Optimal Design and Analysis of Experiments
Y. Dodge, V.V. Fedorov and H.P. Wynn (Editors)

ANALYSIS OF VARIABILITY BY ANALYSIS OF VARIANCE

Friedrich Pukelsheim

Institut für Mathematik, Universität Augsburg

Memminger Strasse 6, D–8900 Augsburg, Fed. Rep. of Germany

Summary

A brief overview of total quality control is given. An essential part of it is data, and statistical methods to evaluate these data. This is exemplified by minimizing process variability using analysis of variance. Signal-to-noise ratios applied to the untransformed data are contrasted with more classical methods applied to appropriately transformed data. The points made are illustrated by example.

1. Introduction

The success of Japanese products in the Western world is proverbial. Much of this success is due to Japanese products having the reputation of being high quality products, as a result of a total devotion to quality. Section 2 presents a brief overview of ideas underlying this total quality control philosophy.

An essential part of this philosophy—or *the* essential part—is data. Hence statistics, as a tool to evaluate the data, plays a prominent role. Following the writings to G. Taguchi data should serve to bring a process onto a given target value, with minimum variability. He utilizes signal-to-noise ratios to achieve this goal (Section 3). An alternative approach proposed by G.E.P. Box is presented in Section 4, and is illustrated by example in Section 5.

The paper concludes with a bibliography of recent publications on the subject, ranging from more mathematically oriented papers, over case studies in various fields, to publications that promote a comprehensive strategy based on a total quality control concept.

2. Total quality control—the Japanese way

Many aspects of the Japanese approach to total quality control can be contrasted with what seems to be customs in the Western world. Total quality control is an active and permanent task; it starts as early as in the off-line stage when the product is being conceived; it is factual, obsessed with data; it is an ongoing dynamic circle of prediction, confirmation, and improvement.

Active versus passive

Western quality control wrongly concentrates on such things as acceptance sampling and sampling schemes, and process control and control charts. This is a passive approach to quality, preventing the worst by screening out bad products. Total quality control wants to build quality into the product from the very beginning, and starts as early as in the product design stage or when the production process is being planned. There is a heavy emphasis on the fact that bad products at the end of the proper manufacturing process may very well originate from the product design or from the way the production process has been set up.

Off-line versus on-line

In the on-line stage when the actual manufacturing is going on it is often too expensive, or even too late, to make adjustments necessary in order for improving quality. It is much preferable to start in the off-line stage, i.e. prior to the on-line stage. Taguchi (1985) makes a point to distinguish *parameter design* associated with the off-line stage, from *tolerance design* associated with the on-line stage. That is to say, in the on-line stage bad quality can often only be improved by tightening the tolerance specifications. In contrast, parameter design in the off-line stage singles out parameters for the manufacturing process that come from a set of equally feasible and equally costly values. A sample run on every possible parameter combination is unnecessary and expensive, and often impossible by the sheer number of parameter combinations. A well-planned experiment provides sufficient information to choose appropriate parameter levels, as advocated in many textbooks on industrial statistics.

Data versus words

Total quality control is obsessed with data. Even the textbook by Ishikawa (1982) which addresses a wide audience emphasizes that data and diagrams are an absolute necessity to pin down the strong and weak points of the production process under discussion. Graves (1986, p. 7) reports from his career as a consultant that "... data be brought to the meetings. Without data there can be endless discussions without resolution of the issues."

Data force us to face facts, they do not in the first place upset the hierarchy that puts a company's employees into a top-to-bottom order. Data activate what Hunter (1986) calls the two untapped resources of western economies, potential information and employee creativity.

Without data a problem soon converts to a philosophical disput, instructions fail to be operational, and commitment evaporates into verbal eloquence.

Dynamic versus static

Total quality control brings improvement, but nobody is perfect. Perpetual improvement is an ongoing challenge, circling around analysis and experiment, prediction and confirmation. This dynamic procedure can only bear fruits when total quality control is part of a company-wide policy that involves all levels of management. Topmanagement must lead all quality efforts, or else data are dumped rather than acted upon. Garvin (1983), with his contraposition of the room airconditioner manufacturers in Japan and the United States, has a story to tell.

