Intro Notes (cont.) 8/10/23

POPULATION
- measurements of entire group of interest
- Data from <u>all</u> quantitative variabe of all
→ the <u>avg. # of hours studied</u> by <u>all stat students in CA</u>
population

SAMPLE N = population n = sample
- part of the population
- smaller section of those being measured
- Data from... Quantitative variable of... ← write for full credit
- <u>Random and representative</u> sampling is desired
→ Data from the avg # of hours studied by 123 students

1.1 #8 HW problem!!!
a. identify the variable
□ average miles per gallon
b. Is the variable quantitative or qualitative
□ quantitative
c. what is the implied population?
□ Data from all new hybrid small cars
of average miles per gallon

- non response...? they don't answer or smt.

SRS - simple random sample ★★★

RNT - random number table (cars on campus)
1. number each car <u>001</u> - 678
2. Drop it like it's hot on a random spot
3. proceed in 1 direction and take 3 digits at a time
4. Keep anything b/w 001 - 678 and toss anything above 678
5. Inspect the 7 vehicles

toss whole #

14

1. # each student 1-33
2. GO TO probability, random int. no repeats
3. lower, upper, sample size (4)
4. DONE! winners = 29, 33, 10, 5

Video Notes: Dr. Nic

- each item equally as likely - unbiased sampling
- there will always be sampling error

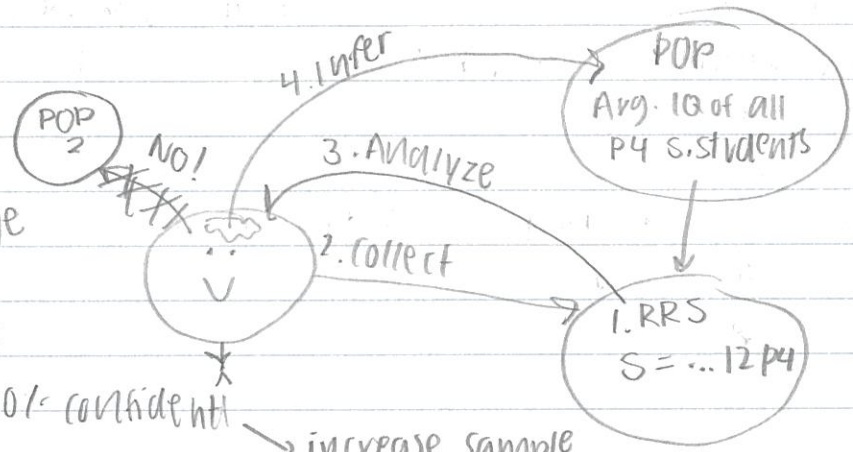
1. Simple Random - unbiased
2. Convienience - biased
3. systematic - (every other 20) - unbiased ★
4. Cluster - possible bias
5. stratified - divide ★

More Notes 8/14/23 (BIASES

INFERENTIAL STATISTICS PIC:

- Draw the process and pathways... ↑n effect

0. complete list
1. random sample (SRS, stratified, cluster, systematic)
2. collect data from random sample (N = pop n = sample size)
3. Analyze
4. Infer



$\bar{X}$ = sample average

M ≈ 143 ± 2.5

↑ 80% confident!

$\bar{X}_1$ = 143.21

$\bar{X}_2$ = 139.78

$\bar{X}_3$ = 147.13

↑ increase sample size for more confidence

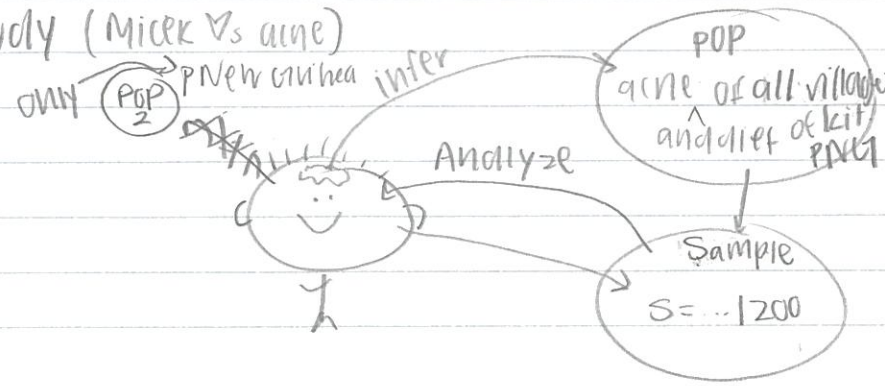
↑n = ↑c

Limits: (JAMHO's coffee one)

