Intro Stats Test A – Data Analysis – Part I   Name __Key__

1. School administrators collect data on students attending the school. Which of the following variables is quantitative?
   A) class (freshman, sophomore, junior, senior)   **B)** grade point average
   C) whether the student is in AP classes   D) whether the student has taken the SAT

3. A professor has kept records on grades that students have earned in his class. If he wants to examine the percentage of students earning the grades A, B, C, D, and F during the most recent term, which kind of plot should he make?
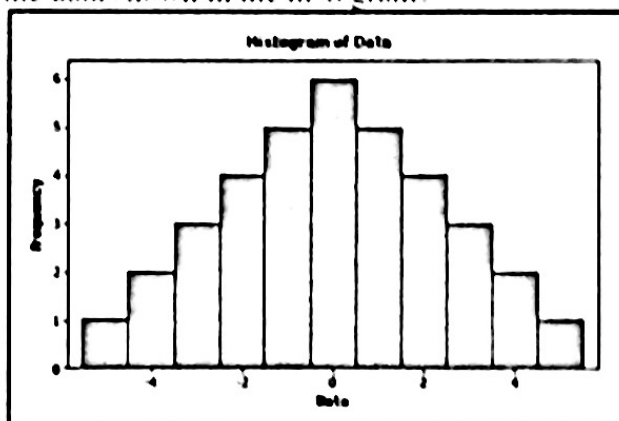   A) boxplot   B) timeplot   C) dotplot   *others are used*
   D) pie chart   E) histogram   *for quantitative data*

4. Which is true of the data shown in the histogram?

Histogram of Data

   I.   The distribution is approximately symmetric. **TRUE**
   II. The mean and median are approximately equal. **TRUE**
   III. The median and IQR to summarize these data summarize the data better than the mean and standard deviation. **FALSE**

   A) I only   B) III only   **C)** I and II
   D) I and III   E) I, II, and III

   Symmetric / No Outliers

   use mean & SD

7. Suppose that a Normal model described student scores in a history class. Parker has a standardized score (z-score) of +2.5. This means that Parker

    A) is 2.5 points above average for the class.

    B) is 2.5 standard deviations above average for the class.

    C) has a standard deviation of 2.5.

    D) has a score that is 2.5 times the average for the class.

    E) None of the above.

Definition of Z-score

9. The five-number summary of credit hours for 24 students in an introductory statistics class is

Min	Q1	Median	Q3	Max
130	150	165	180	220

From this information, we know that

    A) there are no outliers in the data.

    B) there is at least one low outlier in the data.

    C) there is at least one high outlier in the data.

    D) there are both low and high outliers in the data.

    E) None of the above.

$$Q_1 - 1.5(IQR)$$
$$Q_3 + 1.5(IQR)$$

$$IQR = Q_3 - Q_1 = 3$$

$$15 - 1.5(3) = 10.5 \quad \text{No}$$
$$18 + 1.5(3) = 22.5 \quad \text{outliers}$$

10. Which of the following data summaries are changed by adding a constant to each data value?

    I. the mean ✓

    II. the median ✓

    III. the standard deviation ✗

    A) I only   B) III only   **C) I and II**

    D) I and III   E) I, II, and III

Adding changes everything except spread

(SD / IQR)

11. The table shows whether students in an introductory statistics class like dogs and/or cats.

		Like Dogs		
		Yes	No	Total
Like Cats	Yes	194	21	215
	No	110	10	120
	Total	304	31	335

$$\frac{304}{335} \qquad \frac{31}{335}$$

a. What is the marginal distribution (in %) of "liking dogs"?

Yes   No
90.7%.   9.3%.

b. What is the conditional distribution (in %) of "liking dogs" for students who like cats?

Yes   No $\frac{194}{215}$ $\frac{21}{215}$
90.2%.   9.8%.

c. What kind of display(s) would you use to show the association between "liking dogs" and "liking cats"? (Just name the graph(s).) Pie charts, Bar Charts

d. Do "liking dogs" and "liking cats" appear to be independent? Give statistical evidence to support your conclusion.

Either   NO %'s   or   yes b/c %'s are   (NOT ON EXAM) *
         are different =   very close

12. A local plumber makes house calls. She charges $30 to come out to the house and $40 per hour for her services. For example, a 4-hour service call would cost $30 + 4($40) = $190.

a. The table shows summary statistics for the past month. Fill in the table to find out the cost of the service calls.

Statistic	Hours of Service Call	Cost of Service Call
Mean	4.5	$210
Median	3.5	$170
SD	1.2	$48
IQR	2.0	$80
Minimum	0.5	$50

} only multiply by 40

b. This past month, the plumber had a service call that corresponded to a z-score of - -1.50. What was the z-score for the cost of that service call?

