Against All Odds: Inside Statistics

Marsha Davis

Eastern Connecticut State University, 83 Windham Street, Willimantic, CT 06226

Abstract

Given the importance of statistics and statistical reasoning in an increasingly complex and information-rich world, ways must be found to engage students with real-world contexts and activities that support learning the basic elements of statistical thinking and the important concepts that underlie statistical reasoning, particularly concepts that students find difficult. Online resources from the new *Against All Odds (AAO): Inside Statistics* (funded by Annenberg Learner and the Consortium for Mathematics Applications (COMAP)) are designed to do just that. The videos, activities, interactive applets, exercises and written support materials in *AAO* provide a rich learning environment that makes statistics come alive and promotes deeper understanding. This presentation will focus on a video segment, an interactive applet, and related class activities. Results from class testing of *AAO* materials with students at a variety of levels will be shared with participants.

Key Words: Introductory statistics, online resources

1. Introduction

Over the years, the number of sections of *Introductory Statistics* in colleges has greatly increased. The *Common Core State Standards for Mathematics* includes expanded coverage of probability and statistics in the high school curriculum. In addition, 15,000 more students took the AP Statistics test in 2014 compared to the previous year. Unfortunately, the subject of statistics is often viewed as unpleasant, difficult, and/or boring by students. (Garfiled, Hogg, Schau, Whittinghill). Given the importance of statistics and statistical reasoning in an increasingly complex and information-rich world, ways must be found to engage students with real-world contexts and activities that support learning the basic elements of statistical thinking and the important concepts that underlie statistical reasoning, particularly concepts that students find difficult. The new *Against All Odds: Inside Statistics* series, funded by Annenberg Learner, the Consortium for Mathematics and Its Applications (COMAP), and my sabbatic leave spring semester of 2013, is designed to do just that.

Against All Odds: Inside Statistics (*AAO*) is an online resource that consists of 30 modules on topics from descriptive statistics, probability, and inference. Each module consists of a video, an activity (some of which use one of three interactive applets), written support material, and exercises that involve real-world data and use of technology. These online materials can be used either as stand-alone course materials or as supplements to other course materials. While *AAO* is designed to support a standard general-education-requirement introductory statistics course or an AP Statistics course, with careful selection and a bit of creativity, *AAO* materials can be adapted for use with a

wide range of students. For example, some materials from the early modules could be adapted for use with middle school students, while other materials could be adapted for use in a statistics course for mathematics majors, or for teacher professional development (particularly preparing teachers for the statistics and probability content in the *Common Core State Standards for Mathematics*).

2. Materials

This paper will describe sample video segments showing statistics in action, present one of the interactive applets and several activities that rely on that applet, and share sample exercises. These materials can be accessed at:

www.learner.org/courses/againstallodds

Figure 1: Homescreen for *AAO* online materials

2.1 Two-Way Tables Video: The Happiness Survey

Click on the Internet and up pops a survey. Answer the phone, and a consulting firm wants to ask about how you will vote in an upcoming election. Watch the local news and you are asked to respond to their "question of the day." Buy an item online, and you get invited to complete a satisfaction survey. Everywhere we turn, we find more surveys!

Analyzing categorical survey data one question at a time is easy, but rarely very interesting. The more interesting results come from investigating the relationships between the responses to two or more questions – in other words, from analyzing data that can be organized into cross-tabulation tables. However, when faced with two-way tables, college students (even mathematics majors!) struggle with conditional percentages. For example, on a question regarding political affiliation, students often have trouble differentiating between the percentage of men who are Republicans and the

percentage of Republicans who are men. Prior to a formal introduction of this topic, the video described below can help students get a handle on marginal and conditional distributions.

The context for the video is a survey that was administered as part of Somerville, Massachusetts' 2011 annual census. This survey, dubbed the "Happiness Survey" in several news reports, asked Somerville residents to rate their satisfaction with their current lives and with various aspects of Somerville. The video focuses on two of the survey questions:

- How happy do your feel right now?
- How would you rate the beauty or physical setting of Somerville?

In the video, Happiness ratings are boiled down into three categories: Unhappy, So-So, and Happy. Ratings of Somerville's physical beauty are categorized as Bad, OK, and Good. Participants' responses to these two questions are organized into the two-way table shown in Table 1.

Table 1: Results from rating happiness and Somerville's physical beauty

		Physical Beauty			
		Bad	OK	Good	Total
Happiness	Unhappy	90	123	62	275
	So-so	555	972	610	2137
	Happy	541	1426	1406	3373
Total		1186	2521	2078	5785

From here students can easily find marginal distributions of Happiness and Physical Beauty. For example, less than 5% of residents responded that they were unhappy. Slightly more than 20% of residents rated Somerville's physical beauty as bad.

The more interesting question, a question that cannot be answered by the marginal percentages, is whether unhappy people have a more negative view of Somerville's physical beauty than happy people. For that we need to compute the conditional distributions of Physical Beauty for each level of the Happiness variable (Table 2).

Table 2: Conditional distribution of Physical Beauty for each Happiness category

		Physical Beauty			
		Bad	OK	Good	Total
Happiness	Unhappy	32.73%	44.73%	22.55%	100%
	So-so	25.97%	45.48%	28.54%	100%
	Happy	16.04%	42.28%	41.68%	100%

From Table 2, we discover that nearly 33% of unhappy people rated Somerville's physical beauty as bad compared to only around 16% of happy people. Clearly there is linkage between the Happiness and Physical Beauty variables. Upon further inspection, we note that as the level of Happiness goes up, the percentage of Bad ratings goes down while the percentage of Good ratings goes up. The bar chart shown in Figure 2 effectively illustrates these patterns.
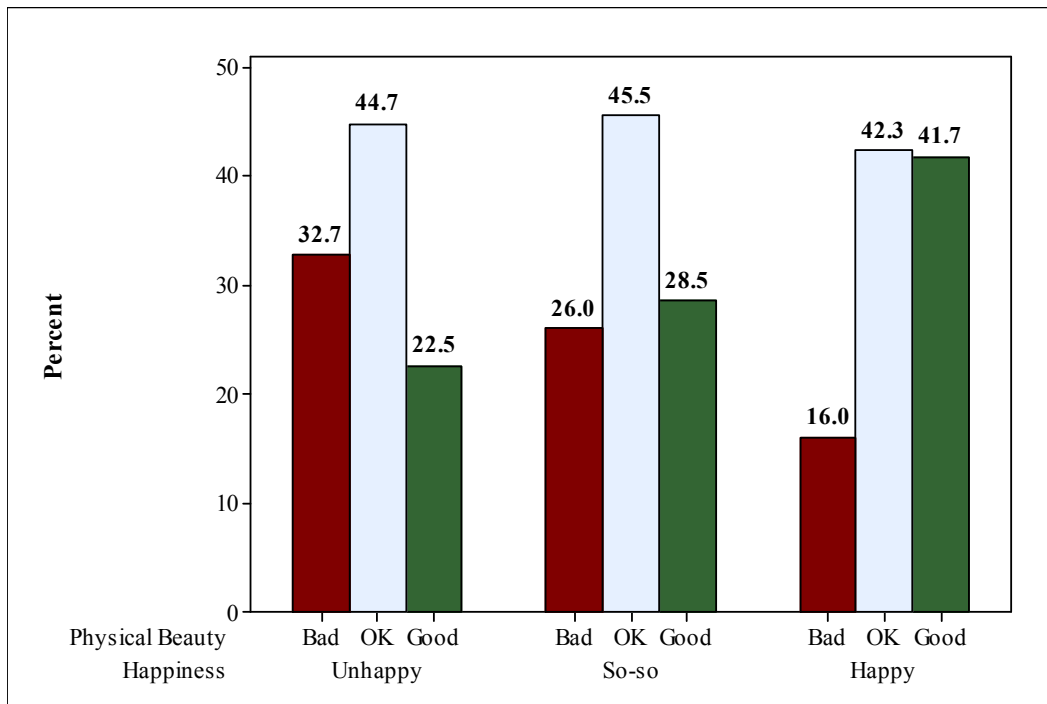

Figure 2: Conditional distribution of Physical Beauty for each level of Happiness

Now that Somerville has identified a connection between resident's happiness levels and their ratings of the city's physical beauty, officials want to dig more deeply into this connection for the next survey in an effort to improve resident's satisfaction with Somerville.

