Dec 1

AP Stat

Objective: Students will explore non linear relationships between Bivariate data.

1. Return Test
2. Activity- Cancer and dice rolling- NOT ENOUGH PEOPLE TODAY
3. Notes recap: pg 259-265
4. HW: see calendar.

Note, some items may be changing on this units calendar.....just be aware.

Chapter 4

More Relationships between data

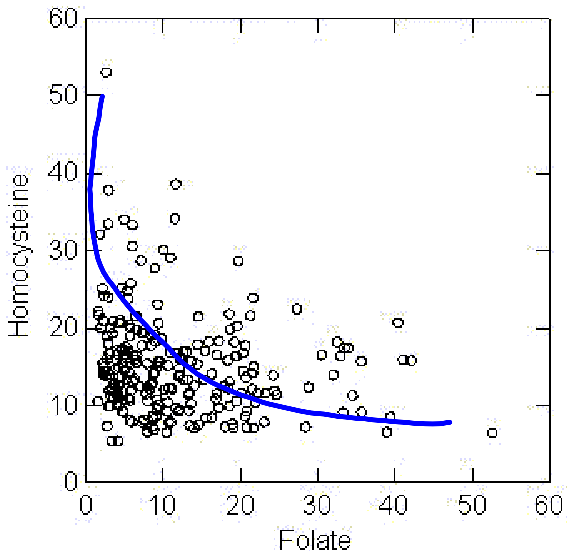
Transforming Data (Re-expressing the data): applying an alternative function to a quantitative variable (log, exponential, square root)-WE do this to use later on in hypothosis testing- which ONLY applies to LINEAR data

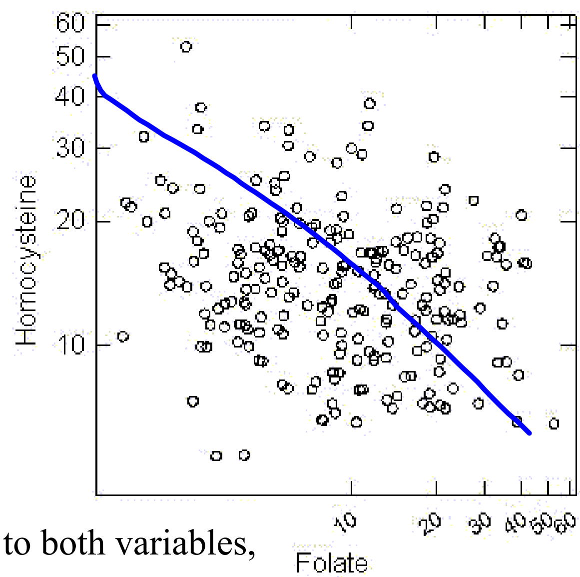
Linear Transformation is more of a change in units. Inches to feet, meters to yards, Celsius to Fahrenheit, etc. (changing the units doesn't change the curve in a relation- to make it linear)

To change the curve, you must use functions that are non linear (log, exponential/power, neg power, sq root, most common use)

There are many reasons to transform data as part of a regression analysis.

- to achieve linearity.
- to achieve homogeneity of variance, that is, constant variance about the regression equation.
- to achieve normality or, at least, symmetry about the regression equation.





When logarithmic transformations are applied to both variables, the distributions of the individual variables are less skewed and their joint distributions is roughly ellipsoidal.
A straight line seem a like reasonable candidate for describing the association between the variables and the variances appear to be roughly constant about the line.

~~Example- together~~....book problem 4.2 pg 265

skip on HW

Also- review/know chart pg 268

Tues 12/5

AP STAT

Objective(s):

1. Students will transform a curved scatter plot using log/ln
2. Students will use the new regression to find predicted values

1. return (remaining) assessment
2. Notes- widdled down to what you need to know- and practice included- CALC
3. Handout- finish for HW

NO NOTES QUIZ tomorrow

Some Nuts and Bolts of Transforming Scatter Plots: (trying to make it a little less complicated)

- Typically we use Ln or Log to transform the scatter plot. Others can be used.
- Determining which one will depend on what you try and see how it works.
- Check the residual plot- it can tell you how well it has straightened out your graph. (by how curved it may appear.
- If its exponentially growing/decaying, log may be the ticket. You can try Ln too. (*should b= 2.7(ish) on exponential regression) (I have found both do pretty mucht the same thing- a and b values migh the different in the equation but the output is same)*

Strategy:

- not all non-linear relationships are exact exponential growth or decay relationships.

