

4

WRITING A DESCRIPTIVE SUMMARY

The purpose of this chapter is to explain how to proceed in order to write a good descriptive report, and how to analyze a frequency table beyond a first-level reading of the percentages, in order to identify the numerical features of the data and to highlight them.

After studying this chapter, the student should know:

- how to proceed when writing a descriptive report to summarize data;
- which measures and charts are appropriate, depending on the measurement level of the variable;
- how to summarize a set of variables that measure a given concept;
- how to analyze a frequency table in detail and identify its important features;
- the difference between a first-level description and an analytical description;
- the criteria for a good descriptive summary.

In Chapter 3, we have seen how to produce simple descriptive statistical measures, as well as simple tables and graphs. We have also seen that the statistical measures to be used depend on the level of measurement of the variable. Now, we would like to see how we can integrate all these elements and produce a synthetic report that describes certain features of a population. For the time being, we will restrict these explanations to univariate descriptions of variables. Later on, you will have to include bivariate descriptions, that is, descriptions of the statistical associations between variables, as well as confidence statements, that is, generalizations from the observed sample to the population as a whole, two statistical topics studied later on in this book. We will also learn how to report the result of a hypothesis testing.

How to Write a Descriptive Report

We will consider two types of report. Basic reports consist in a direct reading of the tables produced by SPSS, and a reformulation in direct, plain language of what the tables say, with accompanying charts as illustrations. There is very little interpretation

in this case. A second level in sophistication consists in writing analytical reports: such reports would highlight the outstanding tendencies that can be seen in the data, and may include a greater degree of interpretation. We will now explore both kinds of reports.

Basic, Direct Reports

Suppose you want to describe the educational level of the individuals included in the **GSS93 subset** data file supplied with the SPSS package. This means that you would like to have some global description that tells you whether the people in your sample tend to have a high level of education or not (this is a description of the central tendency), and whether there is a big polarization, with some people having a lot of education and many others very little (this is a description of the dispersion).

The first thing to do is to see which variables concern education. You will find three such variables in the **GSS93 subset** data file. List them, and list the level of measurement of each.

In this data file, you will find that the three variables are:

- Highest year of schooling completed (scale),
- Highest degree obtained (ordinal, 5 categories), and
- Possession or not of a college degree (ordinal, 2 categories).

Determine what kind of descriptive measures you would use for each. Would you use a frequency table? For which of the variables? Which charts would be more appropriate?

Sometimes you will feel that you are not too sure which type of chart is appropriate. Get SPSS to produce several charts, examine them carefully to see which ones convey a better representation of the distribution of the variable, then select one of them, and paste it into your report.

One of the important pitfalls that you should avoid is to give a lot of tables or charts that are not very useful. You may want to be selective here: select the relevant information, and try to write it in a clear and concise way. For example, SPSS produces tables giving you the number of valid answers. You do not need to include the table itself. You could simply write in brackets ($n = 1500$) when describing the sample, to indicate that your sample contains 1500 individuals. Whenever you discuss or describe the results that relate to one of the variables, if you see that there are a lot of missing answers, add a phrase about the number of valid answers, such as (valid $n = \dots$) and fill in the number of valid answers. Although the number of people in the sample is the same throughout the analysis of this data file ($n = 1500$), the number of valid answers varies a lot. This is why

you have to specify how many *valid* answers you have to a particular question. You do not have to do that for every single question: you report the number of valid answers only when there is a lot of missing data, and the valid percentages differ by several points from the total percentages. It is advisable in this case to report the valid percentages. In some cases it may be relevant to report both the valid and total percentages.

What follows is a set of criteria that define a good descriptive report.

Criteria for a Good Report

THE GENERAL PRESENTATION

Make sure the text is clear, well organized, and concise. If the analysis is long, a cover page may be desirable. Make sure that all the relevant information is in it: a title, your name, the name of the course and the course number, the name of the instructor to which you are presenting it, and the date.

Some of this information, such as your name and the assignment number, could be written in the header of your document (refer to Lab 2 for explanations on the header). The tables and graphs must be printed with the correct identification: a title must be given to every table or graph. If you copy the tables from SPSS with the **Copy...** command (rather than the **Copy Object...** command), you can edit the table, and delete the rows or columns that are not useful or relevant. Also avoid grammatical mistakes: a spell check may be useful, but rely always on a careful reading of your report.

Include in your report a **description of the data file** you are using: its source, the year the survey was conducted, the kind of variables that are found in it, the institution under which it was conducted, etc.

