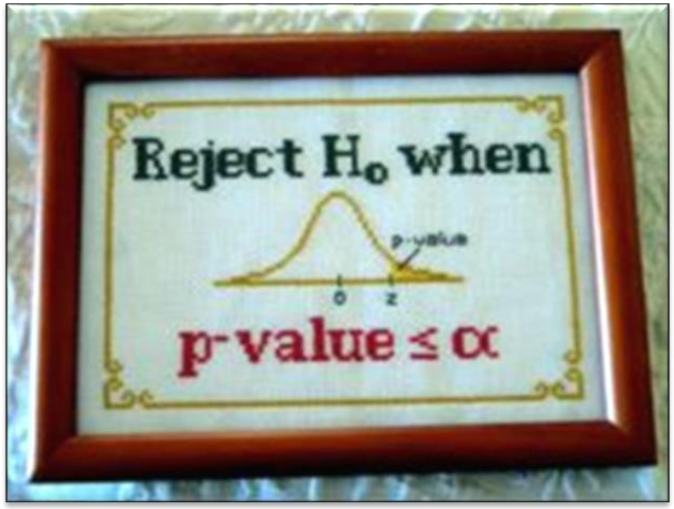


Testagem de Hipótese Nula

- A lógica da testagem de hipótese nula
- A significância estatística e como ela se relaciona com probabilidade condicionada
- Como as distribuições de probabilidade formam as bases dos testes estatísticos
- Os problema associados em utilizar probabilidades como base para conclusões (erros do tipo I e do tipo II)
- Hipóteses unilaterais e bilaterais
- Como escolher o teste apropriado para analisar seus dados

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 = 7cm.

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

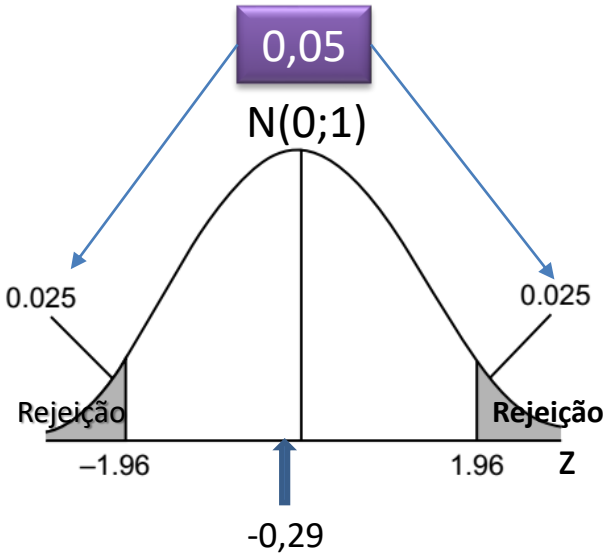
Quatro participantes desse grupo tiveram suas estaturas medidas, cujos valores em centímetro são: 169, 174, 175, 186.

Hipóteses

- H_0: μ = 177
- H_1: μ ≠ 177 (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 = $\frac{\bar{X}-177}{EP} = \frac{176-177}{3{,}5} = -0{,}29$



Decisão

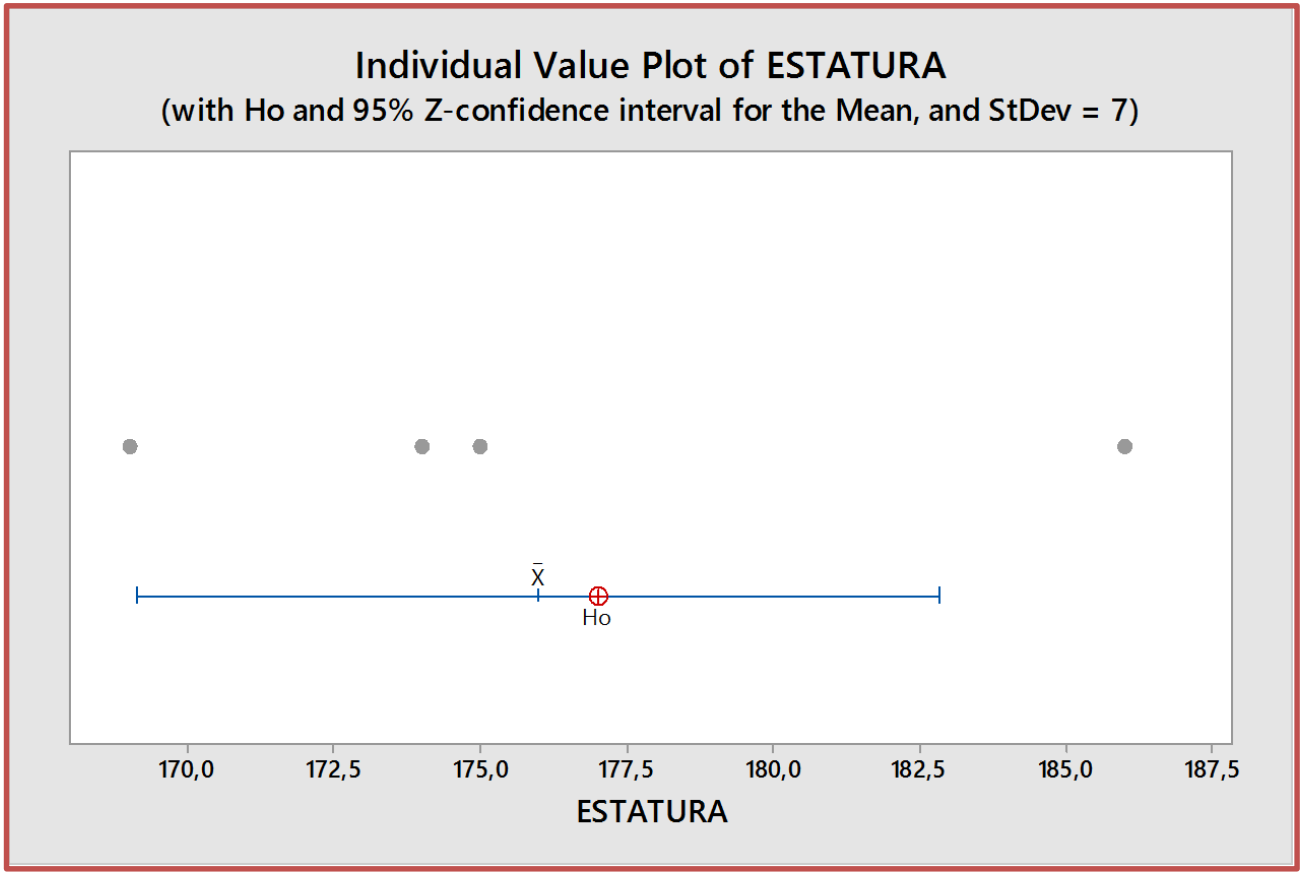
- *Critério do valor crítico*: Como |z| = 0,29 < 1,96, não rejeitar H_0 ou
- *Critério do IC95*: Como IC95 contém 177, não rejeitar H_0 : *O IC95 é a região de não-rejeição de H_0 na escala da variável X*

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 = 177cm hipotetizada

Suposições

- Estatura tem distribuição normal
- Desvio-padrão σ=7cm 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 z = $\frac{\bar{X}-177}{EP} = -0{,}29$

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 = 7cm.

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

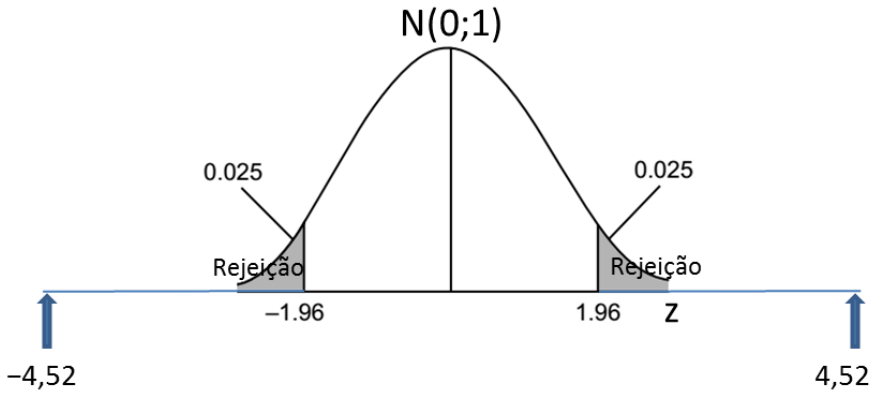
Mil participantes desse grupo tiveram suas estaturas medidas, cujos valores em centímetro são: 169, 174, 175, 186, ...