1. Ethnicity (Japanese Americans)
2. Gender (men)
3. Location (Hawaii)
4. Age (53 years old - avg.)
5. Observational study — Experiment = Causation

↖ self selected, they already drink coffee

Acne Study (Micek ∀s acne)



n = 1200+

S = diet + acne of 1200 villagers of Kitava PNG

*quiz on biases on wed!

BIAS:

ID FLAW

- response bias - lying to sound cool / don't know exactly the answer
  - aka give the wrong answer
- * recall bias - remember 1 specific event (engagement) / could remember WRONG
- nonresponse - plead the fifth
- under coverage Bias
- * convenience sample / selection Bias - asking first 5 people to come ★ KNOW
- voluntary Response Bias
- confirmation Bias
- * Availability Bias - big moment from the past
- Faulty recall - don't accurate remember event details
- Hidden Bias - question worded to elicit a certain response

FIGHT BIAS

- get a LIST
- simple random sample
- * stratified sample - (fr, so, jr, sr) split up grade level b/c no correlation to
- * cluster sample - clusters of locations - randomly pick 3 places or smt location
  - ↓ census location
- systematic sample - every 7th name on LIST

Biases Study Guide
(not from class)

Non-Response Bias:
- the person pleads the fifth aka doesn't answer

Under Coverage Bias:
- when not all members of a population are equally represented
  - lower sample survey given to Democrats

Response Bias:
- the person gives wrong information / doesn't remember small details
  - Leading questions - questions worded to favor one response (satisfied / unsatisfied question)
  - social desirability - answering to meet society's expectations (sounding cool)
  - ~~Voluntary Response Bias - sample members are volunteers (passionate)~~

Convenience Sample / Selection Bias:
- selecting convenient people (asking first ppl you see)
  - disproportionately represented - pple come in groups

Voluntary Response Bias:
- sample members are volunteers (more passionate)

Confirmation Bias:
- seeking out and interpreting data to support views / ideas (prove correct)

Availability Bias:
- remembering 1 big event from the past
  - plane crash makes someone scared of flying

Recall Bias
- the person recalls memories / events inaccurately and give wrong data unintentionally

Levels of Measurement 8/18/23

Nominal - categorical Data (qualitative)
- names, labels, listing
- order doesn't matter!
  - ex: car color, students in a group, etc.

Ordinal - categorical data
- arranged in order (differences don't matter)
- EX: ranking, ratings ; good, okay, bad

Interval - quantitative data
- can be arranged in order AND can differences matter)
- no natural zero or starting point
- EX: Temperature, years, olympic years, 0°F ≠ 0 heat

Ratio! - highest lvl of quantitative data
- can be ordered, differences, division, and starting 0 exists
- EX: distance from OPHS, length of class before getting bored, etc.

QUIZ!!

1. Ages of students (in years)
   - ratio
2. IQs of students
   - interval
3. rankings of student effort (Good better best)
   - ordinal

VIDEO NOTES:

1. Develop a critical mindset
2. Check the source
3. See who else is reporting it
4. Don't take images at face value
5. Check that it "sounds right"

ANOTHER QUIZ!

