Teste para duas médias populacionais

$$H_0 : \mu_1 = \mu_2$$

versus

$$\begin{aligned} H_1 &: \mu_1 \neq \mu_2 && \text{teste bilateral} \\ H_1 &: \mu_1 > \mu_2 && \text{teste unilateral à direita} \\ H_1 &: \mu_1 < \mu_2 && \text{teste unilateral à esquerda} \end{aligned}$$

Nestes casos, observam-se duas amostras de duas populações com médias μ_1 e μ_2. Possíveis casos:

- Dados pareados ou amostras dependentes
- Dados não pareados ou amostras independentes

Teste para dados pareados

Exemplo: Foi conduzido um experimento para estudar o conteúdo de hemoglobina no sangue de suínos com deficiência de niacina. Aplicaram-se 20 mg de niacina em oito suínos. Podemos afirmar que o conteúdo de hemoglobina no sangue diminui com a aplicação de niacina, ao nível de significância de 5%?

Suíno	Antes (A)	Depois (B)
1	13,6	11,4
2	13,6	12,5
3	14,7	14,6
4	12,1	13,0
5	12,3	11,7
6	13,2	10,3
7	11,0	9,8
8	12,4	10,4

Hipóteses:

$$H_0 : \mu_A = \mu_B \quad \textit{versus} \quad H_1 : \mu_A \neq \mu_B$$

Ou ainda,

$$H_0 : \mu_D = 0 \quad \textit{versus} \quad H_1 : \mu_D \neq 0$$

Estatística

$$t = \frac{\bar{D} - 0}{\frac{S_D}{\sqrt{n}}} = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}},$$

a qual segue uma distribuição t de Student com $n-1$ graus de liberdade. $\bar{D}$ corresponde a média amostral da diferença entre os valores de A e B e S_D é o desvio padrão das diferenças.

Exemplo: Foi conduzido um experimento para estudar o conteúdo de hemoglobina no sangue de suínos com deficiência de niacina. Aplicaram-se 20 mg de niacina em oito suínos. Podemos afirmar que o conteúdo de hemoglobina no sangue diminui com a aplicação de niacina, ao nível de significância de 5%?

Suíno	Antes (A)	Depois (B)	Diferenças (D=A-B)
1	13,6	11,4	2,2
2	13,6	12,5	1,1
3	14,7	14,6	0,1
4	12,1	13,0	-0,9
5	12,3	11,7	0,6
6	13,2	10,3	2,9
7	11,0	9,8	1,2
8	12,4	10,4	2,0

$n = 8$

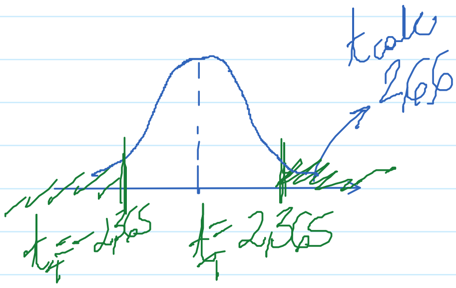
$$\begin{cases} H_0: \mu_D = 0 \\ H_a: \mu_D \neq 0 \end{cases}$$

$\bar{X}_D = 1,15$

$S_D = 1,22247$

$$t = \frac{\bar{D} - 0}{\frac{S_D}{\sqrt{n}}} = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}},$$

$$t = \frac{1,15}{\frac{1,22247}{\sqrt{8}}} = 2,66$$



$t_{tab} = 2,365$
95%, 7gl

Como tcalc > Ttab, rejeita-se Ho ao nível de 5% de significância, sendo assim, há evidências de que a aplicação de niacina diminui a hemoglobina no sangue.

Teste para dados não pareados

Exemplo: Um pesquisador deseja comparar dois meios de cultura diferentes no desenvolvimento de colônias de um certo fungo. Para isso, utilizou 5 placas de Petri com o meio A e 5 com o meio B e colocou o inóculo no centro de cada placa. Após um certo período de tempo, observou as colônias (áreas em cm).

$n_A = 5$

$n_B = 5$

Meio A	22,0	23,2	15,4	22,1	25,0
Meio B	20,5	30,5	26,8	24,8	43,0

Será que as áreas são diferentes?

