

SOLUTIONS

35 points

Statistics 2500, Section 001
Assignment #1: Week of Sept. 21, 2009
Due date: Thursday, Oct. 1, 2009: 5pm

- The use of Minitab is **required** in the questions indicated. If you think you can use it in other questions, feel free to do so.
- Please **staple** the pages of your assignment together.
- Write your **name**, **lab instructor's name** and **day and time** you attend lab on your assignment.
- Assignments are to be passed into the assignment boxes located just to the left of the math/stats department's general office (HH-3003). Please put your assignment in the box that has your lab instructor's name on it:

Melissa (Mon. 9am, 10:30am): **BOX 1**

Vineetha (Mon. 3:30pm, Tues., Thurs. 1:50pm): **BOX 2**

Chithran (Wed. 9am, 10:30am): **BOX 3**

Hubert (Fri. 9am, 10:30am): **BOX 4**

Yunqi (Fri. 2pm): **BOX 5**

- All problem numbers are taken from the textbook *Elementary Statistics* by Bluman and Mayer.

-
1. On Sept. 21, 2009, the VOCM *Question of the Day*, which anyone is welcome to answer on their website, was the following:

Is it time for the provincial government to reconsider the issue of indexing pensions for its public sector pensioners?

At one time during the voting, 1712 voters had given responses: (37%) had answered "yes", 59% had answered "No", and 4% said "Not sure".

- (a) Explain whether or not the 1712 voters are a simple random sample of all voters in this province.
 - (b) Explain whether or not the results could be considered as representative of all voters in this province.
2. #8, p. 22.
 3. #12, p. 23.
 4. #3, p. 70. Also construct a Pareto chart by hand.

5. Refer to the data in #12, p. 40 and complete the following:
- #12, p. 40.
 - Construct a histogram by hand. Describe its shape.
 - Use Minitab to construct a histogram for this data. Does the histogram look exactly the same as the one you constructed in (b)?
6. For the data in #15, p. 67, answer the following:
- Construct a stem-and-leaf plot **by hand**.
 - Construct a stem-and-leaf plot using Minitab.
 - Use Minitab to construct a histogram. Describe the shape of the histogram. Based on its shape, would you expect the mean to be larger than, smaller than, or about the same as the median? Explain.
 - Use Minitab to find the mean, median, range and variance.
 - Use Minitab to construct a boxplot of the data.
7. #38, p. 95.
8. #36, p. 113.
9. #42, p. 113.
10. Consider the following data on the results from one of the Super 7 draws. I've changed the "free ticket" prize to \$2 for this question.

Regular Numbers	Bonus	Numbers Matched	Prize Amount	Number of Winners
5 6 16 36 37 44 46	30	7	\$10,000,000.00	1
		6 + Bonus	\$68,286.80	3
		6	\$1,495.30	137
		5	\$103.50	6,729
		4	\$10.00	143,036
		3 + Bonus	\$10.00	130,276
		3	\$2.00	1,195,440

- Find the sample mean of the prize value awarded.
Hint: If we wrote the prizes in a long list, the number 10,000,000.00 would appear once, 68286.80 would appear 3 times, down to 2 appearing 1,195,440 times.
- The price of a Super 7 ticket is \$2. Your answer in (a) is larger than \$2. Since the mean (average) prize value awarded is greater than the cost of the ticket, does this mean we should buy a Super 7 ticket? Explain.

STATS 2500-001

35 points (1)

SOLUTIONS - Assignment #1

(1) (a) This is not a random sample. In a single random sample, any group of people is as likely to be chosen as another group. This is not the case here. The people who responded to the VOCM question are probably those with strong opinions on the question, those who have easy internet access, or those who simply love VOCM (there could be other reasons). (1)

(b) Since it is not a random sample, these results could not be considered as representative of all NZ voters. (1)

2 #8, p.22.

a ~~Qualitative~~. (O.T)

e Qualitative. (O.T)

b Qualitative ~~Qualitative~~. (O.T)

f Quantitative. (O.T)

c Quantitative. (O.T)

g Qualitative. (O.T)

d Quantitative. (O.T)

3 #12, p.23.

a ~~Systematic~~ Cluster

(O.T) Each part

b Systematic.

c Random (technically, single random).

d Systematic.

