STAT 301 Exam 1 October 14, 2009

1. (9 pts) An article that appeared on `www.usatoday.com` on October 7, 2009 discussed results of a national poll conducted by the Associated Press that asked American adults about their attitudes toward the swine flu vaccine. One finding reported was that 38% of parents surveyed said that they were unlikely to allow their children to be given the swine flu vaccine at school.

The article reported the sample size for the survey to be 1003, but that includes parents and non-parents, and the 38% result given above was only based on parents who were surveyed. The article did report that the margin-of-error for this result is ±5.2 percentage points.

a) (6 pts) Use this information to determine how many parents were sampled in this survey. Assume that the reported margin-of-error is for a 95% confidence interval.)

The margin-of-error is $z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{.38(1-.38)}{n}}$, which we know to equal .052. Solving for n gives: $n = (1.96)^2 \times .38 \times .62 / (.052)^2 \approx 334.7$, so 335 parents must have been in the survey.

b) (3 pts) Would you expect the margin-of-error for questions that were asked of the entire sample be smaller, larger, or the same as ±5.2 percentage points? Explain your answer, without calculating this margin-of-error.

The sample size would be much larger for the entire sample, so the margin-of-error would be smaller. A larger sample produces a smaller margin-of-error because it entails less sampling variability.

2. (36 pts) Do voters make judgments about political candidates based on his/her facial appearance? Can you correctly predict the outcome of an election, more often than not, simply by choosing the candidate whose face is judged to be more competent-looking? Researchers investigated this question in a study published in *Science* (Todorov, Mandisodka, Goren, and Hall, 2005). Participants were shown pictures of two candidates and asked who has the more competent-looking face. Researchers then predicted the winner to be the candidate whose face was judged to look more competent by most of the participants. For the 32 U.S. Senate races in 2004, this method predicted the winner correctly in 23 of them.

a) (1 pt) In what proportion of the 32 U.S. Senate races did the "competent face" method predict the winner correctly?

$23/32 \approx .719$

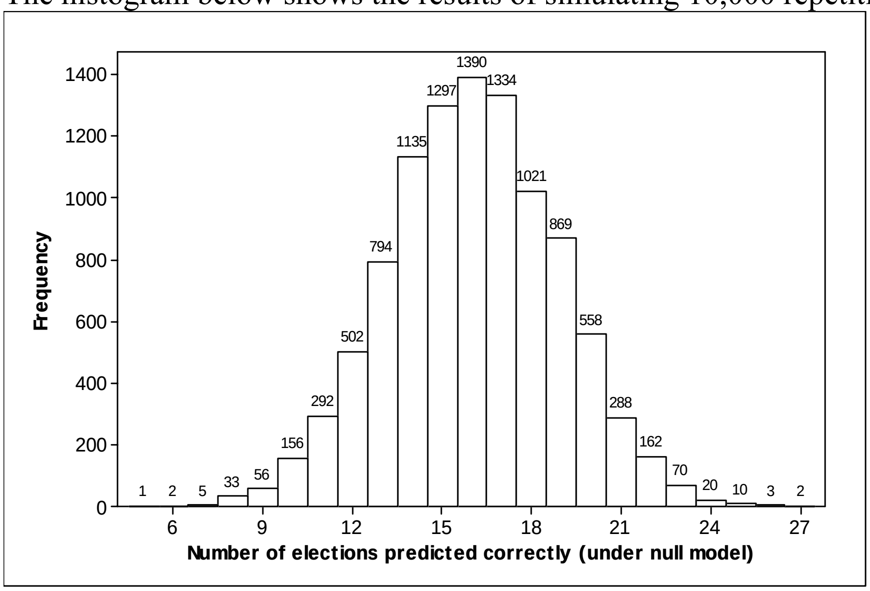
b) (3 pts) Describe (in words) the null model to be investigated with this study.

The null model is that the "competent face" method would successfully predict the winner in 50% of all election races, meaning that this method has no predictive validity.

c) (8 pts) Describe how you could use a coin-flip process to conduct a simulation analysis of this study, in order to assess whether the "competent face" method of predicting election results really does succeed more often than not. Provide enough detail that someone else could implement this simulation and draw the appropriate conclusion.

