

Lista 05 - MQ

Martha Gaudencio

2026-02-01

```
# MQ101 - Lista 05
# Nome: <Martha Gaudencio da Silva>
# RA: <21201510158>
# Data: <2026-01-27>

#Lista de Exercícios 5 - Revisão geral, Teste de Hipóteses e Regressão

knitr:::opts_chunk$set(
  echo = TRUE,
  message = FALSE,
  warning = FALSE,
  dpi = 96
)

library(tidyverse)
```

```
## — Attaching core tidyverse packages ————— tidyverse 2.0.0 —
## ✓ dplyr     1.1.4      ✓ readr     2.1.6
## ✓forcats   1.0.1      ✓ stringr   1.6.0
## ✓ ggplot2   4.0.1      ✓ tibble    3.3.1
## ✓ lubridate 1.9.4      ✓ tidyr    1.3.2
## ✓ purrr    1.2.1

## — Conflicts ————— tidyverse_conflicts() —
## ✘ dplyr::filter() masks stats::filter()
## ✘ dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to be
## come errors
```

```
library(broom)

set.seed(123)

#install.packages("knitr")
install.packages("tidyverse")
```

```
## Warning: o pacote 'tidyverse' está em uso e não será instalado
```

```
install.packages("ggplot2")
```

```
## Warning: o pacote 'ggplot2' está em uso e não será instalado
```

```
install.packages("tibble")
```

```
## Warning: o pacote 'tibble' está em uso e não será instalado
```

```
library("tibble")
library("ggplot2")

n <- 400

dados <- tibble(
  id      = 1:n,
  idade   = round(rnorm(n, mean = 40, sd = 12)),
  sexo    = sample(c("F", "M"), n, replace = TRUE, prob = c(0.55, 0.45)),
  renda   = round(rlnorm(n, meanlog = log(2500), sdlog = 0.5), 0),
  escolarid = sample(c("fundamental", "medio", "superior"),
                      n, replace = TRUE, prob = c(0.30, 0.40, 0.30)),
  ideologia = round(runif(n, 0, 10), 0), # 0 = esquerda, 10 = direita
  apoio_gov = rbinom(n, 1, plogis(-1 + 0.015*(idade - 40) +
                                    0.4*(sexo == "F") +
                                    0.5*(renda > 3000))),
  satisf_gov = pmin(pmax(
    round(3 + 2*apoio_gov + 0.001*(renda - 2500) +
          rnorm(n, 0, 2), 0), 0), 10),
  # 0 = não participou, 1 = participou de protestos

  protesto = rbinom(n, 1, plogis(-2 + 0.3*(ideologia <= 4) - 0.2*apoio_gov))
)

glimpse(dados)
```

```
## Rows: 400
## Columns: 9
## $ id      <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, ...
## $ idade   <dbl> 33, 37, 59, 41, 42, 61, 46, 25, 32, 35, 55, 44, 45, 41, 33, ...
## $ sexo    <chr> "F", "F", "F", "F", "M", "F", "F", "M", "M", "F", "M", ...
## $ renda   <dbl> 4277, 2466, 2459, 1171, 3712, 2250, 1800, 1234, 2152, 1635, ...
## $ escolarid <chr> "medio", "medio", "medio", "superior", "superior", ...
## $ ideologia <dbl> 7, 6, 2, 9, 8, 3, 2, 7, 7, 5, 9, 7, 2, 0, 9, 1, 3, 4, 6, 3, ...
## $ apoio_gov <int> 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, ...
## $ satisf_gov <dbl> 4, 4, 2, 0, 3, 2, 7, 6, 6, 7, 4, 9, 4, 6, 0, 2, 6, 4, 8, ...
## $ protesto <int> 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, ...
```

```
# Exercício 1 - Exploração descritiva e gráficos básicos
```

#1.1. Calcule, para a amostra:

```
summary(dados$idade)
```

```
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
## 10.0   33.0   40.0   40.2   48.0   79.0
```

```
summary(dados$renda)
```

```
##      Min. 1st Qu. Median   Mean 3rd Qu.   Max.  
##      681    1772   2504   2867   3466   9604
```

```
summary(dados$satisf_gov)
```

```
##      Min. 1st Qu. Median   Mean 3rd Qu.   Max.  
##      0.000  2.000  4.000  4.263  6.000 10.000
```

#Para idade - Média: 39.79, Mediana: 39, 1º quartil: 32, 3º quartil: 48.
#Para renda - Média: 2730, Mediana: 2429, 1º quartil: 1781, 3º quartil: 3355.
#Para satisfação governo - Média: 3.897, Mediana: 4, 1ºq: 2, 3º q: 5.

#Desvio-padrão

```
sd(dados$idade, na.rm = TRUE)
```

```
## [1] 11.63932
```

```
sd(dados$renda, na.rm = TRUE)
```

```
## [1] 1588.662
```

```
sd(dados$satisf_gov, na.rm = TRUE)
```

```
## [1] 2.723442
```

#Desvios-padrão de idade: 11.91593; renda: 1349.426; satisfação: 2.38037

#Proporção de respondentes por categoria de sexo e escolaridade;

```
prop.table(table(dados$sexo))
```

```
##  
##      F      M  
## 0.55  0.45
```

```
prop.table(table(dados$escolarid))
```

```
##  
## fundamental     medio     superior  
##      0.2975     0.4050     0.2975
```

```
#Fundamental: 0.2925, médio: 0.3700      superior: 0.3375
#Sexo feminino: 0.55, Masculino: 0,44
```

#Proporção de respondentes com apoio_gov = 1.

```
mean(dados$apoio_gov == 1, na.rm = TRUE)
```