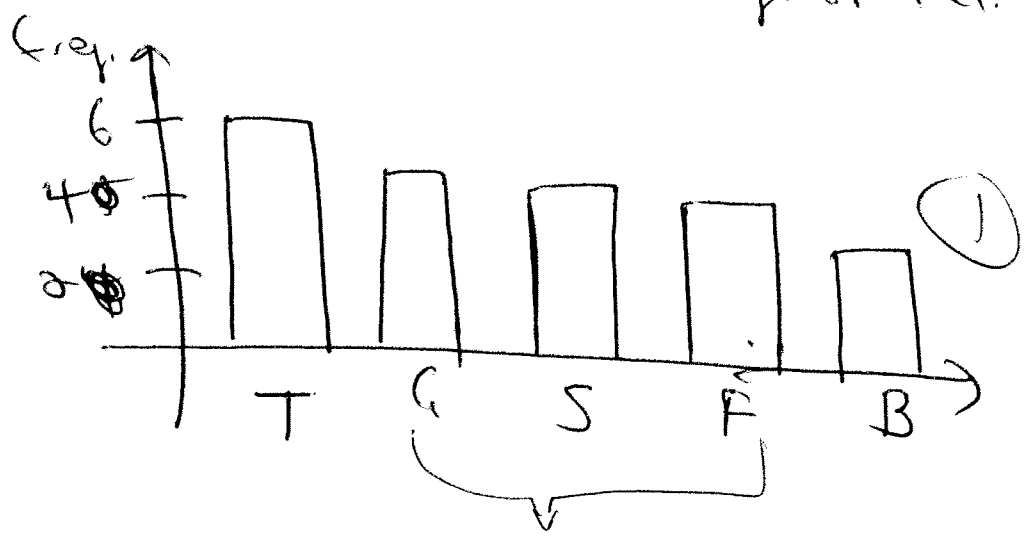
e Stratified.

4 ~~Systematic~~

④ #3, p. 70

Class	Freq.	Rel. freq.
B	4	$4/25 = .16$
G	5	$5/25 = .20$
T	6	$6/25 = .24$
S	5	$5/25 = .20$
F	5	$5/25 = .20$

Pareto chart: Can use freq. or rel. freq.



any arrangement of these 3 is fine.

(H5) p. 40, #12.

(a) Construct distribution with 4 classes, starting at 280.

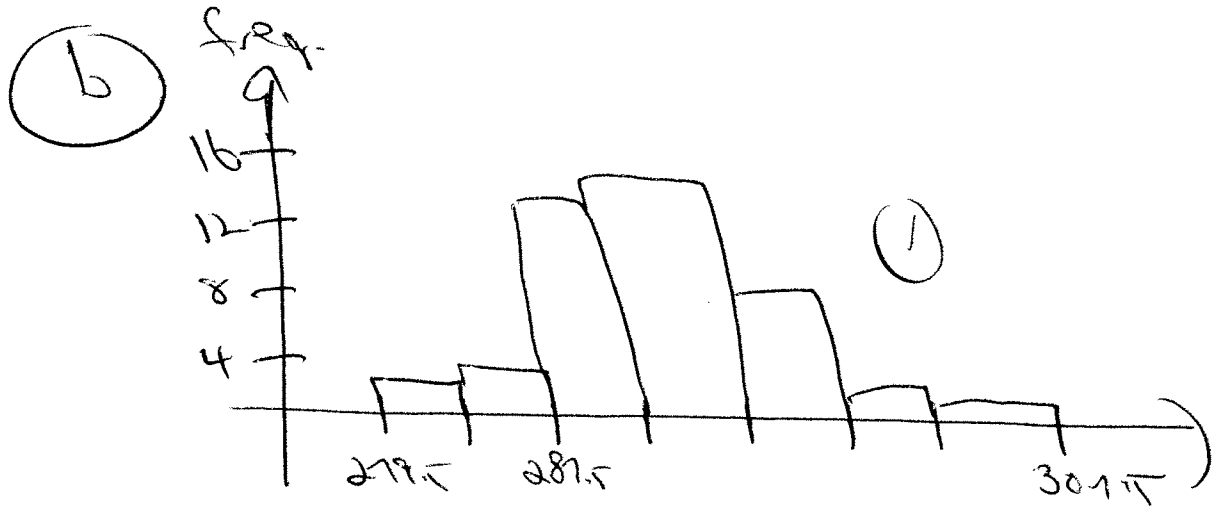
range of data: $305 - 281 = 24$
 or range: $305 - 280 = 25$.

Class Width = $\frac{25}{4} = 3.57$; could try width 3, 4 or 5.

