

Econometrics 1 Project Holidays budget

Frederic ANDRIEU and Nicolas DORIGNY

ENAC

May 21, 2014

Introduction

1 - Model

2 - Model conclusions

3 - Limits of the model

Conclusion



Objectives

- Identify the pertinent parameters for the modelization of travelling behaviours.
- Build a representative model.

Getting the inputs : the survey

The target : travellers

Using a Google Form we have collected 85 answers from people of different age and social background but only 65 were workable.

The screenshot shows a Google Form with the following content:

- Page 1 de 2
- Choix de vacances / Choices for your journeys**
- Les informations demandées dans ce questionnaire sont collectées de manière anonyme et ne nous serviront que pour la durée de notre étude économique. Nous ne conserverons pas ces données pour un quelconque autre usage.
The information asked in this poll are anonymous and will be used only for our economic study. No data will be stored for any other use.
- Vous êtes... / You are...
 - Un homme / A man
 - Une femme / A woman
- Nationalité / Nationality*
- Age*
- Situation de famille / Marital status*
 - Célibataire / Single
 - Divorcé / Divorced
 - Marié / Married
 - En couple / In a relationship

Figure: Screenshot of our Google Form

Expected effects

Explanatory variable	Details	Expected effects
Sex	M/F	?
Nation	Nationality (French/Other)	?
Age	Age (Number)	+
Married	Y/N	-
Couple	In a relationship (Y/N)	+
Single	Y/N	+
Child	Number	-
House	Living in a house (Y/N)	+
Flat	Living in a flat (Y/N)	-
Hall	Living in a residence hall (Y/N)	+
Income	Household income per year	+
Trips	Number of trips in the year	-
Car	Use of car during the trip	?
Plane	Use of airlines during the plane	+

Expected effects (2)

Explanatory variable	Details	Expected effect
Other	Use of other	?
Stu	Student (Y/N)	-
Act	Working (Y/N)	+
Retired	Retired (Y/N)	+
Price	Price of the last trip	+
Nb	Number of trips during the last 12 months	+
Time	Time of preparation	+
Book	Use of travel guide	?
Web	Use of website	?
Agency	Service of travel agency	?

Correlation

We have too many variables to build a model so we use the correlation matrix to see if some can be removed

