

Trabalho de Inferência II

Testes de hipótese

Paulo Ricardo Seganfredo Campana

13 de junho de 2023

O arquivo `test.R` que está definido a função `testar()` pode ser encontrado no [Github](https://github.com/PauloCampana/UFPB/blob/main/P4/inf2/trabalho/test.R)

<https://github.com/PauloCampana/UFPB/blob/main/P4/inf2/trabalho/test.R>

```
library(dplyr)
library(broom)
library(knitr)
source("test.R")
t.test <- function(...) {
  stats::t.test(...) |>
  tidy() |>
  select(-method) |>
  kable(digits = 3)
}
var.test <- function(...) {
  stats::var.test(...) |>
  tidy() |>
  select(-method) |>
  kable(digits = 3)
}
chisq.test <- function(...) {
  stats::chisq.test(...) |>
  tidy() |>
  kable(digits = 3)
}
anova <- function(...) {
  stats::anova(...) |>
  tidy() |>
  kable(digits = 3)
}
```

Capítulo 10

Exemplo 10.1, 10.2, 10.4, 10.5 e 10.7: Está correta a afirmação do fabricante?

Consideremos uma companhia que produz cabos náuticos de determinado tipo. O fabricante garante que a carga de ruptura média de seu produto é de pelo menos 90 kg. Um potencial consumidor, interessado na compra de grande quantidade do produto, decide fazer ensaios de carga de ruptura com 20 espécimes, obtendo uma média amostral de 88.4 kg, supondo que o desvio padrão populacional da carga de ruptura seja conhecido e igual a 10 kg.

```
# H0:  $\mu \geq 90$ 
# H1:  $\mu < 90$ 
testar(
  c(m = 88.4, v = 10^2, n = 20),
  mu0 = 90,
  região_crítica = "menor"
)
```

estatística	p-valor	x-crítico	região crítica	tipo
-0.716	0.237	[86.322, Inf]	menor	z

Supondo sigma desconhecido, $S = 10$

estatística	p-valor	x-crítico	região crítica	tipo
-0.716	0.241	[86.134, Inf]	menor	t

Não rejeita a hipótese nula de que a carga de ruptura média de seu produto é de pelo menos 90 kg, pois o p-valor de 0.237 é maior que o ponto de corte tomado como $\alpha = 0.05$.

Exemplo 10.9

Suponha agora que o verdadeiro valor da média populacional é 89 kg, ou seja, um pouco inferior ao valor da carga média de ruptura afirmada pelo fabricante. Para calcularmos o poder do teste.

```
# poder =  $P(X < x_{crit}; \mu = 89)$ 
pnorm(86.322, 89, sqrt(10^2 / 20))
# [1] 0.115529
```

Exemplo 10.3: Baterias originais ou falsificadas?

A duração da carga das baterias recarregáveis de notebooks de uma certa marca pode ser encarada como uma variável aleatória com distribuição Normal de média 180 minutos e desvio padrão de 40 minutos. Existe uma falsificação do mesmo produto quase perfeita, mas nesse caso a duração da carga da bateria é uma variável aleatória com distribuição Normal de média 150 minutos e o mesmo desvio padrão anterior. Um montador de notebooks recebe 25 baterias dessa marca. Entretanto, ele tem dúvidas quanto à procedência das baterias. Como ele poderia decidir se as baterias são originais ou falsificadas?

Suponha que após submeter à prova as 25 baterias foi encontrada, para a duração de suas cargas, uma média amostral $\bar{X}_{\text{obs}} = 167.4$ min. Qual critério de decisão deve ser adotado se queremos fixar o valor de α em 0.05?

```
# H0:  $\mu = 180$ 
# H1:  $\mu = 150$ 
testar(
  c(m = 167.4, v = 40^2, n = 25),
  mu0 = 180,
  região_crítica = "menor"
)
```

estatística	p-valor	x-crítico	região crítica	tipo
-1.575	0.058	[166.841, Inf]	menor	z

Não rejeita a hipótese nula de que a duração da carga das baterias de notebooks tem média de 180 minutos, pois o p-valor de 0.058 é maior que o ponto de corte tomado como $\alpha = 0.05$.

Exemplo 10.6: Aumentando a pureza de um produto químico

Um produto químico tem seu teor de pureza distribuído conforme uma normal com média 0.72 e desvio padrão 0.02. A fim de aumentar a pureza, o produto é submetido a um tratamento. 16 unidades amostrais do produto são selecionadas de forma aleatória e submetidas a esse tratamento. Em seguida, a pureza de cada unidade é determinada obtendo-se, para elas, uma média aritmética de 0.73. Podemos dizer que o tratamento contribuiu para o aumento da pureza?

```
# H0:  $\mu = 0.72$ 
# H0:  $\mu > 0.72$ 
testar(
  c(m = 0.73, v = 0.02^2, n = 16),
  mu0 = 0.72,
  região_crítica = "maior"
)
```

estatística	p-valor	x-crítico	região crítica	tipo
2	0.023	$[-\text{Inf}, 0.728]$	maior	z

Podemos rejeitar a hipótese nula de que o produto químico tem teor de pureza igual a 0.72, pois o p-valor = 0.023 é menor que o ponto de corte tomado como $\alpha = 0.05$.

