

MRes in Medical Statistics

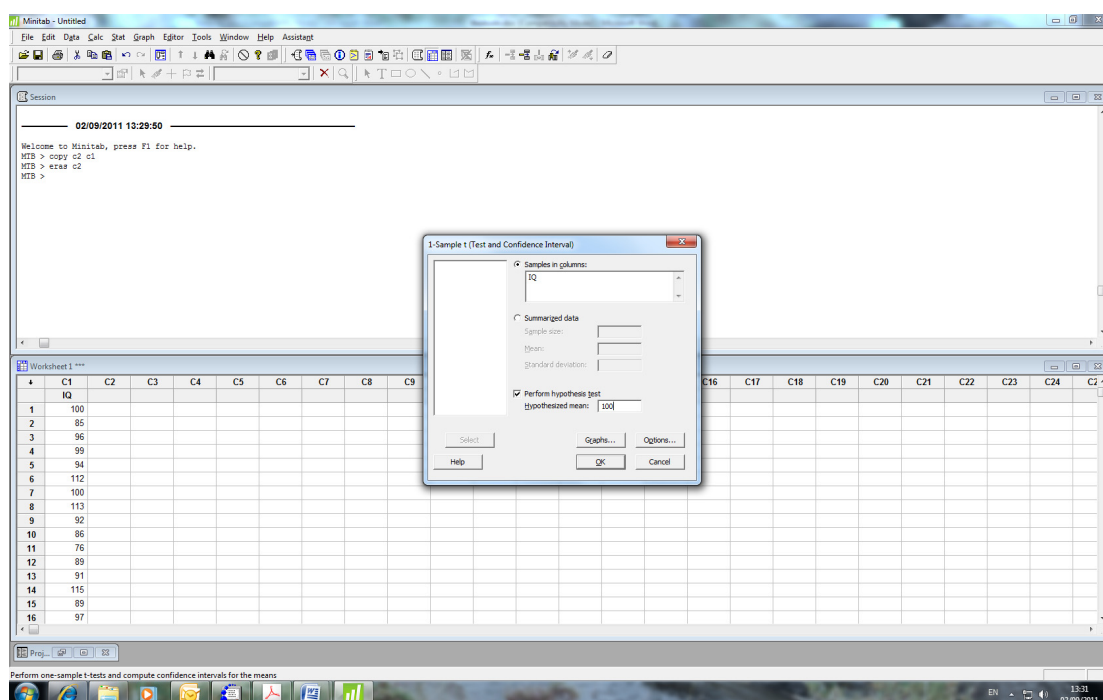
MMB8028

School of Mathematics and Statistics

Practical session on *t*-tests: outline solutions

1.

The IQ values are entered into column C1, named 'IQ' and **1-Sample t...** chosen from the **Basic Statistics** sub-menu of the **Stat** menu. The following screen will be seen and IQ is entered into the **Samples in columns:** box. To test the null hypothesis that the *population* mean is 100, you need to enter 100 into the **Hypothesized mean:** box. You can then click on **OK**.



The resulting screen shows the following in the Session window.

```
MTB > Onet 'IQ';  
SUBC> Test 100.
```

One-Sample T: IQ

Test of $\mu = 100$ vs $\text{not} = 100$

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
IQ	25	95.4400	10.8324	2.1665	(90.9686, 99.9114)	-2.10	0.046

The sample mean is given as 95.44 although it is sensible to quote this to just 1 d.p., namely 95.4 and the 95% confidence interval is (91.0, 99.9).

