A RESPONSE SURFACE APPROACH TO DATA ANALYSIS

IN ROBUST PARAMETER DESIGN

by

Yoon G Kim

Dissertation submitted to the Graduate Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

R. H. Myers, Chairman

K. H. Hinkelmann

G. I. Holtzman

R. G. Krutchkoff

W. R. Pirie

September, 1992

Blacksburg, Virginia

ABSTRACT

It has become obvious that combined arrays and a response surface approach can be effective tools in our quest to reduce (process) variability. An important aspect of the improvement of quality is to suppress the magnitude of the influence coming from subtle changes of noise factors. To model and control process variability induced by noise factors we take a response surface approach. The derivative of the standard response function with respect to noise factors, i. e., the slopes of the response function in the direction of the noise factors, play an important role in the study of the minimum process variance. For better understanding of the process variability, we study various properties of both biased and the unbiased estimators of the process variance. Response surface modeling techniques and the ideas involved with variance modeling and estimation through the function of the aforementioned derivatives is a valuable concept in this study. In what follows, we describe the use of the response surface methodology for situations in which noise factors are used. The approach is to combine Taguchi's notion of heterogeneous variability with standard design and modeling techniques available in response surface methodology.

ACKNOWLEDGEMENT

There are many who have been helpful and supportive throughout my graduate years. First and foremost I would like to thank my parents for their sacrifices and support to make it possible for me to continue my education.

I would like to thank Dr. R. H. Myers, my major professor, who gave me a new meaning of what major professor really is. He has been like a father to me. He has done so much for me, not to mention his many suggestions for this work. He and I wrote this together, working side by side over the phone and by mail - all despite being physically separated by 3,011 miles. He not only produced his share of suggestion but offered counsel and advice in my portions as well. I cannot conceive of tackling this work without him. It is his work as much as mine, and he deserves at least as much credit.

I would like to thank Dr. K. H. Hinkelmann for all his support and encouragement. I owe what I am and what I will be to the professors and friends of the department. My experience of the last four years here at Virginia Tech is something I will treasure for a long time.

The committee members also gave me many valuable suggestions. That this work merits their approval means a great deal to me.

Table of Contents

Chapter 1. Introduction and Overview of Taguchi Methods

"Statistics is concerned with the variability that is evident in any body of data." (Searle, Casella, and McCulloch, 1992). The modern approach to product improvement and optimization can surely borrow on principles from the Japanese quality consultant Genichi Taguchi. Articles by Taguchi (1977, 1988), Taguchi and Wu (1985), Kacker (1985), and many others described and highlighted Taguchi's work. His approach puts more emphasis on *product variability* in the statistical modeling scheme. The root of the idea is the notion that products lack in quality because of inconsistency in performance. This inconsistency is produced by factors that are uncontrollable in the design of the product, e. g., tolerance on design factors, environmental factors, or factors that are a function of usage by the consumer.

There are machines whose handling is somewhat unstable and thus they are undependable. Some machines are tough and easily overcome adverse environments. From this point of view, as a measure of quality, or the goodness of product design, Taguchi introduced the concept of the signal-to-noise ratio (SN-ratio) in the 1960's.

It is important to choose the most stable product design among many possible designs that satisfy the specifications, that is, the kind of design that is not easily influenced by fluctuations in manufacturing environment, changes in components or materials, etc. This kind of design is said to be resistant (robust) to the environment. What we need is not just an emergency measure after some troubles occur, but, from the very design stage, the type of design that can overcome fluctuations in manufacturing conditions, and that will not easily falter when exposed to changes in environmental conditions. Taguchi has proposed parameter design (also known as robust parameter design) using SN-ratio to systematically solve the aforementioned problems. Taguchi proposed how one should suppress the magnitude of the influences coming from subtle changes in environment and/or changes in manufacturing conditions by manipulating the design parameters. Design parameters are also commonly called process variables.

In this chapter we will discuss the role of Taguchi methods in quality engineering. The current and proposed research deals in the extension of response surface methodology (RSM) as an alternative to or an augmentation of the Taguchi approach. The use of RSM begins in Chapter 2.

1.1 Location and Dispersion Effects

In recent years much attention has been paid to the Taguchi methods. These concepts have been successfully applied to improve the quality of industrial processes in Japan (Taguchi, 1988). Taguchi makes use of statistically planned experiments to identify the settings of product and process parameters that reduce the performance variation of a product characteristic around the intended target value. The performance characteristic of a manufactured product is affected by many factors. Some factors affect the mean value of the performance characteristic. These factors are identified as *location effects*. Others might affect the variation of the characteristic. These are called *dispersion effects.* There are certainly some factors that are both location and dispersion effects.

The understanding of dispersion effects and variance modeling has progressed in large part due to Taguchi's preoccupation with reducing variability. Contrary to popular belief, Taguchi neither introduced variance modeling nor invented the notion of squared error loss. However, the attention drawn to these concepts by the Taguchi approach certainly influenced Box and Meyer (1986), Nair and Pregibon (1988), Carroll and Ruppert (1988, 1991), and many others. Since publication of the now classical paper by Bartlett and Kendall (1946), very little had appeared that dealt with modeling and controlling process variance until Taguchi. This important way of thinking will continue to be reflected in courses taught in the university as well as in an industrial setting. Diagnostic work by engineers, e. g., normal probability plotting, now often involves inclusion of dispersion effects. These ideas are natural for engineers. Variance modeling and dispersion effects have the potential of making sequential experimentation more informative, albeit more complicated. More work will emerge in this area.

1.2 Control Factors and Noise Factors (Robust Parameter Design)

Design factors can be classified as *control factors* and *noise factors.* Control factors, often denoted by $\underline{x}$, can be controlled in an experiment and also in a real application. Noise factors, denoted by $\underline{z}$, can be controlled in an experiment, but may not be controllable in the actual process. Noise factors are often environmental factors such as humidity conditions, properties of raw materials, product aging, etc. They often characterize how consumers handle a product. For example, in a study to develop a tasty cake recipe, amount of flour, amount of shortening and number of eggs are control factors, while different cooking time and oven temperature are noise factors (Box and Jones, 1990). Other examples along with corresponding responses are shown below.

Automotive industry	Response	Engine efficiency
	Control factors	Components for building engine part
	Noise factors	Different types of gasoline Outside temperature Different styles of driving
Food/Tobacco industry	Response	Taste of a product
	Control factors	Ingredients
	Noise factors	Amount of milk or water added Baking time Things added by consumers
Petroleum industry	Response	Fuel efficiency
	Control factors	Ingredient concentrations
	Noise factors	Different types of automobile Road conditions
Chemical industry	Response	Performance of a blend
	Control factors	Ingredient concentrations
	Noise factors	Variability in temperature Different types of solvents

The overall objective of quality engineering is to manufacture products that are robust to all sorts of noise factors. *Taguchi's approach is to determine the levels of control factors at which the effect of the noise factors on the performance characteristic is minimized.* This approach has been called *robust parameter design* (RPD) (Taguchi, 1977). "This robustness to noise factors can be incorporated only at the product design stage, which, of course, ideally is the stage in which experimental design is most useful." (Myers, Khuri, and Vining, 1992).

1.3 Orthogonal Arrays

Taguchi's approach has led to performance measures that combine mean response and variability produced by variation in the noise factors. His designs are based on orthogonal arrays, an "inner array" for the control factors and an "outer array" for the noise factors. The inner and outer arrays are cross-classified, resulting in a relatively large experiment. Figure 1.1 provides an illustration of $2^2 \times 2^2$ *crossed array* in which two control factors are crossed with two noise factors, thus establishing 16 experimental points.

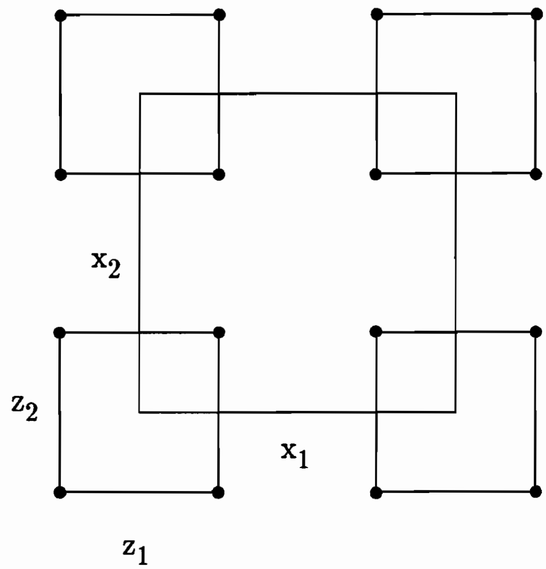


Figure 1.1 Taguchi's $2^2 \times 2^2$ Crossed Array

Taguchi's design approach is to incorporate simultaneously as many factors as possible in an initial experiment. The noise factors involved should be varied over levels likely to occur in the actual process. Taguchi recommends constructing the crossed array by using "orthogonal arrays" (Rao, 1947). Useful orthogonal arrays have been previously developed by Plackett and Burman (1946), Box and Hunter (1961a, 1961b), and many others. A good discussion of constructing orthogonal arrays is in Raghavarao (1971).

1.4 Signal-to-Noise Ratio (SN-Ratio)

Much of Taguchi's focus has been on computing a summary statistic (single performance criterion) at each inner array design point. Quadratic loss functions of the type $E[(y-t)^2]$ where t is some some chosen target value have been used as a possible performance criterion (Taguchi (1977, 1988), Taguchi and Wu (1985)). One of the main points of the Taguchi methods is their heavy reliance on *signal-to-noise ratios.*

In the Taguchi analysis one fits a model of main effects in the control factors with the SN-ratio at each inner array point as the response to be optimized. His principle is to choose the settings of the control factors for which the SN-ratio is maximized. The term "pick the winner" has been used to describe this optimization procedure. This approach essentially involves a "main effects only" model with the SN-ratio being the response. There are three basic scenarios considered by Taguchi in the development of SN-ratios.

- Target is best: The experimenter is interested in achieving a certain target value for the response.
- Larger the better: The experimenter is interested in maximizing the response.
- Smaller the better: The experimenter is interested in minimizing the response.

1. Target is best

Assuming that we can take the mean to the target by manipulation of at least one of the control factors, we use SN-ratio $10log_{10}\eta$, where $\eta = \frac{\overline{y}^2}{s^2}$ for sample responses y_1, y_2, ..., y_n taken at the outer array observations. Kacker (1985) pointed out that in cases where the response variance and mean are independent, one or more factors (adjustment factors) can be used in order to eliminate response bias, that is, the adjustments result in $E(y) = t$. In this case the SN-ratio often used is $10 log_{10}\eta$, where $\eta = \frac{1}{s^2}$ or equivalently $-10 log_{10}s^2$.

2. Larger the better

For characteristics that cannot be negative and bigger is better, one may use SN-ratio $-10 log_{10}\eta$, where $\eta = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{y_i}\right)^2$, where the sum is over the outer array observations.

3. Smaller the better

For characteristics that cannot be positive and smaller is better, again one may use SN-ratio $-10 log_{10}\eta$, where $\eta = \frac{1}{n}\sum_{i=1}^{n} y_i^2$, where the sum is over the outer array observations.

Only the "target is best" case is a true signal-to-noise ratio in a strict sense and the reciprocal of η is the squared coefficient of variation. The base of the logarithm and multiplier do not influence the subsequent analysis and can be chosen strictly for convenience. Taguchi's claim is that these ratios can identify active location and dispersion factors and that their maximization will minimize variability but still achieve desired results on the mean response. In each case there is an accompanying analysis of $\overline{y}$ to identify active adjustment factors. Adjustment factors are those which are known to influence the mean but not variability.

Taguchi's reliance on the SN-ratio to analyze mean and dispersion effects has received considerable criticism. Pignatiello and Ramberg (1985) point out that the use of SN-ratio implies the somewhat bold assumption that a unit increase in

$\log(\overline{y}^2)$ is of equal importance as a unit decrease in $\log(s^2)$. They conclude that it would be better to simply study the variability itself. A general criticism of performance statistics has been voiced by Lucas (1985) and Box (1985). Among the criticism is that the SN-ratios do not adequately separate mean from dispersion. Thus we do not gather enough information about the process.

Example 1.1 (Taguchi's Pick the Winner)

In the following example of color-TV decoded signals, the basic principle of Taguchi's RPD will be illustrated. In the transmission of color-TV signals the quality of the decoded signals is determined by the PSNR's (power signal to noise ratio in electronics engineering) of the image transmitted.

The "process" here refers to different images transmitted (coarse vs. detailed image) and the voltage consumers use after purchasing a TV monitor (100 vs. 200 volts). Let us consider one simple example before going into the details.

The following data are generated by crossing a 3^2 factorial design in the controls with a 2^2 factorial design in the noise factors. This type of design is referred to as a crossed array. The responses (in dB), which indicate the quality of reception of transmitted signals at each factor level, and Taguchi's SN-ratios for "larger the better" case appear in Table 1.1. Larger responses indicate better reception.

Table 1.1

Color-TV Image Data and SN-Ratio

	Coded Levels -1	0	1
x_1 (Number of tabs in a filter)	5	13	21 (tabs)
x_2 (Sampling frequencies)	6.25	9.875	13.5 (MHz)
z_1 (Number of bits of an image)	256		512 (bits)
z_2 (Voltage applied)	100		200 (Volts)

Factor Combination	x_1	x_2	z_1 -1 / z_2 -1	z_1 -1 / z_2 1	z_1 1 / z_2 -1	z_1 1 / z_2 1	SN-ratio
(1)	-1	-1	33.5021	41.2268	25.2683	31.9930	29.9756
(2)	-1	0	35.8234	38.0689	32.7928	34.0383	30.8854
(3)	-1	1	33.0773	31.8435	36.2500	34.0162	30.5485
(4)	0	-1	30.4481	41.2870	15.1493	23.9883	27.1218
(5)	0	0	34.8679	40.2276	27.7724	31.1321	30.2586
(6)	0	1	35.2202	37.1008	33.3280	35.2085	30.9157
(7)	1	-1	21.1553	34.1086	0.7917	15.7450	3.9725
(8)	1	0	27.6736	38.1477	15.5132	25.9873	27.1960
(9)	1	1	32.1245	38.1193	26.1673	32.1622	29.9093

Taguchi's SN-ratio "larger the better" case $-10log_{10}\frac{1}{n}\sum\left(\frac{1}{y_i}\right)^2$ indicates that the reception of the transmitted signals is least sensitive to changes in the noise factors at $(x_1, x_2) = (0, 1)$ combination. Factor combination (6) therefore is a recommended operating condition from the RPD point of view. Notice however, factor combination (2) and to a less degree (3) show similar values of the SN-ratios.

If we recommend only factor combination (6) as the "winner" from this experiment, we must be wasting information available from a model of the experiment. A regression model would allow one to interpolate in the design space. At times "pick the winner" strategy may identify as optimum a location where we don't have data. This can occur because the inner array is typically a highly fractionated factorial.

Example 1.2 (Use of SN-Ratio)

In this example, the experimenter wants to determine the settings of the control factors that are good at hitting the target and robust to changes in noise factors. The experiment was conducted by crossing 2^2 noise factor combinations with each level of nine control factor combinations that are heavy fractions of 3^4 factorials. The nine combinations of control factors are referred to as L_9 by Taguchi.

Table 1.2

Gas Volume Data and SN-Ratio

- Response: y (volume of gas generated (m^3) per hour with a target value 20)
- Four control factors:

	Coded Levels		
	-1	0	1
x_1 (Temperature)	70	80	90 (°C)
x_2 (Pressure)	10	20	30 (psi)
x_3 (Reaction time)	5	6	7 (minutes)
x_4 (Additives)	1	2	3 (%)

- Two noise factors:

	Coded Levels	
	-1	1
z_1 (Humidity)	20	30 (%)
z_2 (Ventilation)	No	Yes

- Summary data and SN-ratios:

x_1	x_2	x_3	x_4	$\bar{y}$	s	$-10 log_{10}\, s^2$
-1	-1	-1	-1	19.1	0.22	13.152
-1	0	0	0	19.4	0.16	15.198
-1	1	1	1	20.1	0.10	20.000
0	-1	0	1	21.7	0.17	15.391
0	0	1	-1	21.2	0.13	17.721
0	1	-1	0	20.9	0.15	16.478
1	-1	1	0	21.8	0.19	14.425
1	0	-1	1	22.9	0.16	15.918
1	1	0	-1	22.6	0.12	18.416

Because of the nature of the experiment, Taguchi's SN-ratio "target is best" $-10log_{10}s^2$ was used for the experiment. The mean responses and the average SN-ratios broken down by the levels of each control factors are shown below.

- Average volume of gas generated takes means into account.

Levels	x_1	x_2	x_3	x_4
-1	19.53	20.87	20.97	20.97
0	21.27	21.17	21.23	20.70
1	22.43	21.20	21.03	21.57

- Average SN-ratio takes variability into account.

Levels	x_1	x_2	x_3	x_4
-1	16.356	14.322	15.182	16.430
0	16.530	16.519	16.575	15.607
1	16.253	18.298	17.382	17.103

Figure 1.1 shows the mean response and the average SN-ratio for each factor. From Figure 1.1 we observe the following.

Factors	Mean	SN-ratio	Conclusion/Recommendation
x_1 (Temperature)	sharp increase	little Change	use as an adjustment factor
x_2 (Pressure)	little change	sharp increase	influences variability
x_3 (Reaction time)	little change	sharp increase	influences variability
x_4 (Additives)	some change	much change	influences variability more than the mean

As Taguchi sees it, Figure 1.1 also suggests "winners" (optimal operating conditions) for each factor. For example, operating x_1 at about 72°C, x_2 at the high level (30 psi), x_3 at the high level (7 minutes) and x_4 at the high level (3%) would yield desirable mean responses with the smallest variability.

By chance, the "winning" factor combination (-1, 1, 1, 1) was included in the original experiment and it happens to have a mean closest to the target (20 m^3/hour) with the smallest variability. A few things are evident from this experiment. The fact that the winner often occurs where we don't have data in the original experiment can be understood clearly from this example. In addition, as can be seen from the way conclusions are made, we have completely ignored possible interactions among control factors as in Example 1.1. Much effort can be directed to understand the process in greater depth. As in Example 1.1, a plot of the mean response contours and process standard deviation contours could be used to graphically predict optimum conditions for the mean with the least variability.

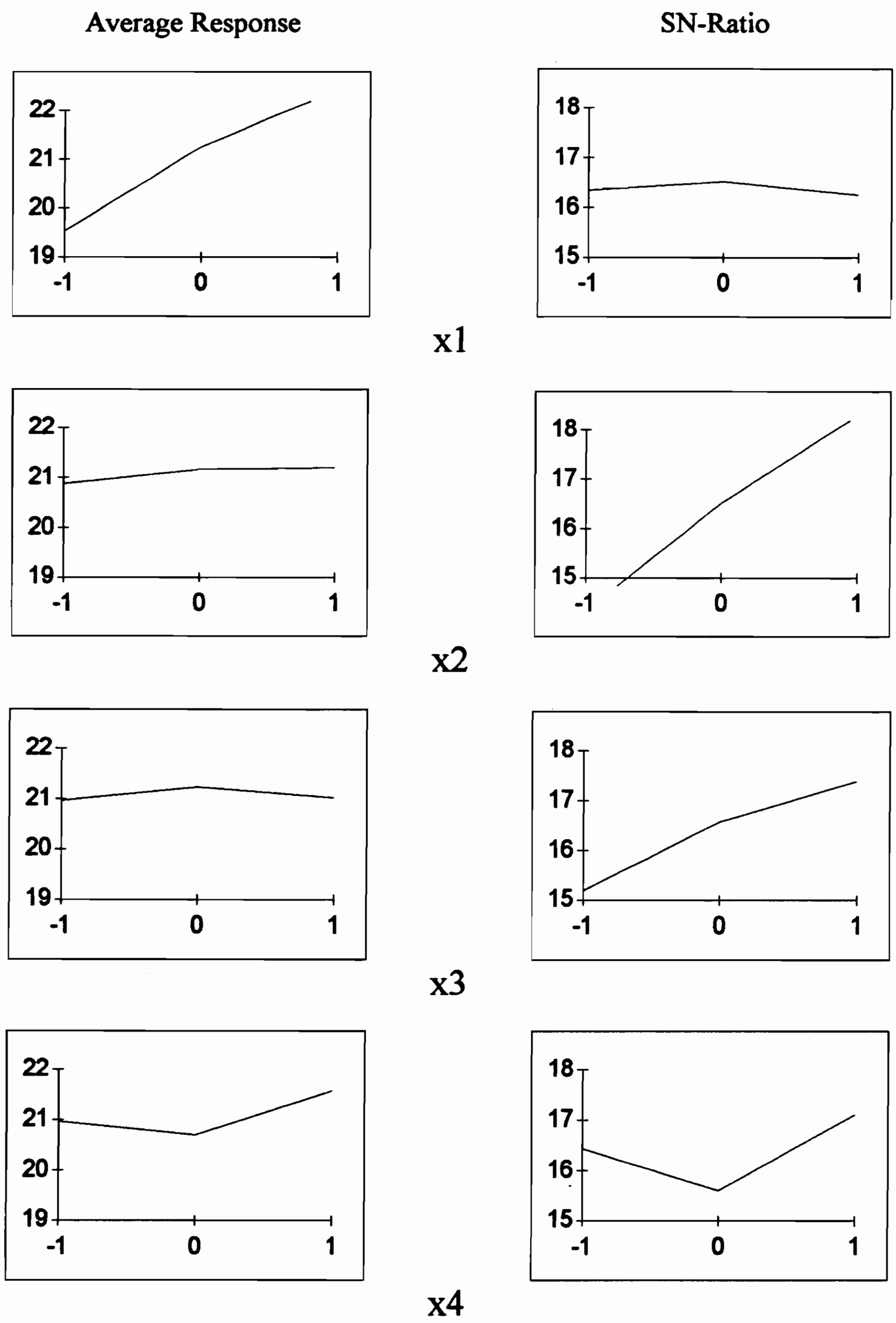


Figure 1.1 Gas Volume Data of Example 1.2

1.5 Potential Areas of Improvement of Taguchi Methods

Some engineers have accepted uncritically the use of Taguchi's SN-ratio for analyzing experimental data. But Taguchi's work certainly leaves some room for improvement. For example, the lack of formal sequential design implementation and limited choice of designs have been widely discussed. Furthermore, heavily fractionated designs suggested by Taguchi often do not allow estimation of interactions among control factors, and his SN-ratio doesn't work efficiently in modeling. These and other related areas for improvement can be summarized as:

1. SN-ratio is unconvincing,
2. there is much preoccupation with finding optimal operating conditions without doing any real statistical modeling,
3. designs do not allow for interactions among control factors,
4. the methodology does not naturally lead to sequential experimentation, and
5. crossed array involves large experiments (experimental extravagance).

Box (1985, 1988), Lucas (1985), and Vining and Myers (1990) pointed out that separate modeling for the mean and variance would result in learning more about the process. This stems from Taguchi's preoccupation with optimization rather than an understanding of the process. Much attention has been paid to the development of alternatives to Taguchi's SN-ratios. The three basic SN-ratios are seen to be transformations of the data but, as a simple example can easily demonstrate, they do not effectively separate the mean and variance induced by the control and noise factors. Taguchi's rigid approach can often lead to incorrect identification of adjustment and dispersion factors. Box (1988) showed how transforming the response can work more efficiently than Taguchi's SN-ratios to make the process mean and process variance independent. Appropriate transformation on the response can provide maximum separation between factors having location effects and those having dispersion effects. This will make the search for the optimum conditions on the mean easier. Response surface methodology as discussed in the following chapter, can also be used to correct the deficiencies of Taguchi's SN-ratio methods.

Chapter 2. Response Surface Analysis Approach

2.1 Alternatives to Taguchi Methods

As an alternative to Taguchi's approach, response surface methodology (RSM) can be used to better understand and improve the process. In improving processes, RSM is a key player by

1. helping to identify the sources of variability,
2. carrying out experiments to find ways to reduce variation,
3. developing and using models and other monitoring procedures to maintain the process after the improvements have been made.

Response surface methodology combines modeling strategy with a systematic approach to variable screening, sequential investigation, and exploration of the region that contain the estimated optimum conditions. RSM puts much emphasis on understanding the process as a system as well as estimation of the optimum conditions. The use of noise factors in modeling and design provides more flexibility than that advocated by "Taguchi methods" in robust parameter design. The approach is to combine Taguchi's notion of heterogeneous variability with standard design and modeling techniques available in RSM.

2.2 Process Variance from the RSM Point of View

In a traditional setting, noise factors are not observed, but rather, are accommodated through the randomization process. When they are used as fixed effects in the experiment, response surfaces for the process mean and variance can be constructed easily. One approach for utilizing the noise factors considers a *fitted model in both the control and noise factors*, although typically in the process the noise factors are actually random variables. Consequently, when their joint effects appear, that is, interactions between noise and control factors are significant, the process variability depends on the specific values of the control factors through the joint effects depicted in the model. This interaction structure provides a reasonable basis for estimating the process variance. The details are described in much of what follows.

As a simple illustration, consider Figure 2.1, which shows an interaction plot of one control and one noise factor case. This easily could have been generated from a fixed-effects model on factors x and z, continuous control and noise factors, respectively.

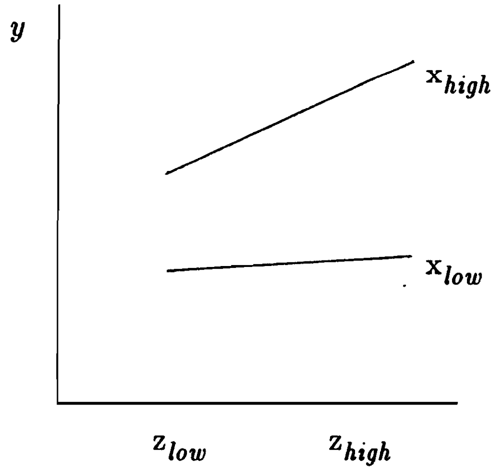
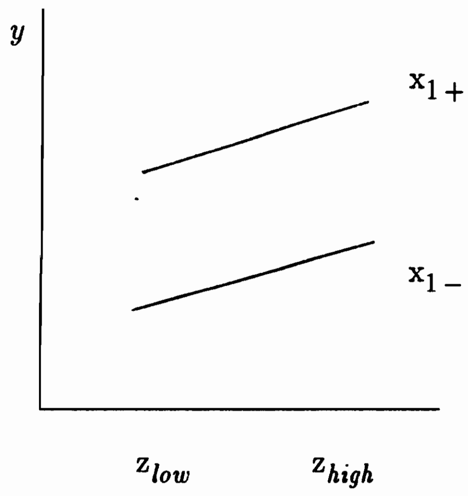


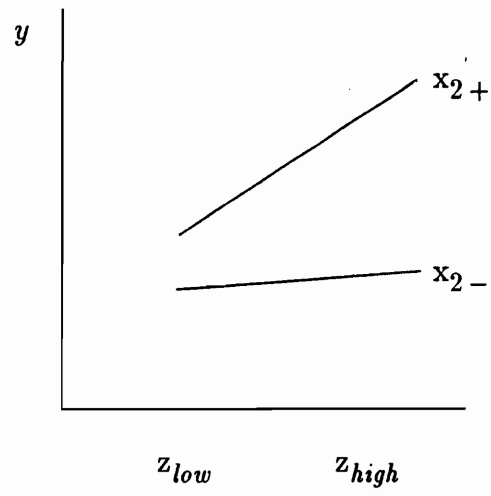
Figure 2.1 Control by Noise Interaction

If the noise factor z were to appear as random in a process, operating at the low level of a control factor clearly results in a minimum process variance. Thus, the fitted model in x and z (called "super model") and the variance of z allows for estimation of the process variance as a function of x. As can be seen from Figure 2.1, the process variance can be viewed as a function of the control factor x. The fact that the process variance can be modeled or controlled by the control factors allows us to try to minimize the process variance in terms of the control factors. That is, we may be able to achieve the minimum process variance by controlling the levels of the control factors.

Another illustration, Figure 2.2, shows an interaction plot for the case of two control factors, x_1 and x_2, and one noise factor.



(a) x_1 has no dispersion effect



(b) x_2 has positive dispersion effect

Figure 2.2 Control by Noise Interaction

Figure 2.2 illustrates that operating either at the high or low level of x_1 would show no difference as far as the process variance is concerned (x_1 has no dispersion effect). Operating at the low level of x_2 would yield smaller process variance than at the high level of x_2 (x_2 has positive dispersion effect). Notice again that the process variance can be viewed as a function of the control factor x. It is important to understand that the variance relationship is not established from a fitted model for the process variance, but through computations made on the fitted model of an experiment in which both x and z are fixed effects.

