Activities and Projects

for the

Freeman Statistics Series

Ron Millard
Shawnee Mission South High School
Graceland University

W. H. Freeman and Company
New York

(Copyright Page)

Table of Contents

Preface

Students, statistics, and learning are the elements that I have combined in the classroom by using projects. All students do not learn in the same manner, and projects that focus on specific learning objectives provide a modality for some to discover statistical concepts that may not have been grasped within the formal classroom setting. Other students, while they grasp the concepts, will expand their understanding of statistical concepts through the use of projects. Learning and motivation for learning have been significantly increased in my classroom as a result of using supporting projects.

Many of the ideas for these projects have been developed through the sharing of ideas with fellow teachers. I thank each of them for giving their ideas to me, as I likewise shared mine with them. Time has erased the original forms of most of these activities, and I find with each new semester that I continue to modify them to fit the needs of each new class of students. Calculators have changed most of the random experiments and provided learning of new methods for random simulations. Computers and the Internet continue to provide resources for data. Students must use these tools in a hands-on environment where they can explore and seek solutions to open-ended questions in order to be prepared for the world they will enter into beyond their education.

The projects in this supplement have been organized to correspond with the topics covered in David Moore's statistics textbooks. I have included a section titled "Notes to Instructors" where the purpose of each project is listed along with any specific items needed to do the project. While many items can be used to generate data, I have found students have an increased motivation for learning by using food items as data collection devices. Buttons may come in all sizes and colors, but, in my classroom, M&M's are more fun, edible, and get the learning task accomplished.

Possible times to do these activities within the class are: a) within each chapter as the topic is discussed, b) at the end of the term for review, c) on shortened class weeks with out time to proceed with an additional unit, and d) as individual projects outside of the class to then be presented to the class in poster sessions or displayed on bulletin boards.

Enjoy the teaching and the students will learn. Enthusiasm is contagious and I make every attempt to spread it. Happy projecting!

January 2003

Notes to Instructors

The following summary provides a statement of purpose for each project and the materials needed.

Project 1 WHAT'S THE COST OF LUNCH?

Purpose:
To gather data and display it using both a graphical and a numeric method.

Materials needed:
None.

Project 2 SPRING BREAK

Purpose:
To look at the relationship of distance traveled and air fares. The student will use the Internet to find the lowest fares for a set of destinations for typical spring break trips and the distance of the flight. Possible Web sites are Map quest, Travelocity, and the airline Web sites, but others may be used at the teacher's discretion. A scatterplot of the data, a least-squares regression line, and correlation will be found. The student will give a persuasive argument for his or her conclusion.

Materials needed:
Access to the Internet.

Project 3 TEST YOUR MEMORY

Purpose:
To work with residuals. Choose a list of people whose names your students recognize. It adds interest for the students to include you in the list. You can find the birth dates of celebrities on the Internet by searching, for example, "Whoopi Goldberg birth date." I am listing a few favorites here:

Drew Bledsoe	02 – 14 – 72
Doug Flutie	10 – 23 – 63
Catherine Zeta–Jones	09 – 25 – 69
Drew Barrymore	02 – 22 – 75
Julia Roberts	10 – 28 – 67
Richard Dreyfus	10 – 29 – 47
Nick Carter (Back Street Boys)	01 – 28 – 80
Whoopi Goldberg	11 – 13 – 49
Ronald Regan	02 – 15 – 11

George W. Bush	07 – 06 – 46
Al Gore	03 – 31 – 48

I suggest that you use at least 10 people with a wide range in ages. Include yourself or some well-known member of the faculty for some class fun.

Materials needed:
People known to the students and their birth dates.

Project 4 POPULATION GROWTH

Purpose:
To use the Internet to find population data, and then use the data to make a prediction of the population for a chosen location in the next census.

Materials needed:
Access to the Internet.

Project 5 WHAT DO STUDENTS DRIVE?

Purpose:
To design a study and gather data.

Materials needed:
Access to a student parking lot.

Project 6 JELLYBEAN

Purpose:
To use capture-recapture random sampling to estimate the number of jellybeans in the jar.

Materials needed:
1 large bag of multicolored jellybeans (with no black)
1 small bag of black jellybeans
1 large clear jar or fish bowl
Utensils for stirring and dipping
Paper cups or napkins

To start, place the multicolored jellybeans in the bowl.

Project 7   FAIR COIN

Purpose:
To observe probability of a repeated event over time. The project may be done individually or as a group where each member of the group uses a different coin.

Materials needed:
Coins for each student.

Project 8   ODD AND EVEN

Purpose:
To explore binomial probability.

Materials needed:
One die to roll
One penny to move as a game piece

An alternative procedure is to use a random number generator on your calculator in place of the die.

Project 9   THE AGE OF A PENNY

Purpose:
To determine the approximate age of pennies in circulation. The distribution of the ages, not the year the coin was minted, will be examined. The central limit theorem will be used to form an interval estimation of the ages for your sample of coins.

Materials needed:
One roll of pennies for each student.

Project 10   AHOY MATES

Purpose:
To develop a sampling procedure, generate data, display the data with an appropriate graph, and perform a hypothesis test of the companies claim. The student is to know procedures for designing an experiment, using a randomizing process to gather data, and conducting a hypothesis test of means.

Materials needed:
One bag of CHIPS AHOY cookies with the label reading “over 1000 chips.”

Project 11 SNAP CRACKLE POP

Purpose:
To compare two distributions using a hypothesis test for the difference of two means, using the P-value approach. This project may be done in groups to access more data in a timely manner.

Materials needed:
None, however the students will need to go out of the classroom to gather their data.

Project 12 TASTE THE DIFFERENCE

Purpose:
To model a taste test to the binomial distribution and perform a hypothesis test for proportions. The student is to gather data and use the binomial probability density function and model the binomial distribution with the standard normal distribution. The TI-83 calculator (or equal), the binomialpdf, and binomialcdf functions will be used.

