Simple Univariate Regression: Part 2

Taro Mieno

August 19, 2020

Deriving the Ordinary Least Squares (OLS) estimates

1. You have collected data with n observations on y and x
2. This random sample is denoted as $\{(y_i, x_i) : i = 1, \ldots, n\}$
3. For each i, we can write:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

The data set and model

Objective

Estimate the impact of lot size on house price

Model

$$price_i = \beta_0 + \beta_1 lotsize_i + u_i$$

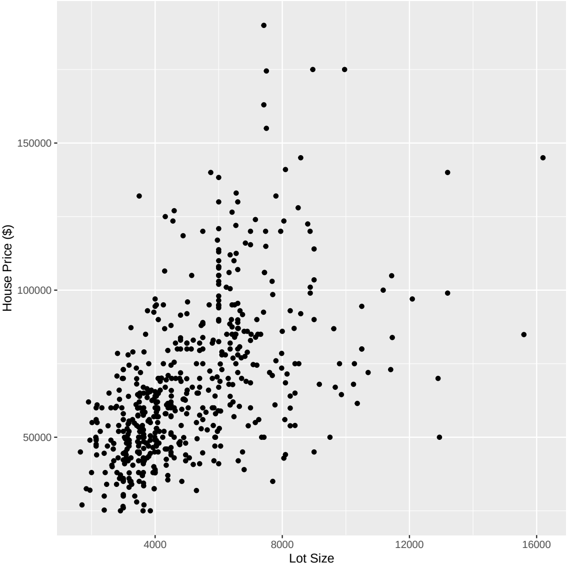
- $price_i$: house price (\$) of house i
- $lotsize_i$: lot size of house i
- u_i: error term (everything else) of house i

Data set we are going to use

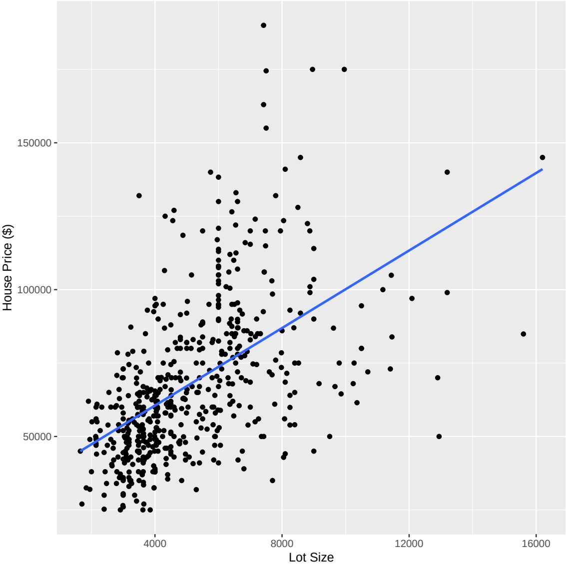
R code: Loading a data set

```
#--- load the AER package  ---#
library(AER) # load the AER package

#--- load the HousePrices data set ---#
data(HousePrices) # load

#--- take a look ---#
head(HousePrices[,1:5])

  price lotsize bedrooms bathrooms stories
1 42000    5850        3         1       2
2 38500    4000        2         1       1
3 49500    3060        3         1       1
4 60500    6650        3         1       2
5 61000    6360        2         1       1
6 66000    4160        3         1       1
```

Random sample and regression



Random sample and regression



- We want to draw a line like this, the slope of which is an estimate of β_1
- A way: Ordinary Least Squares (OLS)

Before talking about OLS

Residuals

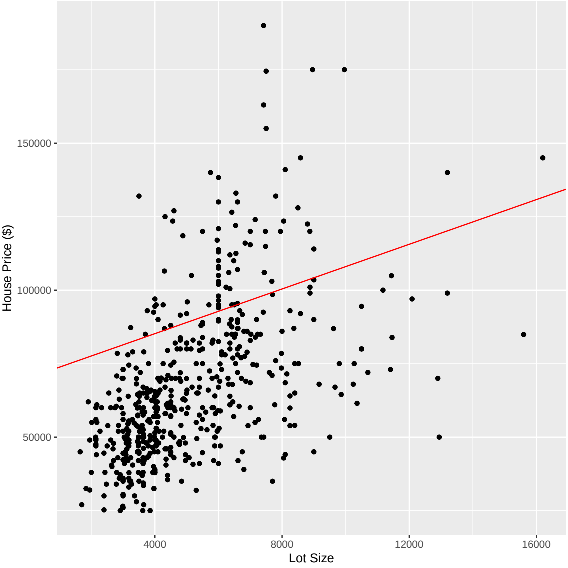
For particular values of $\hat{\beta}_0$ and $\hat{\beta}_1$ you pick, the modeled value of y for individual i is $\hat{\beta}_0 + \hat{\beta}_1 x_i$.

