```
------------------------------------------------------------------------------
      name:  <unnamed>
       log:  /Users/prakritishakya/Documents/Stata/pset1.smcl
  log type:  smcl
 opened on:   9 Nov 2023, 21:29:10

 1 . do "/var/folders/pl/1nqqwb294v9fxc8f33xjhfbw0000gn/T//SD90193.000000"

 2 . /*
   > Title: Causal Methods Problem Set 1
   > Date: 11/02/2023
   > */
 3 .
 4 . /* QUESTION 1
   >
   > A. Y = beta0 + beta1 X + beta2 A
   > B. Y = beta0 + beta1 X + beta2 A + beta3 B + beta4 C
   > C. Y = beta0 + beta1 X + beta2 A
   > D. Y = beta0 + beta1 X + beta2 A + beta3 B
   > E. Y = beta0 + beta1 X + beta2 C + beta3 F
   >
   > */
 5 .
 6 . clear

 7 . cls

 8 . set more off

 9 .
10 . * QUESTION 2
11 .
12 . * QUESTION 2. A)
13 .
```

```
14 . set obs 10
   Number of observations (_N) was 0, now 10.

15 .
16 . gen consp = 70 in 1
   (9 missing values generated)

17 . replace consp = 65 in 2
   (1 real change made)

18 . replace consp = 90 in 3
   (1 real change made)

19 . replace consp = 95 in 4
   (1 real change made)

20 . replace consp = 110 in 5
   (1 real change made)

21 . replace consp = 115 in 6
   (1 real change made)

22 . replace consp = 80 in 7
   (1 real change made)

23 . replace consp = 200 in 8
   (1 real change made)

24 . replace consp = 190 in 9
   (1 real change made)

25 . replace consp = 100 in 10
   (1 real change made)
```

```
26 .
27 . gen inc = 80 in 1
   (9 missing values generated)

28 . replace inc = 100 in 2
   (1 real change made)

29 . replace inc = 120 in 3
   (1 real change made)

30 . replace inc = 140 in 4
   (1 real change made)

31 . replace inc = 160 in 5
   (1 real change made)

32 . replace inc = 180 in 6
   (1 real change made)

33 . replace inc = 200 in 7
   (1 real change made)

34 . replace inc = 220 in 8
   (1 real change made)

35 . replace inc = 240 in 9
   (1 real change made)

36 . replace inc = 260 in 10
   (1 real change made)

37 .
38 . * regressing consumption on income
39 .
```

```
40 . reg consp inc

      Source |       SS           df       MS      Number of obs   =        10
-------------+----------------------------------   F(1, 8)         =      5.80
       Model |  8300.07576         1  8300.07576   Prob > F        =    0.0426
    Residual |  11452.4242         8  1431.55303   R-squared       =    0.4202
-------------+----------------------------------   Adj R-squared   =    0.3477
       Total |     19752.5         9  2194.72222   Root MSE        =    37.836

------------------------------------------------------------------------------
       consp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         inc |   .5015152   .2082796     2.41   0.043     .0212215    .9818088
       _cons |   26.24242   37.37444     0.70   0.503    -59.94319     112.428
------------------------------------------------------------------------------

41 .
42 . * We find that a unit increase in income increases consumption by 0.50 units
   > .
43 .
44 . * QUESTION 2. B)
45 .
46 . * creating column vector of 1's
47 . matrix ones = J(10,1,1)

48 . matrix list ones

   ones[10,1]
        c1
    r1   1
    r2   1
    r3   1
    r4   1
    r5   1
    r6   1
    r7   1
    r8   1
    r9   1
   r10   1
```

```
49 .
50 . * creating matrix of y
51 . mkmat consp, matrix (y)

52 . matrix list y

   y[10,1]
        consp
    r1     70
    r2     65
    r3     90
    r4     95
    r5    110
    r6    115
    r7     80
    r8    200
    r9    190
   r10    100

53 .
54 . * creating matrix of X
55 . mkmat inc, matrix (X)

56 . matrix list X

   X[10,1]
        inc
    r1   80
    r2  100
    r3  120
    r4  140
    r5  160
    r6  180
    r7  200
    r8  220
    r9  240
   r10  260
```

```
57 .
58 . * joining column vector of 1's and matrix X
59 . matrix X = ones, X

60 . matrix list X

   X[10,2]
         c1  inc
    r1    1   80
    r2    1  100
    r3    1  120
    r4    1  140
    r5    1  160
    r6    1  180
    r7    1  200
    r8    1  220
    r9    1  240
   r10    1  260

61 .
62 . * finding beta
63 . matrix beta = ( invsym(X' * X) ) * ( X' * y)

64 . matrix list beta

   beta[2,1]
            consp
    c1  26.242424
   inc  .50151515

65 .
66 . * We find that the beta on income is 0.50 which is the same as the results f
   > rom the previous regression that is a unit increase in income increases cons
   > umption by 0.50 units.
```

```
. 
. * QUESTION 2. C)
. 
. * regress consumption on income
. reg consp inc

      Source |       SS           df       MS      Number of obs   =        10
-------------+----------------------------------   F(1, 8)         =      5.80
       Model |  8300.07576         1  8300.07576   Prob > F        =    0.0426
    Residual |  11452.4242         8  1431.55303   R-squared       =    0.4202
-------------+----------------------------------   Adj R-squared   =    0.3477
       Total |     19752.5         9  2194.72222   Root MSE        =    37.836

------------------------------------------------------------------------------
       consp | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         inc |   .5015152   .2082796     2.41   0.043     .0212215    .9818088
       _cons |   26.24242   37.37444     0.70   0.503    -59.94319     112.428
------------------------------------------------------------------------------

. 
. * predict y_hat and residuals/error
. predict residuals, res

. 
. * call the residuals es
. ren residuals es

. 
. * turn es to a matrix
. mkmat es, matrix (es)

. matrix list es

es[10,1]
            es
 r1  3.6363637
 r2 -11.393939
 r3  3.5757575
 r4 -1.4545455
 r5  3.5151515
 r6 -1.5151515
 r7 -46.545456
 r8  63.424244
 r9   43.39394
r10 -56.636364
```

```
82 .
83 . * square the error terms
84 . gen es_squared = es * es

85 .
86 . * turn squared errors, es_squared to matrix and named it esqr
87 . mkmat es_squared, matrix (esqr)

88 . matrix list esqr

esqr[10,1]
     es_squared
 r1   13.223142
 r2   129.82185
 r3   12.786041
 r4   2.1157026
 r5    12.35629
 r6   2.2956841
 r7   2166.4795
 r8   4022.6348
 r9   1883.0341
r10   3207.6777

89 .
90 . * summarize the squared errors to find the sum
91 . sum es_squared

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
  es_squared |         10    1145.242    1549.616   2.115703   4022.635

92 .
93 . * save the sum in a new scalar
```

```
. scalar error_sum = r(sum)

. dis error_sum
11452.425

.
. * find the squared sigma
. scalar s_squared = error_sum/8

. dis s_squared
1431.5531

.
. * find the variance of beta
. matrix var_beta = s_squared * (invsym (X' * X))

. matrix list var_beta

symmetric var_beta[2,2]
             c1         inc
 c1   1396.8488
inc  -7.3746675    .0433804

.
. * find the standard error of the beta
. scalar stderror_beta = sqrt(var_beta[2,2])

. dis stderror_beta
.20827961

.
. * Therefore, the standard error of the beta is 0.2083.
.
. * QUESTION 2. D)
```

```
112 .
113 . * creating a matrix with squared error terms in diagonal with zeros elsewher
    > e using error term vector esqr (10x10)
114 . matrix mksq_error = diag(esqr)

115 . matrix list mksq_error

    symmetric mksq_error[10,10]
                r1          r2          r3          r4          r5          r6
    > r7          r8          r9         r10
     r1  13.223142
     r2          0   129.82185
     r3          0           0   12.786041
     r4          0           0           0   2.1157026
     r5          0           0           0           0    12.35629
     r6          0           0           0           0           0   2.2956841
     r7          0           0           0           0           0           0  2166.47
    > 95
     r8          0           0           0           0           0           0
    >  0  4022.6348
     r9          0           0           0           0           0           0
    >  0          0  1883.0341
    r10          0           0           0           0           0           0
    >  0          0          0  3207.6777

116 .
117 . * creating the inverse variance-covariance matrix (1x1)
118 . matrix vcm = invsym(X' * X)

119 . matrix list vcm

    symmetric vcm[2,2]
                 c1         inc
     c1   .97575758
    inc  -.00515152    .0000303
```

```
120 .
121 . * creating the "meat" with standard error squares matrix in the middle (1x1)
122 . matrix robust = X' * mksq_error * X

123 . matrix list robust

    symmetric robust[2,2]
                c1        inc
     c1  11452.425
    inc  2622460.7  6.087e+08

124 .
125 . * finding the variance
126 . matrix var = 1.25 * vcm * robust * vcm

127 . matrix list var

    symmetric var[2,2]
                 c1         inc
     c1   865.21963
    inc  -6.8063331   .05508958

128 .
129 . * taking the square root to find robust standard errors
130 . scalar robust_stderror = sqrt(var[1,1])

131 . dis robust_stderror
    29.414616

132 .
133 . * regress consumption on income, robust
134 . reg consp inc, robust

    Linear regression                               Number of obs     =         10
                                                    F(1, 8)           =       4.57
                                                    Prob > F          =     0.0651
                                                    R-squared         =     0.4202
                                                    Root MSE          =     37.836
```

conssp	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
inc	**.5015152**	**.2347117**	**2.14**	**0.065**	**-.039731**	**1.042761**
_cons	**26.24242**	**29.41462**	**0.89**	**0.398**	**-41.5878**	**94.07265**

```
. 
. * The robust standard error is 0.23.
. 
. * QUESTION 2. E) Clustered-standard errors
. 
. * assigning villages to each observation
. gen vil = 0

. 
. replace vil = 2 if _n == 1 | _n == 2 | _n == 5 | _n == 6 | _n == 8
(5 real changes made)

. 
. replace vil = 1 if _n == 3 | _n == 4 | _n == 7 | _n == 9 | _n == 10
(5 real changes made)

. 
. * for village 1, creating X matrix
. 
. mkmat inc if vil == 1, matrix (X1)

. matrix list X1

X1[5,1]
    inc
r1  120
r2  140
r3  200
r4  240
r5  260
```

```
151 .
152 . * joining column vector of 1's and matrix X1
153 . matrix ones = J(5,1,1)

154 . matrix X1 = ones, X1

155 . matrix list X1

    X1[5,2]
         c1  inc
    r1    1  120
    r2    1  140
    r3    1  200
    r4    1  240
    r5    1  260

156 .
157 . * creating matrix for error term for village 1
158 . mkmat es if vil == 1, matrix (es1)

159 . matrix list es1

    es1[5,1]
                es
    r1   3.5757575
    r2  -1.4545455
    r3  -46.545456
    r4    43.39394
    r5  -56.636364

160 .
161 . * "meat" for village 1
162 . matrix vil1 = X1' * (es1 * es1') * X1
```

```
163 .
164 . * for village 2, creating X matrix
165 .
166 . mkmat inc if vil == 2, matrix (X2)

167 . matrix list X2

    X2[5,1]
        inc
    r1   80
    r2  100
    r3  160
    r4  180
    r5  220

168 .
169 . * joining column vector of 1's and matrix X2
170 . matrix ones = J(5,1,1)

171 . matrix X2 = ones, X2

172 . matrix list X2

    X2[5,2]
         c1  inc
    r1    1   80
    r2    1  100
    r3    1  160
    r4    1  180
    r5    1  220

173 .
174 . * creating matrix for error term for village 2
175 . mkmat es if vil == 2, matrix (es2)
```

```
176 . matrix list es2

    es2[5,1]
                es
    r1   3.6363637
    r2  -11.393939
    r3   3.5151515
    r4  -1.5151515
    r5   63.424244

177 .
178 . * "meat" for village 2
179 . matrix vil2 = X2' * es2 * es2' * X2

180 .
181 . * "meat" sum
182 .
183 . matrix meat = vil1 + vil2

184 .
185 . * finding the variance
186 . matrix clvar = (invsym (X' * X)) * meat * (invsym (X' * X)) * 2.25

187 . matrix list clvar

    symmetric clvar[2,2]
                 c1         inc
     c1   729.64105
    inc  -6.2357401   .05329258

188 .
189 . * taking the square root to find robust standard errors
190 . scalar cl_stderror = sqrt(clvar[2,2])
```

```
191 . dis cl_stderror
    .23085186

192 .
193 . * regressing consumption on income and clustering standard errors at village
    >  level
194 . reg consp inc, cluster(vil)

    Linear regression                               Number of obs     =         10
                                                    F(0, 1)           =          .
                                                    Prob > F          =          .
                                                    R-squared         =     0.4202
                                                    Root MSE          =     37.836

                                        (Std. err. adjusted for 2 clusters in vil)
    ------------------------------------------------------------------------------
                 |               Robust
           consp | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
    -------------+----------------------------------------------------------------
             inc |   .5015152   .2308519     2.17   0.275    -2.431736    3.434766
           _cons |   26.24242   27.01187     0.97   0.509    -316.9759    369.4608
    ------------------------------------------------------------------------------

195 .
196 . * The clustered standard error is 0.23.
197 .
198 . * Question 3
199 .
200 . * Question 3.A)
201 .
202 . clear

203 . cls
```

```
204 . set more off

205 . 
206 . * call the dataset
207 . 
208 . use "/Users/prakritishakya/Desktop/Classes/Dataset/hh_98.dta"

209 . 
210 . * regressing log of total expenditures on female microcredit program
211 . reg lexptot progvillf

      Source |       SS           df       MS      Number of obs   =     1,129
-------------+----------------------------------   F(1, 1127)      =      4.03
       Model |  1.06259118         1  1.06259118   Prob > F        =    0.0448
    Residual |  296.797338     1,127  .263351676   R-squared       =    0.0036
-------------+----------------------------------   Adj R-squared   =    0.0027
       Total |   297.85993     1,128  .264060221   Root MSE        =    .51318

------------------------------------------------------------------------------
     lexptot | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   progvillf |   .1298466   .0646421     2.01   0.045     .0030142    .2566789
       _cons |   8.328525   .0626947   132.84   0.000     8.205513    8.451536
------------------------------------------------------------------------------

212 . 
213 . * The regression shows that a unit increase in female microcredit program in
    > creases total expenditure by 12.98 percent.
214 . 
215 . * regressing log of total expenditures on male microcredit program
216 . reg lexptot progvillm

      Source |       SS           df       MS      Number of obs   =     1,129
-------------+----------------------------------   F(1, 1127)      =      2.30
       Model |  .605673329         1  .605673329   Prob > F        =    0.1300
    Residual |  297.254256     1,127  .263757104   R-squared       =    0.0020
-------------+----------------------------------   Adj R-squared   =    0.0011
       Total |   297.85993     1,128  .264060221   Root MSE        =    .51357

------------------------------------------------------------------------------
     lexptot | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   progvillm |  -.0473609   .0312538    -1.52   0.130    -.1086832    .0139613
       _cons |   8.479275   .0242912   349.07   0.000     8.431614    8.526936
------------------------------------------------------------------------------
```

```
217 .
218 . * The regression shows that a unit increase in male microcredit program decr
    > eases total expenditure by 4.73 percent but the result is insignficant. The
    > sign on the coefficient is different for male and female program may be beca
    > use male may prioritize investment rather than spending on household needs.
    > However, due to lack of controls, the coefficients is biased upwards.
219 .
220 . * Question 3.B)
221 .
222 . * regressing log of total expenditures on female microcredit program with co
    > ntrols
223 . reg lexptot progvillf sexhead agehead educhead lnland vaccess pcirr rice whe
    > at milk oil egg
```

Source	SS	df	MS
Model	61.1882456	12	5.09902047
Residual	236.671684	1,116	.212071401
Total	297.85993	1,128	.264060221

Number of obs	=	1,129
F(12, 1116)	=	24.04
Prob > F	=	0.0000
R-squared	=	0.2054
Adj R-squared	=	0.1969
Root MSE	=	.46051

lexptot	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
progvillf	.1120142	.0587021	1.91	0.057	-.0031647	.2271932
sexhead	-.053949	.0481907	-1.12	0.263	-.1485035	.0406056
agehead	.003601	.0011206	3.21	0.001	.0014023	.0057997
educhead	.0481461	.0042581	11.31	0.000	.0397912	.056501
lnland	.1603209	.0293933	5.45	0.000	.1026485	.2179933
vaccess	-.0158758	.0385156	-0.41	0.680	-.091447	.0596954
pcirr	.1684416	.0466194	3.61	0.000	.07697	.2599131
rice	.0033335	.0091713	0.36	0.716	-.0146614	.0213284
wheat	-.039134	.01688	-2.32	0.021	-.0722541	-.0060139
milk	.0203634	.0056485	3.61	0.000	.0092806	.0314463
oil	.0108189	.0035749	3.03	0.003	.0038046	.0178332
egg	.1195542	.0486047	2.46	0.014	.0241872	.2149211
_cons	7.351443	.2286803	32.15	0.000	6.902751	7.800135

```
224 .
225 . * carry out white test for heteroskedasticity
226 . estat imtest, white

    White's test
    H0: Homoskedasticity
    Ha: Unrestricted heteroskedasticity

       chi2(86) = 155.47
    Prob > chi2 = 0.0000

    Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	155.47	86	0.0000
Skewness	37.66	12	0.0002
Kurtosis	9.73	1	0.0018
Total	202.86	99	0.0000