Exemplo 10.8

Determinemos o poder do teste para rejeitar corretamente a hipótese nula $H_0 : \mu = 0.72$ quando o verdadeiro teor médio de pureza do produto é $\mu = 0.725$. Usaremos o nível de significância $\alpha = 0.01$.

estatística	p-valor	x-crítico	região crítica	tipo
2	0.023	$[-\text{Inf}, 0.732]$	maior	z

```
# H0:  $\mu = 0.72$ 
# H1:  $\mu = 0.725$ 
# poder =  $P(X > \text{xcrit}; \mu = 0.725)$ 
pnorm(0.732, 0.725, sqrt(0.02^2 / 16), lower.tail = FALSE)
# [1] 0.08075666
```

Exemplo 10.10: Pentes de memória: estão corretas as especificações?

As especificações dos pentes de memória RAM para computadores fabricados pela Companhia Boa Memória indicam que a porcentagem de pentes defeituosos não excede 5%. Uma amostra de 100 desses pentes apresentou 7 defeituosos. Com base nesse resultado, podemos afirmar que as especificações estão incorretas?

- a) Qual a decisão a ser tomada, ao nível de significância de 0.01?
- b) Qual o nível crítico do teste?
- c) Qual o poder do teste, se $p = 0.12$?

```
# H0: p <= 0.05
# H1: p > 0.05
testar(
  c(m = 0.07, v = 0.05 * 0.95, n = 100),
  mu0 = 0.05, alpha = 0.01,
  região_crítica = "maior"
)
```

estatística	p-valor	x-crítico	região crítica	tipo
0.918	0.179	$[-\infty, 0.101]$	maior	z

Não rejeita a hipótese nula de que a porcentagem de pentes de memória RAM defeituosos não excedem 5%, pois o p-valor = 0.179 é maior que o ponto de corte tomado como $\alpha = 0.01$.

```
# poder = P(X > xcrit; p = 0.12)
pnorm(0.101, 0.12, sqrt(0.12 * (1 - 0.12) / 100), lower.tail = FALSE)
# [1] 0.7206201
```

R 10.1: Velocidade média na estrada

Sergio afirma que Raquel dirige seu carro na estrada a uma velocidade média superior a 100 km/h, enquanto Raquel discorda, afirmando dirigir na estrada a uma velocidade média menor ou igual a 100 km/h. Para dirimir essa controvérsia, Sergio resolve cronometrar o tempo (em minutos) que ela gasta ao volante em 10 viagens, sempre pelo mesmo percurso que liga duas cidades 120 km distantes: 73, 68, 73, 61, 70, 78, 63, 64, 74, 62.

- a) Quem parece ter razão, ao nível de significância de 5%?
- b) Qual o nível crítico?

```
# H0:  $\mu \leq 100$  km/h
# H1:  $\mu > 100$  km/h
R1 <- c(73, 68, 73, 61, 70, 78, 63, 64, 74, 62)
R1 <- 120 / (R1 / 60)
testar(
  c(m = mean(R1), v = var(R1), n = length(R1)),
  mu0 = 100,
  região_crítica = "maior", tipo = "t"
)
```

estatística	p-valor	x-crítico	região crítica	tipo
1.974	0.04	$[-\infty, 105.252]$	maior	t

Podemos rejeitar a hipótese nula de que Raquel dirige na estrada a uma velocidade média menor ou igual a 100 km/h, pois o p-valor = 0.040 é menor que o ponto de corte tomado como $\alpha = 0.05$.

R 10.2: Teste de hipótese simples com modelo Normal de variância conhecida

Queremos testar a hipótese $H_0 : \mu = 50$ contra a alternativa $H_1 : \mu = 52$, onde μ é a média populacional de uma distribuição normal com variância conhecida $\sigma^2 = 16$. Temos então n variáveis aleatórias i.i.d. X, X_2, \dots, X_n , que seguem todas essa lei de probabilidade, e a média amostral \bar{X} será usada como estatística de teste. Como $50 < 52$, é claro que quanto maior for X maiores serão as razões para se rejeitar H_0 , em favor de H_1 . Então o critério de decisão adequado será da forma: rejeitar H_0 , se $X_{\text{obs}} > \bar{x}_c$, e aceitar H_1 , se $X_{\text{obs}} < \bar{x}_c$; onde é uma constante a determinar em função de outras condições a serem especificadas.

- Se $n = 30$ e fixarmos $P(\text{Erro I})$ em $\alpha = 0.01$, quais devem ser o critério de decisão e $\beta = P(\text{Erro II})$?
- Mostre que se o tamanho da amostra for mantido em $n = 30$, ao fazermos com que o ponto de corte X se mova para a esquerda de forma que β diminua, α necessariamente aumentará.
- Quais devem ser o critério de decisão e o tamanho n da amostra para que a probabilidade do Erro II se reduza a $\beta = 0.05$, mantendo a probabilidade do Erro I em $\alpha = 0.01$?

```
# H0: μ = 50
# H1: μ = 52
# xcrit
xc <- qnorm(1 - 0.01, 50, sqrt(16 / 30))
xc
# [1] 51.69892
# beta
pnorm(xc, 52, sqrt(16 / 30))
# [1] 0.3400726
```

$$\begin{aligned} \begin{cases} 0.01 = P\left(Z > \frac{\bar{x}_c - 50}{\sqrt{16/n}}\right) \\ 0.05 = P\left(Z < \frac{\bar{x}_c - 52}{\sqrt{16/n}}\right) \end{cases} &\Rightarrow \begin{cases} \frac{(\bar{x}_c - 50)\sqrt{n}}{4} = 2.326 \\ \frac{(\bar{x}_c - 52)\sqrt{n}}{4} = -1.644 \end{cases} \Rightarrow \begin{cases} \bar{x}_c = 50 + \frac{9.305}{\sqrt{n}} \\ \bar{x}_c = 52 - \frac{6.579}{\sqrt{n}} \end{cases} \\ &\Rightarrow \begin{cases} \frac{15.884}{\sqrt{n}} = 2 \\ \bar{x}_c = 52 - \frac{6.579}{\sqrt{n}} \end{cases} \Rightarrow \begin{cases} \bar{x}_c = 51.17 \\ n = 63.08 \end{cases} \end{aligned}$$

