# REGRESSÃO ESPACIAL

CAMILA TOTTI ANDRADE
CAROLINE DA SILVA
VINÍCIUS LIMA GUIMARÃES

# CONTEÚDOS

# PRESSUPOSTOS DA REGRESSÃO LINEAR E MÉTRICAS DE AVALIAÇÃO

## Linearidade

Para cada incremento em X, deve ter um aumento/diminuição proporcional em Y.

## Independência dos erros

Erros devem ser independentes entre si.

## Normalidade dos erros

Erros devem seguir uma distribuição normal.

## Homoscedasticidade

Erros deve ter variância constante ao longo dos valores da variável Y.

## Ausência de multicolinearidade

Em regressão múltipla, não deve ter correlação entre as variáveis X.

## $R^2$ e $R^2$ ajustado

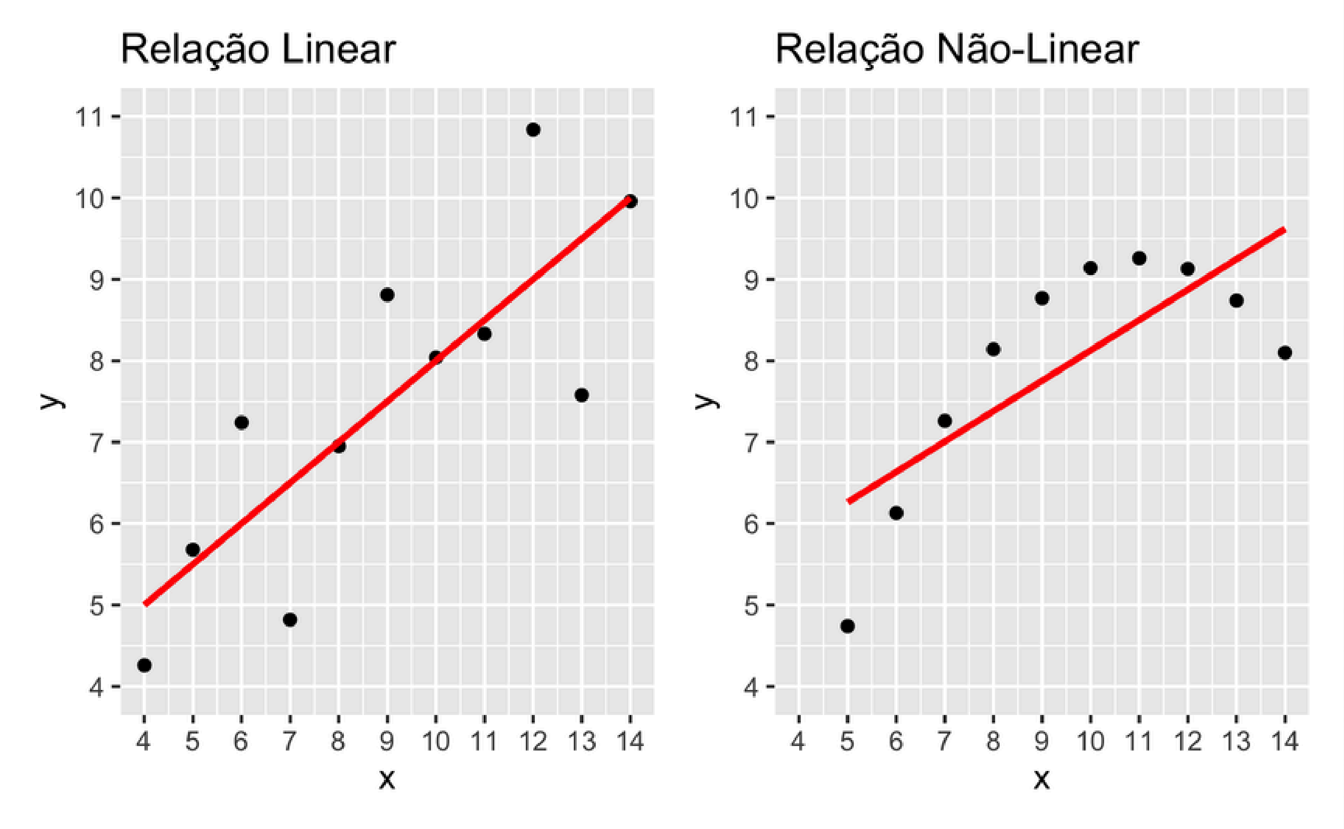
$R^2$ - Quanto da variabilidade de Y é explicada pelo modelo.
$R^2$ ajustado - Correção pelo *n* de Y.

