

Estatística: Ciência encarregada de estudar o problema da coleta, análise e interpretação de observações

Porquê existe a estatística?

Porque existe variação (biológica)

# “Boa estatística é o senso comum refinado”

*We should go through the rigamarole (non-sense) of running statistical tests and figuring our probabilities only if we know very little about the problem we are dealing with (Fisher)*

*An effect that exists at  $p=.05$  doesn't suddenly disappear at  $p=.051$  (Norman and Streiner, 2008)*

*“Surely, God loves the .06 nearly as much as the .05” (Rosnow and Rosenthal, 1989)*

# **Estatística:**

- 1. descritiva:** relacionada à apresentação, organização e sumarização dos dados
- 2. Inferencial:** permite generalizar a partir de nossa amostra de dados para um grande grupo de sujeitos

# Tipos de variáveis (variável: aquilo que está observado ou medido)

1. Discretas: **nominal** (ex.: gênero)

**ordinal** (ex.: escala de avaliação de dor: ausente, leve, moderada, extrema. As diferenças entre as categorias não podem ser consideradas iguais)

2. Contínuas: **Intervalar**: distâncias iguais entre os valores mas o zero é arbitrário (Ex. QI )

**Razão**: distâncias iguais entre os valores mas o zero não é arbitrário. (Ex. Peso )



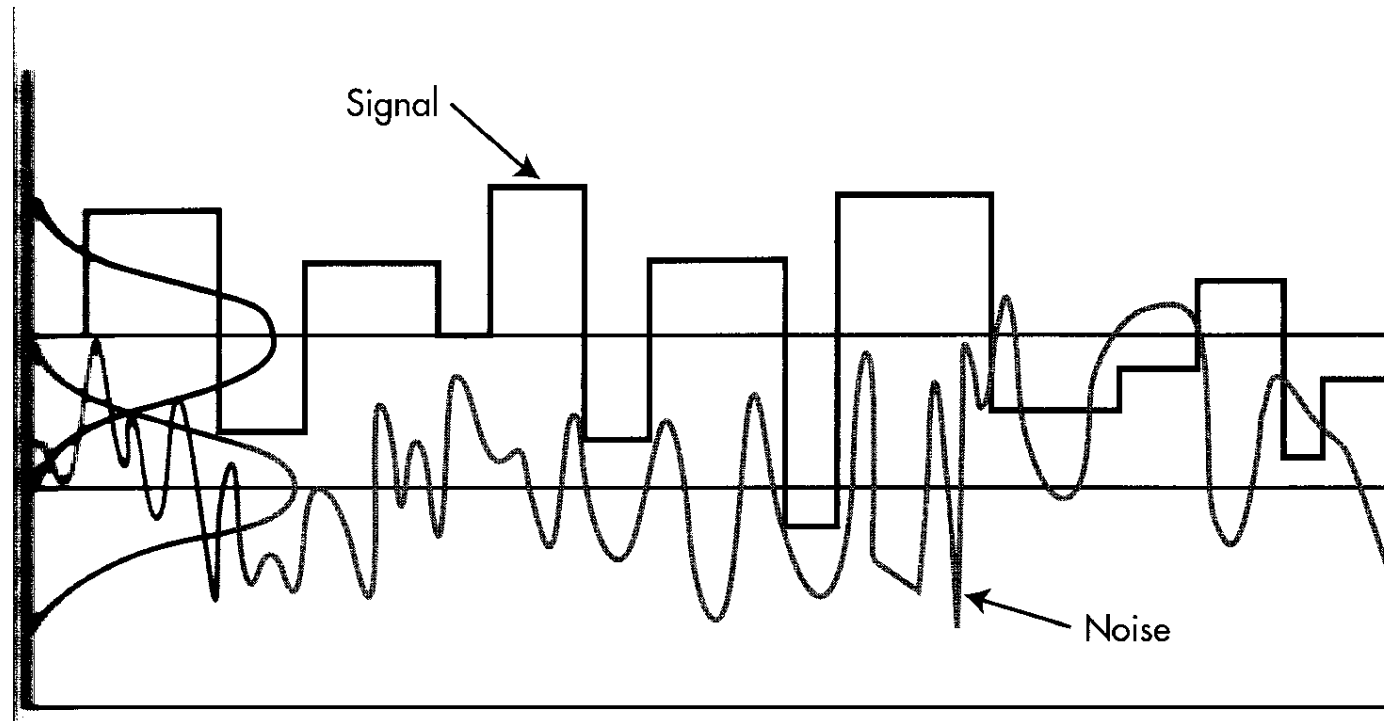
# **Tipos de variáveis** (variável: aquilo que está observado ou medido)

1. Independentes (aquelas sobre o controle do experimentador)

2. Dependentes (aquelas que respondem ao controle do experimentador)

(3. “confounding variables”)

# Princípio Básico da estatística inferencial: relação sinal-ruído



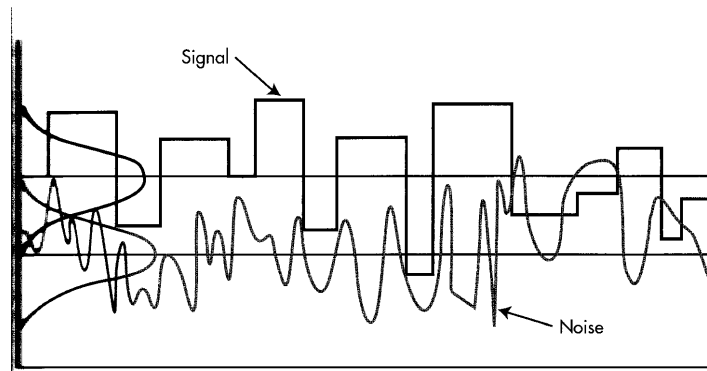
Antes de calcular qualquer estatística é melhor fazer o gráfico!

(ou seja, sempre **por no gráfico**)

O quê mostrar:

1.Todos os valores

2.Medidas centrais e de dispersão dos dados: quais?



# Medidas centrais

Moda

Mediana

Média

Outras: média geométrica ( $\sqrt[n]{x_1 \cdot x_2 \dots x_n}$ ) e média harmônica ( $n / \sum 1/x_i$ )

## Medidas de dispersão

Índice de dispersão: para dados de categoria: de zero (todos os dados em uma categoria) até 1 (distribuição igual entre as várias categorias)

Range (amplitude)

Intervalo (amplitude) interquartil

Variância:  $S^2 = \sum (x_i - \text{média})^2 / N$

Desvio Padrão:  $S = \sqrt{S^2 / N}$

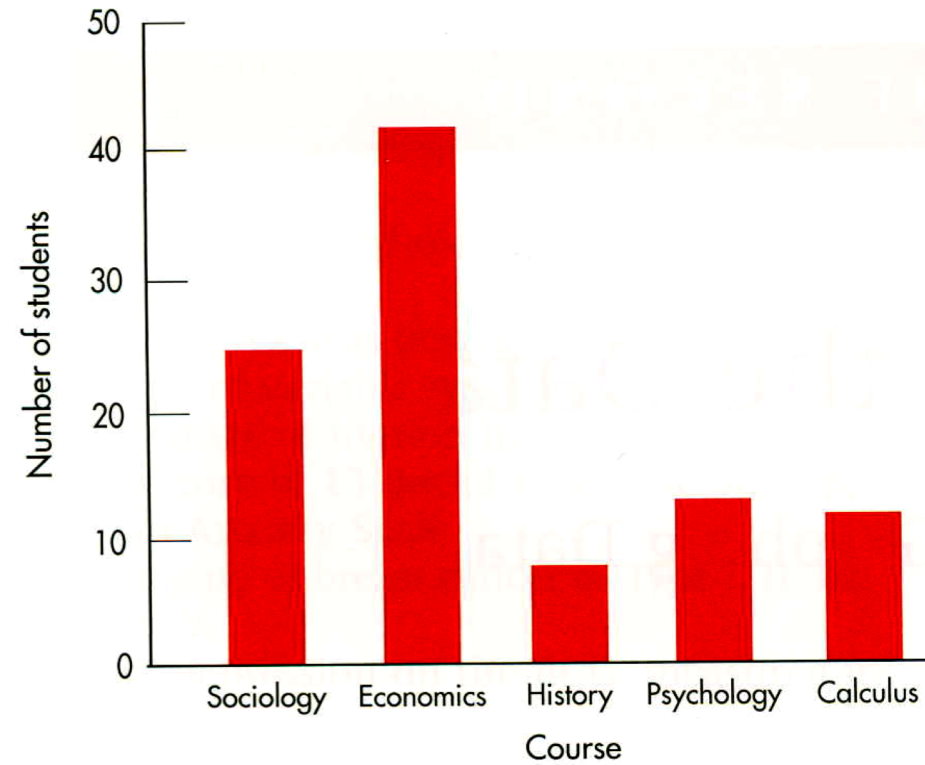
Erro padrão:  $EP = S / \sqrt{n}$

Com variáveis contínuas, podemos ter também 2 outras medidas: skewness (relacionado à simetria da curva) e kurtosis (relacionado ao achatamento da curva)

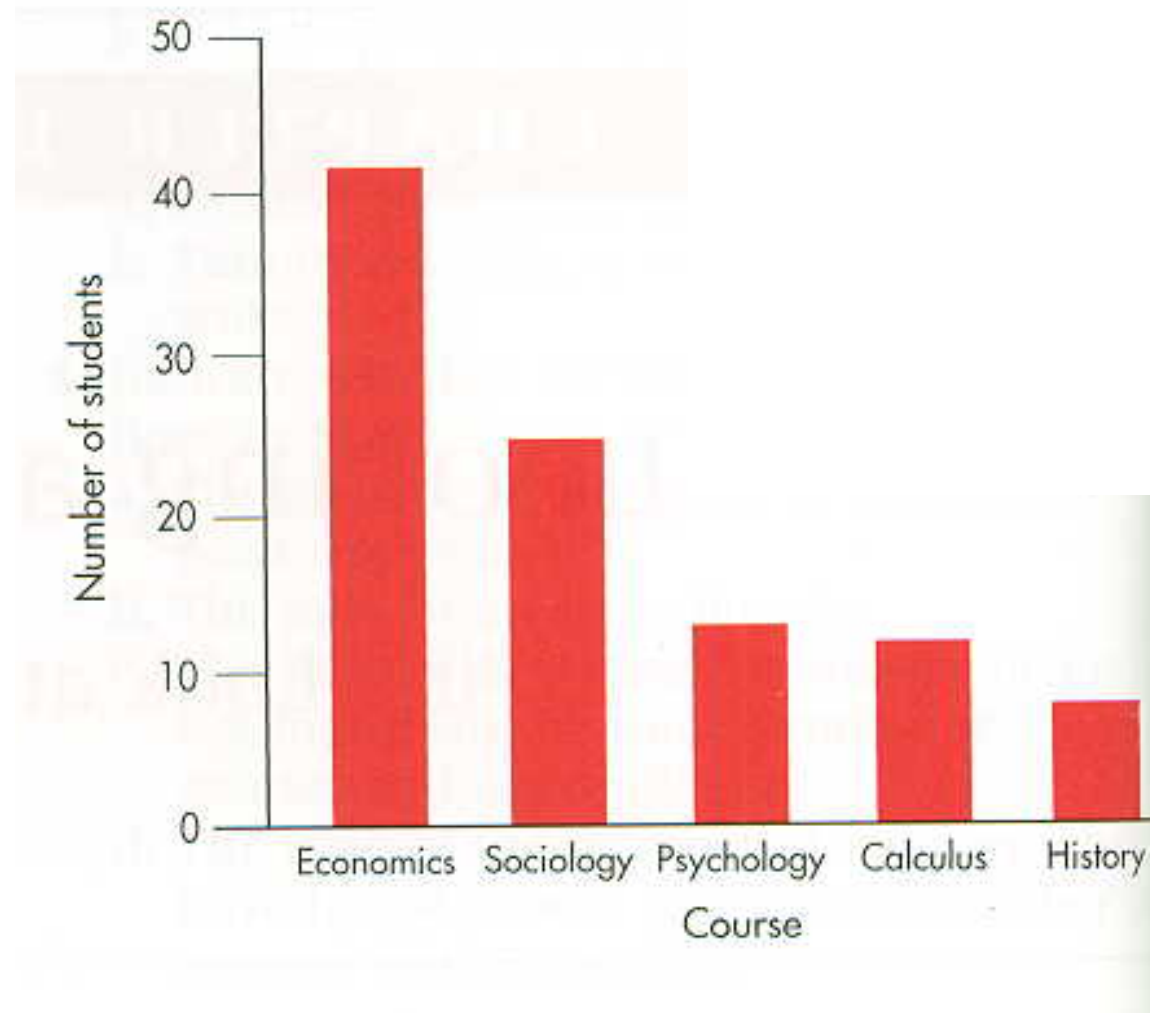
# Exemplos de gráficos: não tão correto

**FIGURE 2-1**

Bar chart of the  
five least popular  
courses



## Exemplos de gráficos: correto

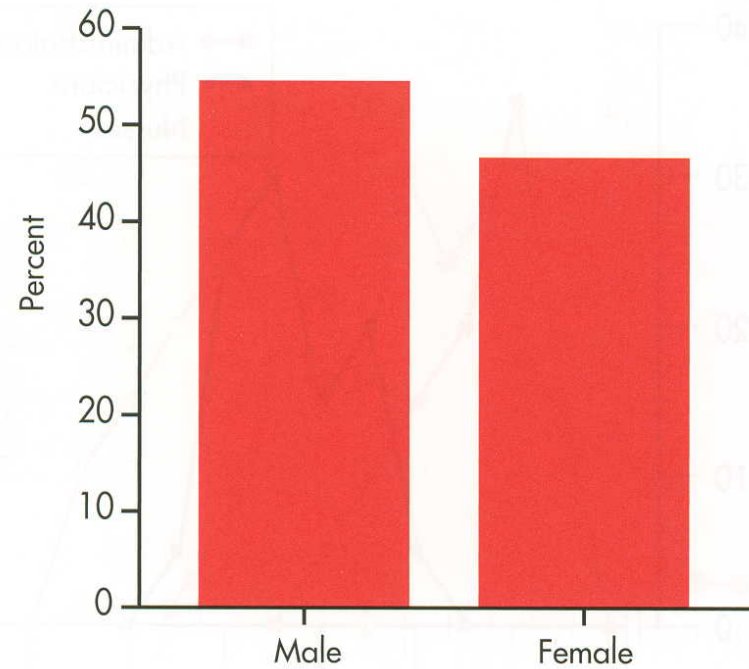




# Exemplos de que não fazer em gráficos

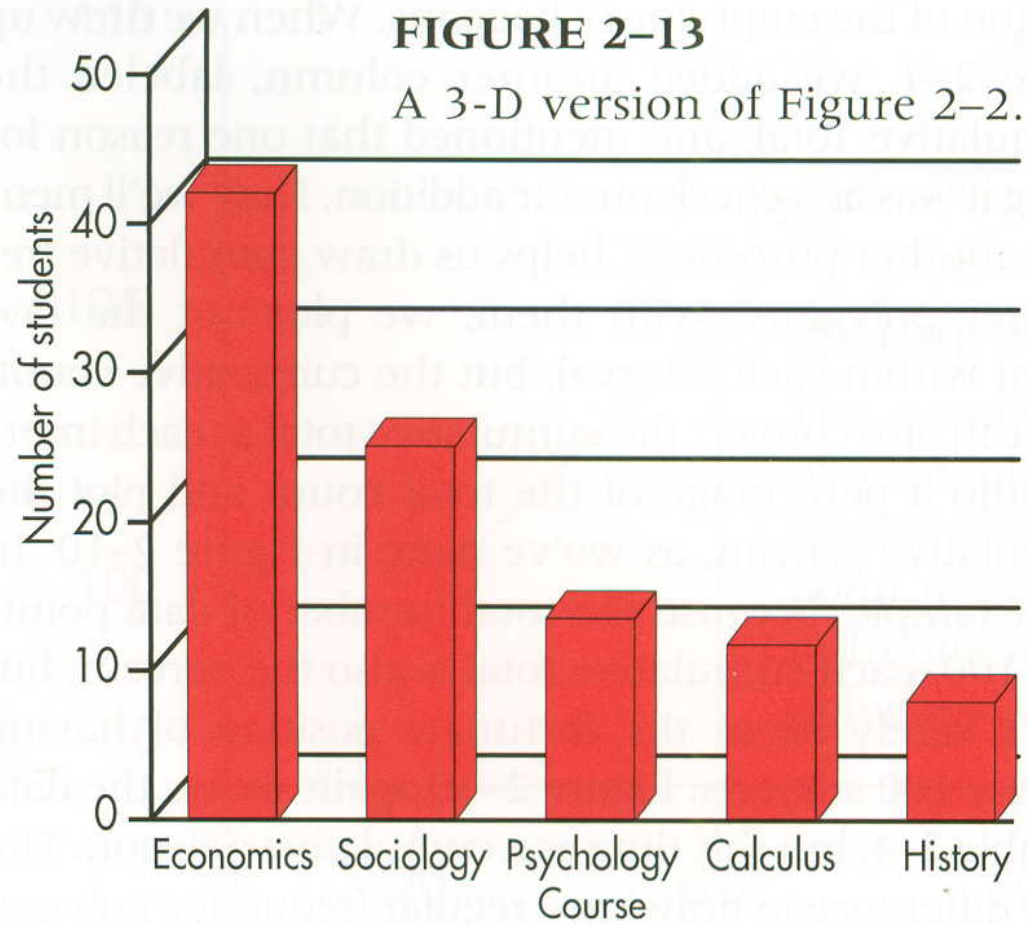
**FIGURE 2-11**

The proportion of males and females in a study.

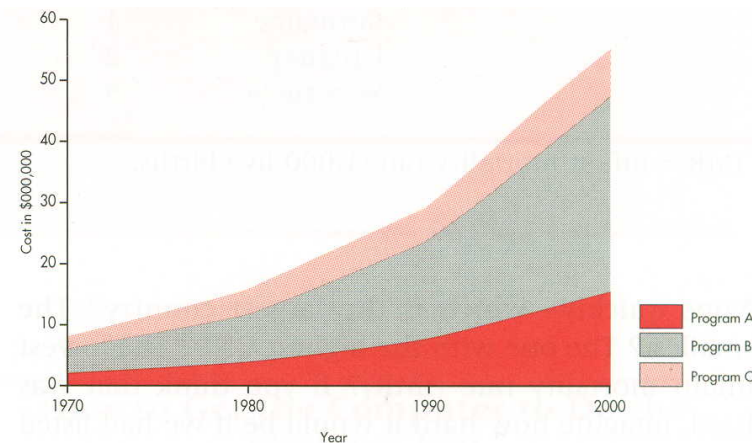
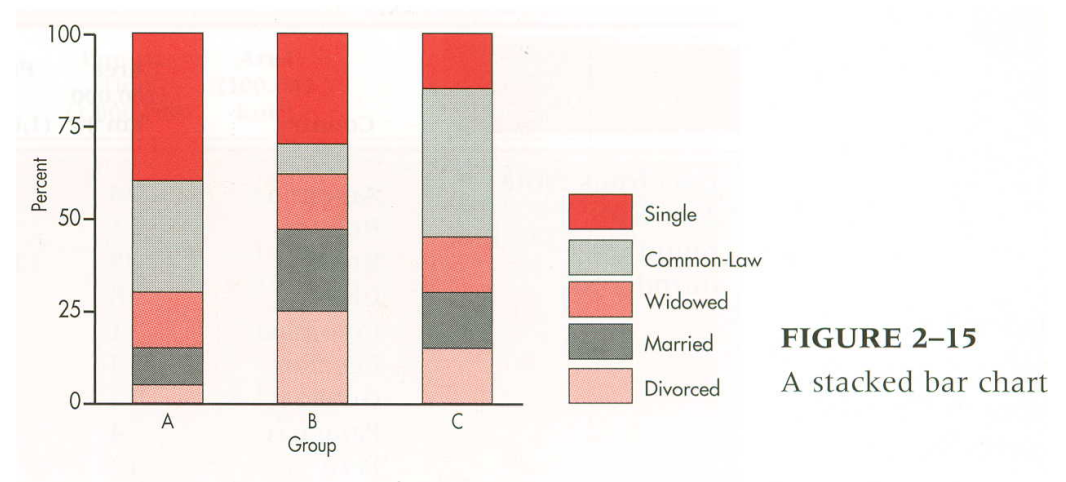
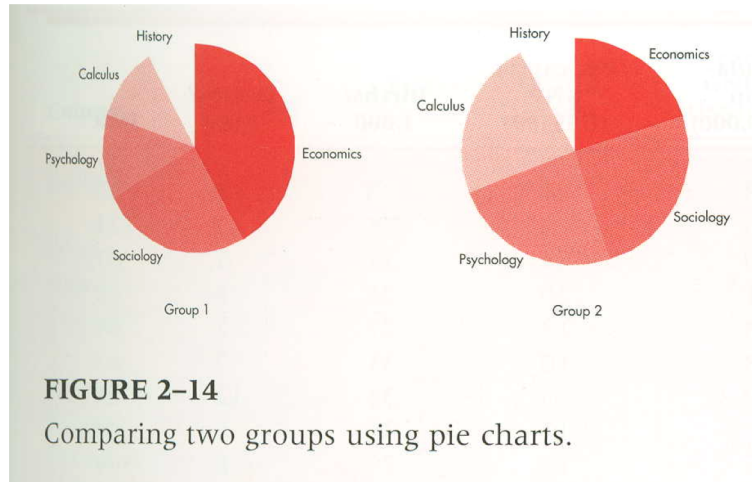


