CHAPTER 2: DATA ANALYSIS AND STATISTICAL CONTROL

1. Introduction

In this chapter we are going to begin the analysis of data, emphasizing the use of four important statistical tools: control charts, runs charts, histograms, and scatter plots. Data analysis presupposes data, and obtaining relevant data is central to our understanding of how well a process is performing. Examples of useful data about a process are:

- Yields. (We want yields to be high.)

- Cycle times, the time for one unit of output to pass through the process. (We want cycle times to be short.)

- The ratio C/P of total cycle time C to that part P of cycle time in which value is actually being added to the product or service. (An ideal C/P would be 1/1; a good value would be 5/1; a typical value would be 100/1.)

- Defect rates, the ratio of defects or errors to the total opportunities for defects or errors. (We want defect rates to be low.)

- Breakdowns or failures. (We want the time between breakdowns or failures to be high.)

- Accidents. (We want the time between accidents to be high.)

Data are most useful when collected systematically through time, because we then can ascertain whether the process is behaving consistently through time -- is in "statistical control" -- and we can assess the success of our efforts to improve it.

"Lightning Data Sets"

To do data analysis, we need to know some simple ideas of statistics. These ideas will be introduced informally and intuitively, with numerous examples based on a variety of data sets, including what we call "lightning data sets":

> Lightning data sets refer to simple yet interesting everyday situations that are easy to understand and relate to.
>
> They resemble data arising in practical applications.
>
> Collection of the data requires very little time, typically a few minutes.
>
> Readers can easily collect their own lightning data sets for practice, and, in so doing, can learn about data collection and can sharpen observational skills.

The Role of *SPSS*

SPSS will be used to carry out all data analyses. Usually ***SPSS*** will display more output than we need at the moment. We must concentrate on what we actually need at the moment and avoid being distracted by the remaining output. **Our aim is to tell you only what you need to know from any display of *SPSS* output, so screen out the rest.** (If there is something important in the remaining output, we will ordinarily get back to it later.)

An urgent exhortation is in order. Most readers will be better able to follow our ***SPSS*** output, especially at the beginning, if they replicate it, hands on the computer keyboard, as they read. (In a few cases we may slightly edit our ***SPSS*** output for pedagogical or display purposes, but basically you can replicate all the analyses, just as they are displayed in the book.)

The reason for this exhortation is this: although the individual steps of any data analysis are simple, **all** steps must be understood if the analysis itself is to be understood. If a gap of understanding does arise, it **may** close of its own accord on further passive reading. But it **may not** close. If it doesn't, it can waste a lot of your time. If you don't straighten it out and if further gaps arise, you can become seriously confused about things that are essentially very simple.

Active reading, replicating the analysis at the keyboard, tends to prevent gaps from arising in the first place. In advocating active reading, we are applying a basic principle of quality: **prevention of defects is preferable to rework.**

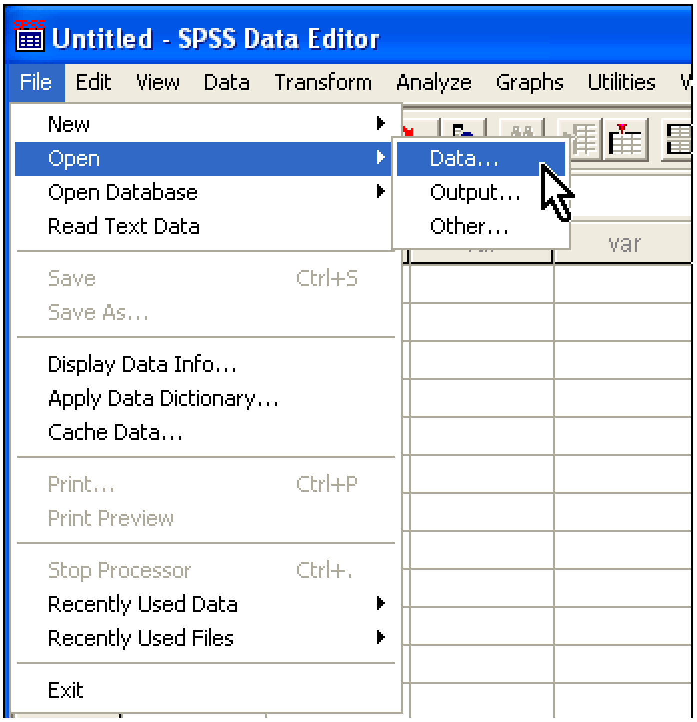
2. A Simple Target Game to Illustrate Statistical Control

We begin with a lightning data set that illustrates the important concept of a **state of statistical control** and introduces first steps of data analysis. One of the authors devised a simple game in which he aimed at a target and tried to come as close to it as possible. In just a few minutes he was able to obtain 30 consecutive data points or "observations" on this game and we use these data points below in order to introduce data analysis.

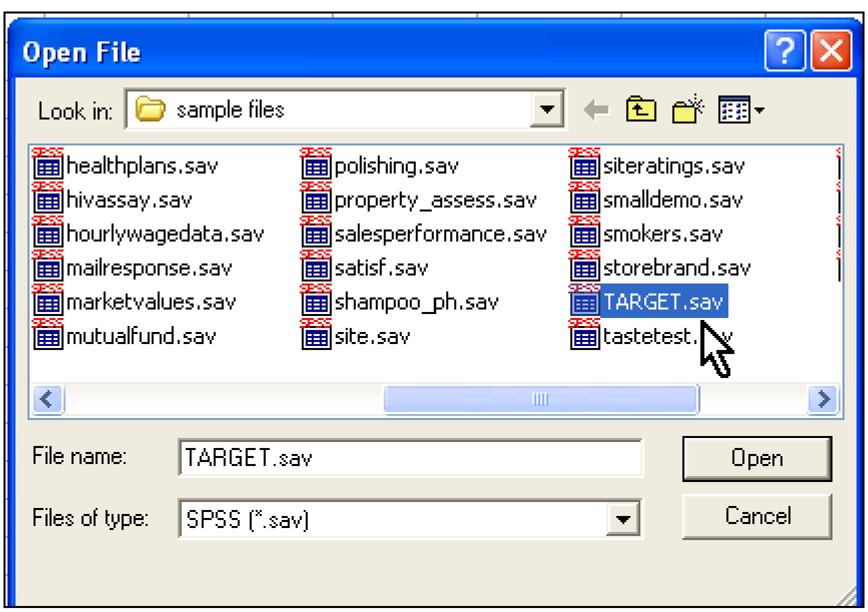
The game contains the essence of many work processes in which we are aiming at a specified output result. We can use data analysis to monitor, and ultimately to improve, our performance in the game. In so doing we use the same tools of data analysis that are useful in all applications of quality management.[1]

[1] To try out the tools for yourself after you have studied this application, you could make up a similar game or you could play a standard game like darts, rifle or bow-and-arrow shooting, or golf. Alternatively, you might also have access to a work process that would generate similar data; for example, a manufacturing process for which the data are deviations of actual from specified dimension of a manufactured part, actual from guaranteed weight of the contents of a jar of peanut butter, actual from nominal concentration of a chemical product, or the actual from promised delivery date of a particular mail order item.

The data that we will use are contained in an ***SPSS*** data file called TARGET.sav. As you see in the image below, we click on **File** on the main menu bar followed by **Open** in the menu itself and by **Data** in the sub-menu. In the future we will write this sequence of steps and other sequences in shorthand form as **File/Open/Data**:

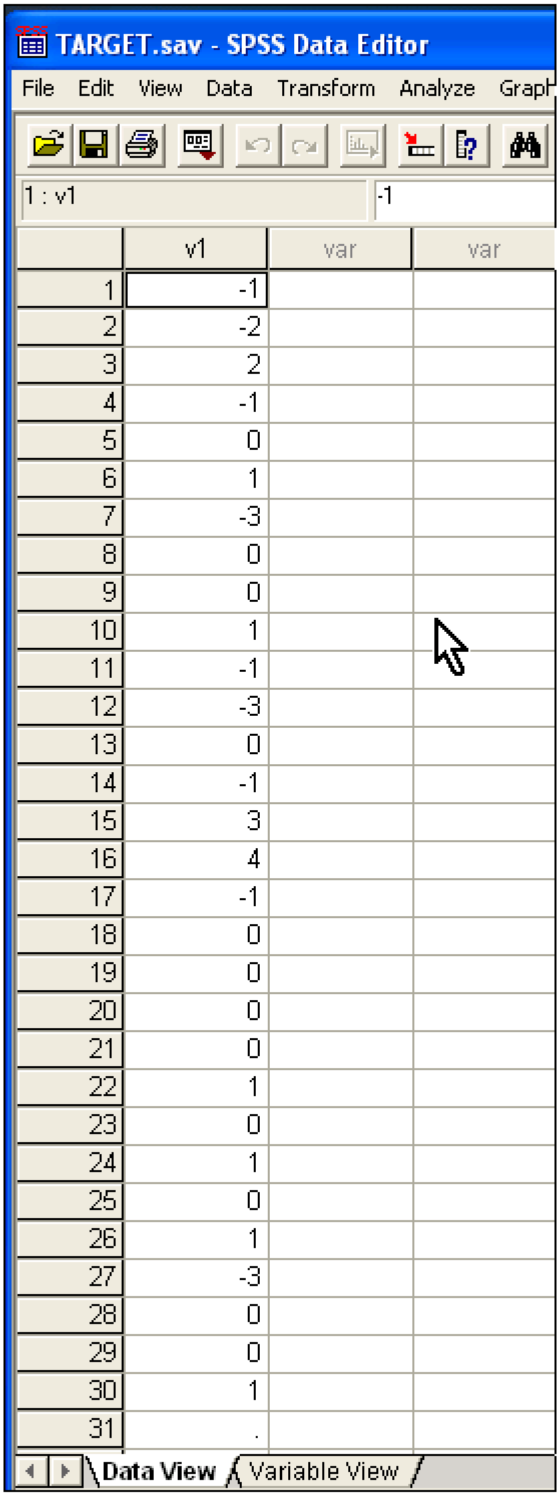


The sequence **File/Open/Data** causes the following window to be displayed:



The same result could have been obtained in an easier way by merely clicking on the  icon on the tool bar.

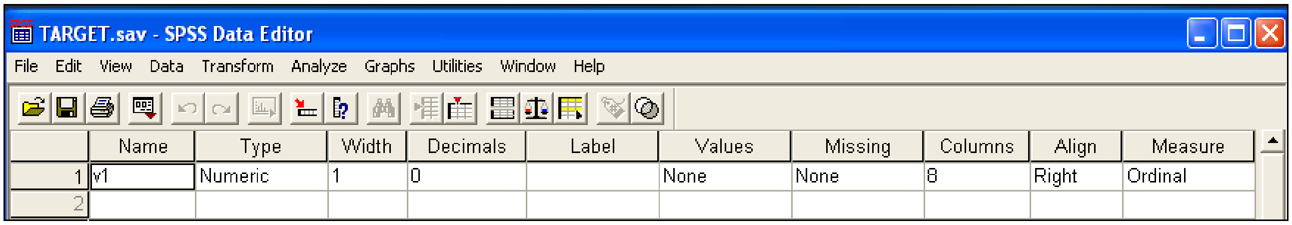
Notice that the file TARGET.sav has been highlighted in the list of available ***SPSS*** data files. We have placed the files that will be used in this textbook with the sample files that came with the student version package, i.e., in the directory **c:\Program Files\SPSS Student\tutorial\sample files**. When we click on the **Open** button the data are displayed in the Data Editor as follows:



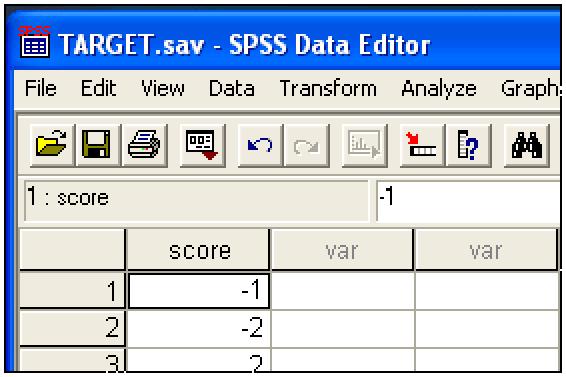
The numbers shown here are the results of an improvised target game that anyone can play in an office or at home. The player simply tosses a Post-it pad on a tiled floor, aiming at a central strip a few feet away. The target strip of tiles runs at right angles to the direction of aim. The tiles in the strip are square and about four inches in diameter. If the Post-it pad lands in the target strip the **score** is zero. If the pad lands on a tile that is two strips beyond the target strip the **score** is +2. If it lands one strip too close, the **score** is -1, etc. If the pad lands on a line between two strips, the strip closer to zero is counted for scoring.

In this particular round thirty tosses were made and recorded. The scores were entered into the Data Editor by hand and then saved as the file TARGET.sav. Because a name for the variable that was recorded was not included in the file, the designation **"v1"** is assigned to the column when the data are displayed. Note that values for many different processes can be stored in the same file, with each column corresponding to a different **variable**. The rows in each column, showing the various data values, are called **cases**.

If we click on the **Variable View** tab at the bottom of the editor, the display changes to the following:



and by clicking on the cell directly under **Name** we can give the variable a name by typing in the word "**score**". Thus the Data Editor in many ways is similar to a spreadsheet. The upper left corner of the editor now looks like this:



What can we learn about the process from this "sample" of 30 observations: How is the player doing?[2]

Ideal performance in the game would be all zeros, right on the target strip. The player didn't do that well. We can judge how well he did by considering two questions:

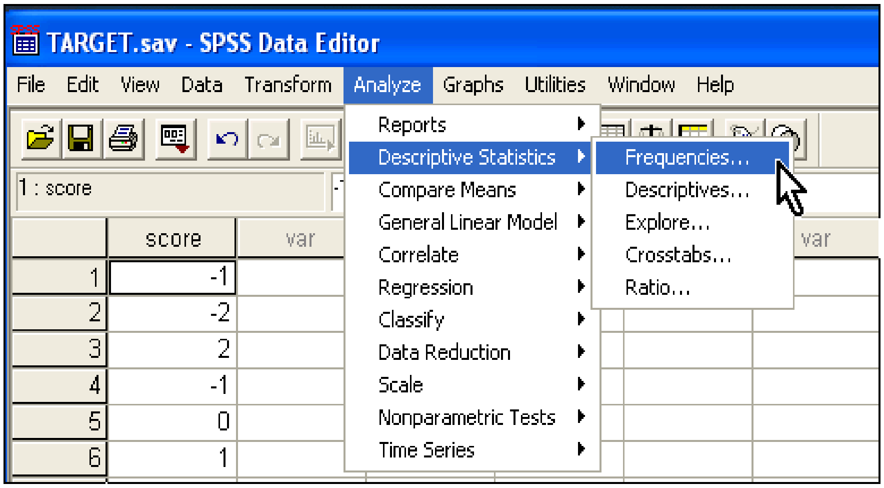
> Did his shots average out on target, that is, was the average of his deviations from target close to zero?
>
> Were his misses (deviations) usually small? That is, was the variability of his shots small?

In managing processes, we often want to be on target on the average with as little variation as possible.

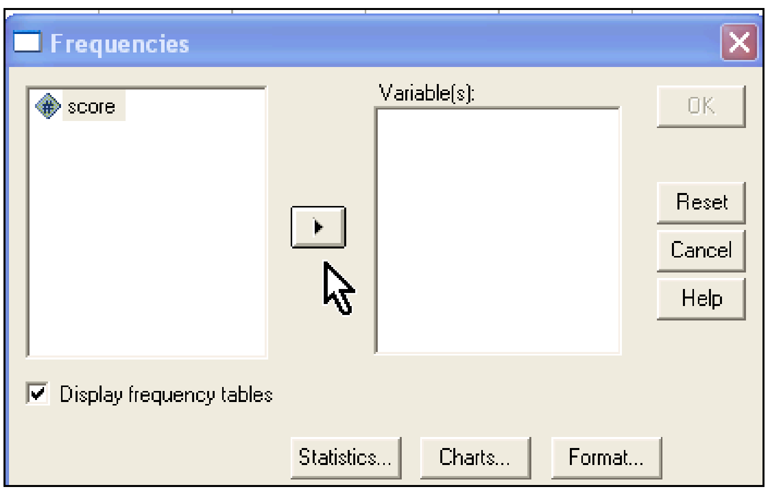
[2]The player could have measured the actual distance in inches from the center of the target strip and obtained more precise data. However, the essence of lightning data sets is speed, and this method of arriving at the measured error is adequate to bring out the needed ideas of process behavior and data analysis.

3. Preliminary Examination of Data

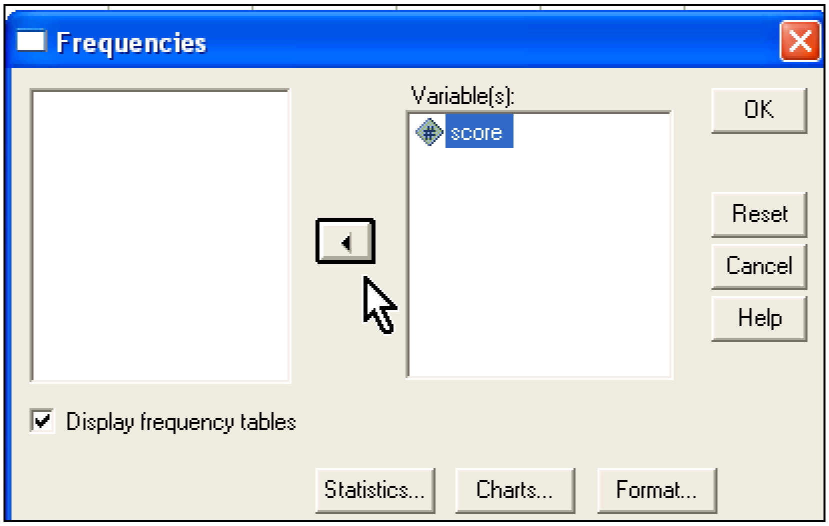
We begin our study of behavior of the variable **score** by performing the ***SPSS*** steps **Analyze/Descriptive Statistics/ Frequencies:**



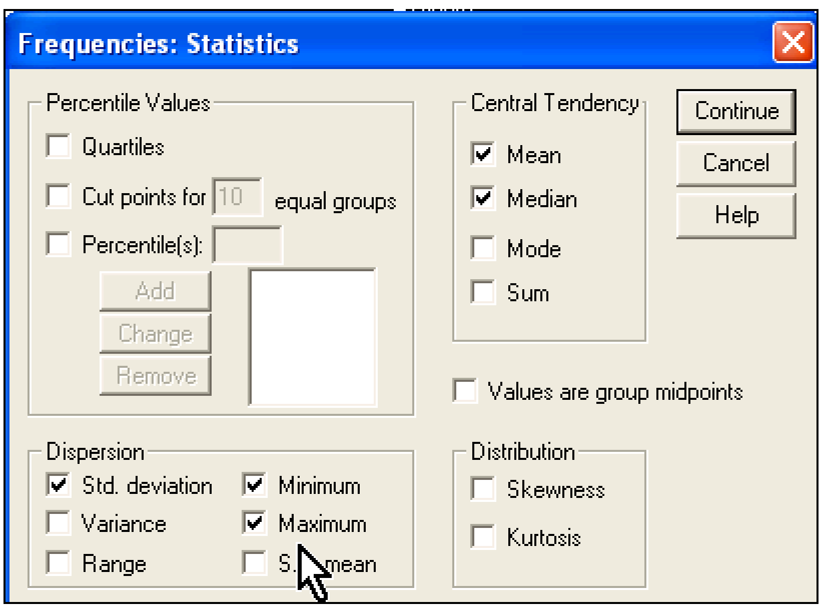
The mouse click results in



We move the variable name **score** into the section for Variable(s) by highlighting it and clicking on the arrow near the center of the window:



Next, we click on the little box at the left labeled **Display frequency tables** to remove the check mark. We then click on the **Statistics** button at the bottom of the display. The latter action brings up the following:



When you first click on the **Statistics** button all of the boxes in the display will be unchecked. You should make your window look exactly like the one above. For now these are all the statistical measures that we need to proceed with our analysis. Clicking on the **Continue** button brings us back to the **Frequencies** window but now, after having selected the statistics that we need, we press **OK** and get the following:

Frequencies

Statistics

SCORE

N	Valid	30
	Missing	0
Mean		-.03
Median		.00
Std. Deviation		1.564
Minimum		-3
Maximum		4

As you read through these various steps to calculate the arithmetic mean and a few other statistics it may all seem tedious and terribly slow, but as soon as you follow these steps on your own PC you will see how lightning fast it really is.

