# TESTE DE AJUSTE, INDEPENDÊNCIA E ANOVA

Aulas 16, 17 e 18

# Objetivos

- Testes de ajuste
- Independência
- Análise de variância (ANOVA)

# TESTE DE AJUSTE

Aula 16

# Objetivos da Aula

- Tratar dados qualitativos (categorias ou atributos)
  - Frequência por categoria
- Testar se uma distribuição de frequência se encaixa numa distribuição afirmada
  - Distribuição qui-quadrado

# Experimentos multinomiais

- Há vários resultados possíveis (mais que dois)
- Os resultados são classificados em **categorias**
- Contar as frequencias observadas e esperadas
- Arranjar em linhas e colunas

- Uma estação de rádio afirma que a distribuição das preferências musicais dos ouvintes na região é a demonstrada abaixo.

| Distribuição das preferências musicais | | | |
|---|---|---|---|
| Clássica | 4% | Antigas | 2% |
| Country | 36% | Pop | 18% |
| Gospel | 11% | Rock | 29% |

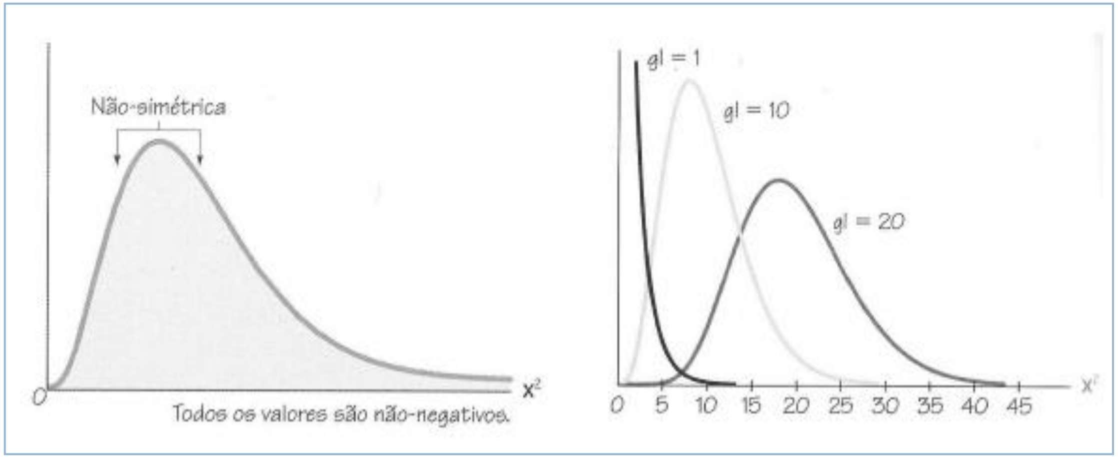
Resultado é classificado em **categorias.**

A probabilidade para cada categoria é fixa.

# Distribuição Qui-Quadrado



Mede a discrepância entre as frequências observadas e esperadas




**Propriedades da Distribuição**

Não é simétrica

Valor sempre maior ou igual a zero

A curva é diferente para diferentes graus de liberdade

# Teste de ajuste qui-quadrado

- Usado para testar se a distribuição de frequência se encaixa na distribuição esperada.
  - Hipótese Nula: distribuição de frequência se encaixa na distribuição especificada.
  - Hipótese Alternativa: distribuição de frequência não se encaixa na distribuição especificada.

# Teste de ajuste qui-quadrado

- Para testar a afirmação da estação de rádio, o executivo pode realizar um teste de ajuste qui-quadrado usando a seguinte hipótese:

$H_0$: A distribuição das preferências musicais na região é 4% clássica, 36% country, 11% gospel, 2% antigas, 18% pop, e 29% rock (afirmação)

$H_a$: A distribuição das preferências musicais difere da distribuição afirmada ou esperada

# Teste de ajuste qui-quadrado

- Para calcular o teste estatístico para o teste de ajuste qui-quadrado, utiliza as frequências observada e esperada.
- A **frequência observada O** de uma categoria é a frequência da categoria observada nos dados amostrais.

- A **frequência esperada *E*** de uma categoria é a frequência *calculada* para a categoria.
  - Obtidas assumindo a distribuição especificada (ou hipotética).
  - A frequência esperada para a *i*° categoria é

$$E_i = np_i$$

onde $n$ é o número de tentativas (tamanho da amostra) e $p_i$ é a probabilidade assumida da *i-ésima* categoria

# Exemplo: encontrando frequências observadas e esperadas

Um executivo de marketing seleciona aleatoriamente 500 ouvintes de rádio de uma região de transmissão e pergunta a cada um se eles preferem música clássica, country, gospel, antiga, pop ou rock. Os resultados podem ser observados na tabela ao lado. Encontre as frequências observadas e as frequências esperadas para cada tipo de música.

| Resultado da pesquisa (n = 500) | |
|---|---|
| Clássica | 8 |
| Country | 210 |
| Gospel | 72 |
| Antigas | 10 |
| Pop | 75 |
| Rock | 125 |

# Solução: encontrando frequências observadas e esperadas

**Frequência observada:** O número de ouvintes da rádio escolhendo um tipo de música em particular.

| Resultados da pesquisa (n = 500) | |
|---|---|
| Clássica | 8 |
| Country | 210 |
| Gospel | 72 |
| Antigas | 10 |
| Pop | 75 |
| Rock | 125 |

Frequência observada

## Frequência esperada: $E_i = np_i$

| Tipo de música | % de ouvintes | Frequência observada | Frequência esperada |
|---|---|---|---|
| Clássica | 4% | 8 | $500(0.04) = 20$ |
| Country | 36% | 210 | $500(0.36) = 180$ |
| Gospel | 11% | 72 | $500(0.11) = 55$ |
| Antigas | 2% | 10 | $500(0.02) = 10$ |
| Pop | 18% | 75 | $500(0.18) = 90$ |
| Rock | 29% | 125 | $500(0.29) = 145$ |

$n = 500$

# Teste de ajuste qui-quadrado

Para usar o teste de ajuste qui-quadrado, é preciso que:

1. As frequências observadas sejam obtidas usando uma amostra aleatória.
2. Cada frequência esperada precisa ser maior que ou igual à 5.