4 will be best (width of 3 won't cover entire range of data; width of 5 will leave no observations in last class).

<u>Classes</u>	<u>Boundaries</u> (2)	<u>Freq.</u>	<u>Rel. Freq.</u>
280 - 283	279.5 - 283.5	3	3/49
284 - 287	↓	4	4/49
288 - 291		14	14/49
292 - 295		15	15/49
296 - 299		8	8/49
300 - 303		3	3/49
304 - 307	303.5 - 307.5	2	2/49

Cumulative Freq. can be included (or not).



Shape : Appears symmetric. (1)

(c) Minitab histogram is attached. (1)

It is not exactly the same as (b).

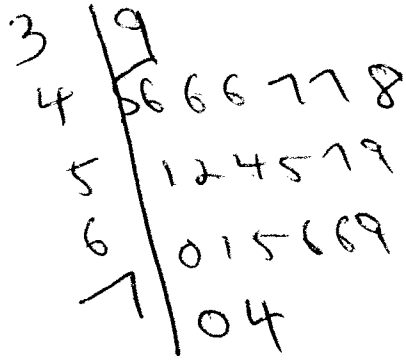
(It has more classes).

Not
needed
in
answer

This illustrates the subjective nature of constructing histograms. You want to pick enough classes to see the main features of the data. The plot in (b), and the one by Minitab, both do this.

(46) AK, p. 67

(a) Stem-and-leaf.



(2)

Given if rows not in increasing order

3|9 represents 39.

(b) Minitab stem-and-leaf is attached (1) for plot (note that it doesn't match (a) exactly. Minitab split all the values between 50 and 59 into 2 rows, etc.)

(c) Histogram is attached - (1) for plot. The shape is tricky. I would say (1) it's right-skewed. Based on the shape, I'd expect mean > median.

(d) Output is attached:
 $\bar{x} = 55.59, s = 9.76, MD = 54.5$ (1)

(e) Boxplot attached. (1)

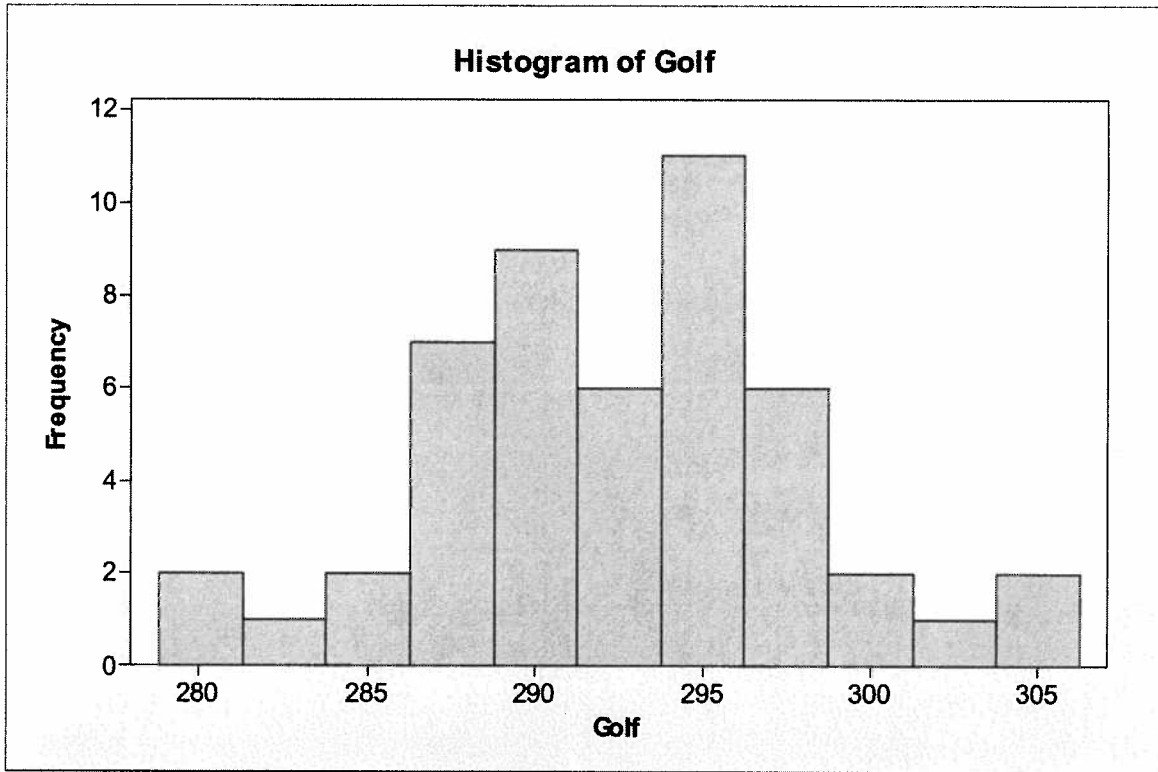
8

7

Minitab Project Report

9/30/2009 3:21:48 PM

Welcome to Minitab, press F1 for help.



