Universidade Federal do Paraná
Campus Avançado de Jandaia do Sul
Estatística

Análise de Variância (ANOVA)

Dr. Landir Saviniec

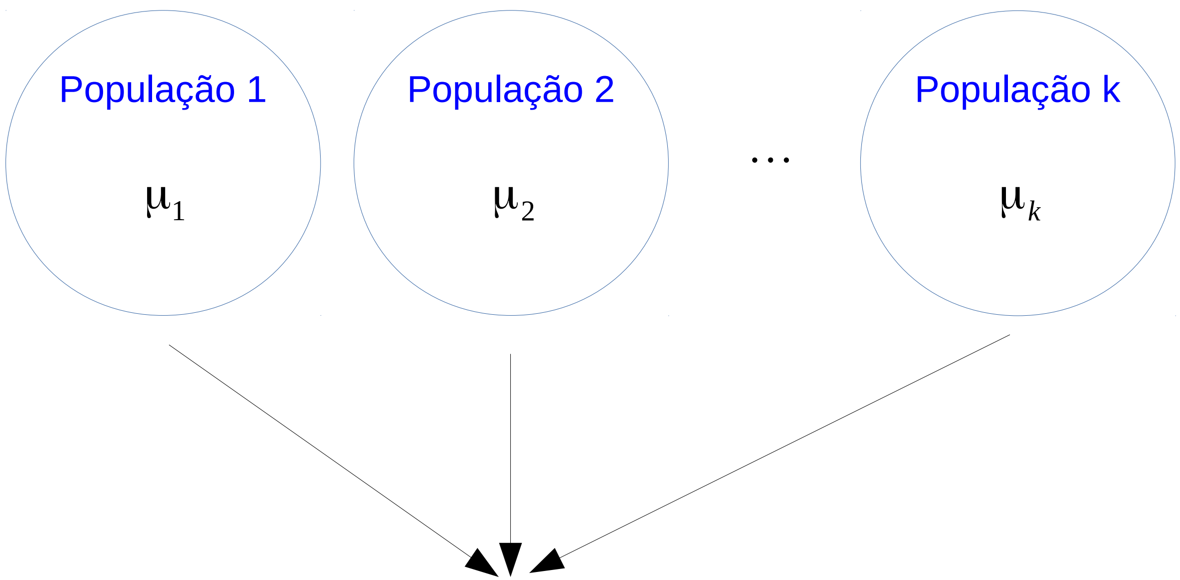
E-mail: landir.saviniec@gmail.com
Homepage: github.com/lansaviniec

Novembro de 2018

ANOVA de um fator

Análise de variância

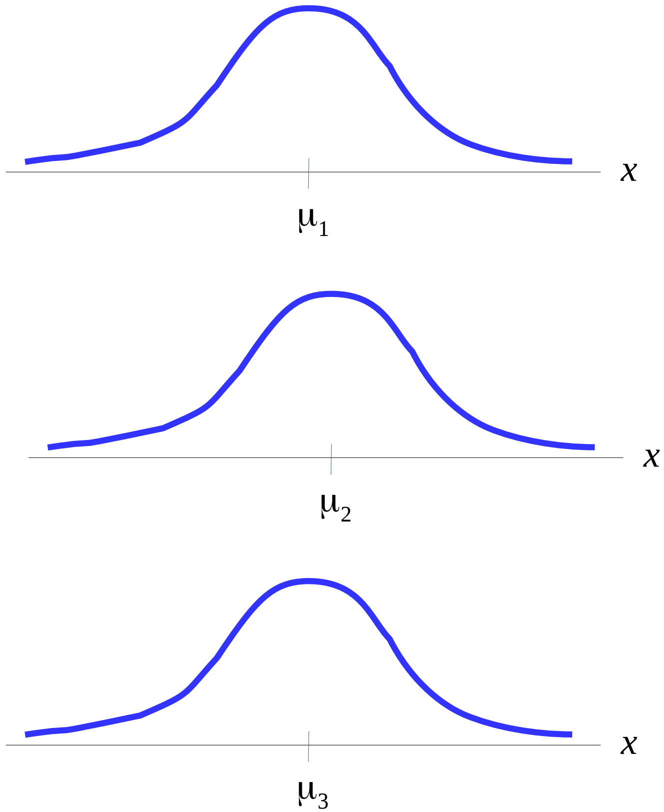
É um **método para teste de igualdade de três ou mais médias populacionais** através da **análise das variâncias amostrais**.



