Forty-Six Men and a Test

Donald J. Wheeler

Forty-six men were given a skills test. This test involved a task that these men performed every day, and at which they were already proficient. Yet there was a desire to determine who was the best. To this end a test was devised and given.

In fact, to better discriminate within this large group, the skills test was given to each person a total of four times. However, in order to eliminate any fatigue effects, the four tests were given on four different days. The four test scores for each individual were then combined to get an average score for each person, and these average scores were used to rank the 46 men.

A simple and straightforward system. A test is given, scores are obtained, and the scores are used to create a ranking of the individuals. But what can we learn by using a process behavior chart with these data?

In order to place data on a process behavior chart it is important to understand the structure of the data, and the different sources of variation present in the context for the data. The 184 test scores described above have two obvious structures—as we go from one value to another the context for the values may change in two ways. For example, we could pick a pair of values that would correspond to different administrations of the test to the same individual. Since differences in scores for a single person represent the natural variation present when a person performs the same task multiple times we shall call this source of variation day-to-day variation.

On the other hand, we could pick a pair of values that correspond to scores obtained on the same day, but which correspond to different individuals. Since the differences in scores between people will represent the difference in the skill levels we shall call this source of variation person-to-person differences.

And of course we could pick a pair that corresponded to different individuals on different days, but since these differences may be explained in terms of the first two sources of variation, we do not need another label.

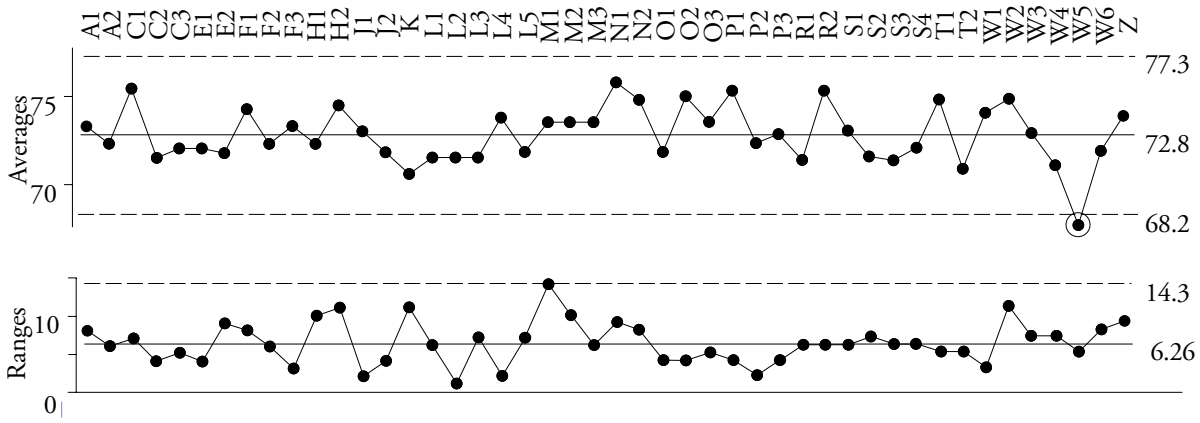
Thus these data contain two main sources of variation: Day-to-day variation and person-to-person differences. The purpose of the skills test is to detect the person-to-person differences. The obstacle to doing this is the day-to-day variation. So we need to filter out the effects of the day-to-day variation before we can detect any person-to-person differences. And we may do this with an Average and Range Chart.

If we combine the four test scores for each person together, and let different people define different subgroups, we will end up with 46 subgroups of size four for our Average and Range Chart.

The Range Chart always checks for consistency within the subgroups. In this case the variation within the subgroups is the day-to-day variation. Thus, the Range Chart will check to see if all 46 individuals show the same amount of variation when they perform the same task repeatedly. If a person should have a range value that is above the upper limit on the Range Chart, then that person could be said to have detectably greater variation in the way he performs the skills test. But those whose range values fall below the limit on the Range Chart can be said to

display the same level of variation in performing this task.

No ranges fall above the limit of 14.3, so we may conclude that all 46 men show the same amount of day-to-day variation in performing this task.



Average and Range Chart for Test Scores

The Average Chart always looks for differences between the subgroups. In this case, when we go from subgroup to subgroup we are changing persons. Thus, the person-to-person differences will show up on the Average Chart.

Since it is the Average Range that is used to compute the limits for both charts, the limits will reflect the uncertainty due to the natural human variation in performing this task. Thus, the limits on the Average Chart define that amount of variation in the average scores that can be attributed to the natural variation inherent in performing this task repeatedly. In order for one person to be said to be detectably above average, or detectably below average, in their proficiency at this task, he will have to have an average that is outside the limits on the Average Chart. Those with averages which fall within the limits must be said to be indistinguishable from each other.

Forty-five of the individuals have averages that fall within the limits of 68.2 to 77.3. These 45 men are indistinguishable in their skill level. They all show the same degree of proficiency. The routine day-to-day variation in the way people perform this skills test is sufficient to explain all ofthe differences between these 45 men.

One man, with an average of 67.5, shows a detectably different level of proficiency. He is different from all the rest. His different level of proficiency cannot be explained by the normal, day-to-day variation that covers all the other person-to-person differences.

Now that you know what these data are telling you, it is time for the rest of the story—these data are the scores for the 1997 Master's Tournament. You can get the data from the sports section of any newspaper for April 14—but then you probably already know who was detectably different from all the rest. His initials are T. W.

Finally, can you see how the Average Chart tells you why Par is 72? And can you express in words what the Average Range of 6.26 represents? Think it over, for the answers to these two questions will help you learn how to use Average and Range Charts more effectively.

The only limitation on the use of process behavior charts is your imagination.