Project 2 = Summary Task
By – Hemant Kshirsagar
Market Segmentation Analysis

Strategic marketing involves developing long-term plans and initiatives that align with the overall business strategy and objectives. It focuses on creating a competitive advantage and building a strong brand reputation that resonates with the target audience.

Tactical marketing involves the implementation of specific marketing tactics and initiatives to achieve short-term goals and objectives. It focuses on executing the strategies outlined in the strategic marketing plan and adapting to changes in the market or competitive landscape.

Both strategic and tactical marketing are necessary for a company to achieve its marketing goals. Strategic marketing provides the big-picture thinking and direction, while tactical marketing puts that thinking into action.

Tactical marketing planning usually covers a period of up to one year. It is traditionally seen to cover four areas: the development and modification of the product in view of needs and desires of the target segment (Product), the determination of the price in view of cost, competition, and the willingness to pay of the target segment (Price), the selection of the most suitable distribution channels to reach the target segment (Place), and the communication and promotion of the offer in a way that is most appealing to the target segment (Promotion).

SWOT analysis can be used to identify the strengths, weaknesses, opportunities, and threats that impact the business as a whole. This analysis can inform the development of long-term plans and initiatives that align with the overall business strategy.

Strengths and weaknesses refer to internal factors such as the company's resources, capabilities, and competitive position. Opportunities and threats refer to external factors such as market trends, consumer behavior, and competition.

The output of a SWOT analysis in strategic marketing can be used to identify key areas of focus for the marketing plan. For example, if the analysis identifies a weakness in the company's digital marketing capabilities, the marketing plan could include initiatives to improve this area.

The segmentation criterion can be one single consumer characteristic, such as age, gender, country of origin, or stage in the family life cycle. Alternatively, it can contain a larger set of consumer characteristics, such as a number of benefits sought when purchasing a product, a number of activities undertaken when on vacation, values held with respect to the environment, or an expenditure pattern.

a concentrated market strategy can be a highly effective way for a company to build a strong brand and differentiate itself from competitors, but it requires careful planning and execution to be successful.

Some benefits of market segmentation:

- Targeted marketing
- Improved customer satisfaction
- Increased profitability
- Competitive advantage
- Resource optimization
- While market segmentation offers many benefits, it is important to note that there are also costs associated with implementing a market segmentation strategy.

Here are some potential costs of market segmentation:

- Research costs
- Implementation costs
- Most important Inaccurate targeting

Chapter 2
Market Segmentation Analysis

The Layers of Market Segmentation Analysis

There are three layers

1. Conducting high quality market segmentation analysis:
   Extracting market segments

2. Enabling high quality market segmentation analysis:
   Collecting good data, exploring data, profiling segments, describing segments

3. Implementation
   Deciding to segment, defining the ideal segment, selecting (the) target segment(s), developing a customized marketing mix, assessing effectiveness and monitoring marketing changes

2.2 Approaches to Market Segmentation Analysis

Two possible systematics are presented, with one based on the willingness of the organization conducting the analysis to make changes in their targeting approach, while the other is based on the segmentation variables used in the analysis.

If market segmentation analysis reveals a promising niche segment, or a promising set of market segments to target with a differentiated market strategy, the organization must develop an entirely new marketing plan in view of those findings.

Very important to note the term a priori segmentation indicates that the decision about what characterizes each segment is made in advance, before any data analysis is conducted

Commonsense segmentation is a basic approach to market segmentation that involves grouping customers based on readily observable characteristics, such as age, gender, income, and location. This method is relatively simple and straightforward, but it may not be the most effective approach for identifying and targeting specific consumer needs or behaviors.

While conducting data-driven market segmentation, data analysts and users of market segmentation solutions often assume that market segments naturally exist in the data. Such naturally occurring segments, it is assumed, need to merely be revealed and described. In real consumer data, naturally existing, distinct and well separated market segments rarely exist.

Market Segmentation Analysis Step-by-Step

There is ten-step approach to market segmentation analysis.

STEP 1 - Deciding what not to segment
Is the market suitable?
Can you make a long-term commitment?

STEP 2 - Specifying the ideal target segment
What would your ideal target segment look like?

STEP 3 - Collecting data
Collect data
(Segmentation and descriptor variables)

STEP 4 - Exploring data
Explore data, pre-process if required.

STEP 5 - Extracting segments
Split consumers into segments
using the segmentation variable.

