CHAPTER 7

STANDARDIZED SCORES

In an example in the preceding chapter, we asked you to select one of four possible sections of a class for your teaching assignment, giving you the mean and standard deviation scores of each section on a language aptitude test. That's a fairly straightforward example where comparing mean scores and standard deviation scores might be useful in making decisions. However, suppose instead that when you were given the values, you found that while two sections had been given the Modern Language Aptitude Test, the other two sections had taken a Short Term Memory Test that a researcher was promoting as a good predictor of language learning ability. Do the scores on the two tests allow you to make comparisons among the four sections now?

If you recall, we have already mentioned that measurement usually uses a set of arbitrary scoring conventions. There is no rule to tell us how to score a given test. That is, you can assign one, two, three points, or any value you prefer, to a correct response on a test. The scores depend on arbitrary scoring decisions set by the test developers. This arbitrariness creates many problems when we want to make comparisons. A score of 90 on a particular test may not have the same interpretation as a 90 on another. To avoid most of these problems and establish criteria for comparing different scores on different tests, we deal with standardized scores rather than raw scores.

Standardized scores are obtained by taking into account the means and standard deviations of any given set of scores and converting them into scores with equal means and equal variances. By converting raw scores to standard scores we can arrive at comparisons which are meaningful.

There are different types of standard scores, but the concept is the same for all of them. The techniques for calculating them are almost identical. Therefore, by understanding what a standard score is, and how we go about finding it, you will be able not only to understand and interpret such scores but also to calculate standard scores for your data.

Before going into the computations, we should first become familiar with the concept of *normal distribution,* for it is the basis of standardized scores.

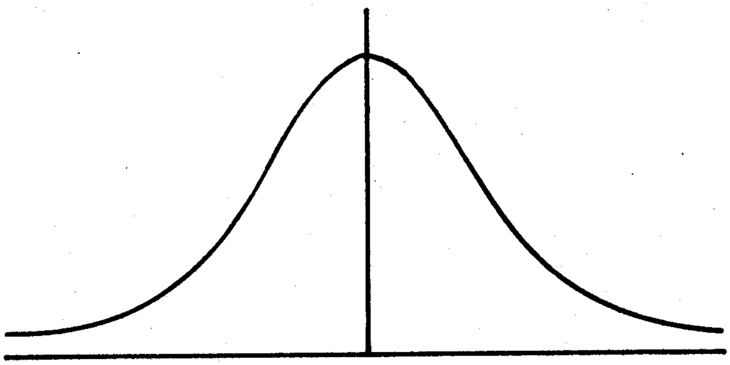
The outcome of almost all human behavior is a normal distribution. No matter what kind of scale is used, no matter what kind of behavior is investigated or what type of data are gathered, the distribution of scores of large samples tends to be normal. For example, if we wanted to find out how long it takes most people to learn 100 new vocabulary items, we could probably get some way of measuring learning time to reach the criterion level for the 100 items. We could then begin tabulating how long it took each person. As we tested more and more people, we would find a curve emerging with most people scoring around a middle point on the distribution (our central tendency point). The other scores would be distributed out from that point. We might find, if we then began adding child subjects, that we would have some learners far off from that point but as we added more and more children, more and more learners of all ages, those differences would gradually become incorporated into our overall curve and, again, most of the scores would cluster around the central point with a smooth flow down from the central point on each side. This normal distribution has special characteristics that allow us to draw important conclusions regardless of the type of data and type of analysis.

This seems logical enough. But, contrary to common sense, the normal distribution does not exist. We never get a completely normal distribution in the real world. Normal distribution is a mathematical concept, an idealization. If you gathered data on a million *S*s, you would get very close to the normal distribution, but we seldom gather data on a million ESL learners, or a million anything else. However, it is usually the case that if the *N* (number of cases) is 30 or more, the distribution of scores of that random sample is close enough to a normal distribution that we will not violate the assumptions of the normal distribution drastically.

