

INFERÊNCIA ESTATÍSTICA (EST0035)

AULA06: TESTES DE HIPÓTESES (PARTE II)

Frederico Machado Almeida
frederico.almeida@unb.br

Departamento de Estatística
Instituto de Exatas
Universidade de Brasília (UnB)

Testes Uniformemente Mais Poderosos

- Os *TMP* apresentado anteriormente pressupõem que as duas hipóteses (H_0 e H_1) sejam ambas simples. Ou seja, que a distribuição da variável aleatória X seja completamente especificada.
- No entanto, em alguns casos, podemos nos deparar com problemas em que a H_0 é simples, mas a H_1 é composta, ou as duas podem ser simultaneamente compostas, como foi apresentado no início desse tópico.
- Nessa parte da matéria proponho uma generalização dos *TMP* introduzidos anteriormente. Ou seja, vamos construir *melhores regiões críticas* RC com tamanho $\alpha \leq \pi_\tau(\theta'')$, com $\theta'' \in \Theta_1$.
- Diferentemente dos *TMP* (que sempre existem), os Testes Uniformemente Mais Poderosos (*TUMP*) podem não existir para determinados casos.

Testes Uniformemente Mais Poderosos

Definição 1

A região crítica RC de um teste τ qualquer é dita ser Uniformemente Mais Poderosa com tamanho α para testar uma H_0 (simples ou composta) e uma H_1 composta, se RC for a melhor região crítica de tamanho α (segundo a definição apresentada nos TMP), e se para qualquer $\theta'' \in \Theta_1$ a melhor região crítica seja sempre a mesma.

- A definição 1 sustenta que, se RC é a melhor região crítica de um TUMP, então, qualquer escolha de $\theta'' \in \Theta_1$ não alterar o tipo de região crítica. Ou seja, *não pode resultar em uma região crítica oposta à anterior.*

Testes Uniformemente Mais Poderosos

Segue da Definição 1 que,

- i O teste deve ter tamanho α .
- ii Deve existir algum $\theta' \in \Theta_0$, tal que, $\mathbb{P}_{H_0}(T(\mathbf{X}) \in RC) = \alpha$.
- iii Para todo $\theta' \in \Theta_0$ dado em (ii) e cada $\theta'' \in \Theta_1$, $\exists k > 0$, tal que,

$$\phi(x) = \begin{cases} 1, & \text{se } R(\mathbf{x}) \leq k, \\ 0, & \text{se } R(\mathbf{x}) > k. \end{cases}$$

- Conforme argumentado anteriormente, um *TUMP* nem sempre vai existir.
- Para caso em que X é proveniente de uma distribuição discreta, a função de decisão pode ser reescrita incluindo o caso em que $R(\mathbf{x}) = k$, que ocorre com probabilidade γ .

Testes Uniformemente Mais Poderosos

Exemplo 1: Sejam X_1, X_2, \dots, X_n cópias *iid*'s de $X \sim \mathcal{N}(\theta, 1)$. Assuma que as hipóteses de interesse seja dadas por: $H_0 : \theta = 0$ contra $H_1 : \theta > 0$. Ou podemos representar as hipóteses de interesse da seguinte maneira, $H_0 : \theta = \theta'$ ($\theta' = 0$) contra $H_1 : \theta = \theta''$ ($\theta'' > 0$). Segue imediatamente que a função de verossimilhança é dada por:

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n (x_i|\theta) = (2\pi)^{-n/2} \exp \left\{ -1/2 \sum_{i=1}^n (x_i - \theta)^2 \right\}.$$

Assim, as funções de verossimilhança sob H_0 e H_1 são dadas por:

$$\text{Sob } H_0: L(\theta'|\mathbf{x}) = (2\pi)^{-n/2} \exp \left\{ -1/2 \sum_{i=1}^n x_i^2 \right\}.$$

$$\text{Sob } H_1: L(\theta''|\mathbf{x}) = (2\pi)^{-n/2} \exp \left\{ -1/2 \sum_{i=1}^n (x_i - \theta'')^2 \right\}.$$

Testes Uniformemente Mais Poderosos

Com base nas quantidades apresentadas anteriormente, a razão de verossimilhanças sob H_0 e H_1 é dada por:

$$\begin{aligned} R(\mathbf{x}) &= \frac{L(\theta'|\mathbf{x})}{L(\theta''|\mathbf{x})} = \frac{(2\pi)^{-n/2} \exp\left\{-1/2 \sum_{i=1}^n x_i^2\right\}}{(2\pi)^{-n/2} \exp\left\{-1/2 \sum_{i=1}^n (x_i - \theta'')^2\right\}} \\ &= \exp\left\{-\theta'' \sum_{i=1}^n x_i + \frac{n}{2} (\theta'')^2\right\}. \end{aligned}$$

Portanto, para $k > 0$, a H_0 será rejeitada à favor da H_1 se

$$R(\mathbf{x}) \leq k \iff \exp\left\{-\theta'' \sum_{i=1}^n x_i + \frac{n}{2} (\theta'')^2\right\} \leq k \iff \sum_{i=1}^n x_i \geq k^* = \left(\log(k) - \frac{n}{2} (\theta'')^2\right) / \theta''.$$

