# REGRESSÃO LINEAR ESPACIAL

## O que é regressão linear?

A regressão linear é uma técnica de análise de dados que prevê o valor de dados desconhecidos usando outro valor de dados relacionado e conhecido. Ele modela matematicamente uma variável resposta ( desconhecida ou dependente) e um conjunto de variáveis explicativas (preditoras, conhecidas ou independente) como uma equação linear: **$y = X\beta + \varepsilon$**. Onde **$y = (y_1, y_2, \ldots, y_n)'$ é um vetor de medidas da variável resposta, ${}_nX_k$ é uma matriz com k variáveis explicativas, $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$ é um vetor de coeficientes explicativos do modelo e $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ é um vetor de erros do modelo. Matricialmente o modelo é representado por:**

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1k} \\ 1 & x_{22} & \ddots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

# REGRESSÃO LINEAR ESPACIAL

## Por que a regressão linear é importante?

Os modelos de regressão linear são relativamente simples e fornecem uma fórmula matemática fácil de interpretar para gerar predições. A regressão linear é uma técnica estatística estabelecida e se aplica facilmente no R e em outros softwares. As empresas a utilizam para converter dados brutos de forma confiável e previsível em inteligência de negócios e insights práticos. Cientistas em muitos campos, incluindo saúde, biologia, ciências comportamentais, ambientais e sociais, usam a regressão linear para realizar análises preliminares de dados e prever tendências futuras. Muitos métodos de ciência de dados, como machine learning e inteligência artificial, usam a regressão linear para resolver problemas complexos.

## Etapas na regressão linear

1. Plotar uma linha e medir a correlação entre (x,y).
2. Estimar os coeficientes da regressão.
3. Identificar a equação de regressão linear.
4. predizer valores para **y** para um valor ou vetor **x** de interesse

# REGRESSÃO LINEAR ESPACIAL

## Exemplo de regressão linear: PLANEJAMENTO DE PROPAGANDAS

considere que o proprietário de um estabelecimento deseja aumentar as vendas investindo em propaganda. Considere também que o gasto neste tipo de publicidade é calculado pelo número de inserções do anúncio na programação de rádio, tv e internet durante o mês. Com o cuidado de mensurar o efeito desses anúncios, o proprietário somou, ao final dos meses em que fez o investimento com o anúncio, o número de unidades vendidas. Uma amostra dos resultados obtidos dos últimos 10 meses está ilustrada a seguir:

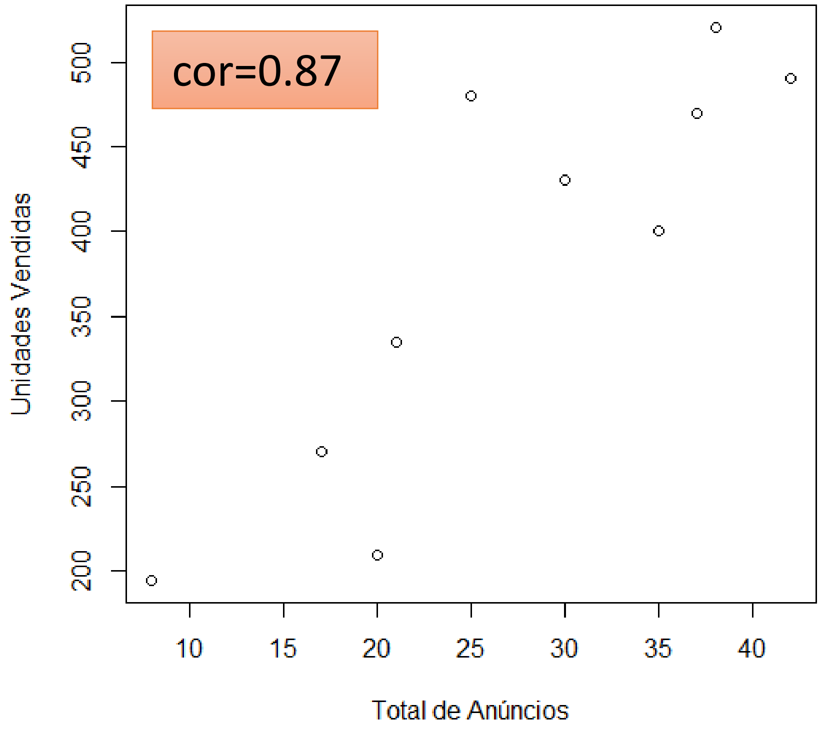
**unidades vendidas y = c(430, 335,520, 490, 470, 210, 195, 270, 400, 480)**

