UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA - IME
DEPARTAMENTO DE ESTATÍSTICA – DEST
Disciplina: Introdução aos Modelos Lineares – MAT D41

ANÁLISE DE REGRESSÃO

VITOR LUIS DOS SANTOS SANTANA FREITAS
DANILO CADENA LIMA
LUCA FERREIRA BARBOZA

SALVADOR
2025

UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA - IME
DEPARTAMENTO DE ESTATÍSTICA – DEST
Disciplina: Introdução aos Modelos Lineares – MAT D41

ANÁLISE DE REGRESSÃO

VITOR LUIS DOS SANTOS SANTANA FREITAS
DANILO CADENA LIMA
LUCA FERREIRA BARBOZA

Relatório apresentado como requisito de avaliação
para a disciplina Introdução aos Modelos Lineares – MAT D41,
ofertada no semestre 2025.1
Docente: Nívea Bispo

SALVADOR
2025

1. INTRODUÇÃO

Desenvolvido como parte das atividades avaliativas da disciplina de Introdução aos Modelos Lineares, este estudo explora as relações entre características de imóveis e seus respectivos valores de aluguel.

A base retrata variáveis como, cidade, aluguel, IPTU, número de garagens, número de quartos, condomínio e área e tem como principal objetivo observar, dentre essas variáveis, qual delas ajuda a explicar melhor o valor do aluguel. Para todos os testes, usamos um nível de significância de 5%.

2. METODOLOGIA

A fim de investigar os fatores associados ao valor do aluguel nas cidades de Campinas e São Paulo, foi utilizado o Modelo de Regressão Linear Simples. Inicialmente, foi realizada uma análise exploratória dos dados por meio de boxplots e gráficos de dispersão, com o objetivo de entender as distribuições das variáveis e identificar possíveis padrões.

Posteriormente, foram utilizadas matrizes de correlação para selecionar, por cidade, as variáveis com uma maior correlação linear com a variável Aluguel. Com base nas informações obtidas, as suas suposições básicas - linearidade, independência, homogeneidade e normalidade dos resíduos - foram avaliadas por meio dos gráficos diagnósticos. Essa abordagem permitiu descrever qual variável influencia no valor do aluguel de imóveis nestas cidades e avaliar o nível de confiabilidade do modelo para essa análise.

3. RESULTADOS E CONSIDERAÇÕES FINAIS

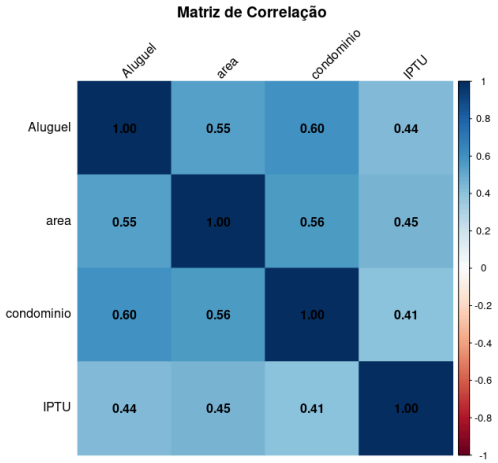

Figura 1: Matriz de correlação das variáveis Aluguel, área, condomínio e IPTU.

Observou-se, na figura 1, que as variáveis área e condomínio apresentaram as maiores correlações positivas com o aluguel. Porém não são correlações consideradas fortes.

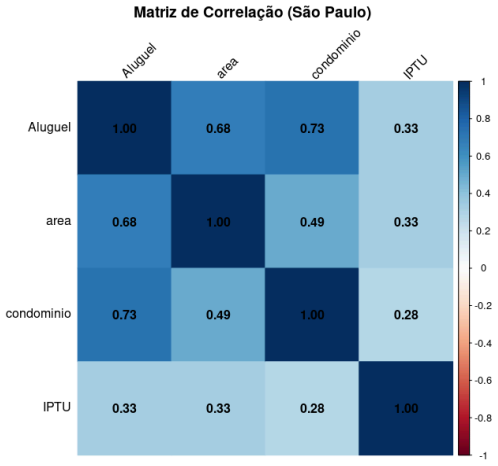

Figura 2: Matriz de correlação das variáveis Aluguel, área, condomínio e IPTU. Somente na cidade de São Paulo.

