Simple Linear Regression

CIVL 7012/8012

Causality and ceteris paribus

- One of the important features of statistical analysis is causality
- What is the causal effect of one variable (education) over another (income)
- Ceteris paribus means “with all other (relevant) factors being equal” what is the causal effect of education over income.

The Question of Causality

- Simply establishing a relationship between variables is rarely sufficient
- Want to know the effect to be considered causal
- If we've truly controlled for enough other variables, then the estimated ceteris paribus effect can often be considered to be causal
- Can be difficult to establish causality

What is Regression Analysis

- Many problems in engineering and science involve exploring the relationships between two or more variables.
- **Regression analysis** is a statistical technique that is very useful for these types of problems.
- Let us consider an example

Hypothetical Example (1)

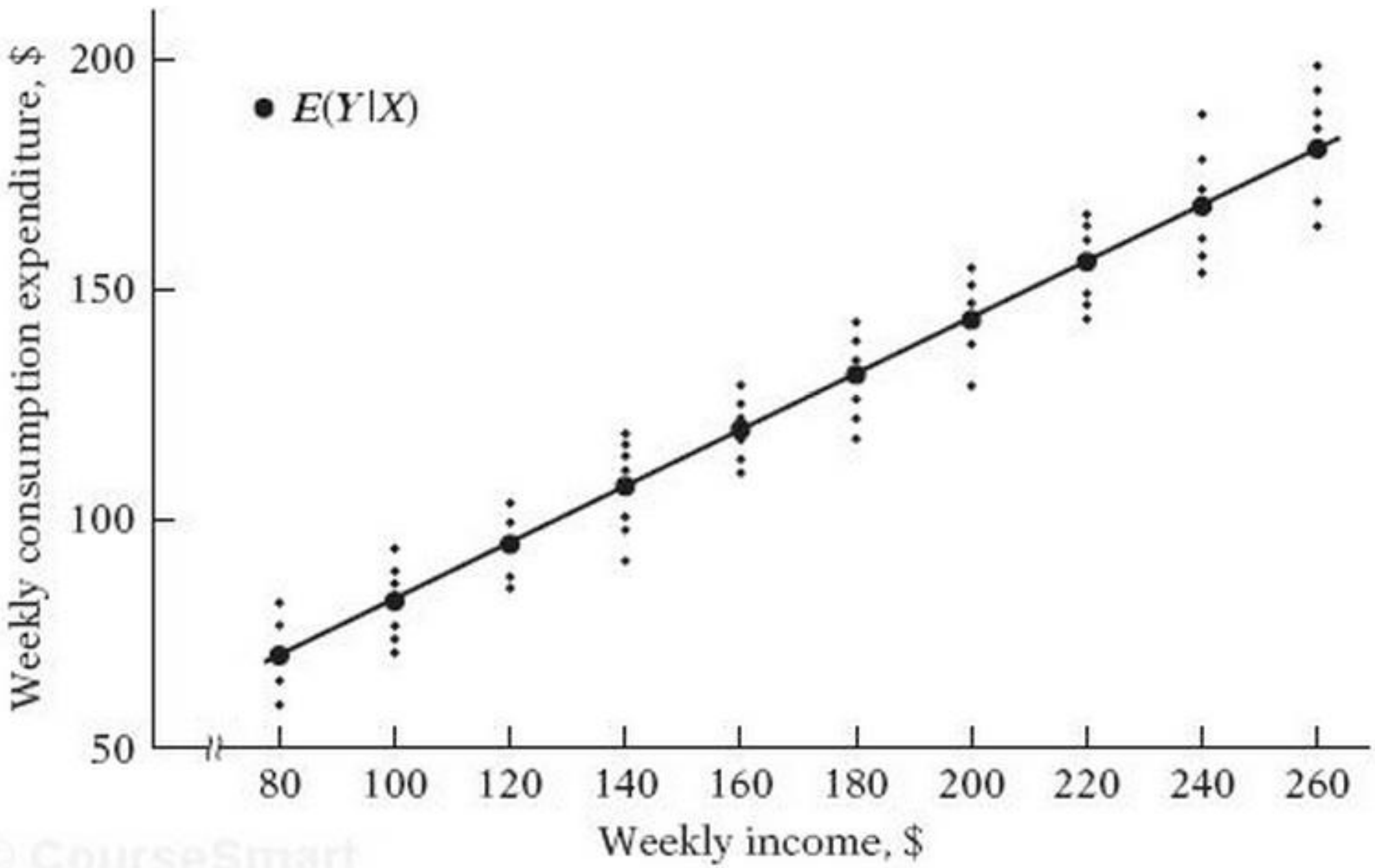
- Let us consider our population be 60 families.
- We collect data on their weekly income (X) and weekly consumption expenditure (Y).
- 60 families are divided into 10 income groups
  - From $80 to $220 in $20 increments

