## Statistical and Econometric Methods

### Assignment #4
### (Discrete Data - Logit Analysis)

Using the information from assignment #3, perform the following:

1. Develop a new model with a price variable in all three choice alternatives. The price variable is created as:

   set price = ((distance/10)/mpg)*1.05

2. Calculate direct elasticities for all continuous variables using the Limdep "effects" command (see software command-file downloads for assignment #3). Briefly comment on your findings.

3. Perform a likelihood ratio test to determine if men and women should be modeled separately. The test statistic is (see page 282 in the text):

$$-2[LL(\beta_T) - LL(\beta_M) - LL(\beta_F)]$$

   where $LL(\beta_T)$ is the log-likelihood at convergence of the model estimated with the data (males and females), $LL(\beta_M)$ is the log-likelihood at convergence of the model using only male data (use Limdep commands: **reject;x11=0$**; then return to full sample with **include;x11=0$**), and $LL(\beta_F)$ is the log-likelihood at convergence of the model using only female data (**reject;x11=1$**; **include;x11=0$**). This statistic is $\chi^2$ distributed with degrees of freedom equal to the summation of the number of estimated parameters in individual male and female models minus the number of estimated parameters in the overall model. The resulting $\chi^2$ statistic provides the probability that the models have different parameters. Confidence levels for this can be read from Table C.3 on page 379 of the text. Briefly comment on your findings.

4. Using all data again (male and female), determine if the price variable should have separate parameters in the alternatives or if they should be the across alternatives. The test statistic is:

$$-2[LL(\beta_{same}) - LL(\beta_{dif})]$$

   where $LL(\beta_{same})$ is the log-likelihood at convergence of the model estimated with the same betas for price and $LL(\beta_{dif})$ is the log-likelihood at convergence using different betas for price. This statistic is $\chi^2$ distributed with degrees of freedom equal to the number of estimated parameters in the "different" model minus the number of estimated parameters in the "same" model.

Variables available for your specification are (in file LOGIT-A1.txt):

| Variable Number | Explanation |
| --- | --- |
| x1 | Route chosen, rows: 1 - arterial, 2 - rural road, 3 - freeway |
| x2 | Arterial row indicator; 1 for arterial row, 0 for others |
| x3 | Rural row indicator; 1 for rural row, 0 for others |
| x4 | Freeway row indicator; 1 for freeway row, 0 for others |
| x5 | Traffic flow rate |
| x6 | Number of traffic signals |
| x7 | Distance in tenths of miles |
| x8 | Seat belts: 1 - if wear, 0 - if not |
| x9 | Number of passengers in car |
| x10 | Driver age in years: 1 - 18 to 23, 2 - 24 to 29, 3 - 30 to 39, 4 - 40 to 49, 5 - 50 and above |
| x11 | Gender: 1 - male, 0 - female |
| x12 | Marital status: 1 - single, 0 - married |
| x13 | Number of children |
| x14 | Annual income: 1 - less than 20000, 2 - 20000 to 29999, 3 - 30000 to 39999, 4 - 40000 to 49999, 5 - more than 50000 |
| x15 | Model year of car (e.g. 86 = 1986) |
| x16 | Origin of car: 1 - domestic, 0 - foreign |
| x17 | Fuel efficiency in miles per gallon |

**Initial model run with price included as a variable:**

```
|-> read;nvar=17;nobs=453;file=C:\Users\cbr95\Box\UDrive Data_flm\00Work-
Purdue\new_laptop\CE697N-disk\LOGIT-A1.txt$
|-> create;cage=86-x15$
|-> create;price=(x7/10)/x17*1.05$
|-> nlogit;lhs=x1;choices=arterial,rural,freeway;model:
    u(arterial)=pricea*price/
    u(rural)=rural*one+pricer*price+cager*cage/
    u(freeway)=freeway*one+pricef*price+cagef*cage
    ;effects:price(arterial,rural,freeway)$
Iterative procedure has converged
Normal exit:  6 iterations. Status=0, F=    .9336640D+02
```

```
-----------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function       -93.36640
Estimation based on N =     151, K =   7
Inf.Cr.AIC  =    200.7 AIC/N =    1.329
----------------------------------------
          Log likelihood R-sqrd R2Adj
Constants only  -124.2267 .2484 .2306
Note: R-sqrd = 1 - logL/Logl(constants)
Root Likelihood:Geom. Mean of P^  .5388
----------------------------------------
Chi-squared[ 5](LR test) =    61.72054
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Number of obs.=   151, skipped    0 obs
--------+--------------------------------------------------------------------
        |                    Standard          Prob.      95% Confidence
     X1|  Coefficient        Error      z    |z|>Z*         Interval
--------+--------------------------------------------------------------------
  PRICEA|   -27.7121***      5.95637   -4.65   .0000    -39.3864  -16.0378
   RURAL|     1.88866**       .95839    1.97   .0488      .01025   3.76706
  PRICER|   -35.9203***      5.94185   -6.05   .0000    -47.5661  -24.2745
   CAGER|      .20382**       .07960    2.56   .0104      .04782    .35982
 FREEWAY|    -2.49579*       1.38370   -1.80   .0713     -5.20779    .21622
  PRICEF|   -21.1303***      5.83513   -3.62   .0003    -32.5669   -9.6936
   CAGEF|      .24885**       .09769    2.55   .0109      .05738    .44031
--------+--------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Jun 12, 2023 at 11:11:26 AM
-----------------------------------------------------------------------------
```

```
Elasticity wrt change of X in row choice on Prob[column choice]
--------+-------------------------
PRICE  |ARTERIAL   RURAL  FREEWAY
--------+-------------------------
ARTERIAL| -6.0598   1.5228   1.5228
   RURAL|  5.5066  -3.1633   5.5066
 FREEWAY|   .8192    .8192  -6.2715
```