DESCRIPTION OF THE VARIABLES UNDER STUDY

Make sure to include in your study all the variables that are relevant for your subject. If there are several variables that address a given topic, use them all to analyze this topic. For instance, 'education' can be measured in several ways. If there are several variables that deal with education, examine the distribution of each.

To describe a variable properly, you must select the appropriate measures. Do not compute the mean of a qualitative variable, because it is meaningless. You may want to use some of the recoded variables, or recode some variables yourself. Do not include a table of frequencies if the variable is quantitative. Such tables are usually quite long, and they are not useful to the reader. If the quantitative variable has been grouped into a small number of categories, a frequency table may be useful, in addition to the descriptive measures used for quantitative variables. Finally, formulate your conclusions in full, grammatically correct sentences that highlight the meaning of your numerical results. An example of a very concise description of the educational level of the people in our sample is given in Insert 4.1.

The appropriate measures to be used are summarized in Table 4.1.

Table 4.1 **Appropriate descriptive measures for the various levels of measurement**

Level of Measurement	Appropriate Statistical Measures	Appropriate Charts
Nominal (categories)	Frequencies, percentages, mode. Ratios, proportions and rates.	Bar charts, pie charts
Ordinal	Frequencies; mode; median. Cumulative frequencies. (If there are many categories, you may compute the mean and median, but the interpretation of the numerical results may be problematic.)	Bar charts; histograms
Numerical scale, ungrouped	Mean, median, mode, range, minimum, maximum standard deviation, interquartile range. (Frequency tables are not useful for this type of measure.)	Histograms, frequency polygons, box plots, time lines
Numerical scale, grouped	Frequency tables, mode. If there are a large number of groups: mean and standard deviation. The mean is usually the mean code of the categories. It can be used for comparative purposes if other samples are grouped in the same way, but it should not be mistaken for the mean of the variable itself. If grouped into a small number of categories, it should be treated like ordinal data.	Histograms, bar charts, pie chart. Box plots may be misleading if the number of categories is small.

Examples of Concise Descriptive Reports

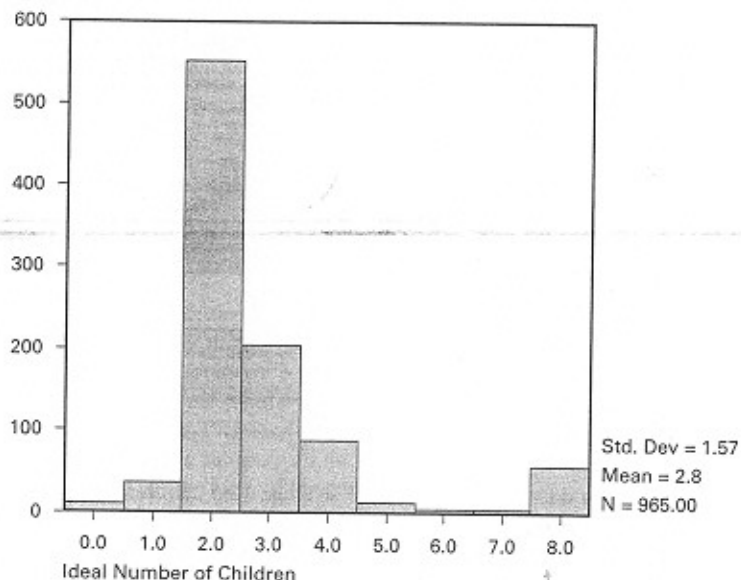
What follows (Insert 4.1) is an example of a short descriptive report, which answers the question: Describe the educational level of the sample given in the file **GSS93 subset** that comes with the SPSS program.

INSERT 4.1 Descriptive report of the educational level of the sample

The data set used here is a subset of the General Social Survey conducted in the US in 1993 ($n = 1500$). There are three variables in this data set that address the issue of education: the highest year of schooling completed (scale), the highest degree obtained (ordinal, 5 categories) and the possession or not of a college degree (ordinal, 2 categories).

The average highest year of schooling completed is 13 years with a standard deviation close to 3 years. The graph below shows the distribution of this variable.

If we compare that situation with the Ideal number of children, we see that the mean for that variable is 2.76 children, but the comparison with the actual number of children is difficult to make, as there are 535 missing answers for that variable (we can assume that only those who had children were asked that question). It is better to examine the histogram of the ideal number of children. Here we see that the mode, or most desirable situation, is by far the situation with two children. Very few people think that one child is the ideal situation.



