AMERICAN JOURNAL OF MATHEMATICAL AND MANAGEMENT SCIENCES

MEASUREMENT SCALES AND RESOLUTION IV DESIGNS: A NOTE

Bert Bettonvil
Catholic University Brabant/Technical University Eindhoven

Jack P.C. Kleijnen
Catholic University Brabant
Tilburg, Netherlands

SYNOPTIC ABSTRACT

Three measurement scales and models are considered, namely standardized, original, and centered. It is argued that the effects at the midpoint (of the experimental area) are relevant. Effects per unit and effects over the whole range are distinguished. Resolution IV designs permit unbiased estimation of these effects, even if quantitative factors are present.

Key Words and Phrases: scaling, standardization, centering, midpoint, quantitative factors.

1990, VOL. 10, NOS. 3 & 4, 309-322
0196-6324/90/030309-14 $19.00

1. INTRODUCTION.

This note discusses two issues:

(i) The use of different scales, namely the original, the standardized, and the centered scales. Each scale gives different values for "the" effects (or regression parameters).

(ii) Resolution IV designs give unbiased estimators of the standardized and the centered main effects, but not of the original main effects.

Textbooks on experimental design do not pay much attention to scaling effects. Their designs are always presented in standardized form; that is, the original factors (say) z_j are scaled so that the standardized factors x_j satisfy $-1 \leq x_j \leq 1$ (also see equation (4) below). The technical literature shows that the interpretation of standardization remains controversial even today; that is, the mathematics is simple, but its meaning takes us outside the area of mathematical proofs and into the fuzzy area of "semantics"; see Arnold (1986, pp. 96-100, 169-175), Box and Draper (1987, p. 667), Box and Hunter (1957, p. 221), Gunst (1983, pp. 2224, 2237-2241), Hocking (1983, pp. 235-236), Hocking (1984), Mendenhall (1963, pp. 221-229, 251-257), Mihram (1972, pp. 359-360), Snee and Marquardt (1984), etc. The social sciences, especially sociology and psychology, pay much more attention to the issue of choosing a correct scale; see Ghiselhi, Campbell and Zedeck (1981), Krantz, Luce, Suppes, and Tversky (1971), Suppes and Zinnes (1963), etc. We show that different scales give different conclusions. Moreover, if interactions are possible, then "Resolution IV" designs (defined in Section 4) are popular in the literature and in practice. However, Seely and Birkes (1984, pp. 84-85) state: "an aspect of the resolution IV property that oftentimes appears to be neglected in the literature [is] ... there is no guarantee that the resolution IV property will carry over to a different parametrization". Actually, Kleijnen (1987, p. 308) states that Resolution IV designs do not preserve their acclaimed property in case quantitative factors are present in the experiment; this note, however, withdraws that statement; for a correct statement see Section 4.

2. THREE MEASUREMENT SCALES.

The following three measurement scales or models are relevant. The original model is

$$y = \gamma_0 + \Sigma_{j=1}^{k} \gamma_j z_j + \Sigma_{j=1}^{k-1} \Sigma_{j'=j+1}^{k} \gamma_{jj'} z_j z_{j'} \tag{1}$$

where z_j $(j=1,\dots,k)$ are the actual values of the factor variables, for which

$$L_j \le z_j \le H_j \quad (-\infty < L_j < H_j < \infty). \tag{2}$$

The standardized model is

$$y = \beta_0 + \Sigma_{j=1}^{k} \beta_j x_j + \Sigma_{j=1}^{k-1} \Sigma_{j'=j+1}^{k} \beta_{jj'} x_j x_{j'} \tag{3}$$

where

$$x_j = \frac{z_j - b_j}{a_j} \text{ with } a_j = \frac{H_j - L_j}{2} \text{ and } b_j = \frac{H_j + L_j}{2} \tag{4}$$

