Especialização em *Data Science* e Estatística Aplicada

Módulo II - Análise estatística de várias populações

Profa. Dra. Tatiane F N Melo

Goiânia, 2024

IME INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
FEN FACULDADE DE ENFERMAGEM
UFG UNIVERSIDADE FEDERAL DE GOIÁS

Aula 2 - Parte 2

1. Continuação: Inferência estatística para duas populações
   - Teste de hipótese para igualdade de duas proporções

2. Análise de aderência e associação
   - Teste de Aderência (*Goodness-of-Fit Test*)

3. Referências Bibliográficas

Teste de hipótese para igualdade de duas proporções

Exemplo 4

- Objetivo: saber se a proporção de vacinados contra COVID-19, com primeira dose, em Aparecidade de Goiânia é igual à proporção de vacinados em Anápolis (Ministério da Saúde - Vacinômetro COVID-19).
- Neste caso, podemos usar o teste de hipótese para igualdade de duas proporções.
- Hipóteses de interesse:

$$H_0 : p_{AG} = p_A \text{ contra } H_1 : p_{AG} \neq p_A,$$

onde p_{AG} é a proporção de vacinados contra COVID-19, com primeira dose, em Aparecida de Goiânia e p_A é a proporção de vacinados, também com primeira dose, em Anápolis.

Teste de hipótese para igualdade de duas proporções

Comparação das proporções de duas populações

- Ao trabalhar com proporções, a escolha do teste adequado depende do tamanho da amostra.
  - Utilizamos o teste z para amostras grandes.
  - Quando a amostra é pequena, usamos a correção de continuidade de Yates, que veremos na próxima aula, ou o Teste Exato de Fisher, que será visto na disciplina de Métodos não paramétricos.

Comparação das proporções de duas populações

Teste z

Para testar a hipótese de que as proporções das duas populações são iguais, aplica-se o teste Z, da seguinte forma:

1. Estabeleça as hipóteses

   H_0: as proporções populacionais são iguais
   H_1: as proporções populacionais são diferentes;

2. escolha o nível de significância: α;

3. calcule a proporção amostral de cada grupo: considere $\widehat{p}_1$ sendo a proporção amostral do grupo 1 e $\widehat{p}_2$ a proporção amostral do grupo 2;

Comparação das proporções de duas populações

Teste z

4. calcule o valor de z_{obs}:

$$z_{obs} = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\overline{p}(1-\overline{p})(\frac{1}{n_1} + \frac{1}{n_2})}},$$

   onde $\overline{p} = (x_1 + x_2)/(n_1 + n_2)$, x_1 e x_2 são os números de sucessos observados nas duas amostras, n_1 e n_2 são os tamanhos das duas amostras.

5. calcule o valor-p:

$$\widehat{\alpha} = 2 \cdot P(Z > z_{obs} | H_0),$$

   onde Z tem distribuição, aproximadamente, $N(0, 1)$;

6. rejeite a hipótese de que as proporções das duas populações são iguais, sempre que o valor-p ($\widehat{\alpha}$) for menor que o nível de significância estabelecido (α).

Comparação das proporções de duas populações

Voltando ao Exemplo 4

Exemplo no R

Testes de aderência, independência e homogeneidade

Os testes de aderência, independência e homogeneidade são testes estatísticos que utilizam a estatística χ^2 (qui-quadrado) para analisar dados categóricos (discretos), mas têm diferentes objetivos e aplicações.

Teste de Aderência

Objetivo

Verificar se a distribuição observada dos dados em uma única variável categórica segue uma distribuição teórica esperada.

Teste de Aderência

- Os testes de aderência podem ser realizados em casos em que a distribuição teórica esperada é a binomial, a Poisson ou qualquer outra distribuição. Também pode ser baseada em dados históricos.
- Aqui, vamos ilustrar mais detalhadamente estes testes com exemplos.
- Abaixo, segue a descrição do teste.

Testes de aderência

Descrição do teste

Consideremos um experimento aleatório onde:

- k é o número de classes;
- O_i é a frequência absoluta observada da i-ésima categoria;
- E_i é a frequência absoluta esperada da i-ésima categoria.