Then, the residual for individual i is:

$$\hat{u}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

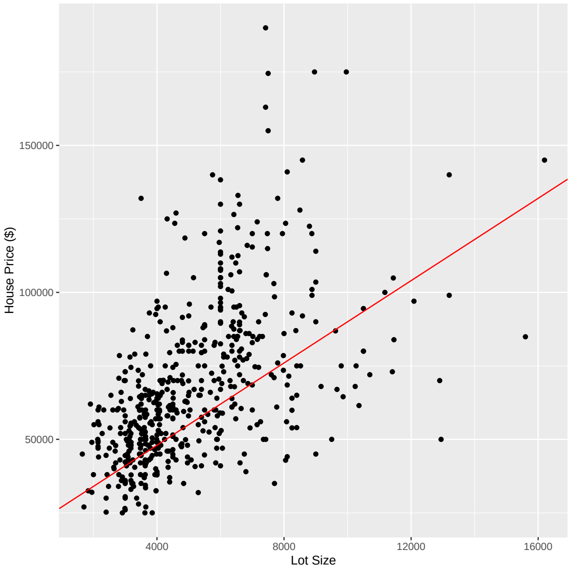
That is, residual is the observed value of the dependent variable less the value of modeled part. For different values of $\hat{\beta}_0$ and $\hat{\beta}_1$, you have a different value of residual.

Residuals



- $\hat{\beta}_0 = 70000$
- $\hat{\beta}_1 = 3.8$

Residuals



- $\hat{\beta}_0 = 20000$
- $\hat{\beta}_1 = 7$

OLS

Idea

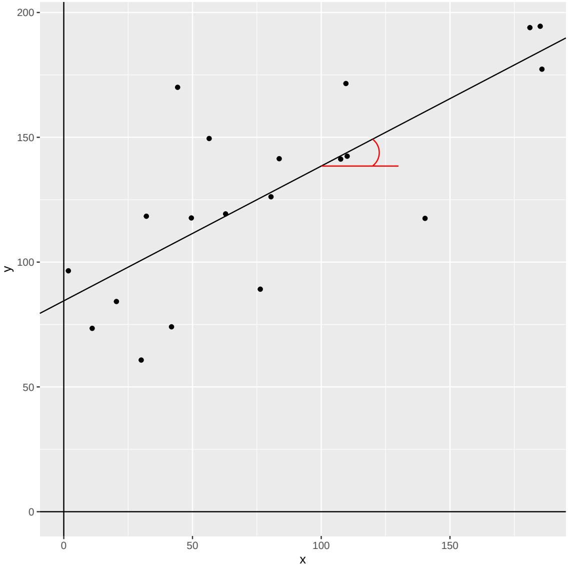
Let's find the value of β_0 and β_1 that minimizes the squared residuals!

Mathematically

$$Min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{n} \hat{u}_i^2,$$

where $\hat{u}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

Ordinary Least Squares (OLS): Visualization



OLS

- Why do we square the residuals, and then sum them up together? What's gonna happen if you just sum up residuals?
- How about taking the absolute value of residuals, and then sum them up?

Deriving OLS estimates

$$Min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2,$$

Steps

1. partial differentiation of the objective function with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$
2. solve for $\hat{\beta}_0$ and $\hat{\beta}_1$

OLS derivation: FOC

$$Min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2,$$

FOC

$$\frac{\partial}{\partial \hat{\beta}_0} = 2\sum_{i=1}^{n}[y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)] = 0$$

$$\frac{\partial}{\partial \hat{\beta}_1} = 2\sum_{i=1}^{n} x_i \cdot [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)] = 0$$

OLS derivation: Blank Sheet 1

OLS derivation: Blank Sheet 2

OLS

OLS estimators: analytical formula

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

OLS

OLS estimators: analytical formula

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Estimators

Specific rules to use available data once you get the data

OLS

OLS estimators: analytical formula

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Estimators

Specific rules to use available data once you get the data

Estimates

Numbers you get once you plug values (your data) into the formula

OLS demonstration in R

Model

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

R code: hard way

```
y <- HousePrices$price
x <- HousePrices$lotsize

#--- beta_1 ---#
b1_num <- sum((x-mean(x))*(y-mean(y)))
b1_denom <- sum((x-mean(x))^2)
b1 <- b1_num/b1_denom
b1

[1] 6.598768
```

OLS demonstration in R

Model

$$price = \beta_0 + \beta_1 lotsize + u$$

R code: easy way

```
#--- run OLS on the above model ---#
# lm(dep_var ~ indep_var,data=data_name)
uni_reg <- lm(price~lotsize,data=HousePrices)
uni_reg
```

```

Call:
lm(formula = price ~ lotsize, data = HousePrices)

Coefficients:
(Intercept)      lotsize
  34136.192        6.599
```

Sample Regression Function (SRF)

Once you have estimated β_0 and β_1, you can form

Sample Regression Function (SRF)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

House Price Example

$$price = 3.4136 \times 10^4 + 6.599 \times lotsize$$

Prediction

Sample Regression Function (SRF)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Prediction

If you plug a value into x in the SRF, you can get a prediction of $E[y|x]$ $(\hat{y})$ (called either fitted value of predicted value)

Prediction

R code: Prediction

```
#--- list the things inside the regression results ---#
ls(uni_reg)

 [1] "assign"        "call"          "coefficients"  "df.residual"
 [5] "effects"       "fitted.values" "model"         "qr"
 [9] "rank"          "residuals"     "terms"         "xlevels"

#--- access fitted values for sample points ---#
uni_reg$fitted.values[1:5]

       1        2        3        4        5
72738.98 60531.26 54328.42 78018.00 76104.35

#--- for values of lotsize that are not in the sample ---#
newdata <- data.frame(lotsize=c(3000,12000,15000))
predict(uni_reg,newdata=newdata)

        1         2         3
 53932.49 113321.40 133117.71
```

Exercise

The impact of lotsize

Your current lot size is 3000. You are thinking of expanding your lot by 1000 (with everything else fixed), which would cost you $5,000$\$. Should you do it? Use R to figure it out.

