Algebra 2 CP
Statistics Assignments

Day 1

1. Determine whether each question is a statistical question. Explain each of your answers.

   a. How many words are in this sentence?
   b. How many words are there in the sentences of our Math text book?
   c. What time do you eat breakfast on weekdays?
   d. What time did you eat breakfast today?
   e. Who was the youngest president?
   f. What were the ages of the presidents at inauguration?
   g. Do more boys in our class like the color purple than girls?
   h. Do girls in our class own more shoes than boys?
   i. What proportion of M&M's in a bag are blue?

2. For each of the statistical questions in #1, identify the population. If it would be difficult to collect data from all members of the population, suggest a way to obtain a random sample.

3. Identify each study as a survey, observational study, or randomized experiment. Explain your reasoning.

   a. A group of students is interested in knowing if the number of times they can make a basketball free throw is related to the color of the basketball. Each student flips a coin to determine whether he will shoot the free throw with a regulation colored basketball or with a blue basketball. This procedure is repeated for 10 free throws for each student. The number of baskets scored with each type of ball is recorded and a statistical analysis is performed.
   b. A teacher is interested in knowing if there is a correlation between attending an SAT Prep class and scores achieved on the SAT Examination. The teacher examines SAT scores from the records of students who took an SAT Prep class and those that did not take an SAT Prep class. A statistical analysis of the data is performed.
   c. Researchers want to know if bathing in Epsom salts can reduce the pain of arthritis. 100 arthritis sufferers are asked to record their pain levels daily for one month. Fifty of the patients are randomly selected to bathe in Epsom salts every day. The other fifty do not bathe in the salts. A statistical analysis of the self-reported pain levels is performed.

4. In #3b, the teacher finds that the students who took the SAT prep course had higher SAT scores, on average, than the students not enrolled in the course. Can you conclude that the SAT course caused the increase in the scores? Explain your reasoning.

5. A researcher wants to investigate whether using tanning beds at least twice a month affects the likelihood of developing skin cancer. Should she conduct a survey, observational study, or experiment? Explain your reasoning.

6. A congressman is considering a bill that would provide government-sponsored insurance. He receives 1152 letters on the issue from his constituents, of which 823 oppose the legislation. He says, "I'm surprised that most people oppose the bill. I would have thought it would be more popular." Do you agree with the congressman's conclusion that the bill is not popular? Explain your reasoning.

Day 2

1. The table below shows the number of cans collected by several homerooms during a canned food drive.

Room	Canned Goods	Room	Canned Goods
A	150	E	10
B	300	F	25
C	55	G	200
D	40	H	100

    a) Find the mean, median and mode of the data.
    b) Explain why the mean is larger than the median for this data set.

2. The table below shows the average yearly rainfall in Los Angeles for the past ten years. The rainfall is measured in inches from July through June.

Year	14-15	13-14	12-13	11-12	10-11	09-10	08-09	07-08	06-07	05-06
Inches	8.52	6.08	5.85	8.69	20.2	16.36	9.08	13.53	3.21	13.19

    a) Find the mean, median and mode of the data.
    b) Which measure of center shows the typical yearly rainfall for the last 10 years?

3. Kaitlyn's scores on her first five algebra tests were 88, 90, 91, 89 and 92. What score must she get on her 6th test so that her mean score will be at least 90?

4. Mrs. Grasel went to the grocery store to buy 5 items at an average cost of $4 per item. She decided to add a chocolate cake which cost $10 to her grocery cart. What is the average cost of the six items she bought?

5. The prices in dollars of paintings sold at an art auction are shown.

1800	750	600	600	1800	1350	300	1200	750	600	750	2700
600	750	300	750	600	450	2700	1200	600	450	450	300

    a. Make a frequency table for the data using 6 intervals from $0 to $3000.
    b. Make a histogram for the data.
    c. What price was paid most often for the artwork?
    d. What was the median price for the artwork?
    e. How many works of art sold for more than $2000?