$$R^2 = 1 - \frac{RSS}{TSS}$$

$$R^2_{ajust} = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

# LINEARIDADE

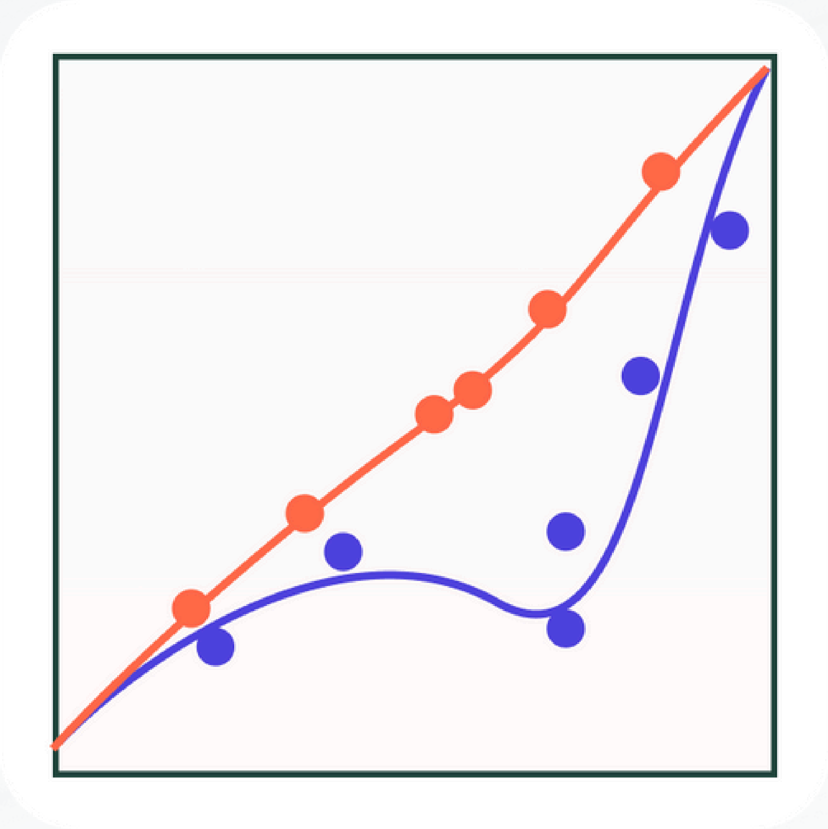
- Definição
- Importância
- Métodos de detecção - Visuais e Estatísticos
- Métodos para solução




**Teste de rainbow**

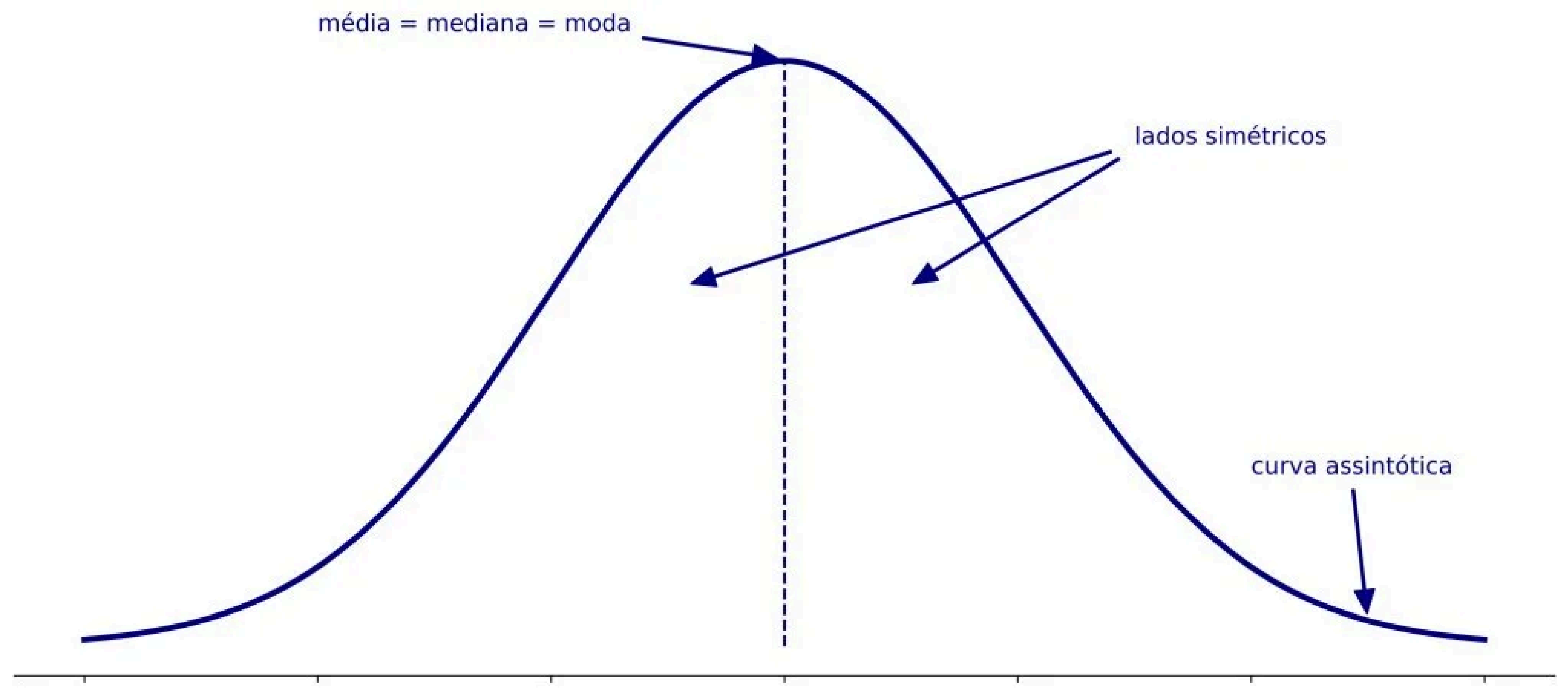
H0: Observações centrais e das extremidades não diferem significativamente em relação à linha ajustada.

H1: Há diferenças significativas entre as observações centrais e as extremidades (p <0,05).