- Se essas condições são satisfeitas, então a distribuição de amostragem para o teste de ajuste qui-quadrado é aproximado por uma distribuição qui-quadrado com k − 1 graus de liberdade, onde k é o número de categorias.
- O teste estatístico para o teste de ajuste qui-quadrado é

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

O teste é sempre um teste unicaudal à direita

Onde O representa a frequência observada de cada categoria e *E* representa a frequência esperada de cada categoria.

| *Em palavras* | *Em símbolos* |
|---|---|
| 1. Identifique a afirmação. Expresse as hipóteses nula e alternativa. | Declare $H_0$ e $H_a$. |
| 2. Especifique o nível de significância. | Identifique $\alpha$. |
| 3. Identifique os graus de liberdade. | g.l. $= k - 1$ |
| 4. Determine o valor crítico. | Use a tabela |

| *Em palavras* | *Em símbolos* |
| --- | --- |
| 5. Determine a região de rejeição. | |
| 6. Calcule o teste estatístico. | $\chi^2 = \sum \frac{(O-E)^2}{E}$ |
| 7. Tome a decisão de rejeitar ou falhar em rejeitar a hipótese nula. | Se $\chi^2$ está na região de rejeição, rejeitar $H_0$. Se não, não rejeitar $H_0$. |
| 8. Interprete a decisão no contexto da afirmação original. | |

# Exemplo: realizando um teste de ajuste

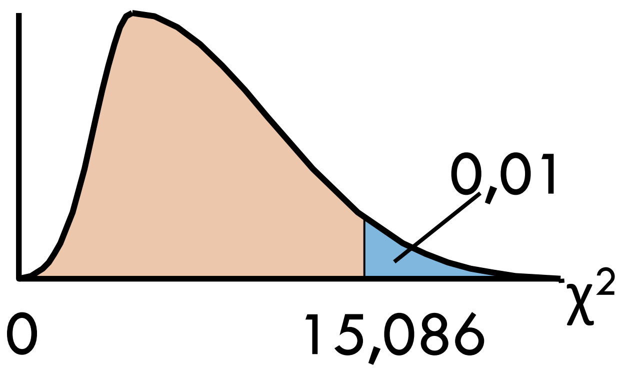
Use os dados das preferências musicais para realizar um teste de ajuste qui-quadrado para testar se as distribuições são diferentes. Use $\alpha = 0{,}01$.

| Distribuição das preferências musicais | |
|---|---|
| Clássica | 4% |
| Country | 36% |
| Gospel | 11% |
| Antigas | 2% |
| Pop | 18% |
| Rock | 29% |

| Resultados da pesquisa (n = 500) | |
|---|---|
| Clássica | 8 |
| Country | 210 |
| Gospel | 72 |
| Antigas | 10 |
| Pop | 75 |
| Rock | 125 |

# Solução: realizando um teste de ajuste

- $H_0$: A preferência musical é 4% clássica, 36% country, 11% gospel, 2% antigas, 18% pop, e 29% rock.
- $H_a$: A preferência musical difere da distribuição afirmada ou esperada.
- $\alpha = 0{,}01$
- g.l. = 6 – 1 = 5




- Teste estatístico:
- Decisão:
- Conclusão:

# Solução: realizando um teste de ajuste

| Tipo de música | Frequência observada | Frequência esperada |
|---|---|---|
| Clássica | 8 | 20 |
| Country | 210 | 180 |
| Gospel | 72 | 55 |
| Antigas | 10 | 10 |
| Pop | 75 | 90 |
| Rock | 125 | 145 |

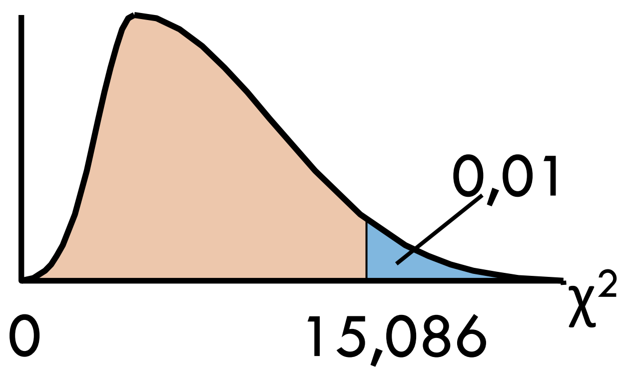
$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(8-20)^2}{20} + \frac{(210-180)^2}{180} + \frac{(72-55)^2}{55} + \frac{(10-10)^2}{10} + \frac{(75-90)^2}{90} + \frac{(125-145)^2}{145}$$

$$\approx 22.713$$

# Solução: realizando um teste de ajuste

- $H_0$: A preferência musical é 4% clássica, 36% country, 11% gospel, 2% antigas, 18% pop, e 29% rock.
- $H_a$: A preferência musical difere da distribuição afirmada ou esperada.
- $\alpha = 0{,}01$
- g.l. $= 6 - 1 = 5$




- Teste estatístico:

$$\chi^2 = 22.713$$

- Decisão: Rejeitar $H_0$

Há evidência suficiente para concluir que a distribuição das preferências musicais difere da distribuição afirmada.

