BS2280 - Econometrics 1

Lecture 2 - Part 1: Analysing Economic Relationships

by Dr Yichen Zhu

Outline

Data Structure

- Data sets for econometric analysis generally come in three forms:
  1) Cross-section data
  2) Time series data
  3) Panel data
- Each gives rise to particular econometric challenges
- Large literature in econometrics devoted to each
- We will focus on cross-sectional data, but useful to be able to understand and recognise each of these different data structure

Data Structure

- Data sets for econometric analysis generally come in three forms:
  1) Cross-section data
  2) Time series data
  3) Panel data
- Each gives rise to particular econometric challenges
- Large literature in econometrics devoted to each
- We will focus on cross-sectional data, but useful to be able to understand and recognise each of these different data structure

Data Structure

- Data sets for econometric analysis generally come in three forms:
  1) Cross-section data
  2) Time series data
  3) Panel data
- Each gives rise to particular econometric challenges
- Large literature in econometrics devoted to each
- We will focus on cross-sectional data, but useful to be able to understand and recognise each of these different data structure

Data Structure

- Data sets for econometric analysis generally come in three forms:
  1) Cross-section data
  2) Time series data
  3) Panel data
- Each gives rise to particular econometric challenges
- Large literature in econometrics devoted to each
- We will focus on cross-sectional data, but useful to be able to understand and recognise each of these different data structure

Cross-sectional Data I

- Observations on different individuals at a single point in time
- Individuals can be people, companies, countries, households, etc.

Person	Year	Income (GBP)	Gender
1	2010	22,000	Male
2	2010	23,265	Male
3	2010	58,000	Female
4	2010	55,998	Male
5	2010	15,000	Male
6	2010	12,350	Female
...	...	...	...

- Note: only the person changes - the year is the same for everyone, i.e. each person is observed in year 2010 only.
- Each person is observed only once in the period under consideration.

Cross-sectional Data I

- Observations on different individuals at a single point in time
- Individuals can be people, companies, countries, households, etc.

Person	Year	Income (GBP)	Gender
1	2010	22,000	Male
2	2010	23,265	Male
3	2010	58,000	Female
4	2010	55,998	Male
5	2010	15,000	Male
6	2010	12,350	Female
...	...	...	...

- Note: only the person changes - the year is the same for everyone, i.e. each person is observed in year 2010 only.
- Each person is observed only once in the period under consideration.

Cross-sectional Data I

- Observations on different individuals at a single point in time
- Individuals can be people, companies, countries, households, etc.

Person	Year	Income (GBP)	Gender
1	2010	22,000	Male
2	2010	23,265	Male
3	2010	58,000	Female
4	2010	55,998	Male
5	2010	15,000	Male
6	2010	12,350	Female
...	...	...	...

- Note: only the person changes - the year is the same for everyone, i.e. each person is observed in year 2010 only.
- Each person is observed only once in the period under consideration.

Cross-sectional Data I

- Observations on different individuals at a single point in time
- Individuals can be people, companies, countries, households, etc.

Person	Year	Income (GBP)	Gender
1	2010	22,000	Male
2	2010	23,265	Male
3	2010	58,000	Female
4	2010	55,998	Male
5	2010	15,000	Male
6	2010	12,350	Female
...	...	...	...

- Note: only the person changes - the year is the same for everyone, i.e. each person is observed in year 2010 only.
- Each person is observed only once in the period under consideration.

Cross-sectional Data II

- Conventional to refer to variables sourced from cross-sectional data with a subscript i
- $Income_i$ refers to the income for each individual i, hence:
  $(i = 1) \rightarrow Income_1 = £22{,}000$
  $(i = 2) \rightarrow Income_2 = £23{,}265$
  and so on ...
- Notations X_i and Y_i in econometrics refer to cross-sectional data.

Cross-sectional Data II

- Conventional to refer to variables sourced from cross-sectional data with a subscript i
- $Income_i$ refers to the income for each individual i, hence:
  $(i = 1) \rightarrow Income_1 = £22{,}000$
  $(i = 2) \rightarrow Income_2 = £23{,}265$
  and so on ...
- Notations X_i and Y_i in econometrics refer to cross-sectional data.

