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 1

1. Continuação: Inferência estatística para duas populações
   - Comparação das variâncias dos grupos
   - Teste t para comparar médias quando as variâncias são iguais (homocedásticas)
   - Teste t para comparar médias quando as variâncias são diferentes (heterocedásticas)

2. Referências Bibliográficas

Relembrando o final da aula passada

- Na aula passada, vimos que quando o interesse é estudar testes de hipótese para comparação de médias, podemos ter dois casos, a saber: (i) dados pareados e (ii) dados não pareados (independentes).
- Estudamos o caso dos dados pareados e agora o objetivo é estudar os testes para dados não pareados ou independentes.
- O teste t para grupos independentes compara as médias de medidas da mesma variável contínua, obtidas de forma independente em cada um de dois grupos.
- Antes de realizar o teste t, é preciso verificar se as variâncias dos grupos são iguais ou diferentes.

Dados não pareados (independentes)

Neste sentido, consideraremos, os seguintes casos:

1º Caso) Comparação das variâncias dos grupos

2º Caso) Teste t para comparar médias quando as variâncias são iguais (homocedásticas)

3º Caso) Teste t para comparar médias quando as variâncias são diferentes (heterocedásticas)

1º Caso) Comparação das variâncias dos grupos

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

1. Estabeleça as hipóteses

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

2. escolha o nível de significância: α;

3. calcule a variância amostral de cada grupo: considere s_1^2 sendo a variância do grupo 1 e s_2^2 a variância do grupo 2;

1º Caso) Comparação das variâncias dos grupos

4. calcule o valor de f_{obs}, dado pela razão entre a maior e a menor variância, ou seja, se $s_1^2 > s_2^2$, então
$$f_{obs} = \frac{s_1^2}{s_2^2};$$
5. calcule o valor-p:
$$\widehat{\alpha} = 2 \cdot min\{P(F > f_{obs}|\mathcal{H}_0), P(F < f_{obs}|H_0)\},$$
onde F tem distribuição F-Snedecor com $(n_1 - 1)$ e $(n_2 - 1)$ graus de liberdade (se $s_1^2 > s_2^2$);
6. rejeite a hipótese de que as variâncias das duas populações são iguais, sempre que o valor-p ($\widehat{\alpha}$) for menor que o nível de significância estabelecido (α).

1º Caso) Comparação das variâncias dos grupos

Exemplo 1

Suponha que estamos interessados em verificar se há uma diferença significativa na variabilidade do número de vacinas aplicadas diariamente em dois municípios do estado de Goiás, a saber: Goiânia e Aparecida de Goiânia, em 2022, $n = 331$ (Ministério da Saúde - Vacinômetro COVID-19).

1º Caso) Comparação das variâncias dos grupos

Continuação do Exemplo 1

- Nesse exemplo, queremos testar:

  H_0: a variabilidade do número de vacinas aplicadas na cidade de Goiânia é igual a variabilidade do número de vacinas aplicadas em Aparecida de Goiânia;

  H_1: a variabilidade do número de vacinas aplicadas na cidade de Goiânia é diferente da variabilidade do número de vacinas aplicadas em Aparecida de Goiânia.

- Ou seja,

$$H_0 : \sigma^2_G = \sigma^2_{AG} \text{ contra } H_1 : \sigma^2_G \neq \sigma^2_{AG},$$

  onde σ^2_G é a variabilidade do número de vacinas aplicadas em Goiânia e σ^2_{AG} é a variabilidade do número de vacinas aplicadas em Aparecida de Goiânia.

1º Caso) Comparação das variâncias dos grupos

Continuação do Exemplo 1

Exemplo 1 no R.

1º Caso) Comparação das variâncias dos grupos

Exemplo 2

Agora vamos supor que estamos interessados em verificar se há uma diferença significativa na variabilidade das idades das mães, de nascidos vivos, entre aquelas que fizeram parto vaginal com aquelas que foram submetidas ao parto cesário. Os dados foram obtidos da Base de dados SINASC, do Estado de São Paulo, em 2023.

Neste caso, foram 28.812 mães que fizeram partos vaginais e 36.718 foram submetidas ao parto cesário. Logo, $n_1 = 28.812$ e $n_2 = 36.718$.

