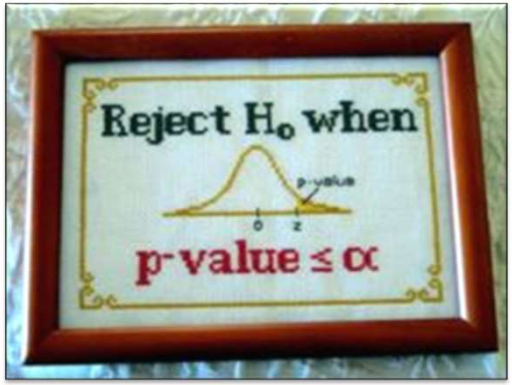


Testagem de Hipótese Nula

- A lógica da testagem de hipótese nula
- Teste estatístico z de uma média populacional e teste qui-quadrado de uma variância populacional: bilateral e unilateral
- Elementos da decisão estatística: nível de confiança, poder prospectivo, tamanho de efeito populacional e tamanho de amostra
- Valor-p, estatística de teste, tamanho de efeito amostral
- Métodos de planejamento do estudo

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

TESTE Z DE MÉDIA POPULACIONAL

Teste z bilateral para uma condição

Não rejeição de H_0

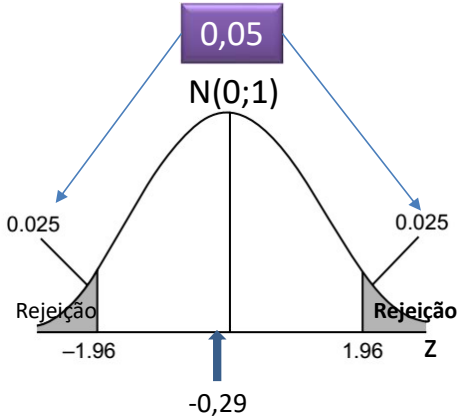
A distribuição da estatura do homem brasileiro de 19 anos de 2016 é normal com desvio-padrão σ = 7 cm.
Testar se a média populacional μ é igual a μ_0 = 177 cm hipotetizada.
Quatro participantes desse grupo de 2020 tiveram suas estaturas medidas, cujos valores em centímetro são: 169, 174, 175, 186.

Hipóteses

- H_0: μ = 177
- H_1: μ ≠ 177 (teste bilateral)

Estatísticas

- $\bar{X}$ = (175+186+169+174)/4 = 176
- EP = $\frac{\sigma}{\sqrt{n}} = 3{,}5$
- IC95(μ) = [176-1,96×3,5; 176+1,96×3,5] = [169,14; 182,86]
- Estatística de teste $z = \frac{\bar{X}-177}{EP} = \frac{176-177}{3{,}5} = -0{,}29$
- Estatística de tamanho de efeito $d = \frac{|\bar{X}-177|}{\sigma} = \frac{|176-177|}{7} = 0{,}14$



Decisão

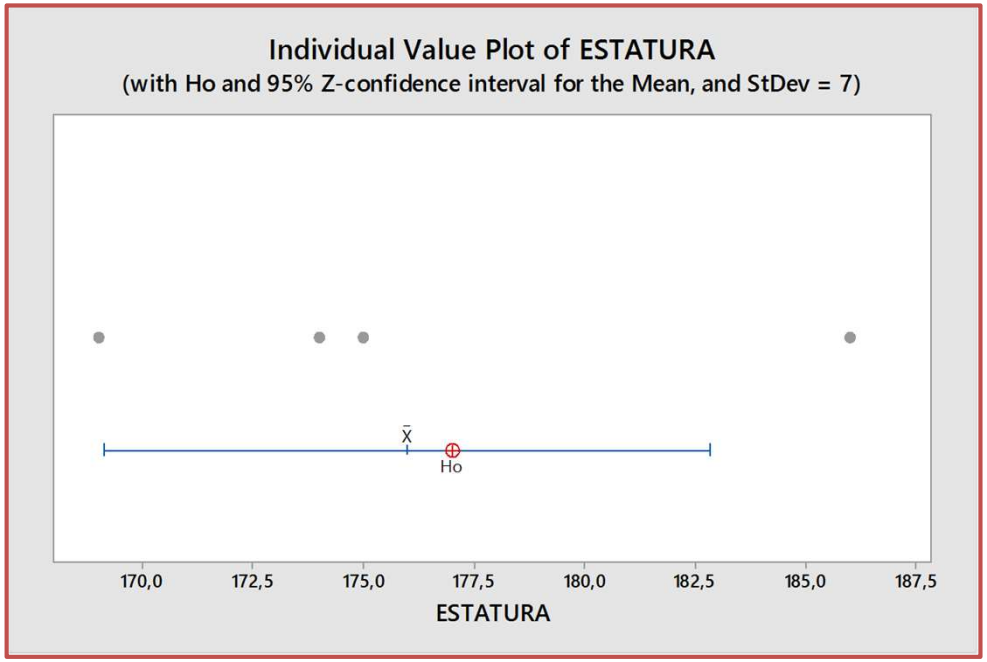
- *Critério do valor crítico*: Como |z| = 0,29 < 1,96, não rejeitar H_0 ou
- *Critério do IC95*: Como IC95 contém 177, não rejeitar H_0

One-Sample Z: ESTATURA

Test of μ = 177 vs ≠ 177

The assumed standard deviation = 7

Variable	N	Mean	StDev	SE Mean	95% CI	Z	P
ESTATURA	4	176,00	7,16	3,50	(169,14; 182,86)	-0,29	0,775



MINITAB 17

Teste z bilateral para uma condição

Teste

- Média populacional μ = 177 cm hipotetizada

Suposições

- Estatura tem distribuição normal
- Desvio-padrão σ = 7 cm conhecido
- n = 4 observações independentes: 169, 174, 175, 186
- Teste bilateral
- Nível de confiança adotado de 95% (ou nível de significância de 5%)

Hipóteses

- H_0: μ – 177 = 0 (ausência de efeito)
- H_1: μ – 177 ≠ 0

Estatísticas

- $\bar{X}$ = (169 + 174 + 175 + 186)/4 = 176
- EP = $\frac{\sigma}{\sqrt{n}} = 3{,}5$
- IC95(μ) = [169,14; 182,86]
- Estatística de teste $\mathrm{z} = \frac{\bar{X}-177}{EP} = -0{,}29$
- Estatística de tamanho de efeito $\mathrm{d} = \frac{|\bar{X}-177|}{\sigma} = 0{,}14$

Decisão

- Como |z| = 0,29 < 1,96, não rejeitar H_0 ou
- Como IC95 contém 177, não rejeitar H_0

Teste z bilateral para uma condição em R

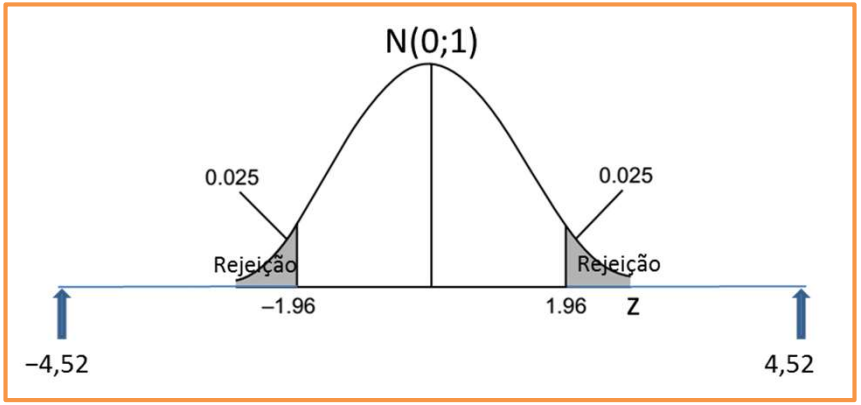
```
library(BSDA)
estatura <- c(169, 174, 175, 186)
BSDA::z.test(x=estatura, sigma.x=7, mu = 177,
             alternative="two.sided", conf.level=.95)
```

```
        One-sample z-Test

data:  estatura
z = -0.28571, p-value = 0.7751
alternative hypothesis: true mean is not equal to 177
95 percent confidence interval:
 169.1401 182.8599
sample estimates:
mean of x
      176
```

Teste z bilateral para uma condição

Rejeição de H_0

A distribuição da estatura do homem brasileiro de 19 anos de 2016 é normal com desvio-padrão σ = 7 cm.
Testar se a média populacional μ é igual a μ_0 = 177 cm hipotetizada pelo pesquisador.
1.000 participantes desse grupo de 2020 tiveram suas estaturas medidas, cujos valores em centímetro são:
169, 174, 175, 186, ...



Hipóteses

- H_0: μ = 177
- H_1: μ ≠ 177

Estatísticas

- $\bar{X}$ = (169 + 174 + 175 + 186 +...)/1000 = 176
- EP = $\frac{\sigma}{\sqrt{n}} = 0{,}22$
- IC95(μ) = [175,57; 176,43]
- Estatística de teste $z = \frac{\bar{X}-1}{EP} = -4{,}52$
- Estatística de tamanho de efeito $d = \frac{|\bar{X}-177|}{\sigma} = 0{,}14$ (muito pequeno)

Decisão

- Como |z| = 4,52 > 1,96, rejeitar H_0 ou
- Como IC95 não contém 177, rejeitar H_0

7

d de Cohen

Effect size	*d*	Reference
Very small	0.01	Sawilowsky, 2009
Small	0.20	Cohen, 1988
Medium	0.50	Cohen, 1988
Large	0.80	Cohen, 1988
Very large	1.20	Sawilowsky, 2009
Huge	2.0	Sawilowsky, 2009

Sawilowsky, S (2009) New effect size rules of thumb. *Journal of Modern Applied Statistical Methods,* 8(2): 467-74.

8

Teste z bilateral para uma condição em R

```
library(BSDA)
set.seed(3)
estatura <- rnorm(mean=176, sd=7, n=1000)
BSDA::z.test(x=estatura, sigma.x=7, mu = 177,
             alternative="two.sided", conf.level=.95)
```

```
        One-sample z-Test

data:  estatura
z = -4.3153, p-value = 1.594e-05
alternative hypothesis: true mean is not equal to 177
95 percent confidence interval:
 175.6109 176.4786
sample estimates:
mean of x
 176.0448
```

Teste z bilateral para uma condição

A distribuição da estatura do homem brasileiro de 19 anos de 2016 é normal com desvio-padrão σ = 7 cm.

Testar se a média populacional μ é igual a μ_0 = 177 cm hipotetizada.

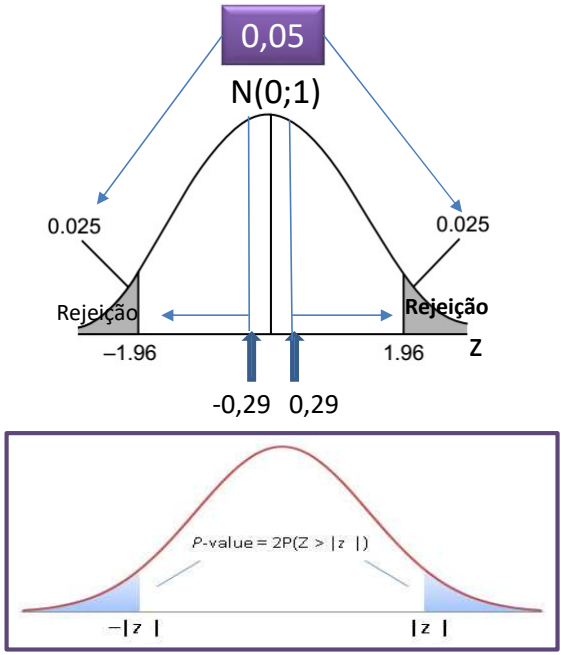
Quatro participantes desse grupo tiveram suas estaturas medidas, cujos valores em centímetro são: 169, 174, 175, 186.