Materials needed:
One one-pound bag of plain M&M's
TI-83 calculators

Project 13 PLAIN AND PEANUT

Purpose:
To analyze more than one data set and the differences in two or more distributions. The project may be used as separate parts for the comparison, or be used with the Chi Square distribution.

Materials needed:
One one-pound bag of plain M&M's
One half-pound bag of plain M&M's
One package of snack size plain M&M's
One one-pound bag of peanut M&M's
One half-pound bag of peanut M&M's
One package of snack size peanut M&M's
Access to the Internet
TI-83 calculators

Project 14 SHORT OR TALL

Purpose:
To find a line of best fit to a paired set of data values, use inferential methods of the linear model, and make prediction intervals.

Materials needed:
None

Project 15 MANY VARIABLES DO PREDICT

Purpose:
To use a computer to find multiple variable equations.

Material needed:
Access to a computer with Excel or Minitab.

Project 16 GROWTH GROUPS

Purpose:
To compare multiple groups using ANOVA.

Materials needed:
Access to a computer with Excel or Minitab.

Project 17 RANDOM ASSIGNMENT

Purpose:
To compare multiple groups using ANOVA.

Materials needed:
Access to a computer with Excel or Minitab
Class rosters with names

Project 18 SKITTLES MACHINES

Purpose:
Check data with statistical process control.

Materials needed:
Fun size bags of Skittles
TI-83 calculator

Project 19 NBA

Purpose:
Check data with statistical process control.

Materials needed:
Access to NBA data (Web site or newspaper)
TI-83 calculator

Project Final

Purpose:
To use five aspects of a statistical study. The project should be three to five pages in length. The student may use either an observational study or a designed experiment. This project is best used as the culminating activity for the class in order to demonstrate overall conceptual learning.

As instructor, approve the students topics before he or she begins the project to assure the appropriateness of the research topic.

Materials needed:
None

Looking at Data–Distributions

Project 1   WHAT'S THE COST OF LUNCH?

How much does a student spend on lunch during school days? The student may choose to go to the local franchise of "Betsy's Burgers," go through the a la carte line, or eat the Type A regular lunch. Any way you slice it, you still are paying for your lunch.

1. You are to gather data from 20 males and 20 females today at lunch and record the data below. Now remember this is not a scientific study, so you can take their word for what they spent for lunch. Record the data and indicate if the lunch was purchased on or off campus.

Student	Female	Amount		Male	Amount
Count	On / Off	$		On/ Off	$
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

2. Make a stemplot of the 40 values in the space below:

3. Are there any outliers in the data set? If so, list them and give a reason to either eliminate them or keep them in the data set.

4. Find the five-number summary of the costs from the above data.

   Low = ____ Q1 = ____ Median = ____ Q3 = ____ High = ____

5. Now let's look at the data as different sets. Make a back-to-back stemplot of the data by gender in the space below.

   Female   Male

6. Make side-by-side boxplots of the Female and Male data sets in the space below.

7. How do the two distributions compare? Who spends more for lunch?

8. Now look at the data of the costs spent for lunch On / Off campus. Make a back-to-back stemplot of the data in the space below.

On Campus Off Campus

9. Make side-by-side boxplots of the above data sets in the space below.

10. How do the two distributions compare? Where does the lunch cost more?

11. By looking at the stemplots and boxplots you have displayed in this project, which display gives you the clearer picture for the comparisons and why?

Looking at Data–Relationships

Project 2 SPRING BREAK

Spring Break is coming and it's time to make your plans. Below is a list of possible destinations for your trip. Give the following information to assist in planning your trip and finding the information needed for this project.

1. You are leaving from ______________________

2. Date of departure ______________________

3. Date of return ______________________

You are to use the Internet to search for the lowest airfares from your town to the destination on the given dates and find the distance from your airport to that city.

4. Complete the following table of information for your Spring Break adventure. The cost is for one person on a roundtrip ticket. A maximum of two stops in route is allowed, for both directions of the trip.

List the Web sites that you use for your search:

DISTANCES ______________________

AIRFARES ______________________

Destination	Distance	Airline	Departing flight number	Roundtrip cost
Miami				
San Diego				
San Juan				
New York City				
Chicago				
Seattle				
Salt Lake City				
Boston				
Honolulu				
Denver				

5. In the space below, make a scatter plot of the distance versus cost. Use the distance as the independent variable on the horizontal axis. Make appropriate scales for the axes.

6. Look at the scatterplot above to answer the following: How is the cost of the trip associated with the distance of the trip?

7. Use a straight edge to approximate a line of best fit to the data. Approximate the slope of the line and the *y* intercept. Write the equation of your estimated line, $Y = a + b X$

8. On a scale of 0 to 1, how good does the line fit the data? O = no fit and 1 = a perfect fit.

9. Enter the data for the distances and the costs in your calculator. Use L1 for the distances and L2 for the cost. Press STAT, CALC, choose LinReg(a + bx) to find the equation of the least-squares regression line.

10. Find the value of the correlation coefficient for the data. The r value can be found as follows on the TI-83: press VARS, Statistics, EQ, r from the menu screen, or you may go to the catalog and select 'DiagnisticOn' then do the LinReg (a + bx) as indicated in question 9.

11. Sketch the least-squares regression line on the scatterplot above and compare it to the line that you guessed. Are they close to the same line?

12. In part 8 you guessed a value for the strength on the fit to your line. Compare it to the value of the correlation coefficient. How close were you?

13. By looking at the scatterplot and the regression line find a place that is a good buy for the miles traveled. Why do you think the airfare is below the expected value for this destination?

14. Use the destination identified in question 13 and any information found in this project to write a persuasive essay (50-100 words) convincing your parents to let you go to your destination of choice for Spring Break.

Looking at Data–Relationships

Project 3 TEST YOUR MEMORY

1. How well do you know the ages of the stars and celebrities that you hear or see on a regular basis? You are to complete the following table with what you believe to be each person's age. After we have completed the list, you will be given their actual ages.

