M2T2

Expect the Unexpected: Probability, Data, and Statistics

Statistics

STATE GOAL 10: Collect, organize and analyze data using statistical methods; predict results; interpret uncertainty using concepts of probability.

Statement of Purpose

We live in a data-filled world. Opinion polls, stock prices, tax rates, crime statistics, scientific studies, and weather reports are increasingly a part of our daily lives. The activities here use a laboratory approach to solving problems in probability. Students gain experience in designing an experiment to solve a problem, carrying out the experiment, analyzing the results and making conclusions that are based on the data. An important part of the work includes preparing a written report on the experiment. This module also engages students in data collection activities—keeping a daily log and designing a survey. These activities are intended to provide students with data about which they have a feeling of ownership, and give opportunities to organize data, make sense of variables and patterns, and write about the findings in convincing ways.

Highlights

Overview: This module deals first with some ideas in probability, then with statistics. As will be shown in this module, the topics of probability and statistics are linked together in many ways.

Connections to the Illinois Learning Standards.

Standard 10.A.—Organize, describe, and make predictions from existing data. Throughout the module, participants are asked to organize their data in the form of a report and make predictions based on the results of data gained through experiments or data they are given.

Standard 10.B.—Formulate questions, design data collection methods, gather and analyze data and communicate findings. Participants in this module are asked to create a survey and gather data. They then analyze the data and use graphs to interpret and communicate their results.

Standard 10.C.—Determine, describe, and apply the probabilities of events. This module begins with several activities involving probability: The Cereal Box Problem, The Rocket Launch, The Traffic Light, and The Dice Game are all explorations of probability.

Table of Contents

Appendices

Minimal:

- Pencils, blank paper, rulers, & protractors
- Graph paper
- Polyhedral Dice
- Chips or checkers or small blocks
  (10 per participant)

Optimal list will include:

- Microsoft Excel® or Clarisworks®, or other spreadsheet software
- An Internet browser

M2T2

Instructor Page

Probability (Models): Design, carry out and report on experiments to solve problems in probability.

A **model** is a representation (either concrete or abstract) of some event, such as getting one out of six different animal cards as a prize in a cereal box.

Expected Value: The Cereal Box Problem

In this module, we differ from the usual treatment of probability. A common approach to the study of probability is to start out with counting methods. For example, if you have a jar of marbles with 5 green and 7 red marbles in it, then the probability of picking a green marble at random is 5 out of 12. However, in many situations, the calculation of the number of favorable outcomes or possible outcomes can be extremely difficult. Nonetheless, it is still possible to find these probabilities using simulation methods.

Simulation methods are important and useful to statisticians, for example, in weather forecasting. Moreover, since probability theory is the study of random processes, exposing students to many examples of circumstances that require judgments under conditions that are uncertain is good preparation for working with theoretical probability.

1. Here we deal with theoretical counting methods (permutations and combinations) only as optional activities, leaving a more thorough treatment until later. (See, for example, Benchmark 10.C.5b, late high school)
2. We introduce the concept of *expected value* before the idea of the likelihood of an event. Our classroom experience, and related research, have demonstrated that expected value is a natural idea, accessible to students, and helps to prepare the way to a more in depth look at probability.
3. We emphasize a 'frequency approach' to probability, but introduce at the same time, through the concept of a 'probabilistic model', the conventional notion of probability as:

$$\textbf{P(Event)} = \frac{\text{Number of favorable outcomes}}{\text{Number of possible outcomes}}$$

Guideposts

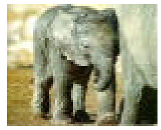
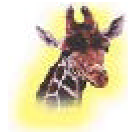
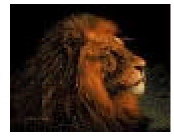
- Introduce the problem and have students make guesses about the number of cereal boxes. Discuss how the problem could be modeled. Since there are six prizes, and each is assumed to be selected with equal likelihood, it is natural to model the problem with six-sided dice.
- The **Expected Value** of an event (such as number of cereal boxes required to get all six prizes) is the average number of desired outcomes for many trials of an experiment.

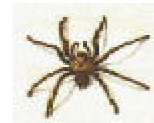
Internet Resources:

An interactive version of this problem is on the Web at <www.mste.uiuc.edu/reese/cereal/

Participant Page

Expected Value: The Cereal Box Problem



Suppose that a cereal company is promoting its product by offering one of six different animal cards as a prize in each of its cereal boxes.

Suppose also that your chances of getting any of the six cards are the same. *About how many* boxes of cereal would you have to buy to get all six cards?

A Solution:

How could we solve this problem? One way would be to go out and start buying cereal, opening up each box, and keeping a record of the different cards we get. Use a tally sheet to keep a record. We are not suggesting that we go and actually buy cereal! Instead, we can do an experiment in class and collect data *just as if* we went out and bought cereal.

How could we do an experiment where we pretend that we are buying a box of cereal, and getting one of the six animal cards by chance? We need a **model**.

Suggestions for a model

- A bag containing six different colored, but equal-sized marbles. Each color represents one of the cards. Take one out, write down its color. Put it back in bag.

- A hat containing six different labeled, but same-sized pieces of paper. Take one out, write down its name of the animal on the paper, and put back (replace) in hat.

- One six-sided die like is used in board games. Each of the sides corresponds to a different animal card.

M2T2

Instructor Page

Illinois Learning Standard Benchmark 10. B.3 Formulate questions, devise and conduct experiments or simulations.

Illinois Learning Standard Benchmark 10 C.3b. Analyze problem situations and make predictions about results

Variations on the Cereal Box Problem

1. Suppose there are eight different animals offered as prizes instead of six. About how many boxes of cereal will it take now to get all colors ?

2. Suppose there are only four colors. Now how many boxes of cereal will it take ?

Expected Value: Cereal Box Problem

Five Step Approach to Solving a Probability Problem

(A sample student response is provided in italics)

1. Model: *Six-sided die* (why? *One for each type of animal)* Of course, other models are possible.

2. Trial: ('shopping trip') *Roll die until all six sides (six colors) are obtained.* Use 'tally sheet' to keep record of colors.

3. Outcome of trial: *Number of rolls of die ('boxes of cereal') to get all six outcomes ('animal cards')*

4. Repeat trials: Do a 'sufficiently large' number of trials (rolls of die or shopping trips). *I did 7 rolls (see tally sheet below).*

5. Find average number of outcomes (boxes):

$$= \frac{\text{Total boxes}}{\text{Number trips}} \quad \frac{136}{7} = 19.4$$

6. Find the average for the whole class. The theoretical expected value is 14.7 and the class average should be in this area.

	Card #1 Alligator	Card #2 Elephant	Card #3 Giraffe	Card #4 Lion	Card #5 Otters	Card #6 Spider	Number Boxes
Trip #1	///	/	/	/////	////	//	16
Trip #2	//	/	//	///	/	///	12
Trip #3	/////	/	//	///	//	//	15
Trip #4	///	//////	/	/////	//////	//////	27
Trip #5	///	//////	/////	/	////	////	23
Trip #6	////	//	///	//////	/	///	19
Trip #7	/	/////	/////	//	///	////// //	24

Total number of boxes = 136

M2T2

Participant Page

Expected Value: The Cereal Box Problem

Suppose that a cereal company is promoting its product by offering one of six different animal cards as a prize in each of its cereal boxes. Suppose also that your chances of getting any of the six cards are the same.

