Universidade Federal do Ceará

Faculdade de Economia

Métodos Quantitativos

Vicente Lima Crisóstomo

Fortaleza, 2020

Sumário

- Introdução
- Estatística Descritiva
- Probabilidade
- Distribuições de Probabilidades
- Amostragem e Distribuições Amostrais
- Estimação
- Testes de Significância
- Análise de Variância
- Teste de Significância para Proporções
- Testes Não Paramétricos
- Correlação e Regressão

Análise de Variância (ANOVA)

- Análise de Variância
  - Testar igualdade de médias de duas ou mais populações
  - Variabilidade pode indicar algo sobre a média

- Hipóteses
  - **H_0: As médias das populações são iguais**

  - **H_1: As médias das populações não são iguais**

Análise de Variância

- Suposições para uso da Análise de Variância
  - Amostras aleatórias e independentes
  - Amostras extraídas de populações normais
  - Populações devem ter variâncias iguais

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \cdots = \sigma_k^2$$

Análise de Variância

- Cálculo da Variância
  - Cálculo da média da amostra
  - Cálculo do quadrado da diferença entre observação e média
  - Somar os quadrados
  - Dividir por (n-1) para variância amostral

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

Análise de Variância

- Fundamento
  - Exame da Variância pode revelar algo sobre médias
    - Pode revelar se as médias populacionais são iguais ou não
  - Para o conjunto de amostras a testar
    - Estima-se a variância populacional por dois processos
    - Se as duas estimativas são iguais
      - Aceita-se H0 (médias são iguais)
    - Se há grande diferença entre as duas estimativas
      - Rejeita-se H0 (médias não são iguais)

Análise de Variância

- Se H0 verdadeira
  - Amostras são provenientes de populações com médias iguais
- Duas formas para
  - Estimar Variância Populacional
    - **Variância entre as médias amostrais**
      - Variância Sistemática
        - **Estimativa "entre" (ou "*between*") (as amostras)**
    - **Média das Variâncias Amostrais**
      - Variância Não Sistemática
        - **Estimativa "dentro" (ou "*within*") (de cada amostra)**

Análise de Variância

- Estimar Variância Populacional
  - Estimar Variância Populacional
    - ***Variância entre as médias amostrais***
      - Variância Sistemática
        - Estimativa "entre" (ou "*between*") amostras
    - Estimativa amostra ***entre*** amostras
      - Permite uma estimativa das variâncias das populações
        - Através de uma distribuição amostral de médias
    - H0 verdadeira significa
      - Amostras provêem da mesma população normal
    - Teorema do Limite Central
      - Distribuição amostral de médias de uma população normal tem distribuição normal

Análise de Variância

- ***Variância entre as médias amostrais***
  - Teorema do Limite Central
    - Distribuição amostral de médias de uma população normal tem distribuição normal
  - Desvio padrão da distribuição amostral (raiz quadrada da variância amostral)
    - Está relacionado com o DP da população

Análise de Variância

- ***Variância entre as médias amostrais***

$$DP\ da\ Distribuicao\ Amostral\ de\ Médias = \frac{DP\ da\ populacao}{\sqrt{Tamanho\ da\ amostra}}$$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \qquad \sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}$$

$$s_{\bar{x}}^2 = \frac{s_x^2}{n} \quad \therefore \quad s_x^2 = n \cdot s_{\bar{x}}^2$$

Análise de Variância

- ***Variância entre as médias amostrais***
- Não se conhece Variância populacional
  - Usa-se a variabilidade (DP ou Variância) amostral para estimar o parâmetro da distribuição da qual se extraiu a amostra
- Tem-se um conjunto de médias amostrais
  - Se H0 é verdadeira significa que médias amostrais provêem da mesma distribuição amostral
- ***Determinação da Variância das médias amostrais***
  - Permite estimar a variância populacional

Análise de Variância

- <u>***Variância entre as médias amostrais***</u>
- <u>***Determinação da Variância das médias amostrais***</u>

