Should alpaca breeders use ‘SD’ or ‘CV’ when evaluating fibre traits?

Paul Vallely, B Soc Sc. AAFT-Europe.

Over the years I have operated AAFT, the question of whether to use Standard Deviation (SD) or Co-efficient of variation (CV) when evaluating fibre traits is undoubtedly one of the most commonly asked questions. It also happens to be one of the issues most plagued by misunderstanding, and consequently, carries the potential to de-rail breeding strategies, particularly those aimed at reducing the incidence of coarse fibres, increasing the level of fibre uniformity, improving the processing performance of fleeces or simply improving the style and handle of fleeces.

In writing this paper, I principally draw upon my professional experience in managing alpaca fibre measurement laboratories over the past 17 years, over 25 years practical experience with managing breeding programs aimed at genetic improvement for fleece traits (alpacas and superfine merino sheep) and having endured two years of statistics at New England University in Australia.

At this point, I should also acknowledge the contrary view put forward by Mr Cameron Holt in a paper recently published in the British Alpaca Society magazine titled ‘SD or CV – What’s it to be’ (Alpaca Magazine, British Alpaca Society. Spring/Summer 2016 ed’s). In his paper, Mr Holt appears to suggest CV is the preferred statistic to be used in fleece trait analysis. My paper should therefore be seen as offering an informed, yet alternate point of view in the spirit of professional debate, the outcome from which should surely be of potential benefit to alpaca breeders.

One of the most useful aspects of fibre testing, is the ability to measure the degree of variation in fibre diameter within a given fleece sample. Variation in fibre diameter is correlated with processing performance of fleeces, fibre alignment and handle of fleeces, micron blow-out, tensile strength and incidence of coarse fibres. Further, a major benefit in using fibre diameter variation is its high level of heritability, meaning breeders are able to achieve significant genetic gains when using this trait as a breeding objective.

World record priced ultrafine bale (AU$10,000) used low SD as selection criteria to ensure superior levels of processing performance

For most alpacas, we find a range of about 25 micron difference between the finest fibres within a sample and the coarsest fibres within a sample. For instance, a test of a midside alpaca fibre sample may reveal an overall fibre diameter average of 25 microns, with a range in average diameter for individual fibres from 15 microns for the finest fibres to 35 microns for the coarsest fibres (although the range normally skews towards the coarse edge).

I might add that we have seen superior alpacas with histograms revealing a range of only 12 microns while we have seen alpacas with a range of over 40 microns between individual fibres within the one midside sample.

I might also add that this range in fibre diameter is repeatable over, say, the fleece's saddle area – but that's another discussion.

The two statistics we refer to when measuring variation of fibre diameter are Standard Deviation (SD) and Co-efficient of variation (CV). In saying this, however, it might also be noted that fibre test histograms provide a graphical representation of fibre diameter variation.

To consider whether to use SD or CV for the purposes of selecting breeding stock, it would be useful to manually calculate SD and CV for two imaginary samples of fibres from two different alpacas.

While we obviously use software programs to calculate these statistics, manually calculating SD and CV will reveal the true relationship these two statistics have with the actual degree of fibre diameter variation found in any fibre sample. For ease of calculation, these imaginary fibre samples will have a ridiculously small number of fibres.

Lets imagine the first sample has 5 fibres, each individual fibre with the following average diameter in microns: 18, 19, 19, 20 & 21. The AFD of this sample is therefore 19.4 microns.

Now starts the heavy mathematics lecture. To calculate SD, we use the following formula (I'll use the above example to illustrate the calculations):

1/ obtain the total of the squares for each of the data values (multiply each micron by itself then add it all up) eg 324 (18 x 18) + 361 + 361 + 400 + 441 = 1887.

2/ square the sum of the data values and divide by the number of values (add up the microns then multiply the total by that same number, then divide by the number of fibres) eg 18 + 19 + 19 + 20 + 21 = 97, thence 97 x 97 divided by 5 = 1881.8

3/ subtract 2/ from 1/, then divide the answer by the number of values less 1 (subtract the answer for 2 from the answer to 1/, then divide by the number of fibres less one) eg 1887 - 1881.8 = 5.2, thence 5.2 divided by 4 = 1.3

4/ obtain the square root of 3/ eg, the square root of 1.3 = 1.14

The SD of the sample is therefore 1.14. I might add that our fibre testing software makes these calculations for over 3000 measurements in well under a second.

Now take a another sample of fibres with exactly the same degree of variation (distribution of AFD for each fibre from the overall average) Lets say the microns of the five fibres are 23, 24, 24, 25 & 26. (AFD of 24.4)

The respective calculations for SD of this second sample are:
1/ 2982
2/ 2976.8
3/ 1.3
4/ <u>1.14</u>

The SD is also 1.14. The SD is the same because they both have precisely the same degree of variation in fibre diameter.

If, on the other hand, we take a sample with a higher degree of variation in the diameter of the fibres, the SD will be higher. For example, our next fibre sample has fibres with diameters of 23, 24, 24, 25 & 29 microns. (AFD of 25.0 microns)

The respective calculations are:
1/ 3147
2/ 3125
3/ 5.5
4/ <u>2.3</u>

The SD is 2.3.

