Scientific Research Publishing

Open Journal of Statistics, 2021, 11, 737-758
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Validating Intrinsic Factors Informing E-Commerce: Categorical Data Analysis Demo

Anthony Joe Turkson*, John Awuah Addor, Douglas Yenwon Kharib

Mathematics, Statistics and Actuarial Science Department, Takoradi Technical University, Sekondi-Takoradi, Ghana
Email: *anthonyjoeturkson@yahoo.com

How to cite this paper: Turkson, A.J., Addor, J.A. and Kharib, D.Y. (2021) Validating Intrinsic Factors Informing E-Commerce: Categorical Data Analysis Demo. *Open Journal of Statistics*, **11**, 737-758.
https://doi.org/10.4236/ojs.2021.115044

Received: July 29, 2021
Accepted: October 9, 2021
Published: October 12, 2021

Abstract

Statistics is a powerful tool for data measurement. Statistical techniques properly planned and executed give meaning to meaningless data. The difficulty some practitioners encounter hinges on the fact that though there are numerous statistical methods available for use in analysis, the extent of their understanding and ease of using these tools for analysis is limited. This study has twofold purpose: firstly, literature on categorical data commonly used in research was reviewed; next, we reported the results of a survey we designed and executed. Categorical data was collected via questionnaire and analyzed to serve as a backbone of the robustness of categorical data. Several conjectures about the independence of the socio-economic variables and e-commence were tested. Some of the factors influencing patronage of e-commerce were identified. It is clear from the literature that as one's academic qualification improves, there is an associated improvement in their preference for e-commerce, but the results revealed otherwise. Size of family was found to influence e-commerce. Both income and social status positively affected patronage in e-commerce. Gender also appeared to affect patronage in e-commerce. 62.3% of staff had patronized e-commerce. This shows that e-commerce patronage was gradually increasing. It is therefore our considered view that policy documents regulating and monitoring the use of e-commerce be developed to increase e-commerce participation across the globe. It is also recommended that the bottlenecks which obstruct patronage in e-commence be addressed so that a lot more staff will develop a positive attitude towards e-commerce.

Keywords

Categorical Data, Chi-Square, E-Commerce, Ordinal Data, Nominal Data

1. Introduction

Statistics has been and is still a very powerful tool for data measurement. If

properly planned and executed, it gives meaning to meaningless numbers. Statistical analysis is therefore a powerful technique that helps to find patterns and differences in the data as well as identify relationships between variables. The nature of the data collected and the design of the study determines the appropriate significance test that should be used [1]. Over the past three to four decades, series of papers and books on categorical data have been published to throw light on the principles governing categorical data. The works of countless scholars and researchers like References [2] and [3] have documented a unified approach to the analysis of categorical data. These set of approaches have a number of significant advantages over traditional methods of analysis. The approaches link the analysis of categorical data to the general linear models and provide a comprehensive and unified scheme for the analysis of multidimensional contingency table. Recent researchers have found the relationship of categorical response variables with one or more explanatory variables [4]. The difficulty some practitioners often encounter hinges on the fact that though there are numerous statistical analysis methods available for use in data analysis, the extent of their knowledge, understanding and ease of using these tools for analysis and application to policy evaluation and research is quite limited.

The purpose of the study is twofold: first and foremost, we reviewed and illustrated categorical data methods commonly used in applied research with emphasis on categorical variables, ordinal and nominal data, contingency table analysis, principles governing use of Chi-square, Cramer's V and modeling procedures. Basically, we wanted our readers to have an in-depth knowledge, cutting-edge understanding and appreciate how to design, collect and use the principles governing categorical data for data analysis. In the second section, we reported the results of a small survey conducted in which we identified some of the common goals of significance testing. In achieving this, we designed and collected categorical data via questionnaire, coded and analyzed it to serve as a backbone of the robustness of categorical data. We tested several claims, conjectures and hypothesis about the independence nature of the socio-economic variables and the response variable.

In particular, we hypothesized that the two categorical variables:

- Patronage of e-commerce and age are independent;
- Patronage of e-commerce and size of family are independent;
- Patronage of e-commerce and region are independent;
- Patronage of e-commerce and gender are independent;
- Patronage of e-commerce and location are independent;
- Patronage of e-commerce and level of education are independent;
- Patronage of e-commerce and social status are independent;
- Patronage of e-commerce and marital status are independent;
- Patronage of e-commerce and religious affiliation are independent;
- Patronage of e-commerce and level of income are independent; and
- Patronage of e-commerce and kind of occupation are independent.

1.1. Merits Associated with Categorical Data Analysis (CDA)

The merits associated with Categorical Data Analysis (CDA), which none other statistical method has included but not limited to the following:

1.1.1. Label Encoder

Researchers can use the CDA method to transform non-numerical labels to numerical labels. This transformation cannot be done with some other statistical methods.

1.1.2. Combining Levels

The CDA procedure has the ability to avoid redundant levels in a categorical variable. It does so by simply combining the different rare levels by applying techniques like the business logic, frequency or response rate.

1.1.3. Dummy Coding

CDA can also adopt dummy coding in converting a categorical input variable into continuous variable. As the name suggests, CDA duplicates variables which represents one level of a categorical variable with one (1) for presence of a level and zero (0) for the absence of a level

1.2. Categorical Variable

Categorical variable is a variable that can take on one of a limited and usually fixed number of possible values, assigning each individual or unit of observation to a particular group or nominal category on the basis of some qualitative properties. According to reference [5], categorical data analysis is the analysis of data where the response variable has been grouped into a set of mutually exclusive ordered or unordered categories. Additionally, reference [5] has defined a categorical variable to be one for which the measurement scale consists of a set of categories that are non-numerical. Reference [6] has also alluded to the fact that categorical variables are used to organize observations into groups that share a common trait. The concept of categorical data was developed by [7] who classified the measurement scale into four categories namely nominal, ordinal, interval and ratio scales. He further prescribed the analysis techniques that are appropriate for the analysis of each of the four measuring scales. Throughout history, researchers have critiqued these prescribed statistical analysis schemes. Reference [8] has proposed a hierarchy under the following classifications to consist of grades, ranks, counted fractions, counts, amounts and balances.

