

# Statistical Analysis Method in the Survey of Physical Social Science

## Xiuying Han

College of physical education, Shandong University of Finance And Economics, Jinan, Shandong, 250014, China

Keywords: Physical Social Survey, Questionnaire, Statistical Analysis.

**Abstract.** Physical social survey questionnaire aims to understand physical social phenomenon. The emphases of survey is statistical analysis. This paper summarizes methods of statistical analysis usually used in physical social survey questionnaire, and gives examples to make it clear.

### Introduction

Social survey is a form of social science research method, which applies questionnaires with the help of self-administered questionnaires and structured interviews, collect data from samples of the ensemble, and analyze these material to understand social phenomena and social rules. Questionnaires in physical social science are self-administered or of question and answer model aiming to comprehend social phenomena of sports an reveal laws governing those phenomena. Form of questions varies with different research questions and objects and questions can be categorized as unstructured questionnaires and structured one according to structures as well as open-ended questionnaires and closed one based on ways of answering questions.

The whole process of a social survey comprises three steps. First, survey design, which solves the questions of how to achieve research goals and how to get information. Second, collecting data. Field operations include interviewing every people of the selected sample and filling out the questionnaires. Third, quantifying and analyzing questionnaires statistically. This step is of primary importance in social survey and is also very complicated. Because total opposite results could be drawn based on different analysis method and understanding of data.

Questionnaire survey can be classified as qualitative and quantitative analysis according to levels of statistical analysis.

## **Qualitative Analysis**

Qualitative analysis is a kind of exploratory research, with the purpose of positioning questions, supplying underlying understanding of questions or defining and finding a solution to the question. Generally speaking, sample of qualitative analysis is of small size with no more than 30 units of data, hence veracity of its research results can be very subtle. In fact, as for qualitative analysis is greatly dependent on statisticians' judgment and their explanation of the materials. It is basically impossible to get the same results from the two researchers in the same study. So qualitative analysis demands high professionality from analysts and prefers statisticians participating in data collection.

## **Quantitative Analysis**

After the initial qualitative analysis of the questionnaire, quantitative analysis will further be applied. The questionnaire should first be quantified, then quantified data will be analyzed. There are simple and complex quantitative analysis according to complexity of analysis methods.

## **Simple Quantitative Analysis**

Simple quantitative analysis is to analyze the questionnaire results in a simple way, with the help of percentage, mean value and frequency. At the present paper, we'll discuss this way of analysis in the following three aspects.



**Quantitative Analysis of Radio Closed Questions.** Closed questions means answers to questions have all been listed by the designers and informants should choose only one answer. For example, through which channels do you get to know about major sports events at home or abroad (choose only one answer)?

- 1. TV
- 2. Radio
- 3. Newspaper
- 4. Magazines
- 5. the Internet
- 6. Talking

We can easily count up the times of choosing each choice in all 60 answers: namely TV=30, Radio=8, Newspaper=6, Magazines=4, the Internet=10, Talking with others=2. The results will be presented in Table 1.

Variables Types	Variable Values	Frequency	Percentage	Cumulative Percentage
TV	1	30	0.50000	0.50000
Radio	2	8	0.13333	0.63333
Newspaper	3	6	0.10000	0.73333
Magazines	4	4	0.06667	0.80000
the Internet	5	10	0.16667	0.96667
Talking	6	2	0.03333	1.00000

Table 1. Main channels for learning about major sports events at home and abroad

Analysis results are displayed clearly in Table 1. Nearly half of the informants get to know about major sports events by watching TV. Table 1 analyzes the whole sample. However, nearly all the questionnaires require comparison between different groups of informants. In this condition, we should employ more complex method, cross-over analysis, which analyze relations between three variables. For instance, in a study about health monitoring in the survey of exercise events, it is found that female martial art exercisers are apparently much more that male ones, which is seemingly inconsistent with the reality. But after the introduction of the variable "age" as the third one, we find that male exercisers under 40 years old accounts for 77.2% of the total, while male exercisers above 40 cover 28.9%. This result is contrary to its female counterpart's, hence when all the data are mixed up and analyzed, we can hardly identify relationship between "sex" and "number of martial art exercisers". When we categorize the group with different age, that implicated relation can be clearly Tabled out, as is displayed in Table 2.

