APR 20 1999

SANDIA REPORT

SAND98–2202
Unlimited Release
Printed April 1999

Statistical Analyses of Scatterplots to Identify Important Factors in Large-Scale Simulations

Jack P.C. Kleijnen, Jon C. Helton

RECEIVED
APR 26 1999
OSTI

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Prices available from (703) 605-6000
Web site: http://www.ntis.gov/ordering.htm

Available to the public from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd
Springfield, VA 22161

NTIS price codes
Printed copy: A06
Microfiche copy: A01

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

SAND98-2202
Unlimited Release
Printed April 1999

Statistical Analyses of Scatterplots to Identify Important Factors in Large-Scale Simulations

Jack P.C. Kleijnen, Tilburg University, 5000 LE Tilburg, The Netherlands

Jon C. Helton, Arizona State University, Tempe, AZ 85287-1804, USA

Abstract

Procedures for identifying patterns in scatterplots generated in Monte Carlo sensitivity analyses are described and illustrated. These procedures are based on attempts to recognize increasingly complex patterns in the scatterplots under consideration and involve the identification of (i) linear relationships with correlation coefficients, (ii) monotonic relationships with rank correlation coefficients, (iii) trends in measures of central tendency as defined by means, medians and the Kruskal-Wallis statistic, (iv) trends in measures of variability as defined by variances and interquartile ranges, and (v) deviations from randomness as defined by the chi-square statistic. As illustrated in a sequence of example analyses with a large model for two-phase fluid flow, the individual procedures can differ in the variables that they do, and do not, identify as having effects on particular analysis outcomes. The example results indicate that the use of a sequence of procedures is a good analysis strategy and provides some assurance that an important effect is not overlooked due to the use of an inappropriate analysis procedure.

Intentionally Left Blank

CONTENTS

1 Introduction 9

2 Test Problems 11

3 Linear Relation: $y = \beta_0 + \beta_1 x$ 25

4 Monotonic Relation: $r(y) = \gamma_0 + \gamma_1 r(y)$ 35

5 Location of y Dependent on x 37

5.1 Common Means: ANOVA F Statistic 37
5.2 Common Locations: Kruskal-Wallis Test 39
5.3 Common Medians: Chi-Square for Contingency Tables 40

6 Dispersion of y Dependent on x 43

6.1 Common Variances: ANOVA F Statistic with Jackknifing 43
6.2 Common Interquartiles: Chi-Square for Contingency Tables 44

7 Distribution of y Dependent on x: Chi-Square for Contingency Tables 47

8 Identification of Important Variables 49

9 Top-Down Correlation 51

10 Comparison of Procedures for Identification of Important Variables 53

10.1 Repository Pressure under Undisturbed Conditions: y = *E0:WAS_PRES* 53
10.2 Brine Inflow under Undisturbed Conditions: y = *E0:BRAALIC* 62
10.3 Repository Saturation under Disturbed Conditions: y = *E2:WAS_SATB* 70
10.4 Repository Pressure under Disturbed Conditions: y = *E2:WAS_PRES* 75

11 Type I and II Errors 85

12 Robustness with Respect to Repeated Independent Samples 91

13 Discussion 97

14 References 99

FIGURES

2.1 Computational grid used in BRAGFLO to represent two-phase flow in 1996 WIPP CCA PA subsequent to a drilling intrusion. Same formulation is used in the absence of a drilling intrusion except that regions 1A, 1B and 1C have the same properties as the regions to either side. 12

2.2 Dependent variables predicted by BRAGFLO model for two-phase flow selected for use in comparison of statistical procedures for identification of patterns in scatterplots: (2.2a) pressure in lower waste panel under undisturbed conditions (*E0:WAS_PRES*), (2.2b) cumulative brine inflow from anhydrite marker beds under undisturbed conditions (*E0:BRAALIC*), (2.2c) saturation in lower waste panel after an E2 intrusion at 1000 yr (*E2:WAS_SATB*), and (2.2d) pressure in lower waste panel after an E2 intrusion at 1000 yr (*E2:WAS_PRES*). 16

2.3 Scatterplots for pressure in lower waste panel under undisturbed (i.e., E0) conditions at 10,000 yr (*E0:WAS_PRES*) versus first four variables selected in stepwise regression analyses with raw and rank-transformed data (Table 2.3): (2.3a) *WMICDFLG*, (2.3b) *HALPOR*, (2.3c) *WGRCOR*, and (2.3d) *ANHPRM*. 19
2.4 Scatterplots for cumulative brine inflow over 10,000 yr from all anhydrite marker beds to repository under undisturbed (i.e., E0) conditions (*E0:BRAALIC*) versus first four variables selected in stepwise regression analysis with rank-transformed data (Table 2.3): (2.4a) *WMICDFLG*, (2.4b) *ANHPRM*, (2.4c) *HALPOR*, and (2.4d) *WGRCOR*. 20
2.5 Scatterplots for brine saturation in lower (i.e., intruded) waste panel at 10,000 yr for an E2 intrusion at 1000 yr (*E2:WAS_SATB*) versus first four variables selected in stepwise regression analysis with rank-transformed data (Table 2.3): (2.5a) *BHPRM*, (2.5b) *WRGSSAT*, (2.5c) *ANHPRM* and (2.5d) *HALPOR*. 22
2.6 Scatterplots for pressure in lower waste panel at 10,000 yr with an E2 intrusion into the lower waste panel at 1000 yr (*E2:WAS_PRES*) versus the three variables (*HALPRM*, *ANHPRM*, *HALPOR*) selected in stepwise regression analysis with raw and rank-transformed data (Table 2.3) and one additional variable (*BHPRM*) identified by examination of scatterplots: (2.6a) *HALPRM*, (2.6b) *ANHPRM*, (2.6c) *HALPOR*, and (2.6d) *BHPRM*. 23

3.1 Graph of $d(b, a) = b(1 - b^2)^{1/2} - a(1 - a^2)^{1/2} > 0$ subject to constraints $0 < a < b < 1$, $a^2 + b^2 < 1$. 33

5.1 Example of the partitioning of the range of x = *HALPOR* into $nX = 5$ classes for y = *E0:WAS_PRES*. 38
5.2 Example of the partitioning of the range of x = *HALPOR* into $nX = 5$ classes and the range of y = *E0:WAS_PRES* into values above and below the median $y_{0.5}$. 40

6.1 Example of the partitioning of the range of x = *HALPOR* into $nX = 5$ classes and the range of y = *E0:WAS_PRES* into values inside and outside the interquartile range $[y_{0.25}, y_{0.75}]$. 45

7.1 Example of the partitioning of the range of x = *HALPOR* into $nX = 5$ classes and the range of y = *E0:WAS_PRES* into $nY = 5$ classes. 48

10.1 Scatterplot for *E0:WAS_PRES* versus *ANHBCVGP*. 55
10.2 Scatterplots *E0:BRAALIC* versus *SHPRMCON* and *HALPRM*. 64
10.3 Scatterplot for *E2:WAS_SATB* versus *WGRCOR*. 72
10.4 Scatterplot for *E2:WAS_PRES* versus *WGRCOR*. 79

11.1 Contour plots for probability of a Type I error, *prob* (Type I | p, n), as a function of p-value, p, and number of unimportant variables, n. 86
11.2 Distribution of p-values and associated Kolmogorov-Smirnov bounds for individual tests and variables in LHS that do not affect *E0:WAS_PRES*, *E0:BRAALIC*, *E2:WAS_SATB* and *E2:WAS_PRES*. 87
11.3 Distributions of p-values for 10 independently generated LHSs: (11.3a) CCs for *E0:WAS_PRES* and (11.3b) RCCs for *E0:WAS_PRES* 90
11.4 Distribution of p-values for 10 independently generated LHSs for CVs test and *E0:BRAALIC* 90

TABLES

2.1 Definition of Dependent Variables Predicted by BRAGFLO Model for Two-Phase Flow Selected for Use in Comparison of Statistical Procedures for Identification of Patterns in Scatterplots 11
2.2 Uncertain Variables Used as Input to BRAGFLO in the Calculation of the Dependent Variables in Table 2.1 (see Table 5.2.1, Helton et al., 1998 and App. PAR, U.S. DOE, 1996, for additional information and a discussion of all 75 variables included in the LHS) 13

2.3 Stepwise Regression Analyses with Raw and Rank-Transformed Data with Pooled Results from Replicates R1, R2 and R3 (i.e., for a total of 300 observations) for Variables *E0:WAS_PRES*, *E0:BRAALIC*, *E2:WAS_SATB* and *E2:WAS_PRES* at 10,000 yr 18

10.1 Comparison of Variable Rankings with Different Analysis Procedures for y = *E0:WAS_PRES*, the Variables in Table 2.2, and a Maximum of Five Classes of Values for Each Variable (i.e., $nX = 5$) 54
10.2 Comparison of Variable Rankings with Correlation Coefficients, Standardized Regression Coefficients and Partial Correlation Coefficients with Raw and Rank Transformed Data for y = *E0:WAS_PRES* 55
10.3 Correlations with Raw and Rank Transformed Data between *WMICDFLG*, *HALPOR*, *WGRCOR*, *ANHPRM* and *SALPRES* 56
10.4 Top-Down Correlation Matrix for Variable Rankings with Different Analysis Procedures for y = *E0:WAS_PRES*, Variables included in Table 10.1, and a Maximum of Five Classes of x values (i.e., $nX = 5$) 57
10.5 Comparison of Variable Rankings for y = *E0:WAS_PRES* Obtained with a Maximum of Five Classes of x Values (i.e., $nX = 5$) and Analytic Determination of p-values with Variable Rankings Obtained with (i) a Maximum of Ten Classes of x values (i.e., $nX = 10$) and Analytic Determination of p-values and (ii) a Maximum of Five Classes of x values (i.e., $nX = 5$) and Monte Carlo Determination of p-values 58
10.6 Comparison of Variable Rankings for y = *E0:WAS_PRES* Obtained with Correlation Coefficients (CCs, RCCs) and Analytic Determination of p-values with Rankings Obtained with Monte Carlo Determination of p-values 61
10.7 Exceedance Probabilities (i.e., p-values) for Common Mean and Common Variance Tests Calculated with Use of Logarithms for y = *E0:WAS_PRES*, the Variables in Table 2.2, and a Maximum of Five Classes of Values for Each Variable (i.e., $nX = 5$) 62
10.8 Comparison of Variable Rankings with Different Analysis Procedures for y = *E0:BRAALIC*, the Variables in Table 2.2, and a Maximum of Five Classes of x Values (i.e., $nX = 5$) 63
10.9 Comparison of Variable Rankings with Correlation Coefficients, Standardized Regression Coefficients and Partial Correlation Coefficients with Raw and Rank Transformed Data for y = *E0:BRAALIC* 65
10.10 Top Down Correlation Matrix for Variable Rankings with Different Analysis Procedures for y = *E0:BRAALIC*, Variables included in Table 10.8, and a Maximum of Five Classes of x values (i.e., $nX = 5$) 66
10.11 Comparison of Variable Rankings for y = *E0:BRAALIC* Obtained with a Maximum of Five Classes of x Values (i.e., $nX = 5$) and Analytic Determination of p-values with Variable Rankings Obtained with (i) a Maximum of Ten Classes of x values (i.e., $nX = 10$) and Analytic Determination of p-values and (ii) a Maximum of Five Classes of x values (i.e., $nX = 5$) and Monte Carlo Determinatin of p-values 67
10.12 Comparison of Variable Rankings for y = *E0:BRAALIC* Obtained with Correlation Coefficients (CCs, RCCs) and Analytic Determination of p-values with Rankings Obtained with Monte Carlo Determination of p-values 69
10.13 Exceedance Probabilities (i.e., p-values) for Common Mean and Common Variance Tests Calculated with Use of Logarithms for y = *E0:BRAALIC*, the variables in Table 2.2, and a Maximum of Five Classes of Values for Each Variable (i.e., $nX = 5$) 69
10.14 Comparison of Variable Rankings with Different Analysis Procedures for y = *E2:WAS_SATB*, the Variables in Table 2.2, and a Maximum of Five Classes of x Values (i.e., $nX = 5$) 71
10.15 Comparison of Variable Rankings with Correlation Coefficients, Standardized Regression Coefficients and Partial Correlation Coefficients with Raw and Rank Transformed Data for y = *E2:WAS_SATB* 72
10.16 Top Down Correlation Matrix for Variable Rankings with Different Analysis Procedures for y = *E2:WAS_SATB*, Variables included in Table 10.14, and a Maximum of Five Classes of x values (i.e., $nX = 5$) 73

10.17 Comparison of Variable Rankings for y = *E2:WAS_SATB* Obtained with a Maximum of Five Classes of x Values (i.e., $nX = 5$) and Analytic Determination of p-values with Variable Rankings Obtained with (i) a Maximum of Ten Classes of x values (i.e., $nX = 10$) and Analytic Determination of p-values and (ii) a Maximum of Five Classes of x values (i.e., $nX = 5$) and Monte Carlo Determination of p-values 74
10.18 Comparison of Variable Rankings for y = *E2:WAS_SATB* Obtained with Correlation Coefficients (CCs, RCCs) and Analytic Determination of p-values with Rankings Obtained with Monte Carlo Determination of p-values 76
10.19 Exceedance Probabilities (i.e., p-values) for Common Mean and Common Variance Tests Calculated with Use of Logarithms for y = *E2:WAS_SATB*, the Variables in Table 2.2, and a Maximum of Five Classes of Values for Each Variable (i.e., $nX = 5$) 77
10.20 Comparison of Variable Rankings with Different Analysis Procedures for y = *E2:WAS_PRES*, the Variables in Table 2.2, and a Maximum of Five Classes of Values for each Variable (i.e., $nX = 5$) 78
10.21 Comparison of Variable Rankings with Correlation Coefficients, Standardized Regression Coefficients and Partial Correlation Coefficients with Raw and Rank Transformed Data for y = *E2:WAS_PRES* 79
10.22 Top Down Correlation Matrix for Variable Rankings with Different Analysis Procedures for y = *E2:WAS_PRES*, Variables included in Table 10.20, and a Maximum of Five Classes of x values (i.e., $nX = 5$) 80
10.23 Comparison of Variable Rankings for y = *E2:WAS_PRES* Obtained with a Maximum of Five classes of x Values (i.e., $nX = 5$) and Analytic Determination of p-values with Variable Rankings Obtained with (i) a Maximum of Ten Classes of x values (i.e., $nX = 10$) and Analytic Determination of p-values and (ii) a Maximum of Five Classes of x values (i.e., $nX = 5$) and Monte Carlo Determination of p-values 81
10.24 Comparison of Variable Rankings for y = *E2:WAS_PRES* Obtained with Correlation Coefficients (CCs, RCCs) and Analytic Determination of p-values with Rankings Obtained with Monte Carlo Determination of p-values 83
10.25 Exceedance Probabilities (i.e., p-values) for Common Mean and Common Variance Tests Calculated with use of Logarithms for y = *E2:WAS_PRES*, the Variables in Table 2.2, and a Maximum of Five Classes of Values for Each Variable (i.e., $nX = 5$) 84

12.1 Comparison of Variable Rankings Obtained with Different Analysis Procedures for Three Independent Samples of Size 100 (Columns AP:R1, AP:R2, AP:R3, where AM ~ CC, RCC, CMN, CL, CMD, CV, CIQ, SI as appropriate), Pooled Sample of Size 300 (Column AP:All), and a Maximum of Five Classes of Values for Each Variable (i.e., $nX = 5$) 92
12.2 Comparison of Variable Rankings Obtained with Common Variances Test with Use of Logarithms for Three Independent Samples of Size 100 (Column CV=R1, CV:R2, CV:R3) and Pooled Sample of Size 300 (Column CV:All) for 1=*EO:BRAALIC* 95

1. Introduction

Sensitivity analysis is now widely recognized as an essential component of studies based on mathematical modeling (e.g., NRC, 1983, 1993, 1994; IAEA, 1989; U.S. EPA, 1993; NCRP, 1996). Here, sensitivity analysis refers to the determination of the effects of uncertain model inputs on model predictions. A number of methods have been proposed for sensitivity analysis, including differential analysis, response surface methodologies, Monte Carlo techniques, and the Fourier amplitude sensitivity test (Iman and Helton, 1988; Helton, 1993; Hamby, 1994).

Monte Carlo techniques probably constitute the most widely used approach to sensitivity analysis due to their flexibility, ease of implementation, and conceptual simplicity. When viewed abstractly, a Monte Carlo sensitivity study involves a vector

$$\mathbf{x} = [x_1, x_2, \ldots, x_{nI}] \tag{1.1}$$

of uncertain model inputs, where each x_i is an uncertain input and nI is the number of such inputs, and a vector

$$\mathbf{y} = f(\mathbf{x}) = [y_1, y_2, \ldots, y_{nO}] \tag{1.2}$$

of model predictions, where f is a function used to represent the model under consideration, each y_j is an outcome of evaluating the model with the input $\mathbf{x}$, and nO is the number of such outcomes. Distributions

$$D_i,\ i = 1, 2, \ldots, nI, \tag{1.3}$$

are used to characterize the uncertainty in each input x_i, where D_i is the distribution assigned to x_i. Correlations and other relationships between the x_i are also possible.

A sampling procedure such as simple random sampling or Latin hypercube sampling (McKay et al. 1979) is used to generate a sample

$$\mathbf{x}_k = [x_{1k}, x_{2k}, \ldots, x_{nI,k}],\ k = 1, 2, \ldots, nS, \tag{1.4}$$

from the population of $\mathbf{x}$'s with the distributions in Eq. (1.3), where nS is the size of the sample. Evaluation of the model under consideration with the sample elements $\mathbf{x}_k$ in Eq. (1.4) then creates a sequence of results of the form

$$\mathbf{y}_k = f(\mathbf{x}_k) = [y_{1k}, y_{2k}, \ldots, y_{n0,k}],\ k = 1, 2, \ldots, nS, \tag{1.5}$$

where y_{jk} is the outcome of evaluating the model with $\mathbf{x}_k$. The pairs

$$(\mathbf{x}_k, \mathbf{y}_k),\ k = 1, 2, \ldots, nS, \tag{1.6}$$

constitute a mapping from model input, $\mathbf{x}_k$, to model output, $\mathbf{y}_k$, that can be explored with various sensitivity analysis techniques to determine how the individual analysis inputs contained in $\mathbf{x}$ (i.e., the x_i's) affect the individual analysis outcomes contained in $\mathbf{y}$ (i.e., the y_j's). Analysis possibilities include regression analysis, correlation analysis, and examination of scatterplots (Iman et al., 1981a, 1981b; Saltelli and Marivoet, 1990; Saltelli and Homma, 1992; Saltelli et al., 1993; Helton, 1993; Hamby, 1994, 1995; Kleijnen, 1998).

Although techniques based on regression analysis and correlation analysis are often successful in identifying the relationships between model input and output embedded in the mapping in Eq. (1.6), in many cases these techniques fail to identify well-defined, but nonlinear, relationships (Iman and Helton, 1988; Saltelli and Homma, 1992; Saltelli et al., 1993; Helton et al., 1998). If the underlying relationship is nonlinear but monotonic, then a rank transformation will linearize the relationship and result in successful sensitivity analyses with regression-based techniques (Iman and Conover, 1979). However, the underlying relationship can be too complex to be linearized in any simple manner. In these cases, sensitivity analysis techniques are needed that can identify patterns in the mapping in Eq. (1.6) without recourse to comparisons with specialized prespecified relationships (e.g., linear or monotonic). The ultimate test of whether or not there is a relationship between an input variable x_i and an output variable y_j lies in determining whether or not the points

$$(x_{ik}, y_{jk}),\ k = 1, 2, \ldots, nS, \tag{1.7}$$

constitute a random pattern conditional on the marginal distributions for x_i and y_j. This presentation will investigate the implications from a sensitivity analysis perspective of a sequence of tests (i.e., hypotheses) for the relationship between x_i and y_j embedded in the mapping in Eq. (1.6). These hypotheses will run from very specific (i.e., a linear relationship) to quite general (i.e., a nonrandom pattern).

The presentation is organized as follows. Example simulation results that will be used to motivate and illustrate the sensitivity analysis procedures are presented in Sect. 2. Then, the procedures to be considered are summarized in Sects. 3-7. Specifically, the following five relationships are proposed as the basis for a sequence of sensitivity tests: (i) Linear relationship: $E(y|x) = \beta_0 + \beta_1 x$, where the subscripts have been dropped from y_j and x_i for notational simplicity (Sect. 3); (ii) Monotonic relation: $E[r(y)|r(x)] = \gamma_0 + \gamma_1 r(x)$, where $r(x)$ and $r(y)$ denote the ranks of x and y, respectively (Sect. 4); (iii) Location (central tendency) of y depends on x (Sect. 5); (iv) Variability (spread) of y depends on x (Sect. 6); and (v) y and x are statistically independent: $p(y|x) = p(y)$, where p denotes the density function for y (Sect. 7). Next, the ranking of variable importance and the use of the Iman and Conover (1987) top-down correlation procedure to compare variable rankings are discussed in Sects. 8 and 9. Then, example analysis results with the indicated procedures are presented in Sect. 10; Type I and Type II errors are discussed in Sect. 11; and the robustness of analysis outcomes for independent samples is discussed in Sect. 12. Finally, a concluding discussion is given in Sect. 13.

2. Test Problems

The test problems considered in this presentation use results obtained in the 1996 performance assessment (PA) for the Waste Isolation Pilot Plant (WIPP) (Helton et al., 1998), which was carried out to support the U.S. Department of Energy's (DOE's) application to the U.S. Environmental Protection Agency (EPA) for the certification of the WIPP for the disposal of transuranic waste (U.S. DOE, 1996). In particular, the test problems involve results (Table 2.1) calculated by the BRAGFLO model (Sect. 4.2, Helton et al., 1998), which was used to represent two phase (i.e., gas and brine) flow in the vicinity of the repository. The BRAGFLO model uses finite difference procedures (Fig. 2.1) to numerically solve a system of nonlinear partial differential equations (Eqs. 4.2.1 - 4.2.6, Helton et al., 1998) and requires a significant amount of computational resources (e.g., 4 to 5 hours of CPU time on a 275 MHz Digital Equipment Corp. (DEC) Alpha with VMS for a single model evaluation).

The 1996 WIPP PA used computational procedures based on Latin hypercube sampling to propagate the effects of subjective (i.e., epistemic) uncertainty through the analysis (Helton et al. 1998). As a result of guidance given by the EPA (U.S. EPA, 1996), the 1996 WIPP PA used a Latin hypercube sample (LHS) of size 300 (Sect. 6.3, Helton et al., 1998) from 75 uncertain variables, of which only 27 were used as inputs to the BRAGFLO model in the calculation of the dependent variables in Table 2.1 (Table 2.2). To provide a test of the robustness of the uncertainty propagation procedures used in the 1996 WIPP PA, the indicated LHS of size 300 was actually generated as 3 independent samples of size 100 each (Sect. 6.4, Helton et al., 1998). Each of these samples was generated with

Table 2.1. Definition of Dependent Variables Predicted by BRAGFLO Model for Two-Phase Flow Selected for Use in Comparison of Statistical Procedures for Identification of Patterns in Scatterplots

E0:WAS_PRES—Pressure (Pa) in lower repository waste panel (region 23, Fig. 2.1) at 10,000 yr under undisturbed (i.e., E0) conditions. Number of sampled variables: 26 (Table 2.2).

E0:BRAALIC—Cumulative brine inflow (m^3) to vicinity of repository over 10,000 yr from anhydrite marker beds (regions 20, 21, 28, Fig. 2.1) under undisturbed (i.e., E0) conditions. Same sampled variables as *E0:WAS_PRES.*

E2:WAS_SATB—Brine saturation (dimensionless) in lower repository waste panel (region 23, Fig. 2.1) at 10,000 yr after a drilling intrusion through the lower waste panel at 1000 yr that does not penetrate pressurized brine in the underlying Castile Formation (i.e., an E2 intrusion). Same sampled variables as *E0:WAS_PRES* plus *BHPRM* (Table 2.2).

E2:WAS_PRES—Pressure (Pa) in lower repository waste panel (region 23, Fig. 2.1) at 10,000 yr after a drilling intrusion through the lower waste panel at 1000 yr that does not penetrate pressurized brine in the underlying Castile Formation (i.e., an E2 intrusion). Same sampled variables as *E2:WAS_SATB.*

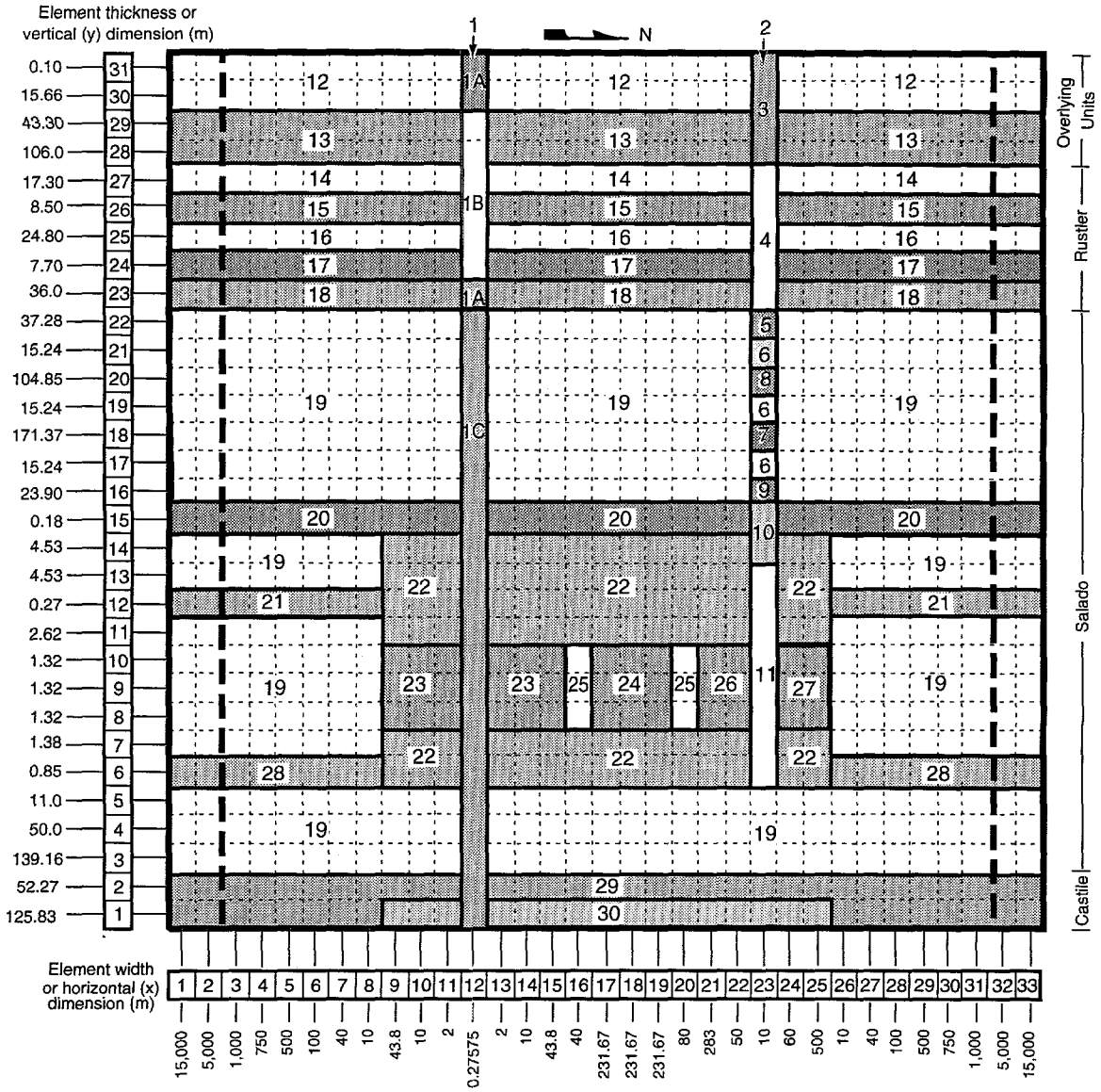

Fig. 2.1. Computational grid used in BRAGFLO to represent two-phase flow in 1996 WIPP CCA PA subsequent to a drilling intrusion. Same formulation is used in the absence of a drilling intrusion except that regions 1A, 1B and 1C have the same properties as the regions to either side.

use of the Iman and Conover restricted pairing technique (Iman and Conover, 1982; Iman and Shortencarier, 1984) to enforce specified correlations between three pairs of variables (the correlated pairs (*ANHCOMP*, *ANHPRM*) and (*HALCOMP*, *HALPRM*) are used in the calculation of the results in Table 2.1 and are described in Table 2.2) and also to ensure that uncorrelated variables had correlations close to zero. The outcome of this sampling was 3 LHSs of size 100 each:

$$R1\text{: } \mathbf{x}_{1k} = [\mathbf{x}_{1k1}, \mathbf{x}_{1k2}, \ldots, \mathbf{x}_{1k75}], \quad k = 1, 2, \ldots, 100 \tag{2.1}$$

$$R2\text{: } \mathbf{x}_{2k} = [\mathbf{x}_{2k1}, \mathbf{x}_{2k2}, \ldots, \mathbf{x}_{2k75}], \quad k = 1, 2, \ldots, 100 \tag{2.2}$$

$$R3\text{: } \mathbf{x}_{3k} = [\mathbf{x}_{3k1}, \mathbf{x}_{3k2}, \ldots, \mathbf{x}_{3k75}], \quad k = 1, 2, \ldots, 100, \tag{2.3}$$

where $\mathbf{x} = [x_1, x_2, \ldots, x_{75}]$ corresponds to the 75 uncertain variables indicated in Table 2.2 and $R1$, $R2$ and $R3$ designate the three replicated (i.e., independently generated) LHSs.

Table 2.2. Uncertain Variables Used as Input to BRAGFLO in the Calculation of the Dependent Variables in Table 2.1 (see Table 5.2.1, Helton et al., 1998, and App. PAR, U.S. DOE, 1996, for additional information and a discussion of all 75 variables included in the LHS)

ANHBCEXP—Brooks-Corey pore distribution parameter for anhydrite (dimensionless). Distribution: Student's with 5 degrees of freedom. Range: 0.491 to 0.842. Mean, Median: 0.644.

ANHBCVGP—Pointer variable for selection of relative permeability model for use in anhydrite. Distribution: Discrete with 60% 0, 40% 1. Value of 0 implies Brooks-Corey model; value of 1 implies van Genuchten-Parker model.

ANHCOMP—Bulk compressibility of anhydrite (Pa^{-1}). Distribution: Student's with 3 degrees of freedom. Range: 1.09×10^{-11} to 2.75×10^{-10} Pa^{-1}. Mean, Median: 8.26×10^{-11} Pa^{-1}. Correlation: –0.99 rank correlation (Iman and Conover, 1982) with *ANHPRM*. Variable 21 in LHS.

ANHPRM—Logarithm of anhydrite permeability (m^2). Distribution: Student's with 5 degrees of freedom. Range: –21.0 to –17.1 (i.e., permeability range is 1×10^{-21} to $1 \times 10^{-17.1}$ m^2). Mean, Median: –18.9. Correlation : –0.99 rank correlation with *ANHCOMP*.

ANRBRSAT—Residual brine saturation in anhydrite (dimensionless). Distribution: Student's with 5 degrees of freedom. Range: 7.85×10^{-3} to 1.74×10^{-1}. Mean, Median: 8.36×10^{-2}.

ANRGSSAT—Residual gas saturation in anhydrite (dimensionless). Distribution: Student's with 5 degrees of freedom. Range: 1.39×10^{-2} to 1.79×10^{-1}. Mean, median: 7.71×10^{-2}.

BHPRM—Logarithm of borehole permeability (m^2). Distribution: Uniform. Range: –14 to –11 (i.e., permeability range is 1×10^{-14} to 1×10^{-11} m^2). Mean, median: –12.5.

Table 2.2. (continued)

HALCOMP—Bulk compressibility of halite (Pa^{-1}). Distribution: Uniform. Range: 2.94×10^{-12} to 1.92×10^{-10} PA^{-1}. Mean, median: 9.75×10^{-11} Pa^{-1}, 9.75×10^{-11} Pa^{-1}. Correlation: –0.99 rank correlation with *HALPRM*.

HALPOR—Halite porosity (dimensionless). Distribution: Piecewise uniform. Range: 1.0×10^{-3} to 3×10^{-2}. Mean, median: 1.28×10^{-2}, 1.00×10^{-2}.

HALPRM—Logarithm of halite permeability (m^2). Distribution: Uniform. Range: –24 to –21 (i.e., permeability range is 1×10^{-24} to 1×10^{-21} m^2). Mean, median: –22.5, –22.5. Correlation: –0.99 rank correlation with *HALCOMP*.

SALPRES—Initial brine pressure, without the repository being present, at a reference point located in the center of the combined shafts at the elevation of the midpoint of MB 139 (Pa). Distribution: Uniform. Range: 1.104×10^7 to 1.389×10^7 Pa. Mean, median: 1.247×10^7 Pa, 1.247×10^7 Pa.

SHBCEXP—Brooks-Corey pore distribution parameter for shaft (dimensionless). Distribution: Piecewise uniform. Range: 0.11 to 8.10. Mean, median: 2.52, 0.94.

SHPRMASP—Logarithm of permeability (m^2) of asphalt component of shaft seal (m^2). Distribution: Triangular. Range: –21 to –18 (i.e., permeability range is 1×10^{-21} to 1×10^{-18} m^2). Mean, mode: –19.7, –20.0.

SHPRMCLY—Logarithm of permeability (m^2) for clay components of shaft. Distribution: Triangular. Range: –21 to –17.3 (i.e., permeability range is 1×10^{-21} to $1 \times 10^{-17.3}$ m^2). Mean, mode: –18.9, –18.3.

SHPRMCON—Same as *SHPRMASP* but for concrete component of shaft seal for 0 to 400 yr. Distribution: Triangular. Range: –17.0 to –14.0 (i.e., permeability range is 1×10^{-17} to 1×10^{-14} m^2). Mean, mode: –15.3, –15.0.

SHPRMDRZ—Logarithm of permeability (m^2) of DRZ surrounding shaft. Distribution: Triangular. Range: –17.0 to –14.0 (i.e., permeability range is 1×10^{-17} to 1×10^{-14} m^2). Mean, mode: –15.3, –15.0.

SHPRMHAL—Pointer variable (dimensionless) used to select permeability in crushed salt component of shaft seal at different times. Distribution: Uniform. Range: 0 to 1. Mean, mode: 0.5, 0.5. A distribution of permeability (m^2) in the crushed salt component of the shaft seal is defined for each of the following time intervals: [0, 10 yr], [10, 25 yr], [25, 50 yr], [50, 100 yr], [100, 200 yr], [200, 10000 yr]. *SHPRMHAL* is used to select a permeability value from the cumulative distribution function for permeability for each of the preceding time intervals with result that a rank correlation of 1 exists between the permeabilities used for the individual time intervals.

SHRBRSAT—Residual brine saturation in shaft (dimensionless). Distribution: Uniform. Range: 0 to 0.4. Mean, median: 0.2, 0.2.

SHRGSSAT—Residual gas saturation in shaft (dimensionless). Distribution: Uniform. Range: 0 to 0.4. Mean, median: 0.2, 0.2.

Table 2.2. (continued)

WASTWICK—Increase in brine saturation of waste due to capillary forces (dimensionless). Distribution: Uniform. Range: 0 to 1. Mean, median: 0.5, 0.5.

WFBETCEL—Scale factor used in definition of stoichiometric coefficient for microbial gas generation (dimensionless). Distribution: Uniform. Range: 0 to 1. Mean, median: 0.5, 0.5.

WGRCOR—Corrosion rate for steel under inundated conditions in the absence of CO_2 (m/s). Distribution: Uniform. Range: 0 to 1.58×10^{-14} m/s. Mean, median: 7.94×10^{-15} m/s, 7.94×10^{-15} m/s.

WGRMICH—Microbial degradation rate for cellulose under humid conditions (mol/kg•s). Distribution: Uniform. Range: 0 to 1.27×10^{-9} mol/kg•s. Mean, median: 6.34×10^{-10} mol/kg•s, 6.34×10^{-10} mol/kg•s.

WGRMICI—Microbial degradation rate for cellulose under inundated conditions (mol/kg•s). Distribution: Uniform. Range: 3.17×10^{-10} to 9.51×10^{-9} mol/kg•s. Mean, median: 4.92×10^{-9} mol/kg•s, 4.92×10^{-9} mol/kg•s.

WMICDFLG—Pointer variable for microbial degradation of cellulose. Distribution: Discrete, with 50% 0, 25% 1, 25% 2. *WMICDFLG* = 0, 1, 2 implies no microbial degradation of cellulose, microbial degradation of only cellulose, microbial degradation of cellulose, plastic and rubber.

WRBRNSAT—Residual brine saturation in waste (dimensionless). Distribution: Uniform. Range: 0 to 0.552. Mean, median: 0.276, 0.276.

WRGSSAT—Residual gas saturation in waste (dimensionless). Distribution: Uniform. Range: 0 to 0.15. Mean, median: 0.075, 0.075.

Once the LHSs in Eqs. (2.1) - (2.3) were generated, BRAGFLO calculations were performed for a variety of cases (Table 6.9.1, Helton et al., 1998). The two cases considered in this presentation are undisturbed (i.e., E0) conditions and a drilling intrusion through the lower waste panel at 1000 yr that does not penetrate pressurized brine in the underlying Castile Formation (i.e., E2 conditions or, in the more detailed descriptions given in Helton et al., 1998, E2 conditions with the intrusion occurring at 1000 yr). Results calculated by BRAGFLO are time-dependent. The time-dependent behavior of the results indicated in Table 2.1 is shown in Fig. 2.2 for replicate *R*1. For simplicity, the technique comparisons contained in this presentation will use the values of the variables at the end points of the individual curves in Fig. 2.2 (i.e., at 10,000 yr). However, nothing prevents analyses at other times and, in general, sensitivity analyses of time-dependent variables should also be time-dependent (Chapts. 7, 8, Helton et al. 1998).

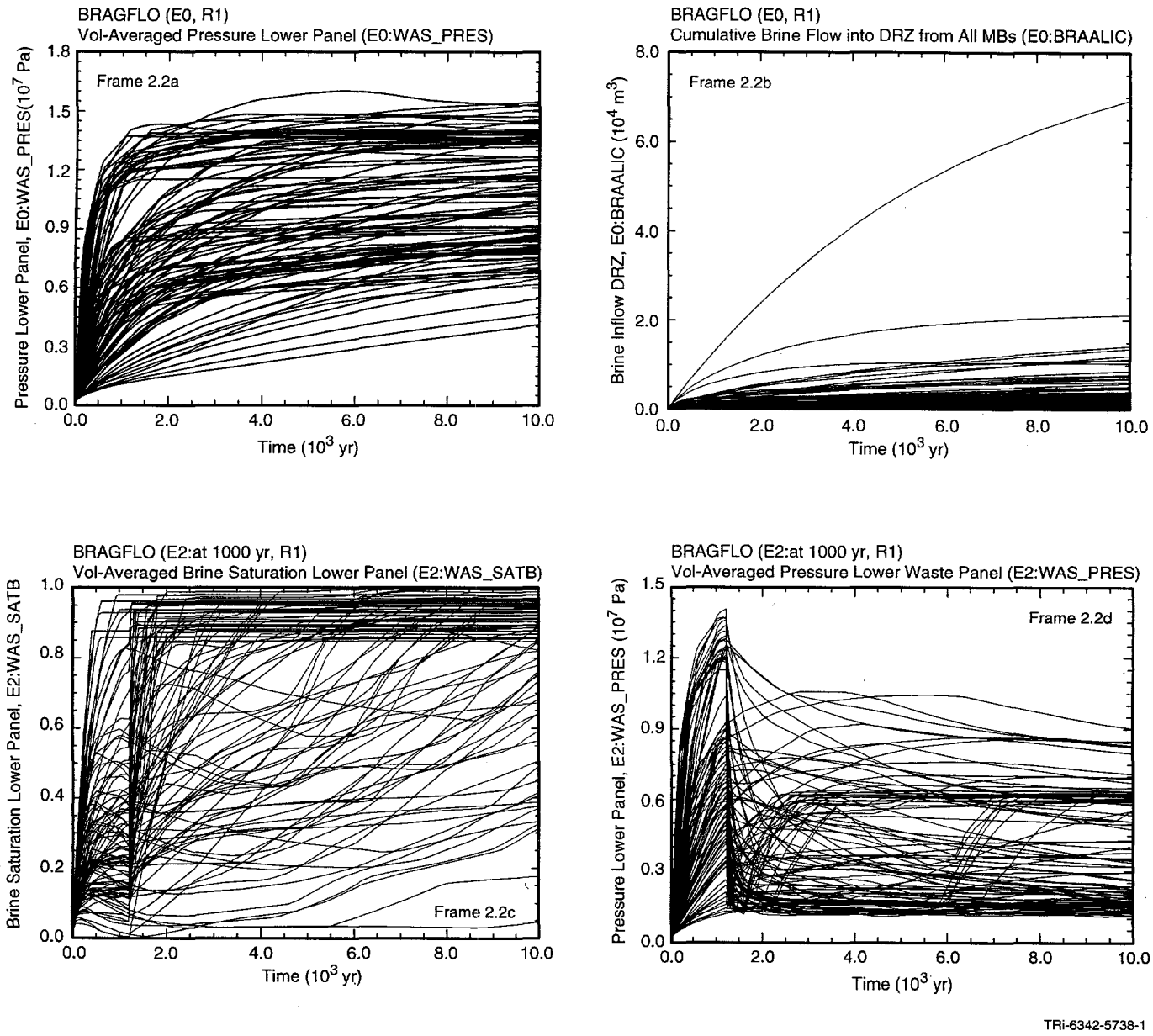

Fig. 2.2. Dependent variables predicted by BRAGFLO model for two-phase flow selected for use in comparison of statistical procedures for identification of patterns in scatterplots: (2.2a) pressure in lower waste panel under undisturbed conditions (*E0:WAS_PRES*), (2.2b) cumulative brine inflow from anhydrite marker beds under undisturbed conditions (*E0:BRAALIC*), (2.2c) saturation in lower waste panel after an E2 intrusion at 1000 yr (*E2:WAS_SATB*), and (2.2d) pressure in lower waste panel after an E2 intrusion at 1000 yr (*E2:WAS_PRES*).

A popular and often effective approach to sensitivity analysis in sampling-based studies is to use regression analysis to estimate the relationships that exist between uncertain (i.e., sampled) variables and predicted variables (Helton, 1993). Specifically, the sampling procedure and associated model evaluations lead to the mapping $(\mathbf{x}_k, \mathbf{y}_k)$, $k = 1, 2, \ldots, nS$, in Eq. (1.6) between sampled variables and model predictions that can be investigated with regression-based techniques.

For perspective and motivation, regression-based results for the variables in Table 2.1 obtained in Helton et al. (1998) with the STEP program (Iman et al., 1980) are presented in Table 2.3 for both raw and rank-transformed data. In the regressions in Table 2.3, a variable was required to be significant at an α-value of 0.02 to enter a regression model and to remain significant at an α-value of 0.05 to be retained in a regression model, although there were no cases of a variable entering and then being dropped from a regression model. As will be seen, the rank-transformation is often an effective procedure for improving the resolution of regression-based sensitivity analyses due to its effect in linearizing nonlinear but monotonic relationships. However, as will also be seen, nonmonotonic relationships can result in patterns that cannot be effectively analyzed with rank-transformed data. It is the need to be able to identify such patterns that forms the motivation for this study.

The analyses in Table 2.3 for repository pressure at under undisturbed conditions (*E0:WAS_PRES*) with raw and rank-transformed data are reasonably effective, with (1) R^2 values of 0.82 and 0.81 for raw and rank-transformed data, (2) the same variables selected in both analyses, and (3) only one minor variation in the order of variable selection (i.e., the order of selection of the last two variables in the regression models is reversed). Scatterplots for the first four variables selected in the regression analyses for *E0:WAS_PRES* are presented in Fig. 2.3. The scatterplots for the first two variables selected in the regression analysis, *WMICDFLG* and *HALPOR*, display well-defined patterns. The pattern for the third variable, *WGRCOR*, is weaker but still detectable. The fourth variable, *ANHPRM*, changes the R^2 values for raw and rank-transformed data by 0.02 and 0.01, respectively, and produces a scatterplot that displays little discernible pattern.

