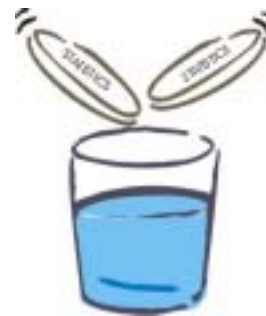


Statistics in Divided Doses



September 2001 Number 3

Assessing the reliability of a sample

Contents

- Factors to consider when assessing a sample
- Confidence intervals
- Standard error of the mean
- Confidence intervals for median values
- Parametric vs. non-parametric methods

Factors to consider when assessing a sample

- **Sample size**
Is it large enough to produce useful conclusions? At a later date, we will consider the question of sample size in greater depth.
- **Method of selection**
 - What are the exclusion and inclusion criteria?
 - What is the setting? e.g. primary vs. secondary care.
 - How have patients been recruited? e.g. advertisement vs. clinic.
- **What are the characteristics of the population?**
Are they likely to be significantly different from the sample? Look at the demographics of the sample – is the population likely to be older, or have more co-morbidity etc.?
- **Sample Data**
Are the sample data appropriate for the purposes of the trial? Are the response variables a sensible choice? Is there wide variability (scatter) between individual values?

What does the term 'variable' mean?

A variable is a measurement taken from subjects within a sample e.g. blood pressure, weight etc. In a broader sense, we can use this term to describe anything that varies within a set of data.

What are the important statistical attributes of a sample?

The mean (or, when appropriate, the median) provides an estimate of the mean (or median) of the parent

population. However, we also need some indication of the variability within our sample e.g. the standard deviation.

Can we measure how closely the sample mean is representative of the population mean?

Yes, by calculating the related *confidence interval* (CI).

Confidence intervals

What is a confidence interval?

A confidence interval is a numerical index of the reliability of the sample mean and is a range of values within which we expect the *true mean of the population* to lie. This is the mean value we would obtain if we were able to measure the entire population. Therefore, the confidence interval provides a measure of the uncertainty surrounding a single sample estimate of the population mean. It is expressed as a range of values around the mean at a selected level of probability which is usually chosen as 95%.

How can the confidence interval be calculated?

For data that follow an approximately Normal distribution, the confidence interval can be calculated from the *standard error of the mean*.

Standard error of the mean

What is a standard error of the mean?

This is a numerical measure of the reliability of that mean.

In *Issue 2* we saw that by adding to or subtracting multiples of the standard deviation from the sample mean we obtain an estimate of the spread of values within a sample.

Similarly, to calculate the confidence interval, we can use the standard error of the mean (SEM). Multiples of this value added to, or subtracted from the mean gives an estimate of the confidence interval. This approach is valid for Normally distributed data, for which a mean value can be calculated. Confidence intervals for a

median value involve a different approach, as we will see later.

How is the standard error of the mean calculated?

By dividing the standard deviation of the sample by the square root of the sample size.

Using our series of height measurements for 20 individuals, the mean value is 170cm and the standard deviation is 8.4cm. The standard error of the mean (SEM) can be calculated as follows:

$$\begin{aligned}\text{SEM} &= \text{SD} \div \sqrt{n} \\ &= 8.4 \div \sqrt{20} \\ &= 1.9\text{cm}\end{aligned}$$

What does the standard error indicate?

In the above example, the standard error is very small compared with the mean. This indicates that the variability of values within the sample is relatively small and that the sample size, although small, is sufficient to provide meaningful data concerning the parent population.

Would the standard error decrease with increasing sample size?

From the above formula we can see that as the sample size (n) increases, the standard error of the mean (SEM) decreases.

How is the standard error of the mean derived?

It is derived from the frequency distribution of sample means. This is best explained using an example.

As before, we can use our fictitious sample of the heights of 20 men. If we measured the heights of a further 20 men from the parent population, it would be unlikely that the mean of our second sample would be identical to the first. If we took a further 98 random samples, each of 20 men, we would be able to calculate a total of 100 mean values which may all differ slightly; some sample means would be closer to the true population mean than others, some would be higher, others lower. If we plotted the mean values against their frequency we would obtain frequency distribution curve, which would be similar in shape to that of the Normal distribution. This distribution would have its own standard deviation with an expected value of $\sigma \div \sqrt{n}$ where;

σ = standard deviation of the variable (i.e. height) in the population

n = size of the sample (in our example n = 20).