Flip a coin 32 times, once for each election race, with Heads indicating that the "competent face" method correctly predicts the winner. Keep track of the number of Heads in these 32 flips, corresponding to the number of correct predictions. Then repeat this process of 32 flips a large number (say, 1000) of times. Count how many of these 1000 repetitions produce 23 or more Heads, and divide by 1000 to obtain the approximate p-value. If this approximate p-value is small (say, less than .05), that indicates that the observed result is unlikely to happen by chance alone if there were no predictive validity to the "competent face" method, which would provide fairly strong evidence that the method does make the correct prediction more than 50% of the time. If this approximate p-value is not so small, that would indicate that the observed result would not be surprising even if the method has no predictive validity, so this would provide little or no evidence that the "competent face" method does any better than flipping a coin.

The histogram below shows the results of simulating 10,000 repetitions of this process:



d) (3 pts) Use these simulation results to approximate the p-value of the test. Indicate how you calculate this value, and also report the value.

The approximate p-value is the proportion of these 1000 repetitions that produce 23 or more correct predictions: (70+20+10+3+2)/10,000 = 105/10,000 = .0105.

e) (6 pts) Fill in the three blanks below to describe how you would calculate the exact p-value based on the binomial distribution:

The p-value is Pr(X $\geq$ 23),

where X has a binomial distribution with n = 32 and π = .5.

The exact p-value (to three decimal places) of this test can be found to be .010.

f) (6 pts) Complete the following sentence to interpret this p-value: The probability is .010 that …

the "competent face" method would make correct predictions in 23 or more of 32 election races, if in fact it has no predictive validity and so its actual probability of a correct prediction is .5.

g) (2 pts) What test decision would you make at the $\alpha = .05$ significance level? Circle your answer. Do not bother to explain.

Reject H_0 Fail to reject H_0

because the p-value is less than .05.

h) (4 pts) Write a paragraph, as if to the researchers, describing what your simulation analysis reveals about whether the data provide strong evidence in support of their conjecture.

.

The simulation reveals that if the "competent face" method really has only a 50% chance of making a correct prediction, then it would be quite unlikely to obtain 23 or more correct predictions in 32 election races. Therefore, the simulation analysis reveals that the researchers' result (23 correct predictions in 32 election races) provides very strong evidence that the "competent face" method really does predict the correct winner more than half the time.

These researchers also predicted the outcomes of 279 races for the U.S. House of Representatives in 2004. The "competent face" method correctly predicted the winner in 189 of those races. Notice that this is a success rate of $189/279 \approx .677$, which is slightly smaller than the success rate with the 32 U.S. Senate races.

i) (3 pts) Nevertheless, the p-value turns out to be orders of magnitude smaller with the House study than with the Senate study. Briefly explain why this makes intuitive sense.

The House study involved a much larger sample size than the House study (279 as opposed to 32). With a much larger sample size, if the probability of a correct prediction were really .5, the sample proportion of correct predictions would vary less around the value .5. So, a sample proportion of correct predictions as large as .677 would be extremely unlikely to happen by chance with a larger sample size.

3. (12 pts) A Gallup Poll conducted on October 1-4, 2009 asked a random sample of 1013 adult Americans: Are you in favor of the death penalty for a person convicted of murder? Suppose that you want to use the result to test whether the proportion of all adult Americans support the death penalty differs from two-thirds.

It turned out that 658 of those interviewed answered "yes."

a) (5 pts) Determine the value of the relevant *z*-score for conducting this test.

The z-score is: $\dfrac{\hat{p}-\pi_0}{\sqrt{\dfrac{\pi_0(1-\pi_0)}{n}}} = \dfrac{\dfrac{658}{1013}-\dfrac{2}{3}}{\sqrt{\dfrac{\frac{2}{3}\left(1-\frac{2}{3}\right)}{1013}}} \approx -1.155.$

b) (2 pts) Based on this *z*-score, is the p-value of the test less than or greater than .05? Explain how you know.

We know that the z-score for a (two-sided) p-value of .05 is ±1.96, so this *z*-score is not nearly as large (in absolute value), so the p-value is greater than .05.

c) (5 pts) Based on the *z*-score and your answer to b), summarize the conclusion that you would draw from this study. Also explain the reasoning process by which this conclusion follows from your previous answers.

The z-score of -1.155 and p-value > .05 indicate that the observed sample result (658/1013 ≈ .650 favoring the death penalty) would not be surprising to occur by chance alone if the population proportion who favor the death penalty is actually two-thirds. So, the survey data provide no reason to reject the hypothesis that two-thirds of all adult Americans support the death penalty.