Name	Your guess of the age	Actual age	Residual	Absolute value of residual	Residual squared
George W. Bush					
Catherine Zeta Jones					
Nick Carter					
Drew Bledsoe					
Whoopi Goldberg					
Al Gore					
Richard Dreyfus					
Ronald Reagan					
Doug Flutie					
Julia Roberts					
Drew Barrymore					
Mr(s). Math					

2. Now that you have filled in your guess, your instructor will give you their actual ages.

3. To see how you have done we will analyze the differences between their actual age and your guess. In the space below draw a coordinate axis system and plot the ordered pairs (X , Y) where X is the actual age and Y is your guess.

4. Now sketch the line Y = X. If your points fall on the line, then you have correctly guessed the ages. If your point is not on the line, then you are to draw a vertical line from your point to the line. The distance of each of these segments is a residual for that guess. List these directed distances in the residual column next to the guess.

5. Did you tend to underestimate, overestimate, or bounce around the line?

6. Sum the residual column. Does your answer tell you how well you guessed the ages overall? Why or why not?

7. There are two choices to consider arriving at a meaningful sum for this data.

    a) Take the absolute value of each value and then find the sum.

    b) Square each residual and then find the sum of the squares.

8. You must decide as a class or a group which procedure above to use as the criteria for the next question.

9. Compare your results with the rest of the class and decide who has the "best" knowledge of the stars' ages.

10. Now use your TI-83 to do this problem. Use the STAT menu to enter the actual age of the star in L1 and your guess in L2.

    - In the STAT PLOT menu, make a scatterplot of the data using L1 as the Xlist and L2 as the Ylist. Choose ZOOM STAT to make a user-friendly window.
    - Use the Y= menu to enter the equation Y1 = X and press the GRAPH key.
    - You should now "see" the graph that you made in the above space.
    - Return to the STAT EDIT menu and define L3 as L2 – L1.
    - Define L4 as ABS(L3).
    - Define L5 as L3 ^ 2.
    - The LIST menu is found by pushing 2nd STAT. Now use the MATH menu to find:

        a) Sum(L4) = _____

        b) Sum (L5) = _____

11. Now that you have completed the above, write your definition of residual in the space below.

Looking at Data–Relationships

Project 4 POPULATION GROWTH

Choose a location that you would like to know the population. This may be a state, a county, a region of the country, or a city. Use the Internet to search the U.S. census reports.

1. List your choice in the space provided.

____

2. You are to research the population given by the U.S. census reports to find the population for each of the previous census reports from 1790-2000. List the populations in the table below:

Year	Population
1790	
1800	
1810	
1820	
1830	
1840	
1850	
1860	
1870	
1880	
1890	
1900	
1910	
1920	
1930	
1940	
1950	
1960	
1970	
1980	
1990	
2000	

3. Use your TI-83 to explore models for the growth of the population over time. Make a plot of the data by choosing STAT PLOT. Use Year as the X List and Population as the Y List for the data set. In the table below give the equation of four possible models of the data set and the value of R^2 for each model. Provide a sketch of the data and graph of the model by using the Graph Link or hand sketch for each of the four models. Attach the graphs to your report for this project.

Model	Equation	R^2

4. Now use the graphs with the supporting information to make a decision on which model best fits the data.

5. Give supporting reasons to your conclusion.

6. Using your model, predict the population in the next census.

Producing Data

Project 5 WHAT DO STUDENTS DRIVE?

What do the students drive to school? Your job, should you decide to accept it, is to design an experiment, gather data, and answer this question.

1. Adjacent to the school you have parking areas. You are to design a method that would produce a representative sample of 30 cars. You may use a block design, a stratified sample, a simple random sample, or other methods you choose. State your design in the space that follows.

2. List the location for each car that you have selected to include in your sample before you venture into the lot.

3. List the results of your sample in the space that follows. You are to identify one more variable to add to the list of characteristics that are given. Write it in the space provided as "other."

Car	Make	Year	Model	Color	Other
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					

4. Summarize the data from your sample to describe a typical car in the lot.

Producing Data

Project 6 JELLYBEAN

How many jellybeans are in the bowl?

1. Make a guess of the number of jellybeans in the bowl and write your guess here.

2. Now it's time to do some "fishing". Each person in the room may use a utensil (the fishing pole) provided by the teacher to remove some jellybeans from the jar. (The total amount removed cannot exceed the number of black jellybeans that you have available.)

3. Replace the jellybeans that were removed with the same number of black jellybeans. (This is similar to the procedure of Wildlife Conservation Agents when the capture, tag, and release wildlife.) Write the number of replacement here.

4. Stir the jellybeans to mix in the black ones.

5. Use the fishing device to remove a sample of the jellybeans from the bowl as in an SRS.

6. Count the jellybeans that you have captured.

   # in total ___________ # of black _________

7. Return your sample to the jar and have each person or group repeat the same sampling method.

8. Write a proportion and solve for the total number of jellybeans in the jar.

9. Now let's see how close we are with our prediction. Everyone participate and divide up the total number of jellybeans and count them! Write down the total number of jellybeans here.

10. Having been through this process, use your knowledge to answer the following question. The Missouri Conservation Commission is to determine the size of the Canadian Geese population in the state of Missouri. You are to design a capture-recapture procedure they could use to estimate the size of the flock.

Project 7 FAIR COIN

Let's flip a coin and record the outcome. As a group, discuss the following items before you begin the experiment:

- How to flip the coin so it is done in the same manner each time.
- Whether or not to turn the coin when you catch it.
- Which side of the coin should start facing up? Alternate heads or tails up, or always the same?
- What do you do if you drop the coin . . . will you count the toss or re-flip?

1. You are to flip your chosen coin according to the agreed upon guidelines above. After each flip you are to record the results as an H (head) or T (tail). Keep track of the total number of heads after each flip and find the percentage of heads after each flip.

   Record the results on the grid that follows:

Trial number	Start up H or T	Outcome H or T	Total number heads	% heads
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				

2. The above data gives you a numerical representation of the behavior of your coin. Make an observation about the percent heads column.