Na figura 2, pode ser visto que área e condomínio continuam sendo as variáveis com as maiores correlações com o aluguel, porém as correlações aumentaram, principalmente condomínio que ficou acima de 0,7.

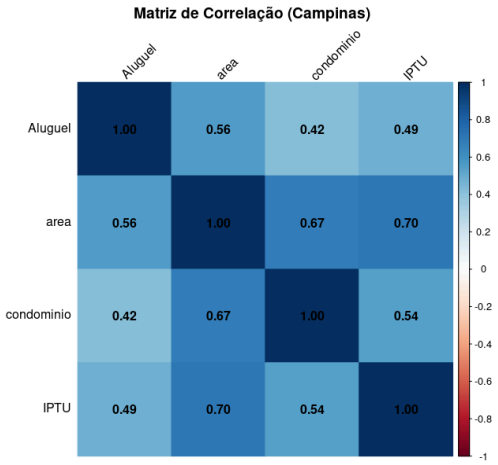

Figura 3: Matriz de correlação das variáveis Aluguel, área, condomínio e IPTU. Somente na cidade de Campinas.

Na figura 3, pode ser visto que agora a área e IPTU são as variáveis com as maiores correlações com o aluguel, as correlações nesse caso são pequenas, assim como no geral, e a maior correlação do aluguel é com a área.

Com base nas análises de correlação preliminares, dois modelos de regressão linear foram desenvolvidos: um para prever o aluguel em São Paulo utilizando o condomínio, e outro para estimar o aluguel em Campinas com base na área do imóvel.

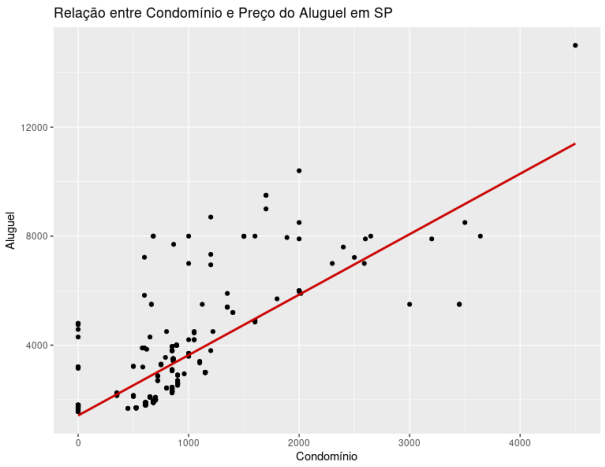

Figura 4: Gráfico de dispersão entre o condomínio e o aluguel na cidade de São Paulo

Tabela 1: Resultados da Regressão Linear para Aluguel em São Paulo

	Coeficiente (β)	Erro Padrão	Estatística t	Valor-p
(Intercepto)	1420.95	111.12	12.79	< 0.001 (***)
Condomínio	2.22	0.11	20.25	< 0.001 (***)

R^2 (Ajustado): 0.5327

Observa-se na figura 4 que os dados possuem uma correlação positiva, com uma tendência de aumento do aluguel conforme o valor do condomínio cresce. No entanto, existem vários pontos que não seguem a reta, ou estão consideravelmente afastados dela, o que sugere a possibilidade de que o modelo linear simples não captura toda a complexidade da relação entre essas variáveis. Além disso, o R^2 ajustado aponta que 53% da variação do aluguel em São Paulo é explicada pelo condomínio

