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 1

1. Distribuições qui-quadrado (χ^2) e F-Snedecor

2. Revisão - Testes de Hipóteses

3. Teste de hipótese para variância

4. Referências Bibliográficas

Distribuição Qui-Quadrado (χ^2)

- A distribuição qui-quadrado (χ^2) é uma distribuição de probabilidade contínua que surge principalmente em problemas de inferência estatística.
- É definida apenas para valores não negativos e possui uma assimetria positiva.
  - A assimetria diminui à medida que o número de graus de liberdade aumenta.
- Esta distribuição depende de um parâmetro chamado graus de liberdade, denotado por k.
  - Para um número de k graus de liberdade, a distribuição é a soma dos quadrados de k variáveis normais padrão independentes.
- Média: $\mathbb{E}(X) = k$

  Variância: $\text{Var}(X) = 2k$
- Notação: χ^2_k - distribuição qui-quadrado com k graus de liberdade.

Distribuição Qui-Quadrado (χ^2)

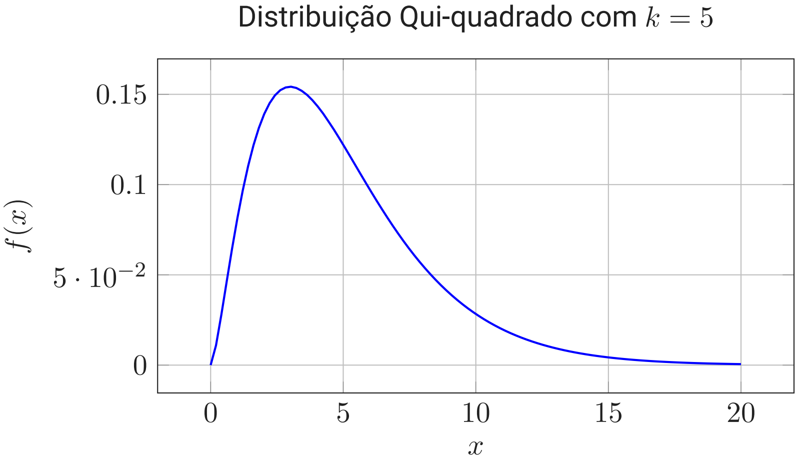
Para uma variável X que segue uma distribuição qui-quadrado com k graus de liberdade temos a seguinte função densidade de probabilidade:

$$f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

onde:

- $x \geq 0$ é a variável aleatória,
- k é o número de graus de liberdade,
- Γ é a função gama.

Distribuição Qui-Quadrado (χ^2)



Distribuição Qui-Quadrado (χ^2)

Exemplo 1

Distribuição χ^2 no R.

Distribuição F-Snedecor (F)

- A distribuição F de Snedecor é uma distribuição de probabilidade que surge frequentemente em testes estatísticos, especialmente em análises de variância (veremos na Aula 4).
- Ela é utilizada para comparar variâncias entre grupos e testar hipóteses sobre a igualdade de variâncias.
- Se F é uma variável aleatória que segue uma distribuição F com m e n graus de liberdade, F é definida como a razão de duas variáveis aleatórias qui-quadrado independentes divididas pelos seus respectivos graus de liberdade.

Distribuição F-Snedecor (F)

- A distribuição F é assimétrica e sempre positiva.
- Média: $\mathbb{E}(X) = \frac{n}{n-2}$ se $n > 2$.

  Variância: Var$(X) = \frac{2n^2(m+n-2)}{m(n-4)(n-2)^2}$ se $n > 4$.
- Notação: $F_{m,n}$ - distribuição F com m e n graus de liberdade.

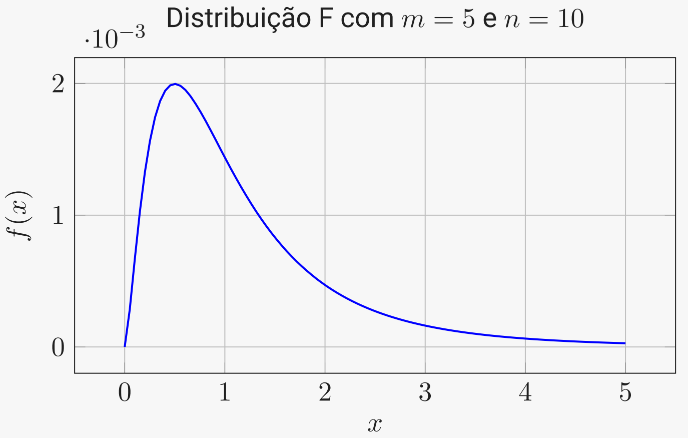
Distribuição F-Snedecor (F)

A função densidade da distribuição F com m e n graus de liberdade é dada por:

$$f(x; m, n) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \cdot \frac{\left(\frac{m}{n}\right)^{m/2} x^{\frac{m}{2}-1}\left(1+\frac{mx}{n}\right)^{-\frac{m+n}{2}}}{n^{m/2}}.$$

onde $\Gamma(\cdot)$ é a função gama.

Distribuição F-Snedecor (F)



Distribuição F-Snedecor (F)

Exemplo 2

Distribuição F no R.

