PAPER #802

Sowing the Seeds

John Gabrosek, gabrosej@gvsu.edu
Grand Valley State University

Mary Richardson, richamar@gvsu.edu
Grand Valley State University

Key Words: Active learning; Experimental design; Matched pairs; Graphical data summaries; Numerical data summaries

Abstract

In this paper we describe an interactive activity that illustrates experimental design and graphical and numerical summaries of matched pairs data. Students generate, collect, and analyze data using techniques taught in an introductory applied statistics course. No calculus skills or prior statistical knowledge is necessary.

1. Introduction

In this paper, we discuss an activity that illustrates experimental design and graphical and numerical summaries of paired data. The experiment can be designed and the data collected in a one-hour class period. With the help of a statistical computing package or the TI-83 calculator, students can complete the graphical and numerical summaries on their own or in two additional one-hour class periods.

We use this activity in an introductory statistical concepts course at the undergraduate level. Our goal is to involve students in the process of designing an experiment to answer a research question, conducting the experiment, and analyzing the data.

2. The Activity

2.1 Background

Not since the quest for fire has a question burned so brightly in the imagination of peoples the world over. You ask, what question is that? But, of course, the question is: "Can a sunflower seed be projected further by a person flicking the seed with his or her finger or by spitting the seed?"

As important as this question is, there has been surprisingly little research done on sunflower seed projection methods. A Google™ search on Sunflower Seed Projection brought up many a website on the edible oil industry but nary a one that supplied an answer to this pressing question. (Get it – edible oil industry and pressing question!) Not to say that many a person hasn't spent time enjoying the simple pleasures of spitting sunflower seeds. Just watch a major league baseball game and see the childlike joy shelling and spitting sunflower seeds brings to the face of even the most jaded player.

As DAVID® says on the front of a bag of Roasted & Salted Sunflower Seeds, "Eat. Spit. Be Happy." Really, what else is there?

2.2 Materials

Students work in groups of three or four. Each group needs a Data Collection Sheet (see Appendix A) and a 50-foot tape measure. (Some students can really spit. The all-time record over eight semesters and ≈ 300 students is 582″.) Each student needs two sunflower seeds in their shells and a copy of the Activity Worksheet (see Appendix B).

2.3 Procedure

We use this activity during the first two weeks of an introductory applied statistics course taken by majors from throughout the university community.

Prior to completing Part 1: Designing the Experiment, students are familiarized with the basic terms of statistical experimental design including response, factor, treatments, control, randomization, replication, and blinding. Students are also given a brief lecture on the difference between matched pairs designs and completely randomized designs. In Part 2: Collecting the Data, students carry out a matched pairs design determined by the class.

Prior to completing Part 3: Summarizing the Data Graphically, students are shown grouped frequency tables and histograms.

Prior to completing Part 4: Summarizing the Data Numerically, students are given a brief lecture on the mean, standard deviation, and the five-number summary. The emphasis of the lecture is on understanding what these measures are and not on their calculation by hand. (This is especially true of the standard deviation.)

2.3.1. Part 1: Designing the Experiment

The research question is: "Can a sunflower seed be projected further by a person flicking the seed with his or her finger or by spitting the seed?" Initially, students design an experiment to answer the research question under two restrictions. The first restriction is that the 30 students in the class are the subjects in the experiment. The second restriction is that a maximum of two sunflower seeds are allowed per student.

Each group develops both a matched pairs design and a completely randomized design. Students answer the following questions to help them develop their designs.

- What is the response variable? What are its units of measurement?
- What is the factor variable? What are its levels?
- Give a step-by-step approach to a **completely randomized design** to determine whether sunflower seeds can be spit or flicked further.
- Discuss how your group's completely randomized design accounts for the four principles of designing an experiment (Control, Randomize, Replicate, Blind).
- Give a step-by-step approach to a **matched pairs design** to determine whether sunflower seeds can be spit or flicked further.
- Discuss how your group's matched pairs design accounts for the four principles of designing an experiment (Control, Randomize, Replicate, Blind).

Students have little trouble identifying the distance traveled by the seed as the response variable and the method of projecting the seed (spit or flick) as the factor variable.

