Water Taste Test Data

M. Leigh Lunsford
Alix D. Dowling Fink
Longwood University

Journal of Statistics Education Volume 18, Number 1 (2010),
www.amstat.org/publications/jse/v18n1/lunsford.pdf

Copyright © 2010 by M. Leigh Lunsford and Alix D. Dowling Fink all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Categorical data, inference for a single proportion, chi-square goodness-of-fit test, chi-square test for independence.

Abstract

In this paper we present data collected from two water taste tests conducted at Longwood University in the fall of 2008. These data are appropriate for performing exploratory data analyses and inference using the test for a single proportion and chi-square tests, including the goodness-of-fit test and the test for independence. The data also provide opportunities for students to consider the cell-count assumptions for chi-square tests, as some tests meet the assumptions and others do not. Finally, the results from the taste tests raise questions about sustainability issues related to consumer preference, bottled water consumption, and its environmental impacts.

1. Introduction

In the fall of 2008 we had the unique opportunity to work with two honors classes on a collaborative research project in which our students conducted water taste tests to answer research questions about preferences for bottled or tap water. The taste tests yielded categorical data appropriate for a test of a single proportion and chi-square goodness-of-fit and independence tests. The data are interesting because they can be used to emphasize the assumptions necessary to conduct these tests, an important pedagogical point when teaching statistics. "As one anthropologist put it, 'I can teach my students how to use chi-square, but I need you to teach them when it is appropriate to use it, and more importantly, when it should not be used at all.'" (Meng 2009)

The data can be used throughout the semester for an introductory level statistics class, beginning with exploratory data analysis, continuing with determination of appropriate statistical procedures for the given research questions, and finally finishing with the interpretation of the results in context. It can also be used for in-class examples or for out-of-class homework or project-type assignments. The data set has utility in more advanced statistics courses as well, and in the spring of 2009 we used it to conduct chi-square tests in our second semester applied statistics class.

Below we describe the motivation for the data collection and the relevance of the data set. In the Experimental Design and Data Collection section we describe the research questions our students posed and their methodology to address those questions. We briefly describe the data set in the Data Description section, and in the Pedagogical Uses section we present ideas for incorporating the data into introductory statistics classes along with some results from the taste tests. These results were interesting, surprising, and at the same time philosophically disappointing to our students. Some of the summary statistics, such as two-way tables, are included for instructors who would like to use the data set but not in its raw format. We also discuss examples from the data for which the assumptions of certain statistical procedures are met and examples for which they are not met. The raw data are more thoroughly described in Appendix A. Although we have cleaned the data for pedagogical purposes, Appendix A does contain some special notes for instructors who would like to use the raw data and perhaps clean it themselves or have their students do it.

Before moving into the detailed description of the experiment, it may be useful to other instructors to understand the context in which these data were collected. This joint honors project brought together students in our Statistical Decision Making (MATH171) and Exploring Science in Our World (GNED261) courses, each of which is a part of our General Education curriculum. As such, most students were first-semester freshmen who were neither science nor mathematics majors. Detailed demographics of the student researchers are provided elsewhere (Fink and Lunsford 2009).

The link between these courses made good sense to us, as we both seek to engage our students in relevant applications of disciplinary content knowledge. The GNED261 course, known to students by its "Power of Water" byline, is a model course for the national Science Education for New Civic Engagements and Responsibilities program (SENCER 2008a, 2008b), and the MATH171 course followed the Guidelines for Assessment and Instruction in Statistics Education (GAISE 2008) and used the Workshop Statistics textbook (Rossman et al. 2008). We built the link between these courses by involving our students in research related to our campus' two-year sustainability theme, thereby challenging them to develop basic research skills within a civically and socially relevant context. Longwood has adopted the widely used Brundtland Commission definition of sustainability: "meeting the needs of the present without compromising the ability of future generations to meet their own needs" (World Commission on Environment and Development 1987). Specifically, we asked our students to focus on the bottled water phenomenon by considering their peers' preference of water types and also researching the costs (economic and environmental) of bottled water consumption. We believe the "green campus" link served to really engage the students in the project and that this

background information may aid in helping other instructors develop similar engagement in their students who also use this data set.

The water taste test blind data is available on the JSE site.

2. Experimental Design and Data Collection

With these issues in mind, students in our MATH171 and GNED261 classes posed the following research questions:

1. Do members of the Longwood community prefer bottled water to tap water?

2. Does brand name affect Longwood community members' preferences for various types of bottled water?

To address the first question the students designed a double-blind water taste test in which subjects were given bottled and tap water and asked to rank their choices from most to least preferred. In order to make the statistical analysis easier, no ties were allowed in the rankings. Each subject was given four identical cups from which to taste: three contained different brands of bottled water (Fiji, Aquafina, and Sam's Choice) and a fourth contained tap water. The subjects did not know which type of water was in each cup. For double-blinding of the experiment the water was poured from four jugs labeled only as A, B, C, and D. The experimenters did not know which label corresponded to which type of water (though this information is provided in Appendix A). Also, the order in which each subject drank the labeled waters was randomized. This test hereafter will be referred to as the "double-blind test."

