

# Trabalho 1-Análise do Dataset *Iris*

Luiz Ledo Mota Melo Junior

August 13, 2024

## 1 Análise Descritiva

- Cálculo das estatísticas descritivas (média, mediana, moda, desvio padrão) para cada uma das variáveis numéricas (`Sepal.Length`, `Sepal.Width`, `Petal.Length`, `Petal.Width`).

- Exemplo:

```
summary(iris)
```

- Visualização das distribuições utilizando histogramas e boxplots.

- Exemplo:

```
hist(iris$Sepal.Length)
boxplot(iris$Sepal.Width ~ iris$Species)
```

## 2 Ajuste de Distribuição

- Ajuste da curva normal sobre os histogramas das variáveis utilizando a função `dnorm`.

- Exemplo:

```
hist(iris$Petal.Width, freq = FALSE, main = "Ajuste Normal")
curve(dnorm(x, mean = mean(iris$Petal.Width), sd = sd(iris$Petal.Width)),
      col = "red", add = TRUE)
```

## 3 Técnicas de *Clustering*

### 3.1 *K-Means*

- Aplicação do algoritmo *k-means* para agrupar as observações em clusters.
- Exemplo:

```
set.seed(123)
kmeans_result <- kmeans(iris[, 1:4], centers = 3)
```

- Visualização dos clusters com base nas variáveis mais significativas.

### 3.2 *Hierarchical Clustering (HC Clustering)*

- Aplicação do *hierarchical clustering* para identificar grupos hierárquicos.
- Exemplo:

```
hc_result <- hclust(dist(iris[, 1:4]))
plot(hc_result)
```

## 4 *Cross-Validation*

- Implementação de *cross-validation* para validar a robustez dos modelos de clustering.
- Exemplo:

```
library(caret)
train_control <- trainControl(method = "cv", number = 10)
kmeans_cv <- train(Species ~ ., data = iris,
  method = "kmeans",
  trControl = train_control)
```

- Avaliação da performance dos modelos através de métricas como *accuracy* e *confusion matrix*.