```
227 .
228 . * With controls, the coefficient on female microfinance program dropped down
    >  by 1.78 percent.
229 .
230 . reg lexptot progvillm sexhead agehead educhead lnland vaccess pcirr rice whe
    > at milk oil egg
```

Source	SS	df	MS
Model	61.4347274	12	5.11956061
Residual	236.425202	1,116	.21185054
Total	297.85993	1,128	.264060221

Number of obs	=	1,129
F(12, 1116)	=	24.17
Prob > F	=	0.0000
R-squared	=	0.2063
Adj R-squared	=	0.1977
Root MSE	=	.46027

lexptot	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
progvillm	-.0637731	.0290828	-2.19	0.029	-.1208363	-.0067099
sexhead	-.0586501	.0480846	-1.22	0.223	-.1529964	.0356963
agehead	.0033787	.0011201	3.02	0.003	.0011811	.0055764
educhead	.0482812	.0042534	11.35	0.000	.0399357	.0566268
lnland	.1643405	.0294918	5.57	0.000	.1064748	.2222062
vaccess	-.0269104	.0383294	-0.70	0.483	-.1021161	.0482953
pcirr	.1553286	.046994	3.31	0.001	.0631222	.2475351
rice	.0065337	.0092747	0.70	0.481	-.0116642	.0247316
wheat	-.0395366	.0168738	-2.34	0.019	-.0726445	-.0064286
milk	.0212032	.0056556	3.75	0.000	.0101064	.0322999
oil	.0105296	.0035735	2.95	0.003	.003518	.0175411
egg	.1066098	.0490947	2.17	0.030	.0102814	.2029382
_cons	7.522192	.2219774	33.89	0.000	7.086652	7.957732

```
231 .
232 . * carry out white test for heteroskedasticity
233 . estat imtest, white

White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

   chi2(87) = 150.57
Prob > chi2 = 0.0000

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	150.57	87	0.0000
Skewness	39.43	12	0.0001
Kurtosis	10.27	1	0.0013
Total	200.28	100	0.0000

```
234 .
235 . * With controls, the coefficient on female microfinance program dropped down
    >  by 1.65 percent.
236 .
237 . * Question 3.C)
238 .
239 . * For both regressions, since p < 0.05, we reject the null hypothesis that t
    > here is homoskedasticity.
240 .
241 . * re-running the regression with robust standard errors
242 . reg lexptot progvillf sexhead agehead educhead lnland vaccess pcirr rice whe
    > at milk oil egg, robust

    Linear regression                               Number of obs     =      1,129
                                                    F(12, 1116)       =      25.83
                                                    Prob > F          =     0.0000
                                                    R-squared         =     0.2054
                                                    Root MSE          =     .46051
```

lexptot	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
progvillf	.1120142	.0590257	1.90	0.058	-.0037995	.227828
sexhead	-.053949	.0565976	-0.95	0.341	-.1649987	.0571008
agehead	.003601	.0011178	3.22	0.001	.0014078	.0057942
educhead	.0481461	.0044472	10.83	0.000	.0394202	.056872
lnland	.1603209	.0325314	4.93	0.000	.0964913	.2241505
vaccess	-.0158758	.0403351	-0.39	0.694	-.095017	.0632654
pcirr	.1684416	.0480318	3.51	0.000	.0741988	.2626843
rice	.0033335	.0085033	0.39	0.695	-.0133509	.0200178
wheat	-.039134	.0168568	-2.32	0.020	-.0722086	-.0060594
milk	.0203634	.0056866	3.58	0.000	.0092058	.0315211
oil	.0108189	.0032319	3.35	0.001	.0044776	.0171602
egg	.1195542	.0495531	2.41	0.016	.0223265	.2167819
_cons	7.351443	.2314452	31.76	0.000	6.897326	7.80556

```
243 .
244 . reg lexptot progvillm sexhead agehead educhead lnland vaccess pcirr rice whe
    > at milk oil egg, robust

    Linear regression                               Number of obs     =      1,129
                                                    F(12, 1116)       =      25.08
                                                    Prob > F          =     0.0000
                                                    R-squared         =     0.2063
                                                    Root MSE          =     .46027

    ------------------------------------------------------------------------------
                 |               Robust
         lexptot | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
    -------------+----------------------------------------------------------------
       progvillm |  -.0637731   .0297973    -2.14   0.033    -.1222381   -.0053082
         sexhead |  -.0586501   .0562908    -1.04   0.298    -.1690978    .0517977
         agehead |   .0033787   .0011098     3.04   0.002     .0012011    .0055563
        educhead |   .0482812   .0044659    10.81   0.000     .0395188    .0570437
          lnland |   .1643405   .0324936     5.06   0.000     .1005852    .2280959
         vaccess |  -.0269104   .0402121    -0.67   0.503    -.1058101    .0519894
           pcirr |   .1553286   .0486838     3.19   0.001     .0598065    .2508508
            rice |   .0065337   .0086591     0.75   0.451    -.0104563    .0235238
           wheat |  -.0395366   .0169433    -2.33   0.020    -.0727808   -.0062923
            milk |   .0212032   .0056546     3.75   0.000     .0101084    .0322979
             oil |   .0105296   .0032401     3.25   0.001     .0041722    .0168869
             egg |   .1066098   .0491095     2.17   0.030     .0102524    .2029671
           _cons |   7.522192   .2324683    32.36   0.000     7.066068    7.978316
    ------------------------------------------------------------------------------

245 .
246 . * We only find very small changes in the coefficients, standard error and th
    > e significance.
247 .
248 . * Question 3.D)
```

```
249 .
250 . * regressing total expenditure on the number of female participants in a hou
    > sehold with controls
251 . reg lexptot dfmfd sexhead agehead educhead lnland vaccess pcirr rice wheat m
    > ilk oil egg

      Source |       SS           df       MS      Number of obs   =     1,129
-------------+----------------------------------   F(12, 1116)     =     24.72
       Model |  62.5500044        12  5.21250037   Prob > F        =    0.0000
    Residual |  235.309925     1,116  .210851187   R-squared       =    0.2100
-------------+----------------------------------   Adj R-squared   =    0.2015
       Total |   297.85993     1,128  .264060221   Root MSE        =    .45919

------------------------------------------------------------------------------
     lexptot | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       dfmfd |    .090514    .028452     3.18   0.002     .0346886    .1463393
     sexhead |  -.0585578   .0479707    -1.22   0.222    -.1526807     .035565
     agehead |   .0033684   .0011168     3.02   0.003      .001177    .0055597
    educhead |    .049832    .004264    11.69   0.000     .0414656    .0581984
      lnland |   .1756242    .029798     5.89   0.000     .1171578    .2340905
     vaccess |  -.0177161    .038246    -0.46   0.643    -.0927584    .0573261
       pcirr |   .1604492   .0465577     3.45   0.001     .0690988    .2517996
        rice |   .0039976   .0091464     0.44   0.662    -.0139485    .0219437
       wheat |  -.0391291   .0168306    -2.32   0.020    -.0721524   -.0061059
        milk |   .0205496   .0056321     3.65   0.000      .009499    .0316003
         oil |   .0099927   .0035712     2.80   0.005     .0029856    .0169997
         egg |   .1164396   .0484775     2.40   0.016     .0213222    .2115569
       _cons |   7.450058   .2200453    33.86   0.000     7.018309    7.881807
------------------------------------------------------------------------------

252 .
253 . * The results show that a unit increase in the number of female microcredit
    > borrowers increases total expenditure by 9.05 percent.
```

```
254 .
255 . * regressing total expenditure on the number of male participants in a house
    > hold with controls
256 . reg lexptot dmmfd sexhead agehead educhead lnland vaccess pcirr rice wheat m
    > ilk oil egg
```

Source	SS	df	MS			
				Number of obs	=	1,129
				F(12, 1116)	=	23.71
Model	60.5057616	12	5.0421468	Prob > F	=	0.0000
Residual	237.354168	1,116	.212682946	R-squared	=	0.2031
				Adj R-squared	=	0.1946
Total	297.85993	1,128	.264060221	Root MSE	=	.46118

lexptot	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
dmmfd	-.0232267	.0357649	-0.65	0.516	-.0934007	.0469473
sexhead	-.056038	.048441	-1.16	0.248	-.1510837	.0390077
agehead	.0034573	.0011227	3.08	0.002	.0012544	.0056602
educhead	.0486288	.0042678	11.39	0.000	.040255	.0570027
lnland	.1567296	.0294925	5.31	0.000	.0988626	.2145967
vaccess	-.0232387	.0383709	-0.61	0.545	-.0985258	.0520485
pcirr	.1679713	.0467009	3.60	0.000	.0763398	.2596028
rice	.0041169	.0092444	0.45	0.656	-.0140216	.0222553
wheat	-.0381749	.0169267	-2.26	0.024	-.0713866	-.0049633
milk	.0206401	.0056634	3.64	0.000	.0095279	.0317523
oil	.0106221	.003582	2.97	0.003	.003594	.0176502
egg	.117315	.0493105	2.38	0.018	.0205633	.2140667
_cons	7.470394	.2210315	33.80	0.000	7.03671	7.904078

```
257 .
258 . * The results show that a unit increase in the number of male microcredit bo
    > rrowers decreases total expenditure by 2.32 percent.
```

```
259 .
260 . * Question 3.E)
261 .
262 . * running regression of total expenditure on the number of female participan
    > ts in a household with controls and cluster effects
263 . reg lexptot dfmfd sexhead agehead educhead lnland vaccess pcirr rice wheat m
    > ilk oil egg, cluster (villid)
```

Linear regression

Number of obs	=	**1,129**
F(2, 3)	=	.
Prob > F	=	.
R-squared	=	**0.2100**
Root MSE	=	**.45919**

(Std. err. adjusted for **4** clusters in **villid**)

lexptot	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
dfmfd	**.090514**	**.0214854**	**4.21**	**0.024**	**.0221378**	**.1588901**
sexhead	**-.0585578**	**.0395607**	**-1.48**	**0.235**	**-.1844576**	**.0673419**
agehead	**.0033684**	**.0006927**	**4.86**	**0.017**	**.0011639**	**.0055728**
educhead	**.049832**	**.0077921**	**6.40**	**0.008**	**.0250342**	**.0746298**
lnland	**.1756242**	**.0413978**	**4.24**	**0.024**	**.0438778**	**.3073705**
vaccess	**-.0177161**	**.045933**	**-0.39**	**0.725**	**-.1638955**	**.1284632**
pcirr	**.1604492**	**.076908**	**2.09**	**0.128**	**-.0843065**	**.4052049**
rice	**.0039976**	**.0150508**	**0.27**	**0.808**	**-.0439006**	**.0518959**
wheat	**-.0391291**	**.0327602**	**-1.19**	**0.318**	**-.1433869**	**.0651286**
milk	**.0205496**	**.0065372**	**3.14**	**0.052**	**-.0002546**	**.0413539**
oil	**.0099927**	**.0033873**	**2.95**	**0.060**	**-.0007874**	**.0207727**
egg	**.1164396**	**.0703207**	**1.66**	**0.196**	**-.1073523**	**.3402315**
_cons	**7.450058**	**.1766659**	**42.17**	**0.000**	**6.887828**	**8.012288**

```
264 .
265 . * Using clustered standard errors increases standard error and decreases the
    >  stastical significance and allows for significance at 1%.
266 .
267 . * running regression of total expenditure on the number of female participan
    > ts in a household with controls and cluster effects
268 . reg lexptot dmmfd sexhead agehead educhead lnland vaccess pcirr rice wheat m
    > ilk oil egg, cluster (villid)

Linear regression                               Number of obs     =      1,129
                                                F(2, 3)           =          .
                                                Prob > F          =          .
                                                R-squared         =     0.2031
                                                Root MSE          =     .46118

                                   (Std. err. adjusted for 4 clusters in villid)
```

lexptot	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
dmmfd	-.0232267	.0363101	-0.64	0.568	-.1387816	.0923283
sexhead	-.056038	.0486577	-1.15	0.333	-.2108884	.0988124
agehead	.0034573	.0005161	6.70	0.007	.0018148	.0050997
educhead	.0486288	.0077954	6.24	0.008	.0238203	.0734374
lnland	.1567296	.0413637	3.79	0.032	.0250919	.2883673
vaccess	-.0232387	.0434608	-0.53	0.630	-.1615504	.115073
pcirr	.1679713	.0775726	2.17	0.119	-.0788992	.4148418
rice	.0041169	.0154714	0.27	0.807	-.04512	.0533538
wheat	-.0381749	.033172	-1.15	0.333	-.143743	.0673932
milk	.0206401	.0066964	3.08	0.054	-.0006709	.0419511
oil	.0106221	.0033907	3.13	0.052	-.0001686	.0214128
egg	.117315	.0777603	1.51	0.229	-.1301531	.3647831
_cons	7.470394	.1708855	43.72	0.000	6.92656	8.014228

```
269 .
270 . * Using clustered standard errors decreases standard error but remains insig
    > nificant.
271 .
272 . * Question 3.F)
273 .
274 . * using bootstrap to estimate standard errors (non-clustered) for female par
    > ticipants
275 . bootstrap, reps(1000) seed(12345) : reg lexptot dfmfd sexhead agehead eduche
    > ad lnland vaccess pcirr rice wheat milk oil egg
    (running regress on estimation sample)

    Bootstrap replications (1,000): .........10.........20.........30.........40..
    > .......50.........60.........70.........80.........90.........100.........11
    > 0.........120.........130.........140.........150.........160.........170...
    > ......180.........190.........200.........210.........220.........230.......
    > ..240.........250.........260.........270.........280.........290.........30
    > 0.........310.........320.........330.........340.........350.........360...
    > ......370.........380.........390.........400.........410.........420.......
    > ..430.........440.........450.........460.........470.........480.........49
    > 0.........500.........510.........520.........530.........540.........550...
    > ......560.........570.........580.........590.........600.........610.......
    > ..620.........630.........640.........650.........660.........670.........68
    > 0.........690.........700.........710.........720.........730.........740...
    > ......750.........760.........770.........780.........790.........800.......
    > ..810.........820.........830.........840.........850.........860.........87
    > 0.........880.........890.........900.........910.........920.........930...
    > ......940.........950.........960.........970.........980.........990.......
    > ..1,000 done

    Linear regression                                       Number of obs =  1,129
                                                            Replications  =  1,000
                                                            Wald chi2(12) = 306.41
                                                            Prob > chi2   = 0.0000
                                                            R-squared     = 0.2100
                                                            Adj R-squared = 0.2015
                                                            Root MSE      = 0.4592
```

lexptot	Observed coefficient	Bootstrap std. err.	z	P>|z|	Normal-based [95% conf.	interval]
dfmfd	.090514	.0276476	3.27	0.001	.0363257	.1447022
sexhead	-.0585578	.0561621	-1.04	0.297	-.1686336	.0515179
agehead	.0033684	.0011907	2.83	0.005	.0010347	.005702
educhead	.049832	.004522	11.02	0.000	.040969	.058695
lnland	.1756242	.0319791	5.49	0.000	.1129463	.238302
vaccess	-.0177161	.040942	-0.43	0.665	-.097961	.0625288
pcirr	.1604492	.0461503	3.48	0.001	.0699963	.2509021
rice	.0039976	.0084018	0.48	0.634	-.0124696	.0204648
wheat	-.0391291	.0172545	-2.27	0.023	-.0729473	-.005311
milk	.0205496	.0055856	3.68	0.000	.0096021	.0314971
oil	.0099927	.0031004	3.22	0.001	.0039159	.0160694
egg	.1164396	.0501807	2.32	0.020	.0180871	.214792
_cons	7.450058	.2244092	33.20	0.000	7.010224	7.889892

```
276 .
277 . * using bootstrap to estimate standard errors (non-clustered) for male parti
    > cipants
278 . bootstrap, reps(1000) seed(12345) : reg lexptot dmmfd sexhead agehead eduche
    > ad lnland vaccess pcirr rice wheat milk oil egg
    (running regress on estimation sample)

    Bootstrap replications (1,000): .........10.........20.........30.........40..
    > .......50.........60.........70.........80.........90.........100.........11
    > 0.........120.........130.........140.........150.........160.........170...
    > ......180.........190.........200.........210.........220.........230.......
    > ..240.........250.........260.........270.........280.........290.........30
    > 0.........310.........320.........330.........340.........350.........360...
    > ......370.........380.........390.........400.........410.........420.......
    > ..430.........440.........450.........460.........470.........480.........49
    > 0.........500.........510.........520.........530.........540.........550...
    > ......560.........570.........580.........590.........600.........610.......
    > ..620.........630.........640.........650.........660.........670.........68
    > 0.........690.........700.........710.........720.........730.........740...
    > ......750.........760.........770.........780.........790.........800.......
    > ..810.........820.........830.........840.........850.........860.........87
    > 0.........880.........890.........900.........910.........920.........930...
    > ......940.........950.........960.........970.........980.........990.......
    > ..1,000 done
```

```
Linear regression                                    Number of obs =  1,129
                                                     Replications  =  1,000
                                                     Wald chi2(12) = 305.42
                                                     Prob > chi2   = 0.0000
                                                     R-squared     = 0.2031
                                                     Adj R-squared = 0.1946
                                                     Root MSE      = 0.4612
```

