



Probabilidades e Estatística 2004/05

Exercício com um teste de ajustamento para a 31^a aula teórica (dia 18/05/05)

LEIC + LERCI + LEE

A24. O número de golos por jogo em 32 jogos de um Campeonato do Mundo de Futebol teve a seguinte distribuição

Nº Golos	0	1	2	3	4	5
Nº Jogos	1	7	11	5	6	2

Teste ao nível de significância de 1% a hipótese de o número de golos por jogo ter distribuição de Poisson.

Nota: A estimativa de máxima verosimilhança de λ , para a amostra dada, é 2.4375.
(*Exame 25 Jun 2002*)

Resolução: seja X a v.a. que designa o número de golos por jogo. Neste exemplo pretende-se testar

$$H_0 : X \sim Poi(\lambda) \quad \text{contra} \quad H_1 : X \sim \text{outra distribuição.}$$

Note-se que λ não é conhecido, pelo que se terá de recorrer ao método da máxima verosimilhança para o estimar. Como a estimativa é fornecida no enunciado, não é necessário deduzi-la. Vamos então calcular as frequências esperadas (estimadas), usando o princípio de invariância dos estimadores de máxima verosimilhança:

$$\hat{e}_i = n\widehat{P(X = i)}$$

onde i designa o valor da classe. Dado que $n = 32$, vem

Nº Golos	o_i	$\hat{e}_i$
0	1	2.80
1	7	6.82
2	11	8.31
3	5	6.75
4	6	4.11
≥ 5	2	3.21

Note-se que a última classe foi alterada; enquanto que nos dados originais foram observados, no máximo, 5 golos num jogo, como a v.a. de Poisson varia em $\mathbb{N}$, é necessário estender a última classe, de forma que a união das classes dê o espaço de valores da distribuição proposta. Idêntico cuidado é necessário ter na classe inferior (aqui não se aplica, uma vez que a primeira classe é o valor 0, o primeiro valor possível para uma v.a. de Poisson).

Se as frequências esperadas forem pequenas, a estatística de teste (que é assintótica) não reflecte de forma adequada a diferença entre os valores observados e esperados. Não existe unanimidade sobre os valores mínimos adequados ao teste de ajustamento. Assumiremos que é necessário que todas as frequências esperadas sejam maiores ou iguais a 5. Neste caso teremos de agrupar as 2 primeiras classes (somando as frequências observadas entre si, e somando as frequências esperadas entre si) e as 2 últimas classes, pelo que a tabela final é da seguinte forma:

Nº Golos	o_i	$\hat{e}_i$
≤ 1	8	9.61
2	11	8.31
3	5	6.75
≥ 4	8	7.33

Então dado que a estatística de teste é:

$$Q = \sum_{i=1}^4 \frac{(O_i - \hat{E}_i)^2}{\hat{E}_i} \sim_{H_0}^a \chi_{(4-1-1)}^2$$

(note-se que há um parâmetro estimado), o valor observado da estatística de teste é:

$$\frac{(8 - 9.61)^2}{9.61} + \frac{(11 - 8.31)^2}{8.31} + \frac{(5 - 6.75)^2}{6.75} + \frac{(8 - 7.33)^2}{7.33} = 1.658$$

Se fixarmos um nível de significância de 1% (o quantil de probabilidade 0.99 é 9.21), então a região crítica é da forma:

$$\text{rejeitar } H_0 \text{ para } q_{obs} > 9.21$$

pelo que, e porque $1.658 < 9.21$, a decisão é de que não há evidências para rejeitar H_0 para $\alpha \leq 0.05$.

Se não fixarmos nenhum nível de significância, de forma que temos de determinar o valor- p , então vem que

$$1 - p = F_{\chi_{(2)}^2}(1.658) \in [0.5, 0.6]$$

donde $p \in [0.4, 0.5]$ pelo que para os níveis usuais de significância não há evidências para rejeitar H_0 .