

ANÁLISE DE DADOS MULTIVARIADOS I - REGRESSÃO

Novembro e dezembro de 2018

Reinaldo Soares de Camargo

ANÁLISE DE DADOS MULTIVARIADOS I - REGRESSÃO

- **Objetivo:**

- Transmitir os principais conceitos de:
 - Regressão linear simples e múltipla;
 - Estimação dos parâmetros via métodos de mínimos quadrados ordinários;
 - Previsão usando modelos de regressão;
 - Intervalos de confiança e testes de hipóteses em modelos de regressão;
 - Estimação via máxima verossimilhança;
 - Modelos de regressão probit e logit.
 - Aplicações práticas em R, para todos os tópicos.

- **Metodologia:**

- Aulas expositivas
- Discussão de estudos aplicados
- Listas de exercícios em grupo de 2 alunos.

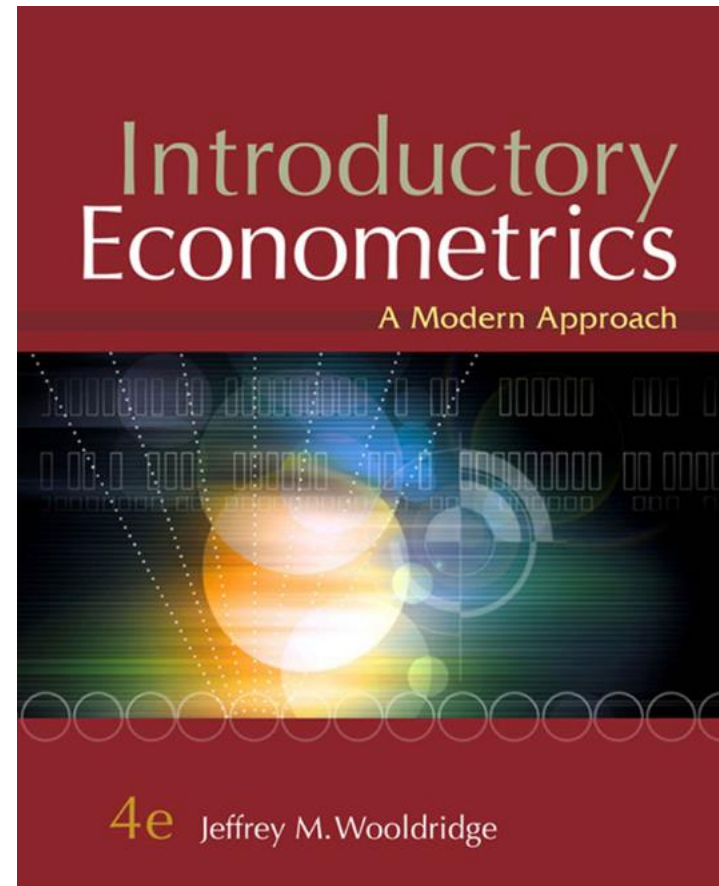
ANÁLISE DE DADOS MULTIVARIADOS I - REGRESSÃO