- If there is still some curvature in the residual plot. This indicates our log transformation didn't work at completely straightening the relationship. Here is a strategy to use when deciding which variable to transform:

1. Try transforming the response variable . If this helps in satisfying the conditions necessary for doing a regression analysis, then proceed with the analysis. If not, try step 2.

2. Try transforming only the explanatory variable. If this helps in satisfying the conditions necessary for doing a regression analysis, then proceed with the analysis. If not, try step 3.

3. Try transforming both the response and explanatory variables. If this helps in satisfying the conditions necessary for doing a regression analysis, then proceed with the analysis.

4. If none of the above worked, then use what looked the best realizing that there are other methods beyond the scope of this class that may have worked.

Note: if one of the above transformations doesn't look any better than the others, you may just want to stick with analyzing the data on the original scale and comment on your analysis.

Do example- handout

Summary: Exponential vs. Power

- If the relationship is exponential then the plot of the log (x) versus y should be roughly linear (or x vs log y). If the relationship between these variables follows a power model, then a plot of log (x) vs. log (y) should be fairly linear.
- **In an exponential model you are transforming one of the variables (usually the response). In a power model you are transforming both.**
- Our eyes are a bad judge of curves so we need to do both transformations, make a scatter plot of each, and compare the residual plot and r values of each to see which did a better job of linearizing the data.
- We can fit exponential growth and power models to data by finding the least-squares regression line for the transformed data, then doing the inverse transformation

Summary of what you need to know

- When data doesn't look straight, try both transformations: (x,y) to (x, logy) or (x, lny) and (logx, logy) or (lnx, lny)- log and natural log are both fine!

- Check which transformation did a better job straightening:
  - Make a scatterplot of each transformation. Do LinReg a+bx to check your r for each. The stronger the r, the better.
  - Also do a residual plot for each transformation to see which better fits the data
  - If your first transformation was better than it's an underlying exponential function fitting your data. If the second transformation was better than it's a power model.
  - Find the regression equation for your original untransformed data

Residual Handout- review together

CHANGE IN HW/CALENDAR

HW: Practice problem in handout #36,39,41,42 plus Frappy problem

No Notes quiz. CHECK FOR UNDERSTANING QUIZ FRIDAY

M and M activity tomorrow

12/6

AP STAT

Objective: *Students will transform graphs to linear models.*

1. check/rev HW

2. M and M Activity

SKILLS QUIZ FRIDAY- short short short quiz

3. HW: Read through (take notes) on sections 4.2 in text.

12/8

AP STAT

1. Questions
2. *CFU- 15 min*
3. Notes on Two way tables- see handout
4. HW- do attached problems and read through section 4.3

NO NOTES QUIZ MONAY

Two way tables

Categorical Data

– if column and row totals are missing- then first calulate those totals. These are **the <u>MARGINAL DISTRIBUTIONS</u>** because they are in the bottom and right margins. Calculating these will help make it easier finding the %s

-percents are often more informative than the counts.

- the marginal distributions alone tell is nothing really about the relationship between the categorical variables (depending on what you are determining)

-Conditional Distribution: "isolated" category

EX:

Calculate Marginal Distributions then Calculate the conditional distribution of responses fo the males and females (what should your bottom margin add too?)

Superpower		Male		Female	
Invisibility	17/30	17	13/30	13	30
Superstrength	3/20	3	17/20	17	20
Telepathy		39		5	44
Fly		36		18	54
Freeze Time		20		32	52
		115		85	200

Which Conditional Distributions should be compare?

- This is determined by what you are looking for.
- Which gives you what you want
- Think about how changes in one variable might explain the other.

ASSOCIATION: We say there is an association between two variables if specific values of one variable tend to occur in common with specific values of another. (Warning- be cautious of lurking variable)

The titanic disaster:

Class of Travel	Survived	Died
First class	197	122
Second Class	94	167
Third class	151	476

But heres another table

	Gender			
	Female		Male	
Class of Travel	Survived	Died	Survived	Died
First Class	97% 140	3% 4	32% 57	118 65%
Second Class	86% 80	13	8% 14	154
Third Class	46% 76	89	75 16%	387

In the movie "Titanic" It suggests that Woman and children boarded life boats first and that passengers in first class were given special treatment.