About how many boxes of cereal would you have to buy to get all six cards?

First, make a guess: ___________

You should become familiar with the following terms:

- **Expected Value**
- **Model**
- **Trial**
- **Outcome**

Five Step Approach to Solving a Probability Problem

1. Model: Do an experiment without having to go out and buy cereal. What model could you use for "buying a box of cereal" and getting one of the six animal cards?

2. Trial: "Purchase" cereal boxes until you have all six animal cards.

3. Outcome of trial: Number of Cereal boxes purchased in one trial.

4. Repeat trials:
Do a 'sufficiently large' number of trials. The more the better – at least 30 – but you can combine with others in your class...

5. Find average number of outcomes (boxes) :

Total number of boxes =
Total number of trips =

Average number of boxes =

6. Find the average for the entire class:

Model of the Cereal Box Problem

	Card #1 Alligator	Card #2 Elephant	Card #3 Giraffe	Card #4 Lion	Card #5 Otters	Card #6 Spiders	Number Boxes
Trip #1							
Trip #2							
Trip #3							
Trip #4							
Trip #5							
Trip #6							
Trip #7							
Trip #8							
Trip #9							
Trip #10							

Write up your results using the lab-report format on the following page.

M2T2

Instructor Page

Optional topic:
A theoretical solution to this problem, available in probability texts[1] is found by the series: $6/1 + 6/2 + 6/3 + 6/4 + 6/5 + 6/6 = 14.7$.

[1] For example, see Feller, W. (1957/1950). An introduction to probability theory and its applications. New York, NY, John Wiley & Sons

Mathematics Laboratory Report: Cereal Box Problem

A. What was the problem to be investigated

The problem was to estimate how many boxes of cereal it would take to get all six different animal cards.

B. What was already known about this problem situation?

It was assumed that each animal card is available in the cereal boxes in equal amounts (there is equal likelihood of getting each of the six cards).

C. How I got the data to solve the problem

I followed the five steps to get the data. The most important part was to find a model, like a number cube (6-sided die) or six differently-colored, same-sized marbles, in a hat.

D. The data that I gathered

I rolled a six-sided die until I got all six outcomes of the die (all six colors). The set of rolls to get all six outcomes is a 'shopping trip'. As a class, we needed to do at least 30 shopping trips.

E. What I found out

I calculated the average number of boxes in a 'shopping trip'. This gives an estimate of the number of boxes of cereal to expect to have to buy to get at least one of each of the six different cards.

Internet Resources:
Check out the Java simulation for this problem at
www.mste.uiuc.edu/reese/cereal/cereal.html

M2T2

Participant Page

Mathematics Laboratory Report:
Cereal Box Problem

A. What was the problem to be investigated?

B. What was already known about this problem situation?

C. How I got the data to solve the problem:

D. The data that I gathered:

E. What I found out:

A Web Page to use with extending this activity is at www.mste.uiuc.edu/reese/cereal/

Instructor Page

Independent Events: Rocket Launch Problem

{See page F-11 for a description of the Rocket Launch Problem}

Optional Topic:
Theoretical solution - At this point, instructors could introduce how to calculate the probability of two or more independent events.
(Independence enables one to multiply the probabilities:
½ x ½ = ¼ .)
Then, students can compare their experimental results with the theoretical probabilities that they can now obtain.

The main idea here is that for independent events, the probabilities multiply. Participants can see this with repeated trials using the model.

Sample Response for the Five Step Approach

1. **Model**: Coin
(need probability of .5 —> heads = "fire" and tails =" not fire")

2. **Trial**: (launch attempt) Toss coin twice (why ? once for each stage)
Record outcomes: (Number of heads)
Repeat ten times (one for each launch attempt)

3. **Results of trial:** In how many of the ten 'launch attempts' did we get lift off?

4. **Repeat:** Repeat at least 30 times.

5. **Average number of 'liftoffs'** Find average number of times the rocket would lift off (out of ten attempted launches).

Guidepost:
You may need to discuss with the participants the reason for using a coin for modeling the problem, or what other models could be used.

A variation on rocket problem:
1. Suppose the rocket has three stages instead of two, each with a probability of firing of .5. In 10 attempts to launch, about how many liftoffs can be expected ?

2. In 10 attempts to launch, about how many liftoffs can be expected if each of three stages fires with probability of .75 ?

3. Suppose the probability of firing is: Stage I .5; Stage II .75; Stage III .80. About how many liftoffs in 10 attempts can be expected?

The Web page below has a Java applet that models this problem.

Internet Resource:
Simulation of this problem at <www.mste.uiuc.edu/activity/rocket>

M2T2

Participant Page

Independent Events: Rocket Launch Problem

Terms to learn:

- **Independence**
- **Theoretical probability**

Suppose that an experimental space rocket has two stages, and that the probability that each stage fires is .5. Assume that each stage fires independently of the other. In order for the rocket to lift off, both stages must fire. In 10 attempts to launch, *about how many,* will result in the rocket lifting off ?

What is your guess ? ______

Five step approach:

1. **Model**: (How to model the situation without actually building a rocket.) :

2. **Trial**: (What will one trial consist of ?)

What is a successful trial?

3. **Results of 10 trials**

_____ _____ _____ _____ _____ _____ _____ _____ _____ _____

_____ _____ _____ _____ _____ _____ _____ _____ _____ _____

How many of your ten launch attempts had successful liftoffs?

4. **Repeat:**

_____ _____ _____ _____ _____ _____ _____ _____ _____ _____

_____ _____ _____ _____ _____ _____ _____ _____ _____ _____

5. **Your average number of 'liftoffs':**

6. **Average for the entire class:**

Instructor Page

Sample Spaces in Probability: What are the chances?

{See page 13 for a description of the Three Child Family Problem}

A data sheet for this activity is available in Appendix B.

The **probability** of an event (like one or more boys in a three child family) is defined as:

$$\textbf{P (Event)} = \frac{\text{Number of favorable outcomes}}{\text{Number of possible outcomes}}$$

In the case of the three child family problem (page 13) we do an experiment to estimate the probability of 1 or more boys.

That is, $\textbf{P(1 or more boys)} = \dfrac{\text{Number of families with 1 or more boys}}{\text{Total number of families}}$

We use the five step approach.

1. **Model:** Coin (Why? *H=boy; T=girl*)

2. **Trial:** Toss coin 3 times *(Why 3? one for each child)*

3. **Results of Trial**: Trial is a success if 1 or more Heads (boys) are obtained in 3 tosses.

4. **Repeat:** Do a large number of trials. At least 10.