The analyses for cumulative brine inflow from all anhydrite marker beds to the repository under undisturbed conditions (*E0:BRAALIC*) are interesting in that the regression with raw data is not particularly effective (i.e., $R^2 = 0.50$ at final step of analysis), while the regression with rank-transformed data is reasonably successful in accounting for the observed uncertainty (i.e., $R^2 = 0.87$). Again, examination of scatterplots shows well-defined patterns for the first two variables, *WMICDFLG* and *ANHPRM*, selected in both regression analyses (Fig. 2.4). Scatterplots for the next two variables, *HALPOR* and *WGRCOR*, selected in the regression analysis with rank-transformed data are also given in Fig. 2.4. The negative effects of these variables, as indicated by the signs of their standardized regression coefficients, are barely discernible in their scatterplots, with these small effects being consistent with observed changes in R^2 values of 0.05 and 0.02 with the entry of *HALPOR* and *WGRCOR*, respectively, into the regression model. In this example, the regression analyses with both raw and rank-transformed data have identified the two

Table 2.3. Stepwise Regression Analyses with Raw and Rank-Transformed Data with Pooled Results from Replicates R1, R2 and R3 (i.e., for a total of 300 observations) for Variables *E0:WAS_PRES*, *E0:BRAALIC*, *E2:WAS_SATB* and *E2:WAS_PRES* at 10,000 yr

	Raw Data, *E0:WAS_PRES*			Rank-Transformed Data, *E0:WAS_PRES*		
Step[a]	Variable[b]	SRC[c]	R^{2}[d]	Variable[b]	SRRC[e]	R^{2}[d]
1	*WMICDFLG*	0.72	0.51	*WMICDFLG*	0.71	0.52
2	*HALPOR*	0.47	0.73	*HALPOR*	0.45	0.73
3	*WGRCOR*	0.25	0.79	*WGRCOR*	0.23	0.79
4	*ANHPRM*	0.13	0.81	*ANHPRM*	0.11	0.80
5	*SHRGSSAT*	0.07	0.81	*SALPRES*	0.07	0.80
6	*SALPRES*	0.06	0.82	*SHRGSSAT*	0.06	0.81

	Raw Data, *E0:BRAALIC*			Rank-Transformed Data, *E0:BRAALIC*		
Step	Variable	SRC	R^2	Variable	SRRC	R^2
1	*ANHPRM*	0.56	0.32	*WMICDFLG*	−0.66	0.43
2	*WMICDFLG*	−0.31	0.42	*ANHPRM*	0.59	0.75
3	*WGRCOR*	−0.16	0.45	*HALPOR*	−0.16	0.80
4	*WASTWICK*	−0.15	0.47	*WGRCOR*	−0.15	0.82
5	*ANHBCEXP*	−0.12	0.49	*HALPRM*	0.14	0.85
6	*HALPOR*	−0.10	0.50	*SALPRES*	0.12	0.86
7				*WASTWICK*	−0.10	0.87

	Raw Data, *E2:WAS_SATB*			Rank-Transformed Data, *E2:WAS_SATB*		
Step	Variable	SRC	R^2	Variable	SRRC	R^2
1	*BHPRM*	0.37	0.12	*BHPRM*	0.59	0.36
2	*ANHPRM*	0.30	0.21	*WRGSSAT*	−0.40	0.52
3	*HALPOR*	0.21	0.25	*ANHPRM*	0.23	0.57
4	*WGRCOR*	−0.19	0.29	*HALPOR*	0.13	0.59
5	*WRGSSAT*	−0.15	0.31	*SHPRMHAL*	−0.12	0.60
6	*WMICDFLG*	−0.14	0.33	*WGRCOR*	−0.10	0.61

	Raw Data, *E2:WAS_PRES*			Rank-Transformed Data, *E2:WAS_PRES*		
Step	Variable	SRC	R^2	Variable	SRRC	R^2
1	*HALPRM*	0.37	0.14	*HALPRM*	0.36	0.13
2	*ANHPRM*	0.24	0.20	*ANHPRM*	0.24	0.19
3	*HALPOR*	0.14	0.22	*HALPOR*	0.14	0.20

[a] Steps in stepwise regression analysis.
[b] Variables listed in order of selection in regression analysis with *ANHCOMP* and *HALCOMP* excluded from entry into regression model because of −0.99 rank correlation within the pairs (*ANHPRM*, *ANHCOMP*) and (*HALPRM*, *HALCOMP*).
[c] Standardized regression coefficients (SRCs) in final regression model.
[d] Cumulative R^2 value with entry of each variable into regression model.
[e] Standardized rank regression coefficients (SRRCs) in final regression model.

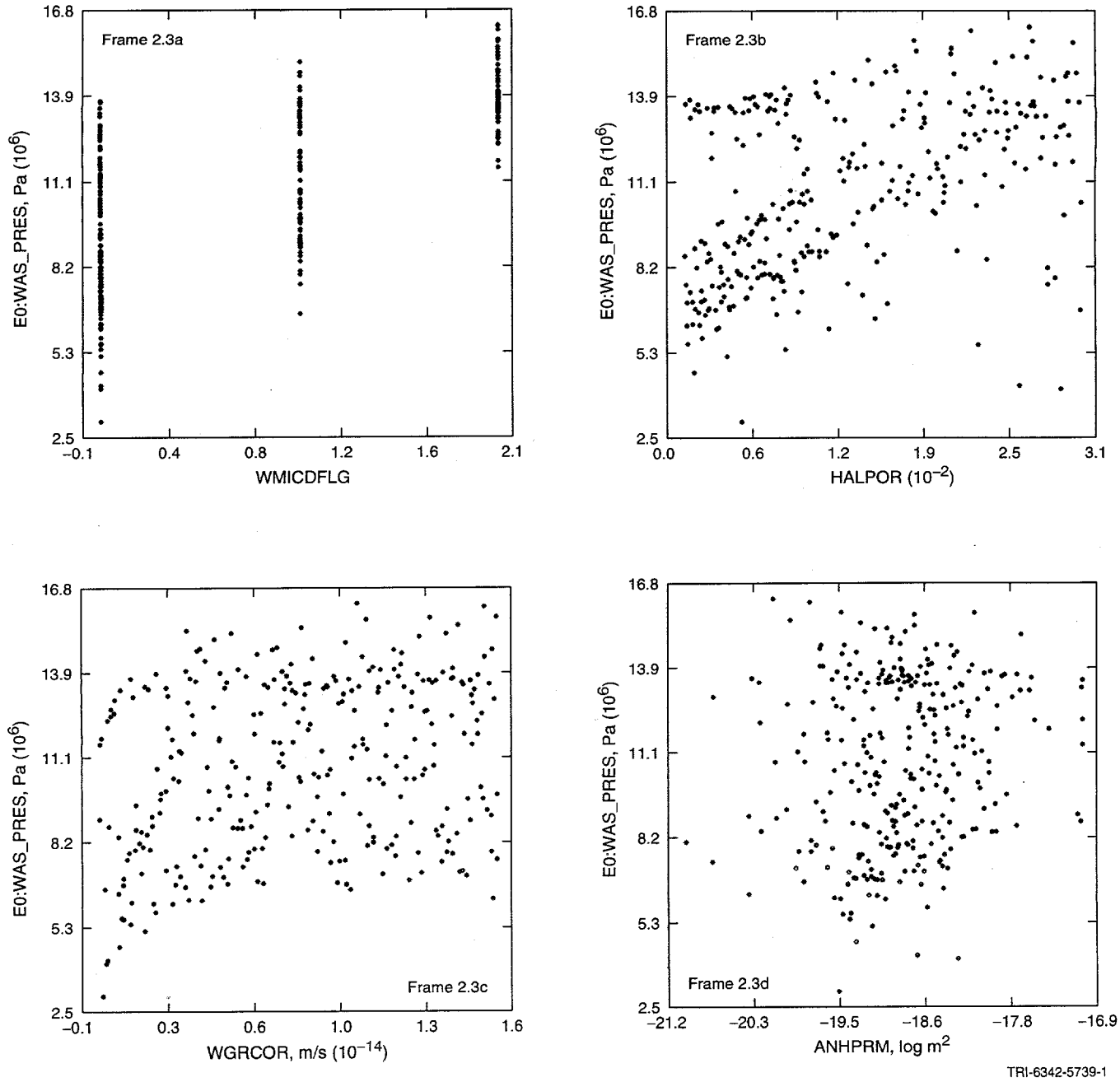

Fig. 2.3. Scatterplots for pressure in lower waste panel under undisturbed (i.e., E0) conditions at 10,000 yr (*E0:WAS_PRES*) versus first four variables selected in stepwise regression analyses with raw and rank-transformed data (Table 2.3): (2.3a) *WMICDFLG*, (2.3b) *HALPOR*, (2.3c) *WGRCOR*, and (2.3d) *ANHPRM*.

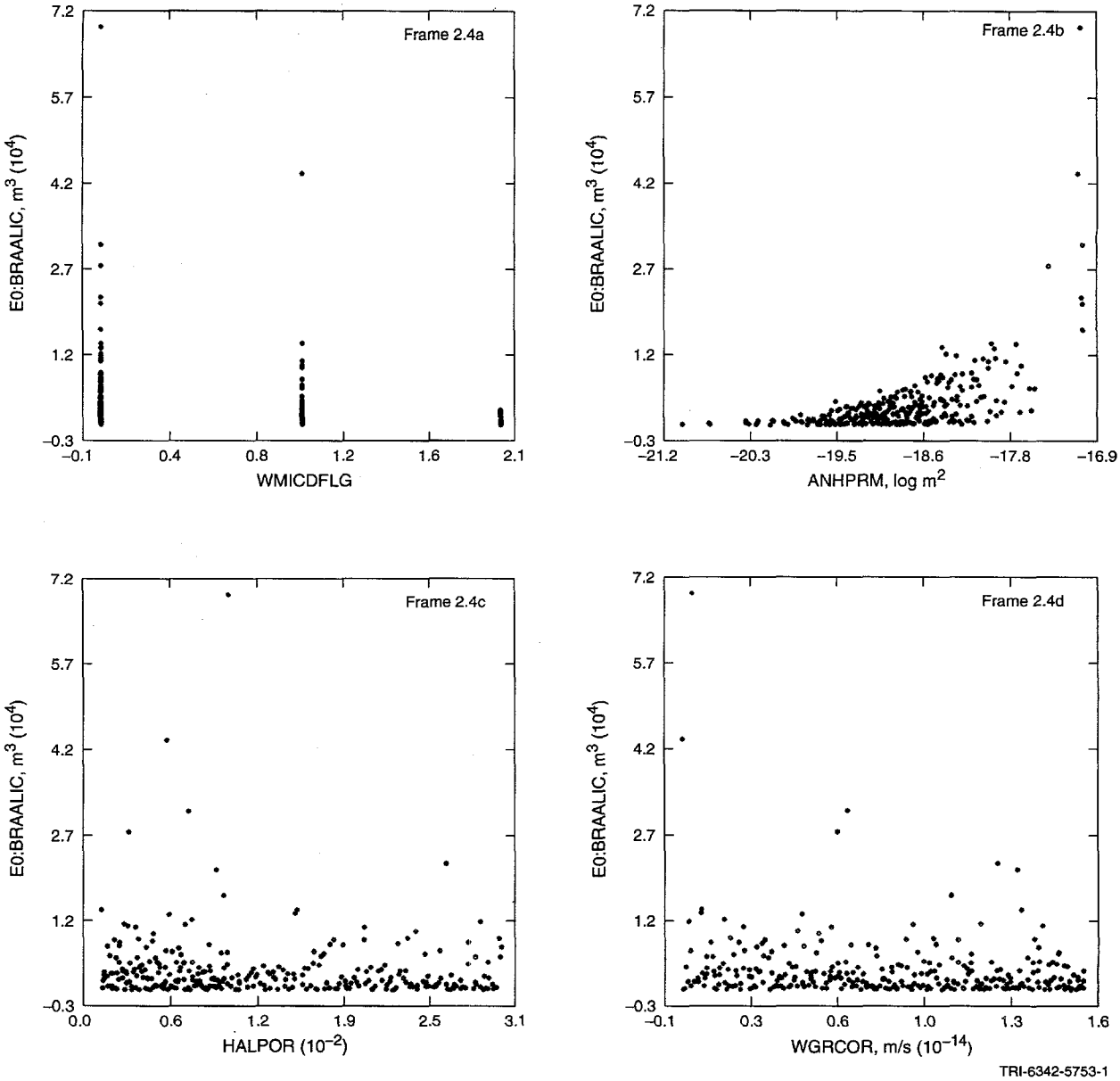

Fig. 2.4. Scatterplots for cumulative brine inflow over 10,000 yr from all anhydrite marker beds to repository under undisturbed (i.e., E0) conditions (*E0:BRAALIC*) versus first four variables selected in stepwise regression analysis with rank-transformed data (Table 2.3): (2.4a) *WMICDFLG*, (2.4b) *ANHPRM*, (2.4c) *HALPOR*, and (2.4d) *WGRCOR*.

dominant variables, *WMICDFLG* and *ANHPRM*. However, the analysis with raw data in isolation would not be very credible due to its low R^2 value.

The regression analysis with raw data for brine saturation in the lower waste panel after an E2 intrusion (*E2:WAS_SATB*) is quite poor, with the final regression model containing 6 variables but having an R^2 value of only 0.33. The regression analysis with rank-transformed data does somewhat better and results in a final regression model with 6 variables and an R^2 value of 0.61. However, an R^2 value of 0.61 is not particularly reassuring with respect to whether or not all the variables giving rise to the observed uncertainty in *E2:WAS_SATB* have been identified. Additional insights on the relationships between the sampled variables and *E2:WAS_SATB* can be obtained by examining scatterplots (Fig. 2.5). The first two variables identified in the regression analysis with rank-transformed data, *BHPRM* and *WRGSSAT*, show well-defined and interacting patterns. In particular, *BHPRM* is the primary determinant of whether or not a high value for *E2:WAS_SATB* occurs; however, given that a high value for *E2:WAS_SATB* occurs, this value is almost completely determined by *WRGSSAT*. Despite the well-defined patterns involving *BHPRM* and *WRGSSAT*, the regression analysis with raw data results in incremental R^2 values of only 0.12 and 0.02 for these two variables, and the regression analysis with rank-transformed data results in incremental R^2 values of only 0.36 and 0.16. The next two variables selected in the regression analysis with rank-transformed data are *ANHPRM* and *HALPOR*. The scatterplot plots for these variables do not show particularly strong patterns, with a stronger pattern actually being shown for the fourth-selected variable, *HALPOR*, than for the third-selected variable, *ANHPRM*. For *E2:WAS_SATB*, the two dominant variables, *BHPRM* and *WRGSSAT*, appear in the regression analyses for both raw and rank-transformed data. However, the R^2 values associated with these regressions (i.e., 0.33 and 0.61) provide little assurance that the dominant variables have been identified. It is only after examination of the associated scatterplots and the development of a physical explanation for the patterns appearing in these plots that some degree of comfort emerges that the dominant variables have indeed been identified.

The final regressions in Table 2.3 are for pressure in the lower waste panel after an E2 intrusion (*E2:WAS_PRES*). The regression analyses with both raw and rank-transformed data perform very poorly and result in final regression models with R^2 values of only 0.22 and 0.20, respectively. Both regression models select *HALPRM*, *ANHPRM* and *HALPOR*, with the scatterplots for these three variables appearing in Fig. 2.6. Examination of these scatterplots does not adequately reveal what is giving rise to the observed uncertainty in *E2:WAS_PRES*. In particular, the uncertainty in *E2:WAS_PRES* does not appear to arise from either *HALPRM*, *ANHPRM* and *HALPOR* individually or from some form of interaction between these variables. At this point in the analysis reported in Helton et al. (1998), a systematic search was made through the scatterplots between *E2:WAS_PRES* and the remaining variables in Table 2.2, with this search revealing that the uncertainty in *E2:WAS_PRES* is dominated by *BHPRM* (Fig. 2.6d). This is disconcerting because the clearly dominant variable

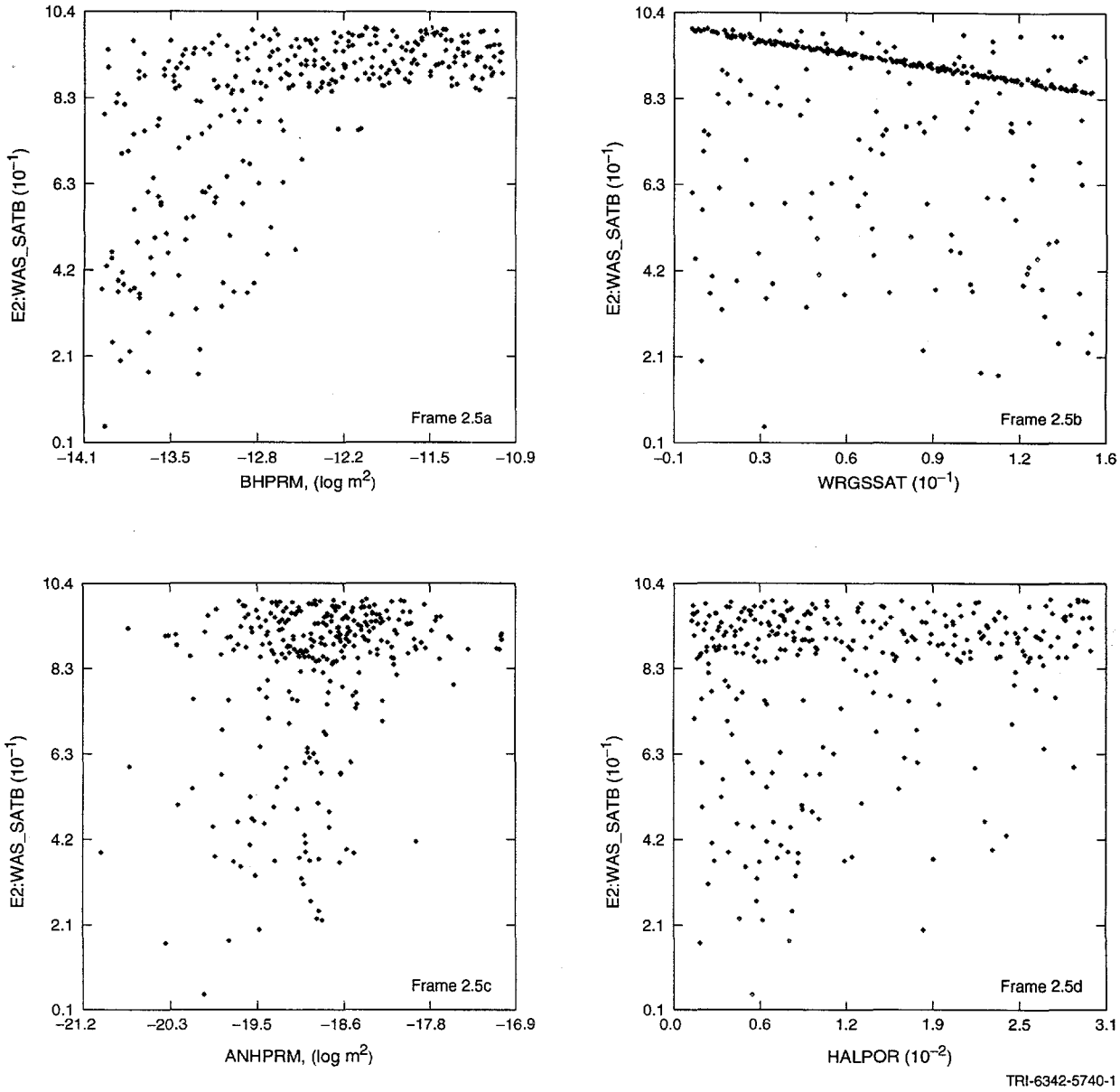

Fig. 2.5. Scatterplots for brine saturation in lower (i.e., intruded) waste panel at 10,000 yr for an E2 intrusion at 1000 yr (*E2:WAS_SATB*) versus first four variables selected in stepwise regression analysis with rank-transformed data (Table 2.3): (2.5a) *BHPRM*, (2.5b) *WRGSSAT*, (2.5c) *ANHPRM* and (2.5d) *HALPOR*.

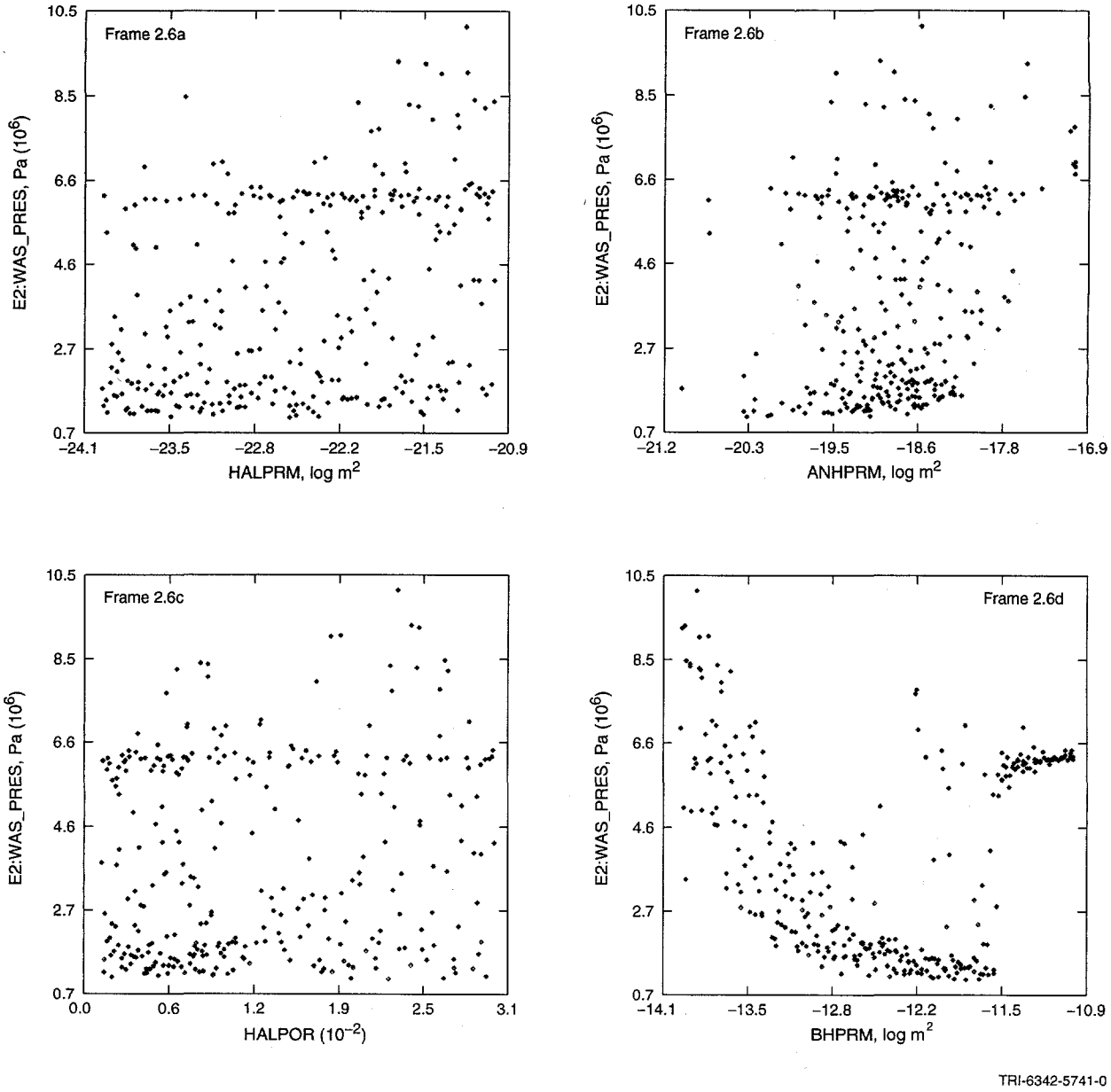

Fig. 2.6. Scatterplots for pressure in lower waste panel at 10,000 yr with an E2 intrusion into the lower waste panel at 1000 yr (*E2:WAS_PRES*) versus the three variables (*HALPRM*, *ANHPRM*, *HALPOR*) selected in stepwise regression analysis with raw and rank-transformed data (Table 2.3) and one additional variable (*BHPRM*) identified by examination of scatterplots: (2.6a) *HALPRM*, (2.6b) *ANHPRM*, (2.6c) *HALPOR*, and (2.6d) *BHPRM*.

with respect to the uncertainty in *E2:WAS_PRES* was not even identified in the regression with raw or rank-transformed data. In contrast, the analyses for *E2:WAS_SATB* included the dominant variables in the regression models even though the R^2 values were low. As an aside, the interesting pattern involving *E2:WAS_PRES* and *BHPRM* in Fig. 2.6d results from two-phase flow in the borehole connecting the waste panel with overlying formations, with gas typically flowing up the borehole and brine typically flowing down the borehole (Helton et al. 1998).

As should be apparent from the regressions in Table 2.3 and the associated scatterplots in Figs. 2.3-2.6, the examination of scatterplots is an important part of sampling-based sensitivity analysis and can reveal patterns that are missed by regression-based procedures. In the development that follows, the variables in Table 2.1 will be used to illustrate a number of procedures for the identification of patterns in scatterplots. These variables were selected to illustrate pattern identification procedures because they constitute a spectrum of analysis possibilities. In particular, regression analysis with both raw and rank-transformed data performs well for *E0:WAS_PRES*; regression analysis with rank-transformed, but not raw, data performs well for *E0:BRAALIC*; regression analysis with neither raw nor rank-transformed data performs particularly well for *E2:WAS_SATB* but both regression models still include the two dominant variables; and regression analysis with raw and rank-transformed data fails to identify the dominant variable with respect to the uncertainty in *E2:WAS_PRES*.

3. Linear Relation: $y = \beta_0 + \beta_1 x$

The coefficients β_0 and β_1 in a first-order polynomial can be estimated with the well-known ordinary least squares procedure. Specifically, $\hat{\beta}_0$ and $\hat{\beta}_1$ are given by (Sect. 2.1, Draper and Smith, 1981)

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, \tag{3.1}$$

where

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_{nS} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_{nS} \end{bmatrix}$$

and the superscript T denotes matrix transpose. The estimated linear regression model then becomes

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x, \tag{3.2}$$

with the coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ deriving from the sampled and calculated values contained in the pairs (x_k, y_k), $k = 1, 2, \ldots, nS$, as indicated in Eq. (3.1).

The apparent strength of the linear relationship in Eq. (3.2) provides one way to assess the relationship between x and y. The linear correlation coefficient ρ_{xy}, which is also called the Pearson correlation coefficient, provides the most commonly used measure to assess the strength of the linear relationship between x and y and is defined by

$$\rho_{xy} = \sigma_{xy} / (\sigma_x \sigma_y), \tag{3.3}$$

where σ_{xy} denotes the covariance between x and y, and σ_x and σ_y denote the standard deviation of x and y, respectively. In turn, ρ_{xy} is estimated by

$$\begin{aligned} \hat{\rho}_{xy} &= \sum_{k=1}^{nS}(x_k - \bar{x})(y_k - \bar{y}) \Bigg/ \left[\sum_{k=1}^{nS}(x_k - \bar{x})^2\right]^{1/2} \left[\sum_{k=1}^{nS}(y_k - \bar{y})^2\right]^{1/2} \\ &= \left(\sum_{k=1}^{nS} x_k y_k - nS\,\bar{x}\,\bar{y}\right) \Bigg/ \left(\sum_{k=1}^{nS} x_k - nS\,\bar{x}^2\right)^{1/2} \left(\sum_{k=1}^{nS} y_k^2 - nS\,\bar{y}^2\right)^{1/2}, \end{aligned} \tag{3.4}$$

where

$$\bar{x} = \sum_{k=1}^{nS} x_k / nS, \ \bar{y} = \sum_{k=1}^{nS} y_k / nS.$$

The quantity $\hat{\rho}_{xy}$ is often called the sample correlation coefficient.

The reason why ρ_{xy}, and hence $\hat{\rho}_{xy}$, provides a measure of the strength of the linear relationship between x and y is not immediately apparent from Eqs. (3.3) and (3.4). Rather, this reason is perhaps best understood in the context of the regression model in Eq. (3.2) with both x and y standardized to variables with a mean of 0 and a standard deviation of 1 by

$$\tilde{x}_k = (x_k - \bar{x}) / \hat{\sigma}_x, \ \tilde{y}_k = (y_k - \bar{y}) / \hat{\sigma}_y, \tag{3.5}$$

where

$$\hat{\sigma}_x = \left[\sum_{k=1}^{nS} (x_k - \bar{x})^2 / (nS-1) \right]^{1/2}, \hat{\sigma}_y = \left[\sum_{k=1}^{nS} (y_k - \bar{y})^2 / (nS-1) \right]^{1/2}.$$

With the preceding standardization, Eq. (3.1) yields the regression model

$$(y - \bar{y}) / \hat{\sigma}_y = 0 + \hat{\rho}_{xy}(x - \bar{x}) / \hat{\sigma}_x = \hat{\rho}_{xy}(x - \bar{x}) / \hat{\sigma}_x. \tag{3.6}$$

Thus, $\hat{\rho}_{xy}$ is the standardized regression coefficient relating x to y. As such, $\hat{\rho}_{xy}$ characterizes the effect that changing x by a fixed fraction of its standard deviation will have on y, with this effect being measured relative to the standard deviation of y.

In addition, the correlation coefficient ρ_{xy}, and hence $\hat{\rho}_{xy}$, provides a measure of the fraction of the variance of y that can be accounted for by x. Again, this is best seen in the context of the regression model in Eq. (3.2), for which the following identity can be established (Sect. 1.3, Draper and Smith, 1981):

$$\sum_{k=1}^{nS} (y_k - \bar{y})^2 = \sum_{k=1}^{nS} (\hat{y}_k - \bar{y})^2 + \sum_{k=1}^{nS} (\hat{y}_k - y_k)^2. \tag{3.7}$$

The summation $\sum_k (\hat{y}_k - \bar{y})^2$ represents the part of the variance of y that can be accounted for by $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$, with the result that

$$R^2 = \sum_{k=1}^{nS} (\hat{y}_k - \bar{y})^2 / \sum_{k=1}^{nS} (y_k - \bar{y})^2 \tag{3.8}$$

represents the fraction of the variance of y that can be accounted by use of x in a linear approximation to y. The preceding quantity is called the R^2 value or the coefficient of determination for x and y. An R^2 value close to 1 indicates that x can account for most of the uncertainty in y; in contrast, an R^2 value close to 0 indicates that a linear relationship involving x accounts for little of the uncertainty in y.

Like the standardized regression coefficient, the R^2 value can be expressed in terms of $\hat{\rho}_{xy}$. The development of the relationship between R^2 and $\hat{\rho}_{xy}$ requires an exact representation for the coefficients $\hat{\beta}_1$ and $\hat{\beta}_0$ in the definition of $\hat{y}_k$ in Eq. (3.8). Specifically, the vector equality in Eq. (3.1) leads to

$$\begin{aligned}
\hat{\beta}_1 &= \frac{-\left(\sum_{k=1}^{nS} x_k\right)\left(\sum_{k=1}^{nS} y_k\right) + nS\sum_{k=1}^{nS} x_k y_k}{nS\sum_{k=1}^{nS} x_k^2 - \left(\sum_{k=1}^{nS} x_k\right)} \\
&= \frac{\sum_{k=1}^{nS} x_k y_k - nS\left(\sum_{k=1}^{nS} x_k / nS\right)\left(\sum_{k=1}^{nS} y_k / nS\right)}{nS\sum_{k=1}^{nS} x_k^2 - nS\left(\sum_{k=1}^{nS} x_k / nS\right)} \\
&= \sum_{k=1}^{nS} (x_k - \bar{x})(y_k - \bar{y}) \Big/ \sum_{k=1}^{nS} (x_k - \bar{x})^2
\end{aligned} \tag{3.9}$$

and

$$\begin{aligned}
\hat{\beta}_0 &= \frac{\left(\sum_{k=1}^{nS} x_k^2\right)\left(\sum_{k=1}^{nS} y_k\right) - \left(\sum_{k=1}^{nS} x_k\right)\left(\sum_{k=1}^{nS} x_k y_k\right)}{nS\sum_{k=1}^{nS} x_k^2 - \left(\sum_{k=1}^{nS} x_k\right)^2} \\
&= \bar{y} - \left(\frac{-\left(\sum_{k=1}^{nS} x_k\right)\left(\sum_{k=1}^{nS} y_k\right) + nS\sum_{k=1}^{nS} x_k y_k}{nS\sum_{k=1}^{nS} x_k^2 - \left(\sum_{k=1}^{nS} x_k\right)^2}\right)\bar{x} \\
&= \bar{y} - \hat{\beta}_1 \bar{x},
\end{aligned} \tag{3.10}$$

where the second equality involving $\hat{\beta}_0$ follows from the substitution

$$\sum_{k=1}^{nS} x_k y_k = \sum_{k=1}^{nS} x_k (y_k - \bar{y} + \bar{y}) \tag{3.11}$$

and some simple algebraic manipulations.

Given the preceding representations for $\hat{\beta}_0$ and $\hat{\beta}_1$,

$$\sum_{k=1}^{nS} (\hat{y}_k - \bar{y})^2 = \sum_{k=1}^{nS} [\hat{\beta}_0 + \hat{\beta}_1 x_k) - \bar{y}]^2 \quad \text{[from Eq. (3.2)]}$$

$$= \sum_{k=1}^{nS} [(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_k) - \bar{y}]^2 \quad \text{[from Eq. (3.10)]}$$

$$= \hat{\beta}_1^2 \sum_{k=1}^{nS} (x_k - \bar{x})^2$$

$$= \left[\sum_{k=1}^{nS} (x_k - \bar{x})(y_k - \bar{y}) \right]^2 \Bigg/ \sum_{k=1}^{nS} (x_k - \bar{x})^2 . \tag{3.12}$$

[from Eq. (3.9)]

Hence, from Eqs. (3.8) and (3.12),

$$R^2 = \left[\sum_{k=1}^{nS} (x_k - \bar{x})(y_k - \bar{y}) \right]^2 \Bigg/ \left[\sum_{k=1}^{nS} (x_k - \bar{x})^2 \right] \left[\sum_{k=1}^{nS} (y_k - \bar{y})^2 \right] = \hat{\rho}_{xy}^2 . \tag{3.13}$$

Thus, the square of the sample correlation coefficient is equal to the fraction of the variance of y that can be accounted for by $\hat{y}$ as defined in Eq. (3.2), and hence by x under a linear transformation.

The preceding has given two interpretations of the correlation coefficient ρ_{xy} between x and y in terms of the regression model in Eq. (3.2). First, the sample correlation coefficient $\hat{\rho}_{xy}$ (i.e., the approximation to ρ_{xy} defined in Eq. (3.4)) can be viewed as the regression coefficient in Eq. (3.2) when x and y are standardized to mean 0 and standard deviation 1. Second, $\hat{\rho}_{xy}$ can be viewed as the square root of the R^2 value for the regression model in Eq. (3.2) (i.e., $\hat{\rho}_{xy}^2 = R^2$). The correlation coefficient can also be viewed as a parameter in a joint normal distribution involving x and y (see Sect. 2.13, Myers, 1990); however, this interpretation is not as intuitively appealing as the two involving the regression model in Eq. (3.2).

When $\hat{\rho}_{xy}$ is close to 1 or -1, an almost linear relationship exists between x and y (see definition of $R^2 = \hat{\rho}_{xy}^2$ in Eq. (3.8)). However, large changes in x may still result in small changes in y if the regression coefficient $\hat{\beta}_1$ in Eq. (3.2) is small. Indeed, the magnitude $|\hat{\beta}_1|$ of $\hat{\beta}_1$ is not a very informative quantity because $|\hat{\beta}_1|$ depends on the units in which x and y are expressed (e.g., changing the units on x from millimeters to kilometers will have a large effect on $|\hat{\beta}_1|$ but no effect on the underlying physical relationships). For this reason, x and y are often standardized to mean 0 and standard deviation 1. As previously discussed, this standardization results in the equality $\hat{\beta}_1 = \hat{\rho}_{xy}$ and also in $\hat{\beta}_1$ characterizing changes in y normalized to $\hat{\sigma}_y$ relative to changes in x normalized to $\hat{\sigma}_x$.

Although $\hat{\rho}_{xy}^2 \doteq 1$ implies a strong linear dependence between x and y, $\hat{\rho}_{xy} \doteq 0$ cannot be used to infer that no relationship exists between x and y (i.e., that x and y are independent). In particular, zero correlations can occur in the presence of a nonmonotonic relationship between x and y. For example, $\rho_{xy} = 0$ for $y = 1 - x^2$ with $-1 \le x \le 1$ and also for $y = \cos x$ with $0 \le x \le 2\pi$. A more interesting example is given by the scatterplot for *BHPRM* in Fig. 2.6. Thus, a linear relationship can be assumed to exist between x and y if $|\hat{\rho}_{xy}|$ is close to 1. Further, linear relationships of lesser strength (i.e., smaller R^2 values) exist for smaller values of $|\hat{\rho}_{xy}|$. For $|\hat{\rho}_{xy}| \doteq 0$, the implication is that no linear relationship exists between x and y.

A significance test can be used to indicate if $\hat{\rho}_{xy}$ appears to be different from 0. For example,

$$t = \hat{\rho}_{xy}(nS-2)^{1/2} / (1-\hat{\rho}_{xy}^2)^{1/2} \tag{3.14}$$

has a t-distribution with $nS-2$ degrees of freedom when x, y are uncorrelated and have a bivariate normal distribution (Press et al., 1992, p. 631). Further,

$$z = \hat{\rho}_{xy}\sqrt{nS} \tag{3.15}$$

is distributed approximately normally with mean 0 and standard deviation 1 when x and y are uncorrelated, x and y have enough convergent moments (i.e., the tails die off sufficiently rapidly), and nS is large (typically > 500) (Press et al., 1992, p. 631). Then,

$$prob(|r|>|\hat{\rho}_{xy}|) = erfc(|\hat{\rho}_{xy}|\sqrt{nS} / \sqrt{2}), \tag{3.16}$$

where $prob(|r|>|\hat{\rho}_{xy}|)$ is the probability that random variation would produce a value r for $\hat{\rho}_{xy}$ larger in absolute value than the observed value $\hat{\rho}_{xy}$ and *erfc* is the complementary error function (Press et al., 1992, p. 631). Significance results obtained with t in Eq. (3.14) converge to those obtained with z in Eq. (3.15) as nS increases. However, as x and y are unlikely to have normal distributions in real analysis problems, results obtained with t and small values of nS should simply be viewed as one form of guidance as to whether or not a linear relationship actually exists between x and y.

If several x_i have scatterplots that appear to have nonzero values for $\hat{\rho}_{x_i y}$, then the relative importance of these x_i can be ordered by the absolute values of $\hat{\rho}_{x_i y}$. This is equivalent to ordering the x_i on the basis of the strength of the linear relationship associated with the pairs (x_{ik}, y_k), $k = 1, 2, \ldots, nS$. This is also equivalent to ordering the x on the basis of p-values obtained from the distributions associated with Eq. (3.14) or (3.15), where the p-value designates the probability that a value for $\hat{\rho}_{xy}$ will be obtained that exceeds the observed value for $\hat{\rho}_{xy}$ in absolute value (i.e., $prob(|r|>|\hat{\rho}_{xy}|)$ in Eq. (3.16)). Actually, the ordering is done on the complements of the p-values because smaller p-values are associated with larger values for $|\hat{\rho}_{x_i y}|$.

Standardized regression coefficients provide a popular way of ranking variable importance (e.g., Chan, 1996; Hamby, 1995; Ma et al., 1993; Ma and Ackerman, 1993; Whiting et al., 1993). However, when the x_i are independent, the standardized regression coefficient for x_i is equal to $\hat{\rho}_{x_i y}$ and so the two rankings are identical. Specifically, the regression model relating y to the x_i has the form

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{nI} \hat{\beta}_i x_i, \tag{3.17}$$

where $\hat{\beta}$ has the same functional form as in Eq. (3.1) with

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_{nI} \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1nI} \\ \vdots & \vdots & & \vdots \\ 1 & x_{nS} & \cdots & x_{nS,nI} \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_{nS} \end{bmatrix}$$

(Sect. 2.6, Draper and Smith, 1981). If the x_{ik}'s have been selected so that the rows of $\mathbf{X}$ are orthogonal (i.e., so that $\mathbf{X}^T\mathbf{X}$ is a diagonal matrix with diagonal elements $d_0, d_1, \ldots, d_{nI}$, which is equivalent to the individual x_i being independent and thus having sample correlations of 0), then

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

$$= \begin{bmatrix} d_0 & 0 & \cdots & 0 \\ 0 & d_1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_{nI} \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{nS,1} \\ \vdots & \vdots & & \vdots \\ x_{1nI} & x_{2nI} & \cdots & x_{nSnI} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{nS} \end{bmatrix} \tag{3.18}$$

and so

$$\hat{\beta}_i = \sum_{k=1}^{nS} x_{ik} y_k / d_k = \sum_{k=1}^{nS} x_{ik} y_k / \sum_{k=1}^{nS} x_{ik}^2. \tag{3.19}$$

Thus, when x_i and y are standardized to mean 0 and standard deviation 1 (see Eq. (3.5)),

$$\hat{\beta}_i = \sum_{k=1}^{nS}(x_{ik}-\bar{x}_i)(y_k-\bar{y}) \Bigg/ \left[\sum_{k=1}^{nS}(x_{ik}-\bar{x}_i)^2\right]^{1/2}\left[\sum_{k=1}^{nS}(y_k-\bar{y})^2\right]^{1/2} = \hat{\rho}_{x_i y} \tag{3.20}$$

and the standardized regression coefficient $\hat{\beta}_i$ and the (sample) correlation coefficient $\hat{\rho}_{x_i y}$ are equal.

Partial correlation coefficients are another popular way of ranking variable importance (e.g., Hamby, 1995; Whiting et al., 1993; Breshears et al., 1992). However, the partial correlation coefficient is just a special form of the sample correlation coefficient. In particular, if least squares techniques are used to determine the coefficients in

$$\hat{x}_j = \hat{\alpha}_0 + \sum_{\substack{i=1\\ i\neq j}}^{nI}\hat{\alpha}_i x_i \text{ and } \hat{y} = \hat{\beta}_0 + \sum_{\substack{i=1\\ i\neq j}}^{nI}\hat{\beta}_i x_i, \tag{3.21}$$

then the partial correlation coefficient $\hat{p}_{x_j y}$ between x_j and y is the sample correlation $\hat{\rho}_{\hat{x}_j \hat{y}}$ determined for the pairs $(x_{jk}-\hat{x}_{jk}, y_k-\hat{y}_k)$, $k = 1, 2, \ldots, nS$. Thus, $\hat{p}_{x_j y}$ is the sample correlation between x_j and y after a correction has been made for the linear effects of the other x_i.

