Gold Open Access
Full contents are available in NepJOL(http://www.nepjol.info/index.php/DSAJ)
DOAJ (https://doaj.org/toc/1994-2672)

Dhaulagiri Journal of Sociology and Anthropology

Answering Research Questions Using Survey Data: An Example from Maryland Agritourism

-Prem Bhandari
University of Maryland Eastern Shore, USA

Article Info
Received: September 14, 2024
Revised received: December 10, 2024
Accepted: December 11, 2024
Available online: December 31, 2024
DOI: https://doi.org/10.3126/dsaj.v18i2.73288

Abstract

This brief paper discusses the concept, the sources, and the types of research questions that can be answered with survey data. In addition, I provided an example of answering a specific question using survey data with step-wise results from univariate (descriptive statistics), bivariate, and multivariate (e.g., binary logistic regression) analysis. I believe this paper will be useful to the young scholars.

Keywords: analysis, evidence, measurement, question

Introduction

In the brief paper, "*Levels of measurement: Foundational basis for quantitative analysis of survey data*," I discussed the concept of the levels of measurement of variables and their applications in choosing an appropriate statistical technique for analyzing survey data (Bhandari, 2022). The paper described the concept, types of levels of measurement, and the application of these measurement levels to the statistical analysis of survey data.

In this paper, I describe a research question, the sources of the research question, and the types of research questions that can be answered using survey data. In addition, I provide an example of answering a research question using survey data.

In my experience of teaching and learning survey data analysis for the past several years, I have found students less clear about framing a research question for their research and answering them empirically using survey data. I hope this paper will be useful to many of them.

What is a Research Question?

A social researcher is interested in answering (exploring, describing, or explaining) a research question for addressing a societal problem. Specifically, a researcher may be interested in empirically answering a question(s) of interest with survey data. A researcher may have an idea, a problem, or an issue in mind, but how to address that problem or answer that question can be a panic step for beginners.

According to Lavrakas (2008), "a research question is an operationalization of the purpose(s) to which a survey project aims" (p. 736). A research problem is the starting point for a research question. According to Alford (1998), a research question points towards a theoretical framework that guides a way to answer the question of interest scientifically. In addition, it seeks empirical evidence with an answer (Alford, 1998). Lavarakas (2008) further clarified that a research question guides a researcher about the entire research design, including the target population, sample, sampling strategy, measurement, data collection, and data analysis.

Broadly, there are two types of research questions – a

https://orcid.org/0000-0002-1705-8511
Email: pbhandari115@gmail.com,

theoretical research question and an empirical research question. A theoretical question is one "...that derives from an unresolved general conceptual issue in the field" (Alford, 1998, p. 26). A few examples are:

- What are the barriers to accessing quality education for rural children?
- How do cultural beliefs and practices influence health-seeking behaviors among individuals of Nepal?
- How do gender norms and roles affect women's participation in labor market?
- How does climate change affect agricultural productivity and food security?
- What are the impacts of urbanization on natural resource management and sustainability?
- How do community-based conservation efforts contribute to biodiversity preservation?

On the other hand, an empirical question is "…one that is answerable from some kind of evidence or data" (Alford, 1998, p. 26). Here, we are concerned about an empirical research question. Examples are provided below.

Examples of Research Questions

Research questions vary by the nature of the investigation.

Exploratory Research

Exploratory research is designed to explore the area under investigation when limited information is available. Exploratory research helps us increase our understanding of a given research topic. Both qualitative and quantitative methods are used to explore a topic. A literature review, case studies, key informant surveys, and focus group discussions are the commonly used approaches. An *exploratory* research question generally poses questions like - What factors (such as age, education, income, etc.) cause something (people) to happen (migrate)? What are the reasons why some individuals migrate and others do not? What are the short-term and long-term effects of migration?

Descriptive Research

Descriptive research is used to describe a population, situation, or phenomenon under investigation. This design answers what, where, when and how questions. Both qualitative and quantitative methods, including a survey design, are used for describing the topic of concern. For *descriptive* studies, we pose questions such as: What is the prevalence of COVID-19 in a given setting? What is the food security status of individuals in a household in a given context?