After completing their group designs, the class decides upon a matched pairs design. The experiment is conducted according to the class design. The typical class design is to have each student flick one sunflower seed and spit one sunflower seed. One point to stress with students is that the order of the factor levels (spit or flick) should be randomly assigned for each student. Most groups want to have everyone spit first and flick second or vice versa. A second point of importance is to control for the effect of wind. Having each group face the same direction mitigates any wind effect.

2.3.2. Part 2: Collecting the Data

Each student is assigned to a group and group members' names are written on a Data Table (see the Activity Worksheet, Appendix B, Part 2). For each group member a random number is generated. If a '2' is generated the person will spit first and then flick. If a '1' is generated the person will flick first and then spit. The First Method and Second Method columns in the Data Table are completed and then we head outside with the tape measures, sunflower seeds, and a cheery disposition.

Students are warned not to spit on any innocent bystanders. Each group member is asked to play each of the following roles once:

- **Spitter/Flicker:** The Spitter/Flicker stands at the start line and projects a sunflower seed by either spitting or flicking. Then, the same person projects a second seed using whatever method was not used first. Next, the Spitter/Flicker becomes the Holder.

- **Holder:** The Holder stands at the start line and holds one end of the tape measure. Next the Holder becomes the Walker.

- **Walker:** The Walker stands at the start line and walks with one end of the tape measure to where the sunflower seed came to rest as determined by the Eyes. Next the Walker becomes the Eyes.

- **Eyes:** The Eyes stands about 15-20 feet from the start line and determines where the sunflower seed comes to rest. It's not that easy to see a sunflower seed fly through the air so students should be told to pay close attention. Next the Eyes becomes the Spitter/Flicker.

Pictures 1, 2, and 3 show students engaged in some spirited spitting and flicking of sunflower seeds.

Picture 1. Spitting.

Picture 2. Flicking.

Picture 3. Measuring.

Example data from a class is given in Table 1. Given the Spit and Flick Distances, students calculate the Differences as Spit minus Flick. Students should be reminded that the Difference can be a negative number.

Table 1. Example class data.

Spit Distances	Flick Distances	Differences[a]	Spit Distances	Flick Distances	Differences
66	176	-110	157	167	-10
94	168	-74	94	186	-92
95	183	-88	138	213	-75
155	178	-23	226	99	127
123	170	-47	113	107	6
94	100	-6	165	242	-77
68	130	-62	157	342	-185
124	141	-17	180	267	-87
265	212	53	154	130	24
229	343	-114	182	265	-83
403	339	64	163	140	23
198	137	61	45	372	-327
75	96	-21	132	105	27
176	160	16	243	54	189
165	44	121			

[a]Differences are calculated as: spit minus flick.

2.3.3. Part 3: Graphically Analyzing the Data

Students begin by graphically exploring the data for The Sunflower Seed Experiment. Students need to be reminded that the overriding goal is to answer the research question: "Can a sunflower seed be projected further by a person flicking the seed with his or her finger or by spitting the seed?"