For the second research question the students designed a water taste test whereby subjects would taste water samples poured from brand-name containers (Fiji, Aquafina, and Sam's Choice) that actually held the same generic store brand water (Wal-Mart Drinking Water, sold in 1-gal jugs). Unfortunately, because the students designed and conducted the experiment, it was not double-blind. Certainly this could have had some effect on the outcome of the experiment. However, multiple students administered the test and they varied the order in which samples were poured. Again subjects were asked to rank their choices from most to least preferred, with ties not allowed. This test hereafter will be referred to as the "deceptive test."

In addition to collecting these preference data for each subject, our students recorded each subject's gender, age, and class ranking (i.e., freshman, sophomore, etc.) as well as what type of water the subject usually drank (i.e., tap, bottled, or filtered) and the subject's favorite brand of bottled water, if any. These additional data helped our students determine if the sample was representative of the Longwood community. They would also allow the students to pursue related research questions such as "Is water preference associated with gender?" or "Is water preference associated with the type of water a person usually drinks?"

Our students conducted their experiments at Longwood's annual Oktoberfest celebration, a venue that afforded them easy access to a large number of research subjects. Both tests were administered in the students' booth, which was one of several dozen at the festival. There were

109 participants in the double-blind test and 104 participants in the deceptive test, each of whom signed an informed consent form prior to participating in the taste tests (a procedure approved in advance by our campus' Human and Animal Subjects Research Review Committee). Most of the participants took both tests. In relying on volunteers at this public event, our students realized that they may not have obtained a random sample from the Longwood community. These issues are discussed in more detail in the Pedagogical Uses section of this paper.

For more details on the coordination and planning of this collaborative student research project, please see Fink and Lunsford (2009).

3. Description of the Data

Two files contain the raw data: Blind_Test.dat and Deceptive_Test.dat. Both of these tab-delimited data files have five common variables collected for each subject who took the test. These are the subject's gender, age (in years), academic class ranking (i.e., freshman, sophomore, junior, senior, or other), the type of water the subject usually drank (i.e. tap, bottled, or filtered), and the subject's favorite brand of bottled water, if any. For each test a ranked preference of water choices was recorded, with ties not allowed. For the double-blind test a string of four letters was recorded to indicate the subject's preference from most to least preferred. For instance, the string BCDA would indicate the subject preferred brand B the most, followed by C, D, and finally A, the least preferred. There are four additional columns of data for each taste test record, one for each level of preference. For the above example, First would be B, Second would be C, Third would be D, and Fourth would be A. Similarly a preference string was also recorded for each subject for the deceptive test, in which the choices were A, F, and S for Aquafina, Fiji, and Sam's Choice Brand, respectively. Thus a string of FAS indicated the subject preferred the taste of the water labeled Fiji the most, followed by Aquafina and then Sam's Choice, the least preferred. Again we recorded additional columns for each preference, such that for this example First would be F, Second would be A, and Third would be S.

More detailed information about the data, including special notes on dealing with data entry issues and data cleaning for pedagogical purposes, appears in Appendix A.

4. Pedagogical Uses

In this section we provide several examples of how the data can be used to illustrate various statistical concepts. We have chosen these examples because we believe they are interesting and can be used in a variety of ways (i.e., as in-class examples, homework problems, or part of a project). We list the examples by statistical content and for each example give pedagogical notes for instructors.

4.1 Data Collection Issues

From a pedagogical point of view we believe the method of data collection is important to discuss with students as it determines what types of conclusions one can draw (i.e., cause and effect) and to which population these conclusions may apply. As mentioned above, our students understood that our reliance on volunteer taste testers may have resulted in our not obtaining a

random sample of the Longwood community. Our data were collected at Oktoberfest, which is an event attended primarily by students involved in extracurricular organizations (such as fraternities, sororities, and clubs), some alumni, a few faculty and staff advisors, and the occasional child. Thus, we had our students conduct some exploratory data analysis to see if their non-random sample of subjects was representative of the larger Longwood population.

To accomplish this, the students considered the variables age, gender, and class ranking. While some faculty, staff, and their children participated in the taste tests, they were a minority, and as a convention their class ranking was recorded as "Other." After examining these data, we felt comfortable enough to extend the scope of our conclusions to the population of the Longwood community that is involved in Oktoberfest. Thus, when we refer to the "Longwood community" throughout the remainder of this paper, we mean the population of the Longwood community that is involved in Oktoberfest. It is an important pedagogical point that students who use this data set do not extend the scope of their conclusions beyond this population.

