5

Application of Statistical Techniques in Food Science: Chemical Analysis Data

5.1 Introduction

There are many applications of statistics in the field of food studies. One of the earliest was in agriculture where R. A. 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 applied science subject. Some examples of the form of statistical applications in food are given in Table 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 on-going. The graphs (control charts) used include limit lines which are set by using other statistical methods, which allow detection of out-of-limit material, e.g. food product packs which 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.

Table 5.1 *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

5.2 Description

5.2.1 The Approach

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. Amore 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 which 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*. 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.). 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.

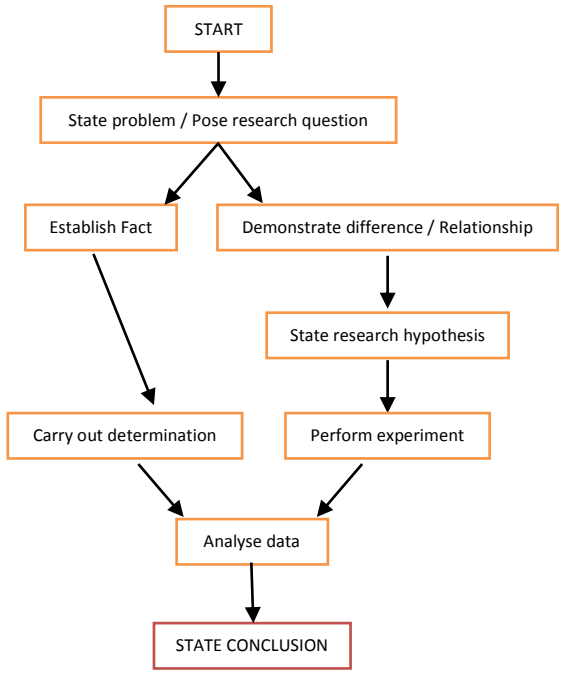

Figure 5.1 *The approach to investigation.*

5.2.2 Chemical Analysis

The chemical analyst is interested in the end result, but also in the *uncertainty* of the estimation; some researchers state that unless a measure of uncertainty is included the results themselves are useless (Mullins 2003). This view could well apply to all scientific measures, but there are still occurrences of it not being adopted for chemical data. Many investigations have taken place to examine error components and to quantify their contribution to uncertainty. Also, cost considerations are included in these studies as reducing uncertainty usually means additional analyses and hence cost in terms of time, resources and personnel. The interest here is in the balance between *gain in certainty*, against the *increased cost* to the laboratory (Lyn *et al.* 2002; FSA 2004a). Another unique aspect of studies in this topic is that the uncertainty is examined

not only for location measures estimate such as the mean, but also for that of the level of variability – thus the uncertainty of the standard deviation is also of interest. *Method proficiency testing* is one aspect of this protocol that has been developed for some common standard methods with measures all focused on uncertainty in analytical chemistry. In addition to analysis coming under this latter umbrella, where analyses such as pesticides are determined at very low level, there are many proximate analyses and 'crude content' methods used for food. These may exhibit higher levels of uncertainty, but their results and in fact, those from any instrumental measure can be subjected to some of the calculations detailed below. Food analysis methods have received special attention via The *Food Analysis Performance Assessment Scheme (FAPAS)*. Patey (1994) described the initial stages and progress of this initiative – there was some improvement, but not for all analyses and all laboratories. A relatively simple check on performance for proficiency testing schemes is based on calculation of a form of *z*-score

$$z = (\text{Test result} - \text{reference value})/\text{sigma}$$

Sigma is a designated 'target value' for the standard deviation of the method data, based on realistic estimates. The larger the discrepancy (error) between the test and the reference is, the larger the *z* value is. The calculation produces *z* in a standardised form – values *equal or less than 2* are required for the laboratories' result to be declared 'satisfactory'.

Accuracy and Bias in Chemical Analysis

As stated above, accuracy of a chemical method is a measure of how close it is to the 'true' value. Variation from the true can occur due to error in the form of bias (Kane 1997). This circumstance can apply to a number of stages in the analysis (Table 5.2)

Table 5.2 *Bias Sources.*

Source
Operator
Lab
Preparation
Run
Method

It is crucial that this source of error is quantified and removed, or at least accounted for in any analytical determination, although this is not always done (O'Donnell and Hibbert 2005). Bias can be calculated as the *error of the mean*, and by the location of the range specified by a *confidence interval*. The 'true value' is represented by *reference samples* or the nearest equivalent.