Testes Uniformemente Mais Poderosos

- Ou seja, segue do resultado anterior que $RC = \{\mathbf{x} : \sum_{i=1}^n x_i \geq k^*\}$.
Desta forma, como para $\forall \theta'' \in \Theta_1$, RC é unicamente definida, então, essa é a *melhor região crítica do TUMP* para as hipótese apresentadas anteriormente.
- Como $\bar{X}_n \sim \mathcal{N}(\theta, 1/n)$, com $\bar{X}_n = \sum_{i=1}^n X_i$, segue que,
 $RC = \{\mathbf{x} : \bar{x} \geq k^{**}\}$, com $k^{**} = k^*/n$. Para algum α dado, k^{**} pode ser obtido, tal que,

$$\alpha = \mathbb{P}_{H_0}(\bar{X}_n \geq k^{**}).$$

- Sendo $RC = \{\mathbf{x} : \bar{x} \geq k^{**}\}$ a melhor região crítica do teste anterior, segue que, $\alpha \leq \pi_{\tau}(\theta'')$.

Testes Uniformemente Mais Poderosos

Exemplo 2: Assuma que X_1, X_2, \dots, X_n denotam cópias *iid*'s de $X \sim \text{Ber}(\theta)$, com $\theta \in \Theta = \{\theta : 0 < \theta < 1\}$. Considere as seguintes hipóteses $H_0 : \theta = \theta'$ ($\theta' = 0,5$) contra $H_1 : \theta = \theta''$ ($\theta'' < 0,5$). Conforme apresentado no Exemplo 8, a função de verossimilhança da distribuição Bernoulli é dada por:

$$L(\theta|\mathbf{x}) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i}.$$

Logo, as funções de verossimilhança avaliadas nos pontos θ' e θ'' são dadas por,

$$\text{Sob } H_0: L(\theta'|\mathbf{x}) = (\theta')^{\sum_{i=1}^n x_i} (1 - \theta')^{n - \sum_{i=1}^n x_i} = 0,5^n.$$

$$\text{Sob } H_1: L(\theta''|\mathbf{x}) = (\theta'')^{\sum_{i=1}^n x_i} (1 - \theta'')^{n - \sum_{i=1}^n x_i}.$$

Testes Uniformemente Mais Poderosos

Segue de forma similar ao Teorema de Neyman-Pearson que,

$$R(\mathbf{x}) = \frac{L(\theta'|\mathbf{x})}{L(\theta''|\mathbf{x})} = \frac{0,5^n}{(\theta'')^{\sum_{i=1}^n x_i} (1 - \theta'')^{n - \sum_{i=1}^n x_i}} = 0,5^n (1/\theta'' - 1)^{\sum_{i=1}^n x_i}.$$

- Desta forma, para $k > 0$ a região crítica será tal que,

$$0,5^n (1/\theta'' - 1)^{\sum_{i=1}^n x_i} < k \iff RC = \{\mathbf{x} : \sum_{i=1}^n x_i < k^*\}, \text{ com}$$

$k^* = \left\lceil \log \left(k \left[\frac{1-\theta''}{0,5} \right]^n \right) \right\rceil / \log(\theta''(1 - \theta''))$. Onde para α dado, k^* é obtido tal que,

$$\alpha = \mathbb{P}_{H_0} \left(\sum_{i=1}^n X_i < k^* \right),$$

com $\sum_{i=1}^n X_i \sim \text{Binom}(n, \theta)$. Logo, RC é a *melhor região crítica do TUMP* para testar H_0 e H_1 , $\forall \theta'' \in \Theta_1$.

Testes Uniformemente Mais Poderosos

Portanto, como estamos perante uma variável aleatória discreta, a função de decisão $\phi(\mathbf{x})$ é dada por:

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{se } R(\mathbf{x}) < k, \\ \gamma, & \text{se } R(\mathbf{x}) = k, \\ 0, & \text{se } R(\mathbf{x}) > k. \end{cases}$$

Com $\gamma = \mathbb{P}_{H_0} \left(\sum_{i=1} X_i = k^* \right)$.

Testes Uniformemente Mais Poderosos

Exemplo 3: Considere novamente o enunciado do *Exemplo 10*, onde X_1, X_2, \dots, X_n são cópias *iid*'s de $X \sim \mathcal{N}(\theta, 1)$. Porém, agora vamos considerar as seguintes hipóteses de interesse $H_0 : \theta = 0$ contra $H_1 : \theta \neq 0$. Ou seja, $H_0 : \theta = \theta'$ ($\theta' = 0$) contra $H_1 : \theta = \theta''$ ($\theta'' \neq \theta'$).