5. **Find probability of success:** Probability = Successes/ Total Tosses

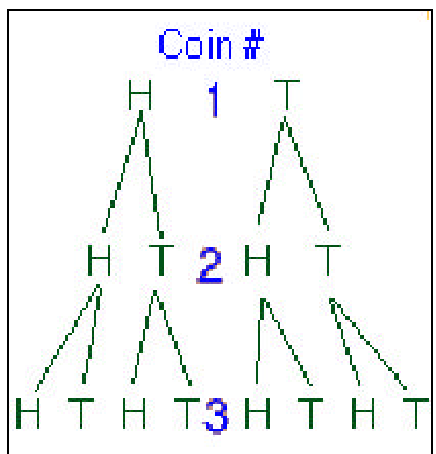
Then write up a laboratory report.

Theoretical solution - Teachers can have students generate a list of all 8 possible outcomes and circle all that would represent a "success":

HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

$$\textbf{P(1 or more boys)} = \frac{\text{Number of families with 1 or more boys}}{\text{Total number of families}} = 7/8$$

Participants may also display the sample space for this problem in the form of a tree diagram like the one at right. Then, students can compare their experimental and their theoretical results .



Tree diagram of the sample space

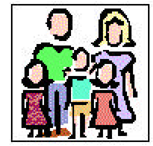
M2T2

Sample Spaces in Probability: What are the chances?

Participant Page

Three child family

There is a new family, with three children, moving in next door.
What are the chances that at least one of the children is a boy?



Five step approach:

1. Model:

2. Trial:

Successful trial (describe):

3. Results of 10 trials:

__ __ __   __ __ __   __ __ __   __ __ __   __ __ __

__ __ __   __ __ __   __ __ __   __ __ __   __ __ __

4. Repeat:

5. Find the probability of success and interpret:

Instructor Page

More Probability: Traffic Lights

{See page 15 for a description of the Traffic Lights Problem}

Five step approach:

1) Model: A light is green for 40 out of 60 seconds. So we need a model that gives 40 successes out of 60 attempts or 4 out of 6 (that is, 4/6 or 2/3).
   We could use a six –sided die:
   Sides 1-4 ==> green
   Sides 5-6 ==> not green

2) Trial: Toss the die twice (Why ? Once for each light)

3) Results of Trial: A successful trial is when we get green for both lights (that is sides 1 – 4 on the die both times).

4) Repeat: Carry out a specified number of trials

5) Find probability of success and then interpret result.

Write up your experiment, including the actual results.

M2T2

Participant Page

More Probability: Traffic Lights

Traffic lights

When Jim and Jane drive to work there are two traffic lights along the way. Each light is on a sixty-second cycle and is green for forty seconds. Assuming each light operates independently of the others, what is the probability that both lights will be green ?

Five step approach:

1. **Model**:

2. **Trial**:

3. **Results of trial:**

4. **Repeat:**

5. **Find probability of success and interpret:**

M2T2

Instructor Page

The Dice Game

Objectives:

1. To recognize the probabilities of sums (such as when rolling a pair of six-sided dice)
2. To recognize outcomes that are not equally likely
3. To describe the sample space of an event
4. To compare theoretical and experimental probabilities

Step 1:

a. Distribute a chart and 10 chips or blocks to each participant. Tell them to distribute the chips in any manner that they like. (They can place any amount of chips on any number, but all 10 chips must be used.)

b. Instructor begins rolling a pair of dice and calling out the sum. Participants can remove 1 chip from the number called, if there was a chip on that number. The first person to remove all their chips is the winner.

c. Repeat the game several times until students begin to notice that the probabilities of the sums are not equal.

Step 2: Discuss the results of the game and talk about what a sample space is.

Step 3: Make a chart showing the theoretical probability of obtaining each possible sum when rolling a pair of dice.

Step 4: Students can practice finding the theoretical probability of rolling a given sum, such as P(sum of 7) = 6/36 = 1/6 and they can compare these theoretical probabilities with their experimental results.

Possible Sums for Two Dice						
SUMS	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

M2T2

Participant Page

The Dice Game

Distribute 10 chips over the "balloons" below. You can place ANY amount of chips on any given number. In the game, a pair of dice will be rolled, and the sum will be called out. If you have a chip on that number, you can remove it.

1 ____________________

2 ____________________

3 ____________________

4 ____________________

5 ____________________

6 ____________________

7 ____________________

8 ____________________

9 ____________________

10 ____________________

11 ____________________

12 ____________________

M2T2

Instructor Page

Introduction to Data Analysis: One Variable Statistics

Before learning a variety of ways to analyze data, students can collect their own data based upon their own daily activities. Several days before beginning this data analysis unit, students can begin charting the amount of time they spend on their various activities (on a log sheet such as the one in Appendix E).

Goal 10 Standard B: Formulate questions, design data collection methods, gather and analyze data and communicate findings

Hints:

While performing some of the calculations in your summary table, you will probably encounter some decimal answers.

We found it easiest to round these answers to whole numbers, although later, columns four and five might not add up to their expected sums (100% or 360°).

This would be a good time to review the rules for rounding with students.

"How I Spend My Time"

Step 1: Have participants classify all their daily activities into at most 6 categories. List these activities under "Type of Activity" in your summary table.

Step 2: Add up the total hours for each activity over the duration of your log sheet.

Step 3: Find the average hours per day by dividing the "Total Hours" by number of days (7 if the week-long chart was used).

Step 4: Next, calculate the percent of time in a day spent on each activity by dividing "Average Hours per Day" by 24 (to express as a percent, move the decimal two places to the right).

Step 5: Find the central angle for a circle graph by multiplying the decimal (percent) in column four by 360°.

Step 6: To begin constructing the circle graph, first draw a radius anywhere within the circle to use as your first side of an angle. Continue with the formation of the circle graph by using a protractor to measure and draw each central angle.

Sample

Summary Table for Log Sheet

Type of Activity	Total Hours	Average Hours per	% of Day	Central Angle
Sleeping	58	58/7=8.28	8/24=.33	.33x360°
School	35	5	21%	76°
Eating	14	2	8%	29°
Homework	11	2	8%	29°
Electronic Media (TV, Movies,	28	4	17%	61°
Socialization/	22	3	13%	47°
TOTAL:	**168**	**23**	**100%**	**361°**

