BS2280 - Econometrics 1

Lecture 3 - Part 1: Properties of OLS

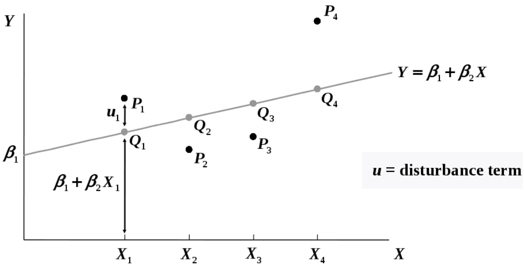
by Dr Yichen Zhu

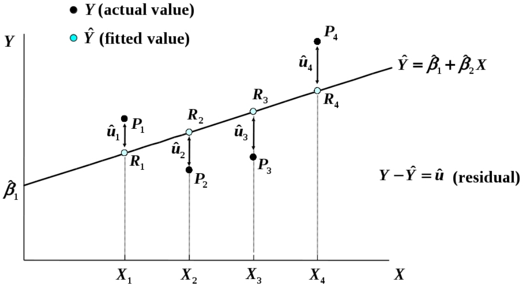
Outline

Intended Learning Outcomes

- Evaluating assumptions of OLS regressions
- Understanding sampling distributions

Review: Simple Linear Regression Model - Population vs. Sample





Population	**Sample**
$Y_i = \beta_1 + \beta_2 X_i + u_i$	$Y_i = \hat{Y}_i + \hat{u}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$
parameters β_1 and β_2	coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$
u_i disturbance term	$\hat{u}_i$ residual

Assumptions behind OLS regressions

- OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ have certain desirable properties which make them highly attractive amongst all estimators
- It is possible to calculate $\hat{\beta}_1$ and $\hat{\beta}_2$ using other methods, but OLS is favoured because of the properties of its estimators
- But these properties (to be discussed later) rely on a set of assumptions we need to make!!!
- Important to understand these assumptions - part of econometrics topics taught later revolves around consequences of these assumptions not holding.

Assumptions behind OLS regressions

- OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ have certain desirable properties which make them highly attractive amongst all estimators
- It is possible to calculate $\hat{\beta}_1$ and $\hat{\beta}_2$ using other methods, but OLS is favoured because of the properties of its estimators
- But these properties (to be discussed later) rely on a set of assumptions we need to make!!!
- Important to understand these assumptions - part of econometrics topics taught later revolves around consequences of these assumptions not holding.

Assumptions behind OLS regressions

- OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ have certain desirable properties which make them highly attractive amongst all estimators
- It is possible to calculate $\hat{\beta}_1$ and $\hat{\beta}_2$ using other methods, but OLS is favoured because of the properties of its estimators
- But these properties (to be discussed later) rely on a set of assumptions we need to make!!!
- Important to understand these assumptions - part of econometrics topics taught later revolves around consequences of these assumptions not holding.

Assumptions behind OLS regressions

- OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ have certain desirable properties which make them highly attractive amongst all estimators
- It is possible to calculate $\hat{\beta}_1$ and $\hat{\beta}_2$ using other methods, but OLS is favoured because of the properties of its estimators
- But these properties (to be discussed later) rely on a set of assumptions we need to make!!!
- Important to understand these assumptions - part of econometrics topics taught later revolves around consequences of these assumptions not holding.

Assumption 1

Model is linear in parameters and correctly specified

- Put simply, this means that the model is always written in a way that β_2 is simply multiplied by X, i.e. $\beta_2 X$

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- An example of a model that is not linear in parameters.
There are other ways in which β_1 and β_2 can interact making the model non-linear in parameters, e.g.:

$$Y_i = \beta_1 X_i^{\beta_2} + u_i$$

Assumption 1

Model is linear in parameters and correctly specified

- Put simply, this means that the model is always written in a way that β_2 is simply multiplied by X, i.e. $\beta_2 X$

