SAMPLING PROCEDURES

In the previous chapter we already learned about how to conduct empirical testing to test a hypothesis in our research by using an experiment. Anyway, you know that when we conduct an experiment, we have to recruit subjects for this task. But what are criteria that we can use to justify who should be selected and how we select them. The main idea of this chapter is about sampling procedures which is the way we select appropriate subject to participate in the study. Detail about the sampling technique will be discussed in the next sections. First, we have to understand what is sample. After that, we will see the steps involved in the sampling process, what are different sampling techniques we can use, and what are some biases that may present in sampling.

Why we need a sample?

Before we get start, it is important to understand the difference between a population and a sample. A *population* refers to a group of objects, people, or entities that have something in common. For example, the population of Thai people encompasses everyone who shares the same characteristic in terms of nationality; in this case, everyone who holds Thai nationality belongs to Thai population. A term population can also be applied to a sub group in which members have common characteristics, for example, a population of small and medium firms, a population of stock traders, a population of salespersons, etc. In particular, when researchers collect data from every single member within the population, they conduct what is called a *census*. Some examples of census are: when a university collects contact information from every enrolled student or when the federal government collects revenue information from all registered firms in the country.

Although data collected from the entire population tend to offer a complete view of information, it is not quite practical due to several reasons. In particular, collecting data from the entire population is difficult, expensive and time-consuming. For example, if Thai population is a main focus in a research, imaging how much time, money, effort, and manpower will be required to collect data from every Thai

citizen from every single area in Thailand. Using the entire population also does not make sense in some experimental research that requires the destruction of test units. For example, when a car manufacturer came up with a new model of vehicles, they must conduct a crash test experiments to observe safety features of a vehicle when it hits a solid concrete wall and various types of barrier. Destructive testing like this only requires a few vehicles in the inventory for the test. It does not make sense to crash test the entire inventory. They assume that a few vehicles that are randomly selected will have similar characteristics and can be the representatives of all vehicles in the entire inventory. What happen to the vehicle sample in the test should also be the same for other vehicles in the inventory.

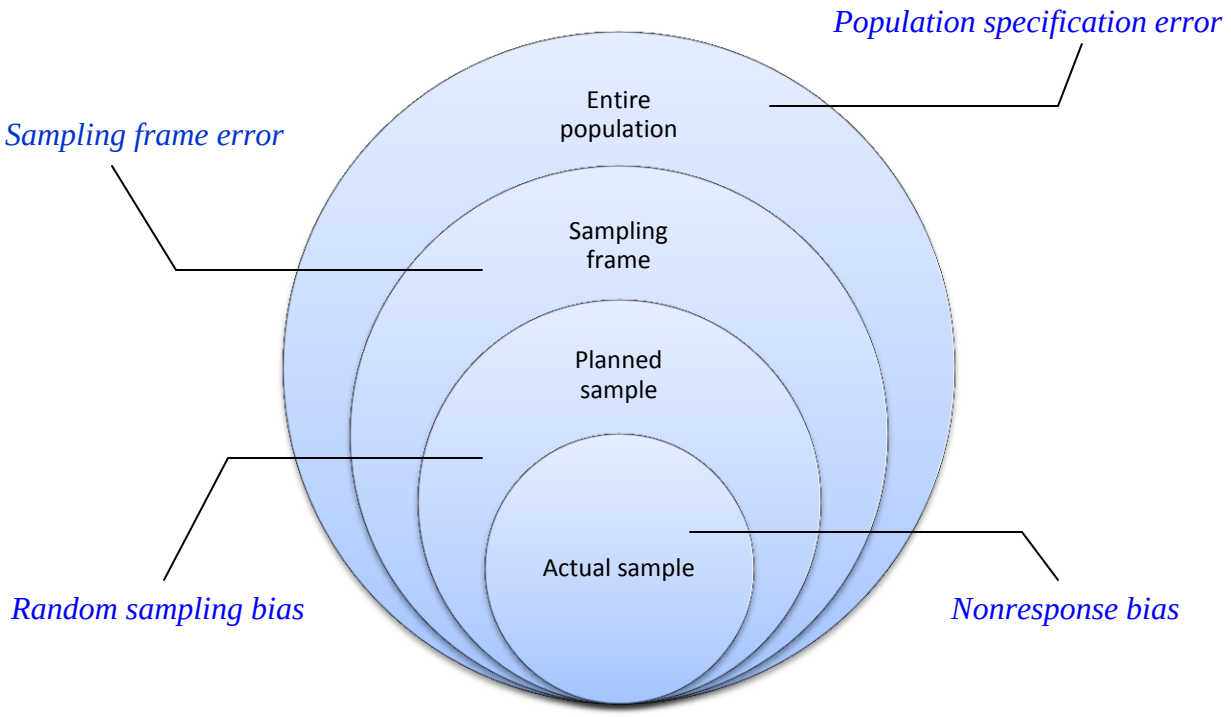
In research, it is widely accepted that data obtained from a proportion of the population, if randomly selected, can be the accurate representative the population of interest. *Sampling* can be defined as the process of by which small unit of objects, people, or entities are selected from the entire population. The small unit of objects, people, or entities that are selected from a population is called *sample*. When a sample that is randomly selected is used in an experiment, the results that are obtained from an experiment using that sample can be generalized back to the population from which they are selected. Going back to the example of experimental research mentioned in the previous chapter, the researcher who aims to investigate whether more homework assignment will affect class performance of students may 'randomly' select a group of students from the entire school as a sample. After the results are obtained (for example, more homework assignment is found to increase class performance of the student sample), the researcher can generalize the findings back to the population of students in that school (for example, to conclude that more homework assignment can increase class performance of all students in that school as well).

Anyway, it is important to note that the external validity of the results that are inferred from a sample to a population tend to be limited to the population from which a sample was drawn. For example, if a student sample was obtained only from a single school, the external validity of the results tends to be limited to that school. When you only use a sample from one school and then claim that the results you obtain should apply to other schools, the external validity of the results can be questionable.

SAMPLING PROCESS

There are several steps involved in the sampling process. These steps, as presented in the following figure, include (1) identify a target population, (2) select a sampling frame, (3) select a planned sample, and (4) obtaining actual sample.

Sampling process and some possible types of error/bias



Identify a target population

The beginning of the sampling process starts from identifying a *target population*. Generally, a target population is identified based on the characteristics of people or entities that match with the objective of the study. For example, when working on a research topic relates to consumer behaviors in online shopping, the characteristics of a target population should be people who have experience in online shopping. In this case, the population of people who regularly purchase products online tends to be appropriate for the study.

However, when the target population is inaccurately identified, the sample that we obtain in this step will suffer from *population specification error*. Referring to the example of research topic about consumer behaviors in online shopping, population specification error may present in a sample if the researcher target on the population of generation X consumers who may like to buy from traditional stores rather than from the Internet.

Select a sampling frame from a target population

After the target population is identified in the first step, we may start to collect data from them. However, there are some difficulties that a researcher may experience. First, the target population that we identify may be too large for us to collect data from all of them. Think about the population of online shoppers; imagine how many people are within that population; it seems to be very huge; probably a hundred thousand or a million. In addition, the contact information of everyone in the population may not be available for us, making it even difficult to gain access to every member in the target population that we identify. For example, although we know that there are hundred thousands of online shoppers in a country, we may not have mailing address or email addresses of everyone to contact them; we may not be able to arrange meeting with everyone in the entire population.

For this reason, the researcher can select the subset of the population as the representative in the second step. The subset that we select directly from the population is called a *sampling frame*. In practice, a sampling frame is a list of people or entities in the population that the researcher *can gain access to them, or can contact them*. Using the same example, if a target population of the study is online shoppers, it would be difficult to have all contact information of everyone who shops online. However, there are some sources that the researcher can obtain the list of consumers that are classified by the product categories they regularly purchase, along with their contact information. Unfortunately, the data like this hardly come for free. In serious marketing research, for example, the researcher may purchase the database that contains the mailing lists and shopping data of consumers who regularly shopped from the online stores (Ellis, 2011). Of course, the database like this may not cover the entire population of online consumers, but

people in the lists can represent a subset of the entire population that the researcher can gain access to, and that is considered the sampling frame of the study.

