Economics 113 Professor Spearot
Introduction to Econometrics
Spring 2008 – Final Exam
Name ________________________

Final Exam – 150 Points

You must answer all the questions. The exam is closed book and closed notes. You may use calculators, but they must not be graphing calculators. Do not use your own scratch paper.

You must show your work to receive full credit

You have plenty of time to finish. Take your time and relax. And, have a safe and wonderful Summer!

Problem 1 (30 Points)

You roll two dice. The first one has THREE sides {1,2,3} and the second one has SIX sides {1,2,3,4,5,6}. Both dice are fair.

a.) Draw and label the Venn diagram describing all possible sample points. **(5 Points)**

b.) What is the probability that you will get a total of four or more points between the two dice? **(5 Points)**

c.) Given that you roll a two with one of the two dice what is the chance that the two dice together will total 4? **(10 Points)**

d.) Given that you roll a three with one of the two dice what is the chance that the two dice together will total a value greater than 4? **(10 Points)**

Problem 2 (90 points)

Suppose that I run the following regression predicting the effects of classroom performance on students' final exam grades:

$$final = \beta_0 + \beta_1 section + \beta_2 mt1 + \beta_3 hwtotal + u$$

Here, *final*, *mt1*, *hwtotal*, *section* are the percent scores on the final, midterm, homework, and section participation, respectively. The results from running this regression are below.

```
. regress final section mt1 hwtotal

      Source |       SS       df       MS              Number of obs =     142
-------------+------------------------------           F(  3,   138) =   26.49
       Model |  12155.6037     3  4051.86791           Prob > F      =  0.0000
    Residual |   21109.933   138  152.970529           R-squared     =  0.3654
-------------+------------------------------           Adj R-squared =  0.3516
       Total |  33265.5367   141  235.925792           Root MSE      =  12.368

------------------------------------------------------------------------------
       final |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     section |   .0795122    .058387     xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
         mt1 |   .4671107   .0669202     xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
     hwtotal |    .235302   .0720338     xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       _cons |   14.46427   7.723413     xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
```

a.) Please interpret the constant. **(5 points)**

b.) I claim that getting a higher grade on homework increases your predicted grade on the final. Conduct a one-sided hypothesis test at the 5% level for the coefficient on *hwtotal*, β_3. Please state your null and alternative hypotheses, and briefly interpret the result. **(10 Points)**

c.) Construct a 99% confidence interval for the coefficient on *section*, β_1. **(10 Points)**

d.) I have reason to suspect that the variability of final exam scores changes with previous performance (homework, midterms, section). What is this called? What can be done about it? What Stata commands are necessary? **(5 Points)**

e.) I want to test the suspicion in ‘d’ rigorously. I run the following regression:

$$\hat{u} = \delta_0 + \delta_1 section + \delta_2 mt1 + \delta_3 hwtotal + \varepsilon$$

Here, $\hat{u}$ is the residual from the regression in ‘a’. The estimates are as follows:

Source	SS	df	MS
Model	7.2760e-12	3	2.4253e-12
Residual	21040.4593	138	152.467096
Total	21040.4593	141	149.223115

Number of obs	142
F(3, 138)	0.00
Prob > F	1.0000
R-squared	0.0000
Adj R-squared	-0.0217
Root MSE	12.348

uhat	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
section	-3.84e-09	.0582909	-0.00	1.000	-.1152588	.1152588
mt1	-1.13e-08	.06681	-0.00	1.000	-.1321036	.1321036
hwtotal	1.12e-08	.0719152	0.00	1.000	-.1421981	.1421982
_cons	1.88e-07	7.710694	0.00	1.000	-15.24638	15.24638

The f-statistic for the full exclusionary test is very low (zero), which implies that the variables of the model tell us very little about the dependent variable. Does the entire procedure outlined above address the assertion in ‘d’? If not, suggest an alternative. What assumption is at play here? **(10 Points)**

f.) I suspect that the return to homework scores is dependent on whether or not you attend sections. To examine this possibility, I run the following regression:

$$final = \beta_0 + \beta_1 section + \beta_2 mt1 + \beta_3 hwtotal + \beta_4 hwtotal * section + u$$

The results from estimating this equation are below:

```
. regress final section mt1 hwtotal hwtotal*section

      Source |       SS       df       MS              Number of obs =     142
-------------+------------------------------           F(  4,   137) =   19.90
       Model |  12225.0771     4  3056.26926           Prob > F      =  0.0000
    Residual |  21040.4597   137  153.579998           R-squared     =  0.3675
-------------+------------------------------           Adj R-squared =  0.3490
       Total |  33265.5367   141  235.925792           Root MSE      =  12.393

------------------------------------------------------------------------------
       final |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     section |   .0842367   .0589235     1.43   0.155    -.0322803    .2007538
         mt1 |   .7592581   .4395156     1.73   0.086    -.1098538     1.62837
     hwtotal |   .4647203   .3486564     1.33   0.185    -.2247237    1.154164
hwtotal*section |  -.0034208   .0050861    xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
       _cons |  -5.333285   30.43569    -0.18   0.861    -65.51776    54.85119
------------------------------------------------------------------------------
```