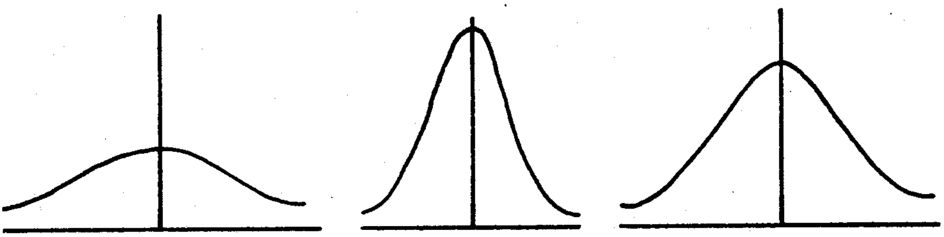
The normal distribution, then, is an idealized model which we can use in dealing with natural behavior. We expect that scores on any behavior will approach a normal distribution. Thus, by comparing the distribution on our data with the normal distribution curve, we can claim that our data matches this expected distribution or, when it does not, that we have obtained a real difference, a difference which is not due to chance but to our treatment.

The normal distribution has three distinct properties that allow us to make inferences about the population in general and our sample of that population in particular. An understanding of these properties is very important and must be kept in mind when carrying out research. They are:

1. The mean, median, and mode in a *normal* distribution are all the same.
2. The first property results in the second characteristic—the shape of the normal distribution is bell-shaped and symmetric.
3. The normal distribution does not have a zero score; the tails never meet the straight line.



This doesn't imply that the means and standard deviations have fixed values in all distributions. They could be quite different and still the distributions would be normal. The following forms are all normal distributions because they have the three properties of a normal distribution:



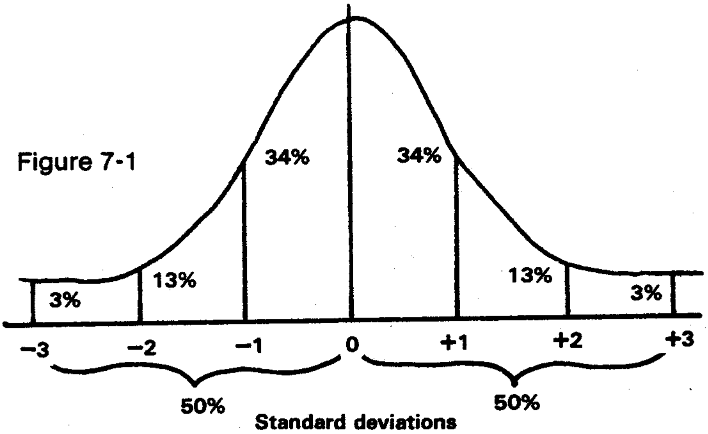
These properties allow us to determine what proportion or what percent of all scores fall in any area of the normal distribution. Since the normal distribution is symmetric, we know that half the scores fall above and half below the midpoint. In the normal distribution we can also visualize the standard deviations from the mean. Between the mean ($\bar{X}$) plus or minus one standard deviation ($\pm 1\ s$) we should find 68% (actually 68.26%) of our observations. Between $\bar{X}$ and $\pm 2s$, 95% of the scores should occur. And between $\bar{X}$ and $\pm 3s$, 99% of the observations are accounted for. The normal distribution, then, looks like Figure 7-1.

By knowing these proportions we can interpret our data. Suppose the mean of the final exam for advanced ESL classes is 500 and the standard deviation is 50. If someone scored 550, we know that he scored one standard deviation above the mean and thus above 84% of the *S*s. Or, if someone scored 400, we know that his score is two standard deviations below the mean (and that he is in trouble).

To interpret any score or performance on any task, we have to tie up the properties of the normal distribution with the concept of standard scores. The standard scores which are most commonly used are called *z scores* and *T scores*.

The easiest way to understand a *z* score is to think of the standard deviation as a measure (a ruler which is one standard deviation long). The *z* score just tells you how many standard deviations above (or below) the mean any score or

Figure 7-1



observation might be. So to compute it, all we have to do is subtract the mean from our individual score. That shows us how high above or below the mean the score is. Then we divide that by the standard deviation to find out how many standard deviations away from the mean we are. That's the *z* score.

$$\boxed{\text{Standard score} = z \text{ score} = \frac{X - \bar{X}}{s}}$$

It's important to remember that if the numerical value of the score is above the mean, the *z* score will be positive and if it is below the mean, the *z* score will be negative.

By finding the *z* score, we can easily say where the raw score falls. For example, if you scored 150 on the psycholinguistics final, a test with a mean of 90 and a standard deviation of 30, your *z* score would be 2.

$$z = \frac{X - \bar{X}}{s} = \frac{150 - 90}{30} = 2$$

This means that your score is two standard deviations above the mean. Looking at the curve, we can then say that you scored as high as or higher than 97% of the students. Most people wouldn't be very delighted if we told them their *z* score on a test was 2, but a *z* score of +2 is an excellent score on a test. −2 is another story!