Cross-sectional Data II

- Conventional to refer to variables sourced from cross-sectional data with a subscript i
- $Income_i$ refers to the income for each individual i, hence:
  $(i = 1) \rightarrow Income_1 = £22{,}000$
  $(i = 2) \rightarrow Income_2 = £23{,}265$
  and so on ...
- Notations X_i and Y_i in econometrics refer to cross-sectional data.

Time-series Data I

- Observations on one individual at several points in time
- Time can be (mili-)seconds, minutes, days, weeks, months, quarters, years, five-year periods, etc.

Person	Year	Income (GBP)	Gender
3	2010	58,000	Female
3	2011	59,740	Female
3	2012	61,533	Female
3	2013	63,378	Female
3	2014	65,280	Female
3	2015	67,433	Female
...	...	...	...

- Note: the same person, but time (year) changes.
- Examples: daily share price of Apple, daily Euro/GBP exchange rate, quarterly GDP growth for the UK.

Time-series Data I

- Observations on one individual at several points in time
- Time can be (mili-)seconds, minutes, days, weeks, months, quarters, years, five-year periods, etc.

Person	**Year**	**Income (GBP)**	**Gender**
3	2010	58,000	Female
3	2011	59,740	Female
3	2012	61,533	Female
3	2013	63,378	Female
3	2014	65,280	Female
3	2015	67,433	Female
...	...	...	...

- Note: the same person, but time (year) changes.
- Examples: daily share price of Apple, daily Euro/GBP exchange rate, quarterly GDP growth for the UK.

Time-series Data I

- Observations on one individual at several points in time
- Time can be (mili-)seconds, minutes, days, weeks, months, quarters, years, five-year periods, etc.

Person	Year	Income (GBP)	Gender
3	2010	58,000	Female
3	2011	59,740	Female
3	2012	61,533	Female
3	2013	63,378	Female
3	2014	65,280	Female
3	2015	67,433	Female
...	...	...	...

- Note: the same person, but time (year) changes.
- Examples: daily share price of Apple, daily Euro/GBP exchange rate, quarterly GDP growth for the UK.

Time-series Data I

- Observations on one individual at several points in time
- Time can be (mili-)seconds, minutes, days, weeks, months, quarters, years, five-year periods, etc.

Person	Year	Income (GBP)	Gender
3	2010	58,000	Female
3	2011	59,740	Female
3	2012	61,533	Female
3	2013	63,378	Female
3	2014	65,280	Female
3	2015	67,433	Female
...	...	...	...

- Note: the same person, but time (year) changes.
- Examples: daily share price of Apple, daily Euro/GBP exchange rate, quarterly GDP growth for the UK.

Time Series Data II

- Conventional to refer to variables sourced from time series data with a subscript t.
- $Income_t$ refers to the income for the individual at time t, hence:
  $(t = 1) \rightarrow Income_{2010} = £58{,}000$
  $(t = 2) \rightarrow Income_{2011} = £59{,}740$
  and so on ...
- Notations X_t and Y_t in econometrics refer to time series data. In other words, values on variables X and Y are observed over time t

Time Series Data II

- Conventional to refer to variables sourced from time series data with a subscript t.
- $Income_t$ refers to the income for the individual at time t, hence:
  $(t = 1) \rightarrow Income_{2010} = £58{,}000$
  $(t = 2) \rightarrow Income_{2011} = £59{,}740$
  and so on . . .
- Notations X_t and Y_t in econometrics refer to time series data. In other words, values on variables X and Y are observed over time t

Time Series Data II

- Conventional to refer to variables sourced from time series data with a subscript t.
- $Income_t$ refers to the income for the individual at time t, hence:
  $(t = 1) \rightarrow Income_{2010} = £58{,}000$
  $(t = 2) \rightarrow Income_{2011} = £59{,}740$
  and so on ...
- Notations X_t and Y_t in econometrics refer to time series data. In other words, values on variables X and Y are observed over time t

Panel Data I

- Observations on many individuals at different points in time
- A combination of cross-sectional and time series data

Person	Year	Income (GBP)	Gender
1	2010	22,000	Female
1	2011	22,240	Female
1	2012	22,442	Female
2	2010	23,265	Male
2	2011	23,963	Male
2	2012	24,682	Male
3	2010	58,000	Female
3	2011	59,740	Female
3	2012	61,533	Female
...	...	...	...