Z = -1.50   b/c Z scores do not change when transforming data.

↑ (Not on exam) *

13. The *World Almanac and Book of Facts 2004* reported the percent of people not covered by health insurance in the 50 states and Washington, D.C., for the year 2002. Computer output gives these summaries for the percent of people not covered by health insurance:

Min	Q1	Median	Q3	Max	Mean	SD
7.9	10.8	13.4	16.7	25.8	13.9	3.6

a. Were any of the states outliers? Explain how you made your decision.

$$IQR = 16.7 - 10.8 = 5.9$$
$$1.5(5.9) = 8.85$$

$$10.8 - 8.85 = 1.95$$
$$16.7 + 8.85 = 25.55$$

At least one outlier (MAX)

b. A histogram of the data is as follows:


Histogram of % not covered
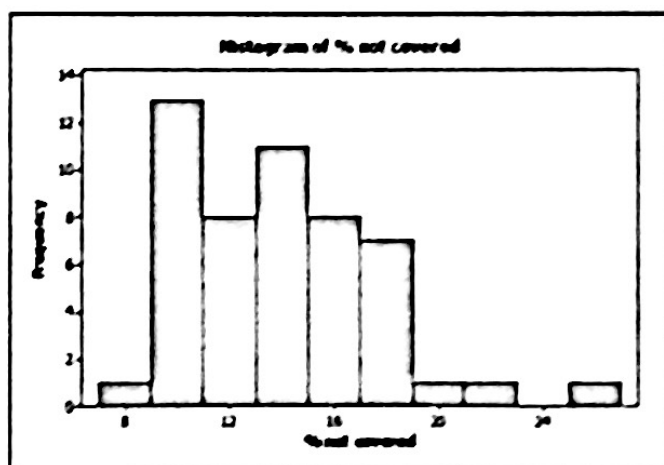
Is it more appropriate to use the mean and standard deviation or the median and IQR to describe these data? Explain.

Median & IQR because the graph is <u>not</u> symmetric (skewed right)

14. Costs for standard veterinary services at a local animal hospital follow a Normal model with a mean of $80 and a standard deviation of $20.

a. Draw and clearly label this model.

$20 $40 $60 $80 $100   120   140
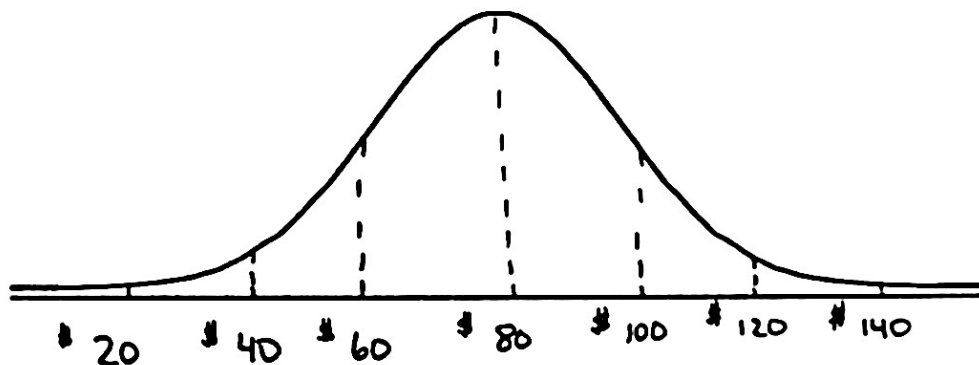
b. Is it unusual to have a veterinary bill for $125? Explain.

$$Z = \frac{125-80}{20} = \frac{45}{20} = 2.25$$ ⟶ look up in chart (unlikely)
(.9878)

1-.9878 = .0122 , so about a 1.22% Chance of a bill being that high

c. What is the IQR for the costs of standard veterinary services? Show your work.

Q_1
(25th percentile)
$Z = -.67$

Q_3
(75th percentile)
$Z = .67$

$$-.67 = \frac{x-80}{20}$$

$X = 66.6$

$$.67 = \frac{x-80}{20}$$

$X = 93.4$

93.4 - 66.6
IQR is $26.80

15. A machine that fills cans with soda fills according to a Normal model with mean 12.1 ounces and standard deviation 0.05 ounces.

a. If the cans claim to have 12 ounces of soda each, what percent of cans are under-filled?

$$Z = \frac{12-12.1}{.05} = -2.00$$
(.0228)

About 2.28% of cans are under-filled

b. Management wants to ensure that only 1% of cans are under-filled.

i. Scenario 1: If the mean fill of the cans remains at 12.1 ounces, what standard deviation does the filling machine need to have to achieve this goal?

Look up (.0100) in chart
$Z = -2.33$

$$-2.33 = \frac{12-12.1}{SD}$$ solve for SD.