You may think that *z* scores are only useful for esoteric research, but that is not the case. Consider the following situation. As a teacher in a teacher-training course, you announced two exams—a midterm and a final. You said that the course grade would depend on performance on the two tests. At midterm you gave a multiple-choice test of 100 items. On the final, you gave a problem-solving test of 10 items. Dion, who worked very hard throughout the term, got a 73 on the midterm and an 8 on the problems. Susan, who could always psych out multiple-choice exams, got an 80 on the multiple-choice test and completely failed the final, getting only a "courtesy mark" of 1 (for writing her name on the paper). As soon as the grades came out, Susan saw that she got a very low grade in the class, and she was doubly annoyed when she saw Dion got a good grade. So, of course, she came to see you, the professor, complaining that her 81 points gave her the same average as Dion's 81. She threatened to sue you, the department, and the university. What will you do? The averages *are* the same, but Susan picked up points when points were cheap and Dion earned points when the earning was hard. So you will have to explain z scores to her. By converting the scores to z scores, it's possible to obtain equal units of measurement despite the fact that the original units of measurement were quite different. As a general rule, then, when you want to compare performance on two tests which have different units of measurement, it's important to convert the scores to z scores.

If you are given z scores, you can also turn the calculations around to get back the raw scores. For example, if someone tells you that your z score was -1 and you'd rather not have to face up to a minus score, you can change it to a score that at least looks better. You do need to know the mean and the standard deviation for the test. Let's say the test developer published the mean as 50 and the standard deviation as 10.

$$z = \frac{X-\overline{X}}{s}$$

$$X-\overline{X} = (s)(z)$$

$$\begin{aligned} X &= (s)(z) + \overline{X} \\ &= (10)(-1) + 50 \\ &= -10 + 50 \\ &= 40 \end{aligned}$$

Now that you understand the concept of standard scores and can compute z scores, let's consider T scores. T scores are another version of z scores with the difference that there is less chance that you will make an error in reporting them. z scores often contain decimal points (a score may be 1.6 standard deviations above or below a mean, rather than just 1 or 2) and they may be either positive or negative. That makes for error in reporting if you are not very careful. With T scores we don't have this problem since the mean of the T distribution is set at 50 instead of at zero as in the z score. And the standard deviation of T scores is 10. To calculate any T score, we first find the z score and then convert it into a T score:

$$\begin{aligned} T \text{ score} &= (s)(z) + \overline{X} \\ &= (10)(z) + 50 \end{aligned}$$

By changing the mean from zero to 50, we end up with positive numbers. And by multiplying the z score by a standard deviation of 10, we also come up with whole numbers rather than fractions. For example, if you got a z score of 2.1 on a test, we could multiply it by 10 to get a whole number 21. Then we add the mean: $21 + 50 = 71$. If your score had been a -3.4, it would become a positive, whole number when we converted it to a T score of 16. This may not seem like an important thing to do. However, when you are converting many, many scores to standard scores you are less likely to become confused if you convert everything into positive, whole numbers.

We do not usually give *S*s their z or T scores. As we mentioned earlier, you probably wouldn't be overjoyed to get back a test with a z score of $+2$ unless you understood what the score really meant. You can imagine how you might feel if you got a test back with a $-.4$ score. While it means that you got very close to the average for the test, most people don't react very well to negative scores. Let's say you took the Graduate Record Exam three times. Your scores were 400, 500, and 600. The mean for the test is 500 and the s is 100. The first time you scored below the mean, the second time at the mean, and the third time above the mean. Your z scores would be -1, 0, and $+1$, respectively. Unless you understood what the scores meant, you would be very disappointed. Even if you did understand, you'd probably rather have scores reported to you as 400, 500, and 600.

Nevertheless, whenever you are comparing or combining scores which have different units of measurement, you must convert them into standard scores. These must be reported along with the means and standard deviations in the first part of the "Results" section of your research paper.