2.3 RSM Analysis with Noise Factors (The Combined Array)

When noise factors are incorporated into the system, the detection of dispersion effects and, indeed, the construction of a model for the process variance can be considerably less difficult. The "crossing" of the orthogonal arrays for the control and noise factors in a product array often results in an exorbitant number of experimental runs (refer to the criticism of Taguchi methods). The RSM and "combined array" approach involves designing an experiment with the control and noise factors in the same array. The design should allow estimation of important model terms in the control factors. The design need not involve crossing of control and noise arrays. A combined array can be more economical, and it includes the crossed array as a special case. Discussions on combined arrays can be found in Myers (1991), Welch and Sacks (1991) and Welch, Yu, Kang, and Sacks (1990), Shoemaker, Tsui, and Wu (1991), and Borkowski and Lucas (1991). The model terms in the noise factors should involve at least main effects. As shown in Figures 2.1 and 2.2, two-factor interactions between control and noise factors can be very important. They not only allow us to manipulate the process variance through the levels of control factors but also represent the components that eventually will produce a response surface for the process variance.

After the data are collected, rather than pursuing the SN-ratio a model is fit using the natural response (or an appropriate transformation, e. g., *log*). The model is used to produce another response surface of the process variance that will be used to investigate process variance. The following are examples of combined arrays and the resulting model that may be fit in the control and noise fixed effects.

Example 2.1 (Combined Array I)

Suppose we have three potential control factors (x_1, x_2, and x_3) and four noise factors (z_1, z_2, z_3, and z_4). It is important to study main effects and two-factor interactions among the control factors. Main effects in the noise factors are to be studied as well as two-factor interactions between control and noise factors. The latter terms are necessary in order that process variance be studied. Recall Figures 2.1 and 2.2. Thirty two experimental runs can be used. The following are possible defining relations that generate an appropriate design

$$I = x_1x_2x_3z_1z_3 = x_1x_2z_1z_2z_4.$$

The resulting $\frac{1}{4}$ fraction of a 2^7 fractional factorial can be used to fit the regression model

$$\hat{y} = b_0 + \sum_{i=1}^{3} b_i x_i + \sum_{i<j}\sum b_{ij} x_i x_j + \sum_{i=1}^{4} c_i z_i + \sum_{i}\sum_{j} d_{ij} x_i z_j. \tag{2.1}$$

Note that the combined array involves 32 runs whereas a crossed array might need 128 ($2^3 \times 2^4$) experimental runs. Standard approaches to model selection or screening can be taken here. If one is certain that lack of fit effects are negligible, t-tests can be applied as well. The dispersion effects exist in the *control factors that produce a significant two-factor interaction with at least one noise factor*. The choice of the metric here may be crucial since further diagnostic work or region seeking may be very simple if the number of interactions among control and control factors is reduced.

Example 2.2 (Combined Array II)

Suppose screening is not necessary and, indeed, the experimenter feels as if he/she is already in the desirable operating region. As a result, a second order model with interaction should be fit in the control factors. Suppose the experiment involves three control (x_1, x_2, and x_3) and two noise factors (z_1 and z_2). Again, two-factor interactions among the control and noise factors are important. There are several designs we can choose from possible response surface designs. One appropriate design might be the central composite design (Myers, 1976).

x_1	x_2	x_3	z_1	z_2	
± 1	± 1	± 1	± 1	± 1	} Resolution V fraction
± 1	0	0	0	0	
0	± 1	0	0	0	
0	0	± 1	0	0	
$\underline{0}$	$\underline{0}$	$\underline{0}$	$\underline{0}$	$\underline{0}$	} Replicated center runs

Axial levels for the noise factors may be used here. An appropriate model might be

$$\hat{y} = b_0 + \sum_{i=1}^{3} b_i x_i + \sum_{i<j}\sum b_{ij} x_i x_j + \sum_{i=1}^{3} b_{ii} x_i^2 + \sum_{i=1}^{2} c_i z_i + \sum_{i}\sum_{j} d_{ij} x_i z_j. \quad (2.2)$$

If the appropriate model is a second order response surface, the resulting estimates will be useful. The replication information will provide an estimate of experimental error for the estimation of the sum of the pure quadratic terms.

2.4 Construction of the Response Surfaces

The purpose of the noise factors is to supply process variance information, and the model as developed in (2.1) and (2.2) will serve that purpose. Suppose we write the general model in the x's and z's as

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\underline{x} + \underline{z}'\underline{\gamma} + \underline{x}'\Delta\underline{z} + \varepsilon, \quad (2.3)$$

where $\underline{x}' = [x_1 \ \ x_2 \ \ \ldots \ \ x_{r_x}]$, $\underline{z}' = [z_1 \ \ z_2 \ \ \ldots \ \ z_{r_z}]$, and ε is iid $N(0, \sigma_\varepsilon^2)$.

The terms $\underline{x}$ and $\underline{z}$ contain linear terms and thus $\underline{x}'\Delta\underline{z}$ contains all two-factor interactions among control and noise factors. The matrix Δ is $r_x \times r_z$ where

r_x is the number of control factors and r_z is the number of noise factors. The method of least squares is used for estimation of coefficients.

The notion of robust parameter design (RPD) deals with the choice of levels of design factors that are more robust to the environmental effects. A response surface for the process variance as a function of the control factors (design factors) can be generated by the use of model (2.3). Although the z's are fixed and controlled in the experiment, it is assumed that the process faces random z's. In accordance with design level centering and scaling, it is assumed that

$$\mathrm{E}(\underline{z}) = \underline{0}$$

$$\mathrm{Var}(\underline{z}) = \mathrm{V}. \tag{2.4}$$

Thus, the mean and variance response surfaces are obtained by taking, respectively, the expectation and variance across $\underline{z}$ of model (2.3) with the operative assumption being that in Equation (2.4). Let us consider an example of a specific form for Equation (2.4). In fact, consider Example 2.1, with the fitted model of Equation (2.1) which was constructed from the two-level design discussed. If we assume

$$\sigma^2_{z_i} = \sigma^2_z,\ i = 1, 2, 3, 4$$

and that the noise factors are uncorrelated (which may not always be a safe assumption) then the estimated response surfaces for the mean and variance are given by

$$\hat{\mu} = b_0 + \sum_{i=1}^{3} b_i x_i + \sum_{i<j}\sum b_{ij} x_i x_j$$

and

$$\hat{\sigma}_y^2 = \sigma_Z^2 \sum c_i^2 + \sigma_Z^2 \sum_i \sum_j d_{ij}^2 x_i^2 + 2\sigma_Z^2 \sum_i \sum_j c_j d_{ij} x_i + 2\sigma_Z^2 \sum_{i<k} \sum_{j=1}^{4} d_{ij} d_{kj} x_i x_k,$$

where b_0, b_i and b_{ij} are ordinary least least squares (ols) estimates of β_0, β_i, and β_{ij}. Similarly c_i and d_{ij} are ols estimates of γ_i and δ_{ij} of the general model (2.3).

For Equation (2.2), it is easier to use matrix notation. From (2.3) we have

$$\hat{y} = b_0 + \underline{x}'\underline{b} + \underline{x}'\hat{B}\underline{x} + \underline{z}'\hat{\underline{\gamma}} + \underline{x}'\hat{\Delta}\underline{z}.$$

The vectors $\underline{b}$ and $\underline{c}$ contain linear terms and matrices $\hat{B}$ and $\hat{\Delta}$ contain quadratic coefficients with off diagonal elements being the interaction coefficients divided by two (see Chapter 3). For the model for the mean, we obtain

$$\hat{\mu} = b_0 + \underline{x}'\underline{b} + \underline{x}'\hat{B}\underline{x}.$$

The estimated variance model is

$$\hat{\sigma}_y^2 = (\hat{\underline{\gamma}}' + \underline{x}'\hat{\Delta})V(\hat{\underline{\gamma}} + \hat{\Delta}'\underline{x}) + \hat{\sigma}_\varepsilon^2,$$

where $\hat{\sigma}_\varepsilon^2$ is the usual mean square error obtained from fitted model (2.3). The response surfaces $\hat{\mu}(x)$ and $\hat{\sigma}_y^2(x)$ can be used together to produce optimum conditions or, more appropriately, an exploration of the experimental region. For the case of the two-level designs (say, Example 2.1) one can use them in a region seeking mechanism. For a more elaborate experiment(say, Example 2.2), the two response surfaces can be used in harmony with

$$\min_{\underline{x}} \hat{\sigma}_y^2(\underline{x})$$

$$\text{subject to } \hat{\mu}(\underline{x}) = \mu_1.$$

For maximum or minimum response problems, several values of μ_1 can be used. Refer to Vining and Myers (1990) for a detailed discussion and Myers and Carter (1973) for discussion on a dual response approach. Also, graphical overlays of contours of constant mean and variance can be very useful in understanding the process.

2.5 Extending RSM Approach to RPD

We can supplement current work on the RSM approach to the concept of robust parameter design (RPD) by extending the response surface analysis to the stationary point of the mean response surface and by analyzing the process variance structure. Extensions of the RSM approach to RPD also allows us to construct a confidence region on the optimal location of the minimum process variance with or without constraints. In addition, the analysis and prediction of future observations can be extended to RPD. In what follows we attempt to analyze the process variance and its role in the robust parameter design and give a detailed description of current research.

Chapter 3. RSM Analysis in Robust Parameter Design

3.1 Analysis of the Location of the Minimum Process Variance

Consider a model in control, $\underline{x}$, and noise, $\underline{z}$, for some process with response y given by

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'\mathrm{B}\underline{x} + \underline{z}'\underline{\gamma} + \underline{x}'\Delta\underline{z} + \varepsilon, \tag{3.1.1}$$

where

$$\varepsilon \sim \mathrm{N}(0,\, \sigma_\varepsilon^2),$$

$$\underline{x}' = [\ \mathrm{x}_1 \quad \mathrm{x}_2 \quad \cdots \quad \mathrm{x}_{\mathrm{r_x}}\ ],$$

$$\underline{z}' = [\ \mathrm{z}_1 \quad \mathrm{z}_2 \quad \cdots \quad \mathrm{z}_{\mathrm{r_z}}\ ],$$

with parameters

$$\underline{\beta}' = [\ \beta_1 \quad \beta_2 \quad \cdots \quad \beta_{\mathrm{r_x}}\ ],$$

$$\mathrm{B} = \begin{bmatrix} \beta_{11} & \frac{1}{2}\beta_{12} & \frac{1}{2}\beta_{13} & \cdots & \frac{1}{2}\beta_{1\mathrm{r_x}} \\ & \beta_{22} & \frac{1}{2}\beta_{23} & \cdots & \frac{1}{2}\beta_{2\mathrm{r_x}} \\ & & \cdots & \cdots & \cdots \\ & \text{(symmetric)} & & & \beta_{\mathrm{r_x r_x}} \end{bmatrix},$$

$$\underline{\gamma}' = [\ \gamma_1 \quad \gamma_2 \quad \cdots \quad \gamma_{\mathrm{r_z}}\ ],$$

and

$$\Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1\mathrm{r_z}} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2\mathrm{r_z}} \\ \cdots & \cdots & \cdots & \cdots \\ \delta_{\mathrm{r_x}1} & \delta_{\mathrm{r_x}2} & \cdots & \delta_{\mathrm{r_x r_z}} \end{bmatrix}.$$

In the above r_x and r_z, respectively, are the number of control and noise factors. Thus $\underline{\beta}$ is $r_x \times 1$, $\underline{\gamma}$ is $r_z \times 1$, B is $r_x \times r_x$, and Δ is $r_x \times r_z$.

Recall that the noise factors, $\underline{z}$, are random in the process and

$$E(\underline{z}) = \underline{0}$$

$$\mathrm{Var}(\underline{z}) = V.$$

The process mean of model (3.1.1) is given by

$$E_{\underline{z}}(y) = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\underline{x}. \tag{3.1.2}$$

Now let us turn our attention to the process variance which is the more important aspect of process control. In order to evaluate the process variance we consider the model of Equation (3.1.1). The variance of the process is given by

$$\mathrm{Var}_{\underline{z}}(y \mid \underline{x}) = (\underline{\gamma}' + \underline{x}'\Delta)V(\underline{\gamma} + \Delta'\underline{x}) + \sigma_\varepsilon^2, \tag{3.1.3}$$

where the variance-covariance matrix of $\underline{z}$'s, V, is

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1r_z} \\ & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2r_z} \\ & & \cdots & \cdots & \cdots \\ & (\mathrm{sym}) & & & \sigma_{r_z}^2 \end{bmatrix}, \tag{3.1.4}$$

where $\sigma_j^2 = \mathrm{Var}(z_j)$ and $\sigma_{kl} = \mathrm{Cov}(z_k, z_l)$, $k \neq l$. If we assume

$$\sigma_{z_j}^2 = \sigma_z^2, \text{ for all } j,$$

and that the noise factors are uncorrelated then Equation (3.1.3) becomes

$$\mathrm{Var}_{\underline{z}}(y \mid \underline{x}) = (\underline{\gamma}' + \underline{x}'\Delta)(\underline{\gamma} + \Delta'\underline{x})\sigma_z^2 + \sigma_\varepsilon^2. \tag{3.1.5}$$

Now, by coding noise factors in a ± 1 metric corresponding to $\pm \sigma_j$, and by assuming that the noise factors are independently and identically distributed with $\sigma_{\underline{Z}}^2 = 1$, we obtain

$$\mathrm{V} = \mathrm{I}. \tag{3.1.6}$$

Thus Equations (3.1.3) and therefore (3.1.5) can be written as

$$\mathrm{Var}_{\underline{Z}}(y \mid \underline{x}) = (\underline{\gamma}' + \underline{x}'\Delta)(\underline{\gamma} + \Delta'\underline{x}) + \sigma_\varepsilon^2. \tag{3.1.7}$$

Rewrite Equation (3.1.7) as

$$\mathrm{Var}_{\underline{Z}}(y \mid \underline{x}) = \underline{l}'\underline{l} + \sigma_\varepsilon^2, \tag{3.1.8}$$

$$\text{where } \underline{l} = \frac{\partial y}{\partial \underline{Z}} = (\underline{\gamma} + \Delta'\underline{x}).$$

The $\underline{l}$'s are *derivatives of the response with respect to the noise factors* and vector $\underline{l}$ consists of linear polynomial functions in the control factors. The presence of the design factors $\underline{x}$ in $\underline{l}$ enables us to reduce the variability of the process through adjusting the control factors. In theory the minimum process variance of the process can be achieved at $\underline{l} = \underline{0}$. Let us now assume that there exists a point $\underline{x}_0'$ for which $\underline{l} = \underline{0}$. Then $\underline{x}_0'$ denotes the location where the process variance becomes minimum. That is, at $\underline{x}_0'$,

$$(\underline{\gamma} + \Delta'\underline{x}_0) = \underline{0}. \tag{3.1.9}$$

Since $\widehat{\underline{l}}$ is distributed as normal with mean $\underline{0}$ and variance-covariance matrix $\mathrm{Var}\,(\widehat{\underline{l}}_{\underline{x}_0'})$ at $\underline{x}_0'$, we have

$$\frac{(\widehat{\underline{l}}_{\underline{x}_0'})'\left[\widehat{\mathrm{Var}}\,(\widehat{\underline{l}}_{\underline{x}_0'})\right]^{-1}(\widehat{\underline{l}}_{\underline{x}_0'})}{\mathrm{r_z}} \sim \mathrm{F}_{\mathrm{r_z},\, \mathrm{dfE}}, \tag{3.1.10}$$

where $\widehat{\underline{l}}$ and $\widehat{\mathrm{Var}}\,(\widehat{\underline{l}})$ are estimates of $\underline{l}$ and $\mathrm{Var}\,(\widehat{\underline{l}})$, respectively. As a result, we get

$$P\{\frac{(\widehat{\underline{l}}_{\underline{x}_0'})' \left[\widehat{\mathrm{Var}}\,(\widehat{\underline{l}}_{\underline{x}_0'})\right]^{-1} (\widehat{\underline{l}}_{\underline{x}_0'})}{r_z} \leq F_{r_z,\ dfE;\ 1-\alpha}\} = 1-\alpha. \tag{3.1.11}$$

Equation (3.1.11) holds since $\mathrm{Var}(\widehat{\underline{l}}) = \sigma_\varepsilon^2 C$, where C contains no random variables. Inside the probability statement of the Equation (3.1.11), $\widehat{\mathrm{Var}}\,(\widehat{\underline{l}}_{\underline{x}_0'})$ is a function of design and s^2, an estimate of σ_ε^2, and also of $\underline{x}_0'$, and dfE is the error degrees of freedom for the fitted model of Equation (3.1.1). The values of $\underline{x}_0'$ that satisfy the probability statement of (3.1.11) lie inside the $(1-\alpha)100\%$ confidence region on the location of the minimum process variance. This concept will be used in an example later in the dissertation.

In what follows, we describe the use of the response surface methodology for situations in which noise factors are used. Indeed, this is the main contribution and the claim to originality for the dissertation.

3.2 Estimation of the Process Variance

We now consider both biased and unbiased estimators of the process variance and study their properties. In much of the development presented in the following sections is the consideration of the process variance as a function of $\underline{l}$.

Consider the process variance of (3.1.8)

$$\mathrm{Var}_{\underline{z}}(y \mid \underline{x}) = \underline{l}'\underline{l} + \sigma_\varepsilon^2. \tag{3.2.1}$$

An intuitive but biased estimator of the process variance to be used in this study is given by

$$\hat{\mathrm{Var}}_{\underline{z}}(y \mid \underline{x}) = \hat{\underline{l}}'\hat{\underline{l}} + s^2, \tag{3.2.2}$$

where s^2 is an estimate of the error variance. Note that this estimator is biased since

$$E\{\hat{\mathrm{Var}}_{\underline{z}}(y \mid \underline{x})\} = \underline{l}'\underline{l} + \sigma_\varepsilon^2 \, \mathrm{tr}(C) + \sigma_\varepsilon^2, \tag{3.2.3}$$

where $\sigma_\varepsilon^2 C = \begin{bmatrix} \mathrm{Var}\,\hat{l}_1 & \mathrm{Cov}\,(\hat{l}_1, \hat{l}_2) & \mathrm{Cov}\,(\hat{l}_1, \hat{l}_3) & \cdots & \mathrm{Cov}\,(\hat{l}_1, \hat{l}_{r_z}) \\ & \mathrm{Var}\,\hat{l}_2 & \mathrm{Cov}\,(\hat{l}_2, \hat{l}_3) & \cdots & \mathrm{Cov}\,(\hat{l}_2, \hat{l}_{r_z}) \\ & & \cdots & \cdots & \cdots \\ & (\mathrm{sym}) & & & \mathrm{Var}\,\hat{l}_{r_z} \end{bmatrix}$ and

tr (C) is a trace of the matrix C. Let us define this matrix C more carefully.

Let

$$\hat{l}_i = (1, \underline{x}')\hat{\eta}_i, \tag{3.2.4}$$

where $(1, \underline{x}') = (1, x_1, x_2, \ldots, x_{r_x})$, $\hat{\eta}_i' = (\hat{\gamma}_i, \hat{\delta}_{1i}, \hat{\delta}_{2i}, \ldots, \hat{\delta}_{r_x i})$, and $i = 1, 2, \ldots, r_z$. Then,

$$\mathrm{Var}\,(\hat{l}_i) = (1, \underline{x}')\mathrm{Var}\,(\hat{\eta}_i)\binom{1}{\underline{x}}, \tag{3.2.5}$$

where

$$\mathrm{Var}\,(\hat{\eta}_i) = \begin{bmatrix} \mathrm{Var}\,\hat{\gamma}_i & \mathrm{Cov}\,(\hat{\gamma}_i, \hat{\delta}_{1i}) & \mathrm{Cov}\,(\hat{\gamma}_i, \hat{\delta}_{2i}) & \cdots & \mathrm{Cov}\,(\hat{\gamma}_i, \hat{\delta}_{r_x i}) \\ & \mathrm{Var}\,\hat{\delta}_{1i} & \mathrm{Cov}\,(\hat{\delta}_{1i}, \hat{\delta}_{2i}) & \cdots & \mathrm{Cov}\,(\hat{\delta}_{1i}, \hat{\delta}_{r_x i}) \\ & & \cdots & \cdots & \cdots \\ & (\mathrm{sym}) & & & \mathrm{Var}\,\hat{\delta}_{r_x i} \end{bmatrix}.$$

Note that $\mathrm{Var}\,(\hat{\eta}_i)$ apart from σ_ε^2 is a submatrix of the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix where $\mathbf{X}$ is a model matrix in a super model (3.1.1) that includes both control factors and noise factors. From Equation (3.2.5) tr (C) can be written as

$$\mathrm{tr}\,(\mathrm{C}) = \sum_{i=1}^{\mathrm{r_z}} \mathrm{Var}\,(\widehat{l}_i)$$

$$= \sum_{i=1}^{\mathrm{r_z}} (1,\, \underline{\mathrm{x}}')\mathrm{Var}\,(\widehat{\eta}_i)\binom{1}{\underline{\mathrm{x}}}$$

$$= \sum_{i=1}^{\mathrm{r_z}} [\mathrm{Var}\,(\widehat{\gamma}_i) + \sum_{j=1}^{\mathrm{r_x}} \mathrm{Cov}(\widehat{\gamma}_i,\, \widehat{\delta}_{ji}) + \sum_{j=1}^{\mathrm{r_x}} \{\mathrm{x}_j^2 \cdot \mathrm{Var}\,(\widehat{\delta}_{ji})\} + \sum_{k \neq l}\sum \mathrm{x}_k \mathrm{x}_l \mathrm{Cov}(\widehat{\delta}_{ki},\, \widehat{\delta}_{li})]$$

$$= (1,\, \underline{\mathrm{x}}')\begin{bmatrix} \sum_{i=1}^{\mathrm{r_z}} \{\mathrm{Var}\ \widehat{\gamma}_i + \sum_{j=1}^{\mathrm{r_x}} \mathrm{Cov}(\widehat{\gamma}_i,\, \widehat{\delta}_{ji})\} & \underline{0}' \\ \underline{0} & \underline{\mathrm{M}} \end{bmatrix}\binom{1}{\underline{\mathrm{x}}},$$

(3.2.6)

where $\underline{\mathrm{M}} =$

$$\begin{bmatrix} \sum_{i=1}^{\mathrm{r_z}} \mathrm{Var}\ \widehat{\delta}_{1i} & \sum_{i=1}^{\mathrm{r_z}} \mathrm{Cov}(\widehat{\delta}_{1i},\, \widehat{\delta}_{2i}) & \sum_{i=1}^{\mathrm{r_z}} \mathrm{Cov}(\widehat{\delta}_{1i},\, \widehat{\delta}_{3i}) & \cdots & \sum_{i=1}^{\mathrm{r_z}} \mathrm{Cov}(\widehat{\delta}_{1i},\, \widehat{\delta}_{\mathrm{r_x}i}) \\ & \sum_{i=1}^{\mathrm{r_z}} \mathrm{Var}\ \widehat{\delta}_{2i} & \sum_{i=1}^{\mathrm{r_z}} \mathrm{Cov}(\widehat{\delta}_{2i},\, \widehat{\delta}_{3i}) & \cdots & \sum_{i=1}^{\mathrm{r_z}} \mathrm{Cov}(\widehat{\delta}_{2i},\, \widehat{\delta}_{\mathrm{r_x}i}) \\ & & \sum_{i=1}^{\mathrm{r_z}} \mathrm{Var}\ \widehat{\delta}_{3i} & \cdots & \sum_{i=1}^{\mathrm{r_z}} \mathrm{Cov}(\widehat{\delta}_{3i},\, \widehat{\delta}_{\mathrm{r_x}i}) \\ & & \cdot & \cdots & \cdots \\ & \text{(sym)} & & & \sum_{i=1}^{\mathrm{r_z}} \mathrm{Var}\ \widehat{\delta}_{\mathrm{r_x}i} \end{bmatrix}$$

apart from σ_ε^2. (3.2.7)

It is easy to see that tr(C) is positive when the model terms of (3.1.1) are orthogonal to each other (i. e., when the design is at least of resolution V). Equation (3.2.3) is an immediate result of the fact that $\hat{\underline{l}}$ is normally distributed with mean $\underline{l}$ and variance-covariance matrix σ_ε^2C.

An alternative estimator of the process variance

$$\widehat{\mathrm{Var}}_{\underline{z}}(y \mid \underline{x}) = \hat{\underline{l}}'\hat{\underline{l}} + \{1 - \mathrm{tr}\ (\mathrm{C})\}\ \mathrm{s}^2 \tag{3.2.8}$$

is unbiased since

$$\begin{aligned} \mathrm{E}[\hat{\underline{l}}'\hat{\underline{l}} + \{1 - \mathrm{tr}\ (\mathrm{C})\}\ \mathrm{s}^2] &= \underline{l}'\underline{l} + \sigma_\varepsilon^2\, \mathrm{tr}(\mathrm{C}) + \{1 - \mathrm{tr}\ (\mathrm{C})\}\sigma_\varepsilon^2 \\ &= \underline{l}'\underline{l} + \sigma_\varepsilon^2, \end{aligned} \tag{3.2.9}$$

where C, the variance-covariance matrix of $\hat{\underline{l}}$ apart from σ_ε^2, and the estimated error mean square s^2 are defined previously. The question now is whether or not one estimator is uniformly better than the other.

A reasonable criterion is the minimization of the mean square error (MSE) as the sum of variance and squared bias, i. e.,

$$\mathrm{MSE} = \mathrm{Variance} + (\mathrm{Bias})^2. \tag{3.2.10}$$

We first consider the case of the biased estimator (3.2.2)

$$\mathrm{MSE}\ (\hat{\underline{l}}'\hat{\underline{l}} + \mathrm{s}^2) = \mathrm{Var}\ (\hat{\underline{l}}'\hat{\underline{l}} + \mathrm{s}^2) + \{\mathrm{Bias}\ (\hat{\underline{l}}'\hat{\underline{l}} + \mathrm{s}^2)\}^2. \tag{3.2.11}$$

Recalling results from distribution theory leads to

$$\mathrm{Var}\ (\hat{\underline{l}}'\hat{\underline{l}} + \mathrm{s}^2) = \mathrm{Var}\ (\hat{\underline{l}}'\hat{\underline{l}}) + \frac{2\sigma_\varepsilon^4}{\mathrm{dfE}}. \tag{3.2.12}$$

Recalling previous result of (3.2.3) gives

$$\text{Bias}\,(\hat{\underline{l}}'\hat{\underline{l}} + s^2) = \sigma_\varepsilon^2\,\text{tr(C)}. \tag{3.2.13}$$

Combining (3.2.12) and (3.2.13) we get

$$\begin{aligned}\text{MSE}\,(\hat{\underline{l}}'\hat{\underline{l}} + s^2) &= \text{Var}\,(\hat{\underline{l}}'\hat{\underline{l}} + s^2) + \{\text{Bias}\,(\hat{\underline{l}}'\hat{\underline{l}} + s^2)\}^2 \\ &= \text{Var}\,(\hat{\underline{l}}'\hat{\underline{l}}) + \sigma_\varepsilon^4\,\{\frac{2}{\text{dfE}} + \text{tr}^2\text{(C)}\}. \end{aligned} \tag{3.2.14}$$

Now consider the unbiased estimator in (3.2.8)

$$\text{MSE}\,[\hat{\underline{l}}'\hat{\underline{l}} + \{1 - \text{tr (C)}\}\,s^2] = \text{Var}\,[\hat{\underline{l}}'\hat{\underline{l}} + \{1 - \text{tr (C)}\}\,s^2].$$

Using the result of (3.2.12) we get

$$\text{Var}\,[\hat{\underline{l}}'\hat{\underline{l}} + \{1 - \text{tr (C)}\}\,s^2] = \text{Var}\,(\hat{\underline{l}}'\hat{\underline{l}}) + \frac{2\sigma_\varepsilon^4\{1 - \text{tr (C)}\}^2}{\text{dfE}}. \tag{3.2.15}$$

Therefore,

$$\begin{aligned}\text{MSE}\,[\hat{\underline{l}}'\hat{\underline{l}} + \{1 - \text{tr (C)}\}\,s^2] &= \text{Var}\,[\hat{\underline{l}}'\hat{\underline{l}} + \{1 - \text{tr (C)}\}\,s^2] \\ &= \text{Var}\,(\hat{\underline{l}}'\hat{\underline{l}}) + \frac{2\sigma_\varepsilon^4\{1 - \text{tr (C)}\}^2}{\text{dfE}}. \end{aligned} \tag{3.2.16}$$

Now in order to compare (3.2.14) and (3.2.16), we see

$$\begin{aligned}&\text{MSE}\,(\hat{\underline{l}}'\hat{\underline{l}} + s^2) - \text{MSE}\,[\hat{\underline{l}}'\hat{\underline{l}} + \{1 - \text{tr (C)}\}\,s^2] \\ &= \sigma_\varepsilon^4\,[\frac{2}{\text{dfE}} + \text{tr}^2\text{(C)} - \frac{2\{1 - \text{tr (C)}\}^2}{\text{dfE}}] \\ &= \frac{\sigma_\varepsilon^4\,\text{tr(C)}}{\text{dfE}}\,[(\text{dfE} - 2)\{\text{tr(C)}\} + 4]. \end{aligned} \tag{3.2.17}$$

When the design is at least of resolution V (i. e., when model terms are orthogonal to each other) with dfE > 1 the right hand side of Equation (3.2.17) exceeds 0. In this case, we see that the unbiased estimator of the process variance gives uniformly smaller mean square error than that of the biased estimator. Therefore, the use of the unbiased estimator is strongly recommended. It turns out that the two estimators often give similar results even when design has resolution less than V or dfE ≤ 1 (saturated design, for example). Examples will illustrate this point.

