



Estatística I

Prof. Fernando de Souza Bastos

fernando.bastos@ufv.br

Departamento de Estatística

Universidade Federal de Viçosa

Campus UFV - Viçosa

Testes χ^2 de Pearson

Teste de Aderência

Teste de Independência

Testes χ^2 de Pearson

Um teste qui-quadrado, também escrito como teste χ^2 , é qualquer teste de hipótese estatística em que a distribuição amostral da estatística de teste é uma distribuição qui-quadrada quando a hipótese nula é verdadeira. O teste qui-quadrado é utilizado para determinar se existe uma diferença significativa entre as frequências esperadas e as frequências observadas em uma ou mais categorias.

Existem três tipos:

- **Teste de aderência:** testa a hipótese da amostra ser proveniente de uma distribuição de probabilidade definida em H_0 . Com essa distribuição definida em H_0 são obtidos as frequências esperadas (E);

Existem três tipos:

- **Teste de aderência:** testa a hipótese da amostra ser proveniente de uma distribuição de probabilidade definida em H_0 . Com essa distribuição definida em H_0 são obtidos as frequências esperadas (E);
- **Teste de homogeneidade:** testa a hipótese H_0 de duas ou mais amostras serem provenientes de uma mesma distribuição de probabilidades. Os valores esperados são obtidos pelo produto da linha marginal e tamanho das amostras;

Existem três tipos:

- **Teste de aderência:** testa a hipótese da amostra ser proveniente de uma distribuição de probabilidade definida em H_0 . Com essa distribuição definida em H_0 são obtidos as frequências esperadas (E);
- **Teste de homogeneidade:** testa a hipótese H_0 de duas ou mais amostras serem provenientes de uma mesma distribuição de probabilidades. Os valores esperados são obtidos pelo produto da linha marginal e tamanho das amostras;
- **Teste de independência:** testa a hipótese H_0 de que a distribuição conjunta é o produto das distribuições marginais, o que só ocorre quando existe independência entre as variáveis aleatórias. No caso de duas variáveis aleatórias organizadas numa tabela de dupla entrada, os valores esperados são obtidos como produto dos valores marginais.

Nos testes chi-quadrado o que muda é só a hipótese envolvida no cálculo dos valores esperados. Para os três tipos de hipótese, a estatística do teste é:

$$\chi_{cal}^2 = \sum_{i=1}^k \frac{(f_{oi} - f_{ei})^2}{f_{ei}}$$

em que f_{oi} e f_{ei} são, respectivamente, as frequências observadas e esperadas. Sendo que sob H_0 a variável aleatória $\chi_{cal}^2 \sim \chi_{\nu}^2$ em que ν são os graus de liberdade.

Teste de Aderência

Temos uma população P e queremos verificar se ela segue uma distribuição especificada P_0 , isto é, queremos testar a hipótese $H_0 : P = P_0$. O procedimento consiste em considerar classes, segundo as quais a variável X , característica da população, pode ser classificada. A variável X pode ser qualitativa ou quantitativa.

Supondo H_0 verdadeira,

$$\chi_{cal}^2 = \sum_{i=1}^k \frac{(f_{oi} - f_{ei})^2}{f_{ei}} \sim \chi_q^2,$$

em que $q = k - 1$ representa o número de graus de liberdade.

Observação: Este resultado é válido para n grande e para $f_{ei} \geq 5, i = 1, 2, \dots, k$.

Regra de decisão: Pode ser baseada no nível descritivo ou valor-P, neste caso

$$valor - p = P(\chi_q^2 \geq \chi_{cal}^2),$$

em que χ_{cal}^2 é o valor calculado, a partir dos dados, usando a expressão apresentada para χ_{cal}^2 . Se para α fixado, $p - valor < \alpha$ rejeitamos H_0 .

Exemplo 14.1 - Morettin e Bussab, 2023

Um dado é lançado 300 vezes, com os resultados dados na próxima Tabela. Por enquanto, considere somente a linha correspondente às frequências observadas. Com os resultados observados, queremos saber se o dado é “honesto”, isto é, se a probabilidade de ocorrência de qualquer face é $1/6$. Ou seja, queremos testar a hipótese

$$H_0 : p_1 = p_2 = \cdots = p_6 = \frac{1}{6},$$

em que $p_i = P(\text{face } i)$, $i = 1, 2, \dots, 6$. Isso equivale a dizer que P_0 segue uma distribuição uniforme discreta.

