Assignment 3

Due Date: 28 May, 2012
Weighting: 25%

- Answering the questions in this assignment should not be your first attempt at these types of questions. It is essential that you work through practice exercises from the tutorial sheets and/or text book first.

- This assignment is important in providing feedback and helping to establish competency in essential skills.

- Answer all the questions. The questions are not of equal weight, and some questions are worth much more than others.

- The questions relate to material up to and including Module 8.

- Before starting this assignment read *Notes Concerning Assignments* under the *Introductory Material* link on the StudyDesk.

- When you are asked to comment on a finding, usually a short paragraph is all that is required.

- Do not copy/paste SPSS output into your assignment unless specifically asked to do so In many cases the SPSS output contains much more information than is required for a correct and complete answer. In those cases just reproducing the output may not attract any marks. **Make sure you report only the information from the SPSS output relevant to your answer.**

- Unless instructed otherwise, show all working and formulae used in calculating confidence intervals and performing hypothesis tests. (Answers may of course be checked where possible using SPSS).

- In order to obtain full marks for any question you must show all working.

- This assessment item consists of 4 questions.

Question 1   (30 marks)

Use the information in the dataset heart.sav to answer the following questions. This file is available under the *Assessment/Assignments* link on the StudyDesk. For parts (c) and (d), do the calculations by hand, using a calculator and the results from part (a).

(a) Use SPSS to calculate the mean and standard deviation of BMI for the **females only** in this sample.

(b) Check the appropriate conditions and assumptions needed to calculate either a confidence interval or hypothesis test around the population mean BMI for women (include an appropriate graph to support your answer).

(c) Use the results in part (a) to estimate the population mean BMI for females in general, using a 95% confidence interval (show all working).

(d) Perform a hypothesis test to see if the population mean BMI for females is less than 26 kg/m^2. In performing this test:

   i. State appropriate hypotheses (define any symbols used).
   ii. Calculate the value of a suitable test statistic for this test.
   iii. Calculate the p -value of this test.
   iv. Write a meaningful conclusion at the 5% level of significance.

(e) Check your answers for part (d) by finding the value of the test statistic and the p-value using SPSS. [Include SPSS output in your answer.]

Question 2   (14 marks)

Note: All calculations in this question should be done by hand and calculator, not using SPSS.

For the past 18 months, sales have been falling consistently in all 1500 franchises of a fast-food chain. As part of a consulting firm you have determined that from a random sample of 120 franchises, 36 showed clear signs of mismanagement.

(a) Estimate, with 95% confidence, the population proportion of franchises suffering from mismanagement.

(b) Check the procedure you used in part (a) is appropriate by checking all the necessary conditions, assumptions and/or 'rules of thumb'.

(c) What is the minimum sample size required if we wish to estimate the population proportion of franchises suffering from mismanagement, to within plus or minus 3%, with 95% confidence? Use a conservative method in determining the sample size.

Question 3   (26 marks)

Use the information in the dataset *heart.*sav to answer the following questions. This file is available under the *Assessment/Assignments* link on the StudyDesk.

A researcher is interested in whether the average systolic BP for 60-64 year olds is higher than that of 50-54 year olds.

(a) Using SPSS and the sample information in heart.sav, determine the sample sizes, means and standard deviations of systolic BP for 50-54 year olds and then for 60-64 year olds.

(b) Use an appropriate graph to compare the distribution of systolic BP for 50-54 year olds with that for 60-64 year olds. Label the axes correctly, include unit of measure and provide an appropriate title. Include your name in the title of your graph. Briefly describe the two distributions.

(c) To answer the question 'Do 60-64 year olds have a higher systolic BP than 50-54 year olds?' perform a hypothesis test by completing the following:

   i.   State appropriate hypothesis, clearly defining all symbols.
   ii. Check the assumptions for carrying out this test.
   iii. Without using SPSS, calculate a suitable test statistic (you can use the results from part (a) in this calculation.
   iv. Without using SPSS, find the p-value of the test.
   v.   Interpret the p-value and describe the outcome of the original question.
   vi. Now use SPSS to check your results for this hypothesis test. Attach or copy and paste the relevant output from SPSS for this test to your assignment.
   vii. Briefly comment on how the test statistic and p-value from SPSS output are similar to or differ from your hand calculations.

Question 4 (30 marks)

A chemist developing insect repellents wishes to know if a newly developed formula gives greater protection from insect bites than the old formula. Eight volunteers had one arm sprayed with the old formula and the other arm sprayed with the new formula. Each volunteer then placed their arms into two chambers filled with the same number of mosquitoes for 5 minutes. The number of bites received on each arm was then recorded (see table below).

Volunteer	Old	New
1	5	3
2	7	2
3	2	4
4	4	5
5	5	3
6	5	2
7	6	3
8	4	1

Is there evidence that there are on average more mosquito bites on the arm using the old formula?

(a) Use a **parametric** test to answer this question by completing the following:
   i. State appropriate hypotheses (define any symbols used).
   ii. State (but do not check) the assumptions for carrying out this test. Describe the assumptions in the context of this question.
   iii. Calculate the value of a suitable test statistic for this test.
   iv. Calculate the p -value of this test.
   v. Interpret the p -value and describe the outcome of the test in the context of this question.
   vi. Use SPSS to carry out the analysis. Copy and paste your SPSS output to your assignment. Do these results agree with those found in part iv.? (Hint: comment on the p-value).

(b) If the assumptions for the test in part (a) were not satisfied, what alternative test could you perform? Perform this **non-parametric** test by completing the following:
   (i) State appropriate hypotheses.
   (ii) Calculate the value of a suitable test statistic for this test. Clearly explain how you calculated the test statistic.
   (iii) Calculate the p -value of this test.
   (iv) Interpret the p -value and describe the outcome of the test in the context of this question.

(c) Compare the results from part (a)v and (b)iv. Do the results agree and why might they differ?