An Investigation of the Association Between Tourist Pre-Trip Planning Time and Length of Trip, Lodging Choice, Tourist Psychographics and Demographics: An Application of Correspondence Analysis and Cramér's V Effect Size

James E. Stoddard, stoddardje@appstate.edu

G. David Shows, showsgd@appstate.edu

Abstract

When performing survey research it is normal to collect descriptive information such as income, gender, highest education attainment, and others. This information is used to categorize the collected research responses into groups. This nominal or ordinal data may also be used to find patterns in the collected data and suggest relationships. As an exploratory research method, it can suggest future research possibilities to confirm these relationships. This paper introduces the use of Correspondence Analysis (CA) as a research technique to suggest possible relationships using nominal or ordinal data, using collected research from a survey performed to measure tourist preferences in the Blue Ridge National Heritage Area of North Carolina. CA reveals possible associations between trip planning time, time spent, and other collected nominal data. We discuss the strengths and weaknesses of this research technique and possible future uses for the marketing researcher.

Introduction

Survey research often includes nominal and ordinal measures as certain questions naturally lend themselves to these measurement levels. A common example includes respondent education attainment level. Other questions that could be measured using a ratio scale, such as respondent income, are also measured using an ordinal scale to increase response rates. Irrespective of the reason(s) why nominal or ordinal measures are used, nonparametric statistical techniques

employing the chi-square statistic are used to examine whether a relationship exists between two categorical variables.

Correspondence Analysis. Among the set of techniques available to process categorical data is correspondence analysis (CA). CA is a nonparametric statistical technique that can be used to analyze categorical data. CA is an exploratory technique that looks for patterns in categorical data using two-way or multi-way tables with each row and column becoming a point on a multidimensional graphical map or bi-plot (Greenacre, 1993; Doey and Kurta, 2011). The goal of the technique is to explain the most variance in the data (called inertia) using the smallest number of dimensions. In this sense then, CA is similar to principal component factor analysis, except for categorical data. Hoffman and Franke (1986) identified several features of CA that contribute to its usefulness to marketing researchers. First, the technique allows for the simultaneous analysis of multiple categorical variables. Second, CA can reveal relationships that would not be detected in a series of pairwise comparisons of variables. Third, CA not only shows that variables are related but also how those variables are related. Finally, CA has very liberal data requirements, necessitating only a rectangular data matrix containing non-negative values.

Effect Sizes. One point that has to be kept in mind when using correspondence analysis to search for relationships between variables is the chi-square statistic's sensitivity to sample size. With a large enough sample size trivial relationships can become statistically significant. As a result, statistically significant relationships do not necessarily mean meaningful relationships.

The concept of effect size was defined by Cohen (1988) as "the degree to which the phenomenon is present in the population" or "the degree to which the null hypothesis is false" where the null hypothesis means that the effect size is zero (p. 9 – 10). Therefore, when the null hypothesis is rejected it is rejected to some specific degree, the effect size in the population.

The phi coefficient is a special case of the product-moment correlation and is used when two variables are dichotomous (i.e., a 2 x 2 contingency table) (Fern and Monroe, 1996). Cramér's V sometimes called Cramér's phi (φc) is a measure of the strength of association between two categorical variables used when one of those variables has more than two categories (i.e., 2 x 3 contingency tables and greater)(Cramér 1946, p. 282, Field 2016). Cramér's V is calculated by taking the square root of the chi-square statistic ($\chi 2$) divided by the sample size (n) multiplied by the minimum dimension of the r by c matrix minus one.

$$V = \sqrt{\frac{\chi^2}{n * \min(r - 1 \; or \; c - 1)}}$$

Cramér's V is a number between 0 and 1 that indicates how strongly two categorical variables are associated. The rule of thumb regarding relationship strength using Cramér's V is as follows (Cohen, 1988 p. 79).

Cramér's V	Strength
Below .10	Weak Relationship
.10 - .30	Moderate Relationship
Above .30	Strong Relationship

The purpose of this exploratory paper is to demonstrate the use of correspondence analysis for categorical data, hypothesis testing using the chi-square statistic and the Cramér's V effect size within the context of a large sample size to demonstrate that traditional hypothesis testing may lead to erroneous conclusions when compared to the effect size estimator.