Gráficos desnecessários

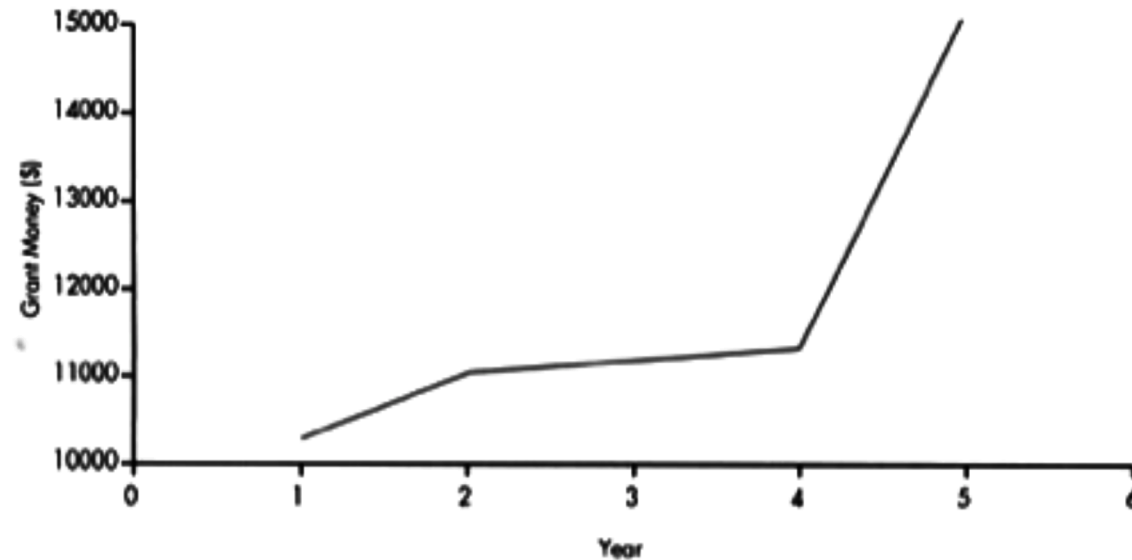
Exemplos de que não fazer em gráficos:  
efeitos especiais 3D



## Exemplos de que não fazer em gráficos: outros efeitos especiais

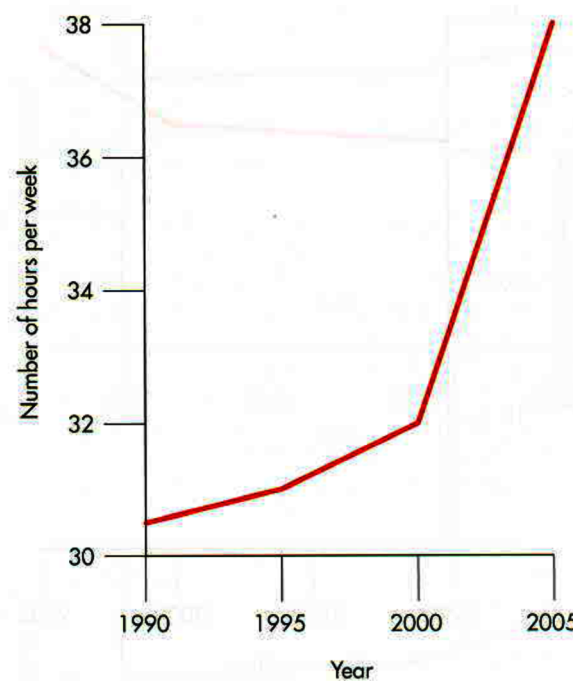


**FIGURE 2-12**  
Grant money per  
year for Dr. X.

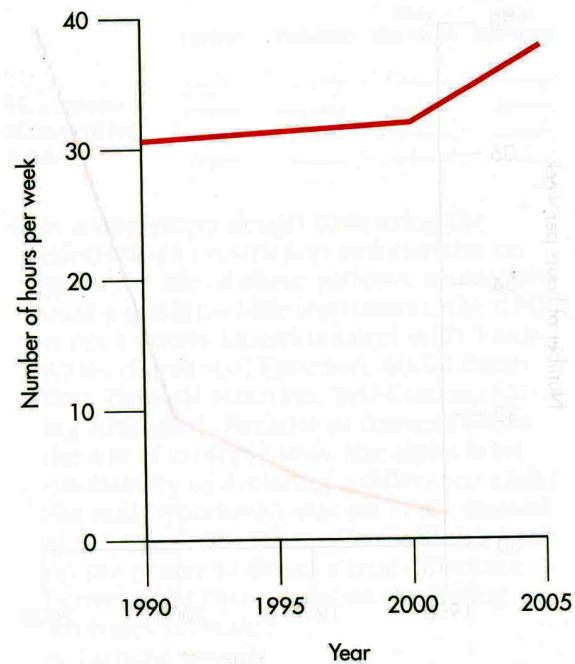


## Problemas da gráfico?

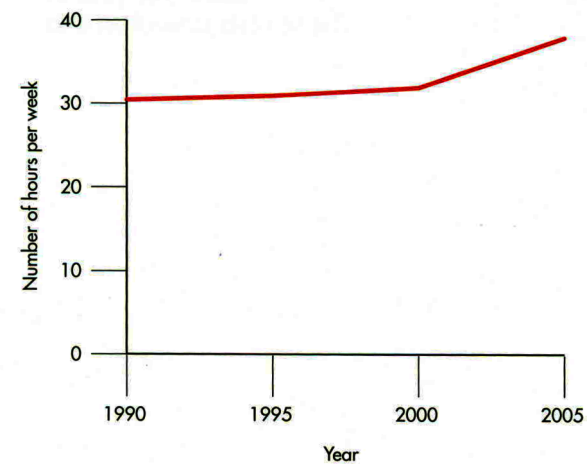
1. Não começa do zero (efeito de aumento parece muito maior)
2. Eixos x e y muito diferentes



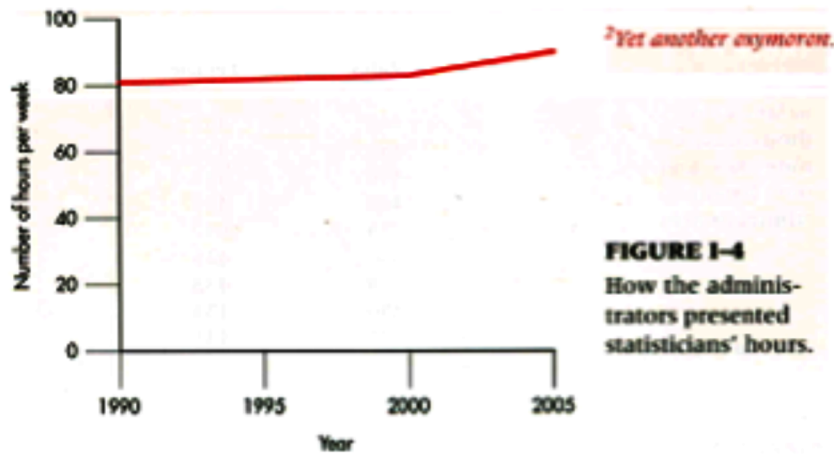
Original



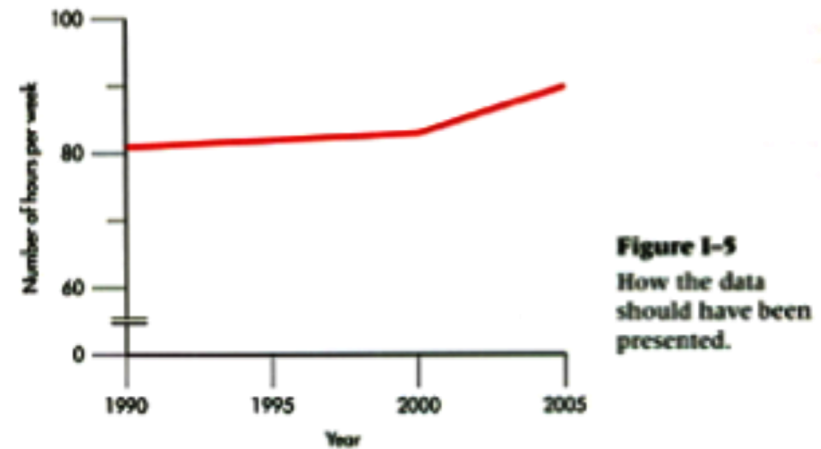
Corrigido (eixo y)



Como deveria ser  
(eixos x e y)

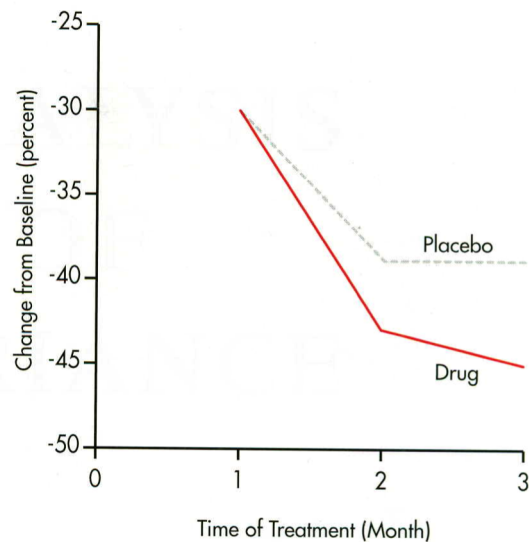


Apresentado



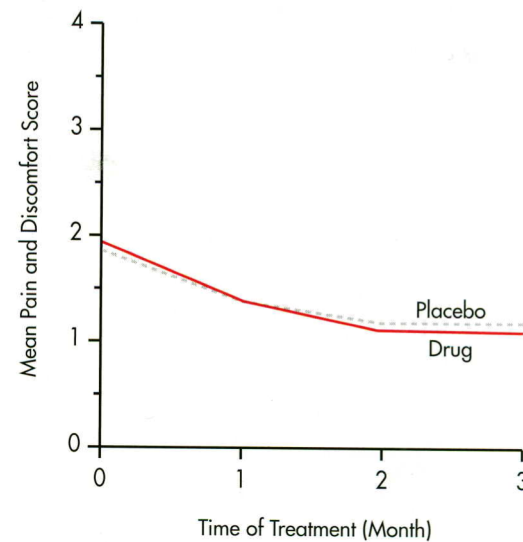
Como deveria ser

Cuidado para não deixar muito “espaço morto”



**FIGURE I-6**  
Improvement as  
reported by a drug  
company.

Apresentado



**FIGURE I-7**  
What actually  
happened.

Como deveria ser

Cuidado delta e percentagem

## Regras de um bom gráfico (Simplificar!)

1. Sempre tentar maximizar a chance do leitor ver o(s) ponto(s) que gostaria de enfatizar
2. Usar a escala adequada (começar do zero quando possível)
3. Sempre colocar claramente nos eixos as unidades que estão sendo usadas
4. Eixos de proporções semelhantes



## Regras de um bom gráfico (Simplificar!)

5. Sempre colocar uma legenda embaixo da figura
6. Representar os dados em gráficos ou tabelas (usualmente uma boa figura é melhor que uma tabela)
7. A figura, com a legenda, deve ser a mais auto-explicativa possível
8. Símbolos claros e suficientemente grandes e em número limitado

**Experimento com labirinto em cruz elevado: droga  
1: veículo, droga 2: diazepam 1 mg/kg, droga 3:  
diazepam 2 mg/kg,**

<b>Droga</b>	<b>% entradas aberto</b>
1,00	15,00
1,00	12,00
1,00	30,00
1,00	25,00
1,00	31,00
1,00	10,00
2,00	20,00
2,00	41,00
2,00	38,00
2,00	36,00
2,00	37,00
2,00	26,00
3,00	44,00
3,00	30,00
3,00	41,00
3,00	29,00
3,00	38,00
3,00	35,00

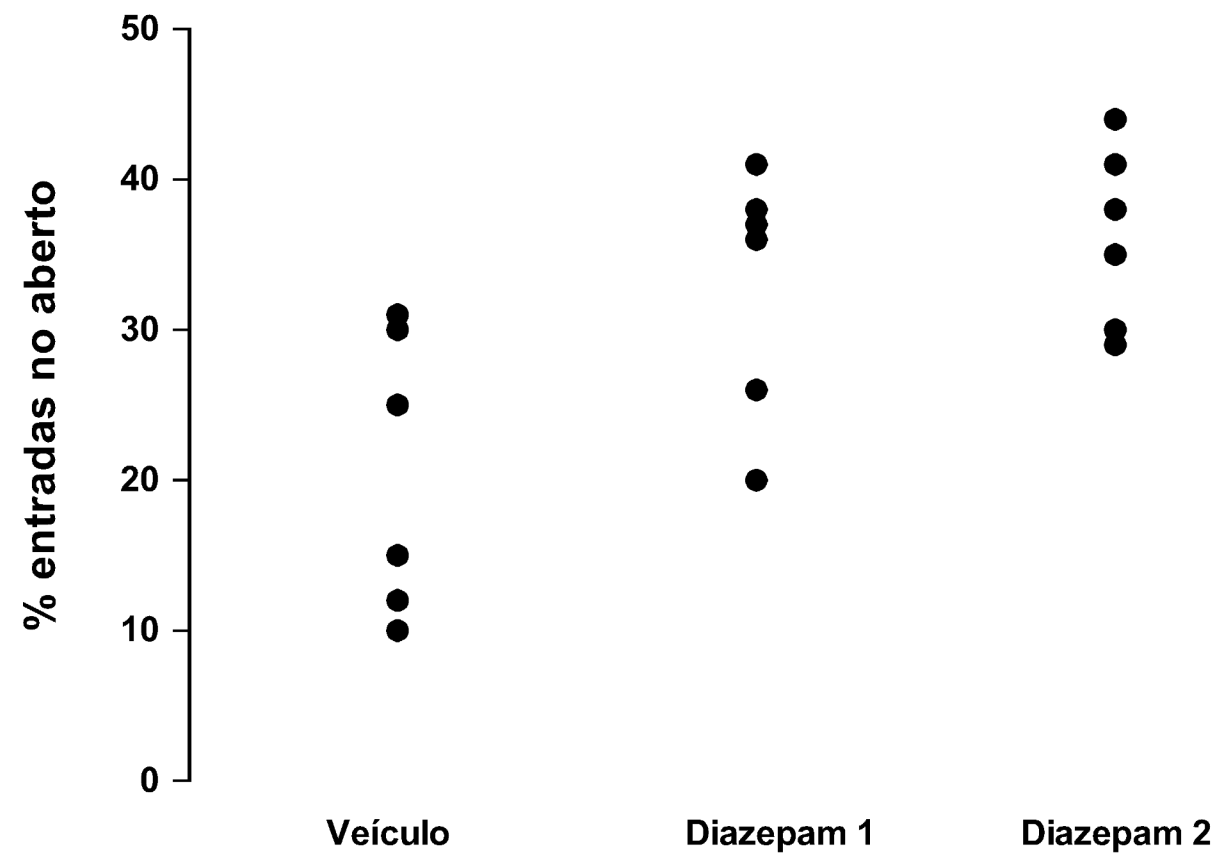


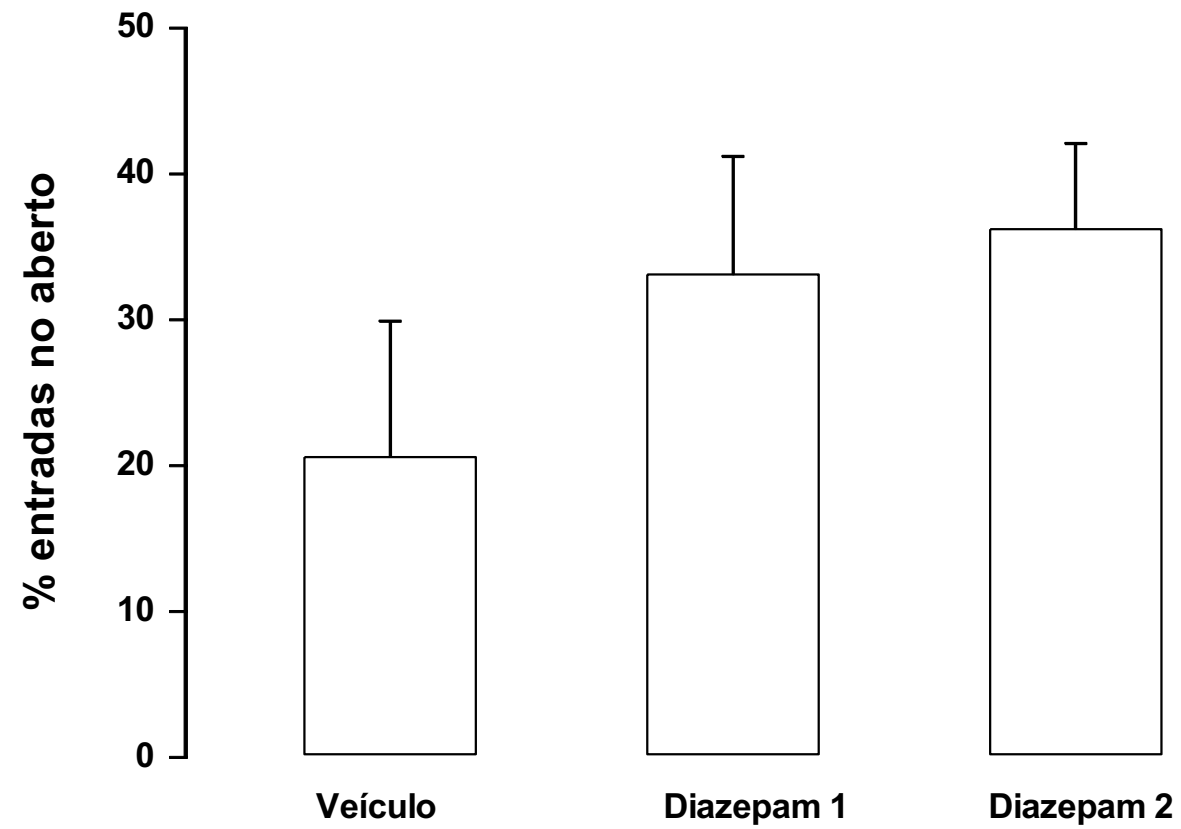
Droga	Média	EPM	n
Veículo	20,5	3,8	6
Diazepam 1 mg/kg	33	3,3	6
Diazepam 2 mg/kg	36,17	2,4	6

Resultados estatísticos:

