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:

$$\ln SAL = \beta_0 + \beta_1\ GENDER + \beta_2\ FINANCE + \beta_3\ MIS + \beta_4\ SCIENCE + \beta_5\ SCIENCE * GENDER + \beta_6\ SCIENCE * FINANCE + \beta_7\ SCIENCE * MIS + \beta_8\ SCIENCE * GENDER * FINANCE + \beta_9 \ln GMAT + \epsilon$$

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.

SOLUTION:

female, not finance or mis, no engineering, no gmat: beta0
female, not finance or mis, engineering, no gmat: beta0 + beta4
female, finance, no-engineering, no gmat: beta0 + beta2
female, finance, engineering, no gmat: beta0 + beta2 + beta4 + beta6
female, mis, no-engineering, no gmat: beta0 + beta3
female, mis, engineering, no gmat: beta0 + beta3 + beta4 + beta7

male, not finance or mis, no engineering, no gmat: beta0 + beta1
male, not finance or mis, engineering, no gmat: beta0 + beta1 + beta4 + beta5
male, finance, no engineering, no gmat: beta0 + beta1 + beta2
male, finance, engineering, no gmat: beta0 + beta1 + beta2 + beta4 + beta5 + beta6 + beta8
male, mis, no engineering, no gmat: beta0 + beta1 + beta3
male, mis, engineering, no gmat: beta0 + beta1 + beta3 + beta4 + beta5 + beta7

a. Carefully interpret each of the ten regression coefficients.

beta0: your base ln salary if you didn't write the GMAT and you are female, in a concentration that is not finance or mis with no engineering/science background.

beta1: the marginal increment to ln salary if you are male and didn't write the GMAT and are in a concentration that is not finance or mis with no engineering/science background.

beta2: the base marginal increment to ln salary you get if you are in a finance concentration
beta3: the base marginal increment to ln salary you get if you have a mis concentration
beta4: the base marginal increment in ln salary you get if you have a science/engineering background
beta5: the addition to the base marginal increment in ln salary you get if you have a science background and are male
beta6: the addition to the base marginal increment in ln salary you get if you have an engineering background and have a concentration in finance
beta7: the addition to the base marginal increment in ln salary you get if you have an engineering background and have a concentration in mis
beta8: the addition to the base marginal increment in ln salary you get if you are male with an engineering background and a finance concentration
beta9: the GMAT score elasticity (of starting salary)

b. Indicate the parameter restrictions and the appropriate test statistic 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, undergraduate background (degree) and test scores.

So, given concentration (MIS, FINANCE) and GMAT scores (GMAT), men and women have the same expected starting salary. That would mean that:

female, not finance or mis, no engineering, no gmat: beta0 = male, not finance or mis, no engineering, no gmat: beta0 + beta1

(So beta1 = 0)

AND

female, not finance or mis, engineering, no gmat: beta0 + beta4 = male, not finance or mis, engineering, no gmat: beta0 + beta1 + beta4 + beta5

(So beta5 = 0)

AND

female, finance, no-engineering, no gmat: beta0 + beta2 = male, finance, no engineering, no gmat: beta0 + beta1 + beta2

(Nothing added here)

AND

female, finance, engineering, no gmat: beta0 + beta2 + beta4 + beta6 = male, finance, engineering, no gmat: beta0 + beta1 + beta2 + beta4 + beta5 + beta6 + beta8

(So beta8 = 0)

AND

female, mis, no-engineering, no gmat: beta0 + beta3 = male, mis, no engineering, no gmat: beta0 + beta1 + beta3

(Nothing added here)

AND

female, mis, engineering, no gmat: beta0 + beta3 + beta4 + beta7 = male, mis, engineering, no gmat: beta0 + beta1 + beta3 + beta4 + beta5 + beta7

(Nothing added here)

We need to test the joint hypothesis that:

H0: beta1 = beta5 = beta8 = 0
H1: not H0

We would construct an F-test testing the restricted and unrestricted models.

– 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.

This means that finance/mis gives you no additional bang for your buck given undergrad degree and GMAT score (but allows differences across gender).

For men:

other concentration, engineering: beta0 + beta1 + beta4 + beta5 = finance, engineering = beta0 + beta1 + beta2 + beta4 + beta5 + beta6 + beta8 = mis, engineering = beta0 + beta1 + beta3 + beta4 + beta5 + beta7

So: we need beta2 = 0, beta6 = 0, beta8 = 0 AND beta3 = 0 , beta7 = 0

other concentration, no engineering: beta0 + beta1 = beta0 + beta1 + beta2 = beta0 + beta1 + beta3

no new stuff is added, here.

For women engineers:

beta0 + beta4 = beta0 + beta2 + beta4 + beta6 = beta0 + beta3 + beta4 + beta7

no new stuff is added, here.