Categorical data consists of counts and not measurements and therefore encapsulates all sampling units obtained by counting. The method does not depend on any assumptions about the parameter of the population and generally assumes that data are measured at the nominal or ordinal levels. It also involves the statistical treatment of categorical response variables to ascertain any intrinsic factor. Reference [6] has posited that the number of groups within any categorical variable should not exceed twenty; in addition, the procedure must be such that the researcher can distinguish between independent (explanatory) and

dependent (response) variables. Categorical variables are pervasive in the social sciences for measuring attitudes and opinions, in biomedical sciences, it measures outcomes such as whether a medical treatment was successful or not; in the behavioral sciences (measures type of mental illness); in epidemiology and public health (measures contraceptive method); in education (measures response of students to examination questions); in marketing (measures consumer preference); in engineering sciences and industrial quality control (measures classification of items to see if they conform to certain standards).

There are basically two common types of hypothesis testing problems that are addressed with categorical data analysis. We might want to find:

1) How well a sample distribution corresponds with a population distribution? (We can hypothesize that the distribution of student dislike of statistics over the last 5 years has not changed and collect sample data to support or refute the claim)

2) An evidence for a relationship between two qualitative variables, thus necessitating the need to analyze a cross-classification of two discrete distributions. (We can hypothesize that there is no relationship between gender and fear of Statistics)

All over the world, people especially researchers love to clump things into categories of all kinds, this tendency of categorization into little bins of distinct categories is an activity that should not be overlooked especially when issues of strategic analysis are involved. In choosing an appropriate statistical method for a categorical data situation, one should consider the measurement scale of both the response (dependent) and explanatory (Independent) variable as well as the form in which the data are best characterized. Categorical data methods apply to situations when the response (dependent) variable is nominal or ordinal. The next section is devoted to the explanation of the nominal and ordinal scales. Researchers have proposed a number of ways of analyzing and interpreting categorical data, prevalent among them is the use of the contingency table. This statistical tool is used to establish whether there is a relationship between two categorical variables.

1.3. Ordinal and Nominal Variables

Qualitative research, qualitative data and qualitative variables are all concerned with opinions, feelings and experiences which can vary from one individual to the other. In this discourse we proffer to use variable or data interchangeably. Qualitative variables are of two main kinds; ordinal and nominal. An in-depth knowledge about and understanding of the concepts of ordinal and nominal variables will go a long way to unravel the mysteries surrounding their use, analysis, interpretation, and drawing of statistical inference. Next, it will help the researcher to decide on the appropriate statistical analysis that must be used for the assigned values.

The first kind, ordinal data have categories which could be ordered or ranked; they cannot be counted. The ordering comes naturally; however, this ordering

does not have a standard scale on which the difference in variable in each scale can be measured. Ordinal variables generally indicate that some subjects or objects are better than others but then, it cannot say by how much better they are because the intervals between the categories are not equal. The position of ordinal variable in the quantitative-qualitative classification is fuzzy, researchers often treat them as qualitative, using methods for nominal variables even though by all standards ordinal variables more closely resemble interval variables than they resemble nominal, this is because they possess quantitative features like greater than or smaller than. As proposed by [5], if S is an ordinal scale that assigns real numbers in $\mathcal{R}$ to the elements of a set P observations then,

$$P \xrightarrow{S} \mathcal{R}$$

such that

$$\iota > j \leftrightarrow S(\iota) > S(j) \quad \text{for all} \quad \iota, j \in P$$

Such a scale S preserves the one-to-one relationship between numerical order values. Under an ordinal scale of measurement, the observed data are ranked in terms of degree to which they possess a characteristic of interest. Instances of this scale of measurement follows: patient condition (good, bad, critical); rank in graduating class (90^{th} percentile, 70^{th} percentile, 50^{th} percentile); type of degree (BSc., MSc., PhD); social status (lower, middle, upper); division of degree (first division, second division-upper or lower, third division, pass and fail); place of settlement (rural, peri-urban urban); customer satisfaction (very poor, poor, not quite sure, good, very good); level of education (no education, junior high school, senior high school, first degree, master's degree, terminal degree); proficiency level (advanced, intermediate, novice). Rating also falls under ordinal variables (respondent can be asked to rate their happiness on a scale of 1 to 5); the agreement – disagreement scale and such other scales. The Likert scale with attitudinal response variables of three, four, five or seven responses can be seen as a partial ordinal variable when there is an inclusion of the option "Neutral or do not know". It is fully ordinal without the inclusion of the "Neutral or do not know" option. Respondents can be asked to rate the food served by a restaurant using the Likert scale. It should be mentioned that ordinal data are not real numbers and therefore cannot be placed on the number line, it is not appropriate to apply any of the rules of basic arithmetic to ordinal data, that is to say, we cannot add, subtract, multiply or divide, this limitation limits us to the type of analysis we can do with such data.

The second kind, nominal data can be defined as data that is used for naming or labelling variables, without any quantitative value. Nominal data involves categories that have no particular order, and which are mutually exclusive. These categories may not require the assignment of numerical values, but only unique identifiers or as the name implies [5]. As indicated by reference [5], nominal data are names or labels put on some variables, however, there is no measure of distance between the values. Classical examples are: Nationality; names of people or things; hair colour or eye; race; gender; qualification; religious affiliation;

brand of soap; brand of vehicle; level of motivation; motives for travelling; marital status and so on. Nominal variables with only two categories are often said to be binary or dichotomous.

A variable's measurement scale determines which statistical methods are appropriate. In the measurement pyramid, ordinal variables are higher than nominal variables. Statistical methods that are used for variables that are at higher levels cannot be used for variables at the lower level, because their categories have no meaningful ordering, on the hand, statistical methods that are used for variables that are at the lower level can be used for variables at higher levels by ignoring the ordering of the categories [9].

1.4. Contingency Table Analysis (R × C)

Studies about complex data involve a combination of non-parametric and model-based testing and estimation procedures. Contingency table (also known as crosstabs or two-way tables) is a type of table in a matrix format that displays the frequency distribution of the variables. Purposefully, a contingency table provides a way of portraying data that can facilitate the computation of probabilities. The table helps in determining conditional probabilities very easily. It displays sample values in relation to two different variables that may be dependent or contingent on one another. The method is synonymous to categorical data analysis. It is widely used in bio-medical, marketing, engineering, social sciences and business research. The major questions addressed by contingency tables are whether the variables under study are independent or not. It also assesses which of two models provides the best explanation for an available data. It is worthy to note that contingency tables are suitable for nominal, ordinal, interval and ratio variables regardless of the number of categories these variables might have. We consider inferences for contingency tables, in particularly, we look at the analysis of two-way tables for the assessment of significant association between two variables and by extension, the analysis of sets of two-way tables for testing conditional independence of two variables. We note the statistics for three cases: Case 1, where both the row (exposure or factor variable) and column (response variable) are nominal. Case 2, when only the column is ordinal and Case 3, when both the row and column variables are ordinal. In these cases, scores are assigned as ranks or integers in ordinal variables.

