MARKETING ANALYTICS MGT7215

Assignment 2

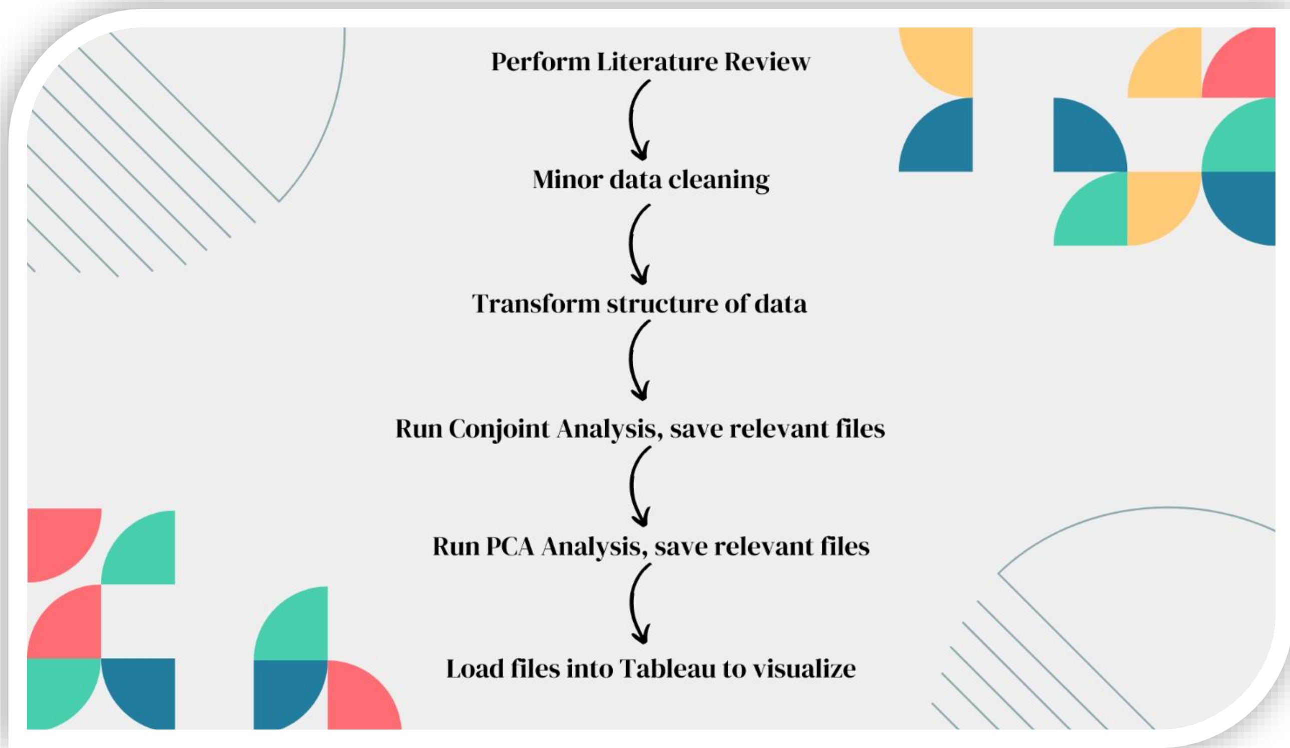
Muhammad Muneeb Ullah Ansari
40426685

Word Count: 2166

Contents

Infographic

The following is an infographic that summarizes the main tasks carried out throughout this assignment.



Introduction and Background

Conjoint analysis and Principal Component Analysis (PCA) are pivotal in modern product design, allowing firms to decode complex consumer preferences and optimize product attributes effectively. Conjoint analysis, for instance, enables the identification of preferred product attributes by analyzing trade-offs consumers make among them, thereby guiding critical product development decisions (Asioli et al., 2014).

This method is often complemented by PCA, which simplifies the interpretation of multivariate data from consumer surveys, reducing dimensionality to uncover underlying patterns in consumer preferences (Almli and Næs, 2018).

These analytical approaches have been instrumental in diverse applications, from determining the design of synthetic rattan sofas through the integration of Kansei engineering and conjoint analysis (Oey et al., 2022), to streamlining the House of Quality process in quality management frameworks (Prasad and Subbaiah, 2011).

Furthermore, these methodologies support the alignment of product designs with strategic market positioning and customer expectations, exemplified in studies like the sensory optimization of food products (Varela and Ares, 2012) and the design of bicycle-related apparel (Tseng et al., 2011).

