Assignment #2 - POSC 300B

We expect the federal government to provide important services. By some method, we have to pay for these services. Taxation is the most commonly used method of payment. The question of who should bear what portion of the tax burden is an important one. If every family paid the same dollar amount in taxes (e.g., $1,000 in federal taxes per year), this would hurt the poor much more than the wealthy. Obviously, $1,000 is a heavier burden for a family that has an income of $30,000 per year than for a family that has an income of $100,000 per year.

From time to time the U.S. Senate has dealt with the question of who should bear how much of the nation's federal tax burden. Some senators are very willing to ask the more affluent to shoulder a greater share of the nation's tax burden. Conversely, other senators are not very willing to ask the more affluent to pay a greater share of the nation's tax burden. What accounts for senator's differing levels of support for asking the wealthy to shoulder a greater share of the nation's tax burden? This is the main purpose of the statistical analysis of Assignment #2. The dependent variable in our study is the percentage of times the senator voted in favor of shifting the federal tax burden more to high income earners. The independent variables are the degree of the senator's conservatism, the senator's political party affiliation and the median family income in the senator's state. Immediately below you will find the names and descriptions of the variables that appear in the computer printouts over pages 195-198. After studying the variable list below, answer the questions on pages 193-194.

Variable List

Variable Name	Operationalization
Tax	The percentage of times the senator voted in favor of shifting the federal tax burden more to high income earners (i.e., taxpayers earning more than the median family income). Scores range from 0% (indicating no support for shifting the federal tax burden more to high income earners) to a high of 100% (indicating 100% support for shifting the federal tax burden more to high income earners). This is the dependent variable.
Cons	The percentage of times the senator voted in favor of positions taken by the Americans for Constitutional Action. Scores range from 0% (least conservative) to 100% (most conservative). If a senator scored "90" it would mean that 90% of the time that senator voted in a conservative direction. The computer reads a score of 90% as 90, <u>not</u> .90. Higher scores denote greater conservatism. Thus, a senator who scores 90% is twice as conservative as a senator who scores 45%.
Party	Dummy variable measuring the senator's political party affiliation (1=Democrat, 0=Republican).
Stinc	Median family income (in thousands of dollars, hence the computer reads 30.1 which represents $30,100) in the senator's state.

Questions

1. Formulate and defend a hypothesis (see pages 1-3) of how each of the three independent variables in equation 1 on page 195 (the senator's degree of conservatism, the senator's political party affiliation and the state median family income in the senator's state), should be related to the percentage of times the senator voted to shift the tax burden more to high income earners (the dependent variable).

For example, suppose you were a Democratic senator who scored 20 on "cons." This would mean you were relatively "liberal" because you were not very conservative (you scored only 20 on "cons" instead of 70 or 80). If you were this senator and you remained a Democrat who scored 20 on "cons" (thus the score on both political party and conservatism remained the same) but the median family income in your state suddenly increased from $33,000 to $37,000, would you be more or less willing to vote in favor of shifting the tax burden more to high income earners? If the increase in the median family income in your state from $33,000 to $37,000 made you more willing to vote in favor of shifting the tax burden to high income earners (from supporting such changes 70% of the time when state median family income was $33,000 to 80% of the time if state median family income increased to $37,000) then you should hypothesize a positive relationship between state median family income and willingness to shift the tax burden to high income earners (because a higher score on state median family income - $37,000 instead of $33,000 - would be associated with a higher score on willingness to more tax high income earners - 80% instead of 70%). If the increase in the median family income in your state from $33,000 to $37,000 made you less willing to vote in favor of shifting the tax burden more to high income earners then you should hypothesize a negative relationship between state median family income and willingness to shift the tax burden more to high income earners (because a higher score on state median family income $37,000 instead of $33,000 - would be associated with a lower score on willingness to tax high income earners). <u>You need a reason for your hypothesis. Why did you hypothesize either a positive or a negative relationship? The middle of the first paragraph on page 64 shows how to formulate and defend a hypothesis (pages 107-108 are useful on understanding multiple independent variables)</u>.

You need to formulate and defend three hypotheses: (1) the relationship between a senator's conservatism and their degree of support for shifting the tax burden more to high income earners; (2) the relationship between the senator's political party affiliation and the senator's degree of support for shifting the tax burden more to high income earners; and (3) for the relationship between the median family income in the senator's state and the senator's degree of support for shifting the tax burden more to high income earners. <u>For conservatism, think about what a conservative philosophy means, not which income groups support conservative candidates. Do not defend one independent variable on the basis of another independent variable. Thus, you can not defend "party" by saying that Republicans are more conservative than Democrats. You already have conservatism (CONS) in the model (see pages 107-108)</u>. <u>Rather, think about which party the poor more vote for. Do not use the computer results from pages 195-196 in formulating the hypotheses (hypothesizing precedes data analysis). Refer to a variable by its operationalization, not its variable name (do not write "TAX"). Your answers must be your own (not part of a "group" answer). Group answers get zero points</u>!!

2. Assess the statistical significance of the "T-STAT^s" (i.e., "t ratios"), except for "C" (the Y intercept), in column 4 of equation 1 on page 195. Review the middle of page 90.

3. Assess the degree of multicollinearity (on multicollinearity see pages 115-123) in equation 1 on page 195. Use the "explained variance test" for multicollinearity on pages 120-121. Remember that multicollinearity is only a potential problem for statistically insignificant independent variables in equation 1 (i.e., the "main equation"). Thus, if an independent variable is statistically significant in equation 1, then you should say that multicollinearity cannot be a problem for that independent variable. This is the only question where you might use the R^2 (on R^2 see pages 109-110), or any other information, in equations 2-4 on pages 195-196.

4. Why is regression preferable to cross tabulation (pages 28-30) and measures of association (pages 31-34) for this assignment? Use reasons 1-4 on pages 35-37. For example, since "CONS" has 101 categories of responses (0 plus 1-100), what problem would we have in using "CONS" in a cross tabulation table (see reason #1 on page 35 - reason #1 only applies to cross tabulation, not measures of association)? Use an example from this assignment for each of the four reasons. Define terms such as "magnitude."

5. Interpret the R^2 (see pp. 109-110) for equation 1 on page 195.

6. Interpret both the unstandardized and standardized coefficients (i.e., the bs) for equation 1 on page 195. The unstandardized coefficients are found under the coefficient column in equation 1 (e.g., -.644). You don't need to interpret "C" (the y intercept, which is 67.38). The unstandardized coefficients are what we have worked with the bulk of the semester (on interpreting unstandardized "b" see pages 107-110). Use the operationalization of each variable. Thus, do not write, "If party increases by one unit ..." What is a "unit" of political party? Look at the operationalization (page 192). Similarly, do not write that support for shifting the tax burden more to high income earners decreases by .644? What is ".644"? Does this mean dollars, percentage points, what? Be specific. Also, do not write that the score on a variable "changed" by some amount. Did it increase or decrease by what specific amount?

To calculate the standardized coefficients you need the unstandardized coefficients (which you just used) and the standard deviation for each of the variables. The standard deviations are as follows: cons 31.24; party .49; stinc 1.52; tax 28.73. Compare the relative importance of conservatism to party affiliation; conservatism to state median family income; and party affiliation to state median family income. Page 128 shows how to compute and interpret standardized coefficients. Show your calculations (i.e., your work) in the paper you submit.