Participant Page

Introduction to Data Analysis: One Variable Statistics

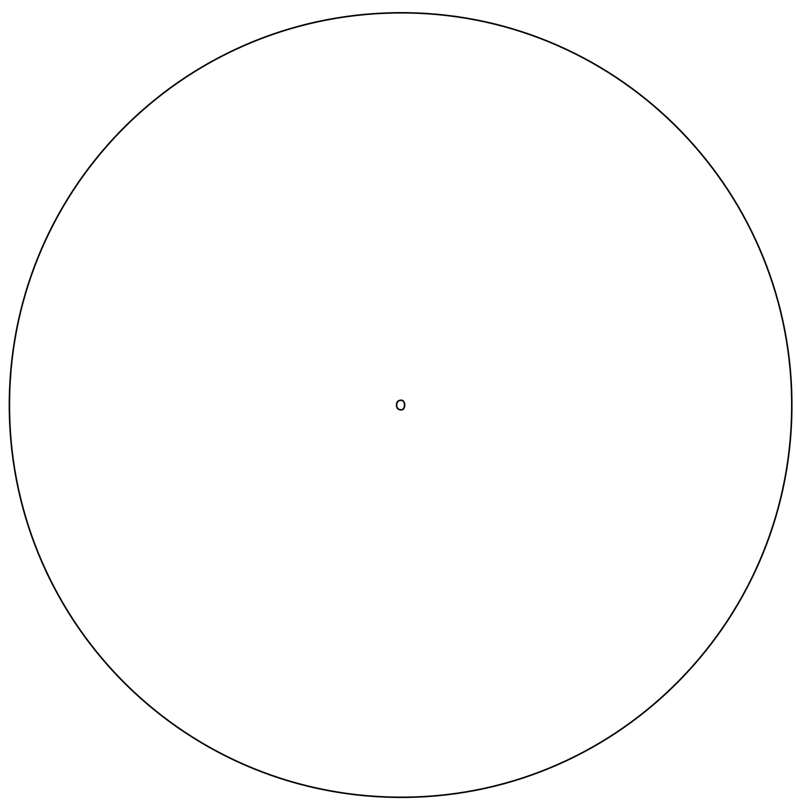
"How I Spend My Time"

1. Look at your log sheet and decide what your five or six major categories are (one category can be Miscellaneous). Then fill in the table **below.**

Summary Table for Log Sheet				
Type of Activity	**Total Hours** (i.e. per week)	**Average Hours per Day**	**% of Day**	**Central Angle**
TOTAL:				

2. Create a circle graph illustrating the amount of time devoted to each of the 5 or 6 categories summarized in the table above.

3. Label each sector of your circle graph with its corresponding activity.



M2T2

Instructor Page

Displaying the Data

Tips:
"A picture is worth a thousand words!"

Using graphs in statistics can be helpful for showing overall trends of data, as well as details about a data set.

In this section, you will see four ways of displaying data: circle graphs, stem and leaf tables, histograms, and box and whisker plots.

Step-by-step guide

⇒ Have participants open a spreadsheet program. Then have them enter their 5 or 6 categories of activities in one row (or one column) and the number of hours they spent on each activity in the next row (or column).

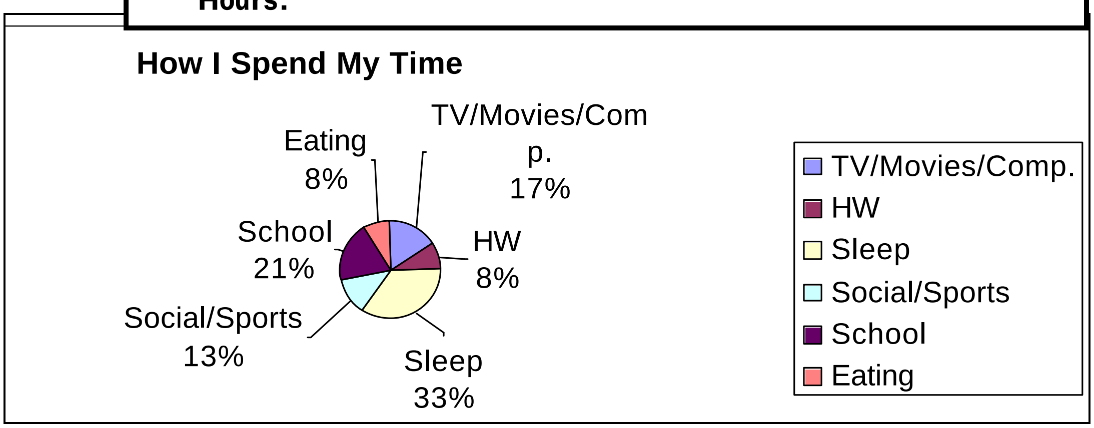
⇒ If participants are working in a Microsoft Excel® spreadsheet, then they may simply follow the instructions on page 21.

⇒ When finished with the computer generated circle graph, participants can brainstorm different ways of creating human circle graphs!

Sample spreadsheet and circle graph created by Excel:

The pie graph below was made with MS Excel® , but it could also be done with Clarisworks® or other spreadsheet programs.

Activity:	**TV/ Movies/**	**HW**	**Sleep**	**Social/ Sports**	**School**	**Eating**
Average Daily Hours:	4	2	8	3	5	2



Participant Page

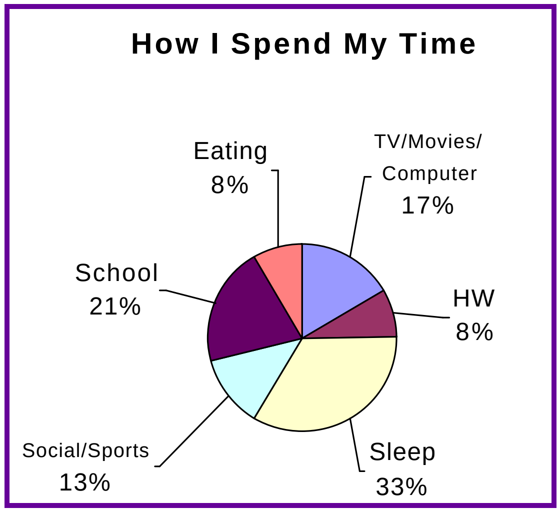
Displaying the Data

Display of How I Spend My Time cont.

You can create a computer generated circle graph and compare it with the one drawn by hand. The following step by step instructions are for an Excel® Spreadsheet:

1. First enter your 5 or 6 categories of activities and your "Average Hours per Day" into two rows in the spreadsheet.

2. Highlight both rows involved (category titles and average daily hours) and select the CHART icon on the Toolbar
   a. Select "Pie" for Chart Type
   b. Press the "Next" button
   c. On the next screen you should see your colored circle graph displayed and your category titles automatically placed into the legend. Press the "Next" button
   d. Type in a title for your chart and change the legend or data labels as desired. *(You might want to select "Show label and percent" , as was done in this sample graph.)*
   e. The last screen will ask if you want to place your chart as an object in the current spreadsheet. Press "Finish".



<u>Another method for producing a circle graph:</u>

Make a "Human Circle Graph" by involving other participants in the class.

One way of doing this is by standing in a circle (representing hours) and using string or yard sticks, etc. to divide into categories.

Instructor Page

Displaying the Data (continued)

In this section, participants read about mean, median, and mode, and look at a stem-and-leaf display of data, a histogram, and a box and whiskers plot.

Tips from a teacher:

"When introducing a stem and leaf table to my class, I asked the students to report the total number of hours they slept in a week according to their log sheets. (One by one, the students just called out their total hours.) As I recorded each student's total on the overhead, I was already displaying it in the form of a stem and leaf table (however, I didn't say anything about what I was doing!). By the time I finished recording all the students' data, most of the students had figured out the format for how I was displaying the data and could then create their own stem and leaf tables."

Later, participants are asked to do the same. But first, here is a detailed explanation of stem and leaf tables, mean, median, mode, range, histograms, and box and whisker plots.

