2 ESTIMATION

Objectives

After studying this chapter you should

- be able to calculate confidence intervals for the mean of a normal distribution with unknown variance;
- be able to calculate confidence intervals for the variance and standard deviation of a normal distribution;
- be able to calculate approximate confidence intervals for a proportion;
- be able to calculate approximate confidence intervals for the mean of a Poisson distribution.

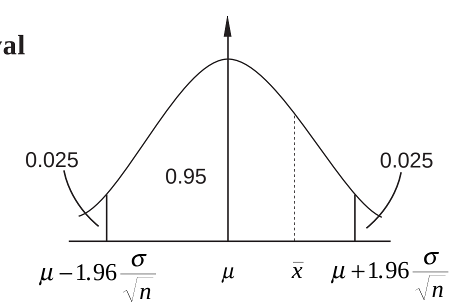
2.0 Introduction

In Chapter 9 of the text *Statistics* the idea of a confidence interval was introduced. Confidence intervals are used when we want to estimate a population parameter from a sample. The parameter may be estimated by a single value (a point estimate) but it is usually preferable to estimate it by an interval which will give some indication of the amount of uncertainty attached to the estimate. In *Statistics*, estimation of the mean of a normal population with known standard deviation was considered.

If $\bar{x}$ is the mean of a random sample of size n from a normal distribution with mean μ and standard deviation σ there is a probability of 0.95 that $\bar{x}$ lies within $1.96\frac{\sigma}{\sqrt{n}}$ of μ. If this is the case the interval

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$$

will contain μ. This interval is called a **95% confidence interval** for μ.



If further samples of size n were taken and the calculation repeated, different intervals would be calculated. 95% of these intervals would contain μ, but 5% would not.

Note: although μ is unknown, it does not vary, it is the intervals that vary. It is possible to calculate 99% or even 99.9% confidence intervals which would be wider than the 95% interval but it is not possible to calculate 100% confidence intervals.

Example

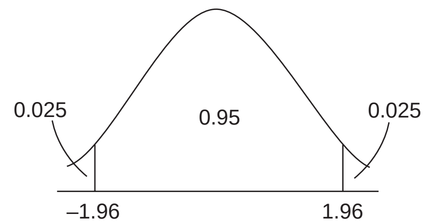
The length of time a bus takes to travel from Chorlton to All Saints in the morning rush hour is normally distributed with standard deviation 4 minutes. A random sample of 6 journeys took 23, 19, 25, 34, 24 and 28 minutes. Find

(i) a 95% confidence interval for the mean journey time,

(ii) a 99% confidence interval for the mean journey time.

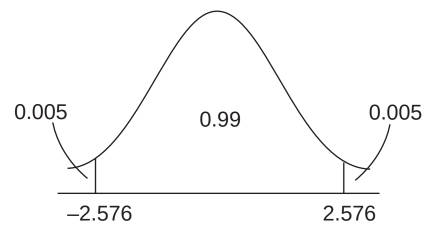
Solution

The sample mean is $\frac{153}{6} = 25.5$



(i) 95% confidence interval is

$25.5 \pm 1.96 \times \frac{4}{\sqrt{6}}$ i.e 25.5 ± 3.20 or $(22.3,\ 28.7)$



(ii) 99% confidence interval is

$25.5 \pm 2.576 \times \frac{4}{\sqrt{6}}$ i.e. 25.5 ± 4.21 or $(21.3,\ 29.7)$

2.1 Confidence interval for mean: standard deviation unknown

The example above of the bus journey times is somewhat unrealistic. How was it known that the bus journey times were normally distributed with a standard deviation of 4 minutes? If so much was known about the distribution how was it that the mean was unknown?

Probably the journey times for similar journeys had been studied and the standard deviation found to be about four minutes. The statement that the standard deviation was 'known' was probably something of an exaggeration. The same argument applies to the normal distribution. However, as you are dealing with the sample mean, no great error will result from assuming a normal distribution unless the distribution is extremely unusual.

If the standard deviation is completely unknown or if you do not wish to accept a rough estimate, there is still a way forward. Provided the sample has more than one member, an estimate of the standard deviation may be made from the sample.