Explanatory Research

An explanatory research design is used to answer 'why' about a question under investigation such as the cause-effect relationship between two or more variables. Experimental (when an experimental stimulus can be randomly assigned or manipulated) and quasi-experimental designs (where an experimental stimulus cannot be randomized) are the commonly used approaches to investigation. Longitudinal (and panel designs) are other examples. If the investigation is *explanatory*, a research question may seek answers such as - Do educated individuals tend to marry late compared to less educated individuals? What household assets (such as natural, human capital, financial capital, cultural capital, etc.) contribute to or explain household food insecurity?

Comparative Research

Researchers often attempt to answer questions comparatively. Comparative research aims to gather information on the differences between two or more research objects based on different variables. Answering such questions allows researchers to identify distinctive characteristics that distinguish one research subject from another. Comparative research questions seek to identify differences between two or more distinct groups, such as men and women, based on one or more variables (e.g., education, employment). Examples of comparative research questions are: What is the difference between men's and women's attitudes towards smaller family size in Nepal? Does migration pattern or remittance behavior vary by caste/ethnicity of individuals in Nepal?; or, Is there a difference in the voting behavior of elderly and young adults in Nepal?

Research questions also vary whether the investigation is cross-sectional or longitudinal (and/or panel) in nature. Comparative studies can be both cross-sectional and longitudinal.

Cross-sectional [Correlational] Studies

A cross-sectional study is generally referred to as a one-time study. Data is collected at one point in time and both dependent and independent variables are measured at the same time. In this case, it is difficult to delineate causality. Thus, the relationships between variables are interpreted as correlations or associations rather than cause-effect relationships. Some examples are: What individual factors are correlated /associated with individual-level outcomes? Does education positively contribute to or is related to/ associated with an individual's migration decision? What household factors are correlated with individual or household-level outcomes? Is land size (positively or negatively) related to/associated with an individual's (or a household's) migration decision? What community-level factors are correlated/associated with individual/ household/community-level outcomes? Is a household's distance from the nearest market positively or negatively associated with individual/household/community level food (in)security?

For *multilevel analysis*, we may pose questions as: What household-level factors contribute to (are correlated /associated with) individual-level outcomes? Or, what community-level factors influence household/individual-

level outcomes?

Longitudinal Studies (multiple time studies)

Longitudinal (or panel) studies utilize long-term data collected over a period. It is more convenient to ascertain a cause-effect relationship between variables. If a study is based on longitudinal (or panel) data, we pose questions as follows: To what extent do changes in individual-level factors/characteristics (age, education, etc.) influence changes in individual-level outcomes (income, migration, etc.)? To what extent do changes in household-level factors influence household-level or individual-level outcomes? To what extent do changes in community-level factors influence community-level or household-level or individual-level outcomes?

Features of a Good Research Question

A good research question has following features:

Clear. A good research question should be clear and concise and can be easily understood.

Relevant. It should be theoretically as well as practically relevant.

Feasible. Posing a question *per se* is not enough. It should be feasible in terms of resources, cost, and time. A research question can be both clear and relevant, but if it is not feasible to answer it, it has no meaning.

Measurable/researchable. A research question should be measurable or researchable. Measurement of concepts or constructs that make up a research question is a big challenge for researchers.

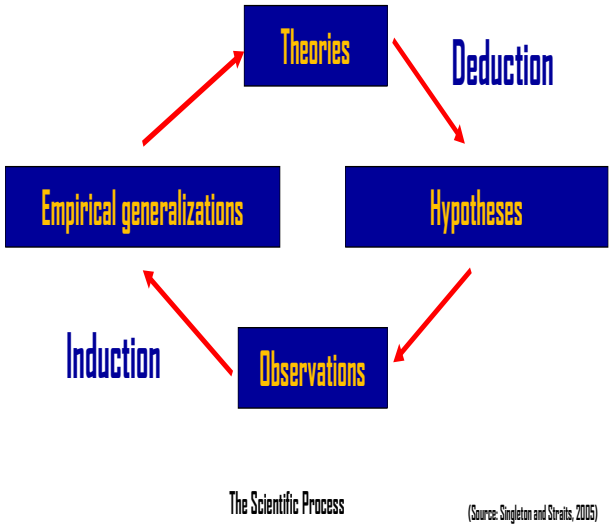
Based on theory or logic. It is convenient to measure a research question that is based on a theory or logic.