and

$$-1 \le x_j \le 1 \quad (j=1,\dots,k). \tag{5}$$

The centered model is

$$y = \delta_0 + \Sigma_{j=1}^{k} \delta_j (z_j - \bar{z}_j) +$$
$$+ \Sigma_{j=1}^{k-1} \Sigma_{j'=j+1}^{k} \delta_{jj'} (z_j - \bar{z}_j)(z_{j'} - \bar{z}_{j'}) \quad (6)$$

where

$$\bar{z}_j = \frac{1}{n} \Sigma_{i=1}^{n} z_{ij} \quad (7)$$

and z_{ij} denotes the value of factor j in observation i ($i=1,\dots,n$). Note that we restrict ourselves to balanced experiments where $\bar{z}_j = (L_j + H_j)/2 = b_j$; in practice most designs are balanced.

If we are interested in prediction only, then it does not matter which model we use: it is easy to prove that $y(\hat{\beta}) = y(\hat{\gamma}) = y(\hat{\delta})$; see Kleijnen (1987, p. 345). However, we may also use the model for explanation, for example, to answer questions like "what is the effect of a change in factor 1?". We distinguish between the "effect of factor 1 when the other factors are at their midpoints $(z_2=\bar{z}_2,\dots,z_k=\bar{z}_k)$" and the "effect of factor 1 when the other factors are at the origin $(z_2=0,\dots,z_k=0)$". This origin may be outside $[L_2, H_2] \times \dots \times [L_k, H_k]$. For example, in a medical experiment z_2 may refer to body temperature and range between 36° C and 42° C. Then the effect of z_1 when $z_2=0$ is of no interest. (We return to this issue in Section 3.)

For the "main" or "first order" effects β_j, γ_j and δ_j equations (8), (9), and (10) hold, which can be seen as follows. From (3) we see that β_j is the effect of factor j at the point where all other factors $j'(j' \neq j)$ are zero in the standardized scale: $x_{j'} = 0$. Likewise (1) shows that γ_j is the effect of factor j if all other factors are zero in the original scale. Finally, (6) shows that δ_j is the effect of factor j if all other factors are at their midpoints $\bar{z}_{j'}$ in the centered scale. Briefly:

$$\beta_j = \frac{dy}{dx_j}\Big]_{\forall j' \neq j(x_{j'}=0)} \tag{8}$$

$$\gamma_j = \frac{dy}{dz_j}\Big]_{\forall j' \neq j(z_{j'}=0)} \tag{9}$$

$$\delta_j = \frac{dy}{dz_j}\Big]_{\forall j' \neq j(z_{j'}=\bar{z}_{j'})} . \tag{10}$$

Upon substitution of (5) into (6) we obtain

$$y = \delta_0 + \Sigma_{j=1}^{k} \delta_j a_j x_j + \Sigma_{j=1}^{k-1} \Sigma_{j'=j+1}^{k} \delta_{jj'} a_j a_{j'} x_j x_{j'}$$

which, combined with (3), implies

$$\beta_j = \delta_j a_j . \tag{11}$$

Equation (11) gives a relationship between the standardized model (3) and the centered model (6). The relationship between the original model (1) and the centered model (6) follows from computing dy/dz_j at the point $z_{j'}=0$ for model (6) and at $z_{j'} = \bar{z}_{j'}$ for model (1) respectively; this yields

$$\gamma_j = \delta_j - \Sigma_{j' \neq j} \delta_{jj'} b_{j'} \tag{12}$$

and

$$\delta_j = \gamma_j + \Sigma_{j' \neq j} \gamma_{jj'} b_{j'} . \tag{13}$$

3. INTERPRETATION.

Before we give an example in Section 4, we interpret the mathematical results of Section 2. First, note that the experimenter is interested in the effect of (say) a temperature change of one degree Centigrade, not in a change of one "standard" degree; only the original and the centered model use Centigrade (see (1) and (6)).