Table 2. Martial art exclusions grouped with age and sex	Table 2. Martial	art exercisers	grouped	with	"age"	and	"sex"
--	------------------	----------------	---------	------	-------	-----	-------

Age	Sex		Total of rows	Number of cases
	Male	Female		
Mixed up	39.2%	60.8%	100%	2215
Under 40	77.2%	22.3%	100%	376
40 and above	28.9%	71.1%	100%	1839

We can clearly understand the power of cross-over analysis from Table 2, and this method can analyze relations between various variables, for instance we can compare data from two more aspects like "profession", "income".

**Quantitative Analysis for Multichoice Questionnaires.** In the present social surveys there exists a lot of such variables as their values are related to several values. For example, when asking which color you like, you may choose red as well as blue and green. If you are asked to sequence the chosen colors in order of preference and your answer is first red, second blue and third green, then it will compromise several choices for the same questions, which are multichoice questions. Now in many sports social surveys of different fields we would face the following problem, i.e. in some city's health monitoring, in order to discuss the affect of lifestyle on health, there is a question: what do you usually do in your free time (choose three items and sort them in order of frequency) and the choices are 0 nothing, 1 physical exercises, 2 playing cards and chess, 3 travel, 4 partying, 5 audiovisual entertainment, 6 doing housework, 7 catching up on sleep, 8 not listed above. We



process the data via SPSS, and results are as listed in Table 3 to Table 5.

Variables	Value of variables	Number of cases	percentage	Cumulative percentage
physical exercises	1	547	27.4%	27.4%
playing cards and chess	2	73	3.7%	31.0%
travel	3	148	7.4%	38.4%
partying	4	173	8.7%	47.1%
audiovisual entertainment	5	532	26.6%	73.7%
doing housework	6	413	20.7%	94.3%
catching up on sleep	7	43	2.2%	96.5%
not listed above	8	71	3.6%	100.0%
	Total	2000	100%	

Table 4. Frequency distribution table of taking physical exercises as the second choice

Variables	Value of variables	Number of cases	percentage	Cumulative percentage
Nothing	0	345	17.3%	17.3%
physical exercises	1	72	3.6%	20.9%
playing cards and chess	2	30	1.5%	22.4%
travel	3	83	4.2%	26.5%
partying	4	155	7.8%	34.3%
audiovisual entertainment	5	568	28.4%	62.7%
doing housework	6	531	26.6%	89.2%
catching up on sleep	7	154	7.7%	96.9%
not listed above	8	62	3.1%	100.0%
	Total	2000	100%	

Table 5. Frequency distribution table of taking physical exercises as the third choice

Variables	Value of variables	Number of cases	percentage	Cumulative percentage
Nothing	0	898	44.9%	44.9%
physical exercises	1	60	3.0%	47.9%
playing cards and chess	2	14	0.7%	48.6%
travel	3	45	2.3%	50.9%
partying	4	85	4.3%	55.1%
audiovisual entertainment	5	192	9.6%	64.7%
doing housework	6	301	15.1%	79.8%
catching up on sleep	7	279	14.0%	93.7%
not listed above	8	126	6.3%	100.0%
	Total	2000	100%	

Table 6. Frequency distribution table of variable sets of multi-response classification

Variables	Value of variables	Number of cases	Percentage Of	Percentage of
			total answers	observed Quantity
Nothing	0	1243	20.7%	62.2%
physical exercises	1	679	11.3%	34%
playing cards and chess	2	117	2.0%	5.9%
travel	3	276	4.6%	13.8%
partying	4	413	6.9%	20.7%
audiovisual entertainment	5	1292	21.5%	64.6%
doing housework	6	1245	20.8%	62.3%
catching up on sleep	7	476	7.9%	23.8%
not listed above	8	259	4.3%	13.0%
	Total	6000	100%	300.0%

As we can see from Table 3, for young people of the investigated city, taking physical exercises (27.4%) and audiovisual entertainment (26.6%) are the two major activities in their free time, with doing housework being the third most popular activities. These three choices covers 80% of the overall choices.