However, when in reality the sampling frame that the researcher select does not appear to be a true representative of the population of interest, *sampling frame error* will present in the data. For example, if it turned out that in reality the majority of consumers that are listed in the database did not shop actively from the Internet, bias caused by sampling frame error would present in the sample because they did not accurately represent the population that the researcher wants to focus.

Select a planned sample from a sample frame

In some case, the amount of people or entities included in the sampling frame is still too large for the researcher to collect data from all of them. For example, the lists of online consumers in the database that the researcher obtained may contain the amount of consumers which is too large to be covered by the research budget. For this reason, the researcher can randomly select another subset of people or entities from a sampling frame to scope down its size to the level that can be surveyed. In this step, the subset that is randomly drawn from a sampling frame is called a *planned sample*. As its name implies, a planned sample is a representative of a population that the researchers plan to collect data from them. In survey research, a planned sample is the group of people whom a researcher will contact them to complete a questionnaire survey.

When selecting a planned sample from a sampling frame, it is necessary that people or entities have to be selected randomly to avoid *random sampling bias*. Without random selection, it can possibly lead to the problem of *undercoverage*, which happens when some entities are inadequately represented in the sample (Squire, 1988). When random selection is performed, every people or entity in the sampling frame will have equal chance to be selected. This method not only prevents subjective bias caused when the researcher uses own judgment to select a sample, but it also prevents the problem of *self-selection bias* that is caused when respondents volunteer for a study. Detail about random sampling methods and self-selection bias will be discussed later in the chapter.

Referring to the example research, if the list of consumers in the database is too large for the researcher to collect the data from all of them, the researcher may decide to select a certain amount of sample out of the total amount in the sampling frame. For example, the researcher may decide to randomly pick only one thousand customers from the sampling frame. These one thousand customers that are randomly selected are considered a planned sample of the study. All people included in this planned sample are considered the representatives of the entire population that will be invited to participate in the survey.

Obtain actual sample

In survey research, after the planned sample is randomly selected from the sampling frame, the researcher can begin surveying people who are included in the planned sample. At this step in survey research, the questionnaires will be distributed to people in the list. However, in reality not everybody in the planned sample is willing to participate; some people may refuse to provide information; some may just ignore the survey. Therefore, the final amount of people who actually provide usable information to the researchers is called an *actual sample*. In practice, the data obtained from an actual sample is the actual data that the researcher will use in data analysis.

In particular, people in a planned sample who refuse to provide usable information or refuse to participate in a study are called *nonrespondent*. The total amount of nonrespondent represents the *nonresponse bias* that may inherit in the data. The reason why nonrespondent is considered a source of bias is because the absence of information from these people may alter the results that researchers would get when complete information was obtained (Groves, 2006). In survey research, the nonresponse bias can be assessed by a response rate that researchers obtain from a survey. A *response rate* is simply calculated by dividing the number of people who answered the survey by the number of people who are in the planned sample (Hawkins, 1975). A response rate is usually expressed in percentage. For example, if the researcher obtained seven hundreds usable surveys out of one thousand surveys that he sent out, he would get the response rate of 700 ÷ 1000 = 70 percent. In particular, the response rate has been proposed as the indicator for the quality of survey data (Atrostic et al, 2001; Babbie, 1990; Rea & Parker, 1997).

The higher the response rate, the lower the chance of nonresponse bias that might present in the data. For this reason, Alreck and Settle (1995, p. 184) recommend the researchers "to do as much as possible to reduce nonresponse and encourage an adequate response rate."

$$\text{Response rate} = \frac{\text{The number of people who answered the survey}}{\text{The number of people in the planned sample}}$$

The question that you may be wondering is what should be the amount that is considered a good response rate. To date, there is no consensus about the threshold. Still, some scholars have proposed the minimum requirement that the researchers need to achieve in social research. For example, Babbie (2007, p. 262) suggested that "a response rate of at least 50 percent is considered adequate for analysis and reporting. A response of 60 percent is good; a response rate of 70 percent is very good". On the other hand, Singleton and Straits (2005, p. 145) argued that "for interview surveys, a response rate of 85 percent is minimally adequate; below 70 percent there is a serious chance of bias".

Despite these minimum requirements suggested by the scholars, the optimal amount of response rate may still depend on the context and type of data collection. For example, data collection from employees in organizations by using questionnaires that are distributed and are collected back in person may yield higher response rate than when the questionnaires are sent and returned by post or by email. In a study that aims to collect company-level data (e.g., surveys are mailed to the executives of the companies, asking them to complete the survey and then mail it back to the researcher), the response rate is normally lower when compared to the survey that aim to collect individual-level data (e.g., surveys are distributed to employees in person and tend recollected by the researcher). In the case of company-level survey, the response rate of about 10-20 percent is considered acceptable in many leading journals.

Moreover, response rate will also be affected by the characteristic of the survey questions. For example, survey topic that contains sensitive questions tend to make a response rate to be lower (Dommeyer, 1985). On the other hand, the survey topics that the respondents are interested tend to obtain higher response rate (Lund & Gram, 1998). The length of the questionnaire is also important, as lengthy questions require more cognitive efforts of respondents to answer them, thereby discouraging respondents from answering the survey (Childers & Ferrell, 1979). For example, Jepson et al (2005) shows that response rate in their pilot study decreased from 60 percent for questionnaires 849 words in length to 16.7 percent for questionnaires over 1,800 words in length. By comparing response rates between three-page and one-page questionnaire, Robert et al (2003) found that the response rate is nearly double in a short version.

Boosting response rate in mail survey

Anyway, some scholars also suggested the technique to increase response rate in mail-survey. The most common techniques are follow-up mailing and monetary incentives (Fox et al, 1988; Yammarino et al, 1991). For the first technique, follow-up mails are sent to the respondents a few days or weeks after the surveys were mailed in order to remind the respondents of the survey. Another follow-up mail can also be sent again if a survey is not returned back after a certain period of time. In addition to following up mailing, researchers can also provide monetary incentives (e.g., cash reward, gift card, discount card, etc.) to persuade respondents to complete and return a survey. Some scholars suggested using lottery incentive in which respondents who return the survey will have a chance to get a grand prize from lottery drawing (Kalantar & Talley, 1999). Anyway, incentives do not necessarily have to be provided to respondents in terms of rewards. Some studies promised to donate a certain amount of money to a charity for each survey that is returned to the researcher (Charoensukmongkol, 2014a, 2014c).

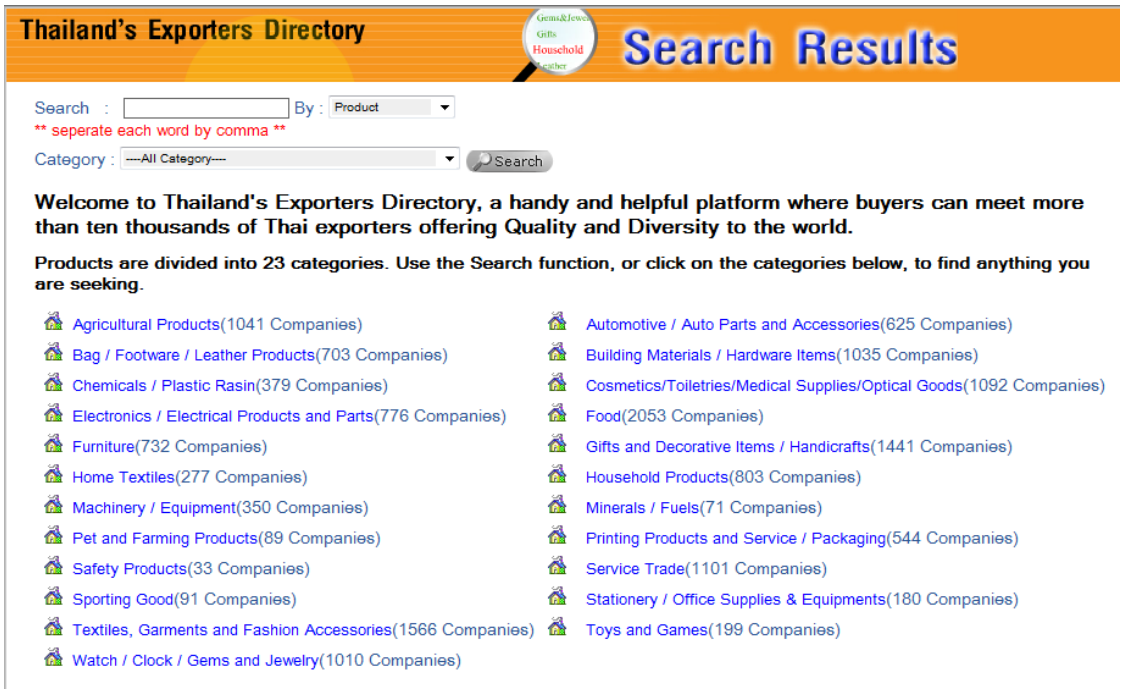
In particular, research suggested that the response rate can be maximized when monetary incentives and follow-up mailing are used in combination (James & Bolstein, 1990; Perneger et al, 1993). Anyway, a study by James and Bolstein (1990) found that although respondents who received incentive tended to put more effort in providing extra detail information to the survey questions, the comments