STEP 6 - Profiling segments
Determine key features
of the extracted market segments.

STEP 7 - Describing segments
Describe segments in detail.

STEP 8 - Selecting (the) target segment(s)
Evaluate segments and select target segment(s).

STEP 9 - Customizing the marketing mix
Develop a customized marketing mix.

STEP 10 - Evaluation and monitoring
Evaluate success, monitor changes.

3.1 Implications of Committing to Market Segmentation

committing to a market segmentation strategy requires a long-term commitment and the willingness and ability to make significant changes and investments. It involves costs associated with research, surveys, packaging, advertisements, and communication messages. It is recommended to only pursue segmentation if the expected increase in sales justifies the cost.

3.2 Implementation Barriers

The first group of barriers relates to a lack of resources, including budget and data quality. The second group of barriers relates to organizational culture, such as a lack of market or consumer orientation, resistance to change, and a lack of training among senior management and the team tasked with segmentation. Croft (1994) developed a short questionnaire to assess the extent of these cultural barriers. Overall, understanding and commitment from senior management is crucial for successful implementation of market segmentation.

3.3 Step 1 Checklist

Task	Who is responsible?	Completed?
Ask if the organisation’s culture is market-oriented. If yes, proceed. If		
no, seriously consider not to proceed.		
Ask if the organisation is genuinely willing to change. If yes, proceed.		
If no, seriously consider not to proceed.		
Ask if the organisation takes a long-term perspective. If yes, proceed.		
If no, seriously consider not to proceed.		
Ask if the organisation is open to new ideas. If yes, proceed. If no,		
seriously consider not to proceed.		
Ask if communication across organisational units is good. If yes,		
proceed. If no, seriously consider not to proceed.		
Ask if the organisation is in the position to make significant		
(structural) changes. If yes, proceed. If no, seriously consider not to		
proceed.		
Ask if the organisation has sufficient financial resources to support a		
market segmentation strategy. If yes, proceed. If no, seriously		
consider not to proceed.		
Secure visible commitment to market segmentation from senior		
management.		
Secure active involvement of senior management in the market		
segmentation analysis.		
Secure required financial commitment from senior management.		
Ensure that the market segmentation concept is fully understood. If it		
is not: conduct training until the market segmentation concept is fully		
understood.		
Ensure that the implications of pursuing a market segmentation		
strategy are fully understood. If they are not: conduct training until the		
implications of pursuing a market segmentation strategy are fully		
understood.		
Put together a team of 2-3 people (segmentation team) to conduct		
the market segmentation analysis.		

Chapter 9

Step 7: Describing Segments

9.1 Developing a Complete Picture of Market Segments

Segment profiling refers to understanding the differences in segmentation variables among market segments. Segmentation variables are chosen early in the market segmentation analysis process, and they are used to extract market segments from the collected data.

This analogy suggests that market segmentation is a long-term commitment, similar to a marriage, where the organization must be willing to make changes and investments. Segment profiling, on the other hand, is about understanding the differences in segmentation variables across market segments, much like going on multiple dates to get to know a potential spouse before committing to a long-term relationship. By profiling and describing market segments, organizations can get a better understanding of their target customers and avoid unexpected issues in the future.

9.2 Using Visualization's to Describe Market Segments

Visualizing differences in descriptor variables using graphical statistics has two key advantages: it simplifies the interpretation of results for both data analysts and users, and it integrates information on the statistical significance of differences, thereby avoiding the over-interpretation of insignificant differences.

This approach is suitable for nominal, ordinal, and metric descriptor variables, and is preferred by marketing managers due to its intuitive nature. Graphical representations are considered to transmit the essence of marketing research results and are processed more efficiently than tabular results.

9.2.1 Nominal and Ordinal Descriptor Variables

When describing differences between market segments using one nominal or ordinal descriptor variable, the first step is to create a cross-tabulation of segment membership with the descriptor variable. This is done by adding segment membership as a categorical variable to the data frame of descriptor variables and using the formula interface of R for testing or plotting.

The advantage of using graphical representations is that it simplifies interpretation and integrates information on the statistical significance of differences. This is preferred by marketing managers as it is more efficient and intuitive than using tabular results.

Very critical on how to use a cross-tabulation to describe differences between market segments in one single nominal or ordinal descriptor variable. In the example using the Australian travel motives data set, we have seen stacked bar chart to visualize the cross-tabulation of segment membership with the gender descriptor variable. The chart shows the segment sizes on the y-axis, with the number of men and women in each segment displayed within each bar. However, comparing proportions of men and women across segments can be complicated if the segment sizes are unequal.