Desta forma, segue do Teorema de Neyman-Pearson que a estatística $R(\mathbf{x})$ tem a seguinte forma,

$$\begin{aligned} R(\mathbf{x}) &= \frac{L(\theta'|\mathbf{x})}{L(\theta''|\mathbf{x})} = \frac{(2\pi)^{-n/2} \exp\left\{-1/2 \sum_{i=1}^n x_i^2\right\}}{(2\pi)^{-n/2} \exp\left\{-1/2 \sum_{i=1}^n (x_i - \theta'')^2\right\}} \\ &= \exp\left\{-\theta'' \sum_{i=1}^n x_i + \frac{n}{2} (\theta'')^2\right\}. \end{aligned}$$

Testes Uniformemente Mais Poderosos

Portanto, observe que, para alguma constante $k > 0$, temos que,
$$\log [R(\mathbf{x})] = -\theta'' \sum_{i=1}^n x_i + \frac{n}{2} (\theta'')^2 \leq \log(k).$$
 Entretanto, como estamos perante uma H_1 bilateral ($\theta'' < 0$ ou $\theta'' > 0$) então, vamos considerar as seguintes situações:

- ❶ Supondo $\theta'' < 0$ implica que, $RC = \{\mathbf{x} : \sum_{i=1}^n x_i \leq k^*\}$, com $k^* = \log(k) - 1$.
- ❷ Supondo $\theta'' > 0$ implica que, $RC = \{\mathbf{x} : \sum_{i=1}^n x_i \geq k^*\}$, com $k^* = 1 - \log(k)$.

Ou seja, valores diferentes de $\theta'' \in \Theta_1$ forneceram regiões críticas diferentes (ou conflitantes). Assim, *NÃO* existe um *TUMP* para testar as hipóteses apresentadas anteriormente.

Testes Uniformemente Mais Poderosos

- Observe que, no Exemplo 11 não existe um TUMP para uma H_1 bilateral. Portanto, definindo hipóteses unilaterais, isto é, (i) $H_1 : \theta < 0$ ou (ii) $H_1 : \theta > 0$ separadamente, existirá para cada caso, a melhor região crítica RC de tamanho α , e o teste será *Uniformemente Mais Poderoso*.

TESTES NÃO-PARAMÉTRICOS

Teste Não-Paramétricos

- Os métodos de inferência abordados até então são chamados de “Métodos Paramétricos”. E para a sua condução, eles requerem que a distribuição no qual as amostras foram geradas seja previamente especificada (distribuição normal preferencialmente).
- Caso a distribuição populacional seja desconhecida, é necessário que o tamanho de amostra seja grande. Pois, nessas circunstâncias o TCL é fundamental para aproximar a distribuição dos dados para a $\mathcal{N}(0, 1)$.
- Quando nenhuma das suposições acima é satisfeita, os métodos não-paramétricos são uma alternativa para contornar as fragilidades encontradas nos modelos paramétricos. Além disso, esses métodos visam analisar contagens em uma tabela de contingência (tabela de dupla entrada).

Teste Não-Paramétricos

- Os modelos paramétricos abordados até o momento possuem as seguintes características principais:
 - 1 Incidem explicitamente sobre um parâmetro de uma ou mais populações, por exemplo, a média: μ , variância: σ^2 , etc.
 - 2 Pressupõem uma forma particular para as distribuições das populações envolvidas, como por exemplo, a normalidade dos dados.
- Os testes que estudaremos em seguida não possuem pelo menos uma das características anteriores e são designados por *testes não-paramétricos* ou *testes adistribucionais*.

Teste Qui-Quadrado

O teste de Qui-quadrado é aplicado quando estão em comparação dois ou mais grupos independentes não necessariamente do mesmo tamanho. A variável deve ser de mensuração nominal. O teste Qui-quadrado não tem equivalente nos paramétricos. Assim, dependendo do tipo de delineamento, o teste pode ser de:

- **Qualidade de Ajuste:** são testes a aferir se um conjunto de dados (discretos ou contínuos) aderem ou não a uma determinada distribuição.
- **Aderência:** neste tipo de teste, testamos se uma determinada amostra é proveniente de uma suposta distribuição de probabilidade.
- **Independência (ou Associação):** neste teste, temos interesse em verificar se as variáveis com distribuições nas marginais são independentes ou não.
- **Homogeneidade:** testamos se uma determinada variável se distribui da mesma forma em várias populações de interesse.

Teste Não-Paramétricos

Confrontando os dois tipos de modelos estatísticos *paramétricos* vs *não-paramétricos*, é possível definir as seguintes características gerais:

- Ambos requerem a recolha de amostras aleatórias.
- Ambos permitem inferir a cerca da população.
- Em geral, os testes não-paramétricos envolvem cálculos mais simples.
- Os testes não-paramétricos usam, por vezes, em substituição dos dados, as suas ordens, pelo que a mediana e a amplitude são as medidas preferenciais de localização e dispersão, respectivamente.
- Os testes não-paramétricos são em geral menos potentes (em consequência do uso das ordens) quando comparados com os testes paramétricos, quando são aplicáveis.
- Sempre que a aplicação dos testes paramétricos não se verifiquem, ou quando as variáveis em estudo não forem quantitativas, devem-se utilizar os testes não-paramétrico, desde que as próprias condições se verifiquem.

Testes de Ajustamento

- Os testes de ajustamento (ou de qualidade de ajuste) permitem aferir a cerca da forma da distribuição de probabilidade da variável em estudo na população, com base na distribuição de *frequências relativas acumuladas*.
- Estes testes possibilitam testar se determinada função de distribuição empírica se *ajusta* a uma determinada distribuição *teórica* (*distribuição alvo*).
- Por exemplo, podemos estar interessados em testar se determinado conjunto de observações para uma certa variável aleatória X permite não rejeitar a hipótese de que tal variável seguem uma distribuição, digamos, exponencial por exemplo.