R 10.4: Testando hipóteses em uma eleição para governador

Em uma pesquisa eleitoral referente ao primeiro turno de uma eleição para governador foram ouvidos $n = 1000$ eleitores selecionados aleatoriamente e entre eles $m = 510$ declararam-se favoráveis ao candidato A . Deseja-se testar a hipótese H_0 , de que a proporção p de eleitores do candidato A é menor ou igual a 0.5 contra a alternativa de que A venceria direto, sem a necessidade do segundo turno.

- a) Qual seria a sua decisão ao nível de significância de 5%? Por quê?
- b) Qual é o p-valor?
- c) Se na realidade $p = 0.55$, qual seria a probabilidade de ser cometido o Erro de tipo II ao ser usada essa mesma regra de decisão? Qual o poder do teste neste caso?

```
# H0: p <= 0.5
# H1: p > 0.5
testar(
  c(m = 0.51, v = 0.5 * 0.5, n = 1000),
  mu0 = 0.5, região_crítica = "maior"
)
```

estatística	p-valor	x-crítico	região crítica	tipo
0.632	0.264	[-Inf, 0.526]	maior	z

Não rejeita a hipótese nula de que a proporção de eleitores do candidato A é menor ou igual a 0.5, pois o p-valor = 0.264 é maior que o ponto de corte tomado como $\alpha = 0.05$.

```
# beta = P(X < xcrit; p = 0.55)
pnorm(0.526, 0.55, sqrt(0.55 * (1 - 0.55) / 1000))
# [1] 0.06356221
```


Capítulo 11

Exemplo 11.1: Comparação de dois catalisadores

Um engenheiro químico responsável por um processo produtivo quer verificar se o fato de se empregar um catalisador recentemente lançado no mercado em vez de empregar o catalisador que usa atualmente provoca um aumento significativo no rendimento do processo. Para isso, ele faz 8 ensaios com o catalisador atual (A) obtendo um rendimento médio de 80.5%, e 10 ensaios com o novo catalisador (B), obtendo rendimento médio de 81.3%.

Supondo que: os desvios padrões populacionais são conhecidos e iguais a 1.5% e 3.8%, respectivamente; e admitindo válidas as suposições de normalidade das v.a.'s envolvidas; a que conclusão se chegaria ao aplicar a esse problema um procedimento de teste de hipóteses?

```
# H0:  $\mu_x \geq \mu_y$ 
# H1:  $\mu_x < \mu_y$ 
testar(
  c(m = 80.5 - 81.3, v = (1.5^2/8 + 3.8^2/10) * (8 + 10), n = 8 + 10),
  região_crítica = "menor"
)
```

estatística	p-valor	x-crítico	região crítica	tipo
-0.609	0.271	[-2.16, Inf]	menor	z

Não rejeita a hipótese nula de que o uso do catalisador recentemente lançado no mercado provoca um aumento significativo no rendimento do processo comparado com o antigo, pois o p-valor = 0.271 é maior que o ponto de corte tomado como $\alpha = 0.05$.

Exemplo 11.2: Conteúdo de enxofre no carvão

Uma siderúrgica recebe carvão mineral de duas mineradoras como matéria-prima para a fabricação de aço. São obtidas aleatoriamente 50 unidades amostrais do produto fornecido pela mineradora A, para as quais se mede o conteúdo de enxofre. Com base nessas 50 medições calculam-se para essa variável uma média amostral de 0.61% e um desvio padrão amostral de 0.058%. Enquanto isso, 60 análises do carvão proveniente da mineradora B nos levam a uma média amostral de 0.68% de enxofre e um desvio padrão amostral de 0.065%. Podemos concluir, ao nível de significância de 1%, que o teor médio de enxofre no carvão proveniente da mineradora B é maior do que no carvão fornecido pela mineradora A?

```
# H0:  $\mu_x = \mu_y$ 
# H1:  $\mu_x < \mu_y$ 
testar(
  c(m = 0.61 - 0.68, v = (0.058^2/50 + 0.065^2/60) * (50 + 60), n = 50 + 60),
  alpha = 0.01,
  região_crítica = "menor"
)
```

estatística	p-valor	x-crítico	região crítica	tipo
-5.965	0	[-0.027, Inf]	menor	z

Podemos rejeitar a hipótese nula de que o teor médio de enxofre no carvão proveniente da mineradora B é igual do que no carvão fornecido pela mineradora A, pois o p-valor = 2.44e-09 é menor que o ponto de corte tomado como $\alpha = 0.01$.

Exemplo 11.3: Resistência à tração do concreto

Considere a comparação entre os concretos fabricados com cimentos das duas marcas diferentes. Os dados relativos às cargas de ruptura (em kg/cm^2) resultantes dos ensaios realizados são os seguintes:

Mistura1	15.4	15.7	14.8	15.7	14.8	15.6	15.4	14.6	15.8	15.3	15.5	15.2
Mistura2	14.7	14.3	15.4	14.5	14.2	15.0	14.8	15.2	15.3	14.9	NA	NA

Com base nesses dados, e supondo que ambas as v.a.'s são normais:

- Teste a hipótese de médias iguais contra médias diferentes relativas às duas misturas, ao nível de significância de 5%.
- Qual é o p-valor neste caso?

```
# H0:  $\mu_x = \mu_y$   
# H1:  $\mu_x \neq \mu_y$   
t.test(mistura1, mistura2)
```

estimate	estimate1	estimate2	statistic	p.value	parameter	conf.low	conf.high	alternative
0.487	15.317	14.83	2.814	0.011	18.983	0.125	0.849	two.sided

Podemos rejeitar a hipótese nula de que as cargas de ruptura das amostras de concreto são idênticas, pois o p-valor = 0.011 é menor que o ponto de corte tomado como $\alpha = 0.05$.

