



INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

An Interactive Excel VBA Example for Teaching Statistics Concepts

Weiyu Tsai, Don G. Wardell,



To cite this article:

Weiyu Tsai, Don G. Wardell, (2006) An Interactive Excel VBA Example for Teaching Statistics Concepts. INFORMS Transactions on Education 7(1):125-135. <https://doi.org/10.1287/ited.7.1.125>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2006, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

An Interactive Excel VBA Example for Teaching Statistics Concepts

Weiyu Tsai

Don G. Wardell

*David Eccles School of Business
University of Utah*

*1645 E. Campus Center Dr. #106
Salt Lake City, Utah 84112-9304*

mgtwt@business.utah.edu

mgtgdw@business.utah.edu

Abstract

It is often challenging for business students to learn abstract statistical concepts and apply these concepts to their work. Three concepts in particular that we have found difficult to communicate effectively are the Central Limit Theorem, interval estimation and hypothesis testing. To improve the effectiveness of teaching these fundamental statistical concepts, we developed a Visual Basic for Applications (VBA) driven Excel spreadsheet that is built around one simple business scenario. The scenario involves setting the filling speed in a cereal filling plant. The faster the filling speed, the larger the variation in cereal box weights and the higher the chance of having an out-of-control filling process. On the other hand, the lower the filling speed, the less efficient the plant is at utilizing capacity. Through interactively finding the optimal filling speed, students are exposed to these key statistics concepts as well as random sampling techniques. Hence, we integrate the illustration of three important statistical concepts in one simple yet practical business scenario. Moreover, the Excel VBA-driven example demonstrates several Excel statistical formulae that are useful to business students. We conducted an in-class open-book quiz to two sections of professional MBA students to assess teaching effectiveness of this interactive example. The results showed that the scores of those using the interactive VBA demo were superior to those exposed to more traditional techniques at 10% significance level. A follow-up on-line feedback survey further supported the usage of the Excel VBA-driven example in enhancing student learning.

1. Introduction

Sampling distributions (including the Central Limit Theorem), interval estimation and hypothesis testing are basic yet important concepts discussed in introductory data analysis courses for undergraduate and graduate business students. Due to the abstraction and complexity of the concepts, students tend to memorize the calculation procedures in completing these courses instead of understanding the concepts underlying these calculations. As a result, many students may leave their classrooms without a solid understanding of the possible applications to their current and/or future jobs.

In this paper we illustrate the usage of a single Excel VBA spreadsheet model that we have found successful in helping students to visualize these concepts. The model introduces and links the three important statis-

tical concepts, namely the Central Limit Theorem, interval estimation and hypothesis testing in one simple business scenario. In addition, there is no need for extra software since Excel is widely used by business school faculty and students. In this familiar worksheet environment, students can enhance their understanding of these concepts by applying the formulae created in the file to their own calculations.

The business scenario in the Excel model is about a cereal filling process. We adapted the scenario from Berenson et al (2003). In this business scenario, students act as a plant manager at a cereal filling plant. This manager is responsible for setting the speed of the cereal filling process during each eight-hour shift. The speed of the filling process is a key factor that affects the variation in the process and results in some boxes to be underfilled and others to be overfilled. The manager knows that the faster the speed, the larger

the variation and the higher chance of creating an out-of-control filling process.

If the process is not working properly, the weight in the boxes could deviate too far from the label weight of 368 grams to be acceptable. Because it is too time-consuming and costly to weigh every single box, the manager must weigh only a sample of boxes and judge whether the cereal-filling process is working properly.

As a plant manager, a student's goal is to utilize the capacity of the cereal-filling plant effectively. In other words, students want to find the maximum cereal filling speed under the condition that the average weight of the total cereal boxes filled during the eight-hour shift is still 368 grams.

In the following section, we review the extant literature related to software tools to visualize statistical principles. Section 3 provides detailed instructions to use the model. In section 4, we provide evidence of teaching effectiveness based on the comparison of the results from a quiz of two professional MBA sections and feedback from a simple anonymous survey. Finally, we conclude in section 5.