5.3 General Analysis

5.3.1 Errors and Measurement Uncertainty

The term, "experimental error" is used extensively in student lab books to account for all manner of unexpected results. While this may be appropriate the error can be allocated to a number of possible sources which can usually be identified as discussed below. *Gross* errors (e.g. a misread balance or grossly incorrect additions /omissions of reagents) are usually *accidental* in nature and with care they can be avoided. In the Kjeldahl analysis an obvious gross error would be seen if there was omission of the catalyst for one of the replicates. Rejection of that value could be considered and there are statistical tests for such "outlier" values. Thus these errors may not affect all measurements in a set and often can be easily detected. Other types of error occur even when the greatest care is taken. *Systematic* errors (e.g. a balance which requires servicing and calibration, unrecognized faulty technique by the analyst, or a method related

systematic error) usually affect all the analyses in a similar manner. The systematic error effect is also known as bias and affects accuracy. Note that even if a balance is calibrated (i.e. set to weigh accurately using certified weights) it may still give an inaccurate reading if the balance model is unable to read beyond ascertain level. Thus the lack of calibration is determinate error, and can be changed, but the other is constant. Calibration improves accuracy and reduces or removes any bias which instruments may have. A blank determination is another aid to detection of a systematic error. Another source of error is detected if the test sample was analysed more than once. Even if gross and systematic errors are absent, repeated measurements may show some variation. These are caused by *random* errors, e.g. small errors in weighing, use of volumetric devices and other analysis instrumentation. Even highly trained analysts using top of the range equipment might not avoid random error. The random error effect in a series of measurements causes the individual results to fall on either side of the mean. They may be accidental in nature but are indeterminate as they are difficult to remove entirely. Random errors affect the precision of the analysis method. These errors can occur at any stage of the analysis and accumulate to produce the overall error. Some errors augment one another whereas others may cancel one another out. The replicate values in Table I are all different and possible error sources could be deduced by examination of each stage of the Kjeldahl analysis. An estimate of error magnitude in the final results can now be calculated.

Table 5.3 *Quality control laboratory data for percentage of crude protein analysis on food product.*

Replicate Number	Percentage of protein Analysis A	(N2 x 6.25) Analysis B
1	7.3	8.4
2	8.5	9.1
3	-	8.7
4	-	8.2
Mean	7.9	8.6

Note: True/most probable value = 8.8 per cent.

5.3.2 Accuracy and Precision in Measurement

Accuracy is the extent of agreement between the determined value and the true or most probable value; and precision is the extent of agreement among a series of measurements of the same quantity. It is important to note that with these terms the presence of one does not automatically imply the other: a high degree of precision does not imply accuracy and vice versa.

(a) Measures of Accuracy

The degree of concordance with the true value can be calculated as the error of the mean which can also be expressed as the relative error of the mean (REM):

$$EM = M - T \quad \%\,REM = \frac{EM \times 100}{T}$$

Where:

EM = error of the mean

T = "true" or "actual" value

M = mean value.

The true value may not be available for unknown samples; unless an independent analysis has been performed giving a confident estimate. If an indication only is required then a rough estimate can be given by "typical" values from text books and/or food product labels. In the food production situation (Table 3), the true expected value can be calculated for quality control purposes from knowledge of the chemical composition of the specified ingredients. Alternatively a standard or control material of known composition can be analysed along with the unknown under the same analysis conditions, thus enabling the above calculation. A suitable crude test material can be made up by

the analyst, formulated from constituent chemicals or purified constituents. Another possibility used in some laboratories is use of a previously analysed material which is kept in stable storage and sampled along with the new samples. For more critical circumstances a CRM (certified reference material) would be required. While many RMs are available, not all common constituents such as nitrogen are found in the certified lists and the matrix (i.e. the physical and chemical "makeup") of the RM may be different from the unknown food sample. There have been some developments to answer these food specific requirements, e.g. the FAPAS (food analysis performance assessment scheme) initiative run by MAFF (Ministry of Agriculture, Fisheries, and Food) which has food product test materials for proximate analyses such as nitrogen protein. The use of a hierarchy of reference standards from secondary RMs to certified RMs and ultimately primary RMs forms part of the traceability chain for chemical composition instigated by VAM. The presence of errors will affect the magnitude of the percentage REM obtained. Assuming the absence of gross and systematic errors then a percentage REM of zero is possible but unlikely due to random errors. Usually negative or positive percentage REM values are obtained representing results which are below or above the true value respectively. These statistics can now be calculated for the data of Table 5.3. As could be easily deduced by inspection of the mean values, both analyses have underestimated percentage protein, and the magnitude of this is shown (Table 5.4) by the negative percentage REMs. Analysis B has a greater agreement with the most probable value.

