Applied Statistics Qualifier Examination
(Part II of the STAT AREA EXAM)
May 23, 2018; 11:00AM-1:00PM

Instructions:

(1) The examination contains 4 Questions. You are to answer 3 out of 4 of them. *** Please only turn in solutions to 3 questions ***

(2) You may use up to 4 books and 4 class notes, plus your calculator and the statistical tables.

(3) NO computer, internet, cell phone, or smart watch is allowed.

(4) This is a 2-hour exam due by 1:00 PM.

Please be sure to fill in the appropriate information below:

I am submitting solutions to QUESTIONS ____ , ____ , and ____ of the applied statistics qualifier examination. Please put your name on every page of your exam solutions, and add page number for solutions to each question individually.

There are ________ pages of written solutions.

Please read the following statement and sign below:
This is to certify that I have taken the applied statistics qualifier and have used no other person as a resource nor have I seen any other student violating this rule.

____________________________________
(Signature)

____________________________________
(Name)
1. The following table gives the group means (5 observations per group) for a two-way ANOVA.

<table>
<thead>
<tr>
<th>Variable 2</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>3.1</td>
<td>5.2</td>
<td>5.9</td>
</tr>
<tr>
<td>High</td>
<td>3.9</td>
<td>1.2</td>
<td>1.9</td>
</tr>
</tbody>
</table>

(a) Given that the MSE was 0.2, would you expect there to be a significant interaction in the ANOVA analysis? Provide justification to your answer.

(b) Suppose now you have a two-way ANOVA where some of the cells have no data because those combination of factor levels is physically impossible to achieve. The following table gives the sample sizes in each cell.

<table>
<thead>
<tr>
<th>Variable 2</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>High</td>
<td>7</td>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

Since we do not know how to analyze two-way ANOVA with empty cells, can you think of a way to analyze this as a one-way ANOVA? Write down the model and interpret the parameters in your model including variance(s) in terms of the original design.
2. Let $Y_i$ be a $Bin(n_i, \pi_i)$ variate for group $i, i = 1, \ldots, N$, with $\{Y_i\}$ independent and $n = \sum_{i=1}^{N} n_i$.
   a. Regard the data as a $N \times 2$ table, i.e. column one for $y_i$ and column two for $n_i - y_i$. Write out the Pearson’s statistic $X^2$. What hypothesis testing can $X^2$ be used for?
   b. When all $n_i = 1$, is it reasonable to use the Pearson’s statistic $X^2$ for testing the model fit? Justify your answer.
   c. Suppose there is a covariate $x_i$ for each group $i$, and consider the model $\logit(\pi_i) = \alpha + \beta x_i$. When all $n_i = 1$, is it reasonable to use deviance to check model fit? Justify your answer.
   d. Consider the model that $\logit(\pi_i) = \alpha + \beta_i$, where $\beta_N = 0$. Given $\{\pi_i > 0\}$, show how to find the MLE of $\{\beta_i, i = 1, \ldots, N - 1\}$. 
3. The dependent variable $Y$ is related to $x_1$ and $x_2$ by the model:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \alpha Z,$$

where $Z$ is $N(0,1)$. All random errors are independent of each other. A research team will examine six settings of the independent variables $(x_1, x_2)$. They will observe $J$ observations with $(x_1, x_2) = (-5,5)$; $J$ observations with $(x_1, x_2) = (-3,-1)$; $J$ observations with $(x_1, x_2) = (-1,-4)$; $J$ observations with $(x_1, x_2) = (1,-4); (x_1, x_2) = (3,-1);$ and $J$ observations with $(x_1, x_2) = (5,5)$. Find the expected value of the corrected sum of squares for the regression of $Y$ on $x_1$ and $x_2$. 
4. A research team has been hired by a state education department to evaluate the teaching of special education students in the state. There is a measure of the quality of a student’s performance called $Y$, the result of a standardized examination. The researchers will use the model

$$Y_{ijr} = \mu + A_i + B_{j(i)} + \sigma_z e_{(ij)r}$$

where $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $r = 1, \ldots, R$. That is, the nested design is balanced. They propose to study $I = 5$ districts selected at random; the state has such a large number of districts that there is no need for finite population correction. That is, the effect of the $i$-th district is random and represented by $A_i$, where $A_i$ are normally and independently distributed random variables with expected value 0 and variance $\sigma^2_A$. The research team proposes to take a random sample of $J$ special education classes from each district. The random variables $B_{j(i)}$ are normal and independently distributed with expected value 0 and variance $\sigma^2_B$. They will select $R$ students at random from each class and measure each student’s performance. Note that the students are nested within classes and that classes are nested within districts. The random variables $Z_{(ij)r}$ are independent and identically distributed normal random variables with mean 0 and variance 1. Each set of random variables is independent of the other sets.

The research team will test $H_0 : \sigma_A^2 = 0$ against the alternative $H_1 : \sigma_A^2 > 0$ at the 0.01 level. What test statistic should they use?

They want to know the probability of a Type II error when $I = 5, \sigma_A^2 = 225, J = 3, \sigma_B^2 = 50, R = 4, \text{ and } \sigma_z^2 = 60$. What is the probability of a Type II error under these conditions with this statistic?