7. Using equation 2 on page 111 and the results from equation 1 on page 195 (i.e., a, b_1, etc.), what percentage of the time would a senator whose conservatism score is 90%, who is a Republican and whose state has a median family income of $12,500 be predicted to vote in favor of the poor on tax legislation (i.e., TAX - see the Variable List on page 192 for the measurement of each of the aforementioned variables)? The example on page 111 is useful. Show your work.

Equation 1 or "Main Equation"

Source	SS	df	MS
Model	54886.5757	3	18295.5252
Residual	26840.2643	96	279.586087
Total	81726.84	99	825.523636

Number of obs	= 100
F(3, 96)	= 65.44
Prob > F	= 0.0000
R-squared	= 0.6716
Adj R-squared	= 0.6613
Root MSE	= 16.721

tax	Coef.	Std. Err.	t	P>|t|	Beta
cons	-.6447205	.0756005	-8.53	0.000	-.7010575
party	11.20792	4.675335	2.40	0.018	.1902963
stinc	-.5600809	1.283164	-0.44	0.663	-.0297112
_cons	67.38277	15.11393	4.46	0.000	.

Note: _cons is called "a," the "y intercept" or the "constant."

Further Note: Beta column contains standardized coefficients

Source	SS	df	MS
Model	54886.5757	3	18295.5252
Residual	26840.2643	96	279.586087
Total	81726.84	99	825.523636

Number of obs	= 100
F(3, 96)	= 65.44
Prob > F	= 0.0000
R-squared	= 0.6716
Adj R-squared	= 0.6613
Root MSE	= 16.721

tax	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
cons	-.6447205	.0756005	-8.53	0.000	-.7947863	-.4946547
party	11.20792	4.675335	2.40	0.018	1.927454	20.48839
stinc	-.5600809	1.283164	-0.44	0.663	-3.10714	1.986979
_cons	67.38277	15.11393	4.46	0.000	37.38186	97.38368

Note: For insignificant independent variables the sign of the coefficient estimates in the 95% confidence interval changes (e.g., stinc)

Equation 2

Source	SS	df	MS
Model	47716.06	2	23858.03
Residual	48917.73	97	504.306494
Total	96633.79	99	976.098889

Number of obs	= 100
F(2, 97)	= 47.31
Prob > F	= 0.0000
R-squared	= 0.4938
Adj R-squared	= 0.4833
Root MSE	= 22.457

cons	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
party	-41.13872	4.688349	-8.77	0.000	-50.44379	-31.83364
stinc	-8.346615	1.500567	-5.56	0.000	-11.32483	-5.368405
_cons	137.4466	14.74029	9.32	0.000	108.1912	166.702

Variable Means, Standard Deviations, Minimums and Maximums

Variable	Obs	Mean	Std. Dev.	Min	Max
tax	100	46.54	28.73193	7	97
cons	100	35.11	31.24258	0	100
party	100	.62	.4878317	0	1
stinc	100	9.205	1.524174	6.1	12.4

Equation 3

Logit Estimate

Number of obs	= 100
LR chi2(2)	= 53.90
Prob > chi2	= 0.0000
Pseudo R2	= 0.4059

Log likelihood = -39.455085

party	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
cons	-.0678678	.0129977	-5.22	0.000	-.0933428	-.0423928
stinc	-.8107697	.2216693	-3.66	0.000	-1.245233	-.376306
_cons	10.55072	2.409466	4.38	0.000	5.828255	15.27319

Regression Estimate

Source	SS	df	MS
Model	10.7694261	2	5.38471307
Residual	12.7905739	97	.131861586
Total	23.56	99	.237979798

Number of obs	= 100
F(2, 97)	= 40.84
Prob > F	= 0.0000
R-squared	= 0.4571
Adj R-squared	= 0.4459
Root MSE	= .36313

party	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
cons	-.0107566	.0012259	-8.77	0.000	-.0131896	-.0083236
stinc	-.1186508	.0251279	-4.72	0.000	-.1685228	-.0687789
_cons	2.089845	.2504194	8.35	0.000	1.592831	2.586858

Equation 4

Source	SS	df	MS
Model	60.1820864	2	30.0910432
Residual	169.805407	97	1.7505712
Total	229.987493	99	2.32310599

Number of obs	= 100
F(2, 97)	= 17.19
Prob > F	= 0.0000
R-squared	= 0.2617
Adj R-squared	= 0.2465
Root MSE	= 1.3231

stinc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
cons	-.0289731	.0052088	-5.56	0.000	-.0393112	-.018635
party	-1.575188	.3335935	-4.72	0.000	-2.237278	-.9130969
_cons	11.19886	.3705936	30.22	0.000	10.46334	11.93439

Multicollinearity [1-(1/VIF column) = R-squared in Eq. 2 - 4]

Variable	VIF	1/VIF
cons	1.98	0.506218
party	1.84	0.542894
stinc	1.35	0.738325
Mean VIF	1.72	

(VIF stands for "variance inflation" - which in this context means increasing the value of the standard error of a regression coefficient - increased variance inflation increases the value of the standard error of the regression coefficient which increases the value of the denominator of the t ratio which decreases the value of the t ratio and reduces the degree of statistical significance of the regression coefficient)

I will discuss this page in class.

obs	CONS	PARTY	STINC	TAX
1	26.00000	1.000000	7.400000	54.00000
2	17.00000	1.000000	7.400000	63.00000
3	43.00000	1.000000	7.400000	41.00000
4	14.00000	1.000000	7.400000	25.00000
5	16.00000	1.000000	10.30000	65.00000
6	8.000000	1.000000	7.500000	97.00000
7	8.000000	1.000000	7.800000	64.00000
8	0.000000	1.000000	9.700000	63.00000
9	7.000000	1.000000	9.700000	53.00000
10	36.00000	1.000000	7.800000	33.00000
11	7.000000	1.000000	11.40000	52.00000
12	12.00000	1.000000	9.700000	87.00000
13	14.00000	1.000000	12.40000	85.00000
14	0.000000	1.000000	9.700000	77.00000
15	4.000000	1.000000	11.80000	55.00000
16	7.000000	1.000000	9.600000	91.00000
17	11.00000	1.000000	9.600000	84.00000
18	13.00000	1.000000	10.70000	47.00000
19	11.00000	1.000000	10.70000	65.00000
20	41.00000	1.000000	10.70000	48.00000
21	9.000000	1.000000	8.500000	84.00000
22	15.00000	1.000000	8.500000	90.00000
23	16.00000	1.000000	8.900000	57.00000
24	4.000000	1.000000	8.900000	82.00000
25	23.00000	1.000000	12.40000	24.00000
26	4.000000	1.000000	9.900000	84.00000
27	0.000000	1.000000	9.900000	77.00000
28	4.000000	1.000000	11.00000	94.00000
29	0.000000	1.000000	10.80000	92.00000
30	0.000000	1.000000	8.200000	79.00000
31	7.000000	1.000000	8.200000	91.00000
32	0.000000	1.000000	9.000000	96.00000
33	4.000000	1.000000	9.000000	97.00000
34	5.000000	1.000000	10.00000	79.00000
35	0.000000	1.000000	10.00000	72.00000
36	0.000000	1.000000	11.00000	82.00000
37	9.000000	1.000000	8.400000	65.00000
38	8.000000	1.000000	11.60000	25.00000
39	0.000000	1.000000	7.500000	92.00000
40	29.00000	1.000000	10.10000	92.00000
41	30.00000	1.000000	8.900000	24.00000
42	7.000000	1.000000	10.10000	89.00000
43	8.000000	1.000000	10.40000	61.00000
44	12.00000	1.000000	10.40000	44.00000
45	0.000000	1.000000	8.900000	83.00000
46	8.000000	1.000000	9.300000	67.00000
47	33.00000	1.000000	7.600000	73.00000
48	47.00000	1.000000	7.800000	30.00000
49	52.00000	1.000000	6.300000	24.00000
50	18.00000	1.000000	6.300000	75.00000

