Improving Stable Processes
Professor Tom Kuczek
Purdue University

Using process knowledge to identify uncontrolled variables and control variables as inputs for Process Improvement
Process may be off Target or Have Excess Variation

• X-double bar is the estimate of the process mean which may be off target.
• Sigma(X) is the estimate of Common Cause Variation.
• Both of these contribute to the Capability of the process, $C_{pk}$. 
Improving Common Cause

- Common causes of variation usually cannot be reduced by trying to explain differences between values when the process is stable.
- Uncontrolled variation and control variables must be understood to partition Common Cause Variation into basic sources.
- Stable processes will require some degree of change to improve Common Cause.
Partitioning Uncontrolled and Controlled Variables into Sources
Sources of variation

Describe where in the preceding flowchart one can find sources of variation:

•
•
•
•
•
Analyze Between and Within Supplier Variation

• Different suppliers are a *between* source of variation.
• Raw materials from a single supplier is a *within* source of variation.
• Control charts can be used to look for special cause between suppliers to reduce variability between suppliers.
Variables in the production process may be *uncontrolled* variables or *control* variables.

*Uncontrolled* variables are variables which may affect the output of the process, but which are not currently controlled.

*Control* variables are variables such as process settings which affect the outcome of the process.
Partitioning Uncontrolled and Controlled Variables into Sources
Controllable and uncontrollable

Where in the flowchart are variables controllable and/or uncontrollable and to what extent?

•

•

•

•
Output Variables

• Output Variables are measurements of the resulting product.
• The chosen measures for the product are measures of the product characteristics important to the customer.
• Customers may be *internal or external* to the organization.
Part I: Reducing Output Variation Around the Target

Output variation of the product may be broken down into two sources:

1. Actual variation of the “true” product characteristic, often around a target value, usually designated by the symbol $\tau$.

2. Variability in the measurement process, which may introduce bias or added variation to the measurement of the characteristic, which occurs in the measurement process itself.
Measuring things

Give examples of products/services that we measure for quality characteristics and how:

•
•
•
•
•
Product Characteristic Variation: Parameter Design

Let us first concentrate on the product characteristic value of interest to our customer. There are two main issues here:

1. To center our product as close to the target value, $\tau$, as possible.

2. To minimize the variation around the target value.
Targets

Give examples of *targets* we want to get close to (sometimes we want to maximize or minimize something):

•

•

•

•
Modeling the Output

In order to put the product characteristic of interest to our customer on target with minimum variation, we need a Statistical Model.

Denoting:

$Y = \text{product characteristic,}$

then we need to model $Y$ as a function of the important inputs to our process.
Input to the Model

Variables important to the Production Process are either *uncontrolled variables* or *control variables*.

*Uncontrolled variables* would include variation in *raw materials* or *environmental* conditions during process operation.

*Control variables* would include any fixed settings for machines involved in the production process.
Control and uncontrolled variables

Give examples of variables we can control and not control in a process:

•
•
•
•
•
Model Form

The model would express product output $Y$, as a function of *uncontrolled* and *control* variables in a form such as:

$$Y = f \left( \text{uncontrolled variables, control variables} \right)$$

where $f(, , )$ generally denotes a simple mathematical function, such a regression model.
Model Building

In order to build even the simplest model for the output variability of $Y$, we need a set of data with the values of the uncontrolled and control variables and the resulting output measure $Y$. We may use either:

- **Exploratory data analysis** using existing data to begin with, or
- **Experimental design**, where we use pre-determined values of the uncontrolled variables (temporarily fixed for the experiment) and control variables to give us an optimal model.
Exploratory data analysis

Where have you seen an existing data set explored for patterns and how?

•

•

•

•

•
Experimental Design

• While generally more expensive than *Exploratory data analysis*, *Experimental Design* provides much better models for estimating the mean and variance of a process.

• In fact, *Experimental Design* can provide ways to use the control settings of a process to compensate for the uncontrolled variation in process inputs and variable operating conditions. This allows us to put a process on target with minimum variation.
Two Main Classes of Designs

• *Taguchi’s Robust Parameter Design*. In the 1980’s, Genichi Taguchi proposed a class of designs which included uncontrolled and control variables in order to choose optimal control settings.

• *Response Surface Methods*. Due to some deficiencies in Taguchi’s approach, *Response Surface Methods* were later proposed as an alternative since they were more cost effective and allowed one to simultaneously model the mean and variance of the process.
Why design?

Either type of design is supposed to improve the capability of the process, $C_{pk}$. What are some reasons capability is important?

•

•

•
The Taguchi Loss Function

Whichever type of design we choose to optimize our product characteristic, $Y$, our primary goal is to find control settings which put our process mean on target, $\tau$, with minimum variance. This means we want to minimize the function:

$$L(Y) = K(Y - \tau)^2$$
Taguchi Loss Function, cont.

On average, since $K$ is a proportionality constant, we need to minimize:

$$E(Y - \tau)^2 = Var(Y) + (E(Y) - \tau)^2$$

where

$$(E(Y) - \tau)$$

is often referred to as the *Bias*, or the average deviation from target.
Why partition?

We partition for two reasons:

• Some control variables allow us to put the process on target.
• Some control variables allow us to control variation.
Taguchi’s Robust Parameter Design

In this design, parameters are inputs to the process which are likely to affect the product quality characteristic, Y. Parameters are divided into two groups:

• **Control variables**—would include any fixed settings for machines involved in the production process.

• **Noise variables**—would include any *uncontrolled variables*, such as environmental conditions which could affect process outcome.
Choosing Parameter Levels

• *Control variables*- levels (settings) of control variables are chosen which span available operating settings.