they provided tend to be more favorable than those who did not receive incentive. This can possibly make the answers to be susceptible to acquiescence bias. In fact, this issue is also consistent with a study by Trice (1984) which found that hotel guesses who received discount incentive tended to give lower negative ratings about the hotel than those who did not receive it. Therefore, the authors warned that researchers may have to take into consideration about the bias that may happen from using incentive, especially when the objective of the survey is to obtain critical responses.

Example of the sampling process in academic research

To gain a practical view of each step in sampling process, let's consider the study that the author conducted using survey data collected from companies in Thailand. In the research titled "Cultural intelligence and export performance of small and medium enterprises in Thailand: Mediating roles of organizational capabilities" (Charoensukmongkol, 2014b), the author aimed to study the benefits of cultural competencies of business owners called 'cultural intelligence' on export performance of their companies. If you recall the steps in sampling process that we just discussed, you will know that the first step is to define the target population. Because the aim of the research is to study export performance of entrepreneurial firms in Thailand, the *target population* of this research is small and medium enterprises (SMEs) that involved in export business.

However, the entire population of Thai SMEs at that time encompassed more than 2.7 million companies, which is too large for the author to cover all of them. In addition, it was difficult for the author to obtain contact information of every single export SMEs in Thailand. For these limitations, the author had to obtain a sampling frame of this entire population. Fortunately, the department of international trade promotion in Thailand offers the online open-access directory that compiles the list of exporters that have registered with the institution.



Thailand's Exporters Directory

Search Results

Search : By : Product

** seperate each word by comma **

Category : ----All Category---- Search

Welcome to Thailand's Exporters Directory, a handy and helpful platform where buyers can meet more than ten thousands of Thai exporters offering Quality and Diversity to the world.

Products are divided into 23 categories. Use the Search function, or click on the categories below, to find anything you are seeking.

- Agricultural Products(1041 Companies)
- Automotive / Auto Parts and Accessories(625 Companies)
- Bag / Footware / Leather Products(703 Companies)
- Building Materials / Hardware Items(1035 Companies)
- Chemicals / Plastic Rasin(379 Companies)
- Cosmetics/Toiletries/Medical Supplies/Optical Goods(1092 Companies)
- Electronics / Electrical Products and Parts(776 Companies)
- Food(2053 Companies)
- Furniture(732 Companies)
- Gifts and Decorative Items / Handicrafts(1441 Companies)
- Home Textiles(277 Companies)
- Household Products(803 Companies)
- Machinery / Equipment(350 Companies)
- Minerals / Fuels(71 Companies)
- Pet and Farming Products(89 Companies)
- Printing Products and Service / Packaging(544 Companies)
- Safety Products(33 Companies)
- Service Trade(1101 Companies)
- Sporting Good(91 Companies)
- Stationery / Office Supplies & Equipments(180 Companies)
- Textiles, Garments and Fashion Accessories(1566 Companies)
- Toys and Games(199 Companies)
- Watch / Clock / Gems and Jewelry(1010 Companies)

BAG/ FOOTWARE/ LEATHER PRODUCTS

Bag / Footware / Leather Products(703 Companies)

Search Company Name : search

Total 703 Record [71 Pages]

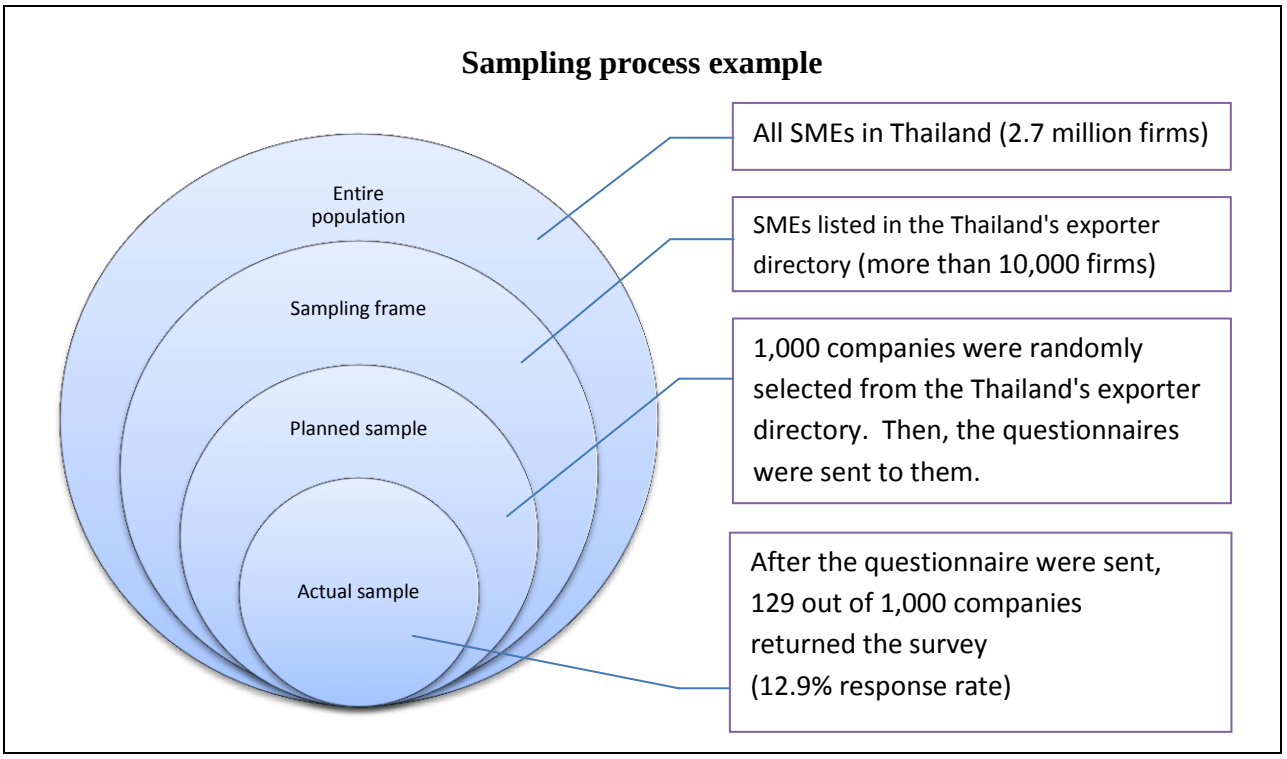
1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Next >> Go Page GO View all

No.	Company Name	Contact	Product	Award
1	1 AM CO., LTD.	MS.TANYAPAT LIMSUNTORN 1112/91-93 PRAKHANONG PLAZA, SUKHUMVIT RD., PRAKHANONG Khlong Toey,Bangkok 10110 rlst@headquarter.co.th Tel.66-2 3919386	1.TEXTILES, GARMENTS AND FASHION ACCESSORIES 2.BAG PRODUCTS 3.DRESSING 4.SHOES	
2	100 STONES	MR.SAJJA AURA-EK 12/19 SOI NAKNIVART 48, NAKINIVART RD., Lat Phrao,Bangkok 10230 100stones@gmail.com Tel.66-86 8230355	1.BRACELET 2.EARRINGS-SILVER 3.KEYCHAIN 4.NECKLACE SILVER	
3	A & T INTERNATIONAL LEATHER CO.,LTD	KHUNKONGPOP CHAIONGKRON 148 MOO 5, SOI TEDSAPAN BANGPU 50, TAIBAN MuangSamut Prakan,Samut Prakan 10280 president_u2@hotmail.com Tel.66-2701 2228	1.STINGRAY BAG 2.STINGRAY BELT 3.STINGRAY LEATHER 4.STINGRAY BRACELET 5.STINGRAY MONEY CLIP	
4	A A HOLDING CO.,LTD.	16/1 SOI PRAPINIT, SUANPLU, SOUTH SATHORN RD., ,Bangkok 10120 Tel.66-2 2855515	1.SHOES-MEN-LEATHER 2.SHOES-LADIES-LEATHER 3.SHOES-CHILDREN-LEATHER	

The directory not only lists Thai exporters classified by industries, but also provides contact information of every company in the list (including the names of the owner, mailing addresses, telephone numbers, and emails). Roughly, there are

more than 10,000 firms that are listed in this directory. In particular, the list of the companies in the directory is considered the representative of the entire population of Thai export SMEs that the author can gain access to. They are considered the *sampling frame* of the study.