The Exact inference for the 2 × 2 table was proposed by [10]. If π denotes the odds ratio, and H_0 and H_1 the null and alternative hypothesis respectively, then the one-tailed p-value for the Fisher's exact test for H_0: $\pi = 1$ against the H_1: $\pi > 1$ is obtained by summing the probabilities corresponding to tables in which the sample odds ratio is at least as large as the observed, or equivalently those tables whose cell count for the first row and first column is at least as large as n_{11}. Fisher showed that conditioning on the row and column margins from the observed table with cell counts (n_{11}, n_{12}, n_{21}, n_{22}) gives the probability of observing $n_{11} = t$ as

$$P\left(n_{11}=t \mid N, N_{1.}, N_{.1}, \pi\right)=\frac{\binom{N_{1.}}{t}\binom{N-N_{1.}}{N_{.1}-t_{11}}\pi^{t}}{\sum_{u}\binom{N_{1.}}{u}\binom{N-N_{1.}}{N_{.1}-u}\pi^{u}}$$

where the index of summation, u, ranges from the maximum of 0 and $N_{1.}+N_{.1}-N$ to the minimum of $N_{1.}$ and $N_{.1}$, the possible values for n_{11} for the given marginal totals. Under H_0, Expression 1 with $\pi = 1$ is the hypergeometric distribution. A two-sided p-value is calculated by summing the probabilities of tables from the reference set whose probabilities are no larger than the probability of the observed table. It is significant to note that this approach is adopted when the sample size is small, that notwithstanding it is also valid for all sample sizes. This method is used because the significance of the deviation from a null hypothesis can be calculated exactly.

1.5. Chi-Square Test of Independence

It has been alluded by reference [11] that the Chi-square statistic is a non-parametric tool designed to analyze group differences when the dependent variable is measured at a nominal level. Like all non-parametric statistics, the Chi-square is robust with respect to the distribution of the data. Specifically, it does not require equality of variances among the study groups or homoscedasticity in the data. It permits evaluation of both dichotomous independent variables and of multiple group studies. Unlike many other non-parametric and some parametric statistics, the calculations needed to compute the Chi-square provide considerable information about how each of the groups performed in the study. This richness of detail allows the researcher to understand the results and thus, to derive more detailed information from this statistic than from many others. It is required that the study groups be independent, otherwise, a different test must be used if the two groups are related.

The primary use of the chi-square test is to examine whether two variables are independent or not. In other words, we want to find out if the two factors are related or not, we say that one variable is "not correlated with" or "independent of" or "not related with" the other if an increase in that variable is not associated with an increase in the other. If two variables are correlated or related, their values tend to move together, either in the same or in the opposite direction. Chi-square examines a special kind of correlation between two nominal variables. Chi-square is a significance statistic, and thus should be followed with a strength test statistic. The Cramer's V is the most common strength test used to test the data when a significant Chi-square result has been obtained. Merits of the Chi-square include its robustness with respect to distribution of the data, its ease of computation, the detailed information that can be derived from the test, its use in studies for which parametric assumptions cannot be met, and its flexibility in handling data from two-group and multiple-group studies. Limitations include its sample size requirements, difficulty of interpretation when there are

large numbers of categories (20 or more) in the independent or dependent variables, and tendency of the Cramer's V to produce relatively low correlation measures, even for highly significant results [11].

The assumptions associated with the chi-square test are fairly straightforward: The data at hand must have been randomly selected (to minimize potential biases) and the variables in question must be nominal or ordinal. Regarding the hypotheses to be tested, all chi-square tests have the same general null and alternative hypothesis. The null hypothesis states that there is no relationship between the two variables, while the alternative hypothesis states that there is a relationship between the two variables. The test statistic follows a chi-square distribution, and the conclusion depends on whether or not the obtained statistic is greater that the critical statistic at a chosen alpha level. The test statistics is:

$$\chi^2 = \sum_{\text{all cells}} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} = \sum_{i,j} \frac{\left(f_{0ij} - f_{eij}\right)^2}{f_{eij}} \tag{1}$$

The test statistic measures the difference between the observed counts and the expected counts assuming independence. Equation (1) is called chi-square statistic because if the null hypothesis is true, then it has a chi-square distribution with $(r-1)(c-1)$ degrees of freedom.

Studies done on this statistic indicates that If the χ^2-statistic is large, it implies that the observed counts are not close to the expected counts if the two variables were independent. Thus, large values of χ^2 give evidence against the H_0, and supports the H_1.

The p-value of the chi-square test is the probability that the χ^2-statistic, is larger than the value we obtained if H_0 is true. Also, if H_0 is true, the χ^2-statistic has chi-square distribution with $(r-1)\times(c-1)$ degrees of freedom.

It is worth noting the following about the Chi-square distribution:

- It is not symmetric;
- All values are positive;
- The shape of the chi-square distribution depends on the degrees of freedom;
- Hypothesis test involving chi-square is usually one-tailed. We aim at finding out if the observed sample distribution significantly differs from the hypothesised distribution. Low values of chi-square indicate that the sample distribution and the hypothetical distribution are similar to each other, high values indicate that the distributions are dissimilar;
- A random variable has a chi-square distribution with N degrees of freedom if it has the same distribution as the sum of the squares of N independent variables, each normally distributed and having expectations 0 and variance 1.

