Teste Qui-Quadrado para Aderência
aula teórica das disciplinas MAE0110 e MAE0116 da USP
"Noções de Estatística"
Ministrante Prof. Dr. Vladimir Belitsky

30 de junho de 2021

Exemplo 1. Um dos mais simples exemplos do tema Teste de Aderência.

Peguei um dado e o lancei 60 vezes. Eis as frequências observadas:

número da face	1	2	3	4	5	6
a frequência absoluta com a qual foi visto o número em 60 lançamentos	8	12	13	9	12	6

Com base nas frequências observadas, quero deduzir se o dado é equilibrado ou desequilibrado, isto é, quero escolher entre

$$
\begin{aligned}
H &: \; p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6 \\
A &: \; \text{pelo menos um dos } p_1, \ldots, p_6 \text{ é diferente de } 1/6
\end{aligned}
$$

sendo que p_i denota a probabilidade de obter a face i $(i = 1, 2, \ldots, 6)$ num lançamento do dado.

Exemplo 1. Um dos mais simples exemplos do tema Teste de Aderência.

Observação 1. *Os p_i's NÃO se referem diretamente às frequencias relativas observadas no experimento descrito. Quer dizer, por exemplo, p_1 NÃO é $8/60$. Dizer*

$$p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6 \tag{1}$$

é a expressão matemática da frase "o dado é equilibrado". Note que "dado equilibrado" é o termo que foi introduzido por mim quando ensinei a Teoria de Probabilidade nas primeiras aulas do curso, e o sentido desse termo foi definido assim: "quando o dado for laçado, a probabilidade de ober a face 1 é $1/6$, a face 2 é $1/6$ e assim por diante até a face 6". O conceito de probabilidade existe para falar daquilo que <u>vai</u> acontecer. É uma das razões pelas quais a expressão (1) não pode se referir ao passado.

Exemplo 1: Por que o princípio de máxima verossimilhança não funciona diretamente.

Observação 2. *A leitura da discussão sobre a não aplicabilidade direta do Princípio de Máxima Verossimilhança pode ser omitida sem o prejuízo para a compreenção do resto da aula.*

Se A fosse mais específica, por exemplo,

$$A^{esp}: \ p_1 = \frac{1.5}{6}, p_2 = p_3 = p_4 = p_5 = \frac{1}{6}, p_6 = \frac{0.5}{6}$$

então poderíamos solucionar o Exemplo 1 assim: aceitaremos ou rejeitaremos H dependendo de se

$$\boldsymbol{P}\left[X_1 = 8,\, X_2 = 12,\, X_3 = 13,\, X_4 = 9,\, X_5 = 12,\, X_6 = 6 \,\middle|\, H\right]$$

for maior, ou, respectivamente, menor que

$$\boldsymbol{P}\left[X_1 = 8,\, X_2 = 12,\, X_3 = 13,\, X_4 = 9,\, X_5 = 12,\, X_6 = 6 \,\middle|\, A^{esp}\right],$$

onde $X_1, X_2, X_3, X_4, X_5, X_6$ denotam as frequências absolutas a serem vistas se formos lançar 60 vezes o dado equilibrado, para o efeito da conta da primeira das probabilidades, ou o dado desequilibrado de acordo com A^{esp} para o efeito da segunda probabilidade.

Exemplo 1: construindo a solução.

Mas como A não afirma nada específico sobre p_i's, então o caminho da solução baseia-se essencialmente na análise dos resultados que obteríamos caso a hipótese nula (H) fosse válida.

Observação 3. *Aquí, assim como em todos os testes de hipóteses desenvolvidos por nós até o momento, o raciocínio segue os seguintes passos: esquecemos temporariamente que temos uma amostra; pensamos nas propriedades que a amostra teria caso fossemos fazer o experimento, separando os efeitos que surgeriam se H_0 fosse válida dos efeito que surgeriam se H_0 fosse inválida; por fim, valtamos à amostra (áquela que foi temporareimante esquecida) e verificamos qual dos dois conjuntos de efeitos ela apresenta, em em função disso decidimos se aceitamos ou rejeitamos H_0.*

As propriedades que a amostra teria serão derivadas com base em **três pilares.**

Exemplo 1: construindo a solução.

