ANALYZING THE TEST DATA

Exercise: Finding Patterns in the Measured Data (Allow 20 to 30 minutes for this exercise.)

Performance measurement projects usually capture large volumes of data. The question is how to make sense of it.

In this exercise, imagine that three controllable factors, A, B and C, all can impact the system response time. We will deliberately not put names to A, B and C yet. (These may be factors like processor speed, size of semiconductor memory, available bandwidth in a network, whether the developers used an optimizing compiler, the intensity of the background noise, etc.) If we give names to A, B and C, biases might creep into our thinking, so we are going to perform this analysis blindly and based strictly on the measured data. We will return later and name A, B and C at the end of this exercise.

If we interested in improving response time, the questions we have are (a) which of these three controllable factors, A, B and C, will give us the biggest "bang for the buck", and (b) how increases or decreases in these factors contribute to improvements in response time.

The answer to the first question depends in part on the relative influence of A, B and C on the response time, which we will examine in this section, and also on other factors, such as the relative costs of changing these factors (e.g., increasingly the processor speed). We have not examined the costs here but we will return to them later.

In addition, I have kept the number of data points in the example small enough to make the computations fairly straightforward: we have measured the response time a total of 8 times, in 8 different situations. In reality, we usually have much more data and would utilize a spreadsheet or a statistical analysis software package to do the number crunching.

We also will assume that A, B and C are all binary, that is, each can be set to either a high or a low value by the testers. We will refer to the response time in seconds as R.

ANALYZING THE TEST DATA
(Continued)

The data collected is as follows:

Test Run	Values of the Factors			Response Time R (seconds)
1.	A lo	B lo	C lo	31.3
2.	A lo	B lo	C hi	10.7
3.	A lo	B hi	C lo	21.6
4.	A lo	B hi	C hi	7.5
5.	A hi	B lo	C lo	30.8
6.	A hi	B lo	C hi	9.2
7.	A hi	B hi	C lo	20.3
8.	A hi	B hi	C hi	6.0

What is this data telling us? What conclusions can we draw?

It helps to have an objective which gives us direction here, so we should answer these questions with other questions: What are we interested in? What do we want to know? These new questions actually were answered earlier: we are interested how the three controllable factors, A, B and C, influence the response time R. As we vary A, B or C, what impact does each have on R?

At first glance the table seems to be a jumble of numbers with no rhyme or reason. Be patient and continue to look at the data and think about it, and in a few minutes patterns will start to emerge.

ANALYZING THE TEST DATA
(Continued)

Suggested Answer to the Exercise: Finding Patterns in the Measured Data

As a follow-up to the original exercise, review and critique this answer, and determine if you (a) agree with the conclusions, and (b) have additional insights based on the collected data.

I can now reveal that A means the rated speed of the processor(s) installed on a server, B means the amount of semiconductor memory on the server, and C means the number of processors – each server can be either a single or dual-processor machine.

The first step is to simply look at the numbers before we whip out a calculator, and see if any relationships come to mind. Normally, at this stage it is important to have a dose of healthy skepticism to help identify and eliminate any questionable data in the set, such as values which are ridiculously high or low. Here, we will simply assume that we have already cleansed the data.

It is important to realize that the relationship between any one of the inputs A, B or C and the outcome R is not necessarily linear (a straight line). With only two data points, it is easy to assume that a relationship between two variables is linear. In other words, we mentally link two points by a straight line. In this situation, there are effectively only two data points for A, B or C – if we hold any two of these factors to fixed values, the curve relating the third factor and the response time R has only two data points. But the small amount of data does not mean straight-line relationships. In this analysis for simplicity we will assume linear relationships, but we want to be conscious that this is a convenient assumption and not a known fact.

We observe that the relationships among the inputs to this table (which are the values of A, B, and C) and the outcomes (the values of R) are all inverse ones – as either A, B, or C increases, the response time R decreases.

The relationships are also all monotonic, which is a mathematician's way of saying that if we increase any one of A, B, or C, or any combination of them, R always decreases. In other words, there is no "going back" where the trend curve reverses direction and R increases again before it decreases. As we change the controllable factors and take any possible path from [A lo; B lo; C lo] to [A hi; B hi; C hi], such as:

[A lo; B lo; C lo] >> [A lo; B lo; C hi] >> [A lo; B hi; C hi] >> [A hi; B hi; C hi], the value of R steadily decreases from step to step along every path. I deliberately doctored the data to make it monotonic in this example, because the patterns tend to be easier to see with monotonic data.