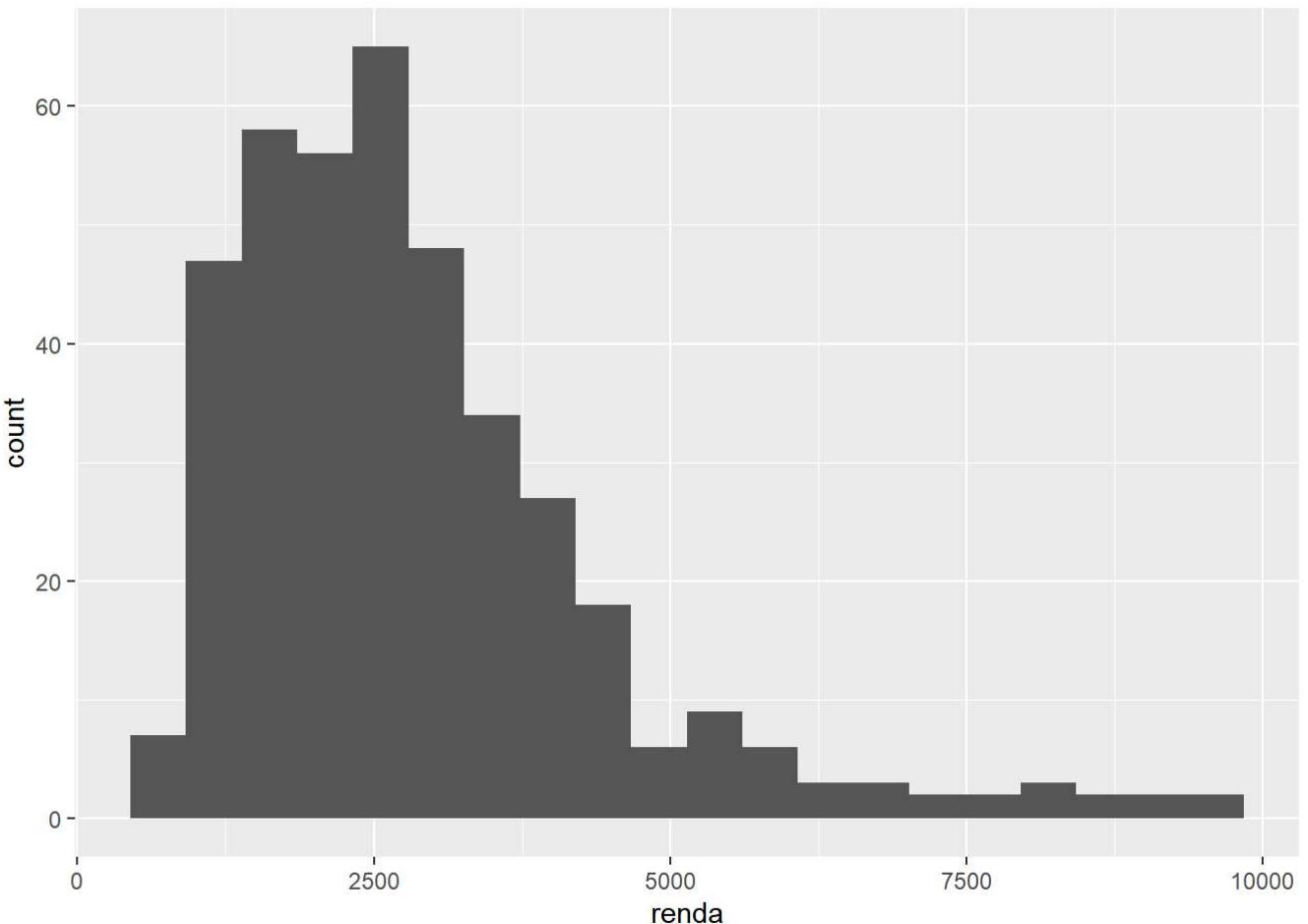
```
## [1] 0.375
```

#Proporção igual a 0.36

#1.2. Construa os seguintes gráficos:

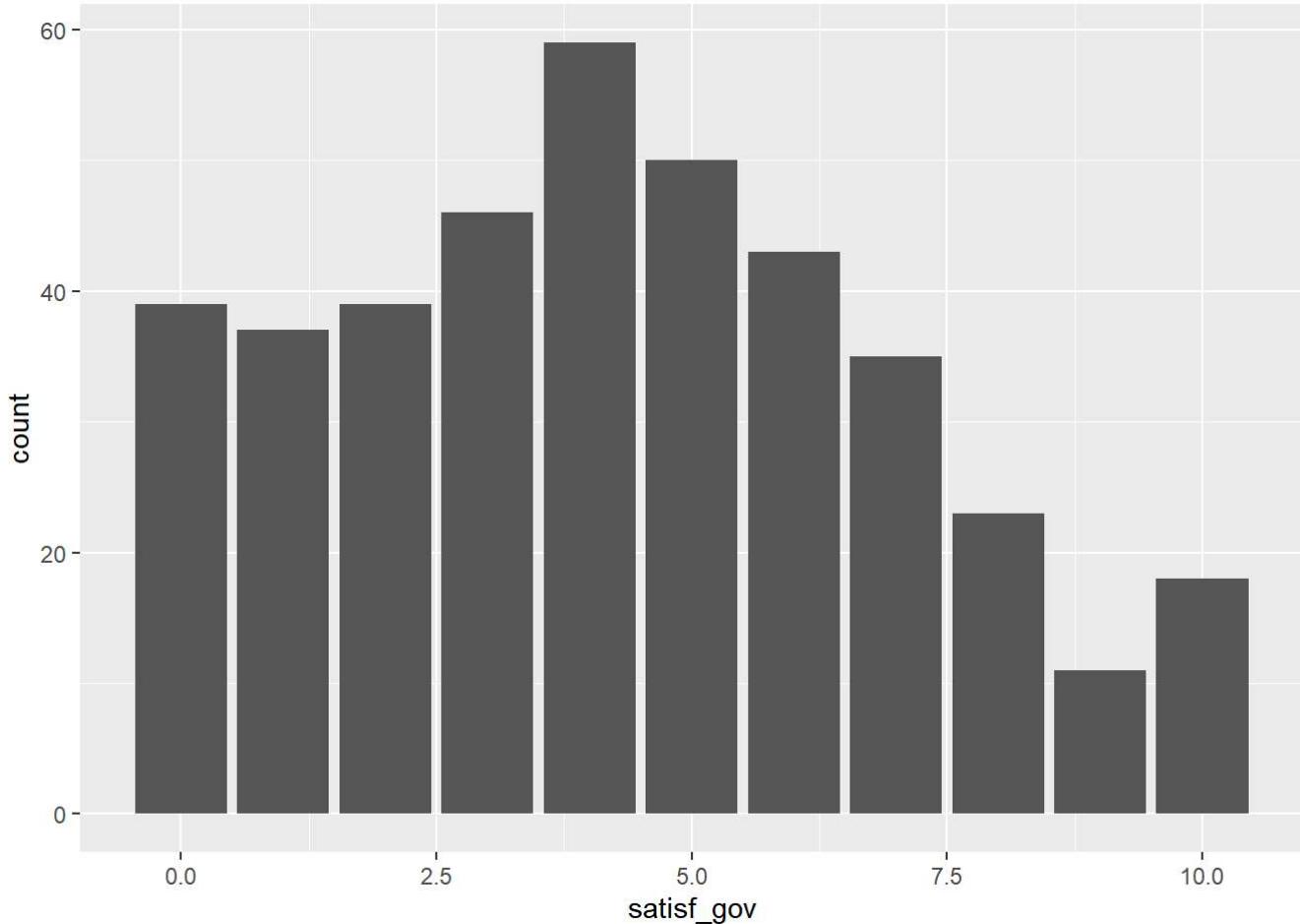
#Um histograma de renda (use pelo menos 20 quebras);

```
ggplot(dados, aes(x = renda)) +
  geom_histogram(bins = 20)
```