Exemplo 11.4: Conteúdo de oxigênio no cascalho

Consideremos. Seja duas porções idênticas de cascalho coletadas a uma dada profundidade; uma das porções foi analisada pelo Laboratório A, e a outra, pelo Laboratório B, ambos usando o mesmo procedimento. A amostra consiste em 25 pares dessas porções, em que cada par é coletado a uma profundidade diferente. A característica a ser medida é o conteúdo de oxigênio no material coletado. Podemos afirmar que, em média, as medições feitas no Laboratório A tendem a resultar maiores do que as feitas no Laboratório B?

O valor da média amostral da diferença é $\bar{\delta} = 0.27$ e o desvio padrão $s_{\Delta} = 0.20$

```
# H0:  $\mu_x \leq \mu_y$ 
# H1:  $\mu_x > \mu_y$ 
testar(
  c(m = 0.27, v = 0.20^2, n = 25),
  região_crítica = "maior", tipo = "t"
)
```

estatística	p-valor	x-crítico	região crítica	tipo
6.75	0	$[-\text{Inf}, 0.068]$	maior	t

Podemos rejeitar a hipótese nula de que as medições feitas no Laboratório A tem resultados iguais aos do Laboratório B, pois o p-valor = $2.78\text{e-}07$ é menor que o ponto de corte tomado como $\alpha = 0.05$.

Exemplo 11.5: Comparando qualidade de produto

No primeiro dia considerado, uma amostra aleatória de 50 peças foi selecionada da linha de produção, resultando em quatro peças não conformes. Suponha que a máquina que produz as peças sofre uma pane. Teme-se, portanto, que isso afete a qualidade do produto. Assim, uma nova amostra de 50 peças é selecionada no dia seguinte com o objetivo de verificar mais uma vez a proporção de peças não conformes. O número obtido nessa segunda amostra foi de seis peças não conformes. Com base nessa informação, podemos afirmar que a proporção de peças não conformes produzidas no segundo dia é superior à do primeiro dia? Use $\alpha = 0.05$.

Seja $\hat{p}_1 = 0.08$ e $\hat{p}_2 = 0.12$, portanto $\hat{p} = 0.1$

```
# H0: p1 >= p2
# H1: p1 < p2
testar(
  c(m = 0.08 - 0.12, v = 0.1 * (1 - 0.1) * (1/50 + 1/50) * (50 + 50), n = 50 + 50),
  região_crítica = "menor",
)
```

estatística	p-valor	x-crítico	região crítica	tipo
-0.667	0.252	[-0.099, Inf]	menor	z

Não rejeita a hipótese nula de que a proporção de peças não conformes produzidas no segundo dia é igual ou inferior a do primeiro dia, pois o p-valor = 0.252 é maior que o ponto de corte tomado como $\alpha = 0.05$.

Exemplo 11.6: Resistência à tração do concreto

Examinemos mais uma vez a comparação entre os concretos fabricados com cimentos das duas marcas, que foi vista no Exemplo 11.3. As cargas de ruptura obtidas experimentalmente são:

Mistura1	15.4	15.7	14.8	15.7	14.8	15.6	15.4	14.6	15.8	15.3	15.5	15.2
Mistura2	14.7	14.3	15.4	14.5	14.2	15.0	14.8	15.2	15.3	14.9	NA	NA

Com base nesses dados, e supondo que ambas as v.a.'s são normais, teste a hipótese nula de variâncias iguais contra a hipótese alternativa de variâncias diferentes, ao nível de significância de 5%.

```
# H0:  $\sigma_1 = \sigma_2$   
# H1:  $\sigma_1 \neq \sigma_2$   
var.test(mistura1, mistura2)
```

estimate	num.df	den.df	statistic	p.value	conf.low	conf.high	alternative
0.923	11	9	0.923	0.886	0.236	3.313	two.sided

Não rejeita a hipótese nula de que a variância das cargas de ruptura das amostras de concreto são idênticas, pois o p-valor = 0.886 é maior que o ponto de corte tomado como $\alpha = 0.05$.

Exemplo 11.7: Comparando três catalisadores

Um engenheiro químico deseja comparar três catalisadores analisando o tempo de reação de um determinado processo químico em cada um deles. Para isso, ele realiza 12 ensaios utilizando o catalisador A, 16 ensaios com o catalisador B e 16 com o catalisador C. Os tempos de reação (em minutos) observados nesses ensaios são apresentados a seguir:

A	48	51	45	53	46	62	44	53	36	40	50	40	NA	NA	NA	NA
B	42	60	47	54	59	53	46	54	36	56	40	54	41	46	47	58
C	47	40	32	36	45	36	39	50	28	52	32	36	41	39	42	28

Teste as hipóteses de médias iguais contra médias diferentes ao nível de significância $\alpha = 5\%$.

```
# H0:  $\mu_1 = \mu_2 = \mu_3$   
# H1:  $\mu_1 \neq \mu_2$  ou  $\mu_1 \neq \mu_3$  ou  $\mu_2 \neq \mu_3$   
lm(tempo ~ catalisador, E7) |> anova()
```

term	df	sumsq	meansq	statistic	p.value
catalisador	2	986.095	493.047	9.396	0
Residuals	41	2151.542	52.477	NA	NA

Podemos rejeitar a hipótese nula de que os tempos de reação dos catalisadores são iguais, pois o p-valor = 4.375e-4 é menor que o ponto de corte tomado como $\alpha = 0.05$.

