Analysis of county demographic information: predicting average income per person

Julia Keating
jhkeatin@andrew.cmu.edu

Notes: The numbering of my tables/figures is off for some reason so I will need to fix that later. This is a super rough draft so any feedback is appreciated!

1 Abstract

In this paper we will analyze how average income per person is related to other variables associated with a county's economic, health and social well-being. Our data consists of selected county demographic information for 440 of the most populous counties in the United States. We will develop a model to predict per capita income from total number of crimes, as well as the best model to predict per capita income from all variables in the dataset using a variety of methods such as the all subsets method. Models will be compared using partial F tests and assessed for validity and goodness of fit through diagnostic plots, R^2 values, VIF's and BIC values. Our final models suggest that the relationship between per capita income and total number of crimes is different in different regions of the country.

2 Introduction

Social scientists are interested in looking at historical county demographic information to learn how average income per person was related to other variables associated with the county's economic, health and social well-being. To provide insight about these relationships, we will answer the following research questions:

1. Looking at the data one pair of variables at a time, which variables seem to be related to which other variables in the data? Which are not? Are all of the relationships what a reasonable person would expect, or are there some surprises? Can you explain these findings in terms of the meanings of the variables?

2. There is a theory that, if we ignore all other variables, per-capita income should be related to crime rate, and that this relationship may be different in different regions of the country (Northeast, North-central, South, and West). What do the data say? Does it matter if you use number of crimes, or(number of crimes)/(population), in your analysis?

3. Find the best model predicting per-capita income from the other variables (including possible transformations, interactions, etc.). Here "best" means a good compromise between

    - Best reflects the social science and the meaning of the variables
    - Best satisfies modeling assumptions
    - Is most clearly indicated by the data
    - Can be explained to someone who is more interested in social, economic and health factors than in mathematics and statistics.

4. A county is a governmental unit in the United States that is bigger than a city but smaller than a state.There are 50 states in the US, plus the District of Columbia, which is usually coded as a 51st state in data like this. There are 48 states represented in the data. There are approximately 3000 counties in the US, and 373 represented in the data set. Should we be worried about either the missing states or the missing counties? Why or why not?

3 Data

Our data is taken from Kutneret al. (2005), and consists of selected county demographic information for 440 of the most populous counties in the United States. Counties with missing data were removed from the dataset. Our dataset contains 17 variables which are defined in Table 1. Our response variable for this analysis is per capita income. Table 2 contains a summary of the quantitative variables except for identification number which is not useful in our analysis. Tables 3, 4 and 5 contain summaries of the 3 categorical variables. We note that the county and state variables contain a large number of unique values (373 and 48, respectively). Thus, these variables are not very useful and we will exclude them in this analysis.

Variable number	Variable name	Definition
1	Identification number	1-440
2	County	County name
3	State	Two-letter state abbreviation
4	Land area	Land area (square miles)
5	Total population	Estimated 1990 population
6	Percent of population aged 18-34	Percent of 1990 CDI population aged 18-34
7	Percent of population 65 or older	Percent of 1990 CDI population aged 65 or older
8	Number of active physicians	Number of professionally active nonfederal physicians during 1990
9	Number of hospital beds	Total number of beds, cribs and bassinets during 1990
10	Total serious crimes	Total number of serious crimes in 1990, including murder, rape, robbery, aggrevated assault, burglary, larceny-theft, and motor vehicle theft, as reported by law enforcement agencies
11	Percent high school graduates	Percent of adult population (persons 25 years old or older) who completed 12 or more years of school
12	Percent bachelor's degrees	Percent of adult population (persons 25 years old or older) with bachelor's degree
13	Percent below poverty level	Percent of 1990 CDI population with income below poverty level
14	Percent unemployment	Percent of 1990 CDI population that is unemployed
15	Per capita income	Per-capita income (i.e. average income per person) of 1990 CDI population (in dollars)
16	Total personal income	Total personal income of 1990 CDI population (in millions of dollars)
17	Geographic region	Geographic region classification used by the US Bureau of the Census, NE (northeast region of US), NC (north-central region of the US), S (southern region of the US), and W (Western region of the US)

Table 1: Variable definitions for CDI data from Kutner et al. (2005)