Special note: For part 3 of this assignment, estimating separate models for men and women is **EXACTLY** the same as estimating one model for everyone but defining all variables for men and women separately. To see this:

The Male model is:

```
|-> Reject;x11=0$
|-> nlogit;lhs=x1;choices=arterial,rural,freeway;model:
    u(arterial)=pricea*price/
    u(rural)=rural*one+pricer*price+cager*cage/
    u(freeway)=freeway*one+pricef*price+cagef*cage
    ;effects:price(arterial,rural,freeway)$
Iterative procedure has converged
Normal exit:  6 iterations. Status=0, F=    .5302549D+02


-------------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable              Choice
Log likelihood function       -53.02549
Estimation based on N =    86, K =   7
Inf.Cr.AIC  =    120.1 AIC/N =    1.396
-----------------------------------------
          Log likelihood R-sqrd R2Adj
Constants only    -73.6498  .2800 .2495
Note: R-sqrd = 1 - logL/Logl(constants)
Root Likelihood:Geom. Mean of P^   .5398
-----------------------------------------
Chi-squared[ 5](LR test) =    41.24854
Prob [ chi squared > value ] =   .00000
Response data are given as ind. choices
Number of obs.=    86, skipped   0 obs
--------+----------------------------------------------------------------------
        |                   Standard              Prob.      95% Confidence
     X1| Coefficient        Error        z     |z|>Z*         Interval
--------+----------------------------------------------------------------------
  PRICEA|   -26.2957***      7.51885    -3.50   .0005    -41.0324   -11.5591
  RURAL|     2.34518*        1.27793     1.84   .0665     -.15951    4.84987
  PRICER|   -37.0939***      7.56657    -4.90   .0000    -51.9241   -22.2637
  CAGER|      .25373**        .10540     2.41   .0161      .04716     .46030
 FREEWAY|    -1.88024        1.71831    -1.09   .2739     -5.24807    1.48759
  PRICEF|   -20.5638***      7.60621    -2.70   .0069    -35.4717    -5.6559
  CAGEF|      .20983          .12850     1.63   .1025     -.04202     .46169
--------+----------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Jun 12, 2023 at 10:20:40 AM
-------------------------------------------------------------------------------

Elasticity wrt change of X in row choice on Prob[column choice]
--------+-------------------------
PRICE   |ARTERIAL   RURAL  FREEWAY
--------+-------------------------
ARTERIAL| -5.6845   1.5532   1.5532
   RURAL|  5.4454  -3.6727   5.4454
 FREEWAY|   .9382    .9382  -6.1092
```