Total quality control goes far beyond the statistical tasks of collecting and analysing data. But it embraces these tasks as an integral part. For a statistician the message is twofold: Statistical methods form a necessary prerequisite, but they are only a technical tool in the strategic kit called total quality control.

3. Signal-to-noise ratios—the Taguchi method

Much of the statistical methods used in the Japanese approach to total quality control is due to Professor Genichi Taguchi, engineer and statistician. There is no such thing as "the Taguchi method," instead Taguchi builds on a profound experience with a wide variety of applied problems.

Box (1986, p. 20) singles out three characteristics: (a) Taguchi stresses that a production process must be close to target, rather than merely being within specification. (b) Process variance must be minimized, besides bringing the process mean to the target value. Planned experiments are run to provide the necessary data. (c) The resulting product must be "robust" not only relative to manufacturing imperfections but also with respect to those noise factors as will be present in the user environment. The necessary data are obtained from a planned experiment that artificially produces variability by simulating the noise factors.

The idea of "design for design (experimental design for engineering design)"—as Basso, Winterbottom and Wynn (1986, p. 73) call it—shifts the emphasis back to the off-line design stage. Of its many facets we shall here discuss item (b), of minimizing process variability while keeping the process mean on target.

Signal-to-noise statistics help analyse a process so that the signal is optimized relative to noise. Taguchi uses a battery of such statistics, each tailored to a specific situation. In his authoritative review Kackar (1985, pp. 183–184) distinguishes three such situations: The smaller the better, the larger the better, and a specific target value is best.

The smaller the better, the larger the better

The three situations have in common that the observations Y take nonnegative values. A target value zero, $\tau = 0$, then manifests the goal 'the smaller the better.' With squared error as loss this leads to minimize the mean squared error,

$$MSE = E[Y^2].$$

Taguchi prefers a decibel scale as being more easily understood by process engineers, thus calling for the maximization of

$$-10 \log MSE.$$

The corresponding sampling quantities then lead to the signal-to-noise statistic

$$-10 \log \left(\sum_{i=1}^{n} \frac{y_i^2}{n} \right).$$

The second situation, the larger the better, is reduced to the first by a transformation from Y to its reciprocal $1/Y$. With this in mind the proposed signal-to-noise statistic is

$$-10 \log \left(\sum_{i=1}^{n} \frac{1/y_i^2}{n} \right).$$

I oppose these extremes on two grounds, a practical and a theoretical one. On the practical side, consider a company that produces a control device which under normal operating conditions has to withstand a certain pressure, 20 bar, say. It is better if the device withstands 25 bar, it is excellent if it withstands 30 bar, and the larger the pressure the better. But it is ridiculous to require that the product withstands 1000 bar, or so. Setting the target at $\tau = 40$ bar would ensure an excellent performance, setting the target at $\tau = \infty$ bar is thoughtless nonsense.

On a theoretical ground it is either Y or $1/Y$ that will have a tractable distribution. Usually the observations Y, or a transformation thereof, are arranged so that they do not too significantly deviate from a normal distribution. If such normality is achieved there is no reason to give it up by a somewhat hasty transition to the reciprocal.

A specific target value is best

When practical situations allow to specify a target value it will be a finite and positive value $\tau = \tau_0$. The mean squared error of the observation from the target value τ_0 is the sum of the squared bias $(\eta - \tau_0)^2$ and the variance σ^2 of the observation Y,

$$MSE = (\eta - \tau_0)^2 + \sigma^2.$$

Taguchi recommends maximization of the signal-to-noise statistic

$$10 \log \left(\frac{\bar{y}^2}{s^2} \right),$$

the reason being that in many technical problems it is reasonable to allow for the sample variance s^2 to vary proportionally to the squared sample average $\bar{y}^2$.

The analysis is much simplified when mean η and variance σ^2 do not depend on each other. Then the squared bias $(\eta - \tau_0)^2$ and the variance σ^2 could be minimized without having to account for a functional dependence.

In any case signal-to-noise statistics are used to classify the experimental factors into

— *control factors* which in the first place account for process variability, and

— *signal factors* which in the first place affect the mean behaviour of the process.

