CHAPTER 7

Confidence Intervals

In chapter 2, it was explained that the location, dispersion and distribution statistics calculated from a sample of data were intended to convey two types of information:

(i) a summary of the behaviour of the data in that sample,

(ii) indications of the characteristics of the population from which the sample was obtained.

In almost all scientific studies it is the second type of information that is of interest.

In section 6, $\bar{Y}$ and S^2 were proposed as estimators of the population mean μ and the population variance σ^2, respectively. Thus for a particular sample, $\bar{y}$ and s^2 could be used as *point estimates* of μ and σ^2 respectively. The term "point estimate" refers to the fact that $\bar{y}$, for example, serves as a location statistic for μ but does not by itself provide an indication of the uncertainty concerning the true value of μ. The parameters μ and σ^2 are fixed but unknown values.

One useful measure of the uncertainty <u>and</u> location of an unknown parameter, such as μ and σ^2, is provided by a *confidence interval*. It is an interval, calculated from the data in the available sample, which has a specified probability of containing that parameter. For example, a 90 per cent confidence interval for μ would be an interval that has a probability of 0.90 of containing μ.

Two points should be noted here:

(i) It is the interval itself, and <u>not</u> the parameter (such as μ or σ^2), that has an associated probability. In fact the interval is a random variable.

(ii) A confidence interval is an interval of uncertainty for a parameter (such as μ or σ^2) and <u>not</u> for an estimator (such as $\bar{Y}$ or S^2) or a statistic (such as $\bar{y}$ or s^2).

A confidence interval for a parameter may be regarded as a range of plausible values for that parameter in the light of the available data. The degree of plausibility is expressed by the probability level associated with the interval. As implied above, different samples of data will give rise to different confidence intervals at the same probability level for the same parameter.

7.1 CONFIDENCE INTERVALS FOR μ

~~In this section it is assumed that the~~ population mean μ is unknown but that a sample of n normally distributed measured values is available. Derivation of a confidence interval expression for μ is considered in two cases.

Case 1 : σ^2 assumed known

this would assume the data is always maintaining the same dispersion characteristics

I am a little confused with the intuition being used in this paragraph

When a process has been operating for a long time under regular monitoring it is often reasonable to assume that the variance, σ^2, of measurements from that process is "known". If the statistic s^2 were calculated from the very large data record for that process, it's associated number of degrees of freedom would be sufficiently large that s^2 would be virtually equal to σ^2. Note that μ might well change from time to time in such a process even though σ^2 remained constant.

If we assume that each of the measured values $Y_1, Y_2, \cdots, Y_n$ in a sample of n observations from this process is independently distributed as $N(\mu, \sigma^2)$, then as stated in section 6.1, $\overline{Y}$ is distributed as $N(\mu, \sigma^2/n)$. This information enables us to derive a confidence interval for μ at any specified level of probability. For illustration we first derive a 95 per cent confidence interval for μ and then develop a more general confidence interval expression for any specified probability level.

For an $N(0,1)$ random variable ~~Z, we know~~ from table 1 that $0.95 = P(-1.96 \le Z \le 1.96)$. Now if $\overline{Y}$ is $N(\mu, \sigma^2/n)$, then $\dfrac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ is $N(0,1)$. Therefore

$$0.95 = P\left(-1.96 \le \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \le 1.96\right), \text{ that is,}$$

$$0.95 = P\left(\overline{Y} - \frac{1.96\sigma}{\sqrt{n}} \le \mu \le \overline{Y} + \frac{1.96\sigma}{\sqrt{n}}\right).$$

First the term "plausible" must be defined in terms of a specific probability level. For illustration we shall use a level of 0.95. A 95 per cent confidence interval for μ in this case is

$$\bar{y} - \frac{1.96\sigma}{\sqrt{n}} \text{ to } \bar{y} + \frac{1.96\sigma}{\sqrt{n}},$$

that is,

$$52.52 - \frac{1.96\sqrt{11.37}}{\sqrt{14}} \text{ to } 52.52 + \frac{1.96\sqrt{11.37}}{\sqrt{14}},$$

or, 50.76 to 54.28.

Case 2 : σ^2 unknown

Although σ^2 is unknown, it can be estimated by S^2. However, because a confidence interval for μ depends upon σ^2, uncertainty concerning the true value of σ^2 must necessarily increase the uncertainty concerning the true value of μ. Thus we would expect the length of any confidence interval for μ in this case to be greater than that of the corresponding confidence interval in the σ^2 known case (case 1).

