6. Measurement Principles

Peter Baxter

Learning objectives

On completion of this topic you should have an understanding of:

- the meaning of common statistical terms used in wool metrology
- the effects of sources of variation on measurement techniques
- the development of standard test methods

Key terms and concepts

Mean, variance, standard deviation, coefficient of variation, standard error, Student's t test, confidence limits, correlation, regression, precision, bias, accuracy, components of variance, variance models

Introduction to the topic

Wool is a very variable material – it varies fibre to fibre, staple to staple, fleece to fleece, etc. Topic 1 – Background to Wool Metrology explains why this variation occurs. In order for measurements to have any meaning in terms of their representing the average properties of the lot of wool, it is necessary to have some understanding of the sources of variation and how they are described. This topic introduces some common statistical terms used in the description of variation, together with the more common techniques used to evaluate the effects of variation and thereby produce relatively robust sampling and measurement methods.

6.1 Statistical terms commonly used in wool metrology

It is important to understand that:

Measurement Data = Information + Error

What we are trying to establish is the information, but what we are starting with is data. There are many sources of variation that contribute to error in measurement. Many of these relate to sampling since the product that we are trying to characterise by measurement is in itself very variable. The sampling regime normally has to be designed to ensure that the most practically homogenous sub-sample is presented for measurement. Even if the sub-sample has been blended to minimise the amount of variation, errors will still arise due to the limitations of the measurement process. These may be caused by many different effects, such as the operator, vagaries in the test specification, the environment, the repeatability of the instrument, or the imprecision of the calibration. This means that no two measurements undertaken on the same sample are likely to give exactly the same result. We therefore have to make use of some basic statistics to characterise the test results. The two main things that we are concerned with are measures of location and of dispersion. The former relates to the value of the measurement (i.e. the "answer"), whereas the latter relates to how variable the result may be.

Mean

The arithmetic mean is the most common statistic of location used in wool metrology. It is calculated by summing all the individual observations or measurements and dividing the sum by the number of items in the sample. It is also known as the average and is calculated in Excel using the function AVERAGE.

As an example, if 4 individual readings of mean fibre diameter (in μm) are 17.9, 18.2, 18.1, and 17.8, then the mean is (17.9+18.2+18.1+17.8)/4 = 18.0 μm.

It should be noted that there are other statistics of location that are useful in some circumstances, such as the geometric mean, median, and mode. These are useful in reducing the effects of outliers or when dealing with non-normal or skewed distributions, but are not extensively used in routine wool metrology.

Variance

The variance is the most fundamental statistic used to represent variability. Variability is represented by the differences of the individual observations from the mean, but it only takes a moment's thought to realize that the sum of the differences from the mean must always be zero, so variability must be measured in a manner that ignores the sign of the differences. The variance is the sum of the squares of the differences between the individual observations and the mean, divided by the number of degrees of freedom. The degrees of freedom concept can become complex, but in univariate statistics is usually the number of observations minus one (n-1).

In the example above, the variance may be calculated as follows:

Mean	observation	difference	square
18.0	17.9	-0.1	0.01
18.0	18.2	+0.2	0.04
18.0	18.1	+0.1	0.01
18.0	17.8	-0.2	0.04
		Sum of squares	0.10
		Divided by (n-1) = variance	0.033

The units of variance are the measurement unit squared, so in the example above, the units are μm². Variance can be calculated in Excel using the function VAR.

Standard deviation

Whilst variance may represent the fundamental statistic of dispersion, it is not the easiest to work with since its units are squared. The standard deviation is the square root of the variance and is an easier statistic to visualise since the units are once again those of the original measurement. In Excel the appropriate function is STDEV. The standard deviation simply represents the amount of variation associated with a group of measurements or observations.

In the example above, the standard deviation of these 4 measurements is the square root of 0.033 = 0.18 μm.

Coefficient of variation (CV)

For some wool properties, the value of the standard deviation varies with the value of the mean. A common example is the standard deviation of fibre diameter, which on average increases as the mean fibre diameter increases. This sometimes makes it difficult to compare levels of variation. The coefficient of variation is a 'normalised' form of the variation and is calculated by dividing the standard deviation by the mean and is expressed as a percentage. It has been shown, for example, that the average coefficient of variation of fibre diameter is close to 19% for fleece samples across a very wide range of diameters, whereas the average standard deviation varies from 2 μm at 12 μm to 10 μm at 44 μm (Baxter & Cottle 1998).

In the example we have been working with, the coefficient of variation of the 4 measurements is 100 * (0.16 / 18.0) = 8.9%.

Standard error and hypothesis testing

There is often confusion amongst students about the difference between standard deviation and standard error. Standard error is associated with a statistic as opposed to standard deviation, which usually describes a group of measurements or observations. Standard error is a method of describing uncertainty or precision of a statistic such as the mean, regression coefficients, correlation coefficient, etc.