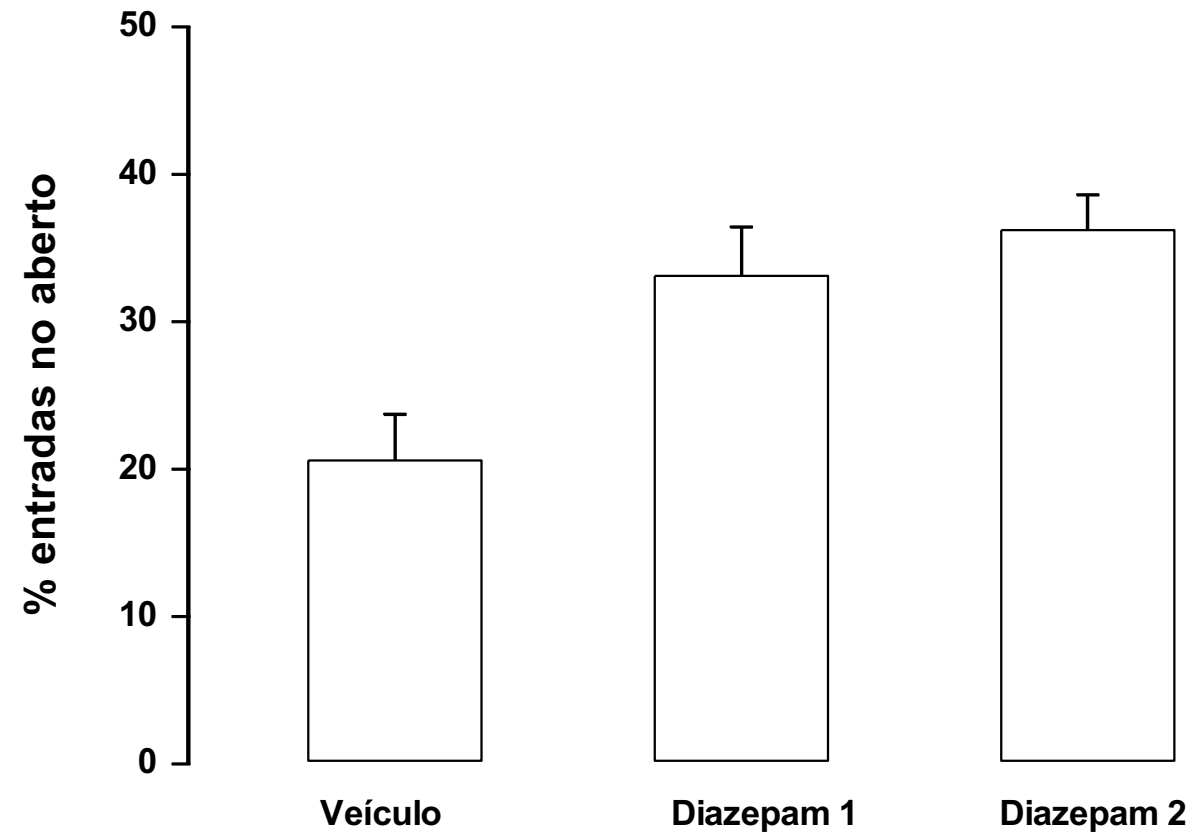
ANOVA  $F_{2,15} = 6,54$ ,  $p = 0,009$

Post-hoc de Duncan: veículo diferente dos demais grupos,  $p < 0,05$

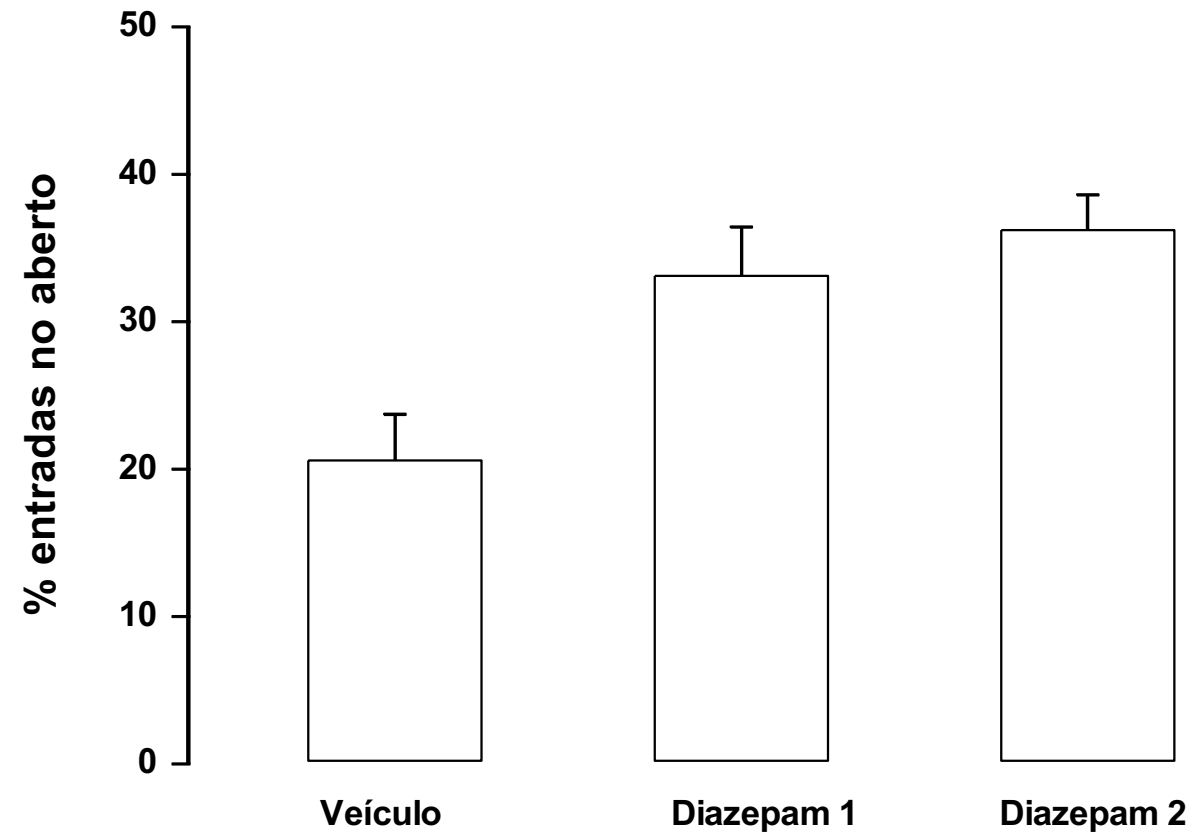


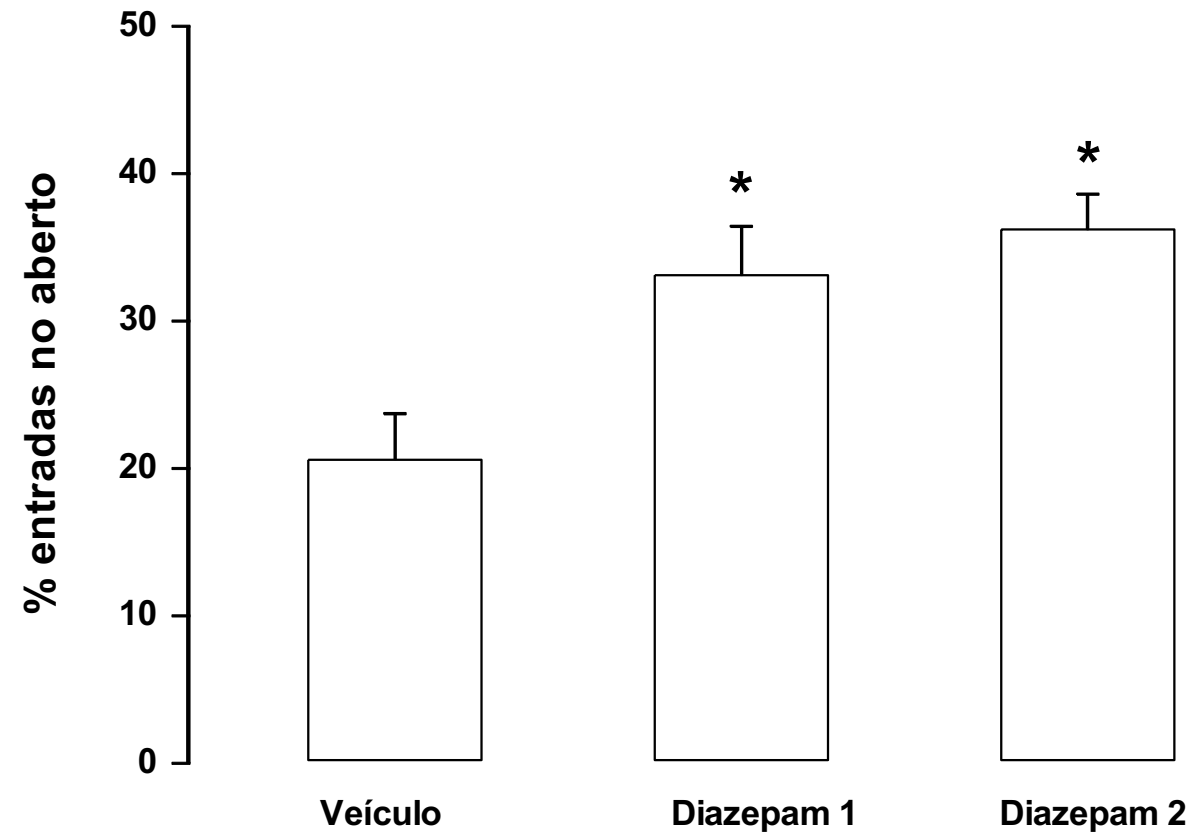


Média + DPM, n=6/grupo



Médias + EPM, n=6/grupo





O que falta?



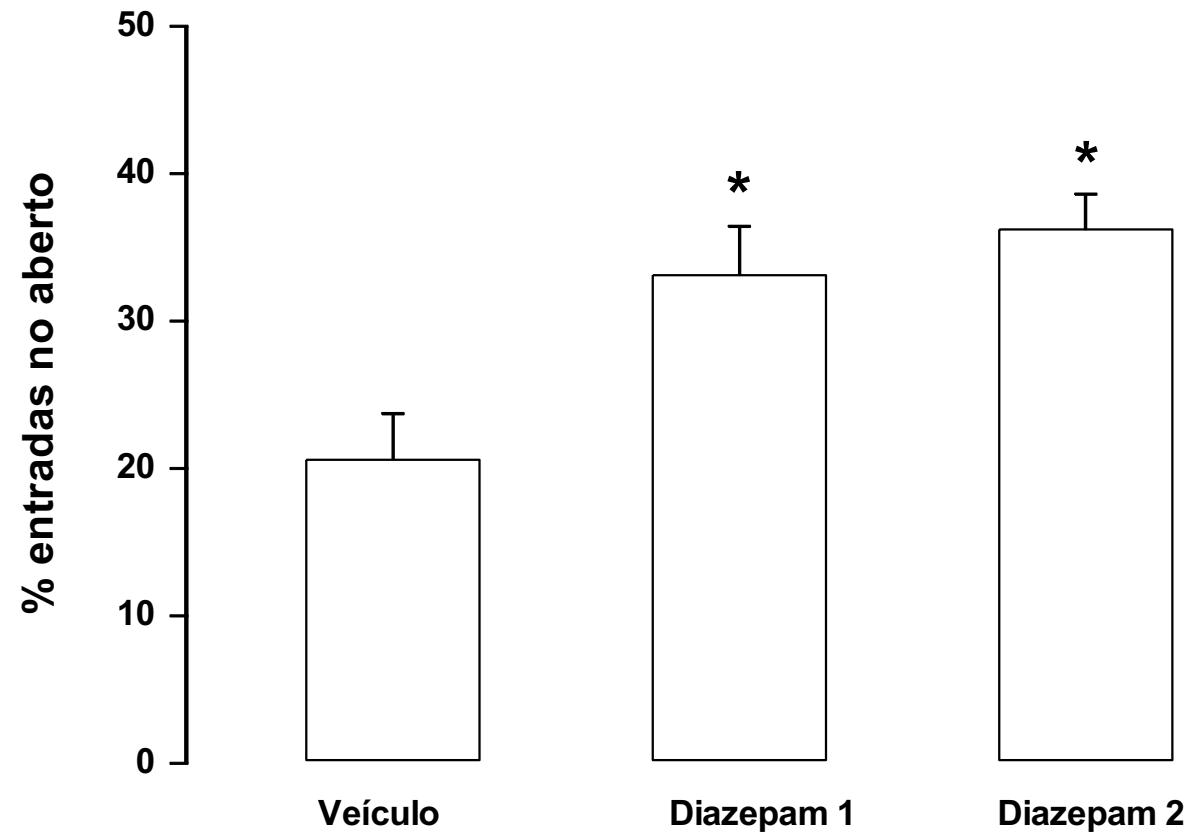


Figura 1, Efeito do diazepam (1 ou 2 mg/kg) na percentagem de entradas de ratos (n=6/grupo) submetidos ao labirinto em cruz elevado, Os dados estão expressos como médias + EPM, Os asteriscos indicam diferença significativa em relação ao grupo veículo (ANOVA seguida de teste de Duncan,  $p < 0,05$ )

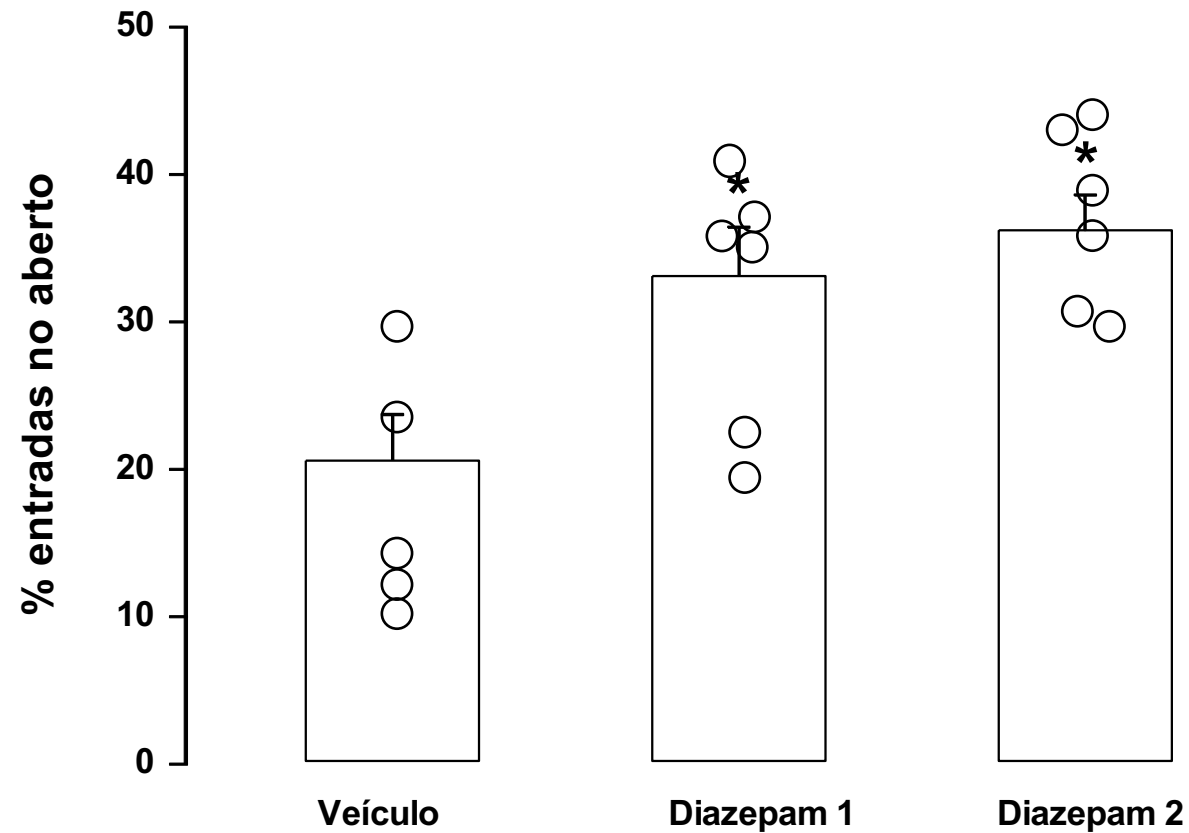
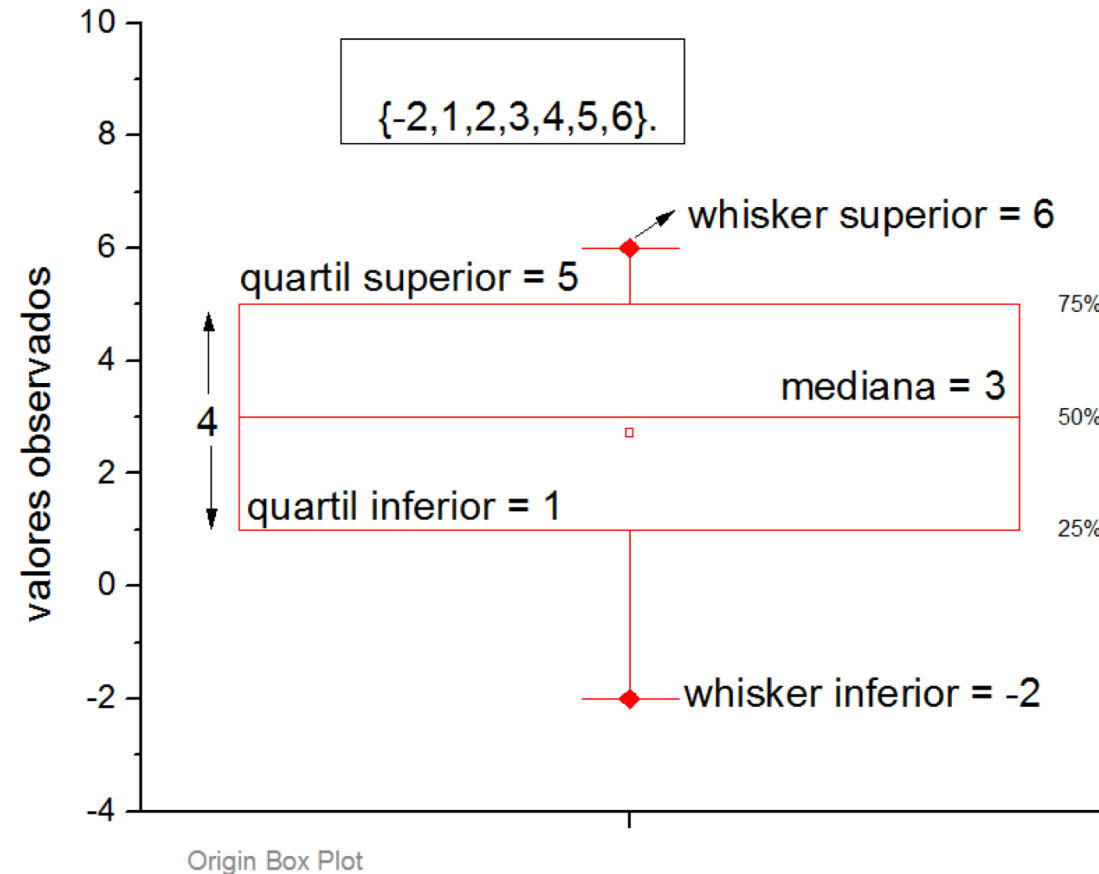
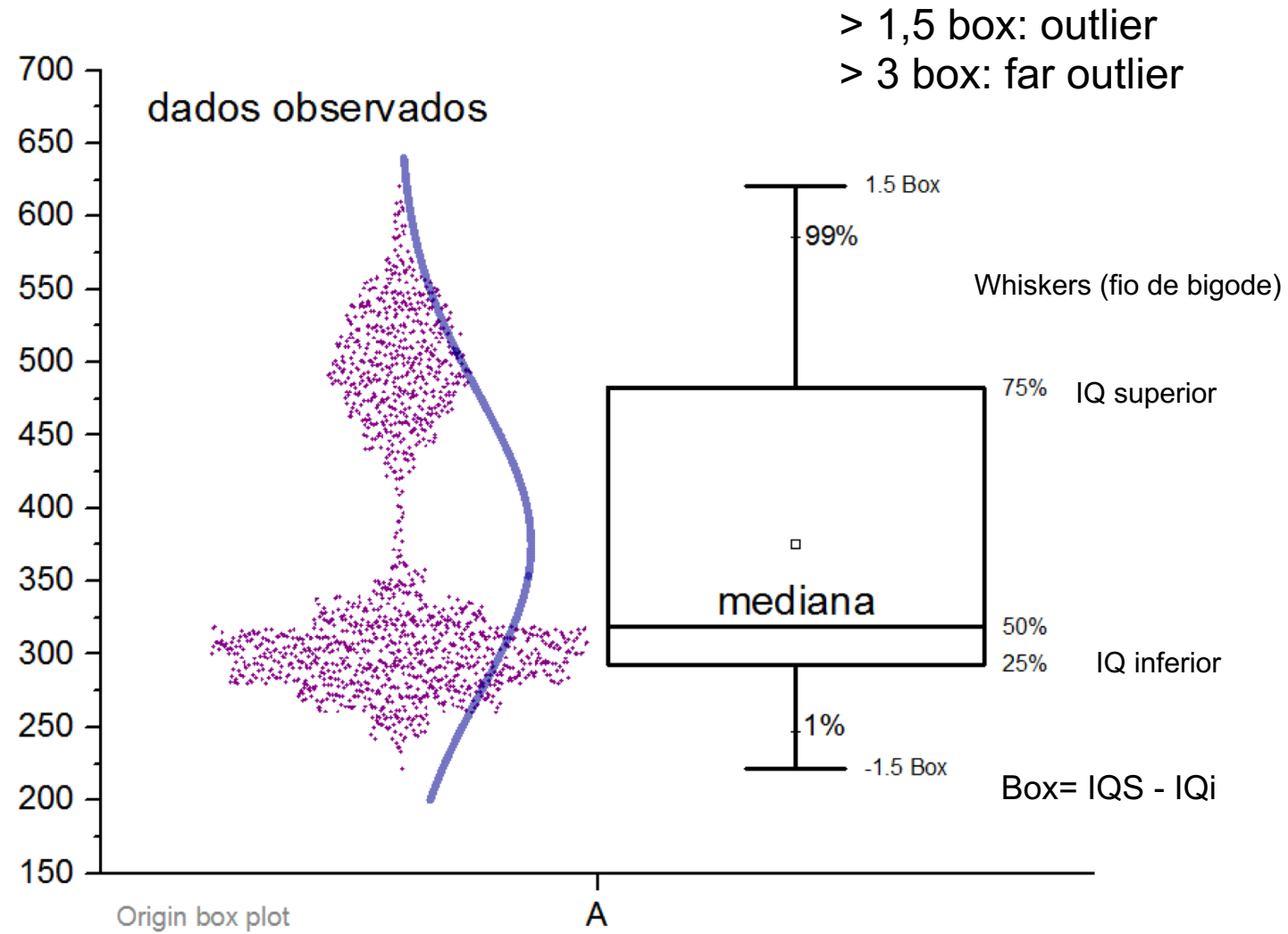


Figura 1, Efeito do diazepam (1 ou 2 mg/kg) na percentagem de entradas de ratos (n=6/grupo) submetidos ao labirinto em cruz elevado, Os dados estão expressos como médias + EPM, Os asteriscos indicam diferença significativa em relação ao grupo veículo (ANOVA seguida de teste de Duncan,  $p < 0,05$ )