2. Literature Review

The idea of using interactive demonstrations and visual representations to help teach statistical principles is not new. For example, the commercial statistics package *Visual Statistics 2.0* (Doane, Mathieson, Tracy, 2001) offers excellent graphical demonstrations and makes statistical concepts vivid and relevant to students. Similarly, Berk and Carey (2004) use Microsoft Excel VBA demonstrations to allow students to visualize concepts such as the CLT, interval estimation and hypothesis testing. These commercial packages, however, often require new textbook adoption and additional software installation. In contrast to these commercial packages, the advancement of Sun Java computing technology has made it possible to develop interactive statistical Java applets to aid in the visualization of statistics concepts on-line for free (West, Ogden, 1998).

In 1998, West and Ogden first developed and shared their six Java applets on the internet in the statistics education community for free. Later, Mills (2002) conducted two extensive literature searches on journal

articles using computer simulation methods to teach statistics in business, economics, mathematics, statistics, social sciences, education, and medicine. In his review, he summarized 48 of the most relevant articles and concluded that statistics educators do agree upon the effectiveness of using interactive simulation programs on the Internet to benefit student learning.

Because of the fast growth in developing interactive statistics Java applets, we made an attempt to canvass the current state. We used Google's search engine to search the topic of "statistics applets" and found 1,840,000 related (English) URL hyperlinks (retrieved on May 20, 2006). The volume reduces to 21,500 hyperlinks related to the Central Limit Theorem, 860 hyperlinks for hypothesis testing, and 296 hyperlinks for all three concepts. From these hyperlinks we found several websites that offers excellent demonstrations of these statistical topics.

We found that Rice University's Virtual Lab in Statistics (Lane, 2006) provides a nice sampling distribution applet that allows users to choose a population (either normal, uniform, skewed or custom) and then draw samples to form the corresponding sampling distribution. The website also provides a confidence interval applet when the sampling is done from a population with a mean of 50 and a population standard deviation of 10. For each of the samples, both 95% and 99% confidence intervals are computed and shown graphically. Users can adjust the sample size, but the number of samples is fixed as 100. Other excellent applets include the Statistical Java from Virginia Tech (Dorai-Raj, Anderson-Cook, Robinson, 2002), the Java applets for visualization of statistical concepts from Katholieke Universiteit Leuven (Michiels, Raeymaekers, 2000), and the Web Interface for Statistics Education (WISE) applets from Claremont Graduate University (Berger et al, 2005).

During our search we noticed that the three statistical concepts are all introduced to students as isolated entities in these software packages and Java applets. Mills (2002) provided a possible explanation of our observation. The reason is that the relationships among CLT, confidence intervals and hypothesis testing are difficult to present using interactive computer programs. The goal of this paper, therefore, is to develop a visual statistical tool in the Excel environment to integrate the three important concepts through one business

related scenario and help students learn to use Excel's statistical functions.

3. Instructions to Use the "Visualize CLT" File

The "Visualize CLT.xls⁽¹⁾" file (which can also be downloaded from CLT⁽²⁾) has two sequential modes: the first for instructors and the second for students. The first time an instructor opens the file, a dialog box appears as shown in Figure 1. The dialog box requests the instructor to enter the maximum stable filling speed (i.e., optimal filling speed) for this cereal plant. The instructor can specify a number between 10,000 and 60,000 (boxes of cereal per eight-hour shift). This is a small set up for this file and gives the instructor the opportunity to customize the file to each class. After providing this input, the instructor saves the file with his/her desired file name and distributes it to the students for class use.

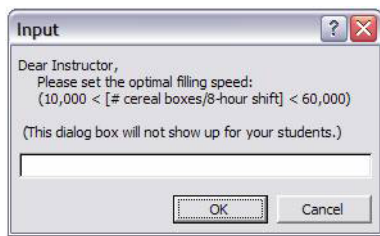


Figure 1: Input Dialog Box for Optimal Filling Speed

Upon providing the file to the students, if necessary the instructor directs them to set the Excel security level to "medium" under the menu item Tools/Macro/Security before opening the file. Once students open the file, they see six worksheet tabs with

a description of the business scenario in the active "Description" worksheet as shown in Figure 2. The steps that students need to take to find the optimal filling speed are also described in the worksheet "Description". We describe these steps below.

