

Lab 2:
Population Parameters, Sampling, Sample Statistics, and the Sampling Distribution.

Objectives:

1. More practice with Excel: equations and functions.
2. Review and illustrate the concept of sampling from a population.
3. Review population and sample descriptive measures.
4. Review and illustrate the concept of a sampling distribution.

Key Terms:

1. Population vs. Sample; Random Samples.
2. Characterizing Distributions: Center and Variation.
3. Population Mean and Population Standard Deviation.
4. Estimation: Sample Mean and Sample Standard Deviation.
5. Sampling Distributions.

Data: The data (used Accord Price data) are in the Excel file: *2004 Accord Prices.xls*. The file is the same one we used last week and is available on the course website – the **Lab Exercises** page. For this week’s lab, you can just open the file you saved last week.

Exercises:

◆ ***Saving Your Excel File***

First, open the Excel file that you worked on last week in lab. Before you do anything else, be sure that you can easily save the file to your folder on the server. You should also save the file to your “U Drive,” a thumb drive, or some other *secure* location. You should be able to access this file for future use.

◆ ***Review: Descriptive Measures for the Population of Used Accord Prices***

1. We worked with these data for used Honda Accords last week in lab. We defined these data as the **population** of all used Honda Accords available on-line on a given day. You should have calculated the **population mean** last week, and you should have calculated **deviations** and **squared deviations** for each observation. You should have then calculated the **population standard deviation**. We also asked for the **5 Number Summary**.
2. We used a couple of the Excel statistical functions last week, *average* and *stdevp*, to check answers. My price data were in column C with labels in row 1, so I typed the expressions: “=AVERAGE(C2:C101)” and “=STDEVP(C2:C101)” to determine the values for the mean and standard deviation. Note that there are two functions for the **standard deviation**; one is for population data (“=STDEVP(..)”) and the other for sample data (“=STDEV(..)”). Be sure you use the correct function for the data you have. Because we had the population data, we calculated the true population parameter values.
3. A quick way to get descriptive measures for a data set is to use the Data Analysis tool. This is presented in Chapter 4, section 4.1 of the Middleton book. Use the **Data Analysis** tool to calculate descriptive measures for the population data. Select **Descriptive Statistics**, enter the input range for the Accord Prices and save the results in a **New Worksheet Ply** that is named: “*Excel’s Stats*.” Review the results that Excel gives you. Which statistic is wrong? Why? You always need to exercise caution when using software. Here, Excel made an assumption because you selected **Descriptive Statistics**. **What was Excel’s assumption?**

◆ ***Sampling from the Population***

1. In statistical analysis, populations are typically too large to collect all values and calculate population parameters, so we **estimate using sample data**. Estimates of the population mean and population standard deviation are obtained using the formulas for the **sample mean** and **sample standard deviation**. What are these two formulas or estimators?

2. While we can calculate the population parameters in this case, let's proceed as if we can't. To *estimate*, we need *sample data*. Use Excel's **Sampling** function found in the **Data** ribbon (click on the **Data** tab). Then look on the far right of the ribbon for **Data Analysis**. Scroll through the list of items in the menu until you find **Sampling**. The **Input Range** is the set of 100 used Accords and we want samples drawn from the population data for the variable *Price*. The **Sampling Method** is **Random**, the **Number of Samples** is 10, and by choosing "New Worksheet Ply," Excel will place the sample in a new worksheet. **Rename** the new worksheet **Samples n = 10**. The sample will be placed in rows 1-10 of column A. Move the sample down to rows 2-11 and name these ten observations "Sample1" in row 1.
3. Calculate the sample mean and sample standard deviation for your sample of size 10. Use the Excel functions (i.e., the functions for the mean and standard deviation) – it's faster. Place the mean and standard deviation on the same sheet.
4. Repeat the process to create a second sample. Be careful, Excel will likely want to over-write your first sample. For **Output options**, choose **Output Range**: and click on the icon. Be sure to click on one cell in the "Sample n=10" worksheet.

◆ *Sample of Size 30.*

1. Repeat the process above to create two samples of size 30. Place the sample in a worksheet titled "**Samples n = 30.**" Compute the sample mean and standard deviation for your sample of size 30.

- ◆ Report your results to the website: <http://courses.umass.edu/resec312/labs/lab2results.html>. Round each sample mean to the nearest whole dollar – no decimal places. Please don't use the **Enter** key to move from one cell to the next (Tab works!) or hit the submit button until you reviewed your estimates. Once you have all estimates entered and you check them for accuracy, hit the Submit button.

◆ *Confidence Intervals*

1. Create a 90% confidence interval for the population mean using your sample mean for your sample of size 30 and place it on the page "**Sample n = 30.**" The formula for the confidence interval is:

$$\bar{Y} \pm z_{\alpha/2} \cdot \frac{\sigma_Y}{\sqrt{n}} .$$

We can apply the z-interval because: (1) the population standard deviation is known (you calculated it last week), and (2) we can feel confident that the sample mean is normally distributed. Why?

2. Create a 90% confidence interval for the population mean assuming we know nothing about the population standard deviation. If we don't know the population standard deviation, we estimate. How? The formula for the confidence interval is:

$$\bar{Y} \pm t_{\alpha/2} \cdot \frac{s_Y}{\sqrt{n}} . \quad (\text{No t-table? No problem. The Excel function: "}=tinv(\text{probability}, \text{df}) \text{ will give}$$

you the t-value you need.

3. Use Excel's **Data Analysis** tool to calculate descriptive statistics for your sample of 30 observations. Check "Summary Statistics" and "Confidence Level for the Mean." Look at the output. What do you think the value labeled "Confidence Level(90%)" represents?

◆ *Review*

Write a couple of paragraphs that summarize what we did today; answer the questions below. (You can put these in your Excel Spreadsheet by using Text Boxes – you'll find an icon for **Text Box** on the **Insert** ribbon.) Compare your **estimates** of the population mean to your neighbors. **What do you observe about the estimates you and your neighbor obtained?** I'm collecting all estimates from the class – I should get about 200 sample means for samples of size 10 and the same number for samples of size 30. What should we expect to see if I create two histograms and we compare the two distributions? Why did I pick a sample size of 30 for the larger samples? Should we find that all the 90% confidence intervals contain the true value for the population mean? **Explain.**