1. (2 points each) Circle either "True" or "False" for each of the following:

a) True False: Suppose that one of the variables in your data set is reaction time. Reaction time is measured in seconds. The mean, median, and standard deviation of reaction time will all be given in seconds. see p223

b) True False: The sum of the deviations of a set of observations should be zero, provided that we don't round the mean. see ex p 222 and notes from class.

c) True False: We survey a sample of people randomly chosen from a large population. Suppose that 43% of our sample answered "Yes" to the survey's "Yes / No" question. Also suppose that our survey has a margin of error of ±3% for 95% confidence. Then there is at most a 5% chance that our sample's result (43%) is more than ±3 percentage points away from the percentage of the entire population that would answer "Yes." see p181

c) True False: Confounding is increased by randomly assigning subjects to the control and treatment groups in an experiment. See p184-185

d) True False: You roll a fair, six-sided die. There will always be exactly ten rolls where a three is rolled in any sixty rolls of this die. When you flip a fair coin 10 times, do you always get 5 heads?

2. (5 points each) A jar of jelly beans has 10 licorice-flavored jelly beans, 20 orange-flavored jelly beans, and 20 lemon-flavored jelly beans. You have shaken the jar and taken a jelly bean. It's a lemon-flavored jelly bean and you don't like this flavor.

(a) You put the lemon-flavored jelly bean back in the jar (with replacement). You shake the jar and draw a jelly bean. What is the probability that you've drawn a lemon-flavored jelly bean? Show your work.

# of jelly beans in jar = 10 + 20 + 20 = 50. Picking any particular jelly bean is equally likely ($\frac{1}{50}$). There are 20 lemon jelly beans in the jar (20 ways that the jelly bean drawn will be a lemon jelly bean).

$$P(\text{lemon jelly bean}) = \frac{20}{50} = \frac{2}{5}$$

(b) Instead of putting the jelly bean back in the jar, you put it in your pocket (without replacement). You shake the jar and draw a jelly bean. What is the probability that you've drawn a lemon-flavored jelly bean? Show your work.

$$\text{\# of jelly beans in jar} = \underset{\text{licorice}}{10} + \overset{\text{orange}}{20} + \underset{\text{lemon}}{(20-1)} = 49$$

$$\text{\# of lemon jelly beans} = 19$$

$$P(\text{lemon jelly bean}) = \frac{19}{49}$$ (choosing any particular jelly bean is equally likely).

Note There is no need to convert these answers to decimals or percents. If you do, note that % means "divide the number by 100."

$$\frac{2}{5} = 0.4 = 0.40 = \frac{40}{100} = 40\%$$

$$0.4\% = \frac{0.4}{100} = \frac{4}{1000} = 0.004, \text{ not } 0.4.$$

3. Consider the following scores on a third grade reading test:

98, 73, 82, 67, 84, 89, 68, 87, 40, 93

Show your work for all calculations.

(a) (5 points) Make a stemplot summarizing this data.

4	0
5	
6	78
7	3
8	2497
9	83

↑ last digit

is OK. Can sort leafs to:

4	0
5	
6	78
7	3
8	2479
9	38

Note: 40|0 read as '400', not 40.

(b) (10 points) Which scores (if any) are outliers? What does this tell you about how the mean (or average) score and median score should compare?

The 40 is away from the rest of the scores, so is an outlier. It has no effect on the median (half the scores are below or at the median, half are at or above the median); for example, if the 40 went up to 67, the median wouldn't change. However, a low outlier pulls down the mean (sum of the scores is lower); if the 40 went up to 67, the mean would increase.

Conclusion: Expect that the mean will be smaller than the median.

(c) (10 points) Find the median score, Q_1, and Q_3. This should be done by hand (rather than having your calculator do it for you).

40 67 68 73 82 | 84 87 89 93 98

$M = \frac{82+84}{2} = 83$

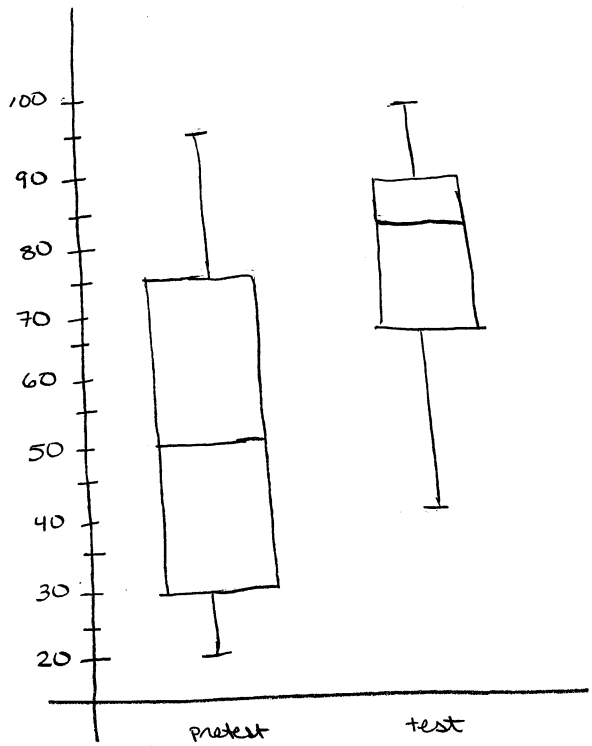
↑ Q_1 (middle obs of 5 obs below median)

↑ Q_3 (middle obs of 5 obs above median)

(d) (5 points) Calculate the mean score on this reading test. If you are entering the scores and pressing the "average" button on your calculator, also write an expression for how you would calculate the mean by hand.

$$\text{mean} = \frac{\sum x_i}{10} = \frac{98+73+82+67+84+89+68+87+40+93}{10} = \frac{781}{10} = 78.1$$

(e) (10 points) Suppose that these students also took a pretest. The five-number summary for the pretest is Min=20, Q_1=30, M=50, Q_3=75, Max=95. Draw a boxplot summarizing the results for the pretest and for the reading test. How do the two sets of results compare? Be specific.



