Modeling and Graphing Organizational Processes in Pursuit of Performance Benchmarks: Methods for Establishing and Evaluating Performance Measures

June 13 1997

William G. Saylor
Deputy Director
Office of Research and Evaluation
Federal Bureau of Prisons

Introduction

Contemporary management philosophies over the last several decades have progressively moved toward statistical process control methods. These methods, also referred to as continuous process improvement (CPI) or continuous quality improvement, are frequently considered to be one of the governing principles of the management strategies referred to as total quality management (TQM). CPI involves the use of statistical tools to convert data into meaningful information. The methods are based on the presumption that measuring (observing and quantifying) phenomena allows for systematic analyses that can lead to insights and a better understanding of a process. A better understanding provides opportunities for creating more control over a process, resulting in a reduction in unwanted process variance and consequently an improvement in the process and the product it produces. (A good introduction to the concepts, principles and technical methods of TQM and CPI can be found in Kiemele and Schmidt, 1993.)

Although CPI was initially more prevalent in engineering and industrial environments it is now common to find these methods employed in many types of organizations, including service oriented organizations such as the Federal Government. The Clinton administration has implemented two related initiatives, the Government Performance and Results Act (GPRA) and the National Performance Review (NPR), that have mandated Federal agencies to establish measurable performance standards and procedures for implementing continuous process improvement. Since 1988, the Bureau of Prisons' Office of Research and Evaluation has provided a tool to assist management with monitoring the organization's performance. The tool, the Key Indicators/Strategic Support System (KI/SSS, Saylor, 1988, 1994), is a PC based, menu driven executive information system designed to provide management with on-demand access to a synthesis of the operational data they require to draw conclusions and make sound decisions. The KI/SSS provides managers with a continuous flow of characterizations and measures of organizational behavior and performance about the Bureau of Prisons.

The purpose of this paper is to use the data from the KI/SSS, in the context of TQM and CPI philosophies, to demonstrate how these methods can provide correctional management with operational definitions of the concepts of measurement thresholds and operational benchmarks. Furthermore, statistical and graphic methods developed to assist in diagnostic assessments of the fit of regression models will be used to define the notion of a process outlier and provide management with an easily accessible and digestible means with which to observe organizational units (e.g., an office, institution, region or any other meaningful collective) that are adding unwanted variance to a process's outcome.

We start by discussing the epistemology of organizational processes, as this determines the validity of any measures and models, and the limits of measurement variance that are deemed acceptable in the modeling of a process. We then discuss the role of theory and the specification of models in achieving parsimony in the measurement set of interest to management. Finally, using models of two organizational processes (related to inmate per capita costs and staff perceptions of institutional crowding) we will illustrate how statistical methods developed to analyze regression residuals (e.g., leverages, outliers, and influence measures) can provide a useful and intuitive means by which management can establish, maintain and monitor organizational performance.

Measurement, and the Role of Theory and Model Specification

Observation plays an essential role in any relationship between theory and phenomena, whether it is in the testing of theory, or in the prediction or explanation of phenomena. Observation usually takes place via measurement. Measurement is an indirect process which employs scales to determine the physical quantities of phenomenal elements. Furthermore, the relevance of information observed is not absolute. Scientific research is undertaken with respect to a specific scientific domain. It is not sufficient to simply have a number of related elements of information (measures), "it must be the case that the bodies of information, related in certain ways, raise problems or pose questions of various sorts, where these problems or questions are considered significant and as such that answers to some of these questions reasonably can be expected..." (Suppe, 1974).

"Theory does not attempt to characterize phenomenal systems in their full complexity; rather, the theory attempts to characterize them in terms of a few parameters abstracted from the phenomenal systems; these abstracted parameters may be idealized in various ways...A theory characterizes what the behavior of phenomenal systems within its intended scope would have been were it the case that only the defining parameters of the theory exerted an influence on its behavior and various idealized conditions were met." (Suppe, 1974). Theory formulation requires an underlying world view to direct the scientist's attention. There is no science without some expectation as to the nature of its subject matter (Laszlo, 1974b). Theory or expectations derived from practical experience provide the means by which to simplify reality in the context of a model, that is, reduce the number of measures to those that are relevant to a specific domain or phenomenal system.