Specific not general. A research question should be specific rather than general in nature. A research question must avoid jargon. If there are complex concepts or constructs such as poverty, well-being, sustainability, etc., they should be operationally defined in a simple way so that they can be measured.

Sources of Research Questions

A problem is the starting point of any research question. A research question originates either from a theory (deductive approach of investigation) or from experience or observation (inductive approach of investigation). If a question under study originates from a theory, it is easier to pose a logical hypothesis based on the theoretical relationships between two variables as explained by the theory. However, if a research question originates from an observation, a researcher observes the problem in a given setting first and gathers empirical evidence. The individual observations are compiled and summarized to make empirical generalizations.

Answering a Research Question: An Example with Maryland Agritourism Survey Data



Below, I attempt to answer a question with survey data (For details about this specific example, you may refer to *Bhandari, P.B., K. Ejiogu, L. B. Karki, E. N. Escobar, N. N. Arbab, and M. T. Kairo. 2024. Factors Associated with the Profitability of Agritourism Operations in Maryland, USA. Sustainability, 16, 1025.* https://doi.org/10.3390/su1603102).

Background

Many small and medium farmers across the U.S. diversify their income streams by opening their farms to visitors for education, recreation, and entertainment – called agritourism. One of the goals of these agritourism operations is to increase the number of visitors to their farms to earn more income through direct sales. However, the understanding of the extent to which the number of visitors influences the profitability of an agritourism operation is limited. This brief paper is designed to address this gap.

Research Question. This example attempts to answer the following question: *Does the number of visitors influence the profitability of agritourism operators in Maryland, USA?* This data was collected in 2022 by the University of Maryland Eastern Shore Extension. I hypothesized that *an increase in the number of visitors increases the profitability of an agritourism operation, net of other controls*.

Measures

The research question points towards examining the relationship between the number of visitors and the

profitability of an operation. Thus, it is expected that the number of visitors will (positively) contribute to profitability. Thus, our outcome/dependent variable is profitability, and the independent variable is the number of visitors to the operation. There may be other factors that may affect the profitability of an operation. Thus, the effect of the number of visitors may be confounded by other variables, which we term as controls (please refer to Bhandari et al. 2024 for details).

Outcome/Dependent Variable

Self-perceived profitability. In the survey, the profitability of an agritourism operation was measured by asking, "Which of the following best describes the economic situation of your farm/agritourism business?" Four response categories were (a) profitable, (b) make some profit, (c) break even, and (d) operate at a loss. For analysis, the outcome variable perceived profitability as reported by agritourism operators is dichotomously coded as profitable (coded 1) vs. otherwise (coded 0).

Explanatory Measure

Number of visitors to the farm. This variable was measured by asking, "Approximately how many visitors come to your agritourism business yearly?" The respondents provided the guestimate of the number of visitors (a ratio scale variable). As the range of the number of visitors greatly varied (with a minimum of 50 to a maximum of 300,000) with a highly skewed distribution, this outcome variable was logged for analysis.

Controls[1]

Length of operation. It is expected that the length of operation may positively contribute to the profitability of a farm. The length of operation was measured in the number of years collected by asking, "How long has your agritourism operation been in business?" The response collected in the month was later converted to the number of years (a ratio scale variable). Another control is the geographic location of the operation. The profitability may also vary depending on where the operation is located. In the survey, the rural or urban location of the farm/enterprise was collected by asking, "Which would you classify the location of your farm/agritourism business as urban, suburban, or rural? The location was grouped into two categories (a nominal scale variable): (a) rural (coded 1) vs. (b) urban/suburban (coded 0).

Analysis

First, descriptive statistics is used to describe the distribution of interest variables. Second, means of interest variables by group were compared using a t-test to examine if the number of visitors and the length of operations (as both ratio scale measures) varied for profitable and otherwise agritourism operations. Similarly, a chi-square test was performed to examine the bivariate association between interest-dependent and independent variable geographic location (because both were categorical or measured on a nominal variables). Next, as the outcome measure, whether an agritourism operation reported profitable or not was measured as a dichotomy; a binary logistic regression technique has been used as a multivariate statistical technique. Results are provided as odds ratios, which can be interpreted as: if the odds ratio is greater than 1, the effect is positive, and every unit increases in the independent variable increases the odds of profitability. Conversely, if the coefficient is less than 1, every unit increase in the independent variable decreases the odds of reporting profitable operation.