- Examples: GDP data on all countries in the world since 1975, daily admissions in every hospital in England in 2016

Panel Data I

- Observations on many individuals at different points in time
- A combination of cross-sectional and time series data

Person	Year	Income (GBP)	Gender
1	2010	22,000	Female
1	2011	22,240	Female
1	2012	22,442	Female
2	2010	23,265	Male
2	2011	23,963	Male
2	2012	24,682	Male
3	2010	58,000	Female
3	2011	59,740	Female
3	2012	61,533	Female
...	...	...	...

- Examples: GDP data on all countries in the world since 1975, daily admissions in every hospital in England in 2016

Panel Data I

- Observations on many individuals at different points in time
- A combination of cross-sectional and time series data

Person	Year	Income (GBP)	Gender
1	2010	22,000	Female
1	2011	22,240	Female
1	2012	22,442	Female
2	2010	23,265	Male
2	2011	23,963	Male
2	2012	24,682	Male
3	2010	58,000	Female
3	2011	59,740	Female
3	2012	61,533	Female
...	...	...	...

- Examples: GDP data on all countries in the world since 1975, daily admissions in every hospital in England in 2016

Panel Data II

- Conventional to refer to variables sourced from time series data with a subscript *it*.
- $Income_{it}$ refers to the income for the individual i at time t, hence:
  $(i = 1, t = 1) \rightarrow Income_{1,2010} = £22{,}000$
  $(i = 1, t = 2) \rightarrow Income_{1,2011} = £22{,}240$
  $(i = 3, t = 2) \rightarrow Income_{3,2011} = £59{,}740$
  and so on ...
- Notations X_{it} and Y_{it} in econometrics refer to panel data set.

Panel Data II

- Conventional to refer to variables sourced from time series data with a subscript *it*.
- $Income_{it}$ refers to the income for the individual i at time t, hence:
  $(i = 1, t = 1) \rightarrow Income_{1,\,2010} = £22{,}000$
  $(i = 1, t = 2) \rightarrow Income_{1,\,2011} = £22{,}240$
  $(i = 3, t = 2) \rightarrow Income_{3,\,2011} = £59{,}740$
  and so on ...
- Notations X_{it} and Y_{it} in econometrics refer to panel data set.

Panel Data II

- Conventional to refer to variables sourced from time series data with a subscript *it*.
- $Income_{it}$ refers to the income for the individual i at time t, hence:
  $(i = 1, t = 1) \rightarrow Income_{1,2010} = £22{,}000$
  $(i = 1, t = 2) \rightarrow Income_{1,2011} = £22{,}240$
  $(i = 3, t = 2) \rightarrow Income_{3,2011} = £59{,}740$
  and so on . . .
- Notations X_{it} and Y_{it} in econometrics refer to panel data set.

Data Structure Summary

Cross-sectional Data	**Time Series Data**	**Panel Data**
Different individuals, single time point	One individual, several points in time	Many individuals, different points in time
Notations X_i and Y_i	Notations X_t and Y_t	Notations X_{it} and Y_{it}
$(i = 1) \rightarrow Income_1 = £22{,}000$	$(t = 1) \rightarrow Income_{2010} = £58{,}000$	$(i = 1, t = 1) \rightarrow Income_{1,\,2010} = £22{,}000$
$(i = 2) \rightarrow Income_2 = £23{,}265$	$(t = 2) \rightarrow Income_{2011} = £59{,}740$	$(i = 1, t = 2) \rightarrow Income_{1,\,2011} = £22{,}240$

Covariance and Correlation

Relationship between two variables

- We often want to analyse the relationship between two variables
- Examples:
  - Shall I attend classes? Attendance and exam marks
  - Shall the government increase numbers of police officers? Police officers and crime rate
  - Shall I go to university? Education and income
- We know already know two numerical measures: Covariance and correlation coefficient

Relationship between two variables

- We often want to analyse the relationship between two variables
- Examples:
  - Shall I attend classes? Attendance and exam marks
  - Shall the government increase numbers of police officers? Police officers and crime rate
  - Shall I go to university? Education and income
- We know already know two numerical measures: Covariance and correlation coefficient

Relationship between two variables

- We often want to analyse the relationship between two variables
- Examples:
  - Shall I attend classes? Attendance and exam marks
  - Shall the government increase numbers of police officers? Police officers and crime rate
  - Shall I go to university? Education and income
- We know already know two numerical measures: Covariance and correlation coefficient