In applying a theory to a phenomenal system the research design determines the relationship between the phenomenal system and the observational system. The phenomenal system determines the limits within which observation can take place and the theory determines how the observation is seen. That is, nature supplies the limit of perception and science provides the interpretation. The relationship between the phenomenal system and the observational system is referred to as the epistemic correlation, or alternatively, the rules of correspondence (Laszlo, 1972b: Margenau, 1950). (Costner (1969) used the term "rules of correspondence" in a related manner to describe the relationship between multiple indicators and a construct or latent variable as a means of contending with measurement error in causal models.)

The issues of epistemology and the rules of correspondence may seem so fundamental and obvious that they need no introduction. Nevertheless, I have provided this brief discussion because, from a philosophy of science perspective, there is nothing cognitive "which would stand for, or point to, physical events with absolute certainty" (Laszlo, 1972a). Although this point may appear too philosophical, it is a point that was also made by one of the quintessential applied scientists. Einstein(1934) expressed the view that "the belief in an external world independent of the perceiving subject is the basis of all natural science" and that one can only grasp this physical reality by speculative means.

The KI/SSS provides a voluminous reservoir of easily accessible measures that can be overwhelming. The KI/SSS contains thousands of fundamental measurement elements, some of which might be directly useful in models, and others that might be used to create measures of interest. If we are to proceed in a systematic, efficient and productive manner then our methods for identifying thresholds and benchmarks (and ultimately institutions that are performing counter to expectations) must be based on parsimonious sets of measures that have been selected for their relevance to specific organizational processes. Decades of experience with SPC and TQM have demonstrated that sufficient knowledge of a process and the use of statistical models can produce an effective method for monitoring organizational performance.

The Illustration

The KI/SSS design is based on the concept of organizational climates. It contains two varieties of climate measures, objective and subjective (Saylor, 1983). The objective measures are a by-product of the organization's operational data needs. These measures are meaningful summaries of individual inmate, staff, and financial units that are culled from the Bureau's operational mainframe MIS with a monthly periodicity. By meaningful summaries I mean summaries that relate to the mission of the Bureau and to the goals and objectives of management. The subjective measures are obtained from administrations of the Prison Social Climate Survey (Saylor, 1983); a survey questionnaire administered to a stratified proportional probability sample of Bureau staff on an annual basis and to inmates on an ad hoc basis.

One of the modules within the KI/SSS is composed of a set of measures selected by the Bureau's executive staff, and is referred to as the executive staff management indicators module (ESMI). These are the measures which this board of directors view as their key (or leading) indicators. Two segments of the module are listed in Figures 1 and 2. In its initial incarnation, the ESMI existed in a paper medium. After the first use of the document it was converted to the electronic document that currently exists in KI/SSS. The KI/SSS version displayed in Figures 1 and 2 has highlighted fields which signify data elements that are beyond the absolute or relative threshold values established by the executive staff. There is also hyper-text associated with the highlighted fields which allows the users to gain additional information about the nature of an extreme value for a particular data element. The ESMI is updated monthly, as is the remainder of the KI/SSS, and is distributed monthly on CD-ROM to all institutions, regions, and most central offices. The ESMI is also used in conjunction with an annual institutional review each spring.

The ESMI measures are oriented around six Bureau goals. Presently, the executive staff have not formally specified any relationships among these measures, at least as they are depicted in the ESMI. In the context of the ESMI, they are essentially six bins of numbers. Although some of the measures are useful in a univariate or contextual sense, wherein relationships among measures are loosely or informally specified, a formal specification of the relationships invokes a collective endorsement and broadens the utility of the measures. Formal specifications of relationships among the measures should be based on the executive staff's expectations derived from theories or operational experiences.

The primary purpose for the research reported here is to explore alternative display and threshold formats that could facilitate, expedite and minimize the burden of institutional performance reviews (based on measures like those contained in the ESMI module) by the Bureau's executive staff. One possibility would be to employ graphic screens that look like meters or instrument panels, similar to those used in cars or in airplanes. This format would provide the benefit of viewing an institution at a time, with respect to all the indicators of interest, which is the principal interest in the current text format display. The graphic format,

EXECUTIVE STAFF MODULE

Medium Security Level. . .

