

How can I create a Q-Q plot in SAS?

Authored by
stats writer

June 23, 2024

RECOMMENDED CITATION

stats writer (2024). *How can I create a Q-Q plot in SAS?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=148362>

A Q-Q plot, or quantile-quantile plot, is a graphical method for assessing the similarity between two distributions. In SAS, a Q-Q plot can be created by using the PROC UNIVARIATE or PROC SGPLOT procedures. The PROC UNIVARIATE procedure allows for the creation of a traditional Q-Q plot, where the data is plotted against theoretical quantiles. The PROC SGPLOT procedure offers more customization options, such as the ability to add reference lines and change the plot style. Both procedures require the use of the ODS (Output Delivery System) to display the plot. Q-Q plots are useful for comparing distributions and identifying potential outliers or deviations from normality.

Create a Q-Q Plot in SAS

A Q-Q plot, short for "quantile-quantile" plot, is used to assess whether or not a set of data potentially came from some theoretical distribution.

In most cases, this type of plot is used to determine whether or not a set of data follows a normal distribution.

If the data is normally distributed, the points in a Q-Q plot will lie on a straight diagonal line.

Conversely, the more the points in the plot deviate significantly from a straight diagonal line, the less likely the set of data follows a normal distribution.

The easiest way to create a Q-Q plot in SAS is to use the PROC UNIVARIATE statement along with the

QQPLOT statement:

```
proc univariate data=my_data noprint;  
qqplot my_variable;  
run;
```

The following examples show how to use this syntax in practice.

Note: We use the NOPRINT statement to suppress all other summary statistics and tables that are automatically generated by the PROC UNIVARIATE statement.

Example 1: Create Q-Q Plot in SAS for Normal Data

The following code shows how to create a Q-Q plot for a dataset that contains 1,000 generated from a with a mean of 10 and standard deviation of 2:

```
/*generate 1000 values that follow normal distribution  
with mean 10 and sd 2 */  
data normal_data;  
do i = 1 to 1000;  
x = 10 + 2*rannor(1);  
output;
```

```
end;
```

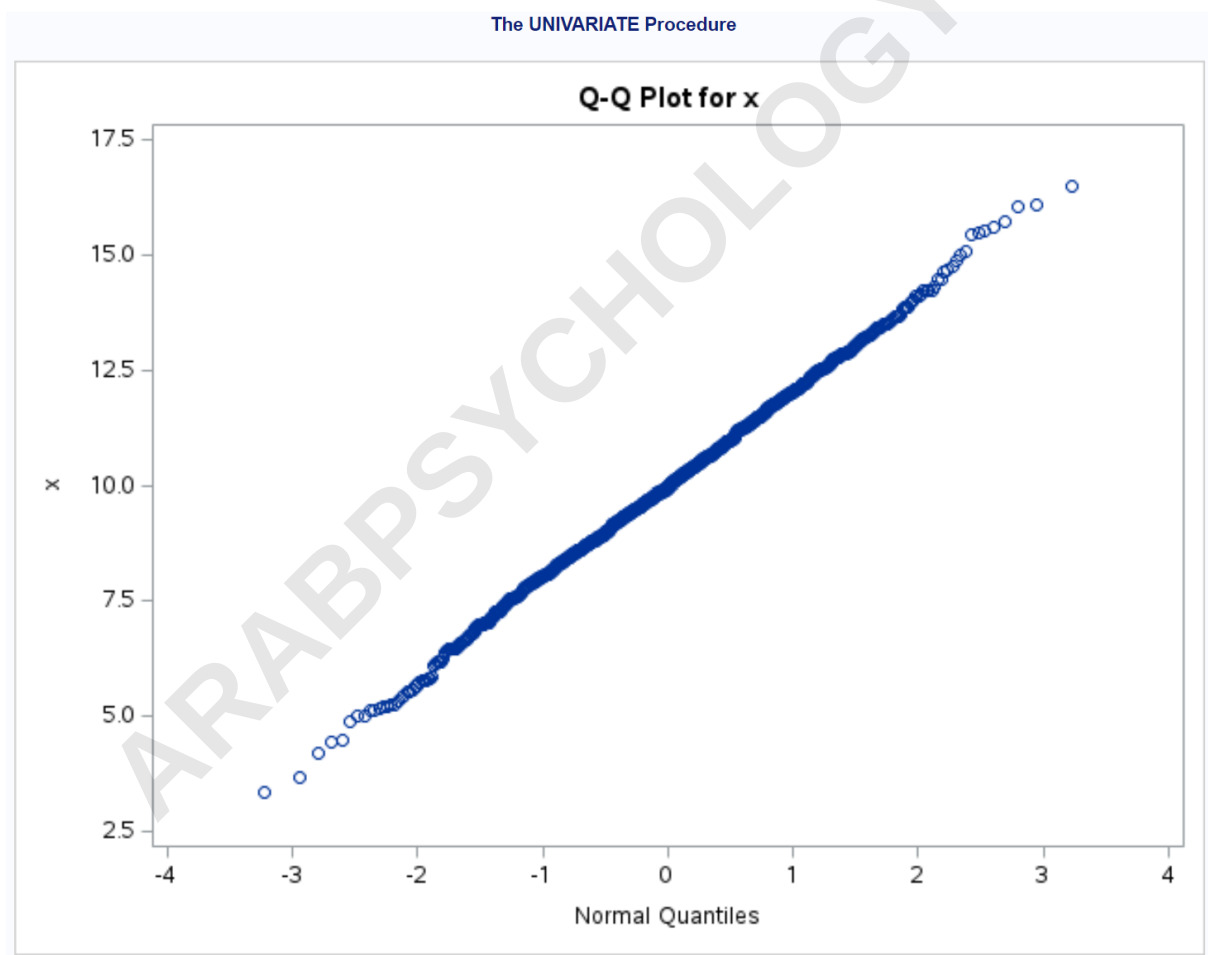
```
run;
```

```
/*create q-q plot*/
```

```
proc univariate data=normal_data noprint;
```

```
qqplot x;
```

```
run;
```