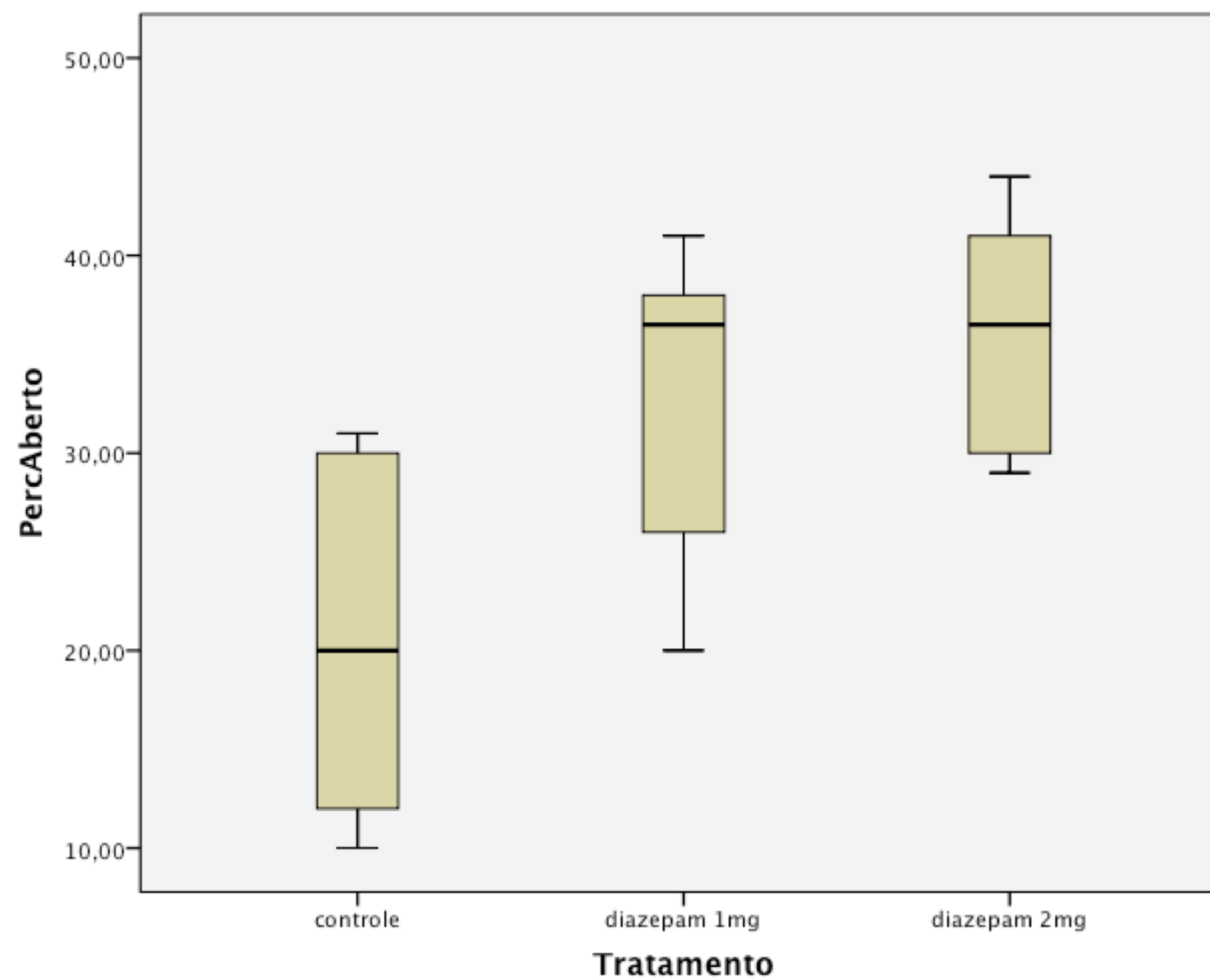
Diagrama de caixa (box-plot): cálculos dos Whiskers: maior ou menor valor que ficar até 1,5 vezes o intervalo interquartil (acima ou abaixo: outliers)



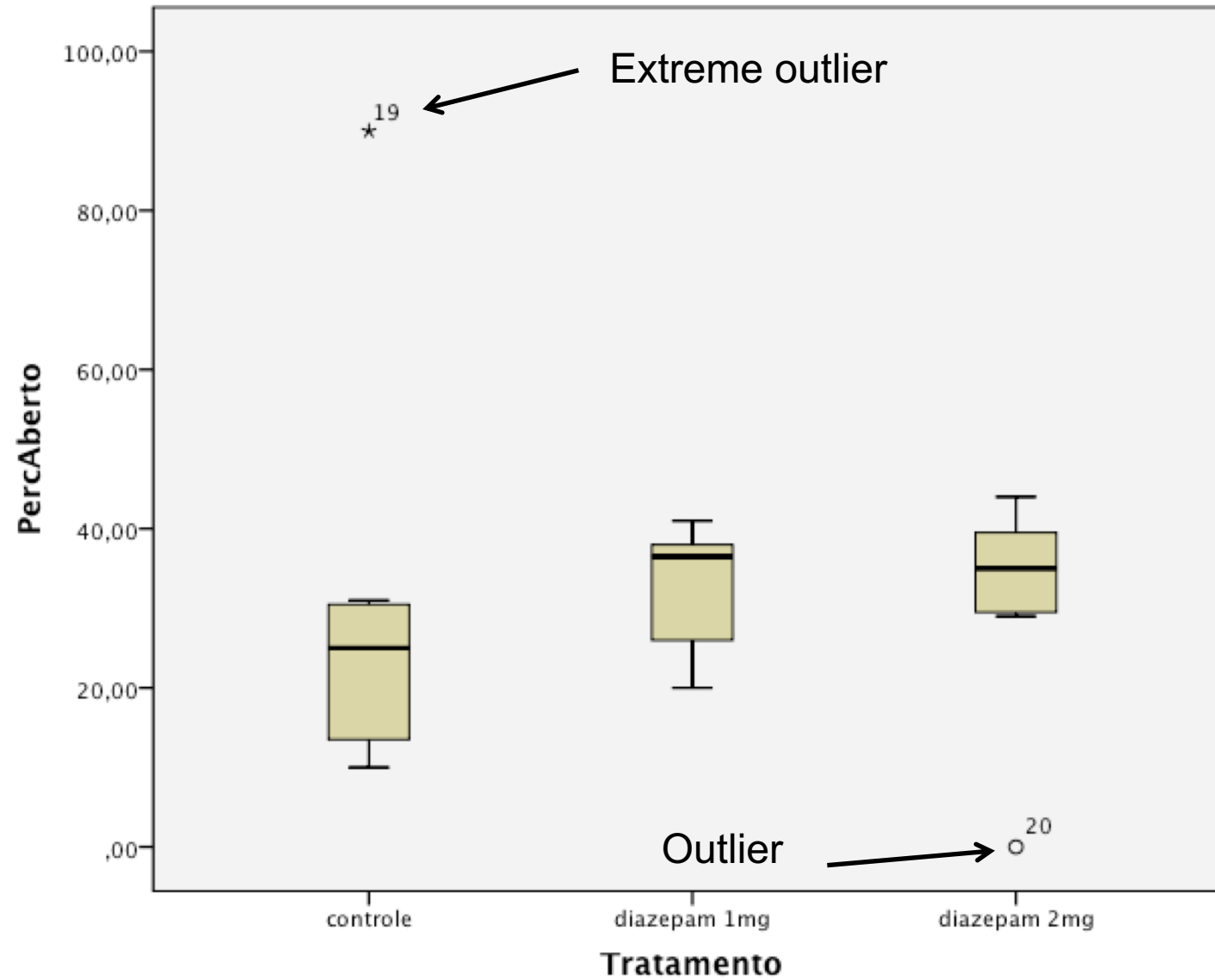
## Diagrama de caixa (box-plot)



Box plot experimento com diazepam no LCE ( $F_{2,5}=6,5$ ,  $p=0,009$ )



Box plot com outliers (ANOVA  $F_{2,17}=0,32$ , NS)



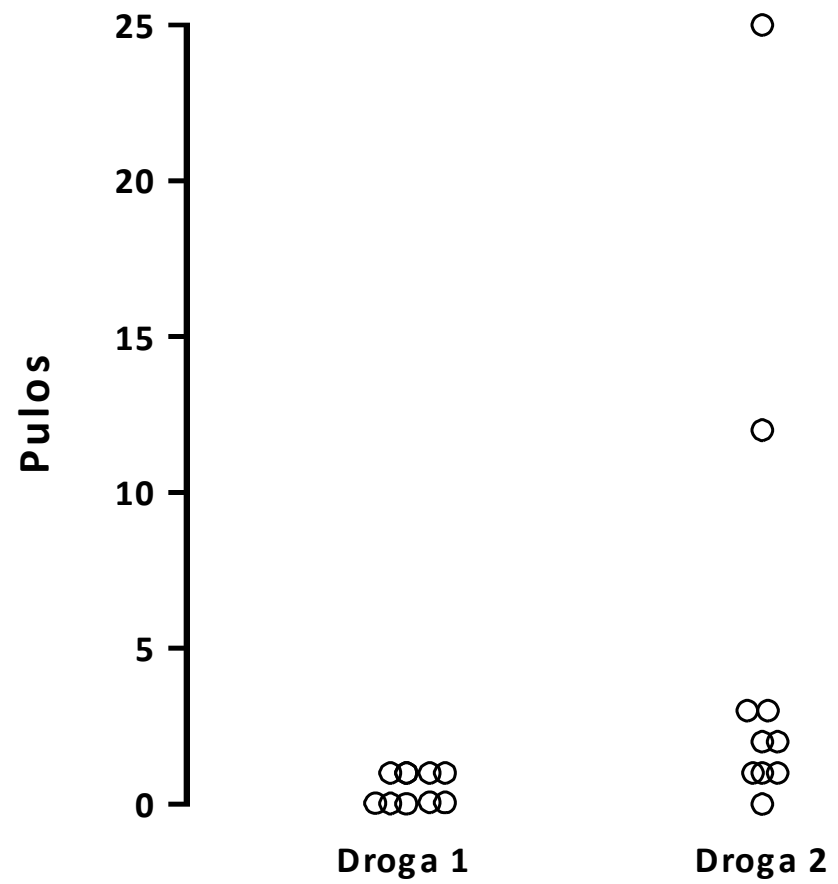
Exemplo 2: número de pulos induzido por NMDA injetado na SCPdl

Droga

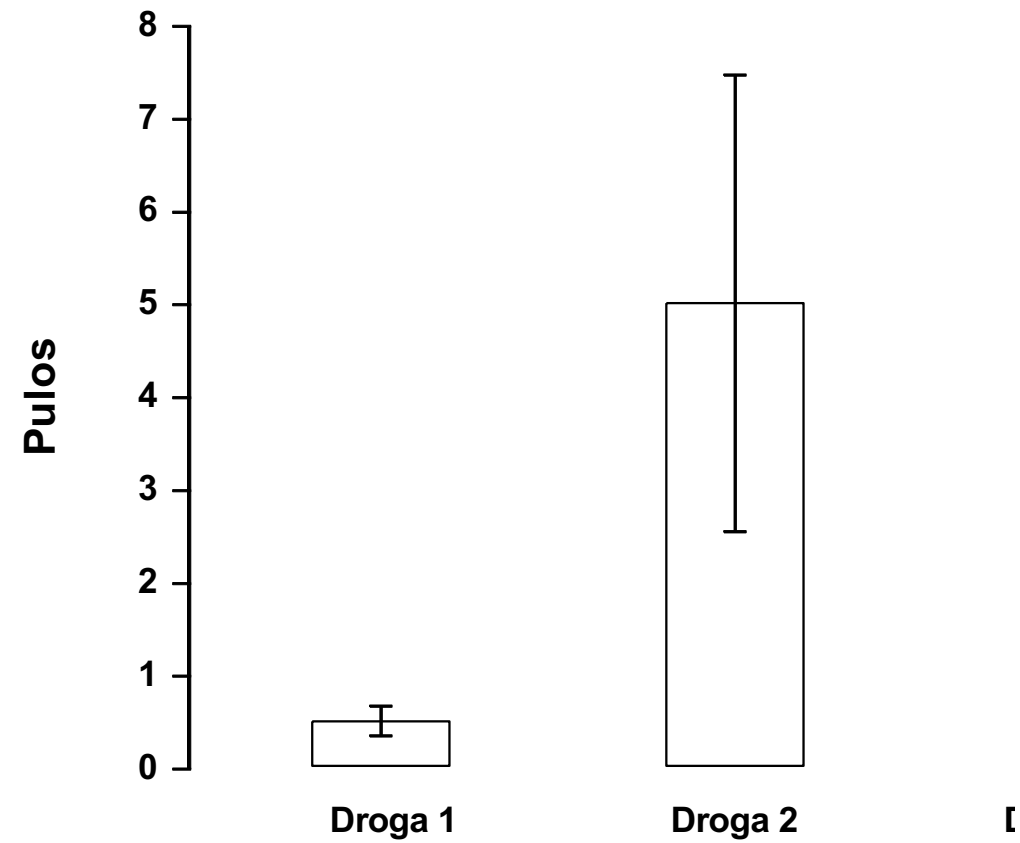
Pulos

1,00	,00
1,00	,00
1,00	,00
1,00	1,00
1,00	1,00
1,00	,00
1,00	,00
1,00	1,00
1,00	1,00
1,00	1,00
2,00	1,00
2,00	1,00
2,00	3,00
2,00	12,00
2,00	2,00
2,00	3,00
2,00	2,00
2,00	25,00
2,00	,00
2,00	1,00

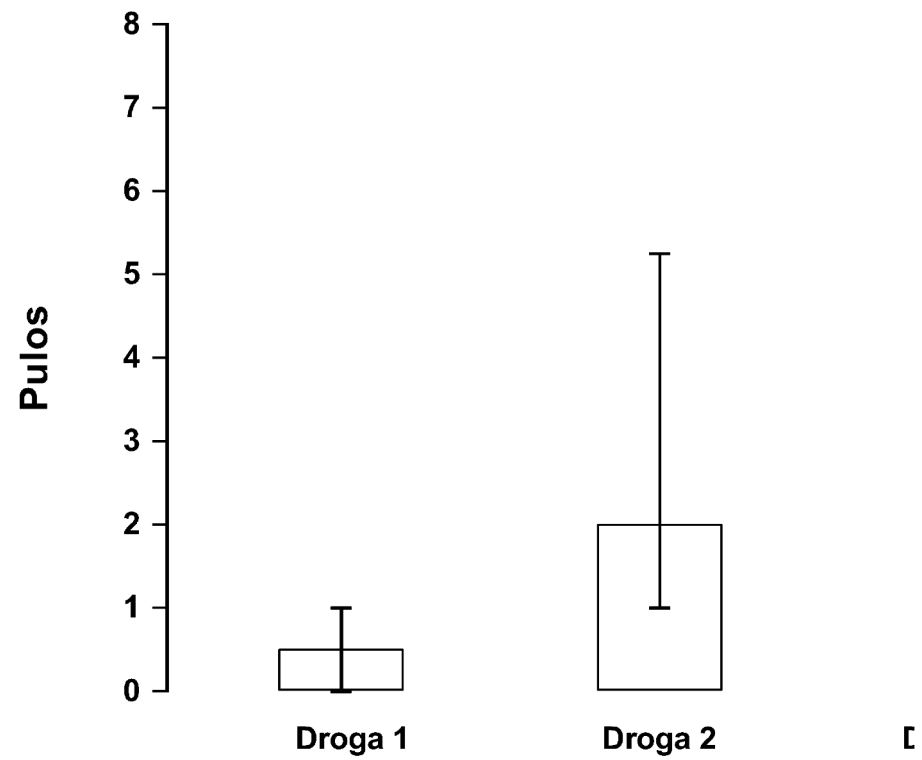




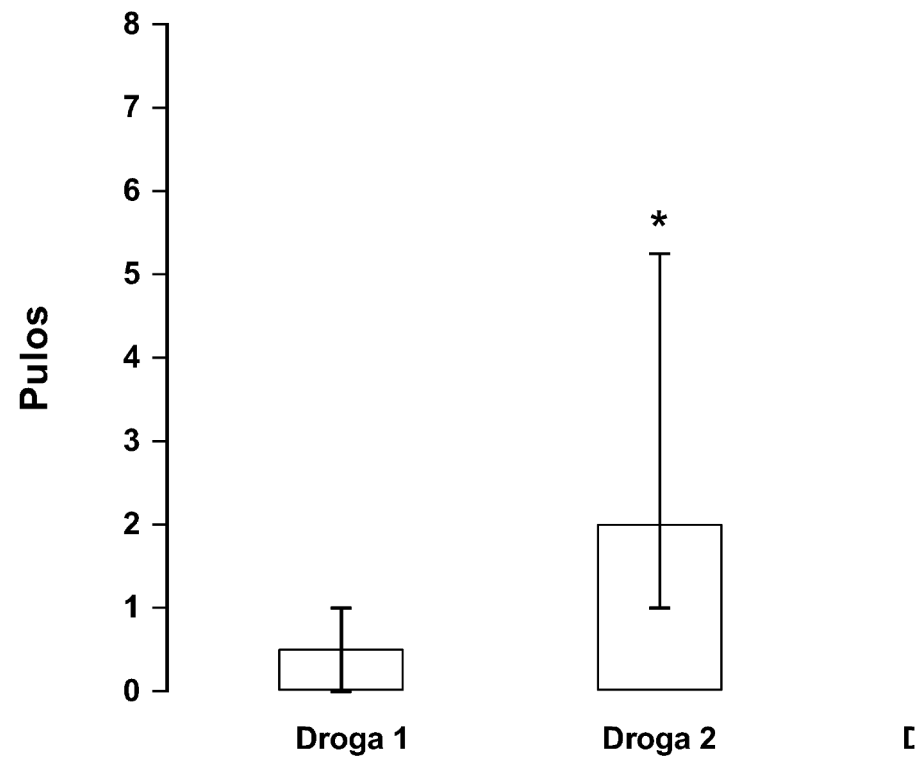


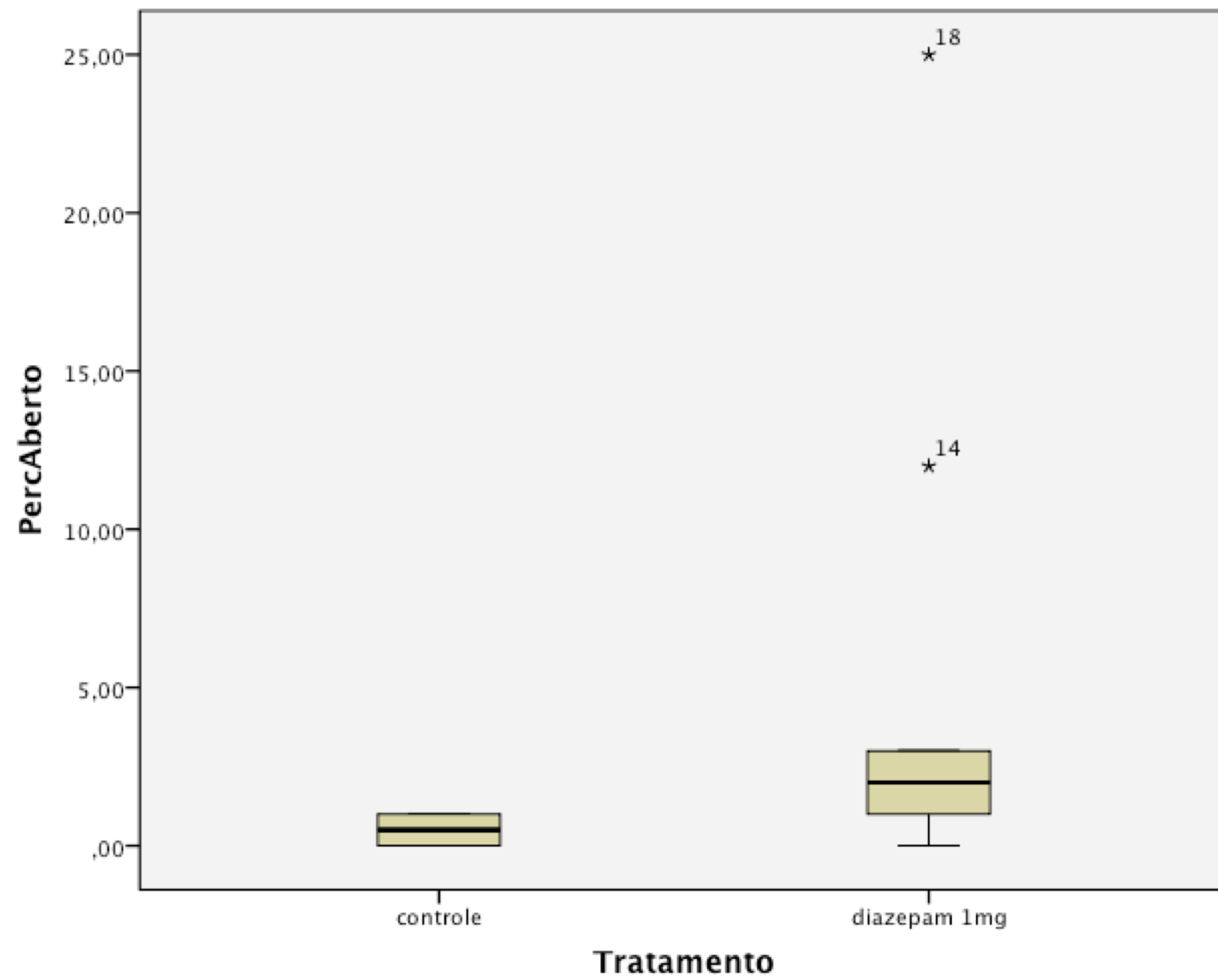


Médias +/- EPM



Medianas +/- intervalos interquartis. O quê falta?