The female model is:

```
|-> Include;x11=0$
|-> reject;x11=1$
|-> nlogit;lhs=x1;choices=arterial,rural,freeway;model:
    u(arterial)=pricea*price/
    u(rural)=rural*one+pricer*price+cager*cage/
    u(freeway)=freeway*one+pricef*price+cagef*cage
    ;effects:price(arterial,rural,freeway)$
Iterative procedure has converged
Normal exit:  7 iterations. Status=0, F=     .3884957D+02

-------------------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable               Choice
Log likelihood function       -38.84957
Estimation based on N =      65, K =   7
Inf.Cr.AIC  =       91.7 AIC/N =    1.411
-----------------------------------------
          Log likelihood R-sqrd R2Adj
Constants only    -50.2237  .2265 .1824
Note: R-sqrd = 1 - logL/Logl(constants)
Root Likelihood:Geom. Mean of P^   .5501
-----------------------------------------
Chi-squared[ 5](LR test) =    22.74821
Prob [ chi squared > value ] =   .00038
Response data are given as ind. choices
Number of obs.=     65, skipped    0 obs
--------+----------------------------------------------------------------------
        |                  Standard              Prob.      95% Confidence
     X1| Coefficient        Error       z     |z|>Z*         Interval
--------+----------------------------------------------------------------------
  PRICEA|   -29.9334***     10.56536    -2.83   .0046   -50.6411    -9.2256
   RURAL|     1.11383        1.77941      .63   .5313    -2.37376    4.60141
  PRICER|   -33.8135***      9.95887    -3.40   .0007   -53.3325   -14.2945
   CAGER|      .12685         .12837      .99   .3231     -.12474     .37844
 FREEWAY|    -3.19389        2.69543    -1.18   .2360    -8.47684    2.08905
  PRICEF|   -22.6908**       9.95188    -2.28   .0226   -42.1961    -3.1855
   CAGEF|      .30265*        .16228     1.86   .0622     -.01541     .62072
--------+----------------------------------------------------------------------
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Jun 12, 2023 at 10:22:51 AM
-------------------------------------------------------------------------------

Elasticity wrt change of X in row choice on Prob[column choice]
--------+--------------------------
PRICE   |ARTERIAL   RURAL  FREEWAY
--------+--------------------------
ARTERIAL| -6.6443   1.4819   1.4819
   RURAL|  5.4744  -2.4882   5.4744
 FREEWAY|   .6740    .6740  -6.7261
```

The full model (all observations) with separately defined parameters for males and females is (note that the results are **EXACTLY** the same as for the separate models):

```
|-> Include;x11=1$
|-> create;if(x11=1)mprice=price$
|-> create;if(x11=0)fprice=price$
|-> create;if(x11=1)mcage=cage$
|-> create;if(x11=0)fcage=cage$
|-> create;if(x11=1)male=1$
|-> create;if(x11=0)female=1$
|-> nlogit;lhs=x1;choices=arterial,rural,freeway;model:
    u(arterial)=priceam*mprice+priceaf*fprice/
u(rural)=ruralm*male+ruralf*female+pricerm*mprice+pricerf*fprice+cagerm*mcage+cage
rf*fcage/
u(freeway)=freewaym*male+freewayf*female+pricefm*mprice+priceff*fprice+cagefm*mcag
e+cagerff*fcage$
Iterative procedure has converged
Normal exit:  7 iterations. Status=0, F=    .9187505D+02


-----------------------------------------------------------------------
Discrete choice (multinomial logit) model
Dependent variable              Choice
Log likelihood function       -91.87505
Estimation based on N =    151, K =  14
Inf.Cr.AIC  =    211.8 AIC/N =    1.402
-----------------------------------------
          Log likelihood R-sqrd R2Adj
Constants only   -124.2267  .2604 .2245
Note: R-sqrd = 1 - logL/Logl(constants)
Warning:  Model does not contain a full
set of ASCs. R-sqrd is problematic. Use
model setup with ;RHS=one to get LogL0.
Root Likelihood:Geom. Mean of P^  .5442
-----------------------------------------
Response data are given as ind. choices
Number of obs.=   151, skipped    0 obs
```

| X1 | Coefficient | Standard Error | z | Prob. \|z\|>Z* | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| PRICEAM | -26.2957*** | 7.51885 | -3.50 | .0005 | -41.0324 | -11.5591 |
| PRICEAF | -29.9334*** | 10.56536 | -2.83 | .0046 | -50.6411 | -9.2256 |
| RURALM | 2.34518* | 1.27793 | 1.84 | .0665 | -.15951 | 4.84987 |
| RURALF | 1.11383 | 1.77941 | .63 | .5313 | -2.37376 | 4.60141 |
| PRICERM | -37.0939*** | 7.56657 | -4.90 | .0000 | -51.9241 | -22.2637 |
| PRICERF | -33.8135*** | 9.95887 | -3.40 | .0007 | -53.3325 | -14.2945 |
| CAGERM | .25373** | .10540 | 2.41 | .0161 | .04716 | .46030 |
| CAGERF | .12685 | .12837 | .99 | .3231 | -.12474 | .37844 |
| FREEWAYM | -1.88024 | 1.71831 | -1.09 | .2739 | -5.24807 | 1.48759 |
| FREEWAYF | -3.19389 | 2.69543 | -1.18 | .2360 | -8.47684 | 2.08905 |
| PRICEFM | -20.5638*** | 7.60621 | -2.70 | .0069 | -35.4717 | -5.6559 |
| PRICEFF | -22.6908** | 9.95188 | -2.28 | .0226 | -42.1961 | -3.1855 |
| CAGEFM | .20983 | .12850 | 1.63 | .1025 | -.04202 | .46169 |
| CAGERFF | .30265* | .16228 | 1.86 | .0622 | -.01541 | .62072 |

```
***, **, * ==>  Significance at 1%, 5%, 10% level.
Model was estimated on Jun 12, 2023 at 11:11:26 AM
-----------------------------------------------------------------------
```