I Will Discuss This Page In Class

It is extremely useful to workout the prediction for senator #1 and see how this relates to the printout of the residuals. As we can see from observation #1 on page 197, senator #1 scored as follows on the three independent variables: conservatism "26" (i.e., the senator voted in the direction favored by the Americans for Constitutional Action 26% of the time), party "1" (i.e., the senator is a Democrat) and state median family income "7.4" (i.e., the median family income in the senator's state is $7,400). Additionally, we know from the results of equation 1 on page 195 that the values of the coefficients (i.e., the unstandardized "b^s") are as follows: cons -.644; party 11.207; and stinc -.560. Furthermore, we know that the value of the Y intercept (i.e., "a" which is listed as "c" on the printout) in equation 1 on page 195 is 67.382. Therefore, the predicted value on Y (i.e., our model's prediction of the percentage of times the senator voted for to shift the tax burden more to high income earners for senator #1 is as follows:

$$[67.382] + [(-.644)(26)] + [(11.207)(1)] + [(-.560)(7.4)] =$$

$$[67.382] + [-16.744] + [11.207] + [-4.144] =$$

$$67.382 - 16.744 + 11.207 - 4.144 = 57.701$$

57.701 is within rounding of the predicted value of Y (i.e., tax) for senator #1 ("fitted" value of 57.683 - see observation #1 on page 199). The residual ("e") for senator #1 (-3.68336 - see observation #1 on page 199) is simply the actual score (54 - see observation #1 on page 199) minus the predicted score (57.683).

If we were trying to predict the degree to which a Republican senator who scored 80 on "CONS" and whose state had a median family income of $7,400 dollars (i.e., only the scores on "PARTY" and "CONS" would be different than the senator above) the prediction would be as follows:

$$[67.382] + [(-.644)(80)] + [(11.207)(0)] + [(-.560)(7.4)] =$$

$$[67.382] + [-51.52] + [0] + [-4.144] =$$

$$67.382 - 51.52 + 0 - 4.144 = 11.718$$

The senator's party and ideology have quite an impact on how likely the senator is to support shifting the tax burden more to high income groups!!!

I Will Discuss This Page In Class


```
                                                     "e"        Y         Ŷ
===============================================================================
              Residual Plot                     obs RESIDUAL   ACTUAL    FITTED
===============================================================================
|             :    *|     :           |           1 -3.68336  54.0000  57.6834
|             :     *     :           |           2 -0.48584  63.0000  63.4858
|             :   * |     :           |           3 -5.72311  41.0000  46.7231
|    *        :     |     :           |           4 -40.4200  25.0000  65.4200
|             :     |*    :           |           5  2.49367  65.0000  62.5063
|             :     |     :    *      |           6  27.7677  97.0000  69.2323
|             :   * |     :           |           7 -5.06430  64.0000  69.0643
|             : *   |     :           |           8 -10.1579  63.0000  73.1579
|             *     |     :           |           9 -15.6449  53.0000  68.6449
|            *:     |     :           |          10 -18.0121  33.0000  51.0121
|             *     |     :           |          11 -15.6927  52.0000  67.6927
|             :     |     : *         |          12  21.5787  87.0000  65.4213
|             :     |     :  *        |          13  22.3804  85.0000  62.6196
|             :     |*    :           |          14  3.84209  77.0000  73.1579
|             :*    |     :           |          15 -14.4029  55.0000  69.4029
|             :     |     :  *        |          16  22.2991  91.0000  68.7009
|             :     |     :*          |          17  17.8780  84.0000  66.1220
|            *:     |     :           |          18 -17.2165  47.0000  64.2165
|             :     *     :           |          19 -0.50590  65.0000  65.5059
|             :     |*    :           |          20  1.83571  48.0000  46.1643
|             :     |     *           |          21  15.9725  84.0000  68.0275
|             :     |     :   *       |          22  25.8408  90.0000  64.1592
|             :   * |     :           |          23 -6.29044  57.0000  63.2904
|             :     |   * :           |          24  10.9729  82.0000  71.0271
|      *      :     |     :           |          25 -32.8171  24.0000  56.8171
|             :     |    *:           |          26  13.5330  84.0000  70.4670
|             :     | *   :           |          27  3.95411  77.0000  73.0459
|             :     |     :  *        |          28  24.1491  94.0000  69.8509
|             :     |     :*          |          29  19.4582  92.0000  72.5418
|             :     | *   :           |          30  5.00197  79.0000  73.9980
|             :     |     : *         |          31  21.5150  91.0000  69.4850
|             :     |     :  *        |          32  22.4500  96.0000  73.5500
|             :     |     :   *       |          33  26.0289  97.0000  70.9711
|             :     |   * :           |          34  9.23372  79.0000  69.7663
|             :     *     :           |          35 -0.98988  72.0000  72.9899
|             :     |   * :           |          36  9.57020  82.0000  72.4298
|             :    *|     :           |          37 -3.08353  65.0000  68.0835
|   *         :     |     :           |          38 -41.9360  25.0000  66.9360
|             :     |     :*          |          39  17.6099  92.0000  74.3901
|             :     |     :       *   |          40  37.7630  92.0000  54.2370
|       *     :     |     :           |          41 -30.2644  24.0000  54.2644
|             :     |     : *         |          42  20.5792  89.0000  68.4208
|             :  *  |     :           |          43 -6.60809  61.0000  67.6081
|           * :     |     :           |          44 -21.0292  44.0000  65.0292
|             :     |   * :           |          45  9.39403  83.0000  73.6060
|             :     *     :           |          46 -1.22418  67.0000  68.2242
|             :     |     : *         |          47  19.9417  73.0000  53.0583
|             :*    |     :           |          48 -13.9202  30.0000  43.9202
|            *:     |     :           |          49 -17.5367  24.0000  41.5367
|             :     |   * :           |          50  11.5428  75.0000  63.4572
|             :*    |     :           |          51 -12.1357  25.0000  37.1357
|             :  *  |     :           |          52 -6.97792  25.0000  31.9779
|     *       :     |     :           |          53 -35.7394  22.0000  57.7394
|             :    *|     :           |          54 -1.79278  25.0000  26.7928
|          *  :     |     :           |          55 -22.4435  21.0000  43.4435
|          *  :     |     :           |          56 -23.7329  21.0000  44.7329
|             : *   |     :           |          57 -11.0912  21.0000  32.0912
```

Assignment #3 - POSC 300-B

One of the most important concepts in international relations is deterrence. A policy of deterrence seeks to convince an adversary by the threat of military retaliation that the costs of resorting to the use of military force to achieve foreign policy objectives will outweigh the benefits. The concept of deterrence has also been extended to include situations where a "defender" will protect a "protege" from an "attacker." Assignment #3 deals with the factors (i.e., independent variables) which likely increase, or decrease, the probability that extended deterrence will be "successful" (i.e., that the "defender" successfully protects the "protege" from being attacked by the potential "attacker"). The data consists of 58 cases of extended deterrence over the 1885-1983 era (see the list of cases on page 216).