Hypothetical Example (2)

- 10 fixed values of X and their corresponding Y values
- Meaning there are 10 Y subpopulations

X->	Weekly Income ($)									
Y ↓	80	100	120	140	160	180	200	220	240	260
Weekly Expenditure ($)	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	0	88	0	113	125	140	0	160	189	185
	0	0	0	115	0	0	0	162	0	191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional Mean of Yi; E(Y/X)	65	77	89	101	113	125	137	149	161	173

Hypothetical Example (3)



Hypothetical Example (4)

- There is a considerably variation in weekly consumption expenditure (Y)
- On the average weekly consumption expenditure (Y) increases as the income (X) increases.
- If we see the mean weekly income level
  - Weekly income level of $80, mean consumption expenditure is $65
  - Similarly for income level of $200, mean consumption expenditure is $137
- Overall we have 10 mean values for 10 subpopulations of Y
- We can call them conditional expected values.

Hypothetical Example (5)

- Symbolically we can denote them as E(Y/X)
- Which reads as expected value of Y given X
- It is crucial to distinguish between conditional and un-conditional expected value of expected weekly expenditure; i.e
  - E(Y/X), and E(Y)
- For all 60 families un-conditional expected value of expected weekly expenditure, i.e. E(Y) is \$7272/60 = \$121.20

Hypothetical Example (6)

- Question " *what is the expected value of weekly consumption expenditure of a family*"
  - $121.20
- Question " *what is the expected value of weekly consumption expenditure of a family whose monthly income is $80*"
- $65
- Conditional mean: E(Y/X=80)
- Question: *"What is the best (mean) prediction of weekly consumption expenditure of a family whose monthly income is $80"*
  - $65

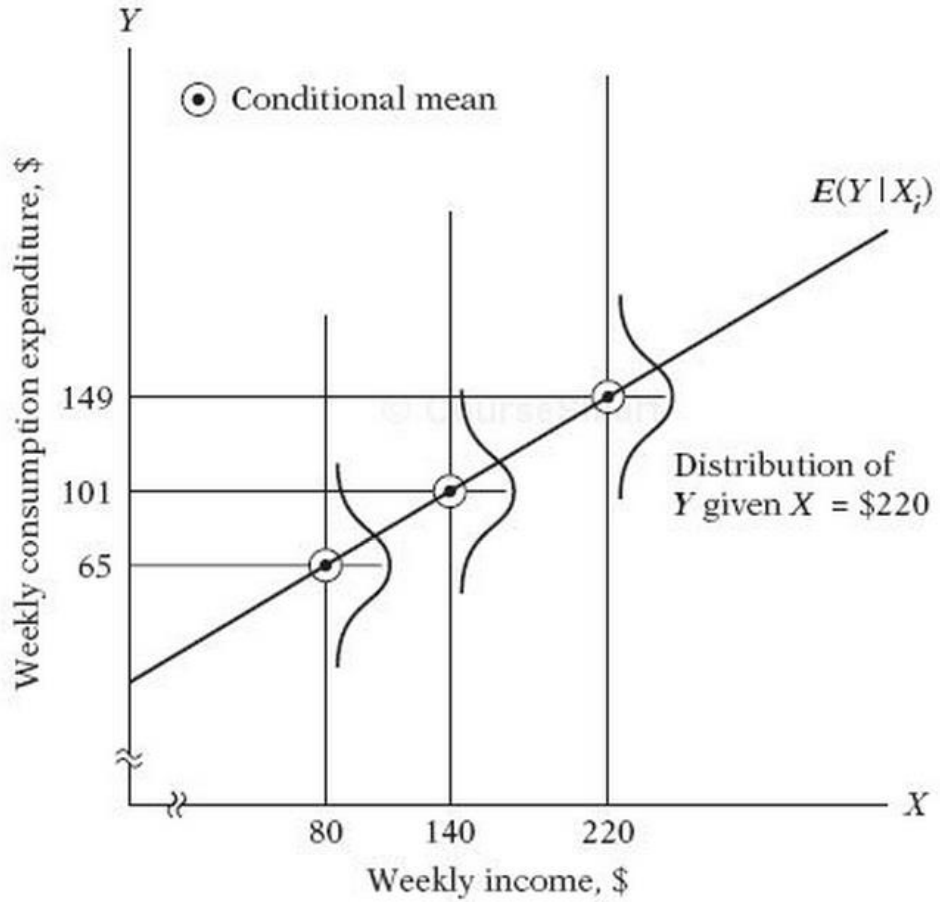
Hypothetical Example (7)