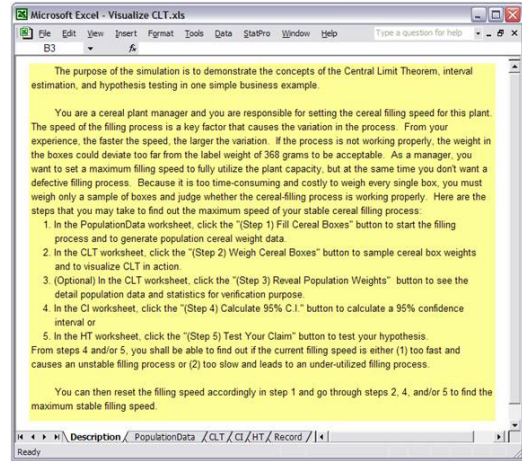


Figure 2: 'Description' Worksheet

3.1. Step 1. Generate Population Data

In the first step, students click on the "(Step 1) Fill Cereal Boxes" button in the worksheet "PopulationData" as shown in Figure 3. An information window with the warning "The original population data, if exists, will be erased!" appears to caution students. After students click OK, another dialogue box appears and requests students to type in a number that indicates number of cereal boxes to be filled (between 10,000 and 60,000 boxes) in the eight-hour shift. Students also select the distribution of the weight of cereal boxes filled in the filling process among normal, uniform, and exponential distributions.

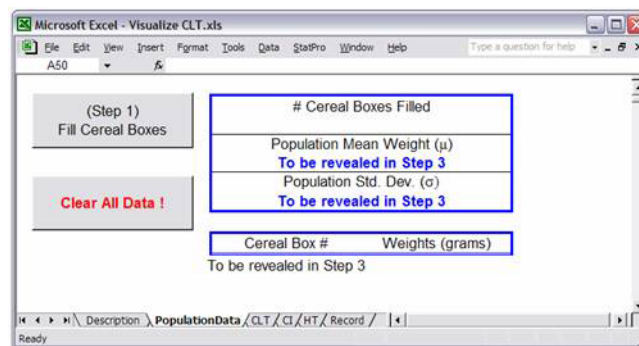


Figure 3: 'PopulationData' Worksheet

(1) <http://ite.pubs.informs.org/Vol7No1/TsaiWardell/VisualizeCLT.xls>

(2) <http://home.business.utah.edu/mgtdgw/CLT/>

Depending on whether the filling speed is larger than the maximum stable filling speed set by the instructor, Excel uses the formulae specified in Table 1 to generate the population data. These formulae make sure that

the population mean weight deviates dramatically away from 368 when the filling speed chosen by a student is larger than the maximum stable filling speed set by the instructor.

Table 1: Formulae Used in Generating Population Data

| Condition | Normal Distribution | | Uniform Distribution | | Exponential Distribution |
|--------------|---------------------|--------------------|------------------------|------------------------|--------------------------|
| | Mean | Standard Deviation | Minimum | Maximum | Mean |
| $x \leq x^*$ | 368 | 2 | $368 - 3.5$ | $368 + 3.5$ | 368 |
| $x > x^*$ | $368 - (x/5000)$ | 5 | $368 - (x/5000) - 8.5$ | $368 - (x/5000) + 8.5$ | $368 - (x/5000)$ |

Note:

1. x is the filling speed chosen by a student.
2. x^* is the maximum stable speed set by an instructor.

When the filling process starts, students are able to see the progress of the filling process. For 12,000 boxes this filling process takes about 10 seconds to complete. The population data can be revealed later in step 3 if desired.

3.2. Steps 2 & 3. Explain Central Limit Theorem

In the second step, students start by taking a sample in order to decide whether the filling process works properly according to the speed they chose. By clicking on the "(Step 2) Weigh Cereal Boxes" button in worksheet "CLT" as shown in Figure 4, students start the sampling process.

