Name: Key

Date:

Algebra

Data Analysis Review Packet

Section 1: Histograms

1. What kind of data do we need to make a histogram?

Univariate!

2. What is the definition of the type of data you identified in number 1?

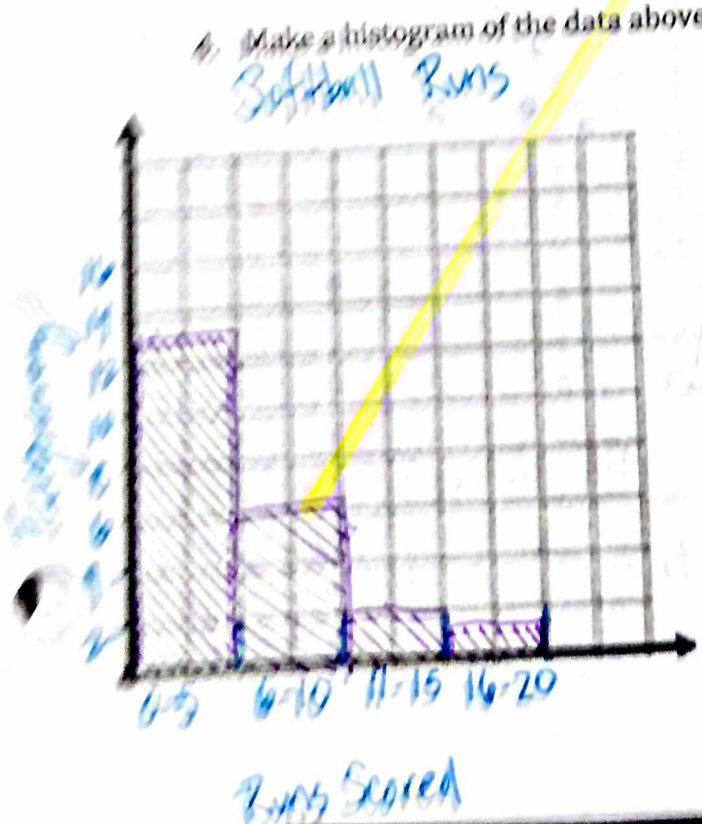
only one question is being asked/answered.

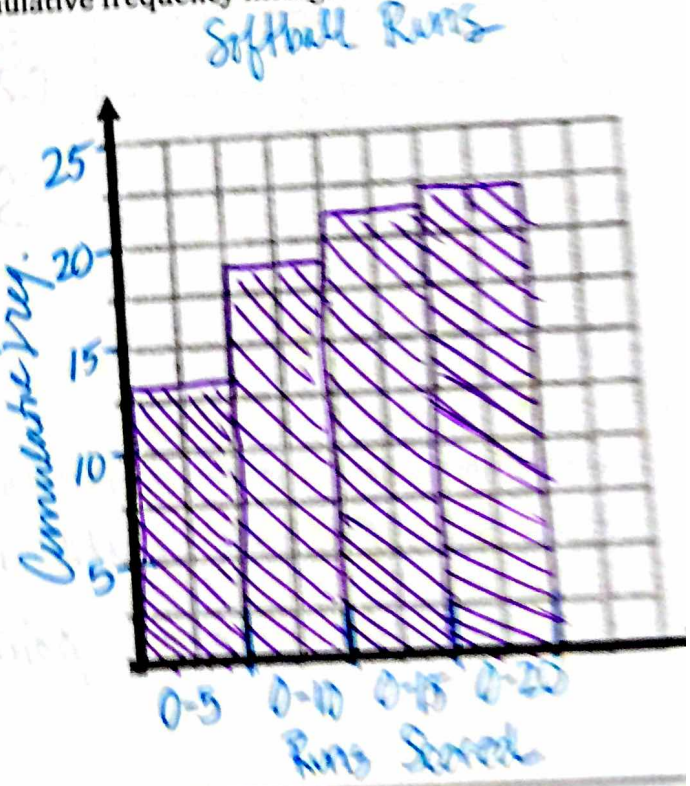
3. The following data set are the number of runs scored by the softball team. Make a table of the data including a tally column, a frequency column and a cumulative frequency column.