# Exercício

O Quadro 1 descreve opiniões sobre qual é o motivo mais importante para se economizar dinheiro. Você trabalha em um empresa de serviços financeiros e quer testar a distribuição que descreve a opinião dos homens. Para testar a distribuição você seleciona aleatoriamente 400 homens e pergunta a cada um qual é o motivo mais importante para eles – economizar para a aposentadoria ou economizar para a educação superior dos seus filhos. Os resultados são exibidos no Quadro 2. Com $\alpha=0{,}05$, teste a distribuição afirmada ou esperada.

**Quadro 1**

| | Aposentadoria | Educação Superior | Não tem certeza |
|---|---|---|---|
| Homens | 44% | 40% | 16% |
| Mulheres | 41% | 46% | 13% |

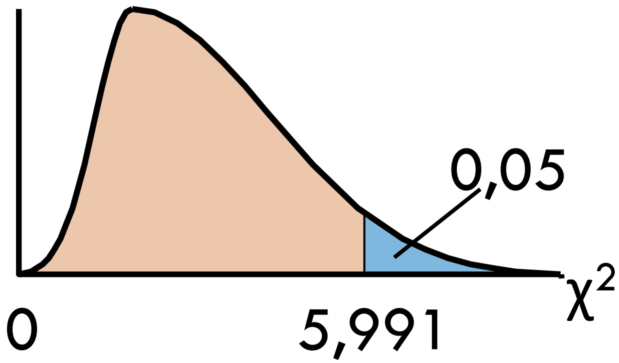
**Quadro 2**

| Resultado da Pesquisa entre homens (n=400) | |
|---|---|
| Aposentadoria | 186 |
| Educação Superior | 143 |
| Não tem certeza | 71 |

# Solução: realizando um teste de ajuste

- $H_0$: A distribuição das opiniões dos homens em relação à economizar para aposentadoria é 44%, para educação superior é 40% e 16% não sabem .
- $H_a$: A distribuição das opiniões dos homens diferem da distribuição afirmada ou esperada.

- $\alpha = 0,05$
- g.l. = 3 – 1 = 2




- Teste estatístico:
- Decisão:
- Conclusão:

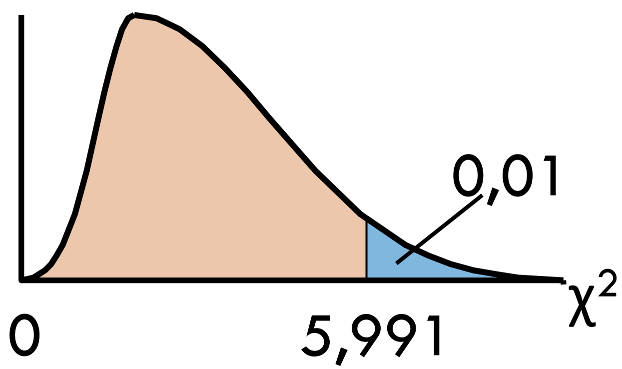
# Solução: realizando um teste de ajuste

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

| Respostas | Frequência observada | Frequência esperada |
|---|---|---|
| Aposentadoria | 186 | 400*0,44=176 |
| Educação Superior | 143 | 400*0,40=160 |
| Não sabe | 71 | 400*0,16=64 |

# Solução: realizando um teste de ajuste

- $\alpha =$ 0,05
- g.l. = 3 – 1 = 2




- Teste estatístico:

  $\chi^2 = 3{,}140$

- Decisão: Não Rejeitar $H_0$

Com 5% de significância, não há evidência suficiente para opor-se a distribuição esperada dos homens.

# Exercício: realizando um teste de ajuste

O fabricante de M&M's afirma que o número de unidades de cores diferentes em embalagens de M&M's de chocolate amargo é uniformemente distribuído. Para testar essa afirmação, você seleciona aleatoriamente uma embalagem que contenha 500 M&M's de chocolate amargo. Os resultados podem ser vistos na tabela. Usando $\alpha = 0{,}10$, faça um teste de ajuste qui-quadrado para testar as distribuições esperadas e afirmadas. O que você pode concluir? *(Adaptado de Mars Incorporated)*

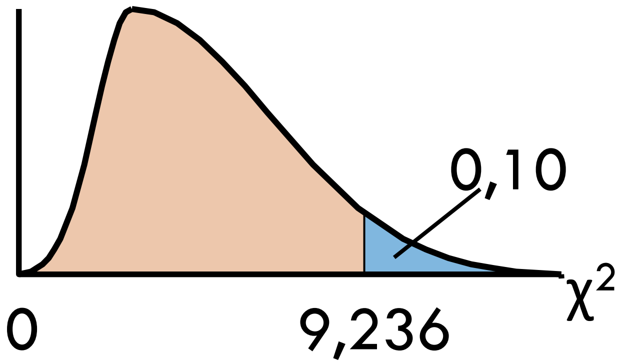
| Cor | Frequência |
|---|---|
| Marrom | 80 |
| Amarelo | 95 |
| Vermelho | 88 |
| Azul | 83 |
| Laranja | 76 |
| Verde | 78 |

$n = 500$

- A afirmação é que a distribuição é uniforme, então as frequências esperadas de cada cor são iguais.
- Para encontrar cada frequência esperada, divida o tamanho da amostra pelo número de cores.
- $\boldsymbol{E = 500/6 \approx 83{,}3}$

