

Chapter 4

Utilization of Machine Readable Data on UN's General Industrial Statistics

Hidekiyo Sakamoto

Introduction

Relational databases allow data to easily be manipulated and displayed in tables without having to prepare complex programs. However, statistical databases are one exception to this rule. In many cases, various footnotes indicating exceptions are appended to statistics data. These footnotes have varying structures, making them difficult to process uniformly. This is because traditional database management systems on the market are designed mainly for processing paperwork and other general duties, and have been developed based on the premise of a data structure with a simple format known as a relation. As a result, these systems are not equipped with data models capable of handling changes in exceptions over time, etc. [2] The "General Industrial Statistics" compiled by the United Nations contains footnotes indicating various exceptions which change with the passage of time. This makes it extremely difficult to process this data.

In the "Industrial Statistics", three-digit International Standard Industrial Classifications (ISIC) are used as the basic classification for industrial sectors. For some countries, however, some data cannot be used with these basic industrial sector classifications. In these cases, data is indicated by unifying multiple three-digit sector classifications. These sector unifications change over time, and there are some cases where sectors unified at one point are noted separately again at a later date. When making time-series and international comparisons using statistics, comparing data with disparate sector

classifications is meaningless and can lead to false conclusions. This chapter investigates changes in the unification and division of industrial sectors used in the "Industrial Statistics" over time.

Section 1 presents an overview of UN industrial statistics publications and describes computer readable data which corresponds to these publications. Section 2 uses Hong Kong and Singapore as examples to examine whether precautions must be taken for changes in the unification of industrial sector classifications over time when using "Industrial Statistics" computer readable data. In addition, a table illustrating changes in the unification and division of industrial sector classifications over time is created using the data for 10 Asian and Pacific countries out of the countries contained in the "Industrial Statistics" in preparation for actually using this data. The method of interpreting this table is described in Section 3.

1. Machine Readable Data

1.1 Industrial Statistics

The 1991 edition of the "Industrial Statistics Yearbook Volume I, General Industrial Statistics" (hereafter called "Industrial Statistics") compiled by the United Nations marked the 25th edition of these statistics, and recorded basic data concerning the industrial activities and industrial structures (Table 1) by industrial sector (Table 2) for the approximately 150 member states.

The "Industrial Statistics" arranges industrial

activity and industrial structure items along the top and industrial sectors along the side of the tables for each country. The years for which data is recorded vary according to the industrial activity and industrial structure item. The 1991 edition contains industrial production indices for the 13 year period from 1979 to 1991, and other data for the 5 year period from 1987 to 1991.

Statistics generally deals with exceptions regarding classification coverage and other items by attaching footnotes. The "Industrial Statistics" also attaches two types of footnotes for coverage of industrial activity and industrial structure items and industrial sectors. The first type of footnote indicates the unification of industrial sector classifications, and is shown by enclosing data in brackets within the statistics tables. Table 3 shows the data for Hong Kong, and provides an example of industrial sectors, ISIC371 and 372 enclosed by brackets with [3.9 3.5 3.2 3.2] indicated for "AVERAGE NUMBER OF PERSONS ENGAGED" item near the center of the table.

The second type of footnote indicates exceptions concerning industrial activities and structures as well as industrial sector classifications. The following footnotes are used for these exceptions.

- (1) Includes vacations and absences due to illness.
- (2) Excludes transport equipment.
- (3) 50 employees or more
- (4) Aggregate of usable data only
- (5) Includes automobile repairs.
- (6) Industrial sector zzz is included in industrial sectors xxx to yyy.

When industrial sector classifications are unified, if the essence of the classification changes to the point that brackets cannot be used, footnotes indicating exceptions such as (6) above are attached in place of brackets. The footnote in Table 3 notes that "a/Major group 352 is included in 351." This indicates that ISIC 352: "Other chemical products" is

unified with ISIC 351: "Industrial chemicals".

1.2 Industrial Statistics Computer Readable Data

The United Nations provides data in magnetic tape form (hereafter, computer readable data is also referred to as "Industrial Statistics") to allow "Industrial Statistics" to be processed by computer. "Industrial Statistics" files are comprised of a data file, footnote 1 file and footnote 2 file.

1.2.1 Data File

The data file is the file containing the main elements of industrial statistics. For data without footnotes, only the data file exists. For data with annotations, relevant information in code form can be obtained from the footnote 1 file. Further, the individual contents of the coded relevant information in the footnote 1 file can be obtained from the footnote 2 file. The data file is comprised of the following items.

- (1) Country code
- (2) Industrial activity and industrial structure code
- (3) Industrial sector classification code
- (4) Year (last two digits)
- (5) Numerical values

Countries are listed only by code in the data file. However, the corresponding country can be understood by referencing Table 4 "Country codes and names". The names corresponding to industrial activity and industrial structure codes can be obtained from Table 5. Regarding numerical values, the units for numerical values must correspond to the industrial activity and industrial structure codes. Table 6 shows units for numerical values which correspond to (2) industrial activity and industrial structure codes. Table 7 lists the industrial sector classification codes described in "1.2.2 Industrial Sector Combination Codes".