Exemplo 11.8: Comparando três tipos de catalisador (cont.)

Queremos saber quais combinações dos três catalisadores apresentam diferenças entre si usando o método LSD de Fisher. Temos $MS_{\text{dentro}} = 52.48$

$$|\bar{Y}_i - \bar{Y}_j| > t_{(1-\alpha/2; N-m)} \sqrt{MS_{\text{dentro}}(1/n_i + 1/n_j)}$$

```
LSD <- function(i, j, alpha = 0.05) {  
  Yi <- E7$tempo[E7$catalisador == i]  
  Yj <- E7$tempo[E7$catalisador == j]  
  esquerda <- abs(mean(Yi) - mean(Yj))  
  direita <- qt(1 - alpha/2, df = 44 - 3) *  
    sqrt(52.48 * (1/length(Yi) + 1/length(Yj)))  
  esquerda > direita  
}  
E8 <- data.frame(grupos = c("AB", "AC", "BC"))  
E8$LSD <- sapply(  
  E8$grupos,  
  \(g) LSD(substr(g, 1,1), substr(g, 2,2))  
)  
E8
```

grupos	LSD
AB	FALSE
AC	TRUE
BC	TRUE

Ou seja, só ocorre diferença significativa entre os grupo A e C, e B e C

Exemplo 11.9: Comparando marcas de carros

Uma revista especializada decide comparar três marcas de carro (A, B, e C) da mesma categoria e na mesma faixa de preço, analisando diferentes itens. Para isso é realizada uma pesquisa entre usuários dos carros, os quais devem indicar o item que consideraram mais relevante ao comprar o carro. Os resultados são apresentados na tabela a seguir. Ao nível de significância de 5%, podemos afirmar que há alguma relação entre a marca do carro e o item considerado mais relevante pelo usuário?

	A	B	C
Maior capacidade do porta malas	68	31	20
Estética e acabamento	20	30	85
Desempenho	40	81	43
Outros	22	38	22

```
# H0: marca de carro e item mais relevante independentes
# H1: marca de carro e item mais relevante dependentes
chisq.test(E9)
```

statistic	p.value	parameter	method
110.404	0	6	Pearson's Chi-squared test

Podemos rejeitar a hipótese nula de que não há alguma relação entre a marca do carro e o item considerado mais relevante pelo usuário, pois o p-valor = 1.68×10^{-21} é menor que o ponto de corte tomado como $\alpha = 0.05$.

Exemplo 11.10: Acidentes de trânsito e dias da semana

A tabela a seguir apresenta a distribuição, por dia da semana, dos acidentes de trânsito ocorridos em 2008 no município de Ribeirão Preto. Os dados foram extraídos do site da Empresa de Trânsito e Transporte Urbano de Ribeirão Preto (Transerp).

segunda	terça	quarta	quinta	sexta	sabado	domingo
2416	2417	2442	2306	2669	2191	1355

Podemos afirmar, ao nível de significância de 5%, que os acidentes ocorrem com a mesma frequência em todos os dias da semana?

```
# H0: acidentes ocorrem com a mesma frequência
# H1: acidentes ocorrem com a diferentes frequências
probs <- rep(1/7, 7)
chisq.test(E10, p = probs)
```

statistic	p.value	parameter	method
476.479	0	6	Chi-squared test for given probabilities

Podemos rejeitar a hipótese nula de que os acidentes ocorrem com a mesma frequência em todos os dias da semana, pois o p-valor = 9.78×10^{-100} é menor que o ponto de corte tomado como $\alpha = 0.05$.

Exemplo 11.11: Número de defeitos e distribuição de Poisson

Os dados a seguir mostram o número de defeitos encontrados em 350 chapas de aço esmaltado de um metro quadrado cada.

Número de defeitos	0	1	2	3	4	5	6	7
Número de chapas	83	124	73	40	20	6	3	1

Testar, ao nível de significância de 5%, se o número de defeitos por metro quadrado segue uma distribuição de Poisson. Qual é o p-valor do teste?

```
# H0: número de defeitos segue distribuição Poisson
# H1: número de defeitos não segue distribuição Poisson
lambda <- sum(defeitos * chapas) / sum(chapas)
probs <- c(dpois(0:7, lambda), 1 - ppois(7, lambda))

chisq.test(c(chapas, 0), p = probs)
```

statistic	p.value	parameter	method
9.176	0.328	8	Chi-squared test for given probabilities

Não rejeita a hipótese nula de que o número de defeitos por metro quadrado segue uma distribuição de Poisson, pois o p-valor = 0.328 é maior que o ponto de corte tomado como $\alpha = 0.05$.

R 11.1: Comparação entre os graus de dificuldade de duas provas de Estatística

Um professor aplicou provas de Estatística a duas turmas no mesmo dia. Os resultados obtidos foram os seguintes:

M	2.5	3.5	4.5	5.0	5.5	5.5	6.0	6.5	7	7.5	7.5	7.5	7.5	8	8	8.5	9.0	9	10	NA
T	3.5	3.5	4.0	4.5	5.0	5.0	5.5	5.5	6	6.5	6.5	7.0	7.0	7	7	7.5	7.5	8	9	9.5

Depois de aplicadas as provas, alguns alunos da Tarde alegaram que a prova da Turma da Manhã tinha sido mais fácil e que, por isso, a Turma da Tarde havia sido prejudicada.

- Teste a hipótese nula de variâncias iguais contra a hipótese alternativa de variâncias diferentes.
- Teste a hipótese nula de médias iguais contra a hipótese alternativa de que a média da Turma da Manhã é maior que a da Turma da Tarde.

```
# H0:  $\sigma_M = \sigma_T$   
# H1:  $\sigma_M \neq \sigma_T$   
var.test(manhã, tarde)
```

estimate	num.df	den.df	statistic	p.value	conf.low	conf.high	alternative
1.356	18	19	1.356	0.516	0.533	3.493	two.sided

Não rejeita a hipótese nula de que as variâncias das notas das duas provas são iguais, pois o p-valor = 0.516 é maior que o ponto de corte tomado como $\alpha = 0.05$.

```
# H0:  $\mu_M = \mu_T$   
# H1:  $\mu_M > \mu_T$   
t.test(manhã, tarde, alternative = "greater", var.equal = TRUE)
```

estimate	estimate1	estimate2	statistic	p.value	parameter	conf.low	conf.high	alternative
0.513	6.763	6.25	0.875	0.194	37	-0.476	Inf	greater

Não rejeita a hipótese nula de que a média das notas das duas provas são iguais, pois o p-valor = 0.194 é maior que o ponto de corte tomado como $\alpha = 0.05$.

R 11.2: Será que a taxa de inflação tende a ser mais baixa nos países mais ricos do que nos países mais pobres?

Os diversos países do planeta foram divididos em dois grupos conforme o seu nível de renda per capita: Grupo A: renda per capita menor ou igual a 2.500 dólares, Grupo B: renda per capita acima de 2.500 dólares.

Queremos comparar as taxas médias de inflação nesses dois grupos de países, por meio de um Teste de Hipótese.

```
# H0:  $\mu_A = \mu_B$   
# H1:  $\mu_A > \mu_B$   
t.test(inflação ~ grupo, R2, alternative = "greater")
```

estimate	estimate1	estimate2	statistic	p.value	parameter	conf.low	conf.high	alternative
27.461	150.435	122.974	2.733	0.004	72.956	10.72	Inf	greater

Podemos rejeitar a hipótese nula de que as taxas médias de inflação dos dois grupos de países são iguais, pois o p-valor = 0.004 é menor que o ponto de corte tomado como $\alpha = 0.05$.

R 11.3: Testando a independência em tabelas de contingência 2x2

Dadas duas variáveis aleatórias discretas X e Y , queremos aplicar o teste Qui-quadrado para testar H_0 : “ X e Y são independentes” contra H_1 : “Existe dependência entre X e Y ”, com base na tabela de contingência a seguir:

$X \setminus Y$	b_1	b_2	Total
a_1	n_{11}	n_{12}	$n_{1\circ}$
a_2	n_{21}	n_{22}	$n_{2\circ}$
Total	$n_{\circ 1}$	$n_{\circ 2}$	n

Para isso será usada a estatística de teste $\sum_{h=1}^H \sum_{k=1}^K \frac{(n_{hk} - e_{hk})^2}{e_{hk}}$, onde $e_{hk} = \frac{n_{h\circ} \times n_{\circ k}}{n}$, para todo par (h, k) , com o critério de decisão usual. Quais das seguintes afirmações estão corretas e quais não estão? Por quê?

- a) Quanto mais próximo de zero estiver o valor da estatística de teste mais motivos se terá para rejeitar a hipótese de independência.
- b) Na expressão anterior e_{hk} representa o número esperado de observações com $X = a_h$, e $Y = b_k$ dados os totais de linha e de coluna, se houver independência total entre X e Y .
- c) Como só há 1 grau de liberdade, uma vez fixados $n_{1\circ}, n_{2\circ}, n_{\circ 1}$ e $n_{\circ 2}$, na montagem de uma tabela como a anterior somente uma das quatro frequências n pode variar livremente.
- d) Se H_0 é falsa, a probabilidade de que a estatística de teste seja inferior ao valor que se obtém da tabela do Qui-quadrado para 1 grau de liberdade e $\alpha = 0.05$ é igual a 0.95.
- e) O procedimento usual de teste é adequado, qualquer que seja o tamanho n da amostra.

- a) Não, pois para rejeitar H_0 com base na estatística do teste, a mesma deve ser alta
- b) Sim, e_{hk} representa o valor de n_{hk} esperado em H_0 .
- c) Sim, pois escolhido uma frequência, as outras podem ser achadas sabendo os totais.
- d) Não, isso é verdadeiro considerando H_0 verdadeira.
- e) Não, para amostras muito pequenas o teste é pouco definitivo.

R 11.4: Concurso público

Em um concurso público promovido por uma empresa estatal, os candidatos às vagas de Engenheiro Civil constituem a nossa população de interesse. Entre eles, os que se submeteram a uma preparação específica para o concurso constituem a subpopulação A, e os que não fizeram essa preparação constituem a subpopulação B. Sejam p_A , a probabilidade de aprovação para um candidato que se preparou e p_B a probabilidade de aprovação para um candidato que não se preparou. Deseja-se testar $H_0 : p_A = p_B$ contra $H_1 : p_A > p_B$. Para isso, foram coletadas amostras aleatórias em ambas as subpopulações, e os resultados obtidos foram os seguintes:

Subpopulação	Tamanho	Aprovados
A	100	34
B	200	43

Se n_A e n_B são os tamanhos de amostra utilizados, e se X_A e X_B , são as respectivas frequências de aprovados, pode ser usada, como alternativa ao que foi visto na teoria do

capítulo, a estatística de teste $Z = \frac{\frac{X_A}{n_A} - \frac{X_B}{n_B}}{\frac{1}{2} \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$.