Methodology

In conjoint analysis, an optimal set of product profiles typically aims for low or no correlation between product attributes hence low collinearity, known as orthogonality. An orthogonal design is statistically efficient, which means that it requires fewer product profiles to estimate the main effects accurately. This efficiency is crucial for avoiding respondent fatigue and keeping the survey length reasonable. In practice, software used for conjoint analysis often includes tools to generate orthogonal designs.

If we consider that the total number of levels for all product attributes, they are as follows:

- Environmental friendliness: 3 levels
- Delivery time: 3 levels
- Service level: 3 levels
- Price: 3 levels
- Quality of material: 2 levels
- Marketing proficiency: 2 levels

Hence the total number of possible product profiles will be the multiple of all the number of levels of each product attribute; 3*3*3*2*2 = 108. The choice to use 18 profiles likely reflects a thoughtful balance between statistical rigor and the need for an efficient, respondent-friendly design.

Before loading the Product Profiles CSV file for the Conjoin Analysis, we deleted the first column representing the number of product profile and adjusted the variable names. Similar cleaning was done with the Conjoint Preference CSV file.

After loading the CSV file as “design” into the R environment, we had to transform the data structure so that the relevant levels (for example "1", "2", "3") represented the relevant labels (for example “"0% CO2 reduction", "30% CO2 reduction”, "50% CO2 reduction" for the case of Environmental Friendliness) so that it is in accordance with the format required for the “caPartUtilities” function.

To calculate the part-worth for each attribute level for each customer, we established the following bases levels:

- Environmental Friendliness: 0% CO2 reduction
- Delivery Time: 14 Days
- Service Level: 5-year warranty
- Price: 1000 GBP
- Quality of Material: Market average
- Marketing Proficiency: Not very proficient and poor communication

To calculate WTP (Willingness to Pay), we first need to calculate the GBP worth of 1 part-worth. That can be done by the following formula:

$$GBP\ worth\ of\ 1\ part-worth = \frac{Difference\ of\ \max\ and\ \min\ price}{Difference\ of\ part-worth\ at\ \max\ and\ min\ price}$$

Multiplying GBP worth of 1 part-worth by individual part-worths of attributes, we can compute the WTP. In Tableau for example for WTP of 14 Days, this can be done by the following calculated field:

```
WTP of 14 Days

AVG([14 Days])*ABS((500/AVG([1500 Gbp])))
```

For part 2, To determine the optimal number of principal components, singular values will be calculated which are the square roots of the eigenvalues of the data's covariance matrix, scaled by the number of observations minus one.

To get a more robust understanding, we calculate PVE (proportion of variance explained) that tells proportion of variance explained by each principal component. That is done by summing the square of all singular values and dividing by the total variance.

The part-worths from a conjoint analysis can be used for market segmentation by grouping customers based on their preferences for different product attributes. Customers with similar part-worths for specific attributes can be segmented into a distinct group. For instance, customers who highly value "30% CO_2 reduction" or a "5-year warranty & free maintenance" might represent an environmentally conscious segment or one that prioritizes long-term service guarantees.

It may be possible that there are many segments in the data. In that case first a broad customer segmentation is performed based on survey responses to identify distinct groups within the market. Each group or segment has its own specific set of preferences and behaviors. After these segments are identified, a separate conjoint analysis is run for each one. This way, the results of the conjoint analysis are more tailored and relevant, as they account for the heterogeneity in preferences across different market segments.

If large number of segments exist, Hierarchical clustering does not require specifying the number of clusters upfront and hence might be a better options compared to K-means clustering. It builds a tree of clusters and can be particularly insightful for understanding nested relationships between different customer segments. Dimensionality reduction methods like PCA (Principal Component Analysis) can be employed as well to simplify the complexity and visualize high-dimensional clustering results more effectively.

Results

To see if the selected combinations of product profiles is optimal, we need to see the correlation between the attribute levels.

