Resumo para P2 de Estatística

(2017)

Autor: Nathan Torquato
Aulas particulares: nathan.mimoso@usp.br ou Nathan Torquato (Facebook)

Sumário

1 Estimação de Parâmetros

Quando estamos interessados num determinado parâmetro de uma população, como sua média ou variância, tentaremos estimá-lo a partir dos dados amostrais. Vamos querer que esses estimadores sejam justos, consistentes e o mais eficientes possível, conforme já foi visto.

A estimativa pode ser pontual, indicando um valor para o parâmetro, ou, como é mais usual, pode consistir num intervalo, associado a uma certa probabilidade do valor real estar nele contido. Mas, antes de tratarmos dos intervalos de confiança, vejamos como se decide qual é o estimador mais adequado.

1.1. Critérios para escolha dos estimadores

Veremos dois deles:

1.1.1. Método da máxima verossimilhança

Nesse método, escolhemos o parâmetro de modo a maximizar a função de verossimilhança. Esse método é o mais utilizado, pois nos leva a estimadores

- justos
- consistentes
- assintoticamente eficientes
- de distribuição assintoticamente normal.

Dadas $X_1, X_2, \ldots, X_n$ com f.p. ou f.d.p. conjunta $f(x_1, x_2, \ldots, x_n; \theta_1, \ldots, \theta_m)$, onde os parâmetros $\theta_1, \ldots, \theta_m$ possuem valores desconhecidos. Essa f é o que chamamos de função de verossimilhança. As estimativas $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$ que maximizam a função de verossimilhança são as estimativas de máxima verossimilhança.

Exemplos

(I) Uma amostra de n elementos é extraída de uma população com distribuição binomial (n,p). Encontre o estimador de máxima verossimilhança (EMV) de p.

Tomando a função de probabilidade

$$f(x) = p^x (1-p)^{1-x}$$

A função de verossimilhança é

$$L(p) = \prod_{i=1}^{n} f(x) = p^{x_1} (1-p)^{1-x_1} \cdot \ldots \cdot p^{x_n} (1-p)^{1-x_n}$$

$$= p^{\sum_{i=1}^{n} x_i} \cdot (1-p)^{n - \sum_{i=1}^{n} x_i}$$

Queremos encontrar os valores de p que maximizam essa função. Para isso, basta fazer $\frac{\partial L(p)}{\partial p} = 0$ e ver o valor de p. Mas, para facilitar as contas, usaremos um artifício matemático: calcularemos $\ln(L(p))$:

$$\ln(L(p)) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

Derivando,

$$\frac{\partial(\ln(L(p)))}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

Igualando à zero e encontrando o estimador de máxima verossimilhança $\hat{p}$:

$$\frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0 \Rightarrow \frac{\sum_{i=1}^{n} x_i}{p} = \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

$$\Rightarrow (1-p)\sum_{i=1}^{n} x_i = p\left(n - \sum_{i=1}^{n} x_i\right) \Rightarrow$$

$$\Rightarrow \sum_{i=1}^{n} x_i - \cancel{p\sum_{i=1}^{n} x_i} = pn - \cancel{p\sum_{i=1}^{n} x_i} \Rightarrow$$

$$\Rightarrow \boxed{\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

(II) O parâmetro μ de uma população com distribuição de Poisson pode ser estimado por $\hat{\mu}$. Mostre.

A função $f(x) = \frac{\mu^x \cdot e^{-\mu}}{x!}$ origina a função de verossimilhança

$$L(\mu) = \prod_{i=1}^{n} \frac{\mu^{x_i} \cdot e^{-\mu}}{x_i!} = \frac{\mu^{\sum_{i=1}^{n} x_i} \cdot e^{-n\mu}}{\prod_{i=1}^{n} x_i!}$$

Ao aplicarmos o ln nela, ficamos com

$$\ln(L(\mu)) = \left(\sum_{i=1}^{n} x_i\right) \ln \mu - n\mu - \ln\left(\prod_{i=1}^{n} x_i!\right)$$

Derivando e igualando à zero:

$$\frac{\partial(\ln(L(\mu)))}{\partial \mu} = 0 \Rightarrow \frac{\sum_{i=1}^{n} x_i}{\mu} - n = 0 \Rightarrow$$

$$\Rightarrow \boxed{\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

(III) Numa população com distribuição exponencial, deseja-se estimar a taxa de ocorrência λ.