Testes de aderência

Descrição do teste

- Definimos

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{\phi},$$

onde $\phi = k - 1 - p$ e p é o número de parâmetros a serem estimados.

- O valor-p é calculado por:

$$\widehat{\alpha} = P(\chi^2 > \chi^2_{obs} | H_0),$$

onde χ^2_{obs} é o valor que a estatística de teste assume.

Testes de aderência

- Para utilizar o software R, é necessário fornecer o vetor com as frequências observadas e as probabilidades (proporções) esperadas.
- Caso não tenhamos as probabilidades esperadas, o R assume uma distribuição teórica uniforme na hipótese nula, isso significa que se espera que todas as k categorias (ou classes) têm a mesma proporção $(1/k)$.

Teste de Aderência

Exemplo 5

Distribuição Binomial: Considere um estudo cujo objetivo é determinar a aceitação de um novo analgésico pelos pacientes. Para isso, 100 médicos selecionaram uma amostra de pacientes para participar do estudo. Cada médico selecionou de 25 pacientes. Após experimentar o novo analgésico por um período de tempo especificado, cada paciente foi questionado se preferia o novo medicamento ou o analgésico utilizado regularmente no passado.

Continuação do Exemplo 5

Tabela 1: Resultados do estudo.

Número de pacientes de 25, que preferem novo analgésico	Número de médicos relatando esse número	Número total de pacientes preferindo novo analgésico, por médico
0	5	0
1	6	6
2	8	16
3	10	30
4	10	40
5	15	75
6	17	102
7	10	70
8	10	80
9	9	81
10 ou mais	0	0
Total	100	500

Teste de Aderência

Continuação do Exemplo 5

- Estamos interessados em determinar se esses dados são ou não compatíveis com a hipótese de que foram extraídos de uma população que segue uma distribuição binomial.

- Ou seja,

  H_0: os dados vieram de uma população que segue uma distribuição binomial;

  H_1: os dados não vieram de uma população que segue uma distribuição binomial.

- Neste caso, é adequado empregarmos um teste de aderência.

Teste de Aderência

Solução do Exemplo 5

- Como o parâmetro binomial, p, não é especificado, ele deve ser estimado a partir dos dados da amostra.
- Um total de 500 pacientes dos 2.500 participantes do estudo disseram que, preferiam o novo analgésico, de modo que nossa estimativa pontual de p é $\widehat{p} = 500/2.500 = 0,20$.

Teste de Aderência

Solução do Exemplo 5

- Sabemos que a função de probabilidade de uma variável aleatória X com distribuição $Bin(n,p)$ é dada por:

$$P(X=k) = \binom{n}{k} \times p^k \times (1-p)^{n-k},$$

onde $k = 0, 1, 2, \ldots, 25.$

Teste de Aderência

Solução do Exemplo 5

- Por exemplo, para encontrar a probabilidade de que, de uma amostra de 25 pacientes, nenhum prefira o novo analgésico, quando na população total a verdadeira proporção que prefere o novo analgésico é 0.2, calculamos $P(X = 0)$. Isso pode ser feito de duas formas:
  - Analiticamente:

$$P(X = 0) = \binom{25}{0} \times 0.20^{0} \times 0.80^{25-0} = \frac{25}{(25-0) \times 0} \times 1 \times 0.80^{25} = 0,0038.$$

  - Usando o software `R`: $P(X = 0)$ = `dbinom(0, 25, 0.2)` = 0,0038.

- Para obter a frequência esperada correspondente, multiplicamos 0,0038 por 100 para obter 0,38.

