MATH 136 - Statistics — Spring 2020

Exam 1

Ch 1, 2, 3

Dr. Jorge Basilio

March_9

jbasilio@glendale.edu

Honesty Pledge

On my honor, by printing and signing my name below, I vow to neither receive nor given any unauthorized assistance on this examination:

NAME (PRINT): Sdutions SIGNATURE: _____

Directions

- YOU ARE ALLOWED TO USE A CALCULATOR ON THIS EXAM. (Ti83/Ti83+/Ti84/Ti84+/Ti84+CE-T, or scientific calculator)
- You have 65 minutes to complete this exam.
- The exam totals **100 points**, with a possible 5 extra-credit points available.
- There are 14 problems, many of them with multiple parts.
- Place all of your belongings in the front of the classroom and I will assign you a seat. Bring with you your writing utensils.
- Cell phones must be turned off and put away in with your items in the front of the classroom.
- Handwriting should be neat and legible. If I cannot read your writing, zero points will be given.
- Make sure to ALWAYS SHOW YOUR WORK; you will not receive any partial credits unless work is clearly shown. *If in doubt, ask for clarification.*
- Leave answers in exact form (as simplified as possible), unless told otherwise.
- Put a box around your final answer where applicable.
- **PLEASE CHECK YOUR WORK!!!**
- If you need extra space, there is extra space on the back of the cover page and clearly indicate that you are continuing your work there in the original location.
- If you finish early, make sure to double- and triple-check your work. If you're done with that, then you may leave.
- I will take attendance at the end of class

Score	Grade

This page is intentionally blank. It may be used for scratch paper. If you wish for me to grade your work on this page, please (i) label the problem you are working on, (ii) box your answer, (iii) indicate in the original problem's location that you will continue your work on this page.

Problem 1: 10 pts (2 pts each)

TRUE or **FALSE** (please spell out/write the entire word for credit).

(a) TRUE A statistics student interviews everyone in his apartment building to determine who owns a cell phone. The sampling technique used is **convenience**.

(b) TRUE The object upon which the response variable is measured is called **an experimental unit**.

(c) TRUE **Experiments** assist the researcher in isolating the causes of the relationships that exist between two variables.

(d) TRUE A **response bias** occurs because we do not obtain complete information about a population.

(e) FALSE Mark retired from competitive athletics last year. In his career as a sprinter he had competed in the 100-meters event a total of 328 times. His average time for these 328 races was 10.24 seconds. The 10.24 seconds is a ~~statistic~~ parameter.

↳ all Mark's races

Problem 2: 4 pts

1 2 3 4 5 6 7 8 9 10 ...

(a) Every fifth adult entering an airport is checked for extra security screening. What **sampling technique** is used?

(A) convenience
(B) cluster
(C) systematic C
(D) simple random
(E) stratified

(b) At a local technical school, five auto repair classes are randomly selected and all of the students from each class are interviewed. What **sampling technique** is used?

(A) convenience
(B) cluster B
(C) systematic
(D) simple random
(E) stratified

note know difference b/w cluster vs stratified

Problem 3: 4 pts

A researcher wanted to determine whether colon cancer was associated with eating meat. He selected a sample of 500 men with colon cancer and an equal number of men without colon cancer. The two groups were matched - in other words they were similar in terms of age, occupation, income, and exercise levels. Histories on the amount of meat consumed over the previous twenty years were obtained for all men. The total amount of meat that each man eaten in the previous twenty years was estimated. The meat consumption was compared for the two groups.

observational

~~(A)~~ cross-sectional; Information is collected at a specific point in time.
(B) cohort; Individuals are observed over a long period of time. B
(C) retrospective; Individuals are asked to look back in time.
~~(D)~~ cohort; Individuals are grouped together and studied in a single day.
~~(E)~~ cross-sectional; Information is collected by many people in a single day.

cohort → men w/ cancer
"groups" → men w/o cancer

Problem 4: 6 pts

Write down the **notation** we use for:

(a) **sample mean**: $\bar{x}$

(b) **sample standard deviation**: s

(c) **sample size**: n

(d) **population mean**: μ

(e) **population standard deviation**: σ

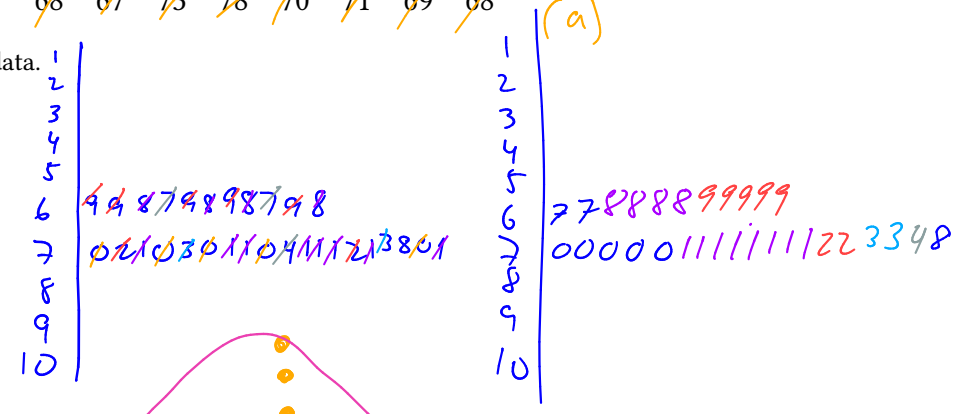
(f) **population size**: N

Problem 5: 10 pts

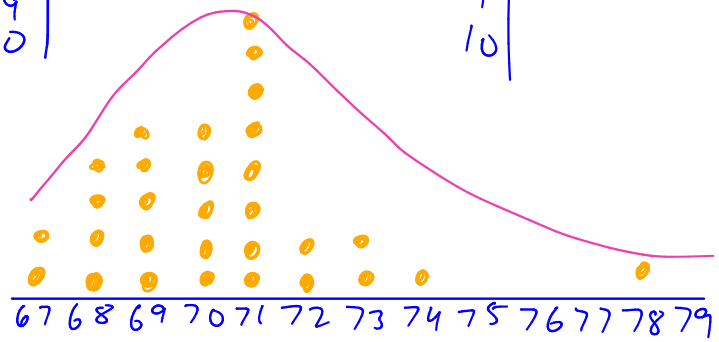
The heights (in inches) of 30 mechanics are listed below.

70	72	71	70	69	73	69	68	70	71
67	71	70	74	69	68	71	71	71	72
69	71	68	67	73	78	70	71	69	68

(a) Create a **stem-leaf plot** display for the data.



(b) Create a **dot plot** display for the data.



(c) What is the **shape of the distribution**.

skewed right

b/c tail is on right on dot plot

Problem 6: 12 pts

units! CCs = chocolate chips

Refer to the data set of number of chocolate chips in a bag of Chip's Ahoy Chewy in the accompanying table to answer the following questions.

21	20	16	17	16	17	20	22	14	20
19	17	20	21	21	18	20	20	21	19
22	20	20	19	16	19	16	15	24	23
10	24								

(a) Find the **sample mean**: $\bar{x} = 19.0$ CCs per bag

(b) Find the **median**: Med = 20 CCs

(c) Find the **mode**: 20 CCs

(d) Find the **sample standard deviation**: $s = 3.0$ CCs

(e) Find the **sample variance**: $s^2 = 9.0$ CCs squared

$s^2 = (3.0)^2 = 9.0$

(f) Find the **minimum**: 10 CCs

(g) Find the **maximum**: 24 CCs

(h) Find the **range**: 14 CCs = max − min = 24 − 10

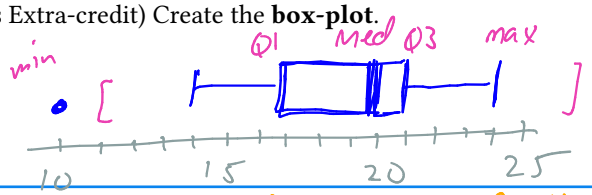
(i) Find the **inter quartile range**: 4 CCs

IQR = $Q_3 - Q_1$ = 21 − 17

(j) (3 pts) Give the **5 number summary**:

min = 10 ccs
Q1 = 17 ccs
Med = 20 CCs
Q3 = 21 CCs
max = 24 CCs

(5 pts Extra-credit) Create the **box-plot**.



outliers? fences LF = Q1 − 1.5 IQR = 11
UF = Q3 + 1.5 IQR = 27

min is outlier!

