4 out of 5 Students Surveyed Would Recommend this Activity (Comparing Chewing Gum Flavor Durations)

Mary Richardson  
Grand Valley State University

Neal Rogness  
Grand Valley State University

Byron Gajewski  
The University of Kansas Medical Center

Abstract

This paper describes an interactive activity developed for illustrating hypothesis tests on the mean for paired or matched samples. The activity is extended to illustrate assessing normality, the Wilcoxon signed rank test, Kaplan-Meier survival functions, two-way analysis of variance, and the randomized block design.

1. Introduction

In this paper, we discuss an activity that revolves around an interactive paired difference experiment. We feel that a strong point of the activity is that the interactive portion requires minimal in-class time. The experiment compares the flavor longevity of two different brands of chewing gum. The experiment was developed for use in an undergraduate general education introductory statistics course. In this course, we use many hands-on interactive activities. We believe that some concepts are difficult for introductory students to learn through lecture and example. We use hands-on explorations of concepts to involve students in the learning process. Cobb (1992, page 9) states: “Shorn of all subtlety and led naked out of the protective fold of educational research literature, there comes a sheepish little fact: lectures don’t work nearly as well as many of us would like to think.” and in describing the Activities-Based Statistics Project, Scheaffer (1996, page vii) states: “Their fast-paced world of action movies, rapid-fire TV commercials and video games does not prepare today’s students to sit and absorb a lecture, especially on a supposedly dull subject like statistics. To capture the interest of these students, teaching must move away from lecture-and-listen to innovative activities that engage students in the learning process.” See Gnanadesikan, Scheaffer, Watkins, and Witmer (1997) for a detailed discussion of utilizing an activities-based approach in an introductory statistics course.

In addition to discussing the use of the activity in the introductory course, we will discuss extensions that can be used in intermediate or upper-level courses. The extensions involve assessing normality, the Wilcoxon signed rank test, Kaplan-Meier survival functions, two-way analysis of variance, and the randomized block design.

1.1 Assessment

The introductory activity to be discussed in Section 2 was assessed in a general education undergraduate course in introductory statistics taught in the summer of 2003 at Grand Valley State University. The class was comprised of nineteen students with majors from many areas. To assess the “likeability” of this activity for learning, questions were administered after the activity was completed. Permission to reproduce answers to these questions was granted by fifteen students.

The randomized block design experiment to be discussed in Section 3.4 was assessed in a non-statistics major graduate course in experimental design taught in the spring of 2003 at the University of Kansas Medical Center. The class was comprised of ten Ph.D. students, with seven from Nursing and three from the Department of Hearing and Speech. Blinded student evaluations (not examined by the instructor until the grades were complete) assessed the “likeability” of the experiment for learning.

2. The Introductory Course Activity

2.1 Background

Prior to completing this activity, students have been exposed to basic experimental design terminology, performed one-sample hypothesis tests on a mean, and hypothesis tests to compare the means of two populations based on independent samples. Dependent samples can be a difficult concept for students to grasp. After discussing several textbook examples of dependent samples, we use this activity to reinforce what we have discussed.

For completeness, we define relevant experimental design terminology.

**Units** are the objects upon which measurements are made or observed.

In an **experiment**, the researcher actively imposes some treatment on the units in order to observe the responses.

The **response variable** measures an outcome of the experiment. It is the variable that is thought to depend on the explanatory variable.

The **explanatory variable** is a variable that is thought to explain or cause the observed outcomes. It is the variable that explains changes in the response variable.

The possible values of the explanatory variable are called the **levels** of that explanatory variable.

A **treatment** is a specific combination of the levels of the explanatory variables.

A **confounding variable** is a variable whose effect on the response variable cannot be separated from the effect of the explanatory variable on the response variable.
A treatment group is a group of experimental units which receive an actual treatment.

Random allocation is a planned use of chance for assigning units to treatments. Randomization tends to produce groups of experimental units that are similar with respect to potential confounding variables.

A single-blind experiment is one in which the units are ignorant of which treatment they receive. A double-blind experiment is one in which neither the units nor those working with the units knows who is receiving which treatment.

A completely randomized design is a design for which independent random samples of experimental units are selected for each treatment.

A crossover design. Cross-over designs account for a “learning” effect or carry-over of the experience by the subject. Accounting for this statistically is out of the scope of this paper. We will assume that there is no carry-over effect. Note that in the presence of multiple sections one might implement a similar design but give the subjects the same brand of gum and one may test day effects to test our hypothesis without utilizing more complex cross-over analyses.

2.2 Procedure

2.2.1 Design

We explain to students that we are going to perform an experiment to determine which of two brands of gum will maintain flavor the longest. We note that the determination of length of gum flavor duration is very subjective. We discuss the fact that some students will undoubtedly judge the gum flavor to have a relatively short duration, while others will judge that the flavor has a long duration. Thus, individual judgment can be expected to be a confounding variable. We note that this confounding variable can be controlled for by blocking (pairing). Each block will be the two flavor duration times for one student.

Assuming a class size of 30 students, the instructor will need 60 small bags, 60 sticks of chewing gum (30 sticks of each of two brands: Brand 1 and Brand 2), 60 sticky notes, and a set of plastic gloves (for sanitary purposes when unwrapping the gum and placing the sticks in a bag). We have used various brands of gum for this activity; both regular and sugar-free stick gums. The only chewing gums we suggest should not be used for this activity are the “intense flavor” gums, whose flavor duration may exceed a typical 50-minute class session. Different choices of gum brands will produce different results in terms of concluding there is/is not a significant difference in the gum flavor longevity. However, this activity can be used irrespective of the conclusion reached from the collected data.

We collect the flavor longevity data during each of two class periods. We ask students to give suggestions for how the experiment should be carried out. For example, should everyone chew Brand 1 on the first day and Brand 2 on the second day? We agree that everyone should not chew the same brand on each day. If everyone chewed the same brand on each day, then order of chewing the brands would be a confounding variable. To prevent chewing order from being confounded with the flavor duration time, we develop a procedure for randomizing the chewing order. One possible approach is to assign each student a two-digit label, and randomly select labels assigning the corresponding students to chew Brand 1 on the first day and Brand 2 on the second day, with the remaining students chewing Brand 2 on the first day and Brand 1 on the second day. Another approach is to have each student flip a coin. If the coin lands heads, that student will chew Brand 1 then Brand 2; otherwise he/she will chew Brand 2 then Brand 1.

We agree that the experiment should be single-blinded. That is, students should not know which gum brand they are chewing. The instructor’s knowledge of which gum brand is being chewed should not introduce bias into the experiment, so there is no need to use double-blinding.

