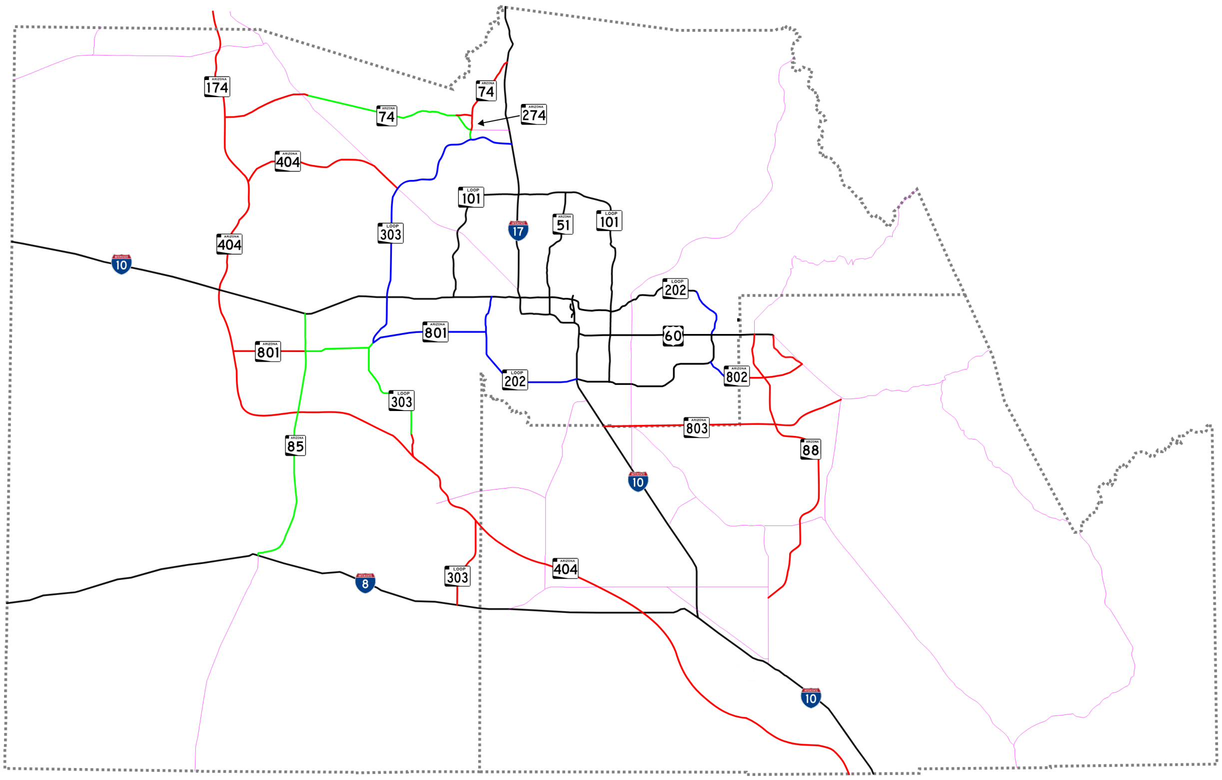
Final Project Report

Gus Lipkin

QMB 3200 ~ Advanced Quantitative Methods



Source:

https://commons.wikimedia.org/wiki/File:Phoenix_Metro_Area_Future_Freeway_System.svg

Introduction

As any good final project should, the final project in QMB 3200: Advanced Quantitative Methods set out to test students' ability to perform statistical analysis and create regressions. The data used is from the NHTS Phoenix-Mesa sub-sample with the goal of creating a "cross-classification matrix of trip generation and linear regression models of person and household trips" (Project Assignment).

Data

The data used was supplied by the professor of the class and was split into two separate files. The first, persontrips.xlsx described an individual person and their trips. The second, hhldtrips.xlsx, described the data at the household level. The person file included data on whether or not the person was a driver and/or a worker, their level of education, their age and sex, whether or not they own or rent their residence, their household income, the number of adults in their house, the number of drivers in their house, and the number of trips that that person made. Because the individual data is not tied to a specific household, the household file repeated some data such as the home ownership status, household income, household size, and the number of drivers. Other household data included the number of vehicles, workers, and adults in the household, and the total number of miles travelled in the household.