Testes de Ajustamento de Kolmogorov-Smirnov

- O teste de Kolmogorov-Smirnov (KS) comumente utilizado para testar a normalidade dos dados, baseia-se na distribuição dos dados observados (*distribuição empírica*), com a função de distribuição teórica (*sob H_0*), e subsequente determinação do ponto onde *existe a maior distância vertical entre ambas as funções*.
- Em linhas gerais, as hipóteses de interesse em um teste de qualidade de ajuste de KS são:

$$H_0 : f(x) = f_0(x) \text{ contra } H_1 : f(x) \neq f_0(x). \quad (1)$$

Ou de forma equivalente, as hipóteses em 1 podem ser reformuladas da seguinte forma:

$$H_0 : X \sim f_0(x) \text{ contra } H_1 : X \not\sim f_0(x). \quad (2)$$

Onde $f(x)$ e $f_0(x)$ denotam as *fdp's* obtidas sob H_1 e H_0 , respectivamente.

Testes de Ajustamento de Kolmogorov-Smirnov

- Designemos por $F_0(x)$ a função de distribuição (*fda*) teórica correspondente a *fdp* obtida sob H_0 . tal que,

$$F_0(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f_0(y) dy.$$

- Denote igualmente por $F(x)$ a *fda* (*empírica*) ou as frequências relativas acumuladas obtidas em uma amostra aleatória de tamanho n . Onde para cada ponto observado x_i , a acumulada é dada por:

$$F(x_i) = \mathbb{P}(X \leq x_i) = \sum_{i=1}^{x_i} \frac{n_i}{n} = \sum_{i=1}^{x_i} f_{ri},$$

onde f_{ri} denota as frequências relativas. E $F(x_i)$ dá-nos a proporção de *valores observados* menores ou iguais a x_i (recorde que n_i denota as frequências absolutas simples).

Testes de Ajustamento de Kolmogorov-Smirnov

- A estatística de teste é dada por (Zar, 1984):

$$\begin{aligned}\mathcal{D}_n &= \max \left\{ \max \left| F(x_i) - F_0(x_i) \right|; \max \left| F(x_{i-1}) - F_0(x_i) \right| \right\} \\ &= \max \left\{ \left| F(x_i) - F_0(x_i) \right|; \left| F(x_{i-1}) - F_0(x_i) \right| \right\}.\end{aligned}$$

- No caso de amostras pequenas, digamos $n \leq 40$ a distribuição amostral exatas da estatística \mathcal{D}_n está tabelada como função de n e do nível de significância.
- Para amostras de dimensão superior, a distribuição assintótica de \mathcal{D}_n está especificada na tabela de KS, em função de α . Assim, RH_0 à favor da H_1 se $\mathcal{D}_n \geq d_{(\alpha, n)}$.

Testes de Ajustamento de Kolmogorov-Smirnov

Exemplo 1: Uma determinada empresa decide promover um funcionário para cargo de chefia. No processo de seleção do funcionário a empresa pretende, antes de mais, averiguar a competência dos funcionários. Para tal, foram selecionados ao acaso 12 funcionários, sendo estes sido sujeitos a uma prova cotada de 0 a 20 pontos. Sabendo que em testes anteriores as classificações apresentaram uma distribuição normal com média $\mu = 12$ e desvio-padrão, $\sigma = 2$ pontos, teste se as classificações deste teste apresentam a mesma distribuição. Ou seja,

$$H_0 : X \sim \mathcal{N}(12, 2^2) \text{ contra } H_0 : X \approx \mathcal{N}(12, 2^2).$$

Para testar a veracidade ou não da H_0 , os dados da amostra de 12 funcionários estão apresentados na tabela a seguir

Testes de Ajustamento de Kolmogorov-Smirnov

i	x_i	n_i	N_i	$F(x_i)$	$F_0(x_i)$	$F(x_i) - F_0(x_i)$	$F(x_{i-1}) - F_0(x_i)$
1	12,1	1	1	0,083	0,520	0,437	0,520
2	12,2	1	2	0,167	0,540	0,373	0,456
3	12,5	1	3	0,250	0,599	0,349	0,432
4	12,8	1	4	0,333	0,655	0,322	0,405
5	13,0	1	5	0,417	0,691	0,275	0,358
6	13,3	1	6	0,500	0,742	0,242	0,325
7	13,6	1	7	0,583	0,788	0,205	0,288
8	15,5	1	8	0,667	0,960	0,293	0,377
9	16,8	2	10	0,833	0,992	0,158	0,325
10	18,0	1	11	0,917	0,999	0,082	0,165
11	18,5	1	12	1,000	0,999	0,001	0,083

Teste Qui-Quadrado de Aderência

O teste qui-quadrado de *aderência* (*também chamado de teste ou prova de ajustamento*), serve para testar a hipótese de que os dados de frequências se distribuem de acordo com alguma teoria postulada. Isto é, o teste objectiva, testar a adequabilidade de um modelo probabilístico a um conjunto de dados observados.

Diferentemente dos casos apresentados até agora, a ideia que norteia esse tipo de teste, é assente na base das frequências esperadas usando a fórmula $E_i = np_i$. Sendo n o número total de observações e p_i a probabilidade associada a cada uma das categorias.

Teste Qui-Quadrado de Aderência

Considere uma tabela de distribuição de frequências, com $k > 2$ categorias de resultados:

Categorias	Frequência Observada
1	O_1
2	O_2
3	O_3
\vdots	\vdots
k	O_k
Total	n

em que O_i é o total de indivíduos observados na categoria i , tal que $(i = 1, 2, \dots, k)$.

