

## Unit 1 - Intro to Statistics

1. Statistics F
  2. Individual C
  3. Variable I
  4. Categorical Variable A
  5. Quantitative Variable H
  6. Population G
  7. Sample K
  8. Census D
  9. Experiment E
  10. Treatment J
  11. ~~Statistical Problem Solving Process~~ B  
**observational study**
- ~~A. Places an individual into one of several groups/categories~~
- ~~B. When someone observes something, but does not influence or interrupt; only describes the behavior of a group or situation~~
- ~~C. The objects described by a set of data; it is the who or the what we are studying~~
- ~~D. A sample survey that attempts to include the entire population in the sample~~
- ~~E. When someone deliberately imposes some treatment on individuals in order to record their responses~~
- ~~F. The art and science of dealing with data~~
- ~~G. Entire group of individuals that are being studied~~
- ~~H. Takes a numerical value where averaging makes sense~~
- ~~I. Any characteristic of an individual~~
- ~~J. Something that researchers give to experimental units/subjects~~
- ~~K. A part of the population that data is actually collected from; this is used to come up with conclusions about the population~~
12. What are the 4 steps of the Statistical Problem-Solving Process? What do you do in each step?
    - a) Ask a question of interest
    - b) Produce Data
    - c) Analyze Data
    - d) Interpret Results
  13. Give 2 examples of categorical variables.
    - a) Zip code
    - b) Gender
  14. Give 2 examples of quantitative variables.
    - a) Test average
    - b) Height

15.

Student Name	Gender	Favorite Subject	Height (inches)	Preferred Communication	Zip Code	Number of Pets
Jessica	Female	Math	64	Text	30062	2
Joel	Male	Lunch	72	Snapchat	30060	4
Tashana	Female	Culinary	65	Snapchat	30060	1
Zach	Male	History	74	Text	30061	2
Daejia	Female	Math	65	Text	30082	0
Bria	Female	English	63	Snapchat	30080	2

a) Identify the individuals. *Students*

b) Identify the categorical variables. *Gender, Fav. Subject, Preferred Communication, Zip code*

c) Identify the quantitative variables. *Height, # of Pets*

16. You want to know the effect of watching TV while studying and final exam scores for high school students. You asked 50 Statistical Reasoning students on Senior Skip Day whether they had the TV on while they were studying for their final. 42% of the students answered "yes".

a. Is this an example of an observational study or experimental study? *observational study*

b. Identify the population. *high school students*

c. Identify the sample. *50 Stat students*

d. If a student answers "yes", what kind of variable is that? *categorical*

e. **How many** students answered "yes"? *.42(50) = 21 students*

g. If you asked a different sample of 50 students, would you expect your result to be 42% again? Why or why not?

*No, different sample → different results*

h. Was the sample a good representation of the population? Why or why not?

*No, 50 students is a small sample AND it was Senior Skip Day.*

17. Determine which brands of orange juice people prefer. The people are randomly chosen at the supermarket and are asked to taste BOTH brands without knowing which brand they are drinking. **Is this an observational study or experiment? Why?**

Experiment; subjects are given orange juice as a treatment

18. Compare the grades on a final math test of 25 students who use calculators and 25 students who do not use calculators. The students decide which group they are in. **Is this an observational study or experiment? Why?**

