IAOS Satellite Meeting on Statistics
for the Information Society
August 30 and 31, 2001, Tokyo, Japan

Key Findings of statistical data survey via Internet

OKUMURA, Fumitaka
Jupiter media Metrix Senior Analyst, Japan

Abstract

The Internet has become a part of social infrastructure in Japan. Today more than 60% of business communication is exchanged via e-mail in Japan. It is estimated about 26.6 million people access Internet by PC and more than 36million by Internet-enabled phones like NTT-DoCoMo i-Mode. (2001.4)

With this explosive penetration of Internet access into Japanese society, data survey via Internet is now becoming more and more popular these days. Online survey, or survey research via Internet, has evident advantages over traditional survey method ; Speed and Cost. However there are 2 challenges on online survey.

This paper raises and discusses a few points for conducting data survey on Internet. In terms of the data accuracy of the online survey, the bias and penetration problems are picked up in first section. Another problem, a huge difference in Internet access environment between US and Japan will be discussed in the next section.

Along with the dramatic advances in the communication technology and the online user demand, the pressure to utilize internet for official statistics seems to increase more and more in almost every country. It would be a great pleasure if what is discussed here in this paper be of some help for everyone to be better prepared for possible IT adaptation for their marketing or researching effort by collecting or using statistical data via Internet on his or her own statistics.

1 Introduction

The Internet is penetrating deep into social life even in Japan. The Internet is also evolving into a major media platform and will permanently alter the entire media landscape. More specifically, as consumers spend more time online-currently, almost 11 hours a week-they are dispensing the time, for the most part, that they formerly spent watching TV.

As a tool for communication, Internet is far smarter than traditional tools like paper or fax, in that by Internet we can communicate interactively, and therefore more effectively using not only texts but also multimedia tools such as voice, pictures, movie clips etc.

No wonder that today more and more research companies are using Internet in their survey projects. Let us call collecting statistical data via Internet "Online survey", and let us define collecting data by traditional methodology using mail, fax, phone, as "Offline survey". Speed and cost are two strong incentives for companies to conduct online survey, now becoming more and more popular. Here is one example that shows how online survey is superior to tradition one in speed and cost.

Table. 1 An Example for Chart of cost, speed in Online survey and Offline survey

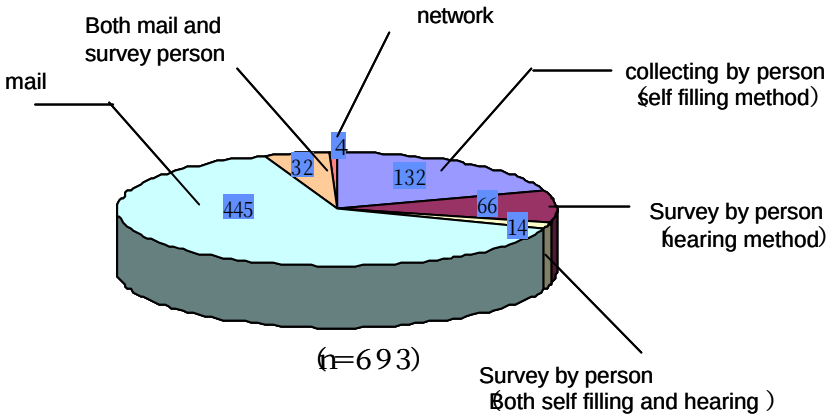
N=500	Mail (paper)	Telephone (CATI)	Internet (Web)
Cost	50(1000$)	50(1000$)	10(1000$)
Lead time	6 month	3 month	2 ～ 3 month
Penetration rate	100 (%)	99 (%)	25 (%)

Source: Jupiter Media Metrix (2000)

Although the pressure to use Internet for statistics data survey will increase, few current Japanese government statistics use online method.

A data shows the Japanese government statistics method. (**Fig. 1**) The data discloses the high-cost structure of Japanese government statistics. According to this data, only 0.5% government official statistics use Information Technology.

Fig. 1 Variation of data collecting method of survey for statistics data



Source: NTT-data Institute of management consulting (2000)

But, it is impossible to use online method for all statistics survey is absolutely true. The problem in terms of the data accuracy of the online survey is its penetration and bias.

The penetration of these Internet accessible devices is obviously low compared to that of telephone, which enjoys almost full penetration into Japanese society. In the first semester of 2001, the penetration rate of PC to individual users is estimated at about 25% in Japan, in sharp contrast to 50% in US. According to Jupiter Media Metrix forecast, the penetration rate of PC in Japan will increase to 50% in 2005, narrowing the gap with US, where the rate is expected to be around 60% in 2005.

With this penetration rate, it will not possible enough to conduct a full-scale statistics survey like population census solely on Internet platform.

However, Singapore used Internet for population census partly in 2000 and minimized their cost dramatically. This indicates there is possibility to utilize Internet for statistics survey as a combination tool even today.