Revisão - Testes de Hipóteses

- **Hipótese:** Em Estatística, uma hipótese é uma afirmação sobre uma propriedade da população.

- **Teste de Hipótese:** É um procedimento padrão para testar uma afirmação sobre uma propriedade da população.

Revisão - Testes de Hipóteses

Componentes de um teste de hipótese:

- Hipótese nula (representada por H_0): É uma afirmação de que o valor de um parâmetro populacional é igual ($\leq$ ou $\geq$) a algum valor especificado.

- Hipótese alternativa (representada por H_1): É a afirmação de que o parâmetro tem um valor que, de alguma forma, difere da hipótese nula.

Revisão - Testes de Hipóteses

Exemplos

- $H_0 : \mu = \mu_0$ contra $H_1 : \mu \neq \mu_0$;
- $H_0 : \mu \geq \mu_0$ contra $H_1 : \mu < \mu_0$;
- $H_0 : \mu \leq \mu_0$ contra $H_1 : \mu > \mu_0$.

1. Se $H_1 : \mu \neq \mu_0$ temos um teste bilateral.
2. Se $H_1 : \mu < \mu_0$ temos um teste unilateral à esquerda.
3. Se $H_1 : \mu > \mu_0$ temos um teste unilateral à direita.

Revisão - Testes de Hipóteses

Regra de decisão

- Em um teste de hipóteses, podemos decidir rejeitar ou não a hipótese nula (H_0). A seguir, apresentamos o cálculo de uma medida crucial na regra de decisão, conhecida como valor-p ($\widehat{\alpha}$), que ajuda a determinar se devemos rejeitar H_0.

- Outra medida utilizada é o nível de significância, que representa a probabilidade máxima de cometer um erro do Tipo I (rejeitar uma hipótese nula verdadeira). Representado por α, é um valor previamente estabelecido que define o limiar para considerar um resultado como estatisticamente significativo.

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

Revisão - Testes de Hipóteses

Valor-p: $\widehat{\alpha}$

Seja T a estatística de teste (variável aleatória) e t_{obs} é o valor que a estatística de teste assume (valor observado). A estatística T será definida de acordo com cada problema.

- **Teste bilateral:** Se a distribuição for simétrica (Normal e t-Student) então

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

  Se a distribuição for assimétrica (Qui-Quadrado e F-Snedecor - veremos a seguir) então

$$\widehat{\alpha} = 2 \cdot min\{P(T > t_{obs}|\mathcal{H}_0), P(T < t_{obs}|H_0)\}.$$

- **Teste unilateral à direita:** $\widehat{\alpha} = P(T > t_{obs}|H_0).$
- **Teste unilateral à esquerda:** $\widehat{\alpha} = P(T < t_{obs}|H_0).$

Testes de Hipóteses - Uma população

Teste de hipótese para variância

- O teste de hipóteses para variância é usado para verificar se a variância populacional (σ^2) é igual a um valor específico.
  - Esse teste é útil quando estamos interessados na variabilidade dos dados.

- Para aplicar o teste para a variância é necessário supor a normalidade da população de onde será extraída a amostra.
  - o entanto, para tamanhos amostrais grandes, a suposição de normalidade pode ser relaxada devido ao Teorema Central do Limite (TCL), que sugere que a distribuição se aproxima de uma normal conforme o tamanho da amostra aumenta, independentemente da distribuição original da população.

Testes de Hipóteses - Uma população

Teste de hipótese para variância

- **Hipóteses:**
  - H_0: A variância populacional (σ^2) é igual a um valor específico (σ_0^2), ou seja,

$$H_0 : \sigma^2 = \sigma_0^2.$$

  - H_1: A variância populacional (σ^2) não é igual a um valor específico (σ_0^2), ou seja,

$$H_1 : \sigma^2 \neq \sigma_0^2 \ \text{ ou } \ H_1 : \sigma^2 > \sigma_0^2 \ \text{ ou } \ H_1 : \sigma^2 < \sigma_0^2.$$

Testes de Hipóteses - Uma população

Teste de hipótese para variância

- **Estatística de teste:** Para uma amostra de tamanho n com variância amostral S^2, a estatística de teste é dada por:

$$Q = \frac{(n-1)S^2}{\sigma_0^2},$$

onde $S^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$.

  - A estatística Q tem distribuição χ^2_{n-1}.

- O valor-p é calculado a partir da distribuição qui-quadrado e usado para decidir se rejeitamos a hipótese nula.

Teste de hipótese para variância

Exemplo 3

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`

Teste de hipótese para variância

Continuação do Exemplo 3

- Considere o conjunto de dados referente à vacinação (3ª dose) contra COVID-19, no estado do Espírito Santo, de março à dezembro de 2023. Suponha que queremos testar se a variabilidade do número de doses aplicadas é igual a um valor específico, digamos σ_0^2 (esse valor depende do que o pesquisador vai definir).

- Aqui, vamos supor $\sigma_0^2 = 1000$.

- Vamos realizar um teste bilateral, ou seja,
    - H_0 : A variância das doses administradas é igual a 1000.
    - H_1: A variância das doses administradas é diferente de 1000.

- Agora vamos resolver este exemplo usando o software 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