Observational study; tests were just compared and students chose the group

## Unit 2 - Data Displays

19. Roundoff Error G ~~A~~ A quantitative display that uses bars that touch, uses a number line, and consistent intervals to show the distribution of data.
20. Pie Chart C ~~B~~ Minimum, first quartile, median, third quartile, and maximum
21. Bar Graph K ~~C~~ A special categorical chart that uses "pie slices" to show relative size of the data.
22. Dot Plot E ~~D~~ ALL of the different values the variable takes on and how often it takes these values
23. Stem and Leaf Plot J ~~E~~ A display of quantitative data on a number line; shows each individual data point
24. Histogram A ~~F~~ The point where about half of the observations are on either sides; sometimes it is better to use the median than the mean due to skewness.
25. Distribution D ~~G~~ The difference between the calculated approximation of a number and its exact mathematical value due to rounding.
26. Center F ~~H~~ The range between the lower quartile and third quartile.
27. Spread M ~~I~~ A value in the data that is much larger or much smaller than the rest of the other data values.
28. Shape N ~~J~~ A display that uses stems and leaves to show the distribution of the data.
29. 5-Number Summary B ~~K~~ A graph made up of bars that don't touch, uses categories, and frequencies to show the distribution of data

30. Box-and-Whisker Plot L

✓ Display used to show quantitative data in a 5-number summary; each section represents 25% of the data

31. Interquartile Range H

✓ Use to describe the variation or how "spread" out the data is displayed

32. Outliers I

✓ Data can be described by symmetric, skewed left, and skewed right

33. What are the 2 ways to display categorical data?

- a) Bar graph
- b) Pie chart

34. What are the 4 ways to display quantitative data?

- a) Dotplot
- b) Boxplot
- c) Histogram
- d) Stem and Leaf Plot

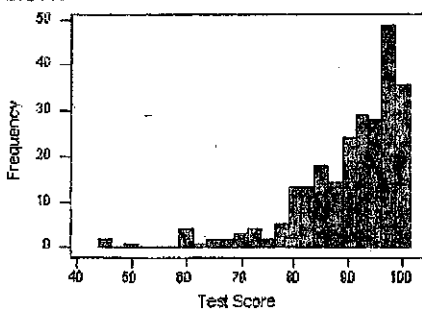
35. List the measures of center. Star (\*) the one that is affected by outliers.

- a) Median
- b) Mean \*
- c) Mode

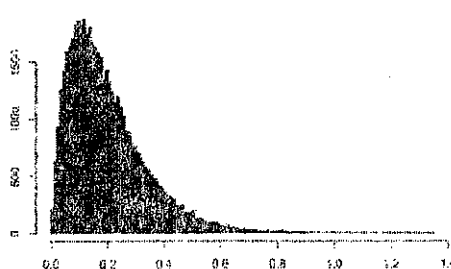
36. List the measures of spread. Star (\*) the ones that are affected by outliers.

- a) Range \*
- b) IQR
- c) St. dev \*

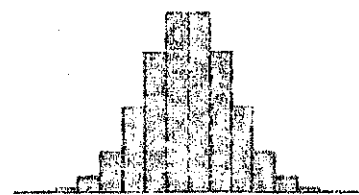
37. Describe the shape and the mean in relation to the median (equal, higher, lower) for the following displays below.



skewed left  
 mean < median  
 ↑  
 lower



skewed right  
 mean > median  
 ↑  
 higher



symmetric  
 mean = median

38. If you created a pie chart, fill out the data below.

Favorite Color	Percentage	Degree
Yellow	26	94°
Blue	14	50°
Red	20	72°
Purple	18	65°
Green	22	79°

b. What other graph could you use to display the "Favorite Color" data?

Bar graph

39. The number of points scored on the last quiz are shown below.

1<sup>st</sup> Block: 83 50 79 80 83 55 71 (42) 94 81 75 69 73 85 100

2<sup>nd</sup> Block: 99 87 (33) 67 100 (25) 75 38 96 67 83 83 79 88 70 79

a. Find the five number summary for both sets of data.

1 <sup>st</sup> Block	2 <sup>nd</sup> Block
min = 42	min = 25
Q <sub>1</sub> = 69	Q <sub>1</sub> = 67
med = 79	med = 79
Q <sub>3</sub> = 83	Q <sub>3</sub> = 87.5
max = 100	max = 100

b. Find the IQR for both sets of data.