The activity for Unit 13, *Two-Way Tables*, is directly connected to this video. David Moore, the content developer for Annenberg's 1980s college-level telecourse *Against All Odds: Inside Statistics*, advocates letting students work on data that is relevant to them. What better way to do just that than to let students conduct their own Happiness survey. I have conducted a Happiness survey on the first day of class for the last several years. In addition to asking students to rate their level of happiness and to rate the beauty of our campus on the first day of class, I have also asked them to give their year (Freshman, Sophomore, Junior, Senior) and gender, which has produced some interesting results. In my introductory statistics course for mathematics majors, it appears that a higher percentage of juniors are happy on the first day of class compared to seniors. Also, as is true of the Somerville survey, a higher percentage of the happy students rated the physical beauty of campus as good compared to the unhappy students.

2.2 Wafer Thickness Interactive Applet

Over 20 years ago, Cobb (1991) provided a direction for improving the teaching of *Introductory Statistics*:

> Almost any course in statistics can be improved by more emphasis on data and concepts, at the expense of less theory and fewer recipes. To the maximum extent feasible, calculations and graphics should be automated. Any introductory course should take as its main goal helping students to learn the basics of statistical thinking.

While students in introductory statistics courses usually focus on describing central tendency and neglect variability (Gould, 2004), Moore (1990) put understanding variability and appropriate ways to quantify and model variability at the core of his description of statistical thinking. Activities based on *AAO*'s Wafer Thickness interactive applet can deepen students' understanding of variability and sources of variability.

The Wafer Thickness applet simulates measuring the thickness of polished silicon wafers used in the manufacturing of microchips. Figure 3 shows a screen capture of the applet in action. In this screenshot, the second of five polished wafers is being measured and the results are posted in real time. As data become available, the applet builds a histogram. Students can compare histograms of up to three samples. The data can be saved in a csv file for export into Excel or statistical software for further analyses.
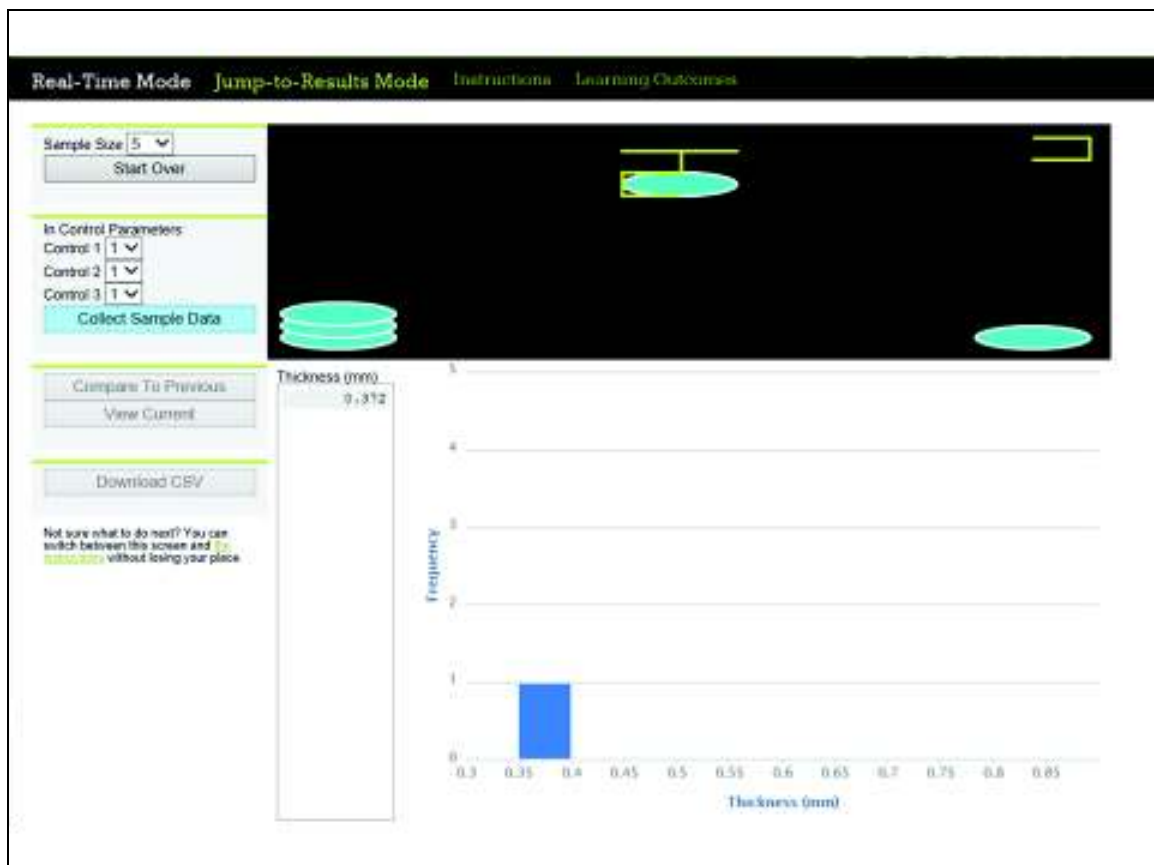