# NORMALIDADE DOS ERROS

- Definição
- Importância - Testes estatísticos e inferências
- Shapiro-Wilk



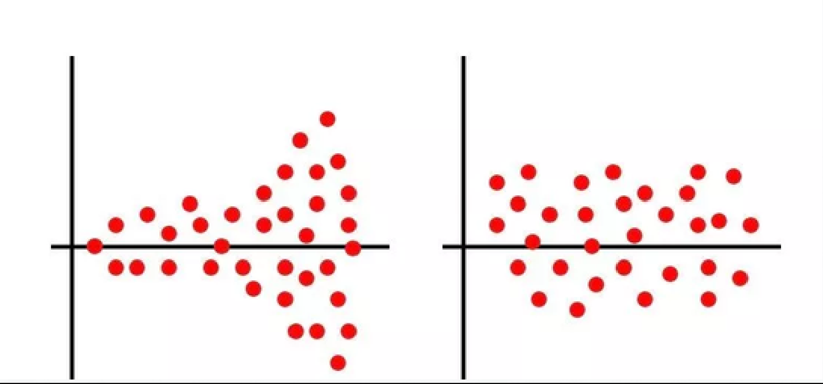

H0: Os resíduos seguem uma distribuição normal ($p > 0{,}05$)

H1: Os resíduos não seguem uma distribuição normal.

$$W = \frac{\sum_{i=1}^{n} a_i {y_i}^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

# HOMOSCEDASTICIDADE

- Definição
- Importância - Testes estatísticos e inferências
- Breusch-Pagan



H0: Os erros possuem variância constante.

H1: Os erros não possuem variância constante.

- Aplica o modelo de regressão
- Estima a variância

$$\sigma^2 = \sum \mu_i^2 / n$$

*where,* *n is the number of observations*

$\sum \mu_i^2$ *is the sum of squared residuals*

Constrói nova variável ri

$$r_i = \mu_i^2 / \sigma^2$$

*That is, each squared residual divided by* $\sigma^2$

Faz a regressão auxiliar entre ri e as variáveis independentes

$$r_i = \alpha_1 + \alpha_2 X_{1i} + \alpha_3 X_{2i} + \alpha_4 X_{3i} + v_i$$

Faz o cálculo de ω e verifica se seu χ2 é maior para o valor crítico, dado nível de significância. Se sim -> H1

$$\omega = 1/2(ESS)$$

$$\omega \sim \chi^2_{d = k-1}$$

# MULTICOLINEARIDADE

- Definição
- Importância - $R^2$ e IC
- Variance Inflation Factor (VIF)
- Técnicas para tratar - Retirada e transformação

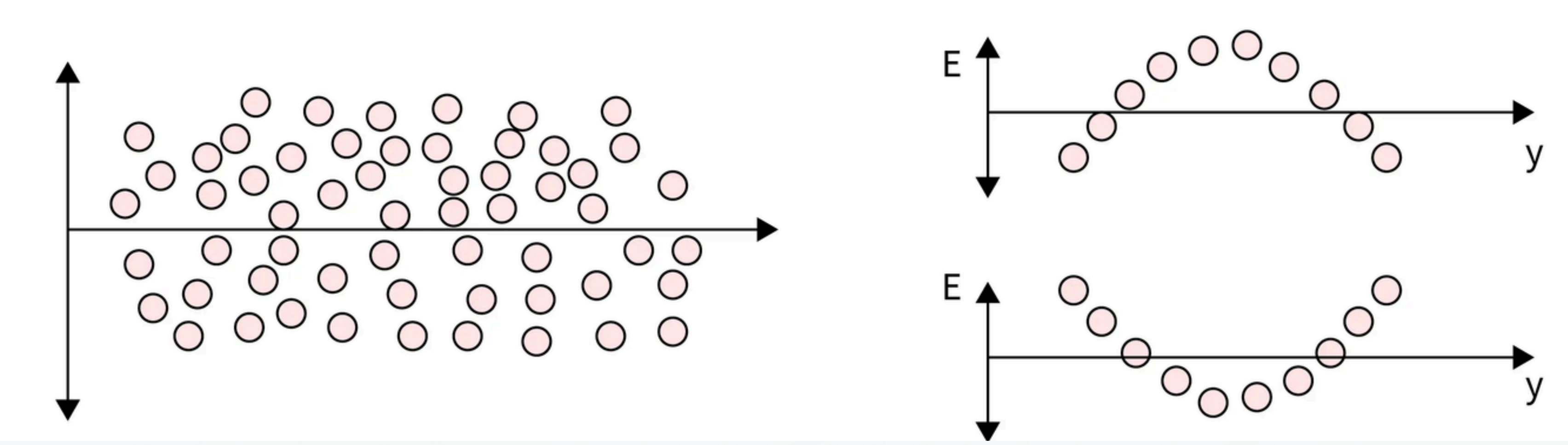
$$VIF_i = \frac{1}{1 - R_i^2}$$

VIF = 1
VIF entre 1 e 5
VIF > 5 ou 10

$R_i^2$ -> coeficiente de determinação obtido ao regredir a variável i em todas as outras variáveis independentes.

# INDEPENDÊNCIA DOS ERROS

- Definição
- Importância - Estrutura não captada
- Durbin-Watson




$$d = \sum_{t=2}^{T} (e_t - e_{t-1})^2 / \sum_{t=1}^{T} {e_t}^2$$

**Durbin-Watson**

H0: Os resíduos não apresentam autocorrelação (ρ = 0)

H1: Os residuso apresentam autocorrelação (ρ ≠ 0).

SQD resíduos consecutivos

SQ dos resíduos

# MORAN

- Mede a autocorrelação espacial nos resíduos de um modelo

- Um índice próximo de zero indica ausência de correlação; valores positivos (0 a +1) sugerem autocorrelação positiva, e negativos (0 a -1) sugerem autocorrelação negativa.

# REGRESSÃO ESPACIAL GLOBAL

- Com a comprovação de dependência espacial dos resíduos a partir do teste de Moran pode-se aplicar a regressão espacial global

- Para incorporar essa dependência na regressão, pode-se utilizar o Spatial Lag (SAR), que atribui a autocorrelação espacial à variável resposta, ou o Spatial Error (CAR), que atribui a autocorrelação ao erro.