3. Let X represent the trial number and Y the percent of heads after X trials. Plot the set of (X, Y) on the grid below.

1																									
.9																									
.8																									
.7																									
.6																									
.5																									
.4																									
.3																									
.2																									
.1																									
0																									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

4. Make a line plot of the above points on the grid by connecting the points as you move from left to right on the grid.

5. The graphical display of the data on the grid gives you a picture of the behavior of your coin over time. What does your graph indicate to you about your coin?

6. By using both the numeric display and the graph, make a conjecture about your coin.

7. If the other members of the group used different coins, compare the graphs of the various coins. State your observations here.

   Penny

   Nickel

   Dime

   Quarter

8. Using what you have learned, consider the following. To start a volleyball match, the official begins with a coin flip. Do you think that you can call it correctly more than 50% of the time if you see the top side of the coin and know that she will not turn it over? Review your results of the coin flip column 2 and column 3. State what you observe.

9. Is this a fair method of starting the match?

Project 8 ODD AND EVEN

Part 1:
The following grid consists of five rows and nine columns. You are to place a penny on the square marked START. With each throw of the die you will move down one row and left one square if the roll is odd or down one row and right one square if the roll is even. One game consists of four rolls of the die. The game will place your penny at some position on the bottom row. You are to play the game 16 times. Place a tally mark on the square in the bottom row position after each game.

But wait! Before you begin playing, you are to predict the final outcome of the 16 games by placing numbers in the bottom row that you anticipate as your final result. Yes, the sum of your guesses in the bottom row is 16.

Now you may play the game. “Let the good times roll!”

				Start				

_____ _____ _____ _____ _____ _____ _____ _____ _____

1. So, how did you do with your predictions? Were you surprised at your results? Discuss symmetry and the cell locations that are impossible.

2. The model used in this game is called the binomial distribution. List the characteristics of a binomial model in the space below.

3. Use the binomial model to find the probability of ending on each of the five possible final positions of the above game board.

   You may use the binomialpdf function on the TI-83 to do your work, or the binomial theorem. Number the final positions 1, 2, 3, 4, 5 from left to right on the bottom row, and give the probability of ending in that position.

   Position 1 _____

   Position 2 _____

   Position 3 _____

   Position 4 _____

   Position 5 _____

4. Compare the results of the game to the result using the probability as computed above. List your observations in the space provided.

Part 2:
Repeat the game with the following changes in the movement rule. Roll a die and move down one row and left one space if the roll is 1 or 2, or move down one row and right one space if the roll is 3, 4, 5, or 6. List your results here.

Position 1 ______

Position 2 ______

Position 3 ______

Position 4 ______

Position 5 ______

1. How are the two games different?

2. How are the two games similar?

Project 9 THE AGE OF A PENNY

Have you ever wondered how long coins stay in circulation? Are you a collector? You each have a tube of pennies. Your first task is to form a distribution of their ages.

1. Organize the data by using a stemplot of the ages. Split the stems to give sufficient stems to the data.

2. What is the shape of the distribution? Why do you think it is this shape?

3. Did you find any outliers?

4. Do you think the distribution of all pennies in circulation is similar to your sample?

5. List the characteristic assumptions for the Central Limit Theorem, and decide if they are satisfied by your distribution.

6. Find the mean and standard deviation of the ages of the pennies in your sample.

   Mean = ________ S.D. = ________ n = ________

7. Compute a 95% confidence interval for the mean ages of pennies.

   _______ $< \mu <$ ______

8. What is the margin of error for your estimate?

   M.E. = _______

9. The president of "COINS UNLIMITED" has just hired you as his chief statistician for his research on the age of pennies. You are charged with the task of estimating the average age of pennies in circulation within one year of age with 99% confidence. How large of a sample would you need to obtain? Use the standard deviation from your sample as your best estimate of the population standard deviation.

10. Consider your roll of pennies as a population and place a scale of the ages on the number line below. Choose 20 pennies at random from your pile of pennies. Find the mean and standard deviation of the sample and compute a 95% confidence interval for the population mean, μ. Draw a line segment for this interval below the number line that you have scaled. Mix up the pennies and repeat the process five times. Do the intervals of your sample capture the value of μ? Why or why not?

____/____/____/____/____/____μ____/____/____/____/____/____

11. On the basis of your research with this project, how would you define the age of a “rare” coin? Give a statistical definition for your choice.

12. What would be the age of the pennies that you would begin to save before they become hard to find? Give a statistical reason for your choice. Consider “rare” as 2% or less of the population.

Introduction to Inference

Project 10 AHOY MATES

The company claims, "There are over 1000 chips in every bag." How can they make such a claim? As a consumer, your job is to design an experiment to check the claim made by the company.

1. Describe a sampling procedure you would use to gather the data. Include how to select: the bags, cookies in the bag, and chips within the cookie.

2. Consider the sample size you think is needed for the problem. Choose a sample size for the number of cookies that will allow for an approximately normal distribution of the sample mean number of chips within a cookie.

3. You are now ready to begin your count. Most students consider this to be a "crummy" problem. State how you decided to "count" the partial chips?

4. Make a graphical display of the distribution of chips within the cookies.

5. Write a 95% confidence interval for the number of chips per cookie?

6. To consider the number of chips in a bag, you must now consider the number of cookies in the bag. While we have not chosen a large number of bags of these cookies, we can make an observation about the number of cookies in the bag. What is it?

7. Using the distribution of chips per cookie, and the number of cookies in the bag as your guide, write a 95% confidence interval for the number of chips in the bag by making a linear transformation of the data. $\mu_{\text{chips in bag}} = n \ \mu_{\text{chips in cookie}}$ and $\sigma_{\text{chips in bag}} = n \ \sigma_{\text{chips in cookie}}$

8. What do you think about the company's claim?

9. The company has just hired YOU as chief statistician. Address the problem with a hypothesis testing procedure. State the hypotheses (Ho and Ha) for the claim made by the company. Report the value of the test statistic and the p-value of the test.

10. Write a word conclusion (approximately 50 words) to the problem that you would submit to the company president.

11. Identify five items relating to statistics that you have learned from this project.

    1)

    2)

    3)

    4)

    5)

You may now eat the data. Got milk?