The following relationship exists between $\hat{p}_{x_j y}$ and the standardized regression coefficient $\hat{\beta}_j$ (i.e., the coefficient associated with the regression model in Eq. (3.17) when x_i and y are standardized to mean 0 and standard deviation 1):

$$\hat{p}_{x_j y} = \hat{\beta}_j[(1-R_j^2)/(1-R_y^2)]^{1/2}, \tag{3.22}$$

where R_j^2 is the R^2 value that results from regressing x_j on y and the x_i, $i = 1, 2, \ldots, nI$ with $i \neq j$, and R_y^2 is the R^2 value that results from regressing y on the x_i, $i = 1, 2, \ldots, nI$ (Iman et al., 1985, Eq. (1)). In the event that the x_i are orthogonal,

$$R_y^2 = \sum_{i=1}^{nI} R_i^2 = \sum_{i=1}^{nI}\hat{\beta}_i^2 = \sum_{i=1}^{nI}\hat{\rho}_{x_i y}^2, \tag{3.23}$$

with the first equality following from Eq. (III-74) of Helton et al. (1991) and the second and third equalities following from Eqs. (3.13) and (3.20). Thus,

$$\hat{p}_{x_j y} = \hat{\beta}_j \left[(1-\hat{\beta}_j^2) / \left(1 - \sum_{i=1}^{nI} \hat{\beta}_i^2 \right) \right]^{1/2} = \hat{\rho}_{x_j y} \left[(1-\hat{\rho}_{x_j y}^2) / \left(1 - \sum_{i=1}^{nI} \hat{\rho}_{x_i y}^2 \right) \right]^{1/2} . \tag{3.24}$$

Because of the inequality

$$b(1-b^2)^{1/2} > a(1-a^2)^{1/2} \tag{3.25}$$

for $a^2 + b^2 < 1$ and $a < b$ (Fig. 3.1), an ordering of variable importance based on $|\hat{p}_{x_j y}|$, $|\hat{\beta}_j|$ or $|\hat{\rho}_{x_j y}|$ will produce the same results when the x_i are orthogonal; further, the values for $\hat{\beta}_j$ and $\hat{\rho}_{x_j y}$ will be the same and generally different from $\hat{p}_{x_j y}$.

Due to the conceptual simplicity of the sample correlation coefficient $\hat{\rho}_{xy}$ and its close relationship to standardized regression coefficients and partial correlation coefficients in the presence of orthogonal values for the x_i's, this study will use $\hat{\rho}_{xy}$ to assess the strength of the linear relationship between x and y. In the presence of small deviations from orthogonality (i.e., the existence of small correlations between the x_i), the three measures will still give similar results. However, in the presence of large deviations from orthogonality, the three measures can give quite different, and possibly misleading, indications of the effects of individual variables.

As noted earlier, $\hat{\rho}_{xy} \doteq 0$ should not be interpreted to mean that no relationship exists between x and y. For example,

$$y = \beta_0 x^{\beta_1} \tag{3.26}$$

results in a low, but nonzero, value for ρ_{xy} even though there is no noise in the relationship between x and y. In this case, a logarithmic transformation will linearize the relationship between x and y. However, such transformations may not exist and, given that they do exist, identifying them is not always easy. For example, logarithmic transformations are not applicable when some of the y values are zero, which is a fairly common analysis situation. One possible transformation of fairly broad applicability is the rank transformation, which is discussed in Sect. 4.

A possible complication in the use of $\hat{\rho}_{xy}$ to identify the existence of a relationship between x and y can be the existence of interactions with other variables. For example, the relationship between y, x_1, and x_2 might be of the form

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2, \tag{3.27}$$

which can also be expressed as

$$y = \beta_0 + \beta_1[1-(\beta_{12}/\beta_1)x_2]x_1 + \beta_2 x_2$$

$$= \beta_0 + \beta_1 x_1 + \beta_2[1-(\beta_{12}/\beta_2)x_1]x_2 .$$

As long as the variation in x_1 is large relative to the variation in $1-(\beta_{12}/\beta_1)x_2$ or the variation in x_2 is large relative to the variation in $1-(\beta_{12}/\beta_1)x_1$, the fact that x_1 or x_2 does indeed have a significant effect on y should be identified by the corresponding value for $\hat{\rho}_{xy}$. Thus, it is not considered necessary to specifically consider interaction effects to identify important variables, although it is certainly possible to calculate $\hat{\rho}_{xy}$ with $x = x_i x_j$ if desired. Further, use of contingency tables to be discussed later (Sect. 7) allows the identification of nonlinear effects without the assumption of a specific model form.

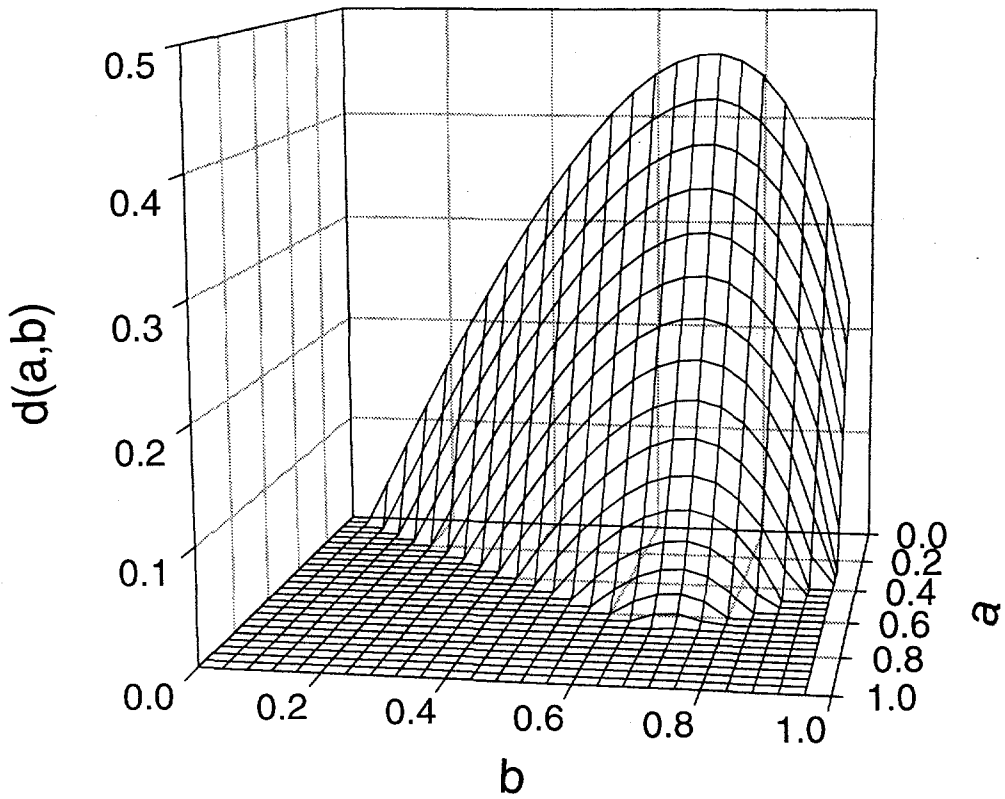

Fig. 3.1. Graph of $d(b, a) = b(1-b^2)^{1/2} - a(1-a^2)^{1/2} > 0$ subject to constraints $0 < a < b < 1$, $a^2 + b^2 < 1$.

Intentionally Left Blank

4. Monotonic Relation: $r(y) = \gamma_0 + \gamma_1 r(y)$

When the relationship between x and y is nonlinear but monotonic, the relationship can be linearized by use of the rank transformation. The idea of the rank transformation is simple. Specifically, the pairs (x_k, y_k) are transformed into a new sequence of pairs

$$[r(x_k), r(y_k)], k = 1, 2, \ldots, nS, \tag{4.1}$$

where (i) the smallest value of x_k is assigned a rank of 1 (i.e., $r(x_k) = 1$), the next largest value of x_k is assigned a rank of 2 (i.e., $r(x_k) = 2$), and so up to the largest value of x_k, which is assigned a rank of nS (i.e., $r(x_k) = nS$), (ii) averaged ranks are assigned to equal values of x_k (e.g., if $x_j = x_k$, $x_l \neq x_j$ for $l \neq j$, k, and $p - 1$ observations have values less than x_j, then $r(x_j) = r(x_k) = (p + p + 1)/2$), and (iii) the assignment of the ranks for y (i.e., $r(y_k)$) is accomplished in the same manner as the assignment of ranks for x.

Rank-transformed data can be analyzed in exactly the same manner as discussed in Sect. 3 for untransformed data. In particular, the strength of the linear relationship between the rank-transformed variables in Eq. (4.1) can be measured with Spearman's rank correlation coefficient for x and y, η_{xy}, which is simply Pearson's correlation coefficient in Eq. (3.4) calculated on ranks. The test for zero rank correlation uses a table of quantiles for $|\hat{\eta}_{xy}|$ (e.g., Table A10, Conover, 1980). For $nS \geq 30$,

$$z = \hat{\eta}_{xy}\sqrt{nS-1} \tag{4.2}$$

approximately follows the normal distribution for $\eta_{xy} = 0$ (p. 456, Conover, 1980), which is very similar to the approximation to the distribution indicated for $\hat{\rho}_{xy}$ in Eq. (3.15). Thus, similarly to Eq. (3.16) for $\hat{\rho}_{xy}$,

$$prob(|r| > |\hat{\eta}_{xy}|) = erfc(|\hat{\eta}_{xy}|\sqrt{nS-1}/\sqrt{2}), \tag{4.3}$$

where $prob(|r| > |\eta_{xy}|)$ is the probability that random variation would produce a value r for η_{xy} larger in absolute value than the observed value $\hat{\eta}_{xy}$ and *erfc* is the complementary cumulative error function (Press et al. 1992, p. 631).

Regression coefficients and partial correlation coefficients can also be calculated with rank-transformed data as discussed in Sect. 3 (e.g., Sanchez and Blower, 1997; Gwo et al., 1996; Helton et al., 1996; Hamby, 1995; Blower and Dowlatabadi, 1994; Whiting et al., 1993; MacDonald and Campbell, 1986). As an aside, the form of the regression model after y and the x_i's have been standardized to mean 0 and standard deviation 1 is

$$(y-\bar{y})/\hat{\sigma}_y = \sum_{i=1}^{nI} (\hat{\beta}_i \hat{\sigma}_{x_i} / \hat{\sigma}_y)(x_i - \bar{x}_i)/\hat{\sigma}_{x_i} \ , \tag{4.4}$$

where $\hat{\beta}_i$ is the regression coefficient obtained with the original (i.e., nonstandardized) values for y and the x_i's. When rank-transformed data are being used and there are no ties in the y or x_i values, then $\hat{\sigma}_{x_i} = \hat{\sigma}_y$ and so the standardized regression coefficient (i.e., $\hat{\beta}_i \hat{\sigma}_{x_i} / \hat{\sigma}_y$) is the same as the original, nonstandardized coefficient (i.e., $\hat{\beta}_i$). Thus, standardization is automatically accomplished by the use of rank-transformed data as long as there are no ties in the y and x values.

Closely related to Spearman's coefficient is Kendall's τ (Conover, 1980, pp. 255 - 260). Because both coefficients give nearly identical significance results, this alternative for identifying monotonic relationships is considered only briefly. Kendall's τ measures the degree of concordance in a set of observations of the form in Eq. (1.7). The pairs (x_r, y_r) and (x_s, y_s) are said to be concordant if both members of one pair are less than the corresponding members of the other pair (i.e., $x_r < x_s$, $y_r < y_s$ or $x_r > x_s$, $y_r > y_s$). Further, the pairs are said to be discordant if the two members in one pair differ in opposite directions from the corresponding members in the other pair (i.e., $x_r < x_s$, $y_r > y_s$ or $x_r > x_s$, $y_r < y_s$). Kendall's τ is estimated by

$$\hat{\tau}_{xy} = (N_c - N_d)/[nS(nS-1)/2], \tag{4.5}$$

where N_c is the number of concordant pairs of observations, N_d is the number of discordant pairs of observations, and $nS(nS-1)/2$ is the total number of pairs $\{(x_r, y_r), (x_s, y_s)\}$ of observations. The statistic $\hat{\tau}_{xy}$ has a distribution that is adequately approximated by the normal distribution for sample sizes as small as $nS = 8$. In contrast, larger samples (e.g., $nS \geq 30$) are required for $\hat{\eta}_{xy}$ to approach a normal distribution, although Monte Carlo sensitivity studies typically use sample sizes larger than $nS = 30$. As estimates for Spearman's coefficient $\hat{\eta}_{xy}$ and Kendall's $\hat{\tau}_{xy}$ produce similar rankings of monotonicity and $\hat{\eta}_{xy}$ is more intuitively appealing because of its close relationship to Pearson's coefficient $\hat{\rho}_{xy}$, this presentation will use $\hat{\eta}_{xy}$ to identify nonlinear but monotonic relationships in scatterplots.

5. Location of y Dependent on x

Tests for two distinct types of patterns in scatterplots were considered in Sects. 3 and 4, with the Pearson correlation coefficient introduced as a means of identifying linear patterns (Sect. 3) and the Spearman correlation coefficient introduced as a means of identifying nonlinear but monotonic patterns (Sect. 4). This section introduces tests for a broader class of patterns. Specifically, patterns are sought where some measure of central tendency for y changes with changing values for x. Linear and monotonic patterns have this characteristic. However, decidedly nonlinear and nonmonotonic patterns can also have this characteristic (e.g., see the scatterplot for *BHPRM* in Fig. 2.6).

The approach taken is to divide the values for x (i.e., x_k, $k = 1, 2, \ldots, nS$) into nX classes and then to test to determine if y has a common measure of central tendency across these classes. Thus, x must be defined on at least a nominal scale to permit the definition of the necessary classes. Classic measures of central tendency are the mean or expected value, $E(y)$, and the median, $y_{0.5}$. The mean is a more widely used measure of central tendency but the median is less sensitive to outliers (e.g., see the Princeton robustness study reported in Andrews et al. (1972)).

Most of the x's under consideration are actually defined on an interval scale (see Table 2.2), and the required classes are obtained by subdividing the range of x into a sequence of mutually exclusive and exhaustive subintervals containing equal numbers of sampled values (Fig. 5.1). A few x's are discrete with unequal probabilities for the individual x values (e.g., see *WMICDFLG* in Fig. 2.3); for these variables, individual classes are defined for each of the distinct values. However, the optimum definition of the classes is not at all apparent, and in practice, some experimentation may be required to determine an appropriate division of the x values into individual classes.

For a given variable x and its nX associated classes, the following statistics will be used to identify apparent deviations from a common central tendency: (i) the ANOVA F statistic for equal means, which requires an interval scale for y (Sect. 5.1), (ii) the Kruskal-Wallis test for common locations, which requires an ordinal scale for y (Sect. 5.2), and (iii) the chi-square test for equal medians, which also requires an ordinal scale for y (Sect. 5.3).

5.1 Common Means: ANOVA *F* Statistic

The first possibility considered is that different classes of x values have different expected values for y. For notational convenience, let q, $q = 1, 2, \ldots, nX$, designate the individual classes into which the values of x have been divided; let $\mathcal{X}_q$ designate the set such that $k \in \mathcal{X}_q$ only if x_k belongs to class q; and let nX_q equal the number of elements contained in $\mathcal{X}_q$ (i.e., the number of x_k's associated with class q). The ANOVA F test is commonly used to test for equivalence of conditional means (Scheffé, 1959):

$$F(nX-1, nS-nX) = \frac{\left[\sum_{q=1}^{nX} nX_q \bar{y}_q^2 - nS\,\bar{y}^2\right] / (nX-1)}{\left[\sum_{k=1}^{nS} y_k^2 - \sum_{q=1}^{nX} nX_q \bar{y}_q^2\right] / (nS-nX)}, \tag{5.1}$$

where $nX - 1$ and $nS - nX$ are the number of degrees of freedom for the numerator and denominator, respectively, $\bar{y}_q = \sum_{k \in X_q} y_k \,/\, nX_q$, and $\bar{y}$ is defined in conjunction with Eq. (3.4).

If the y values conditional on each class of x values are normally distributed with equal expected values, then the statistic $F\,(nX - 1, nS - nX)$ in Eq. (5.1) follows an F distribution with $(nX - 1, nS - nX)$ degrees of freedom. This is the most powerful test for equality of means given that the indicated normality assumptions hold (Scheffé 1959). The probability $prob(F > \hat{F} | \eta_1, \eta_2)$ of exceeding an F statistic of value $\hat{F}$ calculated with (η_1, η_2) degrees of freedom and the indicated normality assumptions can be estimated by

$$prob(F > \hat{F} | \eta_1, \eta_2) = I_v(\eta_2 / 2, \eta_1 / 2), \quad v = \eta_2 / (\eta_2 + \eta_1 \hat{F}), \tag{5.2}$$

where $I_v\,(a, b)$ designates the incomplete beta function (p. 222, Press et al., 1992).

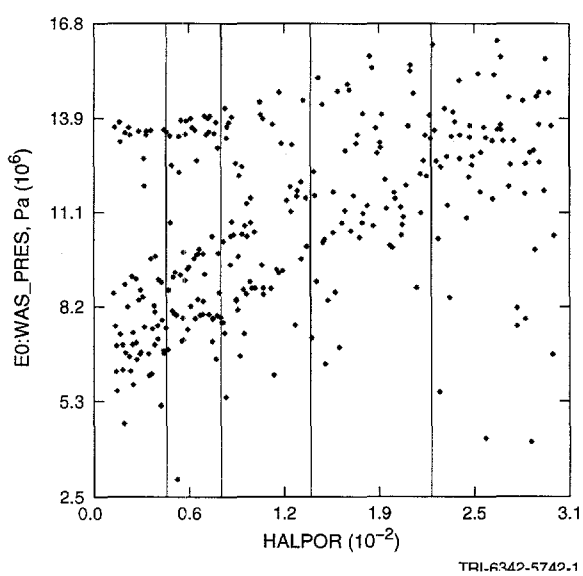

Fig. 5.1. Example of the partitioning of the range of $x = HALPOR$ into $nX = 5$ classes for $y = E0{:}WAS_PRES$.

Although the F-statistic will be used to test for equal expected values across the nX classes into which x has been divided, the y values for each class may not actually follow a normal distribution. If desired, various goodness of fit tests (e.g., chi-square, Kolmogorov-Smirnov, Cramer-von Mises, Anderson-Darling) can be used to test for normality of the y values (pp. 94 - 95, Kleijnen, 1987; D'Agostino and Stephens, 1986). However, the number of observations per class (e.g., 30 or 60 for many of the variables considered in this study) may be too small to provide a powerful test. If a goodness-of-fit test leads to a rejection of the normality hypothesis, then it may be appropriate to apply a normalizing transformation such as the Box-Cox transformation, which includes the logarithmic transformation as a special case (pp. 175 - 185, Kleijnen, 1987). Fortunately, the ANOVA F test is robust with respect to deviations from normality (p. 237, Conover, 1980). For perspective, Monte Carlo estimates of $prob(F > \hat{F} | \eta_1, \eta_2)$ will be presented in Sect. 10.

5.2 Common Locations: Kruskal-Wallis Test

The next measure of common central tendency considered is the Kruskal-Wallis test, which is based on rank-transformed data (p. 229, Conover, 1980). The associated statistic is defined by

$$\hat{T} = \left[\sum_{q=1}^{nX} (R_q^2 / nX_q) - nS(nS+1)^2 / 4 \right] / S^2, \tag{5.3}$$

where

$$R_q = \sum_{k \in X_q} r(y_k), \quad S^2 = \left[\sum_{k=1}^{nS} r(y_k)^2 - nS(nS+1)^2 / 4 \right] / (nS-1)$$

and $r(y_k)$ is defined in conjunction with Eq. (4.1) (p. 230, Conover, 1980).

If the y values conditional on each class of x values have the same distribution, then the statistic $\hat{T}$ in Eq. (5.3) approximately follows a chi-square distribution with $nX - 1$ degrees of freedom (pp. 230 - 231, Conover, 1980). Given this approximation, the probability $prob(T > \hat{T} | nX - 1)$ of obtaining a value T that exceeds $\hat{T}$ in the presence of identical y distributions for the individual classes is given by

$$prob(T > \hat{T} | nX - 1) = Q[(nX-1)/2, \hat{T}/2], \tag{5.4}$$

where Q (a, b) designates the complement of the incomplete gamma function (p. 215, Press et al., 1992). A small value for $prob(T > \hat{T} | nX - 1)$ indicates that the y's conditional on individual classes have different distributions and thus, most likely, different means and medians.

5.3 Common Medians: Chi-Square for Contingency Tables

The final possibility considered is that different classes of x values have different median values for y. The chi-square test for contingency tables can be used to test for this situation (pp. 143 - 178, Conover, 1980). First, the median $y_{0.5}$ is estimated for all nS observations. Specifically,

$$y_Q = \begin{cases} y_{(Q\,nS)} & \text{if } Q\,nS \text{ is an integer} \\ [y_{([Q\,nS])} + y_{([Q\,nS]+1)}]/2 & \text{otherwise} \end{cases} \tag{5.5}$$

where $Q = 0.5$ (Q = 0.25 and 0.75 will be considered in Sect. 6.2) and $y_{(k)}$, $k = 1, 2, \ldots, nS$, denotes the ordering of the y values such that $y_{(k)} \leq y_{(k+1)}$ (p. 14, David, 1970). The individual classes of x values considered in Sects. 5.1 and 5.2 are then divided into subsets of y values that fall above and below $y_{0.5}$ (Fig. 5.2). For class q, let nX_{1q} equal the number of y values that exceed $y_{0.5}$, and let nX_{2q} equal the number of y values that are less than or equal to $y_{0.5}$. The result of this partitioning is a $2 \times nX$ contingency table with nX_{rq} observations in each cell.

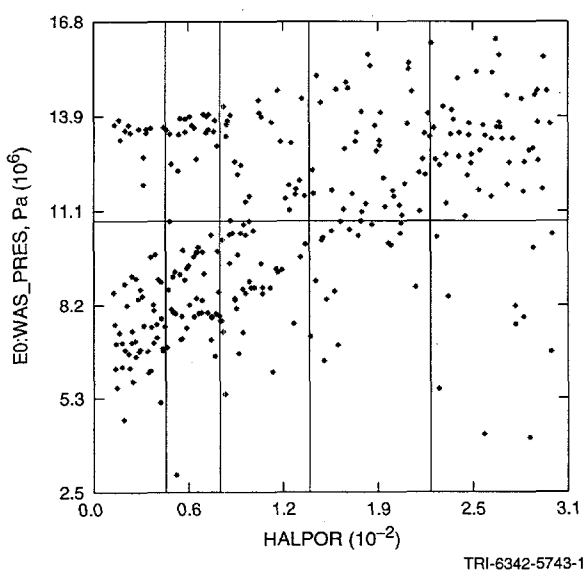

Fig. 5.2. Example of the partitioning of the range of x = *HALPOR* into $nX = 5$ classes and the range of y = *E0:WAS_PRES* into values above and below the median $y_{0.5}$.

The following statistic can be defined:

$$\hat{T} = \sum_{q=1}^{nX} \sum_{r=1}^{2} (nX_{rq} - nE_{rq})^2 \,/\, nE_{rq}, \tag{5.6}$$

where

$$nE_{rq} = \left(\sum_{r=1}^{2} nX_{rq} \right) \left(\sum_{q=1}^{nX} nX_{rq} \right) / \, nS$$

and nE_{rq} corresponds to the expected number of observations in cell (r, q). If the individual classes of x values, $q = 1, 2, \ldots, nX$, have equal medians, then $\hat{T}$ approximately follows a chi-square distribution with $(nX - 1)(2 - 1) = nX - 1$ degrees of freedom (p. 156, Conover, 1980). Thus, the probability of obtaining a value of T that exceeds $\hat{T}$ in the presence of equal medians is given by $prob(T > \hat{T} | nX - 1)$ in Eq. (5.4). To maintain the validity of the chi-square approximation in the analysis of contingency tables, Conover suggests using a partition in which $nE_{rq} \geq 1$ (p. 156, Conover, 1980).

The Kruskal-Wallis rank statistic (Sect. 5.2) also converges to the chi-square statistic with $nX - 1$ degrees of freedom. In a case study in Conover (1980, p. 232), the power of the rank test (i.e., Kruskal-Wallis) exceeded the power of the median test. We interpret this result as follows: the median test only measures whether observations exceed the common median; it does not measure the extent to which individual observations exceed this median (i.e., nominal versus ordinal scale). Thus, the rank test is incorporating more information than the median test.

Intentionally Left Blank

6. Dispersion of y Dependent on x

The preceding section considered techniques for identifying patterns where some measure of central tendency for y changes with changing values for x. In this section, techniques for identifying patterns that involve changes in the dispersion or spread of y with changing values for x are considered. Two measures of dispersion will be considered: the variance σ_y^2 of y, and the interquartile range $y_{0.75} - y_{0.25}$ of y, where $y_{0.75}$ and $y_{0.25}$ represent the 0.75 and 0.25 quantiles of y. The variance is the best known measure of dispersion, and the interquantile range is widely used as a summary of dispersion in box plots (e.g., Helton et al., 1996; Sargent, 1996). The interquartile range is less sensitive to outliers than the variance, as is the case in the analogous situation for medians and means. Two statistics are considered: the ANOVA F statistic with jackknifing in tests for common variances, and the chi-square statistic in conjunction with contingency tables in tests for common interquartile ranges.

6.1 Common Variances: ANOVA F Statistic with Jackknifing

The ANOVA test will use the same classes, $q = 1, 2, \ldots, nX$, of x values introduced in Sect. 5 (Fig. 5.1). Many procedures exist for testing for common variances: five procedures are summarized in Kleijnen (1987, pp. 225 - 227), and 56 procedures are examined in Conover et al. (1981). Additional discussion is also given in Conover (1980, pp. 239 - 250), Hamby (1994, pp. 149 - 150), Piepho (1997) and Wludyka and Nelson (1997). Note that common variances can occur even though the associated mean values are different.

For this analysis, a procedure based on jackknifing is used to indicate if different classes of x values have different variances for y. Jackknifing is a general technique for reducing possible bias in estimators and constructing robust confidence intervals (Efron, 1982; Miller, 1974). The first author has obtained good results with jackknifing in a number of different applications (Kleijnen, 1998). The procedure operates as follows.

The variance σ_{yq}^2 of y conditional on class q is estimated by

$$\hat{\sigma}_{yq}^2 = \sum_{k \in \mathcal{X}_q} (y_k - \bar{y}_q)^2 / (nX_q - 1) \tag{6.1}$$

for $q = 1, 2, \ldots, nX$, where $\mathcal{X}_q$, $\bar{y}_q$ and nX_q are defined in conjunction with Eq. (5.1). Further, an additional nX_q estimators

$$\hat{\sigma}_{yq,-l}^2 = \sum_{\substack{k \in \mathcal{X}_q \\ k \neq l}} (y_k - \bar{y}_{q,-l})^2 / (nX_q - 2) \tag{6.2}$$

of σ^2_{yq} are calculated with individual y's (i.e., y_l) omitted from consideration. The values for $\hat{\sigma}^2_{yq}$ and $\hat{\sigma}^2_{yq,-l}$ can be used to define the so-called pseudo values

$$t_{ql} = nX_q\hat{\sigma}^2_{yq} - (nX_q - 1)\hat{\sigma}^2_{yq,-l}. \tag{6.3}$$

For each class of x values, the resultant values for t_{ql} constitute a sample from a population whose expected value is σ_{yq} in the case of common variances (at least if the x's were generated by random sampling). The ANOVA F test can now be used to test for the equality of the means of the variables t_{ql}. Specifically, the F statistic described in Eq. (5.1) is calculated with the values for t_{ql}, and the corresponding exceedance probability for the resultant F statistic is determined as indicated in Eq. (5.2). In this application, the jackknife procedure is being performed to obtain a particular statistical test and robust confidence interval estimates rather than to reduce bias.

Because variance estimators have long tails to the right, the use of a logarithmic transformation in conjunction with the jackknife procedure may enhance the capability of the procedure to identify different variances for y. Specifically, t_{ql} can be defined by

$$t_{ql} = nX_q \ln(\hat{\sigma}^2_{yq}) - (nX_q - 1)\ln(\hat{\sigma}^2_{yq,-l}), \tag{6.4}$$

and then the procedures defined in Eqs. (5.1) and (5.2) used with this new definition. In this case, the test is for the equality of $\ln(\sigma^2_{yq})$.

A related approach is proposed by Archer et al. (1997), who also use the variability of y to assess the importance of factors in large-scale simulation models. Further, they use an ANOVA-like procedure to decompose the total variability of y into main effects, two factor effects, and higher-order interactions among factors. Finally, they apply bootstrapping, which is closely related to jackknifing.

6.2 Common Interquartiles: Chi-Square for Contingency Tables

A test for common interquartile ranges based on the chi-square statistic for contingency tables is now considered. This test is based on the previously used partitioning of x into q = 1, 2, ..., nX classes, with the hypothesis being that the associated nX interquartile ranges (i.e., $y_{0.75} - y_{0.25}$) are the same (Fig. 6.1). The quantile values $y_{0.25}$ and $y_{0.75}$ are defined by Eq. (5.5) with Q = 0.25 and 0.75. The individual classes of x values are now divided into subsets of y values that fall within and outside the interquartile range. For class q, let nX_{1q} equal the number of y values that fall within the interquartile range, and nX_{2q} equal the number of y values that fall outside the interquartile range. As for the common median test, the result of this partitioning is a $2 \times nX$ contingency table with

nX_{rq} observations in each cell. The statistic in Eq. (5.6) can now be calculated and used in conjunction with the exceedance probability in Eq. (5.4). The interquartile test was suggested by the quantile test mentioned in Conover (1980, p. 174) and, to the best of our knowledge, has not been previously examined in the literature.

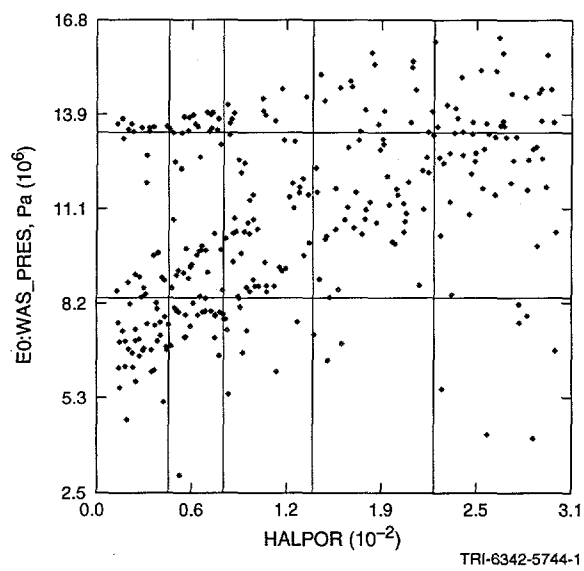

Fig. 6.1. Example of the partitioning of the range of x = *HALPOR* into $nX = 5$ classes and the range of y = *E0:WAS_PRES* into values inside and outside the interquartile range $[y_{0.25}, y_{0.75}]$.

Intentionally Left Blank

7. Distribution of y Dependent on x: Chi-Square for Contingency Tables

The two preceding sections considered procedures for determining if the central tendency of y was dependent on x (Sect. 5) and if the dispersion of y was dependent on x (Sect. 6). In this section, the chi-square test for contingency tables is introduced as a means of determining if the distribution of y is dependent on the distribution of x (i.e., to determine if y is statistically independent of x).

The test will use the same classes, $q = 1, 2, \ldots, nX$, of x values used in Sects. 5 and 6. Further, y is also divided into classes (Fig. 7.1). Thus, y must be defined on at least a nominal scale to permit the definition of the necessary classes. For notational convenience, let p, $p = 1, 2, \ldots, nY$, designate the individual classes into which the values of y have been divided; let $\mathcal{Y}_p$ designate the set such that $k \in \mathcal{Y}_p$ only if y_k belongs to class p; and let nY_p equal the number of elements contained in $\mathcal{Y}_p$. Typically, y is defined on at least an ordinal scale, and the classes are defined by ordering the y and then requiring the individual classes to have similar numbers of elements (i.e., the nY_p are approximately equal for $p = 1, 2, \ldots, nY$).

The partitioning of x and y into nX and nY classes in turn partitions (x, y) into $nX\ nY$ classes (Fig. 7.1), where (x_k, y_k) belongs to class (q, p) only if x_k belongs to class q of the x values (i.e., $k \in \mathcal{X}_q$) and y_k belongs to class p of the y values (i.e., $k \in \mathcal{Y}_p$). For notational convenience, let O_{pq} denote the set such that $k \in O_{pq}$ only if $k \in \mathcal{X}_q$ (i.e., x_k is in class q of x values) and also $k \in \mathcal{Y}_p$ (i.e., y_k is in class p of y values), and let nO_{pq} equal the number of elements contained in O_{pq}. Further, if x and y are independent, then

$$nE_{pq} = (nY_p / nS)(nX_q / nS)nS = nY_p\ nX_q / nS \tag{7.1}$$

is an estimate of the expected number of observations (x_k, y_k) that should fall in class (q, p).

The following statistic can be defined:

$$\hat{T} = \sum_{q=1}^{nX} \sum_{p=1}^{nY} (nO_{pq} - nE_{pq})^2 / nE_{pq}, \tag{7.2}$$

which is the same as the statistic in Eq. (5.6) except for the upper limit on the inner summation. Asymptotically, $\hat{T}$ follows a chi-square distribution with $(nX-1)$ $(nY-1)$ degrees of freedom when x and y are independent. Thus, the probability of obtaining a value of T that exceeds $\hat{T}$ when x and y are independent is given by $prob(T > \hat{T} | (nX-1)(nY-1))$ in Eq. (5.4).

Many other measures can also be used to quantify the degree of dependence between two variables x and y: Cramer's contingency coefficient, Pearson's mean-square contingency coefficient, the phi coefficient, and so on (Conover, 1980, pp. 178-189). However, these techniques do not offer any advantages over the chi-square contingency table approach already discussed.

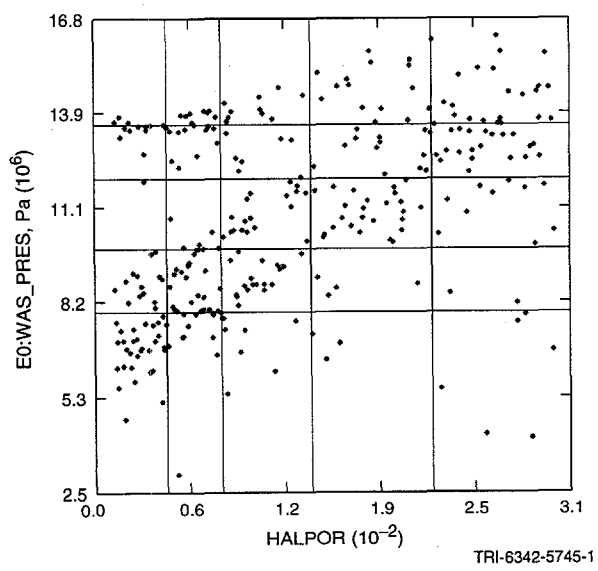

Fig. 7.1. Example of the partitioning of the range of x = *HALPOR* into nX = 5 classes and the range of y = *E0:WAS_PRES* into nY = 5 classes.

8. Identification of Important Variables

The purpose of the statistical procedures under consideration is to identify sampled variables that have significant effects on individual predicted variables. Conceptually, this is equivalent to identifying scatterplots that exhibit some form of deviation from randomness. Once such scatterplots are identified, the analysts' understanding of the model must be called upon to explain the patterns that appear in these plots.

To provide guidance in examining scatterplots, it is useful to have a numerical way to distinguish between variables that appear to have a substantial effect on a predicted outcome and variables that appear to have little or no effect. For a given statistic, the probability that a larger value would occur due to chance variation provides such a measure (i.e., the probabilities in Eqs. (3.16), (4.3), (5.2), (5.4)). These probabilities are often called critical values or p-values and designated by $\hat{\alpha}$ or p. A small critical value indicates that under the assumptions of the test, an outcome equal to or greater than the observed value of the statistic is unlikely to occur due to chance. Thus, the implication is that the pattern in the associated scatterplot arose from some underlying relationship between x and y rather than from chance alone. For a given statistic, the indicated importance of a variable goes up as the value of the corresponding critical value goes down. Thus, an ordering of variables on the basis of the size of their associated critical values provides a way to rank variable importance (i.e., the smaller the critical value, the more important the variable appears to be).

In sensitivity analyses of the type under consideration in this presentation, the distributions for the sampled variables typically characterize subjective (i.e., epistemic) uncertainty (Helton, 1997). Often, the intent of the sensitivity analysis is to identify those variables on which additional research efforts should be expended to reduce the uncertainty in the final outcomes of a large analysis and hence in the decisions based on these outcomes. In this case, the desire may not be to obtain an absolute ranking of variable importance, but rather to prioritize groups of variables for additional research. For example, variables might be divided into the following three groups: Group 1 - important variables that require additional investigation, Group 2 - variables of intermediate importance that may merit additional investigation if time and resources permit, and Group 3 - unimportant variables that do not require additional investigation. One possibility is to define these groups on the basis of critical values (e.g., Group 1 corresponds to variables with $\hat{\alpha} < 0.01$; Group 2 corresponds to variables with $0.01 \leq \hat{\alpha} < 0.05$; and Group 3 corresponds to variables with $0.05 \geq \hat{\alpha}$). However, in practice due to the cost of investigating individual variables, the decision on whether or not to expend resources on the investigation of a particular variable will probably be made on the basis of a number of considerations rather than solely on the basis of a preselected critical value.

Intentionally Left Blank

9. Top-Down Correlation

A number of techniques have been described for the identification of relationships between sampled and predicted variables (Sects. 3-7). These techniques will be applied to four predicted variables (Sect. 2). An important question is the extent to which the different techniques agree in their identification of important variables. A useful tool for assessing such agreement is the top-down correlation introduced by Iman and Conover (1987), which emphasizes agreement/disagreement for the most important variables and places reduced weight on agreement/disagreement for variables of little importance.

The top-down correlation is based on Savage scores:

$$S(h) = \sum_{j=h}^{nI} 1/j, \tag{9.1}$$

where $S(h)$ is the Savage score of a variable of rank h and nI is the number of ranked variables (Eq. (1.1)). Thus, the Savage score $S(1)$ for the most important variable is $S(1) = 1/1 + 1/2 + \ldots + 1/nI$; the Savage score $S(2)$ for the next most important variable is $S(2) = 1/2 + 1/3 + \ldots + 1/nI$; and so on.

Suppose two ranking procedures are under consideration. Further, let h_{1i}, $i = 1, 2, \ldots, nI$, denote the rank for variable x_i obtained with the first procedure, and let h_{2i}, $i = 1, 2, \ldots, nI$, denote the rank for variable x_i obtained with the second procedure. The top-down correlation $\hat{R}_{12}$ for these two tests is defined to be the Pearson correlation coefficient (Eq. (3.4)) associated with the pairs $[S(h_{1i}), S(h_{2i})]$, $i = 1, 2, \ldots, nI$. That is,

$$\hat{R}_{12} = \left[\sum_{i=1}^{nI} S(h_{1i})S(h_{2i}) - nI\right] \Big/ \left[nI - S(1)\right], \tag{9.2}$$

with $S(1)$ defined in Eq. (9.1) and approximately equal to $2.450 + ln[nI + 0.5)/6.5]$ for $nI \geq 7$ (Eq. (3), Iman and Conover, 1987). Large positive values for R_{12} indicate agreement between the two sets of ranks for the most important factors. Exact quantiles for this statistic are given in Iman and Conover (1987, p. 355; also see Iman 1987). Further,

$$z = \hat{R}_{12}\sqrt{nI - 1} \tag{9.3}$$

is distributed approximately normally with mean 0 and standard deviation 1 when the two rankings are uncorrelated and nI is sufficiently large. Under these conditions,

$$prob\left(R > \hat{R}_{12}\right) = 0.5\, erfc\left(\hat{R}_{12}\sqrt{nI-2}/\sqrt{2}\right), \tag{9.4}$$

where $prob\left(R > \hat{R}_{12}\right)$ is the probability that random variation would produce a value R for $\hat{R}_{12}$ larger than the observed value $\hat{R}_{12}$ and *erfc* is the complementary error function (Press et al., 1992, p. 631).

10. Comparison of Procedures for Identification of Important Variables

The following entities and/or associated statistical tests have been introduced for possible use in the identification of patterns in scatterplots, where the given capital letters will be used to identify the associated procedures in the following discussion: correlation coefficients (CCs, Sect. 3), standardized regression coefficients (SRCs, Sect. 3), partial correlation coefficients (PCCs, Sect. 3), rank correlation coefficients (RCCs, Sect. 4), standardized rank regression coefficients (SRRCs, Sect. 4), partial rank correlation coefficients (PRCCs, Sect. 4), common means (CMNs, Sect. 5.1), common locations (CLs, Sect. 5.2), common medians (CMDs, Sect. 5.3), common variances (CVs, Sect. 6.1), common interquartile ranges (CIQ, Sect. 6.2), and statistical independence (SI, Sect. 7). Further, dependent variables with different behaviors have been introduced as examples: *E0:WAS_PRES*, *E0:BRAALIC*, *E2:WAS_SATB*, and *E2:WAS_PRES* (Sect. 2). The results of applying the indicated procedures to these dependent variables are now discussed.

10.1 Repository Pressure under Undisturbed Conditions: y = *E0:WAS_PRES*

The variable y = *E0:WAS_PRES* was included as an example because a linear relationship appears to exist between *E0:WAS_PRES* and several of the sampled variables (Sect. 2). Thus, procedures that can identify linear relationships should work well with *E0:WAS_PRES*, as indeed turned out to be the case (Table 10.1). In particular, tests based on CCs, RCCs, CMNs, CLs, CMDs and SI both identified the same top four variables (i.e., *WMICDFLG HALPOR*, *WGRCOR*, *ANHPRM*) and also assigned these variables the same importance rankings based on p-values. The scatterplots for these variables show a corresponding decrease in the strength of the relationships with *E0:WAS_PRES* (Fig. 2.3). After these four variables, there was little agreement between the individual procedures on the remaining variables, with the p-values for the variables with ranks 5 and above typically close to or above 0.1. The only exception to this was for the test for SI, where *ANHBCVGP* was assigned rank 5 with a p-value of 0.0194. Based on a visual inspection, there appears to be little difference in the distributions of *E0:WAS_PRES* for the two values of *ANHBCVGP*, although the larger value for *ANHBCVGP* (i.e., the value that implies the van Genuchten-Parker model) may result in fewer small values for *E0:WAS_PRES* (Fig. 10.1) The performance of the tests based on measures of dispersion (i.e., CV, CIQ) was disappointing, with CV failing to identify both *HALPOR* and *WGRCOR* based on a p-value cutoff of 0.1 and CIQ failing to identify *WGRCOR* based on the same cutoff.

As discussed in Sect. 3, analyses of variable importance based on CCs, SRCs and PCCs or on RCCs, SRRCs and PRCCs will produce similar results when the independent variables (i.e., the x_i's) are uncorrelated. More specifically, CCs and SRCs are equal; RCCs and SRRCs are equal; orderings of variable importance based on CCs, SRCs and PCCs are the same; and orderings of variable importance based on RCCs, SRRCs and PRCCs are the same. The 24 variables used in the calculation of *E0:WAS_PRES* were assumed to be independent, with the Iman

Table 10.1 Comparison of Variable Rankings with Different Analysis Procedures[a] for y = *E0:WAS_PRES*, the Variables in Table 2.2,[b] and a Maximum of Five Classes of Values for Each Variable (i.e., nX =5)[c]

Variable[d]	CC		RCC		CMN: 1 × 5		CL: 1 × 5	
Name	Rank	*p*-Val	Rank	*p*-Val	Rank	*p*-Val	Rank	*p*-Val
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
HALPOR	2.0	0.0000	2.0	0.0000	2.0	0.0000	2.0	0.0000
WGRCOR	3.0	0.0000	3.0	0.0000	3.0	0.0000	3.0	0.0000
ANHPRM	4.0	0.0241	4.0	0.0268	4.0	0.0195	4.0	0.0187
SALPRES	5.0	0.0855	5.0	0.0664	13.0	0.6283	13.0	0.5672
WGRMICI	17.0	0.7753	20.0	0.8940	23.0	0.9705	23.0	0.9649
SHPRMCON	18.0	0.7878	18.0	0.8618	10.0	0.4099	11.0	0.4878
ANHBCVGP	20.0	0.8084	15.0	0.7686	18.0	0.8062	16.0	0.7686

Variable	CMD: 2 × 5		CV: 1 × 5		CIQ: 2 × 5		SI: 5 × 5	
Name	Rank	*p*-Val	Rank	*p*-Val	Rank	*p*-Val	Rank	*p*-Val
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
HALPOR	2.0	0.0000	12.0	0.3919	2.0	0.0000	2.0	0.0000
WGRCOR	3.0	0.0025	4.0	0.1244	16.5	0.6626	3.0	0.0003
ANHPRM	4.0	0.0663	2.0	0.0042	3.0	0.0007	4.0	0.0049
SALPRES	9.0	0.4932	11.0	0.3723	6.0	0.0868	21.0	0.7554
WGRMICI	24.0	0.9702	24.0	0.8900	5.0	0.0595	13.0	0.3239
SHPRMCON	6.0	0.2674	5.0	0.1287	4.0	0.0244	7.0	0.1487
ANHBCVGP	14.0	0.6442	13.0	0.4752	24.0	1.0000	5.0	0.0194

[a] Table includes only variables that had a *p*-value less than 0.1 for at least one of the procedures under consideration although the variable rankings for a specific procedure are based on the *p*-values obtained for that procedure for all 24 variables included in the analysis (See Footnote b).

