Chapter 7

Using Minitab

7.1 Introduction

Minitab is an easy–to–use statistical package which can carry out a wide variety of statistical tasks. You may use Minitab in many different course during your time as a student, for many different purposes, so it is worth putting a little effort into familiarising yourself with the basics at an early stage. That way, you will be able to adapt to more sophisticated uses of the package later on in your course. There are three basic kinds of object Minitab works with:

- **Data column**
  This is the most frequently used object type in Minitab. Columns are denoted C1, C2, C3, etc. They each store a collection of observations relating to a particular variable.

- **Constants**
  These are denoted K1, K2, K3, etc, and each stores a single number of interest.

- **Data arrays**
  These are matrices of numbers, and are denoted by M1, M2, M3, etc.

All of the data columns, constants and data arrays relating to a particular problem are stored in a working environment called a **worksheet**. Worksheets can be saved and loaded from disk, for later use. Several worksheets may be opened simultaneously, and these can be saved together as a **project**. Minitab has both a command language and a menu driven interface. This module will concentrate on the latter, which is more intuitive, especially for the beginner.

To load Minitab, login to a PC in any University computing cluster. Select Start – Programs – Statistical Software – Minitab 14 for Windows – Minitab 14.

You should now have a spreadsheet ("data window") ready to input data. In Minitab, there are two main windows; the Session window and the Worksheet window. The Worksheet allows you to view and edit the data columns of the current worksheet. It is normally empty on startup, so the first step is to load the data in. Always enter data in the white boxes – the grey boxes are for column titles. Use the arrow keys to move around the worksheet.

7.2 Hypothesis tests for the mean

7.2.1 Tests for one mean

Recall the hypothesis tests for one mean (section 2.3). Here, from a single population, we draw a single sample, and we estimate the population mean μ with $\bar{x}$. We'd then like to see how convincing a proposal for the population mean is, based on the information in our sample. Our null hypothesis is

$$H_0 \quad : \quad \mu = c.$$

The alternative could be

$$\begin{aligned} H_1 &: \mu \neq c, \\ H_1 &: \mu > c \qquad \text{or} \\ H_1 &: \mu < c. \end{aligned}$$

The test statistic we calculate, and the statistical tables we consult to obtain our p–value, depend on whether or not the population variance is known. If the population variance (σ^2) is *known*, we use the test statistic

$$z = \frac{|\bar{x} - \mu|}{\sqrt{\sigma^2/n}}.$$

If the population variance is unknown, we use

$$t = \frac{|\bar{x} - \mu|}{\sqrt{s^2/n}},$$

where s^2 is the *sample* variance.

If the population variance is known, the test statistic is called z; another name for this hypothesis test is therefore the **one–sample z test**. If the population variance is unknown, the test statistic is called t, and we consult tables of probabilities for the t–distribution to obtain our critical value; another name for this hypothesis test is the **one–sample t test**.

Test for one mean in Minitab

This example is the same as that in question 3 of the exercises in chapter 2. A company is in dispute with its workforce. The workers claim that under a new flexitime system they are working longer than the standard 37.5 hours per week. The time cards of 10 workers were selected at random and these showed the following hours worked:

$$35 \quad 40 \quad 45 \quad 41 \quad 36 \quad 37 \quad 39 \quad 38 \quad 42 \quad 32$$

Are the staff working more than a standard week?

We can do this in Minitab! Simply enter the data in column 1 of the worksheet, and then select `Stat – Basic Statistics – 1-Sample t` (we are performing a one sample t test because we are testing for one mean and the population variance is unknown). A box will appear like that in figure 7.1. Entering `C1` in the box labelled `Variables`, and 37.5 in the `Test mean` box will test the null hypothesis against the general two–tailed alternative; since we want to know if the workers are working *longer* than the standard week, we need to test the null hypothesis

$$\begin{aligned} H_0 &: \mu = 37.5 \qquad \text{against} \\ H_1 &: \mu > 37.5. \end{aligned}$$

To do this, select `Options`, and change the `Alternative` to `greater than`. Clicking `OK` twice gives the following output.