(b) Measures of Variability (Precision)

The standard deviation (SD) and the mean absolute deviation (MD) introduced previously are measures of precision. These can be standardized as the percentage coefficient of variation (%CV; also known as the relative standard deviation) and the percentage relative mean deviation (% RMD) respectively:

$$\%\,\mathrm{CV} = \frac{\mathrm{SD} \times 100}{\mathrm{M}}$$

$$\%\,\mathrm{RMD} = \frac{\mathrm{MD} \times 100}{\mathrm{M}}$$

Where

SD = standard deviation

M = mean

MD = mean deviation.

These measures are related (MD is approximately 0.8 times SD). Both are included here as MD is perhaps easier to understand and calculate. In the form above, erroneous comparisons between data sets possessing different measurement scales are avoided, e.g. an MD often for a mean of 10,000 gives a very low percentage RMD (0.1 per cent), but with the same MD for a mean of 100 the RMD is very high (10 per cent). Two other related measures are important. Repeatability is the precision obtained when a method of analysis is repeated under the same conditions, i.e. by the same analyst using the same equipment, on the same sample material, etc. (also referred to as "within laboratory" or "within run" precision). The analyses in Table 3 can be assumed to have been done under repeatability conditions. Reproducibility is the precision obtained when the same method of analysis is repeated on the same test material but under different conditions, i.e. a different analyst, different setoff equipment or a different laboratory or even different method (also known as "between run" or "between laboratory" precision). Its usual to find that repeatability conditions result in greater precision than those of reproducibility. In fact the poor reproducibility shown by different laboratories when analysing the same samples was one of the reasons for instigating the VAM project. The magnitude of the

percentage CV (or percentage MD) will range from zero upwards. "Perfect" precision would produce a CV percentage of zero and although this can occur, more commonly small values are obtained, caused by random error. Large percentage Values may point to gross errors. Note that even if the method is perfectly precise, repeated values could still vary owing to inherent variation within the food material itself. Calculation of precision for the data of Table 5.3 shows that precision is relatively poor in seta (high %CV, %RMD values). Pertinent to these measures is the number of repeated measurements.

Table 5.4 *Accuracy measures for percentage of protein data.*

	Analysis A	**Analysis B**
Number of replicates	2	4
Mean (%)	7.9	8.6
% REM	-10.2	-2.3
Note: Most probable value = 8.8 per cent		

5.3.3 Acceptable Level of Replication

The level of replication is an important consideration as it affects the statistical measures and the cost of the analysis in terms of time and personnel. In practice the costs can limit the degree of replication. For routine analyses with established techniques, modern instruments and trained analysts, minimal replication maybe common, except where the technique is very rapid and low in cost, e.g. as with modern nitrogen analysers based on the Dumas method (2.5 minutes per sample). Thus duplicate determinations or even a single one done along with a reference or standard analysis for the run may be typical. If a single determination is made there is no reference point for error detection. Statistically, the greater the number of determinations is, the more reliable or accurate the result is. Whether or not a low level of replication is acceptable depends on several factors: the experience of the analyst and the laboratory itself; the method of analysis and its history with

respect to the food in question; and the importance of the decisions which are to be based on the results. Certainly, low levels of replication in isolation provide a weak basis for making confident decisions regarding the data obtained, e.g. standard deviation based on only two values is an extremely shaky foundation on which to base further inferences. The difference in magnitude between the SD values (Table 5.5) for two and four replicates, for data sets with similar ranges, illustrates this point. This does not, however, preclude the routine use of duplicates. The final consideration is how to use the calculated measures (Tables 5.4 and 5.5) to answer questions concerning the acceptability of the obtained levels of precision and accuracy

Table 5.5 *Accuracy measures for percentage of protein data.*

	Analysis A	**Analysis B**
Number of replicates	2.0	4
Mean (%)	7.9	8.6
Range (%)	1.2	0.9
MD	0.6	0.3
SD	0.85	0.39
%RMD	0.6	3.9
%CV	10.74	4.55