1.6. Cramer's V (φ_c)

This test is a measure of the strength of association between two categorical variables. The values range from zero (0) to one (1). Cramer's V (φ_c) is a symmetrical measure, what that means is that it does not matter which variables appear

in the columns or rows. When φ_c is zero (weak), it implies there is no association between the variables; when φ_c is exactly one (1), it implies there is a very strong association between the variables. Let a sample of size n of the simultaneously distributed variables S and Y for $i=1,2,\cdots,r; j=1,2,\cdots,k$ be given frequencies, then Cramer's V is computed by taking the square root of the Chi-square divided by the sample size n, and the minimum dimension minus 1.

$$V=\sqrt{\frac{\varphi^2}{\min\left(k-1,r-1\right)}}=\sqrt{\frac{\chi^2/n}{\min\left(k-1,r-1\right)}}.$$

where,

n is the grand total

k is the number of columns

r is the number of rows

φ is the phi coefficient

χ^2 is the value of the Pearson's Chi-square test

The p-value for the significance of V is the same one that is calculated using the Pearson's Chi-square test

The step-by-step procedure for obtaining the phi coefficient is also outlined below in Table 1 below.

1.7. Concepts on E-Commerce

Electronic commerce (E-commerce) was deployed to boost business transactions which involves the selling and buying of information, services, and goods by means of computer telecommunications networks. It usually refers to the trading of goods and services over the Internet. E-commerce consists of business-to-consumer and business-to-business commerce as well as internal organizational transactions that support those activities. With the wide adoption of the Internet and the introduction of the World Wide Web in 1991 and of the first browser for accessing it in 1993, most e-commerce shifted to the Internet. More recently, with the global spread of smartphones and the accessibility of fast broadband connections to the Internet, much e-commerce moved to mobile devices, which also included tablets, laptops, and wearable products such as watches [5]. It has been observed that although traditional methods still exist, online shopping transactions have increased rapidly.

Table 1. Tabular representation of a 2 × 2 matrix for calculating the Cramer's V.

	$Y=1$	$Y=2$	Total
$X=1$	n_{11}	n_{12}	$n_{1.}$
$X=2$	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	n

$\varphi=\frac{n_{11}n_{22}-n_{12}n_{21}}{\sqrt{n_{1.}n_{2.}n_{.1}n_{.2}}}$.

Several studies have documented the determining factors in e-commence, predominant among them are the following: Quality of website; convenience; searching brands; information search; shopping experience; social interactions; information usability; payment systems; security; price; ease of use; satisfaction; reliability of website; secure payments; customization; interaction; internet access; website aesthetics; experience; age; learning capacity; purchasing preferences; consumers' characteristics; contextual factors; perceived uncertainty; perceived benefits; individual differences and technological developments [7] [8] [11]-[21].

Reference [22] has investigated the relationship between the dependent factor "abandonment factors" against four independent variables namely: risk; navigation; finance and purchase and two dummy variables: age and level of education. They aimed at finding the effects of these variables on e-commerce transactions in Nigeria. Interestingly the study revealed that risk, navigation, finance and purchase had significant impact on the abandonment of online purchases. But, age and level of education appeared not to have any significant impact on the abandonment of online purchases.

1.8. Simulation

A simulation study was done on some online goods patronized and the number of customers who had to leave the platform after some time on the platform due to unavailability of the indicated items in stock. This scenario is presented in Table 2. Figure 1 gives a graphical representation of the number of shoppers who left the platform due to unavailability of stock.

2. Methods

This study falls directly under the case study qualitative research design, therefore the rules governing this design were carefully followed. The targeted population of study were all university staff across the globe, but due to logistical constraints, one of the universities in Africa - Ghana, was used as a case study, therefore all adults who were living and working in the Takoradi Technical University community within the period 1st to 30th June 2021, were used as the study

Table 2. Number of Shoppers leaving platform due to unavailability of stock.

Goods	Minutes	Leaving
T-shirt	8.18	65
Bag	9.68	43
Perfume	9.52	71
Sandal	8.93	73
Shoe	9.40	79
Necklace	8.05	69
Wig	8.88	68

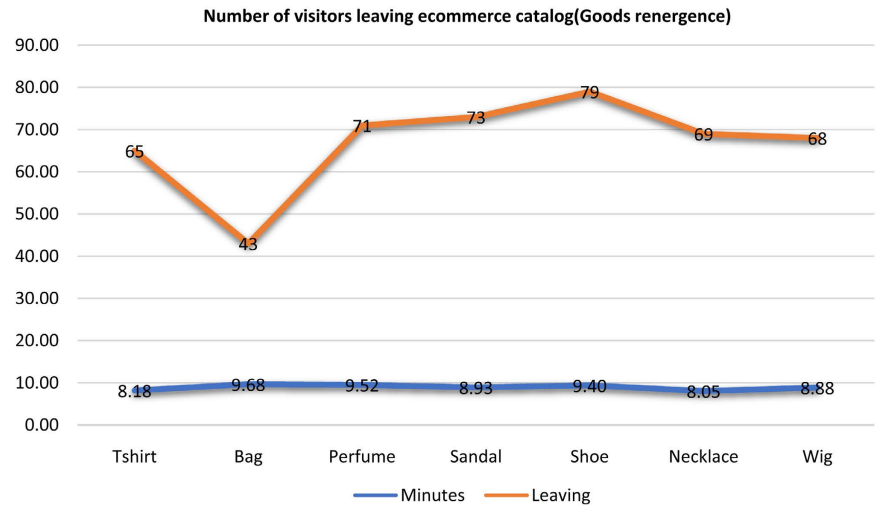

Figure 1. Line graph of shoppers leaving platform due to unavailability of stock items.

units, the conclusions that will be drawn about the study will therefore be delimited to such staff, but results could be used as a generalization of staff in similar institutions. Participants below the age of eighteen were excluded from the study, moreover, students, irrespective of their ages were also excluded. A detailed structured closed-ended questionnaire which covered thirteen thematic areas was developed and pretested. This measuring scale was used so that we could assign a numerical value to otherwise subjective opinions. A stratified two stage cluster sampling technique was employed to obtain the data. The first stage sampling block was achieved by dividing the University community into five blocks, namely: Administration block; applied arts and technology block, applied sciences block; engineering block and the Business block. The second stage was achieved through stratification by age, gender, level of education, religious affiliation, location, occupation, region, level of income, family size, and social status.

From each stratum, the sample-proportional to size method was used to select the appropriate sample size from each of the homogenous groups. Once the sampling units were identified, questionnaires were administered to them. The response variable (also known as the dependent variable) of the study was the use of internet in business transactions (e-commerce). This was unearthed using the closed question "When was the last time you patronized, purchased or ordered goods and services for personal use via the Internet using either your mobile phone or websites.