Derive the return to section attendance. Plug in the estimated coefficients where necessary. Please interpret briefly. **(10 Points)**

g.) What is the homework score which yields a return to section attendance that is equal to zero? Given that homework scores are between 0 and 100, is the return to section attendance always positive? **(10 Points)**

h.) Is there a significant interaction between homework and section attendance? Conduct a two-sided test at the 1% level, stating your null and alternative hypotheses, also briefly interpreting the result. **(10 Points)**

i.) Rather than using interactions as in 'f', I have added in squared terms of homework, *hwtotalsqr*, and section attendance, *sectionsqr*. Their coefficients are β_5 and β_6, respectively.

```
. regress final section mt1 hwtotal hwtotalsqr sectionsqr

      Source |       SS       df       MS              Number of obs =     142
-------------+------------------------------           F(  5,   136) =   15.84
       Model |  12244.9991     5  2448.99983           Prob > F      =  0.0000
    Residual |  21020.5376   136  154.562776           R-squared     =  0.3681
-------------+------------------------------           Adj R-squared =  0.3449
       Total |  33265.5367   141  235.925792           Root MSE      =  12.432

------------------------------------------------------------------------------
       final |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     section |  -.0044826   .2068736    -0.02   0.983    -.4135877    .4046226
         mt1 |   .4710029   .0674799     6.98   0.000     .3375573    .6044485
     hwtotal |   .4722948   .3601857     1.31   0.192    -.2399944    1.184584
  hwtotalsqr |  -.0017249   .0025807    -0.67   0.505    -.0068284    .0033786
  sectionsqr |   .0006319   .0015609     0.40   0.686    -.0024549    .0037186
       _cons |   9.323664    13.5762     0.69   0.493    -17.52409    36.17142
------------------------------------------------------------------------------
```

Which model is preferred, the above model with squared terms, or the original model in 'a'? Please justify your answer. If a test is required, state your null and alternative hypotheses, and test at the 5% level. **(10 Points)**

j.) Suppose that natural ability is an unobserved variable, which does not change over time. I am worried that not including it may be causing omitted variable bias. What technique is appropriate for this problem, and why? **(10 Points)**

Problem 3 (30 Points)

Professor Spearot is getting older. He is worried about a receding hair line. To analyze male hair patterns as a function of demographics, he estimates the following linear probability model using a sample of men:

$$Bald = \beta_0 + \beta_1 \text{Age} + \beta_2 Dad + u$$

Bald takes on the value of 1 if the respondent is bald, and 0 otherwise. *Age* is the Age of the respondent, and *Dad* is an indicator variable taking the value of 1 if the respondent's father is bald and 0 otherwise.

a.) Suppose that β_2 is positive. How do I interpret the estimate of the coefficient on *Dad*, β_2? **(5 Points)**

b.) Suppose that I estimate the model, and I generate predictions for each respondent. Some predictions are negative. Is this sensible? What alternative estimation procedure could remedy this problem? Why? **(10 Points)**

c.) Suppose that Stress, and unobserved variable, increases with age. Stress also leads to a higher likelihood of baldness. What is this called? In what direction is the bias in β_1? **(5 Points)**

d.) Professor Spearot's father is Bald (sorry Dad!). Professor Spearot is 29 years old. Please **derive** the estimating equation required to generate a prediction for somebody with Professor Spearot's characteristics. Please also write the precise STATA commands required to run this regression. **(10 points)**

Extra Credit: (10 Points)

Bob Baden was once a college hockey player (no joke here). Skilled and graceful, he was an offensive weapon.

Suppose that Bob takes three shots at the net. The probability of scoring on the first shot is 0.5. Each time he scores, the probability of scoring on the next shot goes up by 0.1. What is the probability of scoring on the 3rd shot?

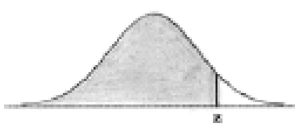
Helpful Formulas

$$\hat{\sigma}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \hat{\mu}_x\right)^2 \qquad \hat{\sigma}_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \hat{\mu}_x\right)\left(y_i - \hat{\mu}_y\right) \qquad \hat{\rho}_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y}$$

$$\hat{\beta}_0 = \hat{\mu}_y - \hat{\beta}_1 \hat{\mu}_x \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(x_i - \hat{\mu}_x\right)\left(y_i - \hat{\mu}_y\right)}{\sum_{i=1}^{n}\left(x_i - \hat{\mu}_x\right)^2}$$

$$R^2 = 1 - \frac{SSR}{SST} \qquad SSR = \sum_{i=1}^{n}\left(\hat{u}_i\right)^2 \qquad SST = \sum_{i=1}^{n}\left(y_i - \hat{\mu}_y\right)^2$$

$$\text{Adj } R^2 = 1 - \frac{\frac{SSR}{n-k-1}}{\frac{SST}{n-1}} \qquad F_{stat} = \frac{\frac{\mathrm{SSR}_{\mathrm{R}} - SSR_{UR}}{q}}{\frac{SSR_{UR}}{n-k-1}}$$



Normal Distribution from -oo to Z

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990