Drawing the bars for men and women next to one another to solve this problem, but this approach can obscure the absolute sizes of the market segments. Another solution is to use a mosaic plot, which shows the sizes of the segments on both the x and y-axes and allows for easy comparison of proportions.

Mosaic plots are useful not only for visualizing tables with two descriptor variables but also for tables with more than two variables. Mosaic plots can also highlight differences between observed and expected frequencies based on the assumption of independence between variables. The color of the cell represents the standardized difference between observed and expected frequencies. Negative differences are colored in red, indicating lower than expected frequencies, while positive differences are colored in blue, indicating higher than expected frequencies. The intensity of the color indicates the absolute value of the standardized difference.

Standardized differences follow a standard normal distribution, and standard normal random variables lie within [-2, 2] with a probability of approximately 95%, and within [-4, 4] with a probability of approximately 99.99%. These standardized differences are equivalent to the standardized Pearson residuals from a log-linear model assuming independence between the variables. This integration of inferential statistics within mosaic plots can help with interpretation.

9.2.2 Metric Descriptor Variables

Conditional plots are well-suited for visualizing differences between market segments using metric descriptor variables. R package lattice generated the segment profile plot in Sect. 8.3.1. In the context of segment description, this R package can display the age distribution of all segments comparatively. Or visualize the distribution of the (Original metric) moral obligation scores for members of each segment.

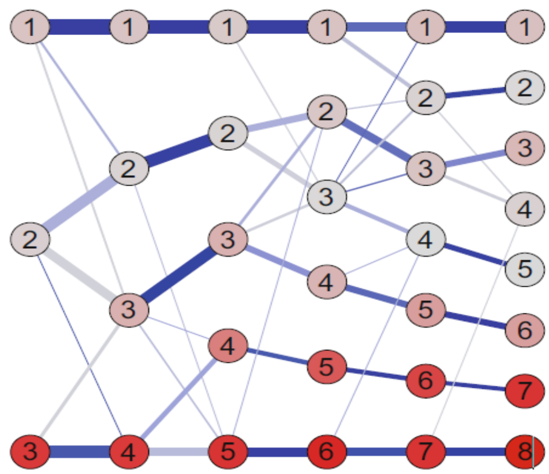
To have segment names (rather than only segment numbers) displayed in the plot, we create a new factor variable by pasting together the word "Segment" and the segment numbers from C6. We then generate a histogram for age for each segment. Argument as table controls whether the panels are included by starting on the top left (TRUE) or bottom left (FALSE, the default).

We can gain additional insights by using a parallel box-and-whisker plot; it shows the distribution of the variable separately for each segment. We create this parallel box-and-whisker plot for age by market segment in R with the following command.

In a modified version of the segment level stability across solutions (SLSA) plot, a metric descriptor variable is plotted using different colors for the nodes. The SLSA plot traces the value of the metric descriptor variable over a series of market segmentation solutions. Each point in the plot represents a segment, and the x-axis indicates the solution number. The y-axis indicates the average value of the metric descriptor variable in each segment.

To generate a modified SLSA plot in R, we can use the package GGally. We first need to compute the average value of the metric descriptor variable for each segment in each solution. We can then use the ggplot function to create the plot, mapping the x-axis to the solution number and the y-axis to the average value of the metric descriptor variable. We can add points for each segment, colored according to the value of the metric descriptor variable. To connect the points representing the same segment across solutions, we can use the geom_line function.

Fig. 9.9 Segment level stability across solutions (SLS_A) plot for the Australian travel motives data set for three to eight segments with nodes coloured by mean moral obligation values



represents the numeric SLS_A value. Light grey edges indicate low stability values. Dark blue edges indicate high stability values.

9.3 Testing for Segment Differences in Descriptor Variables

ne way to test for differences in descriptor variables across market segments is to conduct independent tests for each variable of interest, treating segment membership as a nominal variable. This allows for the testing of associations between the segment membership variable and other nominal or ordinal variables, such as gender, age, or education level. The appropriate statistical test to use depends on the nature of the variables being compared and the research question being addressed.

Segment membership can be treated like any other nominal variable. It represents a nominal summary statistic of the segmentation variables. Therefore, any test for association between a nominal variable and another variable is suitable.

