Part I

Introduction and basics

Chapter 1
Basics and terminology

1.1 Introduction

Food issues are becoming increasingly important to consumers, most of whom depend on the food industry and other food workers to provide safe, nutritious and palatable products. These people are the modern-day scientists and other practitioners who work in a wide variety of food-related situations. Many will have a background of science and are engaged in laboratory, production and research activities. Others may work in more integrated areas such as marketing, consumer science and managerial positions in food companies. These food practitioners encounter data interpretation and dissemination tasks on a daily basis. Data come not only from laboratory experiments but also via surveys on consumers, as the users and receivers of the end products. Understanding such diverse information demands an ability to be, at least, aware of the process of analysing data and interpreting results. In this way, communicating information is valid. This knowledge and ability gives undeniable advantages in the increasingly numerate world of food science, but it requires that the practitioners have some experience with statistical methods.

Unfortunately, statistics is a subject that intimidates many. One need only consider some of the terminology used in statistic text titles (e.g. ‘fear’ and ‘hate’; Salkind 2004) to realise this. Even the classical sciences can have problems. Professional food scientists may have received statistical instruction, but application may be limited because of ‘hang-ups’ over emphasis on the mathematical side. Most undergraduate science students and final-year school pupils may also find it difficult to be motivated with this subject; others with a non-mathematical background may have limited numeracy skills presenting another hurdle in the task.

These issues have been identified in general teaching of statistics, but like other disciplines, application of statistical methods in food science is continually progressing and developing. Statistical analysis was identified, two decades ago, as one subject in a set of ‘minimum standards’ for training of food scientists at

Statistical Methods for Food Science: Introductory Procedures for the Food Practitioner, Second Edition. John A. Bower.

undergraduate level (Iwaoka *et al*. 1996). Hartel and Adem (2004) identified the lack of preparedness for the mathematical side of food degrees, and they describe the use of a quantitative skills exercise for food engineering, a route that merits attention for other undergraduate food science courses.

Unfortunately, for the novice, the subject is becoming more sophisticated and complex. Recent years have seen this expansion in the world of food science, in particular in sensory science, with new journals dealing almost exclusively with statistical applications. Research scientists in the food field may be cognizant with such publications and be able to keep abreast of developments. The food scientist in industry may have a problem in this respect and would want to look for an easier route, with a clear guide on the procedures and interpretation, etc. Students and pupils studying food-related science would also be in this situation. Kravchuk *et al.* (2005) stress the importance of application of statistical knowledge in the teaching of food science disciplines, so as to ensure an ongoing familiarity by continual use.

Some advantages of being conversant with statistics are obvious. An appreciation of the basis of statistical methods will aid making of conclusions and decisions on future work. Other benefits include the increased efficiency achieved by taking a statistical approach to experimentation. Guiding the reader on the path to such knowledge and skills begins with a perusal of the book contents.

What will this book give the reader?

The book will provide the reader with two main aspects of statistical knowledge. One is a workbook of common univariate methods (Part I) with short explanations and implementation with readily available software. Secondly (Part II), the book covers an introduction to more specific applications in a selection of specialised areas of food studies.

1.2 What the book will cover

Chapter 1 introduces the book and gives a summary of how the chapter contents will deal with the various aspects. Accounts of the scope of data analysis in the food field, its importance and the focus of the text lead on to a terminology outline and advice on software and bibliography.

Chapter 2 begins with consideration of data types and defines levels of measurement and other descriptions of data. Sampling, data sources and population distributions are covered.

Chapter 3 introduces the style of the analysis system used with the software and begins with simple analysis for summarising data in graph and table format. Measures including mean, median, mode, standard deviation and standard error are covered, along with various types of graphs. Definitions and application of some of these methods to measures of error, uncertainty and sample character are also given.

Chapters 4–6 cover various aspects of analysis of effects. Firstly, Chapter 4 gives a detailed account of significance testing. Analysis of significant differences, probability and hypothesis testing and its format are described and discussed. The chapter concludes with consideration of types of comparison and factors deciding selection of a test, including assumptions for use of parametric methods. Chapter 5 continues with significance tests themselves, with tests for parametric and non-parametric data, two or more groups, and related and independent groups. Chapter 6 describes effects in the form of relationships as association (cross tabulation) and correlation (coefficients) and their significance. The topic of correlation is then applied in simple regression and prediction.

