**DIRETORIA DE DESENVOLVIMENTO E ENSINO**

**DEPARTAMENTO DE ENSINO SUPERIOR**

**UNIDADE ACADÊMICA DE INFORMAÇÃO E COMUNICAÇÃO**

**CURSO SUPERIOR TECNOLÓGICO EM SISTEMAS PARA INTERNET**

**RELATÓRIO DESCRITIVO E ANALÍTICO DE UM BANCO DE DADOS UTILIZANDO O R COMO FERRAMENTA COMPUTACIONAL**

**DISCENTE: AARON PEDRO SANTANA DOS SANTOS**

João Pessoa, 23 de fevereiro de 2018.

# Introdução

O seguinte trabalho tem como objetivo realizar uma série de análises estatísticas e probabilísticas a partir de um banco contendo dados demográficos quantitativos e qualitativos acerca de uma população. A análise consistirá em uma série de testes conduzidos com o Software livre para análises estatísticas R, testes estes de natureza descritiva e probabilística, buscando maneiras eficientes de apresentar os dados da amostra coletada, quais modelos e métodos analíticos podem ser aplicados e uma breve validação da amostra.

# Metodologia

Para a seguinte análise, foram seguidas as seguintes etapas de forma sequencial: Classificação das variáveis, Criação da amostra, Análises descritivas, Testes de modelos probabilísticos, Testes de hipóteses e intervalos de confiança e por fim realizar Correlação e Regressão Linear.

As variáveis serão classificadas primariamente em dois grupos: Quantitativas e Qualitativas, e estes grupos se subdividem em: Quantitativas discretas e contínuas e Qualitativas Nominais e Ordinais. A criação da amostra foi gerada com uso de funções do R para realizar uma coleta aleatório dos dados o banco. Para as análises descritivas, será realizada distribuições de frequências, cálculo de medidas de posição e gráficos diversos com auxílio do R de maneira automatizada. Para os testes de Modelos probabilísticos serão realizados os testes de Shapiro Wilk, Anderson Darling, Lilliefors e Shapiro Francia, por meio de funções do R. Para a realização dos intervalos de confiança, teste de hipótese, correlação e regressão linear, foram igualmente mediadas por funções do R.

# Resultados encontrados

* 1. Resumo :

Inicialmente, foi gerada a amostrar a partir do banco (com 533 observações) por meio de coleta aleatória, com 60 observações. Na amostra coletada e trabalhada neste relatório a variável “habit\_outro”, não teve nenhum valor válido.

As variáveis do banco foram em grande parte categorizadas como Qualitativas (17) e as demais (5) quantitativas. Grande parte dos participantes registrados na amostra, residem em Recife, tem em média 18 anos, não completaram algum tipo de ensino regular (70%) se considerando de classe baixa (50%).

Nenhuma das variáveis quantitativas analisadas aderiram ao modelo normal nos testes realizados. Os intervalos de confiança traçados com 95% de confiança tinham eram demasiadamente precisos, com um intervalo variando pouco da média. No caso dos testes de hipótese, as variáveis “Casa”, “Recursos”,”Renda\_familiar”, “Renda\_da\_familia” não tiveram a hipótese H0 rejeitada.

Quanto a Correlação e regressão linear entre Renda\_familiar e Renda\_div\_pessoas, ou seja, o número de pessoas na casa e a renda total da família, foi identificada uma correlação positiva (quanto maior o número de pessoas maior a renda), porém, baixa e insignificante, sendo a renda familiar (y) bastante dispersa quando posta em relação à quantidade de pessoas (x), apesar da correlação encontrada.

* 1. Análises das variáveis:

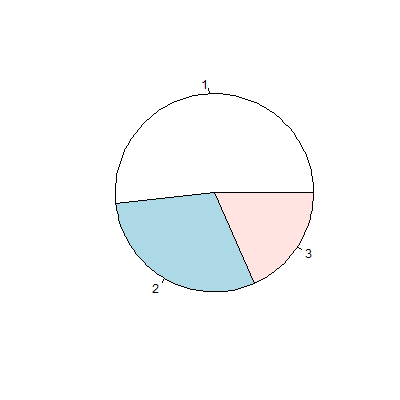
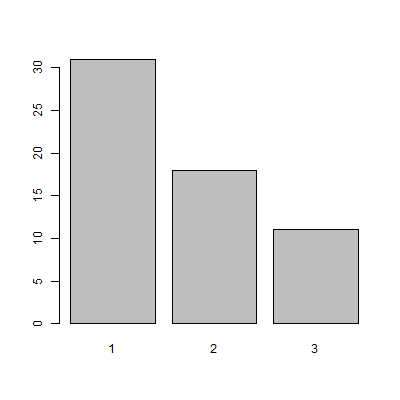
**Variável 2**: NCidade

**Tipo:** Qualitativa Nominal

**Domínio de respostas:** (01) Recife, (02) Vitória , (03) Florianópolis  
**Análises da amostra:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Recife ( 1 )** | **Vitória ( 2 )** | **Florianópolis ( 3 )** | **Total** |
| 31 | 18 | 11 | 60 |

Distribuição de frequência para a Variável 2



Gráficos para a distribuição de Ncidade

**Variável 3**: Local

**Tipo:** Qualitativa Nominal

**Domínio de respostas:**

(01) Unidade de saúde,   
(02) Casa da entrevistada(sozinha)

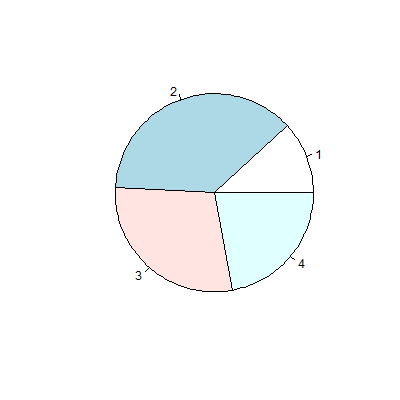
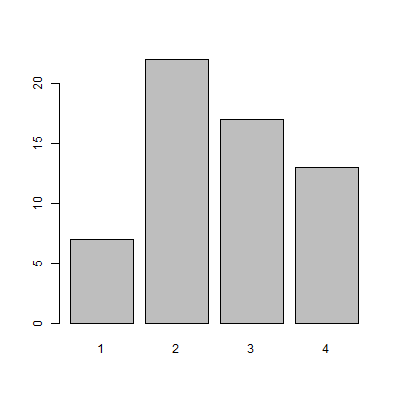
(03) Casa da entrevistada (com presença de outras pessoas)