This assignment also introduces you to the opportunities for model building in political science. Model building in political science is a challenging interplay between theory and data analysis. As a first approach to the assignment, look at the variable list on pages 213-214. Think about which variables are likely to effect the probability that extended deterrence will be successful and the direction of the impact. Thus, after looking at the coding scheme for that particular variable, if the score on the variable in question were to increase (e.g., change from zero to one) would this mean that extended deterrence would be more likely to be "successful" (if so, a "positive" relationship) or less likely to be "successful" (if so, a "negative" relationship)?

Always remember that a model is a simplification of reality. We are trying to examine the effect of the most important independent variables. We are not trying to include every factor that might have some infinitesimally small effect on the dependent variable. Thus, we are striving for a parsimonious model which highlights those factors international relations theory suggests should be most important (e.g., the balance of military power).

While including a great number of independent variables can impede the development of a sound theoretically driven model, it also has important statistical weaknesses. The more insignificant independent variables you include, the more likely you will find a statistically significant relationship when one does not actually exist (i.e., commit a "type I error"). Let me explain. Let us say that in the population (see page 40 on population) an independent variable has no effect on the dependent variable. If you estimate the relationship between the two variables in 20 samples, the relationship would (on average) show up as statistically significant at the .05 level in 1 of the 20 samples. Remember that the .05 level means that 5% of the time (i.e., 1 in 20) you will reject the null hypothesis when the null hypothesis is actually true. If you happened to run your analysis on this one unrepresentative sample, you would have incorrectly concluded that the independent variable had a statistically significant effect on the dependent variable when it really does not. The result was merely the fact that your sample was not representative of the population. For the same reason, if you include 20 insignificant independent variables in an equation one of them should show up as statistically significant when none of them actually are. Each independent variable you add increases the chances that you will incorrectly conclude that an independent variable is statistically significant. Additionally, since there is some chance that an insignificant independent variable will be highly related to a theoretically important independent variable, each additional insignificant independent variable increases the chance of severe multicollinearity. This could lead us to the conclusion that an important independent variable is insignificant when it is actually significant (hence a "type II" error). Therefore, there are good statistical reasons for not including every possible independent variable. Let theory, not data mining, be your guide.

If you are taking a daytime section of 300B, the day this assignment is due read (check the syllabus for the exact date) you will need to submit a list of from three to five independent variables from those on pages 213-214 that you think best explain why deterrence (i.e., the dependent variable "outcom" on page 214) is successful in some instances and not in others. If you are taking 300B at night, you will not turn in a list of independent variables, I will simply hand you a printout. Night courses have too few meetings for me to process your requests. Remember, your work should be done independently. Hence, it should not be the result of a group project. Later (check the syllabus for the date) you will need to submit answers to the following questions:

1. State and defend a hypothesis for each independent variable in the "main equation" in the computer results I give you in class. The dependent variable in your study is "outcom." "Outcom" has two possibilities: either deterrence succeeds (coded "1") or fails (a war results, coded "0"). If there are five independent variables in the computer results for your "main equation," you need to write and defend five hypotheses (do not write a hypothesis for the "y intercept" - i.e., "c" in your computer results). Answer this question with the independent variables in the computer results I give you in class (not the ones you might choose).

You need a "direction" (positive or negative) for the impact of each independent variable on the dependent variable. Look at the coding scheme for the dependent variable (outcom) and each of the independent variables in your "main equation" (see pages 213-214). If the score on one of your independent variables increased (e.g., a score on a dummy variable in your "main equation" suddenly increased from "0" to "1"), should this make deterrence more or less likely to be "successful"? Why?

If the independent variable has two categories of responses (i.e., is a dummy variable), use each category in your hypothesis. For example, if one of your independent variables was whether, or not, the potential attacker uses a democratic form of government, then the hypothesis might be phrased as follows: if the potential attacker uses a democratic form of government deterrence is more likely to be successful than if the potential attacker does not use a democratic form of government. If the independent variable is a percentage the following form would be appropriate: the greater the percentage of the attacker's population who are in the military, the less likely deterrence will be successful (i.e., a "negative" relationship). If the independent variable is a ratio then you need to think through what is in the numerator and what is in the denominator and what an "increase" in the ratio would mean. For example, if I was using a ratio of defender oil reserves to attacker oil reserves the hypothesis might be phrased as follows: the greater the balance of oil reserves favors the defender, the more likely deterrence will be successful. The independent variables used in the previous three examples are just hypothetical (i.e., they are not in the data set you will use). I am just trying to help you think through and phrase a hypothesis.

Do not look at your computer results when generating a hypothesis. A hypothesis is what you expect to happen. Not necessarily what does happen. Be sure you defend the direction of each hypothesis. For example, in the hypothesis in the previous paragraph, why should deterrence be more likely to be successful if the attacker is a democracy rather than a non-democracy? If you do not defend the direction of each hypothesis with a pattern of reasoning you will lose a substantial number of points.

2. Why use logit instead of regression? Re-read pages 150-152 and the discussion of the dependent variable "outcom" in the upcoming variable list (pages 213-214).

3. Interpret the "t ratio" or "T-STAT." (see the middle of page 90 and pages 109-110) for each of the independent variables (omit "C" - the "y intercept" - it is not a variable) in your "main equation" (i.e., the equation which includes all the independent variables). While the "sign" does not effect statistical significance it certainly does effect the degree of support for your hypothesis. Thus, if you hypothesized a "negative" relationship and your results show a "positive" relationship that is statistically significant, this is strong evidence against the hypothesis.
I have an example of answering question 3 on page 208.

4. Assess the degree of multicollinearity among your independent variables. Review pages 115-123 on multicollinearity. Remember that multicollinearity is only a potential problem for statistically insignificant independent variables. So, if some of your independent variables are statistically significant, your first sentence in answering question #4 should mention that multicollinearity cannot be a problem for the statistically significant independent variables. For example, if you use independent variable "maipd," and it is statistically significant, then you should say that since "maipd" is statistically significant, multicollinearity cannot be a problem for it. Then address whether or not (and why) multicollinearity is, or is not, a problem for each of the statistically insignificant independent variables. Use the "deletion method" to test for multicollinearity (pages 121-122). As you know from the readings (pages 121-122), the "deletion method" of testing for multicollinearity requires you to re-estimate the "main equation" (i.e., the equation that has "outcom" as the dependent variable and includes the greatest number of independent variables) omitting one different independent variable each time until each independent variable has been "deleted" once. I have an example of answering question 4 on page 208.

5. Interpret the standardized coefficients in your "main equation" and perform a magnitude assessment (the magnitude assessment will be explained ahead). The calculation of standardized coefficients is shown on pages 127-129 (especially page 208) and 154-155. As you will need the standard deviations of the variables involved, they are provided on page 215. To interpret standardized coefficients follow the discussion on pages 127-129. As your variables and results will differ from those used in the example on pages 127-129, you will need to substitute the values you obtain for those I used. I have an example of calculating and interpreting standardized coefficients using the data set for Assignment #3 on page 209.

