

# Conjoint Analysis

## Multiple Regression Method

## Traditional Conjoint Analysis

A traditional conjoint analysis may be thought of as a multiple regression problem. The respondent's ratings for the product concepts are observations on the dependent variable. The characteristics of the product or attribute levels are observations on the independent or predictor variables. The estimated regression coefficients associated with the independent variables are the part-worth utilities or preference scores for the levels. The  $R^2$  for the regression characterizes the internal consistency of the respondent.

Consider a conjoint analysis problem with three attributes, each with levels as follows:

Brand - A B C

Color - Black Green

Price - \$100 \$150 \$250

For simplicity, let us consider a full-factorial experimental design. A full-factorial design includes all possible combinations of the attributes. There are 18 possible product concepts or cards that can be created from these three attributes:

$3 \text{ brands} \times 2 \text{ colors} \times 3 \text{ prices} = 18 \text{ cards}$

Further assume that respondents rate each of the 18 product concepts on a scale from 0 to 10, where 10 represents the highest degree of preference. See table below.

Card	Brand	Color	Price (\$)
1	A	Black	100
2	A	Black	150
3	A	Black	250
4	A	Green	100
5	A	Green	150
6	A	Green	250
7	B	Black	100
8	B	Black	150
9	B	Black	250
10	B	Green	100
11	B	Green	150
12	B	Green	250
13	C	Black	100
14	C	Black	150
15	C	Black	250
16	C	Green	100
17	C	Green	150
18	C	Green	250

Zero Code conjoint design app uses partial factorial designs to reduce the number of cards. e.g. instead of rating 18 cards, with Zero Code design same information can be collected from respondent using fewer cards ( check how many cards in app) ....

We can use Microsoft Excel to analyze data from traditional conjoint questionnaires. Multiple regression functions come from the Excel Analysis ToolPak add-in.

## Data Organization and Coding

Assume the data for one respondent have been entered into an Excel spreadsheet. Assume the first card is made up of the first level on each of the attributes: (Brand A, Black, \$100). The respondent rated that card a 3 on the preference scale. The second card has the first level on brand and color and the second level on price: (Brand A, Black, \$150). This card gets a 3 on the preference scale. And so on.

Card	Brand	Color	Price (\$)	Preference
1	1	1	100	3
2	1	1	150	3
3	1	1	250	0
4	1	2	100	6
5	1	2	150	3
6	1	2	250	1
7	2	1	100	3
8	2	1	150	3
9	2	1	250	1
10	2	2	100	10
11	2	2	150	4
12	2	2	250	4
13	3	1	100	8
14	3	1	150	5
15	3	1	250	3
16	3	2	100	7
17	3	2	150	5
18	3	2	250	4

After collecting the respondent data, the next step is to code the data in an appropriate manner for estimating utilities using multiple regression. We use a procedure called dummy coding for the independent variables or product characteristics. In its simplest form, dummy coding uses a 1 to reflect the presence of a feature, and a 0 to represent its absence. The brand attribute would be coded as three separate columns, color as two columns, and price as

three columns. Applying dummy coding results in an array of columns. Please see the table below after encoding.

To this point, the coding has been straightforward. But there is one complication that must be resolved. In multiple regression analysis, no independent variable may be perfectly predictable based on the state of any other independent variable or combination of independent variables. If so, the regression procedure could not separate the effects of the confounded variables. We have that problem with the data, since, for example, we can perfectly predict the state of Brand A based on the states of Brand B and Brand C. This situation is called linear dependency.

To resolve this linear dependency, we omit one column from each attribute. It really doesn't matter which column (level) we drop, and for this example we have excluded the first level for each attribute.

Card	Brand B	Brand C	Green	150 \$	250\$	Preference
1	0	0	0	0	0	3
2	0	0	0	1	0	3
3	0	0	0	0	1	0
4	0	0	1	0	0	6
5	0	0	1	1	0	3
6	0	0	1	0	1	1
7	1	0	0	0	0	3
8	1	0	0	1	0	3
9	1	0	0	0	1	1
10	1	0	1	0	0	10
11	1	0	1	1	0	4
12	1	0	1	0	1	4
13	0	1	0	0	0	8
14	0	1	0	1	0	5
15	0	1	0	0	1	3
16	0	1	1	0	0	7
17	0	1	1	1	0	5
18	0	1	1	0	1	4

Even though it appears that one level from each attribute is missing from the data, they are really implicitly included as reference levels for each attribute. The explicitly coded levels are estimated as contrasts with respect to the omitted levels, which are defined as 0.

## **Multiple Regression Analysis**

Microsoft Excel offers a simple multiple regression tool (under Tools + Data Analysis + Regression with the Analysis Toolpak add-in installed). Using the tool, you can specify the preference score (column Y) as the dependent variable (Input Y Range) and the five dummy-coded attribute columns as independent variables (Input X range). You should also make sure a constant is estimated; this usually happens by default (by not checking the box labeled “Constant is zero”).

The mathematical expression of the model is as follows:

$$Y = b_0 + b_1(\text{Brand B}) + b_2(\text{Brand C}) + b_3(\text{Green}) + b_4(\$150) + b_5(\$250) + e$$

where Y is the respondent’s preference for the product concept,  $b_0$  is the constant or intercept term,  $b_1$  through  $b_5$  are beta weights (part-worth utilities) for the features, and e is an error term. In this formulation of the model, coefficients for the reference levels are equal to 0. The solution minimizes the sum of squares of the errors over all observations.

A portion of the output from Excel is illustrated below. Using that output (after rounding to two decimal places of precision), the utilities (coefficients) are the following:

Brand	Color	Price
A = 0.00	Black = 0.00	\$ 100 = 0.00
B = 1.5	Green = 1.67	\$ 150 = -2.33
C = 2.67		\$ 250 = -4.00

The constant or intercept term is 3.94, and the fit for this respondent  $R^2 = 0.77$ . The fit values range from a low of 0 to a high of 1.0. The standard errors of the regression coefficients (betas) reflect how precisely we are able to estimate those coefficients with this design. Lower standard errors are better. The remaining statistics presented in Excel's output are beyond the scope of this chapter and are generally not of much use when considering individual-level conjoint analysis problems.

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.87771841							
R Square	0.77038961							
Adjusted R S	0.67471861							
Standard Err	1.43048813							
Observations	18							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	5	82.3888889	16.4777778	8.05248869	0.00154741			
Residual	12	24.5555556	2.0462963					
Total	17	106.944444						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	3.94444444	0.82589271	4.77597684	0.00045164	2.14497882	5.74391007	2.14497882	5.74391007
X Variable 1	1.5	0.82589271	1.81621654	0.09438581	-0.2994656	3.29946563	-0.2994656	3.29946563
X Variable 2	2.66666667	0.82589271	3.22882941	0.007235	0.86720104	4.4661323	0.86720104	4.4661323
X Variable 3	1.66666667	0.67433857	2.47155766	0.02940946	0.19740913	3.1359242	0.19740913	3.1359242
X Variable 4	-2.33333333	0.82589271	-2.8252257	0.01531077	-4.132799	-0.5338677	-4.132799	-0.5338677
X Variable 5	-4	0.82589271	-4.8432441	0.00040283	-5.7994656	-2.2005344	-5.7994656	-2.2005344

## **Conjoint analysis with multiple regression**

Only in the smallest of problems (such as our 18-card example) would we ask people to respond to all possible combinations of attribute levels. Large full-factorial designs are not practical. Fortunately, Zero Code conjoint app finds efficient fractional-factorial designs. Fractional-factorial designs show an efficient subset of the possible combinations and provide enough information to estimate utilities.