Survey & Questionnaire Design

Authored by: Sean Tomlin

Alex McBride

Gina Sigler

31 August 2019

The goal of the STAT COE is to assist in developing rigorous, defensible test strategies to more effectively quantify and characterize system performance and provide information that reduces risk. This and other COE products are available at www.afit.edu/STAT.

STAT Center of Excellence
2950 Hobson Way – Wright-Patterson AFB, OH 45433

Table of Contents

Executive Summary

Surveys can be a very useful data collection tool when designed and used correctly. Surveys, or questionnaires, are particularly useful when the desired data to be collected comes from a large, hard to reach population of interest. Surveys can easily be mailed or sent electronically, eliminating the need to travel to collect the desired data. As a result, a lot of time, money, and other resources can be saved. They can also be used for exploratory analysis. A well designed survey can give researchers more information about the population of interest, allowing them to narrow down and perfect the scope of their project. However, designing and distributing a survey is a tricky business. This best practice will seek to aid in the creation of a survey from research question to planned analysis.

Keywords: survey, questionnaire, bias, sampling, Likert scale, principal component analysis, factor analysis, cluster analysis

When & Why to Use a Survey

There are several pros for utilizing a survey for data collection. The researcher can reach a large group of people fairly easily. A survey can be used for exploratory purposes when researchers are conducting multiple phases of research. They can be efficient and relatively inexpensive to administer. Furthermore, they contain some level of consistency because each respondent is asked the same-worded questions. This eliminates interviewer bias that can be introduced when different interviewers ask participants questions. When designed correctly, results from a survey can be fairly easy to analyze (Gillham 6).

As with any form of data collection, there are cons associated with surveys. First, the quality of your data is completely dependent on the quality and format of the survey itself. A poorly designed survey will result in poor quality data. There is also the risk of non-response bias. Individuals may incompletely answer the survey questions, or fail to complete it entirely. The responses that are received are not immune to issues either. The accuracy of the data could be questionable depending on several factors, such as misinterpretation of the question, respondents' literacy, and respondents' honesty (Gillham 8). One or more of the aforementioned cons can result in the sample data being unrepresentative of the true population of interest. This is a critical issue because the entire purpose of collecting the sample data is to make inference on a particular population or answer a research question.

Initial Survey Design

In general, survey instruments are cost effective for gathering information from a large group of people. Depending on the design strategy, survey instruments can eliminate the need for travel and are economical for researchers working full-time on other projects. A survey instrument starts with a research question that maps to a testing goal in the STAT process.

Research Questions

Surveys must have a clear research question associated with the developed problem statement. This question will drive the scope and nature of the survey. As you break down your STAT problem statement into goals, consider where a statistical survey may provide value. Surveys and questionnaires are often used to characterize the general attitude or feeling of a populace, or for providing subjective data to test a hypothesis (Gillham 6). Unlike the social sciences, surveys and questionnaires in test and evaluation are free from the burden of characterizing sizeable groups of people, such as national populaces. For this reason, survey instruments are effective for quickly gathering data from populations and can be used to guide development of testing goals, factors, and responses in the STAT process. However, many research questions do not require a survey. For instance, if the research question is to characterize body size of some population, these metrics may be accessible through a fitness database. Consider other ways of answering the research question before constructing a survey. Also, be sure to refine the research question and consider whether a survey will accomplish the testing goal. This includes asking what analysis will need to be completed on the data in order to answer the question. For productive data analysis, survey instruments need clearly defined test variables related to the STAT goal.

Test variables

The research question will help generate your variables of interest in the study. This is important to consider early in the design process, since surveys without well-defined research questions and test variables often answer the wrong question in costly detail. The research question comes from the problem statement and goals outlined in the “understanding requirements” phase of the STAT process (Burke). Survey designers need to identify important variables associated with the question that are measurable, of practical significance, and provide a clear route to data analysis. Visualize the question responses in early development to create a survey that gives you the answers you need. Surveys with strong research questions, clear survey objectives, and measurable responses avoid situations where the collected data cannot be analyzed. Furthermore, test variables are easier to form with an understanding of the target population.

Identify the population of interest

Survey designers must identify the target population and any characteristics that warrant special treatment. Designers need a sample that best represents the population and contributes to the research question. The survey should establish clear communication with the audience to accomplish this. A well-designed survey will consider the audience level during question development. Consider population characteristics such as gender, experience as an operator, or professional background. Proper question interpretation is crucial to gathering relevant information about the population. The survey should also motivate the audience to respond. Try to explain the purpose of the survey and emphasize why their participation is critical to the study. In some cases, compensation may be appropriate to increase response rates. Once the target population is identified, develop a sampling plan to collect responses.

Sampling

Survey sampling will determine what segment of the population takes the survey. Survey sampling is a delicate balance between the amount of information necessary to draw accurate conclusions and the amount of resources available. More data will provide more information about the population of interest. However, collecting more data requires more money, time, and resources. The sample frame determines how well the sample will represent the population. Researchers can only draw inference on the fraction of the population that is selected, hence the importance of choosing a representative sample. Probability sampling procedures are sample designs based on planned randomness that allow the researcher to make probabilistic statements about the estimators derived from the sample (Scheaffer et al. 10). Probability samples include simple random samples, stratified random samples, cluster sampling, and systematic samples. Additional information about sampling can be found in the appendix.

How big a sample needs to be depends on the analysis plan and how adequately population subgroups can be sampled. However, specifying a fraction of the population without considering how a population is grouped is not a good sampling plan. The type of probability sample chosen will better dictate the sample size. A cluster sampling method will require a different amount of sample subjects than a simple random sample, for instance. A STAT expert should be consulted with questions regarding sample size calculation.

Question Types

Survey question development is devoted to developing a measure that leads to analysis. Well-designed, unambiguous questions improve reliability of measured responses. Questions should be clear so that every respondent interprets the question the same way. Response choice should be designed in a way that distributes responses across a continuum (Fowler 152). The following question examples presented in this best practice are related to a survey designed for investigating the use of eye safety glasses in a manufacturing setting. We will look at two distinct types of question types: open-ended and close-ended questions.

Open-ended questions do not have predetermined answers. Unfortunately, this means that analysis is arduous. For the respondent, the response task is longer, possibly increasing bias and reducing response validity (Fowler 91-92). However, open-ended questions may be useful to gain new insights from a population segment. These questions are useful to list problems for further investigation. A common use of the open-ended question is to elaborate on a close-ended question. Respondents are asked for additional comments on key issues. In general, avoid use of open-ended questions unless the researcher is prepared to analyze the responses. Example open-ended question:

How do you use your safety glasses on a daily basis?

Close-ended questions force the respondent to select one or more answers. Most questions in surveys are close-ended to save time and create survey flow. Additionally, close-ended questions can be simple to analyze because there are no written responses. Example close-ended question:

Do you have corrected vision such as glasses or contacts? *Yes* *No*

Researchers might also want to gather specific information about respondent demographics. These questions might be closed- or open-ended depending on the goals. Demographic questions provide information for identifying differences in population and matching segments of the population that behave in similar ways (Alreck 24). Often, these questions are best included last so that respondents are motivated to complete questions more relevant to the study.

Types of Closed-Ended Questions

A simple dichotomy is the standard yes-no question. Clearly this may take other forms: agree-disagree, fair-unfair, or true-false (Alreck 118). The response must be binary with contrasting answers. There tend to be longer lists of these types of questions for effective survey flow.

Do you wear safety glasses while on site? *Yes* *No*

A multiple choice question could force the respondent to pick a single choice out of a response set, or it could allow the respondent to select more than one answer out of the response set. These answers should all be mutually exclusive (Alreck 119).

Select your current age group in years:

__18-24 *__25-39* *__40-59* *__60+*

Select all that apply: Do you experience any of the following while wearing safety glasses?

__Improper fit *__Excessive glare* *__Eye strain and fatigue*

Some questions have discrete answers. Others require the researcher to create a scale that allows the user to answer the question. Scales should be simple and clear. Some examples of scales include verbal frequency, ordinal, forced ranking, linear and numeric, or adjective checklist (Alreck 120-135). There is no established way to select a scale criterion.

Many surveys use the Likert scale to obtain positions on certain issues or conclusions. The response scale represents a wide range of feelings on a global issue.

Please rate your comfort while wearing your safety glasses on the following scale:

1. Very uncomfortable 2. Uncomfortable 3. Neutral 4. Comfortable 5. Very comfortable

(Very uncomfortable) *1* *2* *3* *4* *5* *(Very comfortable)*

Researchers must be able to create statements that are typical of a global issue. Typically, five response choices can characterize the response. Rarely, the researcher may consider seven response choices.

Many rating scales include the 'neutral' response. Avoid an even number of response choices because this forces the respondent into a position even if they are neutral (Alreck 120-122). However, be wary of your wording to ensure that the scale you choose measures the response in a way that answers your research question. Consult a STAT expert for help designing question scales that yield measurable responses. More information regarding rating and ranking scales can be found in the Appendix.

Question Development

Develop questions that avoid bias and answer the intended questions. The following are some general guidelines for good questions:

- Have a narrow focus on key issues
- Keep each item brief
- Write with clarity
- Use familiar vocabulary
- Keep sentences simple
- Avoid specific sources of bias or error
- Use structured questions
- Characterize answers carefully
- Select suitable categories
- Use scales effectively (Alreck 114-115)

A question is narrowly focused on the topic of discussion. There are no confusing questions and the respondent knows what is being asked. Question length should be considered as well, since greater response tasks generally increase bias and error (Alreck 109). Using vocabulary that makes sense to the audience will help provide clarity with the responses. Avoid compound sentences and write in plain language. Next, specific sources of bias or error should be avoided. Response bias can be sidestepped with good question development. Unstructured items ordinarily provide large quantities of poor-quality data (Alreck).

The question intent must be immediately clear to the user to get the best response. There are several ways to develop clear and concise questions. First, choose your wording carefully. Often**,** questions are worded in ways that are interpreted differently by respondents. It's also important to know the context of the question. Furthermore, avoid **"**double-barreled**"** questions, meaning those that ask more than one question in the same sentence. This may confuse the respondent and introduce other unfamiliar contexts to the researcher (Redmiles).

Do you feel your safety glasses are comfortable and functional?

The order of the response choices also matters because the researcher can inadvertently introduce response bias by writing leading questions or nudging the respondent to answer questions a certain

way. In the question below, respondents are being led to believe the glasses might not be as safe as supposed.

How do you feel your safety glasses keep you safe on site?

Moreover, the mode of the survey should be effective for the collection environment. This also means that the survey should be a reasonable length. Typically, completing time should not exceed 30 minutes to respect the respondents' time. Another useful tip is to reuse previously developed questions since they have already been field tested in surveys (Redmiles). Composing meaningful questions requires careful consideration of the respondent's ability to answer the questions. The survey should also debrief the respondent accordingly.

Additional Survey Considerations

- The two main measures developed to assess the quality of a survey are validity and reliability. A survey is valid if it correctly measures what it was intended to measure (Boynton and Greenhalgh 1313). A survey is reliable if it yields the same results from different samples, time periods, researchers, etc. (Boynton and Greenhalgh 1313). Taking these measures into consideration helps ensure any differences in the results exist due to differences in respondents who took the survey, not as a result of flaws in the questionnaire design.
- A survey should conclude with some debriefing for the respondent. The conclusions section should address delicate, sensitive, or intimidating information. Demographic questions can also feel threatening to respondents.
- If possible, distribute your survey to a small test group to be done as a trial run. This can identify mistakes and provide suggestions for better survey format. This helps ensure a well-designed survey that is ready for circulation.
- The survey should have a clear flow of ideas when read by the respondent. The survey should be separated into sections by topic or question type. Each of the questions should be related to one another so that the respondent has a clear understanding of the question being asked. For example, try to include open questions near the end of the survey.

Analysis

Analysis of survey data should be carefully considered throughout the entire process. Questions on the survey should be designed so that a proper analysis will answer the research question. Failure to plan the analysis method from the start means that your survey could provide a large assortment of unhelpful data. While there are plenty of methods to interpret survey results, a selection of common methods are covered below. Please note that some of the more complicated methods will require statistical software to perform.

Descriptive Statistics

One simple method of analysis for survey data is to use descriptive statistics. Depending on the information gathered, results can be presented through graphical or numerical summaries. Histograms, boxplots, scatter plots, or bar charts can be created dependent on whether the data is qualitative or quantitative in nature. If the data is quantitative, it might be useful to calculate a mean, median, or standard deviation for each question. The STAT COE has a best practice available, *Statistics Reference Document Series Part 1: Descriptive Statistics*, which contains more information on each of these graphs and numerical values.

Principal Component Analysis & Factor Analysis

Survey questions have a tendency to be correlated, introducing various issues when trying to analyze the data. Principal component analysis (PCA) is a statistical technique used when the independent variables are highly correlated. Put very simply, PCA replaces the correlated variables with fewer uncorrelated variables (principal components) that retain most of the information from the original variables (Bartholomew et al. 117). Each principal component can account for a portion of the total variation in the original variables. The first few principal components typically account for a large portion of total variation; therefore, the other components can be discarded without losing too much additional information (Bartholomew et al. 117). There are multiple approaches to take to determine which principal components to eliminate:

1. Keep the first k components that explain a "large" portion of total variation
2. Analyze the correlation matrix and keep only the components with an eigenvalue greater than 1
3. Use a scree plot (eigenvalue vs. component number) to find and "elbow" where the eigenvalues start to decrease more slowly
4. Consider if the component has a practical/useful interpretation (Bartholomew 124)

Factor analysis is another statistical technique that tries to identify an underlying, common factor responsible for a group of variables with a large within-group correlation but small between-group correlation. Factor analysis and PCA are similar in that they are both dimension-reduction methods. The subtle difference is that PCA tries to explain as much variation as possible, whereas factor analysis is attempting to account for the correlations between variables (Rencher 409). With factor analysis, you can also construct a model, whereas with PCA, you cannot. Both PCA and factor analysis methods are somewhat controversial among statisticians because they are typically difficult to interpret. Neither method should be used without first consulting a STAT expert.

Cluster Analysis

Cluster analysis is a method used to find patterns in a data set by grouping observations into clusters (Rencher 451). No prior information is needed about any of the groups prior to analysis. Instead, this can be done graphically by simply plotting observations. The goal is to categorize similar observations into groups, but the groups as a whole are dissimilar from each other. Distance is typically used as a metric to measure dissimilarity. The farther apart two observations are, the more dissimilar they are. In order to

perform the analysis, a distance measure and clustering procedure will need to be chosen. Often, this method is performed with factor analysis to minimize the number of survey questions that were measuring the same thing. After deciding the appropriate number of clusters, the clusters should be assessed for both validity and interpretation (Bartholomew 17).

Hypothesis Testing

Another typical method of survey data is hypothesis testing. Hypothesis tests are used to answer a specific question in the form of a hypothesis. For instance, do men and women feel differently about how comfortable the safety glasses are for work? This question could be answered by performing a hypothesis test. The kind of hypothesis test is determined by the underlying assumptions that can be made about the survey data. The STAT COE has a tool called *the Interactive Inferential Statistics Flowchart* that can help determine what method might be appropriate. Often, the comparison can be done with a simple t-test or the non-parametric alternative, the Mann Whitney U Test (also known as the Wilcoxon Rank Sum Test). More details about hypothesis testing can be found in the STAT COE best practice, *Statistical Hypothesis Testing*.

Regression

Regression is an analysis tool that allows you to see the relationship between variables. In particular, regression can be used to see if some independent variables have an effect on a dependent variable. One of the main issues with using regression analysis on survey data is the violation of the typical model assumptions. In this respect, it may be necessary to use another analysis method like principal component analysis or factor analysis to first combat the correlation in the individual survey questions. Instead of using the individual questions as the independent variables, principal component regression uses the principal components from PCA as the regressors. See Burke's two Modeling Building Process best practices for additional information on how to properly implement regression. If used properly, the model will be able to describe which factors are impacting the chosen response and to what degrees.

Conclusion

When used properly, surveys or questionnaires can be a useful tool for gathering information. The process begins by coming up with a well-stated research question. It continues with identification of the population of interest and a sampling plan that will adequately reach that population. Questions should be well thought out and the analysis method documented before beginning distribution. Going through all the steps and working through the STAT process can create an effective tool that can save both time and money in the long term. If you find you need additional help or would like a STAT expert to review a survey, please contact us at COE@afit.edu.