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \bar{x}_j}{k} \qquad s_{\bar{x}}^2 = \left[\frac{\sum_{j=1}^{k}\left(\bar{x}_j - \bar{\bar{x}}\right)^2}{k-1}\right]$$

$$s_{\bar{x}}^2 = \frac{s_x^2}{n} \quad \therefore \quad s_x^2 = n \cdot s_{\bar{x}}^2$$

$$s_{entre/between}^2 = s_b^2 = n \cdot s_{\bar{x}}^2$$

Análise de Variância

- ***Variância entre as médias amostrais***
- ***Determinação da Variância das médias amostrais***
  - Médias iguais, ou quase iguais, apresentarão baixa variabilidade
    - Baixo valor de $\boldsymbol{s_b^2}$
  - Médias diferentes, ou mais dispersas/distantes, apresentarão alta variabilidade
    - Alto valor de $\boldsymbol{s_b^2}$

Análise de Variância

- Estimar Variância Populacional
  - Média das Variâncias Amostrais
    - **Estimativa da variância "dentro" (ou "*within*")**
  - Cada variância amostral representa
    - Variação de cada amostra especificamente
      - Variação **dentro** daquela amostra
  - Uma forma de estimara a Variância Populacional
    - Calcular a Média das variâncias amostrais
      - Proporciona boa estimativa do conjunto de amostras
        - Representa bom número de observações
  - Estimativa da variância baseada na média das variâncias amostrais
    - Estimativa da variância **dentro** (***within***)

Análise de Variância

- Estimar Variância Populacional
  - Cálculo da Estimativa da variância **dentro** (***within***)
    - Médias das variâncias dentro de cada amostra

$$s_w^2 = \frac{s_1^2 + s_2^2 + s_3^2 + \cdots + s_k^2}{k}$$

Análise de Variância

- Razão F
  - Estatística de teste F
    - Comparação com Tabela de valores F

$$Razão\ F = \frac{s_b^2}{s_w^2} = \frac{n \cdot s_{\bar{x}}^2}{\dfrac{s_1^2 + s_2^2 + s_3^2 + \cdots + s_k^2}{k}}$$

Análise de Variância

- Observar termos de F
  - Alto F
    - Grande **numerador** (estimativa **entre**) relativamente ao *denominador* (estimativa *dentro*)
  - Baixo F
    - Pequeno **numerador** (estimativa **entre**) relativamente ao *denominador* (estimativa *dentro*)
  - Numerador (estimativa entre) indica variabilidade entre médias amostrais
  - Denominador (estimativa dentro) indica média da variabilidade das amostras

Análise de Variância

- Exemplo
  - Quatro grupos. As médias dos quatro são iguais?

Observ	g1	g2	g3	g4
1	15,1	14,9	15,4	15,6
2	15,0	15,2	15,2	15,5
3	14,9	14,9	16,1	15,8
4	15,7	14,8	15,3	15,3
5	15,4	14,9	15,2	15,7
6	15,1	15,3	15,2	15,7
Média:	15,2	15,0	15,4	15,6
Variância:	0,088	0,040	0,124	0,032

Análise de Variância

- As médias dos quatro são iguais?
  - 1- Calcular variância **entre** amostras (***between***)
    - Número de observações X Variância das médias amostrais

$$s^2_{entre/between} = s^2_b = n \cdot s^2_{\bar{x}}$$

  - 2- Calcular variância **dentro** das amostras (***within***)
    - Médias das variâncias das amostras

$$s^2_w = \frac{s^2_1 + s^2_2 + s^2_3 + \cdots + s^2_k}{k}$$

  - 3- Calcular a Razão F

$$Razão\ F = \frac{s^2_b}{s^2_w} = \frac{n \cdot s^2_{\bar{x}}}{\frac{s^2_1 + s^2_2 + s^2_3 + \cdots + s^2_k}{k}}$$

  - 4- Comparar F com valor de F crítico da tabela

Análise de Variância

- As médias dos quatro são iguais?
  - 1- Calcular variância **entre** amostras (***between***)
    - Número de observações x Variância das médias amostrais