1. Fake
2. Fake
3. Fake
4. Real
5. Real

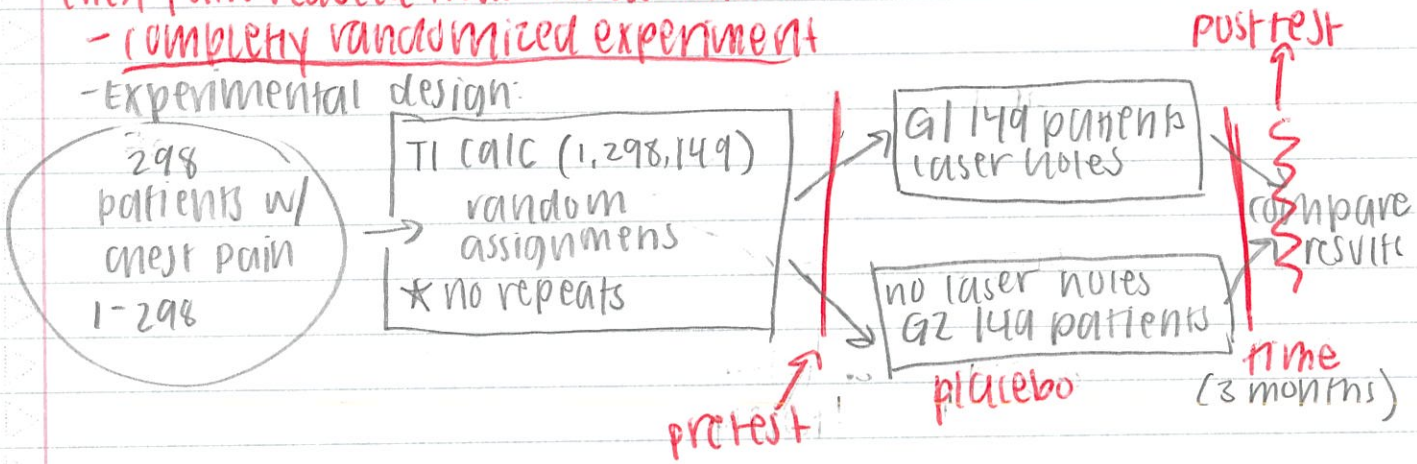
KNOW FOR QUIZ

- Response bias
  - recall
  - intentional vs. unintentional
- voluntary response bias
- non-response bias (who's blame?) – participants
- undercoverage bias
  - who's blame? – researchers
- overcoverage bias
- selection bias
- confirmation bias
- availability bias
- survivorship bias

3.1 Notes 8/24/23

Experiments and observation * skipped in class
- placebo

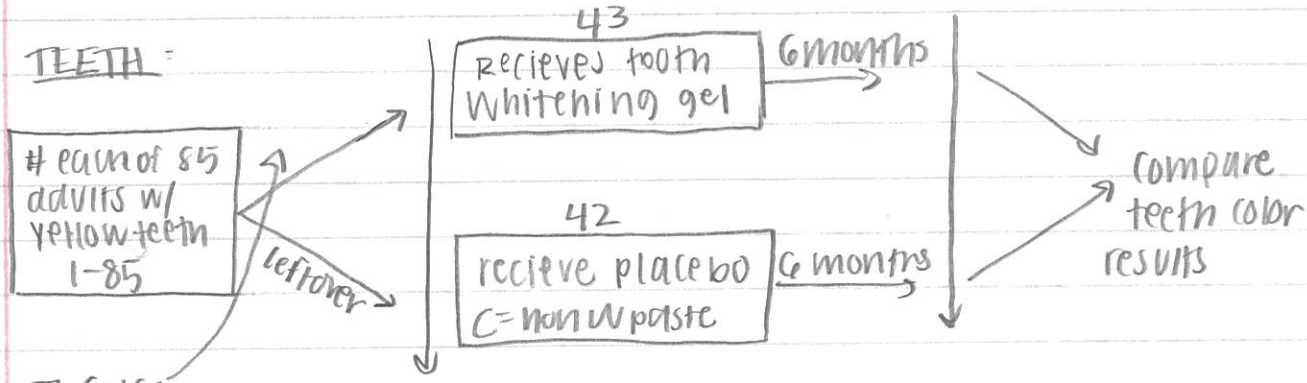
Chest pain reduced from drilled heart?
- completely randomized experiment
- Experimental design:



CHAPTER 1 REVIEW PROBLEMS 8/28/23

86. tooth whitening — double blindness

TEETH:



TI Calc:
randint NOREP
(1, 85, 43)

PRETEST:
- test PH
- before after photos

POST TEST:
- same as pretest

*double blindness (should NOT know)

youtube video problem

population — opionion on athletic budget of ALL parents of children at the large high school