• *Noise variables*- levels are chosen and temporarily fixed for each of the noise variables. These levels are chosen to represent values of the noise variables actually observed during the production process.
Pick a process

What are some

• Control variables:

• Noise variables:
Why Control *and* Noise Variables?

- Noise variables contribute to the response and therefore contribute to Common Cause variation.
- Certain settings of the control variables may minimize the effect of the uncontrolled variation of the noise variables, thereby reducing Common Cause.
- The control variables are used to control the mean of the process and so can be used to put the mean near Target.
Forming Design Arrays

Arrays are combinations of levels of parameters at which we will collect data. These parameter combinations are expressed in matrix form.

- The *inner array* contains the combinations of levels of the *control variables* at which we will collect data.
- The *outer array* contains the combinations of the levels of the *noise variables* at which we will collect data.
Example of Design Array Matrix

The columns of the matrix here represent two parameters. Each parameter has a low (-1) value or a high value (+1).

\[
\begin{bmatrix}
+1 & +1 \\
-1 & +1 \\
+1 & -1 \\
-1 & -1 \\
\end{bmatrix}
\]
Levels

What are some high and low levels for a process?

•

•

•

•
Example, cont.

• The rows of the matrix represent the levels of the parameters at which we take observations. Four rows means that there will be four parameter combinations at which we will take observations. These are called *design points*.

• The columns of the matrix correspond to the different variables in the Array.
Structure of Arrays

The structure of each array is typically either:

- A *factorial design* - this is where all possible combinations of parameters and their respective levels are included.

- A *fractional factorial design* - this is where only a fraction of the combinations of a *factorial design* are used due to cost considerations.
Crossing Arrays

Once the Inner and Outer Arrays are determined, they are then crossed. This means that every combination of the control variables in the experiment will be measured with every combination of the noise variables in the experiment.
Crossing Arrays and Interaction

• When the inner and outer arrays are crossed, then we can determine if the control variables and noise variables interact.

• If interaction(s) exist between control and noise variables, then control variables can be used to reduce the effect of the noise variables on the process, thereby reducing Common Cause.
Why interaction?

If variables interact, we can use control variables to compensate for the variation in noise variables, such as raw material.

Now we can compensate for things we can’t control, like raw material variability, using things we can control, like process settings.
Forms of Interaction

Interaction can take many forms, but two of the most common and important are *antagonism* and *synergy*.

- **Antagonism** occurs when two variables tend to cancel each other out.
- **Synergy** occurs when two variables tend to have a multiplicative effect.
Interaction as Antagonism

![Graph showing antagonism between Drug A and Drug B at different doses.]

- X-Axis: Drug A
  - 0 mg
  - 5 mg
- Y-Axis: Response
  - Drug B
    - 0 mg
    - 10 mg

The graph illustrates the antagonistic interaction between Drug A and Drug B at various doses.
Interaction as Synergy

![Graph showing the interaction between Drug A and Drug B. The Y-axis represents Response, and the X-axis represents Drug A dose (0 mg, 5 mg). The graph shows two lines: one for Drug B at 10 mg and another for Drug B at 0 mg. The lines illustrate the synergistic effect at different doses of Drug A.](image)
Synergy and antagonism

What are examples you have seen of synergy or interaction?

•

•

•
Criticisms of Taguchi

• The Taguchi method of crossing arrays can lead to large and thus, very expensive designs.
• It encouraged a “pick the winner” mentality of choosing from among the control settings in the design array.
• It did not provide fundamental process knowledge to pick “optimal” control settings.
• Fixing levels of *uncontrolled* variables can be impractical.
Response Surfaces Alternative

Goals:

• Model the response, $Y$, as a regression-type function of the control and noise parameters.

• Use recorded data on the distribution of the noise variables to model the mean and variance of the response, $Y$.

• Pick optimal control variable settings to put the process mean on target and minimize the variation due to noise variables.
Simple Example

• $Y = $ Lamina thickness
• $X_1 = $ Viscosity (control variable)
• $X_2 = $ Ambient humidity (noise variable)

Fitted model:

$$
\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_{12} X_1 X_2
$$
Example, cont.

In this case we can choose values of $X_1$ to minimize the Taguchi loss function by using the estimated Mean and Variance of $\hat{Y}$

\[ E(\hat{Y}) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_{12} X_1 \cdot E(X_2) \]

\[ Var(\hat{Y}) = (\hat{\beta}_{12} X_1)^2 \cdot Var(X_2) + \sigma_e^2 \]
Example, cont.

In this particular case we may not get the *lamina thickness* exactly on target when we minimize the loss function due to the estimated relationships. But there are cases where we can use one *control* variable to minimize the variance, then use another *control* variable to put the mean on *target*. *This is the best of all possibilities.*
More Complex Example

Let us suppose that we have two control variables:

\[ X_1, X_2 \quad - \text{Control} \]

and one noise variable:

\[ X_3 \quad - \text{Noise}. \]

We then fit a slightly more complex equation to our data.
Example 2, cont.

Assume our fitted model is now

\[ \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_{23} X_2 \cdot X_3 \]

Now that we have two control variables and one of them interacts with the noise variable, we can use them separately to put the mean on target and to minimize variation.
Example, cont.

Since

\[ E(\hat{Y}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{23} X_2 \cdot E(X_3) \]

and

\[ Var(\hat{Y}) = (\beta_{23} X_2) \cdot Var(X_3) + \sigma_e^2 \]

We can set \( X_2 \) to minimize the variance and set \( X_1 \) to put the mean on target.