In the case of the bus journey times $\hat{\sigma} = 5.09$, this is most easily found by using the σ_{n-1} or s_{n-1} button on a calculator. Alternatively the formula

$$\hat{\sigma}^2 = \sum \frac{(x-\bar{x})^2}{(n-1)}$$

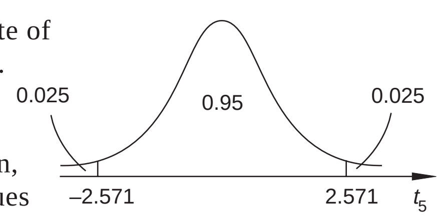
or equivalent can be used.

For a 95% confidence interval, σ known, the formula $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$ was used. Now the known standard deviation σ will be replaced by an estimate $\hat{\sigma}$. There will be some uncertainty in this estimate since, if you started again and took a different sample of the same size, the estimate of $\hat{\sigma}$ would almost certainly be different. It therefore seems reasonable that to allow for this extra uncertainty, the interval should be widened by increasing the figure of 1.96 which came from tables of the normal distribution. How much you need to increase it by has fortunately been calculated for you and is tabulated in tables of the ***t* distribution**.

The required value of *t* will, however, depend on the sample size. If you had a random sample of 1000 bus journeys then you would be able to make a very accurate estimate of the population standard deviation and there would be no real need to change the figure 1.96. However, in this case with a sample of 6 there is quite a lot of uncertainty in the estimate and you may need to make quite a large change. If the sample had been of size 2 there would have been a huge amount of uncertainty and a very large change may have been necessary.

The amount of uncertainty is measured by the **degrees of freedom**. Degrees of freedom are a concept related to the mathematical definition of the **chi-squared distribution**. For your purposes they may be thought of as a measure of the number of pieces of information you have to estimate the standard deviation. It is impossible to use a sample of size one to make an estimate of the standard deviation. The standard deviation is a measure of spread and one item can tell nothing about how spread out a distribution is. A sample of size two gives one piece of information about the spread and a sample of size three gives two pieces. In general, a sample of size *n* gives $n-1$ pieces of information and an estimate made from such a sample is said to be based on $n-1$ degrees of freedom.

In the example there were 6 bus journey times and so the estimate of 5.09 for the standard deviation is based on 5 degrees of freedom. To find a 95% confidence interval you therefore require the upper and lower 0.025 tails of the *t* distribution with 5 degrees of freedom, denoted t_5. As with the standard normal distribution, the *t* distribution is symmetrical about zero and the required values are ± 2.571.



A 95% confidence interval for the mean, making no assumption about the standard deviation, is given by

$$25.5 \pm 2.571 \times \frac{5.09}{\sqrt{6}} \quad \text{i.e. } 25.5 \pm 5.34 \text{ or } (20.2,\ 30.8)$$

Note: For the use of the *t* distribution to be valid the data must be normally distributed. However, small deviations from normality will not seriously affect the results.

In general, the formula is

$$\bar{x} \pm t_{n-1} \frac{\hat{\sigma}}{\sqrt{n}}$$

Example

The resistances (in ohms) of a random sample from a batch of resistors were

2314 2456 2389 2361 2360 2332 2402

Assuming that the sample is from a normal distribution calculate

(i) a 95% confidence interval for the mean,

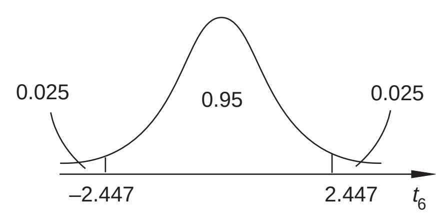
(ii) a 90% confidence interval for the mean.

Solution

The data gives $\bar{x} = 2373.4$ and $\hat{\sigma} = 47.4$.

(i) $t_{6,\ 0.025} = 2.447$, so the

95% confidence interval for the mean is given by



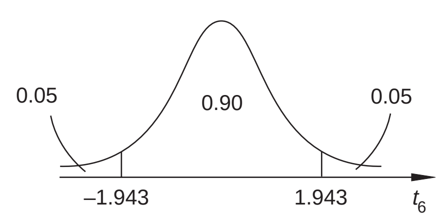
$$2373.4 \pm 2.447 \times \frac{47.4}{\sqrt{7}}$$

$$\Rightarrow \quad 2373.4 \pm 43.8$$

$$\Rightarrow \quad (2330,\ 2417).$$

(ii) $t_{6,\ 0.05} = 1.943$, giving

90% confidence interval for the mean as



$$2373.4 \pm 1.943 \times \frac{47.4}{\sqrt{7}}$$

$$\Rightarrow \quad 2373.4 \pm 34.8$$

$$\Rightarrow \quad (2339,\ 2408).$$

Activity 1

The following data are random observations from a normal distribution with mean 10.

