CEE 123 Transport Systems 3: Planning & Forecasting

Spring 2024: Michael G. Mc Nally (mmcnally-at-uci-dot-edu) [15450]

Homework #5 -- Trip Generation Modeling [S o l u t i o n s]

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].

a. Hypothesize several alternate model structures (via causal arrow diagrams: X -> Y) and then find a valid bivariate trip generation model performing the calculations by hand.

b. Hypothesize several alternate model structures (via causal arrow diagrams: X1, X2 -> 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

Solutions:

a. *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 for these 10 data points, but care should be exercised when developing multivariate models such as Trips=f(Cars, HHSize).*

Sample Hand Calculation Results for Trips(Y)=f[Income(X)] *(any of the four bivariate models could be estimated; hand calcs are shown for only one model (corresponding Excel results are also shown).*

Trips=f(Income)

Obs	Y	X	Y-sq	X-sq	XY
1	4	45.0	16	2025.00	180.0
2	3	40.2	9	1616.04	120.6
3	4	46.5	16	2162.25	186.0
4	5	50.4	25	2540.16	252.0
5	6	57.3	36	3283.29	343.8
6	6	49.8	36	2480.04	298.8
7	7	52.5	49	2756.25	367.5
8	7	55.5	49	3080.25	388.5
9	6	55.8*	36	3113.64	334.8
10	3	42.6	9	1814.76	127.8
sum	51	495.6	281	24871.68	2599.8
mean	5.1	49.56			
std	1.52	5.87			

```
-----------------------------------------------
* HH 9 income was listed as 37.2 in the HW but
  as 55.8 in the Excel spreadsheet (credit will
  be given for either solution if consistent).

b1 = [Sum{XY} - nX'Y']/[Sum{X-sq} - nX'sq]
   = [2599.8-(10)(49.56)(5.1)]/[24871.68-(10)(49.56)(49.56)] =  0.2332

b0 = Y' - b1 X' = 5.1 - (0.2332)(49.56)                       = -6.4586

R = [Sum{XY}-nX'Y']/[Sqrt(Sum{X-sq}-nX'sq) Sqrt(Sum{Y-sq}-nY'sq)]
            [2599.8-(10)(49.56)(5.1)]
  = ---------------------------------------------------       =  0.8978
     Sqrt[24871.68-(10)(49.56)^2] Sqrt[281-(10)(5.1)^2]

R-sq = R(R) = (0.8978)(0.8978)                                 =  0.8061

Sest = Sqrt [(Sum{Y-sq}-b0(Sum{Y})-b1(Sum{XY}))]/(n-k-1)]
     = Sqrt [{281-(-6.4586)(51)-(0.2332)(2599.8)}/(10-1-1)]  =  0.7117

Sb = Sest / [Sx Sqrt(n-1)]  = 0.7117 / [(5.87)(3)]             =  0.0404

t = b1/Sb   =  0.2332/0.0404                                   =  5.7676

Model:
Y (trips) = -6.4586 + 0.2332 X (HHInc)
```

Excel results for Trips=f(Income) ; Trips=f(DUType) ; Trips=f(HHSize):

```
EXCEL: Trips=f(Income)        EXCEL: Trips=f(DUType)        EXCEL: Trips=f(HHSize)
b1 b0     0.23 -6.46          b1 b0    -2.33  8.83          b1 b0     0.82  3.14
se        0.04  2.02          se        0.64  1.07          se        0.71  1.78
R2 SEy    0.81  0.71          R2 SEy    0.63  0.99          R2 SEy    0.14  1.50
F df     33.27  8.00          F, df    13.34  8.00          F, df     1.31  8.00
SSR SSE  16.85  4.05          SSR,SSE  13.07  7.83          SSR,SSE   2.95 17.95
t-stats   5.76 -3.20          t-stats  -3.65  8.26         t-stats    1.15  1.76
```