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \bar{x}_j}{k} = \frac{15{,}2 + 15 + 15{,}4 + 15{,}6}{4} = 15{,}3$$

$$s_{\bar{x}}^2 = \left[\frac{\sum_{j=1}^{k} \left(\bar{x}_j - \bar{\bar{x}}\right)^2}{k-1}\right] = \frac{(15{,}2 - 15{,}3)^2 + (15 - 15{,}3)^2 + (15{,}4 - 15{,}3)^2 + (15{,}6 - 15{,}3)^2}{4-1}$$

$$= 0{,}067$$

$$s_b^2 = n \cdot s_{\bar{x}}^2 = 6 \cdot 0{,}067 = 0{,}402$$

Análise de Variância

- As médias dos quatro são iguais?
  - 2- Calcular variância **dentro** das amostras (***within***)
    - Médias das variâncias das amostras

$$s_w^2 = \frac{s_1^2 + s_2^2 + s_3^2 + s_4^2}{k} = \frac{0{,}088 + 0{,}040 + 0{,}124 + 0{,}032}{4} = \frac{0{,}284}{4} = 0{,}071$$

Análise de Variância

- As médias dos quatro são iguais?
  - 3- Calcular a Razão F

$$F = \frac{s_b^2}{s_w^2} = \frac{n \cdot s_{\bar{x}}^2}{\dfrac{s_1^2 + s_2^2 + s_3^2 + s_k^2}{k}} = \frac{6 \cdot 0{,}067}{\dfrac{0{,}284}{4}} = \frac{0{,}402}{0{,}071} = 5{,}662$$