lexptot	Observed coefficient	Bootstrap std. err.	z	P>|z|	Normal-based [95% conf.	interval]
dmmfd	-.0232267	.0335794	-0.69	0.489	-.0890412	.0425878
sexhead	-.056038	.0567134	-0.99	0.323	-.1671942	.0551182
agehead	.0034573	.0011936	2.90	0.004	.0011179	.0057966
educhead	.0486288	.0044331	10.97	0.000	.0399401	.0573175
lnland	.1567296	.0320375	4.89	0.000	.0939372	.219522
vaccess	-.0232387	.0407761	-0.57	0.569	-.1031583	.056681
pcirr	.1679713	.0462731	3.63	0.000	.0772778	.2586648
rice	.0041169	.0085356	0.48	0.630	-.0126126	.0208464
wheat	-.0381749	.0173242	-2.20	0.028	-.0721297	-.0042202
milk	.0206401	.0055928	3.69	0.000	.0096785	.0316017
oil	.0106221	.0031464	3.38	0.001	.0044554	.0167889
egg	.117315	.0512133	2.29	0.022	.0169389	.2176912
_cons	7.470394	.2264196	32.99	0.000	7.02662	7.914168

```
279 .
280 . * We find that our normal standard errors are overestimated and could cause
    > Type II error.
281 .
282 . * Question 3.G)
283 .
284 . * using bootstrap to estimate standard errors (clustered at village level)
```

```
285 .
286 . * using bootstrap to estimate standard errors (non-clustered) for female par
    > ticipants
287 . bootstrap, reps(1000) seed(12345) : reg lexptot dfmfd sexhead agehead eduche
    > ad lnland vaccess pcirr rice wheat milk oil egg, cluster (villid)
    (running regress on estimation sample)

    Bootstrap replications (1,000): .........10.........20.........30.........40..
    > .......50.........60.........70.........80.........90.........100.........11
    > 0.........120.........130.........140.........150.........160.........170...
    > ......180.........190.........200.........210.........220.........230.......
    > ..240.........250.........260.........270.........280.........290.........30
    > 0.........310.........320.........330.........340.........350.........360...
    > ......370.........380..x......390.........400.........410.........420.......
    > ..430.........440.........450.........460.........470.........480.........49
    > 0.........500.........510.........520.........530.........540.........550...
    > ......560.........570.........580.........590.........600.........610.......
    > ..620.........630.........640.........650.........660.........670.........68
    > 0.........690.........700.........710.........720.........730.........740...
    > ......750.........760.......x.770.........780.........790.........800.......
    > ..810.........820.........830.........840.........850......x..860.........87
    > 0.........880.........890.........900.........910.........920.........930...
    > ......940.........950.........960.........970.........980.........990.......
    > ..1,000 done
    x: Error occurred when bootstrap executed regress.

    Linear regression                                     Number of obs =     1,129
                                                          Replications  =       997
                                                          Wald chi2(12) = 279973.20
                                                          Prob > chi2   =    0.0000
                                                          R-squared     =    0.2100
                                                          Adj R-squared =    0.2015
                                                          Root MSE      =    0.4592

                                              (Replications based on 4 clusters in villid)
```

lexptot	Observed coefficient	Bootstrap std. err.	z	P>|z|	Normal-based [95% conf.	interval]
dfmfd	.090514	.0202646	4.47	0.000	.050796	.1302319
sexhead	-.0585578	.0377071	-1.55	0.120	-.1324624	.0153467
agehead	.0033684	.0006428	5.24	0.000	.0021086	.0046281
educhead	.049832	.0066704	7.47	0.000	.0367583	.0629057
lnland	.1756242	.0365149	4.81	0.000	.1040563	.247192
vaccess	-.0177161	.0420977	-0.42	0.674	-.100226	.0647938
pcirr	.1604492	.0740304	2.17	0.030	.0153523	.3055461

```
        rice |   .0039976   .0127595     0.31   0.754    -.0210105    .0290057
       wheat |  -.0391291   .0352425    -1.11   0.267    -.1082032    .0299449
        milk |   .0205496    .007009     2.93   0.003     .0068123     .034287
         oil |   .0099927   .0057062     1.75   0.080    -.0011913    .0211766
         egg |   .1164396   .0710737     1.64   0.101    -.0228622    .2557414
       _cons |   7.450058   .1777244    41.92   0.000     7.101725    7.798392
------------------------------------------------------------------------------
Note: One or more parameters could not be estimated in 3 bootstrap replicates;
      standard-error estimates include only complete replications.

288 .
289 . * using bootstrap to estimate standard errors (non-clustered) for male parti
> cipants
290 . bootstrap, reps(1000) seed(12345) : reg lexptot dmmfd sexhead agehead eduche
> ad lnland vaccess pcirr rice wheat milk oil egg, cluster (villid)
(running regress on estimation sample)

Bootstrap replications (1,000): .........10.........20.........30.........40..
> .......50.........60.........70.........80.........90.........100.........11
> 0.........120.........130.........140.........150.........160.........170...
> ......180.........190.........200.........210.........220.........230.......
> ..240.........250.........260.........270.........280.........290.........30
> 0.........310.........320.........330.........340.........350.........360...
> ......370.........380..x......390.........400.........410.........420.......
> ..430.........440.........450.........460.........470.........480.........49
> 0.........500.........510.........520.........530.........540.........550...
> ......560.........570.........580.........590.........600.........610.......
> ..620.........630.........640.........650.........660.........670.........68
> 0.........690.........700.........710.........720.........730.........740...
> ......750.........760.......x.770.........780.........790.........800.......
> ..810.........820.........830.........840.........850......x..860.........87
> 0.........880.........890.........900.........910.........920.........930...
> ......940.........950.........960.........970.........980.........990.......
> ..1,000 done
x: Error occurred when bootstrap executed regress.

Linear regression                                     Number of obs =    1,129
                                                      Replications  =      997
                                                      Wald chi2(12) = 202799.36
                                                      Prob > chi2   =    0.0000
                                                      R-squared     =    0.2031
                                                      Adj R-squared =    0.1946
                                                      Root MSE      =    0.4612
```

(Replications based on **4** clusters in **villid**)

lexptot	Observed coefficient	Bootstrap std. err.	z	P>|z|	Normal-based [95% conf.	interval]
dmmfd	**-.0232267**	**.0300179**	**-0.77**	**0.439**	**-.0820608**	**.0356074**
sexhead	**-.056038**	**.0471943**	**-1.19**	**0.235**	**-.1485371**	**.0364611**
agehead	**.0034573**	**.0004914**	**7.04**	**0.000**	**.0024941**	**.0044204**
educhead	**.0486288**	**.0066871**	**7.27**	**0.000**	**.0355223**	**.0617354**
lnland	**.1567296**	**.0363092**	**4.32**	**0.000**	**.0855649**	**.2278943**
vaccess	**-.0232387**	**.0395506**	**-0.59**	**0.557**	**-.1007564**	**.0542791**
pcirr	**.1679713**	**.0741204**	**2.27**	**0.023**	**.022698**	**.3132446**
rice	**.0041169**	**.0132949**	**0.31**	**0.757**	**-.0219406**	**.0301744**
wheat	**-.0381749**	**.0351544**	**-1.09**	**0.278**	**-.1070763**	**.0307265**
milk	**.0206401**	**.0070167**	**2.94**	**0.003**	**.0068876**	**.0343926**
oil	**.0106221**	**.0059566**	**1.78**	**0.075**	**-.0010525**	**.0222968**
egg	**.117315**	**.0749104**	**1.57**	**0.117**	**-.0295067**	**.2641368**
_cons	**7.470394**	**.169568**	**44.06**	**0.000**	**7.138047**	**7.802741**

Note: One or more parameters could not be estimated in 3 bootstrap replicates;
standard-error estimates include only complete replications.

```
291 .
292 . * We find that the normal standard errors are overestimated so this might ca
    > use Type II error.
293 .
294 . * Question 4
295 .
296 . * average treatment effects with regression adjustments
297 . teffects ra (lexptot sexhead agehead educhead lnland vaccess pcirr rice whea
    > t milk oil egg) (dfmfd)

    Iteration 0:  EE criterion =  7.218e-27
    Iteration 1:  EE criterion =  5.362e-31
```

```
Treatment-effects estimation                    Number of obs     =      1,129
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

lexptot	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
dfmfd						
(1 vs 0)	**.0842385**	**.0279233**	**3.02**	**0.003**	**.0295098**	**.1389672**
POmean						
dfmfd						
0	**8.397697**	**.0209225**	**401.37**	**0.000**	**8.35669**	**8.438704**

```
. 
. * Women are likely to spend 8.42 percent more on expenditure.
. 
. * treatment effect on the treated with regression adjustments
. teffects ra (lexptot sexhead agehead educhead lnland vaccess pcirr rice whea
> t milk oil egg) (dfmfd), atet

Iteration 0:  EE criterion =  7.217e-27
Iteration 1:  EE criterion =  2.416e-31

Treatment-effects estimation                    Number of obs     =      1,129
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

lexptot	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATET						
dfmfd						
(1 vs 0)	**.1005064**	**.0282498**	**3.56**	**0.000**	**.0451377**	**.1558751**
POmean						
dfmfd						
0	**8.352572**	**.0223002**	**374.55**	**0.000**	**8.308865**	**8.39628**

```
303 .
304 . * Those women who participated in the microfinance are most likely to spend
    > 10.05 percentage point more on expenditure.
305 .
306 . * average treatment effects with inverse probability weights (IPW)
307 . teffects ipw (lexptot)(dfmfd sexhead agehead educhead lnland vaccess pcirr r
    > ice wheat milk oil egg)

    Iteration 0:  EE criterion =  5.262e-18
    Iteration 1:  EE criterion =  1.075e-30

    Treatment-effects estimation                   Number of obs     =      1,129
    Estimator      : inverse-probability weights
    Outcome model  : weighted mean
    Treatment model: logit

    ------------------------------------------------------------------------------
                 |               Robust
         lexptot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
    -------------+----------------------------------------------------------------
    ATE          |
           dfmfd |
       (1 vs 0)  |   .0794929   .0276263     2.88   0.004     .0253464    .1336393
    -------------+----------------------------------------------------------------
    POmean       |
           dfmfd |
              0  |    8.39528   .0207766   404.07   0.000     8.354558    8.436001
    ------------------------------------------------------------------------------

308 .
309 . * treatment effect on the treated with inverse probability weights (IPW)
310 . teffects ipw (lexptot)(dfmfd sexhead agehead educhead lnland vaccess pcirr r
    > ice wheat milk oil egg), atet

    Iteration 0:  EE criterion =  5.262e-18
    Iteration 1:  EE criterion =  4.404e-31
```

```
Treatment-effects estimation                   Number of obs     =      1,129
Estimator      : inverse-probability weights
Outcome model  : weighted mean
Treatment model: logit
```

lexptot	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATET						
dfmfd						
(1 vs 0)	**.1053432**	**.0283003**	**3.72**	**0.000**	**.0498757**	**.1608107**
POmean						
dfmfd						
0	**8.347735**	**.022749**	**366.95**	**0.000**	**8.303148**	**8.392323**

```
311 .
312 . * The intuition behind both ra and ipw is to control for confounders to find
    >  the average treatment effect in observational studies where the treatment a
    > ssignment is not random.
313 .
314 . * how do the results between ra and ipw compare on the Khandker dataset? The
    >  results from ra and ipw are very similar and only point differences.
315 .
316 . * how are ATE and ATET estimates different? ATE and ATET estimates are diffe
    > rent in a way that the estimates for ATET are bigger since it is for the tre
    > atment effects on the treated.
317 .
318 . * Question 5
319 .
320 . * Question 5.A)
321 .
322 . * (i) propensity score matching for female microcredit borrowers
```

```
323 . capture drop notmatched

324 .
325 . teffects psmatch (lexptot) (dfmfd sexhead agehead educhead lnland vaccess pc
    > irr rice wheat milk oil egg), osample (notmatched)

    Treatment-effects estimation                   Number of obs      =      1,129
    Estimator      : propensity-score matching     Matches: requested =          1
    Outcome model  : matching                                     min =          1
    Treatment model: logit                                        max =          1
    ------------------------------------------------------------------------------
                 |              AI robust
         lexptot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
    -------------+----------------------------------------------------------------
    ATE          |
           dfmfd |
       (1 vs 0)  |   .0689483   .0352761     1.95   0.051    -.0001916    .1380882
    ------------------------------------------------------------------------------

326 .
327 . dis notmatched
    0

328 .
329 . * (ii) propensity score matching for male microcredit borrowers
330 . teffects psmatch (lexptot) (dmmfd sexhead agehead educhead lnland vaccess pc
    > irr rice wheat milk oil egg), osample (notmatched1)

    Treatment-effects estimation                   Number of obs      =      1,129
    Estimator      : propensity-score matching     Matches: requested =          1
    Outcome model  : matching                                     min =          1
    Treatment model: logit                                        max =          1
    ------------------------------------------------------------------------------
                 |              AI robust
         lexptot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
    -------------+----------------------------------------------------------------
    ATE          |
           dmmfd |
       (1 vs 0)  |  -.0251912   .0598302    -0.42   0.674    -.1424562    .0920738
    ------------------------------------------------------------------------------
```

```
331 .
332 . dis notmatched1
    0

333 .
334 . * None of the observation remain unmatched.
335 .
336 . * checking balances in covariates
337 . foreach i in sexhead agehead educhead lnland vaccess pcirr rice wheat milk o
    > il egg {
      2.         dis "`i'"
      3.         ttest `i', by(dfmfd)
      4. }
    sexhead

    Two-sample t test with equal variances
    ------------------------------------------------------------------------------
       Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
    ---------+--------------------------------------------------------------------
           0 |     534    .9157303    .0120325    .2780523    .8920934    .9393673
           1 |     595    .9008403     .012263    .2991277    .8767561    .9249245
    ---------+--------------------------------------------------------------------
    Combined |   1,129    .9078831    .0086105    .2893191    .8909886    .9247775
    ---------+--------------------------------------------------------------------
        diff |              .01489    .0172482               -.0189522    .0487322
    ------------------------------------------------------------------------------
        diff = mean(0) - mean(1)                                      t =   0.8633
    H0: diff = 0                                     Degrees of freedom =     1127

        Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
     Pr(T < t) = 0.8059         Pr(|T| > |t|) = 0.3882          Pr(T > t) = 0.1941
    agehead

    Two-sample t test with equal variances
    ------------------------------------------------------------------------------
       Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
    ---------+--------------------------------------------------------------------
           0 |     534    45.91948    .5895908    13.62452    44.76127    47.07768
           1 |     595     46.0958     .482778    11.77622    45.14764    47.04396
    ---------+--------------------------------------------------------------------
    Combined |   1,129     46.0124     .377334    12.67865    45.27204    46.75276
    ---------+--------------------------------------------------------------------
        diff |           -.1763227    .7560889               -1.659823    1.307178
    ------------------------------------------------------------------------------
        diff = mean(0) - mean(1)                                      t =  -0.2332
    H0: diff = 0                                     Degrees of freedom =     1127
```

```
    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.4078         Pr(|T| > |t|) = 0.8156          Pr(T > t) = 0.5922
```

educhead

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	**534**	**2.945693**	**.1658558**	**3.832669**	**2.619882**	**3.271504**
1	**595**	**1.752941**	**.1236154**	**3.015304**	**1.510165**	**1.995718**
Combined	**1,129**	**2.317095**	**.1034556**	**3.47617**	**2.114108**	**2.520082**
diff		**1.192752**	**.2042385**		**.7920213**	**1.593482**

```
    diff = mean(0) - mean(1)                                      t =   5.8400
H0: diff = 0                                     Degrees of freedom =     1127

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

lnland

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	**534**	**.5006793**	**.0262664**	**.6069746**	**.449081**	**.5522776**
1	**595**	**.2688537**	**.0147783**	**.360481**	**.2398296**	**.2978777**
Combined	**1,129**	**.3785037**	**.0150559**	**.5058872**	**.348963**	**.4080444**
diff		**.2318256**	**.0293683**		**.174203**	**.2894483**