- **Avaliação:**

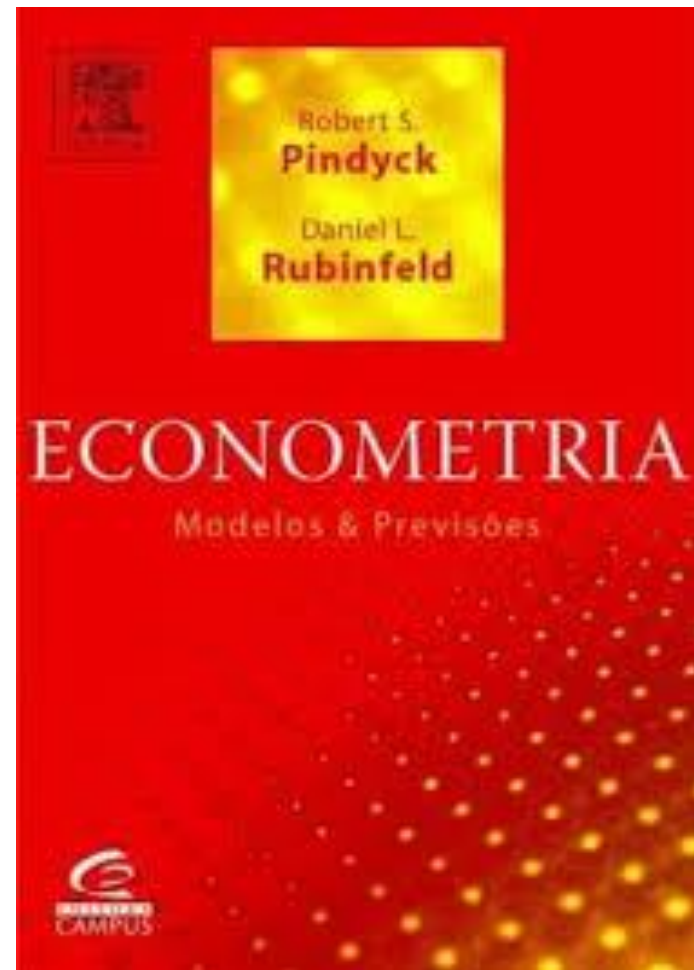
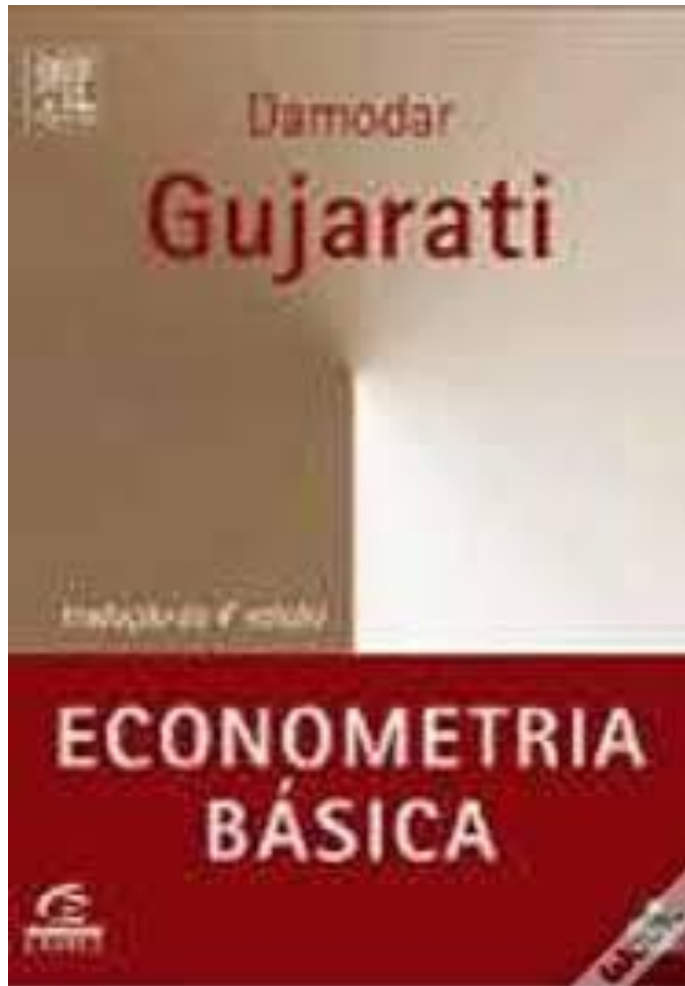
A avaliação do curso será baseada nas listas de exercícios, propostas toda segunda-feira para entrega na segunda-feira da próxima semana, valendo 40% da nota final, e em uma avaliação individual escrita em sala de aula, valendo 60% da nota final. As listas de exercícios devem ser respondidas em grupos de 2 alunos, e entregues até as datas previamente estipuladas. A avaliação final é individual e ocorrerá no penúltimo dia de aula. Para a avaliação em sala de aula será permitido o uso de computadores.

Plano de Aula

Referência básica



Referência básica



Aula 01

01 de novembro (quinta-feira 9h às 12 horas)

- Apresentação do plano de aula
- Metodologia da Econometria
- Tipos de dados

Metodologia da econometria

- **O que é econometria**

Econometria significa literalmente “medida econômica”. Embora a medida seja uma parte importante da econometria, sua finalidade é mais abrangente.

