STAT 2593

Lecture 020 - Probability Plots

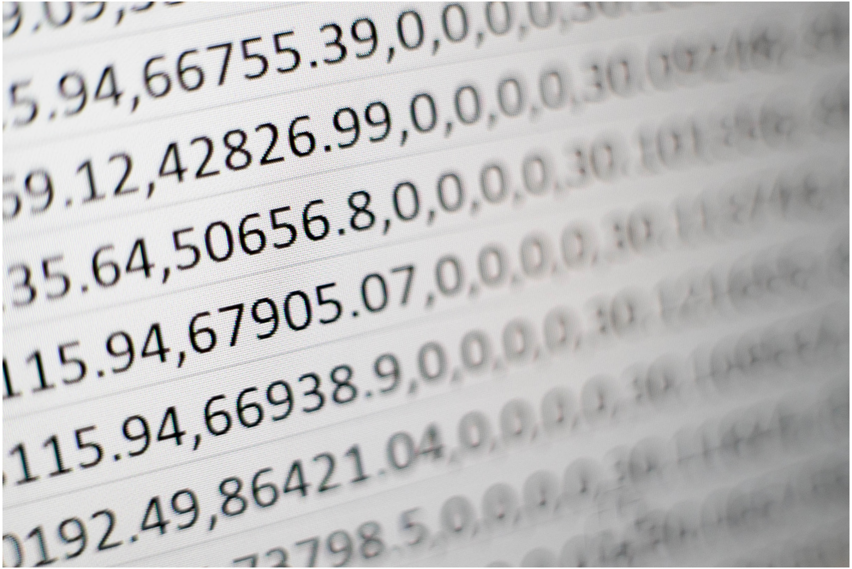
Dylan Spicker

Probability Plots

Learning Objectives

1. Understand the construction and use of a probability plot.

2. Read and draw conclusions from probability plots.


5.94,66755.39,0,0,0,0
59.12,42826.99,0,0,0,0
35.64,50656.8,0,0,0,0
115.94,67905.07,0,0
115.94,66938.9
192.49,86421.04

Probability Plots

- If we want to test whether data follows a particular distribution, we can plot the histogram.

Probability Plots

- If we want to test whether data follows a particular distribution, we can plot the histogram.
  - This is likely to be biased, and impacted by small deviations or choices made during plotting.

Probability Plots

- If we want to test whether data follows a particular distribution, we can plot the histogram.
  - This is likely to be biased, and impacted by small deviations or choices made during plotting.
- Instead, we commonly rely on **probability plots**.

Probability Plots

- If we want to test whether data follows a particular distribution, we can plot the histogram.
  - This is likely to be biased, and impacted by small deviations or choices made during plotting.
- Instead, we commonly rely on **probability plots**.
  - Often these are called QQ plots instead.

Probability Plots

- If we want to test whether data follows a particular distribution, we can plot the histogram.
  - This is likely to be biased, and impacted by small deviations or choices made during plotting.
- Instead, we commonly rely on **probability plots**.
  - Often these are called QQ plots instead.
- The idea with a probability plot is that, if our data are drawn from a particular distribution, the sample percentiles should be approximately equal to the theoretical percentiles of that distribution.

Probability Plots

- If we want to test whether data follows a particular distribution, we can plot the histogram.
  - This is likely to be biased, and impacted by small deviations or choices made during plotting.
- Instead, we commonly rely on **probability plots**.
  - Often these are called QQ plots instead.
- The idea with a probability plot is that, if our data are drawn from a particular distribution, the sample percentiles should be approximately equal to the theoretical percentiles of that distribution.
- If we compare the sample percentiles to the theoretical ones we can assess whether a particular distribution fits.

Sample Percentiles

- If we order the data from smallest to largest, these values roughly correspond to the observed percentiles.

Sample Percentiles

- If we order the data from smallest to largest, these values roughly correspond to the observed percentiles.
  - But which percentiles? Different sources give different answers.

Sample Percentiles

- If we order the data from smallest to largest, these values roughly correspond to the observed percentiles.
  - But which percentiles? Different sources give different answers.
  - In this course we will say that the i-th smallest value is the $100\frac{i-0.5}{n}$-th percentile.

Sample Percentiles

- If we order the data from smallest to largest, these values roughly correspond to the observed percentiles.
  - But which percentiles? Different sources give different answers.
  - In this course we will say that the i-th smallest value is the $100\frac{i-0.5}{n}$-th percentile.
  - If $n = 10$ then the smallest value corresponds to 5%, the largest value to 95%, and so on.

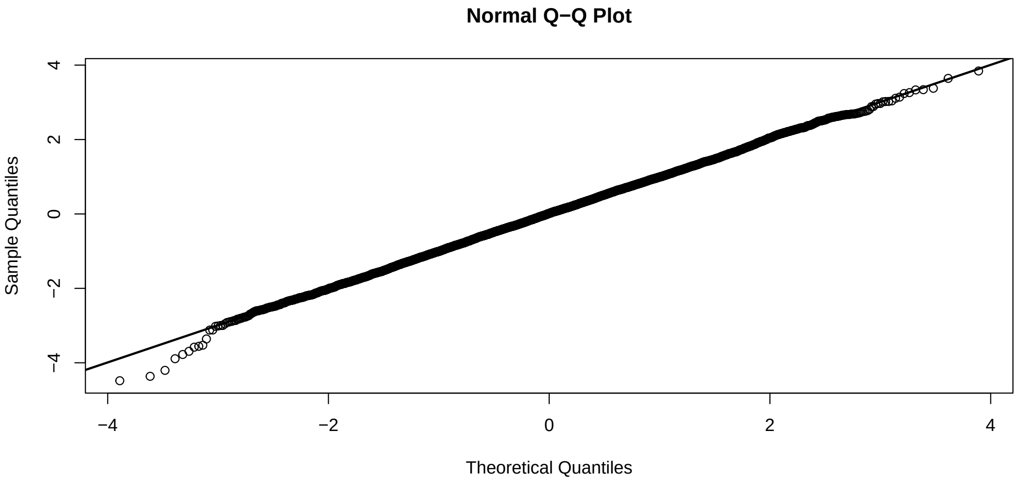
Sample Percentiles

- If we order the data from smallest to largest, these values roughly correspond to the observed percentiles.
  - But which percentiles? Different sources give different answers.
  - In this course we will say that the i-th smallest value is the $100\frac{i-0.5}{n}$-th percentile.
  - If $n = 10$ then the smallest value corresponds to 5%, the largest value to 95%, and so on.
- With the sample percentiles computed, we could then compute the corresponding theoretical percentiles for the distribution that we wish to test against.

Sample Percentiles

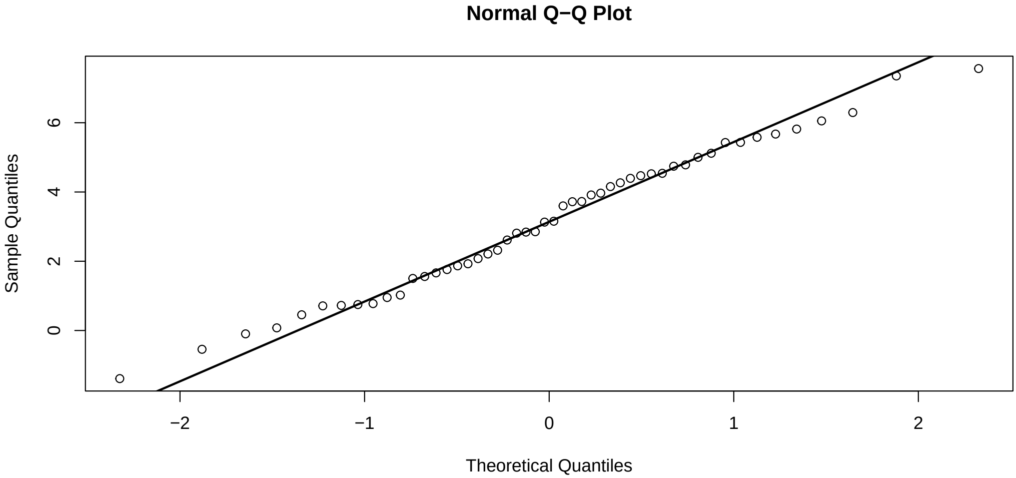
- If we order the data from smallest to largest, these values roughly correspond to the observed percentiles.
  - But which percentiles? Different sources give different answers.
  - In this course we will say that the i-th smallest value is the $100\frac{i-0.5}{n}$-th percentile.
  - If $n = 10$ then the smallest value corresponds to 5%, the largest value to 95%, and so on.
- With the sample percentiles computed, we could then compute the corresponding theoretical percentiles for the distribution that we wish to test against.
- If we simply plot these against one another, then data which follows the distribution should fall along a straight line.