ACTIVITIES

1. You want to compare gains made by a group of *S*s following a special instructional treatment. You have a control group too. At the end of the course, you give a general proficiency test in English to both control and experimental subjects. You then discover that you also have your university's placement scores for some of the *S*s and TOEFL scores for the rest. You want to use these scores as the pretest. What can you do?
2. You took the GRE and Educational Testing Service sent you your z score by mistake. After the shock wore off, you decided to find out what your raw score was. The z score was 2.2. You know the $\bar{X}$ is 400 and the s is 100. Your raw score looks better. What is it?
3. You have changed the form of your Placement Exam. The scoring procedure along with everything else is completely different. The Dean of Foreign Students calls you because she has had "numerous complaints" from foreign students who know they have been placed at the wrong class level. The students obtained scores of 52, 64, 31, and 22 on the new test, which has a mean of 55 and a standard deviation of 2.5. The old test had a mean of 50 and a standard deviation of 2.5. If students scoring $1s$ and above were excused from English classes, those between the mean and $1s$ required to take 1 English class, those between $-1s$ and $\bar{X}$ take 2 classes, and those lower take 3, would the students have been any better off under the old test than under the new one? What would their z scores be? How much English would they have had to take under the old rules?
4. Your media specialist has produced a set of beginning ESL lessons with accompanying tapes. These materials are used by your experimental group, a class of high school students studying English in Japan. Your equivalent control group gets the same material in class, but instead of using the tapes, their Japanese teacher, who is very fluent in English, covers the same material. At the end of the time you hope to be able to compare the pronunciation of the *S*s in the two groups. Unfortunately, two different tests of pronunciation had to be given in the two classes.

They are equivalent tests but the scoring method is different. The experimental group scores are: 54, 56, 33, 39, 22, 38, 48, 42, 41, 38. The control scores were: 88, 89, 100, 79, 75, 78, 88, 84, 91, 93. Convert the scores to *z* scores and *T* scores. At this point does there appear to be any difference between the two groups?

5. You have only ten places in your M.A.—TESL graduate program. The criteria for selecting graduate students for the program are the GPA (grade point average) and a rating on the applicants' tentative research proposals. The GPA's of your 15 applicants are: 3.2, 3.25, 3.25, 3.3, 3.3, 3.3, 3.5, 3.7, 3.8, 3.8, 3.8, 3.9, 3.9, 3.95, 4.0 and their proposal scores (in the same order) are: 14, 12, 9, 12, 16, 18, 9, 12, 14, 14, 14, 12, 9, 8, 16. Convert the scores to standard scores and select the top 10 for admission. Which *S*s are likely to come in to ask for an explanation of why they were not selected? How would you explain the process of selection?
6. Using one of the standard scores that you got in problem 5, show how you can reconstruct the original raw score.

USING COMPUTERS TO DO THE WORK

It's time now to make the computers do some of the work for you. In the beginning, you may feel anxious about dealing with computers. A computer may seem to be a magic box much more mysterious than your desk calculator. However, with a minimal amount of computer literacy, you will be able to put this magic box at your command so that it will do whatever you ask it to do. By following simple directions, you can ask the computer to do computations in less than a second (or even a millisecond) which would have taken you days to accomplish with your desk calculator.

From this chapter on, you will find instructions for computerizing data in ways relevant to the concepts covered in each chapter. However, one word of caution is necessary. If you use these instructions alone, you will never ever understand how the computer does the computations. Therefore, as we emphasized in earlier chapters, you must first fully understand the concepts and then ask the computer to do the labor. If you don't understand the concepts, you will undoubtedly ask the computer to do the wrong things. The computer doesn't know whether your requests are sensible or not; it just does whatever you ask.

One of the most important points to keep in mind when dealing with computers is that they are machines. They don't understand what you mean unless you talk to them in the language in which they have been programmed. For example, if they "read" a period at a certain point, when you meant to type a comma instead, they will never understand that you meant comma, not period. Therefore, absolute accuracy is a prerequisite in working with computers. All your messages to the computer must be accurate; no "typos" are allowed.

To introduce you to the computer, we will give you sample data and ask you to carry out the operations using that sample. Of course, this isn't going to be terribly fascinating data. You can use your own data (which is more desirable, more meaningful for you) following our directions. However, we will use the same data set over and over again, thus maintaining consistency over many different statistical tests.

We will not explain how the computer programs work, simply because it is well beyond the scope of this book (or any other statistics book for that matter).

We don't intend to teach you how to do programming. However, what we will do is use the already available package programs. One such package, in fact the most commonly used one, is the statistical package for social sciences (SPSS). You should buy and use the SPSS Manual. Our exercises and explanations should be regarded as a supplement to (not a replacement for) the Manual. We will start with an explanation of the data and the variables we will use, then work through the program, and finally interpret the printouts which you will get from the computer.