Ast first 100 parents — bad
- voluntary response bias
- convenience bias

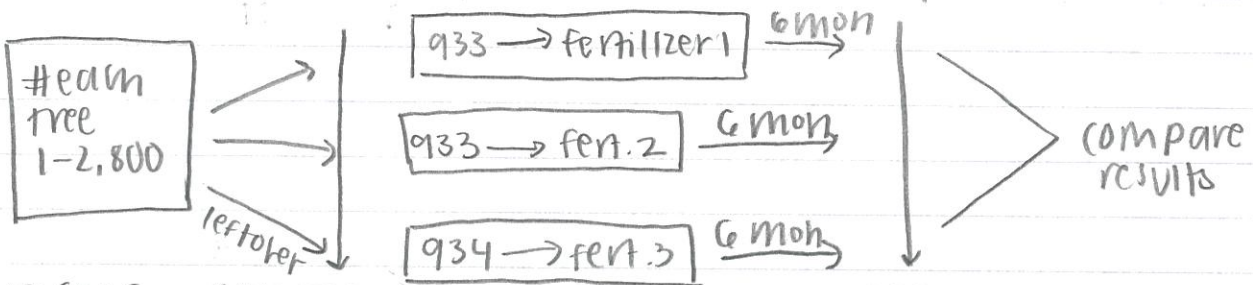
cluster — random group and surveying all from that group
stratified — SRS from each cluster

survey on facebook
- voluntary response bias

random sample ← ① List ② RNT ☆
- no bias

VIDEO #7

A farmer in Hawaii wines to conduct an experiment 3 fert. on 2,800 trees



TI CALC PRE TEST

(1, 2800, 933)

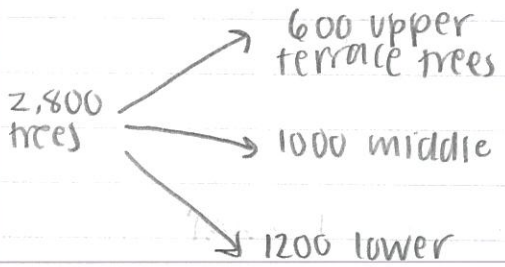
(1, 1867, 933)

leftover trees

POST TEST

* control is recommended

RANDOMIZED BLOCK DESIGN



BLAH BLAH

- trees already planted
- isn't even

shoe lace data 8/30/23

	Laces	No laces	Class:
M	𝍸 𝍸 ||		Kinberg seniors
F	𝍸 𝍸		
M	||||	|	Creason sophomores
F	𝍸 |||	|	
M	𝍸 𝍸 ||||		Myers juniors seniors
F	𝍸 𝍸		

total : 55

ML : 25

MNL : 1

FL : 28

FNL : 1

Good:

- we went to classes with different grade levels
- we went to more than 1 class
- saw real people

Bad:

- not random
- not ALL age groups ~ overcover seniors, undercover freshmen and sophomores
- ~~male females~~
- location ~ convenience
- not consistant samples in each
- no List! ~ not random sample

Shoe Laces 2.0

POP: Data of lace wearing habit from ALL OPHS students 2023-2024

Late August / Early September

statified - gender and grade level

n = 80+

20 freshmen < 10 F, 10 M

20 sophomores < 10 F, 10 M

20 juniors < 10 F, 10 M

20 seniors < 10 F, 10 M

random selection (prob w/ TI-84 calc)

Recorded Lecture Simulation Notes 8/31/23

- carry out experiment many times (count hot streaks!)

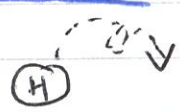
Probability model

- model = procedure for imitating or simulating
- EX: heads/tails for girl or boy being born
- simulation = imitation of of chance behavior based on an experiment that it represents

SIMULATIONS:

1. State the probability
2. Assumptions
3. ★ Assign digits ~ real outcomes
4. Many Reps
5. State conclusion (approximate to the real thing)

EXAMPLE: likelihood of having a girl is 0.938



① 3+ of same coin

Hot streak

② h ≈ t equally likely

50% 50%

- each toss is independent (results don't affect each other)

③ 0, 1, 2, ... 9 RNT 1/10

- independent
- 1 digit = 1 coin

↓

odd = H 1, 3, 5, 7, 9

even = T 0, 2, 4, 6, 8

10 #s = 10 coin flips = 1 rep.

