Chapter 6

Inference for Normal Populations

6.1 Introduction and Basic Concepts

Consider an experiment to investigate the existence of extra-sensory perception (ESP). One subject is in one room and turns cards face up at pre-assigned times from a well-shuffled pack containing cards in equal proportions with the three symbols 'triangle', 'square', and 'three wavy lines'. The other subject is in another room with no contact whatsoever between the two rooms but tries to say the symbols on the cards as they are turned face up. If the two subjects have ESP the success rate should be in considerable excess of one third.
Suppose 4 successes and 1 failure are observed in 5 attempts. Does this provide evidence for ESP? We analyse this problem by answering the following question: If the subject had merely guessed (rather than having ESP),
what is the probability he/she would get at least 4 successes in 5 attempts? Clearly the smaller this probability is, the more evidence there is in favour of ESP and against the hypothesis of 'just guessing'.

The required probability is easily derived as we can assume from the experimental set-up that the five attempts are independent with the same probability of success ~. The number of successes has a Binomial distribution with n=5 and $p = \frac{1}{3}$ so

$$P(4 \text{ or more successes}) = \left(\frac{1}{5}\right)^5 + 5\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^1 = \frac{1+10}{243} = 0.0453$$

This is the probability of doing at least as well as what has actually been observed by just guessing. As it is quite small the result provides *some* evidence against the 'guessing' hypothesis in favour of ESP; that is, because this probability is small either the 'guessing' hypothesis is true and a 'rare' event (4.5% chance) has occurred, or in fact the hypothesis is not true and p (the population parameter of interest) is greater than one third.

Suppose there had been only 3 correct choices out of 4 attempts. Then the probability of doing at least as well by just guessing is now (using the Binomial with $n = 4$ and $p = \frac{1}{4}$)

$$4 \times \left(\frac{1}{3}\right)^3\left(\frac{2}{3}\right)^1 + \left(\frac{1}{3}\right)^4 = \frac{8+1}{81} = \frac{1}{9}$$

So there is a moderate chance (11.1%) of doing at least as well by just guessing. Thus there is *no* evidence in favour of ESP.
The above simple example has all the elements of the significance testing of a hypothesis yet for a Binomial rather than Normal population. We now give our final set of basic definitions for use in this chapter. We must again emphasise that it is *not* important that you learn these by rote. What *is* important is the practical use of them in the analysis of data. You should, however, be

able to use your *own* words to explain these concepts to your non-statistical colleagues and to write brief notes with illustrative examples.

Definition 1 The *null hypothesis* H_0 is a statement about the value of the parameter of interest. A simple null hypothesis specifies the population distribution exactly. We examine the data to see whether they support 'or provide evidence against the null hypothesis H_0. In the ESP experiment the simple null hypothesis is that the subject is guessing randomly, that is, a Binomial distribution with $n = 5$ and $p = 1/3$. A more general null hypothesis is that the population mean takes some specified value: when the population distribution is Normal such a null hypothesis is simple if the variance is known.

The *alternative hypothesis* H_1 describes only the possibilities (there may be many) that we are prepared to consider if H_0 is *not* true.

Definition 2 The *test statistic* for H_0 versus H_1 is a random variable with known (or approximately known) distribution assuming H_0 to be true 'under H_0'. The observed value of the test statistic can indicate departures from H_0 in favour of H_1.

For ESP the number of correct 'guesses' is a test statistic with a Binomial distribution under H_0. In sampling from a Normal population with *unknown* variance a i-statistic is used for testing any specified value of the population mean; and both large positive and negative values of it indicate departures from H_0.

Definition 3 The *P-value* gives he probability of, under H_0, observing a value of the test statistic at least as extreme as the value actually observed, where extremities indicate departures from H_0 in favour of H_1.

In the ESP example, 4 correct out of 5 guesses gives a P-value of 0.0453, which is moderate evidence against H_0 showing the need for further study.

To interpret P-values in a consistent way, we adopt a convention which gives the following interpretations:
$P > 0.1$ very weak or no evidence against the null hypothesis
$0.05 < P < 0.1$ slight or weak evidence
$0.01 < P < 0.05$ moderate evidence
$0.001 < P < 0.01$ strong evidence
$P < 0.001$ very strong or overwhelming evidence,
However the exact interpretation and appropriate action to be taken must obviously vary according to the problem at hand.

