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:

Valores improváveis



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

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

1. 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

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

H1: existe pelo menos alguma diferença

1. Calcula-se a estatística do teste, nesse caso a e o seu respectivo valor-p.
2. 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:

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 para o nosso exemplo. (2.54)

* 1. A tabela do

A tabela do 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 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 que calculamos (seguindo a linha correspondente aos 5 graus de liberdade), se encontra entre os p-valores de 80% a 70%.

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.

1. 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 |  |  |  |  |  |
| A2 |  |  |  |  |  |
| A3 |  |  |  |  |  |
| A4 |  |  |  |  |  |
| Total |  |  |  |  |  |

1. Queremos testar a Hipótese nula de que A é independente de B contra a Hipótese alternativa de A e B não são independentes.
2. Depois, teremos que pensar numa tabela ideal caso A e B fossem independentes. Sob essa hipótese, esperaríamos em cada casela:

(i é o número da linha e j, coluna)

1. Estatística do teste:
2. 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).
3. 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.