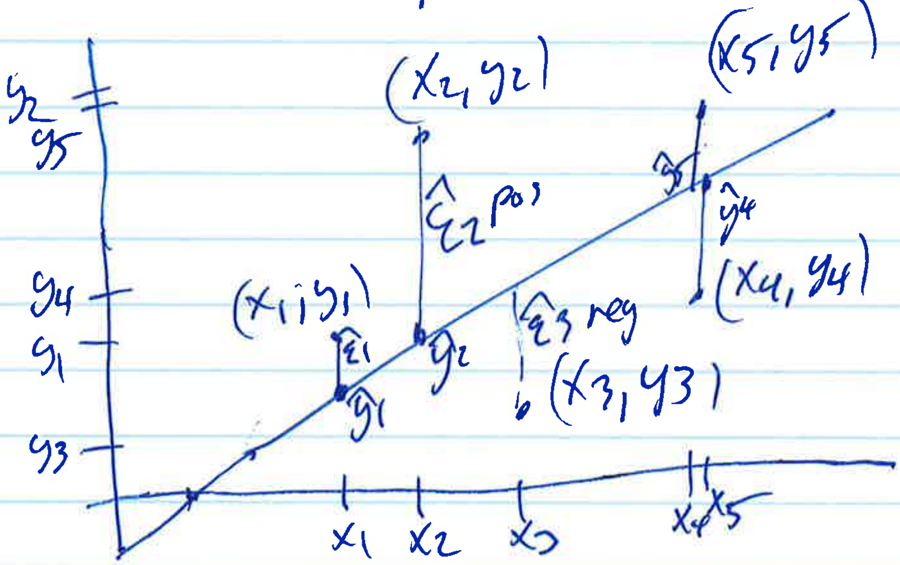
Review

Given data points



And the regression line $y = mx + b$ Stand notation
$y = b_0 + b_1 X$ Stat notation

the explanatory variable is x
the response variable is y

The predicted y_i or $\hat{y}_i$ is what is predicted from the regression line

$$\hat{y}_i = m x_i + b$$ Standard notation

$$\hat{y}_i = b_0 + b_1 X_i$$ Stat notation

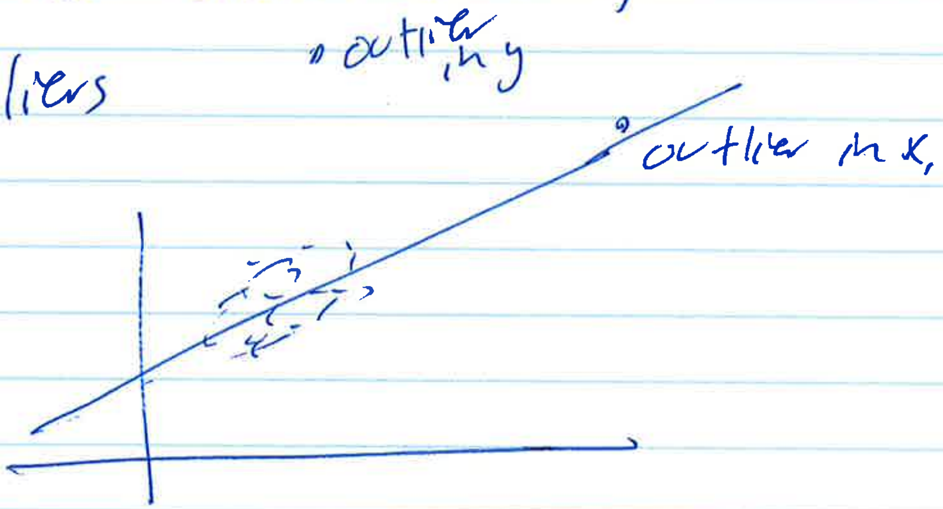
$$\varepsilon_i = y_i - \hat{y}_i$$ residuals

Residual plot – plot residuals against explanatory variable

Should be no pattern if regression line captures overall pattern of data

(Specifically an unstructured horizontal band centered at zero)

Outliers



Influential – Removing obs has a large effect.

Outliers in x tend to be influential

~~Outliers in y too, I think~~

Fact

$$r^2 = \frac{\text{variance of predicted values } y'}{\text{variance of observed values } y}$$

This is what is meant by saying

r^2 is the fraction of the variance in the values of y that is explained by the regression line.

Stat 202 2015S - W7 - Wed

§2.6 The Question of Causation

New

In many studies the goal is to establish that changes to the explanatory variable. Cause changes to the response variabee

Famous saying: Correlation does not imply causation

More basic question: What causes association?

Consider the following observed associations between x and y

1. x: mother's body mass index
   y: daughter's body mass index
2: x: amount of artifical Sweetener (saccharin) in rats diet.
   y: count of tumors in rats bladder
3. x: a student's SAT score as high school senior
   y: a student's first year GPA
4. x: monthly flow of money into stock mutual funds
   y: monthly rate of return for the stock market
5: x: whether a person regularly attends religious services
   y: how long a person lives
6: x: number of years of education a worker has
   y: the worker's income

Links between variables that can cause association

⟶ solid arrow denotes direct causation

⇠⇢ dashed arrow denotes assiation

possibility 1

One possibility is direct causation

X causes y

According to book items 1 and 2 are examples of direct causation

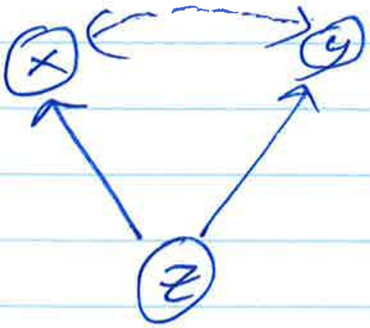
But even when direct causation is present very often it is not the complete explanation of an association

To show x cause y best evidence is to change x hold all other factors fixed and observe y. If y changes we have good reason to think x causes change in y.

Common response

Possibility 2

x and y show a common response response to a third (often lurking) variable z



Items 3 and 4 are exampls of a common response.

Item 3: Student's aptitude
Item 4: Economy

No ~~direct~~ causal link between x and y but association none the less possibility 3

Confounded

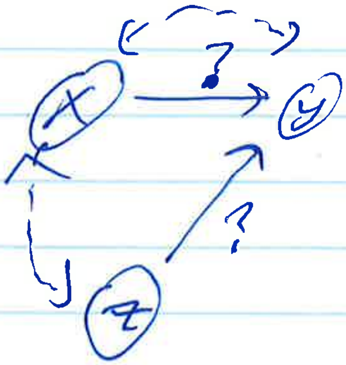
Two variables are confounded when their effects on a response variable cannot be distinguished from each other

The confounded variables may be explanatory variables or may be lurking variables.

Eg Mother / Daughter BMI
Nature vs. Nurture.

When many uncontrolled variables are related to a response variables ask whether confounding prevents you from drawing conclusions about causation

Items 5 and 6 are confounding



How to establish a causal connection between x and y

Only compelling method - carefully designed experiment in which effects of all possible lurking variables are controlled

Needless to say that it is not always possible to do such an experiment

EG: For ethical reasons it is not possible to run an experiment where we force people to smoke.

That would give the best evidence however that smoking causes cancer.

Nevertheless it has been agreed upon that smoking does cause cancer

How? By what Criteria

What criteria allow us to conclude smoking causes cancer absent controlled studies

* The association is strong between smoking and lung cancer
* The association is consistent. Across country and different groups (reduces change that a lurking which might be specific to one group causes association)
* Higher does associated with strong response

x Alleged cause precedes effect in time

x Alleged cause is plausible - experiment in rats show smoke causes cancer,