[b] Table 2.2 contains 27 variables but *BHPRM* was not used in the calculation of E0 results (i.e., *E0:WAS_PRES* and *E0:BRAALIC*) and the variables in the pairs (*ANHPRM, ANHCOMP*), (*HALPRM, HALCOMP*) have a −0.99 rank correlation. As a result, *BHPRM, ANHCOMP* and *HALCOMP* were not included in the analysis, which resulted in 24 variables (i.e., x's) for analysis with each procedure.

[c] Variables *ANHBCVGP, WMICDFLG* in Table 2.2 are discrete with 2, 3 levels, respectively; for these variables, nX = 2, 3. Also, nY = 5 for SI.

[d] Variables are listed in the table based on their ordering with the *p*-values obtained for CCs; thus, the listed rankings for CCs will monotonically increase, which will not in general be the case for the other procedures.

and Conover (1982) restricted pairing technique being used to assure that the correlations between variables were indeed close to zero (see Footnote b to Table 10.1). The outcome, as predicted by theory, was that CCs and SRCs were approximately equal, RCCs and SRRCs were approximately equal, rankings based on CCs, SRCs and PCCs were approximately the same, and rankings based on RCCs, SRRCs and PRCCs were approximately the same (Table 10.2). Approximate correspondence to theory is the best that can be hoped for as the Iman/Conover restricted pairing technique makes the correlations between the sampled variables approximately zero (Table 10.3) rather than exactly zero as is required for strict validity of the theoretical relationships.

Table 10.2 Comparison of Variable Rankings with Correlation Coefficients, Standardized Regression Coefficients and Partial Correlation Coefficients with Raw and Rank Transformed Data for *y = E0:WAS_PRES*

Variable[a]		CC		SRC		PCC	
Name	*p*-Val	Rank	Value	Rank	Value	Rank	Value
WMICDFLG	0.0000	1.0	0.7124	1.0	0.7234	1.0	0.8642
HALPOR	0.0000	2.0	0.4483	2.0	0.4651	2.0	0.7469
WGRCOR	0.0000	3.0	0.2762	3.0	0.2460	3.0	0.5113
ANHPRM	0.0241	4.0	0.1302	4.0	0.1277	4.0	0.2953
SALPRES	0.0855	5.0	0.0993	6.0	0.0639	6.0	0.1526

Variable[b]		RCC		SRRC		PRCC	
Name	*p*-Val	Rank	Value	Rank	Value	Rank	Value
WMICDFLG	0.0000	1.0	0.7229	1.0	0.7207	1.0	0.8564
HALPOR	0.0000	2.0	0.4521	2.0	0.4511	2.0	0.7256
WGRCOR	0.0000	3.0	0.2608	3.0	0.2303	3.0	0.4739
ANHPRM	0.0268	4.0	0.1280	4.0	0.1093	4.0	0.2476
SALPRES	0.0664	5.0	0.1062	5.0	0.0723	5.0	0.1667

[a] Comparison based on variables that had a *p*-value less than 0.1 for CCs. Ranks based on values for CCs, SRCs, PCCs in column "VALUE."

[b] Comparison based on variables that had a *p*-value less than 0.1 for RCCs. Ranks based on values for RCCs, SRRCs, PRCCs in column "VALUE."

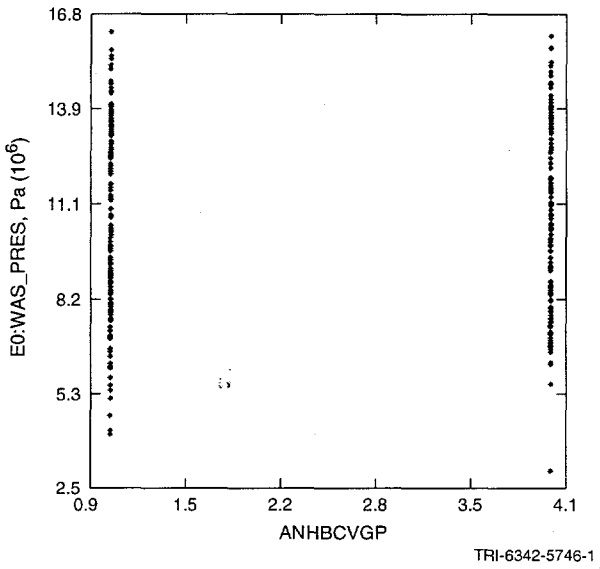

Fig. 10.1 Scatterplot for *E0:WAS_PRES* versus *ANHBCVGP*.

Table 10.3 Correlations with Raw and Rank Transformed Data between *WMICDFLG*, *HALPOR*, *WGRCOR*, *ANHPRM* and *SALPRES*

Raw Data

HALPOR	–0.035			
WGRCOR	0.027	0.022		
ANHPRM	0.001	–0.004	0.013	
SALPRES	0.056	–0.007	0.001	–0.012
	WMICDFLG	*HALPOR*	*WGRCOR*	*ANHPRM*

Rank-Transformed Data

HALPOR	–0.008			
WGRCOR	0.031	0.014		
ANHPRM	0.018	0.005	0.021	
SALPRES	0.053	–0.010	0.001	0.004
	WMICDFLG	*HALPOR*	*WGRCOR*	*ANHPRM*

The large number of procedures under consideration (i.e., CCs, RCCs, CMNs, CLs, CMDs, CVs, CIQs, SI, SRCs, PCCs, SRRCs, PRCCs) can make it difficult to get an overall feeling for the extent to which the individual procedures are agreeing or disagreeing in the identification of important variables. As discussed in Sect. 9, top-down correlation provides a way to compare variable rankings that emphasizes agreement and/or disagreement on the most important variables and places reduced importance on the rankings assigned to less important variables. In particular, top-down correlation gives a compact numeric summary of the comparisons in Tables 10.1 and 10.2 (Table 10.4), with all procedures except for CVs and CIQs showing strong agreement (i.e., top-down correlations close to or equal to one).

The calculation of CMNs, CLs, CMDs, CVs, CIQs and SI in Table 10.1 was based on the division of the range of the variables under consideration into $nX = 5$ intervals of equal probability. Also, the calculation of SI involved the division of the range of *E0:WAS_PRES* into $nY = 5$ intervals of equal probability. In concept, the outcome of the analysis could be quite sensitive to the partitioning selected for use (i.e., the values for nX and nY). To check this, the analysis was repeated with $nX = 10$ for CMs, CLs, CMDs, CVs, CIQs and SI, and $nY = 10$ for SI (Table 10.5). As comparison of the results obtained with $nX = 5$ and $nX = 10$ shows, some changes in variable rankings did take place. For CMNs, CLs and CMDs with $nX = 5$, *ANHPRM* is the fourth ranked variable with p-values of 0.0195, 0.0187 and 0.0663, respectively (Table 10.1); for the same procedures with $nX = 10$, *ANHPRM* is ranked 4, 4 and 7 with p-values of 0.1371, 0.1340 and 0.3398 (Table 10.5). For CVs and CIQs, there are some changes in variable ranking (e.g., CV and CIQ assign *SALPRES* ranks of 11 and 6 with p-values of 0.3723 and 0.0868 for $nX = 5$ (Table

Table 10.4 Top-Down Correlation Matrix for Variable Rankings with Different Analysis Procedures for y = *E0:WAS_PRES*, Variables included in Table 10.1,[a] and a Maximum of Five Classes of x values (i.e., $nX = 5$)

Top-Down Correlation Matrix

	CC	RCC	CMN	CL	CMD	CV	CIQ	SI	SRC	PCC	SRRC
RCC	0.982										
CMN	0.972	0.981									
CL	0.972	0.981	1.000								
CMD	0.972	0.981	1.000	1.000							
CV	0.731	0.740	0.769	0.769	0.769						
CIQ	0.860	0.831	0.872	0.872	0.872	0.705					
SI	0.946	0.967	0.972	0.972	0.972	0.720	0.839				
SRC	0.986	0.996	0.967	0.967	0.967	0.719	0.824	0.963			
PCC	0.986	0.996	0.967	0.967	0.967	0.719	0.824	0.963	1.000		
SRRC	0.996	0.986	0.963	0.963	0.963	0.715	0.840	0.951	0.995	0.995	
PRCC	0.996	0.986	0.963	0.963	0.963	0.715	0.840	0.951	0.995	0.995	1.000

Top-Down Correlation Matrix p Values

	CC	RCC	CMN	CL	CMD	CV	CIQ	SI	SRC	PCC	SRRC
RCC	0.005										
CMN	0.005	0.005									
CL	0.005	0.005	0.004								
CMD	0.005	0.005	0.004	0.004							
CV	0.026	0.025	0.021	0.021	0.021						
CIQ	0.011	0.014	0.011	0.011	0.011	0.031					
SI	0.006	0.005	0.005	0.005	0.005	0.028	0.013				
SRC	0.005	0.004	0.005	0.005	0.005	0.029	0.015	0.005			
PCC	0.005	0.004	0.005	0.005	0.005	0.029	0.015	0.005	0.004		
SRRC	0.004	0.005	0.005	0.005	0.005	0.029	0.013	0.006	0.004	0.004	
PRCC	0.004	0.005	0.005	0.005	0.005	0.029	0.013	0.006	0.004	0.004	0.004

[a] Variable rankings used in calculation of top-down correlation are based on only the 8 variables included in Table 10.1. Specifically, each procedure was used to rank these 8 variables from 1 to 8 (i.e., p-values for CCs, RCCs, CMNs, CLs, CMDs, CVs, CIQs, SI; absolute values of coefficients for SRCs, PCCs, SRRCs, PRCCs); then, top-down correlations were calculated on these rankings.

10.1) and rank *SALPRES* third with p-values of 0.0500 and 0.0077 for $nX = 10$ (Table 10.5)); also, CVs still fail to identify an effect for *HALPOR* (ranked 12 with a p-value of 0.3919 for $nX = 5$ and ranked 20 with a p-value of 0.5800 for $nX = 10$), and CIQs still fail to identify an effect for *WGRCOR* (ranked 16.5 with a p-value of 0.6626 for $nX = 5$ and ranked 23 with a p-value of 0.9429 for $nX = 10$). For SI, *HALPRM* had a rank of 18 with a p-value of 0.6235 for $nX = 5$ and a rank of 3 with a p-value of 0.0036 for $nX = 10$ (Table 10.5). Thus, the partitioning in use can have an effect on the variables identified as affecting the y value under consideration. For perspective, the top-

Table 10.5 Comparison of Variable Rankings for y = *E0:WAS_PRES* Obtained with a Maximum of Five Classes of x Values (i.e., nX = 5) and Analytic Determination of p-values with Variable Rankings Obtained with (i) a Maximum of Ten Classes of x values (i.e., nX = 10) and Analytic Determination of p-values and (ii) a Maximum of Five Classes of x values (i.e., nX = 5) and Monte Carlo Determination of p-values

Variable[a] Name	CMN: 1 × 5[b] Rank	p-Val	CMN: 1 × 10[c] Rank	p-Val	CMNMC: 1 × 5[d] Rank	p-Val
WMICDFLG	1.0	0.0000	1.0	0.0000	2.0	0.0000
HALPOR	2.0	0.0000	2.0	0.0000	2.0	0.0000
WGRCOR	3.0	0.0000	3.0	0.0000	2.0	0.0000
ANHPRM	4.0	0.0195	4.0	0.1371	4.0	0.0214
SHPRMASP	5.0	0.1439	11.0	0.5087	5.0	0.1495
WRBRNSAT	6.0	0.1506	6.0	0.1947	6.0	0.1526
SHRGSSAT	7.0	0.2488	15.0	0.7062	7.0	0.2497
ANRBRSAT	8.0	0.3034	18.0	0.7693	8.0	0.3027
HALPRM	9.0	0.4097	8.0	0.4092	10.0	0.4060
SHPRMCON	10.0	0.4099	12.0	0.5115	9.0	0.4053
SHRBRSAT	11.0	0.4325	10.0	0.4560	11.0	0.4239
WFBETCEL	12.0	0.5694	7.0	0.4034	12.0	0.5645
SALPRES	13.0	0.6283	20.0	0.8300	13.0	0.6378
ANHBCEXP	14.0	0.7116	16.0	0.7465	14.0	0.7035
WASTWICK	15.0	0.7490	22.0	0.8444	15.0	0.7446
ANRGSSAT	16.0	0.7521	13.0	0.6511	16.0	0.7483
SHPRMHAL	17.0	0.7661	14.0	0.6734	17.0	0.7699
ANHBCVGP	18.0	0.8062	19.0	0.8062	18.0	0.7997
SHBCEXP	19.0	0.8100	21.0	0.8342	19.0	0.8099
WRGSSAT	20.0	0.8358	5.0	0.1542	20.0	0.8377
SHPRMCLY	21.0	0.8601	9.0	0.4218	21.0	0.8625
SHPRMDRZ	22.0	0.8726	17.0	0.7562	22.0	0.8755
WGRMICI	23.0	0.9705	23.0	0.9606	23.0	0.9717
WGRMICH	24.0	0.9975	24.0	0.9919	24.0	0.9973
TDC[e]	1.000		0.854		0.970	

Variable Name	CL: 1 × 5 Rank	p-Val	CL: 1 × 10 Rank	p-Val	CLMC: 1 × 5 Rank	p-Val
WMICDFLG	1.0	0.0000	1.0	0.0000	2.0	0.0000
HALPOR	2.0	0.0000	2.0	0.0000	2.0	0.0000
WGRCOR	3.0	0.0000	3.0	0.0002	2.0	0.0000
ANHPRM	4.0	0.0187	4.0	0.1340	4.0	0.0212
SHPRMASP	5.0	0.1237	9.0	0.4376	5.0	0.1277
WRBRNSAT	6.0	0.2042	7.0	0.2838	6.0	0.2053
ANRBRSAT	7.0	0.2710	16.0	0.7391	7.0	0.2710
SHRGSSAT	8.0	0.3153	17.0	0.7495	8.0	0.3167
HALPRM	9.0	0.3923	6.0	0.2725	9.0	0.3901
SHRBRSAT	10.0	0.4625	12.0	0.5456	10.0	0.4575
SHPRMCON	11.0	0.4878	11.0	0.4655	11.0	0.4852
WFBETCEL	12.0	0.5194	8.0	0.3728	12.0	0.5153
SALPRES	13.0	0.5672	20.0	0.8266	13.0	0.5817
SHPRMHAL	14.0	0.6945	13.0	0.5517	14.0	0.6996
SHBCEXP	15.0	0.7390	21.0	0.8301	15.0	0.7399
ANHBCVGP	16.0	0.7686	19.0	0.7686	16.0	0.7654
ANHBCEXP	17.0	0.7703	18.0	0.7594	17.0	0.7658
ANRGSSAT	18.0	0.8272	15.0	0.7298	18.0	0.8209
WASTWICK	19.0	0.8318	22.0	0.8443	19.0	0.8292
WRGSSAT	20.0	0.8826	5.0	0.2088	20.0	0.8839
SHPRMDRZ	21.0	0.8897	14.0	0.7065	21.0	0.8937
SHPRMCLY	22.0	0.9032	10.0	0.4426	22.0	0.9062
WGRMICI	23.0	0.9649	23.0	0.9691	23.0	0.9663
WGRMICH	24.0	0.9865	24.0	0.9894	24.0	0.9839
TDC[e]	1.000		0.861		0.971	

Variable Name	CMD: 2 × 5 Rank	p-Val	CMD: 2 × 10 Rank	p-Val	CMDMC: 2 × 5 Rank	p-Val
WMICDFLG	1.0	0.0000	1.0	0.0000	1.5	0.0000
HALPOR	2.0	0.0000	2.0	0.0000	1.5	0.0000
WGRCOR	3.0	0.0025	3.0	0.0124	3.0	0.0018
ANHPRM	4.0	0.0663	7.0	0.3398	4.0	0.0690
SHPRMASP	5.0	0.2427	14.0	0.6302	5.0	0.2401
SHPRMCON	6.0	0.2674	9.0	0.3725	6.0	0.2718
ANRBRSAT	7.0	0.3386	18.5	0.7532	7.0	0.3329
HALPRM	8.0	0.3883	8.0	0.3614	8.0	0.3967
SALPRES	9.0	0.4932	18.5	0.7532	9.0	0.5058
WRBRNSAT	10.0	0.5037	13.0	0.6163	10.0	0.5180
WRGSSAT	11.0	0.5249	5.0	0.1596	11.0	0.5223
SHRGSSAT	12.0	0.6151	23.0	0.9281	13.0	0.6050
ANHBCEXP	13.0	0.6387	11.0	0.5075	14.0	0.6224
ANHBCVGP	14.0	0.6442	16.0	0.6442	12.0	0.5685
SHRBRSAT	15.0	0.6868	15.0	0.6441	15.0	0.6950
SHPRMDRZ	16.5	0.7358	10.0	0.4311	17.0	0.7283
WFBETCEL	16.5	0.7358	17.0	0.7265	16.0	0.7169
SHPRMCLY	18.0	0.7847	6.0	0.3293	18.0	0.7659
SHPRMHAL	19.0	0.8325	4.0	0.1177	19.0	0.8357
SHBCEXP	21.0	0.9197	22.0	0.9114	22.0	0.9093
ANRGSSAT	21.0	0.9197	12.0	0.5887	20.0	0.9082
WASTWICK	21.0	0.9197	20.0	0.8729	21.0	0.9085
WGRMICH	23.0	0.9554	21.0	0.8930	23.0	0.9439
WGRMICI	24.0	0.9702	24.0	0.9835	24.0	0.9664
TDC	1.000		0.768		0.986	

Variable Name	CV: 1 × 5 Rank	p-Val	CV: 1 × 10 Rank	p-Val	CVMC: 1 × 5 Rank	p-Val
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000
ANHPRM	2.0	0.0042	2.0	0.0172	2.0	0.0031
HALPRM	3.0	0.1184	4.0	0.0844	4.0	0.1095
WGRCOR	4.0	0.1244	6.0	0.1173	3.0	0.1094
SHPRMCON	5.0	0.1287	5.0	0.0929	5.0	0.1201
SHRGSSAT	6.0	0.1466	15.0	0.4691	7.0	0.1411
ANHBCEXP	7.0	0.1539	7.0	0.1928	6.0	0.1393
SHPRMASP	8.0	0.1612	8.0	0.2953	8.0	0.1517
SHPRMCLY	9.0	0.3102	9.0	0.3614	9.0	0.2957
SHBCEXP	10.0	0.3221	11.0	0.4091	10.0	0.3049
SALPRES	11.0	0.3723	3.0	0.0500	11.0	0.3564
HALPOR	12.0	0.3919	20.0	0.5800	12.0	0.3817
ANHBCVGP	13.0	0.4752	16.0	0.4752	13.0	0.4800
WGRMICH	14.0	0.5612	13.0	0.4415	14.0	0.5502
SHPRMDRZ	15.0	0.6067	21.0	0.8635	15.0	0.5942
WASTWICK	16.0	0.6185	10.0	0.3697	16.0	0.6053
WRGSSAT	17.0	0.6398	14.0	0.4670	17.0	0.6237
WRBRNSAT	18.0	0.6632	17.0	0.5542	18.0	0.6588
ANRGSSAT	19.0	0.6761	12.0	0.4391	19.0	0.6666
WFBETCEL	20.0	0.7531	23.0	0.9435	20.0	0.7443
SHRBRSAT	21.0	0.8197	18.0	0.5606	21.0	0.8109
SHPRMHAL	22.0	0.8340	24.0	0.9844	22.0	0.8224
ANRBRSAT	23.0	0.8378	19.0	0.5700	23.0	0.8330
WGRMICI	24.0	0.8900	22.0	0.9219	24.0	0.8854
TDC	1.000		0.892		0.993	

Table 10.5 (Cont.)

Variable Name	CIQ: 2 × 5 Rank	CIQ: 2 × 5 *p*-Val	CIQ: 2 × 10 Rank	CIQ: 2 × 10 *p*-Val	CIQMC: 2 × 5 Rank	CIQMC: 2 × 5 *p*-Val
WMICDFLG	1.0	0.0000	1.0	0.0000	1.5	0.0000
HALPOR	2.0	0.0000	2.0	0.0000	1.5	0.0000
ANHPRM	3.0	0.0007	4.0	0.0112	3.0	0.0005
SHPRMCON	4.0	0.0244	5.0	0.1005	4.0	0.0279
WGRMICI	5.0	0.0595	7.0	0.1719	5.0	0.0565
SALPRES	6.0	0.0868	3.0	0.0077	6.0	0.0893
SHPRMHAL	7.0	0.1801	10.0	0.2993	7.0	0.1729
SHPRMDRZ	8.0	0.1801	14.0	0.4944	8.0	0.1789
WGRMICH	9.0	0.2548	9.0	0.2133	9.0	0.2547
SHPRMASP	10.0	0.3232	8.0	0.1849	10.0	0.3172
SHRGSSAT	11.0	0.3232	13.0	0.4559	11.0	0.3317
SHBCEXP	12.0	0.5249	17.0	0.6441	12.0	0.5281
WASTWICK	13.0	0.5467	12.0	0.4559	13.0	0.5356
WFBETCEL	14.0	0.5918	20.5	0.8729	14.0	0.5948
SHPRMCLY	15.0	0.6387	6.0	0.1426	15.0	0.6264
WGRCOR	16.5	0.6626	23.0	0.9429	16.0	0.6746
WRBRNSAT	16.5	0.6626	20.5	0.8729	17.0	0.6814
HALPRM	18.5	0.6868	11.0	0.3725	19.0	0.7063
ANHBCEXP	18.5	0.6868	16.0	0.6163	18.0	0.7021
ANRGSSAT	20.0	0.7113	18.0	0.6718	20.0	0.7120
WRGSSAT	21.0	0.8557	15.0	0.5887	21.0	0.8508
ANRBRSAT	22.0	0.9197	22.0	0.8930	23.0	0.9122
SHRBRSAT	23.0	0.9554	19.0	0.7265	24.0	0.9426
ANHBCVGP	24.0	1.0000	24.0	1.0000	22.0	0.9010
TDC	1.000		0.917		0.987	

Variable Name	SI: 5 × 5 Rank	SI: 5 × 5 *p*-Val	SI: 10 × 10 Rank	SI: 10 × 10 *p*-Val	SIMC: 5 × 5 Rank	SIMC: 5 × 5 *p*-Val
WMICDFLG	1.0	0.0000	1.0	0.0000	1.5	0.0000
HALPOR	2.0	0.0000	2.0	0.0000	1.5	0.0000
WGRCOR	3.0	0.0003	4.0	0.0073	3.0	0.0003
ANHPRM	4.0	0.0049	5.0	0.0128	4.0	0.0038
ANHBCVGP	5.0	0.0194	8.0	0.1271	5.0	0.0178
WRGSSAT	6.0	0.1229	12.0	0.2786	6.0	0.1196
SHPRMCON	7.0	0.1487	6.0	0.0326	7.0	0.1529
WASTWICK	8.0	0.1850	21.5	0.8743	8.0	0.1829
SHBCEXP	9.0	0.2436	17.0	0.6767	9.0	0.2441
SHPRMHAL	10.0	0.2518	10.0	0.2028	10.0	0.2540
SHPRMASP	11.0	0.2601	11.0	0.2623	11.0	0.2673
SHPRMDRZ	12.0	0.3142	7.0	0.1129	12.0	0.3205
WGRMICI	13.0	0.3239	16.0	0.6363	13.0	0.3252
ANHBCEXP	14.0	0.3438	9.0	0.1768	14.0	0.3472
WFBETCEL	15.0	0.3856	21.5	0.8743	15.0	0.3905
SHRBRSAT	16.0	0.4299	15.0	0.5527	16.0	0.4308
ANRBRSAT	17.0	0.4765	20.0	0.7701	17.0	0.4725
HALPRM	18.0	0.6235	3.0	0.0036	18.0	0.6307
SHRGSSAT	19.0	0.6482	19.0	0.7525	19.0	0.6587
WRBRNSAT	20.0	0.6849	18.0	0.7343	20.0	0.6981
SALPRES	21.0	0.7554	13.0	0.3310	21.0	0.7662
SHPRMCLY	22.0	0.9265	23.0	0.9348	22.0	0.9305
WGRMICH	23.0	0.9437	24.0	0.9709	23.0	0.9429
ANRGSSAT	24.0	0.9763	14.0	0.5316	24.0	0.9791
TDC	1.000		0.812		0.988	

[a] Twenty-four (24) variables included in analysis; see Footnote b to Table 10.1.

[b] Variable rankings obtained with a maximum of five classes of x values (i.e., $nX = 5$) and analytic determination of p-values.

[c] Variable rankings obtained with a maximum of ten classes of x values (i.e., $nX = 10$) and analytic determination of p-values.

[d] Variable rankings obtained with a maximum of five classes of x values (i.e., $nX = 5$) and Monte Carlo determination of p-values.

[e] Top-down correlation with variable rankings obtained with a maximum of five classes of x values (i.e., $nX = 5$) and analytic determination of p-values.

down correlations for results obtained with the two griddings are also given in Table 10.5, with these correlations ranging from 0.854 for (CMN:1 × 5, CMN:1 × 10) to 0.917 for (CIQ:2 × 5, CIQ:2 × 10).

The p-values used to identify important variables in Tables 10.1, 10.2 and 10.5 are calculated with statistical assumptions that are not fully satisfied. For example, in the calculation of p-values for CCs, the sample from the x's consists of three pooled LHSs rather than a random sample (see Eqs. (2.1) - (2.3)), and neither the individual x's nor y = *E0:WAS_PRES* has a normal distribution. A Monte Carlo simulation can be used to assess if the use of formal statistical procedures to determine p-values is producing misleading results. Specifically, 10,000 samples of the form

$$(x_k, y_k),\ k = 1, 2, \ldots, 300, \tag{10.1}$$

can be generated by pairing the 300 values for x (i.e., the 300 values for the particular x under consideration contained in the samples in Eqs. (2.1) - (2.3)) with the 300 predicted values for y (i.e., the 300 values for y that resulted from the use of the sample elements in Eqs. (2.1) - (2.3)). The specific pairing algorithm used was to randomly and without replacement assign an x value to each y value, which is similar to bootstrapping (Efron and

Tibshirani, 1993) except that the sampling is being performed without replacement. This random assignment was repeated 10,000 times to produce 10,000 samples of the form in Eq. (10.1). For a given procedure (i.e., CCs, RCCs, CMNs, CLs, CMDs, CVs, CIQs or SI), each of the 10,000 samples can be used to calculate the value of the statistic used to determine the corresponding p-value. The resulting empirical distribution of the statistic can then be used to estimate the p-value for the statistic actually observed in the analysis. Comparison of the p-value obtained for a given set of statistical assumptions with the p-value obtained from the empirical distribution of the corresponding statistic provides an indication of the robustness of the variable rankings with respect to possible deviations from the assumptions underlying the formal statistical procedures described in Sects. 3-7 for the determination of p-values.

For *E0:WAS_PRES*, the rankings of variable importance with p-values obtained from formal statistical procedures (i.e., CC and RCC in Table 10.6 and CMN, CL, CMD, CV, CIQ and SI in Table 10.5) and the ranking of variable importance with p-values obtained from empirical distributions (i.e., CCMC and RCCMC in Table 10.6 and CMNMC, CLMC, CMDMC, CVMC, CIQMC and SIMC in Table 10.5) are very similar. The largest difference is in the assignment of tied ranks to the most important variables when the empirical distributions of p-values are used (e.g., use of statistical procedures with CCs results in *WMICDFLG*, *HALPOR* and *WGRCOR* being ranked 1, 2 and 3, and use of the empirical distribution of p-values results in these variables being ranked 2, 2 and 2). The tied ranks with the empirical distributions arise because a sample of size 10,000 was used to generate these distributions, with the result that 0.0001 is the smallest nonzero p-value that can be estimated. In contrast, much smaller nonzero probabilities can be estimated with the formal statistical procedures from Sects. 3-7. Overall, the similarity between the exact (i.e., statistically determined) and empirical p-values in Tables 10.5 and 10.6 is quite good, with the two determinations of p-values producing almost identical rankings of variable importance except for the very small (i.e., $< 10^{-4}$) p-values. The associated top-down correlations range from 0.970 for (CMN:1 × 5, CMNMC: 1 × 5) (Table 10.5) to 0.993 for (CV:1 × 5, CVMC: 1 × 5) (Table 10.5). For perspective, a top-down correlation of 0.971 results when 24 variables are under consideration, one procedure has ties (i.e., ranks of 2, 2, 2) on the variables assigned ranks of 1, 2, 3 by the other procedure, and identical ranks are assigned to all other variables (e.g., see (CL:2 × 5, CLMC:2 × 5) in Table 10.5 and (CC, CCMC), (RCC, RCCMC) in Table 10.6).

Approximate 100 $(1 - \alpha)$% confidence intervals for the empirically determined p-values are given by $p \pm x_{1-\alpha/2}\,[p(1-p)/n]^{1/2}$, where p is the estimated p-value, $n = 10{,}000$ is the sample size in use, and $x_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the normal distribution (e.g., 1.96 for a 95% confidence interval (pp. 99-100, Method C, Conover, 1980)). For example, the approximate 95% confidence interval for $p = 0.0815$ (see *SALPRES* for CCMC in Table 10.6) is 0.0815 ± 0.0054, with this interval including the statistically determined value of 0.0855. For most procedures, the 95% confidence intervals on the empirically determined p-values include the statistically determined p-values. The results for CVs tend to show less agreement between the formally and empirically estimated p-values than is the case for the other procedures.

Table 10.6 Comparison of Variable Rankings for $y = E0{:}WAS_PRES$ Obtained with Correlation Coefficients (CCs, RCCs) and Analytic Determination of p-values with Rankings Obtained with Monte Carlo Determination of p-values

Variable[a] Name	CC[b] Rank	CC[b] p-Val	CCMC[c] Rank	CCMC[c] p-Val	Variable Name	RCC Rank	RCC p-Val	RCCMC Rank	RCCMC p-Val
WMICDFLG	1.0	0.0000	2.0	0.0000	*WMICDFLG*	1.0	0.0000	2.0	0.0000
HALPOR	2.0	0.0000	2.0	0.0000	*HALPOR*	2.0	0.0000	2.0	0.0000
WGRCOR	3.0	0.0000	2.0	0.0000	*WGRCOR*	3.0	0.0000	2.0	0.0000
ANHPRM	4.0	0.0241	4.0	0.0236	*ANHPRM*	4.0	0.0268	4.0	0.0250
SALPRES	5.0	0.0855	5.0	0.0815	*SALPRES*	5.0	0.0664	5.0	0.0634
SHRGSSAT	6.0	0.1553	6.0	0.1551	*SHRGSSAT*	6.0	0.2322	6.0	0.2335
WASTWICK	7.0	0.2163	7.0	0.2200	*WFBETCEL*	7.0	0.2408	7.0	0.2469
SHRBRSAT	8.0	0.2226	8.0	0.2222	*WASTWICK*	8.0	0.2726	8.0	0.2758
ANHBCEXP	9.0	0.2369	9.0	0.2379	*SHRBRSAT*	9.0	0.3068	9.0	0.3056
WFBETCEL	10.0	0.2770	10.0	0.2832	*SHPRMASP*	10.0	0.4201	10.0	0.4291
SHPRMCLY	11.0	0.5213	11.0	0.5264	*ANHBCEXP*	11.0	0.4383	11.0	0.4352
HALPRM	12.0	0.5767	12.0	0.5761	*WRBRNSAT*	12.0	0.5519	12.0	0.5581
SHPRMASP	13.0	0.6041	13.0	0.6192	*HALPRM*	13.0	0.6412	13.0	0.6419
WRBRNSAT	14.0	0.6444	14.0	0.6465	*SHPRMCLY*	14.0	0.6812	14.0	0.6848
SHBCEXP	15.0	0.6831	15.0	0.6875	*ANHBCVGP*	15.0	0.7686	15.0	0.7654
ANRBRSAT	16.0	0.7237	16.0	0.7236	*SHBCEXP*	16.0	0.8486	16.0	0.8501
WGRMICI	17.0	0.7753	17.0	0.7772	*SHPRMDRZ*	17.0	0.8599	17.0	0.8596
SHPRMCON	18.0	0.7878	18.0	0.7878	*SHPRMCON*	18.0	0.8618	18.0	0.8644
SHPRMDRZ	19.0	0.7925	19.0	0.7990	*SHPRMHAL*	19.0	0.8710	19.0	0.8785
ANHBCVGP	20.0	0.8084	20.0	0.8016	*WGRMICI*	20.0	0.8940	20.0	0.8934
WRGSSAT	21.0	0.8251	21.0	0.8279	*WGRMICH*	21.0	0.9576	21.0	0.9559
ANRGSSAT	22.0	0.8834	22.0	0.8879	*WRGSSAT*	22.0	0.9848	22.0	0.9863
WGRMICH	23.0	0.9291	23.0	0.9247	*ANRBRSAT*	23.0	0.9964	23.0	0.9973
SHPRMHAL	24.0	0.9474	24.0	0.9459	*ANRGSSAT*	24.0	0.9991	24.0	0.9990
TDC		0.971			*TDC*		0.971		

[a] Twenty-four (24) variables included in analysis; see Footnote b to Table 10.1.

[b] Variable rankings obtained with analytic determination of p-values.

[c] Variable rankings obtained with Monte Carlo determination of p-values.

[d] Top-down correlation between variable rankings obtained with analytic and Monte Carlo determination of p-values.

A possible variant of the common means (CMNs) test is to use logarithmically transformed y-values rather than the original untransformed y-values (Sect. 5.1). Possible rationales for such a transformation are to reduce the effects of extreme values on the estimated mean and to transform y into a variable that more closely follows a normal distribution. For $y = E0{:}WAS_PRES$, use of the logarithmic transformation has little effect on the outcome of the analysis, with both raw and log-transformed y's resulting in the same assignment of ranks to the top four variables (i.e., *WMICDFLG, HALPOR, WGRCOR, ANHPRM*) (Tables 10.5, 10.7).

A possible variant of the common variances (CVs) test is to use t_{ql} as defined in Eq. (6.4) rather than as defined in Eq. (6.3). The rationale for the use of logarithms in Eq. (6.4) is to reduce the effects of extreme values and thus produce more stable variance estimators. For $y = E0{:}WAS_PRES$, use of t_{ql} as defined in Eq. (6.4) rather than Eq. (6.3) had little effect on the outcome of the analysis, with both definitions of t_{ql} resulting in the selection of *WMICDFLG*, *ANHPRM*, *HALPRM* and *WGRCOR* as the top four variables (Tables 10.5, 10.7). Further, both definitions are failing to identify the important effects associated with *HALPOR* (Fig. 2.3).

Table 10.7 Exceedance Probabilities (i.e., p-values) for Common Mean and Common Variance Tests Calculated with Use of Logarithms[a] for y = *E0:WAS_PRES*, the Variables in Table 2.2,b and a Maximum of Five Classes of Values for Each Variable (i.e., $nX = 5$)[c]

Variable Name	CMN: Log,1 × 5 Rank	CMN: Log,1 × 5 p-Val	CMNMC: Log,1 × 5 Rank	CMNMC: Log,1 × 5 p-Val
WMICDFLG	1.0	0.0000	2.0	0.0000
HALPOR	2.0	0.0000	2.0	0.0000
WGRCOR	3.0	0.0000	2.0	0.0000
ANHPRM	4.0	0.0085	4.0	0.0112

Variable Name	CV: Log,1 × 5 Rank	CV: Log,1 × 5 p-Val	CVMC: Log,1 × 5 Rank	CVMC: Log,1 × 5 p-Val
WMICDFLG	1.0	0.0000	1.0	0.0000
ANHPRM	2.0	0.0151	2.0	0.0100
WGRCOR	3.0	0.1051	3.0	0.0672
HALPRM	4.0	0.1116	4.0	0.0786

[a] Log y_k instead of y_k in Eq. (5.1) for common means (CMNs) and t_{ql} as defined in Eq. (6.4) rather than as defined in Eq. (6.3) for common variances (CVs); for each test; table contains variables with p-values less than 0.1.

[b] See Footnote b, Table 10.1.

[c] See Footnote c, Table 10.1.

10.2 Brine Inflow under Undisturbed Conditions: y = *E0:BRAALIC*

The variable y = *E0:BRAALIC* was included as an example because a nonlinear but monotonic relationship appears to exist between *E0:BRAALIC* and several of the sampled variables (Sect. 2). Thus, procedures that can identify monotonic relationships should work well with *E0:BRAALIC* as indicated by the regressions with raw data (R^2 = 0.50) and rank-transformed data (R^2 = 0.87) in Table 2.3. All analysis procedures except CVs identified *ANHPRM* and *WMICDFLG* as the two most important variables, with the variables assigned ranks 1 and 2 changing from test to test (Table 10.8). The scatterplots for both *ANHPRM* and *WMICDFLG* show strong relationships with *E0:BRAALIC* (Fig. 2.4). The CVs test assigned rank 2 to *SHPRMCON* with a p-value of 0.0426, with this variable also being assigned a p-value of 0.0057 and a rank of 3 from the CMNs test. No other tests indicated an effect for this variable, which is consistent with the corresponding scatterplot (Fig. 10.2). Rank 3 was assigned to *HALPRM* for tests based on RCCs (p-value = 0.0014), CLs (p-value = 0.0019), CMDs (p-value = 0.0050) and SI (p-value = 0.0517), with the corresponding scatterplot showing little discernible pattern (Fig. 10.2). Rank 4 was assigned to *WGRCOR* by CCs (p-value = 0.0048), RCCs (p-value = 0.0057), CMNs (p-value = 0.0636) and CLs (p-value = 0.0427), with the corresponding scatterplot indicating a slight tendency for *E0:BRAALIC* to decrease as *WGRCOR* increases (Fig. 2.4). There seems to be little discernible pattern to the ranks assigned to the remaining variables in Table 10.8.

Table 10.8 Comparison of Variable Rankings with Different Analysis Procedures[a] for $y = E0{:}BRAALIC$, the Variables in Table 2.2,[b] and a Maximum of Five Classes of x Values (i.e., $nX = 5$)[c]

Variable Name[d]	CC Rank	CC *p*-Val	RCC Rank	RCC *p*-Val	CMN: 1 × 5 Rank	CMN: 1 × 5 *p*-Val	CL: 1 × 5 Rank	CL: 1 × 5 *p*-Val
ANHPRM	1.0	0.0000	2.0	0.0000	1.0	0.0000	2.0	0.0000
WMICDFLG	2.0	0.0000	1.0	0.0000	2.0	0.0000	1.0	0.0000
WASTWICK	3.0	0.0045	6.0	0.0405	6.0	0.1062	16.0	0.4411
WGRCOR	4.0	0.0048	4.0	0.0057	4.0	0.0636	4.0	0.0427
ANHBCEXP	5.0	0.0095	15.0	0.6490	13.0	0.4467	19.0	0.7146
WFBETCEL	6.0	0.0555	8.0	0.2131	5.0	0.0732	10.0	0.2299
WRBRNSAT	7.0	0.0615	11.0	0.4046	11.0	0.3483	12.0	0.2889
HALPOR	8.0	0.0934	5.0	0.0087	19.0	0.5960	7.0	0.1431
HALPRM	11.0	0.2593	3.0	0.0014	7.0	0.1105	3.0	0.0019
SHPRMDRZ	12.0	0.2910	22.0	0.8392	21.0	0.6935	5.0	0.1060
SHPRMCON	14.0	0.3369	12.0	0.4170	3.0	0.0057	11.0	0.2394
SHRGSSAT	18.0	0.4767	14.0	0.5371	14.0	0.5044	9.0	0.2139
WGRMICI	21.0	0.5809	17.0	0.6663	20.0	0.6466	21.0	0.8966
SHRBRSAT	23.0	0.7329	10.0	0.2767	22.0	0.6946	6.0	0.1174

Variable Name	CMD: 2 × 5 Rank	CMD: 2 × 5 *p*-Val	CV: 1 × 5 Rank	CV: 1 × 5 *p*-Val	CIQ: 2 × 5 Rank	CIQ: 2 × 5 *p*-Val	SI: 5 × 5 Rank	SI: 5 × 5 *p*-Val
ANHPRM	2.0	0.0000	1.0	0.0078	1.0	0.0000	2.0	0.0000
WMICDFLG	1.0	0.0000	13.0	0.4046	2.0	0.0000	1.0	0.0000
WASTWICK	15.5	0.5467	8.0	0.2961	10.0	0.3883	15.0	0.5246
WGRCOR	5.0	0.0231	5.0	0.2125	13.5	0.6868	11.0	0.3644
ANHBCEXP	21.0	0.8088	7.0	0.2321	22.5	0.9554	21.0	0.7776
WFBETCEL	15.5	0.5467	6.0	0.2194	8.0	0.3084	13.0	0.4186
WRBRNSAT	13.0	0.3883	12.0	0.3851	7.0	0.2942	14.0	0.4186
HALPOR	4.0	0.0155	20.0	0.5416	6.0	0.2805	4.0	0.0698
HALPRM	3.0	0.0050	11.0	0.3596	24.0	0.9702	3.0	0.0517
SHPRMDRZ	6.0	0.0306	24.0	0.7101	22.5	0.9554	7.0	0.2202
SHPRMCON	10.0	0.2674	2.0	0.0426	17.0	0.8325	8.0	0.2436
SHRGSSAT	8.0	0.0504	17.0	0.5177	3.0	0.0628	6.0	0.2056
WGRMICI	18.0	0.6868	22.0	0.6096	4.0	0.0780	9.0	0.2863
SHRBRSAT	7.0	0.0362	18.0	0.5347	5.0	0.1395	5.0	0.1917

a, b, c, d See Footnotes a, b, c, d to Table 10.1

Based on knowledge of the model in use, the ordering of variable importance associated with RCCs seems most reasonable, with the signs of the RCCs for the variables ranked 1 through 6 (Table 10.9) corresponding to the effects that these variables should have on *E0:BRAALIC* (i.e., whether *E0:BRAALIC* should increase or decrease as the corresponding variable increases; see Helton et al., 1998, for a discussion of the underlying physics). The procedures that most closely match the variable rankings obtained with RCCs are based on CLs (TDC = 0.897), CMDs (TDC = 0.913) and SI (TDC = 0.838) (Table 10.10) and are procedures that can be expected to perform

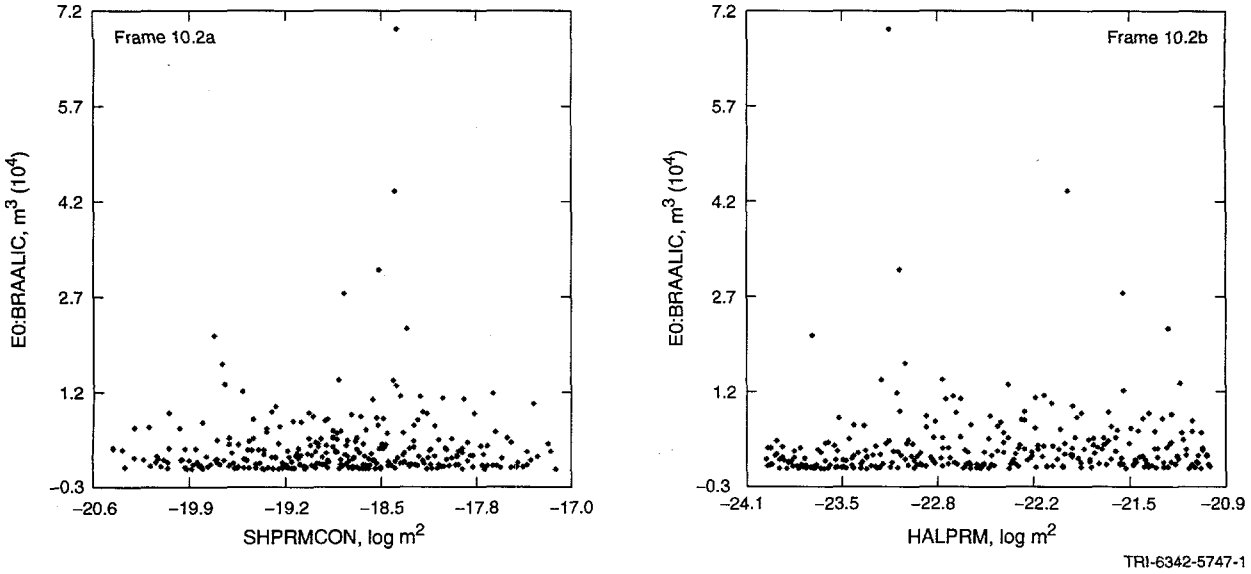

Fig. 10.2 Scatterplots for *E0:BRAALIC* versus *SHPRMCON* and *HALPRM*.

reasonably well in the presence of nonlinear but monotonic relations. The top-down correlation for RCCs and CCs is 0.729 (Table 10.10). Procedures based on measures of dispersion have the poorest agreement with variable rankings based RCCs (i.e., CVs with TDC = 0.301, CIQs with TDC = 0.531) (Table 10.10). Rankings of variable importance based on CCs, SRCs and PCCs are similar, with the rankings based on SRCs and PCCs being identical (Table 10.9). In like manner, rankings based on RCCs, SRRCs and PRCCs are similar, with the rankings based on SRRCs and PRCCs being identical (Table 10.9). The associated top down correlations are correspondingly high (i.e., 0.980 for (CC, SRC), (CC, PCC) and 0.912 for (RCC, SRRC), (RCC, PRCC)) (Table 10.10).