Table 1: Continuous Summary Statistics for the Person File

Variable	Obs	Mean	Std. Dev.	Min	Max
# of Adults	648	2.09	0.67	1	4
# of Drivers	648	2.08	0.84	0	5
Age	642	37.86	23.47	0	88
Household Size	648	3.35	1.66	1	9
# of Person Trips	648	4.60	2.40	1	16

There is nothing particularly surprising about the continuous data for the person file. There are 648 people and six of them declined to share their age. The mean number of adults and mean number of drivers are very similar which would indicate that most drivers are adults. The mean number of trips for each person was 4.6 within a range of one to sixteen and a standard deviation of 2.40.

Table 2: Categorical Summary Statistics for the Person File

Driver Status	Freq.	Percent	Cum.	Household Income	Freq.	Percent	Cum.
Not a Driver	33	6.59	6.59	<=5000	9	1.47	1.47
Driver	468	93.41	100	5000-9999	18	2.94	4.4
Total	501	100		10000-14999	13	2.12	6.53
				15000-19999	28	4.57	11.09
Worker Status	Freq.	Percent	Cum.	20000-24999	16	2.61	13.7

Not a Worker	183	36.6	36.6	25000-29999	63	10.28	23.98
Worker	317	63.4	100	30000-34999	27	4.4	28.38
Total	500	100		35000-39999	46	7.5	35.89
				40000-44999	9	1.47	37.36
Education	Freq.	Percent	Cum.	45000-49999	38	6.2	43.56
Less than HS	52	10.55	10.55	50000-54999	24	3.92	47.47
Greater than HS	127	25.76	36.31	55000-59999	59	9.62	57.1
3	11	2.23	38.54	60000-64999	30	4.89	61.99
4	118	23.94	62.47	65000-69999	30	4.89	66.88
5	30	6.09	68.56	70000-74999	20	3.26	70.15
6	85	17.24	85.8	75000-79999	30	4.89	75.04
7	5	1.01	86.82	80000-99999	45	7.34	82.38
8	65	13.18	100	>=100000	108	17.62	100
Total	493	100		Total	613	100	
Homeowner Status	Freq.	Percent	Cum.	Sex	Freq.	Percent	Cum.
Rent	123	18.98	18.98	Female	328	50.62	50.62
Own	525	81.02	100	Male	320	49.38	100
Total	648	100		Total	648	100	

Like the continuous data, there is nothing particularly spectacular about the categorical variables although it becomes much more visible that not every person answered every question on the survey as the totals are frequently much less than the 648 total people that participated. Over 93% of respondents were drivers, just under two-thirds were workers, just under 90% had a greater than high school education, just over 80% owned their home, sex was split roughly 50/50, and assuming the average household was 3.35 people, roughly 13-23% of households earned at or under the federal poverty level for their household size.

Table 3: Continuous Summary Statistics for the Household File

Variable	Obs	Mean	Std. Dev.	Min	Max
# of Vehicles	297	1.89	1.11	0	7
Household Size	297	2.65	1.43	1	9
# of Drivers	297	1.85	0.79	0	5
# of Workers	297	1.26	1.00	0	5
# of Adults	297	1.91	0.66	1	4
Trip Distance	295	118.37	266.98	0.33	3164

There were 297 households that responded with a mean of just under two vehicles, drivers, and adults per household. This suggests that each driver has their own vehicle in most households.Two households declined to share their trip distance data. The mean trip distance was 118.37 miles within a range of 0.33 to 3164 miles with a standard deviation of 266.98 miles.

Analysis and Discussion

As the data was given in the .xlsx format, it could have been imported directly into Stata. However, I know R better and myself the best and knew that if I tried to recode my variables in Stata, I would somehow mess it up and have to start at the beginning. Thus, all variable recoding was done in R. In addition, some variables such as household size and the number of workers in a household were binned into smaller sizes. To improve the accuracy of the regression models, some new variables were also derived from the existing variables. Finally, a new .csv file was generated for the person and household trip files with the new changes. All of this can be seen in the R Markdown Code.