Results

Descriptive Statistics

Distributions of interest variables are provided in Table 1. The results show that 35.8% of operations reported that they were profitable as compared to 64.2% of operators that reported otherwise.

Table 1 also reveals that the average number of visitors annually was 34,457 (logged 3.85), with a minimum of 50 to a maximum of 300,000 total visitors. The average length of operation was 18.68 years. In addition, 29.5% of the operations were located in urban/suburban areas and 70% in rural Maryland.

Table 1

Descriptive statistics of interest variables (n=67)

Measures	%/Mean	Standard Deviation	Minimum-Maximum
Outcome			
Self-reported profitability: Profitable	35.8		
Otherwise	64.2		
Explanatory measure			
Number of visitors (logged)	3.85	0.79	1.70-5.48
Number of visitors	34,457	56,453	50-300,000
Controls			
Length of operation (years)	18.68	14.85	2.17-60.00
Rural/urban location			
Urban/suburban	29.5		
Rural	70.1		
Total	100.0		

Bivariate Results

Table 2 provides the bivariate associations (t-test results) between self-reported profitability (a dependent variable), the number of visitors (logged), and the length of operation (number of years). In addition, this table also

1. For simplicity, I used only a few variables in this example.

provides the results of cross-tabulation with its Chi-square statistics between self-reported profitability and the rural-urban location of the agritourism operation.

Table 2

Bi-variate associations between dependent (self-reported profitability) and independent variables (n=67)

Measures	Profitable (%) (n=20)	Otherwise(%) (n=47)	Total (n=67)	Statistical Test
Explanatory measure Number of visitors (logged)	4.25	3.62	3.85	t=-3.397 p=0.002 (two-tailed)
Controls:				
Length of operation (years)	25.04	15.12	18.68	t=-2.747 p =0.021 (two-tailed)
Rural/urban location				
Urban/ suburban	30.0	70.0	100.0	Chi-quares=0.420
Rural	38.3	61.7	100.0	Degrees of freedom=1 p=0.517 (two-tailed)

Overall, the bivariate results (t-test results) show that those who reported their operation was profitable employed significantly larger number of employees (4.25 logged numbers) as compared to 3.62 (logged) employees (t=-3.397; p<0.01) for those who reportted otherwise. Regarding the controls, also as expected, those operations that were in business for longer time reported that their business earned a profit (t=-2.747; p<0.05). On the other hand, slightly larger proportions (38.3%) of operations located in rural areas reported profitability as compared to those located in urban/suburban (30%) locations. However, the result was statistically not significant.

Multivariate Results

Do the bivariate level results still hold true when taking into account of other potential confounders, here, the length of operation and rural/urban location of the operation? In other words, does the number of visitors significantly and positively influence the self-reported profitability of an operation, controlling the effects of other potential confounders such as the length of operation and the rural/urban location? A multivariate analysis goes beyond bivariate associations and examines the net effects of interest variables.

Table 3

Multivariate logistic regression results (odds ratios) predicting the relationships between dependent (self-perceived profitability) and independent variable (number of visitors) net of other controls (n=67)

Measures	Model 1	Model 2
Explanatory measure		
Number of visitors (logged)	3.434**	3.359*
Controls:		
Length of operation (years)	-	1.027
Rural/urban location		
Urban/suburban (Reference=0)		
Rural (=1)		2.623
Intercept	0.004***	0.001***
Model Chi-square	11.043	15.179
p	<.001	<.002
df	1	3
Adjusted R-square (Nagelkerke)	0.209	0.278

+p<.10, * p<.05, ** p<.01, ***p<.001.

Table 3 shows that the number of visitors (logged) statistically significantly increased the odds of reporting or experiencing a profitable business (model 1, a model with no controls)[2]. This is the empirical evidence (or the answer) to the above research question. Here, the results are not adjusted for other confounders (or are based on bivariate analysis). This result holds true after controlling the effects of the operation's length and rural-urban location (model 2). This is the empirical evidence (or the answer) to the research question posed above that adjusted for the effects of the length of operation and the rural-urban location of an operation (or is based on multivariate analysis). By adjusting other confounders, we gain more confidence in our results. For a detailed explanation, refer to Bhandari, (2024)[3].