Table 4 implies that the choice ratio of audiovisual entertainment and doing housework are very close, being 28.4% and 26.6% and Table 5 implies that nearly half of informant choose "Nothing"

as their third choice.

And according to the data in fourth row of Table 6, adults of this city who choose audiovisual entertainment and doing housework accounts for 21.5% and 20.8%, and those who choose doing nothing also covers 11.3% in physical exercises. In a word, their free time is relatively dull.

**Quantitative Analysis with Answers of Table.** Tables are the answers for the survey questionnaires. For example, as to the question "how much money do you spend on physical exercises annually?", it is better interval processing the quantified data. The division of interval range is greatly determined by professionality of the researchers. When we use intervals to display distribution of data, we can also employ various statistics to describe results including measurement of value, the mean, median, value of the highest frequency or determination of dispersion as well as range, interquartile range, and standard deviation.

The above mentioned three methods are simple questionnaire analysis. It is of great pity if only simple method in employed to process data, as it will bring about great loss of data, increasing strategic risks, and shallow analysis results.

#### **Complex Quantitative Analysis**

Simple analysis is often employed in analyses of single variable or two variables. Social economic phenomena are complex and changeable and two or three variables are far from satisfactory. In this situation, complex quantitative analysis is needed. There are two common kinds of complex quantitative analysis, namely multivariate analysis and orthogonal design analysis.

**Multivariate Analysis.** Through analysis of observed data, this kind of analysis is to study rules governing interdependence between several variables and evaluate, categorize and judge research objects according to requirements of real problems. In that process, some public factor which are more general or more abstract influence various indicators. There are three types of multivariate analysis commonly used in questionnaire analysis: cluster analysis, principal component analysis, factor analysis.

1) Cluster analysis: there are two kinds of cluster analysis, sample clustering and variable clustering. The main purpose of sample clustering is clustering informants' opinions towards some questions into similar groups according to some rules and analyze these opinions with the help of those groups. Variable clustering aims to choose some variables and do research on some aspects of a phenomenon with the help of SPSS. Variable clustering is very common. Here we will give you an example to describe the role of the variable clustering. For example, through survey we get heptathlon scores of each event of 30 excellent track and field athletes, and analyze track and field heptathlon events with variable clustering namely 1, 100-meter hurdle, 2, high jump, 3, shot put, 4, 200-meter, 5, long jump, 6, javelin throwing, 7, 800-meter. Cluster analysis was carried out for these 7 variables, and analyze them with SPSS, and the get the result:

- 1. 100-meter hurdle, 200-meter, long jump
- 2. shot put, javelin throwing
- 3. high jump
- 4. 800-meter

We can categorize variables through clustering analysis in analyzing questionnaires. For instance, we can classify "1" as "speed events", "2" as "casting events", "3" as "bouncing events", and "4" as "endurance events" so as to be clear about informants advantages and disadvantages events. The greatest strength of clustering analysis is that it is easy to operate and the weakness is that there being no integral theoretical system, so it is difficult to explain.

2) Factor analysis. There are always some basic quality like "attitude", "understanding", "hobby", "ability", "intelligence" which are hard to observe directly. Factor analysis is such a tool that it utilizes these latent variables explaining observable variables.

When researchers design the questionnaire, they actually suppose there existing some structure, and through factor analysis they can test the validity of the hypothesis. The main function of factor analysis is to isolate or extract some public factor from variables in scales. Relationship between those factors and other variables can be studied, usually complex mathematical method is applied to

categorize observed quantities of high correlation into several groups according to some rules. Each group of variables share a public factor; in other words, that variable group is highly correlated with the public factor reflected by The "load" on this factor being obviously large (the absolute value), and these public factors represent basic structure of the scales. We can quantify these factors based on the situation of respondents and analyze them. For example, with the help of factor analysis, we analyze different condition of promoting "healthy economic" development of 31 provinces. 6 typical indicators are selected: 1, GDP per capita (yuan/person), 2, urban PCDI (yuan/person), 3, net income per capita (NIPC)for a rural household(yuan/person),4,number of institutions of higher education, 5, number of health care institutions, 6, per capita public green area (square meter/ person). SPSS analysis results are as follows:

Variables	Variable value	First main factor F1	Second main factor F2	Third main factor F3
GDP per capita	X1	0.942	-0.0260	-0.103
urban PCDI	X2	0.943	0.0258	0.102
NIPC for rural household	X3	0.848	0.211	0.217
number of institutions of higher education	X4	0.3.1	0.836	0.0436
number of health care institutions	X5	-0.129	0.903	-0.0711
per capita public green area	X6	0.0827	0,0339	0.989
Cumulative effectiveness		45.655%	70.486%	87.049%
Names of factors		Factor of net economic income	Factor of social development security	Factor of better environment

Table 7. Factor analysis of healthy economy indicators

Table 7 indicates that the first public factor F1 has high load indicators reflecting different development degree of economic income in the investigated provinces; hence F1 is defined as factor of net economic income. The second factor F2 is highly loaded with indicators reflecting investment and construction of various provinces for meeting the increasing demand of education and medical services. It indicates degree of government's attention on educational and medical construction, so we name it factor of social development security. The third factor reflects in some degree great effort of various provinces on improvement of public environment and community verdurization. It's effectiveness of contribution is 16.563% and cumulative contribution of common factors is 87.049%, which can sufficiently explain development motive of healthy economy generally.

Judging from the above example, the advantage of factor analysis lies in generalizing information as much as possible with one or several comprehensive indicators so as to better summarize research questions and explain more general features and laws.

**Orthogonal Design Analysis.** Besides factor analysis, there is another more complex questionnaire analysis: orthogonal design analysis. The process is as follows:

First orthogonal design analysis is conducted to analyze the questionnaire and try to make it well organized. Secondly, make questionnaire investigation and collect data. Thirdly, collect data and analyze the survey results (analysis of variance in orthogonal experiments can be applied). At last obtain analysis results.

There are many advantages for orthogonal design analysis: first is saving the cost. Comprehensive information can be obtained from limited data. Second, tables are used from design of research to analysis of questionnaire structures. Tables are the most convenient and effective way in multi-factor investigation. Third, orthogonal design analysis is very reliable, and problems selected through orthogonal tables are highly representative.



#### How to Choose Statistical Analysis Method and Problems to be Noticed

#### How to Choose Analysis Method

It is not an easy thing to choose the correct way of questionnaire analysis. Generally speaking, two points need to be borne in the researcher's mind. First, if we only need to discuss the research question in a simple way, qualitative analysis and simple quantitative analysis can be employed. Second, if we need to do the research in a deeper way, and explore the nature of the issue, then complex quantitative analysis is necessary. In complex quantitative analysis, factor analysis is more suitable for attitude questionnaire. If the problem choice needs to be sorted out, it is more appropriate to use cluster analysis. Orthogonal design analysis is used for multiple factors investigation. Of course, these methods can be used in a cross way most of the time. In fact, the proper selection of questionnaire analysis method depend on experience and professional knowledge of the analysts.

#### Problems to be Noticed in Choosing Method

In fact, qualitative and quantitative analysis cannot be completely separated in sports investigation. They are often combined to make them cooperate with each other so as to acquire more accurate, comprehensive and detailed findings. As the paper[1] shows, "one of the important bases of probability design is data and experience judgment and probability statistical theory only provides tools for more detailed analysis and better use of existing data. And create better conditions for people in making accurate judgment, rather than denying experience." Therefore, qualitative analysis should be used to assist and supplement inadequacy of quantitative analysis.

#### References

[1] Bao Chenggang, Discussion on some basic problems in probability analysis methods on geotechnical engineering, Rocks Journal of geoengineering. 1989,1(4).

[2] Yang Miaojie and Jiang Xinghong, Using factor analysis to analyze the development of China's healthy economy Status, Statistics and information Forum. 2003,18(2).

[3] Ke Huixin, Shen hao, Statistical analysis in investigation and research. China Media university press, Beijing, 2005.

[4] Yang Zixu. Construction of orthogonal tables. People's publishing house. Jinan, Shan dong, 2003.