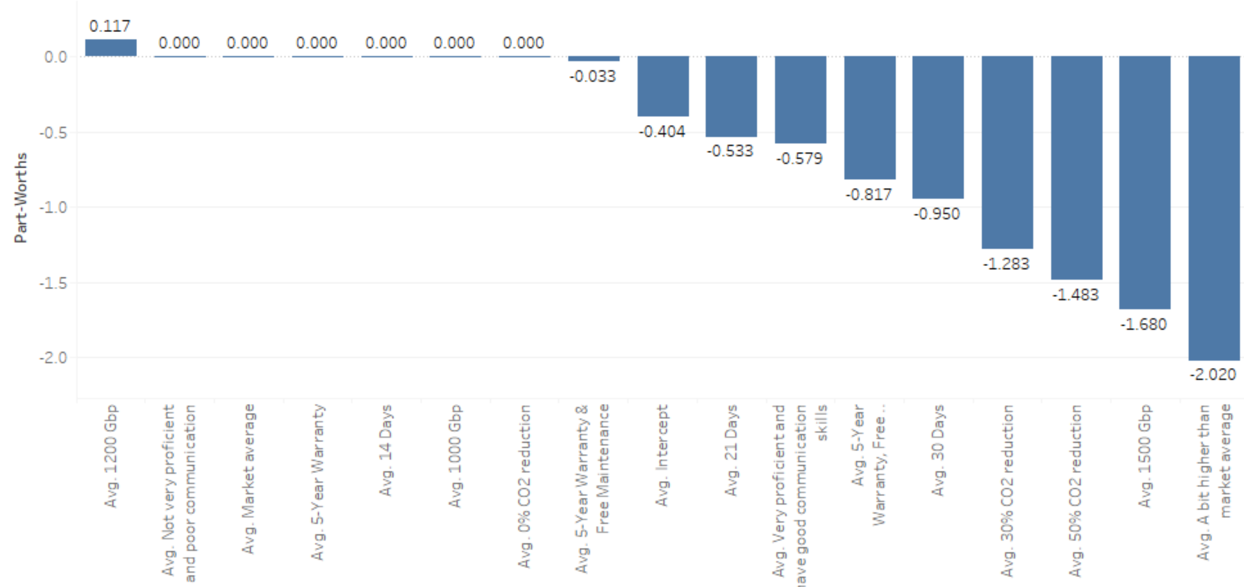
```
> print(cor(caEncodedDesign(design)))
                          Environmental_friendliness Delivery_time Service_Level Price Quality_of_material Marketing_Proficiency
Environmental_friendliness                         1             0             0     0           0.0000000             0.0000000
Delivery_time                                      0             1             0     0           0.0000000             0.0000000
Service_Level                                      0             0             1     0           0.0000000             0.0000000
Price                                              0             0             0     1           0.0000000             0.0000000
Quality_of_material                                0             0             0     0           1.0000000             0.1111111
Marketing_Proficiency                              0             0             0     0           0.1111111             1.0000000
```

The correlations between all different levels are 0 and for the same levels is 1. We need to note that the correlation between quality of material and marketing proficiency is neither 0 nor close to 0, it is approximately 11%.

In conjoint analysis, achieving perfect orthogonality can be challenging, especially as the number of attributes and levels increases. A small correlation between attributes may be acceptable depending on the context of the analysis.

After running the Conjoint Analysis, the following were the results:

Part-Worths for each Product Attribute Level



Avg. 1200 Gbp, Avg. Not very proficient and poor communication, Avg. Market average, Avg. 5-Year Warranty, Avg. 14 Days, Avg. 1000 Gbp, Avg. 0% CO2 reduction, Avg. 5-Year Warranty & Free Maintenance, Avg. Intercept, Avg. 21 Days, Avg. Very proficient and have good communication skills, Avg. 5-Year Warranty, Free Maintenance And Installation, & Upgradeability, Avg. 30 Days, Avg. 30% CO2 reduction, Avg. 50% CO2 reduction, Avg. 1500 Gbp and Avg. A bit higher than market average. The marks are labeled by Avg. 1200 Gbp, Avg. Not very proficient and poor communication, Avg. Market average, Avg. 5-Year Warranty, Avg. 14 Days, Avg. 1000 Gbp, Avg. 0% CO2 reduction, Avg. 5-Year Warranty & Free Maintenance, Avg. Intercept, Avg. 21 Days, Avg. Very proficient and have good communication skills, Avg. 5-Year Warranty, Free Maintenance And Installation, & Upgradeability, Avg. 30 Days, Avg. 30% CO2 reduction, Avg. 50% CO2 reduction, Avg. 1500 Gbp and Avg. A bit higher than market average.

There is some specifically interesting behavior represented in the visualization:

1200 GBP: It is interesting to note that the basis-level is 1000 GBP, hence getting a positive part-worth for 1200 GBP indicates that customers find a higher price of 1200 GBP better. We can say that this is awkward behavior as generally customers would want a lower price for any given product.

5-year warranty and free maintenance: Customers do not value this any more than the base level (5-year warranty) even though it is an extra service provided.