Hipóteses

- H_0: μ = 177
- H_1: μ ≠ 177

Estatísticas

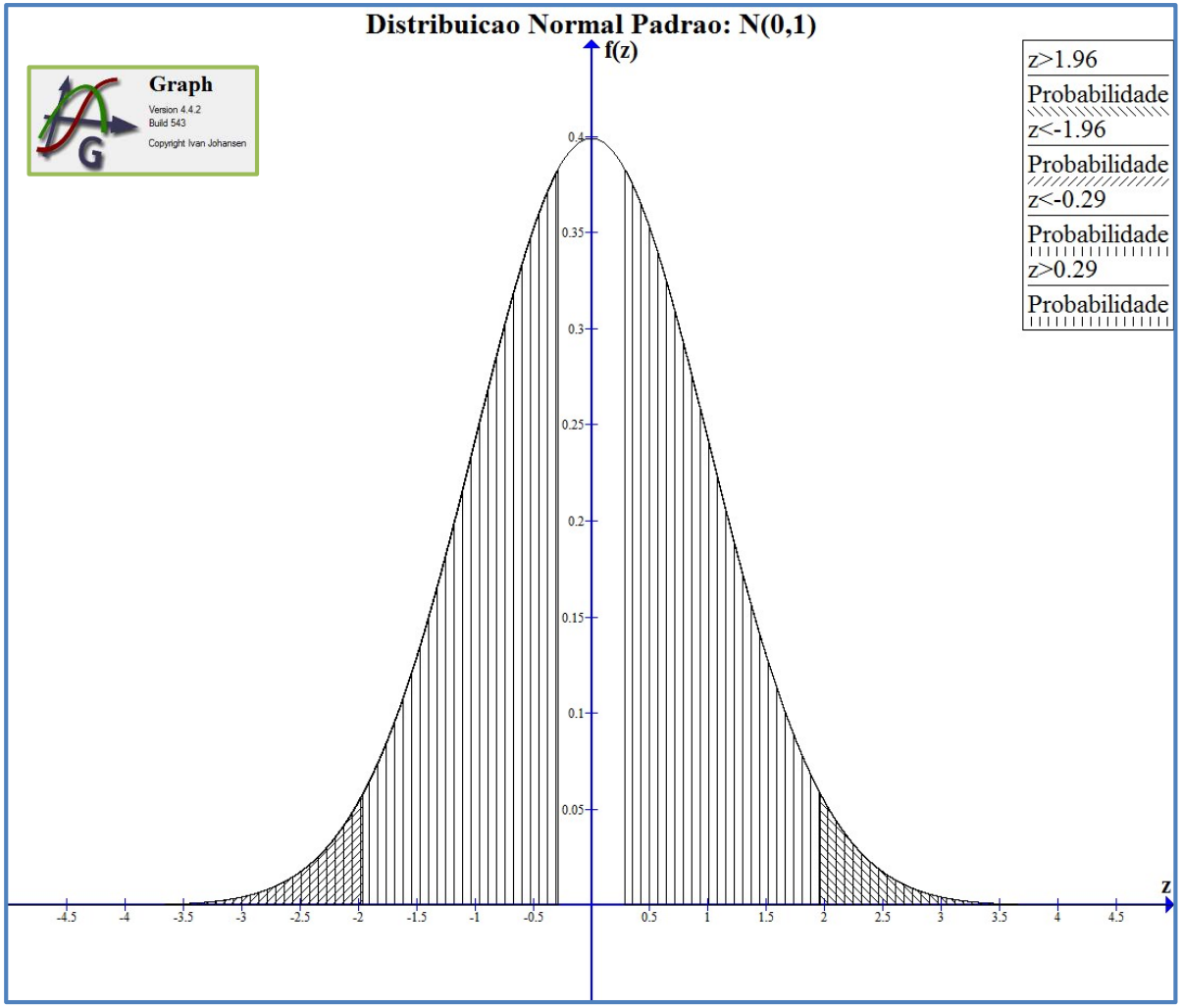
- $\bar{X}$ = (169 + 174 + 175 + 186)/4 = 176
- EP = $\frac{\sigma}{\sqrt{n}} = 3{,}5$
- IC95(μ) = [176-1,96×3,5; 176+1,96×3,5] = [169,14; 182,86]
- Estatística de teste $z = \frac{\bar{X}-177}{EP} = \frac{176-177}{3{,}5} = -0{,}29$
- Estatística de tamanho de efeito $d = \frac{|\bar{X}-177|}{\sigma} = 0{,}14$ (muito pequeno)



Decisão

- Como |z| = 0,29 < 1,96, não rejeitar H_0 ou
- Como IC95 contém 177, não rejeitar H_0 ou
- *Critério do valor-p*: Como o valor-p bilateral = 0,77 = 2*pnorm(-abs(-0.29)) é maior que 5%, não rejeitar H_0

10


Distribuicao Normal Padrao: N(0,1)
f(z)
Graph
Version 4.4.2
Build 543
Copyright Ivan Johansen
z>1.96
Probabilidade
z<-1.96
Probabilidade
z<-0.29
Probabilidade
z>0.29
Probabilidade
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
z
-4.5
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0.5
1
1.5
2
2.5
3
3.5
4
4.5

11

Teste z bilateral para uma condição em R

```
library(BSDA)
estatura <- c(169, 174, 175, 186)
BSDA::z.test(x=estatura, sigma.x=7, mu = 177,
             alternative="two.sided", conf.level=.95)
```

```
        One-sample z-Test

data:  estatura
z = -0.28571, p-value = 0.7751
alternative hypothesis: true mean is not equal to 177
95 percent confidence interval:
 169.1401 182.8599
sample estimates:
mean of x
      176
```


12

Valor-p

- O valor-p é a probabilidade de que a estatística de teste seja igual ou mais extrema que o valor observado na direção prevista pela hipótese alternativa (H_1), presumindo que a hipótese nula (H_0) é verdadeira.
- AGRESTI, A. & FINLAY, B. (2012) *Métodos estatísticos para as Ciências Sociais*. Porto Alegre: PENSO, p. 171.

13

Valor-p

$$\text{p-value} = \frac{\Gamma\left(\frac{(n+1)}{2}\right)}{\sqrt{n \cdot \pi}\,\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^{t} \left(1 + \frac{x^2}{n}\right)^{-\left(\frac{(n+1)}{2}\right)} dx$$

14

Teste z bilateral para uma condição

A distribuição da estatura do homem brasileiro de 19 anos de 2016 é normal com desvio-padrão σ = 7 cm.
Testar se a média populacional μ é igual a μ_0 = 177 cm hipotetizada.

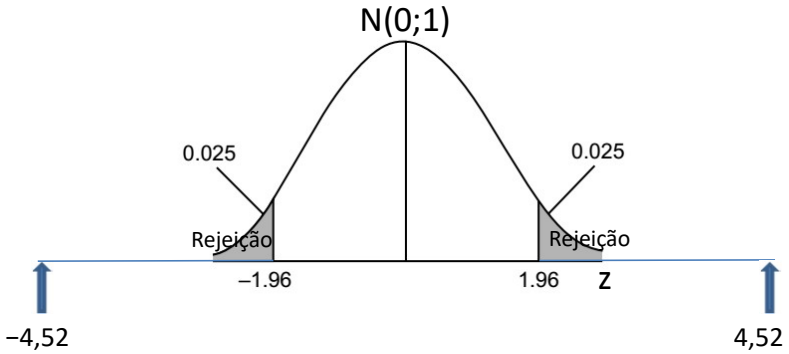
1.000 participantes desse grupo de 2020 tiveram suas estaturas medidas, cujos valores em centímetro são: 169, 174, 175, 186, ...

Hipóteses

- H_0: μ = 177
- H_1: μ ≠ 177

Estatísticas

- $\bar{X}$ = (169 + 174 + 175 + 186 +...)/1000 = 176
- EP = $\frac{\sigma}{\sqrt{n}} = 0{,}22$
- IC95(μ) = [175,57; 176,43]
- Estatística de teste $z = \frac{\bar{X}-177}{EP} = -4{,}52$
- Estatística de tamanho de efeito $d = \frac{|\bar{X}-177|}{\sigma} = 0{,}14$ (muito pequeno)



Decisão

- Como |z| = 4,52 > 1,96, rejeitar H_0 ou
- Como IC95 não contém 177, rejeitar H_0
- Como o valor-p bilateral = 6,18E-06 = 2*pnorm(-abs(-4.52)) é menor que 5%, rejeitar H_0

15

Teste z bilateral para uma condição

Teste

- Média populacional μ = 177 cm hipotetizada

Suposições

- Estatura tem distribuição normal (desnecessária devido ao TCL)
- Desvio-padrão σ = 7 cm conhecido
- n = 1.000 observações independentes
- Teste bilateral
- Nível de confiança de 95% (ou nível de significância adotado de 5%)

Hipóteses

- H_0: μ = 177 (ausência de efeito)
- H_1: μ ≠ 177

Estatísticas

- $\bar{X}$ = (169 + 174 + 175 + 186 +...)/1000 = 176
- EP = $\frac{\sigma}{\sqrt{n}} = 0{,}22$
- IC95(μ) = [175,57; 176,43]
- Estatística de teste $z = \frac{\bar{X}-177}{EP} = -4{,}52$
- Estatística de tamanho de efeito $d = \frac{|\bar{X}-177|}{\sigma} = 0{,}14$ (muito pequeno)

Decisão

- Critério do valor crítico da estatística de teste: Como |z| = 4,52 > 1,96, rejeitar H_0 ou
- Critério do IC95: Como IC95 não contém 177, rejeitar H_0
- Critério do valor-p: Como o valor-p bilateral = 6,18E-06 é menor que 5%, rejeitar H_0

16

Teste z bilateral para uma condição em R

```
library(BSDA)
set.seed(3)
estatura <- rnorm(mean=176, sd=7, n=1000)
BSDA::z.test(x=estatura, sigma.x=7, mu = 177,
             alternative="two.sided", conf.level=.95)
```

```
        One-sample z-Test

data:  estatura
z = -4.3153, p-value = 1.594e-05
alternative hypothesis: true mean is not equal to 177
95 percent confidence interval:
 175.6109 176.4786
sample estimates:
mean of x
 176.0448
```

17

Erro (Flutuação) Amostral

- Devido ao erro amostral, as amostras que utilizamos podem não refletir de forma fiel a população de onde foram retiradas.

18

Conceito do valor-p

- Constitui um dos problemas enfrentados quando conduzimos uma pesquisa o fato de não sabermos qual é o padrão existente na população de interesse.
- De fato, o motivo de realizarmos a pesquisa é, em primeiro lugar, determinar esse padrão.
- Você precisa estar ciente de que, algumas vezes, devido ao erro amostral, obteremos padrões nas amostras que não refletem de forma acurada a população de onde as amostras foram retiradas.
- Assim, precisamos de um algum meio para avaliar a probabilidade de que a amostra selecionada seja um retrato fiel da população.
- Os testes estatísticos nos auxiliam nesta decisão, mas isso ocorre de uma forma não de todo intuitiva.

Problema: População -> Amostra

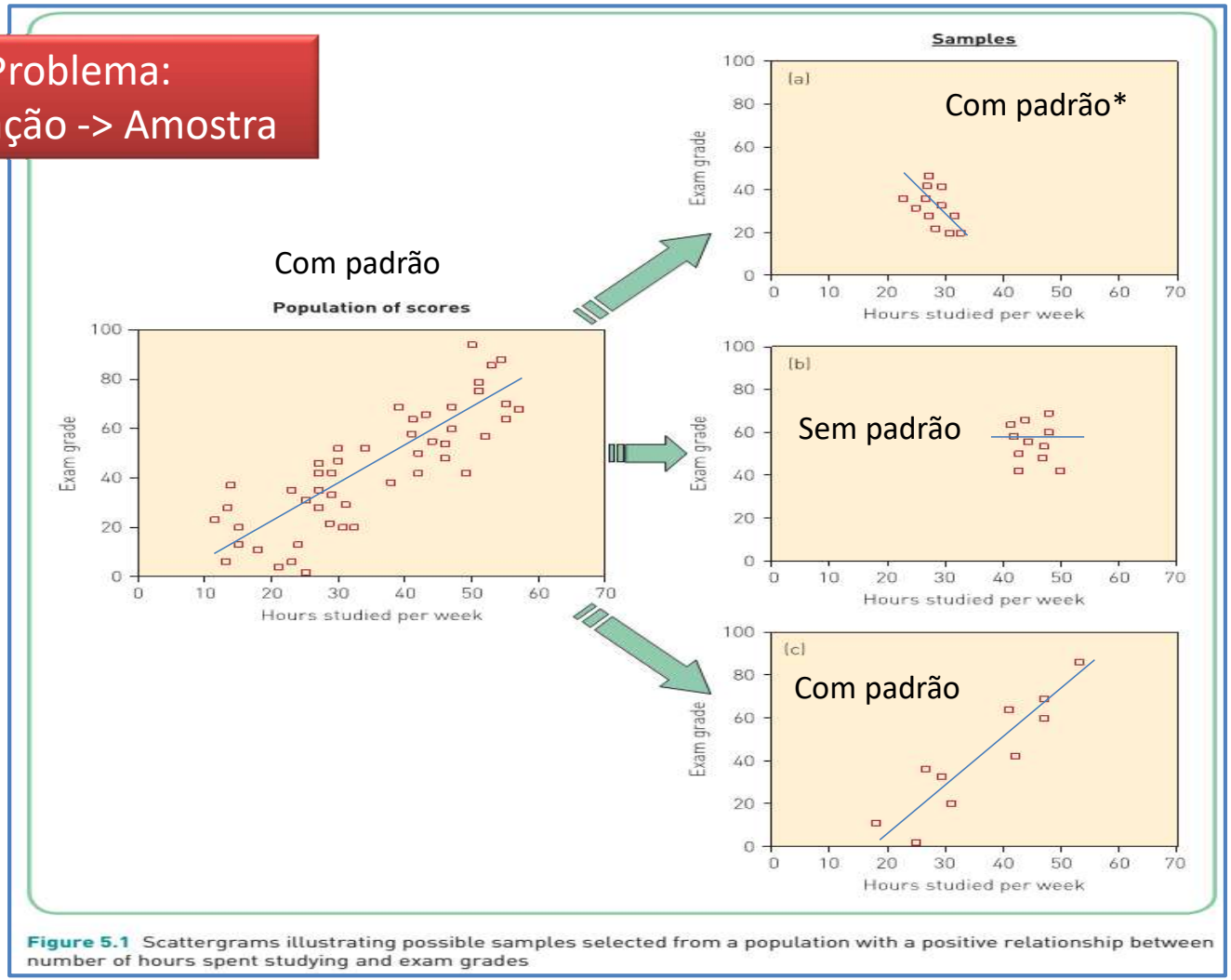


Figure 5.1 Scattergrams illustrating possible samples selected from a population with a positive relationship between number of hours spent studying and exam grades

