Essential statistics for the pharmaceutical sciences

Dr Phil Rowe BSc, MSc, PhD, FRSS

Publisher: John Wiley & Sons Ltd (http://eu.wiley.com)

Published: 2007

ISBN: 978-0-470-03468-2

Additional topics as supplements to the book:

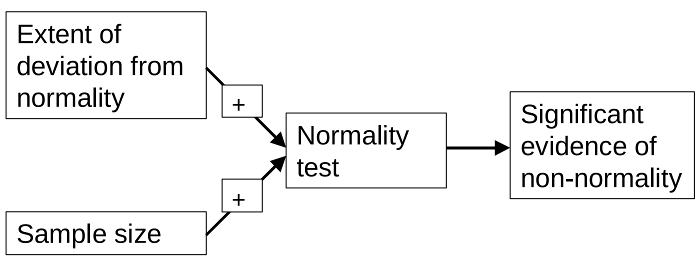
Statistical tests for normal distribution – why you shouldn't do them.

Statistical tests for normal distribution – why you shouldn't do them.

For many statistical tests it is a pre-requisite that the data are consistent with a normal distribution (especially t-tests and analyses of variance). Various statistical tests are available which can allegedly be used to test for normality. These include the Shapiro-Wilk, Ryan-Joiner, Anderson-Darling and the wonderfully named Kolmogorov-Smirnov tests. However, in the book (Chapter 3) I recommended a simple visual inspection of the data rather than any of these formal tests. It was not possible to explain the preference at that stage as this would have required an understanding of statistical power which is not introduced until Chapter 8. Assuming you now understand the idea of power, an explanation is appropriate.

All the available tests have a null hypothesis that the data are normally distributed and a significant outcome is then evidence of non-normality. The likelihood of a significant result depends upon the extent of any non-normality that may be present and the amount of data (See Fig 1).

Figure 1 Statistical test for normal distribution



There are two obvious ways in which inappropriate conclusions can be arrived at.

1) Failure to detect serious non-normality in a small data set.

Even gross non-normality will not produce a significant result if the data set is very small. This would not matter too much if investigators could be relied upon to interpret non-significance properly. If they simply said 'There was no significant evidence one way or the other.' no great harm would be done. But in reality, a non-significant result tends to be taken as a licence to assume that the data are normally distributed and that t-tests and ANOVAs can therefore be relied upon.

A non-significant result never proves that the null hypothesis is true; it just means that it *might* be true. In the current context, we ought to conclude that the data *may be* normally distributed, but they are certainly not proven to be.

2) Mild non-normality within a large data set leading to a significant outcome.

ANOVAs and t-tests are pretty robust and moderate non-normality won't distort their outcomes to an extent that need be of practical concern. However, if the data set is large enough, even the most trivial departure from normality will be detected. Such an outcome probably shouldn't deter us from using a t-test etc, but if you have gone to the trouble of carrying out such a test, are you just going to ignore the result?

Outcomes are dependent upon sample size

Kolmogorov-Smirnov (like all the others) is as much a test of sample size as of normal distribution. Very few data-sets perfectly follow a normal distribution. If your data set is big enough you will detect 'Non-normality' and be deterred from using what might be a perfectly appropriate test. On the other hand, if it's small enough you will be seduced into believing it's normal and end up using a t-test/ANOVA even if you shouldn't.

The visual inspection recommended in Chapter 3 will allow you to avoid any seriously flawed conclusions arising from the inappropriate use of t-tests etc.

Two example data sets

Set out below are two data sets that illustrate the twin perils of normality testing.

Data set 1. Markedly non-normal, but too small to achieve significance.

8.53
25.24
88.10
21.12
41.83
10.45

Mean 32.5 ±29.7

With such a small data set, the usual approach advocated in Chapter 3 - a histogram - is not going to be practical, but a dot plot (Fig 2) is shown.