Exercise

R code: impact of lotsize

```
#--- list the things inside the regression results ---#
ls(uni_reg)

 [1] "assign"        "call"          "coefficients"  "df.residual"
 [5] "effects"       "fitted.values" "model"         "qr"
 [9] "rank"          "residuals"     "terms"         "xlevels"

#--- access the coefficient values  ---#
uni_reg$coefficients

 (Intercept)      lotsize
34136.191565     6.598768

# class(uni_reg)

#--- assess the impact ---#
```

R^2: Goodness of fit

R^2

A measure of how good your model is in predicting the dependent variable (explaining variations in the dependent variable) compared to just using the average of the dependent variable as the predictor

Goodness of fit

You can decompose observed value of y into two parts: fitted value and residual

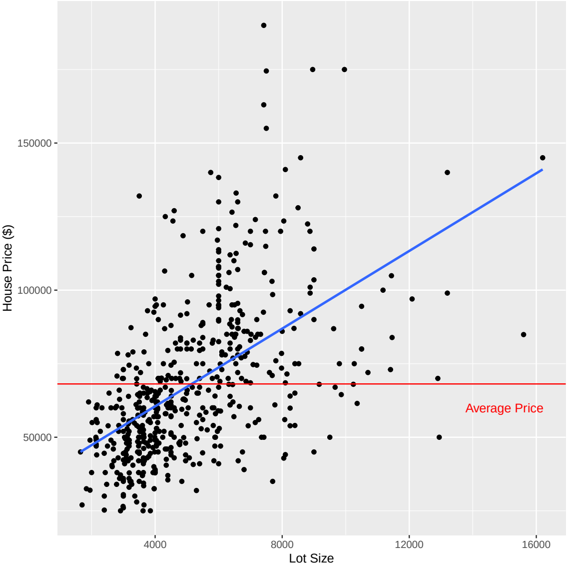
$$y_i = \hat{y}_i(\hat{\beta}_0 + \hat{\beta}_1 x_i) + \hat{u}_i$$

Now, subtracting $\bar{y}$ (sample average of y),

$$y_i - \bar{y} = \hat{y}_i - \bar{y} + \hat{u}_i$$

- $y_i - \bar{y}$: how far away is the actual value of y for ith observation from the sample average $\bar{y}$? (actual deviation from the mean)
- $\hat{y}_i - \bar{y}$: how far away is the predicted value of y for ith observation from the sample average $\bar{y}$? (explained deviation from the mean)
- $\hat{u}_i$: unexplained part

Visualization



- $y_i - \bar{y}$
- $\hat{y}_i - \bar{y}$
- $\hat{u}_i$

R^2: Goodness of fit

total sum of squares (SST)

$$SST \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2$$ Sample variation in y

explained sum of squares (SSE)

$$SSE \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

residual sum of squares (SSR)

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

Goodness of fit

R^2

$$R^2 = SSE/SST = 1 - SSR/SST$$

(the value of R^2 always lies between 0 and 1 as long as an intercept is included in the econometric model)

What does R^2 measure

R^2 is a measure of how much improvement you've made by including independent variable(s) ($y = \beta_0 + \beta_1 x + u$) compared to when simply using the mean of dependent variable as the predictor ($y = \beta_0 + u$)

R^2

Important notes about R^2

- R^2 is of no value if you are interested in finding the causal (ceteris paribus) impact of a variable of interest (More on this later when we discuss bias)
- R^2 is important if your interest lies in predicting y

Small sample property of OLS estimators

Small sample property of OLS estimators

What is an estimator?

- A function of data that produces an estimate (actual number) of a parameter of interest once you plug in actual values of data
- OLS estimators: $\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Small sample property of OLS estimators

What is small sample property?

Properties that hold whatever the size of observation (small or large) prior to obtaining actual estimates (before getting data)

- Put more simply: what can you expect from the estimators before you actually get data and obtain estimates?
- Difference between small sample property and the algebraic properties we looked at earlier?

Small sample property of OLS estimators

Desirable Properties

OLS is only a way of using available information to obtain estimates. Does it have desirable properties?

- Unbiasedness
- Efficiency

As it turns out, OLS is a very good way of using available information!!

Unbiasedness of OLS estimators

What does unbiased mean?