Definition 4: The *standard error* of an estimate (of. a parameter) is the estimated standard deviation of the sampling distribution.
Example: $s/\sqrt{n}$ is the standard error of the sample mean.

Definition 5 : A $100(1-\alpha)\%$ *confidence interval* (C.I.) for an unknown parameter is a random interval which contains the true value of the parameter with probability $1-\alpha$, that is $100(1^{-}$

α)% of all samples will give confidence intervals including the parameter and the remaining 100 α% will 'miss it'. Many confidence intervals have the form of an estimate plus or minus a multiple of its standard error, the multiple being an appropriate percentage point of the Student's t-distribution with $n - 1$ degrees of freedom ($n - r$ in general, if r is the number of independent parameters to be estimated).

The above definitions, like the earlier ones, give a structural framework for this course.

6.2 Inference on a single sample

We assume throughout this section that we have a single random sample from a Normal population and that this assumption has been checked in the light of the observed data by means of a Normal Probability Plot.

6.2.1 Significance Tests of Hypotheses

The following can be thought of as a formal procedure for this important branch of statistical inference:

Assumptions: These should be stated clearly together with graphical or other evidence supporting them. For example the assumption that a set of data is a random sample from a Normal population should be accompanied by a Normal Probability Plot which should show an approximate straight line. In MINITAB this can be performed using the commands NSCORES and then PLOT — see the Handbook .

Null and Alternative Hypotheses A clear statement of both should be given in terms of the population parameter of interest, together with a short verbal interpretation.

Test Statistics: The formula in terms of sample statistics such as mean and standard deviation should be stated with the (sampling) distribution under the null hypothesis. Then the observed value of the test statistic should be calculated to at least three significant figures.

P-value: This should be calculated, usually from tables, to an accuracy of *no more and no less* than *three* significant figures.

Assess evidence: The P-value should be used to form a verbal statement or conclusion regarding the truth or otherwise of the *null* hypothesis. Finally a verbal interpretation of this conclusion should be given for the non-statistician.

Depending on the conclusion reached (if any) the investigator may wish to quote a *confidence interval* for the parameter at the desired level of 90%, 95% or 99%, say. See Section 3.2.2 below.

Example 4 : Articles produced by a manufacturer should have mean length 4 cm. and standard deviation 0.02cm. A test sample of size 10 from a large batch of production has $\bar{x} = 4.01$. Is there evidence that the unknown mean length μ, say, of articles in the batch is unsatisfactory?

The Null hypothesis is $H_0: \mu = 4$ (batch satisfactory) to be tested against the alternative $H_1: \mu \neq 4$ (batch unsatisfactory).

We need a test statistic whose distribution is known under the null hypothesis i.e. assuming H_0 to be true. We know that in general for random samples from a Normal population

$$\overline{X} \sim N(\mu, \frac{\sigma.^2}{n})$$

so under H_0

$$\overline{X} \sim N(4, \frac{(0.02)^2}{10})$$

$$Z = \frac{\overline{X} - 4}{0.02/\sqrt{10}} \sim N(0,1)$$

is standard Normal . Large values of Z (either positive or negative) indicate departures from H_0 in favour of H_1 and the observed value of Z is

$$Z = \frac{4.01 - 4}{0.02/\sqrt{10}} = 1.58$$

so the probability of observing a value of Z at least as extreme as this (the P-value) is

$$P(Z > 1.58) + P(Z < —1.58) = 2 \text{ x } P(Z > 1.58) = 0.1141,$$

'using the symmetry of the Normal distribution.

Thus there is a 11.4% chance of observing this sample result or worse even if the batch is satisfactory. We therefore conclude that there is *no* evidence against the null hypothesis.

Note that this was a 'two-sided' or 'two-tailed' test as the alternative hypothesis is 'two-sided', namely $\mu \neq 4$. Suppose, however, there was a legal requirement of a *maximum* mean length of 4 cm. Then we would not be concerned with the possibility that μ <4. and instead test $H_0: \mu = 4$ versus $H_1: \mu > 4$. We would ask whether there was sufficient evidence in the data to make us worry about failing the requirement, and the test statistic and observed value would be the same as before. Only large *positive* and not negative values of Z would indicate departures in favour of H_1 so the P-value is just $P(Z > 1.58) = 0.057$. Now we have *slight* evidence against H_0 in favour of.H_1 i.e. slight evidence that the batch may fail to meet the legal requirement. This is called a 'one-tailed' or 'one-sided' test as the alternative hypothesis is "one-sided', namely $\mu > 4$.