(04) Outro

**Análises da amostra:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **Total** |
| 7 | 22 | 17 | 13 | 59 |

Distribuição de frequência para a Variável 3

****

Gráficos para a distribuição de Local

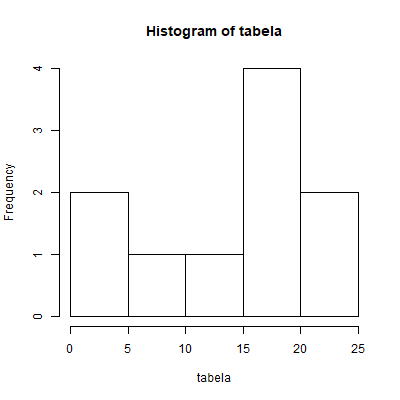
**Variável 4**: Idade

**Tipo:** Quantitativa Continua

**Domínio de respostas:** Números naturais maiores que zero  
**Análises da amostra:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Min.** | **1st Qu.** | **Median** | **Mean** | **3rd Qu.** | **Max.** | **Mode** | **S.d.** | **var.** |
| 15 | 18 | 19.5 | 19.36667 | 21 | 22 | 20 | 1.677029928 | 2.812429379 |

Medidas de dispersão e posição para variável 4



Histograma para frequência de Idade.

**Testes de Normalidade:**

|  |  |  |
| --- | --- | --- |
| **Teste** | **Resultado** | **P-Value** |
| **Shapiro-Wilk** | Rejeitado | 0.0142580766240092 |
| **Anderson-Darling** | Rejeitado | 0.00624507546187314 |
| **Lilliefors (Kolmogorov-Smirnov)** | Rejeitado | 0.00245732403016207 |
| **Shapiro-Francia** | Rejeitado | 0.0250836173196863 |

**Teste de Hipótese:**

*P-value* para h0: 0.0305495907073783 .

Hipótese rejeitada.

**Variável 5**: Cidade

**Tipo:** Qualitativa Nominal

**Domínio de respostas:**

(01) Na cidade onde foi aplicado o questionário

(02) Outra cidade, mas no mesmo estado

(03) Outra cidade, em outro estado, mas na mesma região (04) Outra região

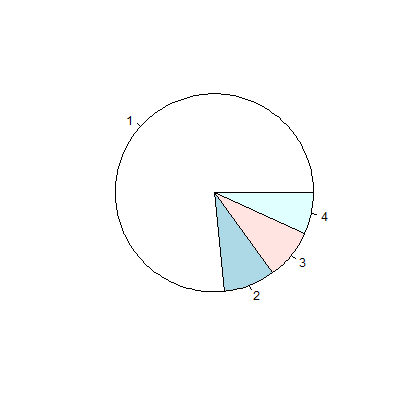
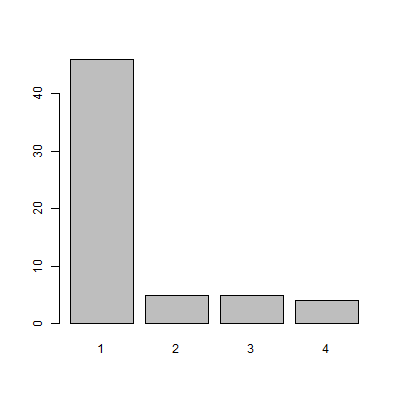
(05) Outro país

(98) Não sei

**Análises da amostra:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** | **98** | **Total** |
| 46 | 5 | 5 | 4 | 0 | 0 | 60 |

Distribuição de frequência para a Variável 5



Gráficos para a distribuição de Cidade

**Variável 6**: Tempo

**Tipo:** Quantitativa Continua Categorizada

**Domínio de respostas:**

(01) Desde que nasceu

(02) Há menos de cinco anos (03) Entre 5 e dez anos

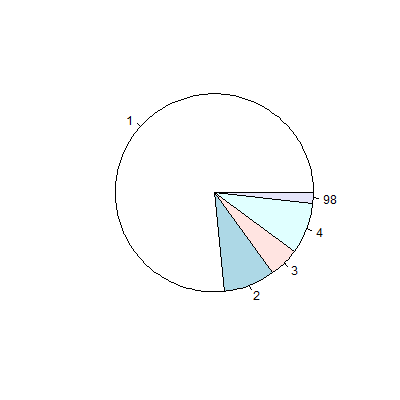
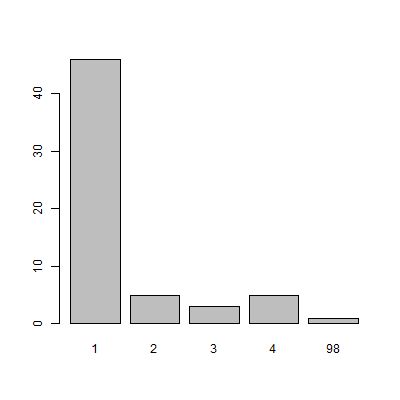
(04) Há mais de dez anos

(98) Não sei

**Análises da amostra:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **98** | **Total** |
| 46 | 5 | 3 | 5 | 1 | 60 |

Distribuição de frequência para a Variável 6



Gráficos para a distribuição de Tempo

**Variável 7**: Bairro

**Tipo:** Qualitativa Nominal

**Domínio de respostas:**

(01) No bairro onde foi aplicado o questionário

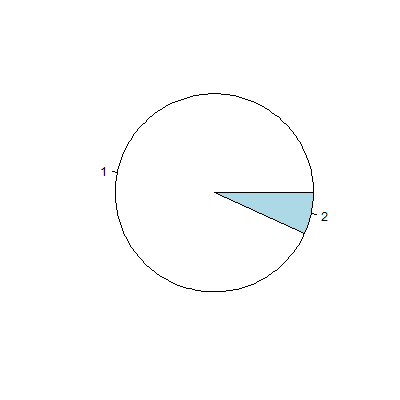
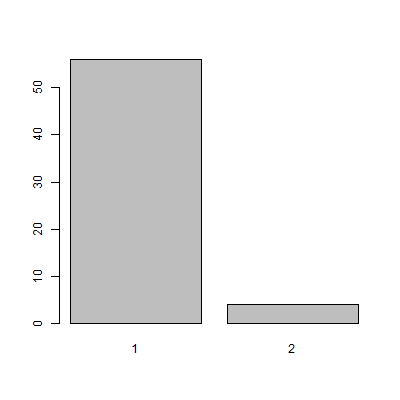
(02) Outro bairro, mas próximo, na mesma região

