CHAPTER 1

Introduction – Choosing the Correct Statistics

1.1 Introduction

Analytical scientists answer customers' questions. The variety of these is enormous and so, therefore, is the range of measurements that have to be made. Perhaps meaningful information can be provided only if a number of test samples are analysed. This might be because there is doubt about the homogeneity of the material or because more confidence is required in the result presented. To answer the customers' questions adequately, the analyst will need to decide on issues such as the number of samples to analyse and on what experiments and numerical comparisons must be done. These decisions require the use of statistics.

Using statistics effectively is an important part of any analytical scientist's job. For any meaningful decision, an analyst must plan the right experiments using properly validated methods, calibrate instruments, acquire the data under appropriately controlled conditions, review and check the results, present the data concisely and make objective decisions based on the data. Every step in this sequence of events is supported by some statistical procedure.

This book describes a range of the most useful statistical procedures for analytical scientists. It is intended to focus on procedure rather than theory and gives detailed instructions on how to carry out particular statistical operations, such as taking averages, evaluating standard deviations and constructing calibration curves. In general, each Chapter builds on topics from previous Chapters. Chapter 2 describes useful graphical methods for examining and presenting data. Chapter 3 provides some essential background information on data and the more common distributions relevant for assessing analytical data. Chapter 4 provides an array of basic statistical techniques, including the most common 'summary statistics' and the main statistical tests used in day-to-day analytical science. Outliers – a common feature of experimental data – are discussed in Chapter 5, which provides methods for detecting outliers and also some appropriate methods for handling data with some outlying values present. Chapters 6 and 7 describe analysis of variance and linear regression. Experimental design – a crucial preliminary activity – is discussed in Chapter 8, as it relies heavily on earlier concepts. Chapters 9–12 consider the statistical aspects of analytical method validation, measurement uncertainty estimation, quality control and proficiency testing, and Chapter 13 provides an introduction to sampling strategies for obtaining representative sets of test samples.

Practical Statistics for the Analytical Scientist: A Bench Guide, 2nd Edition
Stephen L R Ellison, Vicki J Barwick, Trevor J Duguid Farrant

Published by the Royal Society of Chemistry, www.rsc.org

1.2 Choosing the Right Statistical Procedures

Statistics provides a large set of tools for analysing data. Many of the most useful for routine analysis, analytical method development and quality assurance are described in this book. But which tools are appropriate for different tasks?

The following sections form a short guide to the parts of this book which are best suited to solving particular problems. They are arranged as a series of questions that analysts might ask themselves, grouped according to the type of problem.

Throughout this section, questions are denoted by ▷ and answers by ▶.

1.2.1 Planning Experiments

▷ 'How many test samples and replicates do I need?'

▶ Chapter 8, Section 8.5, discusses number of samples or observations for several different situations.

▷ 'In what order do I run my test materials for a single run?'

▶ For a simple experiment in a single run, randomised order (Chapter 8, Section 8.6.1) is simple and effective.

▷ 'How do I minimise operator or run effects in a comparative experiment?'

▶ Either randomise or (better) treat Operator or Run as a 'blocking factor' in a randomised block design (Chapter 8, Section 8.6.3).

▷ 'How do I test for several effects at a time without confusing them?'

▶ Factorial designs (Chapter 8, Section 8.4.4) or, more economically, fractional factorial designs (Chapter 8, Section 8.7.1) can test several effects simultaneously without compromising test power.

▷ 'What experiments help most to improve analytical methods or processes?'

▶ Using fractional factorial designs (Chapter 8, Section 8.7.1) to identify important effects, followed up with optimisation designs (Chapter 8, Section 8.7.2) to locate optimal conditions is often considered the most effective and economical approach.

1.2.2 Representative Test Portions for Large Samples

▷ 'The sample sent to the laboratory is far too big to analyse every item and all the items look the same. Which items should be taken for analysis?'

▶ With a collection of apparently identical objects, simple random sampling is usually appropriate. This is described in Chapter 13, Section 13.4.1. A systematic sampling scheme (Chapter 13, Section 13.4.3) may also be appropriate if there is no reason to suspect regular variation in the laboratory sample.

▷ 'The sample sent to the laboratory is large and there are clearly two or more different types of item present. Which items should be taken for analysis to obtain an average value?'

▶ With two or more identifiable groups, stratified random sampling is often the most effective. This is described in Chapter 13, Section 13.4.2.

1.2.3 Reviewing and Checking Data

▷ 'How do I locate outliers in a set of experimental results?'