Second, the experimenter is interested in the effect of such a unit change at the midpoint of the experimental area ($\bar{z}_j$), not at the origin ($\bar{z}_j = 0$), in general. For example, if in a medical experiment z_j ranges between 36° and 42°, then the effect of any other varable at 0° is irrelevant. A (rare) counter-example is an experiment in which it is technically infeasible to make observations at zero degrees Kelvin (approximately $-273^\circ C$); then we may restrict the experimental area to (say) $-250^\circ \leq z_1 \leq -100^\circ$ and from the experimental results we extrapolate the effect at the point $z_1 = -273^\circ$.

Third, δ_j (and also γ_j) measures the change per unit change of the original variable z_j. The total change as z_j varies over the experimental area $[L_j, H_j]$, equals the product of δ_j and $H_j - L_j$. Consequently, if factor j has a wider range $R_j = H_j - L_j$ than factor j' has, then the total effect of z_j may be higher than the total effect of $z_{j'}$, even if $\delta_j < \delta_{j'}$; also see Kleijnen (1987, pp. 141-142). Now (5) and (11) yield

$$\delta_j(H_j - L_j) = \delta_j\ (2\ a_j) = 2\ \beta_j. \qquad (14)$$

So the unit effect is best measured by δ_j; the relative "importance" (total effect) of a factor is best measured by the relative β's and the "original" effects γ_j are misleading.

Note that the ranges may be unknown at the beginning of an experiment. For example, in Response Surface Methodology (RSM) the quantitative factors are to be optimized. Therefore local experiments are combined with the steepest ascent technique applied to first-order models (where $\beta_{jj'} = \gamma_{jj'} = \delta_{jj'} = 0$). Kleijnen (1988) measures the importance of different factors through $\gamma_j \bar{z}_j$ where $\bar{z}_j$ is the midpoint of z_j in a sequence of local experiments s = 1,2,... (so γ_j and $\bar{z}_j$ should be indexed by s).

Fourth, in the presence of interactions we must qualify our answer to the question "What is the effect of factor j?": we must answer "the effect of factor j, if the values of the other factors are... ". So we should address the following issues:

(i) "Are there significant interactions?" ($\beta_{jj'}, \gamma_{jj'}, \delta_{jj'}$ have identical t-statistics; it does not matter which one is tested; see Kleijnen, van den Burg, and van der Ham (1979, p. 62)).

(ii) "If the other factors are at their midpoints ($z_{j'} = \bar{z}_{j'}$), then the estimated effect of a unit change in factor j is $\hat{\delta}_j$"; see (10). ($\hat{\delta}_j$ equals $\hat{\beta}_j / a_j$; the estimate $\hat{\delta}_j$ is significant if and only if $\hat{\beta}_j$ is (see (11)). If the standard design is orthogonal in x_j and the observations y_i are independent with common variance, then the $\hat{\beta}_j$ are orthogonal; cov $(\hat{\gamma}_j, \hat{\gamma}_{j'})$ equals $b_j b_{j'} \text{cov}(\delta_{jj})$ and is not zero.)

4. RESOLUTION IV DESIGNS.

Box and Hunter(1961, p. 319) defined Resolution IV designs as experimental designs with no main effect confounded with any other main effect or two-factor interaction, but with two-factor interactions aliased with each other.

The statistical literature gives experimental designs expressed in standardized variables x_j. A Resolution IV design yields unbiased estimators $\hat{\beta}_j$ of the standardized main effects β_j. These estimators $\hat{\beta}_j$ do not give unbiased estimators of the original effects γ_j; see (11) and (12) where no estimates of individual interactions are available. However, we saw that the $\hat{\beta}_j$ do give unbiased estimators of δ_j. The estimators $\hat{\delta}_j$ are relevant, not the estimators $\hat{\gamma}_j$, since δ_j measures the unit effect at the midpoint (see (10)) and the midpoint is of interest (not the zero point, except for very special situations; see Section 3). The $\hat{\beta}_j$ are of interest since they measure the total effect over the range R_j; see (14).