Inference for Distributions

Project 11   SNAP CRACKLE POP

Are all cereals created equal? You are to gather data on different brands and types of cereals.

1. Go to the grocery store with pen and paper in hand, and gather some data on rice and oat cereals. Classify the cereal as to the first ingredient listed if it is made from a combination of grains. Record the data on a per-serving basis as follows:

Brand	Type	Calories	Total fat (g)	Sodium (mg)
	Rice			
	Rice			
	Rice			
	Rice			
	Rice			
	Rice			
	Rice			
	Rice			
	Rice			
	Rice			
	Oat			
	Oat			
	Oat			
	Oat			
	Oat			
	Oat			
	Oat			
	Oat			
	Oat			
	Oat			

In the data that you have collected, assume that the data is normally distributed for all cereals.

2. Perform a hypothesis test for the difference of the two means for each of the following questions. List the Ho, Ha, give the observed value of the test statistic, p-value, and write a word conclusion.

   a) Is there a difference in the calorie count for the two types of cereals?

   b) Do the rice cereals have less fat grams per serving than the oat cereals?

   c) Do the rice cereals have more sodium (mg) per serving than the oat cereals?

Inference for Proportions

Project 12   TASTE THE DIFFERENCE

Can you really taste the difference in the brown M&M's?

1. You are to select a partner to perform the following experiment. One person will record, the other will taste. After completion, you will reverse roles and repeat the experiment.

2. Take a small number (approximately 5) of brown M&M's and train your taste buds by eating them slowly one at a time.

3. Now take a non-brown M&M and eat it slowly. Sorry, no more than five tries of knowing each color, brown vs. non-brown. After all, if you eat all the data, we can't run the taste test!

You must trust your partner on this one to record the data, as you are the lucky one that gets to eat your way through.

4. Close your eyes; blindfolding is necessary to prevent looking at the candies.

5. The person recording should choose 10 M&M's of random colors. Make sure that there is a good color mixture, but no more than five brown. One by one, the taster should eat the candies given to you by your partner, and tell him/her if you think it is brown or not brown. The partner is to record the results as correct or incorrect for each trial. Tell the taster if they are correct after each trial. Record the results for each of the 10 trials as T = true for a correct guess and F = false for an incorrect guess in the chart below.

Trial #	True or False
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

6. This experiment consists of independent trials, each of which has two possible outcomes, success or failure, where the probability of success and failure add to one. If a person has no idea as to the color, then the probability of success is .5 for every trial. Count the number of successes that were observed in the 10 trials above. Find the probability of that number of successes if indeed you were guessing. You may use the binomial theorem or use the binomialpdf function on the TI-83 calculator. Express the method of your computation in the space below.

7. Make a probability density function for the random variable of the number of successful guesses in the 10 trials. Use p = .5 and list the results in the table below.

Number	Probability
0	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	

8. Find the expected value of the number of successes for the distribution from the above.

9. Find the standard deviation for the above distribution.

10. Let n equal the number of successes that you observed in your taste test. By using the standard normal distribution, find the P ($X \geq n$). Use the continuity correction factor.

11. On the TI 83, choose the binomialcdf function and find the P ($X \geq n$).

12. Compare your answers to question 10 and question 11. Are they the same or different? Why?

13. So how did you do? Do you think now that you really can tell the difference? Are you willing to bet your leftover M&M's?

14. Switch sides of the table and repeat the taste test with your partner. Record the results here.

15. So who did the better tasting? _____________ Repeat the experiment so that each of you have 20 trials. Let k equal the total successes in the 20 trials, and find P ($X \geq k$) by using the standard normal distribution with $n = 20$ and $p = .5$.

16. Test the hypothesis that you are guessing vs. the alternative that you can tell the difference. Use the p-value techniques to test your hypothesis and write a word conclusion that could be used as evidence in court to support your findings.

Now eat the evidence so that the next hour class is unaware of "how sweet it is" in here!

Inference for Two-Way Tables

Project 13   PLAIN AND PEANUT

What is your favorite color of M&M? Can you really taste the difference? When you buy the bag do you eat them first or last? The real question here is to address the proportions of colors of M&M candies. Over time, the company has changed the colors and proportions.   So, what are the proportions now? You will discover the answers to these and other questions as you proceed with this project.

1. Identify the colors of plain M&M's that are in the package? Name the colors.

2. Before you begin the data collection, state the percentages of colors that you *think* are in the bag.

3. Discuss with the class and determine a procedure for counting the M&M's. Consider the bag sizes in making your choice and the size of the sample. Identify your procedure here.

4. Make a frequency distribution of the counts and give the relative frequency for each.

5. The “color” of M&M is what type of data? __________________

6. Graph the data by using an appropriate chart. State why you choose your type of graph.

7. Repeat the procedure using peanut M&M’s. What are the colors used by the company?

8. Make frequency distributions of the counts and find the relative frequencies.

9. Graph the data using the same type of graph used in the plain M&M data.

10. Compare and contrast the two distributions and graphs.

11. How do we know what the company intends to be the proportions for each color? Consult the Web site for the MARS Company.

12. List the findings from the company Web site for both the plain and peanut M&M's.

    a) Plain
    List the colors and the company's proportion claim.

    b) Peanut
    List the colors and the company's proportion claim.

13. Does the data that you have gathered agree with the company's claims?

14. Identify a procedure that you could use to check the goodness of fits for your data.

15. Identify the Ho and Ha you would use to verify the company's claim.

    Ho:

    Ha:

16. Display in a matrix the observed and expected cell values and perform a Chi-Square analysis for the (a) plain and (b) peanut data.

    a) Plain

Color	Observed	Expected	(O-E)^2 / E

b) Peanut

Color	Observed	Expected	(O-E)^2 / E

17. Write a word conclusion to the hypothesis test you have performed using the 5% level of significance.

    a) Plain

    b) Peanut

18. Now consider the type of M&M and the color. Use the two samples that you counted to test the hypothesis that the type of M&M is independent of the colors as listed. State the hypotheses, the value of the test statistic, the p-value, and write a word conclusion.

