

**475.101/102/107/108 Semester 1 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. (They 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 0001 – 8216. Randomly choose a starting position on the tables. Take 4 digits at a time, rejecting numbers outside the range 0001 – 8216. Also reject numbers that have been selected. Repeat this procedure until 50 different numbers have been selected and then select the 50 people on the list whose name corresponds with these numbers.
- (c) The following are potential sources of error:  
**Sampling error:** Random variation which is present in all schemes that involve sampling.  
**Selection bias:** Only customers who shop between 10am and noon or between 2pm and 4pm are eligible to be selected. These customers are not likely to be representative of all customers. For example, customers who work regular hours are not likely to visit the store during this time.  
**Interviewer effect:** There will probably be differences between how the different employees conduct the surveys leading to different responses. (Especially since the interviewers used were regular workers who may or may not have appropriate skills.)  
**Behavioural considerations:** Some of the questions ask about sensitive issues (such as socio-economic level and employment) and some customers may give answers based on how they want to appear rather than on reality.  
**Non-response bias:** Some customers will not want to complete the questionnaire.  
**Question effects:** The wording of the questions will influence how people respond.  
**Survey format effects:** The use of face-to-face interviews will influence the way people answer the survey. Different responses may result if a different format (such as telephone or postal surveys).
- (d) The results are not very reliable. Only 10% of questionnaires were returned so there was a 90% non-response rate. Principals that felt more strongly about the changes or stress are more likely to respond the survey. The results from the 10% of principals who replied are unlikely to be representative of all principals.
- (e) Mean  $\approx 2.57$ , Standard Deviation  $\approx 3.358$

**Question 2.**

- (a) **Study 1:** Treatment is hair colour. Response is pain tolerance score.  
**Study 2:** Treatment is baking temperature. Response is impact strength.  
**Study 3:** Treatment is diagnostic machine. Response is service time.  
**Study 4:** Treatment is gender. Response is grade.
- (b) **Study 1:** Observational study.  
**Study 2:** Experiment.  
**Study 3:** Experiment.  
**Study 4:** Observational study.

- (c) **Study 1:** This could not be done as an experiment as it would require the experimenter to decide the (natural) hair colour of the subjects.  
**Study 4:** This could not be done as an experiment as it would require the experimenter to decide the gender of the subjects.
- (d) **Study 2:** The only form of blinding possible would be for the technicians measuring the impact strength not to know the baking temperature that was used.  
**Study 3:** No blinding is possible.
- (e) **Study 3:** Blocking was used on the mechanics because the difference in their experience levels was a known source of variation.

**Question 3.**

- (a) **Stem-and-leaf plot for number of parking tickets**

```

3 | 2 4 4
3 | 5 6 7 7 8 8 8 8 9
4 | 0 1 1 2 2 2 3 4
4 | 6 7 7 9 9
5 | 0 2
5 | 8
6 |
6 | 8
  
```

Units: 3 | 2 = 32 parking tickets.

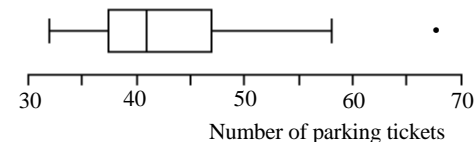
- (b) Sample size = 29. The median is the  $\frac{29+1}{2} = 15^{\text{th}}$  observation. Median = 41.

There are 14 observations below the median so the lower quartile is the  $\frac{14+1}{2} = 7.5^{\text{th}}$  observation. Lower quartile = average of 37 and 38 = 37.5.  
 Similarly, upper quartile = average of 47 and 47 = 47.

Five-number summary: (32, 37.5, 41, 47, 68)

- (c) IQR = 47 – 37.5 = 9.5                      1.5×IQR = 14.25  
 LQ – 1.5×IQR = 37.5 – 14.25 = 23.25              Lower whisker extends to 32  
 UQ + 1.5×IQR = 47 + 14.25 = 61.25              Upper whisker extends to 58  
 68 is an outside value.

