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. The Department Chair says that she wants the final to have about 20% A's, 30% B's, 30% C's, 10% D's and 10% F's. In his class of 80, Balderdash gives 22 A's, 28 B's, 25 C's, 5 D's and 0 F's. Did he follow the guidelines?

   What is the Department Chair's null hypothesis? (In words.)

   _____

   _____

   _____ What is the value of the Chair's test statistic?

   _____ What table does the Chair use? (Include df if appropriate.)

   _____ What P-value does the Chair get? (Give a range if needed.)

   What conclusion does she reach? (In words. Use $\alpha = 0.05$)

   _____

2. The chair wants to decide whether two instructors have the same grading curves. She observes:

	A	B	C	D	F
Balderdash	20	30	30	10	10
Crundel	50	80	50	10	10

   What is the chair's alternative hypothesis? (In words.)

   _____

   _____

_____ What is the value of her test statistic?

_____ What table does she use? (Include df if appropriate.)

_____ What P-value does she get? (Give a range if needed.)

What conclusion does she reach? (In words; use $\alpha = 0.05$)

_____

_____

3. You are doing a linear regression to predict average body temperature from the average daily amount of exercise one does. For a sample of 20 random people, you record their hours per day of exercise for a week, and their body temperatures at noon. You find that the average number of hours of exercise is 0.5 and the sample variance is 2. Your estimate of the intercept as 99.2 and the slope as -.08. And 30% of the variation in Y is explained by X. The standard deviation of the residuals is 0.6 and the standard error of $\hat{\beta}_1$ is 0.04.

_____ Estimate the temperature of someone who exercises for 2 hours per day.

_____ What is upper bound of a two-sided 95% confidence interval on the average temperature of people who exercise 2 hours/day?

_____ What is the upper bound of a two-sided 95% confidence interval on the temperature of Ann, who exercises for 2 hours/day?

_____ What is the value of the correlation coefficient?

_____ What is your significance probability for testing whether exercise is related to body temperature? Give a range if necessary.

4. You want to place a confidence interval (CI) on the correlation between grades and income. For 100 people, you find that the sample correlation is 0.4 You resample 20 times from that sample (with replacement) and find the following correlations:

0.2, 0.3, 0.4, 0.5 0.25, 0.35, 0.45, 0.55, 0.23, 0.33, 0.43, 0.53, 0.21, 0.32, 0.42, 0.52, 0.24, 0.34, 0.44, 0.54

———————— Set a 95% one-sided lower CI with the pivot bootstrap.

———————— What is the probability that Abelard's GPA and income were not included in the first bootstrap sample?

5. Freshmen at UNC work 12.2 hours a week for pay, on average, and the SD is 10.5 hours; at Duke, the average is 10.2 hours and the SD is 9.9 hours. You want to show that Duke students work less. Assume that these data are based on two independent simple random samples, each of size 100. (Subtract Duke from UNC.)

What is the alternative hypothesis (in symbols)?

————————————————————————————————————

———————— What is the value of your test statistic?

———————— What is the P-value of the test statistic? (Give a range if needed.)

What is your conclusion? Consider P-values less than 0.05 as small.

————————————————————————————————————

6. A carnival barker guesses people's weight (WT) from estimates of their height (H) and width (W). Fill in the equation:

$\ln WT = \beta_0 + \beta_1$ ———————— $+\beta_2$ ————————

7. Suppose you want to test whether a new diet pill leads to a five pound weight loss within one month. In fact, the average weight loss is six pounds, and the standard deviation in weight loss is eight pounds.

——————————— For a power of 0.9 with 0.05 Type I error rate, what sample size do you need?

——————————— For $\alpha = 0.05$ and sample size of 100, what is your Type II error?

8. The lifespan of lightbulbs, in months, is exponential with parameter λ. Brand A has $\lambda = 1$, Brand B has $\lambda = 1/2$ and Brand C has $\lambda = 1/3$. Brand A has 30% of the market, Brand B has 20%, and Brand C has the rest.

——————————— Your lightbulb burns out at 1 month. What is your belief about the probability that it is not Brand A?

——————————— How long do you think another bulb from the same package will last?

9. List all, and only, the true statments. ———————————————

    A. As points cluster more tightly around a line, the correlation increases.

    B. The percentile bootstrap is better than the pivot.

    C. Galton invented the idea of eugenics.

    D. In high dimensions, data tends to be too sparse.

    E. Making predictions outside the range of the explanatory variables in your training data is reasonable.

    F. If you make many tests, your true α decreases.

    G. With large samples, you can get results that are statistically significant but not scientifically significant.

    H. You should not plot your data before deciding upon your alternative hypothesis.

    I. The P-value is the probability of observing data that are as more more supportive of the null than the data observed, when the null is true.

    J. All models are wrong, but some models are useful.

10. What is multicollinearity? Here and below, be brief and clear.

11. Why is multicollinearity a problem?

12. How do you calculate a moving average estimate for Y at a point x in nonparametric regression?

13. How does two-fold cross-validation work in nonparametric regression?