Self Assessment of
Statistics Quality

STATISTICS KOREA

Contents



II Guidelines for steps of Self Assessment of Statistics Quality

The summary of Self Assessment of

Statistics Quality

Survey Statistics

Survey Statistics

Chapter 1

The summary of Self Assessment of Statistics Quality

01 Background

Statistics are the basic data for shaping policies of a country. Accurate statistical data give necessary infrastructure for national management. But, distortion of statistics means a distortion of policies. Therefore, the necessity is more required to manage the statistics quality by building up accurate statistical database.

As the necessity for the management of statistical data, Statistics Korea is conducting a quality assessment on overall statistics of the country in order to identify the whole status of statistics quality. However, it takes a great deal of cost and time to conduct a regular assessment of statistics quality by requesting to service outside.

As a better type compared to the above, Statistics Korea has set up laws of self assessment of statistics quality for organization statistics, under the revised Statistics Law, as a measure to conduct a quality assessment of statistics and maximize its effect by encouraging voluntary participation.

02 Meaning of Self Assessment of Statistics Quality

The system of Self Assessment of Statistics Quality (legal basis)

- Article 11 of Statistics Law and 15 of Enforcement Ordinance clarify that the head of a statistics production agency should conduct a quality assessment every year. The law also states that, when a regular assessment of statistics quality, an irregular assessment of statistics quality, and a self assessment of statistics quality are overlapped, a self assessment could be let off, by which the budget and administration incurred from a quality assessment could be minimized.

- Results for self assessment of statistics quality should be submitted to the commissioner of Statistics Korea till the end of December in the year of the assessment. The commissioner of Statistics Korea should inform a statistics production agency of improvements to make sure the results of the assessment are reflected in the quality of statistics.

Table 1-1 Legal articles concerning Self Assessment of Statistics Quality

Article 11 of Statistics Law (Self Assessment of Statistics Quality)

1. The head of a statistics production agency should conduct a quality assessment on agency statistics ('Self Assessment of Statistics Quality' in the following text) every year. But, in the case that statistics producing period is more than one year, a quality assessment can be conducted in the year of statistics production or the next year.
2. The head of a statistics production agency can let a self assessment of statistics quality off when a regular assessment of statistics quality or an irregular assessment of statistics quality is conducted.
3. The head of a statistics production agency should submit the results for a self assessment of statistics quality to the commissioner of Statistics Korea till December 31 in the year of conducting the assessment.
4. The necessary elements regarding the methods and procedures of self assessment of statistics quality are made as Presidential decree.

Article 15 of Enforcement Ordinance (methods and procedures of Self Assessment of Statistics Quality)

1. The commissioner of Statistics Korea should provide the guidelines for Self Assessment of Statistics Quality including the following parts to the head of a statistics production agency to ensure the head of a statistics production agency conducts an assessment of statistics quality for agency statistics ('Self Assessment of Statistics Quality' in the following text) in a voluntary and efficient way.
   ① Standards of Self Assessment of Statistics Quality
   ② Items of Self Assessment of Statistics Quality
   ③ Procedures and methods of Self Assessment of Statistics Quality
   ④ Additional elements for Self Assessment of Statistics Quality
2. The head of a statistics production agency should obey the guidelines for Self Assessment of Statistics Quality clarified in Clause 1.

3. The head of a statistics production agency should set up a plan of Self Assessment of Statistics Quality including an object statistics, period and methods and application plans of the results of Self Assessment of Statistics Quality and submit the above to the commissioner of Statistics Korea till March 31.

4. The commissioner of Statistics Korea should review the plan of Self Assessment of Statistics Quality submitted, clarified in Clause 3. If the commissioner of Statistics Korea thinks the assessment plan needs to be modified or complemented, he or she should inform the head of a statistics production agency of the modifications before Self Assessment of Statistics Quality is conducted.

- A statistics production agency should conduct a self assessment of statistics quality for agency statistics and, without a valid reason, can not deny an assessment.

- Conducting a self assessment of statistics quality carries legal binding force. A statistics production agency should receive a regular assessment of statistics quality or conduct a self assessment of statistics quality. By doing so, it has to try hard to improve the quality for agency statistics .

- The department of producing a statistics can take advantage of the help of statistics experts outside for the matters needing specialized knowledge.

- In writing a self assessment of statistics quality, each question should contain the parts to be checked in order and the results that the data analyzed should be used to improve the quality of agency statistics.

- In particular, the items of a self assessment of statistics quality, five items of quality dimensions, possibly draw parts to be improved.

03 Assessment factors · quality criteria for each dimension

- In conducting a self assessment of statistics quality, the assessment factors and quality criteria for each dimension come down to five elements: relevance, accuracy, timeliness/punctuality, comparability and coherence. These criteria evaluate how well the assessment satisfies five items of quality dimensions.

■ Relevance

- The question of 'relevance' evaluates how many the statistics data satisfied the demand

of a user in terms of coverage, concept, and contents.

- The question of 'relevance' evaluates how meaningful and useful the statistics data are produced to provide.
- The question of 'relevance' evaluates whether a statistics production agency clarifies the purpose of conducting statistics and figures out the demand of users through experts advising meetings and a survey for users' satisfaction, which is reflected in user's demand.

Accuracy

- The question of 'accuracy' evaluates how accurately the characteristics and size of a population are measured.
- The question of 'accuracy' evaluates sampling errors, non-sampling errors, and differences between provisional figures and final figures, and so on.

Timeliness / punctuality

- The question of 'timeliness' evaluates differences between the standard time of framing statistics and the time of releasing results.
- The question of 'punctuality' evaluates how punctually the period for an official announcement is observed.
- As the standard time of framing statistics and the time of releasing results are getting nearer, the assessment gives a high level of timeliness.
- As the scheduled time for an official announcement is punctually observed, the data shows a high level of punctuality.

Comparability

- The question of 'comparability' evaluates how various standards such as identical concepts, classifications, measurement tools and basic data are compared to other statistical data, despite the difference in time and place.
- Even tough the data shows different coverage, according to time and place, if data methods and standards are identical, the data shows a high level of comparability.

Coherence

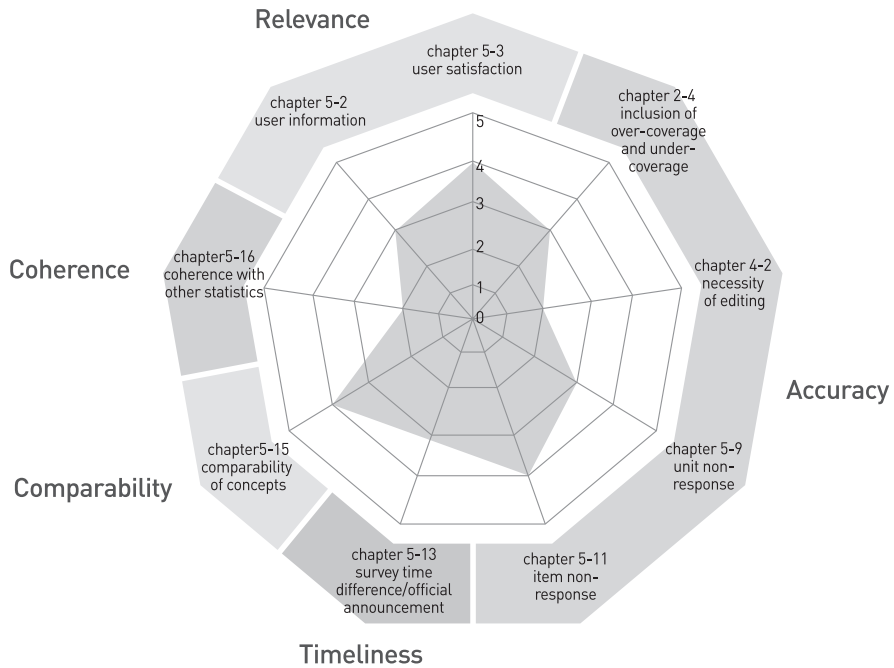
- The question of 'coherence' evaluates how similar data produced by different basic data or making-process for identical economy and society phenomena are presented.
- If data produced by other agencies or showing identical phenomena demonstrate similar results, the data give a high level of coherence.

04 The structure of Self Assessment of Statistics Quality

- Self Assessment of Statistics Quality can be checked, based on the procedures of producing statistics from planning statistics production to post management through 82 items of 7 chapters. However, The matters which are not related to the contents of agency statistics are not checked. Therefore, it means that all 82 items on the checking lists are not filled. The items related to agency statistics are checked.

- Self Assessment of Statistics Quality is required to be conducted only for 5 major items of quality assessment and each chapter consists of each unit for each theme.

Example of making Assessment Diagram of each quality dimension



Survey Statistics

Chapter 2

The procedures of conducting an assessment

01 Time-line for Self Assessment of Statistics Quality

January~March Distribution of Manual and Assessment sheets, setup of basic plan for Self Assessment of Statistics Quality

- Statistics Korea develops and modifies a manual and sheets for Self Assessment of Statistics Quality and then distributes to 159 agencies.
- The agency which conducts a self assessment of statistics quality use the manual as reference, as it sets up plans for a self assessment of quality.
- An assessment agency directly selects target statistics (Clause 1 and 2, Article 11 of Statistics Law excepted-a reason for exception should be submitted to Statistics Korea.
- In setting up a basic plan, items such as a period of assessment and assessment completion and an application of assessment results are specified.

April~May Training for Self Assessment of Statistics Quality(1st training: training as the way of producing statistics, including survey, report, processing)

- The person in charge of an assessment agency participates in training initiated by Statistics Korea.
- The training is preparedly needed to seek out advices of experts or do a self assessment of statistics quality together.
- Preparing basic data related to statistics.

June~August **Conducting a self assessment of statistics quality**

- Finishing assessment sheets by completing an assessment by the finishing day of an assessment planned
- Assessment sheets should be substantial in contents, avoiding false information or letting an assessment item empty.
- Attaching specific data or presenting views during assessment procedures (assignments needing improvements need attaching data or presenting view)
- Assessment should be conducted with the reference of guidelines for Self Assessment of Statistics Quality. In the case that specialized knowledge is needed, an assessment is compiled out of views of experts.

September **Training for Self Assessment of Statistics Quality (2nd Training)**

- Training about questions during assessment procedures or about matters that require attention, as monitoring assessment results.

October **Drawing assignments needing checking, improvements and setting up a application plan**

- After completing assessment results, inconsistency or missing parts in contents should be checked.
- Drawing improvement ideas in each dimension or assessment procedure

