1. (25 points) Using data from the 1992 Health and Retirement Survey, I estimated a Tobit and an OLS model of the determinants of a person's networth (assets-liabilities). **Networth is measured in thousands of dollars and is truncated at zero.** Negative values of networth were possible, but were reported as zero.

<table>
<thead>
<tr>
<th>Networth=dependent variable</th>
<th>Tobit estimates</th>
<th>OLS estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coeffic.</td>
<td>t-stat</td>
</tr>
<tr>
<td>married (dummy=1 if married)</td>
<td>218.49</td>
<td>15.53</td>
</tr>
<tr>
<td>smoker (dummy=1 if smoker)</td>
<td>-80.10</td>
<td>-5.63</td>
</tr>
<tr>
<td>age (age in years)</td>
<td>12.69</td>
<td>9.71</td>
</tr>
<tr>
<td>black (dummy=1 if black)</td>
<td>-127.44</td>
<td>-7.43</td>
</tr>
<tr>
<td>tot_earn (annual earnings)</td>
<td>0.0022</td>
<td>23.24</td>
</tr>
<tr>
<td>educyrs (years of education)</td>
<td>40.21</td>
<td>20.85</td>
</tr>
<tr>
<td>intercept</td>
<td>-1087.19</td>
<td>-13.45</td>
</tr>
<tr>
<td>std error of residual</td>
<td>526.01</td>
<td>502.9</td>
</tr>
<tr>
<td>sample size</td>
<td>7023</td>
<td>7023</td>
</tr>
</tbody>
</table>

For all answers below, describe the basic steps for your calculations. The description need not include the actual numbers, but should make reference to X (the vector of explanatory variables), B (the vector of coefficients on those explanatory variables), etc. Define your notation.

a. Based on the above tobit estimates, what is the probability that a single white non-smoker who is 60 years old, has earnings of $50,000 and 16 years of education has networth less than or equal to zero?

b. In the tobit model, there are two types of networth. The first is the actual value (call it NW*) which could be either positive or negative. The second is the observed value (call it NW) which is set to zero if the person's NW is less than or equal to zero.

   i. What is the predicted value of **actual** networth for the person in (a)?
   ii. What is the predicted value of **observed** networth for the person in (a)?
   iii. What is the marginal effect of smoking on the predicted value of **observed** networth for the person in (a)?

c. What is the probability that the person in (a) will have networth above $200,000?

d. How do the OLS coefficient estimates compare to the Tobit estimates? Are the differences what you would expect? Explain by describing why there is a bias in the OLS estimates for this example.

e. Explain how you could test the null hypothesis that the tobit specification of the networth equation is identical for blacks and whites. Describe the regression(s) that you would run, how you would form the test statistic, the distribution of the test statistic, and under what conditions you would reject the null hypothesis.
2. (20 points) Using data from the April 1993 Current Population Survey, I estimated a probit model of whether a worker receives health insurance from his or her employer. The description of the independent variables is listed in the right-most column of the table below.

<table>
<thead>
<tr>
<th>inhelth</th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>t-statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>wageinc</td>
<td>.0000168</td>
<td>1.47e-06</td>
<td>11.41</td>
<td>(Annual wage income)</td>
</tr>
<tr>
<td>company</td>
<td>.0000566</td>
<td>8.97e-06</td>
<td>6.31</td>
<td>(Number of employees at establishment)</td>
</tr>
<tr>
<td>female</td>
<td>-.3318909</td>
<td>.0464251</td>
<td>-7.15</td>
<td>(Dummy=1 if female)</td>
</tr>
<tr>
<td>black</td>
<td>.1001601</td>
<td>.0957732</td>
<td>1.05</td>
<td>(Dummy=1 if black)</td>
</tr>
<tr>
<td>intercept</td>
<td>.5641257</td>
<td>.0622969</td>
<td>9.06</td>
<td>(constant term)</td>
</tr>
</tbody>
</table>

Log likelihood = -2007.944

a. Compute the probability that a white male earning $40,000 at a firm with 100 employees will be covered by health insurance. Describe the basic steps for your calculations.

b. For the person described in (a), what would be the effect of an additional 100 employees on the probability of being covered by health insurance? Describe the basic steps for your calculations.

c. If you had a sample of 10,000 people with information on all of the independent variables listed above, how could you use the above model to predict the number of people that would have health insurance coverage?
3. (20 points) Using Current Population Survey data from 1980 through 1999, I created a sample of workers that were earning the minimum wage in a given year. Using the panel features of the data, I was able to determine what type of job transition the worker made over the subsequent year. I coded a variable TRANS=1 if the worker stayed in the same industry and occupation, TRANS=2 if the worker switched industry and occupation, and TRANS=3 if the worker left the labor force. I estimated a multinomial logit model to examine how personal characteristics affect the transition probabilities. The results are listed below:

Multinomial regression
Number of obs = 33520
Log likelihood = -33964.201

------------------------------------------------------------------------------
  trans |      Coef.   Std. Err.           t
-------------+----------------------------------------------------------------
 trans=2      |
  female  |  -.1596631   .0282338    -5.66  (dummy=1 if female)
   age    |  -.0358465   .0010145   -35.33  (person's age in years)
  educ1   |   .1185193   .0141759     8.36  (years of education)
  _cons   |    .214764   .0494056     4.35  (intercept)
-------------+----------------------------------------------------------------
 trans=3      |
  female  |  -.0656754   .0290774    -2.26
   age    |  -.0190639   .0008843   -21.56
  educ1   |  -.1139444   .0145286    -7.84
  _cons   |   .2102352   .0501867     4.19
------------------------------------------------------------------------------
(Outcome trans2==1 is the comparison group)

a. What is the estimated probability that a 20 year old male with 12 years of education will
   i. not leave the labor force or change occupation and industry?
   ii. leave the labor force?
   iii. the probability he will change occupation and industry?
Give a brief description of how you calculated your answer.

b. As education increases, what happens to the probability that a worker has trans=1? (I don't need a numerical answer here, just indicate whether the probability would rise or fall and how you determined this. Notice that I am not asking about relative odds.

c. If the "base" category was changed from trans=1 to trans=2, what would the coefficient on female be for trans=2?
ANSWER 4 OR 5 (NOT BOTH)

4. (20 points) In the Health and Retirement Survey, people are asked the following question:
   In deciding how much of their (family) income to spend or save, people are likely to think about different financial planning periods. In planning your (family's) saving and spending, which of the time periods listed in the booklet is most important to you [and your (husband/wife/partner)]?
   1. Next few months
   2. Next year
   3. Next few years
   4. Next 5-10 years
   5. Longer than 10 years

Using the above responses (coded from 1-5), I estimated an ordinal probit model. The results are listed below:

Ordered probit estimates
Number of obs =       6306

Log likelihood = -9290.5534

| horiz | Coef. | Std. Err. | z     | P>|z| | 95% Conf. Interval |
|-------+-------+----------+-------+-------+------------------|
| smokes| -1.148842 | 0.294199  | -3.90 | 0.000 | -.1725461 to -.0572223 |
| age   | -0.0194655 | 0.0027473 | -7.09 | 0.000 | -.02485 to -.0140809 |
| married| 0.2471847 | 0.0290373 | 8.51  | 0.000 | .1902726 to .3040968 |
| black | -0.2586497 | 0.0351229 | -7.36 | 0.000 | -.3274892 to -.1898101 |

_cut1 through _cut4 represent the intercept terms for the ordinal probit.

a. Based on the above, calculate the probability that a 25 year old white smoker will indicate she has a planning horizon of 10 or more years. Give a general description of how you calculated your answer.

b. How could you calculate the expected value of the response variable for the person described in a? Don't do the numerical calculation, just describe how you would do it.
5. (20 points) In the Survey of Consumer Finance data that you used in your last homework, the sample size was 1772 and 230 people were denied credit. Of the 1772 people, 1405 were married and 157 of the married people were denied credit. Of the 367 single people, 73 were denied credit.

Define the following logit model of whether a person is denied credit:

\[
\text{Prob}[CD_i=1] = \frac{\exp(a + \beta S_i)}{1 + \exp(a + \beta S_i)}
\]

where
- \( CD_i \) is a dummy variable that equals one if person \( i \) is denied credit; zero otherwise.
- \( S_i=1 \) is a dummy variable that equals one if person \( i \) is single; zero otherwise.
- \( a \) and \( \beta \) are parameters to be estimated.

a. Using all of the numeric information provided, write out the log-likelihood function for this problem. Be sure that your notation is clear and that the numerical information on sample sizes is used. Ambiguity will be penalized.

b. Solve for the maximum likelihood estimate of \( a \) and \( \beta \). Note: to help you here, note that

\[
\text{if } y = \ln \{\exp(x)/(1+\exp(x))\} \implies \frac{dy}{dx} = \frac{1}{1+\exp(x)}
\]

\[
\text{if } z = \ln \{1/[1+\exp(x)]\} \implies \frac{dz}{dx} = -\exp(x)/[1+\exp(x)]
\]

also, remember that from the chain rule: \( \frac{dz}{db} = (dz/dx)(dx/db) \) and \( \frac{dy}{db} = (dy/dx)(dx/db) \)