(03) Outro bairro, distante

**Análises da amostra:**

|  |  |  |  |
| --- | --- | --- | --- |
| **1** | **2** | **3** | **Total** |
| 56 | 4 | 0 | 60 |

Distribuição de frequência para a Variável 7



Gráficos para a distribuição de Bairro

**Variável 8**: Cor

**Tipo:** Qualitativa Nominal

**Domínio de respostas:**

(01) branca

(02) preta

(03) parda

(04) oriental

(05) indígena

(06) outra

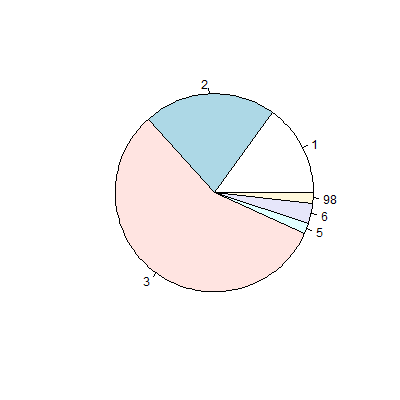
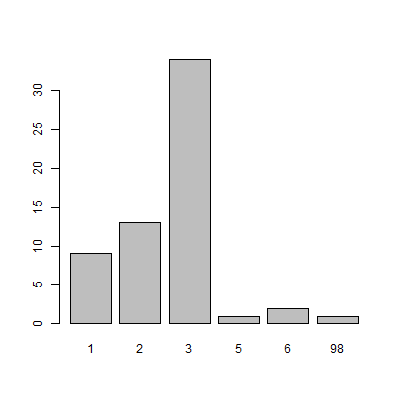
(98) Não sabe

(99) Não respondeu

**Análises da amostra:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **5** | **6** | **98** | **Total** |
| 9 | 13 | 34 | 1 | 2 | 1 | 60 |

Distribuição de frequência para a Variável 8



Gráficos para a distribuição de Cor

**Variável 9**: Religião

**Tipo:** Qualitativa Nominal

**Domínio de respostas:**

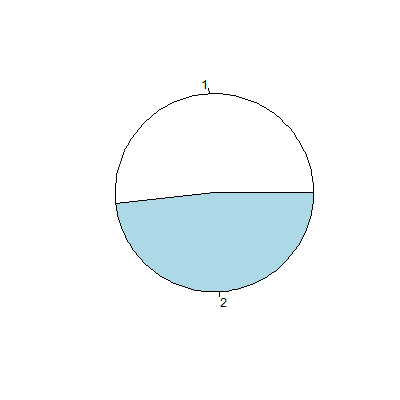
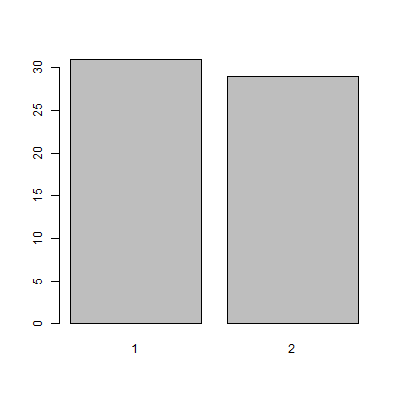
(01) Sim

(02) Não

**Análises da amostra:**

|  |  |  |
| --- | --- | --- |
| **1** | **2** | **Total** |
| 31 | 29 | 60 |

Distribuição de frequência para a Variável 9



Gráficos para a distribuição de Religião

**Variável 10**: Religião\_qual

**Tipo:** Qualitativa Nominal

**Domínio de respostas:**

(01) Católica

(02) Evangélica.

(03) Espírita/Kardecista

(04) Umbanda

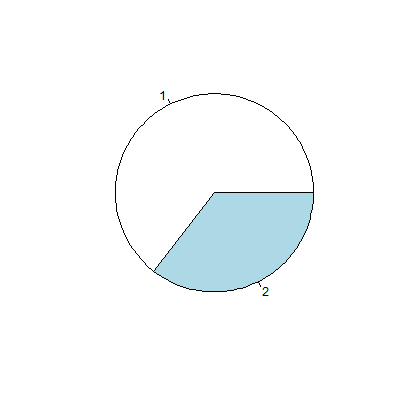
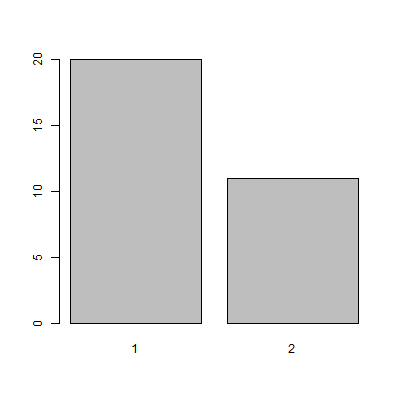
(05) Candomblé

(06) Outra

**Análises da amostra:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **5** | **6** | **Total** |
| 20 | 11 | 0 | 0 | 0 | 31 |

Distribuição de frequência para a Variável 10



Gráficos para a distribuição de Religião\_qual

**Variável 11**: Estudando

**Tipo:** Qualitativa Nominal

**Domínio de respostas:**

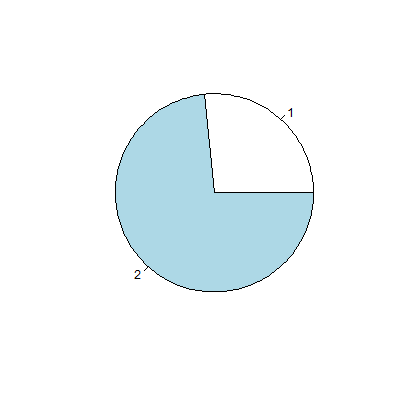
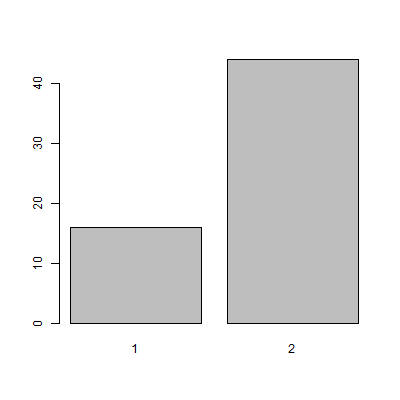
(01) Sim