Variable Name	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Land area	15.0	451.2	656.5	1041.4	946.8	20062.0
Total population	100043	139027	217280	393011	436064	8863164
Percent of population aged 18-34	16.40	26.20	28.10	28.57	30.02	49.70
Percent of population aged 65 or older	3.000	9.875	11.750	12.170	13.625	33.800
Number of active physicians	39.0	182.8	401.0	988.0	1036.0	23677.0
Number of hospital beds	92.0	390.8	755.0	1458.6	1575.8	27700.0
Total serious crimes	563	6220	11820	27112	26280	688936
Percent high school graduates	46.60	73.88	77.70	77.56	82.40	92.90
Percent bachelor's degrees	8.10	15.28	19.70	21.08	25.32	52.30
Percent below poverty level	1.400	5.300	7.900	8.721	10.900	36.300
Percent unemployment	2.200	5.100	6.200	6.597	7.500	21.300
Per capita income	8899	16118	17759	18561	20270	37541
Total personal income	1141	2311	3857	7869	8654	184230

Table 2: Summary of quantitative variables in dataset, excluding id

Figure 3 contains histograms for the quantitative variables excluding identification number. Note that the data for land.area, pop, doctors, hosp.beds, crimes, tot.income and per.cap.income are strongly right-skewed, which suggests we may want to apply transformations these variables later in our analysis.
Figure 3 displays the correlation matrix of the quantitative variables, excluding identification number, as a heatmap. We note that total personal income and total population are highly correlated, which is not surprising because we would generally expect a larger population to have a larger total personal income. We also note that total personal income, total population, active physicians, number of hospital beds and total serious crimes all appear to be fairly highly correlated with eachother. Finally, note that per capita income, our response variable, is not very highly correlated with any variables in the plot. Per capita income is most

Frequency	Number of unique counties
1	334
2	23
3	10
4	3
5	1
6	1
7	1

Table 3: Summary of counties in the dataset

State	Count	State	Count	State	Count	State	Count	State	Count
AL	7	HI	3	MI	18	NM	2	TN	8
AR	2	ID	1	MN	7	NV	2	TX	28
AZ	5	IL	17	MO	8	NY	22	UT	4
CA	34	IN	14	MS	3	OH	24	VA	9
CO	9	KS	4	MT	1	OK	4	VT	1
CT	8	KY	3	NC	18	OR	6	WA	10
DC	1	LA	9	ND	1	PA	29	WI	11
DE	2	MA	11	NE	3	RI	3	WV	1
FL	29	MD	10	NH	4	SC	11		
GA	9	ME	5	NJ	18	SD	1		

Table 4: Summary of states in the dataset

Region	Count
NC	108
NE	103
S	152
W	77

Table 5: Summary of regions in the dataset

strongly correlated with percent bachelor's degrees, percent high school graduates, and percent below poverty level.

4 Methods

We begin by applying log transformations to the variables land.area, pop, doctors, hosp.beds, crimes, tot.income and per.cap.income to address the extreme right skew in the distribution of these variables as shown in Figure 3.

To determine if per capita income should be related to crime rate, we first fit linear models which predict log(per capita income) from log(total serious crimes) as well as from log(total serious crimes) and geographic region, with and without an interaction term between log(total serious crimes) and geographic region. Next, to determine whether using crime rate, where crime rate equals total serious crimes divided by total population, instead of total serious crimes made a difference, we fit linear models which predict log(per capita income) from log(crime rate) in addition to log(crime rate) and geographic region, with and without an interaction term between log(crime rate) and geographic region. We assessed each of the six models using residual plots and compared models using partial F tests in addition to AIC and BIC values to choose the best model.

We then shifted our focus in order to find the best model predicting log(per capita income) using all predictors. We excluded the variables identification number, county and state as they are not useful in our analysis. We also excluded the variables total population and total personal income which are directly related to our response variable, per capita income.

We started by using the all subsets method on all remaining predictors, excluding region temporarily, with transformations as described previously. We fit a new model containing region, the predictors in the model chosen by the all subsets method, and interaction terms between all of these predictors and region. We