Chapter 7 concludes cover of basic material by detailing the nature and terminology of experimental design for simple experiments. Stages in the procedure, such as identification of factors and levels, and sources of experimental error and their elimination are explained. Details of design types for different sample, factor, treatment and replication levels are then described.

Chapters 8 and 9 start the applications part of the book. In Chapter 8, sensory and consumer data are described in terms of level of measurement, sources, sampling via surveys, sensory panels and consumer panels. Summary methods and evaluation of error, reliability and validity in these data sources are considered along with checking on assumptions for parametric nature. Specific methods of analysis are then illustrated for a range of consumer tests and survey data, and for specific sensory tests and monitoring of sensory panels.

Chapter 9 uses a similar approach to instrumental data. They are described in terms of level of measurement, sources and sampling via chemical and physical methodologies in food science. Analytical error, repeatability and accuracy are defined followed by use of calibration and reference materials. An account is then given of specific significance analysis methods for laboratory work results and experiments.

Chapter 10 applies experimental design to formulation procedures in food product development. Identification of factors and levels as ingredients for simple designs is given viewed from the formulation aspect. Decisions on the response and its measurement are described along with the issues in objective versus hedonic responses. Examples of some formulation experiments are used to illustrate the analysis methods and their interpretation.

Chapter 11 deals with the application of the basic methods and experimental design to the case of quality control procedures. Key features of sampling quality measurement are outlined and then application of statistical methods to production is explained. The data forms generated and their analysis are displayed with reference to control charts and acceptance sampling.

Chapter 12 provides some indication of how to take the univariate methods a stage further and apply them to multivariate situations. A selection of commonly used more advanced and multivariate methods are briefly described, with more detail on principal component analysis. The analysis of sensory and instrumental data and how to combine them in the multivariate context concludes.

Despite the above examples and the description of content, many scientists and workers in the food field may still ask why they are required to have such knowledge. This question can be answered by considering the importance and application of the subject in the food field.

1.3 The importance of statistics

Why are statistical methods required?

It is possible to evaluate scientific data without involving statistical analysis. This can be done by experienced practitioners who develop a 'feel' for what the data are 'telling them', or when dealing with small amounts of data. Once data accumulate and time is limited, such judgement can suffer from errors. In these cases, simple statistical summaries can reduce large data blocks to a single value. Now, both the enlightened novice and the experienced analyst can judge what the statistics reveal. Consequent decisions and actions will now proceed with improved confidence and commitment. Additionally, considerable savings in terms of time and finance are possible.

In some instances, decision-making based on the results of a statistical analysis may have serious consequences. Quantification of toxins in food and nutrient content determination rely on dependable methods of chemical analysis. Statistical techniques play a part in monitoring and reporting of such results. This gives confidence that results are valid and consumers benefit in the knowledge that certain foods are safe and that diet regimes can be planned with surety. Other instrumental and sensory measures on food also receive statistical scrutiny with regard to their trustworthiness. These aspects are also important for food manufacturers who require assurance that product characteristics lie within the required limits for legal chemical content, microbiological levels and consumer acceptability. Similarly, statistical quality control methods monitor online production of food to ensure that manufacturing conditions are maintained and that consumer rights are protected in terms of net weights, etc. Food research uses statistical experimental design to improve the precision of experiments on food.

Thus, manufacturers and consumers both benefit from the application of these statistical methods. Generally, statistics provides higher levels of confidence and uncertainty is reduced. Food practitioners apply statistical methods, but ultimately, the consumer benefits.

1.4 Applications of statistical procedures in food science

There are many applications of statistics in the field of food studies. One of the earliest was in agriculture where Fisher used experimental design to partition variation and to enable more precise estimation of effects in crop plot experiments. There was even an early sensory experiment on tea tasting (Fisher 1966), and since then, statistical applications have increased as food science emerged as a distinct

Table 1.1 Some applications of statistics in the food field.

