

Chocolate Chip Cookies as a Teaching Aid

Herbert K. H. LEE

Getting and retaining the attention of students in an introductory statistics course can be a challenge, and poor motivation or outright fear of mathematical concepts can hinder learning. By using an example as familiar and comforting as chocolate chip cookies, the instructor can make a variety of statistical concepts come to life for the students, greatly enhancing learning.

KEY WORDS: Active learning; Teaching statistics; Variability.

1. INTRODUCTION

Teaching introductory statistics at a university can be a major challenge because most of the students in the classroom do not want to be there, and are taking the course only because it is a requirement for their major or it fulfills a distributional requirement. Many students have already heard stories from their friends or even their parents about how hard and unpleasant statistics will be. The challenge is amplified when teaching in a larger lecture format, because of the limited opportunities for personal contact. So it is quite helpful to convince the students to buy in to the course, to get their interest and to connect the course to concepts that they are comfortable with (Singer and Willett 1990; Snee 1993). The use of hands-on activities to facilitate active learning has been well established (Garfield and Ahlgren 1988; Gnanadesikan et al. 1997).

Many introductory textbooks use examples from sports, biology, and economics in an attempt to relate the subject matter to areas of interest to the students, and discussions of the benefits are present in the literature (e.g., Kvam and Sokol 2004). However, I have found that none of these examples are of interest to the whole class, often leaving one-third to one-half of the class uninterested in the examples and still uninterested in the course. Even worse, there can be gender biases in the relative interest in these examples. However, I have found that almost all students like chocolate chip cookies. Other instructors have also documented success in using snack food to motivate statistical education, including M&M's (Dyck and Gee 1998), Reese's Pieces (Rossman and Chance 1999), Hershey's Kisses (Richardson and Haller 2002), and chewing gum (Richardson et al. 2005), although these tend to occur as isolated examples of specific concepts in a course. From a pedagogical standpoint, cookies can go beyond being a gimmick in a single class to becoming a recurring theme that can demonstrate many of the key

concepts in a basic statistics course, allowing students to better see the connections between the different topics. The rest of this article will discuss how the following topics can all be related to cookies:

- variability
- inter-rater agreement and measurement error
- exploratory data analysis: displays of data, outliers
- the Poisson distribution, empirical distributions, extreme values
- sampling distributions
- hypothesis testing: one-sample t tests, two-sample t tests, and analysis of variance
- Bayesian statistics: prior elicitation and prior sensitivity

2. BRINGING COOKIES INTO THE CLASSROOM

By connecting statistical concepts directly to cookies, I have been able to gain the interest of students, even in larger classes. Once the students have found a reason to pay attention, subsequent teaching is much easier. In particular, what is needed is an eye-opening example that the students can relate to. What better than chocolate chip cookies? Consider a bag of mass-produced cookies, such as the brand-name Chips Ahoy or Keebler, or your favorite local store brand. These cookies are all produced to fairly high quality control standards. They are all about the same size. Most students expect that the cookies will all have about the same number of chips. That's what mass-production is all about, right?

2.1 Exploring Variability

Most statistics textbooks have a chapter on exploratory data analysis and graphical displays near the beginning of the book. This is a perfect time to bring in the cookies. I bring in enough cookies for everyone. We discuss the cookies and discuss variability, but the students consistently underestimate the variability. I then pass around the cookies, telling them to each take one cookie and count the chips. Of course, you cannot see all of the chips from the outside, so you need to eat the cookie to find all of the chips.

I have found that the students suddenly become engaged in the course. Almost everyone likes eating cookies. After they have had a chance to count, I have them tell me how many they found and I enter the numbers into my laptop. The students at first tend to think that this is just going to be a long boring process in a large class, but they are soon surprised to hear the numbers coming from their classmates. The variability is far beyond what they expect. For example, in one class we counted

Herbert Lee is Associate Professor, Department of Applied Mathematics and Statistics, University of California, Santa Cruz, 95064 (E-mail: herbie@ams.ucsc.edu). The author thanks Michele Di Pietro and the editor for their helpful comments which led to substantial improvements.