Example

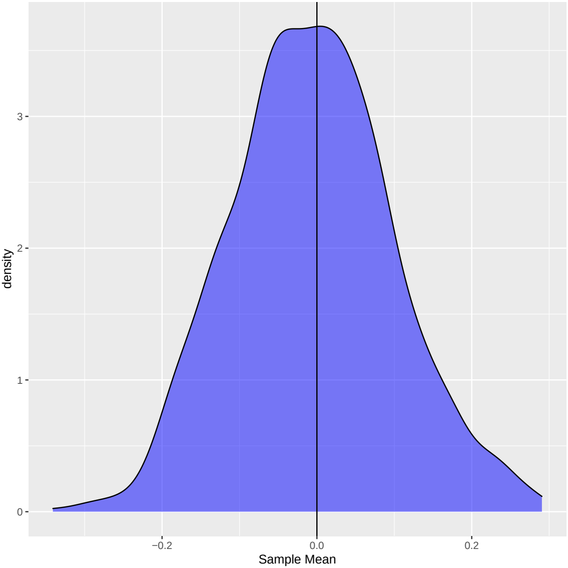
- Consider a problem of estimating the expected value of a single variable, x
- A good estimator is sample mean: $\dfrac{1}{n}\sum_{i}^{n} x_i$

Unbiasedness

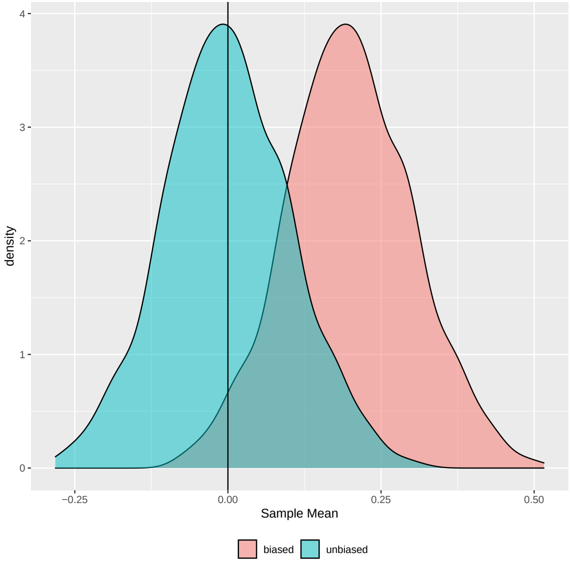
R code: Sample Mean

```
#--- set the number of observations ---#
n <- 100

#--- generate random values ---#
x_seq <- rnorm(n) # Normal(mean=0,sd=1)

#--- calcualte the mean ---#
mean(x_seq)

[1] -0.09058714
```

Unbiasedness: Visualization



Biasedness: Visualization



Unbiasedness of OLS estimators

Unbiasedness of OLS estimators

Under certain conditions, OLS estimators are unbiased. That is,

$$E[\hat{\beta}_1] = E\Big[\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\Big] = \beta_1$$

(We do not talk about unbiasedness of $\hat{\beta}_0$ because we are almost never interested in the intercept. Given the limited time we have, it is not worthwhile talking about it)

Certain Conditions

Condition $SLR.1$: Linear in Parameters

In the population model, the dependent variable, y, is related to the independent variable, x, and the error (or disturbance), u, as

$$y = \beta_0 + \beta_1 x + u$$

Wooldridge, Jeffrey M. (2012-09-26). Introductory Econometrics: A Modern Approach (Upper Level Economics Titles) (Page 45). Cengage Textbook. Kindle Edition.

Certain Conditions

Condition $SLR.2$: Random sampling

We have a random sample of size n, $(x_i, y_i) : i = 1, 2, \ldots, n$, following the population model in equation (1).

(the consequence of the violation of this condition is discussed later)

Non-random sampling

- Example: You observe income-education data only for those who have income higher than $\$25K$
- Benevolent and malevolent kinds:
  - exogenous sampling
  - endogenous sampling

Wooldridge, Jeffrey M. (2012-09-26). Introductory Econometrics: A Modern Approach (Upper Level Economics Titles) (Page 45). Cengage Textbook. Kindle Edition.

Certain Conditions

Condition $SLR.3$: Sample variation in covariates

The sample outcomes on x, namely, $x_i, i = 1, \ldots, n$, are not all the same value.

Wooldridge, Jeffrey M. (2012-09-26). Introductory Econometrics: A Modern Approach (Upper Level Economics Titles) (Page 46). Cengage Textbook. Kindle Edition.

Certain Conditions

Condition $SLR.4$: Zero conditional mean

The error u has an expected value of zero given any value of the explanatory variable. In other words,

$$E[u|x] = 0$$

Along with random sampling condition, this implies that

$$E[u_i|x_i] = 0$$

Wooldridge, Jeffrey M. (2012-09-26). Introductory Econometrics: A Modern Approach (Upper Level Economics Titles) (Page 47). Cengage Textbook. Kindle Edition.

Correlation and Mean Independence

Note

Mean independence of u and x implies no correlation. But, correlation does not imply mean independence.

Mean Independence Implies Correlation (Proof)

$$
\begin{aligned}
Cov(u,x) =& E[(u-E[u])(x-E[x])] \\
=& E[ux]-E[u]E[x]-E[u]E[x]+E[u]E[x] \\
=& E[ux] \\
=& E_x[E_u[u|x]] \text{ (iterated law of expectation)}
\end{aligned}
$$

If zero conditional mean condition ($E(u|x)=0$) is satisfied,

$$Cov(u,x) = E_x[0]=0$$

Good and bad empiricists

Good Empiricists

- have ability to judge if the above conditions are satisfied for the particular context you are working on
- have ability to correct (if possible) for the problems associated with the violations of any of the above conditions
- knows the context well so you can make appropriate judgments

