Law 7930-23
Statistics for Lawyers
Jonathan Klick

Black Tuesday

April 25, 2006
8:30-11:30
Room 241

This final exam consists of 10 questions (plus subparts) on pages 1-4 (note: this page is not included in the numbering). The exam is out of 100 points and the weight of each question is noted.

This is an open book and open note exam. You may use your textbook, class notes, text supplements, commercial outlines, commercial supplements, lucky charms, and any other printed or written materials you believe will be helpful in completing your answers. Surfing the internet is not permitted during the exam period.

You will be given three hours to complete this exam, and you may allocate your time among the questions as you see fit. There is no answer sheet for this exam. You may write as long or as short an answer as you believe is necessary to completely answer the test questions. If you need extra bluebooks, please quietly request them from the proctor. If you finish early, please submit your bluebooks or computer disks to the exam proctor and quietly exit the room.

If you perceive an ambiguity or error in any test question, please proceed to answer it, noting the ambiguity/error and making any reasonable assumptions you believe are necessary to answer the question. Please state these assumptions in your answer and provide your justification for the assumption. If you show your work, you may receive partial credit for an incorrect answer.

Since some students have arranged to take the exam at a later date, recognize that any information you provide them about the exam will adversely affect your own grade because of the law school's grading curve.

Good luck and remember that you're all above average.

1. You want to determine if taking BA appears to improve students' likelihood of passing the bar on the basis of the following information about the most recent graduating class's bar exam performance:

Students taking exam: 200
Students passing exam: 140
Of those passing, 75 percent had BA
30 percent of the students failed AND had BA

What is the probability of passing the bar if the student had BA?   (5 points)

2. If the standard deviations of 3 independent variables are as follows:

A: 15   B: 20   C: 12

What is the standard deviation of the following function:

1/2*A + 1/2*B - 1/2*C   (5 points)

3. In a standard 52 card deck of playing cards (i.e., four suits each of which has an Ace, Jack, Queen, King, and numbers 2-10), what is the probability of:

   a. **Being dealt 4 of a kind (i.e., 4 cards all with the same number or face card) out of the first four cards off the top of the deck?**
   b. **Being dealt a flush (i.e., 5 cards all of the same suit) out of the first five cards off the top of the deck?**
   c. **Being dealt a royal flush (i.e., 10, Jack, Queen, King, Ace all of the same suit) out of the first five cards off the top of the deck?**
   d. **Being dealt a five card hand that does not have a pair (i.e., 2 cards of the same number or face) out of the first five cards off the top of the deck?**
   e. **Pick one of the four subparts above (i.e., 3a, 3.b, 3.c, or 3.d) for me to count double or choose to have each of the subparts count for an equal weight in question 3. Be clear in which you are choosing.**

                                                                (10 points)

4. In most of the studies on the relationship between physical appearance and scores on teaching evaluations, researchers find a positive relationship between measures of physical attractiveness (e.g., measures of facial symmetry, subjective evaluations on ratemyprofessors.com, etc.) and positive scores on teaching evaluations. In general, these studies only control for the sex of the professor and the academic field. Explain how these results may suffer from omitted variables

biases in general terms and provide some well-reasoned hypotheses about what omissions specifically might be generating the bias.

(10 points)

5. A class action case is brought against the major television broadcast networks. The class is composed of parents of obese children and the claim is that television programs aimed at children "caused" the obesity epidemic in this country leading to long term health problems the costs of which will be borne by the parents and their children. As evidence of this, the plaintiffs cite a study with the following findings:

From repeated studies it is known that of children ages 6-11 who watch fewer than 2 hours of television per week, 10 percent are clinically obese (with a standard deviation of 2.5 percent). In the study relied upon by plaintiffs, researchers measured the BMIs of 2,000 children in the 6-11 age range all of whom generally watch more than 2 hours of television per week, finding an obesity incidence of 14.5 percent.

   a. Calculate the 84 percent confidence interval for the "baseline" (i.e., non-TV watching) obesity rate for 6-11 year olds.
   b. Calculate the p value associated with the obesity incidence rate found among the TV watching group.
   c. What statistical argument(s) do you make in this case if you are the defense counsel?
   d. What statistical argument(s) do you make in this case if you are the class counsel?
   e. It turns out that BMIs are not generally normally distributed. Mean BMI tends to be higher than median BMI. Why might that be? Does it cause any problems for the analysis above (explain)? What might a solution be?

(10 points)

6. Which is more important – Type I error or Type II error? Explain.

(10 points)

7. Fill in the missing cells from this regression output (reproduce the whole chart in your bluebook)

Variable	Coefficient	SE	t stat	p value	95% Confidence Interval	
Constant	3.00	0.50				
X1		1.20				2.00
X2		0.08		.001		
X3	5.00			.025		
X4					2.00	12.00
Observations = 10,000						

(15 points)

8. You have the following regression output examining the effect of various variables on the income of 10,000 people in a given year:

Variable	Coefficient	SE
Constant	12,000.00	6,000.00
Years of schooling	5,000.00	3,000.00
Age in years	500.00	375.00
Smoker (= 1 if smoker; = 0 if not a smoker)	-2,000.00	450.00
Female (=1 if woman; 0 otherwise)	-3,000.00	1,500.00
Seniority (years in job)	1,200.00	700.00
$R^2 = 0.125$		

8.a If you believe the true effect of age is non-linear (such that the general experience that comes with age is rewarded in the labor market but, beyond some point, older people get paid less because of discrimination or because of some non-discriminatory reason), how would you model that if you re-ran the regression?

8.b What signs would you expect on your age variable(s) if you implemented your change in 8.a?

8.c Assume that, regarding the variables controlled for above, men and women have exactly the same coefficients (except, of course, for the female, variable). If you re-estimated the regression just on the data for women (and obviously dropped the female variable), what value would your intercept coefficient take?

8.d Arguably "smoker" is a proxy variable in this context. What is a proxy variable and why might you use smoking as one in this case? Given that, what can you say about the causal relationship between smoking and wages?

8.e Does the R^2 suggest this is a "good" model? Explain (note: it will likely be helpful to explain what the R squared represents in the process).

(15 points)

9. One of the enduring questions of social science involves the nature/nurture distinction. That is, how much of an individual's success (defined by various metrics) is attributable to genetics (or at least pre-natal development) as opposed to post-natal environmental conditions (including, for example, family structure, parental investments such as reading to children, early education, etc.).

9.a Why is the nature/nurture question so difficult to address statistically?

9.b Some researchers have attempted to address these problems by examining pairs of twins. For example, one design may look at situations in which one twin is put up for adoption during its infancy, while the other is raised by the biological parents. How can these twin studies help to isolate the relative contributions of nature and nurture in individual success?

9.c Are there any obvious weaknesses inherent in designs like these (i.e., the twin designs discussed in 9.b) that might limit our ability to draw general inferences about the relative contribution of nature and nurture?

(15 points)

10. Discuss the ways in which the practice of regression analysis might differ if your purpose is prediction as opposed to the inference of causal relationships.

(5 points)