Relationship between two variables

- We often want to analyse the relationship between two variables
- Examples:
  - Shall I attend classes? Attendance and exam marks
  - Shall the government increase numbers of police officers? Police officers and crime rate
  - Shall I go to university? Education and income
- We know already know two numerical measures: Covariance and correlation coefficient

Relationship between two variables

- We often want to analyse the relationship between two variables
- Examples:
  - Shall I attend classes? Attendance and exam marks
  - Shall the government increase numbers of police officers? Police officers and crime rate
  - Shall I go to university? Education and income
- We know already know two numerical measures: Covariance and correlation coefficient

Relationship between two variables

- We often want to analyse the relationship between two variables
- Examples:
  - Shall I attend classes? Attendance and exam marks
  - Shall the government increase numbers of police officers? Police officers and crime rate
  - Shall I go to university? Education and income
- We know already know two numerical measures: Covariance and correlation coefficient

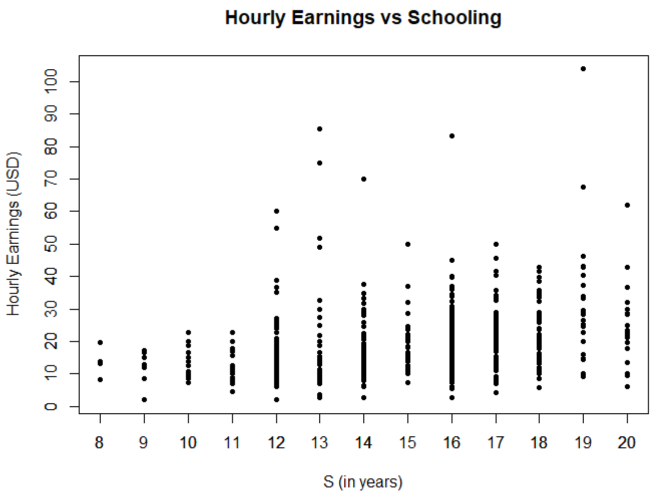
Example: Earnings and education

	S	EARNINGS
1	12	15.00
2	16	83.33
3	18	15.09
4	16	18.36
5	17	15.38
6	14	11.54
7	12	10.53
8	12	17.00
9	12	27.13
10	14	29.81
11	18	20.91
12	16	24.13
13	10	12.50
14	11	17.25
15	12	18.00
16	19	20.00
17	12	7.25

- Snapshot of data on the two variables of interest
- Useful to know summary statistics of these
- Hourly earning range from a minimum of \$2 to a maximum of approx. \$104
- School level ranges from the 8th grade (middle school) to the 20th grade (PhD)

```
> summary(EAWE21.simple)
S              EARNINGS
Min.   : 8.00   Min.   :  2.00
1st Qu.:12.00   1st Qu.: 11.98
Median :15.00   Median : 17.00
Mean   :14.87   Mean   : 19.58
3rd Qu.:17.00   3rd Qu.: 23.93
Max.   :20.00   Max.   :103.85
```

Relationship between hourly wages and education



Covariance

- Estimating the covariance: this measures how closely two variables are move together.

$$s_{XY} = \frac{\sum_{t=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

- Note: We usually use the sample measure rather than the population measure, therefore we divide by $n - 1$
- Result (using R):

```
> cov(EAWE21.simple$EARNINGS, EAWE21.simple$S)
[1] 9.522066
```

Covariance

- Estimating the covariance: this measures how closely two variables are move together.

$$s_{XY} = \frac{\sum_{t=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

- Note: We usually use the sample measure rather than the population measure, therefore we divide by $n-1$
- Result (using R):

```
> cov(EAWE21.simple$EARNINGS, EAWE21.simple$S)
[1] 9.522066
```

Covariance

- Estimating the covariance: this measures how closely two variables are move together.

$$s_{XY} = \frac{\sum_{t=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

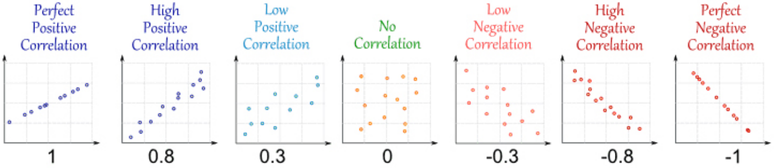
- Note: We usually use the sample measure rather than the population measure, therefore we divide by $n - 1$
- Result (using R):

```
> cov(EAWE21.simple$EARNINGS, EAWE21.simple$S)
[1] 9.522066
```