6. Spanking Children

We have answers for 66% of the respondents, and the rest of the answers are missing. Of those who answered, about three-quarters (73.3%) indicated they either agree or strongly agree with spanking children as a disciplinary measure, while the rest (26.7%) disagree or strongly disagree.

7. Number of Siblings

We see here that the average is 3.7 brothers and/or sisters. If we examine the cumulative frequencies, we see that 60.2% of the respondents come from families of 4 children or less (the respondent plus 3 brothers or sisters), the rest (almost 40%) coming from families with 5 children or more. Comparing that with the number of children people currently have, we see that in general, individuals come from families that are larger than the families they themselves establish, since the average number of children in this sample tends to be much smaller than the number of brothers or sisters respondents have.

Analytical Descriptive Reports

The examples shown above are quite direct, and consist essentially in reporting, almost as is, the information provided in the frequency tables. But a more analytical view would permit a richer reading of such tables. To illustrate what is meant by that we will go into a more detailed – and more analytical – reading of frequency tables.

EXAMPLES OF HOW TO ANALYZE A FREQUENCY TABLE

To make our point clear, we are going to analyze four cases of the same situation, represented by the tables below. They all deal with the frequencies of the variable Political Party Affiliation, taken from the GSS93 subset file. The first table is the one that we get from the actual data in this file. The other three have been modified to illustrate how the analysis can highlight the distribution pattern.

Table 4.2 Political Party Affiliation A

	Frequency	Percent	Valid Percent
Strong Democrat	213	14.2	14.3
Not Str Democrat	298	19.9	20.0
Ind, Near Democrat	180	12.0	12.1
Independent	187	12.5	12.5
Ind, Near Republican	148	9.9	9.9
Not Str Republican	280	18.7	18.8
Strong Republican	168	11.2	11.3
Other Party	17	1.1	1.1
Total valid	1491	99.4	100.0
NA	9	.6	
Total	1500	100.0	

Case A Analysis of Case A (Table 4.2). We see from the table that those who are affiliated with the Democrats (strongly or not strongly) add up to 34.3%, or slightly more than a third. Those who are affiliated with the Republicans add up to 30.1%, or slightly less than a third. The independents add up to 34.5, again a little more than a third. It is interesting to note that the population is almost evenly divided into three groups, and that those who affiliate to neither party are as numerous (or a little more numerous) than those who affiliate with either of the two main parties. We can also notice that, within each of the two main parties, those who do not have a strong affiliation with the party are more numerous than those who have a strong affiliation (for the Republicans: 280:168, or about 7:4, and for the Democrats, 298:213, or about 3:2). The bar chart shown in Figure 4.1 illustrates this situation.

Case B Analysis of Case B (Table 4.3). We see from the table that those who affiliate with the Democrats add up to 42.1%. Those who are affiliated with the Republicans add up to 39.1%, or slightly less than the Democrats. The independents add up only to 17.6%, indicating that there is a strong polarization between the two

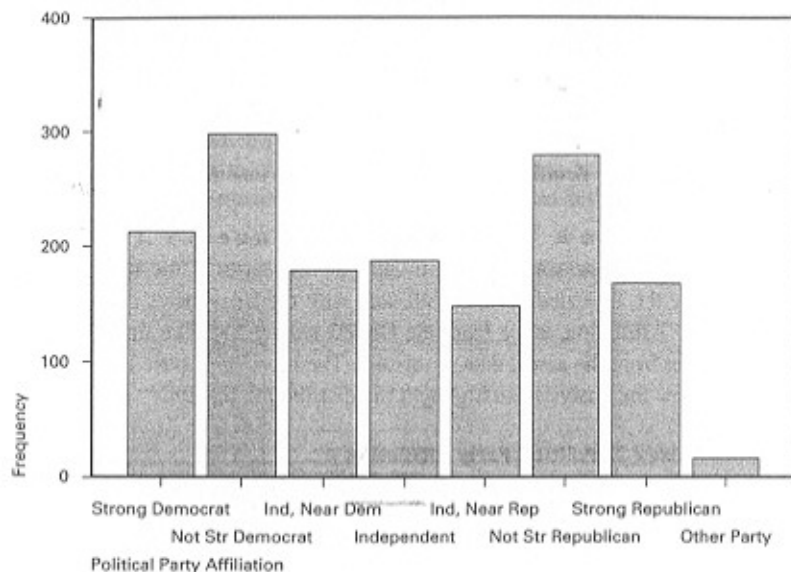


Figure 4.1 Political Party Affiliation

Table 4.3 Political Party Affiliation B

	Frequency	Percent	Valid Percent
Strong Democrat	272	18.2	18.2
Not Str Democrat	356	23.7	23.9
Ind, Near Democrat	122	8.1	8.2
Independent	57	3.8	3.8
Ind, Near Republican	84	5.6	5.6
Not Str Republican	351	23.4	23.5
Strong Republican	232	15.5	15.6
Other Party	17	1.1	1.1
Total valid	1491	99.4	100.0
NA	9	.6	
Total	1500	100.0	