As a numerical example consider an s,S inventory system; that is, when at the end of a time period the inventory is below level s, an order is placed to increase the inventory up to level S. Demand during a time period is a random variable. We may simulate this system, in order to estimate the effects on the out-of-stock probability (y) of the following six factors x_j $(j=1,\dots,6)$:

Factor 1: demand distribution family. The symbol + in Table 1 denotes the exponential distribution with parameter ν or mean $1/\nu$; the symbol - denotes the uniform distribution with domain from $1/(2\nu)$ to $3/(2\nu)$ so that the mean remains $1/\nu$.

Factor 2: distribution of the delivery (or lead) time family: + refers to the exponential distribution with parameter μ, whereas - denotes the Erlang distribution with parameters 4μ and 4 (sum of four exponential distributions); hence the mean lead time is $1/\mu$.

Factor 3: expected demand: - denotes the value 1; + denotes 5.

Factor 4: expected lead time: - means 3, and + denotes 7.

Factor 5: expected time between demands: - denotes 1, and + means 6. The distribution of these times is exponential.

Table 1. Input and output in s, S inventory example.

Run	Input factors x_j with j =						output
	1	2	3	4	5	6	y
1	+	+	+	+	+	+	0.0542
2	+	+	-	+	-	-	0.0301
3	+	-	+	-	+	-	0.0502
4	+	-	-	-	-	+	0.0002
5	-	-	-	+	+	+	0.0000
6	-	-	+	+	-	-	0.7177
7	-	+	-	-	+	-	0.0000
8	-	+	+	-	-	+	0.1925
9	-	-	-	-	-	-	0.0000
10	-	-	+	-	+	+	0.0328
11	-	+	-	+	-	+	0.0328
12	-	+	+	+	+	-	0.1009
13	+	+	+	-	-	-	0.3120
14	+	+	-	-	+	+	0.0000
15	+	-	+	+	-	+	0.6572
16	+	-	-	+	+	-	0.0024

Factor 6: s value: - denotes 3, and + means 10. S is fixed at 25.

Here factors 1 and 2 are qualitative, while factors 3 through 6 are quantitative. We use a Resolution IV design, that is, we use sixteen runs (or factor combinations), namely a 2^{6-2} design with generators 5 = 234 and 6 = 134 so that 23 = 45, 24 = 35, 13 =

Table 2. Standardized versus original versus centered scale in s,S inventory example: no interactions.

Number of the effect j	Estimated standardized main effect $\hat{\beta}_j$	$(\hat{\sigma}_{\hat{\beta}_j})$	Estimated original main effect $\hat{\gamma}_j$	$(\hat{\sigma}_{\hat{\gamma}_j})$	Estimated centered main effect $\hat{\delta}_j$	$(\hat{\sigma}_{\hat{\delta}_j})$
0	0.136	(0.042)*	-0.036	(0.163)	0.136	(0.042)*
1	0.002	(0.042)	0.002	(0.042)	0.002	(0.042)
2	-0.046	(0.042)	-0.046	(0.042)	-0.046	(0.042)
3	0.128	(0.042)*	0.064	(0.021)*	0.064	(0.021)*
4	0.063	(0.042)	0.032	(0.021)	0.032	(0.021)
5	-0.106	(0.042)*	-0.043	(0.017)*	-0.043	(0.017)*
6	-0.115	(0.042)	-0.004	(0.012)	-0.004	(0.012)

* Significant at level 0.05; see standard errors shown in parentheses (df=9).

46, 14 = 36, 12 = 56, 15 = 26, 25 = 34 = 16. Table 2 displays the estimated main effects $\hat{\gamma}_j$ and the overall mean $\hat{\gamma}_0$ assuming all interactions are zero. Table 3 gives results if the seven two-factor interactions in the left-hand side of the preceding confounding relations are assumed to be important. Table 4 gives results if the "right-hand" interactions are important (45, 35, 46, 36, 56, 26, and 34). The estimated standard errors, shown in Tables 2, 3, and 4, are computed from the residual sum of squares; to test significance we use the Student statistic t_ν

Table 3. Standardized versus original versus centered scale in s,S inventory example: "left-hand" interactions.

