



Modeling conversion of Television advertisement for Fast Moving
Consumer Goods (FMCG) – (*Viewer-to-Buyer Conversion*)

BY

WAMBUI SAMUEL KAMANDE

i56/67695/2013

School of Mathematics

College of Biological and Physical Sciences

University of Nairobi

Research project submitted in partial fulfillment of the requirements for the degree of Master of
Science in Social Statistics

2015

DECLARATION

CANDIDATE

I declare that this report is original and has never been presented for any other academic purpose.

Signature _____ Date _____

WAMBUI SAMUEL KAMANDE

SUPERVISOR

This project has been submitted for examination with my approval as supervisor.

Signature _____ Date _____

Ms. ANNE WANG'OMBE

ACKNOWLEDGEMENT

To Sarah Wambui, the most supportive mother a son could ask for, I am eternally grateful. I would not have the motivation and strength for this journey without you.

I would like to thank my Supervisor, Ms. Anne Wang'ombe for the tremendous knowledge impacted and advice, not only through this project, but over the six years that I have been a student at the School of Mathematics.

Very much obliged to Dr. Kenfield Griffith and his wonderful team at mSurvey for the help with data acquisition using the awesome mSurvey SMS platform. What would have taken a few weeks and probably a couple of dollars more took two days and \$52. Grateful!

My classmates of two years at the School of Mathematics made it all a fun journey, over and above the continuous learning from each other; and for that, I am immensely thankful.

To my Data Science colleagues at Nielsen, the encouragement and willingness to listen to my ideas made it much easier for me to have a near perfect balance amidst the endless work load.

Very much obliged.

And to God, for the wisdom, strength and shocking grace through the good and spirit wrenching bad times in the two years, I cannot sufficiently express my gratitude.

To Wambui, Andrew and UNK...

ABSTRACT

There are currently more than 107 TV stations in Kenya (with just over 10 free to air dominating the market), a number that has been growing exponentially since 2001. Further, more than 80% of the country's population has access to a television. Driven by these two factors and the growing economy, advertising revenue for broadcasters has grown threefold from \$107 million in 2007 to \$359 million in 2013. With all this money being invested into TV advertising by companies, there has been a limited availability and exposure to tools for measuring the return of such huge investments for Ad spots. Research companies have developed tools to test ads and define the qualities of good advertisement, but none has zeroed down on estimating the conversion rates of those exposed to advertisement; the probability of audiences being converted to buyers of the advertised product. With a special focus on Fast moving consumer goods, a generalized linear model is obtained to estimate the probability of conversion from "viewers" to "buyers" for those that have been exposed to a particular TV advertisement. Data for 120 residents of Nairobi is collected. Demographic characteristics, social economic status, exposure, purchase habits and motivators data are collected. A multinomial logistic model was constructed using this data, with the response being a three-level multinomial variable – "Will buy", "Will consider buying" and "Will not buy". Six variables significantly influence the conversion of Television ad viewer to buyers – Gender, income/social class, Level of education, total time spent watching TV in a day, main television interest and most important feature of an advertisement. The model was validated by (a) significant test of the overall model, (b) tests of regression coefficients, (c) goodness-of-fit measures, & (d) validation of predicted probabilities. Three methodological issues were highlighted in the discussion: (1) the use of odds ratio, (2) the Hosmer and Lemeshow test extended to multinomial logistic models, and (3) the missing data problem. Believability and relatability of a television advertisement increase the probability of conversion by three times compared to the length/precision aspect. People with either primary or secondary education are also 3 times more likely to be converted compared to those with tertiary education.

CONTENTS

DECLARATION	ii
ACKNOWLEDGEMENT	iii
ABSTRACT	v
INTRODUCTION	8
1.1 BACKGROUND	8
1.1.1 Regression analysis for social sciences	8
1.1.2 Multinomial Logits	11
1.1.4 Television Advertising	12
1.1.3 Regression analysis in TV advertising	15
1.2 Problem statement	15
1.3 Objectives of the study	16
1.3.1 Main objective	16
1.3.2 Specific objectives	16
1.4 Significance of the study	16
LITERATURE REVIEW	18
METHODOLOGY	23
3.1 Ordinary least square Models	23
3.2 Generalized Linear models	24
3.2.1 Random component	25
The Binomial distribution	26
3.2.2 Systematic component:	26
3.2.3 Link function:	27
3.3 Properties of the exponential family of distributions	27
3.4 Logistic Regression Model	32
3.5 Multinomial Distribution	33
3.6 Multinomial logistic regression	34
3.6.1 Estimation of the regression coefficients	38
3.6.2 Independence of irrelevant alternatives	42

3.6.3 Statistical inference on regression coefficients and coefficient interpretation	43
3.7 Analysis Software	46
ANALYSIS	47
4.1 Response variable	47
4.2 Predictor Variables	48
4.3 Parameter Estimation and model fit tests	52
4.4 Interpretation and prediction	56
4.4.1 Consider purchasing Vs Not purchasing	57
4.4.2 Purchasing Vs Not purchasing	59
CONCLUSION AND RECOMMENDATIONS.....	62
5.1 Conclusions and Recommendations.....	62
5.2 Limitations and future research	63
REFERENCES.....	65
Appendix	69

INTRODUCTION

1.1 BACKGROUND

1.1.1 Regression analysis for social sciences

A general goal of regression analysis is to estimate the association between one or more explanatory variables and a single outcome variable. An outcome variable is presumed to depend in a random fashion or could be systematically predicted by the explanatory variable(s). The explanatory variables are thought to independently affect the outcome variable; hence, they are often known as independent variables.

Multiple linear models have several independent/predictor variables which help us understand how these affect other dependent variable(s). We examine through these models, the impact of a variable on the dependent variable when the rest of the predictors are held constant. These simple models are built under a set of assumptions, which must be satisfied prior to fitting. (John P. Hoffmann, 2010)

Regression thus primarily helps us navigate through and establish the following:

- a) The form of the relationship among the outcome and explanatory variables, or what the equation that represents the relationship looks like.
- b) The direction and strength of the relationships based on the valence and size of the slope coefficients.
- c) Which explanatory variables are important and which are not. This is based on evaluation of the measures such as p-values and/or confidence intervals

- d) Prediction of a value or set of values of the outcome variable for a given set of values of the explanatory variables.

Social and economic data often presents researchers with outcome variables that are not linear, such as survey respondents' choices among two or more options, which require different types of non-linear transformations in order to be estimated. The Ordinary Least Square models (Simple and Multiple) are rendered ineffectual and restrictive in such cases, specifically because they only cater for continuous response variables having a normal distribution and the relationship between the response and predictor variables is a strictly simple/identity function.

Generalized Linear Modeling, a flexible generalization of the ordinary least squares regression, is a common technique used to obtain meaningful results in these cases since it allows for transformations and the response variables can have a distribution other than normal. These models help us accommodate binary, ordered and multinomial dependent variables, count data and positive valued continuous distributions (Nelder and Wedderburn, 1972; Hilbe, 1994; Hoffman, 2004). In other terms, generalized linear models allow for extension of the distribution of the response variable to the exponential family of distributions – For a random variable Y with probability distribution function $f(y; \theta)$, where θ is an unknown parameter; Then the distribution is said to belong to the exponential family of distributions if the probability distribution function can be expressed in a generalized form. The normal, Poisson and Binomial (And by extension, multinomial) distributions can be expressed in a generalized format and belong to the exponential family.

This characteristic and adaptability makes the generalized linear models extremely imperative in this field. Because of its flexibility in addressing a variety of statistical problems and the availability of software to fit the models, it is considered a valuable statistical tool and is widely used. The generalized linear model has been referred to as the most significant advance in regression analysis in the past twenty years (Hoffman 2004).

A Generalized Linear Model (GLM) consists of three components:

- a) Random component - This is the response and an associated probability distribution.
- b) Systematic component – This includes explanatory variables and relationships among them.
- c) Link function - This is the part that specifies the relationship between the systematic component or linear predictor and the mean of the response. The link function allows for generalization of the linear models for count, binomial and percent data thus ensuring linearity and constraining the predictions to be within a range of possible values (Guisan, 2002). This ability to handle a larger class of error distributions and data types is one of the key improvements of GLMs that makes them more usable and applicable compared to linear models.

1.1.2 Multinomial Logits

This is based on the multinomial distribution. If we consider a random variable Y_i that may take one of several discrete values, 1, 2, 3... k. Further, we let $P_{ik} = \Pr\{Y_i = k\}$ be the probability that the i-th response falls in the k-th category. Assuming that the response categories are mutually exclusive (no response falls in more than one category) and exhaustive, it follows that the probabilities add up to one for each individual.

The simplest approach is to nominate one of the response categories as the reference, and calculate the log-odds of the rest of the categories relative to the reference. We then let the log-odds be a linear function of the predictors. It is advisable to pick the last category as the reference and calculate the odds that a member (respondent) of the group i falls in category k as opposed to the reference as P_{i1}/P_{iK} . In modeling multinomial logits, we assume that the log-odds of each response follow a linear regression:

$$\eta_{ik} = \log \frac{P_{ik}}{P_{iK}} = \alpha_k + x_i' \beta_k$$

Where α_k is a constant and β_k is a matrix of regression coefficients, for $k = 1, 2, 3 \dots K-1$ (explicit of the vector of 1's).

This model is analogous to a logistic regression model, the only difference being that the probability distribution of the response is multinomial and not binomial. We also have a set of $K-1$ equations as opposed to one, with each category contrasting each of categories 1, 2, 3... $J-1$

with category J. The single logistic regression equation is only a contrast between successes and failures. A key distinguishing element of the multinomial logit model is that there is a single response variable that gives a description of the individual, and not the alternatives that face the individual.

1.1.4 Television Advertising

The very first television advertisement ever broadcasted was on July 1, 1941 during a baseball game on a local New York channel. The 10-second ad advertised Bulova watches and cost a mere four dollars. Due to the overwhelming success of the Bulova advertisement, other companies began to realize that they needed to jump on board with their marketing as well. By 1948, many additional advertisers were using television spots to reach the large audience that owned television sets. Television's spreading popularity merited the formation of the American Association of Advertising Agencies to regulate commercials. Television was so popular during that era that even the movie studios feared that television would dominate all other media.

Locally, in 1959, the Kenya Broadcasting Corporation was founded by the colonial government. By the end of 1962, a transmission station and recording studio had been set up, and television was officially launched the following year, running as an autonomous public corporation. It drew its revenues mostly from advertising. Fast forward to the 1990s and private television stations began as a result of expansion and modernization. Currently, there are more than 12 dominant free to air TV stations in the country, each competing for audiences and advertising dollars subsequently.

With such a large and diverse pool of stations for advertisers to choose from, it is becoming more important for them to evaluate the returns of the money and improve their placement, partly by maximizing on well thought out distribution and placement strategies.

Television advertising for consumer goods is done with two general underlying objectives:

- a) Getting consumer's attention and drawing their attention to the advertised product, thus creating awareness. Building awareness also helps companies to reach out to prospective consumers that either add on to the pool of existing consumers or replace lost ones.
- b) Prompting immediate action (Call to action) – This is a strategic prompt to consumers to purchase the advertised good, subsequently increasing sales volumes, a measurable metric for Return on Investment.