Unbiasedness of OLS estimators

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \\
&= \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad \left[\text{because } \sum_{i=1}^{n}(x_i-\bar{x})\bar{y}=0\right] \\
&= \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{SST_x} \quad \left[\text{where, } SST_x=\sum_{i=1}^{n}(x_i-\bar{x})^2\right] \\
&= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(\beta_0+\beta_1 x_i+u_i)}{SST_x} \\
&= \frac{\sum_{i=1}^{n}(x_i-\bar{x})\beta_0+\sum_{i=1}^{n}\beta_1(x_i-\bar{x})x_i+\sum_{i=1}^{n}(x_i-\bar{x})u_i}{SST_x}
\end{aligned}$$

Unbiasedness of OLS estimators

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\beta_0 + \beta_1 \sum_{i=1}^{n}(x_i - \bar{x})x_i + \sum_{i=1}^{n}(x_i - \bar{x})u_i}{SST_x}$$

Since $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$ and

$$\sum_{i=1}^{n}(x_i - \bar{x})x_i = \sum_{i=1}^{n}(x_i - \bar{x})^2 = SST_x,$$

$$\hat{\beta}_1 = \frac{\beta_1 SST_x + \sum_{i=1}^{n}(x_i - \bar{x})u_i}{SST_x} = \quad \beta_1 + (1/SST_x)\sum_{i=1}^{n}(x_i - \bar{x})u_i$$

Unbiasedness of OLS estimators

$$\hat{\beta}_1 = \beta_1 + (1/SST_x)\sum_{i=1}^{n}(x_i - \bar{x})u_i$$

Taking, expectation of $\hat{\beta}_1$ conditional on $\mathbf{x} = \{x_1, \ldots, x_n\}$,

$$\begin{aligned} \Rightarrow E[\hat{\beta}_1|\mathbf{x}] =& E[\beta_1|\mathbf{x}] + E[(1/SST_x)\sum_{i=1}^{n}(x_i - \bar{x})u_i|\mathbf{x}] \\ =& \beta_1 + (1/SST_x)\sum_{i=1}^{n}(x_i - \bar{x})E[u_i|\mathbf{x}] \end{aligned}$$

So, if condition 4 ($E[u_i|\mathbf{x}] = 0$) is satisfied,

$$\begin{aligned} E[\hat{\beta}_1|\mathbf{x}] =& \beta_1 \\ E_{\mathbf{x}}[\hat{\beta}_1|\mathbf{x}] =& E[\hat{\beta}_1] = \beta_1 \end{aligned}$$

Unbiasedness of OLS estimators

Reconsider the following example

$$price = \beta_0 + \beta_1 \times lotsize + u$$

- $price$: house price ($)
- $lotsize$: lot size
- u: error term (everything else)

Questions

- What's in u?
- Do you think $E[u|x]$ is satisfied?

Unbiasedness of OLS estimators

Important notes (again)

- Unbiasedness property of OLS estimators says NOTHING about the estimate that we obtain for a given sample
- it is always possible that we could obtain an unlucky sample that would give us a point estimate far from β_1, and we can never know for sure whether this is the case.

Variance of OLS estimators

Variance of OLS estimators

Important notes

- OLS estimators are random variables, which means that they have distributions
- OLS estimators have variance (how spread out OLS estimates can be)

Variance

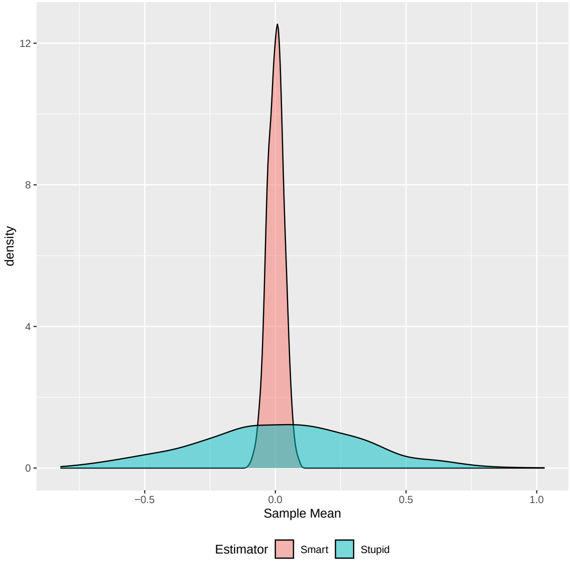
Example:

Consider two estimators of $E[x]$:

$$\theta_{smart} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (n = 1000)$$

$$\theta_{stupid} = \frac{1}{10}\sum_{i=1}^{10} x_i$$

Variance of estimators



Variance of OLS estimators

Variance of OLS estimators

If $Var(u|x) = \sigma^2$ and the four conditions (we used to prove unbiasedness of OLS estimators) are satisfied,

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sigma^2}{SST_x}$$

$Var(u|x) = \sigma^2$

Homoskedasticity

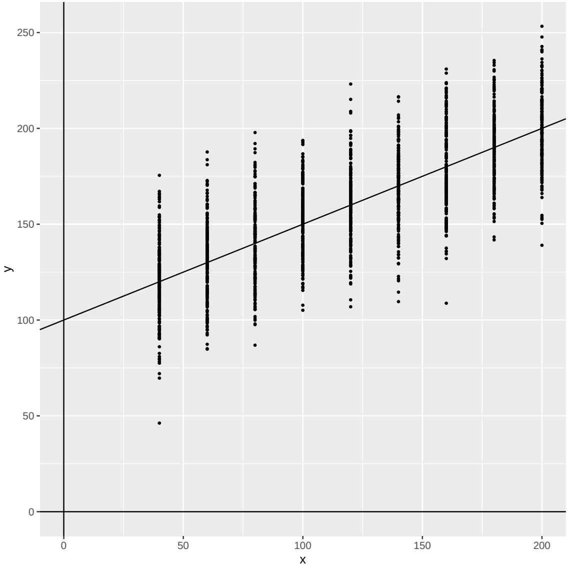
The error u has the same variance give any value of the covariate x

