STATISTICAL TOOLS FOR ANALYSIS

Fundamentals of Developing a Water Quality Monitoring Plan

Michael Schramm

2024-03-05

TWRI, Texas A&M AgriLife Research

Statistical Tools for Analysis

- Base concepts
- Graphical analysis and data exploration
- Statistical design for watershed studies

Base concepts

Exploratory Data Analysis

First step in any data analysis is to plot your data.

- Graphical methods provide quick visual summaries of data.
- Easily interpreted.
- Describes essential information more easily than numbers alone.

Exploratory Data Analysis

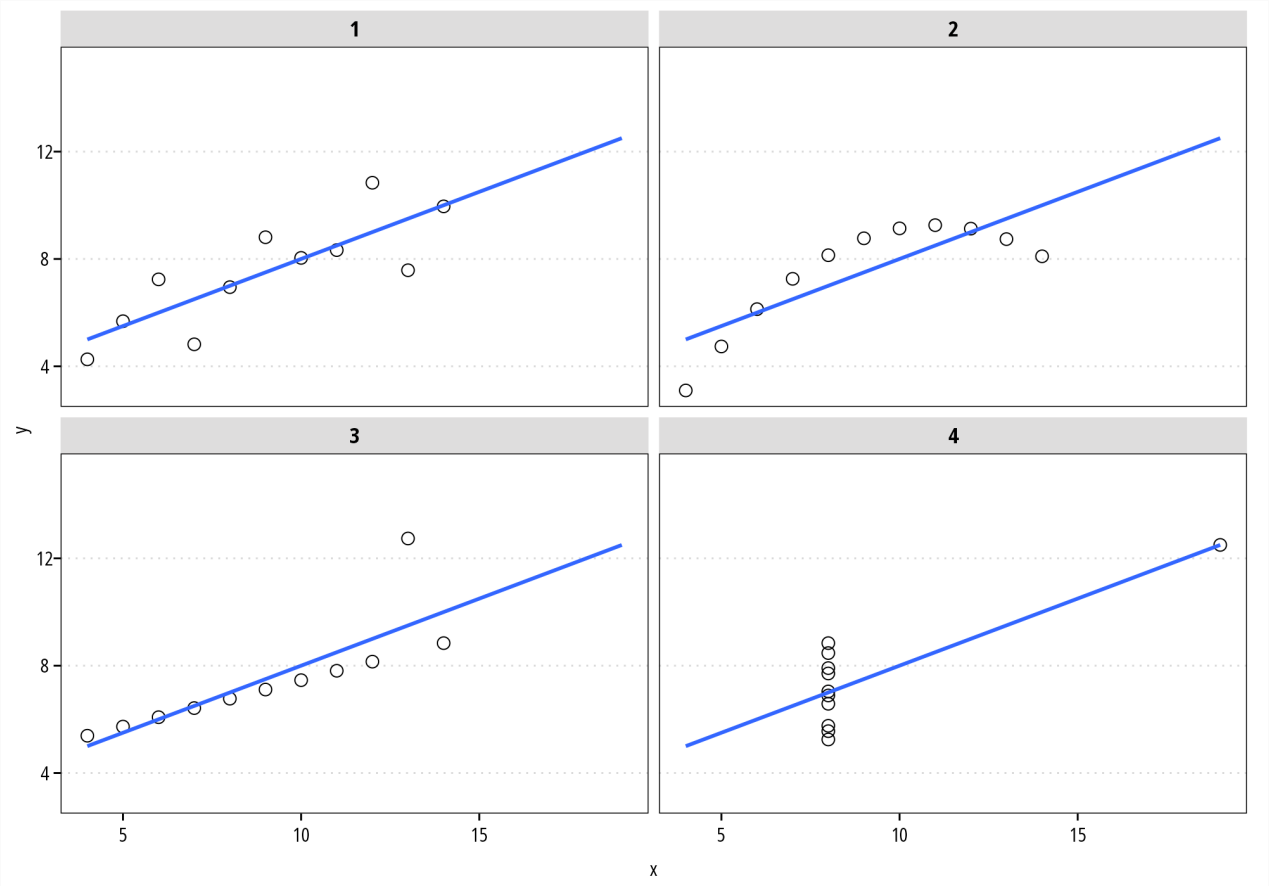


Figure 1: Four different datasets with the same mean, variance, correlation, slope and intercept. Dataset is known as Anscombe's quartet.

Available graphical methods

- Histograms and density plots
- Quantile plots (cumulative density function)
- Boxplots
- Probability plots
- Scatterplots

Histograms and density plots

- Histograms plot the count of observed values within equally spaced bins.
- Displays the distribution, skewness, and variability of the data.
- Density plots are smoothed versions of histograms.

