Econ 533
Problem Set #1
Answer Sheet

1. **Obtain the data set "hybrid.xls" off of the class webpage. This data set contains the following variables on whether or not a series of individuals have chosen to purchase a hybrid car (rather than a standard vehicle):**

Variable	**Description**
BUY	**=1 if the individual buys a hybrid vehicle; =0 otherwise**
MALE	**= 1 if the individual is male; = 0 otherwise**
CONS	**= 1 if the individual belongs to an environmental group (e.g., Nature Conservancy. etc); = 0 otherwise**
FOREIGN	**= if the vehicle being considered if foreign made; = 0 otherwise**
AGE	**= the age of the individual**
COSTD	**= the additional cost of the hybrid vehicle relative to a standard vehicle ($1000's)**
GAS	**= the current price of gasoline**

2. **Using TSP:**

   a. **Summarize the available data to insure that the available variables are within the anticipated ranges.**

   The attached TSP program and output files illustrate how one would read in the data set and summarize the variables using the MSD command.

   b. **Estimate a Linear Probability Model of the probability of purchasing a hybrid vehicle as a function of cost differential (i.e., *costd*). Is the cost differential found to be a significant determinant of the purchase probability? Estimate the predicted probability of purchasing a hybrid vehicle for *costd*=0 and *costd*=10, along with the standard deviation of each probability.**

   The linear probability model does indicate that the cost differential (COSTD) is a significant indicator of probability of purchasing a hybrid vehicle, with the coefficient on COSTD estimated to be -0.28, with a corresponding t-statistic of 14.4 and a P-value of less than (.001). Clearly this coefficient is significant at any reasonable level. The level of the coefficient indicates that each $1000 increase in the cost premium for a hybrid vehicle reduces the probability of purchasing the hybrid by 28%. This is born out by the ANALYZ routine results, showing that the fitted probability of buying a hybrid is 98.9% for when there is no cost difference and -182% when the cost difference reaches $10,000. Of course, the latter figure is nonsense, since probabilities cannot be negative. This highlights one of the limitations of the linear probability model.

   c. **Now estimate a LPM including the other explanatory variables. Is *costd* still a significant determinant of the hybrid purchase probability? Again, estimate the predicted probability of purchasing a hybrid vehicle for *costd*=0 and *costd*=10, along with the standard deviation of each probability. To do**

this, you will have to hold all the other explanatory variables fixed at their means. Are the predicted probabilities from this regression and those from the simpler model comparable? Why or why not?

The second and more complicated linear probability model still shows the cost differential to be a significant predictor of hybrid purchases (with a coefficient of -.281 and a corresponding t-statistic of -14.9. The results also suggest that all of the other factors (except whether or not it is foreign made) impact the hybrid purchase decision. Specifically, we find that males and older individuals are less likely to buy a hybrid, while higher gas prices and belonging to an environmental group both increase the odds of purchasing a hybrid. As to which factors are most important, one indication of that is found by multiplying each of the significant coefficients by the corresponding range of the variable in question. The table below provides the results:

Variable	Coefficient Estimate	Range of Variable	Coeff*Range
MALE	-.091	1	-.091
CONS	.192	1	.192
AGE	-.00175	65	-.114
COSTD	-.281	3	-.843
GAS	.146	1.10	.161

These results suggest that the cost difference has the biggest impact, with gas prices and belonging to an environmental group coming in second. Age and gender have the smallest impacts, though certainly still somewhat important.

The results in part c of the problem are similar to those that we obtained in part b. This need not be the case in general for several reasons. Typically, when you include only a few variables in the model, they will “pick-up” the effects of those variables missing from the model.