For an individual factor the statistical analysis may show that it significantly contributes to process variability (a control factor), or to the mean behaviour of the process (a signal factor), or to neither of those (a nuisance factor), or to both. The latter case makes it particularly clear that the classification into control factors and signal factors is a data dependent decision of the experimenter.

4. Analysis of variability

With the classification into control factors and signal factors in mind Box (1986) proposes an alternative approach based on monotonic data transformations: Find a transformation $Y = f(y)$ that allows easy separation into control factors and signal factors, analyse the transformed data, and check whether the results have a reasonable interpretation within the original data.

According to Box a transformation is successful if it achieves *parsimony*, that is if it minimizes the number of parameters, and if it eliminates *crosstalk*, that is if it permits a separation of location effects and dispersion effects and hence a classification into signal factors and control factors.

Separation by transformation

We summarize the procedure for an experiment consisting of runs $i = 1, \ldots, \ell$ where in each run we observe independent replications y_{ij} for $j = 1, \ldots, r$. The first step then is to find a transformation $Y_{ij} = f(y_{ij})$ such that the transformed data can be modelled with fewer parameters, that their location effects and dispersion effects more clearly separate, and that they more closely conform with the classical linear model assumptions in the sense expounded by Box and Cox (1964). This catalogue of desiderata is quite demanding, and any particular case may well have to terminate with compromise solutions.

Suppose that run i is determined by a vector of experimental conditions x_i, and that the observations Y_{ij} have mean η_i and variance σ_i^2 independent of j. Further assume that the transformation from y_{ij} to Y_{ij} has resulted in a functional independence of mean and variance. Then the approach is to estimate σ_i^2 by

$$S_i^2 = \sum_{j=1}^{r}(Y_{ij} - \overline{Y}_i)^2,$$

and to regress $\log S_i$ on x_i. (The logarithmic transformation from S_i^2 to $\log S_i$ accounts for the functional dependence between normal second and fourth moments, namely σ_i^2 and $2\sigma_i^4$.)

Such an *analysis of variability* leads to a fractional factorial design with unreplicated observations $\log S_i$, and the recent papers by Box and Meyer become relevant. This has to be distinguished from the *analysis of means*, from regressing the replicated observations $Y_{i1}, \ldots, Y_{ir}$ on x_i.

Analysis of variability versus analysis of means

The two analyses entail divergent interests. For the mean analysis the transformation $Y_{ij} = f(y_{ij})$ ought to result in constant variances $\sigma_i^2 = \sigma^2$, in order to approach the assumptions underlying a normal analysis of variance. For the variability analysis equal variances are undesirable, since they carry no information on how the experimental factors influence the process variability. The extreme of equal sample variances S_i^2 prohibits an identification of noise factors. The other extreme, when the sample variances S_i^2 vary significantly from run to run, facilitates identification of noise factors by an analysis of variability, but forces the analysis of means to be at least a *weighted* regression.

5. Free height of leaf springs—the Pignatiello and Ramberg data

Pignatiello and Ramberg (1985) report an experiment to calibrate the free height of leaf springs to eight inches, with experimental factors being (B) furnace temperature, (C) heating time, (D) transfer time, and (E) hold down time. A fifth factor (O), quench oil temperature, is included in the experiment with the understanding that in large scale manufacturing it would appear as an uncontrollable noise factor. The experiment was carried out with a 2_{III}^{7-4} fractional factorial design, in order to analyse the four two-level main effects (B), (C), (D) and (E) and the three interaction effects (BC), (BD), and (CD). In each of the eight runs six replicates were observed, half of them at the low level of the noise factor, O^-, and half of them at the high level of the noise factor, O^+.

Taguchi's signal-to-noise statistic

In order to check the data for normality we center the six replicates of each run at the run mean $\bar{y}_i$. This removes from each run the conjectured mean effect, even though it also introduces a slight dependence between the resulting six centered observations. The remaining array of row-wise centered data is then referred to the corresponding 48 normal scores. The correlation is found to be 0.986, and does not indicate a significant deviation from a normality assumption.

Next we study the assumption of homoscedastic variances. The row-wise sample variances s_i^2, as given by Pignatiello and Ramberg (1985, p. 200), are 0.01 times 9, 7, 0.1, 1, 9, 5, 4, 2. They visibly deviate from an homoscedasticity assumption. Indeed, Bartlett's test statistic takes the value M = 26.29, and this lies in the significant tail of a χ^2-distribution with seven degrees of freedom.

