Teste Qui-Quadrado

A análise de aderência testa a distribuição dos dados:

$$\begin{cases} H_0 : P = P_0 \\ H_1 : P \neq P_0 \end{cases}$$

Em que P_0 é a medida de probabilidade especificada que governaria (sob H_0) os eventos observados.

Neste teste co,paramos a frequência observada com a frequência esperada em k eventos disjuntos e distintos observáveis.

Eventos	1	2	...	k
P_0	P_{01}	P_{02}	...	P_{0k}
E_i	E_1	E_2	...	E_k
O_i	O_1	O_2	...	O_k

Em que observou-se uma amostra de tamanho n. Temos também que E_i é o valor esperado do número de eventos i sob H_0

$$\text{Freq. Esperada} = E_i = P_{0i} \cdot n$$

e $\text{Freq. Observada} = O_i$ é o numero real de eventos i observados na amostra.
A estatística para testar H_0 é

$$\chi^2 = \sum_{i=1}^{k} \frac{(E_i - O_i)^2}{E_i}$$

que, sob H_0 - ou seja, sob a hipótese de que P_0 é de fato a medida de probabilidade que governa o comportamento probabilístico do evento - é aproximadamente

$$\underbrace{\chi^2 \sim \chi^2_{(k-1)}}_{\text{Sob } H_0}$$

*Esse procedimento é confiável sempre que $E_i > 5 \forall i \in \{1, \ldots, k\}$

Exemplo

Considere que queremos verificar se os números sorteados nos concursos da Mega Sena são de fato uniformemente distribuídos.

Nesse caso, analisaremos 60 eventos, cuja probabilidade de cada um seria, caso uniformemente distribuídos, $\frac{1}{60}$.

$$\begin{cases} H_0 : P = P_0 \\ H_1 : P \neq P_0 \end{cases}$$

Em que $P_0(\{i\}) = \frac{1}{60} \forall i \in \{1, 2, \ldots 60\}$
Vamos criar a tabela para as frequências. Consideraremos *a primeira bola* de todos os 2800 sorteios da Mega.

Eventos	1	2	...	60
P_0	$\frac{1}{60}$	$\frac{1}{60}$	...	$\frac{1}{60}$
E_i	$\frac{2800}{60}$	$\frac{2800}{60}$	...	$\frac{2800}{60}$
O_i	42	48	...	55

Portanto,

$$\chi^2 = \sum_{i}^{60} \frac{(46.7 - O_i)^2}{46.7} \overset{a}{\sim} \chi^2_{59}$$

Considerando um nível de significância de $\alpha = 5\%$, calculamos o ponto crítico c tal que

$$P(\chi^2_{59} > c) = 0.05$$

Pelo computador, encontramos $c = 77.93$
Logo, como $\chi^2 = 56.68 < 77.93$, concluímos que, sob H_0, não há

evidências de que o modelo não seja equiprovável a 5% de significância de estatística.

K-Grupos

(Morettin, Pag.404 E.7)
Considere os $n = 30$ dados abaixo que supostamente seguem uma distribuição normal $N(10, 25)$.
(usando os dados do livro já em ordem)

$$\begin{matrix} 1.01 & 1.73 & 3.93 & 4.44 & 6.37 & 6.51 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 14.11 & 14.6 & 14.64 & 14.75 & 16.68 & 22.14 \end{matrix}$$

Queremos testar se os dados de fato se distribuem de acordo com $N(10, 25)$.

$$\begin{cases} H_0 : P = N(10, 25) \\ H_1 : P \neq N(10, 25) \end{cases}$$

Sob H_0, podemos dividir a distribuição normal em k blocos. Escolheremos $k = 4$ delimitado pelos *quartis* teóricos dessa distribuição normal. (Primeiro padronizamos, encontramos os valores pela tabela, então voltamos para nossa normal)

$$\begin{cases} q_1 = 6.63 \\ q_2 = 10 \\ q_3 = 13.3 \end{cases} \overset{\text{Intervalos}}{\Rightarrow} \begin{cases} 1.(-\infty, q_1) \\ 2.[q_1, q_2] \\ 3.(q_2, q_3] \\ 4.(q_3, \infty) \end{cases}$$

Podemos produzir uma tabela com as frequências por intervalo

Eventos	1.	2.	3.	4.
E_i	$0.25 \cdot 30 = 7.5$	7.5	7.5	7.5
O_i	6	9	9	6

$$\chi^2 = \sum_{i=1}^{4} \frac{(7.5 - O_i)^2}{7.5} = 1.2$$

Na χ^2_3 (número de nichos), com nível de significância $\alpha = 0.10$, $c = 6.25$.
Como $\chi^2 = 1.2 < 6.25$, concluímos que não há evidências de que a distribuição dos dados difere de uma $N(10, 25)$ a $\alpha = 10\%$ de significância estatística