Nov 13

AP STAT

1. Check/rev HW
2. Review/recap of notes
3. HW: pg 179-184 #5,7,8,9,11 and read/notes pg 185-188

Chapter 3

Notes Review

Exploring relationships between two variables. ***BIVARIATE DATA***

Is there a relationship and if so what is it?

Categorical variables can now play a role

Review: Response Variable and Explanatory Variable

Response Var: Measures the outcome- *dependant variables*

Explanatory Var: What's being studied, what do we want explained, what explains the respnse variable.-*independent vairables.*

Though Independent and dependent mean something unrelated to this in statistics- which is why we don't really use those terms here.

Ex's: Alcohol and body temp. and Math SAT vs Verbal SAT- whas there a relationship. Can you predict math score if you know the verbal score.

Scatter Plots and Correlation

(do not consider the outlier rule)

A **SCATTER PLOT** is the most effective way to display a relationship between quantitative data

- Be sure to label axes.
- Scale horizontal and vertical axes in uniform intervals
- Explanatory var is usally x, response is usually y
- Outliers are usually deviations that stand out from the rest of the data.
- Categoriacal data/info can be displayed with a different color or mark to determine if a pattern shows up (acts sort of like a third variable)
- can be used to make predictions based on pattern

Directions/Associations

- Positive, negative, or none

Form

- Are there Clusters

Strength

- How close do they follow that form.

Correlation

The correlation coefficient "r"

- Measures the strength of the relationship between the qualitative data
- $-1 \leq r \leq 1$
- Closer it is to 1 or -1, the stronger the relationship.
- pos/neg is directional.
- Non resistant
- not a complete summary of the relationship- its a start.

Example pg 186: body weight vs backpack weight

Body weight(lbs):	120	187	109	103	131	165	158	116
Backpack weight (lbs):	26	30	26	24	29	35	31	28

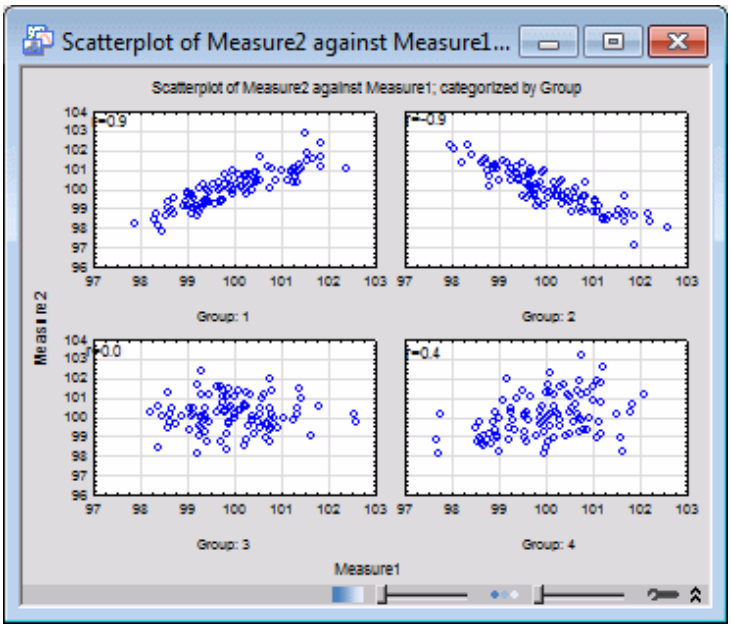
L_1

L_2

Make a scatter plot and determine r.

Diagnostics on

See catalog button


Scatterplot of Measure2 against Measure1...
Scatterplot of Measure2 against Measure1; categorized by Group
r=0.9
r=-0.9
r=0.0
r=-0.4
Group: 1
Group: 2
Group: 3
Group: 4
Measure2
Measure1

AP STAT

Nov 14

1. Check/Review HW/compare with neighbor- what Qs do you have- whole group....(10 min)

2. Activity with scatter plot- using a scatter plot to solve a mystery

3. **HW:** pg 193-197 #15,16,19,21,22,24 read and take notes 199-203 ***(notes quiz)***

mult 2.54 to convert- in to cm

Nov 15

AP STAT

objective: students will write regression and least squares regression equations to make predictions.