Figure 3: Screen shot of Wafer Thickness interactive applet

2.2.1 Introductory Activity: Determining the Optimal Settings

The activity focuses on quality control in the production of the polished wafers. Using the applet, students can set three controls at three different levels: 1, 2, or 3. These controls affect the thickness distribution of the polished wafers. The culminating task asks students to make a recommendation for the control settings so that the product is consistently close to a target thickness of 0.5 mm. This activity has been class tested in a variety of settings: a general-education introductory statistics course, an introductory statistics course for mathematics majors, a graduate level course for in-service teachers, and a middle-grade class. In each case, students worked in groups on the activity. Below are some observations on the results:

- Students were totally engaged in the activity. Each group had their own data, which varied from group to group.
- Students learned how to make a histogram from watching the construction of histograms in real time as data were collected. This reduced the class time spent on the topic of histograms and descriptions of their shape.
- Students observed that in a sample of wafers produced under the same control settings, wafer thickness varied from wafer to wafer.
- Students discovered that the distributions of wafer thicknesses differed from sample to sample even when produced under the same control settings.
- Students were able to observe variation due to changes in the control settings even in the presence of variability from sample to sample.
- In order to compare wafer thickness under different control settings, students created informal measures of both center and spread.
- Students discovered from class discussion that different groups made different recommendations in their culminating task. Students realized that they needed to learn additional statistical techniques before a more consistent recommendation across all groups could be reached.
- Most importantly, students bought into the activity and started the semester ready to learn.