~~4, 10, 2, 9, 6, 4, 5, 7, 8, 1, 18, 4, 3, 7, 13, 15, 8, 9, 1, 3, 5, 2.~~

Runs Scored	Tally	Frequency	Cum. Freq.
0-5	~~𝍸~~ ~~𝍸~~ III	13	13
6-10	~~𝍸~~ I	6	19
11-15	II	2	21
16-20	I	1	22

4. Make a histogram of the data above and a cumulative frequency histogram.

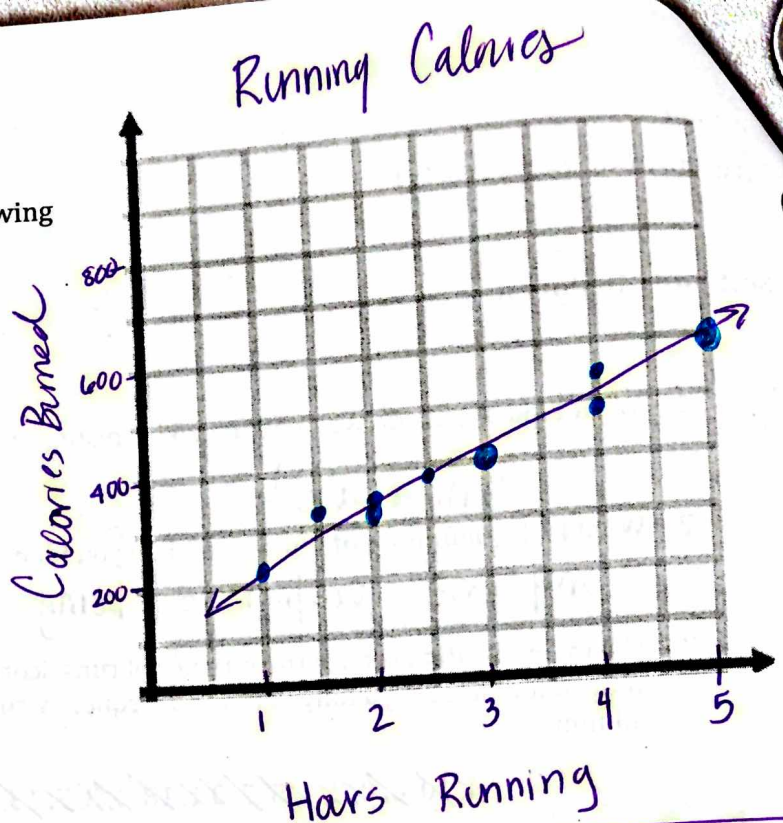




Section 2: Scatterplots

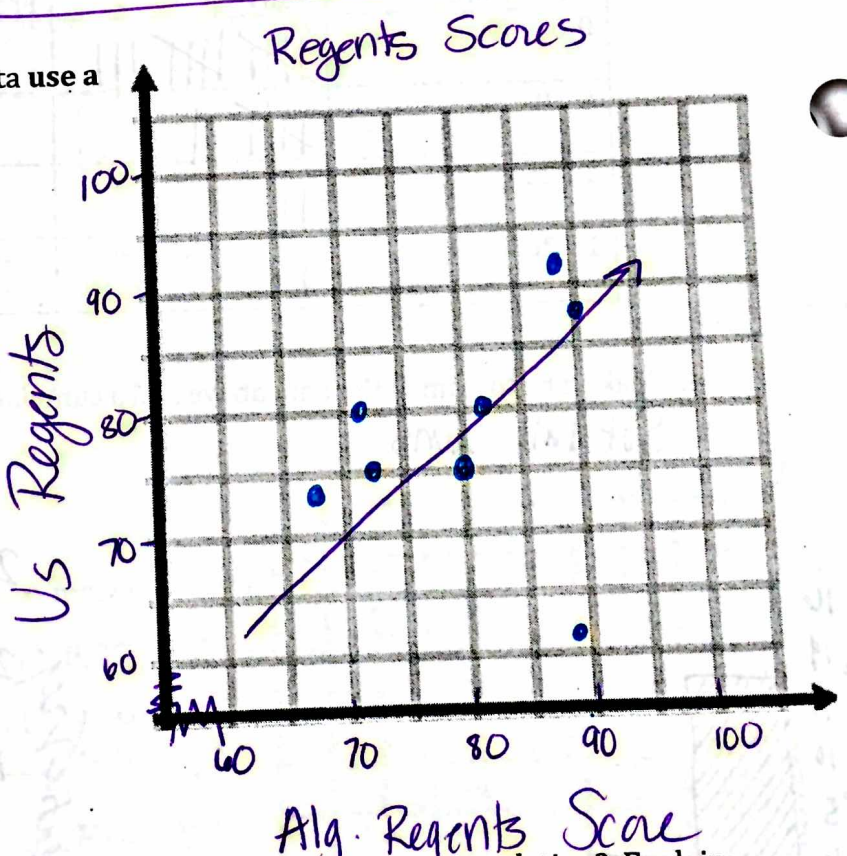
1. Make a scatterplot of the following data:

Hours Running	Calories Burned
3	400
2	350
1.5	290
4	550
1	210
2.5	380
3	420
2	330
4	490
5	600



2. Make a scatterplot of the following data **use a break in scale on both axis:**

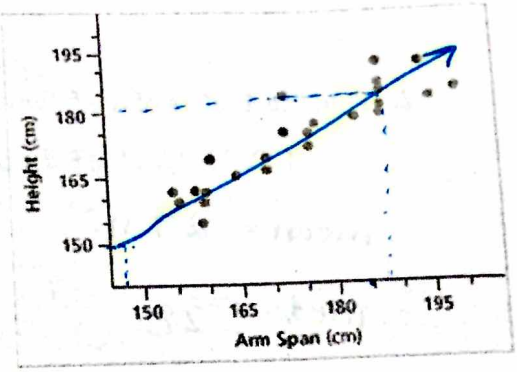
Score on Alg. Regents	Score on US Regents
82	80
89	61
72	75
90	88
71	80
67	73
88	91
80	75



3. Draw a line of best fit on both graphs. Which graph shows a stronger correlation? Explain:

The hours/calories data has stronger correlation because the data points are close to the lobf.

Section 3: Best Fit Line and correlation.



1.

a. Examine the scatterplot to the right. Draw in a line of best fit. What kind of correlation do you see?

b. If someone's arm span is 187.5 cm, approximately how tall are they? ~ 180 cm

c. If a person is 150 cm tall, approximately what is their arm span? ~ 147

d. If you were to approximate the R value as 1, .5, 0, -.5 or -1 which is the best choice? Why?

probably 0.5 - it's definitely positive b/c the correlation is positive, but it can't be 1 b/c the data points are not exactly on the line.

2. You are studying a data set on the amount of time students spent on their cell phones during class and test scores. The data set has strong negative correlation. What does that mean? What happens to a student's test score as the time they spend on their cell phone during class increases?

A strong negative correlation means the more time a student spends on their phone in class the lower their test score will be.

3. Describe the difference between an R value of -.8 and +.8.

one has positive correlation (+.8) and one has negative correlation (-.8).

Section 4: Measures of Central Tendency

1. Define the following mathematical terms in non-mathematical words:
   a. Mean: The average
   b. Median: The middle #
   c. Mode: The most common #
   d. Range: The spread of the data.

2. Find the 4 MCTs of the following data set:

{22, 29, 22, 21, 33, 34, 35, 19, 19, 22, 28, 32, 28, 39, 32, 29, 17, 22, 29}

mean = 27.11 mode = 22

med = 28 Range = 22

3. Find the MCT for the following data set:
1.3, 0.6, 1.8, 2.2. 2.3, 0.9, 1.1, 1.0, 1.8, 2.0, 0.7

mean: 1.43 mode = none

med = 1.3 Rang = 1.7

Section 5: Shifts and Outliers

1. If the data set from number two in section 4 were to incur a shift of plus 5, how would the MCT be effected?

mean: 32.11 , med: 33 , mode: 27 , Rang: 22

2. A data set has mean = X and range = Y. If the data set incurs a shift of minus 10, what will the new mean and ranges be?