1. Notes quiz-HKREILLY

2. check and rev HW probs

3. notes review

4. least sq regression equation (notes)

5: HW : pg 204-205 #29,31 pg 211-213 #33,36,37

pg199-203

Regression Line (our eyeballed line)

- Regression requires an explanatory variable and response variable.
- a line that describes how a response variable changes as an explanatory variable changes
- makes predictions
- mathematical model
- use y=a+bx
- slope doesn't tell you how important a relationship is

our M &M mystery

Extrapolation

using the regression equation to make predictions outside the scope of the data set/range. Usually inaccurate.

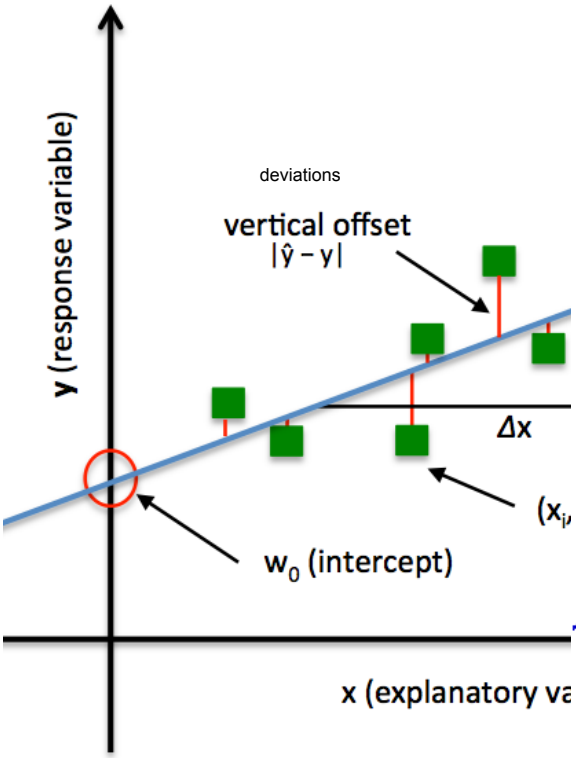
Interpolation

using regression line to make predictions within a range of data

Least Squares Regression Line (LSRL)

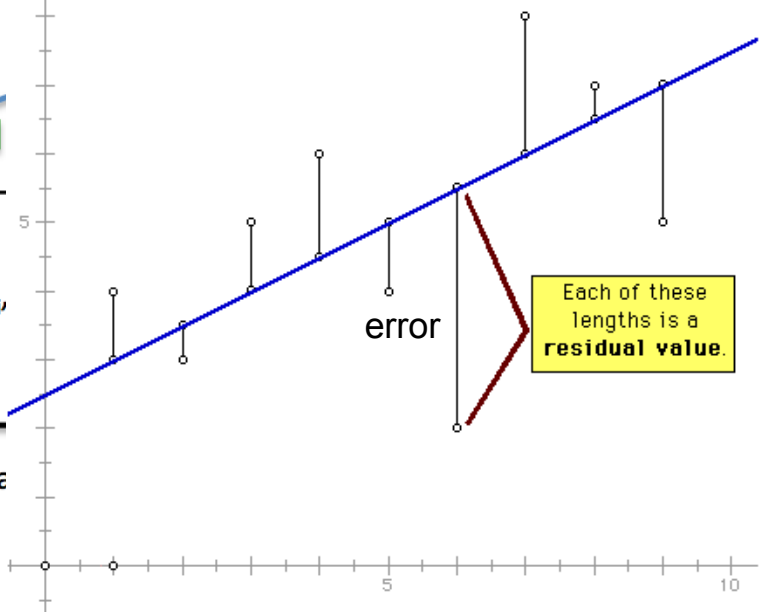
- this is the line that makes the sum of the vertical (y) distances of data points from the line as small as possible (a "true" best fit)
- a way to get a regression line that does not depend on a guess (eyeball)
- since we use form y= a + bx for LRL, use same format but for least squares regression we will use y(hat) or $\hat{y} = a + bx$ LSRL
  - « where $b = r(S_y/S_x)$ and $a = \bar{y} - b\bar{x}$

y hat usually not exact as the actual obser
but still provides good info.



$\hat{y} = w_0 + w_1 x$

Each **residual** is calculated as **the difference between the actual y value and the predicted y value**, where the predicted value is the y-coordinate of the point on the line. If a data point is below the prediction line, its residual is negative, and if a point is above the line, its residual is positive.