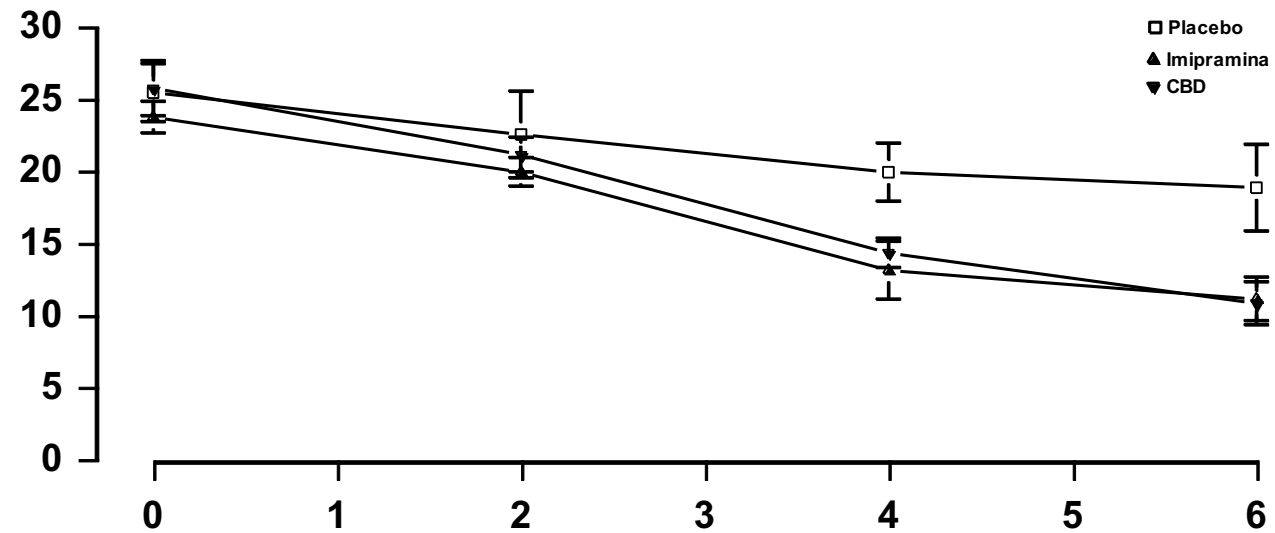
## Exemplo 3

**Descriptives**

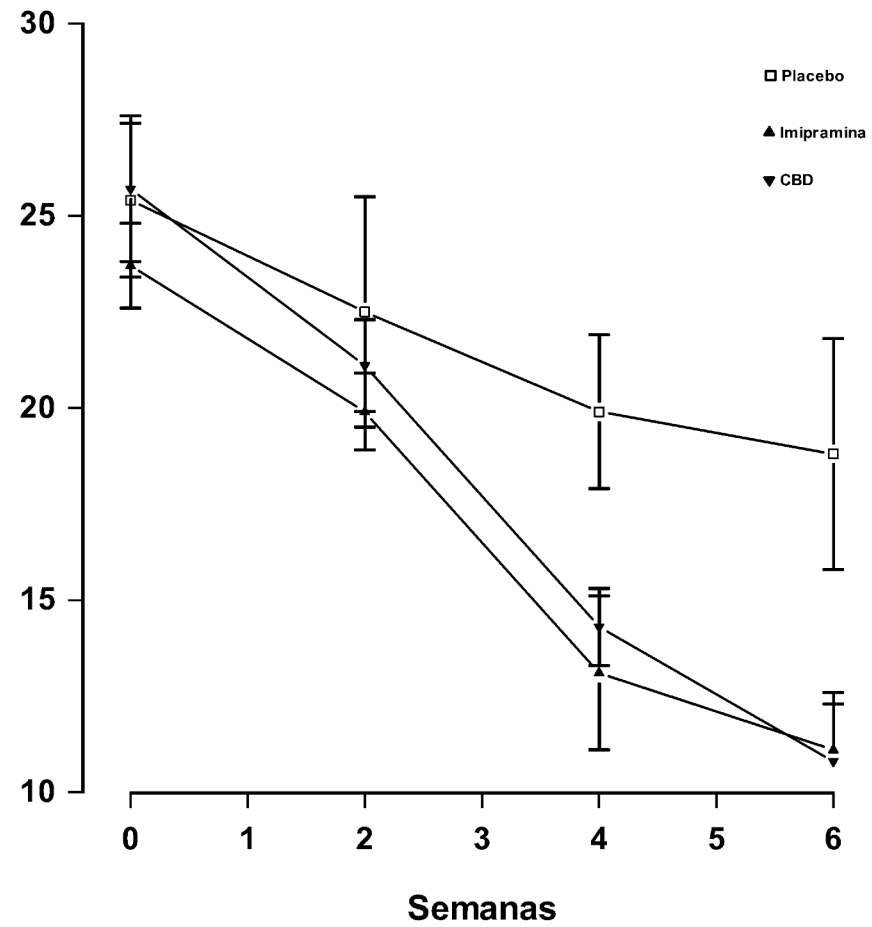
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
basal	1,00	10	25,4000	3,23866	1,02415	23,0832	27,7168	19,00	30,00
	2,00	9	23,6667	3,42783	1,14261	21,0318	26,3015	18,00	28,00
	3,00	11	25,7273	1,95402	,58916	24,4145	27,0400	23,00	29,00
	Total	30	25,0000	2,93610	,53606	23,9036	26,0964	18,00	30,00
segunda semana	1,00	10	22,5000	1,95789	,61914	21,0994	23,9006	19,00	25,00
	2,00	9	19,8889	3,05959	1,01986	17,5371	22,2407	15,00	26,00
	3,00	11	21,0909	1,64040	,49460	19,9889	22,1929	19,00	24,00
	Total	30	21,2000	2,41261	,44048	20,2991	22,1009	15,00	26,00
quarta semana	1,00	10	19,9000	2,76687	,87496	17,9207	21,8793	15,00	25,00
	2,00	9	13,1111	1,76383	,58794	11,7553	14,4669	11,00	16,00
	3,00	11	14,3636	2,69343	,81210	12,5542	16,1731	11,00	20,00
	Total	30	15,8333	3,81543	,69660	14,4086	17,2580	11,00	25,00
sexta semana	1,00	10	18,8000	2,85968	,90431	16,7543	20,8457	15,00	24,00
	2,00	9	11,1111	1,45297	,48432	9,9943	12,2280	9,00	14,00
	3,00	11	10,8182	1,32802	,40041	9,9260	11,7104	9,00	13,00
	Total	30	13,5667	4,23193	,77264	11,9864	15,1469	9,00	24,00

Drogas 1=placebo; 2=imipramina; 3=CBD

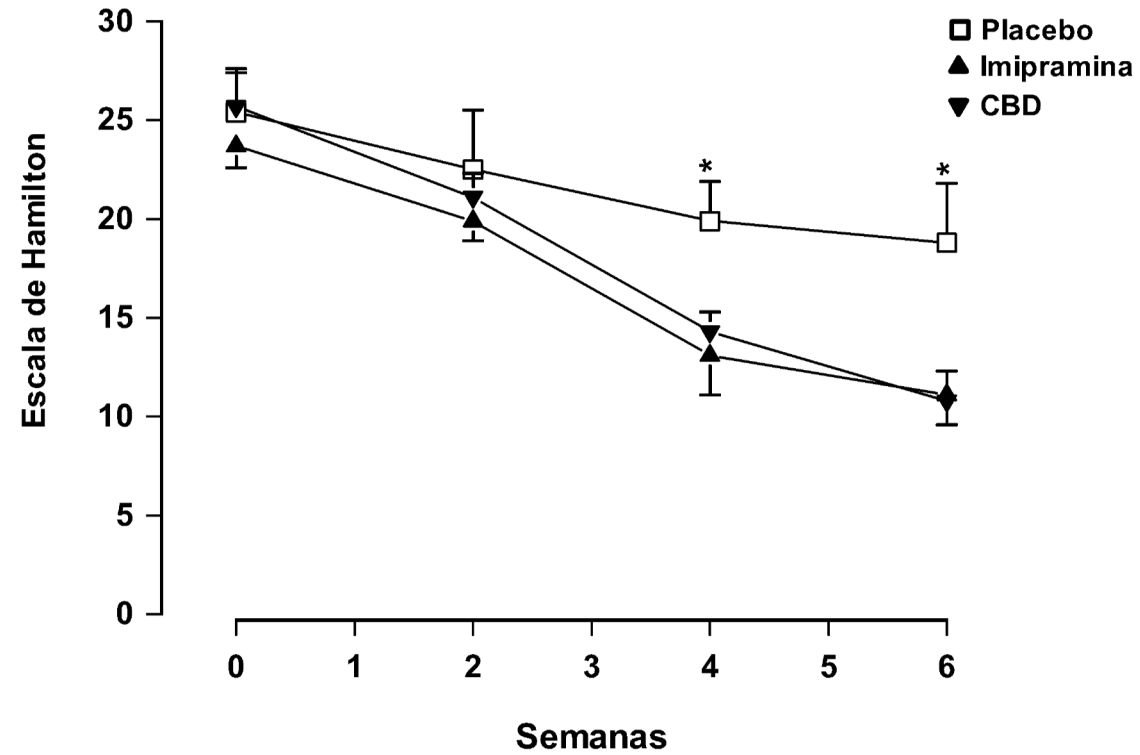
## Gráficos: exemplo 3. Quais os problemas?



# Gráficos: exemplo 3. Quais os problemas?

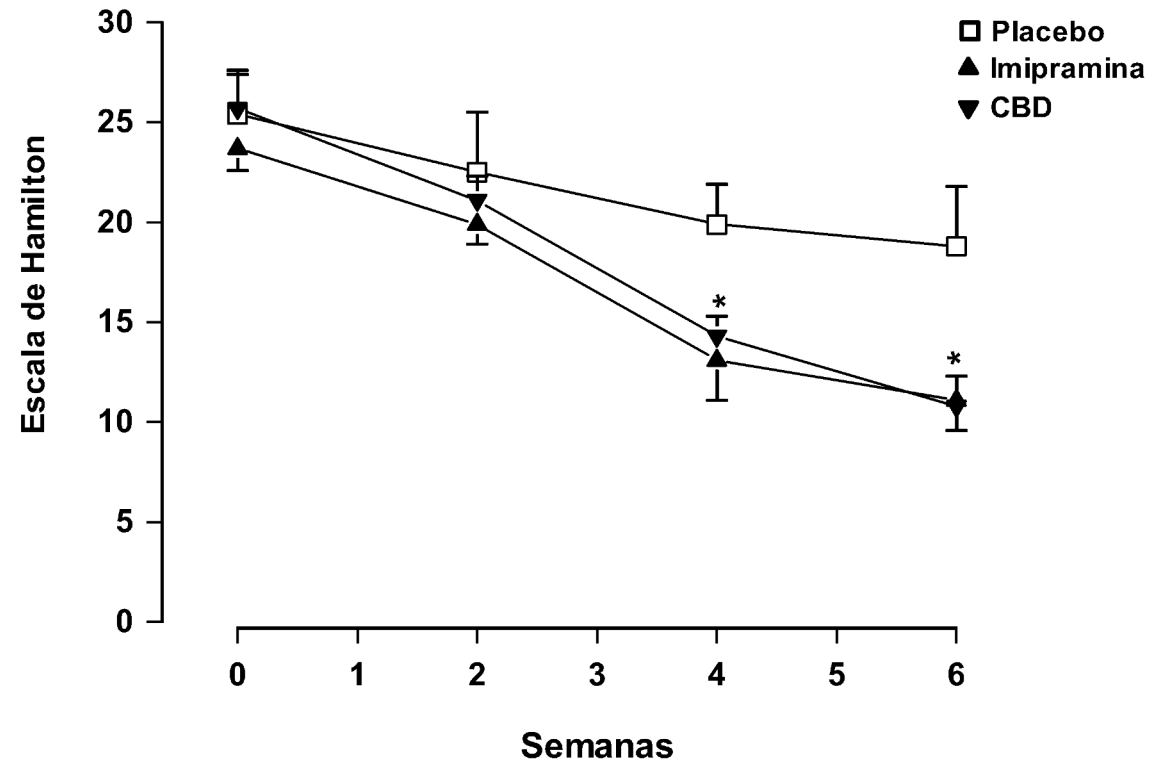


## Gráficos: exemplo 3. Quais os problemas?

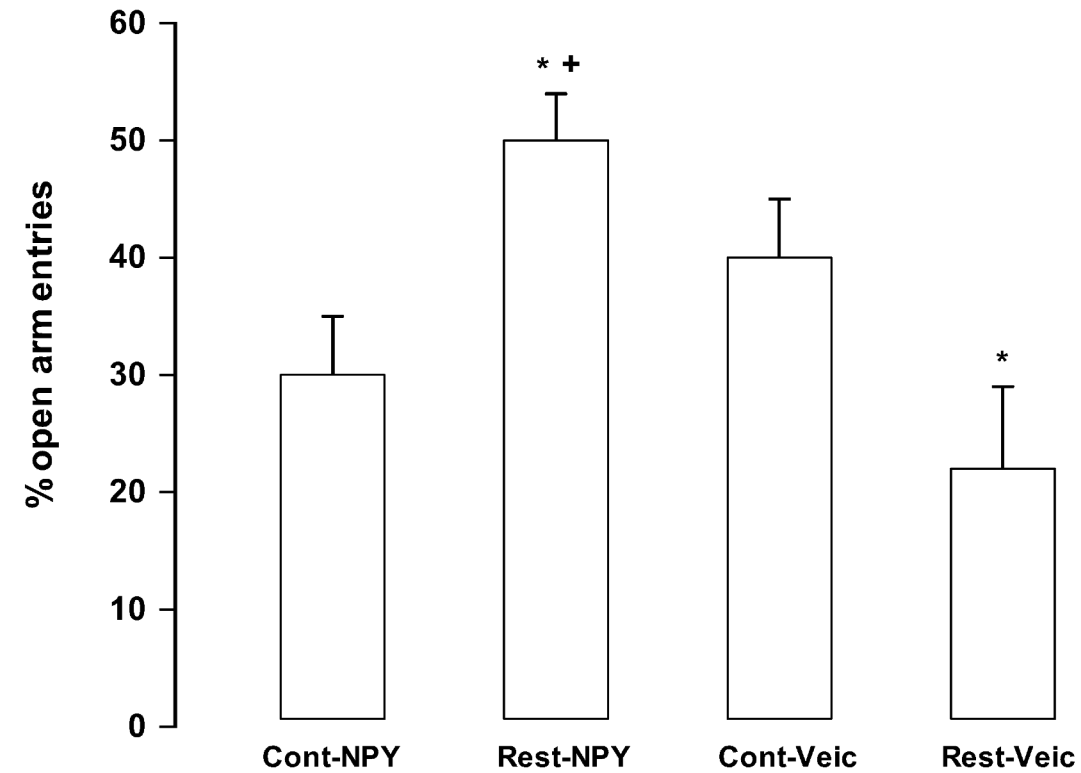




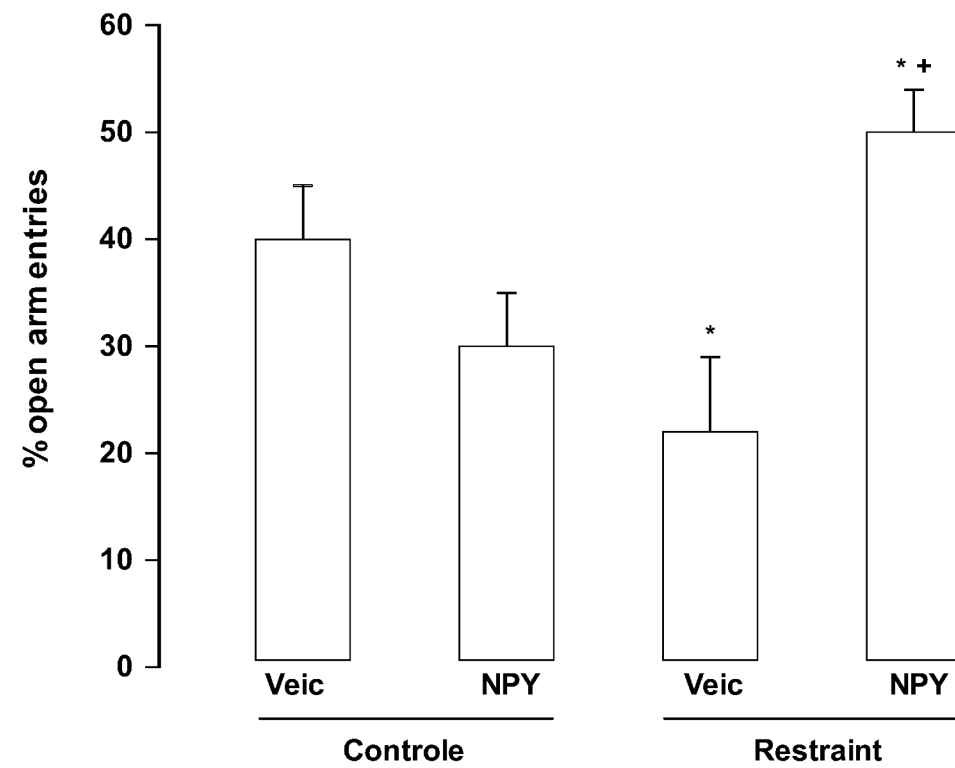
## Gráficos: exemplo 3. At last!



Exemplo 4. Qual o problema na figura? (ou melhor, qual o efeito observado?)



Médias + DP, \* p diferente de cont-veic,; + diferente de cont-NPY, p<0,05



Médias + DP, \* p diferente de cont-veic,; + diferente de cont-NPY,  $p < 0,05$

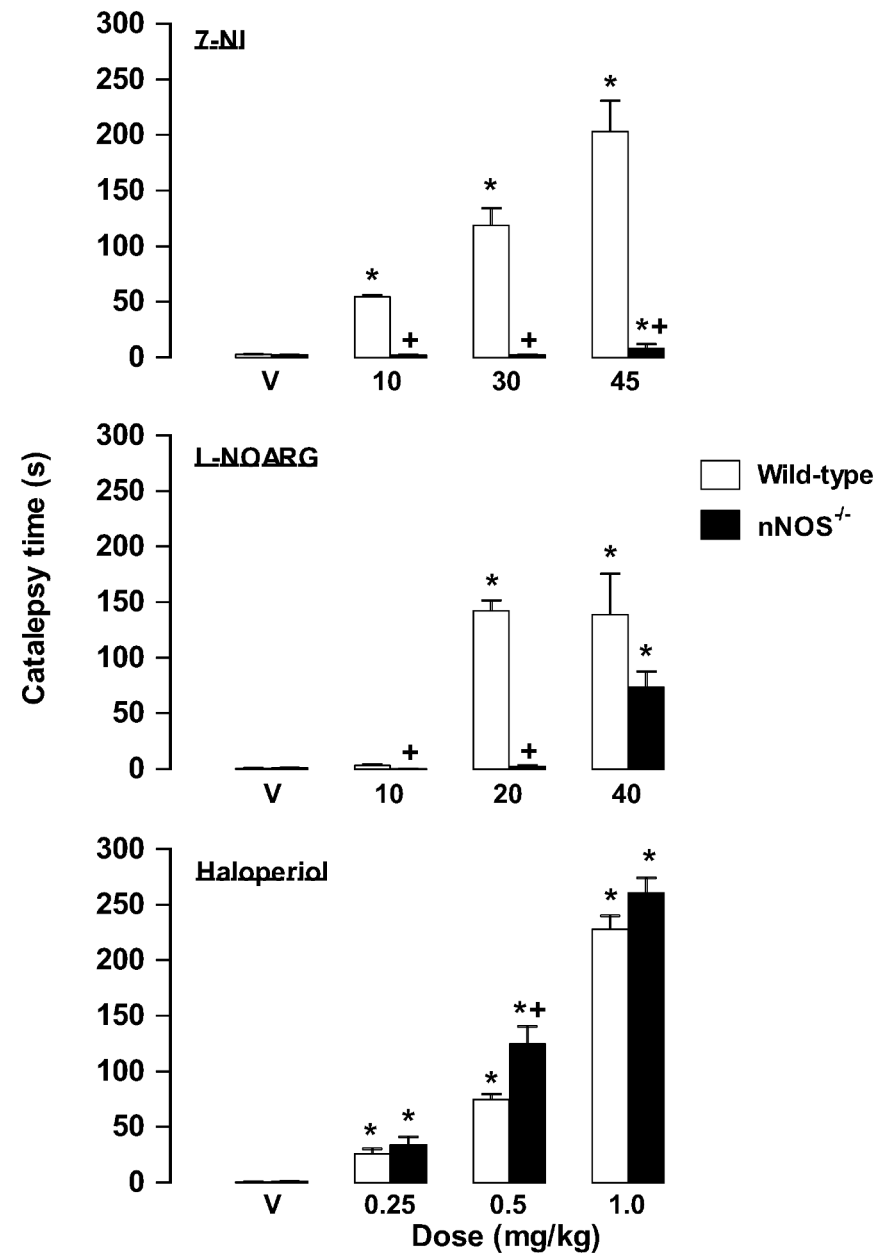
Como colocar estes dados em gráfico (n= 6/grupo)

Strain	doses		7NI		NOARG		Haloperidol	
			Mean	Standard Error of Mean	Mean	Standard Error of Mean	Mean	Standard Error of Mean
Wild type	1,00	Tempode catalepsia 60 min	2,50	,50	,56	,38	,56	,38
	2,00	Tempode catalepsia 60 min	54,50	1,38	3,00	,98	25,63	4,57
	3,00	Tempode catalepsia 60 min	118,83	15,31	142,38	9,28	74,75	4,84
	4,00	Tempode catalepsia 60 min	203,33	27,80	138,71	36,95	227,63	12,35
NOS-/-	1,00	Tempode catalepsia 60 min	2,20	,37	,78	,40	,78	,40
	2,00	Tempode catalepsia 60 min	1,83	,48	,00	,00	33,78	7,20
	3,00	Tempode catalepsia 60 min	2,00	,37	2,25	,96	124,50	15,74
	4,00	Tempode catalepsia 60 min	8,20	3,46	73,25	14,46	260,63	13,67