# LMERR E LMLAG

- Para escolher qual modelo utilizar, existem quatro testes estatísticos de Lagrange Multiplier (LM) para serem aplicados.

| Característica | SAR (Spatial Autoregressive Model) | CAR (Conditional Autoregressive Model) |
|---|---|---|
| Dependência Espacial | Nos valores da variável dependente | Nos erros/resíduos do modelo |
| Teste | LM-Lag | LM-Error |
| Quando usar | Quando as áreas vizinhas se influenciam diretamente (ex: preços de casas influenciando umas às outras) | Quando fatores não observados (ex: características do bairro) causam correlação nos resíduos |
| Exemplo | Preço de casas influenciado por preços de casas vizinhas | Impacto de características não observadas no valor de casas em áreas vizinhas |

# RLMERR E RLMLAG

- Utilizando modelos robustos caso ambos testes anteriores tenham sido significativos:
- RLMerr: Este teste verifica dependência espacial nos erros (ou seja, correlação espacial nos resíduos), mesmo que possa haver uma variável dependente espacialmente defasada ausente no modelo.
- RLMlag: Este teste verifica a presença de uma variável dependente espacialmente defasada, mesmo que possa haver dependência nos erros do modelo.

# AIC E BIC

- Definição
- Importância - Estrutura não captada
- Durbin-Watson

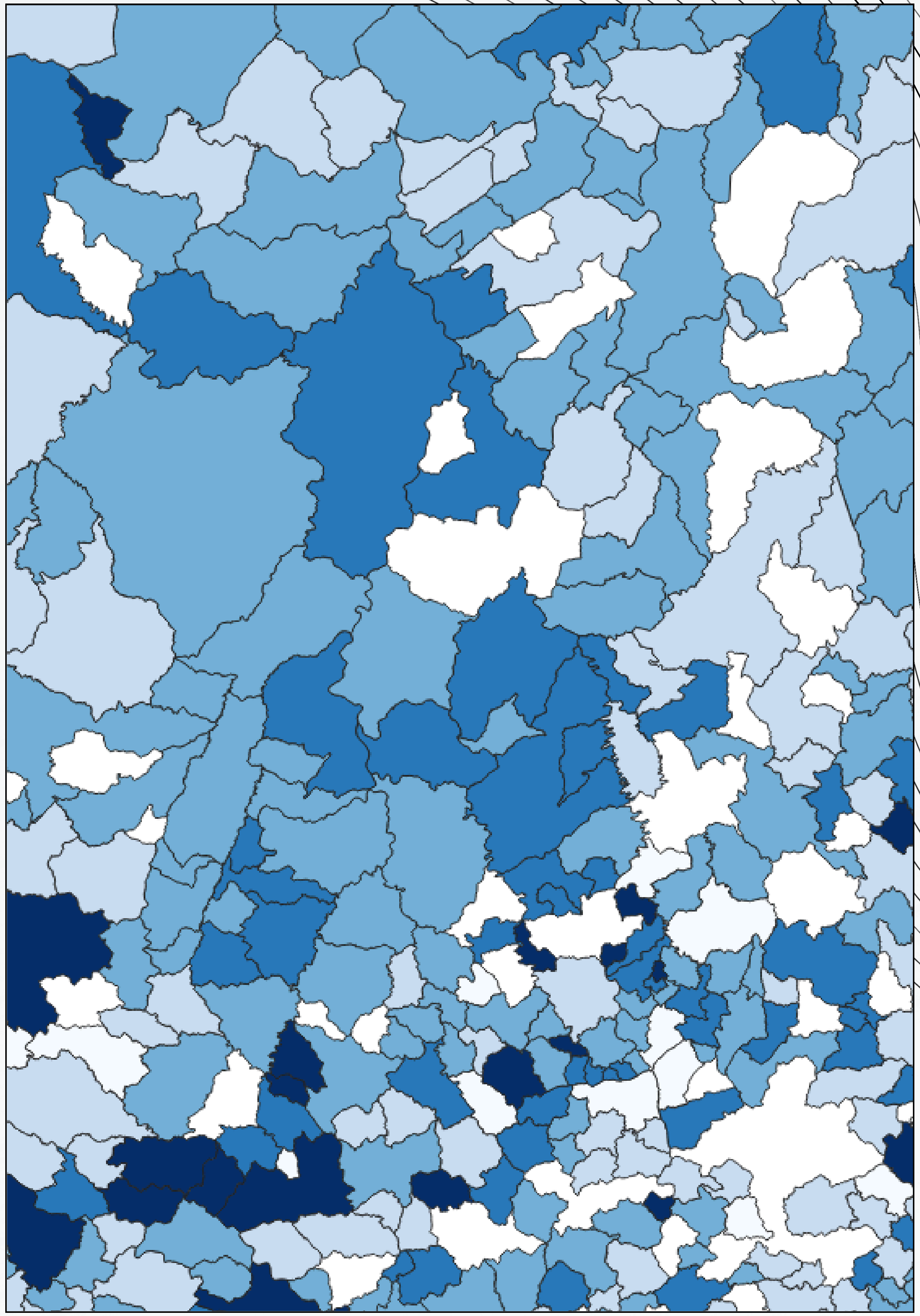
# APRESENTAÇÃO DOS RESULTADOS

- Recorte: Minas Gerais
- Modelo Linear
- Geração das matrizes de vizinhança e de pesos espaciais
- Verificação da autocorrelação espacial dos resíduos

