LAST NAME (Please Print): ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯

FIRST NAME (Please Print): ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯

HONOR PLEDGE (Please Sign): ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯

Statistics 111

Midterm 4

- 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.

I have really enjoyed this semester, and I hope you have too. It has been a lot of work, but the intent was to make it fun as often as possible, and always worthwhile and rewarding.

David Banks

1. What is the purpose of nonparametric regression?

_____

_____

2. Explain how a running line smooth with a window containing five observations works. (2 points)

_____

_____

_____

3. Suppose you are doing cross-validation to assess predictive accuracy in nonparametric regression. But by mistake, some of the (X, Y) pairs that you use to fit the model have been duplicated in the data set. How will this affect your assessment and why? (2 points)

_____

_____

_____

4. In terms of predictive accuracy, what is the implication of the Curse of Dimensionality? (1 point)

_____

_____

5. Consider the following design and data.

run	A	B	C	D	E	F	obs
1	-	-	-	+	+	+	3
2	+	-	-	-	-	+	7
3	-	+	-	-	+	-	-2
4	+	+	-	+	-	-	-4
5	-	-	+	+	-	-	6
6	+	-	+	-	+	-	10
7	-	+	+	-	-	+	1
8	+	+	+	+	+	+	13

What are the three generating relations?

How does one use the three generating relations to find all the interactions that are equal to the identity element (i.e., the defining relations)? (1 point)

What two-way interactions are confounded with A?

_____ What is your numerical estimate of the AB interaction?

In this context, how do we interpret things when a main effect is confounded with a two-way interaction. (2 points)

How would you decide whether an estimate of the A effect is significant?

_____

_____

_____ Give the symbolic name for this design (include the resolution).

6. An investment firm wants to compare the Bloomberg terminal with the Reuters 3000 Xtra. Five traders are chosen from among their employees, and each uses one of the two terminals for two days, and then switches to the other one. The response of interest is how much profit or loss a trader makes on each of the four days.

Complete the following ANOVA table. To minimize error ripple, only the blanks with parentheses will be scored.

Source	df	SS	MS	F
Terminal	_____	10	_____	(_____)
Trader	_____	5	_____	(_____)
Interaction	_____	_____	5	(_____)
Error	_____	_____	_____	
Total	_____	39.17		

_____ What is your critical value for testing whether there is an effect due to Trader, at the 0.05 level?

_____ Is there a Trader effect at the 0.05 level?

When do you use Fisher's LSD?

_____

7. The lifespan of a mayfly, in days, has density $f(t) = 6t^5$ for $0 \le t \le 1$.

   _____ What is the value of the hazard function at $t = 1/2$?

   _____ Which, among the four types discussed in class, is this hazard function?

8. _____ Consider the system shown in the figure. Components A and B fail with probability 0.2, component C fails with probability 0.3, and component D fails with probability 0.4 and E never fails. What is the probability that the system fails?
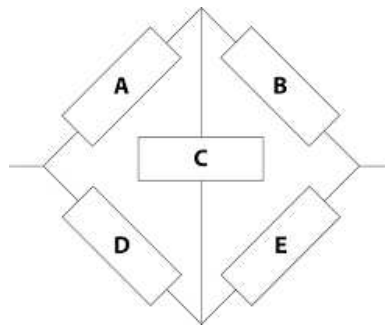
9. _____ A \$10,000 car depreciates by \$1,000 for each year that you own it. An insurance policy that costs \$500/year will reimburse the current value of the car if it is totaled. The probability your car gets totaled in a year is $1/7$. How old is the car when you stop buying insurance?

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

   A. Human beings have increasing hazard rates.
   B. A fractional factorial with resolution II is better than one with resolution III.
   C. The Cox proportional hazards model is good when something wears out at a rate determined by several factors.
   D. People tend to underestimate the risk of being struck by lightning.
   E. The more money someone has, the less they value \$10.
   F. People are about evenly split between risk seekers and risk avoiders.
   G. Cost-benefit analysis in a stochastic world requires statistics.
   H. Roger Boisjoly said "All models are wrong, but some are useful."
   I. Competing risks models are appropriate for most first-person shooter games.
   J. In high dimensions, data tend to concentrate on a subspace.
   K. In regression, extrapolation is safer than interpolation.
   L. Statistics can be hard, but it is useful.