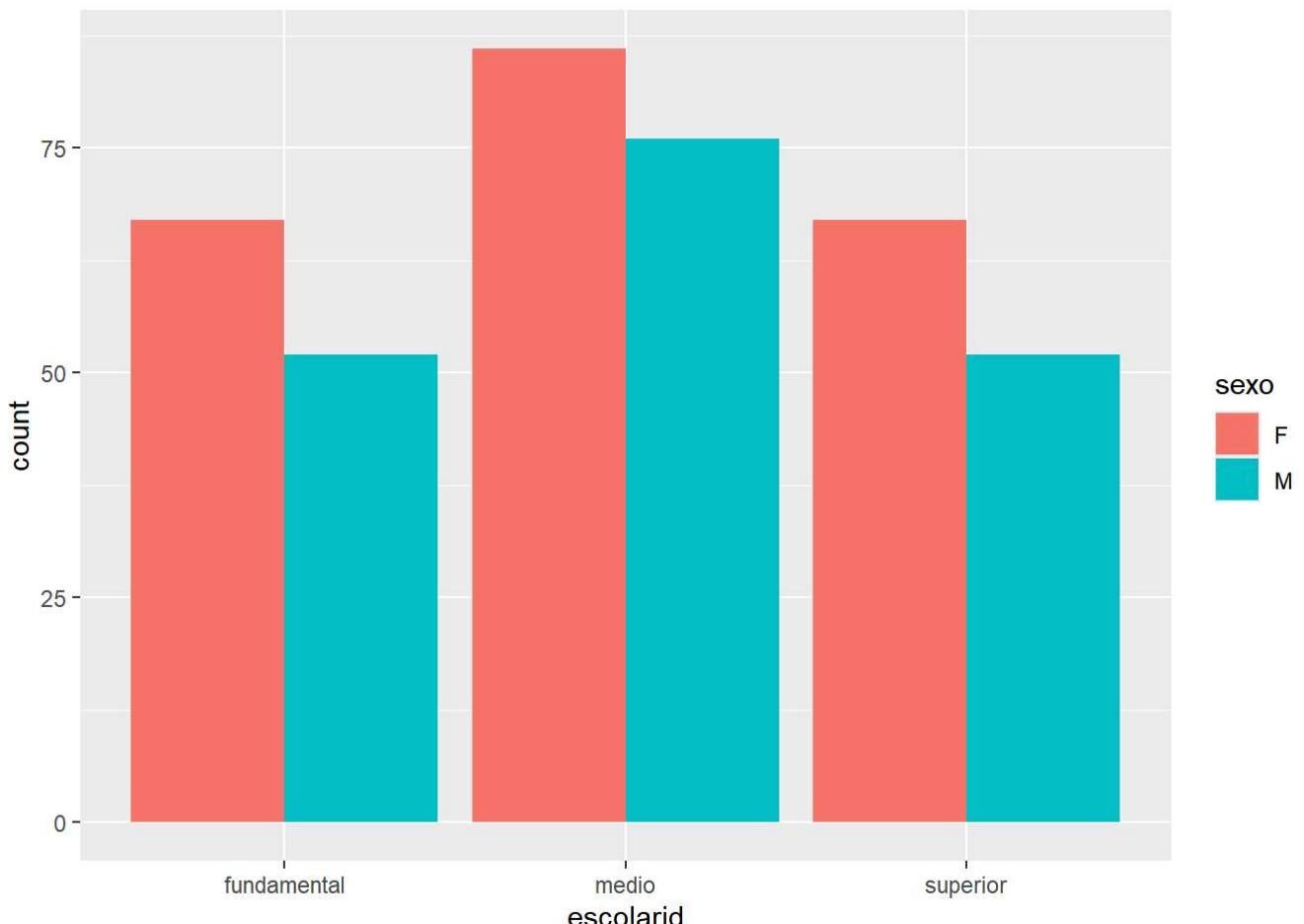


```
ggplot(dados, aes(x = satisf_gov)) +
  geom_bar(bins = 20)
```

```
## Warning in geom_bar(bins = 20): Ignoring unknown parameters: `bins`
```



```
ggplot(dados, aes(x = escolarid, fill = sexo)) +  
  geom_bar(position = "dodge")
```



#Em poucas linhas (em texto, na sua resposta), comente:

#Se a distribuição de renda parece simétrica ou assimétrica;

#A distribuição de renda parece assimétrica pois há variação na distribuição dos dados com concentração em torno de 2500.

#se satisf_gov parece concentrada em valores baixos, médios ou altos;

#Está concentrada em valores médios, com os maiores valores em torno de 4 e 5.

#como se distribuem sexo e escolaridade na amostra.

#O sexo feminino apresenta mais escolaridade que o masculino, nos três níveis de escolaridade apresentados. Os homens apresentam maior concentração no nível de ensino médio completo e as mulheres em ensino superior completo.

#Exercício 2 – Escolha de testes estatísticos

#a) tipo de sexo: categórico - qualitativa

#tipo de apoio_gov: categórico - qualitativa

#teste bivariado sugerido: análise tabular e qui-quadrado

#b) escolaridade: categórico - qualitativa

#apoio_gov: categórico - qualitativa

#teste bivariado sugerido: análise tabular e qui-quadrado

#c) satisf_gov: quantitativa discreta

#apoio_gov (0/1): qualitativa categórica

#teste bivariado sugerido: diferença de médias

#d) satisf_gov: quantitativa discreta

#renda: quantitativa contínua

#teste bivariado sugerido: correção ou regressão simples

#e) satisf_gov: quantitativa discreta

#ideologia: quantitativa contínua

#teste bivariado sugerido: correção ou regressão simples

#Exercício 3 – Teste qui-quadrado: sexo e apoio ao governo

```
tab_sexo_apoio <- table(dados$sexo, dados$apoio_gov)
```

```
tab_sexo_apoio
```

```
##
```

```
##      0   1
```

```
##  F 142  78
```

```
##  M 108  72
```

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

```
##          0      1
##  F  0.568 0.520
##  M  0.432 0.480
```

#3.2. Formule as hipóteses:

#H0: não há associação entre sexo e apoio_gov na população;

#H1: há associação entre sexo e apoio_gov.

