

Lies, Damned Lies, and Teens Who Smoke

Donald J. Wheeler

"Teen use turns upward" read the headline for a graph appearing in *USA Today* on June 21, 1994. The data in the graph were attributed to the Institute for Social Research at the University of Michigan and were labeled as the "percentage of high school seniors who smoke daily." The portion of this graph covering the past ten years is shown below.

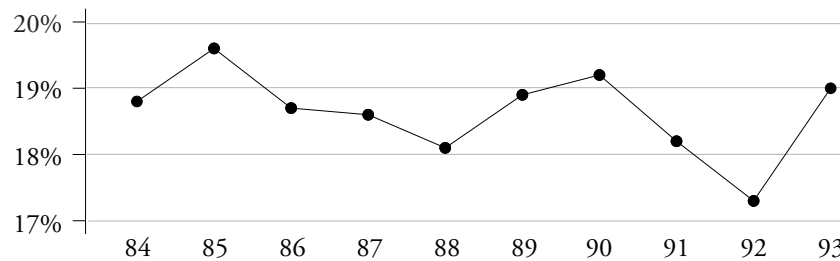


Figure 1 The Percentage of High School Seniors Who Smoke Daily

Each point is the value found in an annual survey. The 1993 value of 19.0% was higher than the 1992 value of 17.3%. This was interpreted to mean that more teenagers were using tobacco now than in the past. But were they?

Before we can make sense of numbers such as these we need to know something about the limitations of data from surveys. First of all, values such as these are subject to variation. Two identical surveys carried out at the same time will rarely yield identical results. Among the sources of variation for survey data are differences in who is interviewed, differences in how they are interviewed, differences in how they respond, and differences in how their responses are reported. Secondly, when a survey is used from year to year there is also the problems of different personnel being used to conduct the survey, and differences in how the questions are perceived in different years.

Therefore, no matter what is being measured, and no matter how carefully it is measured, the statistics will always vary. There will always be some variation in survey results from year to year. Even if nothing is changing, we can expect the value to go up about half the time, and we can also expect the value to go down about half the time. (Look at the past nine years—these values were greater than the previous value four times, and less than the previous value five times.)

So how do we ever detect a change using survey data? If we interpret each and every change in the percentage who smoke daily as a year-to-year difference, how do we know that we are not being misled by the study-to-study variation? If we admit that there is study-to-study variation, how then do we ever know when there has been a change from one year to another?

The answer is that we must first filter out the study-to-study variation, and then look for year-to-year differences. The simplest way to do this is with a process behavior chart (also known as a control chart).

We begin with the yearly values. For the data on high school students who smoked daily, the annual percentages reported for 1984 through 1993 were, respectively:

18.8 19.6 18.7 18.6 18.1 18.9 19.2 18.2 17.3 19.0

The average of these ten values is 18.64 percent. This average is used as a central line, and the ten values are plotted as a time series as shown below. This graph is the beginning of the chart for individual values (also known as an X-Chart).

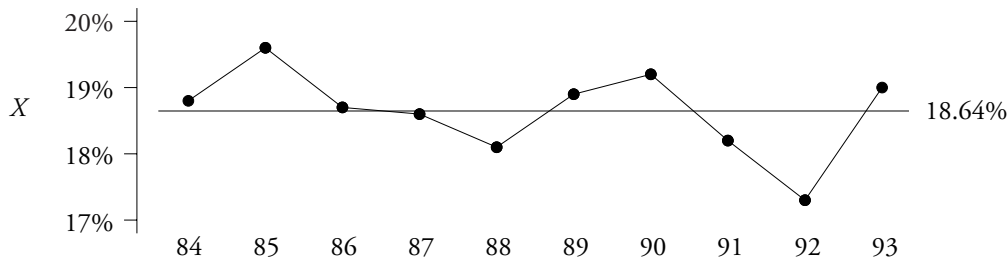


Figure 2 The Percentage of High School Seniors Who Smoke Daily with Average

Since the variation between one year's value and the next will always include the study-to-study variation, we use the year-to-year variation as our guide to how much uncertainty is inherent in the reported results. These year-to-year changes are measured by the differences between successive values (these differences are called moving ranges). The nine moving ranges for these data are:

0.8, 0.9, 0.1, 0.5, 0.8, 0.3, 1.0, 0.9, 1.7

The Average Moving Range is 0.778 percent. We use this Average Moving Range to compute limits for the previous graph.

The limits for our X-Chart are commonly known as Natural Process Limits. They are placed symmetrically on either side of the central line. The distance from the central line to these limits is found by multiplying the Average Moving Range by 2.660. This constant of 2.660 is a scaling factor which converts the raw statistic into the appropriate measure of dispersion.

For these data this distance is:

$$2.660 \times 0.778\% = 2.07\%$$

Thus, the Upper Natural Process Limit is:

$$18.64\% + 2.07\% = 20.71\%$$

Thus, the Lower Natural Process Limit is:

$$18.64\% - 2.07\% = 16.57\%$$

These limits make allowance for routine variation. They are added to the graph to obtain the Chart for Individual Values, more commonly known as the X-Chart.