For now, here are the most important things you need to know about this information:

The **Mean** of -0.03 is the sum of the 30 deviations divided by the sample size, N = 30. We shall use the mean as a measure of the general level of the data. A mean of deviations close to zero, as here, is consistent with the idea that the shots tend to average out on target, the first of the objectives mentioned above.

Std. Deviation (sometimes abbreviated further as SD) -- standard deviation -- of 1.564 is a useful measure of the variability or dispersion of the process -- of how much the player's shots tended to scatter around their general average. The smaller the standard deviation, the better, so long as the mean is close to zero. (Later in this chapter, we shall see how the standard deviation can be interpreted as a measure of dispersion.)

Brief explanations of some of the other output: (1) **Minimum** is the smallest of the 30 observations, -3. (2) **Maximum** is the largest of the 30 observations, +4. (3) **Median**, roughly, is a number that separates the 15 highest observations from the 15 lowest observations, sometimes called the **50th percentile** of the distribution.

4. The Control Chart

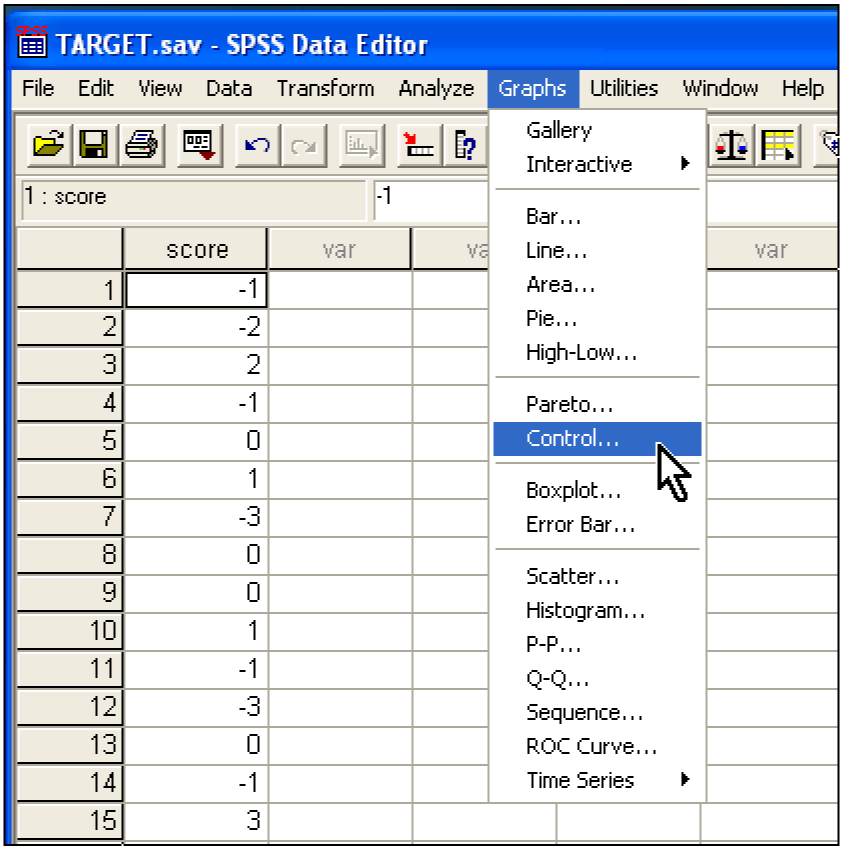
The next step of analysis will be the plotting of a **control chart.**

The heart of a control chart is a plot of the observations in time sequence as points on a graph called a **run chart**: time sequence is plotted horizontally and the values of the observations are plotted vertically. To make a run chart into a control chart, we need:

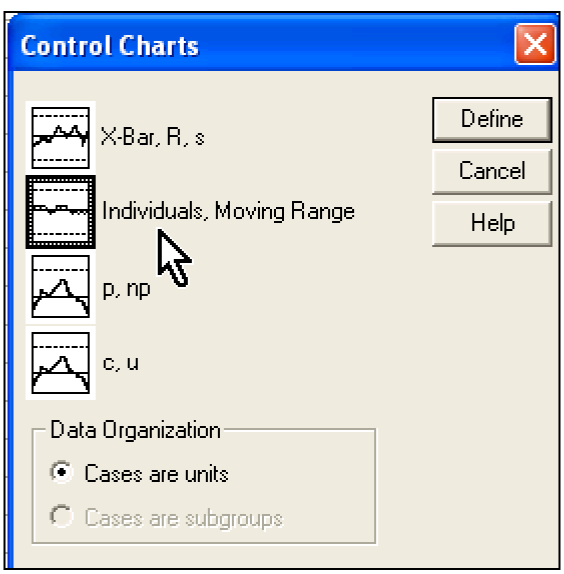
A horizontal centerline, usually placed at the height of the mean;

horizontal lines called **upper and lower control limits,** usually lying above and below all or almost all the data points.

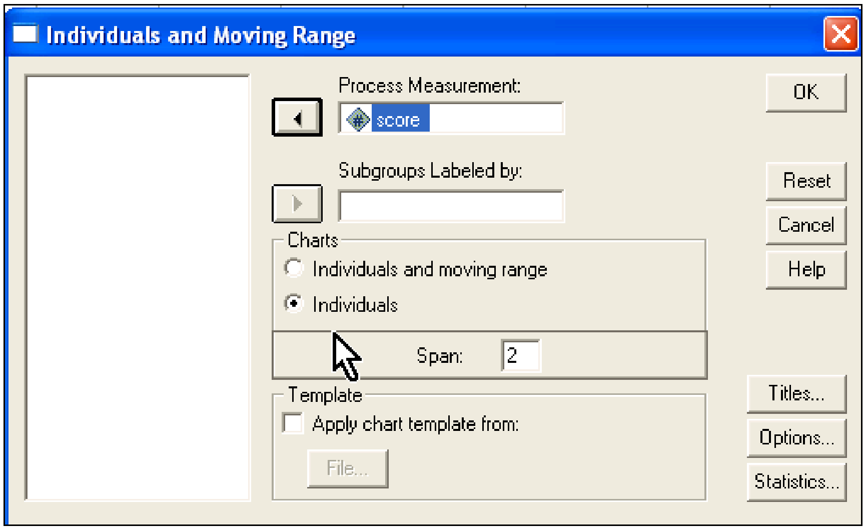
With ***SPSS*** we click **Graphs** on the menu bar and then **Control…** in the menu that drops down:



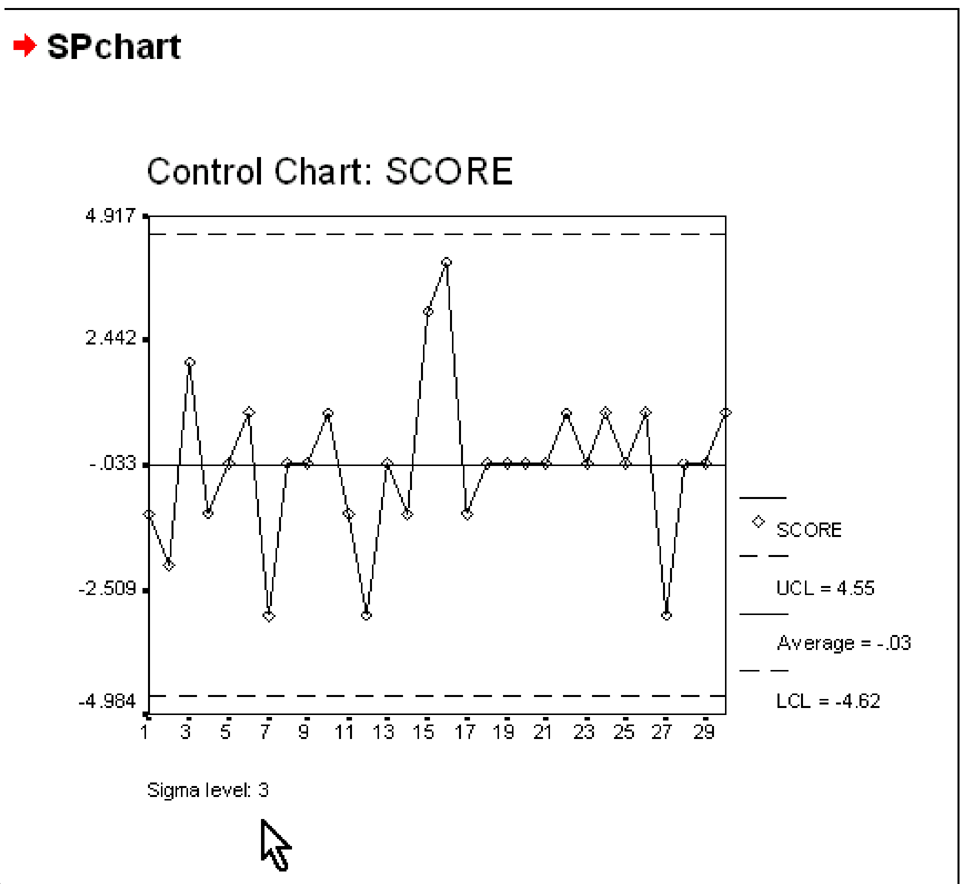
In the window that appears we select **Individuals, Moving Range** and hit the **Define** button**:**



Another window opens in which we move the variable name **score** into the **Process Measurement** space, and we also click in the little circle next to **Individuals** under **Charts**. Everything else in the window is left as it was originally shown.



The **OK** button produces the following high-resolution graph:



:

The control chart above plots the 30 observed scores of the target game in the exact sequence in which they occurred. We see, for example, that 12 of the observations, the zeros, are plotted just a hair above the center line drawn at the mean score of -.03.[3] The upper control limit (the broken line labeled UCL) is approximately at a value above the center line equal to three times the standard deviation, 1.564. This is indicated by the note "sigma level 3" at the bottom of the chart, "sigma" being a special statistical word meaning "standard deviation". We say "approximately" because this chart computes the UCL according to the "moving range method", a common technique used in real-world industrial applications of control charts that gives a result that is slightly different from that of the usual computation. Similarly, the lower control limit (LCL) is approximately three standard deviations below the mean. If you carefully plot two lines that are exactly three standard deviations from the center line you will see that the discrepancies between those lines and the control limits shown in the chart are hardly perceptible.

5. Statistical Control and Visual Checks

We now turn to visual analysis of the control chart for the 30 scores. Visual analysis of a graph is usually a good way to begin data analysis. A physician/statistician named Joseph Berkson coined the humorous expression "interocular traumatic test (ITT)" for visual analysis of graphs. This was to suggest that the graph of the data may convey such a clear visual message that it "hits you between the eyes".

No matter how much you may come to learn about the numerical computations used in data analysis, never forget the merits of the ITT! Among other things, the ITT can alert you to important information that could be missed by numerical computations alone.

Here are salient visual features of our control chart:

As you look across the plot from left to right, the points are irregularly scattered above and below the center line at the mean height. The **trend** of the data is horizontal, neither tending to rise or to fall. We shall refer to this phenomenon as **constant level.**

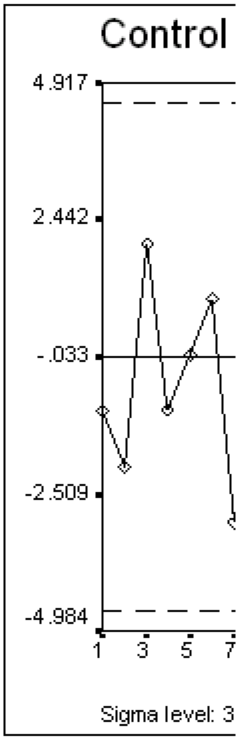
As you look across the plot from left to right, the deviations of points above and below the center line are erratic but they are all comfortably between the upper and lower control limits. The average variability of deviations is thus about the same throughout the plot. We shall refer to this behavior as **constant variance**.

As you look across the plot from left to right, the **direction** and **magnitude** of each deviation about the center line are apparently **unpredictable from examination of the sequential pattern of the deviations that precede it.**

One way to see this final point, unpredictability of sequence, is to put a piece of paper over the plot and to pull the paper slowly to the right, revealing one point at a time. Ask if you

[3] The word "average" is just another term for "mean." Note that the mean of this set of data does not have to be an integer. In this case it is the rounded value of a never ending decimal figure, -0.03333…

can devise a system to predict the next point. For example, suppose that in the graph above you cannot see any of the points to the right of the first seven points in the plot.



You might, for example, conjecture that there is a kind of up-down cyclical pattern in the first seven points. The seventh point looks like a "trough", so you might reason that the eighth point will be a little higher but not up to the center line. It turns out, that the eighth point is right at the center line:

If you continue this process of "unfolding" the data, you will conclude that there is no obvious, certainly no simple, pattern to the direction and magnitude of the deviations. We shall refer to this behavior as **unpredictability of sequence of deviations.**

In summary, we have explained three important visual aspects of the behavior of the data:

The deviations vary about a constant level through time.

The variation of the deviations about the constant level is also about the same through time.

The direction and magnitude of the next deviation is not predictable from the sequence of deviations observed up to the present.

These three visual aspects of the data are the intuitive essence of what is called a **state of statistical control.**

Often people say simply that an in-control process is "stable" as opposed to "chaotic". A better statistical term is "random", and we shall sometimes use this.[4] Whatever the word we choose to use, the idea of statistical control is fundamental to our understanding of data analysis -- and of the statistical theory underlying data analysis.

If a process is in a state of statistical control, the individual deviations above and below the general level elude **exact** predictability. But there is a sense in which we **can** predict: we can predict the **overall pattern** of the data. Specifically, if the process stays in a state of statistical control, the average level of the data will continue at the same general level observed in the past, and **almost all future deviations above and below the general level will fall below the upper control limit (UCL) and above the lower control limit (LCL)**.

6. Runs Check for Statistical Control

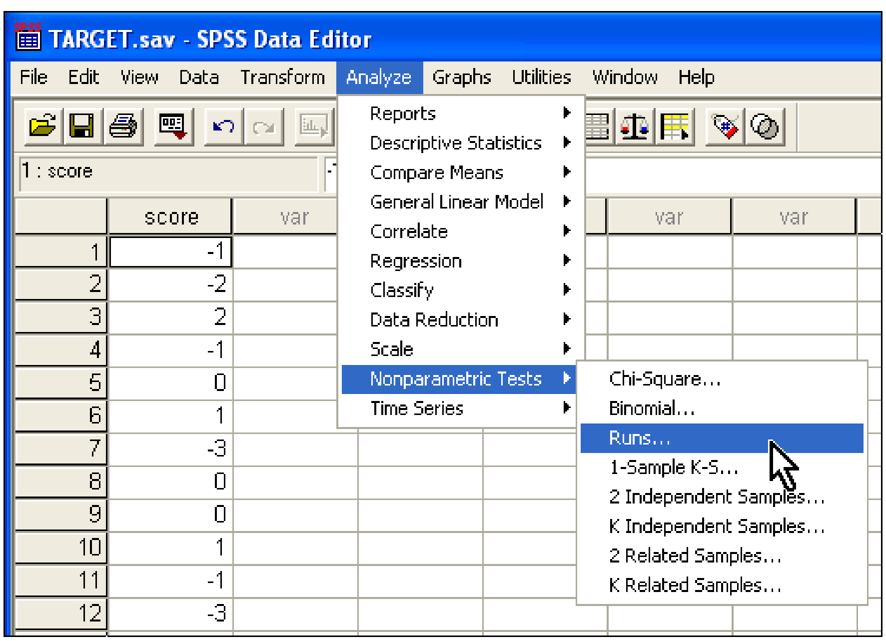
How do we judge whether a process such as the target game is indeed in a state of statistical control? Intuitive visual analysis, as explained above, is always useful; it should come early in data analysis and should never be neglected. But it is nice to have additional guidance from numerical computations suggested by statistical theory. The guide that we will use is called a **runs check** or **runs test**.

A **run** can be defined as a consecutive string of deviations of points either all above the mean level (or some other number) or all below it. In the target example, computation of runs is shown in the ***SPSS*** Data Editor below, which you should follow closely. A deviation above the mean is assigned a plus sign and a deviation below the mean is assigned a minus sign. In this particular example, the mean value is -0.0333. To count runs, we mark off strings of consecutive plus signs and strings of consecutive minus signs. The descriptions of the end points and the run lengths, shown in the second column, were typed in by us, not the ***SPSS*** routine.

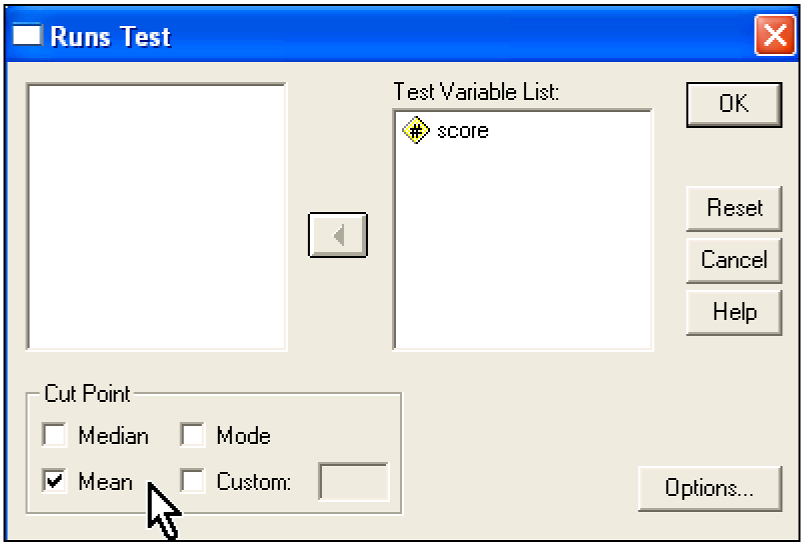
[4] In more technical statistical language the precise term is "independent and identically distributed".

score	runmark
-1	
-2	End of first run, comprising two minuses.
2	End of second run, comprising one plus.
-1	End of third run, comprising one minus.
0	
1	End of fourth run, comprising two pluses.
-3	End of fifth run, comprising one minus.
0	
0	
1	End of sixth run, three pluses.
-1	
-3	End of seventh run, two minuses.
0	End of eighth run, one plus.
-1	End of ninth run, one minus.
3	
4	End of tenth run, two pluses.
-1	End of eleventh run, one minus.
0	
0	
0	
0	
1	
0	
1	
0	
1	End of twelfth run, nine pluses.
-3	End of thirteenth run, one minus.
0	
0	
1	End of fourteenth run, three pluses.

We start off the runs test via the ***SPSS*** sequence **Analyze/Nonparametric Tests/ Runs:**



This opens the following dialog window:



The default cut point in this procedure is the median, but we have chosen the mean to agree with the center line of the control chart. In large samples there will be little difference between the results of the test whether we use median or mean so we will always use the mean.

This is the output that is produced by clicking on the **OK** button:

Runs Test

	SCORE
Test Value[a]	-.03
Cases < Test Value	9
Cases >= Test Value	21
Total Cases	30
Number of Runs	14
Z	.000
Asymp. Sig. (2-tailed)	1.000

a. Mean

We see that there are 14 runs in all. For guidance in making a judgment as to whether the process is in a state of statistical control, we need to know whether 14 runs would be a relatively likely outcome **if the process were in a state of statistical control**. The statistical reasoning goes as follows:

Expected Number of Runs

There were 21 values of **score** that were above the mean, and 9 below. Suppose that we had 30 cards-- 21 marked plus and 9 marked minus-- and that we shuffled the cards thoroughly and counted the number of runs in the shuffled deck, going from, say, top to bottom. Suppose we count 7 runs.

Then we shuffle the deck thoroughly again, and again count the number of runs, this time, say, getting 15 runs.

Imagine this shuffling and counting of runs to be repeated a very large number of times. By recording the number of runs that occurs after each shuffle and noting the total number of such shuffles we can calculate the mean number of runs that occurs in the large number of repetitions. We will call this mean or average value "**the expected number of runs**".

Fortunately, we don't actually have to keep shuffling cards: the expected number of runs can be calculated from probability theory. We will rarely show formulas in this textbook, but in case you are curious it is

[2 (Number of pluses)(Number of minuses)/ Number of observations)] + 1

which in the present example is [2(21)(9)/30] + 1 = 13.6.

Comparison of Actual and Expected

The actual number of runs was 14, and this is close to 13.6. Hence, since the actually observed number is close to what would be expected under the assumption of statistical control, **the runs check is consistent with the assumption that the process is in statistical control.**

"Significance"

Next question: How much would the actual number of runs have to differ from the expected number (either greater or less) to suggest that the assumption of statistical control is seriously challenged by a runs count? To help answer this, we look at the line at the bottom of the runs output above that shows **Asymp. Sig. (2-tailed) 1.000.**

This means that, **if the process is in a state of statistical control,** the chance of getting 14 runs or more, or 13 runs or fewer, is 1.000. This leads to the conclusion that with a process that is in control, having a number of runs that is right on the expected number **or deviates from it by a greater amount** is a sure thing. Therefore the runs check gives us no reason to challenge the assumption that the process is in a state of statistical control.