④ Repeat many times 25+

⑤ State conclusion ≈

Rep1: yes

Rep2: yes

Rep3: yes

Rep4: yes

Rep5: yes

5/5 = 100%

HW PROBLEM! ★ 20. 39

5.56

① state probability
? 10 U.S. about evening exams

② Assumptions
Y/NO not likely
80% - 20%
- independent data

U.S NO
one
around

③ assign digits
0, 1, 2. ... 9
- independent
0-7 = yesss abolish evening exams
8, 9 = NO dont abolish evening exams

④ ten #s at a time
ALL 10

⑤ conclusion

1. 3675958984 NO!
2. 6828822913 NO!
3. No
4. No
5. No ~~0/8~~
6. No 1/9
7. No
8 NO
9. Yes!!!

Polls / surveys 9/8/23

1. Bad
   - verbal / body language X
   - favors the Tokyo olympics X
   - group setting X
   - leading / suggesting words ("Raise your hand", "If you believe", "is better than") X
   - thinking about cities (Tokyo vs. Rio) instead of the olympics X
   - different ages during the olympics X
2. Mediocre
   - leading / suggesting words ("did you enjoy") X
   - nonverbal = good! ✓
   - individual ~ people write their responses ✓
   - city (Tokyo) evokes images X
   - yes vs YESSSS! ← limitation X
3. Good!
   - no city mentioned ✓
   - ranking / more choices than Y/N ✓
   - individual ✓
   - nonverbal ✓
   - even # of #s so people can't choose the middle ✓
   - basic words = good (low / high satisfaction) ✓
   - XX only ordinal data ~ we want ratio data
4. Good
   - Ratio data
   - non verbal
   - collect data throughout the week ✓X

<u>GOOD SURVEY INCLUDES:</u>

- clear, consise
- confidential
- limit anxiety
- motivate a response
- assist accurate answers
- individual
- selection
- avoid bias

Real World Surveys 9/8/2023

- have #s be equally spaced
  1 2 3 4 NOT 1 2 3 4
- short and concise
- no pictures, words are the most powerful drug
- no suggesting / leading questions
- <u>NO</u> group setting

CHAPTER 1:

chapter 1 study guide (test: 9/13/23)

INTRODUCTION:

Population – data from ALL quantitative variable
→ Data from all students OR * Data of ____ from ALL ____

Sample – part of a population
→ Data from *variable*

N = population
n = sample

Random Number Table (steps):
① # each ___ ~~01 - 09~~ 00 – 99
② drop it like it's hot on a spot on the RNT
③ proceed in 1 direction, taking # digits at a time
④ keep #s ___ - ___ and toss any other numbers
⑤ inspect results

TI-84 Calc (steps)
① PROB → randInt NoRep
(___, ___, ___) EX: (1, 22, 4)
↑ lower ↑ upper ↑ n

STATS BASICS:

① collect	② organize
③ Analyze	④ interpret

① COLLECT
Samples (observational + surveys)
Experiments ~ ONLY ONE that shows causation
Census ~ least efficient
Simulations
we secs data.

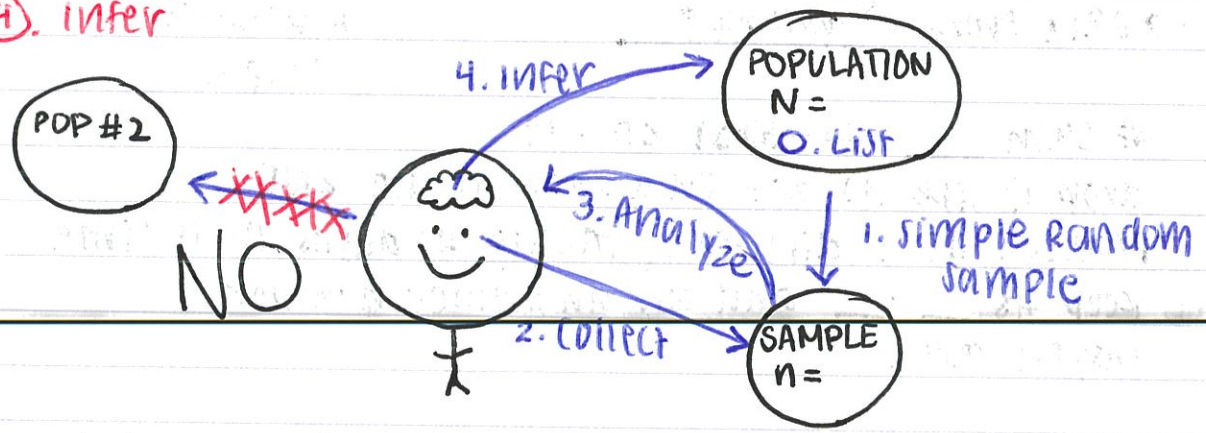
② ORGANIZE
- tables, charts, graphs, etc.