Hipóteses

- H_0: μ = 177
- H_1: μ ≠ 177

Estatísticas

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



Decisão

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

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
```

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.

Valor-p

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

Teste z bilateral para um grupo

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

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

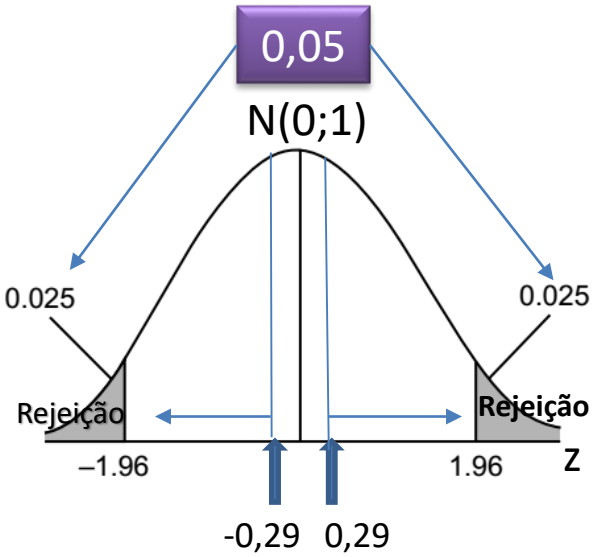
Quatro participantes desse grupo tiveram suas estaturas medidas, cujos valores em centímetro são: 169, 174, 175, 186.

Hipóteses

- H_0: μ = 177
- H_1: μ ≠ 177

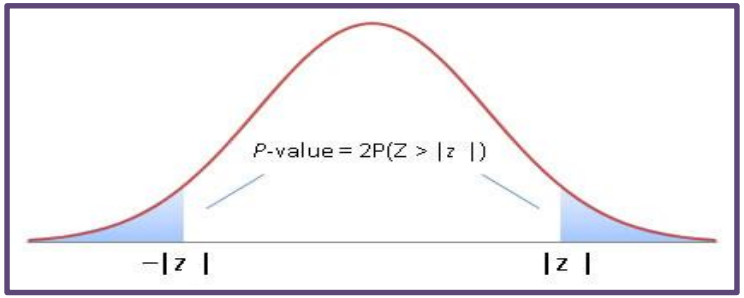
Estatísticas

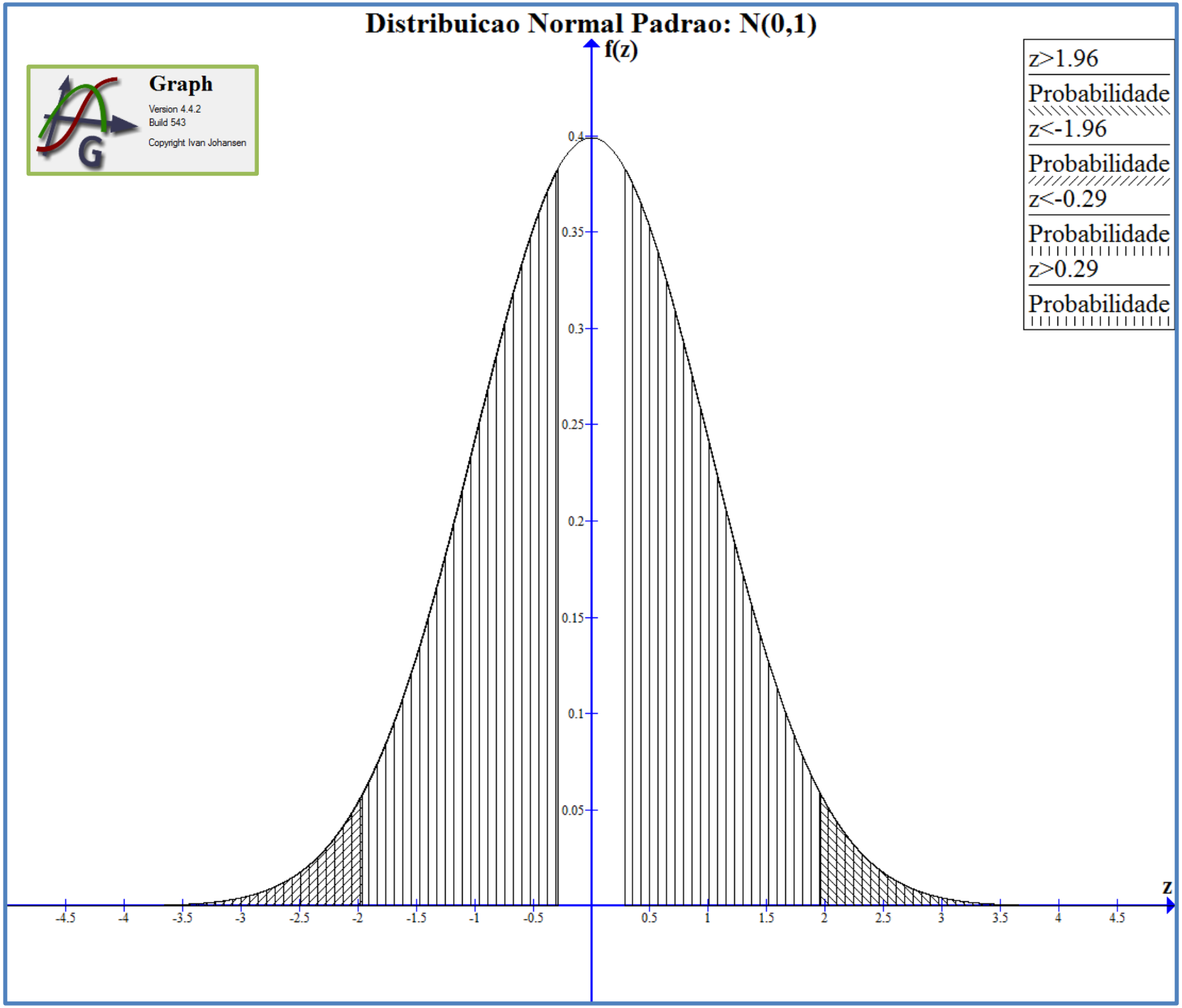
- $\bar{X}$ = (169 + 174 + 175 + 186)/4 = 176
- EP = $\frac{\sigma}{\sqrt{n}} = 3{,}5$
- IC95(μ) = [176-1,96×3,5; 176+1,96×3,5]
  = [169,14; 182,86]
- Estatística de teste z = $\frac{\bar{X}-177}{EP} = \frac{176-177}{3{,}5} = -0{,}29$



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 a probabilidade de escores-z serem mais extremos que 0,29 e -0,29, i.e., o valor-p bilateral = 0,77 = 2*pnorm(-abs(-0.29)) é maior que 5%, não rejeitar H_0




Distribuicao Normal Padrao: N(0,1)
Graph
Version 4.4.2
Build 543
Copyright Ivan Johansen
f(z)
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

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

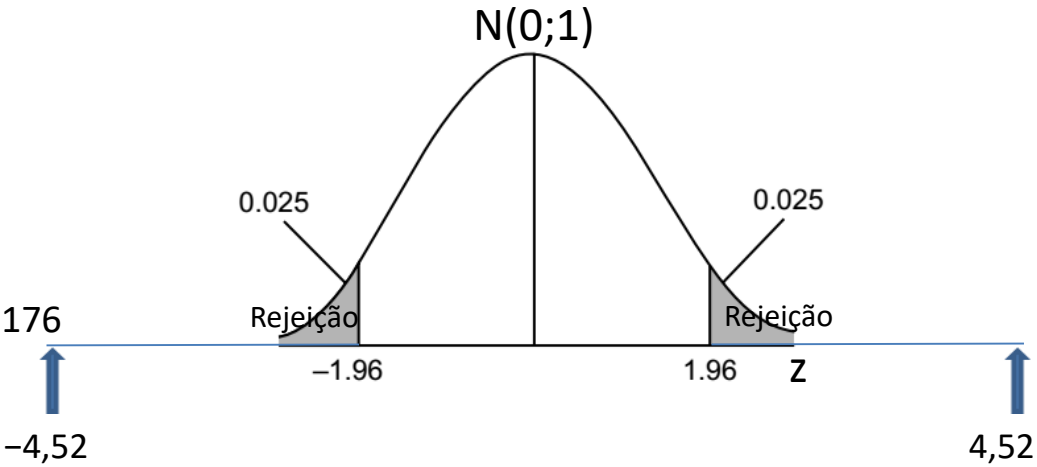
A distribuição da estatura do homem brasileiro de 19 anos de 2016 é normal com desvio-padrão σ=7cm.
Testar se a média populacional μ é igual a μ_0 = 177cm hipotetizada.
Mil participantes desse grupo tiveram suas estaturas medidas, cujos valores em centímetro são:
169, 174, 175, 186, ...

Hipóteses

- H_0: μ = 177
- H_1: μ ≠ 177

Estatísticas

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



Decisão

- Como |z| = 4,52 > 1,96, rejeitar H_0 ou
- Como IC95 não contém 177, rejeitar H_0
- Como a probabilidade de escores-z serem mais extremos que -4,52 e 4,52, i.e., o valor-p bilateral = 6,18E-06 = 2*pnorm(-abs(-4.52)) é menor que 5%, rejeitar H_0

Teste z bilateral para uma condição

Teste

- Média populacional μ = 177cm hipotetizada

Suposições

- Estatura tem distribuição normal
- Desvio-padrão σ=7cm conhecido
- n=1000 observações independentes
- Teste bilateral
- Nível de confiança de 95% (ou nível de significância adotado de 5%)

Hipóteses

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

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 = $\frac{\bar{X}-177}{EP} = -4{,}52$

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 a probabilidade de escores-z serem mais extremos que -4,52 e 4,52, i.e., o valor-p bilateral = 6,18E-06 = 2*pnorm(-abs(-4.52)) é menor que 5%, rejeitar H_0

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 de hipótese nula
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

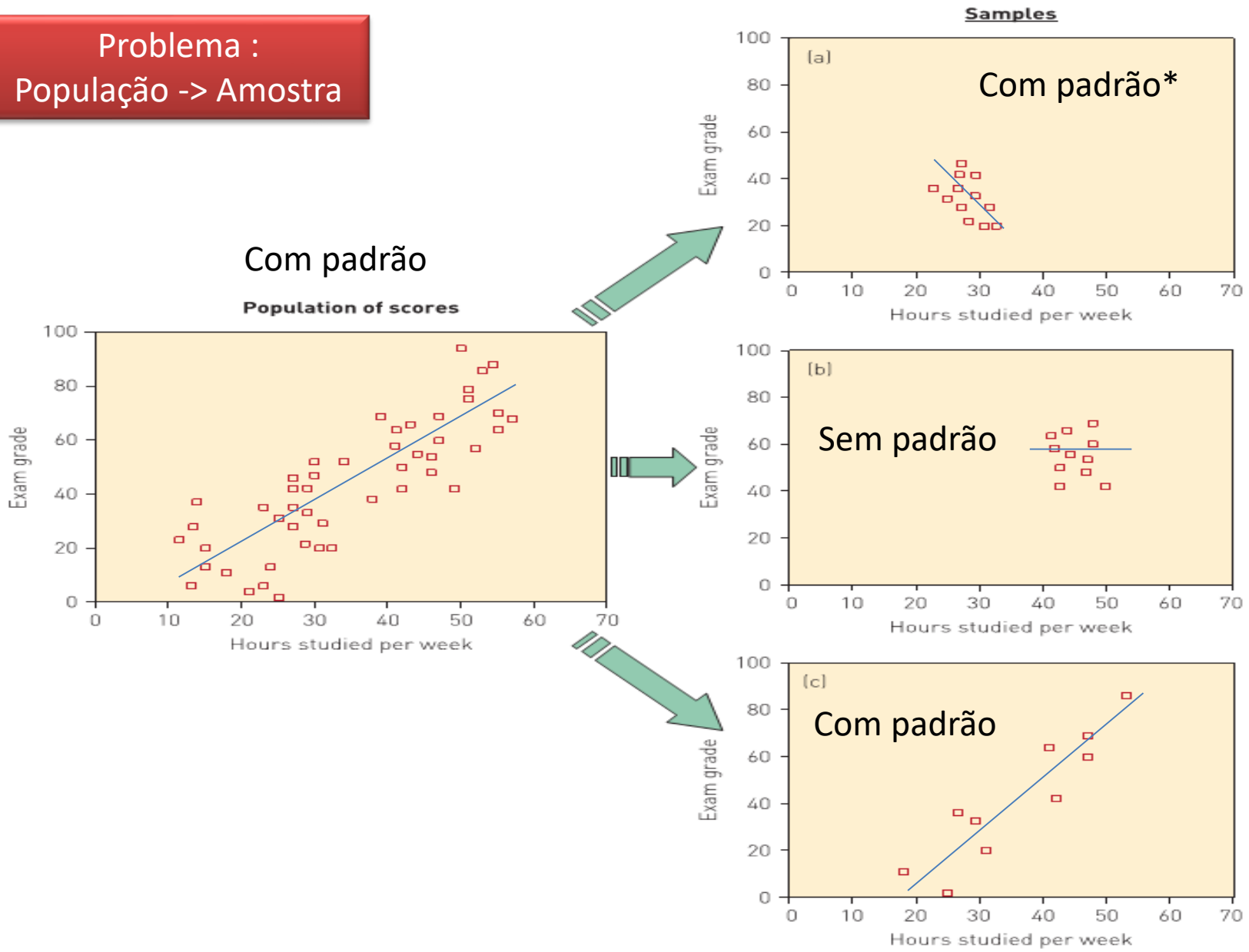


Figure 5.1 Scattergrams illustrating possible samples selected from a population with a positive relationship between number of hours spent studying and exam grades

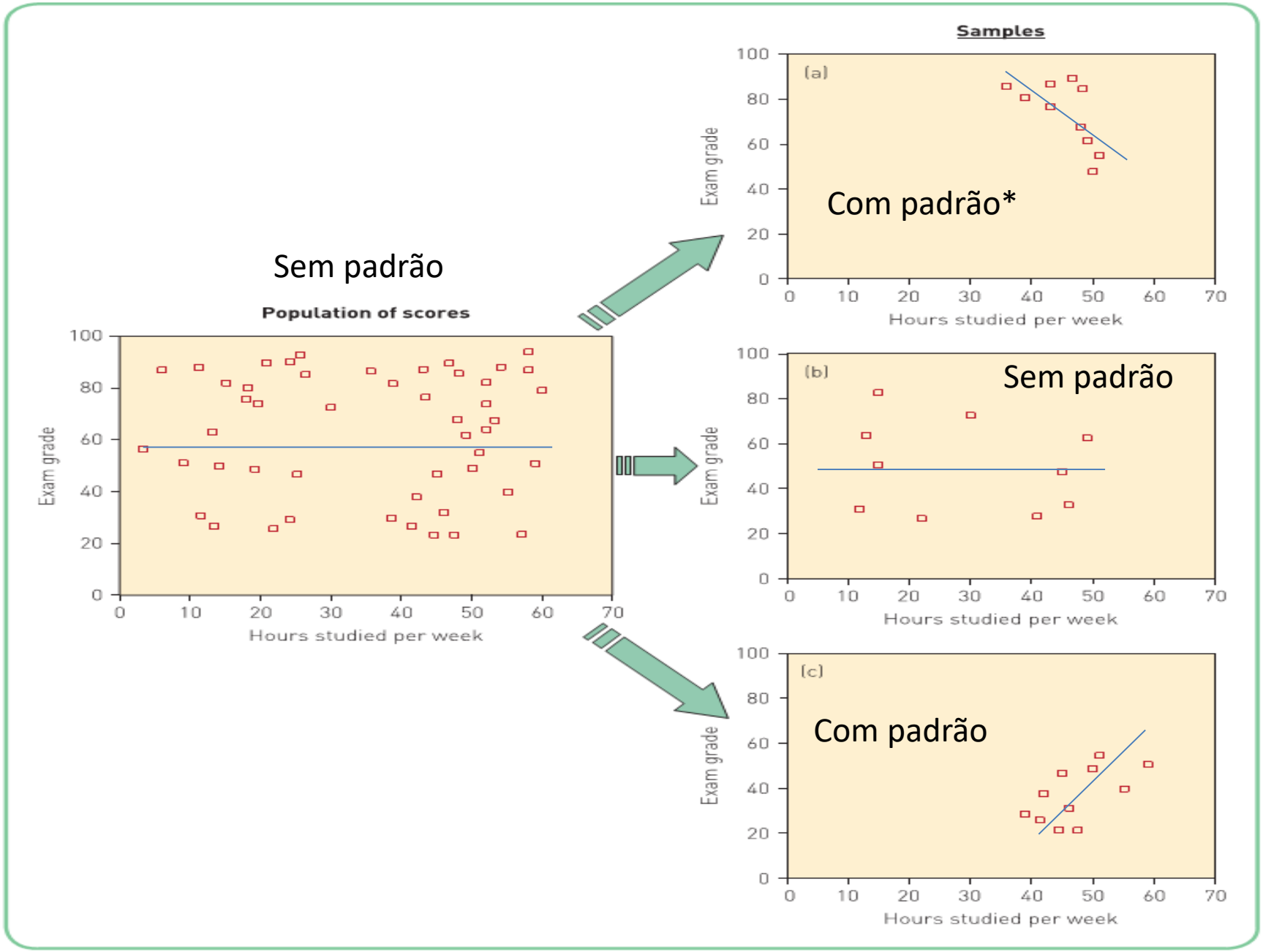


Figure 5.2 Scattergrams illustrating possible samples selected from a population with no relationship between number of hours spent studying and exam grades

Erro (Flutuação) Amostral

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

Problema invertido:
Amostra -> População

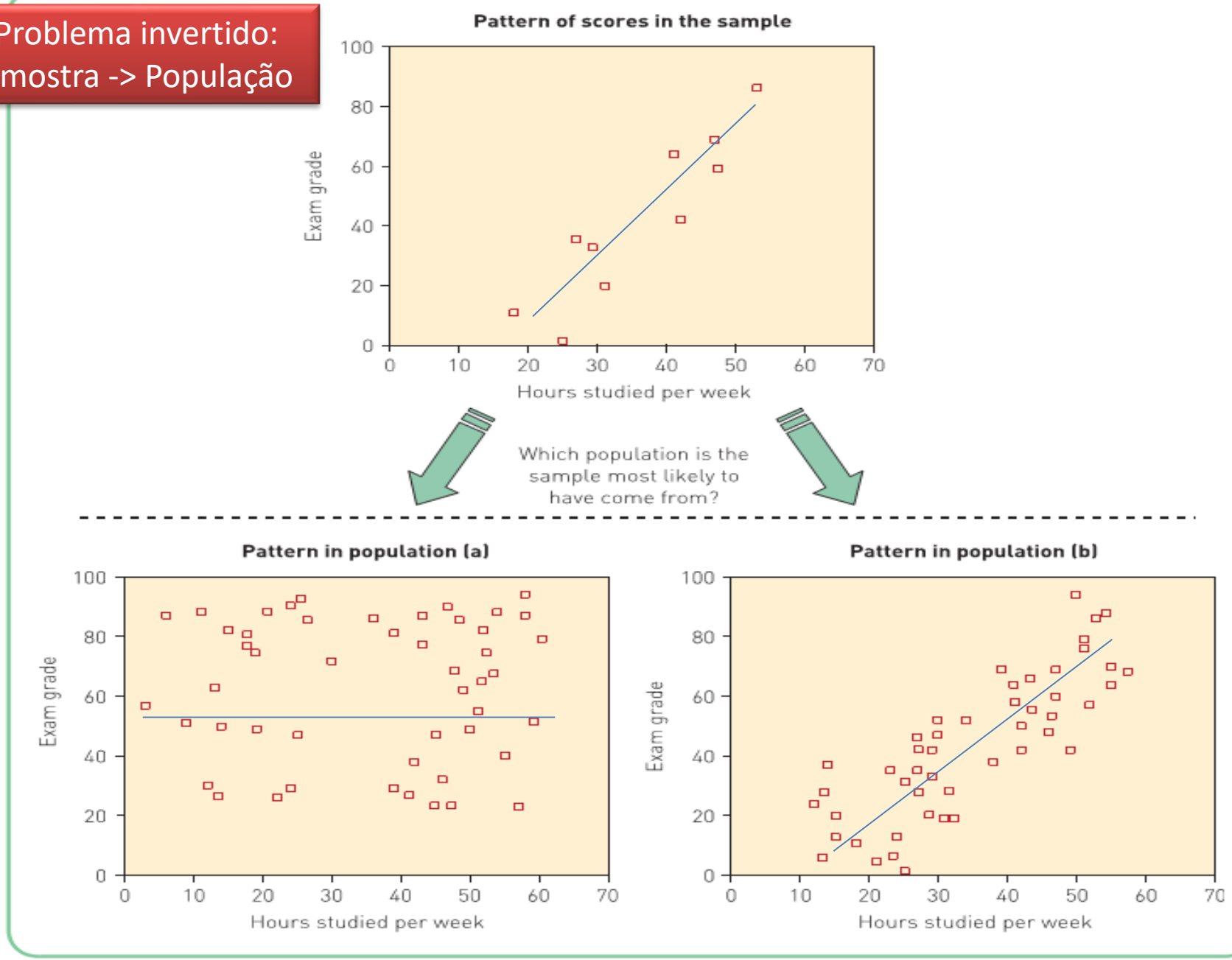


Figure 5.3 Scattergrams illustrating alternative underlying populations when a relationship is observed in a sample

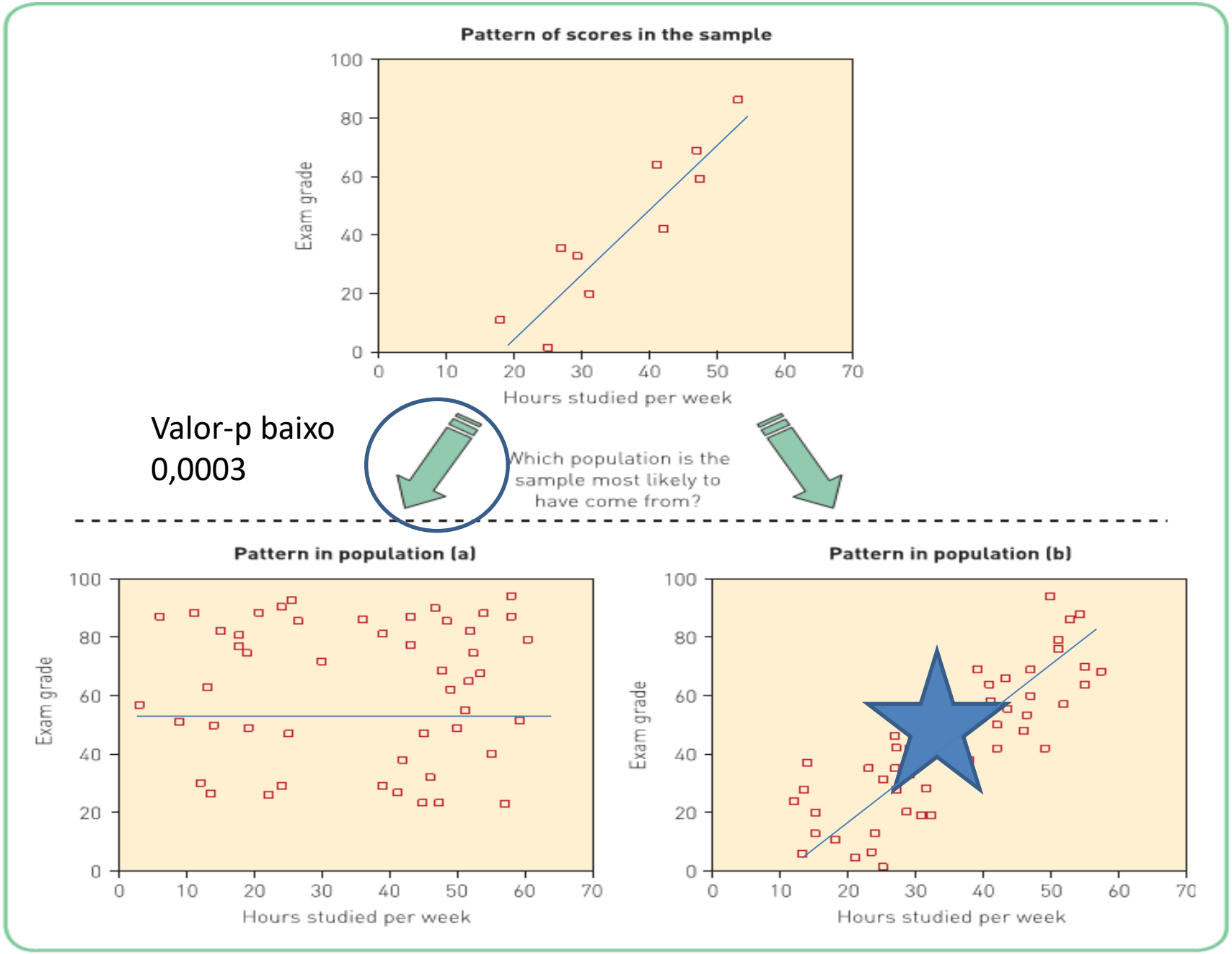


Figure 5.3 Scattergrams illustrating alternative underlying populations when a relationship is observed in a sample

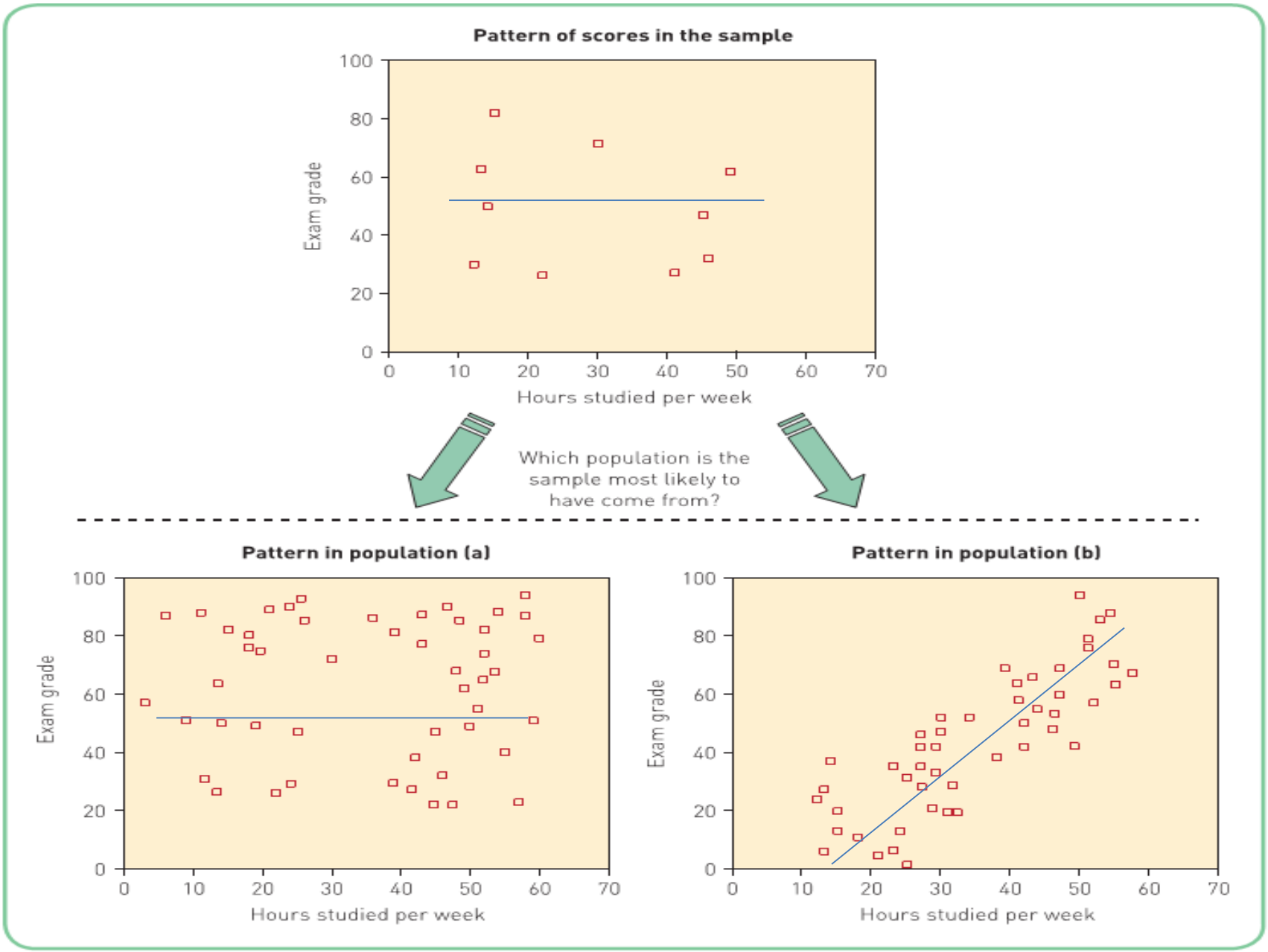


Figure 5.4 Scattergrams illustrating alternative underlying populations when no relationship is observed in a sample

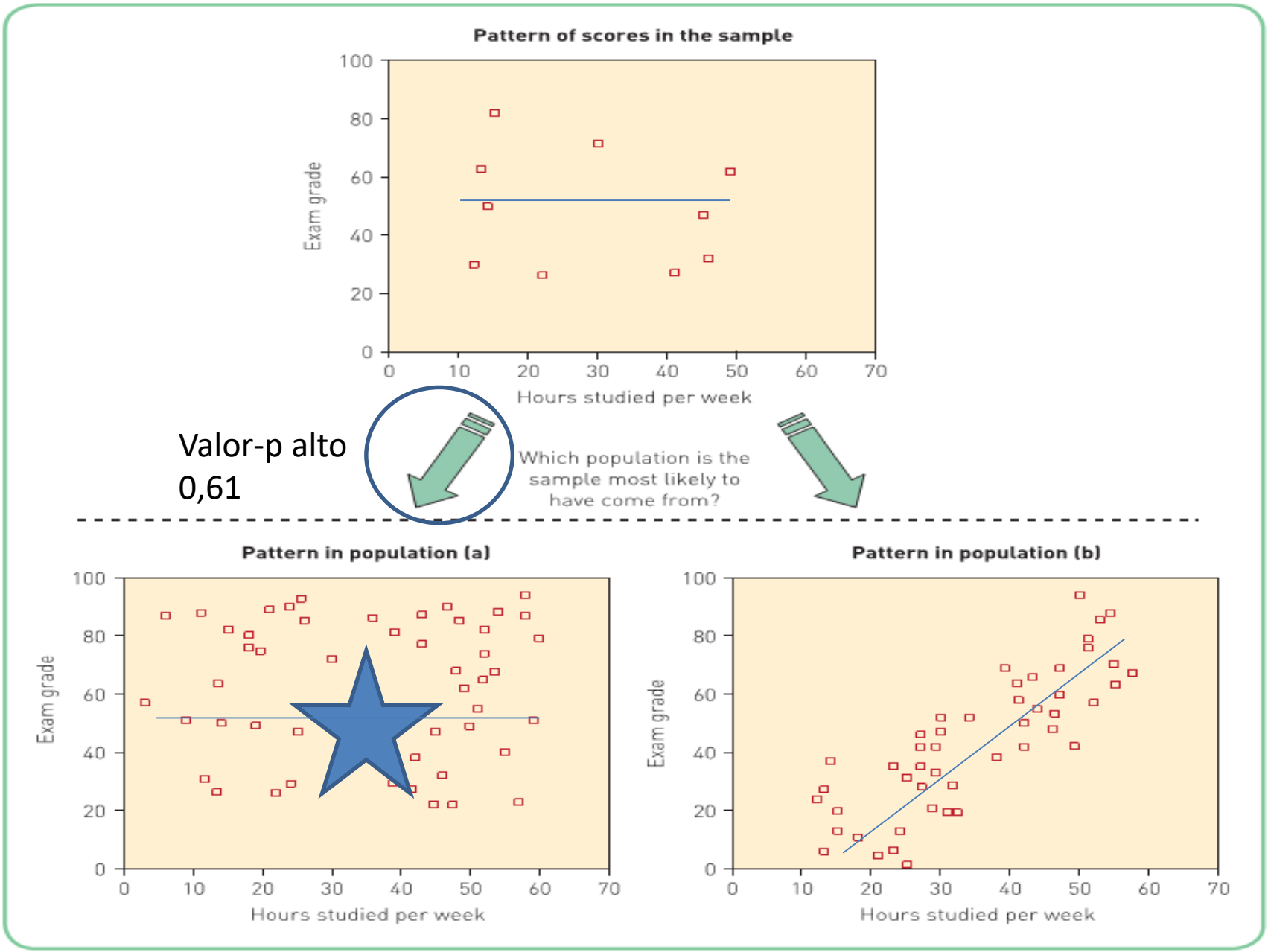


Figure 5.4 Scattergrams illustrating alternative underlying populations when no relationship is observed in a sample

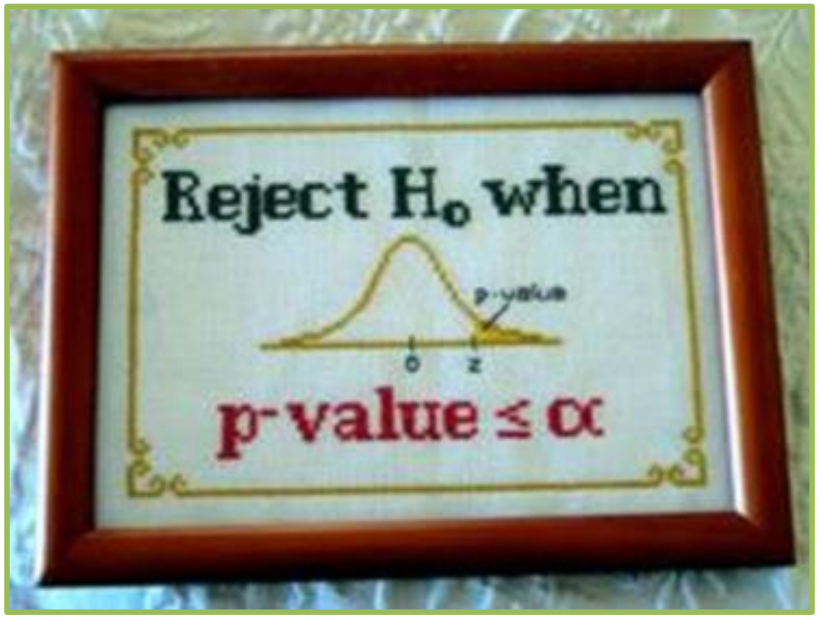
Efeito

- Correlação entre variáveis
- Diferença entre condições

Hipótese nula

- 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 grupos específicos podem estar relacionados entre si.
  - De forma alternativa, pode ser nossa previsão de como grupos específicos de participantes podem ser diferentes entre si ou como um grupo de participantes pode ser diferente quando tem um desempenho sob duas ou mais condições experimentais.

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 $p = \beta$

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.

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

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).

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.

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

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

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

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

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.

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.*

Warnings on Interpreting Results of Hypothesis Testing

PASW Statistics 18: Statistical Procedures Companion, Marija J. Norusis. NJ: Prentice Hall, 2010.

The results of hypothesis testing must be reported carefully.

You can never conclude with certainty' that the null hypothesis is either true or false.

It is easy to make pretentious-sounding claims that are not statistically justified.

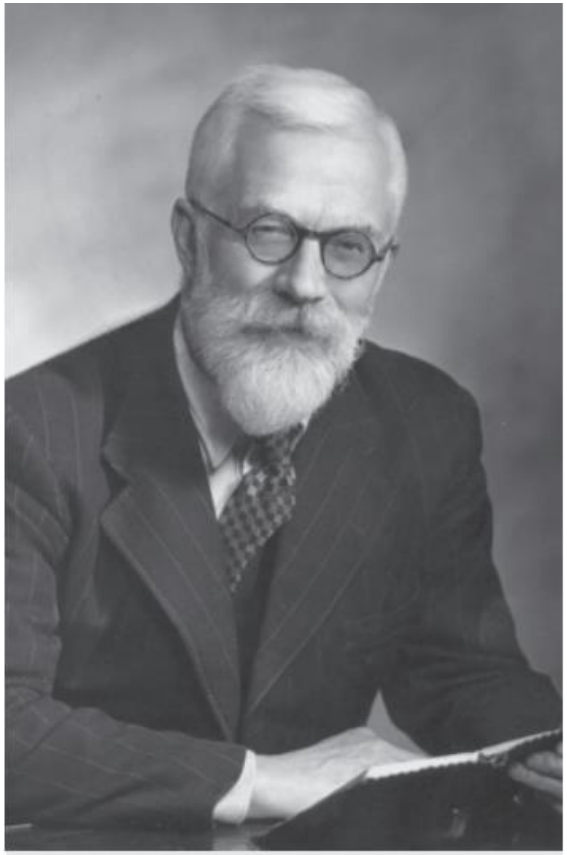
1. Never conclude from a hypothesis test that differences in one variable cause differences in the other unless you are analyzing a carefully controlled experiment. For example, you can't conclude that by staying in school you will decrease your risk of heart disease. It may be that earning more will let you join a health club and exercise. Orit may be that younger people are better educated and also less likely to develop heart disease.
2. Never equate statistical significance with practical significance. When you reject the null hypothesis, it is not necessarily the case that the difference or association you found is important or noteworthy. For large samples, even very small differences in means may be statistically significant. For example, if you find that a particular treatment prolongs life by one week compared to another, even if you've collected enough data to make the difference statistically significant, it is of little practical importance. That's why you should always examine the actual observed differences or the magnitudes of measures of association and focus only on those that are both statistically significant and practically meaningful. Always report the actual difference or coefficient, as well as its observed significance level.