1 <sup>st</sup> Block	2 <sup>nd</sup> Block
<del>83</del> 83 - 69 = 14	87.5 - 67 = 20.5

c. What are the outlier formulas?

$$LF: Q_1 - 1.5(IQR)$$

$$UF: Q_3 + 1.5(IQR)$$

d. Prove any outliers in each class with the formulas.

1 <sup>st</sup> Block	2 <sup>nd</sup> Block
LF: 48 → yes 42	LF: 36.25 → yes 33, 25
UF: 104 → no	UF: 118.25

e. Calculate the mean for each class.

1 <sup>st</sup> Block	2 <sup>nd</sup> Block
74.67	73.06

f. Calculate the standard deviation for each class.

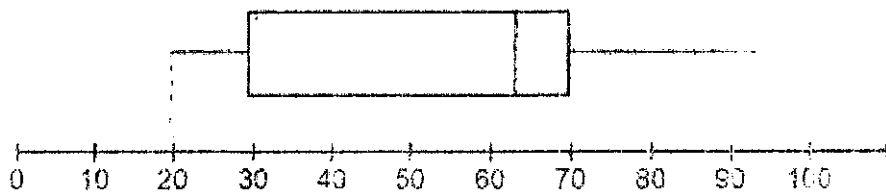
1 <sup>st</sup> Block	2 <sup>nd</sup> Block
15.72	22.86

g. If we added 5 points to every score, what would this do to the mean, median, and standard deviation? Calculate the new mean, median, and standard deviation.

1 <sup>st</sup> Block	2 <sup>nd</sup> Block
mean = 79.67 Sx = 15.72 med = 84	mean = 78.06 Sx = 22.86 med = 84

40.

Mrs. Sanchez's Period 1 Math Quiz Results



- The median grade on the quiz was around: 63
- The minimum grade on the quiz was around: 20
- What percentage of students made lower than a 70%? 75%
- What percentage of students made higher than 63%? 50%
- If 28 students took the quiz, **how many** scored lower than a 29%?  $.25(28) = 7$  students
- If 32 students took the quiz, **how many** scored higher than a 70%?  $.25(32) = 8$  students
- The IQR for the quiz results was around: ~~70~~ 29, 41

**Unit 3 – Data Displays**

41. Percentile C

~~A~~ The normalized value of an observation; shows you how many standard deviations an observation is away from the mean.

42. Z-Score A

~~B~~ Also known as the "68-95-99" Rule; this 'rule' shows how much percentage

43. Normal Distribution E

~~C~~ The percentage that an observation is less than or equal to a certain value.

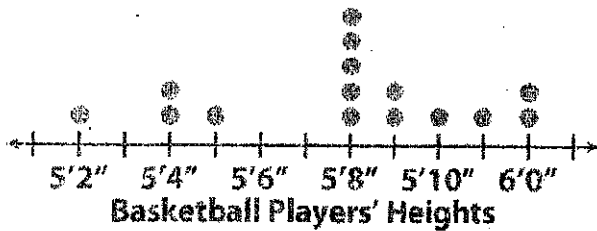
44. Empirical Rule B

~~D~~ The chance of something happening

45. Probability D

~~E~~ This helps show an overall trend when gathering data; it quickly shows if an observation or data point is an outlier or 'abnormal' by using the mean and standard deviation.

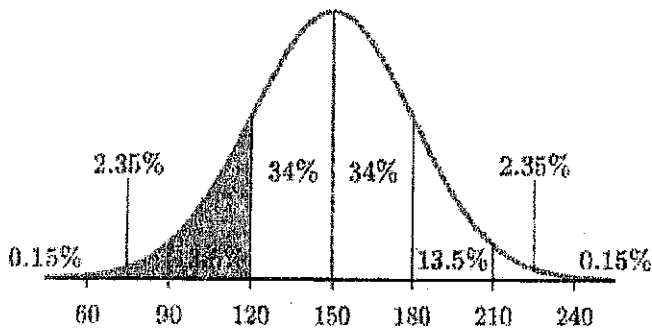
46. What percentile is a 5'8" player in? What does this mean?



$$\frac{9}{15} = 0.6 = 60\%$$

60% of the basketball players are 5'8" or less.