20

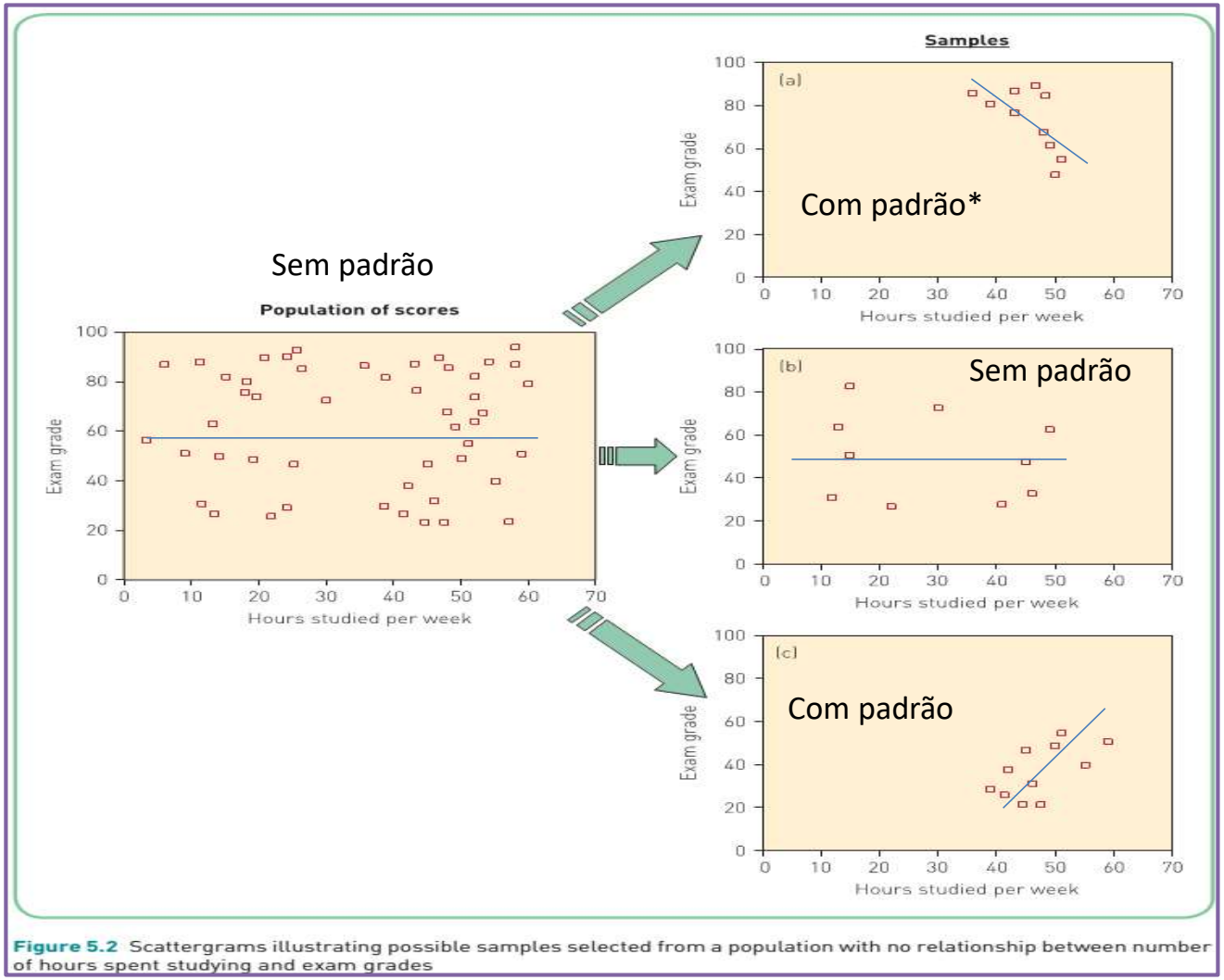


Figure 5.2 Scattergrams illustrating possible samples selected from a population with no relationship between number of hours spent studying and exam grades

21

Problema invertido:
Amostra -> População

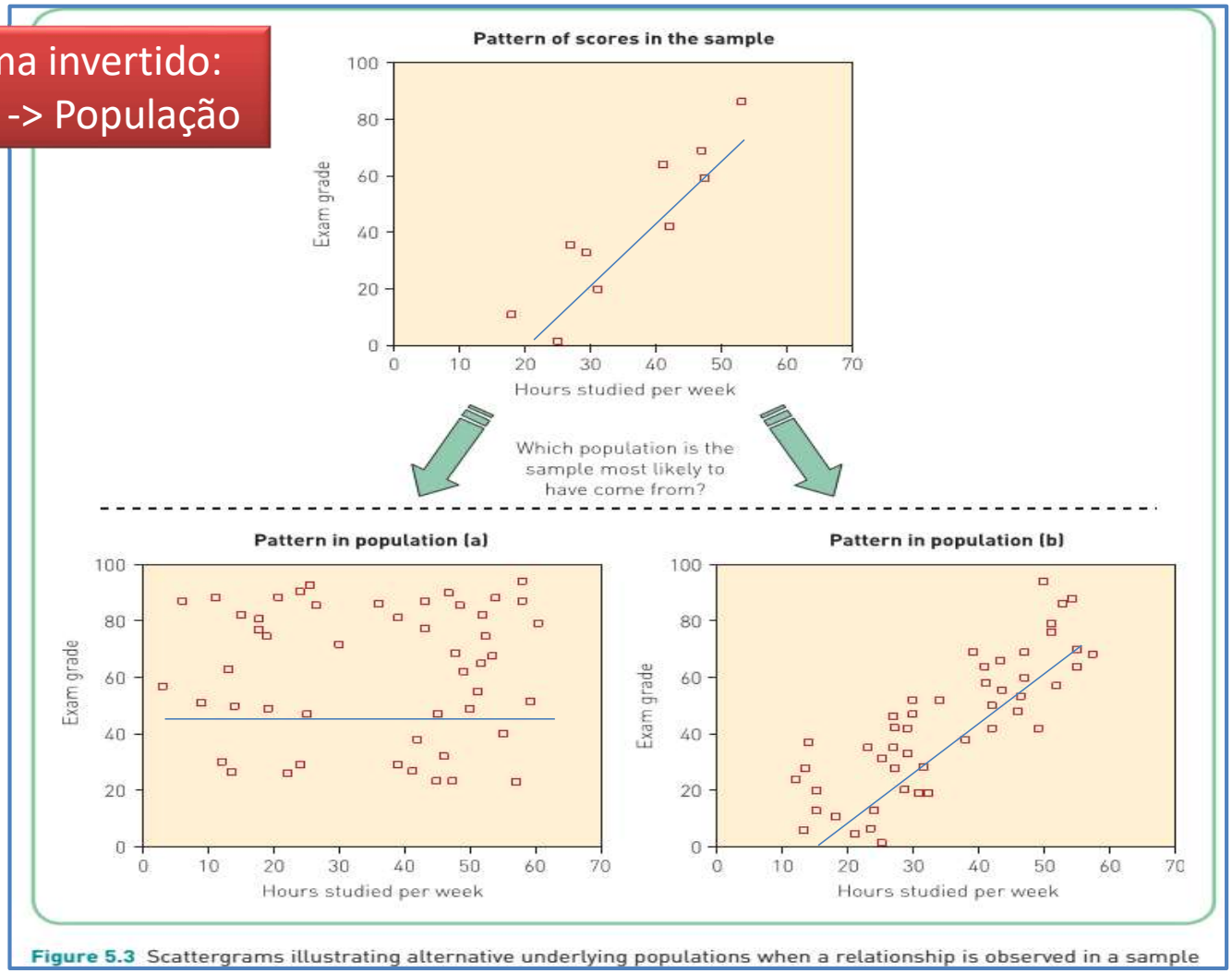


Figure 5.3 Scattergrams illustrating alternative underlying populations when a relationship is observed in a sample

22

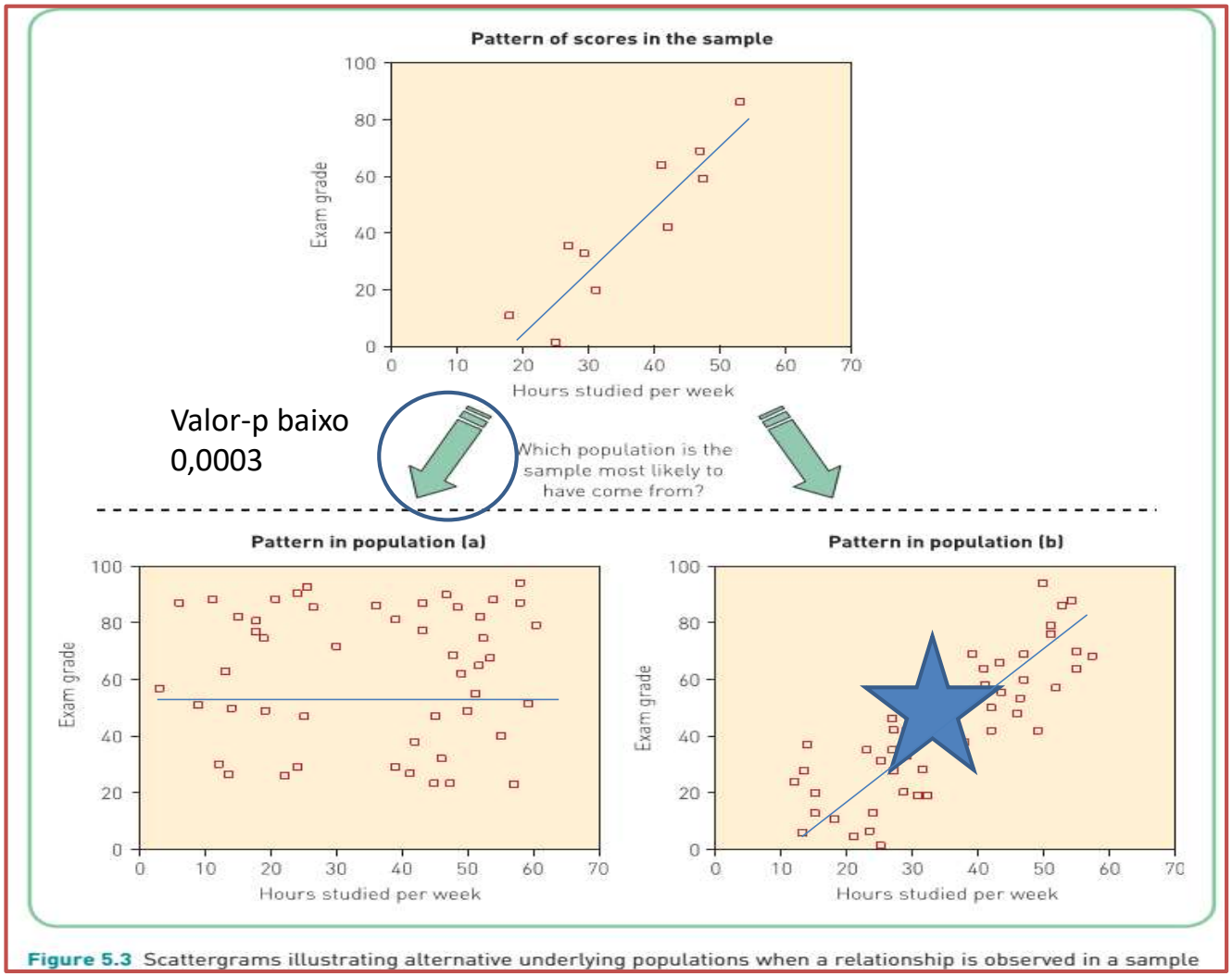


Figure 5.3 Scattergrams illustrating alternative underlying populations when a relationship is observed in a sample