7NI; doses 1=veículo, 2=10, 3=30, 4=45 mg/kg

NOARG; doses 1=veículo, 2=10; 3=20; 4=40 ,mg/kg

Haloperidol; doses 1=veículo, 2=0,25; 3=0,5; 4=1,0 mg/kg

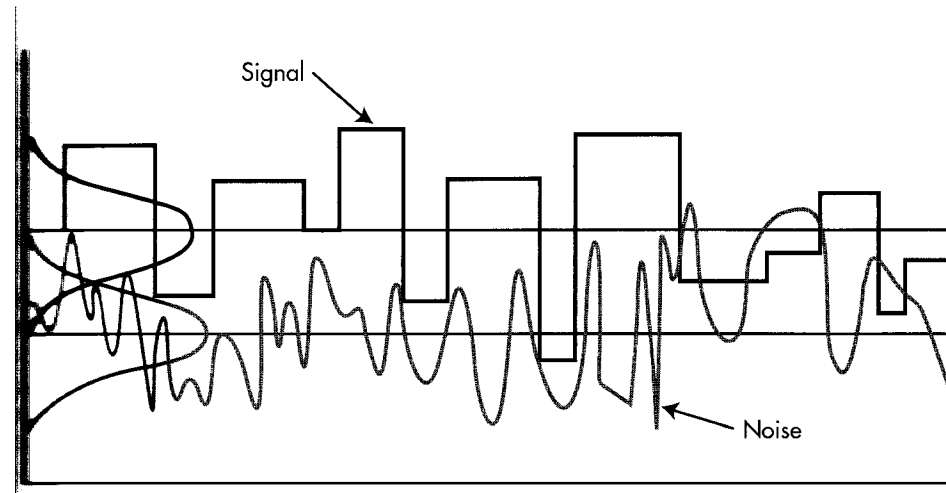




# Inferência estatística

Processo de inferir aspectos de uma população a partir de observações de uma amostra

Princípio geral: relação “sinal-ruído”



# O que é Ciência?

*A Ciência é mais do que um conjunto de conhecimentos, é uma forma de pensar*

*Carl Sagan*

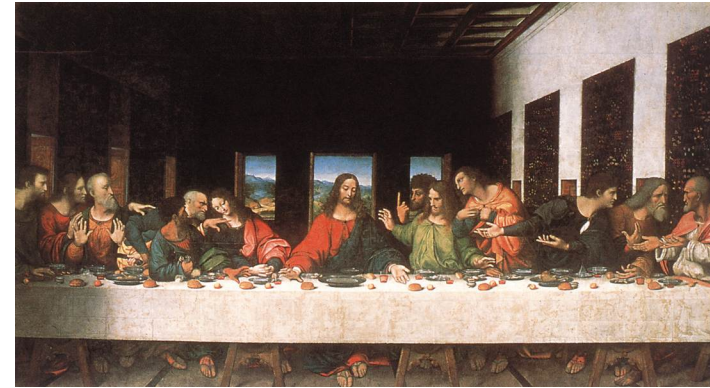


# Ciência:

## 1. Forma de ver e explicar o mundo



Giotto



da Vinci



Blake



Dali

## 2. Utiliza o método científico

# A Ciência não procura apenas descrever os fenômenos, mas também explicá-los

Ex: Babilônicos X Gregos

## Babilônicos:

Cálculos abstratos para  
prever a posição dos  
astros



Sumarizava os fatos



Astrologia

## Eudoxus e Aristóteles

Teoria das esferas



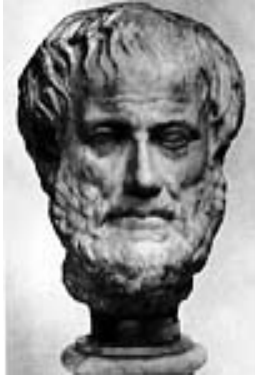
Tanto sumarizava os fatos  
quanto procurava explicá-  
los



Se desenvolveu nas idéias  
modernas sobre o universo

# Pesquisa científica: Formas de raciocínio

Aristóteles (384-322 AC): raciocínio dedutivo (silogismo)



Premissa maior: Todos estudantes vão à escola regularmente

Premissa menor: João é um estudante

Conclusão: João vai a escola regularmente

Francis Bacon (1561-1626): raciocínio indutivo



Premissa maior: João, Maria e José vão à escola regularmente

Premissa menor: João, Maria e José tem boas notas

Conclusão: Ir à escola regularmente resulta em boas notas

Ciência: combina os poderes do raciocínio dedutivo e indutivo no que é chamado **Método Científico**

**Galileu: *propõe hipóteses e as submete a prova experimental***





## John Dewey: Método científico

(1859-1952)

1. Identificar e definir o problema
2. Determinar a hipótese
3. Coletar e analisar os dados
4. Formular conclusões
5. Aplicar as conclusões às hipóteses originais

} Raciocínio  
dedutivo

} Raciocínio  
indutivo

# Teorias e hipóteses

Teoria: Conjunto organizado de princípios ou regras que visam descrever e explicar um certo conjunto de fatos

Hipótese: asserção escrita de forma clara e concisa sobre o que você pensa irá acontecer no seu projeto

Relacionada à teoria, mas:

1. Contém variáveis operacionalmente definidas
2. É apresentada sob uma forma testável

*O que é uma teoria científica?*

# **Teoria científica**

## **1. comprovável (“falsificável”)**

K. Popper: uma teoria científica será sempre conjectural e provisória

(“a verdade é inalcançável, todavia podemos nos aproximar dela por tentativas”)

## **2. Baseado em conhecimento científico prévio**

“Quando nós propomos uma teoria científica geral, a única coisa que podemos estar certos é que – falando em termos absolutos – todas estas teorias são falsas. Elas nada mais são do que verdades parciais e provisórias de que necessitamos, como degraus para descansar, para avançar nossa investigação”

Claude Bernard, 1865



## **Passos “tradicionais” na inferência estatística:**

1. Defina a questão biológica ou hipótese
2. Faça predições específicas sobre o que irá ocorrer se a hipótese for correta
3. Planeje os experimentos/observações para testar as predições
4. Defina as hipóteses de nulidade ( $H_0$ ) e alternativa ( $H_1$ ) de sua predição
5. Decida se o teste será mono ou bi-caudal
6. Determine o tipo de comparação (a) uma amostra; b) duas amostras relacionadas (dependentes); c) duas amostras não-relacionadas (independentes); d) k amostras relacionadas; e) k amostras não-relacionadas; f) uma associação (correlação); g) uma regressão
7. Decida o tipo de variável (nominal, ordinal ou contínua)

## Passos na inferência estatística:

8. Decida o teste a usar
9. Especifique o nível de significância  $\alpha$  (usualmente = 0,05)
10. Se o valor da estatística do teste  $\geq$  ao valor crítico rejeite  $H_0$ ,  
Se for  $<$  que  $H_0$ , aceite  $H_0$  e conclua que os resultados não são significativos (ou suspende julgamento se acreditar que as amostras são muito pequenas) (duas exceções são os testes de Wilcoxon e Mann-Whitney, nos quais o valor da estatística do teste deve ser menor ou igual ao valor crítico)
11. Descreva o resultado da análise: frase descrevendo o resultado (no passado) e o valor do teste, g.l. (ou n) e valor do P, Ex, O diazepam aumentou a percentagem de entradas do braço aberto (quando comparado ao grupo veículo) ( $t=4,3$ , g.l.=12,  $p=0,03$ )

# Distribuição normal

Mais importante na bioestatística

Distribuição contínua (representada por curva de frequência)

Pré-requisito para muitos testes

Muitos fenômenos naturais seguem um distribuição aproximadamente normal

**Importante: Qualquer que seja a distribuição original a distribuição das médias com um  $n^\circ$  razoável de sujeitos será normal**

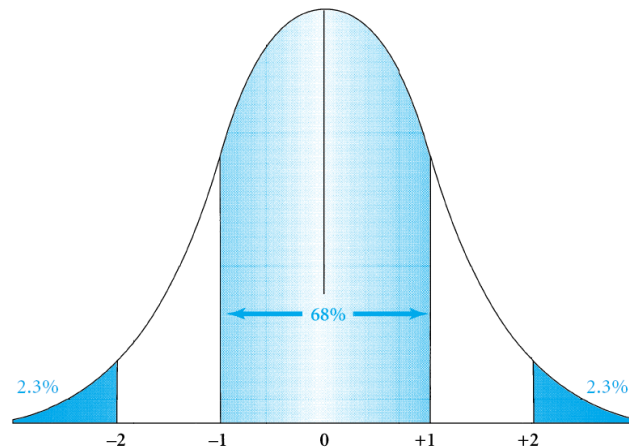


Figure 2-5 The normal distribution (bell-shaped curve).

Características: média=mediana=moda

A média e a variância são independentes

Simétrica, skewness (mede deslocamento para um dos lados)=0, kurtosis (mede achatamento) =0

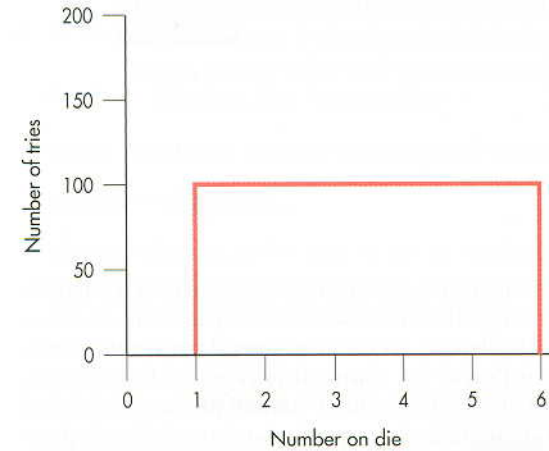
# O teorema do “limite central”



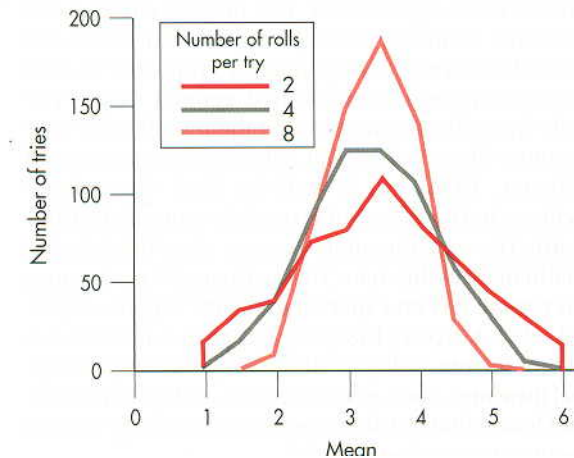
**Qual seria a distribuição dos resultados se rolasse o dado 600 x?**

**Qual seria a distribuição dos resultados se fizesse a média de 2, 4 ou 6 x que rolasse o dado (repetidas 600 x)?**

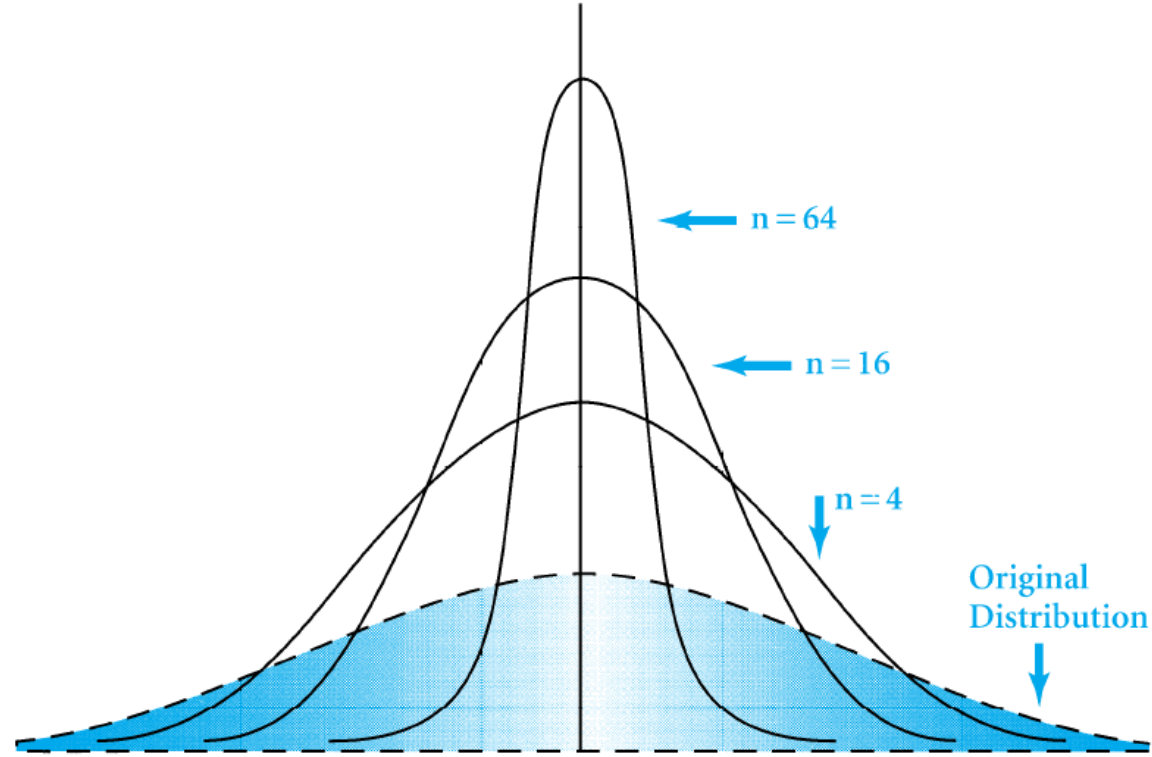
**FIGURE 4-1**  
Theoretical distribution from rolling a die 600 times.



**FIGURE 4-2**  
Computer simulation of averaging the sum of rolling the die 2, 4, and 8 times, each done 600 times.



# O teorema do “limite central”



**Figure 3-1** Original distribution and distribution of means related to sample size.

# Testes de normalidade

VAR00001 Stem-and-Leaf Plot

Frequency Stem & Leaf

3,00 Extremes (= <1,0)  
 3,00 2 . 000  
 4,00 3 . 0000  
 7,00 4 . 0000000  
 9,00 5 . 000000000  
 6,00 6 . 000000  
 4,00 7 . 0000  
 3,00 8 . 000  
 3,00 Extremes (>=9,0)

Stem width: 1,00

Each leaf: 1 case(s)

## Descriptives

		Statistic	Std. Error
VAR00001	Mean	4,9762	,35168
	95% Confidence Interval for Mean	4,2660	
	Lower Bound	5,6864	
	Upper Bound		
	5% Trimmed Mean	4,9735	
	Median	5,0000	
	Variance	5,195	
	Std. Deviation	2,27915	
	Minimum	,00	
	Maximum	10,00	
	Range	10,00	
	Interquartile Range	2,50	
	Skewness	,031	,365
	Kurtosis	-.196	,717

Em relação à normalidade: se skewness e kurtosis forem menores do que duas vezes seus DP, tudo bem

Testes de normalidade

VAR00002 Stem-and-Leaf Plot

Frequency	Stem & Leaf
1,00	0 . 0
1,00	1 . 0
1,00	2 . 0
2,00	3 . 00
2,00	4 . 00
4,00	5 . 0000
4,00	6 . 0000
5,00	7 . 00000
6,00	8 . 000000
12,00	9 . 00000000000000
4,00	10 . 0000

Stem width: 1,00  
Each leaf: 1 case(s)

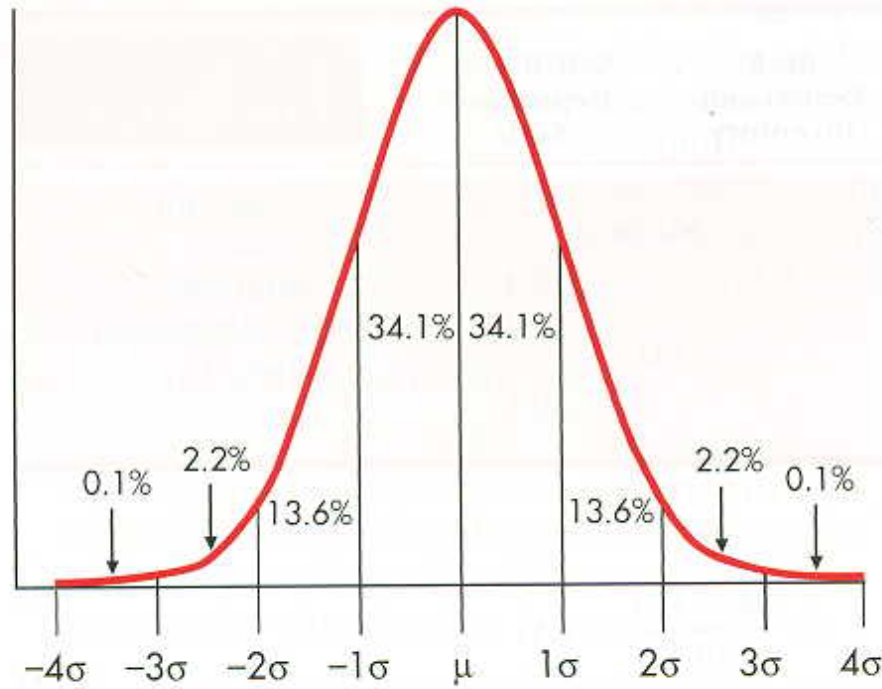
VAR00002	Mean	6,9524	,39590
	95% Confidence Interval for Mean	Lower Bound 6,1528	
		Upper Bound 7,7519	
	5% Trimmed Mean	7,1376	
	Median	8,0000	
	Variance	6,583	
	Std. Deviation	2,56574	
	Minimum	,00	
	Maximum	10,00	
	Range	10,00	
	Interquartile Range	4,00	
	Skewness	-,991	,365
	Kurtosis	,323	,717

Tests of Normality

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
VAR00001	,115	42	,187	,981	42	,682
VAR00002	,182	42	,001	,891	42	,001

a. Lilliefors Significance Correction

Estatística Z= padronização da curva normal =  $(x-\mu)/\sigma$



Permite comparar diferentes medidas



Toda a inferência estatística possui um erro, que pode ser:

Sistemático

e/ou

Aleatório

Erros tipo I e II (níveis  $\alpha$  e  $\beta$  de significância)

“Poder estatístico” ( $1 - \beta$ )

# A relação entre o n e os valores para atingir significância

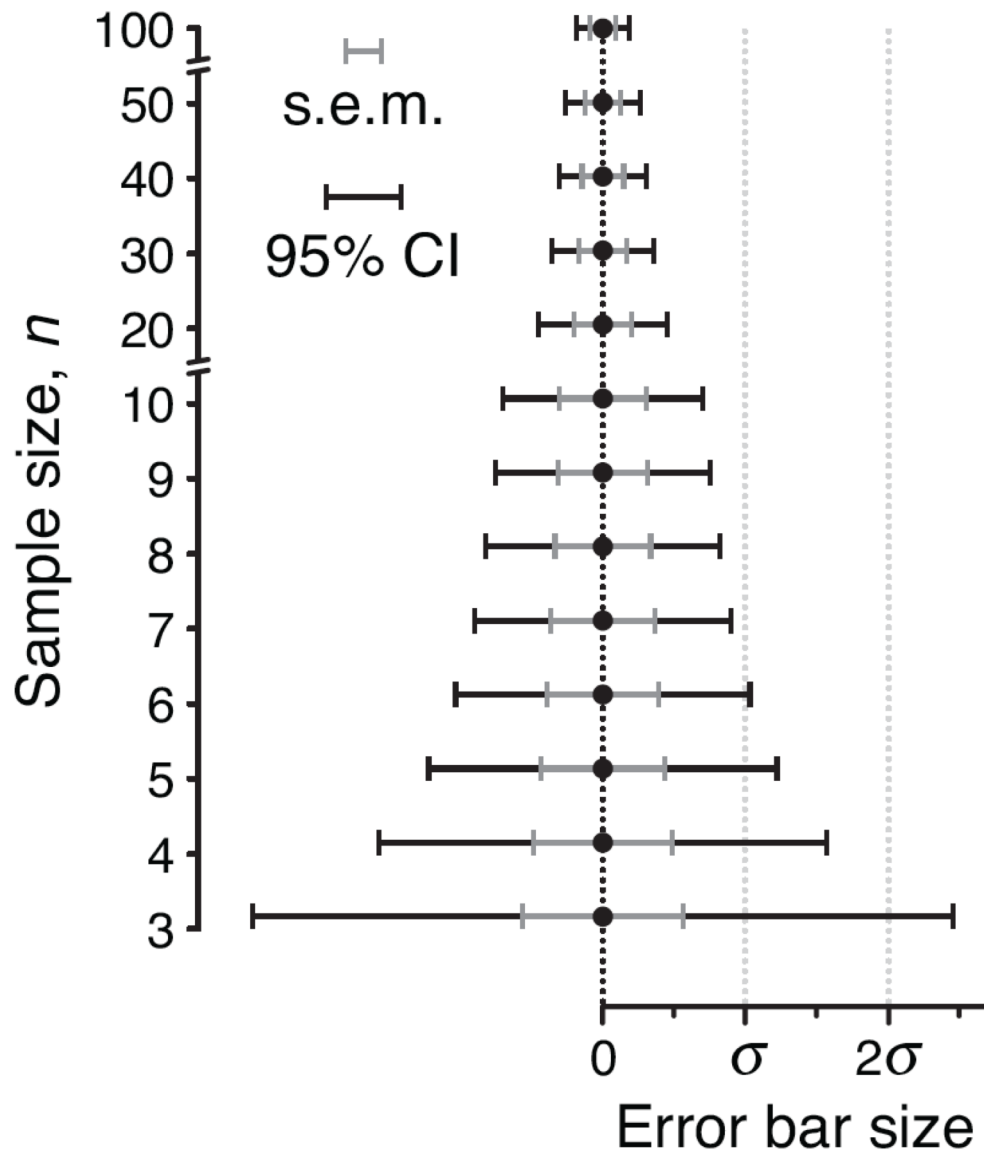
## Relationship of Sample Size and Mean Values to Achieve Statistical Significance

Sample Size	Reader Mean	Population	<i>p</i>
4	110.0	100.0	0.05
25	104.0	100.0	0.05
64	102.5	100.0	0.05
400	101.0	100.0	0.05
2,500	100.4	100.0	0.05
10,000	100.2	100.0	0.05

Com um n muito pequeno não se consegue demonstrar nada,  
e com um n muito grande se consegue demonstrar tudo!