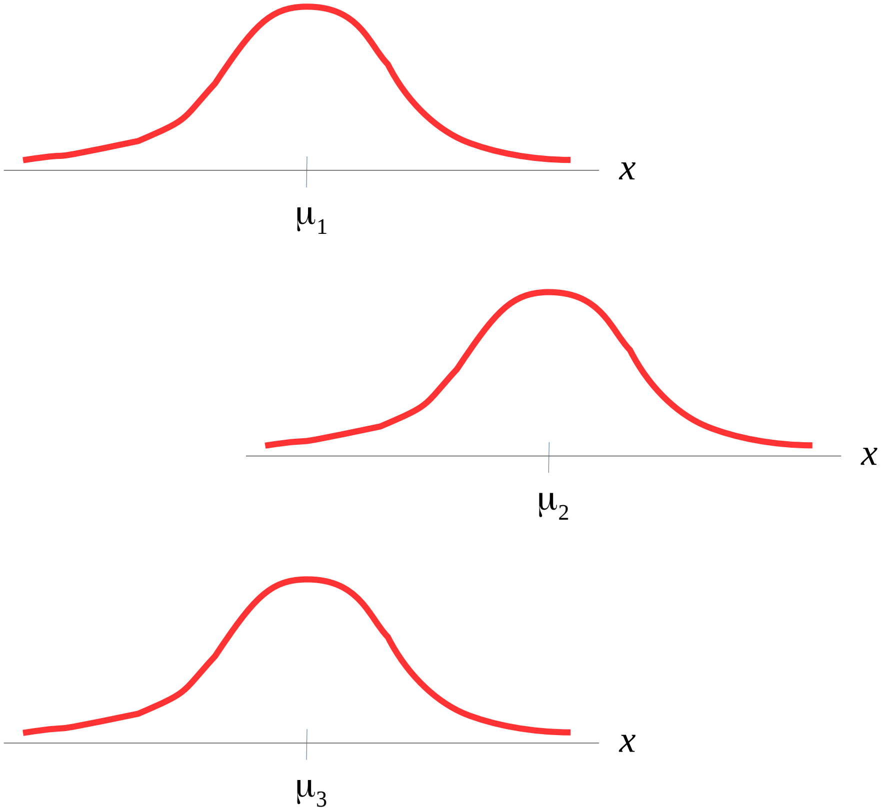
$$H_0: \quad \mu_1 = \mu_2 = \cdots = \mu_k \qquad \text{(as médias são todas iguais)}$$
$$H_1: \quad \text{Pelo menos uma das médias é diferente das demais}$$

Exemplo

Médias aproximadamente iguais



Pelo menos uma média difere das demais



Um fator

Usa-se apenas **uma característica para comparar as populações**.

Exemplo: Suponha as populações de pessoas de três países A, B e C. **Uma única característica (fator)** seria considerar apenas a **altura** das pessoas. **Duas características (fatores)** seria considerar **altura e peso** das pessoas.

Objetivo

Testar se todas as médias populacionais são iguais ou se **pelo menos uma difere das demais,** de forma significativa.

Requisitos

1) As populações devem ter distribuições que são aproximadamente normais. O método funciona bem, a não ser que as distribuições se afastem muito da normal.

2) As populações devem ter variâncias aproximadamente iguais.

3) As amostras devem ser amostras aleatórias simples e independentes.

