

CEE 123 Transport Systems 3: Planning & Forecasting

Winter 2021: Michael G. Mc Nally (mmcnally-at-uci-dot-edu) [15460]

Homework #5 -- Trip Generation Modeling [Due: Monday 10 May 2021]

Problem 1 [20 points]

The data in Table 1 was collected from 10 households (HH). Variables include Household Identification Number, HHID), daily trips per household (Trips), HH Income (HHInc, in \$1,000s), number of cars in the HH (Cars), number of persons in the HH (HHSize), and dwelling unit type (DU, 1=Single Family; 2=Multiple Family). These 10 observations are the **first** set of 10 households in [Table 7](#) [[xls](#)].

- Hypothesize several alternate model structures (via causal arrow diagrams: $X \rightarrow Y$) and then find a valid bivariate trip generation model performing the calculations by hand.
- Hypothesize several alternate model structures (via causal arrow diagrams: $X_1, X_2 \rightarrow Y$) and then find a valid multivariate trip generation model (use any available software -- Excel, TransCAD, or a statistical package -- but please identify the software and include appropriate model output.

Table 1. Sample Household Travel and Demographic Data

HHID	Trips	HHInc	Cars	HHSize	DU
1	4	45.0	2	3	2
2	3	40.2	1	2	2
3	4	46.5	1	1	2
4	5	50.4	2	3	2
5	6	57.3	2	2	2
6	6	49.8	2	3	1
7	7	52.5	1	2	1
8	7	55.5	2	3	1
9	6	55.8	2	3	1
10	3	42.6	1	2	2

Problem 2 [20 points]

Table 2 provides a category distribution of 40 households by number of persons per household (categorized as 1-2 or 3 plus) and HH income (categorized as 45k and under, 45.1 to 60k, or >60k). Each cell contains the total number of trips and the total number of households for the first 40 of 50 data points in **Table 7**.

- Add the remaining 10 households to this table
- Build a *category trip generation model* by computing trip production rates for each cell (*and* for row and column totals) of the matrix. Round to nearest tenth of a trip.

Table 2. Trip Summary (HHs 1-40 only)

HHInc	HHSize		Row Tot
	1-2	3+	
.LE. 45k	24 7	4 1	28 8
45.1 to 60k	53 10	98 15	151 25
.GT. 60k	0 0	66 7	66 7
Column	77	168	245

```

|Total | 17 | 23 | 40 |
+-----+-----+-----+

```

Problem 3 [10 points]

Compare your category model from Problem 2 with the corresponding regression model (see output below).

- Evaluate the regression estimation results statistically.
- Interpret the model coefficients -- what do these values imply?
- Compute regression estimates corresponding to each cell of the category model (use appropriate discrete values). Compare results.

Table 3. Regression Results for Trips versus HHInc and HHSize

----- O R D I N A R Y L E A S T S Q U A R E S -----							
VARIABLE	MEAN	S.D.	OBS	CORREL	HHInc	HHSize	Trips
1. HHInc	50.5080	8.7108	50	HHInc	1.0000	0.8033	0.9499
2. HHSize	2.7400	1.1031	50	HHSize	0.8033	1.0000	0.8356
3. Trips	5.6600	2.4042	50	Trips	0.9499	0.8356	1.0000
* * * * * O R D I N A R Y L E A S T S Q U A R E S * * * * *							
MODEL: Cat.Mod.Compar. DEPENDENT VARIABLE => Trips Produced							
MULTIPLE R	0.9577	* ANOVA *		SUM OF SQR	df	MEAN SQR	F
R-SQUARE	0.9171	REGRESSION		259.74	2	129.87	260.01
ADJ R-SQUARE	0.9136	RESIDUALS		23.48	47	0.50	
S.E. OF EST.	0.7067	TOTAL SS		283.22	49		
VARIABLE NAME	B	BETA	S.E. B	T			
1. HHInc	0.2168	0.7856	0.0195	11.1411			
2. HHSize	0.4458	0.2045	0.1537	2.9009			
Constant	-6.5124						

Problem 4 [10 points]

Using both the category and the regression production models, forecast the number of trips per household for the six household not used in model estimation (households 51-56; see **Table 4**), comparing forecast and observed trip rates.

Table 4. Households for Validation Test

ID	Trips	Income	Cars	HHS	DU	ID	Trips	Income	Cars	HHS	DU
51	6	45.0	2	3	2	54	10	59.4	3	5	1
52	3	40.2	1	2	2	55	8	58.5	3	4	1
53	4	49.5	1	1	2	56	5	43.8	2	2	1

Problem 5 [10 points]

The 50 households were sampled from a study area divided into three zones (TAZs). The associated population-level distributions for these zones are provided in Table 5. Compute the total number of trips produced per zone using your final category model from Problem 2.