Method	Application
Summaries of results	Tables, graphs and descriptive statistics of instrumental, sensory and consumer measures of food characteristics
Analysis of differences and relationships	Research applications on differences in food properties due to processing and storage; correlation studies of instrumental and sensory properties
Monitoring of results	Statistical control of food quality and parameters such as net filled weight
Measurement system integrity	Uncertainty of estimates for pesticides and additives levels in food
Experimental design	Development and applications of balanced order designs in sensory research

applied science subject. Some examples of the form of statistical applications in food are given in Table 1.1.

Preparation of data summaries is one general application of statistics that can be applied across the board. It is one of the simplest applications and can be done manually if necessary, depending on the requirements. A variety of simple graphs and table methods are possible, which allow rapid illustration of results. These summaries are taken further in statistical quality control where measures such as the mean value are plotted 'live', as a process is ongoing. The graphs (control charts) used include limit lines that are set by using other statistical methods, which allow detection of out-of-limit material, e.g. food product packs that are below statutory minimum net weight. Statistical methods can also be applied to evaluate the trustworthiness of data obtained by any method of measurement. This application has been used extensively in evaluation of chemical data generated by analytical laboratories. The statistical analysis provides an evaluation of how dependable the analytical results are. This can range from within-laboratory to between-laboratory comparisons, globally. Enforcement agencies rely on such checks so that they can monitor adherence to legal requirements with confidence.

Food research application brings in analysis of differences and relationships. Here, hypotheses are put forward on the basis of previous work or new ideas and then magnitudes of effects in sample statistics can be assessed for significance, for instance, examination of the change in colour pigment content during frozen storage of vegetables.

Examination of relationships requires that different measurement systems are applied and then compared. There are many examples of this in studies of food where data from instrumental, sensory and consumer sources are analysed for interrelationships.

The process of sampling of items, including food material and consumer respondents, can be controlled using statistical methods, and here, a statistical appreciation of variability is important. Experimental design takes this further, where sources of such variation are partitioned to improve precision or controlled and minimised if extraneous. A common example is the unwanted effect of order of samples in the sensory assessment of foods – design procedures can minimise this.

In fact, *all* the above examples rely on design procedures if the result is to be valid and adequately interpreted. Experimental design is dealt with fully in a later chapter, but an introduction to some aspects of experimentation is important at this point to provide a foundation.

1.4.1 The approach to experimentation

Why are experiments and research necessary?

Progress in food science and all its associated disciplines is underpinned by research activity. New information is gathered by investigations and experiments, and in this way knowledge is advanced. The scientific approach to research and exploration follows an established paradigm called the ***positivism method***. This postulates that events and phenomena are objective and concrete, able to be measured and can be explained in terms of chemical and physical reactions. All scientists are familiar with this viewpoint, which is described as the ***scientific deductive approach*** (Collis and Hussey 2003). It is largely based on ***empirical methods***, i.e. observations from experiments. The scientific style of approach can be used for any type of investigation in any subject.

The procedure uses deduction from theory based on current knowledge. To advance knowledge, experiments can be designed to test advances on existing or new theory, using a ***hypothesis process***. The findings can then be disseminated and knowledge increased. Results are generalised and can be used to establish new theories and to model processes and event reactions, which in turn allows prediction in the formation of new hypotheses. The term ***quantitative research*** is also used in reference to the scientific approach. This strictly refers to the nature of the data generated, but it implies the deductive positivistic viewpoint.

In this process, the researcher is assumed to be objective and detached. Ultimately, the deductive method searches for an explanation on the basis of *cause–effect* relationships. Without such procedures, there would be no progress and they form the foundation of the scientific approach in many food disciplines.

A more recent approach is that of ***phenomenology*** where an ***inductive approach*** can be used to examine phenomena on the basis that they are socially constructed. Theories and explanations are generated and built up from data gathered by methods and techniques such as interviews (Blumberg *et al.* 2005). These methods are often described as ***qualitative***, which again refers to the data that are in the form of words rather than numbers. The modern food practitioner needs to be aware of such data as there are several qualitative methods (e.g. interviews and focus groups) used in sensory and consumer work. Analysis of data from ***qualitative methods*** can be summarised by numerical techniques such as counting the incidence of certain words and phrases, but usually statistical analysis as such is not involved.