This question had four categories namely: Never used it before; One to three months; three to six months; and over six months. There were eleven explanatory or independent variables in the study: Age; gender; marital status; level of education; religious affiliation; location; occupation; region; level of income; family size and social status

A sample size of 300 was enumerated instead of the entire staff population of 900 due to logistical constraints. The data was combed, coded and keyed into

SPSS version 20. Frequency and percentages were first obtained from the respondents according to their status on e-commerce. Pie charts and tables were used to explore the data. Chi-square test of independence was performed to examine the relationship between their e-commerce status and the eleven explanatory variables. Likelihood ratio test was also conducted to ascertain which model was more appropriate in the analysis

Cramer's V measure was used to test the strength of association between two categorical variables.

3. Results

The results of the analysis are presented as Tables.

4. Discussion

Table 3 reveals the percentage of staff who fell into various categories of both the response variable and the eleven explanatory variables. A greater number of staff (33.7%) had not patronized e-commerce. Thirty-three percent (33%) were recent users of that platform while 17% had patronized the platform over the last six (6) months. Most (43.3%) of the staff sampled were less than 28 years old. Almost the same number of males and females were sampled even though the males had a slight urge over the females. Of the number sampled, majority (51.7%) were not married. We infer again that a greater number of staff (44%) were from the Western part of the country, moreover, majority belonged to the Christian community.

With respect to the family or household size we noted that a greater percentage (30.3%) belonged to the 3 or less category. In addition to that, majority lived in the urban community. Regarding the level of Education, a greater proportion had bachelor degrees. With Social status, majority (65.3%) found themselves within the middle class. In respect of the income Level, a lot more people were in the lower middle-income bracket than in the others. Finally, it was also revealed that a greater proportion (40.3%) of staff sampled belonged to the teaching cadre.

Taking a cursory view of Table 3, we note particularly that the three hundred (300) staff selected cuts across all independent variables, namely: Occupation; age; gender; social status; religious status; marital status; level of education; family size; region of descent; place currently living and level of income. We also note that the distribution of staff to the dependent variable which sought to investigate the last time they patronized e-commerce was fairly distributed, this gives us some hope that the further analysis which will be done through the use of the Chi-square statistic will be very reliable and dependable.

Following is the result from the cross tabulation which reported on the Pearson Chi-square significant test, the Likelihood test, the linear-to-linear association, the Phi and the Cramer's V. In addition, a Likelihood ratio test was conducted to ascertain which of the models was appropriate for the significant test.

Table 3. Frequencies and percentages of staff according to e-commerce status.

Explanatory variables Categories	Count	Percentage	
The Last time staff patronized or purchased goods via e-commerce	Never used it before	101	33.7
	One to three months	99	33.0
	Three to six months	49	16.3
	Over Six months	51	17.0
Age	18 - 27	130	43.3
	28 - 37	78	26.0
	38 - 47	57	19.0
	48 and above	35	11.7
Gender	Male	154	51.3
	Female	146	48.7
Marital Status	Single	173	57.7
	Married	114	38.0
	Widow/Divorced	13	4.3
Region of Birth	Ahafo	7	2.3
	Bono	11	3.7
	Ashanti	20	6.7
	Bono East	11	3.7
	Central	43	14.3
	Eastern	14	4.7
	Greater Accra	20	6.7
	Volta	16	5.3
	Western	132	44.0
	Western North	13	4.3
	North of Ghana	13	4.3
Religious Affiliation	Christianity	243	81.0
	Islamic	44	14.7
	Traditionalist	13	4.3
Size of Family	3 or less people	91	30.3
	4 people	79	26.3
	5 people	53	17.7
	6 people	77	25.7
Place of Location	Urban	174	58.0
	Peri-Urban	65	21.7
	Rural	61	20.3

Continued

Level of Education	College	84	28.0
	Bachelor	131	43.7
	Masters	73	24.3
	PhD	12	4.0
Social Status	Lower	37	12.3
	Middle	196	65.3
	Upper	67	22.3
Occupation	Directorate of ICT	19	6.3
	Registry/Transport	70	23.3
	Directorate of Finance	13	4.3
	Lecturer/Professor/Technician/service Person	121	40.3
	Health Directorate	11	3.7
	Directorate of Physical Development	66	22.0
Income Level	High	20	6.7
	Upper-Middle	116	38.7
	Lower-Middle	121	40.3
	Low	43	14.3

Source: Field data.

In interpreting the cross-tabulation results presented in Table 4, we asked ourselves this question: Are the observed counts so different from the expected counts that will warrant the conclusion that a relationship exists between the two variables? Based on this question, we observed the actual values and the expected values within each cell, we noted that the observed values and expected values were quite similar, there were a few discrepancies though, the absolute residual values varied from as small as 0.1 to a high value of 3.5 for the religious affiliation variable; 0.1 to 6.8 for the kind of occupation variable; 0.3 to 8.1 for the education variable and 0.6 to 4.1 for the place of location variable, just to mention a few, with these observations, we can confidently say that the two categorical variables as formulated in the conceptual framework are independent of each other. The chi-square results support this observation with a low Chi-square value (high significance value) of 3.982 (0.679); 6.137 (0.726); and 24.787 (0.735) in the variables: level of education; age and kind of occupation respectively. We notice again that the measures of association provided by the linear-to-linear association were too small (0.077; 0.017; 0.345) and do not approach significance. In reporting the output, Chi-square did the analysis over a two-tailed paradigm, the reason being that it is more difficult to reject our null hypothesis with a two-tailed test than it is with a one-tailed test, a statistically significant result under a two-tailed assumption would also be significant under a one-tailed assumption.

Table 4. Cross tabulation of eleven predictor variables and one response variable and chi-square analysis.