However, the assumption that the population variance σ^2 is known is often unrealistic:

Example 5: A random sample of 5 men had a mean height $\bar{x}$ of 70 inches and a sample

standard deviation *s* of 2 inches. Is there any evidence in these data against the (null) hypothesis that the mean μ of the population is 67 inches? To test $H_0 : \mu =$ '67 versus $H_1 : \mu \neq 67$ we need a test statistic whose distribution is known' under H_0. Such a statistic is

$$T = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t_0 \quad \text{under } H_o$$

that is, Student's 't' with 4 degrees of freedom. The observed value of *T* is 3.35 so to calculate *P* we must refer this value to percentage points of the t-distribution with 4 degrees of freedom (d.o.f.) Now $t_4(0.025) = 2.776$ and $t_4(0.01) = 3.747$ lie on either side of our observed value, so using the symmetry again we have

$0.02 < P < 0.05$.

Alternatively we can use the 2a-values on the second row of Table 3 to arrive at the same answer. We cannot therefore say exactly what the probability of obtaining a value at least as extreme as the one observed is, but we can specify it within a suitable range and this is sufficient to enable us to conclude that there is *moderate* evidence against the null hypothesis. So even this small sample provides evidence. Note that with a one-sided alternative $H_1 : p > 67$ we would have obtained $0.025 > P > 0.01$ which is still classified as 'moderate' evidence.

We may occasionally wish to test a hypothesis about the population variance perhaps as part of checking the assumptions behind a test of the population *mean* based on the Normal distribution (Example 4).

Example 6 :A sample of size $n = 11$ gives $s^2 = 1.5$. Does this provide evidence against $H_0 : \sigma^2 = 1$ versus $H_1 : \sigma^2 > 1$? Here we use the result (2.5) that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2$ ' the chi-square distribution with n – 1 d.o.f. (the same as for Student's 't'.) So a suitable test statistic is $V = \frac{10s^2}{1}$ and we refer

the observed value of 15 to percentage points of the χ^2_{10} distribution given in Table 5. The P-value is $P(V > 15)$ because large values of *V* indicate departures from H_0 in favour of H_1. Reading the values again from either side of the observed value we get

$\chi^2_{10}(0.25) = 12.5489$, $\chi^2_{10}(0.10) = 15.9872 \Rightarrow \quad 0.10 < P < 0.25$.

Hence there is *no* evidence against H_0 and thus no need to doubt our assumption.

The test statistic *V* would be the same if the alternative hypothesis had been $H_1 : \sigma^2 < 1$. However, suppose now $s^2 = 0.5$. Is there evidence? Now *small* values of *V* indicate departures and we look at the *left* rather than the right hand tail of the distribution.

$\chi^2_{10}(0.90) = 4.865 \qquad \chi^2_{10}(0.75) = 6.737$

so $P = P(V < 5)$ lies (just) in the range 0.10 to 0.25, and there is still no evidence against our assumption.

Finally for a two-sided test on the variance with alternative hypothesis $H_1 : \sigma^2 = 1$ we *double* the appropriate one-sided P-value. Thus for both $s^2 = 1.5$ and $s^2 = 0.5$, we have $0.2 < P < 0.5$.

6.2.2 Confidence Intervals

Often we may be asked to *estimate* the population mean p rather than test a hypothesis about it. Or we may have performed a test and found evidence against the null hypothesis casting doubt on our original hypothesised value. We can (and indeed must) give an estimate of uncertainty along with our best estimate of *p,* which is ~, the sample mean.
Using (2.3) we can say that *whatever the value of* μ

$$P[-1.96 \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le 1.96] = 0.95,$$

cross multiplying we get

$$P[-1.96\sigma/\sqrt{n} \le \overline{X} - \mu \le 1.96\sigma/\sqrt{n}] = 0.95$$