Correct, a p-value is a measure of the evidence against the null hypothesis. In this case, a p-value greater than 0.05 suggests that the evidence against the null hypothesis is not strong enough to reject it. Therefore, we conclude that there is not enough statistical evidence to support the claim of significant differences in gender distribution between the Australian travel motives segments.

If the χ2-test rejects the null hypothesis of independence because the p-value is
smaller than 0.05 then mosaic plot can help to visually identify the source of the significant association between the two variables.

The plot uses the area of rectangles to represent the frequency of observations in each category combination. By examining the plot, one can identify which categories are driving the association and whether there is a pattern or trend in the association.

The parallel boxplots can be used to visually compare the distribution of a metric variable across different market segments. Simple statistical tests such as t-tests or ANOVA can then be used to formally test for significant differences in the means or medians of the variable across the segments. The p-values from these tests can be used to determine whether the observed differences are statistically significant.

The Kruskal-Walli's rank sum test is a non-parametric alternative to the analysis of variance (ANOVA) that does not assume the data is normally distributed. Instead, it ranks the observations within each group and compares the median ranks between groups.

Like the ANOVA, it tests the null hypothesis that all groups have the same median value for a given variable. If the p-value is less than 0.05, we reject the null hypothesis and conclude that at least two groups have different median values. The Kruskal-Wallis test can be performed using the "kruskal.test ()" function in R.

After rejecting the null hypothesis of the analysis of variance, we know that at least two market segments differ in their mean moral obligation. Pairwise comparisons between segments can then be conducted to identify which segments differ from each other. One common method for conducting pairwise comparisons is the Tukey HSD (Honestly Significant Difference) test, which is a post-hoc test that adjusts for multiple comparisons. Another method is the Bonferroni correction, which involves adjusting the p-value threshold for each pairwise comparison.

Bonferroni correction is a very conservative approach to correct p-values for multiple testing because it involves multiplying all p-values by the number of tests computed. This can lead to a higher chance of false negatives, which means failing to reject a true alternative hypothesis.

Other methods for correcting p-values, such as the Holm-Bonferroni method used in the example you provided, can be less conservative and more powerful. These methods adjust the p-values based on the number of tests performed, but also take into account the order in which the tests are conducted. This can improve the ability to detect true differences while controlling for the overall false positive rate.

9.4 Predicting Segments from Descriptor Variables

Using regression models for predicting segment membership based on descriptor variables is a powerful approach in market segmentation analysis. By using the segment membership as a categorical dependent variable and the descriptor variables as independent variables, we can identify which descriptor variables are critical in distinguishing between different market segments.

Overall, using regression models for predicting segment membership provides a more comprehensive understanding of the relationship between descriptor variables and market segments. This approach can also be used to validate previously identified segments and to uncover new, previously unknown segments.

Regression analysis is the basis of prediction models. Regression analysis
assumes that a dependent variable y can be predicted using independent variables
The basic regression model is the linear regression model.

The linear regression model assumes that function f (·) is linear, and that y follows a normal distribution with mean f ($x_1, \ldots, x_p$) and variance σ^2. The relationship between the dependent variable y and the independent variables $x_1, \ldots, x_p$ is given by:
$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + E$

where β_0 is the intercept or constant term, $\beta_1, \ldots, \beta_p$ are the coefficients or slopes of the independent variables, and The goal of the linear regression analysis is to estimate the coefficients $\beta_0, \beta_1, \ldots, \beta_p$ that provide the best linear approximation to the relationship between y and $x_1, \ldots, x_p$, in the sense that the sum of the squared.

The estimated coefficients are obtained by minimizing the sum of squared errors (SSE) between the observed values of y and the predicted values of y based on the independent variables $x_1, \ldots, x_p$. The SSE is given by:
$SSE = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2$,

Where n is the number of observations, y_i is the observed value of the dependent variable for observation i, $\bar{y}_i$ is the predicted value of the dependent variable for observation i, and $\bar{y}_i$ is given by:
$\bar{y}_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip}$.

The estimated coefficients can be obtained using various methods, such as ordinary least squares (OLS), maximum likelihood estimation (MLE), or Bayesian estimation. Once the coefficients are estimated, we can use the regression equation to predict the value of the dependent variable for new values of the independent variables. The quality of the prediction can be assessed using measures such as the R-squared or the root mean squared error (RMSE).

The link function used is the logit function, which maps the probability of the dependent variable being a certain value to the entire real number line. The logistic regression model then estimates the probability of the dependent variable being a certain value given the values of the independent variables.

