Math 216
Problem Set 6

1. Stainless steels can be susceptible to stress corrosion cracking under certain conditions. A materials engineer is interested in determining the proportion of steel allow failures that are due to stress corrosion cracking.
   (a) In the absence of preliminary data, how large a sample must be taken so as to be sure that a 98% confidence interval will specify the proportion to within ± 0.05?
   (b) In a sample of 200 failures, 30 of them were caused by stress corrosion cracking. Find a 98% confidence interval for the (population) proportion of failures caused by stress corrosion cracking.

2. During class, 52 students were asked for their preference between Mitt Romney and Barack Obama in the US presidential election this fall. Of these 52 students, 22 of them (42.3%) indicated they would likely vote for Obama. This led to a 95% confidence interval for the proportion of all Vanderbilt students favoring Obama equal to $(.299, .557)$. If the sample proportion had been 90% instead of 42.3%, would the resulting confidence interval been narrower or wider than the the original one? Or would it have had the same width? *Justify your answer.*

3. A recent New York Times / CBS News poll[1] found that 47 percent of Americans disapprove of the 2010 Affordable Care Act (also known as Obamacare). According to the New York Times, "The nationwide poll is based on telephone interviews conducted March 21-25 on landlines and cellphones with 986 adults and has a margin of sampling error of plus or minus three percentage points. " What confidence level did the pollsters use? *Justify your answer.*

4. A rookie is brought to a baseball team on the assumption that he will have a .300 batting average, that is, it is expected that the rookie will get a hit with probability .300 on each of his "at bats." In his first year, he has 300 "at bats" resulting in 80 hits, giving a proportion of hits to "at bats" of .267. Is this compelling evidence that the rookie's true batting average is less than .300? *Justify your answer.*

5. Crash testing is a procedure used to evaluate the ability of an automobile to withstand a serious accident. Suppose a random sample of 24 small cars were subjected to a head-on collison at 40 miles per hour. Of these, 16 were totaled, meaning that the cost of repairs is greater than the value of the car. Another sample of 30 large cars were subjected to the same test, and 10 of them were totaled. Is this sufficient evidence to conclude that there is a difference in the proportions of the two types of cars that are likely to be totaled in such collisons? *Justify your answer.*

6. Two suppliers manufacture a plastic gear used in a laser printer. The impact strength of these gears measured in foot-pounds is an important factor in the life of these parts. A random sample of 50 gears from supplier 1 results in $\bar{x}_1 = 290$ and $s_1 = 6$, while another random sample of 50 gears from supplier 2 results in $\bar{x}_2 = 321$ and $s_2 = 11$.
   (a) Is there evidence to support the claim that supplier 2 provides gears with higher mean impact strength? *Justify your answer.*
   (b) Do the data support the claim that the mean impact stength of gears from supplier 2 is at least 25 foot-pounds higher than that of supplier 1? *Justify your answer.*
   (c) Construct 95% and 99% confidence intervals for the difference in mean impact strength.

[1] Source: `http://thecaucus.blogs.nytimes.com/2012/03/26/poll-47-disapprove-of-health-care-law/`

7. In the last problem set, you looked for a linear relationship between critic and audience scores for 2011 films on the website Rotten Tomatoes. Here's that data set again:

   `http://derekbruff.org/blogs/math216/files/2012/04/Hollywood.csv`

   If we view the scores for these 2011 films as a random sample of all possible film scores on Rotten Tomatoes, do these data support the claim that the average audience score is greater than the average critic score for a film? *Justify your answer.* (If you use R for this problem, you should know that the R functions `mean` and `sd` give you, respectively, the mean and standard deviation of a set of data.)

8. Three different pesticides can be used to control infestation of grapes. It is suspected that pesticide 3 is more effective than the other two. In a particular vineyard, three different plantings of pinot noir grapes are selected for study, each having received one of the three different pesticides.

   Let X_i = the number of bushels of grapes collected from a randomly selected plant treated with pesticide i, where $i = 1$, 2, and 3. Let $\bar{X}_i$ = the average number of bushels of grapes collected per plant from a random sample n_i plants treated with pesticide i, again for i =1, 2, and 3. Consider the random variable $Y = \bar{X}_3 - \frac{1}{2}(\bar{X}_1 + \bar{X}_2)$.

   (a) If μ_Y happens to be positive, what can we conclude about the effectiveness of pesticide 3 relative to the other two pesticides?

   (b) Determine a formula for μ_Y in terms of μ_{X_1}, μ_{X_2}, and μ_{X_3}.

   (c) Determine a formula for σ_Y in terms of σ_{X_1}, σ_{X_2}, σ_{X_3}, n_1, n_2, and n_3. (See page 196 in your textbook if you need a hint.)

   (d) Suppose the study described above yields the following data:

Pesticide	$\bar{x}_i$	s_i	n_i
1	4.6	0.7	100
2	5.2	0.6	120
3	6.1	0.8	130

   Do these data support the claim that pesticide 3 is more effective than the other two? Use an appropriate hypothesis test to answer this question, drawing on your answers to parts (b) and (c).