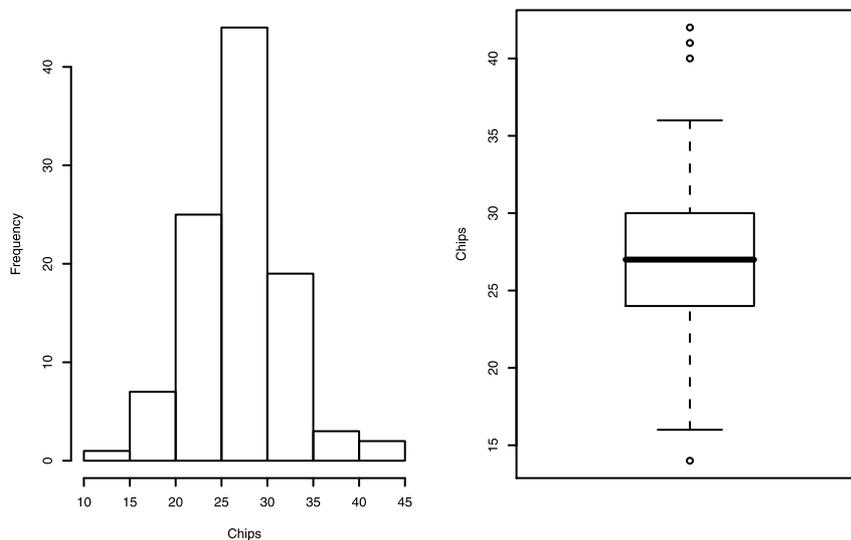


Figure 1. Histogram and boxplot of the number of chips in each of 101 cookies.

101 cookies with the number of chips ranging from 14 to 42 (see Figure 1). Until you stop and think about it, that seems like an awful lot of variability. But if you consider a large batch of well-mixed dough, the number of chips per cookie should be approximately Poisson, so this amount of variability turns out to be reasonable, as further discussed in Section 2.4.

This is a paradigm-shifting moment for most students. They think they understand cookies, yet suddenly realize that there is far more variability than they expected. They may call into question previous cookie-eating experiences. They may wonder if their friends have been getting more chips than they have. But they suddenly have a very tangible example of how variability can affect them personally. And now they have a reason to care about their statistics class. As documented by Sowe (2001), such a “striking demonstration” provokes excitement in the topic and facilitates learning. At this point I can impress upon them that statistics is the study of randomness and variability, and set the stage for many other topics in the course.

2.2 Inter-rater Agreement and Measurement Error

During the counting of the chips, students often ask relevant questions. Two related categories of questions are those about inter-rater agreement and about measurement error. Because of the manufacturing processes, many cookies actually contain one or more broken chips, so students frequently ask “what counts as a chip?” We can discuss how two people might reasonably arrive at different chip counts because of how they count the fractions of chips. If time permits, we can have a class discussion of the counting process, and how different counting techniques can lead to different results. In particular, some students will take larger or smaller bites out of the cookies, leading to different probabilities of missing chips. True to stereotype, I have found engineering students to be more methodical in counting, sometimes going to the extreme of first pulverizing the cookie and separating it into dough bits and chips to ensure a proper count. Thus there are a number of aspects of the process which easily lead into general learning about inter-rater variability.

In an introductory-level course, the discussion of measurement error would typically be kept to a conceptual level, focusing on the ways that people might miss chips, or assumptions made about fractions of chips with which other people might disagree. In a higher-level course, we can have deeper discussions about which probability distribution might be used to model the measurement error. Although the normal distribution is the typical one for error, here we might want a discrete distribution if we are considering only whole numbers of chips, and we expect that people are more likely to fail to see chips (undercount) than to double-count (overcount). Thus we might want a skewed and/or truncated distribution. Prodding the students through a discussion of these concepts can be enlightening for them in learning to relate the theoretical concepts to real-life examples, rather than allowing them to apply rules such as Gaussian measurement error by rote standard.