Subtracting $\overline{X}$ gives

$$P[-\overline{X} - 1.96\sigma/\sqrt{n} \le -\mu \le -\overline{X} + 1.96\sigma/\sqrt{n}] = 0.95$$

and finally multiplying by — 1 gives

,

$$P[\overline{X} + 1.96\sigma/\sqrt{n} \ge \mu \le \overline{X} - 1.96\sigma/\sqrt{n}] = 0.95$$

and this is true whatever the value of p, so we can say that the random interval $(\overline{X} - 1.96\sigma/\sqrt{n},\ \overline{X} + 1.96\sigma/\sqrt{n})$ has a probability of 0.95 of *containing* or *covering* the value of μ*;* that is, 95% of all samples will give intervals (calculated according to this formula) which contain the true value of the population mean. This interval is called a *95% confidence interval for* p. Note that there is no guarantee that any *specific* sample contains μ with 95% probability (and indeed this statement is meaningless as *p* is *not* a random variable): our sample may be one of the 'unlucky 5%'.

Example 4 revisited: The above argument can be used for a 95% confidence interval as we are assuming that the population variance a^2 is known. Thus

$$(\overline{X} - 1.96 \times 0.02/\sqrt{10},\ \ \overline{X} + 1.96 \times 0.02/\sqrt{10})$$

is a 95% confidence interval for p and substituting the observed value $\bar{x}$ = 4.01 we obtain (3.9976, 4.0224).

Example 5 revisited : We can use (2.4) in a similar argument when a^2 is unknown to obtain

$$(\overline{X} - 2.776s/\sqrt{5}, \overline{X} + 2.776s/\sqrt{5})$$

which contains μ with probability 0.95 as $t_4(0.025) = 2.776$ from Table 2, there being only four

degrees of freedom with a sample of size 5. If we substitute the observed values of these statistics the endpoints of this interval become $70 \pm 2.776 \times 2 / \sqrt{5} \Rightarrow (67.517, 72.483)$.

In general a 100(1 — α)% confidence interval for μ is given by

$$\bar{x} \pm \Phi^{-1}(\alpha/2)\frac{\sigma}{\sqrt{n}}$$

where $\Phi^{-1}(\alpha/2)$ denotes an upper percentage point of the standard Normal distribution (Table 1), when σ^2 is known, and given by

$$\bar{x} \pm t^{-1}(\alpha/2)\frac{s}{\sqrt{n}}$$

when σ^2 is unknown.

A confidence interval for σ^2 is

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1}(\alpha/2)}, \frac{(n-1)s^2}{\chi^2_{n-1}(1-\alpha/2)}\right)$$

where we' use both lower and upper percentage points of the chi-square distribution in Table 3.

As an exercise, try substituting $s^2 = 1.5$ (Example 6) for a 90% confidence interval ($\alpha = 0.1$) and see whether the interval contains the original hypothesised value of 1 (or not). How do you interpret your result? Can you conjecture a connection between confidence intervals and hypothesis tests? We will return to this point at the end of this chapter.

6.3 Comparison of two population means

This is perhaps one of the more important topics in the course. The extension to testing more than two means requires more sophisticated techniques, one of which is introduced in Chapter 6 (the Analysis of Variance).

6.3.1 Paired or Independent Samples

There are two seemingly similar but in fact very different data structures that can be encountered here. It is very important that you learn to recognise their differences from a verbal description. Otherwise an inappropriate test may be applied.
The data may consist of

Two independent random samples from two possibly different populations, or

A single random sample of *pairs* of measurements , which could arise either from a random sample of individuals on each of which two possibly similar variables have been measured, or in for example a *matched pairs study* with a random sample of pairs of similar individuals on which the *same* variable was measured.

In the 'cholesterol' data of Example 8 and available in MINITAB as the saved worksheet CHOLEST.MTW, there are *both* data structures. Twenty eight heart-attack patients were measured for cholesterol level 2 days, 4 days and 14 days after the attack and cholesterol level was also measured for a *control* group of 30 patients. In comparing (the population means of) any measurement on the 28 patients with the control group we have two independent samples, assumed to be random, but for the comparison (over time) of any two measurements on the 28 patients, the 'paired' structure clearly applies.
In general, we are testing the *same* hypothesis about two population means denoted here μ_A and μ_B

$H_0: \mu_A = \mu_B$ versus $H_1: \mu_A \neq \mu_B$

or earlier one-sided alternatives

6.3.2 Independent Samples

Notation

	Population *A*	Population *B*	
Population Mean	μ_A	μ_B	unknown
Population Variance	σ^2_A	σ^2_B	unknown
Population Distribution	Normal	Normal	(transformed)
Sample Size	n_A	n_B	both ≥ 2
Sample Mean	$\bar{x}_A$	$\bar{x}_B$	
Sample Variance	s^2_A	s^2_B	

Example 7 Two random samples were independently drawn from two populations, *A* and *B.* Is there evidence in the following data to indicate a difference in the population means?

