

Intro. Econometria Usando R

- Aula 1 -

Prof. Mestre. Omar Barroso Khodr

Instituto Brasileiro de Educação, Pesquisa e Desenvolvimento

O que é Econometria? E como Usar...

- Avalia relações econômicas, testa teorias e embasa decisões políticas/empresariais.
- E.g.:
- Mensuração do impacto de um programa de treinamento profissional nos salários dos trabalhadores.
- Teste de estratégias de investimento (por exemplo, títulos do Tesouro) em relação a teorias econômicas.
- Análise dos efeitos dos gastos de campanha em eleições ou do financiamento escolar no desempenho dos alunos.

O que é Econometria? E como Usar...

- A econometria concentra-se em dados não experimentais (**observacionais**) — coletados passivamente, não por meio de experimentos controlados (comum em ciências sociais devido a restrições de custo/éticas).
- Compara-se com dados experimentais (por exemplo, ambientes de laboratório em ciências naturais).

O que é Econometria? E como Usar...

- Utiliza ferramentas estatísticas (por exemplo, regressão múltipla), mas as adapta para a complexidade dos dados econômicos.
- Desenvolve novas técnicas para abordar problemas econômicos específicos (por exemplo, endogeneidade e causalidade).
- Formula a pergunta (por exemplo, testa uma teoria ou efeito de política).
- Constrói modelos econômicos (por exemplo, maximização da utilidade para equações de demanda).
- Usa dados para testar/estimar relações (por exemplo, como preços/renda afetam a demanda do consumidor).

O termo de Erro

- A principal distinção entre economistas e econometristas é a preocupação com os termos de erro.
- Um economista especificará, por exemplo, que o consumo é uma função da renda, escrevendo $C = f(Y)$, onde C é o consumo e Y é a renda.
- Um econometrista inferirá que essa relação também deve incluir um termo de perturbação (ou erro) e pode alterar a equação para $C = f(Y) + \varepsilon$, onde ε é um termo de erro.
- Sem ε [o termo de erro] o modelo é considerado exato ou determinístico; com o erro a relação é considerada estocástica.
- Uma relação estocástica nem sempre é precisa no sentido de que prevê o valor preciso da variável que está sendo explicada.

O termo de Erro

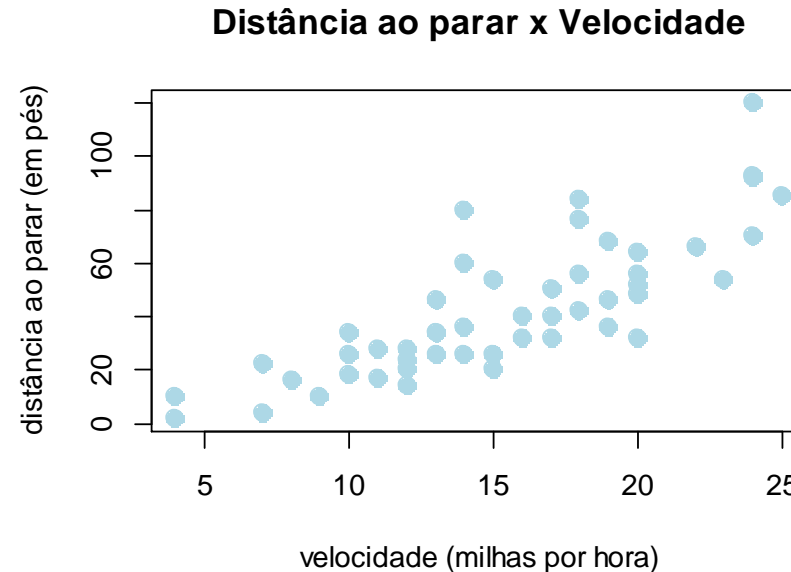
- a existência do termo de erro é justificado de três maneiras.
- 1. **Omissão:** de variáveis e eventos.
- 2. **Erros de Mensura:** Quando as variáveis explicativas não conseguem mensurar corretamente uma relação econômica.
- 3. **Indeterminismo Humano:** A aleatoriedade de ε representa a aleatoriedade do comportamento humano.