In Kenya, TV remains the most effective channel for advertising, even with increasing traction of the internet. In a report by media monitoring firm, Reelforge, companies in Kenya splurged Sh85.8 billion on advertising in 2014, compared to Sh79.2 billion in 2013. Of these, Sh41.8 billion went to TV advertising, accounting for 42% of the total advertising revenue. The increase was partly attributed to an increase in rate cards. Further, Pricewaterhouse Coopers (PwC) Kenya, a leading audit and management firm, estimates that the share of TV in advertising revenue is likely to rise to 50% by 2017 (865 million USD). Kenya is believed to have relatively higher rates of advertising agency commission: a TV ad in Kenya costs more than twice the Nigerian price for the same air time. With manufacturers spending approximately more than 400 million USD on TV advertisements, improvement of how they maximize their impacts is

paramount. Currently, advertising agencies use qualitative and quantitative techniques to test advertisements, pre and post deployment. The qualitative techniques include - Focus Groups, In-depth Interviews, Observations, Post Advertisement Recognition Tests, Ethnography, Projective Techniques and Neurological Networks. Quantitative techniques that are currently used include Persuasion Measures, Electronic Response Measures, Optical Laser Scanning at retail check out, Split Cable Technology and Sample Tracking Studies.

This paper uses statistically rigorous modeling to help establish the numeric chances of getting the consumers that are targeted by the TV ad to make purchases. This, unlike most approaches, can help provide an estimate of the expected impact and subsequent return on investment, even before launch.

1.1.3 Regression analysis in TV advertising

Modeling market response is intended to help scholars and managers understand how consumers individually and collectively respond to marketing activities, and how competitors interact.

Appropriately estimated effects constitute a basis for improved decision making in marketing. (Hanssens, 2005). Much of marketing decision making is of a repetitive or tactical nature. For example, advertising expenditures, sales promotion budgets, shelf space allocations, prices, margins, etc. have to be determined for each period. The consideration of changes in decisions is facilitated by the development of ever more detailed data, whose availability also makes it easier to justify the use of econometric modelling (e.g. bimonthly audit data would not permit the estimation of deal effect curves). And the increasing frequency and amount of marketplace feedback also demands a systematic approach for data analysis.

Standardized models have become important tools to improve the quality of tactical marketing decisions at functional levels such as brand management.

1.2 Problem statement

With 85% TV penetration in Kenya and over \$359 million being spent yearly on advertising, manufacturers of fast moving consumer goods and the advertising agencies have tools to test advertisements before going live, but have no rigorous, tested and definitive tools of measuring the conversion rates and probabilities for the target demographics.

1.3 Objectives of the study

1.3.1 Main objective

The main objective of this study is to model the conversion of viewers of Television FMCG advertisements from mere viewers to buyers of the advertised products.

1.3.2 Specific objectives

- 1. To establish and model the significant factors that significantly affect the conversion of Television FMCG ad viewers to buyers of the advertised products.*
- 2. To estimate the effect of the selected factors on the effect on conversion probabilities.*
- 3. To estimate the probability of viewer-to-buyer conversion for TV FMCG advertisement based on audiences decision to purchase the advertised product, contemplation to purchase or not to purchase at all.*
- 4. To establish the most desirable and optimum qualities of TV FMCG advertisements that lead to higher viewer-to-buyer conversion.*

1.4 Significance of the study

Advertisers – manufacturers and advertising agencies - Measure the expected effect of advertisements that are geared towards increasing sales and tailor them appropriately based on target audience and various preferences

Broadcasters - Measure the impact of FMCG advertisements that run on their channels, thus developing rate cards on a more scientific and statistically driven criterion.

Academics – Lay grounds for extension of modeling of ad conversion into other media and techniques, eventually obtaining an optimal model that collates all platforms and yields estimates for the conversion probabilities of each, providing the ultimate marketing mix.

LITERATURE REVIEW

Peng and Nichols (2003) used multinomial logistic regression to profile adolescents who were at greatest behavioral risks in Indiana University. They applied the model to a data set that they collected themselves. Results showed that gender, intention to drop from the school, family structure, self-esteem, and emotional risk were effective predictors. The model was then validated by the overall significant test, tests of regression coefficients, goodness-of-fit measures and validation of predicted probabilities. Three methodological issues – use of odds ratio, the Hosmer and Lemeshow test extended to multinomial logistic models and the missing data problem were highlighted in the discussion.

Frees (2003) simply lays out the benefits of Generalised Linear Models - Linear modeling theory provides a platform for choosing appropriate linear combinations of explanatory variables to predict a response. In generalized linear models, statisticians have the added ability to widen the class of distributions to allow for handling of other types of non-normal outcomes. This broad class includes, as special cases, the normal, Bernoulli and Poisson distributions. Further, generalized linear models methods are used vastly in various fields due to the fact that the maximum likelihood estimators can be computed quickly through the iterated reweighted least squares technique.

Jain et al. (1994) used a multinomial linear regression model on a marketing data set that was further analyzed by Chen and Kuo (2001S). The subjects under observation based on Nielsen data consist of n=100 households in Springfield, Missouri. The response of interest was the type

of yogurt purchased, consisting of four brands: Yoplait, Dannon, Weight Watchers and Hiland. The households were monitored over a two-year period with the number of purchases ranging from 4 to 185; the total number of purchases was $N=2,412$. The two marketing variables of interest were PRICE and FEATURES. For the brand purchased, PRICE was recorded as price paid, that is, the shelf price net of the value of coupons redeemed. FEATURES was classified as a binary variable, defined to be one if there was a newspaper feature advertising the brand at time of purchase, and zero otherwise. One brand was the omitted alternative. The results obtained showed that every parameter was statistically significantly different from zero. Thus, the parameter estimates were useful when predicting the probability of choosing a brand of yogurt. Moreover, in a marketing context, the coefficients have important substantive interpretations. The results were interpreted to suggest that a consumer is 1.634 times more likely to purchase a product that is featured in a newspaper ad compared to one that is not. For the PRICE coefficient, a one cent decrease in price suggests that a consumer is 1.443 times more likely to purchase a brand of yogurt.

Market response models are intended to help scholars and managers understand how consumers individually and collectively respond to marketing activities, and how competitors interact according to Hansens et al. (2005). Appropriately estimated effects constitute a basis for improved decision making in marketing. In the past years market response models have diffused in the practitioners' community. Leading firms, especially in consumer goods and services, database marketing companies and traditional market research companies develop and use increasingly sophisticated models and analyses. The successful implementation of models

depends on data availability, the methodology used, and other characteristics. It appears that many models appearing in the academic literature have little relation to marketing practice. Such models often deal with specific problems, are more descriptive than prescriptive, and include complexities that reduce the chance of implementation in practice. Academics are challenged to make models that can be easily integrated into real corporate problems, yielding sound solutions.

Randall and Justin (2013) imply in their research the near impossibility of measuring returns on advertising using classical theories to measure the causal impact and conversion probabilities. In their paper, it is explained that statistical evidence from the randomized trials is very weak because individual-level sales are incredibly volatile relative to the per capita cost of a campaign—a “small” impact on a noisy dependent variable can generate positive returns.

According to Shachar and Emerson (2000), an accurate television viewing choice model is a critical working tool for both television network executives, who face difficult programming, scheduling and marketing decisions, and advertisers, who want to get the most from their spend. Shachar and Emerson claim that such a model can help television executives maximize ratings by improving both the scheduling and the characteristics of their shows. In addition it can help advertisers predict ratings and the demographic composition of the audiences. Many researchers have developed such rating models (e.g.; Rust, Kamakura, Wagner & Alpert 1992; Tavakoli & Cave 1996; Meyer & Hyndman 2006), however, the usefulness of these models to advertisers is questionable when there is no accompanying prediction of the frequency of channel switching as an indicator of reduced advertising attention. In this study we develop a model which describes the frequency of channel switching through the use of network loyalty measures.

The demographic and socio-economic characteristics of television viewers play a pivotal role in ascertaining criteria relating to why certain shows are frequently watched. Rao (1975), Webster (1986). There has been extensive research carried out on understanding the motivations and determinants for TV viewing from a British perspective. Tavakoli et al. (1989), Tavakoli & Cave (1996). Various researchers have over the years also looked for ways to examine the relationship between viewers and programs. Zubayr (1999) performed demographic comparisons for repeat viewing rates, finding that repeat viewing rates are slightly higher for women and for older people. It is therefore suggested that demographic and socio-economic effects will also impact on the impact of the advertisements as well as the resultant action or otherwise.

Statisticians have used proportional hazard models in the past to predict online advertisement conversions. Chandler-Pepelnjak (2010) however used a syndicated approach that married the leading edge of online advertising conversion attribution (Engagement Mapping) to the proportional hazard model, eventually producing a tool that can be used to find optimal settings for advertiser models of online conversion attribution.

Grosswindhager (2009) discussed three different dimension reduction approaches to fitting logistic regression models in advertising for classification purposes. For Principal Component Logistic Regression (PCLR), Grosswindhager proposes two different methods on cross validation – R-statistic and PRESS MSE - and the average eigenvalue rule for determining the significant set of principal components. Further he suggests that there exists no ultimate rule for choosing the ultimate the number of Principal Components, and suggests that to take three or at

least two criteria into consideration. Grosswindhager also introduces and discusses General Partial Least Squares.

METHODOLOGY

3.1 Ordinary least square Models

In most research projects many researchers often aim to explain a random variable Y_i with one or more explanatory variables, X_i . An outcome variable is presumed to be systematically explained by one or more predictor variables. Further, the independent variables are assumed to independently impact on the outcome variable.

The simple linear regression model has one predictor/independent variable and takes the form

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon \quad 3.1$$

Where Y is the response variable. X is the predictor variable and ε is the random error. ε is a result of both measurement and non-measurement errors, and is assumed to have a normal distribution with mean μ and variance δ^2 i.e. $\varepsilon \sim N(\mu, \delta)$. The mean of the error is checked for variability, and if found to not be constant, heteroscedascity is inferred. Heteroscedascity must be removed before proceeding.

The multiple linear regression is characterized by more than one predictor/independent variable and takes the form

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon \quad 3.2$$

where $\beta_1, \beta_2, \dots, \beta_k$ are the regression parameters/coefficients.

This equation is said to be linear in terms of the parameter/coefficients and not the variables.

3.2 Generalized Linear models

As discussed in, a general ordinary least squares (OLS) regression model is of the form

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon$$

The response variable $y_i; i=1,2,\dots,n$ is modeled as a linear function of predictor variables $x_j; j=1,2,\dots,k$ as well as an error term. The response variable is a continuous variable that has normal distribution, a result of the normal distribution of the error term.

Generalized Linear Models (GLM) go beyond this in two major respects:

- (i) The response variables can have a distribution other than normal; i.e any distribution within a class of distributions known as “exponential family of distributions”.
- (ii) Instead of having

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon \quad 3.3$$

we can allow for transformation of the equation into

$$g(Y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon \quad 3.4$$

- (iii) Generalized linear models are a flexible generalization of ordinary least squares regression that relates the distribution of the response variable to the systematic portion of the experiment (**the linear predictor**) through a function called the **link function**.

Generalized Linear Models (GLM) have three components:

- (i) Random component
- (ii) Systematic component
- (iii) Link function

3.2.1 Random component

For any random variable Y with probability distribution function $f(y; \theta)$; where θ is an unknown parameter; if the probability distribution function can be expressed in the form

$$f(y; \theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)] \quad 3.5$$

The distribution then is said to belong to the *exponential family of distributions*.

Further, if $a(y) = y$, then the distribution is said to be in *canonical form* and $b(\theta)$ is known as a *natural parameter*.

Some examples of distributions that belong to exponential family of distributions and can be written in canonical form include the normal and Poisson distributions. Below is an illustration of how the Binomial distribution and by extension, the multinomial distribution falls into the exponential family.

The Binomial distribution

$$Y \sim b(n, \theta) \quad 3.6$$

$$f(y; \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y} \quad 3.7$$

$$= \exp \left[y \ln \left(\frac{\theta}{1-\theta} \right) + n \ln(1 - \theta) + \ln \binom{n}{y} \right] \quad 3.8$$

$$a(y) = y; \quad b(\theta) = \ln \left(\frac{\theta}{1-\theta} \right); \quad c(\theta) = n \ln(1 - \theta); \quad d(y) = \ln \binom{n}{y} \quad 3.9$$

3.2.2 Systematic component:

This is the linear predictor part of the model. It is the quantity which incorporates the information about the predictor variables into the model.

