Notas Econometria I

Thiago Oliveira Coelho

11 de fevereiro de 2020

Resumo baseado em (GUJARATI; PORTER, 2011) e (WOOLDRIDGE, 2016)

Sumário

1 1ª Unidade

1.1 Análise de Regressão Simples

É usada para determinar o impacto ceteris paribus entre variáveis. Seu modelo simples é o seguinte:

$$Y = \beta_0 + \beta_1 X + u$$

Aonde:

- Y: Variável dependente;
- X: Variável independente;
- β_0: Intercepto (Valor de Y quando X = 0);
- β_1: Coeficiente do impacto de X em Y;
- u: Erro. Inclui todos os fatores que influenciam Y e não estão explicitados no modelo como variáveis explicativas.

Tanto X quanto u serão tratadas como variáveis aleatórias. Mas para capturar o efeito ceteris paribus das variáveis explicativas nós assumimos que $E(u) = 0$. Com isso podemos derivar também que : $E(u|x) = E(u) = 0$. Isso permite que X tenha um efeito *linear* em Y.

Encontramos assim nossa *Função regressão populacional* [FRP]:

$$E(Y|X) = \beta_0 + \beta_1 X$$

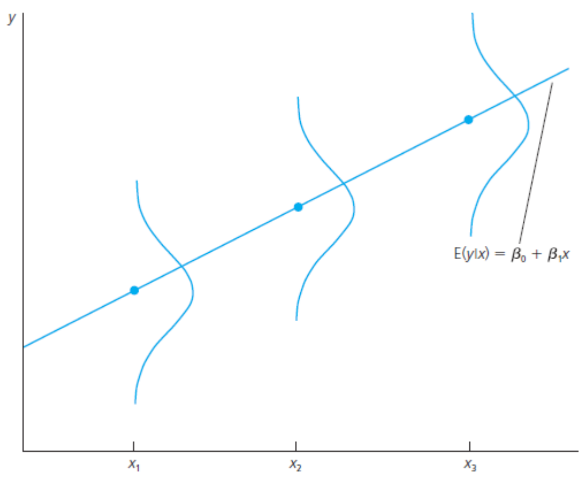

Figura 1 – Fonte: (WOOLDRIDGE, 2016)

Observe que para cada observação de X há uma distribuição de valores de Y, mas os que estão na linha de regressão representam a esperança (média) destes valores para aquele dado X. A linha de regressão é o local geométricos das expectativas condicionais das variáveis dependentes dados certos valores das explicativas.

1.2 Função Regressão Amostral [FRA]

Dada uma amostra da população, existe uma reta de regressão que define esta amostra. Porém não é possível se obter esta reta pura por causa da pertubação estocástica (u). Portanto iramos achar estimativas dos parâmetros:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_i$$

Aonde os betas não são seus valores reais, mas sim os *estimadores* obtidos a partir da amostra, representados pelo chapéu.

1.3 Estimativa de β por mínimos quadrados ordinários

1.3.1 MQO / OLS

Pelas hipóteses podemos deduzir:

$$E(u) = Cov(x, u) = E(Xu) = 0$$

$$u = Y - \beta_0 - \beta_1 X$$

$$E(X(Y - \beta_0 - \beta_1 X)) = 0$$

Montamos o sistema:

$$\frac{\sum_{i=1}^{n} Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i}{n} = 0 \tag{1.1}$$

$$\frac{\sum_{i=1}^{n} X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)}{n} = 0 \tag{1.2}$$

Pelas regras de somatório:

$$\frac{\sum_{i=1}^{n} Y_i}{n} - \frac{\sum_{i=1}^{n} \hat{\beta}_0}{n} - \frac{\sum_{i=1}^{n} \hat{\beta}_1 X_i}{n} = 0$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

Substituindo o valor de β_0 na segunda equação:

$$\sum_{i=1}^{n} X_i(Y_i - (\overline{Y} - \hat{\beta}_1 \overline{X}) - \hat{\beta}_1 X_i = 0$$

Por manipulação algébrica chegamos no seguinte sistema de equações linear (Todos os betas são estimativas):

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (X_i - \overline{X})^2}$$

Percebe que a estimativa de beta 1 chapéu é dada pela covariância entre X e Y sobre a variância de X, portanto o sistema fica:

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \tag{1.3}$$

$$\hat{\beta}_1 = \frac{Cov(X, Y)}{Var(X)} \tag{1.4}$$

1.3.2 Soma dos quadrados dos resíduos

O resíduo de uma observação é dada por:

$$\hat{u}_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1)$$

Existem n resíduos, que nos dizem a diferença entre os valores esperados de Y estimados pelo modelo e os valores das observações.

