

## # Importando os dados

```
library(readxl)
milsa <- read_excel("C:/Users/Desktop/Aula R/MILSA.xlsx")
View(milsa)
```

```
# ----- #
```

## # MANIPULANDO ALGUMAS VARIÁVEIS

```
# ----- #
```

### #ESCOLARIDADE

#Como a variável escolaridade é qualitativa ordinal, podemos indicar para o R que ela deve ser tratada como ordinal.

```
milsa$ESCOLARIDADE <- factor(milsa$ESCOLARIDADE, ordered = TRUE)
levels(milsa$ESCOLARIDADE)
class(milsa$ESCOLARIDADE)
milsa$ESCOLARIDADE
```

### # CRIANDO A VARIÁVEL IDADE

```
milsa$IDADE <- milsa$IDADE_ANOS + (milsa$IDADE_MESES/12)
```

### #Resumindo os dois comandos acima

```
milsa <- transform(milsa,
  ESCOLARIDADE = factor(ESCOLARIDADE, ordered = TRUE),
  IDADE = IDADE_ANOS + (IDADE_MESES/12))
```

```
# ----- #
```

## # VARIÁVEL QUALITATIVA NOMINAL

```
# ----- #
```

### ## Frequência absoluta

```
civil_freq_absoluta <- table(milsa$ESTADO_CIVIL)
civil_freq_absoluta
```

### ## Frequência relativa, calculando manualmente

```
civil_freq_relativa <- civil_freq_absoluta/length(milsa$ESTADO_CIVIL)
civil_freq_relativa
```

### ## Frequência relativa, com a função prop.table()

```
prop.table(civil_freq_absoluta)
```

### ## Frequencia relativa acumulada

```
civil_freq_acum <- cumsum(civil_freq_relativa)
civil_freq_acum
```

```
library(summarytools)
```

```
freq(milsa$ESTADO_CIVIL, cumul = TRUE, totals = FALSE, order = "freq")
```

## #Gráficos

```
barplot(civil_freq_absoluta)
pie(civil_freq_absoluta)
```

## #Moda

```
names(civil_freq_absoluta)[which.max(civil_freq_absoluta)]
```

```
# ----- #
```

```
#  VARIÁVEL QUALITATIVA ORDINAL
```

```
# ----- #
```

## ## Frequência absoluta

```
escolaridade_freq_absoluta <- table(milsa$ESCOLARIDADE)
escolaridade_freq_absoluta
```

## ## Frequência relativa

```
escolaridade_freq_relativa <- prop.table(escolaridade_freq_absoluta)
```

## ## Frequencia relativa acumulada

```
escolaridade_freq_acum <- cumsum(escolaridade_freq_relativa)
escolaridade_freq_acum
```

```
freq(milsa$ESCOLARIDADE, cumul = TRUE, totals = FALSE, order = "level")
```

## #Gráficos

```
barplot(escolaridade_freq_absoluta)
```

```
barplot(escolaridade_freq_relativa)
```

## #Usando a função sort() para ordenar os valores da tabela e fazer o gráfico

### ## Menor para maior

```
barplot(sort(escolaridade_freq_absoluta))
```

### ## Maior para menor

```
barplot(sort(escolaridade_freq_absoluta, decreasing = TRUE))
```

## ## Moda

```
names(escolaridade_freq_absoluta)[which.max(escolaridade_freq_absoluta)]
```

## ## Mediana

```
median(milsa$ESCOLARIDADE) # só funciona para variáveis numéricas
```

## # traz a mediana da codificação do nível

```
median(as.numeric(milsa$ESCOLARIDADE))
```

## #valor correto

```
levels(milsa$ESCOLARIDADE)[median(as.numeric(milsa$ESCOLARIDADE))]
```

```
# ----- #
```

```
#  VARIÁVEL QUANTITATIVA DISCRETA
```

```
# ----- #
```

### ## Frequência absoluta

```
filhos_freq_absoluta <- table(milsa$NUM_FILHOS)
```

```
filhos_freq_absoluta
```

### ## Frequência relativa

```
filhos_freq_relativa <- prop.table(filhos_freq_absoluta)
```

```
filhos_freq_relativa
```

### ## Frequência relativa acumulada

```
filhos_freq_acumulada <- cumsum(filhos_freq_relativa)
```

```
filhos_freq_acumulada
```

### ## Frequência relativa

```
plot(filhos_freq_relativa)
```

### ## Frequência relativa acumulada

```
plot(filhos_freq_acumulada, type = "S")
```

### ## Moda

```
names(filhos_freq_relativa)[which.max(filhos_freq_relativa)]
```

### ## Mediana

```
median(milsa$NUM_FILHOS, na.rm = TRUE)
```

### ## Média

```
mean(milsa$NUM_FILHOS, na.rm = TRUE)
```

### ## Quartis

```
quantile(milsa$NUM_FILHOS, na.rm = TRUE)
```

### ## Máximo e mínimo

```
max(milsa$NUM_FILHOS, na.rm = TRUE)
```

```
min(milsa$NUM_FILHOS, na.rm = TRUE)
```

### ## Amplitude

```
diff(range(milsa$NUM_FILHOS, na.rm = TRUE))
```

### ## Variância

```
var(milsa$NUM_FILHOS, na.rm = TRUE)
```

### ## Desvio-padrão

```
sd(milsa$NUM_FILHOS, na.rm = TRUE)
```

## ## Coeficiente de variação

```
sd(milsa$NUM_FILHOS, na.rm = TRUE)/mean(milsa$NUM_FILHOS, na.rm = TRUE)
```

```
summary(milsa$NUM_FILHOS)
```

```
# ----- #
```

```
#  VARIÁVEL QUANTITATIVA CONTÍNUA
```

```
# ----- #
```

