Cirlin 7

# 1972 CENSUS ESTABLISHMENT SIZE DISTRIBUTION DATA

**ABSTRACT:** This data set contains the 1972 Census statistics by establishment employment size classes for the 2- and 4-digit Census manufacturing industries. Data is also included for "all industries" under the 1-digit code "0". The data is often aggregated to avoid revealing confidential firm information.

**JCL:**
```
//GO.FT08f001 DD DSN=ESTSIZ72,DISP=OLD,UNIT=TAPE6250,
// DCB=(RECFM=FB,LRECL=90,BLKSIZE=~~3420~~),3240
// LABEL=(~~07~~,SL,,IN),VOL=SER=~~DATABK~~
          43                    DTA063
```

**DATA SET ORGANIZATION:** Each record is uniquely associated with a particular size class in a particular industry. Records are sorted according to the following major to minor sequence:

1. Census SIC
2. Size class code

The sort is in ascending sequence, with the following qualification: The 2-digit codes precede the 4-digit codes. The data set begins with an industry with SIC=0 that represents all manufacturing industries.

Unlike the Census size distribution data sets for 1963 and 1967, this data set does not have any SIC codes in the 1900's.

For each industry, there are 12 records--one record for each of ten size classes, one record for establishments covered by administrative records, and one record for industry totals.

The size class codes are as follows:

| Size class | Code |
|---|---|
| Industry totals | 0 |
| 1-4 | 1 |
| 5-9 | 2 |
| 10-19 | 3 |
| 20-49 | 4 |
| 50-99 | 5 |
| 100-249 | 6 |
| 250-499 | 7 |
| 500-999 | 8 |
| 1,000-2,499 | 9 |
| 2,500 and over | 10 |
| establishments | 11 |

covered by administrative records

**# OF RECORDS:** With 471 industries and 12 records per industry, the total number of records is 5652.

**FORMAT:** (I4,I2,I1,F7.0,F7.1,F8.1,2F7.1,4F8.1,F7.1,F8.1)

**VARIABLES:**

| | | |
|---|---|---|
| 1. | Census industrial code (SIC) | |
| 2. | Size class code | |
| 3. | Aggregation code (see note 1 below) | |
| 4. | Number of establishments | |
| 5. | Number of employees | (thousands) |
| 6. | Payroll | ($ millions) |
| 7. | Number of production workers | (thousands) |
| 8. | Man-hours | (millions) |
| 9. | Wages | ($ millions) |
| 10. | Value added | " " |
| 11. | Cost of materials | " " |
| 12. | Value of shipments | " " |
| 13. | Capital expenditures | " " |
| 14. | End-of-year inventories | " " |

**SOURCE:** U.S. Bureau of the Census. Census of Manufactures: vol. 1, parts 1 and 2, 1972. Summary and Subject Statistics, Washington, D.C.: Government Printing Office, 1976.

**NOTES:** 1. Aggregation--As noted earlier, some size classes are aggregated with others to prevent disclosure of confidential firm information. In the Census report, such aggregated data is recorded in the column for the smallest size class of the group and is indicated by underscoring. The subsegent size classes in the grouping are coded by the letter "D", indicated by a zero in the data set. Our corresponding aggregation codes are:

1. All values of the variables 5 through 9 of this size class are aggregates of the data of this and one or more subsequent size classes. (Indicated in the original source by a column of five underscored figures.)

3. The data for variables 5 through 9 of this size class is included i in a previous aggregation. (Indicated in the original source by a

column of five D's.)

2,4. Analogous to codes 1 and 3 for variable 9 (capital expenditures) only. (Indicated in the report by an underscoring and a subsequent D for capital expenditures.)

5. All values in variables 5 through 9 of this size class are aggregates of the data of this and one or more subseqent size classes. In addition, the data for variable 9 (capital expenditures) is included in a previous aggregation. (Indicated in the report by a column of four underscored figures, and a D for capital expenditures.)

6. All values for size class code 11 are suppressed. The data are included in figures for the two smallest size classes.

If both codes 3 and 4 are valid, code 3 is recorded.

2. The size distribution data was summed and checked against the industry totals. A number of errors were found. The Bureau of the Census was able to correct most of these errors including all the major ones.

3. Report forms were not sent to companies that operated only one establishment and which Social Security payrolls indicated had fewer than 10 employees. Actual payrolls (and sales) for 1967 were obtained from administrative records of the Federal Government. The other statistics from these establishments were estimated from industry averages.

4. The Z code (value of variable is greater than zero but rounds to zero at the unit presented) is represented by a -1. Be careful when converting the -1.0's representing the Z-code. There are negative values for some of the variables (e.g., value added).

5. The 1972 data sets contain all the variables that the 1963 data sets contain, plus more, but in different units of measurement. Variables numbered 5, 6, 10, 12, and 13 do not have the same units in both 1963 and 1972. The 1972 and 1967 have the same variables in the same units, but in a different format.

6. The data is often aggregated to avoid revealing confidential firm information. We have developed a technique for disaggregating the aggregated data--the technique is described in separate documentation. The JCL for the data set containing our estimates of the true values of the aggregated data is given below:

```
//GO.FT08f001 DD DSN=SIZ72MOD,DISP=OLD,UNIT=TAPE6250,
// DCB=(RECFM=FB,LRECL=90,BLKSIZE=7290),
// LABEL=(08,SL,,IN),VOL=SER=DATABK
```

7. Unlike the 1963 and 1967 size distribution data sets, there are no SIC's in the 1900's. And in reading the SIC's, note that the 2-digit SIC's are recorded with four digits, e.g., SIC "21" is given as "2100".

circle 8

# 1972 CENSUS ESTABLISHMENT SIZE DISTRIBUTION DATA, DISAGGREGATED

**ABSTRACT:** This data set contains the 1972 Census statistics by establishment employment size classes for the 2- and 4-digit Census manufacturing industries, but differs from the original data in one important respect. In the original data (available in another data set), the data is often aggregated to avoid revealing confidential firm information. We have developed a technique for disaggregating the aggregated data--the technique is described in separate documentation--and this data set contains our estimates of the true values of the aggregated variables.

