To: Dr. Gregory Macfarlane

From: Dr. Gregory Macfarlane

Subject: Trip Generation Lab Solution

Date: February 19, 2014

```
library(knitr)
# make \Sexpr{} output look pretty
knit_hooks$set(inline = function(x) {
  prettyNum(x, big.mark=",")
})
options(width=60)
# set knitr options
opts_chunk$set(cache=TRUE, echo=TRUE,
               tidy=FALSE, size="small", autodep=TRUE)

# load other libraries
library(gdata)
library(apsrtable)
```

This unit has two labs. In the first lab, the students estimate trip generation models from real-world data. In the second, the students apply their estimated coefficients in the Wasatch Front Travel Demand Model.

Generation Models

This lab requires the students to complete two general tasks. In the first, the students estimate a regression model from NHTS household-level data considering multiple covariates. In the second, they estimate a cross-classification model from these same data.

```
load("TripGen.Rdata")

# convert income to continuous
missing.values <- c(-9, -8, -7)
TripGen$HHFAMINC <- unknownToNA (TripGen$HHFAMINC, missing.values)
TripGen$HHINCOME <- ifelse(TripGen$HHFAMINC == 18, 100,
    ifelse(TripGen$HHFAMINC == 17, 90, (TripGen$HHFAMINC * 5 -2.5))
      )

# give levels for rail
TripGen$Rail <- factor(TripGen$RAIL,  labels = c("Rail in MSA", "No Rail in MSA"))

TripGen$HBHUR <- unknownToNA(TripGen$HBHUR, c(-9))
TripGen$HBHUR <- unknownToNA(TripGen$HBHUR, c("x"))
TripGen$UrbanArea <- factor(TripGen$HBHUR, levels = c("S", "U", "TC", "C"),
                            labels = c(" - Suburban"," - Urban",
                                       " - Town and Country", " - Second City"))

#Employment - Household mix
TripGen$EHMix <- TripGen$HTEEMPDN/TripGen$HTPPOPDN
```

The data come from the 2009 National Household Travel Survey, and contain the number of trips made for different types by 106,564 households. Some minor formatting was done to code unknown values as `NA` or to adjust income from a categorical variable to a continuous value. The students may need to do other variable formatting depending on which variables they use.

```
# Workers only
model1.lm <- lm(HBWTrips ~ WRKCOUNT, data = TripGen)
# Vehicles only
model2.lm <- lm(HBWTrips ~ HHVEHCNT, data = TripGen)
#Model 3: with variables such as number of workers, vehicles, type of area and Employment & Housing r
model3.lm <- lm(HBWTrips ~ WRKCOUNT + HHVEHCNT + UrbanArea + EHMix + Rail,
                data = TripGen)
#Model 4: with income of the household
model4.lm <- lm(HBWTrips ~ HHINCOME, data = TripGen)

#Model 5: Adding Income to the variables considered in model 3
model5.lm <- lm(HBWTrips ~ WRKCOUNT + HHVEHCNT + UrbanArea + HHINCOME
                 + EHMix +Rail, data = TripGen)
```

The models are presented in Table 1. The first model estimates the number of work-based trips considering only the number of workers in the household. This model predicts that each worker in a household will increase the number of work trips by on average 0.8238 trips per day. This is somewhat lower than I would initially expect, given that workers typically stereotypically make two HBW trips in a day (home and back). On the other hand, this variable alone only explains 33.21% of the variance in work trips. Perhaps there are some other explanations that the other models can help identify.

At this point the students should discuss the findings of their other models.

One item of note is that a model with only workers and income (Model 4) has a positive coefficient estimate on income, indicating that as the household income increases, this household will make *more* trips, all else equal. But when other hosuehold variables are considered, an increase in income will result in *fewer* trips. This could indicate that there is some correlation between income and the other variables (which would be expected), and also that the income variable may not be coded properly. At any rate, the effect of income on home-based work trips is very small, with a $100,000 increase in income resulting in a decrease of only 0.1 trips in Model 5.