A teoria econômica faz afirmações ou formula hipóteses em sua maioria de natureza qualitativa. Por exemplo, a teoria microeconômica afirma que, tudo o mais constante (*ceteris paribus*) uma redução no preço de uma mercadoria deverá aumentar a quantidade demandada dessa mercadoria. Assim, a teoria econômica postula uma relação negativa ou inversa entre preço e quantidade demandada de uma mercadoria. Mas a teoria em si não fornece qualquer medida numérica da relação entre os dois, ou seja, ela não diz a quantidade que irá aumentar ou diminuir em consequência de uma determinada alteração no preço das mercadorias. Cabe ao econometrista fornecer tais estimativas numéricas. Em outras palavras, a econometria dá conteúdo empírico à grande parcela da teoria econômica.

Metodologia da econometria

- Como os econometristas procedem em sua análise de um problema econômico? Ou seja, qual é sua metodologia? Embora haja diversas escolas do pensamento em matéria de metodologia econométrica, apresentamos aqui a metodologia tradicional ou clássica, ainda predominante na pesquisa empírica em economia e áreas correlatas.
- A metodologia econométrica tradicional segue os seguintes passos:
 1. Formulação da teoria ou da hipótese;
 2. Especificação do modelo matemático da teoria;
 3. Especificação do modelo econométrico da teoria;
 4. Obtenção dos dados;
 5. Estimativa dos parâmetros do modelo econométrico;
 6. Teste de hipótese;
 7. Previsão ou predição;
 8. Utilização do modelo para fins de controle ou política.

1. Formulação da teoria ou da hipótese.

Teoria Keynesiana do Consumo.

“A lei psicológica fundamental.... É que os homens e mulheres, como regra e na média, se dispõem a aumentar seu consumo quando sua renda aumenta, mas não tanto quanto o aumento da renda.”

Keynes postula que a Propensão marginal a consumir (PMgC) – A taxa de variação do consumo – é maior que zero mas menor que 1.

2. Especificação do modelo matemático.

Embora Keynes tenha postulado uma relação positiva entre renda e consumo, ele não especificou a forma precisa da relação funcional entre os dois.

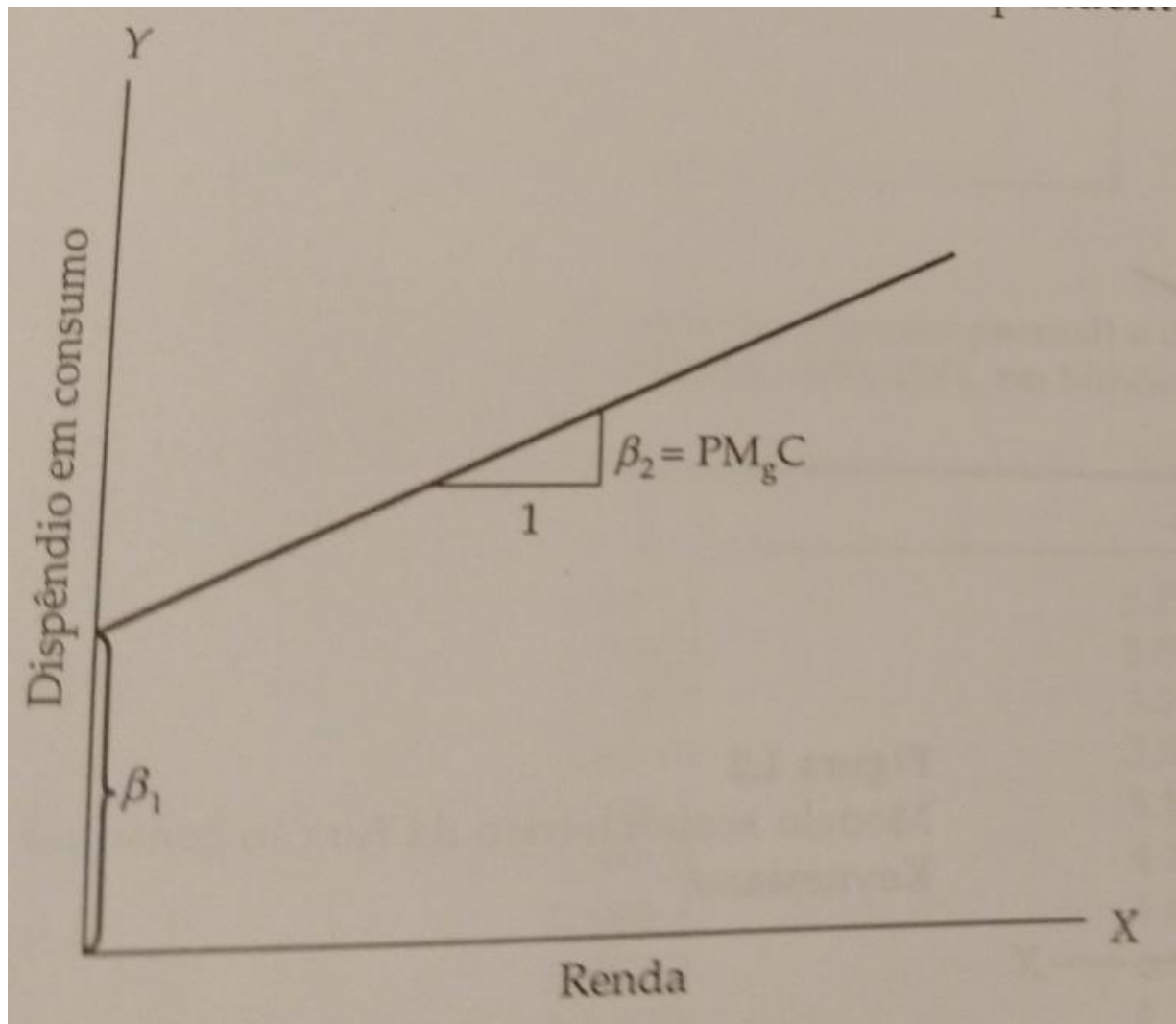
Por simplicidade, os alunos do curso Análise de dados multivariado I – Regressão da ENAP, poderia sugerir a seguinte forma da função consumo keynesiana.