The next thing we might notice is that A does not seem to have much of an influence on R. If we hold the other factors (B and C) to fixed values, either to high or low values, and vary only A, the resulting change in R is minimal.

For example, if we hold B and C both fixed to their low values, as A varies from low to high the response time R decreases from 31.3 seconds to 30.8 seconds. This change in R is statistically insignificant because of the likely imprecision in the measured values of R.

The same observation, that R does not change much as we vary A, holds true if we hold both B and C to different sets of fixed values which are alternatively:

[B lo; C hi], or [B hi; C lo], or [B hi; C hi],

In calculus, we use the notation [dR / dA] to describe the relationship of A and R. "dR" means the incremental change in R, based on or caused by an incremental change in A, which "dA" represents. Mathematicians describe this relationship between A and R as an invariance – R is not affected by the setting for A.

What does this invariance of A and R mean to us?

There are two possible meanings of invariance:

(1) A and R are totally unrelated.

To give a slightly silly example, let's say that A is the processor speed on a machine which is not the one where we are running the test load and measuring response times. Don't laugh, I have seen stranger screw-ups in hectic test labs than this one. For example, a technician could inadvertently swap a slower processor chip for a faster one on the wrong platform, then tells us the test platform has now been upgraded ready for the next response time measurement. The response time won't improve, though, because the faster chip has been installed on a different machine than the test machine.

ANALYZING THE TEST DATA
(Continued)

(2) *A and R are related, but we are working in a flat part of the curve which represents their relationship.* (Remember that A and R may not have a linear relationship.)

Let's say, for example, that the benchmark (i.e., the standard load) we are running is a mix of word processing tasks which are not computationally intensive. We change the processor speed from 2 Ghz (A lo) to 2.5 Ghz (A hi). The response time does not improve despite the faster processor chip, because some other factor than the processor speed is the bottleneck.

By contrast, let's say instead that the low and high values of A are processor speeds of 100 MHz and 125 MHz. One processor is 25% faster than the other, which the same ratio of processor speeds as before (2 Ghz to2.5 Ghz). If we now are in a steep portion of the same curve relating A and R, the response time will dramatically change with the processor speed.

Changes in Behavior

Sometimes there are sharp changes in the slope of a curve – a mathematician calls this point in the curve a cusp. There generally is a specific reason for a cusp which we can fairly easily identify - for example, abrupt changes in behavior (cusps) often occur around bottlenecks, so it is a good idea to see if we can find any cusps. Unfortunately, in this example we have only two points on any cause-and-effect curve (i.e., the curve relating A and R when we hold to both B and C fixed values).

If we look at the relationship of B (the size of the semiconductor memory) and R, the influence is more distinct than with A. If B increases from low to high while we hold A and C constant, the response time R decreases by approximately a third.

The relationship of C and R is the most dramatic – R falls by about two-thirds when C changes from its low value to its high value.

To recap, by examining the data presented in the table we have found that:

- Increasing any one of the factors A or B or C improves the response time, R.
- Increasing any combination of A, B and C also improves the response time.
- Changes in A have virtually no influence on R, in the domain observed.

ANALYZING THE TEST DATA
(Continued)

- o B has a moderate impact (R falls by one-third as B goes from low to high).

- o C has a major impact (R falls by two-thirds).

We can conclude that if we are interested in improving response time, there is no point in increasing A. Increasing B will help, but the big pay-off is in increasing the value of C.

The Cost Factor

One factor we have not considered so far is the relative costs of the alternatives. Let's say we are setting up a network of 100 servers on a tight budget, and that upgrading the semiconductor memory (factor B) will cost an additional $15 per server, while going to a dual-processor machine (factor C) will cost $50 per server. In this situation, the memory upgrade is the more cost-effective option.

There is a hidden assumption behind this claim that the memory upgrade is more cost-effective. Since hidden assumptions can be dangerous, we want to bring it into the light of day.

We can improve response time by one-third for $15 per server, or by two-thirds for $50 per server. The assumption is that there is a linear trade-off between improvements in response time and cost, and this assumption is not necessarily true. Instead, the last spurt of productivity gained from faster system response might be vitally important to us, so we would actually prefer to pay $50 to increase response time by two-thirds.

