Ch 5

#1) AVG = 50 | a) use normal approximation:
SD = 10 | ANS: $T(1.25) = 78.87\%$

Calculate 79% of 25 (#'s in list):

$$\frac{79}{100} \times 25 \approx 19.75 \text{ (round up to } \underline{20}\text{)}$$

b) Actually 1.25 SD's?

ANS: Create an interval: $(AVG - 1.25 * SD,\ AVG + 1.25 * SD)$

$$= (50 - 1.25 \times 10,\ 50 + 1.25 \times 10)$$
$$= (37.5, 62.5)$$

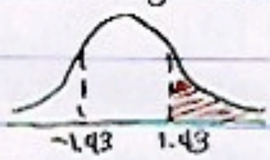
Count how many #s fall in this range: $\underline{17}$

#3 $AVG_{67} = 543$ $SD_{67} = 110$

a) Estimate the ~~#~~ % scoring over 700 in 1967

$$Z_{67} = \frac{700 - 543}{110} = 1.43$$

The table value at 1.43 gives us this 

but we need 

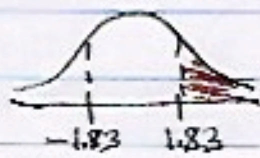
$$= \frac{100\% - T(1.43)}{2}$$

$$= \frac{100 - 84}{2} \approx 8\%$$

b) Estimate % scoring over 700 in 1994

$AVG_{94} = 499$ $SD = 110$

$$Z_{94} = \frac{700 - 499}{110} = 1.83$$



$$= \frac{100 - T(1.83)}{2}$$

$$\approx \frac{100 - 93}{2} \approx 3.5\%$$

Ch. 5

④ $AVG_M = 538$, $AVG_W = 504$ $\Big\}$ $SD = 120$

a) % of men scoring over 700 in 2005?

$$Z_M = \frac{700-538}{120} \approx 1.35$$

$$= \frac{100\% - T(1.35)}{2} \approx 8.85\%$$

About 9% scored over 700 in 2005

b) % of women scoring over 700 in 1994?

$$Z_W = \frac{700-504}{120} = 1.63$$

$$\frac{100 - T(1.63)}{2} \approx \frac{100-89}{2} \approx 5.5\%$$

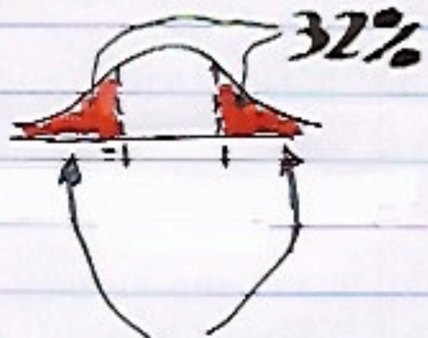
About 5.5% scored over 700 in '94.

#6 AVG = 169, SD = 9, highest score = 178

Did the LSAT scores follow Normal Curve?

ANS: Easiest way to disprove (show it doesn't ~N) is to add: AVG + 1×SD = 169 + 9 = 178, which equals the max. But on pg 68 (and something you should memorize) 68% of the data is within on SD of the avg and 32% of the data is above and below the average. if it is NORMAL

means that



32% of people should have scored above and below but 178 is max

CONCLUSION: NOT NORMAL

Ch.5 #8 a) ADDING/SUBTACTING = SHIFTING the AVG TRUE

b) SD measures spread so adding 7 to each value evenly shifts each number but doesn't affect the spread.

c) DOUBLING — True same logic as a)

D) True because all deviations from the avg are doubled.

e) True

f) When you calculate the SD you square each term so the sign is ~~~~ irrelevant.

Ch.8 #1 ANSWER: D)

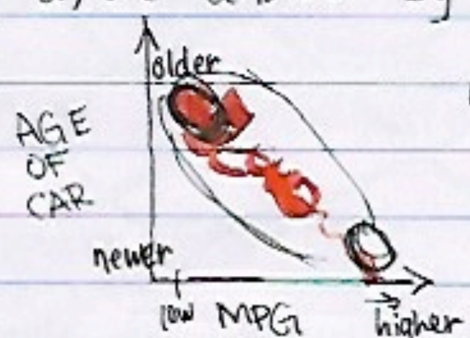
use elimination: can't be (a) because the AVGs for both are centered around 100

can't be (C) because that plot shows highly correlated and here $r \approx 0.6$ (not >0.9)

so we have either (b) or (d). Can't be (b) because the SD is too small

(D) is the logical choice.

