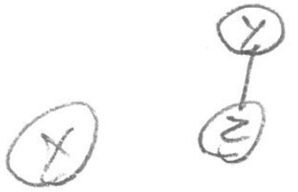
Math 340.03

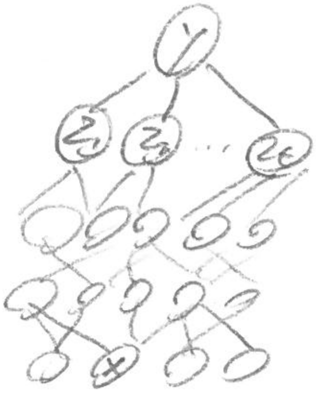
Lec 25 5/14/18

A



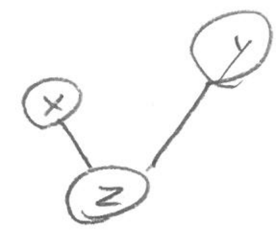
No corr & No caus

B



Corr & Caus

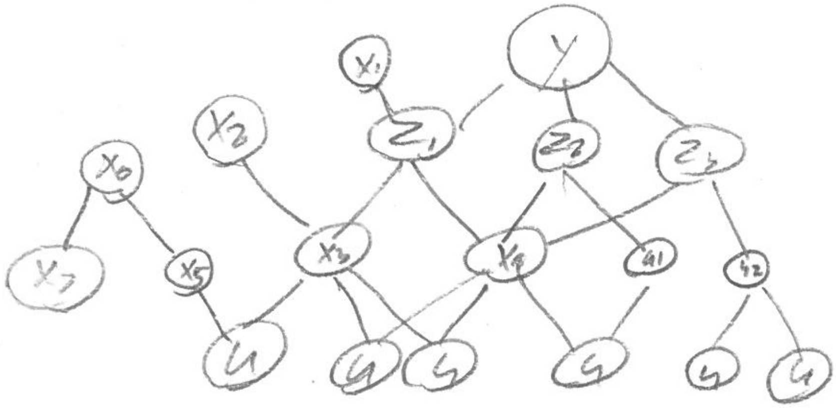
C

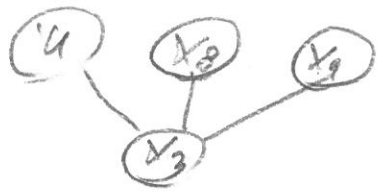


Corr & No Caus.

A common definition:
Causality means if you manipulate X (and nothing else), you will see a change in Y. Which var's below are causal? Corr?

The u's you don't see.





m & Caus: x_3, x_4

nly Corr: x_1, x_5, x_6

Neither: x_7, x_8, x_9

C x: # umbrellas sold by noon

y: # car accidents from 12:00AM → 11:59pm

z: precipitation for the day

Z is the unobserved common cause. It is also called a "lurking variable" because it is (a) unobserved and (b) affects the relationship of x and y when accounted for.

We will see this in a linear regression soon. First, let's go back to

Let's return to the linear model. After the OLS algorithm:

$$g(x) = b_0 + b_1 x$$

What is the interpretation of b_1? Everyone screws this up! Even me! Gotta be very careful

"If x is increased by one unit on its measurement scale, then y is predicted to increase by b_1 units on avg. on its measurement scale assuming the linear model" opt. for least squares.

Very subtle language! Notice the passive voice "is increased". By who?

We amend this to...

"When comparing two observations (A & B) sampled in the same way as the data in $\mathbb{D}$ and where (A) has an x value one unit larger than the x value of (B), (A) is predicted to have a response y_A that differs by b_1 on avg. from y_B assuming the linear model."

This is the best you can do! It's very lame!!

Note: "naturally observed" means unwatched.

Can you say

"If I increase the x value for the obs., its predicted response will increase by b_1 (ass. the lin. mod.)"? Note: this is a causal statement. You are saying due to your manipulation of the variable, there will be a change in pred y.

Usually not ... unless you have convinced the picture looks like (B).

There is no way to prove where x falls in the causal chain without more work. The gold std. to be able to make a causal claim is the randomized controlled experiment.