Primeiro pilar: se H valesse, isto é, se o dado lançado fosse equilibrado, então na situação ideal, iriamos obter

$$
\begin{aligned}
&e_1 = \tfrac{1}{6} \times 60 = 10 \text{ faces "1"} \\
&e_2 = \tfrac{1}{6} \times 60 = 10 \text{ faces "2"} \\
&\ldots \\
&e_6 = \tfrac{1}{6} \times 60 = 10 \text{ faces "6"}
\end{aligned}
$$

A "situação ideal" significa a ausência de quaisquer desvios para mais ou para menos de qualquer um destes valores, quer dizer, aquí a gente imagina que não há flutuação aleatórias de valores que ocorrem devido à presença de aleatoriedade no experimento. A outra visão sobre e_i's (também correta) é que eles correspondem às experanças matemáticas das respectivas frequencias absolutas de aparência das faces em 60 lançamentos. Por essa razão, os e_i's chama-se **valores esperados**.

Exemplo 1: construindo a solução.

Segundo pilar: Mas os desvios desconsiderados acima acontecerão, e para medi-los contruimos a variável aleatória

$$\chi^2 = \frac{(X_1 - e_1)^2}{e_1} + \frac{(X_2 - e_2)^2}{e_2} + \cdots + \frac{(X_6 - e_6)^2}{e_6}$$

onde $X_1, X_2, X_3, X_4, X_5, X_6$ são as variáveis aleatórias que representam as frequências absolutas a serem vistas em n lançamentos de um dado. No presente caso, temos que $n = 60$ e temos os valores de $e_1, \ldots, e_6$ já calculados acima. Então

$$\chi^2 = \frac{(X_1 - 10)^2}{10} + \frac{(X_2 - 10)^2}{10} + \cdots + \frac{(X_6 - 10)^2}{10}$$

Presate a atenção que, de acordo com o programa de raciocínio tracado na Obsrevação 3, estamos olhando no nosso experimento aleatório como se fosse que ele ainda irá acontecer. É por isso que as frequencias absolutas de aparência dos faces do dados são “codificados” agora por variáveis aleatórias.

Exemplo 1: construindo a solução.

Crença do segundo pilar: Cada variável aleatória X_i fica perto de seu valor esperado com grande probabilidade, e faca longe desse valor com probabilidade pequena.
Continuamos com a crença: Se for H quem "rege" as X's, isto é, se o dado lançado for equilibrado, então o valor experado de cada X_i é (de acordo com nossa definição e notação) e_i, e portanto, para cada i, os valores de X_i ficam próximos a e_i com grande probabilidade, e ficam longe de e_i com probabilidade pequena. Isso acarreta no que na maioria dos casos (ou, em outras palavras, com grande probabilidade), o valor observado de χ^2 deve ser pequeno.
Já se for A quem "rege" as X's, então para alguns i, os valores de X_i estarão concentrados ao redor de sua esperança que não é e_i, e portanto na maioria dos casos (ou, em outras palavras, com grande probabilidade), o valor observado de χ^2 deve ser grande.

Exemplo 1: construindo a solução.

A Crença do Segundo Pilar sugere a seguinte

Regra de Decisão (versão preliminar)**:** se $(\chi^2)_{obs}$, valor observado da distância χ^2, for pequeno, devemos aceitar H (aceitar que o dado lançado era equilibrado), já se for grande, devemos aceitar A (aceitar que o dado lançado era desequilibrado).

Observação 4. (Pode ser omitida.) *Observe que a Regra escolhe entre H e A via a maximização de verossimilhança baseada no valor* $(\chi^2)_{obs}$.

Mas como quantificar "grande" e "pequeno" da Regra de Decisão? Por exemplo, imagine que os valores observados das variáveis X's fossem assim:

$$(x_1)_{obs} = 8, (x_2)_{obs} = 12, (x_3)_{obs} = 13,$$
$$(x_4)_{obs} = 9, (x_5)_{obs} = 12, (x_6)_{obs} = 6$$

Exemplo 1: construindo a solução.

o que daria o seguinte valor observado $(\chi^2)_{obs}$ da variável χ^2:

$$
\begin{aligned}
(\chi^2)_{obs} = \; & \frac{(8-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(13-10)^2}{10} \\
& + \frac{(9-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(6-10)^2}{10} = 3,80
\end{aligned}
$$

Então, $3,80$ é um valor pequeno o suficiente para podermos deduzir, por nossa Regra de Decisão, que o dado lançado era equilibrado (em outras palavras, aceitar H), ou talvez, seja um valor grande do ponto de vista desta Regra e, consequentemente, devemos rejeitar H em favor a A e deduzir que o dado era desequilibrado?

A regra que define quem é pequeno e quem é grande será contruída com base no Terceiro Pilar que está explicado em seguida.

Exemplo 1: construindo a solução.