2.3 Exploratory Data Analysis

In an introductory class, this dataset also serves as a good example for graphical displays such as histograms, stem-and-leaf diagrams, and boxplots. Histograms with different bin sizes can be compared. The dataset often contains points that a standard statistical package will mark as potential outliers (see Figure 1), leading to a discussion of what it means to be an outlier and whether or not these points qualify.

2.4 The Poisson Distribution, Empirical Distributions, and Extreme Values

In a class beyond the most introductory level, chip counts serve as a good example for studying the Poisson distribution. As mentioned earlier, the number of chips per cookie made from well-mixed dough should be approximately Poisson. A cookie can be seen as the observation of an interval of a Poisson process (picture the dough being extruded in segments onto the baking sheet) and we can discuss how closely the process meets or misses the assumptions of a Poisson process. Questions such as

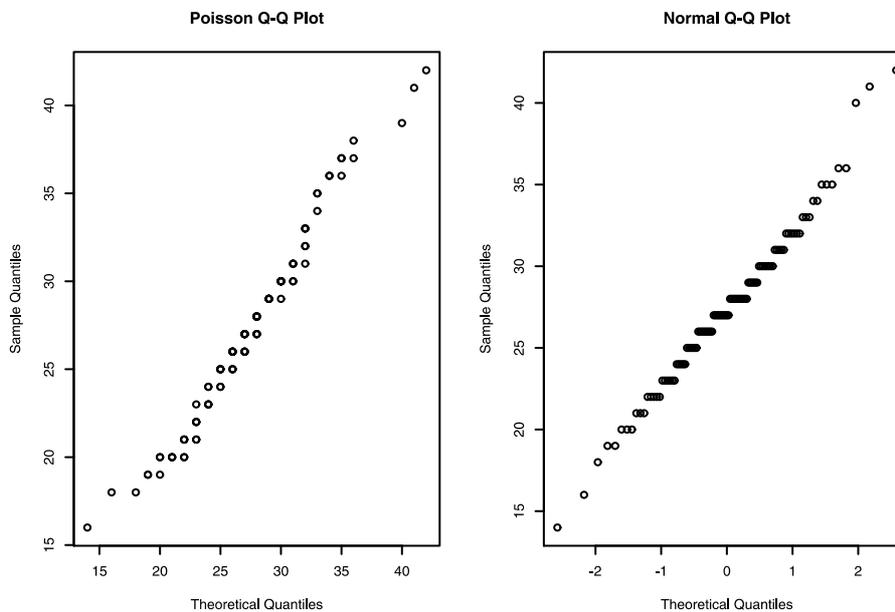


Figure 2. Quantile-quantile plot of chip counts versus theoretical quantiles from a Poisson distribution with the same mean, and a normal probability plot of chip counts.

“If a cookie has 25 chips on average, what is the probability it has exactly 20 chips?” and (with the aid of a table or computer) “What is the probability it has at least 20 chips?” can be asked.

Do the data really match a Poisson distribution? In a lower-level class, I demonstrate by simulation that typically the observed histogram looks similar to what we might expect, generating a number of datasets from Poisson distributions with the same mean and observing their shapes and ranges. In a more advanced class, one could make a quantile-quantile (Q-Q) plot to compare the empirical and theoretical distributions, or even do formal hypothesis tests to compare the distributions. The left panel of Figure 2 shows a Q-Q plot of the data against the theoretical quantiles of a Poisson distribution with mean 27.4, the mean of the observed data. The fit is not bad. The right panel of Figure 2 shows a normal probability plot, illustrating that the normal distribution is a good approximation to the Poisson for sufficiently large means (as in this case), which is just as predicted by the central limit theorem (thinking of a single draw from a Poisson as a sum of independent smaller Poisson draws).