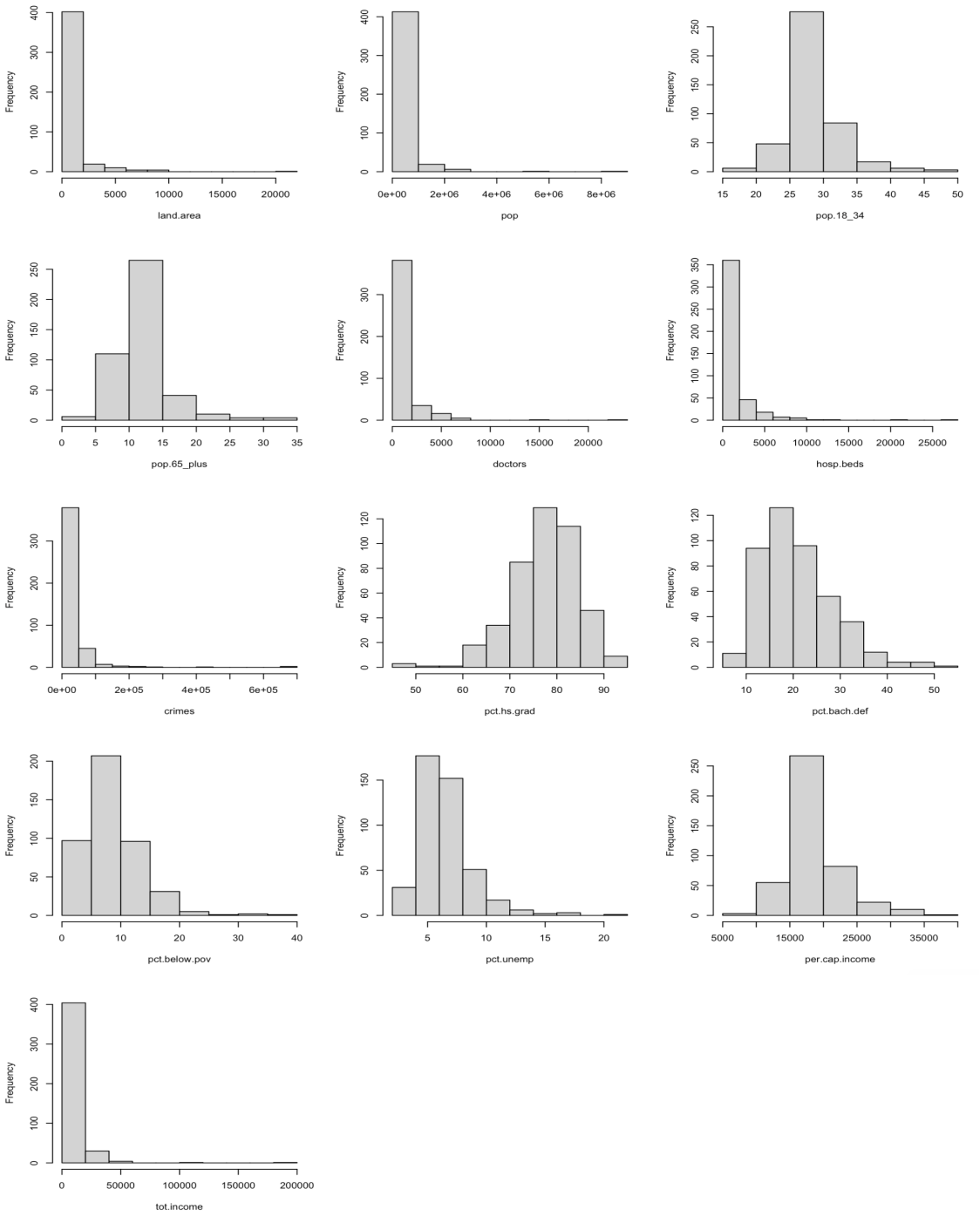

Figure 1: Histograms of quantitative variables in dataset, excluding id

removed interaction terms from the model which did not have statistically significant coefficients for any factor. We compared the resulting model to the model chosen by the all subsets method using a partial F test. We repeated this same process with stepwise regression using the AIC criterion and using the BIC criterion. The final three models were assessed using VIF values and residual plots. Model characteristics were taken into account in addition to AIC and BIC values to choose our final model.

5 Results

We fit a linear model which predicts log(per capita income) from log(crimes), a linear model which predicts log(per capita income) from log(crimes) and region, and a linear model which predicts log(per capita income) from log(crimes), region and the interaction between log(crimes) and region. We then fit the same three models but using log(crime rate) instead of log(crimes), where crime rate is crimes/total.pop. These models