Statistical texts describe how the standard error may be determined for most common statistics. However, it is important to understand the most common use of standard error – in estimating the precision of a mean. This is easily calculated as the standard deviation of the group of measurements, divided by the square root of the number of measurements. In the example above, the standard error of the mean value of 18.0 is 0.18 / sqrt (4) = 0.09 µm.

The standard error is used in judging whether two statistics are similar or likely to be significantly different. In statistics textbooks this subject is covered under "hypothesis testing". The most common application of hypothesis testing in wool metrology is to assess whether the average differences between the paired values of two groups of measurements is likely to be zero. For example, some slightly different treatment might be applied to a processing route, and the requirement is to establish whether the change has a significant effect or not. Whilst there is insufficient scope in this topic to cover this subject in the depth required, the student is referred to any standard text on Basic Statistics. We shall very briefly cover the Student's t test.

Student's T Test

This test assumes that the distribution of observations is normal (see below). The Student's t statistic is used to test a hypothesis. To test whether the average paired differences between two sets of measurements is likely to be zero, the so-called "null hypothesis" is that the mean difference = 0.0. To test this, the actual mean difference is divided by the standard error of the mean, and the resulting t value compared with a standard table of critical values of the t-distribution. Such a table lists the limiting values as function of probability and number of degrees of freedom, and is usually displayed as 'two-tailed' values (in other words it doesn't matter whether the difference is greater or less than the null hypothesis value, whereas a 1-tailed test would be used, for example, if the hypothesis were that the difference was greater than a certain value). For a simple group of paired differences, the number of degrees of freedom is simply the total number of differences minus 1. In 'normal' practice the probability value chosen is 0.05. It should be noted that whilst this value is in common use, there is no theoretical or absolute reason for this choice, and whilst it very often used rigidly in applying the test, common sense should also be used.

In a simple example, the differences between measurements carried out by A method and B method on the same 6 samples are: 0.5, 0.6, -0.2, 1.3, 0.8, and 0.7. The mean difference is therefore 0.62, standard deviation 0.49, and standard error 0.20. The t statistic is 0.62 / 0.20 = 3.1, and from the critical values of the t distribution table, with 5 degrees of freedom, the 0.05 probability value is 2.571 and the 0.01 probability value is 4.032. One could therefore conclude that at between the 0.01 and 0.05 probability levels, the null hypothesis (that the mean difference is zero) is unproven, i.e. it is reasonable unlikely that the mean difference is zero for this small sample set.

The observant student will have noticed that as the number of observations increases, the standard error is likely to decrease, and therefore the power of the test will increase. Students are advised to study this issue, since it is important in the design of experiments – the smaller the average difference that one is examining, and the larger the variance in the measurement system, the greater the number of samples required to prove or disprove a specific hypothesis.

Excel provides a useful function TDIST, which allows the calculation of probability from a given t-value, number of degrees of freedom, and for either the 1-tailed or 2-tailed situation. Whilst there is also a function TTEST, this compares two arrays and is not useful for examining paired differences, which is the most frequent type of test required in wool metrology.

Accuracy

A measure of the closeness of a test result to the true value. The true value of a measured quantity can only be determined by measurement systems that are calibrated by direct reference to primary standards such as length, mass, time, etc.

In practice, whilst this is the technically correct definition of accuracy, the industry may accept a concensus value as the closest that can be practically achieved to the true value. As an example, the airflow instruments are calibrated by reference to the airflow calibration standards that are issued by Interwoollabs. The reference values that are issued for these standards are determined

from the global means of a number of airflow round trials, and therefore the reference values are in fact concensus values. (However, it should also be noted that airflow mean fibre diameters which vary by more a specified amount from the projection microscope mean fibre diameters are excluded from the calibration series, and in this way there remains a linkage to the primary reference of calibrated length.)

The determination of a "true" value becomes more difficult as the processing path for the measurement is extended. In the case of raw wool core testing, for example, which involves core sampling, blending, scouring, drying, etc before specimens are presented for measurement, it is impossible to determine the true value of any of the parameters, and in this case, accuracy can only be estimated from concensus values determined in round trials between competent laboratories.

Bias

Bias is a constant or systematic difference between the true value and average of the measured results. It can only be determined by replicated measurements on blended samples of known properties. It is commonly estimated in wool metrology by using the grand mean values from round trials between competent laboratories. It is also commonly assumed (sometimes incorrectly) that calibration, and, where specified in the test procedure, validation, should eliminate bias. In practice, most wool calibration procedures are not able to completely replicate the entire sample processing path, and some degree of bias is almost inevitable. The aim of all competent laboratories is reduce any such bias to negligible levels.