Table 10.9 Comparison of Variable Rankings with Correlation Coefficients, Standardized Regression Coefficients and Partial Correlation Coefficients with Raw and Rank Transformed Data for *y = E0:BRAALIC*

Variable[a]		CC		SRC		PCC	
Name	*p*-Val	Rank	Value	Rank	Value	Rank	Value
ANHPRM	0.0000	1.0	0.5655	1.0	0.5568	1.0	0.6317
WMICDFLG	0.0000	2.0	–.3210	2.0	–.2931	2.0	–.3878
WASTWICK	0.0045	3.0	–.1639	4.0	–.1451	4.0	–.2075
WGRCOR	0.0048	4.0	–.1628	3.0	–.1669	3.0	–.2370
ANHBCEXP	0.0095	5.0	–.1497	5.0	–.1155	5.0	–.1663
WFBETCEL	0.0555	6.0	–.1105	8.0	–.0757	8.0	–.1098
WRBRNSAT	0.0615	7.0	–.1080	9.0	–.0733	9.0	–.1065
HALPOR	0.0934	8.0	–.0969	6.0	–.0993	6.0	–.1435

Variable[b]		RCC		SRRC		PRCC	
Name	*p*-Val	Rank	Value	Rank	Value	Rank	Value
WMICDFLG	0.0000	1.0	–.6521	1.0	–.6533	1.0	–.8787
ANHPRM	0.0000	2.0	0.5804	2.0	0.5937	2.0	0.8619
HALPRM	0.0014	3.0	0.1850	5.0	0.1443	5.0	0.3817
WGRCOR	0.0057	4.0	–.1598	4.0	–.1509	4.0	–.3963
HALPOR	0.0087	5.0	–.1518	3.0	–.1539	3.0	–.4031
WASTWICK	0.0405	6.0	–.1185	7.0	–.0948	7.0	–.2617

a, b See Footnotes a, b to Table 10.2.

As for *E0:WAS_PRES*, an investigation was carried out to determine if the analysis results obtained for *E0:BRAALIC* are sensitive to the partitioning selected for use (i.e., the values for nX and nY). In particular, the analysis was repeated with $nX = 10$ for CMs, CLs, CMDs, CVs, CIQs and SI, and $nY = 10$ for SI (Table 10.11). As indicated by examination of scatterplots, the two most important variables with respect to *E0:BRAALIC* are *ANHPRM* and *WMICDFLG* (Fig. 2.4). With the exception of CVs, all tests (i.e., CMNs, CLs, CMDs, CIQs, SI) identified *ANHPRM* and *WMICDFLG* as the two most important variables with grids based on both $nX = 5$ and $nX = 10$ (Table 10.11). After these two variables, there is some jumping around in the rankings assigned to the individual variables, although there is sufficient similarity in the results obtained with $nX = 5$ and $nX = 10$ to produce top down correlations that are close to or above 0.9 (Table 10.11). Scatterplots indicate that, after *ANHPRM* and *WMICDFLG*, none of the remaining variables have a very strong effect on *E0:BRAALIC* (Figs. 2.4, 10.2), with the result that the tests are failing to find discernible patterns after these two variables.

Table 10.10 Top Down Correlation Matrix for Variable Rankings with Different Analysis Procedures for y = *E0:BRAALIC*, Variables included in Table 10.8,[a] and a Maximum of Five Classes of x values (i.e., $nX = 5$)

Top-Down Correlation Matrix

RCC	0.729										
CMN	0.841	0.721									
CL	0.589	0.897	0.626								
CMD	0.573	0.913	0.606	0.971							
CV	0.623	0.301	0.820	0.199	0.157						
CIQ	0.581	0.531	0.584	0.526	0.556	0.285					
SI	0.455	0.838	0.531	0.908	0.952	0.072	0.651				
SRC	0.980	0.728	0.839	0.618	0.612	0.630	0.588	0.476			
PCC	0.980	0.728	0.839	0.618	0.612	0.630	0.588	0.476	1.000		
SRRC	0.711	0.912	0.679	0.808	0.877	0.242	0.618	0.817	0.751	0.751	
PRCC	0.711	0.912	0.679	0.808	0.877	0.242	0.618	0.817	0.751	0.751	1.000
	CC	RCC	CMN	CL	CMD	CV	CIQ	SI	SRC	PCC	SRRC

Top-Down Correlation Matrix p Values

RCC	0.004										
CMN	0.001	0.005									
CL	0.017	0.001	0.012								
CMD	0.019	0.000	0.014	0.000							
CV	0.012	0.139	0.002	0.237	0.286						
CIQ	0.018	0.028	0.018	0.029	0.023	0.152					
SI	0.050	0.001	0.028	0.001	0.000	0.398	0.009				
SRC	0.000	0.004	0.001	0.013	0.014	0.012	0.017	0.043			
PCC	0.000	0.004	0.001	0.013	0.014	0.012	0.017	0.043	0.000		
SRRC	0.005	0.001	0.007	0.002	0.001	0.192	0.013	0.002	0.003	0.003	
PRCC	0.005	0.001	0.007	0.002	0.001	0.192	0.013	0.002	0.003	0.003	0.000
	CC	RCC	CMN	CL	CMD	CV	CIQ	SI	SRC	PCC	SRRC

[a] Same as Footnote a to Table 10.4 except for use of 14 variables from Table 10.8

The p-values used to identify important variables in Tables 10.8, 10.9 and 10.11 are calculated with statistical assumptions that are not fully satisfied. As described in conjunction with Eq. (10.1), a Monte Carlo procedure can be used to assess if the use of formal statistical procedures to determine p-values is producing misleading results. The p-values based on formal statistical procedures and on Monte Carlo procedures are very similar, with the associated variable rankings having top-down correlations between 0.987 and 0.995 (i.e., CC and RCC in Table 10.12 and CMN, CL, CMD, CV, CIQ and SI in Table 10.11). The primary difference is that the most important variables (i.e., *ANHPRM* and *WMICDFLG*) tend to be assigned tied-ranks (i.e., 1.5) in the Monte Carlo simulations

Table 10.11 Comparison of Variable Rankings for $y = E0{:}BRAALIC$ Obtained with a Maximum of Five Classes of x Values (i.e., $nX = 5$) and Analytic Determination of p-values with Variable Rankings Obtained with (i) a Maximum of Ten Classes of x values (i.e., $nX = 10$) and Analytic Determination of p-values and (ii) a Maximum of Five Classes of x values (i.e., $nX = 5$) and Monte Carlo Determination of p-values

Variable[a] Name	CMN: 1 × 5[b] Rank	p-Val	CMN: 1 × 10[c] Rank	p-Val	CMNMC: 1 × 5[d] Rank	p-Val
ANHPRM	1.0	0.0000	1.0	0.0000	1.5	0.0000
WMICDFLG	2.0	0.0000	2.0	0.0000	1.5	0.0000
SHPRMCON	3.0	0.0057	4.0	0.0655	3.0	0.0036
WGRCOR	4.0	0.0636	5.0	0.0723	4.0	0.0506
WFBETCEL	5.0	0.0732	10.0	0.2163	5.0	0.0572
WASTWICK	6.0	0.1062	6.0	0.1085	6.0	0.0856
HALPRM	7.0	0.1105	12.0	0.4030	7.0	0.0961
SHBCEXP	8.0	0.1140	3.0	0.0120	8.0	0.0995
SHPRMASP	9.0	0.2745	7.0	0.1970	9.0	0.2729
ANHBCVGP	10.0	0.2943	11.0	0.2943	10.0	0.3063
WRBRNSAT	11.0	0.3483	22.0	0.8366	11.0	0.3482
WGRMICH	12.0	0.4077	18.0	0.6346	12.0	0.4135
ANHBCEXP	13.0	0.4467	8.0	0.1984	13.0	0.4678
SHRGSSAT	14.0	0.5044	13.0	0.4492	14.0	0.5274
SHPRMCLY	15.0	0.5192	20.0	0.7703	15.0	0.5396
WRGSSAT	16.0	0.5378	14.0	0.4570	16.0	0.5649
ANRBRSAT	17.0	0.5595	15.0	0.4797	17.0	0.5871
SALPRES	18.0	0.5804	19.0	0.7158	18.0	0.6108
HALPOR	19.0	0.5960	24.0	0.9361	19.0	0.6414
WGRMICI	20.0	0.6466	17.0	0.6080	20.0	0.6749
SHPRMDRZ	21.0	0.6935	9.0	0.2108	22.0	0.7337
SHRBRSAT	22.0	0.6946	16.0	0.4925	21.0	0.7303
ANRGSSAT	23.0	0.7033	23.0	0.9300	24.0	0.7447
SHPRMHAL	24.0	0.7056	21.0	0.7932	23.0	0.7421
TDC[e]	1.000		0.891		0.987	

Variable Name	CL: 1 × 5 Rank	p-Val	CL: 1 × 10 Rank	p-Val	CLMC: 1 × 5 Rank	p-Val
WMICDFLG	1.0	0.0000	1.0	0.0000	1.5	0.0000
ANHPRM	2.0	0.0000	2.0	0.0000	1.5	0.0000
HALPRM	3.0	0.0019	3.0	0.0052	3.0	0.0013
WGRCOR	4.0	0.0427	8.0	0.2368	4.0	0.0438
SHPRMDRZ	5.0	0.1060	4.0	0.0206	5.0	0.1095
SHRBRSAT	6.0	0.1174	6.0	0.1781	6.0	0.1166
HALPOR	7.0	0.1431	13.0	0.5392	7.0	0.1427
SHBCEXP	8.0	0.1524	5.0	0.0441	8.0	0.1532
SHRGSSAT	9.0	0.2139	7.0	0.1810	9.0	0.2119
WFBETCEL	10.0	0.2299	9.0	0.3390	10.0	0.2328
SHPRMCON	11.0	0.2394	10.0	0.3674	11.0	0.2425
WRBRNSAT	12.0	0.2889	17.0	0.6731	12.0	0.2865
WGRMICH	13.0	0.3021	18.0	0.6854	13.0	0.3015
ANRBRSAT	14.0	0.3431	16.0	0.6694	14.0	0.3455
SHPRMASP	15.0	0.4066	14.0	0.5478	15.0	0.4045
WASTWICK	16.0	0.4411	12.0	0.5178	16.0	0.4363
SALPRES	17.0	0.4535	23.0	0.9222	17.0	0.4536
ANHBCVGP	18.0	0.4930	11.0	0.4930	18.0	0.4914
ANHBCEXP	19.0	0.7146	20.0	0.7230	19.0	0.7218
ANRGSSAT	20.0	0.8110	22.0	0.9036	20.0	0.8081
WGRMICI	21.0	0.8966	24.0	0.9948	21.0	0.9001
WRGSSAT	22.0	0.9035	15.0	0.5638	22.0	0.9085
SHPRMHAL	23.0	0.9367	21.0	0.8705	24.0	0.9392
SHPRMCLY	24.0	0.9385	19.0	0.7203	23.0	0.9387
TDC	1.000		0.941		0.987	

Variable Name	CMD: 2 × 5 Rank	p-Val	CMD: 2 × 10 Rank	p-Val	CMDMC: 2 × 5 Rank	p-Val
WMICDFLG	1.0	0.0000	1.0	0.0000	1.5	0.0000
ANHPRM	2.0	0.0000	2.0	0.0000	1.5	0.0000
HALPRM	3.0	0.0050	3.0	0.0089	3.0	0.0040
HALPOR	4.0	0.0155	8.0	0.1596	4.0	0.0169
WGRCOR	5.0	0.0231	7.0	0.1271	5.0	0.0221
SHPRMDRZ	6.0	0.0306	4.0	0.0215	6.0	0.0275
SHRBRSAT	7.0	0.0362	5.0	0.0282	7.0	0.0347
SHRGSSAT	8.0	0.0504	9.0	0.1849	8.0	0.0486
ANRBRSAT	9.0	0.1193	11.0	0.4559	9.0	0.1163
SHPRMCON	10.0	0.2674	14.0	0.5476	11.0	0.2796
SHBCEXP	11.0	0.2805	6.0	0.0856	10.0	0.2686
WGRMICH	12.0	0.3084	15.0	0.6163	12.0	0.3028
WRBRNSAT	13.0	0.3883	12.0	0.4814	13.0	0.3796
SALPRES	14.0	0.4060	22.0	0.8623	14.0	0.4157
WFBETCEL	15.5	0.5467	19.0	0.8043	16.0	0.5411
WASTWICK	15.5	0.5467	23.0	0.8930	15.0	0.5313
ANHBCVGP	17.0	0.6442	16.0	0.6442	17.0	0.5670
WGRMICI	18.0	0.6868	24.0	0.9429	18.0	0.6800
SHPRMASP	19.0	0.7358	17.0	0.7130	19.0	0.7235
SHPRMCLY	20.0	0.7603	13.0	0.5075	20.0	0.7669
ANHBCEXP	21.0	0.8088	20.5	0.8514	21.0	0.8156
ANRGSSAT	22.0	0.8781	18.0	0.7792	22.0	0.8679
SHPRMHAL	23.5	0.9702	20.5	0.8514	24.0	0.9672
WRGSSAT	23.5	0.9702	10.0	0.4311	23.0	0.9658
TDC	1.000		0.919		0.987	

Variable Name	CV: 1 × 5 Rank	p-Val	CV: 1 × 10 Rank	p-Val	CVMC: 1 × 5 Rank	p-Val
ANHPRM	1.0	0.0078	1.0	0.0010	1.0	0.0000
SHPRMCON	2.0	0.0426	7.0	0.2704	2.0	0.0058
SHBCEXP	3.0	0.1463	2.0	0.0329	3.0	0.0774
ANRBRSAT	4.0	0.1994	5.0	0.1188	4.0	0.1278
WGRCOR	5.0	0.2125	3.0	0.0995	5.0	0.1424
WFBETCEL	6.0	0.2194	13.0	0.4615	6.0	0.1455
ANHBCEXP	7.0	0.2321	4.0	0.1165	7.0	0.1697
WASTWICK	8.0	0.2961	6.0	0.2503	8.0	0.2450
ANHBCVGP	9.0	0.3250	8.0	0.3250	13.0	0.4487
WRGSSAT	10.0	0.3520	10.0	0.3701	9.0	0.3054
HALPRM	11.0	0.3596	15.0	0.5053	10.0	0.3129
WRBRNSAT	12.0	0.3851	21.0	0.5950	11.0	0.3446
WMICDFLG	13.0	0.4046	11.0	0.4046	14.0	0.4577
SALPRES	14.0	0.4395	9.0	0.3528	12.0	0.4360
WGRMICH	15.0	0.5077	17.0	0.5556	15.0	0.5798
SHPRMCLY	16.0	0.5144	22.0	0.6169	17.0	0.5952
SHRGSSAT	17.0	0.5177	20.0	0.5888	16.0	0.5915
SHRBRSAT	18.0	0.5347	14.0	0.4892	18.0	0.6333
SHPRMASP	19.0	0.5412	12.0	0.4393	19.0	0.6552
HALPOR	20.0	0.5416	24.0	0.6974	20.0	0.6583
SHPRMHAL	21.0	0.5895	18.0	0.5857	21.0	0.7538
WGRMICI	22.0	0.6096	16.0	0.5070	22.0	0.8035
ANRGSSAT	23.0	0.6631	23.0	0.6481	23.0	0.9119
SHPRMDRZ	24.0	0.7101	19.0	0.5875	24.0	0.9791
TDC	1.000		0.870		0.995	

Table 10.11 (Cont.)

Variable Name	CIQ: 2 × 5 Rank	CIQ: 2 × 5 *p*-Val	CIQ: 2 × 10 Rank	CIQ: 2 × 10 *p*-Val	CIQMC: 2 × 5 Rank	CIQMC: 2 × 5 *p*-Val
ANHPRM	1.0	0.0000	1.0	0.0000	1.5	0.0000
WMICDFLG	2.0	0.0000	2.0	0.0000	1.5	0.0000
SHRGSSAT	3.0	0.0628	4.0	0.0856	3.0	0.0628
WGRMICI	4.0	0.0780	6.0	0.1719	4.0	0.0757
SHRBRSAT	5.0	0.1395	13.0	0.5341	5.0	0.1382
HALPOR	6.0	0.2805	3.0	0.0235	6.0	0.2710
WRBRNSAT	7.0	0.2942	8.0	0.2803	7.0	0.2917
WFBETCEL	8.0	0.3084	15.5	0.6441	9.0	0.3089
ANHBCVGP	9.0	0.3556	9.0	0.3556	8.0	0.2967
WASTWICK	10.0	0.3883	12.0	0.5341	10.0	0.3859
SHPRMASP	11.0	0.4433	11.0	0.4559	11.0	0.4366
SHPRMCLY	12.0	0.5690	5.0	0.1373	12.0	0.5724
WGRCOR	13.5	0.6868	19.5	0.8930	13.0	0.6669
WRGSSAT	13.5	0.6868	15.5	0.6441	14.0	0.6693
SHBCEXP	15.0	0.7113	7.0	0.1849	15.0	0.7266
ANRBRSAT	16.0	0.7847	18.0	0.8043	16.0	0.7634
SHPRMCON	17.0	0.8325	17.0	0.6579	17.0	0.8557
WGRMICH	18.0	0.8781	19.5	0.8930	18.0	0.8769
ANRGSSAT	19.0	0.8995	21.0	0.9429	19.0	0.8949
SALPRES	20.5	0.9197	23.5	0.9892	21.0	0.9202
SHPRMHAL	20.5	0.9197	22.0	0.9669	20.0	0.9061
SHPRMDRZ	22.5	0.9554	14.0	0.6025	23.0	0.9476
ANHBCEXP	22.5	0.9554	23.5	0.9892	22.0	0.9472
HALPRM	24.0	0.9702	10.0	0.4071	24.0	0.9751
TDC	1.000		0.869		0.987	

Variable Name	SI: 5 × 5 Rank	SI: 5 × 5 *p*-Val	SI: 10 × 10 Rank	SI: 10 × 10 *p*-Val	SIMC: 5 × 5 Rank	SIMC: 5 × 5 *p*-Val
WMICDFLG	1.0	0.0000	1.0	0.0000	1.5	0.0000
ANHPRM	2.0	0.0000	2.0	0.0000	1.5	0.0000
HALPRM	3.0	0.0517	7.0	0.2313	3.0	0.0514
HALPOR	4.0	0.0698	6.0	0.2028	4.0	0.0706
SHRBRSAT	5.0	0.1917	8.0	0.2786	5.0	0.1898
SHRGSSAT	6.0	0.2056	14.0	0.5738	6.0	0.2058
SHPRMDRZ	7.0	0.2202	9.0	0.2955	7.0	0.2221
SHPRMCON	8.0	0.2436	3.0	0.0814	8.0	0.2566
WGRMICI	9.0	0.2863	13.0	0.4075	9.0	0.2825
WGRMICH	10.0	0.3239	20.0	0.8743	10.0	0.3257
WGRCOR	11.0	0.3644	23.0	0.8975	11.0	0.3641
SHBCEXP	12.0	0.4075	12.0	0.3878	12.0	0.4029
WFBETCEL	13.0	0.4186	19.0	0.8034	13.0	0.4175
WRBRNSAT	14.0	0.4186	21.5	0.8863	14.0	0.4221
WASTWICK	15.0	0.5246	15.0	0.6767	15.0	0.5217
ANHBCVGP	16.0	0.5467	4.0	0.1596	16.0	0.5633
SHPRMASP	17.0	0.6111	11.0	0.2955	17.0	0.6044
ANRBRSAT	18.0	0.7324	16.0	0.7155	18.0	0.7286
WRGSSAT	19.0	0.7440	10.0	0.2955	19.0	0.7470
SHPRMCLY	20.0	0.7666	5.0	0.1895	20.0	0.7780
ANHBCEXP	21.0	0.7776	21.5	0.8863	21.0	0.7856
SHPRMHAL	22.0	0.7776	17.0	0.7701	22.0	0.7929
SALPRES	23.0	0.9326	18.0	0.7870	23.0	0.9354
ANRGSSAT	24.0	0.9537	24.0	0.9846	24.0	0.9561
TDC	1.000		0.748		0.988	

a, b, c, d, e See Footnotes a, b, c, d, e to Table 10.5.

because the sample size of 10,000 in use does not allow the estimation of *p*-values less than 0.0001; in contrast, the formal statistical procedures allow the calculation of *p*-values less than 0.0001 and thus result in distinct ranks being assigned to *ANHPRM* and *WMICDFLG*.

A possible variant of the common means (CMNs) test is to use logarithmically transformed *y*-values rather than the original untransformed *y*-values (Sect. 5.1). Use of both raw and logarithmically transformed variables results in *ANHPRM* and *WMICDFLG* being selected as the two most important variables with respect to *E0:BRAALIC* (Tables 10.11, 10.13). Use of logarithmically transformed variables with the CMNs test also results in the identification of *HALPRM* as the third most important variable, with *HALPRM* also assigned a rank of 3 with RCCs, but effectively missed by the CMNs test with raw data (i.e., a *p*-value of 0.1105 and a rank of 7) (Tables 10.11, 10.13). The CMNs test with both raw and logarithmically transformed data assigns rank 4 to *WGRCOR* (Tables 10.11, 10.13). Thus, the use of logarithmically transformed data with the CMNs test results in the identification of one possibly important variable (i.e., *WGRCOR*) missed with the use of raw data.

A possible variant of the common variances (CVs) test is to use t_{ql} as defined in Eq. (6.4) rather than as defined in Eq. (6.3). The logarithmic transformation associated with Eq. (6.4) results in a substantial improvement in that

Table 10.12 Comparison of Variable Rankings for $y = E0{:}BRAALIC$ Obtained with Correlation Coefficients (CCs, RCCs) and Analytic Determination of p-values with Rankings Obtained with Monte Carlo Determination of p-values

Variable[a] Name	CC[b] Rank	CC[b] p-Val	CCMC[c] Rank	CCMC[c] p-Val
ANHPRM	1.0	0.0000	1.5	0.0000
WMICDFLG	2.0	0.0000	1.5	0.0000
WASTWICK	3.0	0.0045	3.0	0.0022
WGRCOR	4.0	0.0048	4.0	0.0029
ANHBCEXP	5.0	0.0095	5.0	0.0115
WFBETCEL	6.0	0.0555	6.0	0.0528
WRBRNSAT	7.0	0.0615	7.0	0.0585
HALPOR	8.0	0.0934	8.0	0.0947
WRGSSAT	9.0	0.1231	9.0	0.1284
WGRMICH	10.0	0.1749	10.0	0.1840
HALPRM	11.0	0.2593	11.0	0.2664
SHPRMDRZ	12.0	0.2910	12.0	0.2983
ANHBCVGP	13.0	0.2927	13.0	0.3063
SHPRMCON	14.0	0.3369	14.0	0.3375
SALPRES	15.0	0.3772	15.0	0.3940
SHBCEXP	16.0	0.4396	17.0	0.4520
ANRGSSAT	17.0	0.4561	16.0	0.4504
SHRGSSAT	18.0	0.4767	19.0	0.4948
ANRBRSAT	19.0	0.4875	18.0	0.4829
SHPRMASP	20.0	0.5725	20.0	0.5916
WGRMICI	21.0	0.5809	21.0	0.5987
SHPRMCLY	22.0	0.6051	22.0	0.6157
SHRBRSAT	23.0	0.7329	23.0	0.7371
SHPRMHAL	24.0	0.7958	24.0	0.8000
TDC[d]		0.987		

Variable Name	RCC Rank	RCC p-Val	RCCMC Rank	RCCMC p-Val
WMICDFLG	1.0	0.0000	1.5	0.0000
ANHPRM	2.0	0.0000	1.5	0.0000
HALPRM	3.0	0.0014	3.0	0.0009
WGRCOR	4.0	0.0057	4.0	0.0044
HALPOR	5.0	0.0087	5.0	0.0084
WASTWICK	6.0	0.0405	6.0	0.0401
SALPRES	7.0	0.1107	7.0	0.1105
WFBETCEL	8.0	0.2131	8.0	0.2107
SHPRMASP	9.0	0.2168	9.0	0.2194
SHRBRSAT	10.0	0.2767	10.0	0.2726
WRBRNSAT	11.0	0.4046	11.0	0.4002
SHPRMCON	12.0	0.4170	12.0	0.4224
ANHBCVGP	13.0	0.4930	13.0	0.4914
SHRGSSAT	14.0	0.5371	14.0	0.5434
ANHBCEXP	15.0	0.6490	15.0	0.6547
SHBCEXP	16.0	0.6617	16.0	0.6576
WGRMICI	17.0	0.6663	17.0	0.6604
ANRGSSAT	18.0	0.6666	18.0	0.6639
WRGSSAT	19.0	0.7237	19.0	0.7247
SHPRMCLY	20.0	0.8111	20.0	0.8101
ANRBRSAT	21.0	0.8156	21.0	0.8137
SHPRMDRZ	22.0	0.8392	22.0	0.8466
WGRMICH	23.0	0.8513	23.0	0.8479
SHPRMHAL	24.0	0.8619	24.0	0.8632
TDC		0.988		

a, b, c, d See Footnotes a, b, c, d in Table 10.6.

Table 10.13 Exceedance Probabilities (i.e., p-values) for Common Mean and Common Variance Tests Calculated with Use of Logarithms[a] for $y = E0{:}BRAALIC$, the variables in Table 2.2,[b] and a Maximum of Five Classes of Values for Each Variable (i.e., $nX = 5$)[c]

Variable Name	CMN: Log, 1×5 Rank	CMN: Log, 1×5 p-Val	CMNMC: Log, 1×5 Rank	CMNMC: Log, 1×5 p-Val
WMICDFLG	1.0	0.0000	1.5	0.0000
ANHPRM	2.0	0.0000	1.5	0.0000
HALPRM	3.0	0.0022	3.0	0.0022
WGRCOR	4.0	0.0284	4.0	0.0286
SHPRMDRZ	5.0	0.0967	5.0	0.1029

Variable Name	CV: Log, 1×5 Rank	CV: Log, 1×5 p-Val	CVMC: Log, 1×5 Rank	CVMC: Log, 1×5 p-Val
ANHPRM	1.0	0.0000	1.0	0.0000
WMICDFLG	2.0	0.0002	2.0	0.0064
SHPRMCON	3.0	0.0019	3.0	0.0403
SHBCEXP	4.0	0.0130	4.0	0.1104
WASTWICK	5.0	0.0144	5.0	0.1140
ANRBRSAT	6.0	0.0189	6.0	0.1341
ANHBCEXP	7.0	0.0290	7.0	0.1699
WRBRNSAT	8.0	0.0304	8.0	0.1711
WFBETCEL	9.0	0.0754	9.0	0.2968
WGRMICH	10.0	0.0930	10.0	0.3384

a, b, c See Footnotes a, b, c to Table 10.7.

WMICDFLG is now identified as an important variable (Table 10.13); in contrast, *WMICDFLG* was missed with raw data as used in Eq. (6.3) (Table 10.11). The associated scatterplot indicates that *WMICDFLG* is a variable that should be identified by any reasonable test (Fig. 2.4).

10.3 Repository Saturation under Disturbed Conditions: $y = E2{:}WAS_SATB$

The variable $y = E2{:}WAS_SATB$ was selected as an example because the regression analyses with raw and rank-transformed data were rather poor (i.e., $R^2 = 0.33$ and 0.61, respectively), although the two most important variables as indicated by scatterplots (i.e., *BHPRM* and *WRGSSAT*) do appear in both regression analyses (Table 2.3, Fig. 2.5). Given the strong patterns displayed in the scatterplots for *BHPRM* and *WRGSSAT* and the discernible but less strong patterns associated with *ANHPRM* and *HALPOR* (Fig. 2.5), procedures that can identify patterns that result from the interaction of two or more variables should work well for *E2:WAS_SATB*. In particular, analyses of variable importance based on RCCs, CLs, CMDs and SI identified *BHPRM* and *WRGSSAT* as the two most important variables (Table 10.14). Analyses based on CCs, CMNs and CVs identified *BHPRM* as the most important variable, but did not identify *WRGSSAT* as the second most important variable; in contrast, CIQs identified WRGSSAT as the most important variable and identified *BHPRM* as the third rather than second most important variable (Table 10.14). Further, CIQs identified *WGRCOR* as the second most important variable (Table 10.14), which seems to be inconsistent with the weakness of the pattern appearing in the associated scatterplot (Fig. 10.3) and also the rankings assigned to *WGRCOR* by other procedures (Table 10.14). The test based on CVs did not identify *WRGSSAT* (i.e., a p-value of 0.1750 and a rank of 9) (Table 10.14).

Given the insights gained from the results of all of the analysis techniques, CCs and RCCs appear to have identified the three dominant variables affecting *E2:WAS_SATB* (i.e., *BHPRM*, *WRGSSAT*, *ANHPRM*) (Table 10.14). However, given the low R^2 values associated with the corresponding regression models with raw and rank-transformed data (Table 2.3), it would be difficult to place much faith in these identifications without results from tests that are less dependent on linear regression models (i.e., CLs, CMDs, CIQs, SI).

As previously observed for *E0:WAS_PRES* and *E0:BRAALIC* (Tables 10.2, 10.9), variable rankings for *E2:WAS_SATB* with CCs, SRCs and PCCs are similar, with SRCs and PCCs producing identical variable rankings (Table 10.15). A similar pattern also occurs for RCCs, SRRCs and PRCCs (Table 10.15).

Top-down correlation provides a formal comparison of the variable rankings obtained with the different procedures (Table 10.16). A considerable amount of variability exists in the rankings obtained with the different techniques. Rankings based on SI, CVs and CIQs appear to have the least agreement with the rankings obtained with other procedures.

Table 10.14 Comparison of Variable Rankings with Different Analysis Procedures[a] for y = *E2:WAS_SATB*, the Variables in Table 2.2,[b] and a Maximum of Five Classes of x Values (i.e., $nX = 5$)[c]

Variable[d]	CC		RCC		CMN: 1 × 5		CL: 1 × 5	
Name	Rank	*p*-Val	Rank	*p*-Val	Rank	*p*-Val	Rank	*p*-Val
BHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
ANHPRM	2.0	0.0000	3.0	0.0001	2.0	0.0000	3.0	0.0001
HALPOR	3.0	0.0006	5.0	0.0269	4.0	0.0124	12.0	0.3437
WGRCOR	4.0	0.0017	6.0	0.1446	6.0	0.0296	11.0	0.3179
WRGSSAT	5.0	0.0081	2.0	0.0000	5.0	0.0143	2.0	0.0000
WMICDFLG	6.0	0.0214	7.0	0.1745	7.0	0.0317	10.0	0.2824
WGRMICH	7.0	0.0838	8.0	0.1842	3.0	0.0021	4.0	0.0059
SHPRMHAL	8.0	0.0996	4.0	0.0225	10.0	0.1586	8.0	0.1528
WRBRNSAT	11.0	0.2350	13.0	0.4950	8.0	0.0801	6.0	0.0270
ANRBRSAT	15.0	0.6402	20.0	0.6645	19.0	0.7070	13.0	0.3977
SHPRMCLY	21.0	0.9020	16.0	0.6137	11.0	0.1743	7.0	0.0972
SHPRMCON	23.0	0.9478	19.0	0.6549	9.0	0.1149	5.0	0.0202

Variable	CMD: 2 × 5		CV: 1 × 5		CIQ: 2 × 5		SI: 5 × 5	
Name	Rank	*p*-Val	Rank	*p*-Val	Rank	*p*-Val	Rank	*p*-Val
BHPRM	1.0	0.0000	1.0	0.0000	3.0	0.0054	2.0	0.0000
ANHPRM	3.0	0.0003	2.0	0.0000	4.0	0.0628	3.0	0.0002
HALPOR	23.5	0.8557	3.0	0.0011	6.0	0.1324	7.0	0.1328
WGRCOR	13.0	0.5037	5.0	0.0067	2.0	0.0019	6.0	0.1010
WRGSSAT	2.0	0.0000	9.0	0.1750	1.0	0.0000	1.0	0.0000
WMICDFLG	8.0	0.2187	6.0	0.0114	15.0	0.5134	8.0	0.1542
WGRMICH	4.0	0.0130	4.0	0.0050	7.0	0.2311	5.0	0.0564
SHPRMHAL	6.0	0.0218	7.0	0.1122	11.0	0.4628	10.0	0.2278
WRBRNSAT	12.0	0.3883	8.0	0.1749	25.0	0.9825	9.0	0.2128
ANRBRSAT	17.0	0.6868	25.0	0.9798	8.0	0.2674	4.0	0.0495
SHPRMCLY	7.0	0.1627	21.0	0.7874	16.0	0.5467	16.0	0.5739
SHPRMCON	5.0	0.0206	22.0	0.8224	10.0	0.3546	12.0	0.4075

a, c, d See Footnotes a, b, c to Table 10.1

b Same as Footnote b to Table 10.1 except that *BHPRM* is used in the calculation of E2 results (i.e., *E2:WAS_SATB* and *E2:WAS_PRES*) and so was included as an independent variable, which resulted in 25 variables (i.e., x's) for analysis with each procedure.

As for *E0:WAS_PRES* and *E0:BRAALIC*, an investigation was carried out for *E2:WAS_SATB* on the effects of using $nX = 10$ rather than $nX = 5$ for CMNs, CLs, CMDs, CVs, CIQs and SI and $nY = 10$ rather than $nY = 5$ for SI (Table 10.17). The results for the highest ranked variables for the two partitionings were similar, with CMNs, CLs, CMDs, CIQs and SI each identifying the same top 3 variables, although the identified variables were not necessarily the same from test to test. For CVs, both partitionings yielded the same top two variables but produced different

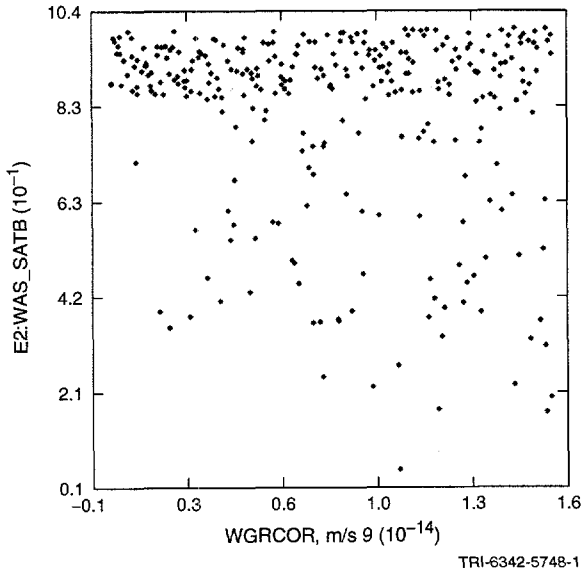

Fig. 10.3. Scatterplot for *E2:WAS_SATB* versus *WGRCOR*.

Table 10.15 Comparison of Variable Rankings with Correlation Coefficients, Standardized Regression Coefficients and Partial Correlation Coefficients with Raw and Rank Transformed Data for *y = E2:WAS_SATB*

Variable[a]		CC		SRC		PCC	
Name	*p*-Val	Rank	Value	Rank	Value	Rank	Value
BHPRM	0.0000	1.0	0.6029	1.0	0.5923	1.0	0.6812
ANHPRM	0.0000	2.0	0.2929	2.0	0.2950	2.0	0.4198
HALPOR	0.0006	3.0	0.1980	4.0	0.1888	4.0	0.2832
WGRCOR	0.0017	4.0	−.1810	3.0	−.1891	3.0	−.2840
WRGSSAT	0.0081	5.0	−.1530	5.0	−.1447	5.0	−.2211
WMICDFLG	0.0214	6.0	−.1329	6.0	−.1357	6.0	−.2044
WGRMICH	0.0838	7.0	0.0998	7.0	0.0928	7.0	0.1437
SHPRMHAL	0.0996	8.0	−.0951	9.0	−.0831	9.0	−.1290

Variable[b]		RCC		SRRC		PRCC	
Name	*p*-Val	Rank	Value	Rank	Value	Rank	Value
BHPRM	0.0000	1.0	0.5989	1.0	0.5904	1.0	0.7011
WRGSSAT	0.0000	2.0	−.4113	2.0	−.3997	2.0	−.5534
ANHPRM	0.0001	3.0	0.2299	3.0	0.2260	3.0	0.3518
SHPRMHAL	0.0225	4.0	−.1319	5.0	−.1148	5.0	−.1870
HALPOR	0.0269	5.0	0.1280	4.0	0.1315	4.0	0.2136

a, b See Footnotes a, b in Table 10.2.

Table 10.16 Top Down Correlation Matrix for Variable Rankings with Different Analysis Procedures for y = *E2:WAS_SATB*, Variables included in Table 10.14,[a] and a Maximum of Five Classes of x values (i.e., $nX = 5$)

Top-Down Correlation Matrix

RCC	0.851										
CMN	0.919	0.790									
CL	0.643	0.815	0.781								
CMD	0.640	0.840	0.763	0.982							
CV	0.947	0.742	0.950	0.609	0.602						
CIQ	0.490	0.631	0.422	0.494	0.503	0.267					
SI	0.551	0.727	0.561	0.702	0.706	0.363	0.840				
SRC	0.988	0.844	0.902	0.646	0.647	0.926	0.530	0.557			
PCC	0.988	0.844	0.902	0.646	0.647	0.926	0.530	0.557	1.000		
SRRC	0.876	0.989	0.812	0.806	0.830	0.762	0.655	0.732	0.859	0.859	
PRCC	0.876	0.989	0.812	0.806	0.830	0.762	0.655	0.732	0.859	0.859	1.000
	CC	RCC	CMN	CL	CMD	CV	CIQ	SI	SRC	PCC	SRRC

Top-Down Correlation Matrix p Values

RCC	0.002										
CMN	0.001	0.004									
CL	0.017	0.003	0.005								
CMD	0.017	0.003	0.006	0.001							
CV	0.001	0.007	0.001	0.022	0.023						
CIQ	0.052	0.018	0.081	0.051	0.048	0.188					
SI	0.034	0.008	0.031	0.010	0.010	0.115	0.003				
SRC	0.001	0.003	0.001	0.016	0.016	0.001	0.039	0.032			
PCC	0.001	0.003	0.001	0.016	0.016	0.001	0.039	0.032	0.000		
SRRC	0.002	0.001	0.004	0.004	0.003	0.006	0.015	0.008	0.002	0.002	
PRCC	0.002	0.001	0.004	0.004	0.003	0.006	0.015	0.008	0.002	0.002	0.000
	CC	RCC	CMN	CL	CMD	CV	CIQ	SI	SRC	PCC	SRRC

[a] Same as Footnote a to Table 10.4 except for use of 12 variables from Table 10.14.

variables with rank 3. After the top three variables, there was often considerable variability in the ranks assigned to the remaining, and typically unimportant, variables. The least agreement between the variable rankings obtained with the two partitionings occurred for SI (i.e., a p-value of 0.746).

As previously described, the p-values used to identify important variables in Tables 10.14, 10.15 and 10.17 were recalculated with a Monte Carlo procedure (Tables 10.17, 10.18). The p-values based on formal statistical procedures and on Monte Carlo procedures are very similar, with the associated variable rankings having top-down correlations between 0.972 and 0.999 (i.e., CC and RCC in Table 10.18 and CMN, CL, CMD, CV, CIQ and SI in Table 10.17). As in previous comparisons, the primary difference is that the most important variables tend to be assigned tied-ranks in the Monte Carlo simulation (e.g., 1.5 for *BHPRM* and *ANHPRM* for CCs in Table 10.18) because p-values less than 0.0001 cannot be estimated with the sample size of 10,000 in use.