1. What do the data tell us about these two suggestions?

2. How does gender effect the relationship between class of travel and survival status? EXPLAIN

SIMPSON PARADOX

An association between two variables that holds for each individual value of a third vaiable can be changed or even reversed when the data of the third variable are combined.

EX: Do medical helicopters save more lives

When you adding the third variable- seriousness of accident, it changed what appeared to be true on the surface. - different percentages to consider.

The seriousness of the accident was a LURKING VARIABLE.-which- until the details were broken down a bit better was "hidden" and harder to interpret.

12/11

AP STAT

1. go over HW
2. Review of notes 4.3- handout
3. HW: pg 312-313 #41-48

12/12

AP Stat

Objective: *Students will apply concepts from chapter 4 (and 3)*

1. warm up- hand out

2. GO over HW answers

3. AP Practice set- project grade. (assigned partner or no partner- option).

Causation

Even when direct causation is present, it is rarely a complete explanation of the association of the association between the two variables

-Even experimentation that shows a strong association/causation in an animal does not mean it carries over to human (rats and saccharin)

- Well controlled experiments will determine causation - as discussed.

Common Response:

A correlation being impacted by something in common:

EX. "good students" tend to have higher SAT math and higher SAT verbal scores. In general, Good students tend to do better than poor students in acedemic measures

Lurking variables: not among the explanatory or response but influences the interpretation of the relationship between the Exp and Res vars

A LURKING variable is one whose effects on the response variable cannot be distinguished from one or more of the explanatory variables in the study, AND IS NOT INCORPORATED INTO THE DESIGN OF THE STUDY. (typically related to observational studies)

Confounding variables- can be lurking. Their effect on the RESPONSE variable can not be distinguished from each other. What lelse could have caused that response?

A confounding variable is one whose effects on the response variable cannot be distinguished from one or more of the explanatory variables in the study.

THE DIFFERENCE

If a variable was measured and included, it's associations between the explanatory and response variables can be determined and (if random assignment was performed) neutralized with methods beyond the AP Syllabus. It is a confounding (or not) variable.

The associations between an unmeasured variable and the explanatory and response variables cannot be determined -- whatever its associations are remain a mystery, and it "lurks" beyond the purview of the investigator.

Confounding refers to a problem that can arise in an experiment, when there is another variable that may effect the response and is in some way tied together with the factor under investigation, leaving us unable to tell which of the two variables (or perhaps some interaction) caused the observed response. For example, we plant tomatoes in a garden that's half-shaded. We test a fertilizer by putting it on the plants in the sun and apply none to the shaded plants. Months later the fertilized plants bear more and better tomatoes. Why? Well, maybe it's the fertilizer, maybe it's the sun, maybe we need both. We're unable to conclude that the fertilizer works because any effect of fertilizer is confounded with any effect of the extra sunshine.

Note that this is why we need to blind some experiments: we don't want knowledge of the treatments on the part of those involved in the experiment itself or evaluating the response to be confounded with any actual effect of the treatments under investigation.

Lurking variables are sometime referred to as "common response". That's where some other variable drives each of the two variables under investigation, making it appear that there's some association between those two variables. A common example is in the strong association between the number of firefighters who respond to a fire and the amount of damage done. One shouldn't conclude that the firefighters may be responsible for the damage; the lurking variable is the size of the fire.

Lurking variables are the risk we face in sampling and observational studies. In an experiment, though, the factor under consideration isn't being driven by some lurking variable, because we are the ones in control there.

time line:

Today- variables
Tomorrow and part of Wed- AP practice
Wed- practice probs
Friday- Review
Monday- MC part of test
Tuesday- open ended part of test
Ch 6 starts Wed....

12/13

AP Stat

Objective: *Students will apply topics on scatter plots and residuals to solve problems*

BIG PACKET- SKIP wksht 4.3 page

1. Warm up- hand out
2. AP practice set- FINISH TODAY
3. Next packet available.

12/15

1. Warm up- handout
2. Work on Packet- check answers on webpage
3. finish packet- study for Test