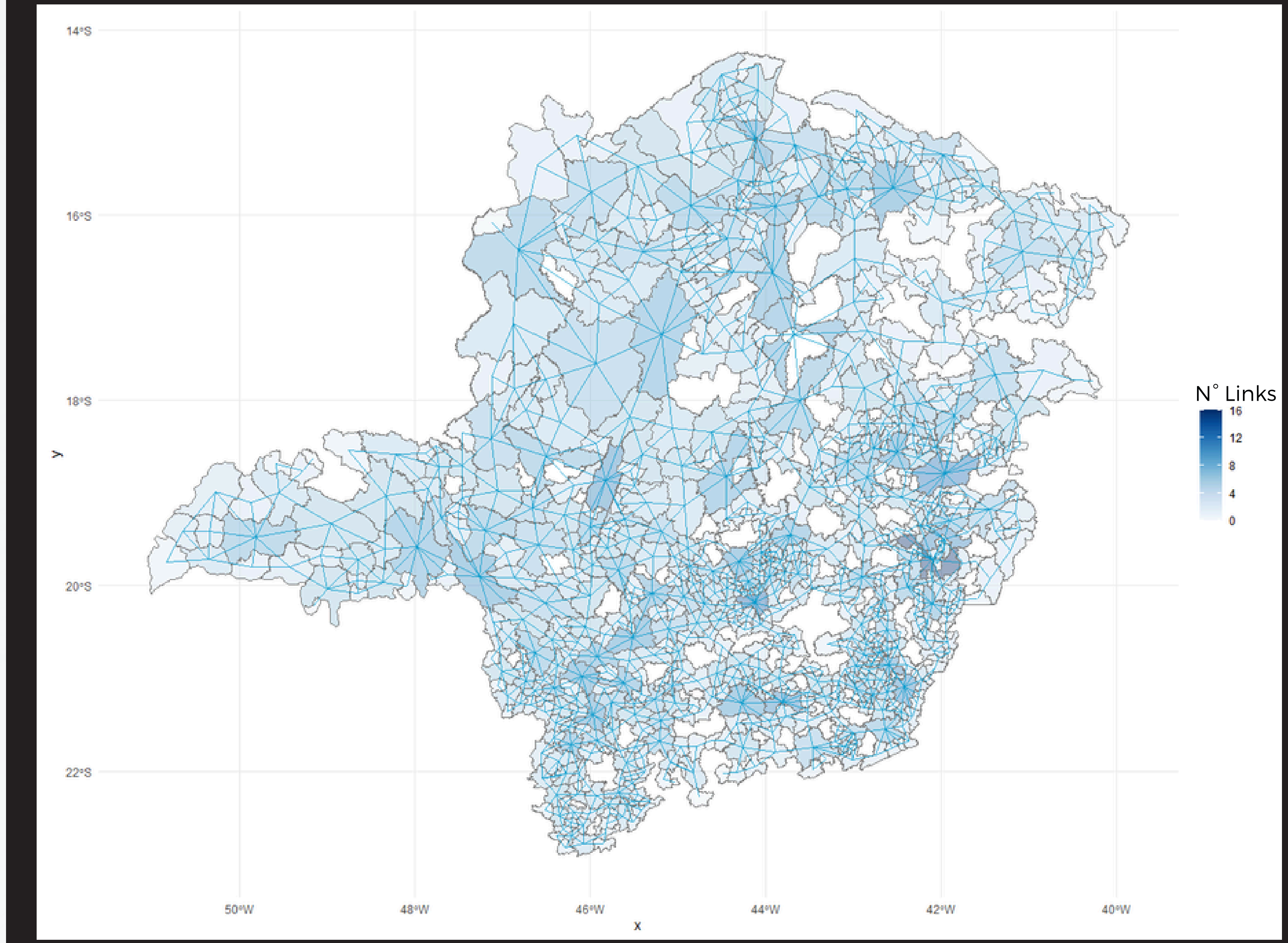
Existe dependência espacial?

- Verificação do modelo espacial mais adequado (SAR ou CAR) - Rao's score (a.k.a Lagrange Multiplier)

- Comparação dos modelos
    - AIC



# MATRIZ DE VIZINHANÇA




```
# Matriz de Vizinhança (Queen)
w <- poly2nb(pl = agua_rede_sf,
             row.names = agua_rede_sf$ID_IBGE
            )

# Matriz de pesos espaciais
wm <- nb2mat(neighbours = w,
             style='B',
             zero.policy = TRUE
            )

rwm <- mat2listw(x = wm,
                 style='W',
                 zero.policy = TRUE
                )
```

# TESTE MORAN

```
        Global Moran I for regression residuals

data:
model: lm(formula = CONSUMO1 ~ RENDAPITA, data = agua_rede_sf, na.action = na.exclude)
weights: rwm

Moran I statistic standard deviate = 4.0288, p-value = 5.607e-05
alternative hypothesis: two.sided
sample estimates:
Observed Moran I      Expectation         Variance
    0.1051184609    -0.0023395626     0.0007114376
```

Significativo!