Table 10.17 Comparison of Variable Rankings for y = *E2:WAS_SATB* Obtained with a Maximum of Five Classes of x Values (i.e., nX = 5) and Analytic Determination of p-values with Variable Rankings Obtained with (i) a Maximum of Ten Classes of x values (i.e., nX = 10) and Analytic Determination of p-values and (ii) a Maximum of Five Classes of x values (i.e., nX = 5) and Monte Carlo Determination of p-values

Variable[a] Name	CMN: 1 × 5[b] Rank	CMN: 1 × 5[b] *p*-Val	CMN: 1 × 10[c] Rank	CMN: 1 × 10[c] *p*-Val	CMNMC: 1 × 5[d] Rank	CMNMC: 1 × 5[d] *p*-Val
BHPRM	1.0	0.0000	1.0	0.0000	1.5	0.0000
ANHPRM	2.0	0.0000	2.0	0.0000	1.5	0.0000
WGRMICH	3.0	0.0021	3.0	0.0053	3.0	0.0022
HALPOR	4.0	0.0124	6.0	0.0546	4.0	0.0116
WRGSSAT	5.0	0.0143	8.0	0.1113	5.0	0.0146
WGRCOR	6.0	0.0296	5.0	0.0343	6.0	0.0294
WMICDFLG	7.0	0.0317	4.0	0.0317	7.0	0.0320
WRBRNSAT	8.0	0.0801	9.0	0.1416	8.0	0.0791
SHPRMCON	9.0	0.1149	10.0	0.1673	9.0	0.1115
SHPRMHAL	10.0	0.1586	7.0	0.0927	10.0	0.1593
SHPRMCLY	11.0	0.1743	13.0	0.2772	11.0	0.1767
HALPRM	12.0	0.2934	12.0	0.2659	12.0	0.2880
WFBETCEL	13.0	0.3777	11.0	0.1777	13.0	0.3834
SALPRES	14.0	0.4811	18.0	0.5665	14.0	0.4812
WGRMICI	15.0	0.5114	22.0	0.7911	15.0	0.5073
SHRBRSAT	16.0	0.5436	15.0	0.4829	16.0	0.5453
SHPRMDRZ	17.0	0.6611	17.0	0.5339	17.0	0.6665
ANHBCEXP	18.0	0.6867	14.0	0.3160	18.0	0.6886
ANRBRSAT	19.0	0.7070	24.0	0.9389	19.0	0.7035
SHRGSSAT	20.0	0.7525	19.0	0.6801	20.0	0.7626
WASTWICK	21.0	0.8552	16.0	0.4959	21.0	0.8486
SHPRMASP	22.0	0.8739	25.0	0.9969	22.0	0.8728
SHBCEXP	23.0	0.8842	21.0	0.7797	23.0	0.8797
ANHBCVGP	24.0	0.8920	23.0	0.8920	25.0	0.8930
ANRGSSAT	25.0	0.8929	20.0	0.7163	24.0	0.8913
TDC[e]	1.000		0.962		0.988	

Variable Name	CL: 1 × 5 Rank	CL: 1 × 5 *p*-Val	CL: 1 × 10 Rank	CL: 1 × 10 *p*-Val	CLMC: 1 × 5 Rank	CLMC: 1 × 5 *p*-Val
BHPRM	1.0	0.0000	1.0	0.0000	1.5	0.0000
WRGSSAT	2.0	0.0000	2.0	0.0000	1.5	0.0000
ANHPRM	3.0	0.0001	3.0	0.0008	3.0	0.0001
WGRMICH	4.0	0.0059	5.0	0.0289	4.0	0.0056
SHPRMCON	5.0	0.0202	6.0	0.0963	5.0	0.0165
WRBRNSAT	6.0	0.0270	4.0	0.0132	6.0	0.0240
SHPRMCLY	7.0	0.0972	11.0	0.3016	7.0	0.0932
SHPRMHAL	8.0	0.1528	8.0	0.1902	8.0	0.1521
WASTWICK	9.0	0.2625	14.0	0.3439	9.0	0.2632
WMICDFLG	10.0	0.2824	9.0	0.2824	10.0	0.2823
WGRCOR	11.0	0.3179	12.0	0.3087	11.0	0.3154
HALPOR	12.0	0.3437	10.0	0.3014	12.0	0.3524
ANRBRSAT	13.0	0.3977	20.0	0.6774	13.0	0.3941
SHPRMDRZ	14.0	0.4990	13.0	0.3203	14.0	0.4985
SHRBRSAT	15.0	0.5183	16.0	0.5305	15.0	0.5152
WFBETCEL	16.0	0.5450	22.0	0.7402	17.0	0.5542
WGRMICI	17.0	0.5515	18.0	0.5774	16.0	0.5445
HALPRM	18.0	0.5728	15.0	0.5070	18.0	0.5734
ANRGSSAT	19.0	0.7586	17.0	0.5626	19.0	0.7587
SALPRES	20.0	0.7684	23.0	0.7880	20.0	0.7732
ANHBCEXP	21.0	0.8201	19.0	0.6201	21.0	0.8244
SHBCEXP	22.0	0.9120	21.0	0.7133	23.0	0.9128
SHPRMASP	23.0	0.9129	25.0	0.9653	24.0	0.9136
ANHBCVGP	24.0	0.9133	24.0	0.9133	22.0	0.9125
SHRGSSAT	25.0	0.9424	7.0	0.1763	25.0	0.9418
TDC	1.000		0.930		0.988	

Variable Name	CMD: 2 × 5 Rank	CMD: 2 × 5 *p*-Val	CMD: 2 × 10 Rank	CMD: 2 × 10 *p*-Val	CMDMC: 2 × 5 Rank	CMDMC: 2 × 5 *p*-Val
BHPRM	1.0	0.0000	2.0	0.0000	1.5	0.0000
WRGSSAT	2.0	0.0000	1.0	0.0000	1.5	0.0000
ANHPRM	3.0	0.0003	3.0	0.0035	3.0	0.0001
WGRMICH	4.0	0.0130	6.0	0.0856	4.0	0.0135
SHPRMCON	5.0	0.0206	7.0	0.1538	5.0	0.0207
SHPRMHAL	6.0	0.0218	5.0	0.0669	6.0	0.0227
SHPRMCLY	7.0	0.1627	15.0	0.5075	7.0	0.1639
WMICDFLG	8.0	0.2187	9.0	0.2187	8.0	0.2133
SHPRMASP	9.0	0.2674	11.0	0.3191	9.0	0.2585
ANHBCVGP	10.0	0.3556	12.0	0.3556	10.0	0.2958
HALPRM	11.0	0.3796	13.0	0.3614	12.0	0.3982
WRBRNSAT	12.0	0.3883	10.0	0.2803	11.0	0.3824
WGRCOR	13.0	0.5037	8.0	0.2133	13.0	0.4932
WASTWICK	14.0	0.5467	20.0	0.8043	14.0	0.5342
SALPRES	15.0	0.5804	19.0	0.7792	15.0	0.5807
ANRGSSAT	16.0	0.6626	17.0	0.7532	16.0	0.6623
ANRBRSAT	17.0	0.6868	22.5	0.8514	17.0	0.6653
SHPRMDRZ	18.0	0.7113	14.0	0.4686	18.0	0.7241
SHRBRSAT	19.0	0.7358	18.0	0.7792	19.0	0.7470
SHRGSSAT	20.0	0.7603	4.0	0.0519	21.0	0.7642
WFBETCEL	21.5	0.7847	22.5	0.8514	22.0	0.8025
WGRMICI	21.5	0.7847	24.0	0.9281	20.0	0.7564
HALPOR	23.5	0.8557	21.0	0.8285	24.0	0.8566
ANHBCEXP	23.5	0.8557	25.0	0.9558	23.0	0.8523
SHBCEXP	25.0	0.9825	16.0	0.6163	25.0	0.9827
TDC	1.000		0.835		0.988	

Variable Name	CV: 1 × 5 Rank	CV: 1 × 5 *p*-Val	CV: 1 × 10 Rank	CV: 1 × 10 *p*-Val	CVMC: 1 × 5 Rank	CVMC: 1 × 5 *p*-Val
BHPRM	1.0	0.0000	1.0	0.0000	1.5	0.0000
ANHPRM	2.0	0.0000	2.0	0.0000	1.5	0.0000
HALPOR	3.0	0.0011	5.0	0.0059	3.0	0.0022
WGRMICH	4.0	0.0050	3.0	0.0043	4.0	0.0082
WGRCOR	5.0	0.0067	4.0	0.0058	5.0	0.0100
WMICDFLG	6.0	0.0114	6.0	0.0114	6.0	0.0140
SHPRMHAL	7.0	0.1122	7.0	0.0765	7.0	0.1208
WRBRNSAT	8.0	0.1749	17.0	0.6414	8.0	0.1806
WRGSSAT	9.0	0.1750	19.0	0.7157	9.0	0.1828
SALPRES	10.0	0.2505	10.0	0.2846	10.0	0.2552
HALPRM	11.0	0.3132	16.0	0.5702	11.0	0.3153
WGRMICI	12.0	0.3367	22.0	0.8053	12.0	0.3400
WFBETCEL	13.0	0.4874	9.0	0.1947	13.0	0.5039
ANHBCEXP	14.0	0.5644	8.0	0.1754	14.0	0.5748
ANHBCVGP	15.0	0.6458	18.0	0.6458	16.0	0.6444
SHBCEXP	16.0	0.6500	11.0	0.3697	15.0	0.6427
SHRGSSAT	17.0	0.7043	23.0	0.8563	17.0	0.7147
SHRBRSAT	18.0	0.7236	15.0	0.5688	18.0	0.7237
WASTWICK	19.0	0.7404	13.0	0.5142	19.0	0.7340
SHPRMASP	20.0	0.7468	24.0	0.9892	20.0	0.7514
SHPRMCLY	21.0	0.7874	12.0	0.4471	21.0	0.7893
SHPRMCON	22.0	0.8224	14.0	0.5288	22.0	0.8257
ANRGSSAT	23.0	0.8434	21.0	0.7810	23.0	0.8462
SHPRMDRZ	24.0	0.8702	20.0	0.7694	24.0	0.8727
ANRBRSAT	25.0	0.9798	25.0	0.9997	25.0	0.9794
TDC	1.000		0.909		0.988	

Table 10.17 (Cont.)

Variable Name	CIQ: 2 × 5 Rank	CIQ: 2 × 5 *p*-Val	CIQ: 2 × 10 Rank	CIQ: 2 × 10 *p*-Val	CIQMC: 2 × 5 Rank	CIQMC: 2 × 5 *p*-Val
WRGSSAT	1.0	0.0000	1.0	0.0000	1.0	0.0000
WGRCOR	2.0	0.0019	2.0	0.0021	2.0	0.0012
BHPRM	3.0	0.0054	3.0	0.0063	3.0	0.0055
ANHPRM	4.0	0.0628	6.0	0.2622	4.0	0.0670
SHRBRSAT	5.0	0.1257	7.5	0.2803	5.0	0.1209
HALPOR	6.0	0.1324	4.0	0.1481	6.0	0.1317
WGRMICH	7.0	0.2311	14.0	0.4814	7.0	0.2359
ANRBRSAT	8.0	0.2674	15.0	0.5075	8.0	0.2613
SHRGSSAT	9.0	0.3232	7.5	0.2803	9.0	0.3186
SHPRMCON	10.0	0.3546	10.0	0.3725	10.0	0.3661
SHPRMHAL	11.0	0.4628	20.5	0.8043	11.0	0.4584
SALPRES	12.0	0.4830	13.0	0.4435	12.0	0.4870
SHPRMASP	13.0	0.4830	17.0	0.5749	13.0	0.4904
HALPRM	14.0	0.5037	12.0	0.4190	14.0	0.5101
WMICDFLG	15.0	0.5134	16.0	0.5134	15.0	0.5185
SHPRMCLY	16.0	0.5467	18.0	0.6025	16.0	0.5455
WASTWICK	17.0	0.5467	23.0	0.8930	17.0	0.5578
WFBETCEL	18.0	0.5690	5.0	0.1719	19.0	0.5747
ANRGSSAT	19.0	0.6387	20.5	0.8043	20.0	0.6237
SHBCEXP	20.0	0.6387	9.0	0.2993	21.0	0.6496
ANHBCVGP	21.0	0.6442	19.0	0.6442	18.0	0.5633
ANHBCEXP	22.0	0.7847	22.0	0.8729	22.0	0.8008
WGRMICI	23.0	0.8995	24.0	0.9114	23.0	0.8932
SHPRMDRZ	24.0	0.9197	11.0	0.3838	24.0	0.9152
WRBRNSAT	25.0	0.9825	25.0	0.9761	25.0	0.9777
TDC	1.000		0.872		0.999	

Variable Name	SI: 5 × 5 Rank	SI: 5 × 5 *p*-Val	SI: 10 × 10 Rank	SI: 10 × 10 *p*-Val	SIMC: 5 × 5 Rank	SIMC: 5 × 5 *p*-Val
WRGSSAT	1.0	0.0000	1.0	0.0000	2.0	0.0000
BHPRM	2.0	0.0000	2.0	0.0000	2.0	0.0000
ANHPRM	3.0	0.0002	3.0	0.0058	2.0	0.0000
ANRBRSAT	4.0	0.0495	19.0	0.8034	4.0	0.0451
WGRMICH	5.0	0.0564	12.0	0.4276	5.0	0.0568
WGRCOR	6.0	0.1010	9.5	0.3310	6.0	0.0963
HALPOR	7.0	0.1328	9.5	0.3310	7.0	0.1271
WMICDFLG	8.0	0.1542	7.0	0.2502	8.0	0.1508
WRBRNSAT	9.0	0.2128	20.0	0.8034	9.0	0.2090
SHPRMHAL	10.0	0.2278	5.0	0.1129	10.0	0.2289
HALPRM	11.0	0.2436	16.0	0.7343	11.0	0.2467
SHPRMCON	12.0	0.4075	6.0	0.1129	12.0	0.4150
SHRBRSAT	13.0	0.4414	14.0	0.6566	13.0	0.4426
WFBETCEL	14.0	0.4530	13.0	0.4687	14.0	0.4611
WGRMICI	15.0	0.5369	21.0	0.8034	15.0	0.5379
SHPRMCLY	16.0	0.5739	24.0	0.9079	16.0	0.5780
SHRGSSAT	17.5	0.7885	11.0	0.3684	17.0	0.7941
ANHBCEXP	17.5	0.7885	22.0	0.8616	18.0	0.7979
SHPRMASP	19.0	0.8197	8.0	0.2955	19.0	0.8277
SALPRES	20.0	0.8296	15.0	0.7155	20.0	0.8326
ANRGSSAT	21.0	0.8296	4.0	0.0103	21.0	0.8379
WASTWICK	22.0	0.8392	23.0	0.8863	22.0	0.8396
ANHBCVGP	23.0	0.9197	18.0	0.7792	23.0	0.9334
SHPRMDRZ	24.0	0.9489	25.0	0.9612	24.0	0.9458
SHBCEXP	25.0	0.9537	17.0	0.7701	25.0	0.9574
TDC	1.000		0.746		0.972	

[a] Twenty-five (25) variables included in analysis; see Footnote b to Table 10.14.

[b, c, d, e] See Footnotes b, c, d, e to Table 10.5.

Again, a possible variant of the common means (CMNs) test is to use logarithmically transformed y-values rather than the original untransformed y-values (Sect. 5.1). Use of both raw and logarithmically transformed variables results in similar rankings of variable importance (Tables 10.17, 10.19). Thus, little is gained in the analysis of *E2:WAS_SATB* with CMNs by the use of logarithmically transformed variables. Similarly, little change in the outcome of the analysis for *E2:WAS_SATB* with CVs took place when t_{ql} as defined in Eq. (6.4) rather than as in Eq. (6.3) was used (Tables 10.17, 10.19).

10.4 Repository Pressure under Disturbed Conditions: $y = E2{:}WAS_PRES$

The variable y = *E2:WAS_PRES* was included as an example because regression analyses with raw and rank-transformed data fail to identify the dominant variable *BHPRM* (Sect. 2). Thus, procedures that can identify nonlinear, nonmonotonic relationships should work well with *E2:WAS_PRES*, which indeed turned out to be the case (Table 10.20). In particular, tests based on CMNs, CL, CMDs, CVs, CIQs and SI all identified *BHPRM* as the most important variable affecting *E2:WAS_PRES* (Table 10.20), which is consistent with the strong pattern appearing in the corresponding scatterplot (Fig. 2.6). In contrast, tests based on CCs and RCCs failed to identify *BHPRM* as an important variable (i.e., p-values of 0.3651 and 0.1704 for CCs and RCCs, respectively) (Table 10.20). Further,

Table 10.18 Comparison of Variable Rankings for $y = E2{:}WAS_SATB$ Obtained with Correlation Coefficients (CCs, RCCs) and Analytic Determination of p-values with Rankings Obtained with Monte Carlo Determination of p-values

Variable[a] Name	CC[b] Rank	CC[b] p-Val	CCMC[c] Rank	CCMC[c] p-Val	Variable Name	RCC Rank	RCC p-Val	RCCMC Rank	RCCMC p-Val
BHPRM	1.0	0.0000	1.5	0.0000	BHPRM	1.0	0.0000	2.0	0.0000
ANHPRM	2.0	0.0000	1.5	0.0000	WRGSSAT	2.0	0.0000	2.0	0.0000
HALPOR	3.0	0.0006	3.0	0.0004	ANHPRM	3.0	0.0001	2.0	0.0000
WGRCOR	4.0	0.0017	4.0	0.0020	SHPRMHAL	4.0	0.0225	4.0	0.0207
WRGSSAT	5.0	0.0081	5.0	0.0088	HALPOR	5.0	0.0269	5.0	0.0287
WMICDFLG	6.0	0.0214	6.0	0.0227	WGRCOR	6.0	0.1446	6.0	0.1480
WGRMICH	7.0	0.0838	7.0	0.0844	WMICDFLG	7.0	0.1745	7.0	0.1750
SHPRMHAL	8.0	0.0996	8.0	0.0998	WGRMICH	8.0	0.1842	8.0	0.1885
SHRBRSAT	9.0	0.1791	9.0	0.1792	ANRGSSAT	9.0	0.2751	9.0	0.2738
WGRMICI	10.0	0.1997	10.0	0.1989	SHRBRSAT	10.0	0.2921	10.0	0.2911
WRBRNSAT	11.0	0.2350	11.0	0.2377	SHRGSSAT	11.0	0.3784	11.0	0.3811
SALPRES	12.0	0.3892	12.0	0.3903	WFBETCEL	12.0	0.4852	13.0	0.4958
SHPRMASP	13.0	0.4801	13.0	0.4837	WRBRNSAT	13.0	0.4950	12.0	0.4928
ANHBCEXP	14.0	0.4911	14.0	0.4945	SHPRMDRZ	14.0	0.4993	14.0	0.5025
ANRBRSAT	15.0	0.6402	15.0	0.6409	WGRMICI	15.0	0.5517	15.0	0.5606
SHRGSSAT	16.0	0.6732	16.0	0.6780	SHPRMCLY	16.0	0.6137	16.0	0.6144
HALPRM	17.0	0.7348	17.0	0.7370	SALPRES	17.0	0.6140	17.0	0.6156
ANRGSSAT	18.0	0.7628	18.0	0.7655	SHPRMASP	18.0	0.6223	18.0	0.6167
WASTWICK	19.0	0.8393	19.0	0.8430	SHPRMCON	19.0	0.6549	19.0	0.6540
ANHBCVGP	20.0	0.8824	20.0	0.8865	ANRBRSAT	20.0	0.6645	20.0	0.6620
SHPRMCLY	21.0	0.9020	21.0	0.9090	SHBCEXP	21.0	0.7907	21.0	0.7907
WFBETCEL	22.0	0.9168	22.0	0.9204	ANHBCEXP	22.0	0.8319	22.0	0.8380
SHPRMCON	23.0	0.9478	23.0	0.9463	ANHBCVGP	23.0	0.9131	23.0	0.9125
SHPRMDRZ	24.0	0.9823	24.0	0.9824	HALPRM	24.0	0.9544	24.0	0.9569
SHBCEXP	25.0	0.9943	25.0	0.9943	WASTWICK	25.0	0.9832	25.0	0.9834
TDC[d]		0.988			*TDC*		0.972		

a Twenty-five (25) variables included in analysis; see Footnote b to Table 10.14.

b, c, d See Footnotes b, c, d to Table 10.6.

tests based on CMNs, CLs, CMDs, CVs, and SI select the variables ranked 2 and 3 from *HALPRM*, *ANHPRM* and *WGRCOR*, while the test based on CIQs assigns ranks 2 and 3 to *WGRCOR* and *SHRGSSAT*, respectively. As indicated by the associated scatterplots, *HALPRM*, *ANHPRM* and *WGRCOR* show barely discernible patterns (Figs. 2.6, 10.4).

Variable rankings for *E2:WAS_PRES* based on CCs, SRCs and PCCs and also on RCCs, SRRCs and PRCCs are the same (Table 10.21). However, these rankings are misleading because they do not include the dominant variable *BHPRM*.

Due to the failure of CCs and RCCs to identify the dominant variable *BHPRM*, there is less agreement as indicated by top-down correlations between the variable rankings obtained with the various analysis procedures for *E2:WAS_PRES* than is the case for *E0:WAS_PRES*, *E0:BRAALIC* and *E2:WAS_SATB* (i.e., compare the top-down correlations and associated p-values in Tables 10.4, 10.10, 10.16 and those in Table 10.22). In particular, variable

Table 10.19 Exceedance Probabilities (i.e., *p*-values) for Common Mean and Common Variance Tests Calculated with Use of Logarithms[a] for $y = E2{:}WAS_SATB$, the Variables in Table 2.2,[b] and a Maximum of Five Classes of Values for Each Variable (i.e., $nX = 5$)[c]

Variable Name	CMN: Log,1 × 5 Rank	CMN: Log,1 × 5 *p*-Val	CMNMC: Log,1 × 5 Rank	CMNMC: Log,1 × 5 *p*-Val
BHPRM	1.0	0.0000	1.5	0.0000
ANHPRM	2.0	0.0000	1.5	0.0000
WGRMICH	3.0	0.0017	3.0	0.0022
HALPOR	4.0	0.0072	4.0	0.0057
WMICDFLG	5.0	0.0263	5.0	0.0254
WGRCOR	6.0	0.0284	6.0	0.0290
WRGSSAT	7.0	0.0833	7.0	0.0801
WRBRNSAT	8.0	0.0914	8.0	0.0871

Variable Name	CV: Log,1 × 5 Rank	CV: Log,1 × 5 *p*-Val	CVMC: Log,1 × 10 Rank	CVMC: Log,1 × 10 *p*-Val
BHPRM	1.0	0.0000	1.0	0.0000
ANHPRM	2.0	0.0107	2.0	0.0112
HALPOR	3.0	0.0174	4.0	0.0159
WGRCOR	4.0	0.0190	3.0	0.0153
WMICDFLG	5.0	0.0257	5.0	0.0262
WGRMICH	6.0	0.0537	6.0	0.0418
SHPRMHAL	7.0	0.0803	7.0	0.0574

[a] See Footnote a, Table 10.7.

[b] See Footnote b, Table 10.14.

[c] See Footnote c, Table 10.1.

rankings with CMNs, CLs, CMDs CVs, CIQs and SI are generally similar (Table 10.22). The exception is the ranking based on CIQs, which shows top-down correlations of 0.429, 0.462 and 0.462 with the rankings obtained with CMNs, CLs and CMDs (Table 10.22). Otherwise, the top-down correlations for the variable rankings obtained with CMNs, CLs, CMDs, CVs, CIQs and SI vary between 0.698 and 1.000 (Table 10.22). In contrast, there is little relationship between the variable rankings obtained with CMNs, CLs, CMDs, CVs, CIQs and SI and with CCs, SRCs, PCCs, RCCs, SRRCs and PRCCs (Table 10.22).

An investigation of the effects of using $nX = 10$ rather than $nX = 5$ for CMNs, CLs, CMDs, CVs, CIQs and SI and $nY = 10$ rather than $nY = 5$ for SI was also carried out (Table 10.23). Each of the indicated procedures with $nX = 5$ and $nX = 10$ identified *BHPRM* as the most important variable. Generally, ranks 2, 3 and sometimes 4 were also assigned to similar variables, although the exact order was not always the same for $nX = 5$ and $nX = 10$. After rank 4, there was considerable variability in the ordering of the variables with $nX = 5$ and $nX = 10$.

Table 10.20 Comparison of Variable Rankings with Different Analysis Procedures[a] for y = *E2:WAS_PRES*, the Variables in Table 2.2,[b] and a Maximum of Five Classes of Values for each Variable (i.e., nX =5)[c]

Variable[d]	CC		RCC		CMN: 1 × 5		CL: 1 × 5	
Name	Rank	p-Val	Rank	p-Val	Rank	p-Val	Rank	p-Val
HALPRM	1.0	0.0000	1.0	0.0000	2.0	0.0000	2.0	0.0000
ANHPRM	2.0	0.0000	2.0	0.0000	3.0	0.0002	3.0	0.0000
HALPOR	3.0	0.0090	3.0	0.0184	5.0	0.0415	5.0	0.0940
SHPRMDRZ	6.0	0.1684	9.0	0.2417	13.0	0.4281	12.0	0.3131
ANHBCEXP	7.0	0.1786	8.0	0.2373	4.0	0.0405	4.0	0.0602
BHPRM	10.0	0.3651	6.0	0.1704	1.0	0.0000	1.0	0.0000
SHRGSSAT	14.0	0.5958	12.0	0.3948	25.0	0.9511	23.0	0.7738
ANRBRSAT	19.0	0.7133	14.0	0.4378	7.0	0.1513	7.0	0.1304
WGRCOR	20.0	0.7676	17.0	0.6560	17.0	0.5428	9.0	0.2242

Variable	CMD: 2 × 5		CV: 1 × 5		CIQ: 2 × 5		SI: 5 × 5	
Name	Rank	p-Val	Rank	p-Val	Rank	p-Val	Rank	Value
HALPRM	2.0	0.0000	2.0	0.0014	11.0	0.4530	2.0	0.0002
ANHPRM	3.0	0.0007	24.0	0.9251	12.0	0.4628	4.0	0.0049
HALPOR	5.0	0.0700	7.0	0.1410	18.0	0.6151	11.0	0.3142
SHPRMDRZ	17.0	0.6868	4.0	0.0298	13.0	0.5037	17.0	0.6111
ANHBCEXP	4.0	0.0595	16.0	0.5178	19.5	0.6868	14.0	0.4414
BHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
SHRGSSAT	22.0	0.8325	14.0	0.3905	3.0	0.0289	5.0	0.0698
ANRBRSAT	6.0	0.0823	22.0	0.7194	4.0	0.0739	10.0	0.2518
WGRCOR	14.5	0.5249	3.0	0.0296	2.0	0.0130	3.0	0.0002

a, b, c, d See Footnotes a, b, c, d to Table 10.14.

The p-values used to identify important variables in Tables 10.20, 10.21 and 10.23 were recalculated with the Monte Carlo procedure described in conjunction with Eq. (10.1) (Tables 10.23, 10.24). The rankings based on analytic determination of p-values and on Monte Carlo determination of p-values are very similar, with the primary difference being the tendency of the Monte Carlo simulation to assign tied ranks to the most important variables due to an inability to estimate p-values less than 0.0001 with the sample of size 10,000 in use.

Again, a possible variant of the common means (CMNs) test is to use logarithmically transformed y-values rather than the original untransformed y-values (Sect. 5.1). Use of both raw and logarithmically transformed variables results in similar rankings of variable importance (Tables 10.23, 10.25). Thus, little is gained in the analysis of *E2:WAS_PRES* with CMNs by the use of logarithmically transformed variables. In contrast, the analysis for *E2:WAS_PRES* with CVs and t_{ql} as defined in Eq. (6.4) with a logarithmic transformation rather than as in Eq. (6.3) without a transformation performed poorly, with the analysis with t_{ql} as defined in Eq. (6.4) failing to identify the dominant variable *BHPRM*. Thus, the use of the logarithmic transformation in Eq. (6.4) has the potential to improve the performance of the CVs test as it did for *E0:BRAALIC* (Tables 10.11, 10.13) and also the potential to degrade performance as is the case for *E2:WAS_PRES*.

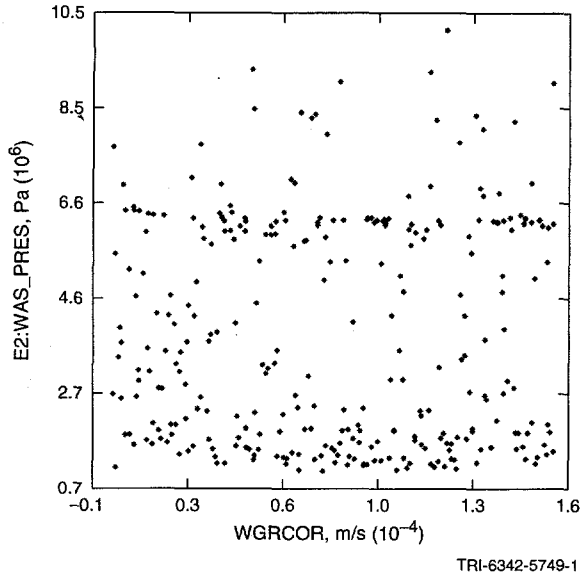

Fig. 10.4. Scatterplot for *E2:WAS_PRES* versus *WGRCOR*.

Table 10.21 Comparison of Variable Rankings with Correlation Coefficients, Standardized Regression Coefficients and Partial Correlation Coefficients with Raw and Rank Transformed Data for *y = E2:WAS_PRES*

Variable[a]		CC		SRC		PCC	
Name	*p*-Val	Rank	Value	Rank	Value	Rank	Value
HALPRM	0.0000	1.0	0.3754	1.0	0.3682	1.0	0.3934
ANHPRM	0.0000	2.0	0.2497	2.0	0.2341	2.0	0.2632
HALPOR	0.0090	3.0	0.1508	3.0	0.1511	3.0	0.1730

Variable[a]		RCC		SRRC		PRCC	
Name	*p*-Val	Rank	Value	Rank	Value	Rank	Value
HALPRM	0.0000	1.0	0.3591	1.0	0.3612	1.0	0.3868
ANHPRM	0.0000	2.0	0.2447	2.0	0.2388	2.0	0.2673
HALPOR	0.0184	3.0	0.1363	3.0	0.1412	3.0	0.1618

[a, b] See Footnotes a, b to Table 10.2.

Table 10.22 Top Down Correlation Matrix for Variable Rankings with Different Analysis Procedures for $y = E2{:}WAS_PRES$, Variables included in Table 10.20,[a] and a Maximum of Five Classes of x values (i.e., $nX = 5$)

Top-Down Correlation Matrix

	CC	RCC	CMN	CL	CMD	CV	CIQ	SI	SRC	PCC	SRRC
RCC	0.967										
CMN	0.398	0.577									
CL	0.378	0.567	0.997								
CMD	0.378	0.567	0.997	1.000							
CV	0.097	0.230	0.698	0.706	0.706						
CIQ	–.427	–.248	0.429	0.462	0.462	0.715					
SI	0.144	0.342	0.798	0.826	0.826	0.850	0.816				
SRC	0.990	0.975	0.423	0.408	0.408	0.072	–.438	0.149			
PCC	0.990	0.975	0.423	0.408	0.408	0.072	–.438	0.149	1.000		
SRRC	0.967	1.000	0.577	0.567	0.567	0.230	–.248	0.342	0.975	0.975	
PRCC	0.967	1.000	0.577	0.567	0.567	0.230	–.248	0.342	0.975	0.975	1.000

Top-Down Correlation Matrix p Values

	CC	RCC	CMN	CL	CMD	CV	CIQ	SI	SRC	PCC	SRRC
RCC	0.003										
CMN	0.130	0.051									
CL	0.143	0.055	0.002								
CMD	0.143	0.055	0.002	0.002							
CV	0.392	0.257	0.024	0.023	0.023						
CIQ	0.886	0.759	0.113	0.096	0.096	0.022					
SI	0.342	0.167	0.012	0.010	0.010	0.008	0.011				
SRC	0.003	0.003	0.116	0.124	0.124	0.420	0.892	0.337			
PCC	0.003	0.003	0.116	0.124	0.124	0.420	0.892	0.337	0.002		
SRRC	0.003	0.002	0.051	0.055	0.055	0.257	0.759	0.167	0.003	0.003	
PRCC	0.003	0.002	0.051	0.055	0.055	0.257	0.759	0.167	0.003	0.003	0.002

[a] Same as Footnote a to Table 10.4 except for use of 9 variables from Table 10.20.

Table 10.23 Comparison of Variable Rankings for y = *E2:WAS_PRES* Obtained with a Maximum of Five classes of x Values (i.e., nX = 5) and Analytic Determination of p-values with Variable Rankings Obtained with (i) a Maximum of Ten Classes of x values (i.e., nX = 10) and Analytic Determination of p-values and (ii) a Maximum of Five Classes of x values (i.e., nX = 5) and Monte Carlo Determination of p-values

Variable[a]	CMN: 1 × 5[b]		CMN: 1 × 10[c]		CMNMC: 1 × 5[d]	
Name	Rank	p-Val	Rank	p-Val	Rank	p-Val
BHPRM	1.0	0.0000	1.0	0.0000	1.5	0.0000
HALPRM	2.0	0.0000	2.0	0.0000	1.5	0.0000
ANHPRM	3.0	0.0002	3.0	0.0000	3.0	0.0002
ANHBCEXP	4.0	0.0405	9.0	0.2063	4.0	0.0419
HALPOR	5.0	0.0415	7.0	0.1914	5.0	0.0438
ANHBCVGP	6.0	0.1130	4.0	0.1130	6.0	0.1072
ANRBRSAT	7.0	0.1513	15.0	0.3538	7.0	0.1513
SHBCEXP	8.0	0.1773	10.0	0.2147	8.0	0.1733
WMICDFLG	9.0	0.2308	13.0	0.2308	9.0	0.2303
WGRMICH	10.0	0.3188	22.0	0.6611	11.0	0.3201
WRBRNSAT	11.0	0.3232	14.0	0.2593	10.0	0.3141
SHPRMASP	12.0	0.3767	18.0	0.4253	12.0	0.3782
SHPRMDRZ	13.0	0.4281	11.0	0.2218	13.0	0.4321
SHPRMCLY	14.0	0.4309	16.0	0.4072	14.0	0.4380
SHPRMHAL	15.0	0.4949	25.0	0.9026	15.0	0.4940
WASTWICK	16.0	0.5125	17.0	0.4218	16.0	0.5197
WGRCOR	17.0	0.5428	12.0	0.2308	17.0	0.5423
WGRMICI	18.0	0.5681	5.0	0.1514	18.0	0.5660
SHPRMCON	19.0	0.6443	23.0	0.7664	19.0	0.6446
SALPRES	20.0	0.6667	8.0	0.2029	20.0	0.6641
ANRGSSAT	21.0	0.7824	19.0	0.4343	21.0	0.7804
WRGSSAT	22.0	0.7937	21.0	0.5797	22.0	0.7885
SHRBRSAT	23.0	0.8724	20.0	0.4427	23.0	0.8731
WFBETCEL	24.0	0.9015	6.0	0.1751	24.0	0.8972
SHRGSSAT	25.0	0.9511	24.0	0.7887	25.0	0.9555
TDC[e]	1.000		0.805		0.988	

Variable	CL: 1 × 5		CL: 1x10		CLMC: 1 × 5	
Name	Rank	p-Val	Rank	p-Val	Rank	p-Val
BHPRM	1.0	0.0000	1.0	0.0000	2.0	0.0000
HALPRM	2.0	0.0000	3.0	0.0000	2.0	0.0000
ANHPRM	3.0	0.0000	2.0	0.0000	2.0	0.0000
ANHBCEXP	4.0	0.0602	10.0	0.2585	4.0	0.0625
HALPOR	5.0	0.0940	13.0	0.3454	5.0	0.0972
ANHBCVGP	6.0	0.1099	6.0	0.1099	6.0	0.1031
ANRBRSAT	7.0	0.1304	11.0	0.2851	7.0	0.1312
SHBCEXP	8.0	0.1919	12.0	0.2878	8.0	0.1887
WGRCOR	9.0	0.2242	4.0	0.1005	9.0	0.2266
WGRMICH	10.0	0.3022	22.0	0.6357	10.0	0.3026
WRBRNSAT	11.0	0.3127	9.0	0.2214	11.0	0.3062
SHPRMDRZ	12.0	0.3131	8.0	0.2149	12.0	0.3182
WGRMICI	13.0	0.3786	5.0	0.1036	13.0	0.3757
WASTWICK	14.0	0.4201	16.0	0.4434	14.0	0.4276
WMICDFLG	15.0	0.4247	15.0	0.4247	15.0	0.4291
SHPRMASP	16.0	0.4655	18.0	0.4688	16.0	0.4713
SHPRMHAL	17.0	0.5170	25.0	0.9158	17.0	0.5116
SALPRES	18.0	0.5210	7.0	0.1879	18.0	0.5222
WRGSSAT	19.0	0.6208	23.0	0.6534	19.0	0.6186
SHPRMCLY	20.0	0.6326	19.0	0.5135	20.0	0.6384
SHPRMCON	21.0	0.6790	21.0	0.6139	21.0	0.6768
ANRGSSAT	22.0	0.7276	17.0	0.4525	22.0	0.7305
SHRGSSAT	23.0	0.7738	24.0	0.6920	23.0	0.7692
WFBETCEL	24.0	0.8482	14.0	0.3540	24.0	0.8468
SHRBRSAT	25.0	0.9199	20.0	0.6008	25.0	0.9230
TDC	1.000		0.828		0.972	

Variable	CMD: 2 × 5		CMD: 2 × 10		CMDMC: 2 × 5	
Name	Rank	p-Val	Rank	p-Val	Rank	p-Val
BHPRM	1.0	0.0000	1.0	0.0000	1.5	0.0000
HALPRM	2.0	0.0000	2.0	0.0001	1.5	0.0000
ANHPRM	3.0	0.0007	3.0	0.0012	3.0	0.0005
ANHBCEXP	4.0	0.0595	8.0	0.2288	4.0	0.0583
HALPOR	5.0	0.0700	5.0	0.1596	5.0	0.0718
ANRBRSAT	6.0	0.0823	14.5	0.4311	6.0	0.0827
WMICDFLG	7.0	0.2187	7.0	0.2187	7.0	0.2133
SHPRMASP	8.0	0.2942	9.0	0.2451	8.0	0.2932
WGRMICH	9.5	0.3711	19.0	0.5075	9.0	0.3636
WASTWICK	9.5	0.3711	13.0	0.4071	11.0	0.3852
WRBRNSAT	11.0	0.3883	12.0	0.3614	10.0	0.3767
SHBCEXP	12.0	0.4628	21.0	0.6163	13.0	0.4534
ANHBCVGP	13.0	0.4884	18.0	0.4884	12.0	0.4104
SHPRMCON	14.5	0.5249	20.0	0.5749	15.0	0.5476
WGRCOR	14.5	0.5249	10.0	0.2622	14.0	0.5225
SHPRMCLY	16.0	0.5918	23.0	0.7265	16.0	0.5937
SHPRMDRZ	17.0	0.6868	11.0	0.3191	17.0	0.6724
WGRMICI	18.5	0.7113	6.0	0.2133	19.0	0.7147
SHPRMHAL	18.5	0.7113	25.0	0.9761	18.0	0.7129
ANRGSSAT	20.0	0.7358	16.0	0.4559	20.0	0.7266
SHRBRSAT	21.0	0.7847	14.5	0.4311	21.0	0.7646
SHRGSSAT	22.0	0.8325	24.0	0.7532	22.0	0.8405
SALPRES	23.0	0.8889	4.0	0.0966	23.0	0.8888
WFBETCEL	24.5	0.9197	17.0	0.4814	24.0	0.9080
WRGSSAT	24.5	0.9197	22.0	0.6993	25.0	0.9091
TDC	1.000		0.800		0.987	

Variable	CV: 1 × 5		CV: 1 × 10		CVMC: 1 × 5	
Name	Rank	p-Val	Rank	p-Val	Rank	p-Val
BHPRM	1.0	0.0000	1.0	0.0012	1.0	0.0000
HALPRM	2.0	0.0014	2.0	0.0201	2.0	0.0021
WGRCOR	3.0	0.0296	3.0	0.0491	3.0	0.0278
SHPRMDRZ	4.0	0.0298	5.0	0.0799	4.0	0.0280
ANHBCVGP	5.0	0.1173	6.0	0.1173	5.0	0.1184
WMICDFLG	6.0	0.1393	8.0	0.1393	8.0	0.1383
HALPOR	7.0	0.1410	10.0	0.1817	7.0	0.1347
SHRBRSAT	8.0	0.1453	15.0	0.3933	6.0	0.1311
SHPRMASP	9.0	0.1741	20.0	0.6816	9.0	0.1667
SHPRMCON	10.0	0.2598	21.0	0.7328	10.0	0.2427
WASTWICK	11.0	0.2823	12.0	0.3183	11.0	0.2689
SHPRMCLY	12.0	0.3595	17.0	0.4008	12.0	0.3483
WFBETCEL	13.0	0.3727	16.0	0.3978	13.0	0.3575
SHRGSSAT	14.0	0.3905	11.0	0.2905	14.0	0.3729
WGRMICI	15.0	0.4555	13.0	0.3219	15.0	0.4343
ANHBCEXP	16.0	0.5178	18.0	0.6371	16.0	0.4974
ANRGSSAT	17.0	0.5840	19.0	0.6807	17.0	0.5694
WRBRNSAT	18.0	0.6510	9.0	0.1809	18.0	0.6418
WGRMICH	19.0	0.7009	23.0	0.7826	19.0	0.6988
WRGSSAT	20.0	0.7041	7.0	0.1302	20.0	0.6997
SHBCEXP	21.0	0.7134	4.0	0.0694	22.0	0.7060
ANRBRSAT	22.0	0.7194	14.0	0.3899	21.0	0.7036
SHPRMHAL	23.0	0.8699	24.0	0.8476	23.0	0.8685
ANHPRM	24.0	0.9251	22.0	0.7599	24.0	0.9175
SALPRES	25.0	0.9938	25.0	0.9958	25.0	0.9940
TDC	1.000		0.824		.0.995	

Table 10.23 (Cont.)

Variable Name	CIQ: 2 × 5 Rank	CIQ: 2 × 5 *p*-Val	CIQ: 2 × 10 Rank	CIQ: 2 × 10 *p*-Val	CIQMC: 2 × 5 Rank	CIQMC: 2 × 5 *p*-Val
BHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000
WGRCOR	2.0	0.0130	5.0	0.0565	2.0	0.0132
SHRGSSAT	3.0	0.0289	2.0	0.0163	3.0	0.0277
ANRBRSAT	4.0	0.0739	4.0	0.0308	4.0	0.0704
SHRBRSAT	5.0	0.2093	17.0	0.5075	5.0	0.2055
WASTWICK	6.0	0.2427	9.0	0.2451	6.0	0.2431
SHPRMASP	7.0	0.2805	6.0	0.1719	7.0	0.2721
WRBRNSAT	8.0	0.2942	13.0	0.3838	8.0	0.2973
ANHBCVGP	9.0	0.3556	12.0	0.3556	9.0	0.3042
SHPRMCLY	10.0	0.4244	17.0	0.5075	10.0	0.4106
HALPRM	11.0	0.4530	20.5	0.6993	12.0	0.4698
ANHPRM	12.0	0.4628	11.0	0.3398	11.0	0.4620
SHPRMDRZ	13.0	0.5037	7.0	0.1917	14.0	0.4968
WFBETCEL	14.0	0.5037	10.0	0.2622	13.0	0.4929
WMICDFLG	15.0	0.5134	19.0	0.5134	16.0	0.5323
SHPRMHAL	16.0	0.5249	17.0	0.5075	15.0	0.5260
ANRGSSAT	17.0	0.5467	15.0	0.4814	17.0	0.5467
HALPOR	18.0	0.6151	22.0	0.7265	18.0	0.6065
WRGSSAT	19.5	0.6868	3.0	0.0235	20.0	0.6803
ANHBCEXP	19.5	0.6868	23.0	0.8285	19.0	0.6626
WGRMICI	21.0	0.7113	8.0	0.2288	21.0	0.7178
SHBCEXP	22.0	0.7358	14.0	0.4559	22.0	0.7240
WGRMICH	23.0	0.7847	24.0	0.8729	23.0	0.7605
SALPRES	24.0	0.8889	20.5	0.6993	24.0	0.8946
SHPRMCON	25.0	0.9702	25.0	0.9865	25.0	0.9764
TDC	1.000		0.754		0.999	

Variable Name	SI: 5 × 5 Rank	SI: 5 × 5 *p*-Val	SI: 10 × 10 Rank	SI: 10 × 10 *p*-Val	SIMC: 5 × 5 Rank	SIMC: 5 × 5 *p*-Val
BHPRM	1.0	0.0000	1.0	0.0000	1.5	0.0000
HALPRM	2.0	0.0002	4.0	0.0082	1.5	0.0000
WGRCOR	3.0	0.0002	2.0	0.0028	3.0	0.0002
ANHPRM	4.0	0.0049	3.0	0.0032	4.0	0.0033
SHRGSSAT	5.0	0.0698	22.0	0.8482	5.0	0.0699
SHBCEXP	6.0	0.1010	15.0	0.3495	6.0	0.0989
WGRMICI	7.0	0.1985	11.0	0.1646	7.0	0.2013
ANHBCVGP	8.0	0.2427	14.0	0.3398	8.0	0.2380
WRBRNSAT	9.0	0.2436	17.0	0.3878	9.0	0.2417
ANRBRSAT	10.0	0.2518	12.0	0.2623	10.0	0.2511
HALPOR	11.0	0.3142	8.0	0.0814	11.0	0.3202
SHRBRSAT	12.0	0.3142	16.0	0.3878	12.0	0.3207
WFBETCEL	13.0	0.3965	5.0	0.0296	13.0	0.3971
ANHBCEXP	14.0	0.4414	18.0	0.4075	14.0	0.4578
WASTWICK	15.0	0.5125	19.0	0.4276	15.0	0.5161
WGRMICH	16.0	0.5492	10.0	0.1043	16.0	0.5509
SHPRMDRZ	17.0	0.6111	23.0	0.8616	17.0	0.6161
SHPRMCLY	18.5	0.6482	21.0	0.7155	18.0	0.6492
ANRGSSAT	18.5	0.6482	13.0	0.2623	19.0	0.6573
SHPRMASP	20.0	0.6482	9.0	0.1043	20.0	0.6658
SALPRES	21.0	0.7089	6.0	0.0478	21.0	0.7323
WRGSSAT	22.0	0.8578	7.0	0.0524	22.0	0.8560
WMICDFLG	23.0	0.8859	25.0	0.9362	23.0	0.8893
SHPRMHAL	24.0	0.9064	24.0	0.8863	24.0	0.9102
SHPRMCON	25.0	0.9898	20.0	0.5316	25.0	0.9933
TDC	1.000		0.735		0.988	

a Twenty-five (25) variables included in analysis; see Footnote b to Table 10.14.

b, c, d, e See Footnotes b, c, d, e to Table 10.5.