Let's interpret multivariate linear models:

$$g(\vec{x}) = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p$$

What is the interpretation of b_1?

"ceteris paribus"

"When comparing two observations A & B sampled in the same way as the data in $\mathbb{D}$ where A has an x_1 value one unit higher than B and <u>all other x_j's are exactly the same</u>, A is predicted to have a response y_A that differs from y_B by b_1 on avg. assuming the linear model optimized for least squares!"

the only thing that's different in the interpretation from the simple linear regression.

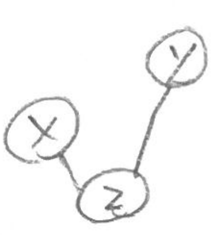
A real thing wrong w/ this.

Imagine x_1: overall GPA, x_2: major GPA

Can you increase x_1 keeping x_2 constant. Yes... but it is weird since x_1, x_2 are <u>highly</u> correlated.

So you can make the statement about b_1, but it may not even be a realistic statement because A and/or B may not be observable in the real world.

Here's where correlation does <u>not</u> imply causation (C)



$y \sim x \Rightarrow$ see an effect of x on y : b_1 is positive

$\Rightarrow y = b_0 + b_1 x$

$y \sim x + z \Rightarrow$ do <u>not</u> see an effect of x on y : b_1 is near zero

$\Rightarrow y = b_0 + b_1 x + b_2 z$

Why? What's the interpretation of b_1? .... If z <u>stays constant</u> and x changes: y doesn't change. Look at the diagram! Z causes y and meanwhile x doesn't have any effect on y!!!

DEMO

$\boxed{\text{END of MATERIAL FOR EXAM}}$

Inference for linear models

$y = h^*(\vec{x}) + \varepsilon$

In the linear model case, $\mathcal{H} = \{\vec{w}\cdot\vec{x} : \vec{w} \in \mathbb{R}^{p+1}\}$

$\Rightarrow y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon$ where the β's are the "true" best weights

and $y = g(\vec{x}) + e = b_0 + b_1 x_1 + \dots + b_p x_p$. We want to see how close b_j's are to β_j's and be able to test if $\beta_j = 0$.

Now make the four lin. regr. assumptions. Need prob. model!!!

(I) $h^* = f$. That is, the true model is linear.

(II) independence of epsilons $\rightarrow$ iid

$$\varepsilon_1, \dots, \varepsilon_n \overset{iid}{\sim} N(0, \sigma^2)$$

(IV) Normality of epsilons

(III) homoskedasticity of epsilons

$\Rightarrow$ Awesome theorems

(1) $\text{Var}[\vec{b}] = \sigma^2 (X^\top X)^{-1} = \begin{bmatrix} s_{b_0}^2 & & & \\ & s_{b_1}^2 & & \\ & & \ddots & \\ & & & s_{b_p}^2 \end{bmatrix}$ Standard errors for each LS weight!!

(2) $\dfrac{b_j}{s_{b_j}} \sim T_{n-(p+1)}$ Student's T distr. $\Rightarrow$ t-tests for $H_0: \beta_j = 0$

i.e. x_j does not affect y linearly

i.e. any signal in x_j?

all equiv.

(3) $\dfrac{\frac{SSR}{(p+1)-1}}{\frac{SSE}{n-(p+1)}} \sim F_{p+1,\, n-(p+1)}$ Fisher's F distr. $\Rightarrow$ omnibus F-test for $H_0: \forall \beta_j = 0$

i.e. No variable affects y linearly

i.e. any signal at all?

(4) $\dfrac{\frac{\Delta SSR}{\Delta p}}{\frac{SSE_{FULL}}{n-(p+1)}} \sim F_{\Delta p,\, n-(p+1)}$ between a full model and a partial model

F-tests for $H_0: \beta_j = 0$ & $\beta_k = 0$ & ... whatever you wish to test

i.e. any signal in a collection of variables? (demo)

(5) $\hat{Y}^* \approx N(\vec{x}^* \cdot \vec{b}, RMSE^2)$

empirical rule we spoke about

& MORE!!!