Análise de Variância

- As médias dos quatro são iguais?
  - 4- Comparar F com valor da tabela

- Distribuição F e

- Tabela de Probabilidades F

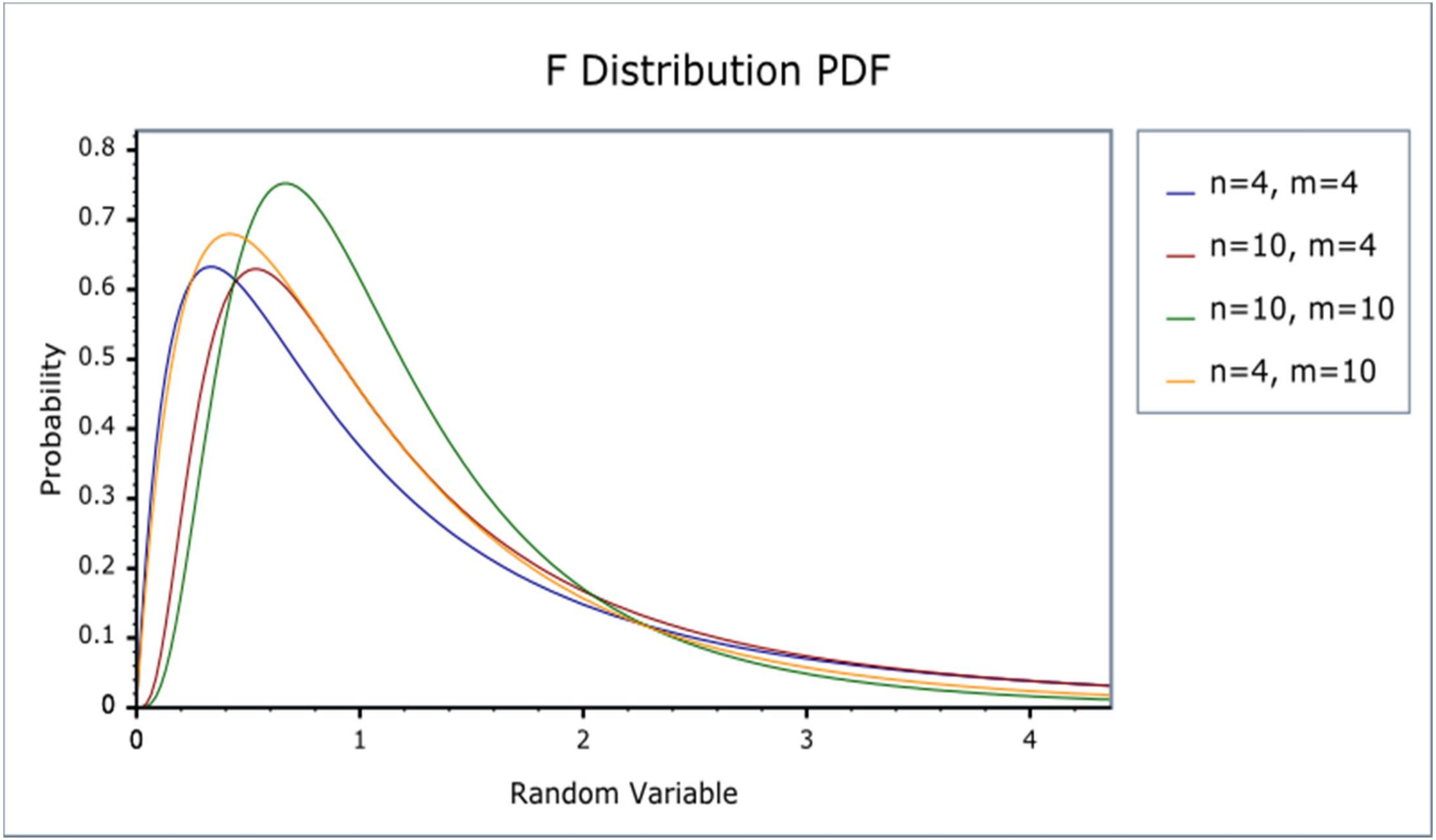
Análise de Variância

- Distribuição F
  - Distribuição F é assimétrica
    - Valor mínimo: 0
    - Sem valor máximo
  - Curva atinge pico próximo à origem
  - Curva aproxima-se do eixo x à medida que F aumenta
  - Distribuição F
    - Dois “tipos” de Graus de Liberdade
      - gl1 = df1 = n1 = GL do numerador
      - gl2 = df2 = n2 = GL do denominador

Análise de Variância

- Distribuição F
  - Uma distribuição F para cada combinação
    - Tamanho de amostra (n) e
    - Número de amostras (k)

    - Há uma distribuição F distinta, por exemplo, para:
      - 3 amostras (k) extraídas de seis observações (n)
      - 5 amostras (k) extraídas de seis observações (n)
      - 6 amostras (k) extraídas de sete observações (n)
  - Tabela de Probabilidades F
  - Tabulam-se valores mais usados


F Distribution PDF
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Probability
0
1
2
3
4
Random Variable
n=4, m=4
n=10, m=4
n=10, m=10
n=4, m=10

Análise de Variância

- Tabela de Probabilidades F
  - Uma distribuição F para cada combinação
    - Tamanho de amostra (n observações) e
    - Número de amostras (k amostras)
  - Distribuição contínua no intervalo [0; +∞]
  - Como os termos da Razão F são ao quadrado, não há valor negativo de F
  - Forma de cada distribuição amostral teórica F depende do número de Graus de Liberdade
    - gl1 = df1 = n1 = GL do numerador; e
    - gl2 = df2 = n2 = GL do denominador

Análise de Variância

- Tabela de Probabilidades F
  - Determinação dos Graus de Liberdade
    - gl1 = df1 = n1 = GL do numerador; e
    - gl2 = df2 = n2 = GL do denominador
  - Determinado pelos cálculos necessários para deduzir cada estimativa da variância populacional
    - Estimativa entre (*between*)
    - Estimativa dentro (*within*)

Análise de Variância

- Tabela de Probabilidades F
  - Cálculos necessários para cada estimativa da variância populacional
    - **Estimativa entre (*between*): numerador**
      - Divisão da soma de quadrados de diferenças pelo número de médias amostrais (k) menos 1
      - gl1 = df1 = n1 = GL do numerador = ***(k – 1)***

$$s_b^2 = n \cdot s_{\bar{x}}^2 = n \cdot \left[\frac{\sum_{j=1}^{k} \left(\bar{x}_j - \bar{\bar{x}} \right)^2}{k - 1} \right]$$

Análise de Variância

- Cálculos necessários para cada estimativa da variância populacional
  - **Estimativa dentro (*within*): denominador**
    - Cada variância amostral resulta da divisão da soma de quadrados de diferenças pelo número de observações (n) menos 1: ***(n – 1)***