In case 1, we used the fact that $\bar{Y}$ was $N(\mu, \sigma^2/n)$ to derive a confidence interval for μ. In case 2, σ^2 is unknown and consequently, values of the unit normal random variable $Z = \dfrac{\bar{Y}-\mu}{\sigma/\sqrt{n}}$ cannot be calculated. If we use $\cancel{\mu}$ S in place of σ in this expression, we no longer have a variable that is $N(0,1)$ but instead a variable that has a t_v p.d.f. where v is the number of degrees of freedom associated with μ. That is, $\dfrac{\bar{Y}-\mu}{S/\sqrt{n}}$ is distributed as t_v.

Following the same line of argument as in case 1, a $100(1-\alpha)$ per cent confidence interval for μ when σ^2 is unknown is the interval $\boxed{\bar{Y} - \dfrac{t_{v,\alpha/2}S}{\sqrt{n}} \text{ to } \bar{Y} + \dfrac{t_{v,\alpha/2}S}{\sqrt{n}}}$ where $t_{v,\alpha/2}$ is the abscissa value of the t_v p.d.f. that leaves an upper tail area of $\alpha/2$. Once again, $(1-\alpha)$ is the probability that this confidence interval contains μ.

What is the definition of a degree of freedom in our case, and why is it taken to be "n-1"? where does the t- distribution come from? is it empirical? will you explain to us how it was derived?

Example 7.2

We shall reconsider example 7.1 with one change. The variance σ^2 is unknown this time but can be estimated by $s^2 = 12.2$, the sample variance of the 14 relative viscosity measurements. This estimate has $14 - 1 = 13$ degrees of freedom.

This time a 95 per cent confidence interval for μ is

$$\bar{y} - \frac{t_{v,0.025}s}{\sqrt{n}} \text{ to } \bar{y} + \frac{t_{v,0.025}s}{\sqrt{n}},$$

that is,

$$52.52 - \frac{t_{13,0.025}\sqrt{12.2}}{\sqrt{14}} \text{ to } 52.52 + \frac{t_{13,0.025}\sqrt{12.2}}{\sqrt{14}},$$

or 50.50 to 54.54 where $t_{13,0.025} = 2.16$.

Is it $S^2 = \frac{\sigma^2}{n-1}\chi^2_{n-1}$?

7.2 Confidence Intervals for σ^2

In section 6.2 it was stated without proof that for independently normally distributed data having a common mean μ and a common variance σ^2, the estimator S^2 defined in equation (6.7) has a p.d.f. of the form $\frac{\sigma^2}{n-1}\chi^2_{n-1}$. This fact can be used to derive a confidence interval expression for σ^2.

I am confused about the form of this p.d.f. what are the variables for the y and x axis?

Since S^2 is distributed as $\frac{\sigma^2}{n-1}\chi^2_{n-1}$, it follows that $\frac{(n-1)S^2}{\sigma^2}$ has a χ^2_{n-1} p.d.f. Then

$$1-\alpha = P\left(\chi^2_{n-1,1-\alpha/2} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{n-1,1-\alpha/2}\right)$$

where $\chi^2_{n-1,1-\alpha/2}$ and $\chi^2_{n-1,\alpha/2}$ are the abscissa values of the χ^2_{n-1} p.d.f. that leave upper tail areas of $1-\alpha/2$ and $\alpha/2$ respectively. $\chi^2_{n-1,1-\alpha/2}$ leaves a lower tail area of $\alpha/2$.

Thus,

$$1-\alpha = P\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right),$$

and so a $100(1-\alpha)$ per cent confidence interval for σ^2 is the interval $\left[\dfrac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} \text{ to } \dfrac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right]$.

Using the more general definition of S^2 given in equation (6.8), a $100(1-\alpha)$ per cent confidence interval for σ^2 is the interval

$$\left[\frac{\nu S^2}{\chi^2_{\nu,\alpha/2}} \text{ to } \frac{\nu S^2}{\chi^2_{\nu,1-\alpha/2}}\right].$$

where ν are the degrees of freedom associated with S^2.

Example 7.3

Using the information from example 7.2, a 95 per cent confidence interval for σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{13,0.025}} \text{ to } \frac{(n-1)s^2}{\chi^2_{13,0.975}},$$

that is,

$$\frac{13(12.2)}{24.74} \text{ to } \frac{13(12.2)}{5.01},$$

or 6.41 to 31.66.

