Two-way Frequency Tables

- two way frequency table- a table that divides responses into categories.
- Joint relative frequency- the number of times a specific response is given divided by the sample.
- Marginal relative frequency- the total number of times a specific response is given divided by the sample.
- Conditional relative frequency- the percent of a joint frequency compared to the subtotal.

★ often indicated by the word "given."

Ex.1 constructing tables.

	High School Diploma	Bachelor's Degree	Master's/ Doctoral Degree	Total
Male	16	46	3	65
Female	12	51	3	66
Total	28	97	6	131

a) find the joint relative frequency of males who have a bachelors degree.

$$\frac{46}{131} = .35 = 35\%$$

b) find the marginal frequency of people with a masters/doctors degree.

$$\frac{6}{131} = .05 = 5\%$$

c) given someone is male, what's the probability of having a high school diploma?

$$\frac{16}{65} = .25 = 25\%$$

Ex.2 two way relative frequency tables

What is your favorite sport to watch on television?				
	Football	Basketball	Baseball	total
Males	40	22	15	77
Females	12	16	45	73
Total	52	38	60	150

a) what percent of girls prefer basketball?

$$\frac{16}{73} = .22 = 22\%$$

b) what is the probability of liking football?

$$\frac{52}{150} = .35 = 35\%$$

c) what's the probability of liking basketball or baseball?

98 ... $.65$ ... 65%

$$\frac{}{150} = .65 = 65\%$$

Measures of center

- is a single measure used to represent the middle value.

Median- is the middle most number

Mean- average number, it is the sum divided by the number of values.

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}$$

Mode- the number(s) that appear most often.

Maximum- the largest value in a data set.

Minimum- the smallest number in a data set.

Measures of spread

- Numbers to describe how far apart certain key values are from each other.

First quartile- median of lower half of data.

Second quartile- median.

Third quartile- median of upper half of data.

range- the difference from the minimum to the maximum.

range = max − min

range = max − min

Interquartile range- the difference between quartile 1 and quartile 3.

$$IQR = Q_3 - Q_1$$

Mean absolute deviation- the average absolute value of the difference between each data set and mean.

$$\frac{\sum_{t=1}^{n} |x_i - \bar{x}|}{n}$$

Ex.1 The April high temperatures for seven years are: ~~77~~, ~~86~~, ~~84~~, ~~93~~, ~~90~~, ~~81~~, and ~~80~~.

77, 80, 81, 84, 86, 90, 93

min $\quad Q_1 \quad Q_2 \quad Q_3 \quad$ max

$$\bar{x} = \frac{77+80+81+84+86+90+93}{7}$$

$$\bar{x} = 84.4$$

range = max − min
= 93 − 77
= 16

$$IQR = Q_3 - Q_1$$
$$= 90 - 80$$
$$\boxed{IQR = 10}$$

$$\frac{\sum_{t=1}^{n} |x_i - \bar{x}|}{n}$$

$|77 - 84.4| = 7.4$

$|80 - 84.4| = 4.4$

$|81 - 84.4| = 3.4$

$|84 - 84.4| = 0.4$

$|86 - 84.4| = 1.6$

$|90 - 84.4| = 5.6$

$|93 - 84.4| = +8.6$

31.4

$$\frac{31.4}{7} = 4.5$$

$$\boxed{MAD = 4.5}$$

Outliers

- Data values that are much less/greater than most of the data set.

extreme values- are values that appear to be outliers.

Steps to determine if a value is an outlier.

- A data value is an outlier if it is less than...
  $Q_1 - IQR(1.5)$
- A data value is an outlier if it is greater than
  $Q_3 + IQR(1.5)$

Ex.1 list of salaries in thousands:

25, 30, 35, 35, 35, 40, 40, 40, 45, 45, 50, 60, 150

35 Q_1 | Q_2 | 47.5 Q_3 | outlier

$IQR = Q_3 - Q_1$
$= 47.5 - 35$

$\boxed{IQR = 12.5}$

$Q_1 - IQR(1.5)$
$35 - (12.5)(1.5)$
16.25

$\boxed{Q_2 = 40 \quad \bar{x} = 48.5}$

$Q_3 + IQR(1.5)$
$47.5 + (12.5)(1.5)$
66.25

★ When there's an outlier, use the median as the measure of center.