5.3.4 Acceptance Level for Precision

The deviation of a set of replicates around the mean depends on the precision of the measurement system and on the degree of variability of the population from which the samples originate. If both are of a completely unknown nature then whether or not to accept a set of replicates cannot be decided easily. Some measure of variability must be established. This can be done by carrying out an initial set of a larger number of replicates than is envisaged for routine use, e.g. at least ten, or if appropriate, by proceeding with duplicate analyses without considering variability until a “data bank” of typical values has been established

from which an estimate of deviation in the form of the standard deviation can be calculated, i.e. comment concerning the "expected variation" for a set of replicates cannot be made until some measure of variability has been established. Once this is available then an error estimate, known as a confidence interval (CI) can be calculated for the population mean of the measurement. It gives a region within which we are confident that the population mean will be located, with a specified probability or "certainty" level. This statistic can be used as an estimate of bias (accuracy) and the width of the interval gives another perspective on precision, as it emphasizes the effect of sample size. To understand a confidence interval we need to appreciate the nature of a population distribution. Put simply, if we know how a population is "mapped out" then it can be used to make estimations based on samples taken from that population. Imagine that the food product (Table 5.3) is analysed a very large number of times for crude protein content and grouped values are plotted on a histogram – then it is likely that a rough inverted cone shape would be obtained (see Figure 5.1). Increasing the number of points would have smoothing effect on the shape and with a very large number bell shaped curve would be obtained. Ultimately with an infinitely large number of values the curve would be smooth

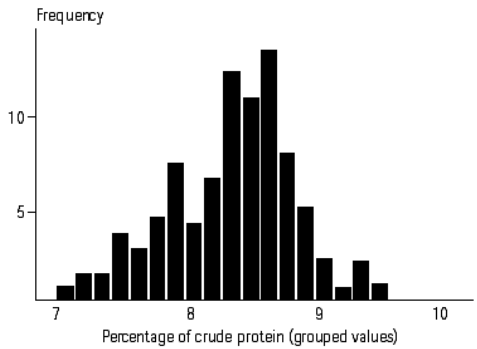

Figure 5.2 *Frequency Distribution of 100 per cent Crude Protein Content Determinations.*

and would represent the population distribution for the measurement. Note that these measurements are the entire same constituent on the same material, using the same technique, etc. The curve shape would be typical of a normal distribution (see Figure 5.2) and similar distributions, "normal" in this case meaning "standard". The curve has certain properties which allow powerful inferential statistics to be performed– the mean (μ), mode and median are centrally located; on either side of the centre the two "tails" are of such shape that more values are clustered towards the centre than at the edges; in terms of variation the proportion of the curve at one or more standard deviations(s) from the mean can be marked and measured. It can be seen that when selecting a random sample from such a population, there is a higher probability of obtaining a percentage protein value within ± 1 standard deviation of the population mean than further away, as there is more area under that region of the curve. In fact approximately 68 per cent of all the values lie in this region, and approximately 95 per cent lie within ± 2 standard deviations. Most chemical and physical measurements on food samples are likely to come from abnormal population. Even if the parent population deviates from normality, statistical theory proves that the distribution of the means of samples from such a population will approximate to normality. Thus this distribution will also possess the above properties and provides the basis for determining the confidence interval for the population mean based on the sample mean. Large sample sizes provide adequate estimates of the population parameters to allow calculation of the confidence interval using the proportions described above. For small samples of the order likely to be used in chemical analysis more appropriate distribution "standard" for making estimates is the *t*-distribution – it is similar in shape and characteristics to the normal distribution but is wider and flatter, having more "spread" (especially for small numbers of samples or replicates). Thus the interval will be wider, reflecting the increased uncertainty. A measure of the degree of confidence must

be specified and it is expressed on a probability scale of zero to 100 per cent, with 100 per cent representing absolute certainty. Unfortunately, choosing the 100 per cent level of confidence would result in an interval of very large width, unusable in practical situations. Usually the 95 or 99 per cent limit is selected, representing high degrees of confidence. The confidence interval limits are calculated using the *t*-value from the *t* distribution based on the number of replicates:

$$95\%\,\mathrm{CI} = \mathrm{M} \pm \mathrm{t} \times \mathrm{SD} / \sqrt{\mathrm{a}}$$

Where

n= number of replicates.

The value of *t* is obtained from statistical tables and its magnitude depends on the confidence level and on the number of samples analysed (more specifically on the degrees of freedom, which is equal to the number of samples minus1). Thus a high confidence level combined with low replication would maximize the *t*-value and the interval width and vice versa. These calculations can now be done for the data of Table 3 and are summarized in Table 5.6.

Thus, assuming no systematic error for Analysis B, the analyst would be confident that 95 per cent of the time, the population mean for percentage of crude protein content would lie between 8.0 and 9.2 per cent. The width of the interval can guide the acceptability of precision. Whether or not it is acceptable depends in turn on how confident the analyst is in the validity of the SD. In set A it is based on only two determinations and gives a confidence interval of 15.2 per cent owing to the large *t*-value and the large SD – in isolation such a wide interval would be unacceptable.