Teste Qui-Quadrado para Aderência

Seja p_i a probabilidade associada à categoria i . O objetivo do teste de aderência é testar as hipóteses:

$$H_0 : p_1 = p_{o1}, \dots, p_k = p_{ok}$$

$$H_1 : \text{existe pelo menos uma diferença.}$$

Sendo p_{oi} a probabilidade especificada para a categoria i , fixada através do modelo probabilístico de interesse. As frequências esperadas E_i são calculadas tal como foi apresentado anteriormente. Isto é, supondo verdadeira a hipótese nula,

$$E_i = n \times p_{oi} \text{ com } i = 1, 2, \dots, k.$$

Teste Qui-Quadrado para Aderência

Categorias	Frequência Observada	Frequência Esperada
1	O_1	E_1
2	O_2	E_2
3	O_3	E_3
\vdots	\vdots	\vdots
k	O_k	E_k
Total	n	n

Uma vez mais, a distância (ou diferença) entre as colunas de frequências é quantificada por meio da estatística qui-quadrado, dada por:

$$\chi_{obs}^2 = \sum_{j=1}^{\ell} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_{gl}^2. \text{ Aqui, } gl = \ell - 1.$$

Obs.: Este resultado é válido para n grande e para $E_i > 5$.

Teste Qui-Quadrado para Aderência

Exemplo 3: Em um celebre experimento de Mendel, foram polinizadas 15 plantas de sementes lisas e albume amarelo com plantas de semente rugosa e albume verde. As plantas resultante desse cruzamento tinham sementes lisas e alburnes amarelo (amarelo-lisa). Cruzando essas plantas entre si, Mendel obteve 556 sementes, distribuídas conforme a tabela abaixo.

Sementes	Frequência Observada (O_i)
Amarelo-lisas	315
Amarelo-lisas	101
Verde-lisas	108
Amarelo-rugosas	32
Total	556

Teste Qui-Quadrado para Aderência

A teoria proposta por Mendel estabelece que a segregação nesse caso deve ocorrer na seguinte proporção: $\frac{9}{16} : \frac{3}{16} : \frac{3}{16} : \frac{1}{16}$.

Pergunta de interesse: Será que os resultados observados no experimento por Mendel estão de acordo com a teoria por ele postulada? Essa pergunta pode ser respondida por meio das seguintes hipóteses:

H_0 : A segregação obedece a lei de Mendel

H_1 : A segregação não obedece a lei de Mendel.

Ou simplesmente

H_0 : $p_1 = 9/16$, $p_2 = 3/16$, $p_3 = 3/16$, $p_4 = 1/16$

H_1 : existe pelo menos um j tal que $p_j \neq p_{0j}$.

Teste Qui-Quadrado para Aderência

A tabela contendo as frequências esperadas é apresentada abaixo.

Sementes	Frequência Observada (O_i)	Frequência Esperada (E_i)
Amarelo-lisas	315	$556 \times 9/16 = 312,75$
Amarelo-lisas	101	$556 \times 3/16 = 104,25$
Verde-lisas	108	$556 \times 3/16 = 104,25$
Amarelo-rugosas	32	$556 \times 1/16 = 34,75$
Total	556	556

Nota: Olhando apenas para a diferença entre as duas frequências, podemos concluir que não há evidências suficientes para rejeitar a hipótese nula. Isto é, que a segregação obedece a lei de Mendel.

Teste Qui-Quadrado para Aderência

No entanto, essa conclusão é mais subjectiva. Para tomar uma decisão mais consistente vamos calcular a estatística de teste e depois decidir em rejeitar ou não a hipótese nula. Ou seja,

$$\chi_{obs}^2 = \sum_{j=1}^{\ell} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(315 - 314,75)^2}{314,75} + \dots + \frac{(32 - 34,75)^2}{34,75} \approx 0,47.$$

Observe que o número de graus de liberdade será $gl = \ell - 1 = 4 - 1 = 3$.

Assim, ao nível de 5% de significância, o valor crítico

$\chi_{vc}^2 = \chi_{(0,05;3)}^2 = 7,82$ (verifique!!!). Para terminar, como $\chi_{obs}^2 < \chi_{vc}^2$, concluímos que não há evidências suficientes para rejeitar a hipótese nula.

Teste Qui-Quadrado de Independência

- De acordo com Triola (2014), a palavra contingente, entre outros significados, refere-se a uma **dependência** de algum outro factor.
- Dessa forma, o termo tabela de contingência é utilizado para testar a independência entre as variáveis linha e coluna.
- Em um **teste de independência**, testa-se a hipótese nula de que, em uma tabela de contingência, as variáveis linha e coluna são independentes, ou seja, não existe relação de dependência entre as variáveis.

Assim, no exemplo anterior podemos estar interessados em verificar se a opinião do entrevistado, sobre esse polêmico assunto (colunas), é influenciada pelo sexo do entrevistador (linhas).

Teste Qui-Quadrado de Independência

Tabelas de Contingências: Também conhecidas como tabelas de frequência de dupla entrada, são compostas de contagens de dados categóricos de duas variáveis:

- Uma variável é utilizada para categorizar linhas
- A outra variável é utilizada para categorizar colunas.

