Caveat: The appearance of a problem on this review sheet doesn't guarantee that a similar problem will appear on the exam. Similarly, the non-appearance of a kind of a problem doesn't imply that it won't appear on the exam. Last, the number of problems on this review and their difficulty is not to be taken as a statement about the exam. Enjoy!

1. Assume that the starting salaries for the Class of 2013 college graduates is (rounded, in $1000's) normally distributed with $\mu = 45$ and $\sigma = 5$. (Based on real data.)

a. Sketch and label the normal curve.

b. In what range must a starting salary be for a graduate to be in (**approximately**) the highest 16% ?

c. In what range must a starting salary be for a graduate to be in (**approximately**) the lowest .15% ?

d. In what range must a starting salary be for a graduate to be in (**approximately**) the middle 95% ?

e. What is the z-score for a graduate who has a starting salary of $42.5?

f. What % of 2013 graduates are making less than $42.5?

2. The Quickness Educational Differentiation test (the QED test) measures how long it takes middle school students to distinguish between different categories of word problems. (Low scores are good.) The QED scores are distributed on the normal curve $N(50, 10)$. Getting in the 33rd percentile or lower will certify the student as "Pretty Darn Quick" (PDQ). What score on the test is required to be classified as PDQ?

3. The Guardian, the British newspaper involved in breaking Edward Snowden's revelations, recently reported that the GCHQ—the British version of the NSA—has been monitoring webcam footage in the UK and the US. However, the practice is disconcerting the analysts because "according to GCHQ's estimate, a lot of it, a confidence interval of 3% to 11%, contains undesirable nudity." (Note: This is based on real data. I rounded the numbers to the closest integers to make the problem easier.)

a. What is the Margin of Error (ME) for the estimate?

b. What is $\hat{p}$?

c. Assume that we've been given a 99% CI and assume that the sample size for the estimate was 270 webcam feeds. Write the formulas for SE($\hat{p}$) and the ME.

d. Is it likely that the true proportion of webcam feeds with "undesirable nudity" is 14% or higher? Explain.

e. Suppose the GCHQ is unwilling to sample more webcam footage for this estimate. How can the ME be shrunk?

f. Suppose now the GCHQ is willing to increase the sample size and wants to shrink the ME to half its current size and maintain a 99% CI. What must be done to achieve this?

g. For your statistics project, you and your group decide to determine the percentage of webcam conversations among college students that contain "undesirable nudity." You imagine it will be higher than the $\hat{p}$ in the sampling done by the GCHQ. Write H_0 and H_a for the appropriate hypothesis test.

h. The Dean of Students is excited by your project (so to speak) and funds your group to do a survey of Wheaton students. You and your group members hang around Watson for a few hours and get 100 responses, with 20 students admitting to "undesirable nudity" in their webcam usage. Are the assumptions for a hypothesis test satisfied? Explain.

i. What's the $\hat{p}$ from your survey?

j. Write the formula for the 95% CI for your results.

k. Below is a hypothetical Minitab read-out from the hypothesis test. Write a sentence or two summing up the results of the hypothesis test.

```
X    N                         P-Value
20   100   0.000000000000000000000000000000000000000000000008
```

4. You are working for an advocacy group as a summer intern and your assignment is to find the mean annual heating costs for at-risk families in the Boston area. You'd like a 95% confidence interval, and you want the margin of error to be $100. A preliminary study indicates that s will likely be about $500.

a. Write a formula for the number of houses n needed to sample for the experiment.

b. Estimate the formula—don't worry about being too precise!

c. It's claimed that the average annual heating cost for at-risk families in the Boston area is $3,300. You believe it is less than that. Write H_0 and H_a for a hypothesis test.

d. You receive funding to survey 64 households about their heating costs, and you find that $\bar{y} = \$3,000$ and s = $80. Write a formula for the 99% CI.

e. In the context of this problem, explain what a 99% CI is.

f. Suppose you end up rejecting H_0 and accepting H_a , and the true average annual heating costs for Boston-area at-risk families is $3,299. Have you committed an error? If so, what type?

5. You decide to collect data on the most burning, controversial question currently facing America: What is the population proportion of people living in the United States have Vanilla as their favorite ice cream flavor?

You randomly and independently poll 100 people living in the United States. You find that 20 of them prefer Vanilla ice cream to all other flavors.

a. Are the assumptions satisfied?

b. What is $\hat{p}$ for your sample?

c. Find the critical value for a 60% confidence interval.

d. Write the formula for a 60% confidence interval. (Some things simplify nicely. You won't lose points if you don't simplify, but I'll have a warm glow of math-teacher-satisfaction if you do simplify as much as you can.)

e. A vanilla-hating cynic says that the true percentage of people living in the United States who prefer Vanilla ice cream is actually 10%. Write the H_0 and the H_a for a hypothesis test using your data.

f. What is the z-score from your data?

g. What is the P-value?

h. What is the outcome of your hypothesis test? Write a conclusion.