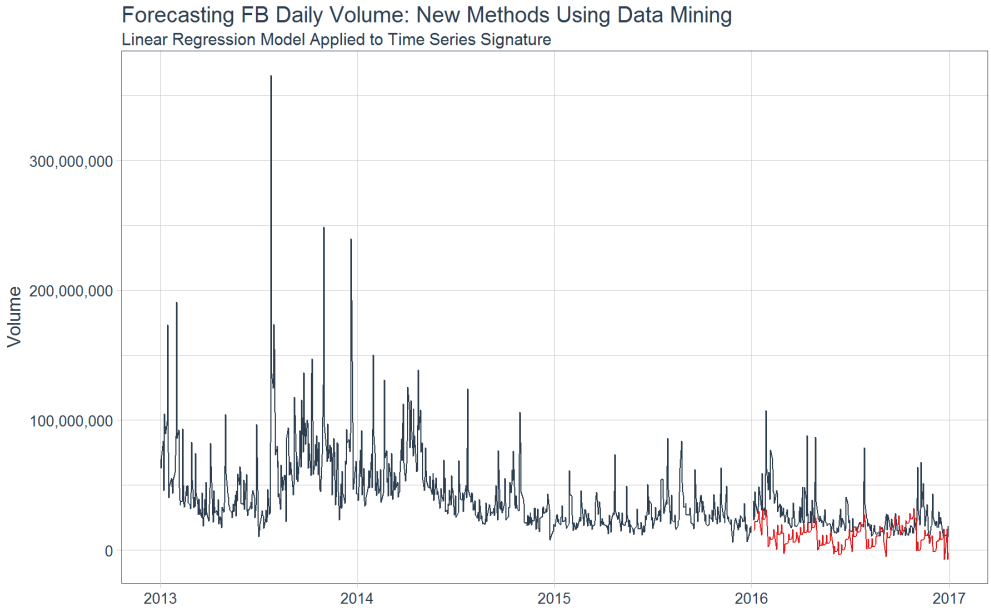
Econometrics III

(菅史彦. 九州大学 2019)

WEEK 1

1. Nonexperimental data

   observational data, or social science data.

2. Data type
   - cross-sectional data 横截面数据
     - sample of different individuals ,at a given point of time
     - often ignore minor time differences
     - observations are more or less independent
     - *applied microeconomics*
   - time series data 时间序列数据
     - one variable(or several variables) over time
     - 

     - serially correlated. trends and seasonality.
     - *applied macroeconomics and finance*
   - pooled cross sections 混合截面数据
     - two or more cross sections combined
     - evaluate policy changes
   - panel/longitudinal data 面板数据
     - same cross sections followed over time = *time series+cross sections*
     - can be used to account for time-invariant un-observables: eg. lags in decision making; city crime statistics ‘s un-observable city characteristics
3. r regression (SLR)

```r
#install Wooldridge package to use datasets
data("crime")
result <- lm(wage ~ educ+exper+training, data = crime)
summary(result)
```

4. definition of causal effect: ***ceteris paribus***

   how does *y* change if variable *x* is changed, but all other relevant factors are held constant?

5. topics
   - OLS
   - IV
   - Fixed Effect & Random Effect Model
   - MLE
   - Probit/Logit
   - Tobit
   - Heckman Two Step

WEEK 2

1. SLR --- *Simple Linear Regression*

   Focus: when is there a causal interpretion?