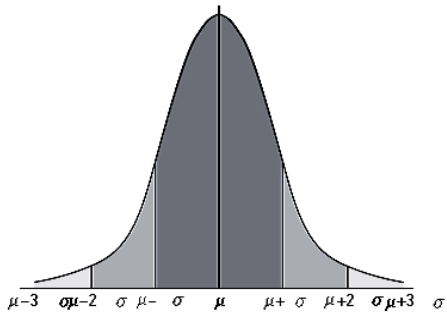

Figure 5.3 *A normal distribution frequency curve.*

Table 5.6 *Confidence intervals for percentage of protein data.*

	Analysis A	**Analysis B**
Number of replicates	2	4
Mean (%)	7.9	8.6
Range (%)	1.2	0.9
SD	0.85	0.39
CI (95%)	0.3-15.5	8.0-9.2
t-value (95% confidence)	12.71	3.18

Analysis B, using four replicates, cuts the interval to 1.2 per cent, obviously more acceptable. This fact seems to condemn low replication but if a confident SD is established initially by a larger number of replicates, or on series of at least six duplicate analyses, then this can be used to calculate a useful statistic, a form of repeatability for subsequent analysis with two replicates:

$$r = t \times \left(\sqrt{2}\right) \times SD$$

Where

r = estimated variability or repeatability which must not be exceeded;

t = value from table based on the larger original number of initial analyses;

SD = standard deviation of original number of repeat determinations under repeatability conditions.

Assuming such circumstances, an additional analysis based on ten crude protein determinations is given below (Table 5.6) along with the calculated *r* statistic. The *t*-value is smaller as it is based on the original ten determinations. Thus we would expect duplicate crude protein determinations to differ by less than 1.1 per cent, so although Analysis A looks more favourable now it could still be rejected on these grounds. Indeed, in the author's experience of the manual Kjeldahl technique on a range of food products, the precision of Analysis B (or better) is more typical and it is likely that a gross error has occurred in Analysis A. Following the above procedure now gives a more definite guide to accepting the level of precision.

5.3.5 Acceptable Level of Accuracy

The magnitude of the percentage of REM or the EM will decide this, but how large should it be before it is regarded as unacceptable? If it's based on comparison with typical values then these can vary by up to 10 per cent or more and this must be borne in mind when gauging accuracy via crude methods. Similarly "most probable" estimates are also approximations. Analysis B (Table 5.5) is within 10 present of the estimated true value whereas Analysis A exceeds this limit. The confidence interval detailed above as a precision check canals be used for accuracy, provided that confident measure of the SD was obtained – if the expected value lies within the interval then this is acceptable. In the example (Table 5.5), both analyses achieve this level of acceptance, but the Analysis A result is rejected because of the very large interval. Determination of crude protein is a proximate technique, and accuracy much beyond that of Analysis B may be an unrealistic target. For a certified RM, a similar procedure can be

applied – the determined interval should contain the certified analysed value. Additionally an interval will be quoted on the certificate – the mean value obtained by the analyst should lie within this interval. This is a more stringent test as the certificate interval is likely to be narrower. In either case, if the determined mean is out with the interval then the result can be viewed as inaccurate and this may indicate the presence of a systematic error. Depending on circumstances, there is some leeway in the decision-making process and individual analysts can decide on acceptable proximity to the precision and accuracy levels. Textbooks on analytical methods may not quote figures for acceptable accuracy and precision. Often it is left to the experience and knowledge of the analyst.

5.4 Actualisation

Study Case: Statistical Technique Application Example – Precision Calculations for Chemical Analysis Data; Source: J. A. Bower 2009.

Data gathered during routine chemical analysis of moisture content in foods were examined for the level of precision. Mean values were in the range 70–72 g/100 g and based on the data bank the population standard deviation was taken as 0.35 g/100 g. A duplicate measure was carried out under repeatability conditions.

Table 5.7 *Repeatability Calculation Data (Excel).*

Data	Moisture Content (g/100g)
Duplicate 1	71.5
Duplicate 2	70.9
Mean	71.2
sd(pop)	0.35
sd (unknown)	0.42
$t_{95\%,\ ldf}$	12.71
repeatability	0.97
repeatability	7.62

Table 5.7 shows the result of the duplicate moisture content determination. Assuming repeatability conditions as defined above, precision (repeatability) can be calculated in two ways.