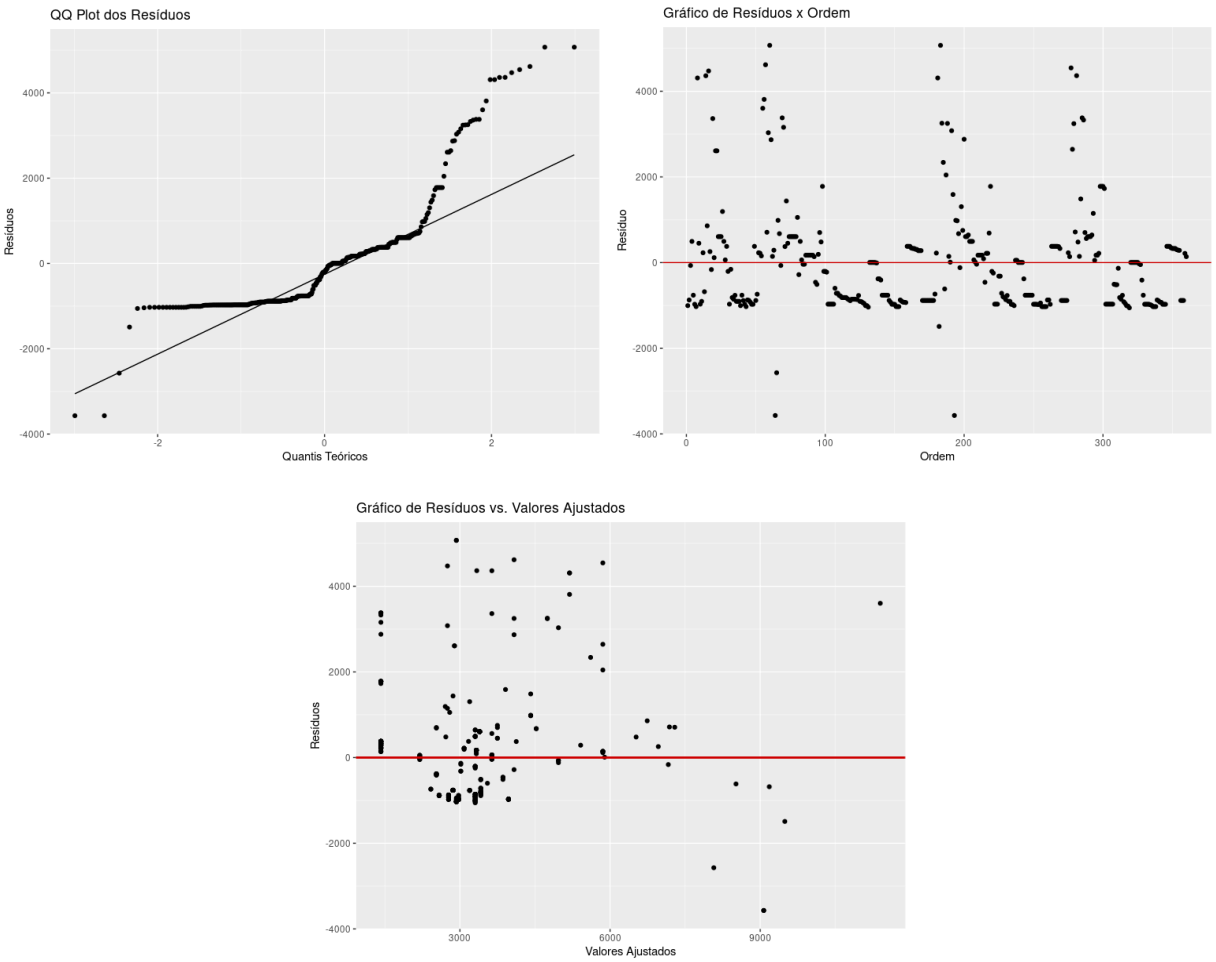

Figura 5: QQ Plot dos resíduos do modelo de São Paulo. **Figura 6:** Gráfico dos resíduos x Ordem do modelo de São Paulo. **Figura 7:** Gráfico dos resíduos x Valores ajustados do modelo de São Paulo.

A análise dos gráficos de resíduos revelou importantes observações sobre os pressupostos do modelo. O gráfico de Resíduos vs. Valores Ajustados (Figura 5) indica a violação da suposição de homocedasticidade, onde a variância dos resíduos não é homogênea. Adicionalmente, o QQ Plot dos Resíduos (Figura 6) indica um desvio da normalidade, principalmente nas caudas da distribuição, sugerindo a presença de alguns outliers ou que a distribuição dos erros não se ajusta perfeitamente a uma normal. Além disso, o gráfico de Resíduos x Ordem (Figura 7) apresentou um leve padrão, sugerindo uma possível violação do pressuposto de independência dos resíduos.