Given a Poisson model, how reasonable is the observed variability? Various aspects of the distribution can be explored. For example, the theoretical quantiles give an indication of what is a reasonable spread. A further step would be to explore the actual distribution of the maximum and minimum, either by deriving the distributions of the order statistics, or through simulation, depending on the level of the class and the amount of class time available.

3. INFERENCE AND BEYOND WITH COOKIES

The value of cookies continues into material on statistical inference. I ask the class how they would decide if a manufacturer is putting enough chips in their cookies. If the manufacturer claims an average of 27 chips per cookie, how would they de-

cide if the manufacturer is truthful, using only a highly variable sample of cookies? Next we can consider comparing brands. Given how much difference there is in the counts for a single brand, how could we possibly hope to say if one brand of cookies has more chips than another? The answer, as we know, is that we need statistical methodology.

3.1 Sampling Distributions

The concept of a sampling distribution can be difficult for the students to grasp until they have a live example (Gourgey 2000), and the cookies work wonderfully. After personally counting cookies from one bag, the students can easily envision that a different bag would have different counts, even if it came from the same population. Indeed, in a larger class, more than one bag is necessary for everyone to get a cookie (bags typically have between 30 and 40 cookies), so by clearly marking the bags, the data can be recorded by bag and the class can see directly that not only do the bags have different averages, but that the variability of the bag averages is much smaller than the variability of the individual cookies. In a smaller class, the Chips Ahoy brand packages can be useful here because the package contains two inner bags, each with half of the cookies, so the two inner bag averages can be compared. The difference between the variability among individual cookies and of bag averages becomes tangible to the students because of the hands-on experience. This facilitates the ensuing standard lessons on sampling distributions.

3.2 Hypothesis Tests

The data are clearly relevant when we get to one-sample t tests. The students now understand that it is not reasonable to expect a bag of cookies to contain an exact number of chips, but that a certain amount of variability is natural. Thus we use

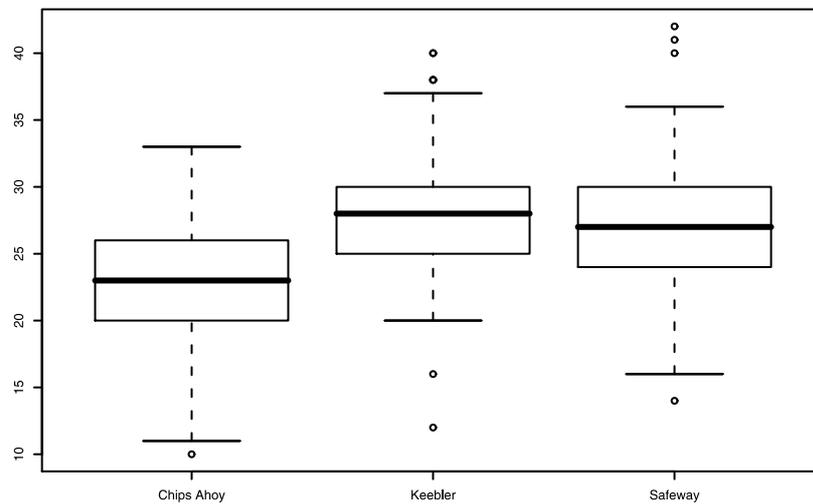


Figure 3. Side-by-side boxplots of chips counts by brand.

a t test to see if the observed difference is large relative to the observed variability.

When we get to two-sample t tests, I bring in a different brand of cookies. Again confronted with the natural variability, the students understand that without statistical tools, they have no way to decide if a difference in means between brands is just random variability or if it is a real difference. Thus they can see why the t test is necessary, and they can relate the cookie data to the parts of the t statistic, comparing variability between the brands to variability among individual cookies. We can also discuss whether or not the assumption of equal variance of the brands makes sense.