Parâmetros

- Associadas a qualquer relação explicativa **estão constantes desconhecidas**, chamadas parâmetros, que vinculam as variáveis relevantes a uma equação. Por exemplo, a relação entre consumo e renda poderia ser especificada como:
- $C = \beta_1 + \beta_2 Y + \varepsilon$
- β_1 e β_2 caracterizam a função de consumo.
- A existência do termo de erro impossibilita o cálculo desses parâmetros, dessa maneira eles devem ser estimados.
- Tentamos estimar os betas de maneira empírica [estatística], no qual, β é referido como um vetor.

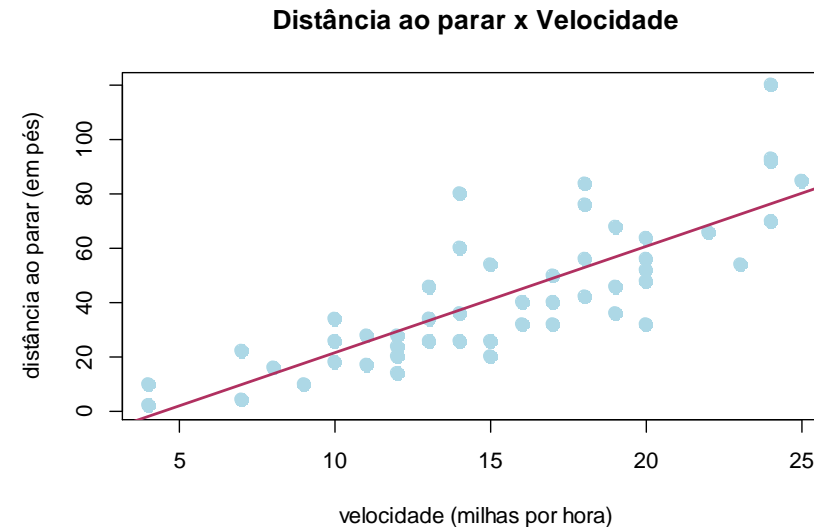
Parâmetros

- Vamos utilizar um exemplo prático, entre a distância e velocidade de um veículo... E como estimar os parâmetros.
- Vamos nomear nossa coluna Y como distância (ao parar) e X como velocidade.
- Utilizaremos o gráfico de dispersão, para verificar a relação entre X e Y (ou distância e velocidade).
- Podemos ver que a relação é positiva entre as duas variáveis.

