Assignment 3: Obtaining zonal totals and trip distribution modeling
Submitted by : 0424042406 , Arnob Protim Roy

Ans No 1

Here is the developed model with independent variables from the previous assignment.

```
summary(model2)
lm(formula = CNTTDHH ~ WRKCOUNT + HHVEHCNT + HHFAMINC + HHSIZE, data = hh_ga)
Coefficients:
            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  0.54266    0.14327   3.788 0.000153 ***
WRKCOUNT     0.42206    0.07259   5.814 6.31e-09 ***
HHVEHCNT     0.16388    0.05147   3.184 0.001458 **
HHFAMINC     0.19835    0.02252   8.808 < 2e-16 ***
HHSIZE       2.12586    0.05114  41.569 < 2e-16 ***
```

So using the model, my modified equation is:

total_trips_per_zone = T(0.54266 + 0.42206* A+ 0.16388*B + 0.19835 *C+ 2.12586*D)

T = total_household in a zone
A = avg_HHWRK of a zone
B = avg_HHVEH in a zone
C = avg_INCOME in a zone
D = avg_HHSIZ in a zone

After analyzing the data with R, the table is generated
(Note : the upper sequence of coefficient is not same as the sequence of table)

	HTAZ	avg_HHSIZ	avg_HHVEH	avg_INCOME	avg_HHWRK	total_households	averagetrip	total_trips
1	39	1.677419	1.193548	8.935484	1.161290	31	6.566705	203.5678
2	72	2.032258	1.870968	16.516129	1.354839	31	9.017368	279.5384
3	175	2.468750	2.187500	11.500000	1.625000	32	9.116237	291.7196
4	521	2.818182	2.181818	16.606061	1.545455	33	10.837363	357.6330
5	522	3.000000	2.114286	18.400000	1.542857	35	11.567547	404.8642
6	906	2.741935	2.225806	9.612903	1.419355	31	9.242168	286.5072

Sum of the total trip is `1823.83`

R Codes for part 1: https://github.com/trewto/CE-6511-Assignments/blob/main/Assignment_3/Assignment3.R

Ans No 2

This is my previous model with a dummy variable.

```
CNTTDHH ~ WRKCOUNT + HHVEHCNT + HHSIZE + HHFAMINC_1 + HHFAMINC_2 +
    HHFAMINC_3 + HHFAMINC_4 + HHFAMINC_5 + HHFAMINC_6 + HHFAMINC_7 +
    HHFAMINC_8 + HHFAMINC_9 + HHFAMINC_10
After Modeling:
```

	Estimate
(Intercept)	2.73730
WRKCOUNT	0.42458
HHVEHCNT	0.16510
HHSIZE	2.12766
HHFAMINC_1	-1.97784
HHFAMINC_2	-1.53642
HHFAMINC_3	-1.78119
HHFAMINC_4	-1.54876
HHFAMINC_5	-1.10593
HHFAMINC_6	-0.96571
HHFAMINC_7	-0.83258
HHFAMINC_8	-0.91071
HHFAMINC_9	-0.21164
HHFAMINC 10	-0.03181

Though it is now seen ,seeing the present excel data that if i would choose the number of cars per household as my dummy variable the model might work better. But as the model with income level as a dummy variable was done by me in a previous assignment, the rest process is done the same.

Also it is noted that the income is distributed in 11 (0 to 10) category. So I take the 11-1 = 10 dummy variable. The effect of the (HHFAMINC_0) will be automatically adjusted by the equation.

So my equation is for the assignment to predict total predicted trip of a zone,

A=2.73730 (Intercept)
B=0.42458 (Coefficient for avg_HHWRK)
C=0.16510 (Coefficient for avg_HHVEH)
D=2.12766 (Coefficient for avg_HHSIZ)

β1=−1.97784
β2=−1.53642
β3=−1.78119
β4=−1.54876
β5=−1.10593

β6=−0.96571
β7=−0.83258
β8=−0.91071
β9=−0.21164
β10=−0.03181
where Hji is the number of households of class j in zone i.

Total expected trip for a zone
= Total_households (A + B * avg_HHWRK + C * avg_HHVEH + D * avg_HHSIZ) +
($\beta 1 * H_{1i} + \beta 2 * H_{2i} + \beta 3 * H_{3i} + \beta 4 * H_{4i} + \beta 5 * H_{5i} + \beta 6 * H_{6i} + \beta 7 * H_{7i} + \beta 8 * H_{8i} + \beta 9 * H_{9i} + \beta 10 * H_{10i}$)

or

= Total_households (2.73730 + 0.42458 * avg_HHWRK + 0.16510 * avg_HHVEH + 2.12766 * avg_HHSIZ) + ($-1.97784 * H_{1i} - 1.53642 * H_{2i} - 1.78119 * H_{3i} - 1.54876 * H_{4i} - 1.10593 * H_{5i} - 0.96571 * H_{6i} - 0.83258 * H_{7i} - 0.91071 * H_{8i} - 0.21164 * H_{9i} - 0.03181 * H_{10i}$)

using the concept on R , according dataset is generate

	1	2	3	4	5	6
HTAZ	39.000000	72.000000	175.00000	521.000000	522.000000	906.000000
avg_HHSIZ	1.677419	2.032258	2.46875	2.818182	3.000000	2.741935
avg_HHVEH	1.193548	1.870968	2.18750	2.181818	2.114286	2.225806
avg_HHWRK	1.161290	1.354839	1.62500	1.545455	1.542857	1.419355
total_households	31.000000	31.000000	32.00000	33.000000	35.000000	31.000000
t_HHFAMINC_1	2.000000	0.000000	0.00000	1.000000	0.000000	1.000000
t_HHFAMINC_2	5.000000	1.000000	0.00000	0.000000	2.000000	0.000000
t_HHFAMINC_3	2.000000	2.000000	1.00000	1.000000	0.000000	3.000000
t_HHFAMINC_4	1.000000	0.000000	1.00000	0.000000	0.000000	3.000000
t_HHFAMINC_5	4.000000	2.000000	1.00000	2.000000	3.000000	1.000000
t_HHFAMINC_6	2.000000	0.000000	1.00000	0.000000	1.000000	1.000000
t_HHFAMINC_7	1.000000	5.000000	1.00000	2.000000	2.000000	9.000000
t_HHFAMINC_8	3.000000	4.000000	6.00000	4.000000	5.000000	5.000000
t_HHFAMINC_9	7.000000	10.000000	4.00000	9.000000	11.000000	7.000000
t_HHFAMINC_10	3.000000	4.000000	16.00000	11.000000	7.000000	0.000000
total_trip_from_zone	188.642520	228.947000	276.25995	308.210500	338.228760	268.213240

(Though original form are in 6 x 16 matrix, but for better view/snapshot purpose i converted it in transpose)
note : t_HHFAMIINC_i refers to the number of households in a particular zone for i income level.

Sum of total trip generated from all the zone is `1608.502` , which is slightly lower than the previous method

R Codes for part 2: https://github.com/trewto/CE-6511-Assignments/blob/main/Assignment_3/Assignment3_part2.R