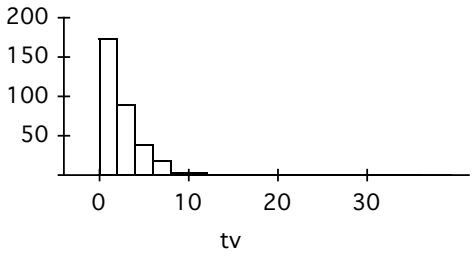
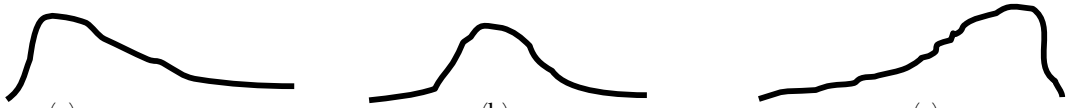
Stat 1 Fall 2008 **Homework 4** **due Thursday, Oct. 2**

Please type as much of your responses as you can. (You may write in formulas and equations by hand.)

(1) In the survey on the first day of class, you were asked how many hours of TV you watch in a typical week. A histogram of the responses is shown to the right. Is the distribution symmetric? If not, which way is it skewed? Which average would be most useful in describing the "center" of this histogram? Explain briefly.



(2) Developing countries often have a relatively low average age. For instance, Belize and Chad each have a mean age of about 24, while the US has a mean age of about 34. Would you expect a histogram of the age of the population of a developing country to look like (a), (b), or (c)? Explain briefly.



(2) Suppose the Philadelphia 76ers basketball team trades for star player Allen Iverson (who is not related to Swarthmore stat professor Phil Everson) and signs him to a contract paying an annual salary of $30,000,000. To make room for this contract under the salary cap (the maximum allowable total salary for a team in the league), the 76ers release all their other players and replace them with 11 players from the Swarthmore freshman basketball team, paying them each the league minimum salary of $19.99.
(2a) What is the mean salary of the team? How many players earn less than the mean? What is the median salary of the team?
(2b) The next season, Iverson gets a raise increasing his salary to $50,000,000, while the other players stay at the same salary. How does this affect the mean salary of the team? How does it affect the median salary?

(3) An old joke is that a certain math professor left Swarthmore to go to Haverford, thereby improving the average quality of both departments. If this statement is true, which college must have a better math department? Why?

(4) A team of seven lightweight male rowers has an average weight of 158.9 pounds. The team plans to race in an eight-man boat, so another rower is needed. If the team needs to maintain an average (mean) weight below 160 pounds, what is the maximum that this additional rower can weigh?

(5) You roll a standard die three times and calculate the SD for the three numbers that come up. What three rolls will give the largest possible SD? What three rolls will give the smallest possible SD? Explain.

(6) You make a histogram of the number of years of teaching experience for all professors at Swarthmore. Is the SD of this histogram roughly 0.7 years, 7 years, or 27 years? Explain briefly.

(7) Among entering students at a certain college, women averaged 680 on the verbal SAT with an SD of 50. Men averaged 600, but had the same SD of 50. Considering all the students together, would the SD of verbal SAT scores likely be less than 50, exactly 50, or larger than 50? (Hint: sketch a picture of the two histograms.)

(over)

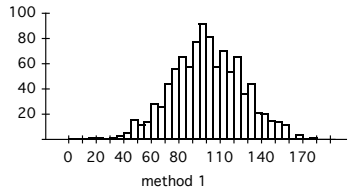
This section deals with the computer simulation we did in class on Thursday, September 25. Our goal was to estimate the number of people in a large class without having to count each person individually. As we saw in class, this problem is analogous to that of estimating the number of Nazi tanks and to estimating the extinction time of a fossil species.

We took a random sample of people from the class, and those people announced the number written on the back of their handout. We then used those numbers to estimate the size of the entire class.

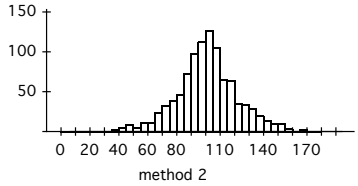
Notation: Let N represent the number of people in the entire class. Let n represent the number of people who were randomly sampled. In our class, we had twenty-five people and sampled five, so $N = 25$ and $n = 5$. A typical sample might be [44, 92, 12, 21, 56].

We compared three methods of estimating N:
Method 1: Take the average of the numbers in the sample and double it
Method 2: Add the size of the first gap to the highest number in the sample
Method 3: Add the size of the average gap to the highest number in the sample

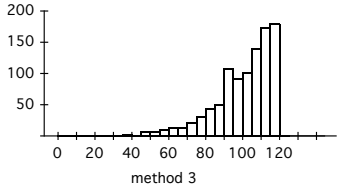
We then used the computer to take samples of size $n = 5$ from a hypothetical large class with size $N = 100$, and kept track of the estimates from each method. The sampling distributions are shown below (these results may differ slightly from what we got in class):



Method 1
Mean of sampling distribution: 100.6 people
SD of sampling distribution: 25.4 people



Method 2
Mean of sampling distribution: 100.6 people
SD of sampling distribution: 21.4 people



Method 3
Mean of sampling distribution: 100.3 people
SD of sampling distribution: 16.1 people

8. In this specific context of estimating N, explain what a sampling distribution is. Why is it important to look at the sampling distributions in comparing the three methods?

9. I claimed in class that method 3 is the "best". What are two criteria that the "best" method should satisfy?

10. Suppose the five people sampled have the numbers [44, 92, 12, 21, 56]. Using Method 1, how many people would you estimate are in the class? Why is this not a good estimate?

11. The drawback in Method 1 pointed out in question (10) can be fixed as follows:
Method 1a: Take the average of the n values in the sample and double it. This is our estimate of N. However, if this estimate is smaller than the largest value in the sample, then use that largest value as our estimate of N instead.
Using the sample given in (3) with Method 1a, how many people would you estimate are in the class?

12. Is Method 1a unbiased? Explain why or why not.

13. Is the SD of Method 1a smaller than, larger than, or the same as that of Method 1? Explain briefly.