6.86	11.53	12.41	12.08	12.80	10.42	8.99	9.55	8.23	5.84
7.59	5.96	10.14	10.12	10.22	10.42	11.83	8.73	11.57	11.83
10.76	9.93	10.63	7.94	12.44	12.49	9.63	9.45	13.40	10.78
13.44	11.85	13.62	13.24	12.56	10.56	10.77	8.51	11.65	9.36
8.12	11.88	11.68	7.36	7.07	10.04	9.55	12.97	10.85	8.58
8.27	9.22	11.36	9.43	8.80	9.07	7.66	13.16	8.34	7.12
3.49	13.04	13.16	11.48	8.30	10.01	10.29	11.78	13.18	8.18
10.00	12.27	14.18	9.91	9.62	7.48	8.50	10.53	13.06	6.74
6.05	9.96	7.51	10.19	9.07	9.29	6.01	12.02	10.04	10.64
9.74	8.23	9.45	5.41	9.68	10.64	6.77	10.76	8.10	10.33
8.34	11.61	9.72	11.24	12.84	6.10	10.78	8.27	8.52	7.42
8.91	8.52	10.66	14.06	9.37	10.44	11.81	9.87	9.78	10.44
10.82	10.10	7.68	11.87	7.49	9.99	8.54	4.65	5.37	8.83
15.00	10.02	9.41	8.16	9.54	9.32	6.15	12.59	12.24	13.02
9.80	8.61	8.92	8.86	11.92	13.01	14.11	11.57	10.46	11.27
8.35	8.95	9.12	7.20	11.20	13.42	13.46	12.80	10.99	10.33
14.31	7.72	9.88	10.57	13.20	11.90	8.48	9.41	7.76	10.35
8.78	9.45	11.48	10.96	7.68	9.26	14.29	8.35	6.80	8.29
8.83	10.72	10.02	11.80	13.56	13.00	10.79	7.51	8.15	10.14
11.02	8.49	9.82	8.97	9.86	7.74	11.81	9.87	10.77	9.18

Starting at any point in the table take a sample of size 3 and calculate $\hat{\sigma}$. Calculate an 80% confidence interval for the mean using the *t*-distribution.

Now take the next sample of 3 and repeat the calculation. Carry on until you have calculated at least 20, and preferably more, intervals. If possible work with a group so that the labour of calculation may be divided up between you.

What proportion of your intervals contain 10, the population mean? Is this approximately the proportion you would have expected?

Recalculate the intervals using z-values, i.e. calculate, for each sample

$$\bar{x} \pm 1.282 \frac{\hat{\sigma}}{\sqrt{3}}.$$

What proportion of these intervals contain 10? Which set of intervals gave results more in line with your expectations?

Exercise 2A

1. Samples of a high temperature lubricant were tested and the temperature (°C) at which they ceased to be effective were as follows:

   235 242 235 240 237 234 239 237

   Calculate a 95% confidence for the mean.

2. In a study aimed at improving the design of bus cabs the functional arm reach of a random sample of bus drivers was measured. The results, in mm, were as follows:

   701, 642, 651, 700, 672, 674, 656, 649

   Calculate a 95% confidence interval for the mean.

3. As part of a research study on pattern recognition a random sample of students on a design course were asked to examine a picture and see if they could recognise a word. The picture contained the word 'technology' written backwards. The times, in seconds, taken to recognise the word were as follows:

   55, 28, 79, 54, 87, 61, 62, 68, 38

   Calculate

   (a) a 95% confidence interval for the mean,

   (b) a 99% confidence interval for the mean.

2.2 Confidence interval for the standard deviation and the variance

In Section 2.1 the standard deviation of a population of bus journey times was estimated from a sample of bus journey times. As was stated, the estimate is subject to uncertainty as, if another sample of the same size were taken, a different estimate would almost certainly result. Provided the data comes from a normal distribution it is possible to estimate the population standard deviation with a confidence interval rather than using a single figure (or point estimate).