The Data Analysis Process

To recap, the data analysis process follows these steps:

(1) We review and cleanse the data if necessary.

(2) We examine the relationship between any one influencing factor, such as A, and the result, R, while holding the other input variables (B and C) to fixed values. After examining the relationship of A and R, we repeat the process for B and R (when we hold A and C to fixed values), and then do the same for C and R too. Effectively, we are tasking a series of different two-dimensional slices through multi-dimensional data.

(3) We look for cusps – abrupt changes of behavior.

(4) We speculate on the likely reasons for the patterns we find. These speculations are hypotheses about the underlying nature of the system and its environment. We are using the hypotheses to build, test and refine a mental model of how the system works.

(5) Based on these hypotheses and the mental model which is in the process of being built, we decide which new data we need to collect in order to shed more light on the situation, i.e., to either confirm or deny the hypotheses.

Populating the Data Table

Another means to the same end, which at first glance seems to avoid a lot of fuss and bother, is to simply vary one factor at a time during the performance measurements. After we have identified the suspects (i.e, the factors which we think are likely to influence the outcome, such as A, B and C), we can simply vary them one at a time while we lock down the other factors to known fixed values. This way, we can more easily see the cause-and-effect relationship which A has with R, for example.

Unfortunately, this apparently straightforward approach tends to be impractical. The reason is that there are often a large number of suspects. Let's say, for example, that we have N factors which we believe are important, and that N is fairly large: let's say N = 27. We also will assume we want to capture several data points for each factor – we will use M to denote the number of different values that a factor can have. Let's say M = 12. Generally each factor will not have exactly the same number of options or variations, but here for simplicity we will assume the number of options (M) is the same for every factor.

We would then take one of these N factors and vary it while the other (N-1) factors are set to fixed values. In all, we would take M sets of measurements for this one factor. We'd have to repeat this process for every factor, in other words, the total numbers of test runs would be (N * M), which in this case is 27 * 12 or 324 test runs. That is a lot of work.

ANALYZING THE TEST DATA
(Continued)

We could represent the test runs by a table, N columns by M rows, and place a check mark at each table location to indicate that we have tested this combination. The checkmarks would fully populate the table. And this table is a simplification – the actual number of distinct testable combinations is M^N, which quickly becomes an awesomely large number when N and M get much above four each.

Even with the huge reduction from M^N to (N * M) testable options, it usually is not feasible to test them all. We only sparsely populate the table with check marks. Instead of 324, we may have enough time and budget for let's say only 15 test runs. This sparse coverage means that the straightforward approach suggested earlier, where we vary only one factor at a time, will not work – there are not enough data points to support it.

A More Tricky Example

In this follow-up variation on the exercise we also will assume that A, B and C all have three settings, that is, each can be set to either a high, middle or low value by the testers. We will assume that we have measured the system response times three times with each of 15 combinations of A, B and C. (There are 27 possible combinations of A, B and C, but we have measured only 15.) In each measurement, the same work load has been processed but the resulting response times vary because of other factors which we do not control, and may not even know.

This section assumes a basic knowledge of calculus, but you should be able to follow it even if your calculus skills are rusty. The data collected is as follows:

Reading	Values of the Factors			Response Time (seconds) Measured Values	Averages	Variations in Response Times
1.	A lo	B lo	C lo	26.4, 37.6, 30.0	31.3	35.8%
2.	A lo	B lo	C md	52.8, 54.0, 45.6	50.8	17.6%
3.	A lo	B md	C lo	42.0, 40.8, 60.0	47.6	40.3%
4.	A lo	B md	C md	72.0, 60.0, 64.8	65.6	18.3%
5.	A md	B lo	C lo	38.4, 51.6, 34.8	41.6	37.2%
6.	A md	B lo	C md	48.0, 44.4, 43.2	45.2	10.6%
7.	A md	B md	C lo	66.0, 56.4, 55.2	59.2	18.2%

ANALYZING THE TEST DATA
(Continued)

8.	A md	B md	C md	46.8, 49.2, 56.4	50.8	18.9%
9.	A md	B md	C hi	52.8, 54.0, 45.6	50.8	17.6%
10.	A md	B hi	C md	42.0, 40.8, 60.0	47.6	40.3%
11.	A md	B hi	C hi	72.0, 60.0, 64.8	65.6	18.3%
12.	A hi	B md	C md	38.4, 51.6, 34.8	41.6	37.2%
13.	A hi	B md	C hi	48.0, 44.4, 43.2	45.2	10.6%
14.	A hi	B hi	C md	66.0, 56.4, 55.2	59.2	18.2%
15.	A hi	B hi	C hi	46.8, 49.2, 56.4	50.8	18.9%

At first glance, this data is meaningless goobledegook. Our intent is to derive what it means.