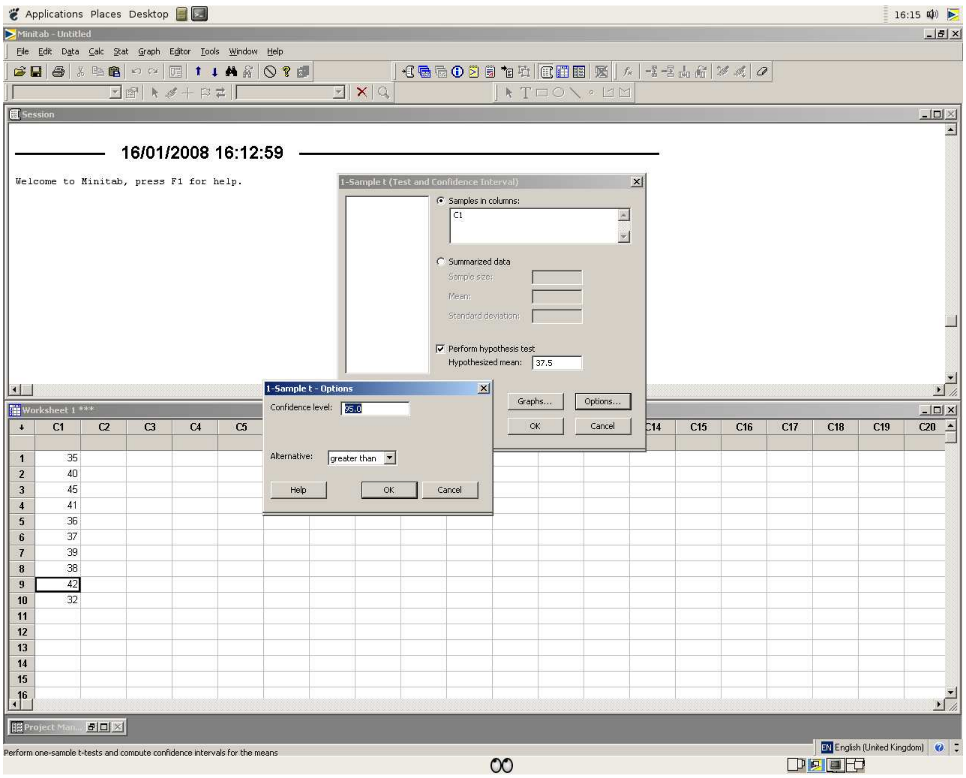

Figure 7.1: Screenshot showing the one–sample t test in Minitab

One-Sample T: Hours worked

```
Test of mu = 37.5 vs mu not > 37.5

 Variable       N  Mean  StDev  SE Mean  95% Lower Bound     T      P
 Hours worked  10  38.5   3.75     1.19            36.33  0.84  0.210
```

The `T` value is our test statistic. Notice that `Minitab` doesn't give you the critical values which we usually obtain from statistical tables; it misses this part out and gives you the *exact* p–value directly (though you need to multiply this by 100 to have it expressed as a percentage). In this example, our p–value is 0.210, or 21%, which is bigger than 10%. Thus, using table 2.1 to interpret this, there is no evidence against the null hypothesis and so we *retain* H_0. It appears that the staff *are not* working more than the standard week, which is the conclusion we reached when the question was done by hand.

Notice the benefit of using a computer package such as `Minitab` over performing the test by hand: instead of just obtaining a *range* for our p–value, we get the *exact* value for p. When the question above was done by hand a few weeks ago, we knew only that the p–value was bigger than 10% – `Minitab` gives us the exact p–value of 21%. *[Oh, and another benefit is that we don't have to use a calculator and a pen and paper!!]*

7.2.2 Tests for two means

If we have **two** independent random samples from **two** populations, we can compare the two sample means. The null hypothesis for such a test is

$$H_0 \;:\; \mu_1 = \mu_2.$$

The alternative hypothesis could be one of

$$\begin{aligned} H_1 &: \mu_1 \neq \mu_2, \\ H_1 &: \mu_1 > \mu_2 \qquad \text{or} \\ H_1 &: \mu_1 < \mu_2, \end{aligned}$$

depending on the question. The formula we use for the test statistic will depend on whether or not both population variances are known. In the case where both *are* known, we use

$$z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

If both are unknown, we use

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{s \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

where s is the "pooled standard deviation" (see section 3.2.2). Thus, similar to the one–sample case, we have a **two–sample z test** if both population variances are known, and

a **two–sample *t* test** if both are unknown. **Test for two means in** `Minitab`
The test for two means in `Minitab` is similar to that for one mean; however, this time we have two columns of data! For example, if we compare the average flight times (in minutes) of two flight companies between Newcastle and Palma, we might observe the following data:

EasyJet	150	147	141	158	155	133	140
Britannia	162	158	163	152	156	149	150

To test for a general difference in the flight times between the two companies, we test the null hypothesis

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \qquad \text{against} \\ H_1 &: \mu_1 \neq \mu_2. \end{aligned}$$

We enter these data in columns 1 and 2 of a fresh `Minitab` worksheet, and select `Stat – Basic Statistics – 2-Sample t`. We select `Samples in different columns` and enter `C1` in `Column 1` and `C2` in `Column 2`. We tick `Assume equal variances` (always tick this), and then click `OK`. Figure 7.2 shows the options in `Minitab`.