Clearly the "known" value, $\sigma^2 = 11.37$, is also a very plausible value for σ^2 on the basis of the sample statistics from the 14 measurements.

CHAPTER 9

Comparisons Between Two Samples

Comparisons between two samples of data are usually made to determine whether they can be considered to have the same parent population. If the parent populations are different for the two samples it is of interest to learn in what respects they differ. Only comparisons between sample means and between sample variances are discussed in these notes and normal p.d.f.'s are assumed. A common example of this type of testing is a comparison of measurements from a modified process with those from a "standard" (i.e. unmodified) process

Two methods of comparison will be discussed,

(i) comparisons between the two groups considered as complete entities without regard to special identification of individual measurements within each group and

(ii) comparisons between individual pairs of measurements, each pair consisting of one measurement from each group.

In both cases the most valid comparisons between the two groups are achieved by subjecting both groups to treatment concurrently. This ensures that all extraneous sources of variation (i.e. those influences other than the deliberately planned differences in treatment) will effect both groups as equally as possible.

9.1 COMPARISONS BETWEEN TWO GROUPS CONSIDERED AS COMPLETE ENTITIES

The general case of n_1 measurements on a standard process and n_2 measurements on a modified process will be considered. If it could be assumed that these $(n_1 + n_2)$ measurements could be made independently of each other, then the order in which the measurements were made would not matter. Often, however, there are time trends or carry-over effects from one measurement to the next which create dependencies among successive measurements. To overcome the effects of possible dependencies (i.e. to "even out" their influence over the complete set of measurements) a *randomized order* should be used in making the measurements. This could be achieved in this case by putting n_1 pieces

How would randomizing the measurements get rid of time dependencies (?)

of paper marked "S" and n_2 pieces of paper marked "M" into a box, mixing well, and selecting each measurement in succession by one draw from the box.

Comparisons of the sample means and sample variances from two groups of measurements will now be described with the aid of example 9.1.

Example 9.1

Eleven measurements were made in randomized order, five on the standard process and six on the modified process, with the following results.

Measurement Number	Standard Process Result	Measurement Number	Modified Process Result
1	62.7	3	64.9
2	61.8	4	62.1
5	63.3	6	60.7
9	65.2	7	63.8
10	60.8	8	65.9
		11	66.7

Summary statistics for the two data sets are

$$n_1 = 5 \qquad n_2 = 6$$
$$\bar{y}_1 = 62.76 \qquad \bar{y}_2 = 64.02$$
$$s_1^2 = 2.75 \qquad s_2^2 = 5.25$$
$$\nu_1 = n_1 - 1 = 4 \qquad \nu_2 = n_2 - 1 = 5$$

9.1.1 Comparison of the sample variances s_1^2 and s_2^2

The two sample variances can be compared to determine whether the two underlying populations have the same population variance, i.e. whether $\sigma_1^2 = \sigma_2^2$. This comparison can be made by a confidence interval (i.e. does a confidence interval for the ratio σ_1^2/σ_2^2 enclose the value 1?) or by a significance test (i.e. is the null hypothesis $\sigma_1^2 = \sigma_2^2$ a reasonable one?).

Considering the confidence interval for σ_1^2/σ_2^2 first, we must use the fact, stated here without proof, that for normally distributed data the ratio $\dfrac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2}$ has an F p.d.f. with ν_1 and ν_2 degrees of freedom. Then

$$1-\alpha = P\left\{F_{v_1,v_2,1-\alpha/2} < \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} < F_{v_1,v_2,\alpha/2}\right\}$$

$$= P\left\{\frac{S_1^2}{S_2^2 F_{v_1,v_2,\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2 F_{v_1,v_2,1-\alpha/2}}\right\}$$

so that a $100(1-\alpha)$ per cent confidence interval for the ratio σ_1^2/σ_2^2 is

$$\left[\frac{S_1^2}{S_2^2 F_{v_1,v_2,\alpha/2}} \text{ to } \frac{S_1^2}{S_2^2 F_{v_1,v_2,1-\alpha/2}}\right].$$

Using $\alpha = 0.10$ in example 9.1, a 90 per cent confidence interval for σ_1^2/σ_2^2 is

$$\left[\frac{2.75}{5.25 F_{4,5,0.05}} \text{ to } \frac{2.75}{5.25 F_{4,5,0.95}}\right].$$

Using table 4 this interval is

$$\left[\frac{2.75}{5.25(5.19)} \text{ to } \frac{2.75}{5.25(1/6.26)}\right]$$

that is, [0.10 to 3.28].