Predictor variables			The Last time patronized or purchased goods via e-commerce				Value Sig.		
			Never before	1 to 3 months	3 to 6 months	Over 6 months			
Size of Family	2 or less	Observed	10	19	3	2	Chi-Square	20.108	0.065
		Expected	11.4	11.2	5.6	5.8	Likelihood	19.965	0.068
	3 people	Observed	17	17	13	10	Linear-Lin.	0.289	0.591
		Expected	19.2	18.8	9.3	9.7			
	4 people	Observed	27	29	12	11			
		Expected	26.6	26.1	12.9	13.4	Phi	0.259	0.065
	5 people	Observed	15	13	12	13	Cramer's V	0.149	0.065
		Expected	17.8	17.5	8.7	9.0			
	6 or more	Observed	32	21	9	15			
		Expected	25.9	25.4	12.6	13.1			
	High	Observed	5	6	3	6	Chi-Square	16.619[a]	0.055
		Expected	6.7	6.6	3.3	3.4	Likelihood	17.724	0.039
Level of Income	Upper-Middle	Observed	31	41	23	21	Linear-Linear	5.140	0.023
		Expected	39.1	38.3	18.9	19.7			
	Lower-Middle	Observed	44	42	21	14			
		Expected	40.7	39.9	19.8	20.6	Phi	0.235	0.055
	Low	Observed	21	10	2	10	Cramer's V	0.136	0.055
		Expected	14.5	14.2	7.0	7.3			
	Lower	Observed	19	6	5	7	Chi-Square	11.369[a]	0.078
		Expected	12.5	12.2	6.0	6.3	Likelihood	11.552	0.073
Social Status	Middle	Observed	61	64	36	35	Linear-Lin.	0.000	1.000
		Expected	66.0	64.7	32.0	33.3			
	Upper	Observed	21	29	8	9	Phi	0.195	0.078
		Expected	22.6	22.1	10.9	11.4	Cramer's V	0.138	0.078
Gender	Male	Observed	56	40	26	32	Chi-Square	8.134[a]	0.043
		Expected	51.8	50.8	25.2	26.2	Likelihood Ratio	8.191	0.042
	Female	Observed	45	59	23	19	Phi	0.165	0.043
		Expected	49.2	48.2	23.8	24.8	Cramer's V	0.165	0.043
Level of Education	College	Observed	34	30	10	10	Chi-Square	11.278	0.257
		Expected	28.3	27.7	13.7	14.3	Likelihood	13.198	0.154
	Bachelor	Observed	40	42	26	23	Linear-Lin.	1.664	0.197
		Expected	44.1	43.2	21.4	22.3			
	Masters	Observed	21	22	13	17			
		Expected	24.6	24.1	11.9	12.4	Phi	0.194	0.257

Continued

	PhD	Observed	6	5	0	1	Cramer's V	0.112	0.257
		Expected	4.0	4.0	2.0	2.0			
	Christian	Observed	80	82	37	44	Chi-Square	3.982[a]	0.679
		Expected	81.8	80.2	39.7	41.3	Likelihood	4.159	0.655
Religion	Islamic	Observed	17	13	10	4	Linear-Lin.	0.077	0.781
		Expected	14.8	14.5	7.2	7.5	Phi	0.115	0.679
	Traditional	Observed	4	4	2	3	Cramer's V	0.081	0.679
		Expected	4.4	4.3	2,1	2.2			
	Single	Observed	50	63	30	30	Chi-Square	7.492[a]	0.278
		Expected	58.2	57.1	28.3	29.4	Likelihood	7.183	0.304
Marital Status	Married	Observed	48	32	17	17	Linear-by-Linear	0.345	0.557
		Expected	38.4	37.6	18.6	19.4	Phi	0.158	0.278
	Divorced	Observed	3	4	2	4	Cramer's V	0.112	0.278
		Expected	4.4	4.3	2.1	2.2			
Age	18 - 27	Observed	44	48	19	19	Chi-Square	6.137[a]	0.726
		Expected	43.8	42.9	21.2	22.1	Likelihood	6.092	0.731
	28 - 37	Observed	21	25	17	15	Linear-Lin	0.017	0.898
		Expected	26.3	25.7	12.7	13.3			
	38 - 47	Observed	23	15	8	11	Phi	0.143	0.726
		Expected	19.2	18.8	9.3	9.7	Cramer's V	0.083	0.726
	Above 48	Observed	13	11	5	6			
		Expected	11.8	11.6	5.7	6.0			
Place of Location	Urban	Observed	56	59	27	32	Chi-Square	11.278	0.257
		Expected	58.6	57.4	28.4	29.6	Likelihood	13.198	0.154
	Peri-Urban	Observed	26	19	10	10	Linear-by-	1.664	0.197
		Expected	21.9	21.5	10.6	11.1			
	Rural	Observed	19	21	12	9	Phi	0.194	0.257
		Expected	20.5	20.1	10.0	10.4	Cramer's V	0.112	0.257
	Directorate of ICT	Observed	6	5	4	4	Chi-Square	15.527[a]	0.414
		Expected	6.4	6.3	3.1	3.2	Likelihood	15.602	0.409
	Registry/Transport	Observed	28	21	13	8	Linear	0.753	0.386
		Expected	23.6	23.1	11.4	11.9			
Kind of work	Directorate of Fin.	Observed	3	7	2	1			
		Expected	4.4	4.3	2.1	2.2	Phi	0.228	0.414
	Lecturer/Professor	Observed	38	46	13	24	Cramer's V	0.131	0.414
		Expected	40.7	39.9	19.8	20.6			
	Health Directorate	Observed	5	2	1	3			
		Expected	3.7	3.6	1.8	1.9			

Continued

	Directorate of Phy. D	Observed	21	18	16	11			
		Expected	22.2	21.8	10.8	11.2			
	Ahafo	Observed	2	3	1	1	Chi-Square	24.787[a]	0.735
		Expected	2.4	2.3	1.1	1.2	Likelihood	24.255	0.760
	Bono	Observed	6	2	1	2	Linear-linear	0.528	0.468
		Expected	3.7	3.6	1.8	1.9			
Region	Ashanti	Observed	7	7	2	4	Phi	0.287	0.735
		Expected	6.7	6.6	3.3	3.4	Cramer's V	0.166	0.735
	Bono East	Observed	4	4	2	1			
		Expected	3.7	3.6	1.8	1.9			
	Central	Observed	12	19	7	5			
		Expected	14.5	14.2	7.0	7.3			
	Eastern	Observed	5	4	3	2			
		Expected	4.7	4.6	2.3	2.4			
	Greater Accra	Observed	6	2	8	4			
		Expected	6.7	6.6	3.3	3.4			
	Volta	Observed	3	7	4	2			
		Expected	5.4	5.3	2.6	2.7			
	Western	Observed	48	44	17	23			
		Expected	44.4	43.6	21.6	22.4			
	Western North	Observed	3	5	2	3			
		Expected	4.4	4.3	2.1	2.2			
	North of Ghana	Observed	5	2	2	4			
		Expected	4.4	4.3	2.1	2.2			

For the Pearson chi-square or simply the Chi-square and the maximum likelihood methods, we noted that in principle as the test statistics value gets larger, the likelihood that the two variables are not independent also increases. If the value is close to one, it suggests that the two variables are not dependent on each other. Under the column heading significance, we infer that the large p-values indicate that the observed values do not differ significantly from the expected values. The linear-by-linear test statistics which test whether two variables correlate with each other was also examined. This measure even though meaningless because there is no logical or numeric relationship to the order of the variables, reveals clearly that the correlation between the two variables was meaningless. Phi, which measures the strength of the association between two categorical variables ranged from 0.115 to 0.287. This means there was a very weak relationship between each of the independent variable and the dependent variables.