Mean: X-10 , Range: y

3. If we were to add the number 4 to the data set in number 2 in section 4, what would happen to the MCT? Write a sentence for each MCT.

mean- be pulled down by the outlier

med- go down 1 #

mode- stay the same

Range- increase alot from the outlier.

4. The following data set are the scores that Jessie earned on her English quizes. Find the MCT and identify any outliers in the data set. How did these outliers effect the MCT.

89, 94, 68, 92, 90, 88, 96, 83

mean: 87.5

med: 89.5

mode: None

Range: 28

68 is an outlier b/c its far away from the rest of the data. The outlier brought down the mean

Section 6: 5 Number Summary

1. What are the 5 numbers in the 5 number summary. What do they mean?

min , Q_1 , med, Q_3 , max

2. Find the 5 number summary for the following data set:

9, 7, 2, 12, 8, 14, 11, 10, 5, 6, 14

min = 2 , Q_1 = 6 , med = 9 , Q_3 = 12 max = 14

3. If a data set has a maximum value of 44, which of the following is a possible value for the third quartile?
   a. 44
   b. 45
   c. 54
   d. ~~24~~

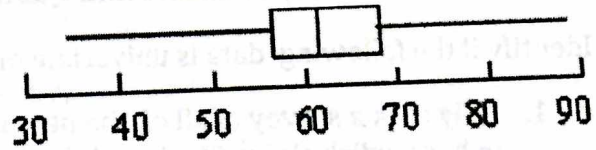
4. If a data set has a first quartile value of 23 and a maximum of 28, what is a possible value for the thrid quartile of the data set?

25 (any # between 23+28)

5. What is the interquarterile range IQR for the data set in number 2?

IQR = 12 − 6 = 6

Section 7: Box and Whisker Plots



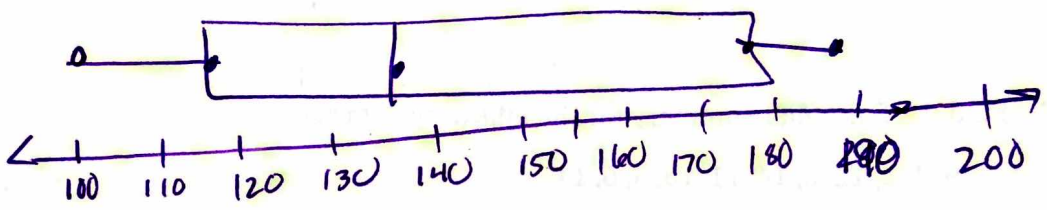
1. Study the scatterplot below of the number of calories in 20 different health bars and answer the questions below.

a. Approximately 75 % of the bars have less than 68 calories.
b. Approximately 100 bars have more than 33 calories.
c. The max number of caloires in the bars is 89.
d. The range of the number of caloris for the data set is about 56.

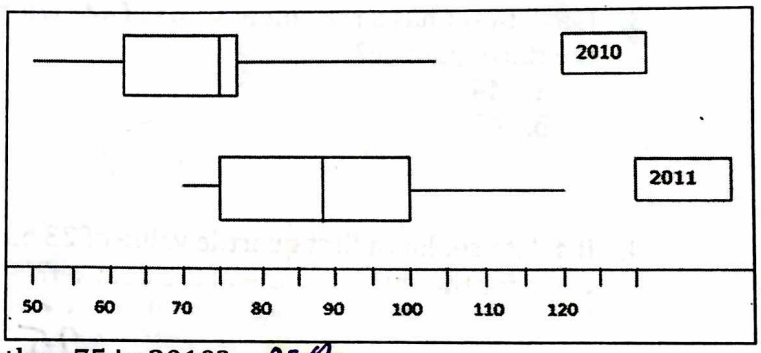
89 − 33

2. Construct a box and whisker plot for the following data set, state the 5#S.
100, 183, 117, 142, 134, 104, 174, 138, 122, 179

min = 100
Q_1 = 117
med = 136
Q_3 = 174
max = 183



3. The comparison box and whisker plot shows the scores for the chemistry regents in 2010 and 2011.



a. Did students do better in 2010 or 2011?