2.2.2 Follow-up Activities

In one follow-up activity, students imported the data into a statistics package. Then they created their own graphic displays and computed numeric summaries to add to their descriptions of the data. Figure 4 shows a graphic display of comparative boxplots created by one group in the introductory statistics course for mathematics majors. The mathematics majors generally tried to collect data from all 27 possible combinations of control settings. When the same activity was given to students in the general-education-requirement introductory statistics course, they quickly discarded combinations of control settings that produced wafers with thickness far from the target thickness or settings that produced polished wafers that had comparatively large variability in thickness.
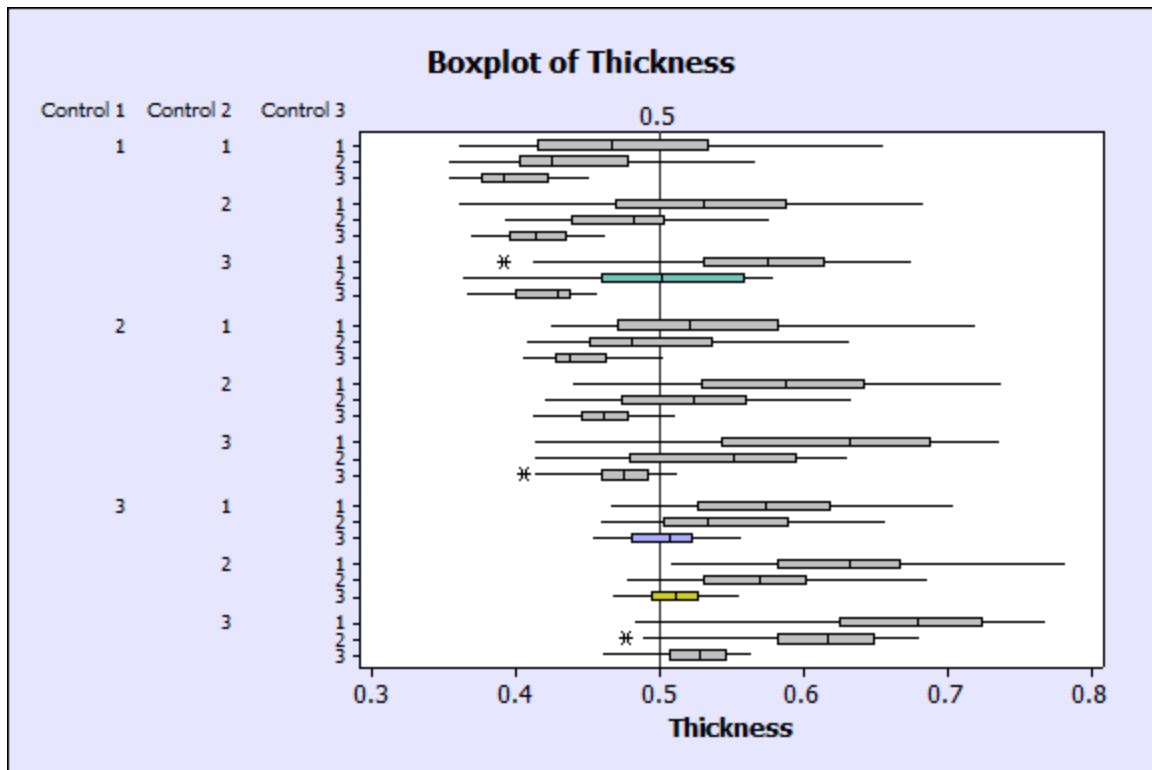
Figure 4: Graphic display of wafer thicknesses under all possible control settings

Even with the added level of information from Figure 4 and the summary statistics in Table 3, there were still some disagreements among groups about the optimal control settings.

Table 3: Summary statistics for three different control settings

Control Settings	Mean	Standard Deviation	Median	IQR
1, 3, 2	0.4953	0.0603	0.5010	0.0980
3, 1, 3	0.5048	0.0262	0.5065	0.0420
3, 2, 3	0.5111	0.0226	0.5105	0.0313

Students discovered that there is not always a definitive "best" solution. In this case, there were several good solutions and the "best" solution depended on whether the group thought it was better to be closest to "on target" or were willing to trade being a little "off target" for some reduced variability.

In the mathematics major course, this activity was extended even further to include a theoretical solution based on the underlying triangular distributions that governed the applet's random generation data:

$$f(x \mid a,b,c) = \begin{cases} 0 & x < a \\ \dfrac{2(x-a)}{(b-a)(c-a)} & a \leq x \leq c \\ \dfrac{2(b-x)}{(b-a)(b-c)} & c < x \leq b \\ 0 & x > b \end{cases}$$

where a is the lower endpoint of the triangle, c is the mode, and b is the upper endpoint.

- Control 1 changed the settings for a: (1) 0.35, (2) 0.40, (3) 0.45.
- Control 2 changed the settings for $(c-a)/(b-a)$: (1) 0.20, (2) 0.50, (3) 0.80.
- Control 3 changed the settings for $(b-a)$: (1) 0.35, (2) 0.25, (3) 0.12.

Students then computed the theoretical mean and standard deviation for the 27 possible triangular distributions and again were faced with the dilemma of selecting the control settings associated with a mean that was closest to the target or trading a little bit of bias for reduced variability.

2.2.3 ANOVA Activity

Data produced by the Wafer Thickness applet was suitable for a project on one-way ANOVA, a topic typically covered at the end of a one semester course or in the second semester of a two-semester sequence. ANOVA is particularly difficult for students to grasp conceptually—here we establish a difference in means by analyzing variation.