JANUARY 1995

GOAL #1: POPULATION MANAGEMENT

CONTEXT

INMATE DAILY PER CAPITA COST(adm)	Institution	Same Sec. Level	Avg. %Over/Under Rated Cap.(ki)
(FY 1994):	44.32	44.32	59.08
(FY 1993):	42.72	42.72	60.54

MEAN SENTENCE LENGTH(ki)	Institution	Same Sec. Level	Sat. Camp	ICC
(Months) (1/95):	142	142	.	.
(1/90):	.	.	.	.

VF: Management administration and planning. [adm]

INMATE POPULATION(ki)	Inmate Pop.	% Over/Under Rated Cap.	Same Sec Lev. Rated Cap.	Sat Camp	ICC
(FY 94):	24312.3	59.08%	59.08%	. %	. %
(FYTD 1/95):	24260	52.08%	52.08%	. %	. %

SOCIAL CLIMATE SURVEY(ore):	% Indicating Crowding 91/ 92/ 93/ 94	Percentile Ranking within BOP 91/ 92/ 93/ 94
- Perceptions of Crowding	39 38 48 38	. . . .
(1544/ 1735/ 1260/ 1394)		
(82%/ 88%/ 86%/ 91%		

Figure 1. Sample Page of the Data Elements From Goal 1 of the KI/SSS Executive Staff Management Indicators Module.

EXECUTIVE STAFF MODULE

Medium Security Level. . .

JANUARY 1995

GOAL #3: SECURITY AND FACILITY MANAGEMENT

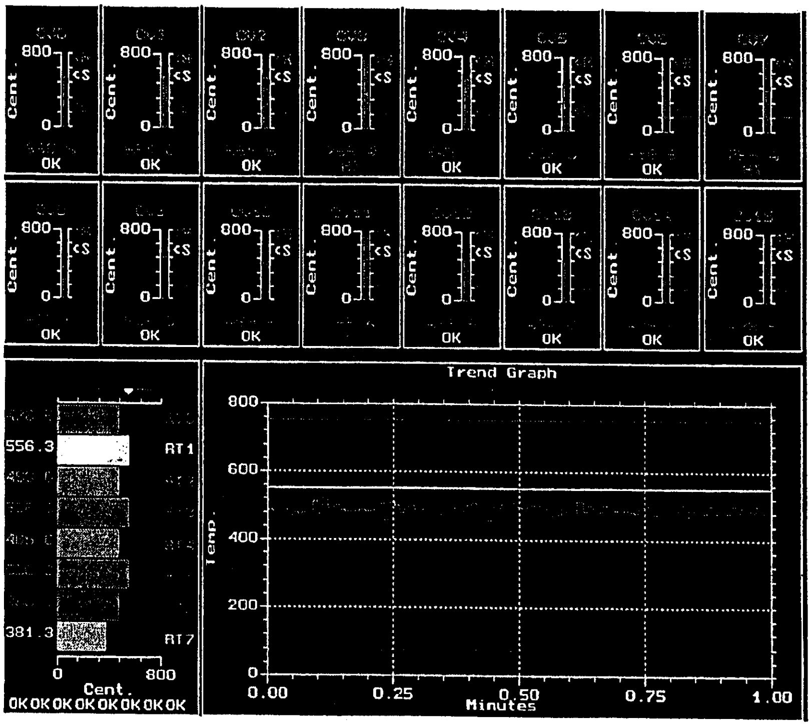
VF: Provide a safe and secure environment for staff and inmates through effective communication of operational concerns. [cor]

SOCIAL CLIMATE SURVEY (ore)	%Indicating Likelihood of Assaults 91/ 92/ 93/ 94	Percentile Ranking within BOP 91/ 92/ 93/ 94
- Staff Perceptions of Inmate Safety (likelihood of assault) (758/ 814/ 636/ 663) (83%/ 88%/ 87%/ 91%)	46 46 48 52	. . . .
- Staff Perceptions of Staff Safety (likelihood of assault) (758/ 814/ 636/ 663) (83%/ 88%/ 87%/ 91%)	23 25 30 33	. . . .