5
0

Stem-and-Leaf Display: PMAge

Stem-and-leaf of PMAge N = 22
Leaf Unit = 1.0

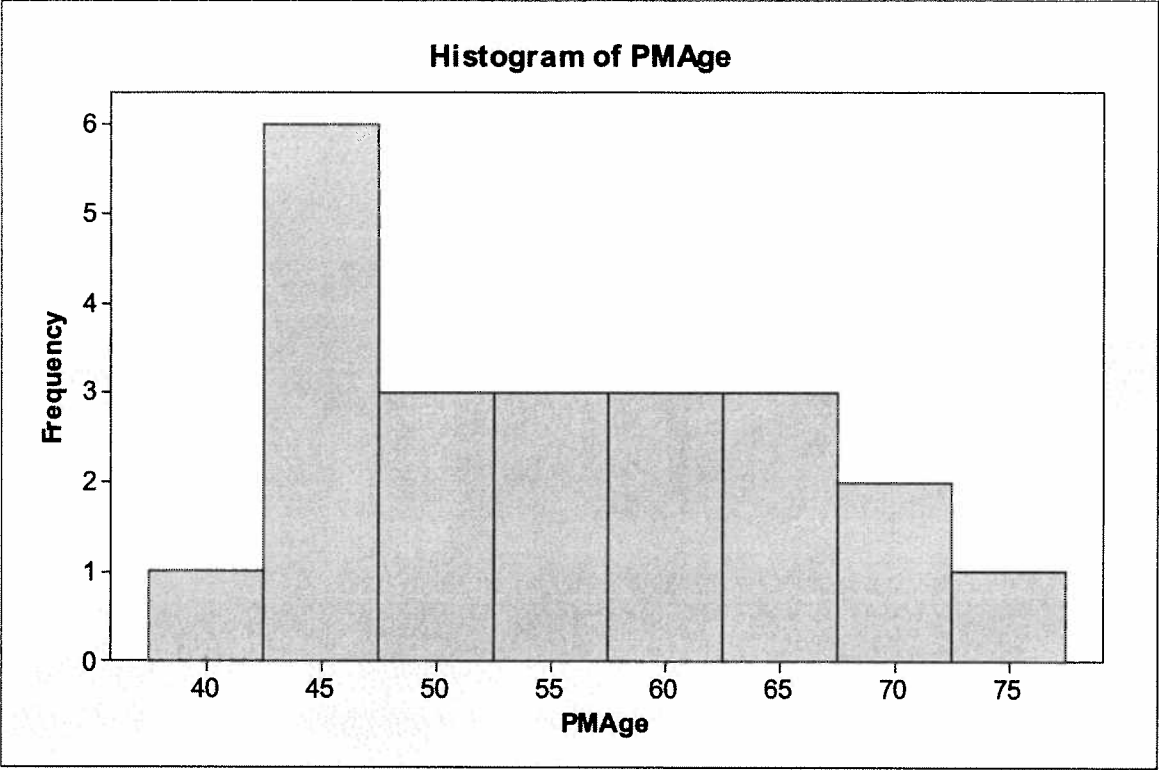
```

1  3  9
1  4
8  4  5666778
11 5  124
11 5  579
8  6  01
6  6  5669
2  7  04

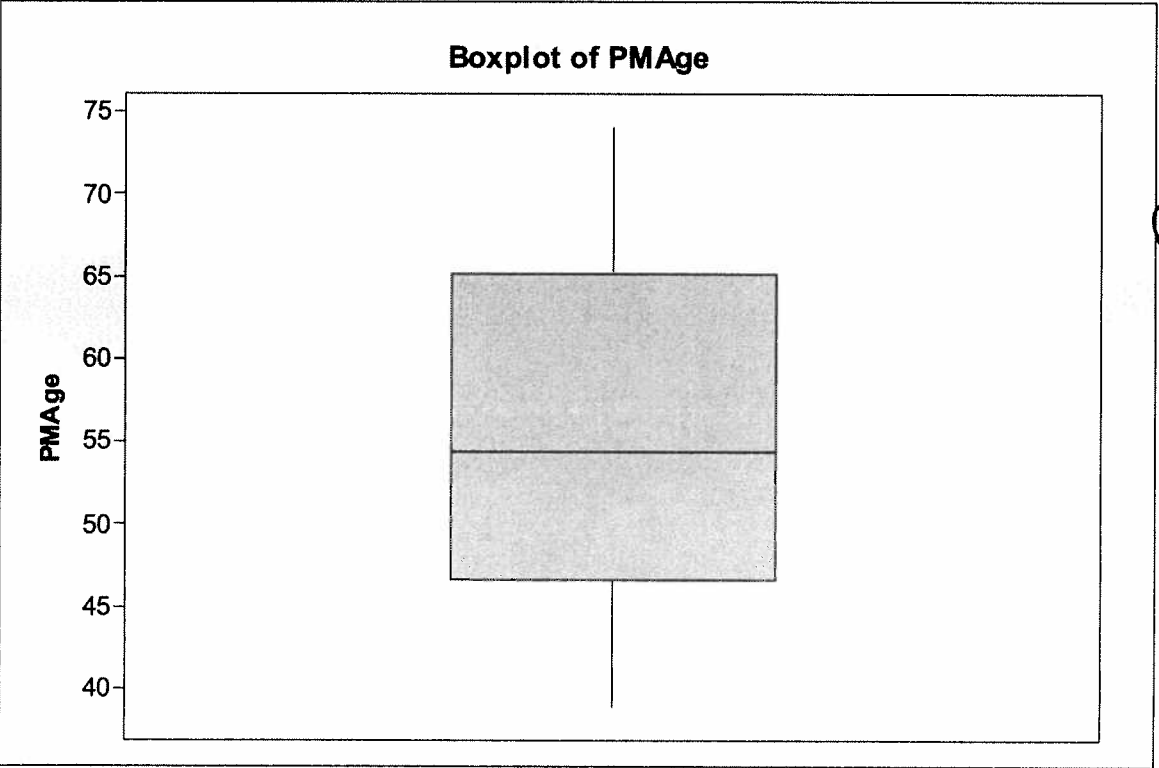
```