To do this you need to use the fact that for a sample of size n from a normal distribution

$$\sum \frac{(x-\bar{x})^2}{\sigma^2}$$

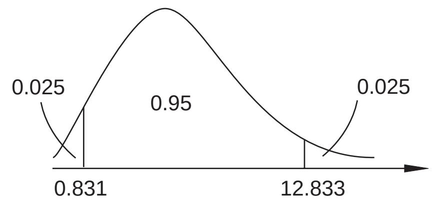
is distributed as

$$\chi^2_{n-1} \quad \text{(chi-squared with } n-1 \text{ degrees of freedom).}$$

$\sum(x-\bar{x})^2$ is equal to $(n-1)\hat{\sigma}^2$ and this is usually the easiest way of calculating it.

In the case of the bus journey times a sample of size 6 gave

$$\hat{\sigma}^2 = 5.09^2 = 25.9\,.$$



If you find the upper and lower 0.025 tails of χ^2, there will be a probability of 0.95 that $\sum\frac{(x-\bar{x})^2}{\sigma^2}$ will lie between them.

A 95% confidence interval for the variance is defined by

$$0.831 < 5\times\frac{25.9}{\sigma^2} < 12.833$$

$$\Rightarrow \quad 0.006415 < \frac{1}{\sigma^2} < 0.0990965$$

$$\Rightarrow \quad 10.09 < \sigma^2 < 155.9$$

and this is a 95% confidence interval for the variance.

Although the variance has a very important role to play in the mathematical theory of statistics, for practical applications the standard deviation is a much more useful statistic. A 95% confidence interval for the standard deviation is found simply by taking the square root of the two limits,

i.e. $\quad 3.2 < \sigma < 12.5$

Note that the interval is not symmetrical about the point estimate which was 5.09. Clearly the interval would not make sense unless it contained the point estimate and this is a useful check on the calculation. However, as the standard deviation cannot be negative but has no upper limit, it would not be expected that the point would lie exactly in the middle.

Is it possible to calculate 100% confidence interval for the standard deviation?

Example

In processing grain in the brewing industry, the percentage extract recovered is measured. A particular brewery introduces a new source of grain and the percentage extract on eleven separate days is as follows:

95.2, 93.1, 93.5, 95.9, 94.0, 92.0, 94.4, 93.2, 95.5, 92.3, 95.4

(a) Regarding the sample as a random sample from a normal population, calculate

(i) a 90% confidence interval for the population variance,

(ii) a 90% confidence interval for the population mean.

(b) The previous source of grain gave daily percentage extract figures which were normally distributed with mean 94.2 and standard deviation 2.5. A high percentage extract is desirable but the brewery manager also requires as little day to day variation as possible. Without further calculation, compare the two sources of grain. (AEB)

Solution

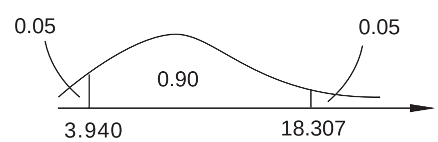
For this data

$$n = 11 \qquad \bar{x} = 94.045 \qquad \hat{\sigma} = 1.34117$$

(a)

(i) 90% confidence interval for variance is given by

$$3.94 < 10 \times \frac{1.34117^2}{\sigma^2} < 18.307$$

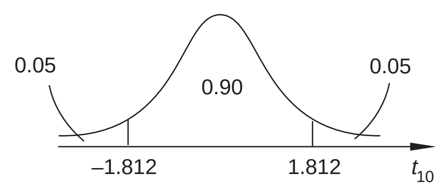


$$\Rightarrow \quad 0.2190 < \frac{1}{\sigma^2} < 1.0178$$

$$\Rightarrow \quad 0.98 < \sigma^2 < 4.57$$

(ii) 90% confidence interval for mean is given by

$$94.045 \pm 1.812 \times \frac{1.34117}{\sqrt{11}}$$



$$\Rightarrow \quad 94.045 \pm 0.733$$

$$\Rightarrow \quad (93.31,\ 94.78)$$

(b) The mean of the previous source of grain was 94.2. This lies in the middle of the confidence interval calculated for the mean of the new source of grain. There is therefore no evidence that the means differ.

The standard deviation of the previous source of grain was 2.5 and hence the variance was $2.5^2 = 6.25$.

This is above the upper limit of the confidence interval for the variance of the new source of grain. This suggests that the new source gives less variability.

Combining these two conclusions suggests that the new source is preferable to the previous source.

Activity 2

Take a sample of size 6 from the data in Activity 1. Calculate an 80% confidence interval for the standard deviation.

