100,0%
		% within OBITO	56,2%	53,8%	55,9%
Total		Count	1171	160	1331
		% within Sexo	88,0%	12,0%	100,0%
		% within OBITO	100,0%	100,0%	100,0%

Resultado

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)	Point Probability
Pearson Chi-Square	,340[a]	1	,560	,611	,308	
Continuity Correction[b]	,248	1	,618			
Likelihood Ratio	,339	1	,560	,611	,308	
Fisher's Exact Test				,611	,308	
Linear-by-Linear Association	,340[c]	1	,560	,611	,308	,057
N of Valid Cases	1331					

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 70,56.

b. Computed only for a 2x2 table

c. The standardized statistic is -,583.

- **Como nenhuma célula tem < 5 posso utilizer X^2, caso contrário utilizaria o valor fornecido pelo Fisher (no caso p=0,611)**
- p= 0,560 (> 0,05), não significativo. Não rejeito a hipótese nula, ou seja, não há ≠ na mortalidade entre homnes e mulheres

“p” significativo x Relevância

- “P” significativo não significa que o resultado é relevante, ou seja, que a diferença tenha relevância

Group Statistics

	Grupo	N	Mean	Std. Deviation	Std. Error Mean
PA	1,00	20	178,6500	8,15814	1,82422
	2,00	20	175,3500	2,85205	,63774

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	95% Confidence Interval of the Difference Upper
PA	Equal variances assumed	21,402	,000	1,708	3[illegible]	,096	3,30000	1,93248	-,61210	7,21210
	Equal variances not assumed			1,708	23,576	,101	3,30000	1,93248	-,69224	7,29224

Group Statistics

	Grupo	N	Mean	Std. Deviation	Std. Error Mean
PA	1,00	40	178,7750	8,11294	1,28277
	2,00	40	175,3500	2,81525	,44513

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	95% Confidence Interval of the Difference Upper
PA	Equal variances assumed	47,754	,000	2,522	78	,014	3,42500	1,35781	,72182	6,12818
	Equal variances not assumed			2,522	48,258	,015	3,42500	1,35781	,69532	6,15468

Redução de 178,7 mmHg para 175,3 mmHg (diferença de 3,4 mmHg), significante, mas provavelmente irrelevante clinicamente!

Poder Estatístico do Teste

Há diferença entre mostrar que o tratamento não tem efeito e falhar em demonstrar o efeito, sobretudo quando há limitações nos estudos !!

Conclusão a partir das observações	Situação real	
	Tratamento tem efeito	Tratamento não tem efeito
Tratamento tem efeito	Positivo-verdadeiro Conclusão correta $1 - \beta$	Falso-positivo Erro do Tipo I (α)
Tratamento não tem efeito	Falso-negativo Erro do Tipo II (β)	Negativo-verdadeiro Conclusão correta $1 - \alpha$

Poder do teste = 1- β (probabilidade de erro tipo II)

“Poder do teste”

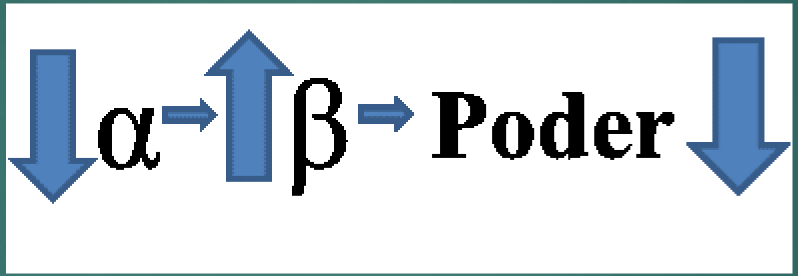
- A capacidade para detectar o efeito de um “tratamento”, ou seja, rejeitar H0
- Assim, se o poder do teste é 0,80, significa que a chance de detectar um positivo verdadeiro é de 80% (rejeitar H0).

