Problem Set 2

EMPIRICAL /COMPUTER WORK

4. [50 points (all questions here are worth 5points each except part (a) which is 10 points)] Important: As usual, your answer should include a printout (can cut and paste into a file. I will show you how to do this) of relevant calculations on the computer (R or other software output) AND a write up of final answers following the sub parts of the question. The data again are the same as for Problem Set 1. Use these data to answer the following questions.

Rural Atlas Dataset:

```
## 'data.frame':    3278 obs. of  96 variables:
##  $ fips                          : int  0 1000 1001 1003 1005 1007 1009 1011 1013 1015 ...
##  $ State                         : chr  "US" "AL" "AL" "AL" ...
##  $ County                        : chr  "United States" "Alabama" "Autauga" "Baldwin" ...
##  $ MedHHInc                      : int  60336 48193 58343 56607 32490 45795 48253 29113 36842 45937
##  $ PerCapitaInc                  : int  31177 25746 27824 29364 17561 20911 22021 20856 19004 23638
##  $ PovertyUnder18Pct             : num  18.4 24.4 19.3 14.7 50.3 27.3 18.5 48.3 33 24.2 ...
##  $ PovertyAllAgesPct             : num  13.4 16.9 13.4 10.1 33.4 20.2 12.8 34.4 21.3 17.7 ...
##  $ UnempRate2018                 : num  NA 3.9 3.6 3.6 5.2 4 3.5 4.7 4.8 4.7 ...
##  $ UnempRate2017                 : num  NA 4.4 3.9 4.1 5.8 4.4 4 4.9 5.5 5 ...
##  $ UnempRate2016                 : num  NA 5.8 5.1 5.3 8.3 6.4 5.4 6.8 6.9 6.5 ...
##  $ UnempRate2015                 : int  NA 6 5 5 8 6 5 7 7 7 ...
##  $ UnempRate2014                 : num  NA 6.8 5.8 6.1 10.5 7.2 6.1 8.8 8.6 8 ...
##  $ UnempRate2010                 : num  NA 10.5 8.9 10 12.3 11.4 9.8 11.8 13.6 11.4 ...
##  $ UnempRate2007                 : num  NA 4 3.3 3.1 6.3 4.1 3.2 9.4 6.2 3.9 ...
##  $ PctEmpChange1018              : num  NA 7.52 6.76 20.42 -11.37 ...
##  $ PctEmpChange1718              : num  NA 1.47 0.23 2.39 1.39 1.75 1.69 -0.94 0.49 0.77 ...
##  $ PctEmpChange0718              : num  NA 1.11 6.1 12.93 -18.01 ...
##  $ PctEmpChange0710              : num  NA -5.96 -0.62 -6.22 -7.49 ...
##  $ PctEmpAgriculture             : num  1.3 1.112 0.606 1.029 4.224 ...
##  $ PctEmpMining                  : num  0.571 0.46 0.315 0.478 0.169 ...
##  $ PctEmpConstruction            : num  6.35 6.41 5.13 7.61 5.63 ...
##  $ PctEmpManufacturing           : num  10.3 14.2 13.3 9.8 23.8 ...
##  $ PctEmpTrade                   : num  14.1 14.5 12.9 16.5 14.7 ...
##  $ PctEmpTrans                   : num  5.1 5.21 7.44 4.84 6.24 ...
##  $ PctEmpInformation             : num  2.107 1.62 1.671 1.376 0.394 ...
##  $ PctEmpFIRE                    : num  6.58 5.6 5.57 7.61 3.67 ...
##  $ PctEmpServices                : num  49 45.4 41.9 45.9 33.3 ...
##  $ PctEmpGovt                    : num  4.67 5.51 11.15 4.84 7.93 ...
