

## Exercícios - Estatística e Delineamento - 2020-21

### 4 Testes $\chi^2$ para tabelas de contingência

- É formulada a hipótese de a cor de ervilheiras ser determinada por um gene e o tipo de superfície ser determinado por outro gene. Admite-se ainda que a cor amarela é dominante da verde e a superfície lisa é dominante da enrugada. Admitindo a segregação independente dos dois genes (isto é, que o alelo herdado do gene da cor é independente do alelo herdado do gene que controla o tipo de superfície), seria de esperar que do cruzamento de ervilheiras heterozigóticas relativamente à cor da semente e tipo de superfície, se observassem as proporções de 9/16 de ervilheiras amarelas lisas, 3/16 amarelas rugosas, 3/16 verdes lisas e 1/16 verdes rugosas.

- Foi realizada uma experiência, cruzando-se ervilheiras heterozigóticas relativamente à cor da semente e tipo de superfície. Os resultados obtidos na descendência são indicados na tabela seguinte.

Cor	Superfície	
	Lisa	Rugosa
Amarelas	556	184
Verdes	193	61

Verifique se os resultados obtidos são compatíveis com as hipóteses genéticas acima formuladas, ao nível de significância  $\alpha = 0.05$ .

- Repita a alínea anterior admitindo que a descendência era 30 vezes maior, mas que as proporções observadas em cada célula da tabela de contingências se mantinham iguais. Comente os seus resultados. Que lição geral se pode extrair destes resultados?
- Cruzaram-se duas linhas puras de cobaias, sendo os progenitores masculinos de pelos curtos e cor amarela e os progenitores femininos de pelos longos e cor branca. A característica pelos curtos é dominante relativamente a pelos longos. Quanto à cor, um genótipo híbrido (isto é, com um alelo de cor amarela e outro de cor branca) terá a cor creme. Admitindo a segregação independente de cada gene, a teoria genética prevê que numa segunda geração ( $F_2$ ) do referido cruzamento, seriam de esperar as seguintes proporções: 6/16 de cobaias de pelo curto e cor creme; 3/16 de pelo curto e cor amarela; 3/16 de pelo curto e cor branca; 2/16 de pelo longo e cor creme; 1/16 de pelo longo e cor amarela; e 1/16 de pelo longo e cor branca.

Uma experiência realizada nestas condições produziu os seguintes resultados:

Pelo	Cor		
	Creme	Amarelo	Branco
Curto	178	93	89
Longo	62	29	31

Verifique se estes resultados são compatíveis com o mecanismo genético referido, ao nível de significância  $\alpha = 0.05$ .

- Considere a experiência descrita no Exercício 3 dos Exercícios Introdutórios, e referente ao enraizamento de estacas semi-lenhosas de oliveiras, submetidas a quatro diferentes tratamentos. Em cada tratamento, foram ensaiadas 60 estacas. Os resultados obtidos foram os seguintes:

Tratamento	Resultado			Total
	Morte	Com calo	Enraizamento	
Sem incisão/sem boro	26	18	16	60
Com incisão/sem boro	32	22	6	60
Sem incisão/com boro	24	24	12	60
Com incisão/com boro	39	19	2	60
Total	121	83	36	240

- (a) Teste se é possível admitir que a distribuição das observações pelas três categorias é igual para os quatro tratamentos.
- (b) Em caso de concluir que diferentes tratamentos estão associados a diferentes distribuições dos resultados, identifique quais as combinações de tratamento/resultado que mais contribuem para essas diferenças. Interprete e comente.
4. Pretende-se estudar a frequência do vírus do enrolamento foliar de tipo 3, um vírus presente na videira em elevadas frequências de ocorrência em todo o mundo. A inexistência deste vírus é condição legal para a selecção de clones. A fim de estudar se a presença do vírus difere consoante as regiões vinícolas, foram aleatoriamente seleccionados 3078 genótipos de várias castas em vinhas antigas em todo o país. Cada genótipo foi classificado, quer de acordo com a região de proveniência, quer pela presença (categoria **positivos**) ou ausência (categoria **negativos**) do vírus. Os resultados obtidos estão na seguinte tabela.

região	negativos	positivos	total
Alentejo e Algarve	124	64	188
Dão	137	52	189
Douro	653	256	909
Lisboa e Vale do Tejo	97	177	274
Vinhos Verdes	471	1047	1518
total	1482	1596	3078