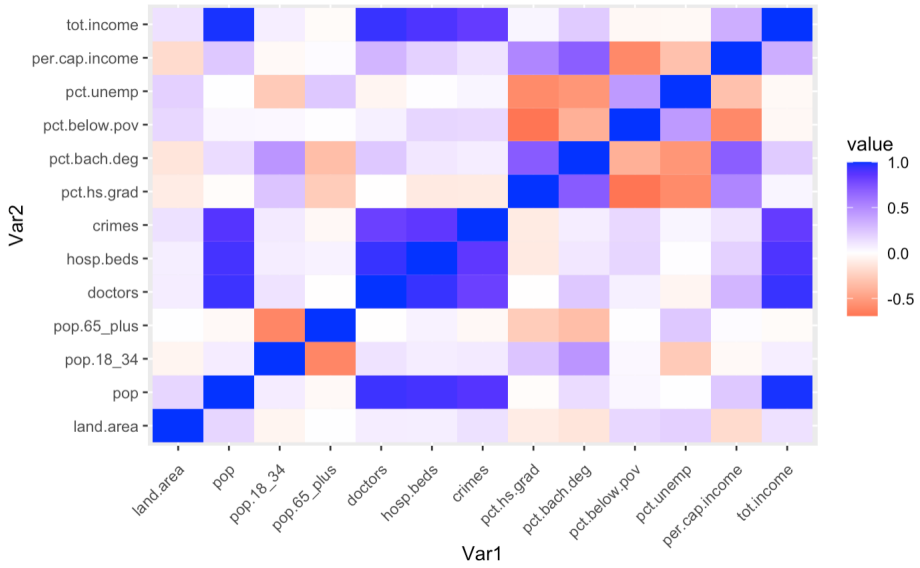

Figure 2: Heatmap of correlation matrix of quantitative variables, excluding id

are shown below.

$$\log(\text{per.cap.income}) \sim \log(\text{crimes}) \tag{1}$$
$$\log(\text{per.cap.income}) \sim \log(\text{crimes}) + \text{region} \tag{2}$$
$$\log(\text{per.cap.income}) \sim \log(\text{crimes}) + \text{region} + \log(\text{crimes}) * \text{region} \tag{3}$$
$$\log(\text{per.cap.income}) \sim \log(\text{crime.rate}) \tag{4}$$
$$\log(\text{per.cap.income}) \sim \log(\text{crime.rate}) + \text{region} \tag{5}$$
$$\log(\text{per.cap.income}) \sim \log(\text{crime.rate}) + \text{region} + \log(\text{crime.rate}) * \text{region} \tag{6}$$

We will first consider models (1), (2) and (3). All three models appear to be equally valid models according to their residual plots. All three models are statistically significant. However, model (3) contains several coefficients which are not statistically significant, unlike models (1) and (2). We also notice that model (1) has a much smaller R^2 value than models (2) and (3), which have very similar R^2 values. A partial F test between models (1) and (2) indicates that we have significant evidence that model (2) is a better fit for the data than model (1), i.e. the region term significantly improves the model. A partial F test between models (2) and (3) indicates that we do not have significant evidence that model (3) is a better fit for the data than model (2). Thus, we choose model (2) as the best model among models (1), (2) and (3).

We will now consider models (4), (5) and (6). Residual plots suggest that model (4) is a valid model, however we see a somewhat concerning clustering pattern between the residuals and fitted values for models (5) and (6). We note that models (5) and (6) are statistically significant while model (4) is not. Also, we note that all coefficients except the intercept in models (4) and (6) are not statistically significant, while all coefficients in model (5) are significant except for one. Model (4) has a very small R^2 value compared to models (5) and (6) which have very similar R^2 values. A partial F test between models (4) and (5) indicates that we have significant evidence that model (5) is a better fit for the data than model (4), i.e. the region term significantly improves the model. A partial F test between models (5) and (6) indicates that we do not have significant evidence that model (6) is a better fit for the data than model (5). Thus, we choose model (5) as the best model among models (4), (5) and (6).

Lastly, we compare models (2) and (5). Model (2) is a statistically significant model and all coefficients in the model are significant as well. Model (2) is a valid model with an R^2 value of 0.1959. Model (5) is a statistically significant model and all coefficients in the model are significant except for one. Model (5) does not appear to be a completely valid model due to the clustering trend between residuals and fitted values, and the value of R^2 for this model is only 0.08814. For these reasons. We choose model (2) as the best model of the six models fit. The residual plots of this model are shown in Figure 5
Next, we applied the all subsets method to all variables except for id, county, state, log.pop and log.tot.income, including region. The model with the lowest BIC value produced by this method is as follows:

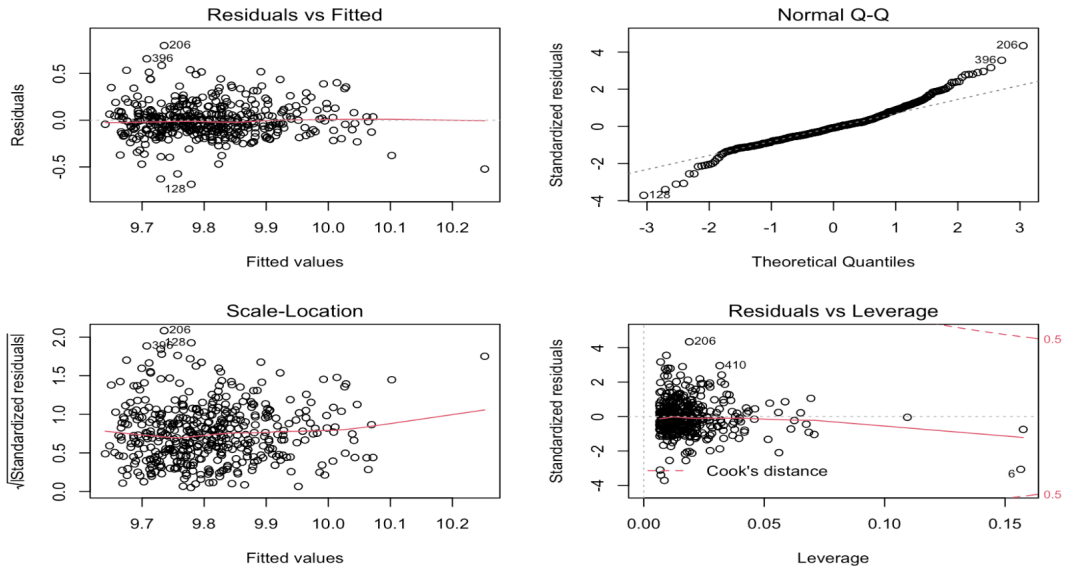

Figure 3: Residual plots for model (2)

$$\log(\text{per.cap.income}) \sim \log(\text{land.area}) + \text{pop.18\_34} + \log(\text{doctors}) + \text{pct.hs.grad} + \text{pct.bach.deg} + \text{pct.below.pov} + \text{pct.unemp} \quad (7)$$

We then fit a model including all predictors in model (7), region, and the interactions between region and all predictors in model (7). We removed all interaction terms which did not contain a significant coefficient for at least one factor of region. The resulting model is as follows:

$$\log(\text{per.cap.income}) \sim \log(\text{land.area}) + \text{pop.18\_34} + \log(\text{doctors}) + \text{pct.hs.grad} + \text{pct.bach.deg} + \text{pct.below.pov} + \text{pct.unemp} + \text{region} + \text{pct.hs.grad} * \text{region} + \text{pct.below.pov} * \text{region} + \text{pct.unemp} * \text{region} \quad (8)$$

A partial F test between models (7) and (8) indicates that we have significant evidence that model (8) is a better fit for the data than model (7), i.e. the region term and the included interactions with region significantly improve the model. The estimated coefficients and their associated standard errors are provided in Table 5 Residual plots for model (8) are shown in Figure 5. In the Residuals vs Fitted plot, we see that the residuals have mean approximately zero and no obvious trends with the fitted values. The standardized residuals are fairly normal according to the Normal Q-Q plot. The Scale-Location plot shows that the square-root of the standardized residuals have fairly constant variance and no obvious trends with the fitted values. Lastly, the Residuals vs Leverage plot does not indicate any high-leverage outliers. Thus, we conclude that model (8) is a valid model. Model (8) is also a good fit for the data because its R^2 value is equal to 0.8615, which is very close to 1.

6 Discussion

We saw in the Data section that total personal income and total population are highly correlated, which is not surprising because we would generally expect a larger population to have a larger total personal income. We also found that total personal income, total population, active physicians, number of hospital beds and total serious crimes all appear to be fairly highly correlated with eachother. This is also not entirely surprising because we would generally expect more populous counties to have more doctors, larger hospitals and a greater number of crimes. Our response variable, per capita income, did not appear to be very highly correlated with any variables, but was most strongly correlated with percent bachelor's degrees, percent high school graduates, and percent below poverty level. This is not surprising because we would generally expect more educated counties to have higher per capita incomes and counties with a high percent of the population with income below the poverty level to have a lower per capita income.

Coefficient	Estimate	Standard Error
Intercept	10.242123935	0.2176557
log(land.area)	-0.038173762	0.0053996
pop.18_34	-0.014934657	0.0010897
log(doctors)	0.057228443	0.0040082
pct.hs.grad	-0.004353194	0.0024515
pct.bach.deg	0.015630966	0.0009715
pct.below.pov	-0.025202882	0.0032612
pct.unemp	0.019739969	0.0046254
regionNE	-0.052006957	0.2707173
regionS	-0.038971766	0.2383516
regionW	1.391048448	0.3408962
pct.hs.grad*regionNE	0.001768418	0.0029293
pct.hs.grad*regionS	0.001152511	0.0025618
pct.below.pov*regionNE	-0.014147323	0.0035826
pct.hs.grad*regionW	-0.001517033	0.0046143
pct.below.pov*regionS	0.007018461	0.0035199
pct.below.pov*regionW	-0.013791967	0.0051811
pct.unemp*regionNE	-0.012984072	0.0070423
pct.unemp*regionS	-0.023113781	0.0061365
pct.unemp*regionW	-0.021735737	0.0065225