3. Never equate failure to reject the null hypothesis with the null hypothesis being true. Your failure to reject the null hypothesis doesn't mean it is true. It could simply mean that you haven't gathered enough evidence to reject it. If your sample size is small, you may fail to reject the null hypothesis even when the population difference is large. That's why it's important, before you embark on a study, to determine how big a sample size you need in order to detect what you consider to be an important difference.
4. Never use the phrase "accept the null hypothesis" because it implies that you believe the null hypothesis is true. You can legitimately reject the null hypothesis, but you can't accept it. (That sounds unfair, but it is so, nonetheless.)
5. Never equate rejection of the null hypothesis with certainty that the null hypothesis is false.
6. Never assign probabilities to the null hypothesis or alternative hypothesis being true or false. The null hypothesis is either true or it is false. You can't know which. The alternative hypothesis is either true or it is false. Again, you can't know which.
7. Never claim that the observed significance level is the probability that the null hypothesis is true. The p value tells you the probability of seeing results at least as extreme as the ones you've observed if the null hypothesis is true. The null hypothesis is either true or it is not.
8. Never claim that the p value is the "probability that the results are due to chance." You can't talk about the probability of the results being due to chance unless you know that the null hypothesis is true.

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

Por que estabelecer α=5%?

- Em muitas situações esse valor fornece um ponto de equilíbrio entre as probabilidades dos erros de tipo I e II: D&R
- Dado um valor de nível de significância, o teste estatístico maximiza o poder estatístico do teste



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

signification

Hankins, MC (2013) Still not signifiant.

http://mchankins.wordpress.com/2013/04/21/still-notsignifiant-2

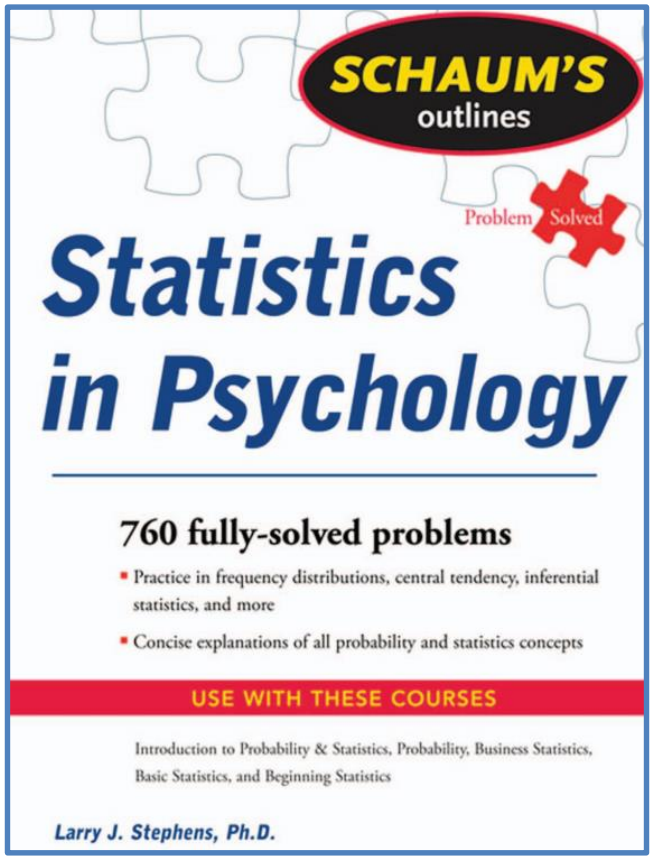
(barely) not statistically significant (p=0.052)
a barely detectable statistically significant difference (p=0.073)
a borderline significant trend (p=0.09)
a certain trend toward significance (p=0.08)
a clear tendency to significance (p=0.052)
a clear trend (p<0.09)
a clear, strong trend (p=0.09)
a considerable trend toward significance (p=0.069)
a decreasing trend (p=0.09)
a definite trend (p=0.08)
a distinct trend toward significance (p=0.07)
a favorable trend (p=0.09)
a favourable statistical trend (p=0.09)
a little significant (p<0.1)
a margin at the edge of significance (p=0.0608)
a marginal trend (p=0.09)
a marginal trend toward significance (p=0.052)
a marked trend (p=0.07)
a mild trend (p<0.09)
a moderate trend toward significance (p=0.068)
a near-significant trend (p=0.07)
a negative trend (p=0.09)
a nonsignificant trend (p<0.1)
a nonsignificant trend toward significance (p=0.1)
a notable trend (p<0.1)
a numerical increasing trend (p=0.09)
a numerical trend (p=0.09)
a positive trend (p=0.09)
uncertain significance (p>0.07)
vaguely significant (p>0.2)
verged on being significant (p=0.11)
verging on significance (p=0.056)
verging on the statistically significant (p<0.1)
verging-on-significant (p=0.06)
very close to approaching significance (p=0.060)
very close to significant (p=0.11)
very close to the conventional level of significance (p=0.055)
very close to the cut-off for significance (p=0.07)
very close to the established statistical significance level of p=0.05 (p=0.065)
very close to the threshold of significance (p=0.07)
very closely approaches the conventional significance level (p=0.055)
very closely brushed the limit of statistical significance (p=0.051)
very narrowly missed significance (p<0.06)
very nearly significant (p=0.0656)
very slightly non-significant (p=0.10)
very slightly significant (p<0.1)
virtually significant (p=0.059)
weak significance (p>0.10)
weakened..significance (p=0.06)
weakly non-significant (p=0.07)
weakly significant (p=0.11)
weakly statistically significant (p=0.0557)
well-nigh significant (p=0.11)

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.: valor t ou correlação).
- A estatística de teste (evidência amostral) será a mesma para qualquer um dos casos para um mesmo conjunto de dados.

CHAPTER 8

Introduction to Hypothesis Testing and the z-Test Statistic



Null Hypothesis (H_0) and Alternative Hypothesis (H_1)

In the above three example research problems, a *treatment* is applied to the members of a sample. In the first example, the treatment is a daily dose of alcohol given to pregnant rats. In the second, rats living in a cool environment is the treatment. In the third example, each person is given a drug for stress. These treatments are being tested to see if they affect the sample members. In the first example research problem, can we conclude that alcohol usage reduces the birth weight of the rats? In the second example, can we conclude that living in a cool environment will increase food consumption? In the third example, does the drug affect response time? In each of the examples the *null hypothesis* is that the treatment has a null or zero effect. We indicate this symbolically in the three examples as: H_0: $\mu = 20$ grams or the alcohol has no affect on birth weight; H_0: $\mu = 12$ grams or the cool temperature has no effect on the food consumption; and H_0: $\mu = 10$ seconds or the drug has no effect on response time. The *alternative hypotheses* are represented as H_1: $\mu < 20$ grams or the alcohol does reduce the average birth weight; H_1: $\mu > 12$ grams or the rats consume more food in the cooler environment; and H_1: $\mu \neq 10$ seconds or the drug affects response time. In each of these situations, μ represents the mean after the treatment has been applied.

Reaching a Decision

What is the probability of getting a sample of size $n = 50$ with a sample mean of $M = 18$ if the sample comes from a population with mean μ equal to 20? It is now that we turn to the results in Chapter 7 on the sampling distribution of the mean. We assume the null hypothesis to be true and calculate the probability of getting a sample mean of 18 or smaller from a population with $\mu = 20$. The distribution of M is normal with mean = 20 and standard error equal to $\sigma/\sqrt{n}$ or $4/\sqrt{50}$, which equals 0.57 as shown in Figure 8.1.

H_0: média pop ≥ v

é **equivalente** a

H_0: média pop = v

vs.

H_1: média pop < v

- Demonstração em
  - GATÁS, Reny R. (1978, p. 220-223) *Elementos de Probabilidade e Inferência*. SP: Atlas.