Bias can arise from a number of causes, such as deviations from or misinterpretation of the sampling or test method, inadequate specification within a test method, differences between instruments, or even variability in the calibration material (Baxter 1999). Nevertheless, wool sampling and test methods are designed to minimise bias, and it is therefore commonly assumed that wool metrology measurements are carried out without bias, and therefore when we discuss errors in sampling and measurement, it is normally assumed that we are referring to **random** errors rather than systematic errors.

Precision

Precision is an indicator of the repeatability of a measurement and is often expressed in terms of a confidence limit (see above).

We are normally concerned with the precision of a test method, and there may be two figures quoted. One (the larger) includes the between-laboratory component of variance, and is the figures used to determine whether results from two different laboratories may be statistically significantly different. (This is called *reproducibility* in European standards). Where the within-laboratory precision is quoted, this can only be used to assess whether two results obtained within the same laboratory on the same bulk sample are likely to be different. (This is commonly referred to as *repeatability* in European standards, although technically this only applies to measurements by the same operator with the same instruments in the same laboratory).

Precision is determined by one or more round trials. The IWTO procedures are outlined in IWTO-0. In wool metrology we are often handicapped by the limited number of competent laboratories available, and with the more important measurements, it is not uncommon for several round trials to have been carried out before a test method can become accepted.

95% Confidence limits

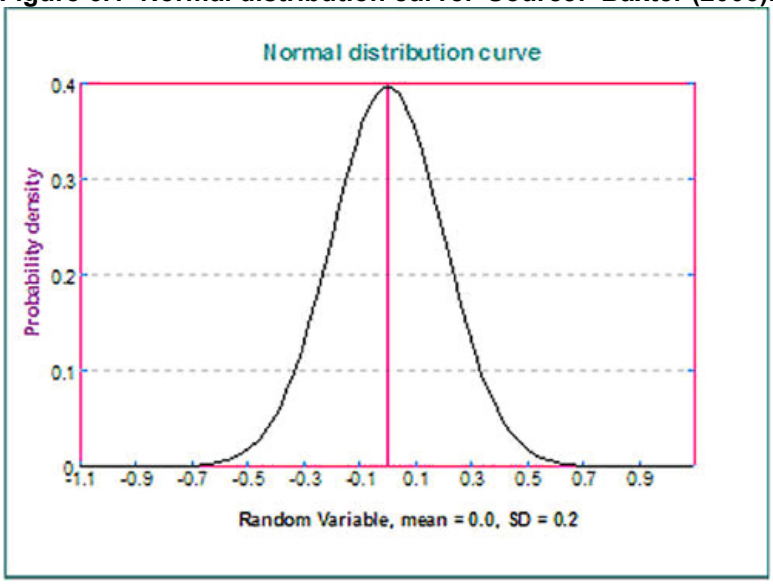
Thus far we have been looking at ways of describing a series of measurements. It is a useful factor that the more measurements we carry out, irrespective of the shape of the distribution from which they originally came, the closer each these statistics (mean, variance, standard deviation and coefficient of variation) tend to approach limiting values which are themselves normally distributed (the central limit theorem). It is convenient to suggest that, in the absence of *bias* (see above), these limiting values are likely to be good estimators of the "true" values for the sample, population, lot, or whatever it is that we are trying to describe.

To take this further we must think in terms of a probability of a single measurement giving a value related to the "true" value. To do this we use the concept of a probability distribution, which simply put, relates the probability that a single measurement will yield a result somewhere near the true value. Luckily, most wool measurements follow a "normal" probability distribution, which is one of the more common probability distributions in nature. (Whilst wool fibre diameter distributions technically tend to follow a lognormal distribution, the degree of non-normality is relatively small and is generally ignored.)

Normal distributions are a family of distributions that have the same general shape. They are symmetric with scores more concentrated in the middle than in the tails. Normal distributions are sometimes described as bell shaped.

The height of a normal distribution can be specified mathematically in terms of two parameters: the mean and the standard deviation. This is very convenient, since if we have values for the mean and standard deviation, we can calculate the probability of values appearing within certain ranges, and these are tabulated in standard texts. For example, 68.3 % of observations fall within ± 1 standard deviation from the mean; 95.5% fall within ± 2 SDs, and 99.7% fall within ± 3 SDs.

Figure 6.1 Normal distribution curve. Source: Baxter (2006).



Once we have assumed that the measurements follow a normal distribution, and we have some information about the variability of that measurement, then we can start to express a level of confidence about an individual measurement representing the "true" value. The two levels of confidence that are normally used in wool metrology are 95% (±1.960 sd) and 99% (±2.576 sd), representing the probability that 95 times out of 100, or 99 times out of 100 respectively, that the measurement will lie within a specified distance of the "true" value (again, in the absence of bias).