The running example goes further. I bring in a third brand to introduce one-way ANOVA. By having a continuing example, it makes the fact that ANOVA is a generalization of a t test more tangible for the students. And by now, the concept of comparing variability within brands to variability across brands is familiar to the students. An example of such an analysis is shown in Figure 3, with side-by-side boxplots of counts by brand, and the following ANOVA table:

Analysis of Variance Table

Response: Chips					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Brand	2	1929.7	964.8	42.435	<2.2e-16
Residuals	326	7412.3	22.7		

In the data from this particular class, the Chips Ahoy cookies turned out to have noticeably fewer chips on average than the other two brands (although this result can vary from class to class).

3.3 Bayesian Statistics

I also use the cookies as an example for prior elicitation in a graduate course in Bayesian statistics. I have discovered that most students do not have a good idea of what the mean should be (they typically guess closer to 10), which leads to an interesting discussion of forming a prior. We can then see how the

prior is updated with cookie data using Bayes’s Theorem, and see what the relative influences of the prior and the data are. This also helps reaffirm the idea that in many cases (including this one) the posterior mean is a weighted average of the prior mean and the data estimate, and here the weights are clearly interpretable. Figure 4 shows an example from a graduate-level class. The dotted line shows the prior elicited from the class via a spirited discussion (we eventually agreed on a gamma distribution with parameters 9 and 1). The dashed line shows the likelihood, and the solid line shows the posterior density. Thankfully the prior was not too concentrated, since the class had misconceptions about the number of chips per cookie. Here the posterior is clearly dominated by the data. As the class’s original prior is often fairly far off the mark, it is natural to ask how much the answer would have changed had a “better” prior been used, so this leads to an exploration of prior sensitivity. Because the students do both the prior elicitation and the data collection, the whole problem becomes much more real for them, and it helps them in thinking about how to relate the theory to real life.

4. ASSESSMENT

Assessment of teaching effectiveness is an important consideration. One important metric is student comprehension, which is classically measured through quizzes and exams, although projects and dynamic assessment techniques have possible advantages (Chance 1997). Another important aspect is that of student perception, for which the traditional approach is that of teaching evaluations filled out by the students at the end of the course. Here our evaluations have three open-ended questions that ask “Please comment on how the instructor’s teaching helped your learning in this course,” “Please suggest how the instructor’s teaching might improve,” and “Other Comments.”

From those questions alone, I get a number of responses that show the value of the cookies, for example: “I especially felt that his examples were relevant and helpful (such as the cookies!);” “I really liked the hands on examples, e.g. the choco-

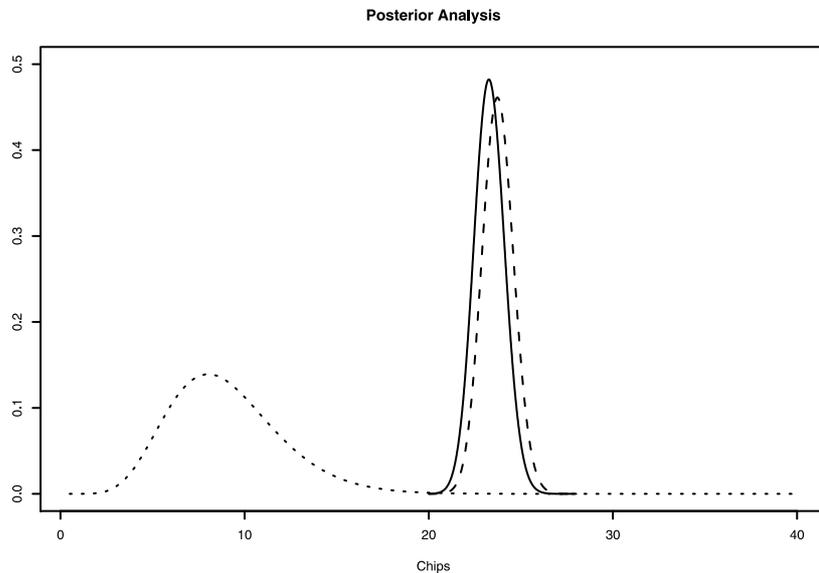


Figure 4. Prior (dotted line), likelihood (dashed line), and posterior (solid line).