Each variation in this table is the spread from the slowest to the fastest response time in a particular row, divided by the average response time for that row. For example, in the first row, the variation is [(37.6 - 26.4) / 31.3], or 35.8%.

Note that although there is only a fairly small amount of data, we need 45 measurements (3 response time measurements for each of the 15 combinations of A, B and C). This means we have to run the performance test suite 45 tines, which is feasible only if we use automated tools.

First Impressions

What is this data telling us? In this case, since there are four variables (including R), we cannot construct a two-dimensional or a three-dimensional graph of this data in order to see the trends. This visual representation is not easily done with four or more dimensions.

By simply examining of the response time data in the prior table, though, we can quickly make the following observations:

- The response time is best (lowest) when A, B and C are all set to their lower values.

- Increasing any one of these three factors, A, B or C, increases the response time.
  - For example, if we move the value of A from low to moderate while B and C are both fixed at their low values, then the response time R increases to 41.6 seconds.
  - On closer inspection, this is true 75% of the time, but there are exceptions to the observation that increasing any one of the three factors increases the response time.

ANALYZING THE TEST DATA
(Continued)

- For example:
  - -- When A and B are fixed at their moderate values, increasing C from low to moderate decreases the response time from 58.2 to 50.8 seconds.
  - -- When B and C are fixed at their moderate values, increasing A from low to moderate decreases the response time from 65.6 to 50.8 seconds.
  - -- When B is fixed at its low value and C is fixed at its moderate value, increasing A from low to moderate decreases the response time from 50.8 to 45.2 seconds.
- So the relationship is not as simple as it appears at first glance -- increasing any combination of A, B and C does not automatically increase R (If R did consistently increase with any increase in combinations of A, B and C, a mathematician would call this a monotonic relationship.)

We might also wonder: it is just a coincidence that the value of 50.8 appears in all three of these exceptions? To proceed with this data analysis, I will get into a little voodoo calculus.

The Influence of these Factors on the Response Time

If A and R are continuous variables, then in calculus we can state the influence of the input factor A on the measured output quantity, R, as:

$$dR/dA.$$

In a similar manner, we can state the influences of B and C as dR/dB and dR/dC respectively.

We can crudely calculate dR/dA using the collected data, as the difference between the values of R when A varies from (a) low to moderate and (b) moderate to high.

XXXXXXXXXXXXXXXXXXX

(a) A Varies from Low to Moderate
(b) A Varies from Moderate to High

Since we have a variety of different measurements of R when A is high, we will take the average response time when A is high. We will do the same when A is low.

In other words, we can calculate the influence of A on R as:

[Average value of R when A is high] - [Average value of R when A is low]

or: [(41.6 + 45.2 + 59.2 + 50.8) / 4] - [(31.3 + 50.8 + 47.6 + 65.6) / 4]

or: dR/dA = 0.37.

Similarly, we can calculate the influence of factors B and C.

This arithmetic reveals that: dR/dB = 13.60, and dR/dC = 8.17.

The conclusions we can draw from this data analysis are as follows:

- If we are interested in reducing R (i.e., speeding the response time), the best way to do it is to minimize factor B, since dR/dB = 13.60.
- Factor A has a only a minor influence on the response time R, since dR/dA = 0.37.

All three of these relationships are positive, that is, the response time R increases as A increases (though very little), and also as B increases or C increases. If the response time *decreased* as any one of these factors increased, then the corresponding influence number for that factor would be negative.

Interactions among the Factors

It is worth examining the combined effects of either A and B together, or A and C, or B and C , or all three (A and B and C together) on the response time R, because these three factors (A, B, and C) probably do not act independently of each other.