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

$$s_w^2 = \frac{\frac{\sum_{i=1}^{n}(x_i - \bar{x}_1)^2}{n-1} + \frac{\sum_{i=1}^{n}(x_i - \bar{x}_2)^2}{n-1} + \dots \frac{\sum_{i=1}^{n}(x_i - \bar{x}_k)^2}{n-1}}{k}$$

Análise de Variância

- Cálculos necessários para cada estimativa da variância populacional
  - **Estimativa dentro (*within*) – denominador**
    - A Média das variâncias obtém-se pelo quociente entre a soma das variâncias amostrais e o número de amostras ***(k)***.
    - O número de Graus de Liberdade da estimativa dentro (*within*) é então: ***k(n - 1)***
    - gl2 = df2 = n2 = GL do denominador = ***k(n - 1)***

Análise de Variância

- gl2 = df2 = n2 = GL do denominador = ***k(n - 1)***

$$s_w^2 = \frac{\frac{\sum_{i=1}^{n}(x_i - \bar{x}_1)^2}{n-1} + \frac{\sum_{i=1}^{n}(x_i - \bar{x}_2)^2}{n-1} + \cdots \frac{\sum_{i=1}^{n}(x_i - \bar{x}_k)^2}{n-1}}{k}$$

$$s_w^2 = \left(\frac{1}{n-1}\right)\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x}_1)^2 + \sum_{i=1}^{n}(x_i - \bar{x}_2)^2 + \cdots + \sum_{i=1}^{n}(x_i - \bar{x}_k)^2}{k}\right]$$

$$s_w^2 = \frac{1}{k(n-1)}\left[\sum_{i=1}^{n}(x_i - \bar{x}_1)^2 + \sum_{i=1}^{n}(x_i - \bar{x}_2)^2 + \cdots + \sum_{i=1}^{n}(x_i - \bar{x}_k)^2\right]$$

Análise de Variância

- Tabela F
  - Uma tabela para cada nível de significância α
    - Algumas tabelas F tabulam valores de F para mais de um nível de significância α
  - **Linha superior**
    - **GL do numerador – estimação entre (*between*)**
      - gl1 = df1 = n1 = GL do numerador = ***(k – 1)***
  - **Coluna à esquerda**
    - **GL do denominador – estimação dentro (*within*)**
      - gl2 = df2 = n2 = GL do denominador = ***k(n – 1)***

Análise de Variância

- Tabela F
  - Para uma tabela de nível de significância α
    - Valores de gl1 ***(k – 1)*** e gl2 ***[k(n – 1)]*** determinam
      - Valor F na célula específica
  - Valor da célula
    - Valor Crítico de F (limite região de aceitação de H_0)
    - Linha, ou valor, limítrofe para o teste de variação entre as médias comparadas
      - Variação devida ao acaso
      - Variação Não devida ao acaso

Análise de Variância

Significação em uma cauda $\mathbf{F}_{\mathbf{sig\ (0,05)}}$
n1 = Graus de Liberdade do Numerador de F (variância entre/between)
n2 = Graus de Liberdade do Denominador de F (variância dentro/within)

	df1 = GL Numerador = (k -1) (variância entre/between)					
df2 = k(n – 1)	***1***	***2***	***3***	***4***	***5***	***6***
1	161	199	216	225	230	234
2	18,5	19	19,2	19,2	19,3	19,3
3	10,1	9,55	9,28	9,12	9,01	8,94
4	7,71	6,94	6,59	6,39	6,26	6,16
5	6,61	5,79	5,41	*5,19*	5,05	4,95
6	5,99	5,14	4,76	4,53	4,39	4,28
7	5,59	4,74	4,35	4,12	3,97	3,87
8	5,32	4,46	*4,07*	3,84	3,69	3,58
9	5,12	4,26	3,86	3,63	3,48	3,37