Joint hypothesis test of:

H0: beta2 = beta3 = beta6 = beta7 = beta8 = 0
H1: not H0

F-test.

– 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.

Male, science, finance: beta0 + beta1 + beta2 + beta4 + beta5 + beta6 + beta8 = beta0 + beta1 + beta2 so we need beta4 = beta5 = beta6 = 0

Joint hypothesis test of:

H0: beta4 = beta5 = beta6 = 0
H1: not H0.

Joint F-test.

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.

Coefficients: yield rate of return on rbi's, walks, errors on salary. Constant gives base level of ln salary if al other things are zero.

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: $\beta 2 <= 0.015$

One sided hypothesis testing.

t-statistic: (0.0139046 - 0.015)/(0.0036653) = -0.30
t-critical value for 5% in one tail: (can use standard normal table since d.o.f. so large) = -1.645. Since we are NOT further out in the tail, we CANNOT reject the null hypothesis in this case.

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

Critical value with 10% in each tail: 1.28

-0.037 +/- 1.28 (0.0117) = (-0.051976, -0.022024)

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

F-test: (ESS/k) / (RSS/N-k-1) = (194.855/3) / (414.4484)/(368-4) = 64.95/1.139 = 57.02 ~ $F_{3, 364}$
The F-critical value for 3 degrees of freedom in the numerator and more than 120 in the denominator for 95% confidence is 2.60. Since our F-statistic is much larger than this, we are further out in the tail and we can soundly reject the null hypothesis that all of the slope coefficients are simultaneously equal to zero.

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, β_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!)

H_0: $\beta_1 - \beta_3 = 0$
H_1: $\beta_1 - \beta_3 \neq 0$

We can do this by constructing a 95% confidence interval around $\beta_1 - \beta_3$. Now, remember that since we don't observe the true beta's, we can use our OLS estimators, instead. Our estimators are NORMALLY distributed, so the random variable which is their difference is also normally distributed. We know that the mean of this new random variable, has a mean of $\beta_1 - \beta_3$ and a variance of what?

The variance is given by:

Var (β_1 hat - β_3 hat) = Var β_1 hat + Var β_3 hat + 2 Cov (β_1 hat, β_3 hat) = $0.003^2 + 0.0117^2$ + 2 (-0.000011) = 0.000009 + 0.00014 - 0.000022 = 0.00012 (approximately)

So, the two sided confidence interval is: 0.0136 - (-0.03723) +/- 1.96(sqrt (0.00012) = (0.025, 0.076). Since 0 is NOT contained in this interval, we can reject the null hypothesis.

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:

$$\ln WAGE = \alpha + \alpha_F\, FE + \alpha_U\, UNION + \alpha_N\, NONWH + \alpha_H\, HISP + \beta_1\, ED + \beta_2\, EX + \beta_3\, EXSQ + \epsilon$$

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).

if both are positive, we see increasing returns to experience. If beta2 is positive and beta3 is negative we have decreasing returns to experience. We expect that beta2 is positive.

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.

This is the equivalent to testing that $\alpha_N = \alpha_H = 0$. To test this, you would run the above, unrestricted regression and you would also run a restricted regression where you would NOT control for race. Then you would construct the following F-test:

$$F = [(SSR\_r - SSR\_ur)/2]/[SSR\_ur]/(N-k-1) \sim F_{2, N-k-1}$$

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, an athlete will test positive even though the athlete has not taken the illegal drug. Suppose that of the 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 used the drug?

P(positive | used drug) = 0.90
P(positive | did not use drug) = 0.02
P(used drug) = 0.12 ==> P(did not use drug) = 0.88
P(did not use drug| positive) = ?

P(did not use drug|positive) = P(positive|did not use drug)*P(did not use drug)/P(positive)

= 0.02*0.88/[0.90*0.12 + 0.02*0.88] = 0.0176/[0.2356] = 0.0747

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 $\mu = \$7.55$.

N = 100
Xbar = 7.5

By the CLT, Xbar ~ N (μ, 6.25/100)

t statistic = (7.5 - 7.25) / (0.25) = 1 ~ t 99 degrees of freedom, so we can use our standard normal table.

We will REJECT the null hypothesis if t-statistic > t-critical value = 1.645. Since our t-statistic is NOT greater than 1.65, we cannot reject the null hypothesis.

When can we reject the null?

Xbar - 7.25/0.25 >= 1.645, or Xbar > 7.66.

P(xbar < 7.66 | true mean = 7.55) = probability of a type II error

P(xbar < 7.66 | true mean = 7.55) = P (z < (7.66 - 7.55)/0.25) = P(z < 0.44) = 1- P (z > 0.44) = 1- 0.33 = 0.67. 67% probability of committing a type II error if the true mean is 7.55.