ESTATÍSTICA

Michelle Hanne Soares de Andrade
michellehanne@cefetmg.br
1º. SEMESTRE 2018

Tabela de Contingência

Tabela de Contingência

Uma tabela de contingência (ou tabela de frequência de dupla entrada) é uma tabela em que as frequências correspondem a duas variáveis. (Uma variável categoriza as linhas, a outra categoriza as colunas)

As tabelas de contingências são de grande importância pois são utilizadas para analisar resultados de pesquisas.

Tabela de Contingência

- **Existem duas maneiras de construir uma tabela de contingência 2 X 2**: 1. Considerar uma amostra e classificá-la simultaneamente segundo duas variáveis. Esse é o caso da abaixo:

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

Tabela de Contingência

- Um total de 1200 homens foram classificados segundo sua resposta (concordo/discordo) e segundo o sexo da pessoa que os entrevistaram (masculino/feminino). **O objetivo do estudo que gerou a Tabela de Contingência é saber se a opinião do entrevistado sobre algo tão polêmico é influenciada pelo sexo do entrevistador.** Essa pergunta pode ser feita de outra maneira: a opinião do entrevistado é independente do sexo da pessoa que o entrevista?

Se há ou não independência entre as duas variáveis de classificação.

Tabela de Contingência

- 2- Considerar dois grupos distintos e então classificá-los segundo uma outra variável. Esse é o caso da Tabela abaixo:

	Hiperatividade		
Droga	*Sim*	*Não*	Total
A	152	48	200
B	132	68	200
Total	284	116	400

Tabela de Contingência

- Num estudo para verificar a eficácia de duas marcas de remédio no controle de hiperatividade, 400 hiperativos foram divididos aleatoriamente em 2 grupos de mesmo tamanho. Um grupo (200 pessoas) tomou a droga A e o outro grupo tomou a droga B. Depois de algum tempo de tratamento, verificou-se, em cada um dos grupos, quantas pessoas ainda estavam com sintomas de hiperatividade, gerando os dados da Tabela acima.

Tabela de Contingência

- Dessa tabela, podemos extrair da informação de que 48 das 200 pessoas que tomaram a droga A (24%) tiveram os sintomas de hiperatividade controlados, enquanto 68 das 200 pessoas que tomaram a droga B (34%) tiveram os sintomas de hiperatividade controlados. **O objetivo do estudo é saber se qual das drogas, A ou B, é a mais eficaz para o controle da hiperatividade.** Essa pergunta pode ser feita de outra maneira: a proporção de pessoas que tiveram a hiperatividade controlada com a droga A é igual à proporção de pessoas que tiveram a hiperatividade controlada com a droga B? Ou ainda podemos perguntar: as proporções são homogêneas?

se as proporções de uma classificação são homogêneas (iguais) ou não em dois grupos.

Tabela de Contingência

- Para isso, usaremos as técnicas de **Teste de Hipóteses**.
- Na situação 1, queremos **testar a hipótese de independência entre as duas variáveis** e, na situação 2, **queremos testar a hipótese de homogeneidade de duas proporções.**

Teste de Independência

- Como em todo teste de hipóteses, devemos estabelecer primeiramente as hipóteses nula e alternativa. Num teste de independência de variáveis usando uma tabela de contingência, essa hipóteses são:

H_0: As variáveis de classificação são independentes;
H_1: As variáveis de classificação não são independentes;

Podemos escrever as hipóteses também da seguinte maneira:

H_0: Não existe associação entre as variáveis de classificação;
H_1: Existe associação entre as variáveis de classificação;

Teste de Independência

Estatística de teste

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

Valores Críticos:

3. Na tabela A – 4 encontram-se os valores críticos, tomando-se graus de liberdade

$$gl = (r-1)(c-1)$$

onde:

r: número de linhas

c: número de colunas.

4. Os testes de hipótese de independência com tabelas de contingência envolvem apenas regiões críticas unilaterais à direita.