```
pnorm(q=18, mean=20, sd=0.57, lower.tail=TRUE)
[1] 0.0002250904
```

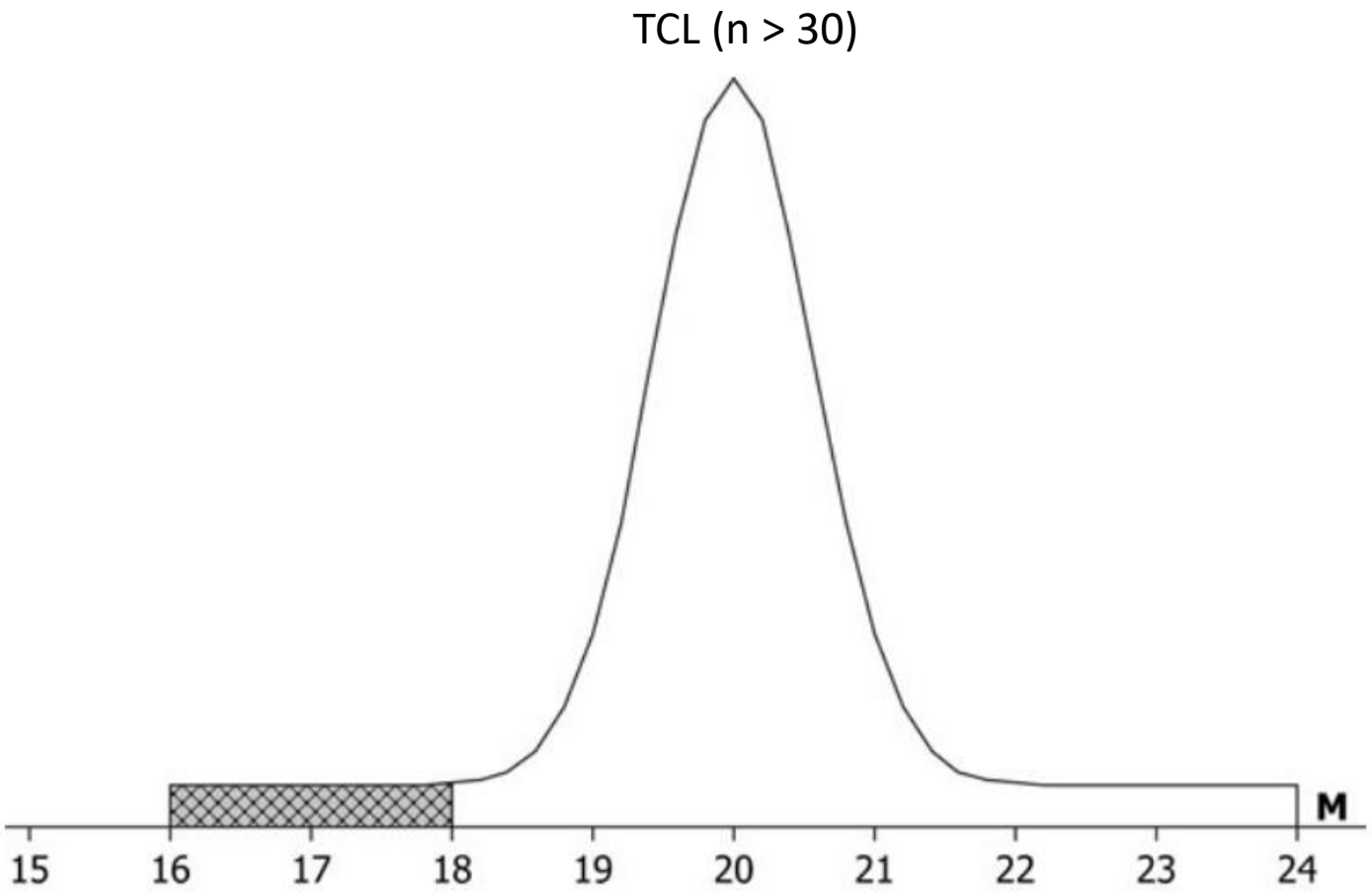


Figure 8.1 Assuming zero affect of the treatment, the left tail area is 0.0002251.

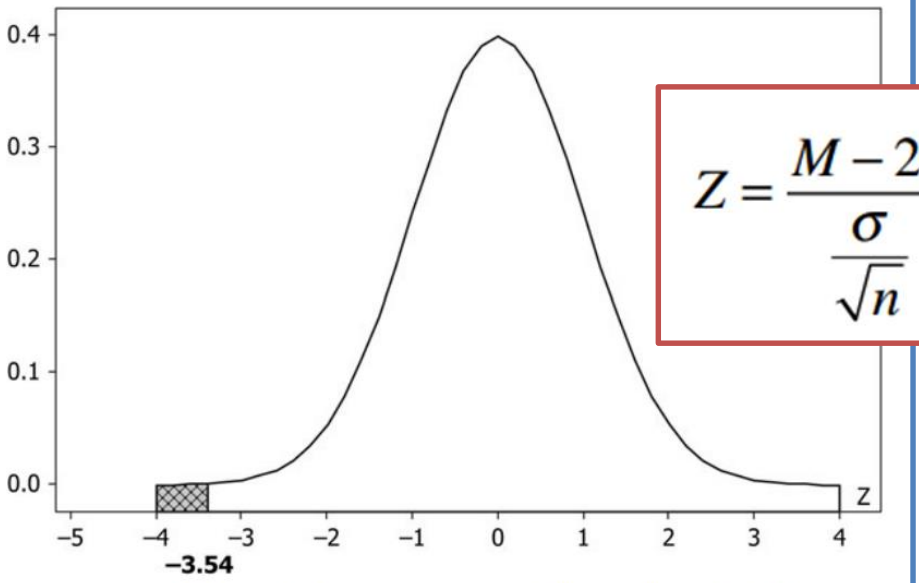


Figure 8.2 Standard normal curve with the p-value = 0.0002.

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

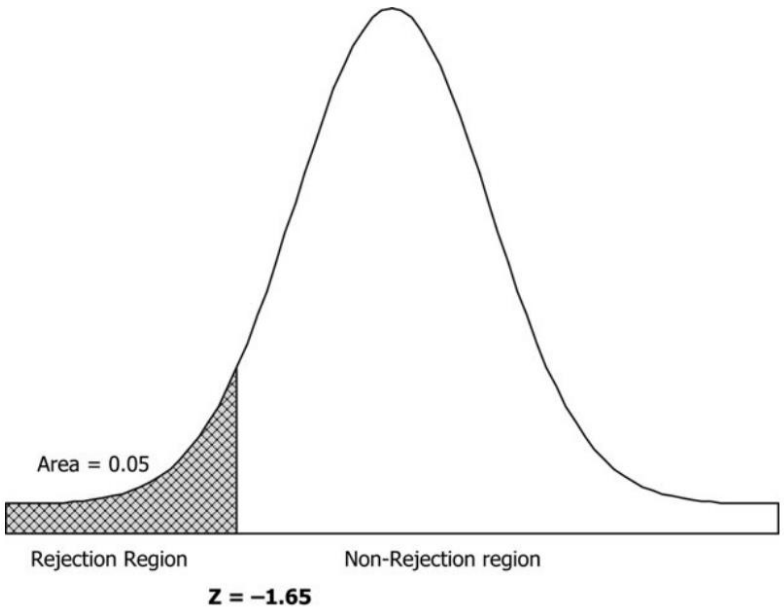
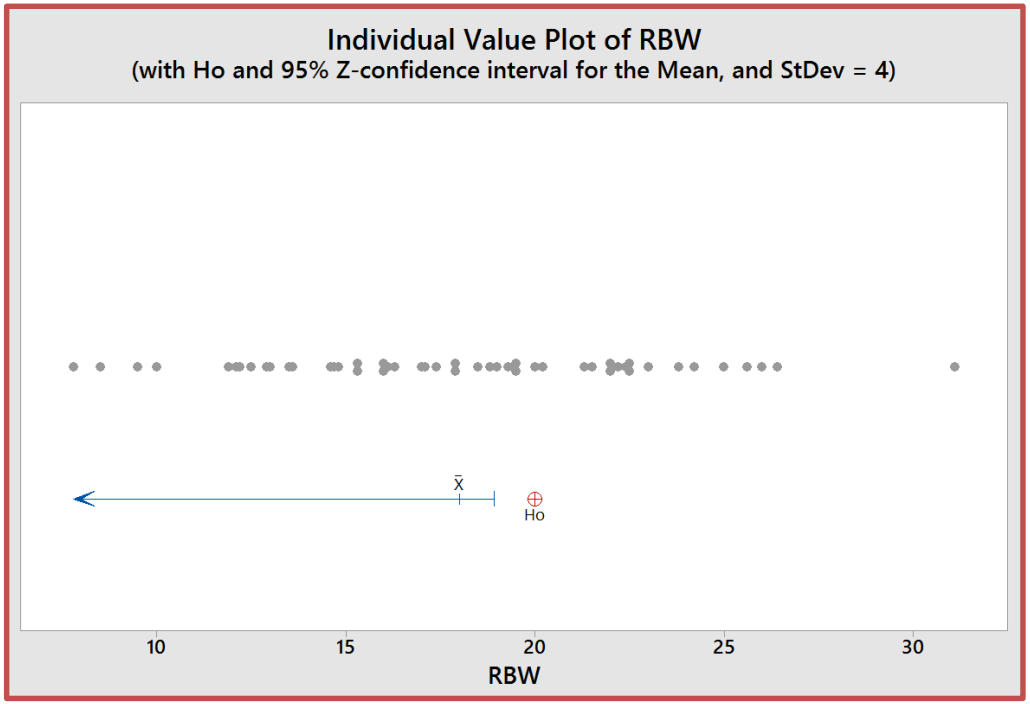


Figure 8.5 Critical value −1.65 is such that area = 0.05 is in the left tail.

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)
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
```