Sample	*A*	*B*
Size	6	5
$\sum_{i-1}^{n} x_i$	297	322
$\sum_{i=1}^{n} x_i^2$	16103	21978
Mean	49.5	64.4
Variance	280.3	310.3
S.E.Mean	6.84	7.88

We can see that the observed difference in sample means is less than two standard errors: not a rigorous test, but an indication that there is likely to be weak or no evidence against the null hypothesis of *no* differences in population mean.

Assuming firstly (and perhaps unrealistically) that σ^2_A and σ^2_B are known, a test can be easily derived using the Normal distribution. As $\overline{X}_A \sim N(\mu_A, \sigma_A^2)$ and $\overline{X}_B \sim N(\mu_B, \sigma_B^2)$ are independent $\overline{X}_A - \overline{X}_B \sim N(\mu_A - \mu_B, \sigma_A^2/n_A + \sigma_B^2/n_B)$ and under H_0 this is a *known* distribution with zero mean. A suitable test statistic is therefore

$$Z = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \sim \text{N(0,1) under } H_0$$

Example 7 revisited: If our assumed values for σ_A and σ_B are 18 and 15 respectively, then the observed value of Z is

$$z = \frac{49.5 - 64.4}{\sqrt{324/6 + 225/5}} = \frac{-14.9}{\sqrt{99}} = -1.4975$$

From Table 3 we obtain $P(Z > 1.49) = 0.06811$, $P(Z > 1.50) = 0.06681$ so using *linear interpolation* $P(Z > 1.4975) = 1/4 \times 0.06811 + 3/4 \times 0.06681 = 0.06713$ approximately, and therefore P for a two-sided test is $2 \times 0.06713 = 0.1342$: no evidence for differences. For a one-sided test we would obtain $P = 0.06713$ which would be *slight* evidence against H_0.

The assumption of known variances is clearly unrealistic yet can be tested (see Example 5), but if there is any doubt concerning our assumed values, we usually proceed with one or both of the following tests:

1 Two Sample t-test with pooled variance estimator Based on the assumption that although the two population variances are unknown they are in fact equal to each other with a common value of o.2, say.

$$Z = \frac{\overline{X}_A - \overline{X}_B}{\sigma\sqrt{1/n_A + 1/n_B}}$$

so we need only estimate σ. The appropriate estimator is a *weighted* average of the two unbiased estimators s^2_A and s^2_B called the *pooled variance estimator* s^2_o where each estimate is weighted by its degrees of freedom:

$$s_o^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}$$

It can be shown that replacing σ by S_O in Z gives a test statistic T say, which has a t-distribution with $n_A + n_B - 2$ degrees of freedom (the sum of the two d.o.f.'s).

Example 7 revisited From the data in the example

$$s_o^2 = \frac{5\times(280.3)+4\times(310.3)}{6+5-2} = -1.435$$

so s_o^2 = 17.14. The observed value of *T* is therefore

$$t = \frac{49.5-64.4}{17.4\sqrt{1/6+1/5}} = -1.435$$

which we refer to Table 3 with $v = 6+5-2 = 9$ For a one-sided alternative $H_1: \mu_A < \mu_B$. We are interested in large *negative* values of *T* so *P* = *P*(*T* < ⁻1.435). Now $t_9(0.1) = 1.383$, $t_9(0.05) = 1.833$ so $0.05 < P < 0.10$, slight evidence against H_0. A two-sided alternative and test would give $0.1 < P < 0.2$, *no* evidence against H_0.