#3.3. Aplique o teste de qui-quadrado em R:

```
chisq.test(tab_sexo_apoio)
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: tab_sexo_apoio
## X-squared = 0.68956, df = 1, p-value = 0.4063
```

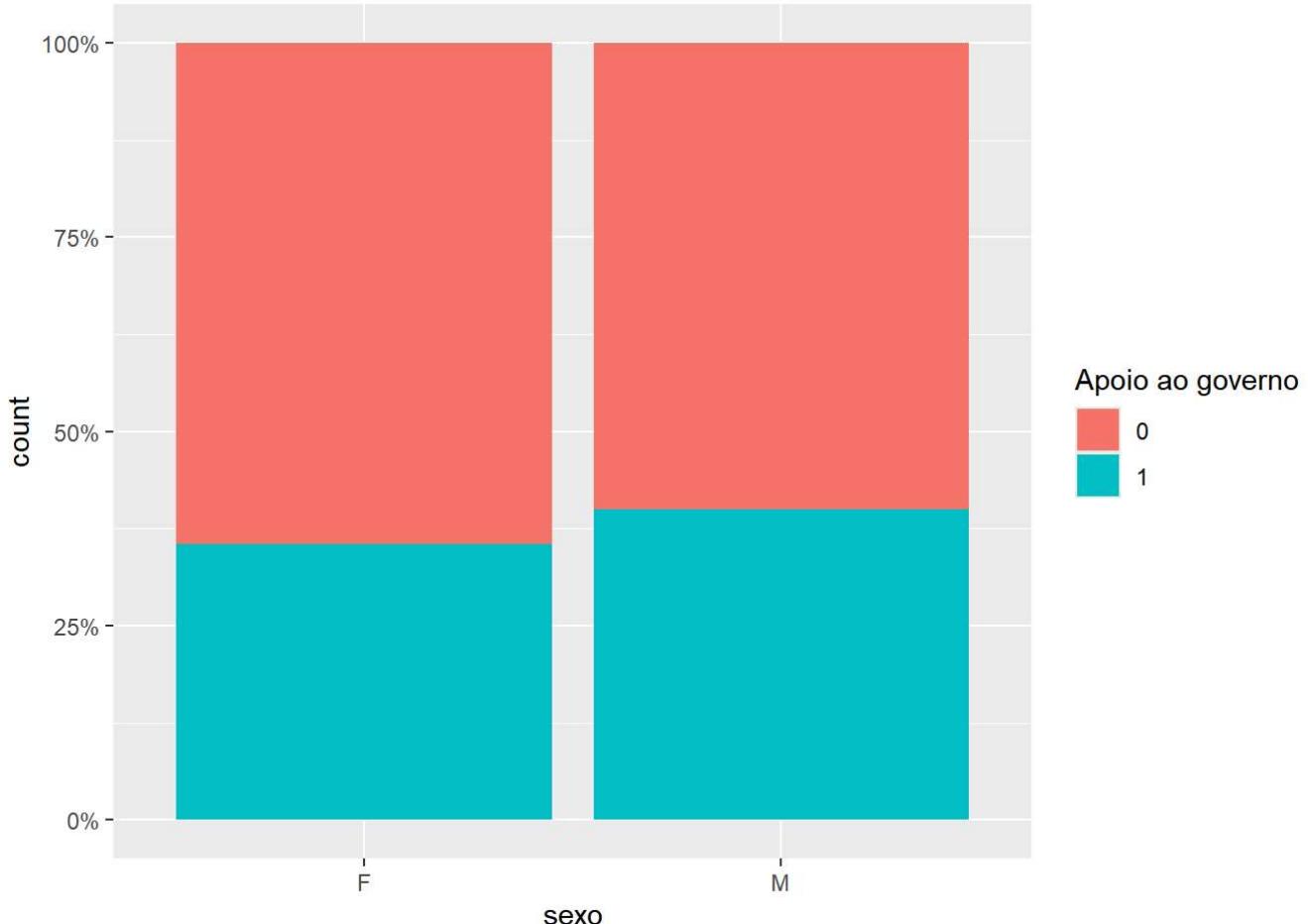
#3.4. Em texto, interprete:

O valor do qui-quadrado é 0,68956. O grau de liberdade é 1, valor-p: 0.4063.

Logo, não se rejeita a H0 pois o valor não é significativo. Assim, não há uma associação entre o sexo e o apoio ao governo na população estudada.

#3.5 - Gráfico

```
ggplot(dados, aes(x = sexo, fill = factor(apoio_gov))) +
  geom_bar(position = "fill") +
  scale_y_continuous(labels = scales::percent) +
  labs(fill = "Apoio ao governo")
```



#É coerente pois não há muita discrepância no apoio ao governo dos sexos # feminino e masculino.

#Exercício 4 – Diferença de médias: renda entre apoiadores e não apoiadores

#4.1. Produza um resumo numérico de renda por apoio_gov #(média, desvio-padrão, tamanho de cada grupo).

```
dados %>%
  group_by(apoio_gov) %>%
  summarise(
    media_renda = mean(renda),
    sd_renda    = sd(renda),
    n           = n()
  )
```

```
## # A tibble: 2 × 4
##   apoio_gov media_renda sd_renda     n
##       <int>      <dbl>     <dbl> <int>
## 1         0      2772.    1577.    250
## 2         1      3025.    1601.    150
```

#4.2. Formule as hipóteses:

```
#H0: μ apoia = μ não apoia
#H1: μ apoia ≠ μ não apoia
```

#4.3. Aplique o teste t para diferença de médias:

```
t.test(renda ~ apoio_gov, data = dados)
```

```
##
## Welch Two Sample t-test
##
## data: renda by apoio_gov
## t = -1.5394, df = 310.01, p-value = 0.1247
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal
## to 0
## 95 percent confidence interval:
## -576.65627 70.40827
## sample estimates:
## mean in group 0 mean in group 1
## 2772.356 3025.480
```