Example 1



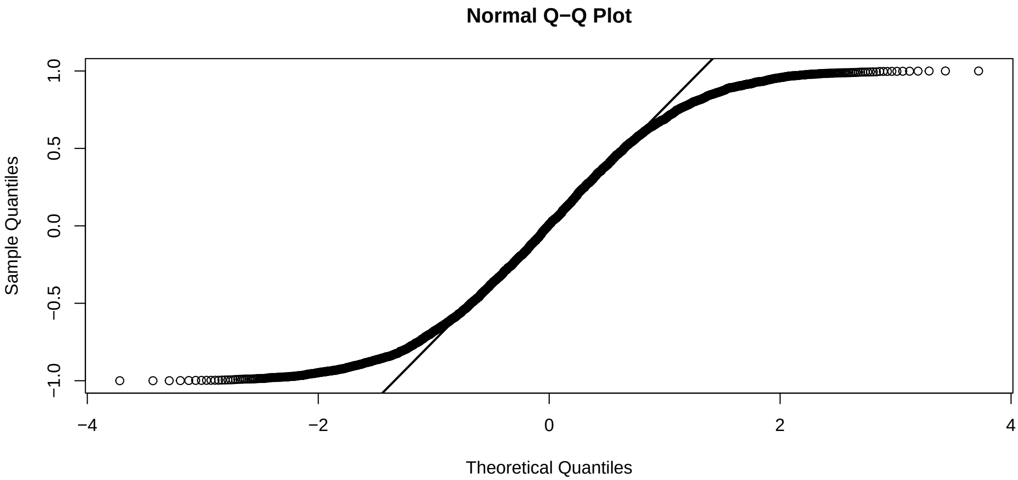
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4
−2
0
2
4

Example 2



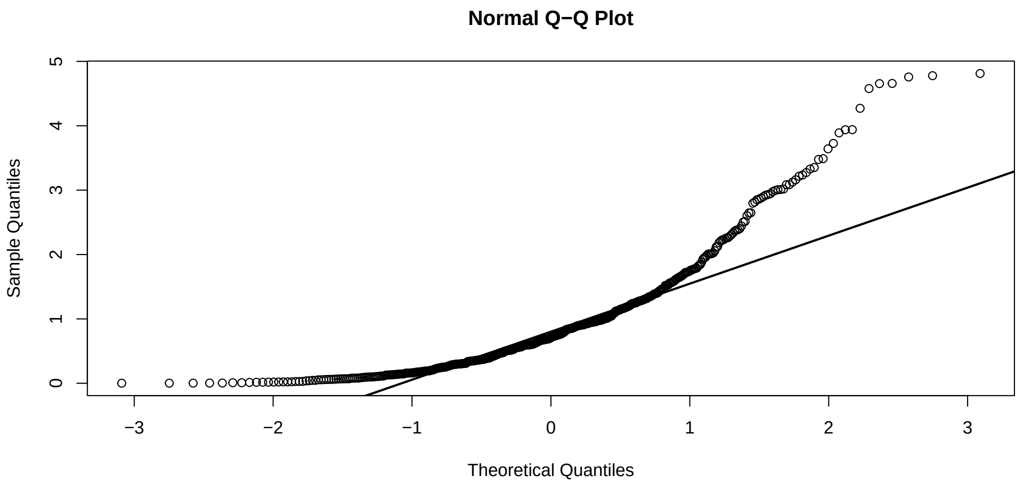
Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
0
2
4
6
−2
−1
0
1
2

Example 3

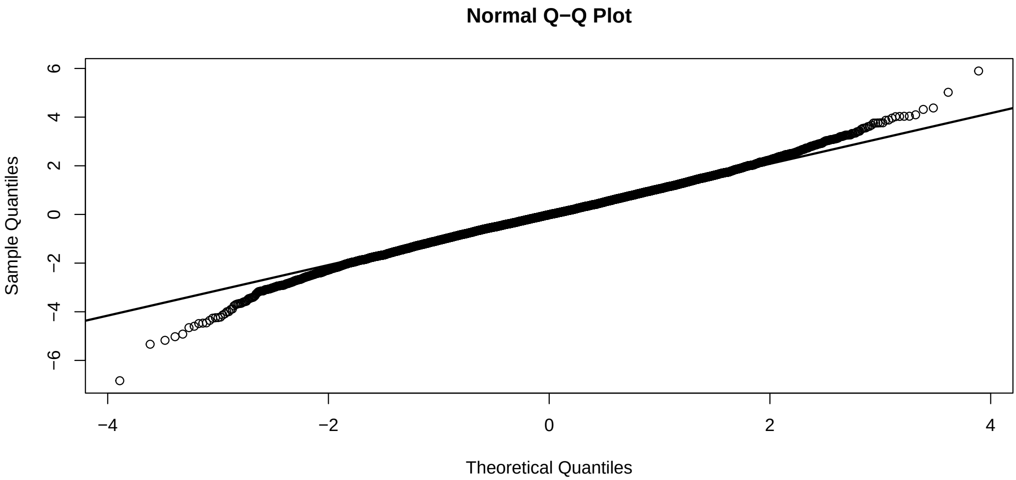


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
1.0
0.5
0.0
−0.5
−1.0
−4
−2
0
2
4

Example 4


Normal Q−Q Plot
Sample Quantiles
0
1
2
3
4
5
−3
−2
−1
0
1
2
3
Theoretical Quantiles

Example 5



Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
6
4
2
0
−2
−4
−6
−4
−2
0
2
4

Summary

- Probability plots plot the sample percentiles against theoretical percentiles.
- Probability plots are useful to determine whether data (approximately) follows a given distribution.
- Data which corresponds to a given distribution should fall on an (approximately) straight line.