Unidade 4 – Testes para duas amostras independentes

Para o caso de populações independentes, emprega-se um teste estatístico não-paramétrico devidamente adequado aos dados a serem analisados, não sendo necessário que as amostras tenham o mesmo tamanho.

Teste Qui-quadrado de independência

Uma outra aplicação do teste qui-quadrado visa estudar relações entre duas ou mais variáveis. Nesse caso, utilizamos tabelas de contingência para a execução do teste. Trabalha-se com frequências observadas distribuídas em classificações, dispondo em uma tabela com h linhas e k colunas.

Hipóteses:

H_o: as variáveis são independentes;

H_1: as variáveis não são independentes, ou seja, elas apresentam algum grau de associação entre si.

A Estatística do teste é a mesma do teste qui-quadrado já apresentado na disciplina

$$\chi_c^2 = \sum_{i=1}^{h} \sum_{j=1}^{K} \frac{\left(fo_{ij} - fe_{ij} \right)^2}{fe_{ij}}$$

fo_{ij} é a frequência observada na linha i e coluna j;

$fe_{ij} =$ frequência esperada $= \dfrac{(\text{Total linha i}) \times (\text{Total coluna j})}{N}$

A restrição $fe_{ij} \geq 5$ também deve ser atendida. Caso não seja, utilizar o teste exato de Fisher.

Regra de decisão

Se $\chi^2_c \geq \chi^2_{\alpha,\nu}$, rejeita-se H_o, $p \leq \alpha$.

Em que :

ν = (h-1)(k-1) graus de liberdade;

h = número de linhas;

k = número de colunas.

Com essas informações, basta buscarmos o valor crítico na tabela da distribuição qui-quadrado.

Exemplo:

Um determinado estudo com 100 fumantes (expostos) e 92 não-fumantes (não expostos/controles) em relação a ocorrência de infarto do miocárdio (IM) apresentou os resultados da Tabela abaixo. Verificar se existe associação entre a exposição (fumo) e o infarto do miocárdio (desfecho) para um nível de significância de α = 5%?

Tabela - Ocorrência de infarto relacionado ao uso de tabaco

	Infarto do miocárdio		
	Presença	*Ausência*	*Total*
Fumantes	*17*	*83*	*100*
Não-fumantes	*6*	*86*	*92*
Total	*23*	*169*	*192*

Resolução

H_o*: Não existe relação entre fumo e infarto;*

H_1*: Existe relação entre fumo e infarto.*

Já temos as 4 frequencias observadas:

$fo_{11} = 17,\ fo_{12} = 83,\ fo_{21} = 6,\ fo_{22} = 86$

Para cada uma delas, devemos encontrar a frequência esperada correspondente, através da fórmula:

$$fe_{ij} = \frac{(\text{Total linha i}) \times (\text{Total coluna j})}{N}$$

Assim,

$$fe_{11} = \frac{(\text{Total linha 1}) \times (\text{Total coluna 1})}{192} = \frac{100 \times 23}{192} = 11.98$$
$$fe_{12} = \frac{(\text{Total linha 1}) \times (\text{Total coluna 2})}{192} = \frac{100 \times 169}{192} = 88.02$$
$$fe_{21} = \frac{(\text{Total linha 2}) \times (\text{Total coluna 1})}{192} = \frac{92 \times 23}{192} = 11.02$$
$$fe_{22} = \frac{(\text{Total linha 2}) \times (\text{Total coluna 2})}{192} = \frac{92 \times 169}{192} = 80.98$$

$$\chi_c^2 = \frac{(17 - 11.98)^2}{11.98} + \frac{(83 - 88.02)^2}{88.02} + \frac{(16 - 11.02)^2}{11.02} + \frac{(86 - 80.98)^2}{80.98} = 4.98$$

Agora vamos a tabela da distribuição qui-quadrado localizar o valor critico. Como temos uma tabela 2x2, temos 1 grau de liberdade. A 5%, o valor crítico é 3,84.

Como 4.98 > 3.94, rejeita-se H0, ou seja, existem evidências para afirmarmos que existe relação entre o fumo e o infarto.