As the population standard deviation (sigma (σ)) is known, then:

Repeatability 95%= z 95% × square root (2) × σ= 1.96 × square root (2) × σ

This is essentially the confidence interval for a duplicate determination, i.e. n = 2. The z value is the confidence level factor based on a normal distribution. In the example, repeatability has a value of 0.97%. Thus, duplicate determinations of moisture by the particular method in the same laboratory, same technician, reagents, etc., should differ by not more than 0.97%. The population sigma can be obtained from previous data as in the example, or by carrying out an initial larger set of determinations to give an improved estimate. If sigma is estimated as the sample *sd*, then:

$$\text{Repeatability } 95\% = t\ 95\%,\ 1\text{d}\ f \times \text{square root } (2) \times sd$$

The confidence level factor is based on the t distribution. This results in a much higher value (>7%), but it is also possible to use a t value based on a larger earlier set to give more representative measure of repeatability, and a narrower interval. In practice, duplicate moistures by oven drying give %*CV*s of <1%, thus the former estimate of repeatability is not usual. Some texts define repeatability as 'within laboratory' in a broader way in that it includes different operators, which is more realistic as it cannot be guaranteed that the same technician will analyse all incoming samples at one session, etc. Repeatability can also be obtained via the 'within variance' estimate in a two-way ANOVA.

5.5 Reproducibility

Repeatability is a critically important measure for a laboratory, but there are many analytical laboratories and there is concern that results can vary depending on which laboratory is used. There are many examples of such discrepancies in the literature (e.g. Thompson 1994), with z-scores (as above) attaining values well above the ±2 limit for some laboratories (±10 in some instances). This has given rise to a further definition of precision. The term is expanded to cover variation between different laboratories – *reproducibility*. This is the variability where all aspects other than the material being analysed are different, i.e. analysis in different laboratories, hence different technicians, reagents, times etc. The definition is calculated in a similar manner to that for repeatability, with the inclusion of the 'different laboratory effect': reproducibility is the magnitude of the interval for 2 determinations by any two laboratories. The calculation reflects the wider source of variation by incorporating the variance of both within- and between-laboratory sources:

Reproducibility 95% (population variance known):= z 95% × square root (2) × square root (variance within+ variance between)

And

Reproducibility 95% (variance estimated): = t 95%, 1 × square root (2) × square root (estimated variance within+ estimated variance between)

Table 5.8 *Estimation of within- and between-laboratory variance using ANOVA (Excel).*

Data	**Moisture content (g/100g)**					
	Lab A	**Lab B**				
	71.1	72.8				
	71.7	72.9				
Mean	71.4	72.85			Estimate	Known
sd(pop)	0.18	0.09	Within variance		0.093	0.065
sd(unk)	0.42	0.07	Between variance		2.10	1.6
$T_{95\%,1df}$	12.71	12.71				
$Reliability_z$	0.50	0.25	$T_{95\%,1df}$			
$Reliability_t$	7.62	1.27	$Reproducibility_z$			
			$Reproducibility_t$			
ANNOVA: single factor						
Summary						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Lab A	2	142.8	71.4	0.19		
Lab B	2	145.7	72.85	0.005		
ANNOVA						
Source of variation		*SS*	*df*	*MS*	*F*	*P - value*	F_{crit}
Between groups		2.103	1	2.10	22.73	0.041	18.51
Within groups		0.185	2	0.093			
Total		2.288	3				

The estimates in the second formula are conveniently obtained by ANOVA, which provides a 'pooled within variance' as well as the 'between variance'. This is shown for two laboratories – laboratory A compared with Laboratory B (Table 5.8). The variance of each laboratory is given and the average of this is the 'pooled within variance' – also shown as the 'within group mean square' (MS). The 'between variance' is the 'between groups MS'. These were used to calculate the reproducibility of these two laboratories. As seen, the reproducibility figures are higher than those of repeatability, especially when the variance estimates are based on two determinations (again, this can be improved using a known or an established more confident variance estimate). The above analysis is limited in that it cannot provide any indication of interaction of laboratories with different concentrations or levels of a particular analyte, e.g. a food with a higher or lower

level of moisture. Inclusion of a second set of replicates from another food type allows interaction to be assessed. Here, reproducibility is determined via estimates of *within*, *between* and *interaction*, which are adjusted to take into account the number of determinations, etc. Reproducibility is still high when more than one concentration of the moisture content is considered, but the differing concentration does not appear to cause a significant effect. Such work is done in inter-laboratory proficiency-testing schemes, and large precision studies can involve ten or more laboratories examining a particular method at several levels of concentration.