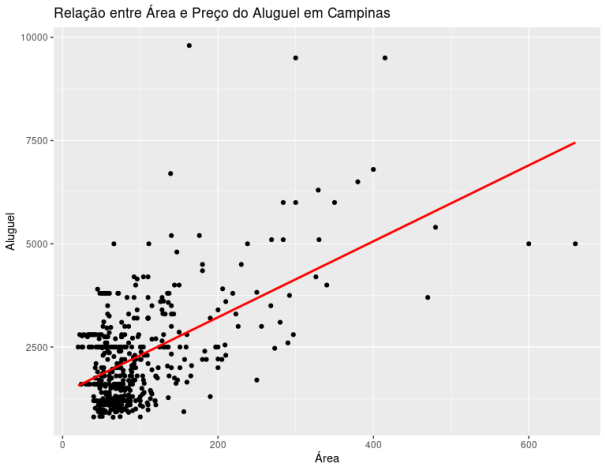

Figura 9: Gráfico de dispersão entre a área e o aluguel na cidade de Campinas

Tabela 2: Resultados da Regressão Linear para Aluguel em Campinas

Variável Preditora	Coeficiente (β)	Erro Padrão	Estatística t	Valor-p
(Intercepto)	1381.31	76.12	18.15	< 0.001 (***)
Área	9.20	0.62	14.77	< 0.001 (***)

R^2 (Ajustado): 0.3074

Na Tabela 2, o modelo de Campinas apresenta $R^2 = 0,3074$, indicando que apenas 30% da variação do aluguel é explicada pela variável Área. Já a figura 9 mostra que área e aluguel em campinas até tem um correlação, porém existem muitos outliers e dados muito distantes da reta.