However, the huge number of more than 10,000 firms in the directory is still too large for the author to survey all of them. Therefore, the author decided to randomly select 1,000 firms from the entire directory and used them as the *planned sample*. After these 1,000 firms were randomly selected, questionnaires along with cover letters were mailed to the owners of those firms based on the contact information provided in the directory. After one month, the total of 129 completely filled surveys was mailed back to the author. This amount accounts for 129 ÷ 1000 = 12.9 percent response rate. The total of 129 responses is considered the *actual sample* that the author used for data analysis.



PROBABILITY SAMPLING AND NON-PROBABILITY SAMPLING

Sampling methods can be broadly classified into two types: (1) probability sampling, and (2) nonprobability sampling. The major difference between these two methods is 'random selection'. In *probability sampling*, sample is selected randomly; every single unit within the population has equal chance to be selected as a sample. On the other hand, for *nonprobability sampling*, sample is not selected using random selection process. Each unit is chosen arbitrarily or based on subjective judgments of the researcher. Detail of each sampling technique will be provided in the next sections. Let's begin with nonprobability sampling.

Nonprobability sampling

As mentioned earlier, nonprobability sampling is a sampling technique that does not rely on random selection process. There are three types of non-probability sampling that has been widely used in research, which are: (1) convenient sampling, (2) judgment sampling, and (3) snowball sampling.

Convenient sampling

Convenient sampling (also known as *haphazard sampling* or *accidental sampling*) is a nonprobability sampling technique in which subjects are selected mainly because they are easily accessed by the researcher. For example, the researcher may distribute the questionnaire randomly to anybody he/she meets at public places (such as the department store, the university, or by the street). By using convenient sampling, the researcher does not concern much about the characteristics (e.g., gender, educational level, income, background or prior experience on some issues) of the people to be selected. Just anyone whom the researcher can approach is selected. Due to the convenience of this sampling method, it allows the researcher to obtain a large numbers of subjects in a short period of time.

However, when the researcher doesn't know in advance about the characteristics of a sample to be selected, it is difficult to know whether a person truly represents the

target population that the researcher requires or not. Thus, the major weakness of convenient sampling is that it is difficult to claim that a sample used in the study is the true representative of the population of interest, thereby making the generalizability of the results becomes questionable. This issue, therefore, can compromise the external validity of the findings. Due to this drawback, convenient sampling is normally used in a pilot study or is used when the objective of the research is not to generalize the results to a larger population.

Judgment sampling

Judgment sampling (or *purposive sampling*) is a nonprobability sampling technique that is quite similar to convenient sampling. The only difference is that, for judgment sampling, the researcher needs to use personal judgment or some predefined criteria to decide who are qualified to be selected as a sample. Instead of picking anyone that the researcher meets haphazardly, the researcher may only select persons who exhibit some characteristics that match with the topic that the researcher aim to study. For example, before a sample is selected, the researcher will have to make sure that a person has adequate knowledge or experience in the issue that the researcher wants to investigate.

In particular, judgment sampling can be used when the number of people who exhibit the characteristics of interest is quite limit. For example, if the researcher want to study about the impact of online game addiction on some personalities of young adults, the suitable sample for this research won't be anybody, but only those who are online gamers (Van Rooij et al, 2011). Sample in this case may be approached haphazardly at some public places such as game centers or internet cafés. Anyway, before deciding whether a particular young adult is qualified to be a sample, the researcher will use some criteria to justify whether he/she has some characteristics that can be a sign of online game addiction or not. For example, the researcher may prescreen a participant based on (a) the amount of time he/she regularly spends on online game, (b) the degree to which he/she feels that playing online game affects daily activities, and (c) the degree of difficulty that he/she experience when trying to quit playing online game. Overall, these can be used as the criteria to justify whether a person that the researcher meet is considered online

game addict (Tejeiro Salguero & Morán, 2002), and whether a person is suitable to be selected as a sample or not.

Anyway, like convenient sampling, the major weakness of judgment sampling is that the generalization of the results to a larger population can be questionable due to the lack of random selection in the sampling process.

Quota sampling

Quota sampling is a nonprobability sampling technique that is quite similar to judgment sampling. Like judgment sampling, a sample obtained from quota sampling is selected based on some predefined characteristics that match with the objective of the research. However, with quota sampling, a sample is classified into subcategories. The proportion of entities to be assigned to each subcategory is also determined by the researcher. For example, when collecting data about online game addiction from college students, the researcher may limit the number of students to be assigned to different classification. In this case, the researcher may require 30 percent of the total sample to be freshman, 30 percent to be sophomore, 20 percent to be Junior, and 20 percent to be senior.

Selecting college student sample using quota sampling

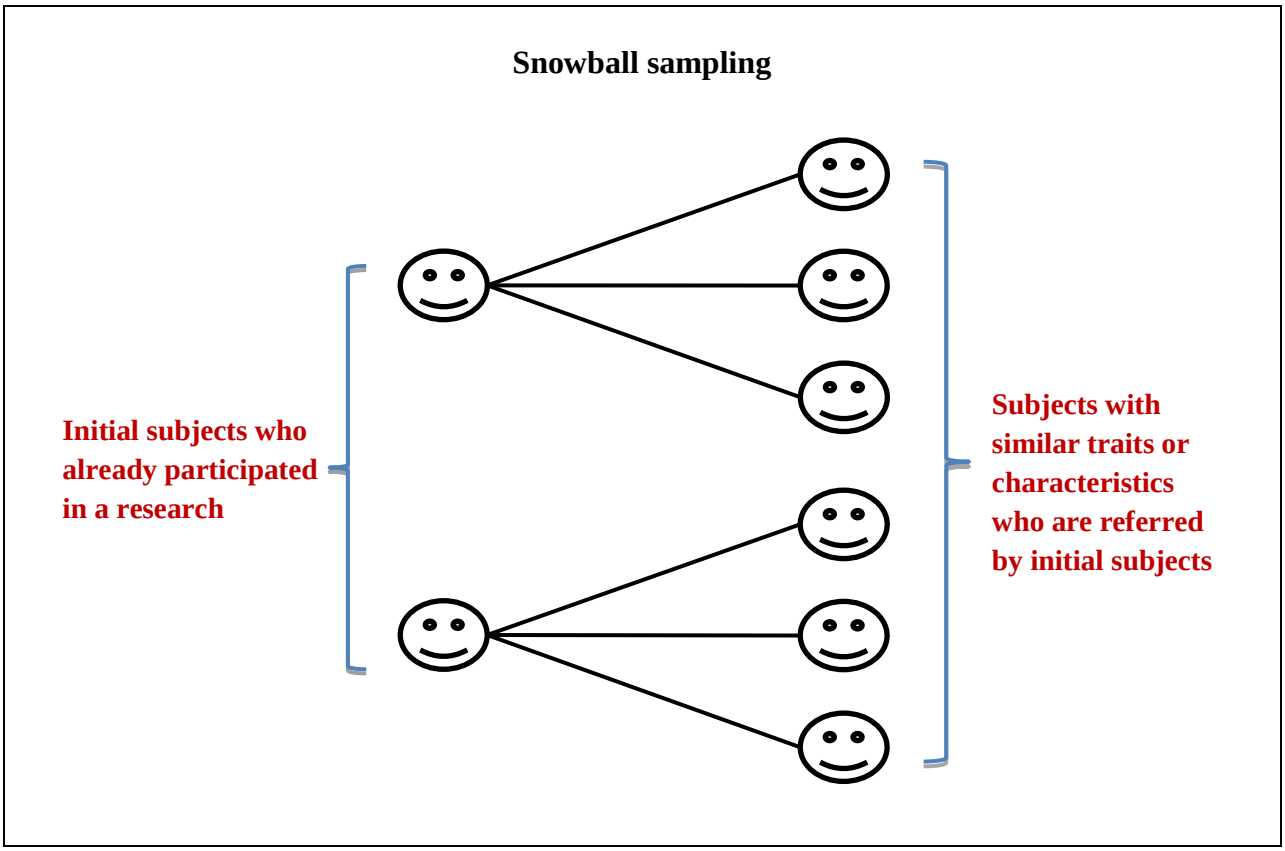
Freshman	Sophomore	Junior	Senior
30%	30%	20%	25%