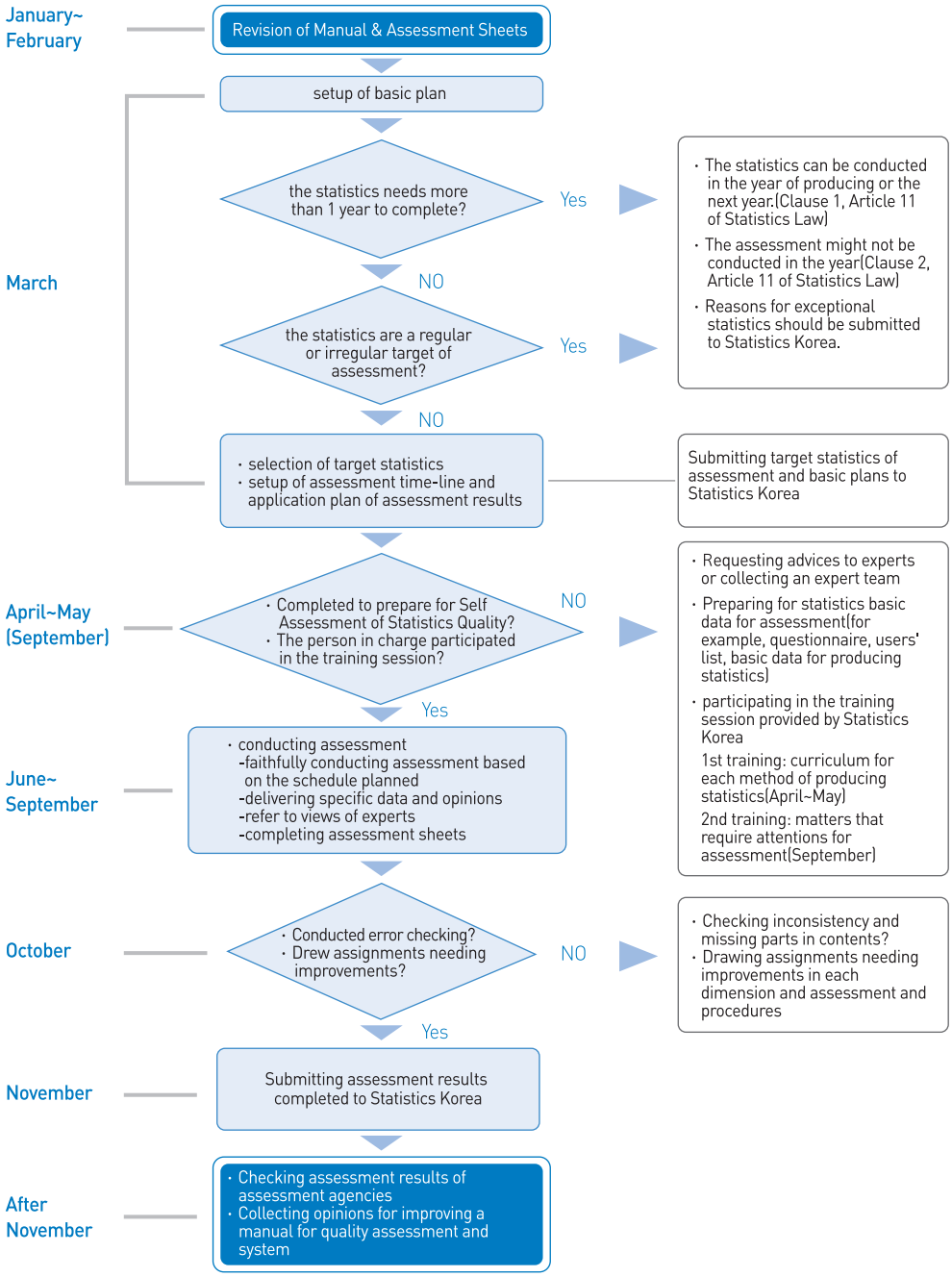
November **Submitting the results of Self Assessment of Statistics Quality**

- After the person in charge of producing statistics checks the assessment contents, a statistics supervisor should do a final check and confirm the data and submit the results of Self Assessment.

After November **Checking assessment results and collecting opinions**

- Statistics checks the results of Self Assessment of Statistics Quality by an assessment agency
- Collecting opinions of assessment agencies for the purpose of improving a manual for Quality Assessment and System.
- Evaluating assessment results, drawing opinions for improvement ideas, and flowing back

Work-flow of Self Assessment of Statistics Quality



Guidelines for steps of Self Assessment of

Statistics Quality

Survey Statistics

Chapter 1

Planning Statistics Production

In order to collect accurate and good data through a survey, it is most important to stipulate the purpose of the survey. The purpose of a survey demonstrates for which the survey is conducted, that is, the direction of a survey, and for whom the survey is conducted, that is, the definition of user groups. That is, a specific description about the purpose of a statistics survey is required because the outset of the survey includes user targets for the data collected from the survey.

Guidelines

- The purpose of producing statistics is stipulated to make sure a survey is efficiently conducted with the given cost fixed for demands of statistics users and statistics agencies.
- The purpose of producing statistics is stipulated after enough reviews by a user advisory group to satisfy the demand of users.
- The purpose of producing statistics should be periodically reviewed in the changing future environment of producing statistics to meet the demand of users. To that end, collecting views of experts could be a solution.
- Statistics producers should cleary demonstrate the legal basis of statistics and the purpose of producing statistics.
- In approving statistics production, the statistics should be conducted based on the statistics period already applied. The change of the period requires the consultation with of Statistics Coordination Division. Then, the modifications should be notified to users.

Guidelines for completing assessment sheets of quality

In order to collect accurate and good data through a survey, it is most important to stipulate the purpose of the survey. The purpose of a survey demonstrates for which the survey is conducted, that is, the direction of a survey, and for whom the survey is conducted, that is, the definition of user groups. That is, a specific description about the purpose of a statistics survey is required because the outset of the survey includes user targets for the data collected from the survey.

1-1 Whether the statistics are announced in the recent year or not

Statistics should be produced, based on the period applied at the time of approval and a thorough plan and preparation for conducting statistics are required. This question checks the recent year when the statistics by a statistics agency is announced.

1-2 Subject of conducting statistics

This question checks whether an agency which gains the approval of conducting statistics is conducting statistics by itself, or the statistics are conducted by contract workers or other service organization or not. In the case that contract workers or other service organization conduct statistics, the information about the year of conducting, the name of the service organization or contract workers, and the parts conducted by another organization, for example, conducting survey and analyzing statistics, should be indicated. For a contract organization to conduct some parts of conducting statistics should be checked as 'conducting statistics by a contract organization' in the question.

1-3 Whether a publication clearly states legal basis and the purpose of conducting statistics or not

Statistics should be conducted based on legal grounds such as laws, regulations, and guidelines. A statistics agency should establish legal basis for planning to conduct statistics and cleary deliver the purpose of conducting statistics in the publication. This question checks whether the above is met or not.

Analyzing the demands of users is related to 'relevance' out of the dimension for statistics quality assessment. Therefore, in order to analyze the demands of users, the information such as who are using the statistics, how many users are using the statistics, and how important the statistics are significant to users, is needed. To that end, first, the user groups

for the statistics and also importance of statistics users need classifying.
In setting up a plan of conducting statistics, users information should be collected and the fact whether the demands of users are appropriate for the statistics should be examined or not. If there is a demand of users which is not satisfied, the cause should be analyzed to make sure the demand is satisfied. At the same time, to analyze the demand of users, it might be considered to operate a user advisory group for statistics.

Guidelines

- In order to analyze the demands of users, the information about as many users as possible should be systematically collected and managed.
- In order to gain the information about user groups, a survey of users' satisfaction or readers of statistics publication are analyzed.
- By operating a user advisory group for statistics, not only the present and future use of statistics should be prepared.
- Through a user advisory group, unsatisfied demands for statistics should be recognized, for which a appropriate plan should be set up.

1-4 Management of Statistics users

In order to analyze the demands of users, the information about as many users as possible should be systematically collected and managed.

A list of a user (form)

name	organization	telephone number	E-mail	users' classification	etc.

※ How to collect a list of user

- A list of user, who purchase statistics publications from the places where publications are distributed for free or for pay
- A list of user, who use statistics data or micro-data from the organization

- A list of user, who participate in a statistics experts meeting or advisory meeting
- A list of user policy customers related to the statistics
- A list of user of library and civil application center
- A list of member of an association related to the statistics
- A list of user, who use the statistics in various ways (for example, people who access statistical data of web site and people who cast questions or request data to an organization by telephone or e-mail)

1-5 Rank the key user of the survey

Key users who make use of the statistics produced from a statistics agency are ranked from 1st to 5th and checked.

1-6 Procedures of seeking out users advices (double selection possible)

When a statistics production agency seeks out users advices concerning statistics, it is checked various routes, for example, which procedures are done, namely, whether statistics are produced by a consultation with a related department or a meeting with user groups or not, whether statistics are produced by the decision of special experts or not, and whether the need of producing statistics is satisfied by a marketing activity or not.

1-7 Users' major requests

Relevance of statistics can be boosted, by figuring out the demand of users for the statistics being produced by a statistics production agency. To that end, the requests of users should be input, based on the importance of requests.

For example: No. 1) Improvement in survey questions
No.2) Timeliness of providing statistics data
No.3) Improvement in methods of supplying data

1-8 Plan to improve users' requests

Here, the plan to improve users' requests, which are presented in 1-7, is described.
For example: In the procedures of producing statistics data last year, unnecessary survey questions should be omitted. The period of supplying statistics data should be reduced. The method of providing micro-data and meta data should shift from off-line to web or e-mail.

Chapter 2

Designing Survey Statistics

The essential items such as the concept, definition and classification standards of statistics should be clearly specified at the design of survey statistics.
The concept of statistics means to beforehand define various characteristics in planning a survey. Those characteristics contain a population, a list of a survey target and classification standards. An unclear description of the concept leads to not reaching the purpose of conducting a survey and giving meaningless statistics results due to confusion over a concept. In addition, the unclear description of a concept leads to no comparisons between two similar statistics due to heterogeneity.
To solve this problem, it is desirable to first review whether there are another statistics similar to the statistics currently being planned or not, when items such as the concept, definition and classification system of a survey are formed. If you find unsatisfactory parts in the existing statistics, a more specific and clearer description of an improved concept should be introduced. If errors, overlaps, or missing parts occur in the classification of statistics, the cause and improvement measures should be delivered and the standard classification system should be preferably used.

Guidelines

- To prevent confusion, in comparing and using a concept, definition and classification system, the domestic and international standard system should be pursed.
- When there is no an existing standard classification standard system, a classification system being commonly used in related statistics.
- When there are no matching parts, compared to other similar a classification standard of statistics, the description of the parts prevents users from being confused.
- To extend use of statistics, it is desirable to subdivide classification the system in the most specific way.

Guidelines of completing quality assessment sheets

2-1 Whether there are applicable statistics for reference in producing statistics or not

You should figure out whether there are statistics similar to the statistics an agency is producing or not and then judge whether the statistics can be applied for quality improvement or not. Among similar statistics, if there are statistics applicable to sampling methods, estimation methods, and data-processing methods, the statistics should be indicated.

2-2 Whether the classification standards are used

You should figure out whether there are a variety of classification standards applied to statistics of a statistics production agency. Even though the same contents are surveyed, applying different classification standards might bring different statistics. Therefore, the application of the standardized classification standards should be used for comparison. For example, in the survey of companies, it is desirable to use KSIC(Korean Standard Industrial Classification) for industrial classification and to use the definition of Statistics Korea for the definition of company. It is better to use International Standard Classification, as the international comparison is needed.

The following items should be checked in the assessment sheets.

1) As the standards approved or used in Ministry of Knowledge Economy, Statistics Korea, Ministry of Land, Transport and Maritime, etc, (for example, KSIC(Korean Standard Industrial Classification), Korean Standard Classification of Occupations of Statistics Korea, Korean Standard Classification of Diseases, and SKTC(Standard Korean Trade Classification) of Statistics Korea, etc.)

2) As the standards delivered by international organizations including ILO, OECD, WHO, UN, IMF, etc, the name of the organization is filled in.

3) As there are no specific domestic or international standards used in producing statistics or the standards presented by expert meetings.

2-3 Whether changes in questionnaire, the number of samples, classification standards and production methods is approved or not

The item should be checked, when there is a change in questionnaire, the number of

samples, classification standards and production methods, which is later approved, to let statistics keep up with reality and boost precision after the approval of producing statistics is confirmed. In this case, the change is required to be approved. Therefore, an agency should request to the Statistics Coordination Division.

2-4 Whether the list of survey targets is accurate or not

Survey statistics should evaluate inclusiveness of survey targets. An obsolete list of survey targets might be possibly too small or too big, which leads to the results of estimation not reflecting the reality. For the list of companies, the change of a business field or the closure of businesses leads to having too large number of survey targets. The change should be indicated to make sure appropriate measures are implemented and survey targets give representation. For the survey of household units, redevelopment or new households, by which the households were not surveyed during Population Census, leads to having too small number of survey targets. It is because the change of the households is not reflected in the survey. Therefore, it should be first evaluated how much the list of targets is suitable for the purpose of producing statistics, which should be reflected in the estimation.

The following items should be checked in the assessment sheets.

1) As immediate improvement is needed for a large gap with a list of targets(more than 20% of the list is not proper for the purpose of statistics)

2) As future improvement is needed, though there is a little gap with a list of targets(more than 10% to less than 20% of the list is not proper for the purpose of statistics)

3) As the list is possibly used, though there is a little gap with a list of targets(more than 2% to less than 10% is not proper for the purpose of statistics)

4) As there is almost no gap with a list of targets, which shows the accuracy of a list of targets(less than 2% of the list is not proper for the purpose of statistics)

5) As there is no difference from a list of targets, which shows the clear accuracy of a list of targets(the list does not show different from the list of targets for the purpose of statistics)

6) As there is no recognition for the difference from a list of targets

2-5 Revision rates of a list of survey targets

A list of survey targets for producing statistics should be arranged prior to data collection. This item checks a complementary ratio before the survey because the need of revising

target information such as business closure, moving, and change of business fields occurs.

