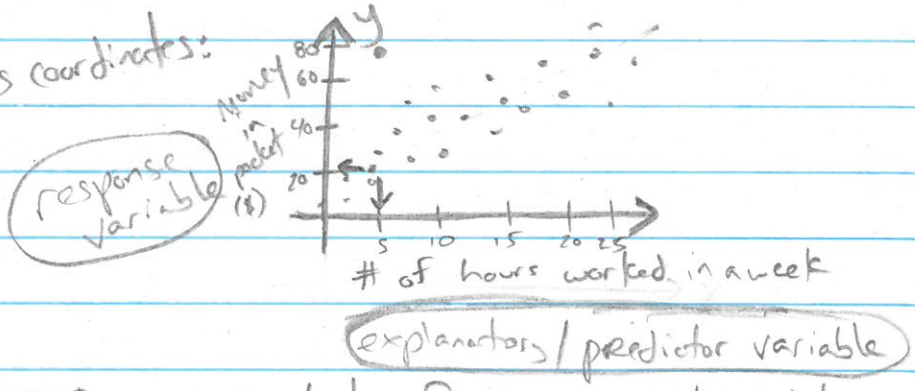
★ Watch Chapter 7 Videos

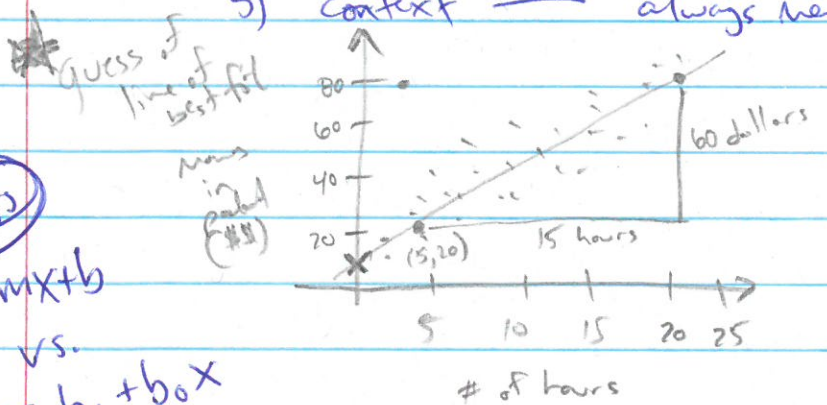
Chapter 7: Scatterplots, Association, & Correlation.

- We start to discuss the relationship between two quantitative variables. Looking for some sort of association between the two. USING SCATTERPLOT

uses coordinates:



- 5 Things to look for in a scatterplot:
  1) direction - positive, negative, or neither
  2) form - straight or curved
  3) strength - tight (strong) or scattered (weak)
  * 4) outliers - don't fit the data.
  5) context — always mention (trend) "It appears that"



$\text{slope} = \frac{\text{rise}}{\text{run}} = \frac{60}{15} = 4$

$y = mx + b$

money = 4 (hours) + b

money = 4 (hours) + 10

$y = 4x + 10$

STATS

$y = mx + b$

vs.

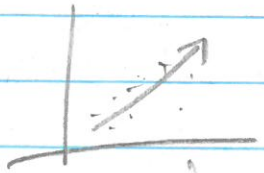
$\hat{y} = b_1 + b_0 x$

Class: Say I worked 30 hours, how much money would we expect me to have given our model?

$4(30) + 10 = \$130$

* Again, things to look for:

Direction



positive

negative

Form:

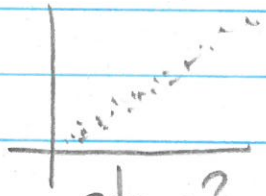
Straight?

Curved?

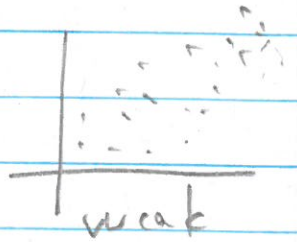
* Make sure you say ASSOCIATION, correlation is an actual number.

No Form?

Strength:



strong?



weak

Check for outliers: if it looks like an outlier, than it probably is.

Making Scatterplot on TI-84!

L_1: # of hours worked	L_2: # in packet (#)
5	20
7	25
10	32
12	40
15	61
20	80

ordered pairs

Make sure same amount of data values.

STAT PLOT → ZOOMSTAT

★ CLASS ACTIVITY: Discuss possible direction, form, strength.

Describe a possible scatterplot

group work:

- Drug Dosage & degree of pain relief
- Consumed calories & weight loss
- Hours of sleep & score on a test
- Shoe size & GPA
- Time for a mile run & age
- Age of car & cost of repairs.

wrap-up

1) Since two variables have different units, keep z-scores in mind.
2) Use association, not correlation (unless ACTUALLY talking about r)

DAY 2: 3) Collect data on heights/weights (or anything)
4) talk about how badly outliers can affect correlation.

