LAST NAME (Please Print): _____

FIRST NAME (Please Print): _____

HONOR PLEDGE (Please Sign): _____

Statistics 111

Midterm 3

- This is a closed book exam.
- You may use your calculator and a single page of notes.
- The room is crowded. Please be careful to look only at your own exam. Try to sit one seat apart; the proctors may ask you to randomize your seating a bit.
- Report all numerical answers to at least two correct decimal places or (when appropriate) write them as a fraction.
- All question parts count for 1 point.

1. Assume IQs have mean 100 and standard deviation 16. You have created a pill that increases IQ by 10 points. You want to prove that your pill works.

   _____ Suppose $\alpha = 0.05$ and $\beta = 0.1$. What sample size do you need?

   _____ Suppose you test 25 people with Type I error rate of 0.01. What is the power of your test?

2. _____ You sample 100 random people in North Carolina and ask them whether they support gay marriage. You find that 40 say yes. Set a 95% one-sided lower confidence bound on the proportion of people in NC who favor gay marriage.

3. Someone asserts that at least 10% more women than men support gay marriage. You want to test this at the 0.05 level. You sample 100 women and find that 60 are supporters, and among 80 men, 30 are supporters.

   Write the alternative hypothesis in symbols. (Subtract men from women.)

   _____

   _____ What is the value of your test statistic?

   _____ What is your P-value?

   _____ What is your critical value?

   _____ Do you reject the null hypothesis?

4. There are 16 statistic professors at Duke. A sample of 10 has an average IQ of 150 with a sample sd of 10. You want to test whether the average IQ of the statistics faculty is greater than 145, with Type I error equal to 0.1. Find your test statistic and critical value.

   ts = _____ cv = _____

5. Someone claims that 40% of Duke students want a career in business, 20% want to become physicians, 20% want to become lawyers, 10% want to be celebrities, and the rest want to become statisticians. A sample of 200 students finds that 70 want a career in business, 40 want to be doctors, 35 want to be lawyers, 15 want to be celebrities, and the rest aspire to become arbiters of truth. You want to test the claim at the 0.05 level.

   In words, write your null hypothesis.

   _____

   _____

   _____ What is the value of your test statistic?

   _____ What is your critical value?

   Give bounds for your significance probability. _____

   _____ Do you reject the null hypothesis?

6. _____ There are three species of Scarlet Snipe. Type A has prevalence 50% and lays a Poisson number of eggs with mean 0.5. Type B has prevalence 25% and lays a Poisson number of eggs with mean 1.5. Type C has prevalnece 25% and its mean is 2.5. You find a Scarlet Snipe nest with 2 eggs. What is your probability for the hypothesis that it is type A?

7. What is multicollinearity, and what does it do? (2 pts.)

   _____

   _____

8. President Broadfoot thinks that freshman, sophomores, juniors and seniors have the same propensity to drink. You decide to test this at the 0.05 level, and gather a sample of 50 freshmen, of whom 40 got drunk the previous night, 80 sophomores, of whom 60 drank the previous night, 50 juniors, of whom 30 drank the previous night, and 40 seniors, of whom 10 drank the previous night.

   In words, what is your null hypothesis?

   _____

   _____

   _____ What is the expected number of freshmen who drink?

   _____ What is the value of your test statistic?

   _____ What is your critical value?

   Give bounds for your P-value. _____

   _____ Do you reject the null hypothesis?

9. You want to predict how long someone lives from the number of hours of exercise they do in a week. From a sample of 15 people, you estimate the intercept as 90 and the slope as -2. The coefficient of determination is 0.6, the sd of the residuals is 9, and the mean number of hours spent exercising is 8 with an sd of 4.

   _____ What is the correlation?

   _____ Set a 95% upper confidence interval on the lifespan of someone who expercises for 4 hours per week.

   _____ Set a 95% upper confidence interval on the average lifespan of people who exercise for 4 hours per week.

10. You want to predict lifespan from the number of hours of exercise (NE), the age of your grandparents (AG), and your income level (IL) in thousands of dollars. From a sample of 15 people, your estimated model is

$$Y = 70 - 2 * NE + 0.1 * AG + 0.05 * IL.$$

The standard errors for the coefficients on NE, AG, and IL are, respectively, 3, 0.2, and 0.01, respectively.

At the 0.05 level, you want to decide whether NE should be in the model. What are your test statistic and critical value(s)?

ts = _____ cv = _____

_____ What is your predicted lifespan for Adelbert, who exercises for two hours per week, has grandparents whose average age is 80, and who earns $100K per year?

_____ Adelbert lives to be 100. What is his residual?

11. List all, and only, the true statements. _____

A. Log transformations help when the data follow a power law.
B. As points cluster more tightly around the line, the coefficient of determination increases.
C. The probabilities of Type I and Type II error add to 1.
D. Heteroscedasticity occurs when the scatterplot is shaped like a cigar.
E. Linear regression assumes that the dependent variable is measured without error.
F. Sir Francis Galton invented eugenics.
G. It is okay to make many tests at the 0.05 level and report those that are significant.
H. With large sample sizes, one rarely finds statistically signficant differences that are not important.
I. In cross-validation, one uses many hold-out samples to estimate predictive accuracy.
J. As sample size increases, so does the power of the test.
K. The correlation between the heights of fathers and sons is about 0.7.
L. In high dimensions, all data sets are sparse.
M. Extrapolation is useful when doing regression inference.