Obs.: Não restrição quanto ao número de categorias em cada tabela de contingência. Isto é, a dimensão da tabela pode ser:

- 2×2 (caso mais simples e mais frequente): tabela com duas linhas e duas colunas.
- $2 \times c$: Tabela com 2 linhas e c colunas.
- $\ell \times 2$: Tabela com ℓ linhas e 2 colunas, ou
- $\ell \times c$: Tabela com ℓ linhas e c colunas.

Nesse caso, ℓ e c devem ser tais que $\ell, c > 2$.

Teste Qui-Quadrado de Independência

- Exemplo 1:** numa pesquisa para avaliar a opinião sobre o aborto, realizada pelo Instituto Eagleton, 1200 homens foram entrevistados, sendo 800 entrevistados por pessoas do sexo masculino e 400 por pessoas do sexo feminino. Desse total de entrevistados, 868 disseram concordar com a seguinte frase *“O aborto é um problema de natureza privada, cuja decisão deve ficar a critério da mulher, sem interferência do governo”*.
- Os resultados, de acordo com a resposta do entrevistado e sexo do entrevistador são resumidas na tabela a seguir:

Tabela 1

Sexo do entrevistador	Resposta	
	Concordo	Discordo
Feminino	308	92
Masculino	560	240

Teste Qui-Quadrado de Independência

Para realizar esse teste de independência, vamos estabelecer as estruturas desse teste de hipótese.

H_0 : As variáveis linha e coluna são independentes

H_1 : As variáveis linha e coluna são dependentes.

Desta forma, a formulação das hipóteses do [Exemplo 1](#), seriam:

H_0 : A opinião do entrevistado sobre o aborto é independente do sexo do entrevistador

H_1 : A opinião do entrevistado sobre o aborto é dependente do sexo do entrevistador

Teste Qui-Quadrado de Independência

Tabela 1

Sexo do entrevistador	Resposta		Total
	Concordo	Discordo	
Feminino	308	92	400
Masculino	560	240	800
Total	868	332	1200

- Sob a hipótese nula de que as variáveis são independentes, qual seria a tabela de contingência que deveríamos esperar no caso da pesquisa da Tabela 1?
- Do total de 1200 homens entrevistados, 868 (72,3%) concordam com a afirmação.
- Portanto, se as variáveis são independentes, deveríamos esperar essa mesma porcentagem tanto entre os que foram entrevistados por pessoas do sexo feminino (400), como entre os que foram entrevistados por pessoas do sexo masculino (800).
- Logo, deveríamos esperar 578,6 ($0,723 \times 800$) homens que concordam com a afirmação e que foram entrevistados por homens.
- A mesma proporção deveria acontecer entre os 400 homens que foram entrevistados por mulheres, ou seja, deveríamos esperar 289,3 ($0,723 \times 400$) homens que concordam com a afirmação e que foram entrevistados por mulheres.

Teste Qui-Quadrado de Independência

- Assim, podemos montar uma tabela esperada, sob a hipótese de independência (hipótese nula), que é mostrada na Tabela 2:

Tabela 2 - Tabela Esperada sob a Hipótese de Independência

Sexo do entrevistador	Resposta		Total
	Concordo	Discordo	
Feminino	$(868/1200) \times 400 = 289,3$	$(332/1200) \times 400 = 110,7$	400
Masculino	$(868/1200) \times 800 = 578,6$	$(332/1200) \times 800 = 221,4$	800
Total	868	332	1200

$$\frac{868}{1200} \approx 72,3\%$$

$$\frac{332}{1200} \approx 27,7\%$$

- Podemos comparar a tabela esperada (Tabela 2) com a tabela observada (Tabela 1) e verificar a distância entre os valores.
 - Se a distância entre o observado e o esperado for “pequena”, então há evidência a favor da hipótese de independência entre as variáveis.
 - Já se a distância entre o observado e o esperado for “grande”, então há evidências suficientes para rejeitar a hipótese de independência entre as variáveis.

Teste Qui-Quadrado de Independência

A questão é: Como medir a distância (ou diferença) entre as frequências observadas e as esperadas sob a hipótese nula?

Teste Qui-Quadrado: É um teste que serve para medir a distância entre as frequências observadas e esperadas em uma tabela de contingência. A notação do teste é χ^2 (lê-se “qui-quadrado”).

Note que, a soma das diferenças “simples” entre as frequências observada e esperadas é nula. Isto é, $\sum (O - E) = 0$. Por essa razão, o teste χ^2 leva em conta a soma ponderada das diferenças quadráticas entre as frequências.

A frequência esperada para uma categoria é a frequência que ocorreria se os dados tivessem realmente a distribuição alegada. Não há requisitos de que as frequências observadas sejam de pelo menos 5.

Teste Qui-Quadrado de Independência

Legenda:

“*O*” representa a *frequência observada* de um resultado encontrada pela tabulação dos dados amostrais.

“*E*” representa a *frequência esperada* de um resultado encontrada supondo-se que a hipótese nula é verdadeira.

Requisitos do teste χ^2

Para aplicar um teste χ^2 é necessário que essas suposições sejam satisfeitas:

- 1 Os dados foram selecionados aleatoriamente.
- 2 Os dados amostrais consistem em contagens para cada uma das diferentes categorias.
- 3 Para cada categoria, as frequência *esperada* é pelo menos 5.