★ Outliers affect the mean and range.

Graphs

Box plot- is a graph that shows the minimum, maximum, Q1, Q2, and Q3.

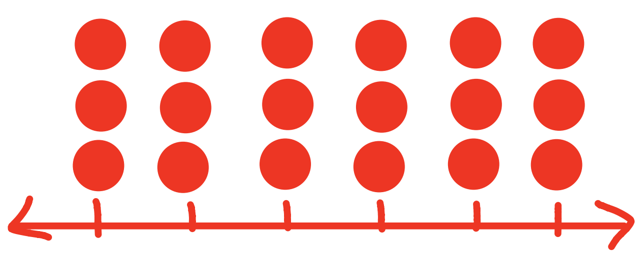
Dot plot- is a graph that uses dots to show the number of times each value in a data set appears in the data set.

Histogram- a bar graph which shows frequency distribution.

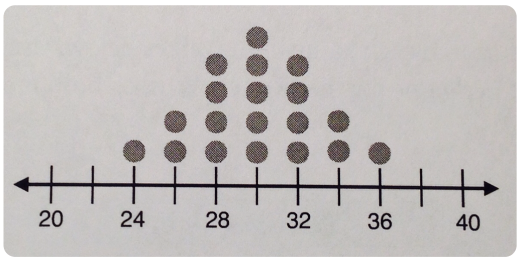
1. Divide the range into even sections.
2. Tally each frequency.

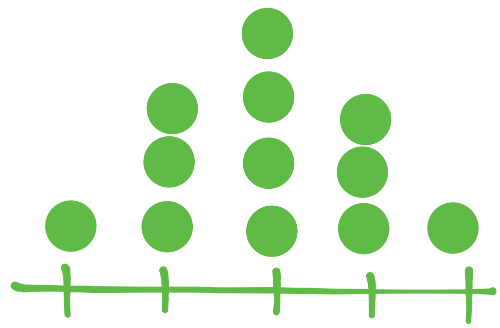
Shape

Uniform- data is evenly distributed.

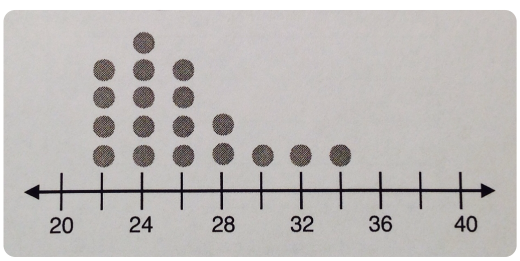


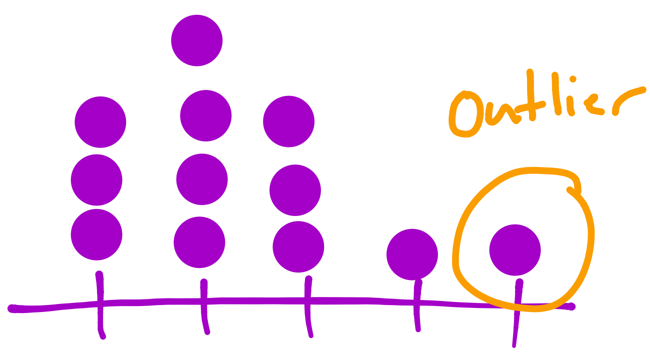
Symmetric-data is centered toward the middle. Also known as normal distribution.



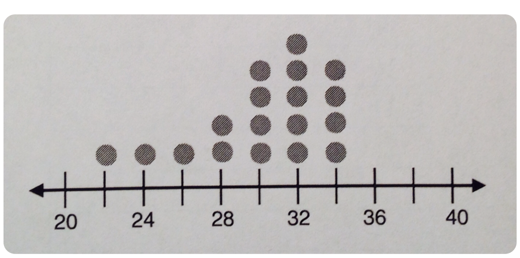


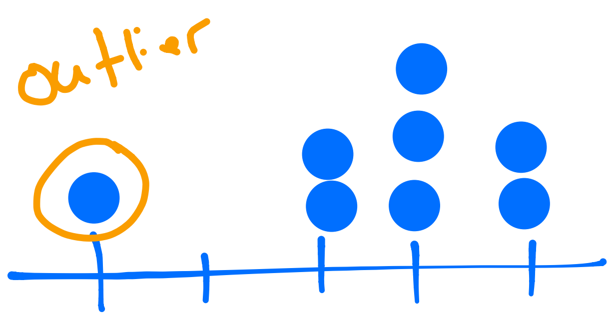
Skewed to the right- describes where the outlier is pulling the data.





Skewed to the left- the outlier(s) is pulling the data left.





Ex.1 Find any outlier then create a dot plot, box plot, and histogram. Then describe the shape.

2, 2, 9, 10, 10, 11, 11, 11, 12

min | 5.5 Q_1 | Q_2 | 11 Q_3 | max

$Q_1 - IQR(1.5)$

$5.5 - 5.5(1.5)$

-2.75

$IQR = Q_3 - Q_1$

$IQR = 5.5$

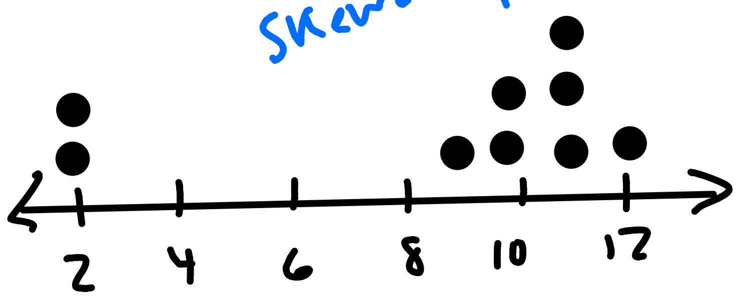
$Q_3 + IQR(1.5)$

$11 + 5.5(1.5)$

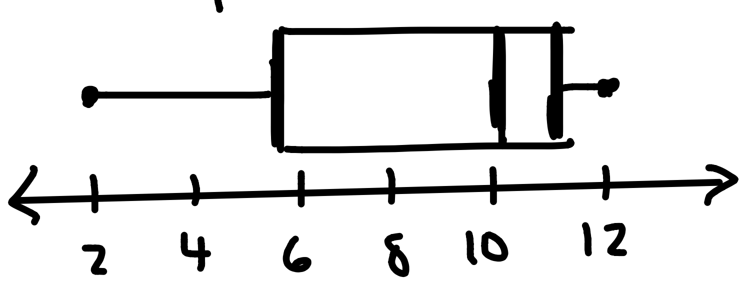
19.25

Dot Plot

skewed left



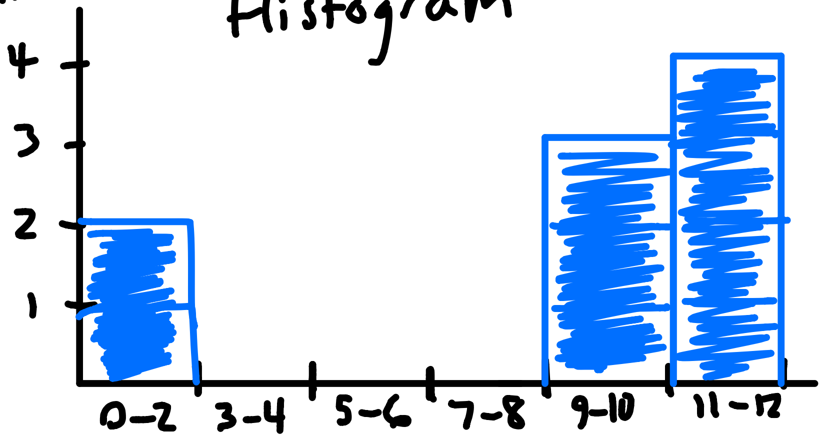
box plot



range = max − min
= 12 − 2
= 10

$\frac{10}{5} = 2$

Histogram



Name ______________________ Date ______________ Class ______________

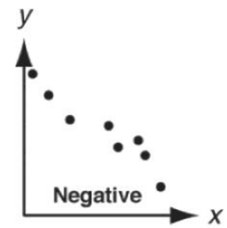
Scatter Plots and Trend Lines Notes

Correlation is one way to describe the relationship between two sets of data.

