**Developments in Business Simulation and Experiential Learning, Volume 25, 1998**

# SUGAR COATED STATISTICS: AN EXERCISE FOR THE FIRST DAY OF CLASS

Denise Potosky, The Pennsylvania State University, Great Valley Graduate Center

## ABSTRACT

In this exercise, students generate data and calculate descriptive statistics about candy they are given. The activity engages students in discussion about collecting and interpreting data, and about their role in using statistics to make decisions. It also establishes a framework for the study of statistics that can be extended into subsequent class sessions.

## INTRODUCTION

Every course in statistics requires that students understand how data are generated and how to calculate measures of central tendency and variability. This activity was designed to enable students to learn about statistics in a "sugar-coated" format (i.e., they get candy).

### Materials Required

- 20 to 50 inexpensive containers (e.g., brightly colored plastic eggs). Plan *for* one or two containers per student in classes of 10 to *50* students. The containers should be in separate sets of colors or groups, e.g., blue, purple, green, yellow, and pink eggs.
- 100 to 300 small candies that fit in the containers, two to ten (or more) candies in each. One color group of containers should contain the same number of candies, e.g., *5* candies in every purple egg. At least one other color should contain considerable variability in the number of candies, e.g., 2 candies in one blue egg, 9 in another, etc.
- For each student, one blank copy of a spreadsheet, which lists the descriptive statistics desired. Students may also need to refer to the appropriate formulas for computing the statistics.
- One copy of the discussion questions (provided below) for each student.

### Guidelines for Instructors

Students may need guidance during the computation stage of the activity. Ideally, the activity establishes an environment in which both mistakes and learning can occur, and one in which computational errors can be corrected quickly. Therefore, it is important that the instructor know how many candies are in each colored container, and has computed the descriptive statistics in advance.

## CONDUCTING THE EXERCISE

The activity proceeds through three stages: Data collection, computation, and interpretation.

I. Data collection. Give each student one or two containers, and explain that each container contains candy. Before eating the candy, students must first count and record the color/classification of the container and the number of candies it contains. Students should record this information on their spreadsheets for every student in the class. When this information has been assembled, the instructor should briefly explain that students have created a "dataset." There are two "variables" of interest in the this dataset: container color/classification, and number of candies. There are N "observations" in the dataset.

II. Computation. The spreadsheet should require that students calculate the sum of all the candies given to the class, the mean number of candies, **"x"** minus the mean for each observation, and several other descriptive statistics. Students should be encouraged to work with partners or in small groups for this part of the activity. When the spreadsheet has been completed, the instructor should confirm the correct answers for each of the computations.

III. Interpretation. There are three objectives that the instructor should keep in mind for this stage of

the activity. One is to encourage students to think about what the numbers they have computed indicate about the dataset. For example, it is often very instructive to ask students why we square deviations from the mean in calculating the variance (i.e., point out that if they sm the x - x differences prior to squaring them, the sum total will equal zero).

A second objective for the interpretation stage is to build self-efficacy with regard to using statistics. This is accomplished by encouraging students to make some decisions and/or guesses based upon what they know from the information they calculated. To this end, the following questions should be presented to students. If time permits, each student should first record his or her response to each question and then compare responses in small groups. These teams of students can formulate their groups' responses and then the class as a whole can discuss them.

1) How many candies do most containers contain?
2) If you were to randomly select a container from the entire set, how many candies would you guess are in the one you select? Why?
3) Describe the characteristics of each color group. Which color is most variable in terms of the number of candies it contains? Which color is most predictable?
4) If you were choosing a particular container and you wanted to maximize the number of candies you would get, what color would you choose? Why?
5) If you wanted to minimize the risk of getting a "below average" container, i.e., one with a relatively low candy count, what color would you choose? Why?

Because they have performed the calculations themselves and can see the descriptive statistics, students usually find answering the above questions surprisingly easy. Yet every student may not have the same answer, which also provides an important learning point: Although there can be "wrong" answers, there may be more than one "right" answer. Some discussion of whether or not the containers and candies in the dataset represent entire populations or just samples may be appropriate. Discussion of the types of errors that may occur during data collection (e.g., someone eats the candy!) and during interpretation (e.g., math errors, sampling errors, different risk tolerances) is also appropriate here.

A third objective of the interpretation stage is to enable students to use colored containers and candies as an analogy to other data they might collect and analyze. The number of candies might represent the number of sales per month, and each color group may be analogous to a different store location. Or the number of candies can represent scores on a test and the container color might be coded as a group identifier or some demographic characteristic. Students should be able to provide other examples of what information a dataset like this one may represent.

## CONCLUSION

The spreadsheet information produced can be used to demonstrate other concepts. For example, later in the course, the instructor can ask if the distribution of candies is skewed, in what direction, and if this matters. Students' answers to these questions lead nicely to a discussion of the empirical rule and probability distributions. Given the groups of containers and the count of candies within, the dataset may also lend itself to analysis of variance and correlation coefficient calculations.

The exercise gets students involved and requires them to share information with each other. It builds upon what they can readily do and what they may already understand about measures of central tendency and probability. As such, this "sugarcoated" exercise may help to reduce some of the anxiety students may feel on the first day of class. At the same time, the exercise enables students to calculate and hopefully internalize some fundamental statistical concepts, and it stimulates their thinking about the use of statistics in making decisions.