In multinomial logistic regression, the dependent variable can take on three or more discrete values (e.g., red, green, or blue). The link function used is also the logit function, but it is applied to each possible outcome of the dependent variable relative to one of the possible outcomes (known as the reference category). The multinomial logistic regression model then estimates the probabilities of each outcome relative to the reference category given the values of the independent variables.

9.4.1 Binary Logistic Regression

The logit link function is commonly used in binary logistic regression models. The logit function is defined as the natural logarithm of the odds ratio:

$\text{logit}(\mu) = \ln(\mu / (1-\mu))$

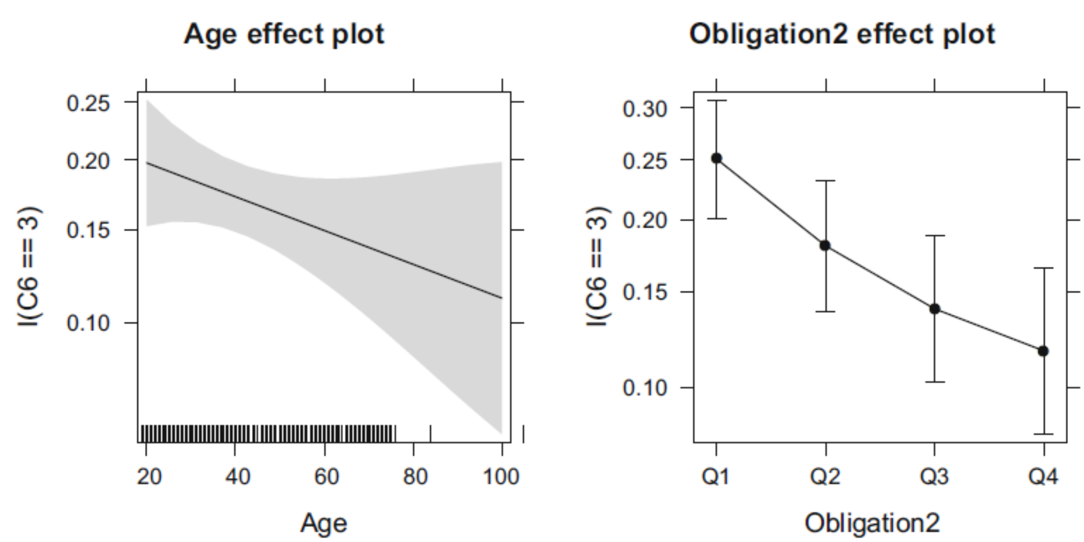
where μ is the probability of success (i.e., 1) for a given binary outcome variable y. The odds ratio can be interpreted as the ratio of the probability of success to the probability of failure (i.e., 0). The logit function maps the probability of success onto a continuous scale from -infinity to infinity, allowing for linear regression analysis to be performed. In binary logistic regression, the goal is to estimate the probability of the binary outcome variable being a success (y=1) or a failure (y=0) as a function of one or more predictor variables.

The intercept in the linear regression model gives the mean value of the dependent variable if the independent variables x1, . . . , xp all have a value of 0. In binomial logistic regression, the intercept gives the value of the linear predictor η if the independent variables x1, . . . , xp all have a value of 0

To calculate the probability of being in segment 3 for a respondent with age 0 and a low moral obligation value, we first need to calculate the value of the linear predictor η using the intercept and coefficients from the logistic regression model. Let's assume the intercept is denoted as β0 and the coefficients for age and moral obligation are β1 and β2, respectively. Then, we have:

η = β0 + β1 * 0 + β2 * low moral obligation value

The regression coefficients in binary logistic regression indicate the change in log odds of success associated with a one-unit increase in the corresponding independent variable while holding other independent variables constant. This means that the coefficient tells us how much the odds of success change for a unit increase in the independent variable, regardless of the level of the other independent variables.



(Figure 9.11) that shows the predicted probability of being in segment 3 based on two independent variables: age and moral obligation. The left plot shows the predicted probability for different ages, while the right plot shows the predicted probability for different levels of moral obligation. The predicted probabilities are shown with 95% confidence intervals or bands, indicating the uncertainty of the estimates.

The output of the Anova function provides a table with the deviance, the degrees of freedom, and the associated p-value for the comparison of the two models. If the p-value is smaller than the chosen significance level (usually 0.05), we reject the null hypothesis that the reduced model (without moral obligation) fits the data equally well as the full model (with moral obligation). In this case, we conclude that including moral obligation significantly improves model fit.