Problem 7: 4 pts

QL QN

Identify whether the given data set is **qualitative** or **quantitative** data:

(a) QL Collection of favorite coffee shops of GCC students

(b) QN Collection of fluid ounces of soda served at GCC cafeteria each day

(c) QN Collection of number of streaming services GCC students subscribe to

(d) QL Collection of favorite courses of GCC students

Problem 8: 4 pts

Identify whether the given data set is **discrete** or **continuous** data:

(a) discrete Number of phone calls a police department receives in a day

(b) discrete (# of) Native language of a tourist

(c) continuous Height of a tree on campus

(d) ~~Highest degree obtained by your mother~~ ← makes no sense (typo)

Problem 9: 4 pts

Identify appropriate **level of measurement** (nominal, ordinal, interval, ratio):

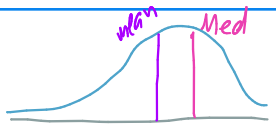
(a) ordinal Movie ratings of one star through five stars

(b) ratio Amount of water used by GCC in a day

(c) interval Year of birth of GCC students

(d) nominal Phone numbers of GCC students

Problem 10: 6 pts



tail pushes mean

Fill-in the blanks.

(a) If a distribution is **skewed left**, then the mean is lees/smaller than the median.

(b) The range is the easiest **measure of dispersion** to calculate by hand.

(c) A **z-score** tells us how many standard deviations away from the mean a data value is.

(d) The **rounding rule** for estimating people tells us to always round up .

(e) The **empirical rule** says that in a normal distribution (bell-shaped) approximately 95% (2σ) of the data lies between $\mu - 2\sigma$ and $\mu + 2\sigma$ (Hint: Use μ and σ in your answers).

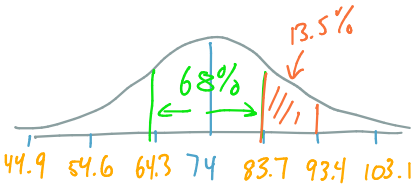
Problem 11: 6 pts

Health care issues are receiving much attention in both academic and political arenas. A sociologist recently conducted a survey of citizens over 60 years of age whose net worth is too high to qualify for government health care but who have no private health insurance.

Suppose the **mean** and **standard deviation** are 74.0 ($\bar{x}$) and 9.7 (s), respectively. Assume that the distribution of ages is bell shaped.

→ can use Empirical Rule

(a) What **percentage** of the respondents will be between 64.3 and 93.4 years old?



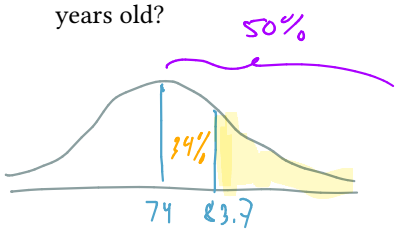
95 − 68 = 27

27/2 = 13.5

68 + 13.5 = 81.5 %

Approximately 81.5% of respondents will be between 64.3 & 93.4 years old.

(b) Assume that 25 citizens over 60 years of age are randomly selected to participate in the survey. **How many** will be over 83.7 years old?



50% − 34% = 16%

25 * 0.16 = 4

Four (4) citizens will be over 83.7 years old in a sample of 25.

Problem 12: 6 pts

Chanel scored 84 on her stat's test where her class had a mean of 66 and a standard deviation of 12. Whereas Luis scored 40 on his stat's test with a mean of 32 and a standard deviation of 5. Who did did better on their test?

Justify your answer with statistical reasoning. Please give your answers in complete sentence(s) $(M \to E)$.

Chanel	Luis
$x = 84$	$x = 40$
$\mu = 66$	$\mu = 32$
$\sigma = 12$	$\sigma = 5$
$z = \frac{84-66}{12}$	$z = \frac{40-32}{5}$
$z = 1.50$	$z = 1.60$

In order to compare chanel's and Luis' performance, we compute z-scores to see how they did relative to their class. Luis did better than Chanel because Luis z-score is higher (1.60 vs 1.50).

Problem 13: 5 pts

Marissa has just completed her second semester in college. She earned a B in her five-hour calculus course, an A in her three-hour social work course, an A in her four-hour biology course, and a C in her three-hour American literature course.

Assuming that an A equals 4 points, a B equals 3 points, and a C equals 2 points, determine Marissa's grade- point average for the semester.

Marissa	w	x
Calc	5 hrs	B 3.0
Social work	3 hrs	A 4.0
Bio	4 hrs	A 4.0
Am Lit	3 hrs	C 2.0

$GPA = \frac{\sum w \cdot x}{\sum w}$

$GPA = \frac{5 * 3.0 + 3 * 4.0 + 4 * 4.0 + 3 * 2.0}{5 + 3 + 4 + 3}$

$= \frac{49.0}{15} = 3.266\ldots$

Marisa's GPA in her second semester is 3.27

Problem 14: 19 pts

The following **frequency distribution** shows the top speed (in kilometers per hour) of all 607 players (except goaltenders) in the 2014 World Cup Soccer Tournament.

