Name: _____

Practice Second Midterm Exam
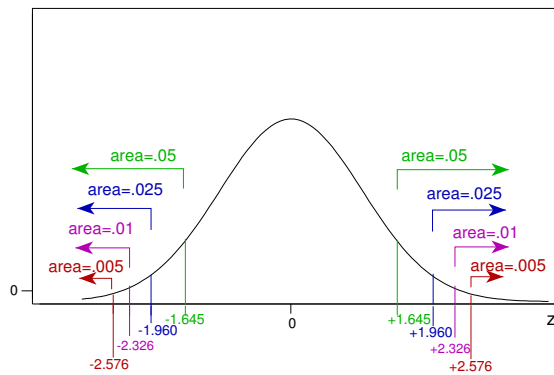
Statistics 200
Dr. Nancy Pfenning

This is a closed book exam worth 150 points. You are allowed to bring a calculator, and 2 two-sided sheets of notes. There are 8 problems, with point values as shown. If you want to receive partial credit for wrong answers, show your work. Don't spend too much time on any one problem.

1. (30 pts.) The proportion of high school seniors who are married is 0.02. Suppose we take a random sample of 1600 high school seniors and look at sample proportion who are married.

   (a) To make sure a standard deviation based on a binomial model is accurate, we check if
      i. the sample is large enough to expect at least 10 each who are and are not married
      ii. the population of interest is more than 16,000 students
      iii. the students were selected without replacement

   (b) To make sure a normal approximation can be used, we check if
      i. the sample is large enough to expect at least 10 each who are and are not married
      ii. the population of interest is more than 16,000 students
      iii. the students were selected without replacement

   (c) Find the mean _____and standard deviation _____of the sample **count** X who are married.

   (d) Find the mean _____and standard deviation _____of the sample **proportion** $\hat{p}$ who are married.

   (e) What is the z score if the sample proportion married is 0.026?

   (f) If the proportion of sampled students who are married is 0.026, what conclusion can we draw?
      i. We are willing to believe that the sample was coming from a population where the proportion married is 0.02.
      ii. We are not willing to believe that the sample was coming from a population where the proportion married is 0.02.
      iii. Results are borderline.

2. (15 pts.) Women's shoulder heights (in inches) are normal with mean 52.5 and standard deviation 2.5.

   (a) According to the 68-95-99.7 Rule, the probability is 0.0015 that a woman's shoulder height would be less than how many inches? _____

   (b) Find the standardized score z for a woman whose shoulder height is 59 inches. _____

   (c) Use the sketch of the tails of the z curve to estimate the probability of a woman's shoulder height being greater than 59 inches: between
   (i) 0 and 0.005 (ii) 0.005 and 0.01 (iii) 0.01 and 0.025 (iv) 0.025 and 0.05 (v) 0.05 and 1

3. (15 pts.) Number of cows X owned by households in a Nepal farming village has this probability distribution:

Cows	0	1	2	3	4
Probability	.78	.11	.06	.03	.02

   (a) Is the distribution (i) left-skewed, (ii) right-skewed, or (iii) approximately normal?

   (b) The distribution of X has mean 0.4 and standard deviation 0.8. What are the mean _____ and standard deviation _____ for sample mean number of cows $\bar{X}$ for a random sample of 64 households?

   (c) Shape of the distribution of sample mean number of cows for samples of 64 households is

      i. exactly normal
      ii. approximately normal because the sample size is fairly large
      iii. skewed just like the distribution of X

4. (10 pts.) Drug manufacturers test the null hypothesis that their anti-depressant does not make teens more suicidal.

   (a) The consequences of a Type II Error (failing to reject the null hypothesis when it is actually false) would be

      i. taking a potentially beneficial drug off the market unnecessarily
      ii. teenagers taking a drug that may make them suicidal
      iii. teenagers having access to a drug that helps their depression and is not harmful

   (b) Suppose the p-value turns out to be small (less than the prescribed cut-off probability α). What type of error could be committed?

      i. Type I (reject null hypothesis even though it is true)
      ii. Type II (fail to reject null hypothesis even though it is false)
      iii. both I and II
      iv. neither I nor II

5. (10 pts.) Assume that 0.50 of all births are boys. Suppose we look at a sample of 20 births in a city each day for 200 days.

   (a) Suppose we set up a 99% confidence interval each day for the proportion of all births that are boys, based on the sample proportion of boys born that day. About how many of the 200 confidence intervals should contain the true proportion 0.50? (i) 0 (ii) 1 (iii) 2 (iv) 10 (v) 20 (vi) 50 (vii) 99 (viii) 100 (ix) 190 (x) 198 (xi) 200

   (b) Suppose that each day we test the true null hypothesis that the population proportion is 0.50 against the two-sided alternative, based on the sample proportion of boys born that day. About how many of these 200 tests should reject at the $\alpha = .01$ level? (i) 0 (ii) 1 (iii) 2 (iv) 10 (v) 20 (vi) 50 (vii) 99 (viii) 100 (ix) 190 (x) 198 (xi) 200

6. (20 pts.) In a representative sample of 400 Pitt students in 2007, the proportion carrying cell phones was 0.95.