Sample data

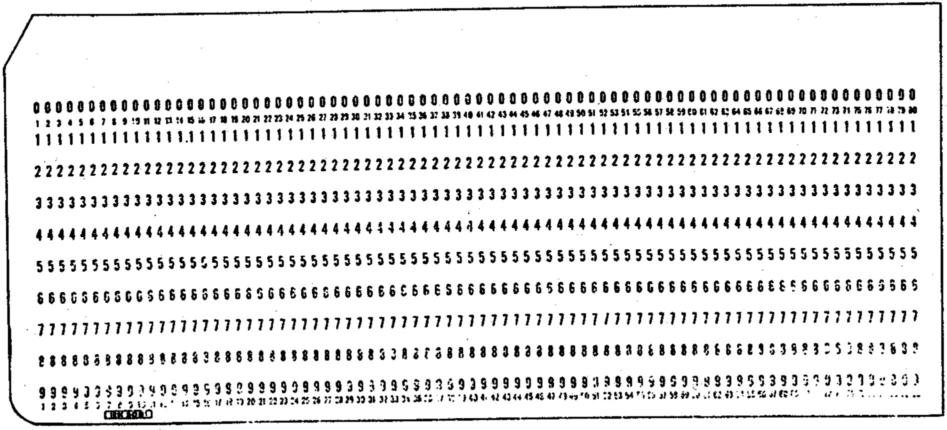
The data to be used in the example programs throughout this book include the following information:

1. Subject variables: The *S*s were 40 incoming foreign students (male and female) who took the UCLA ESL examination. They were from four countries (Iran, Brazil, India, and the Philippines).
2. Test variables: The ESL examination consists of 8 subtests. To simplify the task, we will use only 5 (dictation, cloze, listening comprehension, grammar, and reading).

Transferring the data to cards

The first step is, of course, to find the location of your computer center and the keypunching machines. The keypunching machine is similar to a typewriter; punching cards is similar to typing. At the same location you usually will find computer cards which you can purchase. As you type, the machine punches holes in the appropriate places on the card. Posted near the keypunching machine, you will usually find instructions on how to operate the keypunching machine (how to turn it on and off, how to insert the cards in the machine, etc.). If you have a problem, there is usually a consultant on hand who will help you learn to operate the machine.

Once you have located and learned to operate the keypunching machine, you must transfer the data to cards which the computer can read. You will punch one card for each subject. So, for our data, you will have a total of 40 cards. Each computer card has 80 columns.



Each column can be used for one, independent piece of information about the subject, or combinations of columns can represent a variable (if the score is two digits or more). To type the data, punch them in the following order across the card:

Column(s)	*Variable*
1, 2	Identification number for the *S*s. The ID number usually starts with 01 and ends with the last number of last subject (in this case, 40). The ID number has no mathematical meaning; it's just for identification.
3	Blank. To keep variables separate and easy to read, it helps to leave a blank column between each pair of variables. In our directions, we will use a ____ to mean a blank space. This doesn't mean that you type a ____ but rather that you skip a space.
4	Nationality. As mentioned above, the *S*s were from four different countries. We will represent each level of this nominal variable (nationality) with a number. Each country will have a number. In this case, 1 for Iran, 2 for Brazil, 3 for India, and 4 for the Philippines. Punch a 1 in column 4 if the *S* is from Iran, 2 if from Brazil, etc.

So far we have punched information about the *S*s. Of course, we could have had a column for the sex variable (1 for males, 2 for females, or vice versa). Also, we could have had other columns for other *S* variables such as years in the United States or age. However, to keep the data and our assignments simple, we will only include the *S*s ID number and nationality.

Column(s)	*Variable*
5	Blank
6, 7	Dictation, the score of the *S* on dictation. If the score is below 10, punch a 0 in column 6 (e.g., 09, 08)
8	Blank
9, 10	Cloze, the score of the *S* on cloze
11	Blank
12, 13	Listening comprehension, the *S*'s score on this variable
14	Blank
15, 16	Grammar, the *S*'s score on this subtest
17	Blank
18, 19	Reading comprehension, the *S*'s score on this subtest

For simplicity's sake, we have included only 5 variables from the ESL examination. In your own research, you may have only a few variables to punch or you may have many. Your data is usually first recorded on a data sheet. It helps you use exactly the same order on the data sheet as you plan to use on your cards. The data for our 40 *S*s appear on the following standard IBM data sheets. Punch the 40 data cards.