Teste Z unilateral
Valor-p unilateral (greater)

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

Testar se a média populacional μ é menor ou igual a μ_0 = 177cm hipotetizada pelo pesquisador.

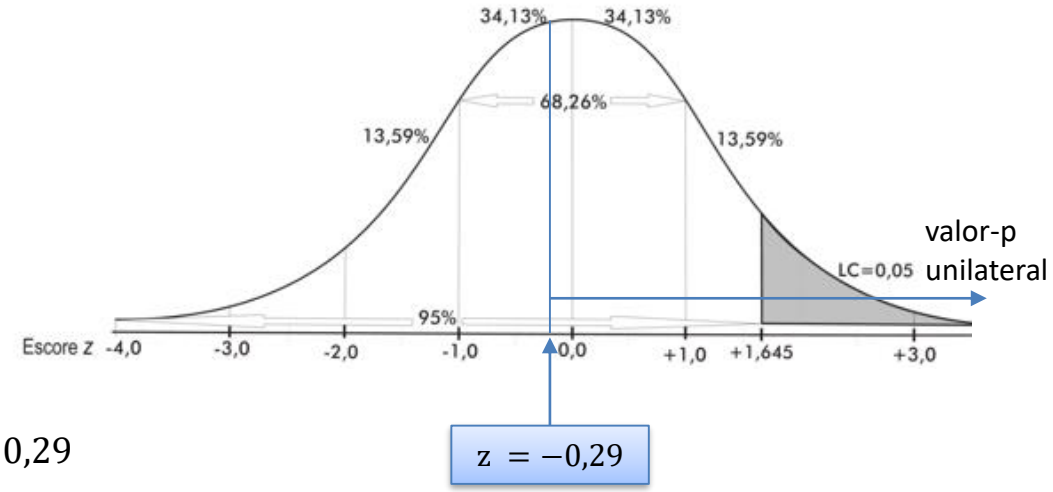
Quatro participantes desse grupo tiveram suas estaturas medidas, cujos valores em centímetro são: 175,186,169,174.