**JCL:**
```
//GO.FT08F001 DD DSN=SIZ72MOD,DISP=OLD,UNIT=TAPE6250,
// DCB=(RECFM=FB,LRECL=90,BLKSIZE=~~3240~~)7290
// LABEL=(~~08~~,SL,,IN),VOL=SER=~~DATADK~~
          44                    DTA063
```

**DATA SET ORGANIZATION:** Each record is uniquely associated with a particular size class in a particular industry. Records are sorted according to the following major to minor sequence:

1. Census SIC
2. Size class code

The sort is in ascending sequence, with the following qualification: The 2-digit codes precede the 4-digit codes. The data set begins with an industry with SIC=0 that represents all manufacturing industries.

Unlike in the 1963 and 1967 Census size distribution data sets, no SIC's in the 1900's are included in the 1972 Census data sets.

For each industry, there are 11 records--one record for each of 10 size classes, plus an additional record for the industry totals. The size class codes are as follows:

| Size class | Code |
|---|---|
| Industry totals | 0 |
| 1-4 | 1 |
| 5-9 | 2 |
| 10-19 | 3 |
| 20-49 | 4 |
| 50-99 | 5 |
| 100-249 | 6 |
| 250-499 | 7 |
| 500-999 | 8 |

| | |
|---|---|
| 1,000-2,499 | 9 |
| 2,500 and over | 10 |

**# OF RECORDS:** 11 records per industry, 471 industries, and hence 5181 total records

**FORMAT:** (I4,I2,I1,F7.0,F7.1,F8.1,2F7.1,4F8.1,F7.1,f8.1)

**VARIABLES:**

| | | |
|---|---|---|
| 1. | Census industrial code (SIC) | |
| 2. | Size class code | |
| 3. | Aggregation code | |
| 4. | Number of establishments | |
| 5. | Number of employees | (thousands) |
| 6. | Payroll | ($ millions) |
| 7. | Number of production workers | (thousands) |
| 8. | Man-hours | (millions) |
| 9. | Wages | ($ millions) |
| 10. | Value added | " " |
| 11. | Cost of materials | " " |
| 12. | Value of shipments | " " |
| 13. | Capital expenditures | " " |
| 14. | End-of-year inventories | " " |

**SOURCE:** U.S. Bureau of the Census, Census of Manufactures: Vol. 1, Parts 1 and 2, 1972. Summary and Subject Statistics. U.S. Government Printing Office, Washington, D.C., 1976.

A copy of the original data is available on magnetic tape; the JCL is given below:

```
//GO.FT08f001 DD DSN=ESTSIZ72,DISP=OLD,UNIT=TAPE6250,
// DCB=(RECFM=FB,LRECL=90,BLKSIZE=3240),
// LABEL=(07,SL,,IN),VOL=SER=DATABK
```

**NOTES:** 1. Aggregation--As noted earlier, all aggregated data has been disaggregated. This data set retains the aggregation coding used in the original data set in order to indicate where the disaggregation took place.

1. All values of the variables 5 through 9 of this size class are aggregates of the data of this and one or more subsequent size classes. (Indicated in the original source by a column of five underscored figures.)

3. The data for variables 5 through 9 of this size class is included i in a previous aggregation. (Indicated in the original source by a column of five D's.)

2,4. Analogous to codes 1 and 3 for variable 9 (capital expenditures) only. (Indicated in the report by an underscoring and a subsequent D for capital expenditures.)

5. All values in variables 5 through 9 of this size class are aggregates of the data of this and one or more subseqent size classes. In addition, the data for variable 9 (capital expenditures) is included in a previous aggregation. (Indicated in the report by a column of four underscored figures, and a D for capital expenditures.)

If both codes 3 and 4 are valid, code 3 is recorded.

2. For some size classes there are no data observations. Variables 3 through 9 in the corresponding records contain zeroes. Missing values are recorded as zeroes.

3. The original size distribution data was summed and checked against the industry totals. A number of errors were found. The Bureau of the Census was able to correct most of these errors including all the major ones.

4. Report forms were not sent to companies that operated only one establishment and which Social Security payrolls indicated had fewer than 10 employees. Actual payrolls (and sales) for 1972 were obtained from administrative records of the Federal Government. The other statistics from these establishments were estimated from industry averages.

5. The Z code (value of variable is greater than zero but rounds to zero at the unit presented) is represented by a -1. Included in this category are variables that have been made too small (less than 0.05) as a result of the disaggregation procedure. There are other variables whose value is sometimes less than zero, e.g., value added. Do not confuse these values with the -1.0 that indicates the Z code.

6. There are two differences between the

organization of these data and the original 1972 data: first, data that were included in the original 1972 Census data for "establishments covered by administrative records" have been omitted in this data set; and seccnd, the format of the data set has been changed.

7. The 1972 data sets contain all the variables that the 1963 data sets contain, plus more, but in different units of measurement. Variables numbered 5, 6, 10, 12, and 13 do not have the same units in both 1963 and 1972. The 1967 and 1972 data sets have the same variables and units, but different formats.

8. Persons desiring to the exact procedure used to disaggregate the aggregated variables should see the document: "The Census Size Distribution Disaggregation Procedure for Undisclosed Data."