In what follows we will show developments using both the biased estimator and the unbiased estimator for the process variance, but the primary focus will be on the unbiased estimator.

3.3 General Stationary Region Analysis for Process Variance

3.3.1 Use of the Unbiased Estimator

In section 3.1, we developed the methodology to find the location of the minimum process variance. Consider now a case in which the location of the minimum process variance, $\underline{x}_0'$, does not exist, i. e., the case when $r_x < r_z$. We cannot use the approach developed in section 3.1 since the system of $\underline{l} = \underline{0}$ has more linear equations than the number of unknown variables (control factors). This problem is now considered from the RSM point of view. Recall that in standard RSM, $\frac{\partial \hat{y}}{\partial \underline{x}} = \underline{0}$ produces a stationary point (point of minimum response, maximum response or saddle point). It turns out that the problem of minimizing the process variance in RPD is greatly simplified when compared to standard

RSM.

To find the location of the estimated minimum process variance when there are more noise factors than control factors, we differentiate the unbiased estimator of the process variance of Equation (3.2.8). Recall from Equation (3.2.6)

$$\mathrm{tr(C)} = (1,\ \underline{x}')\begin{bmatrix} \sum_{i=1}^{r_z}\{\mathrm{Var}\ \hat{\gamma}_i + \sum_{j=1}^{r_x}\mathrm{Cov}(\hat{\gamma}_i,\ \hat{\delta}_{ji})\} & \underline{0}' \\ \underline{0} & \underline{M} \end{bmatrix}\binom{1}{\underline{x}},$$

where $\underline{M}$ is defined in (3.2.7). Therefore,

$$\frac{\partial}{\partial \underline{x}}\,\mathrm{tr}\ (\mathrm{C}) = 2\underline{M}\underline{x}. \tag{3.3.1.1}$$

From Equation (3.1.10) we have

$$\frac{\partial(\hat{\underline{l}}'\hat{\underline{l}})}{\partial \underline{x}} = \frac{\partial}{\partial \underline{x}}(\hat{\underline{\gamma}}' + \underline{x}'\hat{\Delta})(\hat{\underline{\gamma}} + \hat{\Delta}'\underline{x})$$

$$= 2\left(\hat{\Delta}\hat{\underline{\gamma}} + \hat{\Delta}\hat{\Delta}'\underline{x}\right). \tag{3.3.1.2}$$

From the results of (3.3.1.1) and (3.3.1.2) we get

$$\frac{\partial}{\partial \underline{x}}\,[\hat{\underline{l}}'\hat{\underline{l}} + \{1 - \mathrm{tr}\ (\mathrm{C})\}\ s^2] = 2\left(\hat{\Delta}\hat{\underline{\gamma}} + \hat{\Delta}\hat{\Delta}'\underline{x} - s^2\underline{M}\underline{x}\right). \tag{3.3.1.3}$$

Equate (3.3.1.3) to $\underline{0}$ and solve for $\underline{x}$ and obtain

$$(\hat{\Delta}\hat{\Delta}' - s^2\underline{M})\ \underline{x} = -\hat{\Delta}\hat{\underline{\gamma}}. \tag{3.3.1.4}$$

Let the solution of the Equation (3.3.1.4) be $\hat{\underline{x}}_0$, we get

$$\hat{\underline{x}}_0 = -(\hat{\Delta}\hat{\Delta}' - s^2\underline{M})^{-1}\hat{\Delta}\hat{\underline{\gamma}}, \tag{3.3.1.5}$$

if the solution exists. The eigenvalues of the matrix $(\hat{\Delta}\hat{\Delta}' - s^2\underline{M})$ in Equation (3.3.1.5) determine the nature of the stationary point. That is,

(i) If $(\hat{\Delta}\hat{\Delta}' - s^2\underline{M})$ is a positive definite or positive semidefinite matrix then $\hat{\underline{x}}_0$ is indeed the location of the minimum process variance.

(ii) If $(\hat{\Delta}\hat{\Delta}' - s^2\underline{M})$ is a negative definite or negative semidefinite matrix then $\hat{\underline{x}}_0$ is the location of the maximum process variance. This would lead us nowhere.

(iii) If $(\hat{\Delta}\hat{\Delta}' - s^2\underline{M})$ is an indefinite matrix then $\hat{\underline{x}}_0$ is a saddle point and we would need further investigation of the process variance by using ridge analysis as discussed in a later section.

3.3.2 Use of the Biased Estimator

In section 3.3.1, we developed the methodology for the location of the minimum process variance for the case when $r_x < r_z$ using the unbiased estimator (3.2.8) of the process variance. In the following we now consider the same problem using the biased estimator of (3.2.2).

Consider the biased estimator of the process variance

$$\hat{\text{Var}}_{\underline{z}}(y \mid \underline{x}) = \hat{\underline{l}}'\hat{\underline{l}} + s^2. \tag{3.3.2.1}$$

To find the location of the estimated minimum process variance recall first that

$$\frac{\partial(\hat{\underline{l}}'\hat{\underline{l}})}{\partial \underline{x}} = \frac{\partial}{\partial \underline{x}}(\hat{\underline{\gamma}}' + \underline{x}'\hat{\Delta})(\hat{\underline{\gamma}} + \hat{\Delta}'\underline{x})$$

$$= 2\left(\hat{\Delta}\hat{\underline{\gamma}} + \hat{\Delta}\hat{\Delta}'\underline{x}\right). \qquad (3.3.2.2)$$

Equate (3.3.2.2) to $\underline{0}$ and solve for $\underline{x}$, we have

$$\hat{\Delta}\hat{\Delta}'\underline{x} = -\hat{\Delta}\hat{\underline{\gamma}}. \qquad (3.3.2.3)$$

Let the solution of the Equation (3.3.2.3) be $\hat{\underline{x}}_0$, we get

$$\hat{\underline{x}}_0 = -(\hat{\Delta}\hat{\Delta}')^{-1}\hat{\Delta}\hat{\underline{\gamma}}, \qquad (3.3.2.4)$$

if the solution exists. Note that eigenvalues of the matrix $\hat{\Delta}\hat{\Delta}'$ in Equation (3.3.2.4) determine the nature of the stationary point (i. e., $\hat{\underline{x}}_0$). Results from matrix theory show that $\hat{\Delta}\hat{\Delta}'$ is either positive definite or at least positive semidefinite matrix (see Graybill, 1983). This proves that $\hat{\underline{x}}_0$ (if exists) is indeed the location of the minimum process variance. That is, we always obtain the location of the minimum process variance when solution of the Equation (3.3.2.4) exists. Notice the ease of the the analysis compared with when the unbiased estimator is used (see section 3.3.1). Due to the nature of the matrix $\hat{\Delta}\hat{\Delta}'$ the canonical analysis (see Myers, 1976), often performed when the matrix in question (such as $\hat{\Delta}\hat{\Delta}'$) is indefinite, is not necessary here.

By comparing two solutions to the stationary point, i. e., for the biased and unbiased case, the following are evident:

(i) When s^2 is small enough (i. e., model fit is good) the difference between the two solutions (i. e., the location of the estimated minimum process variance when the biased estimator is used compared to the location when the unbiased estimator is used) will be closer together.

(ii) As the size of the experiment increases the locations become closer together.

(iii) The difference between the two locations is largest close to the perimeter of the experimental region.

3.4 Ridge Analysis (RA) for Process Variance

3.4.1 Use of the Unbiased Estimator

In sections 3.1 and 3.3, we developed the methodology to find the location of the stationary point on the process variance, $\underline{x}_0'$. Suppose this location $\underline{x}_0'$ is not the location of the minimum process variance or is far removed from the experimental region (see Example 5.3), we would need some device that could lead us to a more reasonable location inside the experimental region. That is, we need a device that reveals an optimum condition subject to the constraint that the condition lie inside of the experimental region. This device is provided by a method known as a *ridge analysis* (RA) (see Myers (1976), Box and Draper (1987), Khuri and Cornell (1987)).

Let us minimize the process variance of the model (3.1.1),

$$\min_{\underline{x}} \mathrm{Var}_{\underline{z}}(y \mid \underline{x}_i) \tag{3.4.1.1}$$

while restricting ourselves to spheres of varying radii. We use the usual coding of the factors, with the origin of the design at (0, 0, ..., 0), for a point (x_1, x_2, ..., x_{r_x}) on a sphere of radius R,

$$\sum_{i=1}^{r_x} x_i^2 = R^2. \tag{3.4.1.2}$$

Consider the unbiased estimator of $\underline{l}'\underline{l}$, $\hat{\underline{l}}'\hat{\underline{l}} - s^2\,\mathrm{tr}(C)$. Similar to the preceding sections, the problem is to find the stationary (minimum) point of the following second order response system

$$\hat{\underline{l}}'\hat{\underline{l}} - s^2\,\mathrm{tr}(C) = (\hat{\underline{\gamma}}' + \underline{x}'\hat{\Delta})(\hat{\underline{\gamma}} + \hat{\Delta}'\underline{x}) - s^2\,(1,\, \underline{x}')\cdot$$

$$\begin{bmatrix} \sum_{i=1}^{r_z}\{\mathrm{Var}\,\hat{\gamma}_i + \sum_{j=1}^{r_x}\mathrm{Cov}(\hat{\gamma}_i,\, \hat{\delta}_{ji})\} & \underline{0}' \\ \underline{0} & \underline{M} \end{bmatrix} \binom{1}{\underline{x}}, \qquad (3.4.1.3)$$

under the constraint of (3.4.1.2), where $\underline{M}$ is defined in (3.2.7).

We attempt to find the conditions on the factors which *minimize* Equation (3.4.1.3) subject to the constraint given by Equation (3.4.1.2) for various R. After these conditions are found, R can be plotted against the appropriate coordinates x_1, x_2, ..., x_{r_x}, and $\hat{\underline{l}}'\hat{\underline{l}}$.

To minimize $\hat{\underline{l}}'\hat{\underline{l}} - s^2\,\mathrm{tr}\,(C)$ subject to the constraint of (3.4.1.2), consider the function

$$F = \hat{\underline{l}}'\hat{\underline{l}} - s^2\,\mathrm{tr}(C) - \mu(\underline{x}'\underline{x} - R^2) \qquad (3.4.1.4)$$

where μ is a Lagrangian multiplier and $\underline{x}' = (x_1,\, x_2,\, ...,\, x_{r_x})$. Using Equation (3.4.1.3), we get

$$\frac{\partial F}{\partial \underline{x}} = 2\,\{\hat{\Delta}\hat{\underline{\gamma}} + (\hat{\Delta}\hat{\Delta}' - s^2\underline{M})\underline{x} - \mu\underline{x}\}. \qquad (3.4.1.5)$$

Equating Equation (3.4.1.5) to $\underline{0}$, and solving for $\underline{x}$, we get

$$(\hat{\Delta}\hat{\Delta}' - s^2\underline{M} - \mu I_{r_x})\,\underline{x} = -\hat{\Delta}\hat{\underline{\gamma}}. \qquad (3.4.1.6)$$

One can insert predetermined values of μ into Equation (3.4.1.6), solve for x_1, x_2, ..., x_{r_x}, and $\hat{\underline{l}}'\hat{\underline{l}} - s^2 \operatorname{tr}(C)$ from Equation (3.4.1.3). However, it is important to note that the nature of the stationary points depends on the value of μ chosen. It can be shown that a value of μ which is smaller than the smallest eigenvalue of $(\hat{\Delta}\hat{\Delta}' - s^2\underline{M})$ will give rise to an $\underline{x}$ which gives an absolute minimum for a given R. The proof is given in Appendix A. An example of the use of ridge analysis will be given in Chapter 5.

3.4.2 Use of the Biased Estimator

Let us minimize the process variance of the model (3.1.1),

$$\min_{\underline{x}} \operatorname{Var}_{\underline{z}}(y \mid \underline{x}_i) \tag{3.4.2.1}$$

on a sphere of radius R,

$$\sum_{i=1}^{r_x} x_i^2 = R^2. \tag{3.4.2.2}$$

using the biased estimator, $\hat{\underline{l}}'\hat{\underline{l}}$. The problem reduces to finding the stationary (minimum) point of the second order response system

$$\hat{\underline{l}}'\hat{\underline{l}} = \hat{\underline{\gamma}}'\hat{\underline{\gamma}} + 2\,\underline{x}'\hat{\Delta}\hat{\underline{\gamma}} + \underline{x}'\hat{\Delta}\hat{\Delta}'\underline{x}, \tag{3.4.2.3}$$

under the constraint of (3.4.2.2). This is accomplished by solving the system of equations.

$$(\hat{\Delta}\hat{\Delta}' - \mu I_{r_x})\,\underline{x} = -\hat{\Delta}\hat{\underline{\gamma}}. \tag{3.4.2.4}$$

Values of μ smaller than the smallest eigenvalue of $\hat{\Delta}\hat{\Delta}'$ will give rise to an $\underline{x}$

which gives an absolute minimum for a given R. The proof is similar to that in Appendix A and thus will be omitted here.

3.5 General Constraint Set for Process Variance with Confidence Region

When the stationary point is neither maximum nor minimum or occurs outside the experimental region, then the constrained optimization is required.

3.5.1 Use of the Unbiased Estimator

In section 3.4 we have considered minimizing the process variance on a sphere of radius R,

$$\sum_{i=1}^{r_x} x_i^2 = R^2. \tag{3.5.1.1}$$

The constraint set (3.5.1.1) is a special case of a more general constraint set given by

$$\{\underline{x}\colon g(\underline{x}) = C\} \tag{3.5.1.2}$$

Recall the general model (3.1.1) and its corresponding process variance as shown in (3.1.8)

$$\mathrm{Var}_{\underline{z}}(y \mid \underline{x}_i) = \underline{l}'\underline{l} + \sigma_\varepsilon^2. \tag{3.5.1.3}$$

The problem considered first is finding the location of the minimum process variance under the general constraint set of (3.5.1.2). For example, with two control factors, $\{x_1^2 + x_2^2 = 1\}$ is a special constraint set that limits the search alongside a unit circle. Consider the unbiased estimator, $\hat{\underline{l}}'\hat{\underline{l}} - s^2\,\mathrm{tr}(C)$, for $\underline{l}'\underline{l}$ and a Lagrangian form given by

$$L = \{\hat{\underline{l}}'\hat{\underline{l}} - s^2\,\mathrm{tr}(C)\} - \mu\,\{g(\underline{x}) - C\}, \tag{3.5.1.4}$$

where $\hat{\underline{l}} = \widehat{\frac{\partial y}{\partial \underline{z}}} = (\hat{\underline{\gamma}} + \hat{\Delta}'\underline{x}_i)$ as defined previously. To find the location of the minimum process variance under general constraint set consider

$$\frac{\partial L}{\partial \underline{x}} = 2\,\{\hat{\Delta}\hat{\underline{\gamma}} + (\hat{\Delta}\hat{\Delta}' - s^2\underline{M})\underline{x}\} - \frac{\partial}{\partial \underline{x}}\,\mu\,\{g(\underline{x})\}, \tag{3.5.1.5}$$

where $\underline{M}$ is given in (3.2.7). Now equate (3.5.1.5) to $\underline{0}$ and solve for $\underline{x}$, we have a ridge analysis system as discussed in section 3.4. An appropriate choice of μ will again give rise to $\underline{x}$ which will result in the absolute minimum while satisfying the constraint set.

It may be now of interest for the researcher to determine a confidence region on the location of a constrained minimum process variance. Assume $E(\hat{\mu}) \doteq \mu$. Then $\left(\widehat{\frac{\partial L}{\partial \underline{x}_0}}\right) = \left(\widehat{\frac{\partial L}{\partial \underline{x}}}\right)\Big|_{\underline{x}=\underline{x}_0}$ is normally distributed with mean $\underline{0}$ and variance-covariance matrix V^* by assuming that $\hat{\mu}$ is held as a constant. Consequently, a $(1-\alpha)100\%$ confidence region for the constrained minimum process variance location for a fixed μ can be found by

$$P\left\{ \frac{\left(\widehat{\frac{\partial L}{\partial \underline{x}_0}}\right)' \hat{V}^{*-1} \left(\widehat{\frac{\partial L}{\partial \underline{x}_0}}\right)}{r_z} \le F_{r_z,\, dfE;\, 1-\alpha} \right\} = 1-\alpha. \tag{3.5.1.6}$$

Since L of (3.5.1.4) can be often rewritten as a second-order response surface (see section 3.4) we have

$$L = \underset{\sim}{\widehat{\gamma}}'\underset{\sim}{\widehat{\gamma}} + 2\,\underline{x}'\widehat{\Delta}\underset{\sim}{\widehat{\gamma}} + \underline{x}'(\widehat{\Delta}\widehat{\Delta}')\underline{x} - s^2(1,\,\underline{x}')$$

$$\begin{bmatrix} \sum_{i=1}^{r_z}\{\mathrm{Var}\ \widehat{\gamma}_i + \sum_{j=1}^{r_x}\mathrm{Cov}(\widehat{\gamma}_i,\ \widehat{\delta}_{ji})\} & \underline{0}' \\ \underline{0} & \underline{M} \end{bmatrix}\binom{1}{\underline{x}} - \mu\,\{g(\underline{x}) - C\}. \tag{3.5.1.7}$$

One often selects μ smaller than the smallest eigenvalue of $(\widehat{\Delta}\widehat{\Delta}' - s^2\underline{M})$ as in standard ridge analysis. In the probability statement of (3.5.1.6), $\widehat{V}^*$ is a function of the design used, the estimated error mean square s^2, and the estimated location of the minimum process variance $\underline{x}_0'$. Values of $\underline{x}_0'$ that satisfy Equation (3.5.1.6) lie inside the $(1-\alpha)100\%$ confidence region on the location of the minimum process variance while satisfying the constraint set.

3.5.2 Use of the Biased Estimator

We now consider the problem of minimizing the process variance under the general constraint set of (3.5.1.2) by using the biased estimator for $\underline{l}'\underline{l}$, $\widehat{\underline{l}}'\widehat{\underline{l}}$. Consider a Lagrangian form

$$L' = \widehat{\underline{l}}'\widehat{\underline{l}} - \mu\,\{g(\underline{x}) - C\}. \tag{3.5.2.1}$$

To find the location of the minimum process variance under the general constraint set consider

$$\frac{\partial L'}{\partial \underline{x}} = 2\left(\hat{\Delta}\hat{\underline{\gamma}} + \hat{\Delta}\hat{\Delta}'\underline{x}\right) - \frac{\partial}{\partial \underline{x}}\,\mu\,\{g(\underline{x})\}. \tag{3.5.2.2}$$

Now equate (3.5.2.2) to $\underline{0}$ and solve for $\underline{x}$, a ridge system is resulted. As in section 3.5.1 an appropriate choice of μ will give $\underline{x}$ which gives an absolute minimum inside the constraint set.

Next to build a confidence region around the location of the minimum process variance assume $E(\hat{\mu}) \doteq \mu$. Then $\left(\frac{\partial \hat{L}'}{\partial \underline{x}_0}\right) = \left(\frac{\partial \hat{L}'}{\partial \underline{x}}\right)\Big|_{\underline{x}=\underline{x}_0}$ is normally distributed with mean $\underline{0}$ and variance-covariance matrix V^* by assuming that $\hat{\mu}$ is held as a constant. Consequently, a $(1-\alpha)100\%$ confidence region for the constrained minimum process variance location for a fixed μ can be found by

$$P\left\{ \frac{\left(\frac{\partial \hat{L}'}{\partial \underline{x}_0}\right)' \hat{V}^{*-1} \left(\frac{\partial \hat{L}'}{\partial \underline{x}_0}\right)}{r_z} \le F_{r_z,\, dfE;\, 1-\alpha} \right\} = 1 - \alpha. \tag{3.5.2.3}$$

Since L' of (3.5.2.1) can be often rewritten as a second-order response surface (see section 3.4)

$$L' = \hat{\underline{\gamma}}'\hat{\underline{\gamma}} + 2\,\underline{x}'\hat{\Delta}\hat{\underline{\gamma}} + \underline{x}'\hat{\Delta}\hat{\Delta}'\underline{x} - \mu\,\{g(\underline{x}) - C\}, \tag{3.5.2.4}$$

one often selects μ smaller than the smallest eigenvalue of $\hat{\Delta}\hat{\Delta}'$ as in standard ridge analysis. Values of $\underline{x}_0'$ that satisfy the Equation (3.5.2.3) lie inside the $(1-\alpha)100\%$ confidence region on the location of the minimum process variance inside the constraint set.

3.6 Confidence Interval for $\frac{l'l}{\sigma_\varepsilon^2}$ at a Fixed Level of Control Factors

Suppose engineers either already know or have a fairly good estimate of σ_ε^2. For example one may use the standard mean square error from the fit of the model (3.1.1). They might be interested in developing confidence intervals on the ratio of the variability transmitted by the noise factors and the overall variability, that is, confidence intervals on $\frac{l'l}{\sigma_\varepsilon^2}$. This quantity is not only interesting in itself but also plays an important role in the derivation of the tolerance intervals. In a later section we will discuss the tolerance intervals.

Consider a general model of (3.1.1). Recall $\hat{\underline{l}}$, the estimate of $\underline{l}$, given by

$$\hat{\underline{l}} = (\hat{\underline{\gamma}} + \hat{\Delta}'\underline{x}). \tag{3.6.1}$$

Assume that $\hat{l}_i$'s are orthogonal to each other (i. e., γ's and δ's of model 3.1.1 are orthogonal to each other). This gives rise to

$$\frac{\hat{l}'\hat{l}}{\sigma_\varepsilon^2 c} \sim \chi^{2'}(r_Z, \frac{l'l}{2\sigma_\varepsilon^2 c}), \tag{3.6.2}$$

where $\chi^{2'}$ is a noncentral chi-square distribution and $\frac{\hat{l}'\hat{l}}{\sigma_\varepsilon^2 c} = \sum_{i=1}^{r_Z} \frac{\hat{l}_i^2}{\sigma_\varepsilon^2 c^{ii}}$ with c^{ij} being the (i, j)th element of the matrix C. Results from distribution theory leads to

$$\frac{\hat{l}'\hat{l}}{s^2 c} \sim F'(r_Z, \text{dfE}, \lambda), \tag{3.6.3}$$

where F' is a noncentral F distribution with degrees of freedom 1, dfE and the noncentrality parameter $\lambda = \frac{l'l}{2\sigma_\varepsilon^2 c}$. From Expression (3.6.3) we have

$$P\left\{F'_{\{r_Z,\ \text{dfE},\ \lambda;\ 1-\alpha/2\}} \leq \frac{\hat{l}'\hat{l}}{s^2 c} \leq F'_{\{r_Z,\ \text{dfE},\ \lambda;\ \alpha/2\}}\right\} = 1 - \alpha. \tag{3.6.4}$$

If the probability statement (3.6.4) holds for fixed values of $\underline{x}$ and $\frac{l'l}{\sigma_\varepsilon^2}$, then we can say the value lies inside the $(1-\alpha)100\,\%$ confidence region of $\frac{l'l}{\sigma_\varepsilon^2}$. Continuing in this fashion, the $(1-\alpha)100\,\%$ confidence region for $\frac{l'l}{\sigma_\varepsilon^2}$ can be constructed. The computation of the confidence region is not difficult and makes use of the fact that the quantile of the noncentral F-distribution is a monotonically increasing function of the noncentrality parameter λ.

3.7 One-Sided Tolerance Intervals on Future Observations

In this section we discuss the prediction of future process observations through the use of a response surface approach to RPD. This is consistent with quality control concepts such as specification limits and/or process capability considerations.

The subject of statistical tolerance intervals arose and developed in response to engineers' concern with *ordinary tolerance regions.* For example, in a mass-production process, industry-wide specifications might dictate that any component that measures greater than U be considered unsatisfactory. That is to say, a certain variability is tolerated and, indeed, the interval (-∞, U] may arise from design considerations and/or cost break-even points, etc. In fact, the manufacturer may like to know how successful the production process is performing in the sense that he may wish information on the probability P[Y ≤ U], where the random variable Y is the response of interest.

Definition 3.7.1 $(-\infty,\ \hat{\mu}+ks(\underline{x})]$ is p-content upper one-sided tolerance interval at confidence level $(1-\gamma)$ if $\mathrm{P}_{\widehat{\mu}, s(\underline{x})}\left[\mathrm{P}_{y_f}\left\{ y_f \leq \widehat{\mu}+ks(\underline{x})\right\} \geq p\right] \geq 1-\gamma$. A p-content upper one-sided tolerance interval $(-\infty,\ \widehat{\mu}+ks(\underline{x})]$ contains at least $100p\,\%$ of the population or process being sampled with $(1-\gamma)$-confidence level (i. e., probability $1-\gamma$).

A limit is sought so that we can be $(1-\gamma)100\,\%$ sure that at least $100p\,\%$ of the population is below that limit. We wish to compute k so that with probability $(1-\gamma)$ at least a proportion p of the sampled distribution will be below $\widehat{\mu}+ks(\underline{x})$. In the following we consider an upper one-sided tolerance interval of Definition 3.7.1. The result can be easily extended to another form of one-sided tolerance interval as shown in the following definition.

Definition 3.7.2 $[\widehat{\mu}-ks(\underline{x}),\ \infty)$ is p-content lower one-sided tolerance interval at confidence level $(1-\gamma)$ if $\mathrm{P}_{\widehat{\mu}, s(\underline{x})}\left[\mathrm{P}_{y_f}\left\{ y_f \geq \widehat{\mu}-ks(\underline{x})\right\} \geq p\right] \geq 1-\gamma$. A p-content lower one-sided tolerance interval $[\widehat{\mu}-ks(\underline{x}),\ \infty)$ contains at least $100p\,\%$ of the population or process being sampled with $(1-\gamma)$-confidence level.