Hipóteses

$$H_0 : \mu_1 = \mu_2$$

versus

$$\begin{aligned} H_1 &: \mu_1 \neq \mu_2 \quad \text{teste bilateral} \\ H_1 &: \mu_1 > \mu_2 \quad \text{teste unilateral à direita} \\ H_1 &: \mu_1 < \mu_2 \quad \text{teste unilateral à esquerda} \end{aligned}$$

Estatística

- Variâncias iguais ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) desconhecidas:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

com distribuição t de Student com $(n_1 + n_2 - 2)$ graus de liberdade e

$$S_p^2 = \frac{(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2}{n_1 + n_2 - 2}.$$

Estatística

- Variâncias desiguais ($\sigma_1^2 \neq \sigma_2^2$) desconhecidas:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_{X_1}^2}{n_1} + \frac{S_{X_2}^2}{n_2}}},$$

com distribuição t de Student com (ν) graus de liberdade e

$$\nu = \frac{\left(\frac{S_{X_1}^2}{n_1} + \frac{S_{X_2}^2}{n_2}\right)^2}{\frac{\left(\frac{S_{X_1}^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_{X_2}^2}{n_2}\right)^2}{n_2-1}}.$$

Teste de variância de 2 populações

Hipóteses

$H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_a : \sigma_1^2 > \sigma_2^2 \rightarrow$ unilateral

Estatística do teste

					n	Média	Variância
Amostra 1	x_{11}	x_{12}	$\ldots$	x_{1n_1}	n_1	$\bar{x}_1$	s_1^2
Amostra 2	x_{21}	x_{22}	$\ldots$	x_{2n_2}	n_2	$\bar{x}_2$	s_2^2

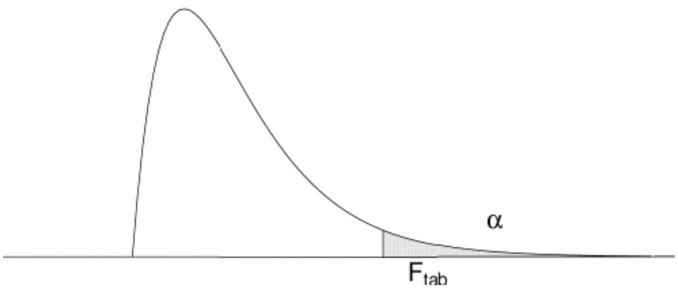
$$F_{calc} = \frac{s_{max}^2}{s_{min}^2}$$ com

$\nu_1 = (n_1 - 1)$g.l. (g.l. do numerador)

$\nu_2 = (n_2 - 1)$g.l. (g.l. do denominador)

Regiões de rejeição

$$H_a : \sigma_1^2 > \sigma_2^2$$



Conclusões

Obs:

Se H_0 é rejeitada $\Rightarrow$ variância heterogênea.

Se H_0 não é rejeitada $\Rightarrow$ variância homogênea.

Teste para dados não pareados

Exemplo: Um pesquisador deseja comparar dois meios de cultura diferentes no desenvolvimento de colônias de um certo fungo. Para isso, utilizou 5 placas de Petri com o meio A e 5 com o meio B e colocou o inóculo no centro de cada placa. Após um certo período de tempo, observou as colônias (áreas em cm).

Meio A	22,0	23,2	15,4	22,1	25,0
Meio B	20,5	30,5	26,8	24,8	43,0

Será que as áreas são diferentes?

homogeneidade de variâncias

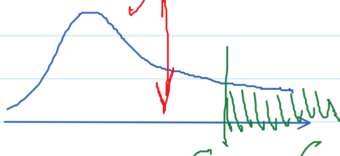
$$\begin{cases} H_0: \sigma_B^2 = \sigma_A^2 \\ H_a: \sigma_B^2 > \sigma_A^2 \end{cases}$$

meio A $\quad n_A = 5 \quad \bar{x}_A = 21,54 \quad S_A^2 = 13,238$

meio A $n_A = 5$ $\bar{X}_A = 21,54$ $S_A = 13,238$

meio B $n_B = 5$ $\bar{X}_B = 29,12$ $S_B^2 = 73,227$

$F_{calc} = \dfrac{73,227}{13,238} = 5,53$ $\nu = 4\,gl$ $\nu = 4\,gl$



Não rejeita Ho e as variâncias são consideradas iguais

$$S_p^2 = \frac{(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2}{n_1 + n_2 - 2}$$