```
    diff = mean(0) - mean(1)                                      t =   7.8937
H0: diff = 0                                     Degrees of freedom =     1127

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

vaccess

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	**534**	**.8501873**	**.0154585**	**.3572224**	**.8198202**	**.8805544**
1	**595**	**.8218487**	**.0156999**	**.382962**	**.7910147**	**.8526828**
Combined	**1,129**	**.8352524**	**.011045**	**.371117**	**.8135815**	**.8569234**
diff		**.0283385**	**.0221159**		**-.0150545**	**.0717315**

diff = mean(**0**) - mean(**1**) t = **1.2814**

H0: diff = 0 Degrees of freedom = **1127**

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = **0.8998** Pr(|T| > |t|) = **0.2003** Pr(T > t) = **0.1002**

pcirr

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	**534**	**.5423426**	**.014232**	**.328879**	**.5143849**	**.5703003**
1	**595**	**.5765465**	**.0137037**	**.3342694**	**.5496329**	**.6034601**
Combined	**1,129**	**.5603686**	**.0098815**	**.3320238**	**.5409804**	**.5797567**
diff		**-.0342039**	**.0197744**		**-.0730027**	**.0045949**

diff = mean(**0**) - mean(**1**) t = **-1.7297**

H0: diff = 0 Degrees of freedom = **1127**

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = **0.0420** Pr(|T| > |t|) = **0.0840** Pr(T > t) = **0.9580**

rice

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	**534**	**10.27356**	**.0680716**	**1.573028**	**10.13984**	**10.40729**
1	**595**	**10.29144**	**.0640179**	**1.561565**	**10.16571**	**10.41716**
Combined	**1,129**	**10.28298**	**.0466161**	**1.566328**	**10.19152**	**10.37445**
diff		**-.0178721**	**.0934084**		**-.201146**	**.1654019**

diff = mean(**0**) - mean(**1**)   t = **-0.1913**
H0: diff = 0   Degrees of freedom = **1127**

Ha: diff < 0   Ha: diff != 0   Ha: diff > 0
Pr(T < t) = **0.4241**   Pr(|T| > |t|) = **0.8483**   Pr(T > t) = **0.5759**

wheat

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	**534**	**7.453635**	**.0378263**	**.8741065**	**7.379328**	**7.527941**
1	**595**	**7.478753**	**.0336961**	**.8219368**	**7.412575**	**7.544931**
Combined	**1,129**	**7.466872**	**.0251998**	**.8467278**	**7.417428**	**7.516316**
diff		**-.0251181**	**.0504901**		**-.1241834**	**.0739471**

diff = mean(**0**) - mean(**1**)   t = **-0.4975**
H0: diff = 0   Degrees of freedom = **1127**

Ha: diff < 0   Ha: diff != 0   Ha: diff > 0
Pr(T < t) = **0.3095**   Pr(|T| > |t|) = **0.6189**   Pr(T > t) = **0.6905**

milk

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	**534**	**10.92194**	**.1518803**	**3.509717**	**10.62359**	**11.2203**
1	**595**	**10.8724**	**.1338738**	**3.265534**	**10.60947**	**11.13532**
Combined	**1,129**	**10.89583**	**.1006472**	**3.381805**	**10.69836**	**11.09331**
diff		**.0495483**	**.2016728**		**-.3461481**	**.4452447**

```
    diff = mean(0) - mean(1)                                      t =   0.2457
H0: diff = 0                                     Degrees of freedom =     1127

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.5970         Pr(|T| > |t|) = 0.8060          Pr(T > t) = 0.4030
```

oil

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	**534**	**39.13095**	**.1710396**	**3.952458**	**38.79495**	**39.46694**
1	**595**	**39.64785**	**.165893**	**4.046564**	**39.32205**	**39.97366**
Combined	**1,129**	**39.40337**	**.1193099**	**4.008882**	**39.16927**	**39.63746**
diff		**-.5169044**	**.2385784**		**-.9850122**	**-.0487965**

```
    diff = mean(0) - mean(1)                                      t =  -2.1666
H0: diff = 0                                     Degrees of freedom =     1127

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0152         Pr(|T| > |t|) = 0.0305          Pr(T > t) = 0.9848
```

egg

Two-sample t test with equal variances

Group	Obs	Mean	Std. err.	Std. dev.	[95% conf.	interval]
0	**534**	**1.935005**	**.0156231**	**.3610246**	**1.904315**	**1.965695**
1	**595**	**1.969769**	**.0157782**	**.3848716**	**1.938781**	**2.000757**
Combined	**1,129**	**1.953326**	**.0111314**	**.3740207**	**1.931486**	**1.975167**
diff		**-.0347637**	**.0222811**		**-.0784809**	**.0089535**

```
    diff = mean(0) - mean(1)                                      t =  -1.5602
H0: diff = 0                                     Degrees of freedom =     1127

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0595         Pr(|T| > |t|) = 0.1190          Pr(T > t) = 0.9405
```

```
338 .
339 . * educhead, lnland, and oil has significant differences between treatment an
    > d control.
340 .
341 . * regressing treatment dummy on outcome and covariates that are difference b
    > etween treatment and control
342 .
343 . reg dfmfd lexptot educhead lnland oil
```

Source	SS	df	MS
Model	**21.6019027**	**4**	**5.40047568**
Residual	**259.824138**	**1,124**	**.231160265**
Total	**281.426041**	**1,128**	**.249491171**

Number of obs	=	**1,129**
F(4, 1124)	=	**23.36**
Prob > F	=	**0.0000**
R-squared	=	**0.0768**
Adj R-squared	=	**0.0735**
Root MSE	=	**.48079**

dfmfd	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lexptot	**.1068631**	**.030292**	**3.53**	**0.000**	**.0474279**	**.1662983**
educhead	**-.0205557**	**.0045525**	**-4.52**	**0.000**	**-.0294879**	**-.0116234**
lnland	**-.2078947**	**.0303684**	**-6.85**	**0.000**	**-.2674799**	**-.1483096**
oil	**.0067718**	**.003577**	**1.89**	**0.059**	**-.0002465**	**.0137901**
_cons	**-.5165626**	**.282748**	**-1.83**	**0.068**	**-1.071336**	**.0382107**

```
344 .
345 . reg dmmfd lexptot educhead lnland oil

      Source |       SS           df       MS      Number of obs   =     1,129
-------------+----------------------------------   F(4, 1124)      =      3.36
       Model |  2.09376776         4  .52344194   Prob > F        =    0.0096
    Residual |  175.036436     1,124  .155726367   R-squared       =    0.0118
-------------+----------------------------------   Adj R-squared   =    0.0083
       Total |  177.130204     1,128  .157030322   Root MSE        =    .39462

------------------------------------------------------------------------------
       dmmfd | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     lexptot |  -.0413688   .0248629    -1.66   0.096    -.0901517    .0074142
    educhead |   .0116205   .0037365     3.11   0.002     .0042891    .0189519
      lnland |  -.0557045   .0249256    -2.23   0.026    -.1046106   -.0067985
         oil |   .0002047   .0029359     0.07   0.944    -.0055558    .0059651
       _cons |   .5305503   .2320727     2.29   0.022     .0752059    .9858947
------------------------------------------------------------------------------

346 .
347 . * The effects on female and male are very different.
348 .
349 . * Question 5.B)
350 .
351 . * (i) covariate matching estimator for female microcredit borrowers
352 . teffects nnmatch (lexptot sexhead agehead educhead lnland vaccess pcirr rice
>  wheat milk oil egg) (dfmfd)

Treatment-effects estimation                   Number of obs      =      1,129
Estimator      : nearest-neighbor matching     Matches: requested =          1
Outcome model  : matching                                     min =          1
Distance metric: Mahalanobis                                  max =          1
------------------------------------------------------------------------------
             |              AI robust
     lexptot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATE          |
       dfmfd |
   (1 vs 0)  |    .098184   .0309151     3.18   0.001     .0375914    .1587765
------------------------------------------------------------------------------
```

```
353 .
354 . * (ii) covariate matching estimator for male microcredit borrowers
355 . teffects nnmatch (lexptot sexhead agehead educhead lnland vaccess pcirr rice
>  wheat milk oil egg) (dmmfd)

Treatment-effects estimation                   Number of obs      =      1,129
Estimator      : nearest-neighbor matching     Matches: requested =          1
Outcome model  : matching                                     min =          1
Distance metric: Mahalanobis                                  max =          1
------------------------------------------------------------------------------
             |              AI robust
     lexptot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATE          |
       dmmfd |
   (1 vs 0)  |  -.0507796   .0368775    -1.38   0.169    -.1230581    .0214989
------------------------------------------------------------------------------

356 .
357 . * The results from covariate matching are different than that from propensit
> y score matching.
358 .
359 . * Question 5.C)
360 .
361 . *
362 . foreach i of numlist 1/5 {
  2.         teffects nnmatch (lexptot sexhead agehead educhead lnland vaccess
>  pcirr rice wheat milk oil egg) (dfmfd), nneighbor(`i')
  3. }

Treatment-effects estimation                   Number of obs      =      1,129
Estimator      : nearest-neighbor matching     Matches: requested =          1
Outcome model  : matching                                     min =          1
Distance metric: Mahalanobis                                  max =          1
------------------------------------------------------------------------------
             |              AI robust
     lexptot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATE          |
       dfmfd |
   (1 vs 0)  |    .098184   .0309151     3.18   0.001     .0375914    .1587765
------------------------------------------------------------------------------
```

Treatment-effects estimation — Number of obs = **1,129**

Estimator : **nearest-neighbor matching** — Matches: requested = **2**, min = **2**, max = **2**

Outcome model : **matching**

Distance metric: **Mahalanobis**

lexptot	Coefficient	AI robust std. err.	z	P>|z|	[95% conf. interval]	
ATE						
dfmfd (1 vs 0)	**.0892153**	**.0290355**	**3.07**	**0.002**	**.0323067**	**.146124**

Treatment-effects estimation — Number of obs = **1,129**

Estimator : **nearest-neighbor matching** — Matches: requested = **3**, min = **3**, max = **3**

Outcome model : **matching**

Distance metric: **Mahalanobis**

lexptot	Coefficient	AI robust std. err.	z	P>|z|	[95% conf. interval]	
ATE						
dfmfd (1 vs 0)	**.0842812**	**.0282422**	**2.98**	**0.003**	**.0289276**	**.1396348**

Treatment-effects estimation — Number of obs = **1,129**

Estimator : **nearest-neighbor matching** — Matches: requested = **4**, min = **4**, max = **4**

Outcome model : **matching**

Distance metric: **Mahalanobis**

lexptot	Coefficient	AI robust std. err.	z	P>|z|	[95% conf. interval]	
ATE						
dfmfd (1 vs 0)	**.073181**	**.027638**	**2.65**	**0.008**	**.0190114**	**.1273505**

Treatment-effects estimation   Number of obs   =   **1,129**
Estimator   : **nearest-neighbor matching**   Matches: requested =   **5**
Outcome model : **matching**   min =   **5**
Distance metric: **Mahalanobis**   max =   **5**

lexptot	Coefficient	AI robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
dfmfd						
(1 vs 0)	**.0721804**	**.027661**	**2.61**	**0.009**	**.0179659**	**.126395**

```
363 .
364 . * what are the tradeoffs using more or fewer nearest neighbors
365 . * how is this tradeoff similar to using nearest neighbors with replacement
366 .
367 . * Question 5.D)
368 .
369 . * What is the difference between Mahalanobis and Euclidean distances in matc
    > hing
370 .
371 . * covariate matching estimator for female microcredit borrowers using Mahala
    > nobis distances
372 . teffects nnmatch (lexptot sexhead agehead educhead lnland vaccess pcirr rice
    >  wheat milk oil egg) (dfmfd), metric(mahalanobis)
```

Treatment-effects estimation   Number of obs   =   **1,129**
Estimator   : **nearest-neighbor matching**   Matches: requested =   **1**
Outcome model : **matching**   min =   **1**
Distance metric: **Mahalanobis**   max =   **1**

lexptot	Coefficient	AI robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
dfmfd						
(1 vs 0)	**.098184**	**.0309151**	**3.18**	**0.001**	**.0375914**	**.1587765**

```
373 .
374 . * covariate matching estimator for female microcredit borrowers using Euclid
> ean distances
375 . teffects nnmatch (lexptot sexhead agehead educhead lnland vaccess pcirr rice
>  wheat milk oil egg) (dfmfd), metric(euclidean)

Treatment-effects estimation                   Number of obs      =      1,129
Estimator      : nearest-neighbor matching     Matches: requested =          1
Outcome model  : matching                                     min =          1
Distance metric: Euclidean                                    max =          2
------------------------------------------------------------------------------
             |              AI robust
     lexptot | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATE          |
       dfmfd |
   (1 vs 0)  |   .0870838   .0312358     2.79   0.005     .0258628    .1483049
------------------------------------------------------------------------------

376 .
377 . * The results are somewhat similar. However, using euclidean metric, the res
> ults are significant at 5% level whereas, using mahalanobis, it is more sign
> ificant at 1% level.
378 .
end of do-file

379 . log close
      name:  <unnamed>
       log:  /Users/prakritishakya/Documents/Stata/pset1.smcl
  log type:  smcl
 closed on:   9 Nov 2023, 21:30:15
------------------------------------------------------------------------------
```

pset1

2023-11-07

```r
# running the library first
library(MatchIt)
library(haven)
library(readstata13)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(lmtest)
```

```
## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
library(sandwich)
library(whitestrap)
```

```
##
## Please cite as:
##
## Lopez, J. (2020), White's test and Bootstrapped White's test under the methodology of Jeong, J., Lee
```

```r
library(estimatr)
library(parameters)
library(clubSandwich)
```

```
## Registered S3 method overwritten by 'clubSandwich':
##   method    from
##   bread.mlm sandwich
```

```r
library(ggplot2)
```

Question 2. A)

```r
# defining multiple vectors
consp <- c(70,65,90,95,110,115,80,200,190,100)
inc <- c(80,100,120,140,160,180,200,220,240,260)

# creating a matrix
x <- cbind(consp, inc)

# print matrix x
print(x)
```

```
##       consp inc
##  [1,]    70  80
##  [2,]    65 100
##  [3,]    90 120
##  [4,]    95 140
##  [5,]   110 160
##  [6,]   115 180
##  [7,]    80 200
##  [8,]   200 220
##  [9,]   190 240
## [10,]   100 260
```

```r
# regressing consumption on income
model <- lm(consp ~ inc)

model
```

```
##
## Call:
## lm(formula = consp ~ inc)
##
## Coefficients:
## (Intercept)          inc
##     26.2424       0.5015
```

Question 2. B)

```r
# change consp and inc to individual matrices
matrix_inc <- as.matrix(inc)
ones <- c(1,1,1,1,1,1,1,1,1,1)
matrix_inc <- as.matrix(cbind(ones,inc))
matrix_consp <- as.matrix(consp)

# transpose of income matrix
t_inc <- as.matrix(t(matrix_inc))

dim(t_inc)
```

```
## [1]  2 10
```

```r
dim(matrix_inc)
```

```
## [1] 10  2
```

```r
# multiply transpose and original consumption matrix
covar <- (t_inc %*% matrix_inc)
dim(covar)
```

```
## [1] 2 2
```

```r
# find the beta
beta <- solve(covar) %*% (t_inc %*% matrix_consp)

beta
```

```
##            [,1]
## ones 26.2424242
## inc   0.5015152
```

Question 2. C)

```r
# predict the regression
forecast <- predict(lm(consp ~ inc))

forecast
```

```
##         1         2         3         4         5         6         7         8
##  66.36364  76.39394  86.42424  96.45455 106.48485 116.51515 126.54545 136.57576
##         9        10
## 146.60606 156.63636
```

```r
# calculate residuals
es <- forecast - consp

# display residuals
es
```

```
##          1          2          3          4          5          6          7
##  -3.636364  11.393939  -3.575758   1.454545  -3.515152   1.515152  46.545455
##          8          9         10
## -63.424242 -43.393939  56.636364
```

```r
# square the residuals
esqr = es * es

# calculate sigma
es_squared = sum(esqr)/8

# display sigma
es_squared
```

```
## [1] 1431.553
```

```r
# calculate the variance of the beta
var_beta <- es_squared * (solve(covar))

# find the standard deviation by taking the square root of  absolute values in the variance matrix
std = sqrt(abs(var_beta))

# display standard deviation
std
```

```
##           ones       inc
## ones 37.374439 2.7156338
## inc   2.715634 0.2082796
```