For predictor j and observation i

$$\eta_{i=\sum_{j=1}^k \beta_j x_{ij}} \quad 3.10$$

It is related to the expected value of the data through the link function

3.2.3 Link function:

This function links the systematic component to the random component. It provides for the relationship between the linear predictor and mean of the distribution function.

Let $E[Y_i] = \mu_i$ and g be a monotonic, differentiable function.

If $g(\mu_i) = \eta_i$

then g is a link function. The relationship is **non-linear**

For this exponential family of distributions that can be written in canonical form, the natural parameter is used as the link function. It is thus known as the canonical link function.

3.3 Properties of the exponential family of distributions

The expressions for the expected value and the variance of $a(y)$ can be found. To obtain these expressions, the following results that apply for any probability density function are used, provided that the order of differentiation and integration are interchangeable. From the basic definition of a probability density function, the area under the curve is equal to 1, so

$$\int f(y; \theta) dy = 1 \tag{3.11}$$

Integration is applicable over all possible values of y , if continuous. (We replace integration with summation if y is discrete)

Differentiating both sides of 3.11 with respect to θ yields

$$\frac{d}{d\theta} \int f(y; \theta) dy = \frac{d}{d\theta} \cdot 1 = 0 \quad 3.12$$

Reversing the order of integration and differentiation gives

$$\int \frac{d f(y; \theta)}{d\theta} dy = 0 \quad 3.13$$

If differentiated twice and the order reversed for 3.13, we obtain

$$\int \frac{d^2 f(y; \theta)}{d\theta^2} dy = 0 \quad 3.14$$

The above results are used in the exponential family of distributions to obtain the properties.

From equation 3.5

$$f(y; \theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)]$$

Thus we have

$$\int \frac{d f(y; \theta)}{d\theta} = [a(y)b' + c'(\theta)]f(y; \theta) \quad 3.15$$

Using 3.13 gives us

$$\int [a(y)b' + c'(\theta)]f(y; \theta) dy = 0 \text{ which simplifies to}$$

$$b'(\theta)E[a(y)] + c'(\theta) = 0 \quad 3.16$$

Since $\int a(y)f(y; \theta) dy = E[a(y)]$ by the basic definition of the expected value, and

$$\int c'(\theta)f(y; \theta) dy = c'(\theta) \text{ from equation 3.12.}$$

A rearrangement of 3.16 gives

$$E[(Y)] = -\frac{c'(\theta)}{b'(\theta)} \quad 3.17$$

Similarly, we can obtain the variance of $a(Y)$ using the second derivative

$$\frac{d^2 f(y; \theta)}{d\theta^2} = [a(y)b''(\theta) + c''(\theta)]f(y; \theta) + [a(y)b'(\theta) + c'(\theta)]^2 f(y; \theta) \quad 3.18$$

We can rewrite the second term of the right hand side as below using the expression for the expectation from equation 3.17

$$[b'(\theta)]^2 \{a(y) - E[a(Y)]\}^2 f(y; \theta)$$

Further, using equation 3.14

$$\int \frac{d^2 f(y; \theta)}{d\theta^2} dy = b''(\theta)E[a(Y)] + c''(\theta) + [b'(\theta)]^2 \text{Var}[a(Y)] = 0 \quad 3.19$$

since $\text{Var}[a(Y)] = \int \{a(y) - E[a(Y)]\}^2 f(y; \theta) dy$ by definition

Rearranging equation 3.19 and subsequently substituting equation 3.17 gives

$$\text{Var}[a(Y)] = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3} \quad 3.20$$

The equations for the expected value and variance of $a(Y)$ (3.17 and 3.20 respectively) are easily verifiable for the Poisson, Normal and Binomial distributions. They are also used to obtain the expected value and variance for other distributions in the exponential family.

The expressions for the expected value and variance of the derivatives of the log-likelihood function can also be obtained.

The log-likelihood function for a distribution in the exponential family is given as $l(y; \theta) = a(y)b(\theta) + c(\theta) + dy$ from equation 3.5.

The derivative for the log-likelihood function with respect to θ is given by

$$U(y; \theta) = \frac{dl(y; \theta)}{d\theta} = a(y)b'(\theta) + c'(\theta)$$

The function U is called the score statistic and it can be regarded as a random variable as it depends on y . i.e.

$$U = a(Y)b'(\theta) + c'(\theta) \tag{3.21}$$

The expected value is given by

$$E(U) = b'(\theta)E(a(Y)) + c'(\theta).$$

Substituting the expression for $E[a(Y)]$ from equation 3.17 yields

$$E(U) = b'(\theta) \left[-\frac{c'(\theta)}{b'(\theta)} \right] + c'(\theta) = 0$$

The variance for U is known as the information, and is denoted by \mathfrak{I} .

$$\mathfrak{I} = \text{Var}(U) = [b'(\theta)^2] \text{var}[a(Y)].$$

Again, substituting the expression for $\text{Var}[a(Y)]$ from equation 3.20 yields

$$\text{Var}(U) = \frac{b''(\theta)c'(\theta)}{b'(\theta)} - c''(\theta) \tag{3.22}$$

The score statistic, U , is to make inference about parameter values in generalized linear models.

Another important property is

$$\text{Var}(U) = E(U^2) = -E(U') \quad 3.23$$

which follows from the general equation for any random variable, and the fact that $E(U)=0$

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

To obtain the second equality, U is differentiated with respect to θ . From equation 3.21

$$U' = \frac{dU}{d\theta} = a(Y)b''(\theta) + c''(\theta)$$

The expected value of U' is thus given by the following after substituting equation 3.17 and then using equation 3.22

$$\begin{aligned} E(U') &= b''(\theta)E[a(Y) + c''(\theta)] \\ &= b''(\theta) \left[-\frac{c'(\theta)}{b'(\theta)} \right] + c''(\theta) \\ &= -\text{Var}(U) = -\mathfrak{I} \end{aligned} \quad 3.24$$

Breslow (1999) lists some of the critical assumptions that underline generalized linear models, many of which apply to any regression model

- Statistical independence of the n observations
- Correct specification of the variance function
- Correct specification of the expected value function
- Correct specification of the link function
- Correct form for the explanatory variables

- Lack of undue influence of individual observations on the fit

Failure of any one of these assumptions may seriously compromise the conclusions of the model and analysis.

3.4 Logistic Regression Model

Logistic regressions are probability models whose origins can be traced in epidemiological studies. They are now commonly used in a wide range of fields including engineering, social studies, economics and Finance. In contrast to the ordinary Least squares linear regression, where the independent variable is continuous in nature, logistic regression deal with categorical in nature variables, which could either have two (binary) or more (Multinomial) categories.

We can define a binary response variable as

$$Y = \begin{cases} 1, & \text{if the outcome is a success} \\ 0, & \text{if the outcome is a failure} \end{cases}$$

If $\Pr(Y=1) = p$ and $\Pr(Y=0) = 1 - p$, then $E[Y] = p$. Y is then said to be a Bernoulli random variable. If there are n such variables, $Y_1, Y_2, Y_3, \dots, Y_n$ that are independent with the probability, $\Pr(Y_i = 1) = p_i, \forall i$, then their joint probability distribution is given by

$$\prod_{i=1}^n (p_i)^{y_i} (1 - p_i)^{1-y_i} = \exp \left[\sum_{i=1}^n y_i \ln \left(\frac{p_i}{1-p_i} \right) + \sum_{i=1}^n (1 - p_i) \right] \quad 3.25$$

This is a member of the exponential family of distributions.

Subsequently, a generalized linear model can be developed with $Y_1, Y_2, Y_3, \dots, Y_n$ as the response variables.

3.5 Multinomial Distribution

The multinomial distribution is an extension of the binomial distribution. Whereas the binomial distribution models nominal data with two categories, the multinomial distribution is used to model nominal data whose outcome has more than two categories.

We consider a random variable Y that has J categories. Further, we let $\pi_1, \pi_2, \pi_3, \dots, \pi_j$ to denote the respective probabilities, with $\pi_1 + \pi_2 + \pi_3 + \dots + \pi_j = 1$. If there are n independent observations of Y which result in y_1 outcomes in the first category, y_2 outcomes in the second category and so on, we let

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_j \end{bmatrix}, \text{ with } \sum_{j=1}^J y_j = n$$

The multinomial distribution is hence given by

$$\begin{aligned} f(\mathbf{y}|n) &= \frac{n!}{y_1!y_2!\dots y_j!} \pi_1^{y_1} \pi_2^{y_2} \dots \pi_j^{y_j} & 3.26 \\ &= \frac{n!}{\prod_{j=1}^J y_j!} \prod_{i=1}^J \pi_i^{y_i} \end{aligned}$$

When $J=2$, then we it simplifies to a binomial distribution since $\pi_2 = 1 - \pi_1$ and $y_2 = 1 - y_1$.

For a multinomial distribution with $1 \leq j, k \leq J$, where $j \neq k$

$$\mu_j = E[y_j] = n\pi_j$$

$$\text{Var}(y_j) = n\pi_j(1 - \pi_j) = \mu_j \left(1 - \frac{\mu_j}{n}\right)$$

$$\text{Cov}(y_j, y_k) = -n\pi_j\pi_k = -\frac{\mu_j\mu_k}{n}$$

3.6 Multinomial logistic regression

For response variables that have no natural order among the categories, like what this paper proposes to use, nominal logistic regression models are used. One category is arbitrarily chosen as the reference category, sometimes with a consideration of how the researcher wants to report. In this case where the response variable has three categories (Buys, considers buying, does not buy), the ‘does not buy’ category will be used as the reference. The logits for the other categories are then defined by

$$\text{logit}(\pi_j) = \log\left(\frac{\pi_j}{\pi_1}\right) = \mathbf{X}_j^T \boldsymbol{\beta}_j, \text{ for } j = 2, \dots, J \quad 3.27$$

The (J-1) logit equations are used simultaneously to estimate the parameters $\boldsymbol{\beta}_j$. In this case, $J=2$. The linear predictors $\mathbf{X}_j^T \mathbf{b}_j$ can be calculated once the parameter estimates \mathbf{b}_j have been obtained. From equation 3.27,

$$\hat{\pi}_j = \hat{\pi}_1 \exp(\mathbf{X}_j^T \mathbf{b}_j) \text{ for } j = 2, \dots, J$$

But $\hat{\pi}_1 + \hat{\pi}_2 + \hat{\pi}_3 + \dots + \hat{\pi}_j = 1$, thus

$$\hat{\pi}_1 = \frac{1}{1 + \sum_{j=2}^J \exp(\mathbf{X}_j^T \mathbf{b}_j)}$$

and

$$\hat{\pi}_j = \frac{\exp(\mathbf{x}_j^T \mathbf{b}_j)}{1 + \sum_{j=2}^J \exp(\mathbf{x}_j^T \mathbf{b}_j)}, \text{ for } j = 2, \dots, J$$

The fitted values for each of the covariate patterns can be calculated by multiplying the estimated probabilities $\hat{\pi}_j$ by the total frequency of the covariate pattern.

The Pearson Chi-squared residuals are obtained by

$$r_i = \frac{o_i - e_i}{\sqrt{e_i}}$$

Where o_i and e_i are the observed and expected frequencies for $i = 1, 2, \dots, N$. N is J times the number of distinct covariate patterns. We can assess the adequacy of the model using the residuals.

