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	obs
1	-	-	-	+	+	3
2	+	-	-	-	-	7
3	-	+	-	-	+	-2
4	+	+	-	+	-	4
5	-	-	+	+	-	6
6	+	-	+	-	+	-10
7	-	+	+	-	-	1
8	+	+	+	+	+	13

What is equal to I? (That is, what are the defining relations?) I = ABD = ACE = BCDE.

Multiplying the columns shows that D=AB and E=AC. There are only two such relations, since this is a one-quarter design. Thus I = ABD = ACE = BCDE.

What effects are confounded with E?

ABDE, AC, BCD

-1.75 What is your estimate of the effect due to E?

(3 - 7 + (-2) - 4 - 6 + (-10) - 1 + 13)/8 = -1.75

Explain what the estimate of the E effect describes (2 points)

It is the average effect at the high level of factor E minus the average effect at the low level of factor E, divided by 2.

How would you decide whether this estimate is significant?

I would look for effect sizes that are outliers, or make a normal probability plot.

2^{5-2}_{III} Give the symbolic name for this design (include the resolution).

There are five factors, but only 1/4 of the possible combinations of levels are observed. So this is 2^{5-2}. The resolution is the the sum of the I and the lowest order interaction with which a main effect is confounded—since D is confounded with AB, this is I+II=III.

2. Suppose that the lifespan of a laptop has cumulative distribution function $x^3/27$ for $0 \leq x \leq 3$.

   What is the survival function of a laptop? $1 - \frac{x^3}{27}$

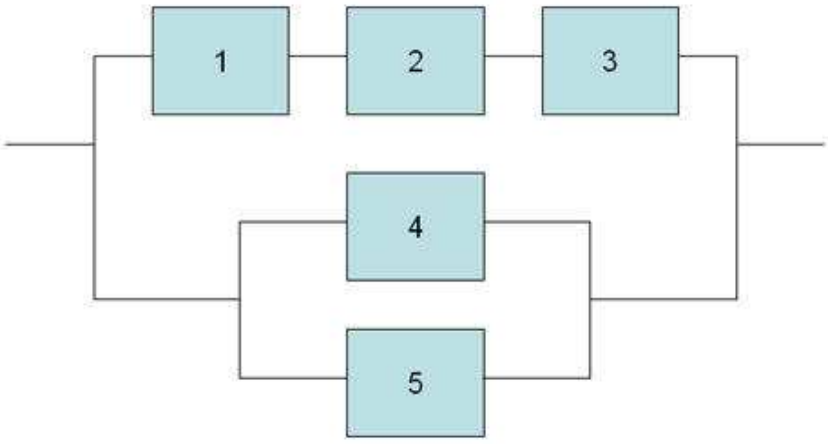
   What is the hazard function of a laptop? $3x^2/(27 - x^3)$

   Grader: Be alert that this fraction might be represented in serveral different ways. $f(x)/[1 - F(x)] = (x^2/9)/[1 - (x^3/27)] = 3x^2/(27 - x^3)$.

   increasing What kind of failure rate does a laptop have? (We discussed four kinds in class.)

   As x increases, the numerator increases while the denominator decreases, so this is an increasing function.

3. Consider a system of the following form:



0.06 Assume that all components in the upper subsystem have have exponential lifespans with $\lambda = 1/2$ and all components in the lower subsystem have exponential lifespans with $\lambda = 1/3$. What is the probability that the system fails in less than one time unit?

The probability that the upper subsystem fails before $x = 1$ is the probability one or more of the three components fail, which is 1 - $\mathbb{P}$[none fail] $= 1 - [\exp(-1/2)]^3 = 0.7769$. And the probability that the lower subsystem fails is the probability that both components fail, which is $[1 - \exp(-1/3)]^2 = 0.0804$. The system fails only if both fail, which has probability 0.7769 * 0.0804 = 0.062.

4. Suppose that the lifespan of a vacuum cleaner follows the Rayleigh distribution, so $F(t) = 1 - \exp(-\theta_0 t - \theta_1 t^2/2)$. You observe three vacuum cleaners: one lasts one year, one lasts two years, and one lasts three years. In the special case when $theta_0 = \theta_1 (= \theta)$, find the numerical value for the maximum likelihood estimate of θ.

