

**Section 17.1: Descriptive Statistics (Day 1)**

**Descriptive Statistics**

**Inferential Statistics**

**STEM-and-LEAF PLOT**

<b>TEAM</b>	Padres	Cubs	Brewers	Astros	Dodgers	Cardinals	Diamondbacks	Reds
<b>At Bats (AB)</b>	103	61	98	102	71	74	98	70
<b>Team</b>	Pirates	Phillies	Mets	Braves	Rockies	Nationals	Giants	Marlins
<b>AB</b>	69	74	65	64	104	70	106	77

Ex 1: Summarize data in a stem-and-leaf plot.

**SAT mathematics scores for 30 seniors**

480 570 670 580 500 570 600 540 690 500  
520 450 570 540 620 450 460 690 580 610  
550 630 460 470 510 630 510 620 470 630

Ex 2: Summarize data in a stem-and-leaf plot.

## AVERAGES – MEASURES of CENTRAL TENDENCY

**Mean** of a set of data is the sum of the data divided by the number of items of data; denoted \_\_\_\_\_

**Median** of a set of data is the middle number of the data in arranged order

**Mode** of a set of data is the item that occurs most often

If there are two modes \_\_\_\_\_

If there are three modes \_\_\_\_\_

If there are more than 3 \_\_\_\_\_

Ex 3: Find the mean, median, and mode for the following set of data.

### **Minutes to read a History test**

1		2 2
2		0 0 0 8
3		0 2 2 3 8 8 9
4		0 0 2 2 4 8
5		2 2 4 4 9

1 | 2 represents 12 minutes

Ex 4:

a) Find the mean, median, and mode for the following set of data.

2      37      41      44      45      46

b) Does the mean represent the data well? Explain.

c) Which measure of central tendency is the best for this set of data?

**Section 17.1: Descriptive Statistics (Day 2)**

**FREQUENCY and RELATIVE FREQUENCY TABLES**

**WORDS PER MINUTE**

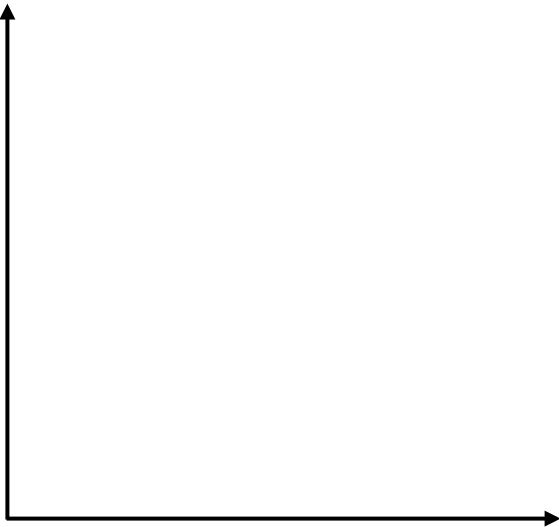
Class	Frequency	Cumulative Frequency
100-149	3	
150-199	14	
200-249	21	
250-299	10	
300-349	2	

**WORDS PER MINUTE**

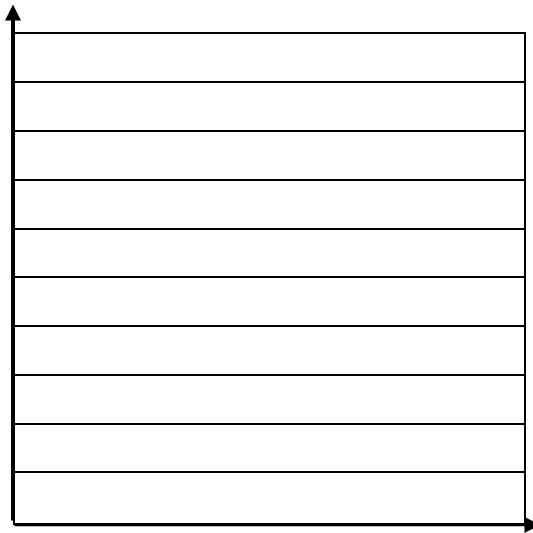
Class	Relative Frequency	Relative Cumulative Frequency
100-149		
150-199		
200-249		
250-299		
300-349		

**HISTOGRAM** – bars are next to each other (NO GAPS)

*Horizontal Bars*



*Vertical Bars*



**FREQUENCY POLYGON**

Vertical Axis:

Horizontal Axis:

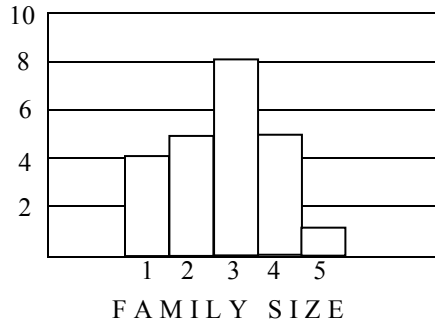
*Polygon begins and ends on horizontal axis*





**Ex 6:**

**Find the three measures of central tendency for the given histogram.**



**Section 17.2: Box and Whisker Plots**

**Example1:** Construct a box-and-whisker plot for the following data.

**Ages of first 41 presidents at the time each took office**

4	2 3
*	6 7 8 9 9
5	0 1 1 1 1 1 2 2 4 4 4 4
*	5 5 5 5 6 6 6 7 7 7 7 8
6	0 1 1 1 2 4 4
*	5 8 9

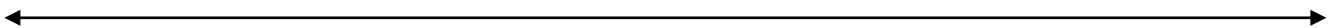
Find the **lower and upper extreme values**. Lower = \_\_\_\_\_ Upper = \_\_\_\_\_

Find the **median**. Median = \_\_\_\_\_

Find the **lower quartile and upper quartile**.

- Do NOT include the \_\_\_\_\_
- Lower Quartile is the median of the \_\_\_\_\_
- Upper Quartile is the median of the \_\_\_\_\_

Lower quartile = \_\_\_\_\_ Upper quartile = \_\_\_\_\_



**Range:** is the \_\_\_\_\_ between the \_\_\_\_\_ and the \_\_\_\_\_

**Interquartile range:** (IQR) is the \_\_\_\_\_ between the \_\_\_\_\_ and the \_\_\_\_\_

**Outliers:** Any data whose distance is \_\_\_\_\_ than  $(1.5 \times \text{IQR})$  away from the UQ and \_\_\_\_\_ than  $(1.5 \times \text{IQR})$  away from the LQ

## Section 17.3: Variability

### Definitions

**Statistic:** is a \_\_\_\_\_ that describes some \_\_\_\_\_ of a set of data

**Mean, Median, Mode:** describe the \_\_\_\_\_ of a set of data

**Range and Interquartile Range:** describe the \_\_\_\_\_ of the data about the center or called \_\_\_\_\_ of \_\_\_\_\_ (how much data varies)

### Example 1:

The following 3 classes have the given test scores.

Class 1

50, 50, 50, 50, 50

Mean =

Range =

Class 2

40, 50, 50, 50, 60

Mean =

Range =

Class 3

10, 40, 50, 60, 90

Mean =

Range =

**OTHER MEASURES OF DISPERSION:** \_\_\_\_\_ and \_\_\_\_\_

**Variance:** Written as \_\_\_\_\_ (sigma squared)

Data:

Mean =

**Standard Deviation:** Written as \_\_\_\_\_

\*It is how much items are dispersed around the mean – NOT an average!

### Example 2:

Five pairs of shoes cost \$10, \$20, \$30, \$40 and \$50.

**Example 3:**

Find the standard deviation for 1, 7, 9, 15

**Example 4:**

Find the mean and standard deviation from the table of data.

$x_i$ = item of data	4	6	10
$f_i$ = frequency	2	5	3

# of distinct data items =

# of total data items =

**Standard Value** (z) of a piece of data

Standard Value = \_\_\_\_\_

OR  $z =$  \_\_\_\_\_

Z give you the number of standard \_\_\_\_\_ between an item of the data and the mean  
(Also called \_\_\_\_\_ or \_\_\_\_\_ score)

**Example 5:**

Use data from example 2 to find the standard value if you had a score of...

1, 7, 9, 15

Score of 1

Score of 7

Score of 9

Score of 15

**Section 17.4: Normal Distribution**

**Types of Distribution**

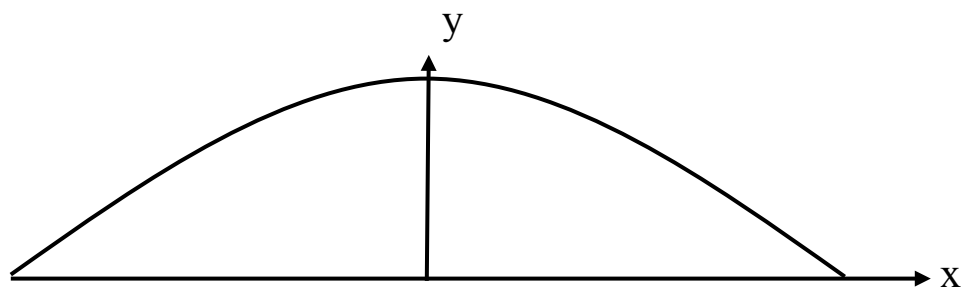
- 1) Uniform                                      2) Skewed                                      3) Normal Distribution

**Example 1:**

Average weight of a newborn baby is 3.6 kg and  $\sigma = 0.5$  kg. Sketch a normal curve showing weights at 1, 2, and 3  $\sigma$  from the mean.

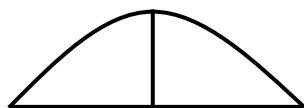
**STANDARD NORMAL DISTRIBUTION**

- ‡ Is a normal distribution with a mean = \_\_\_\_\_ and  $\sigma =$  \_\_\_\_\_
- ‡ Any normal distribution can be related to it using standard values (z-scores)
- ‡ Total area under curve = 1.0
- ‡ Area under curve to the \_\_\_\_\_ of a z-score = proportion of the data having standard values \_\_\_\_\_ than z



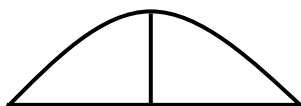
**Example 2:**

a)  $P(2.0) =$



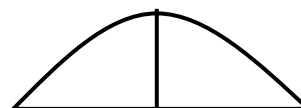
Means area to the left of 2 is \_\_\_\_\_

b)  $P(-1.3) =$



Means the area to the left of  $-1.3$  is \_\_\_\_\_

c)  $P(0.7) =$



Means area to the left of 0.7 is \_\_\_\_\_



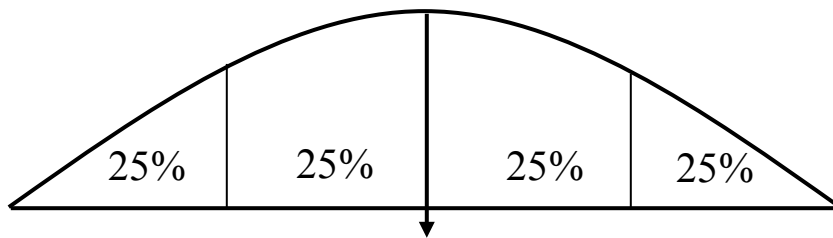
**Example 3:**

At a track meet, the race times for 100m are normally distributed with mean of 11.8 s and  $\sigma = 0.5$  s.

- a) What percent of entrants finished with a time greater than 12 s?
  
  
  
  
  
  
  
  
  
  
- b) If the top 5% received awards, what race time did an entrant need to get an award?
  
  
  
  
  
  
  
  
  
  
- c) What percent of entrants had times between 11.5 s and 12.1 s?

**PERCENTILES**

- ↳ Used for standardized tests
- ↳ 75% percentile means 75% of people taking test had a score \_\_\_\_\_ than or equal to your score



**Example 4:**

The scores on a test are normally distributed with mean = 70 and  $\sigma = 10$ . Find the 90<sup>th</sup> percentile.

## Section 17.5: Inferential Statistics

**Population:** the entire \_\_\_\_\_ of individuals or objects being studied

**Sample:** the \_\_\_\_\_ of the population

**Sample Size:** the \_\_\_\_\_ of objects in the sample

**Sampling:** the \_\_\_\_\_ of selecting a sample that represents the total population

**Sampling Theory:** branch of stats that deals with \_\_\_\_\_ that come up when sample is taken

For example: How to select?

Size?

How reliable are conclusions drawn from the sample?

### **Nonprobability Sampling:**

A procedure for selecting a sample which is NOT a \_\_\_\_\_ process  
(may be biased; different characteristics from population)

- \_\_\_\_\_: the sample selected is easy and convenient
- \_\_\_\_\_: one or more experts select a representative sample based on their subjective judgment
- \_\_\_\_\_: conclusions are based on voluntary responses  
(questions could be biased)

### **Probability Sampling:**

Selecting sample based on \_\_\_\_\_ sample

- \_\_\_\_\_ **Random Sampling:**  
Every element of population has equal and independent chance of being chosen  
  
Example: flip a coin or a random generator
- \_\_\_\_\_ **Random Sampling:**  
Population broken into groups called strata; each stratum has a common property  
  
Example:  
2 strata – male group/female group  
4 strata – freshmen/sophomores/juniors/seniors

**Example 1:**

State whether the sampling procedure described is an example of...

- a. convenience sampling
- b. judgment sampling
- c. sampling by questionnaire
- d. simple random sampling
- e. stratified random sampling

1. A television rating organization mails a questionnaire to subscribers of a television magazine. On the basis of the responses to this questionnaire, it names the most popular shows.
2. A popular financial magazine would like to survey the techniques of successful managers. A managerial consultant is asked to select a group of 50 successful managers to participate in the survey.
3. A school would like to know if students like the hamburger hot dish served at lunch. A school representative stands in front of school and asks the first 200 students who enter school if they like the dish.

**Example 2:**

Maryton High School has 860 students and 56 teachers.

The school newspaper interviewed 80 students and 15 teachers to see whether they were in favor of changing the schedule of the day. The results are listed in the table.

<b>Stratum</b>	<b>Population Size</b>	<b>Sample Size</b>	<b>Number in favor of change</b>
<b>Students</b>	860	80	48
<b>Teachers</b>	40	15	8
<b>Total</b>	900	95	56

Estimate the percent of the school in favor of change.

**Section 17.6: Confidence Intervals for Surveys and Polls**

\_\_\_\_\_ = population proportion having a given characteristic

\_\_\_\_\_ = sample proportion having a given characteristic

**Example 1:**

Find  $\bar{p}$ .

- a) 10 out of 25 doctors recommend daily exercise
- b) 24 out of 250 students favor name change

**Statistic Theorem** states three things are true about the values of  $\bar{p}$ :

- 1. \_\_\_\_\_
- 2. \_\_\_\_\_
- 3. \_\_\_\_\_

NORMAL DISTRIBUTION RULE

**95% CONFIDENCE INTERVAL of p**

## 99% CONFIDENCE INTERVAL of p

### Example 2:

If  $\bar{p} = 0.4$  and  $n = 24$ , find the standard deviation of  $\bar{p}$ .

### Example 3:

Using the information from example 2, also find a 95% and a 99% confidence interval for p.

### Example 4:

Of 400 households surveyed, 144 reported watching Channel 8 news at least twice a week. Find a 99% confidence interval for p, the unknown proportion of households that watch the Channel 8 news at least twice a week.