The Chi-square statistic is given by

$$\chi^2 = \sum_{i=1}^N r_i^2$$

The deviance statistic is defined in terms of maximum values of the log-likelihood function for the fitted model, $l(\mathbf{b})$, and for the maximal model $l(\mathbf{b}_{max})$. The deviance statistic D describes the lack of fit; the larger the value, the poorer the fit. A significant fit is one in which the test gives a non-significant result.

$$D = 2[l(\mathbf{b}_{max}) - l(\mathbf{b})]$$

Alternatively, we can use a null model (one with only the intercept) to test for the adequacy of the model fitted. The hypotheses to carry this test are

H_0 : The null model (intercept only) is a better fit Vs. **H_1** : The fitted model is a better fit

The deviance statistic describes the goodness of fit – the test gives a significant result for a significant fit.

The likelihood ratio chi-square statistic is defined in terms of the maximum value of the log likelihood function for the minimal model. It is given by

$$C = 2[l(b) - l(\mathbf{b}_{min})]$$

We can also evaluate the Pseudo $R^2 = \frac{l(\mathbf{b}_{min}) - l(b)}{l(\mathbf{b}_{min})}$

If the model fits well, then both χ^2 and D have, asymptotically, the distribution χ^2_{N-p} , where N is the number of observations and p is number of parameters estimated. C has the asymptotic distribution $\chi^2_{[p-(J-1)]}$ because the minimal model will have one parameter for each logit defined in equation 3.3.

According to Chapman and Hall (2002), it is often easier to interpret the effects of explanatory factors in terms of odds ratios than the parameters β . More contextually, we consider the three response categories in this project. Further, we consider interpretation of the binary explanatory variable “Remembrance”; if the respondent remembers the last TV advertised they watched, then $x=1$ and if they do not remember, then $x=0$. The odds ratio for remembrance for response j ($j=2,3$) relative to the reference category $j=1$ is

$$OR_j = \frac{\frac{\pi_{jp}}{\pi_{ja}}}{\frac{\pi_{1p}}{\pi_{1a}}}$$

Where π_{jp} and π_{ja} denote the probabilities of response category j ($j=2, 3$) based on whether remembers or does not remember, respectively. Assuming that “Remembrance” is the only explanatory variable, we have the model

$$\log\left(\frac{\pi_j}{\pi_1}\right) = \beta_{0j} + \beta_{1j}x, \quad j = 2,3$$

The log odds are given by

$\log\left(\frac{\pi_{ja}}{\pi_{1a}}\right) = \beta_{0j}$ when $x = 0$, indicating that they do not remember the last TV advertisement watched.

$\log\left(\frac{\pi_{jp}}{\pi_{1p}}\right) = \beta_{0j}$ when $x = 1$, indicating that the respondent remembers the last TV advertisement watched.

The log of the odds ratio can therefore be written as

$$\log(OR_j) = \log\left(\frac{\pi_{jp}}{\pi_{1p}}\right) - \log\left(\frac{\pi_{ja}}{\pi_{1a}}\right) = \beta_{1j}$$

Thus, $(OR_j) = \exp(\beta_{1j})$, which is estimated by $\exp(b_{1j})$. Moreover, if $\beta_{1j} = 0$, then $(OR_j) = 1$, which corresponds to the “remembrance” having no effect.

We can also calculate the 95% confidence intervals for OR_j . They are given by

$\exp[b_{1j} \pm 1.96 \times s.e. (b_{1j})]$, where $s.e. (b_{1j})$ is the standard error of b_{1j} .

Confidence intervals that which do not include 1 correspond to β values that are significantly different from zero.

3.6.1 Estimation of the regression coefficients

We estimate the regression coefficients β_j using the maximum likelihood principle. This principle proceeds by maximization of the multinomial likelihood in equation 3.33. The probabilities π_i are deemed to be functions of the coefficients. This method requires numerical procedures, and Fisher scoring or the Newton-Raphson methods often work well. These methods are typically iterative.

We consider independent random variables $Y_1, Y_2, Y_3, \dots, Y_N$ that satisfy the properties of generalized linear models as described in this chapter. We want to estimate the parameters $\boldsymbol{\beta}$ that are related to the variables $Y_1, Y_2, Y_3, \dots, Y_N$ through $E(Y_i) = \mu_i$ and $\mathbf{X}_i^T \boldsymbol{\beta}$.

For each Y_i , the log-likelihood function is

$$l_i = y_i b(\theta_i) + c(\theta_i) + d(y_i) \tag{3.28}$$

The functions b , c and d are as described in equation 3.5. Additionally,

$$E(Y_i) = \mu_i = -\frac{c'(\theta_i)}{b'(\theta_i)}; \text{Var}(Y_i) = \frac{[b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)]}{[b'(\theta_i)]^3}; \text{ and } g(\mu_i) = \mathbf{X}_i^T \boldsymbol{\beta} = \eta_i$$

Where X_i is a vector with elements $x_{ij}, j = 1, \dots, p$

We denote the log-likelihood function for all the Y_i 's as

$$l = \sum_{i=1}^N l_i = \sum y_i b(\theta_i) + \sum c(\theta_i) + \sum d(y_i)$$

To obtain the maximum likelihood estimator for the parameter β_j we obtain the equation below and proceed to differentiate using the chain rule by considering each right hand term separately

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^N \left[\frac{\partial l_i}{\partial \beta_j} \right] = \sum_{i=1}^N \left[\frac{\partial l_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \beta_j} \right] \quad 3.29$$

$$\frac{\partial l_i}{\partial \theta_i} = y_i b'(\theta_i) + c'(\theta_i) = b'(\theta_i)(y_i - \mu_i)$$

We differentiate equation 3.34 and substitute with the expression for $E(Y_i)$. Next

$$\frac{\partial \theta_i}{\partial \mu_i} = 1 / \frac{\partial \mu_i}{\partial \theta_i}$$

Differentiation of $E(Y_i)$ yields the following from the expression for $Var(Y_i)$,

$$\begin{aligned} \frac{\partial \mu_i}{\partial \theta_i} &= -\frac{c''(\theta_i)}{b'(\theta_i)} + \frac{c'(\theta_i)b''(\theta_i)}{[b'(\theta_i)]^2} \\ &= b' var(Y_i) \end{aligned}$$

From the expression for $g(\mu_i)$, we also have

$$\frac{\partial \mu_i}{\partial \beta_i} = \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_i} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij}$$

The score in equation 3.35 can thus be expressed as

$$U_j = \sum_{i=1}^N \left[\frac{(y_i - \mu_i)}{var(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right] \quad 3.30$$

The variance-covariance matrix for the scores, U_j 's has the terms

$$\mathfrak{F}_{jk} = E[U_j U_k]$$

This forms the information matrix \mathfrak{F} . From equation 3.29,

$$\begin{aligned} \mathfrak{F}_{jk} &= E \left\{ \sum_{i=1}^N \left[\frac{Y_i - \mu_i}{\text{var}(Y_i)} x_{ij} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right] \sum_{l=1}^N \left[\frac{Y_l - \mu_l}{\text{var}(Y_l)} x_{lk} \left(\frac{\partial \mu_l}{\partial \eta_l} \right) \right] \right\} \\ &= \sum_{i=1}^N \frac{E[(Y_i - \mu_i)^2] x_{ij} x_{ik}}{[\text{var}(Y_i)]^2} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \end{aligned} \quad 3.31$$

since $E[(Y_i - \mu_i)(Y_l - \mu_l)] = 0$ for $i \neq l$ as the Y_i 's are independent.

By using $E[(Y_i - \mu_i)]^2 = \text{var}(Y_i)$, we can simplify equation 3.30 into

$$\mathfrak{F}_{jk} = \sum_{i=1}^N \frac{x_{ij} x_{ik}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \quad 3.32$$

This equation is combined with the method of scoring equation which is generalized as

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} + [\mathfrak{F}^{(m-1)}]^{-1} \mathbf{U}^{(m-1)} \quad 3.33$$

Where $\mathbf{b}^{(m)}$ is the vector of estimates for the parameters $\beta_1, \beta_2, \beta_3, \dots, \beta_p$ at the m^{th} iteration.

$[\mathfrak{F}^{(m-1)}]^{-1}$ is the inverse of the information matrix with elements \mathfrak{F}_{jk} given in equation 3.31.

$\mathbf{U}^{(m-1)}$ is the vector of elements given by equation 3.29, all evaluated at $\mathbf{b}^{(m-1)}$.

Multiplying both sides of equation 3.32 by $\mathfrak{F}^{(m-1)}$ gives

$$\mathfrak{F}^{(m-1)} \mathbf{b}^{(m)} = \mathfrak{F}^{(m-1)} \mathbf{b}^{(m-1)} + \mathbf{U}^{(m-1)} \quad 3.34$$

Equation 3.38 makes it possible for \mathfrak{S} to be written as

$\mathfrak{S} = \mathbf{X}^T \mathbf{W} \mathbf{X}$, where \mathbf{W} is the $N \times N$ diagonal with elements

$$w_{ii} = \frac{1}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \quad 3.35$$

The right-hand side expression of equation 3.40 is the vector with elements

$\sum_{k=1}^p \sum_{i=1}^N \frac{x_{ij} x_{ik}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 b_k^{(m-1)} + \sum_{i=1}^N \frac{(Y_i - \mu_i) x_{ij}}{\text{var}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)$ evaluated at $\mathbf{b}^{(m-1)}$ following from equations 3.31 and 3.29

We can therefore write the right-hand side of equation 3.40 as $\mathbf{X}^T \mathbf{W} \mathbf{z}$, where \mathbf{z} has the elements

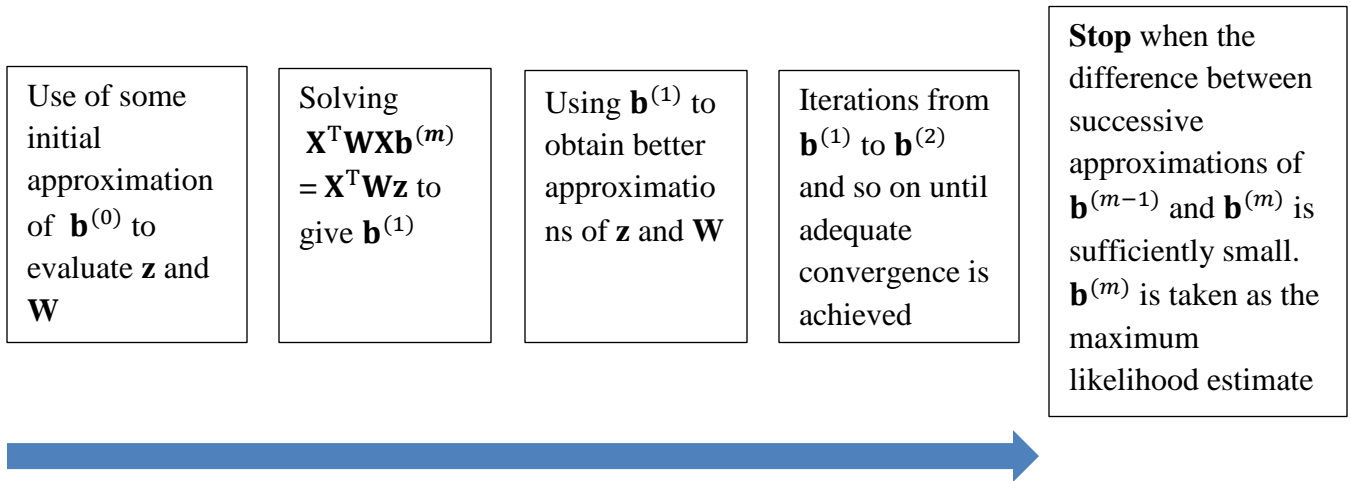
$$z_i = \sum_{k=1}^p x_{ik} b_k^{(m-1)} + (y_i - \mu_i) \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \quad 3.36$$

Both μ_i and $\frac{\partial \mu_i}{\partial \eta_i}$ are evaluated at $\mathbf{b}^{(m-1)}$. The iterative equation 3.33 can hence be written as

$$\mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{b}^{(m)} = \mathbf{X}^T \mathbf{W} \mathbf{z} \quad 3.37$$

This equation takes the same form as the normal equations for the ordinary least squares linear model obtained by weighted least squares. However, this has to be solved iteratively since in general \mathbf{z} and \mathbf{W} depend on \mathbf{b} . Charnes et al. (1976) conclude, based on above, that for generalized linear models, maximum likelihood estimators are obtained by an iterative weighted least squares procedure.

Most statistical packages, including R and SPSS which are used to model in this project, use an efficient algorithm for fitting generalized linear models logit iterative weighted least squares procedure based on equation 3.43. The general procedure that is used by the algorithms is



3.6.2 Independence of irrelevant alternatives

The multinomial Logistic model makes an assumption that is known as the independence of irrelevant alternatives (IIA) assumption. From the model in equation 3.27

$$\log\left(\frac{\pi_j}{\pi_1}\right) = \mathbf{X}_j^T \boldsymbol{\beta}_j, \text{ for } j = 2, \dots, J$$

it is evident that the log odds do not depend on other possible outcomes that are available. In this sense, these alternative outcomes are “irrelevant.” This implies that adding or deleting outcomes does not affect the odds among the remaining outcomes.

Long and Freese (2001), like several other publications, use the example of the red bus/blue bus to illustrate this assumption. Suppose that one has the choice of a red bus or a car to get to work

and that the odds of taking a red bus compared to a car are 1:1. IIA implies that the odds will remain 1:1 between these two alternatives even if a new blue bus company comes to town that is identical to the red bus company except for the color of the bus. Thus, the probability of driving a car can be made arbitrarily small by adding enough different colors of buses. More reasonably, one might expect that the odds of a red bus compared to a car would be reduced to 1:2 since half of those riding the red bus would be expected to ride the blue bus.

The IIA assumption is tested using two procedures. Hausman and McFadden (1984) proposed a Hausmantype test and McFadden, Tye, and Train (1976) proposed an approximate likelihood ratio test that was improved by Small and Hsiao (1985). For both the Hausman and the Small-Hsiao tests, multiple tests of IIA are possible. Assuming that the MNLM is estimated with base category b , $J - 1$ tests can be computed by excluding each of the remaining categories to form the restricted model. By changing the base category, a test can also be computed that excludes b . The results of the test tend to differ depending on the reference category that is used to estimate the model.

3.6.3 Statistical inference on regression coefficients and coefficient interpretation

The asymptotic sampling distribution of the estimators b_j of β_j is the normal distribution i.e.

$$\mathbf{z} = \frac{b_j - \beta_j}{s.e.(b_j)} \sim N(0,1)$$

We use this sampling distribution to construct the 95% confidence interval and test the hypothesis for the significance of β_j

The $100(1 - \alpha)\%$ confidence for β_j is given by

$$b_j \pm Z_{\alpha/2} s.e.(b_j)$$

To test the hypothesis

$$H_0: \beta_j = 0 \quad Vs. \quad H_1: \beta_j \neq 0$$

We can use either of the following test statistics

a. $Z = \frac{b_j}{s.e.(b_j)}$

b. $Z^2 = \frac{b_j^2}{var(b_j)} \sim \chi^2_{(1)}$

If the response variable under evaluation has J categories, we fit $J-1$ binary logistic models for each of the dummy variables. Each model explains the effect of the predictors on the probability of category $\pi_j, j = 2, 3, \dots, J$, in comparison to the reference category π_1 . Each of the $J-1$ models has its own intercept and regression coefficients.

The Wald test, derived from the sampling distribution of the maximum likelihood function,

$$(b - \beta)^T \mathfrak{I}(\mathbf{b})(\mathbf{b} - \beta) \sim \chi^2_{(p)} \quad 3.38$$

is used to evaluate whether or not a specific predictor variable is statistically significant in differentiating between the two categories in each of the underlying binary logistic comparisons. It is possible for a predictor variable with an overall relationship with the response variable to either be or not be statistically significant in differentiating between pairs of groups defined by the response variables.

Each of the models is interpreted individually. Each of the explanatory variables', x_i effect on the odds ratio is calculated by taking the exponent of the corresponding coefficient b_i . The effect is then interpreted by holding all other variables constant.

Prediction of the probabilities is computed as follows

For categories 2,3, ..., J used to fit binary logistic models

$$\hat{\pi}_1 = \frac{1}{1 + \sum_{j=2}^J X_j^T \mathbf{b}_j} \quad 3.39$$

And for the reference category:

$$\hat{\pi}_j = \frac{\exp(X_j^T \mathbf{b}_j)}{1 + \sum_{j=2}^J X_j^T \mathbf{b}_j}, \text{ for } j = 2, \dots, J \quad 3.40$$

3.7 Analysis Software

Data analysis has been carried out using MS Excel, R-GUI and SPSS. Data cleaning, formatting and basic descriptive statistics were done using MS Excel. Model fitting, diagnosis, and prediction were done using R-GUI and SPSS for comparison of results. Both R and SPSS are also used to carry out model significance tests.

ANALYSIS

In this chapter, we analyze the data, fit the best model and interpret the results. The data collected has 12 social-economic and TV watching habits variables, one of which will be the response variable. Data for 120 people is used to estimate the model.

4.1 Response variable

The response variable is classified into three categories based on the respondent's history on purchase of fast moving consumer goods as a result of watching the said good's advertisement on television. The descriptive statistics for the response variable are as below

Table 4.1 Descriptive statistics for the response variable

Variable	Categories	Count	Proportion
Purchase decision (n=114)	No	14	12%
	No, But I considered	60	53%
	Yes	40	35%

About half of the respondents have contemplated purchase of a fast moving consumer good in the past one year after watching the good's advertisement on television but did not do it. The model proposes to use the "No" category as the reference category. Multinomial logistic regression provides an effective and reliable way to obtain the estimated probability of belonging to a specific population (e.g. Buying a product based on its television advertisement) and the estimate of odds ratio of buying or considering to buy compared to not buying at all as a result of

various reasons (Peng, Lee, & Ingersoll, 2002; Peng, Manz, & Keck, 2001; Scott, Mason, & Chapman, 1999). Furthermore, using the multinomial logistic regression procedure yields estimates of the net effects of a set of explanatory variables on the dependent variable can be obtained (Morgan and Teachman,1988). In this case, a set of regression coefficients for social demographic and television watching habits variables will be used to obtain the odds ratio and predicted probabilities.

4.2 Predictor Variables

Table 4.2 Descriptive statistics for predictor variables

Variable	Categories	Count	Proportion
Gender (n=120)	Female	36	30%
	Male	84	70%
Education (n=120)	Tertiary (Degree/Diploma)	82	68%
	Primary/Secondary	38	32%
Marital Status (n=120)	Married	42	35%
	Separated/Divorced	2	2%
	Single	76	63%
Monthly Income (n=120)	Low	70	58%
	Medium	50	42%
Presence of TV in household (n=120)	No	6	5%
	Yes	114	95%
TV Hours per day (n=114)	2 - 4 hours	74	65%
	Less than 1 hour	18	16%
	More than 4 hours	22	19%
Main TV interest (n=114)	International News and Programs	32	28%
	Live events	20	18%
	Local News and Programs	62	54%
Last TV Ad Remembrance (n=114)	No	6	5%
	Yes	108	95%
Key TV Ad feature (n=114)	Entertaining	40	35%
	Short and precise	18	16%

	Believable and Relatable	56	49%
Purchased item living up to standards (n=40)	No	18	45%
	Yes	22	55%
Future Purchase Consideration (n=114)	No	46	40%
	Yes	68	60%

From the administered survey, respondents were, in addition to purchase habits, asked for their social and demographic characteristics. 30% of the respondents were female. 68% of the respondents have tertiary education, while 114 out of the 120 respondents have access to a television set in their household, a key variable as this is one of the basis upon which the model is built – a person must have access to a television set to have visibility of advertisements. Additionally, an underlying caveat is that a person who makes a decision to purchase or has been in consideration must remember the last advertisement they watched – this variable is thus not used in the model but used as an indicator of what variable should be considered.

The Kenya National Bureau of Statistics in 2010 revised the income groups for both Urban and Rural settings in Kenya. These levels are currently as follows:

- Lower income group: Households spending KSh. 23,670 or less per month (they constitute 72.12% of the households).
- Middle-income group: Households spending between KSh. 23,671 up to and including KSh. 120,000 per month (they constitute 24.12% of the households).

- Upper income group: Households spending above KSh. 120,000 per month (they constitute 3.76% of the households).

Respondents in the survey were asked for the income range and grouped into the corresponding groups. 58% of the respondents earn and spend less than 23,670 shillings per month at household level, with the rest spending between 23,671 and 120,000 per month.

Due to the versatility of the advertising industry, one key assumption in regards to television advertising is made, leaving room for further model enhancement. The study and model assumes that all television channels that respondents are exposed to have the same reach and effect, and do not influence the respondent’s perception and subsequent decision to purchase or otherwise.

In order to achieve optimal results for a multinomial logistic model, it is paramount to evaluate the matrix/contingency table obtained from the response variable and each of the predictor variables, and ensure that none of the values is less than two. Furthermore, the number of cases per predictor variable is required to be greater than 15 for model optimization. In this case, due to the limited number of cases, direct entry with evaluation of the most significant variables and the impact they have on the overall significance of the model is done. This narrows the model to 6 predictor variables, which implies about 20 cases per variable.

Table 4.3 Contingency tables of proposed entry predictor variables and the response variable

	Gender	
Purchase Decision	Female	Male

No	6	14
No, But I considered	12	48
Yes	18	22
Education		
Purchase Decision	Primary/Secondary	Tertiary (Degree/Diploma)
No	8	12
No, But I considered	14	46
Yes	16	24
Income Level		
Purchase Decision	Low	Medium
No	14	6
No, But I considered	38	22
Yes	18	22

TV hours per day			
Purchase Decision	2 - 4 hours	Less than 1 hour	More than 4 hours
No	6	4	4
No, But I considered	46	6	8
Yes	22	8	10
Main TV interest			
Purchase Decision	International News and Programs	Live events	Local News and Programs
No	4	4	6
No, But I considered	16	8	36
Yes	12	8	20
Key TV advertisement Feature			
Purchase Decision	Entertaining	Short and precise	Believable and Relatable
No	6	4	4
No, But I considered	20	8	32
Yes	14	6	20

Subsequently, assessment of the overall model fit is done with Significance test of the model log likelihood (Change in -2LL), Measures analogous to R^2 (Cox and Snell R^2 , Nagelkerke R^2), classification matrices as a measure of model accuracy, and Numerical Problems checks.

4.3 Parameter Estimation and model fit tests

Table 4.4 Model Fitting Information

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2 Log Likelihood	Chi-Square	Df	Sig.
Intercept Only	194.778	200.251	190.778			
Final	199.496	254.220	159.496	31.282	18	.027

The initial log likelihood value (190.778) is a measure of a model with no independent variables, i.e. only a constant or intercept. The final log likelihood value (159.496) is the measure computed after all of the independent variables have been entered into the logistic regression.

The difference between these two measures is the model chi-square value that is tested for statistical significance. This test is analogous to the F-test for R^2 or change in R^2 value in multiple regression which tests whether or not the improvement in the model associated with the additional variables is statistically significant. In this problem the model Chi-Square value of 31.282 is also significant at 0.05 since 0.027 < 0.05. We conclude that there is a significant relationship between the dependent variable and the set of independent variables. Still, the sample used for model construction is somewhat low, leading to a high probability of the design affording low power for detecting effects.