    Use the following grid cells to display your data using the columns as types and the rows as colors.

	Plain	Peanut	Total
Red			
Green			
Yellow			
Blue			
Purple			
Total			

Inference for Regression

Project 14   SHORT OR TALL

Do tall people have big feet and short people have small feet?   Your task is to generate a set of data values for the height in inches and their shoe size.

1. In the chart below list at least 15 people and their data. You are to standardize the sizes for male and female.

Name	Height	Shoe size

2. Make a scatterplot of the data where shoe size depends upon the height.

3. Find the line of best fit to the data and write the equation.

4. Make a residual plot of the data here.

5. What does your residual plot indicate to you about the model?

6. Perform a hypothesis test to determine if there is a significant positive slope to the line of best fit. You may use the LINREG t test menu on your TI-83.

    - Write both hypotheses.

    - State the t-value.

    - State the P-value for this test.

    - Write a word conclusion.

7. Suppose Sue Z. Normal walks into your classroom. Find a point estimate for the shoe size of Sue Z. Normal who is 70 inches tall.

8. Find a 95% confidence interval for the shoe size of Sue Z. Normal.

9. You may not have had anyone in your sample that was over seven feet tall, but there are seven footers around, particularly on pro basketball teams. Predict the shoe size for Mr. Guy Tall, who measured in at 7 feet 2 inches for Coach Woy Rilliams at the basketball tryouts.

10. Find a 95% prediction interval for Mr. Guy Tall's shoe size.

11. State why the prediction interval for Mr. Tall is wider than the confidence interval for Sue Z. Normal.

Multiple Regression

Project 15   MANY VARIABLES DO PREDICT

Can you predict your arrival time at school? While there are many variables that could impact this time, we will consider only the two variables of the distance that you are from school and the time that you leave home to go directly to school.

1. Gather information using the time in minutes starting at t = 0 as 7:00 am and give the distance in tenths of a mile. Each member of the class is to record the data and return it tomorrow for class.

Record the information in the table. Use a minimum of 25 sets of values.

Arrival time	Distance	Start time

2. Let Y = the arrival time, X1 = the distance, and X2 = the start time to find a multiple regression equation for the relationship. You are to use a spreadsheet to enter the data and generate a multiple regression output for the variables defined above. Excel or Minitab will output this information. Attach a copy of the output to this report.

3. Write the estimated equation from the output.

4. State the hypothesis tested by the ANOVA F statistic for this problem in words.

5. State your conclusion to the hypothesis based on the computer output.

6. What percent of the variation in arrival time is explained by the distance and start time?

7. If you live 5.3 miles from school and leave at 7:15 am, what is your expected arrival time?

One-Way Analysis of Variance

Project 16 GROWTH GROUPS

As you wander the hallways of the school, you note various heights of students. For this project you are to select a random sample by means of a convenient method to compare the four grade levels of students and their heights.

1. List the heights of a minimum of 15 students of the same gender from each of the four grade levels in the space provided.

Freshman	Sophomore	Junior	Senior

2. Make side-by-side boxplots for the data sets listed above.

3. Summarize the relationship of the boxplots. How are they alike and how are they different?

4. Make a normal quantile plot for the data of each of the four grades, and state your conclusion regarding the normality of each.

    a) Freshman

    b) Sophomore

    c) Junior

    d) Senior

5. Summarize the data for each class in the table below.

	Freshman	Sophomore	Junior	Senior
Mean				
Standard deviation				
Sample size				
Standard error				

6. Assume that the groups have equal standard deviation and perform the one-way ANOVA test. State the hypotheses.

7. State the degrees of freedom for the ANOVA test.

8. Summarize the results in the ANOVA table below.

Source	Degrees of freedom	Sum of squares	Mean sum of squares	F
Groups				
Error				
Total				

9. You may use a computer output to do the computation for this problem and attach a copy of its output to this project.

10. State the critical value of the test statistic. Use an α level of 5%.

11. Summarize the conclusion to the hypothesis test. Include the P-value and state your results in terms of the variables in the problem.

ANOVA

Project 17 RANDOM ASSIGNMENT

In a large school where the students are assigned to their class schedule by computer, it would sound reasonable that the students within all classes of the same topic for example, Algebra II would have a similar distribution of last names.

Your instructor has noted that in each of his classes this year the students seem to be clustered in an alphabetical grouping. Morning classes tend to have the students with last names like Adams, Beltron, and Jones, but the afternoon classes seem to have Smith, Thompson, and Young.

Is this true in your classes? Choose an instructor and make a distribution of the first letter of the last name for each member of the class. You are to use three classes for this project. Recall that we are checking for the assignment of students within the subject matter, so more than one section of a class must exist. Assign a numeric value to each letter of the alphabet as follows: a = 1, b = 2, c = 3, etc.

While the distribution of values will not be normal, the central limit theorem indicates that the sample means will tend to approach a normal distribution as the sample size approaches 30. Use this fact to decide if the students in the classes could be considered a random assignment to the daily class schedule.

1. Make a stemplot of the data for each class in the space below.

2. What relationships, similar or different, do you see in the graphs?

3. Make side-by-side boxplots for the above data in the space below.

4. Compare the plots and state you conclusions.

5. Perform a hypothesis test by comparing the means of the classes using a one way ANOVA on the three distributions. Write the Ho, Ha, give the value of the F statistic, and state the P-value for this test.

6. Write a word conclusion that you could present to the principal of your school regarding the random assignment of students to the various periods of the daily schedule.