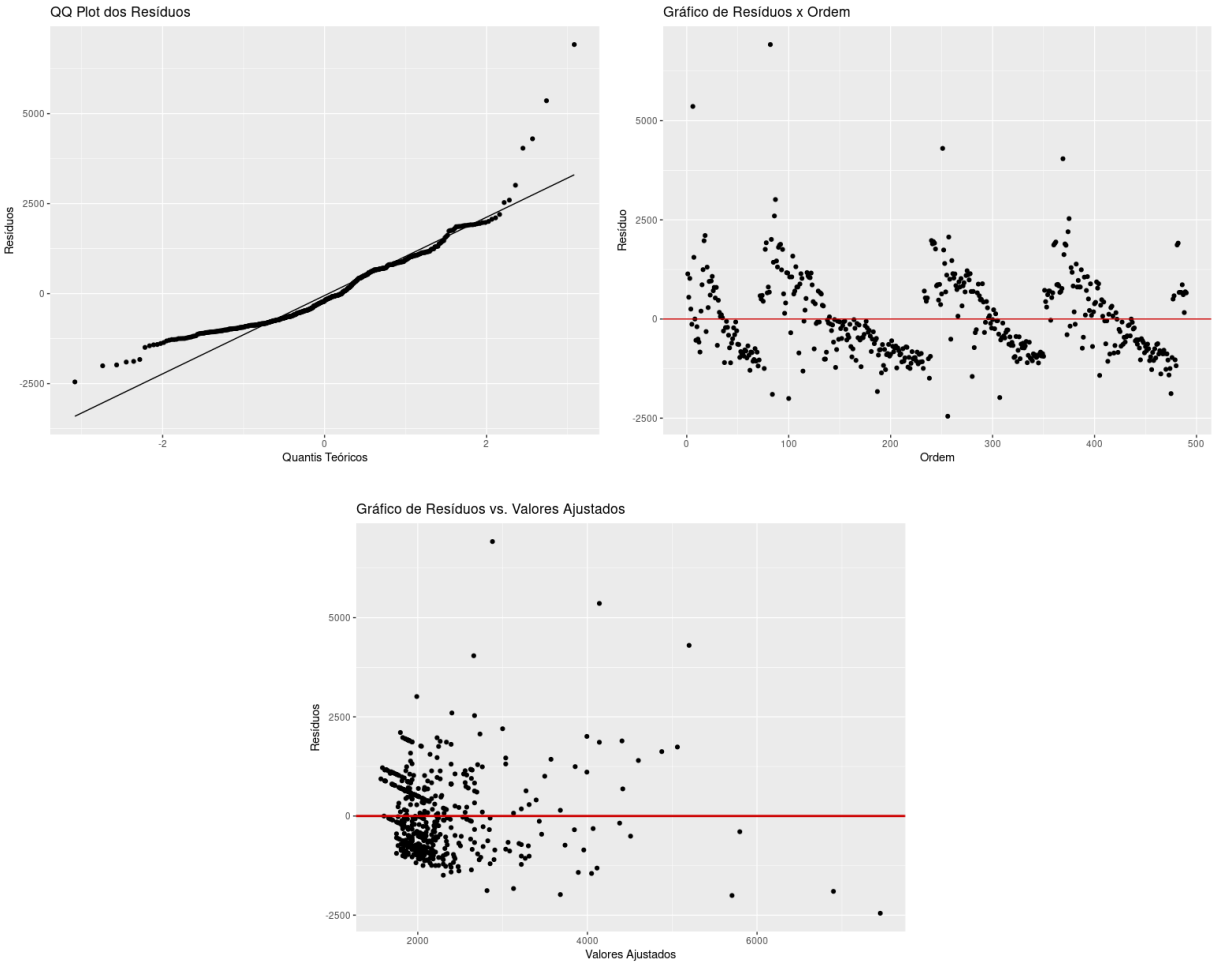

Figura 10: QQ Plot dos resíduos do modelo de São Paulo. **Figura 11:** Gráfico dos resíduos x Ordem do modelo de São Paulo. **Figura 12:** Gráfico dos resíduos x Valores ajustados do modelo de São Paulo.

É possível observar que o gráfico de Resíduos vs. Valores Ajustados (Figura 12) indica claramente a violação do pressuposto de homocedasticidade, dado que a variância dos resíduos não é homogênea. Somado a isso, o QQ Plot dos Resíduos (Figura 10) mostra a violação da suposição de normalidade, principalmente em suas caudas, ainda mais na superior. Para complementar, o gráfico de Resíduos x Ordem (Figura 11) revelou um padrão notável, sugerindo que os resíduos podem não ser independentes dada a presença de aglomerações e tendências nos resíduos ao longo da ordem.

Com base na análise realizada, observa-se que os modelos de regressão ajustados violam as suposições básicas do Modelo de Regressão Linear Simples nas duas cidades analisadas. No caso da cidade de São Paulo, a relação entre as variáveis Condomínio e Aluguel não atende aos pressupostos de linearidade, independência, homogeneidade e normalidade dos resíduos. De forma semelhante, na cidade de Campinas, a regressão entre Área e Aluguel também apresenta violações em todas essas suposições, comprometendo a confiabilidade do modelo.

Esses resultados reforçam a baixa capacidade preditiva dos modelos e indicam que outras variáveis podem influenciar significativamente o valor do aluguel nas duas cidades através de uma outra análise.