Teste Qui-Quadrado de Independência

Obs.: Quando o requisito do item (3) não é satisfeito, aconselha-se agrupar todas as categorias com E'_i s menores que 5.

As tabelas de contingência para as frequências observadas e esperadas são:

Linhas	Colunas		Total
	Coluna 1	Coluna 2	
Linha 1	O_{11}	O_{12}	$O_{1.}$
Linha 2	O_{21}	O_{22}	$O_{2.}$
Total	$O_{.1}$	$O_{.2}$	$O_{..}$

Tabela: Tabela de contingência para frequências observada.

Teste Qui-Quadrado de Independência

Linhas	Colunas		Total
	Coluna 1	Coluna 2	
Linha 1	E_{11}	E_{12}	$\mathbf{E}_{1.}$
Linha 2	E_{21}	E_{22}	$\mathbf{E}_{2.}$
Total	$\mathbf{E}_{.1}$	$\mathbf{E}_{.2}$	$\mathbf{E}_{..}$

Tabela: Tabela de contingência para frequências esperada.

As frequências esperadas para cada célula (i, j) são obtidas multiplicando os totais da linhas com a da coluna e depois dividido pelo total global $\mathbf{O}_{..}$. Isto é,

$$E_{ij} = \frac{\mathbf{O}_{i.} \times \mathbf{O}_{.j}}{\mathbf{O}_{..}}$$

Teste Qui-Quadrado de Independência

Assim,

$$\begin{aligned} E_{11} &= \frac{\mathbf{O}_{1.} \times \mathbf{O}_{.1}}{\mathbf{O}_{..}}; & E_{12} &= \frac{\mathbf{O}_{1.} \times \mathbf{O}_{.2}}{\mathbf{O}_{..}} \\ E_{21} &= \frac{\mathbf{O}_{2.} \times \mathbf{O}_{.1}}{\mathbf{O}_{..}}; & E_{22} &= \frac{\mathbf{O}_{2.} \times \mathbf{O}_{.2}}{\mathbf{O}_{..}} \end{aligned}$$

Observe que, $\mathbf{O}_{..}$ é nada mais do que o número total de observações na amostra n . Isto é, o tamanho de amostra, dado por:

$$\mathbf{O}_{..} = \sum_{i=1}^{\ell} \sum_{j=1}^c O_{ij} = \sum_{i=1}^{\ell} \sum_{j=1}^c E_{ij} = \mathbf{E}_{..}$$

Teste Qui-Quadrado de Independência

O_{11} : representa a frequência observada na linha 1 e coluna 1.

O_{12} : representa a frequência observada na linha 1 e coluna 2.

O_{21} : representa a frequência observada na linha 2 e coluna 1.

O_{22} : representa a frequência observada na linha 2 e coluna 2.

E_{11} : representa a frequência esperada na linha 1 e coluna 1.

E_{12} : representa a frequência esperada na linha 1 e coluna 2.

E_{21} : representa a frequência esperada na linha 2 e coluna 1.

E_{22} : representa a frequência esperada na linha 2 e coluna 2.

Teste Qui-Quadrado de Independência

Uma forma mais simples e elegante de construir uma tabela de contingência, consiste em colocar na mesma tabela os dois tipos de frequência. O que facilita o cálculo das diferenças na hora de computar a estatística de teste.

Linhas	Colunas		Total
	Coluna 1	Coluna 2	
Linha 1	$O_{11} (E_{11})$	$O_{12} (E_{12})$	$O_{1.}$
Linha 2	$O_{21} (E_{21})$	$O_{22} (E_{22})$	$O_{2.}$
Total	$O_{.1}$	$O_{.2}$	$O_{..}$

Teste Qui-Quadrado de Independência

Desta forma, a estatística de teste tem a seguinte forma:

$$\chi_{obs}^2 = \sum_{i=1}^{\ell} \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \sim \chi_{gl}^2.$$

Sendo χ_{gl}^2 a distribuição qui-quadrado com gl graus de liberdade, dado por $gl = (\ell - 1) \times (c - 1)$. Sendo ℓ o número de linhas e c o número de colunas em uma tabela de contingência.

- Quanto mais próximo estiverem as frequências esperadas das observadas (menor distância entre elas), menor é chance de rejeitar a hipótese nula.
- De igual forma, grandes valores da estatística de teste χ_{obs}^2 revelam diferenças significativas entre as frequências observadas e esperadas.

Teste Qui-Quadrado de Independência

Em uma tabela de dupla entrada para **frequências observadas** (Tabela do tipo 2×2) tal como apresentado abaixo,

Linhas	Colunas		Total
	Coluna 1	Coluna 2	
Linha 1	a	b	$a + b$
Linha 2	c	d	$c + d$
Total	$a + c$	$b + d$	n

É possível (e relativamente fácil) mostrar que o valor do χ^2 obtido usando a expressão anterior, pode ser calculado de maneira fácil e algebricamente semelhante.

$$\chi_{obs}^2 = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)}.$$

Teste Qui-Quadrado de Independência

Continuação do exemplo referente a opinião de 1200 homens em relação ao aborto.