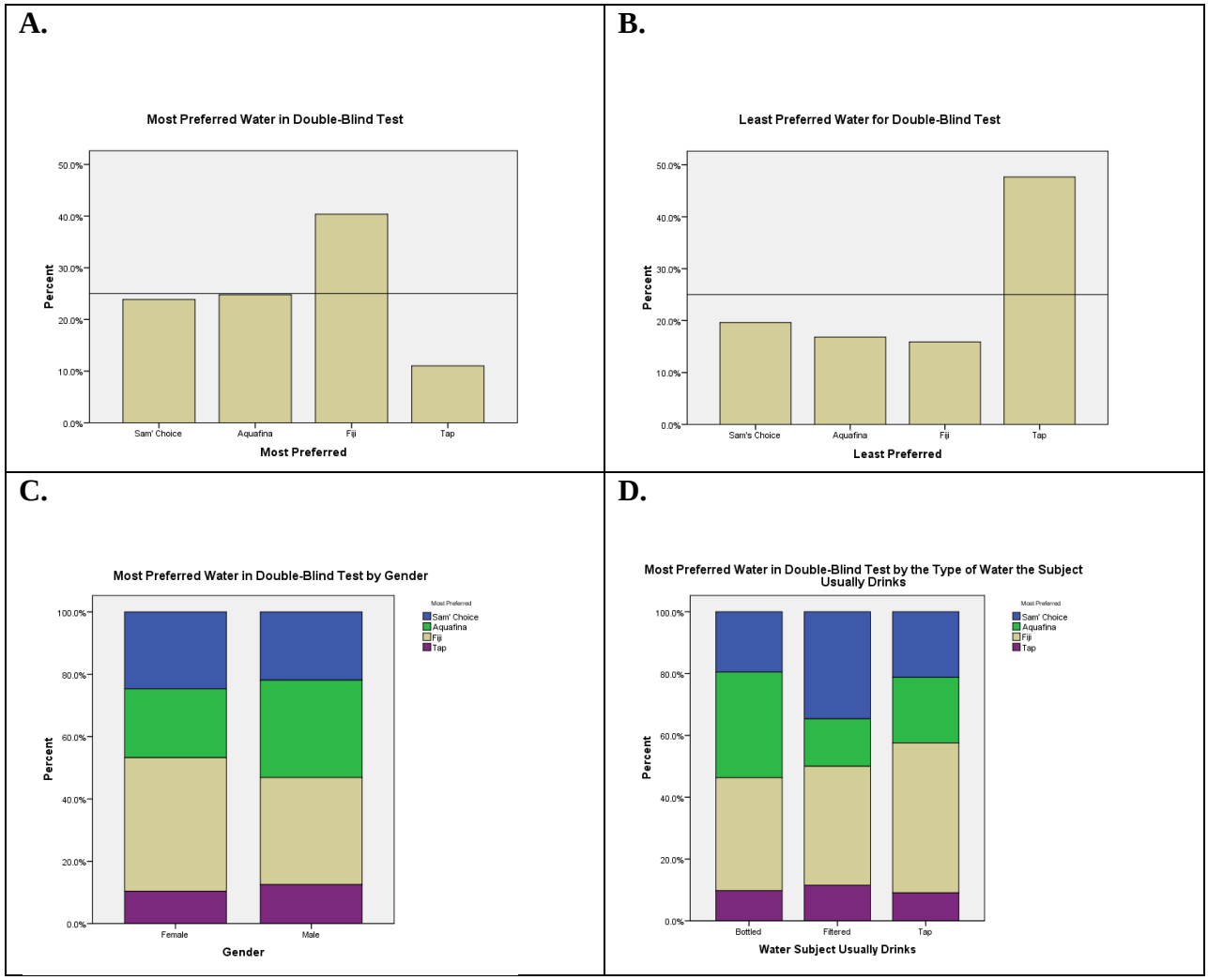
4.2 Data Exploration

Below we provide some graphical displays of the data. An overarching theme in these exploratory graphs is that there is always random variation in data. The discussion of these graphs should get students interested in determining whether they are observing results that occur just due to random variation or results that are unusual and go beyond the expected random variation. We like to bill these discussions as exciting "previews of coming attractions," in that students will soon be learning how to determine numerically if the random variation is more than what is expected.

In Figure 1 below we provide four graphs generated from the double-blind test data set. Graphs A and B show the most preferred and least preferred type of water, respectively. Of the 109 participants in the double-blind test, two did not have the least preferred water recorded and thus we did not use those entries. The horizontal line in each of these graphs corresponds to no preference among the water types (i.e., they are all equally preferred). There are clear patterns here for students to discuss.

We would first recommend plotting these graphs without the horizontal line and discussing with students how they would expect the graphs to look if there was no preference for the various water types. Once students understand that the distributions should be uniform if there are no preferences, then they might want to add the horizontal line to the graph to help them explore the idea of random variation and its ramifications. Specifically, students should discuss if the deviations from the horizontal line are something they might observe just due to random variation or if the variation is beyond what they would expect if there was no preference. In either case there should be clear discussion of the corresponding conclusions they can draw. By the end of this discussion students should be able, without doing any inferential statistics, to make a plausible argument as to whether these data provide some evidence to suggest that bottled water is most preferred by the subjects (in Graph A) or tap water is least preferred (in Graph B).