2011, the 2011 graph is generally higher in 2011.

b. What percentage of students scored higher than 75 in 2010? 25%

What percentage of studetns scored higher than 75 in 2011? 75%

c. Which year has a higher IQR? 2011 Which year has a larger range? 2010

120 - 67 = 53
105 - 50 = 55

Section 8: Univariate/Bivariate and Quanitative/Qualitative

Identify if the following data is univariate or bivariate and if it is Quantitative or Qualitiative.

1. Sally took a survey of all of the people in her english class. She asked them how much money they spent on lunch that day.

Univariate, Quantitative

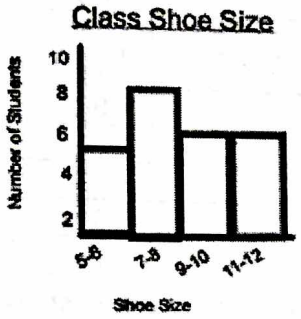
2.

Fav. Color	Frequency
Red	21
Blue	28
Green	37
Purple	25
Pink	18

Univariate, Qualitative

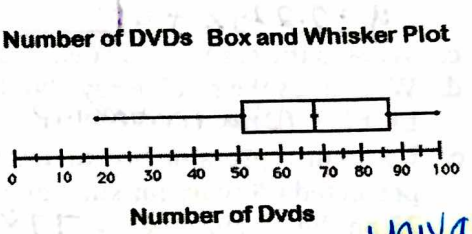
3. Miss P did a survey about how far away students live from school and how long it takes them to get here in the morning. bivariate, quantitative

4.



univariate, Quantitative

5.



univariate, Quantitative

Section 9: Bias

1. You are doing a survey about animal rights. Identify the following as Biased or Unbiased and a survey method or survey question:
   a. Asking every other person who comes out of a library. unbiased method
   b. Asking " Do you think animal rights are imporant?" . unbiased question
   c. Asking every other person who comes out a pets store. . biased method
   d. Asking " don't you think baby animals are the cutest?!" . biased question

2. You are doing a survey about peer tutoring.
   a. Give an example of an unbiased survey method.
   ~~asking kids in office hours~~ asking 2 kids in each advisory.
   b. Give an example of a biased survey method.
   asking students in office hours
   c. Give an example of an unbiased survey question
   "Do you think we should start a peer tutoring group?"
   d. Give an example of a biased survey question
   "Would yu ever attend a nerdy peer tutory session?"

Section 10: Data Analysis with a Calculator

1. Consider the regents data:
   a. Enter the data into your calculator and run a linear regression.
   b. What is the equation of the regression line: $y = 0.224x + 60$
   c. What is the corellation coefficient? .215
   d. What does the r-value say about the line of best fit? not a relationship
   e. Use your regression line to determine the predicted US score for someone who scored a 75 on the Algebr regents 77.8
   f. Use your regression line to predict the algebra score for someone who scored an 79 on the US exam: 81

Score on Alg. Regents	Score on US Regents
82	80
89	61
72	75
90	88
71	80
67	73
88	91
80	75

2. The following data table shows the number of bacteria in a petri dish after a certain number of hours.

Hours	1	2	3	4	5
Bacteria	10	12	17	21	28

   a. Use your calculator to find an **exponential** model that represents the bacteria growth: $y = 7.49 \cdot 1.30^x$

   b. At this growth rate, how many bacteria will there be after one full day?

about 4066

3. The following list is the number of text messages a group of 20 students send in a day.

{ 18, 22, 93, 50, 52, 38, 21, 8, 29, 35, 53, 47, 21, 30, 51, 38, 29, 40, 50, 27}

   a. Enter the data into your calculator and run the 1-var stats
   Mean: 37.65 Med: 36.5 Mode: 38, 50 Range: 85

   Min: 8 Q1: 24.5 Med: 36.5 Q3: 50.3 Max: 93

   b. Is the Mean or the Median a better representation of the data set? Explain why:

the median b/c of the outlier (93)