MQM Comprehensive Examination
Statistics Major

Day 1

Answer both of the following two questions. Begin each response on a new page, and clearly number the question to which you are responding.

1. Statistical significance, testing and the likelihood

   a. What is a p-value or significance value? Describe how the p-value relates to the likelihood or distribution function of a test statistic.

   b. Why do so many researchers present significance values in their reports or research investigations (i.e, what can we learn from a p-value)?

   c. There are many different ways to perform hypothesis tests. Discuss the particular form of hypothesis test known as the Likelihood Ratio Test or LRT. How does the LRT work and why does it make sense? You may discuss either the simple or generalized LRT.

2. In mixed linear models, the fixed effects can be estimated via maximum likelihood independently of the variance components, but only if the data are balanced. If the data are unbalanced, the maximum likelihood estimates of the fixed effects depend upon the unknown variance components. Discuss in some detail how this problem has been successfully addressed (e.g., how have people simultaneously estimated the fixed effects and variance components in the context of unbalanced data?)
Answer two of the following four questions. Begin each response on a new page, and clearly number the question to which you are responding.

3. ANOVA: We wish to know whether or not there are any differences of science scores between students of different race/ethnicity.
   a. What assumptions must be satisfied in order to use and interpret the F-test in ANOVA?
   b. We wish to know whether the mean science score of Asian (Group 1) and White (Group 2) groups are different from the mean of Black (Group 3), Hispanic (Group 4), and Native American (Group 5) groups. Write the hypothesis to be tested for this study using symbols.
   c. Write the orthogonal contrast for the hypothesis in part B.

4. Sampling Distributions: The concept of the sampling distribution is critical to inquiry in statistics and psychometrics. Answer the following questions about the sampling distribution, using a specific statistic of your choice to illustrate your answers.
   a. What is a sampling distribution?
   b. Name two statistical properties you can investigate if you know the form of the sampling distribution of your statistic. Discuss how the sampling distribution informs you about each property.
   c. Describe one method that researchers use to determine the nature of the sampling distribution for a statistic.
5. Paired versus unpaired observations

a. Provide the formulas for the following.
   
i. Two-sample t-test
ii. Paired t-test
iii. t-test for a regression coefficient for a dummy variable

b. What are the critical differences between the statistics in i. and ii.?

c. Under what conditions are i. and ii. identical?

d. Express ii. in terms of iii. (Hint: the sample variance used in ii. can be re-expressed as $s_{y|x} = (1-r_{xy}^2)(s_y)^2$, and $r_{x,y} = (\beta)(s_x / s_y)$).

6. Linear Regression: A simple linear regression is used to predict students’ science test scores from their SES.

a. Write the population model for this regression, defining all the terms in the model.

b. What are the assumptions for this population model?