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

- An example of a model that is not linear in parameters.
  There are other ways in which β_1 and β_2 can interact making the model non-linear in parameters, e.g.:

$$Y_i = \beta_1 X_i^{\beta_2} + u_i$$

Assumption 2

There is some variation in the X variable

- if X is constant in the sample, it cannot account for any of the variation in Y
- The values of X must be different across different observations. Cannot be constant, otherwise cannot run a regression and cannot get estimates $\hat{\beta}_1$ and $\hat{\beta}_2$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

- If X does not vary, $X_i = \bar{X}$ above, and cannot calculate $\hat{\beta}_1$ and $\hat{\beta}_2$

Assumption 2

There is some variation in the X variable

- if X is constant in the sample, it cannot account for any of the variation in Y
- The values of X must be different across different observations. Cannot be constant, otherwise cannot run a regression and cannot get estimates $\hat{\beta}_1$ and $\hat{\beta}_2$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

- If X does not vary, $X_i = \bar{X}$ above, and cannot calculate $\hat{\beta}_1$ and $\hat{\beta}_2$

Assumption 2

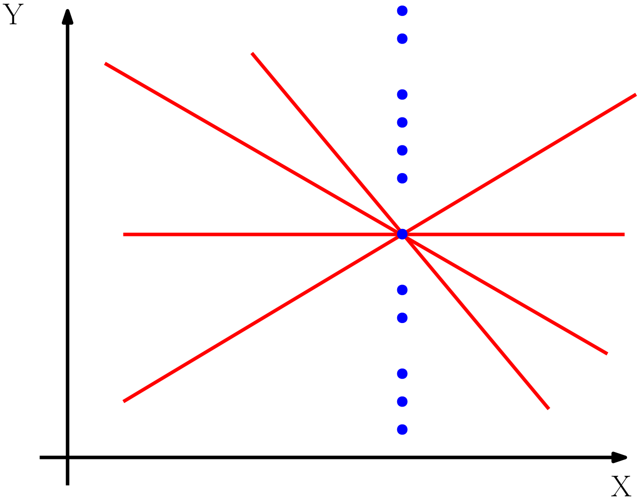
There is some variation in the X variable

- if X is constant in the sample, it cannot account for any of the variation in Y
- The values of X must be different across different observations. Cannot be constant, otherwise cannot run a regression and cannot get estimates $\hat{\beta}_1$ and $\hat{\beta}_2$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

- If X does not vary, $X_i = \bar{X}$ above, and cannot calculate $\hat{\beta}_1$ and $\hat{\beta}_2$

Assumption 2



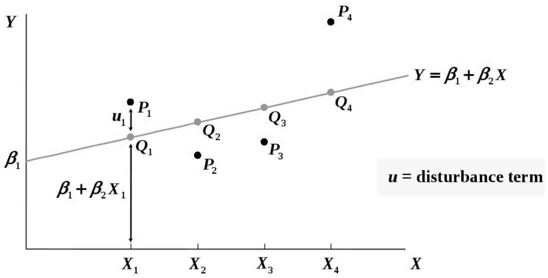
Assumption 3

Disturbance term has zero expectation

- We assume that the expected value of the disturbance term in any observation should be zero.

$$E(u_i) = \mu_u = 0 \text{ for all } i$$

- The disturbance term will sometimes be positive, sometimes negative, but on average will be zero.



Assumption 3

Disturbance term has zero expectation

- We assume that the expected value of the disturbance term in any observation should be zero.

$$E(u_i) = 0 \text{ for all } i$$

- There will be no systematic tendency for some error terms in some observations to be more positive/negative than others
- The above assumption is strictly written as

$$E(u_i | X_1, X_2, ..., X_n) = 0 \text{ for all } i$$

- Dougherty (p.116) shows that with an intercept term β_1 in our regression model, this condition is automatically satisfied

