Measuring Measuring Errors

Backgound

An important part of statistical analysis, perhaps the most important part, is the identification and quantification of sources of variation. For example, in a *t*-test the variation between groups is compared with variation within the groups to see if the former can be explained by the latter alone. The main tool for measuring variation is the standard deviation (SD), although it is sometimes useful to consider the *variance*, which is simply the square of the SD.

Given a set of numbers it is straightforward to compute their SD using the **Descriptive Statistics** command in the **Basic Statistics** item under the **Stat** menu in Minitab, as in figure 1. The six numbers entered there are plasma glucose levels in mmol/l. The "STDEV" of 0.0845 mmol/l measures variation, but variation of what?

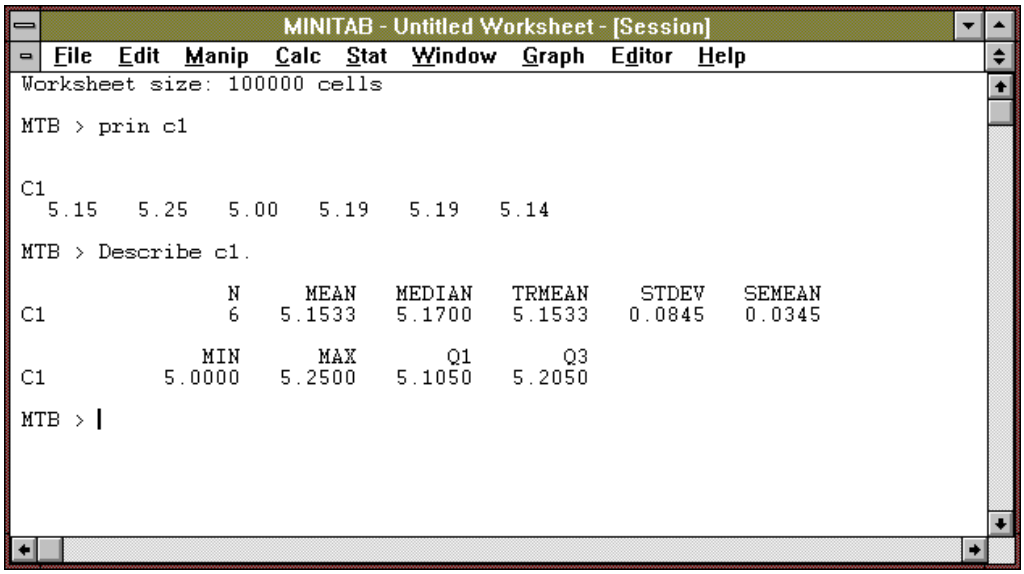

Figure 1: computing an SD in Minitab

The answer is that without further information about how the six numbers were collected a meaningful answer is not possible. If the numbers are from six different patients then the SD will measure the variation between the *observed* plasma glucose levels in those patients. However, these numbers were, in fact, obtained by dividing a single sample from a patient into six aliquots. Thus, each sample contains the same underlying plasma glucose level (assuming the sample was well mixed before it was divided), so the measured variation cannot represent variation in plasma glucose level itself. In fact it measures the error in the assay used to obtain the measurements.

Put more formally, if the true plasma glucose in the sample is X, then the observed values can be written as:

$$X + error_1,\ X + error_2,\ X + error_3,\ X + error_4,\ X + error_5,\ X + error_6.$$

It is clear that the spread of these values does not depend on X and that the SD is simply the SD of the errors. The errors will be random quantities and it is often reasonable to assume these have a Normal distribution as in figure 2. The SD of this

distribution is the SD of the errors, which will be referred to as the *error SD*, its mean will be zero, unless then measuring instrument contains a systematic bias. From the properties of a Normal distribution it follows that for 95% of measurements the

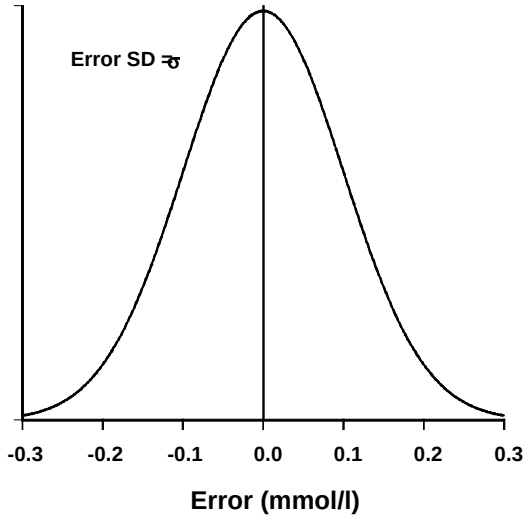

Figure 2: distribution of measurement errors

observed value will differ from the true value by less than 2σ.