Limitations of Covariance

- Size of covariance is difficult to interpret
- Not easy to compare covariances across samples
- We need a standardised measure:

  Correlation Coefficient (ρ)
- Is always between -1 and $+1$
- -1: perfect negative correlation
- 0: no correlation
- +1: perfect positive correlation

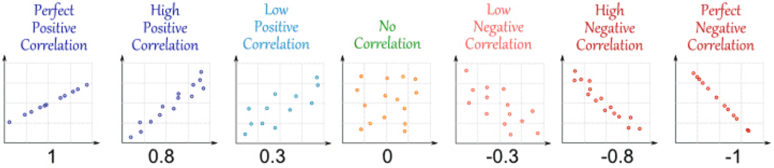


Limitations of Covariance

- Size of covariance is difficult to interpret
- Not easy to compare covariances across samples
- We need a standardised measure:

  Correlation Coefficient (ρ)

- Is always between -1 and $+1$
- -1: perfect negative correlation
- 0: no correlation
- +1: perfect positive correlation

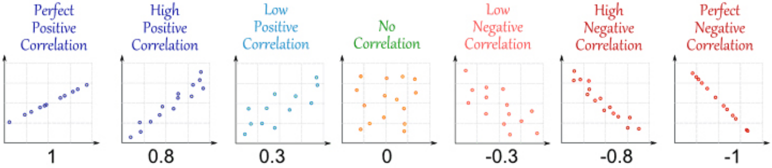


Limitations of Covariance

- Size of covariance is difficult to interpret
- Not easy to compare covariances across samples
- We need a standardised measure:

Correlation Coefficient (ρ)

- Is always between -1 and $+1$
- -1: perfect negative correlation
- 0: no correlation
- +1: perfect positive correlation

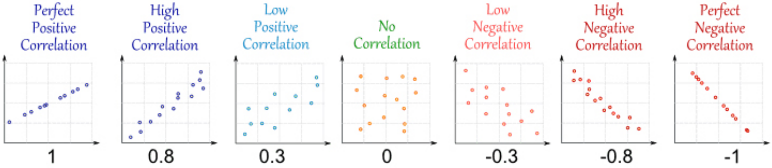


Limitations of Covariance

- Size of covariance is difficult to interpret
- Not easy to compare covariances across samples
- We need a standardised measure:

Correlation Coefficient (ρ)

- Is always between -1 and $+1$
- -1: perfect negative correlation
- 0: no correlation
- +1: perfect positive correlation

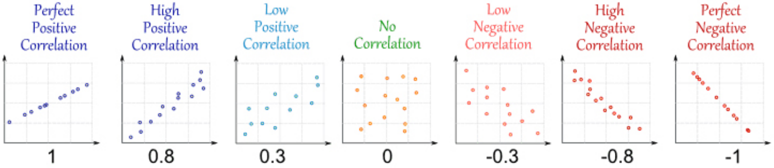


Limitations of Covariance

- Size of covariance is difficult to interpret
- Not easy to compare covariances across samples
- We need a standardised measure:

Correlation Coefficient (ρ)

- Is always between -1 and $+1$
- -1: perfect negative correlation
- 0: no correlation
- +1: perfect positive correlation

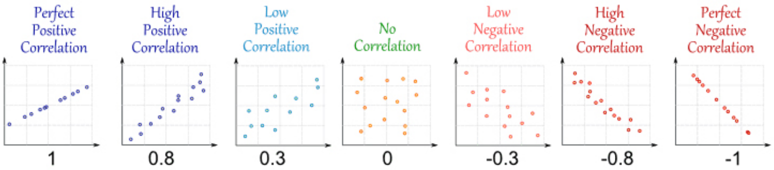


Limitations of Covariance

- Size of covariance is difficult to interpret
- Not easy to compare covariances across samples
- We need a standardised measure:

Correlation Coefficient (ρ)

- Is always between -1 and $+1$
- -1: perfect negative correlation
- 0: no correlation
- +1: perfect positive correlation