Teste de Independência

Freqüência esperada para uma tabela de contingência

$$E = \frac{(total \quad linhas)(total \quad colunas)}{total \quad geral}$$

Teste de Independência – Exemplo 1

- (Livro Estatística Aplicada à Gestão Empresarial – Adriano L. Bruni) Os dados a seguir referem-se ao cruzamento entre as variáveis: possui habilitação e sexo, de 53 funcionários de um escritório de contabilidade.

Tabela: Sexo versus Habilitação

	Habilitado		
	Sim	Não	Total
Feminino	9	12	21
Masculino	25	7	32
Total	34	19	53

Teste de Independência – Exemplo 1

O teste a ser feito é:

$$\begin{cases} H_0: as \quad variáveis \quad são \quad independentes \\ H_1: as \quad variáveis \quad são \quad dependentes \end{cases}$$

Cálculo das freqüências esperadas.

	Habilitado	
	Sim	Não
Feminino	$\frac{21 \cdot 34}{53} = 13,47$	$\frac{21 \cdot 19}{53} = 7,53$
Masculino	$\frac{32 \cdot 34}{53} = 20,53$	$\frac{32 \cdot 19}{53} = 11,47$

Teste de Independência – Exemplo 1

Cálculo da estatística de teste:

	Habilitado	
	Sim	Não
Feminino	$\frac{(9-13{,}47)^2}{13{,}47}=1{,}483$	$\frac{(12-7{,}53)^2}{7{,}53}=2{,}654$
Masculino	$\frac{(25-20{,}53)^2}{20{,}53}=0{,}973$	$\frac{(7-11{,}47)^2}{11{,}47}=1{,}742$

Assim

$$\chi^2=\sum\frac{(O-E)^2}{E}=1{,}483+2{,}654+0{,}973+1{,}742=6{,}852$$

Graus de liberdade: $gl=(r-1)(c-1)=(2-1)(2-1)=1$

Valor crítico: $\chi^2=3{,}841$

Como a estatística de teste > valor crítico, 6,852 > 3,841, então rejeitamos H_0.

As variáveis são dependentes

Teste de Independência – Exemplo 2

Exemplo 4 Um supermercado quer testar ao nível de significância de 5% a hipótese de que o modo de pagamento dos clientes nesse estabelecimento é independente do período do dia em que fazem as compras. Existem três modos de efectuar os pagamentos: por cheque, dinheiro e cartão de crédito.

A seguinte tabela de contingência 3×3 apresenta os resultados obtidos numa amostra de 4000 clientes:

	PERÍODO DO DIA		
MODO DE PAGAMENTO	**Manhã**	**Tarde**	**Noite**
Cheque	750	1500	750
Dinheiro	125	300	75
Cartão de Crédito	125	200	175

Teste de Independência – Exemplo 2

Denotando por A o atributo **Modo de pagamento** e por B o atributo **Período do dia em que faz as compras**, as hipóteses as testar são

H_0: A e B são independentes

H_1: A e B não são independentes

Uma vez que A e B assumem cada uma 3 modalidades, sob H_0, a estatística teste tem distribuição assintótica do Qui-quadrado com $(r-1)(s-1) = (3-1)(3-1) = 4$ graus de liberdade.

Ao nível de significância de 0.05, a região crítica é então $[9.49, +\infty[$

Teste de Independência – Exemplo 2

Como vimos, para obtermos o valor observado da estatística teste, temos de calcular as frequências esperadas:

$$\hat{e}_{ij} = n\hat{p}_{i\cdot}\hat{p}_{\cdot j} = n\frac{o_{i\cdot}}{n}\frac{o_{\cdot j}}{n} = \frac{o_{i\cdot}o_{\cdot j}}{n}$$