	INSTITUTION cnt	INSTITUTION Rate/100	SAME SEC LVL cnt	SAME SEC LVL Rate/100	SAT CAMP cnt	SAT CAMP Rate/100	ICC cnt	ICC Rate/100
TOTAL ASSAULTS W/O WEAPONS (ki)								
Total on staff								
(YTD 1/95):	34	0.14	34	0.14	.	.	.	.
(1994):	399	1.64	399	1.64	.	.	.	.
(1993):	306	1.24	306	1.24	.	.	.	.
(1992):	235	0.98	235	0.98	.	.	.	.
Total on inmates								
(YTD 1/95):	28	0.11	28	0.11	.	.	.	.
(1994):	245	1.01	245	1.01	.	.	.	.

Figure 2. Sample Page of the Data Elements From Goal 3 of the KI/SSS Executive Staff Management Indicators Module.

however, would be far more compact than the current text format, while providing even better capabilities for comparing institutions with respect to various benchmarks or thresholds. A sample of what these types of graphic displays might look like is provided in Graph 1.



Graph 1. Illustration of Meter Style Graphic Display.

This type of graphic display is an improvement over a text display, and I think one worth implementing, even though it still fails in three important respects. First, it fails to provide any means by which to meaningfully and systematically trim the number of measures. Second, it may not be the most efficient display mechanism for providing institutional comparisons (i.e., where an institution resides in the constellation of measures on which the institutions are being evaluated). And third, it still identifies outliers in an essentially univariate context. It is possible, and likely, that some observations could be inliers (i.e., not outliers) on some number of univariate continua and yet be an anomaly in a multivariate context. Conversely, an observation could appear to be an outlier on one or more univariate views and yet be within the range of acceptable operations once all of the relevant factors are taken into consideration. The limitations of the text and meter graphics can be diminished by developing models of operational processes. That is, the method we describe below is not necessarily a replacement for the preceding methods but rather one to augment and enhance these methods.

Our general strategy is to fit models based on our understanding of operational processes and then, using the interactive methods that have been developed for regression diagnostics, identify the facilities that do not adhere to our expectations, given our specification of the model. If our knowledge of the process is sufficient and the model therefore fits the data well, the regression diagnostics will identify institutions that are outliers in a multivariate context. Our use of the interactive diagnostics is consistent with the intended purpose for which they were developed, however our interest is somewhat different. In a typical application a model is fit and the diagnostics are used to determine whether there are any influential observations that are causing the model to fit poorly. In that context the observations identified are either modified (say via a transformation of one or more variables) or eliminated from the model so as to diminish or prevent their influence. The model is then re-estimated. In the present context the identification of observations that are not well explained by the model, and therefore have large residual statistics, is our primary objective. Since the diagnostics are graphical, it is plausible that models could be fit by individuals with research expertise and the residuals from the models stored for subsequent interactive exploration by Bureau managers, most specifically members of the Bureau's executive staff.

The analysis reported here was conducted using code written in Lisp-Stat (Tierney, 1990; also see Tierney, 1995) called R-code (Cook and Weisberg, 1994). Lisp-Stat is an object-oriented statistical computing and dynamic graphic list processor designed for interactive data analysis. R-code was written to implement graphic methods for regression diagnostics. Lisp-Stat (and therefore R-code) provides facilities for statistical estimation and for interactive graphic exploration of two and three dimensional plots. For example, cloud clusters of data can be rotated in three dimensional space and the labels and coordinates of data points identified by clicking on the data points with a mouse pointer.

Table 1 displays the descriptive statistics for all the measures considered in the two models that illustrate our point. The models are estimated at an institutional unit of measure. The response measures are the per capita inmate cost for fiscal year 1993 (percap93) in dollars, and staff perceptions of the level of crowding outside the housing units (e.g., in the dining, recreation, and programs areas) in September of 1993 (crowd93), expressed as the logit

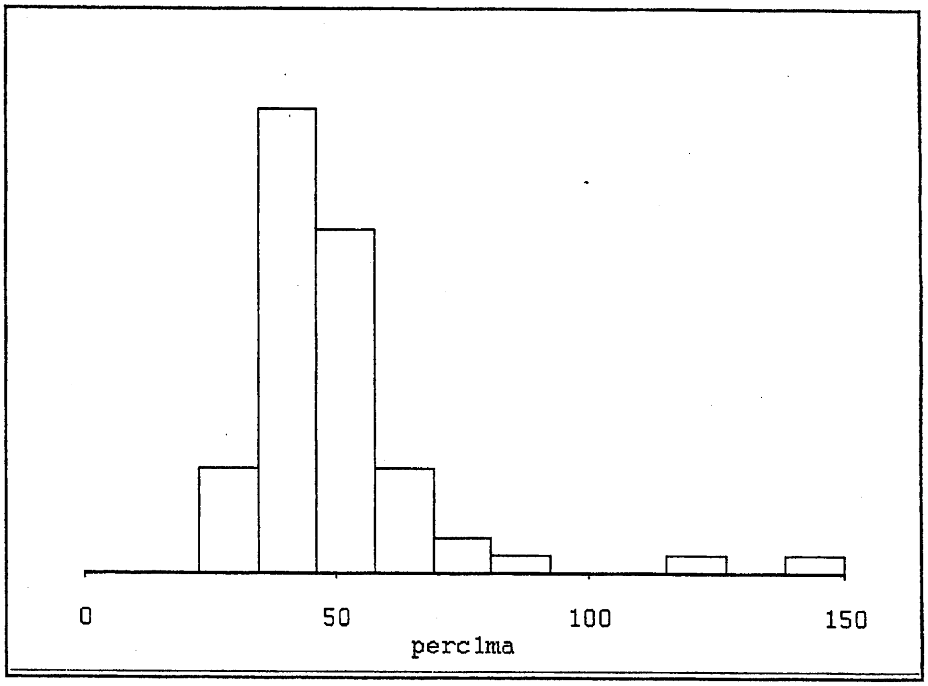
Table 1

Descriptive Statistics for Variables Used in the Models

VARIABLE	LABEL	N	MEAN	MIN	MAX
perc1ma	inmate daily per capita cost '93	65	49.33	0.00	149.09
perc2ma	inmate daily per capita cost '92	65	49.67	30.25	122.31
actv_1mx	inmate-staff ratio '92	68	3.62	1.04	5.86
avp_f1m3	average daily population FY '92	72	950.17	106.40	1996.60
inmtob92	number of inmates in population 11/92	63	1066.71	141.00	2116.00
lslun92	logit of proportion of unclassified inmates 11/92	75	-5.18	-7.26	1.93
lcrwdo3	logit of proportion of staff responding that living conditions are not crowded '93	70	-0.19	-3.31	3.23
lcrwdo2	logit of proportion of staff responding that living conditions are not crowded '92	65	-0.21	-2.56	2.14
linsto3	logit of proportion of staff responding favorably regarding institutional/organ. operations '93	70	0.74	-0.99	2.35
actv_xmx	inmate/staff ratio '93	74	3.87	0.32	6.97
rcp_fxmx	% over/under rated capacity FY '93	74	44.68	-89.02	174.09
ratcap93	facility rated capacity 9/93	70	689.30	254.00	1600.00

of the proportion of staff responding favorably (i.e., the proportion of staff responding that the facility is not crowded). (A logit is the natural log of the odds ratio, in this case the log of the ratio of the proportion of staff responding that the facility is not crowded to the complement of staff responding that the facility is crowded.) The logit is used because it provides a more normal distribution than does the proportion it transforms. The construction of the logit affects only the uniformity of the measure's metric. The interpretation of the logit, therefore, is identical to the interpretation of the proportion. An increase or decrease in the logit corresponds directly to an increase or decrease in the proportion.

Graph 2 displays a univariate histogram for the per capita inmate costs for fiscal year 1993. The graph clearly shows that there are at least two facilities that have very large per capita costs relative to the cost at most facilities. As indicated in Table 1 the average cost is about $50.00 per day while the maximum cost is $150.00 per day.



Graph 2 Histogram of the Per Capita Inmate cost for Fiscal Year 1993.

Table 2 shows the fit of the model for per capita cost 1993 as a function of the per capita costs for fiscal year 1992 and effects vectors for the institution security levels. The model fits quite well and explains almost 90% of the variance in the response measure. As might be expected the lag of per capita cost explains all the variance, since any variation across security level is captured by the lag variable. Graph 3 shows the relationship between

Model name = Percap93, Response = perc1ma
Deleted cases are
(ALF ALM EST FLF FTD GUA MAN MIA)
Coefficient Estimates

Label	Estimate	Std. Error	t-value
Constant	0.838603	5.1622	0.162451
perc2ma	1.027871	0.066531	15.4495
{F}seclevcr[-1]	-5.057273	3.3972	-1.48866
{F}seclevcr[-2]	-2.16236	3.25347	-0.664631
{F}seclevcr[-3]	-1.44811	3.20217	-0.452228
{F}seclevcr[-4]	1.63998	3.78179	0.433652
{F}seclevcr[-5]	aliased		

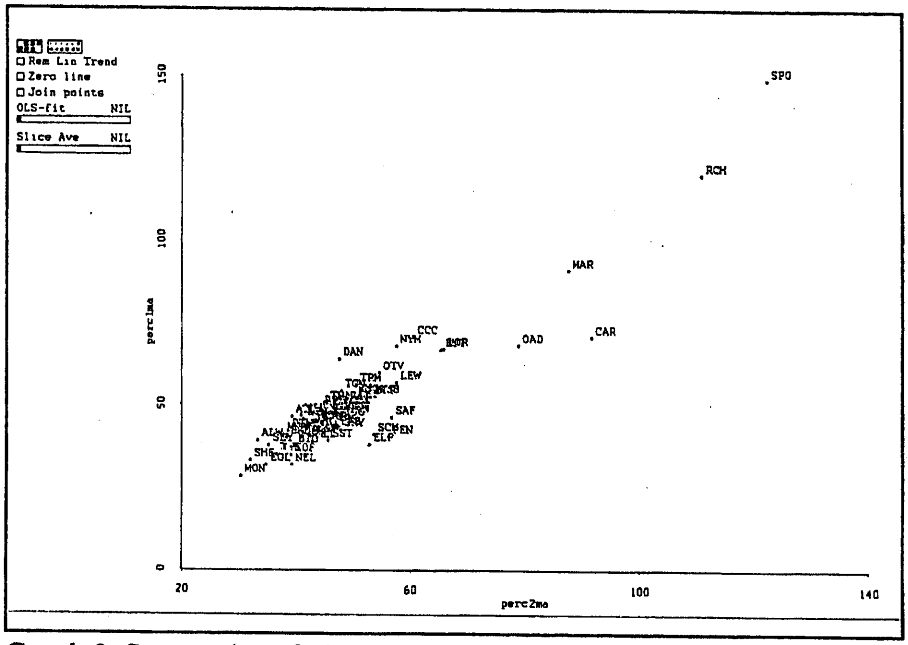
R Squared:	0.87566
Sigma hat:	7.053689
Number of cases:	72
Number of cases used:	64
Degrees of freedom:	58

Model name = Percap93, Response = perc1ma
Deleted cases are
(ALF ALM EST FLF FTD GUA MAN MIA)
Summary Analysis of Variance Table

Source	df	SS	MS	F	p-value
Regression	5	20323	4064.59	81.69	0.0000
Residual	58	2885.76	49.7545		

Table 2. Per Capita Inmate Cost Fiscal Year 1993 as a Function of Fiscal Year 1992 Per Capita Inmate Costs.

the 1992 and 1993 per capita costs. The extreme values in the histogram belong to observations MAR, RCH, and SPG. MAR corresponds to what was, in 1993, the Bureau's most secure facility and the other two correspond to the Bureau's major hospital facilities. If the threshold mechanism identifies only the facilities above a certain absolute or relative level, then these are certainly facilities which will be identified. However, considering the nature of these facilities their extreme values on the per capita scale are to be expected. It would be more informative to have a mechanism which identifies observations that do not fit our expectations.



Graph 3. Scatterplot of Fiscal Year 1992 and Fiscal Year 1993 Inmate Per Capita Costs with all Observations Identified.