Statistics for Quality

Project 18 SKITTLES MACHINES

You have just been hired as the quality control engineer for Skittles. Your job, because you decided to accept it, is to watch the process of the bagging of the fun size bags to insure that they are meeting the company's quality standards is the following ways:

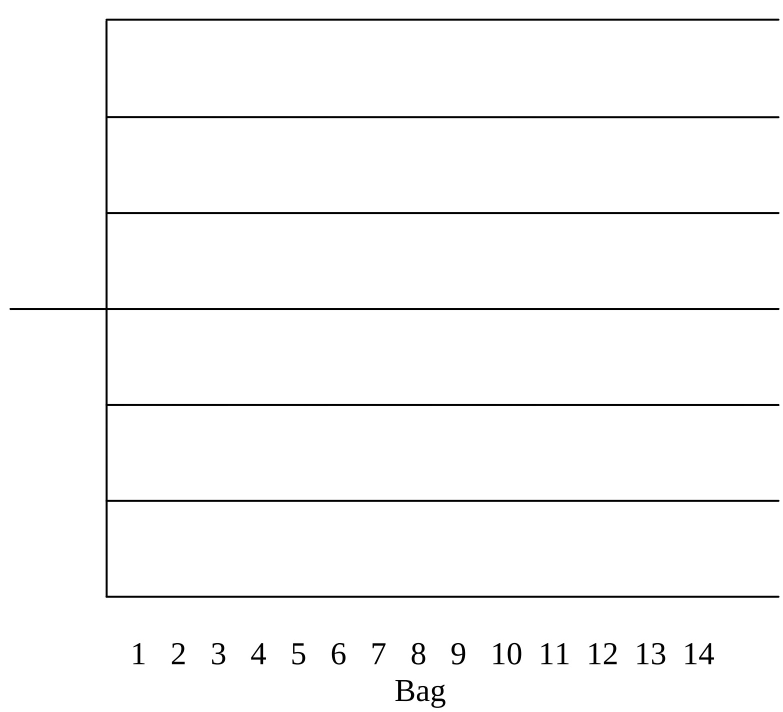
A. Are the number of candies in the bag within the company's guidelines?

B. Are the colors of each candy within the proportion guidelines?

Assume that your bags of fun size Skittles are being selected randomly from the bagging machine at one hour intervals. Select the bags one at a time and count the data to assess if the machine is functioning properly. You are to consider the following three methods of control:

i. Any point outside three standard deviations from the mean.
ii. Any nine consecutive values on the same side of the mean.
iii. Any two of three points more than two standard deviations on the same side from the mean.

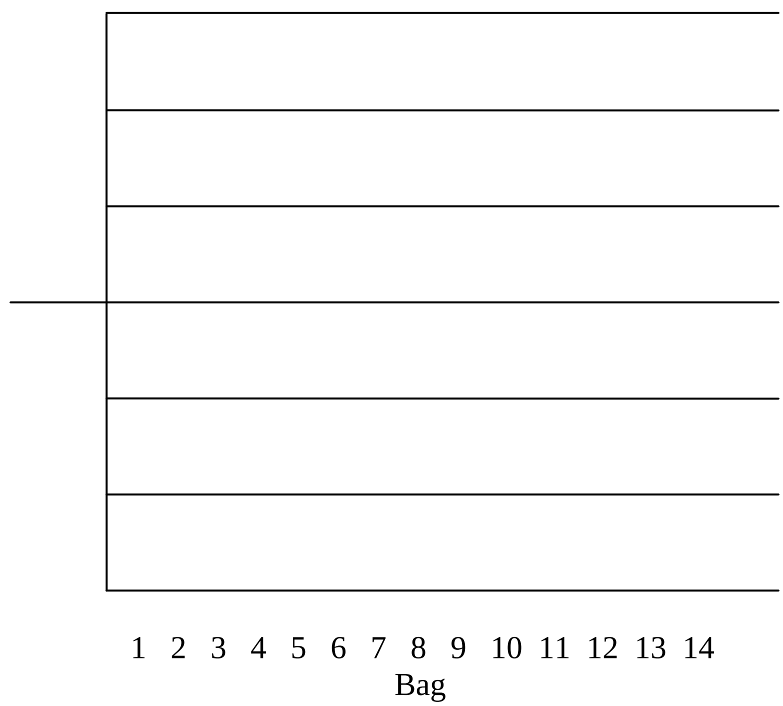
1. In the space below make a quality control chart for the number of Skittles in the bag. Company standards are for the bags to have a normal distribution with mean 18.2 and standard deviation 0.6. Experience has shown that the machine produces bags with equal proportion of the five colors. Label the control values of the vertical axis. Plot the number of candies in each bag, in order selected, on the horizontal scale.



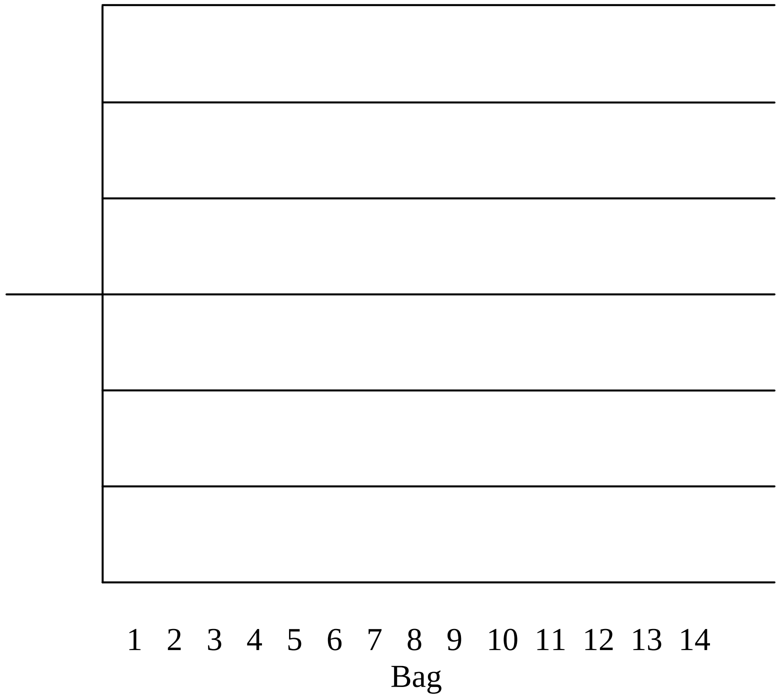
Use the above chart to determine if the process is in control or indicate the point where you have determined it to be out of control.

