

Repetitive Distribution

When data are collected, the information obtained from each member of a population or sample is recorded in the sequence in which it becomes available. This sequence of data recording is random and unranked. Such data, before they are grouped or ranked, are called **raw data**.

Definition

Raw Data Data recorded in the sequence in which they are collected and before they are processed or ranked are called *raw data*.

Suppose we collect information on the ages (in years) of 50 students selected from a university. The data values, in the order they are collected, are recorded in Table(1) below. For instance, the first student's age is 21, the second student's age is 19 (second number in the first row), and so forth. The data in Table below are quantitative raw data.

21	19	24	25	29	34	26	27	37	33
18	20	19	22	19	19	25	22	25	23
25	19	31	19	23	18	23	19	23	26
22	28	21	20	22	22	21	20	19	21
25	23	18	37	27	23	21	25	21	24

Suppose we ask the same 50 students about their student status. The responses of the students are recorded in Table (2) below. In this table, F, SO, J, and SE are the abbreviations for freshman, sophomore, junior, and senior, respectively. This is an example of qualitative (or categorical) raw data.

J	F	SO	SE	J	J	SE	J	J	J
F	F	J	F	F	F	SE	SO	SE	J
J	F	SE	SO	SO	F	J	F	SE	SE
SO	SE	J	SO	SO	J	J	SO	F	SO
SE	SE	F	SE	J	SO	F	J	SO	SO

The data presented in Tables (1) and (2) are also called **ungrouped data**. An ungrouped data set contains information on each member of a sample or population individually

Organizing and Graphing Qualitative Data

This section discusses how to organize and display qualitative (or categorical) data. Data sets are organized into tables, and data are displayed using graphs.

1- Frequency Distributions

A sample of 100 students enrolled at a university were asked what they intended to do after graduation. Forty-four said they wanted to work for private companies/businesses, 16 said they wanted to work for the federal government, 23 wanted to work for state or local

governments, and 17 intended to start their own businesses. Table(3) lists the types of employment and the number of students who intend to engage in each type of employment. In this table, the variable is the *type of employment*, which is a qualitative variable. The categories (representing the type of employment) listed in the first column are mutually exclusive. In other words, each of the 100 students belongs to one and only one of these categories. The number of students who belong to a certain category is called the *frequency* of that category. A **frequency / Repetitive distribution** exhibits how the frequencies are distributed over various categories. Table (3) is called a *frequency distribution table* or simply a *frequency table*.

Variable →	Type of Employment	Number of Students	← Frequency column
	Private companies/businesses	44	
Category →	Federal government	16	← Frequency
	State/local government	23	
	Own business	17	
		Sum = 100	

Definition

Frequency Distribution for Qualitative Data A *frequency distribution* for qualitative data lists all categories and the number of elements that belong to each of the categories.

A sample of 30 employees from large companies was selected, and these employees were asked how stressful their jobs were. The responses of these employees are recorded below, where *very* represents very stressful, *somewhat* means somewhat stressful, and *none* stands for not stressful at all.

somewhat	none	somewhat	very	very	none
very	somewhat	somewhat	very	somewhat	somewhat
very	somewhat	none	very	none	somewhat
somewhat	very	somewhat	somewhat	very	none
somewhat	very	very	somewhat	none	somewhat

Construct a frequency distribution table for these data.

Solution Note that the variable in this example is how stressful is an employee’s job. This variable is classified into three categories: very stressful, somewhat stressful, and not stressful at all. We record these categories in the first column of Table(4). Then we read each employee’s response from the given data and mark a tally, denoted by

the symbol /, in the second column of Table 2.4 next to the corresponding category. For example, the first employee's response is that his or her job is somewhat stressful. We show this in the frequency table by marking a tally in the second column next to the category somewhat. Note that the tallies are marked in blocks of five for counting convenience. Finally, we record the total of the tallies for each category in the third column of the table. This column is called the column of frequencies and is usually denoted by f . The sum of the entries in the frequency column gives the sample size or total frequency.

In Table 2.4, this total is 30, which is the sample size.

Table 2.4 Frequency Distribution of Stress on Job

Stress on Job	Tally	Frequency (f)
Very		10
Somewhat		14
None		6
		Sum = 30

▪ Relative Frequency and Percentage Distributions

The relative frequency of a category is obtained by dividing the frequency of that category by the sum of all frequencies. Thus, the relative frequency shows what fractional part or proportion of the total frequency belongs to the corresponding category. A relative frequency distribution lists the relative frequencies for all categories.

Calculating Relative Frequency of a Category

$$\text{Relative frequency of a category} = \frac{\text{Frequency of that category}}{\text{Sum of all frequencies}}$$

The **percentage** for a category is obtained by multiplying the relative frequency of that category by 100. A *percentage distribution* lists the percentages for all categories.

Calculating Percentage

$$\text{Percentage} = (\text{Relative frequency}) \cdot 100$$

EXAMPLE 2

Determine the relative frequency and percentage distributions for the data of Table (4).

Solution The relative frequencies and percentages from Table (4) are calculated and listed in Table (5). Based on this table, we can state that .333, or 33.3%, of the employees said that their jobs are very stressful. By adding the percentages for the first two categories, we can state that 80% of the employees said that their jobs are very or somewhat stressful. The other numbers in Table 2.5 can be interpreted the same way.

Notice that the sum of the relative frequencies is always 1.00 (or approximately 1.00 if the relative frequencies are rounded), and the

sum of the percentages is always 100 (or approximately 100 if the percentages are rounded).

Table 2.5 Relative Frequency and Percentage Distributions of Stress on Job

Stress on Job	Relative Frequency	Percentage
Very	$10/30 = .333$	$.333(100) = 33.3$
Somewhat	$14/30 = .467$	$.467(100) = 46.7$
None	$6/30 = .200$	$.200(100) = 20.0$
	Sum = 1.000	Sum = 100