Question 2. D)

```r
# creating diagonal matrix with squared individual errors as the principal diagonal elements
matrix_esqr = as.matrix(diag(esqr))

# display diagonal matrix
matrix_esqr
```

```
##           [,1]     [,2]     [,3]     [,4]     [,5]     [,6]     [,7]     [,8]
##  [1,] 13.22314   0.0000  0.00000 0.000000  0.00000 0.000000    0.000    0.000
##  [2,]  0.00000 129.8219  0.00000 0.000000  0.00000 0.000000    0.000    0.000
##  [3,]  0.00000   0.0000 12.78604 0.000000  0.00000 0.000000    0.000    0.000
##  [4,]  0.00000   0.0000  0.00000 2.115702  0.00000 0.000000    0.000    0.000
##  [5,]  0.00000   0.0000  0.00000 0.000000 12.35629 0.000000    0.000    0.000
##  [6,]  0.00000   0.0000  0.00000 0.000000  0.00000 2.295684    0.000    0.000
##  [7,]  0.00000   0.0000  0.00000 0.000000  0.00000 0.000000 2166.479    0.000
##  [8,]  0.00000   0.0000  0.00000 0.000000  0.00000 0.000000    0.000 4022.635
##  [9,]  0.00000   0.0000  0.00000 0.000000  0.00000 0.000000    0.000    0.000
## [10,]  0.00000   0.0000  0.00000 0.000000  0.00000 0.000000    0.000    0.000
##           [,9]    [,10]
##  [1,]    0.000    0.000
##  [2,]    0.000    0.000
##  [3,]    0.000    0.000
##  [4,]    0.000    0.000
##  [5,]    0.000    0.000
##  [6,]    0.000    0.000
##  [7,]    0.000    0.000
##  [8,]    0.000    0.000
##  [9,] 1883.034    0.000
## [10,]    0.000 3207.678
```

```r
# calculate the "meat"
mid_matrix = as.matrix(t(matrix_inc) %*% matrix_esqr %*% matrix_inc)

# calculating robust standard errors

robust_std = 1.25 * (solve(covar) %*% mid_matrix %*% solve(covar))

# taking the square root of the standard errors
robust_std <- sqrt(abs(robust_std))

# display robust standard errors
robust_std
```

```
##           ones       inc
## ones 29.414616 2.6088950
## inc   2.608895 0.2347117
```

```r
# find the robust standard errors in R using regression
robust_model <- coeftest(model, vcov = vcovHC, type = "HC1")

# display the model
robust_model
```

```
##
## t test of coefficients:
##
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 26.24242   29.41462  0.8922  0.39835
## inc          0.50152    0.23471  2.1367  0.06511 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Both the robust standard errors found are 0.23.
```

Question 2. E)

```r
# creating village variable
village <- c(2,2,1,1,2,2,1,2,1,1)

# combining all the variables together
tmatrix <- as.matrix(cbind(village, es, consp, inc))

# taking a subset of village 1 and creating error variable, income variable
tmatrix1 <- subset(tmatrix, village == 1)
es1 <- tmatrix1[, 2]
X1 <- tmatrix1[, 4]
ones <- c(1,1,1,1,1)
X1 <- as.matrix(cbind(ones,X1))

# calculate the "meat" for village 1
mid1 = as.matrix(t(X1) %*% (es1 %*% t(es1)) %*% X1)

# taking a subset of village 2 and creating error variable, income variable
tmatrix2 <- subset(tmatrix, village == 2)
es2 <- tmatrix2[, 2]
X2 <- tmatrix2[, 4]
ones <- c(1,1,1,1,1)
X2 <- as.matrix(cbind(ones,X2))

# calculate the "meat" for village 2
mid2 = as.matrix(t(X2) %*% (es2 %*% t(es2)) %*% X2)

# summing the meat
meat = mid1 + mid2

# calculating the variance
cl_std = 2.25 * (solve(covar) %*% meat %*% solve(covar))

# taking the square root of the standard errors
cl_std <- sqrt(abs(cl_std))

# display clustered standard errors
cl_std
```

```
##           ones       inc
## ones 27.011869 2.4971464
## inc   2.497146 0.2308519
```

```r
# find the clustered standard errors in R using regression
cl_model <- coeftest(model, vcov = vcovHC, cluster = "village")

# display the model
```

```r
cl_model
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 26.24242   37.86113  0.6931   0.5079
## inc          0.50152    0.29656  1.6911   0.1293
```

```r
# Both the clustered standard errors found are 0.23.
```

Question 3. A)

```r
# calling the dataset
gbank <- read_dta("hh_98.dta")

# run regression on total expenditure and female micro credit program
reg1 <- lm(lexptot ~ progvillf, data = gbank)

# display the regression model
reg1
```

```
##
## Call:
## lm(formula = lexptot ~ progvillf, data = gbank)
##
## Coefficients:
## (Intercept)    progvillf
##      8.3285       0.1298
```

```r
# The regression shows that a unit increase in female microcredit program increases total expenditure by

# run regression on total expenditure and male micro credit program
reg2 <- lm(lexptot ~ progvillm, data = gbank)

# display the regression model
reg2
```

```
##
## Call:
## lm(formula = lexptot ~ progvillm, data = gbank)
##
## Coefficients:
## (Intercept)    progvillm
##     8.47927     -0.04736
```

```r
# The regression shows that a unit increase in male microcredit program decreases total expenditure by
```

Question 3. B)

```r
# run regression on total expenditure and female micro credit program with controls
reg3 <- lm(lexptot ~ progvillf + sexhead + agehead + educhead + lnland + vaccess + pcirr + rice + wheat

# display regression model
reg3
```

```
##
## Call:
```

```
## lm(formula = lexptot ~ progvillf + sexhead + agehead + educhead +
##     lnland + vaccess + pcirr + rice + wheat + milk + oil + egg,
##     data = gbank)
##
## Coefficients:
## (Intercept)    progvillf      sexhead      agehead     educhead       lnland
##    7.351443     0.112014    -0.053949     0.003601     0.048146     0.160321
##     vaccess        pcirr         rice        wheat         milk          oil
##   -0.015876     0.168442     0.003333    -0.039134     0.020363     0.010819
##         egg
##    0.119554
```

```r
# conducting the White test
white_test(reg3)
```

```
## White's test results
##
## Null hypothesis: Homoskedasticity of the residuals
## Alternative hypothesis: Heteroskedasticity of the residuals
## Test Statistic: 12.76
## P-value: 0.001695
```

```r
# We see that the p-value is less than 0.05 so we reject the null hypothesis that there is homoskedasti

# run regression on total expenditure and male micro credit program with controls
reg4 <- lm(lexptot ~ progvillm + sexhead + agehead + educhead + lnland + vaccess + pcirr + rice + wheat

# With controls, the coefficient on female microfinance program dropped down by 1.78 percentage point.

# display the regression model
reg4
```

```
##
## Call:
## lm(formula = lexptot ~ progvillm + sexhead + agehead + educhead +
##     lnland + vaccess + pcirr + rice + wheat + milk + oil + egg,
##     data = gbank)
##
## Coefficients:
## (Intercept)    progvillm      sexhead      agehead     educhead       lnland
##    7.522192    -0.063773    -0.058650     0.003379     0.048281     0.164341
##     vaccess        pcirr         rice        wheat         milk          oil
##   -0.026910     0.155329     0.006534    -0.039537     0.021203     0.010530
##         egg
##    0.106610
```

```r
# With controls, the coefficient on female microfinance program dropped down by 1.65 percentage point.

# conducting the White test
white_test(reg4)
```

```
## White's test results
##
## Null hypothesis: Homoskedasticity of the residuals
## Alternative hypothesis: Heteroskedasticity of the residuals
## Test Statistic: 15.37
```

```
## P-value: 0.000461
```

Question 3. C)

```r
# For both regressions, since p < 0.05, we reject the null hypothesis that there is homoskedasticity.

# robust standard errors for female microfinance program
reg5 <- coeftest(reg3, vcov = vcovHC, type = "HC1")

# display regression
reg5
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  7.3514430  0.2314452 31.7632 < 2.2e-16 ***
## progvillf    0.1120142  0.0590257  1.8977 0.0579907 .
## sexhead     -0.0539490  0.0565976 -0.9532 0.3406941
## agehead      0.0036010  0.0011178  3.2215 0.0013119 **
## educhead     0.0481461  0.0044472 10.8261 < 2.2e-16 ***
## lnland       0.1603209  0.0325314  4.9282 9.555e-07 ***
## vaccess     -0.0158758  0.0403351 -0.3936 0.6939536
## pcirr        0.1684416  0.0480318  3.5069 0.0004714 ***
## rice         0.0033335  0.0085033  0.3920 0.6951178
## wheat       -0.0391340  0.0168568 -2.3216 0.0204361 *
## milk         0.0203634  0.0056866  3.5809 0.0003571 ***
## oil          0.0108189  0.0032319  3.3475 0.0008426 ***
## egg          0.1195542  0.0495531  2.4127 0.0159975 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# robust standard errors for male microfinance program
reg6 <- coeftest(reg4, vcov = vcovHC, type = "HC1")

# display regression
reg6
```

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  7.5221921  0.2324683 32.3579 < 2.2e-16 ***
## progvillm   -0.0637731  0.0297973 -2.1402 0.0325520 *
## sexhead     -0.0586501  0.0562908 -1.0419 0.2976785
## agehead      0.0033787  0.0011098  3.0444 0.0023863 **
## educhead     0.0482812  0.0044659 10.8111 < 2.2e-16 ***
## lnland       0.1643405  0.0324936  5.0576 4.959e-07 ***
## vaccess     -0.0269104  0.0402121 -0.6692 0.5034988
## pcirr        0.1553286  0.0486838  3.1906 0.0014596 **
## rice         0.0065337  0.0086591  0.7545 0.4506796
## wheat       -0.0395366  0.0169433 -2.3335 0.0198004 *
## milk         0.0212032  0.0056546  3.7497 0.0001861 ***
## oil          0.0105296  0.0032401  3.2498 0.0011894 **
## egg          0.1066098  0.0491095  2.1709 0.0301521 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# We only find very small changes in the coefficients, standard error and the significance.
```

Question 3. D)

```r
# regress total expenditure on the number of female participants in the household with controls

reg7 <- lm(lexptot ~ dfmfd + sexhead + agehead + educhead + lnland + vaccess + pcirr + rice + wheat + mi

# display regression
reg7
```

```
##
## Call:
## lm(formula = lexptot ~ dfmfd + sexhead + agehead + educhead +
##     lnland + vaccess + pcirr + rice + wheat + milk + oil + egg,
##     data = gbank)
##
## Coefficients:
## (Intercept)        dfmfd      sexhead      agehead     educhead       lnland
##    7.450058     0.090514    -0.058558     0.003368     0.049832     0.175624
##     vaccess        pcirr         rice        wheat         milk          oil
##   -0.017716     0.160449     0.003998    -0.039129     0.020550     0.009993
##         egg
##    0.116440
```

```r
# The results show that a unit increase in the number of female microcredit borrowers increases total e

# regress total expenditure on the number of male participants in the household with controls

reg8 <- lm(lexptot ~ dmmfd + sexhead + agehead + educhead + lnland + vaccess + pcirr + rice + wheat + mi

# display regression
reg8
```

```
##
## Call:
## lm(formula = lexptot ~ dmmfd + sexhead + agehead + educhead +
##     lnland + vaccess + pcirr + rice + wheat + milk + oil + egg,
##     data = gbank)
##
## Coefficients:
## (Intercept)        dmmfd      sexhead      agehead     educhead       lnland
##    7.470394    -0.023227    -0.056038     0.003457     0.048629     0.156730
##     vaccess        pcirr         rice        wheat         milk          oil
##   -0.023239     0.167971     0.004117    -0.038175     0.020640     0.010622
##         egg
##    0.117315
```

```r
# The results show that a unit increase in the number of male microcredit borrowers decreases total exp
```

Question 3. E)

```r
# regress total expenditure on the number of female participants in the household with controls and clu
clreg <- lm_robust (lexptot ~ dfmfd + sexhead + agehead + educhead + lnland + vaccess + pcirr + rice + 

summary(clreg)
```

```
##
## Call:
## lm_robust(formula = lexptot ~ dfmfd + sexhead + agehead + educhead +
##     lnland + vaccess + pcirr + rice + wheat + milk + oil + egg,
##     data = gbank, clusters = villid)
##
## Standard error type:  CR2
##
## Coefficients:
##              Estimate Std. Error t value  Pr(>|t|)   CI Lower CI Upper    DF
## (Intercept)  7.450058  0.1942383 38.3552 0.0000985  6.7857043 8.114412 2.666
## dfmfd        0.090514  0.0213037  4.2488 0.0333387  0.0149725 0.166055 2.529
## sexhead     -0.058558  0.0403197 -1.4523 0.2797348 -0.2267600 0.109644 2.067
## agehead      0.003368  0.0007767  4.3366 0.0338784  0.0005488 0.006188 2.448
## educhead     0.049832  0.0080595  6.1830 0.0133167  0.0215133 0.078151 2.562
## lnland       0.175624  0.0439012  4.0004 0.0411679  0.0151491 0.336099 2.425
## vaccess     -0.017716  0.0505861 -0.3502 0.7595821 -0.2351042 0.199672 2.003
## pcirr        0.160449  0.0835657  1.9200 0.1751855 -0.1516189 0.472517 2.357
## rice         0.003998  0.0170022  0.2351 0.8341558 -0.0629654 0.070961 2.207
## wheat       -0.039129  0.0350440 -1.1166 0.3622013 -0.1668758 0.088617 2.434
## milk         0.020550  0.0072063  2.8516 0.0914157 -0.0074680 0.048567 2.241
## oil          0.009993  0.0041639  2.3998 0.1325329 -0.0071361 0.027121 2.099
## egg          0.116440  0.0736711  1.5805 0.2480863 -0.1845319 0.417411 2.115
##
## Multiple R-squared:   0.21 , Adjusted R-squared:  0.2015
## F-statistic:    NA on 12 and 3 DF,  p-value: NA
```

```r
# Using clustered standard errors decreases standard error and increases the statistical significance a

# regress total expenditure on the number of male participants in the household with controls and clust

clregm <- lm_robust (lexptot ~ dmmfd + sexhead + agehead + educhead + lnland + vaccess + pcirr + rice +

summary(clregm)
```

```
##
## Call:
## lm_robust(formula = lexptot ~ dmmfd + sexhead + agehead + educhead +
##     lnland + vaccess + pcirr + rice + wheat + milk + oil + egg,
##     data = gbank, clusters = villid)
##
## Standard error type:  CR2
##
## Coefficients:
##              Estimate Std. Error t value  Pr(>|t|)  CI Lower CI Upper    DF
## (Intercept)  7.470394  0.1889098 39.5448 9.209e-05  6.823475 8.117313 2.661
## dmmfd       -0.023227  0.0396103 -0.5864 6.086e-01 -0.170388 0.123935 2.373
## sexhead     -0.056038  0.0506044 -1.1074 3.798e-01 -0.266228 0.154152 2.077
## agehead      0.003457  0.0005793  5.9677 1.622e-02  0.001359 0.005555 2.456
## educhead     0.048629  0.0080437  6.0456 1.403e-02  0.020383 0.076874 2.564
## lnland       0.156730  0.0438282  3.5760 5.407e-02 -0.005963 0.319422 2.376
## vaccess     -0.023239  0.0482764 -0.4814 6.776e-01 -0.230207 0.183729 2.008
## pcirr        0.167971  0.0845561  1.9865 1.660e-01 -0.148484 0.484426 2.351
```

```
## rice         0.004117  0.0174997  0.2353 8.340e-01 -0.064719 0.072953 2.210
## wheat       -0.038175  0.0353230 -1.0807 3.748e-01 -0.166579 0.090229 2.443
## milk         0.020640  0.0075350  2.7392 9.860e-02 -0.008683 0.049963 2.239
## oil          0.010622  0.0039013  2.7227 1.075e-01 -0.005502 0.026747 2.088
## egg          0.117315  0.0819144  1.4322 2.811e-01 -0.214728 0.449358 2.134
##
## Multiple R-squared:  0.2031 ,    Adjusted R-squared:  0.1946
## F-statistic:    NA on 12 and 3 DF,  p-value: NA
```

```r
# Using clustered standard errors increases standard error but remains insignificant.
```

Question 3. F)

```r
# bootstrap standard errors of regression of male participants on total expenditure and controls
standard_error(reg8, bootstrap = TRUE, vcov = "HC1", summary = TRUE, iterations = 1000)
```

```
##      Parameter          SE
## 1  (Intercept) 0.228912072
## 2        dmmfd 0.032833469
## 3      sexhead 0.056610496
## 4      agehead 0.001121414
## 5     educhead 0.004474779
## 6       lnland 0.032838151
## 7      vaccess 0.040223185
## 8        pcirr 0.047969706
## 9         rice 0.008626033
## 10       wheat 0.016965973
## 11        milk 0.005681194
## 12         oil 0.003230389
## 13         egg 0.050422357
```