$-2.33 = \frac{-.1}{SD} \cdot SD$ $SD = \frac{-.1}{-2.33} = .043$ ounces

ii. Scenario 2: If the standard deviation is to remain at 0.05 ounces, what mean does the filling machine need to have to achieve this goal?

$$-2.33 = \frac{12-\bar{x}}{.05}$$

Solve for $\bar{x}$ to get 12.12 ounces

16. Explain how much T.V a person watches a week could be measured as a quantitative variable and as a categorical variable:

Quantitative: Hours, Minutes   Categorical: None, Some, A Lot, Extreme
                                                                    etc.

17. You have a bag with 10 number tiles in it and the tiles are numbered from 1 – 10. You are going to select a tile and then record the number, then put the tile back in the back and repeat the process. What is the probability that if you select 500 tiles you would get *more than* 265 odd numbered tiles?

STAT CRUNCH   n = 500
                   P = .5

P(x > 265) ≈ .0828

18. What is the probability of rolling a six-sided die and getting a 4, and then getting a blue marble from a bag that contains 10 red, 7 green, 8 blue and 5 orange marbles?

$$\left(\frac{1}{6}\right)\left(\frac{8}{30}\right) = \frac{8}{180} = \frac{2}{45}$$

19. Given you have a bag that contains 100 number tiles numbered 1 – 100:

a. What would be the probability of selecting one number tile that is either even or *greater than* 75 on one pull?

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

13 #'s
Even & greater
than 75

$$= \frac{50}{100} + \frac{25}{100} - \frac{13}{100} = \frac{62}{100} = \frac{31}{50} \text{ or } .62$$

b. Are these two events disjoint? (Explain: Why or Why Not)

No #'s can be even & greater than 75

20. If you have a drawer with 16 socks in it (10 blue and 6 red), then what would be the probability of selecting two socks at random and getting a matching pair?

B   B   or   R R

$$\left(\frac{10}{16}\right)\left(\frac{9}{15}\right) \quad + \quad \left(\frac{6}{16}\right)\left(\frac{5}{15}\right)$$

$$\frac{90}{240} \quad + \quad \frac{30}{240} = \frac{120}{240} = .5$$

21. Insurance company records indicate that 12% of all teenage drivers have been ticketed for speeding and 9% for going through a red light. If 4% have been ticketed for both, what is the probability that a randomly selected teenage driver has been ticketed for speeding but not for running a red light?

A. 3%
B. 8%
C. 12%
D. 13%
E. 17%

S

R

8% (4%) 5%
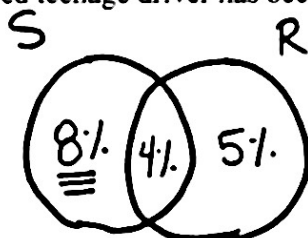
22. Sara is taking a test in her science class and her math class, if the class average in science was a 74% and the standard deviation was 10%, compared to an average of 81% and a standard deviation of 4% in her math class. Which tests did she do better on compared to the rest of the class given her score on the science test was a 90% and he math test score was an 88%? (Explain)
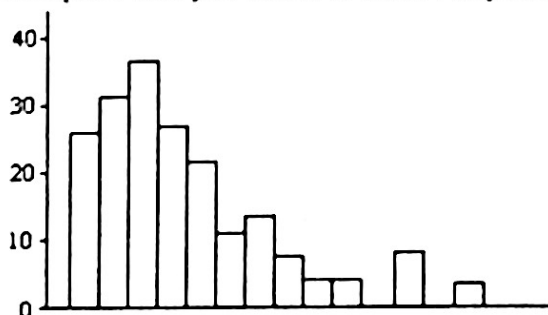
Science
$$Z = \frac{90-74}{10}$$
$$Z = 1.6$$

Math
$$Z = \frac{88-81}{4}$$
$$Z = 1.75$$

Did better in Math

23. In the histogram shown below describe the shape, the best measure for center and spread, and then explain how you think the mean compares to the median.

- Skewed Right
- Mean is best measure of center
- SD " " measure of spread
- Mean > Median

24. You draw two marbles at random from a jar that has 20 red marbles and 30 black marbles without replacement. What is the probability that both marbles are red?