Take further samples of size 6 and repeat the calculation at least 20 times. Work in a group, if possible, so that the calculations may be shared.

The data in Activity 1 came from a normal distribution with standard deviation 2. What proportion of your intervals contain the value 2? Is this proportion similar to the proportion you would expect?

Exercise 2B

1. Using the data in Questions 1, 2 and 3 of Exercise 2A, calculate 95% confidence intervals for the population standard deviations.
2. The external diameter, in cm, of a random sample of piston rings produced on a particular machine were

   9.91, 9.89, 10.12, 9.98, 10.09,
   9.81, 10.01, 9.99, 9.86

   Calculate a 95% confidence interval for the standard deviation. Assume normal distribution.

   Do your results support the manufacturer's claim that the standard deviation is 0.06 cm?
3. The vitamin C content of a random sample of 5 lemons was measured. The results in 'mg per 10 g' were

   1.04, 0.95, 0.63, 1.62, 1.11

   Assuming a normal distribution calculate a 95% confidence interval for the standard deviation.

   A greengrocer claimed that the method of determining the vitamin C content was extremely unreliable and that the observed variability was more due to errors in the determination rather than to actual differences between lemons. To check this 7 independent determinations were made of the vitamin C content of the same lemon. The results were as follows

   1.21, 1.22, 1.21, 1.23, 1.24, 1.23, 1.22

   Assuming a normal distribution, calculate a 90% confidence interval for the standard deviation of the determinations. Does your result support the greengrocer's claim?

2.3 Confidence interval for proportions

An insurance company receives a large number of claims for storm damage. A manager wished to estimate the proportion of these claims which were for less than £500. He examined a random sample of 120 claims and found that 18 of them were for less than £500.

Obviously the best estimate of the proportion of all claims which are for less than £500 is $\frac{18}{120} = 0.15$.

To calculate a confidence interval for the proportion we must observe that, *r*, the number of claims for less than £500 in a sample of size *n*, will be an observation from a binomial distribution. This will have mean *np* and variance $np\,(1-p)$. From *Statistics* you know that a binomial distribution can be approximated by a normal distribution if *n* is large and *np* is reasonably large.

In this case $n = 120$ and *p* is estimated by 0.15. Hence *r* can be approximated by a normal distribution with variance $120 \times 0.15 \times 0.85 = 15.3$ and standard deviation $\sqrt{15.3} = 3.912$.

The proportion of claims for less than £500 is estimated by $\frac{r}{n}$.

This has variance

$$\frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \quad .$$

In this case the proportion can be estimated by a normal distribution with variance

$$\frac{0.15 \times 0.85}{120} = 0.0010625$$

and standard deviation 0.03260.

Hence an approximate 95% confidence interval for the proportion is given by

$$0.15 \pm 1.96 \times 0.0326$$

$$\Rightarrow \quad 0.15 \pm 0.064$$

$$\Rightarrow \quad (0.086,\ 0.214).$$

In general, the formula for an approximate confidence interval for a proportion is

$$\boxed{\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

where $\hat{p}$ is an estimate of *p* and *z* is the appropriate value from normal tables to give the required percentage confidence interval.

Note this formula is valid only if the parameters of the binomial distribution make it reasonable to use the normal approximation.

Example

Employees of a firm carrying out motorway maintenance are issued with brightly coloured waterproof jackets. These come in five different sizes numbered 1 to 5. The last 40 jackets issued were of the following sizes

2 3 3 1 3 3 2 4 3 2 5 4 1 2 3 3 2 4 5 3

2 4 4 1 5 3 3 2 3 3 1 3 4 3 3 2 5 1 4 4

(a) (i) Find the proportion in the sample requiring size 3. Assuming the 40 employees can be regarded as a random sample of all employees, calculate an approximate 95% confidence interval for the proportion, p, of all employees requiring size 3.

(ii) Give two reasons why the confidence interval calculated in (i) is approximate rather than exact.

(b) Your estimate of p is $\hat{p}$.

(i) What percentage is associated with the approximate confidence interval $\hat{p} \pm 0.1$?

(ii) How large a sample would be needed to obtain an approximate 95% confidence interval of the form $\hat{p} \pm 0.1$? (AEB)

Solution

(a) (i) There are 15 out of 40 requiring size 3, a proportion of $\frac{15}{40} = 0.375$. An approximate 95% confidence interval is given by

$$0.375 \pm 1.96 \times \sqrt{\frac{0.375 \times 0.625}{40}}$$

$$\Rightarrow \quad 0.375 \pm 0.150$$

$$\Rightarrow \quad (0.225,\ 0.525).$$

(ii) The confidence interval is approximate because an estimate of p is used (the true value is unknown) and because the normal distribution is used as an approximation to the binomial distribution.

