



## Plano Aula 17 e 18

Markus Stein

### Inferência para várias populações (Bussab e Morettin - seção 15.1)

#### Análise de Variância - ANOVA

Pode ser vista como um teste para três ou mais médias.

- **Exemplo:** Estudar o retorno de **três ou mais** ações durante um dia de operações. Será que existe diferença significativa entre os retornos médios das ações?
- **Exemplo 2:** Gastos em campanha. Suponha que estamos interessados em comparar os gastos diários na campanha para eleições municipais entre três candidatos diferentes. Podemos dizer que os gastos diários médios nas campanhas são iguais?

#### ... aula passada... Distribuição $F$

Teorema (**Distribuição  $F$ , nossa versão**): Seja  $X_1, \dots, X_{n_1}$  uma amostra aleatória da v.a.  $X \sim Normal(\mu_1, \sigma_1^2)$  e  $S_1^2 = \sum_{i=1}^{n_1} (X_i - \bar{X})^2 / (n_1 - 1)$ . Da mesma forma, suponha uma outra amostra  $Y_1, \dots, Y_{n_2}$  uma amostra aleatória da v.a.  $Y \sim Normal(\mu_2, \sigma_2^2)$  e  $S_2^2 = \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2 / (n_2 - 1)$ . Então podemos escrever uma quantidade  $F$  tal que (dadas algumas outras suposições que omitimos aqui)

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1).$$

- Como usar a distribuição de  $F$  para testar médias de duas ou mais populações? **Quais as suposições necessárias? Como interpretar os resultados?**

#### Soma de Quadrados (SQ)

Separamos a soma de quadrados no numerador da variância de uma variável  $Y$  em duas partes:

- uma representará as distâncias entre as médias dos **grupos** - **SQG**;
- a outra as distâncias (ou **resíduos**) entre as observações e a média do seu grupo - **SQR**.
  - Ambas quantidades são estimadas com base nos dados observados.
  - (dadas as suposições necessárias) uma estatística do tipo  $F$  pode ser calculada e interpretada.

#### Modelo Estatístico (Bussab e Morettin - seção 15.3)

#### Tabela de ANOVA

- Exemplo de tabela de análise de variância para duas populações (Bussab e Morettin - seção 15.2.4)



### Percentual de variância explicada (pelo modelo)

Coeficiente de explicação do modelo:  $R^2 = \frac{SQG}{SQT}$ .

### Comparação entre as médias

- Usando intervalos de confiança
  - Exemplos 15.5 e 15.6 de Bussab e Morettin - seção 15.4.
- no slides usamos o limite de decisão  $L_d = 3 \times s_{\bar{x}}$  ( $s_{\bar{x}}$  é o desvio padrão das médias)

### REFERÊNCIA EXTRA

Página ‘Probabilidade e Estatística (EaD)’ da UFRGS

- Capítulo 5 - Inferência para dados numéricos
  - Seção - Comparando muitas médias com ANOVA

---

Ler slides das aulas 17 e 18

Continuar exercícios lista 2-3

Fazer avaliação pontual 2 da área 2 - vale nota!!!

---