Technology Tip (for the TI 73 Graphing Calculator):

The calculator is not able to display a STEM and LEAF TABLE, but it can help you arrange a list of data in order, so that a stem and leaf table would be easier to create by hand:

1. Entering Data: Press the [LIST] button. If other data already exists in the first list, then erase it by moving the cursor up to highlight L1 and pressing [CLEAR] then [ENTER]. Then type all your data in the list.

2. Sorting the Data in Ascending Order: Press [2nd][LIST] and right arrow over to the OPS menu. Select 1: Sort A(. When you press [ENTER], the calculator will be waiting for you to tell it which list to sort. Type [2nd][LIST] again and [ENTER] to indicate L1. Press [ENTER] again and the calculator will just say "Done", but when you look back at List 1, you will see that the data has been sorted in ascending order.

This could be very helpful when making a stem and leaf table, particularly if you are dealing with a very large data set that is not already in any kind of order.

Internet Resource:

An interesting and amusing introduction to descriptive statistics is available on the Web at <www.mste.uiuc.edu/hill/dstat/dstatintro.html>

M2T2

Participant Page

Displaying the Data (continued)

Stem and Leaf Tables, Histograms, and Box and Whisker Plots

When organizing a set of data, such as a list of test scores, it is helpful to first arrange the data in order. One method for displaying data in order from least to greatest in an efficient manner is a stem and leaf table. To make a stem and leaf table, organize the data from least to greatest and put all the tens digits under the "stem" column. Then place all the units digits under the "leaf" column. If a units digit occurs more than once, it needs to be repeated as a separate leaf every time it occurs. The data displayed in the stem and leaf table below are sample test scores from a class: 37, 40, 41, 41, 42, 43, 44, 51, 51, 51, 52, 55, 63, 65, 65, 65, 65, 67, 71, 73, 77, 83, 86, 86, 90.

The ***median***, or the middle number in a set of data which is arranged in order, is useful for showing how one's grades compare to the grades of the rest of the class. Look at the sample stem-and-leaf table:

Sample Stem-And-Leaf

Stem	Leaf
3	**7**
4	**0,1,1,2,3,4,**
5	**1,1,1,2,5**
6	**3,5,5,5,5,7**
7	**1,3,7**
8	**3,6,6**
9	**0**

Suppose your score was 63, the *median* score. The lowest score was 37 and the highest was 90. There were twelve scores that were lower than 63 and twelve scores that were higher.

What does this tell you about how you are doing in relation to the rest of the class?

The ***mode***, or the most frequent or common score, is useful in yet another way. Look again at our stem and leaf table. We see that out of twenty-five people who took the test, four of them scored 65! Our score was very close to the *modal* or most common score.

Find the **mean** for this set of data The mean is the mathematical average found by adding all the elements in a data set and then dividing by the number of elements in the set. How does our 63 compare to the *mean* score?

Which one of these three measures of "central tendency" (the mean, median, or mode) would you rather use to describe your score to your parents?

Instructor Page

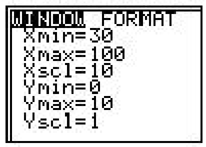
Displaying the Data (continued)

The main idea in the discussion on range, is that the participants should look beyond just the range in interpreting the data.

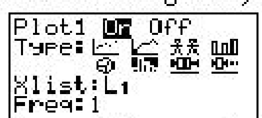
Producing a HISTOGRAM on the TI 73 Graphing Calculator:

1. Enter your data into List 1 (as explained on page 22).

2. Setting the Window: Since a histogram describes how many elements in a list fall within a certain interval, you need to decide how many intervals you want and how wide each interval will be. With the sample test scores, the data ranged from 37 to 90. Therefore, good intervals to use might be 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100. For the calculator to display these seven bars (each ten units wide), you will need to press [WINDOW] and set it up as follows:

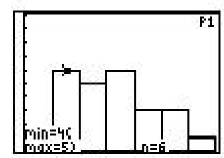


3. Setting up the Plot: Press [2nd][Y=] to get to the Stat Plot menu. Press [ENTER] to select Plot 1 and then press [ENTER] to highlight "ON". Continue to arrow down and over to select the histogram icon, as the type of plot and type L1 for the "Xlist" (the list from which the x values are coming from). Type 1 for the frequency. The screen should look similar to this:

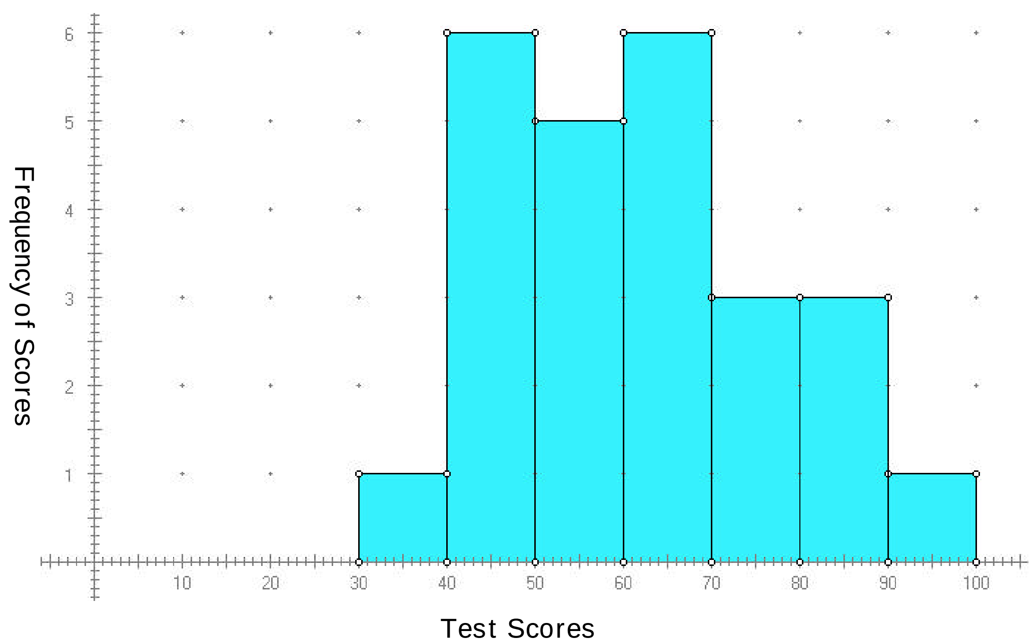


4. Viewing the Graph: Now just press the [GRAPH] button and your histogram should appear like this one:

When you press [TRACE], you can arrow over to each bar of the histogram. The calculator will display a "min" and "max" value, which are the x values representing the left and right edge of each bar. It also displays an "n" value which tells you how many elements (scores) are contained within that bar.



Frequency of Scores Histogram



In the sample histogram at right, how many scores were in the thirties?
1

How many scores were in the sixties, where our score of 63 resides?
6

M2T2

Participant Page

Terms to learn:

- **Stem and Leaf Tables**
- **Mean**
- **Median**
- **Mode**
- **Range**
- **Histogram**
- **Box and Whiskers Plot**

Displaying the Data (continued)

The ***range*** of the data, or the difference between the highest score and the lowest score, is another measure which can easily be obtained from a stem-and-leaf table. Here, the lowest score is 37, while the highest score is 90. The *range* is then 53 (90-37=53) but what does that really tell us? Let's look at another example: Suppose that in the weeks leading up to our big test, we took a series of quizzes. Here are two students' scores:

Zoe scored 5,5,5, and 10 on her four quizzes.
Tareka scored 5,9,9 and 10.