Suppose, on the other hand, that we obtained a much larger absolute discrepancy between actual and expected and that the "significance" number had been 0.02. Two times in 100 does seem rather unusual, so we would then have substantial reason to doubt the assumption that the process is in a state of statistical control.

We call the probability of getting the observed number of runs or a more extreme number (above or below the expected number of runs) the **significance level** of the test. In this case the two-tailed statistical significance of 14 runs when the process is in control is 1.0000-- not considered significant at all.

The 0.05 Rule of Thumb

How do we draw the line between "relatively usual" and "relatively unusual"? A common rule of thumb is 0.05 -- five times in 100 or 1 time in 20 or five percent. Thus:

If the "significance" number turns out to be greater than 0.05, we say that the runs check is consistent with the assumption of statistical control.

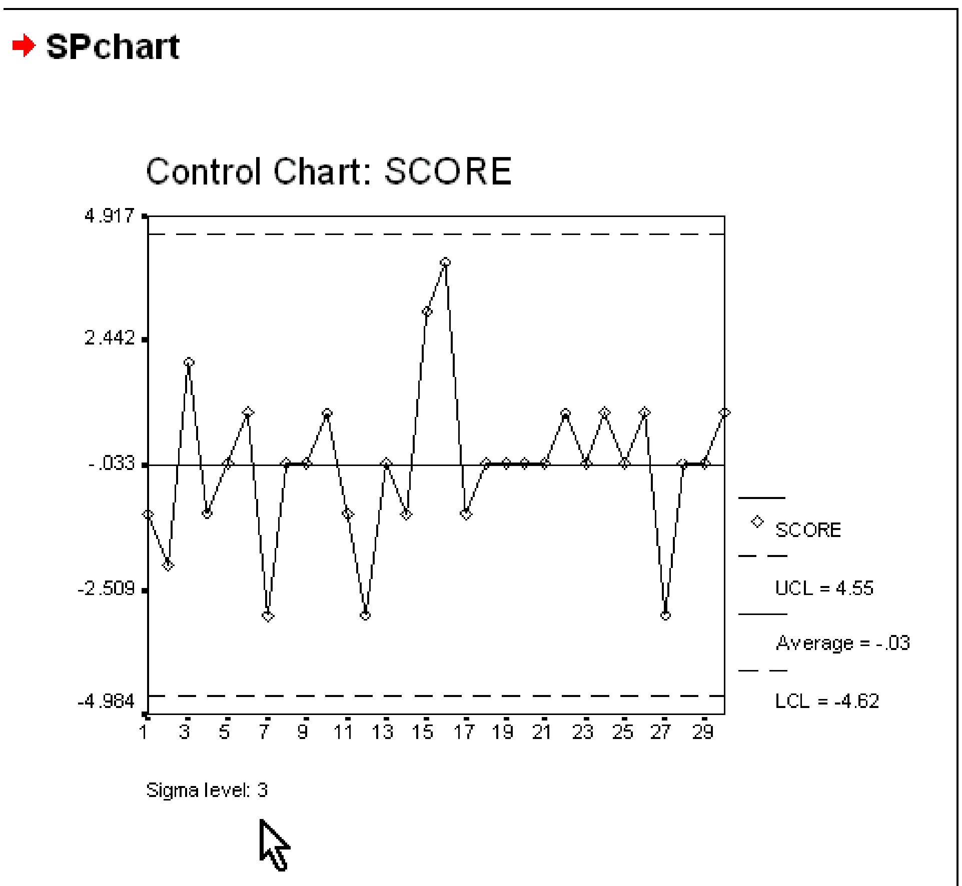
If the number is less than 0.05, we say that the runs check is inconsistent with that assumption, suggesting that the underlying process is **not** in statistical control. There is nothing sacred about the five-percent rule of thumb, and we are free to deviate from it if we wish, but it serves as a rough guide and is widely used in practice.

A synonym for "significance" is **"p-value"**. We shall often use this term.

Remember this: A large deviation in either direction of the actual number of runs from the expected number will result in a p-value (significance level) that is small. The **smaller** the p-value, the **less consistent** the data are with a process that is in statistical control.

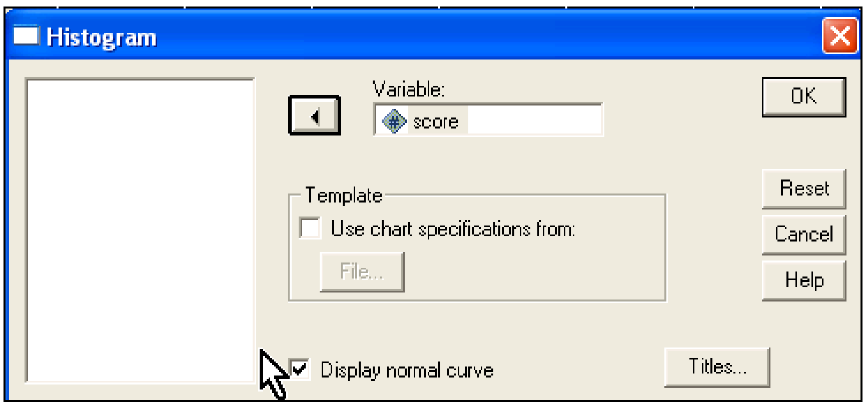
7. **Distribution of the Data**

In our visual data analysis, we omitted a salient feature of the data, one **that has nothing to do with the sequence in which the data arose**. Look again at the control chart:

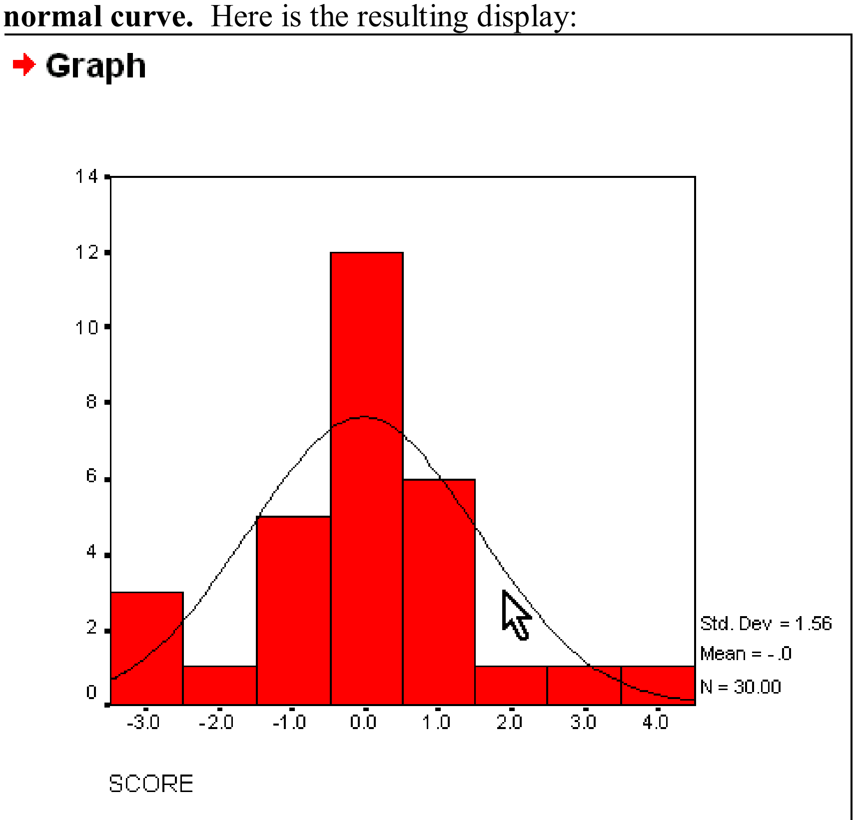


Note that points close to the center line are more frequent than points substantially above or below it. What we have discovered here is a pattern in the data, a pattern related not to the **sequence** of the data but to the **frequency** with which scores of different magnitudes occur. A simple way to display this pattern is shown next. Imagine taking a **squeegee** (the device used for cleaning windows) and holding it to the right of the control chart so that the rubber blade is exactly parallel to the right-hand border of the figure and ready to be pushed to the left. Now, in your imagination, move the squeegee steadily to the left, pushing the data points toward the left-hand border, but without changing their positions above or below the center line. When all of the points have been pushed to the left border, figuratively, they will be collected in stacks located at the appropriate position on the vertical axis of the chart. The number of points in each stack is the frequency for that value on the vertical axis, i.e., the value of **score**.

In ***SPSS*** we perform **Graphics\Histogram** and get the following window:



After the panel appeared we inserted the variable, **score**, and we also checked the box, **Display normal curve.** Here is the resulting display:



A histogram is really nothing more than a bar chart. In terms of our imaginary squeegee applied to the control chart, the stacks of accumulated data points on the left-hand vertical axis have been rotated so that they extend upward, and the stacks have been replaced by bars. The height of each bar indicates the frequency of points at that value of **score**.

You can verify that the sum of the heights of the bars is the total sample size, 30.

The shape of the histogram may bring to mind a concept that most readers have been exposed to: a **normal** or **bell-shaped** curve. The continuous curve in the display above helps in comparing the histogram with the normal distribution. We will get to the normal curve shortly.

An Important Caution

The computation and display of a histogram, as just shown, is valuable mainly when we have good reason to believe that the process being analyzed is in statistical control; that is, when there is no consistent pattern in the **sequence** of the data.

Note that when a histogram is produced (that is, by figuratively pushing the squeegee toward the vertical axis) all information concerning the sequence of observations is lost. In other words, looking at the histogram alone tells you nothing about whether the process is in control or not.

If the process is not in control, then the discussion of the histogram above, and of the normal distribution in the next section, is of little or no interest. For then, in practical data analysis, computation and display of a histogram is at best unnecessary (and therefore wasteful) and at worst misleading. If you find that the observations are not in control, your first priority is to try to find out in what respect they are out of control and, if you can, what are root causes of the departures from statistical control.

8. Normal Distribution

Look at the continuous curve that has been drawn by ***SPSS*** in the histogram above.

Think of this curve as showing the tops of the frequency bars for a very large sample from a process for which the normal curve provided a model for the pattern of variation, and for which the mean -0.03 and standard deviation 1.5643 were the same as we observed in our sample.

The curve is not an artistic fancy; it is described by a precise mathematical formula. (In a moment we shall discuss the underlying rationale for the formula.) What you need to know about the normal curve is summarized in the following statements:

The maximum height of the normal curve is found at the mean, i.e., -0.03.

The height of the curve declines as we move in either direction from the mean. For example, at the mean plus or minus one standard deviation, the height has dropped to about 60 percent of the maximum height.

The right half of the curve is the mirror image of the left half: the curve is "symmetrical" (not "skewed") about the mean.

In the interval that extends from the mean minus one standard deviation to the mean plus one standard deviation, about 2/3 (more precisely 68.27 percent) of the observations will be found.

In the interval running from the mean minus two standard deviations to the mean plus two standard deviations, about 95 percent (more precisely 95.45 percent) of the observations will be found.

In the interval from the mean minus three standard deviations to the mean plus three standard deviations, nearly all (more precisely, 99.73 percent) of the observations will be found.

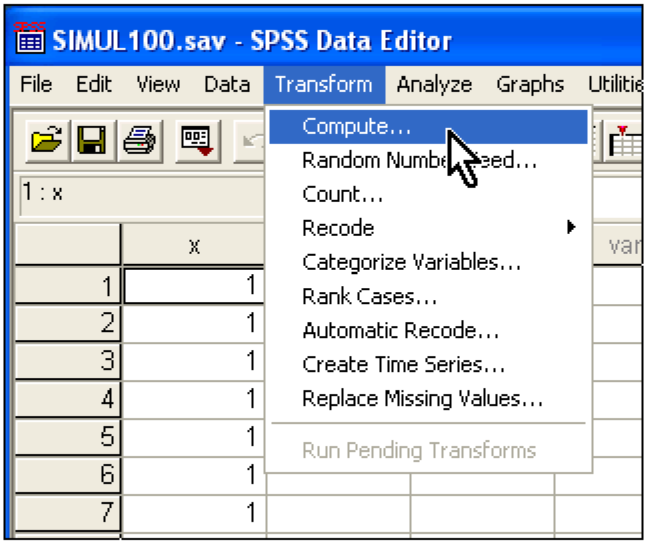
This last point relates to the computation of control limits. If a process is in statistical control **and** normally distributed, only about three points in 1000 will fall either above the upper control limit or below the lower control limit. When in an actual application, a point falls outside the control limits, there is thus some presumption that there has been some basic change in the process. We will develop this idea in Section 12 below.

In the next part of our discussion of the normal distribution it will be necessary to leave ***SPSS*** and to start a new session. Before we do that we must execute the sequence **File/Save** so that the name of the variable, **score,** is permanently stored in the data file TARGET.sav.

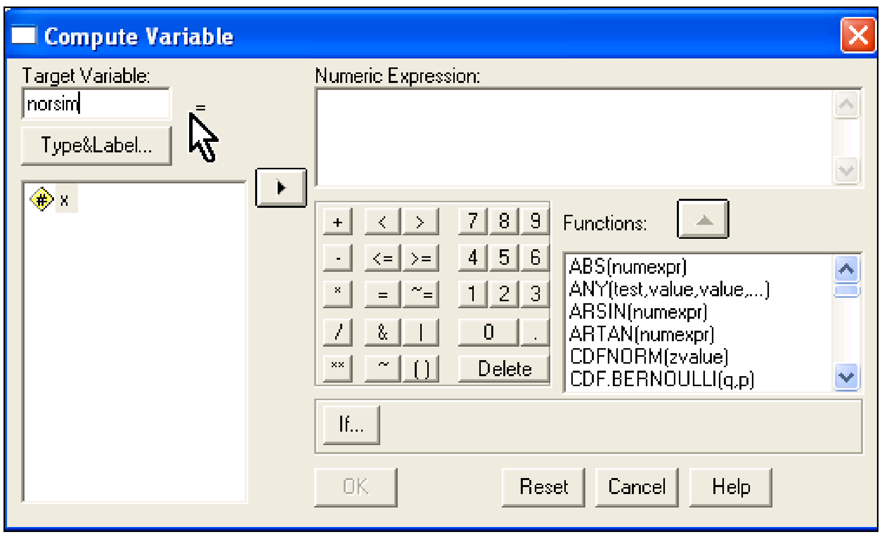
Next, we click on the sequence **File/Open** or, alternatively, the  icon, and we open the file called **SIMUL100.sav**. SIMUL100.sav is a file that was specially created to permit the generation of a set of random data that look just like the results of a particular process that we are interested in studying. We see that the file contains one variable, **x**, with 100 cases, each equal to the integer 1.

We are now going to **simulate** a sample of 100 observations from a normal distribution by using the transform procedures in ***SPSS.*** By "simulation" we mean creating data by a mathematical process, programmed into the software, that behave just as if they were observed empirically as compatible with a normal probability distribution. If we had continued working with the file TARGET.sav we would have been restricted to only 30 observations for our simulation, so we begin with the new file SIMUL100.sav. The variable**, x**, is in the file because ***SPSS*** does not allow the creation of a simulated set of data without data of the desired sample size already in the file. **x** just plays a kind of "dummy" role in helping us to get started.

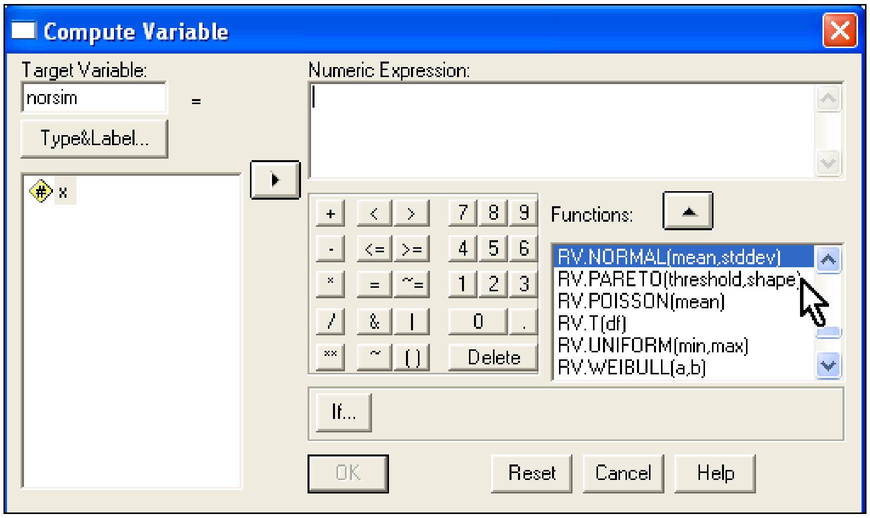
The first step, with the file SIMUL100.sav now open, is to execute the sequence **Transform/Compute…**:



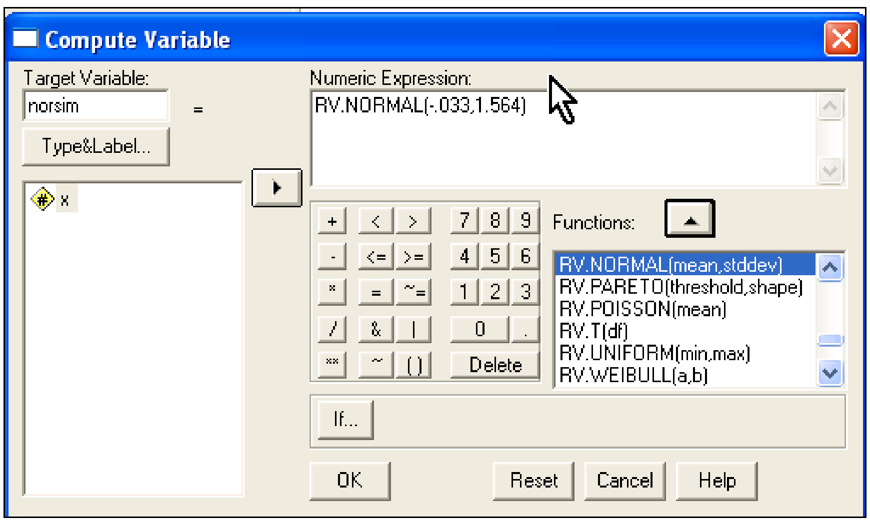
Clicking on **Compute** opens this window:

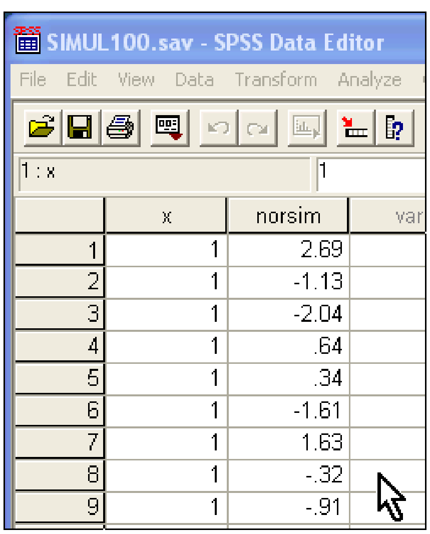


Note that after the window opened we typed in the name **norsim** for the target variable although we could have used any other name that was deemed appropriate. In so doing, we are beginning to state an equation that will cause an operation to be performed that will create a new variable. To complete the equation we have to put something into the blank area under **Numeric Expression:** As shown in the image below, we scroll down through the list under **Functions:** and highlight **RV.NORMAL(mean, stddev)**.



Finally, we click on the function arrow to move the desired expression into the blank dialog area and then we fill in the values for the mean and standard deviation that we observed in the data for score, -0.033, and 1.564.