Variation between individuals and the value of replication

The SD in figure 1 measured the measuring error only because each number was obtained by different measurements of the same sample, that is they differed from one another only because of imprecision in the measurement process. If the six measurements had been on different patients, then the above representation would have had to be rewritten as

$$X_1 + \text{error}_1,\ X_2 + \text{error}_2,\ X_3 + \text{error}_3,\ X_4 + \text{error}_4,\ X_5 + \text{error}_5,\ X_6 + \text{error}_6.$$

where there are now six underlying (i.e. unobserved) true plasma glucose levels, X_1, X_2, X_3, X_4, X_5, X_6: the SD of the observed values will now measure two *components of variation*, namely the variation in the true plasma glucose levels (the Xs), in addition to the measurement error SD, which is an inescapable component of any measured quantity. If the SD of the true glucose level (the Xs) is written as σ_X, say, then the *variance* of the above observations is:

$$\sigma_X^2 + \sigma^2,$$

as shown in figure 3. It is clear that if we only have one measurement on each patient then we can only estimate $\sigma_X^2 + \sigma^2$, we cannot estimate σ and σ_X individually (because the same total variability, which is all this type of sample allows us to see, would be obtained if, e.g. $\sigma_X = 1$ and $\sigma = 2$, or if $\sigma_X = 2$ and $\sigma = 1$).

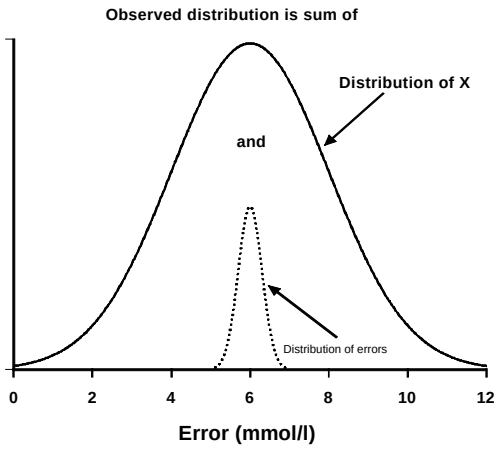

Figure 3: superposition of biological and error variation

Why should this be a problem? It is true that in many circumstances the relevant measure of SD is $\sqrt{(\sigma_X^2 + \sigma^2)}$, which can be obtained from the usual sort of sample. However, there are many circumstances in which it is very important to know the precision of your measurements. One example concerning the use of height velocity is given in the example sheet. In the construction of a new assay, it will be important to know its error characteristics. Sometimes it is necessary to use an assay with a rather high error SD; the influence of the measurement errors can be reduced if the value quoted from the assay is the mean of the measurements on several *replicates*. How many replicates it is sensible to choose will depend on several factors, including a knowledge of the relative sizes of σ_X and σ, as will now be discussed.

The value of replication

One of the reasons for taking the mean of a group of independent observations is that the mean is less variable than an individual observation: if the SD of an observation is σ, then the SD of the mean of n of these observations (known as the standard error) is $\sigma/\sqrt{n}$: so the mean of four independent observations is $\frac{1}{2}\sigma$, of 25 observations is $\frac{1}{5}\sigma$ and so on.

If we return to the original situation, where there were six observations on a single sample, i.e.

$$X + error_1,\ X + error_2,\ X + error_3,\ X + error_4,\ X + error_5,\ X + error_6.$$

then each reading has an SD of $\sqrt{\sigma_X^2 + \sigma^2}$ but their mean does not have an SD of $\sqrt{\sigma_X^2 + \sigma^2}/\sqrt{6}$ because these are not independent observations, the *same* X contributes to each one. In fact the mean of these readings is

$$X + \tfrac{1}{6}(error_1 + error_2 + \cdots + error_6)$$

i.e. the "X part" stays the same because the same value of X appears in each measurement. As such the *variance* (square of the SD) contains the same component for the variation in X but the error variance is reduced because the error on the mean is

the mean of the errors, so *does* have an SD of $\sigma/\sqrt{6}$, giving a variance for the mean of the six readings of:

$$\sigma_X^2 + \tfrac{1}{6}\sigma^2.$$

In general, if a sample is measured n times, the SD of the mean will be $\sqrt{\sigma_X^2 + \tfrac{1}{n}\sigma^2}$.

Although some of the details behind this formula may still be a little opaque (e.g. why it is the *variances* and not the SDs which add), the general implications are sensible. If a sample is remeasured many times and the mean of the replicate measurements is taken, then the SD of the resulting mean gets smaller as the number of replicates gets bigger, essentially because the error variance is effectively reduced. However, the SD of the mean does not get closer and closer to zero, it gets closer and closer to σ_X. This is reasonable because while repeatedly measuring the same sample might be expected to reduce the measuring error, it can do nothing about the underlying biological variation.