b. *BURPP! estimation of best multivariate models: For the first 10 observations, there are strong correlations for HHInc and DUType, weaker for Cars and HHSize. Correspondingly, the associated simple regressions are significant for the first two variables. Multivariable models were estimated for all combinations of two explanatory variable but only the model using HHInc and DUType was a significant model (below).*

```
DEPENDENT VARIABLE => Trips   EXPLANATORY VARIABLES: HHInc, DUType

MULTIPLE R    0.9618    * ANOVA *   SUM OF SQR  df    MEAN SQR    F
R-SQUARE      0.9250    REGRESSION      19.33   2       9.67   43.18
ADJ R-SQUARE  0.9036    RESIDUALS        1.57   7       0.22
S.E. OF EST.  0.4731    TOTAL SS        20.90   9

VARIABLE NAME       B         BETA     S.E. B      T
  1. HHInc        0.1722    0.6627    0.0325    5.2909
  2. DUType      -1.2316   -0.4173    0.3696   -3.3317
     Constant    -1.4614
```

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**.

a. Add the remaining 10 households to this table
b. 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.
c. How would this model be used in Trip Generation?

Table 2a. 40 HH Trip Summary

HHInc	HHSize 1-2	HHSize 3+	Row Tot
.LE.	24	4	28
45k	7	1	8
45.1	53	98	151
to 60k	10	15	25
.GT.	0	66	66
60k	0	7	7
Column	77	168	245
Total	17	23	40

Table 2b. 50 HH Trip Summary

HHInc	HHSize 1-2	HHSize 3+	Row Tot
.LE.	42	4	46
45k	14	1	15
45.1	53	110	163
to 60k	10	17	27
.GT.	0	74	74
60k	0	8	8
Column	95	188	283
Total	24	26	50

Table 2c. Trip Generation Model

HHInc	HHSize 1-2	HHSize 3+	Row Tot
.LE.	3.0	4.0	3.1
45k	[2.8]	[3.9]	[2.9]
45.1	5.3	6.5	6.0
to 60k	[5.4]	[6.5]	[6.1]
.GT.	0.0	9.3	9.3
60k	[8.0]	[9.1]	[9.1]
Column	4.0	7.2	5.7
Total	[3.9]	[7.2]	

Solutions:
*In Table 2b, each cell contains the total number of trips followed by the total number of households, for all 50 households. In Table 2c, each cell contains, first, trips per household from the category model (total trips in a cell divided by total households in a cell) and, second, the corresponding trip rate from the regression model (see Problem 3). For the **regression estimates**, the following values were used for the average value of each cell: HHInc[40k, 52k, 64k], HHSize [1.5, 4.0] and row and column sums are weighted averages of row and column estimates. In this case, there is very good agreement between the two models, with the exception being cell (3,1) which, since there were no observations in the category model, did not have a category estimate.*

Problem 3 [10 points]

