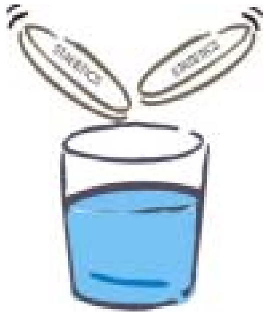
Statistics in Divided Doses



November 2002 Number 5

First steps in analysis - comparing the means of large samples

Contents

Standard error of the mean vs. standard deviation

What is the difference between the standard error of the mean (SEM) and the standard deviation (SD)?

(To derive SD see *Statistics in Divided Doses 2* and SEM see *Statistics in Divided Doses 3*)

The standard deviation is a measure of the variability (scatter) of individual values within a sample or population. The SEM of a sample indicates the accuracy of that mean i.e. how closely it represents the population mean.

$$\text{SEM} = \frac{\text{SD}}{\sqrt{n}} \qquad n = \text{sample size}$$

As n increases, the SEM decreases. It follows that if n is very large, the SEM would be negligible since the size of the sample would approach that of the parent population. Therefore you can assume, if the SEM is small, that the mean value of your sample is close to that of the population from which your sample is drawn.

Comparing the means of large samples

How do we compare the means of two large samples?

Assume that our data are from a population with a Normal distribution. Using a fictitious example, we are interested in the effects of a new drug *Rumbocur* on the blood levels of the enzyme nullotransferase. Blood samples from 90 patients treated with *Rumbocur* are compared with blood samples of 100 untreated controls. We can assume that subjects are well matched for potentially relevant factors such as age and gender.

Mean nullotransferase blood levels (ng/ml);

Rumbocur group = 30.0, SD = 6.5
Control group = 20.0, SD = 5.0.

What is the difference in mean values of nullotransferase levels?

30 - 20 ng/ml = 10 ng/ml.

The standard error of the difference between two sample means

Why calculate a standard error of the difference between two means?

A standard error can be attached to any population parameter. The standard error of the difference between two sample means is a measure of the precision of that difference.

What is the standard error of the difference in sample means?

This is calculated from the following formula;

$$\text{SE (diff)} = \sqrt{\left(\frac{\text{SD}_1^{\,2}}{n_1}\right) + \left(\frac{\text{SD}_2^{\,2}}{n_2}\right)}$$

From our investigation we already know that;
SD_1 (standard deviation of the *Rumbocur* sample) = 6.5 ng/ml.
SD_2 (standard deviation of the control sample) = 5.0 ng/ml.
n_1 (size of the *Rumbocur* sample) = 90
n_2 (size of the control sample) = 100

Substituting in the above equation, we get;

$$\text{SE (diff)} = \sqrt{\left(\frac{6.5^2}{90}\right) + \left(\frac{5^2}{100}\right)}$$

Produced for healthcare professionals by North West Medicines Information Service, The Pharmacy Practice Unit, 70 Pembroke Place, Liverpool, L69 3GF. Editor: Frank Leach. Telephone: 0151 794 8117 E-mail: druginfo@liv.ac.uk

$$= \sqrt{\left(\frac{42.25}{90}\right) + \left(\frac{25}{100}\right)}$$

$$= \sqrt{(0.47 + 0.25)}$$

∴ SE (diff) = 0.85

The 95% confidence interval for this difference

How can we use the standard error of the difference?

To calculate the 95% confidence interval (CI) for the difference in means, we calculate the following;

= difference in means – [1.96 x SE (diff)] to difference in means + [1.96 x SE (diff)]

= 10 - (1.96 x 0.85) to 10 + (1.96 x 0.85)

= 10 - 1.666 to 10 + 1.666

= 8.33 to 11.67

Thus we expect, with a 95% level of probability, that the true difference in the mean nullotransferase values lies between 8.3 and 11.7 ng/ml.

Why have we used the figure 1.96 to calculate the 95% confidence intervals?

Previously we have doubled the standard deviation to approximate the 95% level but in the Normal distribution, 95% of the values around the mean are, to be precise, within 1.96 standard deviations above and below that value.

Testing the null hypothesis and test statistics

What is the null hypothesis?

(See *Statistics in Divided Doses 4* for the definition of the null hypothesis.)
The null hypothesis for the *Rumbocur* study is "there is no difference between patients who receive *Rumbocur* and control patients with respect to blood nullotransferase levels".

What is a test statistic?

A 'test statistic' is a value that we calculate in order to test the null hypothesis.

Which test statistic should we use?

A test statistic is specific for a known theoretical probability distribution. We have assumed that our sample is taken from a population with a Normal distribution. The test statistic for the Normal distribution is referred to as the *z* value (also known as a standard Normal deviate, see below).

$$z = \frac{\text{observed value - hypothesised value}}{\text{SE of observed value}}$$

In our case;

z = (difference in mean nullotransferase levels between *Rumbocur* and control groups) - (expected difference in means if the null hypothesis is true) ÷ (SE of the difference between the mean nullotransferase levels)

Substituting our values

$$z = \frac{10 - 0}{0.85} = 11.8$$

How do we use the test statistic to test the null hypothesis?