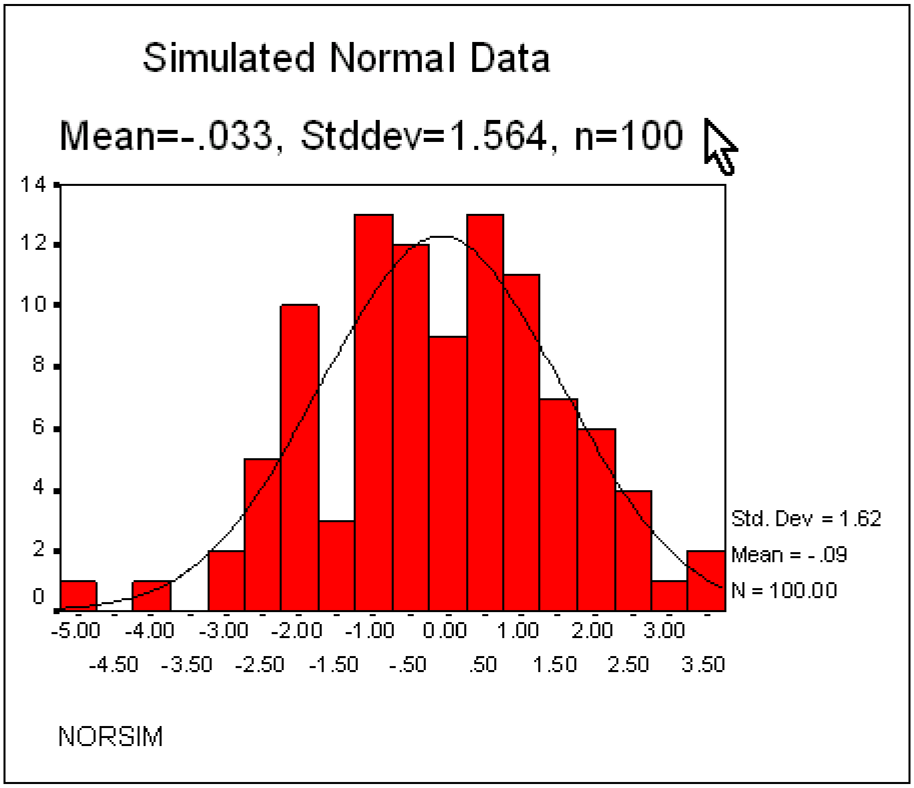




Clicking on the **OK** button produces a new variable, **norsim**, with random values for 100 cases. If you are following along with this discussion by performing the same actions on your personal computer you will not get exactly the same values that we have for **norsim**. In fact if we were to repeat the process we would not get the same values. This is because every time that a transform function beginning with **RV….** is executed the process producing the data starts with a random seed, i.e., a random number generated by the computer that ensures that the data are approximately truly random.

Now is probably a good time for you, especially if you find this discussion a bit overwhelming, to read Chapter 12, "Calculating New Data Values", in the ***SPSS 12.0 Brief Guide***.

We shall now examine the data for **norsim** by means of a histogram. The correct sequence is **Graphs/Histogram…**, and don't forget to check the box for **Display Normal Curve**:



This histogram, unlike our sample distribution of 30 observations, does not concentrate all observations at discrete integer values-- -4, -3, -2, -1, 0, 1, ... The normal distribution is based on the assumption that the data can be precisely measured to any number of decimal places. Thus in this application, the normal distribution is an approximation in two senses: (1) it abstracts from sampling variation (as it does in all applications); (2) it abstracts from discreteness effects when we rounded outcomes to the number of the strip on which the post-it pad landed. The latter approximation often, as here, is not a serious departure from realism. Had we wished, however, we could have avoided the need for it by making precise measures of the distance from a target line in the middle of the target strip.

Note that although we wanted to simulate a process in control with mean and standard deviation exactly the same as our data from the target game, the actual mean and standard deviation of the data that appeared are slightly different, -.09 and 1.62 respectively. This slight discrepancy from the desired values illustrates very nicely the uncertainly in a process that is random, i.e., in control. Later on we shall discuss methods that help us to decide if the process is really in control according to our desired specifications or if it is "out of whack."

Note also that in the display above we have used the **Titles…**button in the dialog box to insert a title that helps the viewer to relate the histogram to our present discussion.

Note well: Simulation is not really a data analytical tool. It is a **demonstration** tool. It only helps in data analysis by showing what data **should look like** under certain assumptions. For example, if we decide that the distribution of **score** looks sufficiently like that of the simulated normal variable **x** with the same mean and standard deviation, then we may conclude that **score** is approximately normally distributed. We make this cautionary statement because students, after reading this section of the chapter, have a tendency to include simulations in their homework assignments in a rather blind and robotic fashion. A simulation should only be run if you intend to use it for purposes of comparison with real data. You will soon see that we have methods for making decisions about the shape of data distributions without having to resort to simulation. Our purpose for discussing **simulated normal data** in this section is primarily to give those who are not so familiar with the concept of a bell-shaped curve a better "feel" for what normal data look like.

9. Why Does the Normal Distribution Sometimes Provide a Smoothed Description of Sample Histograms?

Why does this happen? One theoretical explanation is the following. If each observed measurement is the resultant of many small chance factors acting independently, it is a mathematical fact (Central Limit Theorem) that the normal distribution will give a good description of the histogram of our data.

As applied to bowling the Post-It pads, the "small chance factors": from toss to toss could be small variations in grip, arm motion, release time, aim, friction of the floor as the pad slides, air resistance, and details of technique.

We will see that an assumption like that of a normal distribution, if justified by the behavior of the data, can be very useful in practical applications. For example, we might like to assess the probability that the next toss will be no farther than two standard deviations (1.564 times 2= 3.13 strips) from the target. The normal distribution tells us that a rough answer is about 0.95 (see Section 8 above).

10. Checking for Normality of Distribution

How do we check for approximate normality of our histogram?

Recall, as an example, that for a normal distribution, the interval from the mean minus one standard deviation to the mean plus one standard deviation has about 2/3 (precisely, 68.27 percent) of the data. If normality is appropriate, about 2/3 of the sample data should fall within the corresponding interval defined by the sample mean and standard deviation. We would like to be able to check easily whether this actually happens.

If we use **xbar** to denote the sample mean and **s** to denote the sample standard deviation, we can describe this interval as extending from

xbar - 1s to **xbar + 1s**;

similarly, the interval plus or minus two standard deviations about the mean can be described as extending from

xbar - 2s to **xbar + 2s**.

For checking normality, it would be convenient to denote these two intervals simply by

-1 to +1

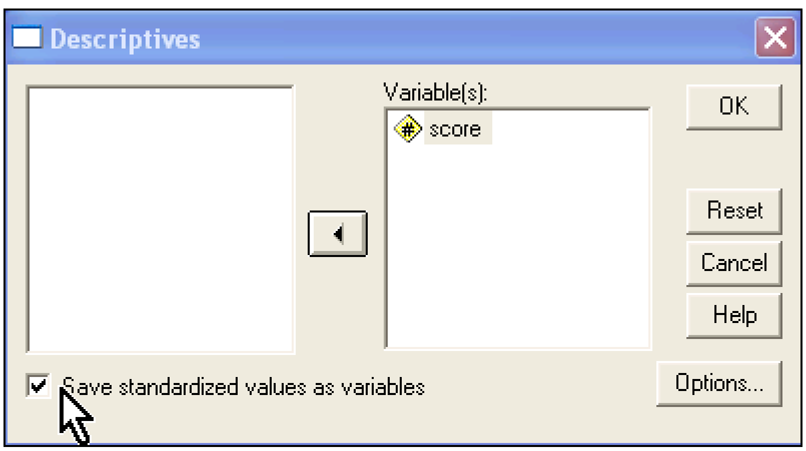
and

-2 to +2.

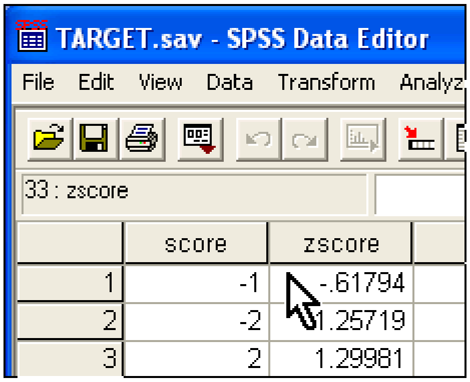
"+1" means "one standard deviation above the mean"; "-2" means "two standard deviations below the mean"; and so on. Such numbers are called "standardized" observations. Mathematically, if **x** represents an original observation, the corresponding standardized observation is:

z=(x-xbar)/s

In ***SPSS*** we have a convenient method that will standardize all of the observations of **score**. We simply perform the sequence **Analyze/Descriptive Statistics/Descriptives…** and when we see this dialog window



we check the little box in the lower left corner labeled **Save standardized values as variables**. After the **OK** button is clicked the usual descriptive table appears which we have already seen. Return, however, to the Data Editor and observe that a new variable, **zscore**, has been created.



You can verify with your pocket calculator, if you care to, that the values of **zscore** are indeed the properly standardized values of **score**. Applying the sequence **Analyze/Descriptive Statistics/ Descriptives…** to **zscore** produces the following table:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Zscore(SCORE)	30	-1.89645	2.57831	.0000000	1.00000000
Valid N (listwise)	30				

The mean is zero and the standard deviation one. By the very nature of the transformation shown above, the mean of a standardized variable **must be** zero and its standard deviation **must be** equal to one. The descriptive statistics also show that for this particular sample of 30 the minimum standardized value is -1.89645, i.e., 1.89645 standard deviations below the mean, and the maximum is +2.57831, or 2.57831 standard deviations above the mean. In other words, transforming a variable to its standardized **z-value** is equivalent to changing to a scale that expresses it in terms of the **signed distance in units of standard deviation from its mean**.

The next step is to apply the sequence **Analyze/Descriptive Statistics/ Frequencies…**to **zscore**, making sure that the **Display frequency tables** box is checked. Here is the result:

Zscore(SCORE)

		Frequency	Percent	Valid Percent	Cumulative Percent	
						xbar - 2s
Valid	-1.89645	3	10.0	10.0	10.0	
	-1.25719	1	3.3	3.3	13.3	
						xbar - 1s
	-.81794	5	16.7	16.7	30.0	
	.02131	12	40.0	40.0	70.0	
	.66056	6	20.0	20.0	90.0	
						xbar + 1s
	1.29981	1	3.3	3.3	93.3	
	1.93906	1	3.3	3.3	96.7	
						xbar + 2s
	2.57831	1	3.3	3.3	100.0	
	Total	30	100.0	100.0		

We have altered the table that was originally produced by ***SPSS*** by drawing lines at **xbar–1s** and **xbar+1s** as well as at **xbar-2s** and **xbar+2s**. Now it is easy to check how many observations fall within one, two, and three standard deviations from the mean. Study the above modification of the distribution. You should note that the counts listed in the column entitled **Frequency** are the same as the heights of the bars of the histogram shown earlier.

Look first at the central interval from **xbar–1s** to **xbar+1s**. You see that there are 5 + 12 + 6 = 23 observations, or 23 out of 30. This comes to 77 percent, a bit higher than 68.3 percent. This is larger than we would normally expect, but it reflects special feature of the target game that is explained in the footnote.[5]

For the interval from **xbar-2s** to **xbar+2s**, we add 3+1+1+1 = 6 to the 23 already counted, so we have 29/30 = 96.7 percent, only slightly larger than 95.5 percent.

Finally, one observation -- 1 in 30 or 3.3 percent is more than 2 standard deviations from the mean. This is not far from 4.5 percent.

Thus we see at a glance that the histogram corresponds roughly to what we would expect from the normal distribution.

A simple visual check of normality is also useful: does the histogram appear to peak in the middle and taper off symmetrically in either direction? If the distribution appears clearly not

---

[5] Our possible scores took only integer values like 2 or -1, whereas the normal curve is defined for continuous measurements. The "target strip" -- score of zero -- was counted whenever the Post-it pad landed inside it or touched the lines between it and the -1 or +1 strips. Effectively this made the target strip wider than the others. The 12 observations of zero exaggerate the player's shooting accomplishment and give slightly too many observations within one standard deviation of the mean. Had he taken the trouble to do so, the player could have measured precise deviations from a line target. To illustrate the idea of a lightning data set, he used the strip approach, even though it led to the need for this footnote.

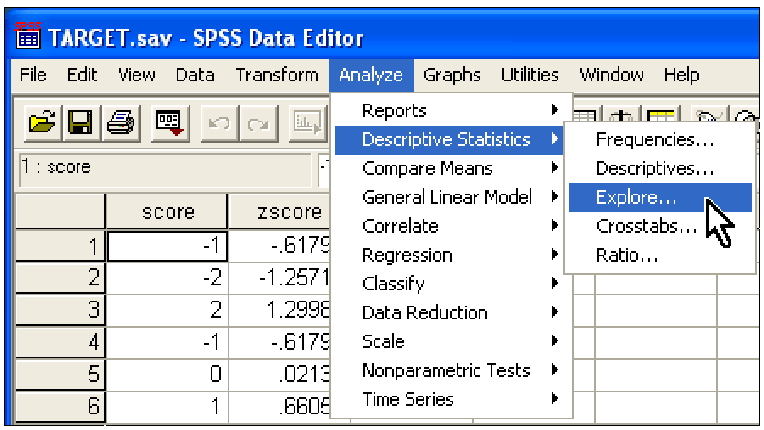
to be symmetrical, this is a visual strike against normality. There is no problem of asymmetry ("skewness") in the present example.

Finally, an observation outside of control limits should be taken seriously as an indication of a departure from the normal distribution. An observation far outside control limits should be taken **very seriously**. Again, there is no problem in the present example.

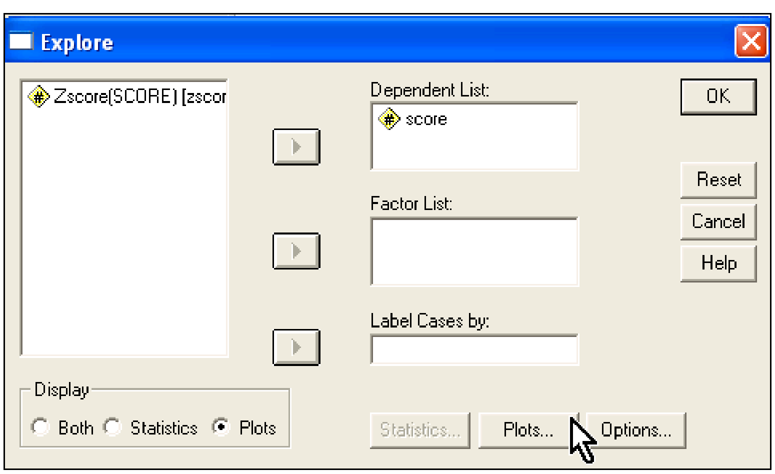
The quick checks just summarized will serve you well in getting the lay of the land for any histogram. Later, we shall consider what can be done when these checks point to trouble.

The Normal Probability Plot

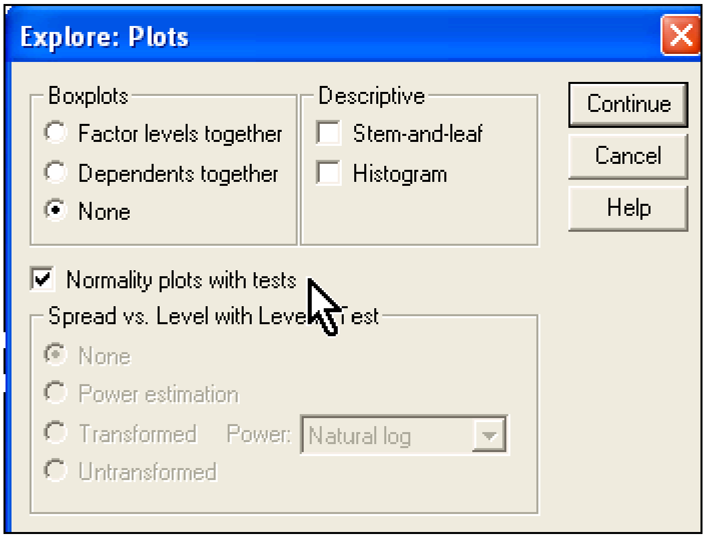
There is another procedure in ***SPSS*** that we can use for comparing a data distribution to that which would be expected under the normal probability model. We access it via the sequence **Analyze/Descriptive Statistics/Explore…:**



In the dialog window that appears there is a button labeled **Plots…** at the bottom toward the right:



We enter **score** into the **Dependent List** and check the **Plots** circle under **Display**. Then when the **Plots…** button is clicked we get the following :



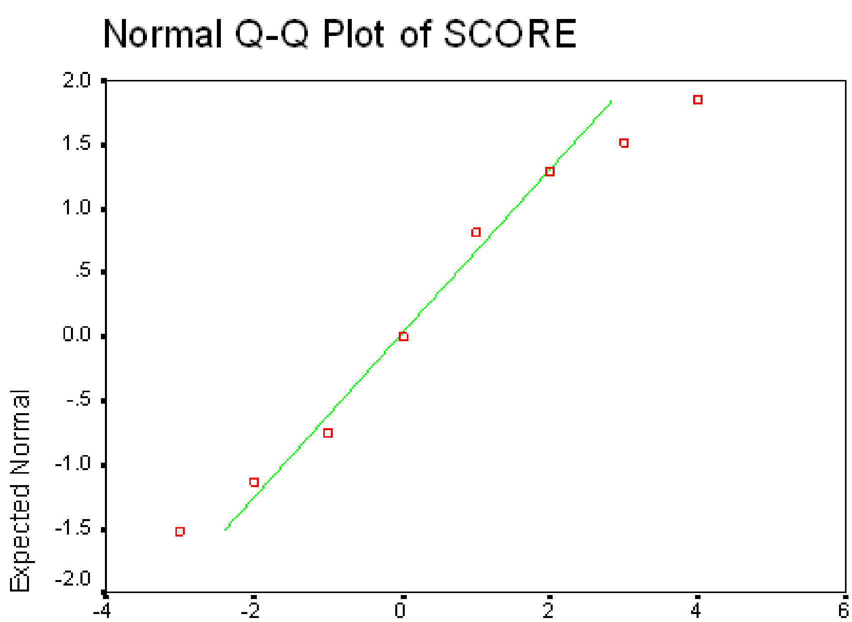
To minimize unnecessary output we have checked the **None** circle under **Boxplots** and we have left all boxes blank except for the one labeled **Normality plots with tests**. Clicking on the **Continue** button returns us to the **Explore** window where **OK** produces this output:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
SCORE	.209	30	.002	.917	30	.023

a. Lilliefors Significance Correction

SCORE



In the image above we have chosen to copy only the central part of the output in ***SPSS***. You see that there are two sets of test results—that for the Kolmogorov-Smirnov test and the Shapiro-Wilk test. The two tests should give equivalent results, but for this textbook we concentrate on the Shapiro-Wilk test. We see that the little squares "□" in the plot form a line that is fairly straight, indicating that the data may be approximately normal. With normally distributed data we also expect to see a Shapiro-Wilk statistic shown in the table of test results that is higher than 0.90. The present value, 0.917, is high, but apparently not high enough given the sample size, because the significance level, 0.023, is less than 0.05. Thus, although visual examination of the histogram and the frequency distribution did not indicate serious deviations from the expected results, **we must formally reject** the hypothesis that the data are normally distributed[6].

A careful explanation of how the Shapiro-Wilk plot works would be too complicated at this point in the textbook. Through a clever mathematical argument it accomplishes the same kind of comparison of actual frequencies with those expected under the normal model that we just performed by looking at the frequency distribution above. Suffice it to say that if the data are compatible with the normal model the plot should be approximately a straight line and the p-value for the Shapiro-Wilk statistic should be greater than 0.05.

To see a more pleasing result let's do the Q-Q plot for the simulation sample of 100 values of **norsim:**

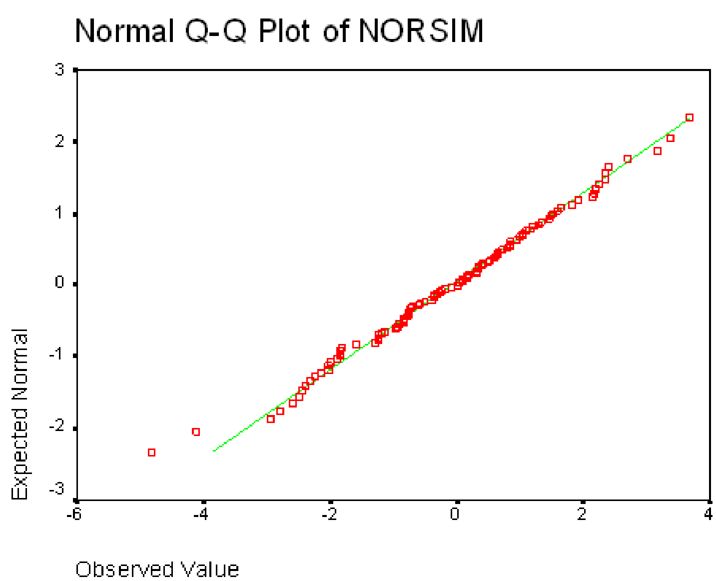
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
NORSIM	.047	100	.200*	.993	100	.909

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

NORSIM



The value of the Shapiro-Wilk statistic in this case is close to one, and the significance level is very large, 0.909. Remember, however that these data were **simulated** to be normal, i.e., created by a mathematical process that is designed to give a very good approximation to normality.

In spite of the low p-value in the plot for **score**, some statisticians might insist that the data ought to be treated as normal because of the general straightness of the plotted line and the fairly symmetric shape of the histogram. How straight must the line be to accept the normal model?

