Statistics Qualifying Exam

9:00 am - 1:00 pm, Monday, August 15, 2016

1. Let us choose at random a point from the interval $(0, 1)$ and let the random variable X_1 be equal to the number which corresponds to that point. Then choose a point at random from the interval $(0, x_1)$, where x_1 is the experimental value of X_1; and let the random variable X_2 be equal to the number which corresponds to this point.

   (a) Make assumptions about the marginal probability density function (pdf) $f_1(x_1)$ and the conditional pdf $f_{2|1}(x_2|x_1)$.

   (b) Compute $P(X_1 + X_2 \geq 1)$.

   (c) Find the conditional mean $E[X_1|x_2]$.

2. Assume the random variable X follows a Poisson distribution with a probability mass function (pmf) as

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots,$$

   where $\lambda > 0$ is the mean parameter.

   (a) Show that the moment generating function (mgf) is $M_X(t) = e^{\lambda(e^t - 1)}$.

   (b) Use the mgf to find $\mathrm{Var}(X)$.

   (c) Suppose that Y_1 has a Poisson distribution with mean λ_1, and that the conditional distribution of Y_2 given $Y_1 = y_1$ is Binomial on y_1 trials with success probability p, $0 < p < 1$. Find the marginal distribution of Y_2.

3. Let $X_1, , X_n$ are identically and independently distributed with the pdf as $f(x|\theta) = \theta x^{\theta-1}$, $\theta > 0$, $0 < x < 1$.

   (a) Find a sufficient and complete statistic for θ.

   (b) Find the maximum likelihood estimator (MLE) of θ. Is the MLE an unbiased estimator for θ? **Show it**.

   (c) Find the minimum variance unbiased estimator (MVUE) of θ.

   (d) Does MVUE of θ achieve the Rao-Cramer lower bound? **Show it**.

4. Let $Y_1 < Y_2 < Y_3 < Y_4$ be the order statistics of a random sample of size $n = 4$ from a distribution with pdf $f(x; \theta) = 1/\theta$, $0 < x < \theta$, zero elsewhere, where $\theta > 0$. The hypothesis $H_0 : \theta = 1$ is rejected and $H_1 : \theta > 1$ is accepted if the observed $Y_4 \geq c$.

   (a) Find the constant c such that the significance level is $\alpha = 0.05$.

   (b) Determine the power function of the test.

5. Consider the following ANOVA table for the multiple linear model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon,$$

where the error term ϵ are independent and identical normal random variables with mean 0 and variance σ^2.

Source	SS	df	MS
Model	2176606	3	725535
X_1	136366	1	136366
$X_2 \mid X_1$	2033565	1	2033565
$X_3 \mid X_1, X_2$	6675	1	6675
Error	985530	48	20532
Total	3162136	51	

(a) Find $SSR(X_2, X_3 \mid X_1)$.

(b) For the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$, using the answer to part a) test the following hypothesis using $\alpha = 0.05$:

$$H_0 : \beta_2 = \beta_3 = 0 \text{ vs. } H_1 : \text{ Not both } \beta_2 \text{ and } \beta_3 = 0.$$

(c) Compute the coefficient of partial determination between Y and X_2 given that X_1 is in the model, i.e., find $R^2_{Y2|1}$. Interpret the result.

(d) With the given information, can a "best" multiple linear regression model be selected using a forward selection procedure for this application? If yes, show your detailed procedure; if no, explain.

6. A clay tile company is interested in studying the effects of cooling temperature on strength. Since the company has five ovens which produce the tiles, four tiles were baked in each oven and then randomly assigned to one of the four cooling temperatures (° C). The data are shown below.

Cooling	Oven					
Temp	1	2	3	4	5	Mean
5	3	10	7	4	3	5.40
10	3	8	12	2	4	5.80
15	9	13	15	3	10	10.00
20	7	12	9	8	13	9.80
Mean	5.50	10.75	10.75	4.25	7.50	7.75

(a) Give an appropriate model to analyze this data, and state the assumptions.

(b) If $MSE = 6.275$, compute the F-statistic to determine if there is a difference among the four cooling temperatures (use $\alpha = 0.05$).

(c) Perform pairwise comparisons for the four cooling temperatures using Tukey's procedure. What are your conclusions?