Typical use of the scientific approach in food studies entails identifying a topic for research or investigation then posing a ***research question(s)***. Deductive reasoning from existing knowledge is examined to develop a ***research hypothesis***.

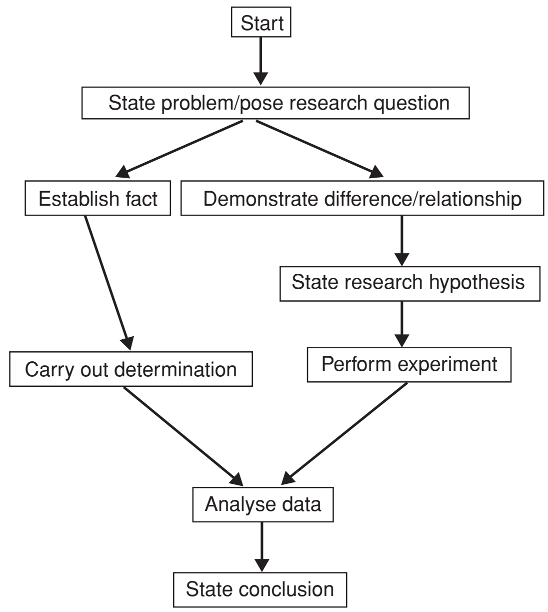

Fig. 1.1 The approach to investigation.

A plan can then be drawn up with an experimental design and specification of measurement system, etc. Data are gathered and then statistical analysis is used to test the hypothesis (quantitative). The scope of the procedure can be from a simple investigation of the 'fact-finding' type, e.g. determination of chemical content values, to a complex experimental design, e.g. a study on the effect of temperature, pressure and humidity levels on the drying properties of a food. In this latter case, the objective would be to identify any significant differences or relationships. Experimental control means that results can be verified and scrutinised for validity and other aspects.

Simple experiments do not usually require stating of hypotheses, etc. In circumstances where differences or relationships are being examined, e.g. 'Does process temperature affect yield of product', a more formal procedure is used or, at least, assumed (Fig. 1.1).

The conclusion of one investigation is not the end of the process as each piece of work leads to new ideas and further studies, etc.

1.5 Focus and terminology

It is important at this point to give some indication of what this text will concentrate on in terms of applications. Data from three main areas are drawn on:

- Instrumental methods of analysis
- Sensory methods
- Consumer tests and measures

These divisions are broad enough in scope to provide examples of a range of statistical applications. The terms above are used throughout this book in reference to data (thus, instrumental data, sensory data, etc.).

Instrumental measures itself can cover any measurement system from chemical and physical analysis to specific food instrumentation methods and process measures, e.g. protein content, Lovibond colour measures and air speed setting. ***Sensory measures*** include all sensory tests used by trained assessors such as discrimination tests and descriptive analysis methods. ***Consumer tests*** include some sensory methods, which are affective or hedonic in nature, e.g. preference ranking. The above systems cover mostly laboratory measurements.

Consumer measures refer to questionnaire measures in surveys, such as consumers' views and opinions on irradiated foods. These consumer applications are usually non-laboratory in nature.

1.5.1 Audience

As stated above, this book is aimed at the food scientist or food practitioner who has to undertake some aspect of data analysis and interpretation. The intention is not to include formulae and mathematical content wherever possible. All calculations are done using appropriate statistical software. In this respect, the text cannot be viewed as a statistical source, but numerous references for those readers who wish more on formulae are given. The emphasis is on providing a basic account of how methods and tests are selected, how to avoid inappropriate use, how to perform the tests and how to interpret the results.

Software packages

All examples of statistical tests and methods are selected from the appropriate menus of a software package usable with ***Microsoft® Windows®.*** Readers will require some familiarity with an analysis package. ***Microsoft® Excel®*** is probably the most accessible and available package that includes statistical analysis and functions. Applications employ the ***Excel 2003*** and ***Excel 2010***[1] versions, as used in Windows for PC. Explanation of the differences between the functions for these versions is detailed in some sections, otherwise the 2010 function is stated followed by the 2003 form in brackets. As far as possible, Excel functions and data analysis add-ins are used, followed by those of ***Minitab*** (Minitab® Statistical Software™; any basic version). Some guidance is given on use of functions and commands for these packages, but the instructions cannot be viewed as comprehensive and some readers may wish to consult texts such as Middleton (2004), Carlberg (2011), and McKenzie *et al.* (1995) and Wakefield and McLaughlin (2005) for advice on Excel and Minitab, respectively. The Excel **Analysis ToolPak** and **Solver Add-ins** are required for some examples. Another useful addition is the ***MegaStat***

[1] ***Excel 2007*** has a similar interface and commands to ***Excel 2010***.