## #Classes para variável contínua

## ## Máximo e mínimo

```
range(milsa$SALARIO)
```

## ## Número de classes estimado, com base no critério de Sturges.

```
nclass.Sturges(milsa$SALARIO)
```

## ## Criando as classes com a função cut(), usando os valores mínimos e máximos dados em range()

```
cut(milsa$SALARIO, breaks = seq(4, 23.3, length.out = 8))
```

## # Retirando a classe NA

```
cut(milsa$SALARIO, breaks = seq(4, 23.3, length.out = 8),  
    include.lowest = TRUE)
```

## ## Tabela com as frequências absolutas por classe

```
salario.freq_absoluta <- table(salario.cut)
```

```
salario.freq_absoluta
```

```
prop.table(salario.freq_absoluta)
```

## #Gráfico

```
hist(milsa$SALARIO)
```

```
hist(milsa$SALARIO, freq = FALSE, labels = TRUE)
```

```
boxplot(milsa$SALARIO)
```

## ## Mediana

```
median(milsa$SALARIO)
```

## ## Média

```
mean(milsa$SALARIO)
```

## ## Quartis

```
quantile(milsa$SALARIO)
```

## ## Máximo e mínimo

```
max(milsa$SALARIO)
```

```
min(milsa$SALARIO)
range(milsa$SALARIO)
```

### ## Amplitude

```
diff(range(milsa$SALARIO))
```

### ## Variância

```
var(milsa$SALARIO)
```

### ## Desvio-padrão

```
sd(milsa$SALARIO)
```

### ## Coeficiente de variação

```
sd(milsa$SALARIO)/mean(milsa$SALARIO)
```

```
# ----- #
```

```
# Qualitativa vs qualitativa
```

```
# ----- #
```

### ## Tabela de frequências absolutas

```
civ_escola_absoluta<- with(milsa, table(ESTADO_CIVIL, ESCOLARIDADE))
```

```
civ_escola_absoluta
```

```
addmargins(civ_escola_absoluta)
```

#Tabelas de frequências relativas são obtidas com prop.table(), mas aqui existem três possibilidades para as proporções em cada casela:

# Em relação ao total geral

# Em relação aos totais por linha (margin = 1)

# Em relação aos totais por coluna (margin = 2)

### ## Frequência relativa global

```
prop.table(civ_escola_absoluta)
```

### ## Frequência relativa por linha

```
prop.table(civ_escola_absoluta, margin = 1)
```

### ## Frequência relativa por coluna

```
prop.table(civ_escola_absoluta, margin = 2)
```

### #Gráficos

```
barplot(civ_escola_absoluta, legend = TRUE)
```

```
barplot(t(civ_escola_absoluta), legend = TRUE)
```

```
barplot(civ_escola_absoluta, beside = TRUE, legend = TRUE)
```

```
barplot(t(prop.table(civ_escola_absoluta)), beside = TRUE, legend = TRUE)
```

```
# ----- #  
# Qualitativa vs quantitativa  
# ----- #
```

## ## Quartis de salario

```
quantile(milsa$SALARIO)
```

## ## Classificação de acordo com os quartis

```
salario.cut <- cut(milsa$SALARIO, breaks = quantile(milsa$SALARIO),  
  include.lowest = TRUE)
```

## ## Tabela de frequências absolutas

```
escola.salario.absoluta <- table(milsa$ESCOLARIDADE, salario.cut)  
escola.salario.absoluta
```

```
prop.table(escola.salario.absoluta, margin = 1)
```

## #Gráfico

```
boxplot(SALARIO ~ ESCOLARIDADE, data = milsa)
```

## #Estatísticas Descritivas

```
with(milsa, tapply(SALARIO, ESCOLARIDADE, mean))
```

```
with(milsa, tapply(SALARIO, ESCOLARIDADE, sd))
```

```
with(milsa, tapply(SALARIO, ESCOLARIDADE, quantile))
```

```
# ----- #  
# Quantitativa vs Quantitativa  
# ----- #
```

## ## Classes de Idade

```
idade.cut <- with(milsa, cut(IDADE, breaks = quantile(IDADE),  
  include.lowest = TRUE))  
table(idade.cut)
```

## ## Classes de salario

```
salario.cut <- with(milsa, cut(SALARIO, breaks = quantile(SALARIO),  
  include.lowest = TRUE))  
table(salario.cut)
```

## ## Tabela cruzada

```
table(idade.cut, salario.cut)
```

```
prop.table(table(idade.cut, salario.cut), margin = 1)
```

```
idade.cut2 <- with(milsa, cut(IDADE,  
                             breaks = quantile(IDADE, seq(0, 1, length = 4)),  
                             include.lowest = TRUE))  
salario.cut2 <- with(milsa, cut(SALARIO,  
                                breaks = quantile(SALARIO, seq(0, 1, length = 4)),  
                                include.lowest = TRUE))  
table(idade.cut2, salario.cut2)  
  
prop.table(table(idade.cut2, salario.cut2), margin = 1)
```

### #Gráfico

```
plot(x = milsa$IDADE, y = milsa$SALARIO)  
plot(SALARIO ~ IDADE, data = milsa)
```

### #Coeficiente de correlação

```
with(milsa, cor(IDADE, SALARIO))  
with(milsa, cor(IDADE, SALARIO, method = "kendall"))  
with(milsa, cor(IDADE, SALARIO, method = "spearman"))
```