EPM e intervalo de confiança são interrelacionados pela estatística t. Em grandes amostras o EPM pode ser interpretado como um IC com nível de confiança de 67%

## Medidas de dispersão e diferenças entre as médias: Influência do n



Relação  
tamanho da  
amostra x  
significância

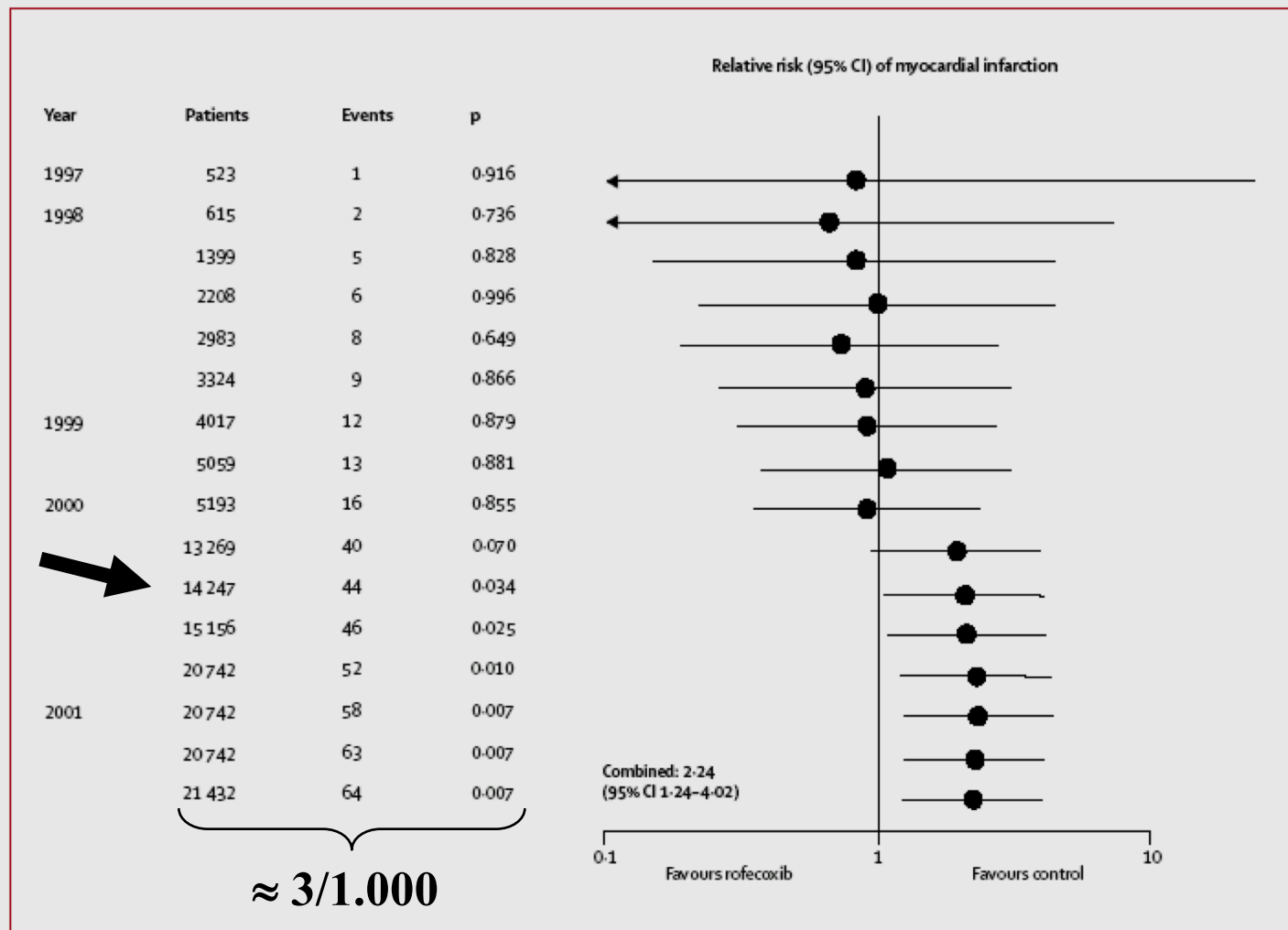


Figure 3: Cumulative meta-analysis of randomised trials comparing rofecoxib with control  
See figure 2 for sequence of trials.

**Risco de infarto do miocárdio por Viox (rofecoxib): meta-análise cumulativa indicava que desde 2000 haviam evidências de efeito adverso importante (Jüni P, Lancet 2004)**

Qual o tamanho ideal da amostra?

Precisa de: diferença esperada,  $\sigma$  estimado, nível  $\alpha$  e  $\beta$  de significância

Ex,:

Qual o n para observar efeito ansiolítico de composto desconhecido no LCE? (aumento na percentagem de entradas nos braços abertos)

Diferença esperada: 10

$\sigma = 8$ ;  $\alpha = 0,05$ ;  $\beta = 0,20$

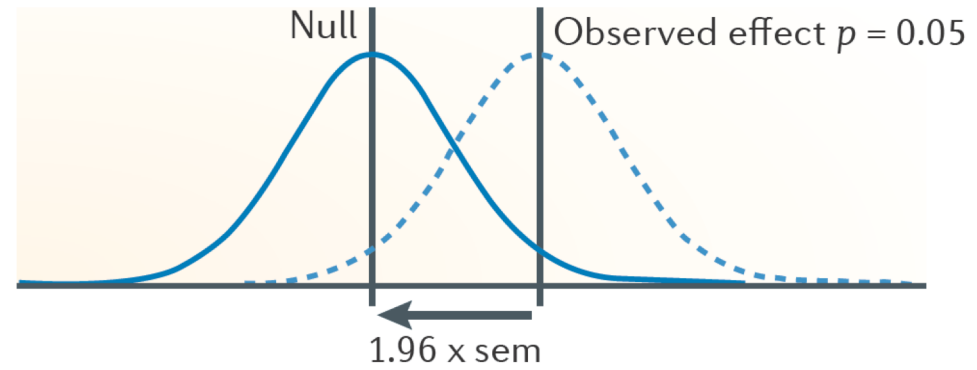
**Regra grosseira para diferença entre médias:  $n = 16 s^2/d^2$**

**entre proporções:  $n = 16 p(1-p)/(p_1 - p_2)^2$ , aonde p é a média das proporções  $p_1$  e  $p_2$**

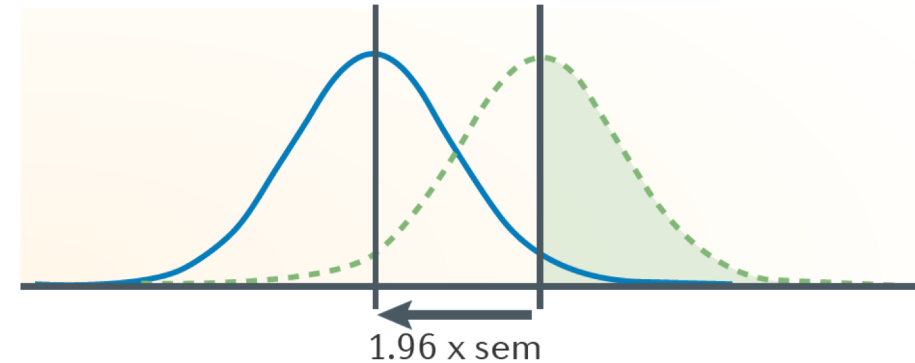
Poder estatístico de um estudo que tenta reproduzir um resultado anterior (considerando que a diferença seja real): chance de não encontrar diferença pode chegar a 50%

Pode aumentar a chance de encontrar diferença aumentando o  $n$ .

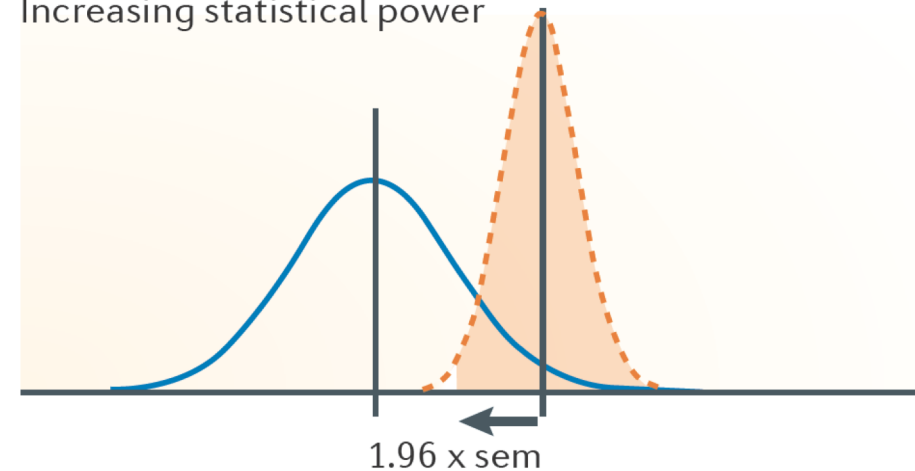
**a** Rejecting the null hypothesis



**b** The sampling distribution

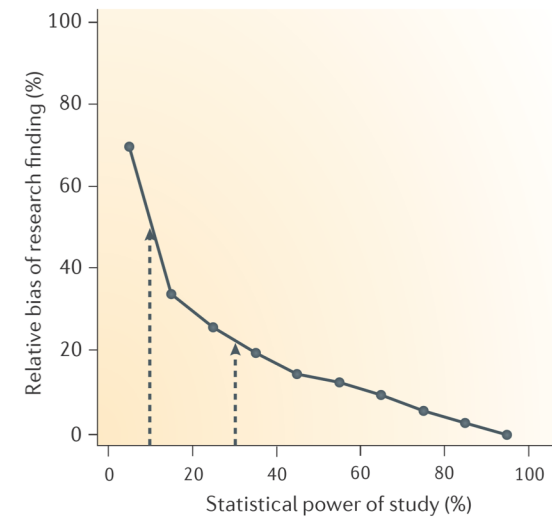


**c** Increasing statistical power



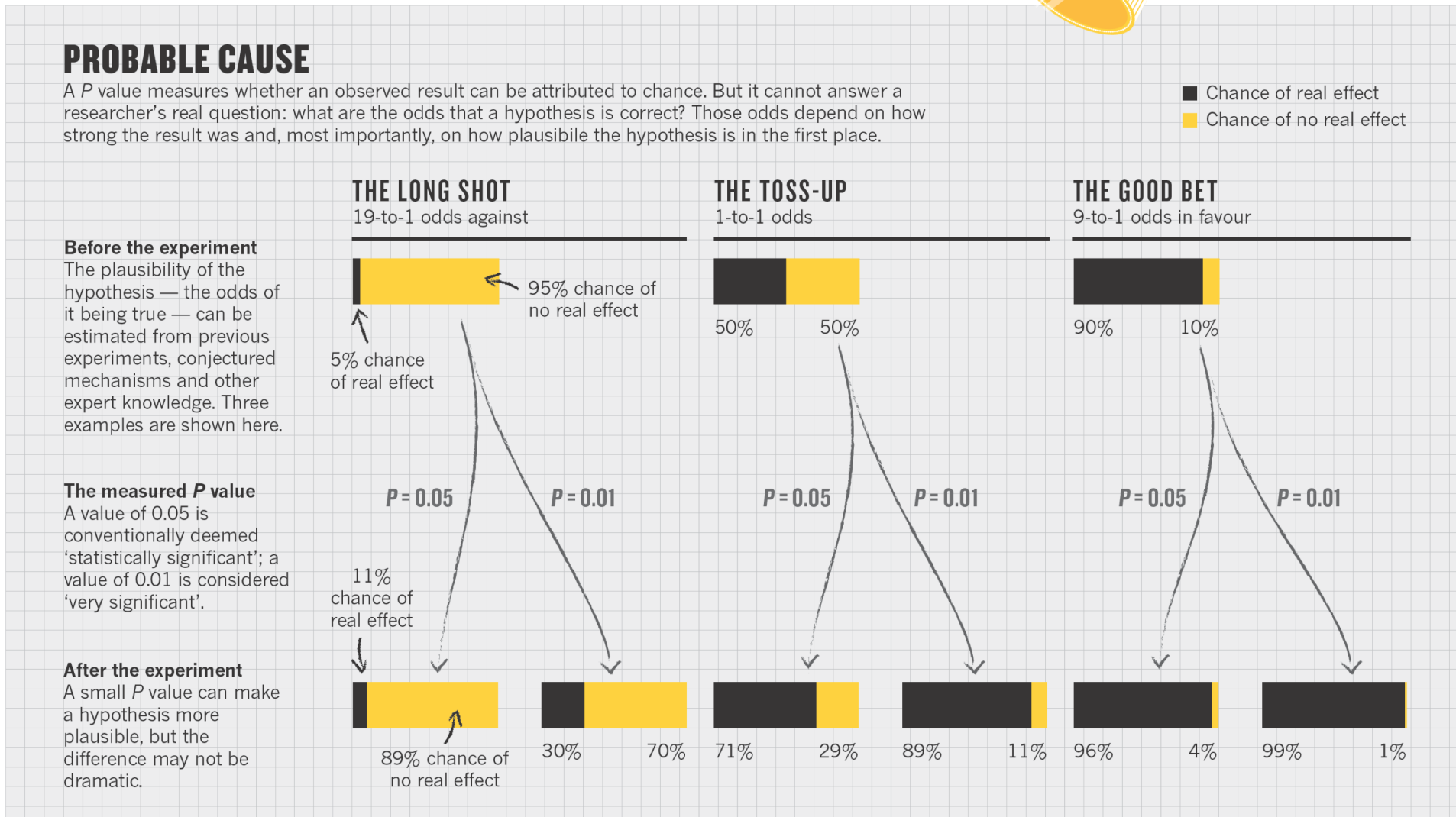
Estudos com amostras pequenas que obtem resultados “significativos” tendem a exagerar o tamanho do efeito, o que pode introduzir bias na literatura e dificultar a reprodutibilidade se os cálculos da amostra de futuros estudos forem baseados neles (em neurociências o poder estimado varia, usualmente, entre 8 a 31%)

*Button et al., Nature Rev Neurosc 2013*



**Figure 5 | The winner's curse: effect size inflation as a function of statistical power.** The winner's curse refers to the phenomenon that studies that find evidence of an effect often provide inflated estimates of the size of that effect. Such inflation is expected when an effect has to pass a certain threshold — such as reaching statistical significance — in order for it to have been 'discovered'. Effect inflation is worst for small, low-powered studies, which can only detect effects that happen to be large. If, for example, the true effect is medium-sized, only those small studies that, by chance, estimate the effect to be large will pass the threshold for discovery (that is, the threshold for statistical significance, which is typically set at  $p < 0.05$ ). In practice, this means that research findings of small studies are biased in favour of inflated effects. By contrast, large, high-powered studies can readily detect both small and large effects and so are less biased, as both over- and underestimations of the true effect size will pass the threshold for 'discovery'. We optimistically estimate the median statistical power of studies in the neuroscience field to be between ~8% and ~31%. The figure shows simulations of the winner's curse (expressed on the y-axis as relative bias of research findings). These simulations suggest that initial effect estimates from studies powered between ~8% and ~31% are likely to be inflated by 25% to 50% (shown by the arrows in the figure). Inflated effect estimates make it difficult to determine an adequate sample size for replication studies, increasing the probability of type II errors. Figure is modified, with permission, from REF. 103 © (2007) Cell Press.

O Valor de P mede a probabilidade do resultado daquele experimento específico decorrer de chance (considerando que  $H_0$  é correta). Ele não, mede, no entanto, a probabilidade da hipótese estar correta. Esta depende também da probabilidade inicial de que seja certa.



Nuzzo, *Nature Rev Neurosc* 2013

25% dos estudos com resultados positivos ( $P < 0,05$ ) podem ser falsos (Hayden, *Nature* 2015)



## Testes Bayesianos versus testes usuais

Testes usuais: medem quão improvável um resultado pode ter ocorrido por acaso

Testes Bayesianos: medem a probabilidade de uma hipótese ser correta considerando os dados obtidos no estudo

Usando a segunda abordagem Jonhson calculou que 17-25% dos resultados com  $P < 0.05$  são falsos.  
Ele sugere usar algo como  $P < 0.005$

*NATURE* | NEWS

## Weak statistical standards implicated in scientific irreproducibility

One-quarter of studies that meet commonly used statistical cutoff may be false.

Erika Check Hayden

11 November 2013

# Testes estatísticos mais utilizados para testar hipóteses

Tipo de Experimento					
Escala de medida	Sujeitos diferentes		Mesmo Sujeito		Associação entre variáveis
	Dois tratamentos	Três ou mais tratamentos	Dois tratamentos	Três ou mais tratamentos	
contínua (com distribuição normal)*	Teste t não-pareado	ANOVA	Teste t pareado	ANOVA de medidas repetidas	Regressão linear de Pearson Correlação
Nominal	Qui-quadrado	Qui-quadrado	Teste de McNemar	Cochrane Q	Risco relativo ou razão de probabilidade
Ordinal	Mann-Whitney	Kruskal-Wallis	Wilcoxon	Friedman	Spearman
Tempo de sobrevivência	Log-rank ou teste de Gehan				

\* Para comparar um média amostral com a da população usar a estatística Z

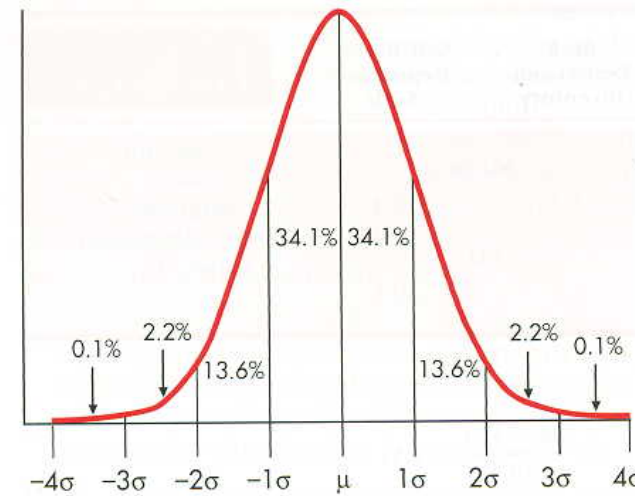
**Estatística  $Z = (x_i - \mu) / (\sigma)$**

Ex:  $Z = 2,00$

Área de curva normal com  
 $Z = 2,00 = 0,4772$

Ou seja, a probabilidade de um valor ser +2 DP diferentes da média populacional é de  $(0,5 - 0,4772) = 0,0228$

Qual valor de Z indicaria a chance de <0,05 do indivíduo ser maior do que a média?  
 $Z = 1,70$  (probabilidade = 0,0446)



**TABLE 4-3**

	<i>z</i>	Area below
A portion of the table of the normal curve	0.00	.0000
	0.10	.0398
	0.20	.0793
	0.30	.1179
	0.40	.1554
	0.50	.1915
	0.60	.2257
	0.70	.2580
	0.80	.2881
	0.90	.3159
	1.00	.3413
	1.00	.3413
	1.10	.3643
	1.20	.3849
	1.30	.4032
	1.40	.4192
	1.50	.4332
	1.60	.4452
	1.70	.4554
	1.80	.4641
	1.90	.4713
	2.00	.4772

# Testes mono e bi-caudais

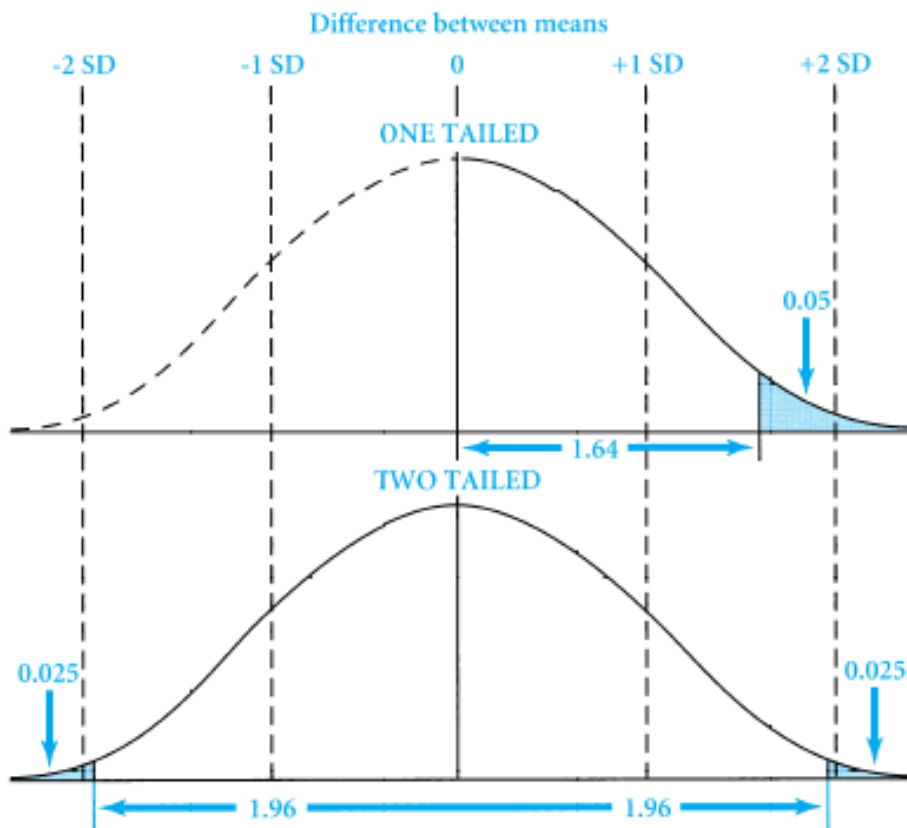


Figure 3-8 Comparison of one-tailed and two-tailed tests of significance.