(McGraw-Hill/Irwin, McGraw-Hill Companies Inc.) add-in for Excel (Bowerman *et al.* 2008), which provides other facilities including some non-parametric tests.

Some of the more advanced design examples and the multivariate applications (Chapter 12) require Excel with more sophisticated facilities, fuller versions of Minitab, or packages such as SPSS (International Business Machines Corporation) and specialist software (Design-Expert(R), Stat-Ease, Inc.). It must be stressed that these are not required for the majority of examples in this book, and that it is appreciated that many readers will not have access to all these packages. Several are available as student versions and this is indicated. Additionally, there is a wealth of statistical tools and facilities on the Internet.

1.5.2 Conventions and terminology

Already in this chapter, new terminology is accumulating, which may cause confusion for some readers. Additionally, certain terms can be used in more than one sense. For example, 'analysis' can refer to chemical analysis or statistical analysis. 'Sample' can be used in more than one way. 'Measure' is a general term, but it signifies the act of carrying out of a determination, etc. Most key terms are defined or explained where they occur and usually the context will aid understanding, but a brief list of common terms and their meaning is presented in Table 1.2.

In addition to software and terminology, etc., indicated earlier, much reference is made to certain key texts and research publications in food science and statistics. Most chapters include at least one of the following sources, which are introduced at this point.

Advice on bibliography

Textbooks on specific basic applications of statistics to food science are not numerous. Some 'standard' texts include Gacula and Singh (1984) on general material for food research and O'Mahony (1986) on sensory applications. There are several texts on multivariate sensory applications. Also, sensory evaluation data analysis is dealt with in other general sensory texts, but only as an add-on section or a chapter in the main contents. For the chemical analyst, there are some general texts of interest such as Miller and Miller (1999), but these are not specific to food. As seen, the publication dates on several of these are more than a decade ago, although they are still used extensively by practitioners. More recent texts in general biosciences are appearing, but they often deal with advanced newer methods. Other useful texts take a more technical view (e.g. Chatfield 1992) or a gentler approach (e.g. Rowntree 2000), but again these are not specific to food science. Ultimately, the food scientist may need to go to journal and article publications to get a particular method, from research journals and the work of organisations such as the Laboratory of the Government Chemist series on Valid Analytical Measurement and the Food Standards Agency. Older texts have the advantage that in some cases the descriptions and statistical analyses are simpler and easier to understand.

Table 1.2 Terminology and conventions.

Term	Meaning
Food practitioner	The person carrying out the food investigation and the statistical analysis (scientist, technologist, researcher, student, pupil, etc.)
Food study	Any investigation on food by the above practitioner
Analysis	Statistical analysis or instrumental analysis
Analyst	Chemical analysis practitioner
Analytical	Chemical analysis
Analyte	Constituent being determined by chemical analysis
End determination	A single measure on a sampling unit or subsample
Observation	As determination
Sample	A selection of sampling units
Sampling unit	An individual object taken from the population
Subsample	A portion of a sampling unit used for an end determination
Instrumental	Instrumental method (covers, chemical, physical, biological, etc.)
Sensory panel	Group of selected, trained assessors
Sensory test	Sensory analysis tests using trained assessors or consumers
Consumer panel	Group of untrained consumers for sensory testing
Consumer test	Sensory measure using consumer respondents
Consumer measure	A measure used in a consumer survey
Consumer	A member of a consumer population
Data	The values as measured
Results	Data that have been summarised and analysed
Assessor	A member of a sensory or consumer panel
Panellist	As assessor
Judge	As assessor
Ranking	Specific use of a ranking test
Rating/scoring	Allocating a measure from an ordinal, interval or ratio scale
Scaling	Allocating a measure from any scale
Variable	A measure that can take any value in a range of possible values
Object	An item upon which measures are made (food material, food machinery, methods, consumer, etc.)
Item	As object

References

Blumberg, B., Cooper, D. R. and Schindler, P. S. (2005) *Business Research Methods*. McGraw-Hill Education, Maidenhead, pp. 18–25.

