University of York

Department of Health Sciences

Applied Biostatistics

Assessment 2006/2007

In this assessment we will use data from the 2006 M.Sc. questionnaire, designed and collected during the Research Methods module. Each student will be given a slightly different version of the data, obtained by dropping a few cases, so that each student's submission should be different. You can download your data set from the website on:

http://www-users.york.ac.uk/~mb55/msc/applbio/assessment/files.htm

The questionnaire was an anonymous, self-administered questionnaire completed by a convenience sample of the students and their friends and family.

The data are as they were entered into SPSS, except that heights and weights have all be put into metres and kilogrammes. The height and weight variables which you should use are hghtm and wghtkg.

Answer each question in the form of a report which could form part of a journal publication. State what method you are using and giving reasons for your choice, give results to an appropriate number of decimal places, and give conclusions based on you analysis.

The assignment should be presented as a 2,000-word report plus relevant figures and tables along with appendices showing the analysis strategy and any other relevant outputs not included in the main report, if necessary. The total should not exceed 35 pages. It should be submitted to Student Services by 16th January 2007.

The Questions

1. Produce charts (or use some other method) to show the distributions of age, height and weight, for male and female respondents separately. Briefly describe what those charts show you about the distribution of the variables.

2. Calculate the body mass index (BMI = weight in Kg divided by height in metres squared) of the respondents and find its mean and standard deviation for male and female respondents. What do these tell you about BMI?

3. How do men and women differ in weight and in BMI? (Include confidence intervals and P values.) What would you conclude?

4. How is BMI related to age? Carry out an appropriate graphical presentation and analysis.

5. Is there any evidence that a diagnosis of diabetes and a diagnosis of high blood pressure are related?

Notes

1. You should check for outliers and distributional assumptions. If you find severe outliers, you can assume that these were data entry errors, and act appropriately.

2. Answers should be brief. For example, do not spend time describing what you did in SPSS, just provide the answer. You should aim to produce something that appears similar in style to something that would appear in a medical journal, such as the *British Medical Journal* (have a look at www.bmj.com).

3. You should use 0.05 as the significance level cut-off, and 95% confidence intervals. Where appropriate, present both p-values and CIs.

4. Think very hard before you cut and paste from SPSS output. This is almost always inappropriate. You will lose marks for inappropriate material.

5. People do badly in these assignments because they do inappropriate statistics. Read your answers, and determine whether they really answer the question.

6. People don't do well in these assignments because they write too much. They worry that they must do the correct thing, and to ensure that they have done the correct thing, they do everything that they can think of. Although this increases the probability that they have done the correct thing, it means that they have certainly done a lot of things which were not correct.