Contents

10

Taguchi methods

Genichi Taguchi developed a methodology also called robust design. It is used in manufacturing to drive variance out of a product. Product quality can be much more sensitive to variance than the mean. A table whose legs are all 3mm short is better than one with just one leg 3mm short, because that one will wobble.

If a product has hundreds or thousands of measured attributes, then variance in their levels can cause unpredictable problems. A door that is slightly too big combined while the frame it has to fit into is slightly too small, becomes a severe problem. Quality is not well described by a rectangular region in the space of attributes.

Our prior chapters were mostly about "getting more". By modeling $\mathbb{E}(Y \mid x)$ we could learn how to get more potatoes or survival or speed or learning by choosing x well. In robust design, we seek x to get a lower value of $\mathrm{var}(Y \mid x)$.

The reason to reduce variance is that we want to get closer to a target. Methods that reduce variance could move the mean response away from the target. In the robust design world there is usually some way to adjust the mean onto target after the variance has been reduced. They call it a ***signal factor***. This is not something that we would anticipate just thinking in terms of distributions and regression models and expected mean squares. It may however be quite obvious to people working on a product. For instance, if the quantity of interest is the thickness of paint on a car, then that amount will be more or less directly prportional to the amount of time that the paint is being sprayed and so the mean thickness can be controlled that way after designing a process to reduce variance. That only really works if adjusting the signal factor does not more than undo the variance reduction we have obtained.

10.1 Context and philosophy

The topic of robust design and Taguchi methods is not purely mathematical or statistical. There are elements of philosophy and it takes advantages of empirically observed phenomena common to many physical systems.

The history of the problem is also an important component. In the 1980s there was significant concern in the US about a loss of quality in manufactured products, especially compared to high quality output from Japan. Methods developed by Taguchi in Japan were introduced to the US, especially at AT&T. This was to some extent returning a favor of Deming who had earlier worked to help Japanese manufacturers improve their quality, some decades earlier.

One of the ideas in robust design is to consider a quadratic loss function instead of specification limits. If the ideal value of a variable Y is a target T then instead of keeping score by the fraction of times that $|Y-T| \leqslant \delta$ or more generally that $L \leqslant T \leqslant U$, we could study instead $k(Y-T)^2$ for some value $k > 0$.

This idea is often accompanied by the experience of Sony making television sets in both the US and Japan. It is described in Phadke (1989) who references the newspaper 'The Asahi' from April 17, 1979. The color density of the US made sets were all within the specification limits but were widely scattered in the interval from L to U. For the Japanese sets, 0.3% were outside the interval $[L, U]$ but the distribution was more sharply peaked around the target level T. Phadke (1989) reports that US consumers thought the Japanese made sets had higher quality. (Somehow I remember the story being lower return rates for Japanese-made sets. That would be harder data than just a survey of perceptions, but I'm unable to find that aspect of the story in print.)

A quadratic loss function has several advantages. It allows one to keep track of continued progress beyond the point where 100% of products are meeting a given set of specs, without having to keep 'moving the goalposts'.

Another reason to avoid using a binary spec to track quality is that those specs may well have been set by some arbitrary tradeoffs between the interests of producers and consumers of a product along with its costs. A spec like $T \pm \delta$ might not really be from a law of nature, but some negotiation based on what is possible or reasonable. Then getting zero defects does not imply perfect quality. A rectangular spec on many features, such as $|Y_j - T_j| \leqslant \delta_j$ for all $j = 1, \dots, J$ is even less likely to be physically appropriate.

Some sort of binary rule is necessary if one has to accept some product and reject others. However continuous data are also more informative than binary data when it comes to understanding and improving a system. Even with a spec such as 0.013 ± 0.001 getting values $(0.01301, 0.01303, 0.01298, 0.01302)$ inspires more confidence than just getting 'OK' four times out of four tries.

The quadratic loss is sometimes motivated as the total cost to society of a product missing its target value. It is difficult to judge the whole cost to society of a product, but perhaps reasonable that a quadratic formula is more useful than a binary one.