Bowerman, B. L., O’Connell, R. T., Orris, J. B. and Porter, D. C. (2008) *Essentials of Business Statistics*, 2nd edn. McGraw Hill International Edition, McGraw-Hill/Irwin, McGraw-Hill Companies, New York.

Carlberg, C. (2011) *Statistical Analysis: Microsoft® Excel 2010.* Pearson Education, QUE Publishing, Indianapolis, IN.

Chatfield, C. (1992) *Statistics for Technology*, 3rd edn. Chapman & Hall, London.

Collis, J. and Hussey, R. (2003) *Business Research*. Palgrave MacMillan, Basingstoke, pp. 46–79.

Fisher, R. A. (1966) *The Design of Experiments*, 8th edn. Hafner, New York.

Gacula, M. C. and Singh, J. (1984) *Statistical Methods in Food and Consumer Research*. Academic Press, Orlando, IL.

Hartel, R. W. and Adem, M. (2004) Math skills assessment. *Journal of Food Science Education*, **3**, 26–32.

Iwaoka, W. T., Britten, P. and Dong, F. M. (1996) The changing face of food science education. *Trends in Food Science and Technology*, **7**, 105–112.

Kravchuk, O., Elliott, A. and Bhandari, B. (2005) A laboratory experiment, based on the maillard reaction, conducted as a project in introductory statistics. *Journal of Food Science Education*, **4**, 70–75.

Malhotra, N. K. and Peterson, M. (2006) *Basic Marketing Research*, 2nd edn. International Edition, Pearson Education, Upper Saddle River, NJ.

Mckenzie, J., Schaefer, R. L. and Farber, E. (1995) *The Student Edition of Minitab for Windows*. Addison-Wesley Publishing Company, New York.

Middleton, R. M. (2004) *Data Analysis Using Microsoft Excel*. Thomson Brooks/Cole Learning, Belmont, CA.

Miller, J. C. and Miller, J. N. (1999) *Statistics and Chemometrics for Analytical Chemistry*, 4th edn. Ellis Horwood, Chichester.

O'Mahony, M. (1986) *Sensory Evaluation of Food – Statistical Methods and Procedures*. Marcel Dekker, New York.

Rowntree, D. (2000) *Statistics without Tears – A Primer for Non-mathematicians*. Pelican Books, London.

Salkind (2004) *Statistics for Those Who (Think They) Hate Statistics*. Sage publications, London.

Wakefield, D. and McLaughlin, K. (2005) *An Introduction to Data Analysis Using Minitab® for Windows*, 3rd edn. Pearson Education, Pearson Prentice Hall, Upper Saddle River, NJ.

Software sources and links

Microsoft© Excel for Windows. Available at www.microsoft.com (accessed 28 February 2013).

Excel add-in: MegaStat® for Microsoft® Excel, McGraw-Hill Higher Education. Available at www.mhhe.com (accessed 28 February 2013).

(also available with Bowerman *et al.* 2008)

Minitab: Minitab® Statistical Software, Minitab Inc. Available at www.minitab.com (accessed 28 February 2013).

Design-Expert(R) software: Stat-Ease, Inc., Minneapolis, MN. Available at www.statease.com (accessed 28 February 2013).

SPSS for Windows Student Edition 13.0, 2006, International Business Machines Corporation, New York, (formerly SPSS Inc.). (Available with Malhotra and Peterson (2006)).

SPSS for Windows, Rel. 15.0.0 2006. International Business Machines Corporation, New York (formerly SPSS Inc.).

Laboratory of the Government Chemist. Available at www.lgc.co.uk (accessed 28 February 2013).

Department of Trade and Industry's VAM. Available at www.vam.org.uk (accessed 28 February 2013).

Food Standards Agency. Available at www.food.gov.uk (accessed 28 February 2013).