Homework 2, due Jan. 31

This handout, plus problems 19, 35, 37, 39, and 42 from Chapter 8.

1 Sampling Distributions and the Central Limit Theorem

1. Go to `http://www.stat.tamu.edu/~west/ph/sampledist.html`.
2. The top plot shows population distribution, with summary parameters shown at left.
   - Change it from Uniform to Normal.
   - Change it from Normal to Custom and try drawing in the box.
   - Change it back to Uniform.
   - Record the mean and median of the default uniform distribution:

     mean _____ median _____
3. The second plot shows sample data from the above population, with summary statistics shown at left.
   - Set $n = 2$, $N = 1$, and check the box by Animate.
   - Click Sample five times, and record the five means below.

     _____ _____ _____ _____ _____
   - Explain why the five sample means are different.
4. The third and fourth plots show the means and the medians of the samples that have been drawn.
   - Record the mean of your five samples means (shown in 3rd plot), and the mean of your five sample medians (shown in 4th plot).

     mean of means _____ mean of medians _____
   - Why aren't these values equal to the population mean and median?
5. Go back to the second plot, and change n to 10, leaving N at 1.
   - Record the five sample means, as well as the mean and median of the sample means.

     _____ _____ _____ _____ _____

     mean of means _____ mean of medians _____
   - Are these five sample means closer to the population mean than the five sample means for $n = 2$? Is this what you expected? Why or why not?

6. Now instead of clicking sample 5 times, we'll use N to take multiple samples.
   - Turn off animation, and set N to 10, and n back to 2. Click Sample to get 10 samples.
   - Change N to 100 and click Sample again to get 100 additional samples.
   - Make a rough sketch of the distribution of sample means.

7. Before you go on, make sure you understand the difference between n and N. Experiment if you need to. Reset the bottom three plots by changing the distribution to something else, and then back to uniform.
   - Explain the difference between n and N in this applet.

8. Now make rough sketches of the distribution of sample means for n equal to 5, 10, and 30. Take as many samples as you think necessary.
   - Describe the changes in these distributions. Comment both on the shape and on the spread. Is this what you expected? Why or why not.
   - How big do you think n needs to be for the distribution of the sample means to be approximately normal?

9. Now change the population distribution to bell shaped. Experiment with different values

of n (start at $n = 3$; 2 gave me errors). Now how big do you think n needs to be for the distribution of the sample means to be approximately normal? (your answer should be 3; why?)

10. Now change the population distribution to custom, and draw in a funny shape.

    - Sketch your distribution here:

    - What is your distribution's mean, median, and standard deviation?

    - Set n to 100 and N to 100. What's the mean and standard deviation of the distribution of sample means?

    - Compare this standard deviation with the population standard deviation by calculating the ratio of sample sd to population sd. What's the ratio?_____It should be close to 0.1. Why?

    - Experiment with different values of n. How big does n need to be for the distribution of the sample means to be approximately normal?

11. Explain the central limit theorem in your own words.

2 The Normal Approximation to the Binomial

One particularly common application of the central limit theorem is the approximation to the binomial distribution.

1. For $X \sim \text{Bin}(n, p)$, state the approximate distribution of $\hat{p} = X/n$ (the sample proportion) for large n.

2. Go to `http://www.stat.tamu.edu/~west/ph/sampledist.html`, Change the population distribution to binary. The proportion of 1s is 0.5 by default.

   - Calculate the population standard deviation for $p = 0.5$. _____
   - Calculate the standard deviation of $\hat{p}$ for $n = 30$. _____
   - Set $n = 30$ and $N = 100$ (with animation off), and sample. What's the mean and standard deviation of the sample proportions (shown in the 4th plot)?

     mean _____ standard deviation _____

3. Go to `http://www.stat.tamu.edu/~west/applets/binomialdemo2.html`. If possible, open this in another window; if you have to use the same window try to remember what the plot of sample counts on the last page looked like.

   - This applet compares the true binomial distribution with the approximating normal. Set p=0.5 and n=30. Why doesn't this plot look exactly the plot of sample counts on the last page?
   - Looking at the new page, how well does the normal approximate this binomial?
   - Change p=0.1, leaving n at 30. How well does the normal approximate this binomial?
   - For very small or very large p, do you need a smaller or larger sample size than if p=0.5?

3 Comments

Please go to `http://www.stat.umn.edu/~arendahl/Teaching/Spring2007-STAT4102/03-CLT.html` (it's also linked from the homework page) and write me a comment about this worksheet. After submitting, you'll get a codeword; write it here: _____.