Pignatiello and Ramberg (1985, p. 200) analyse the apparent variability using Taguchi's signal-to-noise statistic

$$T_i = 10\log_{10}\left(\frac{\bar{y}_i^2}{s_i^2}\right).$$

Notice that numerator and denominator span quite different ranges,

$$\frac{\bar{y}^2_{\max}}{\bar{y}^2_{\min}} = \frac{7.9^2}{7.37^2} = 1.15, \qquad \frac{s^2_{\max}}{s^2_{\min}} = \frac{0.09}{0.001} = 90.$$

Finally an analysis of variance is carried out with responses $T_1, \ldots, T_8$. The percentage of factor sums of squares relative to total sums of squares are tabulated in column 3 of Table 1.

Rank	Source	T	$\widehat{\lambda}=0$	Source	$\widetilde{T}$	$\widehat{\lambda}=-2.66$
1	C	53.80	54.00	B	40.76	33.91
2	CD	16.86	17.16	DO	14.07	! 13.61
3	D	13.07	12.68	BCO	13.52	! 14.43
4	BD	7.46	7.51	CDO	8.31	9.53
5	E	5.42	5.27	CD	5.13	! 5.98
6	BC	3.32	3.32	CO	4.08	! 6.72
7	B	0.07	0.06	(9)	14.13	15.82

Table 1. Faktor ranking by percentage of $SS(\text{factor})/SS(\text{total})$. Left portion has quench oil as noise factor: Taguchi's T and the log transformation $\widehat{\lambda} = 0$ induce identical rankings. Right portion includes main effect and seven interaction effects from quench oil: Pignatiello and Ramberg's $\widetilde{T}$ and the power transformation $\widehat{\lambda} = -2.66$ induce almost identical rankings.

Box-Cox power transformations, with quench oil as noise factor

We now try to find an appropriate transformation from the Box and Cox (1964) power family, $Y = y^\lambda$, see also Box and Draper (1987, pp. 288–291). First we take quench oil temperature O as a noise factor. Then we have six replicates for each of the eight runs. The maximum likelihood estimate of the transformation is $\widehat{\lambda} = -0.034$, calling for a logarithmic transformation of the data,

$$Y_{ij} = \ln y_{ij}.$$

The normal scores test and Bartlett's test are very much the same as for the untransformed data. The correlation of the row-wise centered data with the normal scores is 0.984. Bartlett's test statistic takes the value M = 26.18.

The analysis of variability, of regressing $\log S_i$ on x_i, produces virtually the same relative size of factor sums of squares, and the ranking is identical with that from Taguchi's T. The percentage of factor sums of squares relative to total sums of squares is listed in column 4 of Table 1.

Indeed, Box (1986, part I) makes a point that when the data call for a logarithmic transformation then the analysis of the transformed data and the analysis based on Taguchi's signal-to-noise statistic T are equivalent.

Box-Cox power transformation, with quench oil as controllable factor

When we carry out the analysis under the assumption that quench oil is a controllable factor, following Pignatiello and Ramberg (1985, pp. 204–205), we expand the model by a main effect for O and the seven interactions of O with all previous factors.

The maximum likelihood estimate of the power transformation is $\widehat{\lambda} = -2.66$. Again the normal scores test and Bartlett's test remain unaffected. The correlation of the row-wise centered data with the normal scores is 0.984 as before, and Bartlett's test statistic takes the value M = 26.31.

Finally an analysis of variance is carried out on $\log S_i$. Column 7 of Table 1 shows the relative size of factor sums of squares. Column 6 exhibits the corresponding numbers from the analysis of Pignatiello and Ramberg who choose the signal-to-noise statistic

$$\widetilde{T}_i = 10 \log_{10} s_i^2$$

based on the untransformed data. Again we observe that the size of the factor sums of squares virtually coincide, and induce almost the same factor ranking.

In either case we have not succeeded in transforming the data closer to normal regression assumptions. An analysis of means based on homoscedastic variances should therefore be taken with care.

