LAST NAME (Please Print): KEY

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.

1. Consider the following design and data.

run	A	B	C	D	E	F	obs
1	-	-	-	-	-	-	3
2	+	-	-	-	+	-	17
3	-	+	-	-	+	+	-2
4	+	+	-	-	-	+	-4
5	-	-	+	-	+	+	6
6	+	-	+	-	-	+	20
7	-	+	+	-	-	-	11
8	+	+	+	-	+	-	13
9	-	-	-	+	-	+	2
10	+	-	-	+	+	+	-8
11	-	+	-	+	+	-	6
12	+	+	-	+	-	-	4
13	-	-	+	+	+	-	10
14	+	-	+	+	-	-	-8
15	-	+	+	+	-	+	-3
16	+	+	+	+	+	+	4

What are the defining relations? (1 point)

The two generators and their product, so ABCE, BCDF and ADEF

What interactions are confounded with A?

Multiply A by the defining relations to get BCE, ABCDF, and DEF.

0.31 What is your numerical estimate of the AB interaction?

Multiply the signs of the A and B columns to get the signs for the AB interaction, i.e., + - - + repeating. Multiply the observations by these signs and average, to get 0.3125.

2^{6-2}_{IV} Give the symbolic name for this design (include the resolution).

The highest order interaction that is confounded with main effects III, so III + I = IV.

In this context, how do we interpret things when a main effect is confounded with a two-way interaction?

We assume that the two-way interaction effect is negligible (or 0).

If none of these factors affects the response, then under the usual assumptions, what is the distribution of the effect estimates?

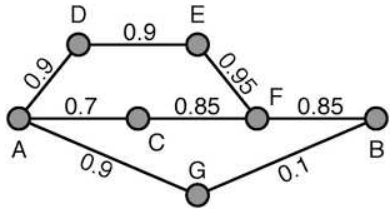
The estimates would be random draws from a normal distribution with mean 0 and unknown standard deviation.

2. 0.67, or 2/3 Let $f(x) = (x/3)\exp(-x^2/6)$ for $x > 0$, and zero else. What is the value of the hazard function when $x = 2$?

Integration shows that the cumulative distribution of this density is $1 - \exp(x^2/6)$, so the survival function is $\exp(x^2/6)$ and thus the hazard function is $x/3$.

Yes Does this have increasing failure rate?

3. _____ Consider the network shown below. If all edges work independently, with probabilities as shown, what is the chance that you can link from A to F?



The ADEF path is open with probabilty a = 0.9 * 0.0 * 0.95 = 0.7695. Path ACF has probability b = 0.595. And path AGBF has probability c = 0.765. The probability that at least one is open is a + b + c - ab - ac -bc + abc = 0.9137.

4. When does multicollinearity occur and why is it bad? (2 pts)

   Multicollinearity occurs when two more explanatory variables are strongly correlated. It reduces predictive accuracy.

5. Explain what ten-fold cross-validation does and how to do it. (5 pts)

   (1) It estimates the predictive accuracy of a fitted model. (1) Divide the training data into tenths at random. (1) Hold out the first tenth, fit the model with the remaining data, and (1) estimate the squared error on the holdout sample. (1) Repeat the process leaving out each tenth, then average the ten estimates.

6. You want to test whether there are differences in IQ between gender and major. You do not have resources to study all possible majors, so, on a whim, you focus upon Statistics, Pyschology, and Economics. You observe two people for each combination of treatment levels. Complete the following table. Only the blanks in the parentheses count.

Source	df	SS	MS	F
Gender	1	10	10	(1.67)
Major	2	5	2.5	(2.5)
Interaction	2	12	6	(6)
Error	6	6	1	
Total	11	33		

   5.14 What is your critical value for the 0.05 level test on Major?

   Yes Do you conclude that Major is significant?

   Because the interaction term is significant.

7. 3 Suppose the value of a \$14,000 car at time t is $14,000/(1+t^2)$ where t is measured in years. Each year, the chance of being totalled is 0.1. At $t = 0$, you purchase a year of insurance for \$100. What is the last year t for which you should purchase insurance?

   At year 0, the expected loss is 14,000 * 0.1 = 1400. At year 1 the expected loss is 7,000 * 0.1 = 700. At year 2 the expected loss is 2800 * 0.1 = 280. At year 3, the expected loss 1400 * 0.1 = 140. At year 4, the expected loss is 823.53 * 0.1 = 82.35. This is less than 100, so stop at year 3.

8. You do an analysis of variance on whether IQs differ by gender. You want to control for the effect of major, and you consider three majors: Statistics, Psychology, and Economics. You take one observation for each gender/major combination. Suppose the sum of squares due to Gender is 20, the sum of squares due to Major is 6, and the total sum of squares is 30. Complete the following table. Only the blanks in parentheses count.

Source	df	SS	MS	F
Gender	1	20	20	(10)
Major	2	6	3	(1.5)
Error	2	4	2	
Total	5	30		

   18.51 What is your critical value for testing the Gender effect?

   $F_{1,2,0.05}$

   4.30, -4.30 If Major were significant at the 0.05 level, what would be your critical value(s) for the least significant difference test?

   $\pm t_{2,0.975}$

   No Was it wise to control for Major?

8 If you had not, what would be your test statistic?

The sums of squres for major would add to the sum of squares for error, and so would the corresponding degrees of freedom. The MSE become 10/4, so the ts is 20/(10/4).

7.71 If you had not, what would be your critical value?

$F_{1,4,0.05}$

9. List all, and only, the true statements. A, B, E

A. Running line smoothers are better than fixed bin-width smoothers.

B. Simmelian ties enforce normative social behavior.

C. In greenlighting the *Challenger* launch, the managers interpolated.

D. Cox proportioal hazards models are appropriate for the lifespans of wild songbirds.

E. Bag-of-words models lose all semantic information.

F. Common risks are overestimated.