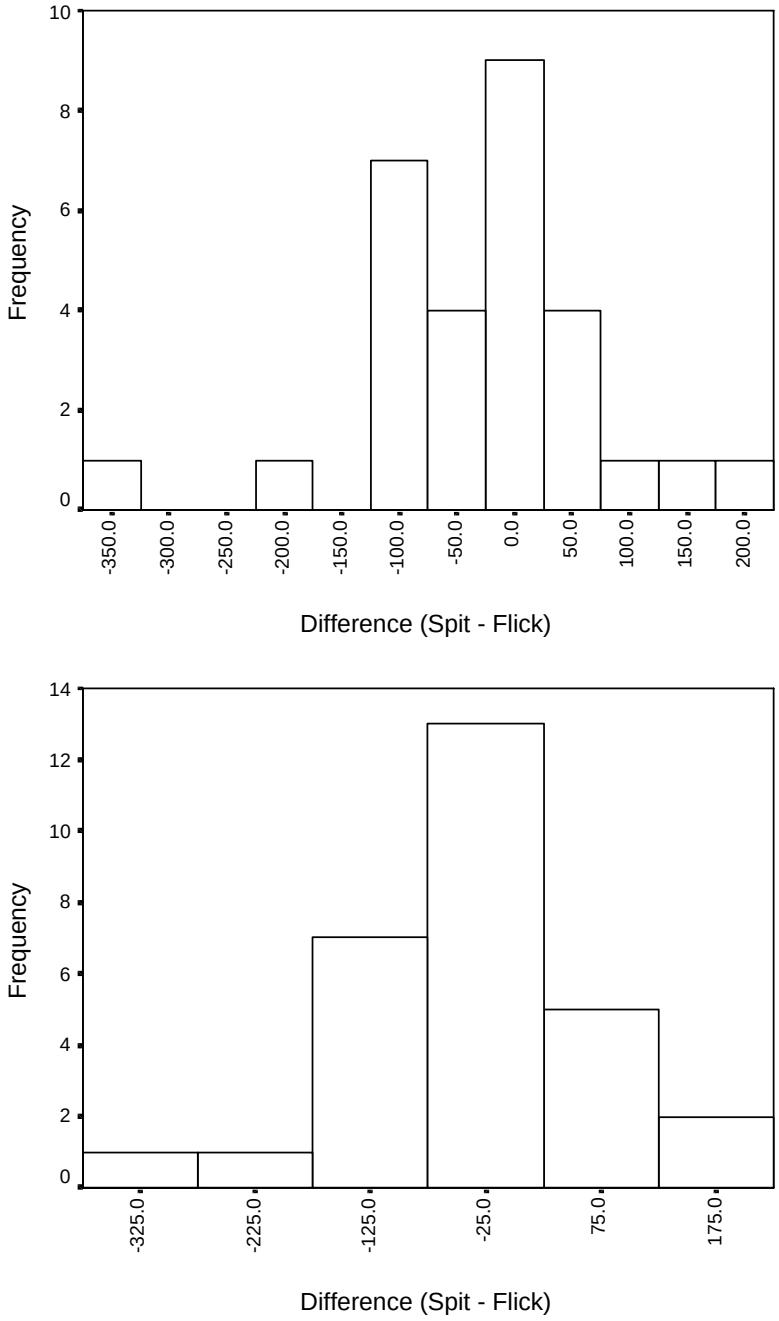
We ask students to complete the Data Collection Sheet by copying the data off of a transparency and finding the Differences as Spit Distance minus Flick Distance. We stress that when data are collected using a matched pairs experiment it is the differences in values between the two treatments that are important. The whole point of doing a matched pairs experiment is to account for the variability from person to person. If each person gets both treatments (here, spits and flicks a sunflower seed), then any special effects for the person will "average out." Thus, the data analysis is done on the Differences and not on the two separate sets of data.

First, students are asked what value the Differences are expected to be close to if neither method is truly better for projecting a sunflower seed.

Students are then asked to construct histograms of the Differences using twelve classes and six classes, respectively. Based on the histograms, students discuss center, spread, shape, and

outliers for the Difference distribution. Students write about the effect that increasing the number of classes has on the appearance of the histogram. In Figure 1, twelve-class and six-class histograms are shown for the example class Differences.

Figure 1. Histograms of example class differences.



Next, students make a stemplot of the Differences. Figure 2 displays a stemplot for the example class Differences.

Figure 2. Stemplot of example class differences.

```
Difference (Spit - Flick) Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     1.00 Extremes    (=<-327)
     1.00       -1 .  8
     2.00       -1 .  11
     8.00       -0 .  67778889
     6.00       -0 .  011224
     5.00        0 .  01222
     3.00        0 .  566
     2.00        1 .  22
     1.00 Extremes    (>=189)

 Stem width:    100.00
 Each leaf:       1 case(s)
```

Students answer the research question based on what they have done in Part 3. Many students have a very shaky understanding of the important role that variability plays in statistics. Emphasizing the spread of the data, as illustrated by the graphs, helps students to see that it takes more than the mean of each projection method to answer the research question. The students need to account for variability when determining whether a real difference exists in the distance a sunflower seed travels by spitting compared to flicking.