Assim, por exemplo,

$$\hat{e}_{11} = (3000 \times 1000)/4000 = 750$$

$$\hat{e}_{12} = (3000 \times 2000)/4000 = 1500$$

$$\hat{e}_{13} = (3000 \times 1000)/4000 = 750$$

Teste de Independência – Exemplo 2

Frequências esperadas

	PERÍODO DO DIA			
MODO DE PAGAMENTO	Manhã	Tarde	Noite	**Totais**
Cheque	750	1500	750	3000
Dinheiro	125	250	125	500
Cartão de Crédito	125	250	125	500
Totais	1000	2000	1000	4000

Valor observado da estatística teste:

$$\chi^2_{obs} = \frac{(750-750)^2}{750} + \frac{(1500-1500)^2}{1500} + \ldots + \frac{(200-250)^2}{250} + \frac{(175-125)^2}{125} = 60$$

Teste de Independência – Exemplo 2

Uma vez que 60 excede o valor crítico 9.49, ao nível de significância de 0.05, rejeitamos a hipótese de que o modo de pagamento é independente do período do dia em que as compras são feitas.

Medidas de Associação

- No teste do Qui-Quadrado apresentado, se for rejeitada a hipótese de independência entre os atributos, pode interessar medir a intensidade da associação entre os mesmos, através de uma medida adequada.
- **Uma vez que a estatística do teste mede o afastamento em relação à hipótese de independência, o seu valor observado também poderá servir para avaliar a força da relação entre os atributos.**

Coeficiente de Contingência Modificado

- O Coeficiente de Contingência Modificado permite quantificar a associação (grau de dependência) entre duas variáveis QUALITATIVAS, a partir da estatística χ^2 vista anteriormente. Sua equação:

$$\boldsymbol{C} = \sqrt{\frac{\boldsymbol{\chi^2}}{\boldsymbol{\chi^2 + N}}} \times \sqrt{\frac{\boldsymbol{K}}{\boldsymbol{K - 1}}}$$

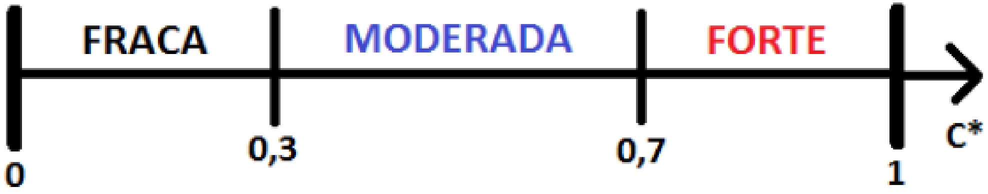
Coeficiente de Contingência Modificado

$$C = \sqrt{\frac{\chi^2}{\chi^2+N}} \times \sqrt{\frac{K}{K-1}}$$

Onde:

- $\chi 2$ é a estatística Qui-Quadrado, calculada a partir das frequências observadas e esperadas (sob a condição de independência) a partir da tabela de contingências.
- N é o número total de observações da tabela de contingências.
- k é o menor número entre o número de linhas e colunas da tabela de contingências.

O Coeficiente de Contingência Modificado varia de zero (completa independência) até 1 (associação perfeita).



Coeficiente de Contingência Modificado

- Valores de C* entre 0 e 0,29 indicam associação fraca entre as variáveis, as frequências dos valores de uma das variáveis aparentemente não são influenciadas pelos valores da outra. Valores de C* entre 0,3 e 0,69 indicam uma associação moderada. Valores acima de C* acima de 0,7 indicam uma associação forte entre as variáveis, há evidência que as frequências dos valores de uma das variáveis foi influenciada pelos valores da outra.

Coeficiente de Contingência Modificado

- Exemplo 2:
- $C = \sqrt{\frac{\chi^2}{\chi^2+N}} \times \sqrt{\frac{K}{K-1}} = \sqrt{\frac{60}{60+400}} \times \sqrt{\frac{3}{3-1}}$ ***=0,4423***
- χ^2 ***= 60***
- ***N= 400***
- ***K=3***

Associação moderada entre as variáveis