The *range* of data for each is 5, but does that tell us how each student did?

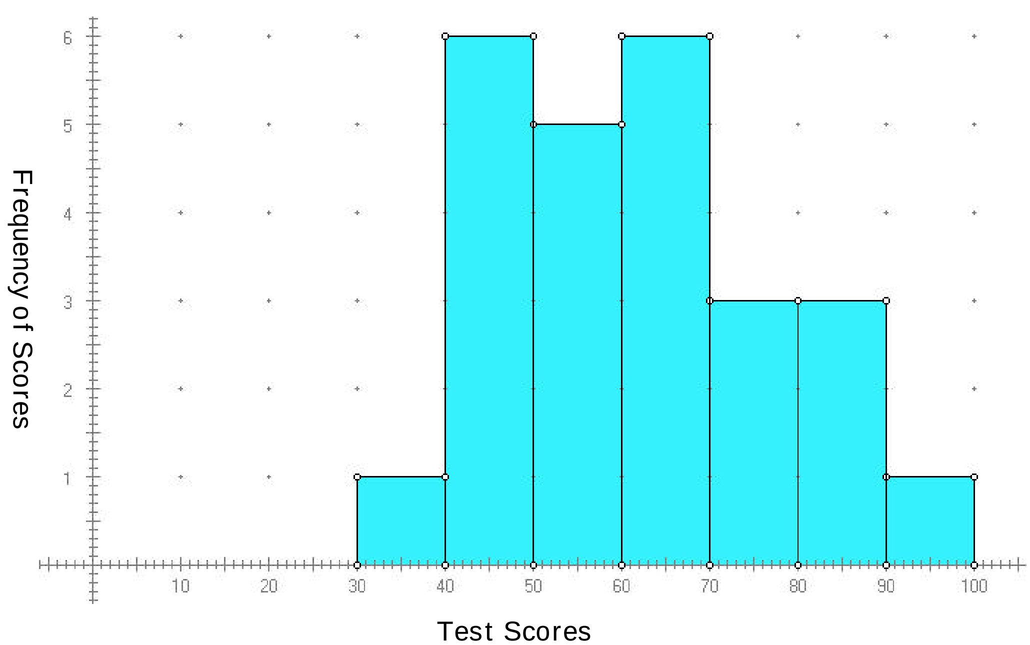
Notice that Zoe scored much lower overall but still has the same *range* as Tareka.

Now, look again at our *stem and leaf table*. Another way of displaying that data is to use a ***histogram***. A histogram is a picture that tells us how pieces of data are distributed over a given set of equally-spaced intervals. It reports how many elements are within each interval. Below, we have a *histogram* that displays the scores on the *x -axis* and the *frequency* of those scores on the *y-axis.*

In the sample histogram at right, how many scores were in the thirties?

How many scores were in the sixties, where our score of 63 resides?

Frequency of Scores Histogram



Analyzing the Data (continued)

Instructor Page

A box and whiskers plot is good for showing the CENTER and the SPREAD for a set of data. This plot also groups the data into fourths (or QUARTILES).

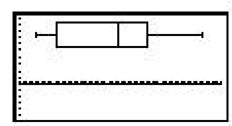
After creating some sample box plots on paper using student data from their time sheets, the teacher can challenge students to create a "Human Box Plot". Students will stand up and organize themselves according to something such as their height or their shoe size.

Producing a BOX and WHISKERS PLOT on the TI 73 Graphing Calculator:

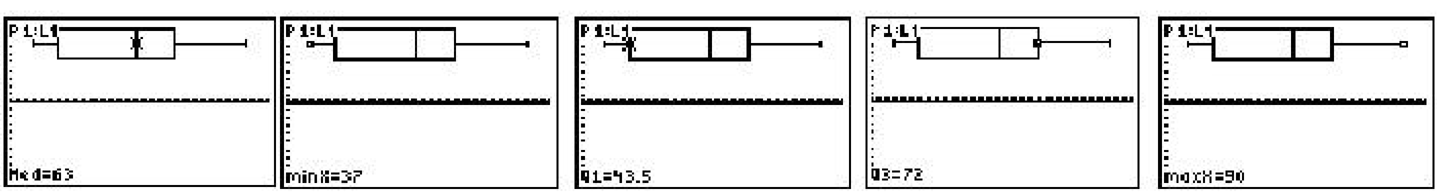
1. Enter your data into List 1 (as explained on page 22).

2. Setting up the Plot: Press [2nd][Y=] to get to the Stat Plot menu. Press [ENTER] to select Plot 1 and then press [ENTER] to highlight "ON". Continue to arrow down and over to select the box plot icon, as the type of plot. Type L1 for the "Xlist" (the list from which the x values are coming from) and then type 1 for the frequency. The screen should look similar to the following:

3. Viewing the Graph: To have the calculator automatically create an appropriate window for your stat plot, press [ZOOM] and select 7: ZoomStat. Press [ENTER] and the box and whiskers plot on your screen (for the test score sample data) should look like the following:



4. Using "TRACE": When you press [TRACE], the cursor will be blinking on the median value. When you press the left and right arrows, the cursor will move to the other four important values on a box plot the minimum, the 1st Quartile, the 3rd Quartile, and the maximum.



Analyzing the Data (continued)

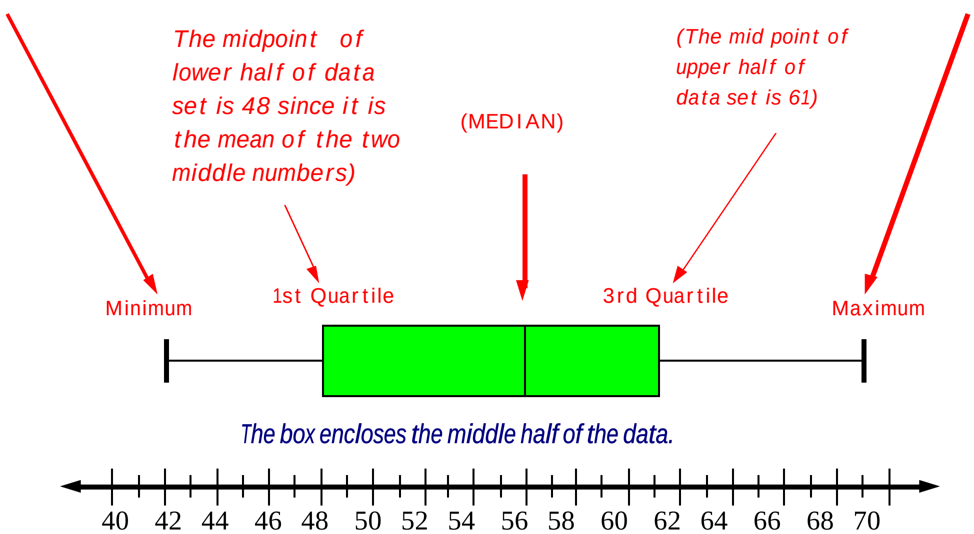
A third method for displaying one variable data is a BOX and WHISKERS PLOT.