Table 4.5 Goodness-of-Fit

	Chi-Square	Df	Sig.
Pearson	140.976	72	.000
Deviance	145.243	72	.000

Table 4.6 Pseudo R-Square

Cox and Snell	.240
Nagelkerke	.281
McFadden	.142

The Cox and Snell R^2 measure operates like R^2 , with higher values indicating greater model fit. However, this measure is limited in that it cannot reach the maximum value of 1, so Nagelkerke proposed a modification that had the range from 0 to 1. We rely upon Nagelkerke's measure as indicating the strength of the relationship. There is a relationship but not very strong based on the Nagelkerke R^2

Table 4.7 Model Classification – Observed Vs. Predicted

Observed	Predicted			
	No	No, But I considered	Yes	Percent Correct
No	2	6	6	14.3%
No, But I considered	2	50	8	83.3%
Yes	0	20	20	50.0%
Overall Percentage	3.5%	66.7%	29.8%	63.2%

The classification matrix in logistic regression serves the same function as the classification matrix in Multinomial Logistic Regression, i.e. evaluating the accuracy of the model.

If the predicted and actual group memberships are the same, i.e. No and No, Consideration and Consideration, or Yes and Yes, then the prediction is accurate for that case. If predicted group membership and actual group membership are different, the model "misses" for that case. The overall percentage of accurate predictions, 63.2% in this case, is the measure of a model that is relied on most heavily for analysis as well as for Multinomial Logistic Regression because it has a meaning that is readily communicated, i.e. the percentage of cases for which the model predicts accurately.

Table 4.8 Parameter estimation

PURCHASED ^a		B	Std. Error	Wald	df	Sig	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
No, But I considered	Intercept	-1.083	1.211	.799	1	.371			
	[Gender=Female]	-.884	.775	1.302	1	.254	.413	.090	1.886
	[Gender=Male]	0 ^b	.	.	0
	[Education=Primary/Secondary]	.304	.957	.101	1	.751	1.355	.208	8.847
	[Education=Tertiary (Degree/Diploma)]	0 ^b	.	.	0
	[Monthly_Income=Low]	-.506	.829	.373	1	.542	.603	.119	3.063
	[Monthly_Income=Medium]	0 ^b	.	.	0
	[TV_hoursDay=2 - 4 hours]	2.637	1.127	5.469	1	.019	13.969	1.533	127.309
	[TV_hoursDay=Less than 1 hour]	.736	1.122	.430	1	.512	2.087	.231	18.831

	[TV_hoursDay=More than 4 hours]	0 ^b	.	.	0
	[Main_TV_interest=International News and Programs]	.062	.791	.006	1	.938	1.064	.226	5.018
	[Main_TV_interest=Lively events]	-1.806	.960	3.537	1	.060	.164	.025	1.079
	[Main_TV_interest=Local News and Programs]	0 ^b	.	.	0
	[Key_Feature=Believable and Relatable]	2.687	1.059	6.440	1	.011	14.689	1.844	117.036
	[Key_Feature=Entertaining]	1.506	1.085	1.928	1	.165	4.511	.538	37.820
	[Key_Feature=Short and precise]	0 ^b	.	.	0
Yes	Intercept	-.886	1.251	.502	1	.479			
	[Gender=Female]	.219	.771	.081	1	.777	1.245	.274	5.647
	[Gender=Male]	0 ^b	.	.	0
	[Education=Primary/Secondary]	1.076	.953	1.274	1	.259	2.932	.453	18.981
	[Education=Tertiary (Degree/Diploma)]	0 ^b	.	.	0
	[Monthly_Income=Low]	-1.096	.819	1.789	1	.181	.334	.067	1.665
	[Monthly_Income=Medium]	0 ^b	.	.	0
	[TV_hoursDay=2 - 4 hours]	1.852	1.113	2.766	1	.096	6.370	.719	56.471
	[TV_hoursDay=Less than 1 hour]	.734	1.077	.465	1	.495	2.084	.252	17.210
	[TV_hoursDay=More than 4 hours]	0 ^b	.	.	0

than 4 hours]									
[Main_TV_interest=International News and Programs]	.374	.808	.214	1	.644	1.453	.298	7.075	
[Main_TV_interest=Lively events]	-.740	.937	.623	1	.430	.477	.076	2.997	
[Main_TV_interest=Local News and Programs]	0 ^b	.	.	0	
[Key_Feature=Believable and Relatable]	1.786	1.091	2.681	1	.102	5.967	.703	50.627	
[Key_Feature=Entertaining]	1.059	1.146	.855	1	.355	2.884	.305	27.242	
[Key_Feature=Short and precise]	0 ^b	.	.	0	

There are many numerical problems that could occur in multinomial logistic regression that are not detected by the statistical package used. These include multi-collinearity among the independent variables, zero cells for a dummy-coded independent variable because all of the subjects have the same value for the variable, and "complete separation" whereby the two groups in the dependent event variable can be perfectly separated by scores on one of the independent variables. All of these problems produce large standard errors (over 2) for the variables included in the analysis and very often produce very large B coefficients as well. If we encounter large standard errors for the predictor variables, we should examine frequency tables, one-way ANOVAs, and correlations for the variables involved to try to identify the source of the problem.

None of the standard errors or B coefficients in the fit model is excessively large, so there is no evidence of a numeric problem with this analysis.

4.4 Interpretation and prediction

The response category “No”, containing those that have not purchased a fast moving good as a result of seeing it being advertised on television, is set as the reference category. We first consider the odds of purchasing compared to not purchasing a good based on its television advertisement at all and interpret, then consider the odds of purchasing relative to not purchasing at all. Each predictor variable’s effect on the response’s odds ratio is interpreted while holding the rest of the variables constant. We then consider a consumer with an assumed set of characteristics and estimate the probability of near conversion or total conversion.

4.4.1 Consider purchasing Vs Not purchasing

Gender (male=0, Female=1)

$b = -0.884$, $\exp(b) = 0.413$; Females are 59% less likely to consider buying a product over not buying it at all based on its television advertisement compared to males.

Education (Tertiary=0, Primary/Secondary=1)

$b = 0.304$, $\exp(b) = 1.355$; Persons with Primary/Secondary education are 36% more likely to consider buying a product over not buying it at all based on its television advertisement compared to those with tertiary education.

Monthly Income (High=0, Medium=1)

$b = -0.506$, $\exp(b) = 0.603$; Persons with low income levels are 40% less likely to consider buying a product over not buying it entirely based on its television advertisement compared to persons with medium monthly income levels

TV hours per day (more than 4 hours=0, {Less than one hour, 2-4 hours} =1)

$b_i = 0.736, \exp(b_i) = 2.087$; Persons who watch less that hour of television everyday are two times more likely to consider purchasing a product over not buying it all based on its television advertisement compared to those that watch television for more than 4 hours a day.

$b_{ii} = 2.637, \exp(b_{ii}) = 13.969$; Persons who watch between two and four hours of television everyday are 13 times more likely to consider purchasing a product over not buying it all based on its television advertisement compared to those that watch television for more than 4 hours a day.

Main Television interest (Locals news/Programs=0, {Live events, International news/Programs} =1)

$b_i = -1.806, \exp(b_i) = 0.164$; Those who are most interested in live events on television are 84% less likely to consider purchasing a product over not purchasing it based on its television advertisement compared to those whose preference is local news and programming.

$b_{ii} = 0.062, \exp(b_{ii}) = 1.064$; On the other hand, those who are most interested in international news and programming television are 6% more likely to consider purchasing a product over not purchasing it based on its television advertisement compared to those whose preference is local news and programming.

Key feature of a television advertisement (Short and precise=0, {Entertaining, Believable and relatable} =1)

$b_i = 1.506, \exp(b_i) = 4.511$; People who look out for the entertainment aspect of a television advertisement and regard it to be the key element are almost 5 times more likely to consider purchase the advertised product over not purchasing it at all compared to those that regard the length and precision of the advertisement.

$b_{ii} = 2.687, \exp(b_{ii}) = 14.689$; Additionally, those that regard the believability and relatability of a good's television advertisement as the most important feature are about 15 times more likely

to consider purchasing as opposed to not purchasing the good entirely compared to those that value the length and precision of the advertisement. Believability and relatability of an advertisement, based on this, emerges as prime in getting people to budge and even consider buying a good based on its advertisement.

4.4.2 Purchasing Vs Not purchasing

Gender (male=0, Female=1)

$b = 0.219$, $\exp(b) = 1.245$; Females are 25% more likely to buy an advertised fast moving good based on its television advertisement compared to males.

Education (Tertiary=0, Primary/Secondary=1)

$b = 1.076$, $\exp(b) = 2.932$; Persons with Primary/Secondary education are 3 times more likely to buy an advertised fast moving good over not buying it at all based on its television advertisement compared to those with tertiary education.

Monthly Income (High=0, Medium=1)

$b = -1.096$, $\exp(b) = 0.334$; Persons with low monthly income levels are 67% less likely to purchase an advertised fast moving good over not buying it entirely based on its television advertisement compared to persons with medium monthly income levels

TV hours per day (more than 4 hours=0, {Less than one hour, 2-4 hours} =1)

$b_i = 0.734$, $\exp(b_i) = 2.084$; Persons who watch less than one hour of television everyday are approximately two times more likely purchase an advertised fast moving good over not buying it all based on its television advertisement compared to those that watch television for more than 4 hours a day.

$b_{ii} = 1.852$, $\exp(b_{ii}) = 6.37$; Persons who watch between two and four hours of television everyday are 6 times more likely to purchase an advertised fast moving good over not buying it

all based on its television advertisement compared to those that watch television for more than 4 hours a day.

Main Television interest (Locals news/Programs=0, {Live events, International news/Programs} =1)

$b_i = -0.74, \exp(b_i) = 0.477$; Those who are most interested in live events on television are about 50% less likely to purchase an advertised fast moving good over not purchasing it based on its television advertisement compared to those whose main preference is local news and programming.

$b_{ii} = 0.374, \exp(b_{ii}) = 1.453$; On the contrary, those who are most interested in international news and programming television are 50% more likely to purchase the advertised fast moving good over not purchasing it based on its television advertisement compared to those whose preference is local news and programming.

Key feature of a television advertisement (Short and precise=0, {Entertaining, Believable and relatable} =1)

$b_i = 1.059, \exp(b_i) = 2.884$; People who value the entertainment aspect of a television advertisement and regard it to be the key element are approximately 3 times more likely to purchase the advertised fast moving good over not purchasing it at all compared to those that regard the length and precision of the advertisement.

$b_{ii} = 1.786, \exp(b_{ii}) = 5.967$; Additionally, those that regard the believability and relatability of a fast moving good's television advertisement as the most important feature are about 6 times more likely to consider purchasing as opposed to not purchasing the good entirely compared to those that value the length and precision of the advertisement. Believability and relatability of an advertisement, based on this, again emerges as prime in getting people to purchase the advertised good.

Based on equations 3.45 and 3.46, we can calculate the probabilities of each response category given a set of predictor variable categories in which a consumer falls.

For the Purchase consideration and Purchase categories, this is given by $\hat{\pi}_1 = \frac{1}{1 + \sum_{j=2}^J X_j^T \mathbf{b}_j}$ while

the probability for the reference category (No) is given by $\hat{\pi}_j = \frac{\exp(X_j^T \mathbf{b}_j)}{1 + \sum_{j=2}^J X_j^T \mathbf{b}_j}$, for $j = 2, \dots, J$.

