

Modelagem Espacial da Criminalidade: Uma Abordagem com Regressão Espacial

Alison Cordeiro Sousa

Contexto

Este estudo investiga como a renda média (INC) e o valor dos imóveis (HOVAL) influenciam as taxas de criminalidade em diferentes regiões, levando em conta um fator frequentemente negligenciado: a proximidade geográfica entre as localidades. Partimos do princípio de que o crime não se distribui aleatoriamente no espaço – áreas com altos índices criminais tendem a influenciar seus entornos, seja por questões sociais compartilhadas, deslocamento de criminosos ou outros mecanismos de transbordamento espacial. Para capturar esses efeitos de vizinhança, comparamos modelos de regressão tradicionais (que tratam cada localidade como independente) com modelos espaciais avançados, que incorporam explicitamente a dependência geográfica entre as observações.

Os resultados revelam padrões intrigantes. O efeito de vizinhança mostrou-se intenso ($Rho = 0.97$) e altamente significativo, indicando que a criminalidade em uma área está fortemente correlacionada com a das regiões próximas. Quando esse fator espacial é considerado, as relações tradicionais se modificam substancialmente: a renda, que no modelo convencional apresentava forte relação negativa com o crime (-0.41), viu seu impacto reduzido em 55% (-0.18) na análise espacial. Já o valor dos imóveis, que mantém uma associação positiva com a criminalidade, mostrou um efeito mais moderado (0.06 contra 0.10 no modelo tradicional). Essas diferenças comprovam que ignorar a dimensão espacial pode levar a conclusões enviesadas – tanto o critério AIC (245.9 contra 290.4) quanto os testes estatísticos confirmam a superioridade dos modelos que incorporam a geografia, revelando dinâmicas que permaneceriam ocultas nas abordagens convencionais.

```
# Limpar ambiente
rm(list = ls())