- O que pode ser dito sobre a distribuição de probabilidade de Z sob H_0 ? Por quê?
- Qual a decisão a ser tomada ao nível de significância de 5% e o seu p-valor?

sob H_0 , Z tem distribuição normal, pois os tamanhos das amostras são suficientemente grandes para essa distribuição assintótica se aproximar da distribuição normal.

```
# H0: pA = pB
# H1: pA > pB
testar(c(
  m = 34/100 - 43/200,
  v = 1/4 * (1/100 + 1/200) * (100 + 200),
  n = 100 + 200
), região_crítica = "maior")
```

estatística	p-valor	x-crítico	região crítica	tipo
2.041	0.021	[-Inf, 0.101]	maior	z

Podemos rejeitar a hipótese nula de que a frequência de aprovados das duas subpopulações são iguais, pois o p-valor = 0.021 é menor que o ponto de corte tomado como $\alpha = 0.05$.

R 11.5: Tensão longitudinal de folhas de papel - ajuste de uma Normal

Suponha que a tensão longitudinal das folhas de papel de um determinado tipo produzidas por uma Companhia foi monitorada a cada 30 minutos, durante uma semana, sendo obtidos os resultados (N) apresentados na seguinte tabela de frequências.

Tensão longitudinal (N)	Frequência observada (n_i)
250 – 260	6
260 – 270	20
270 – 280	64
280 – 290	138
290 – 300	78
300 – 310	26
310 – 320	4
Total	336

Seja $\bar{x} = 285.6$ e $s = 11$, Usando um teste de aderência Qui-quadrado, testar ao nível de significância de 1% se as observações podem ser consideradas provenientes de uma população Normalmente distribuída.

```
# H0: tensão longitudinal segue distribuição normal
# H1: tensão longitudinal não segue distribuição normal
probs <- c(
  pnorm(250, mean = 285.6, sd = 11),
  sapply(seq(250, 310, by = 10),
    \(x) pnorm(x + 10, mean = 285.6, sd = 11) - pnorm(x, mean = 285.6, sd = 11)
  ),
  1 - pnorm(320, mean = 285.6, sd = 11)
)
chisq.test(c(0, 6, 20, 64, 138, 78, 26, 4, 0), p = probs)
```

statistic	p.value	parameter	method
9.467	0.304	8	Chi-squared test for given probabilities

Temos 9 intervalos nesse teste qui-quadrado, e usamos os dados amostrais para estimar dois parâmetros da distribuição, portanto o teste terá 6 graus de liberdade, o quantil para decisão do teste será $qchisq(0.99, df = 6) = 16.811$.

Não rejeita a hipótese nula de que a tensão longitudinal segue distribuição normal, pois a estatística do teste = 9.47 é menos extrema que o ponto de corte tomado como 16.811.

Capítulo 8

Example 8.1

For college-bound seniors in 2009, SAT math scores are normally distributed with a mean of 515 and a standard deviation of 116. You suspect that seniors in the town of Sodor are much brighter than the country as a whole, so you decide to conduct a test. In a random sample of 25 seniors, you find the mean SAT math score to be 555. You assume that the standard deviation of math scores in Sodor is the same as the national standard deviation of $\sigma = 116$. Is this sufficient evidence to conclude that Sodor seniors are smarter, or could a mean score of 555 be attributable to random variability?

```
# H0:  $\mu = 515$ 
# H1:  $\mu > 515$ 
testar(
  c(m = 555, v = 116^2, n = 25),
  mu0 = 515,
  região_crítica = "maior"
)
```

estatística	p-valor	x-crítico	região crítica	tipo
1.724	0.042	[-Inf, 553.161]	maior	z

Podemos rejeitar a hipótese nula de que a média na nota de matemática do teste SAT na cidade de Sodor é igual a média nacional, pois o p-valor = 0.042 é menor que o ponto de corte tomado como $\alpha = 0.05$.

Now, suppose that we are not willing to assume the variability of scores in Sodor is the same as the variability in national scores. We test the same hypotheses, but estimate the standard error from the data, using s/\sqrt{n} instead of σ/\sqrt{n} . Suppose $s = 120$.

estatística	p-valor	x-crítico	região crítica	tipo
1.667	0.054	[-Inf, 556.061]	maior	t

Não rejeita a hipótese nula de que a média na nota de matemática do teste SAT na cidade de Sodor é igual a média nacional, pois o p-valor = 0.054 é maior que o ponto de corte tomado como $\alpha = 0.05$.

Example 8.2

The coffee vending machine at work dispenses 7 oz of coffee in paper cups. The staff suspects that the machine is underfilling the cups. From a sample of $n = 15$ cups, they compute a mean of 6.6 oz and a standard deviation of 0.8 oz. Does this evidence support their suspicions? Assume that the coffee amounts dispensed by the machine are normally distributed.

```
# H0:  $\mu = 7$ 
# H1:  $\mu < 7$ 
testar(
  c(m = 6.6, v = 0.8^2, n = 15),
  mu0 = 7,
  região_crítica = "menor", tipo = "t"
)
```

estatística	p-valor	x-crítico	região crítica	tipo
-1.936	0.037	[6.636, Inf]	menor	t

Podemos rejeitar a hipótese nula de que a máquina de café dispensa em média 7 oz de café, pois o $p\text{-valor} = 0.037$ é menor que o ponto de corte tomado como $\alpha = 0.05$.