(02) Não

**Análises da amostra:**

|  |  |  |
| --- | --- | --- |
| **1** | **2** | **Total** |
| 16 | 44 | 60 |

Distribuição de frequência para a Variável 11



Gráficos para a distribuição de Estudando

**Variável 12**: Estudo

**Tipo:** Qualitativa Ordinal

**Domínio de respostas:**

(01) Nunca estudou e não sabe ler, nem escrever

(02) Nunca estudou, mas sabe ler e escrever

(03) Ensino fundamental incompleto

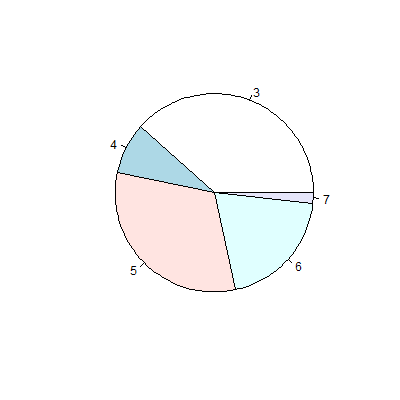
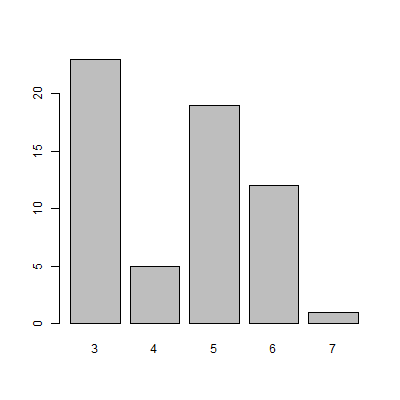
(04) Ensino fundamental completo

(05) Ensino médio incompleto   
 (06) Ensino médio completo   
 (07) Ensino superior incompleto.

**Análises da amostra:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **3** | **4** | **5** | **6** | **7** | **Total** |
| 23 | 5 | 19 | 12 | 1 | 60 |

Distribuição de frequência para a Variável 12



Gráficos para a distribuição de Estudo

**Variável 13**: Casa

**Tipo:** Quantitativa Continua

**Domínio de respostas:** Números naturais maiores que zero, 00 = Mora só  
**Análises da amostra:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Min.** | **1st Qu.** | **Median** | **Mean** | **3rd Qu.** | **Max.** | **Mode** | **S.d.** | **var.** |
| 2 | 3 | 4 | 4.783333333 | 6 | 12 | 3 | 2.083768316 | 4.342090395 |

Medidas de dispersão e posição para variável 13

**Intervalo de confiança:**

95% de confiança de 18.93344 a 19.79989

**Testes de Normalidade:**

|  |  |  |
| --- | --- | --- |
| **Teste** | **Resultado** | **P-Value** |
| **Shapiro-Wilk** | Rejeitado | 9.60103139612951*e-06* |
| **Anderson-Darling** | Rejeitado | 7.72314560219555*e-07* |
| **Lilliefors (Kolmogorov-Smirnov)** | Rejeitado | 8.90038158491331*e-06* |
| **Shapiro-Francia** | Rejeitado | 2.9731169656226*e-05* |

**Teste de Hipótese:**

*P-value* para h0: 0.64087287334763

Hipótese não rejeitada.

**Variável 14**: Qualeotipodehabitacao

**Tipo:** Qualitativa Nominal

**Domínio de respostas:**

(01) casa

(02) apartamento

(03) cômodo

(04) pousada

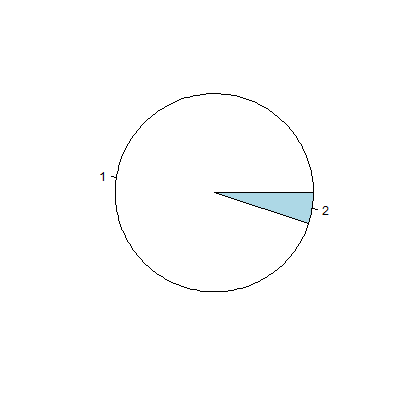
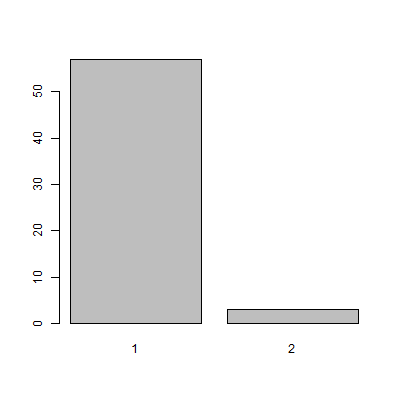
(05) albergue

(06) Outro

**Análises da amostra:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** | **6** | **Total** |
| 57 | 3 | 0 | 0 | 0 | 0 | 60 |

Distribuição de frequência para a Variável 14



Gráficos para a distribuição de Qualeotipodehabitacao

**Variável 15**: Habit\_outro

**Tipo:** Qualitativa Nominal

**Não houveram respostas.**

**Variável 16**: Habit\_alug

**Tipo:** Qualitativa Nominal

**Domínio de respostas:**

(01) próprio (comprado).

(02) próprio (Título de posse adquirido “pósinvasão”).

