Standard error of the mean

When considering the standard deviation we were concerned with making some mathematical statement of the variability of individual observations about the mean of all observations in a particular sample. What we must now examine is the relationship that that or any sample bears to the universe from which it has been drawn.

Plainly, in the fields of clinical and epidemiological research it is very rarely feasible to acquire all possible observations of any one characteristic under study. We usually must be content with collecting a sample which is only one of perhaps hundreds of possible samples available from that universe. Consequently one thing that we require is some idea of the relationship that the mean of our sample bears to the true mean of the universe from which it has been drawn. Put another way, some statement of the error we incur by using our sample mean as representative of the true mean.

By parallel reasoning to that leading to the concept of the normal distribution of individual values described in the first part of this article, we can say, that in the biological context, the means of a large number of successive random samples of equal size from any given universe will tend to fall into a normal distribution about the true mean of that universe. Brief consideration will further render apparent the fact that the size of a sample will be related to the accuracy of its mean as a statement of the true mean.

In other words, the mean of a very large sample is likely to be much nearer to the true mean of the universe than that of a sample of, say, five observations which could quite easily by chance be exceptionally large or small and will give a greatly distorted sample mean relative to the true mean.

This rather lengthy explanation has been given to avoid much simple, though tedious, manipulation of groups of figures to demonstrate the point. In fact, the formula for the calculation of the Standard error of the mean is: 

\[ \text{SE} = \frac{\text{SD}}{\sqrt{n}} \]

(Strictly speaking the standard deviation used in this calculation should be that of the universe: as this is rarely available the SD of the sample is satisfactory providing \( n > 30 \). Where \( n < 30 \) another statistical test needs to be done. This is dealt with later).

**Example**

If we use the figures that have already been calculated for our sample population of RQMS’s in the previous article viz:

- \( n = 200 \)
- Sample mean \( \bar{x} = 42.73 \text{ inches} \)
- \( \text{SD} = 5.86 \text{ inches} \)

Substituting in \( \text{SE} = \frac{\text{SD}}{\sqrt{n}} \) We get: \( \text{SE} = \frac{5.86}{\sqrt{200}} = 0.41 \text{ inches} \)
From the calculated standard error we are able to make statements of probability in exactly the same way as has been explained for the standard deviation. Thus in this instance we may say that the true mean of the universe will lie within the range Sample mean $\bar{x} \pm 2 SE$ or, in our example $42.73 \pm 0.82$ inches, when $p = 0.05$.

Alternatively, introducing the term confidence limit, we may say that we are 95 per cent confident that the true mean lies within the limits of sample mean $\pm 2 SE$. It will readily be seen that this is merely another way of saying the same thing and as one may require a probability of $p = 0.01$ so one may substitute the 99 per cent confidence limit, and so on.

**Further calculations of the standard error**

By exactly similar reasoning to that seen above we are enabled to calculate standard errors in several situations, viz:

*Standard error of a proportion*

By proportion here we usually mean percentage. In any expression of percentage "positive" and "negative" proportions are implied. For example 15 per cent males implies 85 per cent females; 25 per cent red and 75 per cent blue implies 25 per cent red (positive) and 75 per cent not red (negative). These two proportions are referred to as p and q and our object in calculating the standard error of the proportion or $SE_p$ is to ascertain the error involved in assuming the sample proportion representative of the population or universe from which it has been drawn. The formula used is:

$$SE_p = \sqrt{\frac{pq}{n}}$$

*Example*

With a given treatment a sample of 100 patients with a certain condition experience a 10 per cent mortality rate.

Substituting in $SE_p = \sqrt{\frac{pq}{n}}$ We get: $SE_p = \sqrt{\frac{10 \times 90}{100}} = 3$ per cent

On the basis of past experience with this disease and particular treatment, we should not necessarily expect always to see a 10 per cent mortality rate occurring in further samples drawn due to the usual play of chance. We will expect however, that the true mortality in the universe from which our samples have been drawn will lie between 4 and 16 per cent, that is, $p \pm 2 SE$ and we will be wrong in that expectation only once in 20 times.