4. (14 pts) In the mid-1980s, sociologist Shere Hite undertook a study of American women's attitudes toward relationships, love, and sex by distributing 100,000 questionnaires in women's groups. One of the questions was: Do you give more emotional support to your husband or boyfriend than you receive from him? A total of 4500 women returned the questionnaire.

An ABC News/*Washington Post* poll conducted at about the same time surveyed a random sample of 767 women, asking them the same question about emotional support.

a) (5 pts) Which survey would you expect to obtain a more representative sample of the population? Explain briefly.

The ABC News/ *Washington Post* survey was based on a random sample, so it is likely to be representative of the population of American women. The Hite survey was only distributed to women who belong to women's groups, who may not be representative of women in general. Also, the response rate was quite low, suggesting that women with a particularly strong opinion were more likely to respond. So, the Hite survey is likely to be biased in favor of women who were less pleased with their husbands or boyfriends.

Of the 4500 women who returned the Hite questionnaire, 96% said that they gave more emotional support than they received from their husbands or boyfriends. Of the 767 women interviewed in the ABC News/*Washington Post* poll, 44% claimed to give more emotional support than they receive.

b) (6 pts) Using *only* the poll corresponding to your answer to b), determine a 95% confidence interval for the relevant population parameter. (Do this by hand.)

A 95% CI is: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, which is $.44 \pm 1.96\sqrt{\frac{.44(1-.44)}{767}}$, which is .44 ± .035.

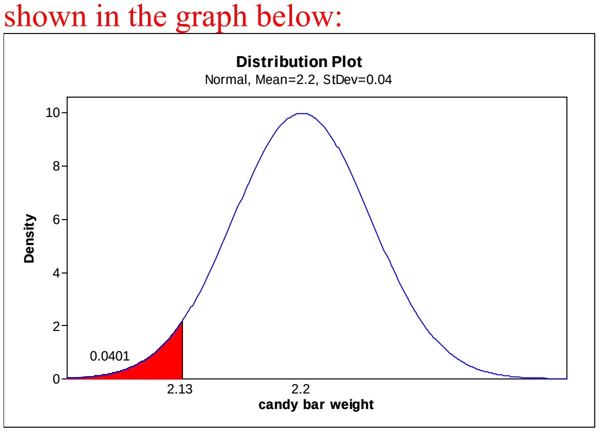
c) (3 pts) Write a sentence interpreting what your confidence interval reveals.

We can be 95% confident that between .405 (40.5%) and .475 (47.5%) of American women feel that they provide more emotional support to their husband/boyfriend than they receive.

5. (12 pts) Suppose that the wrapper of a certain candy bar lists its weight as 2.13 ounces. Naturally, the weights of individual bars vary somewhat. Suppose that the weights of these candy bars vary according to a normal distribution with mean 2.20 ounces and standard deviation 0.04 ounces.

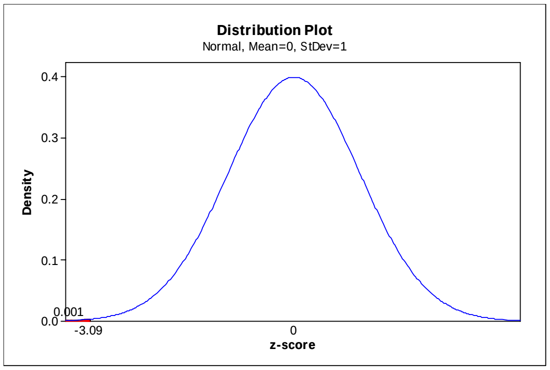
a) (4 pts) Determine what proportion of candy bars weigh less than the advertised weight. Draw a well-labeled sketch to illustrate your calculation.

The probability that a randomly selected candy bar weighs less than 2.13 ounces is .0401, as shown in the graph below:



b) (4 pts) If the manufacturer wants to adjust the production process so that only 1 candy bar in 1000 weighs less than the advertised weight, what should the mean of the actual weights be (assuming that the standard deviation of the weights remains 0.04 ounces)? Explain the process of your solution as well as providing the numerical answer.

