



## MAPA – Material de Avaliação Prática da Aprendizagem

Acadêmico: Lucas Antonio Rocha Santos	R.A. 22134899-5
Curso: Ciências de Dados e Análise de Comportamento	
Disciplina: Modelagem Estatística	
Valor da atividade: 3,00	Prazo: 18/08/2023

### Identificar qual banco de dados utilizará:

R.A:  $2+2+1+3+4+8+9+9+5 = 43$  ( $4 + 3 = 7$ )

Utilizarei o banco de dados Infert do R.

### Aprender sobre o conjunto de dados:

O conjunto de dados Infert é utilizado para analisar a relação entre determinadas variáveis e a fertilidade das mulheres. Ele pode ser usado para criar modelos estatísticos que preveem a fertilidade com base nas informações fornecidas.

#### Variáveis do conjunto infert:

**education:** Nível de educação da mulher, que é uma variável categórica.

**age:** Idade da mulher, representada como uma variável numérica.

**parity:** Número de gestações anteriores, representado como uma variável numérica.

**case:** Um identificador numérico para cada caso no conjunto de dados.

**induced:** Um indicador binário que informa se a gestação atual é induzida ou não



## Análise Exploratória dos Dados:

### Medidas descritivas para as variáveis numéricas "age" e "parity"

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Maps.R\* x

Source on Save Run Source

```

1 library(datasets)
2 library(ggplot2)
3
4 # Carregar conjunto de dados
5 data("infert")
6
7 # Visualizar primeiras linhas do conjunto de dados
8 head(infert)
9
10 # Medidas descritivas para age e parity
11 summary(infert$age)
12 summary(infert$parity)|
13
14

```

12:23 (Top Level) R S

Console Terminal x Background Jobs x

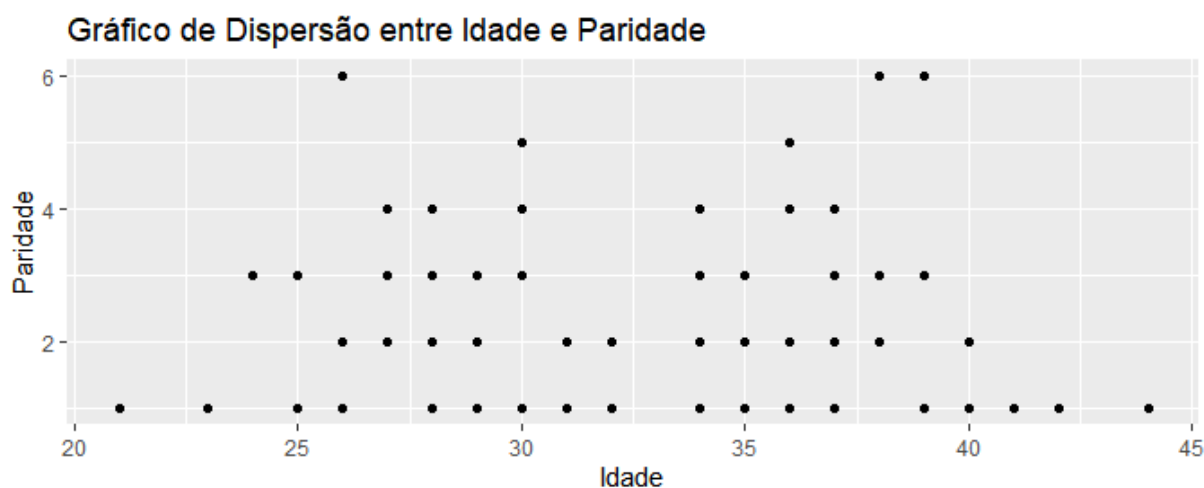
R 4.3.1 ~/

```

> summary(infert$age)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 21.00  28.00   31.00  31.50  35.25   44.00
> summary(infert$parity)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  1.000   2.000   2.093   3.000   6.000
> |

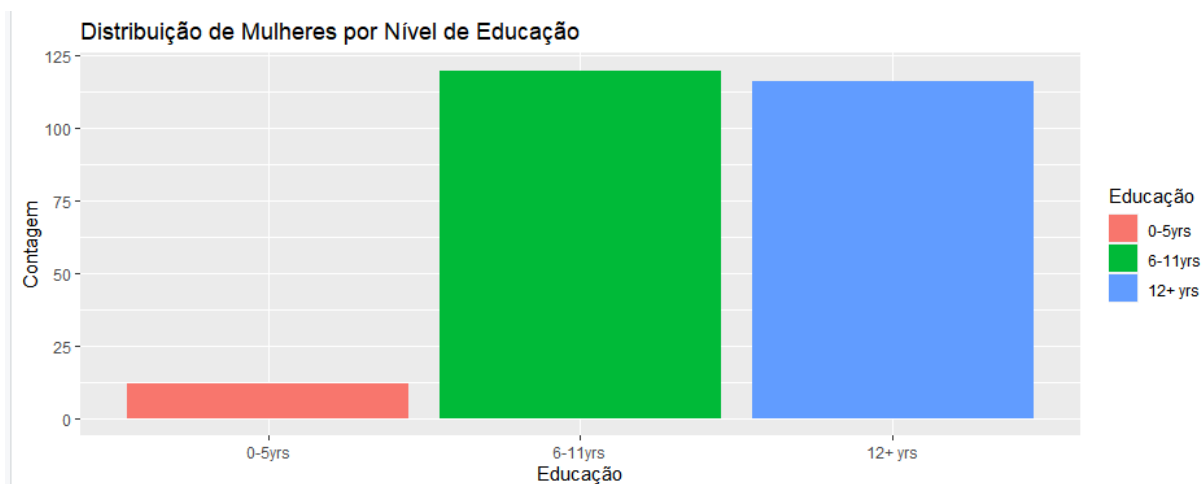
```