- The knowledge of income level may enable us to better predict the mean values of consumption expenditure than if we do not have this knowledge.

- The conditional expectation is an important aspect of regression analysis.

Hypothetical Example (8)

- Population regression line

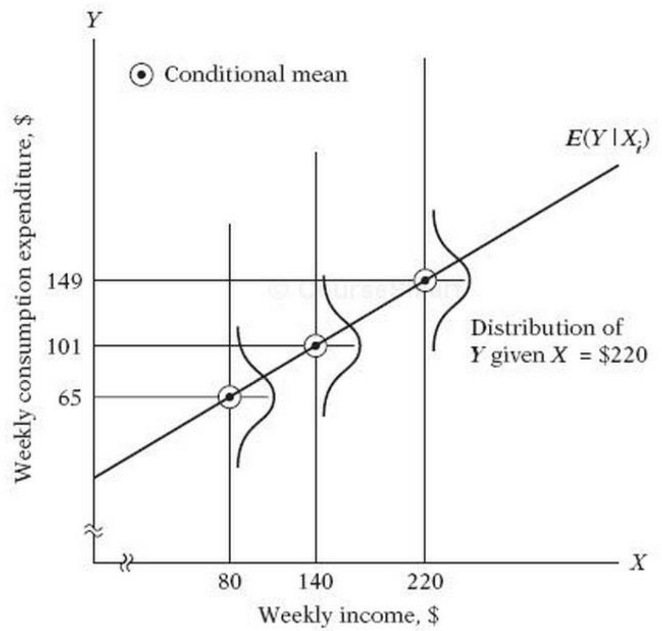


Hypothetical Example (9)

- The dark circles show the conditional mean values of Y against X
- If we join the conditional mean values then we obtain a
  - Population Regression Line (PRL)
  - Also referred as Population Regression Curve or simply Regression Curve
- The adjective *population* comes from the fact that we are dealing in this example with entire population of 60 families. Of curse in reality we can extend this population to many families.

Hypothetical Example (10)

- Geometrically, a PRL is simply the locus of conditional means of the dependent variable (Y) for the fixed values of independent variables (X).
- The PRL passes through these conditional mean values.



Concept of PRF

$$E(Y/X_i) = f(X_i) \text{------------------} (1)$$

Where $f(X_i)$ -> function of the explanatory variable *X*

- Expected conditional mean *Y*; $E(Y/X_i)$ is a function of Xi
- Equation (1) is known as conditional expectation function (*CEF*) or population regression function (*PRF*)
- It suggests that how expected distribution of *Y* given *Xi*.
- Alternatively, how mean or average response of *Y* varies with X

PRF Functional Form (1)

- Which form does f(Xi) assume?
- In reality we do not have the entire population available.
- The functional form of PRF is an empirical question
- For the hypothetical example income was linearly related with expenditure. .
- As first approximation, let us consider that *E(Y/Xi)* is linearly related with *f(Xi)*
- $E\left(\frac{Y}{Xi}\right) = \beta_0 + \beta_1 x$

