Introduction to Statistics: Homework 2

Dichotomous/Nominal Variables and Non-linear Functional Forms

DUE THURSDAY, NOVEMBER 18th

*Please type your responses. For the analysis: copy your regression output and paste it into your document. <u>When you are asked to "interpret the coefficient" you should evaluate the statistical significance of the coefficient as well as the magnitude of the relationship.</u> **<u>Be sure to interpret coefficients in the context of the full regression!</u>***

1. A 1976 dataset provides a way to examine the relationship between a number of individual-level characteristics and monthly wages. The outcome variable is monthly wages. Here are some regression results:

	Coef.	SE	T	P
Black (1=yes)	-241.332	38.472	-6.270	0.000
Lives in the South (1=yes)	-72.764	27.235	-2.670	0.008
Lives in an Urban Area (1=yes)	183.197	27.934	6.560	0.000
Age (in years)	19.314	4.007	4.820	0.000
Constant	243.356	135.558	1.800	0.073

   a. Interpret the coefficient on the constant. [3 points]

   b. Interpret the coefficient on each of the other four variables. [12 points]

   c. What is the expected monthly wage for a 30 year old Black individual who does not live in the South and lives in an urban area? [3 points]

   d. From a theoretical perspective how would you interpret the coefficient on Age? Do you think it is an unbiased estimate of the relationship between age and wages? Why or why not? [8 points]

2. For items in this section use the turnout2008 dataset. The dependent variable we'll focus on is turnout in the 2008 presidential election. [Note: there are more advanced statistical approached for dealing with dichotomous dependent variables. However, they rarely lead to meaningfully different findings. In this case, think about predicted values as the predicted probability that an individual turns out.]

Variable	Description
Turnout	Voted in 2008 (1=yes)
Anycontact	Report being contacted by Democratic or Republican campaign (1=yes)
strength_pid	Strength of party affiliation (0=pure independent; 1="leaner"; 2=weak identifier; 3=strong identifier)
Education	Education (1=no high school; 2=some HS, no diploma; 3=HS diploma; 4=some college; 5=associate degree; 6=bachelor's degree; 7 advanced degree)

a. Estimate a regression model predicting turnout based on whether the respondent reported being contacted by a political party during the campaign.
   i. Interpret the coefficient on the constant. [3 points]
   ii. Interpret the coefficient on the "contact" variable. [3 points]
b. Two of the other factors that political scientists believe might affect turnout are how strongly people feel about the political parties (because more committed partisans are more likely to turn out because they feel more strongly about the outcome) and education (because more educated people are more likely to be knowledgeable about political matters). Estimate a new model adding strength of party affiliation and education to the "contact" model you estimated above.
   i. Interpret the coefficient on the constant. [3 points]
   ii. Interpret the coefficients on each of the three IVs. [9 points]
   iii. Discuss what happened to the coefficient on the "contact" variable when you added the two new variables. Does the change seem substantial? What might explain this change or lack of change? [6 points]
   iv. Do you think the estimate of the effect of being contacted by a campaign is biased? Why or why not? If you think it is biased, what could we do to make it less biased? [8 points]
c. The relationship between education and turnout may not be linear. As we discussed in class, we can assess this possibility by measuring education using a series of indicators. These indicators are included in the dataset. Rerun the analysis from b) above replacing the education measure with these indicator variables. Use "bachelor's degree" as the reference category.
   i. Interpret the coefficient on the constant. [3 points]
   ii. Interpret the coefficients on "advanced degree" and "some college" [6 points]
   iii. Calculate the predicted value (probability of voting) for a weak partisan with a high school diploma who has been contacted by a political party. [4 points]
   iv. Test whether individuals with "some college" were more likely to vote than those with an "associates degree." [4 points]
   v. Test whether individuals with "some college" were more likely to vote than those whose highest level of education is a high school diploma. [3 points]
3. For these items use the literacy dataset. This dataset is from 1970 and has information about countries around the world. The following variables are included:

Variable   Description

LITERACY	Literacy rate (number of literate residents per 1000)
GDPCA	GDP per capita
GDPCA2	GDP per capita (squared)
DEMOCRAC	Democracy (1=yes)

Estimate a model predicting literacy rate with the other three variables in the dataset.

    i. What does the coefficient on GDPCA2 tell us? [3 points]

    ii. If the coefficients on GDPCA and GDPCA2 were both statistically insignificant, should we conclude that there is no relationship between GDP per capita and literacy rates? How could we check this? [6 points]

    iii. Fill in this table (or make one of your own) with predicted literacy rates in *democracies.* [6 points]

GDP per capita	Predicted literacy rate
500	
1000	
1500	
2000	
2500	
3000	

    iv. Use the predicted values you just calculated to draw a simple graph of the relationship between GDP per capita and literacy rate. Describe this relationship in words. [7 points]