(b) (i) Confidence interval is given by

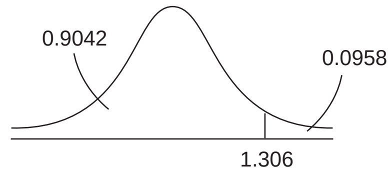
$$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Hence if interval is $\hat{p} \pm 0.1$,

$$0.1 = z\sqrt{\frac{0.375 \times 0.625}{40}}$$

i.e. $z = 1.306$;

this is $(1-2\times 0.096)\times 100 = 81$ per cent confidence interval.



(ii) 95% confidence interval

$$\hat{p} \pm 1.96\sqrt{\frac{0.375\times.625}{n}}.$$

If this is $\hat{p}\pm 0.1$,

$$0.1 = 1.96\sqrt{\frac{0.375\times.625}{n}}$$

$$\Rightarrow \quad \sqrt{n} = 9.4888$$

$$\Rightarrow \quad n = 90.04$$

Sample of size 90 needed.

Exercise 2C

1. When a random sample of 80 climbing ropes were subjected to a strain equivalent to the weight of ten climbers, 12 of them broke. Calculate a 95% confidence interval for the proportion of ropes which would break under this strain.

2. Data from a completed questionnaire were entered into a computer as a series of binary digits (i.e. each digit was 0 or 1). A check on 1000 digits revealed errors in 19 of them. Assuming the probability of an error is the same for each digit entered, calculate a 90% confidence interval for the proportion of digits where an error will be made.

3. A large civil engineering firm issues every new employee with a safety helmet. Five different sizes are available numbered 1 to 5. A random sample of 90 employees required the following sizes

2	4	2	2	2	5	4	5	4	4
4	2	4	3	4	2	3	1	5	4
3	2	3	3	3	4	3	2	4	4
3	4	4	5	3	3	3	2	4	4
2	2	3	2	3	2	3	3	5	4
2	3	4	2	4	3	2	2	3	2
3	4	2	3	4	5	2	3	3	2
4	3	2	2	3	3	3	2	3	4
2	3	2	4	2	3	3	2	2	3

Calculate an approximate 90% confidence interval for the proportion of employees requiring size 2. (AEB)

2.4 Confidence interval for the mean of a Poisson distribution

In an area of moorland, plants of a certain variety are known to be distributed at random at a constant average rate; that is, they are distributed according to the Poisson distribution. A biologist counts the number of plants in a randomly chosen square of area 10 m^2 and finds 142 plants. An approximate confidence interval for the average number of plants in a square of area 10 m^2 can be found by approximating the Poisson distribution by a normal distribution with mean 142 and standard deviation $\sqrt{142}$. This approximation is valid since the mean is large.

In this case a 95% approximate confidence interval is given by

$$142 \pm 1.96\sqrt{142}$$

$$\Rightarrow \quad 142 \pm 23.4$$

$$\Rightarrow \quad (118.6,\ 165.4).$$

In general the formula is

$$m \pm z\sqrt{m}$$

where m is the observed value and z is the appropriate value from normal tables to give the required percentage confidence interval. Remember this is only valid if the mean is reasonably large so that the normal distribution gives a good approximation.

If a confidence interval for the mean number of plants per m^2 was required, the interval calculated for 10 m^2 would be divided by 10. The result would be 14.2 ± 2.34 or $(11.86,\ 16.54)$.

The calculation could have been carried out by finding the mean number of plants observed in 10 areas of 1m^2 and basing the calculation on this. However, there would be no advantage in this as it would give the identical answer.

For example in this case, if 10 separate m^2 had been observed, the mean number of plants in each square would be 14.2 and the distribution would be approximated by normal mean 14.2 standard deviation $\sqrt{14.2}$. Since you now have the mean of 10 observations the calculation for a 95% interval would be

$$14.2 \pm 1.96\frac{\sqrt{14.2}}{\sqrt{10}}$$

$$\Rightarrow \quad 14.2 \pm 2.34, \text{ as before.}$$

It is probably always easiest to work in terms of the total number of events observed and then scale the final answer as required.

