CEE 123 Transport Systems 3: Planning & Forecasting

Spring 2022: Michael G. McNally (e-mail) [15440]

Homework #5 -- Trip Generation Analysis [Due: Friday, May 6th]

Problem 1 [20 points]

The data in **Table 1** was collected from 10 households (HH21-HH30). Variables include Household ID Number, HHID), daily trip productions per household (Trips), HH Income (HHInc, in $1,000s), number of cars in the HH (Cars), persons per HH (HHSize), and dwelling unit type (DU, 1=Single Family; 2=Multiple Family).

a. **Hypothesize** several alternate model structures [e.g., Trips = f(Inc)] and then **estimate** a valid bivariate trip generation model using these 10 data points, performing the calculations using Excel.
b. **Hypothesize** several alternate multivariate structures [e.g., Trips = f(Inc, Cars)] and **estimate** a valid multivariate trip generation model using these 10 data points (use Excel). Include Excel output.

Table 1. Sample Household Travel and Demographic Data

HHID	Trips	Inc	Cars	HHSize	DU
31	5	50.1	2	2	2
32	6	55.5	2	3	1
33	10	61.2	3	4	1
34	4	45.6	1	2	2
35	8	63.9	2	4	1
36	6	55.5	2	2	2
37	5	49.8	2	3	2
38	8	63.9	2	4	1
39	4	42.9	1	1	2
40	4	45.6	1	2	2

Solutions:

(a.1) *All of the possible explanatory variables could be considered potential regressors (DU inversely, the rest directly). There is probably not significant intercorrelation between these variables, but care should be exercised when developing multivariate models such as Trips=f(Cars, HHSize).*

(a.2) *Sample Hand Calculation Results for Trips=f(Income) below.*
Note: Any of the four bivariate models above could be estimated (hand calcs were not required but are shown for one model, bivariate results from Excel are below (results for obs 1-50 are provided in Problem 7).

Table 1. Sample Household Travel and Demographic Data [31-40]

Obs	Trips Y	Inc X	Y-sq	X-sq	XY
31	5	50.1	25	2510.01	250.5
32	6	55.5	36	3080.25	333.0
33	10	61.2	100	3745.44	612.0
34	4	45.6	16	2079.36	182.4
35	8	63.9	64	4083.21	511.2
36	6	55.5	36	3080.25	333.0
37	5	49.8	25	2480.04	249.0
38	8	63.9	64	4083.21	511.2
39	4	42.9	16	1840.41	171.6
40	4	45.6	16	2079.36	182.4
Sum	60	534.0	398	29061.54	3336.3
Mean	6	53.4			
S.D.	2.02	7.79			

```
----------------------------------------------

b1 = [Sum{XY} - nX'Y']/[Sum{X-sq} - n{X'}^2]
   = [3336.3-(10)(53.4)(6)]/[29061.54-(10)(53.4)(53.4)]     =  0.2423

b0 = Y' - b1 X' = 6 - (0.2423)(53.4)                        = -6.9407

R = [Sum{XY}-nX'Y']/[Sqrt(Sum{X-sq}-n{X'}^2) Sqrt(Sum{Y-sq}-n{Y'}^2)]
             [3336.3-(10)(53.4)(6)]
  = --------------------------------------------------      =  0.9185
     Sqrt[29061.54-(10)(53.4)^2] Sqrt[398-(10)(6)^2]

R-sq = R(R) = (0.9592)(0.9592)                              =  0.8437

Sest = Sqrt [(Sum{Y-sq}-b0(Sum{Y})-b1(Sum{XY}))]/(n-k-1)]
     = Sqrt [{398-(-6.9407)(60)-(0.2423)(3336.3)}/(10-1-1)] =  0.8616

Sb = Sest / [Sx Sqrt(n-1)]  = 0.8616 / [(7.79)(3)]          =  0.0369

t = b1/Sb   =  0.2423/0.0369                                =  6.5914

Model (obs 31-40):
Y (trips) = -6.94 + 0.24 X (HHInc)
```

```
-------------------------------         -----------------------------
Excel LINEST of Trips=f(Income):        Excel LINEST of Trips=f(Cars):
-------------------------------         -----------------------------
b1 b0     0.2423 -6.9407                b1 b0    -0.7500  7.5000
se        0.0369  1.9879                se        1.06    2.22
t1 t0     6.5716 -3.4914                t1 t0    -0.7096  3.3828
R2 Sey    0.8437  0.8616                R2 SEy    0.0592  2.11
F df     43.1856  8.                    F df      0.50  8.
SSR SSE  32.0608  5.9392                SSR SSE   2.25   35.75

-------------------------------         -----------------------------
Excel LINEST of Trips=f(DU):            Excel LINEST of Trips=f(HHS):
-------------------------------         -----------------------------
b1 b0     0.40    5.4000                b1 b0    -0.1835  6.5321
se        1.37    2.17                  se        0.66    2.02
t1 t0     0.2917  2.4908                t1 t0    -0.2793  3.2260
R2 SEy    0.01    2.17                  R2 SEy    0.0097  2.17
F df      0.09    8.                    F df      0.08    8.
SSR SSE   0.40   37.60                  SSR SSE   0.37   37.63
```

