How to Create and Interpret Q-Q Plots in Stata: A Step-by-Step Guide

Authored by

stats writer

March 9, 2026

Q-Q plots, short for quantile-quantile plots, are graphical tools used to assess the similarity between two probability distributions. In Stata, these plots can be created by using the "qnorm" command, which generates theoretical quantiles based on a specified distribution, and the "qplot" command, which plots the observed quantiles against the theoretical ones.

To create a Q-Q plot in Stata, the user must first specify the distribution to be compared to the observed data. This can be done by using the "qnorm" command and specifying the distribution's parameters, such as mean and standard deviation. Next, the "qplot" command is used to plot the observed quantiles against the theoretical ones, resulting in a scatter plot.

Interpretation of Q-Q plots in Stata involves visually assessing the points on the plot and comparing them to an ideal straight line. If the points lie close to the line, it indicates that the two distributions are similar. Deviations from the line suggest differences between the two distributions. Additionally, the user can also use statistical tests, such as the Kolmogorov-Smirnov test, to determine the significance of the differences between the distributions.

In summary, Q-Q plots in Stata are created by using the "qnorm" and "qplot" commands and can be interpreted by visually assessing the points on the plot and performing statistical tests. These plots are useful for comparing the distribution of a variable to a known distribution or to another variable's distribution.

Create and Interpret Q-Q Plots in Stata

A Q-Q plot, short for "quantile-quantile" plot, is often used to assess whether or not the residuals in a regression analysis are normally distributed.

This tutorial explains how to create and interpret a Q-Q plot in Stata.

Example: Q-Q Plot in Stata

For this example we will use the built-in *auto* dataset in Stata. We will fit a multiple linear regression model,

using *mpg* and *displacement* as the explanatory variables and *price* as the response variable. We will then obtain the residuals for the model and create a Q-Q plot to see if the residuals following a normal distribution.

Step 1: Load and view the data.

First, we'll load the data using the following command:

```
sysuse auto
```

Next, we'll get a quick summary of the data using the following command:

```
summarize
```

```
. sysuse auto
(1978 Automobile Data)

. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
make	0				
price	74	6165.257	2949.496	3291	15906
mpg	74	21.2973	5.785503	12	41
rep78	69	3.405797	.9899323	1	5
headroom	74	2.993243	.8459948	1.5	5
trunk	74	13.75676	4.277404	5	23
weight	74	3019.459	777.1936	1760	4840
length	74	187.9324	22.26634	142	233
turn	74	39.64865	4.399354	31	51
displacement	74	197.2973	91.83722	79	425
gear_ratio	74	3.014865	.4562871	2.19	3.89
foreign	74	.2972973	.4601885	0	1

Step 2: Fit the regression model.

Next, we'll use the following command to fit the regression model:

regress price mpg displacement

```
. regress price mpg displacement
```

Source	SS	df	MS		
Model	173587098	2	86793549.2	Number of obs	= 74
Residual	461478298	71	6499694.33	F(2, 71)	= 13.35
				Prob > F	= 0.0000
				R-squared	= 0.2733
				Adj R-squared	= 0.2529
Total	635065396	73	8699525.97	Root MSE	= 2549.4

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
mpg	-121.1833	72.78844	-1.66	0.100	-266.3193	23.95276
displacement	10.50885	4.58548	2.29	0.025	1.365658	19.65203
_cons	6672.766	2299.72	2.90	0.005	2087.254	11258.28

Step 3: Calculate the residuals.

Recall that a residual is simply the difference between the predicted response value (as calculated by the estimated regression equation) and the actual response value.

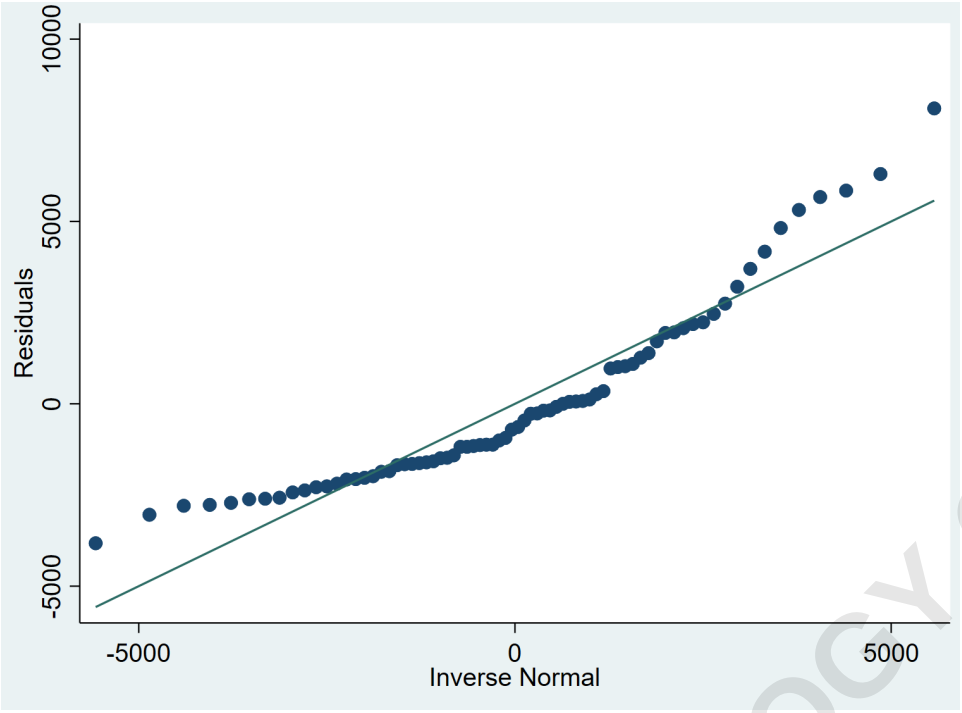
We can obtain the residuals of each prediction by using the residuals command and storing these values in a variable named whatever we'd like. In this case, we'll use the name resid_price:

```
predict resid_price, residuals
```

Step 4: Create the Q-Q Plot.

Now that we have a list of residuals, we can create a Q-Q plot using the qnorm command:

```
qnorm resid_price
```



Step 5: Interpret the Q-Q Plot.

The idea behind a Q-Q plot is simple: if the residuals fall along a roughly straight line at a 45-degree angle, then the residuals are roughly normally distributed. We can see in our Q-Q plot above that the residuals tend to deviate from the 45-degree line quite a bit, especially on the tail ends, which could be an indication that they're not normally distributed.

Although a Q-Q plot isn't a formal statistical test, it offers an easy way to visually check whether or not the residuals are normally distributed.

If it turns out that your residuals deviate severely from the 45-degree line in the Q-Q plot, you may consider performing a transformation on the response variable in your regression, such as using the square root or the log of the response variable.

If the residuals only deviate slightly, you don't need to worry about transforming the response variable as regression is fairly robust to departures from normality.