### Gráfico de dispersão entre idade e paridade

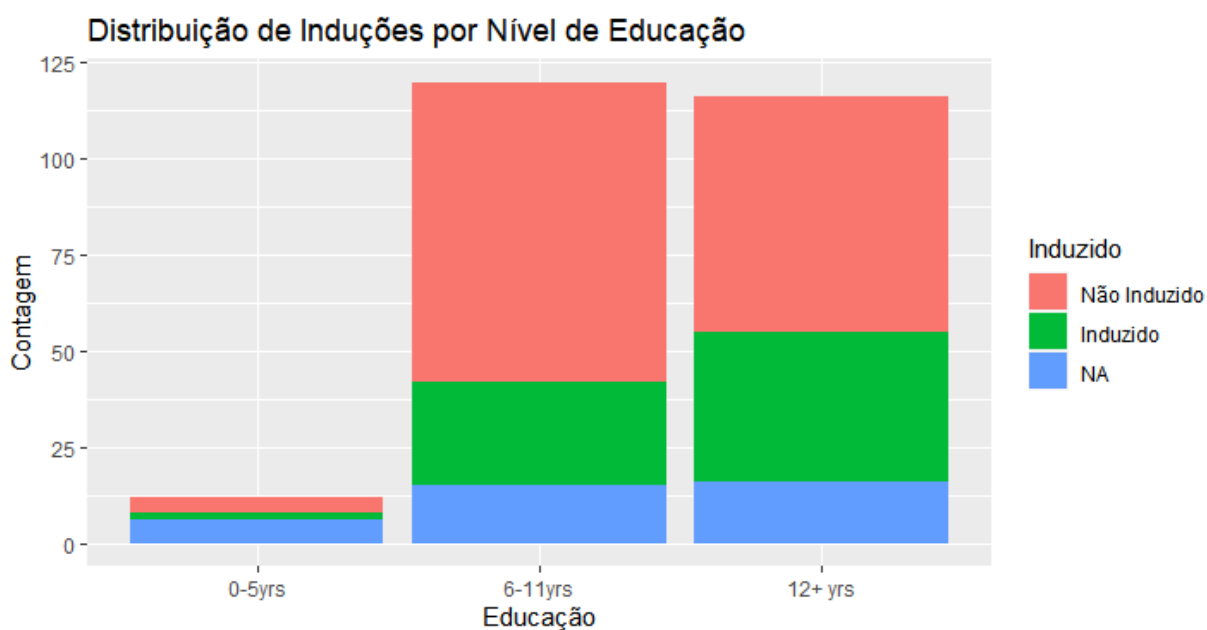




## Gráfico de barras para a distribuição das mulheres em cada nível de educação



## Gráfico de barras empilhadas para mostrar a distribuição de induções em cada nível de educação:





## Construa um Modelo Estatístico

O modelo estatístico pode ser feito usando uma regressão linear simples

### Construção do modelo e informações de convergência

```
41 # Construção do modelo de regressão linear simples
42 modelo <- lm(parity ~ age, data = infert)
43
44 # informações de convergência do modelo
45 summary(modelo)
46
47
```

48:26 (Top Level) R Script

Console Terminal Background Jobs

R 4.3.1 ~ /

```
> modelo <- lm(parity ~ age, data = infert)
> summary(modelo)
```

Call:  
lm(formula = parity ~ age, data = infert)

Residuals:

Min	1Q	Median	3Q	Max
-1.339	-1.043	-0.142	0.784	4.016

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.47114	0.48359	3.042	0.0026 **
age	0.01973	0.01514	1.303	0.1938

---  
Signif. codes:  
0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.25 on 246 degrees of freedom  
Multiple R-squared: 0.006855, Adjusted R-squared: 0.002818  
F-statistic: 1.698 on 1 and 246 DF, p-value: 0.1938

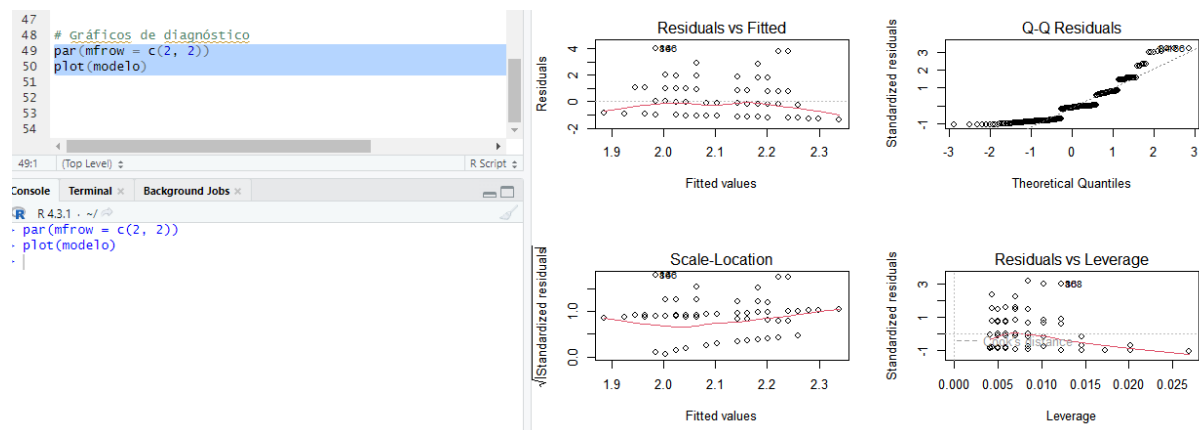
O modelo mostrou que o valor estimado da paridade quando a idade é zero é 1.47114, mas esse valor não faz muito sentido na realidade. O coeficiente estimado para a variável "age" é 0.01973, o que significa que, em média, a paridade aumenta em 0.01973 para cada ano a mais de idade.

O valor-p para o termo de interceptação é baixo (0.0026), indicando que o coeficiente é estatisticamente significativo. No entanto, o valor-p para a variável "age" é maior (0.1938), sugerindo que a idade pode não ser estatisticamente relevante para explicar as variações na paridade.

O coeficiente de determinação ( $R^2$ ) é baixo (0.006855), o que significa que apenas uma pequena parte da variação na paridade é explicada pela idade no modelo. Isso sugere que a idade por si só pode não ser suficiente para explicar as diferenças na paridade entre as mulheres.

Os resultados sugerem que a idade pode não ser um fator estatisticamente significativo para explicar a variação na paridade.

**Para verificar a qualidade do ajuste e a normalidade dos resíduos, pode-se utilizar gráficos de diagnóstico:**



### Teste de Breusch-Pagan para verificar a homogeneidade dos resíduos