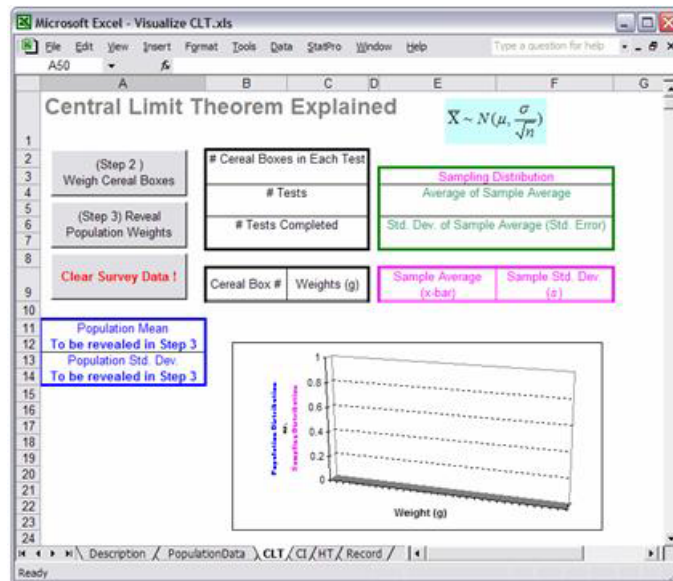


Figure 4: 'CLT' Worksheet

Students are asked to decide both the number of cereal boxes (sample size) in a sample and the number of times that this sampling procedure (sampling times) will be performed. There are no limits on setting these two numbers, but the larger the numbers are, the longer the sampling process takes. As a reference, for a sample size of 10 with 3,000 sampling times, it takes about one minute to complete the sampling process. The cereal box numbers and weights are recorded in the worksheet "CLT" starting from cells B10 and C10. The sample average and sample standard deviation

of the sample is calculated and recorded starting from cells E10 and F10 for each sample. In addition, the average weight of each sample is denoted in the bar chart. Each unit represents one data record of the sample average.

According to the Central Limit Theorem, if the sample size is large enough, the sampling distribution of the sample average follows a normal distribution with mean equal to the population mean and standard deviation equal to the population standard deviation

divided by square root of the sample size. Students can verify the theorem by setting the sampling times to a large number (say, 3,000) each with a sample size of 10, and observe the sampling distribution in the bar chart live.

During the process of drawing a sampling distribution, Excel pauses one second every 25 sampling times in order to allow the sampling distribution to be shown in transit in the bar chart. Interestingly, you may speed up the drawing process by pressing the 'Esc' key *once* when Excel pauses. However, if the 'Esc' key is pressed when Excel is running, you might encounter a warning message that indicates that you are about to interrupt the VBA program. Should it happen, click the "Continue" button in the message window to continue the program.

Once the sampling distribution is completed, Excel calculates the average of the sample averages and the standard deviation of the sample averages in cells E5 and E7. Students then click on the "(Step 3) Reveal Population Weights" button to show the statistics and the distribution of the population data. Please note that in order to make sure that both the population and sampling distributions are visible on the same chart, we normalize the scale of the y-axis using the relative frequency rather than using the original scale of the number of sampling tests. We normalize the scale by dividing the number of boxes in the population distribution by the total number of cereal boxes filled in the 8-hour shift, and dividing the number of sample averages in the sampling distribution by the sampling times.

After going through the process for all three population distributions, students can see in the chart that whether the population distribution is normal, uni-

form, or exponential, the sampling distribution of the sample average is normal with mean equal to population mean and standard error equal to population standard deviation divided by the square root of the sample size as predicted by the Central Limit Theorem. As a rule of thumb, a sample size of 30 or larger is desired in order to verify the Central Limit Theorem. We add a reminder in the sample size input window to tell students to use a number of 30 or larger for an exponentially distributed population. This completes the steps 2 and 3 in explaining the Central Limit Theorem.

Recall that the manager's goal is to find the maximum stable filling speed without knowing the population data. In other words, instructors can delete the "(Step 3) Reveal Population Weights" button to make the population data unavailable. In order to find the answer, students need to know if the speed they choose in step 1 is too fast or too slow, and then make adjustments accordingly. In other words, students are required to apply the concepts of interval estimation or hypothesis testing to make an informative decision based on the cereal boxes they sampled in step 2. In steps 4 and 5 below, we focus on applying the concept of interval estimation and hypothesis testing, respectively, to find the answer.

3.3. Step 4. Explanation of Interval Estimation

The concept of a 95% confidence interval means that if we could perform the sampling procedure many times, say 1000 times, we expect that 950 times the calculated confidence intervals should contain the population mean. In the worksheet "CI" as shown in Figure 5, students have the chance to verify the concept in this step.