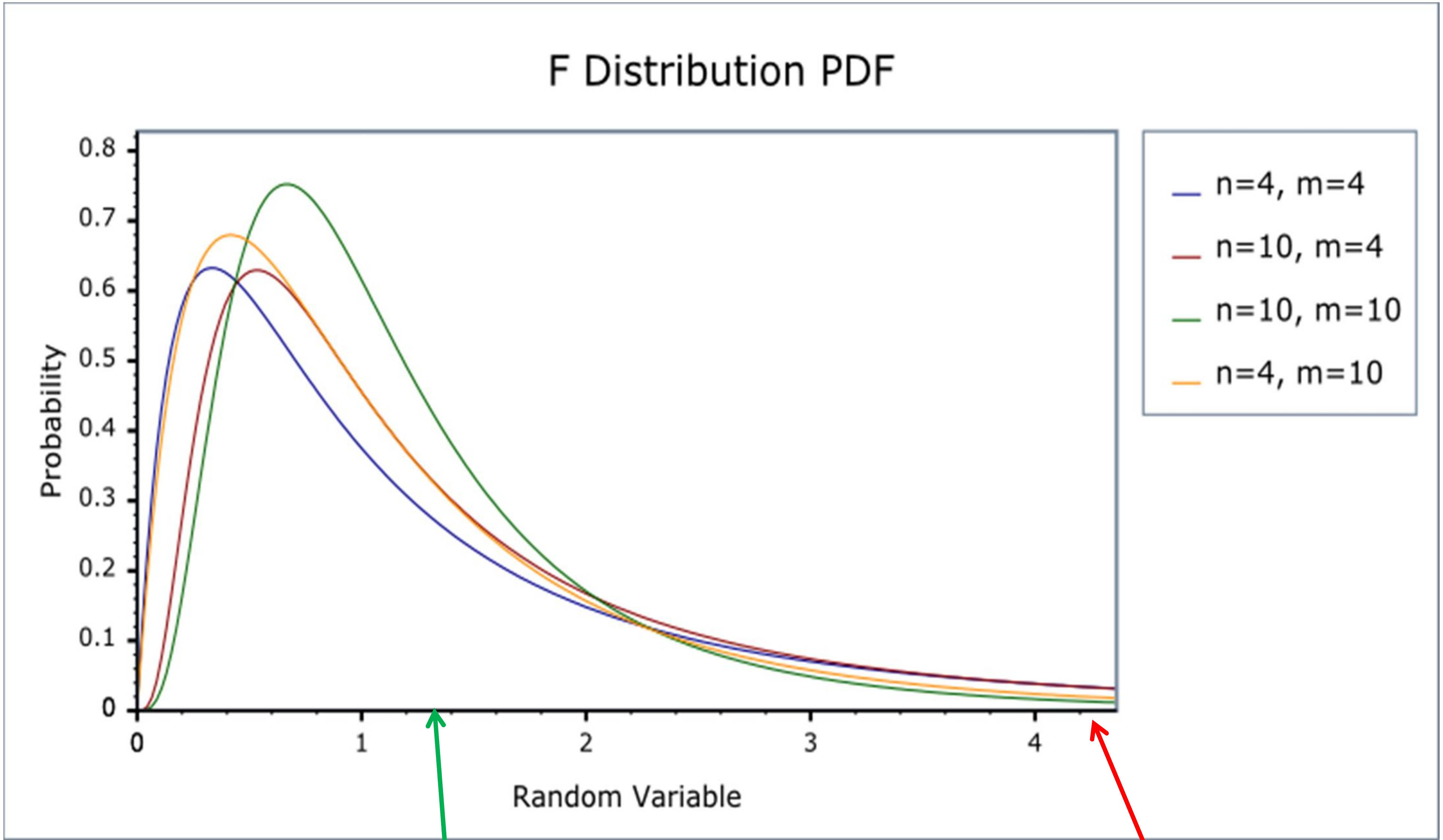
Análise de Variância

- Uso da Tabela F
  - Para as seguintes situações, ver o valor de F ao nível de significância (α) de 5% (0,05)

Número de amostras (k)	GL Numerador (k - 1)	Tamanho amostra (n)	GL Denominador [k(n - 1)]	Valor de F
5	***4***	2	***5***	**5,19**
4	***3***	3	***8***	**4,07**
6	***5***	2	***6***	**4,39**
2	***1***	5	***8***	**5,32**
3	***2***	4	***9***	**4,26**
2	***1***	11	***20***	**4,35**

Análise de Variância

- Hipóteses
  - **H_0: As médias das populações são iguais**
  - **H_1: As médias das populações não são iguais**