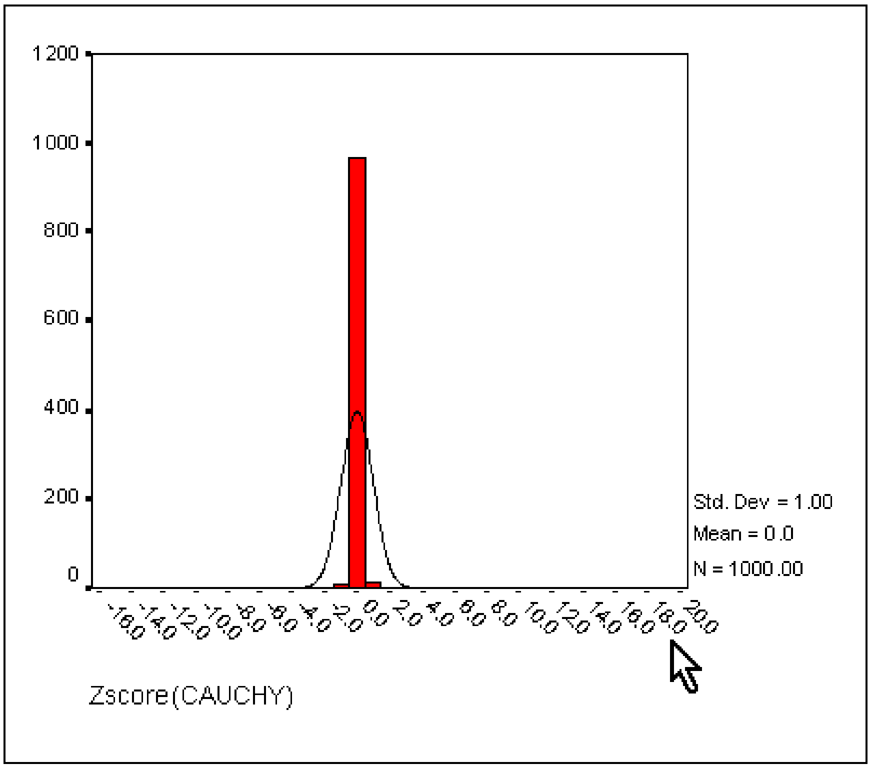
---

[6] Note that the plot was performed on the original variable **score**. If we had used **zscore** we would have obtained the same plot. Thus it is not necessary to standardize before using the Shapiro-Wilk (Q-Q) plot to check on normality.

We have already mentioned that a cutoff value for the p-value of 0.05 is just a rule of thumb, but until you acquire a lot of statistical experience we suggest that you follow it. You will soon learn, however, that in spite of what some people tell you about the "objectivity" of statistical methods, most decisions in data analysis are matters of subjective judgment, experience, and common sense.[7]

As you gain experience and confidence in judging normality, it helps to have in mind cases of data that are clearly **not normal** in the shapes of their distributions. Here is one example of non-normal data. As we move into further chapters we will see more such examples:

Using a certain mathematical formula, I have generated a set of 1,000 cases that simulate a particular random variable that is extremely non-normal. I call the resulting data **cauchy.**



A casual look at the histogram suggests the kind of symmetry that we find with normal data, but there are a few extreme standardized values, **z**, as low as -16 and as high as +20!!
It is those extreme values that cause the normal probability plot to be shaped like a stretched out letter S instead of a straight line:

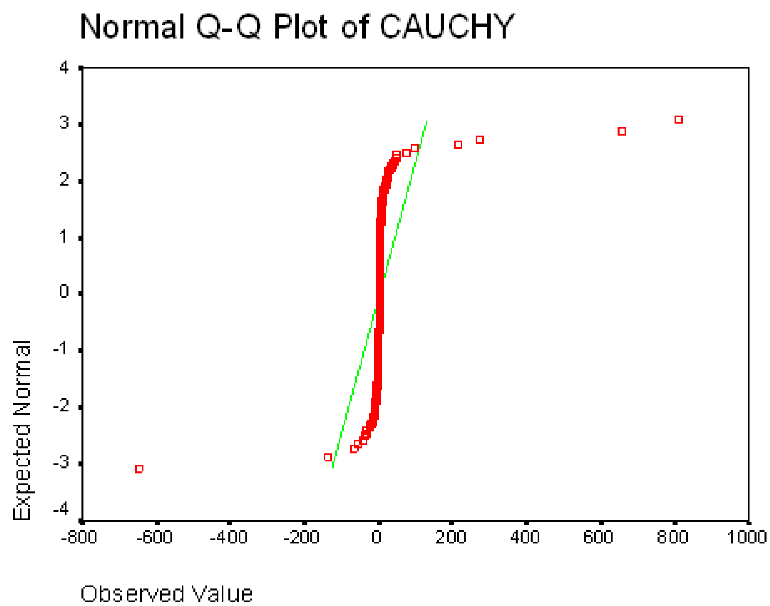
[7]The great French mathematician Laplace once defined probability as "...common sense reduced to calculation," suggesting an analogous definition of statistical analysis as **common sense disciplined by calculation.**

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
CAUCHY	.391	1000	.000	.113	1000	.000

a. Lilliefors Significance Correction

CAUCHY



Notice how low the value of the Shapiro-Wilk statistic is and how small the significance level (p-value). These data are extremely non-normal, but still random and in control. This just shows that not all data are necessarily normally distributed.

11. Statistical Control, Process Capability, and Tampering

The target game, has introduced three important ideas of data analysis:

- How to approach the analysis of time-ordered data, starting from run charts, which graph the data in time sequence, and then looking at control charts, which add a center line and control limits to the run chart. Use of control charts is central to that aspect of quality management called statistical process control or SPC.
- The concept of "statistical control".
- Use of the histogram as a tool for studying the behavior of data from a process that is in a state of statistical control.

In the simple target game, we reached the conclusion that the process appeared to be in a state of statistical control: the deviations from target appeared to vary about a constant level, with roughly constant variance about that level, and with unpredictable deviations from the level.

Further, we decided that the normal distribution gave a roughly satisfactory model for the histogram of deviations, which, in turn, suggested the possibility that only independent chance causes were at work in causing deviations from the target.

Process Capability

If the process is in control, the histogram provides a picture of the estimated process capability, that is, the pattern of results that the process would yield. Given a normal distribution, we can summarize our present process capability in two numbers:

> Our **mean score**, which was -0.033. This suggests that there is no evidence of systematic bias -- a tendency either to undershoot or overshoot the target strip. That is good.
>
> The **standard deviation**, which was 1.564, roughly a strip and one half. The smaller this number the better. The challenge to the player would be to reduce it, while keeping the mean score near zero.

Being in control with no evidence of systematic bias of aim and with a standard deviation of 1.564 doesn't mean that the process is satisfactory or that the player is skilled at the target game. If he were serious about the game -- say in the sense that many golf players are serious about golf -- he would be looking for ways of improving, possibly by practice alone, possibly by experimentation with improved techniques. (Many processes of practical interest can be viewed in the same way: for example, in parts manufacturing, we might be aiming at a dimensional specification, trying to average close to that specification and to minimize the deviations from it of individual parts.)

If we imagine continuation of play of the target game, without any basic change of the player's skill or any sudden changes in the circumstances of play, we would expect that future scores would vary about the center line of our control chart and, with rare exceptions, would stay within the control limits.

Thus to monitor the process, we could continue plotting points to see if future scores do in fact behave in this way. If a point falls outside control limits, this is a strong indication that the process may have changed. Investigation and possibly corrective action are in order; we shall discuss this point later.

But we should not attempt corrective action unless:

- We have a clear indication that there has been a basic change in the process.
- We have good reason to believe that the corrective action may in fact be helpful.

Tampering

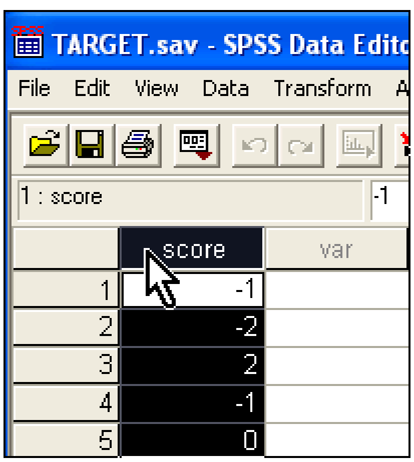
If the player simply made a guesswork change of technique whenever his toss did not land precisely on the target strip, he would be guilty of what W.E. Deming called **tampering**. Deming pointed out that if a process is in a state of statistical control to begin with, tampering is likely to make the process **worse**, not better.

A simple, but not unheard-of, form of tampering can be illustrated by a variation of the target game that was tried after finishing the game reported above. The idea is this. Suppose that one shot lands three strips beyond the target strip, and the player becomes alarmed that he is not aiming properly. To compensate, on the next try, he aims at the **third strip short** of the target strip. Similarly, suppose that on the next try he is **one strip short**, and that he adjusts his aim to the **first strip beyond** the target strip.

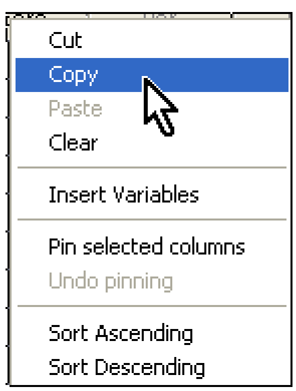
Implicit in these adjustments of aim is the idea that the deviations from target are entirely due to faulty aim, so that corrections of aim need to be made on every trial. Rather than aim consistently at the target strip as he did during the first play of the game, the player tampers with the process unless he lands right on the target strip.

If, in fact, the player's aim is faultless, but chance factors cause deviations that are in statistical control, his tampering will lead to a larger standard deviation of the scores, and thus to a poorer process.

The player followed this tampering policy in the second round of the target game. The scores for 30 tosses of the Post-it pad using this new technique are stored in the file called TARGTAMP.sav. We have named the new variable **scoretam**. Rather than open TARGTAMP.sav directly, we shall first reopen the file TARGET.sav containing the original variable, **score**.



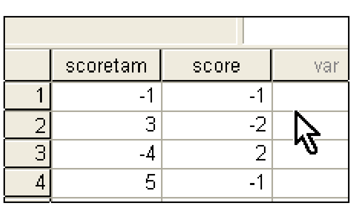
We then place the mouse pointer on the variable name, **score**, at the top of the first column and click with the right mouse button. This causes the column to be selected (blackened) and the following menu to appear:



We left click on **Copy** to write the data for **score** on the clipboard. We then open the file TARGTAMP.sav where we see the data for the new variable, **scoretam**.

Finally, we place the pointer in the name area of the second column and **Paste** in the cases for **score**.

Our **Data Editor** now looks like this:



	scoretam	score	var
1	-1	-1	
2	3	-2	
3	-4	2	
4	5	-1	

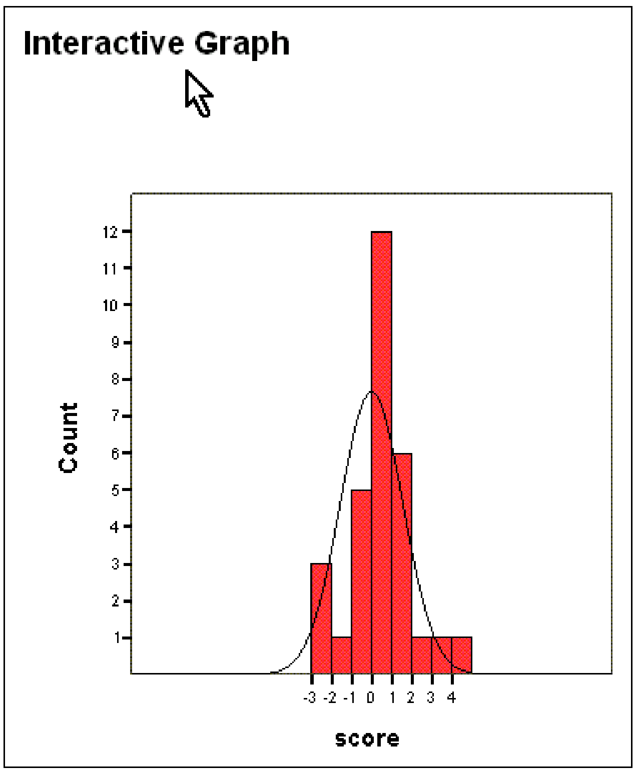
With the two variables, **scoretam** and **score** in the same Data Editor display we can now perform some graphical procedures to compare their distributions.

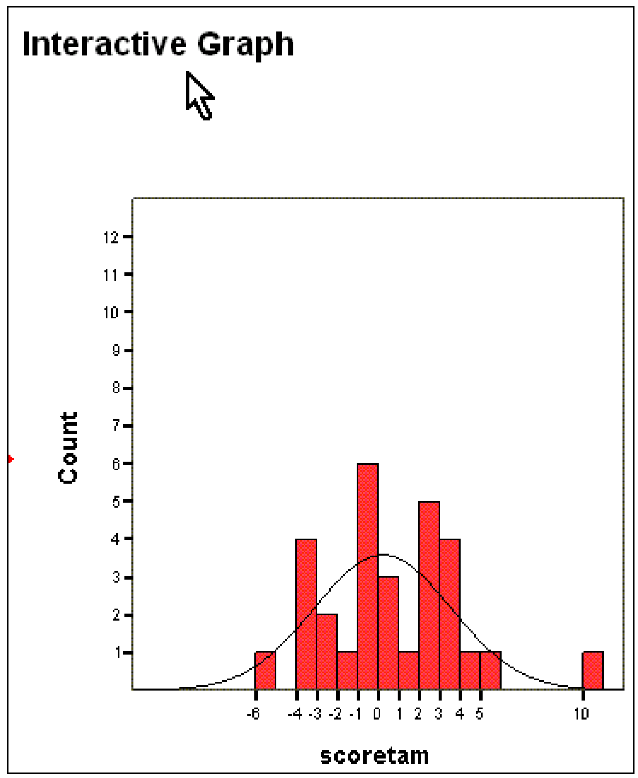
We begin with **Analyze/Descriptive Statistics/ Descriptives…:**

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
SCORETAM	30	-6	10	.20	3.347
SCORE	30	-3	4	-.03	1.564
Valid N (listwise)	30				

The mean for **scoretam** is 0.20 as opposed to -0.03 for **score**, so there is no strong evidence that tampering has made the player's average deviation different from the desired zero. But the standard deviation 3.347 for **scoretam** is substantially larger than the standard deviation 1.564 for **score**, so the shots are now scattering much more widely around the target strip. Note also that the minimum for **scoretam** is lower and the maximum higher, indicating a greater range for the new data. You can visualize the difference by comparing the histograms for the two variables.

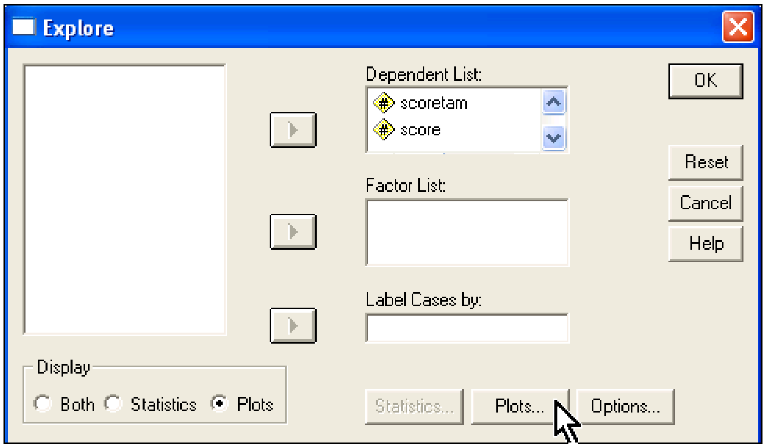




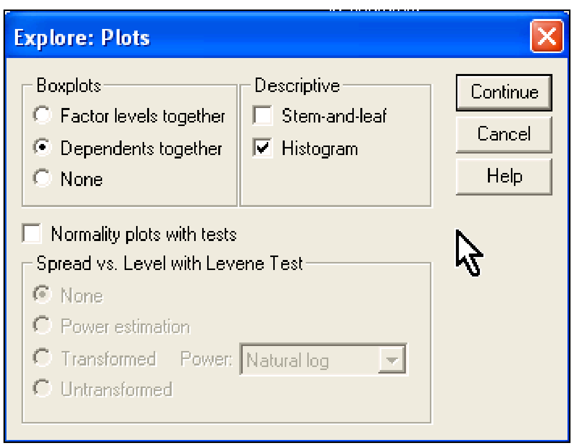
The two histograms above were each made with the sequence **Graphs/Interactive/Histogram…,** a method that allows the user to perform a lot of special tailoring of an image—first through settings on the **Histogram** and **Options** pages within the **Create Histogram** window, followed by further adjustment of the axes that is possible after double clicking on the image and bringing up the drawing tools. To be honest, getting the two histograms to look just right for comparison purposes required considerable trial and error, and we do not recommend that students even attempt to duplicate them unless they have a lot of free time. The documentation provided by ***SPSS*** is not really adequate to explain how to do it.

We see that tampering with the process makes the results more variable (i.e., more spread out), but also the extreme values of **scoretam**, -6 and especially the +10, appear to be separated by gaps from the rest of the data. We are tempted to call these possibly unusual data values "outliers."

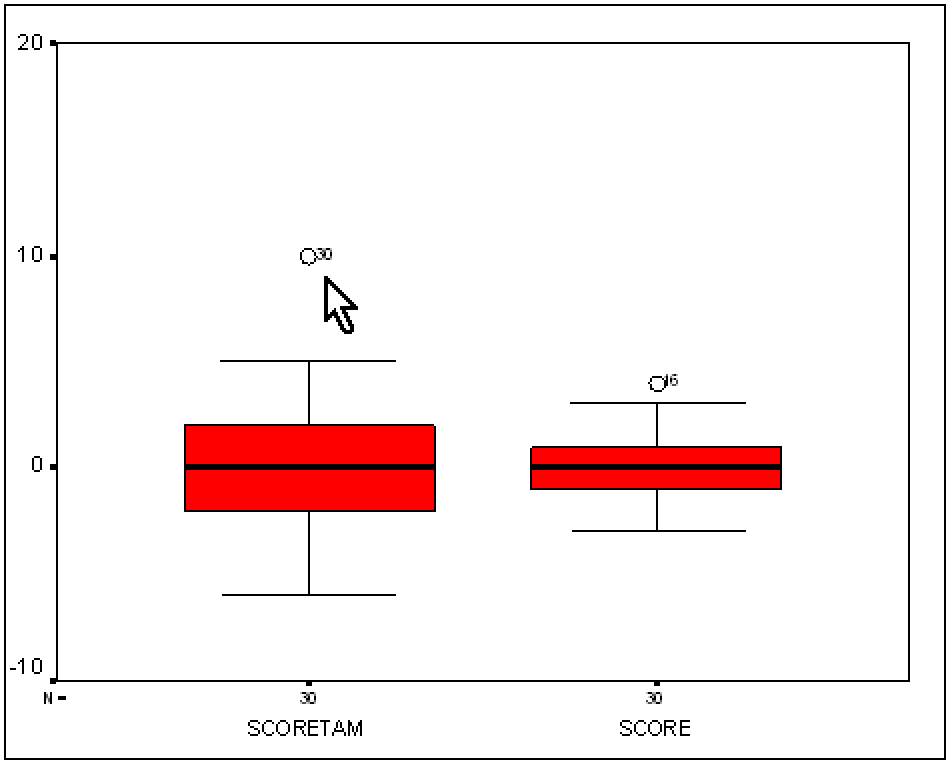
There is another technique available in ***SPSS*** for comparing two data sets that is easier for students to use than interactive graphing. First perform the sequence **Analyze/Descriptive Statistics/ Explore…** In the resulting dialog window we have dragged both variables into the **Dependent List** and marked the **Plots** circle under **Display** in the lower left-hand corner:



In the window below we have checked **Dependents together** under **Boxplots** and **Histogram** under **Descriptive**:

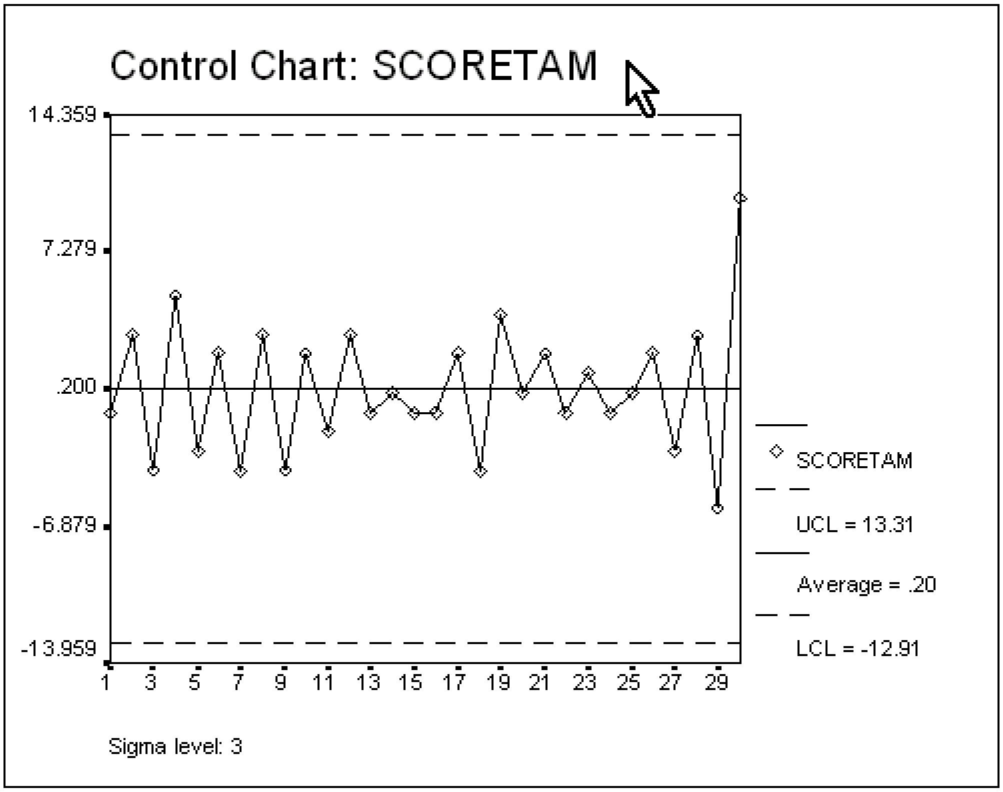


The **Continue** button returns us to the original dialog window where clicking on **OK** produces the desired display. When you do this on your own PC you will see that the resulting histograms are not very useful for comparing the two variables because their scales (over which you have no control) are different. The last graph, however, is a display of **boxplots**:



In the plot above we have lost all information about frequencies, but we can see in a single graphic the relative spreads of the data. The top of the box is the third quartile (the 75th percentile of the data) and the bottom is the first quartile (the 25th percentile); the heavier center line within the box is the median (the 50th percentile) thus the shaded box encloses the middle portion of the data for each variable. Some statisticians call these quartiles the "hinges" of the distribution. The vertical T-shaped lines are called "whiskers" and they indicate the range of typical data values--where "typical" means something like "not unusual". The little circle ○ denotes a data point that is a possible **outlier**. Notice that this nice graphic even tells you the case numbers of the outlying data points. We will say more about outliers later in this chapter.