To make this plot for a set of data, calculate the median and the 1st and 3rd quartiles (midpoints of the top and bottom halves of the data). On a horizontal scale, draw a box which encloses the middle half of the data. Draw a vertical line at the point in the box which corresponds to the median. Draw the "whiskers", which are the horizontal lines that extend from the ends of the box to the least and the greatest numbers in the data set (these show the range of the data).

Sam-

Box and Whiskers Plot for Total Hours of Sleep in a Week

42, 45, 46, 47, 49, 50, 52, 55, 56, 56, 58, 59, 60, 62, 63, 65, 69



M2T2

Instructor Page

Gathering and Organizing Data from Entire Class

<u>Step-by-step guide</u>

⇒ Have the participants look at their log sheets and summary tables and agree upon one activity which they all had in common. As they go around the room and take turns reporting their TOTAL number of hours they spent doing this common activity in a week, the instructor can record all the data in the form of a **stem and leaf table**.

⇒ Ask participants where the numbers in the "stem" are coming from and what the "leaves" mean.

⇒ Then, use the stem and leaf table to calculate the **mean, median, mode, and range** of the data.

⇒ Next, create a **histogram** with the same data. Finally, create a **box and whiskers plot** with the same data.

Participant Page

Gathering and Organizing Data from Entire Class

a. Find an activity on the log sheet summary that all of the participants have in common. As the other participants put their data on the board, make a **stem and leaf table** in the space below.

b. Also create a **histogram** using the same data recorded in the stem and leaf table. Use the space below to make the histogram. Notice that you can compare the histogram to a stem and leaf "turned sideways".

c. Now try a third way of displaying data. Use a **box and whiskers plot.** Clearly label the minimum, 1st quartile, median, 3rd quartile, and maximum

M2T2

Instructor Page

Designing Surveys For Gathering and Analyzing Data

Tips:

Ask participants to brainstorm various types of surveys they have been exposed to in their own lives. What questions were these surveys attempting to answer?

Connect the survey task to students' lives.

Who uses surveys? One example is the U.S. Census Bureau. What kinds of questions are asked on the Census? Why do we have a census at all?

Step-by-step guide

⇒ **Step 1: Identify a question to be explored through the use of a survey.** Question participants about what surveys are and how they are used. What are some questions that we might seek to answer using a survey?

⇒ **Step 2: Design a survey that helps answer your question.** What kinds of questions will be helpful in compiling data that will lend itself to answering the student or group of students' questions? This might be a good time to talk about targeting questions.

⇒ **Step 3: Identify groups of persons to whom the survey will be given and collect the data.** Why does it matter what group of persons will be surveyed? How will the group surveyed affect the outcome of the survey? If I want to find out about the relationship between time spent playing video games and time spent on homework, would I survey a group of senior citizens?

⇒ **Step 4: Do a two variable (bivariate) analysis on the data.** Look at the sample survey in Appendix F and pick pairs of variables that might be related. Construct a scatter plot for the two variables by hand and then by using a spreadsheet.

(See Appendix A for instructions on how to use a spreadsheet to display statistical information.)

Internet Resource:
www.census.gov

Participant Page

Designing Surveys For Gathering and Analyzing Data

Terms to learn:
- **Analysis**
- **Bivariate**
- **Scatter plot**
- **Survey**

Step 1: Identify a question to be explored through the use of a survey.

This should involve the possible relationship between two variables, such as "Do students who spend more time watching TV spend less time doing homework than others?" or "Is it true that the more time students spend on a part-time job, the less they sleep?"

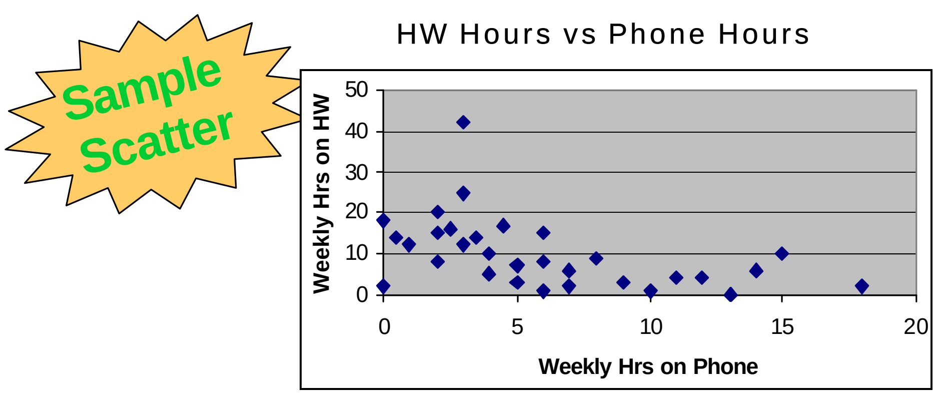
Step 2: Design a survey that helps answer your question.

The class can compile a list of all the survey items needed in order to help answer each student's (or each group of students) question. A sample survey is available in Appendix F .

Step 3: Identify groups of persons to whom the survey will be given and collect the data.

Step 4: Do a two variable (bivariate) analysis on the data.

Do a scatter plot for the two variables related to your original question. *Example: Relationship between number of hours on the phone (X axis – independent variable) and number of hours spent doing homework (Y axis – dependent variable).*
Then determine (by informal exploration) if there appears to be a positive or negative relationship between the two variables as was first predicted.



Step 5: Prepare a report on your data analysis activity.

Be sure to describe the data that were collected and interpret relationships that were found. The presentation of your report should include at least one data table, a graph and a final conclusion based on the data.

This page intentionally blank

M2T2

Participant Page

Using a Spreadsheet to Display Statistical Information

A correlation coefficient is a number between –1 and 1 that describes the strength of a relationship. A zero correlation means no relationship.

These instructions were written for an Excel spreadsheet. However, similar operations can be done with other spreadsheet programs.

A. To calculate the MEAN for a column of data in a spreadsheet:

1st – Place the cursor in the cell where you want the mean to be displayed.

2nd – Go to the Insert menu and down to "Function…" Under Function Category choose "Statistical"

3rd – To calculate the mean, choose "Average" for the Function Name Then press the OK button

4th – Another display box will ask for the cells to be averaged. Type in the first cell of the column from which the data came. Then type a colon and the last cell of the data column. When you press OK, the final command should look like:
=AVERAGE(B2:B32)

B. To calculate the MEDIAN:

Repeat Steps 1 and 2 above.
Then, for the Function Name choose "Median" and press OK.
Type in the cells to be used just as in Step 4 above.
The final command will look like: =MEDIAN(B2:B32)

C. To calculate the MODE:

Repeat Steps 1 and 2 above.
Then, for the Function Name choose "Mode" and press OK.
Type in the cells to be used just as in Step 4 above.
The final command will look like: =MODE(B2:B32)

D. To calculate the CORRELATION COEFFICIENT:

Repeat Steps 1 and 2 from Part A.
Then for the Function Name, choose "CORREL" and press OK
The next display box will ask for the two arrays of cells between which the correlation coefficient will be calculated. Type in the first cell of the first data column, then a colon, and then the last cell for Array 1. Then type in the first and last cells of the second data column for Array 2.