#2 a) we answer by picturing a plot in our mind:



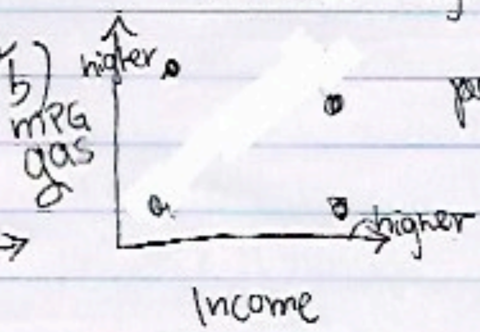
Older cars ~~use fewer miles~~/g

newer cars = more mpg

So correlation is negative

Not a good example

b)



~~positive correlation~~

What kind of cars do Higher Income people drive?

prius, fuel eff. cars

These variables might be low correlated.

Histogram Picture

What percentage of households have incomes \$7,000 to \$15,000?

Range	AREA (height x width)
7–10	10−7 = 3 = width, height = 5 $\boxed{3\times5=15}$ ARE
10–15	width = 15−10 = 5, height ≈ 5.2 (5.2×5=2
	ADD UP AREAS = 15 + 26 = 41%

What % of house holds more than \$20,000?

Range	AREA
20–25	5×2.5 = 7.5
25–50	25×0.25 = 6.25
	ADD UP AREAS: 7.5 + 6.25 = 13.75%

For more practice: problem #4 from chapter 3.

~~Problem #4 from chapter 2.~~

~~100 college students participated in a study~~

~~AVG SYSTOLIC BLOOD PRESSURE~~

~~AVG_S = 120mm with SD_S = 2.7mm~~

~~AVG DIASTOLIC BLOOD PRESSURE~~

~~AVG_D = 80 SD_D = 2.4~~

~~a) In~~

Converting:

Example: AVG_W = 140 lbs SD_W = 27 lbs

AVG_H = 60 inches SD_H = 2.4 inches

a) convert ft → inches and lbs → ounces

1 lb = 16 ounces 1 ft = 12 inches

$140\,\cancel{lbs} \times \frac{16\text{ ounces}}{1\,\cancel{lbs}} = 2240$ ounces $60\,\cancel{inches} \times \frac{1\text{ ft}}{12\,\cancel{in}} = 5$ ft

b) Replacing a value: suppose when I entered the data I listed

15 lbs instead of 150

how can I fix this to calculate the correct avg weight?

① Old $AVG_w = 140 = \frac{sum_{old}}{100}$ $\left(AVG = \frac{sum}{n}\right)$

$\Rightarrow sum_{old} = 140 \times 100 = 1400$

② Subtract the mistake

$1400 - 15 = 1385$

③ Add the ~~mistake~~ correction: $1385 + 150 = 1538$

$AVG_{New} = \frac{sum}{100} = \frac{1538}{100} = 153.8$

Calculating Correlation

~~r = avg(x) * avg(y)~~

x	1, 3, 4, 5, 7
y	5, 9, 7, 1, 13

$AVG_x = \frac{1+3+4+5+7}{5} = 4$

$AVG_y = \frac{5+4+7+1+13}{5} = 7$

We made a table

this column is product o

X	Deviation	Deviations^2	NEW * Stadardized X	Y	Dev	Dev^2	Stand Y *	Product of *s
1	-3	9	$-3/2 = -1.5$	5	-2	4	-0.5	0.75
3	-1	1	$-1/2 = -0.5$	9	2	4	0.5	-0.25
4	0	0	0	7	0	0	0	0
5	1	1	0.5	1	-6	36	-1.5	-0.75
7	3	9	1.5	13	6	36	1.5	2.25