# Solução: realizando um teste de ajuste

- $H_0$: A distribuição de cores diferentes em embalagens de M&M's amargo é uniforme.
- $H_a$: A distribuição de cores diferentes em embalagens de M&M's amargo não é uniforme.
- $\alpha = 0{,}10$
- g.l. = 6 – 1 = 5




- Teste estatístico:
- Decisão:
- Conclusão:

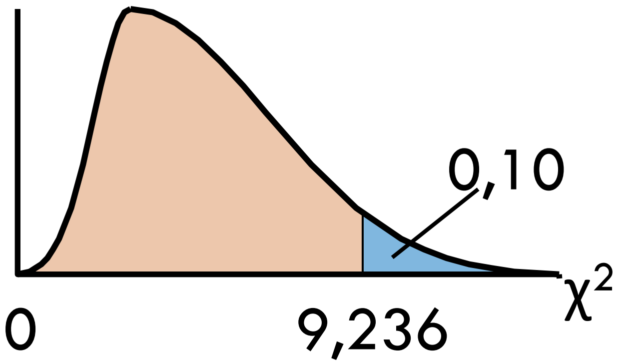
$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

| Cor | Frequência observada | Frequência esperada |
|---|---|---|
| Marrom | 80 | 83.3 |
| Amarelo | 95 | 83.3 |
| Vermelho | 88 | 83.3 |
| Azul | 83 | 83.3 |
| Laranja | 76 | 83.3 |
| Verde | 78 | 83.3 |

$$= \frac{(80-83.3)^2}{83.3} + \frac{(95-83.3)^2}{83.3} + \frac{(88-83.3)^2}{83.3} + \frac{(83-83.3)^2}{83.3} + \frac{(76-83.3)^2}{83.3} + \frac{(78-83.3)^2}{83.3}$$

$$\approx 3.016$$

- $H_0$: A distribuição de cores diferentes em embalagens de M&M's amargo é uniforme.
- $H_a$: A distribuição de cores diferentes em embalagens de M&M's amargo não é uniforme.
- $\alpha = 0{,}10$
- g.l. $= 6 - 1 = 5$




- Teste estatístico:

$$\chi^2 = 3{,}016$$

- Decisão: Não rejeitar $H_0$

Com um nível de 10% de significância, não há evidência suficiente para rejeitar que a distribuição de cores é uniforme.

# Objetivo da Aula

- Usar a distribuição qui-quadrado para testar se uma distribuição de frequência se encaixa numa distribuição afirmada.

# INDEPENDÊNCIA

Aula 17

# Objetivos

- Usar uma tabela de contingência para encontrar as frequências esperadas.
- Usar uma distribuição qui-quadrado para testar se duas variáveis são independentes.

# Tabela de contingência

**Tabela de contingência** ***r*** **×** ***c***

- Mostra as frequências observadas para duas variáveis.
- As frequências observadas são arranjadas nas fileiras *r* e nas colunas c.
- A interseção de uma fileira com uma coluna é chamada de **célula**.

# Tabela de contingência

- Mostra os resultados de uma amostra aleatória de 550 presidentes de empresas classificados por idade e tamanho da empresa(*Adaptado de Grant Thornton LLP, The Segal Company*)

| | Idade | | | | |
|---|---|---|---|---|---|
| Tamanho da empresa | 39 ou menos | 40 - 49 | 50 - 59 | 60 - 69 | 70 ou mais |
| Pequena / Média | 42 | 69 | 108 | 60 | 21 |
| Grande | 5 | 18 | 85 | 120 | 22 |

# Encontrando a frequência esperada

- Assumindo que as duas variáveis são independentes, você pode usar a tabela de contingência para encontrar a frequência esperada para cada célula.
- A frequência esperada para uma célula $E_{r,c}$ numa tabela de contingência é:

$$\text{Frequência esperada } E_{r,c} = \frac{(\text{Soma das fileiras } r)\times(\text{Soma das colunas } c)}{\text{Tamanho da amostra}}$$

# Exemplo: encontrando as frequências esperadas

Encontre a frequência esperada para cada célula na tabela de contingência. Assuma que as variáveis idade e tamanho da empresa são independentes.

| | Idade | | | | | |
|---|---|---|---|---|---|---|
| Tamanho da empresa | 39 ou menos | 40 - 49 | 50 - 59 | 60 - 69 | 70 ou mais | Total |
| Pequena / Média | 42 | 69 | 108 | 60 | 21 | 300 |
| Grande | 5 | 18 | 85 | 120 | 22 | 250 |
| Total | 47 | 87 | 193 | 180 | 43 | 550 |

Totais marginais

# Solução: encontrando as frequências esperadas

$$\text{Frequência esperada } E_{r,c} = \frac{(\text{Soma das fileiras } r) \times (\text{Soma das colunas } c)}{\text{Tamanho da amostra}}.$$

| | Idade | | | | | |
|---|---|---|---|---|---|---|
| Tamanho da empresa | 39 ou menos | 40 - 49 | 50 - 59 | 60 - 69 | 70 ou mais | Total |
| Pequena / Média | 42 | 69 | 108 | 60 | 21 | 300 |
| Grande | 5 | 18 | 85 | 120 | 22 | 250 |
| Total | 47 | 87 | 193 | 180 | 43 | 550 |

$$E_{1,1} = \frac{300 \cdot 47}{550} \approx 25.64$$

| | Idade | | | | | |
|---|---|---|---|---|---|---|
| Tamanho da empresa | 39 ou menos | 40 - 49 | 50 - 59 | 60 - 69 | 70 ou mais | Total |
| Pequena / Média | 42 | 69 | 108 | 60 | 21 | 300 |
| Grande | 5 | 18 | 85 | 120 | 22 | 250 |
| Total | 47 | 87 | 193 | 180 | 43 | 550 |