Chapter 7 - Part II

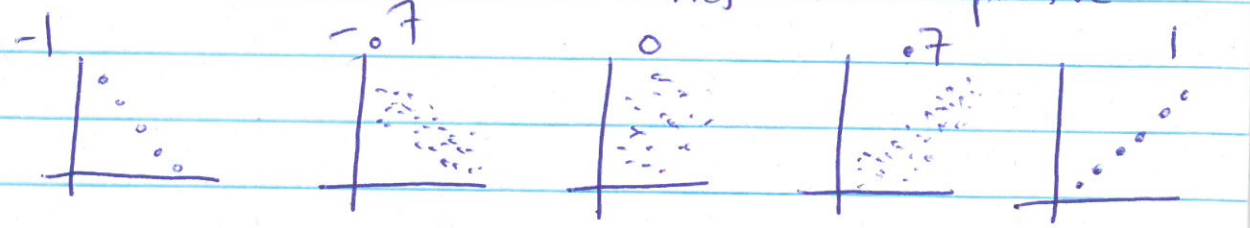
Scatterplots - ① Direction
② strength
③ Form
④ outliers (?)

Corellation - an actual calculation that measures numerically the strength of LINEAR between two quantitative variables.

- We focus on direction & strength to find our corellation coefficient (r)

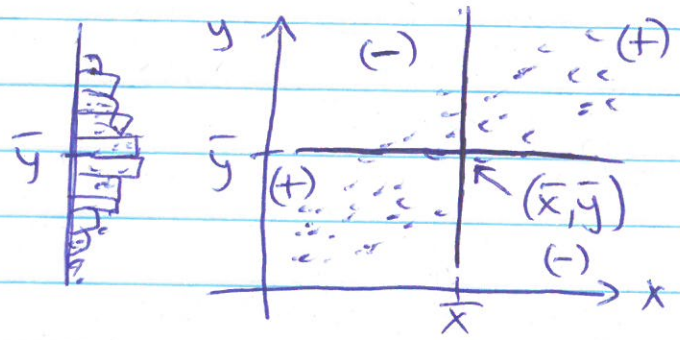
r values range from -1 to 1.

Strong negative to strong positive



STRONGER
(corellations)

How do we calculate it?



insight All values get sucked toward the axes. and fall flat.

More dots in the positive, clear positive association, calculator will calculate an actual value ~~using the sum of all~~ (r)

★ TI-TIPS:

L_1: # of hours of video games	L_2: Grade in my class
12	97
14	75
10	86
13	72
26	64
11	79
6	88
38	45
4	88
X	Y

Put them in L_1 & L_2.
Let's make a scatterplot first in STAT PLOT!

GRAPH → zoomstats.

Ask class: (looks somewhat negative, fairly strong)

- STAT → CALC
  #4: LinReg (ax+b)
  this will give line of best fit
  AND r value!

* LinReg (ax+b) L_1, L_2

$y = ax + b$
$a = -1.32$
$b = 96.81$

$y = -1.32x + 96.81$
(negative line)

$r = -.9$
NICE STRONG CORELLATION.

What does this equation mean? interpret

$\widehat{GRADE} = -1.3$ (# of hours V.G.) + 96

That means if ya play 0 hours of V.G., you should have about a 96. for each hour a week you play, your grade goes down 1.3 pts.

Important Notes about (r):

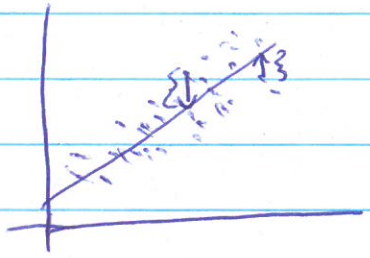
1. r can tell us direction & strength of linear relationship
2. r WON'T TELL YOU FORM!
3. r is always between -1 and 1
4. r is the same for X & Y as it is for Y & X. (switching axes)
5. r is NOT affected by shifting or scaling the data because it is based on z-scores.
6. r has NO UNITS (we don't say -.9 z-scores)
7. r is sensitive to OUTLIERS.

STOP

FORMULA:

$$r = \frac{\sum z_x z_y}{n-1}$$

Residuals - The vertical distances of a point from the line of best fit.



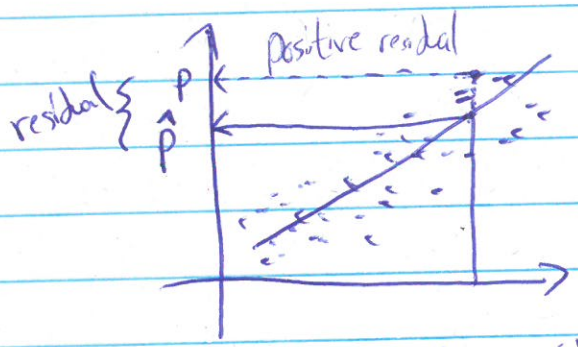
little vertical distances

The residual is the difference between actual value & value predicted by our model.

{residual = A-P} (actual +/- predicted $\hat{y}$)

$\hat{y}$ = y-hat (predicted response by our model)

- positive residual means it is over the line of best fit, ie the model under-predicted.
- negative residual means the opposite, lies below the line of best fit.



Actual - Predicted

$y - \hat{y}$

use the line of best fit to find residuals

- Go over the "What Can Go Wrong" & "What we have Learned" section.

→ Plotting regression line LinReg(ax+b), L_1, L_2, Y_1

{Vars → y-Vars → Y_1}