$$S_p^2 = \frac{4 \cdot 13,238 + 4 \cdot 73,227}{8}$$

$$S_p^2 = 43,26$$

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{21,54 - 29,12}{\sqrt{43,26\left(\frac{1}{5} + \frac{1}{5}\right)}} = \boxed{-1,82}$$

$$\begin{cases} H_o: \mu_A = \mu_B \\ H_a: \mu_A \neq \mu_B \end{cases}$$

t de Student com $(n_1 + n_2 - 2)$ → 8 gl
5 5

não se rejeita Ho

Não se rejeita Ho ao nível de 5% de significância e, portanto, há evidências de que não há diferença entre os meios e cultura para este fungo.

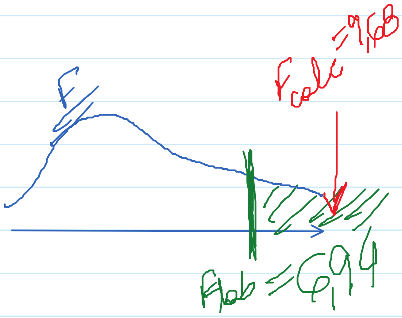
Exercício: As seguintes medidas de *Cytochrome oxidase* foram determinadas em machos de peixes Periplaneta em mm^3 por 10 minutos por miligrama, em um estudo para comparar dois tratamentos, quais sejam: (i) 24 horas após injeção de methoxyclor e (ii) controle, ou seja, sem injeção de methoxyclor.

Tratamento	n	$\bar{x}$	S	S^2
Tratado	5	24,8	0,9	0,81
Controle	3	19,7	2,8	7,84

Verifique se existe efeito significativo da aplicação de methoxyclor quanto às médias de *Cytochrome oxidase*.

Homogeneidade de Variâncias
$$\begin{cases} H_0: \sigma_C^2 = \sigma_T^2 \\ H_a: \sigma_C^2 > \sigma_T^2 \end{cases}$$

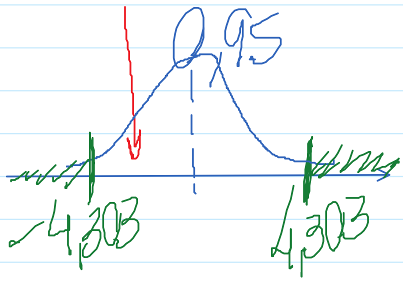
$$F = \frac{7{,}84}{0{,}81} = \boxed{9{,}68} \qquad \nu_1 = 2 \text{ gl} \quad \nu_2 = 4 \text{ gl}$$



Rejeita-Ho e as variâncias são consideradas diferentes

$$\begin{cases} H_0: \mu_C = \mu_T \\ H_a: \mu_C \neq \mu_T \end{cases}$$

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_{X_1}^2}{n_1} + \frac{S_{X_2}^2}{n_2}}} = \frac{19{,}7 - 24{,}8}{\sqrt{\frac{7{,}84}{3} + \frac{0{,}81}{5}}} = \boxed{-3{,}06}$$



na Tabela

t de Student com (ν) graus de liberdade

$$\left(\frac{S_{X_1}^2}{n_1} + \frac{S_{X_2}^2}{n_2}\right)^2 \qquad \left(\frac{7{,}84}{3} + \frac{0{,}81}{5}\right)^2$$

$$\nu = \frac{\left(\frac{S^2_{X_1}}{n_1} + \frac{S^2_{X_2}}{n_2}\right)^2}{\frac{\left(\frac{S^2_{X_1}}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{S^2_{X_2}}{n_2}\right)^2}{n_2 - 1}}.$$

$$\nu = \frac{\left(\frac{7,84}{3} + \frac{0,81}{5}\right)^2}{\frac{\left(\frac{7,84}{3}\right)^2}{2} + \frac{\left(\frac{0,81}{5}\right)^2}{4}} = 2,25 \cong 2gl$$

Não se rejeita Ho ao nível de 5% de significância e, portanto, pode-se dizer que há evidências de que os tratamentos não se diferenciam.