Intercept: A negative intercept indicates that the baseline combination of attributes is less preferred compared to the overall average. However, as all part-worths are negative (except for 1200 GBP Price), this suggests that the respondents generally have a low preference for the levels of attributes tested in the study, including the baseline.

Very proficient and good communication: This level for Marketing proficiency also has a negative part-worth than base level (which is not very proficient and poor communication). Hence for some reason customers prefer poor marketing communication over good communications, this is another interesting insight.

5-Year warrant, free maintenance & upgradeability: This level has the lowest part-worths out of all the levels in its product attribute. This is very interesting because this level offers the most flexibility and still is least preferred by customers.

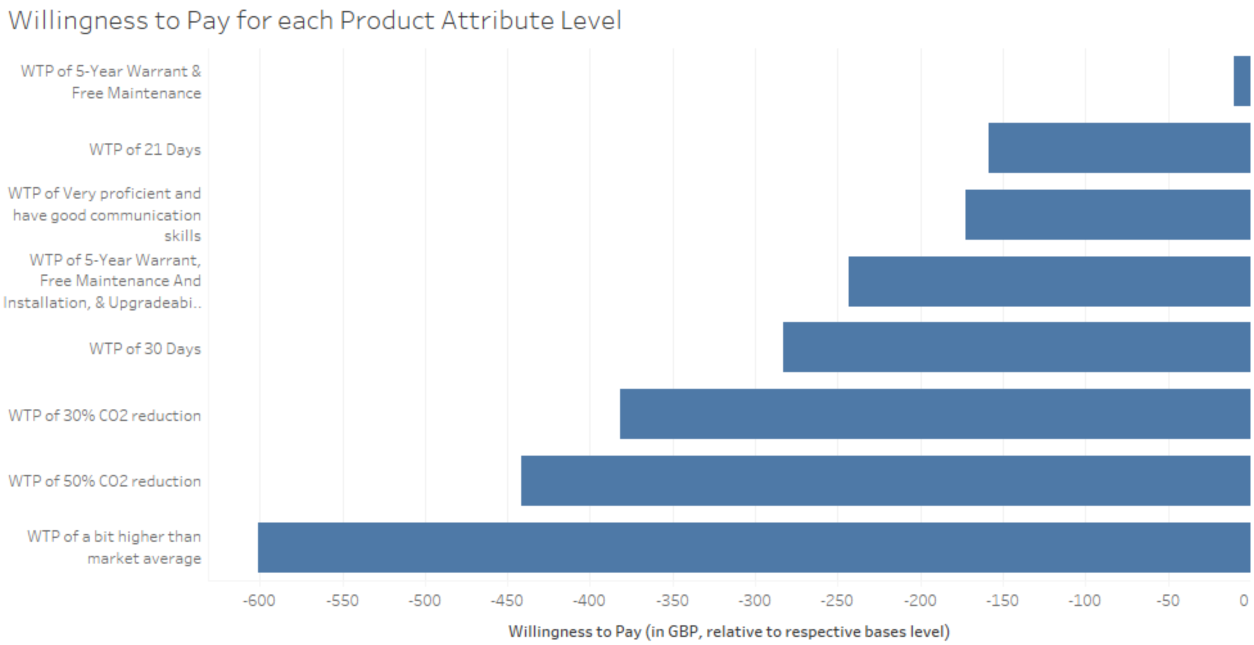
A bit higher than market average: This is the product level for the Quality of material attribute. Since the bases level was “Market average”, customer prefer the quality to be market average rather than higher than market average. This is also very counter intuitive as generally customers would prefer a higher quality.

The GBP wroth of 1 part-worth was as follows:

$$GBP\ worth\ of\ 1\ part-worth = \frac{500\ GBP}{1.680}$$

$$GBP\ worth\ of\ 1\ part-worth = 297.620\frac{GBP}{part-worth}$$

The following is a graphical representation of WTP:



WTP of 21 Days, WTP of 30 Days, WTP of 30% CO2 reduction, WTP of 50% CO2 reduction, WTP of 5-Year Warrant & Free Maintenance, WTP of 5-Year Warrant, Free Maintenance And Installation, & Upgradeability, WTP of a bit higher than market average and WTP of Very proficient and have good communication skills.

Since the WTP for all levels (apart from bases) is negative, this implies that the product levels used as base-levels are the optimal product profile.