References

Alreck, Pamela L., and Robert B. Settle. *The Survey Research Handbook / 3rd Ed.* McGraw-Hill Companies, Inc., 2004.

Bartholomew, David J., et al. *Analysis of Multivariate Social Science Data.* 2nd ed., Taylor & Francis Group, LLC. 2008.

Boynton, Petra M., and Greenhalgh, Trisha. "Hands-on guide to questionnaire research: Selecting, designing, and developing your questionnaire." *British Medical Journal,* vol. 328. 2004. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC420179/pdf/bmj32801312.pdf. Accessed 22 August 2019.

Burke, Sarah et al. "Guide to Developing an Effective STAT Test Strategy V6.0." Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 31 Dec 2018.

Burke, Sarah. "Model Building Process Part 1: Checking Model Assumptions V. 1.1". Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 31 July 2017.

Burke, Sarah. "Model Building Process Part 2: Factor Assumptions." Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 17 July 2018.

Fowler, Floyd J. *Improving Survey Questions: Design and Evaluation*. Sage Publ., 2006.

Fowler, Floyd J. *Survey Research Methods*. Sage, 2005.

Gillham, Bill. *Developing a Questionnaire*. Continuum, 2004.

Kensler, Jennifer. "Statistical Hypothesis Testing." Scientific Test and Analysis Techniques Center of Excellence (STAT COE), 27 August 2019.

Redmiles, Elissa, et al. "A Summary of Survey Methodology Best Practices for Security and Privacy Researchers." 2017.

Rencher, Alvin C., *Methods of Multivariate Analysis.* 2nd ed., John Wiley & Sons, Inc. 2002.

Scheaffer, Richard L., et al. *Elementary Survey Sampling.* 7th ed., Brooks/Cole, Cengage Learning, 2012.

Scientific Test and Analysis Techniques Center of Excellence (STAT COE). "Interactive Inferential Statistics Flowchart." 8 January 2019.

Sigler, Gina. “Statistics Reference Document Series Part 1: Descriptive Statistics.” Scientific Test and Analysis Techniques Center of Excellence (STAT COE), July 2018.

Vegias, Wade M. “Likert-Type Scale Response Anchors.” Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University. 2006.

Appendix A: Survey Sampling

A simple random sample (SRS) is the selection of n sampling units chosen in such a way that each sample of size n has an equal chance of being selected (Scheaffer et al. 76). A SRS can be difficult to conduct because ensuring every unit in the population has an equal chance of being selected is challenging. Stratified sampling is used when you want to ensure multiple, distinct groups are included in your sample so that it is representative of the population. This can be accomplished by dividing the entire population into groups, or strata, and drawing a SRS from each stratum (Scheaffer et al. 116). Cluster sampling has a similar approach as stratified sampling in that you divide population into like groups. However, unlike stratified sampling where you take a SRS within every stratum, a cluster sample draws a SRS of groups and then samples units within only the selected groups (Scheaffer et al. 11). The final sampling discussed here is systematic sampling. A systematic sample is obtained by randomly selecting one element from the population as your starting point, and then selecting every kth element after that (Scheaffer et al. 218). A systematic sample is generally easier to perform than a SRS, and can be more uniformly spread over the population (Scheaffer et al. 219). For more information on survey sampling methods see *Elementary Survey Sampling.*

Appendix B: Likert Scales

As mentioned before, it is important to have the correct wording in questions that are presented to respondents. The goal is the reason out the degrees of opinion on a particular topic. The values in the Likert scale should all be associated with a worded response. This allows for less variation of interpretation from the respondents on exactly what the difference is between a 3 and a 4 in a 5 point scale. There are some commonly used words to go along with questions regarding level of agreement, level of difficulty, or frequency (Vegias).

Level of Agreement:

1-Strongly Disagree *2-Disagree* *3-Neither Agree or Disagree* *4-Agree* *5-Strongly Agree*

Level of Difficulty:

1-Very Difficult *2-Difficult* *3-Neutral* *4-Easy* *5-Very Easy*

Frequency:

1-Never *2-Rarely* *3-Occasionally* *4-A Moderate Amount* *5-A Great Deal*

More information and examples of other possible scales can be found at the following link:
https://www.peru.edu/oira/wp-content/uploads/sites/65/2016/09/Likert-Scale-Examples.pdf