Lastly, the Cramer's V which measures the strength of the association between two categorical variables revealed the following statistics: size of family (0.149); level of income (0.136): social status (0.138); Gender (0.165); level of education (0.112); religion (0.081); marital status (0.112); age (0.08); place of location (0.112); kind of work (0.131) and region of descent (0.166). The values in parenthesis are all small revealing that the association between each of the independent variable and the response variable is very slim.

Table 5 presents the results of the likelihood ratio statistics of the field data. A likelihood ratio statistic reflects the relative likelihood of the data given two competing models. Likelihood ratios provide an intuitive approach to summarizing the evidence provided by an experiment. Basically, the test compares the fit of two models. The null hypothesis states that the smaller model is the best model; It is rejected when the test statistic is large. In other words, if the null hypothesis is rejected, then the larger model is a significant improvement over the smaller one. This statistic is used to test the Null hypothesis (that the reduced model is the true or best model) that all parameters of the effect of reducing the variables are zero. Looking through the results, we note that the effect of reducing the model by each of the variables: social status; income; type of occupation and family size are all zero since their p-values (0.015; 0.002; 0.011 and 0.014 respectively) are all less than 0.05. Thus, we reject the claim that the reduced model is best and accept the fact that a model containing all these variables is the best. With the rest of the variables whose p-values were greater than 0.05, the effect of reducing the model by these variables were different from zero. An attempt to find out from the staff reasons adduced to non-patronage in e-commerce revealed the following as shown in the bar chart represented as Figure 2, we inferred that a greater proportion of the staff, adduced lack of trust as the basis for non-patronage.

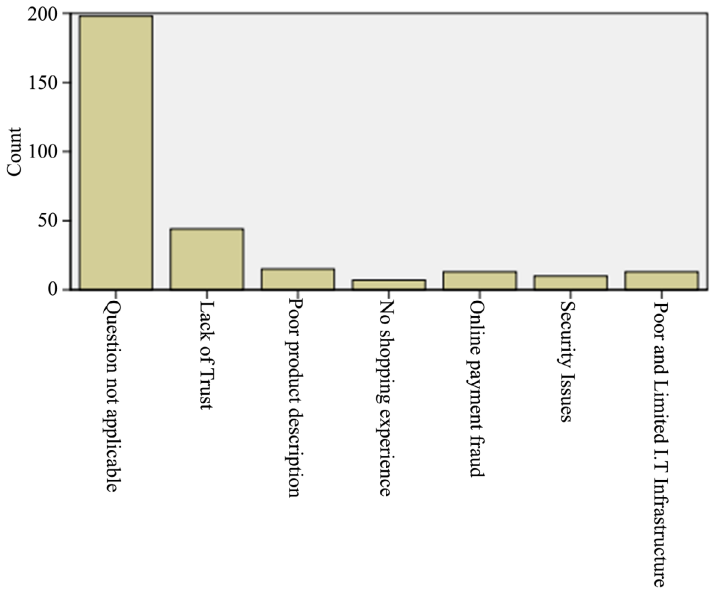

Figure 2. Bar chart showing the reasons why staff refused to patronize E-commerce.

Table 5. Chi-square statistics of the difference in −2log likelihoods between the final model and a reduced model. The reduced model is equivalent to the final model because omitting the effect does not increase the degree of freedom.

Likelihood Ratio Tests				
Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	−2 Log Likelihood of Reduced Model	Chi-square	df	Sig.
Intercept	628.154[a]	0.000	0	
Age	631.268	3.114	9	0.960
Gender	631.791	3.637	3	0.303
Marital status	634.798	6.643	6	0.355
Region	660.463	32.309	30	0.353
Religion	632.720	4.566	6	0.601
Location	630.135	1.981	6	0.921
Level of Education	633.809	5.655	9	0.774
Social Status	643.938	15.784	6	0.015
Income	653.677	25.522	9	0.002
Kind of Occupation	674.633	46.479	27	0.011
Family Size	653.413	25.258	12	0.014

5. Conclusions

From the outset, we sought to review and illustrate categorical data methods commonly used in applied research with emphasis on categorical variables, ordinal and nominal data, contingency table analysis, principles governing use of Chi-square, Cramer's V and modeling procedures. We also aimed at reporting the results of a small survey conducted in which we identified some of the common goals of significance testing.

In the likelihood ratio test, some of the factors influencing the patronage of e-commerce were identified as educational levels. It is clear from the literature that as one's academic qualification improves, there is an associated improvement in their preference for e-commerce, but the results revealed otherwise. Our results therefore contradict studies that have been done on similar topics [23] [24]. The size of the family was another important variable that influences e-commerce. The influence could be negative or positive [25]. In the foregoing analysis, it also came to light that both income and social status positively affect patronage in e-commerce. This revelation supports earlier studies that an improvement in socio-economic conditions has a significant effect on e-commerce. In the cross-tabulation results, gender also appeared to affect patronage in e-commerce. Inferring from **Table 2**, 62.3% of the staff had at one time or the other patronized e-commerce. This revelation shows that e-commerce patronage is gradually increasing. It is therefore our considered view that policy documents regulating and monitoring the use of e-commerce should be developed to increase e-commerce across the globe. It is also recommended that the bottlenecks

which obstruct patronage in e-commence as reveled in **Figure 2** above should be mitigated, so that a lot more staff will develop positive attitude towards patronage of e-commerce.

Acknowledgements

We thank the staff of the university for their time and effort in helping us reach such conclusions.

Funding Information

No funding agency.

Author's Contributions

Anthony Joe Turkson

Prepared the questionnaire, trained the administrators of the questionnaire and supervised the administration and collection of questionnaires, did the simulation, prepared the manuscript according to the author guidelines and edited final work.