Question one asks for the cross-classification matrices of household trip rates by household size, by number of vehicles, and by number of workers, all binned into zero, one, two, three, or more than three. All three of these matrices are in Table 4.

Table 4: Cross-Classification Matrices for Household Trip Rates

Household Size						Number of Workers					Number of Vehicles						
hhldtrips	1	2	3	4	Total	hhldtrips	0	1	2	Total	hhldtrips	0	1	2	3	4	Total
1	1	1	0	2	4	1	2	0	0	2	1	0	2	1	1	0	4
2	14	11	2	2	29	2	10	12	2	24	2	3	11	14	0	1	29
3	8	4	1	1	14	3	5	6	0	11	3	2	7	4	1	0	14
4	11	10	3	2	26	4	9	11	1	21	4	4	17	3	2	0	26
5	3	9	4	1	17	5	9	5	0	14	5	1	8	7	1	0	17
6	11	7	4	3	25	6	11	9	0	20	6	2	12	7	1	3	25
7	4	8	4	2	18	7	2	7	2	11	7	0	6	7	3	2	18
8	2	19	6	2	29	8	9	6	1	16	8	1	8	17	0	3	29
9	0	4	2	2	8	9	1	2	1	4	9	0	1	3	4	0	8
10	0	11	4	1	16	10	3	4	1	8	10	0	3	9	4	0	16
11	2	7	2	1	12	11	2	7	0	9	11	0	4	6	2	0	12
12	0	7	3	2	12	12	3	3	0	6	12	1	4	5	2	0	12
13	0	3	1	4	8	13	0	2	0	2	13	0	0	7	0	1	8
14	0	6	3	3	12	14	6	2	3	11	14	0	4	5	1	2	12
15	0	5	3	2	10	15	1	1	2	4	15	0	2	4	1	3	10
16	1	1	0	5	7	16	1	2	0	3	16	0	1	3	3	0	7

17	0	1	1	4	6
18	0	1	2	4	7
19	0	0	1	4	5
20	0	2	1	3	6
21	0	0	1	4	5
22	0	0	1	4	5
23	0	0	0	1	1
24	0	0	0	5	5
25	0	0	0	2	2
29	0	0	0	2	2
31	0	0	1	0	1
32	0	0	0	1	1
34	0	0	0	1	1
41	0	0	0	2	2
49	0	0	0	1	1
Total	57	117	50	73	297

17	0	1	1	2
18	0	2	1	3
19	0	2	1	3
20	1	1	2	4
21	0	2	0	2
22	1	3	0	4
23	0	1	0	1
24	0	2	1	3
29	0	1	0	1
31	0	1	0	1
32	0	0	1	1
34	0	1	0	1
41	0	1	1	2
49	0	1	0	1
Total	76	98	21	195

17	0	1	4	0	1	6
18	0	1	3	2	1	7
19	0	2	1	2	0	5
20	0	0	3	3	0	6
21	0	0	3	2	0	5
22	0	0	5	0	0	5
23	0	0	0	1	0	1
24	0	1	2	1	1	5
25	0	0	2	0	0	2
29	0	1	0	1	0	2
31	0	1	0	0	0	1
32	0	0	0	0	1	1
34	0	0	1	0	0	1
41	0	0	1	1	0	2
49	0	0	0	1	0	1
Total	14	97	127	40	19	297

Because these outputs are so long, it is difficult to gain any useful information from them at a quick glance. It could be useful to group the number of household trips in bins of five but there is a decent amount of variation in the number of each household at each number of trips that it would not be much more useful than it is now. It could be more useful if the tabulation also included the number of adults and the number of drivers in each household.

We were then asked to find at least two multiple linear regression models of the total person trips. The easiest way to begin building a model in stata is to simply create a regression with all the available variables, including variables that were derived in the R Markdown File. This initial regression, seen in Table 5, has a R^2 value of .058 which is incredibly low.