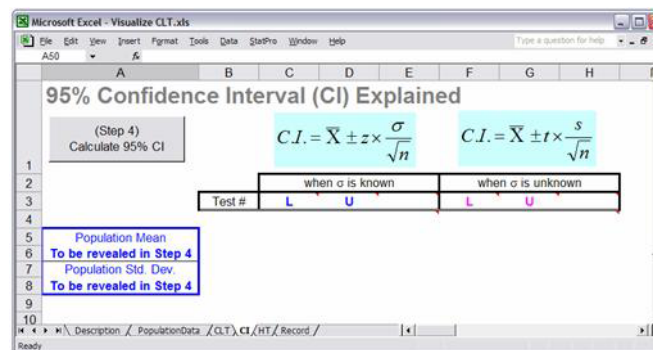


Figure 5: 'CI' Worksheet

In the worksheet "CI" when students click the "(Step 4) Calculate the 95% CI" button, Excel calculates the 95% confidence interval for each of the samples collected previously in step 2. Excel calculates the lower and upper limits of the 95% confidence interval of each sample test according to the sample averages recorded in worksheet CLT starting at cell E10 (refer back to Figure 4). In the CI worksheet, Excel lists the calculated confidence intervals starting from cells C4 and D4 using z-values when the population standard deviation is known; and starting from cells F4 and G4 using t-values if the sample standard deviation is used. For the two types of confidence intervals, cells in columns E and H starting with E4 and H4 indicate 1 if the population mean is in the confidence interval and 0 otherwise (these cells that don't include the population mean are highlighted in red). The formulae in the cells of Test #1 are listed as follows:

Cell C4: =CLT!\$E10-NORMSINV(97.5%)*\$A\$8/SQRT(CLT!\$B\$3)

Cell D4: =CLT!\$E10+NORMSINV(97.5%)*\$A\$8/SQRT(CLT!\$B\$3)

Cell E4: =IF(AND(\$C4<=\$A\$6,\$A\$6<=\$D4),1,0)

Cell F4: =CLT!\$E10-TINV(5%,CLT!\$B\$3-1)*CLT!\$F10/SQRT(CLT!\$B\$3)

Cell G4: =CLT!\$E10+TINV(5%,CLT!\$B\$3-1)*CLT!\$F10/SQRT(CLT!\$B\$3)

Cell H4: =IF(AND(\$F4<=\$A\$6,\$A\$6<=\$G4),1,0)

where the meanings of the cell addresses of the corresponding terms in the CI formulae of Test #1 are

| | |
|-------------|-----------------------|
| \bar{X} : | CLT!\$E10 |
| z: | NORMSINV(97.5%) |
| t: | TINV(5%,CLT!\$B\$3-1) |
| σ : | \$A\$8 |
| s: | CLT!\$F10 |
| n: | CLT!\$B\$3 |

Through viewing the formulae, students can understand how to use statistical functions such as NORMSINV and TINV in Excel to construct confidence intervals.

Excel also calculates the percentage of the sample tests in which their confidence intervals contain the population mean in cells E3 and H3. Instructors can use this calculation as an opportunity to discuss the sample size assumption underlying the confidence interval computations. For example, instructors can divide the class into three groups and ask each group to generate the population of cereal box weights using different distributions in step 1, but use the same sample size of 5 in step 2. They then can compare their resulting percentages of the sample tests calculated in step 4. Instructors can then show that for the cases of either a normal distribution or uniform population (both of which are symmetric), the proportion is very close to 95%. For an exponential distribution, however, the proportion can be much lower than 95% unless a large sample size, e.g. 30, is used. In other words, the sample size needed to create the 95% confidence interval of an exponential population is much larger than that of a normal or uniform population.

Instructors can also ask students to calculate the average width of the confidence intervals for both known σ and unknown s cases. Students can gain more intuition through this exercise and confirm the intuition that the confidence interval of unknown s case is larger than that of the known σ case.

We note that Excel calculates the confidence intervals for *all* samplings conducted in step 2. Therefore, when students test whether the process is working properly, they conduct the sampling procedure *only once* in step 2 and then use the calculated confidence interval to check if 368 is included. We also add a msgbox in the spreadsheet to clarify this concept. Another method to check the filling process is to use hypothesis testing as described in step 5.