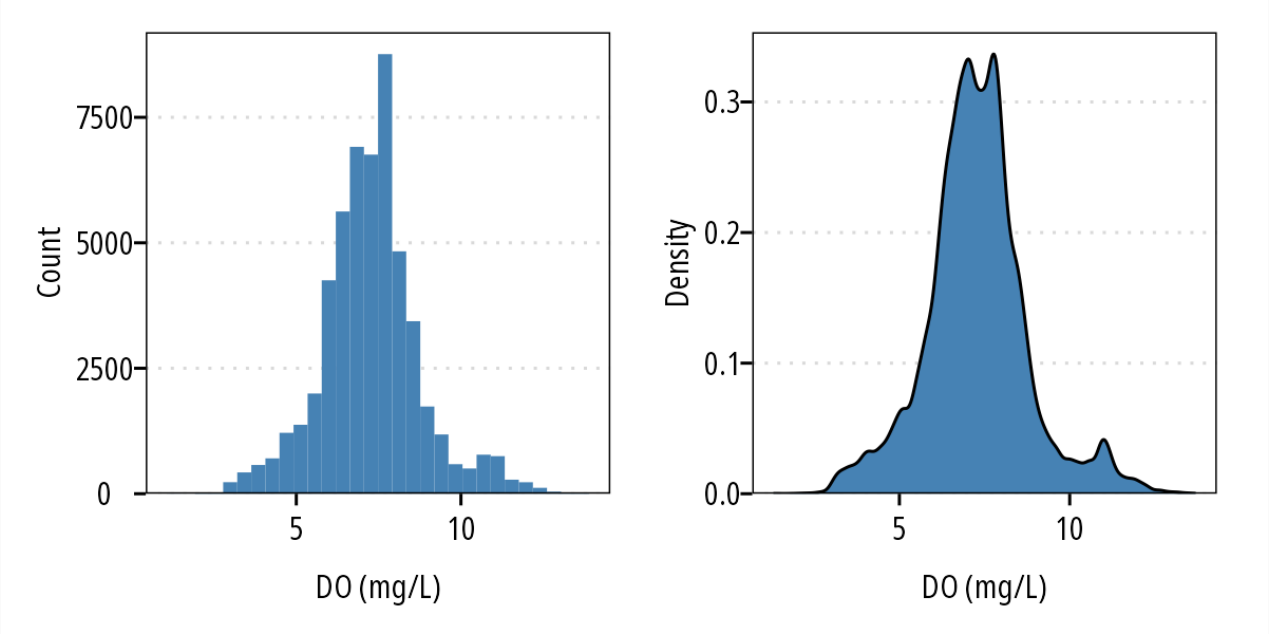


Figure 2: Histogram and density plot of 15-minute DO measurements.

Quantile plots

- Provides information about the distribution of observed values.
- Shows the probability that a random variable will be less than or equal to specific value x.
- Also called empirical cumulative distribution functions (ecdf).
- A flow duration curve is an inverse version of the ecdf using descending ranks instead of ascending ranks.