Solução do Exemplo 5

Número de pacientes de 25, que preferem novo analgésico	Número de médicos relatando esse número (o_i)	No R	Frequências relativas esperadas	Frequências esperadas (e_i)
0	5	`dbinom(0, 25, 0.2)`	0,0038	0,38
1	6	`dbinom(1, 25, 0.2)`	0,0236	2,36
2	8	`dbinom(2, 25, 0.2)`	0,0708	7,08
3	10	`dbinom(3, 25, 0.2)`	0,1358	13,58
4	10	`dbinom(4, 25, 0.2)`	0,1867	18,67
5	15	`dbinom(5, 25, 0.2)`	0,1960	19,60
6	17	`dbinom(6, 25, 0.2)`	0,1633	16,33
7	10	`dbinom(7, 25, 0.2)`	0,1109	11,09
8	10	`dbinom(8, 25, 0.2)`	0,0623	6,23
9	9	`dbinom(9, 25, 0.2)`	0,0295	2,95
10 ou mais	0	`1-pbinom(9, 25, 0.2)`	0,0173	1,73
Total	100		1	100

Teste de Aderência

Solução do Exemplo 5

Observamos que, a primeira frequência esperada é menor que 1, então devemos unir este grupo com o segundo grupo. Quando fazemos isso, todas as frequências esperadas são maiores que 1.

Solução do Exemplo 5

Número de pacientes de 25, que preferem novo analgésico	Número de médicos relatando esse número (o_i)	No R	Frequências relativas esperadas	Frequências esperadas (e_i)
0 e 1	11	`pbinom(1, 25, 0.2)`	0,0274	2,74
2	8	`dbinom(2, 25, 0.2)`	0,0708	7,08
3	10	`dbinom(3, 25, 0.2)`	0,1358	13,58
4	10	`dbinom(4, 25, 0.2)`	0,1867	18,67
5	15	`dbinom(5, 25, 0.2)`	0,1960	19,60
6	17	`dbinom(6, 25, 0.2)`	0,1633	16,33
7	10	`dbinom(7, 25, 0.2)`	0,1109	11,09
8	10	`dbinom(8, 25, 0.2)`	0,0623	6,23
9	9	`dbinom(9, 25, 0.2)`	0,0295	2,95
10 ou mais	0	`1-pbinom(9, 25, 0.2)`	0,0173	1,73
Total	100		1	100

Teste de Aderência

Solução do Exemplo 5

- Agora podemos calcular:

$$\chi^2_{obs} = \frac{(11-2,74)^2}{2,74} + \frac{(8-7,08)^2}{7,08} + \frac{(10-13,58)^2}{13,58} + \ldots + \frac{(0-1,73)^2}{1,73} = 47,678.$$

- Calculando o valor-p no R:

  $\widehat{\alpha}$ = `pchisq(47.678, df = nc-1,lower.tail = FALSE)` = 1.13847e-07 = 0.000000138,

  onde nc é o número de classes final. Aqui, $nc = 10$.

Teste de Aderência

Portanto, rejeitamos a hipótese nula de que os dados vieram de uma distribuição binomial.

Vamos refazer o exemplo no `R`.

Teste de Aderência - Aplicação à dados reais

Exemplo 6

- Agora usaremos os dados referentes à distribuição das doses das vacinas contra COVID-19, em Goiânia, aplicadas em diferentes faixas etárias (Ministério da Saúde - Vacinômetro COVID-19).

- **Objetivo:** verificar se essa distribuição segue uma distribuição esperada baseada nas proporções da população.

- Ou seja,

  H_0: as vacinas são distribuídas proporcionalmente à população em cada faixa etária.

  H_1: as vacinas não são distribuídas proporcionalmente à população em cada faixa etária.

Teste de Aderência - Aplicação à dados reais

Continuação do Exemplo 6

Exemplo no R.

Referências bibliográficas

1. VIEIRA, S. Introdução à Bioestatística, 5ª Edição, Elsevier, 2008.
2. Ministério da Saúde - Vacinômetro COVID-19. `https://infoms.saude.gov.br/extensions/SEIDIGI_DEMAS_Vacina_C19/SEIDIGI_DEMAS_Vacina_C19.html`

Especialização em *Data Science* e Estatística Aplicada

Módulo II - Análise estatística de várias populações

Profa. Dra. Tatiane F N Melo

tmelo@ufg.br

IME
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

FEN
FACULDADE DE ENFERMAGEM

UFG
UNIVERSIDADE FEDERAL DE GOIÁS