University of Waterloo STAT 220 – W. H. Cherry

Assignment 1

A1 – 1. Benjamin Disraeli, *1*st Earl of Beaconsfield, 1804-1881, who was British Prime Minister in 1867 and from 1874 to 1880, is quoted as saying: *There are three kinds of lies: lies, damn lies, and statistics.* Explain, in 80 words or fewer, why Disraeli might have included 'statistics' in this statement.

A1 – 2. The British author and historian, H.G. Wells, 1866-1946, wrote early in this century: *Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write.* Explain, in a total of 100 words or fewer:

(a) what you consider Wells meant by this statement;

(b) whether you consider his prediction to have come true in contemporary Canadian society.

A1 – 3. (a) Referring to Figure 2.8a in the Course Materials, tabulate, as in Table 2.8a, the results of your paper thickness investigating, preferably using the course Text as the paper source.
(If you do *not* use the Text, give a proper citation of the book you use.)

(b) Briefly describe the evidence that your measuring process has *variation.*

(c) Name an important *source* of variation in your measuring process and briefly justify your choice.

(d) Briefly describe the evidence that your measuring process is *inaccurate.*

(e) Name *two* important *sources* of inaccuracy in your measuring process and briefly justify your two choices.

A1 – 4. Referring to Figure 2.2 overleaf, Figure 2.11 and Figure 2.19 in the Course Materials:

(a) In the context of data, what is the *essential* difference between *inaccuracy* and *imprecision*?

(b) What are the three different reasons why data are inaccurate?

(c) What can be done to limit the degree of inaccuracy that can arise in each of these three ways? Explain briefly.

(d) What are the two different reasons why there is variation in data?

(e) What can be done to limit the amount of variation that arises in each of these two ways? Explain briefly.

A1 – 5. Text Exercise 1.6 (page 24): *Popular magazines often rank cities in terms of how desirable .....*

A1 – 6. Text Exercise 1.7 (page 24): *Congress wants the medical establishment to show that progress .....*

A1 – 7. Forty rats, randomly divided into two groups of 20, are used in a psychological learning experiment. First, all the rats are timed as they run a maze. Next, one group of rats is trained to run the maze and then all the rats are again timed as they run it. Finally, the *difference* between the two run times (first – second) is calculated for each rat; the results (in seconds) for the two groups of rats are shown in the tables at the right.

.....TRAINED.....			
4.0	3.2	4.1	4.9
4.2	3.7	4.3	4.2
4.4	3.6	3.5	4.9
5.1	4.5	4.7	5.0
5.6	4.6	5.2	5.5

.....UNTRAINED.....			
–2.1	–2.2	–1.1	–2.5
–1.2	2.0	–2.4	–0.6
1.3	–1.3	–0.2	–2.7
1.4	0.9	2.2	2.1
1.8	2.1	1.1	2.6

(a) Construct separate stemplots of the data for the two groups of rats; use respective sets of stems of 3, 3, 4, 4, 5, 5 and –2, –1, –0, 0, 1, 2.

(b) If the method used to train the rats is effective, which group should have the larger average difference in run times?

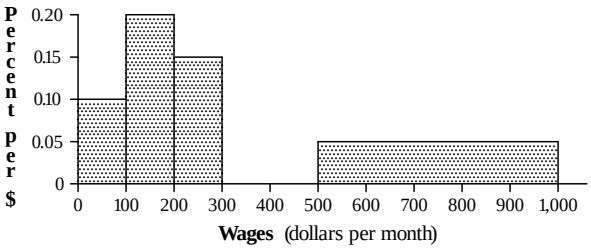
(c) What is actually occurring when the difference between the two run times for a rat is negative?

(d) Explain briefly whether you consider that the data indicate the method of training the rats really was effective.

A1 – 8. (a) Text Exercise 1.15 (pages 26-27): *The histograms in Figure 1.10 display the distributions .....*

(b) Text Exercise 1.16 (page 27): *The number of letters in a word is a measure .....*

A1 – 9. A histogram of monthly wages for part-time employees in a particular industry is shown at the right; no employee earned more than $1,000 per month. The bar over the interval '$300-$500' is missing. Calculate the height of the missing bar and **redraw** the complete histogram; show your calculations clearly.



0205-08-20

(*continued overleaf*)

A1 – 10. The table at the right shows data on the time (in hours) of jury deliberations in fifty criminal cases involving a possible death penalty.

(a) Construct a histogram for these data using six classes.

(b) On the basis of your histogram, explain briefly whether it would be unusual for a jury to deliberate for more than 59.5 hours in such a trial.

15.5	62.5	66.3	63.6	64.8	43.2
51.3	20.7	68.7	50.8	32.7	57.9
58.0	53.7	69.2	23.1	49.3	52.8
42.1	12.3	63.1	33.6	41.7	25.6
29.2	37.9	11.1	51.2	42.3	39.5
21.0	52.8	61.2	31.4	55.4	58.5
61.2	44.0	62.8	17.4	40.2	65.3
46.4	67.9	60.3	26.2	56.9	26.2
				54.0	62.1

A1 – 11. Text Exercise 1.23 (page 30): *The Degree of Reading Power (DRP) test is often used to measure .....*

A1 – 12. (a) A quality engineer in an automobile engine plant measures a critical dimension on each of a sample of crankshafts at regular intervals in the production process. The dimension is supposed to be 224 mm, but some variation will arise due to the production process. The latest measurements are given above at the right. The engineer *codes* these measurements to make them easier to work with; the coded value is the number of thousandths of a millimetre above 223 mm – for example, 224.120 is coded 1120. Give the coded value for each of the measurements in the sample.

224.120	224.001	224.017	223.982
223.989	223.961	223.960	224.089
223.987	223.976	223.902	223.980
224.098	224.057	223.913	223.999

(b) A quality engineer in an auto plant measures the eccentricity of a valve assembly (eccentricity is a measure of how far off centre two supposedly concentric circles are). The data (in inches) are shown at the right. Explain how you would code these data for easier calculation – good coding results in positive whole numbers. Then give the coded value for each of the eccentricities listed.

0.001 084	0.001 131	0.000 887
0.000 639	0.001 216	0.000 903
0.000 977	0.001 088	0.000 940
0.001 069	0.000 667	0.000 536

2005-08-20

Assignment 2 (continued)

A2 – 11. (cont.) (a) Find the values of μ and σ.

(b) What proportion of the bottles filled by the machine contain more than 760 ml?

A2 – 12. The **coefficient of variation** (CV) is defined as shown at the right, where s is the standard deviation and $\bar{y}$ is the average. The CV provides a way of comparing the variation in different data sets in situations where it is inconvenient or inappropriate to base the comparison on the standard deviations or the interquartile ranges. An example of such a situation would be a comparison of weekly milk consumption per family between Canada and the U.S., where the respective averages and standard deviations might be given as 12 and 4 litres for Canada and 3.5 and 1.2 U.S. gallons for the U.S.

$$CV = \frac{s}{\bar{y}} \times 100$$

(a) An experiment is conducted to investigate the effect of a new dog food on the weight gain in puppies during the first eight weeks of life; it finds that the average weight gain in a group of Great Dane puppies is 13.5 kg with a standard deviation of 4.5 kg; the corresponding figures for a group of Chihuahua puppies are 1.35 and 0.70 kg.

(i) Calculate the CV for each group of puppies.

(ii) Explain briefly why a direct comparison of the standard deviations is inappropriate here.

1995-04-20