Heterokedasticity

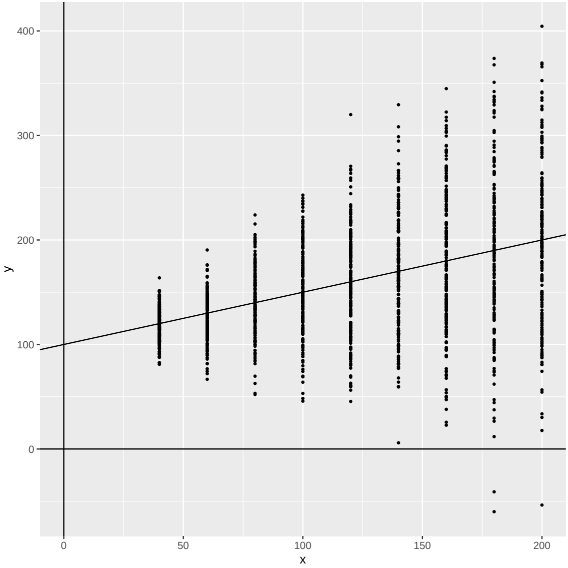
The variance of the error u differs depending on the value of x

$$Var(u|x) = f(x) \tag{1}$$

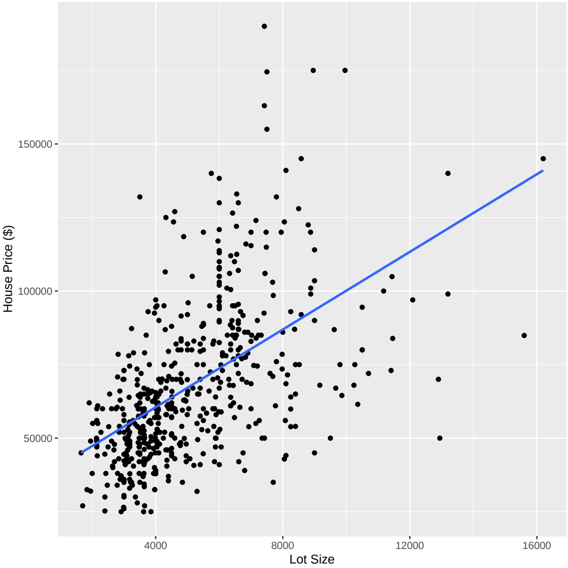
Homoskedastic Error



Heteroskedastic Error



House Price Example



Homoskedasticity Condition (Assumption)

Important notes

- We did NOT use this condition to prove that OLS estimators are unbiased
- In most applications, homoskedasticity condition is not satisfied, which has important implications on:
  - estimation of variance (standard error) of OLS estimators
  - significance test

(A lot more on this issue later)

Derivation of the variance of OLS estimators

$$\hat{\beta}_1 = \beta_1 + (1/SST_x)\sum_{i=1}^{n} d_i u_i$$

Because $Var(\alpha x) = \alpha^2 var(x)$,

$$Var(\hat{\beta}_1|x) = (1/SST_x)^2 Var(\sum_{i=1}^{n} d_i u_i|x)$$

Because of independent random sampling Math Aside,

$$Var(\hat{\beta}_1|x) = (1/SST_x)^2 \sum_{i=1}^{n} d_i^2 Var(u_i|x)$$

If the homoskedasticity assumption is satisfied,

$$\begin{aligned} Var(\hat{\beta}_1|x) &= (1/SST_x)^2 \sum_{i=1}^{n} d_i^2 \sigma^2 \\ &= \sigma^2/SST_x \end{aligned}$$

Math Aside

Go Back

$$
\begin{aligned}
Var(\alpha x) =& \alpha^2 Var(x) \\
Var(x+y) =& Var(x) + 2Cov(x,y) + Var(y)
\end{aligned}
$$

If x and y are independent, $Cov(x, y) = 0$.

Variance of OLS estimators

Variance of the OLS estimators

$$Var(\hat{\beta}_1|x) = \sigma^2/SST_x$$

Variance of OLS estimators

Variance of the OLS estimators

$$Var(\hat{\beta}_1|x) = \sigma^2/SST_x$$

What can you learn from this equation?