$$Y = \beta_1 + \beta_2 X \qquad 0 < \beta_2 < 1$$

Em que Y = despesa de consumo; e X = Renda, e β_1 e β_2 , conhecidos como parâmetros do modelo, são os coeficientes do intercepto e declividade.

Y = Variável dependente; β_1 = Intercepto, ou coeficiente linear; β_2 = Coeficiente angular (declividade)

2. Especificação do modelo matemático.



3. Especificação do modelo econométrico

O modelo matemático da função consumo, tem interesse restrito para os econométricos, uma vez que supõe uma relação exata ou *determinista* entre consumo e renda. Porém as relações econômicas entre variáveis são geralmente inexatas.

Se quisermos obter dados de 500 famílias brasileiras é de se esperar que os valores se situem em torno da média, porém não na média, visto que o consumo não depende somente da renda, não existe essa relação exata, podemos dizer que o consumo depende ainda de variáveis como tamanho da família, idades de seus membros, religião, etc.

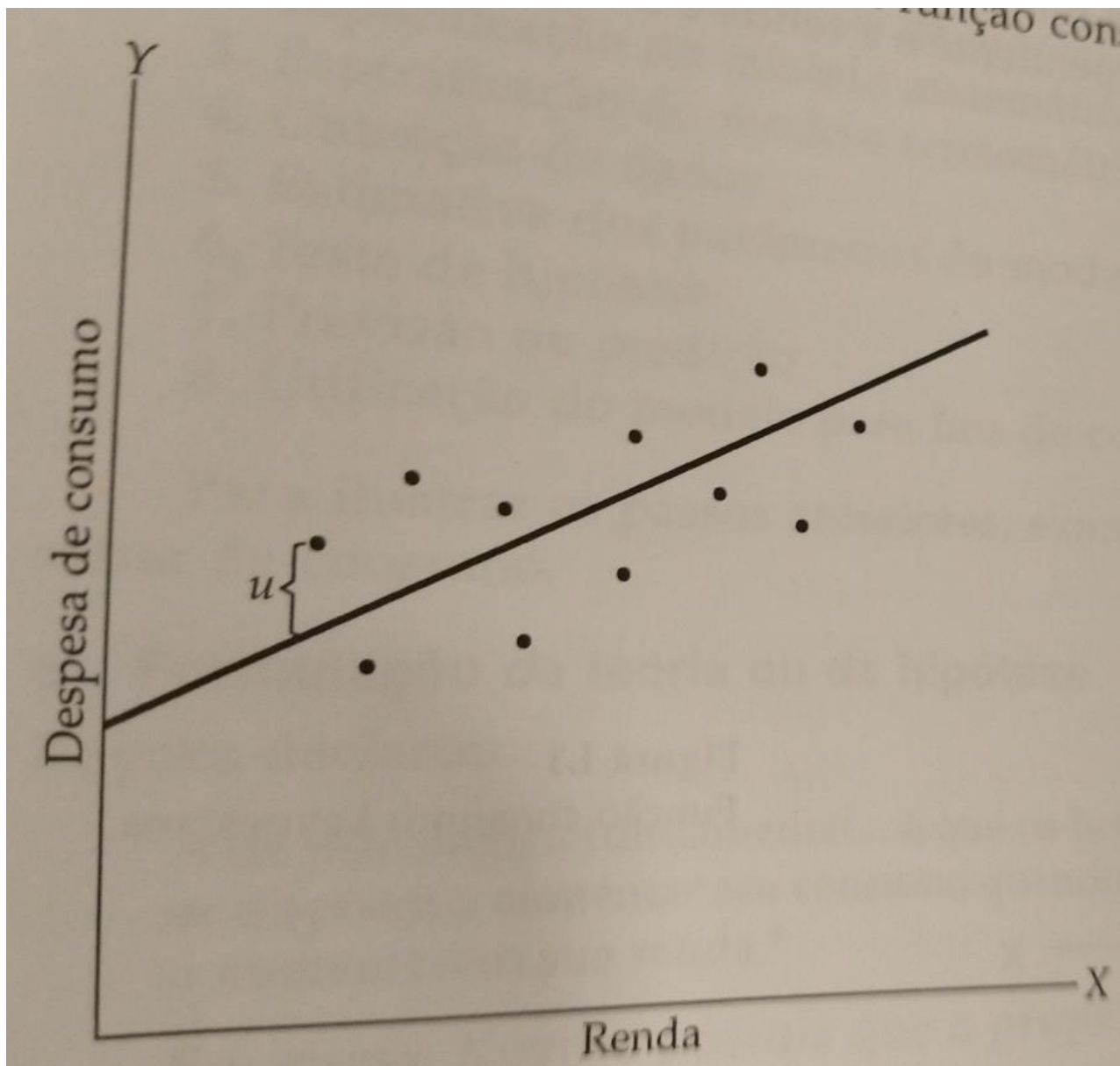
3. Especificação do modelo econométrico

Para considerar relações inexatas entre renda e consumo, o econometrista modificaria a função consumo para

$$Y = \beta_1 + \beta_2 X + u$$