TSP Program:

```
?
?       Problem Set #1
?
?       The first step is to obtain and read in the requisite data
?
smpl    1 1000;
read    (file='hybrid.xls') BUY MALE CONS FOREIGN AGE COSTD GAS;
;
?
?       Question 2a
?       The second step is to compute summary statistics for each
variable to
?       verify that they lie within expected bounds, have the expected
means, etc.
?       At this stage, I will also store the means for each variable,
which will be
?       useful later in answering question 2c.
?
msd     BUY MALE CONS FOREIGN AGE COSTD GAS;
set     mmale   = @mean(2);
set     mcons   = @mean(3);
set     mforeign= @mean(4);
set     mage    = @mean(5);
set     mcostd  = @mean(6);
set     mgas    = @mean(7);
?
?       Question 2b:
?       The next step is to estimate a simple linear probability model.
This is done
?       Using both the OLSQ and the LSQ commands
?       We also need to compute the fitted probabilities for loan
approval
?       for both males and females
?
OLSQ    buy c costd;
frml  eq1 buy = a0 + a1*costd;
param   a0 0 a1 0;
lsq     eq1;
frml    eqa1 PFITA1  = A0 + A1*(0);
frml    eqa2 PFITA2  = A0 + A1*(10);
analyz  eqa1 eqa2;
?
?       Question 2c:
?       In question 2c, you are asked to estimate a more complicated LPM
?       and to compute the correponding fitted probabilities
?
OLSQ    buy c male cons foreign age costd gas;
frml  eq1 buy = b0 + b1*male + b2*cons + b3*foreign + b4*age
                    + b5*costd + b6*gas;
param   b0 0 b1 0 b2 0 b3 0 b4 0 b5 0 b6 0;
lsq     eq1;
frml    eqb1 PFITb1 = b0 + + b1*mmale + b2*mcons + b3*mforeign + b4*mage
                    + b5*(0) + b6*mgas;
frml    eqb2 PFITb2 = b0 + + b1*mmale + b2*mcons + b3*mforeign + b4*mage
                    + b5*(10) + b6*mgas;
analyz  eqb1 eqb2;
```