ANEXO (código R):

```
library(tidyverse)
library(ggplot2)
library(corrplot)
library(lmtest)

###### Geral #####

cor(Aluguel, area)
plot(Aluguel, area)

cor(Aluguel, condominio)
plot(Aluguel,condominio)

plot(Aluguel, IPTU)
cor(Aluguel, IPTU)

plot(Aluguel, n_garagem)
cor(Aluguel, n_garagem)

plot(Aluguel, num_quartos)
cor(Aluguel, num_quartos)

#### Por Variavel em cada cidade ####

### Condominio ####

plot(Aluguel[cidade == "SP"],condominio[cidade == "SP"])
cor(Aluguel[cidade == "SP"],condominio[cidade == "SP"])

plot(Aluguel[cidade == "Campinas"],condominio[cidade == "Campinas"])
cor(Aluguel[cidade == "Campinas"],condominio[cidade == "Campinas"])

### Area ###
plot(Aluguel[cidade == "SP"],area[cidade == "SP"])
```

```
cor(Aluguel[cidade == "SP"],area[cidade == "SP"])

plot(Aluguel[cidade == "Campinas"],area[cidade == "Campinas"])
cor(Aluguel[cidade == "Campinas"],area[cidade == "Campinas"])

### IPTU ###
plot(Aluguel[cidade == "SP"],IPTU[cidade == "SP"])
cor(Aluguel[cidade == "SP"],IPTU[cidade == "SP"])

plot(Aluguel[cidade == "Campinas"],IPTU[cidade == "Campinas"])
cor(Aluguel[cidade == "Campinas"],IPTU[cidade == "Campinas"])

### n_garagem ###
plot(Aluguel[cidade == "SP"],n_garagem[cidade == "SP"])
cor(Aluguel[cidade == "SP"],n_garagem[cidade == "SP"])

plot(Aluguel[cidade == "Campinas"],n_garagem[cidade == "Campinas"])
cor(Aluguel[cidade == "Campinas"],n_garagem[cidade == "Campinas"])

### num_quartos ###
plot(Aluguel[cidade == "SP"],num_quartos[cidade == "SP"])
cor(Aluguel[cidade == "SP"],num_quartos[cidade == "SP"])

plot(Aluguel[cidade == "Campinas"],num_quartos[cidade == "Campinas"])
cor(Aluguel[cidade == "Campinas"],num_quartos[cidade == "Campinas"])



### Escolhendo a variavel condominio para sao paulo e area para campinas ###

### Para SP ###

SP_dep_aluguel <- Aluguel[cidade == "SP"]
sp_ind_condominio <- condominio[cidade == "SP"]



modelo_sp <- lm(SP_dep_aluguel ~ sp_ind_condominio)
summary(modelo_sp)
plot(modelo_sp)

### Teste de Normalidade ###
shapiro.test(modelo_sp$residuals)
shapiro.test(rstandard(modelo_sp))
```

```
###Teste de Homogeneidade ###

bptest(modelo_sp)

### TEste de Linearidade ###
resettest(modelo_sp)

### Teste de Independencia ###
dwtest(modelo_sp)




### Para Campinas ###

Camp_dep_aluguel <-Aluguel[cidade == "Campinas"]
Camp_ind_area <- area[cidade == "Campinas"]

modelo_camp <- lm(Camp_dep_aluguel ~ Camp_ind_area)
summary(modelo_camp)
fitted(modelo_camp)
plot(Camp_ind_area, Camp_dep_aluguel)
plot(modelo_camp)

### Teste de Normalidade ###
hist(modelo_camp$fitted.values)
Camp_dep_aluguel[[1]]
modelo_camp$fitted.values[[1]]
modelo_camp$residuals[[1]]
plot(modelo_camp$fitted.values, modelo_camp$residuals, xlab = "Valores ajustados", ylab = "Residuos")
abline(h = 0, col = "red")
hist(rstandard(modelo_camp))
plot(modelo_camp$fitted.values, rstandard(modelo_camp), xlab = "Valores ajustados", ylab = "Residuos")
abline(h = 0, col = "red")

### Principais testes de Normalidade ###
qqnorm(modelo_camp$residuals,
      main = "Q-Q Plot dos Resíduos (Campinas)",
      xlab = "Quantis Teóricos da Normal",
      ylab = "Quantis Amostrais")
```

```
qqline(modelo_camp$residuals, col = "red", lwd = 2) # Adiciona a linha de referência

hist(modelo_camp$residuals)
shapiro.test(modelo_camp$residuals)



### Teste de homogeneidade ###
plot(modelo_camp, which = 1)
plot(fitted(modelo_camp), residuals(modelo_camp))

bptest(modelo_camp)
summary(modelo_camp)



### Teste de independencia ###

dwtest(modelo_camp)

### Teste de linearidade ###
resettest(modelo_camp)

########################

sp = filter(DadosAluguel_SP1, cidade == "SP")