6. Carlos surveyed 19 friends to find the number of cans of soda they drink in an average week. His results are shown below:

0	0	0	1	1	1	2	2	3	4
4	5	5	7	10	10	10	11	11	

    a. Find the five-number summary for the data.
    b. Make a box plot for the data.

Day 3

1. In order to decide if the local high school needs a bigger parking lot, the district's planning manager gathered the data shown in the chart below.

Cars in the parking lot at 11 a.m.

	M	T	W	T	F
Week 1	150	167	181	156	168
Week 2	135	146	142	166	163
Week 3	155	139	148	156	127
Week 4	158	157	163	138	141

   a. Make a frequency distribution table of the number of cars in the parking lot using 7 intervals from 120 to 190. Draw a histogram for the data in the frequency distribution.
   b. Describe the center and spread of the distribution.
   c. If the current lot has 200 spaces, do you think the high school needs a bigger lot? Why or why not?

2. The number of hits made for the season by the regular players on a little league team are: 31, 15, 22, 18, 12, 34, 26, 28, 19, 26, 17, 29, 26, 22, 31

   a. Find the mean and median of the data
   b. Find the range and IQR of the data
   c. Draw a box plot of the data
   d. Are there any outliers using the 1.5 IQR rule?
   e. A player from another team, who made 28 hits this season, claims that he made more hits than 75% of the team's players. Do you agree? Explain your reasoning.

3. The average life span of some animals commonly found in a zoo are: 1, 7, 7, 10, 12, 12, 15, 15, 18, 20, 20, 20, 25, 40, 100.

   a. Make a histogram of the data using intervals of 10 years.
   b. Describe the center and spread of the distribution.
   c. Do any of the life expectancies appear to be outliers? Explain your reasoning.
   d. Use the 1.5 IQR rule to confirm that 40 an 100 years are outliers.
   e. Make a box plot of the distribution. Clearly identify the outliers.

4. Without using a calculator, find the mean and sample standard deviation of 0, 2, 4, 6, 8. Explain what each number tells you.

Day 4

1. You are one of several judges for a new reality TV show called "*Extreme Bot*" where teams of ten young inventors exhibit robots. The robots are designed to make certain parts which are to be specifically 5 micrometers thick. You are judging quality control and measure a sample of each team's parts. Here is what you discover:

   Team 1's parts measure 4.6, 5.2, 5.1, 4.2, 4.9, 5.3, 5.0, 4.9, 4.8, 4.8 micrometers
   Team 2's parts measure 3.9, 4.8, 5.0, 4.2, 5.1, 5.4, 3.9, 5.2, 4.4, 4.9 micrometers
   Team 3's parts measure 4.6, 4.7, 4.6, 4.6, 4.5, 4.6, 4.5, 4.7, 4.8, 4.9 micrometers

   a. Determine the mean and the median for each team. Based on these findings, to which team would you award the highest score in quality control and why?
   b. Compute the range for each team. What is the range telling you? To which team would you award the highest score if range were used to choose the winner, and why?
   c. Compute the standard deviation for each team. Based upon the standard deviation, to which team would you award the highest score in quality control, and why?
   d. Based upon your total findings, to which team would you award the highest score if the parts can be no more than 5 microns thick?

2. Your little sister has ten songs on her brand new iPod. The device is set to play songs at random with repetitions allowed.

   a. Explain how you could use your calculator to simulate how many songs your sister will need to play until she hears all ten songs.
   b. Conduct your simulation 3 times and record the number of songs played until all ten have been heard.