We consider a female consumer with access to a television in the household and remembers the last fast moving consumer good advertisement she watched on television. The consumer has gone up to university and earns a medium level monthly income. She is employed and is able to watch television for about two and a half hours every day, with her main preference being local news and programming. She treasures the ability of a television advertisement to be believed and directly relatable to her day to day life. We calculate the probability of converting her from a viewer of an advertisement for any particular fast moving consumer good into a buyer of the product.

$$p(y = buy) = \frac{\exp(X_j^T \mathbf{b}_j)}{1 + \sum_{j=2}^J X_j^T \mathbf{b}_j}, \text{ for } j = 3 = \frac{\exp(\beta_{0j} + \beta_{1j} + \beta_{2j} + \beta_{3j} + \beta_{4j} + \beta_{5j} + \beta_{6j})}{1 + \sum_{j=2}^J \exp(\beta_{0j} + \beta_{1j} + \beta_{2j} + \beta_{3j} + \beta_{4j} + \beta_{5j} + \beta_{6j})};$$

$$\begin{aligned} &= \exp(-0.886 + 0.219(1) + 1.076(0) + -1.096(0) + 1.852(1) + 0.374(0) + 1.786(1)) / (1 + \\ &\exp(-0.886 + 0.219(1) + 1.076(0) + -1.096(0) + 1.852(1) + 0.374(0) + 1.786(1)) + \\ &\exp(-1.083 + 0.884(1) + 0.304(0) + -0.506(0) + 2.637(1) + 0.062(0) + 2.687(1))) \end{aligned}$$

= 0.40 There is a 40% chance of converting her from a viewer to a buyer.

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusions and Recommendations

The six response variables used in the model can be grouped into two – Social economic (Gender, education level, Income level) and television watching habits/variables (Tv hours per day, main television interest, most important television advertisement feature). Overall these variables provide a good fit based on all statistical evaluation criteria. Some of the individual variables at sub-model level are not significant and this is mostly attributed to the associated small cell counts.

Males are more likely to just consider purchasing a product but not go ahead with the purchase as opposed to women, who are more likely to be converted from advertisement viewers to buyers. Furthermore, holding other variables constant, persons with Secondary education are more likely to be converted from viewers to buyers compared to those with tertiary education. This could be a case of the more educated being more cautious and critical in the face of decisions regarding the purchase decisions. Controlling for all other variables, medium income earners are more likely to consider purchase of advertised goods or even actually purchasing compared to low income earners. With the middle class in Kenya growing and spurring the growth of various consumer goods industries, this is a key aspect that would measurably target the middle income earners and subsequently increase the conversion rates of advertisements.

People who watch television for an average of between two and four hours a day are also more likely to be converted from viewers to buyers of advertised fast moving consumer goods

compared to those that watch too little (Less than two hours) or too much (more than 4 hours). Further, consumers that are more interested in international programming and events as well as local programming are more likely to be converted compared to those that prefer live events on television.

It is evident based on the results that a believable and relatable television advertisement compared to a short/precise one or even an entertaining one is more likely to provoke a viewer of a good's advertisement to consider purchasing it or actually purchasing it, holding all other variables constant. A believable advertisement is momentous.

Bearing the effect of the social economic and television watching habits variables used in the model on the conversion of viewers to buyers, it is perceptible that it is statistically feasible to measure the conversion probabilities of advertisement and make more effective decisions based on the target audience. Advertisers should thus consider utilization of models such as this in addition to the conventional advertisement testing techniques to achieve more precision and rigor in the various advertisement campaigns. This in turn reaps them more benefit for the advertising money spent to buy space on television.

5.2 Limitations and future research

The results from this model and other researchers' experiences indicate some limitations of the multinomial logistic model fit that pose challenges:

- The residuals cannot be normally distributed (OLS assumption)

- Losing information, data, and power – predictor and sometime even the response categories have to be lumped together especially when the sample is not large enough
- The coding is completely arbitrary i.e. recoding the dependent variable can give very different results.

Moreover, due to the fact data used to generate and test this model is based on only 120 participants, there are limitations on the response variables that can be used. Further research using a bigger pool of responses as well as inclusion of other variables that are assumed in this study is required to have a fully industry and academic certified model. This just forms the premise.

REFERENCES

Long, J. S. and Freese, J. (2006) *Regression Models for Categorical and Limited Dependent Variables Using Stata*, Second Edition. Stata Press.

Hosmer, D. and Lemeshow, S. (2000) *Applied Logistic Regression (Second Edition)*. John Wiley & Sons, Inc.

Agresti, A. (1996) *An Introduction to Categorical Data Analysis*. New York: John Wiley & Sons, Inc.

Chandler-Pepelnjak, J. (2010) *Modeling Conversions in Online Advertisement*. The University of Montana.

Peng, Chao-Ying J. and Nichols, Rebecca N. (2003) Using Multinomial Logistic Models to Predict Adolescent Behavioral Risk. *Journal of Modern Applied Statistical Methods* Spring 2003, Vol. 2, No 1. Indiana University-Bloomington.

Frees, Edward W. (2004) *Longitudinal and Panel Data: Analysis and Applications for the Social Sciences*. Cambridge University Press.

Lewis, Randall A. and Rao, Justin M. (2013) On the Near Impossibility of Measuring the Returns to Advertising. Yahoo! Research

Menard, S. (1995). *Applied logistic regression analysis*. Sage.

Lewis, R., Rao, J. M., and Reiley, D. (2013). chapter Measuring the Effects of Advertising: The Digital Frontier. NBER Press.

Lynch, Scott M. (2013) Using Statistics in Social Research: A Concise Approach. Springer-Verlag New York

Hoffmann, John P. (2010) Linear Regression Analysis: Applications and Assumptions Second Edition. Brigham Young University

Hoffmann, John P. (2003). Generalized linear models: An applied approach. Allyn & Bacon.

Andreß, Hans-Jürgen et al (2013) Applied Panel Data Analysis for Economic and Social Surveys. Springer-Verlag New York

Thorsén, E. (2014) Multinomial and Dirichlet-multinomial modeling of categorical time series. Stockholm University

N.E. Breslow (1996) Generalized Linear Models: Checking Assumptions and Strengthening Conclusions. *Statistica Applicata* 8, 23-41

Annette J. Dobson (2002) An Introduction to Generalized Linear Models, Second Edition. Chapman and Hall

Pierce, D. A., and Schafer, D. W. (1986). Residuals in generalized linear models. *Journal of the American Statistical Association*.

Garson, G. D. (2011). "Logistic Regression", from Statnotes: Topics in Multivariate Analysis.
<http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm>.

Scott, J. and Freese, J. (2001) Regression Models For Categorical Dependent Variables Using STATA. Stata Press

Faisal M. Zahid and Gerhard Tutz (2010) Multinomial Logit Models with Implicit Variable Selection. *Advances in Data Analysis and Classification*. Springer-Verlag New York, Inc.

Predicting How TV Advertising Drives Sales A New ARF Research Project (2013).

<http://mobdev.thearf.com/images/uploads/Predicting%20How%20TV%20Advertising%20Drives%20Sales%20May%202013>. ARF.

Little, John D. C. (1979) Aggregate Advertising Response Models: The State of The Art. Sloan School of Management , M.I.T.

Andreas Diekmann and Ben Jann (2008) Regression Models for Categorical Dependent Variables (Logit, Probit, and Related Techniques). ZA Spring Seminar 2008

Pedhazur, E. J. (1997). Multiple regression in behavioral research: Explanation and prediction (3rd ed.). Harcourt Brace.

Schwab, J. A. (2002). Multinomial logistic regression: Basic relationships and complete problems. <http://www.utexas.edu/courses/schwab/sw388r7/SolvingProblems/>

Fox, J. (1984). Linear statistical models and related methods: With applications to social research. Wiley.

Wei, B. (1998) Exponential Family Nonlinear Models. Springer-Verlag Sinpagore.

Kenya National Bureau of Statistics (KNBS) (2010). The New Consumer Price Index (CPI) Users' Guide.

http://www.knbs.or.ke/index.php?option=com_phocadownload&view=category&download=370:consumer-price-index-cpi-users-guide&id=8:consumer-price-indices-cpi&Itemid=562

Appendix

Collected Data

I d	Ge n d e r	A g e	Education	Marita l_s t a t u s	Month ly_Inc o m e	TV_P r e s e n c e	TV_h o u r s / D a y	Main_TV_i n t e r e s t	R e m e m b r a n c e	Key_F e a t u r e	PURC H A S E D	Living _u p _t o _A d	Future_ c o n s i d e r a t i o n
1	Male	30	Primary/S e c o n d a r y	Married	Mediu m	Yes	More than 4 hours	Local News and Programs	Yes	Entertaining	Yes	Yes	Yes
2	Female	36	Primary/S e c o n d a r y	Married	Mediu m	Yes	More than 4 hours	Local News and Programs	Yes	Believable and Relatable	Yes	No	No
3	Female	31	Primary/S e c o n d a r y	Married	Mediu m	Yes	More than 4 hours	Local News and Programs	Yes	Entertaining	Yes	No	No
4	Male	25	Tertiary (Degree/D i p l o m a)	Single	Mediu m	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	No, But I consider ed	-	No
5	Male	27	Tertiary (Degree/D i p l o m a)	Married	Mediu m	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	No, But I consider ed	-	Yes
6	Male	35	Primary/S e c o n d a r y	Married	Mediu m	Yes	More than 4 hours	Local News and Programs	Yes	Believable and Relatable	No, But I consider ed	-	Yes
7	Female	37	Tertiary (Degree/D i p l o m a)	Married	Mediu m	Yes	Less than 1 hour	Live events	No	Believable and Relatable	No	-	No
8	Male	31	Tertiary (Degree/D i p l o m a)	Married	Mediu m	Yes	More than 4 hours	Live events	Yes	Entertaining	No, But I consider ed	-	Yes
9	Male	26	Primary/S e c o n d a r y	Married	Mediu m	Yes	More than 4 hours	International News and Programs	Yes	Entertaining	No	-	Yes
10	Male	24	Primary/S e c o n d a r y	Single	Mediu m	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	No, But I consider ed	-	Yes
11	Female	27	Tertiary (Degree/D i p l o m a)	Married	Mediu m	Yes	2 - 4 hours	Local News and Programs	Yes	Short and precise	No, But I consider ed	-	Yes
12	Male	31	Tertiary (Degree/D i p l o m a)	Married	Mediu m	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	Yes	Yes	Yes
13	Male	26	Tertiary (Degree/D i p l o m a)	Married	Mediu m	Yes	More than 4 hours	International News and Programs	Yes	Believable and Relatable	Yes	No	No
14	Male	34	Tertiary (Degree/D i p l o m a)	Married	Mediu m	Yes	More than 4 hours	Local News and Programs	No	Believable and Relatable	No	-	Yes
15	Male	33	Tertiary	Married	Mediu m	Yes	2 - 4	International	Yes	Believable	No, But	-	Yes

5	ale	0	(Degree/Diploma)		m		hours	News and Programs		e and Relatable	I considered		
16	Male	26	Tertiary (Degree/Diploma)	Married	Medium	Yes	Less than 1 hour	Local News and Programs	Yes	Entertaining	No, But I considered	-	Yes
17	Female	40	Primary/Secondary	Single	Low	Yes	Less than 1 hour	Local News and Programs	Yes	Believable and Relatable	Yes	Yes	Yes
18	Female	18	Primary/Secondary	Single	Low	Yes	2 - 4 hours	International News and Programs	Yes	Entertaining	Yes	Yes	Yes
19	Female	31	Primary/Secondary	Married	Medium	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	Yes	Yes	Yes
20	Male	36	Primary/Secondary	Married	Low	No			No		No		
21	Male	25	Primary/Secondary	Separated/Divorced	Low	Yes	2 - 4 hours	Live events	Yes	Believable and Relatable	No, But I considered	-	No
22	Male	23	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	No, But I considered	-	No
23	Male	29	Tertiary (Degree/Diploma)	Married	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	No, But I considered	-	No
24	Female	25	Tertiary (Degree/Diploma)	Single	Medium	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	No, But I considered	-	No
25	Male	25	Tertiary (Degree/Diploma)	Single	Medium	Yes	2 - 4 hours	International News and Programs	Yes	Short and precise	No, But I considered	-	No
26	Male	21	Tertiary (Degree/Diploma)	Single	Medium	Yes	More than 4 hours	Local News and Programs	Yes	Believable and Relatable	No, But I considered	-	No
27	Male	25	Primary/Secondary	Single	Low	Yes	2 - 4 hours	International News and Programs	Yes	Entertaining	No, But I considered	-	Yes
28	Female	25	Tertiary (Degree/Diploma)	Married	Low	Yes	2 - 4 hours	International News and Programs	Yes	Short and precise	No	-	No
29	Male	23	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Live events	Yes	Believable and Relatable	No, But I considered	-	Yes
30	Male	31	Tertiary (Degree/Diploma)	Married	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	No, But I considered	-	Yes
31	Female	33	Tertiary (Degree/Diploma)	Married	Low	Yes	More than 4 hours	International News and Programs	Yes	Believable and Relatable	No, But I considered	-	Yes
3	M	3	Primary/S	Married	Low	Yes	2 - 4	International	Yes	Believable	No, But	-	Yes