Estatística amostral (F) muito ***PROVÁVEL*** se H0 é verdadeira

Estatística amostral (F) muito ***IMPROVÁVEL*** se H0 é verdadeira

Análise de Variância

- Tabela F
  - Para certa Razão, ou teste, F calculada(o)
    - Sob determinado nível de significância (α)

    - Se F calculado é inferior ao valor tabelado para gl1 e gl2
      - Não se descarta H0
        - Médias são iguais

    - Se F calculado é superior ao valor tabelado para gl1 e gl2
      - Rejeita-se H0 e aceita-se H1
        - Médias não são iguais

Análise de Variância

- Voltando ao exemplo
  - As médias dos quatro grupos são iguais?
  - 3- Calcular a Razão F

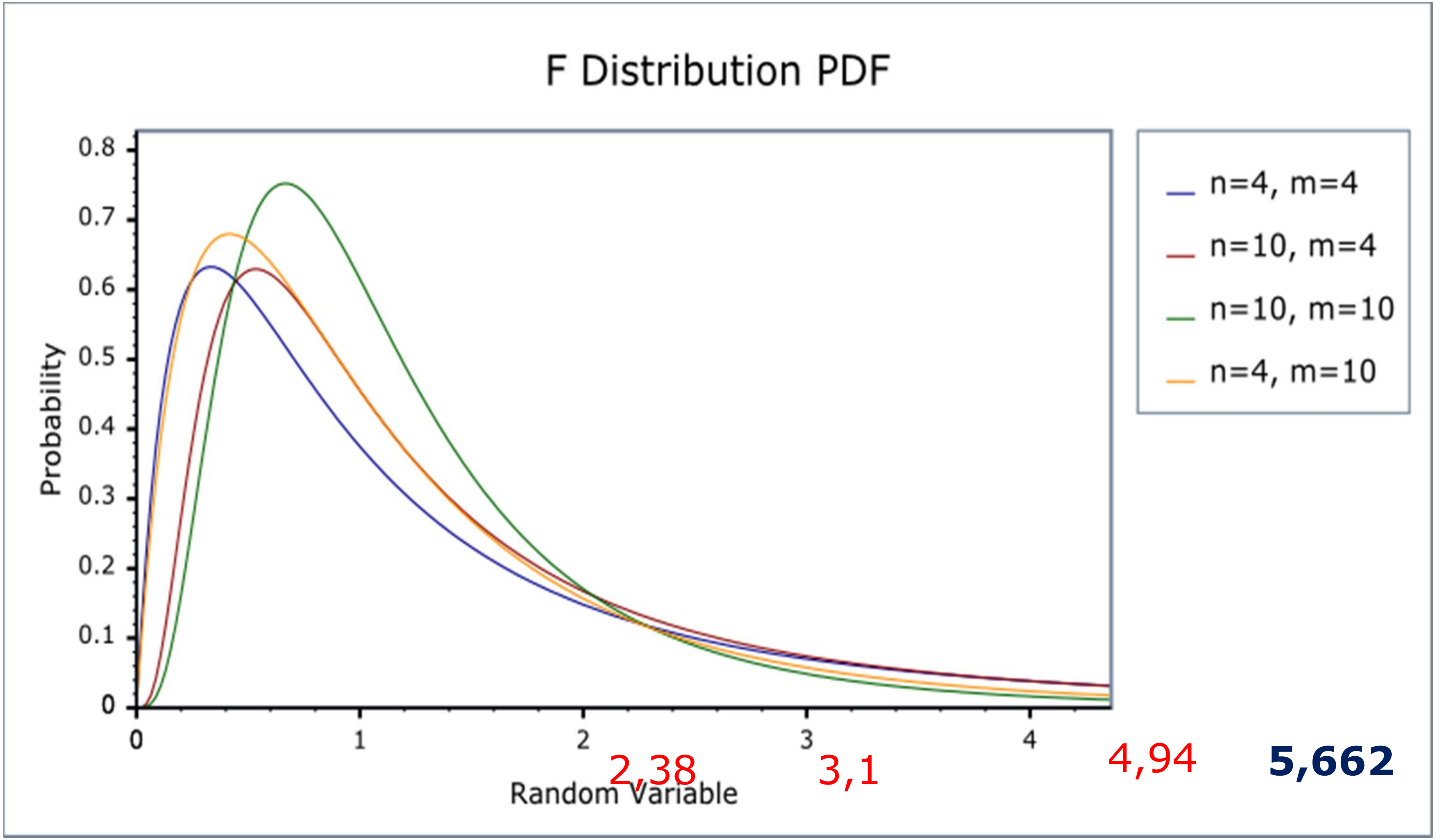
$$F = \frac{s_b^2}{s_w^2} = \frac{n \cdot s_{\bar{x}}^2}{\frac{s_1^2 + s_2^2 + s_3^2 + s_k^2}{k}} = \frac{6 \cdot 0{,}067}{\frac{0{,}284}{4}} = \frac{0{,}402}{0{,}071} = 5{,}662$$

