

# Plano de Análise Estatística de ...

SAP: analise\_dados\_DM\_2020-v01

De: Felipe Figueiredo Para: Diego Rezende Martins

Data: 07/01/2021

## SUMÁRIO

|   |   |
|---|---|
| SUMÁRIO.....                                      | 1 |
| 1 LISTA DE ABREVIATURAS .....                     | 2 |
| 2 INTRODUÇÃO .....                                | 2 |
| 2.1 Contexto .....                                | 2 |
| 2.2 Objetivos .....                               | 2 |
| 2.3 Hipóteses.....                                | 2 |
| 3 LIMPEZA DOS DADOS .....                         | 2 |
| 4 VARIÁVEIS DO ESTUDO.....                        | 3 |
| 4.1 Desfechos primário e secundários .....        | 3 |
| 4.2 Covariáveis .....                             | 3 |
| 5 MÉTODOS ESTATÍSTICOS .....                      | 4 |
| 5.1 Análises estatísticas .....                   | 4 |
| 5.1.1 Análise descritiva .....                    | 4 |
| 5.1.2 Análise inferencial.....                    | 4 |
| 5.2 Significância e Intervalos de Confiança ..... | 4 |
| 5.3 Tamanho da amostra e Poder .....              | 4 |
| 5.4 Softwares utilizados.....                     | 5 |
| 6 REFERÊNCIAS.....                                | 5 |

## Histórico do documento

| Versão | Alterações     |
|--------|----------------|
| 01     | Versão inicial |

## 1 LISTA DE ABREVIATURAS

- DM: Diabetes mellitus
- EVD:
- HAS: Hipertensão arterial sistêmica
- HT: Hipotireoidismo
- IVC:

## 2 INTRODUÇÃO

### 2.1 Contexto

### 2.2 Objetivos

### 2.3 Hipóteses

## 3 LIMPEZA DOS DADOS

Os dados recebidos serão reorganizados em uma tabela retangular, agrupando todos os casos em linhas com uma coluna por variável. A coluna ID será redefinida de modo que haja um ID único para cada participante incluído no estudo, independente do grupo. Com isso, a tabela de dados analíticos apresentará uma linha para cada observação das variáveis, i.e., cada um dos vinte participantes incluídos no estudo.

As seguintes novas variáveis serão criadas a partir para a análise:

- **EVD dif:** diferença entre EVD pré e pós (numérica)
- **Dominante:** se o lado acometido era o lado dominante do participante (dicotômica)

As seguintes variáveis serão limpas para adequação nas análises:

- **Tempo artrose:** convertida em valores numéricos
  - valores possíveis: número (anos) ou célula vazia
- **Comorbidades:** redefinidas como variáveis dicotômicas
  - reorganizadas em colunas independentes
  - valores possíveis: nome da comorbidade ou célula vazia
- **Retorno:** Esta coluna é repetida em todas as quatro planilhas
  - serão utilizados as duas primeiras ocorrências deste dado:
    - Folha 1 - Tabela 1 - Volar
    - Folha 1 - Tabela 2 - Dorsal

A tabela de dados analíticos será criada de acordo com a seguinte estrutura da Tabela 1.

**Tabela 1** Estrutura da nova tabela de dados analíticos

| ID  | Idade | Sexo | Grupo | EVD Pré | EVD Pós | EVD dif | Lado dominante | Lado acometido | Dominante | Artrose | Tempo artrose | Satisfação | Retorno |
|-----|-------|------|-------|---------|---------|---------|----------------|----------------|-----------|---------|---------------|------------|---------|
| 1   |       |      |       |         |         |         |                |                |           |         |               |            |         |
| 2   |       |      |       |         |         |         |                |                |           |         |               |            |         |
| 3   |       |      |       |         |         |         |                |                |           |         |               |            |         |
| ... |       |      |       |         |         |         |                |                |           |         |               |            |         |
| 20  |       |      |       |         |         |         |                |                |           |         |               |            |         |

As colunas de comorbidades serão acrescentadas à tabela de dados, seguindo a estrutura apresentada na Tabela 2.

**Tabela 2** Representação das comorbidades na tabela de dados analíticos

| ID  | (colunas tabela 1...) | HAS | DM | TABAGISMO | HT | IVC |
|-----|-----------------------|-----|----|-----------|----|-----|
| 1   |                       |     |    |           |    |     |
| 2   |                       |     |    |           |    |     |
| 3   |                       |     |    |           |    |     |
| ... |                       |     |    |           |    |     |
| 20  |                       |     |    |           |    |     |

## 4 VARIÁVEIS DO ESTUDO

### 4.1 Desfechos primário e secundários

Serão avaliados quatro desfechos na comparação entre os dois grupos de participantes:

#### desfecho primário

- Diferença na EVD pré e pós nos dois grupos (numérica)

#### desfechos secundários

- Alteração degenerativa (artrose) (dicotômica)
- Grau de satisfação (numérico)
- Retorno à profissão (dicotômica)

### 4.2 Covariáveis

- Idade (em anos)
- Sexo (feminino/masculino)
- Acometimento no lado dominante (direito/esquerdo)
- Tempo até a artrose (em anos)
- Comorbidades (dicotômicas)
  - HAS
  - DM
  - Tabagismo
  - HT
  - IVC

## **5 MÉTODOS ESTATÍSTICOS**

### **5.1 Análises estatísticas**

#### **5.1.1 Análise descritiva**

As variáveis numéricas serão descritas como média e desvio padrão. As variáveis categóricas serão descritas como frequência e proporção. As proporções serão apresentadas como porcentagem.

A análise descritiva será dividida em duas tabelas:

- tabela das características dos participantes do estudo, apresentando as covariáveis descritas na seção 4.2
- tabela descritiva dos desfechos do estudo descritos na seção 4.1

Ambas as tabelas apresentarão os dados estratificados nos dois grupos.

#### **5.1.2 Análise inferencial**

A comparação de variáveis numéricas entre os dois grupos será avaliada com o teste t de Student não pareado. O p-valor do teste t de student será apresentado na tabela descritiva dos resultados

A comparação de variáveis categóricas entre os dois grupos será avaliada com o teste exato de Fisher.

### **5.2 Significância e Intervalos de Confiança**

Todas as análises serão feitas com nível de significância de 5%. Todos os testes de significância e intervalos de confiança calculados serão bilaterais.

### **5.3 Tamanho da amostra e Poder**

O desfecho primário deste estudo é uma variável numérica e será analisada com o teste t de Student. Na ausência de estimativas prévias da média e desvio padrão do desfecho primário (EVD) a análise de poder só pode ser feita usando-se o método de Cohen (Cohen, 1988) de tamanho de efeito padronizado. Esta abordagem utiliza apenas os dados obtidos na amostra de estudo e Cohen sugere a seguinte interpretação de tamanhos de efeito padronizados:

- efeito pequeno:  $d = 0.2$
- efeito médio:  $d = 0.5$
- efeito grande:  $d = 0.8$

Com 10 participantes incluídos em cada grupo, este estudo é capaz de detectar um tamanho de efeito grande ( $d = 0.8$ ) com poder de 39.5% e significância de 5%. Por outro lado, se estipularmos o nível de poder tipicamente encontrado na literatura de 80%, o efeito padronizado precisa ser pelo menos 1.32 para que este estudo possa detectá-lo com significância de 5%.

É pouco provável que esta amostra seja suficiente para detectar um efeito estatisticamente significativo entre os grupos.

#### **5.4 Softwares utilizados**

Esta análise será realizada utilizando-se o software R versão 4.0.3.

#### **6 REFERÊNCIAS**

- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.