**Total de anúncios  x = c(30,21,38,42,37,20,8,17,35,25)**

Dispersão e correlação no R

```
plot(x,y)
cor(x,y)
```

$$\begin{pmatrix} 430 \\ 335 \\ \vdots \\ 480 \end{pmatrix} = \begin{bmatrix} 1 & 30 \\ 1 & 21 \\ \vdots & \vdots \\ 1 & 25 \end{bmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$




**Qual é o valor dos coeficientes de explicação?**

**Como interpretá-los?  Qual é tamanho do erro ?**

# REGRESSÃO LINEAR ESPACIAL

**Solucionando e interpretando o regressão linear no R:**

**modelo_linear <- lm(y ~ x)**

**summary(modelo_linear)**

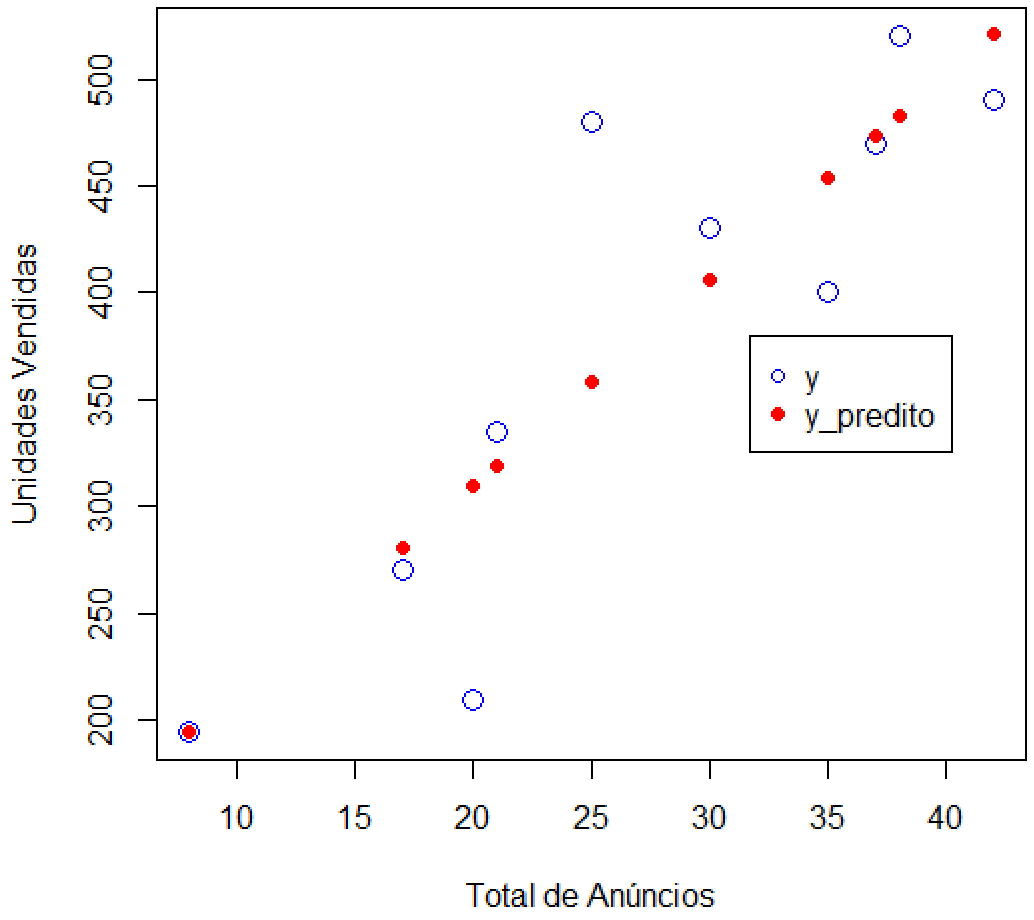
```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-99.775 -26.288  -1.325  21.921 122.126

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  117.377     55.695   2.108 0.068137 .
x              9.620      1.908   5.042 0.000999 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 62.36 on 8 degrees of freedom
Multiple R-squared:  0.7606,    Adjusted R-squared:  0.7307
F-statistic: 25.42 on 1 and 8 DF,  p-value: 0.0009992
```