Estatística de teste utilizada

Calculamos a estatística de teste baseada na **distribuição F**:

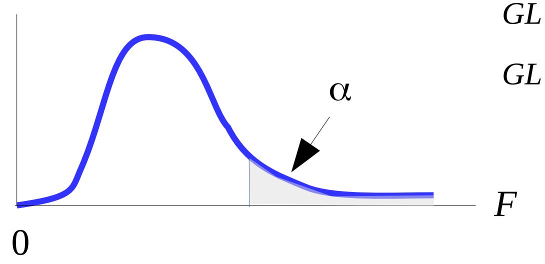
$$F = \frac{\text{variância entre amostras}}{\text{variância dentro das amostras}}$$

Distribuição F

Se extrairmos amostras A e B aleatórias de uma mesma população e calcularmos o escore F abaixo:

$$F = \frac{s_A^2}{s_B^2} \quad , \quad com\ s_A^2 \geq s_B^2$$

Os valores de F seguem uma distribuição assimétrica à direita:



$$GL(numerador) = n_A - 1$$

$$GL(denominador) = n_B - 1$$

Exemplo 1: com amostras de mesmo tamanho

Considere a população de dados {3, 4, 5, 6, 7, 8}:

Todas as amostras são da população dada

Amostra 1	Amostra 2	Amostra 3
7	6	4
3	5	7
6	5	6
6	8	7

$n_1 = 4$, $\bar{x}_1 = 5.5$, $s_1^2 = 3.0$

$n_2 = 4$, $\bar{x}_2 = 6.0$, $s_2^2 = 2.0$

$n_3 = 4$, $\bar{x}_3 = 6.0$, $s_3^2 = 2.0$

variância entre amostras :

$$s_{\bar{x}}^2 = 0.0833$$

$$n\, s_{\bar{x}}^2 = 4\,(0.0833) = 0.3332$$

variância dentro das amostras :

$$s_p^2 = \frac{3.0 + 2.0 + 2.0}{3} = 2.3333$$

Estatística de teste F :

$$F = \frac{n\, s_{\bar{x}}^2}{s_p^2} = \frac{0.3332}{2.3333} = 0.1428$$

Uma das amostras não é da população dada

Amostra 1	Amostra 2	Amostra 3
17	6	4
13	5	7
16	5	6
16	8	7

$n_1 = 4$, $\bar{x}_1 = 15.5$, $s_1^2 = 3.0$

$n_2 = 4$, $\bar{x}_2 = 6.0$, $s_2^2 = 2.0$

$n_3 = 4$, $\bar{x}_3 = 6.0$, $s_3^2 = 2.0$

variância entre amostras :

$$s_{\bar{x}}^2 = 30.0833$$

$$n\, s_{\bar{x}}^2 = 4\,(30.0833) = 120.3332$$

variância dentro das amostras :

$$s_p^2 = \frac{3.0 + 2.0 + 2.0}{3} = 2.3333$$

Estatística de teste F :

$$F = \frac{n\, s_{\bar{x}}^2}{s_p^2} = \frac{120.3332}{2.3333} = 51.5721$$

Exemplo 1: com amostras de mesmo tamanho

Considere a população de dados {3, 4, 5, 6, 7, 8}:

Todas as amostras são da população dada

Amostra 1	Amostra 2	Amostra 3
7	6	4
3	5	7
6	5	6
6	8	7

Uma das amostras não é da população dada

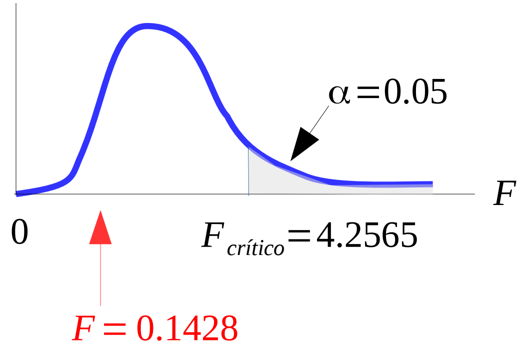
Amostra 1	Amostra 2	Amostra 3
17	6	4
13	5	7
16	5	6
16	8	7

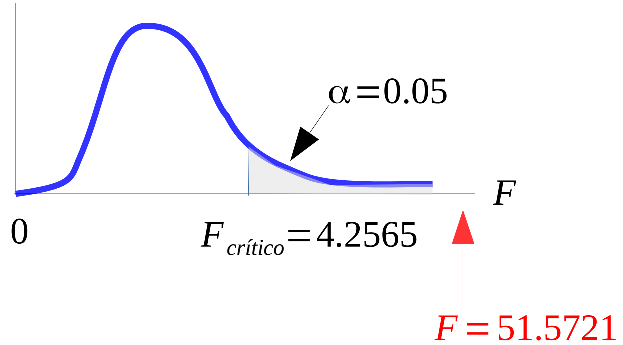
$k=3$ *quantidade de amostras*

$n=4$ *tamanho das amostras*

$GL(numerador) = k-1 = 3-1 = 2$

$GL(denominador) = k(n-1) = 3(4-1) = 9$





Aceita-se H_0. As amostras são de populações que tem as mesmas médias.