Table 5.9 *Reproducibility with interaction (Excel).*

Data	**Moisture content (g/100g)**						
	Lab A	**Lab B**					
s1	73.4	75.1					
	72.7	74.2					
s2	66.5	67.9					
	68.0	67.2					
Variance Estimates							
Within Lab	0.51		=*Within*				
Between Lab	0.24		=*columns – interaction / number of determinations per lab*				
Interaction	0.17		=*interact – within / number in each cell*				
$T_{95\%,4df}$	2.78		*df = within*				
Reproducibility	3.76						
Annova two – factor with replication							
ANNOVA							
Source of variation	*SS*	*df*	*MS*	*F*	*P - value*	F_{crit}	
Sample	83.21	1	83.21	164.76	0.0002	7.71	
Columns	1.81	1	1.81	3.57	0.13	7.71	
Interaction	0.84	1	0.84	1.67	0.26	7.71	
Within	2.02	4	0.51				
TOTAL	87.88	7					

5.6 Discussion

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, manufactures 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.

5.7 General Recommendations

1. Food manufacturers and producers should adhere and comply with national standard bodies through established analytical procedures, regulations and standards.

2. Research scientists should use statistical techniques in experimental and research work and present findings or results that are empirical, accurate and precise. Update information on new statistical software and packages should be adopted and included in educational curriculums for study programmes.

3. Application of statistical techniques in the following areas should be promoted:

   Instrumental measures - covering any measurement system from chemical and physical analysis to specific food instrumentation methods and process measures, e.g. protein content, Lovebird colour measures, air speed setting, etc.

   Sensory measures – to include all sensory tests used by trained assessors such as discrimination tests and descriptive analysis methods. *Consumer tests*– to include some sensory methods, which are affective or hedonic in nature, e.g. preference ranking. Generally systems should cover mostly laboratory measurements.

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

5.8 Conclusion

Food issues are becoming increasingly important to consumers, most of who 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 practitioner 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'; Sal kind 2004) to realise this. Even the classical sciences can have problems. Professional food scientists may have received statistical instruction, but application maybe 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 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 on-going 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 statistical approach to experimentation.

References

[1] Abdullah, B. M. (2007) Properties of five canned luncheon meat formulations as affected by quality of raw materials. *International Journal of Food Science and Technology*, 42, 30–35.

[2] Anonymous (1999) Reference materials update. *VAM Bulletin*, 20, 33.

[3] AOAC (1999) *Official Methods of Analysis of the Association of Official Analytical Chemists*, 15th ed. Association of Official Analytical Chemists, Arlington, VA.

[4] Association of Public Analysts (1986) *A Protocol for Analytical Quality Assurance in Public Analysts' Laboratories*. Association of Public Analysts, London, UK.

[5] Association of Public Analysts, A Protocol for Analytical Quality Assurance in Public Analysts' Laboratories, Blumberg, B., Cooper, D. R. and Schindler, P. S. (2005) *Business Research Methods*. McGraw-Hill Education, Maidenhead, UK, pp. 18–25.

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

[7] BSI (2000) BS ISO 5725:1. *Accuracy (Trueness and Precision) of Measurement Methods and Results – Part 1: General Principles and Definitions*. British Standards Institute, London, UK.

[8] BSI (2002) BS ISO 5725:2. Accuracy (Trueness and Precision) of Measurement Methods and Results – Part 2: Basic Method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method. British Standards Institute, London, UK.

[9] Calcutt, R. and Boddy, R., Statistics for Analytical Chemists, Association of Public Analysts, London, 1986.2 Chapman& Hall, London, 1983.

[10] Chatfield, C. (1992) *Statistics for Technology*, 3rd ed. Chapman & Hall, London, UK. Collis, J. and Hussey, R. (2003) *Business Research*. Palgrave MacMillan, Basingstoke, UK, pp. 46–79.

[11] Fisher, R. A. (1966) *The Design of Experiments*, 8th ed. Hafner, New York.

[12] FSA (Food Standards Agency) (2004a) Optimised Uncertainty at Minimum Cost to Achieve Fitness for Purpose in Food Analysis. FSA Project No. E01034. Available atwww.food.gov.uk.

[13] FSA (Food Standards Agency) (2004b) Pilot Study of Routine Monitoring of Sampling Precision in the Sampling of Bulk Foods. FSA Project No. E01049.

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

[15] Gao, Y., Ju, X. and Jiang, H. (2006) Studies on inactivation of *Bacillus subtilis* spores by high hydrostatic pressure and heat using design of experiments. *Journal of Food Engineering*, 77, 672–679.

[16] Greenfield, H. and Southgate, D. A. T. (2003) *Food Composition Data*, 2nd ed. Food and Agriculture Organisation of the United Nations, Rome, pp. 149–162.

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

[18] Heinonen, M., Valsta, L., Anttolainen, M., Ovaskainen, M., Hyv ̈onen, L. and Mutanen, M. (1997) Comparisons between analytes and calculated food composition data: carotenoids, retinoids, tocopherols, tocotrienols, fat, fatty acids and sterols. *Journal of Food Composition and Analysis*, 10, 3–13.