The following items should be checked in the assessment sheets.

1) As there are a lot of revisions in a list of survey targets(more than 20% revisions out of a list of survey targets)

2) As there are considerable revisions in a list of survey targets(more than 10% to less than 20% of revisions out of a list of survey targets)

3) As there are some revisions in a list of survey targets(more than 5% to less than 10% of revisions out of a list of survey targets)

4) As there are a few slight revisions in a list of survey targets(more than 1% to less than 5% out of a list of survey targets)

5) As there are few revisions in a list of survey targets(less than 1% out of a list of survey targets)

Examples

For the survey of companies, business populations carrying information of change of business fields, new businesses and relocation, should be revised every year. For the survey of households or population, the households or a population carrying information of moving, redevelopment, and new building, should be set up.

Sampling design describes how to extract sampling method, representing a population, which is only applied for survey statistics. Sampling design is an essential procedure, when it comes to extracting samples out of a population as a survey target in a possibly objective way, evaluating accuracy of samples extracted, and deciding what methods are applicable for a survey.

For the above, the followings are considered.

- Selecting samples, considering the characteristics of a population
- Design and contents of questionnaire
- Verification of questionnaire
- Response burden
- Questions of recognizing sex

Guidelines for sampling methods

- In selecting samples, statistical characteristics(sex, industry, region, occupation, etc) should be considered.
- The methods of extracting samples are confirmed, considering such factors as survey convenience, survey costs, and response burden.
- Considering the methods of picking up samples applied in the existing similar survey, a better method of sampling should be applied.
- The methods of sampling applied in the survey should be specified.

Guidelines for determining a sample size

- A sample size is determined to ensure a sample satisfies target precision and its size shows reality to make a survey possible.
- As a similar survey was done before, the information is applied in deciding a sample size.
- A possible allowable error for a survey estimate should be specified.
- Prior to deciding a sample size, you should determine whether estimates of domains, subsets of a population are needed or not.
- In the process of deciding a sample size, as complex sampling design is considered, it is desirable to obtain and use a design effect.

Guidelines for completing quality assessment sheets

2-6 Characteristics considered as for sample selection for statistics production

As for a scientific sample survey, Probability Sampling Methods are generally applied. Samples should be selected, which fully represent a distribution type of a population. Statistics characteristics such as sex, region, occupation, age, and revenue, should be considered to make sure a sample represents its population.

For the above, the followings are considered(double selection available)

1) Occupation

2) Sex

3) Region

4) Age

5) Industrial classification

6) Statistical characteristics not considered

7) Others(other characteristics excepting the above are considered, indicate on the sheet)

2-7 Reasons of not considering statistical characteristics

It is desirable to consider statistical characteristics to make sure samples represent a population. However, as collecting a population is difficult or a specific population does not exist, you might select a method without considering the characteristics. If you extract samples without considering statistical characteristics, you need to describe the details of the reasons, which helps users understand the statistics.
In the assessment sheets, the reasons should be described as followed.

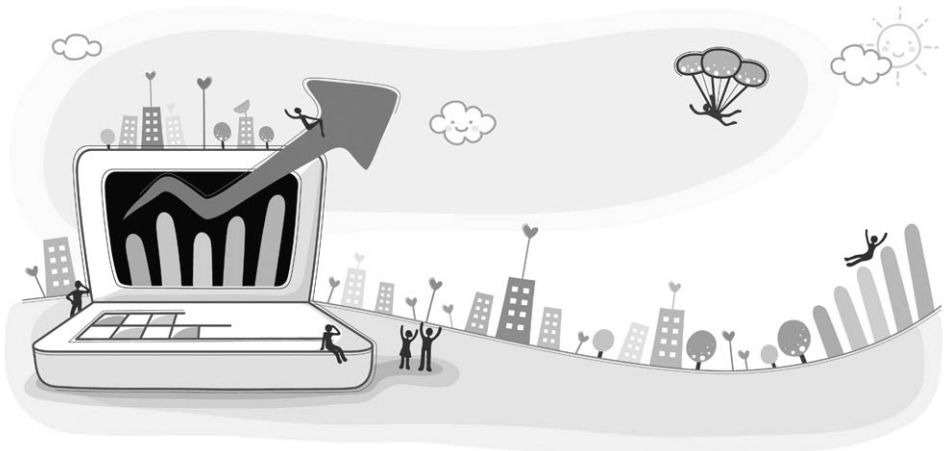
(Example) A sampling method without considering statistical characteristics is applied, because collecting a population for sampling design of the survey is difficult.

In survey statistics, a questionnaire is a tool of measuring the characteristics of survey targets. It is meaningless collecting data by inappropriate questionnaire. Therefore, designing a questionnaire demands great caution. The reasons for measuring errors taking

place in the survey are as followed.

- Survey tool: a questionnaire or measuring tool, which used for data collection, might have some problems, which leads to obtaining false values.
- Respondent: unconsciously or consciously, respondents could make a wrong answers.
- Surveyor: surveyors could affect the answers of respondents.

The first element to be considered in designing a questionnaire is to define what data should be collected. Therefore, it is desirable to design a questionnaire after sufficient discussions between user groups and statistics producers. Survey results are significantly affected by the structure of a questionnaire and respondents' attitude, and the relationship between a surveyor and a respondent. Accordingly, a pilot survey should be conducted to collect the structure of a questionnaire, before the final questions of a survey are fixed. A pilot survey should consider the items such as the form of a questionnaire, logical flow, terms, clearness of classification system, response burden, surveyors and respondents' understanding to survey questions, interview time, sensitivity of survey questions.
In this sense, the following elements should be considered: demands of statistics users and a survey agency, survey purpose and characteristics of a respondent group, situation of survey regions, and demands for data processing.



Guidelines

- Sufficient discussions should be conducted to make sure the demands of a statistics user group are properly reflected.
- A preliminary question for survey questionnaire established in the stage of designing a questionnaire, should be conducted for verification.
- Results table should be completed in advance, considering future data processing.

A questionnaire desirably uses the standardized definition and classification system. The concept, definition, and classification standard used in a questionnaire should be clarified on a related guideline book. To prevent users from feeling confused about the concept different from a general one, a clear description of the concept is required. For example, for a households survey, the definition of 'Household' used in Statistics Korea is 'a unit of sharing livelihood such as cooking and sleeping by one person or above'. In this case, now that the household could contain a person without blood ties, 'family', 'home' and 'domestic establishment has a different meaning from a household.

A questionnaire should be created to make sure all the questions are easily understood and answered by respondents. In addition, for the question which is not easily answered by a respondent, an extra question should be made to help a respondent easily answer to the question. Questionnaire should possibly consist of enclosed questions. For non-enclosed questions, additional open questions should be made. By checking the logical flow of questions, response errors should be prevented.

Guidelines

- The definitions of terms used in a questionnaire should be clearly described, which is recorded to make sure a surveyor can refer to the information.
- Standardized definitions and classification system should be applied in a questionnaire.
- Questionnaire consists of the proven questions with the viewpoint of respondents. In particular, terms, response time, and logical flow should be reviewed.
- At the beginning of a questionnaire, information such as a survey title or topic, survey purpose, requesting cooperation to respondents, survey agency, privacy policy for respondents, and contact information should be reviewed.
- Each title and introduction of each chapter is desirably indicated to make sure a questionnaire is easily identified.
- To call respondents' attention to a questionnaire, a variety of design could be selected and also pictures or illustrations could lead to users' interest.
- In a periodical survey, consistency of questions is essential, so missing parts or new questions from an existing questionnaire should be reviewed.
- The end part of a questionnaire should indicate the expression of thanks for survey cooperation to respondents and also addition space for respondents to fill their comments in should be preferably created.

Guidelines for completing quality assessment sheets

2-8 Procedures for designing a questionnaire (double selection available)

A questionnaire should be created to make sure respondents easily understand and answer to the questions and the flow of questionnaire is produced by an oriented process. To that end, it is desirable to design and verify a questionnaire by taking helps from an internal agency or experts. The procedures of designing a questionnaire could be a barometer of a high quality of survey.

The following items should be checked in the assessment sheets.

1) As a questionnaire is created with the help of a questionnaire expert in the agency,

2) As a questionnaire is created with the help of a questionnaire expert out of the agency,

3) As a questionnaire is created by the person in charge in the agency,

2-9 Questionnaire's Contents (double selection available)

A questionnaire should give trust to respondents. The contents including the necessity of a survey and cooperation request should be included in the questionnaire.

The following items should be checked in the assessment sheets.

1) As the title of a survey to produce statistics is indicated on the questionnaire

2) As the purpose of a survey to produce statistics is indicated on the questionnaire

3) As the legal basis for producing statistics are indicated on the questionnaire

4) As the logo of National Approval Statistics and Production Approval No. are indicated on the questionnaire

5) As cooperation request to respondents for a survey and a thankful expression for survey cooperation are indicated on the questionnaire

6) The name of a survey agency is indicated on the questionnaire

7) Privacy Policy for respondents' information collected in the survey is indicated on the questionnaire

8) Contact information for questions and inquires for statistics is indicated on the questionnaire

Example : contents of a questionnaire

승인(협의)번호
제 30121호

survey agency

Survey in September, 2007

Survey on Loan Types of Financial Institutions

(Trend from July~September, 2007 & Prospect from October~December, 2007)

Name of financial institution		Writer	Division:	Name :
			Telephone:	FAX :

1. The survey has the purposes of using the survey data as basic data to set up the direction of bank monetary policy. If you deliberatively answer to the questions of the survey, I would appreciate for your cooperation.

2. The questionnaire of the survey is produced, considering executives supervising credit of financial institutions. Vice-directors in charge of credit are sure to fill in the survey(if necessary, refer to related-workers' opinions)

3. We would appreciate if you provide reference data for detailed description.

4. Information of the agency related to the survey, I inform you that the survey information will not used for other purpose except for the purpose of the survey.

_____ Agency

※Contact Analysis of Division for inquires related to completing a questionnaire : (02)750-6820 FAX: (02) 750-6892

2-10 Average time for filling in a questionnaire

As for respondents burden, whether the average time to fill in a questionnaire is proper, should be reviewed. As for a general survey, 20 to 40 minutes of the average time is proper. As for a telephone survey, 3 to 5 minutes is appropriate. However, as for a daily-keeping ledger or household ledger which is regularly written on a daily or monthly basis, respondents have a huge burden for response, for which a policy for appropriate compensation should be considered.

The following items should be checked in the assessment sheets.

1) Less than 10 min : a questionnaire consists of brief questions easily answered.
2) 11~20 min: a questionnaire consists of a few questions easily answered.
3) 21~30 min: a questionnaire consists of a reasonable number of questions easily answered without any burden.
4) 31~45 min: a questionnaire consists of a little many questions which needs the attention of respondents.
5) More than 46 min: a questionnaire consists of many questions which need the persuasion of a surveyor.

2-11 Verifying methods of a questionnaire

A pre-survey verifies the response of respondents and the structure of a questionnaire. Along with it, surveyors' opinions are collected to improve a questionnaire. As the methods for verifying the quality of a questionnaire, there are Focus Group Interview, In-Depth Interview, telling thoughts, paraphrasing, etc. A method of dividing into two groups of survey targets and analyzing response values could be considered to improve survey questions.

The following items should be checked in the assessment sheets.

1) As a problem is improved after a pilot survey to a small number of respondents in order to figure out the problems of a questionnaire produced
2) As a problem is improved through qualitative verification for behaviors and understanding of respondents: Focus Group Interview, In-Depth Interview, thoughts-telling interview, paraphrasing interview, etc.

- Focus Group Interview: as to collect data about survey purpose and survey questions, based on a survey of a small number of respondents or the above(5~10 people)
- In-Depth Interview: as to collect data about potential behaviors, attitude, thoughts of respondents for a survey topic and also figure out inner thoughts of respondents
- Thoughts-telling interview: as to collect data about behaviors, intentions, and minds of respondents, which requires respondents to show various behaviors and tell thoughts
- Paraphrasing interview: as to objectively review questionnaire by paraphrasing the thoughts and information respondents already have in mind in a way they easily understand to help understand survey questions from the viewpoint of respondents

3) As survey targets are divided into two groups to analyze response values(separate sample verification)

4) Others(write down another method of verifying another type of a questionnaire)

2-12 Whether there are questions recognizing the distinction of sex or not

This question checks whether there are questions related to the distinction of sex in a questionnaire or not. In order to make a detailed comparison of sex, questions related to the distinction of sex should be applied.

