
Statistics for food science I

John A. Bower

The author

John A. Bower is a Lecturer in Food Science in the Department of Applied Consumer Studies, Queen Margaret College, Edinburgh, Scotland, UK.

Abstract

Begins an introduction to statistics for the food scientist. Aimed at the non-mathematician, discusses the application of the statistical approach and introduces statistical measures for data presentation and summary.

Food science is a multidisciplinary subject, typical curricula including chemistry, bio-chemistry, microbiology, physics, marketing, business studies as well as some mathematical input. The practising food scientist, whether in the laboratory, on the production floor or in the marketing office has the task of interpreting data and making important decisions on his/her conclusions. It is here that the subject of statistics enters. While many undergraduate scientists receive at least one statistics course the instillation of statistical appreciation may be long in the attainment. Indeed, the recent UK STEPS project (statistical education through problem solving) recognizes problems of "motivation" and "a wide spread of numerical skills" when teaching statistics as a support subject to non-mathematicians. Even later, when established in industry, some food scientists may be wary and reluctant about using statistics or attempting the statistical approach to experimentation. This series of articles will take one route through this topic illustrating advantages and pitfalls on the way; with the scientist but "non-statistician" in mind this is not intended to be a statistical handbook. Formulae and theory will be kept to the minimum and readers are referred to the recommended titles in "Further reading" for fuller details.

Where are statistics used in food science?

Statistical procedures can be applied to several areas within food science:

- sensory evaluation;
- instrumental measures of quality;
- correlation of sensory and instrumental measures;
- chemical analysis;
- statistical quality control;
- comparison of nutritional effects of processing.

All of the above involve gathering *data* or making measurements of some kind. Science is based on recording measurements and within food science there can be a wide range of different types (Table I).

As can be seen, statistics deals with the "collecting, organizing and analysing" of data, usually in the form of numbers as above. The particular feature being measured or observed is described as a variable. Each measurement or observation is subject to variation owing to

Table I Data and measurement types in food science

Application	Measurement	Example
Chemical analysis	Mass and volume	g/100g, w/w, w/v
Instrumental quality	Textural properties	Force (newtons) viscosity
	Colour properties	CIE L*a*b* units
Sensory evaluation	Taste panel score sheets	Intensity ratings, hedonic ratings, magnitude ratios, rankings

one or more factors, random or otherwise, some of which contribute to error. Measurements can be single or replicate depending on the circumstances. Statistical methods can also be employed in the way in which the data are obtained (“statistical sampling”).

Let us envisage a typical “data confrontation” situation in which the scientist finds her/himself – the data may be in a report or publication or may have been generated by the scientist. The problem is one of deciding several things: What do these data mean? Are the data valid? Are they convincing enough to progress on to more work?

To provide answers to these questions, and to aid statistical appreciation, a step-wise approach is recommended.

What is the nature of the data?

In many cases the analysis of data depends on the nature of the particular data, i.e. the level of measurement used.

Data types and levels of measurement

Nominal data

This is the lowest level of measurement, where data consists of the incidence of one or more classifications based on a particular attribute, e.g. colour classification of apples as green, red or yellow, thus “a sample of 100 apples was taken from the delivery – 84 were red, 14 were yellow and only 2 were green”. Also known as attribute or category data.

Ordinal data

These are obtained when order is used within a measurement scale. The order can be one of simple rank or can involve use of a scaled list of categories or descriptions, e.g. a hedonic scale (Table II) used in food taste panels.

Any numerical units on an ordinal scale cannot be interpreted in an arithmetical manner, i.e. in the above example a score of 3 cannot be described as being “3 times less liked than a score of 1”.

Table II An ordinal data scale

Like very much	1
Like	2
Like slightly	3
Neither like nor dislike	4
Dislike slightly	5
Dislike	6
Dislike very much	7

Interval data

This is used for measurement scales where the distances or intervals on the scale are more meaningful and operations such as subtraction and addition provide valid information, e.g. a temperature scale where 8°C is 4°C higher than 4°C but 8°C cannot be considered to be twice as hot as 4°C. Other examples are given by taste panel score sheets where an ascending/descending scale describes a specific graded change in some aspect of quality, e.g. the progressive increase in trimethylamine odour in stored fresh fish.

Ratio data

This is the highest level of measurement. Multiplication and division result in valid comparisons of the data, e.g. a food sample weight of 20g is four times heavier than a sample weight of 5g.

What kind of population could they come from?

This is an important feature because it can influence the choice of statistical method to be used to analyse the data. The most important of the population types encountered is the normal population which has certain specific parameters which define it. Many of the data encountered in chemical and physical analyses will be normal, but for data from sensory evaluation assessments this may not be so. There are tests available to check the data for normality, but if in doubt the scientist is recommended to be conservative and assume

non-normality unless experience or advice indicates otherwise.