We ask students to think about to what population the results of this experiment can be generalized. Can the results be generalized to all adults? Is it reasonable to assume that the results obtained from using students as experimental units can be generalized to the entire adult population? Are there characteristics present in college students that would influence their judgment on flavor duration, but not influence the judgment of the adult population? We agree that it seems reasonable to generalize the results to the population of all adults.

We ask students if we can make a cause and effect conclusion based on this experiment. Based upon the results of our experiment, can we conclude that a certain brand of gum results in a longer or shorter flavor duration value than another brand? We agree that our experiment may allow us to make such a conclusion.

It is important to note that we have randomized the experiment by day to eliminate a possible confounding effect between day and brand of gum. Statisticians call this a “cross-over” design. Cross-over designs account for a “learning” effect or carry-over of the experience by the subject. Accounting for this statistically is out of the scope of this paper. We will assume that there is no carry-over effect. Note that in the presence of multiple sections one might implement a similar design but give the subjects the same brand of gum and one may test day effects to test our hypothesis without utilizing more complex cross-over analyses.

We hope, that by participating in this experiment, students will have a much better feel for sampling situations for which data are paired or matched. Students are asked to explain why the two samples of flavor longevity values are not independent.

Here are some examples of correct student answers:

- The same student chewed each type of gum hence the observation of chewing one brand is directly related to the chewing of the other brand since the same person is chewing it.
- The two samples are not independent because they are dependent. The same person is being measured for both gums. One set of data can give you information on how the other set will turn out. The data is dependent upon the person’s judgment and their [sic] judgment should be similar for both types of gum. You can see on a scatterplot that there is a linear relationship showing that the two sets of data are dependent on one another. If one is high the other will most likely be high.

Students are asked to discuss a modification of the data collection scheme for the experiment that would result in independent samples.

Here are some examples of correct student answers:

- In order for this experiment to be independent, two separate groups of people need to give data: one group for one gum, another group for another type of gum. Each group chews only one type of gum, and neither group has anything to do with the other.
- For the data collection scheme to get 2 samples of values that are independent, the students should be randomly assigned to two groups, where each group samples only 1 kind of gum. This would make it so that the observations of one set of data would not influence the other observation’s set.

2.2.2 Data Collection

To implement single-blinding, we remove the gum wrappers, and place each stick in a bag, marking the outside of the bag to indicate Brand 1 or Brand 2. We ask students to remember their chewing order, and on the first day of data collection, at the beginning of the class period, we give each student a bag containing his/her respective gum brand, and one sticky note. For classrooms that have a wall clock, we ask students to keep track of the flavor longevity of their gum to the nearest minute and to record their name, the Brand chewed (1 or 2), and the minutes of flavor duration on their sticky note. For classrooms without a wall clock, we place a mid-sized digital timer in the front of the classroom. We ask students to watch the timer and to record, in minutes, how long the gum flavor lasted. We follow the same procedure on the second day of data collection.

We omit data collected for only one class period (in the case of a student being absent on one of the data collection days). In Figure 1, we have included an example class dataset. These results are for two brands of sugar-free cinnamon stick gum.
2.2.3 Data Analysis

After the flavor duration has been recorded for both gum brands, we give students a Data Sheet (see Appendix A.1) and a Worksheet (see Appendix A.2) that must be completed as homework. We ask students to use a graphing calculator to perform calculations for the data.

Students calculate the differences in flavor duration values (Brand 1 minus Brand 2) and enter the differences into the appropriate column on the Data Sheet. Once these differences have been calculated, students construct a boxplot to display the distribution of the differences. Based on the boxplot, students explain whether they believe that the flavor duration differs for the two gum brands. Figure 2 shows descriptive statistics and a boxplot for the example class differences.

![Boxplot](image)

**Figure 2. Descriptive Statistics and Boxplot of Example Class Differences (differences are formed by: Brand 1 - Brand 2)**

From the boxplot we see that roughly half of the differences are negative and half are positive, indicating that there does not seem to be a difference in the flavor duration of the two brands of gum.

The test statistic formula for performing a hypothesis test on the mean difference for paired samples is introduced. Students perform a hypothesis test to determine if there is a significant difference in the mean number of minutes that the flavor lasts for Brand 1 and Brand 2 gums. The test statistic value for the example class data is 0.73, with a corresponding $P$-value of 0.47, well above any reasonable level of significance. Thus, we cannot conclude that there is a significant difference in the mean flavor duration, in minutes, for Brand 1 and Brand 2 chewing gums. We note that the result obtained here is typical when comparing two brands of sugar-free cinnamon stick gum.

Students give a practical interpretation of the $P$-value that was calculated in performing the hypothesis test. Students construct a 95% confidence interval for the mean difference in flavor duration, in minutes, for gum Brands 1 and 2 (Brand 1 - Brand 2) and explain how the confidence interval gives the same conclusion as the hypothesis test. For the example class differences, the 95% confidence interval is: (-3.39, 7.19).

**Note to the Instructor:** We have found that the safest choice for a gum flavor to use that will not result in any censored data values or data values that are close to the 50-minute cut-off is sugar-free bubble gum flavored stick gum.

2.3 Assessment

2.3.1 Assessing the Activity

After answering questions pertaining to analyzing the differences in the flavor longevity values, students are asked to reflect on what they have learned through completing this activity.

Students are asked if they feel that this activity should be used in future introductory statistics classes. Although we have not found that exactly 80% of students would recommend continuing this activity (i.e. 4 out 5), the good news is that our 95% confidence interval estimate of the percentage of students who would recommend this activity is: (81%, 100%).

Students are asked if they feel that the instructions for completing the activity are clear, and if they do not think the instructions are clear, they are asked to state what they would change in order to make them clear. Overwhelmingly, students respond that they completely understood what was expected of them in completing this activity. Some example student responses are:

- The directions are perfectly clear. They spell out everything.

| Brand 1, Brand 2 | Flavor Duration in Minutes |
|------------------|-----------------------------|
| (10, 35)         | (30, 25)                    |
| (43, 30)         | (45, 40)                    |
| (35, 30)         | (35, 14)                    |
| (40, 23)         | (7, 9)                      |
| (7, 23)          | (13, 30)                    |

**Figure 1. Example Class Data for Sugar-free Cinnamon Stick Gum**
Students are asked if they think that participating in this activity helped them to think about independent samples versus dependent samples. Some example student responses are:

- Yes, this experiment did make me think about independent samples versus dependent samples, especially question 1 which made me stop and think about the differences between them.
- Yes. Obviously, if someone has a 20 minute value for brand 1, they [sic] won't have a 60 minute value for brand 2 and vice versa.
- Yes. I think that any time you physically get involved with learning you can gain more understanding of the topic.