Still, like convenient sampling and judgment sampling, the major weakness of quota sampling is that the generalization of the results to a larger population can be questionable due to the lack of random selection in the sampling process.

Snowball sampling

Snowball sampling (or *chain referral sampling*) is another type of nonprobability sampling that allows the researcher to gain access to a large number of people who are difficult to locate (Penrod et al, 2003). With snowball sampling, the researcher ask the initial subjects to help identify other people who have similar characteristics and to help bring them to participate in a study. For example, if the

researcher want to collect data from online gamers but has difficulty to gain access to a large number of online gamers, he/she can ask the initial subjects who are online gamers to help contacting other online gamers they know and to help bring those people to participate in the study. Of course, those who are online gamers will have a lot of friends who are online gamers as well. They tend to know other gamers who have the same interest more than the researcher does. Thus, using snowball sampling is useful in this circumstance. In particular, the main advantage of snowball sample is that it is simple and cost efficient. It allows the researcher to obtain a large amount of subjects in a short time period. Let's assume that we have 10 initial subjects who agree to help refer 20 friends to participate in the study, eventually we will end up obtaining (10 + (20x10)) = 210 subjects easily.



Although snowball sampling can help gaining access to a large number of subjects, the lack of random selection in the sampling process is a major drawback of snowball sampling. Because the process of sample selection is based on referral, it is difficult for the researcher to know in advance about the characteristics of the subjects who are referred by their friend. Plus, it is possible that the initial subjects

may only refer people whom they know well instead of people who have the right characteristics that the researcher needs. For these reasons, it is very likely that the subjects who are obtained from snowball sampling may not be a true representative of the population of interest. This limitation can make external validity of the research findings become questionable eventually.

Probability sampling

Unlike nonprobability sampling, *probability sampling* is a sampling technique that deploys randomization in the sample selection process. With randomization, every entity will have equal chance to be selected as a sample. In order to understand about the meaning of random sampling, imagine when you have 5 balls with different colors; let's say yellow, red, green, white, and black. If you put those 5 balls in a bag, close your eyes, put your hand in a bag, and then pick one ball from a bag, the chance that each color will be selected will be equal. Using a basic calculation, the chance that each color will be selected is equal to 1 ÷ 5 = 20 percent.

When using probability sampling, you just don't intentionally approach anyone you meet or use your judgment to decide who will be selected. Instead, you just let "chance" or "luck" decides which ones will be selected. Anyway, please note that random selection in probability sampling is not the same as when you randomly select subjects in convenient sampling. Although all of them involve the term 'randomness', fundamentally they are not the same. Just randomly invite people you meet at a public place to participate in the survey cannot be considered probability sampling because you don't have prior knowledge about their characteristics to justify whether they are a right representative of a population of interest or not. To be considered probability sampling, you must have a predefined sampling frame that is the appropriate representative of the population of interest first and then randomly select a sample from a sampling frame.

There are several techniques in probability sampling: (1) simple random sampling, (2) systematic random sampling, (3) stratified sampling, and (4) cluster sampling. Each of them will be discussed in detail as the following.

Simple random sampling

There are several methods that you can utilize to randomly pick a sample. In particular, the method that is widely used in research is called *simple random sampling*. Random selection can be performed in many ways. For example, if you want to select 30 students from 50 students in the class using simple random sampling method, you may prepare a solid box that contains 50 tiny balls inside, 30 of them are green and 20 are red, then you ask each student to blindly pick one ball from the box. Those who pick green balls will be selected as a sample. In this case, because students can't see which color of ball they will get from the box, you just let chance decides who will be selected.

However, the method we just mentioned is practical if you only select small sample from a small population. Anyway, it may not be practical if you have a large number of entities in a population from which you want to select a sample, let's say thousands. In this case, it does not make sense to let thousands of people blindly pick balls from a box. Nowadays, computer software, especially spreadsheet software like MS-Excel, can assist you to draw a sample from the entire list randomly.

To see how it works, let's consider the list of firms in the United Kingdom, as shown in the table below. Due to space limitation, the amount of firms that is shown in the example here is just 30 firms (in fact, the dataset contain more than a thousand firms). Anyway, you can try it later with larger data. The processes are simply the same.

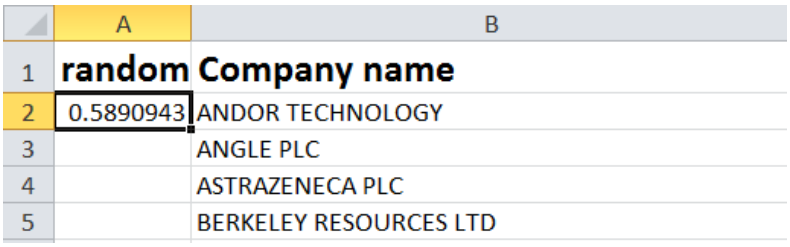
For example, if you have the list of 30 companies, and you want to randomly select only 20 companies from the total, you can use MS-Excel to help you perform the random selection as the following.

	A	B	C
1	**Company name**	**Industry**	**Region**
2	ANDOR TECHNOLOGY	Industrials	N.Ireland
3	ANGLE PLC	Industrials	Southeast
4	ASTRAZENECA PLC	Health Care	London
5	BERKELEY RESOURCES LTD	Basic Materials	Northwest
6	CREST NICHOLSON HLDGS PLC	Consumer Goods	London
7	DAILY INTERNET PLC	Technology	Midlands
8	DP POLAND PLC	Consumer Services	London
9	EW GROUP LTD	Financials	Channel Islands - Guernsey
10	GCM RESOURCES PLC	Basic Materials	London
11	GENEL ENERGY PLC	Financials	Channel Islands - Jersey
12	HIBU PLC	Consumer Services	Southeast
13	INTERNATIONAL GREETINGS	Consumer Goods	East Anglia
14	JPMORGAN ASIAN INV TRUST	Financials	London
15	KEYSTONE INVESTMENT TRUST	Financials	London
16	M.P.EVANS GROUP	Consumer Goods	Southeast
17	NEXT	Consumer Services	Midlands
18	NORISH	Industrials	East Anglia
19	NORTHAMBER	Technology	London
20	NORTHBRIDGE INDUSTRIAL SERVICES	Industrials	Midlands
21	PARITY GROUP	Technology	London
22	PORVAIR PLC	Oil & Gas	East Anglia
23	RED ROCK RESOURCES	Basic Materials	London
24	RENEWABLE ENERGY GENERATION LTD	Utilities	Channel Islands - Jersey
25	SABMILLER	Consumer Goods	London
26	SACOVEN PLC	Financials	Channel Islands - Jersey
27	SHARE PLC	Financials	East Anglia
28	STILO INTERNATIONAL	Technology	Southwest
29	UK MAIL GROUP PLC	Industrials	Southeast
30	WINCANTON	Industrials	Southwest

First, you have to make sure that the names of the company are sorted alphabetically in the beginning. Then, you have to create a new column and name it 'random' (you can name it anything you want to indicate that this column contains random numbers). In MS-Excel, there is a function called RAND() that you can use to generate a random number. You can enter that function at the first cell below the header column.