Table 5: Multiple Linear Regression of Person Trip Count with All Variables, Including Derived

Source	SS	df	MS	Number of obs	458	
				F(15, 442)	1.820	
Model	163.564	15	10.904	Prob > F	.030	
Residual	2651.527	442	5.999	R-squared	.058	
				Adj R-squared	.026	
Total	2815.092	457	6.160	Root MSE	2.449	
pertrips	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
driver	.917	.604	1.520	.129	-.269	2.103
worker	-.202	.303	-.670	.505	-.797	.393

educ	.128	.067	1.910	.057	-.004	.261
hhincttl	.062	.094	.660	.507	-.122	.246
numadlt	-.041	.563	-.070	.943	-1.147	1.065
drvrcnt	-.094	.289	-.320	.746	-.661	.474
r_age	-.005	.009	-.590	.554	-.023	.012
r_sex	-.275	.239	-1.150	.251	-.745	.195
hhsize	-.229	.297	-.770	.441	-.813	.355
homeown	-.500	.344	-1.450	.147	-1.176	.177
incbin	.092	.397	.230	.818	-.688	.871
educbin	.713	.518	1.380	.169	-.305	1.732
incratio	.000	.000	-1.560	.119	.000	.000
adultratio	-.554	1.737	-.320	.750	-3.967	2.859
adult	-1.110	.899	-1.230	.218	-2.876	.657
_cons	5.915	1.630	3.630	.000	2.712	9.118

An astute observer will notice that categorical variables with more than two values such as household income or education were included in their unencoded state. It seemed unlikely that any such variables would have a large impact on the number of trips and so they were used as a signal. Because these same variables had already been binned for the cross-classification matrices, the binned variables were included as well. If any had had an acceptably low p-value, they would have been converted to dummy variables, but this was not the case. If we were to remove every variable from the regression that had a P>t value greater than .05, we would be left with no variables. As such, only the five lowest P>t value variables were kept.

Table 6: Multiple Linear Regression of Person Trip Count with Five Selected Variables

Source	SS	df	MS	Number of obs	461	
				F(5, 455)	3.370	
Model	100.843	5	20.169	Prob > F	.005	
Residual	2722.194	455	5.983	R-squared	.036	
				Adj R-squared	.025	
Total	2823.037	460	6.137	Root MSE	2.446	
pertrips	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
educ	.146	.066	2.22	0.027	.017	.275
driver	.862	.529	1.63	0.104	-.177	1.901
homeown	-.464	.316	-1.47	0.143	-1.085	.158
educbin	.459	.431	1.06	0.288	-.388	1.306
incratio	.000	.000	-0.53	0.596	.000	.000
_cons	3.477	.591	5.88	0	2.315	4.638

Once again, there are many variables with a P>t value greater than .05. Thus, the highest three are removed.

Table 7: Multiple Linear Regression of Person Trip Count with Two Selected Variables

Source	SS	df	MS	Number of obs	493	
				F(2, 490)	7.490	
Model	89.354	2	44.677	Prob > F	.001	
Residual	2922.655	490	5.965	R-squared	.030	
				Adj R-squared	.026	
Total	3012.008	492	6.122	Root MSE	2.442	
pertrips	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
educ	.143	.050	2.86	0.004	.045	.241
driver	1.048	.480	2.18	0.029	.105	1.991
_cons	3.199	.483	6.62	0	2.249	4.148

This final regression with just two variables has a considerably lower R^2 value of just .030 than the initial regression with all variables. However, a regression with just two variables is much easier and less costly to perform.

The next task was to estimate the total household trips using a multiple linear regression and to provide the two best models. Unlike the person trip models, this did not initially include any derived variables. The R^2 of all variables was .4607 and can be seen in Table 8.