Teste de qui-quadrado para duas ou mais proporções

- Tabelas de contingência:
  - Contagem de sobrevivência de enxertos de ameixeiras ⇒ comparar duas épocas de plantio

Distribuição conjunta das frequências das variáveis época de plantio e sobrevivência de enxertos de ameixeiras

Época	Raízes		Total
	Sobreviventes	Mortas	
Fora da primavera	263	217	**480**
Na primavera	115	365	**480**
Total	378	582	**960**

- Tabelas de contingência:
  - Contagem de plantas segregando para dois caracteres: ciclo e virescência (formação de cloroplastos nas pétalas, originando plantas verdes), numa progênie da espécie X

Ciclo	Virescência		Total
	Normal	Virescente	
Tardio	3470	910	4380
Precoce	1030	290	1320
Total	4500	1200	**5700**

- Tabelas de contingência:

A	B		Total
	B_1	B_2	
A_1	π_{11}	π_{12}	$\pi_{1\cdot}$
A_2	π_{21}	π_{22}	$\pi_{2\cdot}$

A_1	π_{11}	π_{12}	$\pi_{1\cdot}$
A_2	π_{21}	π_{22}	$\pi_{2\cdot}$
Total	$\pi_{\cdot 1}$	$\pi_{\cdot 2}$	$\pi_{\cdot\cdot}$

- Hipótese de homogeneidade das duas distribuições binomiais
  - Contagem de sobrevivência de enxertos de ameixeiras $\Rightarrow$ comparar duas épocas de plantio

Distribuição conjunta das frequências das variáveis época de plantio e sobrevivência de enxertos de ameixeiras

Época	Raízes		Total
	Sobreviventes	Mortas	
Fora da primavera	**263**	217	**480**
Na primavera	**115**	365	**480**
Total	378	582	**960**

$$H_0 : \pi_{1j} = \pi_{2j}$$

$$H_a : \pi_{1j} \neq \pi_{2j}$$

H_0: a proporção de sobreviventes na primavera é igual a proporção de sobreviventes fora dela

n_{ij} linha coluna

- Hipótese de independência entre as variáveis
  - Contagem de plantas segregando para dois caracteres: ciclo e virescência (formação de cloroplastos nas pétalas, originando plantas verdes), numa progênie da espécie X

Ciclo	Virescência		Total
	Normal	Virescente	
Tardio	**3470** n_{11}	910 n_{12}	**4380** $n_{1\cdot}$
Precoce	1030 n_{21}	290 n_{22}	1320 $n_{2\cdot}$
Total	**4500**	1200	**5700**
	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot}$

$$H_0 : \pi_{ij} = \pi_{i\cdot}\pi_{\cdot j}$$

$$H_a : \pi_{ij} \neq \pi_{i\cdot}\pi_{\cdot j}$$

H_0: a proporção de elementos classificados na categoria i da variável A e categoria j da variável B é igual ao produto das marginais dessa categoria

Estatística do teste

$$\chi^2_{cal} = \sum_{i=1}^{s}\sum_{j=1}^{r}\frac{(n_{ij}-e_{ij})^2}{e_{ij}},$$

em que

n_{ij} é a frequência observada de elementos na categoria i da variável A e categoria j da variável B,

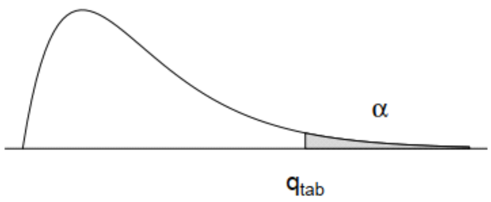
e_{ij} é a frequência esperada de elementos nessa categoria, dada por:

$$e_{ij} = \frac{n_{i.}n_{.j}}{n_{..}},$$

$n_{i.}$, $n_{.j}$ e $n_{..}$ representam as frequências marginais e o total da tabela de contingência a ser analisada.

Distribuição associada

$$\chi^2 \sim \chi^2_{(s-1)\times(r-1)}$$



Rejeita-se H_0 se $\chi^2_{cal} > \chi^2_{tab}$

- Hipótese de homogeneidade das duas distribuições binomiais
  - Contagem de sobrevivência de enxertos de ameixeiras ⇒ comparar duas épocas de plantio

Distribuição conjunta das frequências das variáveis época de plantio e sobrevivência de enxertos de ameixeiras

Época	Raízes		Total
	Sobreviventes	Mortas	
Fora da primavera	**263**	217	**480**
Na primavera	**115**	365	**480**
Total	378	582	**960**

