STAT 505/506   MS COMPREHENSIVE EXAM–TAKE HOME

Released 6 Jan 2017 at 3pm and due 9 Jan 2017 by 9am **OR**
Released 9 Jan 2017 at 3pm and due 12 Jan 2017 by 9am

Instructions: Make sure to read each question carefully and ask for clarification as needed. You are allowed to use any resources from Stat 505/506, the internet and resources from any other course. **If you do use non-Stat 505/506 resources they must be referenced, including internet resources, but you may not receive help from other people except Laura. Do NOT discuss this exam with other students until Thursday 12 Jan 2017!** Please provide your numbered answers separately (you do not need to include the questions) and email your answers to `laura.hildreth@montana.edu` and `jobo@montana.edu` your answers by either 9 Jan 2017 at 9am (if picked up exam on 6 Jan 2017) OR 12 Jan 2017 at 9am (if picked up exam on 9 Jan 2017). **By turning in this exam, you acknowledge that you have completed this exam in accordance with the Student Conduct Code for Academic Honesty found online at** `http://www.montana.edu/policy/student_conduct/academicmisconduct`. Failure to comply with this code will result in an automatic score of 0, failure of the comprehensive exam, and you will be reported for academic dishonesty.

The questions below are based on the article *Borderline Personality and the Detection of Angry Faces* by Hepp et al (2016). The pdf of this article is found at:
`http://journals.plos.org/plosone/article/asset?id=10.1371/journal.pone.0152947.PDF`
and the data are found at
`http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0152947.`
Unless otherwise specified the following questions refer to Study 1.

1. When you look at the data you'll notice that they are not in the format we would like (each row is a subject and each column is a variable–in other words we have a wide data set but would like a long data set). Reformat the data set so it is in a usable format. Provide your R/SAS code and the first 10 rows of the reformatted data set–please feel free to delete variables that are not of interest (which is most of them). [7pts]

2. Describe to what population the authors are able to infer the results of their study. [5pts]

3. In their analysis, the authors used a repeated measures ANOVA. Write out the model used by the authors, defining each parameter in the model, providing the covariance structure of the error term(s), and clearly stating the assumptions being made. [10pts]

4. Reproduce the summary statistics and confidence intervals found in Table 1. [5pts]

5. Explain what is being plotted in Figure 2 to a non-statistician by explaining what information this plot provides, how to interpret the plot, and what this plot suggests about the relationship between the two explanatory variables and the response. [7pts]

6. The authors stated that they removed "[e]xtreme outliers with a reaction time more than 2.5 SD above the sample mean." Explain whether you think this is or is not a reasonable justification for removing observations. [7pts]

7. Fit the model used by the authors. **You do not need to remove the outliers as the authors did**. Because of this and also because it is not clear the estimation method used by

the authors you (likely) will not obtain the exact same results though yours should be close. What we are concerned with in this problem is if you used the correct code in your analysis. [5pts]

8. In Table 2 the authors report the (Greenhouse-Geisser) df for the F-tests of the interaction effect and the main effect. Explain why the denominator df are not the same for the tests of the two main effects. [7pts]

9. Provide the hypotheses being tested by the authors in Hypothesis 2 in terms of the model parameters from Problem 3. [5pts]

10. Recreate the post-hoc analysis conducted by the authors in Hypothesis 1. Clearly explain how the authors obtained their confidence intervals and whether this approach is appropriate. [10pts]

11. The authors chose to log transform the response due to issues with non-normality. Carefully explain if this transformation was "successful." [7pts]

12. Lastly let's look at the analysis (repeated measures ANOVA) used by the authors. Explain the advantages of using a repeated measures ANOVA for this analysis in terms of modeling the covariance structure of the error terms. One potential covariance structure not considered by the authors is a heterogeneous covariance structure such that the variance of the error terms differs for each level of an explanatory variable (i.e. the variance of the error terms for level 1 of the explanatory variable is $sigma_1^2$, the variance of the error terms for level 2 is $sigma_2^2$, and so forth). Choose one of the categorical variables provided in the data set by the authors and sketch out the covariance structure of the error terms under the heterogeneous covariance structure. Fit the model with the heterogeneous covariance structure. [15pts]