

Contents

1 Confidence Intervals

1.1 Large sample distribution of sample mean

1.2 Confidence limits of the mean for large samples

Now that we know the form of the sampling distribution of the mean we can return to the problem of using a sample to define a range which we may reasonably assume includes the true value. (Remember that in doing this we are assuming systematic errors to be absent.) Such a range is known as a confidence interval and the extreme values of the range are called the confidence limits.

The term 'confidence' implies that we can assert with a given degree of confidence, i.e. a certain probability, that the confidence interval does include the true value.

The size of the confidence interval will obviously depend on how certain we want to be that it includes the true value: the greater the certainty, the greater the interval required.

Figure 2.6 shows the sampling distribution of the mean for samples of size n . If we assume that this distribution is normal then 95% of the sample means will lie in the range given by:

$$\mu - 1.96(\sigma/\sqrt{n}) < \mu < \mu + 1.96(\sigma/\sqrt{n}) \quad (2.6)$$

(The exact value 1.96 has been used in this equation rather than the approximate value, 2, quoted in Section 2.2. The reader can use Table A.1 to check that the proportion of values between $z = -1.96$ and $z = 1.96$ is indeed 0.95.)

1.3 Large Sample Confidence Intervals based on the sample mean

In practice, however, we usually have one sample, of known mean, and we require a range for μ , the true value.

Equation (2.6) can be rearranged to give this:

$$\bar{X} - 1.96(\sigma/\sqrt{n}) < \mu < \bar{X} + 1.96(\sigma/\sqrt{n}) \quad (2.7)$$

Equation (2.7) gives the 95% confidence interval of the mean. The 95% confidence limits are $\bar{X} \pm 1.96\sigma/\sqrt{n}$. In practice we are unlikely to know σ exactly. However, provided that the sample is large, σ can be replaced by its estimate, s .

1.4 Example of computations using R

Finding confidence intervals for the mean for the nitrate ion concentrations in Table 2.1.

```
#reading data
x=scan("Table2_1.txt")
#setting the confidence level
CL=0.95
#computing confidence interval
n=length(x)
pm=sd(x)*c(qnorm(0.025),qnorm(0.975))/sqrt(n)
CI=mean(x)+pm
```

1.5 Small-Sample Case ($n \leq 30$)

If the data have a normal probability distribution and the sample standard deviation s is used to estimate the population standard deviation σ , the interval estimate is given by:

$$\bar{X} \pm t_{\alpha/2}s/\sqrt{n}$$

where $\alpha/2$ is the value providing an area of $\alpha/2$ in the upper tail of a Student's t -distribution with $n - 1$ degrees of freedom.

1.6 Confidence limits of the mean for small samples

As the sample size gets smaller, s becomes less reliable as an estimate of σ . This can be seen by again treating each column of the results in Table 2.2 as a sample of size five. The standard deviations of the 10 columns are

$$\{0.009, 0.015, 0.026, 0.021, 0.013, 0.019, 0.013, 0.017, 0.010 \text{ and } 0.018\}.$$

We see that the largest value of s is nearly three times the size of the smallest. To allow for this, equation (2.8) must be modified.

For small samples, the confidence limits of the mean are given by

$$\bar{X} \pm t_{n-1}s/\sqrt{n}$$