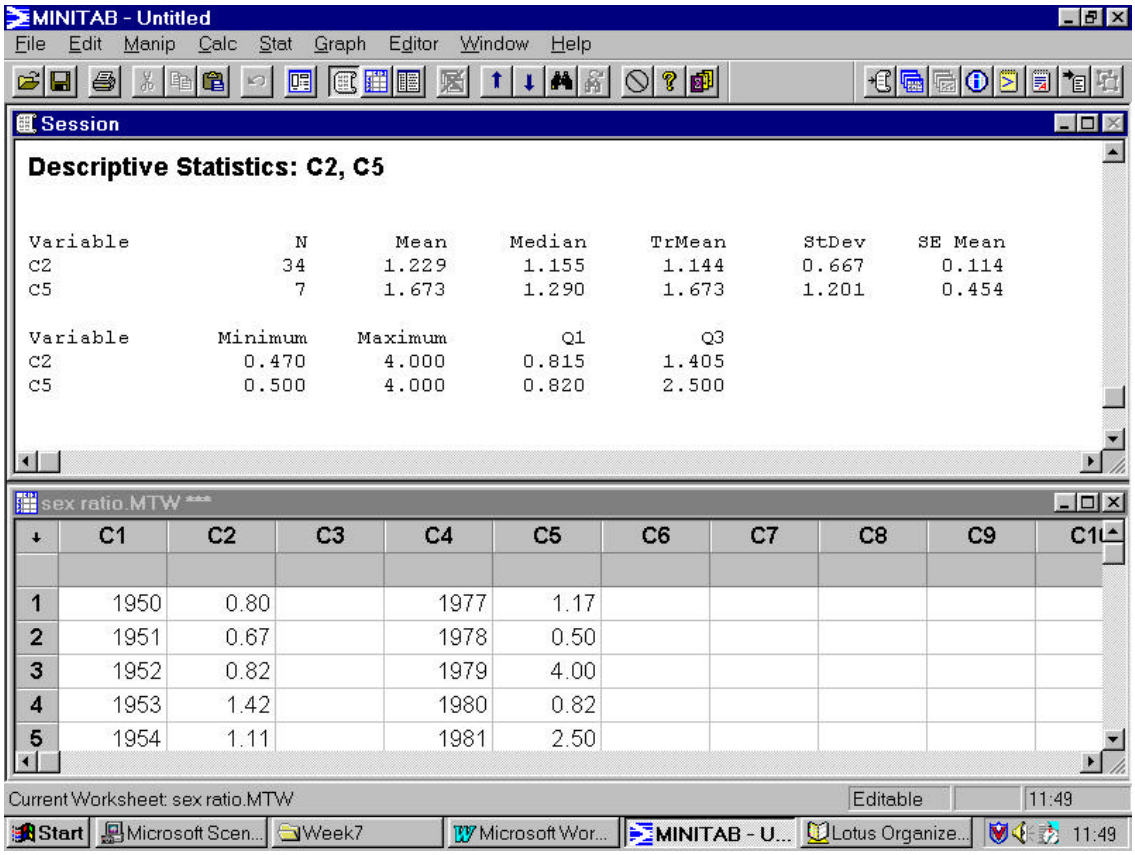
Research Methods 2

Week 7: Exercise Sheet 1

Solution sheet

Question 1

The data from the question have been typed into column C2 and the data from 1977-83 into column C5. Using **Display Descriptive Statistics** the following screen is obtained. (Note that the descriptive statistics for both columns have been found using one application of the command: this has been done by selecting first C2 and then selecting C5 before clicking on **OK**)



For the full data, the 95% confidence interval for the mean sex ratio is computed as