Example : Questions related to the distinction of sex

Survey on International comparisons of youth values

Hello.

National Youth Policy Institute, a government-funded research institute under the Office of Prime Minister, is charge of conducting a variety of surveys and research and developing policy related to Youth.

The institute conducts a survey which compares values of Korea, China, Japan youths, supervised by Ministry of Health and Welfare.

The survey has the purpose of understanding comprehensive values of individual, family, school, society, and country of youths in 3 countries mentioned and making comparisons of them to form a better policy for youths.

Your answers will be statistically processed and protected by Privacy Policy, according to Article 33 and 34 of Statistics Law. We kindly ask you to answer the questions as you think as usual.

Children Youth Policy Department of Ministry of Health and Welfare : 02-2023-8726
National Youth Policy Institute : 02-2188-8886

SQ1. Sex?
① Male ② Female

SQ2. School type?
① Middle school ② Academic high school ③ Special high school

Chapter 3

Collecting Data

Statistical data are divided into survey data, administrative data, and processed data, according to the methods of collecting data. Survey data are data collected by a survey and administrative data are data collected by administrative procedures. processed data are primary data processed by using survey data or administrative data.

Survey data are generally collected by a questionnaire. A questionnaire is divided into face-to-face interview and self-administrative interview, according to the methods of filling in a questionnaire. In addition, a questionnaire, based on the types of a questionnaire, is divided into PAPI(paper assisted personal interviewing), CAPI(computer assisted personal interviewing) and CATI(computer assisted telephone interviewing),etc. CAPI is a survey method where a surveyor obtains survey data, input the data, and identify errors in the field by the input program equipped in the computer. This method has advantages of taking a short time to process data and pursuing accuracy. However, the method has disadvantages of a large burden of early operation costs including developing a program for a perfect survey and input and purchasing computers. Along with it, there is CASI(computer assisted self interviewing), a better type of CAPI where respondents directly input data. CATI is a combination of a computer and a telephone method. The method has a surveyor sit in front of the screen of the computer connected to a telephone, phone a respondent and read questions, and input data on the computer. In this case, As soon as data are input, a logical flow of data is automatically reviwed to recognize errors. In addition, a combination method of survey data and administrative data is called a mixed mode. In this case, the connection of survey data and administrative data needs considering the quality of administrative data.

Guidelines

- In collecting data, a method of boosting the quality of data and collecting data efficiently should be selected.
- As survey data, administrative data, and registered data are used, the need of connection and its method should be reviewed and possible problems at the connection should be fully reviewed to minimize errros.
- As for Self-administrative interview where a respondent fill in a questionnaire firsthand, its guidelines should be made to be easily understood by a respondent.
- As for CAPI, a module for non-response should be applied.
- As for CATI, a surveyor should minutely explain the importance and value of a survey and give simple and clear questions.

Guidelines for completing quality assessment sheets

3-1 Survey method for compiling statistics

A survey method for data collection is divided into interview and self-administrative interview. Self-administrative interview is divided into web survey, CASI(computer assisted self interviewing), and mail survey. Self-administrative survey is divided into paper questionnaire and computer questionnaire, according to the types of a questionnaire. It is desirable to select a survey method which can be most suitable for the reality of a survey in an efficient way.

The following items should be checked in the assessment sheets.

1) Face-to-fact interview by a surveyor

1)-1 As a surveyor uses a computer and input survey contents in the field

1)-2 As a survey is conducted by using a questionnaire out of paper, a traditional media

1)-3 Computer assisted telephone interviewing(general telephone survey included)

2) Self-administrative survey

2)-1 As a respondent fills in the questionnaire on Internet

2)-2 As a respondent fills in the questionnaire in the computer assisted survey

2)-3 As a respondent fills in the paper questionnaire or daily-keeping book

2)-4 Mail survey

3) Others

Self-administrative survey

3-2 Guidelines for a questionnaire

As for a self-administrative survey, respondents should fill in the questionnaire. In this case, a surveyor should provide them with references(advices, examples, major errors) for a questionnaire to increase the understanding of the survey and to prevent from confusion. Therefore, it is most significant to use terms respondents can easily understand, which is different from a surveyor interviewing.

The following items should be checked in the assessment sheets.

1) As references(advices, examples, major error) for a questionnaire are provided

2) As references(advices, examples, major error) for a questionnaire are not provided

3-3 Solution to inquires of respondents (double selection available)

As a respondent can not obtain enough information for survey guidelines while filling in the questionnaire, it is required set up solution policy including using a free call(customer center), and forming solution team for inquires.

The following items should be checked in the assessment sheets.

1) As a free call is available(customer center)

2) As solution team for inquires is formed

3) As Q&A board is open on Internet

4) Others.

CATI or telephone survey

3-4 Usual places for a telephone survey

As for a telephone survey, you should check where to conduct a survey. For example, it should be checked whether a telephone survey is conducted in the office of statistics production agencies, a call center equipped with additional facilities, in the office of outside contractors or in other places or not.

3-5 Statistical calculation

As for a telephone survey, the calculation of final rates of respondents and how many times a surveyor contacts a respondent to obtain answers for a questionnaire, is a barometer of efficiency and accuracy of a telephone survey. In addition, recording a contact time by phone is a good way to increase the response rate of a telephone survey in the future. If you check the final frequency of responses out of the total of contacts, its response rate will be calculated.

The following items should be checked in the assessment sheets.

1) As statistics for the total number of telephone contacts is calculated to obtain a response rate

- Response rate = the number of telephone answered/total number of telephone contacts
- The number of telephone answered indicates that a surveyor completes a telephone survey.

2) As statistics for the total number of telephone contacts is not calculated

3-6 Management of data collected

The management of collected data is directly related to Privacy Policy of respondents. The question checks how the data collected are processed after the completion of a survey. The

data surveyed should be submitted to a survey manager. The manager should keep the data in a safe place.

3-7 Difference between target sample size and sample size surveyed(collected)

In general, there might be a gap between target sample size and sample size surveyed (collected), due to non-response and coverage issues of a list of survey targets(sampling frame). Therefore, you should determine sample size, which should be compared with the actual sample size after a survey is completed. The higher the gap between target sample size and sample size surveyed, the higher survey errors are. For this reason, the results should be necessary to be checked.

The following items should be checked in the assessment sheets.

1) As target sample size stands at 1000, in designing a survey and sample size surveyed less then 950
2) As target sample size stands at 1000, in designing a survey and sample size surveyed more than 950 to less than 1000
3) As target sample size stands at 1000, in designing a survey and sample size surveyed more than 1,000 to less then 1,050
4) As target sample size stands at 1000, in designing a survey and sample size surveyed more than 1,050

3-8 Using survey data as administrative or registered data

In general, a survey statistics is used as administrative or registered data for statistics production. In this case, this question checks whether data are used as administrative or registered data for statistics production or not.

3-9 Types of connecting data with administrative or registered data

A certain type of statistics can be written by using a survey data, administrative or registered data at the same time. In this case, individual objects(unit) of survey data and individual objects of administrative or registered data are connected to obtain statistics. Types of connecting individual objects of both data are as followed: directly connection by using identification number of individual objects and statistically connection by figuring out statistical characteristics of individual objects, and so on.

The following items should be checked in the assessment sheets.

1) As survey data and administrative(registered) data are connected by using each identification number
2) As individual objects are connected by using common variables of survey targets, because there is no identification number of administrative(registered) data
3) Others
4) As not used

3-10 Problems of using other data(administrative & registered data)

When administrative and registered data, except for survey data, are used, you should review what possible problems will emerge. If the use of administrative and registered data causes a serious problem, the connected use should be deliberated.

The following items should be checked in the assessment sheets.

1) As there are some problems for coverage range of the data(representation) that will be connected
2) As there are some difficulties for connection due to the difference of classification
3) As the data that will be connected are not updated, which causes an improper level of timeliness and efficiency
4) As the data that will be connected are overlapped or have a low possibility to succeed due to the numerous inclusion of multiple units(as many objects are connected with one object, not one to one)
5) As there is a difficulty of keeping the time series of data that will be connected
6) Others
7) As there is no problem at the connection

3-11 Data security

In the process of sharing data, the data should not be leaked. In this sense, this question evaluates data security.

The following items should be checked in the assessment sheets.

1) As there is a high possibility of leaking data due to insufficient security data, in sharing administrative and registered data
2) As there is a slight possibility of leaking data, though the data security is satisfying, which need to be improved, in sharing administrative and registered data
3) As there is no possibility of leaking data due to sufficient security data, in sharing administrative and registered data

A surveyor is one of the most essential factors for a survey, which determines the quality of a survey and also statistics. It is quite significant to hire and train surveyors, because the understanding, attitude, and a sense of duty of a surveyor affects the answers of respondents. A man power for a survey is divided into three groups: (1) a survey expert, (2) a survey manager, (3) a surveyor. A survey expert is a kind of expert for survey statistics, who figures out the whole parts of a survey, management, design of a sample questionnaire, and data processing in a detailed way and also understand the various elements of a survey. A survey manager is a person who has practical experiences for a survey and performs surveyors training, pre-checking to the field, supervision of the field, data input, and editing-checking by understanding actual survey procedures. A surveyor is a person who does a survey and input survey data by interviewing respondents in the actual survey field.

Surveyor training should be efficiently conducted in a short period of time. In particular, as the recent development of information technology expands a computer assisted survey, the computer application ability of a surveyor determines the success of a survey. Therefore, a through management ranging from hiring to training surveyors leads to maximizing the ability of a surveyor.

In the process of a survey, in the case that a surveyor can not do a survey, due to accidents or private reasons, some reserved surveyors should be hired to replace the surveyor's job. For the purpose of managing surveyors, reports on business procedures on a daily, weekly, and monthly basis, can be submitted. A survey manager should figure out a variety of problems, in recognizing the weekly and monthly number of households contacted and the number of households survey-completed, to ensure to notify a survey expert of those problems to address problems of survey procedures. As for one-off survey, a short period of training is needed. However, as for annual or quarterly surveys, retraining courses and evaluation for surveyors should be done to improve survey procedures. Along with it, surveyors need to take some time off to get recharged.

Guidelines

- In setting up a survey system, average time it takes a surveyor for a survey, and average number by which a surveyor does a survey on a daily basis, should be considered to form a survey time-line.
- To help the overall understanding of a survey, a variety of education courses should be developed.
- The matters for a surveyor to need to know, including a survey purpose, survey objects, survey periods(the time for completing a survey), Privacy Policy for respondents, and the way of reviewing questionnaire surveyed, should be taught in the training courses.
- Reserved surveyors, excluding surveyors for a survey, should be secured.
- A surveyor should be refer to guidelines and handbook for respondents to help a survey well-done.
- In the process of a survey, as many as information about respondents should be obtained to find out survey problems and their solution.

※ As for the standard of a surveyor, temporary, regular surveyor and public officials in charge of a survey should be classified as a surveyor. Please, answer to the questions, considering the above.

3-12 Guidelines for surveyors

For the purpose of producing accurate statistics, the guidelines surveyors are able to use for a survey any time, are needed.

1) As there are guidelines a surveyor is able to use for a survey

2) As there are no guidelines a surveyor is able to use for a survey

Survey Guidelines

- Survey summary : survey purpose, legal basis, survey object, standard day for survey & survey period, population & survey object, sample size, survey items, survey methods(techniques), survey system & work division, survey and announcement schedule, etc.
- Definitions of major terms and classification standard
- Methods of producing a questionnaire
- Methods of editing a questionnaire(reviewing contents)
- Methods of managing a questionnaire: the matters a surveyor should be observe, ways of responding to non-acceptance and declination to answers
- Methods of managing a questionnaire
- Questions & answers
- Questionnaire

3-13 Methods of training surveyors

For the purpose of high quality survey data collecting, employing talented surveyors and thorough training courses for surveyors should be necessary. That is, a surveyor's attitude and understanding of a questionnaire, ability to encourage a respondent to participate in a survey should be fully taught in the training courses.