The second part of question #5 is to perform a "magnitude assessment." As discussed on pages 155-156, since logit posits a non-linear relationship between each independent variable and the dependent variable, you cannot interpret unstandardized logit coefficients the same way you interpreted unstandardized linear regression coefficients (i.e., "b") on pages 109-110. Thus, you cannot say if one particular independent variable increases by one unit, the dependent variable increases, or decreases, by some particular number of units. However, as pages 155-158 suggest, you can obtain a good idea of how much a change in one of the independent variables effects the dependent variable by performing a "magnitude assessment."

As you will remember from pages 155-158, I chose two variables (the senators liberalism and political party affiliation) and changed the values on these variables (while holding the level of the other independent variables constant - i.e., the same) and then calculated how the probability of a senator voting "yes" on the "Levin Amendment" changed. Since your topic is different than that discussed on pages 150-164, it makes sense to change the procedure slightly. Instead of changing the level of two independent variables (as I did on pages 155-158), I want you to change the level of just one independent variable. Thus, I want you to change the level of one independent variable and see how this changes the probability that deterrence will be "successful."

First, look at the results for your "main equation." Choose the independent variable ("c" the y intercept, or constant, is not an independent variable) with the highest "T-STAT." score (regardless of whether it is positive or negative) and designate it as the "comparison" variable. Now, set the level of each independent variable other than the "comparison" variable at one of the following levels: (1) for dummy variables (i.e., the variable can only attain a value of "1" or "0" - see the "variable list" on pages 213-214 - for the coding scheme for each variable) set them at "1"; (2) for non-dummy variables (i.e., variables that can attain values other than just "1" or "0") set them at their "mean" (i.e., average) level (the means and standard deviations for all variables are shown on page 215). If your "comparison" variable is a dummy variable: set it at "1" and calculate the probability that deterrence will be "successful" (see the example on pages 210-211 ahead). Then set your "comparison" variable at "0" and calculate the probability that deterrence will be "successful." How much did the probability change? If your "comparison" variable is not a dummy variable: set it at its "mean" (i.e., average) value (again, the "mean" value of each variable appears on page 215) and calculate the probability that deterrence will be "successful" (again, see pages 210-211 ahead). Then, set the "comparison" variable one standard deviation higher than the mean value (e.g., if the mean were 3.0 and the standard deviation were 1.0, instead of setting the variable at 3.0 you would now set it at 4.0 - the standard deviation for all variables is shown on page 215) and calculate the probability that deterrence will be "successful." Again, how much does the probability change? Pages 210-211 contain an example of answering this portion of question 5.

6. In Assignment #2 I asked you to interpret the R^2 in the "main" equation (which was equation #1 in Assignment #2). As you will see in your printout, R^2 is not reported for logit. While the reasons are highly technical, the crux of it is that a dependent variable with only two categories of responses is not likely to yield an R^2 of close to 1.0. For this reason, you should use the likelihood ratio index as the measure of model performance for logit. I have an example of calculating and interpreting the likelihood ratio index for Assignment #3 on page 212. Pages 158-159 show you how to calculate and interpret the likelihood ratio index. Obviously, your variable names and results will be different from those on pages 158-159. However, if you just follow the example you should have little trouble. In order to make this simpler, I will tell you that the "log likelihood" value for the equation where "outcom"

is predicted by just a "constant" (thus the equation is: logit outcom c) is -39.336. In order to calculate the "likelihood ratio index" you only need the "log likelihood" value I just mentioned (-39.336) and the "log likelihood" value from your "main" equation. This will be clear when you re-read pages 158-159. I have an example of calculating the likelihood ratio index on page 212.

7. Since you have the order of observations (i.e., the order of the 58 disputes - see page 216), you can undertake an interesting analysis of the residuals. Remember that a residual is the difference between what score the dependent variable actually attained and what score our model predicted it would attain. Since the dependent variable is a dummy variable (the only possible scores are "1" or "0"), the highest absolute value (i.e., disregarding positive or negative sign) a residual could attain would be 1.0 (i.e., the score on the dependent variable was "1" but our model predicted it would be "0," or vice versa). The larger the absolute value of the residual, the less accurate our prediction. Thus, if our model predicts a score on the dependent variable of .9 and the actual score is 1.0, this leaves a small residual (1 -.9 = .1). Alternatively, if our model predicts a score on the dependent variable of .15 and the actual score is 1.0, this results in a much larger residual (1 -.15 = .85). Negative residuals occur when deterrence fails (i.e., the score on "outcom" is "0"). For example, if our model predicts that the score on "outcom" will be .85 (i.e., that there is an 85% chance that deterrence will be successful) and deterrence fails, the residual would be -.85 (because 0 -.85 = -.85).

My question is, can you discover a pattern in the residuals? If so, what does this pattern tell us? Take the five residuals with the largest absolute values (i.e., regardless of whether they are positive or negative) and try to think of what these observations might have in common. For example, are the five largest residuals disproportionately from the earlier, or later, years? If so, what independent variable should you have included to capture this? Perhaps the largest residuals tend to occur when a non-superpower is the "defender." Perhaps the largest residuals typically occur in the same region of the world. Try these, and other possible factors by which to analyze the absolute size of the residuals. As the residuals vary according to the independent variables you select, everyone will have different residuals. See if you can think of an independent variable that is suggested by the residuals but is not in the data set. You could describe what concept this unavailable variable should represent, and better yet, suggest how this variable could be measured.

Page 216 shows the order of the observations, which nations were involved and in what role (i.e., "defender," "attacker," or "protege") in each case. For example, you can see that observation #1 occurred in 1885, the potential "attacker" was Russia, the "protege" (the nation being defended) was Afghanistan, Britain and India were the "defenders" and that deterrence was "successful" (i.e., Britain and India successfully protected Afghanistan from being attached by Russia in 1885). If you can not find a pattern in the residuals tell me what patterns you looked for and what occurrences made you think the pattern you looked for was not the same as the results showed. What I value most is good thinking. Think of a possible pattern that I did not mention above.

8. What do your results suggest concerning international relations theory? The remaining part of this page, as well as your answers to questions 3-7 will be of value in answering question 8. What do your results tell us about why deterrence sometimes succeeds and sometimes fails? If none of your independent variables are statistically significant, at least we should have an idea of what does not effect the probability that deterrence will be successful! Be sure to staple the computer printout you have been using for this assignment to the paper you submit.

Hypothetical Model and Interpretation

The day this assignment is due read (check the syllabus for the exact date) you need to submit a list of from three to five independent variables from those contained over pages 213-214 that you think explain why deterrence is either successful or unsuccessful (i.e., the dependent variable, "outcom" - see page 214). In order to help you perform some of the computations necessary for Assignment #3, I will discuss the results for a possible model. I do not want you to use the model I show ahead. Frankly, I think it is a terrible model. I only use it to show you how to interpret your results. The first (or "top") equation on page 206 is what I have previously referred to as the "main equation." Pages 206-207 show the estimate of my "main equation" and four equations that test for multicollinearity ("deletion equations" 1-4). Just keep reading!! This will all become clear over the next several pages.