2 The complexity to use Internet for data survey

Here we face up to the first problem, the bias problem; by online survey can we keep data accuracy as well as an offline survey? Is there not any bias in there? The concept of bias of statistics is abstract, and it might be quite difficult to tell whether any survey result is biased or not in any sense. While the efforts to define or elaborate on it are important, our case may be also useful for apply information technology and improve the quality of official statistics.

Our firm, Jupiter Media Metrix services Internet data survey and analyze. From 1999, the needs of see and analyze Internet trends were increased.

To provide a solution, we developed measurement tool software, the Real-time Meter, which install in each individual or home sample members' computer operating systems and is engineered to measure usage of virtually any new technology, digital medium or platform. We collect users log data and analyze

online behavior and trends, perspective, the entire digital landscape, including its largest segments: the World Wide Web, proprietary online services (like America Online), and instant messaging and other digital media applications, as well as the software and hardware people own and use.

On these processes, we had accumulated the knowledge and skill to realize and keep strict accuracy on the web survey. There are several findings essential for conducting accurate data survey via Internet.

The first challenge is how to define online users. To date, there is no truly reliable or standardized method of gathering demographic data for online users. We had bad experience when we outsourced a specific online survey to a research company, to find it was far from reliable data without raw data.

We use Sample-based measurement methods for a statistical approach. To prepare a representative sample of the online users as a Web universe, we conduct sample survey independently. The size of the total audience is extrapolated from the sample. The tracking data is collected and extrapolated to represent the current Web universe.

Data integrity requires rigorous sampling methods and innovative measurement technology. We use panel recruitment methodology, Random Digit Dial (RDD). Starting with a pure RDD sample, we take advantage of known addresses to recruit by mail in addition to telephone. To ensure accuracy, the Meter is made on multilingual basis, recording usage of the global digital universe in each user's local language.

The Meter systems have distinct advantages. This system have attached valuable demographic information to their raw measurement data; collected 10,000 users log data per month.

3 Draw statistical data through internet in Japan

Another important factor was how to manage the gap between US and Japan. In Japan, there were 3 big unique characteristics when compared to US.

(1) Most of the Internet users access to WWW in so-called “Always on” environment in U.S, but most of the Japanese users access by “Dialing Up”. In other words US Internet users can access Internet by paying the same monthly charge to telecom companies regardless of how much they connect to Internet. But in Japan, users are charged by NTT or other telecom companies, usually in accordance with how many hours they used telephones line to connect to Internet.
Such “Dial Up users” were nervous about network cost, as telephone fee charged on them could be huge enough when they access WWW as much as they want.

(2) In U.S, the data survey via telephone (CATI) is very familiar, but data survey by telephone is not so common in Japan. Not little Japanese people hesitate or refuse talk over telephone even is several minutes. This is absolutely differs to U.S and Europe. As a result, we invested much to recruit individual sample members' in these circumstances.

(3) In U.S, Most of the Internet users access Internet via PCs, but Japanese use browser-phones as well as PCs, and not small number of Japanese access Internet only by browser-phone. To cover whole Internet activities, surveying only PC users is not enough in Japan. Further more, when IPV6 will prevail, we must pay attention to digital web home device.

4 Toward the Broadband generation

While consumers are just beginning to learn about home broadband options like FTTH, DSL CATV, many online consumers will make the shift from dial-up to broadband.

We define broadband access as Internet connection characterized by persistent (always-on) connectivity, and connection speeds of above 500 kilobits -1.5 Mega bits per second (kbps) and higher in

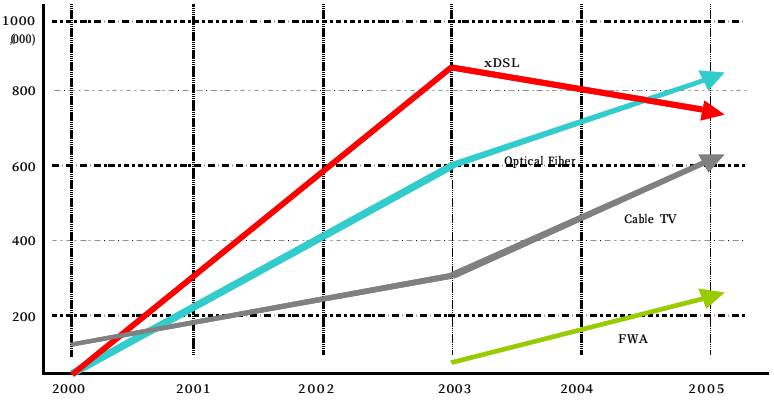
the downstream direction.

Our forecast shows broadband users will explosively increase; at least 2million Internet users will apply broadband access in the end of 2002. The important point of broadband trend is that users' behavior change. Users will be able to use Internet without minding their network cost. On this sentence, (1) network cost problem can be ignorable.

And further more, many kinds of broadband services will come. One example is IP videophone. People will be able to chat through Internet with each other, literally face to face on the PC screen live. This service would be useful to identify respondent, or chat knowing exactly with whom one is talking, and solve (2) problem.

The generation is coming who prefer responding survey via online to offline.

Fig. 3 Broadband access forecasting



Source: Ministry of General affair 2001