③ ANALYZE
- trends, center, outliers, hypothesis tests, etc.

④ INTERPRET
- assumptions, conclusions

INFERENTIAL STATISTICS:

⓪. complete list
①. Random sample (SRS, stratified, cluster, systematic, etc.)
②. collect data from sample
③. Analyze
④. infer



→ LIMITS: (JAMHOs)
- location (Hawaii)
- ethnicity (Japanese Americans)
- gender (male)
- age (old)
- observational study
  - can't conclude any causation (experiment)
  - lifestyles are different

LEVELS OF MEASUREMENT:

categorical
- Nominal – lists, names, labels ~ categorical qualitative
  - order doesn't matter
- Ordinal – categorical
  - order matters, differences don't
  - rankings, ratings

quant. data
- Interval – quantitative data
  - order and differences matter, no natural ZERO
- Ratio – BEST quantitative data
  - order, differences, division matter, and natural ZERO

BIASES:

Non-Response Bias – no response

Response Bias – faulty response from subject (lies, wrong info, etc.)

– Leading questions – duh

– Social desirability – responds in a way to meet social expectations

Convenience / Selection Bias – select easy access people (not complete representation)

Voluntary Response Bias – members volunteer their involvement

Confirmation Bias – seeking data to confirm preconcieved thoughts

Availability Bias – base opinions on 1 big event in the past

Recall bias – person remembers ~~things~~ events inaccurately

→ HOW TO FIGHT BIASES:

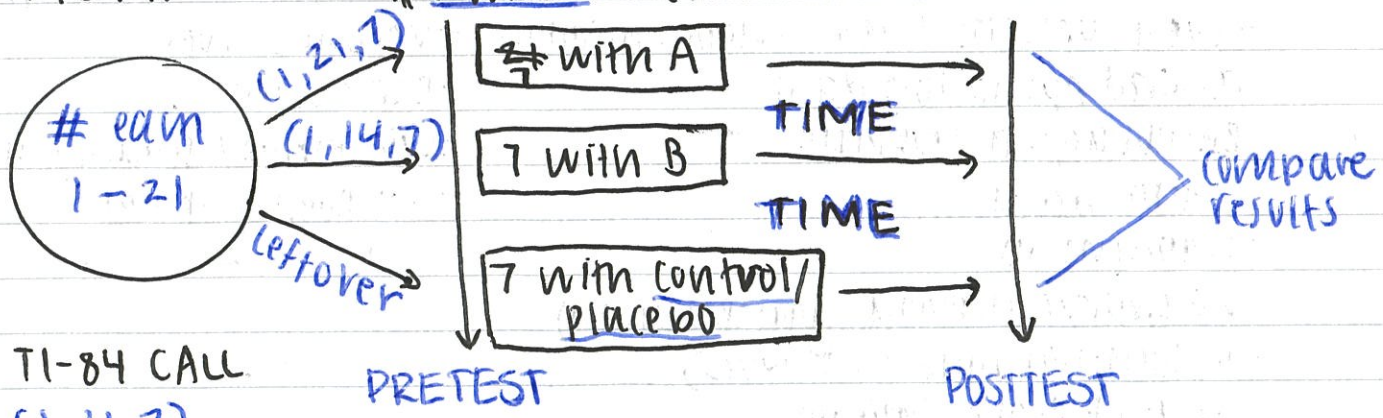
Simple random sample

~~Sta~~ Stratified sample – divide into groups (~~STRATA~~ STRATA) randomly choose a group and do a census on chosen group