(8 pt) (a) Fill out the rest of the table:

Class (Speed km/hr)	Frequency ($\Sigma f = 607$)	Relative Frequency	Cumulative Relative Frequency	Midpoint ($CM = \frac{UCL+LCL}{2}$)	Class Boundary ($CB = \frac{UCL+LCL}{2}$)	$f \cdot x$ (CM)
10-13.9	4	$4/607 = 0.007$	$4/607 = 0.007$	$\frac{10+14}{2} = 12$	$\frac{13.9+14}{2} = 13.95$	48
14-17.9	7	$7/607 = 0.012$	$11/607 = 0.018$	$\frac{14+18}{2} = 16$	17.95	112
18-21.9	17	$17/607 = 0.028$	$28/607 = 0.046$	20	21.95	340
22-25.9	91	$91/607 = 0.15$	$119/607 = 0.197$	24	25.95	2184
26-29.9	282	$282/607 = 0.465$	$401/607 = 0.661$	28	29.95	7896
30-33.9	206	$206/607 = 0.339$	$607/607 = 1.000$	32	33.95	6592

$\Sigma f \cdot x = 17172$ (sum(L3))

L1 L2 L3 = L1 * L3 — use short cuts taught

NOTE: Give only the class boundaries on the right.

NOTE: Give the relative frequency as decimals rounded to the nearest thousands (3 decimal places).

(2 pt) (c) **How many** soccer players were surveyed? (typo fixed)

607 soccer players (except goalies)

(1 pt) (d) Determine the **shape of the distribution** by stating the skewness.

skewed left

"sketch"

(2 pt) (e) What is the **relative frequency** of players that have a top speed of at most 21.9 kilometers per hour?

= 21.9 km/hr or less

same as cumulative frequency up to 21.9

$\frac{4+7+17}{607} = \frac{28}{607} =$ 0.046

(2 pt) (f) **How many** players have a top speed of at least 18 kilometers per hour?

↳ 18 km/hr or more

$17+91+282+206 =$ 596 players have top speeds of 18 km/hr or greater.

(4 pt) (d) Calculate the **mean** top speed of 2014 World Cup Soccer players (except goaltenders):

- $\bar{x} = \frac{\Sigma f \cdot x}{\Sigma f} = \frac{17172}{607} =$ 28.3 km/hr
- $\Sigma f = 607$ — use calculator shortcuts!
- x = use CMs!

Formula Sheet

- $\boxed{\bar{x} = \frac{\sum x}{n}}$ or $\boxed{\bar{x} = \frac{\sum (f \cdot x)}{\sum f}}$
- $\boxed{s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}}}$
- $\boxed{z = \frac{x - \bar{x}}{s}}$ or $\boxed{x = \bar{x} + s \cdot z}$
- $\boxed{\bar{x}_w = \frac{\sum (w \cdot x)}{\sum w}}$
- $\boxed{IQR = Q_3 - Q_1}$
- $CM = \frac{LCL+LCL}{2}$, $CB = \frac{UCL+LCL}{2}$
- k^{th} **Percentile:**

$$\boxed{P_k = \frac{\text{\# scores < given score}}{\text{total \# scores}}}$$

- **Finding the score L given a percentile k:**

$$\boxed{L = \frac{k}{100} \cdot n}$$

  - if L is a decimal, round up
  - if L is whole, then average the k^{th} score and the next higher score

Post Exam Survey

Now that you have finished the exam, please take a few minutes to reflect on how you prepared for the exam and how you think you did. Then answer these questions.

1. When taking the exam I felt
   (a) Rushed. I wanted more time.
   (b) Relaxed. I had enough time.
   (c) Amazed. I had tons of extra time.
2. The week before the test I did all my homework on time: YES NO
3. The week before the test, in addition to the homework I followed a study plan. YES NO
   (a) I think this helped: YES NO
4. The day before the test I spend _____ hours studying and reviewing.
   (a) I think that was enough time: YES NO
5. The night before the test:
   (a) I stayed up very late cramming for the test
   (b) I stayed up very late, but I wasn't doing math
   (c) I didn't need to cram because I was prepared
   (d) I got a good night's sleep so my brain would function well.
6. I think I got the following grade on this test: _____
7. Strategies that worked well for me were (please elaborate):

please fill it out :)

thanks

8. Next time I will do an even better job preparing for the test by: