SECTION 4

ESTIMATION

4.1 INTRODUCTION

In general, we wish to estimate population parameters using suitable sample statistics. Here we will consider:

- using the sample mean $\bar{x}$ to estimate the population mean μ
- using the sample S.D. s to estimate the population S.D. σ
- using the sample proportion $\hat{p}$ to estimate the population proportion p.

A **point estimate** is a single value used to estimate the population parameter. (For example, a sample of four batteries is found to have a mean lifetime of 31.0 hours. Thus 31.0 hours is a point estimate of the mean lifetime of the population of batteries.)

An **interval estimate** gives a range of values which includes the unknown population value with a preassigned probability. Such an interval is a **confidence interval**, the endpoints of the interval are **confidence limits** and the associated probability is the **confidence level**. (For example, if we had found that the 95% confidence interval for the population mean lifetime of the above batteries was 31.0h ± 3h, then we are 95% sure that the range of values 28h to 34h contains the true mean lifetime.)

Exercise 1

Compared to a 95% interval, would you expect the 99% interval to be:

(a) narrower;

(b) wider;

(b) same width?

[Answer: Wider}

In this section we will be concerned with confidence intervals for a mean and for a proportion. The underlying principles are illustrated for the case of estimating a mean.

4.2 SAMPLING DISTRIBUTIONS

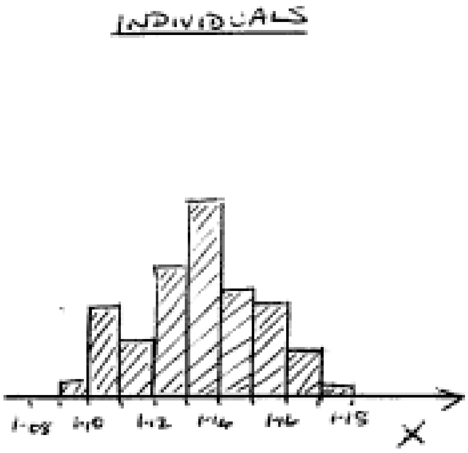
To find how wide an interval should be, we need to know how sample means vary from sample to sample. Consider the following data that consists of 10 samples each of size 5:

Shaft Diameter (inches)

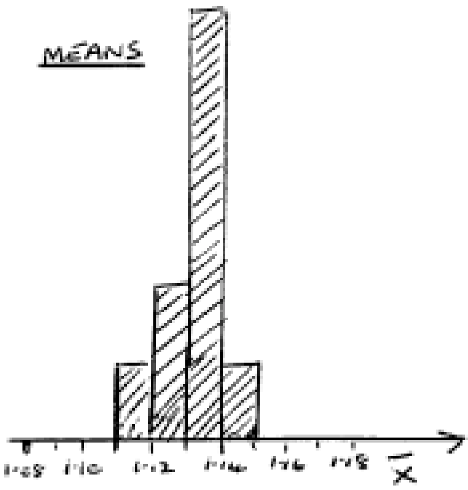
Sample

	1	2	3	4	5	6	7	8	9	10
	1.130	1.141	1.101	1.157	1.136	1.131	1.136	1.150	1.141	1.103
	2.209	1.168	1.092	1.105	1.157	1.121	1.109	1.110	1.129	1.138
	1.101	1.146	1.172	1.169	1.142	1.113	1.126	1.146	1.134	1.130
	1.131	1.143	1.135	1.121	1.130	1.143	1.152	1.112	1.136	1.168
	1.125	1.125	1.133	1.133	1.122	1.153	1.116	1.151	1.129	1.129
($\bar{X}$)	1.119	1.145	1.127	1.137	1.137	1.132	1.126	1.134	1.134	1.134

Shaft Diameter (inches)	Tally	Frequency	%
[1.08-1.09)			(0%)
[1.09-1.10)	|	1	(2%)
[1.10-1.11)	卌 |	6	(12%)
[1.11-1.12)	||||	4	(8%)
[1.12-1.13)	卌 ||||	9	(18%)
[1.13-1.14)	卌 卌 |||	13	(26%)
[1.14-1.15)	卌 ||	7	(14%)
[1.15-1.16)	卌 |	6	(12%)
[1.16-1.17)	|||	3	(6%)
[1.17-1.18)	|	1	(2%)
TOTAL		50	(100%)



Mean Shaft Diameter (inches)	Tally	Frequency	%
[1.08-1.09)			
[1.09-1.10)			
[1.10-1.11)			
[1.11-1.12)	|	1	(10%)
[1.12-1.13)	||	2	(20%)
[1.13-1.14)	卌 |	6	(60%)
[1.14-1.15)	|	1	(10%)
[1.15-1.16)			
[1.16-1.17)			
[1.17-1.18)			
TOTAL		10	(100%)