We can see that the points lie mostly along a straight diagonal line with some minor deviations along each of

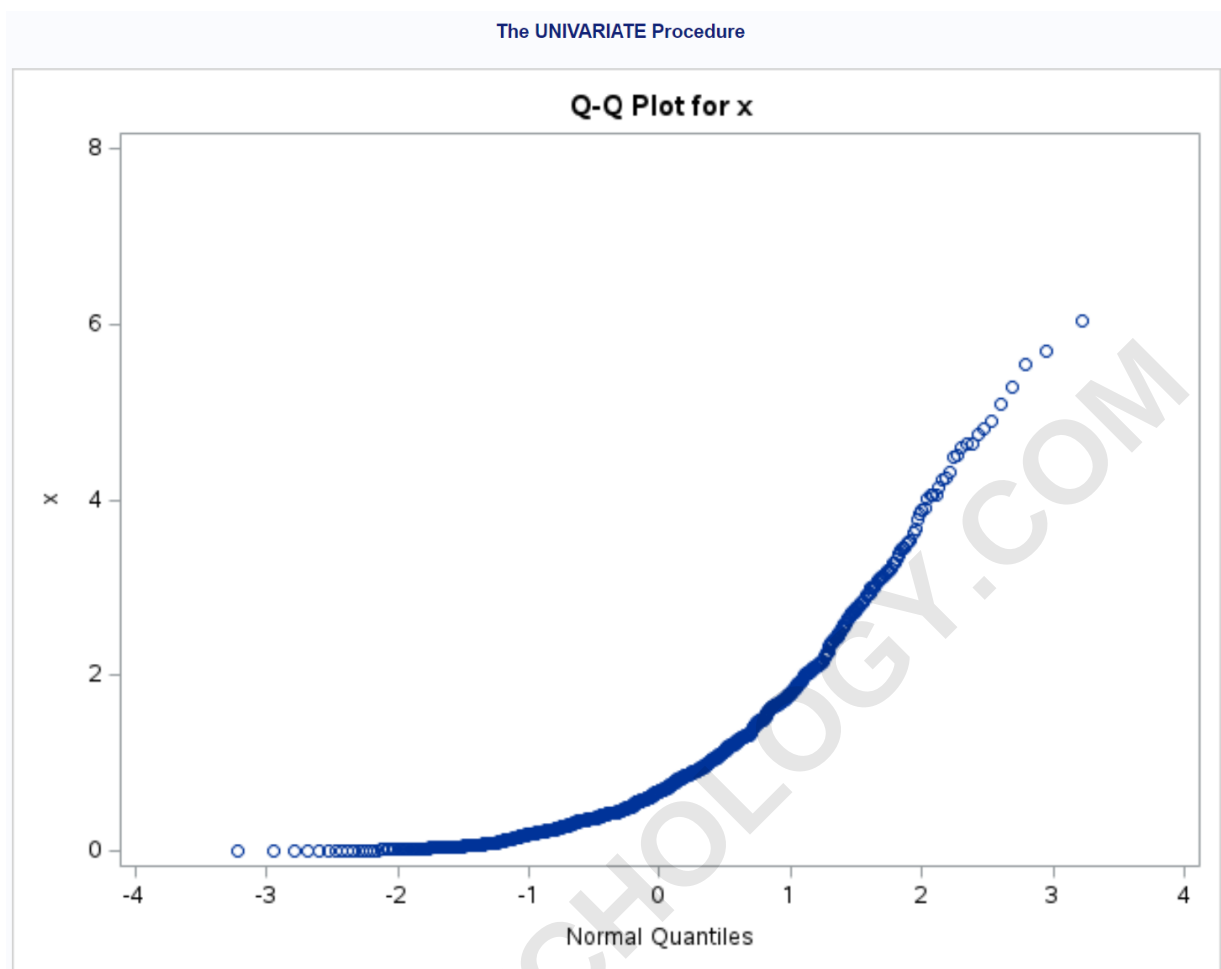
the tails.

Based on this plot, we could safely assume that this set of data is normally distributed.

Example 2: Q-Q Plot for Non-Normal Data

The following code shows how to create a Q-Q plot for a dataset that contains 1,000 observations generated from an :

```
/*generate 1000 values that follow an exponential  
distribution*/  
data exp_data;  
do i = 1 to 1000;  
x = ranexp(1);  
output;  
end;  
run;  
  
/*create q-q plot*/  
proc univariate data=exp_data noprint;  
qqplot x;  
run;
```



We can see that the points deviate significantly from a straight diagonal line. This is a clear indication that the set of data is not normally distributed.

This should make sense considering we specified that the data should follow an exponential distribution.

The following tutorials explain how to perform other common tasks in SAS: