Chapter 5
Survey Design

Case Study: Sampling for Potential Growth Areas

Infinity Auto Insurance is looking to expand their customer base starting with large metropolitan areas across the United States. Because the Latino/Hispanic population is the largest growing population in the United States, Infinity Auto Insurance wishes to better understand this market segment and expand its advertising campaigns to include this population. They have decided to concentrate their strategic efforts to large metropolitan areas in four different states: California, Arizona, Texas, and Florida. For the initial study, they sampled 200 people in Los Angeles, California.

Food for Thought

What other states/cities should they target?
What other demographic segments should they consider?
From what you understand of sampling, what sample size should they use to gather a statistically valid sample?

Possible Answers

What other states/cities should they target?
Houston, Texas; Phoenix Arizona; Miami, Florida; Austin, Texas
What other demographic segments should they consider?
Lower income or less educated individuals

L. Herkenhoff and J. Fogli, *Applied Statistics for Business and Management*   117

Over the years, Infinity Auto Insurance has increased their sample size from 200 to 400, or 600 depending on the depth and detail they wish to obtain from the sample. With a sample size of 400, Infinity Auto Insurance is able to statistically calculate valid information at a 95 % level of confidence (± 5 points); with 600, 95 % level of confidence (± 4 points).

Food for Thought

Would you suggest they gather a larger sample or is their current sample sufficient?

If you were working for Infinity Auto Insurance, how would you go about gathering a valid sample?

What issues would help or hinder you from creating a statistically valid sample?

Possible Answers

Would you suggest they gather a larger sample or is their current sample sufficient?

A sample size of 400 or 600 should be sufficient to capture the necessary data.

If you were working for Infinity Auto Insurance, what types how would you go about gathering a valid sample?

Use a random sample by gathering a database of randomly selected individuals. Contact the individuals in the database randomly to take the survey.

What issues would help or hinder you from creating a statistically valid sample?

Non response, non-random sampling, etc.

Key Concepts

Bias, Bipolar scale, Extreme checking, Likert scale, Nonprobability sampling, Primary data, Probability sampling, Random sampling, Random sampling error, Response rates, Reversed items, Secondary data and Systematic error.

Discussion

This chapter provides a framework for effective survey design. All statistical analyses require high quality input data. The first step in collecting unbiased, reliable and valid survey data is having an unbiased, reliable and valid survey instrument. Several effective examples are provided as well as some of the more common pitfalls to avoid.

Basic Concepts

Primary data is collected by the investigator conducting the research. **Secondary data** is collected by someone other than the investigator. The most common way to generate primary data is through surveys. Surveys provide quick, relatively inexpensive, and efficient means of assessing information about a population. The tasks of determining the survey distribution method, determining the list of questions, and designing the correct format of the survey are essential aspects of the development of a reliable and robust survey.

Most surveys are designed to collect information from a sample of the population. Sampling is the use of a subset of the population to represent the whole population. **Probability sampling**, or **random sampling**, is a sampling technique in which the probability of getting any particular respondent is equally as likely to occur. **Nonprobability sampling** does not meet this criterion and should be used with caution. Regardless of how the data is collected, it is important to minimize errors in the data collection process. This topic is covered in more detail in Chapter 6.

◈ When a selection procedure is biased, taking a larger sample does not help. This just creates the bias on a larger scale.
◈ To minimize bias, an impartial and objective probability method should be used to select the sample.

Although most surveys are conducted to quantify factual information, certain aspects of surveys may be qualitative. Although most surveys are descriptive, they can also be designed to provide causal explanations.

Survey Design

The survey designing process is a very important step to get your future survey produced. The first step is to start planning the purpose of the survey. Start by listing the survey objectives. As you start drafting questions, make sure that your survey items correspond to an initial objective listed. A common mistake is that

Copyright © 2013. Springer. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.
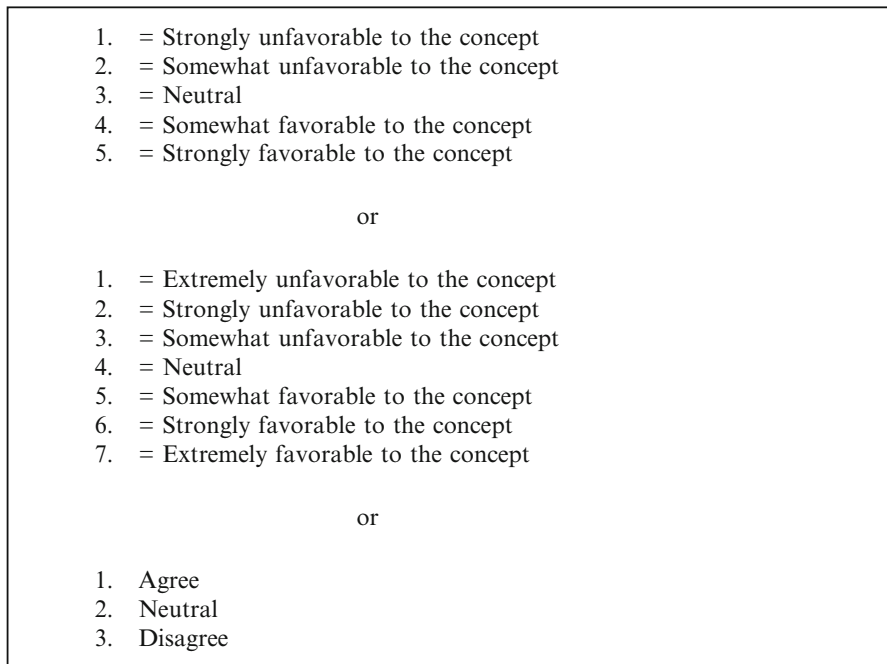
researchers do not get the information needed because they do not ask the right questions in the survey. Finally, keep in mind that demographic questions can be critical to analyze results.

Scale

The **Likert scale** is one of the most common scales you will come across in surveys. This scale was introduced by Likert (England; 1932). Likert scaling is a uni-dimensional scaling method or "summative" scale.

One of the first questions in designing a survey is how many choices should be offered on the Likert scale. The most common are 5 or 7-point Disagree-Agree response scales. Examples are given in Fig. 5.1.

There are tradeoffs between choosing a 5-point scale and a 7-point scale. The 7-point scale provides more choice for the respondent and a more detailed response set for the analyst. However, more choices may lengthen the time to complete the survey, as respondents try to decide on their answer. Too detailed responses may require more time to complete: "Was I really somewhat satisfied or just satisfied the last time I responded to this survey?"

```
1.   = Strongly unfavorable to the concept
2.   = Somewhat unfavorable to the concept
3.   = Neutral
4.   = Somewhat favorable to the concept
5.   = Strongly favorable to the concept


                        or


1.   = Extremely unfavorable to the concept
2.   = Strongly unfavorable to the concept
3.   = Somewhat unfavorable to the concept
4.   = Neutral
5.   = Somewhat favorable to the concept
6.   = Strongly favorable to the concept
7.   = Extremely favorable to the concept


                        or


1.   Agree
2.   Neutral
3.   Disagree
```
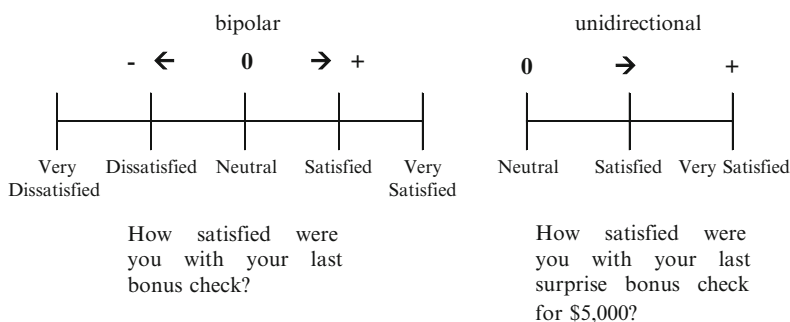
Fig. 5.1 Likert scales

However, having too few choices, less than 5, can be frustrating to the respondents; none of the answers really fit. This smaller scale also limits a more detailed understanding of the responses. A 3-point scale is also given in Fig. 5.1.

The final score for the respondent on the scale is the sum of their ratings for all of the items (this is why it is sometimes called a "summated" scale). On some scales, you will have items that are reversed in meaning from the overall direction of the scale. These are called **reversed items**. You will need to reverse the response value for each of these items before summing for the total.

Bipolar questions offer positive, neutral and negative choices. Whenever you are dealing with bipolar answers, it is important to have an odd number of choices no matter how large the scale. This becomes even more important when using a Likert scale with qualitative data. In this situation, the respondent is forced to decide whether they lean more towards the agree or disagree end of the scale for each item. Figure 5.2 illustrates bipolar and unidirectional Likert scales.

Fig. 5.2 Polarity with Likert scales

In Fig. 5.2 the first bipolar scale is designed to allow respondents to agree, disagree or remain neutral about their level of satisfaction with the recent bonus check. The second unidirectional scale is designed to only capture positive responses. In other words, it is safe to say everyone would be positive to some degree to get a surprise bonus check for $5,000. These scales are easily converted to quantitative scales if they are designed correctly, which allows for more rigorous analysis.

If the bipolar scale had only been a 4-point scale we would not have been able to collect neutral responses. However, often less rigorous surveyors will simply average the responses to synthetically create a neutral response, as shown in Fig. 5.3. This is inaccurate data and should be avoided.
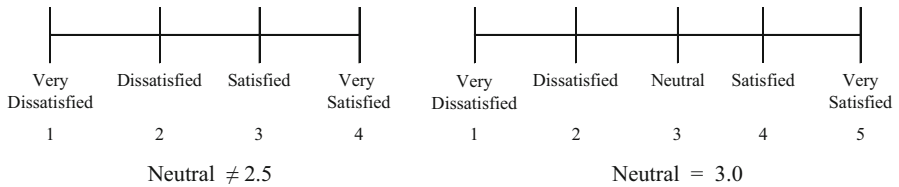
Fig. 5.3 Offering a neutral choice

Types of Questions

Single Response/Select

Single select items are questions that ask the respondent to select one specific response to fulfill the question requirement. For example most demographic questions are single response questions. The question asks for a specific characteristic of the person's demographic status (i.e. age, gender, household income, etc.), and the respondent is allowed to respond with one specific selection.

Example

1. What is your age?

   ○ 18–24
   ○ 25–34
   ○ 35–44
   ○ 45–54
   ○ 55–64
   ○ 65 or over

Because the respondent is a specific age, he/she will respond with only one item.

◈ For range items, such as age, make sure the numbers do not overlap and that the entire range is covered. Otherwise, this might cause confusion and inaccurate responses (i.e. age 18–25, age 25–30, age 30–40, etc.)

Multiple Response/Select

Multiple select questions are ones that ask the respondent to "select all that apply." A respondent can choose one or more responses. Sometimes the option of "*None of the above*" is available when none of the other responses seem to fit.

Example

1. What are your favorite colors?

    ☐ Red
    ☐ Blue
    ☐ Green
    ☐ Yellow
    ☐ Orange
    ☐ None of the above

When a respondent chooses the different responses, he/she is allowed to choose as many responses as they deem fit. However, if the respondent chooses the "*None of the above*" option, it should deselect any previous answer selections for this question.

Structured Questions

Structured questions are questions that have a closed set of responses from which to choose. Structured questions are usually preferred because they make data collection and analysis much easier and they are quicker for the respondent to answer. Structured questions are recommended when you are not trying to capture new ideas or thoughts from the respondent. You need a good understanding of the responses so that you can develop the appropriate answer choices.

Example

Do you have an MBA?
○ Yes
○ No
○ In Progress

Which international city where we have offices would you like to accept an assignment?
○ New York
○ Singapore
○ Dubai
○ Hong Kong
○ Frankfurt
○ Stonetown

For structured questions make sure that the list of answers includes every possible answer and that each of the answers is unique.

To ensure that the data you are collecting is indeed complete, you might want to include *"Other"* as the last answer choice. Including the answer choice *"Don't know"* to a response list provides a valid choice for respondents who are not capable of answering the question.

Do not include answer choices that are irrelevant to the question. You want to keep the survey as short and as easy to complete as possible.

Example of an Irrelevant Answer Choice

To which city should we relocate the head office?

○ San Francisco
○ Atlanta
○ Jacksonville
○ Detroit
○ Disneyland

The last answer choice is not a real city and may be a very popular answer but is certainly not going to add much to your analysis. Although we might think it good to include humorous answers every once and awhile, keep in mind that when someone actually chooses these, you lose some real data.

All of the responses should be consistent so that no single response stands out from the other responses. Consistency ensures that you are not leading respondents to a particular answer by making that answer different from the others.

Example of an Inconsistent Answer Choice

To which city should we relocate the head office?

○ San Francisco
○ Atlanta
○ Jacksonville
○ Detroit
○ Alaska

The last answer choice is not a city but a state. If we really did not want the respondent to choose a city in Alaska this would be the way to ensure that. This confuses the respondent and distracts them from completing the survey in a valid manner.

Ranking and Rating

Surveys often gather opinion information, by way of asking a person's opinion on a topic. Rating or ranking questions capture varying degrees of emotion about a subject. A rating question asks respondents to indicate the degree to which they feel about a certain topic. These are often referred to as Likert scales.

For Example

How do you feel about the new flexible spending account?

Unsatisfied	Somewhat Satisfied	Satisfied	Very Satisfied	Extremely Satisfied
(1)	(2)	(3)	(4)	(5)

A **ranking** question asks respondents to explain how they feel about something by comparing it to other items in a list.

Example

Please rank the following healthcare benefits (1 being highest rank and 4 being the lowest)
___Flexible healthcare savings account
___Employee assistance program
___Long term insurance for retiree healthcare
___Exercise program subsidy

Non-structured (Open-Ended) Questions

Non-structured questions, or open-ended questions, are questions that have no list of answer choices from which to choose. Respondents write in their response to a question.

Example

What do you like best about our new CEO?
_____
_____

Use non-structured questions when you are exploring new ideas and you don't really know what to expect from the respondents. This format gives them an opportunity to provide you with information that you may not have considered before. However, keep in mind that these types of questions take much longer to analyze.

Data

Labels

Another issue involves thinking you have transformed qualitative data into quantitative data because the labels have become numeric. By translating the label "satisfied" to a "3" does not make that data quantitative. Recall from our earlier chapters on descriptive statistics, we are much more limited in what type of analysis we can complete with qualitative data, regardless of whether the labels are numeric or not.

Demographic Data

To make sure the survey sample is representative of a wider population we need to collect demographic information. Demographic items are typically categorical data. Some common demographic variables are presented in Fig. 5.4.

Common Demographic Variables	
Gender	Level of education
Age	Job title
Marital status	Size of organization
Years of experience	Industry
Management level	

Fig. 5.4   Common demographic variables

In addition to those variables listed in Fig. 5.4 and depending on the purpose of the survey, you may also want to collect demographic data on national culture and religion; therefore, make sure you have included all the necessary demographic variables.

Response Rates

It is important to achieve a high response rate; otherwise, we may have non-response bias. In other words, there might be some reason we would suspect

that so many people did not respond. For example, if the survey was 15 pages long, people might not complete the survey because of how much time would be required. Those who do reply may also represent a bias. Perhaps, this group has more time on their hands, they are personal friends with the researcher, or they worked for the manager who initiated the survey. This type of bias weakens the usefulness of the data. Figure 5.5 summarizes some key design elements that may improve response rates.

---

1. Keep the survey as brief as possible.
2. Ensure all questions are clear and easy to understand. Avoid ambiguity ;it creates respondent frustration.
3. Maintain a consistent scale on short surveys; mix it up on longer surveys to keep it interesting for respondents.
4. Provide confidentiality and anonymity.
5. Don't ask the same question in different ways more than three times.
6. Explain what the survey results will help provide.

Fig. 5.5 Response rate factors

Of course, the most influential factor is the subject matter of the survey. If you ask employees to complete a survey on their pay and benefits, there is usually a high response rate, because it is of great interest and valence to employees. However, if you ask employees to complete a survey about a new ad campaign, there may be a very low response rate, because they have been asked to do this every month for the past 6 months.
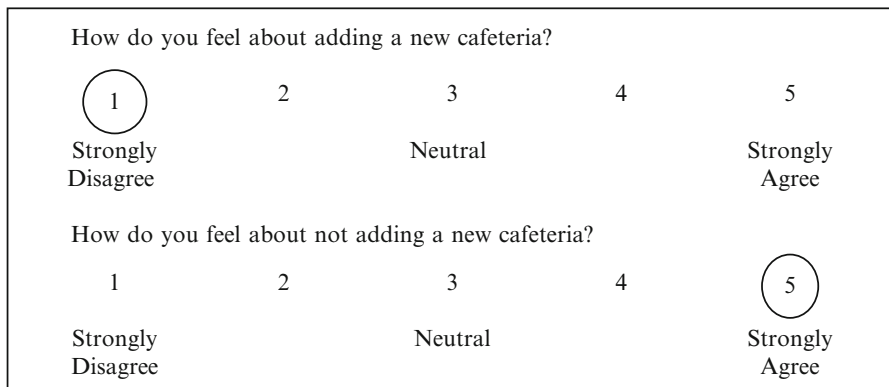
Editing: Data Quality

Before you start generating statistics from your data, you need to have a quick look at the raw data. Generate an Excel output of the raw data to check for unusual patterns in the data. Figure 5.6 gives you some examples.

Respondent		Questions						
		1	**2**	**3**	**4**	**5**	**6**	**7**
1		5	1	5		1	1	5
2		3	2	4		3	4	6
3		3	3	3		5	2	3
4		1	1	1		1	1	1
5		3	1	2		5	5	5
6		3	2	4		3	1	5
7		3	1	5		2	5	2
8		3	3	1		5	2	3

Fig. 5.6 Patterns in response

In Fig. 5.6, there are 7 questions and 8 respondents. There are several patterns in these responses that may concern us.

- Respondent 1 has only checked 1 and 5 on a 5-point Likert scale. This type of response pattern is referred to as **extreme checking** and creates a bias in the data. Some respondents only really agree with something or really disagree; they just cannot take a neutral position.
- Question 1 seems to have an overwhelming "3" or neutral response on the 5-point Likert scale. This may mean the question was worded in a confusing way or there really was no way to disagree or agree with the question. Consider the question "*Over half of our employees were "satisfied" in an earlier version of the survey and this is important. Do you agree or disagree?*" You may not have heard about an earlier version of the survey, let alone the survey results, so how can you agree or disagree. You don't know if the first part of the question is true or not. A remedy in this case would be to insert the word "*If*" at the start of the question.
- Question 4 has a large number of non-responses. Again a source of potential bias.
- Respondent 2 shows a "6" as the response to question 7. This needs to be checked, since the possible answers change only from 1 to 5.
- Respondent 4 seems to have only selected "1" as a response to all questions. This respondent may not have bothered to read any of the questions. If you asked some of the questions with **reversed scales** you can check if the answers are consistent. Consider the example.

How do you feel about adding a new cafeteria?

   (1)   2   3   4   5

Strongly   Neutral   Strongly
Disagree   Agree

How do you feel about not adding a new cafeteria?

    1   2   3   4   (5)

Strongly   Neutral   Strongly
Disagree   Agree

In this example, someone who strongly disagrees in the first question should strongly agree with not adding the cafeteria in the reversed question.

Coding

Coding is the process of identifying and classifying each answer with a numerical
score or other character symbol. Before data can be tabulated, meaningful
categories and character symbols must be established for groups of responses. In
Fig. 5.7 the answers are coded as **1 for Yes**, **2 for No** and **3 for Not sure**.

```
29. Do you belong to a union?
      1 □          Yes
      2 □          No
      3 □          Not sure
```

Fig. 5.7 Basic coding of answers

Open ended questions pose more work in terms of coding as shown in Fig. 5.8.
You need to complete a qualitative word categorization. There may be many
comments concerning "*inequality*" or "*difficult to understand.*" These categories
can then be coded accordingly.

```
30. Do you have any further comments on the new tax incentive?
    _____
```

Fig. 5.8 Coding open-ended questions

Be careful with **reversed scale** questions. As shown in Fig. 5.9, the answers for
question 2 need to be reversed so the high agreement score still gets coded as a
1, rather than the 5.

1. How do you feel about the new vacation policy?				
1	2	3	4	5
Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
2. How do you feel about the new vacation guidelines?				
1	2	3	4	5
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Fig. 5.9 Reversed scale questions without transformation

Reversing the order of the codes for negative statements requires a simple data transformation, so the codes reflect the same direction and order of magnitude as the positive statements. Figure 5.10 shows the same questions with question 2 recoded appropriately

1. How do you feel about the new vacation policy?

1	2	3	4	5
Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

2. How do you feel about the new vacation guidelines?

5	4	3	2	1
Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

Fig. 5.10 Reversed scale questions with transformation

The 0/1 coding is sometimes referred to as dummy variables or indicator variables. One of the most common dummy variables is gender, where we assign a 1 for female and a 0 for male. In some situations, we may have more than 2 categories such as the quarters in the fiscal year or the seasons. In these cases, we use a number of indicator variables always starting with the number "0" as the first category.

Errors in Survey Question Creation

Loaded Questions

This type of question usually produces a socially desirable response or contains emotionally charged words. For example, *"Are you a cheater*?" These types of questions make assumptions about the respondent and tend to evoke certain responses from them. To avoid asking leading questions, use agree/disagree questions that are phrased with words that have neutral connotations and do not assume specific details about the respondent.

Leading Questions

This type of item phrases the question in a way that suggests to the respondent that he/she is supposed to answer in a specific manner (i.e., it leads the respondent to a particular answer). For example, *"Do you agree with Doctors world-wide that say*

obesity has become a chronic pandemic?" To avoid leading questions, remove the leading words or phrases (e.g., the words that make a judgment or imply stupidity if the respondent disagrees). In this example, "*Doctors world-wide*" is leading the respondent to agree with "experts."

To avoid leading questions, phrase all questions in a neutral way. It is a common error for surveyors to word the questions in a manner that reflects their underlying opinion.

Example

Bad question: leading	Good question: neutral
Do you think that the new CEO who was hired from a completely different industry can effectively lead our company? ○ Yes ○ No ○ No opinion	Do you think our new CEO can effectively lead our company? ○ Yes ○ No ○ No opinion

---

The leading question hints that since the new CEO is from another industry this may influence his/her ability to lead. The neutral question presents a better way to phrase this question, and removes the bias.

Double-Barreled Questions

This type of question asks about more than one idea or issue in one question, which may lead to inaccuracies in the attitudes and opinions being measured. For example, "do you agree or disagree that ice cream and cake are great?" The respondent may like ice cream but not cake, vice versa, both, or none. To avoid double-barreled questions ask each idea separately.

◈ Make sure you are only asking one question at a time, this is different from multiple select type questions. These types of questions do not provide you with valid and reliable data.

Example

Bad question: double-barreled question	Good question
How have scientists and administrative staff at your research institute responded to the new parking policy? ○ Satisfied ○ Unsatisfied	How have <u>scientists</u> at your research institute responded to the new parking policy? ○ Satisfied ○ Unsatisfied How have <u>administrative staff</u> at your research institute responded to the new parking policy? ○ Satisfied ○ Unsatisfied

The double barreled question is phrased in such an ambiguous way, that you will not know what the respondent intended with their response (e.g. are the scientists, the administrative staff, or both satisfied/dissatisfied?). The key word (which differentiates the two questions) should be underlined to help ensure that the respondents are reading the questions correctly when the structures are so similar.

Errors in Survey Data Collection

Random Sampling Error

Even with probability sampling methods, statistical errors will occur because of chance error. **Random sampling errors** can be estimated, and sample size can be increased to decrease the effect of this type of error in the data.

Systematic Error

This type of error results from some imperfect aspect of the survey or from a mistake in executing the survey. The two most common forms of **systematic error** are referred to as respondent error and nonresponse error. Nonresponse error occurs when certain groups are no longer represented in your survey because they did not respond. For example, if you were conducting a door-to-door poll in a neighborhood where typically both people work during the day, your sample will only include respondents such as retirees or people on sick leave or vacation.

If your nonresponse is high, you need to rethink your survey questions, format, and/or delivery method. Self-selection bias is another form of error in which allows extreme position to be overrepresented while other differing positions are underrepresented.

Response Bias

There are six specific categories of response bias: acquiescence, auspices, central tendency, extremity, interviewer, and social desirability. Take note that these biases are not mutually exclusive, so you can have several of these biases operating within the same survey. Surveyors should also be alert for culturally based response biases in international business research. One might expect more acquiescence bias within Japanese responses, as they have a more culturally homogeneous society and tend not to contradict their superiors.

- **Acquiescence Bias.** This is a response bias due to the respondents' tendency to agree with questions when having doubt. Some respondents are very agreeable whereas others may be very disagreeable throughout the survey. We typically find positive acquiescence when asking about a new product.
- **Auspices Bias.** Answers may be influenced by the actual group doing the survey. If EarthWatch was conducting a survey about open pit mining the respondent may be inclined to provide answers that align with the mission of the organization.
- **Central Tendency Bias.** This response bias usually occurs with respondents who wish to remain neutral or have no specific opinion. Often, participants who respond this way want to get through the survey quickly and don't care about the subject matter of the survey.
- **Extremity Bias.** Some respondents will indicate extreme answers either in a positive and/or negative manner. These are the folks who only check the extreme ends of the scale, which may cause a bias in the data.
- **Interviewer Bias.** There may be a bias introduced because of the interactions between the interviewer and the interviewee. Often socially acceptable answers are given to impress the interviewer. The physical appearance of the interviewer may also affect the respondents. If the interviewer smiles after a particular answer is given, the respondent may tend to provide similar answers for further positive feedback. The gender of the interviewer may hamper candid responses on gender specific issues. If the interviewer takes too long to ask the question the respondent may feel like their time is being wasted and will start answering abruptly.
- **Social Desirability Bias.** The respondent may either consciously or unconsciously provide an answer to create a favorable impression. For example in asking if someone voted in the last election, you might create a social desirability bias, as most respondents want to be considered as responsible citizens and will report they voted. A more accurate question might be to ask if they know someone that did vote in the last election.

Checklist

✓ **Clearly state your intentions with the research.**

- At the top of your survey, write a brief statement explaining why you are collecting the information.

✓ **Reassure privacy of the data.**

- Reassure each respondent that the information is entirely anonymous and confidential.

✓ **Include instructions with your survey**

- What may seem obvious may not be obvious to respondents; especially, if the survey is being completed outside of your company. Include a short introductory set of instructions at the top of the survey and additional instructions for specific questions, as needed. Don't forget to include how the completed surveys should be submitted.

✓ **Don't ask for personal information unless you need it.**

- Asking individuals to provide you with personal or demographic information may irritate some respondents and prevent them from completing your survey. (i.e., sexual orientation)

✓ **Keep the questions short and concise**

- The wording for survey questions should be short, concise, and easy to understand and complete the question.

✓ **Order/group questions according to subject**

- A good way to minimize respondent distraction is to group questions together by subject. This way your respondents can focus their thoughts and answer a series of questions around these thoughts. However, be careful how you word and order your questions because order effects may also cause response bias.

✓ **Test the survey questionnaire**

- Once you have developed your survey, have a few colleagues or subject matter experts complete the survey to make sure there are no problems.

Excel

Below is an example of a survey written in Microsoft Word. With the example below, you will need to create a data dictionary which will enable you to properly code the data in Microsoft Excel. The actual data dictionary follows the survey.

- **Survey : Trends in Associating**

  Q1. How do you stay connected to others in your industry? (Select all that apply.)
  ☐ Association/society conferences
  ☐ Alumni networks
  ☐ Social networking websites
  ☐ Phone
  ☐ Email
  ☐ Blogs
  ☐ Twitter
  ☐ Other: ___
  ☐ None of the Above

  Q2. How do you use social networking sites? (Select all that apply.)
  ☐ Staying connected with friends and family
  ☐ Communicating with others in my industry
  ☐ Communicating with clients
  ☐ Marketing my company
  ☐ Networking/finding professional opportunities
  ☐ Staying connected to a professional community (i.e. Facebook group)

  Q3. Which of the following social networking sites do you use? (Select all that apply.)
  ☐ Facebook
  ☐ Twitter
  ☐ LinkedIn
  ☐ FourSquare
  ☐ Other_____

  Demographic Information

  D1. What is your gender?
  ○ Male
  ○ Female

  D2. What is your age?

  D3. In which industry do you work?

○ Arts/Entertainment/Recreation	○ Automotive
○ Construction	○ Energy
○ Health Care	○ Hospitality

(continued)

(continued)

○ Information Technology	○ Manufacturing and Distribution
○ Marketing	○ Miscellaneous
○ Professional Services	○ Real Estate
○ Retail	○ Telecommunications
○ Transportation	○ Other_____

D4. How many years have you been in this industry?

A data dictionary assigns a value for each response in a question. For example, Question 1 (Q1) is a multiple select question and may cause the respondent to select multiple items. For this type of question, each response will be categorized as an individual question with the coding being 1, if the item is selected, and 0, if the item is not selected.

- **Data Dictionary**

Below is an example of how a data dictionary would appear in an Excel for the survey example.

ID. Respondent Number		
[this is the number assigned to the respondent for record keeping]		
Q1. How do you stay connected to others in your industry?		
(a) Association/society conferences	1 = Selected	0 = Not Selected
(b) Alumni networks	1 = Selected	0 = Not Selected
(c) Social networking websites	1 = Selected	0 = Not Selected
(d) Phone	1 = Selected	0 = Not Selected
(e) Email	1 = Selected	0 = Not Selected
(f) Blogs	1 = Selected	0 = Not Selected
(g) Twitter	1 = Selected	0 = Not Selected
(h) Other	1 = Selected	0 = Not Selected
h_o.	[text response]	
(i) None of the Above	1 = Selected	0 = Not Selected
Q2. How do you use social networking sites? (Select all that apply.)		
(a) Non-professionally (to stay connected with friends and family)	1 = Selected	0 = Not Selected
(b) Communicating with others in my industry	1 = Selected	0 = Not Selected
(c) Communicating with clients	1 = Selected	0 = Not Selected
(d) Marketing my company	1 = Selected	0 = Not Selected
(e) Networking/finding professional opportunities	1 = Selected	0 = Not Selected
(f) Staying connected to a professional community (i.e. Facebook group)	1 = Selected	0 = Not Selected

(continued)

(continued)

Q3. Which of the following social networking sites do you use? (Select all that apply.)

(a) Facebook	1 = Selected 0 = Not Selected
(b) Twitter	1 = Selected 0 = Not Selected
(c) LinkedIn	1 = Selected 0 = Not Selected
(d) FourSquare	1 = Selected 0 = Not Selected
(e) Other	1 = Selected 0 = Not Selected
e_o.	[text response]

D1. What is your gender?
1 = Male
2 = Female

D2. What is your age?
[text response]

D3. In which industry do you work?
1 = Arts/Entertainment/Recreation
2 = Automotive
3 = Construction
4 = Energy
5 = Health Care
6 = Hospitality
7 = Information Technology
8 = Manufacturing & Distribution
9 = Marketing
10 = Miscellaneous
11 = Professional Services
12 = Real Estate
13 = Retail
14 = Telecommunications
15 = Transportation
16 = Other

D3o. Other:
[text response]

D4. How many years have you been in this industry?
[text response]

✧ **Note**: It is a standard practice to code your questions with a **Q** for the general survey questions, **o** for other, and a **D** for demographic questions.

After creating the data dictionary you can begin entering and coding your response data according to the variables you have established.

The following example includes an example of 10 responses to the previous survey. The question labels are in the first column and are read across each row. Each respondent is a separate column.

ID	1	2	3	4	5	6	7	8	9	10
Q1a	1	0	1	0	1	1	1	1	0	1
Q1b	0	0	0	0	0	0	0	0	1	0

(continued)

(continued)

ID	1	2	3	4	5	6	7	8	9	10
Q1c	1	1	1	1	1	1	1	1	1	1
Q1d	0	1	0	0	0	1	0	0	0	1
Q1e	1	1	0	1	0	1	1	0	1	1
Q1f	0	0	0	0	1	0	0	0	0	1
Q1g	1	1	1	1	0	1	0	1	1	1
Q1h	0	0	0	0	0	0	0	0	0	0
Q1h_o										
Q1i	0	0	0	0	0	0	0	0	0	0
Q2a	1	1	0	1	1	1	1	1	1	1
Q2b	1	1	1	1	1	1	1	1	1	1
Q2c	1	0	0	0	1	0	0	1	0	0
Q2d	1	0	1	0	0	0	0	0	1	0
Q2e	0	1	0	1	0	0	0	0	0	1
Q2f	0	1	0	1	1	0	1	1	1	0
Q3a	1	1	1	1	1	1	1	1	1	1
Q3b	0	1	0	1	1	1	1	1	1	0
Q3c	1	1	1	1	1	1	1	1	1	1
Q3d	0	0	0	0	1	0	0	1	0	0
Q3e	0	0	0	0	0	0	0	0	0	0
Q3e_o										
D1	1	2	2	1	2	1	2	1	1	2
D2	25	32	56	45	22	60	38	42	54	35
D3	4	5	6	7	9	11	1	3	13	16
D3o										Tech
D4	2	5	7	15	2	40	10	24	30	6

After you have entered the data, you can analyze the numbers using descriptive statistics, frequency counts, and various other methods.

Final Thoughts and Activities

Practice Problems and Case Studies

1. Below is an example of a survey. Take a look at the questions below and describe how you can improve them.

> 1. Our in store greeting
>   Unacceptable—Bad—Good—Great—Excellent
>
> 2. Store cleanliness
>   Excellent—Great—Good—Bad—Unacceptable

(continued)

(continued)

3. Our awesome in store merchandise
   Not appropriate—Appropriate—Awesome

4. Our overall store selection and bathrooms
   Unacceptable—Bad—Good—Great—Excellent

5. Wait time at the checkout counter
   Too long—Just Right—Perfect

6. Our sincere "Thank you" upon leaving our store
   Yes/No

7. Your overall store experience
   Awesome—Good—Not Good——Poor

2. We ask the survey question "What do you remember about advertising for the Ford pickup trucks during the Super Bowl?" How should the codes be structured for this question?

Discussion Boards

1. How difficult is it to recruit households for the Nielsen panel? Why?
2. The use of Survey Monkey as an online tool has created some interesting ethical issues. Discuss.

Group Activity

1. Compile a list of at least five Census Bureau activities, and explain what type of information is derived from these surveys.
2. Build your own survey around employee satisfaction and test it out on a small sample to get feedback.

Parting Thought

A statistician is a professional who diligently collects facts and data, then carefully draws confusions about them.

Problem Solutions

Answers may vary but here is an example of how the survey items can be improved
   How good would you rate the following store experiences:

1.   Our in-store greeting
     Very Unacceptable —Unacceptable—Neutral—Acceptable—Very Acceptable

2.   Store cleanliness
     Very Unacceptable —Unacceptable—Neutral—Acceptable—Very Acceptable

3.   Our in store merchandise (loaded question)
     Very Unacceptable —Unacceptable—Neutral—Acceptable—Very Acceptable

4A. Our overall store selection (double-barreled question)
     Very Unacceptable —Unacceptable—Neutral—Acceptable—Very Acceptable

4B. Our bathrooms
     Very Unacceptable —Unacceptable—Neutral—Acceptable—Very Acceptable

5.   Wait time at the checkout counter
     Very Unacceptable —Unacceptable—Neutral—Acceptable—Very Acceptable

6.   Our "Thank you" upon leaving our store (Leading loaded question)
     Very Unacceptable —Unacceptable—Neutral—Acceptable—Very Acceptable

7.   Your overall store experience
     Very Unacceptable —Unacceptable—Neutral—Acceptable—Very Acceptable

   Note: (1) The items all should have the same rating scale; should not switch
between rating scale and yes/no question types. (2) All questions should have the
same directional scales; should not switch between starting with very acceptable to
staring with very unacceptable. Pick a direction and stick with it.

2. This question can be set as:

   • an open ended question, thus the responses should be coded based on
     recurring themes and ideas. OR
   • a nominal multiple response select question, thus the responses would be
     coded as 1 for remember an aspect of the commercial or 0 for no response to
     the item. OR
   • a nominal single select question: the responses can be coded using numbers,
     but the numbers are not scale able and are only categorical labels.