Chapter 6 Normal Probability Distributions
Assessing Normality (6.5)

Objectives

In this lesson, we will learn how to

- decide whether it is reasonable to assume that sample data come from a normally distributed population;
- use a histogram to look for an approximately bell-shaped pattern;
- use a normal quantile plot to assess normality;
- recognize examples of normal, uniform, and skewed distributions;
- understand the idea of formal normality tests and data transformations.

1 Why Assess Normality?

Many important statistical methods require that sample data come from a population with an approximately normal distribution.

In this section, we use two main visual tools:

Two Basic Tools

1. Construct a histogram and check whether it is roughly bell-shaped.
2. Construct a normal quantile plot and check whether the points are reasonably close to a straight line.

2 Normal Quantile Plot

A normal quantile plot, also called a normal probability plot, is a graph of points (x, y), where

- each x-value is from the original sample data;
- each y-value is the corresponding z score expected from the standard normal distribution.

Key Idea

If the data are from a normally distributed population, the points in the normal quantile plot should be reasonably close to a straight line.

3 Procedure for Assessing Normality

Step 1: Use a Histogram

Construct a histogram. If the histogram departs dramatically from a bell shape, conclude that the data do not appear to come from a normally distributed population.

Step 2: Use a Normal Quantile Plot

If the histogram is basically symmetric, use technology to generate a normal quantile plot.

Normal Distribution Criteria

It is reasonable to assume that the population distribution is normal if:

- the points are reasonably close to a straight line, and
- the points do not show a systematic pattern that is not a straight-line pattern.

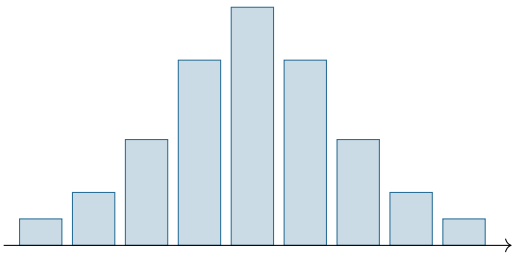
Not Normal Criteria

The population distribution is probably not normal if either of the following occurs:

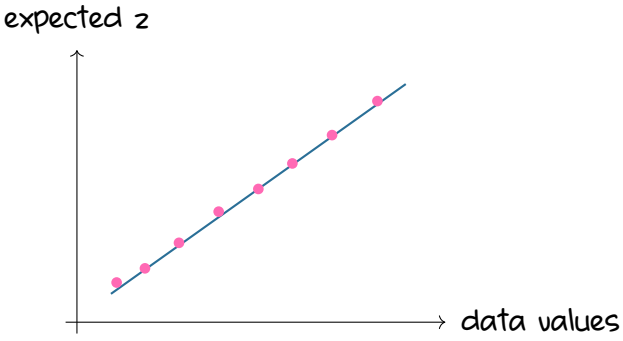
- the points do not lie reasonably close to a straight line;
- the points show a systematic pattern that is not a straight line.

4 Visual Examples

Example 1: Normal Pattern: A sample of IQ scores has a histogram that is close to bell-shaped. The corresponding normal quantile plot has points close to a straight-line pattern.



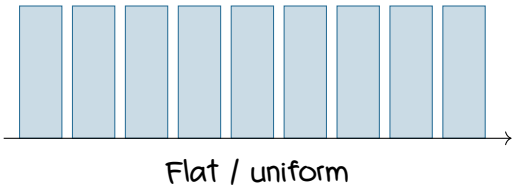
Roughly bell-shaped



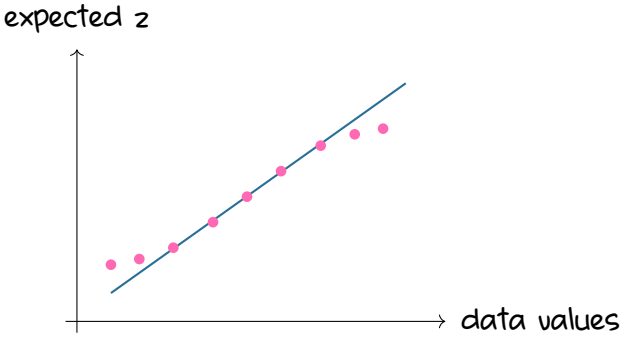
Solution

The histogram is roughly bell-shaped, and the normal quantile plot is close to a straight line. It is reasonable to treat the data as coming from a normally distributed population.