	A	B
1	**random**	**Company name**
2	=RAND()	ANDOR TECHNOLOGY
3		ANGLE PLC
4		ASTRAZENECA PLC
5		BERKELEY RESOURCES LTD

It will generate a random number for you. Note that the random number that you get will be different from what you see here.



	A	B
1	random	Company name
2	0.5890943	ANDOR TECHNOLOGY
3		ANGLE PLC
4		ASTRAZENECA PLC
5		BERKELEY RESOURCES LTD

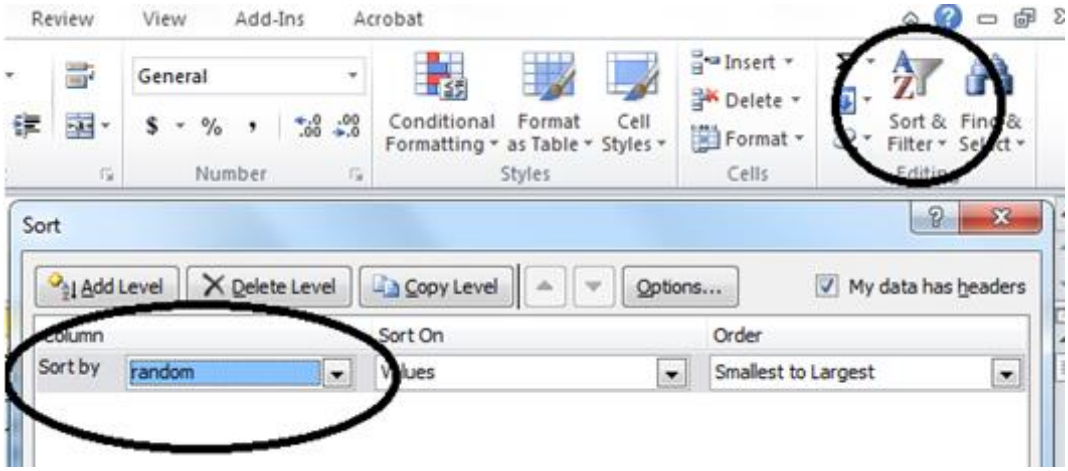
After that, you have to use a mouse to drag the value down until it reaches the last row of the dataset.

	A	B	
1	random	Company name	Ind
2	0.5890943	ANDOR TECHNOLOGY	Indu
3		ANGLE PLC	Indu
4		ASTRAZENECA PLC	Hea
5		BERKELEY RESOURCES LTD	Bas
6		CREST NICHOLSON HLDGS PLC	Con
7		DAILY INTERNET PLC	Tecl
8		DP POLAND PLC	Con
9		EW GROUP LTD	Fina

Then you will get the random number for each entity that you have in the dataset. Anyway, please note again that the random numbers that you obtain will not be the same like this. In fact, the random numbers will keep changing every time you generate and work with them.

	A	B	C	D
1	random	Company name	Industry	Region
2	0.4963512	ANDOR TECHNOLOGY	Industrials	N.Ireland
3	0.51372244	ANGLE PLC	Industrials	Southeast
4	0.41382521	ASTRAZENECA PLC	Health Care	London
5	0.37935191	BERKELEY RESOURCES LTD	Basic Materials	Northwest
6	0.67518692	CREST NICHOLSON HLDGS PLC	Consumer Goods	London
7	0.46969694	DAILY INTERNET PLC	Technology	Midlands
8	0.45357009	DP POLAND PLC	Consumer Services	London
9	0.39933593	EW GROUP LTD	Financials	Channel Islands - Guernsey
10	0.20603692	GCM RESOURCES PLC	Basic Materials	London
11	0.37229154	GENEL ENERGY PLC	Financials	Channel Islands - Jersey
12	0.27085697	HIBU PLC	Consumer Services	Southeast
13	0.89801932	INTERNATIONAL GREETINGS	Consumer Goods	East Anglia
14	0.03138438	JPMORGAN ASIAN INV TRUST	Financials	London
15	0.27648741	KEYSTONE INVESTMENT TRUST	Financials	London
16	0.29798256	M.P.EVANS GROUP	Consumer Goods	Southeast
17	0.0689226	NEXT	Consumer Services	Midlands
18	0.78929683	NORISH	Industrials	East Anglia
19	0.84376806	NORTHAMBER	Technology	London
20	0.79174236	NORTHBRIDGE INDUSTRIAL SERVICES	Industrials	Midlands
21	0.09540814	PARITY GROUP	Technology	London
22	0.37460415	PORVAIR PLC	Oil & Gas	East Anglia
23	0.16247742	RED ROCK RESOURCES	Basic Materials	London
24	0.81251663	RENEWABLE ENERGY GENERATION LTD	Utilities	Channel Islands - Jersey
25	0.70560584	SABMILLER	Consumer Goods	London
26	0.70483396	SACOVEN PLC	Financials	Channel Islands - Jersey
27	0.8129516	SHARE PLC	Financials	East Anglia
28	0.53078538	STILO INTERNATIONAL	Technology	Southwest
29	0.69774509	UK MAIL GROUP PLC	Industrials	Southeast
30	0.06698678	WINCANTON	Industrials	Southwest

After that, you sort data in the field that you created the random numbers.



You can see that now the order of the companies' name is randomly rearranged. They are not listed alphabetically anymore. Because you need only 20 companies, you can select the first 20 companies as your sample, and ignore the rests below them.

	A	B	C	D
1	random	Company name	Industry	Region
2	0.5387895	JPMORGAN ASIAN INV TRUST	Financials	London
3	0.722518	SABMILLER	Consumer Goods	London
4	0.7785859	NORISH	Industrials	East Anglia
5	0.6093906	ASTRAZENECA PLC	Health Care	London
6	0.6606541	DP POLAND PLC	Consumer Services	London
7	0.944875	KEYSTONE INVESTMENT TRUST	Financials	London
8	0.3334197	UK MAIL GROUP PLC	Industrials	Southeast
9	0.3691702	SHARE PLC	Financials	East Anglia
10	0.910281	STILO INTERNATIONAL	Technology	Southwest
11	0.127552	GENEL ENERGY PLC	Financials	Channel Islands - Jersey
12	0.6581474	ANGLE PLC	Industrials	Southeast
13	0.0467235	SACOVEN PLC	Financials	Channel Islands - Jersey
14	0.4151907	HIBU PLC	Consumer Services	Southeast
15	0.8031479	BERKELEY RESOURCES LTD	Basic Materials	Northwest
16	0.8547582	RENEWABLE ENERGY GENERATION LTD	Utilities	Channel Islands - Jersey
17	0.1998391	PARITY GROUP	Technology	London
18	0.7253092	INTERNATIONAL GREETINGS	Consumer Goods	East Anglia
19	0.2607839	M.P.EVANS GROUP	Consumer Goods	Southeast
20	0.4620077	CREST NICHOLSON HLDGS PLC	Consumer Goods	London
21	0.901374	EW GROUP LTD	Financials	Channel Islands - Guernsey
22	0.1194661	PORVAIR PLC	Oil & Gas	East Anglia
23	0.3040025	NEXT	Consumer Services	Midlands
24	0.1678955	WINCANTON	Industrials	Southwest
25	0.2707857	NORTHAMBER	Technology	London
26	0.9729185	NORTHBRIDGE INDUSTRIAL SERVICES	Industrials	Midlands
27	0.8687533	ANDOR TECHNOLOGY	Industrials	N.Ireland
28	0.4948481	GCM RESOURCES PLC	Basic Materials	London
29	0.730352	RED ROCK RESOURCES	Basic Materials	London
30	0.442027	DAILY INTERNET PLC	Technology	Midlands

In fact, this technique can be very handy if you have a huge amount of companies in a list. It can help you quickly and randomly select a certain amount of companies with ease.

Systematic random sampling

In addition to simple random sampling, you may use *systematic random sampling* by selecting every n^{th} unit of entities from the list. For example, when the order of the entities is already random, you may select entities that are in the 5^{th}, 10^{th}, 15^{th}, 20^{th}, 25^{th} of the dataset by using the running number incremented by five, as the following.

