

JMP INTRO® Lab Activities

Lab Activity - Graphically Assessing Normality

Data Set: Fitness

Situation: Mr. Teaque runs the “World Fitness Class Gym” which has over 500 members. Members of this gym must be over 35 but less than 60. He has selected a random sample of members and collected data on their running time and pulse rate. Before any statistical analysis can be completed he must make sure the data can be regarded as a sample from a normal population.

In this lab you will use a data set called “Fitness” which is already a JMP INTRO file. You will be investigating if the data are normal. You will use the 68-95-99.7 rule and the normality plot to check for normality. Begin this activity by opening the file. **Open Date Table** double click on **Sample Data** double click on **Fitness**. *(Your data should now be displayed on a table with 9 columns and 31 rows. The columns represent the sex, age, weight, oxygen consumption, runtime, running pulse, resting pulse, and maximum pulse for the particular subjects.)* The question is whether the values of any of the individual columns are distributed in such a way that they are modeled well by a normal distribution.

First investigate the **weight** data by creating a histogram.

Select **Analyze** ® **Distribution** then select **Weight** select **Y,Column** click **OK**

You now have a histogram with the Quantile table. You may need to scroll down and you will also find the “Moments” table.

From the tables find the following values: (Round to two decimal places).

Mean_____ Standard Deviation _____ Median_____

Minimum_____ Quartile 1 _____ Quartile 3_____ Maximum_____

What are the minimum and maximum values necessary to define a value as an outlier?

Are there any outliers? _____

Describe the general shape of the histogram: *(answer in complete sentences)*

Create a stem and leaf plot

You will need to find out how many data points fall within one standard deviation above and below the mean. You may find the information a number of ways. Use a stem and leaf plot when you have small sets of data, and use the sorting feature when you have larger sets of data.

Method 1 – Stem and Leaf Plot

Open the menu using the red triangle on the histogram graph. Activate the stem and leaf plot along with the histogram graph.

What can you see on the stem and leaf plot that you could not see on the histogram?
(answer in complete sentences)

Using the 68-95-99.7 rule how many data points, for this particular set, would need to be included within one standard deviation above and below the mean to cover 68% of the data? _____

Calculate the mean plus and minus one standard deviation, the interval that should include 68% of the data. _____

What percent of your data is in that interval? _____

Calculate the mean plus and minus two standard deviation, the interval that should include 95% of the data. _____

What percent of your data is in that interval? _____

Calculate the mean plus and minus three standard deviation, the interval that should include 99.7% of the data. _____

From this data set what percent does each data point contribute? _____

What percent of your data is in that interval? _____

From this information would you conclude that the population distribution appears to be approximately normal? Explain.

Method 2 – Sorting Method

A specified column can be sorted so that it appears from smallest to largest. This is done using **Tables® Sort** select **Weight** select **By** select **Sort**. The program will always make a new table with the default name of “Subset of...” There’s a checkbox on the Tables→Sort dialog that allows you to replace the original data.

What student name was listed first before the table was sorted? _____

What student name was listed first after the table was sorted? _____

Normal Quantile Plot

Open the menu using the red triangle on the histogram graph. Activate the **Normal Quantile Plot**.

Do the data follow a diagonal linear pattern?

You will notice in the JMP INTRO that also included in the quantile plot are confidence bands around the diagonal. If a point falls outside this boundary this may be another indicator that the population distribution is not normal. Does this data set have any points outside the confidence band? _____

On the quantile plot what variables the axes represent? (*Recall Quantile plots we have done in class*)

Horizontal _____ Vertical _____

Does the quantile plot support the fact that the population distribution is approximately normal? Justify your answer.

(answer in complete sentences)

Homework:

Use the data of running time and resting pulse and check for normality. Justify your answer using the 68-95-99.7 rule and the quantile plot. Include statistical justification and specific numerical values.

JMP INTRO® Lab Activities

Teacher Notes

Lab Activity - Graphically Assessing Normality

Objectives:

- Use the 68-95-99.7 rule to investigate normality
- Use the Normal Quantile Plot to investigate normality
- Use JMP INTRO as the tool

Time required: 45 minutes

Homework: 45 minutes

Materials:

- Lab Activity 2 student activity directions
- Fitness Data set

Prerequisites:

- Students will have experience checking for normality by hand and/or with graphing calculators. Although the program does not label the axes make sure when graphing by hand or with a calculator the axes are identified.

JMP INTRO Notes:

- The Fitness data is located in the Sample Data folder as a part of JMP INTRO
- Students may need help changing the direction of the histogram. Recall to change a histogram's direction:
use the red triangle menu on the **Weight® Display Option® Horizontal Layout**

Concept Notes:

- It is important for the student to understand that these checks for normality do not guarantee that the data are in fact normal. These verification procedures can only show that normality is plausible and that data that approximately follow the 68-95-99.7 rule and show a linear pattern in the normal quantile plot may not be normally distributed. These check procedures can only show that you have no evidence to think the data are not normal. Discuss with the students the danger in using only the 68-95-99.7 rule with small sets of data.
- Similarly, for small data sets from a normal population, these check may not be accurate even though the population is normal. For example the 68-95-99.7 rule would be difficult to use with a sample size of 10. Each data point would account for 10% of the data and following the two or three standard deviation pattern would not be possible.

JMP INTRO® Lab Activities

Answer Key

Suggested answers for Graphically Assessing Normality:

Mean 77.44 Standard Deviation 8.33 Median 77.45
 Minimum 59.08 Quartile 1 73.03 Quartile 3 82.78 Maximum 91.63

What are the minimum and maximum values necessary to create an outlier?

The random variable of weight must be less than 58.41 or more than 97.41

There are no outliers

Describe the general shape of the histogram.

The histogram appears to be approximately normal with a slight skewness to the right.

What can you see on the stem and leaf plot that you could not see on the histogram?

You can clearly see exactly how many counts are in each particular value.

Calculate the boundaries for the middle 68% of the data set.

$$77.44 + 8.33 = 85.77 .$$

$$77.44 - 8.33 = 69.11.$$

How many data points are included in this 68%? 20

Calculate the boundaries for the middle 95% of the data set. 60.78 to 94.10

How many data points are included in this 95%? 31

Calculate the boundaries for the middle 99.7% of the data set. 52.45 to 102.43

How many data points are included in this 99.7%? 31

From this information would you conclude that this data appears to be normal? Explain

The data appears to be normal. There is 65% of the data set between one standard deviation above and below the mean. There is however 100 % of the data set within 2 standard deviation above and below the mean. Because of the small data set of 31 points each point accounts for 3.2% of the data and this would make using the 68-95-99.7 rule somewhat weak to justify normality.

What student name was listed first before the table was sorted? *Donna*

What student name was listed first after the table was sorted? *Suzanne*

Does this data set have any points outside the confidence band? *no*

On the quantile plot what variables are represented by both axes?

Horizontal *z-score* Vertical *data points*

Does the quantile plot support the fact that the data is normal? Justify your answer.

The quantile plot shows a linear pattern that would justify normality. There are also no points outside the confidence bands around the diagonal.