Assumption 4

The disturbance term is homoscedastic

- We assume that the error term has a constant variance, i.e. its values are drawn from a distribution with constant population variance
- Once sample is generated and OLS regression is run, error term will be greater in some observations, and lower in others
- But it is not expected to be more erratic in some observations than in others

$$E((u_i - \mu_u)^2) = \sigma_u^2 \text{ for all } i$$

since $E(u_i) = \mu_u = 0$

$$E(u_i^2) = \sigma_u^2 \text{ for all } i$$

Assumption 4

The disturbance term is homoscedastic

- We assume that the error term has a constant variance, i.e. its values are drawn from a distribution with constant population variance
- Once sample is generated and OLS regression is run, error term will be greater in some observations, and lower in others
- But it is not expected to be more erratic in some observations than in others

$$E((u_i - \mu_u)^2) = \sigma_u^2 \text{ for all } i$$

since $E(u_i) = \mu_u = 0$

$$E(u_i^2) = \sigma_u^2 \text{ for all } i$$

Assumption 4

The disturbance term is homoscedastic

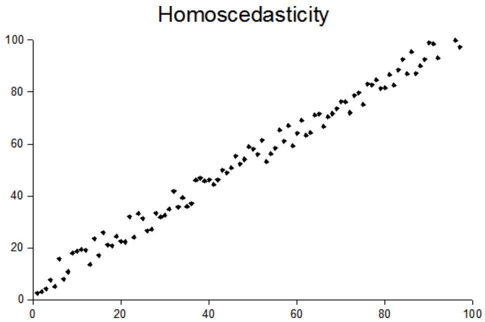
- We assume that the error term has a constant variance, i.e. its values are drawn from a distribution with constant population variance
- Once sample is generated and OLS regression is run, error term will be greater in some observations, and lower in others
- But it is not expected to be more erratic in some observations than in others

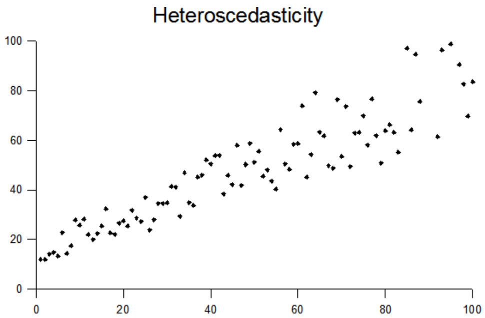
$$E((u_i - \mu_u)^2) = \sigma_u^2 \text{ for all } i$$

since $E(u_i) = \mu_u = 0$

$$E(u_i^2) = \sigma_u^2 \text{ for all } i$$

Homoscedasticity vs Heteroscedasticity

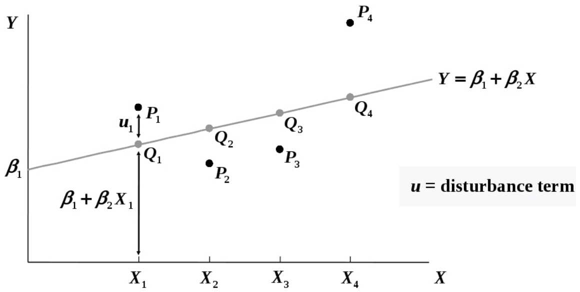




Assumption 5

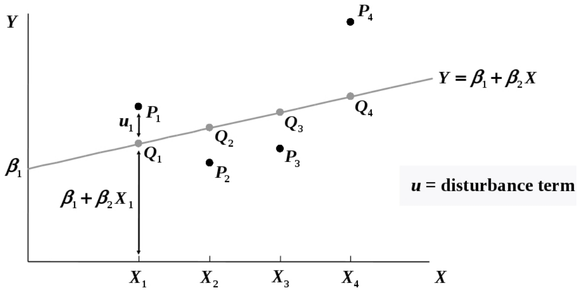
Values of disturbance term have independent distributions

- We assume that the error terms are absolutely independent of each other
- u_i is independently distributed from u_j for all $i \neq j$
- Behaviour/value of the error term in one observation should not affect it's value for another observation