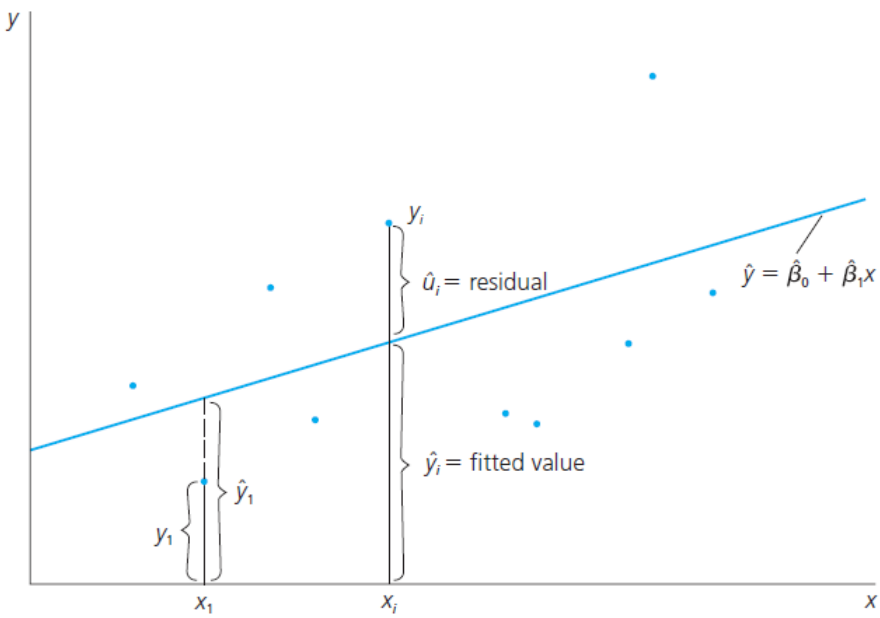

Figura 2 – Fonte: (WOOLDRIDGE, 2016, p. 28)

$$\sum_{i=1}^{n} \hat{u}_i^2$$

A ideia é obter uma reta com o menor resíduo possível, para isso queremos minimizar a soma dos quadrados dos resíduos:

$$\sum_{i=1}^{n} Y_i - \hat{Y}_i$$

Substituímos Y_i e obtemos a equação final:

$$Q = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

Com isso nosso objetivo passa a ser minimizar Q com relação a $\hat{\beta}_0$ e $\hat{\beta}_1$.

Teremos então as condições de 1ª ordem:

$$\frac{dQ}{d\hat{\beta}_0} = 0 \rightarrow -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1)$$

$$\frac{dQ}{d\hat{\beta}_1} = 0 \rightarrow -2\sum_{i=1}^{n}X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1)$$

Podemos ainda fazer uma transformação monotônica das duas equações dividindo ambas por -2 e obter:

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$$

$$\sum_{i=1}^{n}X_i(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)$$

Que equivale ainda as duas hipóteses populacionais a quais chegamos ao tentar deduzir as fórmulas dos estimadores 1.1 e 1.2. Se continuarmos a desenvolver as fórmulas chegamos ao mesmo sistema para determinarmos os estimadores: 1.3 e 1.4. Concluímos então que as fórmulas do estimadores minimizam a soma dos quadrados dos resíduos.

1.4 Reta de regressão

A reta de regressão é aquela que minimiza os resíduos quadrados, que, como vimos anteriormente, é aquela reta que tem como intercepto $\hat{\beta}_0$ e como inclinação $\hat{\beta}_1$. A reta de regressão passa por todos os valores esperados dado x e consegue te dar os valores previstos $\hat{Y}$ para todo valor de variável explicativa, até mesmo aquelas que não possuímos observações.

$$\frac{d\hat{Y}}{dx} = \hat{\beta}_1 \Rightarrow \triangle\hat{Y} = \hat{\beta}_1\triangle X$$

A mudança em $\hat{Y}$ é igual a mudança em X vezes o coeficiente da reta $\hat{\beta}_1$. Esta previsão vale para a população como um todo.

1.5 Forma funcional

Um modelo é linear por causa de seus parâmetros, não por suas variáveis explicativas ou variável dependente. Todos os modelos a seguir são lineares, consequentemente nós podemos utilizar mínimos quadrados para obtermos seus coeficientes:

$$\log Y = \beta_0 + \beta_1 X^2 + u$$

$$Y = \beta_0 + \beta_1\frac{1}{X} + u$$

Temos quatro principais formas funcionais:

1. Linear em Y e X $\rightarrow Y = \beta_0 + \beta_1 X + u$, neste caso $\frac{dY}{dX} = \beta_1$.
2. Log em Y e X $\rightarrow \log Y = \beta_0 + \beta_1 \log X + u$, neste caso $\frac{d \log Y}{d \log X} = \beta_1 \approx \epsilon_{yx}$. β_1 indica a mudança percentual em Y dada alteração de 1% em X.
3. Log em Y, linear em X $\rightarrow \log Y = \beta_0 + \beta_1 X + u$. Multiplicamos β_1 por 100 para obtermos o aumento percentual de Y dada mudança unitária em X.
4. Linear em Y, log em X $\rightarrow Y = \beta_0 + \beta_1 \log X + u$, neste caso β_1 mostra o efeito unitário em Y de uma mudança de 1% de X. Este caso é chamado de semi elasticidade.