For part 2, the following are the results of singular values of principal components:

```
> print(singular_values)
 [1] 14.3129455  9.0644638  4.4094348  2.8909377  2.6319165  2.5611481  2.0477113  1.9519079  1.5454599  1.2700782  0.8266659
```

We see that after the drop of singular values after the third value gets smaller and smaller. Hence we could choose either 2 or 3 principal components. However since our objective is to construct a perceptual map, visualization of 3 principal components might become too complex.

The following are the results of PVE (Proportion of variance explained):

```
> print(pve)
 [1] 0.600763659 0.240951627 0.057017934 0.024508858 0.020313737 0.019236011 0.012296544 0.011172858 0.007004241 0.004730495 0.002004037
```

Since the first to values add up to greater than 0.8, this implies that the first 2 principal components can explain more than 80% of the variation in the data. This makes 2 principal components sufficient to create a perceptual map.

The following are the loading factors of each attribute for each principal component:

```
> print(pca_factors)
    Attribute    factor1     factor2  path
       <char>      <num>       <num> <num>
 1:  Feature1 -0.9319502  0.02625094     1
 2:  Feature2  0.9612188  0.07121589     1
 3:  Feature3  0.9464866 -0.08030095     1
 4:  Feature4  0.8484710  0.40502680     1
 5:  Feature5 -0.7561693  0.44720905     1
 6:  Feature6  0.8897212 -0.23286996     1
 7:  Feature7 -0.5153093 -0.75438614     1
 8:  Feature8 -0.7879428 -0.37712727     1
 9:  Feature9 -0.6039632  0.69910300     1
10: Feature10 -0.5319156  0.75271549     1
11: Feature11  0.5501711  0.67330434     1
```

Since factor 1 or PCA1 describes the largest chunk of the variance in data (60%, where PCA2 describes only 24%), we can focus on the loadings with the highest magnitude in PCA1 to identify what features are most important to customers. Hence features greater than 0.8 in magnitude are:

- Feature2: 0.96
- Feature3: 0.95
- Feature1: 0.93
- Feature6: 0.89
- Feature4: 0.85

The features that have the highest scores in both PC models indicates that the feature is very correlated (+ or -) with features highly loading on that component.

Features with highest values in PC 1:

- Feature2: +0.96
- Feature3: +0.95
- Feature1: -0.93
- Feature6: +0.89
- Feature4: +0.85
- Feature8: -0.79
- Feature5: -0.76

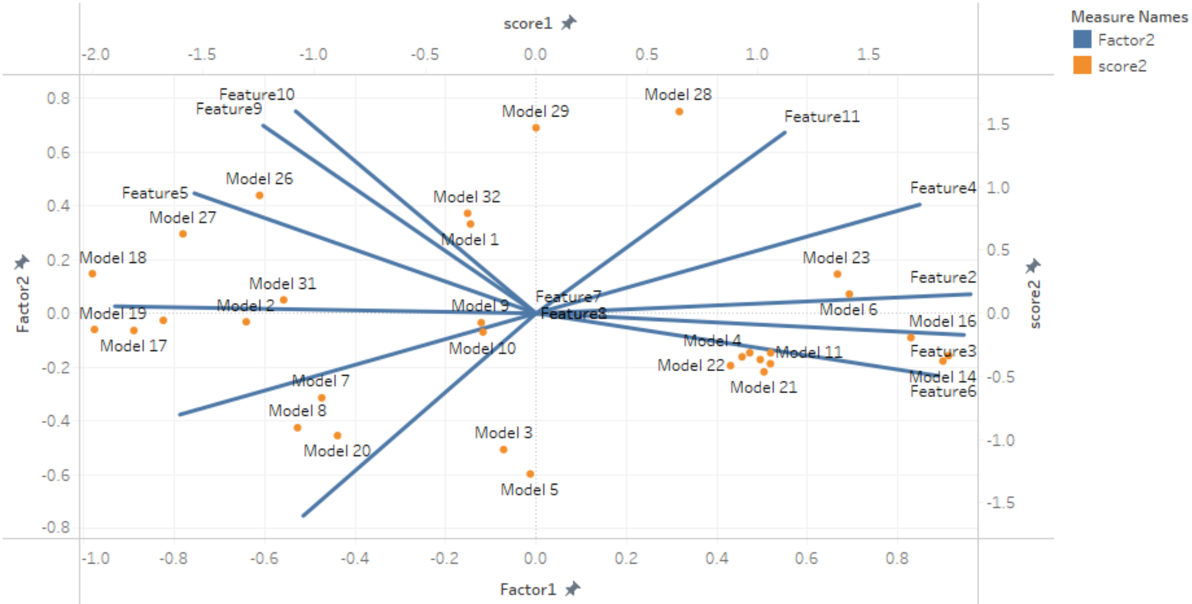
Features with highest values in PC 2:

- Feature7: -0.75
- Feature10: +0.75
- Feature9: +0.70

More features are correlated in PC1 and have generally higher scores than in PC2. This also confirms that PC1 features are more important to customers.

