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)

tripprod %>%
  head(n = 10L) %>%
  flextable() %>%
  theme_booktabs() %>%
  align(align = "center", part = "all") %>%
  autofit(add_w = 0, add_h = 0) %>%
  fit_to_width(6.5)
```

houseid	wthhfin	hhsize	hhvehcnt	numadlt	hhfaminc	wrkcount	HBO	HBSHOP	NHB	HBSOCREC	HBW	-9
30000019	279.14359	2	2	2	03	0	4	0	0	0	0	0
30000288	102.84591	1	2	1	05	0	1	1	4	0	0	0
30000289	243.53625	3	3	2	07	1	2	4	3	1	1	0
30000463	347.59831	2	2	2	06	2	7	2	6	0	0	0
30000465	133.01103	4	2	2	08	2	8	2	9	0	1	0
30000478	119.86888	2	0	2	03	0	2	0	0	0	0	0
30000545	35.71683	2	3	2	06	2	0	2	2	4	2	0
30000770	130.37038	1	1	1	06	1	0	3	1	0	1	0
30000983	147.38400	4	3	4	09	1	4	0	2	1	2	0
30001177	303.80186	2	0	2	04	2	3	2	0	2	2	0

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(6.5)
```

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
1		3,100	0.53	0.86	0.43	0.50	1.43	3.75
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
2		3,894	1.15	1.79	0.69	1.26	2.73	7.62

hhsize	hhfaminc	n	HBW	HBSHOP	HBSOCREC	HBO	NHB	Σ
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
3		1,011	1.83	1.64	0.82	2.06	3.14	9.49
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
4		1,058	1.78	1.75	1.04	4.39	3.84	12.80

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(6.5)
```

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

wrkcount	hhvehcnt	n	HBW	HBSHOP	HBSOCREC	HBO	NHB	Σ
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 4: 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