Figure 1. Graphs from the Double-Blind Test



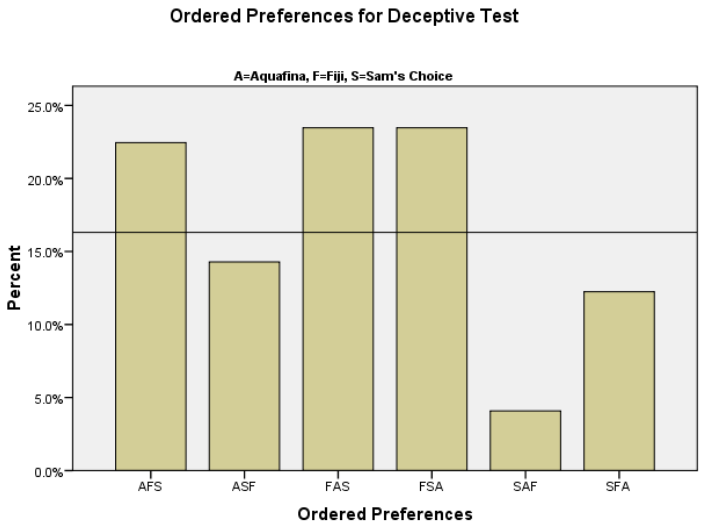
Graphs C and D in Figure 1 are stacked bar charts that examine the relationship between the most preferred type of water and gender and type of water usually consumed, respectively. We think it is a good idea for students to generate these charts from either raw or summarized data (given in the sections below). Again there are many interesting pedagogical points to discuss with students, including how the percentages on the bars are determined and how students would expect these graphs to look if gender or type of water usually consumed have no association with the subject's most preferred type of water in the taste test. For example, consider graph C: if there is no association between these two variables, one would expect the amount of blue (i.e., percent who chose Sam's Choice as their most preferred water type) to be roughly the same for males and females. In this graph they certainly appear to be close to the same percent (in this case 19 out of 77 or 24.7% for the females and 7 out of 32 or 21.9% of the males). This should also be true for each of the other three water types. Clearly these bars have some deviation from equal amounts of each color in the bars (i.e., it appears that males were more likely than females to choose Aquafina as their most preferred water). A pedagogical question to ask and discuss is whether this deviation is something students would expect to see even if there is no association

between the most preferred water and gender. Or, is this deviation enough to suggest that there is an association between gender and most preferred water type? Another important pedagogical point to make is that the percents for each type of water need not be uniform for each gender (i.e., we need not see each color represent about 25% of each bar).

In Figure 2 below we provide a graph generated from the deceptive test data. There were six possibilities for the ordered preferences from the deceptive test: AFS, ASF, FAS, FSA, SAF, and SFA. Recall that AFS, for example, corresponds to a subject choosing Aquafina as her most preferred, then Fiji, and lastly Sam's Choice. Though we had 104 subjects complete this taste test, 5 of these subjects were not deceived by our experiment (the same water was in all of the labeled bottles!) and refused or were unable to give a preferred ordering. We also had one subject who said Fiji only. We removed these 6 subjects to make this graph and also from further analysis because we wanted to conduct simple statistical analyses and thus were only interested in those who did provide a preferred ordering. We do retain these subjects in the raw data set, thus providing an instructor who uses these data an opportunity to discuss when one disregards certain data points and how to go about doing that in a particular software package.

The horizontal line in Figure 2 corresponds to the percents we would expect to see if there were no label effect (i.e., the six orderings should be equally likely). There are many points for discussion here. First, as mentioned above, we would not initially include the horizontal equally preferred line on the graph but instead lead students to discover it. Second, students should notice the trend that, despite the fact that the same water was in all three bottles, Sam's Choice was not chosen as the most preferred water nearly as often as either Fiji or Aquafina. Another discussion point again revolves around chance deviation and its ramifications. By the end of this discussion your students should be able, without doing any inferential statistics, to conjecture whether these results are something we would expect to see via random deviation and thus suggest no label effect or if the deviation from uniformity is enough to suggest a label effect (i.e., subjects choosing preferred water based on label not on taste).

Figure 2. Graph from the Deceptive Test



4.3 Test of a Single Proportion

One of the first inferential techniques learned in introductory statistics is the test for a single population proportion. The preference data collected by our students were interesting in that there are several approaches they could have taken to address the proposed research questions. To address the first research question, "Do members of the Longwood community prefer bottled water to tap water?" students could conduct either a test for a single proportion or a chi-square goodness-of-fit test. One way to approach the test for a single proportion is to reason that if there were no overall preference for tap or bottled water (i.e., either more or less preferred) then we would expect about 25% of the population to choose tap water as the most preferred water since it was one of the four choices. We note that in the context of the sustainability theme in our classes, and related to our students' motivation to show that there was no overall difference in population preference between tap and bottled water, we chose the two-sided alternative hypothesis. Thus, we tested the hypotheses:

$$H_0 : \pi_T = 0.25$$
$$H_a : \pi_T \neq 0.25$$

where π_T is the proportion of the Longwood community that would choose tap water as its most preferred water.

It is also important to emphasize that the conditions for using this test, which can be found in most introductory statistics books, are met (e.g., see Moore 2010, p513, and Rossman, et al. 2008, p346): a simple random sample has been drawn from the population and both the expected number of successes and failures are greater than 10 (i.e., $n\pi_0 > 10$ and $n(1-\pi_0) > 10$). We have discussed the issue of a simple random sample above and since $\pi_0 = 0.25$ and $n = 109$ the expected count condition is clearly met.

