

# Teste $\chi^2$ de aderência

Chuck Norris\*      Arnold Schwarzenegger†

16 de julho de 2013

## Resumo

O teste  $\chi^2$  de aderência é considerado para testar a hipótese de que uma distribuição de probabilidades as frequências de ocorrência de uma variável aleatória. O procedimento de aplicação do teste será descrito e uma aplicação será apresentada.

## 1 Motivação do teste

Um tipo de problema frequentemente encontrado é o de não se conhecer a distribuição de probabilidades de uma variável aleatória  $X$ . No entanto, uma vez observada essa variável aleatória deseja-se testar a hipótese de que uma particular distribuição de probabilidades explica satisfatoriamente a sua ocorrência. Em outras palavras, quer-se testar a adesão de uma distribuição de probabilidades aos valores observados de uma variável aleatória.

## 2 Procedimento de teste

O procedimento para o teste requer que seja observada uma amostra aleatória de tamanho  $n$  da variável aleatória  $X$ . Essas observações são agrupadas em classes. Caso  $X$  seja uma v.a. qualitativa, os níveis observados

---

\* Acadêmico do Curso de Estatística, grr: 12345678.

† Acadêmico do Curso de Estatística, grr: 87654321.

são as classes. No caso de  $X$  ser quantitativa pode-se agrupar os dados em classe tal como se faz para construir um histograma.

Seja  $O_i$  a frequência absoluta observada em cada uma das classes,  $i = 1, \dots, k$ , em que  $k$  é o número total de classes. A partir da distribuição de probabilidade considerada no teste, ou seja, aquela definida na hipótese nula ( $H_0$ :  $X$  tem distribuição tal) aquela para qual vamos aplicar o teste, calcula-se as frequências esperadas da v.a.  $X$ ,  $E_i$ .

A estatística do teste é

$$X_0^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}. \quad (1)$$

No caso de  $H_0$  ser verdadeira,  $X_0^2$  tem distribuição  $\chi^2$  com  $k - 1 - p$  graus de liberdade. O número de parâmetros estimados sob  $H_0$  é representado por  $p$ . A distribuição da estatística sob  $H_0$  é cada mais parecida com a de referência a medida que o tamanho da amostra aumenta. Rejeita-se  $H_0$  quando o valor calculado da estatística for superior ao valor crítico  $\chi_\alpha^2$  para um nível de significância nominal pré-estabelecido  $\alpha$ .

Sugere-se que classes com frequência esperada,  $E_i$ , menor ou igual a 5 sejam combinadas com classes adjacentes de forma que, após serem combinadas, tenham  $E_i > 5$ .

### 3 Aplicação do teste

Para se demonstrar a aplicação do teste serão considerados dados sobre o número de defeitos em placas de circuito impresso. Existe uma forte sustentação teórica relacionada ao processo gerador dos dados que indica que a distribuição do número de defeitos seja Poisson sob certas circunstâncias. Deseja-se testar a aderência da distribuição Poisson. A frequência absoluta observada de defeitos está na tabela abaixo onde o número de classes é  $k = 4$ .

Número de defeitos ( $x_i$ )	Frequência observada ( $O_i$ )
0	32
1	15
2	9
3	4

Sob a hipótese  $H_0$  dos dados terem distribuição Poisson tem-se que calcular as frequências esperadas,  $E_i$ . No entanto, precisa-se estimar o parâmetro  $\lambda$ , com isso  $p = 1$ . A estimativa de  $\lambda$  é obtida por

$$\hat{\lambda} = \frac{\sum_{i=1}^k x_i \cdot O_i}{\sum_{i=1}^k O_i} = 0.75. \quad (2)$$

Assim, os valores esperados são calculados por meio da função de probabilidades da distribuição Poisson. Assim

$$\begin{aligned} E_1 &= n \Pr(X = 0) = 60 \cdot \frac{\exp\{-0.75\} 0.75^0}{0!} = 28.32 \\ E_2 &= n \Pr(X = 1) = 60 \cdot \frac{\exp\{-0.75\} 0.75^1}{1!} = 21.24 \\ E_3 &= n \Pr(X = 2) = 60 \cdot \frac{\exp\{-0.75\} 0.75^2}{2!} = 7.98 \\ E_4 &= n \Pr(X \geq 3) = 60(1 - \Pr(0) - \Pr(1) - \Pr(2)) = 2.46 \end{aligned}$$

A figura ?? representa as frequências observadas e esperadas para o número de defeitos em placas de circuito impresso.

Como o número esperado da quarta classe foi menor que 5, está será combinada com a terceira classe. Então temos a seguinte tabela com frequências observadas e esperadas.