An example of data file records is shown below.

344	06	3211	82	130000000
344	06	3211	83	172000000
344	06	3211	84	185000000
344	06	3211	85	215000000
344	06	3211	86	266000000
344	06	3211	87	314000000

From Table 4, the country code 344 is Hong Kong; from Table 5, the industrial activity and industrial structure code 06 is wage compensation; and from Table 6, the units for the numerical value are understood to be the nation's unit of currency, in this case Hong Kong dollars. According to Table 7, the industrial sector classification code 3211 indicates the spinning and weaving sectors.

As a result, this sample data file indicates that the wage compensation for Hong Kong's spinning and weaving sectors from 1982 to 1987 was HK \$ 130 million, HK \$ 172 million, ... , and HK \$ 314 million.

Numerical values can be obtained for each classification category in the "Industrial Statistics" from the data file. However, explanatory information cannot be obtained from only the data file. The existence of footnotes and their contents can only be known when the footnote 1 and footnote 2 files described below are present.

1.2.2 Industrial Sector Combination Codes

A Table of Industrial Sector Combination Codes which lists code combinations is provided with "Industrial Statistics" computer readable data. These codes indicate the groups containing industrial sectors when industrial sectors can only be used as groups. Table 7 shows the Table of Industrial Sector Combination Codes. An example of these codes is shown below.

0000	2	4	
2000	2		
2000A	2	290	
2000B	2	220	
210A	210	220	
210AB	210	220	353

210AC	210	220	353	354
210B	210	230		
210C	210	220	290	
210D	210	220	230	
2100	210			

There is one exception where industrial sector code 0000 corresponds to sectors from the mining and mining, quarrying sectors (industrial sector code: 2) to the electricity and gas sectors (industrial sector code: 4). Industrial sector code 210A is shown to consolidate the coal mining industry (industrial sector code: 210) and petroleum and gas sectors (industrial sector code: 220) into a single group.

The industrial sector classification codes used in this combination table are used in the data and footnote files.

1.2.3 Footnote 1 File

The footnote 1 file specifies data in the data file which has annotations. The footnote 1 file is also used to match the footnote contents in the footnote 2 file with the data in the data file. The footnote 1 file items are as follows.

- (1) Country code
- (2) Industrial activity and industrial structure code
- (3) Industrial sector classification code
- (4) Year (last two digits)
- (5) Footnote code

Footnote 1 file items (1) to (4) are the same as the respective data file items. For these items, footnote 1 file data which matches with the data in the data file has annotations. An example of footnote 1 file data is shown below.

344	06	3211	84	15
344	06	3211	85	15
344	06	3211	86	15
344	06	3211	87	15

This example indicates that the footnote with the footnote code 15 specified in the footnote 2 file is appended to supplements to wages and salary (industrial activity and structure item code 06) for the spinning and weaving sectors (industrial sector

classification code 3211) in Hong Kong (country code 344) for years 1984 to 1987.

The contents of footnote code 15 cannot be understood with only the footnote 1 file. However, these contents can be obtained from the footnote 2 file.

1.2.4 Footnote 2 File

The footnote 2 file contains a record of footnote contents by country, and the contents of footnote codes specified by the footnote 1 file can be obtained from the footnote 2 file. The items contained in the footnote 2 file are as follows.

- (1) Country code
- (2) Footnote code
- (3) Footnote

Footnote 2 file items (1) and (2) are the same as the respective footnote 1 file items. The contents of footnotes for Hong Kong are listed in Table 8.

As previously mentioned, the footnote 1 file shows that the footnote indicated by footnote code 15 for Hong Kong is appended to supplements to wages and salary (06) for the spinning and weaving sectors (3211) in Hong Kong (344) for years 1984 to 1987. The contents of footnote code 15 can be obtained from the footnote 2 file. According to Table 8, footnote code 15 is listed as

344 Data are incomplete due to confidentiality values.

The relationship between the data file, footnote 1 file and footnote 2 file is shown in Figure. 1

2. Problems When Using Data

Footnotes used with statistics come in various types and formats. As a result, it is difficult to execute statistical data processing as standard processing. Among the various types of footnote data, the processing for footnotes concerning the unification

and division of industrial sector classifications is described below.

When processing "Industrial Statistics", it is important to understand changes in the unification and division of industrial sector classifications over time. Failure to take these changes into consideration will result in the comparison of data with different parameters. In many cases, the data for individual industrial sectors cannot be obtained over a given period and only aggregated values can be used. The true relationship between these data must be understood before actually using the data. This section describes this process using Hong Kong and Singapore as examples.

For Hong Kong, the relationship between industrial sector classifications within a single economic activity and structure item is examined. In addition, the relationship between different industrial activities and structures is also examined. Next, the data for Singapore and Hong Kong are combined to investigate differences between these countries. Then, in Section 3, changes in the unification and division of industrial sector classifications over time are shown for 10 Asian and Pacific countries out of the countries recorded in the "Industrial Statistics" in order to facilitate the actual use of "Industrial Statistics".