# Carregar pacotes necessários
library(spdep)      # análise espacial
library(spatialreg) # modelos espaciais
library(ggplot2)    # gráficos
library(sf)         # manipular coordenadas espaciais

# Definir diretório de trabalho
setwd("C:/Users/PC GAMER/Downloads/data")

# Ler base de dados com variáveis CRIME (dependente),
# INC e HOVAL (explicativas), X e Y (coordenadas)
df <- read.csv("data.csv")
```

```

# Extrair variável dependente
Y <- df$CRIME

# Extrair variáveis independentes
X <- df[, c("INC", "HOVAL")]

# Juntar coordenadas espaciais em matriz
coords <- cbind(df$X, df$Y)

# Converter base para formato espacial sf
df_sf <- st_as_sf(df, coords = c("X", "Y"), crs = 4326)

# Criar estrutura de vizinhança por distância (entre 0 e 10 unidades)
neighbors <- dnearneigh(coords, d1=0, d2=10)

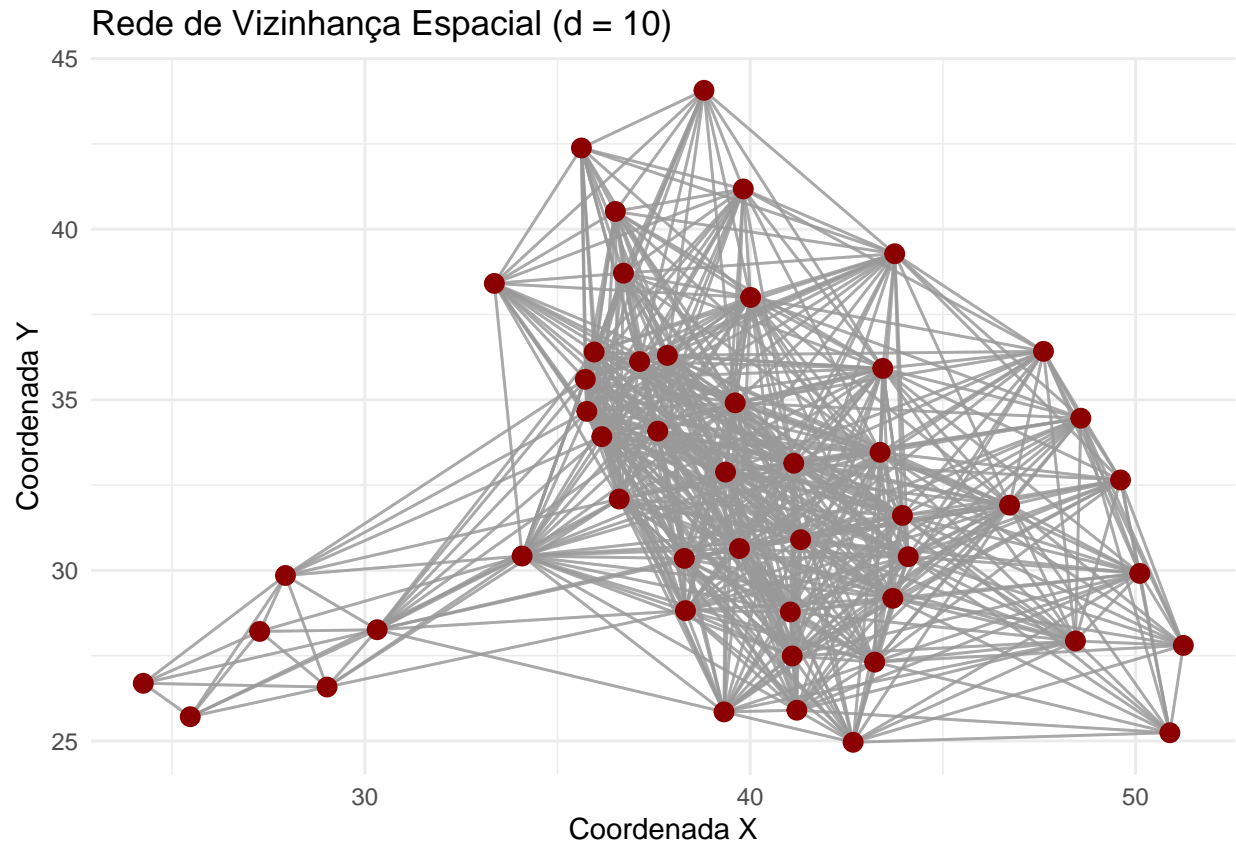
# Gerar matriz de pesos espaciais W
listw <- nb2listw(neighbors, style="W")

# Função personalizada para desenhar gráfico com pontos e linhas entre vizinhos
plot_neighbors <- function(coords, nb, title = "Mapa de Vizinhança Espacial") {
  segments <- list()
  for (i in seq_along(nb)) {
    for (j in nb[[i]]) {
      segments[[length(segments) + 1]] <- data.frame(
        x = coords[i, 1], y = coords[i, 2],
        xend = coords[j, 1], yend = coords[j, 2]
      )
    }
  }
  seg_df <- do.call(rbind, segments)
  coords_df <- as.data.frame(coords)

  ggplot() +
    geom_segment(data = seg_df, aes(x = x, y = y, xend = xend, yend = yend),
      color = "gray60", alpha = 0.6) +
    geom_point(data = coords_df, aes(x = V1, y = V2), color = "darkred", size = 3) +
    theme_minimal() +
    labs(title = title, x = "Coordenada X", y = "Coordenada Y")
}

# Mostrar gráfico de vizinhança
plot_neighbors(coords, neighbors, title = "Rede de Vizinhança Espacial (d = 10)")