A. 0.1551
B. 0.1600
C. 0.2222
D. 0.4444
E. 0.8000

$$\left(\frac{20}{50}\right)\left(\frac{19}{49}\right) = .1551$$

25. You have a bag with 10 number tiles in it and the tiles are numbered from 1 – 10. You are going to select a tile and then record the number, then put the tile back in the back and repeat the process.

   a) What is the probability that if you select 500 tiles you would get *less than* 220 odd numbered tiles?

$$\text{STAT COUNCH}$$

$$n = 500 \qquad p = .5$$

$$P(x < 220) \approx .0032$$

26. What is the probability of flipping a coin and getting tails and then getting a blue marble from a bag that contains 4 red, 6 green, 8 blue and 2 orange marbles?

$$\left(\frac{1}{2}\right)\left(\frac{8}{20}\right) = \frac{8}{40} = \frac{1}{5} = \underline{\underline{.2}}$$

27. If you have a bag that contains 100 number tiles numbered 1 – 100, then what would be the probability of selecting one number tile that is either odd or *greater than* 80 on one pull?

$$P(ODD) + P(\# > 80) - P(ODD \ \& \ \# > 80)$$

$$\frac{50}{100} + \frac{20}{100} - \frac{10}{100} = \frac{60}{100} = \underline{\underline{.6}}$$

28. Sara is taking a test in her science class and her math class, if the class average in science was a 72% and the standard deviation was 8%, compared to an average of 78% and a standard deviation of 2% in her math class. Which tests did she do better on compared to the rest of the class given her score on the science test was a 90% and he math test score was a 83%? (Explain)

$$
\overset{S}{Z = \frac{90 - 72}{8}} = 2.25 \qquad\qquad \overset{M}{Z = \frac{83 - 78}{2}} = 2.5
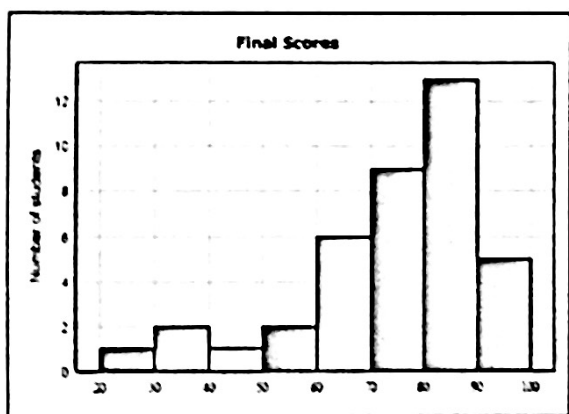$$

$$\text{She did better in Math}$$

29. Describe two variables that would be considered categorical and two variables that would be considered quantitative.

$$\underline{\quad C \quad}$$

Color, Fav. Food
ETC.

$$\underline{\quad Q \quad}$$

Hours of Sleep
Days of School

30. In the histogram shown below describe the shape, the best measure for center and spread, and then explain how you think the mean compares to the median.

- Skewed left
- Median & IQR center & spread

mean < median

31. You play tennis regularly with a friend, and from past experience, you believe that the outcome of each match is independent. For any given match you have a probability of 0.6 of winning. The probability that you win the next two matches is

A. 0.16.
B. 0.36.
C. 0.4.
D. 0.6.
E. 1.2.

$$(.6)(.6) = .36$$

32. Six Republicans and Four Democrats have applied for two open positions on a planning committee. Since all the applicants are qualified to serve, the City Council decides to pick the two new members randomly. What is the probability that both come from the same party?

A) 66/90   B) 42/100   C) 52/100 D) 42/90   E) 52/90

$$\underset{R}{\left(\frac{6}{10}\right)} \underset{R}{\left(\frac{5}{9}\right)} + \underset{D}{\left(\frac{4}{10}\right)} \underset{D}{\left(\frac{3}{9}\right)}$$

$$\frac{30}{90} + \frac{12}{90} = \frac{42}{90} = \boxed{\frac{7}{15}}$$

33. At a school there are 100 students in the Senior Class and:
    10 Students play Baseball, Basketball and Football
    17 Students play Baseball and Football
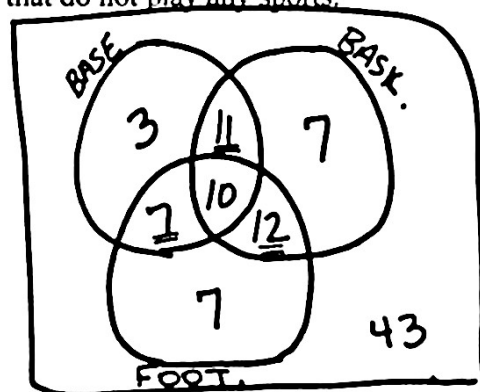   21 Students play Baseball and Basketball
    22 Students play Basketball and Football
    31 Students play Baseball
    36 Students play Football
    40 Students play Basketball

- Draw a Venn diagram to represent this scenario, make sure to include students that do not play any sports.

- What is the probability that you randomly select one student and they do not play any of these sports?

$$\frac{43}{100} = .43$$

- What is the probability of selecting one student at random that plays just Football?

$$\frac{7}{100} = .07$$

- What is the probability that if you select two students, they both play _exactly_ two sports?

$$\left(\frac{30}{100}\right)\left(\frac{29}{99}\right) = \frac{29}{330} \approx .0879$$