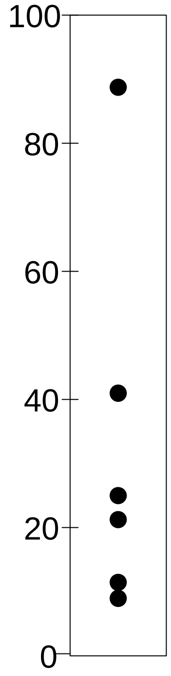


Figure 2
Dot plot of data set 1

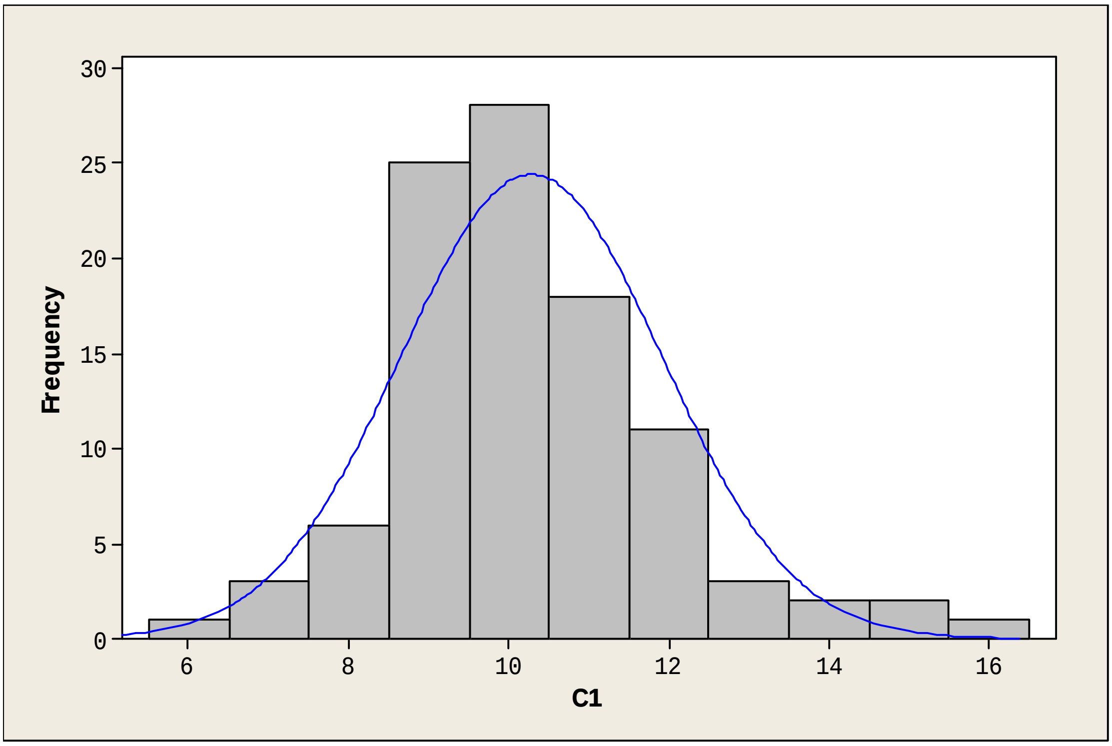
There is too little data to say anything with certainty, but there is a very strong smell of positive skew. We have a cluster of low values and a couple of high ones. It is particularly noteworthy that the two highest values cover a wider range than the lowest five. The degree of skewness in these data would cause an appreciable loss of statistical power if they were included in a t-test or ANOVA. However, if we subjected this data to the Anderson-Darling test in Minitab (*Stat / Basic Statistics / Normality test …*) the P value would be 0.107. These very dodgy data fail to produce a significant result.

Data set 2. Large data set that produces a significant result despite being only marginally non-normal.

The data list is very long (100 observations) and therefore follows this section.

Fig 3 shows that the data are about as close to normal distribution as you will ever see in the real world. For practical purposes, a t-test or ANOVA would cope perfectly satisfactorily with such a data set.

Figure 3. Histogram of data set 2.



However, there is a very slight positive skew to the data – so slight that you would not normally even notice the handful of excess data points in the right hand tail. Unfortunately, this marginal non-normality does get picked up as statistically significant as a result of our large data set (Anderson-Darling test P = 0.028).

Data set 2

10.22
10.96
9.35
12.36
12.84
8.44
10.93
13.64
9.36
9.26
9.89
7.47
8.59
11.21
8.80
14.20
9.35

11.55
9.93
8.96
10.92
9.13
9.92
10.59
11.91
15.56
11.31
9.50
10.33
10.19
12.08
9.81
10.05
10.33
12.56
9.31
12.27
10.38
8.93
8.24
11.49
7.66
10.96
11.00
10.97
8.66
9.12
8.30
9.83
9.64
10.11
14.54
9.41
11.79
12.05
7.17
10.29
10.80
8.19
10.25
10.86
11.15
10.49
10.81
10.19
10.46
8.96

10.84
8.23
9.37
9.35
8.75
6.87
14.71
8.65
10.39
9.52
9.80
12.23
11.94
10.49
12.00
12.88
10.40
10.79
11.62
10.32
9.52
10.27
9.34
6.19
11.14
9.13
9.96
11.09
9.46
9.45
9.12
8.68
9.67

Mean 10.28 ±0.16