Week 6

Zai Rutter

2/21/2022

```r
#######################

library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

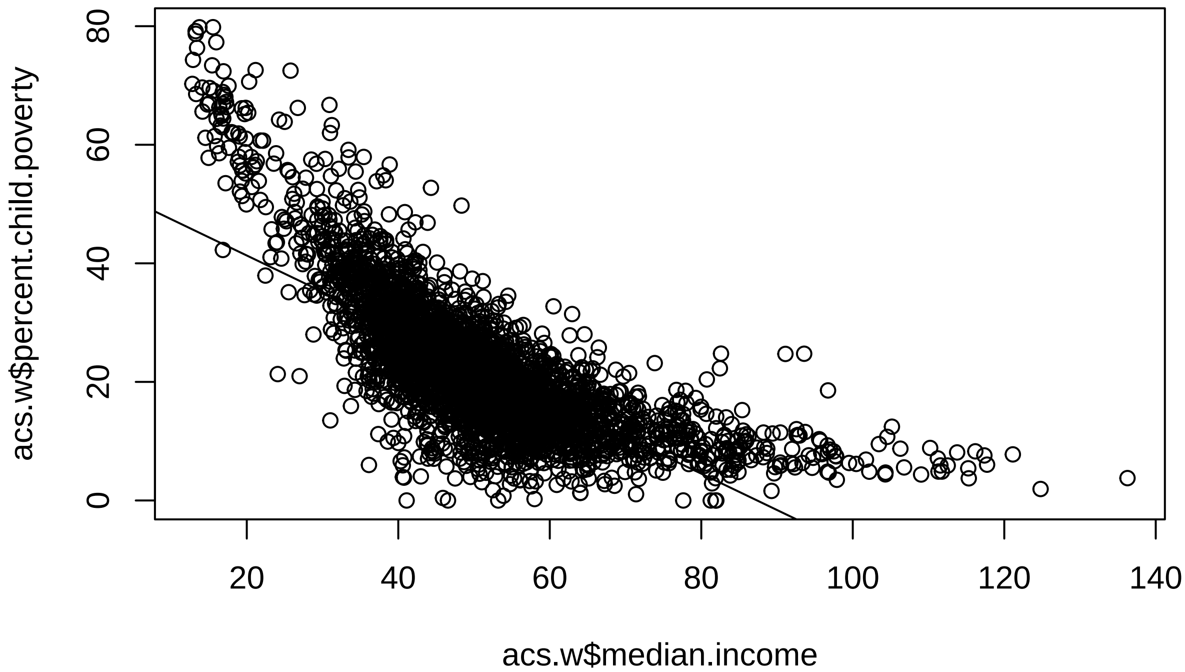
```r
library(readxl)
load("~/Documents/Upenn/Data 310/Week 6/Homework/ACSCountyData.Rdata")
acs.w<-acs
acs.w$census.region <- as.factor(acs.w$census.region)
#######################

## Question 1

## A,B,C,D

acs.w$median.income <- acs$median.income/1000

plot(acs.w$median.income, acs.w$percent.child.poverty)
abline(lm(percent.child.poverty ~ median.income, data =acs.w))
```




```r
# There appears to be a negative relationship, as median income increases, child
# poverty decreases
# R squared - 55% of the variation is explained in the data
p <- ggplot(acs.w, aes(x=median.income, y=percent.child.poverty)) + geom_point() +
  ylim(0,100) +
  labs(x="Median Income (Thousands)", y = "Percent Childen in Poverty") +
  geom_smooth(method = lm, formula = y ~ poly(x, 2), se = FALSE)


## E
## In this new regression with a second order polynomial term, what is the the
## effect of an additional $1000 in median income when median income is at $30k?

## What is the the effect of an additional $1000 in median income when median
## income is at $100k? Does this make theoretical sense?

E<-lm(acs.w$percent.child.poverty ~ poly(acs.w$median.income,2, raw=T), data = acs.w)
coef(E)[2] + 2*(30 * coef(E)[3])
```

```
## poly(acs.w$median.income, 2, raw = T)1
##                               -1.227771
```

```r
## An additional 1000 dollars in median income relates to a .6% decrease in child poverty
## but at 30k, a thousand dollars relates to a 1.2% decrease in child poverty
## This steepening relationship may be a result of the fact that at poorer levels, more money
## has more impact, where a 1k increase on higher median incomes makes up less a percentage of thier in
## leading to a lesser impact, where it averages out to be a total .6% decrease.




## F
```

```r
## A possible confounding variable to this relationship is the unemployment rate,
## which may affect both the median income of a county and the percent of children
## living in poverty.
## Use the cor() function to investigate the relationships
## between median income, unemployment, and child poverty. Based on the pattern of
## correlations,

## what is likely to happen to the coefficient on median.income if you add
## unemployment rate to the first regression model (the one without the polynomial terms)?


cor(acs[,c('percent.child.poverty','median.income',
          'unemployment.rate')],use='pairwise.complete.obs')
```

```
##                       percent.child.poverty median.income unemployment.rate
## percent.child.poverty             1.0000000    -0.7448909         0.6814938
## median.income                    -0.7448909     1.0000000        -0.4990148
## unemployment.rate                 0.6814938    -0.4990148         1.0000000
```

```r
# Unemployment apeears to have a positive relation with percent in poverty (as unemployment grows
# so does child poverty). Unemployment also has a negative relationship ~ -.5 with median income,
# ( As unemployment increases, median income decreases)

## G
# Run this regression with unemployment rate and median income (no polynomial terms),
# and determine the degree to which the coefficient on median.income changes.
# Interpret the other coefficients in the model as well, being sure to adjust
# your language to the fact that there are now multiple indpeendent variables.

