# 475.101/102/107/108
# STATISTICS

## Assignment 1, Semester 2, 2000
## Due: 4pm Thursday, 3rd August

**Instructions for handing in: PLEASE**

1. Use (standard) **A4** sized paper.
2. Number each page in the top centre and **print** your name legibly at the **top righthand corner** of each page with the surname or family name underlined.
3. Attach a *Department of Statistics Assignment Cover Sheet* to the front of the assignment. Staple or clip all pages together in the extreme top left hand corner.
4. Fold the paper lengthwise so that the printed side of the cover sheet faces out.
5. **Print** your name, paper number and assignment number on the outside of the cover sheet. **Print** your ID number on the inside of the cover sheet.
6. **Hand the assignment in** to the appropriate hand-in box in the basement of the Maths/Physics building for the city campus students or in the Tamaki Resource Centre for Tamaki campus students.

**Notes:**

- Statistics is about summarizing, analyzing and communicating information. Communication is an important part of statistics. For this reason you will be expected to write answers which clearly communicate your thoughts. **The mark you receive will be based on your written English** as well as your statistical/technical work.
- Assignment 1 will be marked out of 60 marks, 54 marks for question 1-5 as shown below and 6 marks for communication, style and presentation. Your final mark will be converted to a mark out of 10 which will be recorded towards your coursework.
- **To make your marked assignment easier to find** when they are returned, you could draw a pattern along the edges of the coversheet using coloured pens or put some sort of small sticker on the cover sheet.

**Question 1. [14 marks]**

**(a)** Read §4.6 Course Assessment on page 11 of the Course Information section of the Course Resource Book. The assignment, test and exam marks for four students from this course are shown in the table below. A final mark of at least 50% is ONE of the conditions necessary to pass the course.

| | Assignment Total (/15) | Term Test (/20) | Final Exam (/65) |
|---|---|---|---|
| Student 1 | 4 | 11 | 49 |
| Student 2 | 8 | 9 | 35 |
| Student 3 | 14 | 10 | 28 |
| Student 4 | 12 | Did not sit/no excuse | 37 |

**(i)** Which students fail to qualify for plussage?

**(ii)** Student 2 and Student 3 scored a total of 52% in their coursework and exam (assignments + test + exam). Student 2 passed the paper but Student 3 failed. Why?

**(iii)** Which student would have got a final mark of 75% if they had qualified for plussage, but instead got a final mark of 64%?

**(b)** Briefly describe how you could use a table of random numbers to generate a simple random sample of size 40 from a list of 914 people.

**(c)** Use your calculator to calculate the mean and standard deviation for the following set of data:

4.3, 1.1, 1.9, -1.3, 2.9, 8.4, -2.6, 3.3, 8.2

**(d)** Use your calculator to calculate the mean and standard deviation for the following set of data:

| $x$ | 0 | 0.8 | 1.2 | 1.9 | 3.2 | 5.6 | 8.9 |
|---|---|---|---|---|---|---|---|
| frequency | 8 | 12 | 11 | 6 | 3 | 2 | 1 |

**(e)** The following set of data is the number of minutes a player survived for a new computer game for a sample of 23 players.

8.8, 8.8, 9.1, 9.2, 9.4, 9.4, 9.4, 9.7, 9.8, 9.8, 9.9, 10.2,
10.4, 10.6, 10.7, 11.0, 11.2, 11.4, 12.2, 13.8, 14.0, 15.6, 16.1

**(i)** Calculate the five number summary for the data.
**(ii)** By hand, draw a box plot of the data. Show your calculations.

**Question 2. [12 marks]**

**(a)** A sample survey indicates that the majority of New Zealand university students prefer Burger King to McDonalds. State any four pieces of information about this survey which could help you assess the validity of the resulting statement.

**(b)** The following scheme was proposed to try and find out information about teenagers attitudes towards alcohol, tobacco and marijuana:

> A questionnaire is designed to investigate attitudes towards alcohol, tobacco and marijuana. 5 New Zealand high schools are randomly selected. From each of these high schools, 10 students are randomly selected from each of years 9 to 13 (year 9 students are usually 13 or 14 years old, with each consecutive year group being a year older). A researcher will visit each of the schools and those selected students that are present that day will be gathered together and asked to fill in questionnaires.

