Name: _____   Date: _____

Final Review

- ## Chapter 2: Basic Statistics

Part I:
1.   Give a 5 number summary for the following data set and interpret the results:
The weights (in pounds) of the players on a high school football team:
173   145   205   192   197   227   156   240   172   195
185   208   185   190   167   212   228   190   184

Part II:
1.   Create a frequency distribution chart for the data given in part I with a class width of 5. Then create a histogram and give the shape of the distribution.

Class	Frequency	Midpoint	Relative Frequency	Cumulative Frequency
	N=			

Part III:
1. Find mean, median, mode. Show all work.
2. Which measure of central tendency best describes the data? Prove by finding any mild and/or extreme outliers. Show all work.
3. Create a box-and-whiskers plot. Label the quartiles, minimum and maximum.

The number of wins for each Major League Baseball team in 2003:
101   95   86   77   88   90   86   83   82   49   93   77   75
106   91   86   83   79   88   87

Part IV:
    Find the population mean and population standard deviation for the following set of data. Give your answers using the correct notation. Write your answers on the lines provided.
The age of each Supreme Court justice as of August 20, 2003:
    78   83   73   67   67   63   55   70   65

Part V:
The monthly utility bills in a city have a mean of $70 and a standard deviation of $8.   Three utility bills were given as $60, $71, and $92.
        1.   Identify μ and σ.
        2.   Transform each of the three utility bills into z-scores.
        3.   Interpret the results. Are any z-scores unusual?

Chapter 5: Normal Probability Distributions

1.   Find the area to the right of *z* = -0.83.   _____

2.   Find the area of the region between $z = 1.51$ and $z = -0.34$ _____

3.   Find the probability $P(z<1.45)$. _____

4.   Find the *z*-score corresponding to 0.6443 _____

5.   Find the z-score that has 62.8% of the distribution's _____
        area to the right.

6.   The world's smallest mammal is the Kitti's hog-nosed bat, with a mean weight of 1.5 grams and a standard deviation of 0.25 grams. Assuming the weights are normally distributed, find the probability of randomly selecting a bat that weighs:
        a.   between 1.0 grams and 2.0 grams
        b.   between 1.6 grams and 2.2 grams
        c.   more than 2.2 grams

7.   The monthly utility bills in a city are normally distributed with a mean of $100 and a standard deviation of $12.
        a.   What percent of the utility bills are more than $125?
        b.   Would it be unusual for a person to have a utility bill of more than $125? Explain your reasoning.
        c.   If 300 utility bills are randomly selected, about how many would you expect to be less than $90?

8.   Robert's Golden Retriever weighs 67 pounds. The *z*-score that corresponds to that weight is 1.28. The standard deviation of the weights of all Golden Retrievers is 4.2. Find the average weight of a Golden Retriever. Round to the nearest hundredth.

9.   The annual per capita utilization of oranges in the United States can be approximated by a normal distribution with a mean of 11.7 pounds and a standard deviation of 3 pounds. What annual per capita utilization of oranges represents the third quartile? (Round your *z*-score to the nearest thousandth). Interpret your answer

10.   The weights of bags of cookies are normally distributed with a mean of 1.5 ounces   and a standard deviation of 0.085 ounces. Bags of cookies that have weights in the upper 7.5% are too heavy and must be repackaged. What is the most a bag of cookies can weigh and not need to be repackaged? (Round your *z*-score to the nearest thousandth).

12.   The duration (in minutes) of every power failure at a residence in the last 10 years   is listed below. What is the longest time that the power went out in the lowest 5% of duration? Round your *z*-score to the nearest thousandth and interpret your answer.

18	26	45	75	125	80	33	40	44	49
89	80	96	125	12	61	31	63	103	28

Chapter 3: Probability

1.   Which of the following numbers could explain the probability of a given event? Circle your answer(s).

a.) 1.2   b.) -1   c.) 0   d.) $\dfrac{4}{5}$ e.) 75%

2.   Explain the Law of Large Numbers.

3.   Two cards are selected from a standard deck without replacement. Find the probability that they are both hearts.

4.   A doctor gives a patient a 60% chance of surviving bypass surgery after a heart attack. If the patient survives the surgery, he has a 50% chance that the heart damage will heal. Find the probability that the patient survives surgery and the heart damage heals.

5.   A card is selected from a standard deck. Find the probability that it is not a king or a queen.

6.   A study on left-handed people was done. 55% of the participants were male. Given all males, 12% were left-handed. Given all females, 10% were left-handed. Create a chart to show the results of the survey.
      a. Find the probability of selecting a left-handed person or a male.
      b. Find the probability of selecting a right-handed person or a female.
      c. Find the probability of selecting not a left-handed male.

7.   A card is randomly selected from a standard deck. Find the probability that the card is between 4 and 8 (inclusive) or is a club.

8.   A random sample of 250 working adults found that 37% access the Internet at work, 44% access the Internet at home and 21% access the Internet at both home and work. What is the probability that a person in this sample selected at random accesses the Internet at home or at work?

9.   A pizza shop offers nine toppings. No topping is used more than once. In how many different ways can a three-topping pizza be formed?

10.   An area code consists of three digits. How many area codes are possible if:
      a. There are no restrictions.
      b. The first digit cannot be a 1 or a 0.
      c. What is the probability of selecting an area code at random that ends in an odd number if the first digit cannot be a 1 or a 0?

11.   An employer must hire two people from a list of 13 applicants. In how many ways can the employer choose to hire 2 people?

Chapter 4: Discrete Probability Distributions

1.   Decide whether the random variable *x* is discrete or continuous:
      a. *x* represents the number of pumps in use at a gas station
      b. *x* represents the amount of gas pumped at a gas station

2. Decide whether the following is a probability distribution. Explain your answer:
    a.

x	1	2	3	4	5	6	7
P(x)	0.68	0.14	0.08	0.05	0.02	0.02	0.01

    b.

x	1	2	3	4	5	6	7	8
P(x)	$\frac{1}{80}$	$\frac{2}{75}$	$\frac{1}{10}$	$\frac{12}{25}$	$\frac{27}{20}$	$\frac{1}{5}$	$\frac{2}{25}$	$\frac{1}{120}$

3. Sixty-three percent of adults in the US rent DVDs at least once a month. Consider a random sample of five Americans who are asked if they rent at least one videotape per month.
    a. Construct a binomial distribution
    b. Graph the distribution using a histogram and describe the shape
    c. Find the probability of at least 3 Americans renting DVD's at least once a month.
    d. Find mean and standard deviation.
    e. Which *x* values would be unusual?
    f. Interpret your results

Decide which probability distribution applies to the questions below and then solve:

4. In a typical day, 26% of people in the US with Internet access go online to get news. In a random sample of five people in the US with Internet access, what is the probability that the number going online to get news is at least 4 people.

5. During a promotional contest, a soft drink company places winning caps on one of every six bottles. If you purchase one bottle a day, find the probability that you find your first winning cap on the fourth day.

6. In a recent year, Barry Bonds hit 73 home runs in the 153 games he played. Assume that his home run production stayed at this level the following season. What is the probability that he would hit his first home run on the first or second game of the season?

7. Forty-three percent of adults in the US receive fewer than five phone calls a day. In a random sample of seven adults, what is the probability that the number receiving fewer than five calls a day is at most 3?

Chapter 1: Simulations and Sampling Methods

1. Suppose McDonalds is giving away silly bands as their new happy meal toys. They are putting "M's" in 50% of the happy meals, "cheeseburgers" in 40% of the happy meals, and special "Ronald McDonald-shaped-faces" in 10% of the other happy meals. You want to collect all three silly bands. How many happy meals do you need to buy in order to collect them all?
    a. Conduct a simulation with 5 trials. Be sure to clearly label what you are using to represent the numbers in the random number table

b. Estimate the number of happy meals based on the results of your 5 trials

Collecting Data:
Simple Random Sampling   Cluster Sampling   Convenience Sampling
Stratified Sampling   Systematic Sampling

2.   For quality assurance, every 5th engine part is selected from an assembly line and tested for durability.

3.   A study on attitudes about smoking is conducted at a college. The students are divided by class (freshman, sophomore, junior, and senior). Then a random sample is selected from each class and interviewed.

4.   A student at the beach asks a sample of random people how they feel about water pollution as they walk by.

5.   You assign each student a number and generate random numbers. You then question each student whose number is randomly selected.

6.   You select Loyola University and question each student at the school.

7.   Name three other methods used to collect data besides simulations.

5.4-5.5: Sampling Distributions and Normal Approximations

1.   A study found that the mean migration distance of the green turtle was 2200 kilometers and the standard deviation was 625 kilometers. A sample of 12 green turtles is randomly selected. Assuming that the distances are normally distributed, find the probability that the sample mean of the distance migrated is less than 1900 kilometers.

2.   The mean price of houses in a city is $1.5 million with a standard deviation of $500,000. The house prices are normally distributed. Are you more likely to select one house with a price less than $1.125 million or are you more likely to select a sample of 15 houses with a mean price less than $1.125 million?

3.   A manufacturer claims that the life span of its tires is 50,000 miles. You work for a consumer protection agency and you are testing this manufacturer's tires. Assume the life spans of the tires are normally distributed. You select 100 tires at random and test them. The mean life span is 49,721 miles. Assume standard deviation is equal to 800 miles.
     a. Assuming the manufacturer's claim is correct, what is the probability the mean of the sample is 49,721 miles or less? What do you think of the manufacturer's claim?
     b. Would it be unusual to have an individual tire with a life span of 49,721 miles? Why or why not?

4.   Sixty-five percent of children ages 12 to 17 keep at least part of their savings in a savings account. You randomly select 45 children & ask each if he or she keeps at least part of his or her savings in a savings account. Find the probability that at most 20 children will say yes.

5.   In a survey of adults, 68% thought that DNA tests for identifying an individual an individual were very reliable. You randomly select 24 adults and ask each if he or she thinks DNA tests

for identifying an individual are very reliable. Find the probability that at least 15 people say DNA tests for identifying an individual are very reliable.

6. Fifty-two percent of adults say chocolate chip is their favorite cookie. You randomly select 40 adults and ask each if chocolate chip is his or her favorite cookie. Find the probability that exactly 15 people say chocolate chip is their favorite cookie.

Chapter 6: One-Sample Confidence Intervals

1. In a study of perception, 80 men are tested and 7 are found to have red/green color blindness. Construct a 90% confidence interval estimate of the proportion of all men with this type of color blindness

2.
  a. Among those who take the SAT's, 17 students are randomly selected. This sample has a mean score of 558 and a standard deviation of 139 on the math portion of the SAT. Find a 90% confidence interval estimate of the population mean.
  b. Repeat part (a) assuming the statistics came from a sample of 67 students.
  c. Compare your answers. Why do you think that is?

3. The national average for the number of students per teacher for all US public schools is 15.9. A random sample of 10 school districts from a moderately populated area showed that the mean number of students per teacher was 19.2 with a standard deviation of 4.41. Estimate the true mean number of students per teacher with 99% confidence. How does your estimate compare with the national average?

4. A May 2000 Poll found that 38% of a random sample of 1012 adults said that they believe in ghosts.
  a. Find the margin of error for this poll if we want 90% confidence in our estimate of the percent of American adults who believe in ghosts.
  b. In general, if all other aspects of the situation remain the same, will smaller margins of error involve greater or less confidence in the interval? Why?

5. It is believed that the national average gasoline tax is 44.7 cents. A random sample of 12 states has a mean of 43.5 cents with a standard deviation of 9.24 cents. Does this sample contradict our believed national average? Why or why not? Construct a 95% confidence interval to prove your answer.

6. A recent survey of 8 social networking sites has a mean of 13.1 million visitors for a specific month. The standard deviation was 4.1 million.
  a. Find the 90% confidence interval of the true mean.
  b. Find the 99% confidence interval of the true mean.
  c. Which interval is larger? Why?

7. In a survey of 1,998 randomly selected adults, 24% included loud commercials among the annoying aspects of television. Construct the 99% confidence interval for the percentage of all adults who are annoyed by loud commercials.

Part II: Two-sample confidence intervals

2.	In 2006, 20% of the students at a university with 4000 undergraduate students had an overall grade point average of 3.5 or higher (on a 4.0 scale). In 2010, a random sample of 1100 student records found that 25% had a GPA of 3.5 or higher. Is this evidence of grade inflation? Why or why not? Create a 95% confidence interval to show your answer.

3.	Is it a good idea to listen to music when studying for a big test? In a study conducted by some Statistics students, 62 people were randomly assigned to listen to rap music, music by Mozart, or no music while attempting to memorize objects pictured on a page. They were then asked to list all the objects they could remember. Here are summary statistics for each group:

	Rap	Mozart	No Music
Count	29	20	13
Mean	10.72	10.00	12.77
SD	3.99	3.19	4.73

Create a 90% confidence interval for the mean difference in memory score between students who study Mozart and those who listen to no music at all. Interpret your interval.

4.	A survey of 430 randomly chosen adults found that 21% of the 222 men and 18% of the 208 women had purchased books online. Is there evidence that men are more likely than women to make online purchases of books? Why or why not? Use a 99% confidence interval to prove your answer.

5.	In August 2004, *Time* magazine, reporting on a survey of men's attitudes, noted that "Young men are more comfortable than older men talking about their problems." The survey reported that 80 of 129 surveyed 18 to 24 year old men and 98 of 184 surveyed 25 to 34 year old men said they were comfortable. What do you think? Is Time's interpretation justified by these numbers? Use a 99% confidence interval to prove your answer.

Chapter 7: One-Sample Hypothesis Testing

1.	A company that manufacturers school supplies says that teachers spend a mean of more than $580 of their own money on school supplies in a year. A random sample of the amounts (in dollars) that 24 teachers spend a mean of more than $580 of their own money on school supplies in a year. A random sample of the amounts in dollars that 14 teachers spent on school supplies in a recent year is listed below. At a significance level of 0.10, is there enough evidence to support the company's claim?

523	581	476	600	721	622	575	320	888	875	910	643	596	587

2.	A company that makes cola drinks states that the mean caffeine content per one 12-ounce bottle of cola is 40 milligrams. You work as a quality control manager and are asked to test this claim. During your tests, you find that a random sample of thirty 12-ounce bottles of cola has a mean caffeine content of 39.2 milligrams with a standard deviation of 7.5 milligrams. At a significance level of 0.01, can you reject the company's claim?

3.	A maker of microwave ovens advertises that no more than 10% of its microwaves need repair during the first five years of use. In a random sample of 57 microwaves five years old, 13% needed repairs. At a significance level of 0.04, can you reject the maker's claim that no more than 10% of its microwaves need repair during the first five years of use?

4.   A car company says that the mean gas mileage for its luxury sedan is at least 21 miles per gallon. You believe the claim is incorrect and find that a random sample of twenty-nine has a mean gas mileage of 19 mpg and a standard deviation of 4 mpg. Assume the gas mileage of all the company's luxury sedans is normally distributed. At a significance level of 0.05, test the company's claim. What can you conclude?

5.   A researcher claims that 30% of adults in the US are allergic to trees, weeds, flowers and grasses. In a random sample of 86 adults, 17 say they have such an allergy. At a significance level of 0.05, is there enough evidence to support the researcher's claim?

Chapter 8: Two-Sample Hypothesis Testing

1.   The mean score on a science assessment for 49 randomly selected male high school students was 300.4 and the standard deviation was 1.6. The mean score on the same test for 50 randomly selected female high school students was 290.6 and the standard deviation was 1.5. At a significance level of 0.05, can you support the claim that the mean score on the science assessment for the male high school students was higher than for the female high school students?

2.   In October 2000 the US Department of Commerce reported the results of a large-scale survey on high school graduation. Researchers contacted more than 25,000 Americans aged 24 years to see if they had finished high school; 84.9% of the 12,460 males and 88.1% of the 12,678 females indicated that they had high school diplomas. Does this provide strong evidence that girls are more likely than boys to complete high school?

3.   The National Sleep Foundation asked a random sample of US adults' questions about their sleep habits. One of the questions asked about snoring. Split into two age categories, 48 of the 184 people under 30 said they snored, compared with 317 of the 811 in the older group. At a significance level of 0.10, can you support the claim that the older you are the more you snore?

4.   The table below shoes the scores of 12 randomly selected students the first and second times they took the mathematics SAT. At a significance level of 0.05, is there enough evidence to conclude that the students' SAT scores improved on the second test?

Student	1	2	3	4	5	6	7	8	9	10	11	12
First SAT score	457	419	343	539	394	413	392	421	439	340	493	339
Second SAT score	532	523	427	607	444	490	428	524	532	397	550	357

Chapter 10: Chi-Squared Distributions

1.   State the Null and Alternative Hypotheses for a Chi-Squared Goodness of Fit Test.

2.   State the Null and Alternative Hypotheses for a Chi-Squared Independence Test.

3. Suppose a company printed baseball cards. It claimed that 30% of its cards were rookies; 60%, veterans; and 10%, All-Stars. To determine whether this is true, you randomly selected 200 baseball cards and found that 13 were All-Stars, 111 were veterans and 76 were rookies. Perform a Chi-Square Goodness of Fit test. What can you conclude? Why do you think that is?

4. A bicycle safety organization conducted a study of 996 randomly selected fatal bicycle accidents. The month each accident occurred is shown in the table below. Can you conclude that fatal bicycle accidents are not uniformly distributed by month?

Month	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
Frequency	50	48	81	72	90	101	129	122	89	87	73	54

5. You work for an insurance company and are studying the relationship between types of crashes and the vehicles involved. As part of your study, you randomly select 3207 vehicle crashes and organize the resulting data as shown in the contingency table below. Can you conclude that the type of crash depends on the type of vehicle?

		Type of Vehicle	
Type of Crash	**Car**	**Pickup Truck**	**Large Passenger Van**
Single-vehicle	895	493	45
Multiple-vehicle	1400	336	38

Chapter 9: Correlation and Regression

1. Construct *by hand* a scatter plot of the following data and **interpret** its meaning in context of the problem (choose a scale and label your axis!!).
   The weight (in pounds) and the daily caloric intake of 9 adults:

Weight, x	102	118	125	149	155	180	192	195	211
Calories, y	1500	1570	1800	1720	2000	1880	1900	2230	2520

2. Calculate the correlation coefficient for the following data: _____
   The ages (in years) and the number of hours of sleep in one night of seven adults

Age, x	35	20	59	42	68	38	75
Hours of Sleep, y	7	9	5	6	5	8	4

   a. What type of correlation does set of data have? Be specific.
   b. Is the correlation coefficient significant? Test at the .01 level. Do we have enough evidence to conclude this type of correlation for the entire population? Why or why not?

3. Calculate the regression line for the following set of data. How many hours of sleep per night would you predict a person who is 50 years to get? **Label/Interpret your answer in a complete sentence.**
   The ages (in years) and the number of hours of sleep in one night of seven adults

Age, x	35	20	59	42	68	38	75
Hours of Sleep, y	7	9	5	6	5	8	4

4. Calculate the coefficient of determination for the following set of data. What does that mean? Include explained *and* unexplained proportions in your answer. **Answer in complete sentences.**

The annual per capita sugar consumption (in kilograms) and the average number of cavities of 11 and 12 year old children in seven different countries

Sugar Consumption, x	2.1	5.0	6.3	6.5	7.7	8.7	11.6
Cavities, y	0.59	1.51	1.55	1.70	2.18	2.10	2.43

5. The table below shows median weekly earnings (in US dollars) of full-time male and female workers for five years. Construct a 90% prediction interval for the median number of leisure hours per week when the median number of work hours per week is 500. Use standard deviation equal to 8.226. **Interpret** your results. (formula on the front for margin of error)

Median Weekly Number of Work hours, x	695	679	672	618	557
Median Weekly Number of Leisure hours, y	552	529	511	473	418

6. Use the multiple regression equation below to predict the y-values for the given values. *Show all work for full credit.* **Label/Interpret** your answer in a complete sentence.

An equation that can be used to predict fuel economy (in miles per gallon) for automobiles is

$\hat{y} = 41.3 - 0.004x_1 - 0.0049x_2$ where x_1 is the engine displacement (in cubic inches) and x_2 is the vehicle weight (in pounds).

    a. $x_1 = 305; x_2 = 3750$

    b. $x_1 = 105; x_2 = 2200$