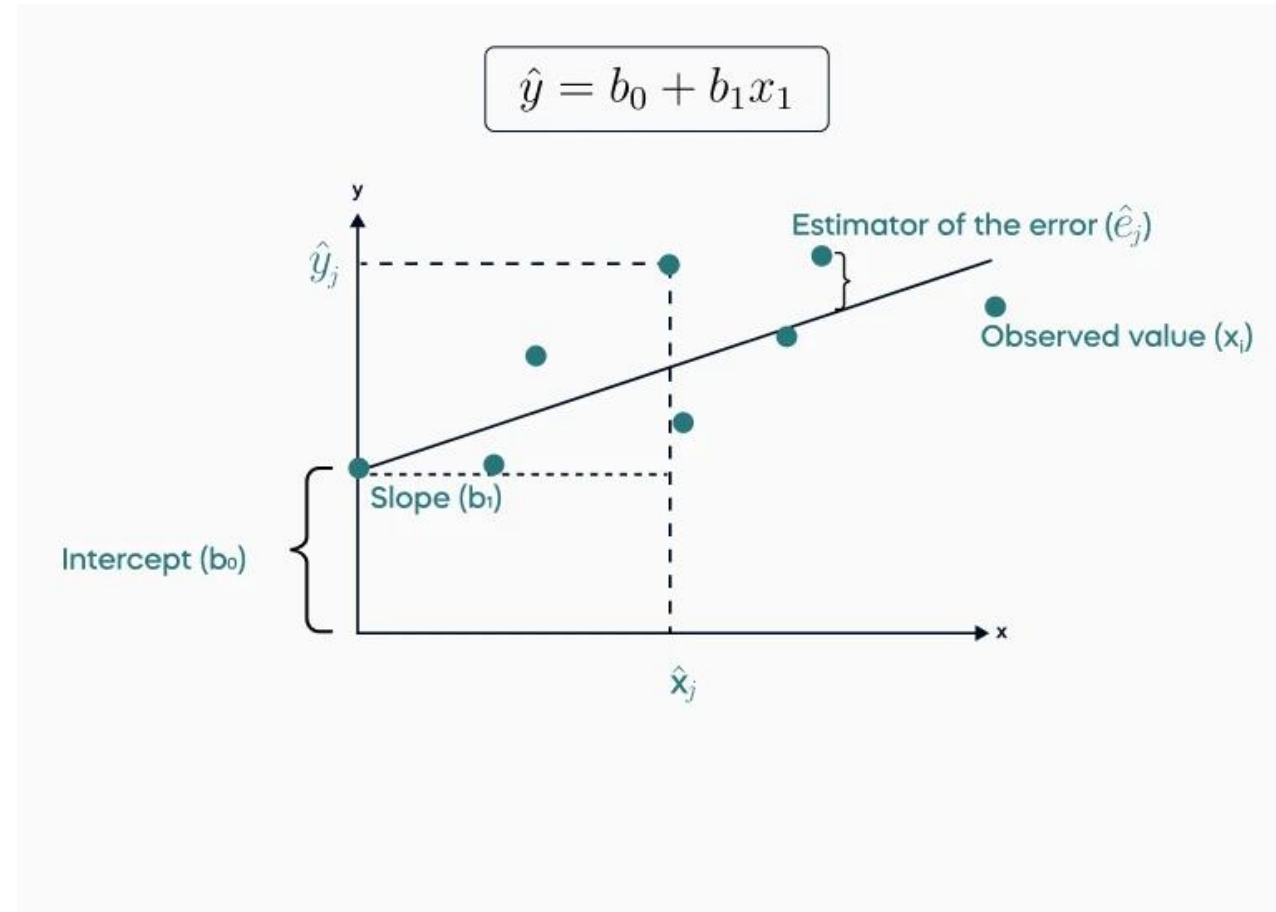


Parâmetros

- Vamos supor que nossa regressão é feita da determinada maneira,
- $D = \beta_1 + \beta_2 V$
- Se tivermos que ajustar uma linha para capturar a relação, provavelmente ela seria desta maneira...
- Acabamos de aprender como representar uma reta, ou seja, com dois números chamados intercepto (β_1) e coeficiente angular (β_2).
- Como podemos escolher β_1 e β_2 como um número concreto?



Parâmetros (nota)



Ajustando a linha

- Lembrando, do termo de erro... Em nosso modelo de distância e velocidade não conseguimos estimar os parâmetros.
- Dessa maneira, faremos um modelo estimativo para que a distância entre os erros e a linha seja diminuída.
- $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$
- \hat{y}_i : O valor que será projetado (no qual representa um vetor em T).
- $\hat{\beta}_n$: Os coeficientes a serem estimados (que será um vetor também).

MQO (Mínimos Quadrados Ordinários)

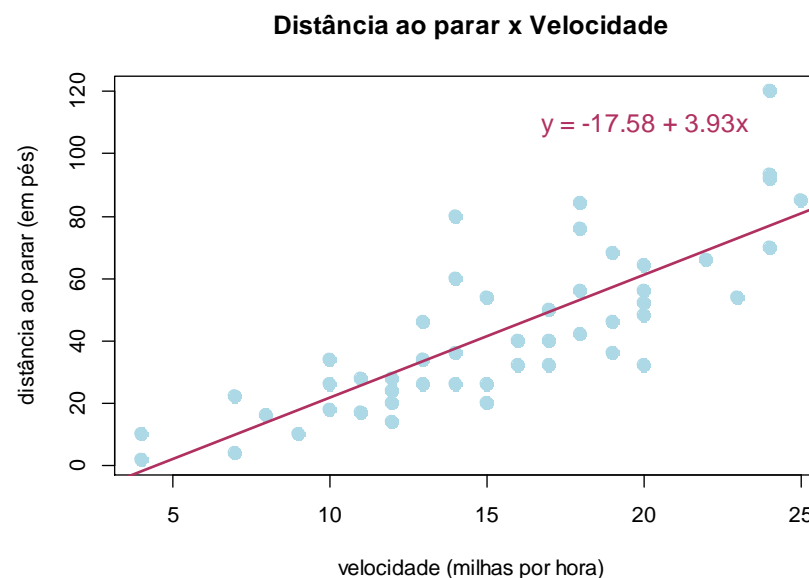
- Assumindo a relação linear...
- $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$
- Podemos estimar os coeficientes.
- Precisamos de;
- N = número de observações.
- $\sum X$
- $\sum Y$
- $\sum XY$
- $\sum X^2$