The output shows – for each independent variable in the model – the test statistic, the degrees of freedom of the distribution to calculate the p-value and the p-value.

In R, you can include multiple independent variables in a binary logistic regression model by adding them to the formula on the right-hand side of the ~ symbol, separated by +. For example, if you have a dataset my data with variables age, gender, and income, and you want to fit a binary logistic regression model to predict a binary outcome y.

Including too many independent variables in a model can lead to overfitting, where the model fits the data too closely and may not perform well on new, unseen data. Model selection methods, such as stepwise regression, can help to prevent overfitting by identifying the most important independent variables to include in the model based on their contribution to model fit as measured by the AIC or other criteria.

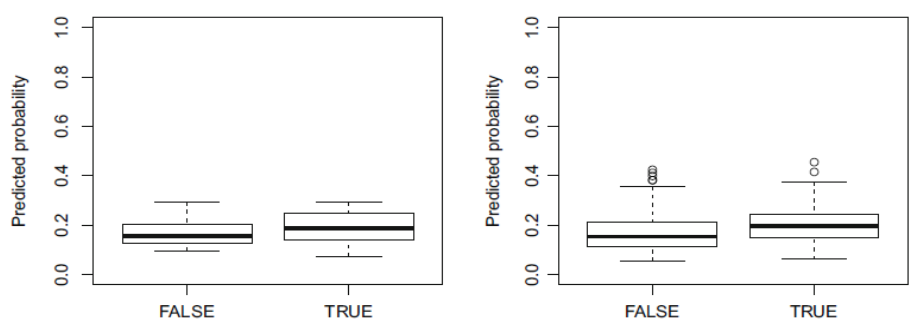


Figure 9.12 is a boxplot comparing the predicted probabilities of segment 3 membership for the two logistic regression models. The y-axis shows the predicted probabilities ranging from 0 to 1. The x-axis shows the true membership in segment 3. The boxes represent the distribution of predicted probabilities for each group. The group TRUE represents respondents who are members of segment 3, and the group FALSE represents all other respondents.

We can see from Fig. 9.12 that the performance of the two fitted models is nowhere close to this optimal case. The median predicted values are only slightly higher for segment 3 in both models. The difference is larger for the model fitted using step, indicating that the predictive performance of this model is slightly better.

9.4.2 Multinomial Logistic Regression

In multinomial logistic regression, the dependent variable y has more than two categories, and the goal is to predict the probability of each category. The logistic function is used as the link function to model the relationship between the independent variables and the probabilities of the different categories. The model estimates a separate set of coefficients for each independent variable for each category of the dependent variable. The coefficients represent the effect of each independent variable on the log-odds of being in a particular category, relative to a baseline category. The baseline category is usually chosen as the one with the highest frequency in the data, or it can be specified by the user.

In a multinomial logistic regression model with k categories, k-1 sets of regression coefficients are estimated, with one category serving as the reference or baseline category. The coefficients for the remaining categories indicate the change in log odds of being in that category compared to the baseline category, given a one-unit increase in the corresponding independent variable.

9.4.3 Tree-Based Methods

Classification and regression trees are a type of decision tree algorithm that recursively partitions the data into subsets based on the values of the independent variables, ultimately leading to the creation of a tree structure with decision nodes and leaf nodes. Each decision node corresponds to a split in the data based on the value of one of the independent variables, and each leaf node represents a predicted outcome for the dependent variable. CARTs are typically built using an algorithm that greedily selects the split that maximally reduces the impurity of the resulting subsets at each decision node.

The resulting tree can be used for prediction by following the branches from the root node to a leaf node and returning the corresponding predicted outcome. CARTs can also be used for variable selection by identifying the most important independent variables based on the number of times they are used in the tree. Interactions between independent variables can be captured by including interaction terms in the model or by allowing for multiple splits at the same decision node.

One common issue with CARTs is overfitting, where the tree is too complex and performs well on the training data but poorly on new, unseen data. Regularization techniques, such as pruning the tree or setting a minimum number of observations per leaf node, can be used to address overfitting. Additionally, ensemble methods such as random forests can be used to improve the stability and performance of CARTs by building multiple trees on bootstrap samples of the data and averaging their predictions.