We use a statistical table to compare the value of the test statistic (*z* value) with values from the known theoretical probability distribution (Normal distribution, see Table 1, below) to obtain the P value (see *Statistics in Divided Doses 4*). The values in the table relate multiples of standard deviations, taken from the Normal distribution, to probabilities.

For the *Rumbocur* study, when z = 11.8, we can say that P < 0.0001. This result is highly statistically significant.

Table 1 Values from the Normal distribution for use in calculating confidence intervals

P value (Probability of the observation) Two tailed	z value (Number of Standard Deviations)
0.5	0.674
0.1	1.645
0.05	1.960
0.01	2.576
0.001	3.291
0.0001	3.891

Do we accept or reject the null hypothesis?

We can reject the null hypothesis as it is extremely unlikely to be true (probability of this is less than 0.0001). Put another way, the null hypothesis would be expected to be true in less than 1 in 10,000 trials.

Does rejecting the null hypothesis prove anything?

Strictly speaking, no. From a statistical standpoint we have reduced our uncertainty to a very low level and if the results prove to be reproducible in further studies, it would be reasonable to extrapolate our findings to clinical practice. In most "real life" clinical trials, such extremely consistent data would be the exception rather than the rule. Frequently, variability in response is large, confidence intervals are wide and P values are correspondingly greater. In these circumstances, the uncertainty that surrounds the outcome is proportionally increased, demanding a cautious interpretation before extrapolation to routine clinical practice.

The Normal probability distribution

What is the difference between frequency and probability distributions?

Frequency distribution - indicates the number of times a specific outcome occurs, using observed data (see *Statistics in Divided Doses 2*).

Probability distribution - indicates the probability of variables attaining specific values using a theoretical model (see *Statistics in Divided Doses 3*).

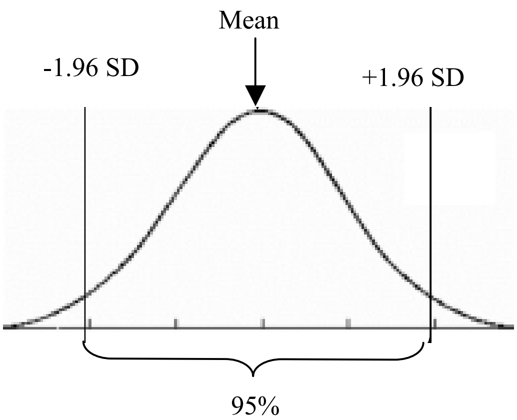
What are the key features of the Normal probability distribution?

- It is a continuous probability distribution in that it relates to a continuous variable (e.g. height), which can have any value, to any level of precision, within defined limits.
- It is a theoretical distribution and is expressed mathematically using defined parameters.
- It extends from minus infinity to plus infinity since it is hypothetical and its limits are not defined.
- The height of the curve is of little practical use as the number of possible values is infinite and the probability of any specific value is 0. It is the area under the curve (AUC) which is of interest. The AUC always equals 1 and represents the probability of all possible values.

What conclusions can we draw from these features?

The area corresponding to a defined limit of values provides the specific probability for those values. Thus, the area defined by 1.96 standard deviations above and below the mean, corresponds to an area that is 95% of the total, and thus values within this area occur with a probability of 95% (see Figure 1).

Figure 1 Normal probability distribution curve



What is the difference between "one-tailed" and "two-tailed " tables of the Normal distribution?

One-tailed P value - the probability of getting a result that is equal to or greater than the observed result. Used only if the direction of an intervention can be specified e.g. a treatment can either have no effect or improve response.

In a one tailed table (not shown), probabilities are given for both the upper and lower tail areas. E.g. when z = 1.96, a P_{upper} is given as 0.025 and P_{lower} as 0.975.

Two-tailed P value - the probability of getting a result that is equal to, greater or less than the observed result. Used when the direction of an intervention is not known e.g. a new drug may either cause a patient to improve or deteriorate.

In a two tailed table (Table 1), only a single P value is given. E.g. when z = 1.96, the P value of 0.05 is the proportion of values outside the range of ± 1.96 standard Normal deviates.

What is the standard Normal deviate?

The Normal distribution is completely defined by its mean (μ) and its standard deviation (σ). Thus, any point along the abscissa (x-axis) of the Normal distribution can be expressed as the distance of a number of standard deviations from the mean. This is known as the standard Normal deviate.

If $\sigma = 2$, $\mu = 10$ and $x = 15$,

$$\text{standard Normal deviate} = \frac{15 - 10}{2} = 2.5$$

This is, in fact, the *z* value that we used in testing the null hypothesis by means of the Normal distribution. The standard Normal deviate is also known as the Normal score.

What is the standard Normal distribution?

A Normal distribution where the mean is 0 and the standard deviation is 1 is called the standard Normal distribution (Figure 2).

Figure 2 Standard Normal distribution