Method

A survey was developed to measure tourist preferences in the Blue Ridge National Heritage Area of North Carolina. The data were collected from tourists to a variety of venues within the Blue Ridge National Heritage area in North Carolina (see Figure 1). The data were collected from seven tourism venues within the Blue Ridge National Heritage Area. In all, 896 responses were collected.

Measures. The categorical measures used in the survey included how far in advance plans were made to visit the area, the number of nights stayed, the type of lodging used, the types of leisure activities engaged in, as well as the demographic variables of gender, education, and income.

Figure 1.

The Blue Ridge National Heritage Area

Advanced planning was an ordinal measure (less than one week, 1 – 2 weeks, 3 – 5 weeks, 6 – 11 weeks, and 12+ weeks). Number of nights stayed was also an ordinal measure (day tripper, 1 – 3 nights, 4 – 7 nights, 8 – 14 nights, 15 – 30 nights, 30+ nights). Lodging establishments were nominal measures (bed & breakfasts, campground or RV, friends/family, hotel/motel chain, independent hotel/motel, rental condo, cabin or house, resort hotel). Local activities were also nominal measures (check all that apply), (art museums & galleries, Blue Ridge Parkway, caverns, festivals & special events, fishing, golfing, Grandfather Mountain, hiking & biking, historic sites, music & theater, orchards, farms or wineries, rafting or canoeing, shopping, Tweetsie Rail Road). Gender was a nominal measure and education and income were ordinal (high school, some college, Bachelor's degree, graduate degree), ($0 – 24,999, $25 – 49,999, $50 – 74,999, $75 – 99,999, $100,000+) respectively.

Results

Relationship #1. Trip Planning Time and Number of Trip Nights

The first exploratory relationship examined the association between trip planning time and length of the trip. Table 1a presents the frequencies (counts) of the number of respondents indicating how long in advance they planned for their trip and their trip length. Table 1b shows the summary from the correspondence analysis. As Table 1b shows, the null hypothesis of no association (independence) between trip planning time and length of the trip was rejected ($\chi 2$ = 332.813, df = 20, p = .000). The strength of the association between trip planning time and length of the trip was strong (Cramér's V = .307). Figure 1 shows the biplot of the association. The greater the length of the trip the more time was spent in trip planning.

Table 1a

Correspondence Table

	nights						
advance	Day tripper	1-3 nights	4-7 nights	8-14 nights	15-30 nights	more than 30 nights	Active Margin
Less than 1 week	91	94	17	4	1	7	214
1 to 2+ weeks	25	76	27	5	0	3	136
3-5+ weeks	15	104	56	7	2	7	191
6-11+ weeks	7	50	46	11	4	8	126
More than 12 weeks	5	42	84	23	9	51	214
Active Margin	143	366	230	50	16	76	881

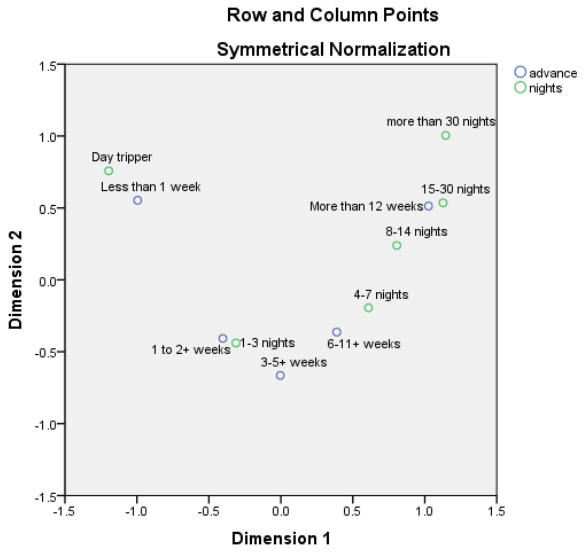
Table 1b:

Summary

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation 2
1	.543	.294			.779	.779	.026	.341
2	.279	.078			.205	.985	.031	
3	.072	.005			.014	.998		
4	.025	.001			.002	1.000		
Total		.378	332.813	.000[a]	1.000	1.000		

a. 20 degrees of freedom

Figure 2



Relationship #2. Trip Planning Time and Lodging Choice

A second correspondence analysis was conducted to see whether there was an association between trip planning time and the type of accommodations the respondents used. Table 2a shows the frequencies and Table 2b shows the results from the correspondence analysis. As Table 2b shows, the null hypothesis of no association between trip planning time and type of accommodation used was rejected ($\chi 2 = 123.451$, df = 24, p = .000). The biplot (Figure 3) shows that tourists that planned for the trip for less than 1 week preferred independent hotels, tourists that planned 1 – 2 weeks preferred camping/RV parks or bed and breakfasts, those that planned 3 – 5 weeks tended to stay with friends/family or chain hotels/motels, tourists that planned 6 – 11 weeks in advance preferred resort hotels, and those than planned 12 weeks or more preferred rental condo or cabins. The relationship between trip planning time and lodging type was moderately strong (Cramér's V = .245).

Table 2a

Correspondence Table

	lodging							
advance	Bed & Breakfast	Campground or RV Park	Friends/Family	Hotel/Motel Chain	Independent Hotel	Rental Condo/Cabin/ House	Resort Hotel	Active Margin
Less than 1 week	9	28	16	30	29	14	2	128
1 to 2+ weeks	11	22	9	19	30	12	4	107
3-5+ weeks	7	19	35	33	34	31	7	166
6-11+ weeks	7	19	16	13	16	34	8	113
More than 12 weeks	6	20	38	14	8	78	7	171
Active Margin	40	108	114	109	117	169	28	685

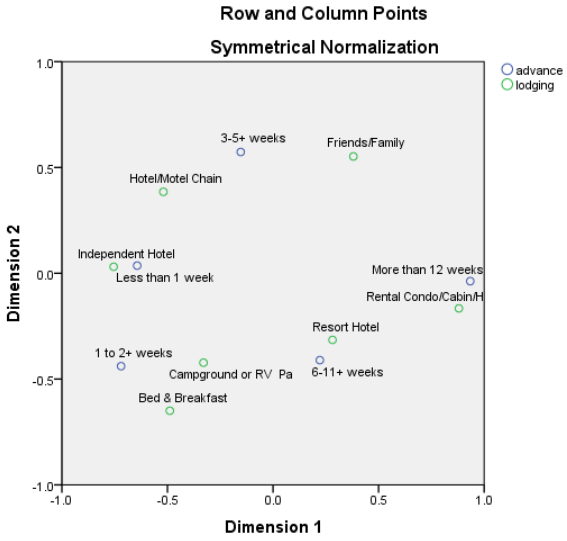
Table 2b

Summary

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation 2
1	.390	.152			.843	.843	.033	.011
2	.138	.019			.106	.949	.037	
3	.085	.007			.040	.989		
4	.044	.002			.011	1.000		
Total		.180	123.451	.000[a]	1.000	1.000		

a. 24 degrees of freedom

Figure 3



Relationship #3. Trip Planning Time and Types of Leisure Activities

A third correspondence analysis was conducted to see whether there was an association between trip planning time and the types of leisure activities tourists preferred. The type of leisure activities question was "check all that apply. As a result, the SAS PROC CORRESP procedure was used and the output is slightly different. Table 3a presents the frequencies and Table 3b shows the output from the correspondence analysis. Ass Table 3b shows, the null hypothesis of no association between trip planning time and type of leisure activities preferred was rejected ($\chi 2 = 105.198$, df = 52, $p < .001$). Visitors that planned their trip less than one week in advance preferred touring caverns, 1 – 2 weeks Tweetsie Rail Road, 3 – 5 weeks Grandfather Mountain and the Blue Ridge Parkway, 6 – 11 weeks hiking & biking, 12 weeks or more were interested in festivals, fishing, rafting, golfing museums, music/theater, rafting and visiting historic sites (Figure 4). The strength of association between trip preplanning time and type of activities tourists were interested in was moderate (Cramér's V = .187).