Table 10.24 Comparison of Variable Rankings for $y = E2{:}WAS_PRES$ Obtained with Correlation Coefficients (CCs, RCCs) and Analytic Determination of p-values with Rankings Obtained with Monte Carlo Determination of p-values

Variable[a] Name	CC[b] Rank	CC[b] p-Val	CCMC[c] Rank	CCMC[c] p-Val
HALPRM	1.0	0.0000	1.5	0.0000
ANHPRM	2.0	0.0000	1.5	0.0000
HALPOR	3.0	0.0090	3.0	0.0098
ANHBCVGP	4.0	0.1123	4.0	0.1072
SHPRMASP	5.0	0.1606	5.0	0.1610
SHPRMDRZ	6.0	0.1684	6.0	0.1670
ANHBCEXP	7.0	0.1786	7.0	0.1795
WGRMICI	8.0	0.1905	8.0	0.1827
WRBRNSAT	9.0	0.2002	9.0	0.1996
BHPRM	10.0	0.3651	10.0	0.3689
SHPRMCLY	11.0	0.3874	11.0	0.3883
SHRBRSAT	12.0	0.4830	12.0	0.4890
WASTWICK	13.0	0.5507	13.0	0.5521
SHRGSSAT	14.0	0.5958	14.0	0.5992
SHBCEXP	15.0	0.6072	15.0	0.6020
WRGSSAT	16.0	0.6104	16.0	0.6098
WMICDFLG	17.0	0.6156	17.0	0.6208
SHPRMHAL	18.0	0.6659	18.0	0.6618
ANRBRSAT	19.0	0.7133	19.0	0.7148
WGRCOR	20.0	0.7676	20.0	0.7617
WGRMICH	21.0	0.8494	21.0	0.8563
SALPRES	22.0	0.9244	22.0	0.9288
WFBETCEL	23.0	0.9457	23.0	0.9453
SHPRMCON	24.0	0.9794	24.0	0.9798
ANRGSSAT	25.0	0.9891	25.0	0.9897
TDC[d]		0.988		

Variable Name	RCC Rank	RCC p-Val	RCCMC Rank	RCCMC p-Val
HALPRM	1.0	0.0000	1.5	0.0000
ANHPRM	2.0	0.0000	1.5	0.0000
HALPOR	3.0	0.0184	3.0	0.0194
ANHBCVGP	4.0	0.1099	4.0	0.1031
WGRMICI	5.0	0.1477	5.0	0.1444
BHPRM	6.0	0.1704	6.0	0.1746
SHPRMASP	7.0	0.1946	7.0	0.1896
ANHBCEXP	8.0	0.2373	8.0	0.2404
SHPRMDRZ	9.0	0.2417	9.0	0.2435
WRBRNSAT	10.0	0.3079	10.0	0.3127
SHPRMCLY	11.0	0.3920	11.0	0.3932
SHRGSSAT	12.0	0.3948	12.0	0.3962
WRGSSAT	13.0	0.4322	13.0	0.4315
ANRBRSAT	14.0	0.4378	14.0	0.4394
ANRGSSAT	15.0	0.4684	15.0	0.4679
SHRBRSAT	16.0	0.5247	16.0	0.5302
WGRCOR	17.0	0.6560	17.0	0.6493
WMICDFLG	18.0	0.7248	18.0	0.7236
SHPRMHAL	19.0	0.7249	19.0	0.7284
WASTWICK	20.0	0.8082	20.0	0.8079
WGRMICH	21.0	0.8865	21.0	0.8916
WFBETCEL	22.0	0.9000	22.0	0.9008
SALPRES	23.0	0.9289	23.0	0.9320
SHBCEXP	24.0	0.9389	24.0	0.9367
SHPRMCON	25.0	0.9918	25.0	0.9918
TDC		0.988		

a Twenty-five (25) variables included in analysis; see Footnote b to Table 10.14.

b, c, d See Footnotes b, c, d to Table 10.6.

Table 10.25 Exceedance Probabilities (i.e., *p*-values) for Common Mean and Common Variance Tests Calculated with use of Logarithms[a] for $y = E2{:}WAS_PRES$, the Variables in Table 2.2,[b] and a Maximum of Five Classes of Values for Each Variable (i.e., $nX = 5$)[c]

Variable	CMN: Log, 1×5		CMNMC: Log, 1×5	
Name	Rank	*p*-Val	Rank	*p*-Val
BHPRM	1.0	0.0000	1.5	0.0000
HALPRM	2.0	0.0000	1.5	0.0000
ANHPRM	3.0	0.0000	3.0	0.0001
ANHBCEXP	4.0	0.0432	4.0	0.0450
HALPOR	5.0	0.0991	5.0	0.1071

Variable	CV: Log, 1×5		CVMC: Log, 1×5	
Name	Rank	*p*-Val	Rank	*p*-Val
HALPRM	1.0	0.0018	1.0	0.0014
WGRCOR	2.0	0.0174	2.0	0.0136
SHPRMDRZ	3.0	0.0513	3.0	0.0387

[a] See Footnote a, Table 10.7.

[b] See Footnote b, Table 10.14.

[c] See Footnote c, Table 10.1.

11. Type I and II Errors

The sensitivity analysis techniques under discussion use p-values to indicate if a relationship appears to exist between an uncertain analysis input and a predicted analysis outcome. Clearly, it is desirable that the techniques identify the inputs that actually affect analysis outcomes (i.e., to avoid Type II errors, which correspond to the failure to identify important variables). As shown by the example analysis in Sect. 10, Type II errors can occur when the test for variable importance is inappropriate for the relationships that exist between analysis inputs and analysis outcomes (e.g., see the analyses for *E2:WAS_PRES* in Sect 10.4). Thus, a good analysis strategy is to use several different tests for variable importance and thus reduce the likelihood of overlooking an important variable (i.e., committing a Type II error).

In addition, it is also important that the techniques not identify inputs as having effects that are not actually present (i.e., to avoid Type I errors, which correspond to the indication of nonexistent effects for unimportant variables). If a variable has no effect on a particular analysis outcome, then the corresponding p-values generated from repeated random sampling should have a uniform distribution on the interval (0, 1) (i.e., $prob\ (\hat{p} < p) = prob\ (\hat{t} > t_p) = p$ and thus $\hat{p}$ has a uniform distribution on (0, 1), where $0 \le p \le 1$, *prob* denotes probability, and t_p and $\hat{t}$ are values of the statistic with p-values of p and $\hat{p}$, respectively). Similarly, if multiple unimportant variables are involved, their p-values from a single sampling should be uniformly distributed on (0, 1). Thus, for a specified p-value (i.e., p) and n unimportant variables, the likelihood *prob* (Type I | p, n) of committing a Type I error (actually, one or more Type I errors) is given by

$$prob\ (\text{Type I} \mid p, n) = 1 - (1 - p)^n, \tag{11.1}$$

with prob (Type I | p, n) increasing as each of p and n increases (Fig. 11.1). Thus, Type I errors cannot be avoided, and their likelihood of occurrence will be defined by Eq. (11.1) provided that the p-values for unimportant variables do indeed follow a uniform distribution.

The LHSs indicated in Eqs. (2.1)-(2.3), and on which the examples in Sect. 10 are based, involved 75 variables. However, 49 of these variables were not used in the calculation of *E0:WAS_PRES* and *E0:BRAALIC*; and 48 of these variables were not used in the calculation of *E2:WAS_SATB* and *E2:WAS_PRES*. Thus, the p-values associated with these variables for the individual tests should have uniform distributions on the interval (0, 1). The Kolmogorov-Smirnov test (Chapt. 6, Conover, 1980) can be used to indicate if the distributions of p-values for these variables for the individual tests do indeed have uniform distributions on (0, 1). In particular, the 0.9 and 0.99 two-sided Kolmogorov-Smirnov bounds around the cumulative distribution function (CDF) for the true distribution (i.e., a uniform distribution on (0, 1) in the present context) are given by $1.22/(n+\sqrt{n/10})^{1/2}$ and $1.63/(n+\sqrt{n/10})^{1/2}$, respectively, where n is the sample size (Table A14, Conover, 1980). For $n = 48, 49$, the corresponding 0.9 and 0.99 bounds are 0.17 and 0.23, respectively.

As 4 dependent variables (i.e., *E0:WAS_PRES, E0:BRAALIC, E2:WAS_SATB, E2:WAS_PRES)* and 8 tests (i.e., CC, RCC, CMN, CL, CMD, CV, CIQ, SI) under consideration, 32 distributions of *p*-values result (Fig. 11.2). Of these, 24 are within the 0.9 bounds. Further, 6 of the 9 distributions that are outside the bounds are for the variable/test pairs (*E0:WAS_PRES,* CC), (*E0:BRAALIC,* CC), (*E2:WAS_SATB*, CC), (*E0:WAS_PRES*, RCC), (*E0:BRAALIC*, RCC), and (*E2:WAS_SATB*, RCC). As results obtained with CCs and RCCs are not independent, the indicated deviations of (*E0:WAS_PRES*, CC) (*E0:BRAALIC*, CC) and (*E2:WAS_SATB*, CC) from a uniform distribution on (0, 1) are not independent of the indicated deviations for (*E0:WAS_PRES,* RCC), (*E0:BRAALIC,* RCC), and (*E2:WAS_SATB*, RCC). The most notable deviation occurs for the pair (*E0:BRAALIC*, CV), with no *p*-values exceeding 0.7 being observed; there is something associated with values of *E0:BRAALIC* that is causing an underrepresentation of large *p*-values for unimportant variables, with this underrepresentation probably deriving from the fact that *E0:BRAALIC* has a few large values and many very small values. Fortunately, the shape of the individual CDFs in Fig. 11.2 does not suggest any tendency for the tests under consideration to produce unusual numbers of small *p*-values; thus, there does not appear to be a tendency to produce excessive numbers of Type I errors in the examples under consideration. However, the results in Fig. 11.2 do suggest that the *p*-values for unimportant variables may not have a uniform distribution on (0, 1). We have not succeeded in finding a completely satisfying explanation for this behavior.

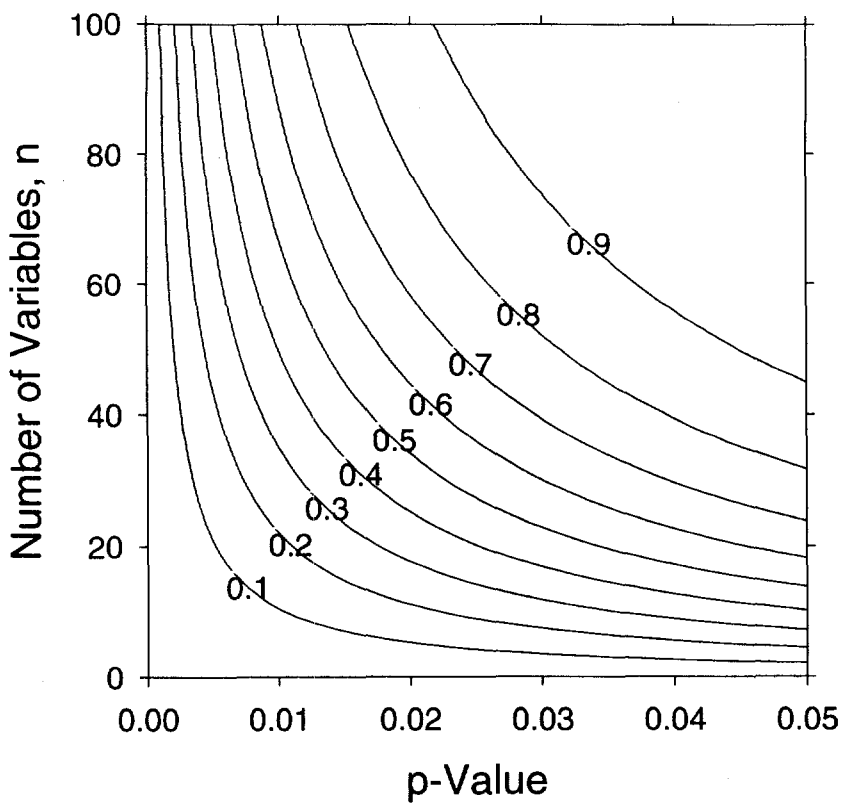

Figure 11.1. Contour plots for probability of a Type I error, *prob* (Type I | *p*, *n*), as a function of *p*-value, *p*, and number of unimportant variables, *n*.

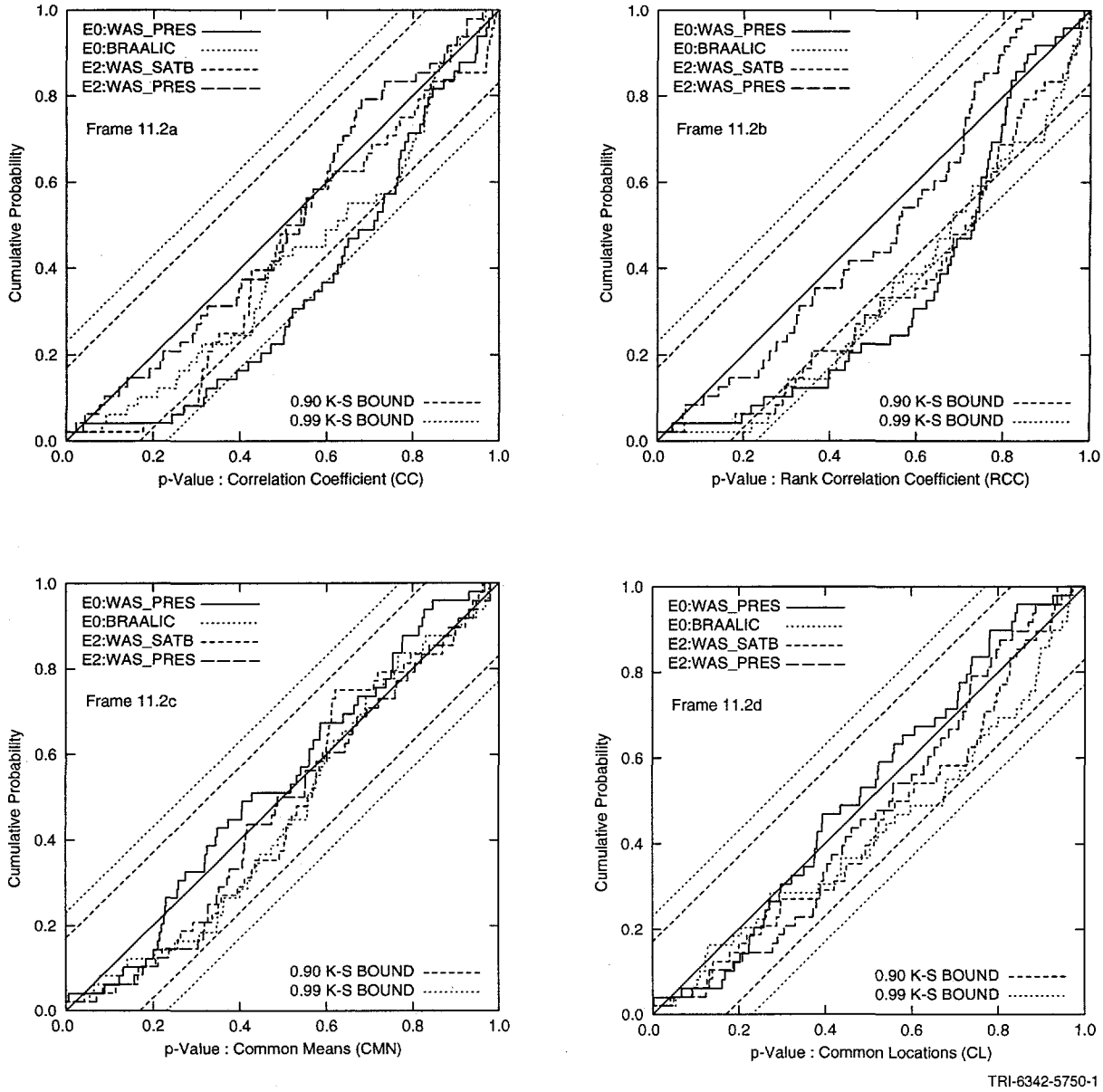

Figure 11.2. Distribution of *p*-values and associated Kolmogorov-Smirnov bounds for individual tests and variables in LHS that do not affect *E0:WAS_PRES*, *E0:BRAALIC*, *E2:WAS_SATB* and *E2:WAS_PRES*.

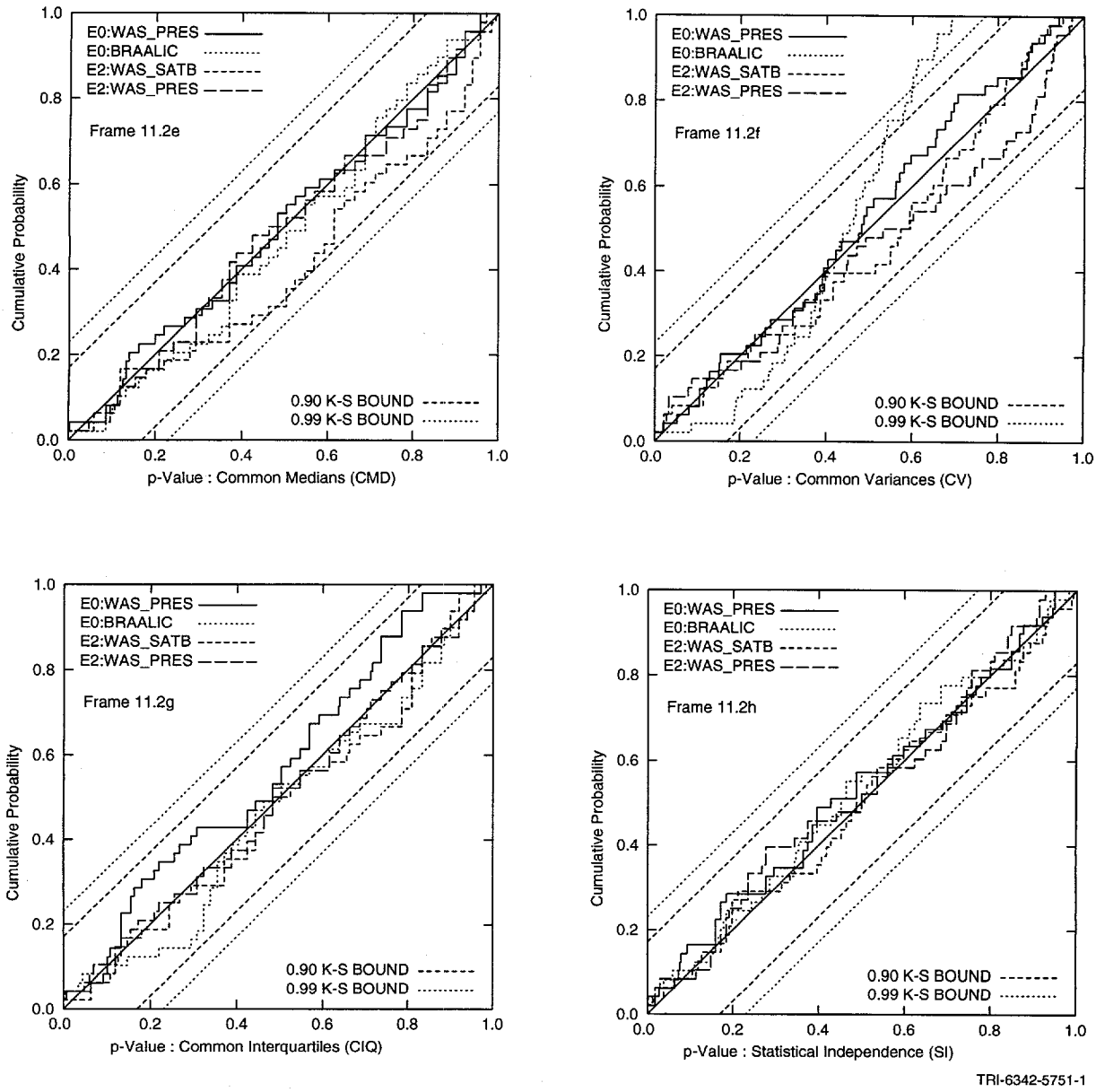

Figure 11.2. Distribution of *p*-values and associated Kolmogorov-Smirnov bounds for individual tests and variables in LHS that do not affect *E0:WAS_PRES*, *E0:BRAALIC*, *E2:WAS_SATB* and *E2:WAS_PRES* (continued).

An additional set of simulations was carried out to provide a check on the reasonableness of the distributions of p-values in Fig. 11.2. In particular, 10 independent LHSs of size 300 were generated with the Iman and Conover (1982) restricted pairing technique from 50 independent variables with uniform distributions on the interval [0, 1]. These LHSs were then associated with the calculated values for *E0:WAS_PRES*, *E0:BRAALIC*, *E2:WAS_SATB* and *E2:WAS_PRES* obtained with the original LHS of size 300 discussed in Sect. 2, and the corresponding distributions of p-values were calculated for the preceding four output variables, each of the eight tests under consideration, and each of the 10 independent LHSs. The p-values were calculated with the previously described analytic procedures. The outcome is 10 CDFs for each of the 32 test/output variable pairs.

If the assumptions of the tests are met and the calculations are implemented correctly, then the CDFs for each test/dependent variable pair should approximate a uniform distribution on [0, 1]. This generally appears to be the case. For example, the original CDFs for *E0:WAS_PRES* and tests based on CCs and RCCs move across the 0.99 Kolmogorov-Smirnov boundary (Figs. 11.2a, b). In contrast, the current exercise with 10 independently generated LHSs produces CDFs of p-values that generally stay within the 0.9 Kolomogorov-Smirnov bounds (Fig. 11.3).

Twenty-nine of the remaining 30 test/output variable pairs produced distributions of p-value CDFs that were similar to the two CDF distributions in Fig. 11.3. The exception to this consistency occurred for *E0:BRAALIC* and the CVs test (Fig. 11.4). For this test/output variable pair, the p-values remain below approximately 0.7, which was also the case in Fig. 11.2f. The variable *E0:BRAALIC* has a large number of values that are effectively zero (Figs. 2.2b, 2.4). As a result, the estimated variances t_{ql} in Eq. (6.3) of Ref. 1 used to define the F statistic for the CVs test do not have a normal distribution for the individual independent variables, and so the associated p-values do not have a uniform distribution on [0, 1] even though the independent variables have no effect on *E0:BRAALIC*.

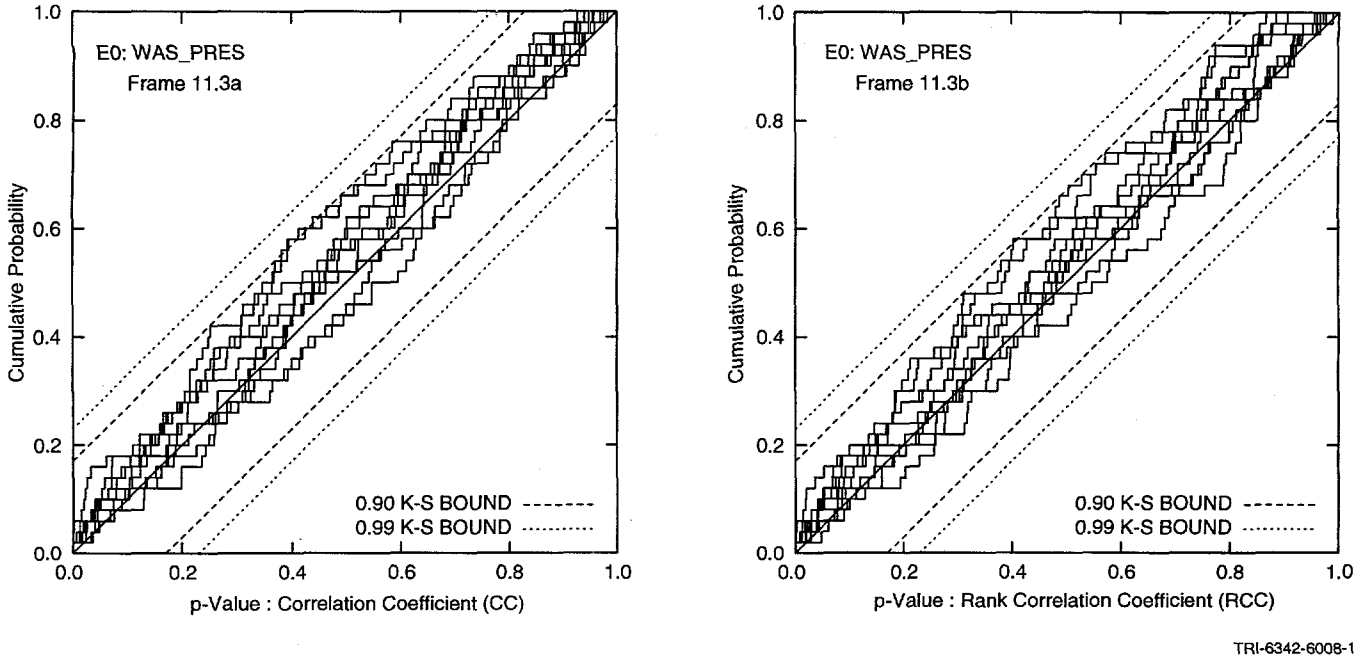

Figure 11.3. Distributions of *p*-values for 10 independently generated LHSs: (11.3a) CCs for *E0:WAS_PRES* and (11.3b) RCCs for *E0:WAS_PRES*.

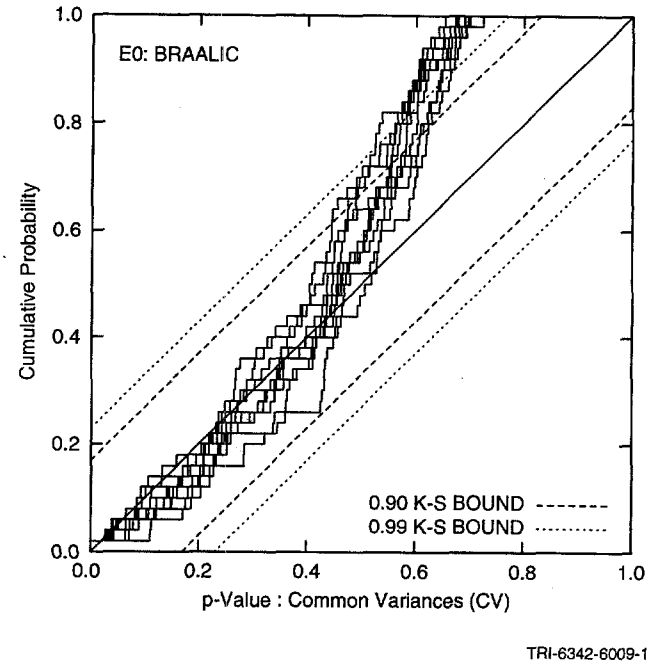

Figure 11.4. Distribution of *p*-values for 10 independently generated LHSs for CVs test and *E0:BRAALIC*.

12. Robustness with Respect to Repeated Independent Samples

The examples in Sect. 10 use a sample of size 300 obtained by pooling the three samples of size 100 each indicated in Eqs. (2.1)-(2.3). The availability of these three independent samples provides a way to examine the robustness of the techniques illustrated in Sect. 10. In particular, the analyses with each of the 8 techniques under consideration can be repeated with the individual samples of size 100. The extent to which the individual samples agree in the identification of important variables then provides an indication of how robust the techniques are with respect to repeated independent samples and also reductions in sample size (Table 12.1).

When comparing the variable selections in Table 12.1, it is important to keep in mind that the likelihood of a Type I error increases rapidly as p-values increase (Fig. 11.1), with 25 variables and a p-value of 0.01 producing a probability of 0.22 of a Type I error as indicated in Eq. (11.1). Further, the p-values for unimportant variables may not always be random on (0, 1) due to various patterns that are imposed on the data by the effects of other variables (Fig. 11.2). Thus, the probabilities in Fig. 11.1 are, at best, only an indication of the likelihood of a Type I error. As a result, the comparison of sets of important variables obtained with different replicates is probably valid only for variables with fairly low p-values. As p-values increase (e.g., >0.01), such comparisons become less and less meaningful.

The overall pattern that emerges from the results in Table 12.1 is that the most important variables identified with the pooled sample of size 300 are also identified as being important with the three individual samples of size 100. In particular, the two most important variables as defined by the size of their p-values are typically the same across all four samples for the individual tests (i.e., CCs, RCCs, CMs, CLs, CMDs, CVs, CIQ, SI), although it should be recognized that the results obtained with the pooled sample are not independent of the results obtained with the individual samples. Hence, the use of a sample size of 300 or 100 made little difference with respect to the variables identified as being most important, although the larger sample size did tend to indicate likely effects for more variables than was the case for the smaller sample size.

The most notable deviations from this consistency occur for the CVs test for *E0:BRAALIC* and *E2:WAS_PRES* and the CIQ test for *E0:BRAALIC*. The variable *E0:BRAALIC* is significantly affected by both *WMICDFLG* and *ANHPRM* (Fig. 2.4). However, as *WMICDFLG* is being missed by the CVs test, it is perhaps not surprising that the individual samples are not producing consistent results. A logarithmic transformation improved the results obtained with the CVs test for *WMICDFLG* with the pooled sample (Table 10.13) and also produced somewhat better results for the individual samples (Table 12.2). The variable *E2:WAS_PRES* is almost completely dominated by *BHPRM* (Fig. 2.6), with this effect being missed by the CVs test for replicate R3; further, although *BHPRM* is identified by

Table 12.1 Comparison of Variable Rankings Obtained with Different Analysis Procedures[a] for Three Independent Samples of Size 100 (Columns AP:R1, AP:R2, AP:R3, where AP ~ CC, RCC, CMN, CL, CMD, CV, CIQ, SI as appropriate), Pooled Sample of Size 300 (Column AP:All), and a Maximum of Five Classes of Values for Each Variable (i.e., $nX = 5$)[b]

Variable Name	CC: All Rank	CC: All *p*-Val	CC: R1 Rank	CC: R1 *p*-Val	CC: R2 Rank	CC: R2 *p*-Val	CC: R3 Rank	CC: R3 *p*-Val
Correlation Coefficients (CCs) for *E0:WAS_PRES*								
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
HALPOR	2.0	0.0000	2.0	0.0000	2.0	0.0000	2.0	0.0001
WGRCOR	3.0	0.0000	3.0	0.0180	3.0	0.0051	3.0	0.0018
ANHPRM	4.0	0.0241	9.0	0.3947	4.0	0.2371	4.0	0.0598
SALPRES	5.0	0.0855	4.0	0.0822	18.0	0.8602	7.0	0.2824
Correlation Coefficients (CCs) for *E2:WAS_SATB*								
BHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
ANHPRM	2.0	0.0000	2.0	0.0003	2.0	0.0281	2.0	0.0033
HALPOR	3.0	0.0006	3.0	0.0884	4.0	0.0706	3.0	0.0159
WGRCOR	4.0	0.0017	6.0	0.1241	3.0	0.0547	5.0	0.0473
WRGSSAT	5.0	0.0081	8.0	0.1367	11.0	0.5224	4.0	0.0175

Variable Name	CC: All Rank	CC: All *p*-Val	CC: R1 Rank	CC: R1 *p*-Val	CC: R2 Rank	CC: R2 *p*-Val	CC: R3 Rank	CC: R3 *p*-Val
Correlation Coefficients (CCs) for *E0:BRAALIC*								
ANHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
WMICDFLG	2.0	0.0000	2.0	0.0016	2.0	0.0060	2.0	0.0000
WASTWICK	3.0	0.0045	4.0	0.0584	9.0	0.2948	4.0	0.0333
WGRCOR	4.0	0.0048	5.0	0.0957	3.0	0.0318	9.0	0.3018
ANHBCEXP	5.0	0.0095	3.0	0.0420	6.0	0.1474	12.0	0.4274
Correlation Coefficients (CCs) for *E2:WAS_PRES*								
HALPRM	1.0	0.0000	1.0	0.0013	1.0	0.0000	1.0	0.0001
ANHPRM	2.0	0.0000	2.0	0.0020	2.0	0.0303	2.0	0.0267
HALPOR	3.0	0.0090	4.0	0.1417	3.0	0.0680	5.0	0.2188
ANHBCVGP	4.0	0.1123	8.0	0.3286	5.0	0.1492	20.0	0.7457
SHPRMASP	5.0	0.1606	10.0	0.3784	16.0	0.5907	6.0	0.3115

Variable Name	RCC: All Rank	RCC: All *p*-Val	RCC: R1 Rank	RCC: R1 *p*-Val	RCC: R2 Rank	RCC: R2 *p*-Val	RCC: R3 Rank	RCC: R3 *p*-Val
Rank Correlation Coefficients (RCCs) for *E0:WAS_PRES*								
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
HALPOR	2.0	0.0000	2.0	0.0000	2.0	0.0000	2.0	0.0001
WGRCOR	3.0	0.0000	3.0	0.0286	3.0	0.0041	3.0	0.0051
ANHPRM	4.0	0.0268	9.0	0.4366	4.0	0.1070	5.0	0.1268
SALPRES	5.0	0.0664	4.0	0.1111	16.0	0.7611	4.0	0.0957
Rank Correlation Coefficients (RCCs) for *E2:WAS_SATB*								
BHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
WRGSSAT	2.0	0.0000	2.0	0.0000	2.0	0.0048	2.0	0.0000
ANHPRM	3.0	0.0001	3.0	0.0013	3.0	0.1182	3.0	0.0335
SHPRMHAL	4.0	0.0225	4.0	0.1842	5.0	0.1243	9.0	0.2595
HALPOR	5.0	0.0269	8.5	0.4570	4.0	0.1236	4.0	0.1398

Variable Name	RCC: All Rank	RCC: All *p*-Val	RCC: R1 Rank	RCC: R1 *p*-Val	RCC: R2 Rank	RCC: R2 *p*-Val	RCC: R3 Rank	RCC: R3 *p*-Val
Rank Correlation Coefficients (RCCs) for *E0:BRAALIC*								
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
ANHPRM	2.0	0.0000	2.0	0.0000	2.0	0.0000	2.0	0.0000
HALPRM	3.0	0.0014	5.0	0.1867	5.0	0.0998	3.0	0.0140
WGRCOR	4.0	0.0057	4.0	0.1772	6.0	0.1383	4.0	0.0570
HALPOR	5.0	0.0087	3.0	0.0980	3.0	0.0396	7.0	0.3723
Rank Correlation Coefficients (RCCs) for *E2:WAS_PRES*								
HALPRM	1.0	0.0000	2.0	0.0110	1.0	0.0000	1.0	0.0001
ANHPRM	2.0	0.0000	1.0	0.0036	2.0	0.0820	2.0	0.0086
HALPOR	3.0	0.0184	4.0	0.1194	6.0	0.2015	5.0	0.2157
ANHBCVGP	4.0	0.1099	11.0	0.2611	4.0	0.1347	18.0	0.8795
WGRMICI	5.0	0.1477	5.0	0.1275	8.0	0.3344	23.0	0.9673

Variable Name	CMN: All,1×5 Rank	CMN: All,1×5 *p*-Val	CMN: R1,1×5 Rank	CMN: R1,1×5 *p*-Val	CMN: R2,1×5 Rank	CMN: R2,1×5 *p*-Val	CMN: R3,1×5 Rank	CMN: R3,1×5 *p*-Val
Common Means (CMNs) for *E0:WAS_PRES*								
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
HALPOR	2.0	0.0000	2.0	0.0002	2.0	0.0000	2.0	0.0010
WGRCOR	3.0	0.0000	3.0	0.0051	3.0	0.0093	3.0	0.0107
ANHPRM	4.0	0.0195	10.0	0.4751	6.0	0.2920	7.0	0.2881
SHPRMASP	5.0	0.1439	21.0	0.8597	5.0	0.1824	12.0	0.5410
Common Means (CMNs) for *E2:WAS_SATB*								
BHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
ANHPRM	2.0	0.0000	2.0	0.0031	2.0	0.0020	3.0	0.0544
WGRMICH	3.0	0.0021	4.0	0.0416	3.0	0.0471	2.0	0.0070
HALPOR	4.0	0.0124	10.0	0.2900	8.0	0.2345	10.0	0.3619
WRGSSAT	4.0	0.0143	5.0	0.0429	20.0	0.5575	6.0	0.1363

Variable Name	CMN: All,1×5 Rank	CMN: All,1×5 *p*-Val	CMN: R1,1×5 Rank	CMN: R1,1×5 *p*-Val	CMN: R2,1×5 Rank	CMN: R2,1×5 *p*-Val	CMN: R3,1×5 Rank	CMN: R3,1×5 *p*-Val
Common Means (CMNs) for *E0:BRAALIC*								
ANHPRM	1.0	0.0000	1.0	0.0014	1.0	0.0000	1.0	0.0000
WMICDFLG	2.0	0.0000	2.0	0.0040	2.0	0.0069	2.0	0.0001
SHPRMCON	3.0	0.0057	12.0	0.3818	8.0	0.3098	7.0	0.1531
WGRCOR	4.0	0.0636	5.0	0.1989	11.0	0.3914	20.0	0.5713
WFBETCEL	5.0	0.0732	10.0	0.3274	18.0	0.6060	12.0	0.2874
Common Means (CMNs) for *E2:WAS_PRES*								
BHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
HALPRM	2.0	0.0000	3.0	0.0288	2.0	0.0016	2.0	0.0027
ANHPRM	3.0	0.0002	2.0	0.0286	6.0	0.1137	5.0	0.1184
ANHBCEXP	4.0	0.0405	6.0	0.1860	5.0	0.1137	4.0	0.0230
HALPOR	5.0	0.0415	16.0	0.5971	4.0	0.0956	15.0	0.6365

Table 12.1 (Cont.)

Variable Name	CL: All,1×5 Rank	CL: All,1×5 *p*-Val	CL: R1,1×5 Rank	CL: R1,1×5 *p*-Val	CL: R2,1×5 Rank	CL: R2,1×5 *p*-Val	CL: R3,1×5 Rank	CL: R3,1×5 *p*-Val
Common Locations (CLs) for *E0:WAS_PRES*								
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
HALPOR	2.0	0.0000	2.0	0.0003	2.0	0.0000	2.0	0.0023
WGRCOR	3.0	0.0000	3.0	0.0112	3.0	0.0093	3.0	0.0179
ANHPRM	4.0	0.0187	6.0	0.3792	6.0	0.2595	8.0	0.3770
SHPRMASP	5.0	0.1237	19.0	0.7696	5.0	0.1901	11.0	0.4537
Common Locations (CLs) for *E2:WAS_SATB*								
BHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
WRGSSAT	2.0	0.0000	2.0	0.0000	3.0	0.0450	2.0	0.0001
ANHPRM	3.0	0.0001	3.0	0.0102	2.0	0.0184	7.0	0.2010
WGRMICH	4.0	0.0059	9.0	0.1714	5.0	0.0979	3.0	0.0206
SHPRMCON	5.0	0.0202	13.0	0.4691	10.0	0.2278	10.0	0.3785

Variable Name	CL: All, 1×5 Rank	CL: All, 1×5 *p*-Val	CL: R1,1×5 Rank	CL: R1,1×5 *p*-Val	CL: R2,1×5 Rank	CL: R2,1×5 *p*-Val	CL: R3,1×5 Rank	CL: R3,1×5 *p*-Val
Common Locations (CLs) for *E0:BRAALIC*								
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
ANHPRM	2.0	0.0000	2.0	0.0000	2.0	0.0000	2.0	0.0000
HALPRM	3.0	0.0019	4.0	0.2667	6.0	0.2321	3.0	0.0125
WGRCOR	4.0	0.0427	6.0	0.3340	10.0	0.3212	13.0	0.4371
SHPRMDRZ	5.0	0.1060	5.0	0.2785	15.0	0.5898	9.0	0.2393
Common Locations (CLs) for *E2:WAS_PRES*								
BHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
HALPRM	2.0	0.0000	4.0	0.1176	2.0	0.0025	2.0	0.0028
ANHPRM	3.0	0.0000	2.0	0.0154	3.0	0.0523	4.0	0.0419
ANHBCEXP	4.0	0.0602	7.0	0.2213	6.0	0.1191	5.0	0.0438
HALPOR	5.0	0.0940	18.0	0.5620	9.0	0.2452	11.0	0.5243

Variable Name	CMD: All, 2×5 Rank	CMD: All, 2×5 *p*-Val	CMD: R1,2×5 Rank	CMD: R1,2×5 *p*-Val	CMD: R2,2×5 Rank	CMD: R2,2×5 *p*-Val	CMD: R3,2×5 Rank	CMD: R3,2×5 *p*-Val
Common Medians (CMDs) for *E0:WAS_PRES*								
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
HALPOR	2.0	0.0000	2.0	0.0001	2.0	0.0000	2.0	0.0123
WGRCOR	3.0	0.0025	5.0	0.1712	3.0	0.0663	4.0	0.1712
ANHPRM	4.0	0.0663	16.5	0.7358	16.0	0.6626	5.0	0.1991
SHPRMASP	5.0	0.2427	16.5	0.7358	7.0	0.2674	6.0	0.2674
Common Medians (CMDs) for *E2:WAS_SATB*								
BHPRM	1.0	0.0000	2.0	0.0000	1.0	0.0000	2.0	0.0001
WRGSSAT	2.0	0.0000	1.0	0.0000	2.0	0.0015	1.0	0.0000
ANHPRM	3.0	0.0003	3.0	0.0073	13.0	0.2674	8.0	0.2674
WGRMICH	4.0	0.0130	9.0	0.1712	6.5	0.0477	4.0	0.0916
SHPRMCON	5.0	0.0206	16.0	0.7358	3.0	0.0244	6.5	0.1991

Variable Name	CMD: All,2×5 Rank	CMD: All,2×5 *p*-Val	CMD: R1,2×5 Rank	CMD: R1,2×5 *p*-Val	CMD: R2,2×5 Rank	CMD: R2,2×5 *p*-Val	CMD: R3,2×5 Rank	CMD: R3,2×5 *p*-Val
Common Medians (CMDs) for *E0:BRAALIC*								
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
ANHPRM	2.0	0.0000	2.0	0.0009	2.0	0.0001	2.0	0.0003
HALPRM	3.0	0.0050	17.5	0.7358	13.0	0.4060	3.0	0.0021
HALPOR	4.0	0.0155	8.5	0.3084	5.0	0.0563	11.0	0.4060
WGRCOR	5.0	0.0231	3.0	0.1257	3.0	0.0244	16.0	0.5918
Common Medians (CMDs) for *E2:WAS_PRES*								
BHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
HALPRM	2.0	0.0000	5.0	0.0663	2.0	0.0113	2.5	0.0289
ANHPRM	3.0	0.0007	3.0	0.0477	4.0	0.0780	5.0	0.1074
ANHBCEXP	4.0	0.0595	2.0	0.0289	3.0	0.0663	2.5	0.0289
HALPOR	5.0	0.0700	15.0	0.5918	5.5	0.1468	21.5	0.8781

Variable Name	CV: All,1×5 Rank	CV: All,1×5 *p*-Val	CV: R1,1×5 Rank	CV: R1,1×5 *p*-Val	CV: R2,1×5 Rank	CV: R2,1×5 *p*-Val	CV: R3,1×5 Rank	CV: R3,1×5 *p*-Val
Common Variances (CVs) for *E0:WAS_PRES*								
WMICDFLG	1.0	0.0000	1.0	0.0016	1.0	0.0058	1.0	0.0004
ANHPRM	2.0	0.0042	17.0	0.8561	2.0	0.0171	2.0	0.0671
HALPRM	3.0	0.1184	14.0	0.6965	10.0	0.4818	9.0	0.2016
WGRCOR	4.0	0.1244	16.0	0.8147	7.0	0.3521	3.0	0.0943
SHPRMCON	5.0	0.1287	22.0	0.9555	3.0	0.0381	10.0	0.2053
Common Variances (CVs) for *E2:WAS_SATB*								
BHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0001	1.0	0.0000
ANHPRM	2.0	0.0000	2.0	0.0030	2.0	0.0018	2.0	0.0134
HALPOR	3.0	0.0011	5.0	0.0243	6.0	0.1228	12.0	0.3601
WGRMICH	4.0	0.0050	3.0	0.0156	7.0	0.1353	3.0	0.1013
WGRCOR	5.0	0.0067	4.0	0.0210	3.0	0.0080	10.0	0.3387

Variable Name	CV: All,1×5 Rank	CV: All,1×5 *p*-Val	CV: R1,1×5 Rank	CV: R1,1×5 *p*-Val	CV: R2,1×5 Rank	CV: R2,1×5 *p*-Val	CV: R3,1×5 Rank	CV: R3,1×5 *p*-Val
Common Variances (CVs) for *E0:BRAALIC*								
ANHPRM	1.0	0.0078	1.0	0.2779	1.0	0.0576	1.0	0.0005
SHPRMCON	2.0	0.0426	13.0	0.4026	6.0	0.0938	22.0	0.6452
SHBCEXP	3.0	0.1463	17.0	0.4412	21.0	0.5811	23.0	0.6840
ANRBRSAT	4.0	0.1994	6.0	0.3463	19.0	0.5557	9.0	0.1909
WGRCOR	5.0	0.2125	11.0	0.3969	9.0	0.4550	15.0	0.4175
Common Variances (CVs) for *E2:WAS_PRES*								
BHPRM	1.0	0.0000	1.0	0.0082	1.0	0.0633	11.0	0.2843
HALPRM	2.0	0.0014	4.0	0.1392	4.0	0.1719	9.0	0.2669
WGRCOR	3.0	0.0296	2.0	0.0329	8.0	0.3272	5.0	0.0741
SHPRMDRZ	4.0	0.0298	3.0	0.0713	5.0	0.2094	13.0	0.3795
ANHBCVGP	5.0	0.1173	20.0	0.7401	9.0	0.4415	1.0	0.0178

Table 12.1 (Cont.)