Students are asked to state why we cannot ignore pairing and analyze paired samples data as if we had two independent samples. Some example student responses are:

- Pairing cannot be ignored because the paired samples are linked by the subjectivity of the person chewing the gums. Whether it is realized or not, each gum is being compared to the other, and this link between the pairs cannot be broken to cause the samples to become independent of one another.
- Because the results are inevitably dependent on each other. There is always something that connects the 2 samples together.
- One set of values wouldn't mean much if not compared to the other set of values. They had too much of an influence on each other.

2.3.2 Assessing Student Learning

Here is an example examination question that we use to determine if our introductory students have achieved competency in distinguishing between independent and dependent samples and the corresponding appropriate hypothesis test to compare means. On the exam, students are allowed the use of formula sheets, but must show the work for the calculation of all test statistics.

**Exam Question:** (Adapted from a question found in McClave and Sincich (2003).)  
In each scenario described below, we are interested in comparing a variable measured for two different groups.

**#1:** A pupillometer is a device used to observe changes in an individual’s pupil dilations as he or she is exposed to different visual stimuli. The Design and Market Research Laboratories of the Container Corporation of America used a pupillometer to evaluate consumer reaction to different silverware patterns for one of its clients. Suppose five consumers were chosen at random and each was shown two different silverware patterns.

| Consumer | 1   | 2   | 3   | 4   | 5   |
|----------|-----|-----|-----|-----|-----|
| Pattern 1 readings | 1.00 | 0.95 | 1.45 | 1.20 | 0.75 |
| Pattern 2 readings | 0.80 | 0.65 | 1.25 | 1.00 | 0.80 |

(a) These samples are (circle one): independent matched (or paired)  
Because ...

(b) Calculate the value of the test statistic for performing a hypothesis test to determine if the data provide significant evidence to indicate that there is a difference in the mean pupillometer readings for the two patterns. Do not perform the hypothesis test, only calculate the value of the test statistic.

**#2:** A pupillometer is a device used to observe changes in an individual’s pupil dilations as he or she is exposed to different visual stimuli. The Design and Market Research Laboratories of the Container Corporation of America used a pupillometer to evaluate consumer reaction to different silverware patterns for one of its clients. Suppose ten consumers were chosen at random. Five of the consumers were shown silverware Pattern 1, and the other five consumers were shown silverware Pattern 2.

| Pattern 1 readings | 1.10 | 0.90 | 1.40 | 1.25 | 0.85 |
| Pattern 2 readings | 1.00 | 0.75 | 1.25 | 1.00 | 0.90 |

(a) These samples are (circle one): independent matched (or paired)  
Because ...

(b) Calculate the value of the test statistic for performing a hypothesis test to determine if the data provide significant evidence to indicate that there is a difference in the mean pupillometer readings for the two patterns. Do not perform the hypothesis test, only calculate the value of the test statistic.

In the table below, we provide an example of student performance on this exam question. Recall that results are presented for fifteen student responses.

**Example Exam Results:**

| Correctly Identified Sampling Scheme for Scenario 1 | Correctly Identified Sampling Scheme for Scenario 2 | Correctly Identified Test Statistic for Scenario 1 | Correctly Identified Test Statistic for Scenario 2 |
|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| 15/15                                      | 15/15                                      | 9/15                                        | 8/15                                        |

The results are quite mixed. A very positive aspect of the results is that every one of the students was able to correctly classify the two sampling scenarios as either independent or dependent. However, a negative aspect of the results is that some students have difficulty identifying the appropriate hypothesis testing procedure to apply, after they have categorized samples as being dependent or independent. We believe that the biggest obstacle that prevents some students from correctly identifying the paired-difference test statistic formula is that, even though they realize they are dealing with dependent samples, they still focus on the fact that they have two samples. They are unable to form differences within the pairs as a first step and then proceed to apply the correct test statistic formula to the differences. Instead, they apply the two independent samples formula.

It is important that the students recognize instances of cases of when to use the paired test statistic and the independent test statistic. However, it is also important for the students to check the condition needed to use the intended probability distribution for the test statistic. This condition is that the distribution for the x-bar values follows a normal curve (or is at least quite close to a normal curve). The operational check for this is either to observe that the original values are essentially normally distributed or that the sample size is reasonably large. Students should be aware that the probability distribution called for in a hypothesis test or confidence interval is not always the correct description of the probabilities involved - certain requirements need to be satisfied in order to appeal to that distribution. This is particularly true for normal, t, chi-squared, and F distributions. Therefore we dedicate the subsequent section to exploring how close the raw data is to being normally distributed.

### 3. Extensions of the Activity
3.1 Extension 1: Assessing Normality

3.1.1 Background

Prior to completing extension 1, students have performed parametric and nonparametric hypothesis tests (one and two sample, independent and dependent) and have been introduced to Q-Q plots. We use this extension to generate a dataset for which students must determine the appropriate statistical procedure to apply. The same materials are used as for the introductory activity. In addition, use of a statistical software package is required.

3.1.2 Procedure

We use the same data collection scheme as for the introductory activity. After the flavor duration values have been recorded for both gum brands, we give students the Data Sheet (see Appendix A.1) and a Worksheet (see Appendix B) that must be completed as homework. In Figure 3, we have included a second example class dataset. For this example dataset, our class size is 25 students, and the results are for two brands of spearmint flavored stick gum.

Students are asked to explain why the two samples are not independent. Students calculate the differences in the number of minutes the flavor lasted for the two gums (Brand 1 minus Brand 2) and enter the differences into the appropriate column on the Data Sheet.

Students calculate the mean, median, and quartiles for the differences and use these calculations to help determine if it can be assumed that the distribution of the differences is a normal distribution. Students construct a stem-and-leaf plot of the differences and check for non-normal features. The mean and median differences are compared, as are the distances from the quartiles to the median. Students use a statistical software package to construct a Q-Q plot of the differences. Students write a summary paragraph to explain whether they believe it can be assumed that the distribution of the differences is a normal distribution. Figure 4 displays results for assessing the normality of the example class differences for the spearmint flavored stick gum.
For the example class differences, the stem-and-leaf plot shows a slightly right skewed distribution. A mean of 7.48 minutes compared to a median of 5.00 minutes also indicates that the distribution is right skewed. The first quartile is 5.00 minutes below the median, while the third quartile is 8.50 minutes above the median, indicating a right skew. However, the Q-Q plot does not show a marked departure from linearity.

The results of the normality assessment will depend on the brands and flavors of gum used. This extension can be used irrespective of the conclusion reached concerning normality.

Students construct a boxplot to display the distribution of the differences. Based on the boxplot, students explain whether they believe that the flavor duration differs for the two gum brands. Figure 5 shows a boxplot of the differences for the spearmint flavored gum. The circle on the boxplot indicates an outlying difference value. An outlier is defined as a difference value that is more than 1.5 times the IQR beyond Quartile 1 or Quartile 3 (where IQR = Quartile 3 - Quartile 1).