1) As there are no training courses for surveyors

2) As there are training courses for surveyors

3) As there are summon(group) training courses for surveyors

4) As there are on-line training cyber courses for surveyors

3-14 Notification for conducting a survey

This question checks whether conducting a survey is informed respondents of by mail, email(official letter), or telephone, prior to conducting a subject or not. Notifying respondents of the survey that will be conducted to encourage people to actively participate in the survey, helps them address a sense of declination and understand for the survey.

The following items should be checked in the assessment sheets.

1) As the survey contents that will be conducted, is notified respondents of in advance by an official letter, email, telephone, and broadcasting, etc.

2) As the survey contents that will conducted is not notified respondents of in advance

3-15 Figuring out the current status of surveyors and survey managers

To evaluate a survey burden of a surveyor, survey schedule, survey costs, you should check how many respondents a surveyor can interview on a daily basis. If too many surveys are alloted to a surveyor, the problem of reliability will be emerging. If too few surveys are alloted to a surveyor, there will be a waste of budget and time. This question checks how rates of survey managers who manage surveyors in survey field and conduct overall checking of questionnaire.
In the assessment, the input of the number of survey targets, survey periods, the number of surveyors and survey managers leads to average numbers of survey targets one surveyor surveys on a daily basis.

$$\text{※ Average numbers of survey targets one surveyor surveys on a daily basis} = \frac{\text{the number of survey targets}}{\text{survey periods(days)} \times \text{the number of surveyors(head)}}$$

Example) 20 survey managers account for 20% out of 100 surveyors

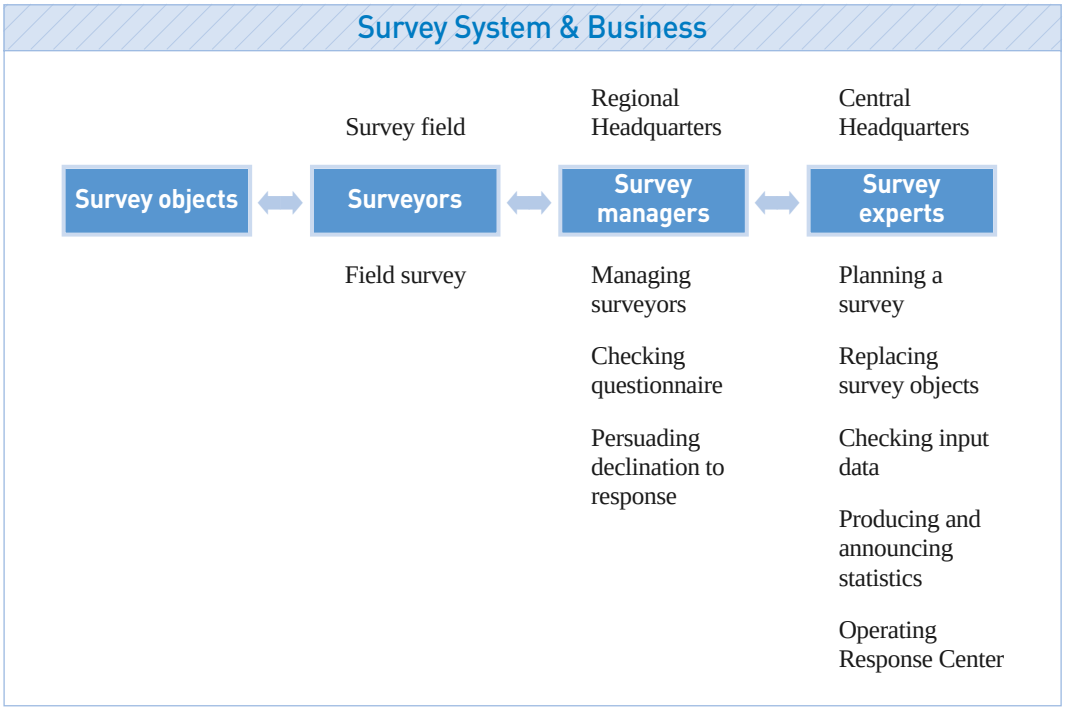
At the step of a field survey to collect data, a surveyor and survey manager visit a survey field and obtain the data about households and companies. In this process, a survey manager should support a surveyor to conduct a survey right. The manager also should send an official letter for survey cooperation, which contains information about a survey purpose, method, period, and size, to ensure to develop trust for a survey. At the same time, a survey manager should keep in close touch with a survey expert, in case for many possible accidents which might happen in a field survey. For example, in the case that man-made events such as a shutdown or strike and also natural disasters such as flood, a survey manager should have a consultation with a survey expert to come up with contingency measures such as replacing a survey object. On top of it, as for an industry survey, it is desirable to have a talk with a person in charge, regarding interview time, survey information, interview location, before visiting an industry. As for households survey, a surveyor should notify households of a visiting day and time to make sure the households are ready for the survey. If a surveyor can not obtain satisfactory responses from households or companies, he or she should inform of revisiting days to make sure to complete a survey. A surveyor should record a variety of information including the atmosphere of a survey field, the attitude of respondents to ensure to use the information as reference in the future.

Guidelines

- A survey manager should check many data needed in a survey field and also support a surveyor to use the data as reference in the field.
- A reporting system where a problem and its measure should be taken, should be set up.
- A surveyed table is checked by a survey manager, which should be sent to the central office, through collecting procedures.
- In a computer assisted survey, input errors should be checked by printing out input data, which ends up being sent to the central computer.

3-16 Countermeasures for problems in the process of a survey

This question checks what kinds of actions surveyors take to solve possible problems, which occurs in the process of a survey. The problems should be solved, based on the standard guidelines. In addition, a surveyor should have a thorough training to ensure he or she does not solve a problem arbitrarily.



3-17 Checking to respondents, as for answers unreasonable

In the process of reviewing contents(editing) for questionnaire surveyed, there might be logically unreasonable survey results, due to a surveyor's faults or inappropriate answers of respondents. For the results, this question check whether there is an additional checking process or not.

Despite the efforts of survey workers participating in the whole process of a survey, such non-response for the survey could occur. Non-responses affect a survey time-line and survey quality. Therefore, measures for survey non-responses should be come up with to keep and manage the quality of a survey stable. There might be a lot of causes for non-

response: no a survey object in the field, declination to response for all questions, no response for a specific question. In general, the occurrence of non-response has the size of a sample size smaller than planned in the beginning, which is a factor to increase dispersion of estimates. Accordingly, it is a primary concern to ensure that non-response does not possibly occur.

Guidelines

- A revisit plan should be set up to reduce the number of non-response.
- In order to encourage respondents to answer a questionnaire, a way might be considered to have respondents choose as many survey options as possible.
- ∫A revisit plan or re-call plan should be built, for unreasonable answers of respondents.

※ Comparison between non-response for a question and non-response for a unit

The type of non-response is divided into non-response for a unit where a respondent does not accept to answer the whole survey itself and non-response for a question where a respondent does not accept to answer some question in the survey. Non-response for a question has no any effective values, which can't be used as information.

3-18 Efforts to prevent non-response for a unit(double selection available)

In the process of a survey, as for the occurrence of non-response for a unit, a surveyor should revisit a respondent in order to reduce non-response. A revisit survey is to minimize the damage of information, due to non-response. This question checks whether a person in charge is doing for the matter or not.

The following items should be checked in the assessment sheets.

1) As there is no non-response

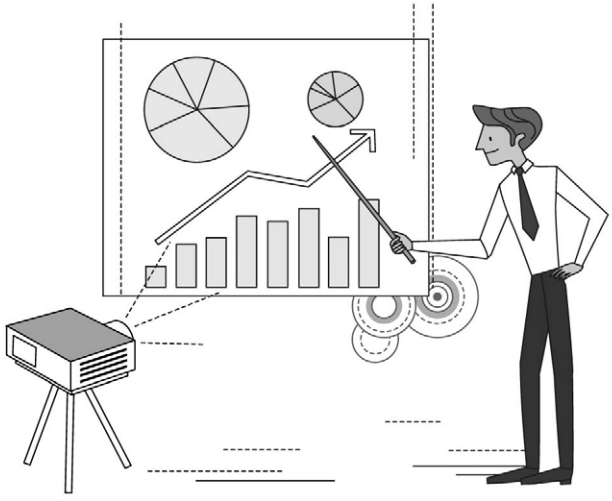
2) As a surveyor revisits once after a specific day passes(as for telephone or computer survey, one more re-call or e-mail)

3) As a surveyor revisits more than twice during the specific period(as for telephone or computer survey, more than twice of re-call or e-mail)

4) As a surveyor visits a big-sized industry or farmhouse, a big regions with high rates of non-response(as for telephone or computer survey, sending email, checking a call or re-mailing)

5) Re-visit survey system should be set up, considering the response pattern from a previous pattern

6) Others



Chapter 4

Data Input and Processing

Data input indicates inputting data by coding process, which a computer easily interprets for analyzing statistics of the data surveyed, after a survey is completed. Therefore, coding process and data input are closely related. Coding system should be divided into the case where there is no matter to answer and the another case where a respondent did not answer a question. That is, the former where a respondent did not belong to a question of a survey and did not answer and the latter where a respondent did not answer a question should be codified as "Not Applicable(N/A)" and Non-response" respectively. For example, as for a single number system, "9 and 0 can be codified as "Non-response" and "N/A" respectively.

Along with it, for a certain reason, a respondent might not answer a question of a survey. For example, "declination to response" or "do not know" are available. For this case, specific codes are necessary to tell the two cases.

As for closed-ended questions, survey data are input on the computer by coding process. However, as for open-ended questions, response data should be re-classified as a few items, which should be codified. In this case, the number of classification items should be properly adjusted and should possibly adopt the standardized system.

Guidelines

- A survey expert should review a data input system in advance to ensure to introduce a input system to minimize input errors with efficiency.
- Coding process of survey data should be conducted based on the coding guidelines. As for open-ended questions, classification codes should be provided together.
- The causes, frequencies, and solutions for data input errors should be recorded to ensure persons in charge of the data can see the data.

Guidelines of completing quality assessment sheets

4-1 Methods of data input(double selection available)

In order to conduct statistical analysis for collected data, it is necessary to input data. Today, a method of automatic data input is widespread due to the development of computer technology. However, when you can not use a computer in reality, a method of passive data input should be utilized. When the latter is adopted, input errors should be checked. A survey expert should set up programs such as CAPI, CATI, and CASI to make sure that input errors are checked.

The following items should be checked in the assessment sheets.

1) As a input worker enters survey data manually after survey completion

2) As data survey and data input are conducted simultaneously, as for computer-assisted surveys such as PDA, CAPI, CASI, and CATI

3) As survey data are input by using electronic works, for example, Magnetic Character Recognition(MCR), Intelligent Character Recognition(ICR), Optical Character Recognition(OCR), Voice Recognition Editing(VRE), and Electronic Data Exchange(EDI)

- Magnetic Character Recognition(MCR) : a method of inputting data by automatically recognizing characters on a questionnaire
- Intelligent Character Recognition(ICR) : a method of inputting data by automatically recognizing characters on a questionnaire through Intelligent Character Recognition

- Optical Character Recognition(OCR) : a method of electronically inputting data recorded on CRC cards
- Voice Recognition Editing(VRE): a method of inputting voice data
- Electronic Data Exchange(EDI): a method of inputting data by Electronic Exchange

4) Others

Reviewing contents(editing) is a procedure of reviewing and correcting data errors, and processing non-response. In the procedures of editing, errors should be corrected based on the following orders.

① As the whole level of unit questions are unreasonable, the data should be removed.

② As errors can be clearly corrected by other information or response values of other questions, the errors should be corrected. In general, the process is called deductive imputation.

③ As for error values can not be right corrected, it should go to non-response.

④ As it is judged that non-response can be imputed, the data should be corrected according to the guidelines to imputation.

Reviewing contents(editing) is classified based on characteristics of works, as follows.

• Format edit

It is a process of checking outward apparence of survey data saved in the file, where missing parts of survey questions should be check and proper input of codified questions should be reviewed.

• Structure edit

It checks whether a response value is proper or not, considering the characteristics of input data, and the effectiveness of survey objects and appropriateness of survey questions.