2.3.4. Part 4: Numerically Analyzing the Data

After completing Part 3, students see the value that graphs have by providing us a picture of the distribution of the data. However, graphical summaries aren't enough. We also need numerical summaries of the data that locate the center of the data set and describe the spread (or variability) in the data.

Students enter the Differences from the Data Collection Sheet into a list in their calculators. Then, they use their calculators to find numerical summaries for the list of Differences. Specifically, students find the five-number summary (minimum, quartile 1, median, quartile 3, maximum), the sample mean, and the sample standard deviation. In Figure 3, we provide numerical summaries for the example class data.

Figure 3. Five-number summary, mean, and standard deviation of example class differences.

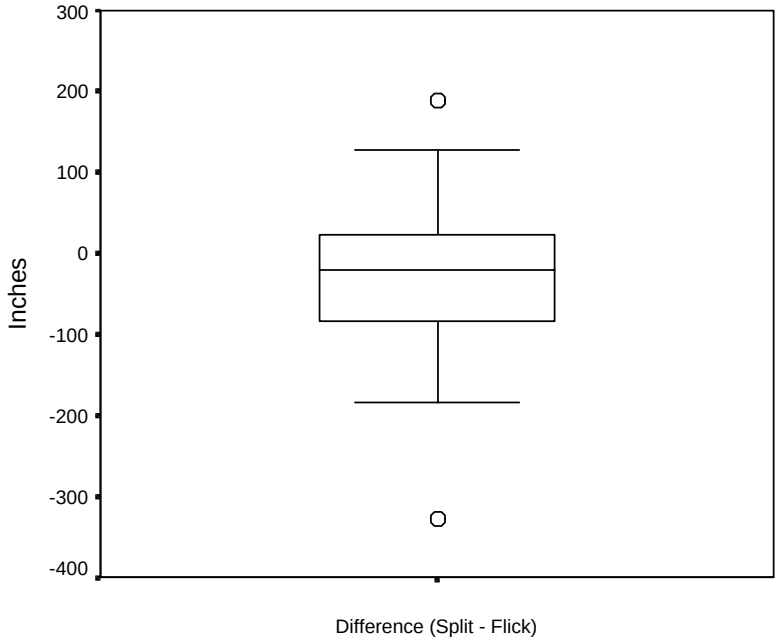
```
mean = -27.14
standard deviation = 99.87
minimum = -327.00
first quartile = -83.00
median = -21.00
third quartile = 24.00
maximum = 189.00
```

Students are asked what they would expect the mean and median to be close to if there is no real difference between Spitting and Flicking. Most of the students correctly respond with "0."

The more difficult question asked is: "Consider the mean and the standard deviation. Together, what do these two numbers tell you about whether or not there is a difference between spitting and flicking?" The goal is to help students to understand that since the standard deviation (99.87) is large relative to the absolute value of the mean (27.14), the mean is not really very far from 0, regardless of its actual value. Thus, there is little evidence that one sunflower seed projection method performs markedly better than the other. This serves as a bridge to sampling distributions and interpretation of the test statistic in a hypothesis test.