Assumption 5

Values of disturbance term have independent distributions



- If the error term is large (positive/negative) in one observation, it should not be equally large (positive/negative) in the next observation

$$\sigma_{u_i u_j} = E(u_i - \mu_u)(u_j - \mu_u) = E(u_i u_j) = E(u_i)E(u_j) = 0$$

disturbance terms have independent distributions vs. dependent distribution

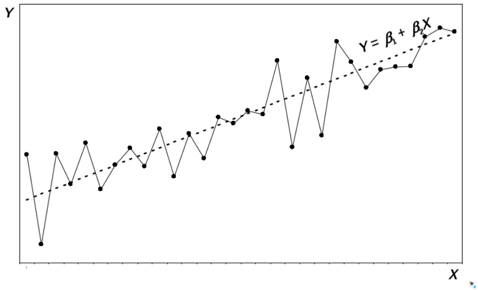


Figure: Values of disturbance terms have independent distributions

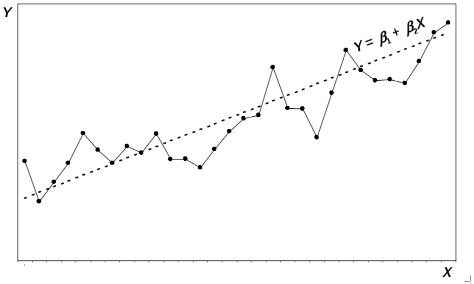
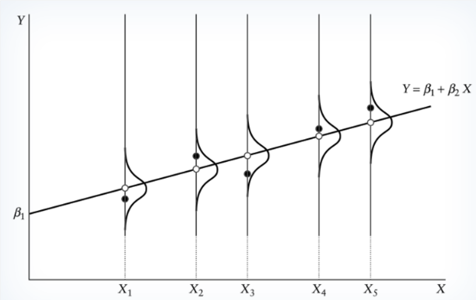


Figure: Values of disturbance terms have dependent distributions

Assumption 6

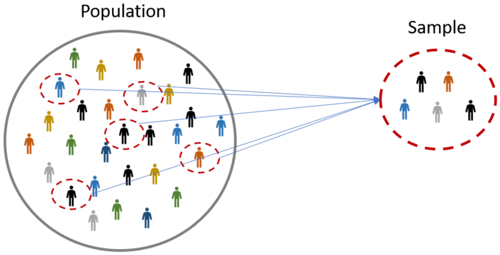
The disturbance term has a normal distribution

- We assume the error terms come from a normal distribution
- This assumption is by virtue of the central limit theorem
- Once we assume that the error term has a normal distribution, we can assume that $\hat{\beta}_1$ and $\hat{\beta}_2$ will also have a normal distribution, allowing us to carry out hypothesis tests on them.



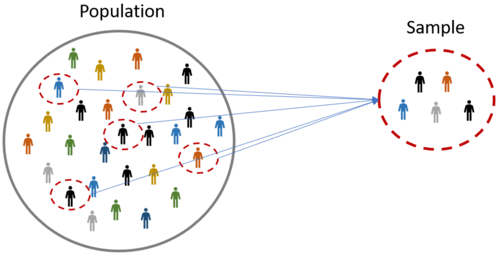
Regression coefficients are random variables

- The OLS regressions coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are random variables
- Different samples from the same population will contain different Y_i, therefore different u_i different values of $\hat{\beta}_1$ and $\hat{\beta}_2$
- Put simply, one sample of data on X and Y will produce one set of values for $\hat{\beta}_1$ and $\hat{\beta}_2$. Another sample of data on X and Y will produce another set of values for $\hat{\beta}_1$ and $\hat{\beta}_2$, and so on ...