$SD_x = \sqrt{\frac{9+1+0+1+9}{5}} = 2$ $SD_y = \sqrt{\frac{4+4+36+36}{5}} = 4$

r is an "average" of the last column

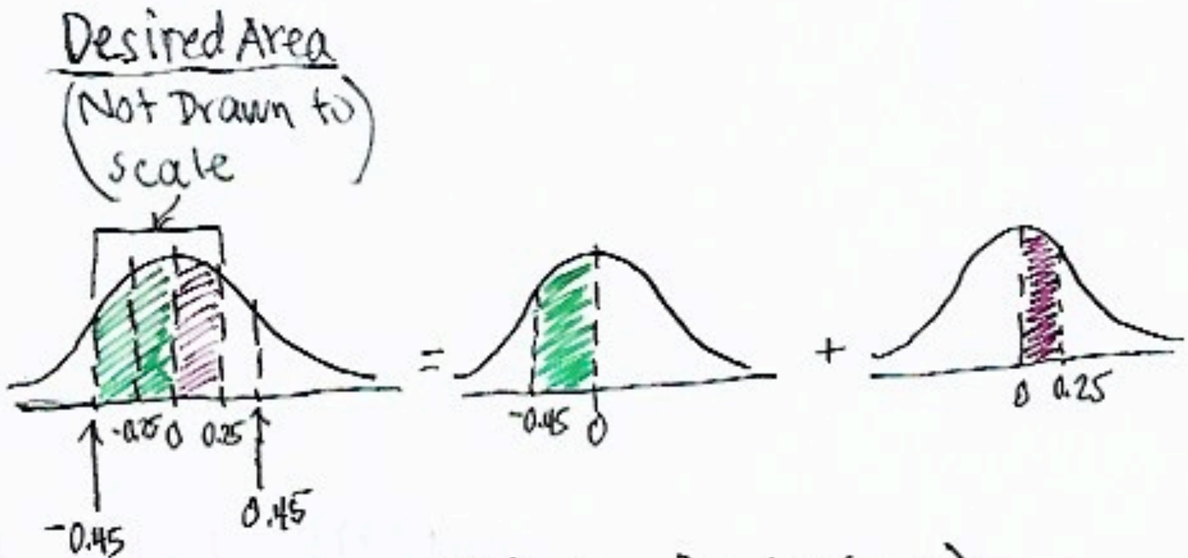
$r \approx \frac{0.75 - 0.25 + 0 - 0.75 + 2.25}{5} = 0.4$

Average watermelons exported from the US is 20 lbs with an SD of 10 lbs. (Assume Normally distributed)

Question:- Approx. what % of watermelons exported are between 15.5 lbs and 22.5 lbs?

$$z_1 = \frac{15.5-20}{10} = -0.45$$

$$z_2 = \frac{22.5-20}{10} = 0.25$$



$$= \frac{1}{2}T(0.45) + \frac{1}{2}T(0.25)$$

$$= \frac{34.73}{2} + \frac{19.74}{2} = \underline{37.105\%}$$

Thus, the % of watermelons exported from the US is about 37%.

Review

Controlled Experiment

treatment group & Control group

given drug/treatment — aren't treated

assigned at random

should be chosen as similar as poss.

- Double-blind Should be done (always best)

  neither subjects, nor doctors should know who is in the treatment group

- Salk/Polio Field Trial

  subjects: grades 1, 2, 3.

  in selected ~~sc~~ school district in US.

  2 million Kids, half were vaccinated.

  Q: Who was the control group? treatment group.

  Responses ~~a~~ were compared to see if treatment made a difference: rates at which children of T/C groups got polio.

  Possible ethical problem? who gets chosen for vaccine group who doesn't?

  Solution: use only kids whose parents consented as treatment group.

  Problem w/ this? Higher incomes tended to have higher rate of parental concent

  bias design since kids of higher incomes had higher clenliness & lower inc. had less cleanliness which made for stronger immune systems

- experiment/ T&C groups should be as similar as poss so that any difference in response could only be ~~due to~~ the result of treatment
- If the T&C groups differ due to some factor other than treatment, the effect of this other factor may be confounded (mixed up) w/ the effect of treatment.
- Confounding is a major source of bias.
- To prevent effects of family income, general health, etc from confounding the effects of the vaccine, designers used randomized control
  ↳ 50/50 "coin toss" probability for any selected individual (kid) to be ~~given a~~ assigned to treatment.
- Placebo: children in control given ~~salt water~~ saline injections
  ↳ so they wouldn't know if they were in control or treat. group
  ↳ prevent response bias
- Double-blinding: subjects nor those who evaluate responses know who is T. or C. group.
  why good? many forms of polio are hard to diagnose... (if diagnostician knew ~~tha~~ someone ~~was~~ in contr. group, might be more likely to diagnose w/ polio when not really polio or vice versa.)

Randomized-control double-blind (best but difficult to achieve