• Range checks

It tells whether there are data beyond the range of effective values of response for each survey question or not. In this case, it is needed to make a guideline to the range of effective values of response for each survey question. Even though response value is included in the

range of effective values, it might be outlier. Therefore, by narrowing the range, it is desirable to check data errors.

● Consistency checks

It checks consistency of response values for related questions. In a questionnaire, there might be filter questions or logically related questions. That is, a 12 year-old person should be checked in the section of "Marriage".

Guidelines

- ▶ As for non-response data or less consistent data, a surveyor should re-check the data by calling or re-visiting respondents.
- ▶ A guideline, a specific procedures of editing, should be prepared.
- ▶ The information about causes and corrections for errors found in the procedure of editing, should be recorded.
- ▶ In the repeated survey, as for the data errors that frequently occur, it is quite difficult to tell the contents in the process of editing. Therefore, it is effective to review with flexibility.

Guidelines for completing quality assessment sheets

4-2 Evaluating the need for reviewing contents(editing) micro-data

Micro-data should be reviewed for missing values, unreasonable data, data lack of consistency. For that end, the quality of the micro-data should be evaluated to check whether the data need editing or not.

The following items should be checked in the assessment sheets.

1) As editing is needed for a number of errors found in the process of editing after checking the micro-data

2) As some errors needs to be corrected after checking the micro-data

3) As editing is not necessary, despite some errors found in the micro-data, because the errors are no significant

4) As the data are fully checked in the process of collecting data by pre-checking system using a computer

5) As editing is not necessary, because there is no error in the micro-data

4-3 Methods of reviewing contents(editing)

The methods of reviewing contents(editing) to check data are as follows: automatic editing by computer, manual editing, and a mixed editing of the two methods. In order to check data consistency, it is desirable to possibly review data contents through automatic editing by computer.

The following items should be checked in the assessment sheets.

1) As automatic editing such as computer programs is done

2) As manual reviewing contents(editing) is done for individual questions with errors by a contents reviewer

3) As automatic editing is entirely done, but, for some questions, manual reviewing contents (editing) is done

4) As editing is not done

4-4 Wether the guidelines are provided for manual editing

As for the use of manual reviewing contents(editing) mentioned on 4-3 question and a mixed editing, stipulated guidelines should be prepared to prevent processed errors. As for automatic editing by computer, there is no need for stipulated guidelines because the logic of checking errors is already programmed in a computer.

The following items should be checked in the assessment sheets.

1) As the guidelines are not provided

2) As the guidelines are provided for some important questions, not the entire questions

3) As the guidelines are provided for the entire questions, in terms of contents, procedures, and methods necessary for reviewing contents(editing)

4-5 Various methods for reviewing contents(editing) (double selection available)

Editing collected data can be done by range check, logic check, statistical check, and comparative check, and so on. Based on the characteristics of the data, various methods for editing can be considered.

The following items should be checked in the assessment sheets.

1) As the range of a variable is checked (for example, the range of age from 0 to 100)

2) As the appropriateness of a question is decided by the logic relation between variables and also the logic flow between logically-dependent questions is checked (for example, if sex is male, answering the question of female diseases should be rechecked and then corrected.)

3) As outliers are corrected, after statistical analysis (for example, through income analysis, if income value shows 100 won, the value should be judged to be outlier and corrected.)

4) As data are corrected by comparing previous data or similar data of the same survey

5) As improper values for specific variables exist, based on past experience

6) others

4-6 Information about major causes for errors

The errors of survey data are divided into sampling errors found in sampling and non-sampling errors found in the process except sampling. In this case, Non-sampling errors are divided into response(measurement) errors, and input errors. So, by checking whether errors occur in the process of collecting data or of inputting data or not, the procedures of survey and data processing should be improved.

The following items should be checked in the assessment sheets.

1) As the cause for errors is not recognized

2) As the errors belong to response(measurement) errors

3) As the errors belong to input errors

4) As the errors belong to a mixed form of response(measurement) errors and input errors

5) Others

4-7 Methods of correcting the errors found in the process of reviewing contents(editing) (double selection available)

The errors found in the process of editing can be corrected by re-contacting respondents or surveyors. As for insignificant variables, the method of deleting the error variable can be considered. In addition, the error values can be replaced by statistical method or empirical imputation.

The following items should be checked in the assessment sheets.

1) As accurate values are obtained by examining a questionnaire with errors or respondents and calling a surveyor or respondent

2) As error value or missing value is replaced by statistical methods(mean imputation, random hot deck imputation, etc.)

3) As error value or missing value is corrected by empirical imputation or other similar response value

4) As you are not sure of the errors(as errors are not corrected)

5) As error value is deleted due to its less importance

6) Others

4-8 Measures to prevent excessive data correction

It is necessary to reviewing contents(editing) to check the errors of a questionnaire. However, it should be careful to excessively correct data errors because it can distort survey data. Therefore, the proper standards for editing should be set up. If necessary, absolute standard values for editing should be set up. It is desirable that the values beyond the standard should not be reviewing contents(editing).

The following items should be checked in the assessment sheets.

1) As optimal reviewing contents(editing) can be done, due to the computer program which can analyze the impact of reviewing contents

2) As data are corrected by errors or missing and then reviewing contents(editing) again

3) As there is no procedure to prevent excessive data correction

4) Others

Imputation means to impute non-response value with the similar value of a respondent in the process of reviewing contents(editing). The imputation needs through steps, so the help of experts are desirably necessary. In general, the tendency of a non-respondent who does not respond to a question and a respondent are different. Therefore, as for missing survey data, the value of a respondent should be not replaced from the general perspective of view. In addition, it is desirable that the imputation for non-response should be done in the central division, not the survey field.

After imputation of non-response, the value should be not treated as response value because imputation value is not different from the actual response value and also increase dispersion of estimated value by rising volatility. there are types of imputations as follows: mean imputation, random hot deck imputation, nearest neighborhood imputation, ratio imputation, and regression imputation, etc.

Guidelines

- Imputation of non-response should be processed in the central division, not the survey field.
- Imputation of non-response should consider the characteristics of data. Imputation method should be stipulated in a report.
- Depending on imputation method, it is effective to do the imputation of non-response by using as many information as possible.
- Imputation value should be indicated in some form in the data.
- Imputation value should be possibly minimized. The rate of imputation should be indicated.

$$\text{• Imputation rate for unit non-response} = \frac{\text{the number of survey objects applying imputation of non-response}}{\text{the number of survey objects}} \times 100$$

$$\text{• Imputation rate for non-response for i question} = \frac{\text{the number of survey objects applying imputation of non-response for i question}}{\text{the number of survey objects}} \times 100$$

$$\text{• Imputation rate for all the questions of non-response} = \frac{\text{the number of questions applying imputation of non-response}}{\text{the number of the entire questions (=the number of questions} \times \text{the number of units)}} \times 100$$

- Imputation computer programs or S/W is possibly used for a general purpose to ensure to be easy to use.

Guidelines for completing quality assessment sheets

4-9 Evaluating the need for editing micro-data

Data used in imputation of non-response for questions.
There are a lot of types of data used in the imputation of non-response. For example, there are the data holding the characteristics of non-respondents. In addition, the same value of respondents or outside data can be used for imputation. In this sense, this question is to check what data are used in the imputation of a question.

The following items should be checked in the assessment sheets.

1) As imputation is done by using the data holding the characteristics of respondents showing non-response for some parts of questions or supplementary information(logical and deductive information)
2) As imputation is done by using the data holding the characteristics of respondents showing non-response for some parts of questions, supplementary information, or predicted data for respondents for the variable
3) As imputation is done by using only predicted data for respondents for the variable, not other data
4) As non-response for a question is not processed
5) Others
6) As there is no value for non-response

4-10 Whether survey field imputation for unit non-response is allowed or not

Arbitrary imputation by surveyor may cause sampling bias. Therefore, as non-response for unit occurs in the process of a survey, survey field imputation should be checked. In case for non-response(missing value) for unit, a proper level of reserved samples should be prepared to ensure imputation is done and checked.

The following items should be checked in the assessment sheets.

1) As a surveyor has survey field imputation applied in an arbitrary way

2) As survey field imputation is used within the range of reserved samples given

3) As weight value for non-response is adjusted

4) As imputation is done by the division in charge(headquarters)

5) As there is no need for imputation, due to no case of unit non-response

6) As imputation is not allowed, even though there is unit non-response

4-11 Whether corrected values are marked or not

In general, corrected(or imputation) values have different characteristics from the micro-data. So, to separate it from the micro-data, it is desirable to mark the corrected values. In addition, The effect of imputation should be calculated by marking the values.

The following items should be checked in the assessment sheets.

1) For example, imputation data and micro-data should be indicated as 1 and 2 respectively. Like this, as the above data should be separately marked to tell their differences

2) As there is no mark of micro-data and imputation data.

4-12 Whether micro-data and imputation data are managed or not

Imputation data and micro-data should be kept at the same time. As the effect of imputation, the micro-data should be used. To check consistency of imputation data, the micro-data should be separately kept.

The following items should be checked in the assessment sheets.

1) As imputation data and the micro-data are separately kept

2) As imputation data are only kept, instead of the micro-data

4-13 Security of data completed to input or transferred

This question is to check how much security of survey data completed to input or transferred is done. Personal or business information should be managed not to be revealed outside.

The following items should be checked in the assessment sheets.

1) As information access is not allowed by coding process, for significant variables out of files input or transferred

2) As personal or business information such as name and birth date out of a questionnaire is not included in the data transferred

3) As there is no specific security

4) Others

Self Assessment of Statistics Quality

Chapter 5

Data Analysis and Quality Assessment

Relevance is the level where the demands of users are examined by the means of expert advisory meetings, or a survey for users satisfaction, which are reflected in statistics data. It shows, in the process of producing statistics, how the concepts(definition, classifications, etc) and the range of survey statistics used to clarify and achieve the purpose of a survey reflect the demands of users.

Guidelines

- A statistics producer collects information of statistics users regarding who users are, where they use it, and how important statistics are to them.
- In order to increase relevance, experts meetings, or a survey for users satisfaction should be conducted on a regular basis to increase relevancy.
- As for a number of demands of users, the demands should be given the order of priority, considering their importance.

Guidelines for completing quality assessment sheets

5-1 Method of evaluating a survey for users satisfaction (double selection available)

A survey for users satisfaction on statistics reflects demands of users and also measures to improve them. Conducting a regular survey of users satisfaction or figuring out demands of outside user groups, increases the relevance between statistics and users and ultimately contributes to the quality of statistics through expanding the use of statistics.

The following items should be checked in the assessment sheets.

1) As a survey of users satisfaction on statistics is conducted on a regular basis

2) As opinions are collected and improvements are returned, from statistics users, on a regular basis

3) As opinions are collected and improvements are returned, from statistics users, on a irregular basis

4) As a survey of users satisfaction and collecting users' opinions are not conducted

5) Others

5-2 Evaluation on sharing information about users

This question evaluates the level of sharing users' information, obtained by using a survey of users satisfaction.

The following items should be checked in the assessment sheets.

1) As there is some information about general users

2) As there is some information about major users

3) As there is some information about general users and general users

4) As there is a lot of information about major users, but a little information about general users

5) As the information about major and general users are thoroughly understood

5-3 Level of meeting general demands of users

This question checks whether statistics reflects demands of users or not. Opinions of users should be possibly and actively reflected.

The following items should be checked in the assessment sheets.

1) As no demands of users are reflected

2) As a few demands of users are met

3) As some demands of users are met

4)As demands of users are entirely met

5)As demands of users are actively met

Accuracy indicates how accurately the characteristics or size of a population for a survey are measured. That is, accuracy shows proximity between true value and estimated value. As errors between two values are smaller, statistical data are more accurate. The errors of survey statistics occur, due to coverage range, sampling, response and non-response, and procedures of producing statistics, etc. The accuracy of processed statistics can be measured by reviewing sampling errors, non-sampling errors and difference between provisional figures and final values. In general, a number of non-response impacts unexpected estimated results and also decreases the estimated results. Therefore, more efforts should be made to reduce the number of non-response because non-response is one of factors to deteriorate the quality of a statistics survey. In addition, CV or RSE, or Mean Square Error should be delivered to ensure users can evaluate the quality of statistics by themselves.