0.23, or $3/13 = \hat{\theta}$

Under the assumption about the equality of the coefficients, $F(t) = 1 - \exp(-\theta t - \theta t^2/2)$ so $f(t) = (\theta + \theta t^2)\exp(-\theta t - \theta t^2/2)$. The likelihood function is

$$\begin{aligned} f(\theta \,|\, t_1, \ldots, t_n) &= \prod_{i=1}^{n} (\theta t_i + \theta t_i^2) \exp(-\theta t_i - \theta t_i^2/2) \\ &= \theta^n * \exp[\sum_{i=1}^{n} -\theta t_i - \theta t_i^2/2)] * \prod_{i=1}^{n} (\theta t_i + \theta t_i^2). \end{aligned}$$

Take the log, to get

$$\ln f(\theta \,|\, t_1, \ldots, t_n) = n \ln \theta - [\sum_{i=1}^{n} \theta t_i + \theta t_i^2/2)] + \text{ stuff that doesn't include } \theta.$$

Now maximize this by taking the derivative wrt θ, so

$$\frac{d}{d\theta} \ln f(\theta \,|\, t_1, \ldots, t_n) = \frac{n}{\theta} - [\sum_{i=1}^{n} t_i + t_i^2/2)],$$

and then setting this to zero and solving. One gets that $\hat{\theta} = n/[\sum_{i=1}^{n} t_i + t_i^2/2)]$. For these numbers, one has $\hat{\theta} = 3/13$.

5. Suppose you have done a one-way ANOVA that looks at sentence lengths (in days) for bar brawlers in Durham County, Alamance County, and Chatham county. Use $\alpha = 0.05$. You have 4 observations from Durham, 6 from Alamance, and 4 from Chatham. Complete the following partial ANOVA table.

Source	df	SS	MS	F
County	2	70	35	3.5
Error	11	110	10	
Total	13	180		

$\pm$ 2.20 What is the critical value for Fisher's LSD in this situation?

It is a t with 11 degrees of freedom and $\alpha = 0.025$.

Grader: Don't mark off if the $\pm$ is omitted.

-1.34 Suppose the average sentence in Durham is 5 days in jail, and the average sentence in Chatham is 8 days in jail. What is your test statistic if you were to use Fisher's LSD?

$ts = \frac{5-8}{\sqrt{10(\frac{1}{4}+\frac{1}{4}}} = -1.34.$

Is it appropriate to use Fisher's LSD in this case? Why or why not?

No, since the ANOVA test was not significant.

6. Assume that flood insurance for your home costs \$100/month, fire insurance costs \$200/month, and a policy that covers both costs \$250/month.

   In your area, the annual probabilty of a flood is 1/100 and the annual probability of a fire is 1/40. You plan to live in your home for 30 years, and your home costs \$100,000.

   \$89,211.57 What is the expected loss from purchasing flood insurance alone?

   Expected loss is the cost of the policy + the expected loss to the home. The cost of the policy is 100 * 12 * 30 = 36,000. The expected loss to the home is 0 if there is a flood and 100,000 if there is a fire. The probability of one or more fires in 30 years is 1 - P[no fire] = $1 - (39/40)^{30} = 0.532$. Thus the expected loss is 89,211.57.

   Note to Grader: Accept anything between 89,205 and 89,215.

What is your best choice? Don't buy insurance.

Using similar logic, the expected loss from the fire insurance is 72,000 + (100,000)*0.260) = 98,000. If one buys both fire and flood together, the house has no loss but the premium totals to 90,000. And if one buys no insurance, the expected loss is 100,000 * P[fire or flood] = 100,000 * [(0.532) + (0.260) - (0.532)(0.260)] = 86,170. So you shouldn't buy the insurance.

7. Which model is better for the lifespan of your goldfish: competing risks or the Cox proportional hazards model?

   Cox proportional hazards model. The goldfish wears out.

8. List all, and only, the true statements. D, F

   A. A fractional factorial with resolution III is better than one with resolution IV.

   B. The exponential distribution has decreasing failure rate.

   C. People tend to overestimate risks that are common.

   D. People evaluate risks in terms of their ability to control them, and whether the risk has high or low dread.

   E. People have a linear perception of the value of money.

   F. Significance probability is the chance of observing data that support the alternative hypothesis as more strongly than the data that were collected, when the null hypothesis is true.

9. You want to know whether there are differences among freshmen, sophomores, juniors and seniors in their performance in a statistics class. And you think there may also be a dormitory effect, so you arbitrarily compare students from Brown, Jarvis and Wilson. This is a balanced two-way ANOVA with two observations on each combination of factor levels.

   Complete the following partial ANOVA table:

Source	df	SS	MS	F
Year	3	6	2	0.33
Residence	2	14	7	3.5
Interaction	6	36	6	3
Error	12	24	2	
Total	23	80		

Mixed Is this a fixed-effects, random-effects, or mixed-effects design?

4.76 What is the critical value for a test of whether Year is significant? ($\alpha = 0.05$.)

This is an F with 3 df in the numerator and 6 in the denominator, since for a mixed-effects table the fixed effect is compared to the interaction term.

0.33 What is the value of the test statistic for deciding whether Year is significant?

Yes Do you conclude that Year is significant at the 0.05 level?

The interaction is significant, so both main effects are signficant.