##  $ UnempRate2011                 : num  NA 9.6 8.4 9 11.5 10.5 8.7 11.6 12.5 10.3 ...
##  $ UnempRate2008                 : num  NA 5.7 5.1 4.6 8.8 5.8 4.7 10.5 8.5 5.7 ...
##  $ UnempRate2009                 : num  NA 11 9.7 9.8 14.3 13.3 10 15.6 16.4 11.1 ...
##  $ UnempRate2012                 : num  NA 8 6.9 7.5 11.5 8.5 6.9 10.4 11.5 8.9 ...
##  $ UnempRate2013                 : num  NA 7.2 6.2 6.6 10.2 7.9 6.3 9.4 10.3 8.8 ...
##  $ PopChangeRate1718             : num  0.72 0.26 0.28 2.54 -1.1 -0.69 0.02 -0.37 -1.1 -0.34 ...
##  $ PopChangeRate1018             : num  5.3 2.14 1.55 19.07 -8.96 ...
```

```
##  $ TotalPopEst2018                : int  NA 4887871 55601 218022 24881 22400 57840 10138 19680 11427
##  $ NetMigrationRate1018           : num  2.28 0.84 0.28 18.14 -8.27 ...
##  $ NaturalChangeRate1018          : num  3.02 1.3 1.26 0.93 -0.69 0.19 0.86 0.65 -0.86 -0.52 ...
##  $ Net_International_Migration_Rat: num  2.28 0.77 0.06 0.74 0.14 0.49 0.27 0.38 0.6 0.39 ...
##  $ PopChangeRate0010              : num  9.71 7.48 24.96 29.8 -5.44 ...
##  $ NetMigrationRate0010           : num  NA 3.3 11.9 26.2 -4.8 ...
##  $ NaturalChangeRate0010          : num  NA 3.3 5.46 3.32 2.29 2.1 3.14 3.2 -0.25 1.62 ...
##  $ Immigration_Rate_2000_2010     : num  NA 1.2229 -0.0102 1.5845 1.8284 ...
##  $ PopDensity2010                 : num  87.4 94.4 91.8 114.7 31 ...
##  $ Under18Pct2010                 : num  24 23.7 26.8 23 21.9 ...
##  $ Age65AndOlderPct2010           : num  13 13.8 12 16.8 14.2 ...
##  $ WhiteNonHispanicPct2010        : num  63.8 67 77.2 83.5 46.8 ...
##  $ BlackNonHispanicPct2010        : num  12.21 26.04 17.58 9.31 46.69 ...
##  $ AsianNonHispanicPct2010        : num  4.69 1.11 0.86 0.74 0.39 0.1 0.2 0.18 0.83 0.7 ...
##  $ NativeAmericanNonHispanicPct201: num  0.73 0.54 0.4 0.63 0.22 0.28 0.5 0.18 0.28 0.4 ...
##  $ HispanicPct2010                : num  16.35 3.88 2.4 4.38 5.05 ...
##  $ MultipleRacePct2010            : num  2.92 1.49 1.59 1.49 0.94 0.89 1.19 0.79 0.77 1.67 ...
##  $ NonHispanicWhitePopChangeRate00: num  1.16 2.51 21.05 25.92 -13.19 ...
##  $ NonHispanicBlackPopChangeRate00: num  11.01 8.2 29.17 18.17 -4.11 ...
##  $ NonHispanicAsianPopChangeRate00: num  42.9 70.8 140.7 152.3 28.9 ...
##  $ NonHispanicNativeAmericanPopCha: num  8.61 19.84 16.67 52.19 -49.58 ...
##  $ HispanicPopChangeRate0010      : num  43 145 115 224 190 ...
##  $ MultipleRacePopChangeRate0010  : num  32 61.3 114.6 85.7 21.7 ...
##  $ ForeignBornPct                 : num  13.4 3.46 2.11 3.25 2.66 ...
##  $ ForeignBornEuropePct           : num  1.494 0.375 0.338 0.742 0.317 ...
##  $ ForeignBornMexPct              : num  3.587 1.078 0.507 0.716 1.313 ...
##  $ NonEnglishHHPct                : num  4.465 1.149 0.75 0.901 0.816 ...
##  $ Ed1LessThanHSPct               : num  12.69 14.68 12.3 9.79 26.91 ...
##  $ Ed2HSDiplomaOnlyPct            : num  27.3 30.9 33.6 27.8 35.5 ...
