1. (25 points) You have data on the starting salary of 200 business school students of the class of 2000, as well as their GMAT test scores, gender, whether they concentrated in finance, management information systems, or some other area, and whether they hold an undergraduate degree in science or engineering. You run the following regression:

$$\begin{aligned} \ln SAL = \; & \beta_0 + \beta_1 \text{ GENDER} + \beta_2 \text{ FINANCE} + \beta_3 \text{ MIS} + \beta_4 \text{ SCIENCE} + \\ & \beta_5 \text{ SCIENCE * GENDER} + \beta_6 \text{ SCIENCE * FINANCE} + \\ & \beta_7 \text{ SCIENCE * MIS} + \beta_8 \text{ SCIENCE * GENDER *FINANCE} + \\ & \beta_9 \ln \text{GMAT} + \epsilon \end{aligned}$$

Where ln SAL and ln GMAT are the natural logarithms of the student's starting salary and GMAT test scores, respectively, and:

GENDER = 1 if student is male, else zero,
FINANCE = 1 if student concentrates in finance, else zero,
MIS = 1 if student concentrates in management information systems, else zero,

(NOTE: you can only have ONE concentration)

SCIENCE = 1 if student has an undergraduate degree in one of the physical sciences or engineering, else zero.

Furthermore, the interaction variables are:

SCIENCE * GENDER = 1 if male and has an undergraduate degree in one of the physical sciences or engineering, else zero,
SCIENCE * FINANCE = 1 if has an undergraduate degree in one of the physical sciences or engineering and is concentrating in finance, else zero,
SCIENCE * MIS = 1 if has an undergraduate degree in one of the physical sciences or engineering and is concentrating in management information systems, else zero,
SCIENCE * GENDER * FINANCE = 1 if has an undergraduate degree in one of the physical sciences or engineering, is male and is concentrating in finance, else zero.

a. Carefully interpret each of the ten regression coefficients. (Hint: write out all the possible combinations for all the types of students.)
b. Indicate the parameter restrictions and the null hypothesis corresponding to a test of the following hypotheses:

  – Business school graduates have the same expected starting salary, regardless of their gender, given their areas of concentration and test scores.

  – The percent effect of area of concentration on starting salary is the same for all possible concentrations, for both males and females, given their undergraduate degree and their GMAT test scores.

– Male business school graduates with an undergraduate science degree who concentrate in finance do no better in terms of starting salary than male business school graduates with an undergraduate degree in a non-science or non-engineering area who concentrate in finance, given GMAT test scores.

2. (45 points) The Major League Baseball Player's Association has hired some Brandeis Economics students to model and estimate player compensation. Using data from the 2003 season, the students come up with the following model:

$$lsal = \beta_0 + \beta_1 rbi + \beta_2 bb + \beta_3 e + \epsilon$$

where lsal is the natural log of the player's compensation, rbi are runners batted in, bb are the number of walks, e are the number of errors committed, and ϵ is the unobserved random error term.

The students decide to estimate this model using OLS and obtain the following STATA output:

```
  Source |       SS       df       MS                  Number of obs =     368
---------+------------------------------               F( x,    x)   =
   Model |  194.855252     3  64.9517508               Prob > F      =
Residual |  414.448406   364  1.13859452               R-squared     =  0.3198
---------+------------------------------               Adj R-squared =  0.3142
   Total |  609.303658   367  1.66022795               Root MSE      =

--------------------------------------------------------
    lsal |      Coef.   Std. Err.       t     P>|t|
---------+----------------------------------------------
     rbi |   .0136069   .0030347      4.484   0.000
      bb |   .0139046   .0036653      3.794   0.000
       e |  -.0372286   .0117379     -3.172   0.002
   _cons |   13.21696   .1011485    130.669   0.000
---------------------------------------------------------
```

a. Carefully interpret each of the four regression coefficients.

b. Formulate and test the hypothesis that the ceteris paribus effect of walks on salary is greater than 1.5% at the 95% confidence level. (H0: $\beta_2 > 0.015$, H1: $\beta_2 \leq 0.015$)

c. Construct the 80% confidence interval around the population parameter on the errors variable, β_3 .

d. Formulate and test for the over-all significance of the model.

3. (20 points) From the regression, given in question 2, above, the covariance between the OLS estimator on rbi's, β_1, and on errors, $\beta_{3,}$ is calculated to be -0.000011 (this was found using the "vce" command in STATA after running the regression on rbi's, walks, and errors,). Using this information, formulate and test the null hypothesis H_0: $\beta_1 - \beta_3 = 0$ against the alternative hypothesis, H_1: $\beta_1 - \beta_3 \neq 0$ (HINT: this question can be thought about as a question on functions of random variables and the construction of confidence intervals. Remember that the OLS estimators for β_1 and β_3 are just random variables!)

4. (15 points) Suppose that we are interested in investigating whether there exists wage discrimination by race and you obtain data on wages, gender, union membership, race, education, and experience. You start your investigation by using a simple dummy variable model that can help you better understand how these characteristics affect wages. In particular, you estimate the following equation:

$$\text{lnWAGE} = \alpha + \alpha_F \text{ FE} + \alpha_U \text{ UNION} + \alpha_N \text{ NONWH} + \alpha_H \text{ HISP} + \beta_1 \text{ ED} + \beta_2 \text{ EX} + \beta_3 \text{ EXSQ} + \epsilon$$

where lnWAGE is the natural log of wages, FE equals 1 if female, 0 otherwise, UNION = 1 if the person is a union member, 0 otherwise, NONWH = 1 if the person is non white, 0 otherwise, HISP = 1 if the individual is of hispanic origin, 0 otherwise, ED is the number of years of education, EX is the number of years of on the job experience, and EXSQ is the number of years of experience squared.

a. Interpret the coefficients on EX and EXSQ (β_2 and β_3).
b. Describe the steps you would take to test the null hypothesis that gender, union status, education, and experience being held FIXED, racial status has no effect on the natural log of wages.

5. (15 points) Use of illegal performance enhancing drugs is a growing problem in the world of sports. During a recent investigation, 100 athletes were suspended from their professional athletic activities on suspicion of using a specific illegal performance enhancing drug that till recently could not be detected using the standard test procedures. A new test has now been devised to detect the use of this particular drug and it is known that the test is 90% reliable when administered to a guilty athlete. In 2% of the cases, however, an athlete will test positive even though they have not taken the illegal drug. Suppose that of the 100 suspended athletes, only 12 were actually using this drug. What is the probability that a given suspended athlete is innocent of using this drug when the test says that they are guilty?

6. (30 points) A recent study by the Brooking's Institute shows that hourly wages earned by college students during summer employment are approximately Normally distributed with a population variance of $6.25. A random sample of 100 college students are surveyed and the sample mean hourly wage is found to be $7.50.

a. Using a 5% significance level, test the null hypothesis H0: $\mu \leq \$7.25$ against the alternative hypothesis, H1: $\mu > \$7.25$

b. Calculate the probability of committing a Type II error for the alternative hypothesis $\mu = \$7.55$.