Let's write a LSRL

Given:

x(mean)= 19.65

Sx= 2.14

y(mean) = 169.29

Sy=10.69

r=.705

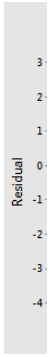
(from eyeball line).

In calc

1. data- use mystery data. input into L1 and L2

2. Stat>Calc>#8> calculate

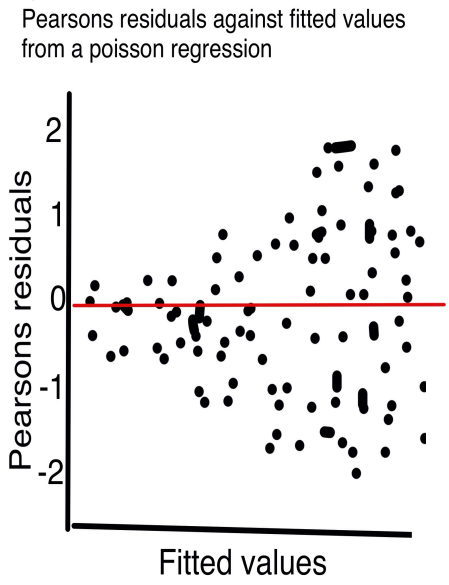
Looking Ahead: Residual plots



being able to see the shape of the residuals is also telling about our prediction equation.

the residual plot is a scatter plot of the residuals against the explanatory var(x/L1). Keeping in mind that the sum of the residuals should be zero or close to zero, the "trend" should be a horizontal line (slope=0).

the residual plot tells us how well our regression equation fits our data.



If fanning of your residuals happens to take place, it indicates that your predicted y values (y hat) will be more inaccurate as your x increases.

and vice versa

Lets make one:

use list from M & M

1. L3: highlight it and type Y1(L1) enter- it will fill your column

2. L4: highlight it and type L2-L3

3. now turn plot 1 off and turn plot 2 on nad select scatter plot with L1(x) and L4(y)

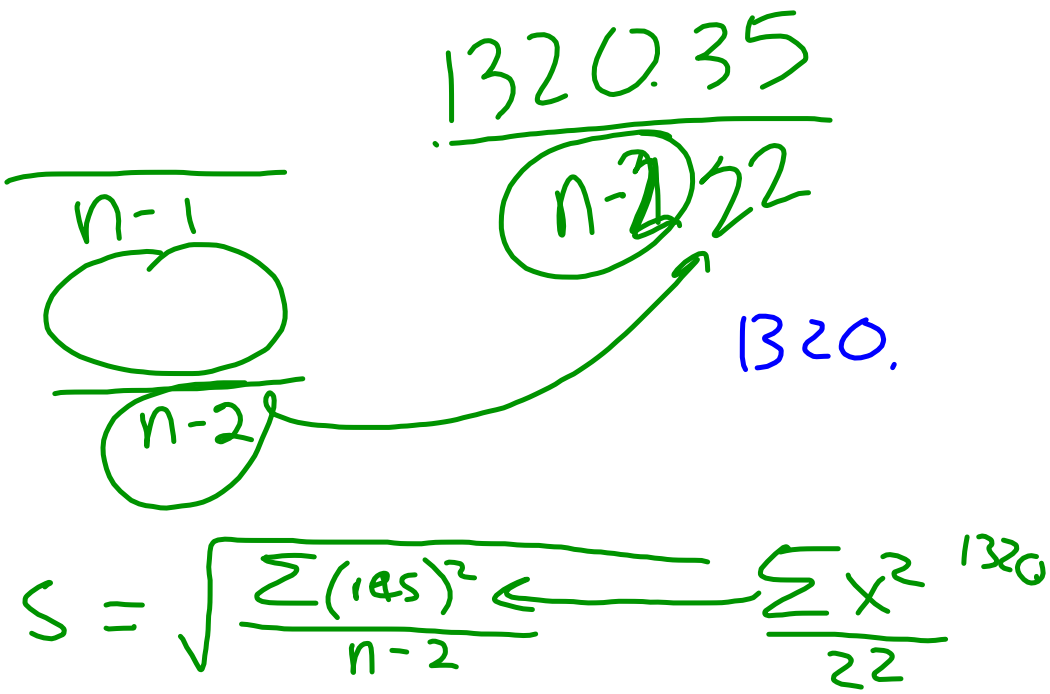
4. Turn off your Y= equations

5. Zoom #9 (zoom stat)

Typical prediction ERROR

we use the standard deviation of the residuals.

use 1-var stats- wait for teacher to explain what to do.