TSP Output File:

```
                     -------------------------------------
                     |        this copy licensed         |
                     |            for use by:            |
                     | TSP/GiveWin 5.0 User   #50AGT0054 |
                     -------------------------------------
                                TSP Version 5.0
                          ( 4/05/05) TSP/GiveWin  4MB
                     Copyright (C) 2005 TSP International
                              ALL RIGHTS RESERVED
                               01/25/08 2:46 PM
               In case of questions or problems, see your local TSP
               consultant or send a description of the problem and the
               associated TSP output to:
                               TSP International
                                P.O. Box 61015
                              Palo Alto, CA 94306
                                      USA
         PROGRAM
COMMAND  ***************************************************************
1  ?
1  ?       Problem Set #1
1  ?
1  ?       The first step is to obtain and read in the requisite data
1  ?
1  smpl    1 1000;
2  read    (file='hybrid.xls') BUY MALE CONS FOREIGN AGE COSTD GAS;
3  ;
3  ?
3  ?       Question 2a
3  ?       The second step is to compute summary statistics for each
   variable to
3  ?       verify that they lie within expected bounds, have the
   expected means, etc.
3  ?       At this stage, I will also store the means for each
   variable, which will be
3  ?       useful later in answering question 2c.
3  ?
3  msd     BUY MALE CONS FOREIGN AGE COSTD GAS;
4  set     mmale   = @mean(2);
5  set     mcons   = @mean(3);
6  set     mforeign= @mean(4);
7  set     mage    = @mean(5);
8  set     mcostd  = @mean(6);
9  set     mgas    = @mean(7);
10  ?
10  ?       Question 2b:
10  ?       The next step is to estimate a simple linear probability
    model. This is done
10  ?       Using both the OLSQ and the LSQ commands
10  ?       We also need to compute the fitted probabilities for loan
    approval
10  ?       for both males and females
10  ?
10  OLSQ    buy c costd;
11  frml eq1 buy = a0 + a1*costd;
12  param   a0 0 a1 0;
13  lsq     eq1;
14  frml    eqa1 PFITA1  = A0 + A1*(0);
15  frml    eqa2 PFITA2  = A0 + A1*(10);
16  analyz  eqa1 eqa2;
17  ?
17  ?       Question 2c:
17  ?       In question 2c, you are asked to estimate a more
    complicated LPM
17  ?       and to compute the correponding fitted probabilities
```

```
17  ?
17  OLSQ    buy c male cons foreign age costd gas;
18  frml eq1 buy = b0 + b1*male + b2*cons + b3*foreign + b4*age
18                    + b5*costd + b6*gas;
19  param   b0 0 b1 0 b2 0 b3 0 b4 0 b5 0 b6 0;
20  lsq     eq1;
21  frml    eqb1 PFITb1 = b0 + + b1*mmale + b2*mcons + b3*mforeign +
    b4*mage
21                    + b5*(0) + b6*mgas;
22  frml    eqb2 PFITb2 = b0 + + b1*mmale + b2*mcons + b3*mforeign +
    b4*mage
22                    + b5*(10) + b6*mgas;
23  analyz  eqb1 eqb2;
         EXECUTION
*******************************************************************************

Current sample:  1 to 1000

                             Univariate statistics
                             =====================

Number of Observations: 1000

                    Mean        Std Dev       Minimum       Maximum
BUY              0.37000        0.48305       0.00000       1.00000
MALE             0.50100        0.50025       0.00000       1.00000
CONS             0.34700        0.47625       0.00000       1.00000
FOREIGN          0.39800        0.48973       0.00000       1.00000
AGE             51.75400       15.86497      25.00000      90.00000
COSTD            2.20167        0.71397       1.00494       3.99961
GAS              2.31259        0.30861       1.80053       2.89942


                     Sum       Variance      Skewness      Kurtosis
BUY            370.00000        0.23333       0.53933      -1.71255
MALE           501.00000        0.25025    -0.0040060      -2.00400
CONS           347.00000        0.22682       0.64380      -1.58870
FOREIGN        398.00000        0.23984       0.41739      -1.82945
AGE          51754.00000      251.69718       0.32608      -0.73829
COSTD         2201.67408        0.50975       0.32749      -0.70467
GAS           2312.59251       0.095242       0.16105      -1.11092


                                     Equation   1
                                     ============

                      Method of estimation = Ordinary Least Squares


Dependent variable: BUY
Current sample:  1 to 1000
Number of observations:  1000

       Mean of dep. var. = .370000      LM het. test = 46.4681 [.000]
  Std. dev. of dep. var. = .483046     Durbin-Watson = 1.98590 [<.424]
Sum of squared residuals = 192.873  Jarque-Bera test = 80.6994 [.000]
   Variance of residuals = .193259   Ramsey's RESET2 = 9.09689 [.003]
Std. error of regression = .439613   F (zero slopes) = 208.152 [.000]
               R-squared = .172575    Schwarz B.I.C. = 602.984
      Adjusted R-squared = .171746    Log likelihood = -596.076

           Estimated    Standard
Variable  Coefficient     Error       t-statistic   P-value
C          .988804       .045087       21.9308       [.000]
COSTD      -.281060      .019481       -14.4275      [.000]
```

```
                              NONLINEAR LEAST SQUARES
                              =======================

EQUATIONS: EQ1

NOTE => The model is linear in the parameters.
Working space used: 10127
                                  STARTING VALUES

                    A0            A1
VALUE          0.00000       0.00000

F= 921.81239653  FNEW= 596.07620194  ISQZ=  0 STEP= 1.       CRIT= 477.76

CONVERGENCE ACHIEVED AFTER   1 ITERATIONS


    2 FUNCTION EVALUATIONS.

Number of observations = 1000     Log likelihood = -596.076
        Schwarz B.I.C. = 602.984

                         Standard
Parameter  Estimate        Error       t-statistic   P-value
A0         .988804       .045087       21.9308       [.000]
A1         -.281060      .019481       -14.4275      [.000]

Standard Errors computed from quadratic form of analytic first derivatives
(Gauss)

Equation: EQ1
Dependent variable: BUY

       Mean of dep. var. = .370000           R-squared = .172575
  Std. dev. of dep. var. = .483046  Adjusted R-squared = .171746
Sum of squared residuals = 192.873        LM het. test = 46.4681 [.000]
   Variance of residuals = .193259       Durbin-Watson = 1.98590 [<.424]
Std. error of regression = .439613

                         Results of Parameter Analysis
                         =============================


                         Standard
Parameter  Estimate        Error       t-statistic   P-value
PFITA1     .988804       .045087       21.9308       [.000]
PFITA2     -1.82180      .152553       -11.9421      [.000]

Wald Test for the Hypothesis that the given set of Parameters are jointly zero:

CHISQ(2) =   916.52660     ; P-value = 0.00000


                                     Equation   2
                                     ============

                      Method of estimation = Ordinary Least Squares


Dependent variable: BUY
Current sample:  1 to 1000
Number of observations:  1000

       Mean of dep. var. = .370000      LM het. test = 27.4922 [.000]
  Std. dev. of dep. var. = .483046     Durbin-Watson = 1.97216 [<.429]
Sum of squared residuals = 178.793  Jarque-Bera test = 57.6937 [.000]
```

```
   Variance of residuals = .180053   Ramsey's RESET2 = 10.8383 [.001]
Std. error of regression = .424327   F (zero slopes) = 50.2698 [.000]
               R-squared = .232979    Schwarz B.I.C. = 582.351
      Adjusted R-squared = .228344    Log likelihood = -558.174
```

Variable	Estimated Coefficient	Standard Error	t-statistic	P-value
C	.733664	.118982	6.16618	[.000]
MALE	-.090711	.026933	-3.36806	[.001]
CONS	.192101	.028252	6.79958	[.000]
FOREIGN	-.033953	.027453	-1.23679	[.216]
AGE	-.174896E-02	.847004E-03	-2.06488	[.039]
COSTD	-.281094	.018860	-14.9045	[.000]
GAS	.146169	.043593	3.35306	[.001]

NONLINEAR LEAST SQUARES
=======================

EQUATIONS: EQ1

NOTE => The model is linear in the parameters.
Working space used: 30493

STARTING VALUES

	B0	B1	B2	B3
VALUE	0.00000	0.00000	0.00000	0.00000

	B4	B5	B6
VALUE	0.00000	0.00000	0.00000

```
F= 921.81239653  FNEW= 558.17423477  ISQZ=  0 STEP= 1.       CRIT= 513.16
```

CONVERGENCE ACHIEVED AFTER   1 ITERATIONS

2 FUNCTION EVALUATIONS.

```
Number of observations = 1000     Log likelihood = -558.174
        Schwarz B.I.C. = 582.351
```

Parameter	Estimate	Standard Error	t-statistic	P-value
B0	.733664	.118982	6.16618	[.000]
B1	-.090711	.026933	-3.36806	[.001]
B2	.192101	.028252	6.79958	[.000]
B3	-.033953	.027453	-1.23679	[.216]
B4	-.174896E-02	.847004E-03	-2.06488	[.039]
B5	-.281094	.018860	-14.9045	[.000]
B6	.146169	.043593	3.35306	[.001]

Standard Errors computed from quadratic form of analytic first derivatives
(Gauss)

Equation: EQ1
Dependent variable: BUY

```
       Mean of dep. var. = .370000           R-squared = .232979
  Std. dev. of dep. var. = .483046  Adjusted R-squared = .228344
Sum of squared residuals = 178.793        LM het. test = 27.4922 [.000]
   Variance of residuals = .180053       Durbin-Watson = 1.97216 [<.429]
Std. error of regression = .424327
```

Results of Parameter Analysis
=============================

Parameter	Estimate	Standard Error	t-statistic	P-value
PFITB1	.988878	.043637	22.6614	[.000]
PFITB2	-1.82207	.147685	-12.3375	[.000]

Wald Test for the Hypothesis that the given set of Parameters are jointly zero:

CHISQ(2) = 982.47644 ; P-value = 0.00000

****

END OF OUTPUT.

MEMORY USAGE: ITEM:	DATA ARRAY	TOTAL MEMORY
UNITS:	(4-BYTE WORDS)	(MEGABYTES)
MEMORY ALLOCATED :	500000	4.0
MEMORY ACTUALLY REQUIRED :	47121	2.3
CURRENT VARIABLE STORAGE :	11158	