```



```
# Ver estatísticas descritivas básicas
summary(Y)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.1783 20.0485 34.0008 35.1288 48.5855 68.8920
```

```
summary(X)
```

```
##      INC      HOVAL
## Min.   : 4.477  Min.   :17.90
## 1st Qu.: 9.963  1st Qu.:25.70
## Median :13.380  Median :33.50
## Mean   :14.375  Mean   :38.44
## 3rd Qu.:18.324  3rd Qu.:43.30
## Max.   :31.070  Max.   :96.40
```

```
# Regressão linear tradicional (sem considerar espaço)
olsreg <- lm(Y ~ INC + HOVAL, data = df)
summary(olsreg)
```

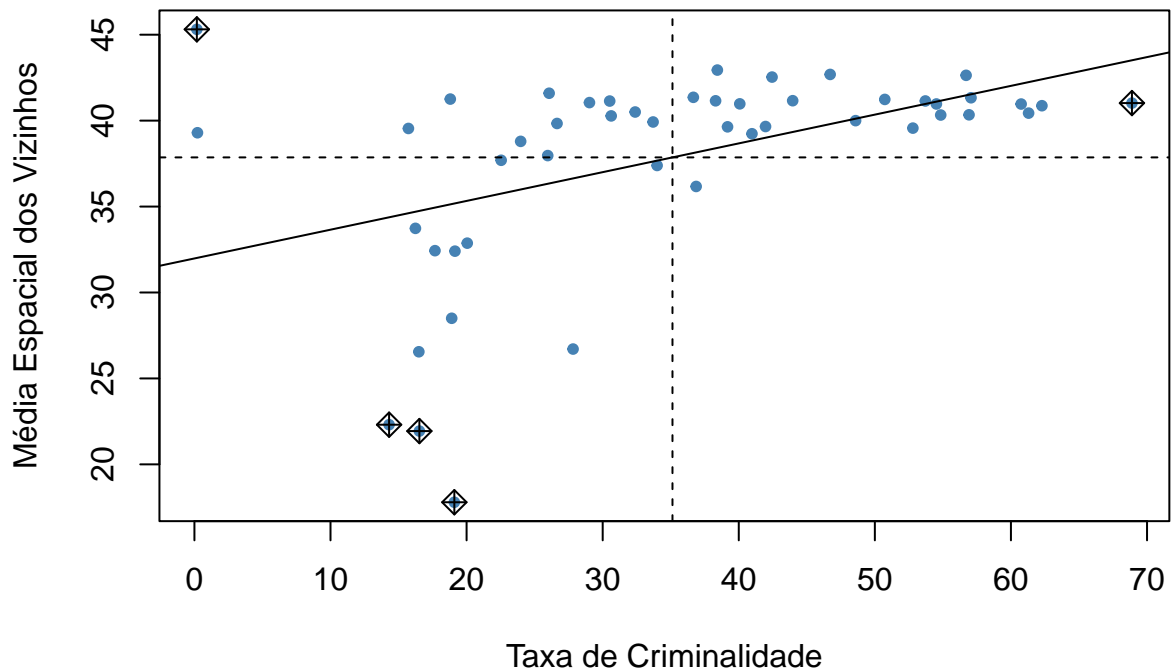
```
##
```

```
## Call:
## lm(formula = Y ~ INC + HOVAL, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.811 -2.690 -0.758  2.150 12.232
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.46582    1.84543  18.676 < 2e-16 ***
## INC         -0.41425    0.13021  -3.181  0.00263 **
## HOVAL        0.10047    0.04022   2.498  0.01612 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.456 on 46 degrees of freedom
## Multiple R-squared:  0.1961, Adjusted R-squared:  0.1611
## F-statistic:  5.61 on 2 and 46 DF,  p-value: 0.006604
```

```
# Teste de Moran I: verifica autocorrelação espacial da variável dependente
moran.test(Y, listw)
```

```
##
## Moran I test under randomisation
##
## data: Y
## weights: listw
##
## Moran I statistic standard deviate = 5.6185, p-value = 9.629e-09
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##      0.167361950      -0.020833333      0.001121945
```

```
# Gráfico de Moran: mostra se valores altos/baixos estão próximos entre si
moran.plot(Y, listw, labels = FALSE, col = "steelblue", pch = 20,
           xlab = "Taxa de Criminalidade", ylab = "Média Espacial dos Vizinhos")
```



```
# Testes Lagrange Multiplier: verificar se modelos espaciais são necessários
lm.RStests(olsreg, listw, test = c("LMlag", "LMerr"))
```

```
##
## Rao's score (a.k.a Lagrange multiplier) diagnostics for spatial
## dependence
##
## data:
## model: lm(formula = Y ~ INC + HOVAL, data = df)
## test weights: listw
##
## RSlag = 89.037, df = 1, p-value < 2.2e-16
##
##
## Rao's score (a.k.a Lagrange multiplier) diagnostics for spatial
## dependence
##
## data:
## model: lm(formula = Y ~ INC + HOVAL, data = df)
## test weights: listw
##
## RSerr = 59.226, df = 1, p-value = 1.41e-14
```

```
# Modelo SAR (Spatial Autoregressive) com efeito de vizinhança no termo defasado (lag)
spatial.lag <- spatialreg::lagsarlm(Y ~ INC + HOVAL, data = df, listw = listw)
summary(spatial.lag)
```

```
##
## Call:spatialreg::lagsarlm(formula = Y ~ INC + HOVAL, data = df, listw = listw)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.38422 -1.59758 -0.46952  1.58598  5.97001
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.230021   1.269203  0.9691 0.332481
## INC          -0.185257   0.073333 -2.5263 0.011529
## HOVAL         0.063364   0.022648  2.7978 0.005145
##
## Rho: 0.9683, LR test value: 46.503, p-value: 9.1457e-12
## Asymptotic standard error: 0.021879
##      z-value: 44.256, p-value: < 2.22e-16
## Wald statistic: 1958.6, p-value: < 2.22e-16
##
## Log likelihood: -117.9492 for lag model
## ML residual variance (sigma squared): 6.2975, (sigma: 2.5095)
## Number of observations: 49
## Number of parameters estimated: 5
## AIC: 245.9, (AIC for lm: 290.4)
## LM test for residual autocorrelation
## test value: 78.987, p-value: < 2.22e-16
```

```
# Modelo SEM (Spatial Error Model) com dependência espacial nos resíduos
spatial.error <- spatialreg::errorsarlm(Y ~ INC + HOVAL, data = df, listw = listw)
summary(spatial.error)
```

```
##
## Call:spatialreg::errorsarlm(formula = Y ~ INC + HOVAL, data = df,
##      listw = listw)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.470235 -1.883781  0.089731  1.608741  5.702873
##
## Type: error
## Coefficients: (asymptotic standard errors)
##              Estimate Std. Error z value Pr(>|z|)
```