The box plot at right shows the results of 100 simulations:

Box Plot
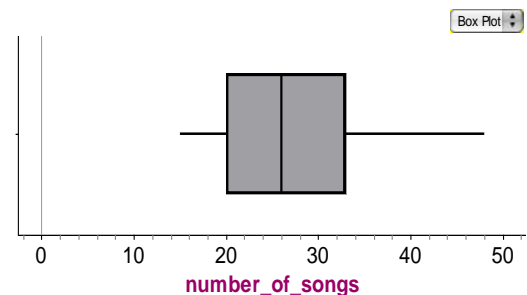
   c. What is the median of the distribution? What does it tell you?
   d. What is the IQR? What does it tell you?
   e. What is Q1? What does it tell you?
   f. What is Q3? What does it tell you?
   g. The maximum value is 48 songs. Use the 1.5 IQR rule to verify that 48 is not an outlier.

   h. Your little sister says she listened to a hundred songs before she heard her favorite. Do you think she is exaggerating? Explain.
   i. What is the (theoretical) probability that your sister will only have to listen to each song one time before hearing all ten?

number_of_songs

Day 5

1. A random sample of the number of hours per week worked by 50 students with part-time jobs is given below.

16	9	10	24	15	10	8	18	22	20
12	20	6	13	9	4	24	14	17	3
5	18	11	8	20	12	10	5	18	6
15	12	28	12	16	26	12	14	4	24
30	16	5	18	21	4	14	8	12	7

a. Find the mean and median of the data
b. Find the standard deviation and IQR of the data
c. Are there any outliers according to the 1.5 IQR rule?
d. Draw a box plot of the data
e. Make a frequency distribution of the data
f. Draw a histogram for the data
g. Is the data positively skewed, negatively skewed or approximately symmetric?
h. Which measure of center would most accurately reflect the number of hours worked by the students?

2. The data below shows the selling prices for 37 homes recently sold in a small neighborhood.

Housing Price	No. of Houses Sold
less than $400,000	0
$400,000--$420,000	1
$420,000--$440,000	3
$440,000--$460,000	7
$460,000--$480,000	8
$480,000--$500,000	6
over $500,000	12

a. Draw a histogram for the distribution table.
b. Is the data positively skewed, negatively skewed or approximately symmetric?
c. If you were a realtor trying to sell a young couple a home in this neighborhood, which measure of center would you be likely to report? Why?

3. The data below shows the weights of a sample of pies made at two bakeries.

Pam's Pie Palace	
Pie	Pie Weight
A	377g
B	392g
C	338g
D	333g
E	418g
F	402g
G	319g
H	324g

Peter's Pie Plaza	
Pie	Weight
A	355g
B	367g
C	340g
D	349g
E	358g
F	344g
G	366g

a. Make a stacked box plot to display the data.
b. Describe the shape, center and spread of each distribution.
c. Based on your findings, where would you prefer to buy your pies? Explain your answer.

Day 6

1. The table at right shows recent composite ACT scores.
   Determine whether the data is positively skewed, negatively
   skewed or approximately normally distributed.

ACT Score	Percent of Students
33-36	1
28-32	9
24-27	19
20-23	29
16-19	27
13-15	12

2. The table shows the age distribution for people living in the United States as determined by the U.S.
   Census Bureau. Determine whether the distribution is
   positively skewed, negatively skewed or approximately
   normally.

Age	Percent
0-19	28.7
20-39	29.3
40-59	25.5
60-79	13.3
80-99	3.2
100+	0.0

*For problems #3 – 5, draw the normal curve and label the mean and ± 1, 2, 3 standard deviations from
the mean. Then answer the questions.*

3. A vending machine usually dispenses about 8 ounces of coffee. The actual amount varies normally with
   a standard deviation of 0.3 ounces.

   a) What percent of the time will you get more than 8 ounces of coffee?
   b) What percent of the time will you get more than 8 ounces of coffee?
   c) What percent of the time will you get between 7.4 and 8.6 ounces?

4. The size of CDs made by a company varies normally with a mean of 120 millimeters and a standard deviation of
   1 mm.

   a) What percent of CDs would you expect to be greater than 120 mm?
   b) If the company manufactures 1000 CDs per hour, about how many of the CDs made in one hour will be
      between 119 and 122 mm?
   c) If the CDs are made for drives which are 122 millimeters wide, about how many CDs per hour will be too lar4ge
      to fit in the drive?

