Noções de Estatística Básica

Testes Qui-Quadrado

- Testes de Aderência
- Testes de Independência

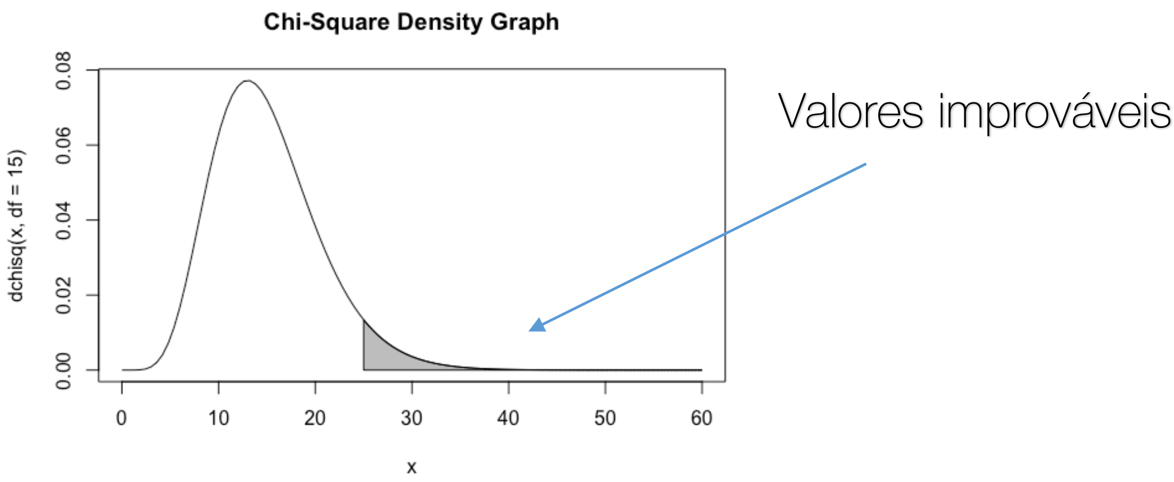
1. Introdução

Como vimos na aula de testes de hipótese, podemos criar uma regra de decisão baseada em uma característica da amostra (estatística). Outra possibilidade é olharmos para o **valor-p** do teste.

O valor-p de um teste pode ser interpretado da seguinte maneira:

Assuma que a hipótese nula é verdadeira.

Sob a hipótese nula, o valor-p indica a probabilidade de se obter um valor igual ou maior que o observado na amostra. Exemplo:



Se o valor-p for uma probabilidade muito pequena (área cinza), podemos verificar a nossa regra de decisão e rejeitar a hipótese nula.

A seguir estudaremos dois tipos diferentes de testes de hipótese que contarão com o auxílio da estatística e também da Distribuição χ^2.

> Você verá que, embora os dois testes façam uso do cálculo dessa estatística, são construídos de forma diferente e também com objetivos diferentes.

2. Testes de aderência

Objetivo: Testar a adequabilidade de um modelo probabilístico a um conjunto de dados observados.

Metodologia:

1. Há um modelo probabilístico teórico (provavelmente uma tabela de probabilidades).

2. Há um conjunto de dados com uma respectiva tabela de distribuição de frequência. (Nosso objetivo é comparar a tabela de distribuição de frequência com o modelo proposto em 1.)

3. Formula-se a hipótese nula: O modelo proposto é adequado para este conjunto de dados. Ou seja,

   H0: p1 = f1 e p2 = f2 e p3=f3 .... pn = fn

4. Hipótese alternativa: O modelo proposto não é adequado para este conjunto de dados.

   H1: existe pelo menos alguma diferença

5. Calcula-se a estatística do teste, nesse caso a χ^2, e o seu respectivo valor-p.

6. Tomamos uma decisão. Rejeitamos ou não a hipótese nula.

Exemplo:

Lançamos um dado 60 vezes.

Vamos testar a hipótese de que o dado seja honesto.