Hipóteses

- H_0: μ = 177
- H_1: μ > 177

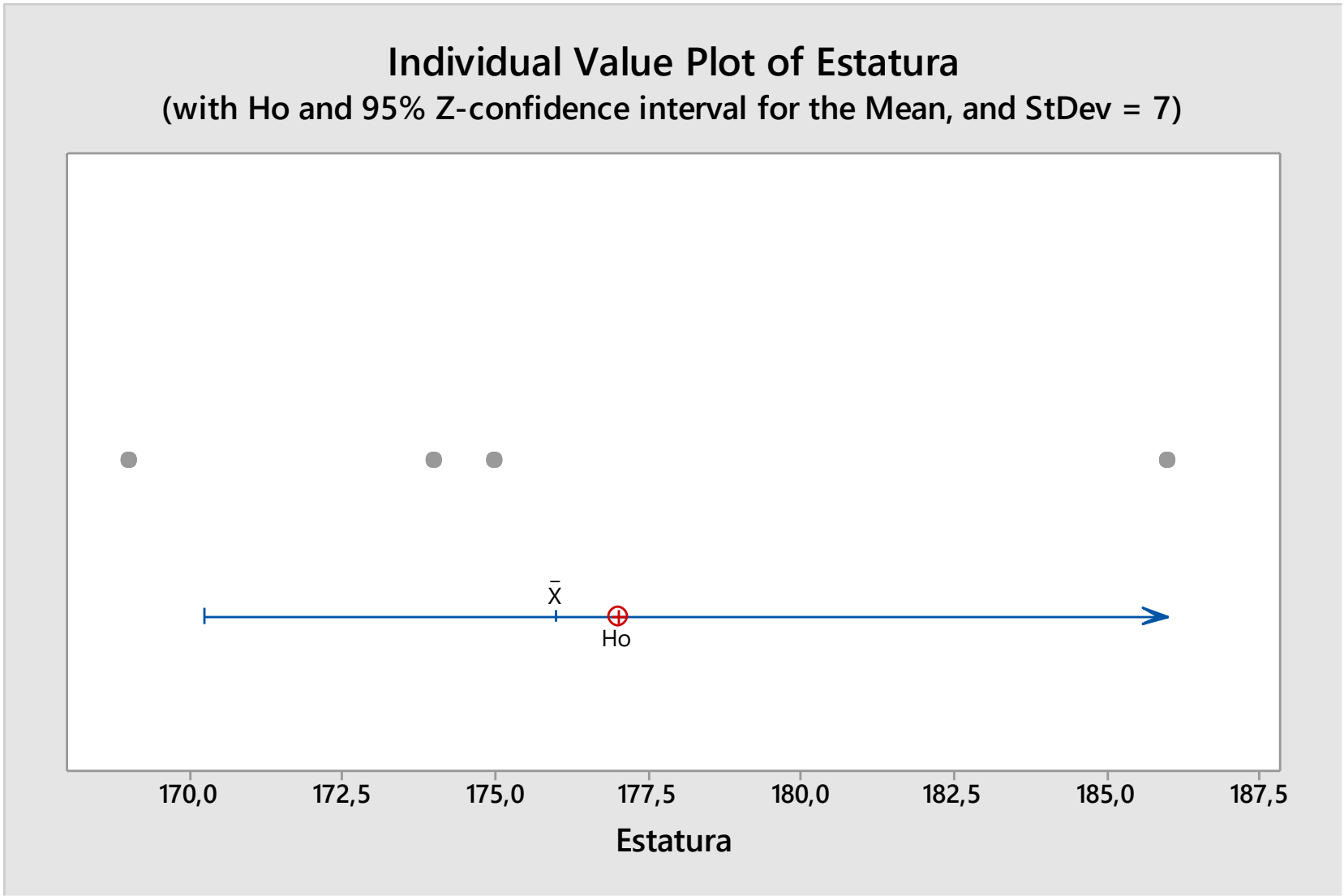
Estatísticas

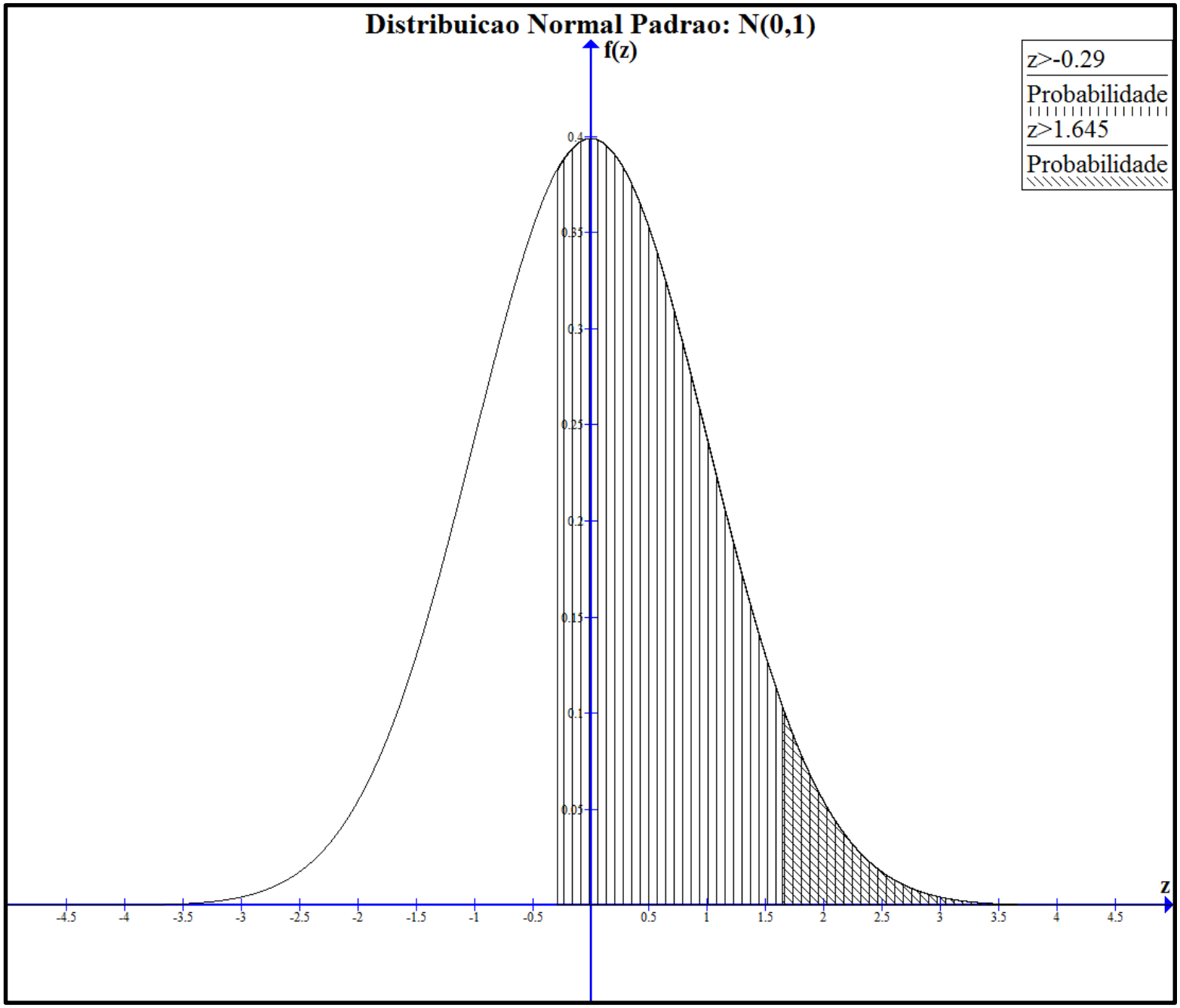
- $\bar{X}$ = (175+186+169+174)/4 = 176
- EP = $\frac{\sigma}{\sqrt{n}} = 3{,}5$
- IC95(μ) = [176-1,64×3,5; ∞]
  = [170,24; ∞]
- Estatística de teste z = $\frac{\bar{X}-177}{EP} = \frac{176-177}{3{,}5} = -0{,}29$



Decisão

- Como z = -0,29 < 1,645, não rejeitar H_0 ou
- Como IC95 contém 177, não rejeitar H_0 ou
- Como a probabilidade de escores-z serem mais extremos à direita de -0,29, i.e., o valor-p unilateral = 0,614 (= pnorm(-0.29, mean=0, sd=1, lower.tail=FALSE)) é maior que 5%, não rejeitar H_0
- O valor-p unilateral NÃO é igual à metade do valor-p bilateral = 0,775; é maior que sua metade: 0,388.


Individual Value Plot of Estatura
(with Ho and 95% Z-confidence interval for the Mean, and StDev = 7)
X̄
Ho
170,0
172,5
175,0
177,5
180,0
182,5
185,0
187,5
Estatura


Distribuicao Normal Padrao: N(0,1)
f(z)
z>-0.29
Probabilidade
z>1.645
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

Teste z unilateral (greater) para uma condição em R

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

```
        One-sample z-Test

data:  estatura
z = -0.28571, p-value = 0.6125
alternative hypothesis: true mean is greater than 177
95 percent confidence interval:
 170.243      NA
sample estimates:
mean of x
      176
```

Teste Z unilateral
Valor-p unilateral (less)

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

Testar se a média populacional μ é menor ou igual a μ_0 = 177cm hipotetizada pelo pesquisador.

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

Hipóteses

- H_0: μ = 177
- H_1: μ < 177

Estatísticas

- $\bar{X}$ = (175+186+169+174)/4 = 176
- EP = $\frac{\sigma}{\sqrt{n}} = 3{,}5$
- IC95(μ) = [0; 176+1,64×3,5]
  = [0; 181,76]
- Estatística de teste z = $\frac{\bar{X}-177}{EP} = \frac{176-177}{3{,}5} = -0{,}29$

Decisão

- Como z = -0,29 > -1,645, não rejeitar H_0 ou
- Como IC95 contém 177, não rejeitar H_0 ou
- Como a probabilidade de escores-z serem mais extremos à esquerda de -0,29, i.e., o valor-p unilateral = 0,388 (= pnorm(-0.29, mean=0, sd=1, lower.tail=TRUE)) é maior que 5%, não rejeitar H_0
- O valor-p unilateral é igual à metade do valor-p bilateral = 0,776.

Teste z unilateral (less) para uma condição em R

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

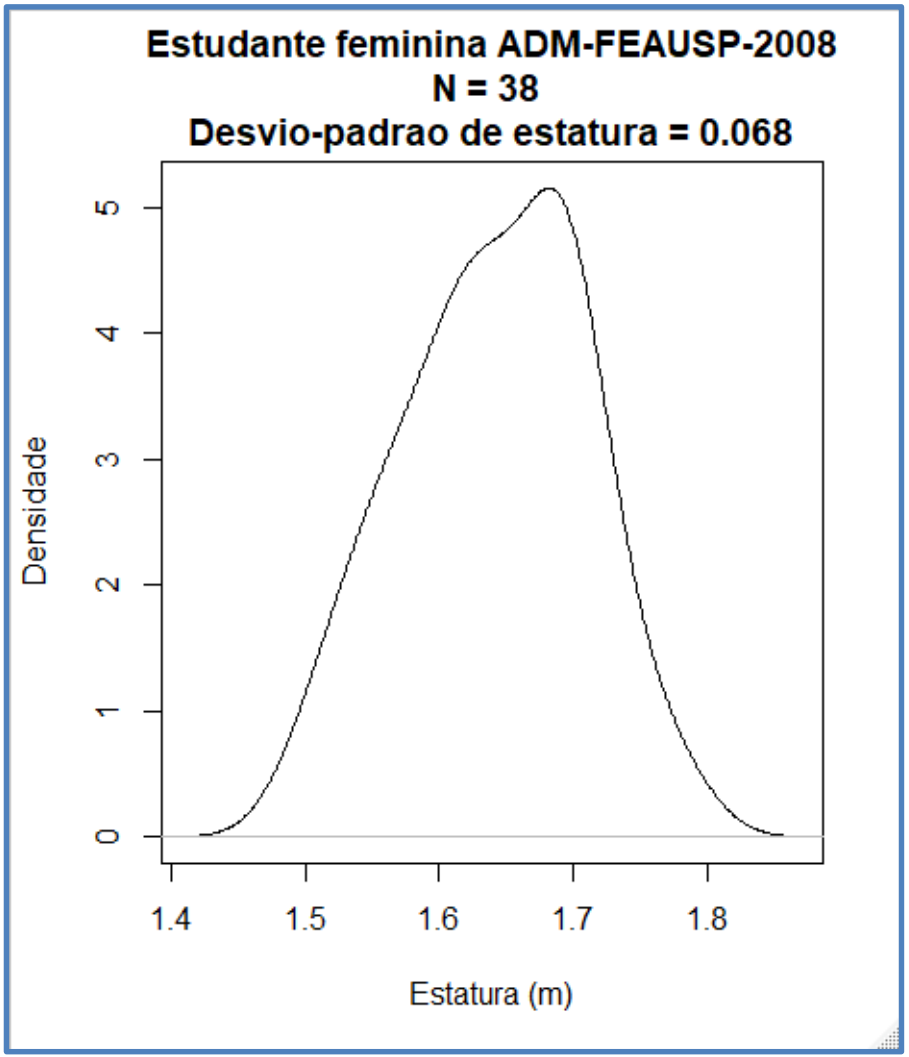
```
        One-sample z-Test

data:  estatura
z = -0.28571, p-value = 0.3875
alternative hypothesis: true mean is less than 177
95 percent confidence interval:
      NA 181.757
sample estimates:
mean of x
      176
```