	Company name	Industry	Region
2	ANDOR TECHNOLOGY	Industrials	N.Ireland
3	ANGLE PLC	Industrials	Southeast
4	ASTRAZENECA PLC	Health Care	London
5	BERKELEY RESOURCES LTD	Basic Materials	Northwest
6	CREST NICHOLSON HLDGS PLC	Consumer Goods	London
7	DAILY INTERNET PLC	Technology	Midlands
8	DP POLAND PLC	Consumer Services	London
9	EW GROUP LTD	Financials	Channel Islands - Guernsey
10	GCM RESOURCES PLC	Basic Materials	London
11	GENEL ENERGY PLC	Financials	Channel Islands - Jersey
12	HIBU PLC	Consumer Services	Southeast
13	INTERNATIONAL GREETINGS	Consumer Goods	East Anglia
14	JPMORGAN ASIAN INV TRUST	Financials	London
15	KEYSTONE INVESTMENT TRUST	Financials	London
16	M.P.EVANS GROUP	Consumer Goods	Southeast
17	NEXT	Consumer Services	Midlands
18	NORISH	Industrials	East Anglia
19	NORTHAMBER	Technology	London
20	NORTHBRIDGE INDUSTRIAL SERVICES	Industrials	Midlands
21	PARITY GROUP	Technology	London
22	PORVAIR PLC	Oil & Gas	East Anglia
23	RED ROCK RESOURCES	Basic Materials	London
24	RENEWABLE ENERGY GENERATION LTD	Utilities	Channel Islands - Jersey
25	SABMILLER	Consumer Goods	London

Another example of systematic random sampling is when lucky people are randomly selected to win some prizes in a concert. In concert tickets, normally each of them has a unique running numbers. You can't decide which running number you will get when you buy a ticket; it is assigned to you randomly. When they want to randomly select lucky people to get some prizes using systematic random sampling, they may announce that anyone who gets the running number in the ticket that ends with 88 are lucky winners. In this case, those who get the running number 188, 288, 388, 488, 588, and so on, are randomly selected.

Stratified sampling

Stratified sampling is the probability sampling technique that is used when subpopulations within an overall population tend to vary. When stratified sampling is used, researchers first divide members of the population into homogeneous subgroups called *strata*. Some types of strata that are commonly used when classifying individuals are demographic characteristics such as gender, age group, race religion, etc. After strata are identified, the researchers randomly select sample to each stratum by using simple random sampling. For example, if you define strata by religion, we will have subpopulations represented by Buddhism, Islam, Roman Catholic, Christian, etc. After that, you randomly select people who belong to each religion to be a sample of that religion stratum. Normally, the sample size of each stratum should be proportionate to the population. When the sample size of each stratum is proportionate to the population, it is called *proportional stratified sample*. However, when the sample size of each stratum is not proportionate to the population, it is called *disproportional stratified sample*.

Operating system strata	Apple iOS	Android
Population of mobile phone users	43%	47%
Proportional stratified sample (1,000 subjects)	430	470
Disproportional stratified sample (1,000 subjects)	500	500

Another example, suppose that we want to collect a sample of mobile phone users. But anyway, mobile phones in the market vary in terms of the operating systems. In particular, there are two leading operating systems that dominate the markets, including Apple iOS and Android, which tend to differ in terms of functions and interface. Now let's assume that the recent statistics showed that 43 percent of the entire population in a country used iOS phones, while another 47 percent used Android phones. If we use stratified sampling, two subgroups or strata that are identified in this case include iOS and Android. If you want *proportional stratified sample*, the sample size that you will obtain for each stratum has to be proportionate to the population. For example, if you plan to obtain 1,000 sample for the study, the sample size of iOS users should be 430 (which is 43 percent of the population); the sample size of Android users should be 470 (which is 47 percent of the population). But if we use *disproportional stratified sample*, the

amount of each group of mobile phone users to be selected does not need to be proportionate to the population. For example, we may decide to select equal samples by randomly select 500 people who use iOS phones and 500 people who use Android phones.

In particular, the main benefit of using stratified sampling is that you can explore the differences in characteristics between subgroups within the population. Because subgroups are identified based on similarity, variability within the subgroups is apparently lower as compared to the variations of the entire population. However, the main disadvantage of using stratified sampling is that it may be difficult to define appropriate strata within a population. In this regard, it is crucial to make sure that there is no overlap in key characteristic of the subgroups that are chosen as strata.

Cluster sampling

Another probability sampling technique is called cluster sampling. *Cluster sampling* is the sampling technique in which population is divided into groups or clusters. After the clusters are identified, we randomly select some of them as representative clusters of the entire population (again, by using simple random sampling). Finally, all members that belong to each cluster are recruited as the sample.

For example, suppose that we want to collect data from police officers in Bangkok. However, obtaining the sample of all police officers by visiting every police station may not be an easy task. Luckily, we know that each district of Bangkok has only one police station. In this case, when using cluster sampling, we first divide the whole Bangkok area into clusters by using district. Apparently, there are totally 50 districts in the Bangkok area as the following.

Bang Bon	Bangkok Noi	Khlong Sam Wa	Phasi Charoen	Saphan Sung
Bang Kapi	Bangkok Yai	Khlong San	Phaya Thai	Sathon
Bang Khae	Bueng Kum	Khlong Toei	Phra Khanong	Suan Luang
Bang Khen	Chatuchak	Lak Si	Phra Nakhon	Taling Chan
Bang Kho Laem	Chom Thong	Lat Krabang	Pom Prap Sattru Phai	Thawi Watthana
Bang Khun Thian	Din Daeng	Lat Phrao	Prawet	Thon Buri
Bang Na	Don Mueang	Min Buri	Rat Burana	Thung Khru
Bang Phlat	Dusit	Nong Chok	Ratchathewi	Wang Thonglang
Bang Rak	Huai Khwang	Nong Khaem	Sai Mai	Watthana
Bang Sue	Khan Na Yao	Pathum Wan	Samphanthawong	Yan Nawa

When the total number of district within the whole area identified, we first randomly select a certain number of districts out of these 50 districts to be the representative clusters. For example, we may end up randomly obtain 10 districts as the cluster sample as the following:

Bang Kapi	Lat Krabang	Chatuchak	Pathum Wan	Bang Na
Prawet	Sathon	Phaya Thai	Bangkok Yai	Yan Nawa

After 10 randomly selected clusters are obtained, all police officer who are located in those districts are included as the sample.

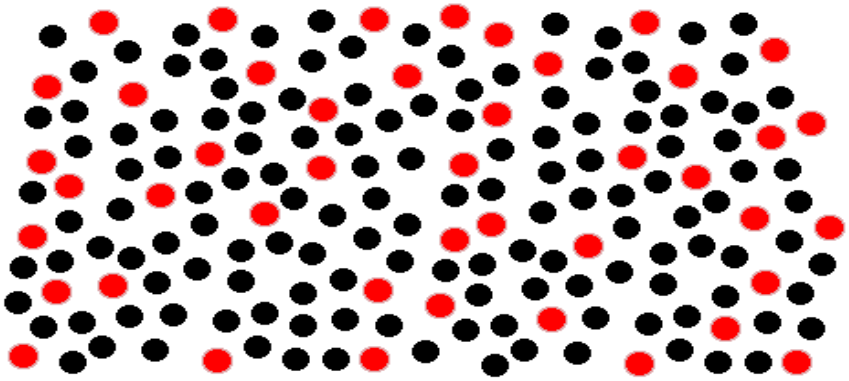
SOME BIAS IN SAMPLE SELECTION

Whether probability sampling or nonprobability sampling is used to obtain sample, bias can still present in the sample. Anyway, the chance to get bias in the sample, tend to be higher in probability sampling than in nonprobability sampling. Two types of bias that commonly present in sample section when random selection is not utilized are (1) random sampling error, and (2) self-selection bias.