(b.2) *Some multiple regression models for the entire data set are summarized in Problem 7.*

Problem 2 [10 points]

A base category distribution of households by number of cars per household (1, 2, or 3+) and type of dwelling unit (1=single family, 2=multiple family) is provided in **Table 2**. Each category cell contains the total number of trips and the total number of households, respectively, but only for the first 40 data points in **Table 7**. Add the remaining 10 households to this table then 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 2A. Trip Summary (All 50 HHs) / Table 2B. Corresponding TG Model

```
  Table 2A. Household Trip Summary   Table 2B. Trip Generation Model
    +-----+-------------+-----+        +-----+-------------+-----+
    |     |Dwelling Type|     |        |     |Dwelling Type|     |
    |Cars +-------------+ Row |        |Cars +-------------+ Row |
    |/HH  | 1=SF | 2=MF | Tot |        |/HH  | 1=SF | 2=MF | Tot |
    +=====+======+======+=====+        +=====+======+======+=====+
```

1	15	46	61
	3	15	18
2	76	62	138
	11	12	23
3+	68	16	84
	7	2	9
Col	159	124	283
Tot	21	29	50

1	5.0	3.1	3.4
Regr	[4.8]	[3.0]	[3.3]
2	6.9	5.2	6.0
Regr	[7.2]	[5.4]	[6.2]
3	9.7	8.0	9.3
Regr	[9.6]	[7.8]	[9.1]
Col	7.6	4.3	5.7
Tot	[7.6]	[4.3]	

In Table 2A, each cell contains the total number of trips followed by the total number of households. In Table 2B, each cell contains trips per household from the category analysis (total trips divided by total HHs, from Table 2A) and, second, the corresponding trip rate estimated from the regression model (see Problem 3).

Problem 3 [10 points]

Compare your category model with the corresponding regression model (below).

a. Evaluate the estimation results statistically.
The model is statistically significant (t- and F-statistics), has a very good R2, is properly signed, and has no indication of multi-colinearity or other estimation problems.

b. Interpret the model coefficients? What do these values imply?
The coefficient of CARS implies that an added vehicle generates 2.37 additional trips per houshold, while multifamily households make on average 1.78 fewer trips per household than a SF household.

c. Compute regression estimates for each cell of the category model. Compare results.
See Table 2B above. Consider these numbers relative to the discussion in Problem 3(b).

```
----------------------------------------------------------------------
--------     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  Cars      DU        Trips
 1. Cars     1.8200  0.7197   50      Cars   1.0000   -0.4425    0.8721
 2. DU       1.5800  0.4986   50      DU    -0.4425    1.0000   -0.6834
 3. Trips    5.6600  2.4042   50      Trips  0.8721   -0.6834    1.0000

 MODEL: Cat.Mod.Compar.   DEPENDENT VARIABLE => Trips Produced

 MULTIPLE R      0.9331   * ANOVA *  SUM OF SQR   df   MEAN SQR    F
 R-SQUARE        0.8706   MODEL SS       246.58    2     123.29  158.17
 ADJ R-SQUARE    0.8651   ERROR SS        36.64   47       0.78
 S.E. OF EST.    0.8829   TOTAL SS       283.22   49

        VARIABLE NAME       B         BETA     S.E. B     T
          1. Cars         2.3665    0.7084    0.1954   12.1095
          2. DU          -1.7840   -0.3700    0.2821   -6.3241
             Constant     4.1718
```

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 54-59; 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
54	10	59.4	3	5	1	57	3	37.5	1	1	2
55	8	58.5	3	4	1	58	5	41.4	1	2	1
56	5	43.8	2	2	1	59	8	51.5	2	4	1

Table 4B. Validation Test Results

HH	Obs. Trips	Cars	DU	Regr. Est.	Table Est.	HH	Obs. Trips	Cars	DU	Regr. Est.	Table Est.
54	10	3	1	9.6	9.7	57	3	1	2	3.0	3.1
55	8	3	1	9.6	9.7	58	5	1	1	4.8	5.0
54	5	2	1	7.2	6.9	59	8	2	1	7.1	6.9

Results for both models are comparable; percent errors for individual households can be high.

Problem 5 [10 points]

The 50 households in question were sampled from a study area divided into 3 zones (TAZs). The associate population-level distributions for these 3 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

TAZ 1

D.U. / Cars	1 SF	2 Other	Row Tot
1	0	0	0
2	50	0	50
3+	50	0	50
Col	100	0	100

TAZ 2

D.U. / Cars	1 SF	2 Other	Row Tot
1	50	50	100
2	100	30	130
3+	50	20	70
Col	200	100	300

TAZ 3

D.U. / Cars	1 SF	2 Other	Row Tot
1	20	80	100
2	20	70	90
3+	10	0	10
Col	50	150	200

Solution

For each TAZ, the number of households in each cell is multiplied by the category model trip rate from Table 3B, summing these results over all types of households in each TAZ.