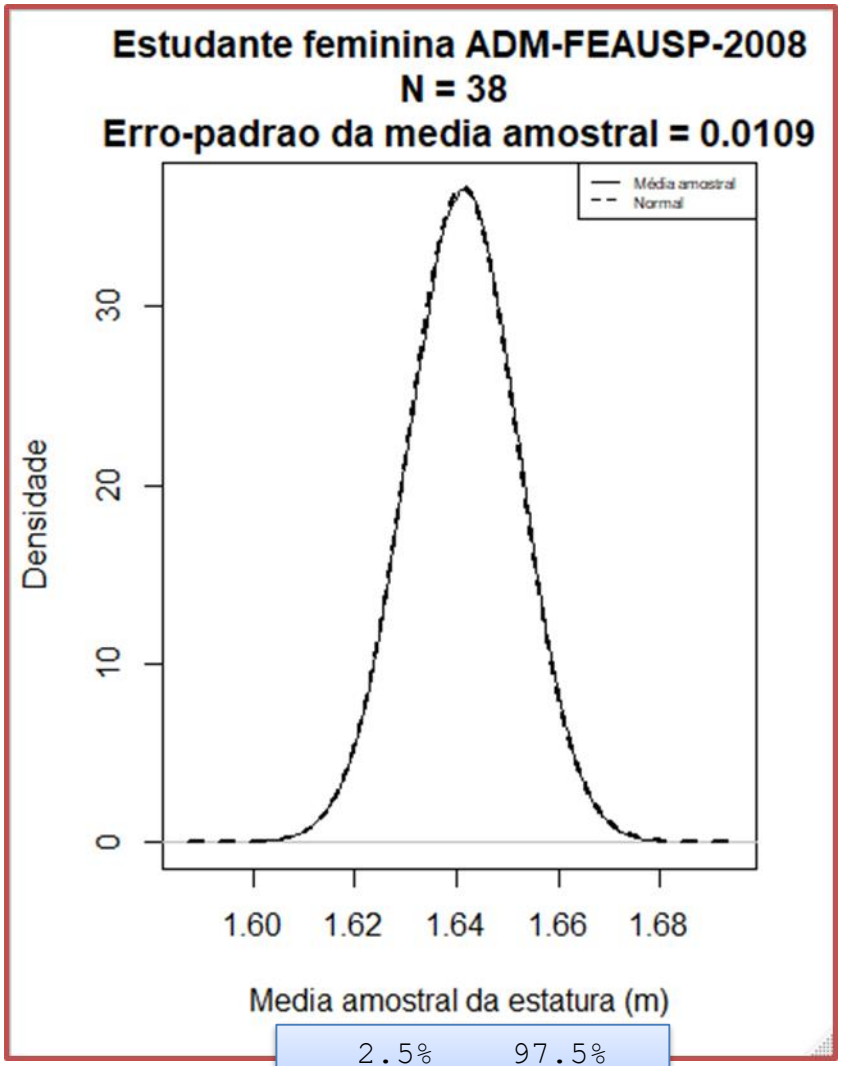
Reamostragem (*bootstrapping*) da média amostral em R

```
library(readxl)
B <- 1e6;  alfa <- 0.05; set.seed(123)
Dados <- readxl::read_excel("Adm2008.xlsx")
Dados <- Dados[, 1:4]
Matriz.Fem <- as.matrix(Dados[Dados$Genero=="Feminino", 3:4])
N.Fem <- nrow(Matriz.Fem)
plot(density(Matriz.Fem[,1], na.rm=TRUE),
     main=paste("Estudante feminina ADM-FEAUSP-2008\nN =", N.Fem,
                "\nDesvio-padrao de estatura =",
                round(sd(Matriz.Fem[,1], na.rm=TRUE), 4)),
     xlab="Estatura (m)", ylab="Densidade")
estat.media.boot.Fem <- replicate(B, mean(sample(Matriz.Fem[,1], replace=TRUE)))
print(mean(Matriz.Fem[,1], na.rm=TRUE))
print(mean(estat.media.boot.Fem, na.rm=TRUE))
quantile(estat.media.boot.Fem, probs=c(alfa/2, 1 - alfa/2))
t.test(Matriz.Fem[,1])$conf.int
plot(density(estat.media.boot.Fem, na.rm=TRUE),
     main=paste("Estudante feminina ADM-FEAUSP-2008\nN =", N.Fem,
                "\nErro-padrao da media amostral =",
                round(sd(estat.media.boot.Fem, na.rm=TRUE), 4)),
     xlab="Media amostral da estatura (m)", ylab="Densidade")
mi <- mean(estat.media.boot.Fem, na.rm=TRUE)
EP <- sd(estat.media.boot.Fem, 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=.5)
```

```
[1] 1.641316
> print(mean(Matriz.Fem[,1], na.rm=TRUE))
[1] 1.641316
> print(mean(estat.media.boot.Fem, na.rm=TRUE))
[1] 1.641323
> quantile(estat.media.boot.Fem, probs=c(alfa/2, 1 - alfa/2))
    2.5%    97.5%
1.620000 1.662368
> t.test(Matriz.Fem[,1])$conf.int
[1] 1.618968 1.663663
attr(,"conf.level")
[1] 0.95
```

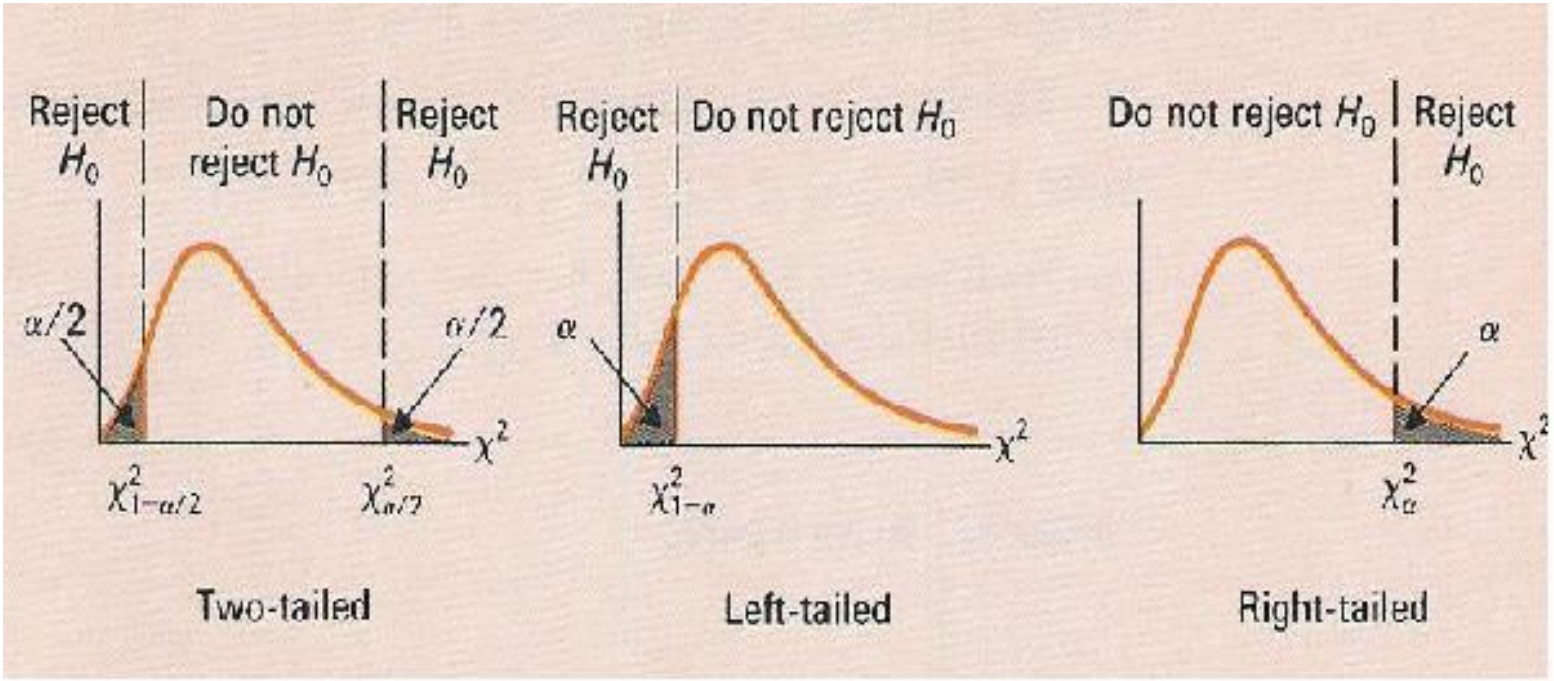
Reamostragem (*bootstrapping*) da média amostral em R






```
    2.5%     97.5%
1.620000 1.662368
```

Teste qui-quadrado de desvio-padrão populacional



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

Teste

- Desvio-padrão σ=7cm hipotetizado

Suposições

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

Hipóteses

- H_0: $\sigma = 7$
- H_1: $\sigma \neq 7$

Estatísticas

- S = 7,16
- GL = 4 – 1 = 3
- Estatística de teste qui-quadrado = 3,14
- IC95(σ) = [4,06; 26,70]

Decisão

- Como o desvio-padrão populacional hipotetizado está dentro do IC95, não rejeitar H_0
- Como a probabilidade dos valores da estatística de teste qui-quadrado serem mais extremos que 3,14, i.e., o valor-p bilateral = 0,741 é maior que 5%, não rejeitar H_0

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

```
library(EnvStats)
estatura <- c(169, 174, 175, 186)
out <- EnvStats::varTest(x=estatura, sigma.squared = 7^2,
        alternative="two.sided", conf.level = 0.95)
print(out)
print(sqrt(out$conf.int)))
```

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

Null Hypothesis:                 variance = 49

Alternative Hypothesis:          True variance is not equal to 49

Test Name:                       Chi-Squared Test on Variance

Estimated Parameter(s):          variance = 51.33333

Data:                            estatura

Test Statistic:                  Chi-Squared = 3.142857

Test Statistic Parameter:        df = 3

P-value:                         0.7402391

95% Confidence Interval:         LCL =  16.4734
                                 UCL = 713.6393

> print(sqrt(out$conf.int))
       LCL        UCL
 4.058744 26.714029
attr(,"conf.level")
[1] 0.95
```

Teste qui-quadrado unilateral (greater) de desvio-padrão populacional para uma condição em R

Teste

- Desvio-padrão σ = 7cm hipotetizado

Suposições

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

Hipóteses

- H_0: $\sigma = 7$
- H_1: $\sigma > 7$

Estatísticas

- S^2 = 7,16
- GL = 4 – 1 = 3
- Estatística de teste qui-quadrado = 3,14
- IC95(σ) = [4,44;∞]

Decisão

- Como o desvio-padrão populacional hipotetizado está dentro do IC95, não rejeitar H_0
- Como a probabilidade dos valores da estatística de teste qui-quadrado serem mais extremos que 3,14, i.e., o valor-p unilateral à direita = 0,371 é maior que 5%, não rejeitar H_0

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

```
library(EnvStats)
estatura <- c(169, 174, 175, 186)
out <- EnvStats::varTest(x=estatura, sigma.squared = 7^2,
                         alternative="greater", conf.level = 0.95)
print(out)
print(sqrt(out$conf.int))
```

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

Null Hypothesis:                 variance = 49

Alternative Hypothesis:          True variance is greater than 49

Test Name:                       Chi-Squared Test on Variance

Estimated Parameter(s):          variance = 51.33333

Data:                            estatura

Test Statistic:                  Chi-Squared = 3.142857

Test Statistic Parameter:        df = 3

P-value:                         0.3701195

95% Confidence Interval:         LCL = 19.70638
                                 UCL =      Inf

> print(sqrt(out$conf.int))
     LCL      UCL
4.439187      Inf
attr(,"conf.level")
[1] 0.95
```

Teste qui-quadrado unilateral (less) de desvio-padrão populacional para uma condição em R

Teste

- Desvio-padrão σ = 7cm hipotetizado

Suposições

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

Hipóteses

- H_0: $\sigma = 7$
- H_1: $\sigma < 7$

Estatísticas

- S^2 = 7,16
- GL = 4 – 1 = 3
- Estatística de teste qui-quadrado = 3,14
- IC95(σ) = [0;20,92]

Decisão

- Como o desvio-padrão populacional hipotetizado está dentro do IC95, não rejeitar H_0
- Como a probabilidade dos valores da estatística de teste qui-quadrado serem mais extremos que 3,14, i.e., o valor-p unilateral à direita = 0,63 é maior que 5%, não rejeitar H_0

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

```
library(EnvStats)
estatura <- c(169, 174, 175, 186)
out <- EnvStats::varTest(x=estatura, sigma.squared = 7^2,
                         alternative="less", conf.level = 0.95)
print(out)
print(sqrt(out$conf.int))
```

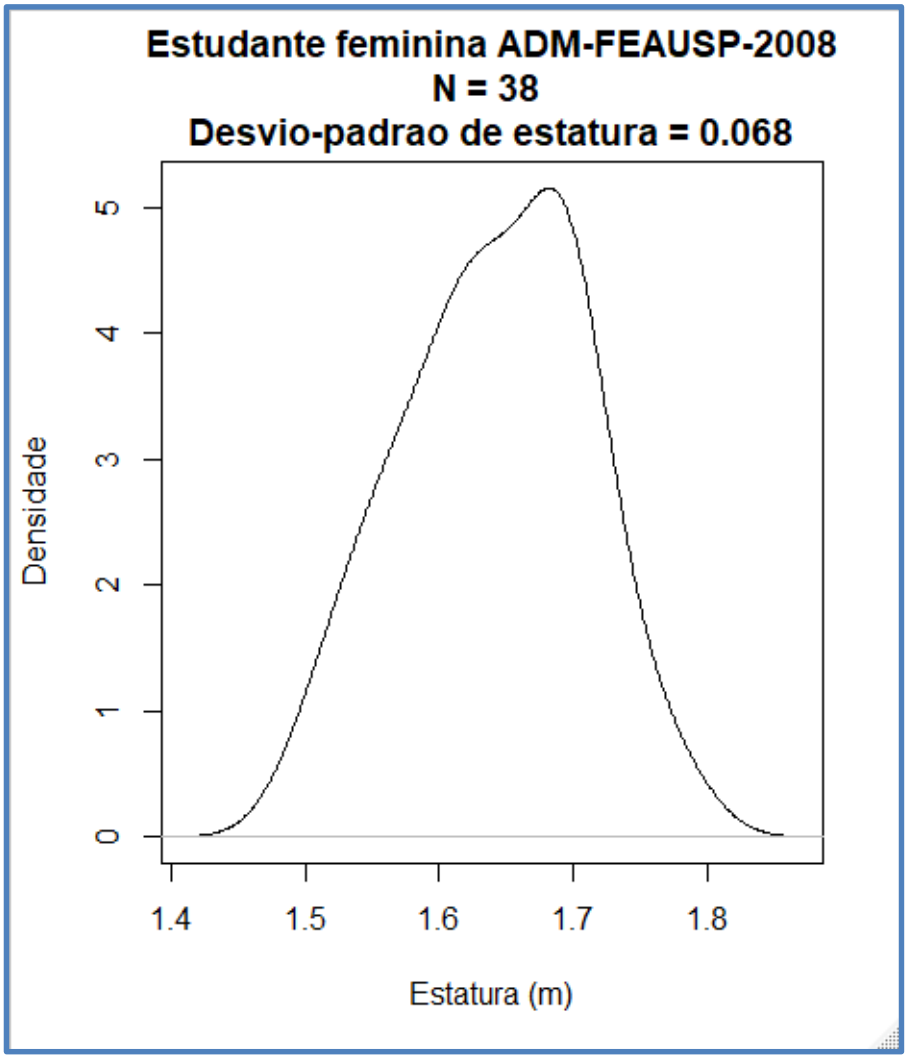
```
Results of Hypothesis Test
--------------------------

Null Hypothesis:                 variance = 49

Alternative Hypothesis:          True variance is less than 49

Test Name:                       Chi-Squared Test on Variance

Estimated Parameter(s):          variance = 51.33333

Data:                            estatura

Test Statistic:                  Chi-Squared = 3.142857

Test Statistic Parameter:        df = 3

P-value:                         0.6298805

95% Confidence Interval:         LCL =   0.0000
                                 UCL = 437.6911

> print(sqrt(out$conf.int))
     LCL      UCL
 0.00000 20.92107
attr(,"conf.level")
[1] 0.95
```