23

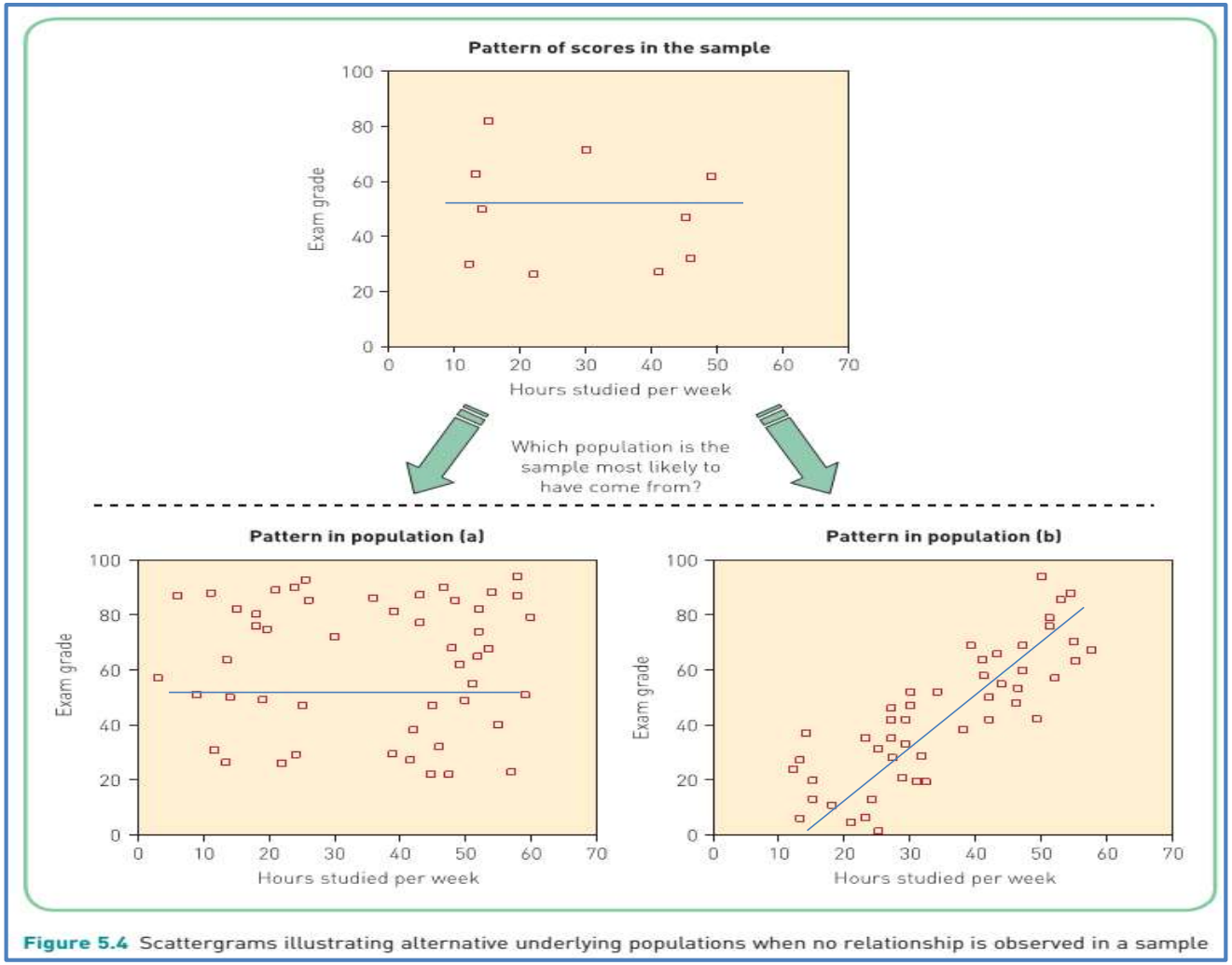


Figure 5.4 Scattergrams illustrating alternative underlying populations when no relationship is observed in a sample

24

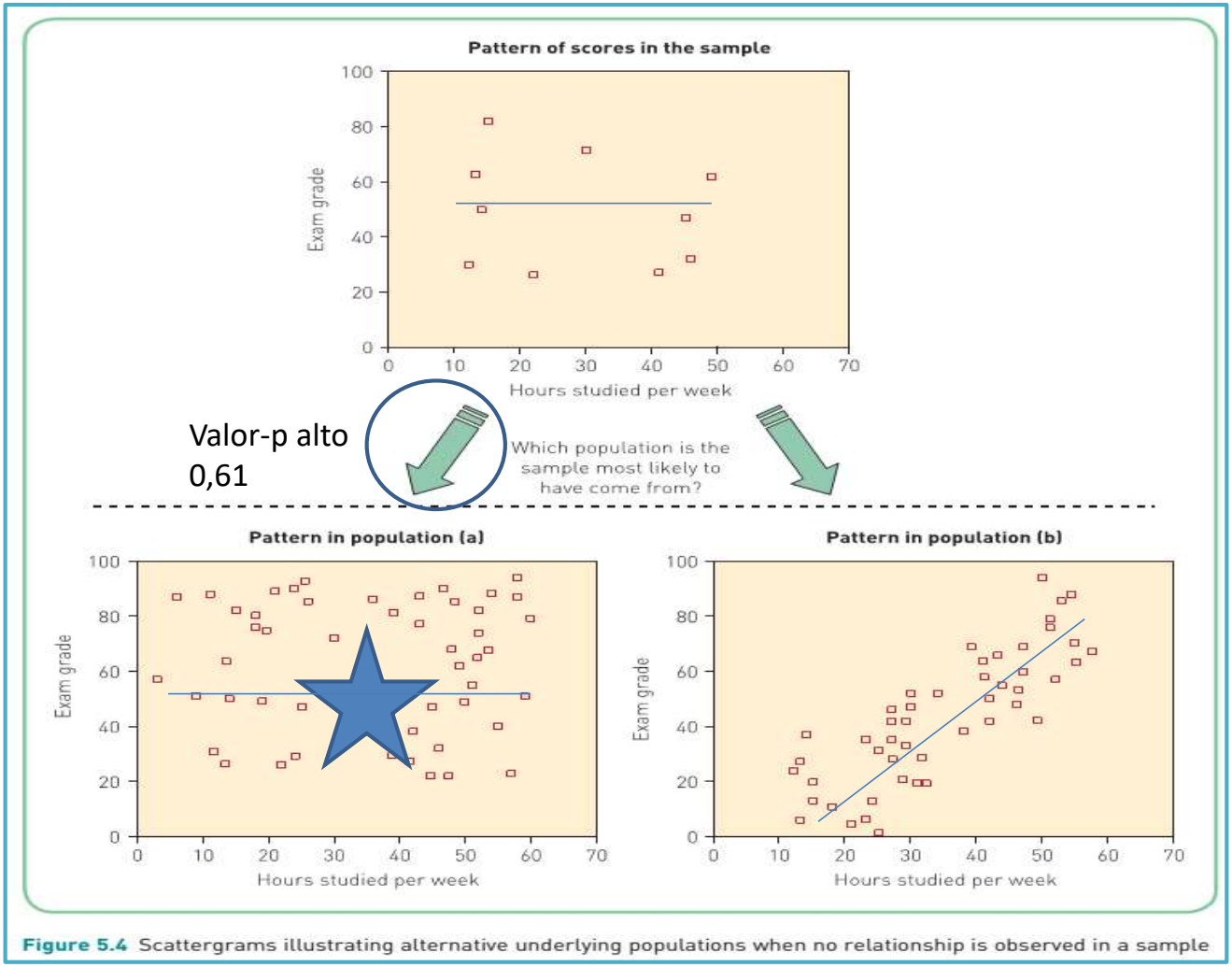


Figure 5.4 Scattergrams illustrating alternative underlying populations when no relationship is observed in a sample

25

Efeito

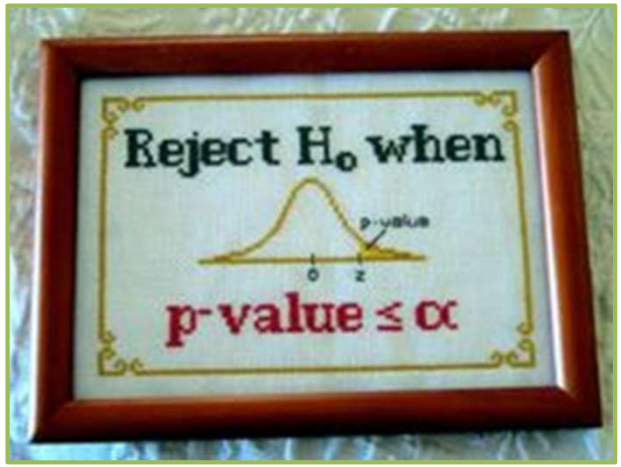
- Correlação entre variáveis
- Diferença entre condições

26

Hipóteses nula e alternativa

- Definição de *hipótese nula* ou H_0
  - A hipótese nula sempre declara que não existe efeito na população.
- Definição de *hipótese de pesquisa* ou alternativa ou H_1 ou H_a
  - A hipótese de pesquisa é a nossa previsão de como condições específicas podem estar relacionadas.

Decisão Estatística vs. Estado da Natureza



Possible Outcomes of the Decision-Making Process

Researcher's Decision	True State of the World: H_0 True	True State of the World: H_0 False
Reject H_0	Type I error $p = \alpha$ = significance level	Correct Decision $p = 1 - \beta$ = Power
Fail to Reject H_0	Correct Decision $p = 1 - \alpha$ = confidence level	Type II error $p = \beta$

28

Teste de Hipótese Nula & Intervalo de Confiança

- Rejeitar a hipótese nula ao nível de significância adotado, alfa, se o valor do parâmetro conjecturado na hipótese nula não pertencer ao intervalo de confiança de 1 – alfa.
- Os critérios de valor-p e do IC são equivalentes.

29

Críticas contra os testes de hipótese nula

- A testagem da hipótese nula é a abordagem dominante na Psicologia e Medicina
- Apesar das críticas à testagem da hipótese nula, isso não significa que tal abordagem deve ser abandonada completamente
- Ao invés disso, devemos ter um entendimento completo de seu significado para podermos nos beneficiar desta tecnologia da decisão
- Além do valor-p, é importante usar o intervalo de confiança e de tamanho de efeito

30

Ecology, 95(3), 2014, pp. 645–651

Recurring controversies about *P* values and confidence intervals revisited

ARIS SPANOS[1]

Department of Economics, Virginia Tech, Blacksburg, Virginia 24061 USA

P *value and the large* n *problem*

A crucial weakness of both the *P* value and the N-P error probabilities is the so-called large *n* problem: there is always a large enough sample size *n* for which any simple null hypothesis. H_0: $\mu = \mu_0$ will be rejected by a frequentist α-significance level test; see Lindley (1957).

31

The large n constitutes an example of a broader problem known as the *fallacy of rejection*: (mis)interpreting reject H_0 (evidence *against* H_0) as evidence *for* a particular H_1; this can arise when a test has very high power, e.g., large n. A number of attempts have been made to alleviate the large n problem, including rules of thumb for decreasing α as n increases; see Lehmann (1986). Due to the trade-off between the Type I and II error probabilities, however, any attempt to ameliorate the problem renders the inference susceptible to the reverse fallacy known as the *fallacy of acceptance*: (mis)interpreting accept H_0 (*no* evidence *against* H_0) as evidence *for* H_0; this can easily arise when a test has very low power; e.g., α is tiny or n is too small.

These fallacies are routinely committed by practitioners in many applied fields. After numerous unsuccessful attempts, Mayo (1996) provided a reasoned answers to these fallacies in the form of a post-data severity assessment.

32

Significância prática

- Mesmo efeitos muito pequenos poderão apresentar significância estatística quando o tamanho da amostra for bem grande
- Para determinar a significância prática a melhor abordagem consiste em obter uma medida do tamanho do efeito, sendo que essa medida não depende do tamanho da amostra
  - E.g.: a correlação de Pearson amostral mede a intensidade da associação linear entre duas variáveis quantitativas e não depende do tamanho da amostra

33

Interpretação errônea do valor-p

- Muitos pesquisadores sem experiência em estatística (e mesmo aqueles com alguma) equiparam o valor-p com o verdadeira tamanho do efeito, i.e., quanto menor o valor-p, mais forte seria, por exemplo, o relacionamento entre duas variáveis; talvez, de fato, quanto mais forte o relacionamento, mais baixo o valor-p, mas não significa que isso necessariamente ocorrerá
- **O valor-p não é a probabilidade de que a hipótese nula seja verdadeira**; de fato, não sabemos qual é a probabilidade de que a hipótese nula seja verdadeira
- 1 – p não é a probabilidade de que a hipótese alternativa seja verdadeira; de fato, não sabemos qual é a probabilidade de que a hipótese alternativa seja verdadeira

JAMA Cardiology | Special Communication

Understanding the Role of *P* Values and Hypothesis Tests in Clinical Research

Daniel B. Mark, MD, MPH; Kerry L. Lee, PhD; Frank E. Harrell Jr, PhD

JAMA Cardiol. doi:10.1001/jamacardio.2016.3312
Published online October 12, 2016.