MQO (Mínimos Quadrados Ordinários)

- Assumindo a relação linear...
- $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$
- N = número de observações.
- $\sum X = 4 + 4 + 7 + 7 + \dots = 770$
- $\sum Y = 2 + 10 + 4 + 22 + \dots = 2149$
- $\sum XY = (4.2) + (4.10) + (7.4) + \dots$
- $\sum X^2 = 4^2 + 4^2 + 7^2 + \dots$

MQO (Mínimos Quadrados Ordinários)

- Assumindo a relação linear...
- $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i$
- Agora temos que calcular $\hat{\beta}_2$ ou nosso coeficiente angular.
- $\hat{\beta}_2 = \frac{\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} \sim 3.93$
- $\hat{\beta}_1 = \frac{\sum y - \hat{\beta}_2 \sum x}{n} \sim -17.58$
- Desta maneira,
- $\hat{y} = -17.58 + 3.93x$



Valores Ajustados e Resíduos

- Como obtemos, $\hat{y} = -17.58 + 3.93x$;
- O próximo passo é calcular os valores da coluna 'x' que está faltando. Ou seja..
- $-17.58 + 3.93(4) = -1.89$; $-17.58 + 3.93(4) = -1,89$; $-17.58 + 3.93(7) = 9.93$;...
- Com isso, obtemos os valores de \hat{y} .
- Logo após, devemos achar os resíduos...
- $e = y - \hat{y}$

Valores Ajustados e Resíduos

- $e = y - \hat{y}$
- Ou seja,
- $2 - (-1.89) = 3.89$; $10 - (-1.89) = 11.98$;...
- Assim, pelo MQO os coeficientes são estimados para minimizar a soma dos resíduos quadrados.
- Ou seja, pela condição de primeira ordem... $\sum e = 0$.
- Isso garante ortogonalidade, ou seja $\sum ex_i = 0$.

Nota sobre Resíduos

- Se, $\sum e \neq 0$... Existem questões que devemos considerar.
- Se a soma dos resíduos não for zero, o modelo pode estar especificado incorretamente (por exemplo, faltando um intercepto ou variáveis-chave).
- Exemplo: Se os resíduos forem majoritariamente positivos, o modelo sistematicamente subestima Y .
- Y ; se forem negativos, ele superestima.
- Resíduos grandes indicam ajuste inadequado para determinadas observações.
- Resíduos pequenos sugerem que o modelo explica bem os dados.

Nota sobre Resíduos

- Se, $\sum e \neq 0$... Existem questões que devemos considerar.
- Podemos ter violações da regressão, como:
- Heteroscedasticidade (variância não constante entre resíduos e valores ajustados); Autocorrelação (séries temporais ou dados em painel); Não-linearidade; Variáveis Omitidas; e, Outliers.

Nota sobre Resíduos

- Se, $\sum e \neq 0$... Existem questões que devemos considerar.
- No exemplo que vimos, a soma dos resíduos é violada. Ou seja,
- Possivelmente, o modelo superestima Y em média.
- Talvez precisa de mais variáveis...
- Talvez tenha vieses indicando erros sistemáticos de previsão.
- Dessa maneira, devemos investigar mais...

Bibliografia

- Wooldridge, J.M. (2013) Introductory econometrics: a modern approach. 5th ed. Michigan State University.