Poder do Teste

- Depende:
  - Tamanho do efeito ou diferença a ser detectada
  - Variabilidade na população ou DP (σ)
  - Tamanho das amostras "n"
  - Tamanho do erro tipo de I
  - Tipo de teste estatístico a ser utilizado

Poder do Teste

Poder do teste= 1- (β Erro tipo II)



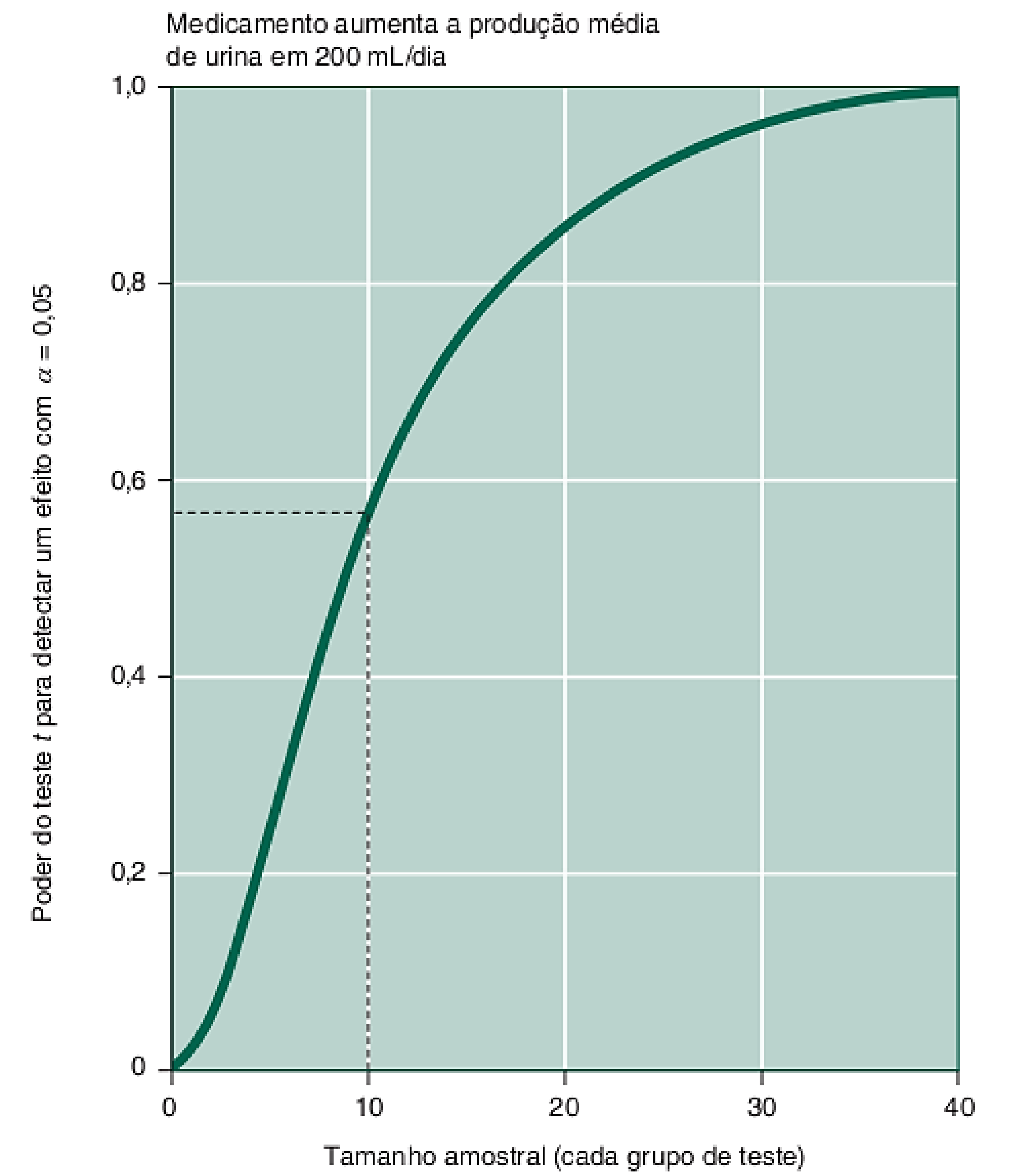
Poder do Teste

- Tamanho do efeito do tratamento (a diferença a se detectar) :
  - QUANTO MAIOR A ≠ a se detectar, MAIOR o PODER.
- Variabilidade na população
  - QUANTO MENOR o desvio-padrão, MAIOR o PODER
- Tamanho da amostra
  - QUANTO MAIOR A TAMANHO DA AMOSTRA, MAIOR o PODER
- ***O único modo de reduzir α e β é aumentando o tamanho das amostras***

Poder do Teste

- ***Daí a razão de ser necessário na elaboração do projeto de pesquisa que se saiba:***
  - ***O tamanho do efeito que se vai considerar relevante***
  - ***O tipo de teste a ser aplicado, para que se possa calcular o tamanho do "n" para determinado "poder do teste"***

