Homework 2: Trip Generation

Hayden Atchley

9/20/2021

Loading the data

I followed the instructions provided in the textbook for loading the data. The following is the R code I used, and the output:

```
hh <- nhts_households %>%
  # filter to MSA size 2, travel on weekday
  filter(msasize == "02", !travday %in% c("01", "07")) %>%
  # select the columns we care about.
  select(houseid, wthhfin, hhsize, hhvehcnt, numadlt, hhfaminc, wrkcount) %>%
  mutate(
    hhsize = ifelse(hhsize > 4, 4, hhsize),
    hhvehcnt = ifelse(hhvehcnt > 3, 3, hhvehcnt),
    wrkcount = ifelse(wrkcount > 2, 2, wrkcount)
  )

trips <- nhts_trips %>%
  # filter to households in the data
  filter(houseid %in% hh$houseid) %>%
  group_by(houseid, trippurp) %>%
  # count up how many trips each household took
  summarise(trips = n()) %>%
  # "spread" the data, filling zero if no trips were taken
  pivot_wider(id_cols = houseid, names_from = trippurp,
              values_from = trips, values_fill = 0)

# function to change NA to 0
nato0 <- function(x) {ifelse(is.na(x), 0, x)}

tripprod <- hh %>%
  # join tables by id field
  left_join(trips, by = "houseid") %>%
  # change all NA values in columns from the trips data to 0
  mutate_at(vars(names(trips)), nato0)
```

I also changed the "income" data type for easier analysis:

```
tripprod$hhfaminc <- as_factor(tripprod$hhfaminc)
tripprod
```

```
## # A tibble: 10,381 x 13
##    houseid  wthhfin hhsize hhvehcnt numadlt hhfaminc wrkcount   HBO HBSHOP   NHB
##    <chr>      <dbl>  <dbl>    <dbl>   <dbl> <fct>       <dbl> <dbl>  <dbl> <dbl>
```

```
##  1 30000019   279.       2        2        2 03              0     4      0     0
##  2 30000288   103.       1        2        1 05              0     1      1     4
##  3 30000289   244.       3        3        2 07              1     2      4     3
##  4 30000463   348.       2        2        2 06              2     7      2     6
##  5 30000465   133.       4        2        2 08              2     8      2     9
##  6 30000478   120.       2        0        2 03              0     2      0     0
##  7 30000545    35.7      2        3        2 06              2     0      2     2
##  8 30000770   130.       1        1        1 06              1     0      3     1
##  9 30000983   147.       4        3        4 09              1     4      0     2
## 10 30001177   304.       2        0        2 04              2     3      2     0
## # ... with 10,371 more rows, and 3 more variables: HBSOCREC <dbl>, HBW <dbl>,
## #   -9 <dbl>
```

2.1

Using this data, we can calculate trip rates by household size and income group:

```
tripprod <- tripprod %>%
  group_by(hhsize, hhfaminc) %>%
  arrange(hhsize, hhfaminc) %>%
  filter(!hhfaminc %in% c("-7", "-8", "-9"))
tripprod$hhfaminc <- as_label(tripprod$hhfaminc)
tripprod %>%
  summarize(
    n = n(),
    HBW = weighted.mean(HBW, wthhfin),
    HBSHOP = weighted.mean(HBSHOP, wthhfin),
    HBSOCREC = weighted.mean(HBSOCREC, wthhfin),
    HBO = weighted.mean(HBO, wthhfin),
    NHB = weighted.mean(NHB, wthhfin),
    "\U03A3" = HBW + HBSHOP + HBSOCREC + HBO + NHB
  ) %>%
  flextable() %>%
  theme_booktabs() %>%
  colformat_double(j = 2:9,digits = 2) %>%
  align(align = "center", part = "all") %>%
  autofit(add_w = 0, add_h = 0) %>%
  fit_to_width(5.5)
```

hhsize	hhfaminc	n	HBW	HBSHOP	HBSOCREC	HBO	NHB	Σ
1	Less than $10,000	323	0.12	0.61	0.35	0.59	0.83	2.51
1	$10,000 to $14,999	333	0.33	0.82	0.31	0.53	1.09	3.08
1	$15,000 to $24,999	512	0.33	0.95	0.43	0.62	1.61	3.94
1	$25,000 to $34,999	476	0.74	0.95	0.50	0.43	1.32	3.94
1	$35,000 to $49,999	503	0.76	0.80	0.37	0.40	1.35	3.68
1	$50,000 to $74,999	499	0.69	0.92	0.35	0.47	2.04	4.47
1	$75,000 to $99,999	261	0.81	1.00	0.70	0.44	1.95	4.91
1	$100,000 to $124,999	135	1.19	0.81	0.36	0.29	1.80	4.45
1	$125,000 to $149,999	58	0.43	0.95	1.33	0.56	1.28	4.55