Looking at the .001 left tail of a standard ($\mu = 0$, $\sigma = 1$) normal distribution, we see that the necessary *z*-score is -3.09, as shown below:



To find the mean weight that will produce a z-score of -3.09, we solve: $-3.09 = (2.13 - \mu)/ .04$. Solving gives a mean of $\mu = 2.13 + 3.09(.04) \approx 2.2536$ ounces.

c) (4 pts) If the manufacturer wants to adjust the production process so that the mean is reduced to 2.15 ounces but only 1 candy bar in 1000 weighs less than the advertised weight, how small does the standard deviation of the weights need to be? Explain the process of your solution as well as providing the numerical answer.

As in the previous question, we need a z-score of -3.09, so now we solve $-3.09 = (2.13 - 2.15) / \sigma$ to obtain: $\sigma = -.02 / -3.09 \approx .0065$ ounces.

6. (17 pts) Suppose that the Human Resources (HR) Director of a company hires you as a statistical consultant to investigate whether employees are abusing the company's sick leave policy by missing work on Mondays and Fridays in order to take a long weekend rather than because of being sick. She offers to provide you with data for a sample of employee sick days, so you can determine the proportion of sick days taken on Mondays and Fridays.

Let the symbol π represent the actual probability that a randomly selected sick day is taken on a Monday or Friday.

a) (2 pts) If employees are equally likely to be sick on any weekday (Mon, Tues, Wed, Thur, Fri), and if no abuse is taking place, what would be the value of π? In other words what proportion of sick days would be taken on a Monday or Friday?

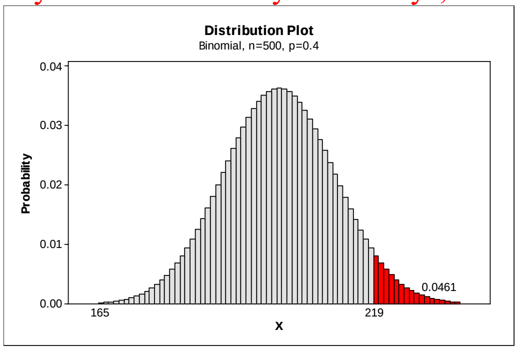
2/5 = .4

b) (2 pts) State (in terms of the symbol π) the appropriate null and alternative hypotheses to be tested.

H_0: $\pi = .4$
H_a: $\pi > .4$

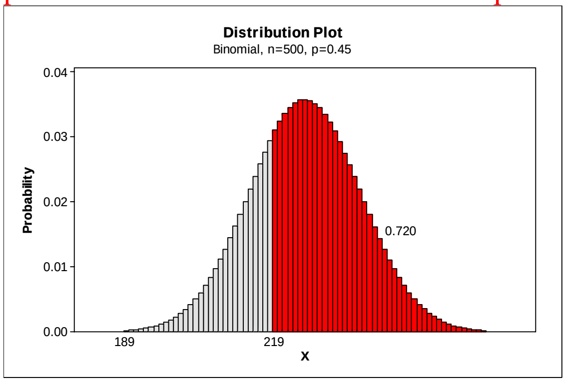
c) (5 pts) Suppose that the HR Director plans to provide you with a sample of 500 sick days. Determine the rejection region, using the $\alpha = .05$ significance level. In other words, determine how many of those 500 sick days would have to be on Monday or Friday in order for you to conclude that employees are taking sick days significantly more often than would be expected by chance? (You may use any appropriate probability distribution. Describe your solution process and report your answer.)

Using a binomial distribution with $n = 500$ and $\pi = .4$, looking for the value that gives a probability of .05 or less in the right tail, we find that the rejection region is for 219 or more sick days to be on Mondays or Fridays, as shown in the graph below:



d) (5 pts) Now suppose that employees really are abusing the policy slightly, and that the probability is actually .45 that a sick day is taken on a Monday or Friday. Determine the power of your test, still using a sample size of 500 and the $\alpha = .05$ significance level. (You may use any appropriate probability distribution. Describe your solution process and report your answer.)

The power is the probability that a binomial distribution with with $n = 500$ and $\pi = .45$ would produce a value of 219 or more. This power turns out to be .720, as shown in the graph below:



e) (3 pts) Explain, as if to the HR Director, what this "power" means.

This power calculation means that if employees really are abusing the policy a bit, to the extent that 45% of all sick days are taken on Mondays or Fridays, then (based on a sample of 500 sick days), we have a .720 probability of correctly rejecting the null hypothesis that only 40% of all sick days are taken on Mondays or Fridays.