> *[Note: there are two caveats which should be entered here. First, it is assumed that if the replicate measurements are taken by dividing a blood sample or a biopsy specimen, then there is a sufficient supply that repeatedly dividing the sample does not produce an amount that is so small that the precision of the assay is affected. Second, it is assumed that the replicate measurements are unaffected by one another; violation of this assumption can have a marked effect, as demonstrated by Voss and colleagues in a study of the measurement of heights (Arch. Dis. Child., 1990, **65**, 1340-44)]*

So, replicating a measurement and taking a mean will effectively reduce the measurement error, but the biological component of the variation is unaffected. As such it may well be sensible to perform a few replications but once the contribution of the error SD to the total is small, it is pointless to take more replicates. How many need to be performed depends on how much of the total SD $\sqrt{(\sigma_X^2 + \sigma^2)}$ is due to biological variation and how much the measurement error. As an example, the effect of replication in two hypothetical cases is shown in table 1: the first case is when σ_X = 5 and σ = 1, a relatively precise measurement, and case 2 is a much noisier one, where σ_X = 5 and σ = 4.

	Total SD	
n	Case 1	Case 2
1	5.09902	6.40312
2	5.04975	5.74456
3	5.03322	5.50757
4	5.02494	5.38516
5	5.01996	5.31037
10	5.00999	5.15752
15	5.00666	5.10555
20	5.00500	5.07937
25	5.00400	5.06360

Table 1: illustration of effect of replication

It can be seen that in case 1 two replicates reduces the total SD by only 1%, so in this case the value of any replication is questionable. In the second case, 2 replicates effect a reduction of 10%, whereas 3 replicates give a 14% reduction. Increasing the replication to 4 and 5 gives reductions of 16% and 17% respectively. However the reduction achieved by replicating 25 times is only 21%, showing that beyond a certain point, extra replication has limited benefits.

In any application the number of replications that are required is a matter of judgment, but calculations such as these provide a basis for striking the balance between increased precion and the extra effort and cost of greater replication.

Estimation of σ and σ_X

So far this note has been concerned with establishing a framework to clarify the issues involved with measurement error and replication. For many purposes it is useful to have an estimate of the error SD σ and in some circumstances (such as assessing the value of replication) it is also helpful to have an estimate of σ_X. We now turn to the methods for obtaining these estimates.

At the start of these notes two forms of data were considered. In the first a single patient or subject was measured repeatedly, so although an estimate of the error SD, σ, could be obtained, no information on between patient variability had been collected, so no estimate of σ_X is available. The second type of sample is when each patient is measured only once, so $\sigma_X^2 + \sigma^2$, but not σ_X and σ individually, can be found. The resolution to this is to collect several samples from each of several patients. The replication within a patient allows σ to be estimated and the replication between patients allows σ_X to be estimated. The balance between how many times to measure each patient and how many patients should be measured varies from case to case. Although it is possible (and indeed often desirable) to measure each patient three or more times, it is simpler to obtain the estimates if each patient is measured just twice and we will restrict consideration to this case.

In table 2 duplicate readings of plasma glucose on each of 20 patients are shown. In symbolic terms, the data from the first patient can be written

$$Y_{11} = X_1 + \text{error}_1,\ Y_{12} = X_1 + \text{error}_2$$

As outlined above, the average of these replicates, i.e. $\frac{1}{2}(Y_{11} + Y_{12})$, has SD $\sqrt{\sigma_X^2 + \frac{1}{2}\sigma^2}$.

The difference between the replicates, $Y_{11} - Y_{12}$, is actually error_1 - error_2, i.e. it does not depend on X, and hence on σ_X, at all. Thus the *variance* of this difference must depend only on σ^2 and it can be shown that it is actually $2\sigma^2$ (see Armitage and Berry, 1994, p.88).

Patient	Reading 1	Reading 2	Patient	Reading 1	Reading 2
1	8.11	7.93	11	5.72	5.78
2	8.42	8.41	12	5.87	5.88
3	4.05	4.25	13	7.47	7.47
4	6.40	6.49	14	6.69	6.62
5	9.13	8.89	15	7.28	7.20
6	8.36	7.91	16	6.79	6.73
7	5.44	5.75	17	9.54	9.73
8	4.65	4.74	18	3.35	3.28
9	6.57	6.50	19	8.01	7.88
10	3.74	3.95	20	5.14	5.23

Table 2: duplicate plasma glucose readings on 20 patients (mmol/l)

Thus the method is simply to form the 20 means and 20 difference of the pairs of glucose measurements in table 2, something which is easily done in the **Mathematical Expressions** part of the **Calc** menu. The SDs and variances of these can then be found in the **Descriptive Statistics** command in the **Basic Statistics** item under the

