Training on Statistical Tools for Research: Stata

Graphical Presentation of Data in Stata

Md Monowar Hossain

Institute of Statistical Research and Training (ISRT)
University of Dhaka
mmhossain1@isrt.ac.bd

September 20, 2025

Graphical presentation of quantitative data

- A quantitative variable can be presented using different graphs among which histogram and boxplot are very popular.
- The former requires construction of frequency table and the later does not.
- A histogram looks like bar diagram, but it differs by the fact that rectangles touch one another and can be of different widths.
- However, in case of histogram, bar height can represent either the frequency or density.

Histogram

A histogram is used to show the **frequency distribution** of a quantitative variable. It looks similar to a bar chart, but the bars are adjacent (no gaps), and widths can vary.

```
hist bwt, frequency ///
title("Histogram for birth weight") ///
subtitle("(in gram)") ///
xtitle("Birth weight") ///
ytitle("Frequency")
```

Histogram (Variations)

- By default, histogram can show **fraction** (relative frequency) instead of raw frequency.
- You can change the **number of bins** or specify **bin width** and **starting point**.
- You may also overlay a **normal curve** to compare with a normal distribution.

```
hist bwt, fraction

hist bwt, bin(8) fraction normal

hist bwt, width(1000) fraction normal

hist bwt, width(1000) start(500) fraction normal
```

Boxplot

Boxplots summarize the distribution using the **five-number summary** (minimum, 1st quartile, median, 3rd quartile, maximum). They also highlight **outliers**.

```
graph box bwt

graph hbox bwt

graph box bwt, over(smoke)
```

- `graph box bwt` → vertical boxplot of birth weight.
- `graph hbox bwt` → horizontal boxplot.
- `graph box bwt, over(smoke)` → compares birth weight by smoking status.

Scatter Plot

Scatter plots show the relationship between **two quantitative variables**. Each dot represents one observation.

```
twoway scatter bwt age
```

- Here, `bwt` (birth weight) is plotted against age (mother's age).

Line Diagram

Line diagrams are used to track **changes over time** (time series data). Each data point is connected with a line, making it easy to see **trends**.

- Example: stock price over days
  - If prices rise, the line slopes upward.
  - If prices fall, the line slopes downward.

Line Diagram (Stata Example)

Using U.S. life expectancy data (1900–1999).

```
use https://www.stata-press.com/data/r17/uslifeexp
/*(U.S. life expectancy, 1900-1999)*/
line le year
line le le_male le_female year
```

- `line le year` → life expectancy over years.
- `line le le_male le_female year` → compares total, male, and female life expectancy trends.

Section 1

Others plot

Density Plot

A smooth curve that estimates the probability density of a variable. Useful for visualizing the shape of the distribution.

```
kdensity bwt, ///
title("Kernel density of Birth weight") ///
xtitle("Birth weight") ytitle("Density")
```

Dot Plot

Displays each observation as a dot along an axis, useful for smaller datasets.

```
dotplot bwt, ///
title("Dot plot of Birth weight") ///
xtitle("Birth weight")
```

Cumulative Distribution Plot (CDF)

Shows the probability that the variable takes a value less than or equal to x.

```
cumul bwt, gen(cumbwt)
line cumbwt bwt, ///
title("Cumulative distribution of Birth weight") ///
xtitle("Birth weight") ytitle("Cumulative probability")
```

Quantile-Quantile (Q-Q) Plot

Checks whether the variable follows a normal distribution by comparing quantiles.

```
qnorm bwt
```

Probability Plot (P-P Plot)

Another normality check, comparing cumulative probabilities of observed data with a normal distribution.

```
pnorm bwt
```

Scatter Plot with Fitted Line

Shows relationship between two quantitative variables with a regression line.

```
twoway (scatter bwt age) (lfit bwt age), ///
title("Birth weight vs Mother's age with fitted line")
```

Scatter Plot with Smoother (Lowess Curve)

Adds a smooth line to capture nonlinear trends in scatter data.

```
twoway (scatter bwt age) (lowess bwt age), ///
title("Scatter plot with Lowess smoother")
```

Scatter Matrix

Displays pairwise scatter plots for multiple variables at once.

```
graph matrix bwt age lwt, ///
title("Scatter plot matrix")
```

Violin Plot

Combines density plot and boxplot to show distribution and summary statistics. *(Requires installation of `vioplot` package from SSC.)*

```
ssc install vioplot
vioplot bwt, over(smoke)
```

Combined Graphs (Overlaying Multiple Types)

Mixes scatter plot, regression line, and confidence bands for deeper insight.

```
twoway (scatter bwt age) (lfitci bwt age), ///
title("Scatter with regression line & CI")
```

THANK YOU