(03) alugado

(04) cedido/emprestado

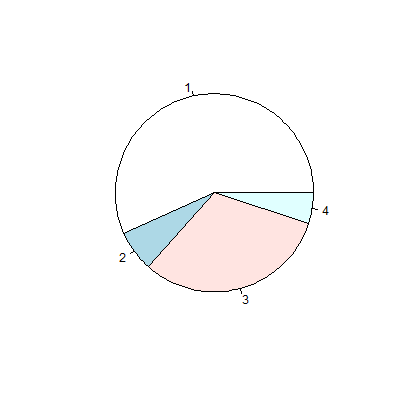
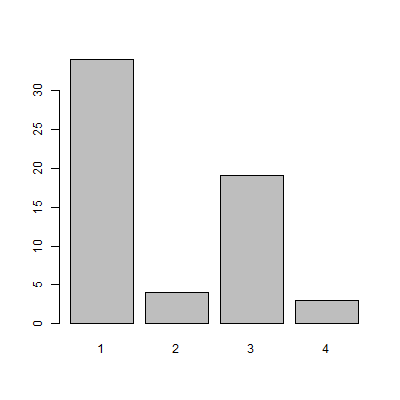
(05) Não sei

(06) Outro

**Análises da amostra:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **5** | **6** | **Total** |
| 34 | 4 | 19 | 3 | 0 | 0 | 60 |

Distribuição de frequência para a Variável 16



Gráficos para a distribuição de Habit\_alug

**Variável 17**: Comodos

**Tipo:** Quantitativa Discreta Categorizada/Qualitativa Ordinal

**Domínio de respostas:**

(01) Não sei

(02) Domicílio coletivo (albergue, pousada etc.)

(03) de 1 a 5

(04) de 6 a 10

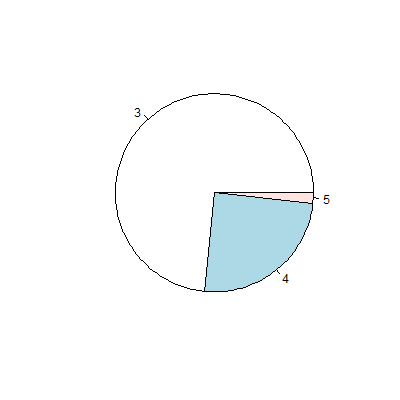
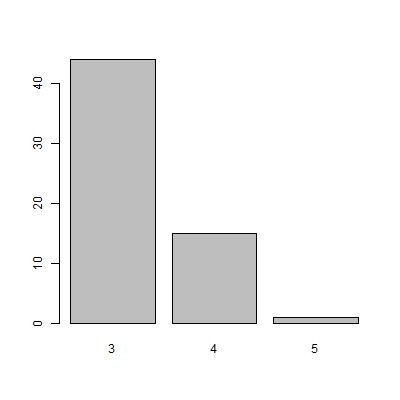
(05) De 10 a 20

(06) Mais de 20

**Análises da amostra:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **1** | **2** | 3 | 4 | 5 | **6** | **Total** |
| 0 | 0 | 44 | 15 | 1 | 0 | 60 |

Distribuição de frequência para a Variável 17



Gráficos para a distribuição de Comodos

**Variável 18**: Trabalho

**Tipo:** Qualitativa Nominal

**Domínio de respostas:**

(01) exerce atividade remunerada.

(02) recebe benefício, pensão, aposentadoria ou bolsa de estudo

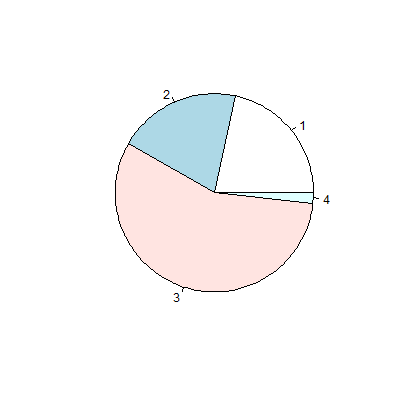
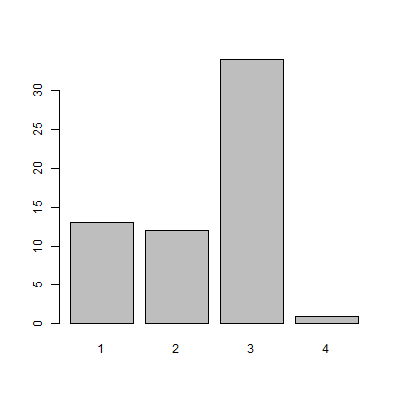
(03) não trabalha, nem recebe benefício.

(04) outro

**Análises da amostra:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **1** | **2** | **3** | **4** | **Total** |
| 13 | 12 | 34 | 1 | 60 |

Distribuição de frequência para variável 18



Gráficos para a distribuição de Trabalho

**Variável 19**: Recursos

**Tipo:** Quantitativa Continua

**Domínio de respostas:** Número pertencente ao grupo dos reais maiores que 0.  
**Análises da amostra:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Min.** | **1st Qu.** | **Median** | **Mean** | **3rd Qu.** | **Max.** | **Mode** | **S.d.** | **var.** |
| 60 | 100 | 234 | 301.12 | 510 | 800 | 100 | 219.9472134 | 48376.77667 |

Medidas de dispersão e posição para variável 19

**Intervalo de confiança:**

95% de confiança de 210.3303 a 391.9097

**Testes de Normalidade:**

|  |  |  |
| --- | --- | --- |
| **Teste** | **Resultado** | **P-Value** |
| **Shapiro-Wilk** | Rejeitado | 0.00278329635927281 |
| **Anderson-Darling** | Rejeitado | 0.00073169230391482 |
| **Lilliefors (Kolmogorov-Smirnov)** | Rejeitado | 0.00811639217130916 |
| **Shapiro-Francia** | Rejeitado | 0.00601815760482751 |

**Teste de Hipótese:**

*P-value* para h0: 0.732666303773859

Hipótese não rejeitada.

**Variável 20**: Renda\_familiar

**Tipo:** Quantitativa Continua

**Domínio de respostas:** Número pertencente ao grupo dos reais maior ou igual a 0.  
**Análises da amostra:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Min.** | **1st Qu.** | **Median** | **Mean** | **3rd Qu.** | **Max.** | **Mode** | **S.d.** | **var.** |
| 1 | 473.75 | 675 | 722 | 922.5 | 2500 | 800 | 427.1510682 | 182458.0351 |

Medidas de dispersão e posição para variável 20

**Intervalo de confiança:**

95% de confiança de 210.3303 a 391.9097

**Testes de Normalidade:**

|  |  |  |
| --- | --- | --- |
| **Teste** | **Resultado** | **P-Value** |
| **Shapiro-Wilk** | Rejeitado | 6.80910596964276*e-05* |
| **Anderson-Darling** | Rejeitado | 0.00105056904912758 |
| **Lilliefors (Kolmogorov-Smirnov)** | Rejeitado | 0.0314160757908708 |
| **Shapiro-Francia** | Rejeitado | 8.99844146009149e-05 |

**Teste de Hipótese:**

*P-value* para h0: 0.151744583660338

Hipótese não rejeitada.