Conclusion

2. Prior to interpreting the results, we should first examine whether the omnibus test (Model Chi-square) is statistically significant or not. If it is statistically significant, then we conclude that the model fits the data and we begin interpreting the results. If the model Chi-square is statistically not significant, then the model does not fit the data. Then, there is no need to interpret the results. Here, the (Model Chi-square for both models 1 and 2) are statistically significant. Next, we also look at the pseudo-R-square, percent of variance explained by the independent variable(s) in the equation. In Model 1, it is 20.9% and for the second model, it is 27.8%.

3. As the purpose was to provide an example of answering a research question with survey data, the interpretation of coefficients of other variables is beyond the scope of this paper. For a detailed investigation and more rigorous analysis results refer to Bhandari et al. (2024).

What is a research question? What types of research questions can be answered using survey data? These are the problems many scholars confront at the beginning of their scholarly careers. This brief paper is designed to answer several questions related to a research question and the use of survey data to empirically answer the question of interest. In addition, this brief paper provides a simple example of using survey data with a limited number of variables with results from univariate distribution (descriptive statistics) to bivariate and multivariate (e.g., binary logistic regression) analysis results. This brief paper will be useful to beginning or novice scholars in their early career stages. The findings from the statistical analysis presented in this brief paper should be considered for learning purposes only.

Declarations

Ethical Conduct of Research:
This writing follows ethical values.

Ethical Approval for the Research:
Ethical approval not required.

Conflict of Interest:
No conflict of interest.

Funding:
No funding.

Acknowledgement:
Feedback received at the initial stage of this article writing from Prof. Madhusudhan Subedi, Patan Academy of Health Science, Nepal were important in improving the quality of this paper.

References

Further Readings (some of the following literatures are not cited in the text):

Alford, R. R. (1998). *The craft of inquiry: Theories, methods, evidence*. Oxford University Press.

Babbie, E. (1990). *Survey research methods.* Wadsworth Publishing Company, Inc.

Bhandari, P. (2022). Levels of measurement: Foundational basis for quantitative analysis of survey data. *Dhaulagiri Journal of Sociology and Anthropology*, 122–126. https://doi.org/10.3126/dsaj.v16i01.50982.

Bhandari, P. B., Ejiogu, K., Karki, L. B., Escobar, E. N., Arbab, N. N., & Kairo, M. T. (2024). Factors associated with the profitability of agritourism operations in Maryland, USA. *Sustainability*, *16*(3), 1025. https://doi.org/10.3390/su160310252

DeMaris, A. (1995). A tutorial in logistic regression. *Journal of Marriage and the Family, 57*: 956-968.

Field, A. (2009). Discovering statistics using SPSS. SAGE.

Fink, A. (2009). How to conduct surveys: A step by step guide. Sage Publications.

Lavrakas, P. J. (2008). Research question. In Paul J. Lavrakas (Ed.) *Encyclopedia of survey research methods.* SAGE.

Ofem, B. I. & A. A. Mchi (2023). Variable conceptualisation and measurement in environmental Research. *Int. J. Methodol.* 2(1): 2–11. https://doi.org/10.21467/ijm.2.1.5991

Singleton, R. A. Jr. & Straits, B. C. (2005). Chapters 1-2, pp. 1-40 in *Approaches to social research*. Oxford University Press.

Stinchcombe, A. (1968). Chapter 2, pp. 15-56 in *Constructing social theories*. Harcourt, Brace and World, Inc.

About Author

Prem Bhandari (https://orcid.org/0000-0002-1705-8511) is a Social Researcher. He has completed Ph.D. in Rural Sociology and Demography from the Pennsylvania State University, USA; M.Sc. in Rural Development Planning from the Asian Institute of Technology, Thailand and B.Sc. in Agricultural Economics from Tribhuvan University, Nepal. Dr. Bhandari is working at the University of Maryland Eastern Shore, Maryland. Prior to this, he worked at the University of Michigan, Michigan, University of Alberta, Canada and Tribhuvan University, Nepal. He also worked with KonTerra Research Group in Washington, D.C. Dr. Bhandari is the founding General Secretary of the Association of Nepalese Agricultural Professionals of Americas (NAPA). He has over 25 years of first-hand experience in designing and implementing large scale social research.
Email: pbhandari115@gmail.com