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 item to which you are responding. **You may not use books or notes when answering these questions.**

1. Inferential procedures
   
   a. Differentiate between inferential procedures based on hypothesis testing and those based on estimation procedures.
   
   b. Explain the role of the sampling distribution in each of the two types of inferential procedures.
   
   c. Indicate whether and how each of the following concepts function in each of these two approaches to inference.
      
      i. Type I and Type II errors
      ii. standard errors
      iii. the null hypothesis
      iv. confidence intervals

2. Consider the independent, identically distributed random variables $X_1$ through $X_n$, with density $f(x)$. Often the transformation $g(X_i)$ is of interest.
   
   a. Describe two situations (e.g., give examples) in which one might be interested in a transformation of the original random variables $X_1$ through $X_n$. Why would the original variables $X_1$ through $X_n$ not be of interest in these situations?
   
   b. Suppose you wanted to obtain a density for the new variables $g(X_1)$ through $g(X_n)$. Describe two different methods you might use to obtain such a distribution. What restrictions apply to the use of each approach (e.g., can it be used for any function $g(x)$, are there conditions on $X$ or $f(x)$, etc.)?
Answer two of the following four questions. Begin each response on a new page, and clearly number the item to which you are responding. **You may not use books or notes when answering these questions.**

3. **Sampling distributions:** Suppose sampling distributions were constructed by generating 5000 random samples of a sample size of 30.
   a. Characterize the shape of the sampling distribution of a mean if the distribution of the population is distributed as a uniform distribution.
   b. Summarize the theorem that is illustrated by the scenario in part A.
   c. Characterize the shape of the sampling distribution of the variance if the distribution of the population is distributed as a normal distribution.

4. **Surveys:** You have been asked to consult with researchers who want to measure reading comprehension of Michigan adolescents. The researchers hope to generalize to all students in grades 6 through 12 in the entire State. The researchers are considering two sampling plans, the simple random sample (SRS) and a two-stage cluster sample with schools serving as clusters, with a sample of 20 students per school.
   a. List at least 2 assumptions that underlie each design.
   b. How likely are these assumptions to be true?
   c. What additional information do you need to evaluate these designs?
   d. Select one of these designs and describe its advantages and disadvantages for the researchers’ purposes.

5. **Explain in detail the important role that latent roots and vectors (eigenvalues and eigenvectors) play in multivariate hypothesis testing theory.**
6. In using the general linear model, programs such as SAS and SPSS allow the user to specify categorical variables as independent variables in the analysis. For example, the CLASS statement is used for declaring categorical variables in SAS.

   a. Explain in detail why such programs produce estimates for multiple estimators (\(\beta\)'s) for each categorical variable.

   b. How would each estimator be interpreted? (You may want to outline a simple model here to use in answering this question.)