```r
# bootstrap standard errors of regression of female participants on total expenditure and controls
standard_error(reg7, bootstrap = TRUE, vcov = "HC1", summary = TRUE, iterations = 1000)
```

```
##      Parameter          SE
## 1  (Intercept) 0.227073034
## 2        dfmfd 0.027701344
## 3      sexhead 0.056174064
## 4      agehead 0.001115049
## 5     educhead 0.004546877
## 6       lnland 0.032778486
## 7      vaccess 0.040128205
## 8        pcirr 0.047865605
## 9         rice 0.008530786
## 10       wheat 0.016865054
## 11        milk 0.005684321
## 12         oil 0.003198162
## 13         egg 0.049553907
```

```r
# We find that our normal standard errors are overestimated and could cause Type II error.
```

Question 3. G)

```r
# bootstrap standard errors of regression of male participants on total expenditure and controls with c
standard_error(reg8, bootstrap = TRUE, vcov = "CR1", vcov_args = list(cluster = gbank$villid), summary =
```

```
##      Parameter           SE
## 1  (Intercept) 0.1699741164
```

```
## 2        dmmfd 0.0361164401
## 3      sexhead 0.0483981475
## 4      agehead 0.0005133471
## 5     educhead 0.0077538486
## 6       lnland 0.0411430772
## 7      vaccess 0.0432290155
## 8        pcirr 0.0771588294
## 9         rice 0.0153888837
## 10       wheat 0.0329950856
## 11        milk 0.0066607114
## 12         oil 0.0033726105
## 13         egg 0.0773456155
```

```r
# bootstrap standard errors of regression of female participants on total expenditure and controls with
standard_error(reg7, bootstrap = TRUE, vcov = "CR1", vcov_args = list(cluster = gbank$villid), summary =
```

```
##      Parameter           SE
## 1  (Intercept) 0.1757236794
## 2        dfmfd 0.0213708110
## 3      sexhead 0.0393496908
## 4      agehead 0.0006890057
## 5     educhead 0.0077505038
## 6       lnland 0.0411770220
## 7      vaccess 0.0456880356
## 8        pcirr 0.0764978617
## 9         rice 0.0149705019
## 10       wheat 0.0325855269
## 11        milk 0.0065023168
## 12         oil 0.0033692755
## 13         egg 0.0699456701
```

```r
# We find that our normal standard errors are overestimated and could cause Type II error.
```

Question 4

```r
# ra

# running model and estimate outcomes on treated
lm_treated_ra <- lm(lexptot ~ sexhead + agehead + educhead + lnland + vaccess + pcirr + rice + wheat + m

summary(lm_treated_ra)
```

```
##
## Call:
## lm(formula = lexptot ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, data = subset(gbank,
##     dfmfd == 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.00479 -0.31675 -0.04016  0.25085  2.16170
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.951888   0.321285  24.750  < 2e-16 ***
## sexhead     -0.084616   0.064947  -1.303  0.19314
```

```
## agehead      0.004109   0.001672   2.457  0.01431 *
## educhead     0.050946   0.006746   7.552 1.67e-13 ***
## lnland       0.045223   0.056338   0.803  0.42247
## vaccess     -0.023544   0.051936  -0.453  0.65048
## pcirr        0.163329   0.064028   2.551  0.01100 *
## rice        -0.005192   0.012849  -0.404  0.68631
## wheat       -0.049210   0.024041  -2.047  0.04111 *
## milk         0.027499   0.008587   3.203  0.00144 **
## oil          0.006592   0.005050   1.305  0.19225
## egg          0.036985   0.068841   0.537  0.59130
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.465 on 583 degrees of freedom
## Multiple R-squared:  0.1283, Adjusted R-squared:  0.1118
## F-statistic: 7.799 on 11 and 583 DF,  p-value: 1.107e-12
```

```r
gbank$lexptot_t1_ra <- predict(lm_treated_ra, gbank)

# running model and estimate outcomes on untreated
lm_untreated_ra <- lm(lexptot ~ sexhead + agehead + educhead + lnland + vaccess + pcirr + rice + wheat +

summary(lm_untreated_ra)
```

```
##
## Call:
## lm(formula = lexptot ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, data = subset(gbank,
##     dfmfd == 0))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.08980 -0.28373 -0.05336  0.23947  2.00672
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.103682   0.303571  23.400  < 2e-16 ***
## sexhead     -0.029244   0.071721  -0.408  0.68363
## agehead      0.002831   0.001514   1.870  0.06210 .
## educhead     0.050039   0.005483   9.126  < 2e-16 ***
## lnland       0.229058   0.035284   6.492 1.98e-10 ***
## vaccess     -0.007449   0.056733  -0.131  0.89559
## pcirr        0.131539   0.068896   1.909  0.05678 .
## rice         0.012206   0.013187   0.926  0.35506
## wheat       -0.029244   0.023724  -1.233  0.21826
## milk         0.013463   0.007598   1.772  0.07699 .
## oil          0.013550   0.005108   2.653  0.00823 **
## egg          0.170177   0.069531   2.447  0.01471 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4512 on 522 degrees of freedom
## Multiple R-squared:  0.3066, Adjusted R-squared:  0.292
## F-statistic: 20.98 on 11 and 522 DF,  p-value: < 2.2e-16
```

```r
gbank$lexptot_t0_ra <- predict(lm_untreated_ra, newdata = gbank)

# creating the difference between treated and untreated
gbank$ATE_ra <- (gbank$lexptot_t1_ra - gbank$lexptot_t0_ra)

# average treatment effects with regression adjustments
t.test(gbank$ATE_ra, data = gbank)
```

```
##
##  One Sample t-test
##
## data:  gbank$ATE_ra
## t = 25.522, df = 1128, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.07776257 0.09071444
## sample estimates:
## mean of x
## 0.0842385
```

```r
# create a subset for just the treated group
gbank_t1 <- gbank[gbank$dfmfd==1,]

# treatment effects on the treated with regression adjustments
t.test(gbank_t1$ATE_ra, data = gbank_t1)
```

```
##
##  One Sample t-test
##
## data:  gbank_t1$ATE_ra
## t = 26.83, df = 594, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.09314945 0.10786341
## sample estimates:
## mean of x
## 0.1005064
```

```r
# ipw

# find logit
logit_ipw <- glm(dfmfd ~ sexhead + agehead + educhead + lnland + vaccess + pcirr + rice + wheat + milk +

summary(logit_ipw)
```

```
##
## Call:
## glm(formula = dfmfd ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, family = binomial,
##     data = gbank)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6629  -1.1969   0.8382   1.0452   2.1547
##
```

```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.410320   1.003560  -1.405 0.159927
## sexhead     -0.029784   0.217550  -0.137 0.891106
## agehead      0.006780   0.005108   1.327 0.184420
## educhead    -0.065693   0.019606  -3.351 0.000806 ***
## lnland      -0.887878   0.147941  -6.002 1.95e-09 ***
## vaccess     -0.264289   0.175109  -1.509 0.131227
## pcirr        0.382918   0.211403   1.811 0.070091 .
## rice        -0.026409   0.041201  -0.641 0.521537
## wheat        0.016489   0.076412   0.216 0.829147
## milk        -0.004249   0.025690  -0.165 0.868645
## oil          0.034651   0.016364   2.118 0.034215 *
## egg          0.279835   0.220247   1.271 0.203889
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1561.8  on 1128  degrees of freedom
## Residual deviance: 1473.2  on 1117  degrees of freedom
## AIC: 1497.2
##
## Number of Fisher Scoring iterations: 4
```

```r
# get estimates of propensity scores on each observations
gbank$logit_pscore <- predict(logit_ipw, newdata = gbank, type = "response")

gbank <- mutate(gbank, lexptot_wt = ifelse(dfmfd==1, lexptot/logit_pscore, lexptot/(1-logit_pscore)))

# running model and estimate outcomes on treated
lm_treated_ipw <- lm(lexptot ~ sexhead + agehead + educhead + lnland + vaccess + pcirr + rice + wheat +

summary(lm_treated_ipw)
```

```
##
## Call:
## lm(formula = lexptot ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, data = subset(gbank,
##     dfmfd == 1), weights = lexptot_wt)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -5.6285 -1.2984 -0.2658  0.9199  9.4096
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.947228   0.331136  24.000  < 2e-16 ***
## sexhead     -0.095855   0.069689  -1.375  0.16951
## agehead      0.004062   0.001730   2.349  0.01917 *
## educhead     0.049886   0.006202   8.043 4.92e-15 ***
## lnland       0.020883   0.045916   0.455  0.64942
## vaccess     -0.028979   0.054546  -0.531  0.59543
## pcirr        0.131301   0.065121   2.016  0.04423 *
## rice        -0.008222   0.012983  -0.633  0.52681
```

```
## wheat       -0.046587   0.024991  -1.864  0.06280 .
## milk         0.026594   0.009135   2.911  0.00374 **
## oil          0.006777   0.005062   1.339  0.18119
## egg          0.081602   0.072722   1.122  0.26228
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.905 on 583 degrees of freedom
## Multiple R-squared:  0.1404, Adjusted R-squared:  0.1242
## F-statistic: 8.656 on 11 and 583 DF,  p-value: 2.776e-14
```

```
gbank$lexptot_t1_ipw <- predict(lm_treated_ipw, gbank)

# running model and estimate outcomes on untreated
lm_untreated_ipw <- lm(lexptot ~ sexhead + agehead + educhead + lnland + vaccess + pcirr + rice + wheat

summary(lm_untreated_ipw)
```

```
##
## Call:
## lm(formula = lexptot ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, data = subset(gbank,
##     dfmfd == 0), weights = lexptot_wt)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5457 -1.2349 -0.3058  0.9073  8.1033
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.986641   0.308733  22.630  < 2e-16 ***
## sexhead     -0.004328   0.068814  -0.063  0.94988
## agehead      0.003656   0.001494   2.447  0.01472 *
## educhead     0.052388   0.005912   8.861  < 2e-16 ***
## lnland       0.220707   0.040063   5.509 5.68e-08 ***
## vaccess     -0.028241   0.054339  -0.520  0.60348
## pcirr        0.151741   0.068161   2.226  0.02643 *
## rice         0.014788   0.012939   1.143  0.25361
## wheat       -0.030830   0.023576  -1.308  0.19155
## milk         0.016347   0.007598   2.152  0.03189 *
## oil          0.014607   0.005300   2.756  0.00606 **
## egg          0.167883   0.069195   2.426  0.01559 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.893 on 522 degrees of freedom
## Multiple R-squared:  0.284,  Adjusted R-squared:  0.2689
## F-statistic: 18.82 on 11 and 522 DF,  p-value: < 2.2e-16
```

```
gbank$lexptot_t0_ipw <- predict(lm_untreated_ipw, newdata = gbank)

# creating the difference between treated and untreated
gbank$ATE_ipw <- (gbank$lexptot_t1_ipw - gbank$lexptot_t0_ipw)

# average treatment effects with regression adjustments
```

```r
t.test(gbank$ATE_ipw, data = gbank)
```

```
##
##  One Sample t-test
##
## data:  gbank$ATE_ipw
## t = 22.699, df = 1128, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.07585031 0.09020399
## sample estimates:
##  mean of x
## 0.08302715
```

```r
# create a subset for just the treated group
gbank_t1_ipw <- gbank[gbank$dfmfd==1,]

# treatment effects on the treated with regression adjustments
t.test(gbank_t1_ipw$ATE_ipw, data = gbank_t1_ipw)
```

```
##
##  One Sample t-test
##
## data:  gbank_t1_ipw$ATE_ipw
## t = 25.145, df = 594, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.09472696 0.11077791
## sample estimates:
## mean of x
## 0.1027524
```

Question 5. A)

```r
# estimate propensity score on yet unmatched data for female microcredit borrowers

f_ps <- glm(dfmfd ~ sexhead + agehead + educhead + lnland + vaccess + pcirr + rice + wheat + milk + oil

summary(f_ps)
```

```
##
## Call:
## glm(formula = dfmfd ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, family = binomial(link = ("probit")),
##     data = gbank)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6672  -1.1991   0.8418   1.0482   2.1981
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.877930   0.617867  -1.421 0.155344
## sexhead     -0.015314   0.134096  -0.114 0.909080
## agehead      0.003973   0.003139   1.266 0.205558
## educhead    -0.040097   0.012010  -3.339 0.000842 ***
```

```
## lnland      -0.542055   0.088269  -6.141  8.2e-10 ***
## vaccess     -0.160977   0.107536  -1.497 0.134404
## pcirr        0.233379   0.130202   1.792 0.073064 .
## rice        -0.015862   0.025464  -0.623 0.533331
## wheat        0.013008   0.047081   0.276 0.782328
## milk        -0.003136   0.015804  -0.198 0.842694
## oil          0.021283   0.010051   2.117 0.034218 *
## egg          0.171692   0.135652   1.266 0.205626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1561.8  on 1128  degrees of freedom
## Residual deviance: 1473.4  on 1117  degrees of freedom
## AIC: 1497.4
##
## Number of Fisher Scoring iterations: 4
```

```r
# estimate propensity score on yet unmatched data for male microcredit borrowers

m_ps <- glm(dmmfd ~ sexhead + agehead + educhead + lnland + vaccess + pcirr + rice + wheat + milk + oil

summary(m_ps)
```

```
##
## Call:
## glm(formula = dmmfd ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, family = binomial(link = ("probit")),
##     data = gbank)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1960  -0.7040  -0.5699  -0.2943   3.1833
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.156928   0.718479  -1.610 0.107344
## sexhead      0.794383   0.218707   3.632 0.000281 ***
## agehead     -0.006529   0.003690  -1.770 0.076807 .
## educhead     0.025272   0.013390   1.887 0.059117 .
## lnland      -0.274878   0.104918  -2.620 0.008795 **
## vaccess      0.068326   0.128130   0.533 0.593856
## pcirr       -0.137786   0.152299  -0.905 0.365620
## rice         0.111413   0.029077   3.832 0.000127 ***
## wheat        0.086570   0.054700   1.583 0.113506
## milk         0.029079   0.019066   1.525 0.127204
## oil         -0.013908   0.011118  -1.251 0.210961
## egg         -0.865880   0.167665  -5.164 2.41e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1113.6  on 1128  degrees of freedom
```

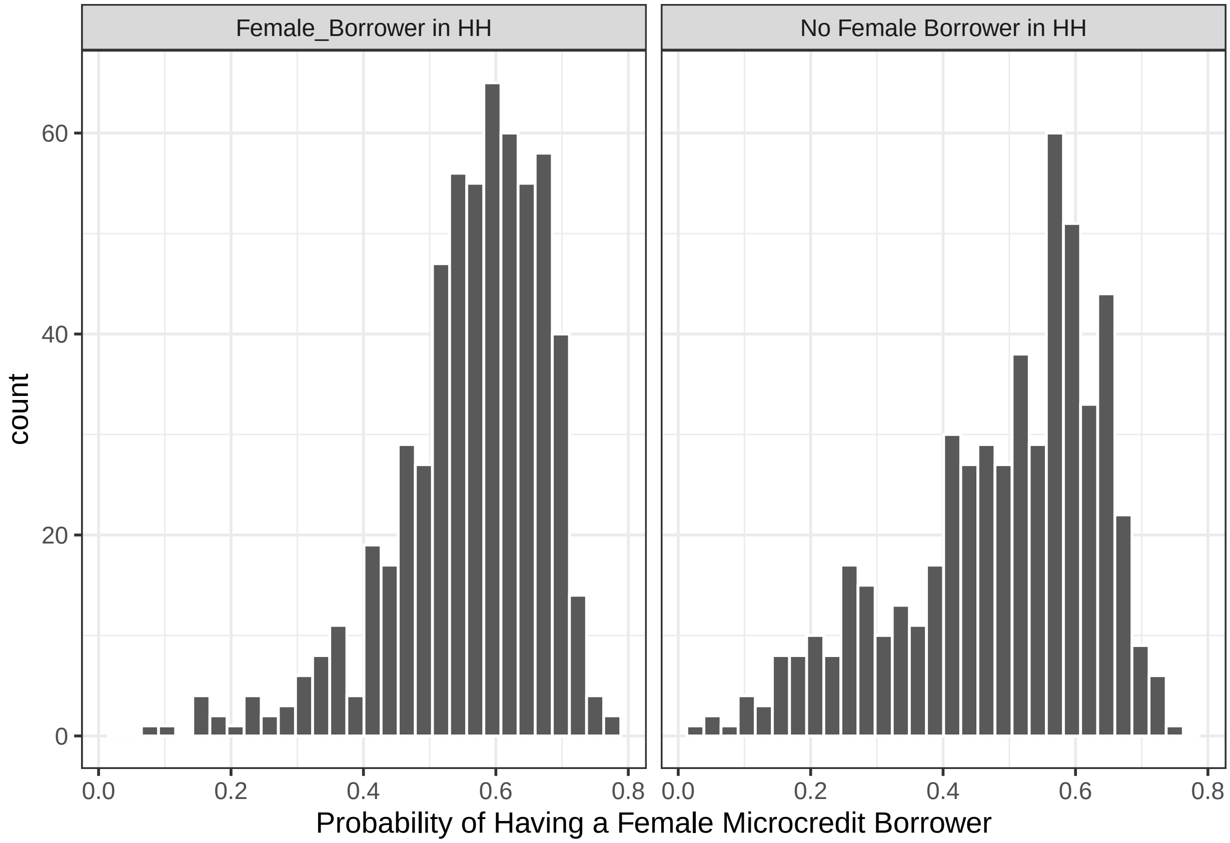
```
## Residual deviance: 1038.1  on 1117  degrees of freedom
## AIC: 1062.1
##
## Number of Fisher Scoring iterations: 5
```