Table 3a

Contingency Table

	museums	parkway	caverns	festivals	fishing	golfing	grandfather	hikeBike	historic	musictheater	orchards	rafting	shopping	tweetsie	Sum
<1week	35	131	80	29	19	12	127	43	47	15	12	19	71	45	685
1-2weeks	22	85	47	25	13	10	84	40	33	21	16	12	48	38	494
3-5weeks	49	125	82	44	18	16	126	60	64	32	23	22	84	43	788
6-11week	27	83	40	29	16	17	85	41	37	24	19	17	64	32	531
12+weeks	89	162	75	98	47	45	160	84	99	63	50	43	126	57	1198
Sum	222	586	324	225	113	100	582	268	280	155	120	113	393	215	3696

Table 3b

Inertia and Chi-Square Decomposition

Singular Value	Principal Inertia	Chi-Square	Percent	Cumulative Percent
0.15458	0.02389	88.314	83.95	83.95
0.04941	0.00244	9.023	8.58	92.53
0.03776	0.00143	5.269	5.01	97.54
0.02648	0.00070	2.592	2.46	100.00
	0.02846	105.198	100.00	

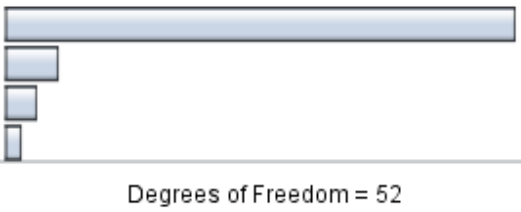
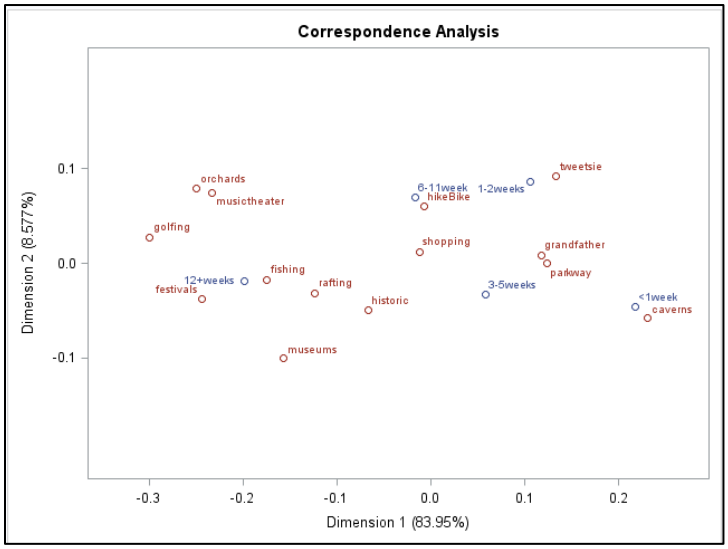


Figure 4



Relationship #4. Trip Planning Time and Gender

A fourth correspondence analysis was conducted to see whether there was an association between pre-trip planning time and respondent gender. Table 4a presents the frequencies and Table 4b the correspondence analysis results. As Table 4b shows, the null hypothesis of no relationship between pre-trip planning time and gender was rejected ($\chi 2 = 10.752$, df = 4, p = .029). No biplot was generated since gender had only two categories. Examination of the frequency chart (Figure 5) shows that males tended to plan trips further in advance than females. The strength of the association between gender and trip planning time was moderate (Cramér's V = .112).

Table 4a

Correspondence Table

	gender		
advance	female	male	Active Margin
Less than 1 week	125	80	205
1 to 2+ weeks	89	44	133
3-5+ weeks	100	89	189
6-11+ weeks	72	50	122
More than 12 weeks	108	102	210
Active Margin	494	365	859