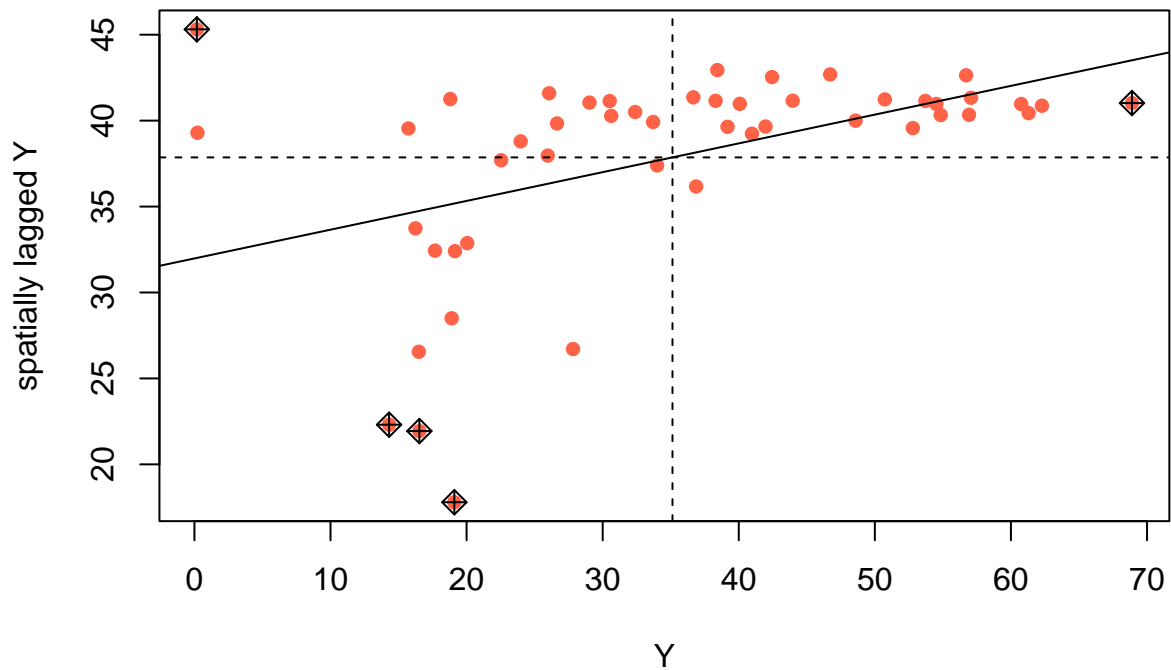
```
## (Intercept) 30.131471 12.486540 2.4131 0.01582
## INC        -0.074357 0.084463 -0.8803 0.37867
## HOVAL       0.040785 0.022495 1.8131 0.06982
##
## Lambda: 0.96882, LR test value: 40.988, p-value: 1.532e-10
## Asymptotic standard error: 0.021535
##      z-value: 44.988, p-value: < 2.22e-16
## Wald statistic: 2023.9, p-value: < 2.22e-16
##
## Log likelihood: -120.7072 for error model
## ML residual variance (sigma squared): 7.0429, (sigma: 2.6539)
## Number of observations: 49
## Number of parameters estimated: 5
## AIC: 251.41, (AIC for lm: 290.4)
```

```
# Criar matriz de vizinhança novamente (com mesmo raio)
nb_dist <- dnearneigh(coords, d1=0, d2=10)
listw_dist <- nb2listw(nb_dist, style="W")

# Repetir o Teste de Moran com a nova matriz
moran.test(Y, listw_dist)
```

```
##
## Moran I test under randomisation
##
## data: Y
## weights: listw_dist
##
## Moran I statistic standard deviate = 5.6185, p-value = 9.629e-09
## alternative hypothesis: greater
## sample estimates:
## Moran I statistic      Expectation      Variance
##      0.167361950      -0.020833333      0.001121945
```

```
# Gráfico com matriz alternativa
moran.plot(Y, listw_dist, labels = FALSE, col = "tomato", pch = 16)
```



```
# Repetir os testes LM para nova matriz de pesos
lm.LMtests(olsreg, listw_dist, test = c("LMlag", "LMerr"))
```

```
##
## Rao's score (a.k.a Lagrange multiplier) diagnostics for spatial
## dependence
##
## data:
## model: lm(formula = Y ~ INC + HOVAL, data = df)
## test weights: listw
##
## RSlag = 89.037, df = 1, p-value < 2.2e-16
##
##
## Rao's score (a.k.a Lagrange multiplier) diagnostics for spatial
## dependence
##
## data:
## model: lm(formula = Y ~ INC + HOVAL, data = df)
## test weights: listw
##
## RSerr = 59.226, df = 1, p-value = 1.41e-14
```



```

# Modelo SAR com matriz de distância
spatial.lag_dist <- spatialreg::lagsarlm(Y ~ INC + HOVAL, data = df, listw = listw_dist)
summary(spatial.lag_dist)