Tomando a função $f(x) = \lambda e^{-\lambda x}$, temos

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda\left(\sum_{i=1}^{n} x_i\right)}$$

Aplicando o ln:

$$\ln(L(\lambda)) = n \cdot \ln \lambda - \lambda \cdot \left(\sum_{i=1}^{n} x_i\right)$$

Ao derivar e igualar a zero, chegamos a:

$$\frac{\partial(\ln(L(\lambda)))}{\partial \lambda} = 0 \Rightarrow \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow$$

$$\Rightarrow \frac{n}{\lambda} = \sum_{i=1}^{n} x_i \Rightarrow \boxed{\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}}$$

- Passo a passo

  i. Escrever a f.p. ou f.d.p

  ii. Calcular $L(\theta) = \prod_{i=1}^{n} f(x_i)$

  iii. Fazer $\ln(L(\theta))$

  iv. Encontrar θ que satisfaz $\frac{\partial(\ln(L(\theta)))}{\partial\theta} = 0$

1.1.2. Método de Bayes

Esse método, usualmente, não cai na prova. Mas, em essência, ele é semelhante ao anterior. Sua grande diferença é levar em conta conhecimento prévio já existente. Isso permite trabalhar com amostras muito pequenas, fazendo esse método e toda a Inferência Bayesiana ganharem grande impulso recentemente.

1.2. Intervalos de confiança

Continuando a matéria da P1, vamos estudar os ICs que ainda não foram vistos.

1.2.1. IC para a variância

Queremos construir um intervalo de confiança ao nível $1-\alpha$, isto é, que possui $1-\alpha$ de probabilidade de conter o valor real do parâmetro, para a variância σ^2 da população.

Ele será dado por

$$\boxed{\frac{(n-1)\,s^2}{\chi^2_{n-1,\,\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)\,s^2}{\chi^2_{n-1,\,1-\alpha/2}}}$$

Exemplo

Uma amostra de 15 elementos retirada de uma população normalmente distribuída forneceu $s^2 = 2{,}56$. Construa um intervalo de 95% de confiança para a variância da população.

Consultando uma tabela estatística, encontramos que $\chi^2_{14;\,2,5\%} = 26,12$ e $\chi^2_{14;\,97,5\%} = 5,63$. Logo,

$$\frac{14 \cdot 2,56}{26,12} \leq \sigma^2 \leq \frac{14 \cdot 2,56}{5,63} \Rightarrow$$

$$\Rightarrow \boxed{1,372 \leq \sigma^2 \leq 6,366}$$

1.2.2. IC para o desvio padrão

De modo análogo ao caso anterior, o desvio padrão σ é dado por

$$\boxed{\sqrt{\frac{(n-1)s^2}{\chi^2_{n-1;\,\alpha/2}}} \leq \sigma \leq \sqrt{\frac{(n-1)\,s^2}{\chi^2_{n-1;\,1-\alpha/2}}}}$$

Exemplo

Usando os dados do exemplo anterior (amostra de 15 elementos com $s^2 = 2,56$), com 95% de confiança, podemos afirmar que

$$\sqrt{1,372} \leq \sigma \leq \sqrt{6,366} \Rightarrow$$

$$\Rightarrow \boxed{1,171 \leq \sigma \leq 2,523}$$

1.2.3. IC para a proporção populacional

Para uma variável aleatória com distribuição binomial, vimos que p é estimado por

$$\hat{p} = \frac{\sum x_i}{n}$$

Para uma amostra suficientemente grande ($n\hat{p} \geq 5$ e $n(1-\hat{p}) \geq 5$), podemos fazer a aproximação pela normal e

$$\boxed{z_{\frac{\alpha}{2}} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Exemplo

Uma moeda abaulada foi jogada 400 vezes, obtendo-se 136 caras. Construa um IC com 95% de confiança para a probabilidade do resultado cara nessa moeda.