Example 2: Uniform Pattern: A sample has a flat histogram. The corresponding normal quantile plot may look somewhat straight in the middle, but it shows a systematic pattern.



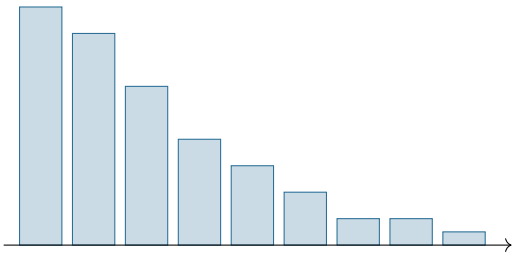
Flat / uniform



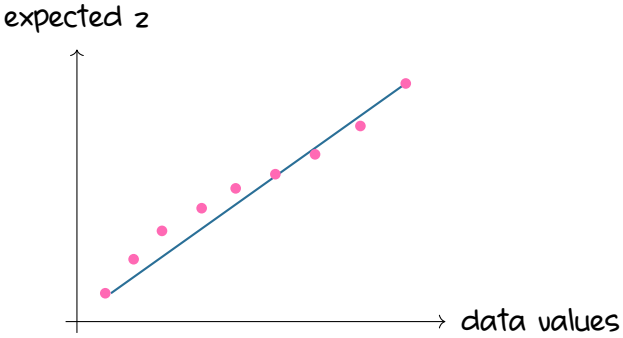
Solution

Because the histogram is flat rather than bell-shaped, and because the normal quantile plot shows a systematic pattern, the data should not be treated as normally distributed.

Example 3: Skewed Pattern: A sample of rainfall amounts is skewed to the right. The normal quantile plot has points that are not close to a straight line.



Skewed right



Solution

Because the histogram is clearly skewed and the normal quantile plot is far from a straight-line pattern, the data do not appear to come from a normally distributed population.

5 Tools for Determining Normality

Histogram and Outliers

If the requirement of normality is not too strict, we can often use a histogram and check for outliers.
If the histogram is roughly bell-shaped and the number of outliers is 0 or 1, we may treat the population as approximately normal.

Normal Quantile Plot

Normal quantile plots are usually generated by technology such as StatCrunch, Excel, Minitab, Statdisk, or a graphing calculator.

Advanced Normality Tests

More advanced tests for normality include:

- chi-square goodness-of-fit test;
- Kolmogorov-Smirnov test;
- Lilliefors test;
- Anderson-Darling test;
- Jarque-Bera test;
- Ryan-Joiner test.

6 Manual Construction of a Normal Quantile Plot

Manual Steps

1. Sort the data from lowest to highest.
2. For a sample size n, find cumulative left areas corresponding to the sorted values.
3. Use the standard normal distribution to find the z scores corresponding to those cumulative left areas.
4. Match each sorted data value with its corresponding expected z score.
5. Plot the points (x, y), where x is the original data value and y is the expected z score.
6. Decide whether the points are reasonably close to a straight line.

Practical Note

In class and on homework, it is usually best to use technology to create normal quantile plots. The main goal is to interpret the plot correctly.

7 Ryan-Joiner Test

The Ryan-Joiner test is one formal test of normality.

Statdisk Normality Assessment

Statdisk can display:

- a histogram;
- a normal quantile plot;
- the number of potential outliers;
- results from the Ryan-Joiner test.

Example 4: Dallas Commute Times: A Statdisk Normality Assessment is performed for 1000 Dallas commute times. The histogram is skewed to the right, and the normal quantile plot is far from a straight-line pattern. What should we conclude?

Solution

The Dallas commute times do not appear to be from a population with a normal distribution. The histogram is not bell-shaped, and the normal quantile plot does not show points close to a straight line.

8 Data Transformations

Many data sets are not normally distributed, but sometimes we can transform the data so that the transformed values are approximately normal.

Log Transformation

One common transformation is to replace each original value x by

$$\log(x).$$

If any original values are 0, we can use

$$\log(x+1)$$

instead.

If the logarithms of the values have a normal distribution, the original values are said to have a lognormal distribution.

Summary

Key Ideas

- Some statistical methods require data from an approximately normal population.
- A roughly bell-shaped histogram supports normality.
- A normal quantile plot supports normality when the points are reasonably close to a straight line.
- A systematic non-straight-line pattern suggests that the data are not normal.
- Formal tests, such as the Ryan-Joiner test, can also help assess normality.
- Some non-normal data can be transformed, such as by using logarithms.