PROGRAM Data List
PROGRAMMER page 1

ID (1–2)	NATION (4)	DICTAT. (6–7)	Cloze (9–10)	Lis.Comp (12–13)	GRAM. (15–16)	READ (18–19)
01	1	22	14	10	29	06
02	1	47	29	27	52	19
03	1	41	22	24	38	11
04	1	25	20	16	31	11
05	1	16	00	10	22	01
06	1	31	28	25	29	18
07	1	35	24	21	45	05
08	1	42	22	21	42	04
09	1	23	08	17	39	08
10	1	27	16	14	32	06
11	2	47	29	26	57	09
12	2	39	16	20	36	17
13	2	06	06	12	19	01
14	2	12	00	07	22	00
15	2	47	36	26	50	19
16	2	38	24	15	28	16
17	2	19	09	21	25	03
18	2	37	23	16	34	07
19	2	14	09	11	25	05
20	2	18	11	15	23	05

PROGRAM Continued
PROGRAMMER page 2

ID (1–2)	NATION (4)	DICTAT. (6–7)	Cloze (9–10)	Lis.Comp (12–13)	GRAM. (15–16)	READ (18–19)
21	3	40	27	18	45	17
22	3	50	36	26	59	25
23	3	49	36	23	52	19
24	3	48	32	23	53	20
25	3	50	35	28	58	21
26	3	50	39	27	61	25
27	3	30	27	15	37	18
28	3	50	34	29	58	23
29	3	48	34	24	57	23
30	3	47	32	26	56	18
31	4	48	32	27	48	22
32	4	35	20	20	41	05
33	4	49	29	26	51	20
34	4	45	29	28	53	19
35	4	50	28	26	56	18
36	4	37	19	24	39	14
37	4	45	29	23	52	19
38	4	50	35	26	58	21
39	4	50	41	26	59	23
40	4	49	33	28	56	23

Running your program

After you have punched the data on the cards, you must give them to the computer. The computer, however, won't know what to do with them unless you also give it instructions telling it what to do. Once you have selected a program to tell it what to do, you will run instructions and data cards through the computer for analysis. For most computers, you will need three sets of cards to do this: job control cards, data control cards, and command cards.

I. Job control cards (often abbreviated JCL). The job control cards tell the computer what program you plan to run and who you are (that you have the right to give it orders). Different computer centers have different ways of letting you give this information, so you should check with your computing center to verify the form of the JCL cards. Most use three job control cards which must be in the following order:

Card A: The program identification card
Card B: The charge number card (the number of your computer account)
Card C: The password card

The form of these cards will be constant through all your work, whatever the analysis. As an example, these cards might be:

For the first card
(starting in Column 1) //____QUICKRUN
For the second card
(starting in Column 1) $ALI240,SPSS,HOSSEIN
For the third card
(starting in Column 1) //____PASSWORD____EVELYN

Remember that we are using a ____ to represent a blank column. It does *not* mean you should punch a ____ but rather that you must leave a column blank.

What these three cards say is the following:

1. The first card identifies the program you want to use.
2. The second card identifies the account number so you will have to insert your own account number in these columns. Following the first comma, SPSS identifies the statistical package the computer must use. This is followed by the user's name, your name.
3. The third card is the secret password by which the computer identifies eligibility to use the account number. The password is an arbitrary word that you choose and keep confidential so that others cannot use your account. In fact, you should change it frequently to be sure that only you and the computer know it. To tell the computer you want to change the password, just punch the new password after the old one and separate the two with a comma:

Card to change the password
(beginning in Column 1): //____PASSWORD____EVELYN,TESL
Afterward, you will use the
following card: //____PASSWORD____TESL

and the computer will know you are authorized to use the account.

Once again, remember that ____ is used to indicate a blank space. It is very important to keep the blank space wherever it is indicated. Otherwise the computer will get upset and your program will not be processed.

Once you have the job control cards ready, you can start punching the second set of cards, the data control cards. The cards do not have one invariate form. They will change according to the number and kinds of variables you have in your research project. However, in our example, we will need those which appear on our data sheets.

II. Data control cards. These cards follow in consecutive order the three job control cards. However, they will differ in an important way from the previous three cards. The *name* of the card starts in column 1 but the *information* begins in column 16. The five data control cards we will start with will tell the computer the name of your experiment, the variables you have selected for analysis, the form in which the data will be given to it (whether cards or tape), the number of *S*s, and directions on how to read the data cards.

