```
                                                   ___  ____  ____  ____  ____(R)
                                                  /__    /   ____/   /   ____/
                                                 ___/   /   /___/   /   /___/
                                                   Statistics/Data analysis

         mpg |  -49.51222   86.15604    -0.57   0.567    -221.3025     122.278
      name:  <unnamed>
       log:  H:\My Drive\Econ 640\Homework 4\Homework 4 Log.smcl
  log type:  smcl
 opened on:   1 Dec 2025, 10:52:41

 1 .
 2 . *Part A*
 3 .
 4 . sysuse auto
   (1978 automobile data)

 5 . reg price weight mpg

         Source |       SS           df       MS      Number of obs   =        74
   -------------+----------------------------------   F(2, 71)        =     14.74
          Model |   186321280         2  93160639.9   Prob > F        =    0.0000
       Residual |   448744116        71  6320339.67   R-squared       =    0.2934
   -------------+----------------------------------   Adj R-squared   =    0.2735
          Total |   635065396        73  8699525.97   Root MSE        =      2514

   ------------------------------------------------------------------------------
          price | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
   -------------+----------------------------------------------------------------
         weight |   1.746559   .6413538     2.72   0.008      .467736    3.025382
            mpg |  -49.51222   86.15604    -0.57   0.567    -221.3025     122.278
          _cons |   1946.069    3597.05     0.54   0.590    -5226.245    9118.382
   ------------------------------------------------------------------------------

 6 . estat hettest

   Breusch-Pagan/Cook-Weisberg test for heteroskedasticity
   Assumption: Normal error terms
   Variable: Fitted values of price

   H0: Constant variance

       chi2(1) =  14.78
   Prob > chi2 = 0.0001

 7 .
 8 . /* Since the p-value is 0.0001, we would reject the null hypothesis that the
   > variance is homoskedastic. This implies that we have are dealing with
   > heteroskedasticity */
 9 .
10 . regress price weight mpg, robust

   Linear regression                               Number of obs     =         74
                                                   F(2, 71)          =      14.84
                                                   Prob > F          =     0.0000
                                                   R-squared         =     0.2934
                                                   Root MSE          =       2514

   ------------------------------------------------------------------------------
                |               Robust
          price | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
   -------------+----------------------------------------------------------------
         weight |   1.746559    .777837     2.25   0.028     .1955963    3.297522
            mpg |  -49.51222    95.8074    -0.52   0.607    -240.5468    141.5223
          _cons |   1946.069   4213.793     0.46   0.646    -6455.995    10348.13
   ------------------------------------------------------------------------------
```

```
11 .
12 . /* The standard errors of the robust regression are larger than just the regular
   > regression. By using robust, we relax the assumption that variance of the errors
   > is constant across all observations. This will lead to higher uncertainty,
   > making the robsut standard errors larger */
13 .
14 . *Part B*
15 .
16 . sysuse nlsw88, clear
   (NLSW, 1988 extract)

17 . reg wage age collgrad

         Source |       SS           df       MS      Number of obs   =     2,246
   -------------+----------------------------------   F(2, 2243)      =     87.71
          Model |  5394.00567         2  2697.00283   Prob > F        =    0.0000
       Residual |  68973.9617     2,243  30.7507631   R-squared       =    0.0725
   -------------+----------------------------------   Adj R-squared   =    0.0717
          Total |  74367.9674     2,245  33.1260434   Root MSE        =    5.5453

   ------------------------------------------------------------------------------