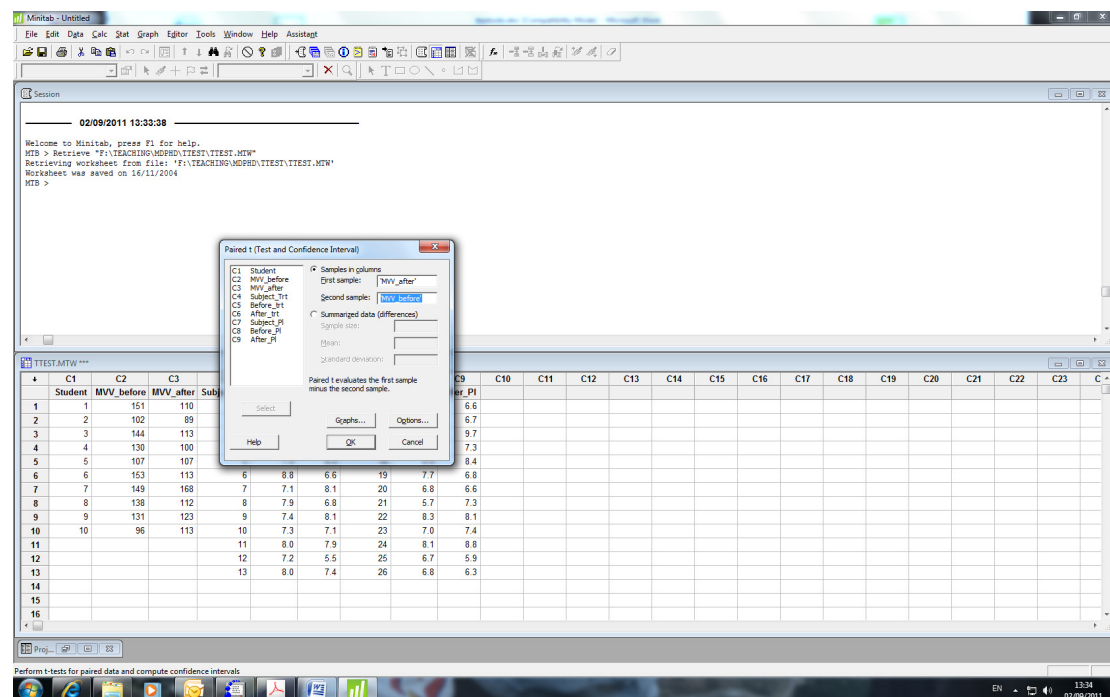
The standard error is 2.1665, so the confidence interval could be calculated by hand as $95.44 \pm 2 \times 2.1665 = 91.107, 99.773$, or using 1.96 in place of 2 (see appendix to handout), 91.194,

99.686. Neither agrees with the interval given by Minitab, as this uses the appropriate multiplier from a t -distribution on 24 degrees of freedom (see appendix), which is 2.0639, and $95.44 \pm 2.0639 \times 2.1665 = 90.968, 99.911$. This is much closer to that given by Minitab: the discrepancy results from rounding to two decimal places.

Note that in the above, figures are quoted to many more decimal places than should be used in a publication. However, it is important to keep as many decimal places as possible during intermediate calculations, with the rounding applied only to the final figures.

The last line of the output performs a t -test of the null hypothesis that $\mu = 100$, which results in a P -value of 0.046. This suggests that a sample mean of 95.44 would only be observed in 4.6% of samples if the population mean were 100, so this casts considerable doubt on the truth of the null hypothesis. It should also be noted that a P -value less than 0.05 could have been predicted as the 95% confidence interval does not include 100. The width of the confidence interval is rather large for IQs and suggests that a larger sample, if it were available, would be useful.

2.



Although there appear to be two samples in this analysis, the MVV before and the MVV after smoking the cigarettes, the samples are closely related. Each value in the post-smoking sample is paired with the pre-smoking value from the same student. The MVVs before smoking show that there is a substantial variation between students in this measure of lung function. Moreover, as interest focuses on the effect of smoking *on each student*, it is appropriate to eliminate this source of variation from the analysis. This is achieved by performing a paired t -test. Click on **Stat** -> **Basic statistics** -> **Paired t**. The above screen is seen and the columns containing the before and after smoking values of MVV are entered in the **First sample** and **Second sample** boxes. The results which appear in the Session Window are as follows.

Paired T-Test and CI: MVV_after, MVV_before

Paired T for MVV_after - MVV_before

	N	Mean	StDev	SE Mean
MVV_after	10	114.80	20.77	6.57
MVV_before	10	130.10	21.24	6.72
Difference	10	-15.30	22.01	6.96

95% CI for mean difference: (-31.05, 0.45)

T-Test of mean difference = 0 (vs not = 0): T-Value = -2.20 P-Value = 0.056

The paired t -test analyses the difference between the paired values, so attention is focussed on the row 'Difference'. The 95% confidence interval for the mean change is given as (-31.05, 0.45) litres/min.

If smoking has no effect on MVV then the population mean of the change in MVV, μ will be 0, and the paired t -test tests the null hypothesis that $\mu=0$. This gives a P -value of 0.056. A value greater than 0.05 was predictable from the fact that the 95% confidence interval included 0, although the fact that it only just does so would also suggest that the P -value will be only slightly larger than 0.05.