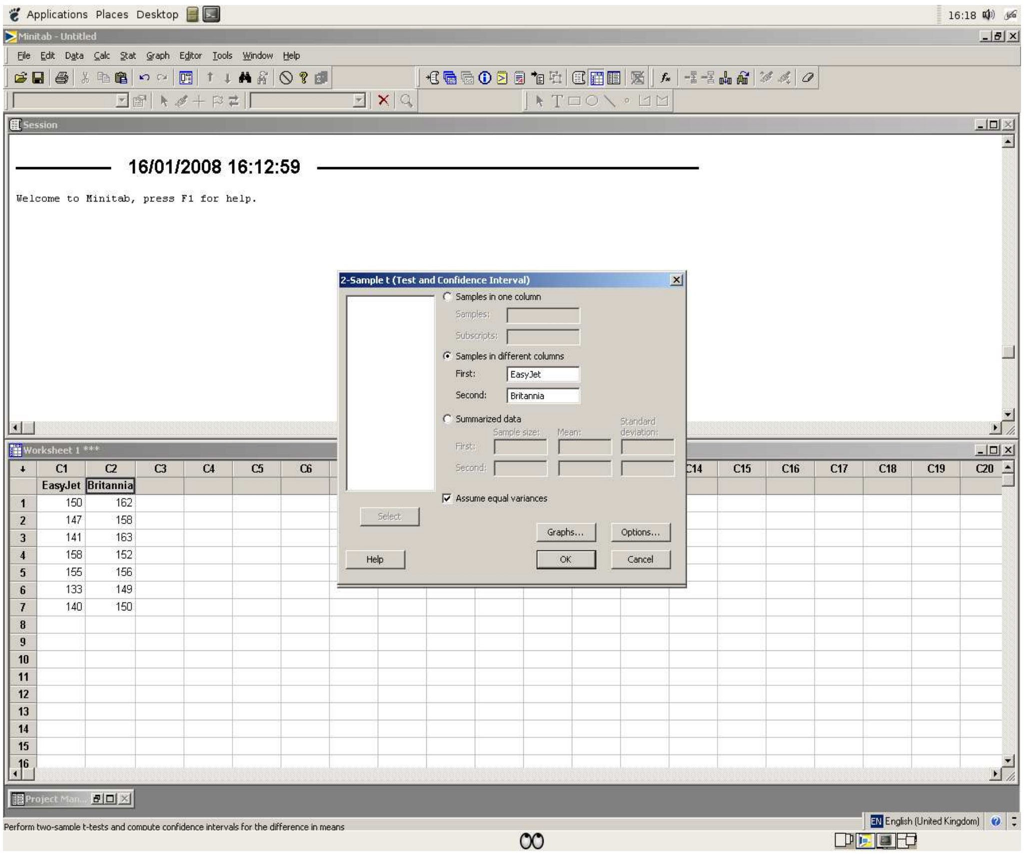

Figure 7.2: Screenshot showing the two–sample t test in `Minitab`

Two-Sample T-Test and CI: EasyJet,Britannia

```
Two-sample T for EasyJet vs Britannia

             N    Mean  StDev  SE Mean
 EasyJet     7  146.29   8.86      3.4
 Britannia   7  155.71   5.62      2.1

Difference = mu(EasyJet) - mu(Britannia)
Estimate for difference:  -9.3
95% CI for difference:  (-18.07, -0.79)
T-test of difference = 0 (vs not = 0): T-value = -2.38 P-value = 0.035 DF = 12
Both use Pooled StDev = 7.42
```

Since the p–value is 0.035, or 3.5%, it lies between 1% and 5% and so we have moderate evidence to reject the null hypothesis in favour of the alternative. It appears the two flight companies *do* have different flight times on this route (and Britannia's look longer).

7.3 Tests of independence using the χ^2 distribution

We can also perform test of independence in `Minitab` (see chapter 5 of these notes). Going back to the employment status and gender example, we have the following contingency table of observed frequencies:

	Permanent	Temporary	Unemployed	Total
Male	100	33	25	158
Female	90	40	22	152
Total	190	73	47	310

The aim here was to test for an association between the two categorical variables. The null and alternative hypotheses are

H_0 : There is no association between employment status and gender versus
H_1 : There *is* an association between employment status and gender.

We simply enter the table the way it is into empty columns of a `Minitab` worksheet (excluding column totals), as shown in figure 7.3. Selecting `Stat` – `Tables` – `Chi-Square Test` will calculate all the expected frequencies and perform the test for us!