- the variance of OLS estimators is smaller (larger) if the variance of error term is smaller (larger)
- the greater (smaller) the variation in the covariate x, the smaller (larger) the variance of OLS estimators
  - if you are running experiments, spread the value of x as much as possible
  - you will rarely have this luxury

Gauss-Markov Theorem:

Gauss-Markov Theorem

Under conditions $SLR.1$ through $SLR.5$, OLS estimators are the best linear unbiased estimators (BLUEs)

Meaning,

No other unbiased linear estimators have smaller variance than the OLS estimators (desirable efficiency property of OLS)

Estimating the error variance

- $Var(\hat{\beta}_1|x) = \sigma^2/SST_x$ will never be known. But, you can estimate it.
- Once you estimate $Var(\hat{\beta}_1|x)$, you can test the statistical significance of $\hat{\beta}_1$ (More on this later)

Estimating the error variance

$$Var(u_i) = \sigma^2 = E[u_i^2]$$
$$\Big(Var(u_i) \equiv E[u_i^2] - E[u_i]^2\Big)$$

- So, $\frac{1}{n}\sum_{i=1}^{n} u_i^2$ is an unbiased estimator of $Var(u_i)$
- What is the problem with this estimator?

Estimating the error variance

We don't observe u_i (error), but we observe residuals ($\hat{u}_i$)

Error and Residual

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i$$

Residuals as unbiased estimators of error

$$\begin{aligned}
\hat{u}_i =& y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \\
\hat{u}_i =& \beta_0 + \beta_1 x_i + u_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \\
\Rightarrow \hat{u}_i - u_i =& (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1) x_i \\
\Rightarrow E[\hat{u}_i - u_i] =& E[(\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1) x_i] = 0
\end{aligned}$$

Estimating the error variance

- We know $E[\hat{u}_i - u_i] = 0$
- so, why don't we use $\hat{u}_i$ (observable) in place of u_i (unobservable)
- $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \hat{u}_i^2$ as an estimator of σ^2?
- Unfortunately, $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \hat{u}_i^2$ is a biased estimator of σ^2

Estimating the error variance

Algebraic property of OLS

$$\sum_{i=1}^{n} \hat{u}_i = 0 \quad \text{and} \quad \sum_{i=1}^{n} x_i \hat{u}_i = 0$$

Estimating the error variance

Algebraic property of OLS

$$\sum_{i=1}^{n} \hat{u}_i = 0 \text{ and } \sum_{i=1}^{n} x_i \hat{u}_i = 0$$

- this means that once you know the value of $n-2$ residuals, you can find the value of the other two by solving the above equations
- so, it's almost as if you have $n-2$ value of residuals instead of n

Estimating the error variance

Algebraic property of OLS

$$\sum_{i=1}^{n} \hat{u}_i = 0 \text{ and } \sum_{i=1}^{n} x_i \hat{u}_i = 0$$

- this means that once you know the value of $n-2$ residuals, you can find the value of the other two by solving the above equations
- so, it's almost as if you have $n-2$ value of residuals instead of n

Unbiased estimator of σ^2

We use $\hat{\sigma}^2 = \dfrac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2$, which satisfies $E[\dfrac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2] = \sigma^2$

The use of $\hat{\sigma}^2$

Since $sd(\hat{\beta}_1) = \sigma/\sqrt{SST_x}$, the natural estimator of $sd(\hat{\beta}_1)$ is

$$se(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2}/\sqrt{SST_x},$$

which is called standard error of $\hat{\beta}_1$. Later, we use $\hat{\beta}_1$ for testing.

Standard Error Estimation

R code: Standard Error

```
summary(uni_reg)



Call:
lm(formula = price ~ lotsize, data = HousePrices)

Residuals:
   Min     1Q Median     3Q    Max 
-69551 -14626  -2858   9752 106901 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 3.414e+04  2.491e+03    13.7   <2e-16 ***
lotsize     6.599e+00  4.458e-01    14.8   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22570 on 544 degrees of freedom
Multiple R-squared:  0.2871,Adjusted R-squared:  0.2858 
F-statistic: 219.1 on 1 and 544 DF,  p-value: < 2.2e-16
```

Functional Form

Functional Form

Functional forms

- transformation of variables is allowed without disturbing our analytical framework as long as the model is linear in parameter.
- transformation of variables change the interpretation of the coefficients estimates

Golas

- present popular functional forms
- use simple calculus to examine the interpretation of the coefficients

Popular Functional Forms

log-linear

$$log(y_i) = \beta_0 + \beta_1 x_i + u_i$$

linear-log

$$y_i = \beta_0 + \beta_1 log(x_i) + u_i$$

log-log

$$log(y_i) = \beta_0 + \beta_1 log(x_i) + u_i$$

Functional form: Log-linear

Model

$$log(y_i) = \beta_0 + \beta_1 x_i + u_i$$

Calculus

Differentiating the both sides wrt x_i,

$$\frac{1}{y_i} \cdot \frac{\partial y_i}{\partial x_i} = \beta_1 \Rightarrow \frac{\Delta y_i}{y_i} = \beta_1 \Delta x_i$$

Interpretation

β_1 measures a percentage change in y_i when x_i is increased by one unit

Functional Form

It is not hard to write the model in a way that y and x are non-linearly related.