Because the value 1 is well inside this interval it can be regarded as a plausible value for the ratio σ_1^2/σ_2^2. That is, this confidence interval has indicated that $\sigma_1^2 = \sigma_2^2$ is a reasonable conjecture.

A significance test of the null hypothesis $\sigma_1^2 = \sigma_2^2$ is usually a one-sided test, the alternative hypothesis being $\sigma_1^2 > \sigma_2^2$ if the sample results show that $s_1^2 > s_2^2$ (or $\sigma_2^2 > \sigma_1^2$ if .) In example 9.1 we shall test the null hypothesis $\sigma_1^2 = \sigma_2^2$ against the alternative hypothesis $\sigma_2^2 > \sigma_1^2$.

If the null hypothesis is true then the significance test can be expressed as follows.

Is $P\left\{\frac{S_2^2\sigma_1^2}{S_1^2\sigma_2^2} > \frac{s_2^2\sigma_1^2}{s_1^2\sigma_2^2}\right\} < \alpha$

Shouldn't this be flipped for the hypothesis $\sigma_2^2 > \sigma_1^2$

are we ever going to go into the proof of why we can use these wierd probability distributions

where s_1^2 and s_2^2 have degrees of freedom v_1 and v_2 respectively. Under the null hypothesis $\sigma_1^2 = \sigma_2^2$, the ratio $\frac{S_2^2\sigma_1^2}{S_1^2\sigma_2^2}$ has an F_{v_2,v_1} p.d.f. and so this question can be posed as follows.

Is $$P\left\{F_{v_2,v_1} > \frac{s_2^2}{s_1^2}\right\} < \alpha ?$$

Using $\alpha = 0.05$ this time so that table 4 can again be employed, it is seen that $F_{5,4,0.05} = 6.26$ and so

$$P\left\{F_{5,4} > \frac{5.25}{2.75}\right\}$$

is clearly greater than 0.05. Thus $\sigma_1^2 = \sigma_2^2$ is a plausible conjecture.

9.1.2 Comparison of the sample means $\bar{y}_1$ and $\bar{y}_2$

There are a number of cases to be considered in comparing the sample means $\bar{y}_1$ and $\bar{y}_2$. Normal p.d.f.'s are assumed for each case. We shall describe both a confidence interval approach and a two-sided significance test approach. Methods of dealing with situations requiring a one-sided significance test should be apparent from the material presented in this section and in section 8.

Case 1: σ_1^2 and σ_2^2 assumed known

If the values of the population variances σ_1^2 and σ_2^2 are known and the individual measurements are mutually independent as described earlier then

$$\left(\bar{Y}_1 - \bar{Y}_2\right) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

why are these added together

Consequently,

$$1 - \alpha = P\left\{-z_{\alpha/2} \le \frac{\left(\bar{Y}_1 - \bar{Y}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\sigma_1^2/n_1} + \sqrt{\sigma_2^2/n_2}} \le z_{\alpha/2}\right\}.$$

so that a $100(1-\alpha)$ per cent confidence interval for the difference in population means $\mu_1 - \mu_2$ is

$$\left[(\bar{Y}_1 - \bar{Y}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \text{ to } (\bar{Y}_1 - \bar{Y}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right].$$

In example 9.1, if $\sigma_1^2 = 3$ and $\sigma_2^2 = 5$, a 95% confidence interval for $\mu_1 - \mu_2$ would be

$$\left[(62.76 - 64.02) - 1.96\sqrt{\frac{3}{5} + \frac{5}{6}} \text{ to } (62.76 - 64.02) + 1.96\sqrt{\frac{3}{5} + \frac{5}{6}}\right],$$

that is, [-3.61 to 1.09]. In this case, since zero is an entirely plausible value for $\mu_1 - \mu_2$, the conjecture that the two processes have the same population mean would be a reasonable one.

A two-sided significance test of the null hypothesis $\mu_1 = \mu_2$ against the alternative hypothesis $\mu_1 \neq \mu_2$ can be expressed as follows :

Is $\quad P\{(\bar{Y}_1 - \bar{Y}_2) \geq |\bar{y}_1 - \bar{y}_2|\} + P\{(\bar{Y}_1 - \bar{Y}_2) \leq -|\bar{y}_1 - \bar{y}_2|\} < \alpha$?