P values and hypothesis testing methods are frequently misused in clinical research. Much of this misuse appears to be owing to the widespread, mistaken belief that they provide simple, reliable, and objective triage tools for separating the true and important from the untrue or unimportant. The primary focus in interpreting therapeutic clinical research data should be on the treatment ("oomph") effect, a metaphorical force that moves patients given an effective treatment to a different clinical state relative to their control counterparts. This effect is assessed using 2 complementary types of statistical measures calculated from the data, namely, effect magnitude or size and precision of the effect size. In a randomized trial, effect size is often summarized using constructs, such as odds ratios, hazard ratios, relative risks, or adverse event rate differences. How large a treatment effect has to be to be consequential is a matter for clinical judgment. The precision of the effect size (conceptually related to the amount of spread in the data) is usually addressed with confidence intervals. *P* values (significance tests) were first proposed as an informal heuristic to help assess how "unexpected" the observed effect size was if the true state of nature was no effect or no difference. Hypothesis testing was a modification of the significance test approach that envisioned controlling the false-positive rate of study results over many (hypothetical) repetitions of the experiment of interest. Both can be helpful but, by themselves, provide only a tunnel vision perspective on study results that ignores the clinical effects the study was conducted to measure.

Table 2. Common Misconceptions About *P* Value

Misconception	Comment
P value equals the probability that the null hypothesis is true.	*P* value is computed by assuming the null hypothesis is true.
P value equals the probability that the observed effect is due to "the play of chance."	*P* value is defined as the probability of a difference (effect) as large as that observed or larger if the null hypothesis is true. Even if the difference observed is consistent with a simple chance mechanism, other more complex explanations are also possible, and nothing in *P* value calculation allows one to conclude that this is the best or most likely explanation for the observed differences.
P value ≤.05 means the null hypothesis is false. *P* value >.05 means the null hypothesis is true.	*P* value is computed assuming the null hypothesis is true. It is not the probability that the null hypothesis is either true or false.
P value ≤.05 identifies a clinically or scientifically important difference (effect). *P* value >.05 rules out a clinically or scientifically important difference (effect).	Clinical or scientific importance of study results is a judgment integrating multiple elements, including effect size (expected and observed), precision of estimate of effect size, and knowledge of prior relevant research. At best, *P* value has a minor role in shaping this judgment.
A small *P* value indicates study results are reliable and likely to replicate.	*P* value provides no information about whether a given study result can be reproduced in a second, replication experiment. There are many other factors that must be considered in judging the reliability of study results. Understanding what works in medicine is a process and not the product of any single experiment.

35

A Dirty Dozen: Twelve *P*-Value Misconceptions

Steven Goodman

The *P* value is a measure of statistical evidence that appears in virtually all medical research papers. Its interpretation is made extraordinarily difficult because it is not part of any formal system of statistical inference. As a result, the *P* value's inferential meaning is widely and often wildly misconstrued, a fact that has been pointed out in innumerable papers and books appearing since at least the 1940s. This commentary reviews a dozen of these common misinterpretations and explains why each is wrong. It also reviews the possible consequences of these improper understandings or representations of its meaning. Finally, it contrasts the *P* value with its Bayesian counterpart, the Bayes' factor, which has virtually all of the desirable properties of an evidential measure that the *P* value lacks, most notably interpretability. The most serious consequence of this array of *P*-value misconceptions is the false belief that the probability of a conclusion being in error can be calculated from the data in a single experiment without reference to external evidence or the plausibility of the underlying mechanism.

Semin Hematol 45:135-140

Table 1 Twelve *P*-Value Misconceptions

1	*If P = .05, the null hypothesis has only a 5% chance of being true.*
2	*A nonsignificant difference (eg, P ≥.05) means there is no difference between groups.*
3	*A statistically significant finding is clinically important.*
4	*Studies with P values on opposite sides of .05 are conflicting.*
5	*Studies with the same P value provide the same evidence against the null hypothesis.*
6	*P = .05 means that we have observed data that would occur only 5% of the time under the null hypothesis.*
7	*P = .05 and P ≤.05 mean the same thing.*
8	*P values are properly written as inequalities (eg, "P ≤.02" when P = .015)*
9	*P = .05 means that if you reject the null hypothesis, the probability of a type I error is only 5%.*
10	*With a P = .05 threshold for significance, the chance of a type I error will be 5%.*
11	*You should use a one-sided P value when you don't care about a result in one direction, or a difference in that direction is impossible.*
12	*A scientific conclusion or treatment policy should be based on whether or not the P value is significant.*

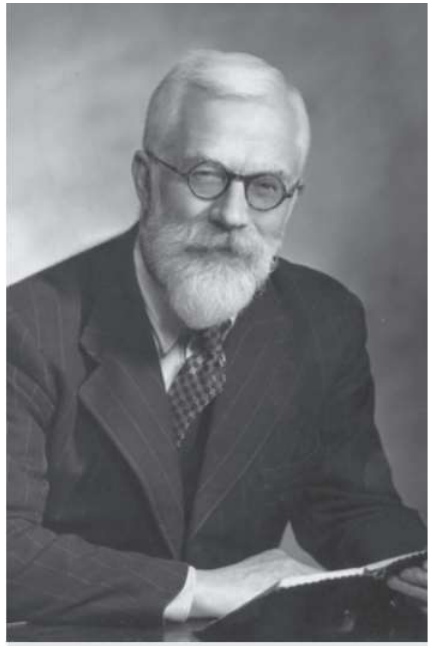
36

Replicação

- A replicação é uma das pedras angulares da ciência
- Se você observa um fenômeno uma vez, então pode ter sido por acaso; se o observa duas, três ou mais vezes, pode estar começando a aprender algo sobre o fenômeno estudado
- Se o seu estudo foi o primeiro neste assunto, é sensato que você trate os resultados com certo grau de cautela

37

Por que estabelecer α = 5%?



The great R. A. Fisher wrote in 1926: "Personally, the writer prefers to set a low standard of significance at the 5 percent point, and ignore entirely all results which fail to reach that level. A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance." (Quoted in Moore, 1979 edition).

It is the fate of a guru that what he sees as a convenient but arbitrary option is taken by followers as written in stone. But it is a philosophy that must be abandoned.

Moore, D. S. (1997) Statistics: Concepts and Controversies, 4th edition. New York: Freeman

38

Testes unilaterais e bilaterais

- Quando a direção do relacionamento ou da diferença (efeito) é especificada, então o teste é unilateral/unicaudal; caso contrário, é bilateral/bicaudal.
- Em geral (mas nem sempre), se você tiver obtido um valor-p para um teste bilateral e quiser saber o valor mínimo correspondente para o teste unilateral, então: $p_{uni} \geq p_{bi}/2$
- Observe que o que deve ser dobrado ou dividido por 2 não é a estatística de teste (e.g.: z).

39

Peso do rato recém-nascido e etilismo

- Um pesquisador deseja investigar se o etilismo durante a prenhez diminui o peso dos ratos recém-nascidos (PRRN).
- PRRN tem distribuição normal com média e o desvio-padrão populacionais sem etilismo na prenhez iguais, respectivamente, a 20 g e 4 g. Essas estimativas foram obtidas de uma amostra de 20 PRRNs independentes.
- 50 fêmeas foram emprenhadas num experimento no qual elas receberam doses diárias de álcool. Um rato recém-nascido de cada fêmea foi selecionado aleatoriamente para compor a amostra de 50 PRRNs independentes.
- A média do PRRN das fêmeas que receberam doses diárias de álcool no experimento é 18 g.

40

Teste z unilateral (*less*) para uma condição

Teste
- Média populacional do PRRN hipotetizada: μ = 20 g

Suposições
- PRRN tem distribuição normal (desnecessária, pois n > 30)
- Desvio-padrão σ = 4 g conhecido
- n = 50 observações independentes (TCL)
- Teste bilateral
- Nível de confiança de 95% (ou nível de significância adotado de 5%)

Hipóteses
- H_0: μ = 20
- H_1: μ < 20 (baixo PRRN)

Estatísticas
- $\bar{X}$ = 18
- EP = $\frac{\sigma}{\sqrt{n}} = 0{,}57$
- IC95(μ) = [0; 18,93]
- Estatística de teste $z = \frac{\bar{X}-2}{EP} = -3{,}54$
- Estatística de tamanho de efeito $d = \frac{|\bar{X}-177|}{\sigma} = 0{,}5$ (*medium* (intermediário))

Decisão
- Critério do valor crítico da estatística de teste: Como |z| = 3,54 > 1,96, rejeitar H_0 ou
- Critério do IC95: Como IC95 não contém 20, rejeitar H_0
- Critério do valor-p: Como o valor-p unilateral (*less*) = 2,09E-3 é menor que 5%, rejeitar H_0

41

$H_0: \mu \geq \mu_0$ é **equivalente** a $H_0: \mu = \mu_0$

vs.

$H_1: \mu < \mu_0$

Demonstração em
GATÁS, RR (1978) *Elementos de Probabilidade e Inferência*. SP: Atlas, p. 220-3.

42

```
pnorm(q=18, mean=20, sd=4/sqrt(50))
[1] 0.000203476
```

TCL (n > 30)

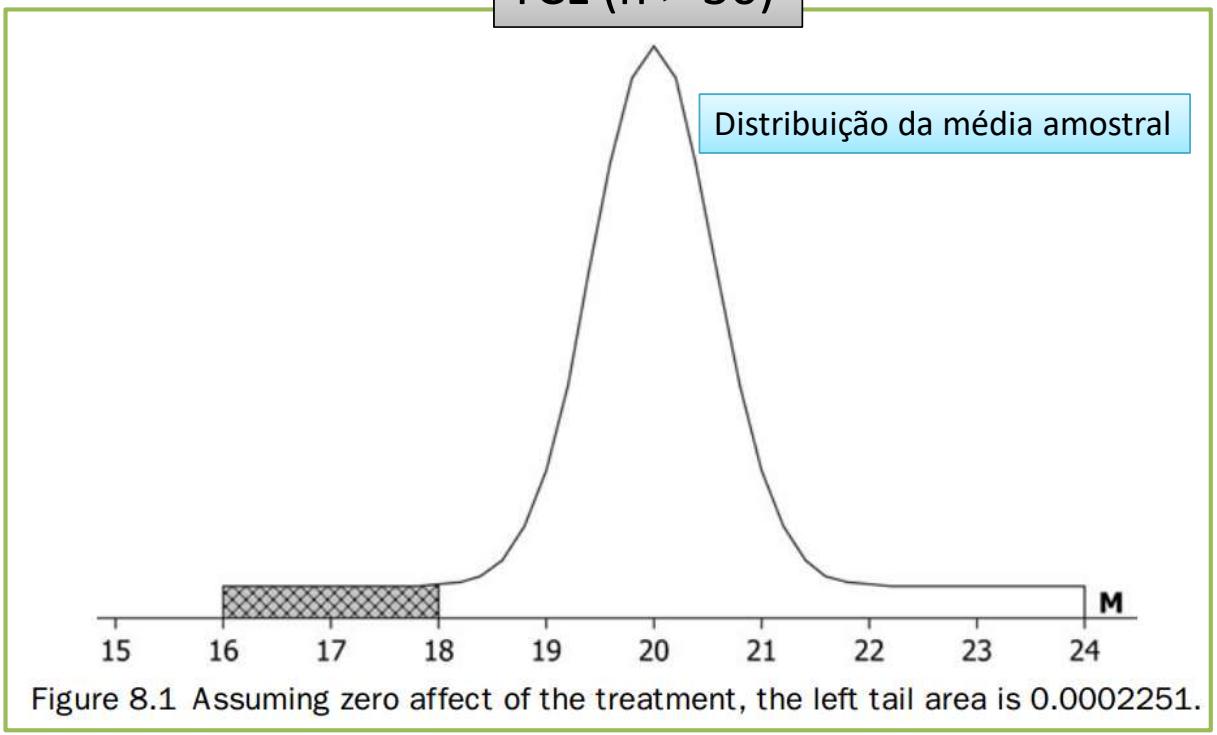
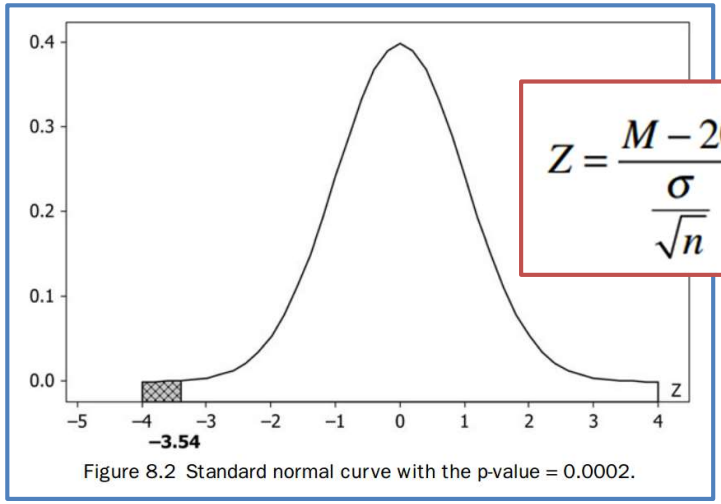


Figure 8.1 Assuming zero affect of the treatment, the left tail area is 0.0002251.