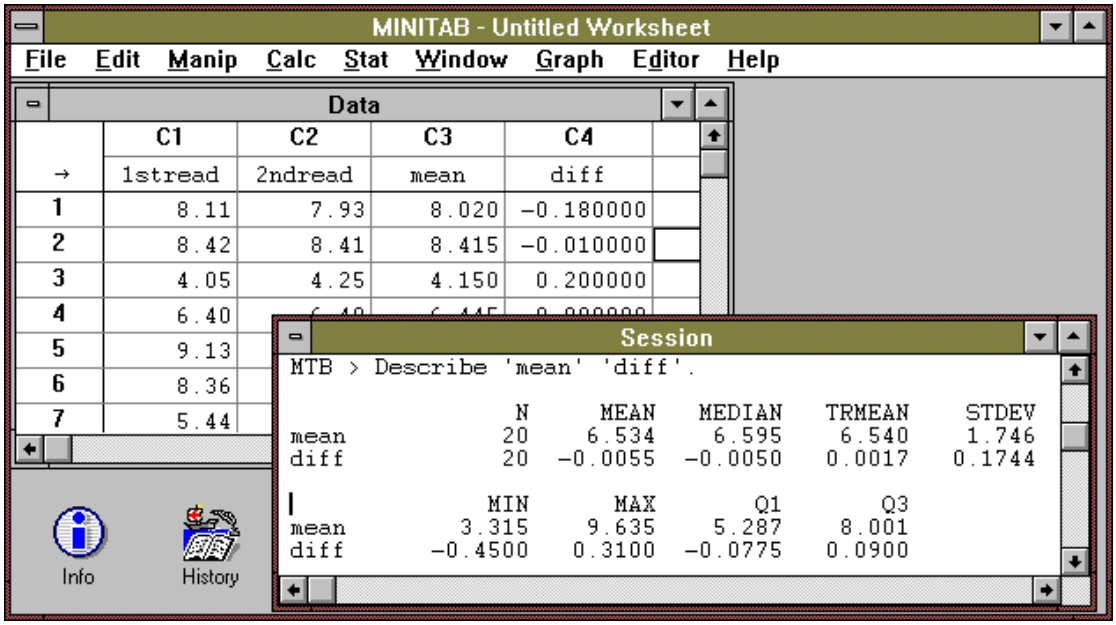

Figure 4: SDs of means and differences

Stat menu. The SD of the 20 differences and means is shown in figure 4

The method can be summarised in the following table, where the above formulae and the estimates are put together:

Quantity	SD	value	variance	value
Means	$\sqrt{\sigma_X^2 + \frac{1}{2}\sigma^2}$	1.746	$\sigma_X^2 + \frac{1}{2}\sigma^2$	3.049
Differences	$\sigma\sqrt{2}$	0.1744	$2\sigma^2$	0.0304

From this, it is easy to estimate σ, as $\sqrt{\frac{1}{2}0.0304} = \sqrt{0.0152}$ = 0.123 mmol/l, that is, measurements of plasma glucose are generally within ±0.246 mmol/l of the true value.

From the first row of the table it is seen that $\sigma_X^2 + \frac{1}{2}\sigma^2$ = 3.049, so σ_X^2 can be found by using the above estimate of σ^2 to find $\frac{1}{2}\sigma^2$ = 0.0076 and then subtract this from 3.049, giving 3.0414, so the estimate of $\sigma_X = \sqrt{3.0414}$ = 1.744 mmol/l.

In this instance the effect of measurement error on the total SD is very small. Clearly it is very unlikely that any application of these values would benefit from any replication.

More complicated situations

The case where two sources of variation, biological and measurement, are disentangled by taking duplicate readings is probably the simplest case for this type of analysis. If each patient had been measured three times then the same quantities, σ_X and σ, would be estimated but using more a more complicated method. This method is

essentially one-way ANOVA although the interpretation of the sums of squares changes from that covered in the note on that topic: details can be found in Armitage and Berry (1994, p.219-222). It is necessary to perform some calculations on the ordinary output of ANOVA to obtain the relevant estimates; if the ANOVA is performed using the **Balanced ANOVA** command in the **ANOVA** sub-menu of the **Stat** menu then options can be set to produce these estimates directly.

A further extension of the method, which is generally known as *random effects ANOVA* or *components of variance*, allows more than two SDs to be estimated: for example when measuring fetal kidney dimensions on ultrasound there will be inter-fetus variation and also measurement error within an ultrasonographer, but there may well be between-ultrasonographer variation too. This additional component of variance can be found by a suitably designed study, in which each fetus is measured repeatedly by each of the ultrasonographers in the study. The details of this are beyond the scope of the present note.

Reference

Armitage, P. and Berry, G. (1994) *Statistical Method in Medical Research* (3rd edn.). Blackwell, Oxford.