Table 8: Multiple Linear Regression of Household Trip Distance with All Variables

Source	SS	df	MS	Number of obs	295	
				F(7, 287)	35.02	
Model	7339.088	7	1048.441	Prob > F	0	
Residual	8591.501	287	29.936	R-squared	0.4607	
				Adj R-squared	0.4475	
Total	15930.5898	294	54.1856797	Root MSE	5.4713	
hhldtrips	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
homeown	.814	.859	0.95	0.345	-.878	2.505
hhvehcnt	.423	.419	1.01	0.313	-.401	1.248
hhsize	3.578	.308	11.61	0	2.971	4.185
drvrcnt	1.766	.808	2.19	0.03	.176	3.355
wrkcount	-.921	.451	-2.04	0.042	-1.808	-.033
numadlt	-2.329	.878	-2.65	0.008	-4.057	-.600
trpmiles	.004	.001	3.43	0.001	.002	.007
_cons	1.003	1.120	0.9	0.371	-1.202	3.207

While this R^2 value is by no means spectacular, it is considerably better than the personal trip count R^2 value. In addition, this time several variables have a P>t value less than .05. Running a regression with those five variables, we have the regression model found in Table 9 with an R^2 value of .4558.

Table 9: Multiple Linear Regression of Household Trip Distance with Five Selected Variables

Source	SS	df	MS	Number of obs	295	
				F(5, 289)	48.41	
Model	7261.482	5	1452.296	Prob > F	0	
Residual	8669.108	289	29.997	R-squared	0.4558	
				Adj R-squared	0.4464	
Total	15930.590	294	54.186	Root MSE	5.4769	
hhldtrips	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
hhsize	3.487	.303	11.51	0	2.891	4.084
drvrcnt	2.245	.730	3.08	0.002	.809	3.681
wrkcount	-.884	.439	-2.01	0.045	-1.748	-.020
numadlt	-2.258	.876	-2.58	0.01	-3.982	-.533
trpmiles	.005	.001	3.79	0	.002	.007
_cons	1.580	1.006	1.57	0.117	-.400	3.559

With a difference in R^2 values of just .0049 between the regression with all variables and the one with five variables, I was fairly certain I could use derived variables to improve the accuracy of the model. After a new model with all original and all existing derived variables was created with an R^2 value of .4695, .0088 higher than the original value, I was certain I could get a higher value. After many combinations of variables and new variables were derived, I settled on the model in Table 10 which had an R^2 value of .4784, .0177 higher than the original.

Table 10: Multiple Linear Regression of Household Trip Distance with Five Selected Variables v2

Source	SS	df	MS	Number of obs	295	
				F(5, 289)	53.02	
Model	7621.646	5	1524.329	Prob > F	0	
Residual	8308.944	289	28.751	R-squared	0.4784	
				Adj R-squared	0.4694	
Total	15930.590	294	54.186	Root MSE	5.362	
hhldtrips	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
hhvehcnt	.835	.343	2.43	0.016	.159	1.510
hhsize	3.061	.296	10.33	0	2.478	3.644

numadlt	-1.761	.683	-2.58	0.01	-3.104	-.417
trpmiles	.016	.003	5.44	0	.010	.021
sizeratio	-.017	.004	-4.4	0	-.024	-.009
_cons	2.810	.983	2.86	0.005	.876	4.744

For the next part of the project, we had to create multiple linear regressions of person trip counts separately for adult males and females. As with the previous models, an initial regression was done with all variables and then variables were slowly eliminated until there was a smaller number of variables left.

Table 11: Multiple Linear Regression of Person Trip Counts for Adult Females with Two Selected Variables

Source	SS	df	MS	Number of obs	251	
				F(2, 248)	4.06	
Model	56.4462227	2	28.2231113	Prob > F	0.0184	
Residual	1724.06374	248	6.95186991	R-squared	0.0317	
				Adj R-squared	0.0239	
Total	1780.50996	250	7.12203984	Root MSE	2.6366	
pertrips	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
driver	1.881463	0.662354	2.84	0.005	0.5769067	3.186019
drvrcnt	-0.2215039	0.2239497	-0.99	0.324	-0.6625897	0.2195818
_cons	3.607884	0.6032482	5.98	0	2.419741	4.796027

Without further reduction to a simple linear regression model, there is no way for the adult female multiple regression model to not have any variables with a P>t value greater than .05. The best model had a R^2 value of .0239 which is very very low.