Na primavera	**115**	365	**480**
Total	378	582	**960**

$$H_0 : \pi_{1j} = \pi_{2j}$$

$$H_a : \pi_{1j} \neq \pi_{2j}$$

H_0: a proporção de sobreviventes na primavera é igual a proporção de sobreviventes fora dela

$H_o: \pi_F = \pi_P$

$H_a: \pi_F \neq \pi_P$

total linha

total coluna

$$e_{ij} = \frac{n_{i.} n_{.j}}{n_{..}}$$

esperado

total geral

	Sob	Mortos	Total
Fora P	189	291	480
Prim	189	291	480
total	378	582	960

2 linhas $(r-1) = 1gl$ ×

2 colunas $(n-1) = 1gl$

$\nu = 1gl$

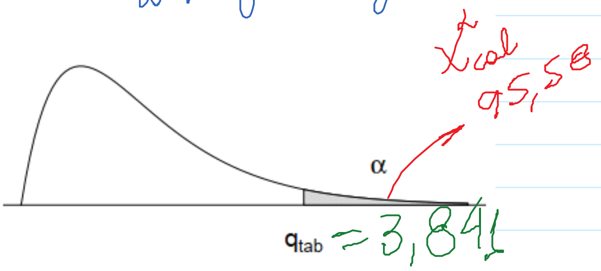
$$\chi^2_{cal} = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \frac{(263-189)^2}{189} + \frac{(217-291)^2}{291} + \frac{(115-189)^2}{189} + \frac{(365-291)^2}{291}$$

$$\chi^2_{calc} = 95,58$$

$$\chi^2 \sim \chi^2_{(s-1)\times(r-1)}$$

$2-1 = 1gl$ $\quad 2-1=1gl$ $\quad = 1gl$

Rejeita-se H_0 se $\chi^2_{cal} > \chi^2_{tab}$



Rejeita-se Ho, e então há evidências de que existe diferença significativa entre as épocas, ou seja, plantar fora da primavera ou na primavera.

- Hipótese de independência entre as variáveis
  - Contagem de plantas segregando para dois caracteres: ciclo e virescência (formação de cloroplastos nas pétalas, originando plantas verdes), numa progênie da espécie X

Ciclo	Virescência		Total
	Normal	Virescente	
Tardio	**3470**	910	**4380**
Precoce	1030	290	1320
Total	**4500**	1200	**5700**

$$\frac{4380 \cdot 4500}{5700} \cong 3458$$

$$H_0 : \pi_{ij} = \pi_{i.}\pi_{.j}$$

$$H_a : \pi_{ij} \neq \pi_{i.}\pi_{.j}$$

H_0: a proporção de elementos classificados na categoria i da variável A e categoria j da variável B é igual ao produto das marginais dessa categoria

$$\begin{cases} H_0: \pi_{ij} = \pi_{i.}\pi_{.j} & \text{(independentes)} \\ H_a: \pi_{ij} \neq \pi_{i.}\pi_{.j} & \text{(não são independentes)} \end{cases}$$

Esperado

	normal	virescente	Total
Tardio	3458	922	4380
Precoce	1042	278	1320
Total	4500	1200	5700

$r = 2$ linhas $(2-1).(2-1) = 1gl$

$s = 2$ colunas

$$\chi^2_{cal} = \frac{(3470-3458)^2}{3458} + \frac{(910-922)^2}{922} + \frac{(1030-1042)^2}{1042} + \frac{(290-278)^2}{278} = 0{,}854$$

$$\chi^2_{tab\,(0,95;\,1gl)} = 3{,}841$$

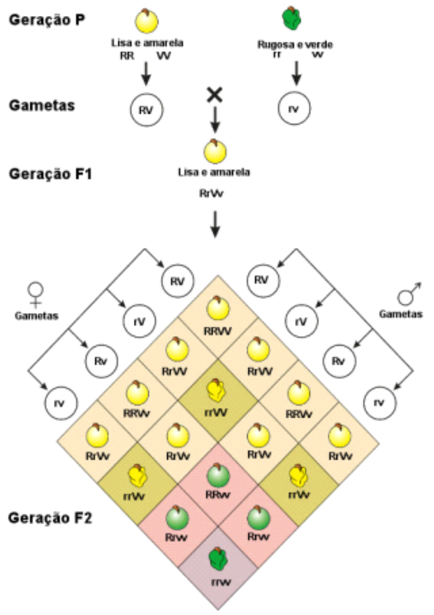
$\chi^2_{cal} < \chi^2_{tab}$ Não rejeita H_0

Não se rejeita Ho ao nível de 5% de significância. Portanto, existem evidências que não existe associação entre virescência e ciclo da planta (são independentes).