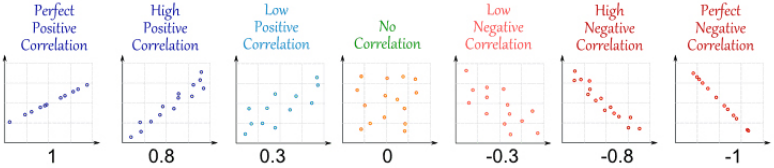


Limitations of Covariance

- Size of covariance is difficult to interpret
- Not easy to compare covariances across samples
- We need a standardised measure:

Correlation Coefficient (ρ)

- Is always between -1 and $+1$
- -1: perfect negative correlation
- 0: no correlation
- +1: perfect positive correlation



Correlation Coefficient

- Formula:

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} \qquad -1 \leq s_{XY} \leq 1$$

- Result (using R):

```
> cor(EAWE21.simple$EARNINGS, EAWE21.simple$S)
[1] 0.2924262
```

- This shows that education and earnings are positively but rather weakly correlated.
- We will require further statistical tools to analyse this relationship!

Correlation Coefficient

- Formula:

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} \qquad -1 \leq s_{XY} \leq 1$$

- Result (using R):

```
> cor(EAWE21.simple$EARNINGS, EAWE21.simple$S)
[1] 0.2924262
```

- This shows that education and earnings are positively but rather weakly correlated.
- We will require further statistical tools to analyse this relationship!

Correlation Coefficient

- Formula:

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} \qquad -1 \leq s_{XY} \leq 1$$

- Result (using R):

```
> cor(EAWE21.simple$EARNINGS, EAWE21.simple$S)
[1] 0.2924262
```

- This shows that education and earnings are positively but rather weakly correlated.
- We will require further statistical tools to analyse this relationship!

Correlation Coefficient

- Formula:

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} \qquad -1 \leq s_{XY} \leq 1$$

- Result (using R):

```
> cor(EAWE21.simple$EARNINGS, EAWE21.simple$S)
[1] 0.2924262
```

- This shows that education and earnings are positively but rather weakly correlated.
- We will require further statistical tools to analyse this relationship!

Notes on Correlation Coefficients

- ρ_{XY} is independent of units of measurement. If earnings were measured in GDP, USD or EUR, correlation still would be 0.29
- The correlation between Y and X is the same as between X and Y, i.e. it does not matter which variable is labelled X and which Y.
- Finding an association between variables might be suggestive, but is rarely sufficient
- Establishing a casual relationship is more desirable
- Ceteris paribus – plays an important role in causal analysis
- Need to control for sufficient number of variables

Notes on Correlation Coefficients

- ρ_{XY} is independent of units of measurement. If earnings were measured in GDP, USD or EUR, correlation still would be 0.29
- The correlation between Y and X is the same as between X and Y, i.e. it does not matter which variable is labelled X and which Y.
- Finding an association between variables might be suggestive, but is rarely sufficient
- Establishing a casual relationship is more desirable
- Ceteris paribus – plays an important role in causal analysis
- Need to control for sufficient number of variables

Notes on Correlation Coefficients

- ρ_{XY} is independent of units of measurement. If earnings were measured in GDP, USD or EUR, correlation still would be 0.29
- The correlation between Y and X is the same as between X and Y, i.e. it does not matter which variable is labelled X and which Y.
- Finding an association between variables might be suggestive, but is rarely sufficient
- Establishing a casual relationship is more desirable
- Ceteris paribus – plays an important role in causal analysis
- Need to control for sufficient number of variables

Notes on Correlation Coefficients

- ρ_{XY} is independent of units of measurement. If earnings were measured in GDP, USD or EUR, correlation still would be 0.29
- The correlation between Y and X is the same as between X and Y, i.e. it does not matter which variable is labelled X and which Y.
- Finding an association between variables might be suggestive, but is rarely sufficient
- Establishing a casual relationship is more desirable
- Ceteris paribus – plays an important role in causal analysis
- Need to control for sufficient number of variables

Notes on Correlation Coefficients

- ρ_{XY} is independent of units of measurement. If earnings were measured in GDP, USD or EUR, correlation still would be 0.29
- The correlation between Y and X is the same as between X and Y, i.e. it does not matter which variable is labelled X and which Y.
- Finding an association between variables might be suggestive, but is rarely sufficient
- Establishing a casual relationship is more desirable
- Ceteris paribus – plays an important role in causal analysis
- Need to control for sufficient number of variables