N.B. Note the difference between s (standard deviation of a single sample) and σ (standard deviation of the population).

The standard deviation of the means of several samples ($\sigma \div \sqrt{n}$) is known as the standard error of the mean to distinguish it from the standard deviation of the individual observations. Thus it is a rather hypothetical

concept and we normally use a single sample to estimate its value.

How do we calculate the confidence interval from the standard error?

To calculate the confidence interval, we add or subtract approximately twice the standard error, to and from the calculated mean. Therefore, using our height example in which SEM = 1.9cm, the 95% confidence interval would be 170 ± 3.8 cm i.e. approximately 166 to 174 cm.

Why do we use an approximate calculation in the above sample?

If we want to use a 95% level of probability, the precise value of the multiple is 1.96 since 95% of values in a Normal distribution lie within the range ± 1.96 times the standard deviation around the mean. Thus a multiple of 2 is an approximation. If we wish to be precise, a multiple of 2.09 should be used when n = 20. This multiple can be found from the t distribution, which will be discussed later.

Do confidence intervals enable us to compare sample means?

Yes. Even though two mean values may differ, extensive overlap of their respective confidence intervals may suggest that the difference is not statistically significant. This will be discussed further in a future issue.

What does the probability value attached to a confidence interval tell us?

The probability value is a measure of our certainty that the true population mean lies within the confidence interval. Hence a 95% confidence interval implies that we can expect the confidence interval to contain the true population mean 95% of the time and not to contain it 5% of the time. If we are more demanding, we can calculate a 99.7% confidence interval by using a multiple of three standard errors in our calculation.

Why is the confidence interval useful within a clinical trial report?

It enables us to estimate the reliability of the mean variables cited in the report.

For example, if the mean systolic blood pressure measured in a group of patients treated with an antihypertensive drug was 150mmHg (95% CI 130 - 180mmHg), we can assume, with a 95% level of confidence, that the true mean of the blood pressures of a treated population would be within the limits of 130 to 180mmHg. But a mean systolic blood pressure of between 130 to 180mmHg is compatible with both a normotensive and hypertensive state and might lead us to question the efficacy of the treatment. In order to assess the *real effect* of drug treatment in reducing blood pressure, we would wish to know what pre-treatment values were for comparison.

Confidence intervals for median values

How is a confidence interval for a median value calculated?

By ranking the data and obtaining the appropriate confidence intervals from a standard table.

How can data in a sample be ranked?

By setting out the individual values in ascending order. Using our example of the heights of 20 men, we would have;

150, 155, 162, 165, 167, 167, 167, 167, 170, 170, 170, 170, 172, 172, 172, 175, 177, 180, 180, 187.

As we have an even number of values (i.e. $n = 20$) there is no middle value. Therefore, we calculate the median by adding the 10th and 11th values together (i.e. 170 and 170) and dividing by 2, to give 170cm. Using a table of ranks for confidence intervals of the median (found in statistical books) we see that a 95% confidence interval would lie between the 6th and 15th ranks i.e. between 167 and 172cm.

Compare these results with our previous calculations, using the same data, resulting in a mean value of 170cm and confidence intervals of 166-174cm. We can see that the *mean* and the *median* values are the same and that the confidence intervals similar. This implies that our sample has been taken from a population with a Normal distribution. If these values were different it would imply a skewed frequency distribution.

In calculating a confidence interval from ranked data we are using a *non-parametric* method of analysis. This method of analysis makes no assumptions about the distribution characteristics of the data.

Parametric vs non parametric analysis

What are parametric and non-parametric methods of analysis?

The difference between these methods lies in underlying assumptions made about the data. Parametric analysis makes the assumption that the data are from a population with a known *probability distribution*. Non-parametric (also known as "distribution-free") analysis makes no such assumptions.

What is a probability distribution?

This is used to calculate the theoretical probability of a particular value occurring and is the theoretical equivalent of the relative frequency distribution. When we measured the heights of 20 men, we illustrated our results using a histogram. This showed the relative frequency distribution for each height i.e. how many men had a particular height. We were able to calculate the mean and standard deviation of the frequency distribution for the sample. We were also able to predict certain characteristics of the population from which our sample was drawn e.g. that 95% of men in that population would probably be within a certain height

range. For many such data, we are able to make the assumption that the distribution of values within the parent population follows a defined theoretical form. If this assumption is reasonable (we cannot prove it) then it is reasonable to use parametric methods of analysis.