Beginning in Column 1	*Beginning in Column 16*
Card D:	EXAMPLE___PROBLEM
RUN___NAME	(Use any name you like to identify your resesearch)
Card E:	
VARIABLE___LIST	NATION,DICTAT,CLOZE,LIS,GRAM,READ (You may abbreviate your variables to only the first letter if you wish; (i.e., N,D,C,L,G,R. You can only use 8 characters for each variable.)

However you decide to abbreviate your variables on this card, you *must* enter them in the same order as they have been punched on the data cards. And they *must* be separated by a comma with *no* blank spaces. If you change the order in listing the variables, the computer will assign the names to the wrong data. If you forget to put in a comma or leave blank spaces, the computer will probably tell you that you have made an error.

Card F:	CARD
INPUT___MEDIUM	(You're using cards, not tape)
Card G:	40
N___OF___CASES	(N of cases refers to number of *S*s)
Card H:	
INPUT___FORMAT	FIXED(3X,F1.0,5F3.0)

The input format card tells the computer how to find your variables on the cards. Each word in the format means:

FIXED	This means that all the cards have the same data in the same order.
3X	The letter X means "to skip." The number before X refers to the number of columns to be skipped. In this case, the first two columns of the data card contain the *S*s' ID numbers (which we won't need for analysis) and the third column is blank. So, the computer has been told to skip the first three columns.
F1.0	The letter F means "to read." The number after F refers to the number of columns to be read. For example, we said that the computer will skip the first three columns and start reading column 4. The variable in column 4 is NATION (*S*'s nationality). So the

computer will read only column 4 to find that variable. The decimal after the number refers to the decimal points in the data. Since we do not have a decimal point in the data, we put in a 0. Otherwise, we would have put a 1 or 2 depending on the number of decimals.

5F3.0 We know that F means "to read." The number after it means the number of columns to be read as one variable. In this case we have a blank and a two-digit number, so we need a 3 for 3 columns. Since all the variables have a blank and 2 digits (3 columns each) and there are 5 variables, we punch in 5 times 3 columns. Since the blank to the left of the two-digit score will be read as a 0, it will have no effect on the scores (020 = 20).

Instead of giving the computer the instructions to read three columns five times (5F3.0), we could have punched the card as: FIXED(3X,F1.0,1X,F2.0,1X,F2.0,1X,F2.0,1X,F2.0,1X,F2.0). This reads as: skip the first 3 columns, read column 4 as one variable, skip 1 column, read the next 2 columns as one variable, and so forth. But for economy and simplicity, we can use 5F3.0 instead.

After the data control cards are punched, the third set of cards needs to be punched. These are called command cards.

III. Command cards. The command cards refer to the specific analysis you want the computer to run. For example, our first assignment on the computer is to get the descriptive statistics (mean, standard deviation, variance, range, etc.) for each variable. The command card which asks the computer to do this is called "condescriptive."

Beginning in column 1	*Beginning in column 16*
Card I:	DICTAT____TO____READ
CONDESCRIPTIVE	

This means that we want to get the descriptive statistics for the variables starting with dictation and ending with reading (including all the variables between the two).

After this command card, we must ask the computer for the information we want, i.e., the statistics we want. So, we use a card which simply says "statistics." The number of statistics available is designated in the manual. However, to get all possible information, it is safe to ask for all statistics. (Sometimes there is so much information given that it will save you money and time to specify exactly what you want—for example, means and standard deviations.)

Beginning in column 1	*Beginning in column 16*
Card J:	ALL
STATISTICS	

Then you need a card to tell the computer to read the data according to the information you have provided:

Card K: READ____INPUT____DATA

Now you put everything together—first the 3 job control cards, then the 5 data control cards, then the 3 command cards, and then your 40 data cards (one for each of your *S*s).

Then, remembering how stupid the computer is, you know you must tell it that that's all there is. It takes two cards to do this:

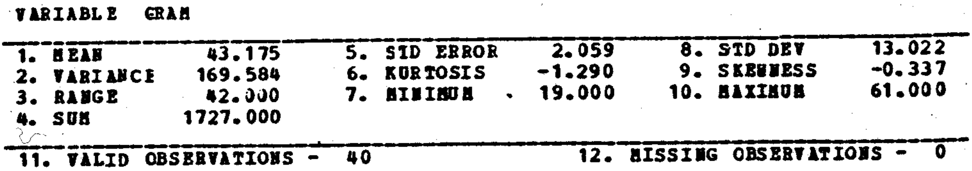
Card L: (beginning in column 1) FINISH

Card M: (beginning in column 1) / /

To summarize, you will have 40 data cards and 13 cards to tell the computer what the cards are, who they belong to, and what it should do with them. They must be in the form and order shown on page 77.