There is another interesting consequence of the tampering: the tampered-with process is not in statistical control. Positive and negative deviations tend to alternate, and there are far too many runs. Study the following control chart to be sure that you understand what is happening:



Observe the up-and-down alternation about the center line. There are too many jumps above and below for the process to be in control. This is confirmed by the following result of **Analyze/Nonparametric tests/ Runs…:**

Runs Test

	SCORETAM
Test Value[a]	.20
Cases < Test Value	17
Cases >= Test Value	13
Total Cases	30
Number of Runs	26
Z	3.697
Asymp. Sig. (2-tailed)	.000

a. Mean

With the formula that we showed earlier, you can verify that if the process is in a state of statistical control the expected number of runs is 15.7. Considering all of the possible numbers of runs that could occur in random shuffling of the data, is the observed number, 26, so large that we doubt randomness?

SPSS indicates that the two-tailed significance level of the test is 0.000. It is not really zero, but since the number is rounded, its exact value must be less than 0.0005. Thus we can say that the probability is less than 0.0005 that 26 runs or a more extreme result in either direction from the expected value of 15.7 could occur **if the process is random. We therefore reject the hypothesis that the process is in control (random).**[8] There is an appendix to this chapter that discusses the runs test in further detail and which should help you to understand better the meaning of the significance level (p-value).

The lesson of this second target game is this: if a process is in control, don't tamper with it by trying to compensate for individual deviations above or below the center line. So long as the points stay between the control limits (and our other criteria for statistical control are met), benign neglect is the best policy unless you have some good insight into how the process works and how it might be improved.

In the meanwhile, continued plotting of points on the control chart will permit you to check that the process continues in a state of statistical control. This is the **monitoring** or **surveillance** function of statistical process control.

The type of tampering illustrated in the second target game may seem extreme, but it happens in practice. The senior author of this book, as a young soldier in 1944, fell into this trap as he tried to zero-in the sights on his M1 rifle. After each firing, he adjusted the sights to exactly compensate for the deviation of the previous round from the bullseye of the target. His assumption was that if he fired very carefully from the prone position, which is extremely steady, the only source of deviation from the bullseye would be the adjustment of the sights. After an entire afternoon of firing, he was no closer to being properly zeroed-in than he had been at the start. The successive shots wandered back and forth across the bullseye in much the same way as the longs and shorts from the target strip in the second target game.

A Common and Disastrous Form of Tampering

A more common type of tampering could be demonstrated by a variation of the target game:

- Keep the aim point constant until a miss of three or more strips is recorded.
- If a short of 3 or more strips occurs, lengthen the aim point, by the full amount of the miss.
- If a long of 3 or more strips occurs, shorten the aim point by the full amount of the miss.

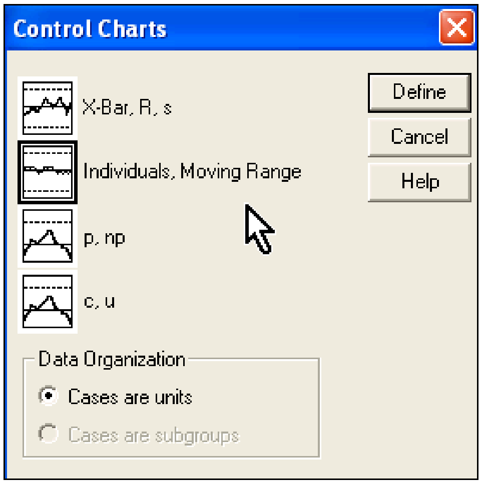
You might wish to try this out. Here, however, we shall present a practical business application, contained in a data file called CANTWELL.sav. A company president asked an operator to obtain internal diameter measurements on successive bushings produced by a grinder. The tolerance specifications were from 1.536 to 1.538, and there had been problems with excess variability when

---

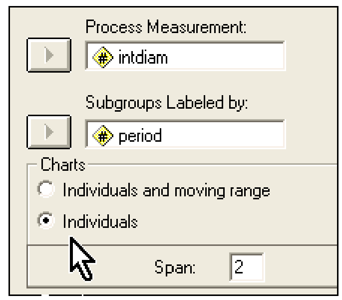
[8] There is no law that states that you **must** reject the hypothesis of randomness; but if you do not do so when confronted with the evidence of 26 runs (when only 15.7 are expected), then you are saying, in effect, "I know that this is a very rare event, but I am not yet ready to doubt that the process is in control." You may have very compelling reasons for taking such a stand in favor of randomness, especially if you have a lot of past experience with the process, but you must be prepared to be considered a radical by others who see the same test results.

the bushing was assembled into a truck engine. The operator was instructed to run off 40 consecutive measurements, and to note any machine adjustments made. Review of the first 40 readings showed that the operator was making frequent adjustments in the attempt to keep the process within the tolerance specifications. He was asked to take 40 more readings with minimal adjustments. After the 47th reading, no adjustments at all were made.

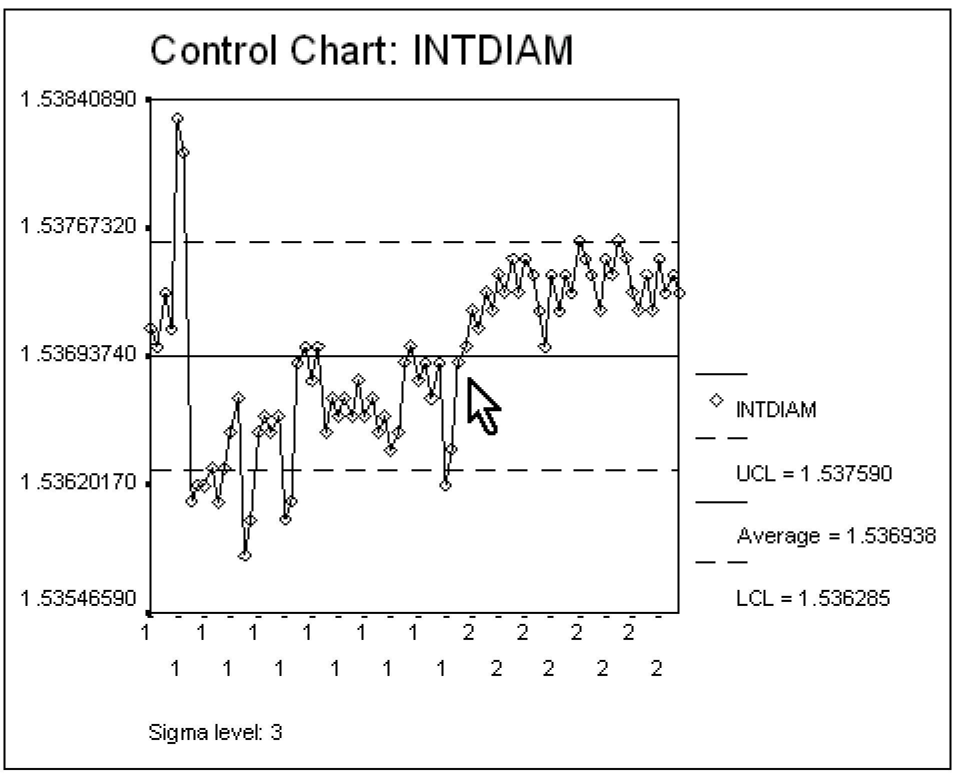
Using the ***SPSS*** command sequence **Graphs/Control…**, we need to make a run chart of the data. The first dialog window gives us some choices. We highlight **Individuals, Moving Range:**



Note that the file CANTWELL.sav contains a second variable called **period** that has integer values, 1 or 2, indicating whether the case is among the first 47 or after when readjustment stopped. Before clicking on **OK**, the interior section of the next dialog box should look like this:

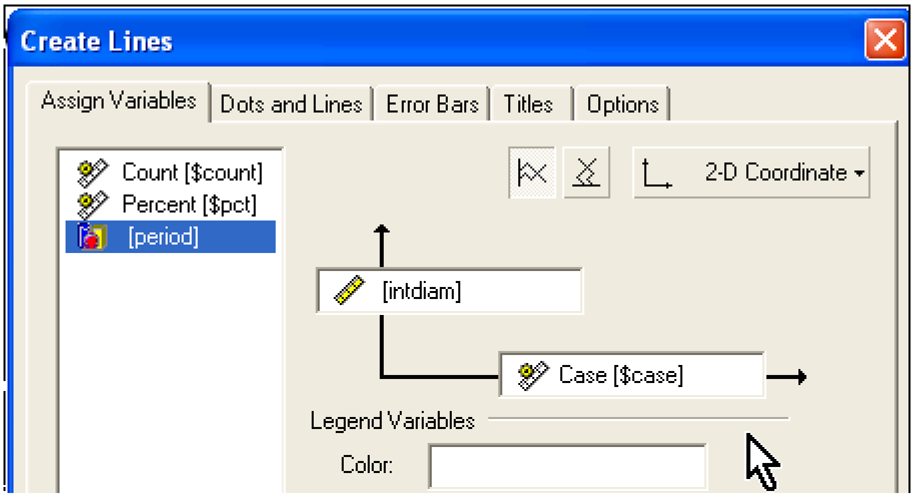


Here is the resulting control chart:

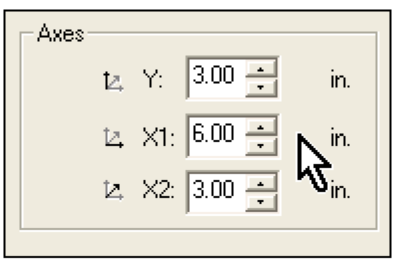


There are wide variations in the early readings, when the operator was frequently readjusting the machine setting. After observation 47 (note pointer), when readjustment ceased, the variations are much narrower. The mean level of those final 33 readings is a bit higher than the midpoint of the upper and lower specification limits, 1.537. A small, once-and-for-all, adjustment could be made to compensate for the fact that the mean of the final 33 readings is slightly higher than 1.537. But even without that adjustment, it is apparent that the machine can provide very low variability and will provide measurements consistently inside the specification limits of 1.536 and 1.538, **if only it is left alone**!

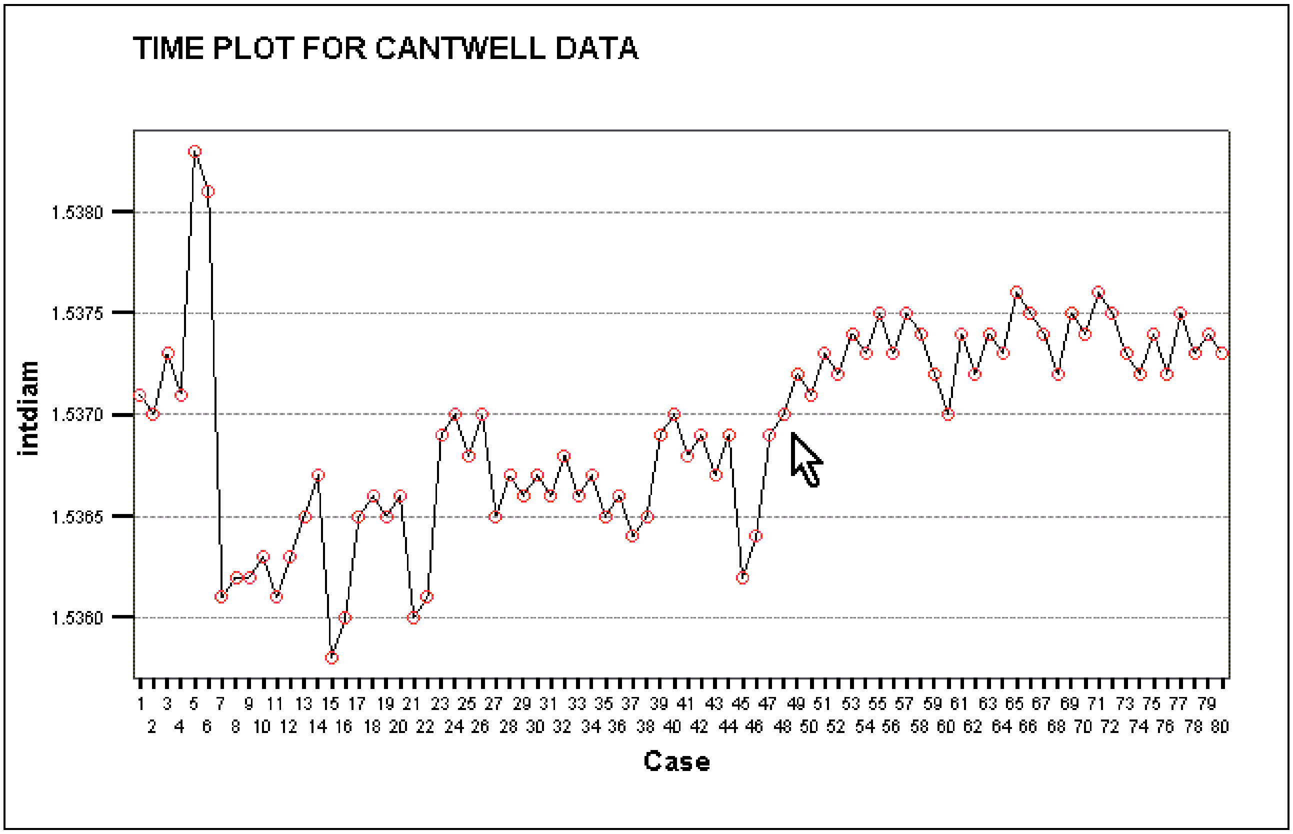
Just for the record, here is an alternative plot produced by **Graphs/ Interactive/Line…** that is wider and easier to look at. The initial dialog window is set up like this,



and after clicking on the **Options** tab we change the **X1 axis** from 3 to 6 inches:



We leave it to you readers to figure out how to add the title and the dots at each value by using another tab in the dialog box. Furthermore, the grid lines were added by double-clicking on the plot after it was produced, using the right mouse button and other steps. Again, we caution that interactive graphing, although yielding results that are aesthetically pleasing, is not absolutely necessary for doing either your homework or a data analysis project. Thus we recommend ignoring these techniques unless you have a lot of free time to play around with ***SPSS***.



The practical importance of this little study was great. This particular bushing had been causing serious performance problems in truck engines because of the wide variations of its internal diameters, which led to many bushings outside of the specification limits.

Here is a brief summary of the two kinds of tampering that we have illustrated:

- Second target game and zeroing-in the rifle: try to compensate or correct fully and immediately for deviations from aim point.

- Bushing diameters: whenever a result appears to be "unusual", make a substantial adjustment in the hopes of correcting the process aim, whatever it is that caused the unusual result.

If the process is in control to begin with, both forms of tampering will make the process worse because they increase process variability.

"Hot Streaks and Slumps"

Tampering is more widespread than these two examples might suggest. We illustrate by sports examples, which have many parallels both in business and in personal life. Both individual and team sports performance can vary enormously through time. A star basketball player may hit 5 three-point attempts in a row, then miss 11 of the next 14. A baseball team may win 11 games in a row, then lose 8 of the next 10. The nearly irresistible temptation is to try to fix whatever it is that has gone wrong. The basketball player may even hesitate to try good three-point opportunities, or, worse, start to tamper with the shooting technique that has served him well in the past. The coach may even consider benching him. (Most coaches are too smart for this.)

The owners of the baseball team may start to think of firing the manager. (Many owners may not be too smart for this.)

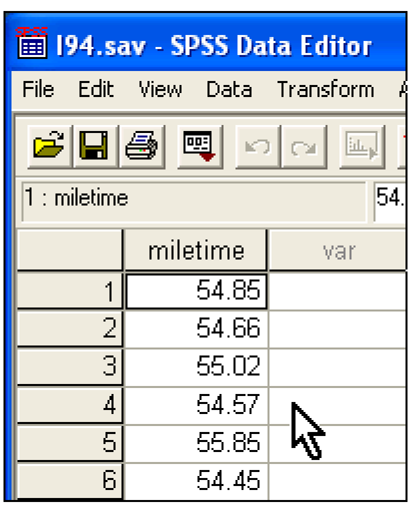
There is a lot of evidence to suggest that -- at least for professional athletes and teams -- sports performance measures through time are usually in a state of statistical control or a close approximation to statistical control, in spite of all appearances to the contrary. Appearances to the contrary are largely due to the statistical fact that random occurrences of successes and failures lead to much more variation in percentages of success in short sequences of trials than people would expect intuitively. (We shall later illustrate by a data set of shots made and missed in a game between the Chicago Bulls and the Boston Celtics.)

In the three-point shot example, change of shooting technique would be in order only if there is good reason to think that a particular change should improve performance, not just because recent performance has been disappointing. (In a later chapter, we will present statistical tools for systematic experimentation with process changes.)

12. "Outliers" and "Special" or "Sporadic" Causes

What happens if a point is found to be outside of control limits, above or below? A statistical term for such a point is "outlier". It is always reasonable to investigate outliers to see if a special or sporadic cause can be found. We'll look at a simple example from a lightning data set in the file, I94.sav, containing a variable named **miletime**. The observer's notes are as follows:

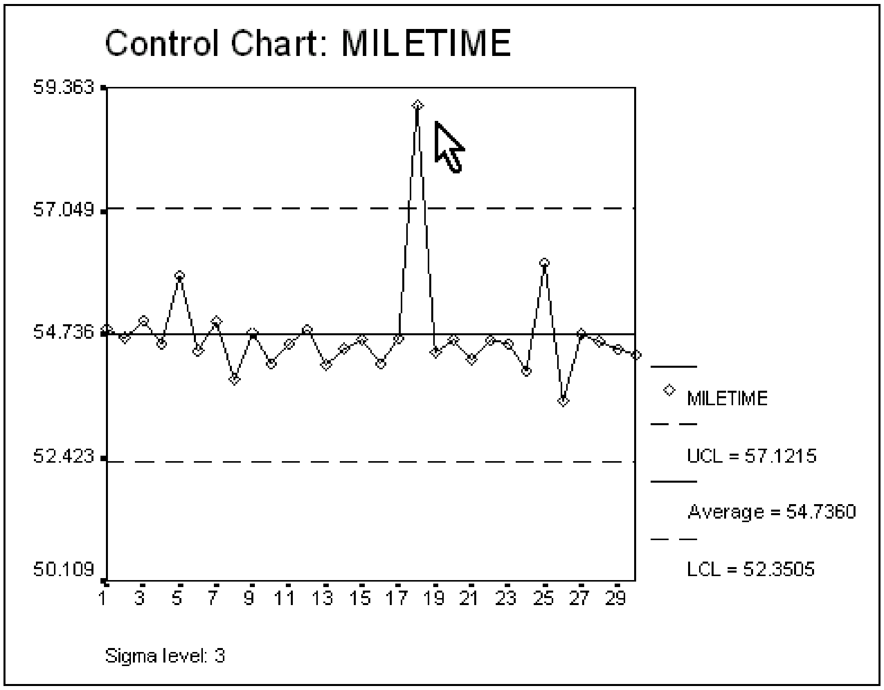
> Time between successive mile markers on I94 southbound, north of Wisconsin Dells, Wisconsin, 7 September 1993. Driving with cruise control set at about 66 mph. Level or gently rolling terrain. A traffic situation required brief disengagement of cruise control on mile 3. Total time 27:22.08 mean 54.74.