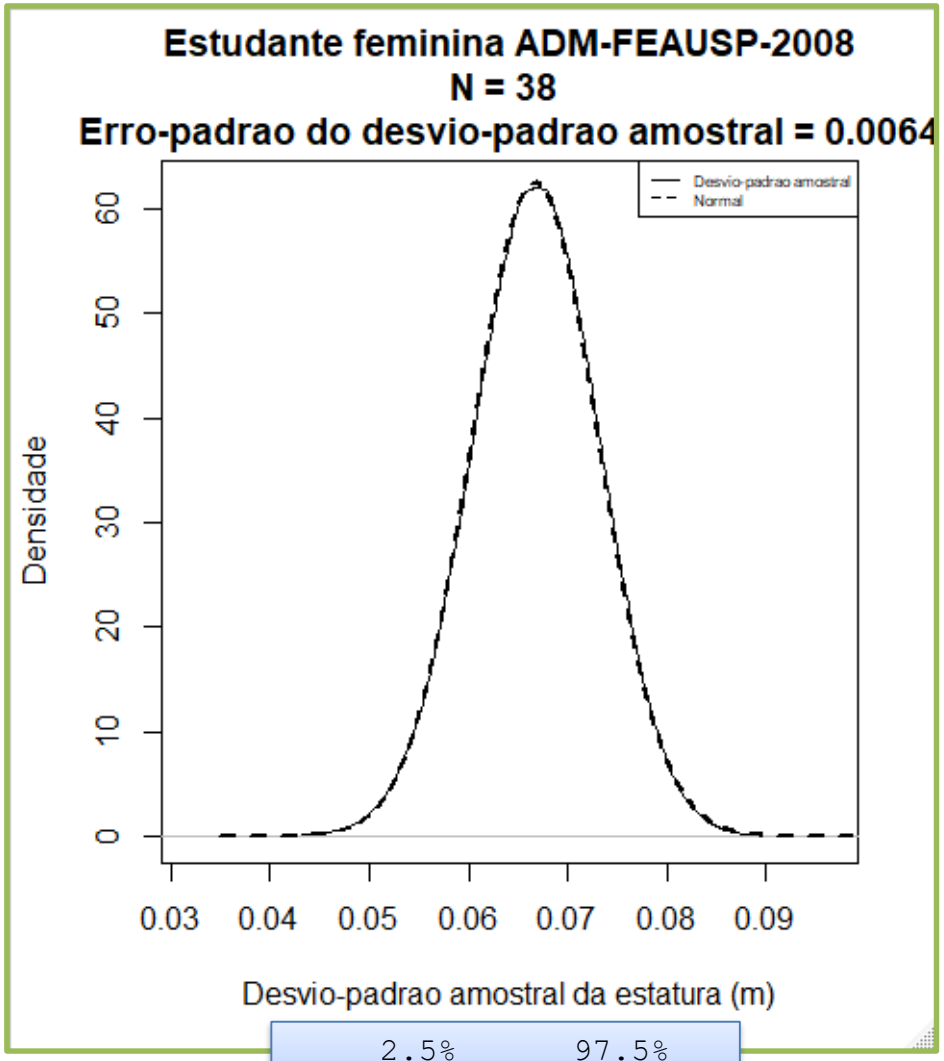
Reamostragem do desvio-padrão amostral em R

```
library(readxl)
library(EnvStats)
B <- 1e6; alfa <- 0.05; set.seed(123)
Dados <- readxl::read_excel("Adm2008.xlsx")
Dados <- Dados[, 1:4]
Matriz.Fem <- as.matrix(Dados[Dados$Genero=="Feminino", 3:4])
N.Fem <- nrow(Matriz.Fem)
plot(density(Matriz.Fem[,1], na.rm=TRUE),
     main=paste("Estudante feminina ADM-FEAUSP-2008\nN =", N.Fem,
                "\nDesvio-padrao de estatura =",
                round(sd(Matriz.Fem[,1], na.rm=TRUE), 4)),
     xlab="Estatura (m)", ylab="Densidade")
estat.media.boot.Fem <- replicate(B, sd(sample(Matriz.Fem[,1], replace=TRUE)))
print(sd(Matriz.Fem[,1], na.rm=TRUE))
print(mean(estat.media.boot.Fem, na.rm=TRUE))
quantile(estat.media.boot.Fem, probs=c(alfa/2, 1 - alfa/2))
ICDP <- EnvStats::varTest(Matriz.Fem[,1])$conf.int
cat("IC95%(DP.Fem) = [", round(sqrt(ICDP[1]),4), ";", round(sqrt(ICDP[2]),4), "]\n")
plot(density(estat.media.boot.Fem, na.rm=TRUE),
     main=paste("Estudante feminina ADM-FEAUSP-2008\nN =", N.Fem,
                "\nErro-padrao do desvio-padrao amostral =",
                round(sd(estat.media.boot.Fem, na.rm=TRUE), 4)),
     xlab="Desvio-padrao amostral da estatura (m)", ylab="Densidade")
mi <- mean(estat.media.boot.Fem, na.rm=TRUE)
EP <- sd(estat.media.boot.Fem, 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=.5)
```

```
> print(sd(Matriz.Fem[,1], na.rm=TR
[1] 0.06798931
> print(mean(estat.media.boot.Fem,
[1] 0.0667759
> quantile(estat.media.boot.Fem, pr
      2.5%      97.5%
0.05413486 0.07916033
> ICDP <- EnvStats::varTest(Matriz.
> cat("IC95%(DP.Fem) = [", round(sq
IC95%(DP.Fem) = [ 0.0554 ; 0.088 ]
```

Reamostragem do desvio-padrão amostral em R
Bootstrap_DesvioPadrao.R






```
     2.5%      97.5%
0.05413486 0.07916033
```


That's all Folks!

Introductory Statistics

THOMAS H. WONNACOTT

Associate Professor of Mathematics
University of Western Ontario

RONALD J. WONNACOTT

Professor of Economics
University of Western Ontario

1969

JOHN WILEY & SONS, INC.

New York · London · Sydney · Toronto

It can be proven, in general, that[8]

Não-rejeitada

> H_0 is accepted if and only if the relevant confidence interval contains H_0 (9-36)

[8] For a general algebraic proof (rather than geometric interpretation) for (9-36), consider the basis of both the confidence interval and hypothesis test. (We illustrate with the normal test of $\bar{X}$, but our remarks are equally valid for most tests.) With 95% probability,

$$\left| \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right| < 1.96 \tag{9-37}$$

In deciding whether to accept the null hypothesis μ_0, we first fix μ_0, and then see whether the observed $\bar{X}$ satisfies this inequality.

In constructing a confidence interval, we first observe $\bar{X}$; then the values of μ which satisfy (9-37) form our confidence interval. μ_0 will be in the confidence interval if and only if the hypothesis μ_0 is accepted, for in both cases we have

$$\left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| < 1.96$$

(c) The Confidence Interval as a General Technique

The reader may ask: "Doesn't (9-36) reduce hypothesis testing to a very simple adjunct of interval estimation?" In a sense this is true. Whenever a confidence interval has been constructed, it can immediately be used to test any null hypothesis: the hypothesis is accepted if and only if it is in the confidence interval. To emphasize this point, we can restate (9-36) in an equivalent form:

> A confidence interval may be regarded as just the set of acceptable hypotheses. (9-38)

9-4 THE RELATION OF HYPOTHESIS TESTS TO CONFIDENCE INTERVALS

(a) Two-sided Hypothesis Tests

In this section we shall reach a very important conclusion: a confidence interval can be used to test *any* hypothesis; in fact, the two procedures are equivalent. We illustrate with an example.

Suppose a firm has been producing a light bulb with an average life of 800 hours. It wishes to test a new bulb. A sample of 25 new bulbs has an average life of 810 hours ($\bar{X}$), with a standard deviation of 30 hours (s). Noting that because of our small sample we should use the t, rather than the normal distribution, we can either

1. Test the hypothesis

$$H_0: \mu_0 = 800 \tag{9-30}$$

against the alternative

$$H_1: \mu \neq 800 \tag{9-31}$$

H_0 may be accepted[7] at the 5% level of significance if

$$|\text{observed } t| = \left| \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \right| \leq t_{.025} \tag{9-32}$$

i.e., if $\mu_0 - 2.06\, s/\sqrt{n} \leq \bar{X} \leq \mu_0 + 2.06\, s/\sqrt{n}$.

Given our sample s, along with our hypothesis μ_0, this condition becomes

$$788 \leq \bar{X} \leq 812 \tag{9-33}$$

Since our observed $\bar{X}$ (810) does fall within this interval, μ_0 is acceptable. This is shown in Figure 9-9*a*.

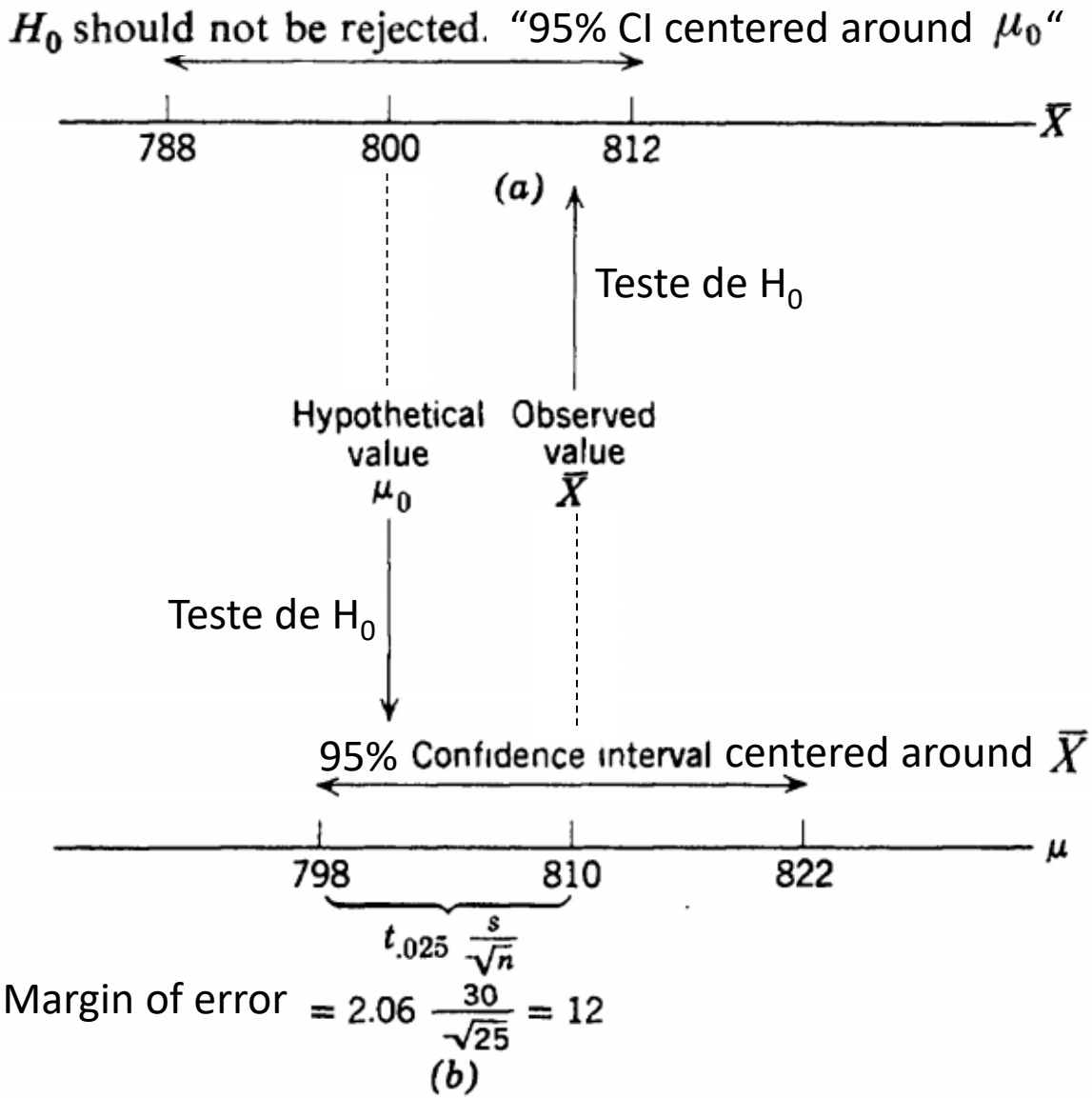


FIG. 9-9 Comparison of two-sided hypothesis test with confidence interval (using a sample with $\bar{X} = 810$ and $s = 30$). (*a*) Test of $H_0: \mu_0 = 800$ versus $H_1: \mu \neq 800$. (*b*) Confidence interval for μ.