1º Caso) Comparação das variâncias dos grupos

Continuação do Exemplo 2

- Aqui, o teste de interesse é:

  H_0: a variabilidade das idades das mães que fizeram parto vaginal é igual a variabilidade das idades daquelas que foram submetidas ao parto cesário;

  H_1: a variabilidade das idades das mães que fizeram parto vaginal é diferente a variabilidade das idades daquelas que foram submetidas ao parto cesário.

- Ou seja,

$$H_0 : \sigma_V^2 = \sigma_C^2 \text{ contra } H_1 : \sigma_V^2 \neq \sigma_C^2,$$

  onde σ_V^2 é a variabilidade das idades das mães que fizeram parto vaginal e σ_C^2 é a variabilidade das idades das mães que foram submetidas ao parto cesário.

1º Caso) Comparação das variâncias dos grupos

Continuação do Exemplo 2

Exemplo 2 no R.

1º Caso) Comparação das variâncias dos grupos

Exemplo 3

Consideremos os dados referentes ao número de vacinas aplicadas contra COVID-19 (Ministério da Saúde - Vacinômetro COVID-19) em 20 dias do mês novembro de 2022, nos municípios de Formosa e Catalão - estados de Goiás. Neste caso, temos que o tamanho amostral é igual a $n = 20$.

1º Caso) Comparação das variâncias dos grupos

Continuação do Exemplo 3

- Aqui, o teste de interesse é:

  H_0: a variabilidade no número de doses aplicadas no município de Formosa é igual à variabilidade no número de doses aplicadas no município de Catalão;

  H_1:a variabilidade no número de doses aplicadas no município de Formosa é diferente à variabilidade no número de doses aplicadas no município de Catalão.

- Ou seja,

$$H_0 : \sigma^2_{Form} = \sigma^2_{Cat} \text{ contra } H_1 : \sigma^2_{Form} \neq \sigma^2_{Cat},$$

  onde σ^2_{Form} é a variabilidade no número de doses aplicadas no município de Formosa e σ^2_{Cat} é a variabilidade no número de doses aplicadas no município de Catalão.

1º Caso) Comparação das variâncias dos grupos

Continuação do Exemplo 3
Exemplo 3 no R.

2º Caso) Teste t para comparar médias quando as variâncias são iguais (homocedásticas)

Quando o teste F indica a não rejeição da hipótese de igualdade entre variâncias, podemos considerar que as variâncias são iguais. Para calcular o valor de t, siga estes passos:

1. estabeleça as hipóteses (H_0 e H_1);
2. estabeleça o nível de significância (α);
3. calcule a média de cada grupo ($\bar{x}_1$ e $\bar{x}_2$);
4. calcule a variância de cada grupo (s_1^2 e s_2^2);
5. calcule a variância ponderada, dada pela fórmula:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

2º Caso) Teste t para comparar médias quando as variâncias são iguais (homocedásticas)

6. calcule o valor de t_{obs}, pela seguinte fórmula:

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})s_p^2}};$$

7. calcule o valor-p:

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

onde T é uma variável aleatória com distribuição t-Student com $(n_1 + n_2 - 2)$ graus de liberdade, n é o tamanho da amostra.

2º Caso) Teste t para comparar médias quando as variâncias são iguais (homocedásticas)

Voltando Exemplo 2

- Nesse exemplo, não rejeitamos a hipótese de igualdade entre as variâncias, então podemos usá-lo neste caso.
- Agora queremos testar:

H_0: as idades médias das mães que fizeram parto vaginal e cesário são iguais.

H_1: as idades médias das mães que fizeram parto vaginal e cesário são diferentes.

2º Caso) Teste t para comparar médias quando as variâncias são iguais (homocedásticas)

Voltando Exemplo 2

Vamos reescrever as hipóteses, matematicamente, da seguinte forma:

$$H_0 : \mu_V = \mu_C \text{ contra } H_1 : \mu_V \neq \mu_C,$$

onde μ_V é a idade média das mães que fizeram parto vaginal e μ_C é a idade média das mães que foram submetidas ao parto cesário.

2º Caso) Teste t para comparar médias quando as variâncias são iguais (homocedásticas)

Voltando Exemplo 2

Exemplo no R.