PRF Functional Form (2)

$$E\left(\frac{Y}{Xi}\right) = \beta_0 + \beta_1 x + u$$

- The linearity means one unit increase in *x* changes the expected value of *y* by the amount of β_1
- What about the disturbance term *u*?
- Since u represents all unobservable variables, and they are random in nature as well, we need to establish a relationship between x and u
- Otherwise we will not be able to estimate β_0 and β_1

The Disturbance Term (1)

- Before we state how u and x are related, we can make one assumption about u
  - As long as intercept β_0 is included in the equation, nothing is lost by assuming that the average value of u in the population is zero.
- i.e. E(u) = 0 ---(2)
- Eq. (2) suggests that the distribution of unobserved factors in the population is zero.
- In the hypothetical example, we can say that average education of all 60 families will be zero (deviation from the mean, some positive and negatives..)
- We can normalize unobserved factors to zero.

The Disturbance Term (2)

- We can now turn into relationship between u and x.
- Since u and x are random variables, correlation coefficient seems an obvious measure to quantify their relationship.
- If u and x are uncorrelated then correlation coefficient is zero’
- But u may be correlated with functions of x such as x2, x3, etc.
- Therefore correlation poses problems for deriving statistical properties.
- A better assumption would be expected value of u given x (or the conditional distribution)

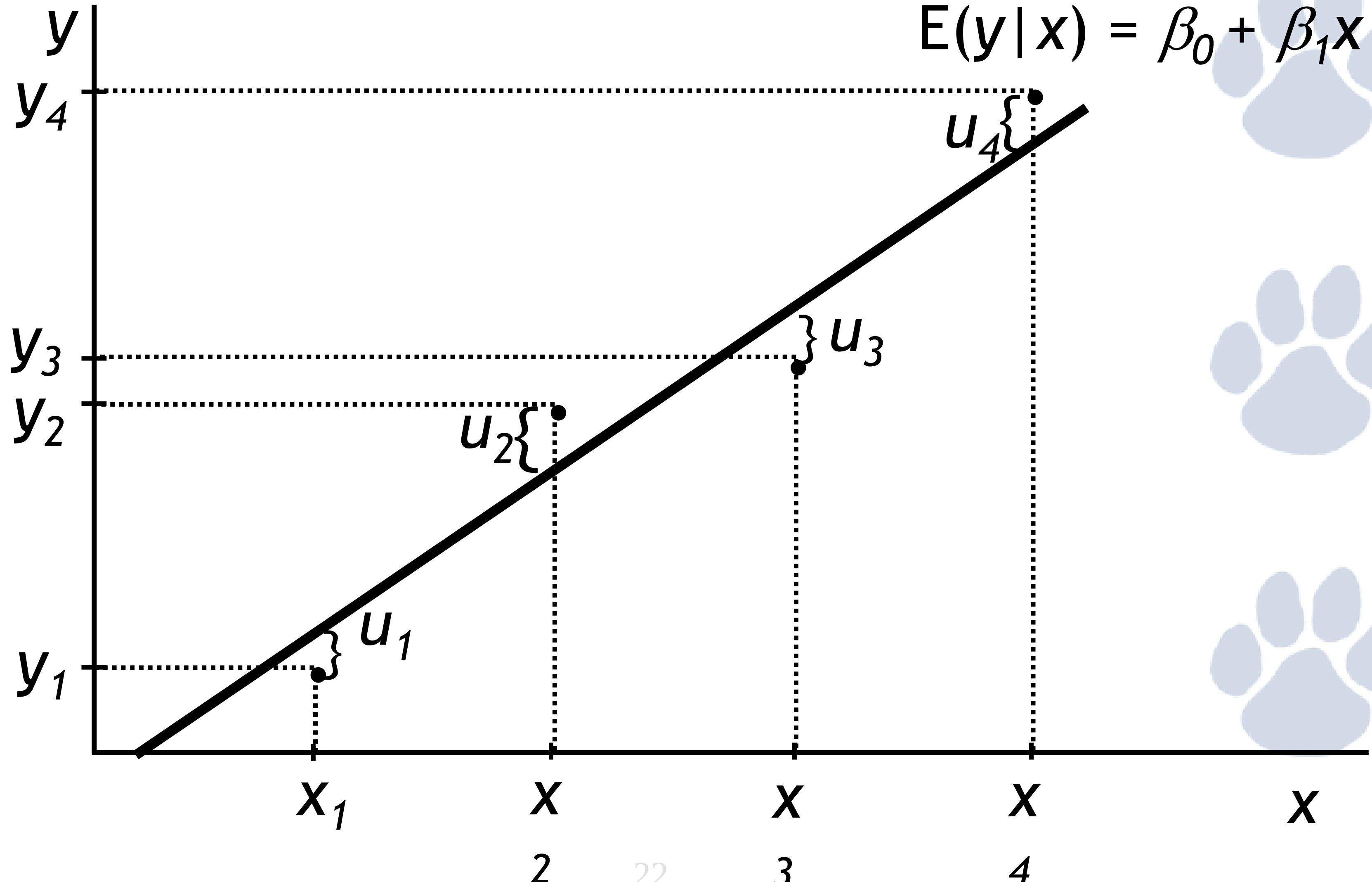
The Disturbance Term (3)