Overall, the class did better on the test than the pretest. (For example, the low score on the test is above Q_1 on the pretest and Q_1 for the test is above the median of the pretest. And the median of the test is above Q_3 of the pretest). Also notice how much more spread out the pretest scores are than the test scores, especially in the box (Q_1 to Q_3) part of the plot.

Comment Both sets of results should be plotted on the same graph to facilitate comparison (and so that both are plotted on the same scale).

4. (5 points each) Mrs. Mason teaches at a school that has a total of 327 students. This school just received twenty children's tickets for a concert. The principal has asked Mrs. Mason to choose which students get the tickets. Mrs. Mason decided to choose the students randomly by using a table of random digits.

(a) Mrs. Mason has an alphabetical roster of all of the students at the school. How should she assign labels so that she can use a table of random digits to select which students receive the concert tickets? In particular, what label would the first child on the roster receive? What label would the last child on the roster receive?

Each label needs to have the same number of digits (see p 173). Need at least three digits:

1st child:	001
2nd child:	002
⋮	⋮
327th child:	327

(other possibilities:
1st: 000 ... last: 326
1st: 101 ... last: 427
etc)

Note If Mrs. Mason wanted to use the table more efficiently, she could assign each child three labels; for ex,

1st child:	000	327	654
2nd child:	001	328	655
⋮			
last child:	326	653	980

(Every child should be assigned the same number of labels, to give each child the same chance of being selected.)

(b) After discarding labels that didn't match any of the labels for her school's students, Mrs. Mason has read off the following twenty labels from a table of random digits:

123, 320, 179, 320, 005, 163, 230, 079, 087, 005,
323, 070, 273, 019, 001, 173, 003, 109, 154, 227

Based on this list of labels, will the first child on the roster be given a concert ticket? Is Mrs. Mason done reading labels from her table of random digits? Briefly explain your answers.

1st child gets ticket based on list? Yes, if 1st = 001. No, if 1st = 000 (ie, depends on how the labels were assigned).

Done reading labels from table? No. There are duplicate labels in this list (for example, 320 and 005). Although 20 labels were read off, there are fewer than 20 different labels — so fewer than 20 children have been chosen to receive a free concert ticket.

Note If she assigned more than one label to each child, she'd also need to be sure that the labels read off belong to 20 different children.

5. Suppose that you ask 3200 registered voters the following question:

Are you in favor of increasing taxes to pay for an after-school arts program?

All 3200 voters answered your question and 1072 of these voters answered "Yes."

(a) (10 points) The people you asked were selected randomly from all of the registered voters in Chicago. Write a press release summarizing your result about Chicago voters. Be sure to include your confidence level and margin of error.

('Yes') $\frac{1072}{3200} = 0.335 \cong 34\%$ Margin $= \frac{100\%}{\sqrt{3200}} \cong 0.018 = 1.8\%$

It looks like the election won't be bringing us a tax increase to support city after-school arts programs. A recent poll indicates that only 34% of Chicago voters are in favor of the proposed tax increase. Our poll has a margin of error of ±2% for 95% confidence.

Note The margin of error is relative to the sample percent, not to the 1072 people who said 'yes'.

(b) (5 points) The people you asked were all of the registered voters in your school district. Write a press release summarizing your result about the voters in your school district. Do you need to include a confidence level and margin of error for this result? Explain.

If the election were held today, a district-wide tax increase for after-school arts programs would be defeated. Our poll of all of the voters in our school district shows that only 34% are in favor of this proposed tax increase.

This isn't a sample; all of the voters in the district answered the question. $\frac{1072}{3200}$ will give us the population percentage (ie, we aren't using a sample percentage to estimate the population percentage). Confidence level and margin of error not needed here (we're 100% confident that the population percentage is 33.5% with a margin of error of 0%).

6. (5 points each) When rolling a pair of six-sided dice. you are twice as likely to roll a seven (*i.e.*, the numbers on the two dice sum to seven) as you are to roll a four. Let's see if this is also true when rolling a pair of four-sided dice. (The faces of a four-sided die are numbered 1, 2, 3, and 4. Each of the four faces is equally likely to be the face rolled.)

(a) What is the probability of rolling a four with a pair of four-sided dice? Show your work.

\# of possible rolls

1st die	2nd die
4 choices	4 choice

$4 \times 4 = 16$

All rolls equally likely, so $P(\text{specific roll}) = \frac{1}{16}$

\# of ways to roll a four:

1st	2nd
1	3
2	2
3	1

3 ways

$P(\text{roll a four}) = \frac{3}{16}$

(b) What is the probability of rolling a seven with a pair of four-sided dice? Show your work.

From (a) $P(\text{specific roll}) = \frac{1}{16}$

\# of ways to roll a seven:

1st	2nd
3	4
4	3

2 ways

$P(\text{roll a seven}) = \frac{2}{16} = \frac{1}{8}$

(c) With a pair of four-sided dice, are you twice as likely to roll a seven as you are to roll a four? Explain.

No. $P(\text{roll a seven}) : P(\text{roll a four}) = 2/16 : 3/16 = 2:3 \neq 2:1$

<u>Alt</u> $P(\text{roll a seven}) < P(\text{roll a four})$ since $\frac{2}{16} < \frac{3}{16}$

So rolling a seven is <u>less</u> likely than rolling a four and certainly is <u>not</u> twice as likely as rolling a four.