1.6 R^2

Podemos criar uma lista das propriedades mais importantes da MQO amostral vistas até agora:

1. Soma dos resíduos é igual a zero: $\sum \hat{u}_i = 0$;
2. Covariância amostral entre x e resíduos na amostra é zero: $\sum X_i \hat{u}_i = 0$;
3. Os pontos $(\overline{X}, \overline{Y})$ sempre estarão na reta de regressão, dado que:
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_i \Rightarrow X = \overline{X} \rightarrow Y = \overline{Y}$$
4. Covariância amostral entre $\hat{Y}$ e $\hat{u}$ é zero;
5. A média amostral de $\hat{Y}$ é igual a média amostral de Y: $\overline{\hat{Y}} = \overline{Y}$;
6. $Y = \hat{Y} + \hat{u}$;

Podemos então, a partir da propriedade 6, chegar nas seguintes abreviações:

Soma dos Quadrados totais:
$$SQT = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \tag{1.5}$$
Soma dos quadrados explicados:
$$SQE = \sum_{i=1}^{n} (\hat{Y}_i - \overline{Y})^2 \tag{1.6}$$
Soma dos quadrados dos resíduos:
$$SQR = \sum_{i=1}^{n} (\hat{u}_i)^2 \tag{1.7}$$
Por fim, temos que:
$$SQT = SQE + SQR \tag{1.8}$$

O R^2, também chamado de coeficiente de determinação, pode ser definido de diferentes formas:

1. *Quanto da variância de Y é explicada por X, em termos proporcionais*;

2. Relação entre variância de Y e de $\hat{Y}$;

3. Proporção da variância de Y explicada pelo modelo de regressão;

$$R^2 = \frac{SQE}{SQT} = 1 - \frac{SQR}{SQT} \tag{1.9}$$

logo:

$$0 \leq R^2 \leq 1$$

R^2 seria igual 1 somente em duas ocasiões:

1. SQR = 0, o que ocorre somente se todos os erros dão zero. Isso significaria que $\hat{Y}_i = Y_i$ para todo i, todos os valores reais de Y estariam em cima da reta de regressão;

2. SQE = SQT, o que significa que seu modelo amostral se iguala perfeitamente a realidade. Isso novamente implicaria que $\hat{Y}_i = Y_i$ para todo i.

Se mudarmos a forma funcional, alteramos o valor do R quadrado.

2 2ª Unidade

2.1 Eliminação de viés nos estimadores

Algumas das precauções que podem ser tomadas para termos estimadores não enviesados:

1. Amostragem aleatório: é aquela que representa a população inteira, e não somente grupos específicos. Para isso todos os membros de população tem de ter alguma chance de participar da amostra;

2. Auto seleção: quando partes da população pode intencionalmente decidir participar da amostragem;

3. Média condicional de erro igual a zero: $E(u|X) = 0$. Não há correlação entre o erro e X, isso é importante para garantirmos o efeito ceteris paribus;

4. As variáveis X e Y não podem ter somente um único valor para toda a amostra;

Suponha que desejamos estimar um parâmetro populacional θ. Seja $\hat{\theta}$ o estimador de θ, $\hat{\theta}$ não seria enviesado se:

$$E(\hat{\theta}) = \theta \tag{2.1}$$

O MQO garante que dadas as inúmeras amostras diferentes possíveis a partir da população, os estimadores não são enviesados. Contanto que as quatro hipóteses passadas sejam respeitadas:

$$E(\hat{\beta}_0) = \beta_0 \tag{2.2}$$

$$E(\hat{\beta}_0) = \beta_1 \tag{2.3}$$

2.2 Confiabilidade e variância

Nossa principal hipótese é a da *homocedasticidade do erro*. Ela nos diz que:

$$Var(u|x) = \sigma^2 \tag{2.4}$$

2.2.1 Variância dos estimadores

Sob a hipótese de homocedasticidade, e tendo σ^2 como a variância populacional, as variâncias dos estimadores são:

$$Var(\hat{\beta_1}) = \frac{\sigma^2}{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2} \tag{2.5}$$

$$Var(\hat{\beta_0}) = \frac{\sigma^2}{\sum(x_i - \overline{x})^2} \frac{\sum\limits_{i=1}^{n} x_i^2}{n} \tag{2.6}$$