Notes on Correlation Coefficients

- ρ_{XY} is independent of units of measurement. If earnings were measured in GDP, USD or EUR, correlation still would be 0.29
- The correlation between Y and X is the same as between X and Y, i.e. it does not matter which variable is labelled X and which Y.
- Finding an association between variables might be suggestive, but is rarely sufficient
- Establishing a casual relationship is more desirable
- Ceteris paribus – plays an important role in causal analysis
- Need to control for sufficient number of variables

Examples

- How much would your grade increase by 1 extra hour of study?
- How much effect does education have on wages?
- What is the price elasticity of demand for smart phones?
- How much would the price of a stock change with a change in the previous year's profit?
- Does your IQ level increase with the amount of protein intake?

Examples

- How much would your grade increase by 1 extra hour of study?
- How much effect does education have on wages?
- What is the price elasticity of demand for smart phones?
- How much would the price of a stock change with a change in the previous year's profit?
- Does your IQ level increase with the amount of protein intake?

Examples

- How much would your grade increase by 1 extra hour of study?
- How much effect does education have on wages?
- What is the price elasticity of demand for smart phones?
- How much would the price of a stock change with a change in the previous year's profit?
- Does your IQ level increase with the amount of protein intake?

Examples

- How much would your grade increase by 1 extra hour of study?
- How much effect does education have on wages?
- What is the price elasticity of demand for smart phones?
- How much would the price of a stock change with a change in the previous year's profit?
- Does your IQ level increase with the amount of protein intake?

Examples

- How much would your grade increase by 1 extra hour of study?
- How much effect does education have on wages?
- What is the price elasticity of demand for smart phones?
- How much would the price of a stock change with a change in the previous year's profit?
- Does your IQ level increase with the amount of protein intake?

The Simple Regression Model

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

Simple Linear Regression Model

- The regression model hypothesizes a mathematical relationship between variables
- We start with only two variables of interest (Y and X)
- We hypothesise that X affects Y (and not the other way round)
- More formally:

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

β_1 is where the line cuts the Y axis (intercept)
β_2 is the slope of the line (gradient)

Simple Linear Regression Model

- The regression model hypothesizes a mathematical relationship between variables
- We start with only two variables of interest (Y and X)
- We hypothesise that X affects Y (and not the other way round)
- More formally:

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

β_1 is where the line cuts the Y axis (intercept)
β_2 is the slope of the line (gradient)

Simple Linear Regression Model

- The regression model hypothesizes a mathematical relationship between variables
- We start with only two variables of interest (Y and X)
- We hypothesise that X affects Y (and not the other way round)
- More formally:

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

β_1 is where the line cuts the Y axis (intercept)
β_2 is the slope of the line (gradient)

Simple Linear Regression Model

- The regression model hypothesizes a mathematical relationship between variables
- We start with only two variables of interest (Y and X)
- We hypothesise that X affects Y (and not the other way round)
- More formally:

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

β_1 is where the line cuts the Y axis (intercept)
β_2 is the slope of the line (gradient)

Simple Linear Regression Model

- There are different ways of numbering coefficients of a regression model.

Dougherty & some textbooks	Other textbooks
$Y = \beta_1 + \beta_2 X$	$Y = \beta_0 + \beta_1 X$

- We use Dougherty's convention for numbering parameters
- β_1 represents the intercept term and β_2 the slope of the line

Simple Linear Regression Model

- There are different ways of numbering coefficients of a regression model.

Dougherty & some textbooks	Other textbooks
$Y = \beta_1 + \beta_2 X$	$Y = \beta_0 + \beta_1 X$

- We use Dougherty's convention for numbering parameters
- β_1 represents the intercept term and β_2 the slope of the line

Simple Linear Regression Model

- There are different ways of numbering coefficients of a regression model.

Dougherty & some textbooks	Other textbooks
$Y = \beta_1 + \beta_2 X$	$Y = \beta_0 + \beta_1 X$

- We use Dougherty's convention for numbering parameters
- β_1 represents the intercept term and β_2 the slope of the line

Student task

1. Assume that $\beta_1 = 1$ and $\beta_2 = 2$. Draw the regression line.
2. Show how the regression line will change when $\beta_1 = 0$
3. Show how the regression line will change when $\beta_2 = -1$