Number of the effect $j(jj')$	Estimated standardized main effect $\hat{\beta}_{j(jj')}$	$(\hat{\sigma}_{\hat{\beta}_{j(jj')}})$	Estimated original main effect $\hat{\gamma}_{j(jj')}$	$(\hat{\sigma}_{\hat{\gamma}_{j(jj')}})$	Estimated centered main effect $\hat{\delta}_j$	$(\hat{\sigma}_{\hat{\delta}_{j(jj')}})$
0	0.136	(0.007)*	-0.036	(0.026)	0.136	(0.007)*
1	0.002	(0.007)	0.045	(0.023)	0.002	(0.007)
2	-0.046	(0.007)*	0.205	(0.023)*	-0.046	(0.007)*
3	0.128	(0.007)*	0.064	(0.003)*	0.064	(0.003)*
4	0.063	(0.007)*	0.032	(0.003)*	0.032	(0.003)*
5	-0.106	(0.007)*	-0.043	(0.003)*	-0.043	(0.003)*
6	-0.015	(0.007)	-0.004	(0.002)	-0.004	(0.002)
23	-0.054	(0.007)*	-0.027	(0.003)*	-0.027	(0.003)*
24	-0.099	(0.007)*	-0.049	(0.003)*	-0.049	(0.003)*
13	0.002	(0.007)	0.001	(0.003)	0.001	(0.003)
14	-0.015	(0.007)	-0.008	(0.003)	-0.008	(0.003)
12	0.007	(0.007)	0.007	(0.007)	0.007	(0.007)
15	-0.005	(0.007)	-0.002	(0.003)	-0.002	(0.003)
25	0.055	(0.007)*	0.022	(0.003)*	0.022	(0.003)*

* Significant at level 0.05; see standard errors shown in parentheses (df=2).

with degrees of freedom ν equal to 16-7 and 16-14 respectively. Tables 2, 3, and 4 show that in a Resolution IV design the standardized main effects $\hat{\beta}_1$ through $\hat{\beta}_6$ and the centered main

Table 4. Standardized versus original versus centered scale in s,S inventory example: "right-hand" interactions.

Number of the effect $j(jj')$	Estimated standardized main effect $\hat{\beta}_{j(jj')}$	$(\hat{\sigma}_{\hat{\beta}_{j(jj')}})$	Estimated original main effect $\hat{\gamma}_{j(jj')}$	$(\hat{\sigma}_{\hat{\gamma}_{j(jj')}})$	Estimated centered main effect $\hat{\delta}_j$	$(\hat{\sigma}_{\hat{\delta}_{j(jj')}})$
0	0.136	(0.007)*	-0.242	(0.061)	0.136	(0.007)*
1	0.002	(0.007)	0.002	(0.007)	0.002	(0.007)
2	-0.046	(0.007)*	-0.036	(0.014)	-0.046	(0.007)*
3	0.128	(0.007)*	0.079	(0.012)*	0.064	(0.003)*
4	0.063	(0.007)*	0.026	(0.010)	0.032	(0.003)*
5	-0.106	(0.007)*	0.065	(0.010)*	-0.043	(0.003)*
6	-0.015	(0.007)	-0.002	(0.007)	-0.004	(0.002)
45	-0.054	(0.007)*	-0.011	(0.001)*	-0.011	(0.001)*
35	-0.099	(0.007)*	-0.020	(0.001)*	-0.020	(0.001)*
46	0.002	(0.007)	0.000	(0.001)	0.000	(0.001)
36	-0.015	(0.007)	-0.002	(0.001)	-0.002	(0.001)
56	0.007	(0.007)	0.001	(0.001)	0.001	(0.001)
26	-0.005	(0.007)	-0.002	(0.002)	-0.002	(0.002)
34	0.055	(0.007)*	0.014	(0.002)*	0.014	(0.002)*

* Significant at level 0.05; see standard errors shown in parentheses (df=2).

effects $\hat{\delta}_1$ throught $\hat{\delta}_6$ are not affected by the inclusion of two-factor interaction; the original main effects $\hat{\gamma}_1$ through $\hat{\gamma}_6$, however, are influenced.