To analyze the results for the "hypothetical model" I would first look at the results for the "main equation" (top of page 206). Looking at the "T-STAT." column for the results for the "main equation," I see that none of the four independent variables (i.e., "maipd," "min," "vppa" and "milall") have a "T-STAT." with an absolute (disregard positive or negative sign) of over 2.0. Therefore, none of my hypotheses are supported. Let me also mention that even if one of the "T-STAT^s" was over 2.0, this does not necessary mean my hypothesis is supported. Although I did not use the variable "dd," let me show you how this might have applied to that variable. From page 213, you know that "dd" stands for diplomatic defeat of the defender in the most recent confrontation

Main Equation

LOGIT // Dependent Variable is OUTCOM
Date: 4-14-1996 / Time: 8:42
SMPL range: 1 - 58
Number of observations: 58
Convergence achieved after 3 iterations

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	-0.0752274	0.6083717	-0.1236536	0.9021
MAIPD	0.1536330	0.0939154	1.6358657	0.1078
MIN	0.0591765	0.6642536	0.0890871	0.9293
VPPA	1.061E-05	2.028E-05	0.5230687	0.6031
MILALL	-1.1309406	0.7208192	-1.5689658	0.1226

Log likelihood	-37.430472
Cases with OUTCOM = 1	34
Cases with OUTCOM = 0	24

Deletion Equation #1 (omitting "maipd")

LOGIT // Dependent Variable is OUTCOM
Date: 4-14-1996 / Time: 8:43
SMPL range: 1 - 58
Number of observations: 58
Convergence achieved after 2 iterations

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	0.6499811	0.4315520	1.5061480	0.1379
MIN	-0.1971676	0.6291111	-0.3134066	0.7552
VPPA	-2.580E-06	1.780E-05	-0.1449004	0.8853
MILALL	-0.4871229	0.5705723	-0.8537443	0.3970

Log likelihood	-38.862568
Cases with OUTCOM = 1	34
Cases with OUTCOM = 0	24

Deletion Equation #2 (omitting "min")

LOGIT // Dependent Variable is OUTCOM
Date: 4-14-1996 / Time: 8:44
SMPL range: 1 - 58
Number of observations: 58
Convergence achieved after 3 iterations

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	-0.0519035	0.5495637	-0.0944450	0.9251
MAIPD	0.1518555	0.0918067	1.6540788	0.1039
VPPA	1.090E-05	2.003E-05	0.5440843	0.5886
MILALL	-1.1344677	0.7199729	-1.5757090	0.1209

Log likelihood	-37.434446	
Cases with OUTCOM = 1		34
Cases with OUTCOM = 0		24

Deletion Equation #3 (omitting "vppa")

LOGIT // Dependent Variable is OUTCOM
Date: 4-14-1996 / Time: 8:45
SMPL range: 1 - 58
Number of observations: 58
Convergence achieved after 2 iterations

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	0.0640506	0.5470437	0.1170850	0.9072
MAIPD	0.1343751	0.0856385	1.5690974	0.1225
MIN	0.1153033	0.6567081	0.1755777	0.8613
MILALL	-0.9679159	0.6441863	-1.5025403	0.1388

Log likelihood	-37.569969	
Cases with OUTCOM = 1		34
Cases with OUTCOM = 0		24

Deletion Equation #4 (omitting "milall")

LOGIT // Dependent Variable is OUTCOM
Date: 4-14-1996 / Time: 8:46
SMPL range: 1 - 58
Number of observations: 58
Convergence achieved after 2 iterations

VARIABLE	COEFFICIENT	STD. ERROR	T-STAT.	2-TAIL SIG.
C	-0.0081591	0.5942361	-0.0137304	0.9891
MAIPD	0.0734816	0.0753407	0.9753231	0.3337
MIN	0.1143897	0.6444960	0.1774871	0.8598
VPPA	-3.442E-06	1.758E-05	-0.1957442	0.8455

Log likelihood	-38.746842	
Cases with OUTCOM = 1		34
Cases with OUTCOM = 0		24

between the defender and the potential attacker. Now, all other factors being equal (e.g., the same balance of military power), if the defender had "won" in the last diplomatic confrontation with the attacker (i.e., a score of "0" on "dd") and this were changed to a diplomatic "defeat" (i.e., a score of "1" on "dd") I would hypothesize that deterrence (i.e., "outcom") would now be less likely to succeed (i.e., to decrease from "1"- a success to "0"- a failure). This seems likely because if the defender "lost" in the last confrontation between the defender and the potential attacker, the potential attacker would have less respect for the defender than if the defender had won. Therefore, the attacker would probably be less "deterred" by the defender. If so, then I should of hypothesized a negative relationship between "dd" and "outcom" (i.e., that as a score increased on "dd" - from "0" to "1" - then the score on "outcom" would be more likely to decrease - from "1" to "0"). If I had used "dd" (which I did not - this is only a hypothetical) and the sign on the coefficient for "dd" in the "main equation" was "positive," then having a "T-STAT" greater than 2.0 is actually strong evidence against the hypothesis. That is, we are very certain that there is a positive association when I had hypothesized in answering question #1 that there would be a negative association. If this occurred, the result would definitely not support my hypothesis.

<u>Question #3 asks to interpret the "T STAT.s" or "t ratios" in my main equation on page 206</u>. Looking at the "T STAT.s" in the "main equation" on page 206, it is clear that since none of them has an absolute value (either positive or negative) of at least 2.0, none of the independent variables has a statistically significant impact on the probability that deterrence will be successful.

<u>Question #4 says to assess the degree of multicollinearity in my "main equation" on page 206</u>. Since none of the independent variables are statistically significant, perhaps multicollinearity is causing one, or more, of them to be statistically insignificant. This is the reason that I estimated the four "deletion equations" that follow the "main equation" (see pages 206-207). Each of these four equations re-estimates the "main equation" but deletes one of the four independent variables. The idea behind this procedure is that if two of the independent variables are strongly related, when one of them is "deleted" (i.e., omitted) from the "main equation" the one that remains may become statistically significant. So, what I need to do is to see if any independent variable in any of the four additional equations (i.e., "deletion equation #1," "deletion equation #2," etc.) has a "T-STAT." with an absolute value of 2.0 or greater. Looking at the four deletion equations (pages 206-207), I see that no independent variable was ever statistically significant in any of the four deletion equations. Therefore, I would conclude that multicollinearity was not a serious problem. Perhaps my theories were poor. If, for example, "vppa" had been statistically significant (i.e., had a "T-STAT." with an absolute value of over 2.0) in several of the "deletion equations," I could have said that perhaps "vppa" was insignificant in the "main equation" because of multicollinearity. Even if "vppa" had been statistically significant in several of the "deletion equations" this would pertain to "vppa" only. Hence, I could not say that, for example, "min" was effect by multicollinearity. But, in my case, none of this applies. I simply can not make the case that any of my independent variables were insignificant in the "main equation" due to multicollinearity.