*Standard error of a difference between proportions*

As we have seen already, samples from any given universe will vary one from the other by play of chance within the broad limits of the range of possible values in that universe. What we now need to do is to assess the probability of, given their difference, two samples being in fact drawn from the same universe. To do this we make an assumption, the null hypothesis, that they do not differ other than by the degree permitted within the limits of chance variation. The standard error of the difference $SE_{p_1 - p_2}$, is calculated from $\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$.
Example

If we had 2 samples of patients, the first being identical with that seen in the previous example with a 10 per cent mortality rate and the second, a sample of 200, with a similar condition but on different treatment, with a 20 per cent mortality. The observed difference is thus $20\% - 10\% = 10\%$. If our null hypothesis is correct this difference has arisen merely by chance. To test the hypothesis we calculate:

$$SE(p_1 - p_2) = \sqrt{\frac{10 \times 90}{100} + \frac{20 \times 80}{200}} = 4.12 \text{ per cent}$$

Now, we have the $SE(p_1 - p_2)$ as 4.12 per cent but our observed difference of 10 per cent is more than twice this figure therefore we can say that this observed difference would occur less than once in 20 times by chance; by convention it is significant $p < 0.05$ and we have “disproved” our null hypothesis.

Standard error of a difference between means

Means of samples, like proportions, vary by chance within the imposed limits of the values of the universe from which they are drawn and we have the same problem of deciding whether an observed difference between the means of two samples is merely such a chance difference or a “real” difference, indicating that the samples are from different universes. The formula used to calculate the standard error of a difference between sample means requires the prior calculation of the standard deviations of the samples (strictly the SD of the universe from which the samples have been drawn is required but as this is manifestly not available we may use the sample SD's without incurring serious error):

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}$$

Example

Suppose a new ration scale had been introduced and soldiers from an infantry battalion had been fed on this scale for a trial period. At the end of the trial their mean weight was compared to that of another battalion fed on the old scale. The results were:

Battalion on new scale ($n = 548$), $\bar{x}_1 = 73.5$ kg, SD = 9.4 kg

Battalion on old scale ($n = 426$), $\bar{x}_2 = 71.0$ kg, SD = 10.2 kg

So, we observed a difference between means of (73.6 - 71.0 kg) 2.5 kg. Is this difference a reflection of a true weight gain in the men of the battalion on the new scale or is it merely a chance variation. Again, we assume the null hypothesis: that is, in reality $\bar{x}_1 - \bar{x}_2 = 0$ and we test this hypothesis by calculating:

$$SE(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{9.4^2}{426} + \frac{10.2^2}{548}} = \sqrt{0.397} = 0.63 \text{ kg}$$

We may now see that the observed difference of means (2.5 kg) is about four times the calculated standard error and as such is highly unlikely to represent a chance difference. The null hypothesis is disproved.
In all these examples the exact statistical significance of the normal deviates e.g.,
\[ \frac{\bar{x}_1 - \bar{x}_2}{\text{SE}(\bar{x}_1 - \bar{x}_2)} \]
can be found by reference to standard statistical tables found in publications such as "Statistical Tables for Students" by J. S. Fowlie.

**Student's t test**

**Background**

In the calculation of the standard error of a difference between means we have made two assumptions; first, that inconsiderable error is incurred by using the standard deviation of the sample rather than that of the universe and, second, that the difference between the means of many successive pairs of examples will fall into a normal distribution about the true mean. This is all very well but there are drawbacks. The first is that if we have only a small number of observations, say less than 30, the assumption that the standard deviation derived from them will be reasonably representative of the standard deviation of the universe of values becomes untenable. Further, it can be shown mathematically that, with small numbers of values, the other assumption that the difference between means will be distributed normally no longer holds.

**Student's t test**

The powerful mind of W. S. Gossett (writing under the pseudonym of "Student") solved the problem for us. The steps are:

a. To calculate and pool the variance (see the first article under standard deviation for a reminder of variance).

b. To use these variances to calculate the standard error of the difference between the means.

c. Finally, to calculate the ratio between the actually observed difference and the standard error.

The interpretation of this ratio (which, the reader will observe is just like the normal deviate referred to above) in terms of probability must be taken from a table of t values.

**Formula**

The formula commonly used is:

\[ t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\text{pooled variance}}{n_1} + \frac{\text{pooled variance}}{n_2}}} \]

**Example**

Two groups of obese patients were placed on a similar diet. In addition Group I were given the latest wonder drug for slimmers while Group II received a placebo. The calculation of the mean loss in kilograms after three months can be seen in Table I.

We therefore observe a marked difference between the groups in the mean weight losses, that is, 7.06 - 4.73 = 2.33 kg. Is this a real difference? To start the calculation of the value of t we will need to calculate the pooled variance.