Teste de qui-quadrado de aderência

Primeira lei de Mendel ou Lei da Segregação: As características dos indivíduos são determinadas por pares de fatores, os quais se separam na formação dos gametas, indo apenas um fator para cada gameta.
A Primeira Lei de Mendel aplica-se para o estudo de uma única característica.

Segunda Lei de Mendel ou Lei da Segregação Independente dos Genes: Os pares de fatores para duas ou mais características segregam-se de forma independente na formação dos gametas.



- Mendel queria saber como ocorria a transmissão de duas ou mais características em simultâneo.
- Cruzou plantas de sementes amarelas e lisas com plantas de sementes verdes e rugosas.
- Esperava que a geração F_1 teria 100% de sementes amarelas e lisas, pois são características de caráter dominante.
- Fez autofecundação da geração F1, pois imaginava que surgiriam sementes nas proporções 9:3:3:1.

$$\left.\begin{array}{l}\text{ervilhas com sementes amarelas lisas}\\ \text{ervilhas com sementes verdes rugosas}\end{array}\right\} \Rightarrow \text{ervilhas amarelas lisas } (F_1)$$

$$\text{Autofecundação} \Rightarrow F_2 \left\{\begin{array}{l}\text{amarelas lisas } (9/16)\\ \text{verdes lisas } (3/16)\\ \text{amarelas rugosas } (3/16)\\ \text{verdes rugosas } (1/16)\end{array}\right.$$

58 / 9

- Aplicação à teoria Mendeliana

$$\left.\begin{array}{l}\text{ervilhas com sementes amarelas lisas}\\ \text{ervilhas com sementes verdes rugosas}\end{array}\right\} \Rightarrow \text{ervilhas amarelas lisas } (F_1)$$

$$\text{Autofecundação} \Rightarrow F_2 \left\{\begin{array}{l}\text{amarelas lisas } (9/16)\\ \text{verdes lisas } (3/16)\\ \text{amarelas rugosas } (3/16)\\ \text{verdes rugosas } (1/16)\end{array}\right.$$

Frequências observadas das quatro classes fenotópicas geradas por autofecundação de plantas dihíbridas da F_1

Tipos de ervilhas	Frequências observadas	Frequências esperadas sob H_0
Amarelas lisas	315	$312{,}75 = 556 \times (9/16)$
Verdes lisas	108	$104{,}25 = 556 \times (3/16)$
Amarelas rugosas	101	$104{,}25 = 556 \times (3/16)$
Verdes rugosas	32	$34{,}75 = 556 \times (1/16)$
Total	556	556

Avaliar se o padrão de segregação dos caracteres envolvidos segue aquele proposto pela segunda lei de Mendel.

H_0: $\pi_1 = 9/16, \quad \pi_2 = 3/16; \pi_3 = 3/16, \quad \pi_4 = 1/16$

H_a: pelo menos uma das igualdades é falsa

Estatística do teste

$$\chi^2_{cal} = \sum_{i=1}^{m} \frac{(n_i - e_i)^2}{e_i},$$

em que

m é o número de categorias da variável qualitativa n_i é a frequência observada

e_i é a frequência esperada , supondo a hipótese nula verdadeira

Rejeita-se H_0 se $\chi^2_{cal} > \chi^2_{tab}$

$$\begin{cases} H_0: \pi_1 = 9/16,\ \pi_2 = 3/16,\ \pi_3 = 3/16,\ \pi_4 = 1/16 \\ H_a: \text{pelo menos uma das igualdades é falsa} \end{cases}$$

$$\chi^2_{calc} = \frac{(315-312{,}75)^2}{312{,}75} + \frac{(108-104{,}25)^2}{104{,}25} + \frac{(101-104{,}25)^2}{104{,}25} + \frac{(32-34{,}75)^2}{34{,}75}$$

$$\chi^2_{calc} = 0{,}47$$

$\chi^2_{tab} = 7{,}815$
$(3gl;\ 0{,}95)$

Não se rejeita Ho ao nível de 5% de significância e, portanto, existem evidências de que o modelo de Mendel explica os resultados obtidos.