Para cada 10 anúncio a mais, espera-se aumentar as vendas em 9.62*10 = 96 unidades vendidas. Se não investir em anúncios, espera-se vender 117 unidades. O erro médio do Modelo é 62.36 com poder explicativo de 0.7307 (70,07%). Isto é, através dos anúncios conseguimos explicar 70,07% das variações nas vendas.

## Predição no  modelo de regressão linear no R: y_preditor = 117.38 + 9.62x

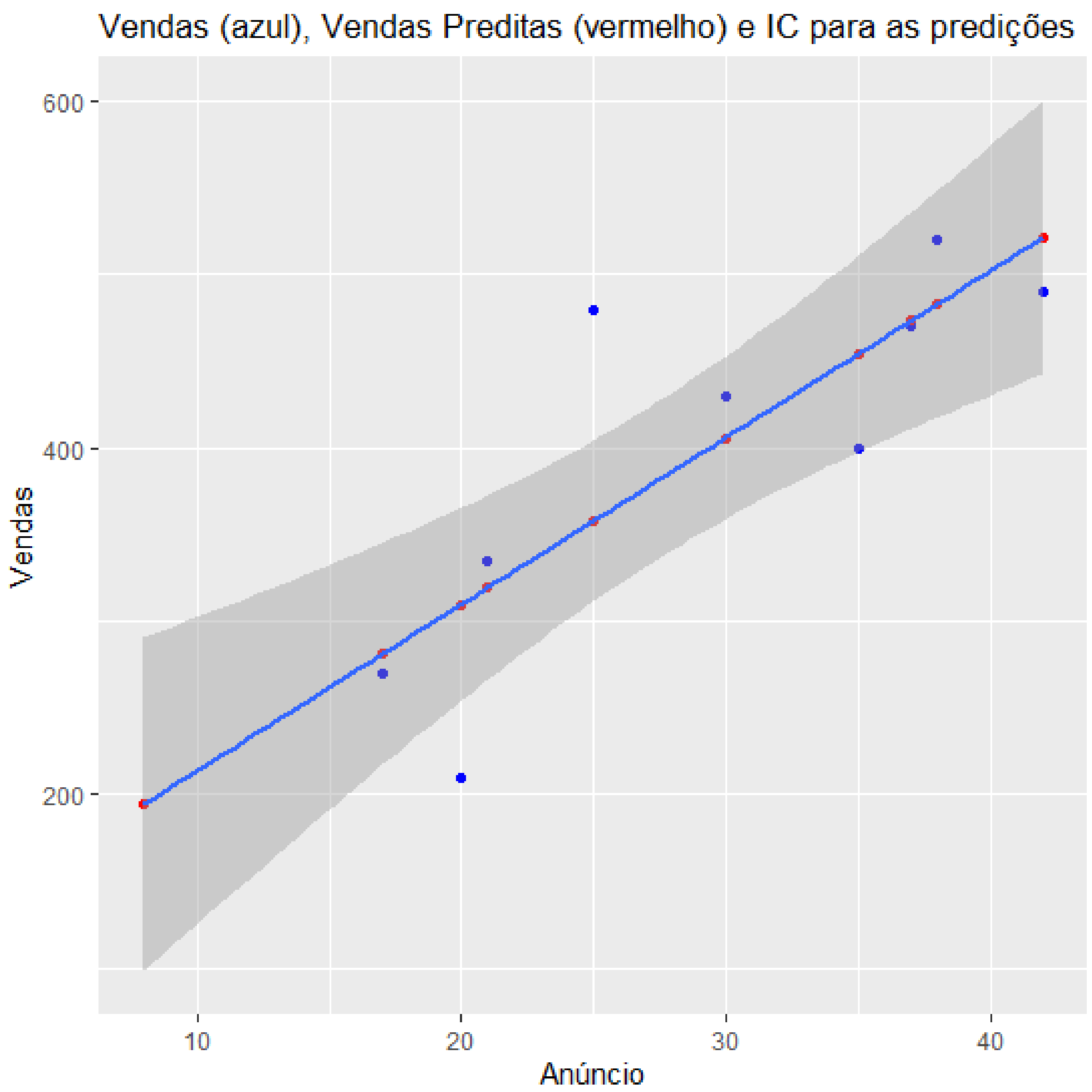
```
amostra_teste <- data.frame(x=x)
y_predito <- predict(modelo_linear, amostra_teste)
y_predito
plot(x,y,xlab="Total de Anúncios",ylab="Unidades Vendidas",type="p",col="blue",pch=1,cex=1.5)
points(x,y_predito,type="p",col="red",pch=16)
legend(locator(1),c("y","y_predito"),pch=c(1,16),col=c("blue","red"))
```

