Instructions: Please answer each of the following questions

Important: Very few of these responses should be simple statements or one-word answers. Be sure to adequately elaborate on each response through explanation and clarification as if you were teaching this to someone with no background in the subject.

Question I: Bivariate Hypothesis Testing

1. In no more than four sentences (may require less) define/explain each of these terms common in bivariate hypothesis testing procedures.

   a) Null hypothesis
   b) Alternative hypothesis
   c) Test statistic (Critical value)
   d) Region of Rejection
   e) Type I Error
   f) Type II Error
   g) Statistical Power ($1 - \beta$)
   h) One-tailed vs. Two-tailed test
   i) Confidence Level
   j) T-test
   k) ANOVA
   l) Among group variation
   m) Total variation
   n) Within group variation

2. Answer the following questions as they pertain to examination of differences in means between more than one group.

   a) The T-test, ANOVA, and Chi-Square test are all way so of detecting variable associates via examinations of groups differences and associations. In what instance would you expect each of the three tests to be used?

   b) Pertaining to the first two tests listed above, how are the formal null hypotheses stated? What are the meanings of these formal statements?
c) What distributions are each the test-statistics examined against? Why are the three distributions different?

d) Suppose you must choose the one- or two-tailed version pertain to certain tests mentioned above. In what cases would a one-tail test appropriate? In what case would a two-tail test be appropriate? Why?

Question II: Multivariate Regression Analysis

**OLS (see attached output)**

Familial disruption has been linked to higher levels of social disorganization and crime rates in research in the area of ecological criminology. However, levels of familial disruption have also been shown to be significantly related to regional differences in crime rates. Using county level data, the attached output has been compiled to test for the potential effects of being a Southern County (“south”) and the county level percent divorced (“pctdiv”) on the index crime rate of the county (“indexrt”).

Interpret the output by detailing the results of the analysis and referring to the appropriate tables in your attempt to answer this question. Be sure to properly, and formally, interpret all appropriate statistics from the output.

In doing so, focus on three basic research questions:

1) What are the basic assumptions of the OLS regression approach? How are each tested in this case? … does this data violate any of these assumptions?

2) Is there a statistical relationship between the % divorced in a county and the crime rate in the county? How do you know? Interpret the slope and Beta coefficients associated with this IV.

3) Is that relationship mediated by the differentials in crime rates between counties in the South and the rest of the country? How do you know? Interpret the slope and Beta coefficients associated with this IV.

4) Do the IV’s significantly predict the county level index crime rate? How can we tell the magnitude of model fit? Interpret any appropriate statistics.

**Logistic (see attached output)**
Using survey data associated with neighborhood conditions, fear, and demographics suppose an analysis of one’s fear of their neighborhood was conducted. In the dataset there are a series of variables, including a binary indicator of neighborhood fear (1 = “ever” feeling unsafe in one’s neighborhood in reference to 0 = never feeling unsafe). For this question then, we are predicting ever feeling unsafe in one’s neighborhood by race (being white), gender (being male) and by age.

Interpret the output by detailing the results of the analysis and referring to the appropriate tables in your attempt to answer this question. Be sure to properly, and formally, interpret all appropriate statistics from the output.

In doing so, focus on three basic research questions/directives:

1) What are the basic assumptions of the Logistic regression approach? How does this differ from the OLS approach?... and what inherent violations of the OLS approach make using the Logistic Regression approach necessary (hint: refer to violations of OLS assumptions)?

2) Is there a statistical relationship between likelihood of feeling unsafe in one’s neighborhood (binary form) based on the respondent’s race (being black), gender (being male), age, and employment status (full time worker)? How do you know in each case?

3) Interpret the results as they relate to each of the IV’s. Report your responses in technical form of odds ratios for interpretation of the statistical relationships.

4) Do the IV’s significantly improve our ability to predict the likelihood of feeling unsafe in one’s neighborhood? How can we tell from the model fit statistics?