6. Bibliography

Books, Special Journal Issues, Symposia

Box, G.E.P. and Draper, N.R., *Empirical Model Building and Response Surfaces*, (John Wiley & Sons, New York, 1987) xiv+669 pp.

Center for Quality and Productivity Improvement Report Series, University of Wisconsin–Madison, 610 Walnut Street, Madison, Wisconsin 53705, USA

Communications in Statistics, Theory and Methods, Vol. 14, No. 11, Special Issue on Conference on Advances in Statistics in Quality Control (1982) 2575–2844.

Ishikawa, K., *Guide to Quality Control*, (Asian Productivity Organization, UNIPUB Kraus International Publications, New York, 1982) vii+225 pp.

Journal of Quality Technology, Vol. 17, No. 4 (1985) 175–246.

Supplier Symposium on Taguchi Methods, April 10, 1984, 200 pp., Second Supplier Symposium on Taguchi Methods, November 20, 1984, 184 pp., Third Supplier Symposium on Taguchi Methods, October 8, 9, 1985, 458 pp., distributed by the American Supplier Institute, Inc., Center for Taguchi Methods, Six Parklane Blvd., Suite 411, Dearborn, MI 48126-2618, USA.

Taguchi, G., *On-line Quality Control during Production* (Japanese Standards Association, distributed by the American Supplier Institute, 1981) 154 pp.

Taguchi, G. and Wu, Y., *Off-line Quality Control* (Central Japan Quality Control Association, distributed by the American Supplier Institute, 1985) 111 pp.

Articles and Case Studies

Bartlett, M.S. and Kendall, D.G., The statistical analysis of variance-heterogeneity and the logarithmic transformation, *Journal of the Royal Statistical Society, Supplement*, **8** (1946) 128–138.

Basso, L., Winterbottom, A. and Wynn, H.P., A review of the "Taguchi Methods" for off-line quality control, *Quality and Reliability Engineering International*, **2** (1986) 71–79.

Box, G.E.P., Studies in quality improvement: Signal-to-noise ratios, performance criteria and statistical analysis, part I: Report No. 11, 17 pp., part II: Report No. 12, 23 pp., Center for Quality and Productivity Improvement, University of Wisconsin–Madison, USA (1986).

Box, G.E.P. and Cox, D.R., An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, **26** (1964) 211–243; discussion 244–252.

Box, G.E.P. and Meyer, R.D., Analysis of unreplicated factorials allowing for possibly faulty observations, MRC Technical Summary Report No. 2799, 12 pp., Mathematics Research Center, University of Wisconsin–Madison, USA (1985).

Box, G.E.P. and Meyer, R.D., An analysis for unreplicated fractional factorials, *Technometrics*, **28** (1986) 11–18.

Box, G.E.P. and Meyer, R.D., Dispersion effects from fractional designs, *Technometrics*, **28** (1986) 19–27.

Bryce, G.R. and Colette, D.R., Optimizing wafer processing via statistical design, *Semiconductor International*, March (1984) 71–77.

Burgam, P.M., Design of experiments – the Taguchi way, *Manufacturing Engineering*, May (1985) 44–46.

Fuller, F.T., Eliminating complexity from work: improving productivity by enhancing quality, Report No. 17, 18 pp., Center for Quality and Productivity Improvement, University of Wisconsin–Madison, USA (1986).

Graves, S., A process for consulting for improvement in quality and productivity, Report No. 21, 18 pp., Center for Quality and Productivity Improvement, University of Wisconsin–Madison, USA (1986).

Gaddum, J.H., Lognormal distributions, *Nature*, **156** (1945) 463–466.

Garvin, D.A., Quality on the line, *Harvard Business Review*, **61** (1983) 65–75.

Hendrix, C.D., What every technologist should know about experimental design, *Chemtech*, March (1979) 167–174.

Hogg, R.V. et al, Statistical education for engineers: An initial task force report, *The American Statistician*, **39** (1985) 168–175.

Hunter, W.G., Managing our way to economic success: two untapped resources, Report No. 4, 24 pp., Center for Quality and Productivity Improvement, University of Wisconsin–Madison, USA (1986).

Johnson, M. and Lee, K., Optimizing of a photoresist process using statistical design of experiments, *Solid State Technology*, September (1984) 281–284.