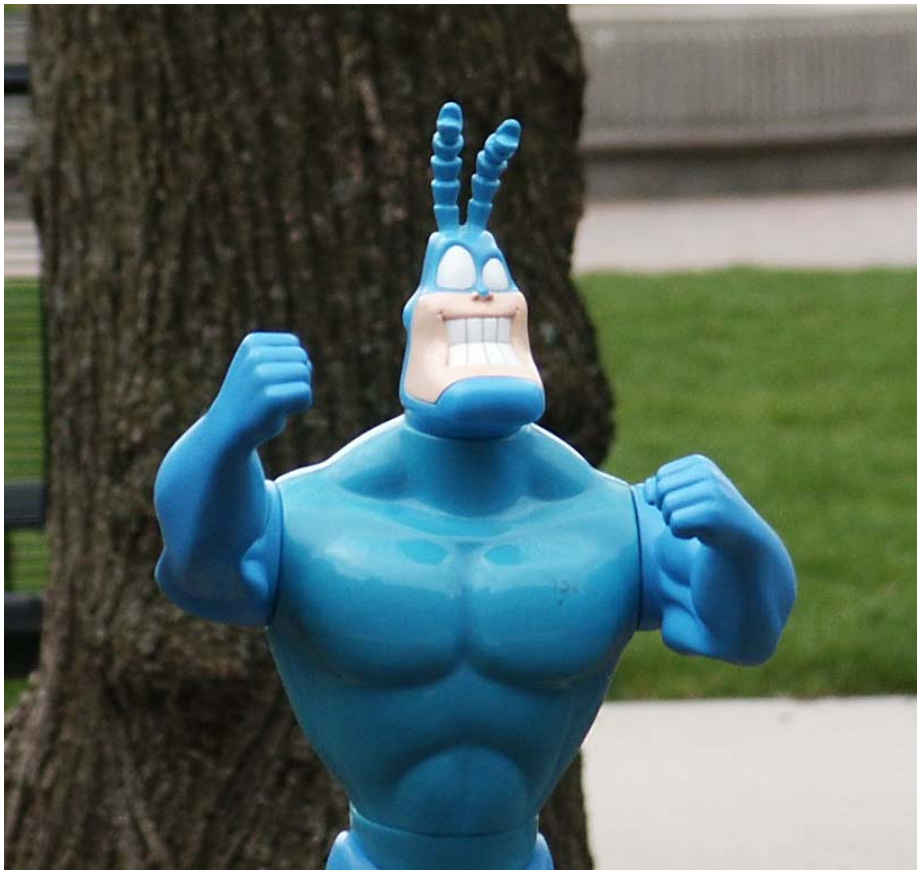
Now that students have calculated the five-number summary, they can make a simple boxplot (no outliers displayed) or a modified boxplot (outliers displayed). Figure 4 shows a modified boxplot constructed from the example class data. Students describe where the boxplot should be centered if no real difference exists between the sunflower seed projection methods. The center can be compared to the width of the box and students quickly see that since 0 is near the middle of the box this suggests no real difference between the methods.

Figure 4. Boxplot of example class differences.



The next two questions in Part 4 deal with the effect of an outlier (an atypically large or small data value) on numerical summaries. Students imagine that The Tick is a member of their class. (For those of you unfamiliar with The Tick, he is a big blue superhero with really good lung capacity. Unfortunately, he's rather clumsy and not especially bright.)

Picture 4. The tick.



The Tick spits the sunflower seed an incredible 1000″, but drops the seed off his finger so that his flicking distance is only 4″. Students investigate the effect that including The Tick's difference (996) in the dataset has on the five-number summary, mean, and standard deviation. Figure 5 shows numerical summaries for the example class data, including The Tick's Difference.

Figure 5. Five-number summary, mean, and standard deviation of example class differences with positive outlier.

```
mean = 6.97
standard deviation = 211.01
minimum = -327.00
first quartile = -83.00
median = -19.00
third quartile = 27.00
maximum = 996.00
```

Actually The Tick isn't too bright. He faced the wrong way when spitting the sunflower seed. His sidekick Arthur (a man in a moth suit) convinces you that The Tick's spit value should be –1000″ not 1000″. Students investigate the effect that including The Tick's difference (-1004) in the dataset has on the five-number summary, mean, and standard deviation. Figure 6 shows numerical summaries for the example class data, including The Tick's Difference.

Figure 6. Five-number summary, mean, and standard deviation of example class differences with negative outlier.

```
mean = -59.70
standard deviation = 203.56
minimum = -1004.00
first quartile = -87.00
median = -22.00
third quartile = 24.00
maximum = 189.00
```

Students are not surprised to find that the mean is pulled toward the outlier while the median and the quartiles are left relatively unaffected. Many students are surprised that the standard deviation increases drastically regardless of whether the outlier is atypically large or small. This provides a nice opportunity for the teacher to review the formula for the variance and the effect of squaring on the standard deviation. Lastly, students speculate on the effect that data skewed to the right and data skewed to the left have on the mean and median.

3. Conclusions

This activity provides a uniquely fun and interesting way to illustrate the research process. The focus is on answering a research question. Students design a simple experiment, conduct the experiment, and analyze the data both graphically and numerically. The research question and experimental protocol are simple enough to use the project early in an applied statistics course that has no statistics prerequisites.