After calculating PCA scores for each model, the following is the perceptual map.

Perceptual Map



Sum of Factor1 and sum of score1 vs. Factor2 and score2. Color shows details about Factor2 and score2. The marks are labeled by Attribute and Model. Details are shown for Path.

Since it is a little difficult to see some labels:

- Feature1 is the line almost parallel to the x-axis towards the left.
- Feature8 is the first line below Feature1.
- Feature7 is the first line below Feature8.

Some important insights from the perceptual map:

- Models like 4,22,21,11 are almost similar as they are cluttered together. There are many groups of models that consumers perceive as similar. Also some are pairs like Model 6 and 23 etc.
- The models that are the most unique to customers are 26,27,28,29 as they are not cluttered to closely to any other models.
- Features that are not related to each other are 10 and 11, 9 and 11, 7 and 10, 7 and 9 (they form an angle close to 90 degrees with each other).
- Features that are almost identical are 9 and 10, 2 and 3 (to some extent), 3 and 6 (to some extent) as they are closest.
- Features that are on the opposite sides of the y-axis (left and right) are negatively related to each other if they form an angle greater than 90 degrees.
- Features that form an angle less than 90 degrees and are not very close are positively related.

Conclusion

For this assignment, the product profiles that were created did not result in perfect non-collinearity. Even though at one instance the correlation between two product attribute was as low as 11%, it was not close to 0. This leaves slight room for issues of multicollinearity where preference of one product attribute could erroneously show preference of another product attribute.

The preferences of customers seen in terms of part-worths were a little awkward as customers at times associated more preference to product levels with lesser objective utility (For example customers preferred 5-year warranty much more compared to 5-year warranty + installation and even 5-year warranty + installation + upgradeability). This could indicate to a unique kind of customer with an unusual customer behavior.

Usually as Price levels increase, the part-worths for them decrease because intuitively customers would want a product at the lowest price. However in this case, customers preferred a medium price over the lowest price. This could mean that the product profiles made with a medium price were objectively more attractive not because of the price but because of the other product attribute levels present in them.

An intercept close to zero or negative and all other part-worths negative implied that customers did not like the product profile with base levels of product attributes and even more strongly disliked other product attribute levels. However these part-worths were calculated on average, meaning there were customers that liked a certain product level much more than others which resulted in a higher variance of individual part-worths of customers. With such high variance of individual part-worths, this implies that either the opinions of customers are objectively very varied throughout the whole customer base in the market that it operates, or the customers sampled for the preference survey are not an actual scaled representation of the market. To address this, another study can take the approach of better sampling methods.

References

Asioli, D., Næs, T., Granli, B.S. et al., 2014. Consumer preferences for iced coffee determined by conjoint analysis: An exploratory study with Norwegian consumers. International Journal of Food Science & Technology. Available at: https://ifst.onlinelibrary.wiley.com/doi/abs/10.1111/ijfs.12485 [Accessed 26 April 2024].

Almli, V.L. and Næs, T., 2018. Conjoint analysis in sensory and consumer science: Principles, applications, and future perspectives. Methods in Consumer Research, Volume 1. Elsevier. Available at: https://www.sciencedirect.com/science/article/pii/B9780081020890000194 [Accessed 26 April 2024].

Oey, E., Soputro, C.S., Wibowo, S.Y. et al., 2022. Integration of Kansei engineering and conjoint analysis to product design-a case study of a synthetic rattan sofa manufacturer. International Journal of Business Excellence. Available at: https://www.inderscienceonline.com/doi/abs/10.1504/IJBEX.2022.121581 [Accessed 26 April 2024].

Prasad, K.D. and Subbaiah, K.V., 2011. Prioritization of customer needs in house of quality using conjoint analysis. Center for Quality. Available at: http://www.cqm.rs/2011/cd/5iqc/pdf/087.pdf [Accessed 26 April 2024].

Tseng, C.C., Lin, J.Y., Chen, M.S., Lin, M.C., 2011. The use of conjoint analysis in the design of bicycle-related clothing. Journal of the Textile Institute. Available at: https://www.tandfonline.com/doi/abs/10.1080/10170669.2011.579477 [Accessed 26 April 2024].