Of course, this is a minimal amount of data. The first time you try to prepare your data for the computer, it may seem like an onerous task, but it becomes very easy and automatic after a couple of times. Our data are banal, but that is not the point. The point is that we want you to feel at ease with the computer so that you can soon run many programs without the stress or anxiety you may feel the first time you try.

After you run your program, you will get a printout back from the computer. Just as you had to put your messages into a form the computer could read and understand, you must now learn to read the message the computer sends back to you. We believe that the most important part of analysis is to be able to interpret the printout and understand what the numbers on it mean. Therefore, for each sample example, a part of the printout will be reproduced here and each element will be explained.



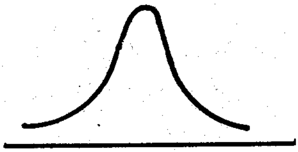
```
VARIABLE  GRAM
---------------------------------------------------------------------------------
 1. MEAN          43.175     5. STD ERROR      2.059      8. STD DEV       13.022
 2. VARIANCE     169.584     6. KURTOSIS      -1.290      9. SKEWNESS      -0.337
 3. RANGE         42.000     7. MINIMUM       19.000     10. MAXIMUM       61.000
 4. SUM         1727.000
---------------------------------------------------------------------------------
11. VALID OBSERVATIONS -  40                  12. MISSING OBSERVATIONS -  0
```

The first page of the printout just gives you back the information that you gave the computer. Then the computer gives you information on each of the variables. It gives us 12 pieces of information about each:

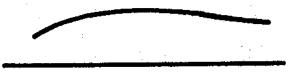
1. MEAN	$\overline{X} = \Sigma X/N$
2. VARIANCE	s^2 (standard deviation squared)
3. RANGE	$X_{\text{highest}} - X_{\text{lowest}}$
4. SUM	The sum of scores of all 40 *S*s on this variable.

Order	*Card Type*	*Card Form*	
		Column 1	Column 16
1	Job Control Cards	//_QUICKRUN	
2		$ALI240,SPSS,HOSSEIN	
3		//_PASSWORD_EVELYN	
4	Data Control Cards	RUN_NAME	EXAMPLE_PROBLEM
5		VARIABLE_LIST	NATION,DICTAT,CLOZE,LIS,GRAM,READ
6		INPUT_MEDIUM	CARD
7		N_OF_CASES	40
8		INPUT_FORMAT	FIXED(3X,F1.0,5F3.0)
9	Command Cards	CONDESCRIPTIVE	DICTAT_TO_READ
10		STATISTICS	ALL
11		READ_INPUT_DATA	
12 to 51		Your 40 data cards	
52		FINISH	
53		//	

5. STD ERROR	The standard error of the means (this concept will be explained in later chapters) $s_{\overline{X}} = s / \sqrt{N}$
6. KURTOSIS	This refers to a measure of peakedness or flatness of the distribution. A highly peaked distribution will have a positive value of kurtosis; a flat distribution will have a negative value of kurtosis. So this number gives you information on the shape of the distribution. A normal distribution will have a kurtosis value of zero or close to zero



Peaked



Flat

7. MINIMUM	The lowest score on this variable
8. STD DEV	The standard deviation (s)
9. SKEWNESS	The value should be zero or close to zero to indicate normality of distribution. Outliers may skew the distribution either positively or negatively. This value, then, gives you information on the shape of the distribution
10. MAXIMUM	The highest score on this variable
11. VALID OBSERVATION	The number of *S*s included in the analysis
12. MISSING OBSERVATIONS	The number of *S*s for which there was no data on this variable. The sum of 11 and 12 must equal the total number of subjects

If you are not sure about the meanings of each of these terms, check back through the chapters and review the terms carefully. Some of them, of course, will not be discussed in detail until later in the volume. But now you are better prepared to notice them when they do occur.

We hope that you have been able to carry out this first computer assignment successfully. It is possible that your computer has special quirks which make it necessary to program some of the cards (particularly the job control cards) slightly differently. If you have any problems at all, don't be afraid to ask for assistance. If the consultant is not available, ask another student to help. Remember that they were "beginners" once themselves. Most people who use computers grow to love them and so are happy to help you learn to use them too.

Suggested further reading for this chapter: Ebel, and The SPSS Manual.