Tabela 1: Resultados do lançamento de um dado 300 vezes. (Morettin e Bussab, 2023)

Ocorrência (i)	1	2	3	4	5	6	Total
Freq. Observada	43	49	56	45	66	41	300
Freq. Esperada	50	50	50	50	50	50	300

H_0 : O dado é honesto

H_1 : Não H_0

$$\chi_{cal}^2 = \frac{(43 - 50)^2}{50} + \dots + \frac{(41 - 50)^2}{50} = 8.96$$

```
(x2_t <- qchisq(0.05, df, lower.tail = FALSE)) #No R
```

$$\chi_{tab}^2 = 11,0705$$

```
(pvalor <- pchisq(x2_c,df,lower.tail = FALSE))
```

$$\chi_{tab}^2 = 0,1106703$$

Conclusão: Não rejeitamos H_0 ao nível de 5% de significância.

Exercício

Dia	Seg	Ter	Qua	Qui	Sex
Acidentes	32	40	20	25	33

Deseja-se testar, ao nível de 5%, se os acidentes estão uniformemente distribuídos durante a semana.

Hipóteses:

- H_0 : Frequências uniformes ($p_i = 1/5$).
- H_1 : Pelo menos um dia difere.

Pergunta: Com base no teste qui-quadrado, há evidências para rejeitar H_0 ?

Resolução

Cálculos

- Total de acidentes: 150
- Esperado por dia: $E_i = 150/5 = 30$
- Estatística do teste:

$$\chi_{cal}^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 7.933$$

- Graus de liberdade: $k - 1 = 4$
- Valor-p = $P(\chi_4^2 > 7.933) \approx 0.094$

Conclusão: Como $p = 0,094 > 0,05$, não rejeitamos H_0 . Não há evidências suficientes para afirmar que as frequências diferem da distribuição

Teste de Independência

O teste de independência supõe a existência de duas v.a.'s X e Y , e os valores de amostras delas são classificados segundo categorias, obtendo-se uma tabela de dupla entrada. Quere-mos testar a hipótese que X e Y são independentes.

Exemplo 14.3 - Morettin e Bussab, 2023

Uma companhia de seguros analisou a frequência com que 2.000 segurados (1.000 homens e 1.000 mulheres) usaram hospitais. Os resultados estão na Tabela abaixo. A hipótese a testar é que o uso de hospital independe do sexo do segurado.

```
> M<-data.frame(uso_hospital=c("usaram_hospital",  
+                             "nao_usaram_hospital")  
+               ,homens=c(100,900),mulheres=c(150,850),  
+               row.names = 1)  
> M
```

	homens	mulheres
usaram_hospital	100	150
nao_usaram_hospital	900	850

Exemplo

Uma companhia de seguros analisou a frequência com que 2.000 segurados (1.000 homens e 1.000 mulheres) usaram hospitais. Os resultados estão na Tabela abaixo. A hipótese a testar é que o uso de hospital independe do sexo do segurado.

```
> Xsq <- chisq.test(M, correct = FALSE)
> Xsq
```

Pearson's Chi-squared test

data: M

X-squared = 11.429, df = 1, p-value = 0.0007232

```
> Xsq$observed    # observed counts (same as M)
```

	homens	mulheres
usaram_hospital	100	150
nao_usaram_hospital	900	850

```
> Xsq$expected    # expected counts under the null
```

	homens	mulheres
usaram_hospital	125	125
nao_usaram_hospital	875	875

```
> Xquad <- (((100-125)^2)/125)+(((150-125)^2)/125)+  
            (((900-875)^2)/875)+(((850-875)^2)/875)
```




```
> Xsq <- chisq.test(M, correct = TRUE)
> Xsq
```

Pearson's Chi-squared test with Yates' continuity correction





data: M

X-squared = 10.976, df = 1, p-value = 0.000923

Referências

-  Bastos, Fernando de Souza (2025). ***Apostila Interativa***. Disponível online: <https://ufvest.shinyapps.io/ApostilaInterativa/>.
-  Ferreira, Eric Batista e Marcelo Silva de Oliveira (2020). ***Introdução à Estatística com R***. Editora Universidade Federal de Alfenas. URL: https://www.unifal-mg.edu.br/bibliotecas/wp-content/uploads/sites/125/2021/12/32-EBR_Unifal.pdf.
-  Meyer, Paul L (1982). ***Probabilidade: aplicações à estatística***. Livros Técnicos e Científicos.

Referências ii

-  Montgomery, D. C. e G. C Runger (2016). ***Estatística Aplicada E Probabilidade Para Engenheiros***. 6ª ed. São Paulo: Grupo Gen-LTC.
-  Morettin, P.A. e W.O Bussab (2023). ***Estatística básica***. 10ª ed. São Paulo: Editora Saraiva.
-  Peternelli, Luiz Alexandre (s.d.). ***Apostila (EST 106)***. Formato slide Disponível no PVANet - Moodle.
-  Zeviane, Walmes (jul. de 2011). ***Rídiculas - Dicas Curtas sobre o R***. URL: <https://ridiculas.wordpress.com/2011/07/04/os-testes-chi-quadrado/>.