That is,

is $\quad P\left\{Z > \dfrac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right\} + P\left\{Z < -\dfrac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right\} < \alpha$

For example 9.1 with $\sigma_1^2 = 3$ and $\sigma_2^2 = 5$,

$$P\left\{Z > \frac{|62.76 - 64.02|}{\sqrt{\frac{3}{5} + \frac{5}{6}}}\right\} + P\left\{Z < \frac{-|62.76 - 64.02|}{\sqrt{\frac{3}{5} + \frac{5}{6}}}\right\} = P\{Z > 1.05\} + P\{Z < 1.05\}$$

$$= 0.29$$

For commonly used significance levels this is not a "small" probability and so the null hypothesis that $\mu_1 = \mu_2$ is not rejected.

Case 2: σ_1^2 and σ_2^2 unknown

In this case it is necessary to test the conjecture that $\sigma_1^2 = \sigma_2^2$ before attempting to compare the two sample means $\bar{y}_1$ and $\bar{y}_2$. Methods for testing the equality of the two population variances have been described on the preceding pages. We now consider the two possible outcomes in turn.

Case 2a:

If $\sigma_1^2 = \sigma_2^2$ is a reasonable conjecture, then the two sample variances s_1^2 and s_2^2 can be *pooled* to provide the best available estimate of the common population variance which we shall denote as a σ^2. This pooled sample variance estimate is

$$s_p^2 = \frac{\nu_1 s_1^2 + \nu_2 s_2^2}{\nu_1 + \nu_2}.$$

and it has $\nu = \nu_1 + \nu_2$ degrees of freedom. Because s_p^2 has a higher number of degrees of freedom than either s_1^2 or s_2^2, it is a more reliable estimate of σ^2. For example 9.1 it has been shown that $\sigma_1^2 = \sigma_2^2$ is a reasonable conjecture and so

$$\begin{aligned} s_p^2 &= \frac{4(2.75) + 5(5.25)}{4+5} \\ &= 4.14 \end{aligned}$$

is a pooled estimate of the common population variance σ^2 with 9 degrees of freedom.

Now, analogous to the p.d.f. of $\bar{Y}$ in a one sample case, it can be shown that for independent measurements

$$\frac{(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{s_p^2\left[\dfrac{1}{n_1} + \dfrac{1}{n_2}\right]}}$$

is distributed as t_ν where s_p^2 is the pooled variance estimator having ν degrees of freedom.

Consequently a $100(1-\alpha)$ per cent confidence interval for $\mu_1 - \mu_2$ in this case is

$$\left[(\bar{Y}_1 - \bar{Y}_2) - t_{v,\alpha/2}\sqrt{s_p^2\left[\frac{1}{n_1}+\frac{1}{n_2}\right]} \text{ to } (\bar{Y}_1 - \bar{Y}_2) + t_{v,\alpha/2}\sqrt{s_p^2\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}\right].$$

From the data given in example 9.1 a 95 per cent confidence interval for $\mu_1 - \mu_2$ is

$$\left[(62.76-64.02) - t_{9,0.025}\sqrt{4.14\left[\frac{1}{5}+\frac{1}{6}\right]} \text{ to } (62.76-64.02) + t_{9,0.025}\sqrt{4.14\left[\frac{1}{5}+\frac{1}{6}\right]}\right],$$

that is, [-4.05 to 1.53]. Again, because zero is contained well inside this confidence interval, the conjecture that $\mu_1 = \mu_2$ is a reasonable one.

A two-sided significance test of the null hypothesis $\mu_1 = \mu_2$ against the alternative hypothesis $\mu_1 \neq \mu_2$ can be expressed as follows for this case,

is $$P\left\{t_v > \frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{s_p^2\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}}\right\} + P\left\{t_v < \frac{-|\bar{y}_1 - \bar{y}_2|}{\sqrt{s_p^2\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}}\right\} < \alpha\,?$$

For example 9.1,

$$P\left\{t_9 > \frac{|62.76-64.02|}{\sqrt{4.14\left[\frac{1}{5}+\frac{1}{6}\right]}}\right\} + P\left\{t_9 < \frac{-|62.76-64.02|}{\sqrt{4.14\left[\frac{1}{5}+\frac{1}{6}\right]}}\right\} \geq 0.20$$

from table 2. This probability is not one normally judged as "small" and so the null hypothesis $\mu_1 = \mu_2$ is not rejected.