Temos a probabilidade $\hat{p} = \frac{136}{400} = 0,34$ de obter cara num lançamento. Dado $\alpha = 5\%$, $z_{2,5\%} = 1,96$ e

$$0,34 - 1,96\sqrt{\frac{0,34 \cdot 0,66}{400}} \leq p \leq 0,34 + 1,96\sqrt{\frac{0,34 \cdot 0,66}{400}}$$

$\Rightarrow \boxed{p = 0,340 \pm 0,046}$

(2) Testes de Hipótese

Usamos os testes de hipótese para... Adivinhem... testar hipóteses, isto é, para poder decidir entre duas afirmações contraditórias. Essas afirmações são a hipótese nula (H_0) e a hipótese alternativa (H_1). Ao realizar o teste, nossa conclusão será por rejeitar ou não H_0, de acordo com a significância dos dados amostrais obtidos. Assim sendo, há dois erros que podemos cometer

- Rejeitar H_0, sendo H_0 verdadeira → chamado de erro tipo I
- Não rejeitar H_0, sendo H_0 falsa → chamado de erro tipo II

A probabilidade de cometer o erro tipo I (α) é denominada nível de significância.

2.1. Testes para a média

Dada uma variável normal X, analisada por meio de uma amostra n, suponhamos as hipóteses

$$H_0 : \mu \geq \mu_0$$
$$H_1 : \mu < \mu_0$$

Teremos que, fixado α, a região crítica (conjunto de valores de $\bar{x}$ que leva à rejeição de H_0) é

$$\bar{x} < \mu_0 - z_\alpha \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow$$

$$\Rightarrow \underbrace{\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}}_{z_{calculado}} < \underbrace{-z_\alpha}_{z_{crítico}}$$

Todo teste de hipótese consistirá na comparação de dois valores: um obtido a partir dos dados da amostra e outro consultado nas tabelas estatísticas.

Para o caso da média, de acordo com as hipóteses formuladas, teremos as seguintes situações:

- σ conhecido

A estatística de teste será

$$\boxed{Z_{calc} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$$

Os casos possíveis são

Hipóteses	Rejeitar H_0 se
$H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$	$Z_{calc} < -Z_{crit}$
$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$	$Z_{calc} > Z_{crit}$
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$\lvert Z_{calc} \rvert > Z_{crit} = Z_{\alpha/2}$

- σ desconhecido

Neste caso, usaremos

$$\boxed{t_{calc} = \frac{\bar{x} - \mu}{s/\sqrt{n}}}$$

A análise subsequente é análoga

Hipóteses	Rejeitar H_0 se
$H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$	$t_{calc} < -t_{n-1;\alpha}$
$H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$	$t_{calc} > t_{n-1;\alpha}$
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$\lvert t_{calc} \rvert > t_{n-1;\alpha/2}$

Exemplo

Um técnico de basquete está analisando a possibilidade de contratar um novo ala para sua equipe. Ele recebeu a informação de que este jogador tem uma média de pontos por partida superior a 16 e uma média de assistências por partida superior a 8. Para verificar esta informação, ele analisou os dados das últimas 10 partidas (ver tabela abaixo) com nível de significância de 5%. As afirmações sobre pontuação e assistências são aceitáveis?

Partida	Pontos	Assistências
1	18	11
2	19	12
3	13	14
4	16	17
5	23	8
6	16	15
7	13	15
8	19	13
9	13	13
10	19	13

- Para testar a afirmação sobre a pontuação, formulamos

$$H_0: \mu \leq 16$$
$$H_1: \mu > 16 \rightarrow$$ a hipótese alternativa é sempre a informação que queremos verificar

Da amostra, obtemos $n = 10$, $s = 3{,}315$, $\bar{x} = 16{,}9$.