Figure 5. Boxplot of Example Class Differences (differences are formed by: Brand 1 - Brand 2)

From the boxplot we see that roughly 75% of the differences are positive, indicating that the flavor duration of the Brand 1 gum appears to last longer than that of Brand 2.

Students conduct an appropriate statistical hypothesis test to determine if there is a significant difference in the flavor duration for Brand 1 and Brand 2 gums. The hypothesis testing procedure that is applied will depend on whether the distribution of differences is judged to be non-normal.

For the example class data, since the normality checks indicate that it may not be safe to assume that the population of differences is normal, students might apply the Wilcoxon signed rank test. For these data, the Wilcoxon signed rank test statistic produces a P-value of 0.001, well below any reasonable level of significance. It is therefore concluded that there is a significant difference in the typical flavor duration of the two gum brands. Students give a practical interpretation of this P-value.

3.2 Extension 2A: One Sample Censoring

3.2.1 Background

Prior to completing extensions 2A and 2B, students should have performed hypothesis tests. The same materials are used for this extension as for the introductory activity. In addition, use of a statistical software package is required.

3.2.2 Procedure

For this censored case we collect data for only one brand of gum. In order to obtain a single right-censored sample, rather than starting the data collection at the beginning of a 50-minute class period, we start at approximately 10 minutes after the beginning of the period. We do not tell students that it is our goal to obtain censored data. We introduce the data collection scheme and we mention to students that, if in their opinion, the gum still has flavor at the end of the class period they should record ‘still has flavor’ on their sticky note. In Figure 6, we have included an example class dataset. For this example dataset, our class size is 25 students, and the results are for one brand of sugar-free cinnamon flavored stick gum.

During the next classroom period, we give students the flavor longevity data values, using a ‘40c’ to indicate censoring at 40 minutes. We introduce the concept of censored data and note that many of the longevity values are censored. We ask students to view the flavor longevity values as the survival lengths of experimental units in a study. We discuss with students how the censoring might affect the analysis of the data and introduce the concept of a survival function and discuss the interpretation of a Kaplan-Meier life table and survival function. Students are given a Worksheet (see Appendix C.1) that must be completed as homework. The Worksheet formally introduces students to introductory terminology pertaining to censoring and survival analysis.
Students enter the data into a statistical software package and generate the Kaplan-Meier life table. Students generate a plot of the Kaplan-Meier survival function. In Figure 7, we have included the Kaplan-Meier life table and survival function for the example class data.

| Time | Status     | Cumulative Survival | Standard Error | Number of Events | Number Remaining |
|------|------------|----------------------|----------------|------------------|-----------------|
| 7.00 | Not Censored | 0.9200               | 0.0543         | 2                | 23              |
| 13.00| Not Censored  | 0.8800               | 0.0650         | 3                | 22              |
| 20.00| Not Censored  | 0.8000               | 0.0800         | 5                | 20              |
| 22.00| Not Censored  | 0.7600               | 0.0854         | 6                | 19              |
| 27.00| Not Censored  | 0.7200               | 0.0898         | 7                | 18              |
| 30.00| Not Censored  | 0.6400               | 0.0960         | 9                | 16              |
| 31.00| Not Censored  | 0.6000               | 0.0980         | 10               | 15              |
| 32.00| Not Censored  | 0.5600               | 0.0993         | 11               | 14              |
| 35.00| Not Censored  | 0.4800               | 0.0999         | 13               | 12              |
| 40.00| Not Censored  | 0.4000               | 0.0980         | 15               | 10              |
| 40.00| Censored          |                      |                |                  |                 |
| 40.00| Censored          | 0.0800               | 0.0854         | 6                | 19              |
| 40.00| Censored          | 0.0898               | 0.0960         | 9                | 16              |
| 40.00| Censored          | 0.0980               | 0.0993         | 11               | 14              |
| 40.00| Censored          | 0.0999               | 0.1000         | 13               | 12              |
| 40.00| Censored          | 0.1000               | 0.1000         | 15               | 10              |

Number of Cases: 25  Censored: 10  (40.00%)  Events: 15

Figure 7. Survival Table and Survival Function for Example Censored Class Data

Students interpret the survival table and the plot of the survival function. An estimated 75% of the chewers consider their gum to still have flavor at 25 minutes. However, by 35 minutes, this percentage drops to approximately 50%. Approximately 40% of the chewers consider the gum to maintain flavor for at least 40 minutes.

3.3 Extension 2B: Two Sample Censoring

3.3.1 Background

Prior to completing extension 2B students should have completed extension 2A. This extension does not require additional materials or data collection. The use of a statistical software package is required.

3.3.2 Procedure

For this censored case we utilize data for two brands of gum. We use the data collected for Brand 1 in extension 2A to represent the survival times of experimental units under Treatment 1. We use the data collected for Brand 2 from one of our introductory classes (considering any duration value that exceeds 40 minutes to be censored) to represent the survival times of experimental units under Treatment 2.

We discuss with students that our task is to determine which of two treatments is more effective in prolonging the survival time of the experimental units. We note that we are assuming that the two treatment groups (the chewers of Brands 1 and 2 gums) are both representative of their respective populations. We give students the data, along with a Worksheet (see Appendix C.2) that must be completed as homework. In Figure 8, we have included example class datasets. For these example datasets, our class size is 25 students for Brand 1 and 18 students for Brand 2. The results are for two brands of sugar-free cinnamon flavored stick gum.
In extension 2A, students used a statistical software package to construct the Kaplan-Meier life table and survival function for the Brand 1 flavor duration values. Students use a statistical software package to construct the Kaplan-Meier life table for the Brand 2 flavor duration values. Students plot Kaplan-Meier survival functions for the flavor duration values of both gum brands on the same graph. In Figure 9, we include a plot of the Kaplan-Meier survival functions for the example class data.

Students compare the overall survival rates for the two brands and examine the plot of the survival functions to determine if it appears that one of the brands tends to have longer flavor duration values than the other brand. At 10 minutes, the percentage of chewers still detecting flavor in the gum is approximately 95% for Brand 2 and 90% for Brand 1. At 15 minutes, the percentage still detecting flavor drops to 88% for Brand 1 and stays at 95% for Brand 2. However, at 20 and 25 minutes, the percentages still detecting flavor are higher for Brand 1, by approximately 2% and 8%, respectively. The median percentage for Brand 2 occurs at approximately 28 minutes, while the median for Brand 1 is at approximately 35 minutes. The percentage still detecting flavor at 40 minutes and beyond is approximately 40% for Brand 1 and only 22% for Brand 2.

Students use a statistical software package to perform a hypothesis test to determine if there is a significant difference in the flavor duration values for the two brands. Figure 10 shows test statistics and $P$-values for the Log Rank and Wilcoxon (Breslow’s generalized Wilcoxon) test procedures.