This page intentionally blank

Participant Page

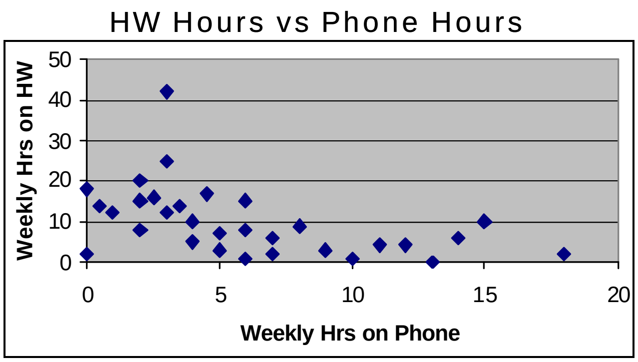
Using a Spreadsheet to Display Statistical Information (continued)

The display box should look like: Array 1: B2:B32
Array 2: C2:C32

When you press OK, the answer should appear in the cell and the final command looked like: =CORREL(B2:B32,C2:C32)

When the correlation coefficient was calculated in the above manner for this set of data and for the scatter plot shown below, the answer was:

CORREL = -0.47715

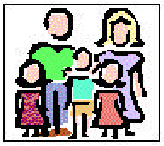


Check out the Web page “Guessing Correlations” at
www.stat.uiuc.edu/~stat100/java/GCApplet/GCAppletFrame.html

	Hrs on phone	Hrs on HW
	0	2
	1	12
	4	5
	2	15
	2.5	16
	6	15
	3	42
	4	10
	2	8
	6	1
	10	1
	15	10
	2	20
	13	0
	0	18
	9	3
	18	2
	7	2
	3	25
	4.5	17
	5	7
	6	8
	12	4
	5	3
	3	12
	0.5	14
	8	9
	7	6
	3.5	14
	11	4
	14	6
Mean	6.03	10.038
Median	5	8
Mode	2	2

This page intentionally blank

APPENDIX B



Data sheet for three child family

	Child 1	Child 2	Child 3	Success?
Trial #1				
Trial #2				
Trial #3				
Trial #4				
Trial #5				
Trial #6				
Trial #7				
Trial #8				
Trial #9				
Trial #10				

This page intentionally blank

Birth Month Problem

If you choose 4 people at random, what is the probability that at least two were born in the same month?
(not necessarily the same year)

OPTIONAL BIRTH MONTH MODEL

How can we model this situation using a random number program on the TI-73, TI 82 or the random number generator on the TI-83?

What will a trial consist of based on the model?

How will you know if the trial is a success?

How many trials should you do?

- Make a guess!!! _______
- What model should we use? ___________________
- We will run the experiment in class. What will be one trial?
- How will we know if it is successful?
- Keep track of the data and then compute the probability.

TRIAL	RESULTS	SUCCESSFUL?
1		
2		
3		
4		
5		
6		
7		

Probability (Number of successes divided by the number of trials)

= ________

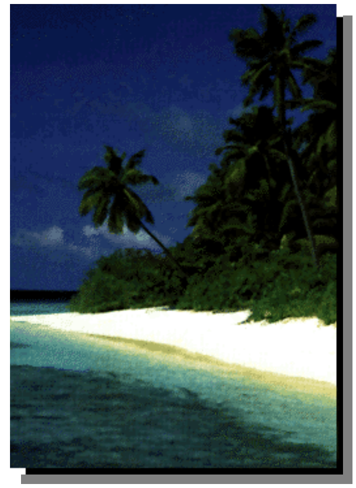
- Do you want to change your original guess? Why?

- Now calculate the average probability of success for the entire class?

This page intentionally blank

Hermit's epidemic:

Six hermits live on an otherwise deserted island. An infectious disease strikes the island. The disease has a 1-day infectious period and after that the person is immune (cannot get the disease again).



Assume one of the hermits gets the disease. He randomly visits one of the other hermits during his infectious period. If the visited hermit has not had the disease, he gets it and is infectious the following day. The visited hermit then visits another hermit.

The disease is transmitted until an infectious hermit visits an immune hermit and the disease dies out.

There is one hermit visit per day.

Assuming this pattern of visits continues, how many hermits, on average, can be expected to get the disease? *(From Travers, K. J. et. al. Using Statistics, Addison Wesley, 1986, page 67).*

Internet Resource:

For an interactive simulation of this problem (and a discussion of its theoretical solution) go to the following web site:

< http://www.mste.uiuc.edu/users/hill/ev/hermitprob.html>

This page intentionally blank

APPENDIX E

"How I Spend My Time" Log Sheet

Time of Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
12 Midnite							
1 am							
2 am							
3 am							
4 am							
5 am							
6 am							
7 am							
8 am							
9 am							
10 am							
11 am							
12 noon							
1 pm							
2 pm							
3 pm							
4 pm							
5 pm							
6 pm							
7 pm							
8 pm							
9 pm							
10 pm							
11 pm							

This page intentionally blank

Sample Class Survey

Please answer as honestly and accurately as possible:

1. In what month were you born? _______________
2. How many brothers do you have? ______
3. How many sisters do you have? ______
4. On average, about how many HOURS A WEEK do you...
   a. Sleep _____
   b. Watch TV _____
   c. Talk on the phone _____
   d. Spend time doing homework _____
   e. Work at a part-time job _____
   f. Spend time on the computer or electronic games ____
   g. Play basketball _____
   h. Play other sports _____
   i. Eat _____
5. How many pets do you have? _____
6. How many books do you read a MONTH on average? _____
7. What is your favorite school subject? __________
8. What is your grade in school (6, 7, or 8)? ______

This page intentionally blank

APPENDIX G

Glossary of Terms

Box and Whiskers Plot: A box and whiskers plot is good for showing *center* and the *spread* for a set of data. This plot also groups the data into fourths (or quartiles).

Histogram: A bar graph that shows the frequency of the occurrence of certain data.

Mean: The sum of a set of numbers divided by the number of numbers in the set. The mean is often called the "average".

Median: The middle number in a series of numbers where they are arranged in order (such as least to greatest or greatest to least).

Mode: The number in a set of data that occurs most frequently.

Independence: When one event or variable does not affect the probability of occurrence of another event or variable.

Range: The difference between the largest number in a set of data and the smallest number in that set.

Scatter Plot: A graph that displays quantitative data as points in two or more dimensions. Scatter plots are used to visualize relationships between variables.

Stem and Leaf Plot: A way to display data where certain digits are used as "stems" and others as "leaves".

Theoretical Probability: The number obtained by dividing the number of successful outcomes by the total number of outcomes.

Variable: A quantity that varies such as the number of hours of television watched per week.

(From Travers, K. J. et. al. Using Statistics, Addison Wesley, 1986, page 449-451).

Statistics

Send questions on these modules to m2t2@mail.mste.uiuc.edu