Em que u é conhecido como termo de perturbação ou erro, é uma variável aleatória (estocástica) que possui propriedades probabilísticas bem definidas. O termo u pode representar todos os fatores que afetam o consumo das famílias, mas que não são considerados explicitamente (na equação).

3. Especificação do modelo econométrico



4. Obtenção dos dados

Para estimar modelos econométricos precisamos de dados para obter os valores numéricos β_1 e β_2 . Consideremos os seguintes dados.

Y=(despesa de consumo pessoal) e X = PIB, 1980-1991
EUA em bilhões de dólares.

Ano	Y	X
1980	2.447,1	3.776,3
1981	2.476,9	3.843,1
1982	2.503,7	3.760,3
1983	2.619,4	3.906,6
1984	2.746,1	4.148,5
1985	2.865,8	4.279,8
1986	2.969,1	4.404,5
1987	3.052,2	4.539,9
1988	3.162,4	4.718,6
1989	3.223,3	4.838,0
1990	3.260,4	4.877,5
1991	3.240,8	4.821,0

5. Estimativa dos parâmetros do modelo econométrico

De posse dos dados, precisamos estimar os parâmetros do modelo, o procedimento concreto de estimativa de parâmetros será tema de nossas próximas aulas, por hora observe a técnica estatística de **análise de regressão** é a principal ferramenta para obter as estimativas.

$$\hat{y} = -231,80 + 0,7194X$$

O acento circunflexo em y indica que se trata de uma estimativa.