parties, with less than 1 person out of 5 not affiliated to one of these two parties. We can also notice that, within a party, those who are not strongly affiliated with the party are more numerous than those who are (for the Republicans 23.4% vs. 15.5%, or a ratio of about 3:2, and for the Democrats 23.7% vs. 18.1%, or a ratio of about 4:3). The bar chart in Figure 4.2 illustrates this situation, and the polarization between the two parties is clearly visible.

Case C Analysis of case C (Table 4.4). We see from the table that those who are affiliated with the Democrats add up to 35.6%, or slightly more than a third. Those

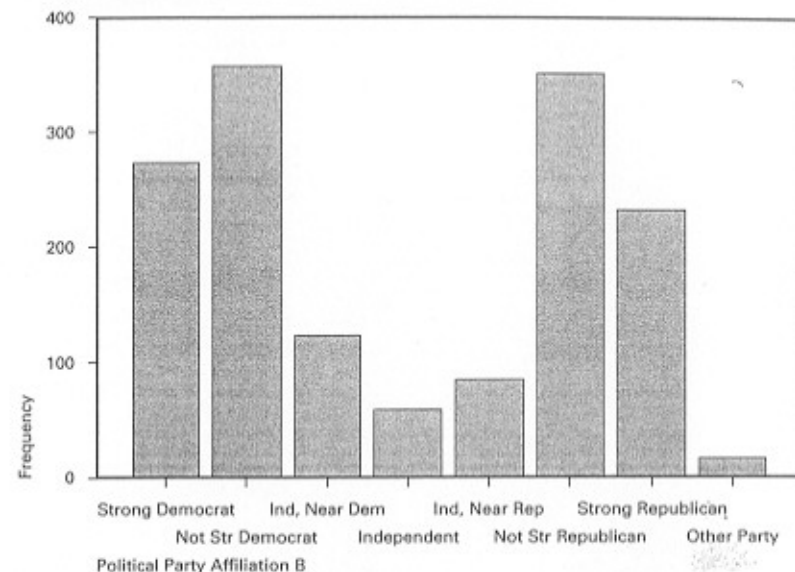


Figure 4.2 Political Party Affiliation B

Table 4.4 Political Party Affiliation C

	Frequency	Percent	Valid Percent
Strong Democrat	292	19.5	19.6
Not Str Democrat	236	15.7	15.8
Ind, Near Democrat	188	12.5	12.6
Independent	93	6.2	6.2
Ind, Near Republican	165	11.0	11.1
Not Str Republican	233	15.5	15.6
Strong Republican	267	17.8	17.9
Other Party	17	1.1	1.1
Total valid	1491	99.4	100.0
NA	9	.6	
Total	1500	100.0	

who are affiliated with the Republicans add up to 33.5%, or about a third. The independents add up to 29.9%. Thus, the population is almost evenly split between the three groups, with the Democrats only slightly ahead of the Republicans. Notice that, within each party, those who are strongly affiliated with the party are more numerous than those who are not (a ratio of 4:3 for the Democrats, and a ratio of 6:5 for the Republicans). This is illustrated in Figure 4.3.

Case D Analysis of case D (Table 4.5). We see from the table that this is a situation of weak polarization between the Republicans and the Democrats. The Democrats attract 42.8% of the population, while the Republicans only get 30% of the