The precision of wool measurements is often expressed as the 95% confidence level (95%CL). This is determined by calculating the total variance associated with the sampling and measurement process and converting this to a standard deviation. We know, from the shape of the normal distribution, that 95% of all measurements in a normal distribution will lie within ± 1.96 standard deviations of the mean, so the 95%CL is simply obtained by multiplying the standard deviation by 1.96.

So, for example, in Table D3 of IWTO-12, the test method for using the Sirolan-Laserscan, the total variance of the method for aqueous scoured cores of less than 26.0 µm mean fibre diameter is calculated to be 0.0364 µm². The 95%CL is therefore 1.96 * sqrt (0.0364) = 0.37 µm

A confidence interval is a contiguous range of values within which the “true” value of the statistic will be found with some predetermined probability. So in the case of a Laserscan result of, say, 20.0 µm mean fibre diameter, the 95% confidence interval using these figures would be 19.63 to 20.37 µm.

Confidence limits can be calculated in Excel using the function CONFIDENCE.

For a detailed and relatively technical exposition on the practical application of confidence limits to measurement processes, refer to the 'Guide to the Expression of Uncertainty in Measurement'. This is regarded as the reference document by accreditation authorities.

Correlation

Correlation measures the intensity of association between a pair of variables. It is mathematically related to *regression*, but is not the same thing. In correlation we are concerned about whether two variables might covary – that is, vary together; whereas in regression analysis we are trying to describe the dependence of one variable on another independent variable. In the first case we are exploring possible associations, in the 2nd, we are concerned with modelling a circumstance where one variable is known to have an effect on another, and might, therefore, be used to predict another.

As a simple example, we might use correlation to discover that there is a degree of association between standard deviation and mean fibre diameter. We would be unlikely to use this association as a method to build a model to predict mean fibre diameter, but, we might, under some circumstances want to use regression to build a model to predict standard deviation from a mean fibre diameter measurement. We would, therefore, be using mean fibre diameter as the independent variable, and standard deviation as the dependent variable – in correlation there is no such distinction – all variables are "equal".

It is also not uncommon to find that two variables might be associated (as with a high correlation coefficient), but only because a third variable affects both of them. In other words the correlation is not causative, and we should be careful to avoid such assumptions, although the temptation is always on hand.

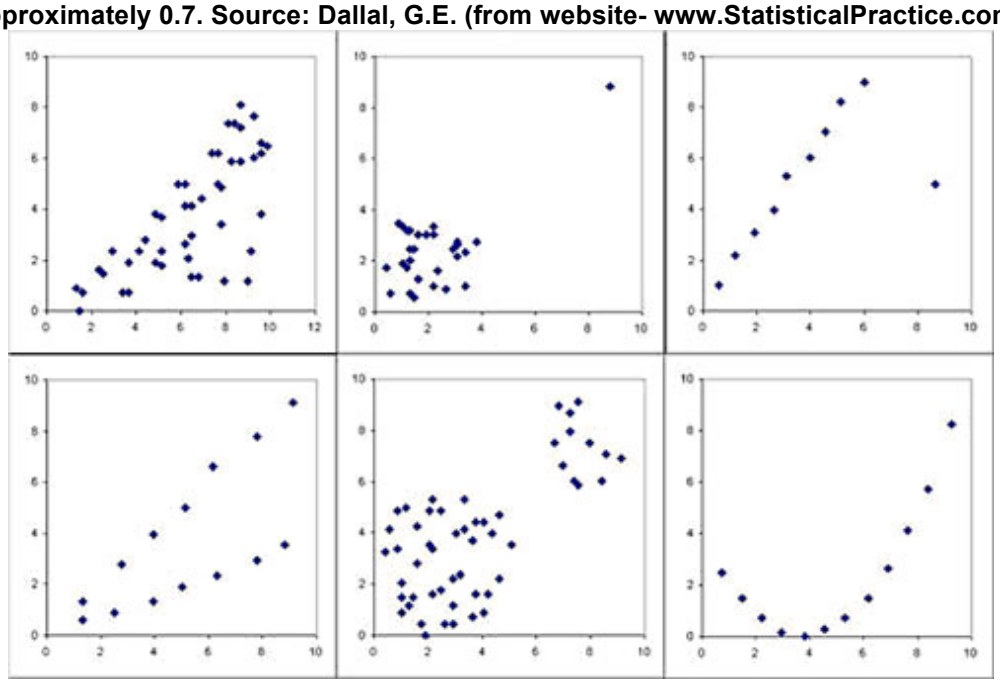
There are a number of correlation statistics in use, but we need not be concerned with them here. The most common is the Pearson product-moment correlation. The method of calculation is shown in all standard texts and a tool exists in Excel (CORREL or PEARSON).