Thus there is no evidence that smoking has had an effect on the MVV but the confidence interval shows that the mean MVV certainly does not increase materially and might have decreased by as much as 31 l/min. If 31 l/min is a physiologically important change in MVV then it will probably be necessary to repeat the study with a larger sample. If 31 l/min is unimportant, then the study has, to a large extent, excluded any important difference due to smoking.

An alternative way to perform this test is to compute explicitly the change in MVV. This is most easily done by typing the command

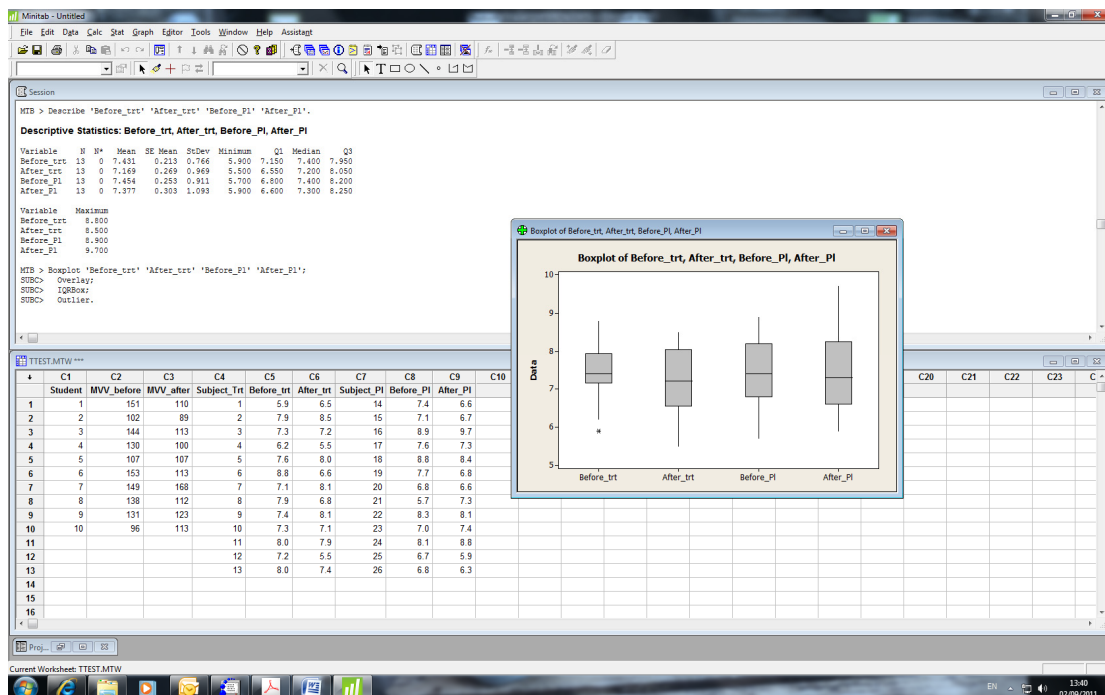
```
Let c20=c3-c2
```

in the Session window (make sure the Minitab prompt MTB> has been enabled[†]). Then choose **1-sample t...** where you chose **Paired t...** above. Put C20 into the **Samples in columns:** box and enter 0 in the **Test mean:** box. Try this and compare your answers with those obtained above.

[†] Make the Session window the active window and click on **E**ditor and make sure **E**nable commands is checked