O coeficiente de declividade (angular) que representa neste exemplo a PMgC foi de 0,7194, sugerindo que um aumento de um dólar na renda real provocará, em média um aumento de aproximadamente 72 centavos no consumo.

O intercepto -231,80 representa o consumo para quem não possui renda ($x = 0$), geralmente não possui significância econômica.

6. Teste de hipótese

Supondo que o modelo ajustado seja uma boa aproximação da realidade, temos de desenvolver critérios adequados para descobrir se as estimativas obtidas satisfazem as expectativas da teoria que está sendo testada.

Como dito anteriormente, Keynes supunha que a PMgC era positiva porém menor do que 1. Em nosso exemplo obtivemos 0,72. Antes de aceitarmos esse resultado com confirmação da teoria, devemos averiguar se esta estimativa está suficientemente abaixo de 1 para nos convenceremos de que não se trata de uma ocorrência casual ou uma peculiaridade dos dados.

Em outras palavras precisamos testar se 0,72 é estatisticamente menor do que 1? Se for, pode sustentar a Teoria de Keynes.

Essa confirmação ou rejeição da teoria econômica com base na evidência da amostra se baseia em um ramo da teoria estatística conhecido como inferência (teste de hipótese).

7. Previsão ou predição

Se o modelo escolhido confirmar a hipótese ou a teoria em consideração, podemos usá-lo para prever os valores futuros da variável dependente Y, com base em valores futuros conhecidos ou esperados da variável explicativa X.

Suponha uma expectativa de PIB real de US\$ 6 bilhões em 1994. Qual previsão de consumo em 1994?

$$\hat{y} = -231,80 + 0,7194 (6.000) = 4.084,6$$

8. Uso do modelo para fins de controle ou elaboração de Políticas Públicas.

Suponha que o governo acredite que um nível de gastos de US\$ 4.000 (bilhões de dólares de 1987) manterá a taxa de desemprego no patamar de cerca de 6,5% (segundo estimativas agência de estatísticas EUA). Que nível de renda garantirá o montante almejado de consumo?

$$\begin{aligned}\hat{y} &= -231,80 + 0,7194X \\ 4.000 &= -231,80 + 0,7194X \\ X &= 5.882\end{aligned}$$

Ou seja, um nível de renda de US\$ 5.882 (bilhões) dada uma PMgC de 0,72, produzirá um gasto de US\$ 4.000 bilhões.

Como sugere esses cálculos um modelo estimado pode ser usado para fins de controle ou políticas econômica. Combinando políticas fiscais e monetária apropriadas, o governo pode manipular a variável de controle X para produzir o nível desejado da variável alvo Y.

Aplicações em R.

- 1) Importe o arquivo **dados_consumo_renda_eua_1980_1991.xlsx**, criando um dataset com o nome dados.
- 2) Esboçe um gráfico das variáveis **X e Y**;
- 3) Obtenha as estatísticas descritivas das variáveis X e Y;
- 4) Estime um modelo de regressão simples, e identifique na saída do computador as estimativas apresentada na aula de hoje.

Aplicações em R.

1) Importe o arquivo **dados_consumo_renda_eua_1980_1991.xlsx**, criando um dataset com o nome **dados**.

2) Esboçe um gráfico das variáveis **X** e **Y**;
plot(dados\$x, dados\$y)

3) Obtenha as estatísticas descritivas das variáveis **X** e **Y**;
Summary(dados)

4) Estime um modelo de regressão simples, e identifique na saída do computador as estimativas apresentada na aula de hoje.

```
ex1 <- lm( Y ~ X, data = dados)  
summary(ex1)
```

Tipos de dados

- Dados experimentais
 - . Demanda por um automóvel;
 - . Satisfação de usuários de serviços públicos.
- Dados não experimentais
 - **Séries temporais:** Dados coletados ao longo de intervalos discretos de tempo: PIB, Inflação, Cotação de Ações na Bovespa, etc.
 - **Dados de Cortes:** Dados coletados sobre unidades de amostra em determinado período: Renda média dos brasileiros extraídas do Censo 2000 ; taxa de matrícula no ensino fundamental em 2018.
 - **Dados de Painel:** Dados que acompanham microunidades individuais ao longo do tempo: pesquisa para acompanhar o desempenho de alunos desde o ensino fundamental até a conclusão do ensino superior

Obrigado.