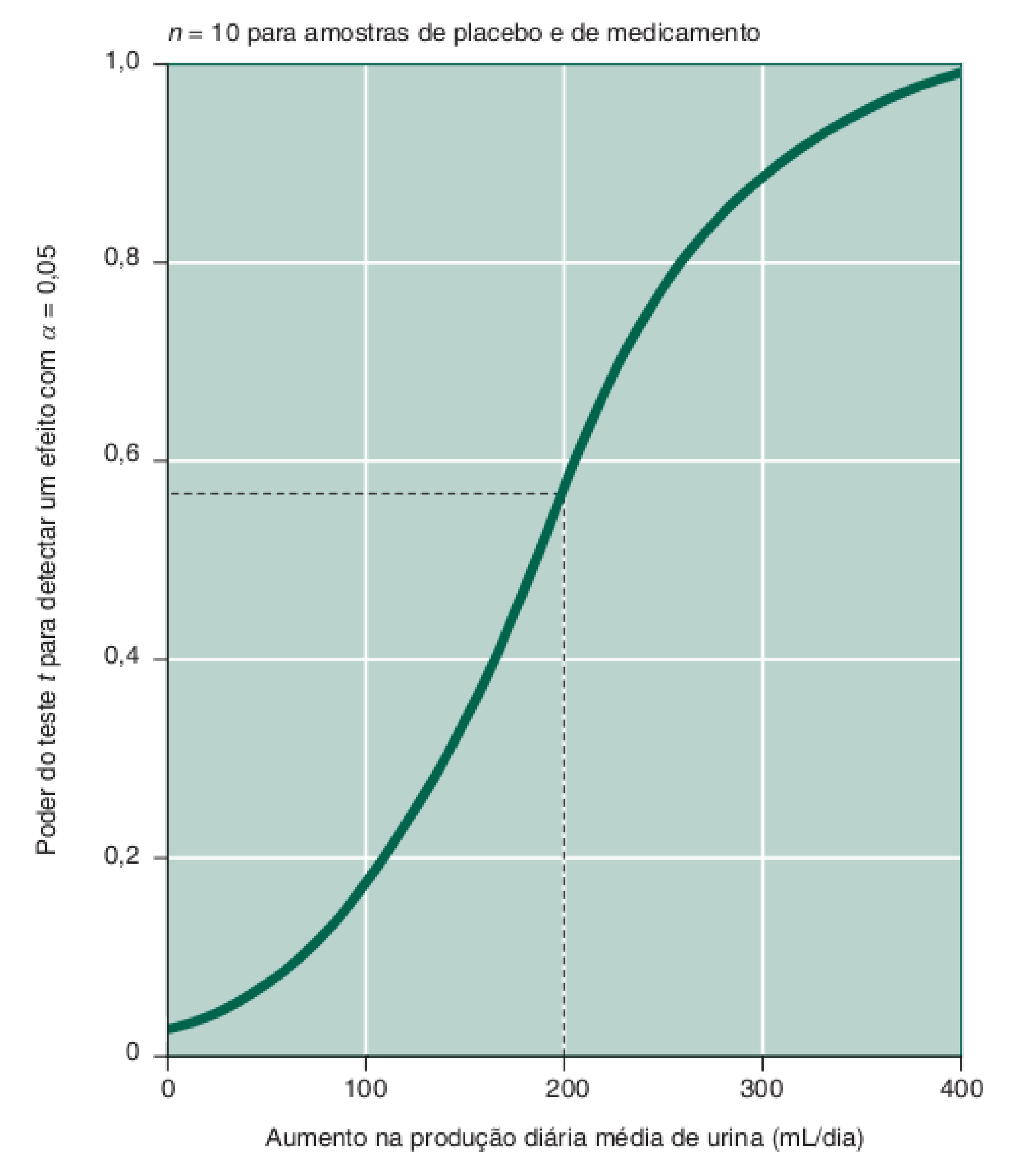
TAMANHO DA AMOSTRA E PODER DO TESTE



- Experimento
- Diurético aumenta volume urinário
- n=10 em cada grupo
- Resultado:
- Diferença de 200ml no vol. urina entre grupos

- **Qual o poder do teste "t" para detectar tal ≠ com este tamanho de amostras ("n")?