We can, in fact, establish theoretically that

- $\overline{X}$'s *have the* <u>*same mean*</u> *as individual values*

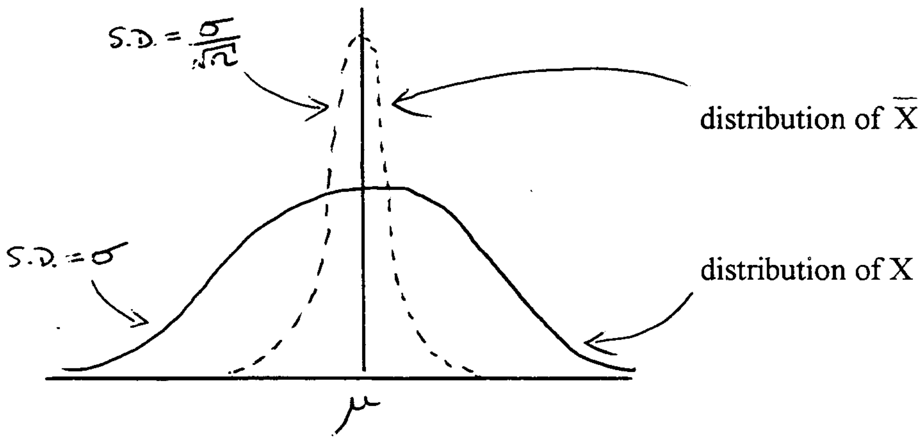
- *The standard deviation of the* $\overline{X}$'s *is equal to the standard deviation of the individuals divided by* $\sqrt{n}$

$$\text{standard deviation of } \overline{X} = \frac{\text{standard deviation of individual}}{\sqrt{n}}$$

$$= \boxed{\frac{\sigma}{\sqrt{n}}}$$

This is called the ***standard error*** *of the mean.*

- <u>*The distribution of* $\overline{X}$ *is normal*</u> *if the* X's *are normal and is approximately normal if not (provided the sample is large enough - roughly* $n \geq 30$*).*



Thus, if we want to estimate μ, an observation from the distribution of $\overline{X}$ is more likely to be close to the true value μ than an observation from the distribution of X. Also since the spread of the distribution is $\sigma / \sqrt{n}$, the larger the sample size n, the smaller the spread and the more tightly packed the distribution will be around μ. This justifies the intuitively sensible view that if you want to measure something, you'll get a more accurate idea of the true value if you average over several readings - and the more readings you take, the more accurate the estimate will be.

4.3 CONFIDENCE INTERVALS FOR A MEAN

4.3.1 σ known

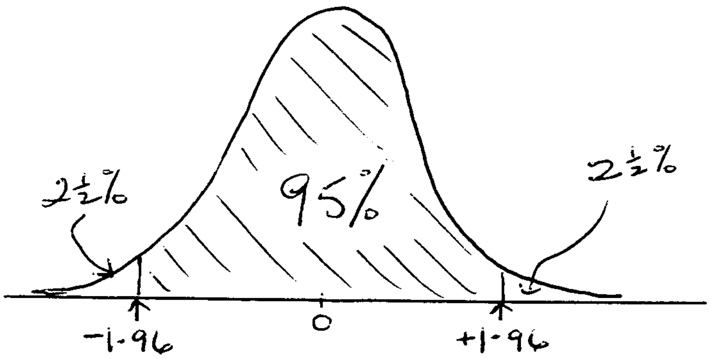
In the section above, it was stated that the sampling distribution of the sample mean $\overline{X}$ was normal with mean μ and standard error $\sigma / \sqrt{n}$. (Exactly, if the original population is normal; approximately, if the sample is large, n ≥ 30.)

Thus

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

has a standard. normal distribution i.e. a normal distribution with mean 0 and standard deviation 1.

From standard normal tables, we know that 95% of the values of such a variable are in the range –1.96 to +1.96.



For any normal distribution, 95% of the values are within 1.96 standard deviations of the mean - i.e. in the range

$$(\text{mean}) \pm 1.96 \times (\text{standard deviation})$$

Thus, since $\overline{X}$ has standard deviation (or standard error) $\sigma / \sqrt{n}$, we can define the 95% confidence interval for μ as

$$\boxed{\overline{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}}$$

Exercise 2

Write down the 90% and 99% confidence intervals for μ.

The 90% CI for μ is $\bar{x} \pm 1.6449\frac{\sigma}{\sqrt{n}}$

The 99% CI for μ is $\bar{x} \pm 2.5758\frac{\sigma}{\sqrt{n}}$

In general, if the population standard deviation σ is known, the 100(1 – α)% confidence interval for μ is

$$\boxed{\bar{x} \pm z\frac{\sigma}{\sqrt{n}}}$$

where Z is the standard normal value with an area in the right hand tail of α/2.

Exercise 3

The supplier of an electronic probe states that the standard deviation of measurements given by the probe is 0.6 microns. The probe is used to measure the radial play in a bearing and 36 measurements of a particular bearing give a mean reading of 8 microns.

Then the 95% confidence interval for the true RP of this bearing is

$$8 \pm 1.96\left(\frac{0.6}{\sqrt{36}}\right)$$

= 8 ± 0.20 microns

(or 7.80 to 8.20 microns).

Interpretation

We are 95% sure that the range 7.80 microns to 8.20 microns includes the true radial play of this bearing.

Exercise 4

Find the 90% and 99% confidence intervals for the true RP of the bearing in Exercise 3.

The 90% CI for μ is $\bar{x} \pm 1.6449 \frac{\sigma}{\sqrt{n}}$

$$= 8 \pm 1.6449 \frac{0.6}{\sqrt{36}}$$

$$= 8 \pm 0.16 \text{ microns}$$

or <u>7.84 microns</u> to <u>8.16 microns</u>

The 99% CI for μ is $\bar{x} \pm 2.5758 \frac{\sigma}{\sqrt{n}}$

$$= 8 \pm 2.5758 \frac{0.6}{\sqrt{36}}$$

$$= 8 \pm 0.26 \text{ microns}$$

or 7<u>.74 microns</u> to <u>8.26 microns</u>

4.3.2 σ not known

It will usually be the case that the population standard deviation σ is <u>not</u> known. It can then be estimated by the sample standard deviation

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

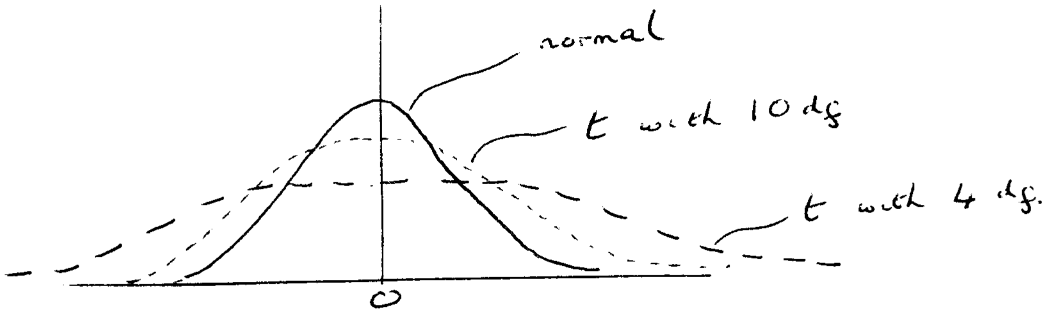
Provided the sample size is **large** (n at least 30), s can then be used in place of σ in the formulae of Section 4.3.1 to give approximate confidence intervals.

However, if only a **small** sample is available ($n < 30$), an adjustment should be made using the fact that

$$\frac{\bar{x} - \mu}{s/\sqrt{n}}$$

does not have a standard normal distribution but has **Student's *t* distribution with (n-1) degrees of freedom.**

Like the standard normal, *t* distributions are bell-shaped and symmetrical about zero but they are 'flatter' than the normal. Percentage points of the *t* distributions are well tabulated.



Exercise 5

From *t* tables, find the value of *t* with the following right hand tail areas (α) and degrees of freedom (ν).

(a)	$\alpha = 0.05$	$\nu = 12$
(b)	$\alpha = 0.05$	$\nu = 4$
(c)	$\alpha = 0.025$	$\nu = 28$
(d)	$\alpha = 0.01$	$\nu = 28$
(e)	$\alpha = 0.005$	$\nu = 50$

[Answers: (a) 1.782 (b) 2.132 (c) 2.048 (d) 2.467 (e) 2.682 (by interpolation)]

(Note that the values in the bottom row where $\nu = \infty$ imply that *t*-values are then the same as the values from a normal distribution.)

From the above, we can define the 100(1 – α)% confidence interval for μ when σ is unknown as:

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

where t_{n-1} is the *t* value with (n-1) degrees of freedom and a right hand tail area of $\alpha/2$.

Note

This interval again assumes that the original distribution is normal.

4.3.3 **Which Interval to Use?**

The CI for μ is of the form $\bar{x} \pm (somevalue) \times \frac{SD}{\sqrt{n}}$

Where *somevalue* is either (z or t) and SD is either (σ or s).

Which to use (if any) depends on

- whether the original distribution of X is normal (at least approximately)
- whether σ is known
- the sample size (n).

The following table is one way of summarising the choice:

	X is normal		*X is not normal*	
Population SD σ is known	$n \geq 30$	**z and σ**	$n \geq 30$	**z and σ**
	$n < 30$	**z and σ**	$n < 30$	–
σ is not known so sample SD s used	$n \geq 30$	**z and *s***	$n \geq 30$	**z and *s***
	$n < 30$	**t and *s***	$n < 30$	–

($n \geq 30$) : sample is 'large'
($n < 30$) : sample is 'small'

Note that the intervals described here cannot be used if X is not normal and the sample is small.

In summary, use *t*-values if the sample is small and σ is not known.

Exercise 6

25 readings of the power output from an electronic device are found to have a mean of 54.62 MHz and a standard deviation of 5.34 MHz. The 95% confidence interval for the true power output for this device is then:

$$\bar{x} \pm t_{24} \frac{s}{\sqrt{n}}$$

$$= 54.62 \pm 2.064 \times \left(\frac{5.34}{\sqrt{25}} \right)$$

$$= 54.62 \pm 2.204 \text{ MHz}$$

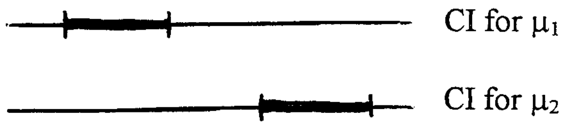
(i.e. 52.416 MHz to 56.824 MHz)

(Note: since the sample size is 'small' (n < 30) we need to assume that power output has approximately a normal distribution.)

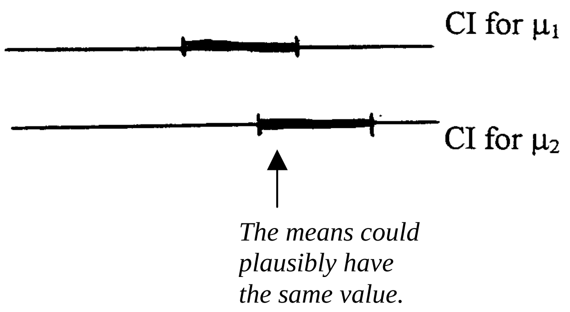
4.3.4 Comparing Means

If we wish to compare the means of 2 populations, one way of doing it is to take a sample from each population and construct 2 confidence intervals.

If the intervals <u>do not</u> overlap then this suggests that the <u>true means really are different</u>.



If the intervals <u>do</u> overlap then the data are <u>consistent with the true means being the same</u>.
In other words, we have no evidence that they really do differ.



Note

Means from more than 2 populations can be compared in a similar fashion. This is rather a crude approach however. There are more sophisticated ways of handling the problem.

4.4 CONFIDENCE INTERVALS FOR A PROPORTION

We now wish to estimate a population proportion p using a proportion $\hat{p}$ from a sample of size n.

For example, if 180 out of 300 people interviewed say they support some new European legislation, what does that tell us about the proportion in the whole population who support it?

It can be shown that if the sample is large enough (typically $n \geq 30$), then a sample proportion $\hat{p}$ has approximately a normal distribution centred around the population proportion p. Furthermore, the standard error of this distribution can be estimated by

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

Thus, by analogy to the construction of a confidence interval for a mean, we can define the 95% confidence interval for p as:

$$\boxed{\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Similarly, 90% and 99% CI's can be obtained by replacing 1.96 with 1.64 and 2.58 respectively.

To answer the above question, if $\hat{p} = \frac{180}{300} = 0.60$ then the 95% CI for p is

$$0.60 \pm 1.96\sqrt{\frac{(0.60)(0.40)}{300}}$$

$$= 0.60 \pm 0.055$$

(or 60% ± 5.5%)

Interpretation

We are 95% confident that the range 54.5%-65.5% contains the percentage of the whole population who support the legislation.

Note that this interval does not fall below 50%, so we are reasonably confident that a majority of the population support the legislation.

Exercise 7

A series of tests was carried out to evaluate the breaking strength of steel beams. One experiment involved testing 50 beams to destruction by applying a load of 4000Nmm^{-2} to each and observing how many of them buckled. It turned out that 18 of the 50 buckled under that load. Construct a 90% confidence interval for the proportion of all beams of this type that would buckle under such a load.

$$\hat{p} = \frac{18}{50} = 0.36$$

90% CI is

$$\hat{p} \pm 1.64\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$= 0.36 \pm 1.64\sqrt{\frac{(0.36)(0.64)}{50}}$$

$$= 0.36 \pm 0.111 \qquad \text{(or 36\% ± 11.1\%)}$$

i.e. $0.249 \rightarrow 0.471$ (or 24.9% → 47.1%)

Note that this is a very wide interval - mainly because the sample is so small.

4.5 CONFIDENCE INTERVALS AND SAMPLE SIZE

4.5.1 Estimating Means

An important question is: 'how large should the sample be to get an estimate with some required accuracy?'

For example, in Exercise 3, 36 readings enabled the true measurement to be estimated as 8 ± 0.20 microns with 95% confidence. How many readings should be taken if it is required to estimate to within 0.10 microns with 95% confidence (rather than to within 0.20 microns)?

Essentially we now need to work backwards. We require

$$\bar{x} \pm 1.96 \times \frac{\sigma}{\sqrt{n}} = \bar{x} \pm 0.10$$

In this case, σ is known to be 0.6, so we require n (the sample size) so that:

$$1.96 \times \frac{0.6}{\sqrt{n}} = 0.10$$

This equation can now be rearranged and solved for n.

i.e. $1.96 \times 0.6 = 0.10\sqrt{n}$

or $\sqrt{n} = 11.76$

Then $n = (11.76)^2$

$= 138.3$

So at least 139 readings would need to be taken to estimate the measurement with that degree of accuracy.

4.5.2 Estimating Proportions

In Exercise 7, a sample of 50 steel beams enabled the proportion of the population of beams that would buckle under the given load estimated to within 0.111 (with 90% confidence). In that study, the point estimate was 0.36 (or 36%).

How many beams would need to be tested in order to estimate the proportion to within 0.03?

For 90% confidence, using the value of $\hat{p} = 0.36$ obtained from the initial tests, we require n so that:

$$1.64\sqrt{\frac{(0.36)(0.64)}{n}} = 0.03$$

i.e. $$1.64\sqrt{(0.36)(0.64)} = 0.03\sqrt{n}$$

$$0.7872 = 0.03\sqrt{n}$$

Thus $$\sqrt{n} = \frac{0.7872}{0.03} = 26.24$$

Then $$n = (26.24)^2$$

$$= 688.5$$

i.e. At least 689 beams should be tested.

Notes: Using the more accurate value of z = 1.6449 instead of 1.64 gives $n \geq 693$. *In practice, for either value of z, we would probably round up to 700.)*

<u>Exercise 8</u>

How many beams should be tested if it is required to estimate the proportion to within 0.05 with 99% confidence?

We require n so that $2.58\sqrt{\frac{(0.36)(0.64)}{n}} = 0.05$

Rearranging this gives $n = \left[\frac{2.58\sqrt{(0.36)(0.64)}}{0.05}\right]^2$

$$= 613.5$$

Thus **at least 614** should be tested.

(Note If the more accurate z = 2.5758 is used, we would get $n \geq 612.$)

4.5.3 **Preliminary Information**

In both cases above, some information is needed before the sample size calculation can be carried out.

In the case of estimating a mean, an estimate of σ is required. This would often be obtained from a small sample. The resulting value of n would then be interpreted liberally.

In the case of estimating a proportion, either:

- estimate p from a pilot study or small experiment as in the above example
- use a value of $\hat{p} = 0.5$ which is the most pessimistic case (i.e. results in the largest sample size). This will at least give an upper limit to the sample size required.

e.g. in the example of 4.5.2, if $\hat{p} = 0.5$ is used instead of the value obtained from the initial test ($\hat{p} = 0.36$), we would require n so that:

$$1.64\sqrt{\frac{(0.5)(0.5)}{n}} = 0.03$$

which leads to n = 747.1. So with this 'pessimistic' estimate of p, at least 748 beams are required, compared to the 689 resulting from using the estimate $\hat{p} = 0.36$.

Note

If this estimate were required with 95% confidence then, replacing 1.64 with 1.96 in the above equation leads to a value of n of at least 1068. Along similar lines, market research/opinion poll results are often stated loosely to be *'accurate to within 3 points'* – journalese for *'the 95% confidence interval for the proportion is ± 0.03'* This calculation explains why so many such surveys involve about 1000 respondents.