(a) Give a 95% confidence interval for the proportion of all Pitt students who carry a cell phone, rounding to the nearest hundredth (two decimal places).

(b) If a much larger sample were taken, then the interval would be
(i) wider (ii) narrower (iii) not enough information

(c) If a higher level of confidence were desired, then the resulting interval would be
(i) wider (ii) narrower (iii) not enough information

(d) Based on your interval in (a), can we assert that more than 90% of all Pitt students carry a cell phone?
(i) yes (ii) borderline (iii) no, not at all

(e) A test is carried out to see if the proportion of all Pitt students carrying cell phones could be 0.90 against the alternative that the proportion is greater than 0.90, and the p-value is found to be 0.0004. This means that (choose one):

i. If the proportion of all Pitt students with cell phones is **equal to** 0.90, then .0004 is the probability that a random sample of 400 Pitt students would have a proportion with cell phones **equal to** 0.95.

ii. If the proportion of all Pitt students with cell phones is **equal to** 0.90, then 0.0004 is the probability that a random sample of 400 Pitt students would have a proportion with cell phones **at least as large as** 0.95.

iii. If the proportion of all Pitt students with cell phones is **greater than** 0.90, then 0.0004 is the probability that a random sample of 400 Pitt students would have a proportion with cell phones **equal to** 0.95.

iv. If the proportion of all Pitt students with cell phones is **greater than** 0.90, then 0.0004 is the probability that a random sample of 400 Pitt students would have a proportion with cell phones **at least as large as** 0.95.

7. (30 pts.) Age (in years) at first conception was recorded for 7 captive male Beluga whales who sired offspring at various aquariums in the U.S.

   (a) The multiplier for a 95% confidence interval in this case is
       (i) less than 2 (ii) more than 2 (iii) equal to 2

   (b) A 95% confidence interval is found to be $(10.5, 15.5)$. Which one of the following is the only correct interpretation of the interval?

       i. The sample mean has a 95% probability of falling in this interval.
       ii. The population mean has a 95% probability of falling in this interval.
       iii. 95% of all captive male Belugas are between 10.5 and 15.5 years old when first offspring is sired.
       iv. 95% of sampled whales were between 10.5 and 15.5 years old when first offspring is sired.
       v. We are 95% sure that sample mean falls in this interval.
       vi. We are 95% sure that population mean falls in this interval.

   (c) State the null and alternative hypotheses to test if the mean age of all captive male Belugas, when first offspring was sired, could be 13.7 years.

   (d) Keeping the confidence interval in mind, tell which one of these is the actual p-value for the test in part (c): (i) 0.52 (ii) 0.052 (iii) 0.0052

   (e) Choose the best option from the following conclusions:

       i. The P-value is not small; we have evidence that the population mean differs from 13.7.
       ii. The P-value is not small; we have proven that the population mean equals 13.7.
       iii. The P-value is not small; the population mean may equal 13.7.
       iv. The P-value is small; we have evidence that the population mean differs from 13.7.
       v. The P-value is small; we have proven that the population mean equals 13.7.
       vi. The P-value is small; the population mean may equal 13.7.
       vii. The P-value is borderline; results are inconclusive.

   (f) If the alternative hypothesis had been one-sided, the p-value would be
       (i) the same (ii) twice as much (iii) half as much

8. (20 pts.) In order to determine if overall prices were lower at either of two supermarkets, consumer reseachers checked for prices of a sample of six items (frozen peas, dozen eggs, wheat bread, cocoa cereal, decaf coffee, sandwich bags) at both stores and carried out a test in MINITAB.

```
                   N       Mean      StDev    SE Mean
Walmart            6       1.365     0.519     0.212
GiantEagle         6       1.977     1.445     0.590
Difference         6      -0.612     1.275     0.521
95% CI for mean difference: (-1.950, 0.727)
T-Test of mean difference = 0 (vs not = 0): T-Value = -1.17   P-Value = 0.293
```

(a) Was this a paired study? (i) yes (ii) no (iii) there is no way to tell

(b) Does the output above convince us that the prices differ overall for the two stores? Use $\alpha = 0.05$ as your cut-off, and circle the best answer below.

   i. Yes, definitely.
   ii. Yes, but just barely.
   iii. No, not quite.
   iv. No, not at all.

(c) The shape of the histogram below suggests that

   i. our conclusions should be correct because the data are roughly normal
   ii. our conclusions may be incorrect because the data show extreme skewness
   iii. our conclusions may be incorrect because the data are symmetric but bimodal

(d) The position of the confidence interval suggests

   i. the difference is significant because the confidence interval is centered way to the right of zero
   ii. the difference is significant because the confidence interval is centered way to the left of zero
   iii. the difference is not significant because zero is well-contained in the confidence interval
   iv. the difference is borderline because the confidence interval just barely includes zero

(e) Which of these would tend to produce a smaller p-value? (Circle any that apply.) (i) if more items were sampled (ii) if mean of sampled differences were closer to zero (iii) if standard deviation of sampled differences were larger

Histogram of Differences