Recall the general model of (3.1.1)

$$y=\beta_0+\underline{\mathrm{x}}'\underline{\beta}+\underline{\mathrm{x}}'\mathrm{B}\underline{\mathrm{x}}+\underline{\mathrm{z}}'\underline{\gamma}+\underline{\mathrm{x}}'\Delta\underline{\mathrm{z}}+\varepsilon,$$

where $\underline{\mathrm{x}}$ and $\underline{\mathrm{z}}$ denote control factors and noise factors, respectively. The upper one-sided p-content tolerance interval at confidence level $(1-\gamma)$ is given in definition 3.7.1. We have $s(\underline{\mathrm{x}})=\widehat{\mathrm{Var}}(y)^{1/2}=\left(\widehat{\underline{l}}'\widehat{\underline{l}}+\widehat{\sigma}_{\varepsilon}^2\right)^{1/2}$, $\widehat{\underline{l}}=\widehat{\underline{\gamma}}+\widehat{\Delta}'\,\underline{\mathrm{x}}$, and $\widehat{\sigma}_{\varepsilon}^2=$ estimated MSE and assume $\sigma_{\underline{\mathrm{z}}}^2=1$. Note that $s(\underline{\mathrm{x}})$ is a function of $\underline{\mathrm{x}}$. Definition 3.7.1 can be written as

$$P_{\hat{\mu}, s(\underline{x})}[P_{y_f}\left\{\frac{y_f-(\beta_0+\underline{x}'\underline{\beta}+\underline{x}'B\underline{x})}{\sqrt{\underline{l}'\underline{l}+\sigma_\varepsilon^2}} \le \frac{\hat{\mu}-(\beta_0+\underline{x}'\underline{\beta}+\underline{x}'B\underline{x})+ks(\underline{x})}{\sqrt{\underline{l}'\underline{l}+\sigma_\varepsilon^2}}\right\} \ge p] \ge 1-\gamma. \tag{3.7.1}$$

Letting K_p denote the pth quantile point of the standard normal distribution expression, (3.7.1) can be written as

$$P_{\hat{\mu}, s(\underline{x})}\left\{\frac{\hat{\mu}-(\beta_0+\underline{x}'\underline{\beta}+\underline{x}'B\underline{x})+ks(\underline{x})}{\sqrt{\underline{l}'\underline{l}+\sigma_\varepsilon^2}} \ge K_p\right\} \ge 1-\gamma. \tag{3.7.2}$$

The left hand side of Expression (3.7.2) can be written as

$$P_{\hat{\mu}, s(\underline{x})}\left\{\hat{\mu}-(\beta_0+\underline{x}'\underline{\beta}+\underline{x}'B\underline{x}) \ge K_p\sqrt{\underline{l}'\underline{l}+\sigma_\varepsilon^2} - ks(\underline{x})\right\}$$

$$= P_{\hat{\mu}, s(\underline{x})}\left\{\frac{\hat{\mu}-(\beta_0+\underline{x}'\underline{\beta}+\underline{x}'B\underline{x})}{\sigma_\varepsilon\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}} \ge \frac{K_p\sqrt{\underline{l}'\underline{l}+\sigma_\varepsilon^2} - ks(\underline{x})}{\sigma_\varepsilon\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\right\}. \tag{3.7.3}$$

Denote $Z = \dfrac{\hat{\mu}-(\beta_0+\underline{x}'\underline{\beta}+\underline{x}'B\underline{x})}{\sigma_\varepsilon\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}$ as a standard normal variate, and define

$$\lambda = \frac{K_p\sqrt{\underline{l}'\underline{l}+\sigma_\varepsilon^2}}{\sigma_\varepsilon\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}} = \frac{K_p\sqrt{\frac{\underline{l}'\underline{l}}{\sigma_\varepsilon^2}+1}}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}. \tag{3.7.4}$$

Expression (3.7.3) can now be written as

$$P_{\hat{\mu}, s(\underline{x})}\left\{Z-\lambda \ge \frac{-k\sqrt{\hat{\underline{l}}'\hat{\underline{l}}+\hat{\sigma}_\varepsilon^2}}{\sigma_\varepsilon\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\right\}$$

$$= P_{\hat{\mu}, s(\underline{x})}\left\{Z-\lambda \ge \frac{-k}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\cdot\sqrt{\frac{\hat{\underline{l}}'\hat{\underline{l}}}{\sigma_\varepsilon^2}+\frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}}\right\}$$

$$= P_{\hat{\mu}, s(\underline{x})}\left\{\frac{Z+\lambda}{\sqrt{\frac{\hat{\underline{l}}'\hat{\underline{l}}}{\sigma_\varepsilon^2}+\frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}}} \le \frac{k}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\right\}. \tag{3.7.5}$$

To find the underlying distribution of the left hand side denominator of the above expression assume that $\widehat{l}_i$'s are orthogonal to each other (see section 3.6) and recall from (3.6.2)

$$\frac{\widehat{l}'\widehat{l}}{\sigma_\varepsilon^2 \mathrm{c}} \sim \chi^{2'}(\mathrm{r_Z}, \frac{l'l}{2\sigma_\varepsilon^2 \mathrm{c}}). \tag{3.7.6}$$

Patnaik (1949) obtained the chi-square approximation to the above noncentral chi-square distribution as

$$\chi^{2'}(\mathrm{r_Z}, \frac{l'l}{2\sigma_\varepsilon^2 \mathrm{c}}) \doteq \rho {\chi^2}_{\nu_1}, \tag{3.7.7}$$

where $\rho = \dfrac{\mathrm{r_Z} + \frac{l'l}{\sigma_\varepsilon^2 \mathrm{c}}}{\mathrm{r_Z} + \frac{l'l}{2\sigma_\varepsilon^2 \mathrm{c}}}$ and $\nu_1 = \dfrac{(\mathrm{r_Z} + \frac{l'l}{2\sigma_\varepsilon^2 \mathrm{c}})^2}{\mathrm{r_Z} + \frac{l'l}{\sigma_\varepsilon^2 \mathrm{c}}}$. Furthermore we have

$$\frac{\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \frac{\chi^2_{\mathrm{dfE}}}{\mathrm{dfE}}. \tag{3.7.8}$$

By equating the first two moments on both sides of

$$\mathrm{c}\rho\chi^2_{\nu_1} + \frac{\chi^2_{\mathrm{dfE}}}{\mathrm{dfE}} \doteq d\chi^2_{\nu_2} \tag{3.7.9}$$

we obtain the approximation to $\frac{\widehat{l}'\widehat{l}}{\sigma_\varepsilon^2} + \frac{\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}$ by $d\chi^2_{\nu_2}$, where $d = \dfrac{(\mathrm{c}\rho)^2\nu_1 + \frac{1}{\mathrm{dfE}}}{(\mathrm{c}\rho)\nu_1 + 1}$ and $\nu_2 = \dfrac{(\mathrm{c}\rho\nu_1 + 1)^2}{(\mathrm{c}\rho)^2\nu_1 + \frac{1}{\mathrm{dfE}}}$. For the more general case of the distribution of linear combinations of χ^2 variates, refer to Satterthwaite (1946) and Welch (1956). Using the final form of (3.7.9) expression (3.7.5) can be written as

$$\mathrm{P}_{\widehat{\mu}, s(\underline{x})}\left\{\frac{\mathrm{Z} + \lambda}{\sqrt{\chi^2_{\nu_2}/\nu_2}} \le \frac{k\sqrt{d\nu_2}}{\sqrt{\underline{x}'(\mathrm{X}'\mathrm{X})^{-1}\underline{x}}}\right\}$$

$$= \mathrm{P}_{\widehat{\mu}, s(\underline{x})}\left\{t'_{\nu_2, \lambda} \le m\right\}$$

$$\geq 1-\gamma, \tag{3.7.10}$$

where $m = \dfrac{k\sqrt{d\nu_2}}{\sqrt{\underline{x}'(\mathrm{X}'\mathrm{X})^{-1}\underline{x}}}$.

The noncentrality parameter λ of (3.7.4) can be estimated in two different ways: either by replacing $\underline{l}'\underline{l}$ and σ_ε^2 with their point estimates, which will eventually lead to an approximate tolerance interval, or by substituting limit values for $\frac{l'l}{\sigma_\varepsilon^2}$ found by the confidence interval procedure (see section 3.6). The other parameters, d and ν_2 can be estimated in a similar fashion. If one substitutes the upper $(1-\alpha)100\%$ confidence interval for $\frac{l'l}{\sigma_\varepsilon^2}$, the derived inequality (3.7.10) would be satisfied.

From inequality (3.7.10) m and, therefore, k can be obtained subsequently. The procedure can be summarized as follows:

1. estimate all related model terms as well as σ_ε^2,
2. pick a point of interest $\underline{\mathrm{x}}$, and decide confidence levels p and γ,
3. calculate $\underline{x}'(\mathrm{X}'\mathrm{X})^{-1}\underline{x}$ for $\underline{\mathrm{x}}$ chosen,
4. calculate the noncentrality parameter λ from (3.7.4),
5. calculate parameters d and ν_2 from (3.7.9),
6. with predetermined γ calculate m from (3.7.10),
7. calculate k from (3.7.10).

In this development, the tolerance interval has made use of the biased estimator of the process variance. The same development can be followed easily if we used the unbiased estimator. For example, when the unbiased estimator for the process variance is used, we would have $s(\underline{\mathrm{x}}) = [\,\widehat{\underline{l}'\underline{l}} + \{1-\mathrm{tr}(\mathrm{C})\}\hat{\sigma}_\varepsilon^2 + \hat{\sigma}_\varepsilon^2]^{1/2}$

and therefore (3.7.5) would be modified to

$$P_{\hat{\mu},\, s(\underline{x})}\left\{\frac{Z+\lambda}{\sqrt{\dfrac{\hat{l}'\hat{l}}{\sigma_\varepsilon^2}+\{2-\mathrm{tr(C)}\}\dfrac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}}} \le \frac{k}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\right\}. \tag{3.7.11}$$

Similar to the development of (3.7.9), we find the following approximation

$$c\rho\chi^2_{\nu_1}+\frac{\{2-\mathrm{tr(C)}\}\chi^2_{\mathrm{dfE}}}{\mathrm{dfE}} \doteq d'\chi^2_{\nu_2'}, \tag{3.7.12}$$

where $d'=\dfrac{(c\rho)^2\nu_1+\dfrac{\{2-\mathrm{tr(C)}\}^2}{\mathrm{dfE}}}{(c\rho)\nu_1+\{2-\mathrm{tr(C)}\}}$ and $\nu_2'=\dfrac{[c\rho\nu_1+\{2-\mathrm{tr(C)}\}]^2}{(c\rho)^2\nu_1+\dfrac{\{2-\mathrm{tr(C)}\}^2}{\mathrm{dfE}}}$. Using this approximation (3.7.11) can be written as

$$\begin{aligned} &P_{\hat{\mu},\, s(\underline{x})}\left\{\frac{Z+\lambda}{\sqrt{\chi^2_{\nu_2'}/\nu_2'}} \le \frac{k\sqrt{d'\nu_2'}}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\right\} \\ &= P_{\hat{\mu},\, s(\underline{x})}\left\{t'_{\nu_2',\,\lambda} \le m'\right\} \\ &\ge 1-\gamma, \end{aligned} \tag{3.7.13}$$

where $m'=\dfrac{k\sqrt{d'\nu_2'}}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}$. The computational procedure is the same as when the biased process variance is used except that d, ν_2, and m are replaced by d', ν_2', and m', respectively.

Another Method of Computation

When actual data are used, the probability statement of (3.7.5) as shown below

$$P_{\widehat{\mu},\, s(\underline{x})}\left\{\frac{Z+\lambda}{\sqrt{\dfrac{\widehat{l'l}}{\sigma_\varepsilon^2}+\dfrac{\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}}} \le \frac{k}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\right\} \ge 1-\gamma$$

can be solved numerically by using multiple integration. This allows us to avoid approximating the underlying distribution of $\dfrac{\widehat{l'l}}{\sigma_\varepsilon^2}+\dfrac{\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}$ to $d\chi^2_{\nu_2}$ (see 3.7.9). This method involves writing out the probability density function of $\dfrac{Z+\lambda}{\sqrt{\dfrac{\widehat{l'l}}{\sigma_\varepsilon^2}+\dfrac{\widehat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}}}$ using the technique of transformations of variables. The details can be seen in Appendix C.

When R^2 is close to 0.8 or higher it appears that the two methods of computation of the tolerance intervals give similar results.

3.8 Two-Sided Tolerance Intervals on Future Observations

Definition 3.8.1

$[\widehat{\mu}-ks(\underline{x}),\ \widehat{\mu}+ks(\underline{x})]$ is p-content two-sided tolerance interval at confidence level $(1-\gamma)$ if $P_{\widehat{\mu},\, s(\underline{x})}\left[\, P_{y_f}\left\{\, \widehat{\mu}-ks(\underline{x}) \le y_f \le \widehat{\mu}+ks(\underline{x})\right\} \ge p\,\right] \ge 1-\gamma$. A p-content two-sided tolerance interval $[\widehat{\mu}-ks(\underline{x}),\ \widehat{\mu}+ks(\underline{x})]$ contains at least $100p\,\%$ of the population or process being sampled with $(1-\gamma)$-confidence level.

We want to compute k so that with probability $(1-\gamma)$ at least a proportion p of the sampled distribution lies inside $[\widehat{\mu}-ks(\underline{x}),\ \widehat{\mu}+ks(\underline{x})]$. To compute the p-content two-sided tolerance interval at confidence level $(1-\gamma)$, we show the case of one control factor and one noise factor. The results can be easily extended to cases with multiple control factors and noise factors.

For the computing details refer to Appendix B. As in the case of one-sided tolerance intervals, when the parameter estimation is very good, this procedure delivers fairly reasonable tolerance intervals. From the last Equation (B.23), k for the two-sided tolerance intervals can be obtained. The procedure can be summarized as follows:

1. estimate all related model terms and σ_ε^2,
2. pick a point of interest $\underline{x}$, and decide confidence levels p and γ,
3. find h_0 such that $g(h_0) = p$ (B.17),
4. calculate k from (B.23).

The one-sided tolerance interval would be appropriate in the "larger the better" case or "smaller the better" case, while the two-sided tolerance interval would be appropriate for the "target is best" case.

Chapter 4. RPD with Categorical Noise Factors

4.1 Estimation of the Minimum Process Variance and Mean Model with Categorical Noise Factors

Many times the noise factors are categorical, such as, different types of automobiles, different operators, different storage conditions, etc. In this section we apply what has been developed thus far to experiments where noise factors are discrete (binomial in particular) instead of continuous variates. The analysis can be greatly simplified by introducing an indicator variable.

Consider a model in control, $\underline{x}$, and noise, $\underline{z}$, for some process with response y given by

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\underline{x} + \underline{z}'\underline{\gamma} + \underline{x}'\Delta\underline{z} + \varepsilon, \tag{4.1.1}$$

where $\varepsilon = N(0,\ \sigma_\varepsilon^2)$. Now suppose we have two noise factors, z_1 and z_2. Suppose noise factors z_1 and z_2, each has three categories, with known probabilities of occurrence in the process or in the field. For example, noise factor z_1 may be three different types of operators and z_2 may be three different operating conditions in which the equipment is operated. In such a case, model (4.1.1) can be rewritten as

$$\begin{aligned} y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\underline{x} + \gamma_1 I_1 + \gamma_2 I_2 + \gamma_3 I_3 + \gamma_4 I_4 + \delta_{11}x_1 I_1 + \delta_{12}x_1 I_2 + \delta_{13}x_1 I_3 \\ + \delta_{14}x_1 I_4 + \delta_{21}x_2 I_1 + \delta_{22}x_2 I_2 + \delta_{23}x_2 I_3 + \delta_{24}x_2 I_4 + \varepsilon, \end{aligned} \tag{4.1.2}$$

where

$$I_1 = \begin{cases} 1, & \text{for operator type 1} \\ 0, & \text{otherwise,} \end{cases} \qquad I_2 = \begin{cases} 1, & \text{for operator type 2} \\ 0, & \text{otherwise,} \end{cases}$$

$$I_3 = \begin{cases} 1, & \text{for condition 1} \\ 0, & \text{otherwise,} \end{cases} \qquad I_4 = \begin{cases} 1, & \text{for condition 2} \\ 0, & \text{otherwise.} \end{cases}$$

Model (4.1.2), which uses indicator variables, is equivalent to the previous model (4.1.1). Indicator variables I_1 and I_2 partitions the first noise factor z_1. Similarly, Indicator variables I_3 and I_4 partitions the second noise factor z_2. Note that these four indicator variables, I_1, I_2, I_3 and I_4, are Bernoulli trials with success probabilities p_1, p_2, p_3, and p_4, respectively. This model can be considered as a special case of the general super model of (3.1.1). In matrix notation (4.1.2) can be written in general as

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\underline{x} + \underline{I}'\underline{\gamma} + \underline{x}'\Delta\underline{I} + \varepsilon, \tag{4.1.3}$$

where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$,

$$\underline{x}' = [\ x_1 \quad x_2 \quad \cdots \quad x_{r_x}\ ],$$

$$\underline{I}' = [\ I_1 \quad I_2 \quad \cdots \quad I_{r_I}\ ],$$

with parameters $\underline{\beta}' = [\ \beta_1 \quad \beta_2 \quad \cdots \quad \beta_{r_x}\ ],$

$$B = \begin{bmatrix} \beta_{11} & \frac{1}{2}\beta_{12} & \frac{1}{2}\beta_{13} & \cdots & \frac{1}{2}\beta_{1r_x} \\ & \beta_{22} & \frac{1}{2}\beta_{23} & \cdots & \frac{1}{2}\beta_{2r_x} \\ & & \cdots & \cdots & \cdots \\ & (\text{sym}) & & & \frac{1}{2}\beta_{r_x r_x} \end{bmatrix},$$

$$\underline{\gamma}' = [\ \gamma_1 \quad \gamma_2 \quad \cdots \quad \gamma_{r_I}\ ],$$

and $$\Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \cdots & \delta_{1r_I} \\ \delta_{21} & \delta_{22} & \cdots & \delta_{2r_I} \\ \cdots & \cdots & \cdots & \cdots \\ \delta_{r_X1} & \delta_{r_X2} & \cdots & \delta_{r_Xr_I} \end{bmatrix}.$$

In the above r_I is the number of indicator variables. Thus $\underline{\beta}$ is $r_X \times 1$, $\underline{\gamma}$ is $r_I \times 1$, B is $r_X \times r_X$, and Δ is $r_X \times r_I$.

Recall that the indicator variables, $\underline{I}$, are Bernoulli trials and therefore, we have

$$E(\underline{I}) = [\, p_1 \quad p_2 \quad \cdots \quad p_{r_I} \,]'$$

$$\mathrm{Var}(\underline{I}) = V = \begin{bmatrix} V_1 & & \underline{0} & \\ & V_2 & & \\ & & \cdots & \\ & \underline{0} & & V_{r_I} \end{bmatrix}, \tag{4.1.4}$$

where V_i's are block matrices. For example, V_1 can be written as

$$V_1 = \begin{bmatrix} p_1(1-p_1) & -(p_1)(p_2) & \cdots & \cdots \\ & p_2(1-p_2) & -(p_2)(p_3) & \cdots \\ & & \cdots & \cdots \\ & (\mathrm{sym}) & & p_l(1-p_l) \end{bmatrix},$$

where the size of a block, l, is the number of indicator variables generated by partitioning of the first noise factor. Each block corresponds to the indicator variable generated by partitioning of noise factors. The diagonal elements of the variance-covariance matrix V of (4.1.4) are variances of a Bernoulli trial, $p_i(1-p_i)$, and the off-diagonal elements of each "block" submatrix are covariances among Bernoulli trials, $-p_ip_j$, where p_i and p_j are success probabilities of indicators I_i and I_j $(i \neq j)$, respectively, and each submatrix corresponds to the indicator variable generated by partitioning of noise factors.

The process mean of model (4.1.3) is given by

$$E_{\underline{z}}(y) = [\beta_0 + \{E(\underline{I})\}'\underline{\gamma}] + \underline{x}'[\underline{\beta} + \Delta\{E(\underline{I})\}] + \underline{x}'B\underline{x}, \tag{4.1.5}$$

where $E(\underline{I})$ is given in (4.1.4). The reader should notice that the response surface model for the mean with categorical noise factors is quite different from the case where the noise factors are all continuous. Now let us turn our attention to the process variance. In order to evaluate the process variance we consider the model of Equation (4.1.3). The process variance is given by

$$\mathrm{Var}_{\underline{z}}(y\,|\,\underline{x}) = (\underline{\gamma}' + \underline{x}'\Delta)V(\underline{\gamma} + \Delta'\underline{x}) + \sigma_\varepsilon^2, \tag{4.1.6}$$

where the variance-covariance matrix of $\underline{I}$'s, V, is given in (4.1.4). Rewrite Equation (4.1.6) as

$$\mathrm{Var}_{\underline{z}}(y\,|\,\underline{x}) = \underline{l}'V\underline{l} + \sigma_\varepsilon^2, \tag{4.1.7}$$

$$\text{where } \underline{l} = \frac{\partial y}{\partial \underline{I}} = (\underline{\gamma} + \Delta'\underline{x}).$$

The $\underline{l}$'s are derivatives of the response with respect to the indicator variables (partitioning of noise factors) and vector $\underline{l}$ consists of linear polynomial functions in the control factors as before.

As a biased estimator of the process variance we have

$$\hat{\mathrm{Var}}_{\underline{\mathrm{Z}}}(y \mid \underline{\mathrm{x}}) = \hat{\underline{l}}'\mathrm{V}\hat{\underline{l}} + \mathrm{s}^2, \tag{4.1.8}$$

where s^2 is an estimate of the error variance. This estimator is biased since

$$\mathrm{E}\{\hat{\mathrm{Var}}_{\underline{\mathrm{Z}}}(y \mid \underline{\mathrm{x}})\} = \underline{l}'\mathrm{V}\underline{l} + \sigma_\varepsilon^2 \,\mathrm{tr}(\mathrm{CV}) + \sigma_\varepsilon^2, \tag{4.1.9}$$

where C is the variance-covariance matrix of $\hat{\underline{l}}$ apart from σ_ε^2 and tr (CV) is a trace of the matrix $\mathrm{C} \times \mathrm{V}$. From Equation (4.1.9) we see that an unbiased estimator of the process variance can now be written as

$$\hat{\mathrm{Var}}_{\underline{\mathrm{Z}}}(y \mid \underline{\mathrm{x}}) = \hat{\underline{l}}'\mathrm{V}\hat{\underline{l}} + \{1 - \mathrm{tr}\,(\mathrm{CV})\}\,\mathrm{s}^2. \tag{4.1.10}$$

This is unbiased since

$$\begin{aligned}\mathrm{E}[\hat{\underline{l}}'\mathrm{V}\hat{\underline{l}} + \{1 - \mathrm{tr}\,(\mathrm{CV})\}\,\mathrm{s}^2] &= \underline{l}'\mathrm{V}\underline{l} + \sigma_\varepsilon^2 \,\mathrm{tr}(\mathrm{CV}) + \{1 - \mathrm{tr}\,(\mathrm{CV})\}\sigma_\varepsilon^2 \\ &= \underline{l}'\mathrm{V}\underline{l} + \sigma_\varepsilon^2. \end{aligned} \tag{4.1.11}$$

4.2 Comparison of the Mean Square Error between the Biased Estimator and the Unbiased Estimator of the Process Variance with Categorical Noise Factors

Consider the process variance of (4.1.7)

$$\mathrm{Var}_{\underline{\mathrm{Z}}}(y \mid \underline{\mathrm{x}}) = \underline{l}'\mathrm{V}\underline{l} + \sigma_\varepsilon^2. \tag{4.2.1}$$

In the preceding section 4.1, we see that the biased estimator of the process variance when indicator variables are used (i. e., noise factors are categorical) is

$$\hat{\mathrm{Var}}_{\underline{\mathrm{Z}}}(y \mid \underline{\mathrm{x}}) = \hat{\underline{l}}'\mathrm{V}\hat{\underline{l}} + \mathrm{s}^2, \tag{4.2.2}$$

and the unbiased estimator of the process variance is

$$\hat{\mathrm{Var}}_{\underline{\mathrm{Z}}}(y \mid \underline{\mathrm{x}}) = \hat{\underline{l}}'\mathrm{V}\hat{\underline{l}} + \{1 - \mathrm{tr}\,(\mathrm{CV})\}\,\mathrm{s}^2. \tag{4.2.3}$$

To find tr (CV), recall from (3.2.4)

$$\hat{l}_i = (1, \underline{x}')\hat{\eta}_i,$$

where $(1, \underline{x}') = (1, x_1, x_2, ..., x_{r_X})$, $\hat{\eta}_i' = (\hat{\gamma}_i, \hat{\delta}_{1i}, \hat{\delta}_{2i}, ..., \hat{\delta}_{r_X i})$, and $i = 1, 2, ..., r_I$. Then,

$$\text{Var}\,(\hat{l}_i) = (1, \underline{x}')\text{Var}\,(\hat{\eta}_i)\binom{1}{\underline{x}}. \tag{4.2.4}$$

Note that $\text{Var}\,(\hat{\eta}_i)$ apart from σ_ε^2 is a submatrix of the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix where $\mathbf{X}$ is a model matrix of a general model (4.1.3) that includes both control factors and indicator variables. Recall matrices $\sigma_\varepsilon^2\text{C}$ of Equation (4.1.9) and V of Equation (4.1.4). From Equation (4.2.4) tr (CV) can be written as

$$\begin{aligned}
\text{tr (CV)} &= \sum_{\substack{i,\, j\\ i \neq j}}^{r_I} \left\{\text{Var}(\hat{l}_i)\text{Var}(\text{I}_i) + \text{Cov}(\hat{l}_i, \hat{l}_j)\text{Cov}(\text{I}_i, \text{I}_j)\right\} \\
&= \sum_{\substack{i,\, j\\ i \neq j}}^{r_I} \left\{(1, \underline{x}')\text{Var}\,(\hat{\eta}_i)\binom{1}{\underline{x}}\, p_i(1-p_i) + \text{Cov}(\hat{l}_i, \hat{l}_j)\text{Cov}(\text{I}_i, \text{I}_j)\right\} \\
&= \sum_{\substack{i,\, j\\ i \neq j}}^{r_I} \left\{p_i(1-p_i)\text{Var}\,(\hat{\gamma}_i) + \text{Cov}(\hat{\gamma}_i, \hat{\gamma}_j)\text{Cov}(\text{I}_i, \text{I}_j)\right\} \\
&\quad + \sum_{i=1}^{r_I}\sum_{k=1}^{r_X} \left\{p_i(1-p_i)\, x_k^2\, \text{Var}\,(\hat{\delta}_{ki})\right\} + \sum_{\substack{i,\, j\\ i \neq j}}^{r_I}\sum_{k=1}^{r_X} x_k^2\, \text{Cov}(\hat{\delta}_{ki}, \hat{\delta}_{k,i+1})\text{Cov}(\text{I}_i, \text{I}_j) \\
&= (1, \underline{x}')\begin{bmatrix} c^* & \underline{0} \\ \underline{0} & \underline{M}^* \end{bmatrix}\binom{1}{\underline{x}},
\end{aligned} \tag{4.2.5}$$

where $$c^* = \sum_{\substack{i,\, j\\ i \neq j}}^{r_I} \left\{p_i(1-p_i)\text{Var}\,(\hat{\gamma}_i) + \text{Cov}(\hat{\gamma}_i, \hat{\gamma}_j)\text{Cov}(\text{I}_i, \text{I}_j)\right\}$$

and

$$\underline{M}^* = \begin{bmatrix} M_1 & 0 & \cdots & \cdots & 0 \\ & \cdots & \cdots & \cdots & \cdots \\ & & M_k & \cdots & \cdots \\ & & & \cdots & \cdots \\ & (\text{sym}) & & & M_{r_X} \end{bmatrix}$$

with

$$M_k = \sum_{i=1}^{r_I} \left\{ p_i(1-p_i)\mathrm{Var}\,(\hat{\delta}_{ki}) \right\} + \sum_{\substack{i,\,j \\ i \neq j}}^{r_I} \mathrm{Cov}(\hat{\delta}_{ki},\, \hat{\delta}_{k,i+1})\mathrm{Cov}(I_i,\, I_j). \tag{4.2.6}$$

We will again address the question "which estimator is better?" by looking at the minimization of the mean square error (MSE), which is defined as

$$\mathrm{MSE} = \mathrm{Variance} + (\mathrm{Bias})^2. \tag{4.2.7}$$

We first consider the case of the biased estimator (4.2.2)

$$\mathrm{MSE}\,(\hat{\underline{l}}'V\hat{\underline{l}} + s^2) = \mathrm{Var}\,(\hat{\underline{l}}'V\hat{\underline{l}} + s^2) + \{\mathrm{Bias}\,(\hat{\underline{l}}'V\hat{\underline{l}} + s^2)\}^2. \tag{4.2.8}$$

We have

$$\mathrm{Var}\,(\hat{\underline{l}}'V\hat{\underline{l}} + s^2) = \mathrm{Var}(\hat{\underline{l}}'V\hat{\underline{l}}) + \frac{2\sigma_\varepsilon^4}{\mathrm{dfE}}. \tag{4.2.9}$$

From (4.1.9) we see that

$$\mathrm{Bias}\,(\hat{\underline{l}}'V\hat{\underline{l}} + s^2) = \sigma_\varepsilon^2\,\mathrm{tr}(CV). \tag{4.2.10}$$

Combining (4.2.9) and (4.2.10) we get

$$\begin{aligned} \mathrm{MSE}\,(\hat{\underline{l}}'V\hat{\underline{l}} + s^2) &= \mathrm{Var}\,(\hat{\underline{l}}'V\hat{\underline{l}} + s^2) + \{\mathrm{Bias}\,(\hat{\underline{l}}'V\hat{\underline{l}} + s^2)\}^2 \\ &= \mathrm{Var}(\hat{\underline{l}}'V\hat{\underline{l}}) + \sigma_\varepsilon^4\,\{\frac{2}{\mathrm{dfE}} + \mathrm{tr}^2(CV)\}. \end{aligned} \tag{4.2.11}$$

Now consider the case of the unbiased estimator (4.2.3)

$$\mathrm{MSE}\,[\hat{\underline{l}}'V\hat{\underline{l}} + \{1 - \mathrm{tr}\,(CV)\}\,s^2] = \mathrm{Var}\,[\hat{\underline{l}}'V\hat{\underline{l}} + \{1 - \mathrm{tr}\,(CV)\}\,s^2]. \tag{4.2.12}$$

Using the result of (4.2.9) we have

$$\text{Var}\,[\hat{\underline{l}}'\text{V}\hat{\underline{l}} + \{1 - \text{tr (CV)}\}\ \text{s}^2] = \text{Var}(\hat{\underline{l}}'\text{V}\hat{\underline{l}}) + \frac{2\sigma_\varepsilon^4\{1 - \text{tr (CV)}\}^2}{\text{dfE}}. \tag{4.2.13}$$

Therefore,

$$\begin{aligned}\text{MSE}\,[\hat{\underline{l}}'\text{V}\hat{\underline{l}} + \{1 - \text{tr (CV)}\}\ \text{s}^2] &= \text{Var}\,[\hat{\underline{l}}'\text{V}\hat{\underline{l}} + \{1 - \text{tr (CV)}\}\ \text{s}^2] \\ &= \text{Var}(\hat{\underline{l}}'\text{V}\hat{\underline{l}}) + \frac{2\sigma_\varepsilon^4\{1 - \text{tr (CV)}\}^2}{\text{dfE}}. \end{aligned} \tag{4.2.14}$$

Now in order to compare (4.2.11) and (4.2.14), we look at

$$\begin{aligned}&\text{MSE}\,(\hat{\underline{l}}'\text{V}\hat{\underline{l}} + \text{s}^2) - \text{MSE}\,[\hat{\underline{l}}'\text{V}\hat{\underline{l}} + \{1 - \text{tr (CV)}\}\ \text{s}^2] \\ &= \sigma_\varepsilon^4\,[\frac{2}{\text{dfE}} + \text{tr}^2(\text{CV}) - \frac{2\{1 - \text{tr (CV)}\}^2}{\text{dfE}}] \\ &= \frac{\sigma_\varepsilon^4\,\text{tr(CV)}}{\text{dfE}}\,[(\text{dfE} - 2)\{\text{tr(CV)}\} + 4]. \end{aligned} \tag{4.2.15}$$

When tr(CV) is positive and $\text{dfE} > 1$ the right hand side of Equation (4.2.15) exceeds 0. In this case, we see that the unbiased estimator of the process variance uniformly gives smaller mean square error than the biased estimator. Again the use of the unbiased estimator is strongly recommended. It turns out that the two estimators often show similar results even when $\text{dfE} \leq 1$. An example will be shown later to illustrate this point.