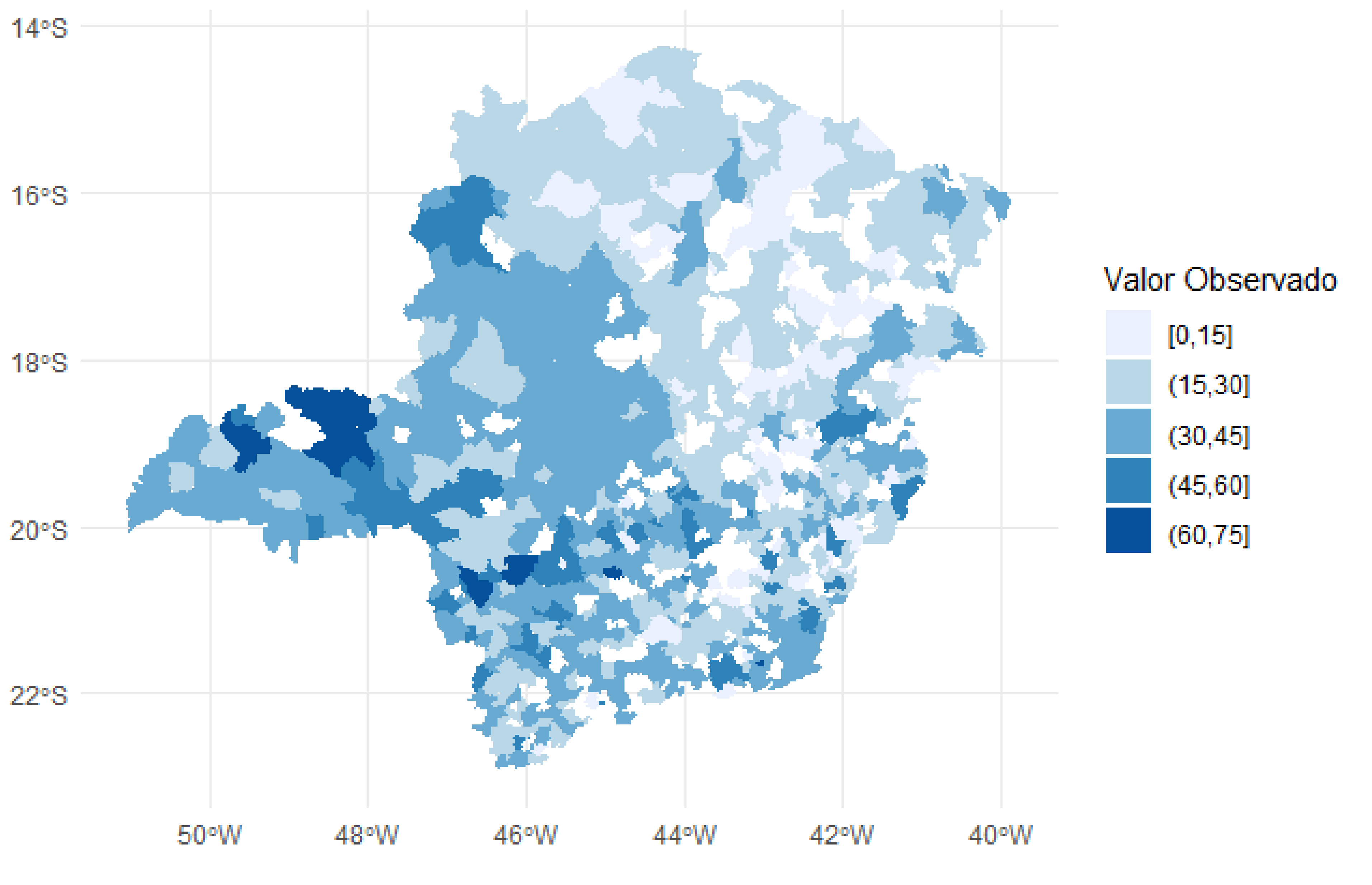
# TESTE LM

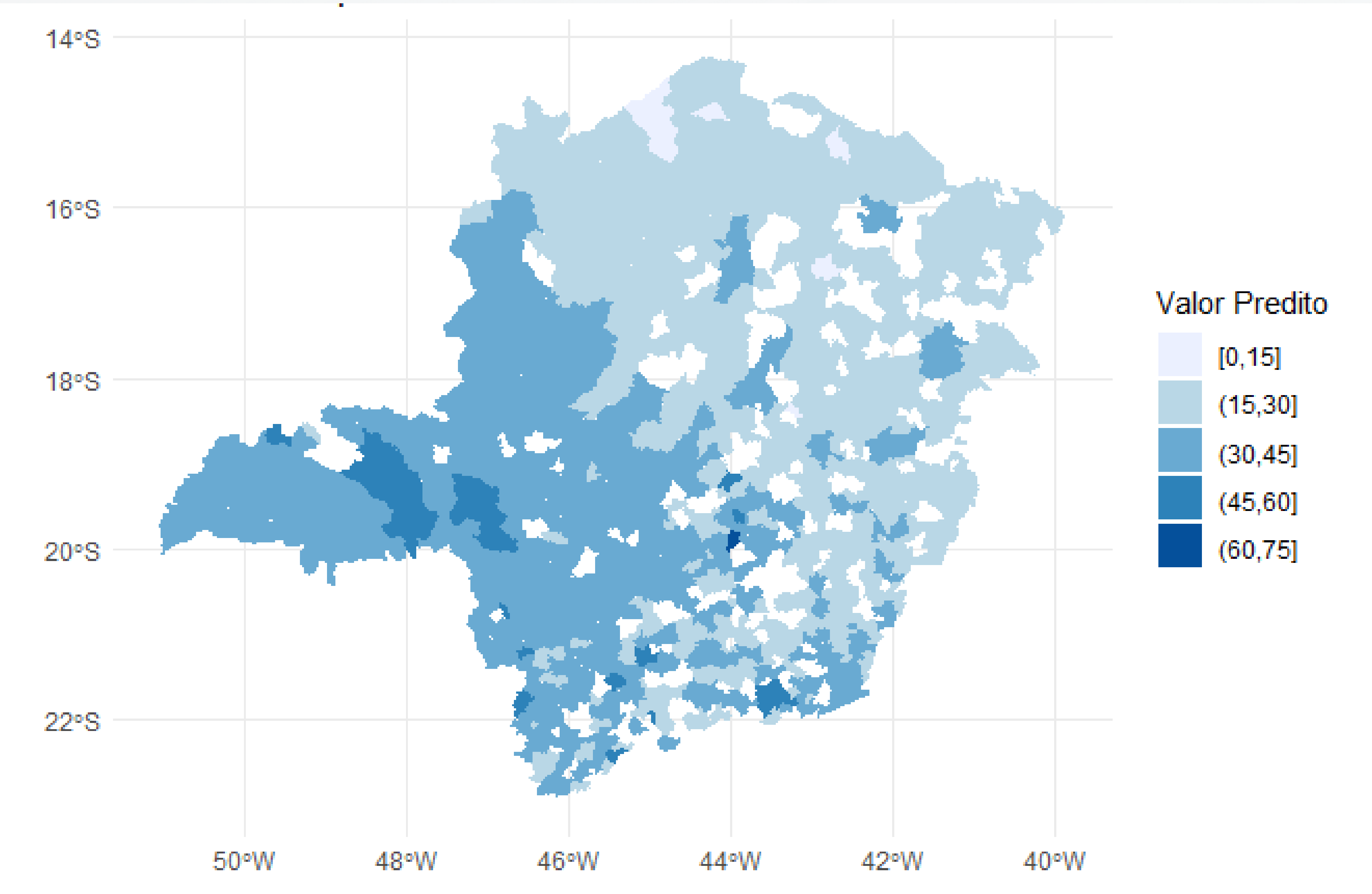
| Teste | Estatística | p.valor |
|---|---|---|
| **RSerr** | 15.4144 | 8.63E-05 |
| **RSlag** | 27.0407 | 1.99E-07 |
| **adjRSerr** | 0.0374 | 8.47E-01 |
| **adjRSlag** | 11.6638 | 6.37E-04 |