Tree constructing algorithms differ with respect to:
• Splits into two or more groups at each node (binary vs. multi-way splits)
• Selection criterion for the independent variable for the next split
• Selection criterion for the split point of the independent variable
• Stopping criterion for the stepwise procedure
• Final prediction at the terminal node

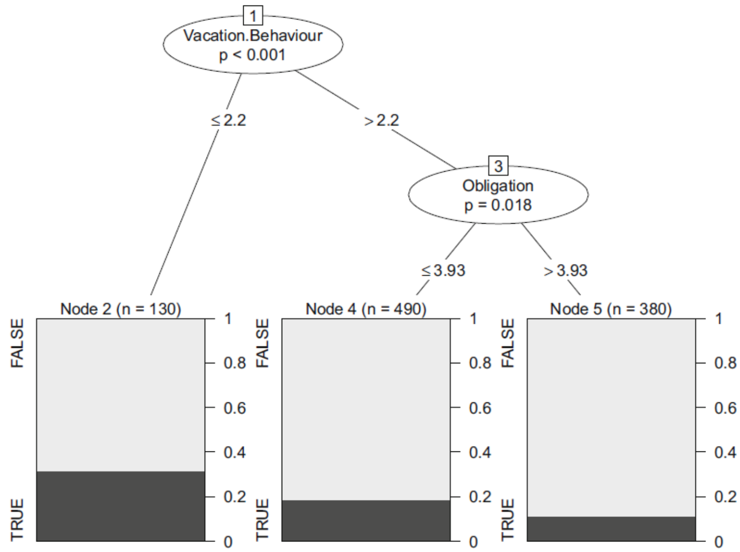


Fig. 9.15 Conditional inference tree using membership in segment 3 as dependent variable for the Australian travel motives data set

This is a correct summary of the output describing the fitted classification tree shown in Fig. 9.15. The tree is split into two nodes at the root node using the independent variable VACATION.BEHAVIOUR, and then split again at node 3 using the independent variable OBLIGATION. The predicted value for terminal node 2 is FALSE, and for terminal nodes 4 and 5 it is a mix of TRUE and FALSE. The output also provides information on the number of consumers in each terminal node and the proportion of wrongly classified respondents. Finally, the output notes the number of inner nodes and terminal nodes in the tree.

The fitted classification tree for segment 6 is more complex than that for segment 3; the number of inner and terminal nodes is larger. The stacked bar charts for the terminal nodes indicate how pure the terminal nodes are, and how the terminal nodes differ in the proportion of segment 6 members they contain.

The tree algorithm tries to maximize these differences. Terminal node 11 (on the right) contains the highest proportion of consumers assigned to segment 6. Node 11 contains respondents with the highest possible value for moral obligation, and a NEP score of at least 4. We can also fit a tree for categorical dependent variables with more than two categories with function classification tree.

Here, the dependent variable in the formula on the left is a categorical variable. C6 is a factor containing six levels; each level indicates the segment membership of respondents.

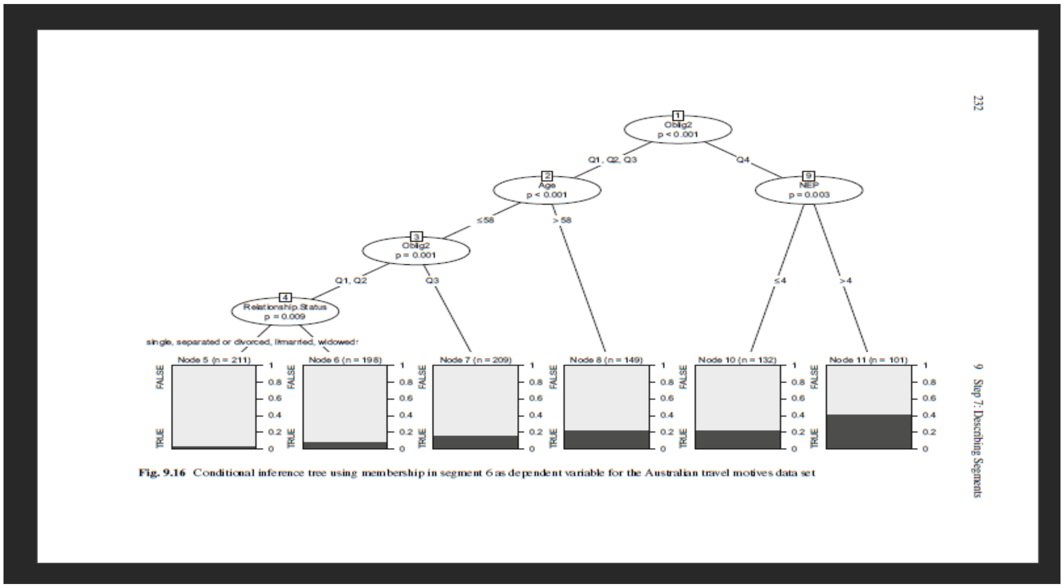


Fig. 9.16 Conditional inference tree using membership in segment 6 as dependent variable for the Australian travel motives data set