Discuss **four** potential sources of errors with this sampling scheme. [**Note:** Simply naming the sources of error is not sufficient to get any marks. You must discuss how or why each source of error is potentially present.]

**(c)** A common variable that is requested in surveys is *income*. Two possible questions asking about income are:

1 – What is your current income (in thousands of dollars)?

2 – Please tick the box that corresponds to your current income level:
o Under $20,000 o $20,000 - $39,999 o $40,000 or more

**(i)** For each of the above questions about income, state whether they would lead to a qualitative or quantitative variable.

**(ii)** Which of the above questions about income would give more useful information? Briefly justify your answer.

**(iii)** Which of the above questions about income would be likely to give a higher response rate? Briefly justify your answer.

**(iv)** Suppose that, one variable measured in a survey was age (to the nearest month). For each of the two questions about income, what would be an appropriate graphical tool for presenting the relationship between age and income?

**Question 3. [9 marks]**

Consider the following studies:

**Study 1:** The manager of a plastics company wishes to compare the quality of plastic champagne flutes when using raw materials from two different suppliers. The company has four plastic moulding machines of differing ages and levels of reliability. Each of the four moulding machines is used to make 50 champagne flutes using raw materials from supplier A and 50 using raw materials from supplier B. A quality score is assessed for each of the champagne flutes.

**Study 2:** A market researcher is interested in whether the response rates to consumer surveys differ between using two different incentive methods: the first where respondents are entered into a draw for a single $500 prize and the second where there are five draws each for a $100 prize. For the next survey the company conducted, 1000 households were randomly allocated to one of the two different incentive methods. The percentage of respondents for the two groups was recorded.

**Study 3:** A medical researcher is interested in whether the level of success for a particular burns treatment depends on whether the patient has suffered first or second degree burns. Over 8 weeks, patients at an accident and emergency centre who have suffered either first or second degree burns are given the treatment and followed up the next day. A success score based on alleviation of pain is assessed for each patient.

**(a)** For each study, describe what "treatment" is being compared and what "response" is being measured to compare the treatments.

**(b)** Which of the studies would be described as experiments and which would be described as observational studies?

**(c)** For the studies that are observational, could an experiment have been carried out instead? If not, briefly explain why not.

**(d)** In which of the studies has blocking been used? Briefly describe what was blocked and why it was blocked.

**Question 4. [7 marks]**

The medical profession is rich with anecdotes about surgeons, many of national and international repute, using colourful language in the operating theatre. Legend has it that the language of the mildest mannered and pious surgeon, once he or she is gowned and gloved, undergoes a transformation. A light-hearted study was carried out into the extent of foul language use by surgeons and reported in the British Medical Journal. ("*Surgeons swear when operating: fact or myth?*", F Fausto Palazzo, Orlando J Warner, British Medical Journal, 18 December, 1999.)

One hundred consecutive elective operations under general anaesthesia performed at a single hospital were assessed for the incidence of swearing by the operating surgeon. Without the surgeon's knowledge, a swearing score was recorded. Swear words were classified into three groups: 1 point for mild swear words; 2 points for bodily function swear words; and 3 points for so called four letter words. For strings of swearwords, the highest scoring obscenity alone was counted. The resulting data is presented in the table below:

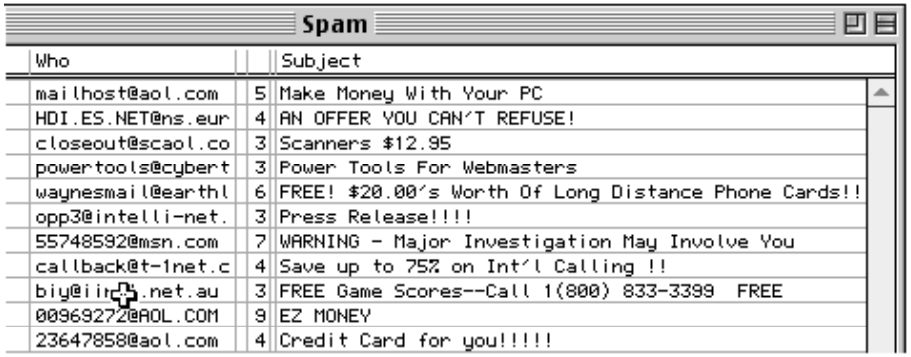
| | Ear, Nose and Throat | Orthopaedics | Urology | General Surgery | Gynaecology |
|---|---|---|---|---|---|
| Number of Operations | 19 | 23 | 20 | 29 | 9 |
| Operating Time (minutes) | 655 | 1190 | 935 | 1715 | 335 |
| Swearing Points | 2 | 41 | 6 | 38 | 7 |

**(a)** Enter the data into an *Excel* spreadsheet. On your spreadsheet, calculate the number of minutes operating time per swearing point for the five classes of surgeon.

**(b)** Following the guidelines for tabular presentation (Section 2.2.2 of the text), redraw the above table. Assume that we are primarily interested in the number of minutes operating time per swearing point.
**Note:** - You **MUST** use *Excel* to draw the table. [Refer to pages 1-4 of *Excel* notes in Course Resource book.]

**(c)** What is the biggest problem with the design of this study?

**Question 5. [12 marks]**



Spam

| Who | | Subject |
|---|---|---|
| mailhost@aol.com | 5 | Make Money With Your PC |
| HDI.ES.NET@ns.eur | 4 | AN OFFER YOU CAN'T REFUSE! |
| closeout@scaol.co | 3 | Scanners $12.95 |
| powertools@cybert | 3 | Power Tools For Webmasters |
| waynesmail@earthl | 6 | FREE! $20.00's Worth Of Long Distance Phone Cards!! |
| opp3@intelli-net. | 3 | Press Release!!!! |
| 55748592@msn.com | 7 | WARNING - Major Investigation May Involve You |
| callback@t-1net.c | 4 | Save up to 75% on Int'l Calling !! |
| biy@iin.net.au | 3 | FREE Game Scores--Call 1(800) 833-3399 FREE |
| 00969272@AOL.COM | 9 | EZ MONEY |
| 23647858@aol.com | 4 | Credit Card for you!!!!! |

A lecturer in the statistics department receives junk e-mail (also known as spam). These are unsolicited e-mails usually offering (unbelievable) deals. Records of the number of junk e-mails were received were kept for a sample of weeks preceding the February 1998 power crisis in central Auckland and for a sample of weeks after the power crisis. During the crisis the university e-mail system was out of action for a week and e-mail messages sent to the lecturer at this time were sent back with a message that they were undeliverable.

The resulting data is given below:

**Before Power Crisis**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 19 | 16 | 22 | 24 | 14 | 25 | 12 | 8 | 17 | 27 | 25 | 14 | 17 |
| 10 | 8 | 21 | 5 | 13 | 21 | 14 | 7 | 22 | 9 | 10 | 12 | 13 |

**After Power Crisis**

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 7 | 8 | 8 | 13 | 12 | 4 | 5 | 5 | 7 | 10 | 13 | 5 |
| 8 | 5 | 5 | 10 | 11 | 1 | 10 | 5 | 8 | 7 | 6 | 6 | 9 |
| 9 | 5 | 6 | 8 | 17 | 9 | 5 | 1 | 9 | 6 | 13 | 5 | |

**(a)** By hand, construct a back-to-back stem-and-leaf plot of the data.

**(b)** Enter the data into *MINITAB*. You should enter the first column of data as the number of e-mails and the second column of data as a label variable. [Refer to pages 1-5 of *MINITAB* notes in Course Resource book.]

**(c)** Produce side-by-side box plots of the data. Produce summary statistics for the two groups of data.

**(d)** Using your plots, in plain English, **briefly** comment on the data in terms of the original story.