Mono-caudal: especifica, previamente, a direção da possível diferença de interesse

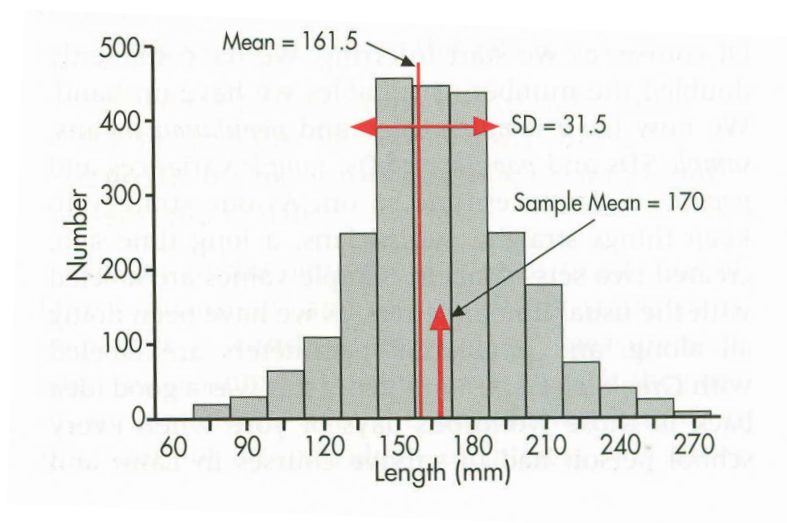
Bi-caudal: não especifica, previamente, a direção da possível diferença de interesse

Exemplo com uma amostra: Estudo realizado com 100 homens que usaram um produto vegetal chamado “mangro” que aumentaria o tamanho peniano.

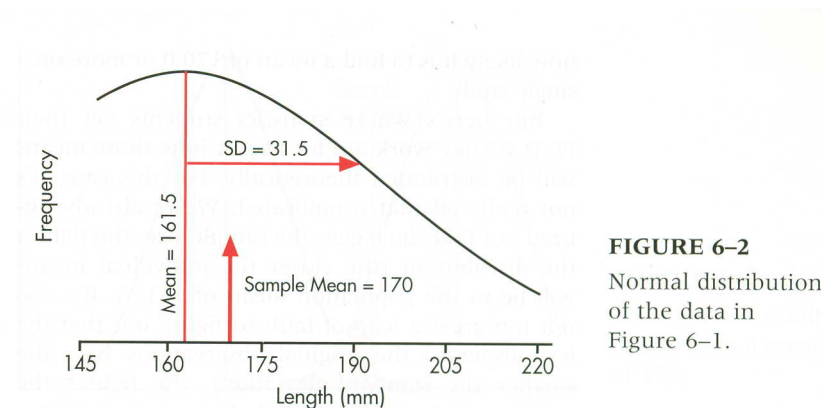
Resultados: média: 170mm

População: média: 161,5, DP: 31,5

Pergunta: O Mangro funciona?



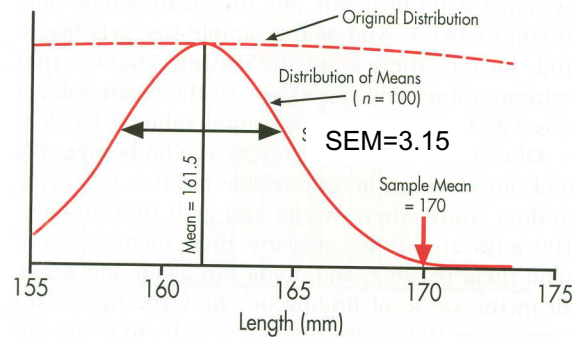
Resultados da população



Curva normal da população  
baseada na média e desvio  
padrão

A pergunta, no entanto, não é se um valor (ex. 170) de um particular indivíduo é diferente da média populacional mas sim qual seria a chance de uma média de 170 de uma amostra de 100 sujeitos retiradas de uma população com média de 161,5 ser diferente da média populacional?

**FIGURE 6-3**  
The distribution of means of  $n = 100$  from the population in Figure 6-1.



Erro padrão:  $EP = S/\sqrt{n}$

$$Z = \frac{170 - 161,5}{3,15} = 2,7$$

Portanto: O **desvio padrão** indica a proximidade de um indivíduo da média amostral

O **erro padrão** indica, para um dado tamanho amostral, a proximidade das médias de amostras repetidas da média populacional.

Estatística  $Z = (x - \mu) / (\sigma / \sqrt{n})$

$Z = 170 - 161,5 / 3,15$

$Z = 2,70$

Área de curva normal com  
 $z = 2,70 = 0,4965$

Ou seja, a probabilidade da  
média obtida ser diferente da  
populacional é de  $(0,5 - 0,4965) = 0,0035$

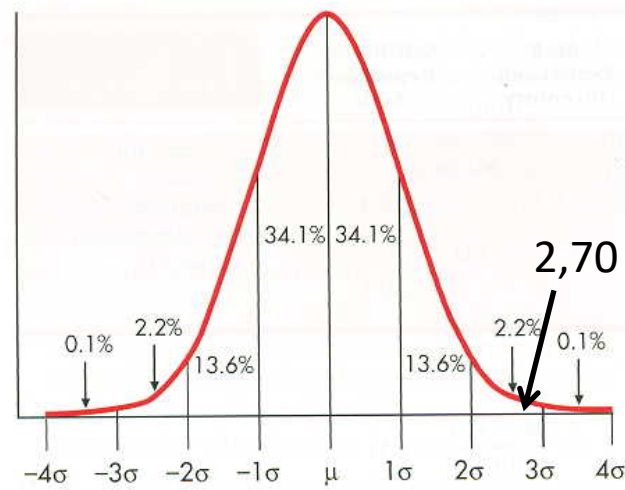


TABLE 4-3

	<i>z</i>	Area below
A portion of the table of the normal curve	0.00	.0000
	0.10	.0398
	0.20	.0793
	0.30	.1179
	0.40	.1554
	0.50	.1915
	0.60	.2257
	0.70	.2580
	0.80	.2881
	0.90	.3159
	1.00	.3413
	1.00	.3413
	1.10	.3643
	1.20	.3849
	1.30	.4032
	1.40	.4192
	1.50	.4332
	1.60	.4452
	1.70	.4554
	1.80	.4641
	1.90	.4713
	2.00	.4772

## Significância:

1. Estatística: os dados permitem rejeitar  $H_0$  e concluir pela existência de diferença
2. Biológica: O resultado tem relevância biológica
3. Prática: eles justificam uma determinada ação prática



Intervalo de confiança: medida da representatividade da média da amostra em relação a média populacional

IC (95%)= média amostral  $\pm$  (1,96 x EPM)

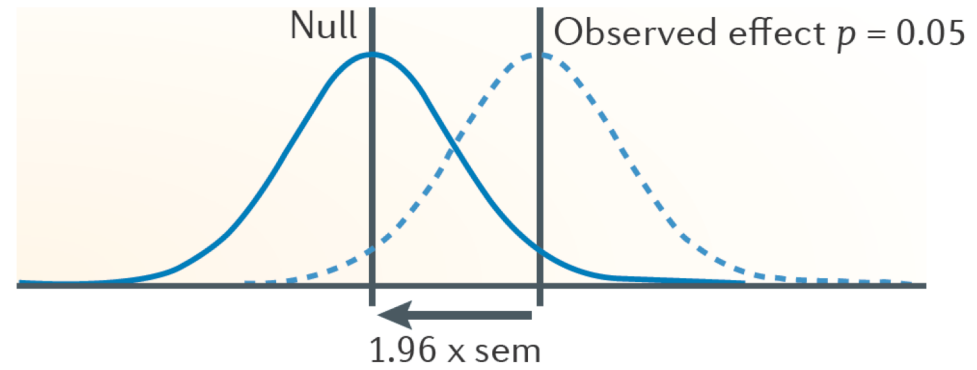
Quando se conhece o desvio padrão da população.

Se essa é desconhecida, utiliza-se a distribuição t (se  $n > 30$ , t aproxima-se de Z)

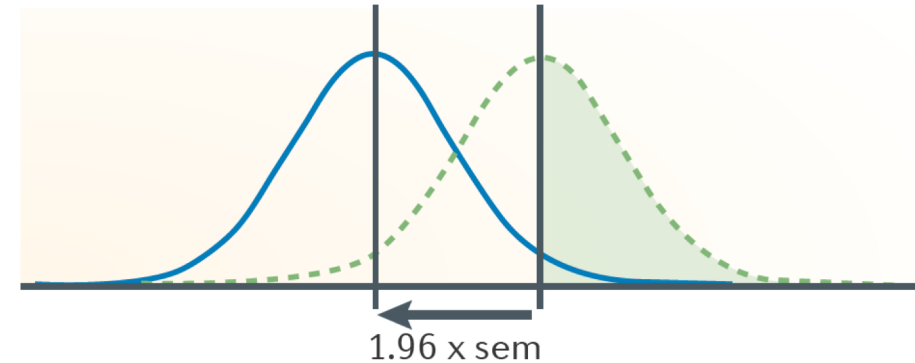
Poder estatístico de um estudo que tenta reproduzir um resultado anterior (considerando que a diferença seja real): chance de não encontrar diferença pode chegar a 50%

Pode aumentar a chance de encontrar diferença aumentando o n.

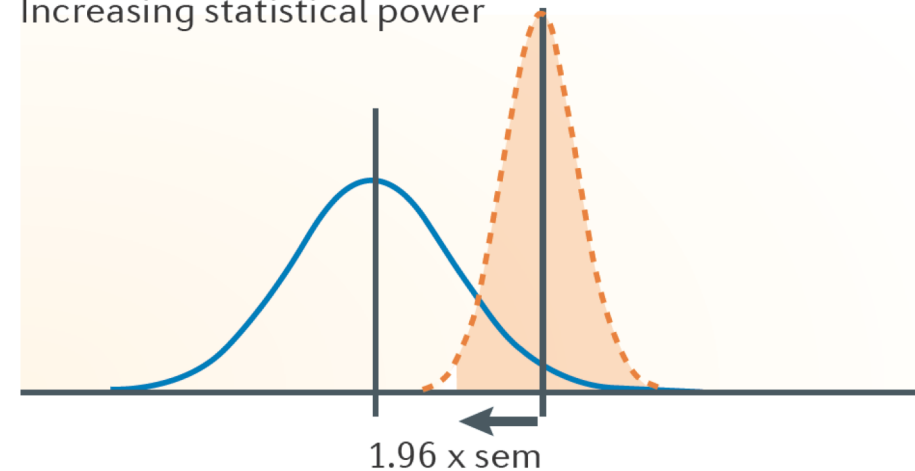
**a** Rejecting the null hypothesis



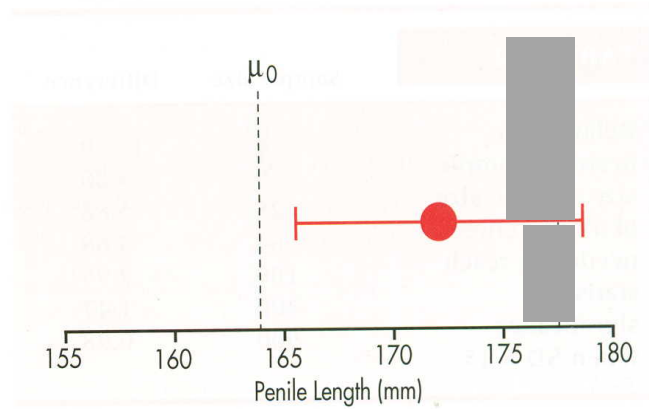
**b** The sampling distribution



**c** Increasing statistical power

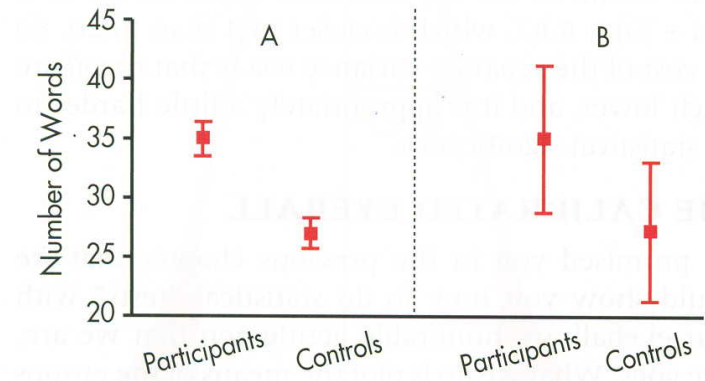


Outra forma de mostrar os resultados: como intervalo de confiança



**FIGURE 6-8**  
Another way of  
showing the  
95% CI.

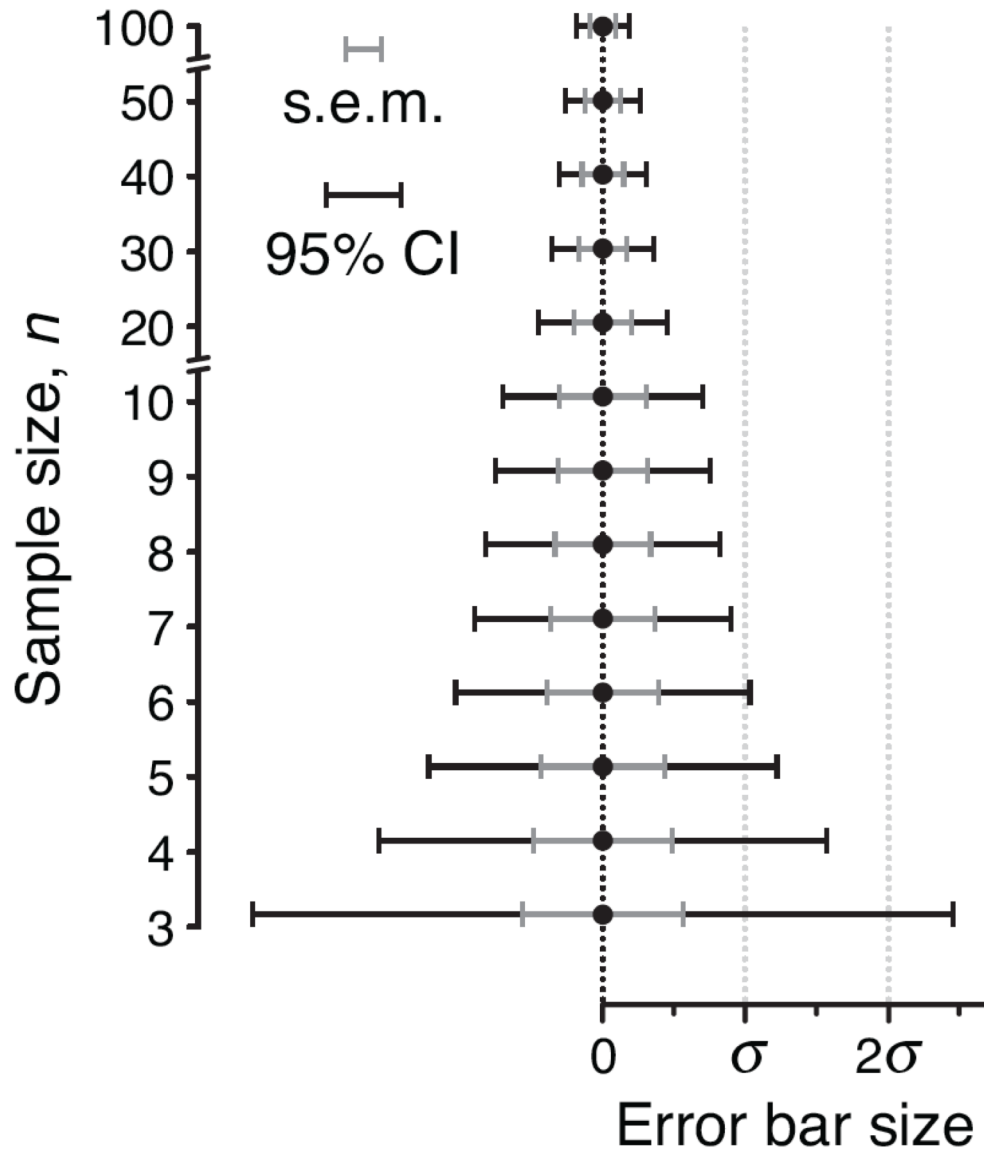
IC entre a amostra e  
a população



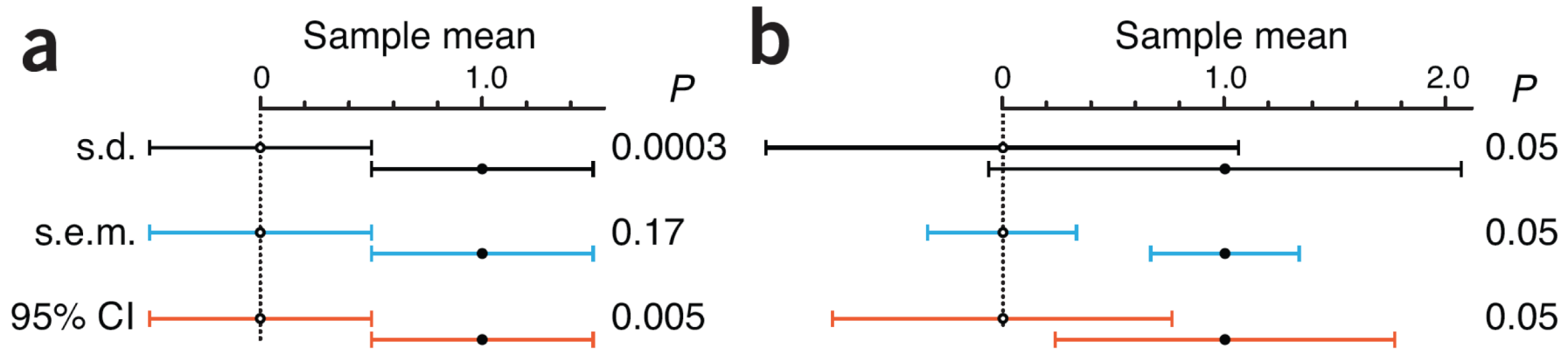
IC entre duas  
amostras

## Medidas de dispersão e diferenças entre as médias

EPM e intervalo de confiança são interrelacionados pela estatística t. Em grandes amostras o EPM pode ser interpretado como um IC com nível de confiança de 67%



# Relação entre diferenças entre médias e medidas de dispersão e significância




Quando é diferente?

Regra grosseira:

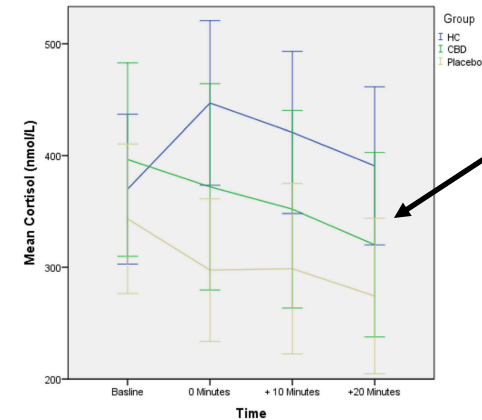
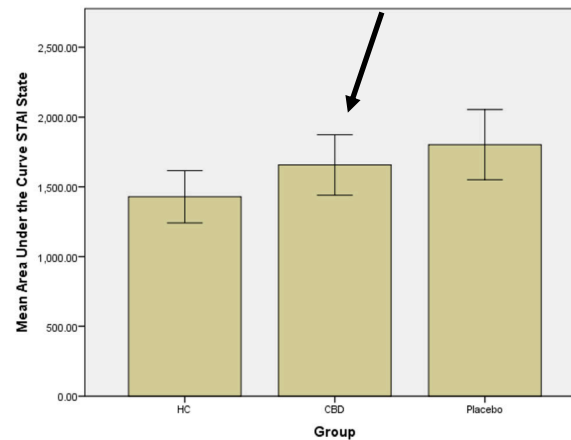
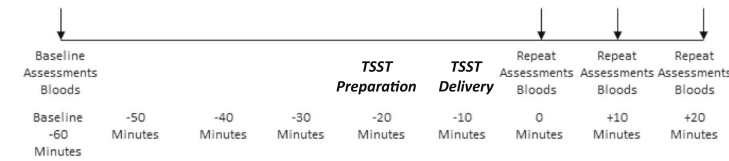
1. Se as barras do IC não se sobrepõem:  $p < 0,01$
2. Se a sobreposição for menor do que 50% das barras,  $p < 0,05$
3. Se sobreposição das barras for maior do que 50% do IC: não existe diferença significativa

# Effects of short-term cannabidiol treatment on response to social stress in subjects at clinical high risk of developing psychosis

Psychopharmacology 2020

E. Appiah-Kusi<sup>1</sup> • N. Petros<sup>1</sup> • R. Wilson<sup>1</sup> • M. Colizzi<sup>1,2</sup> • M. G. Bossong<sup>1,3</sup> • L. Valmaggia<sup>4,5</sup> • V. Mondelli<sup>5,6</sup> • P. McGuire<sup>1,5</sup> • S. Bhattacharyya<sup>1,5</sup> 

Sub-chronic (600 mg P.O. daily/7days)  
CBD treatment attenuated the  
anxiogenic and blunted cortisol  
responses to the Tryer social stress  
test in subjects with high risk of  
developing psychosis



## Impropriedades comuns em testes de hipótese

1. Falha em rejeitar a hipótese de nulidade não significa aceitá-la
2. Os resultados são altamente significantes  $p=0,000$
3. Os resultados são de significância fronteira (borderline)  $p=0,058$
4. O efeito foi maior no estudo A ( $p=0,013$ ) do que no estudo B ( $p=0,04$ ) (depende do  $n$ )