If we have no reason to believe that the population variances are different, then the above procedure is valid, but the assumption of equal variances nevertheless should be checked. So we test $H_o^{'}: \sigma_A^{\,2} = \sigma_B^2$ versus $H_o^{'}: \sigma_A^{\,2} \neq \sigma_B^2$ and it can be shown that a suitable test statistic is

$$F = s_A^2 \big/ s_B^2 \sim F_{n_B-1}^{n_A-1} \quad \text{under } H_o^{/}$$

which is Snedecor and Fisher's F-distribution, whose percentage points are given in Table 6. Note that the degrees of freedom in the numerator and denominator are the superscript and subscript respectively. It can also be shown that

$$F^{-1} = s_B^2 \big/ s_A^2 \sim F_{n_B-1}^{n_A-1} \text{ under } H_o^{/}$$

Clearly values of *F* (or F^{-1}) which are very much greater or very much less than 1 provide evidence against $H_o^{/}$.

Example 7 revisited: The observed value of *F* is 280.3/310.3 which is 0.9033 so *P* = 2 x $P(F_4^5 < 0.9033)$. However, Table 4 gives only upper tails so we write instead $P = 2\times P(F_3^4 > 1/0.9033)$ Thus we refer 1.107 *not* 0.9033 to Table 4 and obtain *P* > 2 x 0.10 and there is clearly no evidence against $H_o^{/}$. Thus our assumption has been validated in the sense that there is no evidence against it.

But what if this F-test does give evidence (strong or weak) against the assumption of equal variances embodied in $H_o^{/}$? It does not make sense now to 'pool' estimates of variance and we must instead adopt an approximate procedure. MINITAB uses this procedure as the default so care must indeed be taken (see below).

Example 8: ('Cholesterol data') The data are:

	2-day *(A)*	Control *(B)*
n	28	30
$\bar{x}$	253	193.13
s	47.71	22.30
$s/\sqrt{n}$	9.02	4.07

An examination of the standard errors shows there is likely to be evidence for differences in the two population means but before we embark on a formal test (using the 'pooled' estimate) we should check the assumption of equal variances, particularly as the sample variances seem worryingly different. The observed value of F is $(47.71)^2/(22.30)^2 = 4.5774$ and this should be referred to F^{27}_{29}. From Table 9 we have $F^{20}_{29}(0.001) = 3.54$, $F^{40}_{29}(0.001) = 3.12$, so $3.54 > F^{27}_{29}(0.001) > 3.12$ and as our observed value is greater than this, $P < 0.002$. There is very strong evidence *against* equal variances so the following approximate procedure should be used:

2. Two Sample t-test: approximate procedure

If we replace both population variances in Z on page ( )by their sample estimators, the test statistic becomes

$$T = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{s_A^2/n_A + s_B^2/n_B}}$$

but the distribution of T^* is not known exactly. However it can be approximated by a Student's t-distribution with ν degrees of freedom, where

$$\nu = \frac{(s_A^2/n_A + s._B^2/n_B)^2}{.\left(\frac{s_A^4/n_A^2}{n_A - 1} + \frac{s_B^4/n_B^2}{n_b - 1}\right)}$$

An easier way to remember this formula is by writing it as

$$\nu^* = \frac{v_A + v_B)^2}{v_A^2/(n_A - 1) + v_B^2/(n_B - 1)}$$

,where $v_A = s_A^2/n_A$ *and* $v_B = s_B^2/n._B$ are the estimated variances of the sample means $\bar{X}_A$ and $\bar{X}_B$ respectively. Note that v is equal to $n_A + n_B - 1$ as in the 'pooled' case only when $s_A^2 = s_B^2$ and $n_A = n_B$

Example 8 revisited $v_A = (9.02)^2$, $v_B = (4.07)^2$

$$t^* = \frac{97.9253^*}{9.02^4/27 + 4.07^4/29} = 37.66$$

which is not an integer but we can still use Tables by suitable interpolation. The observed value of T^* is

$$t^* = \frac{60.8.}{\sqrt{(9.02)^2 + (4.07)^2}} = 6.144$$

which we refer to Table 4. We find that $t_{30}(0.0005) = 3.646$ and $t_{40}(0.0005) = 3.551$, so for a one-sided test $P < 0.0005$, as large positive values of T^* indicate departures from H_0 in favour of $H_1 : \mu_A > \mu_B$. There is very strong evidence against the hypothesis that the means are equal.

Note that if we had wrongly assumed the population variances were equal and used the 'pooled' procedure, the observed value of T would be 6.286 (very similar) but that the degrees of freedom would be 56 (much larger, giving a more powerful test). This is a good argument for using the 'pooled' procedure whenever possible, but see the MINITAB handbook for a counter-argument.