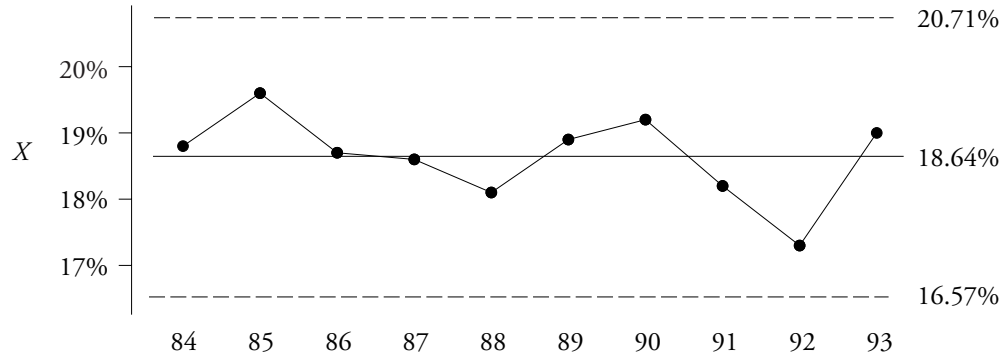


Figure 3 The X Chart for the Percentage of High School Seniors Who Smoke Daily

Before a yearly value can be said to represent a change in the use of tobacco by teenagers, it will have to either exceed the upper limit or fall below the lower limit. Since none of these values fall outside these limits, any statement about changes in the percentage of teens who smoke is questionable.

But wait—the change between the last two values, where the percentage jumped from 17.3% to 19.0%, is the biggest change during the past 10 years. Surely this should mean something.

To see if this is the case we can place the moving ranges on a second chart. The Average Moving Range of 0.778 will be the central line, and the upper limit will be found by multiplying the Average Moving Range by the scaling factor of 3.27. This results in an upper limit for the moving ranges of 2.54.

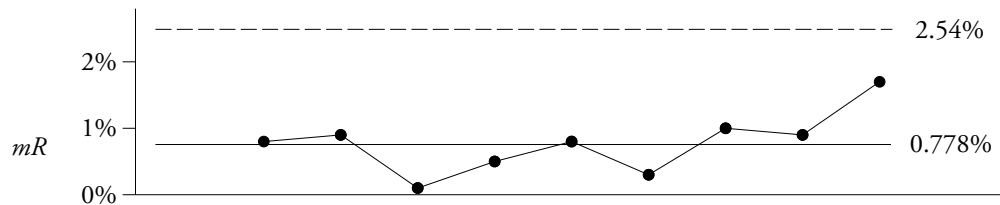


Figure 4 The mR Chart for the Percentage of High School Seniors Who Smoke Daily

The last value on this Moving Range Chart shows the “jump” between 1992 and 1993. This moving range of 1.7% does not fall above the upper limit of the Moving Range Chart. Thus, once again, the “jump” from 17.3% to 19.0% does not qualify as a clear-cut signal.

So, what can we say about the percentage of teenagers who smoke daily? Just this: there is no evidence that the percentage of teenagers who smoke has increased. Neither is there any evidence that this percentage has decreased during the past 10 years. The only headline for these data that has any integrity is “No change in teen use of tobacco.” Anything else is just propaganda.

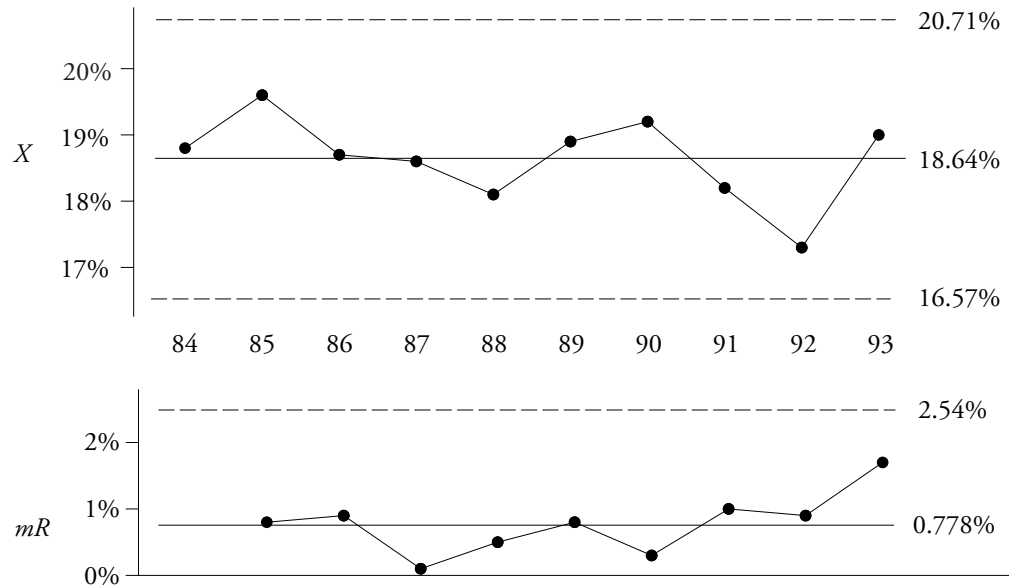


Figure 5 The Complete XmR Chart for the Percentage of High School Seniors Who Smoke Daily

So how do you avoid being persuaded by propaganda? The beginning is to realize that while all data contain noise, only some data contain signals. If you do not know how to separate the probable noise from the potential signals you are susceptible to being misled by the noise in the data. Others may use data to mislead you—or you may even mislead yourself. Process behavior charts like the XmR Chart illustrated here are the simplest way to separate signals from noise.

By the way, did you read the article about how the Trade Deficit soared last April? Oh, well, that's another story—or is it?