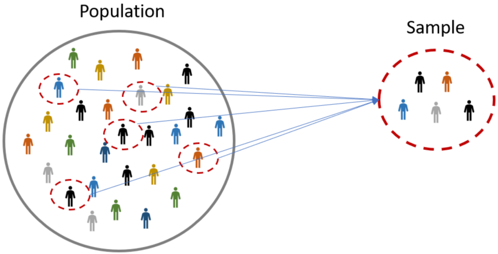
Regression coefficients are random variables

- The OLS regressions coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are random variables
- Different samples from the same population will contain different Y_i, therefore different u_i different values of $\hat{\beta}_1$ and $\hat{\beta}_2$
- Put simply, one sample of data on X and Y will produce one set of values for $\hat{\beta}_1$ and $\hat{\beta}_2$. Another sample of data on X and Y will produce another set of values for $\hat{\beta}_1$ and $\hat{\beta}_2$, and so on ...



Regression coefficients are random variables

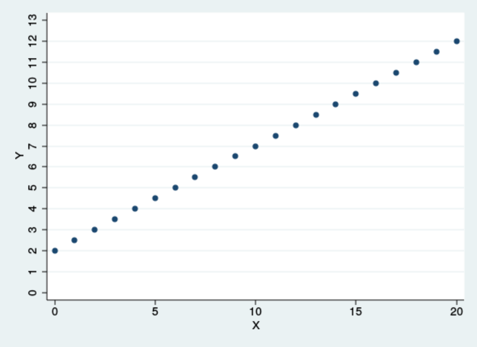
- The OLS regressions coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ are random variables
- Different samples from the same population will contain different Y_i, therefore different u_i different values of $\hat{\beta}_1$ and $\hat{\beta}_2$
- Put simply, one sample of data on X and Y will produce one set of values for $\hat{\beta}_1$ and $\hat{\beta}_2$. Another sample of data on X and Y will produce another set of values for $\hat{\beta}_1$ and $\hat{\beta}_2$, and so on . . .

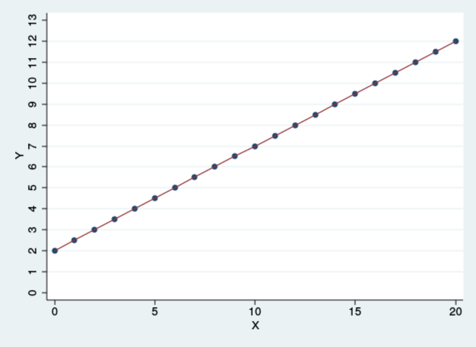


How do sample estimates relate to population parameters?

- Monte Carlo Simulation: a controlled artificial experiment
- Rationale: Create artificial data and set population β_1 and β_2, e.g.:

$$Y_i = 2 + 0.5X_i + u_i$$

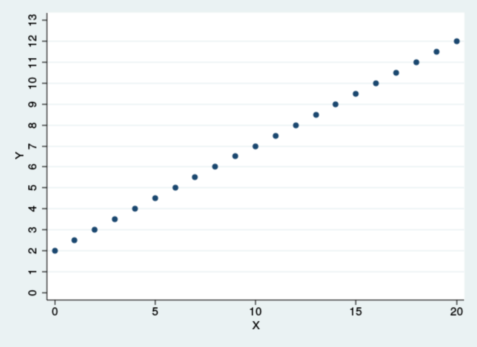


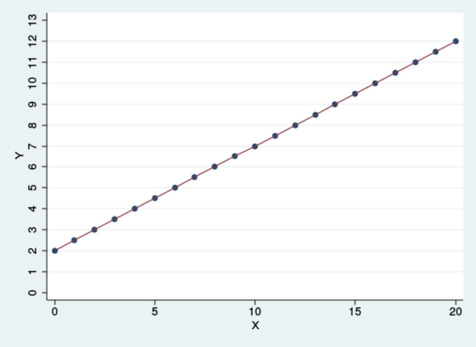


How do sample estimates relate to population parameters?