2. The bagging machine is fed by machines that produce each color of the candy. You wish to determine if each of them is remaining in control throughout the day. You are to make a control chart for the proportion of each color. Assume that the colors are of equal proportion (p = .2). Make the control limits for the proportion and then plot the proportions from each bag by color, as you sample.

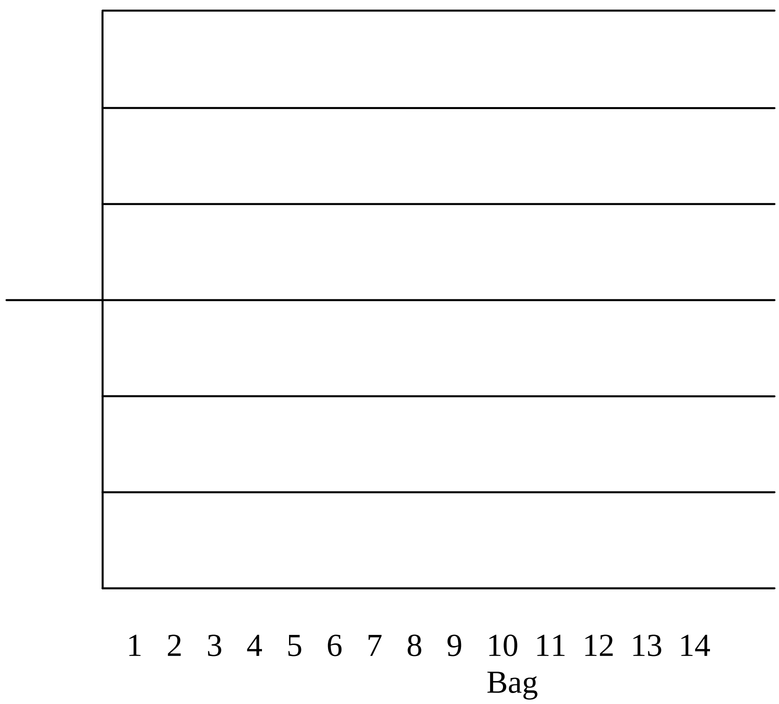
Color RED



Color Purple



Color Orange



Color Yellow

1 2 3 4 5 6 7 8 9 10 11 12 13 14
Bag

Color Green

1 2 3 4 5 6 7 8 9 10 11 12 13 14
Bag

3. Write your conclusions about the performance quality of the bagging machine and each of the color production machines in the space below as a report to your boss on the day's activity.

A) Bagging machine

B) Red

C) Purple

D) Orange

E) Yellow

G) Green

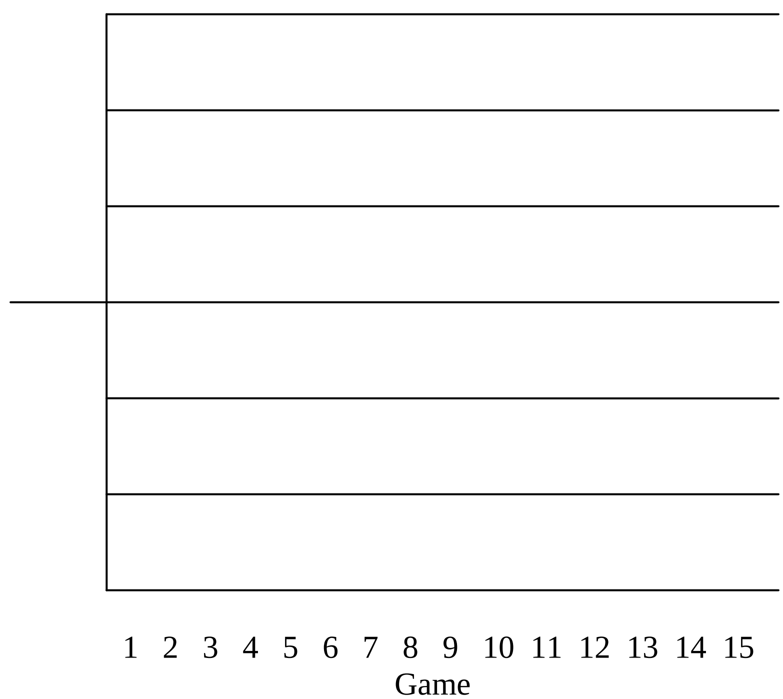
Project 19   NBA

You are to select a team from the NBA to do the following study by making a control chart for your team. As an avid fan of basketball, you have decided to chart the points scored by your team, the ____________________________. You may use the Web site www.nba.com to find the data for this project as a retrospective study, or chart the next 15 games that your team plays by checking the papers on a daily basis.

Past experience has shown the NBA teams average 98 points per game with a standard deviation of 3 points per game. You expect your team to be consistent. Consider the following three measures of them experiencing a significant change:

i. Any game outside three standard deviations of the mean.
ii. Any nine in a row on the same side of the mean.
iii. Any two of three more than two standard deviations on the same side of the mean.

Team ________________________________



____

List the date for each game for reference.

Write a summary of your team's performance over the past 15 games. Note any trends that appear on your graph. If anything happened to any player on the team during the time or if any other significant event(s) took place, which could have accounted for the change in team performance, please reference them in your report.

Research Project

Project Final

Within this project you must demonstrate five different aspects of a statistical study. The project should be three to five pages in length. You may use either an observational study or a designed experiment. Your instructor must approve your topic of research before you begin.

1. Pose a question that you can answer with data (this must be pre-approved).

2. Explain how you will select the data or design the experiment to gather the data. Give a rationale for your choices.

3. List the data values you obtain and choose an appropriate graphical display for the data. The descriptive methods are to relate to the question that you have asked.

4. Use a hypothesis testing procedure to analyze the data. Give supporting reasons for the choice of the procedure used. Summarize your conclusion in terms of the question that you posed. Include the p-value of the test.

5. The final paragraph should summarize what you have done in the form of a letter to your Grandma or Grandpa, using language that she or he will understand.