Lab Activity: Normal Distribution

In this lab activity, you will create a histogram and compute descriptive statistics, z-scores, and normal probabilities for a built-in dataset in Statcato. Using the computer-generated results, you will determine normality and unusual values. You will also model the data using a normal distribution and compute various probabilities.

Student Learning Outcomes

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

- Create histograms for data in Statcato
- Compute descriptive statistics, z-scores, and normal cumulative and inverse cumulative probabilities in Statcato
- Recognize the normal probability distribution and apply it appropriately

Preliminary

Read Chapter 6 The Normal Distribution in:

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

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

normal distribution, standard normal distribution, z-scores, critical value

Loading Data

This lab uses a dataset built into Statcato (*Eye Gain Ratio Measurements for Schizophrenic Patients*, Schizophrenic Eye-Tracking Data in Rubin and Wu (1997) Biometrics).

STATCATO Here you will load the built-in dataset.

- Go to `File > Load Dataset`. Or click the icon in the toolbar.
- Under Built-in Datasets, select the **Choose a sample dataset** radio button. Select "Eye Gain Ratio Measurements for Schizophrenic Patients" in the drop-down menu.
- Click **Load Dataset**.

The data should now be in Data window. In the following steps, you will be using the schizophrenic eye gain ratios in C1.

Creating a Histogram

You will now construct a histogram for the data.

STATCATO

Go to `Graph > Histogram`.

- For **Graph Variables**, choose **C1** in list box. Choose **Frequency** under **Heights of bars represent**.
- For **X-axis**
    - Choose the option **Provide the number of classes, minimum, and maximum**.
        - Number of bins: 10
    - For **Label**, enter the word *eye gain ratios*.
- In **Y-axis** panel, enter the word *frequency* for **Label**.
- In **Other Options** panel, enter *Histogram for Schizophrenic Eye Gain Ratios* in the **Title** text box.
- Click **OK**.

The generated histogram should be shown in a separate window. You can copy the graph by going to `Graph > Copy Graph to Clipboard` and then paste it the provided space to **LR: Histogram**.

Computing Descriptive Statistics

You will compute the mean and standard deviation of the data.

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

- In the **Input Variable(s)**: text box, enter **C1**.
- In the **Statistics** panel, select the following statistics: Mean, Standard deviation.
- Click **OK**.

The selected descriptive statistics should now be displayed in the Log. Copy them to **LR: Descriptive Statistics**. Also record the mean and standard deviation below.

* $\bar{x}$ = ____________________ s = ____________________

Computing z-scores

Next you will find the z-scores for the data.

Go to `Data > Standardize`.

- **Input Column(s)**: C1
- **Store results in column(s)**: C2
- Select **Subtract mean and divide by standard deviation** (z-score)
- Click **OK**.

The z-scores for the Schizophrenic eye gain ratios are now in column C2. We consider values to be **unusually low** if they have z-scores less than -2 and **unusually high** if they have z-scores greater than +2. Copy the data values that are unusual (those with z-scores less than -2 or greater than +2) to **LR: z-scores**.

Using the Normal Distribution

Now you will model the distribution of schizophrenic eye gain ratios using the statistics of the sample data you computed above. You will model it with a normal distribution that has a mean and standard deviation the same as the sample mean and standard deviation:

$X \sim N(\bar{x}, s) \sim N($__________,__________$)$ fill in $\bar{x}$ and s from above (*)

Using this normal distribution model, you will compute various probabilities.

Find the probability that a randomly chosen schizophrenic eye gain ratio is less than 0.8.

This means finding P(X < 0.8). You can find this cumulative probability in Statcato.

STATCATO Finding Normal Cumulative Probability

Go to `Calculate > Probability Distributions > Normal`.

- **Distribution Parameters**
  - Mean: enter the value of $\bar{x}$ (*)
  - Standard deviation: enter the value of *s* (*)
- **Compute**: Cumulative probability
- **Input(s)**: Select Constant. Enter 0.8 in the text box.
- Click **Compute**.

Record the result in **LR: Normal Distribution**.

Find the probability that a randomly chosen schizophrenic eye gain ratio is at least 0.5.

This means finding P(X ≥ 0.5) = 1 – P(X < 0.5). Find the normal cumulative probability P(X < 0.5) in Statcato in order to obtain P(X ≥ 0.5). Record the result in **LR: Normal Distribution**.

Find the probability that a randomly chosen schizophrenic eye gain ratio is between 0.8 and 1.

This means finding P(0.8 < X < 1) = P(X < 1) – P(X < 0.8). Find the normal cumulative probabilities in Statcato and record the result in **LR: Normal Distribution**.

Find the 60th percentile of schizophrenic eye gain ratios.

The 60^{th} percentile is the number *k* such that P(X < *k*) = 0.6. You can find *k* (the inverse cumulative probability of 0.6) in Statcato.

Finding Normal Inverse Cumulative Probability

Go to `Calculate > Probability Distributions > Normal`.

- **Distribution Parameters**
  - Mean: enter the value of $\bar{x}$ (*)
  - Standard deviation: enter the value of s (*)
- **Compute**: Inverse cumulative probability
- **Input(s)**: Select Constant. Enter 0.6 in the text box.
- Click **Compute**.

Record the result in **LR: Normal Distribution**.

Discussion

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

1. Describe the shape of the histogram of schizophrenic eye ratios.
2. Do you think the normal distribution model you used to approximate the population of schizophrenic eye ratios is good? Is so, why? If not, how would you improve the model?
3. How many unusual values are there in the sample data? What is the percentage of sample values that are usual? Is it what you expect? Explain.
4. Suppose you are creating a study of schizophrenic patients. You want to exclude patients who have eye gain ratios that are within the bottom 3% and top 3%. What would be the two critical values separating the rejected eye gain ratios from the others? Show your work.