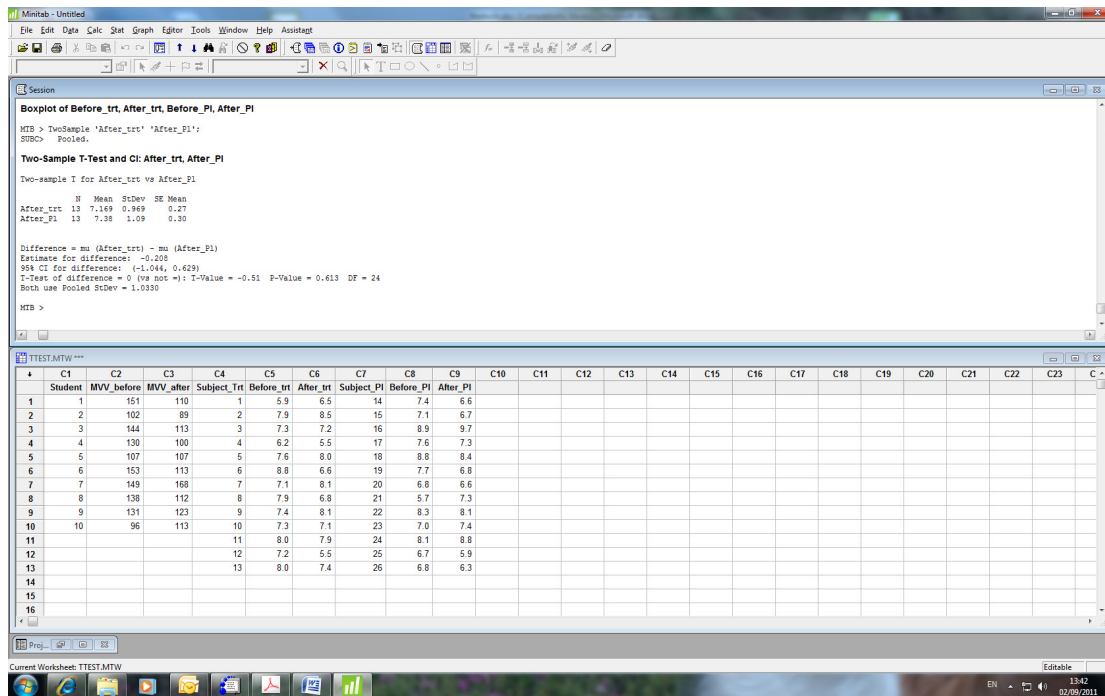
3.

The data can be summarised in various ways. The screen below shows boxplots for each of the columns given in the question, together with means and other summary quantities. The summary statistics have been obtained by using Display Descriptive Statistics, as in question 1. The Boxplots have been obtained from the Graph menu, clicking on Boxplots... and then selecting Multiple Y's Simple and entering the cholesterol concentrations in the Graph variables: box.



The problem both the graphical and numerical displays is that neither shows the structure of the data. The four sets of values comprise two pairs, namely patients given placebo, measured before and after administration, and patients given active treatment and also measured before and after.

When analysing data from a clinical trial it must always be borne in mind that the principal aim is to compare the different treatment groups. It is therefore appropriate to compare the two sets of values obtained after giving the placebo or the active treatment. This can be done using an unpaired t-test, which is found as the 2-Sample t... entry on the Basic Statistics item of the Stat menu. The result is show on the following screen. The 2-Sample t dialogue box has a check-box marked Assume equal variances. This is present because Minitab will perform a variant of the *t*-test which does not assume equal variances. This variant has dubious statistical properties and the box for Assume equal variances should be checked. In cases where this seems unreasonable, alternatives ought to be sought and some of these will be described later in the course.



This shows a P -value of 0.613, indicating that there is no evidence that there is a difference between the population mean cholesterol concentrations in the two treatment groups. This is emphasised by the mean difference of -0.21 mmol/l and a 95% confidence interval of $(-1.04, 0.63)$ mmol/l. This confidence interval indicates that there could be clinically important differences between placebo and the active treatment but that this has not been established by this analysis

What role, if any, do the observations taken at the beginning of the trial have? These will be unaffected by any effect the treatment may have and they provide information on the natural variation in cholesterol concentration between these patients. Inter-patient variation such as this is of no direct help in evaluating the difference between the treatments. As such, the values recorded at the beginning of the trial might allow this source of variation to be removed from the analysis. Consequently an alternative analysis is to compare the change in cholesterol concentration over the course of the trial between the two treatment groups.

The first step is to calculate the change in cholesterol concentration for both treatment groups. To do this most easily, make sure the Minitab prompt is showing in the Session window and then type

