Math 263: Excel Assignment 5(Sampling Distributions) New Haven Firefighters

You must do the computer work for this assignment yourself, although you may certainly talk to other people. Answers which appear to be copied will be treated as an integrity violation.

1. In November and December 2003, the City of New Haven gave tests to firefighters who wanted to be promoted to one of the seven captain vacancies and eight lieutenant vacancies. In the top 15 scores on the captains' exam, there were no African-Americans and three Hispanics; the others were white. On the lieutenants' exam, the top 13 scorers were white; the 14th was African American.

The New Haven city charter requires that firefighter positions be filled by one of the top three scorers on the test, and the city is under pressure to increase diversity of its services. In March 2004, the Civil Service Commission refused to certify the results of the test and the City of New Haven decided not to use the results of the exam.

In June 2004, a group of white firefighters (including one Hispanic) sued the City of New Haven for discrimination, saying they had not been promoted because they were white. The case was turned down in September 2006 and the verdict appealed to the **2nd US Circuit Court of Appeals** in 2008.

The case received a lot of scrutiny partly because one of the three judges on the 2nd US Circuit was Sonia Sotomayor, the first Hispanic and the third woman on the US Supreme Court. (She started there in the summer of 2009.[1]) The Circuit court turned down the appeal; Judge Sotomayor supported the city of New Haven in not using the test.

To be appointed to the US Supreme Court, Judge Sotomayor had to be confirmed by the Senate.[2] During the hearings, her impartially was questioned, partly on the basis of her role in the New Haven firefighters case. Frank Ricci, the lead plaintiff in the case, testified in full firefighter uniform. He was accompanied by Ben Vargas, the Hispanic firefighter who was part of the group suing New Haven.[3]
The case went to the Supreme Court and on June 29, 2009, a 5-4 verdict in favor of the firefighters (and against the city) was handed down.

1" Description and photo:
http://www.cbsnews.com/stories/2009/10/05/national/main5363816.shtml

2" http://www.washingtonpost.com/wp-dyn/content/article/2009/07/13/AR2009071301154.html and http://www.washingtonpost.com/wp-dyn/content/article/2009/08/06/AR2009080601706.html

3" http://www.washingtonpost.com/wp-dyn/content/article/2009/07/16/AR2009071603767.html

In this assignment you will see how likely it is that the results of the exam could have happened randomly. We will look at the data from the lieutenants' exam, where the 13 top places went to whites. We want to figure out the probability that this would happen if the rankings from the exam were assigned randomly.

There are two ways to answer this kind of question: ***theory*** and ***simulation***. The theory you will see in class: it is the Central Limit Theorem. Simulation means you get the computer to draw samples randomly and see what happens; that's what you will do in this assignment. In Problem 1 you will do the simulation; in Problem 2 you will compare theory and simulation.

The data is in *FireFighters-NewHaven_Simulation.xlsx.* Go to the *Lieutenants* page.

Calculating the Population Parameters

(a) Find the number of Whites, Blacks, Hispanics among the candidates and put the results in E2, F2, G2. The race of the candidates are listed in Column B as W, B, H. You can count the numbers of each race by sorting or by using the command = COUNTIF(array, "W") or any other way you like.

Note: The counting has been started for you in cell E2. The =COUNTIF function checks in a block of cells (called the array) for a "W" or "B" or "H" or whatever you type. (The quote marks " " are necessary.) To use this function, see what has been done in E2 and copy the formula across to F2 and G2 and edit as needed.

(b) Find the proportion of each race among the candidates. Do this by dividing the counts from part (a) by the total number of candidates.

The simulation uses Excel to draw random samples of size 13. In Columns J to AL, the simulation has been started. First you will look at the simulation and see what each part is doing; then you will extend the simulation to get your own values to approximate a sampling distribution.

Note: The simulation is "live". Pressing F9 makes Excel select a new random sample.

How the Simulation is Set Up

(c) Look at K7:W7. Press F9 a few times. What are these cells doing? *Look at the formula inside a cell to see how it is done.*

(d) Look at Z7:AL7. Press F9 a few times. What are these cells doing? *Look at the formula inside a cell to see how it is done.*

(e) Now look at G7:H7. What are these cells doing? Do they change as you press F9? *Look at the formula inside a cell to see how it is done.*

Extending the Simulation to 500 Samples (Or More!)

Copy all the simulation cells downward (that is, columns J through AL) until you have at least 500 samples. (That will take you to Row 506.) Make sure it works by pressing F9. Now do the same for columns G and H.

Every time you press F9 you get a new simulation of 500 samples of size 13.

The Sampling Distribution of the Proportion of Whites in Sample

(f) For this simulation, what is the meaning of p, p, and n? (Give a verbal or numerical answer, as appropriate.)

(g) What is meant by the sampling distribution of p in this context? (Give a one sentence verbal answer.) Why is information might it give you about the firefighters' case?

Approximating the Sampling Distribution from your Simulated Data

(h) Explain why the values in Column H can be considered representative of the sampling distribution of the proportions of whites in a sample of 13.

(i) You are going to analyze the proportions in Column H. Before doing this, "freeze" a copy of your data on a new page so that you are dealing with one particular simulation. To do this, copy Column H (just this column is enough) into a *new page* in your spreadsheet using *Paste Values* from under the *Paste* menu. (Or you can use *PasteSpecial* and select *Values*.)

Note: "Paste values" pastes only the values and not the formulas that led to the values, which means this simulation will not change next time you press F9. Do this in a new page or you will kill the original simulation.

(j) What is the mean and standard deviation of the proportions in your "frozen" Column H? *These are your approximations to the mean and standard deviation of the Sampling Distribution.*

(k) Make a histogram of the values in your values in "frozen" Column H. Use bins of width 0.1 from 0 to 1. Label the histogram's axes. (Select the graph and edit.) Paste the histogram into your solutions. *This is your approximation to the Sampling Distribution.*

(l) In your "frozen" simulation, what proportion of the samples had more than 0.9 whites? *This is the proportion of values in your top bin of the histogram.*

(m) What does your answer to part (l) tell you about the results on the fire fighter's test?

2. This problem asks you to compare the approximate distribution that you generated with predictions from the Central Limit Theorem.

How Does the Simulated Sampling Distribution Compare with What Theory Predicts?

(a) What does the Central Limit Theorem (CLT) say the sampling distribution should look like? (The CLT expects the sample size to be above 30, so we may not see a very good match, but we'll look at this anyway.)

(b) What does the CLT says should be the mean and standard deviation of the sampling distribution?

(c) Sketch the sampling distribution predicted by the CLT. Label the horizontal axis and show the mean and points where graph approximately touches the horizontal axis.

(d) What does the CLT predict should be the proportion of samples with proportions of whites above 0.9?