Table 6: Estimated coefficients for model (8)

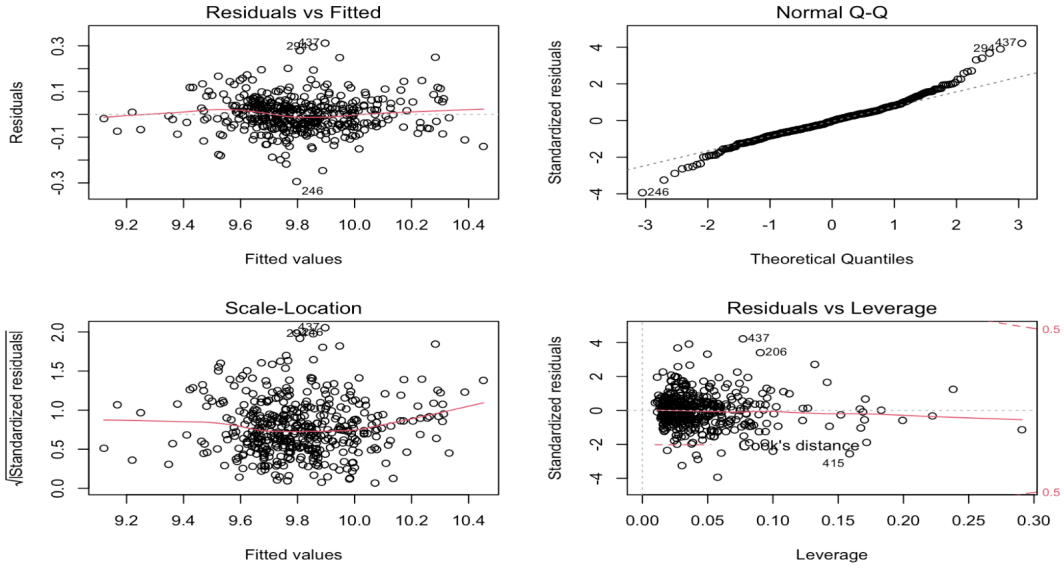

Figure 4: Residual plots for model (8)

We found that the best model predicting per capita income from total number of crimes included region but not the interaction between total number of crimes and region. Using crime rate instead of total number of crimes in the model did not make a significant difference. Our model suggests that the relationship between total number of crimes and per capita income is different in different regions of the country, with per capita income being smallest in the southern region and largest in the north-eastern region for a set total number of crimes. However, for a set change in the total number of crimes, our model estimates that the change in per capita income is the same regardless of region.

We found that the best model predicting per capita income from all variables in the dataset was model (8). We chose this model as our best model because it was a valid model which was a good fit to the data and was not as complex as models produced by other methods.

Our model would likely be improved by collecting more data. The dataset we used may not necessarily be representative of all counties in the US. For example, note that the state Texas contains 254 counties while the state Pennsylvania contains only 67 counties, but our dataset contains almost the same number of counties for

Texas and Pennsylvania (28 and 29 respectively). Collecting data on more counties would help ensure that our dataset is representative of all counties in the US. In addition, a larger dataset might allow us to include the variable state in our model which might improve the fit.

7 References

Technical Appendix

Read in the data and summarize quantitative variables

```
cdi <- read.delim("cdi.dat", header = TRUE, sep=" ")
head(cdi)
summary(cdi)
```

Summary of county

```
c <- data.frame(table(cdi$county))
unique(c$Freq)
length(which(c$Freq==1))
length(which(c$Freq==2))
length(which(c$Freq==3))
length(which(c$Freq==4))
length(which(c$Freq==5))
length(which(c$Freq==6))
length(which(c$Freq==7))
```