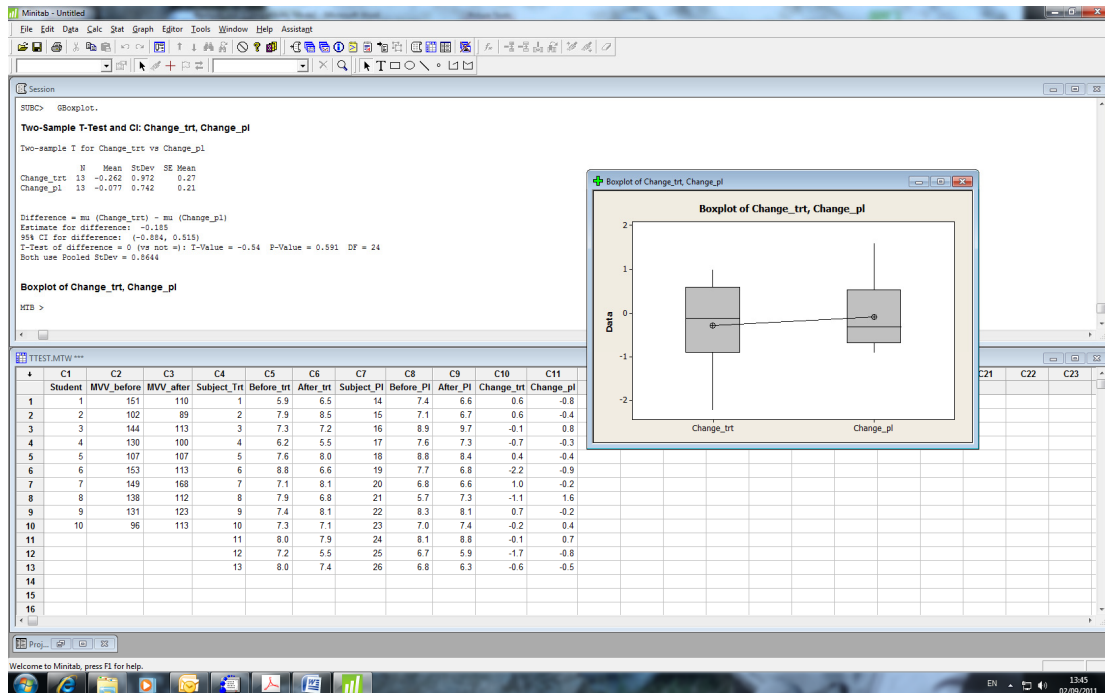
```

let c10=c6-c5
let c11=c9-c8

```

Columns c10 and c11 could then be named appropriately, perhaps as **Change_trt** and **Change_pl** respectively.

It is important to realise that although this alternative approach requires the calculation of differences, as is also required in the paired t -test, the analysis still compares two unrelated groups and therefore an *unpaired* t -test must be applied to the two sets of differences. Summary statistics of the changes and the t -test comparing them are shown in the following screen. In addition, boxplots for the change in each treatment group are given. This output is obtained in the same way as for the above analysis based only on the final cholesterol concentrations. The boxplots are added by clicking on the **Graphs...** box in the **2-sample t...** dialogue box and then selecting **Boxplots of data**.



The standard deviations of the change are smaller than the corresponding quantities for the post-treatment values, indicating that some inter-patients variation has been removed by considering the change in concentration. However, difference is not striking.

The t -test gives $P=0.591$, so again there is no evidence of a difference between the treatment groups. The estimate of a difference between the groups is -0.185 mmol/l, with a 95% confidence interval from $(-0.884, 0.515)$ mmol/l. The estimate of the difference between the groups is similar to that obtained from the post-treatment values alone, namely -0.21 mmol/l. This is as it should be – the two differ only by the difference in the baseline means and the randomization used to form the groups implies that these should be similar. The confidence interval based on the changes is narrower, as some of the between patient variation has been eliminated by considering change from baseline.

Therefore even the analysis based on within-patient changes leaves doubt about the value of the treatment.

An Erroneous Analysis

Data with this structure are often erroneously analysed in the following way. The differences within each patient are computed, as in the above analysis, but then a *paired* t -test is applied to each treatment group separately. If the columns containing the differences are both entered in the Samples in columns: box of the 1-sample t ... dialogue box then output shown below is obtained.

One-Sample T: Changes_trt, Change_pl

Test of $\mu = 0$ vs $\text{not} = 0$

Variable	N	Mean	StDev	SE Mean	95% CI	T
Changes_trt	13	-0.261538	0.971715	0.269505	(-0.848740, 0.325663)	-0.97
Change_pl	13	-0.076923	0.741793	0.205736	(-0.525184, 0.371338)	-0.37

Variable	P
Changes_trt	0.351
Change_pl	0.715

The approach is then to compare the P -values in the two groups. The above results would imply that there is no difference between the treatments as the non-significant P -values show that the population mean change in each group is 0. If one group had a significant P -value but the other not, then a difference would have been established.

This is flawed because it is based on interpreting a non-significant P -value as evidence that the null hypothesis is true. This is not correct: a non-significant P -value simply means that it has not been possible to establish a difference, which could be because there isn't one, or because we have not looked hard enough.