##  $ Ed3SomeCollegePct              : num  20.8 21.7 21.5 22 18.2 ...
##  $ Ed4AssocDegreePct              : num  8.28 8.16 7.63 9.62 7.34 ...
##  $ Ed5CollegePlusPct              : num  30.9 24.5 25 30.7 12 ...
##  $ AvgHHSize                      : num  2.63 2.55 2.59 2.63 2.54 2.97 2.76 2.74 2.81 2.49 ...
##  $ FemaleHHPct                    : num  12.7 14.78 11.41 9.43 19.37 ...
##  $ HH65PlusAlonePct               : num  10.5 11.2 11.4 12.8 13.9 ...
##  $ OwnHomePct                     : num  63.8 68.6 73.3 72.9 62.5 ...
##  $ TotalPop25Plus                 : int  216271644 3276637 36757 143022 18434 15859 39475 7377 13958
##  $ ForeignBornAfricaPct           : num  0.6347 0.17 0 0.0713 0 ...
##  $ TotalPopACS                    : int  321004407 4850771 55036 203360 26201 22580 57667 10478 2012
##  $ TotalOccHU                     : int  118825921 1856695 21054 76133 9191 6916 20690 3670 7050 450
##  $ ForeignBornAsiaPct             : num  4.085 1.091 0.849 0.595 0.424 ...
##  $ TotalHH                        : int  118825921 1856695 21054 76133 9191 6916 20690 3670 7050 450
##  $ ForeignBornCentralSouthAmPct   : num  5.557 1.597 0.654 1.266 1.863 ...
##  $ ForeignBornCaribPct            : num  1.2993 0.1378 0.1581 0.4406 0.0305 ...
##  $ TotalPop2010                   : int  308745538 4779736 54571 182265 27457 22915 57322 10914 2094
##  $ LandAreaSQMiles2010            : num  3531905 50645 594 1590 885 ...
##  $ Net_International_Migration_200: num  NA 54447 -4.5 2239.5 530.5 ...
##  $ TotalPopEst2012                : int  313993272 4815564 54936 190143 27174 22664 57570 10607 2068
##  $ TotalPopEst2013                : int  316234505 4830460 54713 194886 26944 22516 57611 10551 2037
##  $ TotalPopEst2010                : int  309338421 4785448 54754 183111 27330 22872 57373 10878 2094
##  $ TotalPopEst2014                : int  318622525 4842481 54876 199189 26758 22541 57521 10665 2034
##  $ TotalPopEst2011                : int  311644280 4798834 55208 186540 27350 22747 57554 10677 2087
##  $ Net_International_Migration_201: int  7042455 36623 32 1364 38 111 156 41 125 460 ...
##  $ NaturalChange1018              : int  9338302 62441 692 1697 -188 43 491 71 -180 -612 ...
```

```
##  $ TotalPopEst2015               : int  321039839 4853160 54838 202995 26294 22562 57522 10400 2017
##  $ TotalPopEst2016               : int  323405935 4864745 55242 207712 25819 22576 57517 10381 2002
##  $ TotalPopEst2017               : int  325719178 4875120 55443 212619 25158 22555 57827 10176 1989
##  $ NetMigration1018              : int  7042455 39982 155 33214 -2261 -515 -24 -811 -1082 -3588 ...
##  $ TotalPopEstBase2010           : int  308758105 4780138 54574 182264 27457 22920 57321 10911 2094
##  $ Metro2013                     : int  NA NA 1 1 0 1 1 0 0 1 ...
##  $ Metro2003                     : int  NA NA 1 0 0 1 1 0 0 1 ...
```