Segundo essa hipótese, esperamos que:

	1	2	3	4	5	6
Freq. Esperada	10	10	10	10	10	10

Mas ao lançarmos o dado, observamos:

	1	2	3	4	5	6
Freq. Observada	12	11	7	8	12	10

A estatística do teste pode ser calculada a partir da fórmula:

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

Em que S é o número de categorias de possíveis resultados. No nosso caso, observamos 6 categorias diferentes (1,2,3,4,5,6).
Calcule o χ^2 para o nosso exemplo. (2.20)

2.1 A tabela do χ^2

A tabela do χ^2 funciona de forma diferente da tabela da Normal padrão.

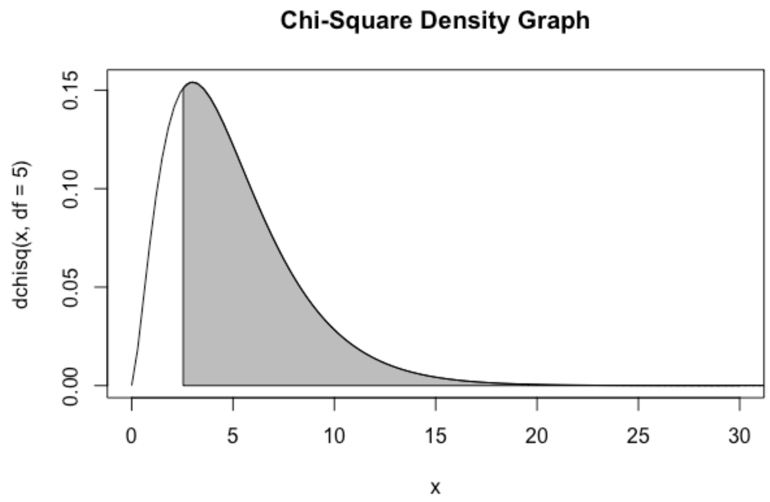
Primeiramente, observe que a primeira coluna da esquerda é referente ao Grau de Liberdade.

Se a hipótese nula for verdadeira, pode-se demonstrar que a estatística χ^2 tem uma distribuição qui-quadrado com s-1 graus de liberdade.

No nosso caso, temos 5 graus de liberdade.

Em seguida, podemos verificar que o χ^2 que calculamos (seguindo a linha correspondente aos 5 graus de liberdade), se encontra entre os p-valores de 90% a 80%.

Ou seja, algo como:



Ou seja, vemos que a estatística observada não nos fornece evidências para descartar a hipótese nula de que o dado seja honesto.

3. Testes de Independência

Objetivo: Verificar se há independência entre duas variáveis medidas nas mesmas unidades amostrais.

O que isso significa?

Lembre-se da aula de análise bidimensional (Aula 3).

Nós já trabalhamos com esse problema ao calcularmos uma medida de associação para duas variáveis qualitativas. Retome o exemplo caso necessário.

Metodologia:

1. Em geral, teremos uma tabela de dupla entrada mais ou menos assim:

   Ai e Bi são categorias das variáveis A e B.

A/B	B1	B2	B3	B4	Total
A1	O_{11}	O_{12}	O_{13}	O_{14}	$O_{1.}$
A2	O_{21}	O_{22}	O_{23}	O_{24}	$O_{2.}$
A3	O_{31}	O_{32}	O_{33}	O_{34}	$O_{3.}$
A4	O_{41}	O_{42}	O_{43}	O_{44}	$O_{4.}$
Total	$O_{.1}$	$O_{.2}$	$O_{.3}$	$O_{.4}$	$O_{..}$

2. Queremos testar a Hipótese nula de que A é independente de B contra a Hipótese alternativa de A e B não são independentes.

3. Depois, teremos que pensar numa tabela ideal caso A e B fossem independentes. Sob essa hipótese, esperaríamos em cada casela:

$$E_{ij} = \frac{O_{i.}O_{.j}}{n}$$

   (i é o número da linha e j, coluna e n, a soma total)

4. Estatística do teste:

$$\chi^2 = \sum_{i=1}^{s}\sum_{j=1}^{r}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

5. Supondo H0 verdadeira, teremos que essa estatística segue uma distribuição Qui-quadrado com (s-1)x(r-1) graus de liberdade (em que s é o número de categorias de A e s é o número de categorias de B).

6. Tomamos uma decisão. Rejeitamos ou não a hipótese nula.

   Obs: A tabela da distribuição de qui-quadrado pode ser utilizada da mesma forma, basta calcular a estatística do teste.