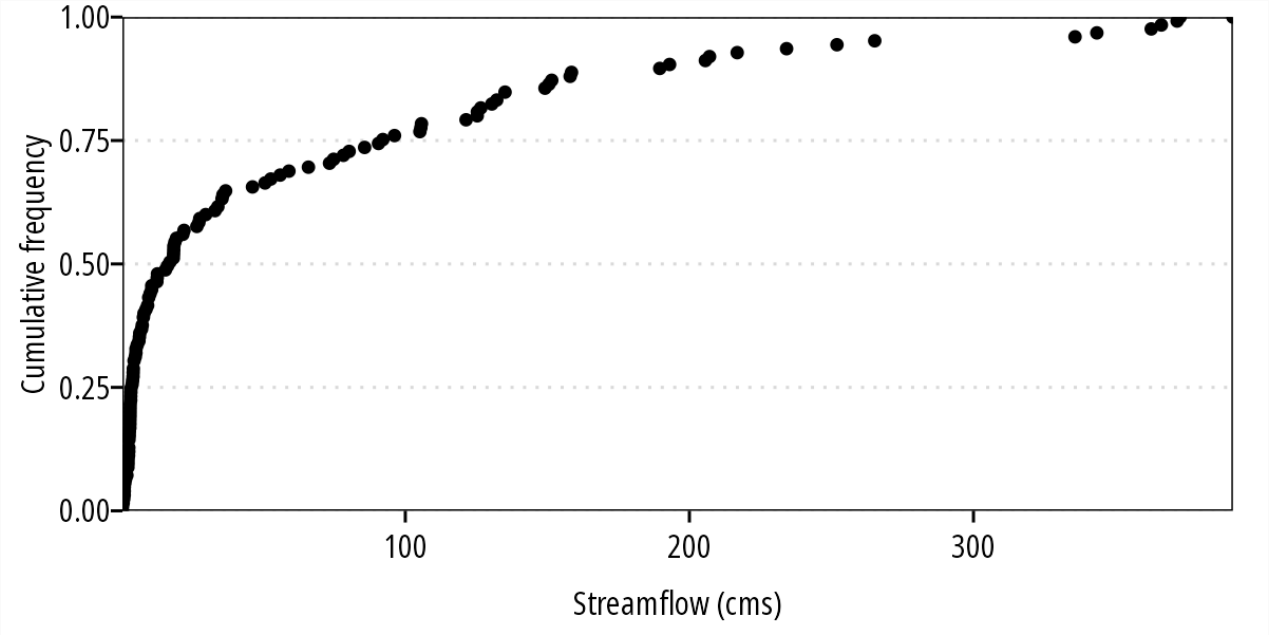


Figure 3: Quantile plot of 2-years of mean daily streamflow values

Boxplots

Boxplots are concise displays of the median, variation, skew, and outliers. These can also be used to compare attributes between datasets or sites.