# MODELOS

| Métricas | LINEAR | | CAR | | SAR | |
|---|---|---|---|---|---|---|
| | (Intercept) | RENDAPITA | (Intercept) | RENDAPITA | (Intercept) | RENDAPITA |
| **Estimativa** | 6.3187 | 0.0452 | 7.3299 | 0.0433 | 3.1723 | 0.0395 |
| **Erro.Padrão** | 1.0175 | 0.0019 | 1.1230 | 0.0021 | 1.1872 | 0.0022 |
| **z.valor** | 6.2099 | 23.4712 | 6.5270 | 20.5334 | 2.6721 | 17.9942 |
| **p.valor** | 9.39E-10 | 5.66E-89 | 6.71E-11 | 0.00E+00 | 7.54E-03 | 0.00E+00 |
| **AIC** | **4738.2** | > | **4725.6** | > | **4715.5** | |

# MODELOS

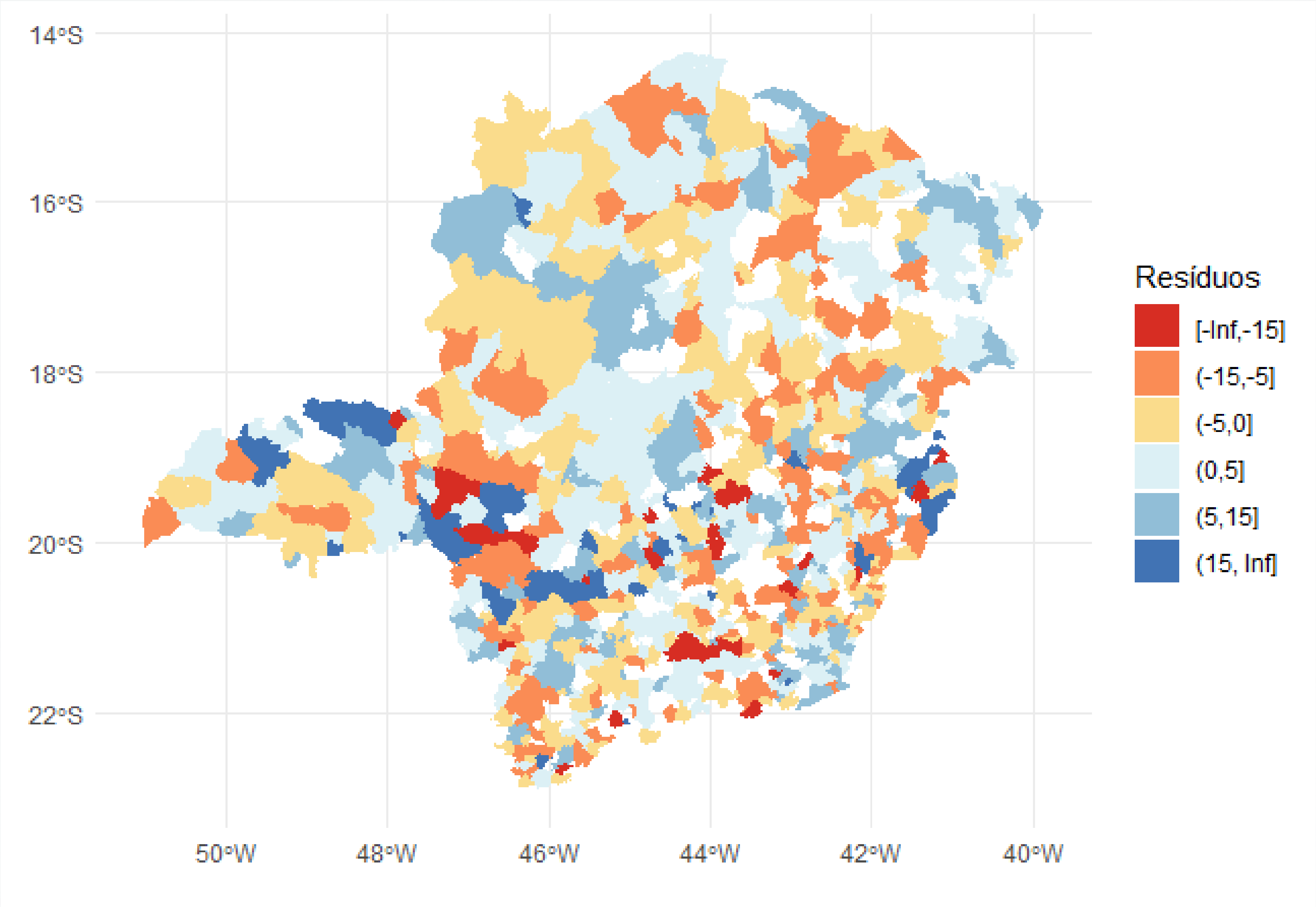



