

MAS131/231: Introduction to Probability

Assessed Project (Semester 1, 2004/05)

Before starting the project (and before asking any questions about the project!), read all of this document very carefully (preferably twice!). Everything you need to know is explained in this document.

The data for this project represent some of the finds from an archaeological dig. The data provide information on every *limpet scoop* (a small stone tool) found in the *midden* at *Cnoc Coig* on *Oronsay*. The first and second columns give the lengths and widths (respectively) of the scoops (in centimetres), and the third and fourth the grid coordinates (East and North, respectively, in arbitrary units), of where the scoops were found on the site.

The general and specific questions of interest are:

1. How many scoops were found on the site? Are the scoops all of the same type (in terms of size), and if not, how many different types of scoops are there? How many of each type are there?
2. What are the characteristics of each scoop type? Do these types differ in only one dimension, or in both? Draw histograms, calculate summary statistics, *etc.*
3. Are the scoops all of the same general shape? In other words, is there a relationship between length and width for any or all of the different scoop types?
4. Were the different types of scoop found in different parts of the site? That is, are the different types of scoop spatially separated? Where is each type of scoop located on the site?

This project constitutes 10% of the assessment for the first Semester. Your report should present a coherent account of the problems posed, the analysis performed and the conclusions drawn — it should not be simply a list of answers. The report should be posted into the box in the Foyer by the General Office before 4pm on **Monday 15th November** — late submissions may not be marked. The report should be no more than **8 sides of A4** including all figures, tables and plots. It should be word processed, with figures integrated into the text, and should be stapled at the top left hand corner. Make sure that you spell-check your report, and that a **font size of 12 points** is used throughout. Some pointers to Microsoft Word information and documentation can be found on the module web page, along with some pointers to background information relating to the data set. Ensure that your name, your student number and the name of the data set you have analysed is clearly marked on the front of your report.

Each student has their own data to analyse. The data is simulated, but the simulation model is based on the results of an actual survey. The data sets are available on the UCS Windows XP network, in the directory `Q:\131\PROJECT\` (remember that drive `Q:` is the course data drive you should have mapped during induction week). Make sure you **analyse the data you have been assigned**. Information on which data set you should analyse **has been emailed to you** and posted on the Module newsgroup `ncl.mas.131`. These are the only ways you can find out which data set you should analyse — do not ask me or the other tutors which data set to analyse.

P.T.O.

You should analyse the data, answering the questions posed using only the data description techniques from Chapter 1 of the course. In particular, you will use the methods from Practical 1. You should not carry out any formal probabilistic or statistical analyses (*eg.* testing for differences between groups).

You should carry out the work for this project independently, in your own time. You can use the Oracle PC Cluster in Merz Court, or any other open UCS Windows XP Cluster at any time when it is not being used for teaching purposes. There will not be any special supervised sessions for this project. As a rough guide, you should expect to spend around 8–10 hours on this project (it is 10% of a 10 credit module, which is supposed to be 100 hours of work).

Note that the tutorial slots are intended mainly for discussion of the Exercises, and not for discussion of the project — the project is to be undertaken in parallel with the rest of the course, and is designed to develop your ability to work independently. If you have a general question about the project, post it to the module newsgroup. You should check the newsgroup for information related to the project — I may post extra information to the newsgroup.

Getting started

The data is stored in a plain text (ASCII) file, and not a Minitab worksheet, and hence should be imported into Minitab by entering the command:

```
MTB > READ 'Q:\131\PROJECT\PROJnnn.DAT' C1-C4
```

into the session window, immediately after starting a new Minitab session and enabling commands (see Practical 1), where *nnn* is the number you have been assigned. Please **type this command very carefully**; in particular, the ' character is that next to ; on a standard keyboard, and there is a space before the first ' and after the second '. You also need to replace the *nnn* with your number! Also, you will need to have completed the computer practical before attempting to do any of this — in particular, you will need to know how to enable and enter Minitab commands.

You should save the resulting Worksheet (or Project) to your H: drive (*My Documents*) for future use. Once you have decided on how many different types of scoop you have, and how you can categorise them (Hint: **look at the length and width distributions**), you should create an extra column in your worksheet containing an integer from 1 to the number of types you have, denoting the type of each particular scoop. Obviously, it doesn't matter which type you call 1, which you call 2, *etc.* You should create this extra column by *coding* the variable which discriminates between the groups.

For example, suppose you have numbers in column 6 (C6) of a worksheet which fall into 3 different groups: 0–10, 10–15.5 and 15.5–20. You can create a new column of groups in C7 in the following way. From the Minitab menus select Manip → Code → Numeric to Numeric... In the resulting dialog box enter C6 next to From:, and by To: enter C7. Then in the column marked Original enter 0:10 in the top cell, 10:15.5 in the cell below and 15.5:20 in the cell below that. In the column marked New, enter 1, 2 and 3 for the corresponding groups. Obviously, **you will need to adapt this example to your own data** once you have decided how to discriminate between scoop types, and how many groups you have. You will probably want to put your scoop type code in C5.

Once you have created a coded column representing scoop type, you will be able to use it as a categorical grouping variable in order to **obtain summary statistics and plots separately for each type of scoop**. If you want, you can *unstack* (see Minitab Help) the data using your scoop type variable, but this is not necessary.