	ACT	AGE	AGENCY	BOOK	CAR	CHLD	COUPLE	FLAT	HALL	HOL1	HOUSE	INCOME	NATION	NB	OTHER	PLANE	PRICE	RETIRED	SEX	SINGLE	STUD	TIME	WEB	
ACT	1																							
AGE	0.2508	1																						
AGENCY	0.2186	0.3197	1																					
BOOK	-0.0236	-0.0907	-0.2253	1																				
CAR	0.1465	0.0477	-0.2357	0.0127	1																			
CHLD	0.2856	0.8038	0.0851	0.0470	0.1345	1																		
COUPLE	-0.1465	-0.2231	0.0000	-0.0127	-0.0385	-0.2306	1																	
FLAT	0.1548	0.0357	0.1380	-0.1961	0.0390	-0.1185	0.1592	1																
HALL	-0.2415	-0.3549	-0.1867	0.1327	-0.0880	-0.2957	-0.0377	-0.8013	1															
HOL1	0.1481	-0.0723	0.0860	0.0914	0.1823	-0.0861	0.0203	0.0427	-0.0657	1														
HOUSE	0.1481	0.5162	0.0860	0.0914	0.0810	0.0610	-0.1823	-0.2958	-0.3638	0.0394	1													
INCOME	0.6365	0.5978	0.1840	0.0411	0.1497	0.6923	-0.1908	0.1122	-0.3594	-0.0112	0.4045	1.0000												
NATION	-0.0044	-0.1852	0.1003	0.1100	0.0203	-0.2814	0.0610	-0.0427	0.1651	0.1207	-0.1995	-0.2710	1											
NB	0.0738	-0.0883	0.1851	-0.1817	0.1370	-0.0571	-0.1405	-0.1214	0.1621	-0.0481	-0.0721	-0.0699	0.0425	1										
OTHER	-0.1597	-0.0711	-0.1855	0.1985	-0.0967	-0.0780	0.1657	-0.2814	0.2751	-0.1142	-0.0050	-0.0726	0.0050	-0.1161	1									
PLANE	0.0910	0.2154	0.1429	0.1227	-0.5433	0.1114	-0.0884	0.0454	-0.0515	-0.0676	0.0123	0.0542	0.0876	-0.2149	-0.1717	1								
PRICE	0.1132	0.3237	0.1774	0.1000	-0.0550	0.1484	0.0548	0.0480	-0.0800	0.2368	0.0525	0.2379	0.0288	-0.1260	-0.0941	0.4304	1							
RETIRED	-0.0882	0.8576	0.2117	-0.1122	0.0296	0.6108	-0.1796	-0.0164	-0.2304	-0.0794	0.3967	0.3186	-0.1601	-0.0527	0.0198	0.1974	0.3106	1						
SEX	-0.0111	-0.0233	-0.0987	-0.2289	0.1776	0.0060	-0.3044	-0.2357	0.3512	-0.0987	-0.1989	-0.0329	0.0987	0.1250	-0.1923	-0.0308	0.0620	0.0433	1					
SINGLE	0.0137	-0.2619	-0.0802	0.0337	0.0000	-0.3050	-0.8189	-0.1230	0.2281	0.0230	-0.1761	-0.1763	0.0766	0.1653	-0.0635	-0.0239	-0.1773	-0.2376	0.3164	1				
STUD	-0.8425	-0.8871	-0.3091	0.0617	-0.1457	-0.5934	0.2295	-0.1269	0.3395	-0.0905	-0.3464	-0.7389	0.0905	-0.0426	0.1314	-0.1878	-0.2667	-0.4823	-0.0136	0.1163	1			
TIME	-0.0473	-0.0857	-0.0207	0.1527	-0.2062	-0.1254	0.0556	0.1040	-0.0882	0.0896	-0.0519	-0.0343	0.0685	-0.0422	-0.0079	0.1695	0.4787	-0.0940	-0.1837	0.0023	0.0929	1		
WEB	-0.1341	-0.1828	-0.4492	-0.2672	0.0997	-0.2324	0.1795	-0.1391	0.1990	0.1142	-0.1041	-0.2484	-0.0050	-0.0426	0.1156	-0.0325	0.0252	-0.0198	0.2606	-0.0522	0.1301	-0.1168	1	

Removing : RETIRED and STUDENT

Linear regression

After a first linear regression we find the following equation :

$$\begin{aligned} PRICE_i = & \beta_0 + \beta_1 ACT + \beta_2 AGE + \beta_3 AGENCY + \beta_4 BOOK + \beta_5 CAR + \\ & \beta_6 CHILD + \beta_7 COUPLE + \beta_8 FLAT + \beta_9 HOL1 + \beta_{10} HOUSE + \\ & \beta_{11} INCOME + \beta_{12} NATION + \beta_{13} NB + \beta_{14} OTHER + \beta_{15} PLANE + \\ & \beta_{16} SEX + \beta_{17} SINGLE + \beta_{18} TIME + \beta_{19} WEB \end{aligned}$$

We obtain

$$R^2 = 0.656$$

β_i	Value	Prob	β_i	Value	Prob
β_0	-1969.6	0.1518	β_{10}	-86.1	0.8882
β_1	-495.9	0.3469	β_{11}	0.02	0.1050
β_2	51.9	0.0433	β_{12}	-364.3	0.4167
β_3	673.8	0.3634	β_{13}	1.0	0.8400
β_4	251.9	0.4470	β_{14}	128.4	0.6966
β_5	718.6	0.0504	β_{15}	1280.3	0.0008
β_6	-879.4	0.1292	β_{16}	657.4	0.0587
β_7	39.6	0.9692	β_{17}	-553.3	0.5901
β_8	-208.4	0.5352	β_{18}	21.8	0.0000
β_9	971	0.0356	β_{19}	205.3	0.6198