For some quantities Y that cannot be negative this smaller they are the

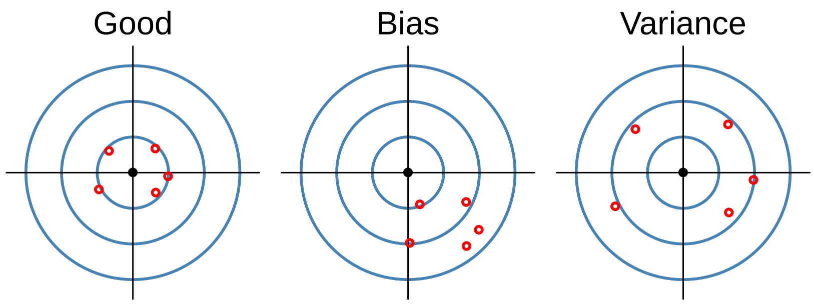

Figure 10.1: An archery example to illustrate bias and variance. The first archer is quite good, or at least best. The second has higher bias. The third has higher variance.

better they are. These then have a target of $T = 0$ and are scored by kY^2. For other quantities, the larger they are, the better, and those are scored via kY^{-2}. That is smaller values of $1/Y$ are better.

10.2 Bias and variance

A great convenience of a quadratic loss function is that it splits into variance and bias squared. Given n observations we have

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - T)^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 + (\bar{Y} - T)^2.$$

The first, variance term, is generally harder to control. A common analogy is made to a coach in archery as shown in Figure 10.1. The coach can might be able to help the second archer with one or two suggestions about where to aim. The third archer might have issues that involve many variables, just like the noise in a model is often the result of many uncontrolled variables.

Whether the archery model is appropriate to manufacturing is not something we can prove philosophically or mathematically. It might be known to hold empirically in some problems. Then the strategy in robust design is to first find a way to reduce the variance. Then later we fix any bias created by the first step assumed to be the hard one (without causing the variance to go back up).

One model to explain robust design has

$$Y = f(x_1, \dots, x_k,\, z_1, \dots, z_n) + \varepsilon = f(\boldsymbol{x}, \boldsymbol{z}) + \varepsilon$$

where $\boldsymbol{x} = (x_1, \dots, x_k)$ is a vector of variables that we can control and $\boldsymbol{z} = (z_1, \dots, z_r)$ is a vector of ***noise variables*** that we cannot control. At first this sounds like it will be tough to experiment on variables we cannot control. What is happening is that those are variables we can control in an experiment but not afterwards. In an example mentioned by Shin Taguchi in a round table discussion (Nair et al., 1992) a company making a paper feeder has noise variables including "paper type, paper size, paper warp, paper surface, paper alignment, stack height, roller wear, and humidity." They can vary all of those things in their experiment to understand how they affect paper jams, but they cannot control them in their customers' uses once the product has been sold. Variables $\boldsymbol{x}$ are different. They might describe how the paper feeder was constructed. In an automotive application, $\boldsymbol{z}$ might include the weather that a car will be driven in.

In robust design, we get to choose $\boldsymbol{x}$ but our product has to work very generally for lots of $\boldsymbol{z}$. We might formulate the problem as

$$\min_{\boldsymbol{x}} \mathrm{var}(Y \,|\, \boldsymbol{x})$$

where the randomness in Y is from an assumed distribution on $\boldsymbol{z}$ and from ε. Then change some ***signal variable*** to adjust $\mathbb{E}(Y \mid \boldsymbol{x}) = T$. That variable cannot be one of the z_j. It must be one of the x_j or some other variable either implicitly or explicitly in $f(\cdot)$. A good example is picking $\boldsymbol{x}$ to minimize variance of the paint thickness on a car. Then if the mean is off target, adjust the spray time.

There is a very famous example about the Ina Tile company. See Phadke (1989). Their tiles came out of the kiln in unequal sizes and quality. One of the noise factors was the position of a tile in the kiln. The tiles were packed within the kiln. No matter how you do that some tiles will be central and some at the edge. So that position z could be a noise factor that you don't want to affect product quality. After doing a robust design experiment they found that they were able to reduce the variability of their tiles. The solution even involved using less of an expensive ingredient and more of an inexpensive one and so the result was greater quality at lower cost.