$$E_{1,2} = \frac{300 \cdot 87}{550} \approx 47.45$$

$$E_{1,3} = \frac{300 \cdot 193}{550} \approx 105.27$$

$$E_{1,4} = \frac{300 \cdot 180}{550} \approx 98.18$$

$$E_{1,5} = \frac{300 \cdot 43}{550} \approx 23.45$$

| | Idade | | | | | |
|---|---|---|---|---|---|---|
| Tamanho da empresa | 39 ou menos | 40 - 49 | 50 - 59 | 60 - 69 | 70 ou mais | Total |
| Pequena e Média | 42 | 69 | 108 | 60 | 21 | 300 |
| Grande | 5 | 18 | 85 | 120 | 22 | 250 |
| Total | 47 | 87 | 193 | 180 | 43 | 550 |

$$E_{2,1} = \frac{250 \cdot 47}{550} \approx 21.36 \qquad E_{2,2} = \frac{250 \cdot 87}{550} \approx 39.55 \qquad E_{2,3} = \frac{250 \cdot 193}{550} \approx 87.73$$

$$E_{2,4} = \frac{250 \cdot 180}{550} \approx 81.82 \qquad E_{2,5} = \frac{250 \cdot 43}{550} \approx 19.55$$

# Teste qui-quadrado para independência

- Usado para testar a independência de duas variáveis.
- Pode determinar se a ocorrência de uma variável afeta a probabilidade de ocorrência da outra variável.

# Teste qui-quadrado para independência

Para usar teste qui-quadrado para independência, preencher as seguintes condições:

1. Frequências observadas devem ser obtidas usando uma amostra aleatória.
2. Cada frequência esperada deve ser maior que ou igual a 5.

# Teste qui-quadrado para independência

- Distribuição de amostragem para o teste qui-quadrado para independência é aproximado pela distribuição qui-quadrada com: ($r$ − 1)(c − 1) graus de liberdade, onde $r$ e c são o número de fileiras e colunas, respectivamente, de uma tabela de contingência.
- O **teste estatístico** para o teste qui-quadrado para independência é:

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

O teste é sempre um teste unicaudal à direita

Onde O representa as frequências observadas e $E$ representa a frequência esperada.

# Teste qui-quadrado para independência

| *Em palavras* | *Em símbolos* |
| --- | --- |
| 1. Identifique a afirmação. Expresse as hipóteses nula e alternativa. | Expresse $H_0$ e $H_a$. |
| 2. Especifique o nível de significância. | Identifique $\alpha$. |
| 3. Identifique os graus de liberdade. | g.l. $= (r - 1)(c - 1)$ |
| 4. Determine o valor crítico. | Use a tabela 6 no apêndice B. |

# Teste qui-quadrado para independência

| *Em palavras* | *Em símbolos* |
| --- | --- |
| 5. Determine a região de rejeição. | |
| 6. Calcule o teste estatístico. | $\chi^2 = \sum \frac{(O-E)^2}{E}$ |
| 7. Tome a decisão de rejeitar ou falhar em rejeitar a hipótese nula. | Se $\chi^2$ está na região de rejeição, rejeite $H_0$. Se não, falhe em rejeitar $H_0$. |
| 8. Interprete a decisão no contexto da afirmação original. | |

# Exemplo: realizando um teste de independência $\chi^2$

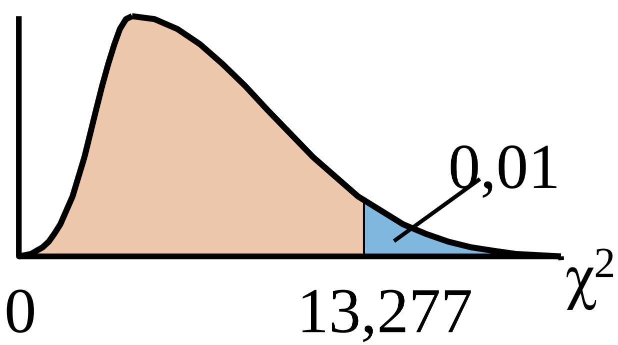
Usando a tabela de contingência de idade/empresas, você pode concluir que as idades dos presidentes das empresas têm relação com o tamanho da empresa? Use $\alpha = 0,01$. As frequências esperadas são mostradas entre parênteses.

| | Idade | | | | | |
|---|---|---|---|---|---|---|
| Tamanho da empresa | 39 ou menos | 40 - 49 | 50 - 59 | 60 - 69 | 70 ou mais | Total |
| Pequena/ Média | 42 (25,64) | 69 (47,45) | 108 (105,27) | 60 (98,18) | 21 (23,45) | 300 |
| Grande | 5 (21,36) | 18 (39,55) | 85 (87,73) | 120 (81,82) | 22 (19,55) | 250 |
| Total | 47 | 87 | 193 | 180 | 43 | 550 |

# Solução: realizando um teste de ajuste

- $H_0$: Idades dos presidentes são independentes dos tamanhos das empresas.
- $H_a$: Idades dos presidentes não são independentes dos tamanhos das empresas.