Summary: Inference on $\mu_A - \mu_B$

A. **Both σ_A^2 and σ_B^2 known** (or approximately known when both sample sizes n_A, n_B large, say 50, say $\geq$ 50 and using s_A^2 for σ_A^2 and s_B^2 for σ_B^2

1. $H_o : \mu_A = \mu_B$ versus $H_1 : \mu_A \neq \mu_B$
2. test statistic $T = \dfrac{\overline{X}_A - \overline{X}_B}{\sqrt{s_A^2/n_A + s_B^2/n_B}} \sim N(0,1)$ under H_o
3. $P = 2 \times P(Z > |z|)$ where z is observed value (Table 1)
4. $100(1 - \alpha)\%$ Confidence Interval

$$\bar{x}_A - \bar{x}_B \pm \Phi^{-1}(\alpha/2)\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

using Table 1 for percentage points $\Phi^{-1}(\alpha/2)$

B. Both σ_A^2 and σ_B^2 unknown that is the population variances are unknown but are equal**.**

1. $H_o : \mu_A = \mu_B$ versus $H_1 : \mu_A \neq \mu_B$
2. test statistic $Z = \dfrac{\overline{X}_A - \overline{X}_B}{s_o\sqrt{1/n_A + 1/n_B}} \sim t_{n_A+n_B-2}$ under H_o

   where s_o^2 is the *pooled estimate of population variance* given by

$$s_o^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}$$

3. $P = 2 \times P(T > |t|)$ where t is observed value (Table t)
4. $100(1 - \alpha)\%$ Confidence Interval

$$\bar{x}_A - \bar{x}_B \pm t_{n_A+n_B-2}(\alpha/2)\sqrt{1/n_A + 1/n_B}$$

C. Both Population Variances Unknown (and *not* assumed to be equal).

1. $H_o : \mu_A = \mu_B$ versus $H_1 : \mu_A \neq \mu_B$

2. test statistic $Z = \dfrac{\overline{X}_A - \overline{X}_B}{\sqrt{s_A^2/n_A + s_B^2/n_B}} \sim t_v$ under H_o

   where v is determined by

$$v = \frac{(s_A^2/n_A + s.{}_B^2/n_B)^2}{\left(\dfrac{s_A^4/n_A^2}{n_A - 1} + \dfrac{s_B^4/n_B^2}{n_b - 1} \right)}$$

3. P = 2 x $P(T >$ | t |) where t is observed value (Table 4)

4. 100(1 a)% Confidence Interval

$$\bar{x}_A - \bar{x}_B \pm t._{v^*}(\alpha/2)\sqrt{s_A^2/n_A + s_B^2/n_B}$$

6.3.3 Paired Samples

Example 9: Sixteen patients sampled at random were assigned as matched pairs to two treatments, treatment *A* being assigned to a random member of each pair. A response was measured and the data were:

A	B	X (difference)
14.0	13.2	+ 0.8
5.0	4~7	+ 0.3
8.6	9.0	0.4
11.6	11.1	+ 0.5
12.1	12.2	⁻ 0.1
5.3	4.7	+ 0.6
8.9	8.7	+ 0.2
10.3	9.6	+ 0.7

If we had assumed that the two samples were independent, and performed a two sample t-test, the observed value of the test statistic would be 0.205 on 14 degrees of freedom so there would have been certainly *no* evidence for difference in means (check this as an exercise). The large patient-to-patient variability within each treatment group would have obscured or *masked* any difference in the means (if indeed there were any differences.) But of course this (invalid) analysis ignores the valuable pairing information.

As the parameter of interest is still the difference in population means $\mu_A - \mu_B$, we look at differences between the members of a pair. In this example, only two differences are negative, so do the data provide sufficient evidence against $H_o : \mu_A - \mu_B = 0$ in favour of $H_1 : \mu_A - \mu_B \neq 0$.

Note that the parameter of interest is equivalently the mean of the population of differences (defined on each pair) so we test H_0 by a one-sample procedure based on the i-distribution (assuming the variance of the differences σ^2, say, is unknown). Under H_0 and the Normality assumption (which can be checked by say, a Normal probability plot of the differences) , the differences are a random sample from a Normal population with zero mean and variance σ^2, so a suitable test statistic is

$$T = \frac{\overline{X}}{s/\sqrt{n}} \sim t_{n-1} \quad \text{under } H_o$$

where n is the number of *pairs* in the data, and the sample mean $\overline{X} = \overline{X}_A - \overline{X}_B$ and sample variance $s^2 \neq s_A^2 + s_B^2$ refer to the sample of *differences X*.