Students enjoy spitting and flicking the sunflower seeds. (I mean who wouldn't?) The activity is memorable enough so that the teacher can refer to it by name to introduce new topics throughout the course.

The Sunflower Seed experiment can be extended to include a discussion of statistical inference for matched pairs data. The data illustrate the importance of using a matched pairs design when the variability from subject to subject is large. If the instructor teaches two sections of the same course, then she could have one class collect the data using a matched pairs design and the other class using a completely randomized design. This would provide students with an example rich enough to compare the two methods.

Appendix A Data Collection Sheet

Complete the Data Table below by copying the data off of the transparency and finding the Differences as Spit Distance – Flick Distance.

Student	Spit Distance	Flick Distance	Difference (Spit minus Flick)
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			

Appendix B Activity Worksheet - Sowing the Seeds

Purpose: The purpose of "Sowing the Seeds" is to design an experiment to collect data on the distance that a sunflower seed can be projected by spitting or flicking and then use graphical and numerical techniques to summarize the quantitative data.

Introduction: Not since the quest for fire has a question burned so brightly in the imagination of peoples the world over. What question is that, you ask? But, of course, the question is: "Can a sunflower seed be projected further by a person flicking the seed with his or her finger or by spitting the seed?"

As important as this question is, there has been surprisingly little research done on sunflower seed projection methods. A Google™ search on Sunflower Seed Projection brought up many a website on the edible oil industry but nary a one that supplied an answer to this pressing question. (Get it – edible oil industry and pressing question!) Not to say that many a person hasn't spent time enjoying the simple pleasures of spitting sunflower seeds. Just watch a major league baseball game and see the childlike joy shelling and spitting sunflower seeds brings to the face of even the most cynical player.

Let us begin our quest for an answer to which method performs better – spitting or flicking.

Part 1: Designing the Experiment

Your goal is to design an experiment to determine whether sunflower seeds can be spit or flicked further. You have the 30 students in our class to use in your experiment as well as a maximum of 2 sunflower seeds per student. Initially, you will work as a group to develop two group designs. Then, we will decide upon a class design for the experiment. Work through questions 1-6 as a group to develop your two designs.

Group Designs

1. What is the response variable? What are its units of measurement?
2. What is the factor variable? What are its levels?
3. Give a step-by-step approach to a **Completely Randomized Design** to determine whether sunflower seeds can be spit or flicked further.
4. Discuss how your group's **Completely Randomized Design** accounts for the four principles of designing an experiment (Control, Randomize, Replicate, Blind).
5. Give a step-by-step approach to a **Matched Pairs Design** to determine whether sunflower seeds can be spit or flicked further.
6. Discuss how your group's **Matched Pairs Design** accounts for the four principles of designing an experiment (Control, Randomize, Replicate, Blind).

Class Design - We will perform the experiment using a Matched Pairs Design that we come up with as a class.

7. In a bulleted list, write down the key points of the class **Matched Pairs Design**.

Part 2: Collecting the Data

Each of you has been assigned to a group. Each group member will play each of the following roles once:

Spitter/Flicker: The Spitter/Flicker stands at the start line and projects a sunflower seed by either spitting or flicking. Then the same person projects a second seed using whatever method was not chosen first. Next, the Spitter/Flicker becomes the Holder.

Holder: The Holder stands at the start line and holds one end of the tape measure. Next the Holder becomes the Walker.

Walker: The Walker stands at the start line and walks with one end of the tape measure to where the sunflower seed came to rest as determined by the Eyes. Next the Walker becomes the Eyes.

Eyes: The Eyes stands about 15-20 feet from the start line and determines where the sunflower seed comes to rest. It's not that easy to see a sunflower seed fly through the air so pay close attention. Next the Eyes becomes the Spitter/Flicker.

Begin by writing your group members' names in the Data Table below. Then, for each group member generate a random number, either 1 or 2. If 2 is generated the person will spit first and then flick. If 1 is generated the person will flick first and then spit. Complete the columns First Method and Second Method in the Data Table.