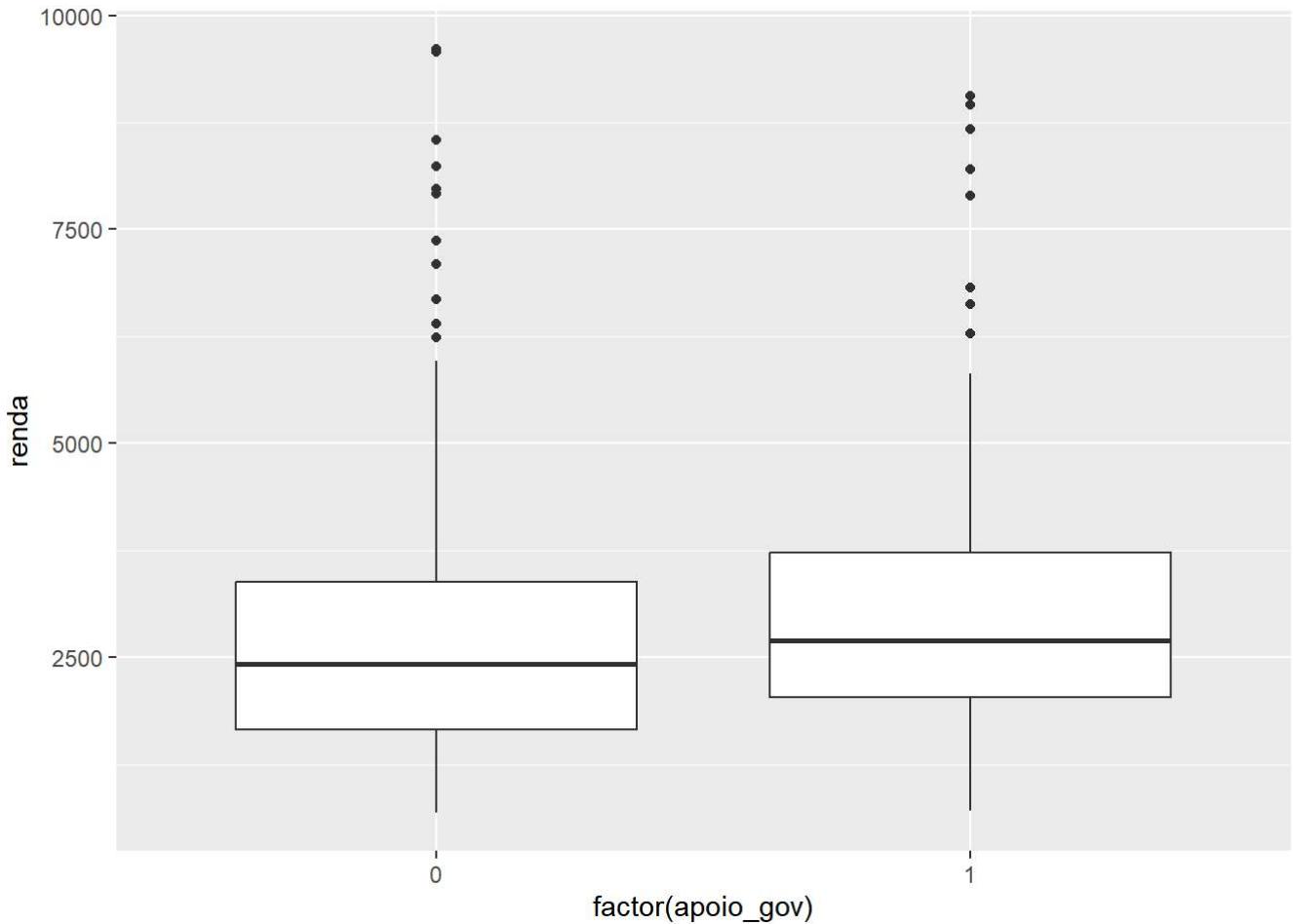
#4.4. Em texto, interprete:

```
#a estimativa da diferença de médias;
#o intervalo de confiança de 95%;
#o valor-p;
#a conclusão sobre H0 a 5%.
```

#A estimativa de diferença de médias é de -253.124, valor-p = 0.1247 e
#intervalo de confiança vai de -576.65627 a 70.40827, incluindo o zero,
#então não há uma diferença real entre os grupos. Não rejeitamos então H0.

#4.5. Construa um boxplot

```
ggplot(dados, aes(x = factor(apoio_gov), y = renda)) +
  geom_boxplot()
```



```
#O gráfico confirma o resultado do teste de hipóteses pois não há
#uma diferença significativa entre os dois grupos de apoio ao governo
#no que tange a renda.
```

#Exercício 5 – Correlação: renda, ideologia e satisfação com o governo

#5.1. Calcule a matriz de correlações de Pearson

```
dados %>%
  select(renda, ideologia, satisf_gov) %>%
  cor(use = "complete.obs")
```

```
##                   renda   ideologia  satisf_gov
## renda      1.00000000  0.02053989  0.59336969
## ideologia  0.02053989  1.00000000 -0.07964692
## satisf_gov 0.59336969 -0.07964692  1.00000000
```

#5.2. Para cada par de variáveis, com base na matriz:

#indique o sinal da correlação (positivo/negativo);

#comente, de forma qualitativa, a força da associação (fraca, moderada, forte),
#usando como referência aproximada:

$|r| \approx 0.1$ (fraco),
 $|r| \approx 0.3$ (moderado),
 $|r| \geq 0.5$ (mais forte).

#Renda e Ideologia: positiva; associação fraca;
#Renda e satisfação gov: positiva; associação forte;
#Ideologia e Satisfação gov: negativa; associação fraca.