4.3 General Stationary Region Analysis and Ridge Analysis using the Unbiased Estimator of the Process Variance with Categorical Noise Factors

To find the location of the estimated minimum process variance we differentiate the unbiased estimator of the process variance of Equation (4.1.11). From (4.1.7) we have $\hat{\underline{l}} = (\hat{\underline{\gamma}} + \hat{\Delta}'\underline{x})$ and by using a chain rule, we see that

$$\frac{\partial}{\partial \underline{x}}(\hat{\underline{l}}'V\hat{\underline{l}}) = [\frac{\partial \hat{\underline{l}}}{\partial \underline{x}'}]' \frac{\partial}{\partial \hat{\underline{l}}}(\hat{\underline{l}}'V\hat{\underline{l}})$$

$$= 2(\hat{\Delta})(V\hat{\underline{l}}) = 2\hat{\Delta}V(\hat{\underline{\gamma}} + \hat{\Delta}'\underline{x}) = 2(\hat{\Delta}V\hat{\underline{\gamma}} + \hat{\Delta}V\hat{\Delta}'\underline{x}). \tag{4.3.1}$$

Recall from (4.2.5) that

$$\mathrm{tr}\,(CV) = (1, \underline{x}')\begin{bmatrix} c^* & \underline{0} \\ \underline{0} & \underline{M}^* \end{bmatrix}\binom{1}{\underline{x}}.$$

Therefore,

$$\frac{\partial}{\partial \underline{x}}\mathrm{tr}(CV) = 2\begin{bmatrix} x_1 \sum_i^{r_I}\left\{p_i(1-p_i)\mathrm{Var}(\hat{\delta}_{1i}) + \sum_{i \neq j}^{r_I}\mathrm{Cov}(\hat{\delta}_{ki}, \hat{\delta}_{ki+1})\mathrm{Cov}(I_i, I_j)\right\} \\ x_2 \sum_i^{r_I}\left\{p_i(1-p_i)\mathrm{Var}(\hat{\delta}_{2i}) + \sum_{i \neq j}^{r_I}\mathrm{Cov}(\hat{\delta}_{ki}, \hat{\delta}_{ki+1})\mathrm{Cov}(I_i, I_j)\right\} \\ \cdots \\ x_{r_x} \sum_i^{r_I}\left\{p_i(1-p_i)\mathrm{Var}(\hat{\delta}_{r_x i}) + \sum_{i \neq j}^{r_I}\mathrm{Cov}(\hat{\delta}_{ki}, \hat{\delta}_{li})\mathrm{Cov}(I_i, I_j)\right\} \end{bmatrix} = 2\underline{M}^*\underline{x}. \tag{4.3.2}$$

From (4.3.1) and (4.3.2) we have

$$\frac{\partial}{\partial \underline{x}}[\hat{\underline{l}}'V\hat{\underline{l}} + \{1 - \mathrm{tr}(CV)\}\, s^2] = 2\left(\hat{\Delta}V\hat{\underline{\gamma}} + \hat{\Delta}V\hat{\Delta}'\underline{x} - s^2\underline{M}^*\underline{x}\right). \tag{4.3.3}$$

Equate (4.3.3) to $\underline{0}$ and solve for $\underline{x}$, we have

$$\hat{\Delta}V\hat{\underline{\gamma}} + (\hat{\Delta}V\hat{\Delta}' - s^2\underline{M}^*)\,\underline{x} = \underline{0}. \tag{4.3.4}$$

Therefore, as a solution to the Equation (4.3.4), we get

$$\hat{\underline{x}}_0 = -(\hat{\Delta}V\hat{\Delta}' - s^2\underline{M}^*)^{-1}\hat{\Delta}V\hat{\underline{\gamma}}, \tag{4.3.5}$$

if the solution exists. The eigenvalues of the matrix $(\hat{\Delta}V\hat{\Delta}' - s^2\underline{M}^*)$ in Equation (4.3.5) again determine the nature of the stationary point.

Next, let us minimize the process variance of the model (4.1.3),

$$\min_{\underline{x}} \mathrm{Var}_{\underline{z}}(y \mid \underline{x}) \tag{4.3.6}$$

while restricting ourselves to spheres of varying radii. We use the usual coding of the factors, with the origin of the design at (0, 0, ..., 0), for a point $(x_1, x_2, ..., x_{r_x})$ on a sphere of radius R,

$$\sum_{i=1}^{r_x} x_i^2 = R^2. \tag{4.3.7}$$

Consider the unbiased estimator of $\underline{l}'V\underline{l}$, $\hat{\underline{l}}'V\hat{\underline{l}} - s^2\,\mathrm{tr}(CV)$. Similar to what we have done previously, the problem reduces to finding the stationary (minimum) point of the second order response system

$$\begin{aligned}\hat{\underline{l}}'V\hat{\underline{l}} - s^2\,\mathrm{tr}(CV) &= (\hat{\underline{\gamma}}' + \underline{x}'\hat{\Delta})V(\hat{\underline{\gamma}} + \hat{\Delta}'\underline{x}) - s^2\,\mathrm{tr}(CV) \\ &= \hat{\underline{\gamma}}'V\hat{\underline{\gamma}} + 2\,\underline{x}'\hat{\Delta}V\hat{\underline{\gamma}} + \underline{x}'(\hat{\Delta}V\hat{\Delta}')\underline{x} \\ &\quad - s^2(1, \underline{x}')\begin{bmatrix} c^* & \underline{0} \\ \underline{0} & \underline{M}^* \end{bmatrix}\binom{1}{\underline{x}},\end{aligned} \tag{4.3.8}$$

under the constraint of (4.3.7).

We should attempt to find the conditions on the factors which *minimize* Equation (4.3.8) subject to the constraint given by Equation (4.3.7) for various R. After these conditions are found R can be plotted against the appropriate coordinates $x_1, x_2, ..., x_{r_x}$, and $\hat{\underline{l}}'V\hat{\underline{l}}$. To minimize $\hat{\underline{l}}'V\hat{\underline{l}} - s^2\,\mathrm{tr}\,(CV)$ subject to the constraint of (4.3.7), consider the function

$$F = \hat{\underline{l}}'V\hat{\underline{l}} - s^2\,\mathrm{tr}(CV) - \mu(\underline{x}'\underline{x} - R^2) \tag{4.3.9}$$

where μ is a Lagrangian multiplier and $\underline{x}' = (x_1,\ x_2,\ \ldots,\ x_{r_x})$. Using Equations (4.3.8) and (4.3.2), we get

$$\frac{\partial F}{\partial \underline{x}} = 2\,\{\hat{\Delta}V\hat{\underline{\gamma}} + (\hat{\Delta}V\hat{\Delta}' - s^2\underline{M}^*)\underline{x} - \mu\underline{x}\}. \tag{4.3.10}$$

Equating Equation (4.3.10) to $\underline{0}$, and solving for $\underline{x}$, we get

$$(\hat{\Delta}V\hat{\Delta}' - s^2\underline{M}^* - \mu I_{r_x})\,\underline{x} = -\hat{\Delta}V\hat{\underline{\gamma}}. \tag{4.3.11}$$

One can insert predetermined values of μ into Equation (4.3.11), solve for x_1, x_2, ..., x_{r_x}, and $\hat{\underline{l}}'V\hat{\underline{l}} - s^2\,\mathrm{tr}(CV)$ from Equation (4.3.8). Thus this method generates stationary (minimum) points on spheres of varying radii. Again, the nature of the stationary points depends on the value of μ chosen. Values of μ which are smaller than the smallest eigenvalue of $(\hat{\Delta}V\hat{\Delta}' - s^2\underline{M}^*)$ will give rise to an $\underline{x}$ which gives an absolute minimum on the sphere of radius R. The proof is similar to that in Appendix A and thus will be omitted here.

Next, to find the location of minimum process variance under general constraint set

$$\{\underline{x}\colon g(\underline{x}) = C\} \tag{4.3.12}$$

using an unbiased estimator for $\underline{l}'V\underline{l}$, $\hat{\underline{l}}'V\hat{\underline{l}} - s^2\,\mathrm{tr}(CV)$, consider a Lagrangian form

$$L = \{\hat{\underline{l}}'V\hat{\underline{l}} - s^2\,\mathrm{tr}(CV)\} - \mu\,\{g(\underline{x}) - C\}. \tag{4.3.13}$$

To build a confidence region around the constrained minimum process variance location consider

$$\frac{\partial \mathrm{L}}{\partial \underline{\mathrm{x}}} = 2\,\{\widehat{\Delta}\mathrm{V}\widehat{\underline{\gamma}} + (\widehat{\Delta}\mathrm{V}\widehat{\Delta}' - \mathrm{s}^2\underline{\mathrm{M}}^*)\underline{\mathrm{x}}\} - \frac{\partial}{\partial \underline{\mathrm{x}}}\,\mu\,\{g(\underline{\mathrm{x}})\}. \tag{4.3.14}$$

Now equate (4.3.14) to $\underline{0}$ and solve for $\underline{\mathrm{x}}$. We have a ridge system. An appropriate choice of μ will give $\underline{\mathrm{x}}$ which gives an absolute minimum inside the constraint set. As before $\left(\frac{\widehat{\partial \mathrm{L}}}{\partial \underline{\mathrm{x}}_0}\right) = \left(\frac{\widehat{\partial \mathrm{L}}}{\partial \underline{\mathrm{x}}}\right)\Big|_{\underline{\mathrm{x}}=\underline{\mathrm{x}}_0}$ is normally distributed with mean $\underline{0}$ and variance-covariance matrix V^* by assuming that $\widehat{\mu}$ is held as a constant and $\mathrm{E}\,(\widehat{\mu}) \doteq \mu$. Consequently, a $(1-\alpha)100\%$ confidence region for the constrained minimum process variance location for a fixed μ can be found by

$$\mathrm{P}\,\{\frac{\left(\frac{\widehat{\partial \mathrm{L}}}{\partial \underline{\mathrm{x}}_0}\right)'\widehat{\mathrm{V}}^{*-1}\left(\frac{\widehat{\partial \mathrm{L}}}{\partial \underline{\mathrm{x}}_0}\right)}{\mathrm{r}_\mathrm{I}} \le \mathrm{F}_{\mathrm{r}_\mathrm{I},\ \mathrm{dfE};\ 1-\alpha}\} = 1-\alpha. \tag{4.3.15}$$

Since L of (4.3.13) can be often rewritten as a second-order response surface

$$\mathrm{L} = \widehat{\underline{\gamma}}'\mathrm{V}\widehat{\underline{\gamma}} + 2\,\underline{\mathrm{x}}'\widehat{\Delta}\mathrm{V}\widehat{\underline{\gamma}} + \underline{\mathrm{x}}'(\widehat{\Delta}\mathrm{V}\widehat{\Delta}')\underline{\mathrm{x}} - \mathrm{s}^2(1,\,\underline{\mathrm{x}}')\begin{bmatrix} \mathrm{c}^* & \underline{0} \\ \underline{0} & \underline{\mathrm{M}}^* \end{bmatrix}\binom{1}{\underline{\mathrm{x}}} - \mu\,\{g(\underline{\mathrm{x}}) - \mathrm{C}\}, \tag{4.3.16}$$

one often selects μ smaller than the smallest eigenvalue of $(\widehat{\Delta}\mathrm{V}\widehat{\Delta}' - \mathrm{s}^2\underline{\mathrm{M}}^*)$ as in standard ridge analysis. In the probability statement of (4.3.15), $\widehat{\mathrm{V}}^*$ is a function of the design used, s^2, and $\underline{\mathrm{x}}_0'$. Values of $\underline{\mathrm{x}}_0'$ that satisfy the Equation (4.3.15) lie inside the $(1-\alpha)100\%$ confidence region on the location of the absolute minimum process variance inside the constraint set.

What have been described in this section was developed by using the unbiased estimator for $\underline{l}'\mathrm{V}\underline{l}$, $\widehat{\underline{l}}'\mathrm{V}\widehat{\underline{l}} - \mathrm{s}^2\,\mathrm{tr}(\mathrm{CV})$. Similar results can be declared with biased estimator, $\widehat{\underline{l}}'\mathrm{V}\widehat{\underline{l}}$. For example, consider the stationary region analysis by using the biased estimator. The estimated location of the minimum process variance location would be

$$\hat{\underline{x}}_0 = -(\hat{\Delta}V\hat{\Delta}')^{-1}\hat{\Delta}V\hat{\underline{\gamma}}. \tag{4.3.17}$$

This can be compared to (4.3.5), the solution when the unbiased estimator was used. Similarly, using the biased estimator under the constraint of (4.3.7) would give a location that satisfies the following Equation

$$(\hat{\Delta}V\hat{\Delta}' - \mu I_{r_x})\,\underline{x} = -\hat{\Delta}V\hat{\underline{\gamma}}. \tag{4.3.18}$$

This location can be again compared to (4.3.11), the expression when the unbiased estimator was used. Finally, using the biased estimator under the general constraint set of (4.3.12) would result in a location that satisfies

$$2\,\{\hat{\Delta}V\hat{\underline{\gamma}} + (\hat{\Delta}V\hat{\Delta}')\underline{x}\} - \frac{\partial}{\partial \underline{x}}\,\mu\,\{g(\underline{x})\} = 0. \tag{4.3.19}$$

Compare this with (4.3.14), the location when the unbiased estimator was used. A $(1-\alpha)100\%$ confidence region for the location of the minimum process variance under the general constraint set can be written similar to (4.3.15) with appropriate derivative and variance-covariance matrix. Comparison between two locations, the locations of the estimated minimum process variance when the unbiased estimator is used compared to the location when the biased estimator is used, leads us to similar conclusions as those described in section 3.3.

4.4 One-Sided Tolerance Intervals on Future Observations with Categorical Noise Factors

In this section we discuss construction of p-content upper one-sided tolerance interval at confidence level $(1-\gamma)$ as discussed in Definition 3.7.1. That is, we look for an interval of the form $(-\infty,\ \widehat{\mu}+ks(\underline{x})]$ so that we can be $(1-\gamma)100\,\%$ sure that at least $100p\,\%$ of the population is below the limit. We wish to compute k so that with probability $(1-\gamma)$ at least a proportion p of the sampled distribution will be below $\widehat{\mu}+ks(\underline{x})$.

Assume a general model of (4.1.3) with categorical noise factors

$$y=\beta_0+\underline{x}'\underline{\beta}+\underline{x}'\mathrm{B}\underline{x}+\underline{\mathrm{I}}'\underline{\gamma}+\underline{x}'\Delta\underline{\mathrm{I}}+\varepsilon,$$

where $\underline{x}$ and $\underline{\mathrm{I}}$ denote control factors and indicator variables (i. e., partitions of binomial noise factors), respectively. The upper one-sided p-content tolerance interval at confidence level $(1-\gamma)$ is

$$\mathrm{P}_{\widehat{\mu},\,s(\underline{x})}\left[\,\mathrm{P}_{y_f}\left\{y_f\ \le\ \widehat{\mu}+ks(\underline{x})\right\}\ \ge\ p\,\right]\ \ge\ 1-\gamma, \tag{4.4.1}$$

where the estimated mean $\widehat{\mu}=[\widehat{\beta}_0+\{\mathrm{E}(\underline{\mathrm{I}})\}'\widehat{\underline{\gamma}}]+\underline{x}'[\widehat{\underline{\beta}}+\widehat{\Delta}\{\mathrm{E}(\underline{\mathrm{I}})\}]+\underline{x}'\widehat{\mathrm{B}}\underline{x}$, the square root of the estimated process variance $s(\underline{x})=\widehat{\mathrm{Var}}(y)^{1/2}=[\,\widehat{\underline{l}}'\mathrm{V}\widehat{\underline{l}}+\widehat{\sigma}^2_\varepsilon\{1\text{-}\mathrm{tr}(\mathrm{CV})\}]^{1/2}$, $\widehat{\underline{l}}=\widehat{\underline{\gamma}}+\widehat{\Delta}'\,\underline{x}$, and $\widehat{\sigma}^2_\varepsilon=$ estimated MSE from the model. Note that $s(\underline{x})$ is a function of $\underline{x}$. Inequality (4.4.1) can be written as

$$\mathrm{P}_{\widehat{\mu},\,s(\underline{x})}\,[\mathrm{P}_{y_f}\left\{\frac{y_f-\mathrm{E}(y_f)}{\sqrt{\underline{l}'\mathrm{V}\underline{l}+\sigma^2_\varepsilon}}\le\frac{\widehat{\mu}-\mathrm{E}(y_f)+ks(\underline{x})}{\sqrt{\underline{l}'\mathrm{V}\underline{l}+\sigma^2_\varepsilon}}\right\}\ge p]\ge 1-\gamma, \tag{4.4.2}$$

where $\mathrm{E}(y_f)=[\beta_0+\{\mathrm{E}(\underline{\mathrm{I}})\}'\underline{\gamma}]+\underline{x}'[\underline{\beta}+\Delta\{\mathrm{E}(\underline{\mathrm{I}})\}]+\underline{x}'\mathrm{B}\underline{x}$ is the true mean (see 4.1.5). Using the pth quantile point of the standard normal distribution Expression (4.4.2) can be written as

$$P_{\widehat{\mu}, s(\underline{x})}\left\{\frac{\widehat{\mu} - E(y_f) + ks(\underline{x})}{\sqrt{\underline{l}'V\underline{l} + \sigma_\varepsilon^2}} \geq K_p\right\} \geq 1 - \gamma, \tag{4.4.3}$$

where K_p is the pth quantile of the standard normal distribution. This development follows very closely the one in section 3.7. In this case, we will use the unbiased estimator for the process variance. The left hand side of Expression (4.4.3) can be written as

$$P_{\widehat{\mu}, s(\underline{x})}\left\{\widehat{\mu} - E(y_f) \geq K_p\sqrt{\underline{l}'V\underline{l} + \sigma_\varepsilon^2} - ks(\underline{x})\right\}$$

$$= P_{\widehat{\mu}, s(\underline{x})}\left\{\frac{\widehat{\mu} - E(y_f)}{\sigma_\varepsilon\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}} \geq \frac{K_p\sqrt{\underline{l}'V\underline{l} + \sigma_\varepsilon^2} - ks(\underline{x})}{\sigma_\varepsilon\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\right\}. \tag{4.4.4}$$

Denote $Z = \dfrac{\widehat{\mu} - E(y_f)}{\sigma_\varepsilon\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}$ as a standard normal variate, and define

$$\lambda = \frac{K_p\sqrt{\underline{l}'V\underline{l} + \sigma_\varepsilon^2}}{\sigma_\varepsilon\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}} = \frac{K_p\sqrt{\dfrac{\underline{l}'V\underline{l}}{\sigma_\varepsilon^2} + 1}}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}. \tag{4.4.5}$$

Expression (4.4.4) can now be written as

$$P_{\widehat{\mu}, s(\underline{x})}\left\{Z - \lambda \geq \frac{-k\sqrt{\widehat{\underline{l}}'V\widehat{\underline{l}} + \widehat{\sigma}_\varepsilon^2\{1\text{-tr}(CV)\}}}{\sigma_\varepsilon\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\right\}$$

$$= P_{\widehat{\mu}, s(\underline{x})}\left\{\frac{Z + \lambda}{\sqrt{\dfrac{\widehat{\underline{l}}'V\widehat{\underline{l}}}{\sigma_\varepsilon^2} + \dfrac{\widehat{\sigma}_\varepsilon^2\{1\text{-tr}(CV)\}}{\sigma_\varepsilon^2}}} \leq \frac{k}{\sqrt{\underline{x}'(X'X)^{-1}\underline{x}}}\right\} \geq 1 - \gamma\,. \tag{4.4.6}$$

When actual data are used, the probability statement (4.4.6) can be solved numerically by multiple integration (see section 3.7). In addition, the left hand side can be approximated by a noncentral t. The noncentrality parameter λ of

(4.4.5) can be estimated by replacing $\underline{l}'V\underline{l}$ and σ_ε^2 with their point estimates, which will eventually lead to an approximate tolerance interval. The procedure can be summarized as follows:

1. estimate all model terms and σ_ε^2,
2. pick a point of interest $\underline{x}$, and decide confidence levels p and γ,
3. calculate $\underline{x}'(X'X)^{-1}\underline{x}$ for $\underline{x}$ chosen,
4. calculate the noncentrality parameter λ from (4.4.5),
5. with predetermined γ calculate k from (4.4.6),

4.5 Estimation of the Minimum Process Variance and Mean Model with Mixed Noise Factors – Some Continuous and Others Categorical

We now look at the case where some of the noise factors are continuous and others are discrete (binomial, in particular). As an illustration consider a model in two control factors, x_1 and x_2, and four noise factors, z_1, z_2, z_3 and z_4. Suppose that z_1 and z_2 are continuous and z_3 and z_4 are binomial random variates. Furthermore, suppose that both z_1 and z_2 indicate three possible categories to choose from. We can write a model:

$$\begin{aligned} y=&\beta_0+\beta_1x_1+\beta_2x_2+\beta_{12}x_1x_2+\beta_{11}x_1^2+\beta_{22}x_2^2+\gamma_1z_1+\gamma_2z_2+\gamma_1^*I_1+\gamma_2^*I_2+\gamma_3^*I_3+\gamma_4^*I_4\\ &+\delta_{11}x_1z_1+\delta_{12}x_1z_2+\delta_{11}^*x_1I_1+\delta_{12}^*x_1I_2+\delta_{13}^*x_1I_3+\delta_{14}^*x_1I_4\\ &+\delta_{21}x_2z_1+\delta_{22}x_2z_2+\delta_{21}^*x_2I_1+\delta_{22}^*x_2I_2+\delta_{23}^*x_2I_3+\delta_{24}^*x_2I_4+\varepsilon, \end{aligned} \tag{4.5.1}$$

where $\varepsilon = N(0, \sigma_\varepsilon^2)$ and indicator variables are

$$I_1 = \begin{cases} 1, & \text{if } z_3 \text{ is in category 1} \\ 0, & \text{otherwise,} \end{cases} \qquad I_2 = \begin{cases} 1, & \text{if } z_3 \text{ is in category 2} \\ 0, & \text{otherwise,} \end{cases}$$

$$I_3 = \begin{cases} 1, & \text{if } z_4 \text{ is in category 1} \\ 0, & \text{otherwise,} \end{cases} \qquad I_4 = \begin{cases} 1, & \text{if } z_4 \text{ is in category 2} \\ 0, & \text{otherwise.} \end{cases}$$

Indicator variables I_1 and I_2 partitions the binomial noise factor z_3. Similarly, Indicator variables I_3 and I_4 partitions second binomial noise factor z_4. Note that the four indicator variables, I_1, I_2, I_3 and I_4, are Bernoulli trials with a success probability $\frac{1}{3}$ for each. This model can be considered as a special case of the super model of (3.1.1). In matrix notation model (4.5.1) can be written as

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\underline{x} + \underline{z}^{*\prime}\underline{\gamma}^* + \underline{x}'\Delta\underline{z}^* + \varepsilon, \tag{4.5.2}$$

where $\underline{x}' = \begin{bmatrix} x_1 & x_2 \end{bmatrix}$, $\underline{z}^{*\prime} = \begin{bmatrix} z_1 & z_2 & I_1 & I_2 & I_3 & I_4 \end{bmatrix}$,

with parameters $\underline{\beta}' = \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix}$, $B = \begin{bmatrix} \beta_{11} & \frac{1}{2}\beta_{12} \\ \frac{1}{2}\beta_{12} & \beta_{22} \end{bmatrix}$,

$\underline{\gamma}^{*\prime} = \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_1^* & \gamma_2^* & \gamma_3^* & \gamma_4^* \end{bmatrix}$, and $\Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{11}^* & \delta_{12}^* & \delta_{13}^* & \delta_{14}^* \\ \delta_{21} & \delta_{22} & \delta_{21}^* & \delta_{22}^* & \delta_{23}^* & \delta_{24}^* \end{bmatrix}$.

In the above r_x, r_z and r_I, respectively, are the number of control factors, number of the continuous noise factors and number of the indicator variables. Thus $\underline{\beta}$ is $r_x \times 1$, $\underline{\gamma}^*$ is $(r_z+r_I) \times 1$, B is $r_x \times r_x$, and Δ is $r_x \times (r_z+r_I)$, in general.

The mean model can be written as

$$E_{\underline{z}}(y) = [\beta_0 + E(\underline{z}^*)'\underline{\gamma}^*] + \underline{x}'[\underline{\beta} + \Delta E(\underline{z}^*)] + \underline{x}'B\underline{x}, \tag{4.5.3}$$

where

$$E(\underline{z}^*) = [0 \;\; 0 \;\; E(I_1) \;\; E(I_2) \;\; E(I_3) \;\; E(I_4)]' = [0 \;\; 0 \;\; \tfrac{1}{3} \;\; \tfrac{1}{3} \;\; \tfrac{1}{3} \;\; \tfrac{1}{3}]'.$$

Note that

$$\mathrm{Var}(\underline{z}^*) = \mathrm{V} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 \\ & & \frac{2}{9} & -\frac{1}{9} & 0 & 0 \\ & & & \frac{2}{9} & 0 & 0 \\ & & & & \frac{2}{9} & -\frac{1}{9} \\ & (\mathrm{sym}) & & & & \frac{2}{9} \end{bmatrix} \qquad (4.5.4)$$

and V is of dimension $(\mathrm{r_Z}+\mathrm{r_I}) \times (\mathrm{r_Z}+\mathrm{r_I})$ in general assuming that $\sigma^2_{\mathrm{Z}_1} = \sigma^2_{\mathrm{Z}_2} = 1$. The diagonal elements of the variance-covariance matrix in (4.5.4) are variances of noise factors, $\mathrm{p}_i \times (1 - \mathrm{p}_i)$ for indicator variables, and the off-diagonal elements of each "block" submatrix are covariances of a Bernoulli trial, $\text{-}\mathrm{p}_i\mathrm{p}_j$, where p_i and p_j are success probabilities of indicators I_i and I_j $(i \neq j)$, respectively, and each submatrix corresponds to the indicator variable generated by partitioning of noise factors. In general, we write

$$\mathrm{E}(\underline{z}^*) = [\,0 \;\; \cdots \;\; 0 \;\; \mathrm{p}_1 \;\; \mathrm{p}_2 \;\; \cdots \;\; \mathrm{p}_{(\mathrm{r_z}+\mathrm{r_I})}\,]'$$

$$\mathrm{Var}(\underline{z}^*) = \mathrm{V} = \begin{bmatrix} 1 & 0 & \cdots & 0 & & & & \\ & \cdots & \cdots & \cdots & & \underline{0} & & \\ & & \cdots & \cdots & & & & \\ & & & 1_{\mathrm{r_z}} & & & & \\ & & & & \mathrm{V}_1 & & & \\ & & & & & \mathrm{V}_2 & & \\ & & (\mathrm{sym}) & & & & \cdots & \\ & & & & & & & \mathrm{V}_{(\mathrm{r_z}+\mathrm{r_I})} \end{bmatrix}, \qquad (4.5.5)$$

where V_i's are block matrices defined as before (see 4.1.5). Now let us turn our attention to the process variance. In order to evaluate the process variance consider the model of Equation (4.5.2).

$$y = \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\underline{x} + \underline{z}^{*\prime}\underline{\gamma}^* + \underline{x}'\Delta\underline{z}^* + \varepsilon,$$

where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. From the above model the process variance is given by

$$\mathrm{Var}_{\underline{z}}(y \mid \underline{x}) = (\underline{\gamma}^{*\prime} + \underline{x}'\Delta)V(\underline{\gamma}^* + \Delta'\underline{x}) + \sigma_\varepsilon^2, \tag{4.5.6}$$

where the variance-covariance matrix of $\underline{z}^*$'s, V, is given in (4.5.5). Rewrite Equation (4.5.6) as

$$\mathrm{Var}_{\underline{z}}(y \mid \underline{x}) = \underline{l}'V\underline{l} + \sigma_\varepsilon^2, \tag{4.5.7}$$

$$\text{where } \underline{l} = \frac{\partial y}{\partial \underline{z}^*} = (\underline{\gamma}^* + \Delta'\underline{x}).$$

Vector $\underline{l}$ consists of linear polynomial functions in the control factors as before.