5. The shelf life of a particular snack chip is normally distributed with a mean of 180 days and a standard deviation of
   30 days.

   a) About what percent of chips will last between 150 and 210 days?
   b) About what percent of chips will last between 180 and 210 days?
   c) About what percent of chips will last less than 90 days?
   d) About what percent of chips will last more than 210 days?

Day 7

1. The following frequencies are obtained when two dice are rolled and their sum is found:

Sum	2	3	4	5	6	7	8	9	10	11	12
Freq.	1	2	3	4	5	6	5	4	3	2	1

    a. Draw a histogram of the distribution.
    b. The mean of the distribution is 7 and the standard deviation is about 2.5. Is the data normally distributed? Explain your answer.

2. The scores on a math final at a large university were normally distributed with a mean of 85 and a standard deviation of 4.
    a. Draw the normal curve and label the mean and $\pm$ 1, 2, 3 standard deviations
    b. What percent of students scored above 93?
    c. What percent of students scored below 81?
    d. Use your calculator to find the percentage of students who scored between 70 and 90.
    e. If 1000 students took the final, about how many scored 95 or above?

3. The useful life of a car battery is normally distributed with a mean of 100,000 miles and a standard deviation of 10,000 miles.
    a. Draw the normal curve and label the mean and $\pm$ 1, 2, 3 standard deviations
    b. About what percent of batteries will last between 90,000 and 110,000 miles?
    c. Use your calculator to find the percentage of batteries that will last less than 85,000 miles.
    d. What is the probability that your next car battery will last more than 105,000 miles?

4. The maker of M&M candies claims that the proportion of yellow candies in a bag of M&Ms are approximately normally distributed with a mean of 0.15 with a standard deviation of 0.03.
    a. Would you be surprised to open a bag of M&M's and find no yellow candies? Explain.
    b. Would you be surprised to open a bag of M&M's and find 1/3 of the bag is yellow? Explain.
    c. About what percent of the time would you expect to open a bag of candies and find 0.10 to 0.20 yellow candies?

Day 8

1. A random sample of 45 students at WHS were asked if they would prefer a later start time for the school day. Twenty of the students said "yes."
   a. What is the sample proportion of students answering "yes"? Report your answer as a decimal.
   b. If a different random sample of 45 students were asked the same question, do you think you'd get the same sample proportion? Explain why or why not.
   c. The margin of error for a 95% confidence interval for this survey is about 0.15. What is the 95% confidence interval for the true proportion of WHS students who wish to have a later school start time?
   d. Explain what is meant by a 95% confidence interval.
   e. The school paper reports that more than 50% of students want a later start time. Comment on the accuracy of this statement.

2. A statistics student wants to know the true proportion of yellow candies in a bag of M&M's. She opens a bag of M&M's and finds that 14 of the 75 candies are yellow.
   a. What is the sample proportion of M&M's that are yellow? Report your answer as a decimal.
   b. Is the sample proportion likely to be the same as the true population proportion of yellow M&M's? Why or why not?
   c. The student determines that the margin of error for a 95% confidence interval is about 0.09. Determine a 95% confidence interval for the true proportion of yellow M&M's.
   d. Explain what is meant by a 95% confidence interval.
   e. The M&M website reports that while there is a lot of variation in the color of M&M's in various bags, on average the proportion of yellows is 0.13. Comment on the accuracy of this statement.

3. The girls' basketball coach is curious about the average (mean) height of WHS girls. She measures a random sample of 30 girls and measures a mean height of 55".
   a. Is the mean height of a different random sample of 30 girls at WHS likely to be 55"? Explain.
   b. The standard deviation of girls' heights is about 2". Use this value to find the Standard Error of repeated measurements of the mean height of 30 girls.
   c. Use the Standard Error from (b) to find the Margin of Error for a 95% confidence interval.
   d. Determine a 95% confidence interval for the mean height of WHS girls.
   e. Explain what it means to be 95% confident.