The data show that 12 out of 109 participants chose tap water as their first choice, which yields a sample proportion of 0.11 and a *p*-value less than 0.001 (p = 0.00074). Thus it appears that there is some overall preference being observed. From Graph A and B in Figure 1, students might conjecture that there appears to be a preference for bottled water. Pedagogically, it is important for students to understand why $\pi_T = 0.25$ in the null hypothesis and that they cannot conclude there is a preference for bottled water because they are doing a two-sided test.

Finally, if prior to seeing the graphs in Figure 1 the students conjectured that bottled water would be preferred to tap water, then it would be plausible to use the one-sided alternative hypothesis $\pi_T < 0.25$ (p = 0.00037). Here, because of the one-sided alternative, students can conclude there is an overall population preference for bottled water over tap.

4.3.1 Additional Teaching Notes and Comments

We note that the hypotheses above refer to the proportion of the Longwood community and not to preferences of individuals in the Longwood community (i.e., we are not hypothesizing that

individuals have no preference). Indeed we did not allow for individuals to have the response of no preference since we forced them to rank their preferences. Thus our reference class for our assertion is the Longwood community and we are testing about an overall population proportion, not individual proportions. This distinction, while subtle, is important.

4.4 Chi-Square Goodness of Fit Test

The single proportion test, discussed above in section 4.3, is limited in that it we can only draw a conclusion about the proportion of the Longwood community that would choose tap water as a first preference. This presents an instructor with a nice segue into the chi-square goodness-of-fit test. If one is interested in a test that gives more information about the distribution of first preference among all the types of water used in the taste test, then the chi-square goodness-of-fit test can be used to determine if the first choice has a particular distribution. Again assuming there are no overall preferences, the four different types of water should be equally likely to be chosen. If it has not already been done, this is a good time to add the horizontal line at 25% to Graph A in Figure 1 and discuss the variation that students see from this line. This should help students understand the hypotheses:

$$H_0 : \pi_T = \pi_A = \pi_F = \pi_S = 0.25$$
$$H_a : At\ least\ one\ \pi_i\ is\ different\ from\ 0.25.$$

where π_T, π_A, π_F, and π_S are the proportions of the Longwood community that would choose tap, Aquafina, Fiji, and Sam's Choice water, respectively, as the most preferred water.

Again, it is important to check the assumptions of this test: a simple random sample has been drawn from the population and the expected counts in each cell are at least 5. We have addressed the issue of a simple random sample above and clearly our expected counts will be greater than 5 since we have 109 subjects.

The data (see Table 1 below) give a chi-square test statistic of 18.89 with 3 degrees of freedom and a *p*-value of 0.000288.

Table 1. Data for the Chi-Square Goodness-of-Fit Test for First Preference in the Double-Blind Test

Type of Water	**Tap**	**Aquafina**	**Fiji**	**Sam's Choice**	**Total**
Observed	12	27	44	26	109
Expected	27.25	27.25	27.25	27.25	109
Contribution to Test Statistic	8.534	0.002	10.296	0.057	18.889

Because this is a significant result, it is worthwhile to discuss with students what features of the distribution contribute most to the variation away from uniformity. This is also an ideal time to revisit Graph A in Figure 1 to visualize the variation. Most of the contribution to the chi-square test statistic comes from Fiji (10.30) and Tap (8.53), with Fiji being chosen as first preference more than 25% of the time and tap being chosen less than 25% of the time. Finally, it is

important that students correctly interpret these results. In particular, the test reveals that the Longwood community appears to not be choosing their first preference of water randomly in the taste test, thus there are some preferences. From the graphical results we can conjecture that bottled water is preferred to tap.

Students also were interested in seeing which type of water was the least preferred in the taste test. Graph B in Figure 1 seems to indicate that tap water was overwhelmingly chosen as the least preferred water. Using the same logic as the above goodness-of-fit test, students should be able to derive the hypotheses:

$$H_0 : \pi_T = \pi_A = \pi_F = \pi_S = 0.25$$
$$H_a : At\ least\ one\ \pi_i\ is\ different\ from\ 0.25.$$

where π_T, π_A, π_F, and π_S are the proportions of the Longwood community that would choose tap, Aquafina, Fiji, and Sam's Choice water, respectively, as their least preferred water. The data (see Table 2 below) give a chi-square test statistic of 29.64 with 3 degrees of freedom and a *p*-value of 0.000002.

Table 2. Data for the Chi-Square Goodness-of-Fit Test for the Least Preferred Type of Water in the Double-Blind Test

Type of Water	**Tap**	**Aquafina**	**Fiji**	**Sam's Choice**	**Total**
Observed	51	18	17	21	107
Expected	26.75	26.75	26.75	26.75	107
Contribution to Test Statistic	21.984	2.862	3.554	1.236	29.636

This is a very significant result with most of the contribution to the value of the chi-square test statistic coming from the tap water. We think it is an excellent idea to revisit Graph B in Figure 1 to visualize the connection between the chi-square goodness-of-fit test and the graphical representation of the data. Also, as mentioned above, we think it is important for students to interpret these results in context. That is, the Longwood community is not likely to equally choose among the four types of water as its least preferred.

It is worth noting that our students found these results to be a bit disturbing because of the background knowledge they had obtained regarding bottled water and its relevance to our campus' sustainability theme. At the risk of seeming insensitive, we should note that we were somewhat pleased with that outcome. We sought to provide an authentic research experience for our students (though modest in scope), and results that do not fit one's hopes are certainly a reality of research. Additionally, these results, which students viewed as less than ideal, prompted a much richer discussion about the significant challenges of true sustainability than had the actual results matched their desired outcomes.