Tabela 1 - Observada

Sexo do entrevistador	Resposta		Total
	Concordo	Discordo	
Feminino	308	92	400
Masculino	560	240	800
Total	868	332	1200

Tabela 2 - Esperada

Sexo do entrevistador	Resposta		Total
	Concordo	Discordo	
Feminino	289,3	110,7	400
Masculino	578,6	221,4	800
Total	868	332	1200

- Estatística de Teste:

$$\chi^2_{obs} = \frac{(o_{11} - e_{11})^2}{e_{11}} + \frac{(o_{12} - e_{12})^2}{e_{12}} + \frac{(o_{21} - e_{21})^2}{e_{21}} + \frac{(o_{22} - e_{22})^2}{e_{22}}$$

$$\chi^2_{obs} = \frac{(308 - 289,3)^2}{289,3} + \frac{(92 - 110,7)^2}{110,7} + \frac{(560 - 578,6)^2}{578,6} + \frac{(240 - 221,4)^2}{221,4}$$

$$\chi^2_{obs} = 6,53$$

Teste Qui-Quadrado de Independência

- A estatística calculada χ^2_{obs} pode ser considerada uma diferença grande, sob a hipótese de que as variáveis são independentes (hipótese nula)?
- Para decidirmos se 6,53 é um valor grande, vamos recorrer à **Distribuição de Probabilidade** de χ^2 , sob a suposição de que a hipótese nula é verdadeira.
- Grandes valores da estatística de teste formam a Região de Rejeição (RR) da hipótese nula. Esses valores encontram-se na região extrema direta da distribuição qui-quadrado.

Teste Qui-Quadrado de Independência

- Tal como foi mencionado anteriormente, a distribuição qui-quadrado também depende dos graus de liberdade, assim como a distribuição t-Student apresentada nos tópicos anteriores. Assim, a forma de determinar (ou consultar) os valores críticos na distribuição qui-quadrado, é similar com aquela usada para a distribuição T.
- Para o presente exemplo, temos $\ell = 2$ (duas linhas) e $c = 2$ (duas colunas). Desta forma, o número de graus de liberdade será:
$$g/l = (\ell - 1) \times (c - 1) = (2 - 1) \times (2 - 1) = 1.$$
- Escolhendo os níveis de significância $\alpha = 0,05$ e $\alpha = 0,01$ os valores críticos χ^2_{vc} 's são obtidos por meio da interseção entre a linha correspondente a $g/l = 1$ grau de liberdade e área a direita igual a 0,05 e 0,01. Fornecendo $\chi^2_{(1;0,05)} = 3,841$ e $\chi^2_{(1;0,01)} = 6,635$.

Teste Qui-Quadrado de Independência

A regra de decisão segue a mesma abordagem apresentada nos testes paramétricos. Isto é, rejeitamos a hipótese nula à favor da hipótese alternativa se $\chi_{obs}^2 > \chi_{vc}^2$. Ou se $\text{valor-p} < \alpha$, onde:

$$\text{valor-p} = P(\chi_{gl}^2 > \chi_{obs}^2).$$

Conclusão: A hipótese nula será rejeitada apenas quando for considerado o nível de significância de 5% já que nesse caso $\chi_{obs}^2 = 6,53 > \chi_{vc} = 3,841$. No caso de $\alpha = 1\%$, concluímos que há evidências suficientes para afirmar que a opinião do entrevistado sobre o aborto é independente do sexo do entrevistador.

Usando a tabela de distribuição qui-quadrado, o valor-p é dado por:

$$\text{valor-p} = P(\chi_{gl=1}^2 > 6,53) = 0,01061 < 0,05.$$

Teste Qui-Quadrado de Independência

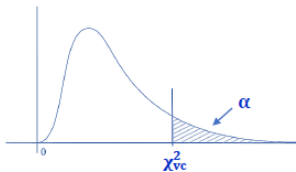


Tabela 2 × 2

Graus de liberdade = $(2 - 1) \times (2 - 1) = 1$

$\alpha = 5\%$

$\alpha = 1\%$

$$\chi^2_{vc} = 3,841$$

$$\chi^2_{vc} = 6,635$$

Área à Direita do Valor Crítico										
Graus de Liberdade	99.5% 0.995	99% 0.99	97.5% 0.975	95% 0.95	90% 0.90	10% 0.10	5% 0.05	2.5% 0.025	1% 0.01	0.5% 0.005
1	-	-	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.835	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Teste Qui-Quadrado de Independência

- Hipóteses nula e alternativa

$$\begin{cases} H_0: \text{a opinião do entrevistado sobre o aborto é independente do sexo do entrevistador} \\ H_1: \text{a opinião do entrevistado sobre o aborto é dependente do sexo do entrevistador} \end{cases}$$

- Estatística de teste

$$\chi^2 = \frac{(308 - 289,3)^2}{289,3} + \frac{(92 - 110,7)^2}{110,7} + \frac{(560 - 578,6)^2}{578,6} + \frac{(240 - 221,4)^2}{221,4} = 6,53$$

- Nível de significância e Região crítica

Graus de liberdade = $(2 - 1) \times (2 - 1) = 1$

$\alpha = 5\%$

$\chi^2_{vc} = 3,841$



- **Conclusão:** Existe evidência estatística, ao nível de 5% de significância, a favor da hipótese de que a opinião do entrevistado sobre o aborto é dependente do sexo do entrevistador.