3.4. Step 5. Explain Two-sided Hypothesis Testing

We finally demonstrate the concept of hypothesis testing by testing whether the average weight of all cereal boxes under the chosen speed is still 368 grams. Students click the "(Step 5) Test Your Claim" button in "HT" worksheet as shown in Figure 6 to submit their claimed weight.

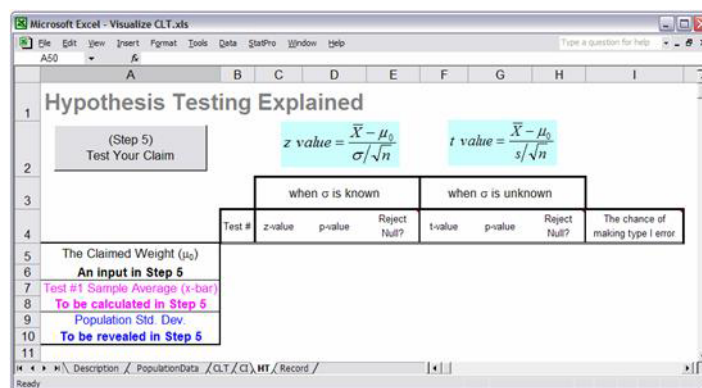


Figure 6: 'HT' Worksheet

Excel then calculates the corresponding z-value and p-value if the population standard deviation is known or the corresponding t-value and p-value if the population standard deviation is unknown; and decides whether to reject the null hypothesis or not. The formulae used for the first sample test are as follows:

Cell C5: $= (CLT!E10 - A\$6) / (A\$10 / \text{SQRT}(CLT!B\$3))$

Cell D5: $= 2 * (1 - \text{NORMSDIST}(\text{ABS}(C5)))$

Cell E5: $= \text{IF}(D5 < 5\%, 1, 0)$

Cell F5: $= (CLT!E10 - A\$6) / (CLT!F10 / \text{SQRT}(CLT!B\$3))$

Cell G5: $= \text{TDIST}(\text{ABS}(F5), CLT!B\$3 - 1, 2)$

Cell H5: $= \text{IF}(G5 < 5\%, 1, 0)$

In addition, if the submitted claimed weight is 368 grams, Excel also shows the two-sided null and alternative hypotheses and the corresponding p-value for test #1 in a graph. We position the graphs next to the case that s is known to correspond to the normalization formulae shown in the graphs. Instructors can use the graphs to explain the concepts of standardization, z-value, p-value, and type I error. When the chosen filling speed is less than or equal to the optimal filling speed set by the instructor, then the population mean is equal to the suggested hypothesized value of 368. In that case, Excel also calculates the simulated type I error using the following formula in cell I5: $= \text{AVERAGE}(H:H)$, which finds the relative frequency of the error. Students can compare this number and see how close it is to the theoretical value of 5%.

In contrast, when the filling process is out of control because a student chooses a filling speed larger than

the instructor's optimal speed, the population mean is not equal to the hypothesized value of 368 (it is lower). In that case, if the students choose to test the hypothesis, the null hypothesis $\mu = 368$, it is impossible to make a type I error because the null hypothesis is not true. Consequently, Excel reports "0.00%" as the type I error in the HT worksheet.

In the case when a student clicks the "(Step 5) Test your claim" button and inputs a value other than 368, the spreadsheet has no way of knowing whether or not the null hypothesis is true without referring back to the formulae used in Table 1. Therefore, Excel won't be able to calculate the type I error correctly. If Excel does refer to Table 1 to find out if the null hypothesis is true and calculates the corresponding type I error, it is difficult to explain to students where the value of the type I error came from without showing the underlying formulae used to generate the population data. Thus, we let Excel simply reply "N/A". The comment in Cell I4 contains the following message to note to students how this simulated type I error is reported.

"If you test the value of 368 in your claim, Excel calculates the simulated type I error according to column H when σ is unknown. If you test other values, Excel reports 'N/A' in the cell."