Why is a confidence interval calculated for the mean of a Poisson distribution only an approximation?

Exercise 2D

1. Cars pass a point on a motorway during the morning rush hour at random at a constant average rate. An observer counts 212 cars passing during a 5 minute interval. Calculate
   (a) a 95% confidence interval for the mean number of cars passing in a 5 minute interval,
   (b) a 95% confidence interval for the mean number of cars passing in a one minute interval,
   (c) a 90% confidence interval for the mean number of cars passing in a two minute interval,
   (d) a 99% confidence interval for the mean number of cars passing in an hour.
2. The number of times a machine needs resetting on a night shift follows a Poisson distribution. On three randomly selected nights it was reset 9, 5 and 11 times. Calculate a 95% confidence interval for the average number of times it needs resetting per night.
3. The number of a certain type of organism suspended in a liquid follows a Poisson distribution. 10 cc of the liquid are found to contain 35 of the organisms. Calculate
   (a) a 90% confidence interval for the mean number of organisms per 10 cc,
   (b) a 95% confidence interval for the mean number of organisms per cc,
   (c) a 99% confidence interval for the mean number of organisms per 100 cc.

   A further 10 cc of the liquid were examined and found to contain 26 of the organisms. Modify your answers to (a), (b) and (c) to take account of this additional data.

2.5 Miscellaneous Exercises

1. The development engineer of a company making razors records the time it takes him to shave, on seven mornings, using a standard razor made by the company. The times, in seconds, were

   217, 210, 254, 237, 232, 228, 243

   Assuming that this may be regarded as a random sample from a normal distribution, calculate a 95% confidence interval for the mean.
   (AEB)
2. A car insurance company found that the average amount it was paying on bodywork claims was 435 with a standard deviation of 141. The next eight bodywork claims were subjected to extra investigation before payment was agreed. The payments, in pounds, on these claims were

   48, 109, 237, 192, 403, 98, 264, 68.

   (a) Assuming the data can be regarded as a random sample from a normal distribution, calculate a 90% confidence interval for
      (i) the mean payment after extra investigation,
      (ii) the standard deviation of the payments after extra investigation.

   (b) Explain to the manager whether or not your results suggest that the distribution of payments has changed after special investigation and comment on her suggestion that in future all claims over £900 should be subject to special investigation. (AEB)
3. A car manufacturer purchases large quantities of a particular component. Tests have shown that 3% fail to function and that, of those that do function, the mean working life is 2400 hourswith a standard deviation of 650 hours. The manufacturer is particularly concerned with the large variability and the supplier undertakes to improve the design so that the standard deviation is reduced to 300 hours.
   (a) A random sample of 310 of the new components contained 12 which failed to function. Calculate an approximate 95% confidence interval for the proportion which fail to function.
   (b) A random sample of 3 new components tested had working lives of 2730, 3120 and 2300 hours.

Assuming that the claim of a standard deviation of 300 hours is correct and that the lives of the new components follow a normal distribution, calculate

(i) a 90% confidence interval for the mean working life of the components,

(ii) how many components it would be necessary to test to make the width of a 90% confidence interval for the mean just less than 100 hours.

(c) Lives of components commonly follow a distribution that is not normal. If the assumption of normality is invalid in this case, comment briefly on the amount of uncertainty in your answers to (b)(i) and (b)(ii).

(d) Using all the information available, compare the two designs and recommend which one should be used. (AEB)

4. The resistances (in ohms) of a sample from a batch of resistors were

2314, 2456, 2389, 2361, 2360, 2332, 2402.

Assuming that the sample is from a normal distribution,

(a) Calculate a 90% confidence interval for the standard deviation of the batch.

Past experience suggests that the standard deviation, σ, is 35 ohms.

(b) Calculate a 95% confidence interval for the mean resistance of the batch

(i) assuming $\sigma = 35$,

(ii) making no assumption about the standard deviation.

(c) Compare the merits of the confidence intervals calculated in (b). (AEB)

5. Packets of baking powder have a nominal weight of 200 g. The distribution of weights is normal and the standard deviation is 7 g. Average quantity system legislation states that, if the nominal weight is 200 g,

(i) the average weight must be at least 200 g,

(ii) not more than 2.5% of packages may weigh less than 191 g,

(iii) not more than 1 in 1000 packages may weigh less than 182 g.