| Mês | Vendas | Vendas preditas |
|---|---|---|
| 1 | 430 | 405.9737 |
| 2 | 335 | 319.3947 |
| 3 | 520 | 482.9328 |
| 4 | 490 | 521.4123 |
| 5 | 470 | 473.3129 |
| 6 | 210 | 309.7748 |
| 7 | 195 | 194.3362 |
| 8 | 270 | 280.9152 |
| 9 | 400 | 454.0731 |
| 10 | 480 | 357.8743 |



**Plotando o Gráfico dos valores observados e Preditos para Vendas.**

```
##--construindo um data.frame dos dados
ml <- data.frame(x=x, y=y, y_predito = y_predito)

##--plotanto o grafico de vendas e preditos
ggplot(data=ml, aes(x=x, y=y, y_predito=y_predito)) +
geom_point(aes(x=x, y=y), colour = "blue") +
geom_point(aes(x=x, y=y_predito), colour = "red") +
geom_smooth(method = "lm", formula = y ~ x) +
labs(title = "Vendas (azul), Vendas Preditas (vermelho)
e IC para as predições", y = "Vendas", x = "Anúncio")
```



# REGRESSÃO LINEAR ESPACIAL

**O que é regressão linear Espacial?**

É um modelos de regressão que incorporam a dependência espacial nos dados através de uma estrutura de matriz de vizinhança. Os modelos mais usuais são:

**Modelo SAR para dependência global (Spatial AutoRegressivo): $y = \rho Wy + X\beta + \varepsilon$.**

**Modelo CAR para dependência global (Condicional AutoRegressivo):**

**$E(y|y_{[-i]}) = X\beta + \rho W(y - X\beta)$, W é a matriz Binária linha padronizada**

**Modelo de Regressaõ Ponderada (GWR - Geographically Weighted Regression):**

**$y(s) = X\beta(s) + \varepsilon$, um beta para cada local s dado por: $\beta(s) = (X'W(s)X)^{-1}XW(s)y$**

**A AULA CONTINUA NO ARQUIVO: regressão_espacial_doctor**

# Conclusão da Obra de Hoje!

A partir da análise de regressão espacial, Modelos SAR, CAR e GWR, além da análise do correlograma de Moran para o resíduo da regressão usual, foi possível identificar um padrão de dependência de 1ª ordem local para os casos de leucemia em NY8 1978-1982, sendo o modelo CAR o mais adequado com menor AIC. O uso da inteligência espacial nos evidenciou que o padrão de casos foi alterado após o ajuste por variáveis preditoras. Após uma análise minuciosa, percebemos através do Modelo CAR que o fator de exposição (proximidade da fonte TCE) foi o que mais influenciou o número de casos. Por fim, o gráfico de valores observado pelo ajustado e resíduos ajustados versus a predição evidenciaram visualmente um bom ajuste do modelo CAR estimado o qual foi utilizado para elaboração do mapa da predição de riscos relativos de casos, permitindo assim, identificar distritos isolados de alto risco.

# PERGUNTAS?

- Obrigado !