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

Conteúdo Programático

- Distribuições qui-quadrado (χ^2) e F-Snedecor (Aula 1).
- Revisão - Testes de hipóteses (Aula 1).
- Teste de hipótese para variância (Aula 1).
- Inferência estatística para duas populações (Aulas 1 e 2).
- Análise de aderência e associação: testes de aderência, homogeneidade e independência (Aulas 2 e 3).
- Análise de variância de um fator (Aula 4).

Aula 1 - Parte 2

1. Teste de hipótese para igualdade de duas médias
   - Dados pareados
   - Dados não pareados (independentes)

2. Referências Bibliográficas

Teste de hipótese para igualdade de duas médias

Comparação das médias de duas populações

- Em muitas situações práticas é importante comparar 2 populações, baseados em dados fornecidos por amostras dessas populações.
- Aqui, analisaremos casos de comparações de médias de duas populações normais (X e Y).
- Em geral, faremos testes sobre a diferença entre duas médias populacionais:

$$H_0 : \mu_x - \mu_y = \mu_d,$$

sendo que na maioria dos casos $\mu_d = 0$, o que significa que estaremos testando a igualdade entre as médias:

$$H_0 : \mu_x = \mu_y.$$

Comparação das médias de duas populações

Vamos considerar dois casos de comparação de médias:

1 - Dados pareados ou emparelhados (populações correlacionadas): Dizemos que os dados são pareados se o pesquisador adotar um dos seguintes métodos para seu trabalho:

   (i) medir a mesma variável nas mesmas unidades, antes e depois de uma intervenção;

   (ii) recrutar participantes da pesquisa aos pares, ou parear os participantes por idade, sexo, estágio da doença. Depois, administrar o tratamento em teste a um dos participantes de cada par, escolhido ao acaso, e ao outro, o tratamento convencional;

   (iii) medir a mesma variável em gêmeos ou outro tipo de par, como mãe e filho.

2 - Dados não pareados (populações não correlacionadas).

Ensaio com dados pareados: duas medidas obtidas em cada indivíduo

Exemplo 4

Para verificar se duas drogas diferentes, usadas como antitussígenos (bloqueadores de tosse), alteram o tempo de sono, foi feito um ensaio com nove voluntários. Eles tomaram um dos antitussígenos na primeira noite e o outro, na noite seguinte. Foi registrado o tempo de sono de cada voluntário, nas duas noites consecutivas. A proposta consiste em comparar as médias de tempo de sono obtidas com cada antitussígeno.

Ensaio com dados pareados: medidas feitas em pares de unidades

Exemplo 5

Para verificar se uma droga é eficiente na inibição do crescimento de tumores, foram injetadas células cancerosas em 14 ratos similares. Em seguida, os tumores foram medidos e foram formados pares de ratos com tumores de mesmo tamanho. Por sorteio, um rato de cada par recebeu a droga (grupo tratado), enquanto o outro foi mantido como controle. A ideia é comparar as médias dos tamanhos de tumores de ratos tratados e de ratos controles.

Dados pareados

Pressupostos

1. **Independência das Observações:**
   - As observações em cada par devem ser independentes umas das outras.
   - Deve haver independência entre os pares, mas as observações dentro de cada par estão relacionadas.

Dados pareados

Pressupostos

2. **Normalidade das Diferenças:**
    - As diferenças entre os pares (subtraindo um valor do outro dentro de cada par) devem ser aproximadamente normalmente distribuídas.
    - Isso pode ser avaliado por meio de testes de normalidade ou inspeção gráfica (ex: histograma ou gráfico Q-Q das diferenças).
    - Este pressuposto é mais crítico para amostras pequenas ($n < 30$). Para amostras maiores, o teste que será utilizado (teste t) é robusto a desvios da normalidade devido ao Teorema Central do Limite.

Dados pareados

Pressupostos

3. **Escala dos dados:**
    - Os dados devem ser em uma escala que permita a subtração significativa entre os pares. Ou seja, sejam quantitativos (discretos ou contínuos).

Dados pareados

O que fazer?

Quando temos dados pareados, aplicamos o teste t. Mas o pareamento deve ter algum tipo de lógica; não basta ter duas amostras com o mesmo número de dados. Para fazer o teste t,

1. estabeleça as hipóteses;
2. escolha o nível de significância α;
3. calcule as diferenças entre todas as observações pareadas:

$$d = x_2 - x_1;$$

Dados pareados

O que fazer?

4. calcule a média dessas diferenças:

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n};$$

5. calcule a variância dessas diferenças:

$$s^2 = \frac{\left(\sum_{i=1}^{n} d_i^2\right) - \frac{\left(\sum_{i=1}^{n} d_i\right)^2}{n}}{n-1};$$

Dados pareados

6. calcule o valor que a estatística de teste assume, denominado por t_{obs}:

$$t_{obs} = \frac{\bar{d}}{\sqrt{\frac{s^2}{n}}};$$

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 $(n-1)$ graus de liberdade, n é o tamanho da amostra.

  - Sempre que o valor-p $(\widehat{\alpha})$ for menor que o nível de significância estabelecido (α), rejeite a hipótese nula, caso contrário, não rejeite H_0.

Aplicação à dados reais - Tamanho da amostra maior que 30

Exemplo 6

Vamos usar dados reais obtidos no site do Ministério da Saúde - Vacinômetro COVID-19. Estes dados são referentes às doses aplicadas da vacina contra COVID-19 e estão disponíveis em:

`https://infoms.saude.gov.br/extensions/SEIDIGI_DEMAS_Vacina_C19/SEIDIGI_DEMAS_Vacina_C19.html`

Exemplo 6 - Continuação

Objetivo

O interesse é verificar se há uma diferença significativa entre as médias do número de vacinas aplicadas nas primeiras e segundas doses, nos municípios do estado de Goiás.