           wage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
   -------------+----------------------------------------------------------------
            age |  -.0643324   .0382481    -1.68   0.093    -.1393378     .010673
       collgrad |   3.612098   .2752221    13.12   0.000     3.072381    4.151815
          _cons |   9.430184   1.503989     6.27   0.000     6.480828    12.37954
   ------------------------------------------------------------------------------

18 . reg wage age collgrad, cluster(occupation)

   Linear regression                               Number of obs     =      2,237
                                                   F(2, 12)          =      14.91
                                                   Prob > F          =     0.0006
                                                   R-squared         =     0.0733
                                                   Root MSE          =     5.5487

                                (Std. err. adjusted for 13 clusters in occupation)
   ------------------------------------------------------------------------------
                |               Robust
           wage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
   -------------+----------------------------------------------------------------
            age |  -.0669641   .0431162    -1.55   0.146    -.1609062    .0269781
       collgrad |   3.632842   .6751054     5.38   0.000     2.161914     5.10377
          _cons |   9.539055   2.092498     4.56   0.001     4.979894    14.09822
   ------------------------------------------------------------------------------

19 .
20 . /* The cluster-robust standard errors are a lot larger than the OLS standard
   > errors. This is because they allow for correlation within each cluster which
   > causes the variance to increase as a result of the dependency. */
21 .
22 . *Part C*
23 .
24 . bcuse cps91, clear

   Contains data from http://fmwww.bc.edu/ec-p/data/wooldridge/cps91.dta
    Observations:         5,634
       Variables:            24                  20 May 2002 11:05
   -----------------------------------------------------------------------------------
   Variable      Storage   Display    Value
       name         type    format    label      Variable label
   -----------------------------------------------------------------------------------
   husage          byte    %8.0g                 husband's age
   husunion        byte    %8.0g                 =1 if hus. in union
   husearns        int     %8.0g                 hus. weekly earns
   huseduc         byte    %8.0g                 husband's yrs schooling
   husblck         byte    %8.0g                 =1 if hus. black
```

```
hushisp         byte    %8.0g                 =1 if hus. hispanic
hushrs          byte    %8.0g                 hus. weekly hours
kidge6          byte    %8.0g                 =1 if have child >= 6
earns           float   %8.0g                 wife's weekly earnings
age             byte    %8.0g                 wife's age
black           byte    %8.0g                 =1 if wife black
educ            byte    %8.0g                 wife's yrs schooling
hispanic        byte    %8.0g                 =1 if wife hispanic
union           byte    %8.0g                 =1 if wife in union
faminc          float   %9.0g                 annual family income
husexp          byte    %8.0g                 huseduc - husage - 6
exper           byte    %8.0g                 age - educ - 6
kidlt6          byte    %8.0g                 =1 if have child < 6
hours           int     %9.0g                 wife's weekly hours
expersq         int     %8.0g                 exper^2
nwifeinc        float   %9.0g                 non-wife inc, $1000s
inlf            byte    %8.0g                 =1 if wife in labor force
hrwage          float   %9.0g                 earns/hours
lwage           float   %9.0g                 log(hrwage)
-------------------------------------------------------------------------------------
Sorted by:

25 . gen lnexper = ln(exper)
   (20 missing values generated)

26 . reg hrwage lnexper
```