Appendix

Code for Conjoint Analysis

```
#####################

# Conjoint Analysis #

#####################



## Load Packages and Set Seed

library(conjoint)

set.seed(1)


## Set up attributes and levels as a list

attrib.level <- list(Environmental_friendliness = c("0% CO2 reduction", "30% CO2 reduction", "50% CO2 reduction"),

                     Delivery_time = c("14 Days", "21 Days", "30 Days"),

                     Service_Level = c("5-year warranty", "5-year warranty & free maintenance", "5-year warranty, free maintenance and installation, & upgradeability"),

                     Price = c("1000 GBP", "1200 GBP", "1500 GBP"),

                     Quality_of_material = c("Market average", "A bit higher than market average"),

                     Marketing_Proficiency = c("Not very proficient and poor communication", "Very proficient and have good communication skills"))



#Load Product Profiles, Deleted the first column, added _ between spaces of variable names

design <- read.csv(file.choose())


str(design)

design[] <- lapply(design, factor)

str(design)
```

```
design <- within(design, {

  Environmental_friendliness <- factor(Environmental_friendliness, levels = c("1", "2", "3"), labels = c("0% CO2 reduction", "30% CO2 reduction", "50% CO2 reduction"))

  Delivery_time <- factor(Delivery_time, levels = c("1", "2", "3"), labels = c("14 Days", "21 Days", "30 Days"))

  Service_Level <- factor(Service_Level, levels = c("1", "2", "3"), labels = c("5-year warranty", "5-year warranty & free maintenance", "5-year warranty, free maintenance and installation, & upgradeability"))

  Price <- factor(Price, levels = c("1", "2", "3"), labels = c("1000 GBP", "1200 GBP", "1500 GBP"))

  Quality_of_material <- factor(Quality_of_material, levels = c("1", "2"), labels = c("Market average", "A bit higher than market average"))

  Marketing_Proficiency <- factor(Marketing_Proficiency, levels = c("1", "2"), labels = c("Not very proficient and poor communication", "Very proficient and have good communication skills"))

})


# Check the results

str(design)



# Manually adding attributes

attr(design, "out.attrs") <- list(

  dim = c(3, 3, 3, 3, 2, 2),  # Number of levels in each factor, ensure these numbers match your data

  dimnames = list(

    Environmental_friendliness = c("Environmental_friendliness=0% CO2 reduction", "Environmental_friendliness=30% CO2 reduction", "Environmental_friendliness=50% CO2 reduction"),

    Delivery_time = c("Delivery_time=14 Days", "Delivery_time=21 Days", "Delivery_time=30 Days"),

    Service_Level = c("Service_Level=5-year warranty", "Service_Level=5-year warranty & free maintenance", "Service_Level=5-year warranty, free maintenance and installation, & upgradeability"),

    Price = c("Price=1000 GBP", "Price=1200 GBP", "Price=1500 GBP"),

    Quality_of_material = c("Quality_of_material=Market average", "Quality_of_material=A bit higher than market average"),

    Marketing_Proficiency = c("Marketing_Proficiency=Not very proficient and poor communication", "Marketing_Proficiency=Very proficient and have good communication skills")
```

```
  )
)

# Check the structure to see if the attributes are added
str(design)

## Check for correlation in design
print(cor(caEncodedDesign(design)))

## Run the conjoint analysis study

## Read in the survey preference results, delete the first column, save as csv.
pref <- read.csv(file.choose()) ## Choose the file named Conjoint_Preference_Cleaned.csv
str(pref)

## Set up attributes and levels as a vector and Estimate the part-worths for each respondent

attrib.vector <- data.frame(unlist(attrib.level,use.names=FALSE))



temp <- caPartUtilities(pref, design, attrib.vector)
str(temp)
summary(temp)

colnames(attrib.vector) <- c("levels")
part.worths <- NULL

for (i in 1:ncol(pref)) {
  temp <- caPartUtilities(pref[,i], design, attrib.vector)
```

```
## Base Case: Environmental.friendliness 0%, Delivery.time 14 days, Service.Level 5 year warranty, price 1000GBP, Quality.of.material average, Marketing.Proficiency poor

base_Environmental_friendliness <- temp[,"0% CO2 reduction"]

base_Delivery_time <- temp[,"14 Days"]

base_Service_Level <- temp[,"5-year warranty"]

base_Price <- temp[,"1000 GBP"]

base_Quality_of_material <- temp[,"Market average"]

base_Marketing_Proficiency <- temp[,"Not very proficient and poor communication"]


## Adjust Intercept

temp[,"intercept"] <- temp[,"intercept"] - base_Environmental_friendliness - base_Delivery_time - base_Service_Level -

  base_Price - base_Quality_of_material - base_Marketing_Proficiency


## Adjust Coefficients for each attribute

## Environmental_friendliness

L1 <- length(attrib.level$Environmental_friendliness)

for (j in 1:L1) { temp[,j] <- temp[,j] - base_Environmental_friendliness }

## Delivery_time

L2 <- length(attrib.level$Delivery_time) + L1

for (k in (L1+1):L2) { temp[,k] <- temp[,k] - base_Delivery_time }

## Service_Level

L3 <- length(attrib.level$Service_Level) + L2

for (l in (L2+1):L3) { temp[,l] <- temp[,l] - base_Service_Level }

## Price

L4 <- length(attrib.level$Price) + L3

for (m in (L3+1):L4) { temp[,m] <- temp[,m] - base_Price }

## Quality_of_material
```

```
  L5 <- length(attrib.level$Quality_of_material) + L4
  for (n in (L4+1):L5) { temp[,n] <- temp[,n] - base_Quality_of_material }
  ## Marketing_Proficiency
  L6 <- length(attrib.level$Marketing_Proficiency) + L5
  for (n in (L5+1):L6) { temp[,n] <- temp[,n] - base_Marketing_Proficiency }

  part.worths <- rbind(part.worths, temp)
}

rownames(part.worths) <- colnames(pref)
summary(part.worths)
print(part.worths)

## Export part-worths from analysis
write.csv(part.worths, "Conjoint_Partworths_Assignment.csv", row.names = FALSE) ## Name the file Conjoint_Partworths_Assignment.csv
```