- (a) Diga, justificando, se considera mais adequado analisar os dados utilizando um teste de homogeneidade ou um teste de independência.
- (b) Efectue o teste que escolheu na alínea anterior, explicitando as hipóteses e estatística do teste, a natureza da região crítica e o nível de significância a usar. Sabendo que o valor calculado da estatística é 469.1039 e que se verificam as condições de Cochran, qual é a sua conclusão? Comente.
- (c) Calcule a contribuição da região do Douro para o valor da estatística do teste. Comente.
5. Um estudo analisou a adaptabilidade de pinheiros mansos de seis diversas proveniências (Marrocos, Alcácer do Sal, Itália, Turquia, Viseu e Espanha) nas condições edafo-climáticas de Alcácer do Sal. O estudo consistiu em plantar 150 árvores de cada proveniência e, ao fim de dois anos, contar o número de árvores mortas e sobreviventes, a fim de avaliar se as probabilidades de mortalidade são idênticas para as várias proveniências. Eis os resultados obtidos.

	mortas	sobreviventes
Marrocos	30	120
Alcácer, Portugal	36	114
Itália	23	127
Turquia	67	83
Viseu, Portugal	21	129
Espanha	36	114
Total	213	687

- (a) Descreva formalmente um teste  $\chi^2$  adequado para responder ao objectivo do estudo. Sabendo que o valor da estatística de Pearson calculada foi  $X_{calc}^2 = 51.2762$ , indique e discuta as suas conclusões. Discuta a validade da distribuição assintótica da estatística do teste.
- (b) Identifique a proveniência que mais contribuiu para o valor calculado da estatística do teste, e calcule a respectiva contribuição para o valor de  $X_{calc}^2$  indicado na alínea anterior. Comente.
6. Num estudo sobre a susceptibilidade de variedades de oliveira a fungos causadores de gafa, inocularam-se azeitonas de 4 variedades de oliveira (Galega, Cobrançosa, Negrinha e Picual) com um isolado de *Colletotrichum acutatum* (PT170). Após 24 horas, escolheram-se ao acaso 600 esporos presentes nos frutos de cada variedade, tendo sido registados o número de esporos não germinados, o número de esporos germinados sem formação de apressórios e o número de esporos germinados com formação de apressórios. Os resultados obtidos são indicados na tabela seguinte.

	Não germinados	Germ. sem apressórios	Germ. com apressório	Total
Galega	197	303	100	600
Cobrançosa	267	216	117	600
Negrinha	419	128	53	600
Picual	391	189	20	600
Total	1274	836	290	2400

- (a) Pretende-se saber se cada possível resultado da germinação tem, em todas as variedades, igual probabilidade. Qual o tipo de teste adequado? Justifique, e estime o número esperado de observações em cada entrada da tabela, caso seja verdade a hipótese nula correspondente.
- (b) Descreva o teste e as suas conclusões ( $\alpha = 0.05$ ), sabendo que o valor calculado da estatística do teste é  $X_{calc}^2 = 259.7168$ . Em particular, diga, justificando, se considera válida a distribuição assintótica da estatística do teste.
- (c) A parcela que mais contribui para o valor calculado da estatística de teste corresponde à variedade Galega e resultado “não germinado”. Calcule e comente o valor desta parcela.
7. Num ensaio sobre a variedade de macieira Bravo de Esmolfe em três diferentes terrenos, comparou-se a produtividade (em número de frutos, um ano após a plantação) de macieiras de três diferentes proveniências, a fim de determinar se haveria uma eventual relação entre proveniência e tipos de terreno. As proveniências eram a Beira Litoral, a Beira Interior e Trás-os-Montes. A fim de garantir a comparabilidade de resultados, foram demarcados três diferentes terrenos com igual dimensão e igual número de *árvores* de cada proveniência. Observou-se o número de *frutos* de cada proveniência, num total de 1262 frutos, com os resultados indicados na tabela seguinte:

Proveniência	Terreno 1	Terreno 2	Terreno 3
Beira Litoral	85	137	186
Beira Interior	76	112	214
Trás-os-Montes	67	140	245

- (a) Qual a probabilidade estimada de um fruto escolhido ao acaso ser: (i) duma árvore proveniente de Trás-os-Montes; e (ii) proveniente do Terreno 1?
- (b) Sabendo que o valor calculado da estatística de Pearson adequada ao problema sob estudo é  $X_{calc}^2 = 10.305$ , identifique e concretize os passos do teste que considere adequado para testar se há relação entre proveniência e tipos de terreno. Utilize os níveis de significância:  $\alpha = 0.05$  e  $\alpha = 0.01$ . Comente. O que pode dizer sobre o valor de prova (*p-value*) correspondente?
- (c) Qual a contribuição dos frutos provenientes de Trás-os-Montes, no Terreno 1, para o valor calculado da estatística de teste?