Summary of state

```
table(cdi$state)
```

Summary of region

```
table(cdi$region)
```

Histograms of untransformed quantitative variables

```
par(mfrow=c(2,3))
hist(cdi$land.area,main=NULL,xlab="land.area")
hist(cdi$pop,main=NULL,xlab="pop")
hist(cdi$pop.18_34,main=NULL,xlab="pop.18_34")
hist(cdi$pop.65_plus,main=NULL,xlab="pop.65_plus")
hist(cdi$doctors,main=NULL,xlab="doctors")
hist(cdi$hosp.beds,main=NULL,xlab="hosp.beds")
hist(cdi$crimes,main=NULL,xlab="crimes")
hist(cdi$pct.hs.grad,main=NULL,xlab="pct.hs.grad")
hist(cdi$pct.bach.deg,main=NULL,xlab="pct.bach.deg")
hist(cdi$pct.below.pov,main=NULL,xlab="pct.below.pov")
```

```
hist(cdi$pct.unemp,main=NULL,xlab="pct.unemp")
hist(cdi$per.cap.income,main=NULL,xlab="per.cap.income")
hist(cdi$tot.income,main=NULL,xlab="tot.income")
```

Heatmap of correlation matrix

```
library(reshape2)
cdinumeric <- cdi[,-c(1,2,3,17)]
corgraph <-function(df) {
  cormat <- cor(df)
  melted_cormat <- melt(cormat)
  ggplot(data = melted_cormat, aes(x=Var1, y=Var2, fill=value)) +
    geom_tile() +
    theme(axis.text.x = element_text(angle = 45,vjust=0.9,hjust=1)) +
    scale_fill_gradient2(low="red",mid="white",high="blue")
  }

corgraph(cdinumeric)
```

Histograms of transformed quantitative variables

```
cdi.updated <- cdi
par(mfrow=c(2,3))
hist(log(cdi$land.area),main=NULL,xlab="log.land.area")
cdi.updated$land.area <- log(cdi$land.area)

hist(log(cdi$pop),main=NULL,xlab="log.pop")
cdi.updated$pop <- log(cdi$pop)

hist(cdi$pop.18_34,main=NULL,xlab="pop.18_34")

hist(cdi$pop.65_plus,main=NULL,xlab="pop.65_plus")

hist(log(cdi$doctors),main=NULL,xlab="log.doctors")
cdi.updated$doctors <- log(cdi$doctors)

hist(log(cdi$hosp.beds),main=NULL,xlab="log.hosp.beds")
cdi.updated$hosp.beds <- log(cdi$hosp.beds)

hist(log(cdi$crimes),main=NULL,xlab="log.crimes")
cdi.updated$crimes <- log(cdi$crimes)

hist(cdi$pct.hs.grad,main=NULL,xlab="pct.hs.grad")

hist(cdi$pct.bach.deg,main=NULL,xlab="pct.bach.deg")

hist(cdi$pct.below.pov,main=NULL,xlab="pct.below.pov")

hist(cdi$pct.unemp,main=NULL,xlab="pct.unemp")
```

```
hist(log(cdi$per.cap.income),main=NULL,xlab="log.per.cap.income")
cdi.updated$per.cap.income <- log(cdi.updated$per.cap.income)

hist(log(cdi$tot.income),main=NULL,xlab="log.tot.income")
cdi.updated$tot.income <- log(cdi$tot.income)

cdi.updated <- cdi.updated[,-c(18,19,20,21)] %>%
  rename(log.pop = pop,
         log.land.area = land.area,
         log.doctors = doctors,
         log.hosp.beds = hosp.beds,
         log.crimes = crimes,
         log.tot.income = tot.income,
         log.per.cap.income = per.cap.income
         )
```

Heatmap of correlation matrix after transformations applied

```
cdi.updatednumeric <- cdi.updated[,-c(1,2,3,17)]
corgraph(cdi.updatednumeric)
```

Relationship between log(crimes) and log(per.cap.income)

```
plot(cdi.updated$log.crimes,cdi.updated$log.per.cap.income)
```

Find model to predict log(per.cap.income) from log(crimes)

```
linmod0 <- lm(log.per.cap.income~log.crimes,data=cdi.updated)
linmod1 <- lm(log.per.cap.income~log.crimes+region,data=cdi.updated)
linmod2 <- lm(log.per.cap.income~log.crimes*region,data=cdi.updated)
summary(linmod0)
summary(linmod1)
summary(linmod2)
anova(linmod0,linmod1)
anova(linmod1,linmod2)
par(mfrow=c(2,2))
plot(linmod0)
plot(linmod1)
plot(linmod2)
```

Find model to predict log(per.cap.income) from log(crime rate)

```
cdi.updated$log.per.cap.crime <- log((cdi$crimes)/(cdi$pop))
hist(cdi.updated$log.per.cap.crime)
linmod3 <- lm(log.per.cap.income~log.per.cap.crime,data=cdi.updated)
linmod4 <- lm(log.per.cap.income~log.per.cap.crime+region,data=cdi.updated)
linmod5 <- lm(log.per.cap.income~log.per.cap.crime*region,data=cdi.updated)
```

```
summary(linmod3)
summary(linmod4)
summary(linmod5)
anova(linmod3,linmod4)
anova(linmod4,linmod5)
par(mfrow=c(2,2))
plot(linmod3)
plot(linmod4)
plot(linmod5)
cdi.updated$log.per.cap.crime <- NULL
```