43



$$Z = \frac{M - 20}{\frac{\sigma}{\sqrt{n}}} = \frac{18 - 20}{\frac{4}{\sqrt{50}}} = -3.54$$

Figure 8.2 Standard normal curve with the p-value = 0.0002.

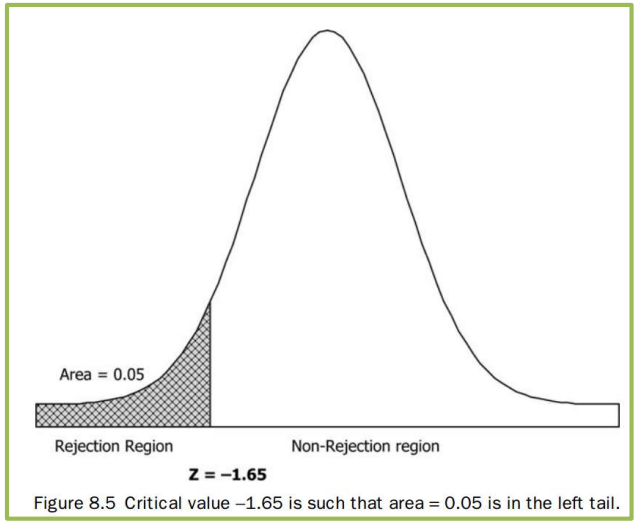


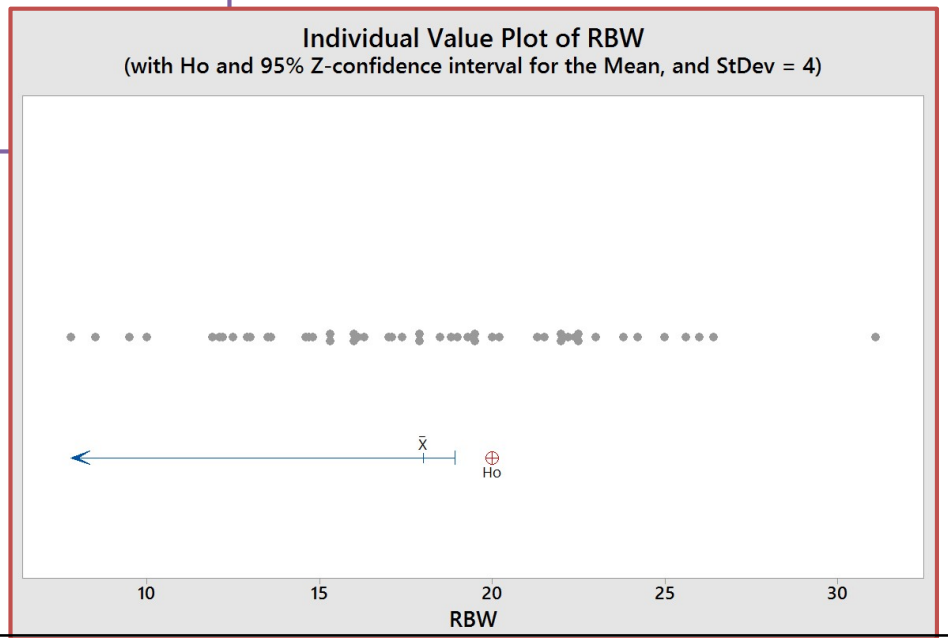
Figure 8.5 Critical value –1.65 is such that area = 0.05 is in the left tail.

44

ID	RBW
1	12,2
2	7,8
3	16,1
4	22,4
5	22,2
6	19,5
7	14,7
8	22,0
9	9,5
10	12,1
11	17,9
12	25,0
13	19,0
14	18,8
15	16,0
16	16,3
17	31,1
18	15,3
19	12,9
20	17,1
21	20,0
22	16,0
23	11,9
24	17,0
25	12,5
26	8,5
27	13,5
28	20,2
29	22,5
30	25,6
31	23,8
32	21,5
33	19,5
34	14,6
35	14,8
36	10,0
37	21,3
38	13,0
39	17,9
40	19,3
41	23,0
42	22,0
43	18,5
44	26,0
45	24,2
46	26,4
47	13,6
48	15,3
49	17,4
50	22,5

```
library(readxl)
library(BSDA)
Dados <- readxl::read_excel("Table 8.2 RawBirthWeight.xls")
BSDA::z.test(x=Dados$RBW, sigma.x=4, mu = 20,
             alternative="less", conf.level=.95)
```

```
        One-sample z-Test

data:  Dados$RBW
z = -3.5285, p-value = 0.000209
alternative hypothesis: true mean is less than 20
95 percent confidence interval:
       NA 18.93447
sample estimates:
mean of x
   18.004
```

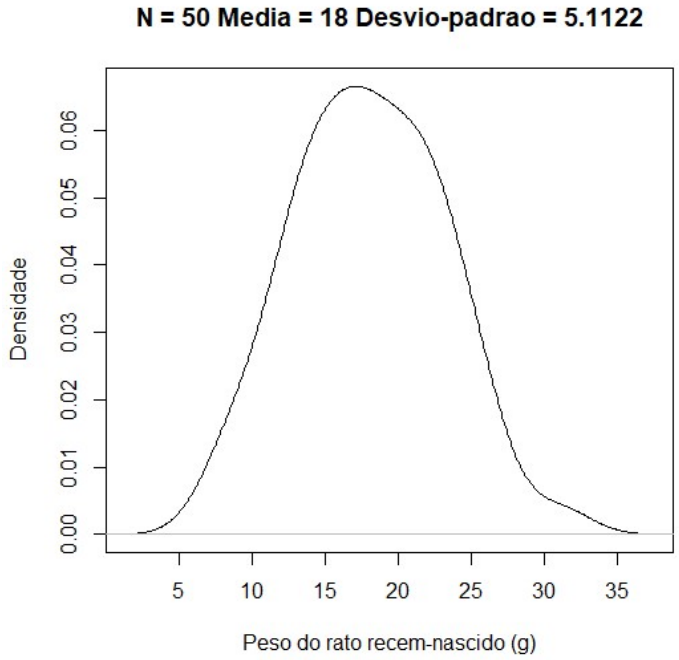


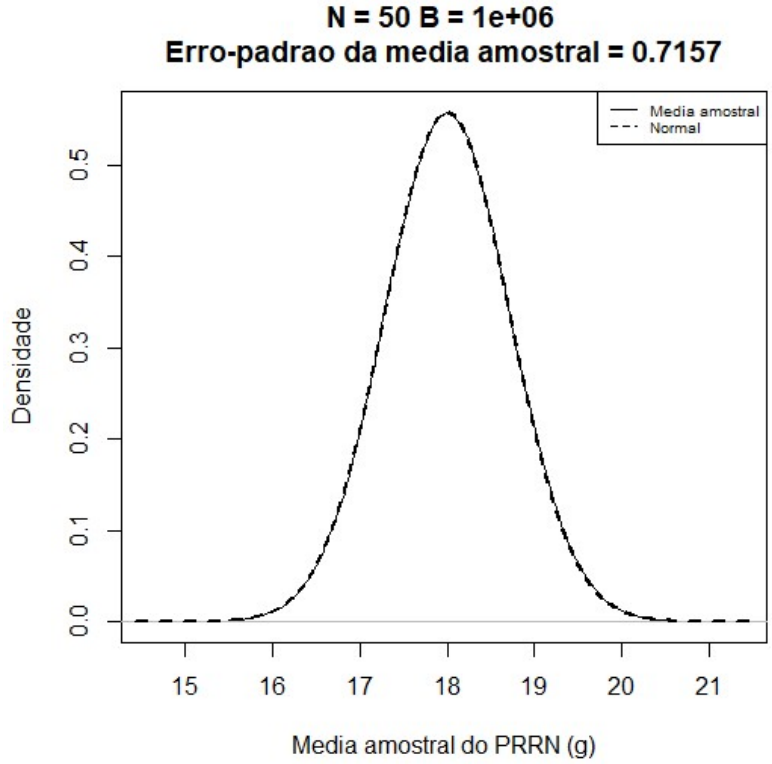
45

Reamostragem (*bootstrapping*) da média amostral em R

```
library(readxl)
library(BSDA)
B <- 1e6; alfa <- 0.05; set.seed(123)
Dados <- readxl::read_excel("Table 8.2 RawBirthWeight.xls")
PRRN <- as.matrix(Dados[,2])
N <- nrow(PRRN)
plot(density(PRRN, na.rm=TRUE),
     main=paste("N =", N,
                "Media =", round(mean(PRRN,na.rm=TRUE),1),
                "Desvio-padrao =",
                round(sd(PRRN, na.rm=TRUE), 4)),
     xlab="Peso do rato recem-nascido (g)", ylab="Densidade")
PRRN.media.boot <- replicate(B, mean(sample(PRRN, replace=TRUE)))
print(mean(PRRN, na.rm=TRUE))
print(mean(PRRN.media.boot, na.rm=TRUE))
round(quantile(PRRN.media.boot, probs=1-alfa),2)
print(BSDA::z.test(x=Dados$RBW, sigma.x=4, mu = 20,
                        alternative="less", conf.level=.95))
plot(density(PRRN.media.boot, na.rm=TRUE),
     main=paste("N =", N, "B =", B,
                "\nErro-padrao da media amostral =",
                round(sd(PRRN.media.boot, na.rm=TRUE), 4)),
     xlab="Media amostral do PRRN (g)", ylab="Densidade")
mi <- mean(PRRN.media.boot, na.rm=TRUE)
EP <- sd(PRRN.media.boot, na.rm=TRUE)
x <- seq(from=mi-5*EP, to=mi+5*EP, by=1e-5)
y <- dnorm(x, mean=mi, sd=EP)
lines(x,y,lwd=2,lty=2)
legend("topright", c("Media amostral","Normal"), lty=1:2, cex=.6)
```

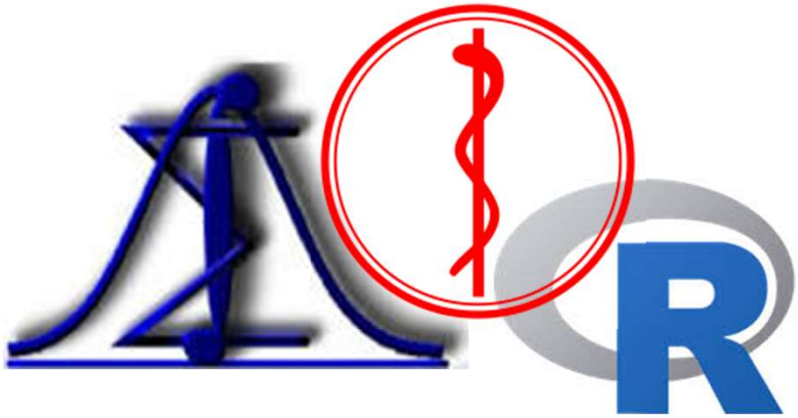
Reamostragem (*bootstrapping*) da média amostral em R






```
95 percent confidence interval:
NA 18.93447
```

LI95%	LS95%
0	19.19

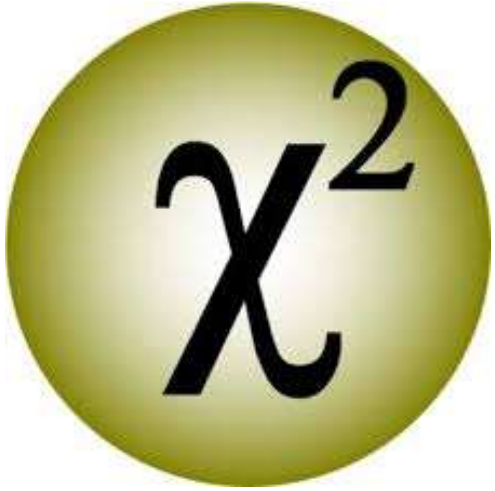


TESTE T DE ESCORE-Z POPULACIONAL
&
INTERVALO DE CONFIANÇA DE PREVALÊNCIA

Intervalo de confiança da proporção de baixo PRRN em R

```
library(readxl)
library(DescTools)
media_sem.etilismo <- 20
dp_sem.etilismo <- 4
n_sem.etilismo <- 20
n_com.etilismo <- 50
Dados <- readxl::read_excel("Table 8.2 RawBirthWeight.xls")
Dados$RBW_z <- (Dados$RBW - media_sem.etilismo)/dp_sem.etilismo
# Comparing an Individual's Test Score Against Norms Derived from Small Samples
# Crawford & Howell - 1998
nBPRRN <- sum(Dados$RBW_z < -1.77)
resultado <- DescTools::BinomCI(x=nBPRRN, n=n_com.etilismo, method="wilson")
round(resultado, digits=4)

#      est lwr.ci upr.ci
#[1,] 0.18 0.0977  0.308
```

50



TESTE QUI-QUADRADO DE DESVIO-PADRÃO POPULACIONAL

Teste qui-quadrado bilateral de desvio-padrão populacional para uma condição em R

