

Assignment 1 — Due September 12, 2003

1. Introduction and practice with \sum notation. Consider the following two sets of data:

x:	4.47	9.75	1.41	8.26	4.17	7.63	5.28	2.66
y:	77.41	76.51	77.02	77.61	77.53	77.34	76.98	77.11

- (a) Evaluate the following:

- i. $\sum_{i=1}^8 x_i$
- ii. $\sum_{i=1}^8 x_i^2$
- iii. $(\sum_{i=1}^8 x_i)^2$
- iv. $\sum_{i=1}^8 y_i$
- v. $\sum_{i=1}^8 x_i y_i$
- vi. $(\sum_{i=1}^8 x_i)(\sum_{i=1}^8 y_i)$
- vii. $\sum_{i=1}^8 a x_i$ with $a = 7$
- viii. $a \sum_{i=1}^8 x_i$ with $a = 7$
- ix. $\sum_{i=1}^8 a$ with $a = 4$
- x. Comment on the relationship between (ii) and (iii).
- xi. Comment on the relationship between (v) and (vi).
- xii. Comment on the relationship between (vii) and (viii).

- (b) Before making any further calculations, which sample, x or y, do you think has the larger mean? Calculate \bar{x} and \bar{y} and compare.

- (c) Before making any further calculations, which sample, x or y, do you think has the larger variance? Calculate s^2 for each sample and compare.

- (d) Verify numerically that, except for rounding error, the $n = 8$ values satisfy the following:

- i. $\sum_{i=1}^8 (x_i - \bar{x}) = 0$
- ii. $\sum_{i=1}^8 (x_i - \bar{x})^2 = \sum_{i=1}^8 x_i^2 - n(\bar{x})^2 = \sum_{i=1}^8 x_i^2 - \frac{(\sum_{i=1}^8 x_i)^2}{n}$

2. The following table contains measurements of the survival time in seconds of fourth-instar milkworm larvae poisoned with sodium arsenate at a dose of .10 mg/gm.

292	309	267	245	298	241	254	256	275	226	287	280	339
298	283	366	300	315	280	240	291	286	230	285	218	270
293	244	293	285	330	284	274	307	235	225	279	280	286
314	258	215	299	241	294	311	246	215	232	346	250	308

- (a) Construct a stem and leaf display for these data, and comment on its shape.
- (b) Using your results in (a), find the sample median, range, interquartile range, and 40th percentile.
- (c) Construct a box plot for these data.

3. Below are measurements on stem volume (in cubic centimeters) of 2-year-old seedlings. One group was propagated from virus-infected buds whereas the other was propagated from healthy buds. (Bliss, p. 228)

Healthy			Infected				
1870	1324	1446	1121	408	184	16	741
1325	1759	1652	170	991	711	734	202
1364	1515	1065	893	742	335	444	

Present a useful display (displays) of these data and find the sample median, mean, standard deviation, and coefficient of variation for each group. What tentative conclusions might be drawn about the effect of the virus?

4. A University Hospital was interested in determining whether or not a new type of pain killer had any side effects of nausea. They advertised on the campus of their university for students willing to take part in the study; a small amount of cash was promised to each participant. The resultant sample was comprised of 130 students. What is the population in this case? To what extent could one extrapolate the inference to the public at large? If your goal were to make inference to the public at large, how might you change the procedure for obtaining the sample? (There is no “right” answer to this question. The purpose of the question is to get you to think about the potential difficulties that arise in determining the target population connected with a given study. Present a *brief* discussion.)

Readings:

- Course Notes: Chapters 1 and 2