- ► Inspection using a dot plot (Chapter 2, Section 2.2) or box plot (Chapter 2, Section 2.8) is often the quickest way to find possible outliers. Follow up suspect values with *outlier tests* (see below).
- ▷ 'How do I decide whether an apparent outlier is a chance occurrence or requires closer examination?'
- ► Perform either Dixon's test or Grubbs' test (G') for a single outlier (Chapter 5, Sections 5.2.3 and 5.2.4, respectively).
- ▷ 'How do I check two possible outliers simultaneously?'
- ► Perform the Grubbs' tests G'' and G''' for outlying pairs (Chapter 5, Section 5.2.4).
- ▷ 'If I have data from several different sources, how do I find out if one of the sources has an abnormally high (or low) spread of data?'
- ► If you have a set of standard deviations, s, or variances, s^2, you can carry out the Cochran test following the procedure in Chapter 5, Section 5.2.5, to check whether one of the variances is abnormal.
- ▷ 'If I have outliers among my data, what should I do?'
- ► The appropriate action to take on finding outliers when using outlier tests is discussed in Chapter 5, Section 5.2.2. Alternatively, if outliers are likely, consider using robust statistics as described in Chapter 5, Section 5.3.

1.2.4 Reporting Results – Summarising and Describing Data

- ▷ 'How can I best summarise my data?'
- ► (a) If you have a set of analytical results from a single source (such as one method, location or time), it is normally sufficient to give the number of results, the mean and the standard deviation. Confidence limits or range may be included. All these statistics are described in Chapter 4, Section 4.1.
  (b) If you have a set of results from several different sources and it is not appropriate to treat all the data as a single set, each set can be summarised separately as in (a). For multiple groups of data on the same material, it may also be useful to calculate within- and between-group standard deviations as described in Chapter 6, Section 6.3.2.2. For data on precision of methods, estimation of repeatability and reproducibility as described in Chapter 9, Section 9.2.2, is also useful.
- ▷ 'It looks like the results spread out more at higher concentrations. Is there a simple way of describing this?'
- ► Often, analytical precision is approximately proportional to concentration when far from the detection limit. This can be conveniently summarised across modest concentration ranges using the relative standard deviation or coefficient of variation described in Chapter 4, Section 4.1.4.4.

1.2.5 Decisions About Differences and Limits

1.2.5.1 Comparing Results with Limits

- ▷ 'I have a single measurement; how do I decide whether my result indicates that the limit has been exceeded?'
- ► This question is now most commonly addressed by considering the measurement uncertainty associated with the observation. If the result exceeds the limit by more than the expanded uncertainty (Chapter 10, Section 10.2.7), the limit is considered exceeded.

▷ 'I have made a number of measurements on a material; how do I decide if my results indicate that the limit has been exceeded?'
▶ Refer to Chapter 4, Section 4.2.3.1 (Comparing the mean with a stated value) and carry out a one-tailed test to see if the mean significantly exceeds the limit.
▷ 'The limit is zero. What do I do?'
▶ *Either* carry out a significance test (Chapter 4, Section 4.2.3.1) to see if the mean result significantly exceeds zero, *or* compare the result with the *critical value* described in Chapter 9, Section 9.5.1.

1.2.5.2 Differences Between Sets of Data

▷ 'Is there a significant difference between my two sets of results?'
▶ Use the procedure in Chapter 4, Section 4.2.3.2 (Comparing the means of two independent sets of data).
▷ 'Is my new method performing to the same (or better) precision than the method previously used?'
▶ If you have a set of data on the same material for each method, then compare the variances using the *F*-test (Chapter 4, Section 4.2.4).

1.2.6 Calibrating Instruments

▷ 'I have calibrated an instrument by measuring the instrument response for a set of solutions of differing concentrations of an analytical standard. What do I do next?'
▶ Refer to Chapter 7, Section 7.1, on linear regression. Calculate the regression line as described in Chapter 7, Section 7.1.4. Examine the residuals (Chapter 7, Section 7.1.5) to see whether the data are adequately represented by the chosen line. You should also examine the regression statistics (Chapter 7, Sections 7.1.6 and 7.1.8) to make sure the calibration is adequate.
▷ 'How do I calculate the concentration of a given test material from the calibration line?'
▶ Measure the instrument response (or responses, if replicated) for the given sample solution, then use the gradient and intercept of the calibration line as described in Chapter 7, Section 7.1.7.
▷ 'How do I estimate the statistical uncertainty of a concentration calculated from a regression line?'
▶ Refer to Chapter 7, Section 7.1.7, to calculate the standard error on the predicted concentration.
▷ 'How do I fit data to a curved line?'
▶ Some moderately curved calibration lines can be effectively fitted using polynomial regression as shown in Chapter 7, Section 7.2. Special forms (such as immunoassay response) are better treated with specialist software to fit the correct type of curve.

1.2.7 Describing Analytical Method Performance

Method performance is normally described using a range of parameters, such as precision (including repeatability and reproducibility), bias, recovery and detection and quantitation limits.

These are usually determined during method validation, which is discussed further below and in more detail in Chapter 9. Simple summary statistics for describing bias and dispersion are also covered in Chapter 4.