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

a. Evaluate the regression estimation results statistically.
b. Interpret the model coefficients -- what do these values imply?
c. 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
```

Solutions:

a. Evaluate the estimation results statistically.
The model is statistically significant (t- and F-statistics), has a very good R2, is properly signed, and only minor

indication of multi-colinearity. Care must be exercised given the negative value of the constant since this implies that smaller, low income households would make a negative number of trips, thus, the model is only valid for the range of data used in model estimation.

b. Interpret the model coefficients? What do these values imply?
The coefficient of HHSize implies that an added person in a household generates 0.45 additional trips per household, while each additional $1000 of HH income generates an average 0.22 more trips per HH.

c. Compute regression estimates corresponding to each category model cell. Compare.
See Table 2c. Consider these numbers relative to the discussion in part (b).

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

Table 4b. Results of Validation Test

HH	Obs. Trips	HH Size	HH Inc	Regr. Est.	Table Est.	HH	Obs. Trips	HH Size	HH Inc	Regr. Est.	Table Est.
51	6	3	45.0	4.6	4.0	54	10	5	59.4	8.6	6.5
52	3	2	40.2	3.1	3.0	55	8	4	58.5	8.0	6.5
53	4	1	49.5	4.7	5.2	56	5	2	43.8	3.9	3.0

Results for both models are similar but not extremely close. In the Category Model, categorization of both HHInc and HHSize makes application difficult since the model reflects averages for categories and not actual HHInc or HHSize). The percent errors for individual households can be high, so a measure of overall fit should be estimated.

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

HHSiz / HInc	1-2	3-5	Row Tot
LE 45	0	0	0
45-60	0	60	60
GT 60	0	40	40
Col	0	100	100

TAZ 2

HHSiz / HInc	1-2	3-5	Row Tot
LE 45	40	40	80
45-60	40	80	120
GT 60	20	80	100
Col	100	200	300

TAZ 3

HHSiz / HInc	1-2	3-5	Row Tot
LE 45	30	70	100
45-60	70	20	90
GT 60	0	10	10
Col	100	100	200

For each TAZ, the number of households in each cell is multiplied by the category model trip rate from Table 2b, summing these results over all types of households in each TAZ.

Table 5d. Trip Production Summary

TAZ	HH	Trips	Total
1	100	0(3.0)+ 0(5.3)+ 0(0.0)	
		+ 0(4.0)+ 60(6.5)+ 40(9.3) =	762
2	300	40(3.0)+ 40(5.3)+ 20(5.7)	
		+40(4.0)+ 80(6.5)+ 80(9.3) =	1756 + 114
3	200	30(3.0)+ 70(5.3)+ 0(0.0)	
		+70(4.0)+ 20(6.5)+ 10(9.3) =	964
Tot	600		3482 3596

Note: An average of 5.7 trips/HH over 3 TAZs
Note: Since no rate was available for cell (3,1), the table average was used, adding 114 trips.

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:

$A_j = 1.5\ POP_j + 3.0\ EMP_j$

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
2	300	1100	400
3	200	600	100
Tot	600	2000	500

HH = total households

POP = total population

EMP = total employment

Table 6b. Demographic Data Summary

TAZ	HH	POP	EMP	Prod.	Trip Attractions	Total	Norm A
1	100	300	0	762	1.5(300)+3.0(0) =	450	360
2	300	1100	400	1870	1.5(1100)+3.0(400) =	2850	2277
3	200	600	100	964	1.5(600)+3.0(100) =	1200	959
Tot	600	2000	500	3596		4500	3596

The "Prod" column contains the trip production totals from Table 5d. The trip attraction column provides the calculations using the estimated regression equation provided. The "Norm A" column contains the normalized (balanced) attractions. These numbers are the raw attractions multiplied by the balancing factor of 0.7991 (the ratio of total productions, 3496, to total attractions, 4500).

Table 7. Household Travel Survey Data Not Shown in Solutions

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).

Solutions: Available to those who complete the Extra Credit Problem.

Appendix
BURPP! output for 2 variable regressions using first 10 data points.

```
--------------------------------------------------------------------------
----------    O R D I N A R Y   L E A S T   S Q U A R E S    ----------
--------------------------------------------------------------------------

--- BASIC STATISTICS ---            --- CORRELATIONS ---
VARIABLE NAME   MEAN    S.D.  OBS    Trips     HHInc     Cars      HHSize    DUType
 1. Trips     5.1000  1.5239   10    1.0000    0.8978    0.4801    0.3754   -0.7907
 2. HHInc    49.5600  5.8665   10    0.8978    1.0000    0.6030    0.3186   -0.5634
 3. Cars      1.6000  0.5164   10    0.4801    0.6030    1.0000    0.8001   -0.2500
 4. HHSize    2.4000  0.6992   10    0.3754    0.3186    0.8001    1.0000   -0.4308
 5. DUType    1.6000  0.5164   10   -0.7907   -0.5634   -0.2500   -0.4308    1.0000