Teste

- Desvio-padrão σ = 4 g hipotetizado

Suposições

- PRRN tem distribuição normal (desnecessária, pois n > 30)
- n = 50 observações independentes de PRRN (TCL)
- Teste bilateral
- Nível de confiança de 95% (ou nível de significância adotado de 5%)

Hipóteses (não é teste da suposição do teste z: é teste do efeito do etilismo no PRRN)

- H_0: $\sigma = 4$
- H_1: $\sigma \neq 4$

Estatísticas

- S = 5,11
- GL = 50 – 1 = 49
- Estatística de teste qui-quadrado = 80
- IC95(σ) = [4,27; 6,37]

Decisão

- Como o desvio-padrão populacional hipotetizado NÃO está dentro do IC95, rejeitar H_0 ou
- Como o valor-p bilateral = 0.00674 é menor que 5%, rejeitar H_0

Intervalo de confiança do CV de Pearson

- CV sem etilismo: 4/20 = 0.2
- CV com etilismo: 5.11/18 = 0.284

```
$method
[1] "Corrected cv with Basic Bootstrap 95% CI"
$statistics
     est lower upper
  28.302 22.75 33.61
```

Intervalo de confiança do CV de Pearson em R

```
library(readxl)
library(cvcqv)
media_semetilismo <- 20
dp_semetilismo <- 4
Dados <- readxl::read_excel("Table 8.2 RawBirthWeight.xls")
print(CV_semetilismo <- dp_semetilismo/media_semetilismo)
cvcqv::cv_versatile(Dados$RBW,na.rm=TRUE,digits=3,
                    method="basic",correction = TRUE)
```

Teste qui-quadrado bilateral de desvio-padrão populacional para uma condição em R

```
library(readxl)
library(EnvStats)
dp_0 <- 4
Dados <- readxl::read_excel("Table 8.2 RawBirthWeight.xls")
out <- EnvStats::varTest(x=Dados$RBW, sigma.squared = dp_0^2)
print(out)
print(sqrt(out$conf.int))
print(sd(Dados$RBW,na.rm=TRUE))
```

```
      LCL       UCL
4.270402 6.370494
attr(,"conf.level")
[1] 0.95
```

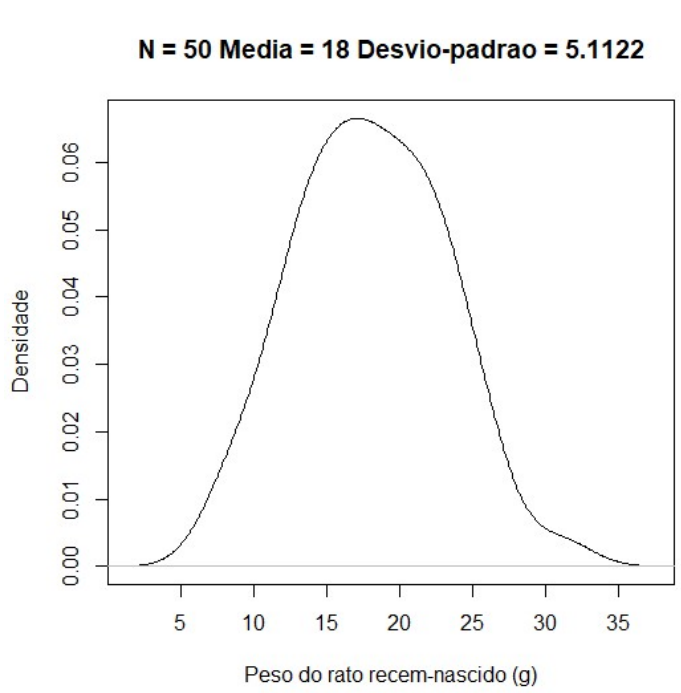
```
Results of Hypothesis Test
--------------------------

Null Hypothesis:                 variance = 16

Alternative Hypothesis:          True variance is not equal to 16

Test Name:                       Chi-Squared Test on Variance

Estimated Parameter(s):          variance = 26.13468

Data:                            PRRN[, 1]

Test Statistic:                  Chi-Squared = 80.03745

Test Statistic Parameter:        df = 49

P-value:                         0.006746438

95% Confidence Interval:         LCL = 18.23633
                                 UCL = 40.58319
```

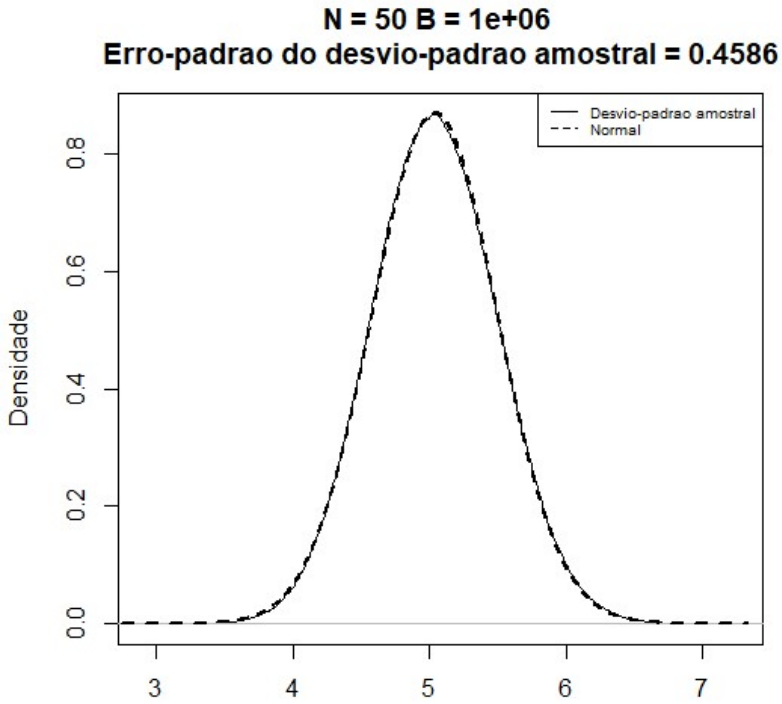
55

Reamostragem do desvio-padrão amostral em R

```
library(readxl)
library(EnvStats)
B <- 1e6; alfa <- 0.05; set.seed(123)
Dados <- readxl::read_excel("Table 8.2 RawBirthWeight.xls")
PRRN <- as.matrix(Dados[,2])
N <- nrow(PRRN)
plot(density(PRRN, na.rm=TRUE),
     main=paste("N =", N,
                "Media =", round(mean(PRRN,na.rm=TRUE),1),
                "Desvio-padrao =",
                round(sd(PRRN, na.rm=TRUE), 4)),
     xlab="Peso do rato recem-nascido (g)", ylab="Densidade")
PRRN.dp.boot <- replicate(B, sd(sample(PRRN, replace=TRUE)))
print(dp <- sd(PRRN, na.rm=TRUE))
print(mean(PRRN.dp.boot, na.rm=TRUE))
quantile(PRRN.dp.boot, probs=c(alfa/2, 1 - alfa/2))
ICDP <- EnvStats::varTest(PRRN[,1])
plot(density(PRRN.dp.boot, na.rm=TRUE),
     main=paste("N =", N, "B =", B,
                "\nErro-padrao do desvio-padrao amostral =",
                round(sd(PRRN.dp.boot, na.rm=TRUE), 4)),
     xlab="Desvio-padrao amostral do PRRN (g)", ylab="Densidade")
mi <- mean(PRRN.dp.boot, na.rm=TRUE)
print(EP <- sd(PRRN.dp.boot, na.rm=TRUE))
x <- seq(from=mi-5*EP, to=mi+5*EP, by=1e-5)
y <- dnorm(x, mean=mi, sd=EP)
lines(x,y,lwd=2,lty=2)
legend("topright", c("Desvio-padrao amostral","Normal"), lty=1:2, cex=.6)
```

Reamostragem do desvio-padrão amostral em R






```
      LCL      UCL
4.270402 6.370494
attr(,"conf.level")
[1] 0.95
```

2.5%	97.5%
4.16	5.95

Misuses of correlation and regression analyses in orthodontic research: The problem of mathematical coupling

Yu-Kang Tu,[a] Zararna L. Nelson-Moon,[b] and Mark S. Gilthorpe[c]
Leeds, United Kingdom

Introduction: The aim of this article was to encourage good practice in the statistical analyses of orthodontic research data. Our objective was to highlight the statistical problems caused by mathematical coupling (MC) in correlation and regression analyses. These statistical problems are among the most common pitfalls in orthodontic research when exploring associations among clinical variables. This article will show why these problems arise and how they can be avoided and overcome. **Methods:** Four orthodontic journals were electronically and manually searched for articles that used correlation and regression analyses. Studies that seemed to suffer from MC in their statistical analyses were identified and carefully examined. **Results:** Several examples from our search illustrate that MC in correlation and regression analyses can potentially cause misleading results. More appropriate statistical methods are available and should be used to eliminate confusing results and improve any subsequent interpretations. Because many clinical and radiographic variables used in orthodontic research are correlated due to direct or indirect MC, interpretation of studies in the literature needs to be cautious. **Conclusions:** Correlation and regression analyses are useful tools in orthodontic research when their assumptions and limitations are recognized. However, greater care is required in formulating research questions and experimental designs. It is prudent to seek statistical advice when orthodontic research involves complex data analyses. (Am J Orthod Dentofacial Orthop 2006;130:62-8)

58

Correlating change with baseline (MC)

Suppose x is the pretreatment (baseline) value and y the posttreatment (follow-up) value, the Pearson correlation between change $(x - y)$ and the pretreatment value (x) is: [30]

$$Corr[x - y,\ x] = \frac{s_x - r_{xy}s_y}{\sqrt{s_x^2 + s_y^2 - 2r_{xy}s_xs_y}}$$

where s_x^2 is the variance (square of the standard deviation) of the observation x, s_y^2 is the variance of y, and r_{xy} is the correlation between x and y. The correlation between change and baseline is often positive if r_{xy} is positive, which is likely when repeated measurements are made on the same subjects. This is a typical example of MC.[9,10] If r_{xy} is close to zero—ie, there is poor correlation between pretreatment and posttreatment values—the spurious positive association between baseline and change is large; if r_{xy} is far from zero, the spurious association is less marked, although still present. Poor correlation between x and y could occur because either the treatment is unpredictable or the measurement errors are large, or both. Whatever the cause, a poor correlation between x and y only exacerbates the effects of MC; moreover, if r_{xy} is close to zero, the actual dependency of change on baseline values does not exist or becomes negligible.[31]

	Final	Inicial	Delta
A	74	3	71
A	68	4	64
A	77	5	72
B	76	2	74
B	80	4	76
C	87	3	84
C	91	7	84
M=	79,00	4,00	75,00
DP=	7,83	1,63	7,19
r(F,I)=	0,48		
r(I,D)=	-0,30		

	Final	Inicial	Delta
A	74	3	71
A	68	4	64
A	77	5	72
B	76	2	74
B	80	4	76
C	87	3	84
C	91	7	84
M=	79,00	4,00	75,00
DP=	7,83	1,63	7,19
r(F,I)=	0,00		
r(I,D)=	0,20		

Oldham's method

In 1962, Oldham noticed the problem of MC in correlation/regression when seeking whether there was a relationship between treatment effect and baseline disease severity in patients with hypertension.[30] He warned against the common practice of regressing (or correlating) change $(x - y)$ with baseline (x), as the null hypothesis—that correlation coefficient or regression slope is zero—is no longer valid. He suggested that the change $(x - y)$ should be regressed on the arithmetic mean of the pretreatment and posttreatment values. The Pearson correlation between change and the mean is: [30]

$$Corr[x - y,\ (x + y)/2] = \frac{s_x^2 - s_y^2}{\sqrt{(s_x^2 + s_y^2)^2 - 4r_{xy}^2 s_x^2 s_y^2}},$$

where s_x^2, s_y^2, and r_{xy} are as defined previously.

The rationale behind Oldham's method is that x and y are 2 repeated measurements made on the same subject on successive occasions, so that their variances should be almost identical if there were no intervention and no other time-related biologic variation between measurement occasions.[30]

The variance ratio test

Based on the same assumptions as those for Oldham's method, the variance ratio s_x^2/s_y^2 has also been proposed as an appropriate test, by assessing the equality of the correlated variances,[31] yielding a statistic that follows the t distribution with $n - 2$ degrees of freedom,[32] and which is nonsignificant if the variances are similar:

$$t = \frac{(s_x^2 - s_y^2)\sqrt{n-2}}{2s_x s_y \sqrt{1 - r_{xy}^2}}$$

where s_x^2, s_y^2, and r_{xy} are as defined previously.

In the statistical literature, Oldham's method is sometimes called the Pittman-Morgan test, because Pittman and Morgan independently solved the problem of testing the equivalence of 2 correlated variances in 1939.[35,36] A limitation of Oldham's method and the variance ratio test is that the measurement errors of 2 repeated measurements need to be constant, but this is not always the case. For instance, in many circumstances, pretreatment or posttreatment values that have a greater mean might be accompanied by greater measurement error or greater biological variation. Even when a treatment effect is constant across all levels of the pretreatment values, the difference in measurement error or biological variation due to the change in mean outcome will give rise to the misleading impression that treatment effects depend on the baseline disease severity.