##
## Call:
## spatialreg::lagsarlm(formula = Y ~ INC + HOVAL, data = df, listw = listw_dist)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.38422 -1.59758 -0.46952  1.58598  5.97001
##
## Type: lag
## Coefficients: (asymptotic standard errors)
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.230021   1.269203  0.9691 0.332481
## INC          -0.185257   0.073333 -2.5263 0.011529
## HOVAL         0.063364   0.022648  2.7978 0.005145
##
## Rho: 0.9683, LR test value: 46.503, p-value: 9.1457e-12
## Asymptotic standard error: 0.021879
##      z-value: 44.256, p-value: < 2.22e-16
## Wald statistic: 1958.6, p-value: < 2.22e-16
##
## Log likelihood: -117.9492 for lag model
## ML residual variance (sigma squared): 6.2975, (sigma: 2.5095)
## Number of observations: 49
## Number of parameters estimated: 5
## AIC: 245.9, (AIC for lm: 290.4)
## LM test for residual autocorrelation
## test value: 78.987, p-value: < 2.22e-16

# Modelo SEM com matriz de distância
spatial.error_dist <- spatialreg::errorsarlm(Y ~ INC + HOVAL, data = df, listw = listw_dist)
summary(spatial.error_dist)

##
## Call:spatialreg::errorsarlm(formula = Y ~ INC + HOVAL, data = df,
##      listw = listw_dist)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.470235 -1.883781  0.089731  1.608741  5.702873
##
## Type: error
## Coefficients: (asymptotic standard errors)

```

```

##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 30.131471  12.486540  2.4131  0.01582
## INC        -0.074357   0.084463 -0.8803  0.37867
## HOVAL       0.040785   0.022495  1.8131  0.06982
##
## Lambda: 0.96882, LR test value: 40.988, p-value: 1.532e-10
## Asymptotic standard error: 0.021535
##      z-value: 44.988, p-value: < 2.22e-16
## Wald statistic: 2023.9, p-value: < 2.22e-16
##
## Log likelihood: -120.7072 for error model
## ML residual variance (sigma squared): 7.0429, (sigma: 2.6539)
## Number of observations: 49
## Number of parameters estimated: 5
## AIC: 251.41, (AIC for lm: 290.4)

```