  * * * * * * *   O R D I N A R Y   L E A S T   S Q U A R E S   * * * * * * *

  DEPENDENT VARIABLE => Trips

  MULTIPLE R      0.9028     * ANOVA *    SUM OF SQR  df     MEAN SQR    F
  R-SQUARE        0.8150     REGRESSION        17.03   2         8.52   15.42
  ADJ R-SQUARE    0.7622     RESIDUALS          3.87   7         0.55
  S.E. OF EST.    0.7432     TOTAL SS          20.90   9

            VARIABLE NAME       B         BETA     S.E. B     T
              1. HHInc        0.2250    0.8662    0.0445    5.0507
              2. HHSize       0.2168    0.0995    0.3738    0.5802
                 Constant    -6.5710


  * * * * * * *   O R D I N A R Y   L E A S T   S Q U A R E S   * * * * * * *

  DEPENDENT VARIABLE => Trips

  MULTIPLE R      0.9011     * ANOVA *    SUM OF SQR  df     MEAN SQR    F
  R-SQUARE        0.8120     REGRESSION        16.97   2         8.49   15.12
  ADJ R-SQUARE    0.7583     RESIDUALS          3.93   7         0.56
  S.E. OF EST.    0.7491     TOTAL SS          20.90   9

            VARIABLE NAME       B         BETA     S.E. B     T
              1. HHInc        0.2483    0.9559    0.0534    4.6539
              2. Cars        -0.2843   -0.0963    0.6061   -0.4690
                 Constant    -6.7516


  * * * * * * *   O R D I N A R Y   L E A S T   S Q U A R E S   * * * * * * *

  DEPENDENT VARIABLE => Trips

  MULTIPLE R      0.9618     * ANOVA *    SUM OF SQR  df     MEAN SQR    F
  R-SQUARE        0.9250     REGRESSION        19.33   2         9.67   43.18
  ADJ R-SQUARE    0.9036     RESIDUALS          1.57   7         0.22
  S.E. OF EST.    0.4731     TOTAL SS          20.90   9
```

```
          VARIABLE NAME       B         BETA      S.E. B      T
            1. HHInc        0.1722    0.6627    0.0325    5.2909
            2. DUType      -1.2316   -0.4173    0.3696   -3.3317
               Constant    -1.4614


 * * * * * * *   O R D I N A R Y   L E A S T   S Q U A R E S   * * * * * * *

 DEPENDENT VARIABLE => Trips

 MULTIPLE R      0.7916     * ANOVA *    SUM OF SQR  df     MEAN SQR    F
 R-SQUARE        0.6267     REGRESSION        13.10   2         6.55    5.88
 ADJ R-SQUARE    0.5200     RESIDUALS          7.80   7         1.11
 S.E. OF EST.    1.0558     TOTAL SS          20.90   9

          VARIABLE NAME       B         BETA      S.E. B      T
            1. HHSize       0.0930    0.0427    0.5577    0.1668
            2. DUType      -2.2791   -0.7723    0.7552   -3.0180
               Constant     8.5233


 * * * * * * *   O R D I N A R Y   L E A S T   S Q U A R E S   * * * * * * *

 DEPENDENT VARIABLE => Trips

 MULTIPLE R      0.4803     * ANOVA *    SUM OF SQR  df     MEAN SQR    F
 R-SQUARE        0.2307     REGRESSION         4.82   2         2.41    1.05
 ADJ R-SQUARE    0.0109     RESIDUALS         16.08   7         2.30
 S.E. OF EST.    1.5156     TOTAL SS          20.90   9

          VARIABLE NAME       B         BETA      S.E. B      T
            1. Cars         1.4737    0.4994    1.6309    0.9036
            2. HHSize      -0.0526   -0.0241    1.2045   -0.0437
               Constant     2.8684
```

Last Updated: 11 May 2024