Code for PCA

```
########################################
## Perceptual and Preference Mapping #
########################################

## Load Packages and Set Seed
library(data.table)
set.seed(1)

## Read in perceptions data
per <- read.csv(file.choose()) #Load PCA data
str(per)
summary(per)

## Run Princple Components Analysis on Perceptions

pca <- prcomp(per[,2:length(per)], retx=TRUE, scale=TRUE)
print(pca)

## Perceptual Map Data - Attribute Factors and CSV File

attribute <- as.data.table(colnames(per[,2:length(per)])); setnames(attribute, 1, "Attribute")
print(attribute)

# Calculate the singular values
singular_values <- pca$sdev * sqrt(nrow(per) - 1)
print(singular_values)
#Observe where the singular values drop significantly, that is the number of PCs that you should have.
```

```
total_variance = sum(singular_values^2)

pve = (singular_values^2) / total_variance

print(pve)

#Proportion of variance explained

#1st PC exlpains 60% and 2nd PC explains 24% of the data, which is +80%, Hence 2 PCs are best.


factor1 <- pca$rotation[,1]*pca$sdev[1]; factor2 <- pca$rotation[,2]*pca$sdev[2]; path <- rep(1, nrow(attribute))

print(factor1) #Loading Factors of PCA1, can tell which feature is most important

print(factor2)


pca_factors <- subset(cbind(attribute, factor1, factor2, path), select = c(Attribute, factor1, factor2, path))

print(pca_factors)


pca_origin <- cbind(attribute, factor1 = rep(0,nrow(attribute)), factor2 = rep(0,nrow(attribute)), path = rep(0,nrow(attribute)))

print(pca_origin)


pca_attributes <- rbind(pca_factors, pca_origin)

print(pca_attributes)


write.csv(pca_attributes, "perceptions_attributes_assignment2.csv", row.names = FALSE) ## Name file perceptions_attributes_assignment2.csv


## Perceptual Map Data - Brand Factors and CSV File


score1 <- (pca$x[,1]/apply(abs(pca$x),2,max)[1])

print(score1)


score2 <- (pca$x[,2]/apply(abs(pca$x),2,max)[2])
```

```
print(score2)

pca_scores <- subset(cbind(per, score1, score2), select = c(Model, score1, score2))
print(pca_scores)

#Observe what models have the highest scores in both PC models or score1 and score2. High scores in that specific PC indicates that that specific model
#is highly correlated with the attributes heavily loading on this component. + or - sign will indicate if this relation is positive or negative.

write.csv(pca_scores, "perceptions_scores_assignment2.csv", row.names = FALSE) ## Name file perceptions_scores_assignment2.csv
```