The $P$-values for the test procedures are both above any reasonable level of significance. Thus, we cannot conclude that there is a significant difference in the flavor duration values for the two brands.

### 3.4 Extension 3: Two-Way Analysis of Variance

#### 3.4.1 Background

Prior to completing extension 3, students have performed both a one-way and a two-way analysis of variance. Further, students have experience assessing the presence of interaction and performing multiple comparisons. After discussing several textbook examples involving two-way analysis of variance, we use this activity to reinforce what we have discussed.

We explain to students that it is desired to perform an experiment that involves a quantitative response variable and at least two qualitative attributes, all of which are related to gum. Students are divided into groups and asked to brainstorm potential variables of interest. The various ideas are collected via the whiteboard and the merits of each are discussed. Potential qualitative variables include flavor of gum, sugar vs. sugar-free, brand of gum, and type of piece of gum (i.e., stick vs. tablet). Potential quantitative variables include a rating of the gum flavor, the flavor intensity, the texture of the gum, and the length of flavor duration.

For this example, the independent variables chosen were Flavor of gum (spearmint vs. winterfresh) and Piece of gum (stick vs. tablet) and the response variable was the...
Texture of gum (using a rating scale where 0 = very soft, pliable and 10 = very hard, rubbery). To reduce confounding, all four gums were manufactured by the same company and were all sugar-free.

Assuming a class size of 30 students, the instructor will need 120 small paper cups, 120 pieces of chewing gum (30 sticks of each of two flavors: spearmint and winterfresh), three sets of playing cards, 120 sticky notes (preferably 30 of four different colors), 30 index cards, and a set of plastic gloves (for sanitary purposes when unwrapping gum and placing the pieces in a cup). In addition, use of a statistical software package is required.

3.4.2 Procedure

To facilitate performing the experiment, four colors of sticky notes are utilized. Each color of sticky note is assigned an order of administration. For instance, blue might represent the first treatment and yellow might represent the fourth treatment. The four sticky notes are affixed to an index card, always in the same color order. The playing cards are sorted into sets of four so that each set contains a heart, a spade, a club, and a diamond. Each suit is randomly assigned to one of the four treatments (four combinations of Flavor and Piece of gum).

During a class period before the data collection begins, each student is given an index card and asked to write his or her name on each of the four sticky notes. Then each student is given a set of cards and asked to shuffle the cards. When the top card is flipped over, the students are told to write the name of the suit on the first sticky note. This process is repeated using the remaining cards and sticky notes. After class, the sticky notes are removed from the index cards and each is affixed to a small paper cup. A piece of unwrapped gum, corresponding to the suit on the sticky note, is placed in a cup and the cup is stapled shut.

The instructor assigns a treatment number to each of the four combinations of Flavor and Piece and on the first day of data collection, those cups with the color of sticky notes corresponding to Treatment One are taken to class. A Ratings Sheet is passed out (see Appendix D), as are the cups. Each student is asked to claim the cup with his or her name on it and to record his or her name on the Ratings Sheet, along with the suit indicated on the sticky note. All students begin chewing at the same time. Students are told to record a texture score on the data collection sheet at the time that the chewed gum is discarded. This process is repeated on the subsequent three class periods. The color of sticky notes is useful in helping to determine which gum to chew in the event that a student is absent on a data collection day and has more than one cup in the set of cups being passed around. In Figure 11, we have included an example class dataset.

Table 1. Descriptive statistics for the response variable Texture based upon the four treatments.

| PIECE  | FLAVOR     | N  | Mean | Std. Deviation |
|--------|------------|----|------|----------------|
| Tablet | Spearmint  | 21 | 2.71 | 1.793          |
| Tablet | Winterfresh| 22 | 2.41 | 1.563          |
| Stick  | Spearmint  | 22 | 5.27 | 2.051          |
| Stick  | Winterfresh| 23 | 4.87 | 2.181          |

We ask students to input the class data into a statistical software package and perform a two-way analysis of variance using the response variable Texture and the main effects of Flavor and Piece of gum. Students are instructed to prepare a report of their analysis and are asked to address several questions in their report (see Appendix E for the questions).

For each treatment, students are asked to generate the mean and standard deviation of the Texture scores. In Table 1, we have included these descriptive statistics for the four treatments.

Table 2. Two-way ANOVA summary table for the response variable Texture.

| Source        | Type III Sum of Squares | df | Mean Square | F     | P-value |
|---------------|-------------------------|----|-------------|-------|---------|
| Model         | 1439.424                | 4  | 359.856     | 97.959| 0.000   |
| PIECE         | 138.399                 | 1  | 138.399     | 37.675| 0.000   |
| FLAVOR        | 2.757                   | 1  | 2.757       | 0.750 | 0.389   |
| PIECE*FLAVOR  | 0.053                   | 1  | 0.053       | 0.014 | 0.905   |
| Error         | 308.576                 | 84 | 3.674       |       |         |

Students are asked to generate an ANOVA summary table and make a conclusion regarding whether Flavor of gum and Piece of gum interact to affect the mean Texture score. Students are also asked to generate an interaction plot and discuss whether the conclusion regarding interaction agrees with what the plot shows. The profile plot in Figure 12 shows the absence of interaction between the independent variables Piece and Flavor, which is consistent with the information given in Table 2 ($F = 0.01, P = 0.91$).
R Squared = 0.823 (Adjusted R Squared = 0.815)

Figure 12. Profile Plot of Piece (Tablet vs. Stick) and Flavor (Spearmint vs. Winterfresh) for the Response Variable Texture.

In the absence of interaction, students are asked to use the appropriate generated P-values to make conclusions regarding whether Flavor or Piece of gum are significant. Students are asked to compare means corresponding to the levels of the significant factors.

Table 2 shows that Flavor is not significant with respect to Texture ($F = 0.75, P = 0.39$), whereas Piece is significant ($F = 37.68, P < 0.001$), with the mean texture for stick gum (5.07) significantly higher than the mean texture for tablet gum (2.56) (see Table 3 and Table 4).