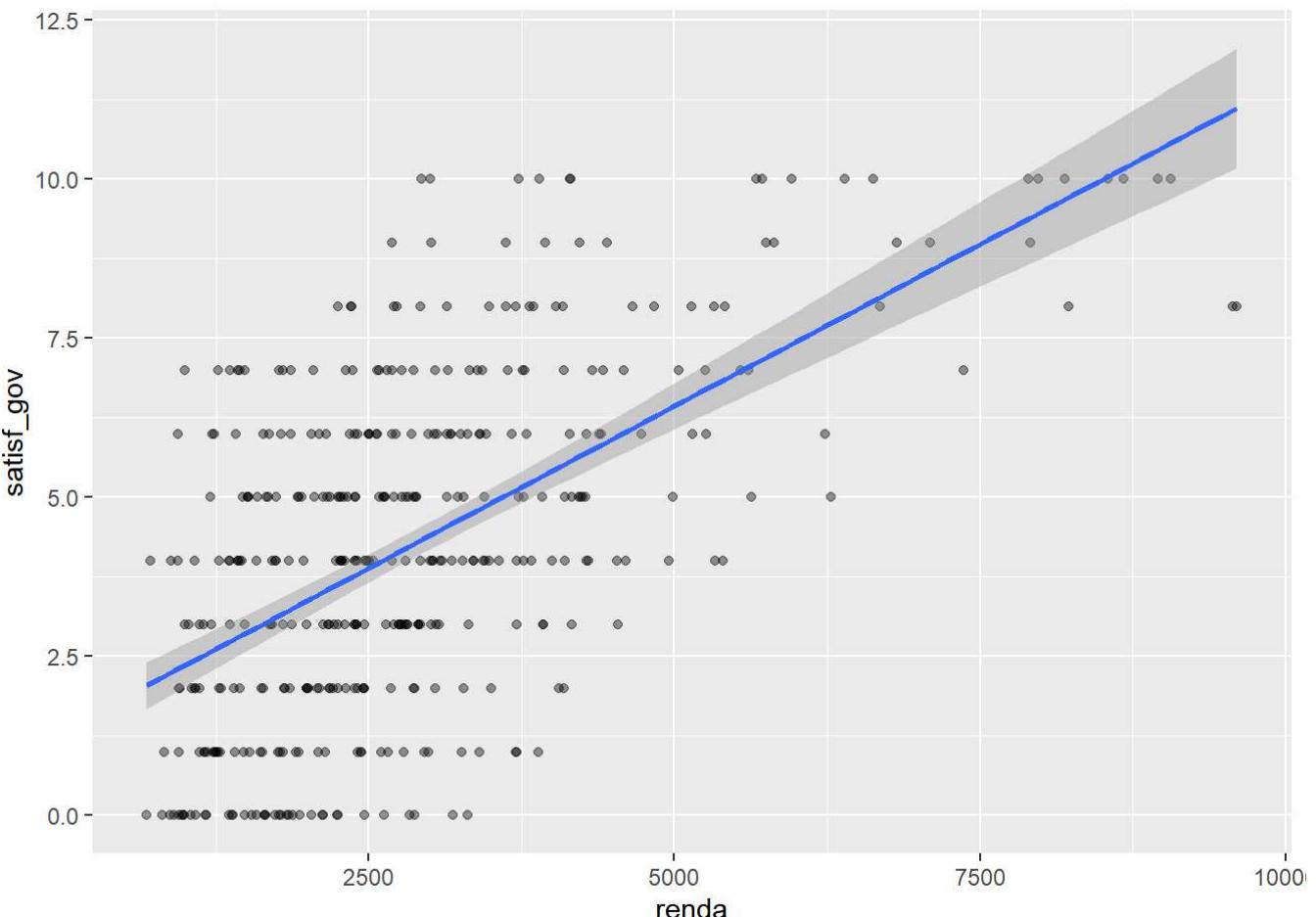
#5.3 Gráfico de dispersão

```
ggplot(dados, aes(x = renda, y = satisf_gov)) +  

  geom_point(alpha = 0.4) +  

  geom_smooth(method = "lm", se = TRUE)
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



#5.4. Comente o padrão visual do gráfico

#O gráfico representa a associação forte, uma vez que o gráfico tem uma #tendência positiva mostrando que o aumento da renda converge com aumento #da satisfação com o governo.

#Exercício 6 – Regressão Linear simples: satisfação e renda

#6.1

```
mod1 <- lm(satisf_gov ~ renda, data = dados)
summary(mod1)
```

```
##
## Call:
## lm(formula = satisf_gov ~ renda, data = dados)
##
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -4.7149 -1.6354 -0.1648  1.5664  5.6727 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 1.346e+00  2.267e-01   5.938  6.3e-09 ***
## renda       1.017e-03  6.917e-05  14.706 < 2e-16 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 2.195 on 398 degrees of freedom
## Multiple R-squared:  0.3521, Adjusted R-squared:  0.3505 
## F-statistic: 216.3 on 1 and 398 DF,  p-value: < 2.2e-16
```

#6.2. Em texto, interprete:

#Há uma relação positiva entre renda e satisfação com o governo, comprovada #pela regressão Linear simples. O coeficiente angular indica que para cada #aumento de 1.000 na renda, aumenta em média 1 ponto na satisfação com o governo. #Ainda assim, mesmo que a renda importe, outros fatores podem influenciar #nessa satisfação.

#6.3 Broom

```
tidy(mod1)
```

```
## # A tibble: 2 × 5
##   term      estimate std.error statistic p.value
##   <chr>      <dbl>     <dbl>      <dbl>    <dbl>
## 1 (Intercept) 1.35      0.227      5.94  6.30e- 9
## 2 renda       0.00102   0.0000692   14.7   2.08e-39
```

#As estimativas (0.0010) mostraram que a cada aumento na renda, a satisfação sobe.
#O erro-padrão (0.000069) é bem pequeno, indicando que o cálculo é preciso.
#A estatística t deu um valor alto (14.70) e o valor-p é praticamente zero,
#o que prova que o resultado possui confiabilidade.

#6.4 Gráfico

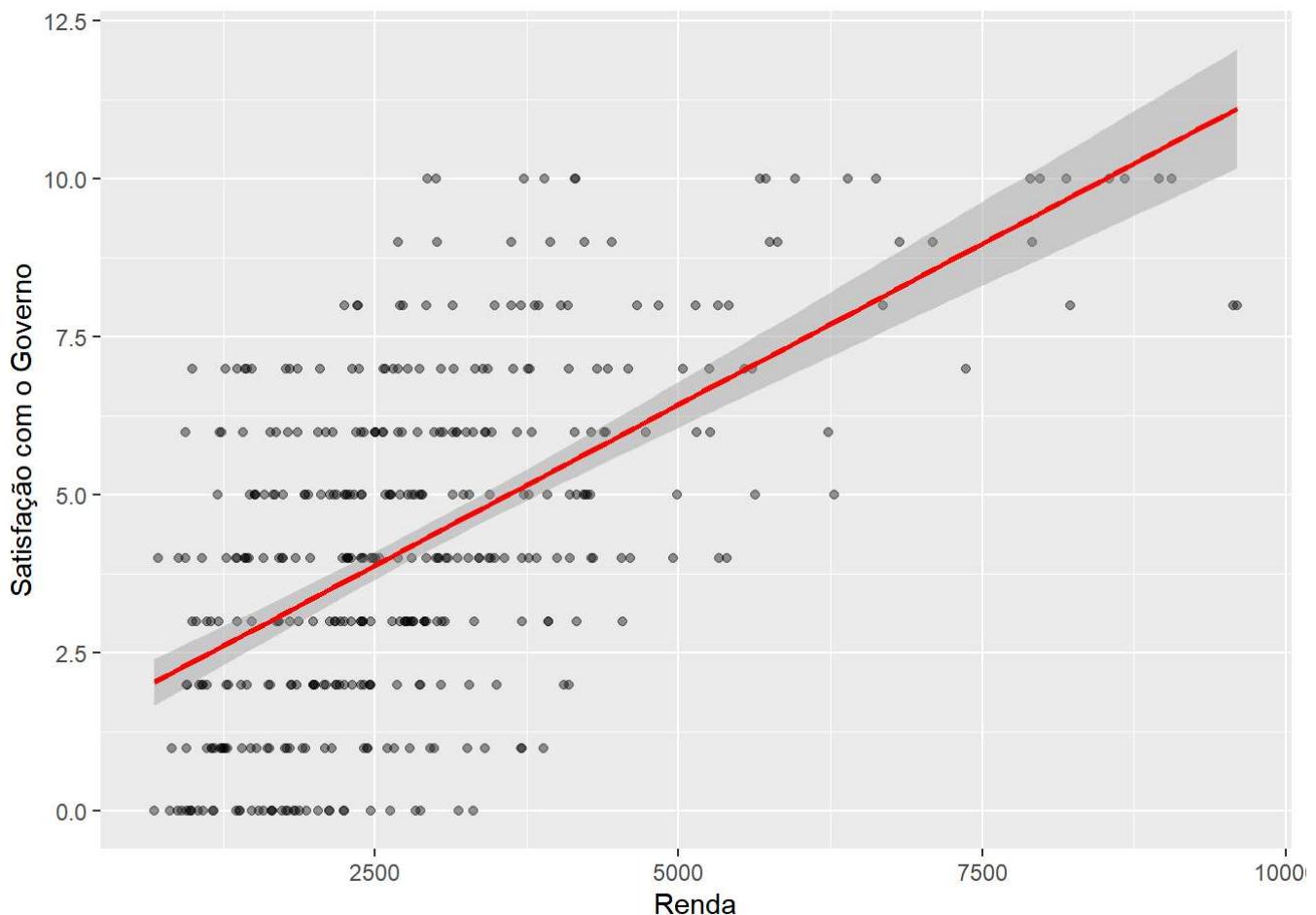
```
ggplot(dados, aes(x = renda, y = satisf_gov)) +  