Guidelines

- ▶ CV, confidence interval, and MSE for the characteristics of statistics are delivered.
- ▶ Bias and volatility of major statistics values should be evaluated, compared with a previous survey.
- ▶ Sampling errors, measurement and processing errors, non-response, model specification errors, the impacts of errors should be described for users to take advantage of information.

Guidelines for completing quality assessment sheets

5-4 Values delivered for accuracy of estimated values

By delivering elements of measuring accuracy of statistics such as CV, relative standard errors, MSE, and confidence interval, users should be able to evaluate accuracy of statistics.

Example : Accuracy Index for major items of Household Trend Survey by Statistics Korea

12.주요 항목별 표본 오차 / 최근년 / 소득~실내장식품 - Windows Internet Explorer

http://www.kosis.kr/ups3/service/ch_header.jsp

KOSIS 온라인 간행물 / [가계동향조사연보]

12.주요 항목별 표본 오차 / 최근년 / 소득~실내장식품

○ 기준시점 [년:2008]
○ 수록기간 [분기:2003. 1/4-2008. 4/4, 년:2003-2008]

시점 조회 엑셀

	전국가구 표준오차	상대표준오차	근로자가구 표준오차	상대표준오차	근로자외가구 표준오차	상대표준오차	도시가구 표준오차	상대표준오차	근로자가구 표준오차	상대표준오차	근로자외가구 표준오차	상대표준오차
2008 (단위:원,%) 소득	23 845	0,708	29 611	0,780	28 840	1,031	27 970	0,801	33 539	0,861	33 926	1,164
경상소득	22 604	0,712	27 715	0,771	27 019	1,036	26 620	0,807	32 040	0,866	31 735	1,165
근로소득	22 245	1,032	26 921	0,821	17 605	2,710	25 970	1,141	31 076	0,918	21 258	3,046
가구주 근로소득	-	-	-	-	-	-	-	-	-	-	-	-
배우자 근로소득	-	-	-	-	-	-	-	-	-	-	-	-
기타가구원 근로소득	-	-	-	-	-	-	-	-	-	-	-	-
사업소득	11 006	1,669	4 706	4,429	20 848	1,475	12 551	1,871	5 279	4,852	24 061	1,636
재산소득	2 987	3,982	2 216	5,326	6 029	5,027	3 502	4,373	2 626	5,890	7 186	5,505
이전소득	4 956	1,787	4 406	2,626	9 308	2,191	5 389	1,987	4 985	3,019	9 927	2,351
비경상소득	6 090	3,105	8 990	4,181	8 657	4,550	6 284	3,264	7 692	4,000	10 215	5,299
가계지출	19 650	0,717	24 194	0,819	24 809	1,010	23 323	0,821	27 760	0,914	29 695	1,156
소비지출	14 379	0,628	17 304	0,713	18 706	0,891	16 971	0,715	19 721	0,791	21 913	0,996
식료품	2 865	0,490	3 408	0,553	3 837	0,706	3 330	0,548	3 932	0,619	4 369	0,769
곡류및식빵	274	0,601	340	0,770	430	0,906	292	0,632	379	0,851	441	0,907
육류	337	0,713	400	0,864	400	1,029	396	0,810	460	0,964	570	1,169
낙농품	208	0,875	271	1,045	273	1,304	236	0,961	306	1,146	303	1,410
어개류	277	0,784	330	0,991	431	1,135	309	0,862	375	1,101	481	1,250
채소해조류	227	0,540	283	0,695	329	0,750	259	0,598	323	0,771	372	0,822
과실류	270	0,726	326	0,884	379	1,012	313	0,815	372	0,972	443	1,145

5-5 Range of major coefficients of variation

As for delivering CV to evaluate the accuracy of estimated values, this question checks the level of CV, which depends on the scale and purpose of a survey. CV less than 5% is satisfactory.

The following items should be checked in the assessment sheets.

1) As CV for major items is high(more than 20%)

2) As CV for major items is medium(10% to less than 20%)

3) As CV for major items is low(5% to less than 10%)

4) As CV for major items is quite low(less than 5%)

5-6 Types of Weight values used for survey analysis

It is necessary to adopt and estimate a proper weight value in order to obtain an accurate estimated values from a sampling survey. This question is to examine which weight values are used in producing statistics.

The following items should be checked in the assessment sheets.

1) As a weight value is used(write down the type of the weight value)

2) As a weight value is not used

5-7 Whether a weight value is adjusted or not

Re-weighting is to adjust a original weight value by using a supplementary information, in order to reduce faults of a list of survey objects(sampling frame) or non-response errors, in calculating estimated values. For this, post-stratification, regression estimation, and ratio estimate can be used. This question is to check whether a weight value is adjusted or not and which method is used.

- Post-stratification : a method of adjusting a weight value to make sure composition ratio of population known is equivalent of composition ratio of a sample

- Regression estimation : a method of increasing the efficiency of estimated quantity by using supplementary variables of a population related to interest variables
- Ratio estimate: a method of increasing the efficiency of estimated quantity by using supplementary variables of a population related to interest variables

5-8 Types of errors adjusted by revision

The errors found in a sampling survey are sampling errors and non-sampling errors. To reduce the errors, revision using supplementary information is adopted.

The following items should be checked in the assessment sheets.

1) As revision using supplementary information, is adopted to reduce non-response errors
2) As revision using supplementary information, is adopted to reduce sampling errors
3) As revision using supplementary information, is adopted to reduce measurement errors of questionnaire or respondents
4) As revision is not adopted
5) As revision is adopted to reduce other types of errors

5-9 Non-response rate for a survey unit

This question evaluates the level of non-response rate. The rate of non-response depends on the characteristics and difficulty of a survey. In general, a survey conducted by a government organization has a tendency to show a low rate of non-response. In addition, one of ways to lower non-response rate is active promotion.

In the assessment, questions should be marked, considering the following standard.

(Table 5-1) Range of unit non-response

Low rate of non-response	5% to less than 20%
Middle rate of non-response	20% to 40%
High rate of non-response	more than 40%

$$\text{※ Non-response rate for unit} = \frac{\text{the number of survey objects of non-response}}{\text{the number of survey objects}} \times 100(\%)$$

5-10 Whether the types of unit non-response are fragmented or not

In the procedure of a survey, unit non-response must occurs. The fragmentation of unit non-response gives users a higher quality of statistics information. The types of non-response are classified: a object unqualified for a survey, not a survey object, declination to response, impossible to contact, and absence of a respondent, etc.

The following items should be checked in the assessment sheets.

1) As a survey object was first selected as a sample, but, at the time of the survey, the object is unqualified for the survey, due to business closure or change of business field(this is no a survey object)

2) Declination to response

3) Absence of a respondent(impossible to meet)

4) Impossible to have a normal contact, due to hospitalization, mental weakness, drunkenness

5) Others(write down the case)

5-11 Level of non-response rate in each question

Question non-response is non-response for survey questions. A respondent might not respond to Individual questions such as income, age, tax payment, and non-response. It is desirable to evaluate the rate of non-response for each question to secure and evaluate reliability of survey data. Question non-response is divided into response aversion where most questions are no answered and partial non-response where easy questions are answered and sensitive or difficult questions are not answered. As for response aversion, it is desirable to set the level of a permissible range of non-response which can be used as effective information before data processing.

$$\text{※ Non-response rate for i question} = \frac{\text{the number of survey objects of non-response for i question}}{\text{the number of survey objects}} \times 100(\%)$$

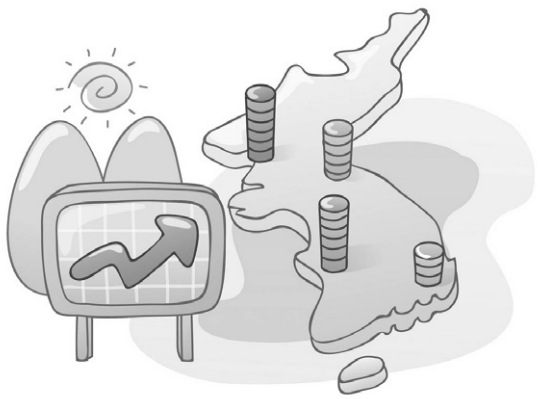
In this case, the highest rate of non-response for each question should be examined and be marked.

Timeliness means to produce and announce statistics during a specific period of time. As a period of surveying objects(standard time of producing) and a period of announcing survey results near, the statistics shows a high level of timeliness. Users want to timely and efficiently use statistics information. So, timeliness is one of most important demands for users.

Timeliness of statistics or information is a concept related to statistics' reflection to reality, which indicates the difference between(a period of surveying objects and a period of announcing survey results). Punctuality is about observing a pre-announcement time-line. A system to pre-announce a time-line for statistics, is implemented to ensure users can notice the announcement time-line in advance. As a survey observes the pre-announcement time-line, it shows a high level of punctuality.

Guidelines

- Planned announcement time-line should be observed. If there are some gaps, its cause and improvement measure should be delivered.
- The difference between a period of surveying objects and a period of announcing survey results should be evaluated.



Guidelines for completing quality assessment sheets

5-12 Meeting the deadline to publish a report

It is necessary to check the difference between an expected day when a statistics report(army) is published and an actual day for publication. In this case, the difference of one month is regarded as 30 days.

Example : Official announcement of statistics publication by Statistics Korea

5-13 Difference between a time of surveying objects and a period of announcing final results

A time of surveying objects and the day of announcing final results should be described in a statistics report or publication to ensure users can employ statistics results with a high level of timeliness.

The following items should be checked in the assessment sheets.
(It is based on the following table of a period of producing statistics)

	annual statistics	semi-annual statistics	quarter statistics	monthly statistics
1) considerable difference ⬅	more than 24 months	more than 12 months	more than 180 days	more than 60 days
2) a slight long difference ⬅	18 to 24 months	9 to 12 months	136 to 180 days	46 to 60 days
3) normal ⬅	12 to 18 months	6 to 9 months	91 to 135 days	31 to 45 days
4) a slight little difference ⬅	9 to 12 months	4 to 6 months	60 to 90 days	20 to 30 days
5) quite little difference ⬅	less than 9 months	less than 4 months	less than 60 days	less than 20 days

Example : Difference between a time of surveying objects and the day of announcing results

▶ A period or a point of time of surveying objects

- Standard time of a survey : as of December 31, 2007
- Period of surveying object : January 1, 2007 to December 31, 2007(for one year)

▶ The standard time of the survey of mining · manufacturing as of 2007 is December 31, 2007. A estimated announcement of its survey results is made on October 31, 2008(Statistics/notice board/news data/etc.) A difference between the survey point and its announcement indicates about 10 months. Therefore, timeliness of the survey of mining · manufacturing is categorized into "a slight little difference" of the table above.

5-14 Measures of protecting private information of individual objects in providing statistics information

In announcing statistics data in public, the information for individuals or specific businesses should be not revealed, which requires a special attention. For that, a thorough standard for announcing statistics data should be prepared to ensure private information is not revealed. If protection objects(sensitive information) is revealed, a measure against it will have to be come up with.

Comparability is about the possibility to compare data, in terms of a identical concept, classification, measurement tool and procedures, and basic data, and so on, even though the data shows the gap of a geographical field. Comparability is to see whether it is possible to compare data with a different country, city, and year or not. It is necessary to apply the international standard classification and evaluation methods to increase the international comparability.
Comparability is largely divided into three. First, geographical comparability is to have populations with different geographical objects for similar surveys measuring identical phenomena. Comparability for time is about comparing the difference between two points of time of surveying objects. Field comparability is about comparing two surveys with a similar purpose of statistics in a statistical field.

Guidelines

- Statistics for international comparison should use the international standards.
- As for time-series, as the period of surveying objects is cut off, the difference between concepts and methods before and after measurement should be described.
- To evaluate comparability of other statistics, concepts, definitions, population composition, sampling methods, classification standards, and methods of producing statistics should be described.

5-15 Comparability between similar statistics in terms of statistics design

This question is to evaluate comparability between similar statistics.

The following items should be checked in the assessment sheets.