late chip counting in the three different types of cookies. It made the material fun and easy to understand;" and "Loved the examples—good connections to other subjects and brought 'statistics to life' with cookie examples." The students thus respond positively to active learning methods. They also provide comments showing that such examples do make a difference in gaining their attention and assisting with learning a subject about which they had initial apprehensions: "The approach for teaching this usually 'dry' class is very good. Cookies in class are very good;" "Cookies were a great way to get students to come to class, get into the material, and understand it;" "I hate math and I was dreading the idea of taking stats . . . I actually learned what I needed to;" and ". . . made a dull subject (statistics) interesting . . . I enjoyed this class and I didn't think that I would."

Comments under the improvement question rarely relate to the cookies, with the memorable exceptions being "have more examples with healthier types of food" and "Vegan friendly class experiments, please."

5. CONCLUSIONS

One of the goals of a teacher of introductory statistics is to have their students think about statistical concepts in their lives outside of the classroom. By using the familiar chocolate chip cookie, one can demonstrate variability in an eye-opening way to the students. This demonstration causes them to look at cookies differently, and thus brings an important statistical concept directly into their lives. It is then much easier for them to apply these concepts to other examples, and they are much more likely to be interested in the rest of the course.

[Received November 2006. Revised June 2007.]

REFERENCES

- Chance, B. L. (1997), "Experiences with Authentic Assessment Techniques in an Introductory Statistics Course," *Journal of Statistics Education*, 5, 3.
- Dyck, J. L., and Gee, N. A. (1998), "A Sweet Way to Teach Students About the Sampling Distribution of the Mean," *Teaching of Psychology*, 25, 192–195.
- Garfield, J., and Ahlgren, A. (1988), "Difficulties in Learning Basic Concepts in Probability and Statistics: Implications for Research," *Journal for Research in Mathematics Education*, 19, 44–63.
- Gnanadesikan, M., Scheaffer, R. L., Watkins, A. E., and Witmer, J. A. (1997), "An Activity-Based Statistics Course," *Journal of Statistics Education*, 5, 2.
- Gourgey, A. F. (2000), "A Classroom Simulation Based on Political Polling To Help Students Understand Sampling Distributions," *Journal of Statistics Education*, 8, 3.
- Kvam, P. H., and Sokol, J. (2004), "Teaching Statistics with Sports Examples," *INFORMS Transactions on Education*, vol. 5. Available online at <http://ite.pubs.informs.org/Vol5No1/KvamSokol/>.
- Richardson, M., and Haller, S. (2002), "What is the Probability of a Kiss? (It's Not What You Think)," *Journal of Statistics Education*, 10, 3.
- Richardson, M., Rogness, N., and Gajewski, B. (2005), "4 out of 5 Students Surveyed Would Recommend this Activity (Comparing Chewing Gum Flavor Durations)," *Journal of Statistics Education*, 13, 3.
- Rossmann, A. J., and Chance, B. L. (1999), "Teaching the Reasoning of Statistical Inference: A 'Top Ten' List," *College Mathematics Journal*, 30, 4, 297–305.
- Singer, J. D., and Willett, J. B. (1990), "Improving the Teaching of Applied Statistics: Putting the Data Back Into Data Analysis," *The American Statistician*, 44, 223–230.
- Snee, R. (1993), "What's Missing in Statistical Education," *The American Statistician*, 47, 149–154.
- Sowey, E. R. (2001), "Striking Demonstrations in Teaching Statistics," *Journal of Statistics Education*, 9, 1.