  geom_point(alpha = 0.4) +  

  geom_smooth(method = "lm", se = TRUE, color = "red") +  

  labs(x = "Renda", y = "Satisfação com o Governo")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



#O gráfico confirma as informações anteriores. A reta sobe em direção à direita, em uma correlação positiva, o que combina com o sinal positivo do coeficiente (0.0010). O valor da renda é significativo porque o p-valor é pequeno, confirmando que a relação entre ganhar mais e estar mais satisfeita é real na amostra selecionada e se reflete no gráfico.

#Exercício 7 – Diagnóstico simples do modelo

#7.1

```
dados_diag <- augment(mod1)  

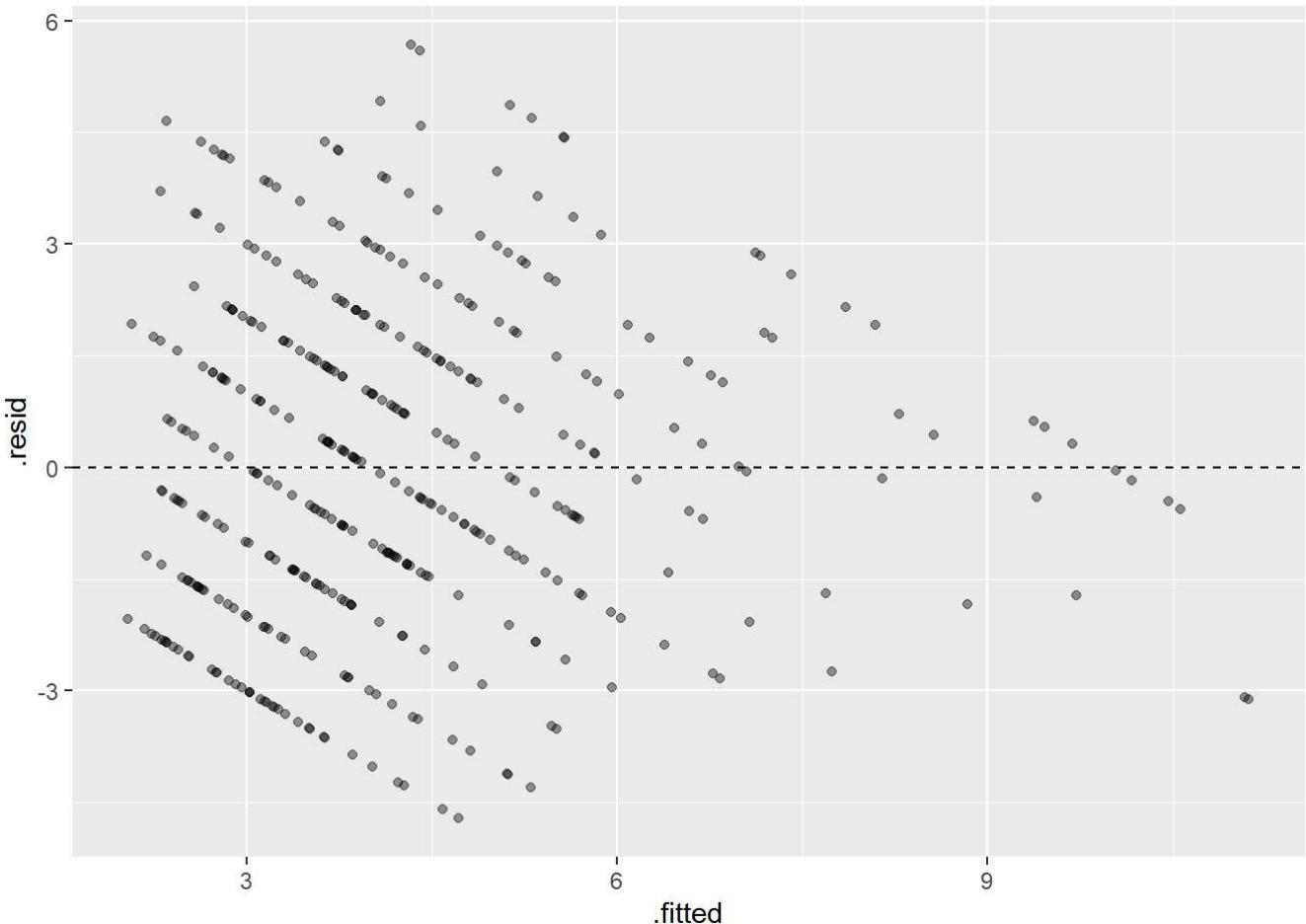
glimpse(dados_diag)
```

```
## Rows: 400
## Columns: 8
## $ satisf_gov <dbl> 4, 4, 2, 0, 3, 2, 7, 6, 6, 7, 4, 9, 4, 6, 0, 2, 6, 4, 8...
## $ renda      <dbl> 4277, 2466, 2459, 1171, 3712, 2250, 1800, 1234, 2152, 1635, ...
## $ .fitted     <dbl> 5.696488, 3.854315, 3.847195, 2.537024, 5.121763, 3.634597, ...
## $ .resid      <dbl> -1.6964884, 0.1456848, -1.8471947, -2.5370241, -2.1217629, ...
## $ .hat        <dbl> 0.004473474, 0.002659902, 0.002665529, 0.005357314, 0.00320...
## $ .sigma      <dbl> 2.196038, 2.197683, 2.195734, 2.193984, 2.195106, 2.196159, ...
## $ .cooksdi    <dbl> 1.348244e-03, 5.890268e-06, 9.489750e-04, 3.617343e-03, 1.5...
## $ .std.resid   <dbl> -0.77464598, 0.06646169, -0.84269626, -1.15896373, -0.96821...
```

#7.2.