At this point, it should be clear that SD is the true and unbiased indicator of variation – and let me assure you it is.

By the way, one standard deviation will effectively represent about 66% of the total variation in a sample. For instance, if the average diameter is 20 microns and the SD is 2 microns, then 66% of the total variation in the sample will lie between 18 and 22 microns, generally speaking.

This then brings us to CV.

CV is simply what percentage the SD is of the overall average. Therefore, we calculate CV by dividing the SD by the AFD and then multiply by 100. A simple example being that if the average fibre diameter of a sample is 20 microns, and the SD is 2 microns, then the CV is 10% as 2 is 10% of 20.

Using the formula for calculating CV, we find the first sample above has a CV of 5.9%, (1.14 divided by19.4 x 100) while the second sample above has a lower CV of 4.7% (1.14 divided by 24.4 x 100) even though there is exactly the same degree of variation. This is where the problem lies.

Because of the way we calculate CV, it means the higher the average fibre diameter, the lower the CV. Or put another way, a low CV may be due to a high fibre diameter rather than a low degree of variation.

At this point, I draw the reader's attention to the divergence of opinion presented in this paper and the opinion expressed in Mr Holt's paper regarding the formula for calculating SD. Mr Holt states the formula for SD as being 'CV x micron divided by 1'. (P10) As shown in this paper, however, CV can only be calculated once the relevant sample's SD has been determined. It is therefore suggested Mr Holt's formula appears implausible as it relies on quite the opposite in that he claims SD can only be calculated once CV is determined.

‘Practising what is preached’ – A fleece from the author’s own superfine merino flock. The ewes consistently produce fleeces such as this one due to a breeding program that prioritises SD. This fleece had AFD of 15.4 microns with SD of 2.2 microns.

I should also add that Mr Holt’s formula is in fact quite at odds with the standard formula for SD (‘Statistics Today – A Comprehensive Introduction’ Byrkit, D. University of West Florida. 1987 P69).

So why do we calculate CV?

Co-efficient of variation is a commonly used statistic that enables us to compare the degree of variation between two different types of data. To illustrate, lets take the example of a foreign exchange dealer who might want to compare the movement in the US Dollar against the movement of the Euro over a period of time. In this case, a one US dollar variation is very different to a one euro variation, so we use CV to bring the variation in the two currencies to a comparable statistic. It can also be used where we need to directly relate the degree of variation to the average, for example, estimating the likely variation on the return on investment for purchasing two investments of quite different purchase amounts. In these cases, CV effectively brings everything to a level playing field.

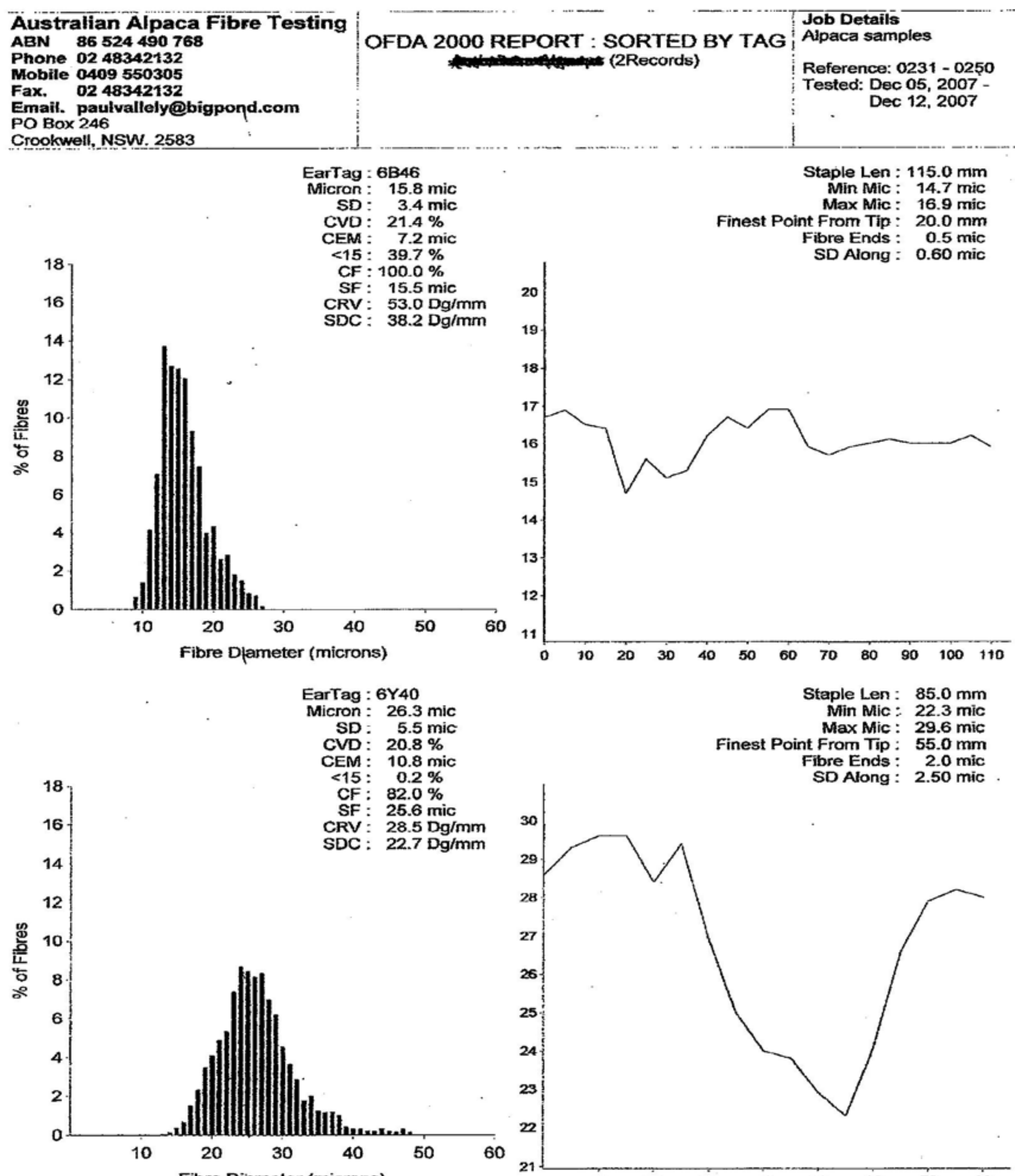
Comparing the degree of variation in fibre diameter for two different alpacas is not a case of two different data sets, nor is it impacted by a direct relationship with the average diameter. When comparing the fleece traits of two alpacas with regard to variation, we simply want to compare how much variation is evident – that is, what is the range in diameter of the fibres between the two alpacas. To expect anything more is beyond the purpose of SD or CV.

