

Spring 2019

BUAN 6337: Predictive Analytics using SAS

A Report of

Group Assignment 5 on Smartphone Data Conjoint Analysis



**THE UNIVERSITY
OF TEXAS AT DALLAS**

Submitted by Group 7

Vinay Singh	2021441554
Vaibhav Shrivastava	2021434681
Megan Malisani	2021440151
Pragati Mishra	2021434655
Ishan Jain	2021426222
Erhao Liang	2021435949

Under the Guidance of

Prof. Shervin Shahrokhi Tehrani

Overview

The dataset **smart.csv** has data on a conjoint analysis of important attributes of a smartphone by consumers. Each observation denotes the consumer decision on a set of choices. The description of columns is as follows:

1. **ID:** = Subject's ID
2. **Question:** = The number of the question which has been asked from the subject
3. **Alternatives:** = The alternative index in the corresponding question
4. **UsingFORInternet:** = denotes will subject uses the phone only as a phone (i.e., "no") or as a device to connect to internet (i.e., "yes")
5. **InternetConnetivity:** = How well the alternative's capability is in connecting to Internet
6. **Design:** = How well the alternative's design
7. **OperatingSystem:** = What is the operating system of the phone.
8. **Price:** = The price of the alternative
9. **Choice:** = Which alternative was chosen by the subject

Summary Statistics

Distribution of internet connectivity as per choice (0/1)

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of InternetConnectivity by Choice			
	InternetConnectivity	Choice		
		0	1	Total
Excellent		2001	982	2983
		22.23	10.91	33.14
		67.08	32.92	
		33.35	32.73	
Good		2139	854	2993
		23.77	9.49	33.26
		71.47	28.53	
		35.65	28.47	
Normal		1860	1164	3024
		20.67	12.93	33.60
		61.51	38.49	
		31.00	38.80	
Total		6000	3000	9000
		66.67	33.33	100.00

<u>Internet Connectivity</u>	<u>Not Buy%</u>	<u>Buy%</u>
Excellent	22.23	10.91
Good	23.77	9.49
Normal	20.67	12.93
Total	66.67	33.33

Number of Normal phone buyer is highest and that of good is least.

Distribution of Design as per choice (0/1)

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of Design by Choice			
	Design	Choice		
		0	1	Total
Fancy		2811	1688	4499
		31.23	18.76	49.99
		62.48	37.52	
		46.85	56.27	
Normal		3189	1312	4501
		35.43	14.58	50.01
		70.85	29.15	
		53.15	43.73	
Total		6000	3000	9000
		66.67	33.33	100.00

<u>Internet Connectivity</u>	<u>Not Buy%</u>	<u>Buy%</u>
Fancy	31.23	18.76
Normal	35.43	14.58
Total	66.67	33.33

Number of people buying fancy phone is more than that of people buying normal phone.

Summary of Operating system as per choice (0/1)

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of OperatingSystem by Choice			
	OperatingSystem	Choice		Total
		0	1	
Andriod		1561	1444	3005
		17.34	16.04	33.39
		51.95	48.05	
		26.02	48.13	
Mac		2037	948	2985
		22.63	10.53	33.17
		68.24	31.76	
		33.95	31.60	
Windows		2402	608	3010
		26.69	6.76	33.44
		79.80	20.20	
		40.03	20.27	
Total		6000	3000	9000
		66.67	33.33	100.00

<u>Internet Connectivity</u>	<u>Not Buy%</u>	<u>Buy%</u>
Android	17.34	16.04
Mac	22.63	10.53
Windows	26.69	6.76
Total	66.67	33.33

Number of people buying Android is highest and buying windows phone is least.

Without Heterogeneity

There are 200 respondents ID information with 3000 Observations in our dataset consisting of customer's choices of alternative of smartphone combination. We get our estimates result by a logistics model based on potential utility assumption.

Total number of combination = 3(Price) * 3(InternetConnectivity) * 2(Design) * 3(OperatingSystem) = 54 choices.