For the deceptive test we could perform the same type of chi-square goodness-of-fit tests on first and least preferences as discussed for the double-blind test. However, because we only have three choices of water, we have six possible ordered preferences for the taste test. Figure 2

shows the distribution of these six ordered preferences. Given that we have 98 subjects with complete data, our expected count for each ordered preference is 16.333 (16.667%) under the assumption that all of the preferences should be equally likely (since the same type of water was in the bottles and only the labels were different). Notice that 16.7% is where the horizontal line is drawn in Figure 2. Thus our sample meets the cell-count assumption for conducting a chi-square goodness-of-fit test on the ordered preferences. Assuming there is no label effect, we have the hypotheses:

$$H_0 : \pi_{AFS} = \pi_{ASF} = \pi_{FAS} = \pi_{FSA} = \pi_{SAF} = \pi_{SFA} = 1/6$$
$$H_a : \textit{At least one } \pi_i \textit{ is different from } 1/6.$$

where π_{AFS} is the proportion of the Longwood community that would chose Aquafina, Fiji, and Sam's Choice as their ordered preference from most to least preferred, and analogously for the others. The data (see Table 3 below) give a chi-square test statistic of 18.204 with 5 degrees of freedom and a *p*-value of 0.0027.

Table 3. Data for the Chi-Square Goodness-of-fit Test for Ordered Preferences for the Deceptive Test

Ordered Preference	AFS	ASF	FAS	FSA	SAF	SFA	Total
Observed Counts	22	14	23	23	4	12	98
Expected Counts	16.333	16.333	16.333	16.333	16.333	16.333	98
Contribution to Test Statistic	1.966	0.333	2.721	2.721	9.313	1.150	18.204

Again, since this is a significant result it is good to discuss with students what features of the distribution contribute the most to the variation away from uniformity and how these features manifest themselves in Figure 2. Most of the contribution to the chi-square test statistic comes from the SAF value (9.313 of 18.204 in the table above). Finally, it is important that students correctly interpret these results. In particular, the test reveals that the label on the bottle has some effect on the Longwood community's preference choices. We think it is interesting to further discuss this result by having students think about and conjecture as to why this may be the case, though it is important that students understand that this test does not provide any reasons for the results or "proof" for their conclusion.

4.4.1 Additional Teaching Notes and Comments

Because the samples for these tests and those in section 4.3 are essentially the same, conducting the tests brings up the important issue of multiple testing. This is not typically covered in most introductory statistics books, but could be mentioned in the introductory class, and should be discussed in more advanced classes when using this or any other data set for multiple testing.

Also, one item we overlooked was the possible effect of the order in which subjects tasted water, in either experiment, on preferences. Collecting these data could be easily incorporated into these experiments but, unfortunately, we did not do this.

Finally, students might also wonder why we did not conduct a chi-square goodness-of-fit test on the ordered preferences for the double-blind test. There we had four choices and hence 24 ordered preferences. With only 109 subjects, this leads to an expected count of 4.54 in every cell, which does not meet one of the standard criteria for conducting a chi-square goodness-of-fit test (i.e., each expected cell count is 5 or more).

4.5 Chi-Square Test for Independence

From the chi-square goodness-of-fit test for most preferred water for the double-blind test above, we observed that the Longwood community appeared to not be choosing their first preference randomly. In particular we conjectured there was a preference for bottled water over tap water. Now we ask if these preferences are dependent on (or associated with) gender. This leads to the chi-square test for independence. We will test the hypotheses:

$$H_0: \textit{Gender and First Preference are independent}$$
$$H_a: \textit{Gender and First Preference are not independent}$$

Again the assumptions for running this test need to be addressed. Specifically students should note that one of the cells (Tap and Male) has an expected count less than five (see Table 4 below). However, because this is only 12.5% of the cells, we can safely use the test. Many texts state that if at most 20% of the expected frequencies are less than 5, and all are greater than or equal to 1, it is acceptable to conduct the test (see Weiss 2005, p674).

The data (see Table 4 below) give a chi-square test statistic of 1.336 with 3 degrees of freedom and a *p*-value of 0.721. Here we have a result that is not significant. At this point it might be worthwhile to revisit the stacked bar graph in Figure 1 and see that, while the proportions in each gender bar are not exactly the same, they are similar enough to be due to random variation. Thus we can conclude that the data provide no evidence that the preferences we have observed are associated with gender.

Table 4. Two-Way Table of Gender and Most Preferred Type of Water in the Double-Blind Test. Expected counts are given in parentheses below the observed counts.

	First Preference Observed (Expected)				
Gender	Tap	Aquafina	Fiji	Sam's Choice	Total
Female	8 (8.5)	17 (19.1)	33 (31.1)	19 (18.4)	77
Male	4 (3.5)	10 (7.9)	11 (12.9)	7 (7.6)	32
Total	12	27	44	26	109

Our students also were interested in determining if there was an association between the type of water a person usually consumes and choice of most preferred water in the double-blind test. Below in Table 5 we see a two-way table of these data. Note that there are only 100 subjects here because we did not include 9 of the subjects whose type of water they usually consumed was classified as NA (see the Special Notes in Appendix A).