Example 8.3

About 13% of the population is left-handed. A biologist suspects that the scientific community is not like the general population in terms of handedness. He will conduct a study, and if the P-value from his test is less than 5%, he will conclude that the evidence supports his theory. He queries 200 scientists and finds that 36, or 18%, are left-handed. Does this data support the biologist's theory?

Aproximando o teste de proporções pela distribuição normal temos:

```
# H0: p = 0.13
# H1: p != 0.13
testar(
  c(m = 0.18, v = 0.13 * (1 - 0.13), n = 200),
  mu0 = 0.13
)
```

estatística	p-valor	x-crítico	região crítica	tipo
2.103	0.036	[0.083, 0.177]	bilateral	z

Ou utilizando um teste exato:

```
2 * (1 - pbinom(36-1, size = 200, prob = 0.13))
# [1] 0.05334989
```

Com o teste exato, não rejeita a hipótese nula de que a comunidade científica tem maior proporção de canhotos, pois o p-valor = 0.053 é maior que o ponto de corte tomado como $\alpha = 0.05$. Porém, aproximando pela normal podemos ter um resultado diferente para esse teste.

Example 8.11

A company claims that only 3% of people who use their facial lotion develops an allergic reaction (rash). You are a bit suspicious of their claims since you think a higher proportion of people at your college are allergic to this lotion. You query a random sample of 50 people and ask them to try the lotion. If more than three people develop the rash, you will send a nasty email to the company CEO. What is the probability that you make a type I error?

```
# α = P(X > 3; p = 0.03)
1 - pbinom(3, size = 50, prob = 0.03)
# [1] 0.06275993
```

Temos a probabilidade do erro tipo I de 0.062.

Example 8.14

An analyst draws a random sample of size 8, X_1, X_2, \dots, X_8 from a distribution with pdf $f(x; \theta) = (\theta + 1)x^\theta$, $0 \leq x \leq 1$, $\theta > 0$. She wants to test $H_0: \theta = 2$ versus $H_A: \theta > 2$. As a decision rule, she records Y , the number of observations greater than or equal to 0.88. She rejects H_0 if $Y \geq 5$. What is the probability of a type I error?

Sendo a regra de decisão $Y \geq 5$ em que Y é a soma de ensaios de Bernoulli, Y terá distribuição binomial de tamanho 8, resta saber a probabilidade de cara experimento.

$$p = P(X_i \geq 0.88; \theta = 2) \int_{0.88}^1 (\theta + 1)x^\theta dx = x^3 \Big|_{0.88}^1 = 1 - 0.88^3 = 0.318$$

```
# α = P(Y >= 5; p = 0.318)
# α = 1 - P(Y < 5; p = 0.318)
# α = 1 - P(Y <= 4; p = 0.318)
1 - pbinom(4, size = 8, prob = 1-0.88^3)
# [1] 0.07361352
```

Temos a probabilidade do erro tipo I de 0.073

Example 8.20

Suppose X_1, X_2, \dots, X_9 is drawn from the exponential distribution with pdf $f(x; \lambda) = \lambda e^{-\lambda x}$. To test $H_0: \lambda = 8$ versus $H_A: \lambda = 10$

$$L(\lambda) = \prod_{i=1}^9 \lambda e^{-\lambda X_i} = \lambda^9 e^{-\lambda \sum X_i}$$

$$\lambda(X) = \frac{8^9 e^{-8 \sum X_i}}{10^9 e^{-10 \sum X_i}} = 0.8^9 e^{2 \sum X_i} < k \implies \sum_{i=1}^9 X_i < c$$

Se X_i tem distribuição exponencial de parâmetro λ , então $\sum_{i=1}^n X_i$ tem distribuição Gama de parâmetros n e λ .

$$\alpha = 0.05 = P\left(\sum_{i=1}^9 X_i < c; \lambda = 8\right) \implies c = \text{qgamma}(0.05, 9, 8) = 0.596$$

Rejeitamos H_0 se $\sum_{i=1}^9 X_i < 0.596$.

Example 8.21

Let X_1, X_2, \dots, X_n be n independent Bernoulli trials with parameter p . Suppose we wish to test $H_0: p = 0.4$ versus $H_A: p = 0.5$.

$$L(p) = \prod_{i=1}^n p_i^X (1-p)^{1-X_i} = p^{\sum X_i} (1-p)^{n-\sum X_i} = (1-p)^n \left(\frac{p}{1-p} \right)^{\sum X_i}$$

$$\lambda(X) = \frac{(0.6)^n \left(\frac{2}{3} \right)^{\sum X_i}}{(0.5)^n (1)^{\sum X_i}} = 1.2^n (2/3)^{\sum X_i} < k \implies \sum_{i=1}^n X_i > c$$

Se X_i tem distribuição Bernoulli de parâmetro p , então $\sum_{i=1}^n X_i$ tem distribuição Binomial de parâmetros n e p .

$$\alpha = 0.05 = P \left(\sum_{i=1}^9 X_i > c; p = 0.4 \right) \implies c = \text{qbinom}(1 - 0.05, n, 0.4)$$

Para o caso $n = 10$ temos que $c = \text{qbinom}(1 - 0.05, 10, 0.4) = 7$, portanto rejeita H_0 se $\sum_{i=1}^9 X_i > 7$.