47. Using the Normal Curve below about average weight of Middle School boys, answer the following questions:



a. What is the mean weight of a Middle School boy? 150 lbs

b. What is the standard deviation of a Middle School boy? 30 lbs

c. What percentage of Middle School boys have a weight lower than 180 lbs? 84%

d. What percentage of Middle School boys have a weight higher than 210 lbs? 2.5%

e. What weight separates the lowest 16%? ~~120~~ 120 lbs.

f. What weight separates the highest 0.15%? 240 lbs

g. If 200 boys were surveyed, **how many** boys weight outside of 90 and 240 lbs?

$$200(.0265) = 5.3 \rightarrow 5 \text{ boys}$$

48. How much percentage of data should be between the following standard deviations on a normal curve?

a.  $\pm 1$  standard deviations = 68%

b.  $\pm 2$  standard deviations = 95%

c.  $\pm 3$  standard deviations = 99.7%

49. Your SAT score was a z-score of -1.5. What does that mean?

Your SAT score was 1.5 standard deviations below the mean.

50. What is the z-score formula? 
$$z = \frac{X - \bar{X}}{S_x}$$
  
observation  
mean  
st. deviation

51. Bob made an 86 on his Stat test that had a mean of 88 and a standard deviation of 5. Sally scored a 78 on her Biology test that had a mean of 84 and a standard deviation of 4. Who did relatively better?

$$\text{Bob} \Rightarrow \frac{86-88}{5} \\ z = -0.4$$

$$\text{Sally} \rightarrow \frac{78-84}{4} \\ z = -1.5$$

Bob did relatively better

52. The next three questions use the normal standard table with the following information: The distribution of Statistic final grades is approximately normal with mean 72 and standard deviation 2.5.

a. What percentage of students' final grade was a 62 or lower?

$$z = \frac{62-72}{2.5} = -4 \quad 0\%$$

b. What percentage of students' final grade was an 80 or higher?

$$z = \frac{80-72}{2.5} = 3.2 \quad 1 - .9993 = .0007 \Rightarrow .07\%$$

c. What percentage of students' final grade was between a 70 and 79?

$$z = \frac{70-72}{2.5} = -0.8 \quad z = \frac{79-72}{2.5} = 2.8 \quad p = .9974 - .2119 \\ p = .7855 \Rightarrow 78.55\%$$



d. What grade did a student get if they have a z-score of 2.1?

$$2.1 = \frac{x-72}{2.5} \Rightarrow 5.25 = x-72$$
$$77.25 = x$$

e. What grade separates the **top** 10%?

$$P = 1 - 0.10 = .90$$

↓

$$Z = 1.28$$

$$1.28 = \frac{x-72}{2.5} \Rightarrow 3.2 = x-72$$

$$75.2 = x$$

f. What grade is in the **bottom** 7.21%?

$$P = .0721$$

↓

$$Z = -1.46$$

$$-1.46 = \frac{x-72}{2.5} \Rightarrow -3.65 = x-72$$

$$68.35 = x$$

### Unit 3.5 - Hypothesis Testing

53. If  $\alpha = 0.01$  and  $p = 0.04$ , what would you do with the null hypothesis? (reject or fail to reject)

fail to reject

54. If  $\alpha = 0.05$  and  $p = 0.04$ , what would you do with the null hypothesis? (reject or fail to reject)

reject

55. Bjorn claims that his son can eat 4 cookies at once. His wife saw that his son ate only 1 cookie yesterday. What conclusion would you draw if the significance level is set at 0.03 and a standard deviation of 1.5?

$$z = \frac{1-4}{1.5} = -2$$

$$p = .0228$$

Reject the null hypothesis;  
we do have enough evidence  
to show Bjorn's son eats less  
than 4 cookies