4. A teacher wonders if an on-line video is effective in improving student knowledge about histograms. In one class period, she gives a random sample of 12 students a pre-test, then shows them the video, and then gives the students a similar post-test. The difference in post-test scores minus pre-test scores is shown:

13	6	8	7	8	10	0	3	15	4	11	9

   a. Draw a histogram of the differences in test scores and describe its shape. Use an interval width of 2.
   b. Calculate the mean increase in test score for this sample of 12 students. Do you think a different random sample would have given the same mean?
   c. The teacher estimates that the standard deviation of test scores is about 4. Use this value to find the Standard Error and the Margin of Error for a 95% confidence interval
   d. Determine a 95% confidence interval for the mean increase in test scores.
   e. Can the teacher conclude that watching the video will improve the average test scores of her students by at least 5 points?

Day 9

In 1978 researchers published a study in Science magazine reporting an experiment where a chimpanzee named Sarah was shown videos of eight different scenarios of a human being faced with a problem. After being shown each video, Sarah was presented with two photographs, one of which depicted a possible solution to the problem. For example, a problem might be a human trying to reach an object that is just out of reach. The picture of the correct solution would show the human stepping on a box to be able to reach further. In the experiment, Sarah picked the correct photograph seven out of eight times.

a. Give two possible explanations for why Sarah got seven of the eight answers right.
b. If Sarah were simply guessing, for how many of the eight problems would you expect her to correctly identify the correct solution?
c. Suppose that you are designing a simulation to study the distribution of possible outcomes of the experiment. Assuming that Sarah was choosing the pictures randomly, how could you use a coin to model Sarah's responses?
d. A simulation of 100 tosses of eight coins resulted in the distribution of the number of correct answers shown below. Describe the shape, center, and spread of the distribution.
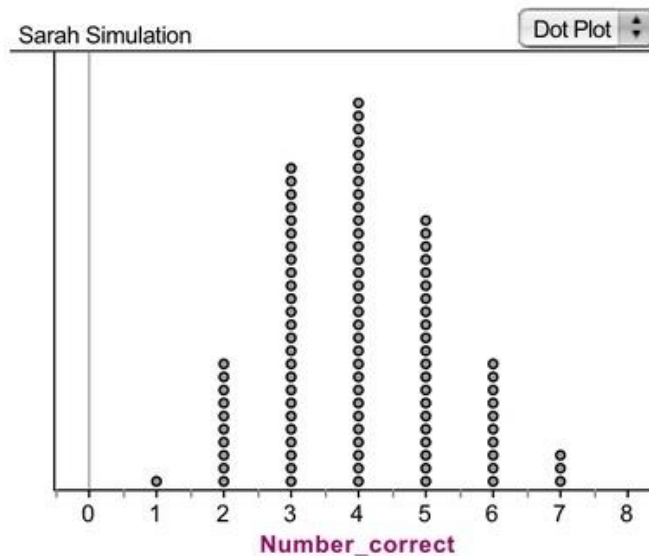
e. Use the dot plot to determine the probability of obtaining 7 heads in eight coin tosses.
f. Now perform a binomial calculation to determine the theoretical probability of obtaining exactly 7 heads in eight coin tosses. (Hint: we did this type of calculation in Lesson 12-9) Do your theoretical results agree with the simulation?
g. Perform a binomial calculation to determine the theoretical probability of obtaining exactly 4 heads in eight coin tosses. Do your results agree with the simulation?
h. Does the distribution provide evidence that Sarah was recognizing correct solutions and not just guessing? Explain your answer.
i. Sarah had been raised in captivity and had participated in many experiments using videos and photographs. Based on the results of this study, could you conclude that all chimpanzees can solve problems of this nature? Why or why not?