We choose the reference attribute levels as:

InternetConnectivity=Normal; Design=Normal; OperatingSystem=Andriod;

Here is the Utility Formula:

$$U_{ij} = \beta X + \varepsilon = V_{ij} + \varepsilon_{ij}$$

$$V_{ij} = \beta Price_i + \alpha_{interConn\cdot good} I_{i,interConn\cdot good} + \alpha_{interConn\cdot excellent} I_{i,interConn\cdot excellent} \\ + \alpha_{i,Design\cdot Fancy} I_{i,Design\cdot Fancy} + \alpha_{i,OS\cdot Mac} I_{i,OS\cdot Mac} + \alpha_{i,OS\cdot Mac} I_{i,OS\cdot Mac}$$

Analysis of Conditional Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
InternetConnectivity	Excellent	1	-0.3061	0.0611	25.0844	<.0001
InternetConnectivity	Good	1	-0.5345	0.0624	73.4956	<.0001
Design	Fancy	1	0.4767	0.0509	87.8359	<.0001
OperatingSystem	Andriod	1	0.8107	0.0601	181.8833	<.0001
OperatingSystem	Windows	1	-0.7184	0.0655	120.3060	<.0001
Price		1	-0.00433	0.000173	623.6288	<.0001

Internet Connectivity: Normal internet connectivity has higher utility than excellent and good option.

Design: Fancy design is having higher utility than other (normal) available option.

Operating system: Android has higher utility as compared with Mac and Windows available option.

Price: Price has negative utility. As price increases, customer willingness to buy will decrease.

Internet	Design	Operating system	Price
Normal	Fancy	Android	Low-price

With Heterogeneity

There are 200 respondents ID information with 3000 Observations in our dataset consisting of customer's choices of alternative of smartphone combination. We get our estimates result by a logistics model based on potential utility assumption.

Total number of combination = 3(Price) * 3(InternetConnectivity) * 2(Design) * 3(OperatingSystem) = 54 choices.

We choose the reference attribute levels as:

InternetConnectivity=Normal; Design=Normal; OperatingSystem=Andriod; UsingFORInternet=No.

Here is the Utility Formula:

$$\begin{aligned}
 U_{ij} &= \beta X + \varepsilon = V_{ij} + \varepsilon_{ij} \\
 V_{ij} &= \beta Price_i + \alpha_{interConn\cdot good} I_{i,interConn\cdot good} + \alpha_{interConn\cdot excellent} I_{i,interConn\cdot excellent} \\
 &\quad + \alpha_{i,Design\cdot Fancy} I_{i,Design\cdot Fancy} + \alpha_{i,OS\cdot Mac} I_{i,OS\cdot Mac} + \alpha_{i,OS\cdot Mac} I_{i,OS\cdot Mac} \\
 &\quad + I_{i,UsingFORInternet} \times (\alpha_{interConn\cdot good} I_{i,interConn\cdot good} \\
 &\quad + \alpha_{interConn\cdot excellent} I_{i,interConn\cdot excellent} + \alpha_{i,Design\cdot Fancy} I_{i,Design\cdot Fancy} + \alpha_{i,OS\cdot Mac} I_{i,OS\cdot Mac} \\
 &\quad + \alpha_{i,OS\cdot Mac} I_{i,OS\cdot Mac})
 \end{aligned}$$

Analysis of Conditional Maximum Likelihood Estimates

Parameter			DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
InternetConnectivity	Excellent		1	-0.6876	0.0741	86.1290	<.0001
InternetConnectivity	Good		1	-0.9075	0.0761	142.1633	<.0001
Design	Fancy		1	0.4156	0.0608	46.7811	<.0001
OperatingSystem	Mac		1	-0.8199	0.0725	127.7311	<.0001
OperatingSystem	Windows		1	-1.5469	0.0803	371.2232	<.0001
Price			1	-0.00416	0.000206	407.5933	<.0001
InternetC*UsingFORIn	Excellent	yes	1	1.3643	0.1409	93.7633	<.0001
InternetC*UsingFORIn	Good	yes	1	1.3070	0.1412	85.7038	<.0001
Design*UsingFORInter	Fancy	yes	1	0.2581	0.1154	4.9990	0.0254
Operating*UsingFORIn	Mac	yes	1	-0.1078	0.1361	0.6273	0.4284
Operating*UsingFORIn	Windows	yes	1	-0.1391	0.1546	0.8088	0.3685
Price*UsingFORIntern	yes		1	-0.00117	0.000407	8.1974	0.0042