Então, $t_{calc} = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{16{,}9 - 16}{3{,}315/\sqrt{10}} = 0{,}859$

Já $t_{crit} = t_{9;5\%} = 1{,}833$

Como $t_{calc} < t_{crit}$, não rejeitamos H_0, ou seja, a média de pontos não é superior a 16 ao nível de significância de 5%

- Analogamente para as assistências

$$H_0: \mu \leq 8$$
$$H_1: \mu > 8$$

Da amostra, $\bar{x} = 13{,}1$ e $s = 2{,}470$:

$$t_{calc} = \frac{13{,}1 - 8}{2{,}47/\sqrt{10}} = 6{,}529$$

$$t_{crit} = 1{,}833$$

Como $t_{calc} > t_{crit}$, rejeitamos H_0 e a média de assistências é maior que 8 ao nível de 5% de significância.

2.2. Teste para uma variância populacional

Os testes para a variância populacional são executados com a mesma lógica do anterior, mas usaremos χ^2 no lugar de z/t.

A estatística usada é

$$\chi^2_{calc} = \frac{(n-1)s^2}{\sigma_0^2}$$

↳ valor usado na formulação das hipóteses

Os limites da região crítica são $\chi^2_{n-1;1-\alpha}$ ou $\chi^2_{n-1;\alpha}$ (ou ambos, trocando α por $\alpha/2$)

Temos

Hipóteses	Rejeitar H_0 se
$H_0: \sigma^2 \leq \sigma_0^2$ $H_1: \sigma^2 > \sigma_0^2$	$\chi^2_{calc} > \chi^2_{n-1;\alpha}$
$H_0: \sigma^2 \geq \sigma_0^2$ $H_1: \sigma^2 < \sigma_0^2$	$\chi^2_{calc} < \chi^2_{n-1;1-\alpha}$
$H_0: \sigma^2 = \sigma_0^2$ $H_1: \sigma^2 \neq \sigma_0^2$	$\chi^2_{calc} < \chi^2_{n-1;1-\alpha/2}$ OU $\chi^2_{calc} > \chi^2_{n-1;\alpha/2}$

Exemplo

Com os dados da amostra abaixo é possível afirmar, ao nível de significância de 5%, que a variância populacional é superior a 12,4?

220	218	202
210	206	208
206	207	218

As hipóteses testadas devem ser

$$H_0: \sigma^2 \leq 12,4$$
$$H_1: \sigma^2 > 12,4$$

A partir da amostra, sabemos $n = 9$, $s^2 = 41,77$ e $\bar{x} = 210,6$. Rejeitaremos H_0 se $\chi^2_{calc} > \chi^2_{8,5\%}$. Vejamos:

$$\chi^2_{calc} = \frac{(n-1)\,s^2}{\sigma_0^2} = \frac{8 \cdot 41,77}{12,4} = 26,95$$

$$\chi^2_{8;5\%} = 15,507$$

$\therefore$ Rejeitamos H_0. Sim, ao nível de significância de 5%, a variância populacional é maior que 12,4.

2.3. Teste para uma proporção populacional

A proporção populacional pode ser estimada por $\hat{p} = \frac{\sum x_i}{n}$, conforme foi visto (se $np > 5$ e $n(1-p) > 5$). A estatística de teste é

$$\boxed{Z_{calc} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}}$$

Hipóteses	Rejeitar H_0 se
$H_0: p \geq p_0$ $H_1: p < p_0$	$Z_{calc} < -Z_{crit} = -Z_\alpha$
$H_0: p \leq p_0$ $H_1: p > p_0$	$Z_{calc} > Z_{crit}$
$H_0: p = p_0$ $H_1: p \neq p_0$	$\lvert Z_{calc} \rvert > Z_{\alpha/2}$

Exemplo

Uma empresa fez uma campanha de comunicação e deseja agora avaliar o desempenho desta iniciativa. Nos critérios atuais, uma campanha é considerada de sucesso se conseguir atingir, pelo menos, 30 % do seu público alvo. Em uma amostra de 40 indivíduos,

20 afirmaram que viram as peças publicitárias.