Cluster sample – divide into groups (STRATA) and randomly select a few from each strata

Systematic sample – selecting every _th person

EXPERIMENTS: * control is recommended



TI-84 CALL

(1, 21, 7)

↓

–7 every time and renumber

– Blind? ~ surveyer knows ONLY

– doubleblind? ~ both don't know

SIMULATIONS:

1. state the probability
2. Assumptions
3. Assign digits
4. Many Reps
5. Conclusions

1. basketball free throws 70% makes them make 5/5 miss 3?
2. MONEYYY $$$ = 70% miss = 30%
   - each basket is independent (previous basket doesn't effect the next)
3. 0-6 = make 7-9 = miss - assume info on court/ player

4.

	1. NO	6. NO	11. NO	16. NO	21. NO
	2. NO	7. NO	12. NO	17. NO	22. NO
96746	3. NO	8. NO	13. Yes	18. Yes	23. No
12149	4. Yes	9. Yes	14. Yes	19. ~~Yes~~ NO	24. No
	5. NO	10. NO	15. NO	20. NO	25. No

MISS 3+ /5 = YES

5/25 = 1/5

71868

3/5

5. Conclusion - 20% won't make it

SURVEYS:

GOOD:
- basic wording
- even options (cant choose middle)
- remind ppl to log data
- encourage a response
- ratio data
- anonymous
- no groups around while asking
- no location bias (1 ▽ NYC)
  - ▫ don't evoke images

BAD:
- asking in a groupsetting
- verbal survey
- expressions
- suggestive / leading questions
- not evenly spaced
  1 2 3 4

REVIEW BEFORE THE TEST 9/13/23

Leading questions:

Implication questions - think of consequences to current and past events

Linked statements - association to things you previously said

assumptive question - act likes something is true to make people believe it's true

tag questions - short questions attached (right? isn't it?)

coercive question

* last few steps

- consider adding incentives (prizes)
- Take your survey for a test drive
- get help

SURVEY DESIGN NOTES 9/18/23

- at least a few questions should have ratio data ~ not just ordinal
- try not to use "not sure" or just ~~bata~~ basic "Yes or No"
- have an opening paragraph informing anonymoty? and purpose of the survey
- don't have a list in one question "students, teachers, parents"
  - split up question into a, b, c

SURVEY INFORMATION:

GOOD:

- say <u>who</u> is doing the study
- emphasize <u>importance</u> of response
- <u>incentive</u> ($2!)
- "cannot replace you!"
- ratio data
- anonymous
- having "a, b, c, d" questions

BAD:

- pictures ✓
- "not sure" as an option

WRITING SURVEY QUESTIONS:

GOOD:

- use language that's understandable
- remind them what the subject is
- more options
- placement of question
  - not having Q's influence future questions
- diverse reviews ★★
- pretest / test drive

BAD:

- assume that they know information on something
- offering too much information
- leading questions
- double-negatives
- <u>acquiescence bias</u> ~ people want to be agreeable

SEMESTER 1 PROJECT

~~LAGES AGAIN~~?

- well-written = ~~MOP~~ IMPORTANT! 4-5 ratio questions!
  - ☐ look at examples for help
- good planning ✱
- have DETAILS DETAILS DETAILS in explanation (struggles)
- be honest! (Did you actually ═ get everybody)
- limitations
  - ☐ race - more or less
  - ☐ wealthy ~ how does that affect laces ???!!!
- conclusions - why did you get what you got as a conclusion
- short intro! 3 lines is good
  - ☐ bold randomly selected, confidencial, can't be replaced, bag, and incentive
- don't start with the toughest questions and not the LAST question either

1. Never give up
2. Never too old to chase your dreams
3. It looks like a solitary sport but it's ~~atem~~ a team

Do you and your parents regularly watch the news
- ☐ no rating "yes or no"
- ☐ what is "regularly"
- ○ "discuss" ???
- ☐ "news"?

Rate your level of agreement that you and your parents talk about local news

Please chart the amount of time (in minutes) you and a trusted adult 5+ min minimum discuss the news this week

☐ sun ☐ mon ☐ tues ☐ wed ☐ thurs ☐ fri ☐ sat