- The conditional distribution of u over x is
- E(u/x) = E(u)-----------------------------------(3)
- Equation (3) suggests that average value of u does not depend on the value of x
- If equation (3) holds true then we can say that *u is mean independent of x*
- By combining equation (2) and (3) we can state the zero conditional mean assumption,
- E(u/x) = 0 ------------------------------------(4)

The Disturbance Term (4)

- Let us see an example; in an effort to determine income as a function of education, we can state that $Income = \beta_0 + \beta_1 education + u$
- Let us say *u* is same as innate ability
- If *E(ability/8)* represents average ability for the group of the population with 8 years of education
- Similarly, If *E(ability/16)* represents average ability for the group of the population with 16 years of education
- As per equation (3) *E(ability/8)* = *E(ability/16)* =0
- As we can not observe innate ability, we have no way of knowing whether or not average ability is same for all education levels.
- So for all unobserved factors we consider that *E(u/x) = 0*
- So the PRF is always $E\left(\frac{Y}{Xi}\right) = \beta_0 + \beta_1 x$

PRF and The Disturbance Term



Deriving OLS Estimates (*1*)

- Basic idea of regression is to estimate the population parameters from a sample
- Let $\{(x_i, y_i): i = 1, \ldots, n\}$ denote a random sample of size n from the population
- For each observation in this sample, it will be the case that
- $y_i = \beta_0 + \beta_1 x_i + u_i$

Deriving OLS Estimates (2)

- To derive the OLS estimates we need to realize that our main assumption of $E(u|x) = E(u) = 0$ also implies that
- $Cov(x,u) = E(xu) = 0$
- Why? Remember from basic probability that $Cov(X,Y) = E(XY) - E(X)E(Y)$

Deriving OLS Estimates (3)

- We can write our 2 restrictions just in terms of x, y, β_0 and β_1 , since $u = y - \beta_0 - \beta_1 x$

- $E(y - \beta_0 - \beta_1 x) = 0$
- $E[x(y - \beta_0 - \beta_1 x)] = 0$

- These are called moment restrictions

Deriving OLS using M.O.M.

- The method of moments approach to estimation implies imposing the population moment restrictions on the sample moments

- What does this mean? Recall that for E(X), the mean of a population distribution, a sample estimator of E(X) is simply the arithmetic mean of the sample

Deriving OLS using M.O.M. (1)

- We want to choose values of the parameters that will ensure that the sample versions of our moment restrictions are true
- The sample versions are as follows:

$$n^{-1}\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

$$n^{-1}\sum_{i=1}^{n}x_i\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0$$

Deriving OLS using M.O.M. (2)

- Given the definition of a sample mean, and properties of summation, we can rewrite the first condition as follows

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

or

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

More Derivation of OLS

$$\sum_{i=1}^{n} x_i \left(y_i - \left(\bar{y} - \hat{\beta}_1 \bar{x} \right) - \hat{\beta}_1 x_i \right) = 0$$

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

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

So the OLS estimated slope is

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

$$\text{provided that} \sum_{i=1}^{n}(x_i - \bar{x})^2 > 0$$