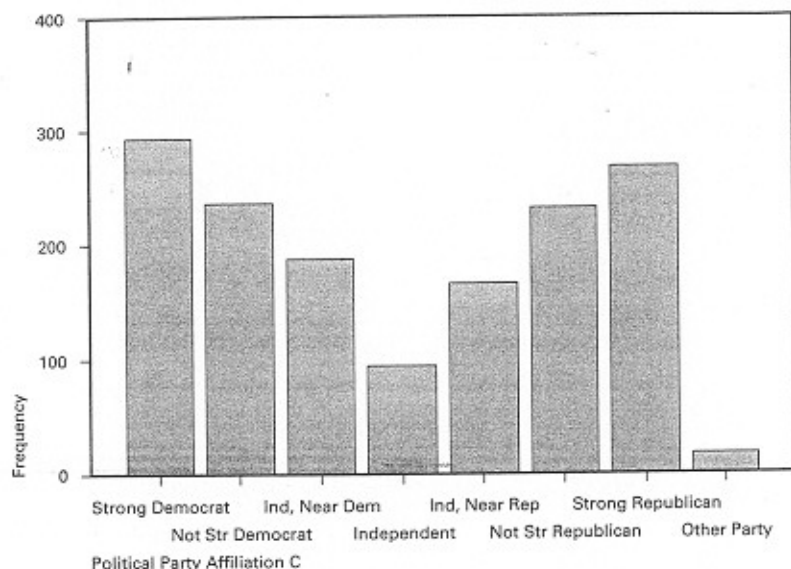


Figure 4.3 Political Party Affiliation C

support, almost 13 points behind the Democrats. The independents add up to 26.0% of the population. Notice that, within each party, those who are strongly affiliated with the party are the majority, with a ratio of about 4:3 for the Democrats and about 5:4 for the Republicans, a situation illustrated by Figure 4.4.

Table 4.5 Political Party Affiliation D

	Frequency	Percent	Valid Percent
Strong Democrat	356	23.7	23.9
Not Str Democrat	282	18.8	18.9
Ind, Near Democrat	188	12.5	12.6
Independent	116	7.7	7.8
Ind, Near Republican	84	5.6	5.6
Not Str Republican	202	13.5	13.5
Strong Republican	246	16.4	16.5
Other Party	17	1.1	1.1
Total valid	1491	99.4	100.0
NA	9	.6	
Total	1500	100.0	

As we have seen, the short descriptive paragraphs that follow each table do not simply report the frequencies. We have tried to highlight the specific features of each situation by answering the following questions: Is there a polarization

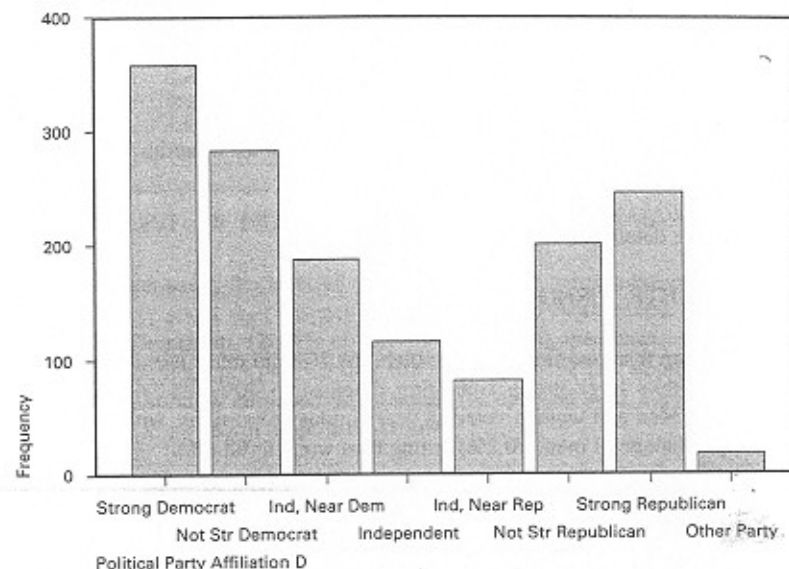


Figure 4.4 Political Party Affiliation D

between the two parties? Is one of them clearly more popular than the other? Is there a large proportion of independents? How is the level of mobilization *within* each party? We answered that last question by providing the ratio of those who feel a strong affiliation to the party compared to those who do not feel a strong affiliation.

A descriptive report that does that systematically is more analytical than one where the percentages are flatly reported as is. Insert 4.3 illustrates such a report.

INSERT 4.3 Description of the Voting Behaviour and of the Political Tendencies of a Sample of US Residents

The data summarized here come from a (non-representative) sample of 1500 individuals, which is a subset of the General Social Survey conducted in the US in 1993.

Four variables deal with our topic: Voting in 1992 Election, Political Party Affiliation, Think of self as Liberal or Conservative, and Political outlook. All four variables are measured at the nominal level. An examination of the frequency tables shows that the last variable is a recode of the third one, as explained below.