The ANOVA activity collects data from three experiments. In Experiment 1, control 1 is varied from level 1 to 2 to 3 while controls 2 and 3 are set at level 2. In Experiments 2 and 3, controls 2 and 3, respectively, are varied while the other controls are held fixed at level 2. For each experiment, the sample size is 10 so that students can observe the data being collected in real time. The data can be exported for analysis with spreadsheet or statistical software. Because the applet generates random data, each student (or group of students) works with different data. However, the conclusions are usually as given below:

- The mean thicknesses of wafers produced under the three different settings of control 1 differ significantly.
- The mean thicknesses of wafers produced under the three different settings of control 2 do not differ significantly.
- For control 3, the underlying assumption of equal standard deviations for ANOVA is not satisfied.

After individuals or groups have completed the activity, there is opportunity to share results. Students can observe that the same experiment results in different values for the F-statistic. However, the conclusions generally turn out to be the same. (It is possible that

one individual/group's conclusions could differ from the majority due to sampling variability, though I have yet to observe this happening.)

2.3 Exercises/Review Questions

In addition to videos and activities, the online materials include an overview of each topic along with exercises and review questions that give students an opportunity to practice what they have learned. Generally exercises and review questions are based on real data. An example of two review questions from Unit 11, *Fitting a Line to Data*, based on the data in Table 4 follows.

Table 4: Data on femur bone length and height

Femur Length (mm)	Height (cm)
447	168
444	168
470	175
459	170
482	178
520	191
464	175
470	172
482	182
462	178
522	193
461	171
422	160
520	185
476	180
508	183
477	173
504	175
547	189
508	198

1. Femur bone lengths (mm) and heights (cm) from a random sample of 20 males appears in Table 4. These data are from the Forensic Data Bank at the University of Tennessee.
   a. We would like to predict the height of a male given the length of his femur bone. Which variable is the explanatory variable and which is the response variable? Explain.
   b. Enter the data into columns of a computer software spreadsheet (or calculator lists). Make a scatterplot of the data. Describe the pattern in your scatterplot.

Does the pattern appear linear or nonlinear? Is the association between the variables positive or negative?

c. Fit a least-squares line to the data and report its equation. Overlay a graph of the line on your scatterplot from (b).

d. Do there appear to be any outliers? If so, identify the point(s). Do you think these are mistakes or real data values? Explain.

2. Return to your work from question 1.

a. Interpret the slope and *y*-intercept in the context of these data. Do these quantities make sense in the given context?

b. Two femur bones presumed to be from two men are measured and their lengths differ by 5 mm. Use the least-squares regression equation to predict the difference in heights between the two men.

c. Predict the height of a male whose femur length is 475 mm. Is this a reasonable height for a man? (Convert your answer to feet and inches. Recall there are 2.54 centimeters per inch.)

d. The femur length of a boy measures 250 mm. Predict the height of the child. Explain why this prediction might not be trustworthy.

3. Conclusion

The new online *Against All Odds*: *Inside Statistics* series will breathe life into any introductory statistics course. Providing students with videos showing statistics in action and activities in which students can collect and analyze their own data can engage students and support their learning of difficult statistical concepts. The *AAO* written materials include summaries of statistical techniques/concepts as well as exercises/review questions that provide opportunities for students to review and put into practice what they have learned.

References

Cobb, George W. (1991). "Teaching statistics: more data, less lecturing," *Amstat News*, December 1991, pp. 1, 4.

Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics*. www.corestandards.org/Math.

Garfield, J.& Hogg, B. & Schau, C. & Whittinghil, D. (2000). Best Practices in Introductory Statistics. Position paper prepared for the Undergraduate Statistics Education Initiative.

Garfunkel, S. (Project Manager) & Davis, M. (Chief content Advisor/Writer), & Villiger, M. (Writer/Producer/Director). (2013). *Against All Odds: Inside Statistics*. Washington, DC: Annenberg Learner. www.learner.org/courses/againstallodds/.

Gould, R. (2004). Variability: One Statistician's View. Statistics Education Research Journal, 3(3), pp. 7 – 16.

Moore, D. (1990). Uncertainty. In L. Steen (Ed.). On the Shoulders of Giants: New Approaches to Numeracy, pp 95 – 138. Washington. DC: National Academy Press.