# BM7001 Frameworks for Scientific Research

Statistics Assessment – Spring 2020

# General Instructions

These questions require you to select and run appropriate statistical tests using SPSS and to write a fully detailed commentary explaining all of your reasoning and rationale. You must include all appropriate information from the SPSS Output file for each question and EMBED the output WITHIN the commentary at the appropriate place. See 'An Introduction to SPSS Editing' document on WebLearn detailing how to do this and two references (all in Statistics Tutorial Sheets folder). The written commentary must:

* be word-processed
* ***only*** cover the points specified in the questions
* not exceed 2000 words in total

For each question detail what statistical test(s) were run, showing all relevant output generated and explain why this test(s) is appropriate? Marks will be awarded for (i) selection of correct procedures ***including those needed to verify test assumptions*** (ii) correct data entry and execution of tests (iii) overall presentation of the answer so that it contains only relevant material, in a logical order, neatly set out, without repetition.

For the commentary, marks will be awarded for (i) general presentation and use of English (ii) grasp of statistical theory and concepts (iii) understanding of the SPSS Output.

Mark allocation is as follows: Q1 20%; Q2 25%; Q3 25%; Q4 30%.

**The work must be submitted online in WebLearn by 3pm on 22nd April 2020. You are reminded about the University’s Regulations on Plagiarism and Academic Misconduct, outlined in the Module Booklet. A sample question with model answer is available on Weblearn.**

**Q1**

It is thought that a new drug X might affect muscle metabolism. An experiment was performed using 18 men volunteers, 9 of whom were randomly selected to take a pill containing drug X and 9 who were randomly chosen to take a pill which contained placebo. 2 hours later the 18 men underwent arm exercise tests to measure their respiratory exchange ratio (RER) i.e. the ratio of CO2 produced to O2 consumed. The values obtained for the RER is shown below.

**Placebo** **Drug X**

105 96

119 99

100 94

97 89

96 96

101 93

94 88

95 105

98 88

What can be deduced by statistical analysis about the effect of drug X on the RER observed in this study?

**20 marks**

**Q2**

Dataset Q2\_Spring20.xls contains the weights in pounds for an equal number of students studying for 5 different degree courses at a London University, obtained by random sampling. Using statistical analysis what can be said about the weight differences, on average for the five different courses?

**25 marks**

**Q3**

Various factors are hypothesised to be associated with the likelihood of receiving surgical treatment in localised, non-small cell lung cancer (NSCLC) patients. Data on 970 patients with primary localised NSCLC were collected from the Nebraska Cancer Registry between 2005 and 2009. Some characteristics of the patients, who either had or did not have surgery, are shown in the Table below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **No surgery (n=279)** | **Surgery (n=691)** | **Total (n=970)** |
| **Sex** |  |  |  |
| Male | 146 | 329 | 475 |
| Female | 133 | 362 | 495 |
|  |  |  |  |
| **Age (years)** |  |  |  |
| 0-65 | 28 | 220 | 248 |
| 65-75 | 80 | 278 | 358 |
| >75 | 171 | 193 | 364 |
|  |  |  |  |
| **Residency** |  |  |  |
| Rural | 146 | 334 | 480 |
| Urban | 133 | 357 | 490 |
|  |  |  |  |
| **Histologic type** |  |  |  |
| Squamous cell carcinoma | 94 | 192 | 286 |
| Adenocarcinoma | 88 | 363 | 451 |
| Large cell carcinoma | 10 | 10 | 20 |
| Other | 87 | 126 | 213 |
|  |  |  |  |
| **Comorbidity count** |  |  |  |
| No comorbidity | 32 | 170 | 202 |
| 1 comorbidity | 104 | 269 | 373 |
| >=2 comorbidity | 143 | 252 | 395 |

Using statistical analysis investigate whether there is evidence that the occurrence of surgery is related to the sex of patient, their age, place of residence, histologic type of the carcinoma or comorbidity count of the patient.

Discuss your results and any significant association found. For any significant association explain the nature of the association.

**25 marks**

**Q4**

A study was conducted to investigate whether the length of a baby could be explained by the three explanatory variables:

* Weight of the baby
* Height of the mother
* Height of the father

The file Q4\_Spring20.xls details the baby length in inches for 42 babies as well as the baby weight in pounds, the mother’s height in inches and the father’s height in inches.

Using the data contained within this file, determine by statistical analysis which variables have an impact on baby length, and the extent of the relationship.

**30 marks**