Day 10

The winners of the Academy Award for Best Actor and Best Actress in the last ten years are shown below:

Year	Actor	Age	Movie
2005	Jamie Foxx	37	*Ray*
2006	Philip Seymour Hoffman	38	*Capote*
2007	Forest Whitaker	45	*The Last King of Scotland*
2008	Daniel Day Lewis	50	*There Will Be Blood*
2009	Sean Penn	48	*Milk*
2010	Jeff Bridges	60	*Crazy Heart*
2011	Colin Firth	50	*The King's Speech*
2012	Jean Dujardin	39	*The Artist*
2013	Daniel Day-Lewis	55	*Lincoln*
2014	Matthew McConaughey	44	*Dallas Buyers Club*

Year	Actress	Age	Movie
2005	Hilary Swank	30	*Million Dollar Baby*
2006	Reese Whitherspoon	29	*Walk the Line*
2007	Helen Mirren	61	*The Queen*
2008	Marion Cotillard	32	*La Vie en Rose*
2009	Kate Winslet	33	*The Reader*
2010	Sandra Bullock	45	*The Blind Side*
2011	Natalie Portman	29	*Black Swan*
2012	Meryl Streep	62	*The Iron Lady*
2013	Jennifer Lawrence	22	*Silver Linings Playbook*
2014	Cate Blanchett	44	*Blue Jasmine*

a. Find the five-number summary for the ages of the Best Actors and construct a box-plot of the age distribution.

b. Find the five-number summary for the ages of the Best Actresses and construct a box-plot of the age distribution.

c. Use your results from a) and b) to compare the shape, center, and spread of each age distribution.

d. Find the mean and standard deviation of the men's and women's ages. Calculate the difference between the men's mean age and the women's mean age.

e. Does your data suggest that for the last ten years, the winners of the Best Actress award are younger than the winners of the Best Actor award? Explain your answer.

f. The graph below shows the probability model for the difference in the mean of the actors' and actresses' ages based on the hypothesis that there really is no age difference between the genders. For each simulation, the 20 different ages of the award winners are assigned to the two genders at random and the difference between the mean ages is determined. Use the graph to comment on whether the observed difference of 7.8 indicates that there is a significant difference between the actors' and actresses' mean age. There are a total of 200 differences shown in the simulation.

Measures from Scrambled Academy Awards — Dot Plot — diffOfMeans
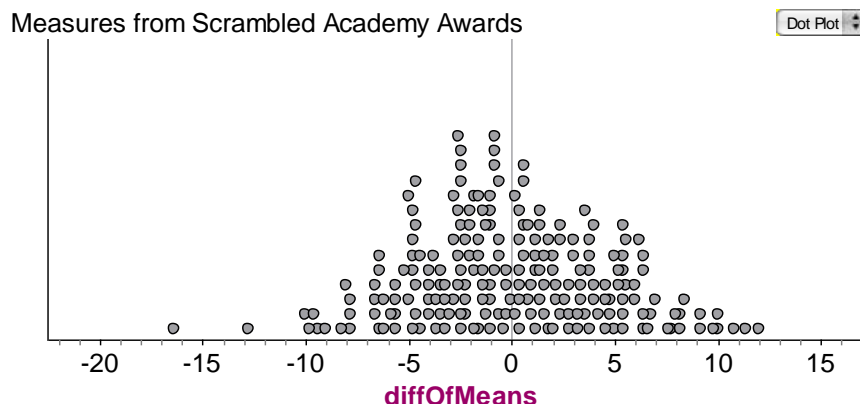
Day 11 Review

1. Identify each study as a survey, observational study, or experiment.

    a. A biologist studying a new fertilizer applies it to his 15 favorite plants in his garden.
    b. A radio sports broadcaster asks his listeners to call in and tell him their favorite basketball team.
    c. A parent who wants to know how much time his child spends on the phone examines the phone records during Spring Break.

2. For each study above, determine whether randomness was utilized appropriately. Explain a possible consequence if randomness was not utilized. Then suggest a <u>better method</u> for collecting the data.

3. A researcher examines data from a random sample of 27 countries and determines that in countries where more chocolate is consumed, people generally live longer. Can he conclude that eating chocolate causes people to live longer? Explain your reasoning.