Positive Correlation

Data: As one set **increases,** the other set **increases.**

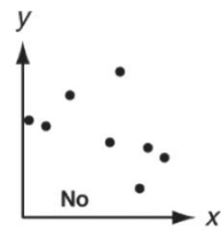
Graph: The graph **goes up** from left to right.



Negative Correlation

Data: As one set **increases,** the other set **decreases.**

Graph: The graph **goes down** from left to right.

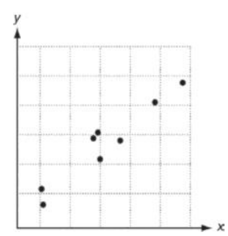


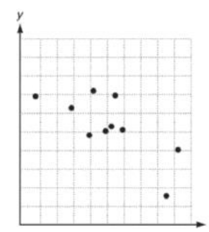
No Correlation

Data: There is **no relationship** between the sets.

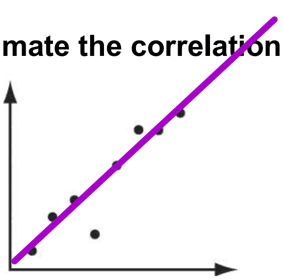
Graph: The graph has **no pattern.**

Example	Correlation	Correlation Coefficient (*estimated*)
1st graph *above*	strong positive	+1
2nd graph *above*	strong negative	–1
3rd graph *above*	no correlation	0
4th graph *beside*	weak positive	+0.5
5th graph *beside*	weak negative	–0.5



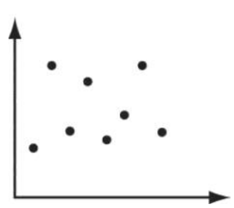


Estimate the correlation coefficient for each scatter plot as –1, –0.5, 0, 0.5, or 1.

1. 

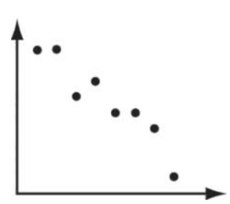
positive correlation

r= .7

2. 

No correlation

r=0

3. 

negative correlation

r=-.8

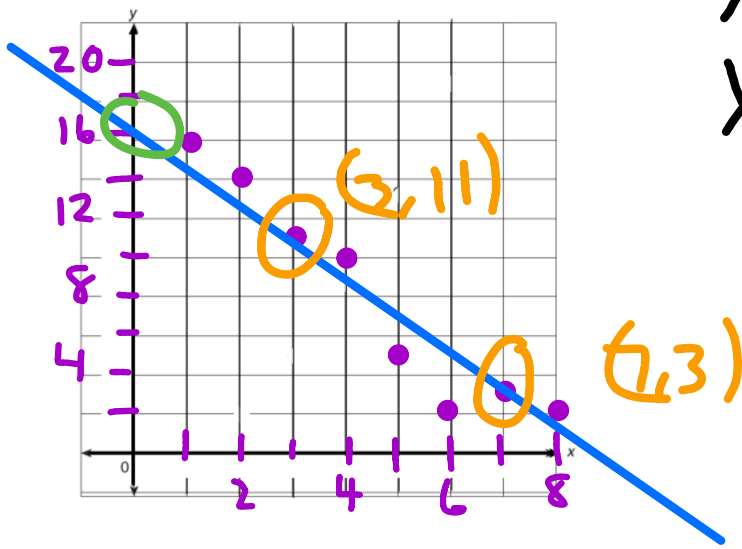
Name ________________________________ Date _______________ Class_______________

Fitting a Linear Model to Data Notes

The table shows the relationship between two variables. Identify the correlation, sketch a line of fit, and find its equation.

x	1	2	3	4	5	6	7	8
y	16	14	11	10	5	2	3	2

Step 1 Make a scatter plot of the data.



$y=ax+b$

$y=-2.2x+17.8$

$r=-.96$

Strong negative correlation

Step 2 Use a straightedge to draw a line.
There will be some points above and some below the line.

Step 3 Choose two points on the line to find the equation:

Step 4 Use the points to find the slope:

$$m=\frac{y_2-y_1}{x_2-x_1}=\frac{11-3}{3-7}=\frac{8}{-4}=\boxed{-2}$$

Step 5 Find the y-intercept:

$b=16$

Step 6 Write the equation:

$y=mx+b$ $\boxed{y=-2x+16}$