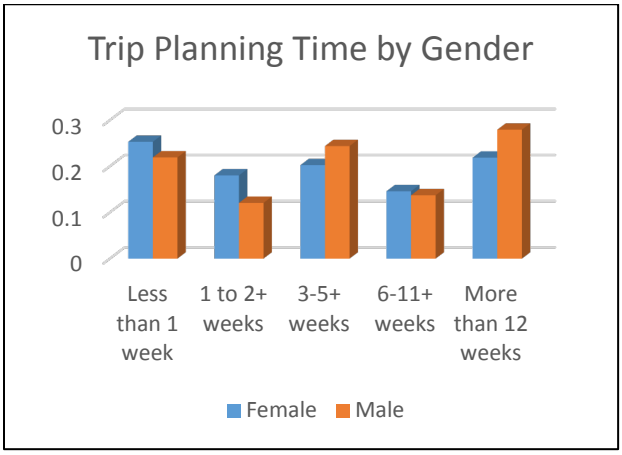
Table 4b

Summary

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value
					Accounted for	Cumulative	Standard Deviation
1	.112	.013			1.000	1.000	.034
Total		.013	10.752	.029[a]	1.000	1.000	

a. 4 degrees of freedom

Figure 5



Relationship #5. Trip Planning Time and Educational Attainment

A fifth correspondence analysis explored whether there was a relationship between pre-trip planning time and tourists' educational attainment level. Table 5a presents the frequencies and Table 5b the correspondence analysis results. As Table 5b shows, the null hypothesis of no relationship between pre-trip planning time and educational attainment was not rejected ($\chi 2$ = 13.429, df = 12, p = .339). The strength of the association between pre-trip planning time and educational attainment level was also weak (Cramér's V = .064).

Table 5a

Correspondence Table

advance	educat: High School	Some College	Bachelor's Degree	Graduate Degree	Active Margin
Less than 1 week	29	70	51	49	199
1 to 2+ weeks	19	44	35	26	124
3-5+ weeks	24	44	66	42	176
6-11+ weeks	15	31	38	35	119
More than 12 weeks	24	60	57	57	198
Active Margin	111	249	247	209	816

Table 5b

Summary

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia: Accounted for	Proportion of Inertia: Cumulative	Confidence Singular Value: Standard Deviation	Confidence Singular Value: Correlation 2
1	.109	.012			.724	.724	.035	.027
2	.067	.004			.270	.994	.035	
3	.010	.000			.006	1.000		
Total		.016	13.429	.339[a]	1.000	1.000		

a. 12 degrees of freedom

Relationship #6. Trip Planning Time and Income

A final correspondence analysis sought to determine whether an association existed between pre-trip planning time and tourist income level. Table 6a presents the frequencies and table 6b the correspondence analysis results. As the correspondence analysis results show, the null hypothesis of no association between trip pre-planning time and tourist income was rejected ($\chi 2 = 30.452$, df = 16, p = .016). Generally speaking, those tourists with a higher income plans for the trip further in advance. However, the strength of this association was only moderate to weak (Cramér's V = .101). The biplot shows the mixed results (Figure 7).

Table 6a

Correspondence Table

advance	income					
	$24,999 or less	$25-$49,999	$50-$74999	$75-$99,999	$100,000 or more	Active Margin
Less than 1 week	27	64	52	16	28	187
1 to 2+ weeks	20	26	27	23	22	118
3-5+ weeks	18	34	48	31	35	166
6-11+ weeks	12	32	22	22	20	108
More than 12 weeks	12	48	43	27	40	170
Active Margin	89	204	192	119	145	749

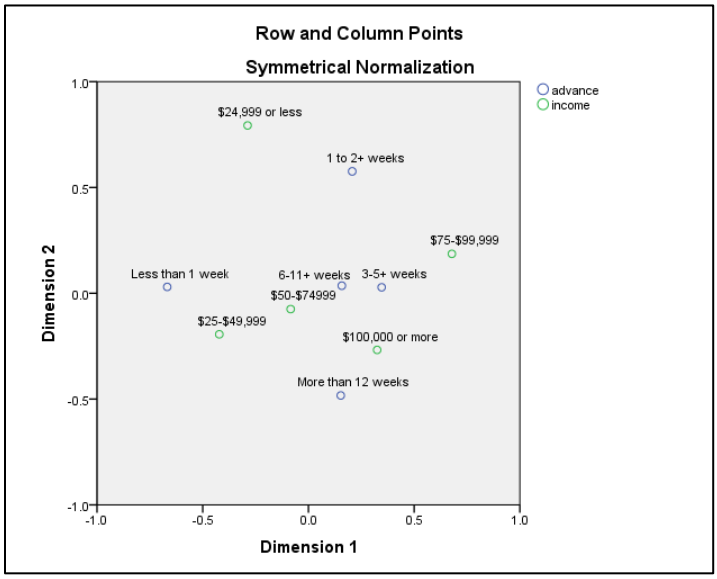
Table 6b

Summary

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation 2
1	.154	.024			.581	.581	.033	-.025
2	.106	.011			.275	.856	.036	
3	.073	.005			.130	.987		
4	.023	.001			.013	1.000		
Total		.041	30.452	.016[a]	1.000	1.000		