- $\alpha = 0,01$
- g.l. = $(2 – 1)(5 – 1) = 4$




- Teste estatístico:
- Decisão:

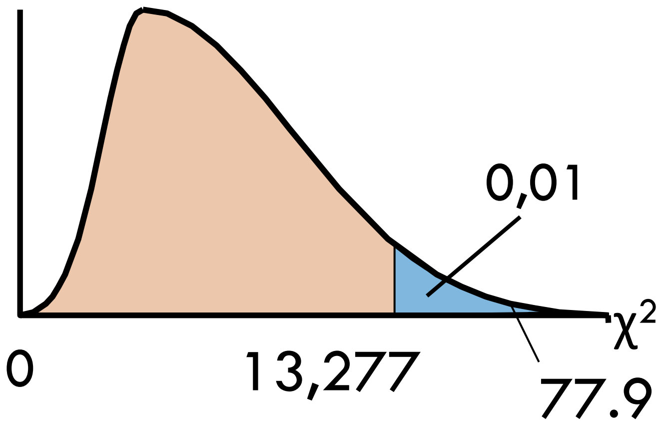
$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

$$= \frac{(42-25.64)^2}{25.64} + \frac{(69-47.45)^2}{47.45} + \frac{(108-105.27)^2}{105.27} + \frac{(60-98.18)^2}{98.18} + \frac{(21-23.45)^2}{23.45}$$

$$+ \frac{(5-21.36)^2}{21.36} + \frac{(18-39.55)^2}{39.55} + \frac{(85-87.73)^2}{87.73} + \frac{(120-81.82)^2}{81.82} + \frac{(22-19.55)^2}{19.55}$$

$$\approx 77.9$$

- $H_0$: Idades dos presidentes são independentes dos tamanhos das empresas.
- $H_a$: Idades dos presidentes não são independentes dos tamanhos das empresas.
- $\alpha = 0{,}01$
- g.l. $= (2 - 1)(5 - 1) = 4$




- Teste estatístico:

$$\chi^2 = 77.9$$

- Decisão: Rejeitar $H_0$

Existe evidência para concluir que as idades dos presidentes dependem dos tamanhos das empresas.

# Exercício



Transtorno obsessivo compulsivo (TOC): Os resultados de uma amostra aleatória de pacientes com TOC tratados com um remédio ou com um placebo podem ser observados na tabela de contingência. Com $\alpha$=0,10, você pode concluir que o tratamento está relacionado ao resultado? Você recomendaria usar o remédio como parte de um tratamento do transtorno obsessivo-compulsivo?

| | **Tratamento** | |
|---|---|---|
| Resultado | Remédio | Placebo |
| Melhoraram | 39 | 25 |
| Não apresentaram melhora | 54 | 70 |

# Exercício



A tabela abaixo mostra a distribuição dos horários do dia que mortes em rodovias aconteceram no ano 2010. Os resultados de um estudo recente sobre 627 mortes em rodovias selecionadas aleatoriamente podem ser visto na tabela 2. Com $\alpha=0{,}01$, a distribuição mudou?

| Resultado em 2010 | | Resultado do estudo | |
|---|---|---|---|
| Horário do dia | Porcentagem | Horário do dia | Frequência |
| Da meia-noite às 6 h da manha | 32% | Da meia-noite às 6 h da manha | 224 |
| Das 6 h ao meio-dia | 15% | Das 6 h ao meio-dia | 128 |
| Do meio dia às 18h | 24% | Do meio dia às 18h | 115 |
| Das 18h à meia-noite | 29% | Das 18h à meia-noite | 160 |

# Exercício



Você trabalha em um departamento de educação continuada de uma faculdade e quer determinar se as razões dadas por trabalhadores para continuar estudando está relacionada ao tipo de trabalho. No seu estudo, você coleta aleatoriamente os dados exibidos na tabela de contingência. Com $\alpha$=0,01, você pode concluir que a razão e o tipo de trabalho são dependentes? Como você usaria essa informação nos seus esforços de marketing?

| | Razão | | |
|---|---|---|---|
| Tipo de trabalho | Profissional | Pessoal | Profissional e pessoal |
| Técnico | 30 | 36 | 41 |
| Outros | 47 | 25 | 30 |

# Objetivos

- Usar uma tabela de contingência para encontrar as frequências esperadas.
- Usar uma distribuição qui-quadrado para testar se duas variáveis são independentes.

# ANOVA

Aula 18

# Objetivos

- Usar a análise de variância com um fator para testar afirmações envolvendo três ou mais médias.
- Introduzir análise de variância com dois fatores.

# Análise da variância com um fator (ANOVA unidirecional)

- Técnica de teste de hipótese usada para comparar médias de três ou mais populações.
- Hipóteses:
  - $H_0$: $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$ (todas as médias populacionais são iguais).
  - $H_a$: Pelo menos uma das médias é diferente das outras.

# Análise da variância com um fator (ANOVA unidirecional)

Em um teste ANOVA unidirecional, as seguintes condições devem ser preenchidas:

1. Cada amostra deve ser aleatoriamente selecionada de uma população normal ou aproximadamente normal.
2. As amostras devem ser independentes entre si.
3. Cada população deve ter a mesma variância.

# Análise da variância com um fator (ANOVA unidirecional)

$$\text{Teste estatístico} = \frac{\text{Variância entre amostras}}{\text{Variância dentro das amostras}}$$

1. A variância entre as amostras $MS_B$ mede a diferença relacionada ao tratamento dado à cada amostra

   Chamado de **média entre quadrados.**

2. A variância dentro das amostras $MS_W$ mede as diferenças relacionadas aos lançamentos dentro da mesma amostra e decorre do erro de amostragem.

   Chamado **média dos quadrados internos**

# Teste de análise de variância de um fator

- Se as condições para uma análise de variância de um fator são satisfeitas, então a distribuição amostral para o teste é aproximada pela distribuição *F*.
- O teste estatístico é:

$$\boxed{F = \frac{MS_B}{MS_W}}$$

- Os graus de liberdade para o teste *F* são:

  $\text{g.l.}_N = k - 1$ e $\text{g.l.}_D = N - k$.