This completes step 5 as well as the use of this Excel file. Through this five-step exercise, students are able to show statistically if the cereal filling speed they chose causes an unstable filling process. If the current process is under control, students should rerun the experiment and increase the filling speed in order to find the optimal filling speed for the cereal plant. Otherwise, they should decrease the filling speed.

To help students keep track of their input filling speeds, the file also includes a "Record" worksheet as shown in Figure 7. This worksheet automatically records the fill rates and distributions that students

choose in their experiments. In each run of the experiment, the worksheet also records the sample statistics as well as the confidence interval and the p-value of the first set of sampling data.

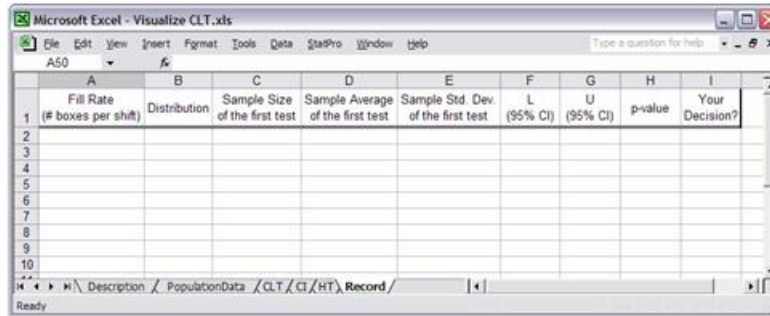


Figure 7: 'Record' Worksheet

This worksheet provides an excellent opportunity for students to use the file in an integrative way and forces them to interpret the results correctly. For example, to enhance students' understanding of these statistical principles, instructors may suggest that students, as an exercise, compare the differences in their output analyses based on different input populations generated from the normal, uniform, and exponential distributions. In addition, instructors can use this file to assess students' understanding of the concept of hypothesis testing. First, instructors create another file with a different optimal filling speed. Instructors then delete the Step-3 and Step-4 buttons (or the CI worksheet altogether) in this newly created file and ask students to find the optimal filling speed.

4. Evidence of Teaching Effectiveness

We introduced the concepts of interval estimation and hypothesis testing to students in two sections of a professional MBA Data Analysis and Decision Making course. We used the Excel Visual CLT file in section A with 84 students and excluded the file in section B with 87 students. One week after we covered the topics, we gave an in-class open-book quiz (provided in the Appendix 1) with 4 related questions totaling 40 points to both sections. For section A, the average grade on the quiz was 31.4 with a standard deviation of 8.5. For section B, the average grade on the quiz was 29.3 with a standard deviation of 11.2. The two-sample t-test statistic for sample mean difference was 1.396 and the corresponding the p-value was 0.082. Therefore, the average grade difference of 2.1 between the two sections was significant at 0.10 level, but not at the 0.05 level. The lack of significance at the .05 level

may be due to the fact that it was an open-book quiz and students were able to find correct formulae in a reasonable amount of time. Nevertheless, we can at least conclude that the Excel Visual CLT file does have some positive impact on students' understanding of the concepts of interval estimation and hypothesis testing. In addition, we received many positive comments in a simple follow-up open-ended question online survey (provided in the Appendix 2) without any negative responses. We especially liked the following vivid comment:

"I wanted to tell you that the simulation was awesome! I guess I'm a nerd, but I thought it was pretty cool - I was having way too much fun watching the graph grow. It really helps visualize what is going on."

5. Conclusion

In this paper, we describe a simple example with a VBA-driven Excel spreadsheet to teach the key statistics concepts of the Central Limit Theorem, interval estimation, and hypothesis testing. Through this exercise, students are also exposed to random sampling techniques. In addition, instructors can easily redesign the final answer in class and assess students' understanding using this interactive file. An in-class open-book quiz was administrated to students and the results showed a significant difference at the 0.10 significance level between those who were taught using the spreadsheet vs. those who were taught through lecture alone. A follow-up on-line feedback question further supported the usage of the Excel macro in student learning.

6. Acknowledgement

The authors would like to thank the Guest Editor and two anonymous referees for their thorough comments and suggestions that greatly improved the paper. The first author would also like to express deep appreciation on learning VBA tools from many excellent speakers presenting at the INFORMS 2004 Teaching of Management Science Workshop coordinated by Dr. Stephen G. Powell. In addition, the authors gratefully acknowledge that funding for the project was provided by a summer research grant from David Eccles School of Business, University of Utah.