Table 12: Multiple Linear Regression of Person Trip Counts for Adult Males with Two Selected Variables

Source	SS	df	MS	Number of obs	225	
				F(2, 222)	6.78	
Model	68.5110244	2	34.2555122	Prob > F	0.0014	
Residual	1121.8712	222	5.05347386	R-squared	0.0576	
				Adj R-squared	0.0491	
Total	1190.38222	224	5.31420635	Root MSE	2.248	
pertrips	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
worker	-0.7339333	0.342072	-2.15	0.033	-1.408057	-0.0598095
educbin	1.545391	0.4874094	3.17	0.002	0.5848494	2.505932
_cons	3.792042	0.5062309	7.49	0	2.794409	4.789675

The model for adult males was over twice as good as measured by the R^2 value with .0576 which is still very very low. In this case, however, both variables used had P>t values less than .05 by a wide margin.

Conclusion

When creating a simple or multiple linear regression, it seems logical that your models can only be as good as your data. Through this exploration, however, we have seen that by using existing variables to create new, derived variables, you can improve your models. Through this method, I found that personal trips are best estimated by a person's education level and whether or not they were a driver. When specifically looking at adults, the results were best when using a person's driver status and the number of drivers in the household for women and a person's employment status and binned education level for men. Of course, this only applies to the NHTS Phoenix-Mesa sub-sample as when the data source changes, so would the best variables for the models.

Appendix

R Markdown Code

```
---
title: "R Notebook"
output: html_notebook
---

```{r}
library(tidyverse)
library(readxl)

dfPerson <- read_xlsx("../Final Project/persontrips.xlsx")
dfHousehold <- read_xlsx("../Final Project/Hhldtrips.xlsx")
dfPerson
dfHousehold

#remove -1
dfPerson$driver[dfPerson$driver == -1] <- NA
dfPerson$educ[dfPerson$educ == -1 | dfPerson$driver == -7] <- NA
#driver == 1, no == 0
dfPerson$driver[dfPerson$driver == 2] <- 0
#worker == 1, not worker == 0
dfPerson$worker[dfPerson$worker == 2] <- 0
#male == 1, female == 0
dfPerson$r_sex[dfPerson$r_sex == 2] <- 0
#homeown == 1, rent == 0
dfPerson$homeown[dfPerson$homeown == 2] <- 0
dfHousehold$homeown[dfHousehold$homeown == 2] <- 0
#remove -7, -8, and -9 from hhincttl in both files
dfPerson$hhincttl[dfPerson$hhincttl >= -9 & dfPerson$hhincttl <= -7] <- NA
dfHousehold$hhincttl[dfHousehold$hhincttl >= -9 & dfHousehold$hhincttl <= -7] <- NA
#fix sub-zeros
dfPerson$r_age[dfPerson$r_age < 0] <- NA
dfPerson$worker[dfPerson$worker < 0] <- NA
dfPerson$educ[dfPerson$educ < 0] <- NA
dfHousehold$trpmiles[dfHousehold$trpmiles < 0] <- NA

dfHousehold$incBin <- case_when(
  dfHousehold$hhincttl <= 6 ~ 1,
  dfHousehold$hhincttl >= 7 & dfHousehold$hhincttl <= 11 ~ 2,
  dfHousehold$hhincttl >= 12 & dfHousehold$hhincttl <= 16 ~ 3,
  dfHousehold$hhincttl >= 17 ~ 4)