Source	SS	df	MS
Model	2.93774269	1	2.93774269
Residual	162316.271	3,274	49.5773582
Total	162319.209	3,275	49.5631171

Number of obs	=	3,276
F(1, 3274)	=	0.06
Prob > F	=	0.8077
R-squared	=	0.0000
Adj R-squared	=	-0.0003
Root MSE	=	7.0411

hrwage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lnexper	.0459342	.1886995	0.24	0.808	-.3240469	.4159152
_cons	10.25202	.5408182	18.96	0.000	9.191647	11.3124

```
27 .
28 . /* The coefficient on lnexper is 0.0459. This means that a 1% change in
   > experience is associated with a change in hourly wages of 0.000459. */
29 .
30 . reg hrwage exper
```

Source	SS	df	MS
Model	109.6416	1	109.6416
Residual	162457.898	3,284	49.4695184
Total	162567.54	3,285	49.4878356

Number of obs	=	3,286
F(1, 3284)	=	2.22
Prob > F	=	0.1367
R-squared	=	0.0007
Adj R-squared	=	0.0004
Root MSE	=	7.0335

hrwage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
exper	-.0184498	.0123929	-1.49	0.137	-.0427484	.0058488
_cons	10.72282	.268537	39.93	0.000	10.19631	11.24934

```
31 .
32 . /* If we were to regress without logging experience then the coefficient would
   > be -0.0184. This would mean that a one unit-change in experince would be
   > associated with a -0.0184 change in horuly wages. However, this implies that
   > increasing experience would diminish wages. By logging experience, the
   > diminishing effects of experience is accounted for */
33 .
34 . *Part D*
35 .
36 . bcuse bwght2, clear

Contains data from http://fmwww.bc.edu/ec-p/data/wooldridge/bwght2.dta
 Observations:         1,832
    Variables:            23                  24 May 2002 08:47
-------------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------------
mage            byte    %10.0g                mother's age, years
meduc           byte    %10.0g                mother's educ, years
monpre          byte    %10.0g                month prenatal care began
npvis           byte    %10.0g                total number of prenatal visits
fage            byte    %10.0g                father's age, years
feduc           byte    %10.0g                father's educ, years
bwght           int     %10.0g                birth weight, grams
omaps           byte    %10.0g                one minute apgar score
fmaps           byte    %10.0g                five minute apgar score
cigs            byte    %10.0g                avg cigarettes per day
drink           byte    %10.0g                avg drinks per week
lbw             byte    %9.0g                 =1 if bwght <= 2000
vlbw            byte    %9.0g                 =1 if bwght <= 1500
male            byte    %9.0g                 =1 if baby male
mwhte           byte    %9.0g                 =1 if mother white
mblck           byte    %9.0g                 =1 if mother black
moth            byte    %9.0g                 =1 if mother is other
fwhte           byte    %9.0g                 =1 if father white
fblck           byte    %9.0g                 =1 if father black
foth            byte    %9.0g                 =1 if father is other
lbwght          float   %9.0g                 log(bwght)
magesq          int     %9.0g                 mage^2
npvissq         int     %9.0g                 npvis^2
-------------------------------------------------------------------------------------
Sorted by:

37 . logit lbw cigs mage

Iteration 0:  Log likelihood =  -134.8245
Iteration 1:  Log likelihood = -133.21762
Iteration 2:  Log likelihood = -132.74617
Iteration 3:  Log likelihood = -132.74034
Iteration 4:  Log likelihood = -132.74034

Logistic regression                                     Number of obs =  1,722
                                                        LR chi2(2)    =   4.17
                                                        Prob > chi2   = 0.1244
Log likelihood = -132.74034                             Pseudo R2     = 0.0155

------------------------------------------------------------------------------
         lbw | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        cigs |    .041738   .0315485     1.32   0.186     -.020096     .103572
        mage |   -.067341   .0422063    -1.60   0.111    -.1500638    .0153818
       _cons |  -2.307042    1.20524    -1.91   0.056    -4.669269     .055186
------------------------------------------------------------------------------
```

```
38 . margins, dydx(cigs)

   Average marginal effects                                Number of obs = 1,722
   Model VCE: OIM

   Expression: Pr(lbw), predict()
   dy/dx wrt:  cigs
```

	dy/dx	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
cigs	.0006189	.0004809	1.29	0.198	-.0003236	.0015613

```
39 .
40 . /* The marginal effect is about 0.0006. This means that a one-unit change in
   > average cigarettes smoked during pregnancy can affect the probability that a
   > baby will be born with low weight by 0.0006
   >
   > For an OLS regression, the coefficient would describe the direct change in the
   > dependent variable if there was a one-unit change in the predictor variable.
   > However, the coefficient of a logit model would describe the log probability of
   > an outcome. The reason why logit coefficients are less straightforward is
   > because the effect relies on the base probability of an event happening despite
   > the coefficients staying the same
   > */
41 .
42 . log close
         name:  <unnamed>
          log:  H:\My Drive\Econ 640\Homework 4\Homework 4 Log.smcl
     log type:  smcl
    closed on:   1 Dec 2025, 10:52:46
```