Paramater	Parameter-2	Estimate	Probability
Internet connectivity	Excellent	-0.6876	0.503
Internet connectivity	Good	-0.9075	0.404
Design	Fancy	0.4156	1.515
Operating System	Mac	-0.8199	0.440
Operatinf system	Windows	-1.5469	0.213
Price		-0.00416	0.996
Internet connectivity*Using for internet	Excellent	1.3643	3.913
Internet connectivity*Using for internet	Good	1.307	3.695
Design*Using for Internet	Fancy	0.2581	1.294
Operating *using for internet	Mac	-0.1078	0.898
Operating *using for internet	Windows	-0.1391	0.870
Price*using for internet	Yes	-0.00117	0.999

Factors play important role in choosing a smart phone:

1) With considering the factor of using for internet as NO:

Internet Connectivity: Normal internet connectivity has higher utility than excellent and good option.

Design: Fancy design is having higher utility than other (normal) available option.

Operating system: Android has higher utility as compared with Mac and Windows available option.

Price: Price has negative utility. As price increases, customer willingness to buy will decrease.

So, If using for internet is NO, then following factors are most important among their categories:

Internet	Design	Operating system	Price
Normal	Fancy	Android	Low-price

2) With considering the factor of “using for internet” as YES:

Internet connectivity: Excellent internet connectivity has higher utility than other two option.

Design: Fancy design is having higher utility than other available option.

Operating system: Android has higher utility than other available option.

Price: Price has negative utility. As price increases, customer willingness to buy will decrease.

So, If using for internet is YES, then following factors are most important among their categories:

Internet	Design	Operating system	Price
Excellent	Fancy	Android	Low-price

Conclusion:

Firm should make phone with design-fancy, operating system-android, keeping price low. Now customer who use internet need internet connectivity as excellent and those who doesn't use internet need as normal internet phone.

So, considering both groups, below combination will work best:

Internet	Design	Operating system	Price
Excellent	Fancy	Android	Low-price

Appendix

```

/* reading the file */
PROC IMPORT OUT= WORK.Assign5
            DATAFILE= "H:\My SAS Files\Assign 5\smartphone.csv"
            DBMS=CSV REPLACE;
            GETNAMES=YES;
            DATAROW=2;
RUN;

PROC FREQ DATA=Assign5;
    TABLES InternetConnectivity*Choice ;
RUN;

PROC FREQ DATA=Assign5;
    TABLES Design*Choice ;
RUN;

PROC FREQ DATA=Assign5;
    TABLES OperatingSystem*Choice ;
RUN;

PROC FREQ DATA=Assign5;
    TABLES InternetConnectivity*Design*Price*OperatingSystem*Choice ;
RUN;

/* Partial Factorial Design with Multiple Observations per Individual */
proc logistic data = Assign5;
    strata id question;
    class InternetConnectivity (ref = 'Normal') Design (ref = 'Normal') OperatingSystem
(ref = 'Mac') / param= ref;
    model choice (event = '1') = InternetConnectivity Design OperatingSystem price;
run;

/* We can take into account the heterogeneity of people in using For Internet*/
proc logistic data = Assign5;
    strata id question;
    class UsingFORInternet (ref='no') InternetConnectivity (ref = 'Normal') Design (ref =
'Normal') OperatingSystem (ref = 'Mac') /
param= ref;
    model choice (event = '1') = InternetConnectivity Design OperatingSystem price
InternetConnectivity*UsingFORInternet
Design*UsingFORInternet OperatingSystem*UsingFORInternet Price*UsingFORInternet;
run;

```