<u>Question #5 asks you to interpret the standardized and unstandardized coefficients. I'll answer the standardized portion first</u>. From page 128, we know that the standardization formula is as follows:

$$standardized\ b = (unstandardized\ b)\left(\frac{standard\ deviation\ of\ x}{standard\ deviation\ of\ y}\right)$$

Verbally, the standardization formula says that in order to "standardized b," first take the "unstandardized b" and multiply it by the ratio of the standard deviation of that independent variable to the standard deviation of the dependent variable. So, the first step would be to take the unstandardized coefficients from the "main equation." Looking at the top of page 206 we see that the unstandardized coefficients are as follows: maipd .153; min .059; vppa 1.061E-05 (the "E-05" is scientific notation meaning to move the decimal point 5 places to the left - thus, 1.061E-05 becomes .00001061); and milall -1.131. The standardization formula tells me that I also need the standard deviation for each variable. Looking at page 215, I see that the standard deviations are as follows: maipd 3.95; min .451; vppa 16042; milall .5. We also need the standard deviation for outcom which is .497. So, applying the standardization formula:

For maipd:
[.153][(3.95)/(.497)] = [.153][7.947] = 1.21

For min:
[.059][(.451)/(.497)] = [.059][.907] = .053

For vppa:
[.00001061][(16042)/(.497)] = [.00001061][32277.6] = .342

For milall:
[-1.131][(.5)/(.497)] = [-1.131][1.006] = -1.137

As we learned on pages 127-129, we then take the ratio of the absolute values of the standardized coefficients for each possible comparison. Thus, comparing maipd to the other independent variables, we could say: maipd is approximately 23 times as important as min (because 1.21/.053 = 22.83); maipd is approximately 3.5 times as important as vppa (because 1.21/.342 = 3.538); and maipd is approximately the same importance as milall (because 1.21/1.137 = 1.064 - remember to disregard the minus sign on -1.137, thus it is just 1.137). Since we have already compared maipd and min, we can simply compare min to vppa, and milall as follows: min is approximately one-seventh as important as vppa (because .053/.342 = .154; min is approximately one twentieth as important as milall (because .053/1.137 = .046). Since vppa has already been compared to maipd and min, we need only compare it to milall. We find that vppa is only one-third as important as milall (because .342/1.137 = .301). Since milall has already been compared to each other independent variable, we have completed the standardized portion of question #5.

Now for the magnitude assessment. From the discussion in the last paragraph on page 155 and the first paragraph on page 156, I know that I can not make the same magnitude statement for each coefficient (i.e., each "b") in a logit equation that I did for the linear regression equation in Assignment #2 or that we learned on pages 109-110 (just keep reading). For example, the coefficient value (or "b") for "maipd" in the "main equation" on the top of page 206 (.153) can not be interpreted to say if "maipd" increases by one unit, then the probability that deterrence will be successful increases, on average, by .153 units. From pages 202-203, I know that I am suppose to choose the independent variable with the highest "T-STAT." as my "comparison" variable. Okay, looking at the results for my "main equation" on page 206, I see that "maipd" has the "T-STAT." with the highest absolute value (1.635). The purpose of this is to see how much effect a change in the value of "maipd" has on the probability that deterrence will be successful if each other independent variable is set at a typical level (for non-dummy variables this is the mean and for dummy variables we will use "1") and not changed (just keep reading, it will become clear). Remember, the "y intercept" (i.e., "C" on your printout) is not an independent variable. Looking at page 214, I find that "maipd" is not a dummy variable (maipd can range between 1 and 10). So, the instructions for question #5 tell me that if my "comparison" variable (maipd in my case) is not a dummy variable, I should first set it at the mean level and then "compare" it to a situation in which the "comparison" variable is set one standard deviation above the mean. Looking at page 215, I find that the mean score on "maipd" is 5.103 (meaning that typically the protege imports slightly over half of its imported armaments from the defender) and the standard deviation of maipd is 3.95. As the variable list on page 214 tells us, "min" is a dummy variable, so let us set it at "1" (meaning the protege possesses strategic minerals). Looking again at page 214 we find that vppa is a non-dummy variable. So, as we were directed in question 5, let us set it at the mean value. According to page 215 the mean score on vppa is 12355 (meaning that the average protege had a population of approximately 12,355,000 people). Page 214 shows that milall is a dummy variable, so let us set it at "1" (meaning that there is a military alliance between the defender and the protege). So, as you read on pages 155-158, we take the "y intercept" (i.e., "C" -.075 - see page 206) and add it to the product of the coefficients from the main equation multiplied by the values for the appropriate variables mentioned above. So, looking at the "main equation" on page 206, the "y intercept" (i.e., "C") is -.075; the coefficient in the main equation for "maipd" is .153 which we now multiply times the mean score of 5.103; the coefficient for "min" in the main equation is .059 which we multiply by 1 (since min is a dummy variable); the coefficient for "vppa" is .00001061 which we multiply by the mean score of 12355; the coefficient for "milall" is -1.130 which we multiply by 1 (since milall is a dummy variable). Thus, we have the following:

$$-.075 + [(.153)(5.103)] + [(.059)(1)] + [(.00001061)(12355)] +$$

$$[(-1.130)(1)]$$ which when multiplying inside the parentheses yields the following:

$$-.075 + .780 + .059 + .131 + (-1.130) = -.235$$

Now from pages 157-158, we know that -.235 becomes the exponent of base "e" (which we know from past reading is equal to approximately 2.71828) in the equation:

probability = $e^{-.235}/(1 + e^{-.235})$

Using the e^x key on a scientific calculator (your scientific calculator may differ - also watch for "INV," "Shift" or "2nd Function" keys) we find that $e^{-.235}$ = .790 (i.e., $2.71828^{-.235}$ = .790).

Putting this value in the above formula we have: .790/(1 +.790) = .790/1.790 = .44

This tells us that if the protege has an average percentage of its imported armaments coming from the defender, the protege possesses strategic minerals, the protege is of average population (i.e., about 12,355,000 people) and does have a strategic alliance with the defender, there is approximately a 44% chance that if the protege is threatened, the defender will be able to successfully deter the threat.

Now the "comparison" is to see how much this probability increases or decreases if the comparison variable (maipd in our case) is set one standard deviation above the mean. Thus, the computation would be the same as before except that instead of multiplying the coefficient of maipd (.153) by 5.103 we now multiply it by one standard deviation above the mean. Since the mean is 5.103 and the standard deviation is 3.95 (see page 215), this means that we now multiply .153 by 9.053 (5.103 + 3.95 = 9.053). Had my "comparison" variable been a dummy variable, I would have performed the computations with that variable set at "1" and then at "0." So, let us proceed. All we need to do is to change 5.103 to 9.053 and proceed as we did previously.

-.075 + [(.153)(9.053)] + [(.059)(1)] + [(.00001061)(12355)] +

[(-1.130)(1)] which when multiplying inside the parentheses yields the following:

-.075 + 1.385 + .059 + .131 + (-1.130) = .37

So,

probability = $e^{.37}/(1 + e^{.37})$ = 1.45/(1 + 1.45) = 1.45/2.45 = .59

This tell us that, assuming that the protege continued to possess strategic minerals, was of average population size for a protege and had a military alliance with the defending nation, if the protege raised the percentage of its imported armaments from the defender from approximately 50% to approximately 90% (from a score of approximately "5" to "9" on "maipd"), the probability that deterrence would be successful increases from approximately 44% to approximately 59%. That concludes the answer to question #5.

Question 6 asks you to calculate and interpret the likelihood ratio index. The likelihood ratio index is analogous to R^2 except that it is designed for limited dependent variable models such as logit (see pages 158-159). From pages 158-159, we know that the formula for the likelihood ratio index is as follows:

LRI = 1 - [(log likelihood from the main equation)/(the log likelihood from an equation using just a constant to predict the dependent variable)]

From page 206 we know that the log likelihood from the main equation is -37.43. Additionally, from page 203, we know that the log likelihood for the equation where "outcom" is predicted by only a constant (thus no independent variables) is -39.336. Inserting these numbers into the LRI formula yields:

LRI = 1 - [(-37.43)/(-39.336)]

LRI = 1 - [.95]

LRI = .05

An LRI (likelihood ratio index) score of .05 might be interpreted to mean that the model has a low degree of explanatory power. However, note the caution that I mentioned on page 159 concerning the use of the LRI statistic to assess explanatory power. Since none of the independent variables is statistically significant (and there does not seem to be high multicollinearity) it is not surprising that my LRI statistic is low. Perhaps I should re-think my model. I told you at the outset I thought this was a poor model. It was only an example.