- Monte Carlo Simulation: a controlled artificial experiment
- Rationale: Create artificial data and set population β_1 and β_2, e.g.:

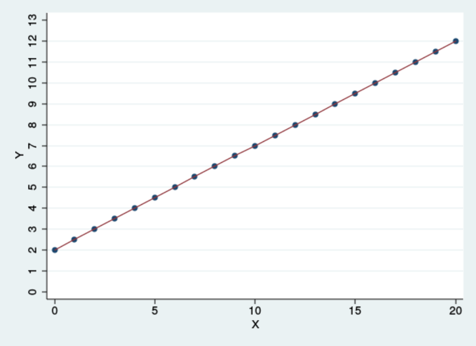
$$Y_i = 2 + 0.5X_i + u_i$$

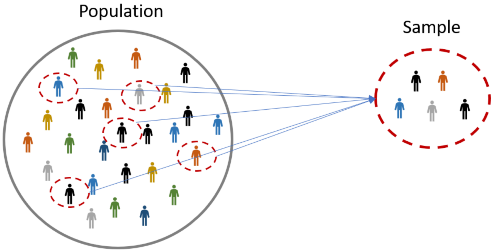




“Real” Population linear relationship

We take random samples from this population and estimate β_1 and β_2



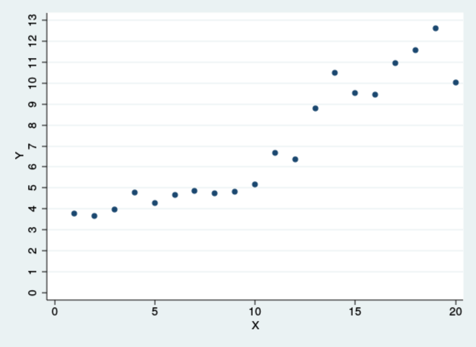


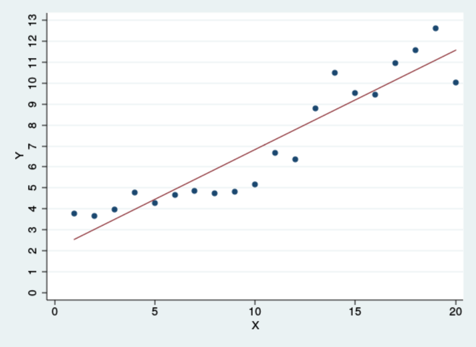
Experiment 1: get a sample with 20 observations

- We start with a sample that contains 20 observations, with values of X as whole numbers ranging from 1-20.
- u_i values are drawn from a standard normal distribution (i.e. with mean 0 and variance 1)

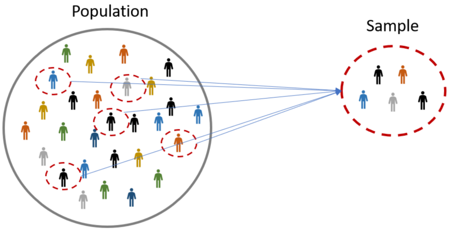
X	u	Y	X	u	Y
1	1.26	3.76	11	-0.82	6.68
2	0.64	3.64	12	-1.62	6.38
3	0.46	3.96	13	0.31	8.81
4	0.77	4.77	14	1.49	10.49
5	-0.23	4.27	15	0.03	9.53
6	-0.35	4.65	16	-0.56	9.44
7	-0.65	4.85	17	0.45	10.95
8	-1.26	4.74	18	0.60	11.60
9	-1.70	4.80	19	1.11	12.61
10	-1.83	5.17	20	-1.96	10.04