**

TAMANHO DA DIFERENÇA E PODER DO TESTE



- Experimento
- Diurético X aumenta o volume urinário?

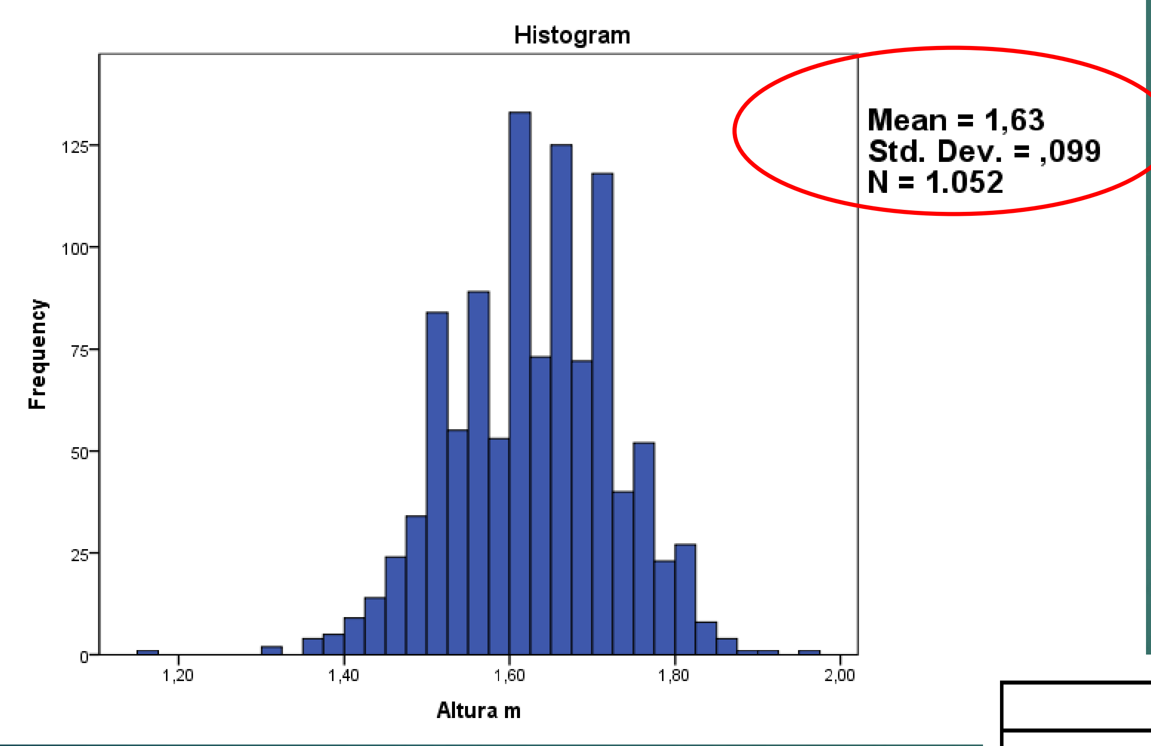
- n=10 em cada grupo
- Resultado:
- Diferença de 200ml no vol. urina entre grupos