Table 3. Descriptive statistics for the response variable Texture based upon the Flavor of gum.

| FLAVOR  | N  | Mean | Std. Deviation |
|---------|----|------|----------------|
| Spearmint TEXTURE | 45 | 4.02 | 2.304 |
| Winterfresh TEXTURE | 45 | 3.67 | 2.256 |

Table 4. Descriptive statistics for the response variable Texture based upon the Piece of gum.

| FLAVOR  | N  | Mean | Std. Deviation |
|---------|----|------|----------------|
| Tablet TEXTURE | 45 | 2.56 | 1.666 |
| Stick TEXTURE | 45 | 5.07 | 2.104 |

At this point we ask the students to step back from these conclusions and consider the statistical assumptions that validate the inference made from the ANOVA table. We remind the students that there are three assumptions. The first is that values from the populations for each cell come from a normal distribution, second that those distributions have a common value for their variances, and third that the responses are independent. Students can investigate the first two assumptions by observing a Q-Q plot of the residuals (which for brevity we skip) and the summary statistics presented in Table 1 respectively. The standard deviations are fairly close; assuming equal variances is not unreasonable. Independence is the assumption in clear violation and can be shown to students by considering the following question:

- Student A and Student B are each exposed to Treatment 1 and Treatment 2. With each treatment a score is obtained. Is it reasonable to assume that Student A’s score for a given treatment is independent of Student B’s score for the same treatment? Why or why not? Is it reasonable to assume that Student A’s score for Treatment 1 is independent of Student A’s score for Treatment 2? Why or why not?

Given their answers to this question, students are asked to discuss what concerns they may have about using a two-way analysis of variance design for the gum texture ratings data. In addition, they are asked to suggest possible alterations to the experimental design to correct for this concern. Readers may have noted that the experiment discussed in this section is not a true completely randomized design. Rather, the design can be viewed as a randomized block design with each student as a block.

After discussing the results for a two-way analysis of variance, we ask students to reformat the class data and use a statistical software package to run a randomized block design (blocking on student) using the response variable Texture and the main effects of Flavor and Piece of gum. Students are instructed to prepare a report of their analysis and are asked to address questions [2], [3], and [4] from the Appendix E questions sheet.

In Figure 13, we modify the example class texture data to include fixed blocking. There are missing values in the data, but we include only those cases for which there are complete data ($n = 20$), as dealing with missing data is beyond the scope of this activity.

0 = very soft, pliable, ..., 10 = very hard, rubber
Table 5 shows that there is no significant interaction between the independent variables Piece and Flavor ($F = 0.005, P = 0.94$). Flavor is not significant with respect to Texture ($F = 1.11, P = 0.30$), whereas Piece is significant ($F = 44.47, P < 0.001$), with the mean texture for the stick gums significantly higher than the mean texture for the tablet gums. Table 6 displays estimates for individual treatment means.

Table 5. ANOVA summary table for randomized block design.

| Source          | Type III Sum of Squares | df | Mean Square | F     | P-value |
|-----------------|-------------------------|----|-------------|-------|---------|
| Model           | 1378.388                | 23 | 59.930      | 23.622| 0.000   |
| PIECE           | 112.812                 | 1  | 112.812     | 44.466| 0.000   |
| FLAVOR          | 2.813                   | 1  | 2.813       | 1.109 | 0.297   |
| PIECE*FLAVOR    | 0.013                   | 1  | 0.013       | 0.005 | 0.944   |
| SUBJECT         | 145.237                 | 29 | 7.644       | 3.013 | 0.000   |
| Error           | 144.613                 | 57 | 2.537       |       |         |
| Total           | 1523.000                | 80 |             |       |         |

R Squared = 0.905 (Adjusted R Squared = 0.867)

Table 6. Descriptive statistics for the response variable Texture based upon the treatment.

| Treatment       | Mean  | Std. Error | 95% Confidence Interval |
|-----------------|-------|------------|-------------------------|
| Stick Spearmint | 5.100 | 0.356      | 4.387, 5.813            |
| Tablet Spearmint| 2.750 | 0.356      | 2.037, 3.463            |
| Tablet Winterfresh| 2.350 | 0.356    | 1.637, 3.063            |
| Stick Spearmint | 4.750 | 0.356      | 4.037, 5.463            |

Figure 13. Example Class Texture Scores for Randomized Block Design
This experiment was done in the graduate course and in answering a multiple response question “The following teaching strategy assisted in my learning: class experiments” three students “agreed” and seven “strongly agreed.”

4. Conclusions

This activity has a wide range of possible uses and extensions. It can be used in intermediate or upper-level applied courses as well as an introductory course, an applied course for the health professions, or an introductory biostatistics course.

Students at all levels enjoy participating in this activity and develop an interest in analyzing the data to determine whether there is a significant difference in the flavor duration of the brands (or types) of chewing gum.

The activity provides introductory students with a concrete example of paired or matched samples. In an intermediate course, the activity provides a paired dataset for which students must determine the appropriate statistical procedure to apply. Altering the data collection scheme allows the instructor to use the activity to introduce a basic analysis of right-censored data and to discuss the comparison of survival rates. Introducing different gum types and flavors gives the instructor an opportunity to discuss principles of experimental design and allows students to interactively generate a dataset that can be analyzed using a two-way analysis of variance or a randomized block design.

One obvious conclusion from this paper is that the gum experiment looks promising for assisting beginning non-mathematical college students in understanding the difference between paired and independent data. In addition, the use of the later activities look promising given the enthusiastic response in the graduate course. Now hopefully at least 4 out of 5 readers will give these activities or modifications of them a try in the class!

Appendix A: Introductory Activity

pdf version of Appendix A: A.1 Data Sheet

A.1 Data Sheet

| Student | Brand 1 Flavor Duration (minutes) | Brand 2 Flavor Duration (minutes) | Difference in Flavor Duration (Brand 1 - Brand 2) |
|---------|----------------------------------|----------------------------------|-----------------------------------------------|
A.2 Worksheet

Which Gum Lasts Longer?

Background: (Background taken from: Brandt (2001) “FORMULATION CHALLENGE: CONFECTIONERY - A STICKY Situation,” found online at: www.preparedfoods.com/CDA/ArticleInformation/features/BNP_Features_Item/0,1231,114008,00.html.) “Gum chewing dates back to ancient civilizations. Ancient Greeks chewed mastic tree resin, ancient Central American Mayans chewed chicle, and American Indians chewed gum made from spruce tree resin. This gum was eventually replaced by paraffin wax gum. Today’s chewing gums are made mostly of synthetic materials.”

“Long-lasting flavor is one of the ‘holy grails’ of the chewing gum industry. For most chewing gums today, flavor lasts about 12 to 13 minutes as a standard…”

Problem: We want to determine if there is a significant difference in the mean number of minutes it takes for two different brands of chewing gum to lose their flavor.

Instructions: The Data Sheet contains the gum data that was collected in class (the length of time, in minutes, that the flavor lasted for Brand 1 and Brand 2 gums).

1. a. Explain why our two samples (the number of minutes that the flavor lasted for chewing gum Brand 1 and chewing gum Brand 2) are not independent.
   b. Explain how our data collection scheme would have to be changed for this experiment in order to get two samples of flavor duration values that are independent. Be sure to clearly explain why the samples would be independent, under your modified data collection scheme.

2. Calculate the differences in the number of minutes the flavor lasted for the two gums (Brand 1 minus Brand 2). Enter the differences into the appropriate column on the Data Sheet.