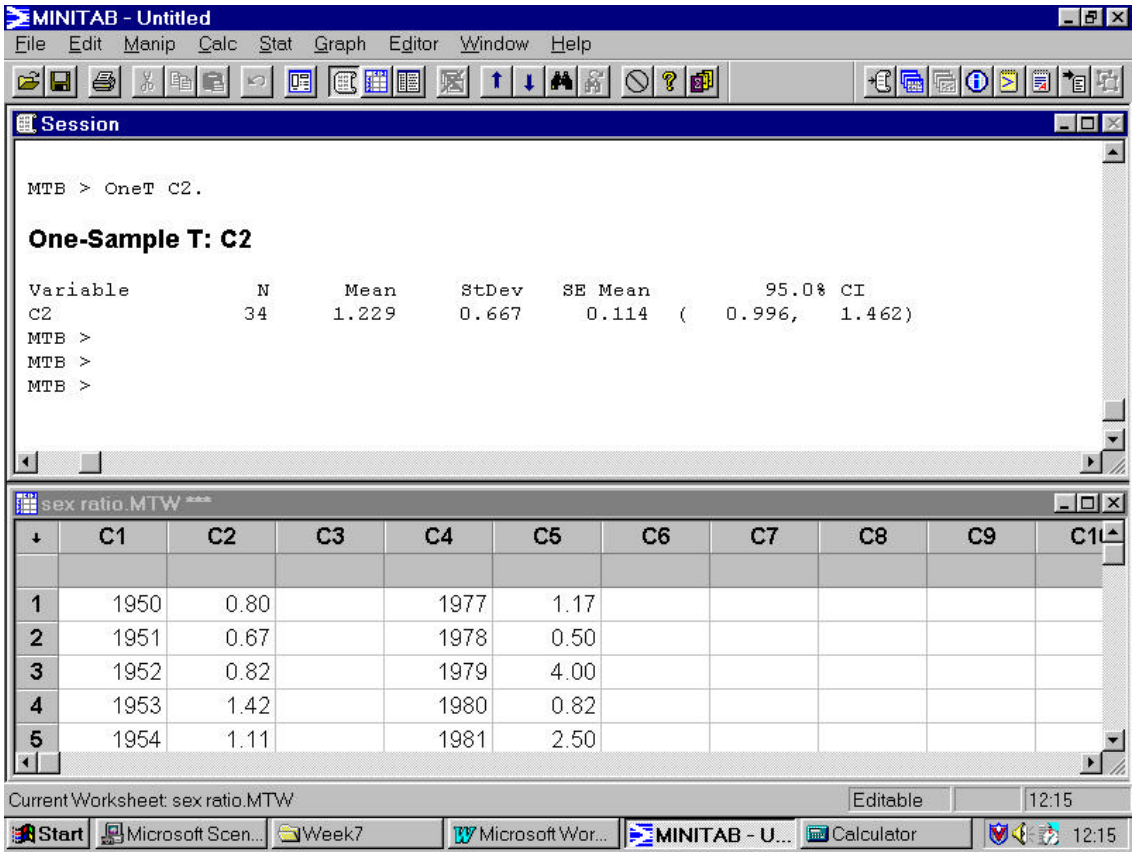
Lower limit = $1.229 - 2 \times 0.114 = 1.001$ Upper Limit = $1.229 + 2 \times 0.114 = 1.457$

For the data from 1977-83, the 95% confidence interval is found as

Lower limit = $1.673 - 2 \times 0.454 = 0.765$ Upper Limit = $1.673 + 2 \times 0.454 = 2.581$

Question 2

Clicking the sequence **Stat** -> **Basic Statistics** -> **1-sample t...** and selecting C2 into the **Variables:** box and clicking **OK** gives the following screen.



The output from the command repeats the values of the sample mean and SD and the SE and also produces the limits of the confidence interval.

Repeating the procedure with column C5 gives the interval (0.561, 2.784)