10.3 Taylor expansion

Suppose that we set targets $\boldsymbol{x}_0$ for $\boldsymbol{x}$ and $\boldsymbol{z}_0$ for $\boldsymbol{z}$. Under random assignment $\mathbb{E}(\boldsymbol{z}) = \mu$ which is not necessarily $\boldsymbol{z}_0$ and $\mathrm{var}(\boldsymbol{z}) = \mathrm{diag}(\sigma_1^2, \dots, \sigma_n^2)$. Here we are making a great simplification that components z_j are uncorrelated. We could weaken that. With this assumption a Taylor expansion gives

$$f(\boldsymbol{x}_0, \boldsymbol{z}) \doteq f(\boldsymbol{x}_0, \boldsymbol{z}_0) + \sum_{j=1}^n \frac{\partial}{\partial z_j} f(\boldsymbol{x}_0, \boldsymbol{z}_0)(z_j - z_{0j})$$

and then

$$\mathrm{var}(f(\boldsymbol{x}_0, \boldsymbol{z})) \approx \sum_{j=1}^n \Bigl(\frac{\partial}{\partial z_j} f\Bigr)^2 \sigma_j^2.$$

Since $\boldsymbol{z}$ and hence σ_j^2 is out of our control, our best chance to reduce this variance is to reduce $(\partial f/\partial z_j)^2$ through our choice of $\boldsymbol{x}_0$. That is, we want to reduce the sensitivity of our output to fluctuations in the input.

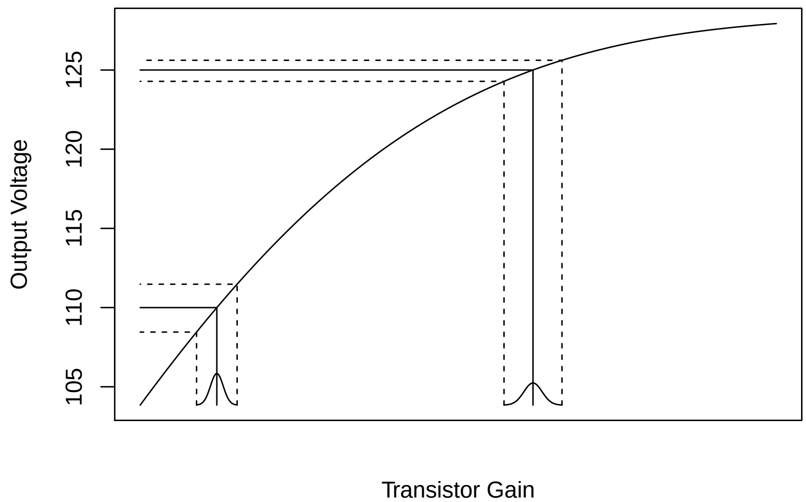

Figure 10.2: Hypothetical curve modeled after the account in Phadke (1989) of power supply design.

Phadke (1989) has an example where z is gain of a transistor Y is output voltage of a power supply. Figure 10.2 illustrates his setting. The horizontal axis depicts a transistor gain quantity z (without units). There is a curve that translates this gain into output voltage $Y = f(z)$. There is a value z_{110} that gives $Y = f(z_{110}) = 110$, the desired value, but in that region z has some noise that gets amplified by the steep slope $f'(z_{110})$ to give a very noisy voltage. Moving to a higher value of z, such as z_{125} with $f(z_{125}) = 125$ leads to less noisy Y. This happens despite increased variance in z there because $f'(z_{125})$ is much smaller than $f'(z_{110})$. In order to bring the system on target he uses a resistor that reduces the voltage from a value centered around 125 to one centered around the target $T = 110$. The impact of this signal factor does not change variance due to linearity of the voltage versus resistance relationship.

10.4 Inner and outer arrays

Taguchi's strategy is to choose an experiment varying $\boldsymbol{x}$ in n_1 runs. This is called the ***inner experiment***. For each of those n_1 runs he has a second experiment in the noise factors $\boldsymbol{z}$ using n_2 runs. That is the ***outer experiment***. The total experiment then has $n_1 \times n_2$ runs, yielding y_{ij} for $i = 1, \dots, n_1$ and $j = 1, \dots, n_2$.

