Lab Activity: The Central Limit Theorem

In this lab activity, you will explore the properties of the Central Limit Theorem.

Student Learning Outcomes

By the end of this chapter, you should be able to do the following:

- Generate groups of random samples from a list of data values in Statcato

- Compute sample mean and standard deviation in Statcato

- Understand the properties of the Central Limit Theorem and apply them appropriately

Preliminary

Read Chapter 7 The Central Limit Theorem in:

  Illowsky, Barbara, and Susan Dean. <u>Collaborative Statistics</u>. Connexions. 2 Mar. 2010 <http://cnx.org/content/col10522/1.37/>.

Make sure you understand the following **key terms** (**LR:Key Terms**):

  Central Limit Theorem for Sample Means, standard error of the mean, Central Limit Theorem for Sums, Law of Large Numbers

Entering Data

In this activity, you will use the ages of U.S. presidents at inauguration. The 44 ages are as follows:

57, 61, 57, 57, 58, 57, 61, 54, 68, 51, 49,

64, 50, 48, 65, 52, 56, 46, 54, 49, 51, 47,

55, 55, 54, 42, 51, 56, 55, 51, 54, 51, 60,

62, 43, 55, 56, 61, 52, 69, 64, 46, 54, 47

This data set is also available on the Statcato web site. You can use the load dataset tool in Statcato to obtain the data:

Loading an Online Dataset

Go to `File > Load Dataset`.   Or click the icon in the toolbar.

- Under Online Datasets, enter the web address: http://statcato.org/labs/data/ages.csv. Select "Comma-separated values" in the dataset file type drop-down menu.

- Click **Load Dataset**.

The 44 ages should now be in column C1.

Generating Groups of Random Samples

Now you will generate groups of 40 random samples of size 2, 5, and 10 from the list of 44 ages.

![STATCATO] Generating 40 random samples of size 2

Go to `Data > Generate Random Data > Sample from Column`.

- For **Sample from Column**, select **C1 ages** in the drop-down menu
- In **Store Samples in** text box, enter: c2-c3
- Select **Sample with Replacement**
- **Number of Samples to Generate**: 40
- Click OK

Columns C2 and C3 contain the 40 random samples of size 2. Each pair of two samples in C2 and C3 in each row represents a sample.

![STATCATO] Generating 40 random samples of size 5

Click the item **Random Samples from Column** in Dialog History (or go to `Data > Generate Random Data > Sample from Column`).

- For **Sample from Column**, select **C1 ages** in the drop-down menu
- In **Store Samples in** text box, enter: c5-c9
- Select **Sample with Replacement**
- **Number of Samples to Generate**: 40
- Click OK

Columns C5 to C9 contain 40 random samples of size 5.

![STATCATO] Generating 40 random samples of size 10

Click the item **Random Samples from Column** in Dialog History (or go to `Data > Generate Random Data > Sample from Column`).

- For **Sample from Column**, select **C1 ages** in the drop-down menu
- In **Store Samples in** text box, enter: c11-c20
- Select **Sample with Replacement**
- **Number of Samples to Generate**: 40
- Click OK

Columns C5 to C9 contain 40 random samples of size 5.

Computing Sample Means

You will compute the mean of each random sample. The mean of samples of size 2 will be placed in column C4, that of size 5 in column C10, and that of size 10 in column C21. First, you will label those columns:

TATCATO Labeling Columns for Means

- Click on the cell right below column label C4 (in row Var). Enter **means of size 2**.

- Similarly, label C10 **means of size 5**.

- Label C21 **means of size 10**.

Next, you will compute the mean of each sample.

TATCATO Compute Means of Random Samples of Size 2.

Go to `Statistics > Basic Statistics > Row Statistics`.

- In the **Input Variables(s)** text box, enter c2-c3.

- For **Statistic**, select **Mean**.

- **Store Results In**: c4

The means of sample groups of size 2 should now be in column C4.

TATCATO Compute Means of Random Samples of Size 5.

Click the item **Row Statistics** in Dialog History (or go to `Statistics > Basic Statistics > Row Statistics`).

- In the **Input Variables(s)** text box, enter c5-c9.

- For **Statistic**, select **Mean**.

- **Store Results In**: c10

The means of sample groups of size 5 should now be in column C10.

TATCATO Compute Means of Random Samples of Size 10.

Click the item **Row Statistics** in Dialog History (or go to `Statistics > Basic Statistics > Row Statistics`).

- In the **Input Variables(s)** text box, enter c11-c20.

- For **Statistic**, select **Mean**.

- **Store Results In**: c21

The means of sample groups of size 10 should now be in column C21.

Computing Means and Standard Deviations of Sample Means

You will compute the means and standard deviations of the original data as well as of the three groups of sample means.

TATCATO Compute the Mean of Original Sample Data and Sample Means.

Go to `Statistics > Basic Statistics > Descriptive Statistics`.

- **Input Variables(s)**: c1 c4 c10 c21 (separated by a space)
- For **Statistics**, select **Mean** and **Standard Deviation**.
- Click **OK**

The means and standard deviations should now be displayed. Copy the descriptive statistics to **LR: Means and Standard Deviations**.

Creating Histograms

You will now construct a histogram for original data (C1) and each set of sample means (C4, C10, C21).

TATCATO Creating Histograms of C1, C4, C10, and C21

Repeat the following steps for each of the four columns (C1, C4, C10, C21):

Go to `Graph > Histogram`.

- For **Graph Variables**, choose the desired column in the list box. Choose **Frequency** under **Heights of bars represent**.
- For **X-axis**
  - o Choose the option **Provide the class width and the minimum**.
    - ▪ Class width: 3
    - ▪ Minimum: 40
- In **Other Options** panel, enter a descriptive title in the **Title** text box.
- Click **OK**.

You can copy the graph by going to `Graph > Copy Graph to Clipboard` and then paste it to **LR: Histograms**.

Discussion

Answer the following questions in **LR: Discussion**.

1. State the theoretical distributions for the sample means, according to the Central Limit Theorem.

n	$\overline{X} \sim N(\mu,\sigma)$	
	μ	σ
2		
5		

10		

How does the theoretical distribution compare to the empirical distributions?

2. Compare and contrast the four histograms you have made. Describe the similarities and differences. How do they demonstrate the Central Limit Theorem for Sample Means?

3. Find the probability that a randomly selected president has an age more than 50 at inauguration.

4. Find the probability that the sample average age at inauguration for 5 presidents is more than 50.

5. Find the probability that the sample average age at inauguration for 10 presidents is between 50 and 60.