Table 5. Population Distribution of Households (HHInc by HHSize)

TAZ 1				TAZ 2				TAZ 3			
HHSiz	1-2	3-5	Row	HHSiz	1-2	3-5	Row	HHSiz	1-2	3-5	Row
HInc			Tot	HInc			Tot	HInc			Tot
LE 45	0	0	0	LE 45	40	40	80	LE 45	30	70	100
45-60	0	60	60	45-60	40	80	120	45-60	70	20	90
GT 60	0	40	40	GT 60	20	80	100	GT 60	0	10	10
Col	0	100	100	Col	100	200	300	Col	100	100	200

Problem 6 [10 points]

The other side of the trip generation stage is estimating trip attractions. The following regression-based total trip attraction model was estimated for the region:

$$A_j = 1.5 \text{ POP}_j + 3.0 \text{ EMP}_j$$

Table 6 provides regional demographic information. Compute total attractions and compare these results with the estimates for total productions from Problem 5. Since every trip has a production and an attraction, normalize the attractions so that the total equals total productions.

Table 6. Demographic Data Summary

TAZ	HH	POP	EMP	
1	100	300	0	HH = total households
2	300	1100	400	POP = total population
3	200	600	100	EMP = total employment
Tot	600	2000	500	

Problem 7 [Optional: 10 points Extra Credit for CEE123]

The following regression results summarize an attempt to build a home-to-work trip production model. Fill in the blanks, interpret the parameters, and discuss the results, and select a significant model (if any).

VARIABLE	NAME	MEAN	S.D.	OBS
1. Population	POP	4090.90	3624.10	29
2. Labor Force	LABF	1630.93	1424.85	29
3. HH Income	INC	6068.55	1337.82	29
4. Employment	EMPL	735.14	853.31	29
5. Work Origins	TRIP	1324.93	1207.45	29

Correlation Matrix:

1. POP	1.0000	0.9971	0.0061	0.4894	0.9652
2. LABF	0.9971	1.0000	-0.0002	0.5069	0.9693
3. INC	0.0061	-0.0002	1.0000	-0.2717	0.0271
4. EMPL	0.4894	0.5069	-0.2717	1.0000	0.4778
5. TRIP	0.9652	0.9693	0.0271	0.4778	1.0000
Variables:	POP	LABF	INC	EMPL	TRIP

(a) ==== OLS MODEL #1 : DEPENDENT VARIABLE => TRIP =====

MULTIPLE R _____ * ANOVA * df SUM OF SQR MEAN SQR F

R-SQUARE		REGRESSION	38354310.0		
ADJ R-SQUARE	0.9372	RESIDUALS	2470519.0		
S.E. OF EST.		TOTAL SS			
VARIABLE NAME	B	BETA	S.E. B	T	
1. LABF	0.8214	0.9693	0.0401		
Constant					

Model #1 Comments:

(b) ==== OLS MODEL #2 : DEPENDENT VARIABLE => TRIP =====

MULTIPLE R	0.9694	* ANOVA *	df	SUM OF SQR	MEAN SQR	F
R-SQUARE		REGRESSION		38363630.0		
ADJ R-SQUARE	0.9351	RESIDUALS				
S.E. OF EST.		TOTAL SS				
VARIABLE NAME	B	BETA	S.E. B	T		
1. POP	-0.0672	-0.2016	0.2094			
2. LABF	0.9917	1.1702	0.5326			
Constant						

Model #2 Comments:

(c) Which, if any, is the better model? Why? Would POP by itself make a good model?

Table 7. Household Travel Survey Data

ID	Trips	Income	Cars	HHS	DU	ID	Trips	Income	Cars	HHS	DU
1	4	45.0	2	3	2	26	10	59.4	3	5	1
2	3	40.2	1	2	2	27	8	58.5	3	4	1
3	4	46.5	1	1	2	28	5	40.8	1	2	1
4	5	50.4	2	3	2	29	8	54.3	2	4	1
5	6	57.3	2	2	2	30	9	61.5	3	4	1
6	6	49.8	2	3	1	31	5	50.1	2	2	2
7	7	52.5	1	2	1	32	6	55.5	2	3	1
8	7	55.5	2	3	1	33	10	61.2	3	4	1
9	6	55.8	2	3	1	34	4	45.6	1	2	2
10	3	42.6	1	2	2	35	8	63.9	2	4	1
11	5	46.8	2	2	2	36	6	55.5	2	2	2
12	7	50.4	2	3	1	37	5	49.8	2	3	2
13	6	52.8	2	2	1	38	8	63.9	2	4	1
14	4	43.2	1	1	2	39	4	42.9	1	1	2
15	5	49.2	2	3	2	40	4	45.6	1	2	2
16	5	49.2	2	3	2	41	8	60.6	3	4	2
17	8	60.0	3	5	2	42	7	55.5	2	3	1
18	3	39.0	1	1	2	43	5	48.6	2	3	2
19	6	51.9	2	2	2	44	3	40.5	1	2	2
20	9	63.0	3	4	1	45	2	37.5	1	2	2
21	11	67.8	3	5	1	46	3	40.8	1	1	2
22	5	49.5	2	3	2	47	2	37.5	1	2	2
23	11	67.5	3	5	1	48	3	41.4	1	2	2
24	2	35.4	1	2	2	49	2	35.1	1	2	2
25	7	57.3	2	3	1	50	3	40.8	1	2	1

1. HH ID = household number
 2. Trips = number of daily trips per household
 3. HHInc = mean Household income (in \$1000s)
 4. Cars = number of cars per HH
 5. HHSiz = Household size (persons per HH)
 6. DUTyp = Dwelling Unit Type (1=single family; 2=multiple family)
-

Last Updated: 30 April 2021