Table 5b. Trip Production Summary

TAZ	HH	Trips	Total
1	100	0(5.0)+ 50(6.9)+50(9.7) + 0(3.1)+ 0(5.2)+ 0(8.0) =	830
2	300	50(5.0)+100(6.9)+50(9.7) +50(3.1)+ 30(5.2)+20(8.0) =	1896
3	200	20(5.0)+ 20(6.9)+10(9.7) +80(3.1)+ 70(5.1)+ 0(8.0) =	947
Tot	600		3673

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:

$A_j = 1.5\ POP_j + 2.9\ EMP_j$

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 6A-B. Demographic Data Summary

TAZ	HH	POP	EMP
1	100	300	0
2	300	1100	400
3	200	600	100
Tot	600	2000	500

HH = total households

POP = total population

EMP = total employment

Solutions
The "Prod" column contains the trip production totals from Table 5B. The trip attraction column provides the calculations using the estimated regression equation provided. The "Norm A" column contains the normalized (balanced) attractions: the raw attractions multiplied by the balancing factor of 0.8642 (the ratio of total productions, 3673, to total attractions, 4250).

Table 6B. Attraction Model Application and Balancing

TAZ	HH	POP	EMP	Prod.	Trip Attractions	Total	Norm A
1	100	300	0	830	1.5(300)+2.9(0) =	450	371
2	300	1100	400	1896	1.5(1100)+2.9(400) =	2810	2320
3	200	600	100	947	1.5(600)+2.9(100) =	1190	982
Tot	600	2000	500	3673		4450	3673

Problem 7 [20 points]

Concisely analyze each of the following bivariate (Table 7A) and multivariate (Table 7B) regression total trip production models which were estimated using the household travel diary data set (see Table 8 below). Discuss which models are better and why.

```
Table 7A. Bivariate Regression Summary (variables defined in Table 8)
-----------------------------------------------------------------
Model     A       B       C       D
-----------------------------------------------------------------
INC     0.79
(t)   (21.05)
CARS            2.91
(t)           (12.35)
HHS                     1.79
(t)                   (10.68)
```

DU				-3.30
(t)				(-6.49)
Cnst	-7.58	0.36	0.79	10.87
R	0.95	0.87	0.84	-0.68
R-Sq	0.90	0.76	0.70	0.47
S.E.	0.76	1.19	1.32	1.77
F	443.10	152.52	114.06	43.12

Table 7B. Multiple Regression Summary

Model	E	F	G	H	I	J	K
INC	0.65	0.62	0.49	0.50	0.58		0.65
(t)	(8.99)	(8.61	(7.17)	(7.25)	(10.49)		(11.04)
CARS	0.66	0.27	0.68	0.87		2.08	
(t)	(2.27)	(0.79	(2.17)	(3.52)		(5.87)	
HHS		0.36	0.16		0.39	0.23	0.45
(t)		(1.95)	0.97		(2.82)	(0.96)	(2.95)
DU			-0.95	-1.02	-0.79	-1.68	
(t)			(-4.11)	(-4.59)	(-3.46)	(-5.57)	
Cnst	-6.40	-6.17	-2.80	-2.66	-3.98	3.90	-6.45
R	0.95	0.96	0.97	0.97	0.97	0.93	0.96
R-Sq	0.91	0.92	0.94	0.94	0.93	0.87	0.92
S.E.	0.73	0.71	0.61	0.61	0.63	0.88	0.70
F	243.32	173.16	178.90	238.53	219.11	105.58	261.55

Solution: ONLY a summary is provided but full statistical evaluation is required.

Each bivariate model is significant. Model A is best on most statistics but has a negative constant which is considered by some to be problematic in applications. Models B and C are very good; model D is good, but a has a relatively lower R-squared. The coefficients of the explanatory variables may be interpreted as producing 0.79, 2.91, and 1.79 more trips for each additional unit of income, cars, and household size, and reducing the number of trips per HH by 3.30 for a MF household versus a SF household, all other things constant.

Each multiple variable model tested yielded significant F-scores and very good R-squares. All models which included income also had negative constants. The variables CARS and HHS cannot be in the same model due to multicolinearity (models F, G, and J are eliminated with low t-scores). Overall, models E, H, I, and K are significant. Since CARS and DU are often used in regional models as important policy veriable, model H may be the overall best model, although dropping DU from the model doesn't significantly effect overall performance (see Model E), and produces a simpler model structure. A model with just CARS and DU (estimated for the full sample in Problem 3) might prove quite good if estimated for this sub-sample.

Table 8. Household Travel Survey Data used in Problems 1-7

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	37.2	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 trip productions per household
3. HHInc = mean Household income (in $1000s)
4. Cars = number of cars per HH
5. HHS = Household size (persons per HH)
6. D.U. = Dwelling Unit Type (1=single family; 2=multiple family)

Last Updated: 11 May 2022