Figure 9.17 discusses the similarities and differences between the classification tree with a binary dependent variable and one with a categorical dependent variable. The main difference is the bar charts at the bottom, which show the proportion of respondents in each segment for each terminal node. The optimal situation is to have nearly all consumers in each node assigned to the same segment, or only a small number of segments. Node 7 in Figure 9.17 is a good example of this, with high proportions of members of segments 1 and 5 and low proportions of members of other segments.

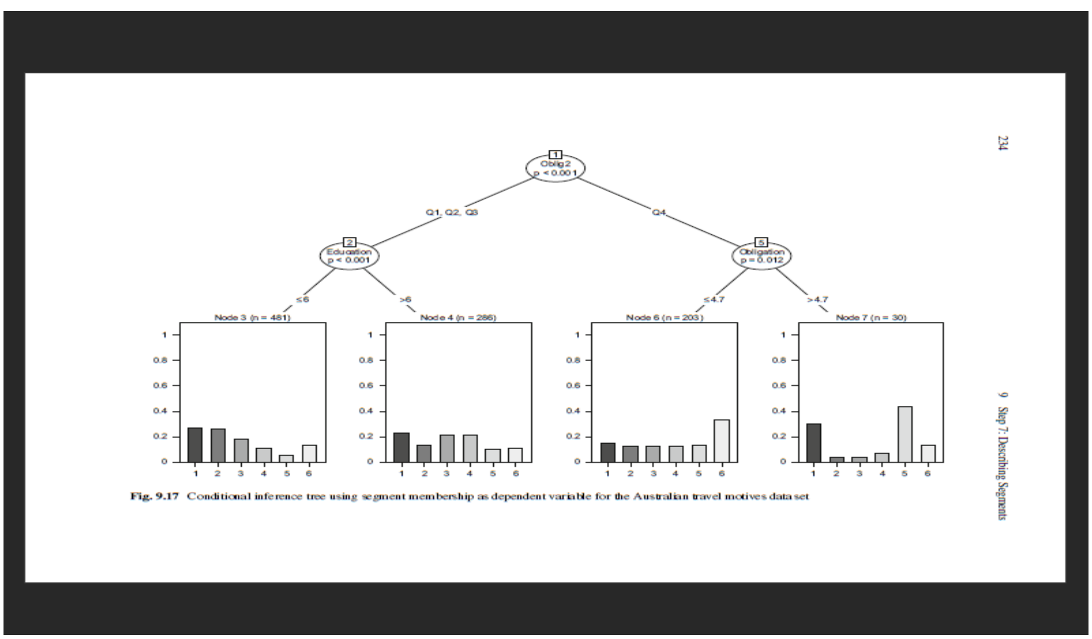


Fig. 9.17 Conditional inference tree using segment membership as dependent variable for the Australian travel motives data set

9.5 Step 7 Checklist

Task	Who is responsible?	Completed?
Bring across from Step 6 (profiling) one or a small number of market segmentation solutions selected on the basis of attractive profiles.		☐
Select descriptor variables. Descriptor variables are additional pieces of information about each consumer included in the market segmentation analysis. Descriptor variables have not been used to extract the market segments.		☐
Use visualisation techniques to gain insight into the differences between market segments with respect to descriptor variables. Make sure you use appropriate plots, for example, mosaic plots for categorical and ordinal descriptor variables, and box-and-whisker plots for metric descriptor variables.		☐
Test for statistical significance of descriptor variables.		☐
If you used separate statistical tests for each descriptor variable, correct for multiple testing to avoid overestimating significance.		☐
"Introduce" each market segment to the other team members to check how much you know about these market segments.		☐
Ask if additional insight into some segments is required to develop a full picture of them.		☐

https://drive.google.com/file/d/1kWtbYEruUSLyxfmRyNKQ9GQ55PJSxDOX/view?usp=share_link