**Variável 21**: Renda\_div\_pessoa

**Tipo:** Quantitativa Continua

**Domínio de respostas:** Número pertencente ao grupo dos reais maior ou igual a 0.  
**Análises da amostra:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Min.** | **1st Qu.** | **Median** | **Mean** | **3rd Qu.** | **Max.** | **Mode** | **S.d.** | **var.** |
| 2 | 3 | 4 | 4.50877193 | 5 | 12 | 3 | 2.062616291 | 4.254385965 |

Medidas de dispersão e posição para variável 21

**Testes de Normalidade:**

|  |  |  |
| --- | --- | --- |
| **Teste** | **Resultado** | **P-Value** |
| **Shapiro-Wilk** | Rejeitado | 4.16770403725866*e-06* |
| **Anderson-Darling** | Rejeitado | 2.94880746379014*e-07* |
| **Lilliefors (Kolmogorov-Smirnov)** | Rejeitado | 1.0790545185988*e-05* |
| **Shapiro-Francia** | Rejeitado | 1.41670598404273*e-05* |

**Teste de Hipótese:**

*P-value* para h0: 0.000960572615672153

Hipótese rejeitada.

**Variável 22**: Renda\_da\_familia

**Tipo:** Quantitativa Continua

**Domínio de respostas:** Número pertencente ao grupo dos reais maior ou igual a 0.  
**Análises da amostra:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Min.** | **1st Qu.** | **Median** | **Mean** | **3rd Qu.** | **Max.** | **Mode** | **S.d.** | **var.** |
| 0 | 123.5 | 168.5 | 172.3928571 | 201 | 533 | 200 | 96.36156883 | 9285.551948 |

Medidas de dispersão e posição para variável 22

**Testes de Normalidade:**

|  |  |  |
| --- | --- | --- |
| **Teste** | **Resultado** | **P-Value** |
| **Shapiro-Wilk** | Rejeitado | 0.00366362255604043 |
| **Anderson-Darling** | Rejeitado | 0.012665908578953 |
| **Lilliefors (Kolmogorov-Smirnov)** | Rejeitado | 0.00854981174569407 |
| **Shapiro-Francia** | Rejeitado | 0.00854981174569407 |

**Teste de Hipótese:**

*P-value* para h0: 0.0748966526858987

Hipótese não rejeitada.

**Variável 23:** Classe

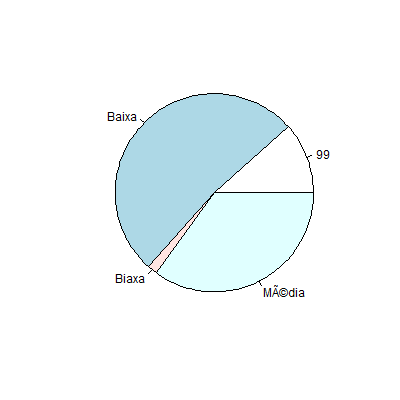
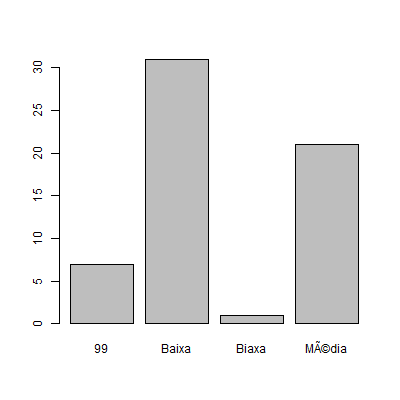
**Tipo:** Qualitativa Nominal

**Domínio de respostas:**

[Digitar resposta] ou (99) não respondeu  
**Análises da amostra:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **99** | **Baixa** | **Biaxa** | **Média** | **Total** |
| 7 | 31 | 1 | 21 | 60 |

Distribuição de frequência para variável 23



Gráficos para a distribuição de Classe

**Variável 24:** Civil

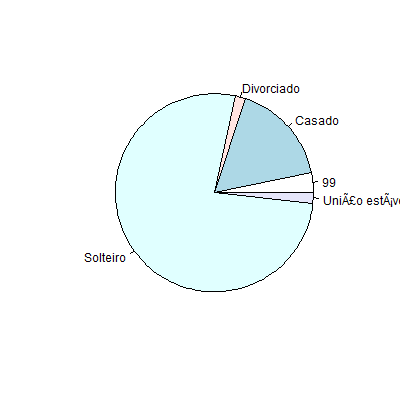
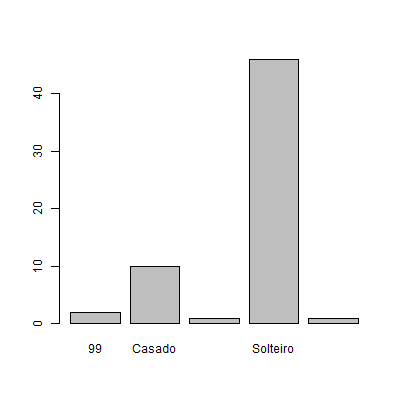
**Tipo:** Qualitativa Nominal

**Domínio de respostas:**

[Digitar resposta] ou (99) não respondeu  
**Análises da amostra:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **99** | **Casado** | **Divorciado** | **Solteiro** | **União estável** | **Total** |
| 2 | 10 | 1 | 46 | 1 | 60 |

Distribuição de frequência para variável 24



Gráficos para a distribuição de Civil

* 1. Correlação e Regressão Linear :

Correlação e Regressão Linear Entre as variáveis Renda\_familiar e Renda\_div\_pessoa (Renda total e o número de pessoas da família)