**Box plot for number of parking tickets**



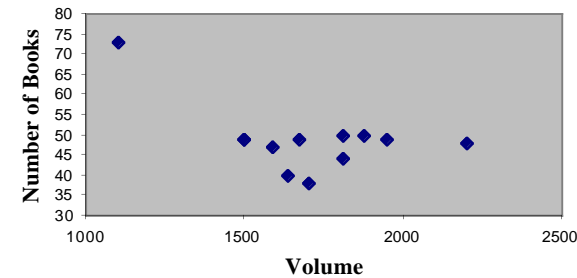
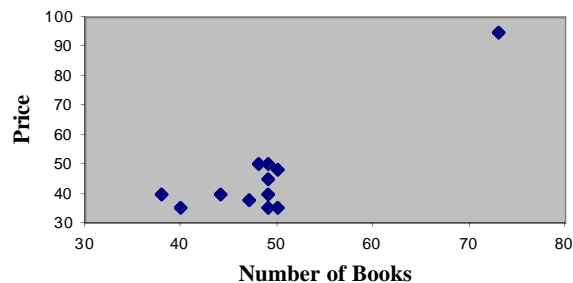
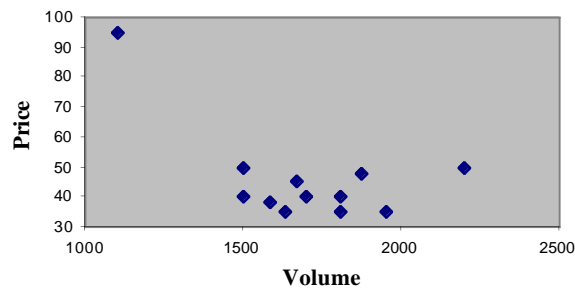
- (d) Based on this data, the number of parking tickets issued on working days is centred around 41. The distribution of the number of tickets issued is right (positively) skewed, with the bulk of the data being in the high 30's and low 40's with a few higher values. The most tickets issued on a day was 68.

#### Question 4.

(a) & (b) Information for available

Backpack Code	Volume (Cubic inches)	Number of Books	Average Volume per book	Price (US\$)
CM	1800	50	36	35
MR	2000	49	41	35
JC	1600	40	40	35
PM	1600	47	34	38
MM	1800	44	41	40
EW	1700	38	45	40
EA	1500	49	31	40
MP	1700	49	35	45
DS	1900	50	38	48
BL	2200	48	46	50
RP	1500	49	31	50
CB	1100	73	15	95
<b>Average</b>	<b>1700</b>	<b>49</b>	<b>36</b>	<b>46</b>

(c)



- (d) There does not appear to be any evidence that larger volume backpacks easily hold any more books than the smaller volume backpacks. There also does not appear to be any trend relating the volume of the backpack to price. All the more expensive backpacks (over \$40) can all hold at least 48 books, but there are also cheap backpacks that hold more than 48 books.

It is difficult to believe that the data for the CB backpack is correct as a 5-inch by 7-inch by half-inch book would have a minimum volume of  $5 \times 7 \times 0.5 = 17.5$  cubic inches, while according to the data the books in the CB backpack take up an average of only 15 cubic inches per book.

Other data that could be assessed on the backpacks to help make comparisons would include quality of workmanship, strength of materials used, ease of use and visual appeal.

Based on maximising the number of books carried and minimising the price, the CM and MR appear to be the best buys. They are the cheapest of the backpacks, but have 49-50 book carrying capacity.

#### Question 5.

##### Study 1:

The treatment is a **qualitative** variable (which category of hair colour).

The response is a **quantitative** variable (measure of pain tolerance on a scale from 0-100).

The appropriate tool for looking at the relevant relationship is plotting the data for each category against the same scale using either dot plots or box plots.

##### Study 2:

The treatment is a **quantitative** variable (baking temperature).

The response is a **quantitative** variable (impact strength).

The appropriate tool for looking at the relevant relationship is plotting the data on a scatter plot.

##### Study 3:

The treatment is a **qualitative** variable (whether or not diagnostic machine was used).

The response is a **quantitative** variable (service time).

The appropriate tool for looking at the relevant relationship is plotting the data for each category against the same scale using either dot plots or box plots.

##### Study 4:

The treatment is a **qualitative** variable (which gender: Male or Female).

The response is a **qualitative** variable (which grade: A, B, C or D).

The appropriate tool for looking at the relevant relationship is to cross-tabulate the data to form a two-way table of counts and then plot the appropriate proportions.