hhsize	hhfaminc	n	HBW	HBSHOP	HBSOCREC	HBO	NHB	Σ
1	$150,000 to $199,999	42	1.31	0.31	0.21	0.26	1.63	3.72
1	$200,000 or more	57	0.49	0.48	0.49	0.71	1.27	3.44
2	Less than $10,000	122	0.73	2.58	0.48	1.32	4.21	9.31
2	$10,000 to $14,999	119	0.86	2.50	0.58	1.09	2.39	7.43
2	$15,000 to $24,999	268	0.65	1.79	0.53	1.65	3.04	7.65
2	$25,000 to $34,999	398	1.12	1.59	0.69	0.93	1.76	6.09
2	$35,000 to $49,999	582	1.19	1.99	0.81	1.29	2.36	7.63
2	$50,000 to $74,999	882	1.11	1.56	0.77	1.25	2.61	7.31
2	$75,000 to $99,999	744	1.52	1.64	0.58	1.33	3.14	8.22
2	$100,000 to $124,999	504	1.32	1.59	0.85	1.20	3.04	8.01
2	$125,000 to $149,999	275	1.22	2.01	0.64	1.24	3.17	8.30
2	$150,000 to $199,999	259	1.93	1.63	1.52	0.71	2.84	8.63
2	$200,000 or more	264	1.55	1.95	1.16	0.94	2.89	8.49
3	Less than $10,000	51	1.15	1.41	0.23	2.65	1.96	7.40
3	$10,000 to $14,999	40	0.67	3.11	0.45	1.51	4.17	9.91
3	$15,000 to $24,999	83	1.35	0.88	0.33	2.39	2.20	7.15
3	$25,000 to $34,999	97	1.54	1.66	0.96	1.36	2.08	7.61
3	$35,000 to $49,999	119	2.33	1.50	1.03	2.17	2.54	9.58
3	$50,000 to $74,999	222	2.01	1.82	0.88	2.15	3.66	10.52
3	$75,000 to $99,999	182	2.26	1.48	0.77	1.62	3.23	9.36
3	$100,000 to $124,999	130	1.73	1.85	1.49	2.29	6.10	13.47
3	$125,000 to $149,999	87	2.57	2.93	1.31	1.39	2.45	10.65
3	$150,000 to $199,999	68	2.15	1.07	1.45	0.94	2.94	8.55
3	$200,000 or more	66	2.11	2.25	1.53	2.45	3.41	11.76
4	Less than $10,000	50	0.62	1.53	1.19	4.63	3.06	11.02
4	$10,000 to $14,999	28	0.98	2.37	0.67	2.33	1.87	8.22
4	$15,000 to $24,999	70	1.20	1.72	0.43	7.13	5.20	15.70
4	$25,000 to $34,999	100	1.55	1.98	1.06	5.20	3.12	12.91
4	$35,000 to $49,999	124	1.94	1.42	0.98	3.50	4.36	12.20
4	$50,000 to $74,999	222	1.94	1.72	1.32	3.89	4.10	12.97
4	$75,000 to $99,999	197	2.31	1.13	0.96	3.94	3.40	11.75
4	$100,000 to $124,999	164	2.32	2.22	1.04	4.12	3.80	13.50
4	$125,000 to $149,999	103	2.41	2.06	1.29	4.81	5.24	15.80
4	$150,000 to $199,999	100	1.50	2.50	2.09	5.97	6.64	18.69
4	$200,000 or more	88	1.99	1.16	2.59	3.17	4.89	13.80

hhsize	hhfaminc	n	HBW	HBSHOP	HBSOCREC	HBO	NHB	Σ

Generally, this data seems to make sense: a larger household size will produce more trips, and a higher income roughly will produce more trips as well. In most of the household sizes, the HBW (and total) trip rates peak somewhere around the $150,000 range, which makes sense as after a certain level of income, possibly fewer trips are needed as more executive positions may be easier to do remotely. The HBSOCREC trips also increase generally with income, which makes sense, though there are quite a few specific numbers that don't follow this trend. Most of these are from groups with small sample sizes however, so the exact numbers may not represent a realistic average.

2.2

We can use a similar process to calculate trip rates by household workers and vehicles:

```
tripprod %>%
  group_by(wrkcount, hhvehcnt) %>%
  arrange(wrkcount, hhvehcnt) %>%
  summarize(
    n = n(),
    HBW = weighted.mean(HBW, wthhfin),
    HBSHOP = weighted.mean(HBSHOP, wthhfin),
    HBSOCREC = weighted.mean(HBSOCREC, wthhfin),
    HBO = weighted.mean(HBO, wthhfin),
    NHB = weighted.mean(NHB, wthhfin),
    "\U03A3" = HBW + HBSHOP + HBSOCREC + HBO + NHB
  ) %>%
  flextable() %>%
  theme_booktabs() %>%
  colformat_double(j = 3:9,digits = 2) %>%
  align(align = "center", part = "all") %>%
  autofit(add_w = 0, add_h = 0) %>%
  fit_to_width(5.5)
```

wrkcount	hhvehcnt	n	HBW	HBSHOP	HBSOCREC	HBO	NHB	Σ
0	0	318	0.01	0.83	0.42	0.94	0.87	3.06
0	1	1,652	0.01	1.36	0.61	1.47	1.71	5.17
0	2	1,240	0.06	2.03	0.99	1.46	2.25	6.79
0	3	472	0.03	2.47	1.01	1.90	2.52	7.92
1	0	79	0.50	1.70	0.16	0.59	0.95	3.90
1	1	1,387	1.00	1.13	0.51	1.49	2.25	6.38
1	2	1,268	1.21	1.58	0.79	2.14	2.72	8.43
1	3	760	1.03	1.89	1.13	2.10	3.06	9.22
2	0	8	3.79	2.43	1.12	1.97	1.64	10.94
2	1	153	2.07	1.92	0.35	1.85	3.06	9.24
2	2	1,360	2.30	1.53	0.74	2.32	3.43	10.32
2	3	1,310	2.59	1.53	1.22	2.54	4.22	12.09

This data also roughly makes sense, as trips go up with number of workers and number of vehicles. However, there are some unexpected data points: as an example, a 2-worker 0-vehicle (2W0V) household makes almost twice as many HBW trips as a 2W1V household. I can think of two things that might contribute to this though: the very small number of observations for the 2W0V category (which can introduce statistical error), and that with a car it is much more likely for workers to stop at a different activity on their way to or from work than if they were relying on transit or walking/biking. Overall, the data seems to be mostly good, but may require some tweaking to achieve more realistic numbers.

2.3

We can calculate the standard deviation in trip rates by household size/vehicle count and number of workers/vehicle count:

```
size <- tripprod %>%
  group_by(hhsize, hhvehcnt) %>%
  arrange(hhsize, hhvehcnt) %>%
  summarise(HBW = wtd.sd(HBW, wthhfin)) %>%
  pivot_wider(names_from = hhsize, values_from = HBW)
emp <- tripprod %>%
  group_by(wrkcount, hhvehcnt) %>%
  arrange(wrkcount, hhvehcnt) %>%
  summarise(HBW = wtd.sd(HBW, wthhfin)) %>%
  pivot_wider(names_from = wrkcount, values_from = HBW)

left_join(size, emp, by = "hhvehcnt") %>%
  flextable() %>%
  delete_part(part = "header") %>%
  add_header(hhvehcnt = "Household Vehicles",
             `1.x` = "1", `2.x` = "2", `3` = "3", `4` = "4",
             `0` = "0", `1.y` = "1", `2.y` = "2") %>%
  add_header(hhvehcnt = "",
             `1.x` = "Size", `2.x` = "Size", `3` = "Size", `4` = "Size",
             `0` = "Workers", `1.y` = "Workers", `2.y` = "Workers" ) %>%
  merge_h(part = "header") %>%
  theme_booktabs() %>%
  colformat_double(j = 2:8, digits = 2) %>%
  align(align = "center", part = "all") %>%
  autofit(add_w = 0, add_h = 0) %>%
  vline(j = c("hhvehcnt", "4")) %>%
  set_caption("Weighted standard deviations of HBW trips by HH size/vehicles and HH workers/vehicles")
```

Table 3: Weighted standard deviations of HBW trips by HH size/vehicles and HH workers/vehicles

	Size				Workers		
Household Vehicles	1	2	3	4	0	1	2
0	0.32	0.63	1.04	1.73	0.10	0.81	2.03
1	0.85	1.23	1.15	1.36	0.16	0.95	1.69
2	1.08	1.49	1.54	1.46	0.40	1.03	1.56
3	0.76	1.53	1.80	1.85	0.22	0.97	1.76