Again, I draw the reader’s attention to a divergence of opinion presented in this paper and the paper presented by Mr Holt in that he states “Comparing say a 15 micron (alpaca) with a 20, a 25 and a 30, the standard deviation would not be a reliable method for comparison” (p10).

As mentioned above, a variation in the diameter of individual fibres within a sample of, say, 40 microns cannot be viewed differently if considering such a variation on a ’25 micron’ alpaca compared to that of a ’20 micron’ alpaca. To accept the principle put forward by Mr Holt would immerse the breeder in an unnecessarily complicated method of genetic or commercial appraisal of alpacas.

At this point, I might mention that fleeces with low average fibre diameter have a propensity to also exhibit low degrees of variation in fibre diameter. I should also mention this has nothing to do with statistical calculations, but again, this is another discussion.

We shall now put the mathematics lecture behind us and look at two actual fibre tests using OFDA2000 technology on midside samples taken from two alpacas.



64B6 has an average fibre diameter of just 15.8 microns. It has an SD of 3.4 microns indicating very low variation with fibre diameter within the sample. As can be seen with the narrow histogram, this very low degree of variation is clearly evident. The CV, however, is 21.4% as 3.4 is 21.4% of 15.8.

This alpaca is considered likely to be genetically superior for breeding towards high quality, uniform and soft handling fleeces, although obviously, subjective appraisal of the alpaca would also need to be considered before making such a judgment. If using CV, this alpaca might not even appear on your radar.

Moving our attention now to alpaca 6Y40, we find it has much higher average fibre diameter at 26.3 microns. Further, we find much higher degree of variation in fibre diameter with an SD of 5.5 microns and a histogram to match. Even to the untrained observer, this alpaca has a far greater degree of variation than the previous alpaca, yet the CV is lower at 20.8%.

The reason the CV is lower has absolutely nothing to do with the degree of variation, but because the first alpaca has a much lower average fibre diameter.

Based on its fleece traits, this alpaca runs the risk of regressing a herd's genetics due to its high variation in fibre diameter and evidence of very coarse fibres. If a breeder relied on CV, this is an example of where they would make a serious mistake.

Moving away from these examples, but referring back to Mr Holt's paper, he cites the wool industry's use of CV when cataloguing sale consignments as a reason for the alpaca community to adopt CV as a selection trait.

Over the past few years, I have combined my professional roles as wool producer and owner of fibre testing laboratories to provide consultancy to the superfine wool industry.

One project was to consider changing the use of CV in sale catalogues to the use of SD. While fibre metrology technicians accepted my position on the use of SD, it was considered that given wool consignments within a sale catalogue were compared with other sale lots of similar fibre diameter, the use of CV was not problematic. While we accepted this position, it clearly does not present itself as evidence in support of the use for CV with regard to animal breeding and husbandry.

In conclusion, and after having studied Mr Holt's paper, I remain firmly of the opinion that using CV may result in the infusion of inferior genetics into a breeder's herd, the effects of which may take many generations to effectively remove.

The message is abundantly clear – SD is the preferred statistic when evaluating fibre traits.

Aug 2016.