3. Construct a boxplot of the differences.
   \[ \text{min} = _____ \quad \text{quartile 1} = _____ \quad \text{median} = _____ \quad \text{quartile 3} = _____ \quad \text{max} = _____ \]

   Boxplot:
   
   -40 -30 -20 -10 0 10 20 30 40

   Based on the boxplot, would you conclude that there is a difference in the number of minutes that flavor is retained for Brand 1 and Brand 2 gums? Explain.

4. Explain why we cannot make a formal statistical conclusion that compares the flavor durations of the two brands of gum based solely on an examination of the boxplot constructed in Question 3.

5. Recall, that in order to test \( H_0: \mu = \mu_0 \), we use the test statistic: \( t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \), where \( \mu_0 \) is the null hypothesized value for \( \mu \), \( \bar{x} \) is the mean of the sample, \( s \) is the standard deviation of the sample, \( n \) is the sample size, and \( n - 1 \) degrees of freedom are used for the test.

   Let \( \mu_0 \) be the mean difference in the number of minutes that flavor is retained (Brand 1 minus Brand 2).

   To test: \( H_0: \mu_d = \mu_{d_0} \), based on a simple random sample of \( n_d \) differences from the population, we use the test statistic: \( t = \frac{\bar{x}_d - \mu_{d_0}}{s_d/\sqrt{n_d}} \), where \( \bar{x}_d \) is the mean of the sample differences, \( s_d \) is the standard deviation of the sample differences, and \( n_d - 1 \) degrees of freedom are used for the test.

   Perform a hypothesis test to determine if there is a significant difference in the mean number of minutes that the flavor lasts for Brand 1 and Brand 2 gums.

   1. \( H_0: \)
   \[ H_{\alpha}: \]

   2. test statistic \( t = \frac{\bar{x}_d - \mu_{d_0}}{s_d/\sqrt{n_d}} \)

   3. \( P\)-value =

   4. conclusion =

   6. Give a practical interpretation of the \( P\)-value that you calculated in Question 5.
7. Construct a 95% confidence interval for $\mu_1 - \mu_2$ (use the formula: $\bar{x} \pm t_{\alpha/2} \frac{s_p}{\sqrt{n}}$). Explain how the confidence interval gives the same conclusion as the hypothesis test performed in Question 5.

8. a. Would you recommend that this activity be used in future introductory statistics classes? Yes _____ No _____
   b. Do you think that the instructions for this activity are clear? Why or why not?
   c. Do you think that participating in this activity helped you to think about independent samples versus dependent samples? Why or why not?
   d. Why can we not ignore pairing and analyze paired samples data as if we had two independent samples?

---

**Appendix B: Assessing Normality Worksheet**

**Which Gum Lasts Longer?**

**Problem:**

We want to determine if there is a significant difference in the typical flavor duration, in minutes, for two different brands of chewing gum.

**Instructions:**

The Data Sheet contains the gum data that was collected in class (the length of time, in minutes, that the flavor lasted for Brand 1 and Brand 2 gums).

1. Explain why our two samples (the number of minutes that the flavor lasted for chewing gum Brand 1 and chewing gum Brand 2) are **not** independent.
2. Calculate the differences in the number of minutes the flavor lasted for the two gums (Brand 1 minus Brand 2). Enter the differences into the appropriate column on the Data Sheet.
3. Calculate the following quantities for the differences:
   - $\text{min} = \text{____}$
   - $\text{q1} = \text{____}$
   - $\text{median} = \text{____}$
   - $\text{q3} = \text{____}$
   - $\text{max} = \text{____}$
   - $\text{mean} = \text{____}$
4. Now, determine if it is safe to assume that the distribution of the differences is a normal distribution.
   a. Construct a stem-and-leaf plot of the differences and check for non-normal features such as gaps, outliers, or pronounced skewness. Do you detect any non-normal features?
   b. Compare the mean difference to the median difference. Recall that, in a normal distribution, the mean and the median will be roughly the same. Are the mean and the median roughly the same? If not, what do the values indicate about the shape of the distribution of the differences?
   c. Compare the distance from the difference quartiles to the median difference. Recall that, in a normal distribution, the first quartile and the third quartile will be approximately the same distance from the median. Are the quartiles approximately the same distance from the median? If not, what do the distances indicate about the shape of the distribution of the differences?
   d. Use a statistical software package to construct a Q-Q plot of the differences. Does the Q-Q plot indicate that the differences do not have a normal distribution? Explain your answer.
5. In your opinion, is it safe to assume that the distribution of the differences is a normal distribution? Write a summary paragraph to explain your answer.
6. Construct a boxplot of the differences.
   - $\text{min} = \text{____}$
   - $\text{q1} = \text{____}$
   - $\text{median} = \text{____}$
   - $\text{q3} = \text{____}$
   - $\text{max} = \text{____}$
   **Boxplot:**

   ![Boxplot Image]

   Based on the box plot, would you conclude that there is a difference in the number of minutes that flavor is retained for Brand 1 and Brand 2 gums? Explain.

7. Conduct an appropriate statistical hypothesis test to determine if the typical flavor duration differs for chewing gum Brands 1 and 2. The test procedure that you use will depend on your answer to Question 5.
   1. $H_0$:
   2. $H_a$:
   2. formula for test statistic =
calculated value of test statistic = 3.

3. P-value =

4. conclusion =

8. Give a practical interpretation of the P-value that you calculated in Question 7.

Appendix C: Censoring Worksheets

pdf file for Appendix C.1: One Sample Censoring Worksheet

C.1 One Sample

How Long does the Gum Last?

Problem:

We want to determine the flavor duration, in minutes, for a certain brand of chewing gum. However, some of our data values have been censored at 40 minutes. How can we analyze this data?

Background: (Collett (1996); Lang and Secic (1997))

In analyzing the gum flavor duration data, we are analyzing the times to an event (the event that the gum flavor has expired). Survival analysis is a common application of time-to-event analysis. Estimates can be obtained of the probability of survival (the event does not occur) as a function of time from a starting point. Any event occurring at the end of some time interval, such as the death of a medical patient, or the failure of a part in a piece of equipment, can be viewed as the event in a survival analysis.

Survival analysis requires special statistical methods due to the fact that the event of interest may not yet have occurred when the data analysis is performed. When data are collected on occurrence times for an event of interest and the data includes events that have not yet occurred, these data are said to be right-censored. Survival analysis methodology can incorporate right-censored data. Note that the gum flavor duration values are right-censored (at 40 minutes).

One way to summarize survival data is through estimates of the survival function. The survival function, S(t), is the probability that the survival time is greater than or equal to time, t. That is S(t) = \( P(\text{survival time} \geq t) \).