- Quatro amostras de seis observações cada
  - Calcular GL
    - Número de amostras (k) = 4
      - GL Numerador (k - 1) = 4 - 1 = ***3 (GL1 = df1 = n1)***
    - Tamanho amostra (n) = 6
      - GL Denominador [k(n - 1)] = 4 * (6 – 1) = ***20 (GL2)***

Número de amostras (k)	**GL Numerador (k - 1)**	Tamanho amostra (n)	**GL Denominador [k(n - 1)]**	Valor de F (0,10)
4	**3**	6	**20**	***2,38***

Número de amostras (k)	**GL Numerador (k - 1)**	Tamanho amostra (n)	**GL Denominador [k(n - 1)]**	Valor de F (0,05)
4	**3**	6	**20**	***3,1***

Número de amostras (k)	**GL Numerador (k - 1)**	Tamanho amostra (n)	**GL Denominador [k(n - 1)]**	Valor de F (0,01)
4	**3**	6	**20**	***4,94***


F Distribution PDF
Probability
Random Variable
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
n=4, m=4
n=10, m=4
n=10, m=10
n=4, m=10
2,38
3,1
4,94
5,662

Análise de Variância

- *F calculado* = ***5,662***
  - Supera F crítico ao nível de 0,01
    - Muito significativo

    - F calculado é superior a valor F tabelado
      - Rejeita-se H0 e aceita-se H1
        - Médias dos quatro grupo são diferentes
        - Amostras não são oriundas da mesma população

Análise de Variância

- Resumo:

1- Calcular variância **entre** amostras (***between***) - Numerador

- Número de observações (n) X Variância das médias amostrais

$$s^2_{entre/between} = s^2_b = n \cdot s^2_{\bar{x}}$$

2- Calcular variância **dentro** das amostras (***within***) - Denominador

- Médias das variâncias das amostras

$$s^2_w = \frac{s^2_1 + s^2_2 + s^2_3 + \cdots + s^2_k}{k}$$

3- Calcular a Razão F

$$Razão\ F = \frac{s^2_b}{s^2_w} = \frac{n \cdot s^2_{\bar{x}}}{\frac{s^2_1 + s^2_2 + s^2_3 + \cdots + s^2_k}{k}}$$

4- Comparar F calculado com F crítico da tabela

Análise de Variância

- Se amostras de tamanhos distintos
- Ajuste das fórmulas
  - Estimativa entre (*between*)

$$s_b^2 = \frac{n_1(\bar{x}_1 - \bar{\bar{x}})^2 + n_2(\bar{x}_2 - \bar{\bar{x}})^2 + \cdots + n_k(\bar{x}_k - \bar{\bar{x}})^2}{k-1}$$

  - Estimativa dentro (*within*)
    - gl2 = df2 = $(n_1 + n_2 + ... + n_k) - k$

$$s_w^2 = \frac{\frac{\sum_{i=1}^{n_1}(x_i - \bar{x}_1)^2}{n_1 - 1} + \frac{\sum_{i=1}^{n_2}(x_i - \bar{x}_2)^2}{n_2 - 1} + \cdots \frac{\sum_{i=1}^{n_k}(x_i - \bar{x}_k)^2}{n_k - 1}}{k}$$