Intervalo de confiança do CV de Pearson em R

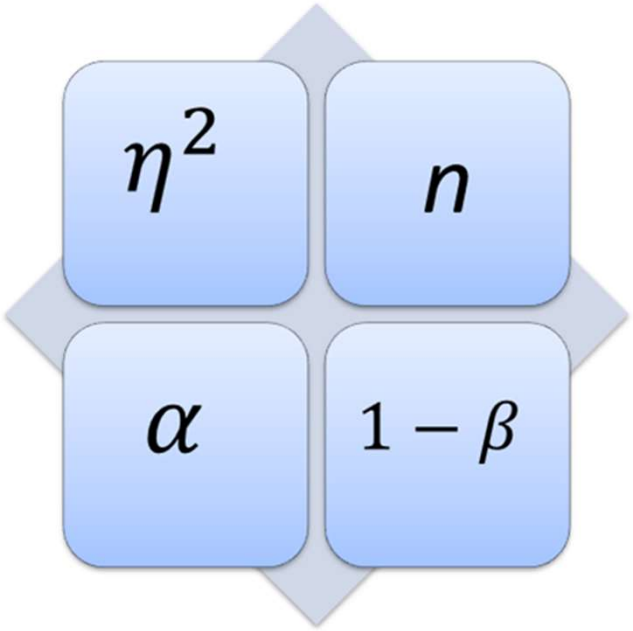
- CV sem etilismo: 4/20 = 0.2 (Não pertence ao IC95 do CV)
- CV com etilismo: 5.11/18 = 0.284

```
$method
[1] "Corrected cv with Basic Bootstrap 95% CI"
$statistics
     est lower upper
  28.302 22.75 33.61
```

62

Intervalo de confiança do CV de Pearson em R

```
library(readxl)
library(cvcqv)
media_semetilismo <- 20
dp_semetilismo <- 4
Dados <- readxl::read_excel("Table 8.2 RawBirthWeight.xls")
print(CV_semetilismo <- dp_semetilismo/media_semetilismo)
cvcqv::cv_versatile(Dados$RBW,na.rm=TRUE,digits=3,
                    method="basic",correction = TRUE)
```



PLANEJAMENTO DE ESTUDO

64

GPower

Tabela 4.1

Tipo de análise da potência	O que se pretende determinar	O que se deve especificar
Análise da potência *a priori*	n	α, $1 - \beta$ e ES
Análise da potência *post hoc*	$1 - \beta$	α, ES e n
Análise da potência de compromisso	α e $1 - \beta$	ES, n e $\frac{\beta}{\alpha}$
Análise de sensibilidade	ES mínimo	α, $1 - \beta$ e n
Análise do critério	α	$1 - \beta$, ES e n

COELHO, JP *et al*. (2008) *Inferência Estatística: com utilização do SPSS e G*Power*. Lisboa: Sílabo.

INTERNATIONAL REVIEW OF SOCIAL PSYCHOLOGY

Perugini, M., et al. (2018). A Practical Primer To Power Analysis for Simple Experimental Designs. *International Review of Social Psychology*, 31(1): 20, 1–23, DOI: https://doi.org/10.5334/irsp.181

RESEARCH ARTICLE

A Practical Primer To Power Analysis for Simple Experimental Designs

Marco Perugini, Marcello Gallucci and Giulio Costantini

Power analysis is an important tool to use when planning studies. This contribution aims to remind readers what power analysis is, emphasize why it matters, and articulate when and how it should be used. The focus is on applications of power analysis for experimental designs often encountered in psychology, starting from simple two-group independent and paired groups and moving to one-way analysis of variance, factorial designs, contrast analysis, trend analysis, regression analysis, analysis of covariance, and mediation analysis. Special attention is given to the application of power analysis to moderation designs, considering both dichotomous and continuous predictors and moderators. Illustrative practical examples based on G*Power and R packages are provided throughout the article. Annotated code for the examples with R and dedicated computational tools are made freely available at a dedicated web page (https://github.com/mcfanda/primerPowerIRSP). Applications of power analysis for more complex designs are briefly mentioned, and some important general issues related to power analysis are discussed.

Keywords: power analysis; effect size; moderation; sensitivity analysis; uncertainty

66

LIMITS OF RETROSPECTIVE POWER ANALYSIS

PATRICK D. GERARD,[1] Experimental Statistics Unit, Box 9653, Mississippi State University, Mississippi State, MS 39762, USA
DAVID R. SMITH, U.S. Geological Service, Biological Resources Division, Leetown Science Center, Kearneysville, WV 25430, USA
GOVINDA WEERAKKODY, Department of Mathematics and Statistics, Box 9715, Mississippi State University, Mississippi State, MS 39762, USA

Abstract: Power analysis after study completion has been suggested to interpret study results. We present 3 methods of estimating power and discuss their limitations. We use simulation studies to show that estimated power can be biased, extremely variable, and severely bounded. We endorse the practice of computing power to detect a biologically meaningful difference as a tool for study planning but suggest that calculation of confidence intervals on the parameter of interest is the appropriate way to gauge the strength and biological meaning of study results.

1998 ***JOURNAL OF WILDLIFE MANAGEMENT 62(2):801–807***

Key words: interpretation of results, noncentrality parameter, statistical power, study design.

Table 1. Expected power when power is estimated with the observed test statistic based on simulated comparisons between 3 populations where sample size was 20/population. Six values of power (0.05, 0.11, 0.34, 0.66, 0.90, 0.98) were created by setting $\alpha = 0.05$ and altering population means. Population means were distributed normally with unit variance. Three estimators of power were calculated for each replication; estimators are discussed in detail in the text. Expected power, E(P), and expected upper and lower bounds for the 95% confidence interval (CI) were computed by averaging over 500 replications. For each replicate and the simple plug-in (P_p) and corrected (P_{bc}) estimators, the bounds for the 95% CI were the appropriate percentiles from a bootstrap sample of size 400. For the percentile estimator (P_{mu}), the 95% CI was based on exact methods.

True power	Simple plug-in (P_p)		Corrected (P_{bc})		Percentile (P_{mu})	
	E(P_p)	95% CI	E(P_{bc})	95% CI	E(P_{mu})	95% CI
0.05	0.236	(0.072, 0.870)	0.126	(0.052, 0.790)	0.163	(0.051, 0.671)
0.11	0.271	(0.081, 0.890)	0.151	(0.055, 0.820)	0.194	(0.055, 0.717)
0.34	0.455	(0.133, 0.948)	0.309	(0.073, 0.912)	0.375	(0.074, 0.870)
0.66	0.662	(0.259, 0.981)	0.538	(0.153, 0.966)	0.601	(0.148, 0.950)
0.90	0.844	(0.462, 0.995)	0.770	(0.326, 0.991)	0.811	(0.309, 0.987)
0.98	0.948	(0.678, 0.999)	0.915	(0.561, 0.998)	0.934	(0.533, 0.998)

Análise de poder retrospectivo (*plug in*) de ANOVA unifatorial independente balanceada (Gerard *et al*. (1998)

Analise de poder retrospectivo de ANOVA unifatorial independente balanceada.R

```
library(MBESS)
k <- 3
n <- 20
alfa <- 0.05
dfn <- k - 1
dfd <- k*(n - 1)
Fcrt <- qf(1-alfa, dfn, dfd)
Fobs <- Fcrt*0.99
eta2 <- dfn*Fobs/(dfn*Fobs+dfd)
eta2lims <- MBESS::ci.pvaf(Fobs, dfn, dfd, k*n, 1-alfa)
f2 <- eta2/(1-eta2)
f2.ll <- eta2lims$Lower.Limit.Proportion.of.Variance.Accounted.for/
  (1-eta2lims$Lower.Limit.Proportion.of.Variance.Accounted.for)
f2.ul <- eta2lims$Upper.Limit.Proportion.of.Variance.Accounted.for/
  (1-eta2lims$Upper.Limit.Proportion.of.Variance.Accounted.for)
ncp.p <- dfd*f2 # ou dfn*Fobs
ncp.p.ll <- dfd*f2.ll
ncp.p.ul <- dfd*f2.ul
cat(paste("N =", k*n, "\tFcrt =", round(Fcrt,2),"\tFobs =", round(Fobs,2),"\n"))
poder.p <- 1-pf(Fcrt,dfn, dfd, ncp.p)
cat(paste("\tPoder.p =", round(poder.p,3),"\n"))
poder.p.ll <- 1-pf(Fcrt,dfn, dfd, ncp.p.ll)
cat(paste("\tPoder.p.ll =", round(poder.p.ll,3),"\n"))
poder.p.ul <- 1-pf(Fcrt,dfn, dfd, ncp.p.ul)
cat(paste("\tPoder.p.ul =", round(poder.p.ul,3),"\n"))
sink()
```

```
N = 60        Fcrt = 3.16   Fobs = 3.13
        Poder.p = 0.579
        Poder.p.ll = 0.05
        Poder.p.ul = 0.967
```

Bacchetti *BMC Medicine* 2010, **8**:17
http://www.biomedcentral.com/1741-7015/8/17

BMC Medicine

DEBATE Open Access

Current sample size conventions: Flaws, harms, and alternatives

Peter Bacchetti

Abstract

Background: The belief remains widespread that medical research studies must have statistical power of at least 80% in order to be scientifically sound, and peer reviewers often question whether power is high enough.

Discussion: This requirement and the methods for meeting it have severe flaws. Notably, the true nature of how sample size influences a study's projected scientific or practical value precludes any meaningful blanket designation of <80% power as "inadequate". In addition, standard calculations are inherently unreliable, and focusing only on power neglects a completed study's most important results: estimates and confidence intervals. Current conventions harm the research process in many ways: promoting misinterpretation of completed studies, eroding scientific integrity, giving reviewers arbitrary power, inhibiting innovation, perverting ethical standards, wasting effort, and wasting money. Medical research would benefit from alternative approaches, including established *value of information* methods, simple choices based on cost or feasibility that have recently been justified, sensitivity analyses that examine a meaningful array of possible findings, and following previous analogous studies. To promote more rational approaches, research training should cover the issues presented here, peer reviewers should be extremely careful before raising issues of "inadequate" sample size, and reports of completed studies should not discuss power.

Summary: Common conventions and expectations concerning sample size are deeply flawed, cause serious harm to the research process, and should be replaced by more rational alternatives.

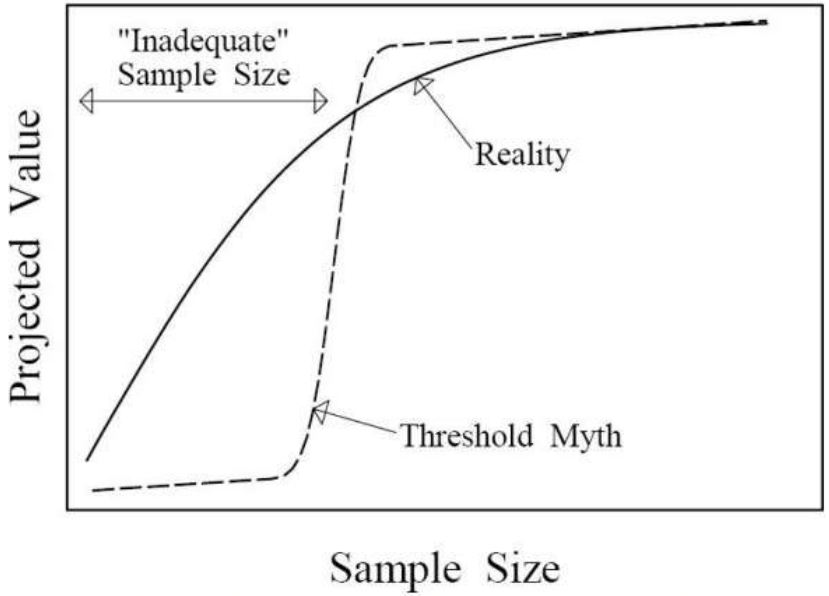


Figure 1 Qualitative depiction of how sample size influences a study's projected scientific and/or practical value. A threshold shaped relationship (dashed line) would create a meaningful distinction between adequate and inadequate sample sizes, but such a relation does not exist. The reality (solid line) is qualitatively different, exhibiting diminishing marginal returns. Under the threshold myth, cutting a sample size in half could easily change a valuable study into an inadequate one, but in reality such a cut will always preserve *more* than half of the projected value.

70

Referências

- COELHO, JP et al. (2008) Inferência Estatística: com utilização do SPSS e G*Power. Lisboa: Sílabo.
- DANCEY, C & REIDY, J (2019) *Estatística sem Matemática para Psicologia.* 7ª ed. Porto Alegre: Penso.
- STEPHENS, LJ (2009) *Statistics in Psychology*. NY: McGraw-Hill, Schaum's Outline Series.

72