Thus none of the comparisons of the two sample means $\bar{y}_1$ and $\bar{y}_2$ have indicated any significant difference between the population means of the standard and modified processes.

do this on it? does this problem arise because the "t" distribution can no longer be used?

Case 2b:

If σ_1^2 and σ_2^2 are unknown and a comparison of the two sample variances indicates that $\sigma_1^2 = \sigma_2^2$ is *not* a reasonable conjecture, then no exact procedure exists for testing the equality of the two population means. Several approximate procedures have been proposed for this problem, known as the Behrens-Fisher problem, and one of these, due to Cochran, is now described.

An approximate two-sided test of significance of the null hypothesis $\mu_1 = \mu_2$ can be made by referring the statistic,

$$\frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

to a "d" distribution.

The abscissa value d_α which leaves an upper tail area of size α is

$$d_\alpha = \frac{t_{v_1,\alpha}\left(s_1^2/n_1\right) + t_{v_2,\alpha}\left(s_2^2/n_2\right)}{s_1^2/n_1 + s_2^2/n_2}.$$

If in example 9.1, s_2^2 had been 25.25 instead of 5.25, but the values of all the other statistics had remained unchanged, then a significance test of the null hypothesis $\sigma_1^2 = \sigma_2^2$ against the alternative hypothesis $\sigma_1^2 < \sigma_2^2$ would show that

$$P\left\{F_{5,4} > \frac{25.25}{2.75}\right\}$$

is clearly less than 0.05 so that $\sigma_1^2 = \sigma_2^2$ is not a plausible conjecture.

To test the null hypothesis $\mu_1 = \mu_2$ against the alternative hypothesis $\mu_1 \neq \mu_2$, the following question must be answered.

Is $$P\left\{d > \frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]}}\right\} + P\left\{d < \frac{-|\bar{y}_1 - \bar{y}_2|}{\sqrt{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]}}\right\} < \alpha\ ?$$

Now $$\frac{|62.76 - 64.02|}{\sqrt{\left[\frac{2.75}{5} + \frac{25.25}{6}\right]}} = -0.58.$$

If a significance level of 0.05 is used, then

$$d_{0.025} = \frac{t_{4,0.025}\left(2.75/5\right) + t_{5,0.025}\left(25.25/6\right)}{\left(2.75/5\right) + \left(25.25/6\right)}.$$
$$= 2.59$$

Thus $P\{d > 0.58\} + P\{d < -0.58\}$ is clearly greater than 0.05, implying that the null hypothesis $\mu_1 = \mu_2$ is a reasonable conjecture.

Where did this come from

The two population means can also be compared in this case using the following $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$,

$$\left[(\bar{y}_1 - \bar{y}_2) - d_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \text{ to } (\bar{y}_1 - \bar{y}_2) + d_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right].$$

9.2 TESTING FOR A DIFFERENCE IN POPULATION MEANS BY PAIRED COMPARISONS BETWEEN TWO GROUPS

Considerable gain in sensitivity can be achieved in tests comparing sample means from two groups by eliminating extraneous sources of variation from the comparisons. An effective way to do this for comparisons between two groups is to arrange for measurements to be made in pairs, each pair comprising one measurement from each group. With proper planning the measurements within each pair will have as much in common as possible, except of course for any inherent difference between the two processes, which is the quantity to be detected. Each pair is called a "block" and, as discussed later in this course, blocking can be extended for comparisons among several groups.

In example 9.2 two types of insulating material are being tested to determine whether they have a difference in mean impact strength. Only two measurements can be carried out each day. One obvious blocking arrangement in this situation is to test each type of material once per day. The order in which the two types of material are tested should be selected at random each day to eliminate any possible bias arising from a time trend. With this pairing arrangement daily comparisons of test results would be expected to be sensitive to any difference in the mean impact strengths between the two products even though extraneous factors such as testing personnel and environmental conditions might vary from day to day. In general, tests conducted on the same day are more likely to share common conditions than tests conducted on different days.

Blocking is often arranged by time, as in example 9.2, by location, or by other test conditions. Treatments on paper products, for example, can be blocked using the location of paper specimens from the sheet. Specimens located close to one another would be expected to be more similar than specimens located far apart, thereby allowing any differences in the treatments to be more evident.