Another example is found in Kleijnen, van den Burg, and van der Ham (1979) where a harbor was simulated: the t statistic (with 128 degrees of freedom) for one of the standardized factors is 30.85; for the original factor this value is only 0.06.

5. CONCLUSION.

Experimenters should center their independent variables as in (6), because unit effects δ_j at the midpoint $\bar{z}$ are of interest, in general. Resolution IV designs give unbiased estimators of these "centered" main effects δ_j (even if the factors are quantitative so that the standard designs must be linearly transformed as in (5)). These designs also give unbiased estimators of the standardized effects β_j which measure the total effects over the ranges R_j. When optimizing, the importance of different factors is measured through $\gamma_j \bar{z}_j$ where $\bar{z}_j$ is the midpoint of z_j in the local experiment, and γ_j is the local original effect.

REFERENCES

Arnold, S.F. (1981). *The Theory of Linear Models and Multivariate Analysis*. John Wiley & Sons, Inc., New York.

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

Box, G.E.P. and Hunter, J.S. (1957). Multi-factor experimental designs for exploring response surfaces. *Annals of Mathematical Statistics*, 28, 195-241.

Box, G.E.P. and Hunter, J.S. (1961). The 2^{k-p} fractional factorial designs, Part I. *Technometrics*, 3, 311-351.

Ghiselli, E.E., Campbell, J.P. and Zedeck, S. (1981). *Measurement Theory for the Behavioral Sciences*. W.H. Freeman and Company, San Francisco.

Gunst, R.F. (1983). Regression analysis with multicollinear predictor variables: Definition, detection, and effects. *Communications in Statistics*, *12A*, 2217-2260.

Hocking, R.R. (1983). Developments in linear regression methodology: 1959-1982 (and discussion). *Technometrics*, *25*, 219-249.

Hocking, R.R. (1984). Letter to the editor. *Technometrics*, *26*, 299-301.

Kleijnen, J.P.C. (1987). *Statistical Tools for Simulation Practitioners*. Marcel Dekker, Inc., New York.

Kleijnen, J.P.C. (1988). Simulation and optimization in production planning: a case study. Report number 308, Catholic University Brabant, Tilburg, The Netherlands.

Kleijnen, J.P.C., van den Burg, A.J., and van der Ham, R.Th. (1979). Generalization of simulation results: practicality of statistical methods. *European Journal of Operational Research*, *3*, 50-64.

Krantz, D.H., Luce, R.D., Suppes, P., and Tversky, A. (1971). *Foundations of Measurement, I*. Academic Press, New York.

Mendenhall, W. (1968). *Introduction to Linear Models and the Design and Analysis of Experiments*. Wadsworth Publishing Company, Belmont, California.

Mihram, G.A. (1972). *Simulation: Statistical Foundations and Methodology*. Academic Press, New York.

Seely, J. and Birkes, D. (1984). Parametrizations and resolution IV. *Experimental Design, Statistical Models, and Genetic Statistics*, edited by K. Hinkelmann, Marcel Dekker, New York, 77-94.

Snee, R.D. and Marquardt, D.W. (1984). Collinearity diagnostics depend on the domain of prediction, the mode, and the data. *The American Statistician*, *38*, 83-87.

Suppes, P. and Zinnes, J.R. (1963). Basic measurement theory. *Handbook of Mathematical Psychology, Volume I*, edited by R.D. Luce, R.R. Bush, and E. Galanter. John Wiley & Sons, Inc., New York, 4-74.

Received 9/9/87; Revised 11/23/89.