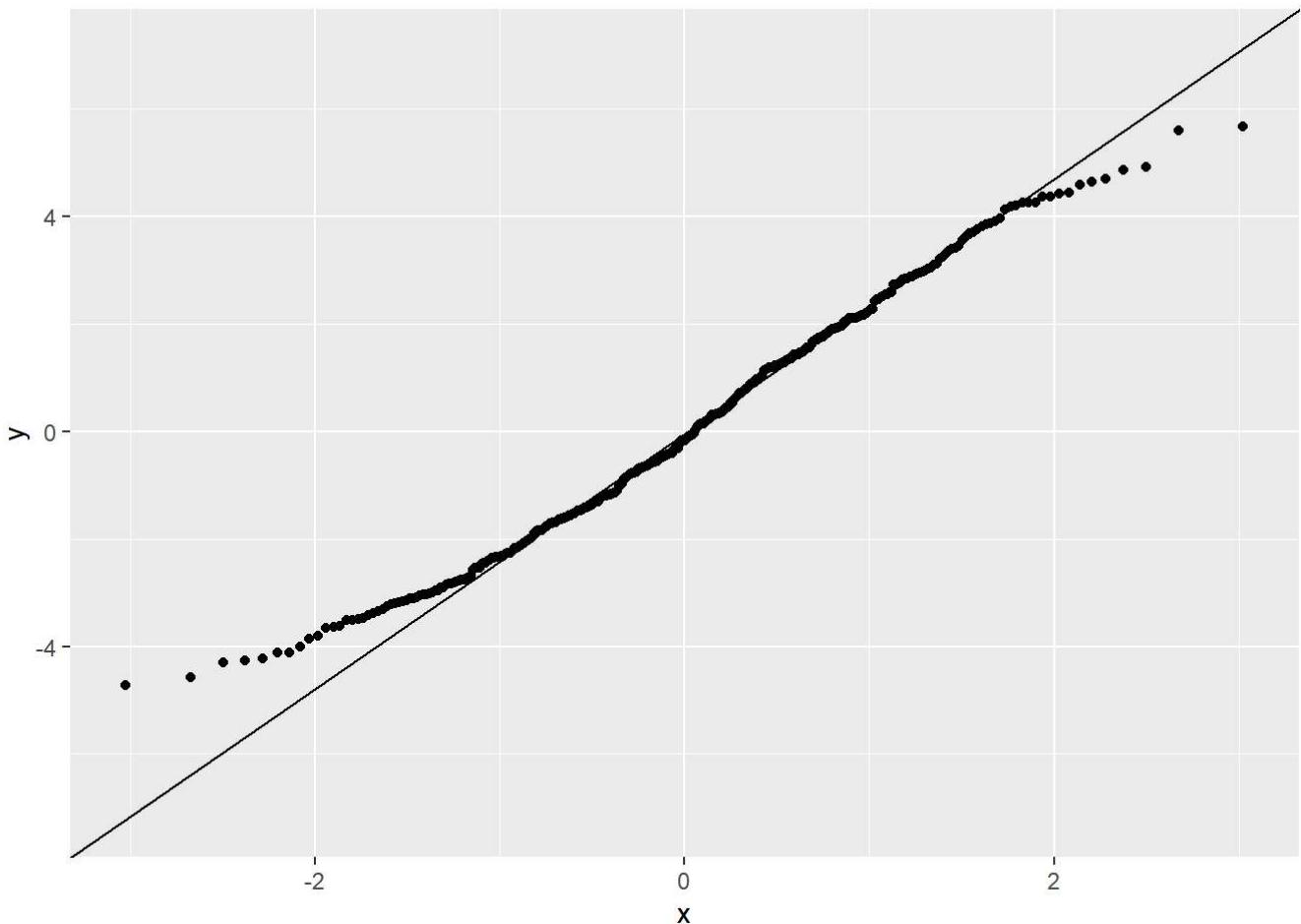
#Gráfico de resíduos vs valores ajustados

```
ggplot(dados_diag, aes(x = .fitted, y = .resid)) +
  geom_point(alpha = 0.4) +
  geom_hline(yintercept = 0, linetype = "dashed")
```



#Gráfico QQ-plot dos resíduos.

```
ggplot(dados_diag, aes(sample = .resid)) +
  stat_qq() +
  stat_qq_line()
```



#7.3 O primeiro gráfico mostra os pontos mais espalhados ao redor da Linha zero, sem formar um desenho de "funil" claro. Isso indica que não há indícios fortes de heterocedasticidade. Já no QQ-plot, os pontos seguem de perto a linha diagonal, o que mostra que os resíduos têm uma distribuição muito próxima da Normal.

#7.4 É válido olhar para esses gráficos porque eles permitem confirmar a validade e confiança dos dados, além de confirmar se não há dados 'viciados' probabilisticamente nas amostras.

#Exercício 8 – Regressão com variável dummy e diferença de médias

#8.1

```
mod2 <- lm(satisf_gov ~ apoio_gov, data = dados)
summary(mod2)
```

```

## 
## Call:
## lm(formula = satisf_gov ~ apoio_gov, data = dados)
## 
## Residuals:
##    Min     1Q Median     3Q    Max 
## -5.680 -1.680 -0.412  1.588  6.588 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept)  3.4120    0.1578  21.624 <2e-16 ***
## apoio_gov    2.2680    0.2577   8.802 <2e-16 ***
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.495 on 398 degrees of freedom
## Multiple R-squared:  0.1629, Adjusted R-squared:  0.1608 
## F-statistic: 77.48 on 1 and 398 DF,  p-value: < 2.2e-16

```

#8.2 e 8.3 As médias de satisfação são 3,41 para quem não apoia e 5,68 para quem apoia o governo. Esses valores batem com o modelo: o intercepto é a média do primeiro grupo (3,41) e o coeficiente é a diferença de 2,26 entre eles.
#Somando os dois, o valor é o da média do grupo que apoia.

#8.4

```
t.test(satisf_gov ~ apoio_gov, data = dados)
```

```

## 
## Welch Two Sample t-test
## 
## data: satisf_gov by apoio_gov
## t = -8.9438, df = 329.69, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -2.766844 -1.769156
## sample estimates:
## mean in group 0 mean in group 1 
##            3.412            5.680

```

#O valor-p de apoio_gov na regressão equivale ao do t.test e as médias também batem (3,41 e 5,68). Isso acontece porque a regressão com uma variável dummy e o teste t são caminhos diferentes para chegar ao mesmo Lugar:
#ambos provam que a diferença de satisfação entre os grupos é real e estatisticamente significante.