The intervals found in questions 1 and two are as follows

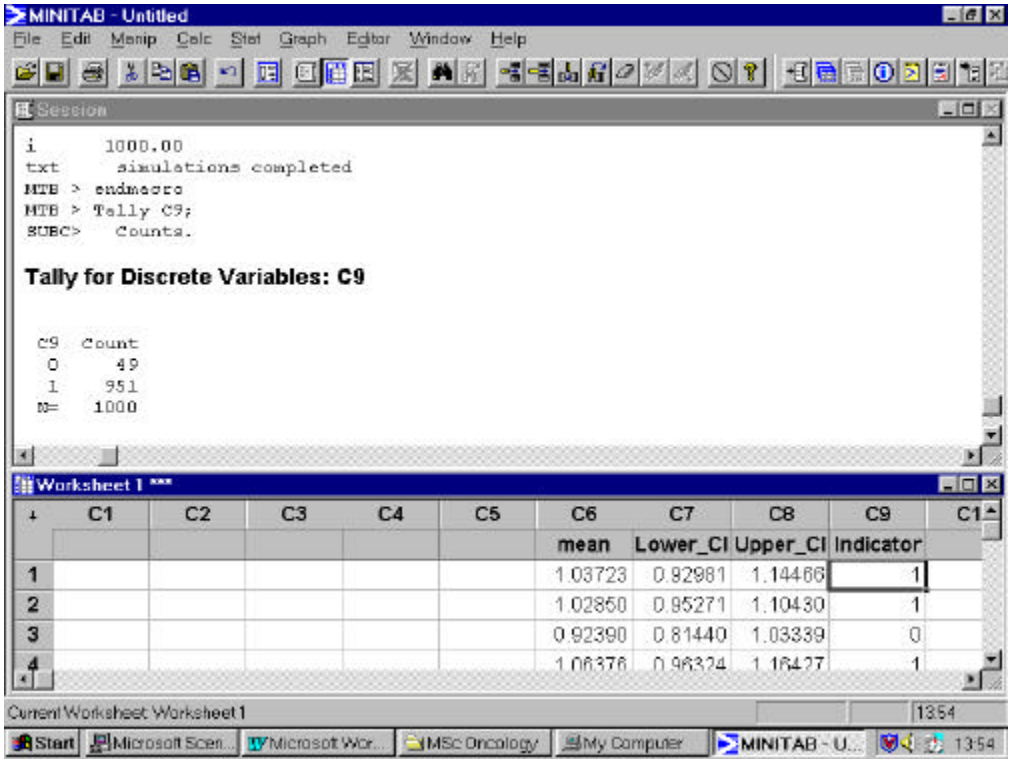
Method	Full data	Data 1977-83
Question 1	(1.001, 1.457)	(0.765, 2.581)
Question 2	(0.996, 1.462)	(0.561,2.784)

Two comments are made.

1. The intervals based on the reduced sample are much wider than those for the full sample. It is not surprising that we should have a better idea about the location of the mean of a population if we base the inference on 34 rather than just 7 points.
2. The methods in the two questions produced very similar answers when applied to the full sample. When applied to the reduced sample there is a greater discrepancy. This relates to the point made in the note 'Confidence intervals: a technicality' in the study document for this week. For small samples smaller than about 25-30 the method based on mean $\pm$ 2 $\times$ SE gives intervals that are too narrow and the exact method obtained from Minitab directly should be used.

Question 3

Entering the command in the question gives the following screen, once the Tally command has been applied to the indicator column, C9.



Note: in the Data Window the population mean, 1.05, is between the upper and lower limits in rows 1, 2 and 4 but not in row 3. Hence column C9 contains 1s on rows 11, 2 and 4 and 0 in row 3. Note also that the column names have been added for clarity

The claim for a 95% confidence interval is that on 95% of occasions the interval will include the population mean. Of course, in practice, it will not be known whether or not a given interval will include the population mean, as this quantity is unknown. It is simply that it is likely to include the population mean - indeed, with probability 95%. However, if data are generated from a known population, the population mean *is* known and for each of many similarly generated intervals it can be decided whether or not the interval includes that population mean.

From the above, the number of 1s is 951, so out of 1000 intervals, only 49 fail to include the population mean. This is close the value of 95% $\times$ 1000 = 950 intervals containing the mean.

Trying various different populations and sample sizes should give similar results. The percentage of intervals containing the population mean, using the same mean and SD above but with different sample sizes and different numbers of samples, are given below.

	Number of sample generated		
Sample size	100	1000	10000
10	0.94	0.955	0.9470
20	0.95	0.957	0.9505
100	0.98	0.943	0.9517

Question 4

The 95% confidence interval for the sex ratio, computed using the Minitab command from the whole sample is (0.996, 1.462). It is probably sensible to round this to 2 decimal places, giving (1.00, 1.46).

The sample mean (to 2 d.p.) is 1.23. If we had to estimate the mean sex ratio by a single value, then we would use this value. A more complete estimate, which acknowledges the imprecision which necessarily arises from basing our inference on a sample of 34 values, is to say that we believe the population mean sex ratio is between 1.00 and 1.46, and we have a 5% chance of being in error in this assertion.

Confidence intervals are valuable in two respects. First, we have a reasonable belief that the sex ratio is certainly not below 1 or above 1.5 (to 1 d.p.), i.e. the values outside the interval have reasonably been excluded. Second, our sample has not allowed us to pin the population mean down any better than between 1 and 1.5. In other words, a confidence interval allows us to quantify the imprecision in our knowledge in a way that an estimate based on a single value cannot do. Moreover, it is often the case that the intervals are wider than we might have imagined, i.e. our intuitions are often over-confident.

End of solution sheet