How about this model?

$$log(wage) = \beta_0 + \beta_1 educ + u$$

Interpretation of β_1? Differentiating both sides with respect to $educ$,

$$\frac{1}{wage}\frac{\partial wage}{\partial educ} = \beta_1 \Rightarrow \frac{\Delta wage}{wage} = \beta_1 \Delta educ$$

Interpretation

If education increases by 1 year ($\Delta educ = 1$), then wage increases by $\beta_1 * 100\%$ ($\frac{\Delta wage}{wage} = \beta_1$)

Log-linear model

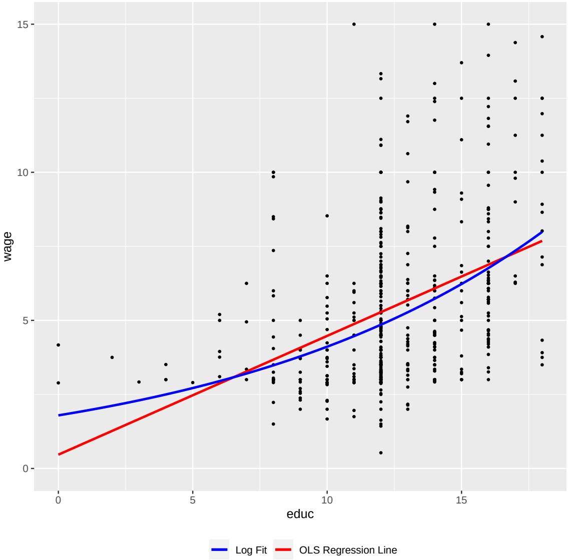
If you estimate $log(wage) = \beta_0 + \beta_1 educ + u$,

$$\widehat{log(wage)} = 0.584 + 0.083educ$$

With $u = 0$,

$$\widehat{wage} = e^{0.584+0.083educ}$$

Visualization



Functional form: Linear-log

Model

$$y_i = \beta_0 + \beta_1 log(x_i) + u_i$$

Calculus

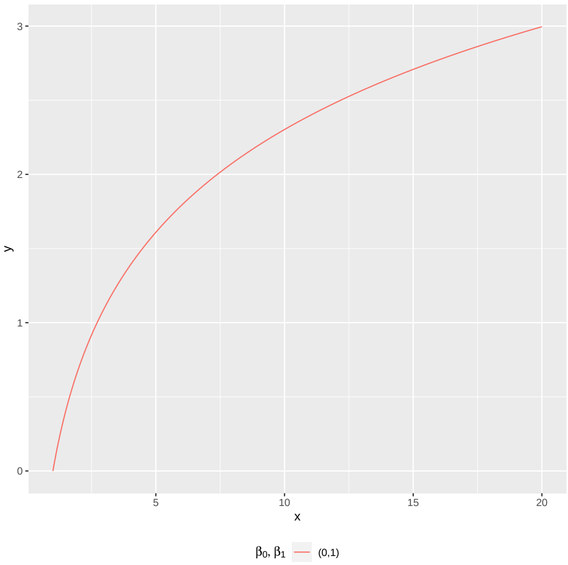
Differentiating the both sides wrt x_i,

$$\frac{\partial y_i}{\partial x_i} = \beta_1 / x_i \Rightarrow \Delta y_i = \beta_1 \frac{\Delta x_i}{x_i}$$

Interpretation

When x increases by 1%, y increases by β_1

Visualization



Functional form: Log-log

model

$$log(y_i) = \beta_0 + \beta_1 log(x_i) + u_i$$

Calculus

Differentiating the both sides wrt x_i,

$$\frac{\partial y_i}{y_i} / \frac{\partial x_i}{x_i} = \beta_1 \Rightarrow \frac{\Delta y_i}{y_i} = \beta_1 \frac{\Delta x_i}{x_i}$$

Interpretation

A percentage change in x would result in a β_1 percentage change in y_i (constant elasticity)

Simple Linear Regression

- In these models, the dependent variable and independent variable are non-linearly related, how come are these models called simple linear model?
- "linear" in simple linear model means that the model is linear in parameter, but not in variable

Non-linear (in parameter) Models

Examples: non-linear models

$$y_i = \beta_0 + x_i^{\beta_1} + u_i$$
$$y_i = \frac{x_i}{\beta_0 + \beta_1 x_i} + u_i$$

Notes

Transformation of the dependent and independent variables would not affect the OLS results we have seen as long as the model is linear in parameter

Appendix

R code to generate the unbiasedness figure go back

```
#--- set seed ---#
set.seed(244598)

#--- set the number of observations ---#
n <- 100
B <- 1000
mean_st <- rep(0,B)
for (i in 1:B){
  #--- generate random values ---#
  x_seq <- rnorm(n) # Normal(mean=0,sd=1)

  #--- calcualte the mean ---#
  mean_st[i] <- mean(x_seq)
}
plot_data <- data.table(mean_st)
g_unsbiased <- ggplot(data=plot_data) +
  geom_density(aes(x=mean_st),fill='blue',alpha=0.5) +
  xlab('Sample Mean')
```

R code to generate the variance figure

go back

```
set.seed(38495)
B <- 1000
theta_smart <- rep(0,B)
theta_stupid <- rep(0,B)
for (i in 1:B){
  x <- rnorm(1000)
  theta_smart[i] <- mean(x)
  theta_stupid[i] <- mean(x[1:10])
}

smart_data <- data.table(value=theta_smart,type='Smart')
stupid_data <- data.table(value=theta_stupid,type='Stupid')
plot_data <- rbind(smart_data,stupid_data)

g_var <- ggplot(data=plot_data) +
  geom_density(aes(x=value,fill=type),alpha=0.5,size=0.5) +
  scale_fill_discrete(name='Estimator') +
  xlab('Sample Mean') +
  theme(
    legend.position='bottom'
    )
ggsave('./variance.pdf',width=4,height=2.8)
```