Problem Set 1
Solutions

Economics 310 – Professor Menzie Chinn
Fall 2003
University of Wisconsin-Madison

1.26 a. The population of interest is the set of all adults in the Minhang District, a suburb of Shanghai, China.

b. The sample size was 3,423 + 3,593 = 7,016.

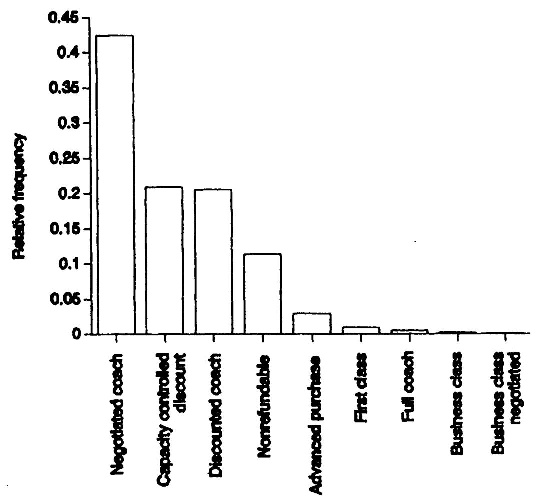
c. The study made inferences about all "people in China."

d. Since only those in the Minhang District were sampled, the results may not be characteristic of all Chinese people. The group of people surveyed was from a very small group of people in China. The people in the Minhang District may be quite different from the population of China in general.

2.6 a. The data in the table are probably from a sample. The population of all domestic airline tickets is so large that it would be extremely difficult to obtain information on all tickets.

b. The Pareto diagram is:

Almost half (.425) of all domestic airline tickets are negotiated coach. Approximately 20% (.209) of all domestic airline tickets are capacity controlled discount tickets and another 20% (.206) are discounted coach tickets. These three categories account for 84% of all domestic airline tickets.



c. Yes. The proportion of discounted tickets is .206 + .425 + .001 + .029 + .209 = .870. Thus, 87% of the tickets are discounted.

2.16 a. The highest proportion of test scores (.25) fell in the measurement class 7.5–9.5

b. The proportion of scores between 3.5 and 5.5 is .15

c. The proportion of test scores higher than 11.5 is .10 + .05 + .05 = .20

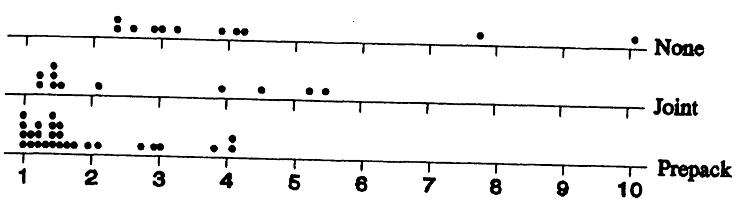
d. The proportion of the 100 students who scored less than 5.5 is .05 + .15 = .20

2.24 a. A stem-and-leaf display is as follows, where the stems are the units place and the leaves are the decimal places:

Stem	Leaves
1	⓪⓪⓪0①1②②222③④④④4444⑤⑤55⑥79
2	①144⑥7⑨9
3	⓪028⑨9
4	1112⑤
5	24
6	
7	⑧
8	
9	
10	1

b. A little more than half (26/49 = .53) of all companies spent less than 2 months in bankruptcy. Only two of the 49 companies spent more than 6 months in bankruptcy. It appears then, in general, the length of time in bankruptcy for firms using "prepacks" is less than that of firms not using "prepacks."

c. A dot diagram will be used to compare the time in bankruptcy for the three types of "prepack" firms:



d. The circled times in part a correspond to companies that were reorganized through a leverage buyout. There does not appear to be any pattern to these points. They appear to be scattered about evenly throughout the distribution of all times.

- Problem X.1

(a) $$\frac{\sum_{i=1}^{5} x_i}{\sum_{i=1}^{5} y_i} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{y_1 + y_2 + y_3 + y_4 + y_5}$$

$$= \frac{8+9+15+0+3}{5+1+(-1)+8+8}$$

$$= \frac{35}{21}$$

$$= \frac{5}{3} \approx 1.67$$

(b) $$\sum_{i=1}^{5} \frac{x_i}{y_i} = \frac{x_1}{y_1} + \frac{x_2}{y_2} + \frac{x_3}{y_3} + \frac{x_4}{y_4} + \frac{x_5}{y_5}$$

$$= \frac{8}{5} + \frac{9}{1} + \frac{15}{-1} + \frac{0}{8} + \frac{3}{8}$$

$$= \frac{-161}{40} = -4.025$$

Note that this is *NOT* the same as (a). Summing over the whole fraction gives a different result than summing over the numerator and denominator separately.

(c) $$\frac{\sum_{i=1}^{6}|x_i|}{|\sum_{i=1}^{5}y_i|} = \frac{|x_1|+|x_2|+|x_3|+|x_4|+|x_5|+|x_6|}{|y_1+y_2+y_3+y_4+y_5|}$$

$$= \frac{|8|+|9|+|15|+|0|+|3|+|-2|}{|5+1+(-1)+8+8|}$$

$$= \frac{8+9+15+0+3+2}{|21|}$$

$$= \frac{37}{21} \approx 1.762$$

It is important to be clear that the absolute value operator is *inside* the summation in the numerator, but *outside* the summation in the denominator. This makes a difference.

2.30 Assume the data are a sample. The mode is the observation that occurs most frequently. For this sample, the mode is 15, which occurs three times.