A random sample of 30 packages had the following weights:

218	207	214	189	211	206	203	217	183	186
219	213	207	214	203	204	195	197	213	212
188	221	217	184	186	216	198	211	216	200

(a) Calculate a 95% confidence interval for the mean weight.

(b) Find the proportion of packets in the sample weighing less than 191 g and use your result to calculate an approximate 95% confidence interval for the proportion of all packets weighing less than 191 g.

(c) Assuming that the mean weight is at the lower limit of the interval calculated in (a), what proportion of packets would weigh less than 182 g?

(d) Discuss the suitability of the packets from the point of view of the average quantity system. A simple adjustment will change the mean weight of future packages. Changing the standard deviation is possible but very expensive. Without carrying out any further calculations, discuss any adjustments you might recommend. (AEB)

6. A car manufacturer introduces a new method of assembling a new component. The old method had a mean assembly time of 42 minutes with a standard deviation of 4 minutes. The manufacturer would like the assembly time to be as short as possible and to have as little variation as possible. He expects the new method to have a smaller mean but to leave the variability unchanged. A random sample of assembly times, in minutes, taken after the new method had become established was

27, 19, 68, 41, 17, 52, 35, 72, 38.

A statistician glanced at the data and said she thought the variability had increased.

(a) Suggest why she said this.

(b) Assuming the data may be regarded as a random sample from a normal distribution, calculate a 95% confidence interval for the standard deviation. Does this confirm the statistician's claim or not?

(c) Calculate a 90% confidence interval for the mean using a method which is appropriate in the light of your answer to (b).

(d) Comment on the suitability of the new process. (AEB)

7. Stud anchors are used in the construction industry. Samples are tested by embedding them in concrete and applying a steadily increasing load until the stud fails.

(a) A sample of 6 tests gave the following maximum loads in kN

27.0, 30.5, 28.0, 23.0, 27.5, 26.5

Assuming a normal distribution for maximum loads, find 95% confidence intervals for

(i) the mean,

(ii) the standard deviation.

(b) If the mean was at the lower end and the standard deviation at the upper end of the confidence intervals calculated in (a), find the value of k which the maximum load would exceed with probability 0.99.

Safety regulations state that the greatest load that may be applied under working conditions is $\frac{(\bar{x} - 2\hat{\sigma})}{3}$ where $\bar{x}$ is the mean and $\hat{\sigma}^2$ is the unbiased estimate of variance calculated from a sample of 6 tests. Calculate this figure for the data above and comment on the adequacy of this regulation in these circumstances.

(AEB)

8. A campaign to combat the economic devastation caused to coalfield communities by pit closures employed a researcher. The campaign organisers wished to know the proportion of redundant miners who were able to find alternative employment within a year of becoming redundant.

(a) The researcher found that the probability of a redundant miner visited at home refusing to answer a questionnaire is 0.2.

What is the probability that on a day when he visits twelve redundant miners at home

(i) 3 or fewer will refuse to answer the questionnaire,

(ii) exactly 3 will refuse to answer the questionnaire,

(iii) at least 10 will agree to answer the questionnaire?

(b) The researcher decided to try a postal survey and as a pilot scheme sent out 70 questionnaires to randomly selected redundant miners. There were 26 completed questionnaires returned. Calculate an approximate 95% confidence interval for the proportion of redundant miners who would return a completed questionnaire.

(c) Of the 26 who replied, 10 had obtained employment. Of all redundant miners who would reply to a questionnaire, a proportion p have obtained emplyment. Approximately how many replies would be necessary to obtain 95% confidence interval of width 0.1 for p?

(d) Using the results of (b) and (c) estimate approximately how many letters should be sent out to give a high probability of obtaining sufficient replies to calculate the confidence interval in (c). Explain your answer. (AEB)

9. It is known that repeated weighings of the same object on a particular chemical balance give readings which are normally distributed with mean equal to the mass of the object. Past esperience suggests that the standard deviation, σ, is 0.25 mg. Seven repeated weighings gave the following readings (mg).

19.3, 19.5, 19.1, 19.0, 19.8, 19.7, 19.4

(a) Use the data to calculate a 95% confidence interval for σ.

(b) Calculate a 95% confidence interval for the mass of the object assuming $\sigma = 25$ mg.

(c) Calculate 95% confidence interval for the mass of the object, making no assumption about σ, and using only data from the sample.

(d) Give two reasons for preferring the confidence interval calculated in (b) to that calculated in (c).