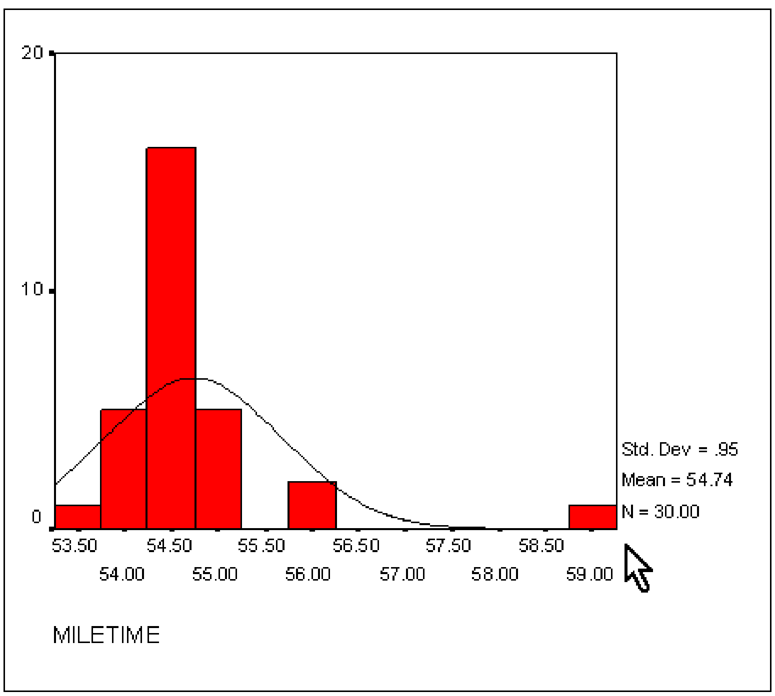


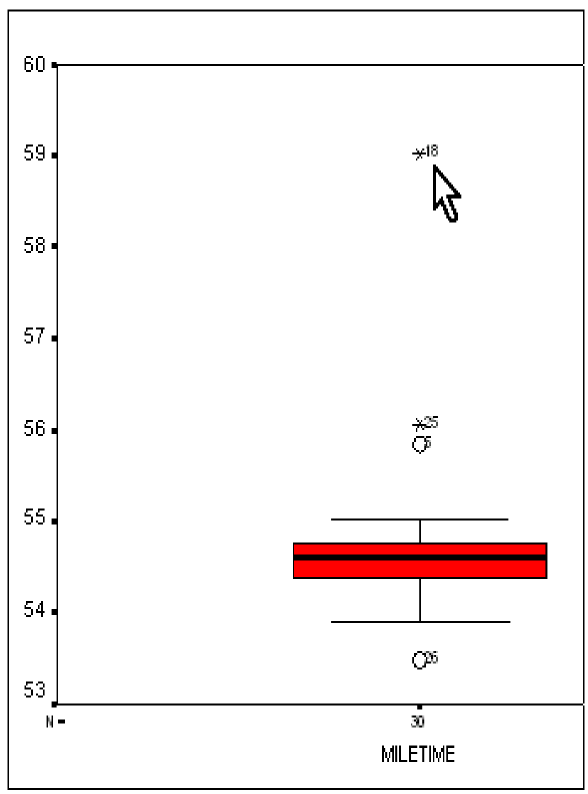
We only display the first few cases in the set of 30. Here are the descriptive statistics followed by a control chart, histogram and boxplot:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
MILETIME	30	53.49	59.02	54.7360	.94632
Valid N (listwise)	30				







Observation 18 is an example of an individual observation that is "out of control". It is called an "outlier" because it is not easily reconciled with the rest of the data. Mile 18 is far outside control limits. If the process were in control and approximately normally distributed, the chance that a single observation will fall outside the three standard deviation control limits is only about 3 in 1000. Mile 18 is much more extreme than three standard deviations: it is more like five than three standard deviations. It is therefore prudent to look for special causes.

The search for special causes will not always be rewarded: either the special cause was present but could not be found, or the outlier was simply one of those improbable things that happens about three times in 1000 when a process is in statistical control.

Experience suggests, however, that the search for special causes is often rewarded. When a special cause can be found, it can often be corrected, or precautions taken so that it does not recur. (In case the outlier was a good outcome -- such as an unusually high yield of a refining process -- steps can be taken to assure that the special cause **always** occurs.)

In practice, our search for a special cause should begin as soon as we become aware of the possibility that one is present. The search trail often becomes faint as time passes. Memories fade, written records are often inadequate.

However, this lightning data set was collected as a byproduct of a vacation trip, and we didn't have time to turn around, check exactly where it was, and pursue detective work. But we can offer useful speculation as to what the special cause might have been.

- It might have reflected a gross timing error by the data recorder. However, she was using a Casio Lap Memory 30 digital watch, which recorded the individual and cumulative times for each mile, and all these were internally consistent to the hundredth of a second. Furthermore, the watch itself is extremely accurate, gaining only a few seconds in a month. The recorder’s reflexes are far from perfect, but the **maximum** error from this source is likely to be no more than a couple of tenths of a second. (In Section 13 below, we provide some evidence on this.)

- Moreover, if the recorder had somehow been 4 seconds too slow in pressing the split button at the end of mile 18, the next mile time should have been several seconds **less** than the general level of about 55 seconds. You can see from the above listing of the file that the time for mile 19 was 54.39 seconds, just slightly under the mean. Also, the preceding mile, 17, was 54.67 seconds, also close to the mean.

- By similar reasoning, the evidence would also seem to rule out the possibility that someone had simply put the marker for mile 18 more than one hundred yards from where it belonged.

- Something may have gone wrong with the cruise control, and the car may have been going much less than 66 mph for mile 18 -- it would have to have suddenly slowed to an average of 61 mph during that mile. However, there was a fair amount of traffic around us, all going steadily at about 65 plus mph, the terrain was flat, and we had no impression of loss of speed. Nor had we noted any recent problems with the cruise control. And even if the cruise control did intermittently fail, it seems odd that the failure had a substantial effect in only one of the 30 miles that were measured.

So we are left with a mystery that could be resolved only by more effort than would be worthwhile for a lightning data set that was primarily intended to have instructional value. But there could be interesting practical consequences. Suppose that the error was not in the timing, the cruise control, or in placement of the mile markers. Rather, suppose that in the original planning of the expressway, "mile 18" had been incorrectly surveyed as a mile when it was substantially longer. The cost of construction of a mile of expressway is many millions of dollars. The contractor who built that mile could have lost a great deal of money!

As we have already noted, an observation outside control limits is termed an **outlier**. It is good statistical practice generally to look for outliers, and to pay careful attention to them when they are found.

Many people are inclined simply to ignore and discard outliers with the rationale that outliers are measurement errors. Outliers may, in fact, reflect only measurement errors, but they may also reflect special causes of which we would otherwise not be aware. Even if outliers are the result of measurement errors, it is a good idea to find out why the error occurred so that similar errors can be prevented in the future. Outliers, like defects on personal quality checklists, are friends, because they potentially have much to tell us. They should not be treated as unwelcome strangers who should be thrown out the door.

13. Two Further Applications

To give further illustrations of lightning data sets and further opportunities to review the key concepts of data analysis introduced in this chapter, we offer two further examples.

Timing the Passage of Ten Seconds

The first application, like that of Section 12, entails use of a digital watch, but the purpose is different. When running or swimming races are timed by hand, different timers will often disagree by one or more 10ths of a second in an event as short as the 60 yard dash.

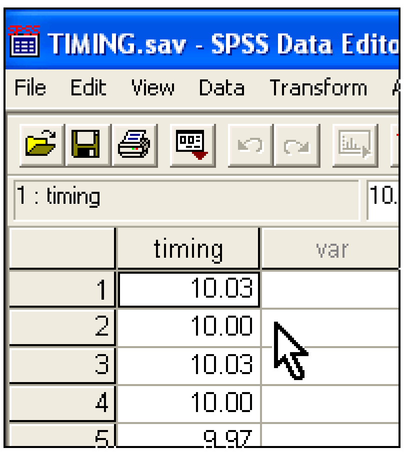
In a running race, for example, the timer tries to start the watch at the flash of the starting pistol and to stop it at first contact with the finish tape. Obtaining accurate results by this process is harder than it sounds. One problem is anticipating either of these two signals and pressing the starting button too soon, or being inattentive and pressing the button substantially too late. But even if the timer gives the mental command to press the button at exactly the time the signal is perceived, there is a delay between the perception of the signal and the actual pressing of the button, due to delays of human reflexes. On average, these delays are on the order of 0.2 seconds, about half the time it takes a fast ball in baseball to travel from the pitcher's mound to home plate. But the delays are variable, so the delay in starting the watch does not exactly cancel out the delay in stopping the watch.

Accurate measurements are necessary in quality management, and an application based on timing measurement provides a good illustration as to how measurement processes can be studied. We shall use a lightning data set to obtain a simple application. We simply used one watch to time how long it took another watch to advance ten seconds. In this example, both watches were digital. The timing watch had a stopwatch feature and provided a readout to the hundredth of a second. The watch to be timed was digital, and the starting signal occurred when it displayed a starting signal (00 seconds or 30 seconds); the ending signal occurred when it displayed an ending signal (10 seconds or 40 seconds). (The watch to be timed could have been an analog watch, as long as it had a sweep hand. Then the signals would occur as the sweep hand touched the tick marks for 00, 10, 30, and 40 seconds.)

The following ***SPSS*** displays show our experience in recording and analyzing these data. The data are reasonably compatible with the assumptions of statistical control and normality. You should study the analysis carefully. (You may also try a similar experiment on your own: your own data will help you to internalize that idea that abstract concepts like "statistical control" and "normality" have real-world relevance. You can also see whether your timing was more accurate than ours!)

Here are the observer's notes for the variable named **timing** contained in TIMING.sav:

> 30 consecutive timings of ten seconds on one stopwatch by a second stopwatch. Done on an airplane, March 23,1993. Two readings per minute taken on the even minute and the minute plus 30 seconds, with recording in between. Thus the data collection took only 15 minutes. Data in seconds and hundredths. Watches were CASIO digitals, which display timings to 1/100th of a second. They are highly accurate, so the main source of error is in the reflexes of the person doing the timing.

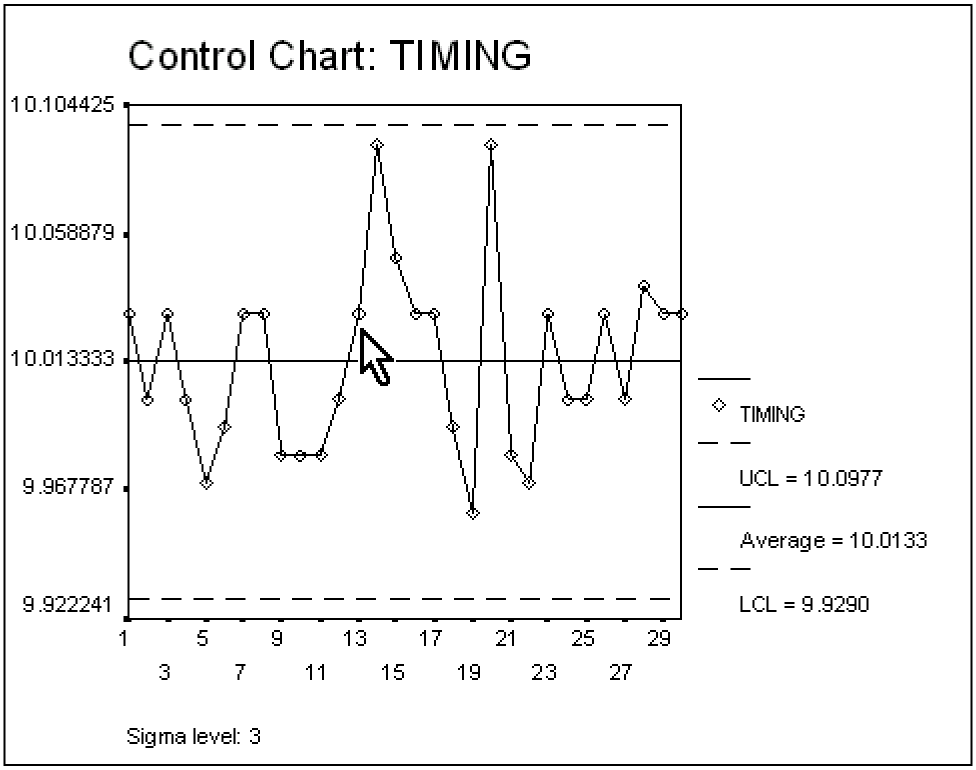


Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
TIMING	30	9.96	10.09	10.0133	.03220
Valid N (listwise)	30				

We see that the mean, 10.0133 is close to the theoretical value, 10, and that the standard deviation is only 0.0322, both suggesting that the measurement process is accurate.

We next use ***SPSS*** to make a control chart:



The appearance of the control chart does not exhibit any indication of predictability, but, although the trend is level, there does seem to be some increase in variability (Cases 13-25) and then settling down thereafter.

Runs Test

	TIMING
Test Value[a]	10.0133
Cases < Test Value	15
Cases >= Test Value	15
Total Cases	30
Number of Runs	15
Z	-.186
Asymp. Sig. (2-tailed)	.853

a. Mean

The runs test shows that the actual number of runs is very close to the expected number under the hypothesis of a random (in control) process. Thus the significance level (p-value) is very high--much greater that 0.05--, implying that we **should not reject** the hypothesis of randomness.

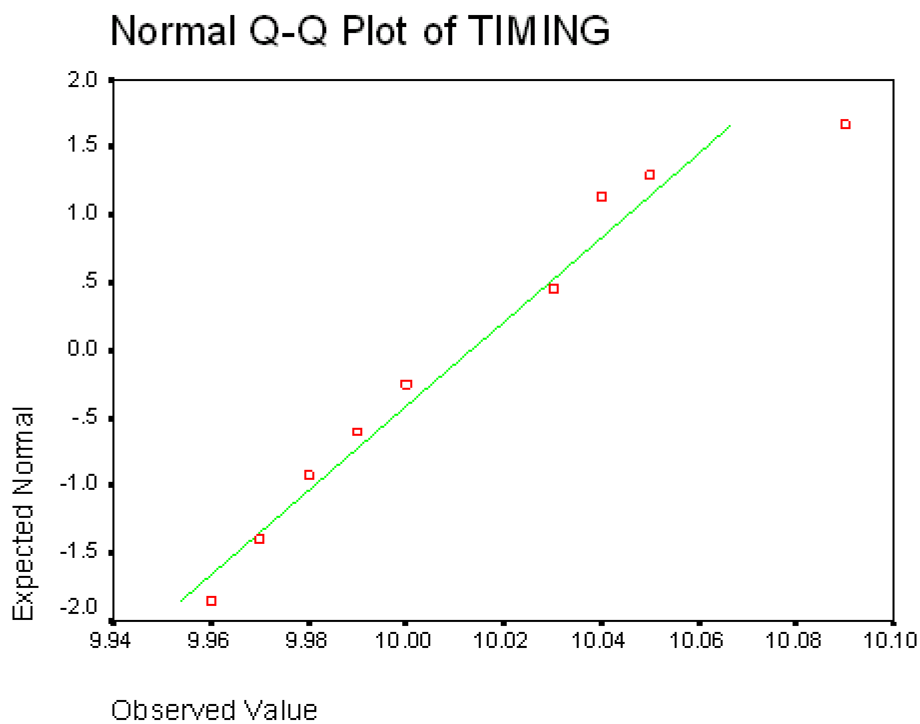
Now for the test to see if we can safely assume that the data are normally distributed:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
TIMING	.198	30	.004	.911	30	.016

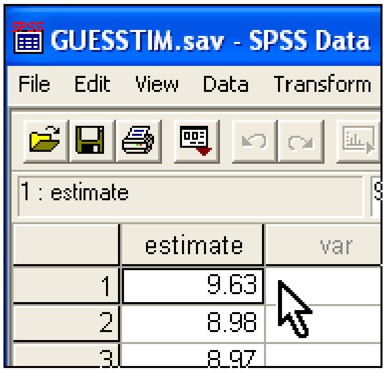
a. Lilliefors Significance Correction

TIMING



You may be surprised to see that the Q-Q and the Shapiro-Wilk test do not support normality. (Make sure that you understand what it is in the output that supports this conclusion.) Although the data may in fact be approximately normal and the plot is just overly sensitive to the small sample size, even if the data are not normal they can still be in control, i.e., random and independent. Repeating we have said before, random data may be explained by one of any number of probability models, the normal being just one of them. It is true, however, that many processes are well described by the normal model, but certainly not all.

Here is a second example. The same observer who gathered the lightning data in TIMING.sav tried to see how well he could estimate subjectively the passage of 10 seconds of time. Each guess was automatically measured by starting a digital stopwatch, stopping it when the observer thought that the 10 seconds had elapsed, and recording the time that was showing on the watch. The data are stored in a file called GUESSTIM.sav and the variable of interest is **estimate**.