The results presented in Table 5 raise an important issue because three of the cells, or 25% of the expected counts, are less than five, and thus the cell-count requirements to use the chi-square test of independence are not met. This could possibly be resolved by combining the Filtered and Tap categories or combining the Bottled and Filtered categories. In both cases you would have only one cell of eight (12.5%) with an expected count less than five, and in both cases you do not get a significant result.

Table 5. Two-Way Table of Water Usually Consumed and Most Preferred Type of Water in the Double-Blind Test. Expected counts are given in parentheses below the observed counts.

	First Preference Observed (Expected)				
Type of Water Usually Consumed	Tap	Aquafina	Fiji	Sam's Choice	Total
Bottled	4 (4.1)	14 (10.2)	15 (16.8)	8 (9.8)	41
Filtered	3 (2.6)	4 (6.5)	10 (10.7)	9 (6.2)	26
Tap	3 (3.3)	7 (8.2)	16 (13.5)	7 (7.9)	33
Total	10	25	41	24	100

4.5.1 Additional Teaching Notes and Comments

We note that although we did not include the 9 subjects classified as NA in the data for Table 5, they are included in the data files with this article. As mentioned previously, instructors can clean these data for their students or use this as an opportunity to show students how to do this with a particular statistics package.

5. Conclusion

In this paper we have presented data collected from two water taste tests conducted at Longwood University in the fall of 2008 along with suggested uses of the data for introductory level statistics classes. The data allow for the illustration of important pedagogical points when analyzing categorical data and conducting tests of inference. We hope that instructors will find these data useful for their introductory level statistics classes.

The results are also interesting in the context of sustainability, a growing issue on many college campuses and at the national level, and can be used in considering broad questions about

sustainability and the specific issue of bottled water. For instance, can we afford to ship small volumes of water around the country (and the world) when we live in a country that readily supplies its citizens with clean drinking water? Important discussions can develop from the examination of the data. Should preference and convenience trump issues of environmental impact? For example, even if students prefer bottled water to tap, should they consider switching to a more sustainable option? We think the civic and social issues related to these data provide a very engaging context in which students can learn basic statistical concepts.

Acknowledgments

The authors wish to thank their academic departments, the Cormier Honors College for Citizen Scholars, the students in GNED261 and MATH171, and the many volunteers who assisted with this research. This manuscript benefited from thoughtful reviews by Marcus Pendergrass and the *Journal of Statistics Education*'s Datasets and Stories referees.

Appendix A

Key to Variables in Water Taste Test Data Files

Two tab-delimited data files are given: Blind_Test.dat and Deceptive_Test.dat. A header line that contains the variable names is included in each data file. Both data files have five common variables collected for each subject who took the test. In Table 6 below we describe the common variables included in both files in the order they occur. In Table 7 we describe the additional variables recorded for the double-blind test, and in Table 8 we describe the additional variables recorded for the deceptive test. Following the tables we have a few special notes about how we resolved certain data-recording problems and how we cleaned the data for pedagogical purposes.

Table 6. Variables Common to Both Data Files

Variable	Label in Header Line	Description	Values
1	Gender	Gender of subject	M or F for Male or Female, respectively
2	Age	Age of subject	Integer values
3	Class	Academic class ranking of subject	F, SO, J, SR, O for freshman, sophomore, junior, senior, and other, respectively
4	UsuallyDrink	Type of water the subject usually consumes	B, F, T, NA for bottled, filtered, tap, and not applicable, respectively. See the Special Notes below regarding the NA values.
5	FavBotWatBrand	Favorite bottled water brand	Bottled water brands spelled with all capital letters (e.g., AQUAFINA, DEER PARK, NONE). Note that some have blanks in them so be sure to use tab-delimited formatting.

Table 7. Additional Variables Recorded for the Double-Blind Test

Variable	**Label in Header Line**	**Description**	**Values**
6	Preference	List of ordered preferences from the double-blind test	Four-letter strings containing the letters A, B, C, and D. For example if Preference = BCAD, then the subject chose water type B as his first preference, C as his second, A as his third, and D as his fourth, or least preferred, water type. For decoding purposes A, B, C, and D correspond to Sam's Choice, Aquafina, Fiji, and Tap water, respectively.
7	First	The water type chosen as the first preference	A, B, C, or D. Note that if Preference = BCAD then First = B. The same correspondence between the letters and the brands is used as the Preference variable.
8	Second	The water type chosen as the second preference	A, B, C, or D. Note that if Preference = BCAD then Second = C. The same correspondence between the letters and the brands is used as the Preference variable.
9	Third	The water type chosen as the third preference	A, B, C, or D. Note that if Preference = BCAD then Third = A. The same correspondence between the letters and the brands is used as the Preference variable.
10	Fourth	The water type chosen as the fourth preference or least preferred	A, B, C, or D. Note that if Preference = BCAD then Fourth = D. The same correspondence between the letters and the brands is used as the Preference variable.