1) As it is impossible to compare statistics, due to a big gap of definitions, classification and methods of producing
2) As there are many limitations to compare statistics, due to a gap of definitions, classification and methods of producing
3) As it is possible to compare statistics after describing limitations, despite some gaps of definitions, classification and methods of producing
4) As it is possible to compare statistics, despite a little gap of definitions, classification and methods of producing
5) As it is possible to compare statistics, despite few gap of definitions, classification and methods of producing
6) As there are no statistics to compare

Coherence indicates how similar statistics data, produced by different basic data or methods of producing, in terms of economic and social phenomena, are. For example, estimated data and final data, annual data and quarter(monthly) data might be produced by different data sources and methods of producing. If these data show similar results, the statistics data are judged to have a high level of coherence. Coherence is different from comparability and compares actual statistics values.

Guidelines

- Estimated and final values for major values should be compared and the specific difference should be described.
- Coherence should compare and review a survey purpose, production methods, and standards and also describe matters that require attention.

Guidelines for completing quality assessment sheets

5-16 Coherence evaluation of similar statistics

As there are similar elements such as a survey purpose, and classification standards between two statistical data, the data can be checked in terms of coherence. Coherence evaluates simple similarity of statistical values or trend values.

The following items should be checked in the assessment sheets.

1) As there are big differences between all statistical indexes

2) As there are big differences between most statistical indexes

3) As there are big differences between some statistical indexes

4) As there are a few differences between most statistical indexes

5) As there are a few differences between all statistical indexes

6) As it is impossible to analyze coherence, due to no statistical values to compare

Survey Statistics

Chapter 6

Documentation and Data Supply

Documentation means to record all statistical activities to produce statistics in a statistics production agency. Its purpose is to deliver the entire procedures of a survey and its results to have users understand easily in the most efficient way. Therefore, documents of survey data needed for production and documents providing various information and related techniques, needed for survey procedures should be divided.

In this chapter, we focus on the documentation for statistics production. Documents for statistics production include statistics publications or reports. Statistics publications cover all entire procedures ranging from a survey to results announcement. That is, all the contents such as a survey purpose, a survey period, a survey agency, sampling design, processing non-response, estimation, results announcement, and quality management, are included in the documents. Reports covers all specific description of major items of each survey procedure. For example, there is a report on sampling design, a reports on quality management, and a reports on processing non-response, etc. Thus, documentation means to supply documents about the entire contents of a survey to ensure users can easily figure out statistics information.

Guidelines

▶ In the process of documentation, the following information should be included.

survey purpose, population, sampling design, variables, survey method and technique, method of statistics production, changes in survey design, demands of users, etc.

Guidelines for completing quality assessment sheets

6-1 Methods of announcing statistics(double selection available)

Statistics data can be supplied with the following types of reports, Internet publication, supplying CD-ROM, and public data, etc. In this case, the question is to check what type of supplying data are used. A general system needs to be managed for the purpose of users' easy access to data and data use, and improving statistical application, etc.

6-2 Data supply through a statistical publication and web-site

When statistics are approved with a statistical publication, the title of a publication should be described. However, in reality, it is rare to keep the rule. Data supply on the web-site requires special attention to ensure users are not difficult to find data necessary for them.
As for a statistical publication, its title should be described for its related question. As for data supply on the web-site, Internet address should be provided for its related question.

6-3 Data supply to international organizations(OECD, UN, IMF, WHO)

It is checked whether statistical data such as a statistical index or table, or values are offered to international organizations or not, which should be marked. As for data supply to an international organization, the data should be briefly explained. When data are supplied to many international organizations, the user of organizations and data contents have be described.

6-4 Manuals(guidelines) used for statistics production

This question is to check whether, in the process of producing statistics, a statistics production agency has manuals(guidelines) such as a survey manual, data input manual, editing manual, business manual(business handbook, work handbook). manuals(guidelines) needed for each step of production should be described by a production agency. Additional contents should be immediately corrected.

6-5 Matters to be improved for a higher quality of statistics (double selection available)

This question is to check whether there are some matters to be improved to raise quality of statistics. It covers all procedures of producing statistics including data collection, surveyor's education and training, questionnaire, survey design, data processing and analysis, data management, and data announcement, etc.

The following items should be checked in the assessment sheets.

1) As there is no matter to be improved
2) As there is need to make up for methods of collecting data
3) As there is need to make up for matters regarding data input and processing
4) As there is need to make up for survey methodology or estimation methods
5) As there is need to make up for expertise of workers in charge of managing a survey
6) As there is need to make up for matters regarding change or improvement of software used for data processing
7) As there is need to make up for matters regarding statistics publications
8) As there is need to make up for methods of announcing statistics
9) As there is need to make up for methods of managing statistical data
10) Others

According to Article 30 of the Statistics Law, a statistics production agency should not provide statistical data in a way to tell a specific individual, corporate, and group. However, a statistics production agency can provide data in a way to tell a specific individual, when the agency use the data as the sample of a sampling survey. Therefore, a data provider should consider this point to ensure the demands of users are reflected at a maximum level, based on their purpose of using data. A statistics production agency should set up a policy of offering an appropriate data about statistics use to ensure it provides consistent data. In particular, it requires special attention, due to its sensitivity, for a strategy to revise errors of micro-data and non-response or provide micro-data where individuals can be distinguished.

Guidelines

- A set of policies are needed to give data suitable for users by providing various types of data and to give users easy access to data.
- Statistical data providers should provide various programs to ensure users can have easy search for statistical information.
- In offering micro-data, data providers should deliver specific layouts of data to ensure users can use the data without difficulty.

Guidelines for completing quality assessment sheets

6-6 Which types of statistical results are supplied to users(double selection available)

Data should be provided to users to ensure users can use the data for education and other purposes. In this case, the type of data should be marked in a related-question.

6-7 Whether data are registered in KOSIS and e-country index or not

It should be identified whether statistical data are registered in KOSIS or e-country of Statistics Korea or not. When data are registered, the number of e-country indices for the data should be marked in a related-question.

6-8 Meta DB contents for data registered in KOSIS

When statistical data are registered in KOSIS of Statistics Korea, meta DB(explanation data for statistics) should be written to provide detailed statistical information to users. For that, the contents of meta DB should be marked

1) Title of statistics
2) Type of statistics (methods of producing statistics, field, kind)
3) Legal basis for producing statistics)
4) Purpose of producing statistics(survey purpose)
5) Survey period
6) Survey objects
7) Survey method
8) Survey contents(survey questions)
9) Survey system
10) Application classification(classification standards and system, etc.)
11) Survey period(object period/survey standard time, survey period, etc.)
12) Survey history(change of contents after approval)
13) Methods of collecting data
14) Methods of processing data(input, aggregation, internal check, error, correction, etc.)
15) Methods of announcing statistical results(announcement methods and period)
16) Announcement coverage
17) Announcement period(news, publications, data announcement period)
18) Title of publications
19) Data search(news, Kosis, digital publication, etc.)
20) Glossary for major terms
21) Use notice
22) Contact information of a person in charge
23) Others

6-9 Whether changes of contents are recorded in statistical publications or not

When information regarding standard revision, changes in methods of producing statistics or surveying, and provisional or final values are recorded(announced), specific explanations and limitations should be recorded to help users understand data. In particular, it is desirable to analyze and announce the difference of results before or after revision or change of methods of producing.

The following items should be checked in the assessment sheets.

1) As a detailed explanation on changes for statistical procedures, standard revision, or announcement of estimated values and use notice are clearly described

2) As there is no explanation on provisional figures or estimates

Example : User Notice

For Users

1. The report is on the data of classifying and aggregating individual construction companies according to 『Korean Standard Industrial Classification』 The data might not be equivalent of data from associations of comprehensive or professional agencies, equipment, facilities, and information communication business.

2. Starting surveys as of 1999, a survey unit is changes from business fields registered to company units. In addition, the 8th revised 『Korean Standard Industrial Classification』 is applied. So, time series of data might not be connected to reports made before 1998.

3. As for heating construction companies, there are many number of the companies. But, the budget of construction is not sizable. So, considering these points, the survey that was conducted every year, is conducted in the year of ending 0 and 5 starting 2006.

   - In 2006, a survey for heating construction companies was not conducted. Therefore, the time series of data in 2006 is inconsistent with the data surveyed before 2005.

6-10 Whether there is a statistical table regarding questions of recognizing sex or not

This question is to check whether there are questions of recognizing sex in the stage of producing statistics or statistical tables or not. It also examines whether questions of recognizing sex are applied in the survey questions for a more detailed comparison or not, in terms of sex, or whether survey data are reflected in statistical results or not.

Example: Results of analyzing questions of recognizing sex

(unit: %, head)

		Total(N)	It should be done	It had better be done	If a respondent does not want, he or she does not have to do	X^2 value
Korea	male student	1,074	25.8(277)	48.5(521)	25.7(276)	253.241**
	female student	1,055	8.3(88)	33.9(358)	57.7(609)	
China	male student	928	33.6(312)	30.3(281)	36.1(335)	85.127**
	female student	1,055	21.3(225)	22.0(232)	56.7(598)	
Japan	male student	948	15.2(144)	44.3(420)	40.5(384)	30.585**
	female student	977	9.4(92)	38.6(377)	52.0(508)	

**p < 0.01

Chapter 7

Post-management

A person in charge in a statistics production agency should possess expertise for his or her own field. A frequent movement to divisions is likely to undermine continuance and professionalism of statistics-related work. In this sense, officials of the central government and local autonomous organizations are quite lack of expertise. Statistics production is a specialized field, different from general administrative affairs. Therefore, persons in charge of statistics should take a supplementary course on a regular basis. Along with it, seminars or workshops on new analytic theories or survey methodology should be held to improve expertise on statistics. If possibly, hiring prominent statistics personnel is taken into consideration.

Guidelines

- Persons in charge of statistical affairs should have an opportunity to take a supplementary course and have access to new analytic theories or survey methodology on a regular basis.
- The head or manager of a statistics production agency should improve working environment to have statistical workers participate in various education programs.

Guidelines for completing quality assessment sheets

7-1 Check in every step of producing statistics

This question is to check all procedures of statistics a statistics production agency is producing, ranging from design to survey, data processing, analysis, and data supply. Bo doing so, errors which occur at a changing environment of producing statistics can be prevented. In addition, a chance to apply new technique can be created.

7-2 Whether a field survey or post-check is conducted or not

A statistics production agency should express its thanks to respondents for responding to a statistical survey after the survey is completed. In order to check whether there are a respondent a surveyor did not visit, to prevent a surveyor from entering wrong information, and to check the accuracy of a survey, a statistics production agency should re-visit or re-call a survey respondent.

7-3 A list of information submitted by a contractor after survey completion

A statistics production agency, which has an agency produce statistics, should establish a measure to receive and manage a list of survey information such as tables surveyed, input data, and results analyzed, from its contractor. Items submitted should be marked for a related question.

7-4 Requests for statistical improvement and measures implemented

This question is to check whether there are requests for improvement through the National Assembly, the Board of Audit and Inspection, and the media in the last three years or not, and to check which measures are implemented, if requests are taken. By taking an active response to requests for improvement, the quality of statistics can be boosted.

7-5 Current status of using information resources related to statistics production

This question is to check the current status of applying information resources to statistics

production by a statistics production agency. If it is difficult to answer this question, ask for help to a person in charge of information of an agency. For this is important for developing a system of a general-purpose survey, kindly ask for your active cooperation.

Entry Example

■ **System name**

- Statistics production :
  all systems required to produce statistics included. Not only many survey systems, but also Meta-data management system, population management system, and electronic survey system included.
  (Example) Industrial statistics analysis system, Population survey system, Company population system

- Statistics management :
  a system to manage statistical micro-data and aggregated data
  (Example) Statistics micro-data management system, Statistics DW system, National statistics DB management system

- Statistics service :
  a system to supply statistical micro-data and aggregated data
  (Example) Statistical micro-data service system, KOSIS(Korean Statistical Information System), ECOS(Economic Statistics System)

■ **Operation equipment**

- Entering major specifications and DBMS of server equipment used at each step of statistical work
  (Example) Major specifications : IBM SERVER, 5.6Ghz × 32. 256GB
  DRMS : Oracle, My SQL, DB2, etc.

■ **Statistics analysis package**

- Entering a package used to analyze statistics
  (Example) SERVER type(or PC type) SAS, SPSS, R, STATA, etc

Self Assessment of Statistics Quality