(d) Suppose the company believes there is a jump in the strength at 12.5° C but otherwise cooling temperature has no effect (i.e., step function $- - -\_\_\_$). To test this, we find the following contrasts:

$$\mathbf{C}_1 = (1, -1, 0, 0)$$

$$\mathbf{C}_2 = (0, 0, 1, -1)$$

$$\mathbf{C}_3 = (1, 1, -1, -1)$$

Test these contrasts *simultaneously* using the Bonferroni's method with $\alpha = 6\%$.

7. An engineer suspects that the surface finish of a metal part is influenced by the feed rate and the depth of cut. He selects three feed rates and four depths of cut. He then conducts a factorial experiment and obtains the following data.

	Depth of Cut (in)			
Feed rate (in/min)	1	2	3	4
1	74, 64, 60	79, 68, 73	82, 88, 92	99, 104, 96
2	92, 86, 88	98, 104, 88	99, 108, 95	104, 110, 99
3	99, 98, 102	104, 99, 95	108, 110, 99	114, 111, 107

Some summary statistics are give below.

```
Grand mean: 94.333

feed        mean      |     depth   mean
1          81.583     |     1       84.778
2          97.583     |     2       89.778
3         103.833     |     3       97.889
                      |     4      104.889
------------------------------------------------
feed    depth       mean
1       1           66.000
1       2           73.333
1       3           87.333
1       4           99.667
2       1           88.667
2       2           96.667
2       3          100.667
2       4          104.333
3       1           99.667
3       2           99.333
3       3          105.667
3       4          110.667
```

Suppose the following statistical model is used to fit the data.

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \epsilon_{ijk}; k = 1, 2, 3$$

where τ_i ($i = 1, 2, 3$) and β_j ($j = 1, 2, 3, 4$) are the effects of feed rate and cut depth, and $(\tau\beta)_{ij}$ are their interactions. For parameter estimation, we impose the following constraints as in the lecture notes: $\sum_i \tau_i = \sum_j \beta_j = \sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0$. The ANOVA of the data was done in SAS and the output is shown.

```
Dependent Variable: finish
                          Sum of
Source      DF          Squares        Mean Square     F Value   Pr > F
Model       11         5842.667         531.151515       18.49    <.0001
Error       24          689.333          28.722222
Co Total    35         6532.000

ource       DF          Type I SS      Mean Square   F Value    Pr > F
feed         2         3160.500        1580.250        55.02     <.0001
depth        3         2125.111         708.370        24.66     <.0001
feed*depth   6          557.056          92.843         3.23     0.0180
```

(a) What are the estimates of τ_1 and $(\tau\beta)_{22}$?

(b) Test if the interaction between feed rate and cut depth is significant. To get full credits, give a test statistic, the corresponding p-value and your conclusion.

(c) If we plan to perform pairwise comparison for *all* treatment combinations, which procedure should we use? What is the corresponding critical difference (using $\alpha = 5\%$)?

(d) Use the Bonferroni method to compare the following treatments (i.e. the level combinations of speed and depth): (2,3), (2,4), (3,3) and (3,4), *pairwisely* (Use $\alpha = 6\%$). Calculate the critical difference and report your results of comparison. You can report the result as we have seen in SAS output by labeling significantly different combinations with different Latin letter.

8. Answer the following questions.

(a) An engineer is interested in the effects of cutting speed (A), tool geometry (B), and cutting angle (C) on the life (in hours) of a machine tool. Two levels of each factor are chosen, and three replicates of a 2^3 factorial design are run. The results follow:

factor			replicate		
A	B	C	I	II	III
-	-	-	22	31	25
+	-	-	32	43	29
-	+	-	35	34	50
+	+	-	55	47	46
-	-	+	44	45	38
+	-	+	40	37	36
-	+	+	60	50	54
+	+	+	39	41	47

Please complete the missing values for the degrees of freedom (DF) and Source in the following tables indicated by "???".

Source	DF	Sum of Squares
Model	???	
Error	???	
Corrected Total	???	

Source	DF	Type I SS
A	???	
???	???	
???	???	
???	???	
???	???	
???	???	
???	???	

(b) Suppose I run a 2^{6-2} design using

$$I = ABCDE = CDEF = ABEF.$$

If I include all main effects and two-way interactions in my model statement, fill in the following ANOVA table with both Type I and Type III degrees of freedom.

Source	Type I df	Type III df
A		
B		
C		
D		
E		
F		
AB		
AC		
AD		
AE		
AF		
BC		
BD		
BE		
BF		
CD		
CE		
CF		
DE		
DF		
EF		
Error		
Total		