From the `Minitab` output, we see that our p–value is 0.529, or 52.9%. Thus, we retain the null hypothesis; it appears that there is no association between employment status and gender (again, exactly the same as when we performed this test by hand; compare `Minitab`'s calculations to those on pages 45–47).

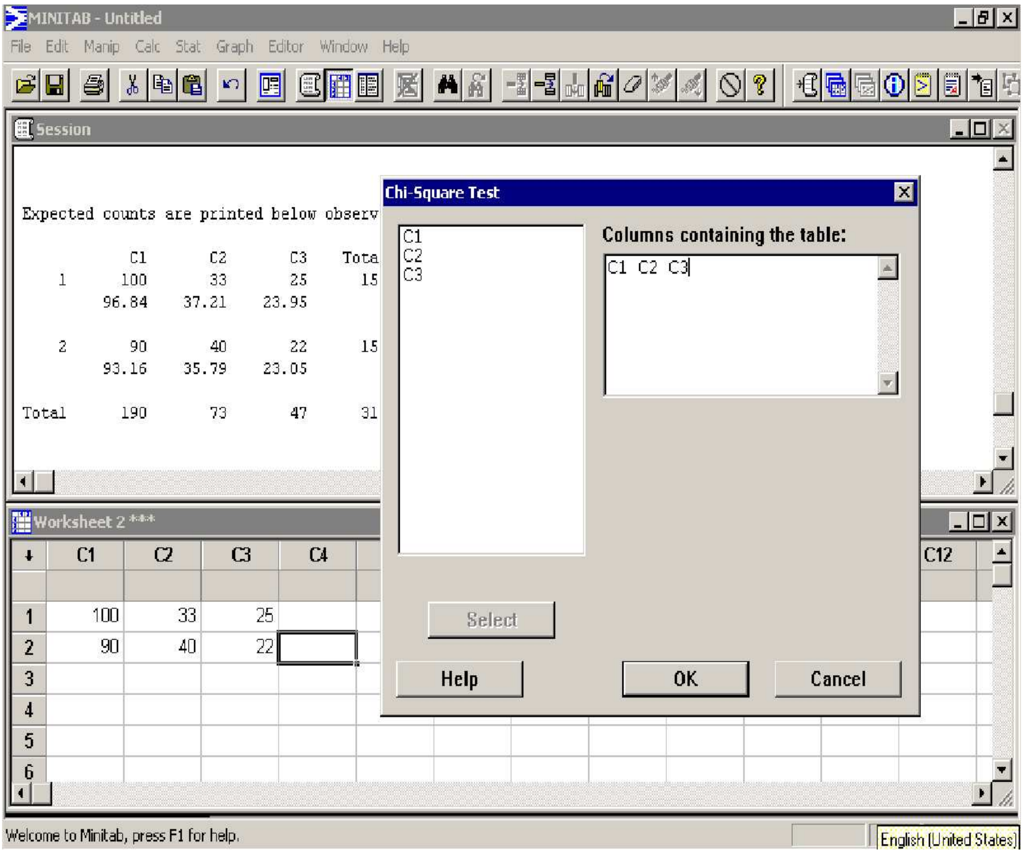

Figure 7.3: Screenshot showing the chi–squared test in Minitab

Chi-Square Test: C1,C2,C3

```
Expected counts are printed below observed counts

            C1      C2      C3   Total
     1     100      30      25     158
         96.84   37.21   23.95

     2      90      40      22     152
         93.16   35.74   23.05

 Total     190      73      47     310

   Chi-Sq = 0.103 + 0.476 + 0.046 + 0.107 + 0.494 + 0.047 = 1.273

DF = 2, P-Value = 0.529
```

7.4 Correlation and linear regression

Last week we looked at how to calculate correlation coefficients and perform a regression analysis on paired data. Well, guess what – `Minitab` can do that as well! Consider the ice cream sales data from last week. If we enter the average temperatures in column 1 of a new `Minitab` worksheet, and ice cream sales in the other, we can use `Minitab`'s correlation and regression options to do everything we did by hand in last week's lecture.

For example, to calculate the correlation coefficient, we select `Stat` – `Basic Statistics` – `Correlation` and enter `C1` and `C2` in the `Variables` box. Clicking `OK` will give the output shown below:

```
Correlations: C1,C2

Pearson correlation of C1 and C2 = 0.983
P-Value = 0.000
```

To find the regression equation, we select `Stat` – `Regression` – `Regression`, enter our Y–variable as the `Response` (i.e. the one you'd probably want to make predictions of – in this case, sales) and the X–variable in `Predictors`. `Minitab` will give quite a lot of output, most of which is beyond the scope of this course. However, the regression equation will be stated near the top of the output, and should match up with what we did by hand in last week's lecture. Try this one yourself before next week!