Correlation coefficients vary from -1 to +1, with -1 being a perfect negative association and +1 being a perfect positive association and 0 being no association.

A point to be aware of is that correlation coefficients increase as the range of data increases even though the same degree of association may exist. Whilst correlation analysis is a useful preliminary data exploration tool, it should always be followed up by plotting any variables which you may think have a useful degree of association – one outlier can cause the correlation coefficient to be high even though the degree of association between two variables is poor.

Finally, correlation coefficients are linear associations. If you have a perfect quadratic association between two variables it will come out of a correlation analysis poorly. To repeat: **plot the variables**.

Figure 6.2 Illustration of 6 scatterplots all with very similar correlation coefficients of approximately 0.7. Source: Dallal, G.E. (from website- www.StatisticalPractice.com).



Regression

In general terms regression is used to allow one variable (the dependent variable) to be predicted from another (the independent variable). There may be more than one independent variable (for example the TEAM equations used to predict processing performance from core and staple test data). Usually the independent variable(s) is(are) deliberately varied (or samples selected) to give as wide a range of values as possible.

The most common form of regression used in wool metrology is linear regression, where the dependence of Y on X is expected to follow the form:

$$Y = A + B * X$$

Where A and B are the intercept (or constant) and slope coefficient respectively. In Excel these values may be calculated using the INTERCEPT, SLOPE, or LINEST functions.

However, it is by no means the rule that variables must vary linearly with each other. For example, when examining the relationship between components of precision and the main variable, it may be found that a non-linear function may fit the data better (see, for example, Stubbs & Marler 1994)

When calculating regressions, it commonly the practice to report the R^2 value. In common parlance this describes the proportion of the total variance explained by the regression. It is effectively the square of the correlation coefficient and therefore the same warnings apply to its use. When using regression to predict a variable, it always more helpful to quote the standard error of regression (SE) since this can be compared across different data sets with confidence.

One further example of regression should be mentioned. Normally regression is used to predict one variable from another, but it is sometimes the case that two different measurement systems are to be compared on the same samples. In that case both X and Y variables contain measurement errors, (whereas the normal regression functions assume no error in X), and therefore more specialised techniques must be used in this instance. These are outlined in detail in IWTO-0.

6.2 The effects of sources of variation on measurement techniques

Variances are additive

As we have already seen, errors can be divided into two categories – systematic (i.e. bias) and random. Test methods are designed to try and minimise the causes of systematic errors, by using calibration methods that attempt to standardise individual measurement systems. However, random errors will always be present for sampling and measurement of such a widely variable material such as wool. Much of what follows relates to random errors.

Each step in the process of sampling and measurement has associated random error. Each source of error can be described by its associated variance. In some cases, of course, the errors are relatively small – nowadays, most electronic instruments used for measurements in the laboratory are very precise and often their contribution to overall error is almost negligible.

Whilst it may be tempting to consider that random errors might balance out – sometimes positive, sometimes negative, and indeed this may be true on any individual measurement, and we expect it to be so if a large enough number of measurements are averaged, in reality on other occasions they may also be additive, and thus when we consider the effects of all the possible error contributions, we would end up with a wide range of possible values for the result. How do we quantify this? The normal method is to think in terms of variances, the fundamental measure of variability.

The first point to understand is that variances are additive as long as they are independent. Variances cannot be negative since they are squares of some quantity. Each step along the process therefore contributes some variance which must be added into the total. We call these the components of variance.

Components of variance

The wool metrology literature is full of references to components of variance. One may easily get the impression that these are fundamental physical attributes. Generally, nothing may be further from the truth. They are the results of modelling the sampling and measurement process. The modelling may be as simple or as complex as the original developers or standard-writers thought appropriate.

The simplest way of breaking down total variance is to consider two components only: within-laboratory and between-laboratory. Many standards now only quote these two components, and indeed there are good arguments for keeping it this simple. If a standard is prescriptive enough about how the samples are to be taken, blended, sub-sampled, processed, and specimens selected and measured, then why would one be interested in knowing any more than this?

The counter to this view is that standards are usually evolving. If continued development work is to be carried out in order to simplify, improve or even substitute some step in the process, it greatly helps if the components of variance can be broken down into constituent contributions from each step in the process, so that any change to an individual step can be objectively referenced to the variability in the original method.

IWTO and AS/NZS test methods are generally helpful in this regard. If the original variance model is not written into the standard (e.g. staple length and strength sampling and sub-sampling variance is described in IWTO-7), then usually there is a reference to the original work from which the information can be extracted. This is not always the case with other standards.

So, how does one develop a variance model to describe a particular system? This usually starts with a logical breakdown of the sampling, sub-sampling and measurement method, followed by some replicated measurements on a suitably wide range of wools to estimate each of the components of variance. An example of such an analysis carried out on airflow measurement is reported in Baxter and Houghton (1990).

