

Exercícios - Estatística e Delineamento - 2015/16

1 Testes χ^2 baseados na estatística de Pearson

1. Galinhas homozigóticas de penas brancas são cruzadas com galos homozigóticos de penas pretas, produzindo uma primeira geração em que a conjugação dos dois alelos (de penas brancas e de penas pretas) produz penas de cor azul. Na segunda geração (F_2), e segundo a teoria genética, seria de esperar que $1/4$ dos pintos tenha penas brancas, $1/4$ tenha penas pretas e $2/4$ tenha penas azuis. Foi realizada uma experiência nos moldes acima indicados, verificando-se na segunda geração 36 pintos de penas brancas, 32 pintos de penas pretas e 73 de penas azuis. Verifique se estes valores observados são compatíveis com a teoria genética (utilize um nível de significância $\alpha = 0.05$).
2. Um investigador afirma que numa determinada cultivar de videira, com um esquema de condução padronizado, o número médio de cachos por pé é igual a 4. É sugerido que se modele o número de cachos por pé através duma distribuição de Poisson com parâmetro $\lambda = 4$. A fim de avaliar essa hipótese, foram amostrados 200 pés e contados os cachos em cada pé, tendo sido obtidos os resultados indicados na tabela. Teste a hipótese indicada com base nestes resultados.

No. cachos	0	1	2	3	4	5	6	7	8	>8
No. de pés	2	20	29	47	54	29	14	4	1	0

3. Uma determinada cultura encontra-se atacada pelo fungo F. Foram seleccionadas aleatoriamente 100 plantas e classificadas quanto à intensidade da infestação. A intensidade da infestação é medida da seguinte forma: a planta é dividida em quatro partes iguais e conta-se o número de partes onde se manifesta a infestação. No quadro figuram o número de plantas classificadas em cada uma das classes de intensidade. Teste se é admissível considerar que uma distribuição Binomial é um modelo aceitável para descrever a intensidade do ataque do referido fungo. Utilize os níveis de significância 0.05 e 0.01, comentando os resultados nos dois casos. Discuta o valor de prova (*p-value*) associado ao valor da estatística. Comente.

Grau de ataque	0	1	2	3	4
No. de plantas	15	20	40	18	7

4. A porosidade de cascas de ovo pode ser medida em termos do número de poros por mm^2 de superfície. Uma experiência aleatória selecciona 320 bocados de casca e procede-se à contagem do número de poros em $1mm^2$ de cada bocado. Os resultados obtidos são dados na tabela de frequências. Será admissível considerar que a variável aleatória que indica o número de poros por milímetro quadrado de casca segue uma distribuição de Poisson? Justifique adequadamente. Calcule o valor de prova associado ao valor calculado da estatística.

No. de poros	0	1	2	3	4	5
Frequência	18	74	139	70	17	2

5. Em estudos relativos à distribuição espacial de insectos é frequente dividir uma determinada zona numa grelha de quadrados de área pré-fixada, procedendo-se à contagem do número de insectos (da espécie sob estudo) por quadrado. Em certos casos, verifica-se que é possível modelar este tipo de experiências considerando que a variável aleatória X , que representa o número de insectos

por quadrado, segue uma distribuição de probabilidades Binomial Negativa. Esta distribuição de probabilidades tem dois parâmetros, m e p , e a função de probabilidades que a caracteriza é:

$$P[X = j] = (1 - p)^j p^m \binom{m + j - 1}{j} \quad \forall j \in \mathbb{N}_0 .$$

Numa experiência deste tipo inspeccionaram-se 100 quadrados, e anotaram-se as contagens de insectos por quadrado, dadas no quadro. Perante estes dados, considera adequada a hipótese de o número de insectos por quadrado ser modelado pela distribuição Binomial Negativa com parâmetros $p = 0.8$ e $m = 5$?

No. de insectos	0	1	2	3	≥ 4
Frequência	35	29	21	8	7

6. A descendência resultante dum cruzamento de ervilheiras heterozigóticas relativamente à cor da semente e tipo de superfície foi a indicada na tabela seguinte.

Côr	Superfície	
	Lisa	Rugosa
Amarelas	556	184
Verdes	193	61

Segundo a teoria genética, 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), desse cruzamento seriam esperadas as proporções de 9/16 de ervilheiras amarelas lisas, 3/16 amarelas rugosas, 3/16 verdes lisas e 1/16 verdes rugosas. Verifique se os resultados obtidos são compatíveis com estas hipóteses, ao nível de significância $\alpha = 0.10$.

7. 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	Côr		
	Creme	Amarelo	Branco
Curto	178	93	89
Longo	62	29	31

Verifique se estes resultados são compatíveis com a teoria genética, ao nível de significância $\alpha = 0.05$.

8. 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.
9. Pretende-se determinar se a distribuição de frutos por várias classes de calibre difere em três regiões nacionais de produção desses frutos: Bombarral, Alentejo e Setúbal. Foram definidas oito classes de calibre, e utilizaram-se no estudo os frutos de cada região que estavam disponíveis: 19 do Bombarral, 155 do Alentejo e 100 de Setúbal. Obtiveram-se as seguintes distribuições:

Região	Calibre								Total
	≤ 40	40 – 45 ⁺	45 – 50 ⁺	50 – 55 ⁺	55 – 60 ⁺	60 – 65 ⁺	65 – 70 ⁺	> 70	
Bombarral	0	0	0	3	4	6	5	1	19
Alentejo	0	7	37	59	36	15	1	0	155
Setúbal	2	40	45	10	3	0	0	0	100
Total	2	47	82	72	43	21	6	1	274

- (a) Teste se é possível considerar que a distribuição de calibres nas três regiões é idêntico.
- (b) Caso opte por admitir diferenças nas distribuições, identifique as combinações de calibres e regiões mais responsáveis por essas diferenças e interprete os seus resultados.
10. Na região escocesa de Caithness, 5387 pessoas foram classificadas de acordo com a sua cor de cabelo (5 classes) e de olhos (4 classes). A tabela de contingências resultante está disponível numa *data frame* de nome `caith`, no pacote MASS do R. Efectue um teste χ^2 para estudar se existe independência entre estes dois factores de classificação. Comente os resultados do teste.
11. Uma direcção regional necessita identificar, na sua região, três tipos de ocupação de solo agrícola: cultura de sequeiro, cultura de regadio e não cultivado. Para tal, envia equipas ao terreno para obter essa informação, o que acarreta custos elevados. Uma pequena empresa propõe fornecer, por um preço muito menor, um mapa de ocupação do solo, obtido por análise de imagens de satélite. Para avaliar se o mapa da empresa é útil, os técnicos da direcção regional escolhem ao acaso 100 parcelas de terreno da região e comparam na tabela abaixo a classificação realizada pelas equipas no terreno (linhas) e a classificação fornecida pelo mapa da empresa (colunas).

	não cultivado	sequeiro	regadio
não cultivado	16	15	4
sequeiro	15	22	3
regadio	0	5	20

Efectue um teste de independência à tabela e comente os seus resultados.