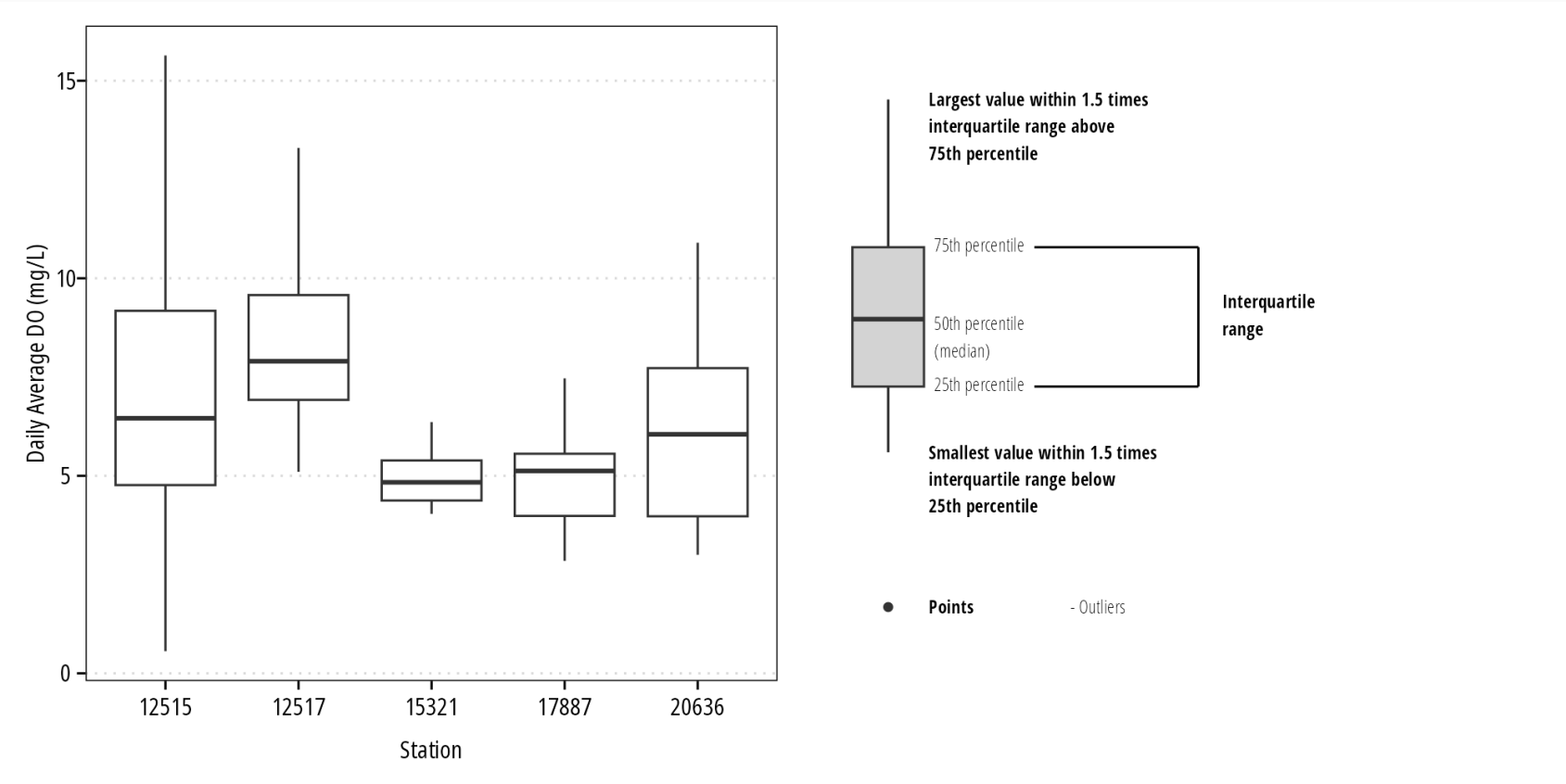
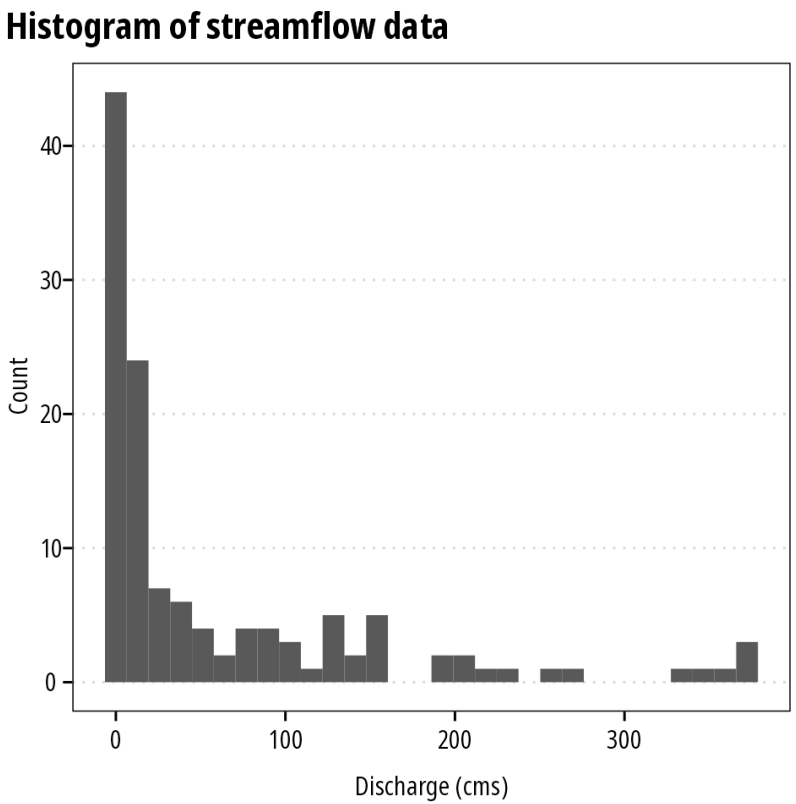
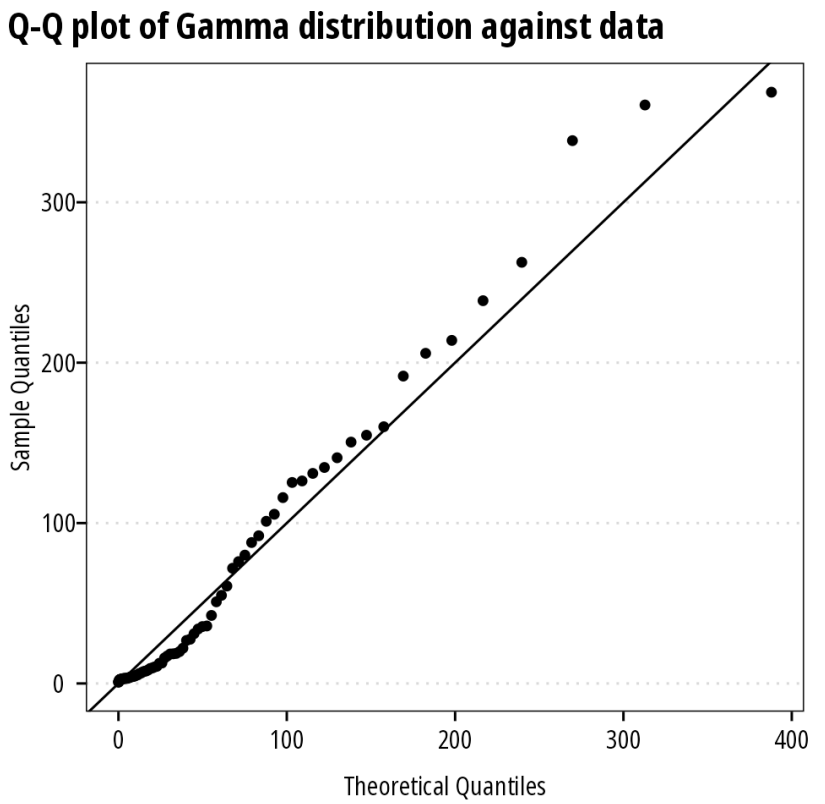


Figure 4: Boxplots of dissolved oxygen concentrations at 5 sites.

Probability plots

Also called a quantile-quantile (Q-Q) plot. This is the quantile plot generated earlier plotted against quantiles from a theoretical distribution. These are used to evaluate how well the data fits against distributions such as the normal, log-normal, or gamma distribution.





Scatterplots

Typically used to plot two continous variables against each other, or a continuous variable over time.