```r
# assign each observation a propensity score for female and male microcredit borrowers

prs_f <- data.frame(pr_score = predict(f_ps, type = "response"), dfmfd = f_ps$model$dfmfd)

prs_m <- data.frame(pr_score = predict(m_ps, type = "response"), dmmfd = m_ps$model$dmmfd)

head(prs_f)
```

```
##    pr_score dfmfd
## 1 0.6212929     1
## 2 0.3723999     1
## 3 0.5130929     1
## 4 0.6219732     1
## 5 0.4338388     1
## 6 0.4988592     0
```

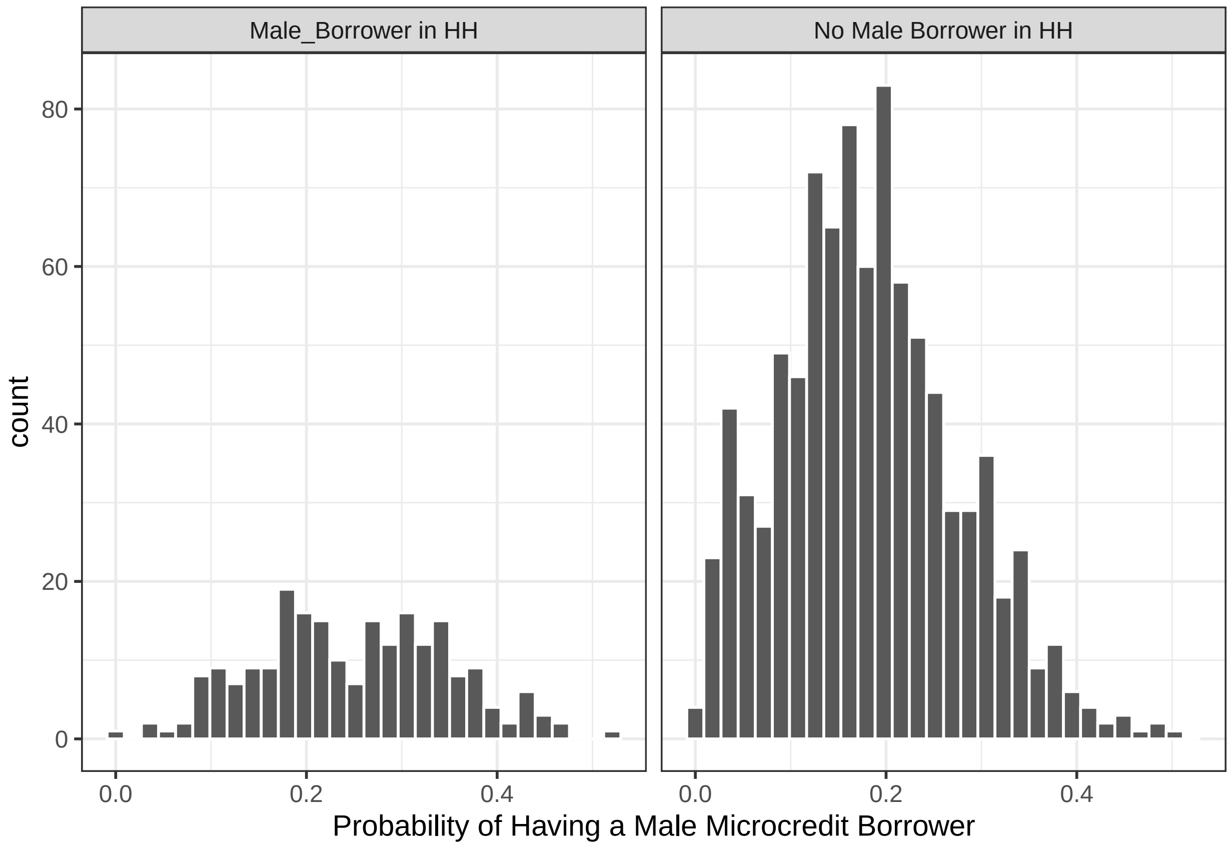
```r
hh_98cb <- cbind(gbank, prs_f$pr_score)
hh_98cbm <- cbind(gbank, prs_m$pr_score)

# examine the region of common support in unmatched data for female microcredit borrowers

labs <- paste(c("Female_Borrower in HH", "No Female Borrower in HH"))
prs_f %>%
  mutate(dfmfd = ifelse(dfmfd == 1, labs[1], labs[2])) %>%
  ggplot(aes(x = pr_score)) +
  geom_histogram(color = "white") +
  facet_wrap(~dfmfd) +
  xlab("Probability of Having a Female Microcredit Borrower") +
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```r
# examine the region of common support in unmatched data for male microcredit borrowers
labs <- paste(c("Male_Borrower in HH", "No Male Borrower in HH"))
prs_m %>%
  mutate(dmmfd = ifelse(dmmfd == 1, labs[1], labs[2])) %>%
  ggplot(aes(x = pr_score)) +
  geom_histogram(color = "white") +
  facet_wrap(~dmmfd) +
  xlab("Probability of Having a Male Microcredit Borrower") +
  theme_bw()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```r
# checking balances in covariates

hh_98_cov <- c('sexhead', 'agehead', 'educhead', 'lnland', 'vaccess', 'pcirr', 'rice', 'wheat','milk',

# carry out t-tests to see if differences are significant

attach(gbank)
t.test(sexhead ~ dfmfd)
```

```
##
##  Welch Two Sample t-test
##
## data:  sexhead by dfmfd
## t = 0.86669, df = 1125.6, p-value = 0.3863
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.01881904  0.04859905
## sample estimates:
## mean in group 0 mean in group 1
##       0.9157303       0.9008403
```

```r
t.test(agehead ~ dfmfd)
```

```
##
##  Welch Two Sample t-test
##
## data:  agehead by dfmfd
## t = -0.23139, df = 1059.8, p-value = 0.8171
```

```
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -1.671584  1.318939
## sample estimates:
## mean in group 0 mean in group 1
##        45.91948        46.09580
```

```r
t.test(educhead ~ dfmfd)
```

```
##
##  Welch Two Sample t-test
##
## data:  educhead by dfmfd
## t = 5.7661, df = 1010, p-value = 1.077e-08
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  0.7868373 1.5986662
## sample estimates:
## mean in group 0 mean in group 1
##        2.945693        1.752941
```

```r
t.test(lnland ~ dfmfd)
```

```
##
##  Welch Two Sample t-test
##
## data:  lnland by dfmfd
## t = 7.6921, df = 847.64, p-value = 4.017e-14
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  0.1726711 0.2909802
## sample estimates:
## mean in group 0 mean in group 1
##       0.5006793       0.2688537
```

```r
t.test(vaccess ~ dfmfd)
```

```
##
##  Welch Two Sample t-test
##
## data:  vaccess by dfmfd
## t = 1.2862, df = 1125.3, p-value = 0.1986
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.01489186  0.07156891
## sample estimates:
## mean in group 0 mean in group 1
##       0.8501873       0.8218487
```

```r
t.test(pcirr ~ dfmfd)
```

```
##
##  Welch Two Sample t-test
##
## data:  pcirr by dfmfd
## t = -1.7312, df = 1117.5, p-value = 0.08369
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.07296898  0.00456121
## sample estimates:
## mean in group 0 mean in group 1
##       0.5423426       0.5765465
```

```r
t.test(rice ~ dfmfd)
```

```
##
##  Welch Two Sample t-test
##
## data:  rice by dfmfd
## t = -0.19126, df = 1112.1, p-value = 0.8484
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.2012212  0.1654770
## sample estimates:
## mean in group 0 mean in group 1
##        10.27356        10.29144
```

```r
t.test(wheat ~ dfmfd)
```

```
##
##  Welch Two Sample t-test
##
## data:  wheat by dfmfd
## t = -0.49583, df = 1095.5, p-value = 0.6201
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.12451624  0.07428002
## sample estimates:
## mean in group 0 mean in group 1
##        7.453635        7.478753
```

```r
t.test(milk ~ dfmfd)
```

```
##
##  Welch Two Sample t-test
##
## data:  milk by dfmfd
## t = 0.24473, df = 1091.7, p-value = 0.8067
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.3477054  0.4468020
## sample estimates:
## mean in group 0 mean in group 1
##        10.92194        10.87240
```

```r
t.test(egg ~ dfmfd)
```

```
##
##  Welch Two Sample t-test
##
## data:  egg by dfmfd
## t = -1.5656, df = 1124.8, p-value = 0.1177
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.078330251  0.008802826
## sample estimates:
## mean in group 0 mean in group 1
##        1.935005        1.969769
```

```r
detach(gbank)

# educhead, lnland, and oil has significant differences between treatment and control.

# omitting missing values for female
hh_98_nomiss <- hh_98cb %>%
  select(lexptot, dfmfd, one_of(hh_98_cov)) %>%
  na.omit()

# omitting missing values for male
hh_98_nomissm <- hh_98cbm %>%
  select(lexptot, dmmfd, one_of(hh_98_cov)) %>%
  na.omit()

# using propensity score matching to match using nearest neighbor for female
mod_match <- matchit(data = hh_98_nomiss, dfmfd ~ sexhead + agehead + educhead + lnland + vaccess + pci
```

```
## Warning: Fewer control units than treated units; not all treated units will get
## a match.
```

```r
summary(mod_match)
```

```
##
## Call:
## matchit(formula = dfmfd ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, data = hh_98_nomiss,
##     method = "nearest", distance = "glm", discard = "both", caliper = c(0.1,
##         lnland = 0.5, educhead = 2), ration = 1, estimated = "ATT")
##
## Summary of Balance for All Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance        0.5627        0.4872          0.6615     0.5870    0.1509
## sexhead         0.9008        0.9157         -0.0498          .    0.0149
## agehead        46.0958       45.9195          0.0150     0.7471    0.0312
## educhead        1.7529        2.9457         -0.3956     0.6190    0.0762
## lnland          0.2689        0.5007         -0.6431     0.3527    0.1170
## vaccess         0.8218        0.8502         -0.0741          .    0.0283
## pcirr           0.5765        0.5423          0.1023     1.0330    0.0411
## rice           10.2914       10.2736          0.0114     0.9855    0.0160
## wheat           7.4788        7.4536          0.0306     0.8842    0.0152
## milk           10.8724       10.9219         -0.0152     0.8657    0.0145
## oil            39.6479       39.1309          0.1277     1.0482    0.0185
## egg             1.9698        1.9350          0.0903     1.1365    0.0283
##          eCDF Max
## distance   0.2209
## sexhead    0.0149
## agehead    0.0833
## educhead   0.1567
## lnland     0.1824
## vaccess    0.0283
```

```
## pcirr      0.0985
## rice       0.0457
## wheat      0.0272
## milk       0.0408
## oil        0.0838
## egg        0.0542
##
## Summary of Balance for Matched Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance        0.5456        0.5367          0.0779     1.0564    0.0272
## sexhead         0.9106        0.9129         -0.0079          .    0.0024
## agehead        45.1482       45.1082          0.0034     0.6573    0.0370
## educhead        2.0188        2.2494         -0.0765     0.9742    0.0169
## lnland          0.2946        0.3054         -0.0299     0.9123    0.0079
## vaccess         0.8518        0.8494          0.0061          .    0.0024
## pcirr           0.5628        0.5474          0.0461     1.0453    0.0288
## rice           10.3272       10.3057          0.0138     1.0704    0.0132
## wheat           7.4753        7.4809         -0.0068     0.8430    0.0120
## milk           10.8906       10.7680          0.0375     0.9876    0.0193
## oil            39.3813       39.3192          0.0153     1.2353    0.0145
## egg             1.9588        1.9377          0.0548     1.2385    0.0316
##          eCDF Max Std. Pair Dist.
## distance   0.0706          0.0822
## sexhead    0.0024          0.5432
## agehead    0.1035          1.0676
## educhead   0.0541          0.5478
## lnland     0.0471          0.2518
## vaccess    0.0024          0.5842
## pcirr      0.0706          1.0043
## rice       0.0541          1.1034
## wheat      0.0353          1.0780
## milk       0.0400          1.0497
## oil        0.0588          0.9109
## egg        0.0518          1.0167
##
## Sample Sizes:
##           Control Treated
## All           534     595
## Matched       425     425
## Unmatched     105     165
## Discarded       4       5
```

```r
# using propensity score matching to match using nearest neighbor for male
mod_match1 <- matchit(data = hh_98_nomissm, dmmfd ~ sexhead + agehead + educhead + lnland + vaccess + pc

summary(mod_match1)
```

```
##
## Call:
## matchit(formula = dmmfd ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, data = hh_98_nomissm,
##     method = "nearest", distance = "glm", discard = "both", caliper = c(0.1,
##         lnland = 0.5, educhead = 2), ration = 1, estimated = "ATT")
##
## Summary of Balance for All Data:
```

```
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance        0.2519        0.1811          0.6523     1.2798    0.1861
## sexhead         0.9773        0.8911          0.5783          .    0.0862
## agehead        44.0364       46.4906         -0.2089     0.8353    0.0430
## educhead        2.7409        2.2145          0.1434     1.1516    0.0345
## lnland          0.3243        0.3916         -0.1700     0.5617    0.0346
## vaccess         0.8500        0.8317          0.0513          .    0.0183
## pcirr           0.5532        0.5621         -0.0252     1.1482    0.0266
## rice           10.5316       10.2228          0.1731     1.4065    0.0513
## wheat           7.5296        7.4517          0.1006     0.8055    0.0344
## milk           10.6085       10.9654         -0.1070     0.9682    0.0284
## oil            39.4259       39.3979          0.0064     1.2295    0.0218
## egg             1.8509        1.9781         -0.3710     0.8270    0.0836
##          eCDF Max
## distance   0.3022
## sexhead    0.0862
## agehead    0.1217
## educhead   0.0790
## lnland     0.0639
## vaccess    0.0183
## pcirr      0.0528
## rice       0.1258
## wheat      0.0756
## milk       0.0861
## oil        0.0531
## egg        0.1483
##
## Summary of Balance for Matched Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## distance        0.2422        0.2416          0.0060     1.0166    0.0029
## sexhead         0.9758        0.9710          0.0324          .    0.0048
## agehead        44.6039       44.2560          0.0296     0.8495    0.0156
## educhead        2.4203        2.2319          0.0513     1.1899    0.0177
## lnland          0.3007        0.2869          0.0348     1.0664    0.0168
## vaccess         0.8551        0.8164          0.1082          .    0.0386
## pcirr           0.5554        0.5221          0.0949     1.0239    0.0348
## rice           10.4341       10.5109         -0.0431     1.3500    0.0355
## wheat           7.5252        7.6315         -0.1372     0.8211    0.0478
## milk           10.6863       10.6733          0.0039     1.1356    0.0245
## oil            39.4175       39.4960         -0.0180     1.3855    0.0238
## egg             1.8641        1.8805         -0.0477     0.9891    0.0150
##          eCDF Max Std. Pair Dist.
## distance   0.0242          0.0155
## sexhead    0.0048          0.0972
## agehead    0.0773          1.0223
## educhead   0.0483          0.5197
## lnland     0.0870          0.2593
## vaccess    0.0386          0.7576
## pcirr      0.0773          1.0650
## rice       0.0966          0.9173
## wheat      0.1256          1.0994
## milk       0.0918          1.0621
## oil        0.0531          0.9421
## egg        0.0290          0.8530
```

```
##
## Sample Sizes:
##           Control Treated
## All           909     220
## Matched       207     207
## Unmatched     700      12
## Discarded       2       1
```

```r
# creating a dataset of successful matches for female
dta_m <- match.data(mod_match)