Components of variance for some wool measurements are discussed in Chapter 15 of the Australian Sheep and Wool Handbook (see reading Cottle 2000.pdf).

Combining components of variances

Whilst it has already been pointed out that variances are additive, it is not necessarily the case that the total variance is simply the sum of each individual component of variance. If a process is exactly replicated, then the variance of the average value from that replication is reduced, assuming that the measurements are independent (i.e. the result of one measurement does not influence the result of the next measurement). Thus if we carry out two measurements on one specimen, and the variance associated with each measurement is s^2, the variance associated with the mean value from the two measurements is $s^2/2$. This very important rule has often been the key to improving the precision of test methods.

During the course of developing a test method, and during the analyses of variance, it may become quite clear that one step in the process is the largest contributor to the total variance. The most logical step, if that particular process cannot easily be improved, is to invoke replication at that step. Thus we see, for example, that whilst the largest contributor to the imprecision of staple sampling is the variance between staples in a grab (190 mm^2 for fleece – equivalent to a standard deviation of staple length of nearly 14mm on one staple), this can be reduced to manageable proportions if 60 staples are sub-sampled and the average result used (effectively reducing this component down to 3.2 mm^2 for fleece – equivalent to an SD of 1.8 mm on the average of 60 staples).

This process is codified in appendix D of IWTO-0.

Development of a variance model

There is a wide variety of variance models used in existing IWTO test methods, but from the perspective of providing examples of this process, many of these have been developed over a number of years, and tracking down each of the steps in the development of the variance models is not always straightforward. I have therefore chosen to use a more recent example in which all the necessary information for the model is contained in one document.

The OFDA2000 is an instrument developed for measuring the mean fibre diameter and diameter-length profile on greasy staples usually drawn from a fleece sample or directly from the animal. The establishment of a variance model for this sampling and measurement process presents some interesting situations, not least because the measurements are carried out on the greasy fibres. The reading material (Baxter 2001.pdf) illustrates the logic and processes used to develop two variance models. These are not definitive, but were considered useful at the time to examine methods of improving the precision of the measurement process.

It has subsequently been confirmed that the estimated precision from one of these models (individual fleece model) was reasonably close to the actual precision observed in comprehensive field trials funded and organised by Australian Wool Innovation. It is instructive to review another variance model that was developed to assess this technology in the field and to compare it with alternative measurement methods (Marler & Baxter 2004). It becomes clear that there is no unique variance model that completely describes even one technology, let alone one measurement system.

Precision versus cost

It should be clear by now that it is often possible to develop a test method to a stage that the required precision can be achieved. However, during that process, a number of alternative sampling, sub-sampling or measurement replication scenarios might have needed to be examined. Clearly, the more replication is required in order to achieve satisfactory precision, the greater the cost involved in carrying out the test. This always needs to be borne in mind.

An example that relates to this issue is the examination of farm lots for dark and medullated fibres. The original method for this test, developed by AWTA Ltd and SARDI, was very labour-intensive and required two operators to each examine 20 small fibre specimens of approximately 0.5g in mass, requiring in total 272 minutes of operator time per sample, at a cost of A$150. Further research by AWTA Ltd and CSIRO produced a much improved sample presentation system so that only 4 specimens each of 5g in mass needed to be prepared and measured. The cost was reduced to $A 39.70 (AWTA Ltd., CSIRO, AWI 2004).

6.3 The development of standard test methods

It should now be clear that the development of a standard test method, certainly in wool metrology, depends intimately on understanding the sources of variance in the sampling and measurement processes. A systematic approach would tend to follow this path:

- Identify the need, in relatively unambiguous terms. A test method has to satisfy a need – why are we measuring this property? How will the results be used? What are the commercial implications? What is the required precision? How available is the method to be (in-house, national, international)? Is the technology available or under development? Is there a reference measurement against which the results will be compared? If so, what are the commercial tolerable differences in terms of bias and precision?
- How is the sampling to be carried out? Is specialized equipment needed? What are the components of variance relevant to this?
- If sub-sampling is required, how is this to be achieved? Is specialized equipment needed? What are the components of variance relevant to this?
- Is any sample processing required (e.g. scouring, drying, carding)? What biases might be encountered and why? How is this to be incorporated into the calibration system? Are there any variables that may give rise to components of variance that may need to be separated out?
- What equipment is to be used for the measurement? What biases might be encountered and why? How is this to be incorporated into the calibration system? Are there any variables that may give rise to components of variance that may need to be separated out? What would be the effects on the method of this equipment becoming unobtainable?
- What are the costs associated with achieving the required precision? Can these be reduced?