Variable Name	CIQ: All, 2×5 Rank	CIQ: All, 2×5 *p*-Val	CIQ: R1,2×5 Rank	CIQ: R1,2×5 *p*-Val	CIQ: R2,2×5 Rank	CIQ: R2,2×5 *p*-Val	CIQ: R3,2×5 Rank	CIQ: R3,2×5 *p*-Val
	Common Interquartile (CIQ) for *E0:WAS_PRES*							
WMICDFLG	1.0	0.0000	1.0	0.0057	1.0	0.0012	1.0	0.0001
HALPOR	2.0	0.0000	2.0	0.1257	2.0	0.0206	2.0	0.0061
ANHPRM	3.0	0.0007	8.5	0.4628	3.0	0.0342	5.0	0.0780
SHPRMCON	4.0	0.0244	17.0	0.6626	16.0	0.4628	10.0	0.3546
WGRMICI	5.0	0.0595	13.5	0.5918	6.5	0.1468	17.0	0.8088
	Common Interquartile (CIQ) for *E2:WAS_SATB*							
WRGSSAT	1.0	0.0000	1.0	0.0001	1.0	0.0001	1.0	0.0000
WGRCOR	2.0	0.0019	9.0	0.2311	4.0	0.0563	6.0	0.1468
BHPRM	3.0	0.0054	2.0	0.0206	13.5	0.3546	2.0	0.0289
ANHPRM	4.0	0.0628	13.0	0.4628	6.0	0.1074	5.0	0.1074
SHRBRSAT	5.0	0.1257	11.0	0.2674	5.0	0.0663	20.0	0.6626

Variable Name	CIQ: All,2×5 Rank	CIQ: All,2×5 *p*-Val	CIQ: R1,2×5 Rank	CIQ: R1,2×5 *p*-Val	CIQ: R2,2×5 Rank	CIQ: R2,2×5 *p*-Val	CIQ: R3,2x5 Rank	CIQ: R3,2x5 *p*-Val
	Common Interquartile (CIQ) for *E0:BRAALIC*							
ANHPRM	1.0	0.0000	17.0	0.7358	6.5	0.1468	1.0	0.0001
WMICDFLG	2.0	0.0000	2.0	0.0392	1.0	0.0321	2.0	0.0057
SHRGSSAT	3.0	0.0628	13.5	0.5918	5.0	0.1074	5.5	0.2674
WGRMICI	4.0	0.0780	5.0	0.2311	3.0	0.0663	7.0	0.3084
SHRBRSAT	5.0	0.1395	7.0	0.3084	20.0	0.8088	24.0	0.8781
	Common Interquartile (CIQ) for *E2:WAS_PRES*							
BHPRM	1.0	0.0000	1.0	0.0002	1.0	0.0021	2.0	0.0061
WGRCOR	2.0	0.0130	2.0	0.1074	7.0	0.2674	7.0	0.1468
SHRGSSAT	3.0	0.0289	11.0	0.4628	2.0	0.1074	11.0	0.3546
ANRBRSAT	4.0	0.0739	22.5	0.8781	14.5	0.5918	10.0	0.3084
SHRBRSAT	5.0	0.2093	18.5	0.7358	8.0	0.4060	1.0	0.0051

Variable Name	SI: All, 5×5 Rank	SI: All, 5×5 *p*-Val	SI: R1,5×5 Rank	SI: R1,5×5 *p*-Val	SI: R2,5×5 Rank	SI: R2,5×5 *p*-Val	SI: R3,5×5 Rank	SI: R3,5×5 *p*-Val
	Statistical Independence (SI) for *E0:WAS_PRES*							
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
HALPOR	2.0	0.0000	2.0	0.0034	2.0	0.0000	2.0	0.0000
WGRCOR	3.0	0.0003	13.5	0.4884	4.0	0.0316	9.5	0.2687
ANHPRM	4.0	0.0049	4.5	0.1785	8.5	0.2202	3.0	0.1010
ANHBCVGP	5.0	0.0194	3.0	0.1712	13.0	0.3546	18.0	0.7358
	Statistical Independence (SI) for *E2:WAS_SATB*							
WRGSSAT	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
BHPRM	2.0	0.0000	2.0	0.0000	2.0	0.0000	2.0	0.0001
ANHPRM	3.0	0.0002	3.0	0.0316	7.5	0.0415	7.0	0.2954
ANRBRSAT	4.0	0.0495	14.0	0.4530	5.0	0.0275	23.5	0.9134
WGRMICH	5.0	0.0564	11.0	0.3856	9.0	0.0895	8.5	0.3540

Variable Name	SI: All, 5×5 Rank	SI: All, 5×5 *p*-Val	SI: R1,5×5 Rank	SI: R1,5×5 *p*-Val	SI: R2,5×5 Rank	SI: R2,5×5 *p*-Val	SI: R3,5×5 Rank	SI: R3,5×5 *p*-Val
	Statistical Independence (SI) for *E0:BRAALIC*							
WMICDFLG	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
ANHPRM	2.0	0.0000	2.0	0.0003	2.0	0.0001	2.0	0.0000
HALPRM	3.0	0.0517	7.5	0.3540	21.5	0.7776	5.0	0.1137
HALPOR	4.0	0.0698	17.5	0.7089	8.0	0.2202	17.0	0.7440
SHRBRSAT	5.0	0.1917	23.0	0.8392	10.0	0.2687	3.0	0.0540
	Statistical Independence (SI) for *E2:WAS_PRES*							
BHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000	1.0	0.0000
HALPRM	2.0	0.0002	10.0	0.2954	3.0	0.1137	7.0	0.2202
WGRCOR	3.0	0.0002	2.0	0.0100	18.0	0.7089	2.0	0.0362
ANHPRM	4.0	0.0049	14.0	0.3856	9.0	0.3856	5.0	0.1785
SHRGSSAT	5.0	0.0698	17.5	0.5615	16.0	0.6359	11.0	0.3856

a Twenty-four (24) variables included in analysis for *E0:WAS_PRES* and *E0:BRAALIC* (see Footnote b to Table 10.1); twenty-five (25) variables included in analysis for *E2:WAS_SATB* and *E2:WAS_PRES* (see Footnote b to Table 10.14); for each test and dependent variable, top five variables based on their ordering with *p*-values obtained from pooled sample of size 300 are included in table.

b See Footnote c, Table 10.1.

Table 12.2 Comparison of Variable Rankings Obtained with Common Variances (CVs) Test with Use of Logarithms[a] for Three Independent Samples of Size 100 (Column CV:R1, CV:R2, CV:R3) and Pooled Sample of Size 300 (Column CV:All) for *y = E0:BRAALIC*[b]

Variable	CV: All,1×5		CV: R1,1×5		CV: R2,1×5		CV: R3,1×5	
Name	Rank	*p*-Val	Rank	*p*-Val	Rank	*p*-Val	Rank	*p*-Val
ANHPRM	1.0	0.0000	1.0	0.0000	1.0	0.0000	2.0	0.0000
WMICDFLG	2.0	0.0002	10.0	0.0251	7.0	0.0035	1.0	0.0000
SHPRMCON	3.0	0.0019	11.0	0.0257	5.0	0.0022	22.0	0.7184
SHBCEXP	4.0	0.0130	15.0	0.0528	19.0	0.2129	21.0	0.6442
WASTWICK	5.0	0.0144	13.0	0.0387	4.0	0.0002	17.0	0.3413

[a] See Footnote a, Table 10.7 for description of test.
[b] See Footnote a, Table 12.1.

the CVs test as the most important variables affecting *E2:WAS_PRES* for replicate R2, the *p*-value is high (i.e., 0.0633). The CIQ test misses the effect of *ANHPRM* on *E0:BRAALIC* for replicates R1 and R2, with this behavior probably resulting from the large number of zero and near-zero values associated with *E0:BRAALIC* (Fig. 2.4).

An important point that emerges from the results for the individual replicates is that consistency across independent analyses does not necessarily imply that these analysis are properly identifying the dominant variables with respect to the dependent variable under consideration. For example, all four analyses with both CCs and RCCs identify *HALPRM* and *ANHPRM* as being the two most important variables with respect to *E2:WAS_PRES* (Table 12.1) and completely fail to identify the dominant role played by *BHPRM* (Fig. 2.6). For *E2:WAS_PRES*, the three replicates are producing similar patterns, which in turn are producing similar outcomes when analyzed with CCs and RCCs.

Intentionally Left Blank

13. Discussion

Sensitivity analysis is an essential component of model development, assessment and application. Monte Carlo procedures are widely used in sensitivity studies to develop a mapping between uncertain model inputs and associated model results that can then be explored with regression-based techniques. Unfortunately, regression-based techniques sometimes fail to identify important patterns in the mapping (i.e., scatterplot) between model input and model results. The reason for this is simple; the relationships between model inputs and model results can be too complex to be identified by the linear relationships that most regression analyses are predicated on.

The likelihood of a successful sensitivity analysis can be increased by using a number of different procedures to identify relationships between model inputs and model results. With this strategy, a relationship that is missed by one procedure may be identified by another procedure. Fortunately, the post-processing of model results that is involved in the identification of patterns in scatterplots is relatively inexpensive from a computational perspective, and so the use of a number of different procedures does not present a significant burden.

In this presentation, a sequence of procedures for identifying patterns in scatterplots is described and illustrated. These procedures are based on attempts to recognize increasingly complex patterns in the scatterplots under consideration and involve the identification of (i) linear relationships with correlation coefficients, (ii) monotonic relationships with rank correlation coefficients, (iii) trends in measures of central tendency as defined by means, medians and the Kruskal-Wallis statistic, (iv) trends in measures of variability as defined by variances and interquantile ranges, and (v) deviations from randomness as defined by the chi-square statistic. As illustrated in a sequence of example analyses with a large model for two-phase fluid flow, the individual procedures can differ in the variables that they do, and do not, identify as having effects on particular analysis outcomes. The example results indicate that the use of a sequence of procedures is a good analysis strategy and provides some assurance that an important effect is not overlooked due to the use of an inappropriate analysis procedure.

The procedures under consideration identify patterns in scatterplots that in some sense appear to be nonrandom. However, they provide no explanation for why these patterns exist. Once such patterns are identified, it is the responsibility of the appropriate modelers and analysts to develop explanations for them. If such explanations cannot be developed, then the possibility exists that an error is present in the analysis. For this reason, well-designed sensitivity analyses provide both a way to develop insights with respect to the problem under consideration and also a way to check the conceptual and computational implementation of the problem.

Intentionally Left Blank

14. References

Andrews, D.F., P.J. Bickel, F.R. Hampel, P.J. Huber, W.H. Rogers, and J.W. Tukey. 1972. *Robust Estimates of Location: Survey and Advances.* Princeton, NJ: Princeton University Press; Ann Arbor, MI: Available from UMI, Books on Demand Program.

Archer, G.E.B., A. Saltelli, and I.M. Sobol. 1997. "Sensitivity Measures, ANOVA-like Techniques and the Use of Bootstrap," *Journal of Statistical Computation and Simulation.* **58**(2), 99-120.

Blower, S.M., and H. Dowlatabadi. 1994. "Sensitivity and Uncertainty Analysis of Complex Models of Disease Transmission: an HIV Model, As an Example," *International Statistical Review.* **62**(2), 229-243.

Breshears, D.D., T.B. Kirchner, and F.W. Whicker. 1992. "Contaminant Transport Through Agroecosystems: Assessing Relative Importance of Environmental, Physiological, and Management Factors," *Ecological Applications.* **2**(3), 285-297.

Chan, M.S. 1996. "The Consequences of Uncertainty for the Prediction of the Effects of Schistosomiasis Control Programmes," *Epidemiology and Infection.* **117**(3), 537-550.

Conover, W.J. 1980. *Practical Nonparametric Statistics.* 2nd ed. New York: John Wiley & Sons, Inc.

Conover, W.J., M.E. Johnson, and M.M. Johnson. 1981. "A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data," *Technometrics.* **23**(4), 351-361.

D'Agostino, R.B., and H.A. Stephens, eds. 1986. *Goodness-of-Fit Techniques.* New York, NY: Marcel Dekker, Inc.

David, H.A. 1970. *Order Statistics.* New York, NY: John Wiley & Sons.

Draper, N.R., and H. Smith. 1981. *Applied Regression Analysis.* 2nd ed. New York: John Wiley & Sons, Inc.

Efron, B. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans.* CBMS-NSF Regional Conference Series in Applied Mathematics 38. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Efron, B., and R.J. Tibshirani. 1993. *An Introduction to the Bootstrap.* Monographs on Statistics and Applied Probability 57. London; New York: Chapman & Hall.

Gwo, J.P., L.E. Toran, M.D. Morris, and G.V. Wilson. 1996. "Subsurface Stormflow Modeling with Sensitivity Analysis Using a Latin-Hypercube Sampling Technique," *Groundwater.* **34**(5), 811-818.

Hamby, D.M. 1994. "A Review of Techniques for Parameter Sensitivity Analysis of Environmental Models." *Environmental Monitoring and Assessment.* **32**(2), 135-154.

Hamby, D.M. 1995. "A Comparison of Sensitivity Analysis Techniques," *Health Physics.* **68**(2), 195-204.

Helton, J.C. 1993. "Uncertainty and Sensitivity Analysis Techniques for Use in Performance Assessment for Radioactive Waste Disposal," *Reliability Engineering and System Safety.* **42**(2-3), 327-367.

Helton, J.C. 1997. "Uncertainty and Sensitivity Analysis in the Presence of Stochastic and Subjective Uncertainty," *Journal of Statistical Computation and Simulation.* **57**(1-4), 3-76.

Helton, J.C., J.W. Garner, R.D. McCurley, and D.K. Rudeen. 1991. *Sensitivity Analysis Techniques and Results for Performance Assessment at the Waste Isolation Pilot Plant.* SAND90-7103. Albuquerque, NM: Sandia National Laboratories.

Helton, J.C., D.R. Anderson, B.L. Baker, J.E. Bean, J.W. Berglund, W. Beyeler, K. Economy, J.W. Garner, S.C. Hora, H.J. Iuzzolino, P. Knupp, M.G. Marietta, J. Rath, R.P. Rechard, P.J. Roache, D.K. Rudeen, K. Salari, J.D. Schreiber, P.N. Swift, M.S. Tierney, and P. Vaughn. 1996. "Uncertainty and Sensitivity Analysis Results Obtained in the 1992 Performance Assessment for the Waste Isolation Pilot Plant," *Reliability Engineering and System Safety.* **51**(1), 53-100.

Helton, J.C., J.E. Bean, J.W. Berglund, F.J. Davis, K. Economy, J.W. Garner, J.D. Johnson, R.J. MacKinnon, J. Miller, D.G. O'Brien, J.L. Ramsey, J.D. Schreiber, A. Shinta, L.N. Smith, D.M. Stoelzel, C. Stockman, and P. Vaughn. 1998. *Uncertainty and Sensitivity Analysis Results Obtained in the 1996 Performance Assessment for the Waste Isolation Pilot Plant.* SAND98-0365. Albuquerque, NM: Sandia National Laboratories.

IAEA (International Atomic Energy Agency). 1989. *Evaluating the Reliability of Predictions Made Using Environmental Transfer Models.* Safety Series Report No. 100. Vienna, Austria: International Atomic Energy Agency. Sandia WIPP Central Files WPO#49330.

Iman, R.L. 1987. "Tables of the Exact Quantiles of the Top-Down Correlation Coefficient for n = 3(1)14," *Communications in Statistics—Theory and Methods.* **16**(5), 1513-1540.

Iman, R.L., and W.J. Conover. 1979. "The Use of the Rank Transform in Regression," *Technometrics.* **21**(4), 499-509.

Iman, R.L., and W.J. Conover. 1982. "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables," *Communications in Statistics: Part B. Simulation and Computation.* **B11**(3), 311-334.

Iman, R.L., and W.J. Conover. 1987. "A Measure of Top-Down Correlation," *Technometrics*. **29**(3), 351-357.

Iman, R.L., and J.C. Helton. 1988. "An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models," *Risk Analysis*. **8**(1), 71-90.

Iman, R.L., and M.J. Shortencarier. 1984. *A FORTRAN 77 Program and User's Guide for the Generation of Latin Hypercube and Random Samples for Use With Computer Models.* NUREG/CR-3624, SAND83-2365. Albuquerque, NM: Sandia National Laboratories.

Iman, R.L., J.M. Davenport, E.L. Frost, and M.J. Shortencarier. 1980. *Stepwise Regression with PRESS and Rank Regression (Program and User's Guide).* SAND79-1472. Albuquerque, NM: Sandia National Laboratories.

Iman, R.L., J.C. Helton, and J.E. Campbell. 1981a. "An Approach to Sensitivity Analysis of Computer Models: Part I. Introduction, Input Variable Selection and Preliminary Variable Assessment," *Journal of Quality Technology*. **13**(3), 174-183.

Iman, R.L., J.C. Helton, and J.E. Campbell. 1981b. "An Approach to Sensitivity Analysis of Computer Models: Part II. Ranking of Input Variables, Response Surface Validation, Distribution Effect and Technique Synopsis," *Journal of Quality Technology*. **13**(4), 232-240.

Iman, R.L., M.J. Shortencarier, and J.D. Johnson. 1985. *A FORTRAN 77 Program and User's Guide for the Calculation of Partial Correlation and Standardized Regression Coefficients*. SAND85-0044; NUREG/CR-4122, Albuquerque, NM: Sandia National Laboratories.

Kleijnen, J.P.C. 1987. *Statistical Tools for Simulation Practitioners*. Statistics, Textbooks and Monographs **76**. New York: Marcel Dekker.

Kleijnen, J.P.C. 1998. "Chapter 6 Experimental Design for Sensitivity Analysis, Optimization, and Validation of Simulation Models," *Handbook of Simulation: Principles, Methodology, Advances, Application, and Practice*. Ed. J. Banks. New York, NY: John Wiley & Sons. 173-223.

Ma, J.Z., and E. Ackerman. 1993. "Parameter Sensitivity of a Model of Viral Epidemics Simulated with Monte Carlo Techniques. II. Durations and Peaks," *International Journal of Biomedical Computing*. **32**(3-4), 255-268.

Ma, J.Z., E. Ackerman, and J.-J. Yang. 1993. "Parameter Sensitivity of a Model of Viral Epidemics Simulated with Monte Carlo Techniques. I. Illness Attack Rates," *International Journal of Biomedical Computing*. **32**(3-4), 237-253.

MacDonald, R.C., and J.E. Campbell. 1986. "Valuation of Supplemental and Enhanced Oil Recovery Projects with Risk Analysis," *Journal of Petroleum Technology*. **38**(1), 57-69.

McKay M.D., W.J. Conover, and R.J. Beckman. 1979. "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*. **21**(2), 239-245.

Miller, R.G. 1974. "The Jackknife - A Review," *Biometrika*. **61**, 1-15.

Myers, R.H. 1990. *Classical and Modern Regression with Applications*. Boston, MA: Duxbury Press.

NCRP (National Council on Radiation Protection and Measurements). 1996. *A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination*. NCRP Commentary No. 14. Bethesda, MD: National Council on Radiation Protection and Measurements.

NRC (National Research Council). 1983. *Risk Assessment in the Federal Government: Managing the Process*. Committee on the Institutional Means for Assessment of Risks to Public Health, Commission on Life Sciences. Washington, DC: National Academy Press.

NRC (National Research Council). 1993. *Issues in Risk Assessment*. Committee on Risk Assessment Methodology, Board on Environmental Studies and Toxicology, Commission on Life Sciences. Washington, DC: National Academy Press.

NRC (National Research Council). 1994. *Science and Judgment in Risk Assessment*. Committee on Risk Assessment of Hazardous Air Pollutants. Washington, DC: National Academy Press.

Piepho, H.-P. 1997. "Tests for Equality of Dispersion in Bivariate Samples - Review and Empirical Comparison," *Journal of Statistical Computation and Simulation*. **56**(4), 353-372.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. 1992. *Numerical Recipes in Fortran: The Art of Scientific Computing*. 2nd ed. Cambridge, England; New York: Cambridge University Press.

Saltelli, A., and T. Homma. 1992. "Sensitivity Analysis for Model Output. Performance of Black Box Techniques on Three International Benchmark Exercises," *Computational Statistics and Data Analysis*. **13**, 73-94.

Saltelli, A., and J. Marivoet. 1990. "Non-Parametric Statistics in Sensitivity Analysis for Model Output. A Comparison of Selected Techniques," *Reliability Engineering and System Safety*. **28**(2), 229-253.

Saltelli, A., T.H. Andres, and T. Homma. 1993. "Sensitivity Analysis of Model Output. An Investigation of New Techniques," *Computational Statistics and Data Analysis*. **15**(2), 211-238.

Sanchez, M.A., and S.M. Blower. 1997. "Uncertainty and Sensitivity Analysis of the Basic Reproductive Rate. Tuberculosis as an Example," *American Journal of Epidemiology*. **145**(12), 1127-1137.

Sargent, R.G. 1996. "Some Subjective Validation Methods Using Graphical Displays of Data," *1996 Winter Simulation Conference Proceedings, Coronado, CA, December 8-11, 1996.* Eds. J.M. Charnes, D.J. Morrice, D.T. Brunner, and J.J. Swain. New York, NY: Association for Computing Machinery; Piscataway, NJ: The Institute of Electrical and Electronics Engineers. 345-351.

Scheffé, H. 1959. *The Analysis of Variance.* New York, NY: John Wiley & Sons, Inc.

U.S. DOE (U.S. Department of Energy). 1996. *Title 40 CFR Part 191 Compliance Certification Application for the Waste Isolation Pilot Plant.* DOE/CAO-1996-2184: Carlsbad, NM: U.S. Department of Energy, Waste Isolation Pilot Plant, Carlsbad Area Office.

U.S. EPA (Environmental Protection Agency). 1993. *An SAB Report: Multi-Media Risk Assessment for Radon. Review of Uncertainty Analysis of Risks Associated with Exposure to Radon.* EPA-SAB-RAC-93-014. Washington, DC: Environmental Protection Agency, Science Advisory Board.

U.S. EPA (Environmental Protection Agency). 1996. "40 CFR Part 194: Criteria for the Certification and Re-Certification of the Waste Isolation Pilot Plant's Compliance with the 40 CFR Part 191 Disposal Regulations; Final Rule," *Federal Register.* **61**(28), 5224-5245.

Whiting, W.B., T.-M. Tong, and M.E. Reed. 1993. "Effect of Uncertainties in Thermodynamic Data and Model Parameters on Calculated Process Performance," *Industrial and Engineering Chemistry Research.* **32**(7), 1367-1371.

Wludyka, P.S., and P.R. Nelson. 1997. "Analysis of Means Type Tests for Variances Using Subsampling and Jackknifing," *American Journal of Mathematical and Management Sciences.* **17**(1-2), 31-60.

Intentionally Left Blank

WIPP
UC721 - DISTRIBUTION LIST
SAND98-2202

Federal Agencies

US Department of Energy (4)
Office of Civilian Radioactive Waste Mgmt.
Attn: Deputy Director, RW-2
Acting Director, RW-10
Office of Human Resources & Admin.
Director, RW-30
Office of Program Mgmt. & Integ.
Director, RW-40
Office of Waste Accept., Stor., & Tran.
Forrestal Building
Washington, DC 20585

Yucca Mountain Site Characterization Office
Director, RW-3
Office of Quality Assurance
Attn: Project Director
P. O. Box 30307
Las Vegas, NV 89036-0307

US Department of Energy
Research & Waste Management Division
Attn: Director
P.O. Box E
Oak Ridge, TN 37831

US Department of Energy (5)
Carlsbad Area Office
Attn: K. Klein
G. T. Basabilvazo
D. Galbraith
M. McFadden
J. A. Mewhinney
P.O. Box 3090
Carlsbad, NM 88221-3090

US Department of Energy
Office of Environmental Restoration and
Waste Management
Attn: M. Frei, EM-30
Forrestal Building
Washington, DC 20585-0002

US Department of Energy (3)
Office of Environmental Restoration and
Waste Management
Attn: J. Juri, EM-34, Trevion II
Washington, DC 20585-0002

US Department of Energy
Office of Environmental Restoration and
Waste Management
Attn: S. Schneider, EM-342, Trevion II
Washington, DC 20585-0002

US Department of Energy (2)
Office of Environment, Safety & Health
Attn: C. Borgstrom, EH-25
R. Pelletier, EH-231
Washington, DC 20585

US Department of Energy (2)
Idaho Operations Office
Fuel Processing & Waste Mgmt. Division
785 DOE Place
Idaho Falls, ID 83402

US Environmental Protection Agency (2)
Radiation Protection Programs
Attn: M. Oge
ANR-460
Washington, DC 20460

Timothy M. Barry
Chief, Science – Policy, Planning, and
Evaluation
PM 223X U.S. EPA
Washington, DC 20460

Norman A. Eisenberg
U.S. Nuclear Regulatory Commission
Office of Nuclear Material Safety and
Safeguards
Division of Waste Management
Washington, D. C. 20555

Boards

Defense Nuclear Facilities Safety Board
Attn: D. Winters
625 Indiana Ave. NW, Suite 700
Washington, DC 20004

Nuclear Waste Technical Review Board (2)
Attn: Chairman
J.L. Cohon
2300 Clarendon Ste. 1300
Arlington, VA 22203-3367

State Agencies

Attorney General of New Mexico
P.O. Drawer 1508
Santa Fe, NM 87504-1508

Environmental Evaluation Group (3)
Attn: Library
7007 Wyoming NE
Suite F-2
Albuquerque, NM 87109

NM Environment Department (3)
Secretary of the Environment
1190 St. Francis Drive
Santa Fe, NM 87503-0968

NM Bureau of Mines & Mineral Resources
Socorro, NM 87801

Laboratories/Corporations

Battelle Pacific Northwest Laboratories
Battelle Blvd.
Richland, WA 99352

Dr. Pamela Doctor
Battelle Northwest
P.O. Box 999
Richland, WA 99352

Los Alamos National Laboratory
Attn: B. Erdal, INC-12
P.O. Box 1663
Los Alamos, NM 87544

M. D. McKay
F600
Los Alamos National Laboratory
Los Alamos, NM 87545

Max Henrion
The Decision Laboratory (IDSR)
4984 El Camino Real, Suite 105
Los Altos, CA 94022

Tech Reps, Inc. (3)
Attn: J. Chapman (1)
Loretta Robledo (2)
5000 Marble NE, Suite 222
Albuquerque, NM 87110

Westinghouse Electric Corporation (5)
Attn: Library
J. Epstein
J. Lee
B. A. Howard
R. Kehrman
P.O. Box 2078
Carlsbad, NM 88221

S. Cohen & Associates
Attn: Bill Thurber
1355 Beverly Road
McLean, VA 22101

B. John Garrick
PLG Incorporated
4590 MacArthur Blvd., Suite 400
Newport Beach, CA 92660-2027

Christopher G. Whipple
ICF Kaiser Engineers
1800 Harrison St., 7th Floor
Oakland, CA 94612-3430

Dr. Kenneth T. Bogen
LLNL/ENV SCI DIVL – 453
P. O. Box 808
Livermore, CA 94550

David E. Burmaster
ALCEON Corporation
P. O. Box 382669
Harvard Square Station
Cambridge, MA 02238-2669

Scott Ferson
Applied Biomathematics
100 North Country Road
Setauket, NY 11733

Dr. Robert J. Budnitz
Future Resources Associates
2039 Shattuck Avenue, Suite 402
Berkeley, CA 94704

Dr. Tom Cotton
JK Research Associates
2650 Park Tower Drive, Suite 800
Vienna, VA 22180

Dr. John Kessler
Electronic Power Research Institute
3412 Hillview Avenue
Palo Alto, CA 94304-1395

D. Warner North
Decision Focus Incorporated
650 Castro Street, Suite 300
Mountain View, CA 94041-2055

S. David Sevougian
Duke Engineering and Services
CRWMS M&O
1180 Town Center Drive
Las Vegas, NV 89134

Michael B. Gross
Michael Gross Enterprises
21 Tradewind Passage
Corte Madera, CA 94925

Beta Corporation Int.
Attn: E. Bonano
6613 Esther NE
Albuquerque, NM 87109

Center for Nuclear Waste Regulatory Analysis
(CNWRA)
Southwest Research Institute
Attn: B. Sagar
P. O. Drawer 28510
620 Culebra Road
San Antonio, TX 78284

Banda S. Ramarao
Duke Engineering and Services
9111 Research Blvd.
Austin, TX 78758

Senes Oak Ridge, Inc (2)
Center for Risk Analysis
Attn: Steve Bartell
F. Owen Hoffman
102 Donner Drive
Oak Ridge, TN 37810

National Academy of Sciences, WIPP Panel

Tom Kiess (15)
Staff Study Director
GF456
2101 Constitution Ave.
Washington, DC 20418

Universities

University of New Mexico
Geology Department
Attn: Library
141 Northrop Hall
Albuquerque, NM 87131

University of Washington
College of Ocean & Fishery Sciences
Attn: G. R. Heath
583 Henderson Hall, HN-15
Seattle, WA 98195

Prof. G. E. Apostolakis
Department of Nuclear Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139-4307

Prof. V. M. Bier
Department of Industrial Engineering
University of Wisconsin
Madison, WI 53706

Prof. M. Elisabeth Pate-Cornell
Department of Industrial Engineering and Management
Stanford University
Stanford, CA 94305

Prof. C. Frey
Department of Civil Engineering
Box 7908/NCSU
Raleigh, NC 27659-7908

Prof. Yacov Y. Haimes
Center for Risk Management of Engineering Systems
D111 Thornton Hall
University of Virginia
Charlottesville, VA 22901

Prof. D. B. Hattis
CENTED
Clark University
950 Main Street
Worcester, MA 01610

Prof. Ali Mosleh
Center for Reliability Engineering
University of Maryland
College Park, MD 20714-2115

Prof. T. G. Theofanous
Department of Chemical and Nuclear Engineering
University of California
Santa Barbara, CA 93106

Prof. Steve Hora
Institute of Business and Economic Studies
University of Hawaii, Hilo
523 W. Lanikaula
Hilo, HI 96720-4091

Prof. Thomas E. McKone
School of Public Health
University of California
Berkeley, CA 94270-7360

Prof. Herschel Rabitz
Princeton University
Department of Chemistry
Princeton, NJ 08544

Prof. Robert L. Winkler
Fuqua School of Business
Duke University
Durham, NC 27708-0120

F. E. Haskin
Department of Chemical and Nuclear Engineering
University of New Mexico
Albuquerque, NM 87131

Alan Gutjahr
Department of Mathematics
New Mexico Institute of Mining and Tech.
Socorro, NM 87801

Thomas H. Pigford
Department of Nuclear Engineering
4159 Etcheverry Hall
University of California
Berkeley, CA 94720

C. John Mann
Department of Geology
245 Natural History Bldg.
1301 West Green Street
University of Illinois
Urbana, IL 61801

Frank W. Schwartz
Department of Geology and Mineralogy
Ohio State University
Scott Hall
1090 Carmack Rd.
Columbus, OH 43210

David M. Hamby
University of Michigan
109 Observatory St.
Ann Arbor, MI 48109-2029

Rodney C. Ewing
Nuclear Engineering and Radiological Science
University of Michigan
Ann Arbor, MI 48109-2104

David Okrent
Mechanical and Aerospace Engineering Department
University of California
48-121 Engineering IV Building
Los Angeles, CA 90095-1587

Libraries

Thomas Brannigan Library
Attn: D. Dresp
106 W. Hadley St.
Las Cruces, NM 88001

Government Publications Department
Zimmerman Library
University of New Mexico
Albuquerque, NM 87131

New Mexico Junior College
Pannell Library
Attn: R. Hill
Lovington Highway
Hobbs, NM 88240

New Mexico State Library
Attn: N. McCallan
325 Don Gaspar
Santa Fe, NM 87503

New Mexico Tech
Martin Speere Memorial Library
Campus Street
Socorro, NM 87810

WIPP Public Reading Room
Carlsbad Public Library
101 S. Halagueno St.
Carlsbad, NM 88220

Foreign Addresses

Atomic Energy of Canada, Ltd. (2)
Whiteshell Laboratories
Attn: B. Goodwin
T. Andres
Pinawa, Manitoba, CANADA R0E 1L0

Dr. Arnold Bonne
Acting Head of the Waste Technology Section
Division of Nuclear Fuel Cycle and Waste Management
International Atomic Energy Agency
P.O. Box 100
A-1400 Vienna
AUSTRIA

Clandio Pescatori
AERI/NEA/OECD
LeSeine St. Germain
12 Boulevard des iles
92130 Issy-les-Moulineaux
FRANCE

Francois Chenevier (2)
ANDRA
Parc de la Croix Blanche
1-7 rue Jean Monnet
92298 Chatenay-Malabry Cedex
FRANCE

Claude Sombret
Centre d'Etudes Nucleaires de la Vallee Rhone
CEN/VALRHO
S.D.H.A. B.P. 171
30205 Bagnols-Sur-Ceze
FRANCE

Commissariat a L'Energie Atomique
Attn: D. Alexandre
Centre d'Etudes de Cadarache
13108 Saint Paul Lez Durance Cedex
FRANCE

Ghislain de Marsily
University Pierre et Marie Curie
Laboratorie de Geologie Applique
4, Place Jussieu
T.26 – 5e etage
75252 Paris Cedex 05
FRANCE

Bundesanstalt fur Geowissenschaften und Rohstoffe
Attn: M. Langer
Postfach 510 153
D-30631 Hannover
GERMANY

Bundesministerium fur Forschung und Technologie
Postfach 200 706
5300 Bonn 2
GERMANY

Gesellschaft fur Anlagen und Reaktorsicherheit (GRS)
Attn: B. Baltes
Schwertnergasse 1
D-50667 Cologne
GERMANY

Forschunginstitute
GRS (2)
Attn: Eduard Hofer
B. Krzykacz-Hausmann
Forschungsgelande Nebau 2
D-85748 Garching
GERMANY

Tamas Turanyi
Eotvos University (ELTE)
Po.O. Box 32
H –1518 Budapest
HUNGARY

Jan Marivoet
Centre D'Etudes de l'Energie
Nucleaire
Boeretang 200
B-2400 MOL
BELGIUM

Andrea Saltelli
European Commission
Via Fermi
21020 Ispra (VA)
ITALY

Prof. I. Papazoglou
Institie of Nuclear Technology-Radiation
Protection
N.C.S.R. Demokritos
Aghia Papakevi
153-10 Athens
GREECE

European Commission (3)
Attn: Francesca Campolongo
Karen Chan
Stefano Tarantola
JRC Ispra, ISIS
I – 21020 Ispra
ITALY

Enrico Zio
Politecnico di Milano
Via Ponzio 34/3
I –20133 Milan
ITALY

Ricardo Bolado
Polytechnical University of Madrid
Jose Gutierrez Abascal, 2
E- 28006 Madrid
SPAIN

David Rios Insua
University Rey Juan Carlos
ESCET-URJC, C. Humanes 63
E –28936 Mostoles
SPAIN

Shingo Tashiro
Japan Atomic Energy Research Institute
Tokai-Mura, Ibaraki-Ken, 319-11
JAPAN

Toshimitsu Homma
Nuclear Power Engineering Corporation
3-17-1 Toranomon, Minato-Ku
Tokyo 1015
JAPAN

Netherlands Energy Research Foundation ECN
Attn: J. Prij
3 Westerduinweg
P.O. Box 1
1755 ZG Petten
THE NETHERLANDS

Prof. Roger Cooke
Dept. of Mathematics
Delft University of Technology
P.O. Box 5031 2800 GA Delft
THE NETHERLANDS

Louis Goossens
Safety Science Group
Delft Univ. of Technology
Kanaalweg 2b
2628 EB Delft
THE NETHERLANDS

Prof. J.P.C. Kleijnen
Department of Information Systems
Tilburg University
5000 LE Tilburg
THE NETHERLANDS

A. Seebregts
ECN PO Box 1
1755 ZG Petten
THE NETHERLANDS

Willem Van Groenendaal
Tilburg University
P.O. Box 90153
NL – 5000 Le Tilburg
THE NETHERLANDS

Svensk Karnbransleforsorjning AB
Attn: F. Karlsson
Project KBS (Karnbranslesakerhet)
Box 5864
S-102 48 Stockholm
SWEDEN

Prof. S. E. Magnusson
Lund University
P.O. Box 118
22100 Lund
SWEDEN

Prof. Christian Ekberg
Chalmers University of Technology
Dept. of Nuclear Chemistry
S-41296 Goteborg
SWEDEN

Nationale Genossenschaft fur die Lagerung
Radioaktiver Abfalle (2)
Attn: S. Vomvoris
P. Zuidema
Hardstrasse 73
CH-5430 Wettingen
SWITZERLAND

AEA Technology
Attn: J. H. Rees
D5W/29 Culham Laboratory
Abington, Oxfordshire OX14 3DB
UNITED KINGDOM

AEA Technology
Attn: W. R. Rodwell
044/A31 Winfrith Technical Centre
Dorchester, Dorset DT2 8DH
UNITED KINGDOM

Daniel A. Galson
Galson Science Ltd.
35, Market Place
Oakham
Leicestershire LE15 6DT
UNITED KINGDOM

David Draper
University of Bath
Claverton Down
Bath BA2 7AY
UNITED KINGDOM

AEA Technology
Attn: J. E. Tinson
B4244 Harwell Laboratory
Didcot, Oxfordshire OX11 ORA
UNITED KINGDOM

Prof. Marian Scott
Dept. of Statistics
University of Glasgow
Glasgow G12 BQW
UNITED KINGDOM

Prof. Simon French
School of Informatics
University of Manchester
Coupland 1
Manchester M13 9pl
UNITED KINGDOM

Arthur Jones
Nat. Radio. Prot. Board
Chilton, Didcot
Oxon OX110RQ
UNITED KINGDOM

Prof. M. Newby
Dept. of Acturial Sci & Statistics
City University
Northhampton SQ
London EC1V OHB
UNITED KINGDOM

Prof. Russell Cheng
University of Kent at Canterbury
Cornwallis Building
Canterbury, Kent, CT2 7NF
UNITED KINGDOM

B.G.J. Thompson
20 Bonser Road
Twickenham
Middlesex, TW1 4RG
ENGLAND

Other

Leonard F. Konikow
US Geological Survey
431 National Center
Reston, VA 22092

Dr. Bob Andrews
1280 Town Center Dr.
Las Vegas, NV 89314

Dr. Alison Cullen
2125 North 90th Street
Seattle, WA 98103

P.S. Price
129 Oakhurst Road
Cape Elizabeth, ME 04107

Dr. Gareth Parry
19805 Bodmer Ave
Poolesville, MD 200837

Dennis Powers
Star Route Box 87
Anthony, TX 79821

Internal

MS	Org.	
0735	6115	P. B. Davies
0737	6831	E. J. Nowak
0737	6833	J. R. Tillerson
0737	6833	E. H. Ahrens
0779	6849	D. R. Anderson
0779	6848	H. N. Jow
0771	6800	M. Chu
0733	6832	J. T. Holmes
1395	6821	M. Marietta
0771	6000	W. D. Weart
0771	6800	F. W. Bingham
0773	6821	G. K. Froehlich
1395	6821	F. D. Hansen
0776	6850	S. M. Howarth
0779	6848	K. W. Larson
0778	6850	C. D. Leigh
0773	6851	P. N. Swift
0734	6803	L. D. Bustard
1395	6860	R. Waters
0720	6804	K. B. Sorenson
0715	6805	C. E. Olson
0715	6805	R. L. Hunter
0715	6805	P. Kaplan
1395	6811	S. Y. Pickering
1395	6821	B. L. Baker
1395	6821	M. K. Knowles
1395	6821	M-A. Martell
1399	6850	P. E. Sanchez
1395	6821	Y. Wang
0733	6832	S. G. Bertram
0733	6832	L. H. Brush
0733	6832	R. V. Guzowski
0733	6832	H. W. Papenguth
0733	6832	A. C. Peterson
0733	6832	M. D. Siegel
0778	6851	K. M. Trauth
0779	6821	K. Aragon
0779	6848	J. E. Bean
0779	6848	F. J. Davis
0779	6848	L. J. Dotson
0779	6848	K. M. Economy
0779	6848	J. N. Emery
0779	6848	J. C. Helton (10)
0779	6848	J. D. Johnson
0779	6848	R. D. McCurley
0779	6848	J. D. Miller
0779	6848	D. O'Brien
0779	6848	H. C. Ogden
0779	6848	J. A. Rollstin
0733	6832	L. C. Sanchez
0779	6848	J. D. Schreiber
0779	6848	M. J. Shortencarier
0779	6848	L. N. Smith
1395	6848	M. S. Tierney
0779	6848	M. Williamson
0779	6848	T. Zimmerman
0779	6849	J. W. Berglund
0779	6849	P. J. Chen
0779	6849	M. E. Fewell
0779	6849	G. A. Freeze
0779	6849	J. W. Garner
0733	6832	L. S. Gomez
0779	6821	T. Hadgu
0779	6849	H. J. Iuzzolino
0779	6849	M. Lavenue
0779	6849	M. E. Lord
0779	6849	L. J. Rahal
0779	6849	J. L. Ramsey
0779	6849	R. P. Rechard
0779	6849	D. Rudeen
0779	6849	A. Schenker
0779	6849	A. H. Treadway
0779	6849	P. Vaughn
0779	6849	M. Wallace
0779	6849	J. A. Jones
0779	6849	J. H. Saloio
1399	6850	A. Orrell
0778	6851	H. A. Dockery
0778	6851	G. E. Barr
0778	6851	R. Aguilar
0778	6851	J. H.Gauthier
0778	6851	R. J. MacKinnon
0778	6851	M. L. Wilson
0778	6851	C. T. Stockman
1399	6853	J. J. Danneels
0735	6115	A. R. Lappin
0735	6115	R. L. Beauheim
0735	6115	T. F. Corbet
0735	6115	B. Holt
0735	6115	L. Meigs
1044	7511	A. Reiser
0731	6811	K. Hart (2)

0731	4415	NWM Library (10)
9018	8940-2	Central Technical Files
0899	4916	Technical Library (2)
0619	15102	Review and Approval Desk, For DOE/OSTI