[19] Hill, A. R. C. and von Holst, C. (2001) A comparison of simple statistical methods for estimating analytical uncertainty, taking into account predicted frequency distributions. *Analyst*, 126, 2044-2052.

[20] 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.

[21] J ̈ulicher, B., Gowik, P. and Uhlig, S. (1999) A top-down in-house validation based approach for the investigation of the measurement uncertainty using fractional factorial experiments. *Analyst*, 124, 537–545.

[22] Kane, J. S. (1997) Analytical bias: the neglected component of measurement uncertainty. *Analyst*, 122, 1283–1288.

[23] 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.

[24] Levermore, R. J. and McLean, B. D. (2002) Development of a Stable Sulphated Food Reference Material, Campden and Chorleywood Food Research Association Group. FSA Project code E01041. Available at www.food.gov.uk.

[25] Lyn, J. A., Ramsey, M. and Wood, R. (2002) Optimised uncertainty in food analysis: application and comparison between four contrasting 'analyte–commodity' combinations. Analyst, 127, 1252–1260.

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

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

[28] Mart´ın-Diana, A. B., Rico, D., F´ıas, J. M., Barat, J. M., Henehan, G. T. M. and Barry-Ryan, C. (2007) Calcium for extending the shelf life of fresh whole and minimally processed fruits and vegetables: a review. *Trends in Food Science and Technology*, 18, 210–218.

[29] May, N. S. and Chappell, P. (2002) Finding critical variables for food heating. *International Journal of Food Science and Technology*, 37, 503–515.

[30] Miller, J. C. and Miller, J. N. (1993) *Statistics for Analytical Chemistry*, 3rd ed. Ellis Horwood, Chichester, UK.

[31] Miller, J. C. and Miller, J. N. (1999) *Statistics and Chemometrics for Analytical Chemistry*, 4th ed. Ellis Horwood, Chichester, UK.

[32] Moreira, M. D. R., Ponce, A. G., Del Valle, C. E. and Roura, S. I. (2006) Ascorbic acid retention, microbial growth, and sensory acceptability of lettuce leaves subjected to mild heat shocks. *Journal of Food Science*, 71(2), S188–S192.

[33] Mullins, E. (2003) *Statistics for the Quality Control Chemistry Laboratory*. Royal Society for Chemistry, Cambridge, UK.

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

[35] Miller, J. C. and Miller, J. N. (1999) *Statistics and Chemometrics for Analytical Chemistry*, 4th ed. Ellis Horwood, Chichester, UK.

[36] Nielsen, S. S. (2003) *Food Analysis*, 3rd ed. Kluwer Academic/Plenum Publishers, New York.

[37] O'Donnell, G. E. and Hibbert, D. B. (2005) Treatment of bias in estimating measurement uncertainty. *Analyst*, 130, 721–729.

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

[39] Patey, A. (1994) the food analysis performance assessment scheme. *VAM Bulletin*, 11, 12–13.

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

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

[42] Severini, C., Baiano, A., De Pilli, T., Carbone, B. F. and Derossi, A. (2005) Combined treatments of blanching and dehydration: study on potato cubes. *Journal of Food Engineering*, 68, 289–296.

[43] Therdthai, N., Zhou, W. and Adamczak, T. (2002) Optimisation of the temperature profile in bread baking. *Journal of Food Engineering*, 55, 41–48.

[44] Thompson, M. (1994) Proficiency testing in analytical laboratories – the international harmonised protocol. *VAM Bulletin*, 11, 4–5.

[45] Villavicencio, A. L. C. H., Ara újo, M. M., Fanaro, G. B., Rela, P. R. and Mancini-Filho, J. (2007) Sensorial analysis evaluation in cereal bars preserved by ionizing radiation processing. *Radiation Physics and Chemistry*, 76, 1875–1877.

[46] Wakefield, D. and McLaughlin, K. (2005) *An Introduction to Data Analysis Using Minitab R for Windows*, 3rd ed. Pearson Education, Inc., Pearson Prentice Hall, Upper Saddle River, NJ.

[47] Williams, A. A., Rogers, C. A. and Collins, A. J. (1988) Relating chemical/physical and sensory data in acceptance studies. *Food Quality and Preference*, 1(1), 25-31.

[48] XiuRong, P. (1995) A view on the traceability of certified values of chemical composition RMs. *VAM Bulletin*, 12(reference material special), 18–19.

[49] Yann, D., Didier, H. and Daniel, B. (2005) Utilisation of the experimental design methodology to reduce browning defects in hard cheeses technology. *Journal of Food Engineering*, 68, 481–490.