The comparison of population means from two groups using paired comparisons will be demonstrated using example 9.2.

Example 9.2

The following data are the results of paired testing of two insulating materials. The order of the two tests each day was determined by the flip of a coin.

Day	Impact Strength of Material A	Impact Strength of Material B	w = Difference in Impact Strengths
1	1.25	1.01	0.24
2	1.16	0.89	0.37
3	1.33	0.97	0.36
4	1.15	0.95	0.20
5	1.23	0.94	0.29
6	1.20	1.02	0.18
7	1.32	0.99	0.33
8	1.28	1.06	0.22
9	1.21	0.98	0.23

Although from these results there is little doubt that the population mean impact strength for material A is greater than that for material B (the probability of obtaining 9 positive differences in 9 trials being $0.5^9 = 0.00195$ if $\mu_A = \mu_B$) it is of interest to determine how large the difference in population means might be on the basis of these data.

Denoting each daily difference in measurement impact strength as w where $w_i = y_{Ai} - y_{Bi}$, $i = 1, \cdots, 9$ the sample mean of these differences is

$$\begin{aligned} \bar{w} &= \frac{\sum_{i=1}^{9} w_i}{9} \\ &= 0.258 \qquad (= \bar{y}_A - \bar{y}_B) \end{aligned}$$

and the sample variance of these differences is

blocking

$$\begin{aligned} s_w^2 &= \frac{\sum_{i=1}^{9} (w_i - \bar{w})^2}{8}. \\ &= 0.00359 \end{aligned}$$

A $100(1-\alpha)$ per cent confidence interval for $\mu_A - \mu_B$, the difference in population means, in this case (variance of the daily differences unknown) is

$$\left[\bar{w} - \frac{t_{8,0.025} s_w}{\sqrt{9}} \text{ to } \bar{w} + \frac{t_{8,0.025} s_w}{\sqrt{9}} \right].$$

That is, for $\alpha = 0.05$,

$$\left[0.258 - \frac{2.306\sqrt{0.00359}}{3} \text{ to } 0.258 + \frac{2.306\sqrt{0.00359}}{3} \right]$$

or [0.212 to 0.304]. This interval could be regarded as a range of plausible values for the difference $\mu_A - \mu_B$.

To demonstrate the increase in sensitivity achieved by pairing, a comparison of sample means is now made using each group of measurements as a complete entity (i.e. using the methods of section 9.1). For the data in example 9.2,

$n_A = 9$ $\qquad$ $n_B = 9$

$\bar{y}_A = 1.237$ $\qquad$ $\bar{y}_B = 0.979$

$s_A^2 = 0.00415$ $\qquad$ $s_B^2 = 0.00246$

$\nu_A = n_A - 1 = 8$ $\qquad$ $\nu_B = n_B - 1 = 8$

First of the null hypothesis a $\sigma_A^2 = \sigma_B^2$ is made against the alternative hypothesis $\sigma_A^2 > \sigma_B^2$.

$$P\left\{F_{8,8} > \frac{0.00415}{0.00246}\right\}$$

is clearly greater than 0.05 (see table 4) so that $\sigma_A^2 = \sigma_B^2$ is a reasonable conjecture. A pooled estimate of the common variance σ^2 is then

$$s_p^2 = \frac{8(0.00415) + 8(0.00246)}{8+8}$$
$$= 0.00331$$

with 16 degrees of freedom.

A 95% confidence interval for $\mu_A - \mu_B$ is

$$\left[(\bar{y}_A - \bar{y}_B) - t_{16,0.025} s_p \sqrt{\frac{1}{9} + \frac{1}{9}} \text{ to } (\bar{y}_A - \bar{y}_B) + t_{16,0.025} s_p \sqrt{\frac{1}{9} + \frac{1}{9}}\right],$$

that is, [0.201 to 0.315].

This 95 per cent confidence interval is wider than that obtained using paired comparisons because $s_p^2\left(\frac{1}{9} + \frac{1}{9}\right)$, the estimated variance of $(\bar{y}_A - \bar{y}_B)$, includes day to day variation whereas $s_w^2/9$ the estimated variance of $\bar{w}$, does not. Notice that the effect of this increase in variance more than offsets the decrease in t value caused by the increased number of degrees of freedom in the unpaired comparison test in determining the length of a confidence interval for $\mu_A - \mu_B$.