<http://www.amstat.org/publications/jse/v10n1/mills.html>

West, R. W. and R. T. Ogden. (1998) "Interactive Demonstrations for Statistics Education on the World Wide Web," *Journal of Statistics Education* [Online], Vol. 6, No. 3, pp. 1-8,
<http://www.amstat.org/publications/jse/v6n3/west-ogedn.html>

References

- Berger, D. E. (2005), The Web Interface for Statistics Education (WISE) applets. Retrieved on May 20, 2006 from
<http://wise.cgu.edu/>
- Berenson, M. L., D. M. Levin, and T. C. Krehbiel, (2003), *Basic Business Statistics: Concepts and Applications*, 9th Edition, Pearson-Prentice Hall, NJ.
- Berk, K.N. and P. Carey, (2004), *Data Analysis with Microsoft Excel*, 2nd Edition, Thomson Brooks/Cole, Pacific Grove, CA.
- Doane, D. P., K. Mathieson, and R. L. Tracy, (2001), *Visual Statistics 2.0*, McGraw-Hill, Boston, MA.
<http://www.mhhe.com/business/opsci/doane/home.htm>
- Dorai-Raj, S., C. Anderson-Cook, and T. Robinson, (2002), The Statistical Java. Retrieved on May 20, 2006 from
<http://www.stat.vt.edu/~sundar/java/applets>
- Land, D. M. (2006) Java Applets for Sampling Distribution and Confidence Interval. Retrieved on May 20, 2006 from
http://www.ruf.rice.edu/~lane/stat_sim
- Michiels, S., and B. Raeymaekers, (2000), Visualization of and Experimentation with Statistical Concepts. Retrieved on May 20, 2006 from
<http://www.kuleuven.ac.be/ucs/java>
- Mills, J. D. (2002), "Using Computer Simulation Methods to Teach Statistics: A Review of the Literature," *Journal of Statistics Education* [Online], Vol. 10, No. 1, pp. 1-26,

Appendix

Appendix I - Quiz Questions

Question 1

The SEC requires companies to file annual reports concerning their financial status. It is impossible for SEC to audit every account receivable in a company's filing. Therefore, suppose SEC audit a random sample of 49 accounts receivable invoices for one company and find a sample average of \$128 and a sample standard deviation of \$53. What is the 99% confidence interval for the average amount of an accounts receivable invoice of this company?

Solution

Question 2

Pineapple Corporation maintains that their cans have always contained an average of 6 ounces of fruit. The production group believes that the mean weight has changed. The drained weights in ounces for a sample of 15 cans of fruit from Pineapple Corporation show a sample average of 6.137 with a sample standard deviation of 0.203. Use an appropriate hypothesis test to determine if the data show evidence of a change in mean weight at a significance level of 0.05.

Solution

Question 3

A new potential supplier claims that the average time to failure (m) of his IC chips is at least 1000 hours. We sample 100 chips to test his claim and find that the sample mean is calculated to be 950 hours with sample standard deviation 500 hours. Can we reject his claim at a 10% significance level?

Solution

Question 4

A supplier for a critical airplane component claims that the population mean diameter of the supplied component is 2.5 mm with the population standard deviation of 0.6 mm. You take a sample of 25 parts, and find the sample mean to be 2.25 mm. What is the probability of finding a sample mean of 2.25 or smaller if the supplier's claim is true?

Solution

Appendix II - Survey Question

Anonymous WebCT Online Feedback Form

I seek your comments on the Visualize CLT Excel file that we used last week in this course. Please respond to the following questions.

1. Do you think the "Visualize CLT.xls⁽³⁾" Excel file enhances your learning in these statistical concepts (Central Limit Theorem, Confidence Interval, and Hypothesis Testing)?
2. Please list any bug you find in the Excel file.

Results

The anonymous survey was posted in one WebCT course section with 84 students for 10 days. We received 71 responses with no negative comments. The response rate was 85%.

⁽³⁾ <http://ite.pubs.informs.org/Vol7No1/TsaiWardell/VisualizeCLT.xls>