All subsets method on all variables except id, county, state, log.pop, log.tot.income, and region

```
library(leaps)
library(car)
cdi.good <- cdi.updated[,-c(1,2,3,5,16,17)]
all.subsets <- regsubsets(log.per.cap.income~.,cdi.good,nvmax=14)
s <- summary(all.subsets)
d <- data.frame(s$rss,s$bic)
d
```

Coefficients when subset size is 7 (which corresponds to minimum BIC value)

```
coef(all.subsets,7)
```

Summary, VIFs and residual plots of model chosen by all subsets method

```
all.subsets.mod <- lm(log.per.cap.income~log.land.area+pop.18_34+log.doctors+pct.hs.grad+pct.bach.deg+p
summary(all.subsets.mod)
vif(all.subsets.mod)
par(mfrow=c(2,2))
plot(all.subsets.mod)
```

Stepwise AIC and BIC method on all variables except id, county, state, log.pop, log.tot.income, and region

```
stepAIC(lm(log.per.cap.income~.,data=cdi.good),direction="both",k=2)
n <- dim(cdi.updated)[1]
stepAIC(lm(log.per.cap.income~.,data=cdi.good),direction="both",k=log(n))
```

Summary, VIF's and residual plots for model chosen by stepwise AIC method

```
stepAIC.mod <- lm(log.per.cap.income ~ log.land.area + pop.18_34 + pop.65_plus + log.doctors + pct.hs.g
summary(stepAIC.mod)
```

```
vif(stepAIC.mod)
par(mfrow=c(2,2))
plot(stepAIC.mod)
```

Summary, VIF's and residual plots for model chosen by stepwise BIC method

```
stepBIC.mod <- lm(formula = log.per.cap.income ~ log.land.area + pop.18_34 + log.doctors + pct.hs.grad +
summary(stepBIC.mod)
vif(stepBIC.mod)
par(mfrow=c(2,2))
plot(stepBIC.mod)
```

Lasso method on all variables except id, county, state, log.pop, log.tot.income, and region

```
result <- cv.glmnet(as.matrix(cdi.good[,-c(11)]),cdi.good[,11])
plot(result)
c(lambda.1se=result$lambda.1se,lambda.min=result$lambda.min)
cbind(coef(result),coef(result,s=result$lambda.1se),coef(result,s=result$lambda.min))
```

Now we look at all models above but including interactions with region

First, all subsets

```
cdi.good$region <- cdi.updated$region
all.subsets.mod.tmp <- lm(log.per.cap.income~(log.land.area+pop.18_34+log.doctors+pct.hs.grad+pct.bach.
#summary(all.subsets.mod.tmp)
all.subsets.mod.r <- lm(log.per.cap.income~log.land.area+pop.18_34+log.doctors+pct.hs.grad+pct.bach.deg
summary(all.subsets.mod.r)
vif(all.subsets.mod.r)
par(mfrow=c(2,2))
plot(all.subsets.mod.r)
anova(all.subsets.mod,all.subsets.mod.r)
```

Stepwise AIC

```
stepAIC.mod.tmp <- lm(log.per.cap.income ~ (log.land.area + pop.18_34 + pop.65_plus + log.doctors + pct
# summary(stepAIC.mod.tmp)
stepAIC.mod.r <- lm(log.per.cap.income ~ log.land.area + pop.18_34 + pop.65_plus + log.doctors + pct.hs
summary(stepAIC.mod.r)
vif(stepAIC.mod.r)
par(mfrow=c(2,2))
plot(stepAIC.mod.r)
anova(stepAIC.mod,stepAIC.mod.r)
```

Stepwise BIC

```
stepBIC.mod.tmp <- lm(formula = log.per.cap.income ~ (log.land.area + pop.18_34 + log.doctors + pct.hs.g
# summary(stepBIC.mod.tmp)
stepBIC.mod.r <- lm(formula = log.per.cap.income ~ log.land.area + pop.18_34 + log.doctors + pct.hs.grad
summary(stepBIC.mod.r)
vif(stepBIC.mod.r)
par(mfrow=c(2,2))
plot(stepBIC.mod.r)
anova(stepBIC.mod,stepBIC.mod.r)
```