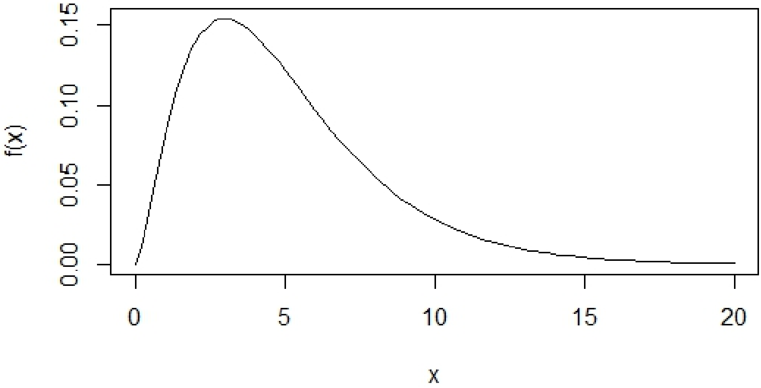
O terceiro pilar da solução baseia-se na análise da distribuição da variável aleatória χ^2 no caso quando as variáveis X's envolvidas na sua construção, possuem as distribuições regidas pela hipótese nula, H.

Tal análise (que não será apresentada no nosso curso) disse o seguinte: A distribuição da variável aleatória χ^2, sob a validade de H, pode ser bem aproximada pela distribuição chamada **Qui-Quadrado com 5 graus de liberdade**, cuja expressão analítica é conhecida e cujo formato está na figura da transparência seguinte.

No que se segue, vamos omitir e esquecer que esta distribuição é só uma aproximação para a variável que nos interessa, e vamos pensar como se fosse que a função Qui-Quadrado com 5 graus de liberdade é a verdaderira distribuição de χ^2 no caso quando H for verdadeira.

Exemplo 1: construindo a solução.

$$\chi^2 = \frac{(X_1 - 10)^2}{10} + \frac{(X_2 - 10)^2}{10} + \cdots + \frac{(X_6 - 10)^2}{10}$$



Exemplo 1: construindo a solução.

Do terceiro pilar deduzem-se as seguintes propriedades da distribuição da variávela aleatória χ^2 sob a validade da hipótese H (as propriedades abaixo listadas é o que vemos olhando o gráfico da figura apresentada na transparência anterior):

(a) os valores pequenos ocorrem com probabilidade pequena;

(b) os valores que ocorrem com as maiores probabilidades são aqueles nos quais apoia-se a corcova da função;

(c) os valores maiores que os do apóio da corcova ocorrem com probabilidades pequenas, e quanto mais afasta-se a cauda à direita da corcova, menor a probabilidade de ocorrência de seus valores.

Exemplo 1: construindo a solução.

Os três pilares supracitados sustentam a seguinte

Regra de Decisão (vesão definitiva): Escolher um limiar ℓ do suporte da cauda da distribuição que a variável aleatória χ^2 possui quando H é válida,
e
caso $(\chi^2)_{obs}$, valor observado de χ^2, for menor que ℓ, devemos aceitar H (aceitar que o dado lançado era equilibrado),
já se $(\chi^2)_{obs}$ for maior que ou igual a ℓ, devemos aceitar A (aceitar que o dado lançado era desequilibrado).

Notamos que há a possibilidade do
Erro da Regra de Decisão: mesmo lançando um dado equilibrado, a Regra pode decidir que este é desequilibrado.

Devido à construção, a probabilidade deste erro é a cauda da distribuição de χ^2 à direita do corte ℓ.

Exemplo 1: construindo a solução.

O erro supracitado, chama-se **Erro do tipo I**.
Sua probabilidade chama-se **nível de significância** e denota-se tipicamente por α.
O semi-eixo $[\ell, \infty)$ chama-se **a região crítica**.
Obviamente, existe o erro do outro sentido:

> se na realidade lança-se um dado desequilibrado, a Regra pode decidir que ele era equilibrado.

Este chama-se **Erro do Tipo II**. Mas sua probabilidade não é expressada por um valor (pois a hipótese alternativa não especifica o desequilíbrio do dado). Por isto que o "erro no outro sentido" não será refletido em nossas respostas. (Na realidade, ele pode ser estudado e os resultados deste permitem comparar o "poder" de diversas regras.)

Passos de execução do Teste de Aderência.

(i) Identificar n, o número de observações (tamanho da amostra).

(ii) Identificar K, o número de classes de observações.

