Math 70: Fundamentals of Statistics

Farhan Sadeek

Spring 2026

Contents

1 QQ Plots (Quantile-Quantile Plots)

1 What is a QQ Plot?

A **QQ plot** (quantile-quantile plot) is a graphical tool used to assess whether a dataset follows a particular theoretical distribution, most commonly the normal distribution. It works by plotting the **quantiles** of the observed data against the quantiles of the theoretical distribution.

If the data follow the theoretical distribution, the points in the QQ plot will fall approximately along a straight line. Deviations from this line indicate departures from the assumed distribution.

2 Construction

Given a sample of n observations $x_1, x_2, \ldots, x_n$, we construct a **normal QQ plot** as follows:

1. **Order the data.** Sort the observations in increasing order to obtain the **order statistics**:

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}.$$

2. **Compute theoretical quantiles.** For each $i = 1, 2, \ldots, n$, compute the theoretical quantile

$$q_i = \Phi^{-1}\left(\frac{i - 0.5}{n}\right),$$

   where Φ^{-1} is the **quantile function** (inverse CDF) of the standard normal distribution. The shift by 0.5 is a continuity correction that avoids mapping to $\pm\infty$ at the endpoints.

3. **Plot.** Plot the points $(q_i, x_{(i)})$ for $i = 1, \ldots, n$, with theoretical quantiles on the horizontal axis and sample quantiles on the vertical axis.

3 Interpretation

3.1 Ideal Case: Data are Normal

If $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, then the order statistics are approximately

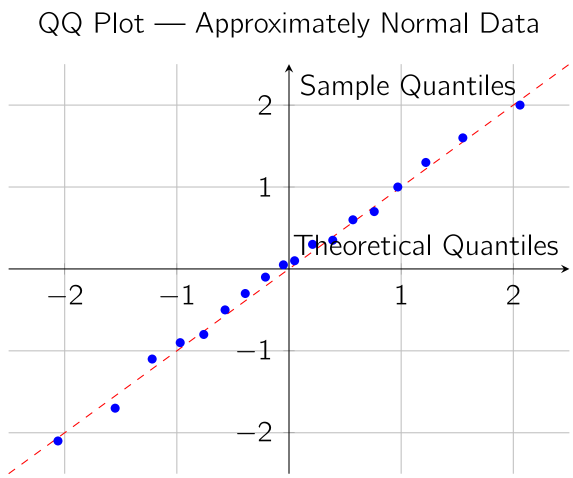
$$x_{(i)} \approx \mu + \sigma \cdot q_i.$$

This means the QQ plot will be approximately linear with slope σ and intercept μ. A perfectly normal sample produces a straight line with slope equal to the standard deviation and y-intercept equal to the mean.

3.2 Common Departures

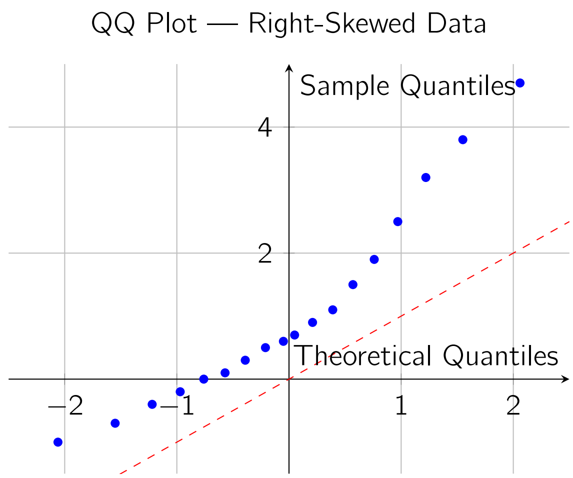
Pattern	Interpretation
Points follow a straight line	Data are approximately normal
S-shaped curve (concave up then concave down)	Heavy tails (leptokurtic)
Reverse S-shape	Light tails (platykurtic)
Curve bending upward on the right	Right skew (positive skew)
Curve bending downward on the left	Left skew (negative skew)
Staircase pattern	Discrete or rounded data

4 Example: Normal Data



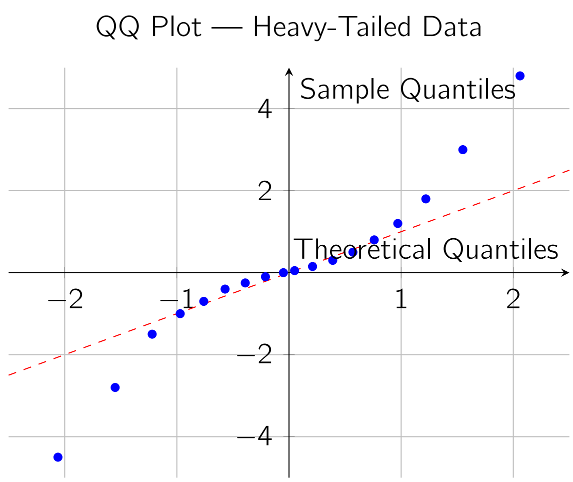
The points hug the reference line closely, suggesting the data are well-approximated by a normal distribution.

5 Example: Right-Skewed Data



The upward bend on the right side indicates that the upper tail of the data is heavier than a normal distribution—the data are **right-skewed**.

6 Example: Heavy-Tailed Data



The S-shape (points below the line on the left and above the line on the right) indicates **heavy tails**—the data have more extreme values than a normal distribution would predict. This is characteristic of distributions like the t-distribution with few degrees of freedom.

7 Mathematical Justification

The QQ plot exploits the **probability integral transform**. If $X \sim F$ for some continuous CDF F, then $F(X) \sim \text{Uniform}(0,1)$. Equivalently, if $U \sim \text{Uniform}(0,1)$, then $F^{-1}(U) \sim F$.

For a sample $x_{(1)} \leq \cdots \leq x_{(n)}$, we expect

$$x_{(i)} \approx F^{-1}\left(\frac{i-0.5}{n}\right).$$

If $F = \Phi$ (the standard normal CDF), this gives us

$$x_{(i)} \approx \Phi^{-1}\left(\frac{i-0.5}{n}\right) = q_i.$$

So plotting $x_{(i)}$ vs. q_i should yield points near the line $y = x$ when the data are standard normal, or near the line $y = \mu + \sigma x$ when the data are $N(\mu, \sigma^2)$.

8 QQ Plot vs. Histogram

While histograms are useful for visualizing the shape of a distribution, QQ plots are generally **more sensitive** to departures from normality, especially in the tails. The advantages include:

- **No binning artifacts**: histograms depend on bin width choice; QQ plots do not.
- **Better tail assessment**: the tails of the distribution are spread out and easy to inspect.
- **Direct comparison**: you compare directly against the theoretical distribution rather than making a visual judgment about bell-curve shape.

9 General QQ Plots

Although the normal QQ plot is the most common, the same idea applies to *any* reference distribution F_0. Replace Φ^{-1} with F_0^{-1} and plot $(F_0^{-1}((i-0.5)/n), x_{(i)})$. Common choices include:

- **Exponential QQ plot**: to check if data follow an exponential distribution.
- **Uniform QQ plot**: to check if data are uniformly distributed.
- t**-distribution QQ plot**: to check for heavy-tailed behavior with a specific degrees of freedom.
- **QQ plot of two samples**: plot the quantiles of one sample against the quantiles of another to compare their distributions directly (no theoretical reference needed).