- Onde *k* é o número de amostras e *N* é a soma dos tamanhos das amostras.

# Teste estatístico para teste ANOVA de um fator

| ***Em palavras*** | ***Em símbolos*** |
| --- | --- |
| 1. Encontre a média e a variância de cada amostra. | $\bar{x} = \frac{\sum x}{n} \quad s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$ |
| 2. Encontre a média de todas as entradas em todas as amostras (a grande média). | $\bar{\bar{x}} = \frac{\sum x}{N}$ |
| 3. Encontre a soma dos quadrados entre as amostras. | $SS_B = \sum n_i (\bar{x}_i - \bar{\bar{x}})^2$ |
| 4. Encontre a soma dos quadrados dentro das amostras. | $SS_W = \sum (n_i - 1) s_i^2$ |

# Análise da variância com um fator (ANOVA unidirecional)

| *Em palavras* | *Em símbolos* |
|---|---|
| 5. Encontre a variância entre as amostras. | $MS_B = \frac{SS_B}{k-1} = \frac{SS_B}{\text{d.f.}_\text{N}}$ |
| 6. Encontre a variância dentro das amostras. | $MS_W = \frac{SS_W}{N-k} = \frac{SS_W}{\text{d.f.}_\text{D}}$ |
| 7. Encontre o teste estatístico. | $F = \frac{MS_B}{MS_W}$ |

# Análise simples do teste de variância de um fator ANOVA

| *Em palavras* | *Em símbolos* |
|---|---|
| 1. Identifique a afirmação. Expresse as hipóteses nula e alternativa. | Expresse $H_0$ e $H_a$. |
| 2. Especifique o nível de significância. | Identifique $\alpha$. |
| 3. Identifique os graus de liberdade. | $\text{g.l.}_N = k - 1$ $\text{g.l.}_D = N - k$ |
| 4. Determine o valor crítico. | Use a tabela |

# Análise da variância com um fator (ANOVA unidirecional)

| *Em palavras* | *Em símbolos* |
|---|---|
| 5. Determine a região de rejeição. | |
| 6. Calcule o teste estatístico. | $F = \dfrac{MS_B}{MS_W}$ |
| 7. Tome a decisão de rejeitar ou falhar em rejeitar a hipótese nula. | Se $F$ está na região de rejeição, rejeitear $H_0$. Se não, não rejeitar $H_0$. |
| 8. Interprete a decisão no contexto da afirmação original. | |

# Tabela de resumo ANOVA

- Tabela para resumir os resultados de um teste com um fator ANOVA.

| Variação | Soma dos quadrados | Graus de liberdade | Médias quadradas | $F$ |
|---|---|---|---|---|
| Entre | $SS_B$ | $g.l._N$ | $MS_B = \frac{SS_B}{d.f._N}$ | $\frac{MS_B}{MS_W}$ |
| Interna | $SS_W$ | $g.l._D$ | $MS_W = \frac{SS_W}{d.f._D}$ | |

# Exemplo: teste ANOVA simples

Um médico pesquisador quer determinar se há uma diferença na média de tempo que três tipos de analgésicos levam para aliviar a dor de cabeça. Várias pessoas que sofrem com dores de cabeça são selecionadas aleatoriamente e tomam um dos três medicamentos. Cada pessoa diz o tempo (em minutos) que o medicamento começou a fazer efeito. Os resultados são mostrados no próximo slide. Com $\alpha = 0{,}01$, você pode concluir que as médias de tempo são diferentes? Suponha que cada população de tempo de alívio seja normalmente distribuída e que a população de variâncias seja igual.

| 1 | 2 | 3 |
|---|---|---|
| 12 | 16 | 14 |
| 15 | 14 | 17 |
| 17 | 21 | 20 |
| 12 | 15 | 15 |
|  | 19 |  |

| Medicação 1 | Medicação 2 | Medicação 3 |
|---|---|---|
| 12 | 16 | 14 |
| 15 | 14 | 17 |
| 17 | 21 | 20 |
| 12 | 15 | 15 |
| | 19 | |
| $\bar{x}_1 = \frac{56}{4} = 14$ | $\bar{x}_2 = \frac{85}{5} = 17$ | $\bar{x}_3 = \frac{66}{4} = 16.5$ |
| $s_1^2 = 6$ | $s_2^2 = 8.5$ | $s_3^2 = 7$ |

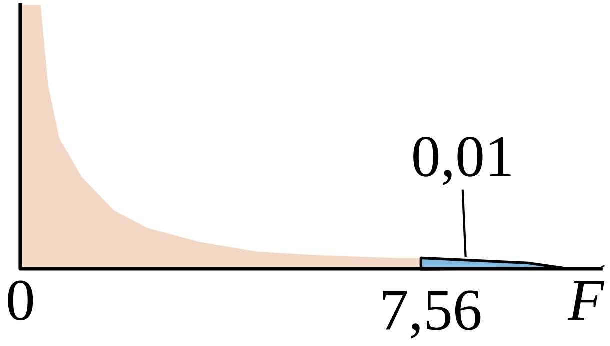
**Solução:**

$k = 3$ (3 amostras)

$N = n_1 + n_2 + n_3 = 4 + 5 + 4 = 13$ (soma dos tamanhos das amostras)

# Solução: teste ANOVA simples

- $H_0$: $\mu_1 = \mu_2 = \mu_3$
- $H_a$: Pelo menos uma média é diferente
- $\alpha = 0{,}01$
- $g.l._N = 3 - 1 = 2$
- $g.l._D = 13 - 3 = 10$
- Região de rejeição:




- Teste estatístico:
- Decisão:

Para encontrar o teste estatístico, o seguinte deve ser calculado:

$$\bar{\bar{x}} = \frac{\sum x}{N} = \frac{56+85+66}{13} \approx 15.92$$

$$MS_B = \frac{SS_B}{\mathrm{d.f.}_\mathrm{N}} = \frac{\sum n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1}$$

$$= \frac{4(14-15.92)^2 + 5(17-15.92)^2 + 4(16.5-15.92)^2}{3-1}$$

$$\approx \frac{21.92}{2} = 10.96$$

Para encontrar o teste estatístico, o seguinte deve ser calculado:

$$MS_W = \frac{SS_W}{\text{d.f.}_D} = \frac{\sum (n_i - 1)s_i^2}{N - k}$$

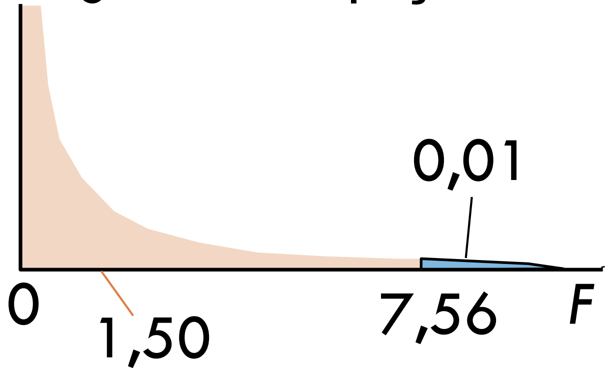
$$= \frac{(4-1)(6) + (5-1)(8.5) + (4-1)(7)}{13-3}$$

$$= \frac{73}{10} = 7.3$$

$$F = \frac{MS_B}{MS_W} = \frac{10.96}{7.3} \approx 1.50$$

## Teste estatístico:

- H0: $\mu_1 = \mu_2 = \mu_3$
- Ha: Pelo menos uma média é diferente
- $\alpha =$ 0,01
- g.l.$_N$= 3 – 1 = 2
- g.l.$_D$= 13 – 3 = 10
- Região de rejeição:




$$F = \frac{MS_B}{MS_W} \approx 1.50$$

- Decisão: Não rejeitar $H_0$

Não há evidência suficiente no nível de significância 1% para concluir que há uma diferença no tempo médio que os três analgésicos levam para aliviar as dores de cabeça.

# Exercício



Um analista de vendas quer determinar se há uma diferença na média de vendas mensais de uma empresa em quatro regiões diferentes. Vários vendedores de cada região são selecionados aleatoriamente e cada um fornece suas quantias de vendas (em milhares de reais) do mês anterior. Os resultados são mostrados na tabela. Com $\alpha$=0,05, o analista pode concluir que há uma diferença na média mensal de vendas entre as regiões de vendas? Suponha que cada população de vendas seja normalmente distribuída e que as variâncias populacionais são iguais.

| Norte | Leste | Sul | Oeste |
|---|---|---|---|
| 34 | 47 | 40 | 21 |
| 28 | 36 | 30 | 30 |
| 18 | 30 | 41 | 24 |
| 24 | 38 | 29 | 37 |
| | 44 | | 23 |
| $\bar{x}_1 = 26$ | $\bar{x}_2 = 39$ | $\bar{x}_3 = 35$ | $\bar{x}_4 = 27$ |
| $s_1^2 = 45{,}33$ | $s_2^2 = 45$ | $s_3^2 = 40{,}67$ | $s_4^2 = 42{,}5$ |

# ANOVA bidirecional

**Análise de variância bidirecional**

- Técnica de teste de hipótese usada para testar o efeito de duas variáveis independentes, ou **fatores,** em uma variável dependente.

**Exemplo:**

- Suponha que um médico pesquisador queira testar os efeitos do gênero e tipo de medicamente na média do período de tempo que analgésicos levam para aliviar a dor.

| Tipo de medicação | | Gênero: Homem | Gênero: Mulher |
|---|---|---|---|
| | I | Homens tomando tipo I | Mulheres tomando tipo I |
| | II | Homens tomando tipo II | Mulheres tomando tipo II |
| | III | Homens tomando tipo III | Mulheres tomando tipo III |

# Hipóteses de ANOVA bidirecional

**Efeito principal**

- O efeito de uma variável independente na variável dependente.

**Efeito de interação**

- O efeito das duas variáveis independentes na variável dependente.

**Hipóteses para efeitos principais:**

- $H_0$: O gênero não tem efeito na média do tempo que leva para um analgésico aliviar a dor.
- $H_a$: O gênero tem efeito na média do tempo que leva para um analgésico aliviar a dor.
- $H_0$: O tipo de medicação não tem efeito na média do tempo que leva para um analgésico aliviar a dor.
- $H_a$: O tipo de medicação tem efeito na média do tempo que leva para um analgésico aliviar a dor.

**Hipóteses para efeitos de interação:**

- $H_0$: Não há efeito de interação entre gênero e tipo de medicação no tempo médio que leva para um analgésico aliviar a dor.
- $H_a$: Há um efeito de interação entre gênero e tipo de medicação no tempo médio que leva para um analgésico aliviar a dor.

# ANOVA bidirecional

- Um teste ANOVA bidirecional calcula um teste estatístico *F* para cada hipótese.
- É possível rejeitar nenhuma, uma, duas, ou todas as hipóteses nulas.
- Pode-se usar uma ferramenta tecnológica como o MINITAB para realizar um teste ANOVA bidirecional.

# Objetivos

- Usar análise de variância com um fator para testar afirmações envolvendo três ou mais médias.
- Introduzimos a análise de variância com dois fatores.