2.1 Changes in the Unification and Division of Industrial Sectors Over Time for a Single Industrial Activity and Structure in a Single Country

Data concerning the number of establishment-/enterprises in Hong Kong is shown in Table 5. According to this data, industrial chemicals (351) and other chemical products (352) are consolidated into a single sector from 1976. Accordingly, the data for the individual industrial sectors cannot be used for these years, and care must be taken when processing the data before and after 1976 together. In addition,

pottery, china, etc (361), glass and glass products (362) and nonmetal products, etc. (369) were unified from 1976 to 1982, but the data from each industrial sector was available again from 1983. As shown by this example, there are also cases where industrial sectors are unified and then later divided again allowing the data for each industrial sector to be used. Similarly to glass and glass products, iron and steel (371) and nonferrous metals (372) were also unified and then later divided again.

In this manner, since there is not just a single pattern for changes in industrial sector classifications, data must be carefully examined before commencing data processing.

2.2 Disparities in the Unification and Division of Industrial Sectors Between Multiple Industrial Activities and Structures for a Single Country

Multiple items are sometimes necessary when using data. For example, when obtaining the output per establishment/enterprise, both establishment/enterprise data and output data are necessary. In these cases, if the industrial sector classifications for the various items differ, the data cannot be processed.

Fig. 3 shows the relationship between the unification and division of industrial sectors for output in Hong Kong. In this figure, industrial chemicals (351) and other chemical products (352) are unified from 1981 to 1983 and from 1989 to 1990, and differ from the data for the number of factories. Accordingly, regarding the number of factories and gross production, data for industrial chemicals (351) and other chemical products (352) can be used together only for years 1981 to 1983 and 1989 to 1990. Care should also be taken for pottery, china, etc (361) and glass and glass products (362), etc. as the data for these items also differs from the data for the number of factories.

2.3 Comparing Multiple Countries

International comparisons are difficult to make when the definitions for the same items differ. In the "Industrial Statistics", industrial sector classifications are unified and divided differently even for the same industrial activities and structures depending on the country. Disparities in industrial sector classifications for the number of establishments/enterprises in Hong Kong and Singapore are shown below.

Fig. 2 shows the number of establishments/enterprises in Hong Kong, and Fig. 4 shows the number of establishments/enterprises in Singapore. For Singapore, pottery, china, etc (361) and glass and glass products (362) are unified from 1975. On the other hand, pottery, china, etc (361), glass and glass products (362) and nonmetal products, etc. (369) are unified from 1976 for Hong Kong. In this case, both the unified sectors and the point in time of unification differ. In addition, steel (371) and nonferrous metals (372) are unified from 1970 to 1982 for Hong Kong. However, these sectors are not unified for Singapore from 1970. Various other cases are also possible depending on the countries being compared, necessitating that due caution be taken before processing these data.

3. Changes in the Unification and Division of Industrial Sector Classifications Over Time for Asian and Pacific Countries

This section takes 10 Asian and Pacific countries: China, Hong Kong, Indonesia, Japan, Korea, Malaysia, the Philippines, Singapore, Thailand and U.S. from among the countries contained in the "Industrial Statistics" and shows the changes in the unification and division of industrial sectors over time by industrial activity and structure item for these countries. Note that industrial sectors which are not unified are not shown.

3.1 How to Interpret the Table

Using China as an example, industrial sectors were unified for the following seven industrial activities and structures.

- (1) Number of companies
- (2) Average number of employees
- (3) Wages and salaries of employees
- (4) Quantity of electricity consumed
- (5) Output
- (6) Value added
- (7) Gross fixed capital formation (total)

Looking at the number of companies, while ISIC 311, 313 and 314 were recorded separately from 1981 to 1991, these three industrial sector classifications were recorded as a single sector from 1977 to 1980. ISIC 353 and 354, and ISIC 371 and 372 are also recorded in a similar manner. When the division and unification of industrial sectors occurs consecutively in this manner, the link is indicated with a solid line.

Examining the average number of employees shows that while ISIC 351, 352, 353 and 354 are recorded as individual industrial sectors from 1985 to 1986, these sectors are recorded in a unified manner from 1977 to 1980. In this case, since unification and division are not performed consecutively, the link is shown with a dotted line to make the table easier to interpret. ISIC 361B is the unified sector for 361, 362

and 369. In this table, however, only 361B appears and the individual sectors 361, 362 and 369 are not shown. This indicates that individual sectors are not used for the entire period, or in other words that only the unified sector is used.

Conclusion

When processing computer readable statistics data, exceptions such as footnotes and other explanatory data cannot be ignored. In particular, since economic statistics data is often compared over time, changes in exceptions over time must be correctly understood. Exceptions indicating the unification and division of industrial sector classifications and other items also appear in the "Industrial Statistics". It is clear that improperly processing the unification and division of these classifications will lead to faulty conclusions. In addition, comparing different classifications is meaningless. Therefore, used classifications must be checked to determine how they change over time before starting to process data. Classification patterns vary not only for each country, but for items within a single country as well.

This chapter illustrated the changes in the unification and division of industrial sectors over time by industrial structure and activity item for 10 Asian and Pacific countries.