Monday 11/20

AP STAT

Objective: *Students will use the coefficient of determination to describe LSRL and variation*.

1. NOTES QUIZ: HKREILLY

2. Check/rev HW- questions?

3. Review of notes

4. HW: pg 227-229 #44,48 Pg 230-233 #49,51,52,53,58

QUIZ WEDNESDAY

AP STAT

Notes

The coefficient of determination r^2

R sq is our correlation coefficient r.....squared.

This value essentially describes the percent variation that is due to the LSRL that we use to make predictions.

The higher the value- the better....

For example:

if $r^2 = .82$, then 82% of the variation is due to the given regression equation. The remaining "variation" of 18% would be due to the individual or subject. 82% is predictable from equation. The other 18%is variable for some other reason

if $r^2 = .60$ that means our equation is relating the data "well" only about 60 percent of the time. the remaining 40 percent is individual which may say the LSRL isn't too great at helping make prediction or relating the two variables

Caution:

Association is not causation.

That is, just because a data set is characterized by having a large r-squared value, it does not imply that x causes the changes in y.

4 Facts about Least squares regression

1. The distinction between explanatory and response variables is essential

2. There is a close connection between correlation and slope of the least squares regression line

3. The least squares regression line always passes through $(\bar{x},\bar{y})$

4. r^2 is the fraction of variation in the values of y that is explained by the least squares regression of y on x.

other stuff:

- Since r^2 is a proportion, it is always a number between 0 and 1.
- If $r^2 = 1$, all of the data points fall perfectly on the regression line. The predictor x accounts for all of the variation in y!
- If $r^2 = 0$, the estimated regression line is perfectly horizontal. The predictor x accounts for none of the variation in y!

Nov 21

AP STAT

Objective: *Students will review topics in scatterplots, correlation, LSRL, and r-sq*

1. check HW and go over any questions

2. handout- practice for quiz- key will be on webpage

3. Be aware that the packet is not the end all- review all of sections 3.1,3.2- you will need to EXPLAIN answers on quiz

Nov 23

AP Stat

7:25-8:25

10:55-11:55

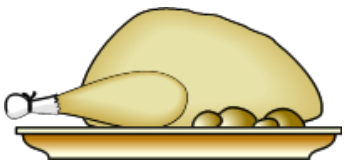
QUIZ! Pencil, calc

spread out

Have a great Thanksgiving weekend!

There is HW

on your calendar



Nov 27

AP Stat

TEST WEDNESDAY 11/29

1. Section 3.3 Recap

2. Practice: pg 238 #62, pg 244 #70,72

Section 3.3

Correlation and Regression Wisdom

-correlation and regression describe linear relationships

-Extrapolation- unreliable predictions, predictions make no sense

-Correlation- nonresistant (bc of SD- if points added changes SD)

Outliers and Influential Observations:Subjective: how one point can change things LSRL or r. the graph/visual is important.

Outlier:

- in real data gathering-check for error in collection/testing
- not all outliers are influential
- outliers are unusual data: in y direction, large residuals

  in x direction will not have large residuals

Influential Data:

- influential data will have small residuals on x-not noticeable-thus scatter plot is crucial.
- pulls LSRL toward them (in x) so will not be noticeable on the residual plot- this means it will increase y intercept of line and decrease slope
- data point is influential if removing it changes the results in calculations
- If questionable point, find LSRL for both (with point and without point)- if changes more than slightly, then the point is influential

Making Decisions: Keep or throw away

may need more data for that influential data or decide to exclude it.

(Slow talker example)

Lurking Variables:

- can make correlation and regression misleading
- not among explanatory or response variables but have influence in interpreting relationship between the initial variables/testing variables
- Lurking var can also show a nonsense relationship
- Lurking var can also disguise a relationship

Remember Association does not imply causation.

Other Odds and ends:

remember that correlation (r) is based on z scores- which are already standardized. Therefore, changing units of measure in an equation does not change the correlation

Review:

pg 250-255

#77,78,79,85