**Correlação:**

0.1856019 com P-value 0.1708

**Regressão Linear:**

**Estimativa:** 543.75

**Erro Padrão:** 116.16  
**P-Value:** 0.1708

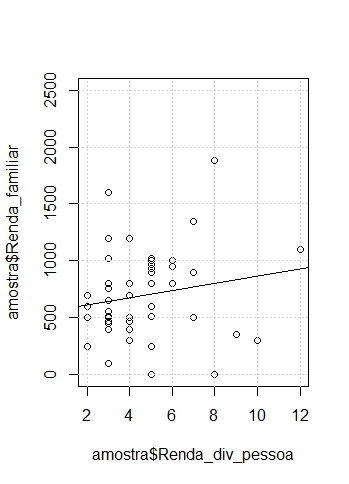


Gráfico de dispersão demonstrando a Regressão Linear

# Linhas de código

O código foi escrito para que seja executado por inteiro, e assim a análise é gerada e impressa no console. Ainda assim, estruturado de forma que possa ser executado pelas etapas descritas para realização do trabalho, mas é importante que seja executado sequencialmente, pela dependência de dados e funções.

Antes de executá-lo, é necessário preencher a variável chamada “caminho\_projeto” com o path do diretório windowns onde se encontra pasta do projeto (com \\ ao invés de \ ) Isso é necessário para garantir o devido carregamento dos arquivos como também a exportação dos gráficos e tabelas realizados na etapa 2.

Apesar do código estar descrito abaixo, é recomendado que seja executado a partir do arquivo .R na pasta FONTES do projeto, pasta essa que também contém a amostra trabalhada.

# ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO #

### INSIRA AQUI O CAMINHO DO DIRETORIO EM QUE SE ENCONTRA O PROJETO

# (O CAMINHO ATÉ O DIRETORIO QUE CONTÉM AS PASTAS "FONTE", "DOCUMENTOS", "DADOS BRUTOS", "ANÁLISES")

caminho\_projeto = "C:\\Users\\Aaron\\Desktop\\rel\_est"

# ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO # ATENÇÃO #

############Implementação de funções de uso genérico.

#install.packages('nortest')

#library(nortest)

#Função de moda:

Mode <- function(x) {

ux <- unique(x)

ux[which.max(tabulate(match(x, ux)))]

}

#Função Tira NA

tiraNa = function(lista){

lista2 = c()

for( n in lista){

if (!is.na(n)){

lista2 = c(lista2,n)

}

}

return(lista2)

}

#Função CV:

cv<-function(x){coef<-sd(x)/mean(x)\*100

return(coef)}

##################################### DEFINIÇÕES DO BANCO E AMOSTRA############################################

##Carregando Banco de dados da população

dadosAaron = read.table(paste( caminho\_projeto, "\\DADOS BRUTOS\\Banco\_de\_dados\_Trabalho\_tratado.txt", sep=""), header = T, sep ="\t", encoding="latin1")

##Gerando amostra com 60 elementos

#amostra = dadosAaron[sample(nrow(dadosAaron), 60),]

##Escrevendo amostra em disco:

#write.table(amostra, file="~/Documentos/Code/R/PROJETO ESTATISTICA-20180206T113737Z-001/PROJETO ESTATISTICA/FONTE/amostraAaron.txt", sep = "\t")

##Lendo amostra de disco:

amostra = read.table(paste(caminho\_projeto,"\\FONTE\\amostraAaron.txt", sep = ""), header = T, sep = "\t")

#######################################FUNÇÕES DA ETAPA 1#############################################################

#Funções de análise descritiva de variáveis qualitativas

geraAnaliseQuali = function(lista,nome,caminho)

{

tabela = table(lista)

caminhoCsv = paste(caminho,"\\",nome, ".csv",sep="")

caminhoPng1 = paste(caminho,"\\",nome,"\_pie.png",sep="")

caminhoPng2 = paste(caminho,"\\",nome,"\_bar.png",sep="")

print("#")

print(paste("Variável: ",nome))

tryCatch({

png(file=caminhoPng1, width = 400, height = 400 )

pie(tabela)

dev.off()

}, error=function(cond){

print("###ERRO AO EXPORTAR GRÁFICO PIZZA###")

print(cond)

print("#")

})

tryCatch({

png(file=caminhoPng2,width = 400, height = 400)

barplot(tabela)

dev.off()

}, error=function(cond){

print("###ERRO AO EXPORTAR GRÁFICO DE BARRAS###")

print(cond)

print("#")

})

tabela["Total"] = sum(tabela)[1]

tabela = t(tabela)

tryCatch({

write.csv(tabela,caminhoCsv, row.names = FALSE)

}, error=function(cond){

print("###ERRO AO EXPORTAR CSV###")

print(cond)

print("#")

})

print(tabela)

}

#Funções de análises descritivas Quantitativa

geraAnaliseQuant = function(lista,nome,caminho) {

lista = tiraNa(lista)

tabela = summary(lista)

tabela["Mode"] = Mode(lista)

tabela["S.d."] = sd(lista)

tabela["var."] = var(lista)

tabela["C.V"] = cv(lista)

caminhoCsv = paste(caminho,"\\",nome,".csv",sep="")

caminhoPng = paste(caminho,"\\",nome,"\_hist.png",sep="")

print("#")

print(paste("Variável: ", nome))

tryCatch(

{

write.csv(t(tabela),caminhoCsv, row.names = F)

}, error=function(cond){

print("###ERRO AO EXPORTAR CSV###")

print(cond)

print("#")

}

)

tryCatch(

{

png(file=caminhoPng, width = 400, height = 400 )

hist(tabela)

dev.off()

}, error=function(cond){

print("###ERRO AO EXPORTAR GRÁFICO HISTOGRAMA###")

print(cond)

print("#")

})

print(tabela)