Algumas conclusões que podem ser retiradas a partir disto:

1. Quão maior a quantidade da amostra (n), a variância tenderá a ser menor. Isso se dá por haverem mais valores para serem somados nos denominadores;

2. Quanto maior a variância de X na amostra, menor a variação dos estimadores.

Como u é variável aleatória, sua variância é dada por:

$$Var(u|x) = E(u^2|x) - [E(u|x)]^2 \tag{2.7}$$

$$Var(u|x) = E(u^2) = \sigma^2 \tag{2.8}$$

Como é impossível obter os valores reais de σ^2, iremos calcular ele a partir da melhor informação que temos: os resíduos.

$$\hat{\sigma^2} = \frac{\sum\limits_{i=1}^{n} u_i^2}{n-2} \tag{2.9}$$

Sendo $(n-2)$ o número de *graus de liberdade* dos estimadores. O grau de liberdade representa o número de observações amostrais independentes. Este valor é obtido pelo número de observações n menos o número de restrições independentes (β_1, β_2). Portanto a regra que podemos estabelecer é a seguinte, sendo k o número de variáveis do modelo:

$$G^{o} liberdade = n - k \tag{2.10}$$

Com nossa estimativa de σ^2 podemos agora calcular as variâncias estimadas dos coeficientes:

$$\hat{Var}(\hat{\beta_1}) = \frac{\hat{\sigma^2}}{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2} \tag{2.11}$$

$$\hat{Var}(\hat{\beta_0}) = \frac{\hat{\sigma^2}}{\sum(x_i - \overline{x})^2} \frac{\sum\limits_{i=1}^{n} x_i^2}{n} \tag{2.12}$$

E também o *erro padrão* dos estimadores:

$$Ep(\beta_1) = \frac{\sigma^2}{\sqrt{\sum\limits_{i=1}^{n} x_i^2}} \tag{2.13}$$

$$Ep(\hat{\beta}_0) = \sqrt{\frac{\sum\limits_{i=1}^{n} x_i^2}{n\sum\limits_{i=1}^{n} x_i^2}}\sigma \tag{2.14}$$

Para entendermos o valor dos graus de liberdade, temos que lembrar o processo de minimização dos quadrados dos resíduos. Pelo *MQO*, buscamos minimizar os resíduos do nosso modelo:

$$min \sum_{i=1}^{n} (Y_i - \beta_0 + \beta_1 X + u) \tag{2.15}$$

Temos as seguintes condições de primeira ordem:

$$\sum_{i=1}^{n} (Y_i - -\beta_0 \beta_1 X_i) = 0 \tag{2.16}$$

$$\sum_{i=1}^{n} X_i (Y_i - -\beta_0 \beta_1 X_i) = 0 \tag{2.17}$$

$$\tag{2.18}$$

Se relembrarmos que: $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i = 0$. Então as equações 2.16 e 2.17 podem ser reescritas da seguinte forma:

$$\sum_{i=1}^{n} \hat{u}_i = 0 \tag{2.19}$$

$$\sum_{i=1}^{n} X_i \hat{u}_i = 0 \tag{2.20}$$

$$\tag{2.21}$$

Essas condições implicam duas *restrições* ao modelo, e por isso teremos $N - 2$ graus de liberdade. Os graus de liberdade garantem que a variância estimada é não enviesada.

2.3 Regressão sem intercepto

Supondo que o verdadeiro modelo é: $Y = \beta_1 X + u$ e que as hipóteses usuais se aplicam. Nossa função regressão populacional seria a seguinte:

$$E(Y|X) = \beta_1 X \tag{2.22}$$

Teríamos então uma reta com valor médio 0 quando $X = 0$. E se tivéssemos uma regressão sem β_1, teríamos então uma reta com somente a média de Y, visto que, na equação 1.3 se $\beta_1 = 0 \rightarrow \beta_0 = \overline{Y}$.

2.4 Inferência

Como temos várias amostras possíveis dada certa população consequentemente possuímos vários $\hat{\beta}$ possíveis dependendo da amostra utilizada para estimação. Então podemo abstrair que tais valores de $\hat{\beta}$ possuem um distribuição. Nossa nova hipótese especifica as distribuições do modelo:

$$u \ N(0, \sigma^2) \tag{2.23}$$

$$y \ N(\beta_0 + \beta_1 X, \sigma^2) \tag{2.24}$$

$$\hat{\beta} \ N(\beta, Var(\hat{\beta})) \tag{2.25}$$

(GUJARATI; PORTER, 2011)

Referências

GUJARATI, Damodar N; PORTER, Dawn C. **Econometria Básica-5**. [S.l.]: Amgh Editora, 2011.

WOOLDRIDGE, Jeffrey M. **Introductory econometrics: A modern approach**. [S.l.]: Nelson Education, 2016.