Unidade 2 – Testes para uma amostra

Teste Qui-Quadrado de Aderência

É um teste utilizado para comprovar se determinada freqüência observada (f_o) difere significativamente da freqüência esperada (f_e) baseada na hipótese H_0. No geral a frequência esperada é especificada por uma **distribuição de probabilidade.** Para tal, é necessário dividir uma variável em duas ou mais categorias.

Admitindo então que a distribuição da variável em estudo seja descrita por um modelo teórico de probabilidade (Uniforme, Poisson, Binomial, etc...), verifica-se o grau de aderência dos dados amostrais ao modelo.

Hipóteses:

$H_o: fo_i = fe_i$ (a amostra foi extraída de uma população que segue uma determinada distribuição);

$H_1: fo_i \neq fe_i$ (a amostra não foi extraída de uma população que segue uma determinada distribuição).

Se houver então uma concordância entre a frequência observada (fornecida pelos dados do experimento, por exemplo) e a frequência esperada (fornecida pelo modelo probabilístico que queremos testar a aderência), aceitaremos a hipótese nula.

A estatística de teste é:

$$\boxed{\chi^2_{cal} = \sum_{i=1}^{k} \frac{\left(fo_i - fe_i \right)^2}{fe_i}}$$

Em que:

fo_i: freqüência observada na categoria i;

fe_i: freqüência esperada na categoria i, com base na hipótese H_o;

k = número de categorias.

A estatística χ^2_{cal} tem distribuição qui-quadrado com k-1 graus de liberdade.

Procedimento:

a) Estabelecer o nível de significância;
b) Distribuir as frequências observadas nas categorias. O somatório dessas frequências deve ser igual a n;
c) Com base em H_0, determinar as freqüências esperadas (fe_i) para cada uma das k categorias e observar o seguinte:

Observação 1 - Quando k > 2, o teste de Qui-Quadrado não deve ter mais de 20% das frequências esperadas abaixo de 5 e nenhuma freqüência esperada igual a zero; Em casos de fe_i menores que 5, podemos somar categorias adjacentes para eliminar o problema. Porém isso acarretará na perda de graus de liberdade, pois perderemos categorias.

Se temos por exemplo:

Categorias	fo_i	fe_i
0	10	9,5
1	20	20,5
2	8	8,5
3	2	1,5

A categoria 3 tem fe_i = 1,5. Como temos k=4 categorias e 4 fe_i, essa frequência esperada menor que 5 representa 25% (1/4). Não poderíamos aplicar o teste dessa forma. Observe que originalmente temos nesse caso v=4-1 = 3 graus de liberdade.

Para proceder corretamente com o teste nesse caso, poderíamos agrupar as categorias 2 e 3, somando as frequências:

Categorias	fo_i	fe_i
0	10	9,5
1	20	20,5
2 e 3	8 + 2 = 10	8,5 + 1,5 = 10

Dessa forma passamos a ter k= 3 categorias e v=3-1 = 2 graus de liberdade.

Observação 2 - quando k = 2, pode-se utilizar o teste somente se as fe_i são maiores que 5. Nesse caso não podemos agrupar as categorias pois teríamos grau de liberdade igual a 0. Se estas condições não forem satisfeitas, aplica-se o teste Binomial;

d) calcular o valor de χ^2_{cal};

e) Regra de decisão:

Com o auxílio da tabela da Distribuição Qui-Quadrado encontramos o valor de $\chi^2_{\alpha;v}$.

Se $\chi^2_{cal} \geq \chi^2_{\alpha;v}$ rejeitamos a hipótese nula, pois estará na região de rejeição do teste (p-valor $\leq \alpha$).

Exemplo: Verificar se podemos afirmar que os dados abaixo se ajustam a uma distribuição Poisson com $\lambda = 1,38$. Utilize α = 0.05.

Número de acidentes (categorias)	0	1	2	3	4	5
Número de dias (frequências observadas)	25	19	10	9	4	3

Resolução:

$$\begin{cases} H_0: \text{se ajustam a uma Poisson} \\ H_1: \text{não se ajustam a uma Poisson} \end{cases}$$

Nesse caso, queremos testar a aderência a uma distribuição Poisson com $\lambda = 1,38$.

Para tal, devemos então encontrar as frequências esperadas correspondentes a cada categoria tomando por base o cálculo de probabilidades envolvendo o modelo Poisson:

$$P(\mathrm{X} = \mathrm{k}) = \frac{e^{-\lambda}\lambda^k}{k!}$$

$$P(\mathrm{X} = 0) = \frac{e^{-1,38}1,38^0}{0!} = 0,25 \qquad P(\mathrm{X} = 1) = \frac{e^{-1,38}1,38^1}{1!} = 0,35$$

$$P(\mathrm{X} = 2) = 0,24 \qquad P(\mathrm{X} = 3) = 0,11$$

$$P(\mathrm{X} = 4) = 0,04 \qquad P(\mathrm{X} = 5) = 0,01$$

Note que precisamos das fe_i e por enquanto o que temos são as probabilidades. Para encontrar as frequências basta multiplicarmos as probabilidades encontradas por n, que nesse caso é igual a 70 (soma das frequências observadas)

$$f_{e0} = 0,25 \times 70 = 17,5 \qquad f_{e1} = 0,35 \times 70 = 24,5$$
$$f_{e2} = 0,24 \times 70 = 16,8 \qquad f_{e3} = 0,11 \times 70 = 7,7$$
$$f_{e4} = 0,04 \times 70 = 2,8 \qquad f_{e5} = 0,01 \times 70 = 0.70$$

Teremos então:

Número de acidentes (categorias)	0	1	2	3	4	5
fo_i	25	19	10	9	4	3
fe_i	17,5	24,5	16,8	7,7	2,8	0,70

Temos duas categorias com fe_i menores que 5, que equivalem a mais de 20% das k=6 categorias. Dessa forma devemos agrupar as adjacentes para eliminarmos esse problema:

Número de acidentes (categorias)	0	1	2	3, 4 e 5
fo_i	25	19	10	16
fe_i	17,5	24,5	16,8	11.2

Agora, aplicando a estatística do teste:

$$\chi^2_{cal} = \sum_{i=1}^{k} \frac{(fo_i - fe_i)^2}{fe_i} = \frac{(25-17,5)^2}{17,5} + \frac{(19-24,5)^2}{24,5} + \frac{(10-16,8)^2}{16,8} + \frac{(16-11,2)^2}{11,2} = 9,26$$

Como ficamos com 4 categorias, temos v=3 graus de liberdade e α = 0.05. Pela tabela da distribuição qui-quadrado temos:

$\chi^2_{0.05;3} = 7,81$ e então $\chi^2_{cal} \geq \chi^2_{0.05;3}$, logo rejeitamos H_0 e não existem evidencias para que se afirme que os dados sigam uma distribuição Poisson com $\lambda = 1,38$.