- **Considerando este "n" qual o poder do teste "t" para detectar a ≠ observada ?**

Intervalo de Confiança das Médias e da Diferença das Médias

Intervalo de confiança (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

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

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

Intervalo de confiança (IC)

- Condições necessárias para interpretação correta do IC
  - Amostra deve ser aleatoriamente selecionada da população
  - A distribuição da população é “normal”
  - Todos os indivíduos são da mesma população e selecionados de forma independente

Intervalo de confiança da média

A verdadeira média populacional estará dentro do interval de X ± 1,96 erro-padrão em 95% dos casos e de X ± 2,58 erro-padrão em 99% dos casos

$$\text{lower boundary of confidence interval} = \overline{X} - (1.96 \times \text{SE})$$
$$\text{upper boundary of confidence interval} = \overline{X} + (1.96 \times \text{SE})$$

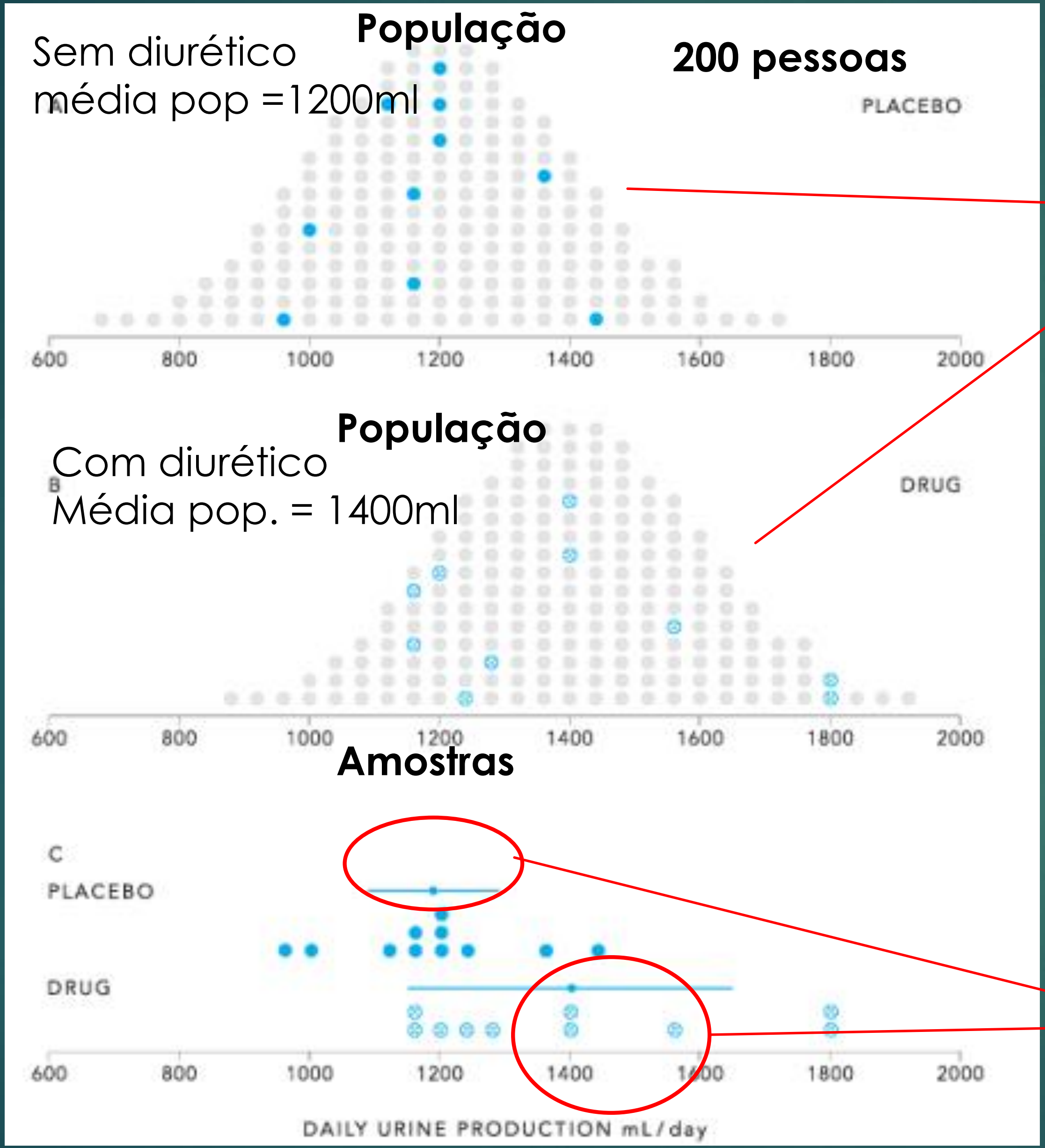
$$\overline{x} \pm z\frac{s}{\sqrt{n}}$$