At each level of the inner experiment, the method computes

$$\eta_i = -10 \log_{10}\Bigl(\frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(y_{ij} - \bar{y}_{i\bullet})^2\Bigr).$$

This is the sample variance over the outer experiment recorded in decibels. Recall that $10\log_{10}(\cdot)$ converts a quantity to decibels. A logarithmic scale is convenient because any model we get for $\mathbb{E}(\eta)$ when exponentiated to give a variance (or inverse of a variance) will never give a negative value. It is also more plausible that the factors in $\boldsymbol{x}$ might have mutiplicative effects on this variance than an additive effect.

In its most basic form Taguchi's method finds settings that maximize the signal to noise ratio η. It is often an additive model in the components of $\boldsymbol{x}$. In the "bigger the better" setting, the analysis is of

$$\eta_i = -10\log_{10}\Bigl(\frac{1}{n_2}\sum_{j=1}^{n_2}\frac{1}{y_{ij}}\Bigr).$$

while in the "bigger the better" setting, the analysis is of

$$\eta_i = -10\log_{10}\Bigl(\frac{1}{n_2}\sum_{j=1}^{n_2}y_{ij}\Bigr).$$

The experimental designs are usually orthgonal arrays like the following one at 3 levels:

0	0	0	0
0	1	1	2
0	2	2	1
1	0	1	1
1	1	2	0
1	2	0	2
2	0	2	2
2	1	0	1
2	2	1	0

This design is called an orthogonal array because in any pair of columns each of the 9 possible combinations appears the same number of times, i.e., once. We will see more orthogonal arrays when we look at computer experiments. We recognize this as a 3^{4-2} design because it handles 4 variables at 3 levels each using only 9 runs not 81.

It is usual for the outer experiment to also be an orthogonal array. Then the design is made up of an ***inner array*** and an ***outer array***.

There is a good worked example of robust design in Byrne and Taguchi (1987) (which seems not to be available online). The value Y was the force needed to pull a nylon tube off of a connector. This pull-off force was studied with an inner array varying 4 quantities at three levels each (interference, wall thickness, insertion depth and % adhesive). The outer array was a 2^3 experiment in the time, temperature and relative humidity during conditioning. There were thus 72 runs in all.

10.5 Controversy

There were some quite heated discussions about which aspects of Taguchi's robust design methods were new and which were good. The round table discussion in Nair et al. (1992) includes many points of view and dozens of references.

One issue is whether it is a good idea to use that combination of inner and outer arrays. There every value of $\boldsymbol{x}$ is paired with every value of $\boldsymbol{z}$. It might be less expensive to run a joint factorial experiment on $(\boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R}^{k+n}$ instead. As usual, costs come into it. If the cost is dominated by the number of unique $\boldsymbol{x}$ runs made, then a split plot structure like Taguchi uses is quite efficient. If changing $\boldsymbol{z}$ and changing $\boldsymbol{x}$ cost the same then a joint factorial experiment and the corresponding analysis will be more efficient.

Another issue is whether an analysis of signal to noise ratios is the best way to solve the problem. In the roundtable discussion James Lucas says "The designs that Taguchi recommends have the two most important characteristics of experimental designs: (1) They have factorial structure, and (2) they get run." That is, an analysis that is not optimal may be better because more people can understand it and use it.

Bibliography

Byrne, D. M. and Taguchi, S. (1987). The taguchi approach to parameter design. *Quality progress*, 20(12):19–26.

Nair, V. N., Abraham, B., MacKay, J., Box, G., Kacker, R. N., Lorenzen, T. J., Lucas, J. M., Myers, R. H., Vining, G. G., Nelder, J. A., Phadke, M. S., Sacks, J., Welch, W. J., Shoemaker, A. C., Tsui, K. L., Taguchi, S., and Wu, C. F. J. (1992). Taguchi's parameter design: a panel discussion. *Technometrics*, 34(2):127–161.

Phadke, M. S. (1989). *Quality engineering using robust design.* Prentice Hall, New York.