Cross-classification Models

The students should also estimate cross-classification models of trips with categories for the household size (number of adults) and vehicle availability. The code for this is straightforward, as it is just a conditional mean. The students first need to cap the categories.

```
TripGen$adultclass <- ifelse(TripGen$NUMADLT > 6, 6, TripGen$NUMADLT)
TripGen$vehclass <- ifelse(TripGen$HHVEHCNT > 3, 3, TripGen$HHVEHCNT)
```

In class we used the `aggregate` function. I wonder if there is a way in the new `dplyr` package, but I haven't looked yet. At any rate, the cross-classification results are given in 2. What is most obvious about these results is that for households with many adults and no vehicles, the numbers become very strange or unbelievable. This is likely due to the fact that these households are rare and underrepresented in the NHTS. Perhaps a different way of finding these rates would be more appropriate.

```
triptypes <- c("HBOTrips", "HBWTrips", "NHBTrips")
cctables <- list()

for(i in triptypes){
    table <- aggregate(TripGen[,i],
                       by = list(TripGen$adultclass,
                                 TripGen$vehclass),
```

Table 1: Estimated Trip Generation Models

```r
coefnames <- c("(Intercept)", "\\# Workers", "\\# Vehicles",
               "Urban", "Town \\& Country", "Second City",
               "Employment Mix", "No Rail in MSA", "Income (kUSD)")
apsrtable(model1.lm, model2.lm, model3.lm, model4.lm, model5.lm,
          stars="default", digits=3, Sweave=TRUE, coef.names=coefnames)
```

	Model 1	Model 2	Model 3	Model 4	Model 5
(Intercept)	-0.006	0.125^{***}	-0.101^{***}	0.247^{***}	-0.055^{***}
	(0.005)	(0.007)	(0.011)	(0.008)	(0.013)
# Workers	0.824^{***}		0.812^{***}		0.827^{***}
	(0.004)		(0.004)		(0.004)
# Vehicles		0.321^{***}	0.023^{***}		0.035^{***}
		(0.003)	(0.003)		(0.004)
Urban			0.024^{*}		0.016
			(0.010)		(0.011)
Town & Country			0.010		0.002
			(0.008)		(0.009)
Second City			0.042^{***}		0.031^{***}
			(0.009)		(0.009)
Employment Mix			0.021^{***}		0.026^{***}
			(0.005)		(0.006)
No Rail in MSA			0.043^{***}		0.035^{***}
			(0.008)		(0.009)
Income (kUSD)				0.009^{***}	-0.001^{***}
				(0.000)	(0.000)
N	106564	106564	106555	97565	97557
R^2	0.332	0.078	0.333	0.055	0.327
adj. R^2	0.332	0.078	0.333	0.055	0.327
Resid. sd	1.035	1.216	1.035	1.249	1.054

Standard errors in parentheses

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 2: Cross-classification Results

	0 Vehicles	1 Vehicles	2 Vehicles	3 + Vehicles
HBW Trips				
1 Adult	0.1077	0.2805	0.407	0.4444
2 Adults	0.4181	0.3971	0.8306	1.009
3 Adults	0.8065	0.9045	1.131	1.708
4 Adults	1.291	1.367	1.369	2.067
5 Adults	1.778	2	1.804	2.638
6 + Adults	4	3	3	2.851
NHB Trips				
1 Adult	0.5868	1.379	1.606	1.736
2 Adults	1.186	1.936	2.73	3.104
3 Adults	1.919	2.603	2.894	3.594
4 Adults	2.455	2.107	3.173	4.2
5 Adults	1.444	3	3.411	4.793
6 + Adults	0	6.667	4.364	4.797
HBO Trips				
1 Adult	0.4742	0.7682	0.893	0.9367
2 Adults	1.178	1.535	1.935	2.141
3 Adults	1.667	2.458	2.609	2.58
4 Adults	2.036	3.1	3.33	3.116
5 Adults	1.333	4.741	4.339	3.854
6 + Adults	0	5	4.636	4.676

```
                        FUN = mean)
    cctables[[i]]$matrix <- matrix(table$x, nrow=6, ncol=4)
}
```