3º Caso) Teste t para comparar médias quando as variâncias são diferentes (heterocedásticas)

Quando as variâncias são diferentes, para comparar duas médias, aplica-se o teste t, da seguinte forma:

1. estabeleça as hipóteses (H_0 e H_1);
2. estabeleça o nível de significância (α);
3. calcule a média de cada grupo ($\bar{x}_1$ e $\bar{x}_2$);
4. calcule a variância de cada grupo (s_1^2 e s_2^2);

3º Caso) Teste t para comparar médias quando as variâncias são diferentes (heterocedásticas)

5. calcule o valor de t_{obs}, pela seguinte fórmula:

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}};$$

3º Caso) Teste t para comparar médias quando as variâncias são diferentes (heterocedásticas)

6. calcule o valor-p:

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

onde T é uma variável aleatória com distribuição t-Student com g graus de liberdade (usar a parte inteira), onde

$$g = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}.$$

3º Caso) Teste t para comparar médias quando as variâncias são diferentes (heterocedásticas)

Voltando Exemplo 1

- Nesse exemplo, rejeitamos a hipótese de igualdade entre as variâncias, ou seja, indicação de variâncias diferentes.
- Agora queremos testar:

  H_0: o número médio de vacinas aplicadas na cidade de Goiânia é igual ao número médio de vacinas aplicadas em Aparecida de Goiânia, em 2022.

  H_1: o número médio de vacinas aplicadas na cidade de Goiânia é diferente do número médio de vacinas aplicadas em Aparecida de Goiânia, em 2022.

3º Caso) Teste t para comparar médias quando as variâncias são diferentes (heterocedásticas)

Voltando Exemplo 1

Matematicamente, temos as hipóteses:

$$H_0 : \mu_G = \mu_{AG} \text{ contra } H_1 : \mu_G \neq \mu_{AG},$$

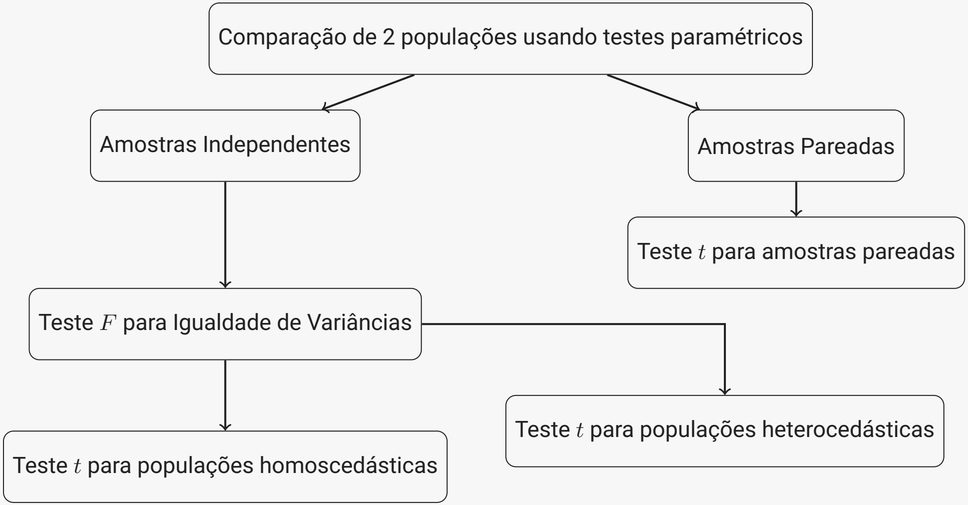
onde μ_G é o número médio de vacinas aplicadas na cidade de Goiânia e μ_{AG} é o número médio de vacinas aplicadas em Aparecida de Goiânia.

3º Caso) Teste t para comparar médias quando as variâncias são diferentes (heterocedásticas)

Voltando Exemplo 1

Exemplo no R.

Resumo - Comparação de dois grupos



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

3. Base de dados SINASC, Arquivos de nascidos vivos no município de São Paulo. https://www.prefeitura.sp.gov.br/cidade/secretarias/saude/epidemiologia_e_informacao/nascidos_vivos/index.php?p=306422, Ano: 2023.

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