Rejeita-se H_0. Pelo menos uma amostra é de uma população com média diferente das demais.

De onde surgiu a fórmula para calcular a variância entre as amostras?

Pelo teorema do limite central, temos:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad \sigma_{\bar{x}}^2 = \left(\frac{\sigma}{\sqrt{n}}\right)^2 \quad \Rightarrow \quad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \quad \Rightarrow \quad \sigma^2 = n\,\sigma_{\bar{x}}^2$$

Logo, a **variância entre amostras** ou variâncias das médias amostrais é dada por:

$$variância\ entre\ amostras = n\ s_{\bar{x}}^2$$

Onde:

$$A\ variância\ populacional\ \sigma_{\bar{x}}^2\ e\ estimada\ pela\ variância\ das\ médias\ amostrais\ s_{\bar{x}}^2$$

Exemplo 2

As amostras a seguir apresentam valores de compras de clientes de três lojas A, B e C de uma mesma rede de lojas.

A: 350, 380, 450, 500, 555, 540, 590, 602, 620, 630
B: 520, 405, 540, 580, 610, 650, 555, 590, 645, 615
C: 800, 750, 770, 830, 850, 860, 790, 775, 815, 845

Aplique um teste de hipótese com nível de significância de 5% para verificar se o **gasto médio por compra é o mesmo nas três lojas**.

$$\alpha = 0.05 \qquad n = 10$$
$$\bar{x}_A = 521.7 \qquad s_A^2 = 9902.23$$
$$\bar{x}_B = 571.0 \qquad s_B^2 = 5221.11$$
$$\bar{x}_C = 808.5 \qquad s_C^2 = 1405.83$$

Passo 1: Formulamos as hipóteses nula e alternativa:

$$H_0: \quad \mu_1 = \mu_2 = \mu_2 \qquad \left(\textit{as médias são todas iguais} \right)$$
$$H_1: \quad \textit{Pelo menos uma das médias é diferente das demais}$$

Exemplo 2

Passo 2: Calculamos a estatística de teste para os dados amostrais:

variância entre amostras :

$$s_{\bar{x}}^2 = 23515.16$$

$$n\, s_{\bar{x}}^2 = 10\,(23515.16) = 235151.6$$

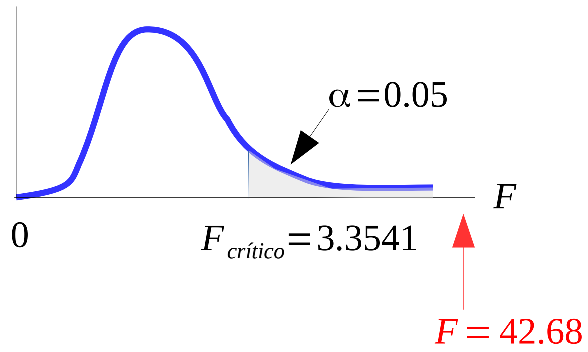
variância dentro das amostras :

$$s_p^2 = \frac{9902.23 + 5221.11 + 1405.83}{3} = 5509.73$$

Estatística de teste F :

$$F = \frac{n\, s_{\bar{x}}^2}{s_p^2} = \frac{235151.6}{5509.73} = 42.68$$

Passo 3: Analisamos os valores críticos na distribuição F:



$k = 3$ *quantidade de amostras*

$n = 10$ *tamanho das amostras*

$$GL(numerador) = k - 1 = 3 - 1 = 2$$

$$GL(denominador) = k(n-1) = 3(10-1) = 27$$

Exemplo 2

Passo 4: Decidimos se aceitamos ou rejeitamos a hipótese nula:

Como a estatística de teste caiu na região de rejeição, rejeitamos a hipótese nula. Logo, há evidências de que em pelo menos uma das lojas o gasto médio por compra é diferente das demais.