The Kaplan-Meier procedure is a nonparametric (distribution-free) method of estimating survival rates at each point in time. Kaplan-Meier is said to be nonparametric since it does not require specific assumptions to be made about the underlying distribution of the survival times.

Note to the Instructor: For additional references that discuss basic biostatistics concepts, see: Lachin (2000) and Rosner (2000).

Instructions:

1. In the space provided below, record the gum flavor longevity data values that were collected in class. Recall, that '40c' denotes right-censoring at 40 minutes.

2. Enter the data into a statistical software package (be careful and make sure that you have correctly entered the data -- i.e., make sure that you have handled the censored data values properly) and generate the Kaplan-Meier life table. Use complete sentences to interpret the life table. Be sure to state the approximate median flavor duration value.

3. Use a statistical software package to generate a plot of the Kaplan-Meier survival function. Use complete sentences to give an interpretation of the survival function.

pdf file for Appendix C.2: Two Sample Censoring Worksheet

C.2 Two Sample

Which Gum Lasts Longer?

Problem:

We want to compare the flavor durations, in minutes, for two brands of chewing gum. However, some of our data values have been censored at 40 minutes. How can we analyze this data?

Background:

We have seen how to analyze data for a single right-censored sample. In order to compare two right-censored samples, we will extend what we have learned for the one sample case.

Instructions:

1. Below are two right-censored samples. These samples give flavor duration values, in minutes, for two brands of sugar-free cinnamon stick gum. Some of the data values have been censored at 40 minutes (denoted by '40c').

\[
\begin{array}{cccccc}
\text{Flavor Duration in Minutes} & \text{40c} & 40 & 22 & 35 & 30 \\
\text{Brand 1 (n=25)} & 8 & 16 & 20 & 28 \\
35 & 7 & 40c & 40c & 35 & 40c & 22 & 40 & 25 \\
40c & 40 & 20 & 20 & 27 & 17 & 26 & 40c & 30 \\
40c & 40c & 7 & 40c & 31 & 35 & 40c & 35 \\
\end{array}
\]

Page 15 of 17
2. You have previously constructed the Kaplan-Meier life table for the Brand 1 flavor duration values. Use a statistical software package to construct the Kaplan-Meier life table for the Brand 2 values. Using complete sentences, compare and contrast the two tables.

3. You have previously constructed the Kaplan-Meier survival function for the Brand 1 flavor duration values. Use a statistical software package to construct and plot Kaplan-Meier survival functions for the flavor duration values of both of the gum brands. Be sure to plot both survival functions on the same graph. Using complete sentences, compare the overall survival rates for the two brands. Does it appear that one of the brands tends to have longer flavor duration values than the other brand?

4. We can use a statistical hypothesis test to compare the survival (or failure) times for two or more samples. Two of the different tests for censored data that are available are the Log Rank test and the Wilcoxon test (Breslow’s generalized Wilcoxon test). Use a statistical software package to perform one of these tests in order to determine if there is a significant difference in the flavor duration values for the two brands.

Appendix D: Ratings Sheet

pdf of Appendix D: Ratings Sheet

Name ________________________________

1. The gum being evaluated (circle one):

2. Evaluate the flavor intensity of the gum (circle one):

    (0 = no intensity, 10 = extreme intensity)

    0 1 2 3 4 5 6 7 8 9 10

3. Evaluate the flavor of the gum (circle one):

    (0 = no flavor, 10 = extremely flavorful)

    0 1 2 3 4 5 6 7 8 9 10

4. The point at which you would discard the gum: (record the length of time, in minutes, you have chewed the gum).

    ___________ minutes

5. At the point at which you would discard the gum, evaluate the texture of the gum (circle one):

    (0 = very soft, pliable, 10 = very hard, rubbery)

    0 1 2 3 4 5 6 7 8 9 10

Appendix E: Two-Way Analysis of Variance Questions Sheet

pdf of Appendix E: Two-Way Analysis of Variance Questions Sheet

1. For each treatment, generate the mean and standard deviation of the texture scores.

2. Generate an ANOVA summary table and an interaction plot. Using the appropriate generated P-value, make a conclusion regarding whether Flavor of gum and Piece of gum interact to affect the mean Texture score. Does your conclusion agree with what your graph shows?

3. If you concluded that Flavor of gum and Piece of gum interact to affect the mean Texture score, use a multiple comparisons procedure to compare all pairs of the treatment means.

4. If you did not conclude that Flavor of gum and Piece of gum interact to affect the mean Texture score, use the appropriate generated P-value to make a conclusion regarding whether the treatment means are equal. If you conclude that the treatment means are not equal, conduct tests of two null hypotheses that the mean texture score is the same at each level of Flavor of gum and at each level of Piece of gum. If the test for either Flavor or Piece of gum is significant, compare the pair of means corresponding to the levels of the significant factor.

Acknowledgments

The authors gratefully acknowledge the helpful comments and suggestions of the editor, the associate editor, and the referees during the preparation of this manuscript.

References

Brandt, L. A. (2001), “FORMULATION CHALLENGE: CONFECTIONARY - A STICKY Situation,” [Online],
www.preparedfoods.com/CDA/ArticleInformation/features/BNP_Features_Item/0,1231,114008,00.html

Cobb, G. (1992), “Teaching Statistics,” in Heeding the Call for Change: Suggestions for Curricular Action, ed. L. Steen, MAA Notes, 22, 3-43.

Collett, D. (1996), Modelling Survival Data in Medical Research, New York: Chapman and Hall.
Gnanadesikan, M., Scheaffer, R., Watkins, A., and Witmer, J. (1997), “An Activity-Based Statistics Course,” *Journal of Statistics Education* [Online], 5(2).

Lachin, J. M. (2000), *Biostatistical Methods: The Assessment of Relative Risks*, New York: John Wiley and Sons.

Lang, T. A. and Secic, M. (1997), *How to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers*, Philadelphia: American College of Physicians.

McClave, J. T. and Sincich, T. (2003), *Statistics*, 9th edition, Upper Saddle River, New Jersey: Prentice Hall.

Rosner, B. (2000), *Fundamentals of Biostatistics*, 5th edition, Pacific Grove, California: Duxbury.

Scheaffer, R. (1996), *Overview for Activity-Based Statistics: Instructor Resources*, New York: Key Curriculum Press; Springer.

Mary Richardson  
Department of Statistics  
Grand Valley State University  
Allendale, MI 49401  
U.S.A.  
richamar@gvsu.edu

Neal Rogness  
Department of Statistics  
Grand Valley State University  
Allendale, MI 49401  
U.S.A.  
rognessn@gvsu.edu

Byron Gajewski  
The University of Kansas Medical Center  
Kansas City, KS 66160  
U.S.A.  
bgajewski@kumc.edu