How were the data gathered/obtained?

Collection of data – sampling

Data are obtained by taking samples from a parent population. It is possible to take all samples in, or members of, a population, but usually the sample is a small fraction of the total population. In food science the population can be discrete in nature or continuous, e.g. taking three cans every hour from a canning line is a discrete sample; taking weighed samples from a large batch of a food ingredient is a continuous sample, where the sample could be one of an almost infinite number of possible weights. Irrespective of sample type the sample must be representative of the whole, that is, it must provide a valid measure of the particular attribute being examined or give a true reflection of the situation which exists in the population. To assess the data for a representative sample, check whether or not any of the following procedures were followed:

- Was sample selection done on a *random* basis, where every possible member or part of the population had an equal chance of being selected? For example: a food factory produces 6,000 packs of frozen meals a day; 40 packs per day are to be taken out for examination – if this is done using a fixed timetable, the sample is likely to be invalid. A truly random method could involve using random number tables or a random number generator to decide on 40 individual time slots per day during production. At least the sampling should be spaced across any possible source of variation, e.g. if there are five production runs per day then eight samples are randomly selected from each.
- Was the sample continuous in nature and if so was the sample size indirectly proportional to the population homogeneity? For example: a large silo of flour is sampled daily for moisture content analysis. The production process used for the flour ensures a well mixed fine particle powder, very homogeneous in character – duplicate 5g samples per hour may suffice. This contrasts with a batch of meat filling which contains ten separate ingredients more heterogeneous in compositional content, and requiring five 50g samples per hour.

- Was the sample obtained from a taste panel? If so, then the food samples which were tested should themselves have been selected by an appropriate sampling technique, but the number of times each sample is subjected to the test is often decided by the number of tasters. Were the numbers of tasters and tastings sufficient for the type of test? Was it an expert panel or a consumer panel? The statistical significance of taste panel results can be influenced by these points and they are therefore important considerations in the planning of these tests.

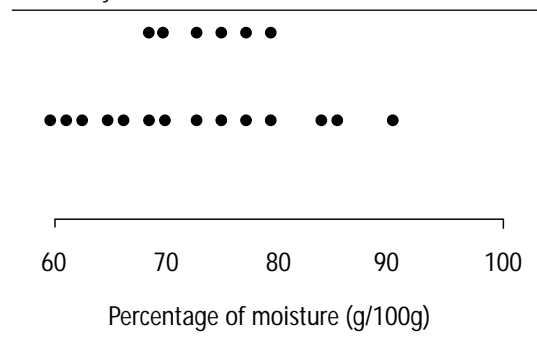
How are the data presented?

Organization of data

Some indication of how the data are presented or displayed should now be sought. The simplest method could be a list of numbers. Data are difficult to interpret in this form and require organization. Graphical presentation by actual plotting of the results on a graph would yield more information, but the *histogram* and *pie-chart* yield diagrams which are perhaps visually more informative. The columns or bars on the histogram and the slices on the pie-chart represent individual items of data. Many other techniques are possible such as the “blob chart” (also the cross chart or dot chart) where the data values are marked as blobs or crosses on an appropriate scale. This is a simple, rapid procedure which can quickly reveal a general trend in the nature of the data (Figure 1).

The above methods of presentation would become cumbersome with many data items. Putting the data into groups or classes, then displaying the frequency of each group can make large data sets more visually assessable, thus forming a *grouped frequency distribution*.

Figure 1 Blob chart of the percentage of moisture content of 20 ready meals



This may also be planned before collection of data and involves grouping or classing the individual data items into smaller ranges or bands, then displaying the incidence or frequency of occurrence for the groups. For graphical display of frequency distributions the groups should be of equal width.

Up until now we have considered the individual data items or groups which have not been treated to yield *derived* data (as distinct from the *raw* data). Much more useful information can be gained by a variety of procedures which essentially summarize the data and reduce the amount of numerical values to be examined (data reduction).

How have the data been summarized?

Initial analysis of data

Range

This is a very simple procedure, reducing the data to two values: the minimum and the maximum.

Mean

This is the average value, universally recognized, reduction of the data, no matter how large the sample, to a single value. Obtained by addition of all individual values to give the total or sum, then dividing by the sample number:

$$\text{Mean} = \frac{\text{Sum of all the data values}}{\text{The number of data values}}$$

Median

This is the value in the data set with an equal number of values above and below. In the case of an even data set the “middle two” values are averaged.

Mode

This is the most frequently occurring value.

The latter three forms of analysis have reduced the data to a single value, each characterized by its location on the measurement scale relative to the range of values, giving rise to the term “measures of central tendency” (although the mode may not occur as centrally as the mean and median). Caution is recommended in taking these reduced data measures at face value and they should be viewed along with other data summaries and presentation such as a frequency distribution, which can reveal important information such as the presence of a bimodal response.

Interquartile range

The range of the data values gives an indication of how the data are spread (no indication is given by mean, median and mode). The interquartile range is obtained by the difference between the upper and lower quartiles of the data, where the upper quartile is given by the value with a quarter of the values above (three-quarters below) and the lower quartile is the value with one-quarter below.

The procedures above will have resulted in a smaller number of values to look at, but to begin assessing their validity some way of expressing another dimension of the data is required.

How does the data set vary or what is the data “spread”?

Measures of variation

The above calculated values provide little or no indication of how the data set varies. The mean, median and mode give no indication and can thus be misleading if examined in isolation. The range gives some insight into the degree of variation but is based on only two values and if one or both of these varies greatly in comparison with the other values in the data set then again a false impression can be created. More powerful measures of variation are required. It is here that the first indications of *validity* regarding the data can be detected.

The mean deviation (also known as the average deviation)

As the mean gives a measure of central tendency, a measure of variation of the whole data set can be obtained by calculating the difference between each value and the mean. This gives a derived data set which if summed should total zero as both positive and negative values will be present. In the case of the mean deviation this drawback is overcome by simply ignoring the sign of each deviation (i.e. the absolute value is taken), summing the values then dividing by the number of values:

$$\text{Mean} = \frac{\text{Sum of the absolute deviations of each data value from the mean}}{\text{The number of data values}}$$

The magnitude of the mean deviation will be directly proportional to the degree of variation in the data, i.e. the more the data vary the greater the mean deviation.

The standard deviation and variance

Another way of solving the problem of a zero value for summed deviations is to square them. Summing the squares then dividing by the number of values would give a measure of the sample variance for the set of values. The variance would not be in the same numerical units as the data values, but the square root of the variance would, giving the sample standard deviation, the most well-known and widely-used statistical measure of variation. The inclusion of the qualifying adjective “sample” is perhaps confusing, but it is a necessary distinction. It has been proven by statistical method that calculation of variance and standard deviation by use of the sample data number provides information relating to the sample only as distinct from the population from which the sample comes. Usually it is intended to use the data to obtain information concerning the population and thus amended formulae are used (the same arguments apply to the calculation of the mean deviation but this statistic is more used in analytical data analysis where the sample variation is more under scrutiny):

$$\text{Variance} = \frac{\text{Sum of the squared deviations of each value from the mean}}{\text{The number of data values less one}}$$

and

$$\text{Standard deviation} = \sqrt{\text{variance}}$$

The mean deviation and the standard deviation are both measures of *precision* and should follow a similar trend for different data sets. The standard deviation is more widely used in general statistics, one reason being that the variance value is more sensitive to large deviations than the mean deviation.

A related measure is now appropriate.

How do the data vary from the true value, i.e. how accurate is the result?

To enable calculation of a statistic for accuracy the actual true value must be known, i.e. the “correct” result. In some cases this is available as in, for example, a sensory difference test where the organizer knows that one

sample is different and thus accuracy is “in-built”, although this view is a simplification. A more obvious problem relating to accuracy is found in chemical analysis where the analyst has an unknown sample for analysis. As part of any professional protocol calibration via known analyte standards and analysis of a reference sample of known analyte content should be included, that is, the analyst analyses material of known composition along with the unknown. These precautions form part of the traceability procedures endorsed by the DTI’s valid analytical measurement initiative. The analysed reference mean can then be compared with the certificated true value and the error of the mean can be calculated:

$$\text{Error of the mean} = \frac{\text{Deviation of the experimental mean from the true mean}}{\text{...}}$$

Note that it is not possible to calculate the error of the mean for the unknown sample unless reliable independent analysis data are available, but the reference error will give an indication.

Of course the big question regarding both measures of precision and accuracy is: what magnitude of precision and accuracy is acceptable?

The answer to this question involves looking at specific data sources and must wait until next time.

Further reading

- Bender, E.F., Douglass, L.W. and Kramer, A. (Eds), *Statistics for Food and Agriculture*, Food Products Press, New York, NY, 1988.
- Cohen, S.S., *Practical Statistics*, Edward Arnold, Sevenoaks, 1988.
- Gacula, M.C. and Singh J., *Statistical Methods in Food and Consumer Research*, Academic Press, Orlando, IL, 1984.
- Miller, J.C. and Miller, J.N., *Statistics for Analytical Chemistry*, 3rd ed., Ellis Horwood, Chichester, 1993.
- Rees, D.G., *Essential Statistics*, 2nd ed., Chapman & Hall, London, 1991.
- Rowntree, D., *Statistics without Tears – A Primer for Non-mathematicians*, Pelican Books, Harmondsworth, 1981.
- Wynne, J.D., *Learning Statistics*, Macmillan, New York, NY, 1982.