(iii) Identificar a distribuição teórica por classes de acordo com o enunciado; esta distribuição são valores, os quais denotamos aquí genericamente por

$$(p_1)_{teor}, (p_2)_{teor}, \ldots, (p_K)_{teor}$$

(iv) Escrever as hipóteses

$$\begin{array}{rcl} H & : & p_1 = (p_1)_{teor}, \ldots, p_K = (p_K)_{teor} \\ A & : & \text{pelo menos um dos } p_i \text{ é diferente de } (p_i)_{teor} \end{array}$$

(v) Calcular as frequências esperadas:

$$e_1 = n \times (p_1)_{teor}, \ldots, e_K = n \times (p_K)_{teor}$$

Passos de execução do Teste de Aderência; continuação

(vi) Identificar as frequências observadas

$$(x_1)_{obs},\ (x_2)_{obs},\ \ldots,\ (x_K)_{obs}$$

do experimento descrito no enunciado.

(vii) Calcular a observação $(\chi^2)_{obs}$ que a variável χ_2 assume para as frequências observadas; a fórmula é:

$$\left(\chi^2\right)_{obs} = \sum_{i=1}^{K} \frac{((x_i)_{obs} - e_i)^2}{e_i}$$

(viii-a) Identificar o nível de significânica α. Na vida real, você receberá este da pessoa interessada no teste; no nosso curso, você receberá o nível de significância junto com o enunciado do problema.

(ix-a) A partir da tabela da distribuição Qui-Quadrado com $K-1$ graus de liberdade determinar o valor de ℓ como o limiar que "recorta" cauda de peso α desta distribuição.

Passos de execução do Teste de Aderência; continuação

(viii-b) Identificar o valor do corte ℓ. Na vida real, você receberá este da pessoa interessada no teste; no nosso curso, você receberá ℓ junto com o enunciado do problema.

(ix-a) A partir da tabela da distribuição Qui-Quadrado com $K - 1$ graus de liberdade determinar a probabilidade da cauda da distribuição que se encontra à direita do limiar ℓ. Esta probabilidade é nível de significância α do teste.

(x) Comparar $(\chi^2)_{obs}$ com ℓ e concluir:

$$\begin{array}{l} \text{aceitar } H \text{ caso } (\chi^2)_{obs} < \ell; \\ \text{aceitar } A \text{ caso } (\chi^2)_{obs} \geq \ell. \end{array}$$

(xi) Formular a conclusão: O Teste de Aderência baseado na distribuição Qui-Quadrado indicou que, ao nível de significância α, os dados amostrais confirmam/não confirmam H.

Execução do Teste de Aderência para os dados do Exemplo 1.

No exemplo, imagine que os valores observados das variáveis X's foram assim:

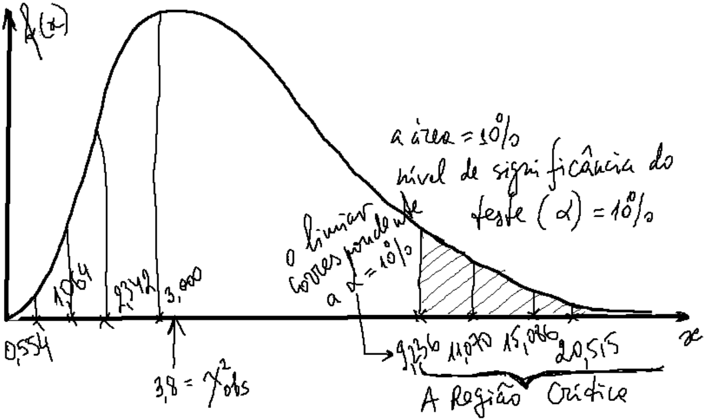
$$(x_1)_{obs} = 8, (x_2)_{obs} = 12, (x_3)_{obs} = 13,$$
$$(x_4)_{obs} = 9, (x_5)_{obs} = 12, (x_6)_{obs} = 6$$

o que dá o seguinte valor observado $(\chi^2)_{obs}$ da variável χ^2:

$$\begin{aligned}(\chi^2)_{obs} = \quad & \frac{(8-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(13-10)^2}{10} \\ & + \frac{(9-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(6-10)^2}{10} = 3,80\end{aligned}$$

Imagine que desejamos que o teste de aderência seja proferido com o nível de significância $\alpha = 10\%$. Consultando a Tabela da Distribuição Qui-Quadrado, na sua linha correspondete a G.L.=5 (5=6-1), acamo o valor do limiar ℓ: ele é $9,236$. Como $(\chi^2)_{obs} < \ell$, aceitamos H; isto é, a amostra indica que lançavamos dados equilibrado

Execução do Teste de Aderência para os dados do Exemplo 1



A distribuição Qui-Quadrado; o que precisa saber sobre ela.