Remember that:

\[ V = \frac{\Sigma x^2 - \left( \frac{\Sigma x}{n} \right)^2}{n - 1} \]
Pooled variance = \[ \frac{\sum x_1^2 - \left( \frac{\sum x_1}{n_1} \right)^2}{n_1 - 1} + \frac{\sum x_2^2 - \left( \frac{\sum x_2}{n_2} \right)^2}{n_2 - 1} \]

The calculation of pooled variance is shown in Table II.

### Table I
Results after three months in terms of kilograms of weight loss

<table>
<thead>
<tr>
<th>Group I, n = 8</th>
<th>Group II, n = 11</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.5</td>
<td>16.0</td>
</tr>
<tr>
<td>7.0</td>
<td>2.0</td>
</tr>
<tr>
<td>11.0</td>
<td>1.5</td>
</tr>
<tr>
<td>11.5</td>
<td>9.5</td>
</tr>
<tr>
<td>2.5</td>
<td>3.0</td>
</tr>
<tr>
<td>1.0</td>
<td>6.5</td>
</tr>
<tr>
<td>6.5</td>
<td>8.5</td>
</tr>
<tr>
<td>4.5</td>
<td>1.0</td>
</tr>
<tr>
<td>( \Sigma x_1 ) = 56.5</td>
<td>( \Sigma x_2 ) = 52.0</td>
</tr>
</tbody>
</table>

The respective means are: \( \bar{x}_1 = \frac{56.5}{8} = 7.06 \) kg, \( \bar{x}_2 = \frac{52.0}{11} = 4.73 \) kg.

### Table II
The calculation of pooled variance

<table>
<thead>
<tr>
<th>Group I</th>
<th>Group II</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_1 )</td>
<td>( x_2 )</td>
</tr>
<tr>
<td>( x_1^2 )</td>
<td>( x_2^2 )</td>
</tr>
<tr>
<td>12.5</td>
<td>16.0</td>
</tr>
<tr>
<td>156.25</td>
<td>256.00</td>
</tr>
<tr>
<td>7.0</td>
<td>2.0</td>
</tr>
<tr>
<td>49.00</td>
<td>4.00</td>
</tr>
<tr>
<td>11.0</td>
<td>1.5</td>
</tr>
<tr>
<td>121.00</td>
<td>2.25</td>
</tr>
<tr>
<td>11.5</td>
<td>1.5</td>
</tr>
<tr>
<td>132.25</td>
<td>2.25</td>
</tr>
<tr>
<td>2.5</td>
<td>3.0</td>
</tr>
<tr>
<td>6.25</td>
<td>9.00</td>
</tr>
<tr>
<td>1.0</td>
<td>6.5</td>
</tr>
<tr>
<td>1.00</td>
<td>42.25</td>
</tr>
<tr>
<td>6.5</td>
<td>2.0</td>
</tr>
<tr>
<td>42.25</td>
<td>4.00</td>
</tr>
<tr>
<td>1.0</td>
<td>8.5</td>
</tr>
<tr>
<td>1.00</td>
<td>72.25</td>
</tr>
<tr>
<td>4.5</td>
<td>0.5</td>
</tr>
<tr>
<td>20.25</td>
<td>0.25</td>
</tr>
<tr>
<td>( \Sigma x_1 = \frac{56.5}{8} ) = 7.06</td>
<td>( \Sigma x_2 = \frac{52.0}{11} ) = 4.73</td>
</tr>
</tbody>
</table>

So:

\[
\frac{\left( \frac{\sum x_1}{n_1} \right)^2}{n_1} = \frac{56.5^2}{8} = 399.0 \quad \frac{\left( \frac{\sum x_2}{n_2} \right)^2}{n_2} = \frac{52^2}{11} = 245.82
\]

Substituting into the equation for pooled variance:

\[
\frac{528.25 - 399.0 + 483.5 - 245.82}{7 + 10} = 21.58
\]
Expressing the original equation in terms of $t^2$ (to avoid the square root until the end of the calculation) we get: $t^2 = \frac{(x_1 - x_2)^2}{(pooled\ variance)\ \frac{n_1}{(pooled\ variance)\ \frac{n_2}}}$

Therefore, substituting: \[
\frac{(7.06 - 4.73)^2}{21.58 - 21.58} = 1.165, \ t = 1.08
\]

Consulting the table. We have now calculated that $t = 1.08$ and we must consult the appropriate table (an extract is given below). First however we need the number of degrees of freedom that the original values allow. This implies the number of independent values contributing to the calculation of $t$. As a simple example to illustrate the point—if we had the equation $5 + 2 + 7 = 14$, knowing the right hand side of the equation and any two of the digits on the left hand side allows us to calculate the third or dependent digit. So, this equation may be said to have two degrees of freedom. Applying this concept to our original example we calculate degrees of freedom from $(n_1 + n_2) - 2$, so $(8 + 11) - 2 = 17$.