The sample mean is:

$$\bar{x} = \frac{\sum x}{n} = \frac{18 + 10 + 15 + 13 + 17 + 15 + 12 + 15 + 18 + 16 + 11}{11} = \frac{160}{11} = 14.545$$

The median is the middle number when the data are arranged in order. The data arranged in order are: 10, 11, 12, 13, 15, 15, 15, 16, 17, 18, 18. The middle number is the 6th number, which is 15.

2.40 a. the stem-and-leaf display is:

6	0	000009
8	1	25
(2)	2	45
7	3	13
5	4	0
4	5	
4	6	2
3	7	057

b. The median is the middle number once the data are arranged in order. The data arranged in order are: 0, 0, 0, 0, 0, 9, 12, 15, 24, 25, 31, 33, 40, 62, 70, 75, 77.

The middle number or the median is 24.

c. The mean of the data is $\bar{x} = \frac{\sum x}{n} = \frac{77 + 33 + 75 + \cdots + 31}{17} = \frac{473}{17} = 27.82$

d. The number occurring most frequently is 0. The mode is 0.

e. The mode corresponds to the smallest number. It does not seem to locate the center of the distribution. Both the mean and the median are in the middle of the stem-and-leaf display. Thus, it appears that both of them locate the center of the data.

2.50 a. Range = 42 − 37 = 5

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{7935 - \frac{199^2}{5}}{5-1} = 3.7 \qquad s = \sqrt{3.7} = 1.92$$

b. Range = 100 − 1 = 99

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{25{,}795 - \frac{303^2}{9}}{9-1} = 1{,}949.25 \qquad s = \sqrt{1{,}949.25} = 44.15$$

c. Range = 100 − 2 = 98

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{20{,}033 - \frac{295^2}{8}}{8-1} = 1{,}307.84 \qquad s = \sqrt{1{,}307.84} = 36.16$$

2.54 a. Range = 3 − 0 = 3

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{15 - \frac{7^2}{5}}{5-1} = 1.3 \qquad s = \sqrt{1.3} = 1.1402$$

b. After adding 3 to each of the data points,

Range = 6 − 3 = 3

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{102 - \frac{22^2}{5}}{5-1} = 1.3 \qquad s = \sqrt{1.3} = 1.1402$$

c. After subtracting 4 from each of the data points,

Range = −1 − (−4) = 3

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{39 - \frac{(-13)^2}{5}}{5-1} = 1.3 \qquad s = \sqrt{1.3} = 1.1402$$

d. The range, variance, and standard deviation remain the same when any number is added to or subtracted from each measurement in the data set.

2.68 a. From the printout, the mean is 40.0555556 and the standard deviation is 2.1770812. Both of these measures are measured in the same units as the original data, which is miles per gallon.

b. Since the sample mean is a good estimate of the population mean, the manufacturer should be satisfied. The sample mean is 40.0555556 which is greater than 40.

c. The range of the data set is 45 − 35 = 10. Using Chebyshev's Rule, the range should cover approximately 6 standard deviations. Thus, a good estimate of the standard deviation would be 10/6 = 1.67. Using the Empirical Rule, the range should cover approximately 4 standard deviations. Thus, a good estimate of the standard deviation would be 10/4 = 2.5 The given standard deviation is 2.2 which is between these two estimates. Thus, it is a reasonable value.

d. , the frequency histogram is (the relative frequency histogram would have the same shape):

```
Midpoint   Count
      35       1  *
      36       1  *
      37       3  ***
      38       3  ***
      39       4  ****
      40       9  *********
      41       7  *******
      42       4  ****
      43       2  **
      44       1  *
      45       1  *
```

Yes, the data appear to be mound-shaped.

e. Because the data are mound-shaped, we can use the Empirical Rule. We would expect approximately 68% of the data within the interval $\bar{x} \pm s$, approximately 95% of the data within the interval $\bar{x} \pm 2s$, and approximately all of the data within the interval $\bar{x} \pm 3s$.

f. The interval $\bar{x} \pm s$ is 40.056 ± 2.177 or (37.879, 42.233). Twenty-seven of the observations fall in this interval or 27/36 = .75 or 75%. This number is a little larger than 68%.

The interval $\bar{x} \pm 2s$ is 40.056 ± 2(2.177) or (35.702, 44.410). Thirty-four of the observations fall in this interval or 34/36 = .94 or 94%. This number is very close to 95%.

The interval $\bar{x} \pm 3s$ is 40.056 ± 3(2.177) or (33.525, 46.587). Thirty-six of the observations fall in this interval or 36/36 = 1.00 or 100%. This number is the same as all of the observations.

2.74 a. Since it is given that the distribution is mound-shaped, we can use the Empirical Rule. We know that 1.84% is 2 standard deviations below the mean. The Empirical Rule states that approximately 95% of the observations will fall within 2 standard deviations of the mean and, consequently, approximately 5% will lie outside that interval. Since a mound-shaped distribution is symmetric, then approximately 2.5% of the day's production of batches will fall below 1.84%.

b. If the data are actually mound-shaped, it would be extremely unusual (less than 2.5%) to observe a batch with 1.80% zinc phosphide if the true mean is 2.0%. Thus, if we did observe 1.8%, we would conclude that the mean percent of zinc phosphide in today's production is probably less than 2.0%.

2.76 a. $z = \frac{x - \bar{x}}{s} = \frac{40 - 30}{5} = 2$ (sample) 2 standard deviations above the mean.

b. $z = \frac{x - \mu}{\sigma} = \frac{90 - 89}{2} = .5$ (population) .5 standard deviations above the mean.

c. $z = \frac{x - \mu}{\sigma} = \frac{50 - 50}{5} = 0$ (population) 0 standard deviations above the mean.

d. $z = \frac{x - \bar{x}}{s} = \frac{20 - 30}{4} = -2.5$ (sample) 2.5 standard deviations below the mean.