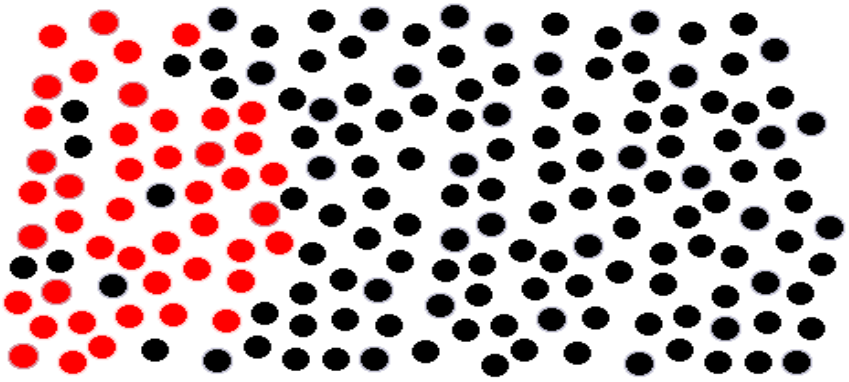
Random sampling error

Random sampling error or *random sampling bias* is the common problem that happens when the sample is not the representative of the population of interest. This problem normally occurs when probability sampling techniques are not used in sample selection. When sample is not randomly selected, it is more likely that

the sample that is obtained will come from a portion of the population rather than the whole population. When random sampling error presents in research, it can substantially lower the validity of research results due to the mismatch between the characteristics of the sample and the population of interest.



Random selection makes sure that every entity in the entire population has equal chance to be selected. It maximizes the chance that the sample obtained can be the representative of the entire population.



Without random selection, there is higher possibility that the sample may only come from some portion of the population that are not the true representative of the entire population.

The common solution to help minimize random sampling error is to use probability sampling to select a sample. In addition, increasing the sample size can also help lessen the concern of random sampling error. In particular, when sample size is large, it ensures that more people in the population are covered in the data.

Self-selection bias

Self-selection bias happens when people select themselves to participate in a study instead of being selected by the researcher. This circumstance can create bias because people who volunteer to be a sample may not be a true representative of the population of interest.

Basically, there are several reasons why some people select themselves to participate in a study. For example, it can happen when the researcher solicits participation by providing financial incentive. In this case, people who participate may not have underlying characteristics that match the purpose of a study, but they just participate simply because they can get compensated from doing so. This can possibly make the results obtained from a study to be misleading.

Another problem that might happen from self-selection is that some certain characteristics of a participant can be over-present. In particular, this occurs when the majority of people who join a study tend to be those who have strong interest about the research subject rather than those who feel indifferent about it. To make sense of this issue, let's consider this example. On the dining tables of many leading restaurant chains you can see a short survey that asks customers to rate the level of satisfaction they have toward several aspects of dining experience at the restaurant. Evidently, in normal that not so many customers are willing to do a survey, let alone paying attention to it (in fact, we may do it if the restaurant offers some discount for us). However, let's assume that one day the waiter expressed rude behaviors to you. When this happened, it is very likely that you would feel more urge to pick up the customer satisfaction survey on the table and worked on it. And of course, you would response negatively to every question in a survey. Obviously, when you have a strong feeling about the topic being asked in a survey, you self-select yourself to participate. But unfortunately, the information obtained from self-selected participants in this situation may not be their true opinions. Especially, when people have a strong attitude toward the topic being asked, the answers they provided can be very extreme at only one end. In this case, the only feedback that a restaurant received would only come from the opinions of customers who were dissatisfied with the services rather than the opinions of other regular customers.

REFERENCES

Alreck, P., & Settle, R. (1995). *The Survey Research Handbook*. New York, NY: McGraw-Hill.

Atrostic, B. K., Bates, N., Burt, G., & Silberstein, A. (2001). Nonresponse in U.S. government household surveys: consistent measures, recent trends, and new insights. *Journal of Official Statistics, 17*(2), 209–226.

Babbie, E. (2007). *The Practice of Social Research* (11th ed.). Belmont, CA: Wadsworth.

Babbie, E. R. (1990). *Survey Research Methods*. Belmont, CA: Wadsworth.

Charoensukmongkol, P. (2014a). Benefits of Mindfulness Meditation on Emotional Intelligence, General Self-Efficacy, and Perceived Stress: Evidence from Thailand. *Journal of Spirituality in Mental Health, 16*(3), 171–192.

Charoensukmongkol, P. (2014b). Cultural intelligence and export performance of small and medium enterprises in Thailand: Mediating roles of organizational capabilities. *International Small Business Journal*. doi: 10.1177/0266242614539364

Charoensukmongkol, P. (2014c). The effect of software piracy on research and development intensity at the country level: do developed countries and emerging economies suffer the same impact? . *Global Business and Economics Review, 16*(3), 253-268. doi: 10.1504/GBER.2014.063064

Childers, T. L., & Ferrell, O. C. (1979). Response Rates and Perceived Questionnaire Length in Mail Surveys. *Journal of Marketing Research, 16*(3), 429-431.

Dommeyer, C. J. (1985). Does response to an offer of mail survey results with questionnaire interest? *Journal Market Research Society, 27*(1), 27–38.

Ellis, B. (2011). The banks' billion-dollar idea. Retrieved November 5, 2014, from http://money.cnn.com/2011/07/06/pf/banks_sell_shopping_data

Fox, R. J., Crask, M. R., & Kim, J. (1988). Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response. *Public Opinion Quarterly, 52*(4), 467-491. doi: 10.1086/269125

Groves, R. M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly, 70*(5), 646-675. doi: 10.1093/poq/nfl033

Hawkins, D. F. (1975). Estimation of Nonresponse Bias. *Sociological Methods & Research, 3*(4), 461-488. doi: 10.1177/004912417500300406

James, J. M., & Bolstein, R. (1990). The effect of monetary incentives and follow-up mailings on the response rate and response quality in mail surveys. *Public Opinion Quarterly, 54*(3), 346-361. doi: 10.1086/269211

Jepson, C., Asch, D. A., Hershey, J. C., & Ubel, P. A. (2005). In a mailed physician survey, questionnaire length had a threshold effect on response rate. *Journal of Clinical Epidemiology, 58*(1), 103-105. doi: http://dx.doi.org/10.1016/j.jclinepi.2004.06.004

Kalantar, J. S., & Talley, N. J. (1999). The Effects of Lottery Incentive and Length of Questionnaire on Health Survey Response Rates: A Randomized Study. *Journal of Clinical Epidemiology, 52*(11), 1117-1122. doi: http://dx.doi.org/10.1016/S0895-4356(99)00051-7

Lund, E., & Gram, I. T. (1998). Response rate according to title and length of questionnaire. *Scandinavian Journal of Public Health, 26*(2), 154-160. doi: 10.1177/14034948980260020401

Penrod, J., Preston, D. B., Cain, R. E., & Starks, M. T. (2003). A Discussion of Chain Referral As a Method of Sampling Hard-to-Reach Populations. *Journal of Transcultural Nursing, 14*(2), 100-107. doi: 10.1177/1043659602250614

Perneger, T. V., Etter, J.-F., & Rougemont, A. (1993). Randomized Trial of Use of a Monetary Incentive and a Reminder Card to Increase the Response Rate to a Mailed Health Survey. *American Journal of Epidemiology, 138*(9), 714-722.

Rea, L., & Parker, R. (1997). *Designing and conducting survey research* (2nd ed.). San Francisco, CA: Jossey-Bass.

Robert, S., David, O., Bruce, H., & Dan, C. (2003). The effect of questionnaire length on participant response rate: A case study in the U.S. cabinet industry. *Forest Products Journal, 53*, 33-36

Singleton, R., & Straits, B. (2005). *Approaches to Social Research* (4th ed.). New York, NY: Oxford University Press.

Squire, P. (1988). Why the 1936 Literary Digest poll failed. *Public Opinion Quarterly, 52*(1), 125-133.

Tejeiro Salguero, R. A., & Morán, R. M. B. (2002). Measuring problem video game playing in adolescents. *Addiction, 97*(12), 1601-1606. doi: 10.1046/j.1360-0443.2002.00218.x

Trice, A. D. (1984). Hotel ratings: IV. Recruitment and incentives. *Psychological reports, 55*(1), 58.

Van Rooij, A. J., Schoenmakers, T. M., Vermulst, A. A., Van Den Eijnden, R. J. J. M., & Van De Mheen, D. (2011). Online video game addiction: identification of addicted adolescent gamers. *Addiction, 106*(1), 205-212. doi: 10.1111/j.1360-0443.2010.03104.x

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior: A meta-analysis. *Public Opinion Quarterly, 55*(4), 613–639.