4. In 2001 baseball player Barry Bonds broke Mark McGwire's record of 70 home runs in a single season. Here are Bonds' home run counts for the years 1986 (his rookie year) to 2004:

1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
16	25	24	19	33	25	34	46	37	33

1996	1997	1998	1999	2000	2001	2002	2003	2004	
42	40	37	34	49	73	46	45	45	

    a. Make a frequency table and histogram of the number of home runs using intervals of 5 from 15 to 75.
    b. Describe the shape, center and spread of the distribution. Does Bonds' record of 73 home runs appear to be an outlier?
    c. Find the five-number summary of the data and use it to make a box plot.
    d. Does the 1.5* IQR rule identify 73 as an outlier? Show your work!

5. Your math teacher has designed a strange type of multiple choice test. For the test, you will answer the first question. If you get it right, then you are done with the test. If you get the question wrong, then you must answer a second question. You will continue answering (different) questions until you get a question right. Assume that there are 4 answer choices for each question.
    a. If you have not studied at all for the test, what is the probability you get the first question correct?
    b. Explain how you could use your calculator's random integer function to generate test answers.
    c. Use your calculator to see how many questions you will need to answer until you get a question correct. You may assume that the correct answer to each question is "A."
    d. Repeat the simulation in (c) five times and report your results.
    e. When your friend took the test, it took him 10 questions until he got one correct. Does this seem unusual? Explain.

6. Assume that the number of home runs scored each year by a baseball player is normally distributed with a mean of 37 and standard deviation of 13.
    a. Draw a normal probability curve and label $\pm$ 1, 2 , 3 standard deviations from the mean.
    b. What is the probability the baseball player will hit between 24 and 63 home runs during a season?
    c. Use your calculator to find the probability the player will hit more than 70 home runs.

7. The mean SAT Math score for a random sample of 25 WHS students is 560.
   a. Would another random sample of 25 students be likely to have the same mean score?
   b. The College Board reports that SAT scores are approximately Normally distributed with a standard deviation of 90. Find the standard error of the mean for a sample of 25 students.
   c. Determine a 95% confidence for the true mean SAT Math score for WHS students.
   d. Explain what is meant by a 95% confidence interval.
   e. A math teacher brags that on average students at WHS score at least 600 on the SAT. Comment on the accuracy of this statement.

8. A student claims to "know nothing" about probability. He takes a 10 question True/False quiz on probability and gets 9 questions correct.
   a. Give two possible explanations for how the student got 9 out of 10 answers correct.
   b. If the student were simply guessing, how many of the ten problems would you expect him to get correct?
   c. Suppose that you are designing a simulation to study the distribution of possible outcomes of guessing correct answers on a 10-question quiz. Assuming that the student was guessing the answers randomly, how could you use a coin to model the student's responses?
   d. A simulation of 100 tosses of ten coins resulted in the distribution of the number of correct answers shown below. The counts for each interval are shown above the bin. Describe the shape, center, and spread of the distribution.
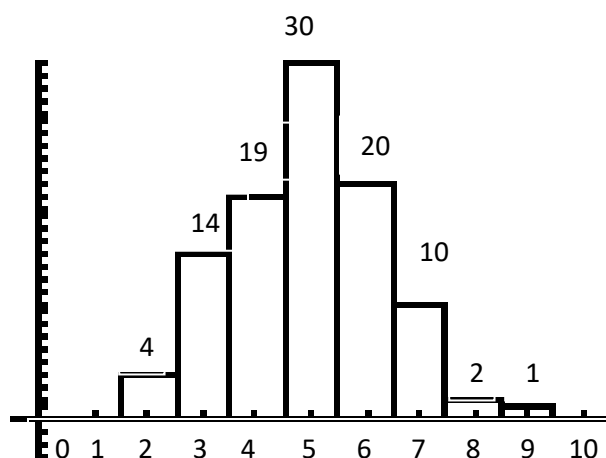
   e. Use the histogram to determine the approximate probability of obtaining 9 heads in ten coin tosses.
   f. Do you think the student really "knew nothing" about the topic of probability? Use the results from (e) to explain your answer.