- Aqui, estamos olhando para duas medições diferentes (primeira e segunda dose) nos mesmos municípios do estado de Goiás. Neste caso, é adequado usarmos o teste t para dados pareados.

Exemplo 6 - Continuação

Hipóteses de interesse

H_0: Não há diferença entre o número médio de vacinas aplicadas nas primeiras e segundas doses para os municípios analisados, ou seja,

$$H_0 : \mu_d = 0,$$

onde $\mu_d = \mu_1 - \mu_2$ é a diferença entre as médias das quantidades de vacinas aplicadas na primeira e na segunda dose.

H_1: Há diferença entre o número médio de vacinas aplicadas nas primeiras e segundas doses para os municípios analisados.

$$H_1 : \mu_d \neq 0.$$

Exemplo 6 - Continuação

- As unidades experimentais são os 246 municípios de Goiás. Então, $n = 246$.
- Os dados são referentes ao número de vacinas aplicadas em cada munícipio (primeiras e segundas doses)
- Aqui, a premissa de normalidade, das diferenças entre as observações pareadas, não é necessária, por causa do tamanho de n. Isto se deve ao Teorema Central do Limite, que afirma que a distribuição da média das amostras se aproxima de uma distribuição normal à medida que o tamanho da amostra aumenta.

Exemplo 6 - Continuação

> Exemplo 6 no R.

Aplicação à dados reais - Tamanho da amostra menor que 30

Exemplo 7

O dados reais usados aqui, também foram obtidos no site do Ministério da Saúde (Vacinômetro COVID-19).

Agora o tamanho da amostra é $n = 28$, então a premissa de normalidade, das diferenças entre as observações pareadas, é necessária.

Exemplo 7 - Continuação

Objetivo

O interesse é verificar se há uma diferença significativa entre as médias do número de vacinas aplicadas nas primeiras e segundas doses, na cidade de Goiânia, em novembro de 2021. Especificamente, para a vacina Pfizer.

- Aqui, também estamos olhando para duas medições diferentes (primeira e segunda dose) na mesma cidade (Goiânia). Então, é adequado usarmos o teste t para dados pareados.
- As unidades experimentais são 28 dias do mês de novembro de 2021.

Exemplo 7 - Continuação

Hipóteses de interesse

H_0: Não há diferença entre o número médio de vacinas aplicadas nas primeiras e segundas doses, na cidade de Goiânia, em novembro de 2021, ou seja,

$$H_0 : \mu_d = 0,$$

onde $\mu_d = \mu_1 - \mu_2$ é a diferença entre as médias das quantidades de vacinas aplicadas na primeira e na segunda dose.

H_1: Há diferença entre o número médio de vacinas aplicadas nas primeiras e segundas doses, na cidade de Goiânia, em novembro de 2021.

$$H_1 : \mu_d \neq 0.$$

Exemplo 7 - Continuação

> Exemplo 7 no R.

Dados não pareados (independentes)

Existem várias situações em que o pesquisador retira amostras de populações independentes, como por exemplo:

- comparar o nível de ansiedade de meninos e meninas no primeiro dia de aula;
- comparar dois grupos de pessoas, um grupo submetido a um novo tratamento enquanto o outro grupo é submetido a tratamento convencional.

Ensaio com dados não pareados: comparação de grupos independentes

Exemplo 8

Para saber se determinado produto faz nascer cabelos em pessoas calvas, um dermatologista pode fazer um ensaio clínico: um grupo de pessoas calvas recebe o tratamento em teste – grupo tratado –, enquanto um grupo de pessoas calvas recebe um placebo – grupo controle.

Dados não pareados (independentes)

Pressupostos

1. **Independência das Observações:**
    - As observações dentro de cada grupo e entre os grupos devem ser independentes umas das outras.
    - Cada grupo deve ser uma amostra aleatória da população.

Dados não pareados (independentes)

Pressupostos

2. **Normalidade:**
    - Os dados dentro de cada grupo devem ser normalmente distribuídos.
    - Este pressuposto é crítico para amostras pequenas ($n < 30$). Para amostras maiores, o teste t é robusto a desvios da normalidade (Teorema Central do Limite).

Dados não pareados (independentes)

Pressupostos

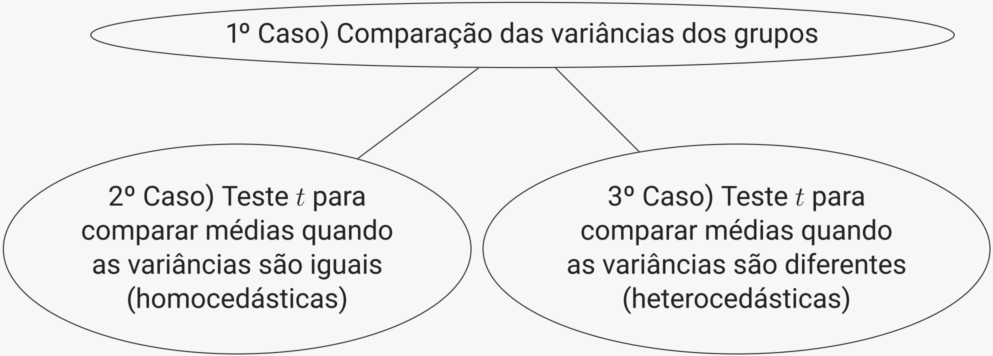
3. **Escala dos dados:**
   - Os dados devem ser variáveis quantitativas, que podem ser discretas ou contínuas.

Dados não pareados (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, na próxima aula, os seguintes casos:



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