Model conclusions

- The probability of the variables give their importance in the model.
- We can draw conclusions

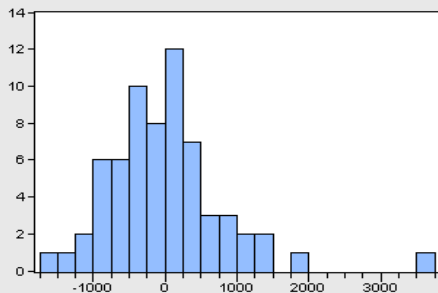
Significant parameters

Explanatory variable	Effect
Sex	+ 657 if boy
Age	+ 52 per year
Child	- 879 per child
Income	+ 10 per 1000 revenue
Car	+ 718 if use of car
Plane	+ 1280 if use of plane
Time	+ 22 per day of preparation

Analysis

- Increase of income affects slightly the budget
- The marital status has no effect
- The sex has a strong effect
- The living conditions has no effect

The resid do not follow a normal law



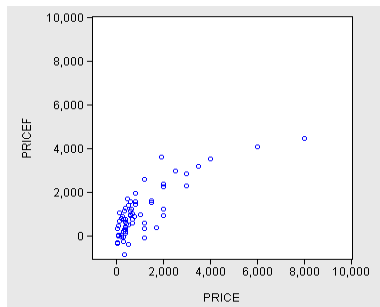
Series: RESID	
Sample 1 65	
Observations 65	
Mean	1.02e-13
Median	-19.60355
Maximum	3563.328
Minimum	-1691.715
Std. Dev.	814.7067
Skewness	1.346086
Kurtosis	7.453168
Jarque-Bera	73.33759
Probability	0.000000

Heteroskedasticity

Our case is at the limit so we reject the hypothesis

Heteroskedasticity Test: White				
F-statistic	1.844096	Prob. F(19,45)	0.0468	
Obs*R-squared	28.45478	Prob. Chi-Square(19)	0.0751	
Scaled explained SS	44.00444	Prob. Chi-Square(19)	0.0009	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 05/20/14 Time: 17:35				
Sample: 1 65				
Included observations: 65				
	Coefficient	Std. Error	t-Statistic	Prob.
C	351375.3	1819551.	0.193111	0.8477
ACT^2	-328773.6	742511.9	-0.442786	0.6600
AGE^2	32.48558	400.2567	0.081162	0.9357
AGENCY^2	-740557.5	1126556.	-0.657364	0.5143
BOOK^2	149277.5	504985.6	0.295608	0.7689
CAR^2	820673.0	545623.5	1.504101	0.1395
CHILD^2	-168261.9	280409.2	-0.600058	0.5515
COUPLE^2	656073.5	1526976.	0.429655	0.6695
FLAT^2	-756574.0	509030.7	-1.486303	0.1442
HOL1^2	2060146.	681806.0	3.021602	0.0041
HOUSE^2	-391019.6	930677.1	-0.420145	0.6764
INCOME^2	1.68E-05	0.000117	0.144014	0.8861
NATION^2	-952443.9	695488.1	-1.369461	0.1777
NB^2	5.261451	39.75431	0.132349	0.8953
OTHER^2	-331665.2	503175.2	-0.659145	0.5132
PLANE^2	1359519.	543488.1	2.501469	0.0161
SEX^2	595908.5	520614.3	1.144626	0.2584
SINGLE^2	-338646.8	1521325.	-0.221285	0.8259

Comparison model vs reality



- High budgets are not well modeled
- Some computed budgets are negative

Solutions to improve the model

Deleting the high budgets

We don't have enough data around the high budgets to build a good model

Problems with some negative budgets

We have to check the calculation of the probabilities and factors for some variables

Improvements

By deleting two observations of high budgets, the new model gives

$$R^2 = 0.77$$