Data Table

Student Name	First Method (S = Spit, F = Flick)	Distance of 1st Method (inches)	Second Method (S = Spit, F = Flick)	Distance of 2nd Method (inches)

Now let's head outside with those sunflower seeds and tape measures. Have your group complete the Data Table by measuring the distance to the nearest inch.

Part 3: Summarizing the Data Graphically

We will first explore the data from our matched pairs design for the sunflower seed experiment graphically. Remember that our overriding goal is to answer the research question: "Can a sunflower seed be projected further by a person flicking the seed with his or her finger or by spitting the seed?"

We will now summarize the data that we collected. Complete the Data Collection Sheet by copying the data off the transparency and finding the Differences as Spit Distance – Flick Distance.

When data are collected using a matched pairs experiment it is the differences in the two treatments that are important. The whole point of doing a matched pairs experiment is to account for the variability from person to person. If each person gets both treatments (here, spits and flicks a sunflower seed), then any special effects for the person will "average out." Thus, our data analysis will be done on the differences and not on the two separate sets of data.

Use the data from the Data Collection Sheet to complete the following.

1. If neither method is truly better for projecting a sunflower seed, then what would we expect the Differences to be close to?

2. Construct a frequency table of the Differences using approximately 12 classes.

3. Make a histogram of the Differences using the frequency table from #2.

4. Construct a frequency table of the Differences using approximately 6 classes.

5. Make a histogram of the Differences using the frequency table from #4.

6. Describe how you figured out the width of the classes for your 12-classes and 6-classes frequency tables.

7. Describe the center, spread, shape, and outliers for the histogram in #3.

8. Describe the center, spread, shape, and outliers for the histogram in #5.

9. What effect does increasing the number of classes have on the center, spread, shape, and outliers?

10. Make a stemplot of the Differences.

11. Can sunflower seeds be spit or flicked further or are the distances typically about the same? Justify your answer referring to what you did in questions 1 through 10.

Part 4: Summarizing the Data Numerically

Graphical summaries are very helpful because they give us a picture of the distribution of the data. However, graphical summaries aren't enough. We also need numerical summaries of the data that locate the center of the data set and describe the spread (or variability) in the data.

Enter the Differences from the Data Collection Sheet into a list in your calculator.

Use your calculator to help you to answer the following questions.

1. Find the five-number summary, the sample mean, and the sample standard deviation for the Differences. If there is no real difference between Spitting and Flicking, what should the mean and median be close to?

2. Consider the mean and the standard deviation. Together, what do these two numbers tell you about whether or not there is a difference between spitting and flicking?

3. From the five-number summary, make a boxplot of the Differences. Describe the shape of the Differences. If there is no real difference between Spitting and Flicking, where should the boxplot be centered?

4. Consider the boxplot. What does it tell you about whether or not there is a difference between spitting and flicking?

5. Consider the Differences. Suppose The Tick is a member of our class. (For those of you unfamiliar with The Tick, he is a big blue superhero with really good lung capacity. Unfortunately, he's rather clumsy.) The Tick spits the sunflower seed an incredible 1000″, but drops the seed off his finger so that his flicking distance is only 4″.

   (a) What is the Difference (Spit minus Flick) for The Tick's data? Add his Difference to the data set.

   (b) Recalculate the five-number summary, the mean, and the standard deviation with The Tick's data included.

   (c) What effect does an atypically large outlier have on the mean, standard deviation, and each value of the five-number summary?

6. Actually The Tick isn't too bright. He faced the wrong way when spitting the sunflower seed. His sidekick Arthur (a man in a moth suit) convinces you that The Tick's spit value should be –1000″ not 1000″.

   (a) What is the new Difference for The Tick's data? Replace his Difference from #5(a) with the new Difference in the data set.

(b) Recalculate the five-number summary, the mean, and the standard deviation with The Tick's new data included.

(c) What effect does an atypically small outlier have on the mean, standard deviation, and each value of the five-number summary?

7. What effect does being skewed right have on the mean and median of a data set? How will they compare to each other?

8. What effect does being skewed left have on the mean and median of a data set? How will they compare to each other?