dfPerson$incBin <- case_when(
  dfPerson$hhincttl <= 6 ~ 1,
  dfPerson$hhincttl >= 7 & dfPerson$hhincttl <= 11 ~ 2,
```
```

```
  dfPerson$hhincttl >= 12 & dfPerson$hhincttl <= 16 ~ 3,
  dfPerson$hhincttl >= 17 ~ 4)

dfHousehold$sizeBin <- case_when(
  dfHousehold$hhsize == 1 ~ 1,
  dfHousehold$hhsize == 2 ~ 2,
  dfHousehold$hhsize == 3 ~ 3,
  dfHousehold$hhsize > 3 ~ 4)

dfHousehold$vehBin <- case_when(
  dfHousehold$hhvehcnt == 0 ~ 0,
  dfHousehold$hhvehcnt == 1 ~ 1,
  dfHousehold$hhvehcnt == 2 ~ 2,
  dfHousehold$hhvehcnt == 3 ~ 3,
  dfHousehold$hhvehcnt > 3 ~ 4)

dfHousehold$wrkBin <- case_when(
  dfHousehold$wrkcount == 0 ~ 0,
  dfHousehold$wrkcount == 1 ~ 1,
  dfHousehold$wrkcount > 2 ~ 2)

dfHousehold$vehRatio <- dfHousehold$hhvehcnt / dfHousehold$drvrcnt
dfHousehold$driverRatio <- dfHousehold$drvrcnt / dfHousehold$hhvehcnt
dfHousehold$distanceRatio <- dfHousehold$trpmiles / dfHousehold$drvrcnt
dfHousehold$sizeRatio <- dfHousehold$trpmiles / dfHousehold$hhsize
dfHousehold$adultRatio <- dfHousehold$numadlt / dfHousehold$hhsize

dfPerson$educBin <- 0
dfPerson$educBin[dfPerson$educ > 1] <- 1

dfPerson$incRatio <- (dfPerson$hhincttl * 5000) / dfPerson$hhsize
dfPerson$adultRatio <- dfPerson$numadlt / dfPerson$hhsize

dfPerson$adult <- 0
dfPerson$adult[dfPerson$r_age >= 18] <- 1

dfPerson
dfHousehold

write_excel_csv(dfPerson, "../Final Project/dfPerson.csv")
write_excel_csv(dfHousehold, "../Final Project/dfHousehold.csv")
```
````

### Stata Do-File for the Person File

```
capture log close
log using final_project_person, replace
```
```

```
import delimited "/Users/guslipkin/Documents/Fall2020/QMB 3200 ~ Advanced
Quantitative Methods/Final Project/dfPerson.csv"
clear all

* Summary statistics
summ numadlt drvrcnt r_age hhsize pertrips
tabulate driver
tabulate worker
tabulate educ
tabulate hhincttl
tabulate r_sex
tabulate homeown

* Multiple linear regression of total person trips
regress pertrips driver worker educ hhincttl numadlt drvrcnt r_age r_sex hhsize
homeown incbin educbin incratio adultratio adult
regress hhldtrips hhsize drvrcnt wrkcount numadlt trpmiles
regress hhldtrips homeown hhvehcnt hhsize drvrcnt wrkcount numadlt trpmiles
vehratio driverratio
regress hhldtrips hhvehcnt hhsize wrkcount numadlt trpmiles vehratio driverratio
regress hhldtrips hhvehcnt hhsize wrkcount numadlt trpmiles vehratio
regress hhldtrips hhvehcnt hhsize numadlt trpmiles sizeratio

* Multiple linear regression of total person trips for males and females
* Female
regress pertrips driver worker educ hhincttl numadlt drvrcnt r_age hhsize homeown
incbin educbin incratio adultratio if r_sex==0 & adult==1
regress pertrips driver incratio drvrcnt if r_sex==0 & adult==1
regress pertrips driver drvrcnt if r_sex==0 & adult==1

* Male
regress pertrips driver worker educ hhincttl numadlt drvrcnt r_age hhsize homeown
incbin educbin incratio adultratio if r_sex==1 & adult==1
regress pertrips worker educ r_age hhsize educbin adultratio if r_sex==1 & adult==1
regress pertrips worker educbin if r_sex==1 & adult==1

log close
```

### Stata Do-File for the Household File

```
capture log close
log using final_project_household, replace
import delimited "/Users/guslipkin/Documents/Fall2020/QMB 3200 ~ Advanced
Quantitative Methods/Final Project/dfHousehold.csv"
clear all

* Summary statistics
```

```
summ hhvehcnt hhsize drvrcnt wrkcount numadlt trpmiles
tabulate hometown
tabulate hhincttl

* Cross-classification matrix
tabulate hhldtrips sizebin
tabulate hhldtrips vehbin
tabulate hhldtrips wrkbin

* Multiple linear regression models of total household trips
regress hhldtrips homeown hhvehcnt hhsize drvrcnt wrkcount numadlt trpmiles
regress hhldtrips hhsize drvrcnt wrkcount numadlt trpmiles
regress hhldtrips hhvehcnt hhsize numadlt trpmiles sizeratio

log close
```