Joiner, B.L., The key role of statisticians in the transformation of North American industry, *The American Statistician*, **39** (1985) 224–227.

Jugle, D.B., Comments on "Off-line Quality Control in Integrated Circuit Fabrication Using Experimental Design", with a reply by M.S. Phadke, R.N. Kackar, D.V. Speeney and M.J. Grieco, *The Bell System Technical Journal*, (1984) 1405–1409.

Kackar, R.N., Some orthogonal arrays for screening designs, Technical Report, Bell Laboratories, Holmdel, N.J., USA (1982).

Kackar, R.N. Off-line quality control, parameter design, and the Taguchi method, *Journal of Quality Technology*, **17** (1985) 176–188.

Kackar, R.N., Taguchi's quality philosophy: analysis and commentary, *Quality Progress*, December (1986) 21–29.

Kackar, R.N. and Shoemaker, A.C., Robust design: a cost-effective method for improving manufacturing processes, *AT&T Technical Journal*, **65** (1986) 39–50.

Katz, L.E. and Phadke, M.S., Macro-quality with micro-money, *Record*, November (1985) 22–28.

León, R.V., Shoemaker, A.C. and Kacker, R.N., Performance measures independent of adjustment, an explanation and extension of Taguchi's signal-to-noise ratios, *Technometrics*, **29** (1987) 253–265; discussion 266–285.

Lin, K.M. and Kacker, R.N., Optimizing the wave soldering process, *Electronic Packaging & Production*, February (1986) 108–115.

Marquardt, D.W., Statistical consulting in industry, *The American Statistician*, **33** (1979) 102–107.

Meyer, R.D., Studies in quality improvement: analysis of fractional factorials, Report No. 16, 19 pp., Center for Quality and Productivity Improvement, University of Wisconsin–Madison, USA (1986).

Meyer, R.D., Further details of an analysis for unreplicated fractional factorials, Report No. 22, 33 pp., Center for Quality and Productivity Improvement, University of Wisconsin–Madison, USA (1987).

Meyer, R.D. and Box, G.E.P., Identification of active factors in unreplicated fractional factorial experiments, Report, No. 23, 23 pp., Center for Quality and Productivity Improvement, University of Wisconsin–Madison, USA (1987).

Montgomery, D.C., Friedman, D.J. and Heikes R.G., Experimental design in process design, development, and optimization, Typescript, University of Washington FU-10, Seattle, Washington 98195, USA (1986) 27 pp.

Nachtsheim, C.J., Tools for computer-aided design of experiments: Review and comment, *Journal of Quality Technology*, **19** (1985) 132–160.

Pao, T.W., Phadke, M.S. and Sherrerd, C.S., Computer response time optimization using orthogonal array experiments, Paper prepared for ICC–85, IEEE Communications Society, Chicago, IL, USA, June (1985) 6 pp.

Phadke, M.S., Quality engineering using design of experiments, *Proceedings of the American Statistical Association*, Section on Statistical Education, August (1982) 11–20.

Phadke, M.S., Kackar, R.N., Speeney, D.V. and Grieco, M.J., Off-line quality control in integrated circuit fabrication using experimental design, *The Bell System Technical Journal*, **62** (1983) 1273–1309.

Pignatiello, J.J. and Ramberg J.S., Discussion of Kackar (1985), *Journal of Quality Technology*, **17** (1985) 198–206.

Taguchi, G., Performance analysis design, *Int. J. Prod. Res.*, **16** (1978) 521–530.

Taguchi, G. and Phadke, M.S., Quality engineering through design optimization, Conference Record, GLO–BECOM 84 Meeting, IEEE Communications Society, Atlanta, GA, USA (1984) 1106-1113.

Upperman, P.M., Industrial statistics, *Statistica Neerlandica*, **39** (1985) 203–217.

OPTIMAL DESIGN AND ANALYSIS OF EXPERIMENTS

edited by

Yadolah DODGE

University of Neuchâtel
Switzerland

Valeri V. FEDOROV

International Institute
for Applied Systems Analysis
Austria

Henry P. WYNN

City University, London
England

1988

NORTH-HOLLAND
AMSTERDAM · NEW YORK · OXFORD · TOKYO