

JMP® Introductory Lab Activities

Activity 4: Graphically Assessing Normality



Data Set: Fitness.jmp

Summary

Mr. Teaque runs the World Fitness Class Gym, which has more than 500 members. Members of this gym must be older than 35 but younger 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 JMP data set called **Fitness**. You will investigate whether the data are normal using the 68-95-99.7 rule and the normality quantile plot. Use this document as a worksheet.

The Fitness Data

Open the file **Fitness.jmp** (go to **Help > Sample Data**, then click **See an Alphabetical List of all Sample Data Files**).

The data table has nine columns and 31 rows. The columns represent the name, sex, age, weight, oxygen consumption (Oxy), runtime, running pulse, resting pulse and maximum pulse for the particular subjects.

The question of interest is whether the values in each column are distributed in such a way that they are modeled well by a normal distribution.

Explore the Weight Data

First, investigate the **Weight** data by creating a histogram. Select **Analyze > Distribution**, select **Weight** for the **Y, Columns**, and click **OK**.

You now have a histogram with the **Quantiles** and **Summary Statistics** tables.

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*).

Investigating Normality Using the 68-95-99.7 Rule

You will need to find out how many data points fall within one, two and three standard deviations above and below the mean. You will explore two ways of finding this information:

- Method 1: Stem and leaf plot (when you have small data sets)
- Method 2: Data sorting (when you have larger data sets)

Method 1 – Stem and Leaf Plot

From the Distribution output window for Weight, select **Stem and Leaf** from the **red triangle** next to **Weight**.

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

In this data set what percent does each data point contribute?

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 deviations (the interval that should include 95% of the data).

What percent of your data is in that interval?

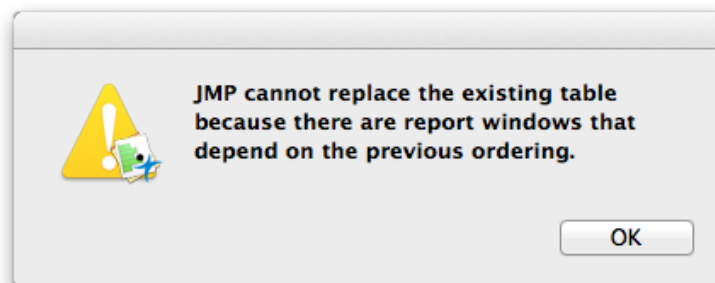
Calculate the mean plus and minus three standard deviations (the interval that should include 99.7% of the data).

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 (again, *answer in complete sentences*).

Method 2 – Sorting Method

A specified column can be sorted so that it appears from smallest to largest. To do this, **return to the data table**. Right click on a column, and select **Sort**. Note: If any analyses are open, you'll receive the following message:



Click **OK**. JMP will make a new table with the sorted data.

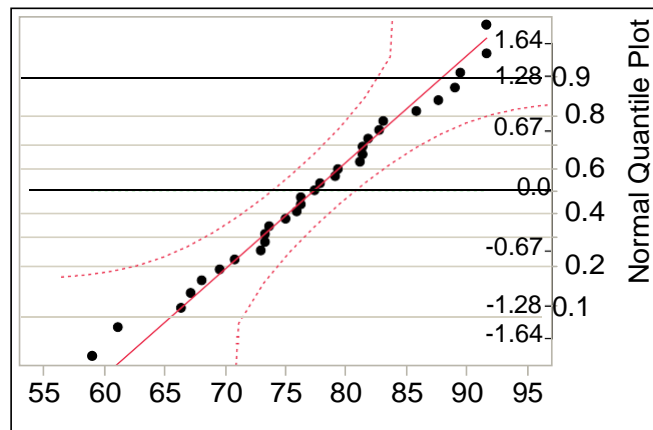
With the data sorted we could now apply the 68-95-99.7 rule to investigate normality using this sorted data (we won't do this exercise).

Normal Quantile Plot

Return to the Distribution output for Weight. If you closed the output, go to **Analyze > Distribution** and rerun the analysis. Hint: Click the **Recall** button in the Distribution dialog window to remember what you did the last time.

Select **Normal Quantile Plot** from the **red triangle** next to **Weight**.

Do the data follow a diagonal linear pattern?



You will notice that the normal quantile plot includes 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 normal quantile plot, what variables do the axes represent?

Horizontal:

Vertical:

Does the normal quantile plot support the assumption that the population distribution is approximately normal? Justify your answer (*answer in complete sentences*).

Homework

*Investigate the normality of the running time (**Runtime**) and resting pulse (**RstPulse**) data.*

Justify your answer using the 68-95-99.7 rule and the normal quantile plot.

Include statistical justification and specific numerical values.