The observed value of T is

$$t = \frac{0.325}{0.413/\sqrt{8}}$$

on 7 d.o.f., so referring to Table 7, we have $0.025 < P < 0.05$ for a. one-sided test. We conclude that there is *moderate* evidence for H_1 in favour of H_0. So by pairing we have achieved a more precise comparison of the two treatments. There is a generalization of pairing to more than two treatments, called *blocking* in experimental design but this is beyond the scope of this course.

6.6.4 Confidence Intervals and Tests

The detailed derivation of confidence intervals is *not* required in this section. Usually a confidence interval is quoted for the difference in means (or mean difference) once a 'significant' result is obtained from a test of H_0, say when the P-value is less than 0.05.

Example *7* revisited: A 95% confidence interval for $\mu_A - \mu_B$ is

$$\overline{x}_A - \overline{x}_B \pm t_{n_A+n_B-2}(\alpha/2)s_o\sqrt{1/n_A + 1/n_B}$$

where $\alpha = 0.05$ sousing Table 2 and the sample data, we have —14.9 ± 17.14 x 2.262 x 0.6055 = ⁻14.9±23.48 = (⁻38.4,8.6). Note that this interval includes

the value zero, confirming our previous result that there was no evidence for differences in the population means. We can think of any value within the interval as a 'plausible' value, as it would not be 'rejected' as a null hypothesis value with a significance level of $\alpha = 0.05$.

Similarly for intervals under other assumptions (see the summary at the end of the last section).
In general$_1$ the $100(1 - \alpha)\%$ confidence interval contains every value (and only those values) of the unknown parameter of interest such that the P-value is greater than or equal to a in a two-sided test of the null hypothesis that the unknown parameter takes that value.

Exercise: Review the remaining examples in this Chapter and explore the connections by computing appropriate confidence intervals.

6.7 Exercise

1 . For the ROCKET data of Example 1 in Chapter 1 and Question 3, Exercise1:

(a) test the hypothesis that the mean thrust is 1002 against a general alternative;

(b) give a 95% and a 99% confidence interval for the mean thrust and comment;

(c) check your calculations using MINITAB and comment (using an appropriate plot) on the validity of your assumptions.

2 For a random sample of size 7 from a Normal distribution with $\bar{x} = 3.47$,

(a) test the hypothesis that the population mean is 3, against a one-sided alternative that it is greater than 3 on the assumption that the population variance is 0.5;

(b) check this assumption if in fact $s^2 = 1$.

3 Two machines *A* and *B* fill bottles with fluid. Six bottles were measured carefully from a large production run for each machine. The results (in fluid ounces) were
A: 16.03 16.01 16.04 15.96 16.05 15.98
B: 16.02 16.03 15.97 16.04 15.96 16.01
Is there any evidence that the average amounts filled by the two machines differ? State any assumptions you make. Perform a non-graphical test for one of these assumtions.

4. Using the PULSE data and MINITAB answer the following questions *for those who 'ran in place' (RAN=1):*
Is there evidence for differences in mean *increase* in pulse rate between

(a) smokers and non-smokers?

(b) males and females?

Give a TABLE of the two factors with the number (sample size), mean and standard deviation of pulse rate increase in each cell. Comment.

5 Ten patients who suffered from insomnia were examined in a medical study to determine the effect of a sedative. Each patient received both the sedative and a placebo (control drug) for a two-week period, the drugs being administered in random order, and there was a 'cooling-off' period of one week in between the two two-week periods. Neither the patient nor the drug administrator knew which drug was being taken. The average

number of hours of sleep per night was recorded for each patient for each drug and the results were:

Patient	1	2	3	4	5	6	7	8	9	10
Sedative	1.3	1.1	6.2	3.6	4.9	1.4	6.6	4.5	4.3	6.1
Placebo	0.6	1.1	2.5	2.8	2.9	3.0	3.2	4.7	5.5	6.2

(a) Is there any evidence of differences showing that the sedative has a beneficial effect on patients?

(b) Give a 99% confidence interval for the effect.

(c) State clearly the assumptions you make and include a check.