   Assumptions:

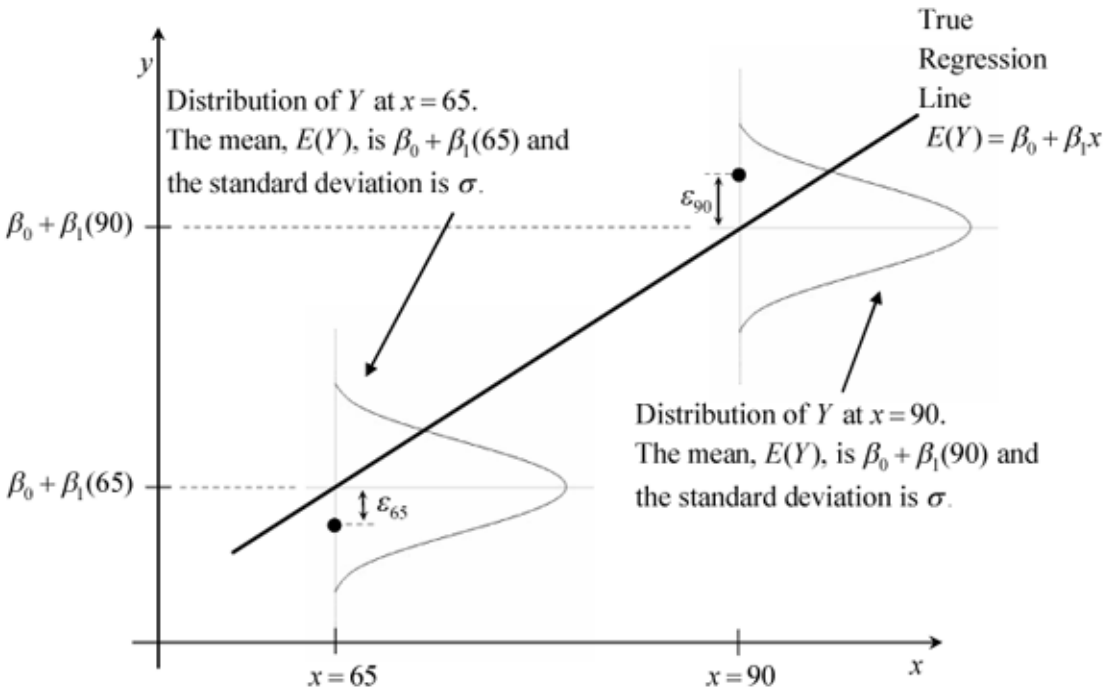
   1. Conditional Expectation independence

      $E(u|x) = 0$ (which is a function of x)

      x does not contain any information about mean of u.

      - PRF *Population Regression Function*

$$\begin{aligned} E(y|x) &= E(\beta_0 + \beta_1 x + u|x) \\ &= \beta_0 + \beta_1 x + E(u|x) \\ &= \beta_0 + \beta_1 x \end{aligned}$$



$< Empty\ Math\ Block >$

   - Sample analogue
     - regression residuals

$$\hat{u} = y^i - \hat{y}_i$$

minimize sum of squared residuals:

$$min\Sigma\hat{u_i}^2 \rightarrow \hat{\beta}_0, \hat{\beta}_1$$

- OLS estimates *Ordinary Least Squares Estimates*

  by minimizing SSR, we estimates β_0 and β_1.

- error terms

  fitted regression line - unknown population regression line.

  differ from residuals

2. R introduction

```r
data("gpa1")
#subset dataset gpa1, we only use 3 columns
gpa1small <- gpa1[,c("colGPA","hsGPA","ACT")]
summary(gpa1small)
#variation and correlation index(pearson)
var(gpa1small)
cor(gpa1small)
#standard deviation
sd(gpa1small$ACT)
#histogram
hist(gpa1small$colGPA)
#scatter plots
plot(gpa1small$colGPA,gpa1small$hsGPA)
#rgl package
install.package("rgl")
library("rgl")
plot3d(gpa1small$colGPA,gpa1small$hsGPA)# this function must
contain 3 vars.
#linear regression model
result <- lm(gpa1small$colGPA~gpa1small$hsGPA+gpa1small$ACT)
result <- lm(colGPA~hsGPA+ACT,data=gpa1small)
summary(result)
#obtain bata1 by coefficients index
beta1<-result$coefficients[2]
```