These are just some of the questions that might be asked along the way. Normally this process takes a reasonable length of time to evolve, since many questions just won't have answers in the beginning. Realistically, most major wool test methods take at least 2 to 3 years to get to an acceptable stage where round trialling can begin. By that stage many of the questions will have been answered, but probably not all. One hopes that most of the variables have been identified, and that the robustness of the system has been tested. ASTM D4853 provides a good guide to this process.

By this stage a draft test method should have evolved. Round trialling to establish performance characteristics cannot begin without a common set of instructions, and whilst this may not be in the exact form of the final test method, it is good practice to adopt this format as early as possible, since it allows procedural difficulties to be highlighted. There are a number of guides to the preparation of test methods – both ASTM and ISO publish guidelines, but the reality is that most wool test methods will end up having to satisfy members of IWTO, and hence IWTO-0 is the most appropriate and detailed guide and is therefore included in the reading material (IWTO-0 App B.pdf, IWTO-0 App D.pdf.)

It should be clear that the final precision of the method cannot be established until round trials have been completed involving a number of laboratories, preferably in different countries. It is not uncommonly the case that these may identify differences between laboratories that need to be reduced in order to achieve a commercially-satisfactory level of precision. Once that tuning has been achieved, generally a further round trial is necessary before the method can be accepted internationally. It's a hard road!

Readings [3]

The student is strongly recommended to obtain a standard textbook on basic statistics. There are many to choose from, and some include CD-ROMs with example datasets, links to websites and free statistical software. Chapters on describing data, the normal probability distribution, sampling methods, hypothesis testing, linear regression and correlation usually adequately cover fundamental concepts that are necessary to an understanding of the basic statistics of measurement principles. A typical example is:

Basic Statistics for Business and Economics, D A Lind, R D Mason and W G Marchal, Publ. Irwin McGraw-Hill, 3rd ed 2000, ISBN 0-07-366062-0

Readings with * are available on CD:

1. *Mandel, J. 1984, The Precision and Accuracy of Measurements, in the Statistical Analysis of Experimental Data, Dover Publications Inc, Mineola, N.Y., U.S.A. chapter 6, part 1, pp 102-117 and part 2, pp 118-130. Mandel, of the National Bureau of Standards in Washington, USA, wrote one of the clearest texts on the statistical analysis of experimental data in 1964. Chapter 6 deals explicitly with the measurement and expression of experimental data and covers much of the basic statistical concepts required in this course.
2. Teasdale, D. 2000, Wool Preparation, Marketing and Processing, in Australian Sheep and Wool Handbook, ed D.J. Cottle 4th Edition, WRONZ Developments, Christchurch, New Zealand, pp 309-348. There are few texts which deal specifically with the Australasian wool industry, and this one provides a good overview of the main aspects. Chapter 15 puts much of the content of this course within the context of basic wool metrology, and quotes specific examples where components of variance are discussed in terms of the practical consequences of applying these concepts to the more common tests.
3. *Baxter, P., 2001, Precision of measurement of diameter and diameter-length profile, of greasy wool staples on-farm, using the OFDA2000 instrument, Wool Technology and Sheep Breeding, vol 49 (1), pp. 42-52. Reprinted from the 10th International Wool Textile Conference, Germany, Nov, 2000. This paper is used as a relatively recent example where analysis of variance has been used to explore several components of variance that were believed to impact on the precision of measurement of diameter of greasy wool samples. It is useful to compare the approach used in this paper with that used in Reading 4.
4. *Marler, J.W. and Baxter, P. 2004. The 2003 Australian Wool Innovation On-farm fibre measurement instrument evaluation trial. Part 1: Accuracy and Precision Trials, IWTO CTF 01 May 2004, Evian. This paper is provided as a direct contrast to Reading 3. Both cover the use of components of variance to assess the precision of measurement of diameter of individual animals. The specific focus of this paper was to compare measurement systems, and to assess how they may be used to assess the precision of measurement of the diameter of a whole fleece rather than simply the precision of measurement of a fleece sample, which hitherto had been the prime focus of the publications on this topic.
5. *IWTO, IWTO-0 (2003), Procedures for the development, review, progression or relegation of IWTO test methods and draft test methods Appendix B, Presentation of supporting technical data. Appendix D, Statistical methods IWTO-0 is one of the few standards publications that specifically deal with the development of test methods and the testing of equivalence between test methods. It was developed specifically for the wool industry and is required reading for anyone working on test methods or changes to test methods that might eventually load to international standardisation.

Activities

@ Available on WebCT

Multi-Choice Questions

@ Submit answers via WebCT

Useful Web Links

@ Available on WebCT

Assignment Questions


Choose ONE question from ONE of the topics as your assignment. Short answer questions appear on WebCT. Submit your answer via WebCT

Summary

Summary Slides are available on CD
This topic covers the essential statistical concepts necessary to understand the basic principles of measurement, together with a brief overview on the effects of sources of variance on measurement techniques. It introduces the general subject of components of variance and how these may be used to build up a variance model of the sampling and measurement process. Finally there is an overview of how a test method becomes reality, from genesis in a need, to publication as an international standard.

References

ASTM, ASTM D4853 (1991). Standard guide for reducing test variability

AWTA Ltd, CSIRO, and AWI (2004). Development of an improved test for the detection of dark and medullated fibres in presale core samples. IWTO RWG 5, Shanghai

Baxter, B. P. and M. Houghton (1990). Components of variance in airflow testing of raw wool - hierarchical blending effects. IWTO, Cavtat, Yugoslavia

Baxter, B. P. and D. J. Cottle (1998). The use of midside fleece fibre diameter distribution measurements in sheep selection. Wool Technology and Sheep Breeding 46(2): 154-171

Baxter, P. (1999). The effects of calibration errors on between-laboratory and between-instrument diameter comparisons. IWTO, Nice

Baxter, P. (2001). Precision of measurement of diameter and diameter-length profile, of greasy wool staples on-farm, using the OFDA2000 instrument. Wool Technology and Sheep Breeding 49 (1): 42-52

Dallal, G.E., (undated). The little handbook of statistical practice. Statistics online tutorials available from http://www.StatisticalPractice.com

ISO, (1993). Guide to the expression of uncertainty in measurement. ISBN 92-67-10188-9

IWTO, IWTO-0 (2003), Procedures for the development, review, progression or relegation of IWTO test methods and draft test methods,

Appendix B, Presentation of supporting technical data

Appendix D, Statistical methods

IWTO, IWTO-7 (2000), Sub-sampling staples from grab samples

IWTO, IWTO-12 (2003), Measurement of the mean and distribution of fibre diameter using the Sirolan-Laserscan Fibre Diameter Analyser

Marler, J.W. & Baxter, P. (2004) The 2003 Australian Wool Innovation On-farm fibre measurement instrument evaluation trial. Part 1: Accuracy and Precision trials, IWTO CTF 01, Evian

Stubbs, C.G. and Marler, J.W. (1994). Analysis of Interwoollabs airflow round trials on wool tops, Report 12, IWTO New Delhi.

Glossary of terms

Accuracy	A measure of the closeness of a test result to the true value
Bias	A systematic difference between test results and their corresponding true values
Coefficient of variation	A measure of variability exhibited within a group of values. It expresses the standard deviation as a percentage of the mean.
Component of variance	In a variance model, one of the discrete portions of the total variance associated with a specific sampling or measurement aspect.

Confidence interval	The absolute range within which the true result is expected to lie within the stated probability. The 95% confidence interval is equal to the mean minus 95% CL to mean plus 95%CL
Confidence limits	An expression of the precision of a test result or the mean of a group of results. It is usually associated with a stated probability, normally 95%. The 95% confidence limits are the range of values within which the true value is expected to occur 95 times out of 100.
Correlation coefficient	A measure of the degree of association between two variables. It is normally calculated as the product-moment correlation coefficient, and varies between -1 and +1, with -1 being a perfect correlation with a negative slope, 0 being no correlation, and +1 being a perfect correlation with appositive slope.
Error	Error is the difference between an individual measurement results and the true value. Errors may be divided into systematic errors (which give rise to bias), and random errors (which are the main contributors to imprecision)
Mean	Arithmetic average. The mean is calculated by summing the individual measurements and dividing by the number of measurements
Normal distribution curve	A continuous curve which is symmetrical about the mean and for which the height is a function of the mean and standard deviation only. The mean ± 1, 2 and 3 standard deviations contains 68.27%, 95.45% and 99.73% of the observations respectively.
Precision	An indicator of the repeatability of measurement. It is often expressed in terms of the confidence limits.
Regression	A series of techniques for establishing mathematical relationships between one variable and another
Sample	In the case of wool, the portion drawn by appropriate methods from a lot, consignment or delivery
Standard deviation	A measure of dispersion of individual results. Standard deviation is expressed in the units of measurement
Standard error	A measure of the uncertainty in a mean value. It is equal to the standard deviation of the individual measurements divided by the square root of the number of measurements
Sub-sample	A randomly-drawn portion, representative of the sample, used for a specific test measurement
True value	The absolute value of a characteristic for a bulk of material is almost always unknown. Measurements of the characteristic are, in the absence of bias, normally distributed about the true value with a variance this is also unknown in a particular case. The mean of a set of un-biased measurements is the best estimate of the true value.
Variance	The variance is a measure of the dispersion of values about a mean. It is calculated from the sum of the squares of the deviations from the mean and is expressed in units of measurement squared
Variance model	A mathematical model which expresses the total variance of measurement system or part of a measurement system in terms of the components of variance of identifiable contributors of error