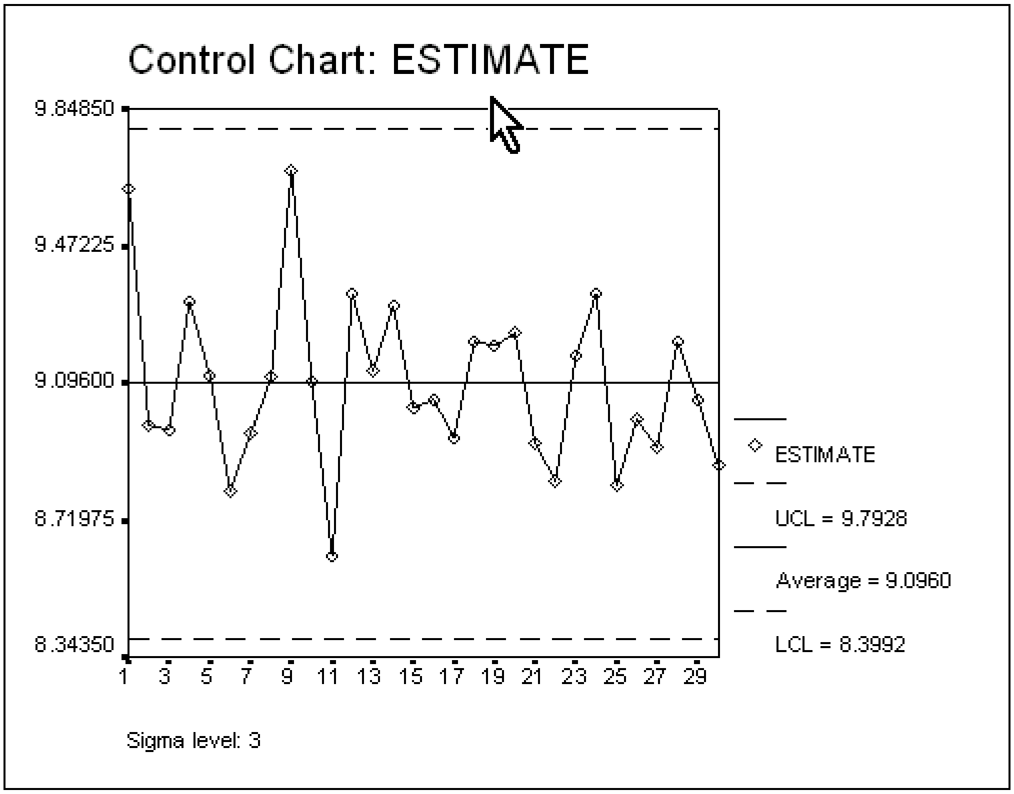


After entering the contents of GUESSTIM.sav in Data Editor, we apply the analytical steps that we just performed on **timing** above to the new variable, **estimate**.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
ESTIMATE	30	8.62	9.68	9.0960	.23202
Valid N (listwise)	30				

We see immediately that the guessing process is biased downward. The mean of **estimate** is, 9.096, close to a full second low. Note that the maximum is 9.68, indicating that our estimator never even got as high as 10 seconds, the target time. Note also that the standard deviation is 0.232, larger than the spread for **timing** in the previous example. Let's see if the process appears to be in control:



Runs Test

	ESTIMATE
Test Value[a]	9.0960
Cases < Test Value	15
Cases >= Test Value	15
Total Cases	30
Number of Runs	14
Z	-.557
Asymp. Sig. (2-tailed)	.577

a. Mean

The chance under randomness of getting 14 or fewer runs, or a result equally extreme in the opposite direction, is 0.577-- much higher than 0.05-- so we consider the data to be consistent with a process that is in control. The same conclusion is drawn from the control chart. We must not forget, however, that although apparently in control, **estimate** varies around a mean that is considerably lower than the target 10 seconds. This is a clue that some special cause has taken hold.[9]

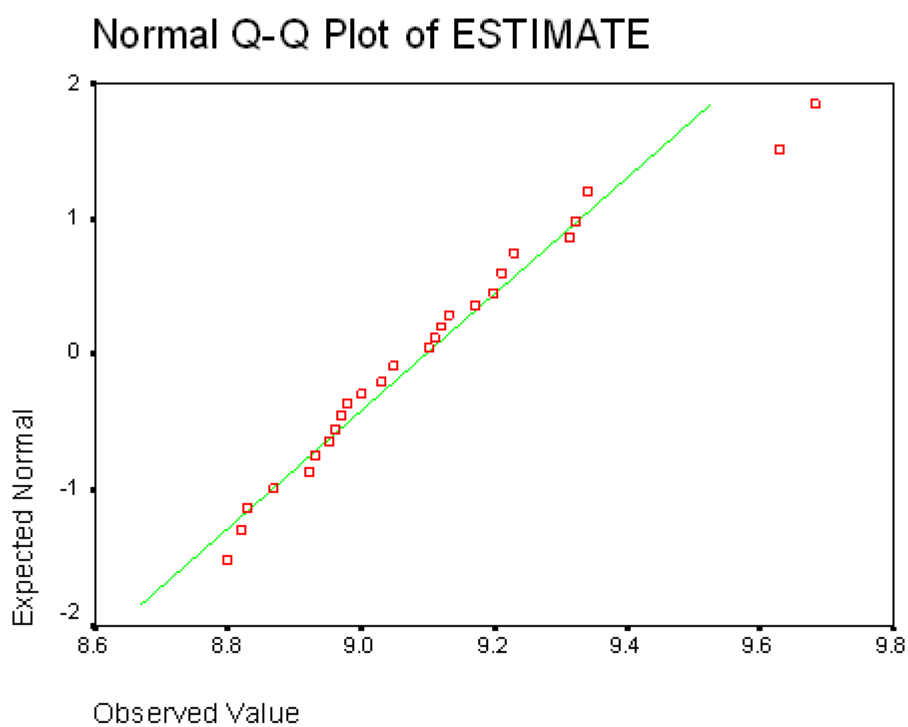
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ESTIMATE	.082	30	.200*	.965	30	.406

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

ESTIMATE



The Q-Q plot above and the significance level for the Shapiro-Wilk test indicate that if the mean and the standard deviation do not change, the normal probability model may be used to predict future estimates.

9We will never be able to say for sure what that special cause is. Apparently our estimator systematically perceived time to be passing faster than it actually did. We can only conjecture about the systematic way that the mean would change if more wine had been consumed on the flight.

14. Checking for Bias

In three of the four examples of this chapter, there was a known target value for the process:

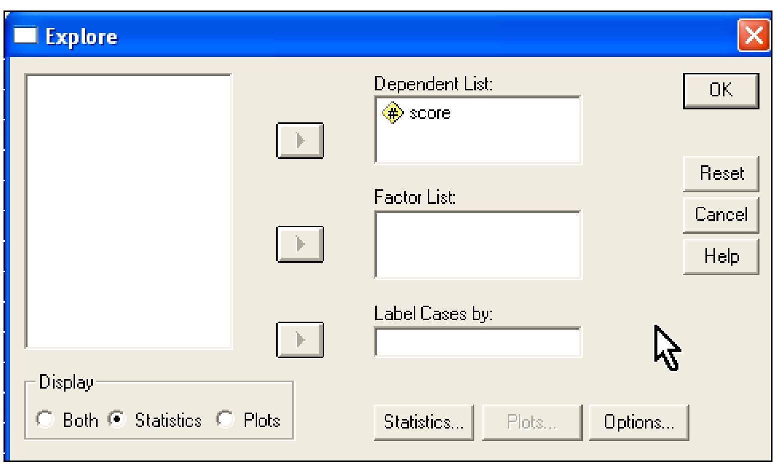
- In the target game, the target value was 0, the numerical value of the target strip at which the player was aiming.
- In the timing example, the target value was 10 seconds.
- In the example of guessing the passage of 10 seconds, the target value was also 10 seconds.

Whenever a process has a target value, one desirable criterion for process performance is **zero bias**: on average, the measurements of process performance should equal the target value. We can quickly provide a rule of thumb for deciding whether the data support the idea of zero bias. (In a later chapter, we shall discuss the statistical testing ideas on which the rule of thumb is based. They are essentially the same as in the discussion of runs in Section 6 of this chapter, where we explained the term "significance".)

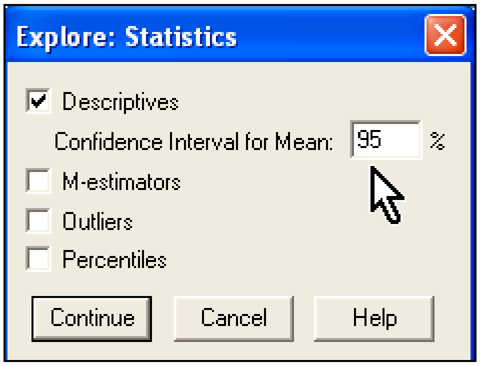
In the following discussion we are going to use the ***SPSS*** procedure **Analyze/Descriptive Statistics/ Explore…**

Target Game

For the data in TARGET.sav we set up the Explore window like this,



and after clicking on the **Statistics…** button at the bottom, we make sure that the next open window looks like this:



Clicking on the **Continue** button and the **OK** button yields the following table:

Descriptives

			Statistic	Std. Error
SCORE	Mean		-.03	.286
	95% Confidence Interval for Mean	Lower Bound	-.62	
		Upper Bound	.55	
	5% Trimmed Mean		-.07	
	Median		.00	
	Variance		2.447	
	Std. Deviation		1.564	
	Minimum		*	
	Maximum		4	
	Range		7	
	Interquartile Range		2.00	
	Skewness		.175	.427
	Kurtosis		1.089	.833

In this game the target value is 0, and the mean of 30 sample tosses is -0.03, seemingly very close to zero. If we had observed, however, another sample of 30, the mean would probably **not be the same**. Imagine repeating the sampling process of 30 tosses, time after time, each time recording the results and calculating the mean deviation from the target. It can be proved mathematically that the long-run average of means from this repetitive sampling will be the "true mean error", which may or may not be close to zero.

We need to decide on the basis of **only one sample** whether or not our estimated mean, -0.03 is close enough to zero that we can conclude that the "true mean error" is zero. To do this, we need an estimate of the spread of the theoretical distribution of estimated means that would result if we repeated the sampling of 30 over and over. This is the purpose of **Std. Error**, the **standard error of the mean**, which is reported as 0.286. **Std. Error** is defined as follows, where "**SQRT**" is an abbreviation for "square root."

Std. Error = Std.Deviation/ SQRT(N)
= 1.564/ SQRT(30)
= 1.564/ 5.477 = 0.286 .

Note that the standard error increases with the standard deviation of the individual observations: uncertainty is greater if the individual readings vary more. Further, the standard error decreases as the sample size increases: uncertainty is less as the number of observations increases. That is, if our sample size were infinite there would be no error at all concerning the mean, as long as we could measure accurately.

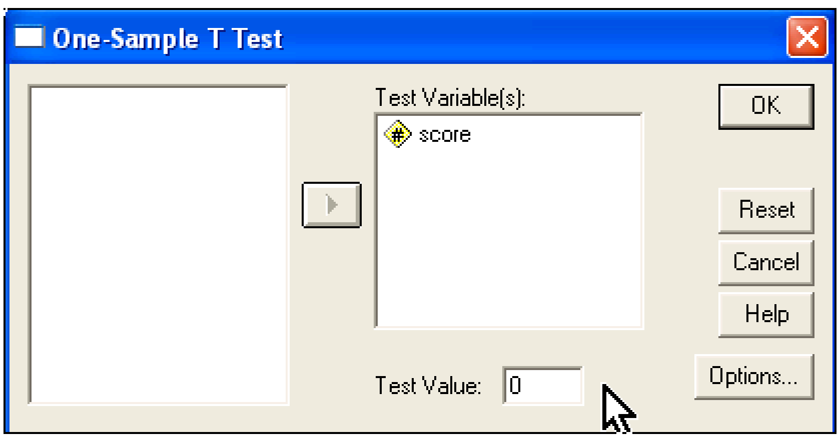
Here is a rule of thumb for deciding if the mean is sufficiently close to the target and how to apply it:

- Calculate the absolute deviation of the **mean** from the target value, **0**: **│ -0.03 - 0 │** = **0.03.**

- Divide 0.03 by **Std. Error** = 0.286: **0.03/0.286 = 0.11**

- If the result is less than **2**, judge the sample departure from the target value to be insignificant; otherwise judge it to be significant. In this example, then, the player displayed no significant tendency toward bias, that is, to consistently overshoot or undershoot.

You don't actually have to carry out the above arithmetic. ***SPSS*** gives the answer with the sequence of steps **Analyze/Compare Means/One-Sample T Test...** Executing the sequence yields the following dialog window:



Note that we have to indicate the value that we are testing against in the lower box for **Test Value**.

The window is set up to test the hypothesis that the underlying mean of **score** is zero against the alternative hypothesis that the mean deviates from zero in either direction, negative or positive. When the **OK** button is actuated the following results appear:

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
SCORE	30	-.03	1.564	.286

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
SCORE	-.117	29	.908	-.03	-.62	.55

The report **t = -0.117** corresponds to the result calculated above. (We understated the value slightly because of rounding.) The output also gives the "p-value" (significance) equal to **0.908**: if the player really had no bias in his bowling of the post-it pad (that is, if the true mean is zero), a mean departure of 0.03 or more in either direction from zero would occur about 91 percent of the time, not at all an unusual happening. Because the observed result is not unusual under the hypothesis that the mean is zero, we conclude that **we should not reject that hypothesis**.

Finally, we note that the descriptive statistics report a 95% confidence interval for the mean with a lower limit of -0.62 and an upper limit of 0.55. There is a great temptation to interpret this as saying that the probability is 0.95 that the true mean error lies between **-0.62 and 0.55.**

Unfortunately, that is not a correct statement-- the confidence coefficient, 95%, is not a probability that applies to the interval -0.62 to 0.55. Rather, it expresses our confidence in the method in the sense that it is the probability **before** we actually generate and observe the sample that an interval computed by this method **will capture** the mean. After the sample is observed and the interval calculated we really do not know whether the mean is in there or not, and the traditional approach to confidence interval estimation does not permit us to assign a probability to that event **after the results are seen**. We only know that **before we did the calculation** we had a lot of confidence that it would yield correct results.

If this seems strange and unsatisfying, you should accept the fact that statistical methods seldom, if ever, enable us to make definitive statements or to eliminate uncertainty altogether. With the right data and the right analytical approach, however, we can add to our accumulated knowledge about a process in a way that **reduces our uncertainty**. There are two main points to remember from this example:

- A close approximation to the 95% confidence interval can be hand-calculated by adding and subtracting two times **Std. Error** from **mean**. This calculation yields

-0.03 ±2(0.286) = -0.602 to 0.542,

which is not quite wide enough to agree with the exact results above, but close.

The fact that the confidence interval contains zero is consistent with our conclusion from the one sample t-test that the bias in the target game is negligible

Now we shall consider the example using the data in TIMING.sav. If you execute the one-sample T- test yourself, don't forget to enter 10 for the target value:

Timing 10 Seconds

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
TIMING	30	10.0133	.03220	.00588

One-Sample Test

	Test Value = 10					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
TIMING	2.268	29	.031	.0133	.0013	.0254

Std Error Mean is quite small because there was not much variability in the sample observations. This yields a 95% confidence interval of the bias from .0013 to .0254. The fact that the desired value, zero, is not in that interval means that have to admit to a slight upward bias in the timing.

The large t-value, **2.268**, and the correspondingly small p-value, **.031**, confirm the indications of the confidence interval.

There conclude that there is evidence of a small, but "significant" bias in the process.[10]

[10] One of the many problems with traditional significance tests is that sample results may be "statistically significant" but not "practically significant". Because Std. Error Mean decreases with increasing sample size, it follows that with a big enough sample **any estimated bias** different from zero will be statistically significant. But is it practically significant? Does it matter in the sense of leading to waste and inefficiency?

Guessing 10 Seconds

In this example the target was also 10 seconds. Here is the output from ***SPSS***:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
ESTIMATE	30	9.0960	.23202	.04236

One-Sample Test

	Test Value = 10					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
ESTIMATE	-21.341	29	.000	-.9040	-.9906	-.8174

We see that the 95% confidence interval for the bias runs from -.9906 to -.8174, clearly much lower than zero. The results of the one-sample t-test confirm our findings of a possibly serious bias of underestimation.

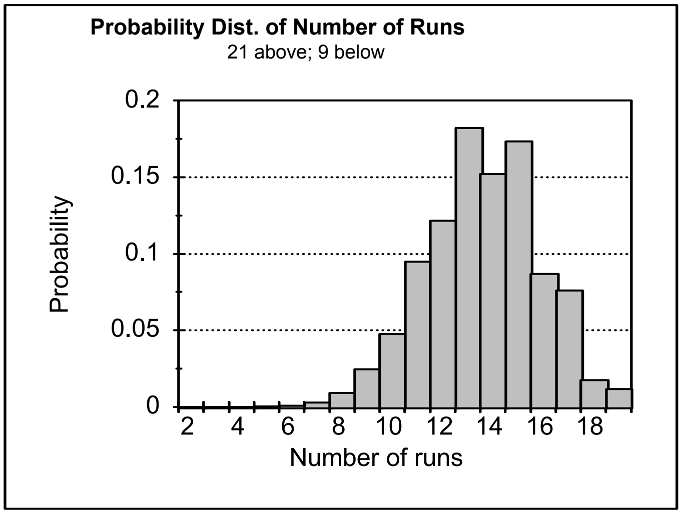
Note, however, that although the t-statistic is very large and negative, the significance level is not really .000. The output should say <.0005, but rounded up to zero. Any value of t is possible, no matter how improbable.

Appendix: FURTHER REMARKS ON THE RUNS TEST

Starting on page 2-13 of ***STM***, we have discussed the runs test for the score data in TARGET.sav. Note that there were 21 observations above the mean and 9 below. These numbers, 21 above and 9 below, plus the total sample size, 30, are the basis for the calculations of the expected number of runs, 13.6, and also the p-value (significance level) that is displayed by ***SPSS***. Perhaps the following will help you to better understand the reasoning behind all of this:

If the observations of **score** are random, then any possible sequence of the 21 aboves and 9 belows has the same chance of occurring as any other. The total number of such scramblings is given by a mathematical formula (the number of combinations). The formula tells us that the number of different ways to get 21 aboves and 9 belows in 30 observations is very large--it is 14,307,150. We also discussed in class that the number of runs can range from as few as 2 up to as many as 19. Thus each of those 14,307,150 arrangements of 21 aboves and 9 belows has a number of runs associated with it. Through further application of mathematics we can calculate the chance of each different number of runs occurring **if the data are random**. The table below shows the probabilities of each outcome, assuming randomness. (Note that the probabilities sum to one.) The histogram on the next page provides a visual impression of the way that the total chance of occurrence is distributed over the possible numbers of runs.

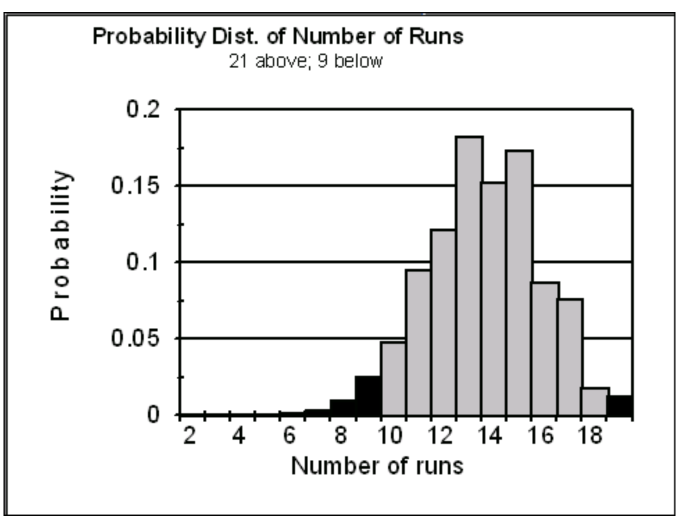
RUNS	PROBABILITY
2	0.0000001
3	0.0000020
4	0.0000224
5	0.0001454
6	0.0007437
7	0.0029747
8	0.0089242
9	0.0245416
10	0.0474099
11	0.0948197
12	0.1213692
13	0.1820539
14	0.1517116
15	0.1733846
16	0.0866923
17	0.0758558
18	0.0176094
19	0.0117396
	1.0000000



Even though the distribution is skewed to the left, one might venture to say that the fit to a normal curve is not too bad, but we are dealing with exact probabilities here so we need not worry about a normal approximation.

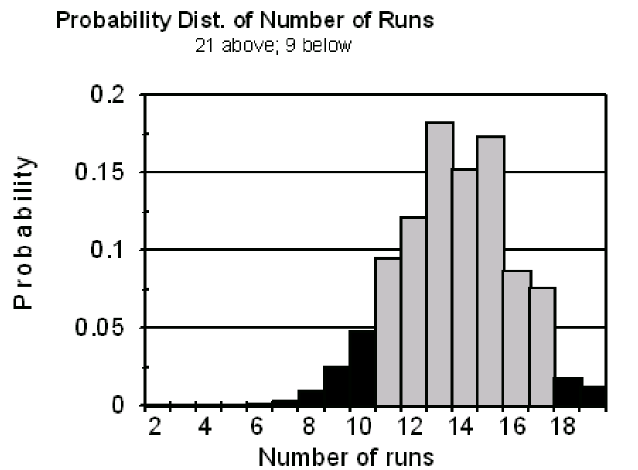
Now, we saw 14 runs in the particular pattern of 21 aboves and 9 belows that was displayed by **score**. The question is, "How compatible is this pattern with what is to be expected if the data are really random?" And we answer the question in terms of how unusual it is for 14 runs to occur. In fact, we allow for an equally unusual result in the other direction from the expected number of runs, 13.6. So we ask, "What is the chance of having 14 or more runs or 13 or fewer runs **if the data are random?**", and since 14 or more or 13 or fewer cover every possibility, that chance is reported as equal to one. We call the result the **p-value** or **significance level** in a two-tailed test. It follows that since the chance of obtaining what we got or an even more extreme result is a sure thing if the data are random, we conclude that the data are probably random-- at least as far as the runs test can show it.

Suppose, however, that we had seen a different arrangement of the 21 aboves and 9 belows and that the I-chart had shown only 9 runs. We reason as follows: 9 runs is 5 fewer runs that we expect under randomness (13.6 rounds to 14) and 19 runs is an equal distance above the expected number. Thus we need to calculate the probability of 9 or fewer or 19 or more, and by adding the appropriate numbers in the table above we obtain a p-value of 0.049. We have shaded in the corresponding bars in the histogram below:



As explained in the textbook on page 2-17, there is a rule of thumb that says if the significance level (p-value) is less than 0.05, we say that the runs check is inconsistent with the assumption of statistical control (randomness). Since we are just below that cutoff point, we report the results as "statistically significant" and we reject the assumption of randomness.

Let's look at one more possible outcome in the I-chart. Suppose that instead of too few runs, the control chart had shown 18 runs, indicating a great deal of alternation from above to below the mean. How significant is that result? We see from the table and shaded area below that the p-value is now 0.114, about 11 chances in 100, but greater than the rule of thumb, 5 chances in 100. Most researchers would conclude that the number of runs is a little high, but not high enough to conclude that the process is out of control.



Remember this: A large deviation of the actual number of runs from the expected number in either direction will result in a p-value (significance level) that is small. The **smaller** the p-value, the **less consistent** the data are with a process that is in statistical control (random). Small p-values lead to rejection of the hypothesis of randomness; if the p-value is large we should not reject. We call a result with a small p-value "statistically significant". (If this language is confusing don't blame us. Statisticians, and researchers in general, have been talking like that for many years and their habits are probably impossible to change.)