(a) Run a regression to determine the impact of the 2013 unemployment rate (UnempRate2013) on the per capita income (PerCapitaInc) in a county. What is the estimated slope? Explain what this number means in words in terms of the unemployment rate and in terms of per capita income. Also indicate if the relationship is statistically significant at the 10%, 5%, and 1% levels. For this first pass, use homoskedastic standard errors.

```
model1 <- lm(formula = UnempRate2013 ~ PerCapitaInc, data = rural_atlas_merged)

summary(model1)
```

```
##
## Call:
## lm(formula = UnempRate2013 ~ PerCapitaInc, data = rural_atlas_merged)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.6689 -1.7598 -0.1889  1.3143 16.5830
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.415e+01  1.835e-01    77.1   <2e-16 ***
## PerCapitaInc -2.541e-04  6.905e-06   -36.8   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.635 on 3269 degrees of freedom
##   (7 observations deleted due to missingness)
## Multiple R-squared:  0.2929, Adjusted R-squared:  0.2927
## F-statistic:  1354 on 1 and 3269 DF,  p-value: < 2.2e-16
```

(b) Re-run the regression from part (a) but this time use heteroskedastic standard errors. Are your coefficients the same as in part (a)? Why? Are your standard errors (of your betas) the same as in part (a)? Why?

(c) Run the same regression as in part (b) but now also include the additional regressors percentage of the population that is college-educated (Ed5CollegePlusPct), percentage of the population that is black (BlackNonHispanicPct2010), and percentage of the population that is Hispanic (HispanicPct2010). Now, what is the estimated impact of unemployment rate in 2013 on per capita income? Also indicate if the relationship is statistically significant at the 10%, 5%, and 1% levels? Make sure that you are using heteroskedastic standard errors.

(d) Provide economic/econometric intuition as to why the impact of the unemployment rate's impact on per capita income changed between parts (b) and (c). Note that I am asking you to think about the context (and hence the "story" behind these data).

(e) Construct a 95% confidence interval for the slope coefficient on UnempRate2013 in (c). Write out your calculations. Clearly indicate how this confidence interval relates to whether UnempRate2013 is statistically significant or not in this context by relating your answer to your constructed confidence interval.

(f) You recall from problem set 1 that both the means of per capita income and of unemployment rate in 2013 are quite different across metro and nonmetro areas. You therefore want to explore this in more detail. Run the regression from (c) using only metro areas in 2013 (Metro2013==1). [Hint: You need to restrict the data based on a criterion before running the regression.] Now, what is the estimated effect of the 2013 unemployment rate on per capita income and also indicate if the relationship is statistically significant at the 10%, 5%, and 1% levels? Make sure that you are using heteroskedastic standard errors.

(g) Now, run the regression from (c) using only non-metro areas in 2013 (Metro2013==0). [Hint: You need to restrict the data based on a criterion before running the regression.] Now, what is the estimated effect of the 2013 unemployment rate on per capita income and also indicate if the relationship is statistically significant at the 10%, 5%, and 1% levels? Make sure that you are using heteroskedastic standard errors.

(h) What did you learn from the comparison between results in parts (f) and (g)? Explain your answer. Note that I again am asking you to think about the context (and hence the "story" behind these data).

(i) Return to the full sample. Now, run a regression to determine the impact of changing the percentage of the population which is college educated (Ed5CollegePlusPct) on the per capita income (PerCapitaInc) in a county. Include controls for the unemployment rate in 2010 (UnempRate2010), percentage of the population that is black (BlackNonHispanicPct2010), and now also include a dummy variable for metro status (Metro2013). Now, what is the estimated impact of percentage with a college education on per capita income? Also indicate if the relationship is statistically significant at the 10%, 5%, and 1% levels? Make sure that you are using heteroskedastic standard errors.