6 6

8



(S) (B)



(S) (B)

① # 38, p. 95

I did this in Minitab, but it could be done by hand.

① Values: 10, 20, 30, 40, 50; $\bar{x} = 30$

② ~~Subtract~~ Add 10; $\bar{x} = 40$

③ Subtract 10; $\bar{x} = 20$

④ Multiply by 10; $\bar{x} = 300$

⑤ Divide by 10; $\bar{x} = 3$

⑥ 0.5
Each.

⑦ We see that, in each case, the amount by which we changed each value is exactly the change we saw in the mean; adding 10 to each value caused \bar{x} to go up by 10, etc. ⑧ 0.5

Descriptive Statistics: PMAge

Variable	N	N*	Mean	StDev	Variance	Median
PMAge	22	0	55.59	9.76	95.21	54.50

Handwritten notes: $\bar{x} = 55$, MD, (6) d

Descriptive Statistics: Num38

Variable	N	N*	Mean	StDev	Variance	Median
Num38	5	0	30.00	15.81	250.00	30.00

(7d)

Descriptive Statistics: Num38+10

Variable	N	N*	Mean	StDev	Variance	Median
Num38+10	5	0	40.00	15.81	250.00	40.00

(7d)

Descriptive Statistics: Num38-10

Variable	N	N*	Mean	StDev	Variance	Median
Num38-10	5	0	20.00	15.81	250.00	20.00

(7b)

Descriptive Statistics: Num38x10

Variable	N	N*	Mean	StDev	Variance	Median
Num38x10	5	0	300.0	158.1	25000.0	300.0

(7d)

Descriptive Statistics: Num38/10

Variable	N	N*	Mean	StDev	Variance	Median
Num38/10	5	0	3.000	1.581	2.500	3.000

(7d)

Descriptive Statistics: PMAge

Variable	N	N*	Mean	StDev	Variance	Median
PMAge	22	0	55.59	9.76	95.21	54.50

Large handwritten scribble over the table and extending below it.