Table 8. Additional Variables Recorded for the Deceptive Test

Variable	Label in Header Line	Description	Values
6	Preference	List of ordered preferences from the deceptive test	Three-letter strings containing the letters A, F, and S (and also the string NONE), where A, S, and F, represent Aquafina, Sam's Choice, and Fiji brand water, respectively. For example, if Preference = SAF, then the subject chose Sam's Choice as her first preference, Aquafina as her second, and Fiji as her third, or least preferred, water type.
7	First	The water type chosen as the first preference	A, S, F, or blank Note that if Preference = SAF then First = S. If Preference = NONE then First is blank.
8	Second	The water type chosen as the second preference	A, S, F, or blank Note that if Preference = SAF then Second = A. If Preference = NONE then Second is blank.
9	Third	The water type chosen as the third preference or least preferred	A, S, F, or blank Note that if Preference = SAF then Third = F. If Preference = NONE then Third is blank.

Special Notes:

As with any real-life data collection, we encountered problems with how data were recorded on the data sheets in the field and subsequently needed to make decisions regarding how these data would be recorded electronically. Below we detail the problems and our resolutions to the problems. Most of these changes were made to make the data usable for pedagogical purposes. We would be happy to provide copies of the completed data collection sheets to any interested parties.

In the Class column, some of our student experimenters recorded a generic S instead of SR or SO for the academic class of the subject. By examining the Age variable we believe we were able to make an educated guess to which academic class (either SR or SO) those subjects belonged, as most of our students are of traditional collage age. We also had four instances of age being recorded as 21+. In these cases we just made the age 21. Lastly, there was one age that was recorded on a data sheet as 2, which we were certain was incorrect. However, since most of our analyses do not involve age, we left that number as it was recorded. If using the variable Age in any analysis, one might want to consider eliminating that particular line of data.

For subjects who gave more than one answer to the question of the type/brand of water usually consumed, we used the first type cited by the subject. For instance, on our data sheets from the field we have a few entries recorded as B/F, B & T, T/F, T/B, etc. These entries were coded as B, B, T, and T, respectively, in the data files associated with this paper. There were only six of these types of entries on the field data sheets for the deceptive test and seven for the double-blind

test. On the deceptive test we also had one subject who apparently did not answer this question so we entered NA for the type of water she usually consumed. For the double-blind test we also classified the following answers as NA: NA (n = 1), W (1), All (3), Any (1), None (2), and blank (1).

In the deceptive test, one subject cited "AQUAFINA or SAM'S" for her favorite brand of bottled water. Following the convention above, since the subject said AQUAFINA first, we listed that as her favorite (last line of data set). Also, on all sheets, if a blank, NA, or N/A was recorded for the variable Favorite Brand of Water, that record was entered into our Excel spreadsheet as "NONE." We also endeavored to type in all brand names with all capital letters.

For the subjects who could not decide on a preference for either test, we listed NONE in the ordered preference field and left the remaining fields (First, Second, etc.) blank. In the deceptive test we did have one subject who could only give her most preferred. For this subject we recorded F in the ordered preference field, F in the First field, and left the remaining fields blank.

In the double-blind test we had a preference recorded as CABC (line 7 of the raw data file). We decided to leave the ordered preference as is, place a C for the First variable but leave the remaining individual preferences blank. On line 32 we had a preference recorded as CADA on the data sheet. We again decided to leave the ordered preference as is, place a C for the First variable but leave the remaining individual preferences blank. Note that if an analysis were done with the string of ordered preferences, then these two data points should be removed.

The water taste test blind data is available on the JSE site.

References

Fink A. D. and M. L. Lunsford. 2009. "Bridging the Divides: Using a Collaborative Honors Research Experience to Link Academic Learning to Civic Issues," *Honors in Practice*, National Collegiate Honors Council, Volume 5, 97-113.

GAISE. 2008. GAISE College Report. http://www.amstat.org/education/gaise/GAISECollege.htm. Accessed 17 Dec 2008.

Meng, X. L. 2009. "Desired and Feared – What Do We Do Now and Over the Next 50 Years?" *The American Statistician*, Vol. 63, No. 3, 202-210.

Moore, D. S., 2010. *The Basic Practice of Statistics*. Fifth edition. W. H. Freeman and Company, New York, NY, USA.

Rossman, A. J., B. L. Chance, and J. Barr von Oehsen. 2008. *Workshop statistics: discovery with data and the graphing calculator*. Third edition. Key College Publishing, Emeryville, California, USA.

SENCER. 2008a. Program home page. http://www.sencer.net/. Accessed 16 Dec 2008.

SENCER. 2008b. SENCER Ideals. <http://www.sencer.net/About/pdfs/SENCERIdeals.pdf>. Accessed 16 Dec 2008.

Weiss, N. A. 2005. *Introductory Statistics*. Seventh Edition. Pearson/Addison Wesley, Boston, USA.

World Commission on Environment and Development. 1987. Our Common Future. Oxford: Oxford University Press.

M. Leigh Lunsford
Department of Mathematics and Computer Science
Longwood University
Farmville, VA 23909
lunsfordml@longwood.edu

Alix D. Dowling Fink
Department of Biological and Environmental Sciences
Longwood University
Farmville, VA 23909
finkad@longwood.edu