Ao afirmar que a campanha é um sucesso, qual a probabilidade de estar cometendo um erro?

O teste a ser realizado é

$$H_0: p \leq 0{,}3$$
$$H_1: p > 0{,}3$$

Neste caso, o que queremos saber é a probabilidade de rejeitar H_0, sendo H_0 verdadeira. Assim, buscamos o nível de significância do teste: o α.

Supondo $np \geq 5$ e $n(1-p) \geq 5$. H_0 será rejeitada se $Z > Z_\alpha$:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\overset{\hat{p} = 20/40}{0{,}5} - 0{,}3}{\sqrt{\frac{0{,}3 \cdot 0{,}7}{40}}} = 2{,}76 = Z_\alpha$$

Procurando na tabela, $P(0 \leq Z \leq Z_\alpha) =$

$P(0 \leq Z \leq 2{,}76) = 0{,}4971$

Logo, $\alpha = 0{,}5 - 0{,}4971 = 0{,}0029 = \boxed{0{,}29\%}$

2.4. Comparação de duas médias

Corresponde ao caso em que temos duas amostras, a princípio, de populações diferentes. A partir disso, compararemos parâmetros dessas populações e também poderemos verificar se as amostras podem ser provenientes da mesma população.

Usaremos a diferença das médias como base para nossos cálculos.

O primeiro ponto a ser checado é se os dados são emparelhados, isto é, relacionados dois a dois segundo algum critério que introduz uma influência marcante entre os diversos pares.

2.4.1. Dados emparelhados

Usamos $\Delta = \mu_1 - \mu_2$ como variável de teste. O resto, fazemos como se fosse uma média normal, lembrando de usar o desvio-padrão de Δ, simbolizado por s_Δ.

Exemplo

Um dispositivo para redução do consumo de combustível está sendo testado em veículos. Para cada veículo, foi medido o consumo com e sem o dispositivo durante uma semana de uso. Os dados estão abaixo. Ao nível de 5% de significância, é possível afirmar que o dispositivo reduz o consumo de combustível?

Veículo	Consumo médio (km/l) Sem disp.	Com disp.
1	8	17
2	15	19
3	11	15
4	7	17
5	11	16
6	12	17
7	22	16
8	14	14
9	11	14
10	15	14
Média	12,6	15,9
Desvio	4,25	1,66

Os dados, por associarem o consumo de um mesmo veículo com e sem o dispositivo, são emparelhados. Vamos usar a diferença neste teste:

Veículo	Δ
1	9
2	4
3	4
4	10
5	5
6	5
7	-6
8	0
9	3
10	-1

Média $\bar{\Delta} = 3,30$

Desvio $s_\Delta = 4,715$

Façamos o teste:

$$H_0: \Delta \leq 0$$
$$H_1: \Delta > 0$$

- $t_{calc} = \dfrac{3,30 - 0}{4,715/\sqrt{10}} = 2,213$

- $t_{crit} = t_{9;5\%} = 1,833$

Como $t_{calc} > t_{crit}$, rejeito H_0. Portanto, o consumo é menor com o uso do dispositivo.

2.4.2 Dados não-emparelhados

Neste caso, não faz sentido calcular as diferenças Δ. Além disso, as amostras poderão ter tamanhos diferentes. Essencialmente, há três casos distintos

- σ igual e conhecido

Usamos $$z_{calc} = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta}{\sigma\sqrt{1/n_1 + 1/n_2}}$$

n_1, n_2 ↓ tamanho das amostras

- σ igual e desconhecido

Calculamos $$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Usamos $$t_{calc} = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta}{\sqrt{s_p^2\left(1/n_1 + 1/n_2\right)}}$$

- σ desiguais e desconhecidos

Usamos $$t_{calc} = \frac{(\bar{x}_1 - \bar{x}_2) - \Delta}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

O número de graus de liberdade de t_{crit} será

$$\nu = \frac{(w_1 + w_2)^2}{w_1^2/(n_1 + 1) + w_2^2/(n_2 + 1)} - 2 ,$$

onde $w_1 = \frac{s_1^2}{n_1}$ e $w_2 = \frac{s_2^2}{n_2}$.