1.2.8 Analytical Method Validation

▷ 'How do I test for significant method bias using a certified reference material?'

▶ Experiments for bias are described in Chapter 9, Section 9.3. Statistical tests are usually based on the *t*-tests described in Chapter 4, Section 4.2.3.

▷ 'How do I test for bias compared to a reference method?'

▶ This question is answered in Chapter 9, Section 9.3.1.2. It is worth remembering that paired experiments (Chapter 8, Section 8.6.2) and the corresponding paired test in Chapter 4, Section 4.2.3.3, provide sensitive tests for differences between methods.

▷ 'How do I estimate the repeatability standard deviation?'

▶ The repeatability standard deviation (see Chapter 9, Sections 9.2.1 and 9.2.2) can be estimated from repeated measurements on identical test materials using the same method, the same operator and within short intervals of time. It can also be estimated in conjunction with reproducibility (see below).

▷ 'How do I estimate the reproducibility standard deviation?'

▶ The reproducibility standard deviation (Chapter 9, Sections 9.2.1 and 9.2.2) can be estimated from the results of repeated measurements on identical test materials using the same method, but in different laboratories. Usually, reproducibility is determined using simple nested experiments (Chapter 8, Section 8.4.3) and calculated from ANOVA tables as described in Chapter 9, Section 9.2.2.

▷ 'How do I determine the detection limit?'

▶ Experiments and calculations for characterising detection capability are discussed in detail in Chapter 9, Section 9.5.

▷ 'How do I assess linearity?'

▶ Linearity is assessed visually using scatter plots (Chapter 2, Section 2.9) of both raw data and residuals from linear regression (Chapter 7, Section 7.1.5). Objective tests for linearity are described in Chapter 7, Section 7.1.9.

▷ 'How do I carry out a ruggedness study?'

▶ Experiments and calculations for ruggedness studies usually use fractional factorial designs (Chapter 8, Section 8.7.1); their application to method validation is discussed in Chapter 9, Section 9.7.

▷ 'What do I do if I need to compare more than two methods?'

▶ Refer to Chapter 6, which describes analysis of variance (ANOVA). The type of ANOVA test depends on the information available. For results on the same material for each method, one-way ANOVA as described in Section 6.3 is appropriate. When several materials have been tested, two-factor ANOVA (Section 6.5) is appropriate.

1.2.9 Analytical Quality Control

▷ 'How do I set up a quality control (QC) chart?'

▶ Obtain data from the analysis of a stable check sample repeated over time (ideally several weeks or months) to set up either a Shewhart chart (Section 11.2) or a CuSum chart (Section 11.3).

▷ 'What is the best chart to detect unusual individual results?'

▶ A Shewhart chart (Chapter 11, Section 11.2) is often the most useful chart for identifying short-term deviations such as individual analytical run problems.

▷ 'How do I set up a chart to detect small changes over time?'

▶ A CuSum chart (Chapter 11, Section 11.3) is particularly good for identifying small, sustained shifts in a process or for retrospective examination to locate the time of a change.

1.2.10 Testing Laboratory Performance – Proficiency Testing

▷ 'How did we do?'

▶ Review the scores from your last proficiency test (PT) following Chapter 12, Section 12.3. Large z-scores indicate poorer performance.

▷ 'How do PT scores work?'

▶ PT scores compare the difference between a laboratory's result and an assigned value with some measure of the expected spread of results for competent laboratories. The most common are z-scores and Q-scores described in Chapter 12, Section 12.2.2.

▷ 'What does a z-score outside ± 2 indicate?'

▶ Usually, a z-score over 2 is an indication that the laboratory result was unusually distant from the assigned value. If above 3, investigative and corrective actions are nearly always required (see Chapter 12, Section 12.3).

▷ 'Is there a way of keeping track of our laboratory performance over time?'

▶ Chapter 12, Section 12.4, gives some useful pointers.

▷ 'We had the best score in our last PT round. Does that make us the best laboratory?'

▶ Sadly, no. Ranking in PT schemes is not a good indication of performance. Chapter 12, Section 12.5 explains why.

1.2.11 Measurement Uncertainty

▷ 'What is measurement uncertainty?'

▶ Measurement uncertainty is a single number that summarises the reliability of individual measurement results. It is defined and discussed in detail in Chapter 10.

▷ 'Is there a quicker way of calculating measurement uncertainty?'

▶ A particularly useful method using an ordinary spreadsheet is described in Chapter 10, Section 10.3.1.

▷ 'I already have the published repeatability and reproducibility data for my analytical method. Can I use this for uncertainty estimation?'

▶ Repeatability and reproducibility data can be used as the basis for uncertainty estimation in testing laboratories. The general principle is described briefly in Chapter 10, Section 10.3.2 and references to some relevant international documents (some freely available) are provided.