Extract from table of $t$ distribution

<table>
<thead>
<tr>
<th>Degrees of freedom</th>
<th>.10</th>
<th>.05</th>
<th>.02</th>
<th>—</th>
<th>—</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>*1.75</td>
<td>2.13</td>
<td>2.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.72</td>
<td>2.09</td>
<td>2.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Examining our table we find 15 and 20 degrees of freedom but no 17. Undismayed we look between the rows where 17 ought to be finding a group of figures near (although in this case, not very) to our calculated 1.08. It can be seen that it falls to the left of column under $p = .10$. Therefore in our case $p > .10$ and we have found no significant difference between the mean weights of our patients.

The $t$ test in the "before and after" situation

Often in clinical studies a small group of patients with a specific condition is observed before and after some specific treatment. Sometimes the treatment will effect an apparent improvement. What we need to know is whether this improvement is in truth due to our therapy or due to chance variation. As we are dealing with the same people twice rather than with two groups we can avoid the use of pooled variance and merely calculate the ratio of the observed difference to its own standard error to find the value of $t$.

Example

Nine patients with severe obstructive pulmonary disease were treated with a long-acting bronchodilator and the results of measurement of $FEV_{1.0}$ before and after therapy are shown in Table III together with the calculation of $t$. 
D. A. Moore

Table III

Results from nine patients treated with a long-acting bronchodilator

<table>
<thead>
<tr>
<th>Before treatment FEV₁₀ litres</th>
<th>After treatment FEV₁₀ litres</th>
<th>Observed difference x</th>
<th>x²</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.8</td>
<td>3.4</td>
<td>0.6</td>
<td>0.36</td>
</tr>
<tr>
<td>1.6</td>
<td>1.7</td>
<td>0.1</td>
<td>0.01</td>
</tr>
<tr>
<td>1.9</td>
<td>1.9</td>
<td>0.0</td>
<td>0</td>
</tr>
<tr>
<td>1.1</td>
<td>1.0</td>
<td>0.1</td>
<td>0.01</td>
</tr>
<tr>
<td>2.0</td>
<td>2.9</td>
<td>0.9</td>
<td>0.81</td>
</tr>
<tr>
<td>1.3</td>
<td>3.2</td>
<td>1.2</td>
<td>1.44</td>
</tr>
<tr>
<td>2.3</td>
<td>1.8</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>2.5</td>
<td>2.5</td>
<td>0.2</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>0.3</td>
<td>0.09</td>
</tr>
</tbody>
</table>

n = 9

\[ \Sigma x = 3.7 \quad \Sigma x^2 = 3.01 \]

Mean observed difference \( \frac{(\Sigma x)}{n} = \frac{3.7}{9} = 0.41 \)

Standard deviation \[ \sqrt{\frac{\Sigma x^2 - (\Sigma x)^2}{n-1}} = \sqrt{\frac{3.01 - (3.7)^2}{9}} = 0.43 \]

Standard error \[ \frac{SD}{\sqrt{n}} = \frac{0.43}{\sqrt{9}} = 0.14 \]

\[ t = \frac{\text{observed mean difference}}{\text{SE of difference}} = \frac{0.41}{0.14} = 2.93 \]

Entering our \( t \) distribution table with 8 degrees of freedom \((n - 1)\) we can see from the extract below that our calculated value for \( t \) indicates that \( 0.02 > p > 0.01 \) and therefore the improvement observed in these patients is statistically significant.

Extract from table of \( t \) distribution

<table>
<thead>
<tr>
<th>Degrees of freedom</th>
<th>.10</th>
<th>.05</th>
<th>.02</th>
<th>.01</th>
<th>—</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.86</td>
<td>2.31</td>
<td>2.90</td>
<td>*3.36</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the next article the \( \chi^2 \) test and correlation and regression coefficients will be dealt with.

(to be continued)