}

# Função que gera resultado da etapa 1

resultado\_1 = function(caminho) {

geraAnaliseQuali(amostra$NCidade, "V2\_Ncidade", caminho)

geraAnaliseQuali(amostra$Local, "V3\_Local", caminho)

geraAnaliseQuant(amostra$Idade, "V4\_Idade", caminho)

geraAnaliseQuali(amostra$Cidade, "V5\_Cidade", caminho)

geraAnaliseQuali(amostra$Tempo, "V6\_Tempo", caminho)

geraAnaliseQuali(amostra$Bairro, "V7\_Bairro", caminho)

geraAnaliseQuali(amostra$Cor, "V8\_Cor", caminho)

geraAnaliseQuali(amostra$ReligiÃ.o, "V9\_Religião", caminho)

geraAnaliseQuali(amostra$ReligiÃ.o\_qual, "V10\_Religião\_qual", caminho)

geraAnaliseQuali(amostra$Estudando, "V11\_Estudando", caminho)

geraAnaliseQuali(amostra$Estudo, "v12\_Estudo", caminho)

geraAnaliseQuant(amostra$Casa, "V13\_Casa", caminho)

geraAnaliseQuali(amostra$Qualeotipodehabitacao, "V14\_QualTipoDeHabitação", caminho)

geraAnaliseQuali(amostra$Habit\_outro, "V15\_Habit\_outro", caminho)

geraAnaliseQuali(amostra$Habit\_alug, "V16\_habit\_alug", caminho)

geraAnaliseQuali(amostra$Comodos, "V17\_Comodos", caminho)

geraAnaliseQuant(amostra$Trabalho, "V18\_trabalho", caminho)

geraAnaliseQuant(amostra$Recursos, "V19\_Recursos", caminho)

geraAnaliseQuant(amostra$Renda\_familiar, "V20\_Renda\_familiar", caminho)

geraAnaliseQuant(amostra$Renda\_div\_pessoa, "V21\_Renda\_div\_Pessoa", caminho)

geraAnaliseQuant(amostra$Renda\_da\_familia, "V22\_Renda\_da\_familia", caminho)

geraAnaliseQuali(amostra$Classe, "V23\_Classe", caminho)

geraAnaliseQuali(amostra$Civil, "V24\_Civil", caminho)

}

print("ANÁLISE DESCRITÍVAS E MEDIDAS DE POSIÇÃO DAS VARIÁVEIS:")

resultado\_1(paste(caminho\_projeto, "\\Analises", sep = ""))

##########################################FUNÇÕES ETAPA 2 ####################################################

analisaP = function(teste) {

if(teste["p.value"] < 0.05){

return( paste("Rejeitado no ",teste["method"]," com p-value de: ", teste["p.value"]) )

}else {

return( paste("Aceito no ",teste["method"]," com p-value de: ", teste["p.value"]) )

}

}

testesModelo = function(lista) {

result = c( analisaP(shapiro.test(lista)), analisaP(ad.test(lista)), analisaP(lillie.test(lista)), analisaP(sf.test(lista)) )

return(result)

}

print("TESTE DE NORMALIDADE DAS VARIÁVEIS:")

print("Variável Idade:")

testesModelo(amostra$Idade)

print("Variável Recursos:")

testesModelo(amostra$Recursos)

print("Variável Casa:")

testesModelo(amostra$Casa)

print("Variável Renda\_familiar:")

testesModelo(amostra$Renda\_familiar)

print("Variável Renda\_div\_pessoa")

testesModelo(amostra$Renda\_div\_pessoa)

print("Variável Renda\_da\_familia")

testesModelo(amostra$Renda\_da\_familia)

##########################################FUNÇÕES ETAPA 3 #######################################################

teste\_hip = function(h0,h1){

teste = t.test(h0,h1)

if(teste$p.value < 0.05){

return(paste("P-value para h0: ",teste$p.value,". Hipotese rejeitada."))

}else {

return(paste("P-value para h0: ",teste$p.value,". Hipotese não rejeitada"))

}

}

print("TESTES DE HIPOTESE DAS VARIÁVEIS E INTERVALOS DE CONFIANÇA:")

print("Variável Idade:")

print("Intervalo de confiança:")

t.test(amostra$Idade)

teste\_hip(amostra$Idade, dadosAaron$Idade)

print("Variável Recursos:")

print("Intervalo de confiança:")

t.test(amostra$Recursos)

teste\_hip(amostra$Recursos, dadosAaron$Recursos)

teste\_hip(amostra$Casa, dadosAaron$Casa)

print("Variável Renda\_familiar:")

print("Intervalo de confiança:")

t.test(amostra$Renda\_familiar)

teste\_hip(amostra$Renda\_familiar, dadosAaron$Renda\_familiar)

print("Variável Renda\_da\_familia:")

print("Intervalo de confiança:")

t.test(amostra$Renda\_da\_familia)

teste\_hip(amostra$Renda\_da\_familia, dadosAaron$Renda\_da\_familia)

print("Variável Renda\_div\_pessoa:")

print("Intervalo de confiança:")

t.test(amostra$Renda\_div\_pessoa)

teste\_hip(amostra$Renda\_div\_pessoa, dadosAaron$Renda\_div\_pessoa)

##########################################FUNÇÕES ETAPA 4 ######################################################

print("TESTE DE CORRELAÇÃO ENTRE NÚMERO DE PESSOAS E A RENDA FAMILIAR:")

cor.test(amostra$Renda\_div\_pessoa, amostra$Renda\_familiar)

print("REGRESSÃO LINEAR ENTRE NÚMERO DE PESSOAS E A RENDA FAMILIAR:")

regressaoLinear = lm(amostra$Renda\_familiar ~ amostra$Renda\_div\_pessoa )

summary(regressaoLinear)

plot(amostra$Renda\_div\_pessoa,amostra$Renda\_familiar)

grid(regr)

abline(regrecaoLinear)

regr

# Conclusão

A seguinte análise demonstrou principalmente a importância de uma amostra bem coletada, visto que alguns dos valores não passaram nos testes de hipótese como também as divergências encontradas nos testes de regressão e correlação entre as diferentes amostras criadas nas outras aplicações da análise. Apesar disso, quanto aos dados qualitativos, a amostra com 60 observações pode ser bastante representativa, sendo relativamente equivalente ao banco de dados os resultados já apresentados ao longo desta análise.

O Software R foi indispensável para realização da análise, visto sua natureza, possibilitando a automatização de tarefas repetitivas, mas ainda assim, o software não pareceu estar alinhado à certos padrões adotados em diversas linguagens de programação, dificultando algumas operações e/ou tornando-as relativamente improdutivas. Apesar dos pontos negativos, a ferramenta se mostrou útil e eficiente, com bastantes funcionalidades ainda a se explorar.