Assim, IC depende de n, da média e do DP. Portanto, cada amostra fornecerá um IC diferente.

Intervalo de confiança (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, "irreal"**)

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

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

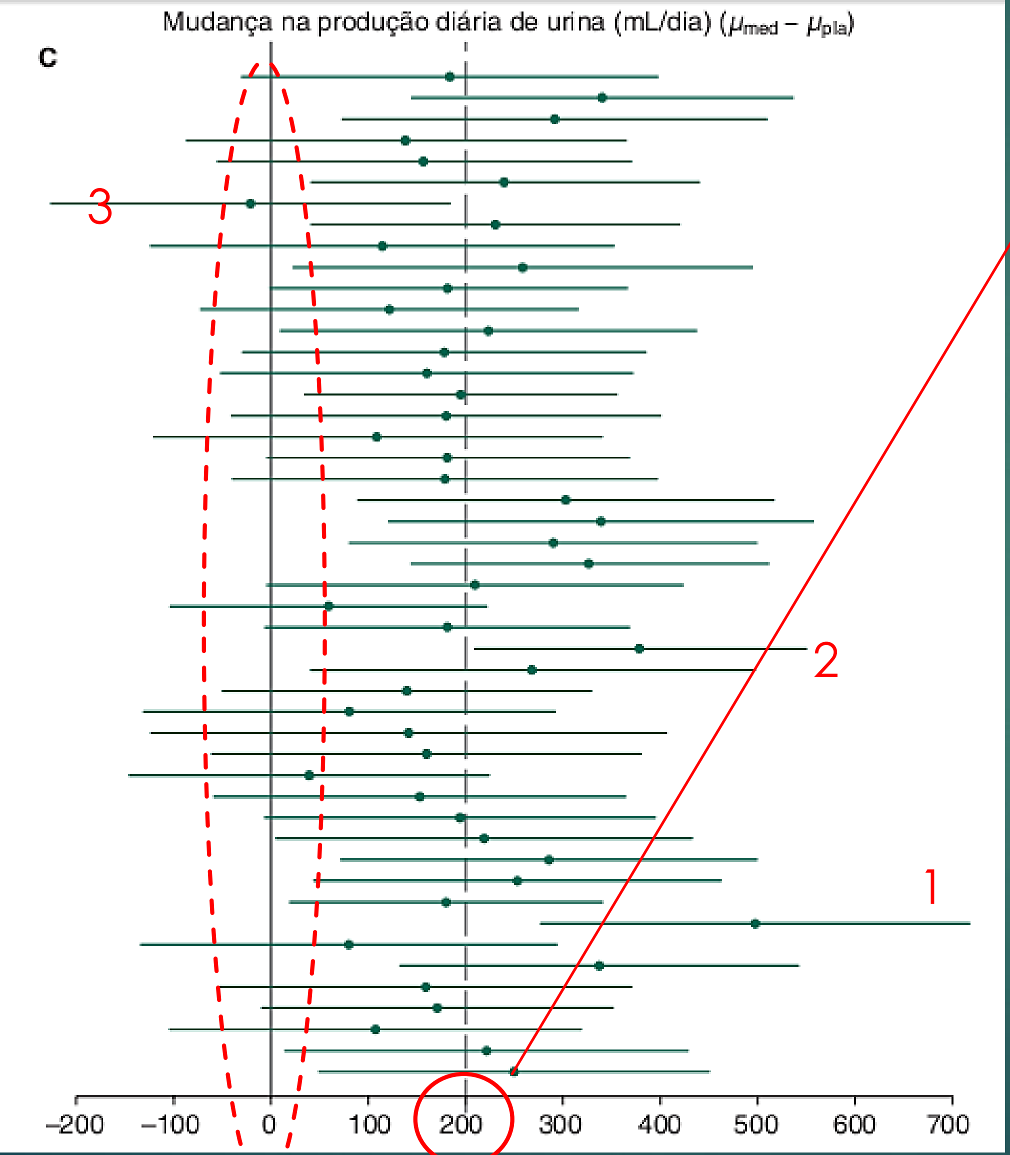
- Para nosso exemplo, utilizando IC da ≠ de médias observadas de 95%:

  31ml/dia < ≠ de médias < 409ml/dia (inclui os 200ml da ≠ observada)

- Podemos estar 95% confiantes de que o diurético aumenta a diurese entre 31 e 409 ml/dia. (não inclui valor “0”, portanto significativo para p<0,05)

Diferentes amostra ⟶ Diferentes IC 95%

Diferença média e IC 95% de 50 amostras



Verdadeira ≠ médias populacional (200ml/d)

Com excessão de 3 amostras (1,2,3), todas as demais incluem o valor 200ml

Das 50 amostras em 22 o IC inclui o valor **"0"(zero)**, portanto **não podemos descartar H0 (amostras são semelhantes)**. Desde que 22 (45% de 95%) incluem 0 a probabilidade de detectar a ≠ é 1-β= 0.55 (55%)

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

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

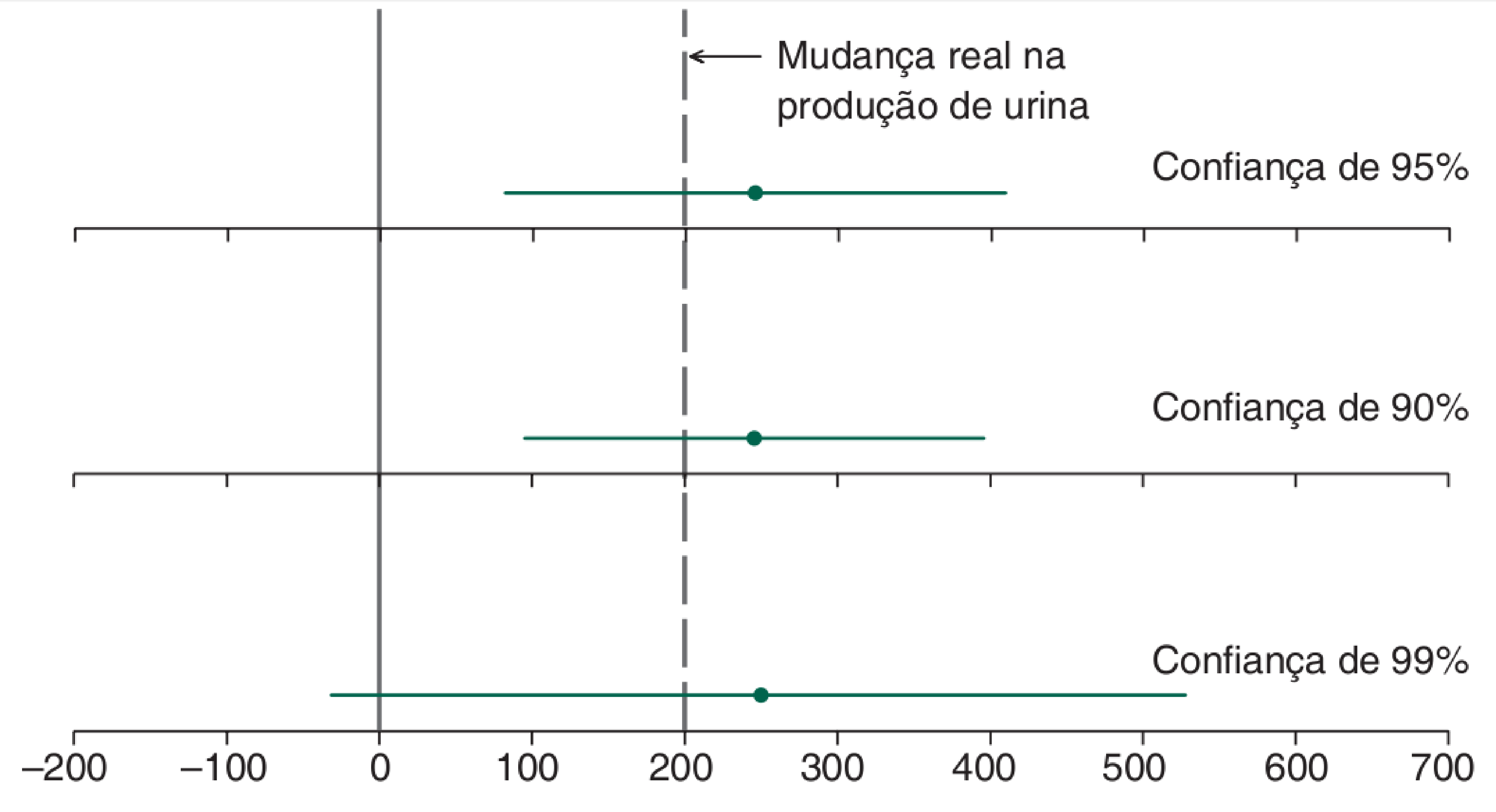
95% confiança de que a medicação diminuia PA entre 1,2 a 6,9mmHg

- **Permite testar hipótese**
  - Se a ≠ nas média amostrais incluiu o valor zero temos 95% confiança de que a ausência de ≠ esta entre os resultados possíveis, portanto a ≠ é não-significativa

Conclusões

- Um determinado intervalo de confiança poderá ou não incluir o valor **Verdadeiro da média**, e a chance disso ocorrer poderá ser de 1%, 5%, 10%, etc. (a), conforme se o intervalo é de 99%, 95% e 90%, e assim por diante
- **Se o IC da diferença entre as médias INCLUI O VALOR ZERO não se pode rejeitar H0 para p < a (efeito é não sifnificativo).**
- **Se o IC NÃO INCLUI o valor zero rejeita-se H0 para p < a. (a=5% se o IC é de 95% e 1% se IC é de 99%)**
- **Além de possibilitar rejeitar ou não H0, IC nos fornece o tamanho do efeito**
- Portanto, se o resultados da ≠ alcança significância estatística (p<0,05) **mais devido ao tamanho da amostra** do que devido ao tamanho do efeito (tamanho ≠), o IC nos mostrará isso.

Aumentando a o nível de confiança para se ter maior chance de englobar o efeito real (de p<0,05 para p<0,01, p.e), mais amplo será o intervalo de confiança.



- Embora calculado de **forma um pouco diferente as conclusões acerca do IC 95% ou 99% também são válidadas para proporções e comparações entre elas.**