2	ale	0	Secondary				hours	News and Programs		e and Relatable	I considered		
33	Male	24	Primary/Secondary	Single	Low	No			No		No		
34	Male	27	Primary/Secondary	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	No, But I considered	-	Yes
35	Male	25	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Live events	Yes	Entertaining	Yes	Yes	Yes
36	Male	21	Tertiary (Degree/Diploma)	Single	Low	Yes	Less than 1 hour	International News and Programs	Yes	Entertaining	Yes	Yes	Yes
37	Male	24	Tertiary (Degree/Diploma)	Single	Medium	Yes	2 - 4 hours	Live events	Yes	Believable and Relatable	Yes	Yes	Yes
38	Male	23	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	International News and Programs	Yes	Short and precise	Yes	Yes	Yes
39	Female	26	Tertiary (Degree/Diploma)	Single	Medium	Yes	2 - 4 hours	Live events	Yes	Believable and Relatable	Yes	Yes	Yes
40	Male	23	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	Yes	Yes	Yes
41	Male	26	Primary/Secondary	Single	Medium	Yes	2 - 4 hours	Live events	Yes	Believable and Relatable	Yes	No	No
42	Female	22	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	Yes	No	No
43	Male	27	Tertiary (Degree/Diploma)	Single	Medium	Yes	More than 4 hours	Local News and Programs	Yes	Entertaining	Yes	No	No
44	Female	24	Primary/Secondary	Single	Low	Yes	Less than 1 hour	International News and Programs	Yes	Believable and Relatable	Yes	No	No
45	Female	26	Tertiary (Degree/Diploma)	Single	Medium	Yes	Less than 1 hour	International News and Programs	Yes	Short and precise	Yes	No	No
46	Male	24	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Short and precise	Yes	No	No
47	Male	29	Primary/Secondary	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	No, But I considered	-	No
48	Male	22	Primary/Secondary	Single	Low	No			No		No		
49	Male	25	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Short and precise	No, But I considered	-	No
50	Male	24	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	International News and Programs	Yes	Entertaining	No, But I considered	-	No
51	Female	25	Tertiary (Degree/Diploma)	Single	Low	Yes	Less than 1 hour	Live events	Yes	Entertaining	No	-	No

52	Male	22	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Live events	Yes	Believable and Reliable	No, But I considered	-	Yes
53	Female	20	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Short and precise	No, But I considered	-	Yes
54	Female	23	Tertiary (Degree/Diploma)	Single	Low	Yes	Less than 1 hour	Local News and Programs	Yes	Believable and Reliable	No, But I considered	-	Yes
55	Female	24	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Reliable	No, But I considered	-	Yes
56	Male	21	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	No	-	No
57	Male	22	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	International News and Programs	Yes	Entertaining	No, But I considered	-	Yes
58	Male	23	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	No	Short and precise	No	-	Yes
59	Male	23	Tertiary (Degree/Diploma)	Single	Low	Yes	Less than 1 hour	International News and Programs	Yes	Believable and Reliable	No, But I considered	-	Yes
60	Male	24	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Reliable	No, But I considered	-	Yes
61	Male	30	Primary/Secondary	Married	Medium	Yes	More than 4 hours	Local News and Programs	Yes	Entertaining	Yes	Yes	Yes
62	Female	34	Primary/Secondary	Married	Medium	Yes	More than 4 hours	Local News and Programs	Yes	Believable and Reliable	Yes	No	No
63	Female	32	Primary/Secondary	Married	Medium	Yes	More than 4 hours	Local News and Programs	Yes	Entertaining	Yes	No	No
64	Male	25	Tertiary (Degree/Diploma)	Single	Medium	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	No, But I considered	-	No
65	Male	27	Tertiary (Degree/Diploma)	Married	Medium	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Reliable	No, But I considered	-	Yes
66	Male	26	Primary/Secondary	Married	Medium	Yes	More than 4 hours	Local News and Programs	Yes	Believable and Reliable	No, But I considered	-	Yes
67	Female	26	Tertiary (Degree/Diploma)	Married	Medium	Yes	Less than 1 hour	Live events	No	Believable and Reliable	No	-	No
68	Male	35	Tertiary (Degree/Diploma)	Married	Medium	Yes	More than 4 hours	Live events	Yes	Entertaining	No, But I considered	-	Yes

69	Male	40	Primary/Secondary	Married	Medium	Yes	More than 4 hours	International News and Programs	Yes	Entertaining	No	-	Yes
70	Male	24	Primary/Secondary	Single	Medium	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	No, But I considered	-	Yes
71	Female	40	Tertiary (Degree/Diploma)	Married	Medium	Yes	2 - 4 hours	Local News and Programs	Yes	Short and precise	No, But I considered	-	Yes
72	Male	33	Tertiary (Degree/Diploma)	Married	Medium	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	Yes	Yes	Yes
73	Male	26	Tertiary (Degree/Diploma)	Married	Medium	Yes	More than 4 hours	International News and Programs	Yes	Believable and Relatable	Yes	No	No
74	Male	39	Tertiary (Degree/Diploma)	Married	Medium	Yes	More than 4 hours	Local News and Programs	No	Believable and Relatable	No	-	Yes
75	Male	31	Tertiary (Degree/Diploma)	Married	Medium	Yes	2 - 4 hours	International News and Programs	Yes	Believable and Relatable	No, But I considered	-	Yes
76	Male	26	Tertiary (Degree/Diploma)	Married	Medium	Yes	Less than 1 hour	Local News and Programs	Yes	Entertaining	No, But I considered	-	Yes
77	Female	40	Primary/Secondary	Single	Low	Yes	Less than 1 hour	Local News and Programs	Yes	Believable and Relatable	Yes	Yes	Yes
78	Female	18	Primary/Secondary	Single	Low	Yes	2 - 4 hours	International News and Programs	Yes	Entertaining	Yes	Yes	Yes
79	Female	31	Primary/Secondary	Married	Medium	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	Yes	Yes	Yes
80	Male	38	Primary/Secondary	Married	Low	No			No		No		
81	Male	25	Primary/Secondary	Separated/Divorced	Low	Yes	2 - 4 hours	Live events	Yes	Believable and Relatable	No, But I considered	-	No
82	Male	23	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	No, But I considered	-	No
83	Male	29	Tertiary (Degree/Diploma)	Married	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	No, But I considered	-	No
84	Female	25	Tertiary (Degree/Diploma)	Single	Medium	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	No, But I considered	-	No
85	Male	25	Tertiary (Degree/Diploma)	Single	Medium	Yes	2 - 4 hours	International News and Programs	Yes	Short and precise	No, But I considered	-	No
86	Male	21	Tertiary (Degree/Diploma)	Single	Medium	Yes	More than 4 hours	Local News and Programs	Yes	Believable and Relatable	No, But I considered	-	No

											ed		
87	Male	25	Primary/Secondary	Single	Low	Yes	2 - 4 hours	International News and Programs	Yes	Entertaining	No, But I considered	-	Yes
88	Female	25	Tertiary (Degree/Diploma)	Married	Low	Yes	2 - 4 hours	International News and Programs	Yes	Short and precise	No	-	No
89	Male	23	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Live events	Yes	Believable and Relatable	No, But I considered	-	Yes
90	Male	29	Tertiary (Degree/Diploma)	Married	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	No, But I considered	-	Yes
91	Female	30	Tertiary (Degree/Diploma)	Married	Low	Yes	More than 4 hours	International News and Programs	Yes	Believable and Relatable	No, But I considered	-	Yes
92	Male	37	Primary/Secondary	Married	Low	Yes	2 - 4 hours	International News and Programs	Yes	Believable and Relatable	No, But I considered	-	Yes
93	Male	24	Primary/Secondary	Single	Low	No			No		No		
94	Male	27	Primary/Secondary	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	No, But I considered	-	Yes
95	Male	25	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Live events	Yes	Entertaining	Yes	Yes	Yes
96	Male	21	Tertiary (Degree/Diploma)	Single	Low	Yes	Less than 1 hour	International News and Programs	Yes	Entertaining	Yes	Yes	Yes
97	Male	24	Tertiary (Degree/Diploma)	Single	Medium	Yes	2 - 4 hours	Live events	Yes	Believable and Relatable	Yes	Yes	Yes
98	Male	23	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	International News and Programs	Yes	Short and precise	Yes	Yes	Yes
99	Female	26	Tertiary (Degree/Diploma)	Single	Medium	Yes	2 - 4 hours	Live events	Yes	Believable and Relatable	Yes	Yes	Yes
100	Male	23	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	Yes	Yes	Yes
101	Male	26	Primary/Secondary	Single	Medium	Yes	2 - 4 hours	Live events	Yes	Believable and Relatable	Yes	No	No
102	Female	22	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	Yes	No	No
103	Male	27	Tertiary (Degree/Diploma)	Single	Medium	Yes	More than 4 hours	Local News and Programs	Yes	Entertaining	Yes	No	No
104	Female	24	Primary/Secondary	Single	Low	Yes	Less than 1 hour	International News and Programs	Yes	Believable and Relatable	Yes	No	No
105	Female	26	Tertiary (Degree/Diploma)	Single	Medium	Yes	Less than 1	International News and	Yes	Short and precise	Yes	No	No

5	le		iploma)				hour	Programs					
106	Male	24	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Short and precise	Yes	No	No
107	Male	29	Primary/Secondary	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	No, But I considered	-	No
108	Male	22	Primary/Secondary	Single	Low	No			No		No		
109	Male	25	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Short and precise	No, But I considered	-	No
110	Male	24	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	International News and Programs	Yes	Entertaining	No, But I considered	-	No
111	Female	25	Tertiary (Degree/Diploma)	Single	Low	Yes	Less than 1 hour	Live events	Yes	Entertaining	No	-	No
112	Male	22	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Live events	Yes	Believable and Relatable	No, But I considered	-	Yes
113	Female	20	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Short and precise	No, But I considered	-	Yes
114	Female	23	Tertiary (Degree/Diploma)	Single	Low	Yes	Less than 1 hour	Local News and Programs	Yes	Believable and Relatable	No, But I considered	-	Yes
115	Female	24	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	No, But I considered	-	Yes
116	Male	21	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Entertaining	No	-	No
117	Male	22	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	International News and Programs	Yes	Entertaining	No, But I considered	-	Yes
118	Male	23	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	No	Short and precise	No	-	Yes
119	Male	23	Tertiary (Degree/Diploma)	Single	Low	Yes	Less than 1 hour	International News and Programs	Yes	Believable and Relatable	No, But I considered	-	Yes
120	Male	24	Tertiary (Degree/Diploma)	Single	Low	Yes	2 - 4 hours	Local News and Programs	Yes	Believable and Relatable	No, But I considered	-	Yes