As a biased estimator of the process variance we see

$$\widehat{\mathrm{Var}}_{\underline{z}}(y \mid \underline{x}) = \underline{\hat{l}}'V\underline{\hat{l}} + s^2, \tag{4.5.8}$$

where s^2 is an estimated error mean square. This estimator is biased since

$$E\{\widehat{\mathrm{Var}}_{\underline{z}}(y \mid \underline{x})\} = \underline{l}'V\underline{l} + \sigma_\varepsilon^2\,\mathrm{tr}(CV) + \sigma_\varepsilon^2. \tag{4.5.9}$$

From Equation (4.5.9) it is obvious that an unbiased estimator of the process variance can now be written as

$$\widehat{\mathrm{Var}}_{\underline{z}}(y \mid \underline{x}) = \underline{\hat{l}}'V\underline{\hat{l}} + \{1 - \mathrm{tr}\,(CV)\}\,s^2. \tag{4.5.10}$$

Constructed in the same way as in section 4.2, tr(CV) can be rewritten as

$$\mathrm{tr}(CV) = (1, \underline{x}')\begin{bmatrix} c^* & \underline{0} \\ \underline{0} & \underline{M}^* \end{bmatrix}\binom{1}{\underline{x}}, \tag{4.5.11}$$

where $$c^* = \sum_{\substack{i,\,j \\ i \neq j}}^{r_{z^*}} \left\{p_i(1-p_i)\mathrm{Var}\,(\hat{\gamma}_i^*) + \mathrm{Cov}(\hat{\gamma}_i^*,\,\hat{\gamma}_j^*)\mathrm{Cov}(z_i^*,\,z_j^*)\right\}$$

and

$$\underline{M}^* = \begin{bmatrix} M_1 & 0 & \cdots & \cdots & 0 \\ & \cdots & \cdots & \cdots & \cdots \\ & & M_k & \cdots & \cdots \\ & & & \cdots & \cdots \\ & (\mathrm{sym}) & & & M_{r_x} \end{bmatrix}$$

with

$$M_k = \sum_{i=1}^{r_{z^*}} \left\{p_i(1-p_i)\mathrm{Var}\,(\hat{\delta}_{ki})\right\} + \sum_{\substack{i,\,j \\ i \neq j}}^{r_{z^*}} \mathrm{Cov}(\hat{\delta}_{ki},\,\hat{\delta}_{k,i+1})\mathrm{Cov}(z_i^*,\,z_j^*). \qquad (4.5.12)$$

For details see section 4.2. To compare the error mean square between the biased estimator (4.5.8) and the unbiased estimator (4.5.10) we have

$$\mathrm{MSE}\,(\hat{\underline{l}}'\mathrm{V}\hat{\underline{l}} + s^2) - \mathrm{MSE}\,[\hat{\underline{l}}'\mathrm{V}\hat{\underline{l}} + \{1 - \mathrm{tr}\,(\mathrm{CV})\}\,s^2]$$

$$= \frac{\sigma_\varepsilon^4\,\mathrm{tr(CV)}}{\mathrm{dfE}}\,[(\mathrm{dfE}-2)\{\mathrm{tr(CV)}\} + 4]. \qquad (4.5.13)$$

When tr(CV) is positive and $\mathrm{dfE} > 1$ the unbiased estimator of the process variance will have smaller mean square error than the biased estimator.

The general stationary region analysis by the unbiased estimator finds $\hat{\underline{x}}_0$ as the location of the estimated minimum process variance given by

$$\hat{\underline{x}}_0 = -(\hat{\Delta}\mathrm{V}\hat{\Delta}' - s^2\underline{M}^*)^{-1}\hat{\Delta}\mathrm{V}\hat{\underline{\gamma}}^*. \qquad (4.5.14)$$

For computational details see section 4.3.

Ridge analysis by the unbiased estimator of $\underline{l}'\mathrm{V}\underline{l}$, $\hat{\underline{l}}'\mathrm{V}\hat{\underline{l}} - s^2\mathrm{tr(CV)}$, finds

$$(\hat{\Delta} V \hat{\Delta}' - s^2 \underline{M}^* - \mu I_{r_x})\,\underline{x} = -\hat{\Delta} V \hat{\underline{\gamma}}^* \tag{4.5.15}$$

as the location of the minimum process variance under the constraint of

$$\sum_{i=1}^{r_x} x_i^2 = R^2. \tag{4.5.16}$$

Finding the location of the minimum process variance under a general constraint set, construction of $(1-\alpha)100\%$ confidence region for the location of the minimum process variance for a fixed μ, and one-sided tolerance intervals on future observation can also be argued in a similar fashion as described in section 4.3.

Chapter 5. Examples

We shall now show examples to illustrate the methodology we have developed thus far. For the data of Example 1.1 we first do the analysis through the use of some of the RSM techniques obtained in the previous sections. Included are the estimated super model, the estimated mean response model, the estimated process variance model, the estimated location of the minimum process variance, the estimated one-sided tolerance intervals. Comments about the analysis are also included.

Example 5.1 (Confidence Region on the Location of the Minimum Process Variance and One-Sided Tolerance Intervals)

Refer to the color-TV image data of Table 1.1. The model to be used is

$$y = \beta_0+\beta_1x_1+\beta_2x_2+\beta_{12}x_1x_2+\beta_{11}x_1^2 + \beta_{22}x_2^2 + \gamma_1z_1+\gamma_2z_2+\delta_{11}x_1z_1 +\delta_{12}x_1z_2+\delta_{21}x_2z_1+\delta_{22}x_2z_2+\varepsilon.$$

The process variance of y given x_1 and x_2 becomes

$$\begin{aligned}\mathrm{Var}_{\underline{z}}(y|\underline{x}) &= \{\gamma_1^2 + \gamma_2^2 + \delta_{11}^2x_1^2 + \delta_{12}^2x_1^2 + \delta_{21}^2x_2^2 + \delta_{22}^2x_2^2 + 2\gamma_1(\delta_{11}x_1 + \delta_{21}x_2) \\ &\quad + 2\gamma_2(\delta_{12}x_1 + \delta_{22}x_2) + 2(\delta_{11}\delta_{21} + \delta_{12}\delta_{22})x_1x_2\}\, \sigma_z^2 + \sigma_\varepsilon^2 \\ &= \left(l_1^2 + l_2^2\right)\sigma_z^2 + \sigma_\varepsilon^2,\end{aligned}$$

where $l_1 = \left(\frac{\partial y}{\partial z_1}\right) = \gamma_1 + \delta_{11}x_1 + \delta_{21}x_2$ and $l_2 = \left(\frac{\partial y}{\partial z_2}\right) = \gamma_2 + \delta_{12}x_1 + \delta_{22}x_2$.

An analysis of variance was performed and the output appears below.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	11	2503.70542	227.60958	413.132	0.0001
Error	24	13.22248	0.55094		
C Total	35	2516.92791			

Root MSE	0.74225	R-square	0.9947
Dep Mean	30.59240	Adj R-sq	0.9923
C.V.	2.42626		

Variable	DF	Parameter Estimate	S. E.	t for Parameter=0	Prob > |t|
INTERCEP	1	33.388881	0.27662068	120.703	0.0001
x1	1	-4.175204	0.15151139	-27.557	0.0001
x2	1	3.748096	0.15151139	24.738	0.0001
x12	1	3.348494	0.18556279	18.045	0.0001
x11	1	-2.327671	0.26242542	-8.870	0.0001
x22	1	-1.867046	0.26242542	-7.115	0.0001
z1	1	-4.075519	0.12370853	-32.945	0.0001
z2	1	2.985436	0.12370853	24.133	0.0001
x1z1	1	-2.324121	0.15151139	-15.340	0.0001
x1z2	1	1.932154	0.15151139	12.753	0.0001
x2z1	1	3.268287	0.15151139	21.571	0.0001
x2z2	1	-2.072946	0.15151139	-13.682	0.0001

□

All the model terms are highly significant, explaining almost all of the variation in y. Plots of the residuals against predicted values, normal probability plot of the residuals show no apparent violations in model assumptions.

$$\text{Denote } \widehat{l}_1 = \left(\frac{\widehat{\partial y}}{\partial z_1}\right) = \widehat{\gamma}_1 + \widehat{\delta}_{11}x_1 + \widehat{\delta}_{21}x_2 \text{ and } \widehat{l}_2 = \left(\frac{\widehat{\partial y}}{\partial z_2}\right) = \widehat{\gamma}_2 + \widehat{\delta}_{12}x_1 + \widehat{\delta}_{22}x_2.$$

To build a confidence region on the estimated minimum process variance point $(x_{10}, x_{20}) = (-0.874336, 0.625237)$, where both $\widehat{l}_1$ and $\widehat{l}_2$ become zero,

(i) draw two lines: $\widehat{l}_1 = 0$ and $\widehat{l}_2 = 0$ that intersect. (Figure 5.1).

(ii) Let $\underline{x}_0' = (x_{10}, x_{20})$ be the point where two lines $\widehat{l}_1 = 0$ and $\widehat{l}_2 = 0$ meet each other. Then, $\widehat{l}_1(\underline{x}_0') = \widehat{l}_2(\underline{x}_0') = 0$. Note that $\left\{ \widehat{l}_1(\underline{x}_0'),\ \widehat{l}_2(\underline{x}_0')\right\}' \sim \underline{N}(\underline{0}, V)$, where the estimated variance-covariance matrix $\widehat{V}$ is

$$\widehat{V} = \widehat{Cov}\left\{\widehat{l}_1(\underline{x}_0'),\ \widehat{l}_2(\underline{x}_0')\right\}' =$$

$$\begin{bmatrix} \widehat{Var}\,(\widehat{\gamma}_1+\widehat{\delta}_{11}x_{10}+\widehat{\delta}_{21}x_{20}) & \widehat{cov}\,(\widehat{\gamma}_1+\widehat{\delta}_{11}x_{10}+\widehat{\delta}_{21}x_{20},\ \widehat{\gamma}_2+\widehat{\delta}_{12}x_{10}+\widehat{\delta}_{22}x_{20}) \\ \text{(symmetrical)} & \widehat{Var}\,(\widehat{\gamma}_2+\widehat{\delta}_{12}x_{10}+\widehat{\delta}_{22}x_{20}) \end{bmatrix}$$

$$= \begin{bmatrix} \widehat{Var}\,(\widehat{\gamma}_1+\widehat{\delta}_{11}x_{10}+\widehat{\delta}_{21}x_{20}) & 0 \\ 0 & \widehat{Var}\,(\widehat{\gamma}_1+\widehat{\delta}_{12}x_{10}+\widehat{\delta}_{22}x_{20}) \end{bmatrix}$$

$$= 0.0229557 \times \begin{bmatrix} (0.666667+x_{10}^2+x_{20}^2) & 0 \\ 0 & (0.666667+x_{10}^2+x_{20}^2) \end{bmatrix}$$

The 95% confidence region can be found from (3.1.11)

$$P\left\{\frac{(\widehat{\underline{l}}_{\underline{x}_0'} - \underline{0})' \left[\widehat{Var}\,(\widehat{\underline{l}}_{\underline{x}_0'})\right]^{-1} (\widehat{\underline{l}}_{\underline{x}_0'} - \underline{0})}{2} \ \le\ F_{2,\ dfE;\ 0.95}\right\} = 0.95.$$

This leads to

$$\frac{(\widehat{\gamma}_1 + \widehat{\delta}_{11}x_{10} + \widehat{\delta}_{21}x_{20})^2 + (\widehat{\gamma}_2 + \widehat{\delta}_{12}x_{10} + \widehat{\delta}_{22}x_{20})^2}{(0.0153038 + 0.0229557x_{10}^2 + 0.0229557x_{20}^2)} \le 6.8056.$$

Figure 5.1 shows two solid lines $\widehat{l}_1 = 0$, $\widehat{l}_2 = 0$ and a 95% confidence region (shaded area) on the location of the minimum process variance (-0.8743, 0.6252). Clearly it is of interest to maximize the response and obtain a consistent response. The estimated mean response surface is

$$E_{\underline{z}}\widehat{(y)} = 33.388881 - 4.175204x_1 + 3.748096x_2 + 3.348494x_1x_2 - 2.327671x_1^2 - 1.867046x_2^2.$$

Figure 5.2 shows the mean response contours with a stationary point for the mean responses at (-0.49, 0.56). Figure 5.3 shows the mean response contours and process standard deviation response contours superimposed. Since we want to maximize the response, one-sided tolerance interval of definition 3.7.2 will be appropriate here. Figure 5.4 shows the contour plot of the one-sided tolerance intervals ($p = 0.95$ and $\gamma = 0.05$). The computation was done by using the IMSL subroutines for multiple integration as described in section 3.7. In particular, five locations were selected to compare the intervals. Those selected five points are: (1) minimum process variance point, (2) and (3) two boundary points of the minimum variance confidence region, (4) stationary point for the mean responses, and (5) the point selected by Taguchi's SN-ratio. The results are summarized below.

Selected Locations		Estimated Mean	Tolerance Interval
(1)	(-0.8743, 0.6252)	35.0441	32.1724
(2)	(-0.24, 1.0)	35.3343	31.9464
(3)	(-1.0, 0.4)	35.0975	31.7985
(4)	(-0.493, 0.562)	35.4705	30.8947
(5)	(0, 1)	35.2699	30.9046

In calculating one-sided tolerance intervals, point estimates for $\underline{l}'\underline{l}$ and σ_ε^2 were used for $\frac{l'l}{\sigma_\varepsilon^2}$ (see section 3.7). Note that it is the minimum process variance point (1) that guarantees the highest tolerance interval. Both locations (2) and (3) are boundary points of the 95% confidence region on the location of the minimum process variance, and they show higher tolerance interval values than both the stationary point for the mean responses and Taguchi's location even though their estimated means are lower. Location (2) shows higher tolerance interval than location (3). It appears that this is due to the fact that the estimated mean at (2) is higher than that of (3). Tolerance interval values drop significantly when we move away from the minimum process variance point. The estimated location of the minimum process variance, therefore, the estimated location of the largest one-sided tolerance interval, (-0.8743, 0.6252), is quite different from (0, 1), the location where Taguchi's SN-ratio was maximized (refer to Example 1.1).

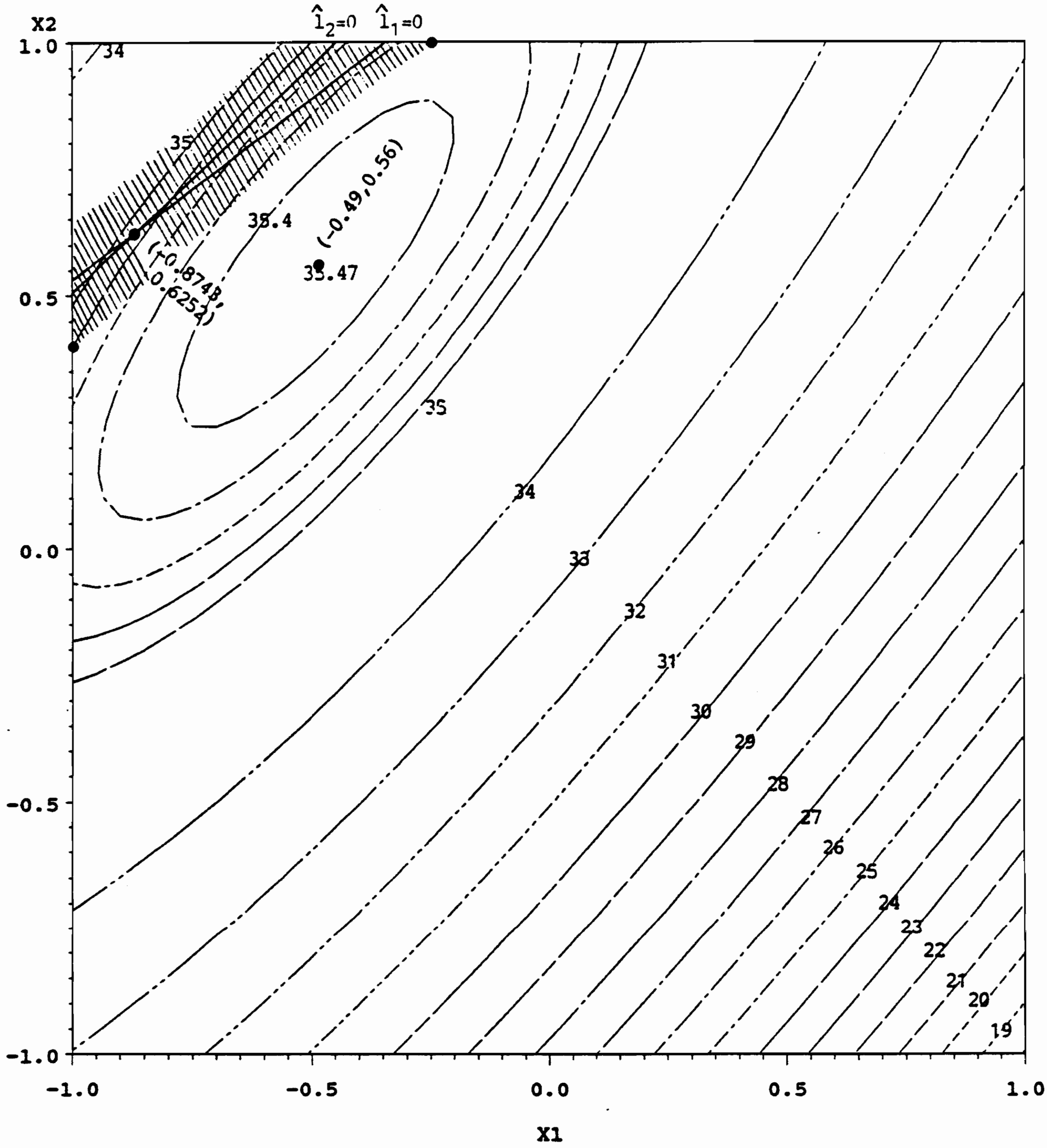


Figure 5.1 Estimated Minimum Process Variance Location and Confidence Region

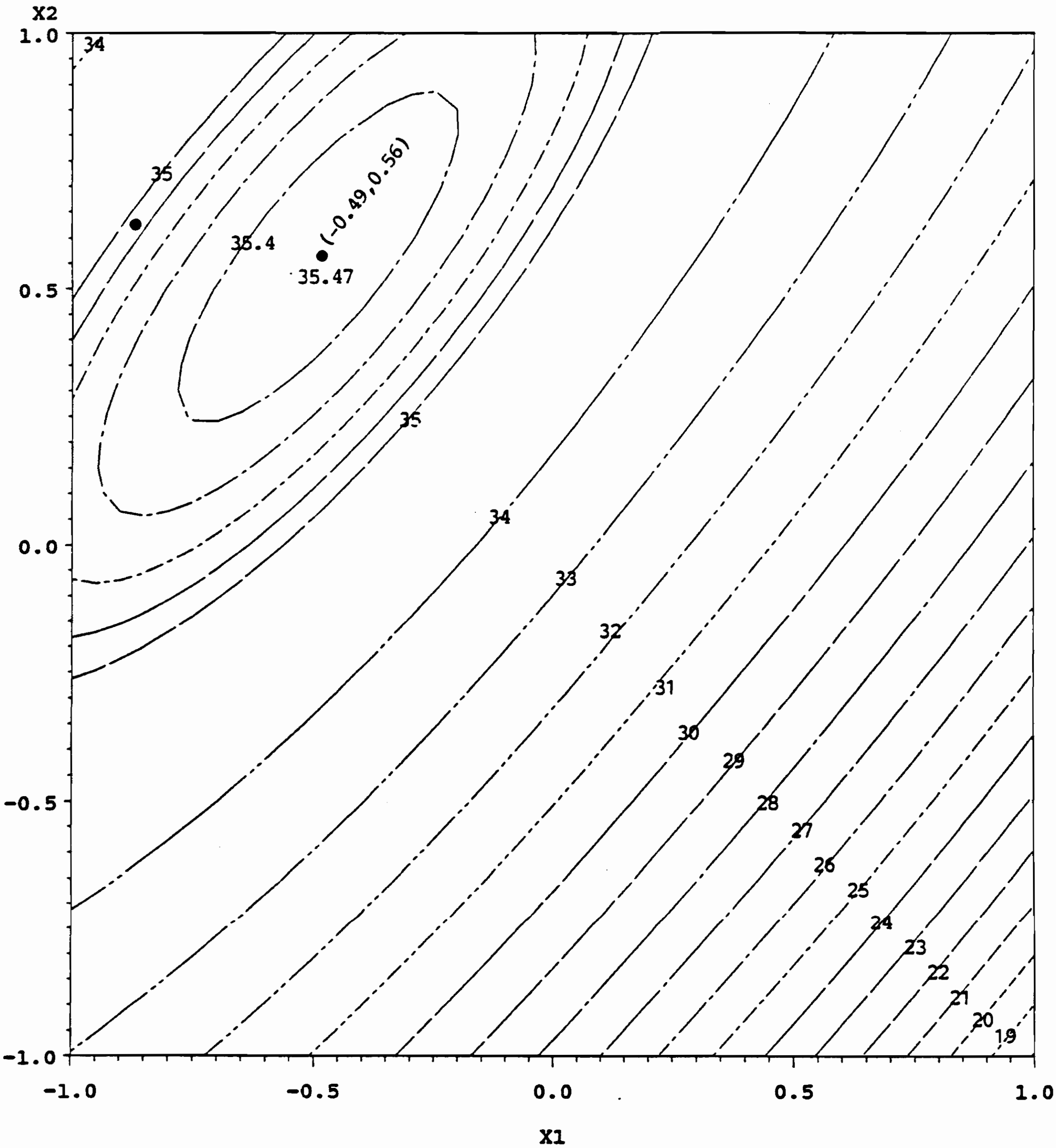


Figure 5.2 Estimated Mean Responses and Stationary Point

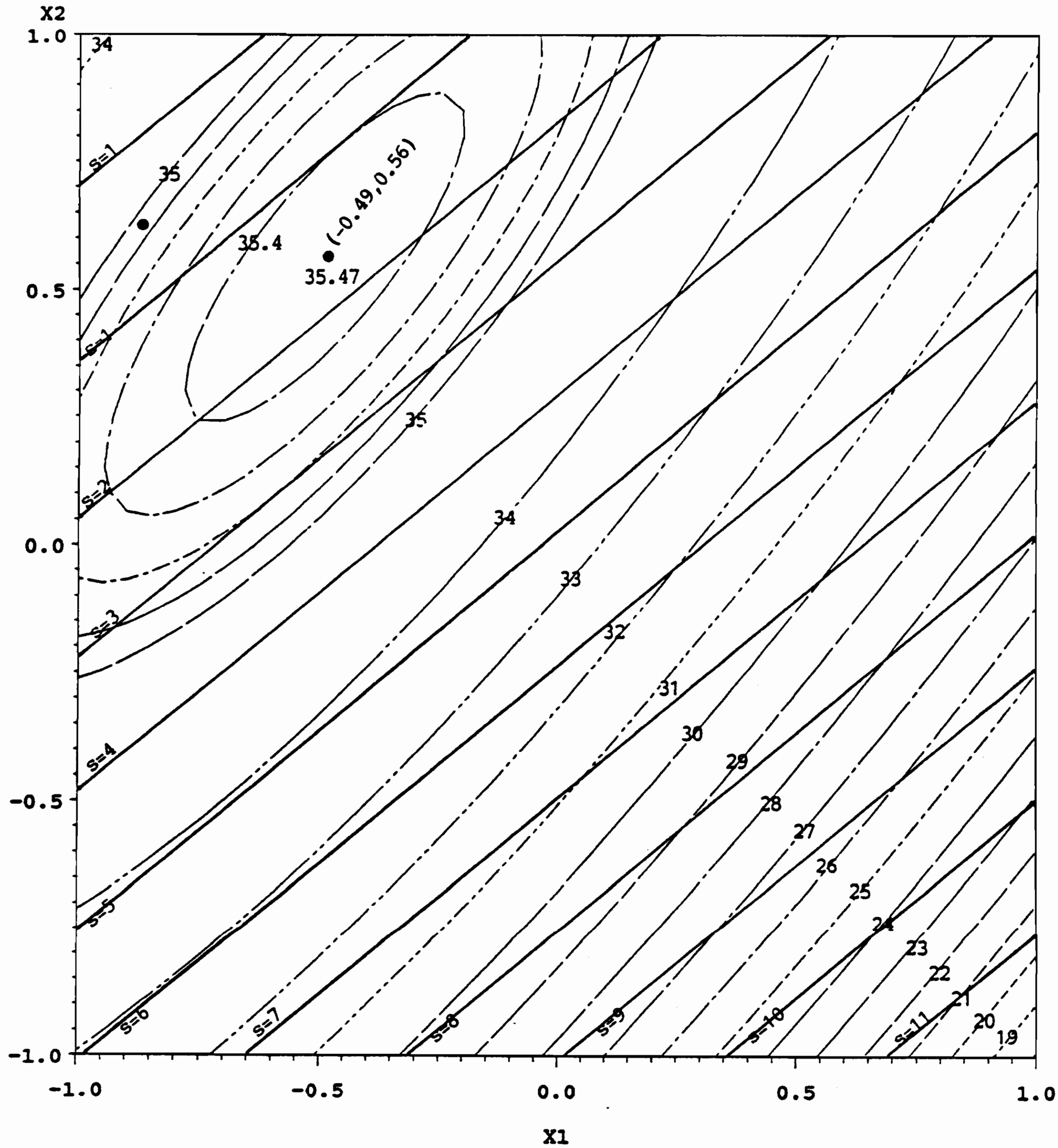


Figure 5.3 Estimated Mean Responses and Process Standard Deviation

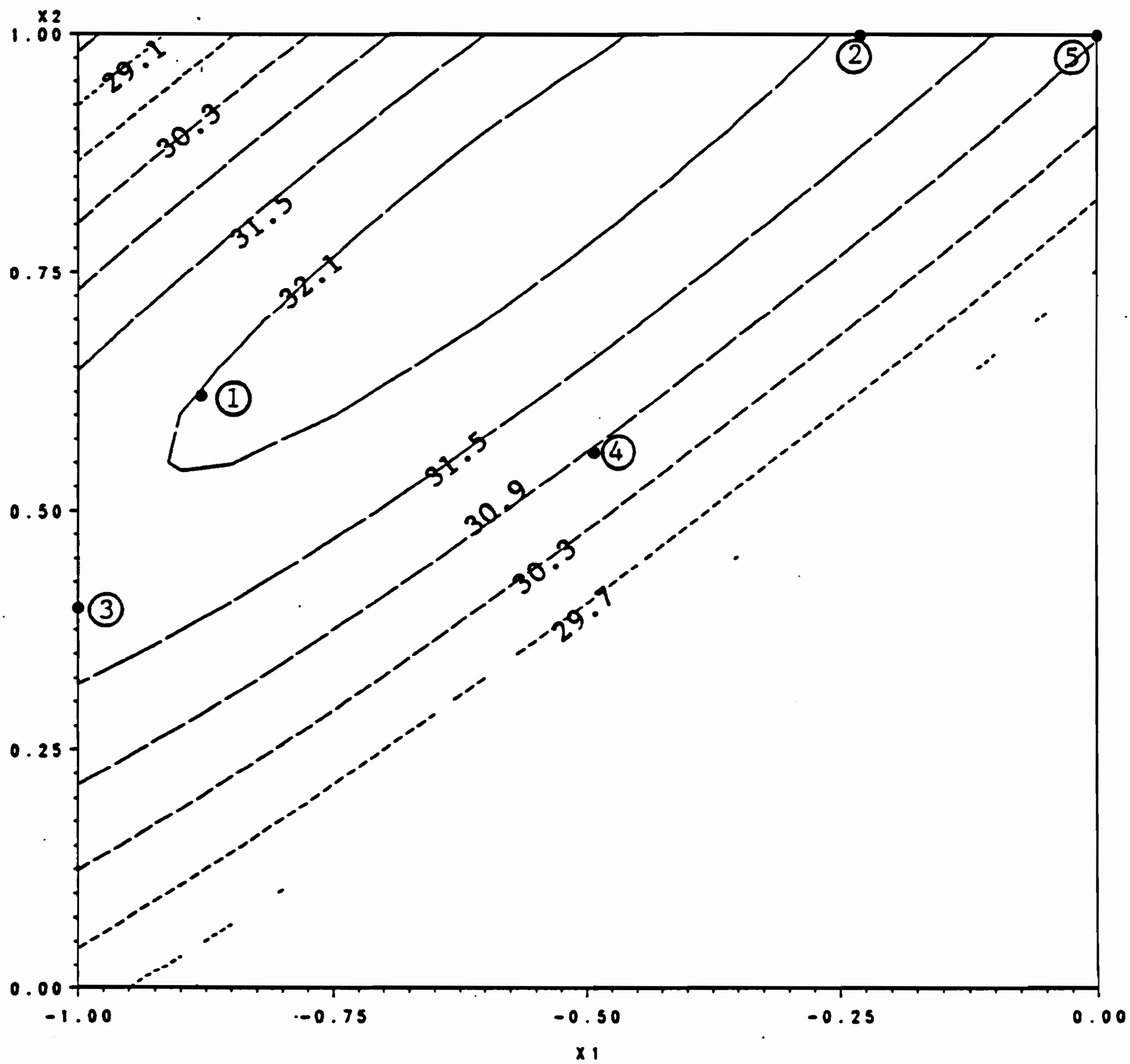


Figure 5.4 One-Sided Tolerance Limit Values and Selected Locations

Example 5.2 (General Stationary Region Analysis)

We now illustrate the stationary region analysis developed in section 3.3 with a numerical example. Suppose there are two control factors, x_1 and x_2, and three noise factors, z_1, z_2 and z_3. All five factors are quantitative and are varied in a 23-run central composite design. For more justification of the design refer to Example 2.2. The design consists of 16-run fractional factorial part ($\frac{1}{2}$ of 2^5 factorial with resolution V), four axial runs and three center runs. Assume that it is of importance to find operating conditions that produce good reproducibility. The data are shown below.