Let's examine the interaction between A and C. (For brevity, we will not examine the other interactions, such as between B and C.) In calculus, the effect which C has on the relationship between A and R is expressed as: d/dC (dR/dA).

ANALYZING THE TEST DATA
(Continued)

We can calculate this interaction of A and C approximately, by measuring it with B fixed at two different values, first high and then low, and then averaging the resulting numbers from these sceanarios.

We calculate the relationship of A and R when C is fixed at its high value as follows.

- The relationship of A and R when C is fixed at its high value and B also is fixed at its high value is:
  - When A is high, R = 50.8; when A is low, R = 65.6.
  - In other words, dR/dA when C is high and B is high is (50.8 - 65.6) or -14.8

- The relationship of A and R when C is fixed at its high value and B is fixed at its low value is:
  - When A is high, R = 45.2; when A is low, R = 50.8.
  - In other words, dR/dA when C is high and B is low is (45.2 - 50.8) or -5.6

\

- Thus, the relationship of A and R when C is fixed at its high value is the average of the numbers when B is high and low: [-14.8 - 5.6] / 2, or -10.7.

Similarly, we calculate the relationship of A and R when C is fixed at its low value as follows.

- The relationship of A and R when C is fixed at its low value and B is fixed at its high value is:
  - When A is high, R = 59.2; when A is low, R = 47.6.
  - In other words, dR/dA when C is low and B is high is (59.2 - 47.6) or + 11.6.

- The relationship of A and R when C is fixed at its low value and B is fixed at its low value is:
  - When A is high, R = 41.6; when A is low, R = 31.3.
  - In other words, dR/dA when C is low and B is low is (41.6 - 31.3) or +10.3.

- Thus, the relationship of A and R when C is fixed at its low value is the average of the numbers when B is high and low: [11.6 + 10.3] / 2, or + 11.05.

ANALYZING THE TEST DATA
(Continued)

To restate these findings, dR/dA is - 10.7 when C is high and + 11.05 when C is low. (The difference between these numbers, 0.35, does not exactly match the earlier value for dR/dA of 0.37 because of rounding errors.)

However, these findings appear to contradict the earlier one that A has a only a minor influence on the response time R, since dR/dA was computed earlier as 0.37. In fact, the situation seems to be more complicated than that. The relationship of A and R is significantly negative when C is high, but significantly positive when C is low. In other words, high values of C means that increasing A significantly reduces the response time, but at low values of C the effect is just the opposite.

To compute the effect which C has on the relationship between A and R, [d/dC (dR/dA)], we need to know the high and low values of C, or at least the difference between them.

So we will introduce new information at this time, sand reveal the meaning of C: C is the number of concurrent active users of the system, and the low and high values of C are 1 user and 2 users respectively. We calculate the interaction of A and C as follows:

$$d/dC\ (dR/dA) = [\ (-10.7 - 11.05) / (2-1)\ ],\ \text{or} - 21.75$$

.

ANALYZING THE TEST DATA
(Continued)

Assumptions about the Data

With only two settings available for each variable (high or low), we have assumed that there are linear relationships among the four variables, A, B, C and R. In other words, we can graph the relationships as straight lines. In fact, these are unlikely to be linear. With more intermediate data points between the lows and the highs we can be more sophisticated and use nonlinear models for describing the relationships among the variables. (Calculus assumes that the smaller the scale becomes, i.e., the smaller the differences between the low and high values, the more likely it is that the relationships will be linear. Fans of fractal theories might disagree.)

Curve-fitting software can give us a view of the shapes of the curves connecting the multiple intermediate data points and thus a better sense of what those relationships are likely to be (e.g., the trend line could look similar to an exponential curve). In practice we would look at a sampling of two-dimensional slices through the multi-dimensional data, for example, we might look a the various curves relating A to R for different fixed values of B and C, to gain a feel for the relationship.

The conclusions drawn from the data are true at least within the high-to-low ranges of A, B and C which we used in this analysis, but may not necessarily be true outside these ranges.

Now let's put some names to A, B and C, and see if we can figure out the reasons for this system's behavior. I took this data from a performance measurement project where A, B and C meant the following:

Factor	Meaning	Low Value	High Value
A	Background Noise	No e-mail	E-mail arriving at the rate of one every 5 minutes
B	-------------		
C	Number of Concurrent Active Users	1 user	20 users