Exemplo

Um comprador deve escolher entre dois fornecedores de parafusos. A escolha do fornecedor será baseada na resistência dos parafusos. Uma amostra de cada fabricante foi analisada e os resultados obtidos estão na tabela abaixo.

	Amostra 1	Amostra 2
Tamanho	8	6
Média	56,38	64,5
Desvio padrão	0,74	3,15

Devemos verificar qual fornecedor entrega a maior resistência média:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 > \mu_2 \text{ ou } \mu_1 < \mu_2$$

Os dados não são emparelhados e não sabemos nada dos desvios-padrão: nem seu valor, nem se são iguais. Assim, usaremos o último método visto

- $t_{calc} = \dfrac{(\bar{x}_1 - \bar{x}_2) - \Delta}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \dfrac{56,38 - 61,5}{\sqrt{0,74^2/8 + 3,15^2/6}} = -6,187$

- $w_1 = \dfrac{s_1^2}{n_1} = \dfrac{0,74^2}{8} = 0,0925$ e $w_2 = 1,6538$

- $\nu = \dfrac{(0,0925 + 1,6538)^2}{0,0925^2/(8+1) + 1,6538^2/(8+1)} - 2 = 5,59 \Rightarrow 5 < \nu < 6.$

Tanto para $\nu = 5$ quanto para $\nu = 6$, teremos $t_{calc} < -t_{\nu,\alpha}$. Logo, H_0 será rejeitada e $\mu_2 > \mu_1$, pois sua média é maior.

2.5. Comparação de duas variâncias

Usaremos aqui a distribuição F, que possui a propriedade $F_{\nu_1;\nu_2;1-\alpha} = F^{-1}_{\nu_2;\nu_1;\alpha}$

A variável de teste será $F_{calc} = \dfrac{s_1^2}{s_2^2}$ ou $F_{calc} = \dfrac{\text{máx}(s_1^2, s_2^2)}{\text{mín}(s_1^2, s_2^2)}$, no caso de testar $\sigma_1^2 \neq \sigma_2^2$. Portanto,

Hipóteses	Rejeitar H_0 se
$H_0: \sigma_1^2 \leq \sigma_2^2$ $H_1: \sigma_1^2 > \sigma_2^2$	$F_{calc} > F_{n_1-1;\, n_2-1;\, \alpha}$
$H_0: \sigma_1^2 \geq \sigma_2^2$ $H_1: \sigma_1^2 < \sigma_2^2$	$F_{calc} < F_{n_1-1;\, n_2-1;\, 1-\alpha}$
$H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$	$F_{calc} > F_{\nu_i,\nu_j;\, \alpha/2}$

Exemplo

Duas amostras, com 10 e 15 elementos, extraídas de populações normais possuem $s_1^2 = 6,34$ e $s_2^2 = 18,7$, respectivamente. Com $\alpha = 5\%$, devemos aceitar que as variâncias são iguais?

As hipóteses são $H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

Temos que

$$F_{calc} = \frac{\text{máx}\,(s_1^2, s_2^2)}{\text{mín}\,(s_1^2, s_2^2)} = \frac{18,7}{6,34} = 2,950$$

Como $\nu_1 = 14$ e $\nu_2 = 9$,

$$F_{14;9;97,5\%} = F^{-1}_{9;14;2,5\%} = \frac{1}{3,21} = 0,312$$

$$F_{14;9;2,5\%} = 3,80$$

Portanto, não devemos rejeitar H_0.

Aceitamos que as variâncias são iguais.

2.6. Comparação de duas proporções

Sendo $H_0: p_1 - p_2 = \Delta$, a variável de teste será

$$Z_{calc} = \frac{(\hat{p}_1 - \hat{p}_2) - \Delta}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

que compararemos com $-z_\alpha$ ou z_α, conforme o caso anteriormente estudado.

Se formos testar a igualdade entre os parâmetros, então Z_{calc} é simplificada para

$$Z_{calc} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$