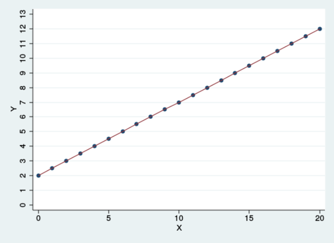
Plotting the sample data

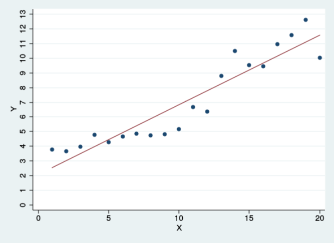




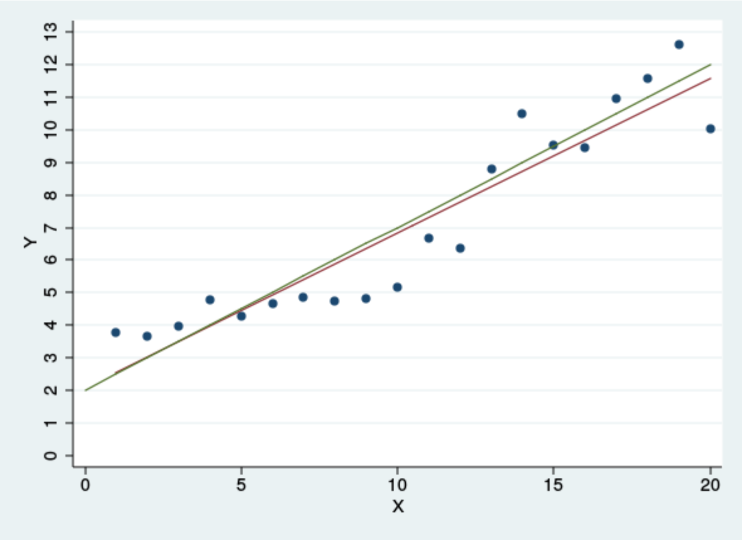
Sample estimated regression line vs. population regression line







Sample estimated regression line vs. population regression line



- The "Real" population model is

$$Y_i = 2 + 0.5X_i + u_i$$

- Sample estimated model: from the above table, regression of Y on X gives $\hat{\beta}_1 = 2.05$ and $\hat{\beta}_2 = 0.48$, i.e.:

$$\hat{Y}_i = 2.05 + 0.48X_i$$

- Note: $\hat{\beta}_1$ is an overestimate of β_1 (i.e. $2.05 > 2$) but $\hat{\beta}_2$ slightly underestimates β_2 (i.e. $0.48 < 0.5$)

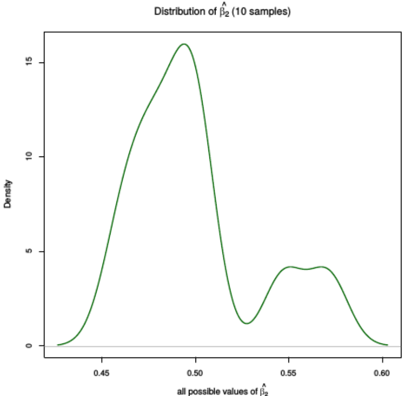
Simulating Sample Distribution

- The "Real" population model is

$$Y_i = 2 + 0.5X_i + u_i$$

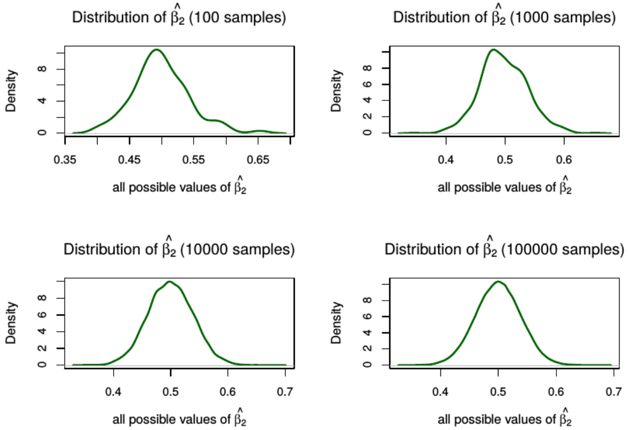
- Take 10 samples from population and estimate $\hat{\beta}_1$ and $\hat{\beta}_2$ for each sample:

Sample	$\hat{\beta}_1$	$\hat{\beta}_2$
1	2.05	0.48
2	1.77	0.50
3	1.45	0.57
4	1.52	0.55
5	2.15	0.48
6	1.86	0.50
7	2.52	0.47
8	2.19	0.50
9	2.25	0.46
10	1.86	0.49



Simulating Sample Distribution

- See distributions of $\hat{\beta}_2$ with 100, 1000, 10,000, 100,000 samples.
- Greater sample - average value of the $\hat{\beta}_2$ converges onto the true population $\beta_2 = 0.5$



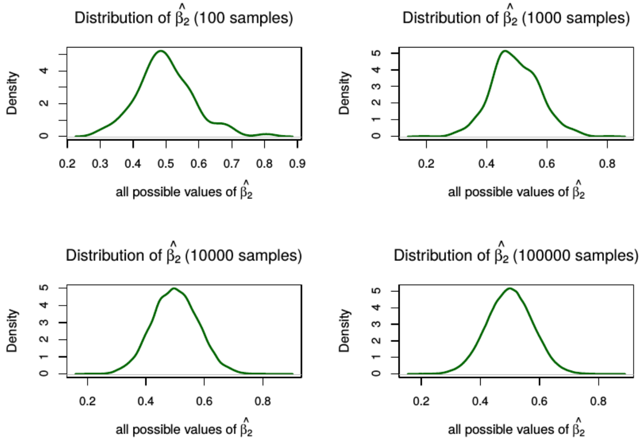
Experiment 2: get a sample with 20 observations again

- We consider another experiment
- We get a sample that contains 20 observations again, with values of X as whole numbers ranging from 1-20.
- But we now **double** the values of the disturbance terms

X	u	Y	X	u	Y
1	2.52	5.02	11	-1.64	5.86
2	1.28	4.28	12	-3.24	4.76
3	0.92	4.42	13	0.62	9.12
4	1.54	5.54	14	2.98	11.98
5	-0.46	4.04	15	0.06	9.56
6	-0.70	4.30	16	-1.12	8.88
7	-1.30	4.20	17	0.90	11.40
8	-2.52	3.48	18	1.20	12.20
9	-3.40	3.10	19	2.22	13.72
10	-3.66	3.34	20	-3.92	8.08

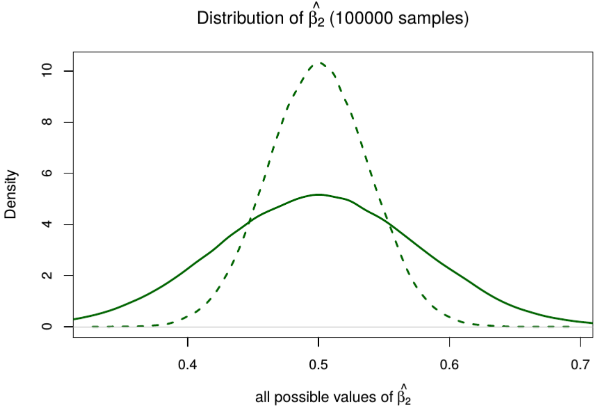
Simulating Sample Distribution

- See distributions of $\hat{\beta}_2$ with 100, 1,000, 10,000, 100,000 samples.
- Note the limits of the values on the y-axis



Compare experiment 1 and 2

- Comparison of the distributions of $\hat{\beta}_2$ in above two experiments
- Doubling of u_i doubles std deviation of distribution.



- Sampling distributions: dotted line u_i, bold line $2 \times u_i$

Application

- In practice:
  - We do not know the true population values β_1 and β_2
  - We will collect only **one** sample from which we run only **one** regression
  - We will get only one set of values for $\hat{\beta}_1$ and $\hat{\beta}_2$ from that regression
- Whatever the mean value of that sampling distribution, our estimated $\hat{\beta}_2$ will appear somewhere on the x-axis
- Key question: where on the x-axis is our sample value of $\hat{\beta}_2$ located?
- This is where we need to rely on statistical theory - hypothesis testing!!!