OBS	x1	x2	z1	z2	z3	y
1	1	-1	-1	-1	-1	25.5521
2	-1	1	-1	-1	-1	20.3121
3	-1	-1	1	-1	-1	15.5759
4	-1	-1	-1	1	-1	20.0077
5	-1	-1	-1	-1	1	3.9085
6	1	1	1	-1	-1	22.3759
7	1	1	-1	1	-1	10.3277
8	1	1	-1	-1	1	18.5485
9	1	-1	1	1	-1	41.8315
10	1	-1	1	-1	1	26.3723
11	1	-1	-1	1	1	13.2441
12	-1	1	1	1	-1	12.6715
13	-1	1	1	-1	1	20.2523
14	-1	1	-1	1	1	23.7641
15	-1	-1	1	1	1	19.2679
16	1	1	1	1	1	11.8279
17	-2	0	0	0	0	18.5517
18	2	0	0	0	0	28.6356
19	0	-2	0	0	0	30.8721
20	0	2	0	0	0	22.9501
21	0	0	0	0	0	12.1989
22	0	0	0	0	0	11.2298
23	0	0	0	0	0	12.2307

□

A model containing all main effects and two-factor interactions was fitted to the data:

$$y = \beta_0+\beta_1 x_1+\beta_2 x_2+\beta_{12} x_1 x_2+\beta_{11} x_1^2 + \beta_{22} x_2^2 + \gamma_1 z_1+\gamma_2 z_2+\gamma_3 z_3$$
$$+\delta_{11} x_1 z_1+\delta_{12} x_1 z_2+\delta_{13} x_1 z_3+\delta_{21} x_2 z_1+\delta_{22} x_2 z_2+\delta_{23} x_2 z_3+\varepsilon.$$

The process variance of y given x_1 and x_2 becomes

$$\mathrm{Var}_{\underline{z}}(y|\underline{x}) = \underline{l}'\underline{l} + \sigma_\varepsilon^2,$$

$$\text{where } \underline{l} = \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} = \underline{\gamma} + \Delta'\underline{x} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} + \begin{bmatrix} \delta_{11} & \delta_{21} \\ \delta_{12} & \delta_{22} \\ \delta_{13} & \delta_{23} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Here, it is assumed that z_1, z_2 and z_3 are uncorrelated and that $\sigma_{z1}^2=\sigma_{z2}^2=\sigma_{z3}^2=1$. Parameter estimates and t statistics are shown below.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	14	1492.93719	106.63837	196.932	0.0001
Error	8	4.33198	0.54150		
C Total	22	1497.26917			

Root MSE	0.73587	R-square	0.9971
Dep Mean	19.23952	Adj R-sq	0.9920
C.V.	3.82476		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	12.322930	0.32908891	37.446	0.0001
x1	1	2.270325	0.15020785	15.115	0.0001
x2	1	-1.730167	0.15020785	-11.518	0.0001
x12	1	-3.885000	0.18396629	-21.118	0.0001
x11	1	2.899517	0.16709569	17.352	0.0001
x22	1	3.728879	0.16709569	22.316	0.0001
z1	1	2.156900	0.18396629	11.724	0.0001
z2	1	0.002800	0.18396629	0.015	0.9882

z3	1	-1.966800	0.18396629	-10.691	0.0001
x1z1	1	2.185000	0.18396629	11.877	0.0001
x1z2	1	-1.955000	0.18396629	-10.627	0.0001
x1z3	1	-1.795000	0.18396629	-9.757	0.0001
x2z1	1	-2.885000	0.18396629	-15.682	0.0001
x2z2	1	-2.865000	0.18396629	-15.574	0.0001
x2z3	1	3.055000	0.18396629	16.606	0.0001 □

To find the location of the estimated minimum process variance we see that

$$\hat{\underline{l}} = \begin{bmatrix} \hat{l}_1 \\ \hat{l}_2 \\ \hat{l}_3 \end{bmatrix} = \begin{bmatrix} 2.1569 + 2.185x_1 - 2.885x_2 \\ 0.0028 - 1.955x_1 - 2.865x_2 \\ -1.9668 - 1.795x_1 + 3.055x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

has no solution for x_1 and x_2. Therefore, we turn to the stationary region analysis approach developed in section 3.3. Let us first find the location of the minimum process variance using a biased estimator for the process variance (see section 3.3). Recall the solution of (3.3.2.3),

$$\hat{\underline{x}}_0 = -(\hat{\Delta}\hat{\Delta}')^{-1}\hat{\Delta}\hat{\underline{\gamma}}.$$

Using the parameter estimates shown above we get an estimated minimum process variance location as

$$\hat{\underline{x}}_0 = \begin{bmatrix} -0.513641 \\ 0.350352 \end{bmatrix}.$$

Let us now find the location of the minimum process variance using the unbiased estimator for the process variance (see section 3.3). Recall the solution of (3.3.1.4),

$$\hat{\underline{x}}_0 = -(\hat{\Delta}\hat{\Delta}' - s^2\underline{M})^{-1}\hat{\Delta}\hat{\underline{\gamma}},$$

$$\text{where } \underline{M} = \frac{3}{16}I_2 \text{ and } s^2\underline{M} = \begin{bmatrix} 0.10153125 & 0 \\ 0 & 0.10153125 \end{bmatrix}.$$

Using the parameter estimates shown above we get an estimated minimum process variance location as

$$\hat{\underline{x}}_0 = \begin{bmatrix} -0.517903 \\ 0.350710 \end{bmatrix}.$$

This is indeed the location of the minimum process variance since $\hat{\Delta}\hat{\Delta}' - s^2\underline{M}$ is a positive definite matrix with

$$\hat{\Delta}\hat{\Delta}' - s^2\underline{M} = \begin{bmatrix} 11.716744 & -6.186375 \\ (\text{sym}) & 25.762944 \end{bmatrix}.$$

Note that the two locations, one by using a biased estimator and the other by using the unbiased estimator, are not much different. An extraordinarily small s^2 (i. e., very good fit of the model to the data) has caused such a result.

Example 5.3 (Ridge Analysis)

We first illustrate the ridge analysis developed in section 3.4. Suppose there are two control factors, x_1 and x_2, and three noise factors, z_1, z_2 and z_3. All five factors are quantitative and are varied in a 23-run central composite design. For justification of the design see Example 2.2. The design consists of 16-run fractional factorial part ($\frac{1}{2}$ of 2^5 factorial with resolution V), four axial runs and three center runs. Assume that we want to find operating conditions that minimize the process variance while restricting ourselves on a sphere of radius $\sqrt{2}$. That is, the constraint is given as

$$\sum_{i=1}^{r_x} x_i^2 = 2.$$

A model containing all main effects and two-factor interactions was fitted to the data:

$$y = \beta_0+\beta_1 x_1+\beta_2 x_2+\beta_{12}x_1x_2+\beta_{11}x_1^2 + \beta_{22}x_2^2 + \gamma_1 z_1+\gamma_2 z_2+\gamma_3 z_3$$

$$+\delta_{11}x_1z_1+\delta_{12}x_1z_2+\delta_{13}x_1z_3+\delta_{21}x_2z_1+\delta_{22}x_2z_2+\delta_{23}x_2z_3+\varepsilon.$$

The process variance of y given x_1 and x_2 becomes

$$\mathrm{Var}_{\underline{z}}(y\,|\,\underline{x}) = \underline{l}'\underline{l} + \sigma_\varepsilon^2,$$

$$\text{where } \underline{l} = \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} = \underline{\gamma} + \Delta'\underline{x} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} + \begin{bmatrix} \delta_{11} & \delta_{21} \\ \delta_{12} & \delta_{22} \\ \delta_{13} & \delta_{23} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

It is assumed that z_1, z_2 and z_3 are uncorrelated and that $\sigma_{z1}^2=\sigma_{z2}^2=\sigma_{z3}^2=1$. The data, parameter estimates and t statistics are shown below.

OBS	x1	x2	z1	z2	z3	y
1	1	-1	-1	-1	-1	30.0250
2	-1	1	-1	-1	-1	30.0007
3	-1	-1	1	-1	-1	49.8009
4	-1	-1	-1	1	-1	43.4717
5	-1	-1	-1	-1	1	44.1905
6	1	1	1	-1	-1	31.3911
7	1	1	-1	1	-1	16.0333
8	1	1	-1	-1	1	35.3823
9	1	-1	1	1	-1	30.3383
10	1	-1	1	-1	1	36.3417
11	1	-1	-1	1	1	36.1355
12	-1	1	1	1	-1	30.1289
13	-1	1	1	-1	1	41.3179
14	-1	1	-1	1	1	22.7125
15	-1	-1	1	1	1	43.2415
16	1	1	1	1	1	39.1733
17	-2	0	0	0	0	46.1502

18	2	0	0	0	0	36.0689
19	0	-2	0	0	0	47.3903
20	0	2	0	0	0	31.4659
21	0	0	0	0	0	30.8109
22	0	0	0	0	0	30.7499
23	0	0	0	0	0	30.9655

□

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	14	1455.75931	103.98281	113.022	0.0001
Error	8	7.36020	0.92003		
C Total	22	1463.11951			

Root MSE	0.95918	R-square	0.9950
Dep Mean	35.36029	Adj R-sq	0.9862
C.V.	2.71259		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	30.381985	0.42895808	70.827	0.0001
x1	1	-2.925279	0.19579168	-14.941	0.0001
x2	1	-4.135579	0.19579168	-21.122	0.0001
x12	1	2.855256	0.23979486	11.907	0.0001
x11	1	2.595620	0.21780450	11.917	0.0001
x22	1	2.175257	0.21780450	9.987	0.0001
z1	1	2.736381	0.23979486	11.411	0.0001
z2	1	-2.325944	0.23979486	-9.700	0.0001
z3	1	2.331581	0.23979486	9.723	0.0001
x1z1	1	-0.277844	0.23979486	-1.159	0.2800
x1z2	1	0.893481	0.23979486	3.726	0.0058
x1z3	1	2.574056	0.23979486	10.734	0.0001
x2z1	1	1.998919	0.23979486	8.336	0.0001
x2z2	1	-1.429556	0.23979486	-5.962	0.0003
x2z3	1	1.547419	0.23979486	6.453	0.0002

□

Note first that the minimum process variance occurs at (0.0097, -1.5038) which lies outside the experimental region. We need to look for a location of the estimated minimum process variance on a sphere of radius $\sqrt{2}$. We first use a biased estimator for the process variance (see section 3.4). Let us minimize

$$\hat{\underline{l}}'\hat{\underline{l}}$$

under the constraint (a sphere of radius $\sqrt{2}$)

$$\sum_{i=1}^{2} x_i^2 = 2.$$

From the printout we see that

$$\hat{\underline{l}} = \begin{bmatrix} \hat{l}_1 \\ \hat{l}_2 \\ \hat{l}_3 \end{bmatrix} = \begin{bmatrix} 2.736381 - 0.277844x_1 + 1.998919x_2 \\ -2.325944 + 0.893481x_1 - 1.429556x_2 \\ 2.331581 + 2.574056x_1 + 1.547419x_2 \end{bmatrix}.$$

Recall Equation (3.4.2.4)

$$(\hat{\Delta}\hat{\Delta}' - \mu I_2)\,\underline{x} = -\hat{\Delta}\hat{\underline{\gamma}}.$$

Eigenvalues of the matrix $\hat{\Delta}\hat{\Delta}'$ turn out to be 10.167984 and 5.7670985. To find the location of the minimum process variance we replace μ with values smaller than the smallest eigenvalue (5.7670985). We find that when $\mu = -0.313$

$$\hat{\underline{x}}_0 = \begin{bmatrix} -0.015622 \\ -1.414139 \end{bmatrix}$$

is the location of the minimum process variance on a sphere of radius $\sqrt{2}$.

Let us now find the location of the minimum process variance on a sphere of radius $\sqrt{2}$ using the unbiased estimator of the process variance (see section 3.4). That is, we want to minimize

$$\hat{\underline{l}}'\hat{\underline{l}} - s^2\,\mathrm{tr}\,(C)$$

under the constraint (a sphere of radius $\sqrt{2}$)

$$\sum_{i=1}^{2} x_i^2 = 2.$$

Recall Equation (3.4.1.6)

$$(\hat{\Delta}\hat{\Delta}' - s^2\underline{M} - \mu I_{r_x})\, \underline{x} = -\hat{\Delta}\hat{\underline{\gamma}},$$

$$\text{where } \underline{M} = \frac{3}{16} I_2 \text{ and } s^2\underline{M} = \begin{bmatrix} 0.172505625 & 0 \\ 0 & 0.172505625 \end{bmatrix}.$$

Eigenvalues of the matrix $(\hat{\Delta}\hat{\Delta}' - s^2\underline{M})$ turn out to be 9.9954788 and 5.5945929. To find the location of the minimum process variance we replace μ with values smaller than the smallest eigenvalue (5.5945929). We find that when $\mu = -0.486$

$$\hat{\underline{x}}_0 = \begin{bmatrix} -0.015645 \\ -1.414053 \end{bmatrix}$$

is the location of the minimum process variance on a sphere of radius $\sqrt{2}$. This is indeed the location of the minimum process variance under the constraint since

$$\hat{\Delta}\hat{\Delta}' - s^2\underline{M} - \mu I_2 = \begin{bmatrix} 7.8147643 & 2.1504744 \\ (\text{sym}) & 8.7473075 \end{bmatrix}$$

is a positive definite matrix. Notice that the two locations, one by using a biased estimator and the other by using the unbiased estimator, are almost identical. An extraordinarily small s^2 (i. e., very good fit of the model to the data) has caused such a result.

Figure 5.5 shows a trace of the location of the estimated minimum process variance at varying radii (i. e., on concentric circles).

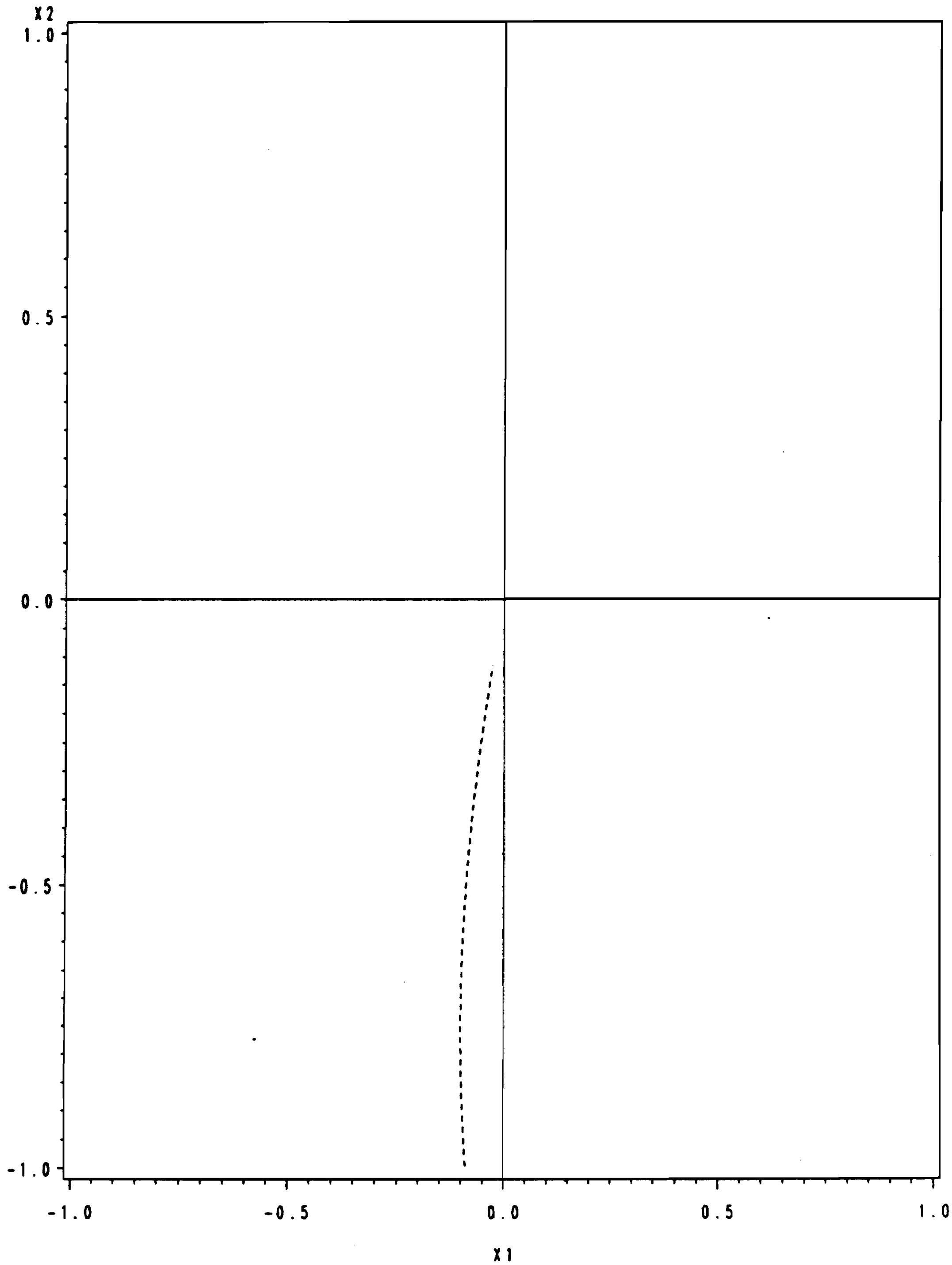


Figure 5.5 Trace of a Location of the Estimated Minimum Process Variance

Example 5.4 (Categorical Noise Factors)

In the following we illustrate the stationary region analysis using an unbiased estimator of the process variance when noise factors are binomial. The methodology was developed in section 4.3. Suppose there are two control factors, x_1 and x_2, and two noise factors, z_1 and z_2. Furthermore, suppose z_1 denotes three different types of operators and each one of them has the same probability of being chosen and z_2 denotes three different operating conditions in which an equipment is operated. It is assumed that z_1 and z_2 are uncorrelated. By using indicator variables a model can be written as

$$y_i = \beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\beta_{12} x_{1i} x_{2i} + \gamma_1 I_{1i}+\gamma_2 I_{2i}+\gamma_3 I_{3i}+\gamma_4 I_{4i}$$
$$+ \delta_{11} x_{1i} I_{1i}+\delta_{12} x_{1i} I_{2i}+\delta_{13} x_{1i} I_{3i}+\delta_{14} x_{1i} I_{4i}$$
$$+ \delta_{21} x_{2i} I_{1i}+\delta_{22} x_{2i} I_{2i}+\delta_{23} x_{2i} I_{3i}+\delta_{24} x_{2i} I_{4i} + \varepsilon_i,$$

where

$$I_1 = \begin{cases} 1, & \text{if } z_1 = \text{operator type 1} \\ 0, & \text{otherwise,} \end{cases} \qquad I_2 = \begin{cases} 1, & \text{if } z_1 = \text{operator type 2} \\ 0, & \text{otherwise,} \end{cases}$$

$$I_3 = \begin{cases} 1, & \text{if } z_2 = \text{first condition} \\ 0, & \text{otherwise,} \end{cases} \qquad I_4 = \begin{cases} 1, & \text{if } z_2 = \text{second condition} \\ 0, & \text{otherwise.} \end{cases}$$

Note that these four indicator variables, I_1, I_2, I_3 and I_4, are Bernoulli trials with same success probabilities $\frac{1}{3}$ at each. The model can be written as

$$y_i = \beta_0 + \underline{x}_i'\underline{\beta} + \underline{x}_i' B \underline{x}_i + \underline{I}_i'\underline{\gamma} + \underline{x}_i'\Delta \underline{I}_i + \varepsilon_i,$$

where $\underline{x}_i' = [\ x_{1i} \quad x_{2i}\ ]$, $\underline{I}_i' = [\ I_{1i} \quad I_{2i} \quad I_{3i} \quad I_{4i}\ ]$, with parameters

$\underset{\sim}{\beta}' = [\ \beta_1 \quad \beta_2\ ]$, $\quad B = \begin{bmatrix} 0 & \frac{1}{2}\beta_{12} \\ \frac{1}{2}\beta_{12} & 0 \end{bmatrix}$, $\quad \underset{\sim}{\gamma}' = [\ \gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4\ ]$, and

$\Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} \\ \delta_{21} & \delta_{22} & \delta_{23} & \delta_{24} \end{bmatrix}$. We have $r_X = 2$, $r_I = 4$, $\underset{\sim}{\beta}$ is $r_X \times 1$, $\underset{\sim}{\gamma}$ is $r_I \times 1$, B is $r_X \times r_X$, and Δ is $r_X \times r_I$. Note also that

$$E(\underset{\sim}{I}) = \frac{1}{3}\underset{\sim}{1}_4 = \frac{1}{3}[1 \ \ 1 \ \ 1 \ \ 1]'$$

$$\mathrm{Var}(\underset{\sim}{I}) = V = \begin{bmatrix} \frac{2}{9} & -\frac{1}{9} & 0 & 0 \\ -\frac{1}{9} & \frac{2}{9} & 0 & 0 \\ 0 & 0 & \frac{2}{9} & -\frac{1}{9} \\ 0 & 0 & -\frac{1}{9} & \frac{2}{9} \end{bmatrix}.$$

The mean model can be written as

$$E_{\underset{\sim}{Z}}(y_i) = \beta_0 + \underset{\sim}{x}_i'\underset{\sim}{\beta} + \underset{\sim}{x}_i' B \underset{\sim}{x}_i + \{E(\underset{\sim}{I})\}'\underset{\sim}{\gamma} + \underset{\sim}{x}_i'\Delta\{E(\underset{\sim}{I})\},$$

and the process variance can be written as

$$\mathrm{Var}_{\underset{\sim}{Z}}(y \mid \underset{\sim}{x}_i) = \underset{\sim}{l}'V\underset{\sim}{l} + \sigma_\varepsilon^2,$$

$$\text{where } \ \underset{\sim}{l} = \frac{\partial y}{\partial \underset{\sim}{I}} = (\underset{\sim}{\gamma} + \Delta'\underset{\sim}{x}_i).$$

A combined array, $\frac{1}{2}$ fraction of 2^6 factorial design for the controls and noise factors, was used for the experiment. The data, parameter estimates and t statistics are shown below.

OBS	x1	x2	I1	I2	I3	I4	y
1	1	-1	0	0	0	0	30.5
2	-1	1	0	0	0	0	36.1
3	-1	-1	1	0	0	0	34.3
4	-1	-1	0	1	0	0	25.0
5	-1	-1	0	0	1	0	43.7
6	-1	-1	0	0	0	1	42.0
7	1	1	1	0	0	0	25.9
8	1	1	0	1	0	0	36.7

9	1	1	0	0	1	0	40.1
10	1	1	0	0	0	1	38.3
11	1	-1	1	1	0	0	27.8
12	1	-1	1	0	1	0	28.2
13	1	-1	1	0	0	1	26.1
14	1	-1	0	1	1	0	25.8
15	1	-1	0	1	0	1	23.8
16	1	-1	0	0	1	1	24.2
17	-1	1	1	1	0	0	16.0
18	-1	1	1	0	1	0	37.8
19	-1	1	1	0	0	1	36.4
20	-1	1	0	1	1	0	41.6
21	-1	1	0	1	0	1	40.2
22	-1	1	0	0	1	1	51.1
23	-1	-1	1	1	1	0	25.7
24	-1	-1	1	1	0	1	24.1
25	-1	-1	1	0	1	1	42.8
26	-1	-1	0	1	1	1	33.5
27	1	1	1	1	1	0	35.3
28	1	1	1	1	0	1	33.5
29	1	1	1	0	1	1	36.9
30	1	1	0	1	1	1	47.7
31	1	-1	1	1	1	1	21.5
32	-1	1	1	1	1	1	41.9

□

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	15	2115.64719	141.04315	38.244	0.0001
Error	16	59.00750	3.68797		
C Total	31	2174.65469			

Root MSE	1.92041	R-square	0.9729
Dep Mean	33.57812	Adj R-sq	0.9474
C.V.	5.71922		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	34.384375	0.75910811	45.296	0.0001
x1	1	-2.290625	0.75910811	-3.018	0.0082
x2	1	-0.165625	0.75910811	-0.218	0.8300
x12	1	1.765625	0.33948347	5.201	0.0001
I1	1	-5.381250	0.67896693	-7.926	0.0001
I2	1	-4.643750	0.67896693	-6.839	0.0001
I3	1	5.068750	0.67896693	7.465	0.0001
I4	1	3.343750	0.67896693	4.925	0.0002

x1I1	1	1.393750	0.67896693	2.053	0.0568
x1I2	1	4.881250	0.67896693	7.189	0.0001
x1I3	1	-2.931250	0.67896693	-4.317	0.0005
x1I4	1	-3.131250	0.67896693	-4.612	0.0003
x2I1	1	-3.131250	0.67896693	-4.612	0.0003
x2I2	1	3.431250	0.67896693	5.054	0.0001
x2I3	1	3.593750	0.67896693	5.293	0.0001
x2I4	1	3.718750	0.67896693	5.477	0.0001

□

To find the location of the estimated minimum process variance we see

$$\underline{\hat{l}} = \begin{bmatrix} \hat{l}_1 \\ \hat{l}_2 \\ \hat{l}_3 \\ \hat{l}_4 \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_1 + \hat{\delta}_{11} + \hat{\delta}_{21} \\ \hat{\gamma}_2 + \hat{\delta}_{12} + \hat{\delta}_{22} \\ \hat{\gamma}_3 + \hat{\delta}_{13} + \hat{\delta}_{23} \\ \hat{\gamma}_4 + \hat{\delta}_{14} + \hat{\delta}_{24} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

has no solution for x_1 and x_2. Therefore, we turn to the stationary region analysis approach developed in section 4.3. We see that

$$C = \begin{bmatrix} \frac{1}{16}(1+x_1^2+x_2^2) & \frac{1}{8}(1+x_1^2+x_2^2) & \frac{1}{8}(1+x_1^2+x_2^2) & \frac{1}{8}(1+x_1^2+x_2^2) \\ & \frac{1}{16}(1+x_1^2+x_2^2) & \frac{1}{8}(1+x_1^2+x_2^2) & \frac{1}{8}(1+x_1^2+x_2^2) \\ & & \cdots & \cdots \\ & (\text{sym}) & & \frac{1}{16}(1+x_1^2+x_2^2) \end{bmatrix}$$

apart from σ_ε^2, and

$$CV = (1+x_1^2+x_2^2)\begin{bmatrix} \frac{1}{16} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{16} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{16} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{16} \end{bmatrix}\begin{bmatrix} \frac{2}{9} & -\frac{1}{9} & 0 & 0 \\ -\frac{1}{9} & \frac{2}{9} & 0 & 0 \\ 0 & 0 & \frac{2}{9} & -\frac{1}{9} \\ 0 & 0 & -\frac{1}{9} & \frac{2}{9} \end{bmatrix}.$$

Therefore, we have tr(CV) = 0. We find the location of the estimated minimum process variance using an unbiased estimator for the process variance (see section 4.3). Recall the solution of (4.3.5),

$$\hat{\underline{x}}_0 = -(\hat{\Delta}V\hat{\Delta}' - s^2\underline{M}^*)^{-1}\hat{\Delta}V\hat{\underline{\gamma}},$$

where $\underline{M} = \underline{0}$. Using the parameter estimates shown above we get an estimated minimum process variance location as

$$\hat{\underline{x}}_0 = \begin{bmatrix} 1.0596671 \\ -0.547439 \end{bmatrix}.$$

This is indeed the location of the minimum process variance since

$$\hat{\Delta}V\hat{\Delta}' - s^2\underline{M}^* = \begin{bmatrix} 6.2631318 & 1.4519646 \\ \text{(sym)} & 10.15594 \end{bmatrix}$$

is a positive definite matrix.