Variable List

border: the protege and the potential attacker share a common border (dummy variable: 1 = yes, 0 = no)

capit: in the most recent case of extended deterrence by the defender the deterrence failed and the defender did not intervene with military force to protect the protege [dummy variable: 1 = yes, the preceding is what occurred, 0 = the preceding did not occur (because the most recent extended deterrence was successful, it was unsuccessful but the defender intervened militarily to protect the protege, or there was no past case of extended deterrence by this defender)]

dd: diplomatic defeat of the defender in the most recent confrontation between the defender and the potential attacker (dummy variable: 1 = yes, 0 = no)

dpdbd: diplomatic put-down by defender in the most recent confrontation between the defender and the potential attacker (dummy variable: 1 = yes, 0 = no)

fbfds: firm-but-flexible diplomatic strategy (dummy variable: 1 = yes, 0 = no). A "no" would include either a strategy of bullying or no diplomatic strategy (likely due to an absence of negotiations and diplomatic interaction between the defender and the potential attacker).

ftbdp: The percentage of the defender's total foreign trade accounted for by the protege. (not a dummy variable)

ibf: immediate balance of forces: the ratio of defender and protege forces to potential attacker forces. If the ratio is "1" this means that the defender and the protege have military forces equal to those of the potential attacker. A ratio of 1.5 means that the defender and protege have 1.5 times (i.e., 150% of) the military forces of the attacker. A ratio of .7 means that the defender and protege have only .7 (or 70%) of the military capability of the attacker. Thus, "ibf" is not a dummy variable.

ltbf: Same as "ibf" except that this is the ratio of military capability of the defender and the protege to potential attacker forces over the long-term (i.e., the ability to fight a war for years) rather than over the short-term (i.e., the ability to fight a war for several months). See the description of "ibf" on page 213. Thus, "ltbf" is not a dummy variable.

maipd: military arms imported by protege from defender as a percentage of the protege's total arms imports. A scale from 1 to 10 with 1 = 0-10% and 10 = 91-100% (not a dummy variable).

milall: military alliance between defender and protege (dummy variable: 1 = yes, 0 = no)

min: protege possesses strategic minerals (dummy variable: 1 = yes, 0 = no)

nw: nuclear weapons: defender possession of operational nuclear weapons (dummy variable: 1 = yes, 0 = no)

outcom: outcome of each case of extended deterrence (dummy variable: 1 = success, 0 = failure). This is the dependent variable. For deterrence to be "successful," the attacker must both fail to acquire the protege and not engage in sustained combat to try to attain the protege. Only if both conditions are met is deterrence considered "successful."

pme: reciprocal military escalation (dummy variable: 1 = yes, 0 = no). A "no" would mean that one side escalated but the other side did not.

stale: stalemate in the most recent confrontation between the defender and the potential attacker (dummy variable: 1 = yes, 0 = no)

succes: most recent case of extended deterrence by the defender (dummy variable: 1 = successful, 0 = either failure or no past case of extended deterrence)

vppa: value of the protege to the potential attacker: population of the protege in thousands (e.g., a score of 800 would mean that the population of the protege was 800,000 - not a dummy variable)

Variable Means and Standard Deviations

Variable Name	Mean	Standard Deviation
border	.672	.473
capit	.069	.256
dd	.138	.348
dpdbd	.293	.459
fbfds	.448	.502
ftbdp	1.745	2.424
ibf	1.209	1.204
ltbf	1.968	2.046
maipd	5.103	3.95
milall	.431	.500
min	.276	.451
nw	.259	.442
outcom	.586	.497
pme	.517	.504
stale	.259	.442
succes	.466	.503
vppa	12355	16042

Table A-1. Cases of Attempted Extended-Immediate Deterrence, 1885–1984

Case	Year	Potential Attacker	Protégé	Defender	Outcome[a]
1	1885	Russia	Afghanistan	Britain, India	S
2	1885–86	Bulgaria	Serbia	Austria-Hungary	S
3	1886	Greece	Turkey	Britain	S
4	1894	Japan	Korea	China	F
5	1897	Greece	Crete	Britain, Turkey	S
6	1898	France	Egyptian Sudan	Britain	S
7	1902–3	Germany	Venezuela	United States	S
8	1903–4	Japan	Korea	Russia	F
9	1903–4	Colombia	Panama	United States	S
10	1905–6	France	Morocco	Germany	F
11	1905–6	Germany	France	Britain	S
12	1906	Turkey	Egypt	Britain	S
13	1908–9	Serbia, Russia	Austria-Hungary	Germany	S
14	1911	Italy	Tripoli	Turkey	F
15	1911	France	Morocco	Germany	F
16	1911	Germany	France	Britain	S
17	1912–13	Serbia, Russia	Austria-Hungary	Germany	S
18	1913	Romania	Bulgaria	Russia	S
19	1913	Bulgaria	Greece	Serbia	F
20	1913	Serbia	Albania	Austria-Hungary	S
21	1914	Austria, Germany	Serbia	Russia	F
22	1914	Russia, Serbia	Austria-Hungary	Germany	F
23	1914	Germany	France	Britain, Russia	F
24	1914	Germany	Belgium	Britain	F
25	1921	Panama	Costa Rica	United States	S
26	1922	Turkey	Greece	Britain	S
27	1935	Italy	Ethiopia	Britain	F
28	1935–36	Japan	Outer Mongolia	Soviet Union	S
29	1937	Soviet Union	Manchukuo	Japan	S
30	1938	Soviet Union	Manchukuo	Japan	F
31	1938	Germany	Czechoslovakia	Britain, France	F
32	1938–39	Italy	Tunisia	France, Britain	S
33	1939	Germany	Poland	Britain, France	F
34	1940–41	Soviet Union	Finland	Germany	S
35	1946	Soviet Union	Iran	United States	S
36	1946	Soviet Union	Turkey	United States	S
37	1948	Soviet Union	West Berlin	United States, Britain	S
38	1950	China	Taiwan	United States	S
39	1950	United States	North Korea	China	F
40	1954–55	China	Quemoy-Matsu	United States	S
41	1957	Turkey	Syria	Soviet Union	S
42	1961	Iraq	Kuwait	Britain	S
43	1961	North Vietnam	Laos	United States	S
44	1961	India	Goa	Portugal	F
45	1961–62	Indonesia	West Irian	Netherlands	F
46	1964–65	Indonesia	Malaysia	Britain	F
47	1964–65	North Vietnam	South Vietnam	United States	F
48	1964–65	United States	North Vietnam	China	S
49	1967	Israel	Syria	Egypt	F
50	1967	Turkey	Cyprus	Greece	F
51	1970	Syria	Jordan	Israel	S
52	1971	India	Pakistani Kashmir	China	S
53	1974	Turkey	Cyprus	Greece	F
54	1975	Morocco	Western Sahara	Spain	F
55	1975	Guatemala	Belize	Britain	S
56	1977	Guatemala	Belize	Britain	S
57	1979	China	Vietnam	Soviet Union	F
58	1983	Libya	Chad	France	S

[a]S = successful deterrence, F = failure of deterrence.