John Awuah Addor

Was part of the team that pre-tested the questionnaire, corrected the questionnaire, administered the questionnaire, coded the data, entered into the SPSS software, analyzed the data and interpreted the results.

Douglas Yenwon Kharib

Was part of the team that pre-tested the questionnaire, corrected the questionnaire, administered the questionnaire, coded the data, entered into the SPSS software, analyzed the data and interpreted the results.

Ethics

All ethical issues have been addressed. In particular, the reasons behind the questionnaire were clearly explained.

Conflicts of Interest

None declared.

References

[1] Lazar, J., Feng, J. and Hochheiser, H. (2017) Research Methods in Human-Computer Interaction. 2nd Edition, Morgan Kaufmann, Cambridge.

[2] Cox, D.R. (1970) The Analysis of Binary Data. Methuen, London.

[3] McFaffan, D. (1974) The Measurement of Urban Travel Demand. *Journal of Public Economics*, **3**, 303-328. https://doi.org/10.1016/0047-2727(74)90003-6

[4] Preissor, J.S. and Koch, G.G. (1997) Categorical Data Analysis in Public Health Annual Review. *Annual Review of Public Health*, **18**, 15-82. https://doi.org/10.1146/annurev.publhealth.18.1.51

[5] Velleman, P.F. and Wilkinson, L. (1993) Nominal, Ordinal, Interval and Ration Typology Are Misleading. *The American Statistician*, **47**, 65-72.

https://doi.org/10.1080/00031305.1993.10475938

[6] Imrey, P.B. and Koch, G. (2005) Categorical Data Analysis (Encyclopedia of Biostatistics). John Wiley & Sons, Hoboken. https://doi.org/10.1002/0470011815.b2a10011

[7] Kadı, F. and Peker, C. (2015) Analyzing the Factors Affecting E-Commerce in Turkey. *International Journal of Management, Accounting and Economics*, **2**, 1319-1339.

[8] Wolfinbarger, M. and Gilly, M.C. (2001) Shopping Online for Freedom, Control, and Fun. *California Management Review*, **43**, 34-55. https://doi.org/10.2307/41166074

[9] Agresti, A. (1992) A Survey of Exact Inferences for Contingency Tables. *Statistical Science*, **7**, 131-177. https://doi.org/10.1214/ss/1177011462

[10] Fisher, R.A. (1922) On the Interpretation of X^2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society*, **85**, 87-94. https://doi.org/10.2307/2340521

[11] Mchugh, M.L. (2013) The Chi-Square Test of Independence. *Biochemia Medica*, **23**, 143-149. https://doi.org/10.11613/BM.2013.018

[12] Lin, G.T. and Sun, C.C. (2009) Factors Influencing Satisfaction and Loyalty in Online Shopping: An Integrated Model. *Online Information Review*, **33**, 458-475. https://doi.org/10.1108/14684520910969907

[13] Padmavathy, C., Swapana, M. and Paul, J. (2019) Online Second-Hand Shopping Motivation-Conceptualization, Scale Development, and Validation. *Journal of Retailing and Consumer Services*, **51**, 19-32. https://doi.org/10.1016/j.jretconser.2019.05.014

[14] Shin, J.I., Chung, K.H., Oh, J.S. and Lee, C.W. (2013) The Effect of Site Quality on Repurchase Intention in Internet Shopping through Mediating Variables: The Case of University Students in South Korea. *International Journal of Information Management*, **33**, 453-463. https://doi.org/10.1016/j.ijinfomgt.2013.02.003

[15] Kim, J. and Lennon, S.J. (2013) Effects of Reputation and Website Quality on Online Consumers' Emotion, Perceived Risk and Purchase Intention. *Journal of Research in Interactive Marketing*, **7**, 33-56. https://doi.org/10.1108/17505931311316734

[16] Park, E.J., Kim, E.Y., Funches, V.M. and Foxx, W. (2012) Apparel Product Attributes, Web Browsing, and e-Impulse Buying on Shopping Websites. *Journal of Business Research*, **65**, 1583-1589. https://doi.org/10.1016/j.jbusres.2011.02.043

[17] Rahman, M.A., Islam, M.A., Esha, B.H., Sultana, N. and Chakravorty, S. (2018) Consumer Buying Behavior towards Online Shopping: An Empirical Study on Dhaka City, Bangladesh. *Cogent Business & Management*, **5**, Article ID: 1514940. https://doi.org/10.1080/23311975.2018.1514940

[18] Nagar, K. (2016) Drivers of E-Store Patronage Intentions: Choice Overload, Internet Shopping Anxiety, and Impulse Purchase Tendency. *Journal of Internet Commerce*, **15**, 97-124. https://doi.org/10.1080/15332861.2016.1148971

[19] Alkan, O., Kucukoglu, H. and Tutar, G. (2021) Modeling of the Factors Affecting E-Commerce Use in Turkey by Categorical Data Analysis. *International Journal of Advanced Computer Science and Application*, **12**, 1-11. https://doi.org/10.14569/IJACSA.2021.0120113

[20] Watson, K.B. (2014) Categorical Data Analysis. In: Michalos, A.C., Ed., *Encyclopedia of Quality of Life and Well-Being Research*, Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_291

[21] Alkan, Ö., Oktay, E. and Genç, A. (2015) Determination of Factors Affecting the

Children's Internet Use. *American International Journal of Contemporary Research*, **5**, 57-67.

[22] Toyin, O. and Damilola, O.T. (2012) Abandonment Factors Affecting E-Commerce Transactions in Nigeria. *International Journal of Computer Applications*, **46**, 41-47.

[23] Stranahan, H. and Kosiel, D. (2007) E-Tail Spending Patterns and the Importance of Online Store Familiarity. *Internet Research*, **17**, 421-434. https://doi.org/10.1108/10662240710828076

[24] Koyuncu, C. and Lien, D. (2003) E-Commerce and Consumer's Purchasing Behavior. *Applied Economics*, **35**, 721-726. https://doi.org/10.1080/0003684022000020850

[25] Leong, L.-Y., Jaafar, N.I. and Ainin, S. (2018) Understanding Facebook Commerce (F-Commerce) Actual Purchase from An Artificial Neural Network Perspective. *Journal of Electronic Commerce Research*, **19**, 75-103. https://doi.org/10.1037/t67487-000