#8 #36, p.113.

$$\bar{x} = 2.60, \quad s = 0.15.$$

Find range in which at least 88.89% of data will lie.

Chebyshev's Rule: At least a proportion of $(1 - \frac{1}{k^2})$ of data falls in $(\bar{x} - ks, \bar{x} + ks)$.

Step 1: Find k .

What range with at least 88.89% of data; need k such that

$$(1 - \frac{1}{k^2}) = 0.8889$$

$$\frac{1}{k^2} = 1 - 0.8889 = 0.1111$$

$$k^2 = 9 \Rightarrow \boxed{k \geq 3} \quad (1)$$

Step 2: Find interval:

$$\begin{aligned} (\bar{x} - ks, \bar{x} + ks) &= (2.6 - 3(0.15), 2.6 + 3(0.15)) \\ &= \boxed{(2.15, 3.05)} \quad (1) \end{aligned}$$

So at least 88.89% of data falls within 2.15 and 3.05 range.

~~42~~ A42, p.113.

12

$$\bar{x} = 53, s = 2.8$$

- (a) what percentage work more than 58.6 hours per week?

Chebyshev's Rule: we have \bar{x}, s ; need k .

~~3~~ To find k , need to evaluate

$(\bar{x} - ks, \bar{x} + ks)$ for different k values until $\bar{x} + ks = 58.6$.

It turns out: $k = 2$ ①

Therefore, at least

① $(1 - \frac{1}{2^2}) = 0.75 = 75\%$ of data falls in interval

$$(\bar{x} - 2s, \bar{x} + 2s) = (53 - 2(2.8), 53 + 2(2.8)) = \boxed{(47.4, 58.6)}$$

① That means at most $(100 - 75) = 25\%$ fall outside of this range (below 47.4 or above 58.6).

No points for this part.

Strictly speaking, we can't say what proportion falls above 58.6 without making some assumptions.

If we assume bell-shaped curve, then about $25/2 = \boxed{12.5\%}$ would be above 58.6

3

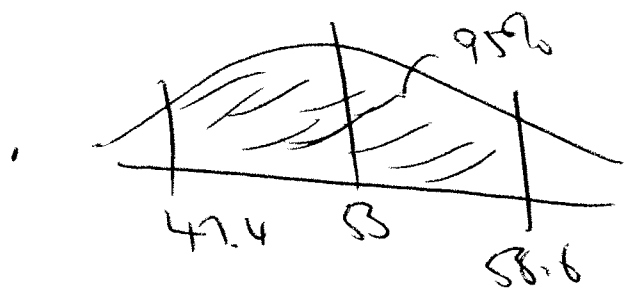
(42) b) Assume bell-shaped: Empirical Rule applies.

Try 2nd part of Rule: 95% of data fall within $(\bar{x} - 2s, \bar{x} + 2s)$.

$$\bar{x} - 2s = (47.4)$$

$$\bar{x} + 2s = 58.6 \text{ from lat.}$$

So 95% of data fall within (47.4, 58.6). (✓)



Since curve is symmetric, we'd expect about

$$\frac{5\%}{2} = \boxed{2.5\%} \text{ above } 58.6 \quad (✓)$$

10.0 Consider the following data on the results from one of the Super 7 draws. I've changed the "free ticket" prize to \$2 for this question.

Regular Numbers	Bonus	Numbers Matched	Prize Amount	Number of Winners
5 6 16 36 37 44 46	30	7	\$10,000,000.00	1
		6 + Bonus	\$68,286.80	3
		6	\$1,495.30	137
		5	\$103.50	6,729
		4	\$10.00	143,036
		3 + Bonus	\$10.00	130,276
		3	\$2.00	1,195,440

(a) Find the sample mean of the prize value awarded.

$$\bar{X} = \frac{1(10000000) + 3(68286.80) + \dots + 1195440(2)}{1 + 3 + \dots + 1195440} = \frac{16230368}{1475642} = 11.00$$

(b) The price of a Super 7 ticket is \$2. Your answer in (a) is larger than \$2. Since the mean (average) prize value awarded is greater than the cost of the ticket, does this mean we should buy a Super 7 ticket? Explain.

It looks like the average prize amount is higher than the price of a ticket, so it seems like we should always buy a ticket: spend \$2, win around \$11!

What are we missing? This argument completely ignores the millions of people who buy a Super 7 ticket, but don't win anything. (What we'd really like to see is the mean amount won for **everyone** who bought a ticket, including those who won nothing).

11

2.1