Há uma família de funções chamadas **distribuições Qui-Quadrado**.

As funções da família dependem de um parâmetro chamado **número de graus de liberdade**. O nome reflete uma propriedade intrínseca por trás da construção da funções, mas esta não será abordada na presente aula.

Suas expressões analíticas são conhecidas, mas o cálculo de integrais envolvendo as funções é complicado. Para simplificar, há tabela que para cada inteiro G, relaciona o limiar de corte ℓ e a área da cauda recortada por ℓ da distribuição Qui-Quadrado com G graus de liberdade.

Tabela 1 - Distribuição Qui-Quadrado (χ^2)

Corpo da tabela fornece os valores de ℓ tais que $P(\chi^2 \geq \ell) = \alpha$

G.L.	α=99%	90%	80%	70%	10%	5%	1%	0.1%
1	0.000	0.016	0.064	0.148	2.705	3.841	6.635	10.828
2	0.020	0.211	0.446	0.713	4.605	5.992	9.210	13.816
3	0.115	0.584	1.005	1.424	6.251	7.815	11.345	16.266
4	0.297	1.064	1.649	2.195	7.779	9.488	13.277	18.467
5	0.554	1.610	2.342	3.000	9.236	11.070	15.086	20.515
6	0.872	2.204	3.070	3.828	10.645	12.592	16.812	22.458
7	1.239	2.833	3.822	4.671	12.017	14.067	18.475	24.322
8	1.647	3.490	4.594	5.527	13.362	15.507	20.090	26.125
9	2.088	4.168	5.380	6.393	14.684	16.919	21.666	27.877
10	2.558	4.865	6.179	7.267	15.987	18.307	23.209	29.588
11	3.054	5.578	6.989	8.148	17.275	19.675	24.725	31.264
12	3.571	6.304	7.807	9.034	18.549	21.026	26.217	32.910
13	4.107	7.042	8.634	9.926	19.812	22.362	27.688	34.528
14	4.660	7.790	9.467	10.822	21.064	23.685	29.141	36.123
15	5.229	8.547	10.307	11.721	22.307	24.996	30.578	37.697

Observação: Uma tabela mais detalhada da distribuição Qui-Quadrado está disponibilizada junto com listas de exercícios referentes ao tema Teste Qui-Quadrado.

A distribuição Qui-Quadrado; o que você precisa saber sobre ela.

A distribuição Qui-Quadrado com $K-1$ graus de liberdade serve como uma boa aproximação para a distribuição da variável aleatória

$$\chi^2 = \sum_{i=1}^{k} \frac{(X_i - n \times valor_i)^2}{n \times valor_i}$$

sob a validade da hipótese

$$H: \; p_1 = valor_1, \cdots, p_k = valor_k$$

Quanto maior o tamanho da amostra (por exemplo, no caso do Exemplo 1, tal tamanho é 60), melhor a qualidade da aproximação. Não abordaremos a questão da qualidade de aproximação e usaremos esta para qualquer n.

A distribuição Qui-Quadrado; o que você precisa saber sobre ela.

Devido ao uso da aproximação pelas funções Qui-Quadrado, surge a seguinte particularidade na aplicação da supradescrita Regra de Decisão para Teste de Aderência:
se na amostra $(x_1)_{obs}, \cdots, (x_K)_{obs}$ há valores menores que 5, então as correspondentes classes devem ser agrupadas algumas às outras, para eliminarmos tais valores.
Por exemplo, se em $n = 60$ lançamentos de um dado observamos as frequências $(x_1)_{obs} = 13$, $(x_2)_{obs} = 4$, $(x_3)_{obs} = 10$, $(x_4)_{obs} = 7$, $(x_5)_{obs} = 18$, $(x_6)_{obs} = 8$, então (um dos possíveis caminhos de agrupamento) é "juntar" as classes das faces "1" e "2". Isto transforma a hipótese

$$H: \ p_1 = \cdots = p_6 = \frac{1}{6} \text{ em } H: \ p_{1,2} = \frac{2}{6}, p_3 = \cdots p_6 = \frac{1}{6},$$

e transforma a amostra para ser

$$(x_{1,2})_{obs} = 17, (x_3)_{obs} = 10, (x_4)_{obs} = 7, (x_5)_{obs} = 18, (x_6)_{obs} = 8$$