$x_i$	$O_i$	$E_i$
0	32	28.32
1	15	21.24
2+	9	10.44

A estatística do teste é calculada por

$$X_0^2 = \frac{(32 - 28.32)^2}{28.32} + \dots + \frac{(13 - 10.44)^2}{10.44} = 2.94. \quad (3)$$

O grau de liberdade é  $k - 1 - p = 1$  e o valor correspondente para  $\alpha = 0.05$  é  $\chi_1^2 = 3.84$ . Uma vez que  $X_0^2 \leq \chi_1^2$  não rejeitamos  $H_0$  de que a distribuição do número de defeitos em placas de circuito impresso seja Poisson. A figura ?? ilustra a distribuição de referência com a posição do valor crítico e da estatística calculada do teste.

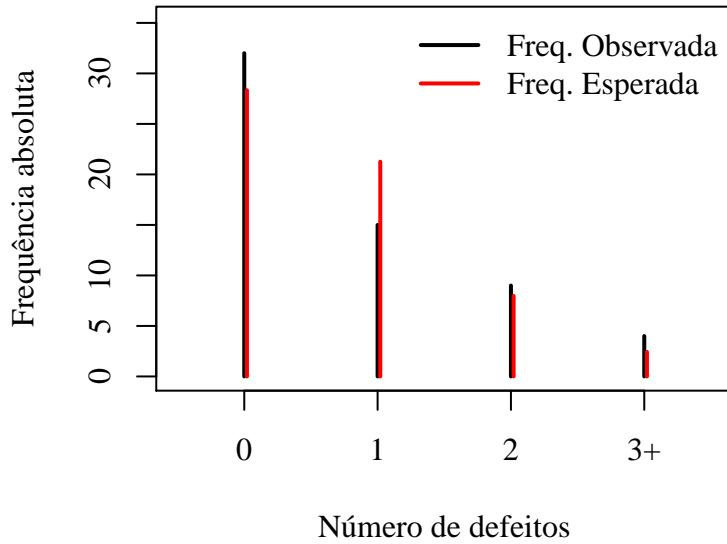


Figura 1: Frequências observadas e esperadas para o número de defeitos em placas de circuito impresso.

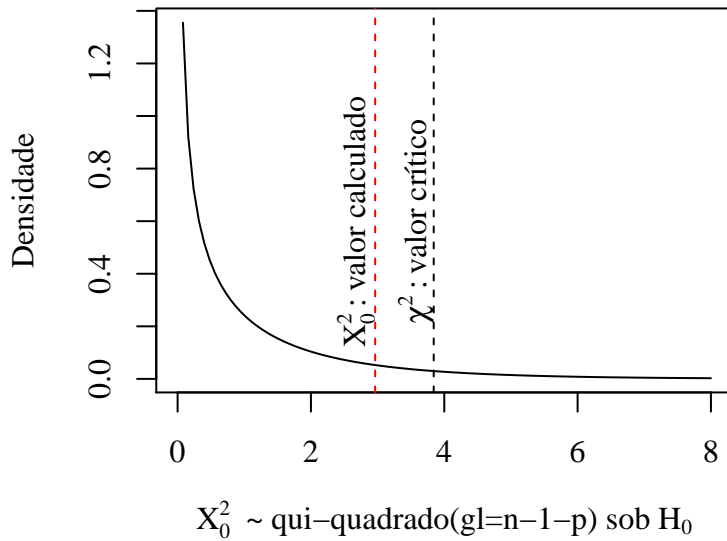


Figura 2: Função densidade da distribuição qui-quadrado com 1 grau de liberdade. Linhas verticais representam o valor crítico e da estatística calculada do teste.

## 4 Sistematização do procedimento

A seguir a sequência de etapas de descreve de forma objetiva (ou procedural) a aplicação do teste  $\chi^2$  de aderência.

1. Defina um nível nominal de significância  $0 < \alpha < 1$ ;
2. Defina uma distribuição de probabilidade em  $H_0$  para  $X$ ;
3. Observe uma amostra aleatória de tamanho  $n$  de  $X$ ;
4. Classifique os valores observados em  $k$  classes e obtenha as frequências observadas  $O_i$ ;
5. Sob  $H_0$  obtenha as frequências esperadas  $E_i$ . Combine classes adjacentes de forma que  $E_i > 5 \forall i$ . Denote por  $p$  o número de parâmetros estimados para obter  $E_i$ .
6. Encontre o valor crítico,  $\chi_{\alpha}^2$ , correspondente à  $\alpha$  na distribuição qui-quadrado com  $k - 1 - p$  graus de liberdade;
7. Calcule a estatística do teste

$$X_0^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i};$$

8. Rejeite  $H_0$  se  $X_0^2 > \chi_{\alpha}^2$ , caso contrário aceite.

## 5 Considerações finais

Para um bom desempenho do teste, ou seja, operar com o nível nominal de significância estabelecido, supoe-se que a amostra seja grande e que os valores esperados sejam maiores que 5. No R pode-se usar a função `chisq.test()` para aplicar esse teste.

## Referências

- [1] Montgomey, D. C. **Estatística aplicada e probabilidade para engenheiros**. 4 ed., LTC, 2007, 493 p.