**intercepto** **coeficiente** **Termo de Dependência Espacial** **resíduo**

CONSUMO1 = 3.1723 + 0.0395 * RENDAPITA + 0.2046 * Dependência Espacial + Residual

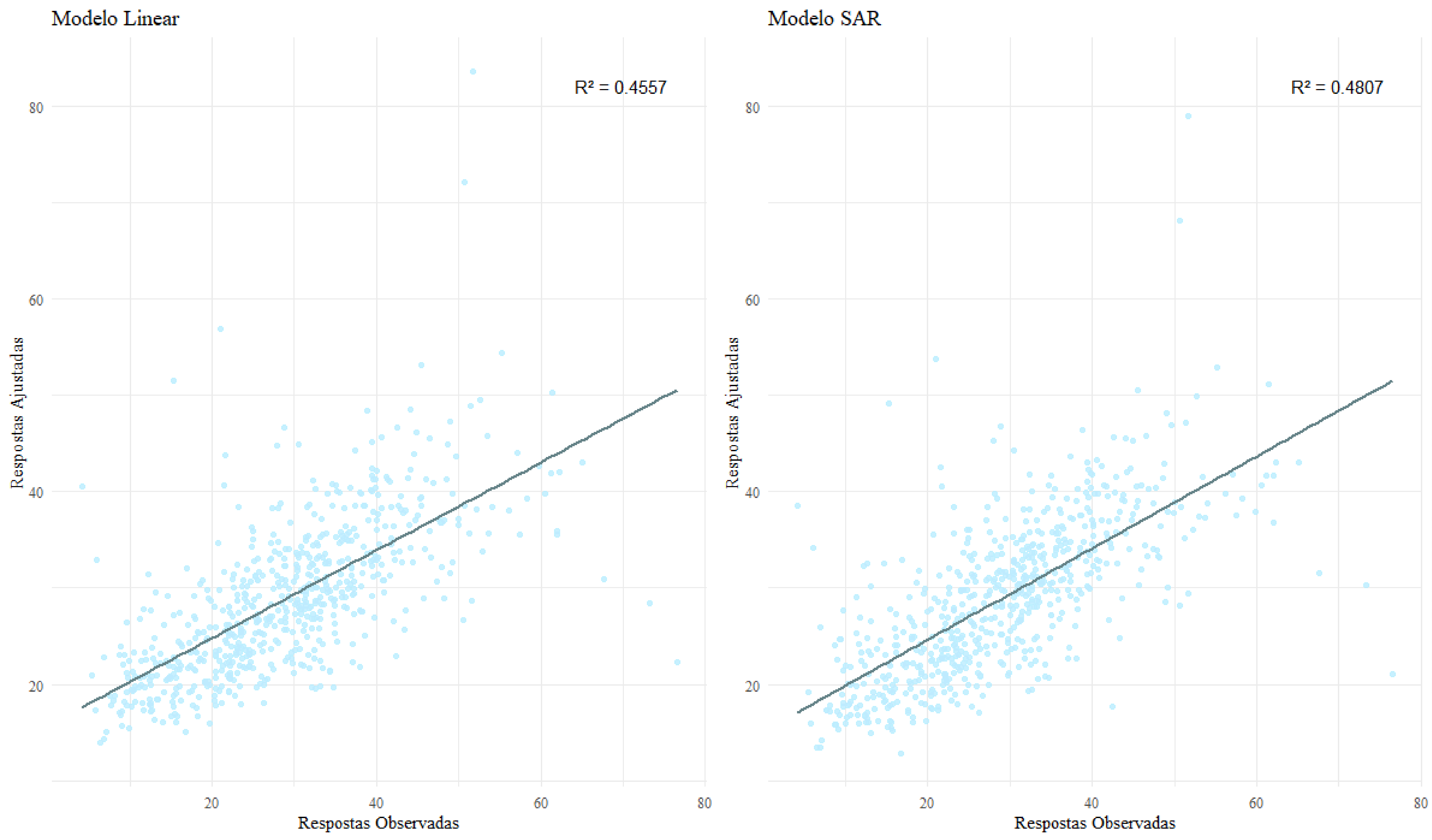
# MODELOS




**Variância dos Resíduos**

| Linear | SAR |
| --- | --- |
| 76.22 | 72.72 |

# MODELOS




Coeficiente de Determinação $R^2$

| Linear | SAR |
|---|---|
| 0.4557 | 0.4807 |

# CONCLUSÕES

*Ambos os modelos indicam que o rendimento per capita tem um impacto positivo no consumo e que as interações espaciais (dependência espacial) entre as regiões precisam ser consideradas ao modelar o consumo.*