# creating a dataset of successful matches for male
dta_m1 <- match.data(mod_match1)

# obtaining ATT on the matched data for female
t.test(lexptot ~ dfmfd, data = dta_m)
```

```
##
##  Welch Two Sample t-test
##
## data:  lexptot by dfmfd
## t = -3.0805, df = 847.95, p-value = 0.002133
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.1721071 -0.0381451
## sample estimates:
## mean in group 0 mean in group 1
##        8.359695        8.464821
```

```r
# obtaining ATT on the matched data for male
t.test(lexptot ~ dmmfd, data = dta_m1)
```

```
##
##  Welch Two Sample t-test
##
## data:  lexptot by dmmfd
## t = 0.32575, df = 405.24, p-value = 0.7448
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.07442333  0.10398739
## sample estimates:
## mean in group 0 mean in group 1
##        8.410672        8.395890
```

```r
# regress treatment dummy on outcome and covariates that are different between treatment and control

lm(lexptot ~ dfmfd + educhead + lnland + oil, data = dta_m)
```

```
##
## Call:
## lm(formula = lexptot ~ dfmfd + educhead + lnland + oil, data = dta_m)
##
## Coefficients:
## (Intercept)        dfmfd     educhead       lnland          oil
##    7.991774     0.117853     0.050424     0.133277     0.005437
```

```r
lm(lexptot ~ dmmfd + educhead + lnland + oil, data = dta_m1)
```

```
##
## Call:
## lm(formula = lexptot ~ dmmfd + educhead + lnland + oil, data = dta_m1)
##
## Coefficients:
## (Intercept)        dmmfd     educhead       lnland          oil
##     7.46349     -0.02227      0.03729      0.15187      0.02077
```

```r
# The effects on female and male are very different.
```

Question 5. B)

```r
# covariate matching estimator for female microcredit borrowers
mod_match3 <- matchit(data = hh_98_nomiss, dfmfd ~ sexhead + agehead + educhead + lnland + vaccess + pc
```

```
## Warning: Fewer control units than treated units; not all treated units will get
## a match.
```

```r
summary (mod_match3)
```

```
##
## Call:
## matchit(formula = dfmfd ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, data = hh_98_nomiss,
##     method = "nearest", distance = "mahalanobis", ratio = 1,
##     estimated = "ATT")
##
## Summary of Balance for All Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## sexhead         0.9008        0.9157         -0.0498          .    0.0149
## agehead        46.0958       45.9195          0.0150     0.7471    0.0312
## educhead        1.7529        2.9457         -0.3956     0.6190    0.0762
## lnland          0.2689        0.5007         -0.6431     0.3527    0.1170
## vaccess         0.8218        0.8502         -0.0741          .    0.0283
## pcirr           0.5765        0.5423          0.1023     1.0330    0.0411
## rice           10.2914       10.2736          0.0114     0.9855    0.0160
## wheat           7.4788        7.4536          0.0306     0.8842    0.0152
## milk           10.8724       10.9219         -0.0152     0.8657    0.0145
## oil            39.6479       39.1309          0.1277     1.0482    0.0185
## egg             1.9698        1.9350          0.0903     1.1365    0.0283
##          eCDF Max
## sexhead    0.0149
## agehead    0.0833
## educhead   0.1567
## lnland     0.1824
## vaccess    0.0283
## pcirr      0.0985
## rice       0.0457
## wheat      0.0272
## milk       0.0408
## oil        0.0838
## egg        0.0542
##
## Summary of Balance for Matched Data:
```

```
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## sexhead         0.9026        0.9157         -0.0439          .    0.0131
## agehead        46.0187       45.9195          0.0084     0.7604    0.0305
## educhead        1.7734        2.9457         -0.3888     0.6117    0.0745
## lnland          0.2805        0.5007         -0.6109     0.3730    0.1105
## vaccess         0.8539        0.8502          0.0098          .    0.0037
## pcirr           0.5945        0.5423          0.1561     1.0311    0.0595
## rice           10.3371       10.2736          0.0407     1.0730    0.0115
## wheat           7.4302        7.4536         -0.0285     0.9102    0.0098
## milk           10.7864       10.9219         -0.0415     0.7984    0.0205
## oil            39.5529       39.1309          0.1043     1.1132    0.0143
## egg             1.9230        1.9350         -0.0313     0.9294    0.0162
##          eCDF Max Std. Pair Dist.
## sexhead    0.0131          0.0689
## agehead    0.0805          0.6585
## educhead   0.1498          0.6633
## lnland     0.1723          0.9619
## vaccess    0.0037          0.0881
## pcirr      0.1255          0.4782
## rice       0.0300          0.5108
## wheat      0.0225          0.4695
## milk       0.0506          0.4901
## oil        0.0674          0.4694
## egg        0.0337          0.4807
##
## Sample Sizes:
##           Control Treated
## All           534     595
## Matched       534     534
## Unmatched       0      61
## Discarded       0       0
```

```
# creating a dataset of successful matches for female
dta_m3 <- match.data(mod_match3)

# obtaining ATT on the matched data for female
t.test(lexptot ~ dfmfd, data = dta_m3)
```

```
##
##  Welch Two Sample t-test
##
## data:  lexptot by dfmfd
## t = -0.42794, df = 1057.1, p-value = 0.6688
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.07505290  0.04817764
## sample estimates:
## mean in group 0 mean in group 1
##        8.447977        8.461414
```

```
# covariate matching estimator for male microcredit borrowers

mod_match4 <- matchit(data = hh_98_nomissm, dmmfd ~ sexhead + agehead + educhead + lnland + vaccess + p

summary (mod_match4)
```

```
##
## Call:
## matchit(formula = dmmfd ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, data = hh_98_nomissm,
##     method = "nearest", distance = "mahalanobis", ratio = 1,
##     estimated = "ATT")
##
## Summary of Balance for All Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## sexhead         0.9773        0.8911          0.5783          .    0.0862
## agehead        44.0364       46.4906         -0.2089     0.8353    0.0430
## educhead        2.7409        2.2145          0.1434     1.1516    0.0345
## lnland          0.3243        0.3916         -0.1700     0.5617    0.0346
## vaccess         0.8500        0.8317          0.0513          .    0.0183
## pcirr           0.5532        0.5621         -0.0252     1.1482    0.0266
## rice           10.5316       10.2228          0.1731     1.4065    0.0513
## wheat           7.5296        7.4517          0.1006     0.8055    0.0344
## milk           10.6085       10.9654         -0.1070     0.9682    0.0284
## oil            39.4259       39.3979          0.0064     1.2295    0.0218
## egg             1.8509        1.9781         -0.3710     0.8270    0.0836
##          eCDF Max
## sexhead    0.0862
## agehead    0.1217
## educhead   0.0790
## lnland     0.0639
## vaccess    0.0183
## pcirr      0.0528
## rice       0.1258
## wheat      0.0756
## milk       0.0861
## oil        0.0531
## egg        0.1483
##
## Summary of Balance for Matched Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## sexhead         0.9773        0.9773          0.0000          .    0.0000
## agehead        44.0364       44.2318         -0.0166     0.9766    0.0182
## educhead        2.7409        2.4091          0.0904     1.1420    0.0218
## lnland          0.3243        0.3220          0.0057     0.9383    0.0186
## vaccess         0.8500        0.8500          0.0000          .    0.0000
## pcirr           0.5532        0.5707         -0.0497     1.0054    0.0216
## rice           10.5316       10.4537          0.0437     1.0765    0.0186
## wheat           7.5296        7.4604          0.0893     1.2025    0.0136
## milk           10.6085       10.5470          0.0184     1.0335    0.0167
## oil            39.4259       39.5705         -0.0332     0.9971    0.0084
## egg             1.8509        1.8609         -0.0291     1.0207    0.0076
##          eCDF Max Std. Pair Dist.
## sexhead    0.0000          0.0000
## agehead    0.0682          0.5444
## educhead   0.0545          0.3429
## lnland     0.0773          0.4391
## vaccess    0.0000          0.0000
## pcirr      0.0500          0.1597
## rice       0.0545          0.1443
```

```
## wheat      0.0318          0.2244
## milk       0.0500          0.1254
## oil        0.0227          0.1477
## egg        0.0227          0.1457
##
## Sample Sizes:
##           Control Treated
## All           909     220
## Matched       220     220
## Unmatched     689       0
## Discarded       0       0
```

```r
# creating a dataset of successful matches for male
dta_m4 <- match.data(mod_match4)

# obtaining ATT on the matched data for female
t.test(lexptot ~ dmmfd, data = dta_m4)
```

```
##
##  Welch Two Sample t-test
##
## data:  lexptot by dmmfd
## t = -0.35349, df = 426.43, p-value = 0.7239
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.10445885  0.07261365
## sample estimates:
## mean in group 0 mean in group 1
##        8.396819        8.412741
```

```r
# The results from covariate matching are different than that from propensity score matching.
```

Question 5. C)

```r
for (x in 1:5) {
  print(x)
mod_model <- matchit(data = hh_98_nomiss, dfmfd ~ sexhead + agehead + educhead + lnland + vaccess + pci

summary(mod_model)

# creating a dataset of successful matches for female
dta <- match.data(mod_model)

# obtaining ATT on the matched data for female
t.test(lexptot ~ dfmfd, data = dta)
}
```

```
## [1] 1

## Warning: Fewer control units than treated units; not all treated units will get
## a match.

## [1] 2

## Warning: Fewer control units than treated units; not all treated units will get
## a match.

## [1] 3
```

```
## Warning: Fewer control units than treated units; not all treated units will get
## a match.

## [1] 4

## Warning: Fewer control units than treated units; not all treated units will get
## a match.

## [1] 5

## Warning: Fewer control units than treated units; not all treated units will get
## a match.
```

```r
# The trade offs for using more nearest neighbors is that the coefficients will be biased whereas, trad
```

Question 5. D)

```r
# matching using euclidean distance
mod_euclidean <- matchit(data = hh_98_nomiss, dfmfd ~ sexhead + agehead + educhead + lnland + vaccess +
```

```
## Warning: Fewer control units than treated units; not all treated units will get
## a match.
```

```r
summary(mod_euclidean)
```

```
##
## Call:
## matchit(formula = dfmfd ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, data = hh_98_nomiss,
##     method = "nearest", distance = "euclidean", ratio = 1, estimated = "ATT")
##
## Summary of Balance for All Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## sexhead         0.9008        0.9157         -0.0498          .    0.0149
## agehead        46.0958       45.9195          0.0150     0.7471    0.0312
## educhead        1.7529        2.9457         -0.3956     0.6190    0.0762
## lnland          0.2689        0.5007         -0.6431     0.3527    0.1170
## vaccess         0.8218        0.8502         -0.0741          .    0.0283
## pcirr           0.5765        0.5423          0.1023     1.0330    0.0411
## rice           10.2914       10.2736          0.0114     0.9855    0.0160
## wheat           7.4788        7.4536          0.0306     0.8842    0.0152
## milk           10.8724       10.9219         -0.0152     0.8657    0.0145
## oil            39.6479       39.1309          0.1277     1.0482    0.0185
## egg             1.9698        1.9350          0.0903     1.1365    0.0283
##          eCDF Max
## sexhead    0.0149
## agehead    0.0833
## educhead   0.1567
## lnland     0.1824
## vaccess    0.0283
## pcirr      0.0985
## rice       0.0457
## wheat      0.0272
## milk       0.0408
## oil        0.0838
## egg        0.0542
##
## Summary of Balance for Matched Data:
```

```
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## sexhead         0.9026        0.9157         -0.0439          .    0.0131
## agehead        46.0187       45.9195          0.0084     0.7604    0.0305
## educhead        1.7734        2.9457         -0.3888     0.6117    0.0745
## lnland          0.2805        0.5007         -0.6109     0.3730    0.1105
## vaccess         0.8539        0.8502          0.0098          .    0.0037
## pcirr           0.5945        0.5423          0.1561     1.0311    0.0595
## rice           10.3371       10.2736          0.0407     1.0730    0.0115
## wheat           7.4302        7.4536         -0.0285     0.9102    0.0098
## milk           10.7864       10.9219         -0.0415     0.7984    0.0205
## oil            39.5529       39.1309          0.1043     1.1132    0.0143
## egg             1.9230        1.9350         -0.0313     0.9294    0.0162
##          eCDF Max Std. Pair Dist.
## sexhead    0.0131          0.5326
## agehead    0.0805          0.2530
## educhead   0.1498          0.5440
## lnland     0.1723          1.2399
## vaccess    0.0037          0.5677
## pcirr      0.1255          0.8716
## rice       0.0300          0.7340
## wheat      0.0225          0.8333
## milk       0.0506          0.5033
## oil        0.0674          0.4055
## egg        0.0337          0.6618
##
## Sample Sizes:
##           Control Treated
## All           534     595
## Matched       534     534
## Unmatched       0      61
## Discarded       0       0
```

```r
# creating a dataset of successful matches for female
dta_euclidean <- match.data(mod_euclidean)

# obtaining ATT on the matched data for female
t.test(lexptot ~ dfmfd, data = dta_euclidean)
```

```
##
##  Welch Two Sample t-test
##
## data:  lexptot by dfmfd
## t = -0.42794, df = 1057.1, p-value = 0.6688
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.07505290  0.04817764
## sample estimates:
## mean in group 0 mean in group 1
##        8.447977        8.461414
```

```r
# matching using mahalanobis distance
mod_mhs <- matchit(data = hh_98_nomiss, dfmfd ~ sexhead + agehead + educhead + lnland + vaccess + pcirr
```

```
## Warning: Fewer control units than treated units; not all treated units will get
## a match.
```

```r
summary(mod_mhs)
```

```
##
## Call:
## matchit(formula = dfmfd ~ sexhead + agehead + educhead + lnland +
##     vaccess + pcirr + rice + wheat + milk + oil + egg, data = hh_98_nomiss,
##     method = "nearest", distance = "mahalanobis", ratio = 1,
##     estimated = "ATT")
##
## Summary of Balance for All Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## sexhead         0.9008        0.9157         -0.0498          .    0.0149
## agehead        46.0958       45.9195          0.0150     0.7471    0.0312
## educhead        1.7529        2.9457         -0.3956     0.6190    0.0762
## lnland          0.2689        0.5007         -0.6431     0.3527    0.1170
## vaccess         0.8218        0.8502         -0.0741          .    0.0283
## pcirr           0.5765        0.5423          0.1023     1.0330    0.0411
## rice           10.2914       10.2736          0.0114     0.9855    0.0160
## wheat           7.4788        7.4536          0.0306     0.8842    0.0152
## milk           10.8724       10.9219         -0.0152     0.8657    0.0145
## oil            39.6479       39.1309          0.1277     1.0482    0.0185
## egg             1.9698        1.9350          0.0903     1.1365    0.0283
##          eCDF Max
## sexhead    0.0149
## agehead    0.0833
## educhead   0.1567
## lnland     0.1824
## vaccess    0.0283
## pcirr      0.0985
## rice       0.0457
## wheat      0.0272
## milk       0.0408
## oil        0.0838
## egg        0.0542
##
## Summary of Balance for Matched Data:
##          Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## sexhead         0.9026        0.9157         -0.0439          .    0.0131
## agehead        46.0187       45.9195          0.0084     0.7604    0.0305
## educhead        1.7734        2.9457         -0.3888     0.6117    0.0745
## lnland          0.2805        0.5007         -0.6109     0.3730    0.1105
## vaccess         0.8539        0.8502          0.0098          .    0.0037
## pcirr           0.5945        0.5423          0.1561     1.0311    0.0595
## rice           10.3371       10.2736          0.0407     1.0730    0.0115
## wheat           7.4302        7.4536         -0.0285     0.9102    0.0098
## milk           10.7864       10.9219         -0.0415     0.7984    0.0205
## oil            39.5529       39.1309          0.1043     1.1132    0.0143
## egg             1.9230        1.9350         -0.0313     0.9294    0.0162
##          eCDF Max Std. Pair Dist.
## sexhead    0.0131          0.0689
## agehead    0.0805          0.6585
## educhead   0.1498          0.6633
## lnland     0.1723          0.9619
## vaccess    0.0037          0.0881
```

```
## pcirr      0.1255          0.4782
## rice       0.0300          0.5108
## wheat      0.0225          0.4695
## milk       0.0506          0.4901
## oil        0.0674          0.4694
## egg        0.0337          0.4807
##
## Sample Sizes:
##           Control Treated
## All           534     595
## Matched       534     534
## Unmatched       0      61
## Discarded       0       0
```

```r
# creating a dataset of successful matches for female
dta_mhs <- match.data(mod_mhs)

# obtaining ATT on the matched data for female
t.test(lexptot ~ dfmfd, data = dta_mhs)
```

```
##
##  Welch Two Sample t-test
##
## data:  lexptot by dfmfd
## t = -0.42794, df = 1057.1, p-value = 0.6688
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.07505290  0.04817764
## sample estimates:
## mean in group 0 mean in group 1
##        8.447977        8.461414
```

```r
# The results are somewhat similar. However, using euclidean metric, the results are significant at 5%
```