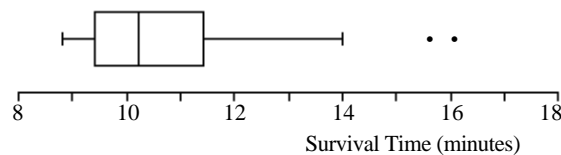


475.101/102/107/108 Semester 2 2000
Assignment 1 Solutions

Question 1.

- (a) (i) Student 1 failed to qualify for plussage as he/she got less than half marks for the assignments. (He/she got 4/15 and needed at least 7.5/15.)
 Student 4 failed to qualify for plussage as he/she did not sit the test and had no acceptable excuse.
- (ii) Student 3 scored $\frac{28}{65} \times 100\% \approx 43\%$ in the final exam. A minimum of 45% is required.
- (iii) If Student 1 was awarded plussage then their final mark would have been their exam percentage of $\frac{49}{65} \times 100\% \approx 75\%$.
- (b) Number the people on the list from 001 – 914. [Randomly choose a starting position on the tables.] Take 3 digits at a time, rejecting numbers outside the range 001 – 914. Also reject numbers that have been selected. Repeat this procedure until 40 different numbers have been selected and then select the 40 people on the list whose name corresponds with these numbers.
- (c) Mean ≈ 2.91 , Standard Deviation ≈ 3.751
- (d) Mean ≈ 1.49 , Standard Deviation ≈ 1.710
- (e) (i) Five number summary = (8.8, 9.4, 10.2, 11.4, 16.1)

(ii) Boxplot of Game Players Survival Times



IQR = 11.4 – 9.4 = 2.0
 LQ – 1.5×IQR = 9.4 – 3.0 = 6.4
 UQ + 1.5×IQR = 11.4 + 3.0 = 14.4
 15.6 and 16.1 are outside values

1.5×IQR = 3.0
 Lower whisker extends to 8.8
 Upper whisker extends to 14.0

Question 2.

- (a) The following are additional pieces of information that would help judge the reliability of the survey results:
- what form did the survey take (e.g. face-to-face interview, phone survey, internet, write-in).
 - what questions were asked in the survey.
 - how many people were surveyed.
 - what locations were surveyed.
 - how were respondents chosen.
 - how many people chose not to answer the survey or could not be contacted.
 - when was the survey carried out.
 - what were the actual results of the surveys (the specific proportions give each answer).
 - what is the margin of error of the survey.
 - who conducted the survey.

- (b) The following are potential sources of error:
- Sampling error:** Random variation which is present in all schemes that involve sampling.
- Selection bias:** Only high school students are eligible to be selected. Older teenagers that have left high school cannot be interviewed. Also, only five schools were surveyed. They are unlikely to form a representative sample of teenagers from throughout New Zealand.
- Behavioural considerations:** Some of the questions ask about sensitive issues (such as marijuana use) and some students may give answers based on how they want to appear rather than on reality.
- Non-response bias:** Selected students that were not at school at the time of the survey could not respond. The absentee students may have different opinions in general to the other students (e.g. due to sickness, truancy or having sporting/cultural commitments).
- Question effects:** The wording of the questions will influence how people respond.
- Interviewer effects:** The attitude of the person giving the surveys to the group may influence how the students respond.
- Survey format effects:** Asking the students to fill in the survey while sitting in a group may influence how they answer. Also, it is possible that students may be worried about anonymity since the person conducting the survey watches them fill in the questionnaire.
- (c) (i) Question 1 would lead to quantitative data, question 2 will lead to qualitative data.
 (ii) Question 1 will give more detailed and therefore more useful information.
 (iii) Question 2 is more likely to give a higher response rate as people may be happier to answer with an income group rather than give a specific value.
 (iv) Age is a quantitative variable.
 For question 1, we would have age (quantitative) versus income (quantitative) so would use a scatter plot.
 For question 2, we would have age (quantitative) versus income group (qualitative) so would use either dot plots, stem-and-leaf plots or box plots, depending on sample sizes.

Question 3.

- (a) **Study 1:** Treatment is supplier. Response is quality score.
Study 2: Treatment is incentive method. Response is response rate.
Study 3: Treatment is degree of burns. Response is success score.
- (b) **Study 1:** Experiment.
Study 2: Experiment.
Study 3: Observational study.
- (c) **Study 3:** This could not be done as an experiment as it would require the experimenter to decide the level of burns a patient has. This would be immoral and unethical.
- (d) **Study 1:** Blocking was used on the moulding machines because the difference in their reliability levels was a known source of variation.

Question 4.

(a) & (b)

	Number of Operations	Operating Time (minutes)	Swearing Points	Operating Minutes per Point
Ear, Nose and Throat	19	660	2	328
Urology	20	940	6	156
Gynaecology	9	340	7	48
General Surgery	29	1700	38	45
Orthopaedics	23	1200	41	29

- (c) Only one hospital is used and consecutive operations studied. Using multiple hospitals and randomly selecting operations would lead to more representative data.

Question 5.

(a)

Stem-and-leaf plot of number of junk e-mails

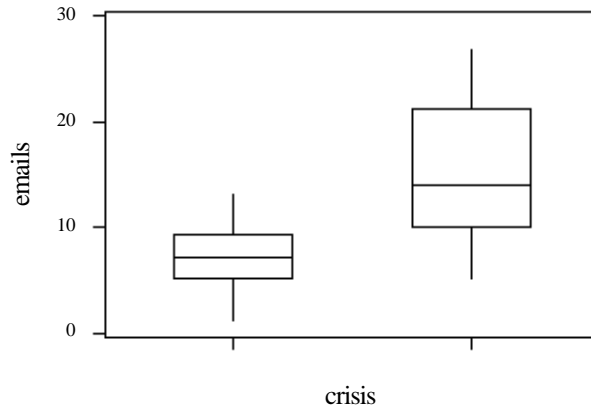
Before Power Crisis		After Power Crisis	
	0	1 1	
	0		
5	0	4 5 5 5 5 5 5 5 5	
7	0	6 6 6 6 6 7 7 7	
9 8 8	0	8 8 8 8 8 9 9 9 9	
0 0	1	0 0 0 1	
3 3 2 2	1	2 3 3 3	
4 4 4	1		
7 7 6	1	7	
9	1		
1 1	2		
2 2	2		
5 5 4	2		
7	2		

Units: 0 | 1 = 1 e-mail

(b),(c) Descriptive Statistics

Variable	crisis	N	Mean	Median	TrMean	StDev
emails	after	38	7.553	7.000	7.500	3.318
	before	26	15.58	14.00	15.54	6.29

Variable	crisis	SE Mean	Minimum	Maximum	Q1	Q3
emails	after	0.538	1.000	17.000	5.000	9.250
	before	1.23	5.00	27.00	10.00	21.25



- (d) Before the power crisis, the number of e-mails received each week varied from 5 to 27, averaging around 15.6 e-mails. After the power crisis the average number of junk e-mails per week dropped to 7.6, about half what it was. The variability of the number of junk e-mails also had a large decrease. After the power crisis most weeks had between 4 and 13 junk e-mails, with a couple of weeks getting only 1 junk e-mail and only one week with more than 13. The distribution of e-mails was slightly right skewed before the power crisis.