summary(lm(percent.child.poverty ~ median.income, data =acs.w))
```

```
##
## Call:
## lm(formula = percent.child.poverty ~ median.income, data = acs.w)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.376  -4.879  -0.690   4.021  35.773
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   53.584098   0.511495  104.76   <2e-16 ***
## median.income -0.613294   0.009686  -63.31   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.943 on 3216 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.5549, Adjusted R-squared:  0.5547
## F-statistic:  4009 on 1 and 3216 DF,  p-value: < 2.2e-16
```

```r
summary(lm(acs.w$percent.child.poverty ~ acs.w$median.income + acs.w$unemployment.rate,
           data = acs.w))
```

```
##
## Call:
## lm(formula = acs.w$percent.child.poverty ~ acs.w$median.income +
##     acs.w$unemployment.rate, data = acs.w)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.365  -3.902  -0.518   3.228  41.328
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             36.662660   0.638421   57.43   <2e-16 ***
## acs.w$median.income     -0.443836   0.009438  -47.03   <2e-16 ***
## acs.w$unemployment.rate  1.369099   0.038039   35.99   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.707 on 3215 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.6827, Adjusted R-squared:  0.6825
## F-statistic:  3459 on 2 and 3215 DF,  p-value: < 2.2e-16
```

```r
G<- lm(acs.w$percent.child.poverty ~ acs.w$median.income + acs.w$unemployment.rate,
   data = acs.w)

coef(G)
```

```
##             (Intercept)     acs.w$median.income acs.w$unemployment.rate
##              36.6626595              -0.4438365               1.3690988
```

```r
# With the first equation, child poverty had a -.6% realtion with Median income.
# when we add the unemployment rate we see that child poverty had a -.44% realtion with Median income,
# Indicating that holding all things constant when unemployment is considered, child poverty has less
# negative relationship with median income


## H
# Another possible confounding variable is the census region people are living in.
# For example, living in the south could be associated with both lower average
# incomes and more child poverty. Create an indicator variable for the 4 census
# regions (or change the variable into a factor variable) and then

# re-estimate the regression with median income and unemployment to take into
# account what region each county is in. Interpret the coefficients from this regression.
table(acs.w$census.region)
```

```
##
##   midwest northeast     south      west
##      1055       217      1422       448
```

```r
summary(lm(acs.w$percent.child.poverty ~ acs.w$median.income + acs.w$unemployment.rate +
           acs.w$census.region, data=acs.w))
```

```
##
## Call:
## lm(formula = acs.w$percent.child.poverty ~ acs.w$median.income +
##     acs.w$unemployment.rate + acs.w$census.region, data = acs.w)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.492  -3.760  -0.362   3.239  34.828
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 33.280821   0.635661  52.356  < 2e-16 ***
## acs.w$median.income         -0.397512   0.009289 -42.794  < 2e-16 ***
## acs.w$unemployment.rate      1.193781   0.045132  26.451  < 2e-16 ***
## acs.w$census.regionnortheast  1.825657   0.474129   3.851  0.00012 ***
## acs.w$census.regionsouth     3.546065   0.269091  13.178  < 2e-16 ***
## acs.w$census.regionwest      0.622232   0.358791   1.734  0.08297 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.204 on 3134 degrees of freedom
##   (80 observations deleted due to missingness)
## Multiple R-squared:  0.6364, Adjusted R-squared:  0.6358
## F-statistic:  1097 on 5 and 3134 DF,  p-value: < 2.2e-16
```

```r
# ~64% of the variance can be described in our model
# Median income negatively affects child poverty in the midwest region by -.39%
# Unemployet positivley affects child poverty in the midwest region by 1.19%
# in relation to the midwest, child poverty is postively affected more by all other regions


## I
## It's possible that the effect of median income is different conditional on
## whether a county is urban or not.

## Create an indicator variable for whether a
## county is urban (population density greater or equal to 1000) or not.

acs.w$urban <- acs.w$population.density >= 1000

## Interact this variable with median income in the regression with unemployment
## rate and census region indicators. Interpret the coefficients on median income,
## the urban indicator, and the interaction term.


summary(lm(percent.child.poverty ~ median.income *urban + unemployment.rate +
             census.region, data=acs.w))
```

```
##
```

```
## Call:
## lm(formula = percent.child.poverty ~ median.income * urban +
##     unemployment.rate + census.region, data = acs.w)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.320  -3.650  -0.318   3.244  33.928
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              35.86332    0.67829  52.873  < 2e-16 ***
## median.income            -0.44208    0.01019 -43.384  < 2e-16 ***
## urbanTRUE                -6.34458    1.75620  -3.613 0.000308 ***
## unemployment.rate         1.13193    0.04483  25.247  < 2e-16 ***
## census.regionnortheast    1.29103    0.46993   2.747 0.006044 **
## census.regionsouth        3.29040    0.26607  12.367  < 2e-16 ***
## census.regionwest         0.63727    0.35325   1.804 0.071327 .
## median.income:urbanTRUE   0.16178    0.02537   6.376 2.09e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.106 on 3132 degrees of freedom
##   (80 observations deleted due to missingness)
## Multiple R-squared:  0.648,  Adjusted R-squared:  0.6472
## F-statistic: 823.6 on 7 and 3132 DF,  p-value: < 2.2e-16
```

```
## When rual median income has a negative .44 affect on child poverty.
## The 'urbanTrue' coef is not relevant because it is result when meidan income is NULL
# and that makes no sesne. So We use the last coef 'median.income:urbanTRUE' to see
# the affect when meidan income and urban have affect on child poverty which is a pos .16%
```