campinas = filter(DadosAluguel_SP1, cidade == "Campinas")

#### Matriz de Correlação ###
colunas_interesse_geral = DadosAluguel_SP1 %>%
  select(Aluguel, area, condominio, IPTU)

matriz_cor_geral <- cor(colunas_interesse_geral)
corrplot(matriz_cor_geral, method = "color", addCoef.col = "black", tl.col = "black", title = "Matriz de Correlação", tl.srt = 45, mar = c(0, 0, 2, 0))

  # SP #
colunas_interesse_sp = sp %>%
  select(Aluguel, area, condominio, IPTU)

matriz_cor_sp <- cor(colunas_interesse_sp)
corrplot(matriz_cor_sp, method = "color", addCoef.col = "black", tl.col = "black", title = "Matriz de Correlação (São Paulo)", tl.srt = 45, mar = c(0, 0, 2, 0))
```

```
  # Campinas #
colunas_interesse_campinas = campinas %>%
  select(Aluguel, area, condominio, IPTU)

matriz_cor_campinas <- cor(colunas_interesse_campinas)
corrplot(matriz_cor_campinas, method = "color", addCoef.col = "black", tl.col = "black", title = "Matriz de Correlação (Campinas)", tl.srt = 45, mar = c(0, 0, 2, 0))

### Modelo São Paulo  ###

cor(sp$Aluguel, sp$condominio)

modelo_sp <- lm(sp$Aluguel ~ sp$condominio)
summary(modelo_sp)

ggplot(data = sp,aes(x = condominio, y = Aluguel)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red3") +
  labs(title = "Relação entre Condomínio e Preço do Aluguel em SP",
       x = "Condomínio",
       y = "Aluguel")

  ## QQPLOT ##

ggplot(modelo_sp, aes(sample = modelo_sp$residuals)) +
  stat_qq() +
  stat_qq_line() +
  labs(title = "QQ Plot dos Resíduos",
       x = "Quantis Teóricos",
       y = "Resíduos")

  ## RESIDUOS X ORDEM ##

residuos_x_ordem_sp <- data.frame(Indice = 1:length(modelo_sp$residuals),
                                  Residuos = modelo_sp$residuals)

ggplot(data = residuos_x_ordem_sp, aes(x = Indice, y = Residuos)) +
  geom_point(color = "black") +
  geom_hline(yintercept = 0, color = "red3") +
  labs(
    title = "Gráfico de Resíduos x Ordem",
    x = "Ordem",
```

```
    y = "Resíduo")

  ## RESIDUOS X VALORES AJUSTADOS ##

ggplot(data = modelo_sp, aes(x = modelo_sp$fitted.values, y = modelo_sp$residuals)) +
  geom_point(color = "black") +
  geom_hline(yintercept = 0, color = "red3", linewidth = 1) +
  labs(
    title = "Gráfico de Resíduos vs. Valores Ajustados",
    x = "Valores Ajustados",
    y = "Resíduos"
  )



### Modelo Campinas  ###
cor(campinas$Aluguel, campinas$area)

modelo_campinas <- lm(campinas$Aluguel ~ campinas$area)
summary(modelo_campinas)

ggplot(data = campinas,aes(x = area, y = Aluguel)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(title = "Relação entre Área e Preço do Aluguel em Campinas",
       x = "Área",
       y = "Aluguel")

## QQPLOT ##

ggplot(modelo_campinas, aes(sample = modelo_campinas$residuals)) +
  stat_qq() +
  stat_qq_line() +
  labs(title = "QQ Plot dos Resíduos",
       x = "Quantis Teóricos",
       y = "Resíduos")

## RESIDUOS X ORDEM ##

residuos_x_ordem_campinas <- data.frame(Indice = 1:length(modelo_campinas$residuals),
                                        Residuos = modelo_campinas$residuals)

ggplot(data = residuos_x_ordem_campinas, aes(x = Indice, y = Residuos)) +
```

```
  geom_point(color = "black") +
  geom_hline(yintercept = 0, color = "red3") +
  labs(
    title = "Gráfico de Resíduos x Ordem",
    x = "Ordem",
    y = "Resíduo")

## RESIDUOS X VALORES AJUSTADOS ##

ggplot(data = modelo_campinas, aes(x = modelo_campinas$fitted.values, y = modelo_campinas$residuals))
+
  geom_point(color = "black") +
  geom_hline(yintercept = 0, color = "red3", linewidth = 1) +
  labs(
    title = "Gráfico de Resíduos vs. Valores Ajustados",
    x = "Valores Ajustados",
    y = "Resíduos"
)
```