Chapter 6. Future Research

It has become obvious that combined arrays and an RSM approach can be effective tools in our quest to reduce (process) variation. We would still like to study more about how, if at all, one attempts to estimate process variation if one uses a combined array and assumes some joint distribution for the noise factors. Actually, the question is not really how, since one can do this directly using a fitted (super) model and the assumed noise distribution. The concern is that the estimated process variation depends very much on the appropriateness of the model being fit. Issues regarding sequential design, appropriateness of variance model, and assumptions on noise variables are listed below.

1. In pursuing the optimal (say, maximum) mean response, sequential experiments along the steepest ascent path to move from one region to a more fruitful region are an attractive tool for the experimenter. Since we have already developed a good approach for fixed control and noise factors, we might develop a $(1-\alpha)100\%$ confidence region along the path of the steepest ascent with the process variance taken into account. We might take two different approaches:
   (i) use $\sqrt{\underline{l}'\underline{l}}$ or its Taylor series expanded form in pursuing the path of the steepest ascent, or
   (ii) compute the steepest ascent based on $\hat{y} - ks(\mathrm{x})$ as shown in one-sided tolerance intervals (Definition 3.7.2) in "larger the better" case. For

"smaller the better" case we will be looking into the steepest descent on $\hat{y} + ks(\mathrm{x})$.

2. In the "super model" (as model (3.1.1) is often called), if interactions are so important in analyzing the process variance, would there be an appropriate rule for the interaction terms to be included in the model? It would be useful if we can develop a certain yardstick analogous to C_p statistic (Myers, 1990) in regression analysis, or some standard model selection criteria. Consider Equation (3.1.8),

$$\underline{l} = \frac{\partial y}{\partial \underline{z}} = (\underline{\gamma} + \Delta' \underline{x}).$$

   (i) If we leave out Δ terms (that is, model misspecification) then a certain amount of bias of the estimated process variance would result.

   (ii) If we force in Δ terms (that is, over parametrizing) then we would have an inflated variance of the estimate of process variance.

   As has been well known, modeling for the variability is not easy and traditional ordinary least squares estimates often do not show much significance, thus resulting in only few terms in the Δ matrix. Instead of a traditional t-test, we might be able to develop some other convenient rules involving a mean square error type of criterion.

Appendix A. Choice of a Lagrangian Multiplier in Equation (3.4.1.6)

Recall Equation (3.4.1.6)

$$(\hat{\Delta}\hat{\Delta}' - s^2\underline{M} - \mu I_{r_X})\ \underline{x} = -\hat{\Delta}\hat{\underline{\gamma}}. \tag{A.1}$$

Denote the characteristic roots of $\hat{\Delta}\hat{\Delta}' - s^2\underline{M}$ by λ_i ($i = 1, 2, ..., r_X$, with r_X being the number of control factors). The λ_i's are such that

$$|\ \hat{\Delta}\hat{\Delta}' - s^2\underline{M} - \lambda_i I\ | = 0. \tag{A.2}$$

Consider the following quadratic form

$$\underline{u}'(\hat{\Delta}\hat{\Delta}' - s^2\underline{M} - \mu I)\underline{u} = \underline{u}'(\hat{\Delta}\hat{\Delta}' - s^2\underline{M})\underline{u} - \mu\underline{u}'\underline{u}. \tag{A.3}$$

Since $(\hat{\Delta}\hat{\Delta}' - s^2\underline{M})$ is real symmetric there exists an orthogonal transformation

$$\underline{u}' = v'P' \tag{A.4}$$

such that
$$\underline{u}'(\hat{\Delta}\hat{\Delta}' - s^2\underline{M})\underline{u} = v'D(\lambda_i)v, \tag{A.5}$$

and
$$\mu\underline{u}'\underline{u} = \mu v'P'Pv = \mu v'v, \tag{A.6}$$

$$\text{where } D(\lambda_i) = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ & & \cdots & \cdots \\ & (\text{sym}) & & \lambda_{r_X} \end{bmatrix}.$$

From (A.5) and (A.6), (A.3) can be written as

$$\underline{u}'(\hat{\Delta}\hat{\Delta}' - s^2\underline{M} - \mu I)\underline{u} = v'D(\lambda_i - \mu)v. \tag{A.7}$$

Therefore, if μ is chosen smaller than all of the r_X characteristics roots, (A.7) becomes bigger than zero, resulting in a positive definite $(\hat{\Delta}\hat{\Delta}' - s^2\underline{M} - \mu I)$. Recalling results from calculus, one sees that this gives a minimum at $\underline{x}$ on R. □

Appendix B. Computation of Two-Sided Tolerance Intervals

Assume a simple model

$$y = \beta_0 + \beta_1\,\mathrm{x} + \delta\,\mathrm{xz} + \varepsilon,$$

where x and z denote control factor and noise factor, respectively. The p-content tolerance interval at confidence level $(1-\gamma)$ is

$$\mathrm{P}_{\hat{\mu},\,s(x)}[\mathrm{P}_{y_f}\left\{(\mathrm{b}_0 + \mathrm{b}_1\mathrm{x}) - ks(\mathrm{x}) \le y_f \le (\mathrm{b}_0 + \mathrm{b}_1\mathrm{x}) + ks(\mathrm{x})\right\} \ge p] \ge 1-\gamma, \tag{B.1}$$

where $s(\mathrm{x}) = (\,\hat{l}^2 + \hat{\sigma}_\varepsilon^2)^{1/2}$ and $\hat{l} = \hat{\delta}\,\mathrm{x}$, $\hat{\sigma}_\varepsilon^2 =$ MSE from the model. Note that $s(\mathrm{x})$ is a function of x, the x location where we want to predict future y_f. Expression (B.1) can be written as

$$\mathrm{P}\left\{\,\mathrm{A}_\mathrm{X} \ge p,\ \ ^\forall\mathrm{x}\right\} = 1-\gamma,$$

where $\mathrm{A}_\mathrm{X} = \Phi\left\{\dfrac{(\mathrm{b}_0+\mathrm{b}_1\mathrm{x})+ks(\mathrm{x}) - (\beta_0+\beta_1\mathrm{x})}{\sqrt{l^2+\sigma_\varepsilon^2}}\right\} - \Phi\left\{\dfrac{(\mathrm{b}_0+\mathrm{b}_1\mathrm{x}) - ks(\mathrm{x}) - (\beta_0+\beta_1\mathrm{x})}{\sqrt{l^2+\sigma_\varepsilon^2}}\right\}$,

and

$$\Phi(\mathrm{Z}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\mathrm{Z}} \mathrm{e}^{-\frac{1}{2}t^2}\,\mathrm{dt}.$$

Without loss of generality assume $\beta_0 = \beta_1 = 0$, $\sqrt{l^2 + \sigma_\varepsilon^2} = 1$ $(\sigma_\varepsilon^2 \doteq 1)$. Denote

$$\mathrm{A}^*_{\mathrm{b}_0,\,\mathrm{b}_1,\,s(\mathrm{x})} = \min_{\mathrm{X}}\,[\Phi\left\{\,(\mathrm{b}_0 + \mathrm{b}_1\mathrm{x}) + ks(\mathrm{x})\right\} - \Phi\left\{\,(\mathrm{b}_0 + \mathrm{b}_1\mathrm{x}) - ks(\mathrm{x})\right\}]. \tag{B.2}$$

The problem is choosing k so that

$$\mathrm{P}\left\{\,\mathrm{A}^*_{\mathrm{b}_0,\,\mathrm{b}_1,\,s(\mathrm{x})} \ge p,\ \ ^\forall\mathrm{x}\right\} = 1-\gamma. \tag{B.3}$$

Equation (B.3) can be written as a conditional expectation as

$$E_{b_0,\, b_1}[\, P\left\{ \left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right) \middle| b_0, b_1 \right\}] = 1 - \gamma \tag{B.4}$$

The integrand can be expressed in a power series form as follows:

$$\begin{aligned}
P\left\{ \left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right) \middle| b_0, b_1 \right\} &= P\left\{ \left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right) \middle| b_0, b_1 \right\} \Big| (b_0 = 0,\ b_1 = 0) \\
&+ b_0 \frac{\partial}{\partial b_0} P\left\{ \left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right) \middle| b_0, b_1 \right\} \Big| (0,\ 0) \\
&+ b_1 \frac{\partial}{\partial b_1} P\left\{ \left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right) \middle| b_0, b_1 \right\} \Big| (0,\ 0) \\
&+ \frac{1}{2} [b_0^2 \frac{\partial^2}{\partial b_0^2} P\left\{ \left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right) \middle| b_0, b_1 \right\} \Big| (0,\ 0) \\
&+ 2 b_0 b_1 \frac{\partial^2}{\partial b_0 \partial b_1} P\left\{ \left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right) \middle| b_0, b_1 \right\} \Big| (0,\ 0) \\
&+ b_1^2 \frac{\partial^2}{\partial b_1^2} P\left\{ \left(A^*_{b_0, b_1, s(x)} \geq p\right) \middle| b_0, b_1 \right\} \Big| (0,\ 0)] + o(b_0^2,\ b_1^2).
\end{aligned} \tag{B.5}$$

From the definition of A_X, changing the signs of b_0 and x simultaneously does not alter A_X, we have $A^*_{b_0,\ b_1,\ s(x)} = A^*_{-b_0,\ b_1,\ s(x)}$. That is, $A^*_{b_0,\ b_1,\ s(x)}$ is symmetric in b_0 about 0 for each b_1 and $s(x)$. Therefore,

$$P\left\{ \left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right) \middle| b_0, b_1 \right\} = P\left\{ \left(A^*_{-b_0,\ b_1,\ s(x)} \geq p\right) \middle| -b_0, b_1 \right\}, \tag{B.6}$$

$$\frac{\partial}{\partial b_0} P\left\{ \left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right) \middle| b_0, b_1 \right\} \Big| (0,\ 0) = 0. \tag{B.7}$$

Since simultaneous changing of signs b_1 to $-b_1$ and x to $-x$ also leaves A_X unchanged. This gives $A^*_{b_0,\ b_1,\ s(x)} = A^*_{b_0,\ -b_1,\ s(x)}$. That is, $A^*_{b_0,\ b_1,\ s(x)}$ is symmetric in b_1 about 0 for each b_0 and $s(x)$. We get

$$P\left\{ \left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right) \middle| b_0, b_1 \right\} = P\left\{ \left(A^*_{b_0,\ -b_1,\ s(x)} \geq p\right) \middle| b_0, -b_1 \right\}, \tag{B.8}$$

$$\frac{\partial}{\partial b_1} P\left\{\left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right)\Big| b_0, b_1\right\}\Bigg|_{(0,\ 0)} = 0. \tag{B.9}$$

Also, since simultaneous sign changes of b_0 and b_1 leaves A_x unchanged $A^*_{b_0,\ b_1,\ s(x)}$ is symmetric in b_0 and b_1 about 0 for each x. Therefore,

$$\frac{\partial^2}{\partial b_0 \partial b_1} P\left\{\left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right)\Big| b_0, b_1\right\}\Bigg|_{(0,\ 0)} = 0. \tag{B.10}$$

From Equations (B.7), (B.9) and (B.10), Equation (B.5) can be written as

$$\begin{aligned} P\left\{\left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right)\Big| b_0, b_1\right\} &= P\left\{\left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right)\Big| b_0, b_1\right\}\Bigg|_{(0,0)} \\ &+ \frac{1}{2} b_0^2 \frac{\partial^2}{\partial b_0^2} P\left\{\left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right)\Big| b_0, b_1\right\}\Bigg|_{(0,\ 0)} \\ &+ \frac{1}{2} b_1^2 \frac{\partial^2}{\partial b_1^2} P\left\{\left(A^*_{b_0, b_1, s(x)} \geq p\right)\Big| b_0, b_1\right\}\Bigg|_{(0,\ 0)} + o(b_0^2,\ b_1^2), \end{aligned} \tag{B.11}$$

and

$$\begin{aligned} E_{b_0, b_1}[P\left\{\left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right)\Big| b_0, b_1\right\}] &= P\left\{\left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right)\Big| b_0, b_1\right\}\Bigg|_{(0,\ 0)} \\ &+ \frac{1}{2} E(b_0^2) \frac{\partial^2}{\partial b_0^2} P\left\{\left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right)\Big| b_0, b_1\right\}\Bigg|_{(0,\ 0)} \\ &+ \frac{1}{2} E(b_1^2) \frac{\partial^2}{\partial b_1^2} P\left\{\left(A^*_{b_0, b_1, s(x)} \geq p\right)\Big| b_0, b_1\right\}\Bigg|_{(0,\ 0)} + o(b_0^2,\ b_1^2) \\ &= P\left\{\left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right)\Big| b_0, b_1\right\}\Bigg|_{(0,\ 0)} \\ &+ \frac{1}{2} \frac{1}{n} \frac{\partial^2}{\partial b_0^2} P\left\{\left(A^*_{b_0,\ b_1,\ s(x)} \geq p\right)\Big| b_0, b_1\right\}\Bigg|_{(0,\ 0)} \\ &+ \frac{1}{2} \frac{1}{\sum x_i^2} \frac{\partial^2}{\partial b_1^2} P\left\{\left(A^*_{b_0, b_1, s(x)} \geq p\right)\Big| b_0, b_1\right\}\Bigg|_{(0,\ 0)} + O(\tfrac{1}{n}), \end{aligned}$$

$$\text{assuming } \sum_{i=1} x_i^2 = O(n). \tag{B.12}$$

Equating (B.11) and (B.12), we get

$$\mathrm{P}\left\{ \mathrm{A}^*_{\mathrm{b}_0,\ \mathrm{b}_1,\ s(\mathrm{x})} \geq p \right\} \doteq \mathrm{P}\left\{ \left(\mathrm{A}^*_{\mathrm{b}_0,\ \mathrm{b}_1,\ s(\mathrm{x})} \geq p \right) \Big| \mathrm{b}_0 = \frac{1}{\sqrt{n}},\ \mathrm{b}_1 = \frac{1}{\sqrt{\sum x_i^2}} \right\}, \tag{B.13}$$

The problem now reduces to choosing k so that

$$\mathrm{P}\left\{ \left(\mathrm{A}^*_{\mathrm{b}_0 = \frac{1}{\sqrt{n}},\ \mathrm{b}_1 = \frac{1}{\sqrt{\sum x_i^2}},\ s(\mathrm{x})} \geq p \right) \right\} = 1 - \gamma. \tag{B.14}$$

By definition this is equivalent to

$$\min_{\mathrm{x}} \left[\Phi \left\{ \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{\sum x_i^2}} \mathrm{x} + ks(\mathrm{x}) \right\} - \Phi \left\{ \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{\sum x_i^2}} \mathrm{x} - ks(\mathrm{x}) \right\} \right] = p. \tag{B.15}$$

Let the solution of (B.15) be x^*, where x^* satisfies the following transcendental equation

$$\frac{\partial}{\partial \mathrm{x}} \left[\Phi \left\{ \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{\sum x_i^2}} \mathrm{x} + ks(\mathrm{x}) \right\} - \Phi \left\{ \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{\sum x_i^2}} \mathrm{x} - ks(\mathrm{x}) \right\} \right] = 0. \tag{B.16}$$

The solution x^* will be a function of $ks(\mathrm{x})$ and $\mathrm{A}^*_{\frac{1}{\sqrt{n}},\ \frac{1}{\sqrt{\sum x_i^2}},\ s(\mathrm{x})}$ will be monotonically non-decreasing function of $ks(\mathrm{x})$ since by letting $\mathrm{A}^* = g\,(ks(\mathrm{x}))$ we see $g(0) = 0$ and $g(\infty) = 1$. There exists h_0 such that

$$g(h_0) = p. \tag{B.17}$$

If k is chosen such that

$$\mathrm{P}\left\{ ks(\mathrm{x}) > h_0 \right\} = 1 - \gamma, \tag{B.18}$$

then

$$\mathrm{P}\left\{ \left(\mathrm{A}^*_{\frac{1}{\sqrt{n}},\ \frac{1}{\sqrt{\sum x_i^2}},\ s(\mathrm{x})} \geq p \right) \right\} = 1 - \gamma. \tag{B.19}$$

Once h_0 is found numerically k can be obtained easily from the distribution theory

$$s^2 = \hat{l}^2 + \hat{\sigma}_\varepsilon^2 = \hat{l}^2 + \frac{\chi^2_{(df = n-2)}}{(n-2)}, \tag{B.20}$$

since
$$(n-2)\hat{\sigma}_\varepsilon^2 \sim \chi^2_{(n-2)}. \tag{B.21}$$

From (B.18)

$$\mathrm{P}\left\{s^2(\mathrm{x}) > \frac{h_0^2}{k^2}\right\} = 1 - \gamma. \tag{B.22}$$

Therefore,

$$\mathrm{P}\left\{\hat{l}^2 + \frac{\chi^2_{(n-2)}}{(n-2)} \geq \frac{h_0^2}{k^2}\right\} = 1 - \gamma. \tag{B.23}$$

From (B.23) k, therefore, two-sided tolerance intervals can be obtained. □

Appendix C. Computation of the Inequality (3.7.5)

The probability statement of (3.7.5)

$$\mathrm{P}_{\hat{\mu},\, s(\underline{x})}\left\{\frac{\mathrm{Z}+\lambda}{\sqrt{\frac{\widehat{l'l}}{\sigma_\varepsilon^2}+\frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}}} \le \frac{k}{\sqrt{\underline{x}'(\mathrm{X}'\mathrm{X})^{-1}\underline{x}}}\right\} \ge 1-\gamma \tag{C.1}$$

can be solved numerically by multiple integration (see section 3.7). To do so, we make use of the technique of the transformations of variables as shown below.

The one-to-one transformations can be written as

$$y_1 = \frac{\mathrm{Z}+\lambda}{\sqrt{\mathrm{X}_1+\mathrm{X}_2}},\ y_2 = \mathrm{X}_1 \text{ and } y_3 = \mathrm{X}_2, \tag{C.2}$$

where $\mathrm{Z} \sim \mathrm{N}(0,\, 1)$, $\mathrm{X}_1 \sim c\rho\chi^2_{\nu_1}$ and $\mathrm{X}_2 \sim \frac{\chi^2_{\mathrm{dfE}}}{\mathrm{dfE}}$. See section 3.7 for the values of c, ρ and ν_1. The inverse transformation can be written as

$$x_1 = y_2,\ x_2 = y_3 \text{ and } z = y_1\sqrt{y_2+y_3}-\lambda, \tag{C.3}$$

with the absolute value of the Jacobian of the transformation equal to $\sqrt{y_2+y_3}$. Since Z, X_1 and X_2 are mutually independent, the joint probability distribution function can be written as

$$f(z_1,x_1,x_2) = \frac{1}{\sqrt{2\pi}\,\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\mathrm{dfE}}{2})(2c\rho)^{\nu_1/2}(\frac{2}{\mathrm{dfE}})^{\mathrm{dfE}/2}}\, \mathrm{e}^{-\frac{z^2}{2}-\frac{x_1}{2c\rho}-\frac{(\mathrm{dfE})x_2}{2}}\, x_1^{\frac{\nu_1}{2}-1}\, x_2^{\frac{\mathrm{dfE}}{2}-1}, \tag{C.4}$$

where $-\infty < z < \infty$ and $0 < x_1,\ x_2 < \infty$. The joint probability distribution function of y_1, y_2 and y_3 can be found by

$$g(y_1, y_2, y_3) = f(y_1\sqrt{y_2+y_3} - \lambda,\ y_2,\ y_3) \cdot \sqrt{y_2+y_3}. \tag{C.5}$$

This allows us to find the marginal probability distribution function of the left hand side of (3.7.5) by

$$g(y_1) = \int_0^\infty \int_0^\infty g(y_1, y_2, y_3)\, dy_2\, dy_3. \tag{C.6}$$

Therefore, values of k in (C.1) can be found by

$$\int_{-\infty}^{\frac{k}{\sqrt{\underline{x}'(\mathrm{X}'\mathrm{X})^{-1}\underline{x}}}} g(y_1)\, dy_1 \geq 1 - \gamma. \tag{C.7}$$

□

References

Bartlett, M. S., and Kendall, D. G. (1946), "The Statistical Analysis of Variance Heterogeneity and the Logarithmic Transformation," *JRSS*, B8, pp. 128-138.

Borkowski, J., and Lucas, J. M. (1991), "The Analysis of Mixed Resolution Designs," paper presented at the Joint Statistical Meetings, Atlanta, GA, August.

Box, G. E. P. (1985), "Discussion of Off-Line Quality Control, Parameter Design, and the Taguchi Methods," *Journal of Quality Technology*, 17, pp. 198-206.

Box, G. E. P. (1988), "Signal-to-Noise Ratios, Performance Criteria, and Transformations (with Discussion)," *Technometrics*, Vol. 30, pp. 1-40.

Box, G. E. P., and Draper, N. R. (1987), *Empirical Model-Building and Response Surfaces*, John Wiley & Sons, New York.

Box, G. E. P., and Hunter, J. S. (1961a), "The 2^{k-p} Fractional Factorial Designs I," *Technometrics*, Vol. 3, pp. 311-351.

Box, G. E. P., and Hunter, J. S. (1961b), "The 2^{k-p} Fractional Factorial Designs I," *Technometrics*, Vol. 3, pp. 449-458.

Box, G. E. P., and Jones, S. (1990), "Designing Products that are Robust to the

Environment," paper presented at the Joint Statistical Meetings, Anaheim, CA, August.

Box, G. E. P., and Meyer, R. D. (1986), "Dispersion Effects from Fractional Designs," *Technometrics*, Vol. 28, pp. 19-27.

Carroll, R. J., and Ruppert, D. (1988), *Transformation and Weighting in Regression*, Chapman and Hall, New York, NY.

Carroll, R. J., and Ruppert, D. (1991), "Prediction and Tolerance Intervals with Transformation and/or Weighting," *Technometrics*, Vol. 33, No. 2, pp. 197-210.

Graybill, F. A. (1983), *Matrices with Applications in Statistics*, Wadsworth, Belmont, CA.

Kacker, R. N. (1985), "Off-Line Quality Control, Parameter Design, and the Taguchi Method," *Journal of Quality Technology*, 17, pp. 176-209.

Khuri, A. I., and Cornell, J. A. (1987), *Response Surfaces*, Marcel Dekker, New York.

Lucas, J. M. (1985), Comment on "Off Line Quality Control, Parameter Design, and the Taguchi Method," by R. N. Kacker, *Journal of Quality Technology*, 17, pp. 195-197.

Myers, R. H. (1991), "Response Surface Methodology in Quality Improvement," Commun. Statist. - Theory & Meth., Vol. 20, No. 2, pp. 457-476.

Myers, R. H. (1990), *Classical and Modern Regression with Applications*, PWS-KENT Publishing Company, Boston, MA.

Myers, R. H. (1976), *Response Surface Methodology*, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA.

Myers, R. H., and Carter, W. H., Jr. (1973), "Response Surface Techniques for Dual Response Systems," *Technometrics*, Vol. 15, pp. 301-317.

Myers, R. H., Khuri, A. I., and Vining G. G. (1992), "Response Surface Alternatives to the Taguchi Robust Parameter Design Approach," *The American Statistician*, Vol. 46. No. 2, pp. 131-139.

Nair, V. N., and Pregibon, D. (1988), "Analyzing Dispersion Effects from Replicated Factorial Experiments," *Technometrics*, Vol. 30, pp. 247-257.

Patnaik, P. B. (1949), "The Non-Central χ^2- and F-Distributions and Their Applications," *Biometrika* 36, pp. 202-232.

Pignatiello, J. J., and Ramberg, J. S. (1985), "Discussion (of Kacker, 1985)," *Journal of Quality Technology*, 17, pp. 198-206.

Plackett, R. L., and Burman, J. P. (1946), "The Design of Optimum Multifactorial Experiments," *Biometrika* 33, pp. 305-325.

Raghavarao, D. (1971), *Construction and Combinatorial Problems in Design of Experiments*, John Wiley & Sons, New York.

Rao, C. R. (1947), "Factorial Experiments Derivable from Combinatorial Arrangements of Arrays," *JRSS*, B9, pp. 128-137.

Satterthwaite, F. (1946), "An Approximate Distribution of Estimates of Variance Components," *Biometric Bulletin* 2, pp. 110-114.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, John Wiley & Sons, New York.

Shoemaker, A. C., Tsui, K. L., and Wu, C. F. J. (1991), "Economical Experimentation Methods for Robust Design," *Technometrics*, Vol. 33, pp. 415-427.

Taguchi, G. (1977), *Experimental Design* (Vol. 1 & 2), Maruzen Publishing Co., Tokyo, Japan.

Taguchi, G. (1988), *SN-Ratio for the Quality Evaluation.* Japanese Standards Association, Tokyo, Japan.

Taguchi, G., and Wu, Y. (1985), "Introduction to Off-Line Quality Control," Central Japan Quality Control Association, Nagoya, Japan.

Vining, G. G., and Myers, R. H. (1990), "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach," *Journal of Quality Technology*, 22, pp. 38-45.

Welch, B. L. (1956), "On Linear Combinations of Several Variances," *Journal of the American Statistical Association* 51, pp. 132-148.

Welch, W. J., and Sacks, J. (1991), "A System for Quality Improvement Via Computer Experiments," *Commun. Statist. - Theory & Meth.*, Vol. 20, No. 2, pp. 477-495.

Welch, W. J., Yu, T. K., Kang, S. M., and Sacks, J. (1990), "Computer Experiments for Quality Control by Parameter Design," *Journal of Quality Technology*, 22, pp. 15-22.

Vita

The author was born in October 7, 1958 to Kirak Kim and Bunki Kim. He grew up in many different places. He graduated from Tongnai High School in Pusan, S. Korea. In February of 1981, he graduated from Seoul National University in Seoul, Korea, with a Bachelor of Science degree in Computer Science and Statistics.

In February of 1984, he received his Master of Science degree in Statistics at the Graduate School of Seoul National University.

In August of 1984, he entered the Korean Army and after six months of service he was discharged with an award as one of three best Second Lieutenant.

In June of 1987, he received another Master of Science degree in Statistics at the Department of Mathematics and Statistics of Wright State University in Dayton, Ohio.

In September of 1992, he completed the requirements for a Doctor of Philosophy degree in Statistics at Virginia Polytechnic Institute and State University.

The author is a member of the American Statistical Association and the Mu Sigma Rho honorary societies.

Since August of 1992, the author has been teaching statistics classes for undergraduate students at Humboldt State University in Arcata, California as an assistant professor.