a. 16 degrees of freedom

Figure 7



Summary

The results provide an interesting illumination to tourists that visit the west North Carolina area around the Blue Ridge Parkway. There is a strong association between planning the trip to the Blue Ridge and the length of stay. The greater the length of the trip the more time was spent in trip planning. Trip planning time also saw a moderately strong association when it came to lodging; the greater amount of time planning, the more relatively upscale accommodations. Greater time planning is associated with interactive activities including hiking, biking, canoeing or kayaking, and golfing. There is a moderate association between earlier trip planning and gender; planning earlier is more likely in males. There is a weak association between pre-trip planning time and education level. Finally, there is a moderate association between income and pre-trip planning.

Considering this research, tourism bureaus in the Blue Ridge area might wish to provide a free online trip planning application, or an Android or IOS app, which provide the consumers with a directory of the activities and resources available. Predictive analytics could monitor both their planning window and their time online planning, and "push" possible options based upon these two parameters. Options for the applications could including automatic reservations, a planner for their week of travel, and suggested other festivals or activities for the tourists to visit.

Conclusion

The above results provide an excellent example of the strengths associated with correspondence analysis and the broad range of information that is attainable from nominal and ordinal data. Much of the time we use such categorical information in general descriptive terms; in this example, we are given predictive correlations with are both interesting and worthy of further research. Correspondence analysis is a desirable tool for exploratory study, providing quantitate output for future study using more robust experimental techniques. Like most exploratory techniques, it provides what, but not why. Like good exploratory techniques, it provides sufficient and significant data to properly focus future exploration.

As mentioned, with a large enough sample size trivial relationships can become statistically significant. Stated again, statistically significant relationships do not necessarily mean meaningful relationships. Conversely, this is a good technique to consider when dealing with limited sample sizes; if an association exists between two nominal or ordinal categories with smaller data collections, it supports associations that deserve further study.

Standard marketing tools for the researcher include t-tests, ANOVAs, structural equation modeling, partial least squares, MANOVAs, and others. Researchers sometimes limit themselves to research techniques that they show competence in, or invest in research procedures that are currently topical. Proficient marketing researchers should invest themselves in a wide array of techniques, adapting to the output given to them by the data. Correspondence analysis can be considered a valuable tool when probing to find associations and to greater focus work for future research.

References:

Cohen, Jacob (1988), Statistical Power Analysis for the Behavioral Sciences, 2ed., Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Cramér, Harald (1946), Mathematical Methods of Statistics, NJ: Princeton University Press.

Doey, Laura and Jessica Kurta (2011), "Correspondence Analysis Applied to Psychological Research." Tutorials in Quantitative Methods for Psychology, 7 (1), 5–14.

Fern, Edward F. and Kent B. Monroe (1996), "Effect –Size Estimates: Issues and Problems in Interpretation, Journal of Consumer Research, 23 (2), 89-105.

Field, Andy (2016), Discovering Statistics Using IBM SPSS Statistics, Thousand Oaks, CA: Sage Publications, Inc.

Greenacre, Michael J., (1993), Correspondence Analysis in Practice, London, UK, Academic Press, Harcourt Brace and Company.

Hoffman, Donna L. and George R. Franke (1986), "Correspondence Analysis: Graphical Representation of Categorical Data in Marketing Research." Journal of Marketing Research, 23 (August), 213-227.

Liebetrau, Albert M. (1983), Measures of Association, Thousand Oaks, CA: Sage Publications, Inc.

Keywords**:** correspondence analysis, research, exploratory research, statistical associations, nominal data, ordinal data, categorical research

Track: Marketing Research

ID#: 1312