Table 3 displays the final model for the per capita costs for fiscal year 1993. The model fits good and explains about 85% of the variance in the per capita costs of institutions. We have excluded the three institutions with the extreme costs, since they are consistent with our expectations and would prevent us from observing the more subtle outliers. The model now includes four additional measures, all of which are also lagged (1992). The t-values can be interpreted as partial standardized coefficients and provide the relative importance of each measure, where importance is defined by the amount of explained variance that would be reduced if a variable were dropped from the model (Bring, 1994). As might be expected, the lag of the per capita cost is still the most important in explaining the response variance, followed by the inmate to staff ratio, the logit of the proportion of inmates who are unclassified (indicating they are in a detention status, which is apparently more costly to the Bureau), a

Model name = Percap931, Response = perc1ma
Deleted cases are
(ALF ALM EST FLF FTD GUA MAN MAR MIA RCH SPG)
Coefficient Estimates

Label	Estimate	Std. Error	t-value
Constant	47.0247	7.19166	6.53878
perc2ma	0.516452	0.0736662	7.010712
actv_1mx	-5.4154	0.988759	-5.47697
avp_f1m3	-0.00705418	0.00349597	-2.017802
inmtob92	0.00718203	0.00281066	2.55528
lslun92	0.871515	0.219963	3.9621

R Squared:	0.855455
Sigma hat:	3.98852
Number of cases:	72
Number of cases used:	61
Degrees of freedom:	55
Cp (rel to full model):	2.911

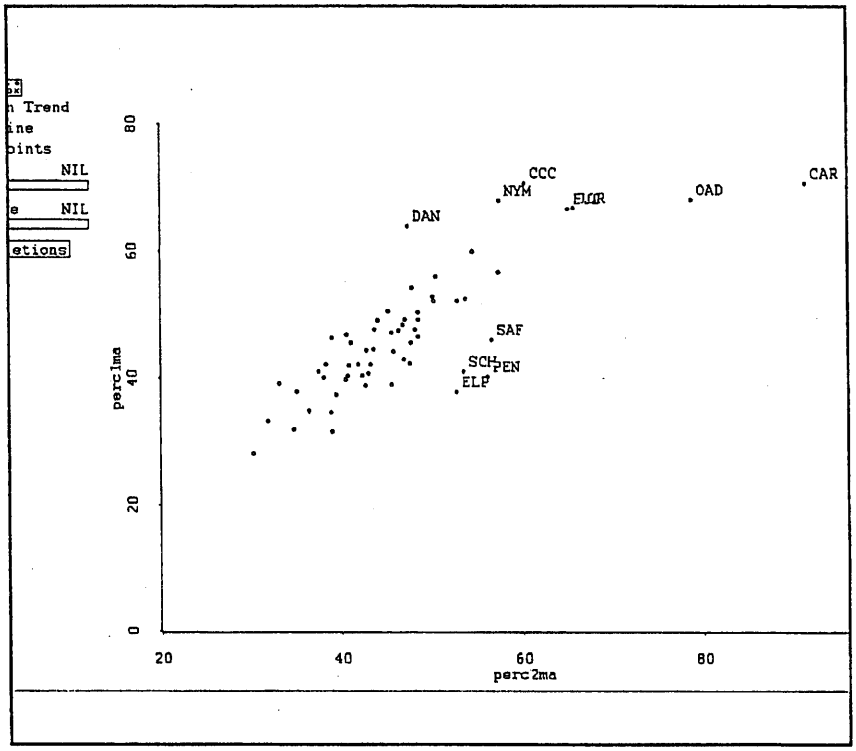
Model name = Percap931, Response = perc1ma
Deleted cases are
(ALF ALM EST FLF FTD GUA MAN MAR MIA RCH SPG)
Summary Analysis of Variance Table

Source	df	SS	MS	F	p-value
Regression	5	5178.19	1035.64	65.10	0.0000
Residual	55	874.955	15.9083		

*****All Wald Tests exceed 1.96

Table 3. Final Model for Fiscal Year 1993 Per Capita Inmate Costs.

count of the number of inmates in population at the end of fiscal year 1992, and the average daily population over the course of fiscal year 1992, respectively.

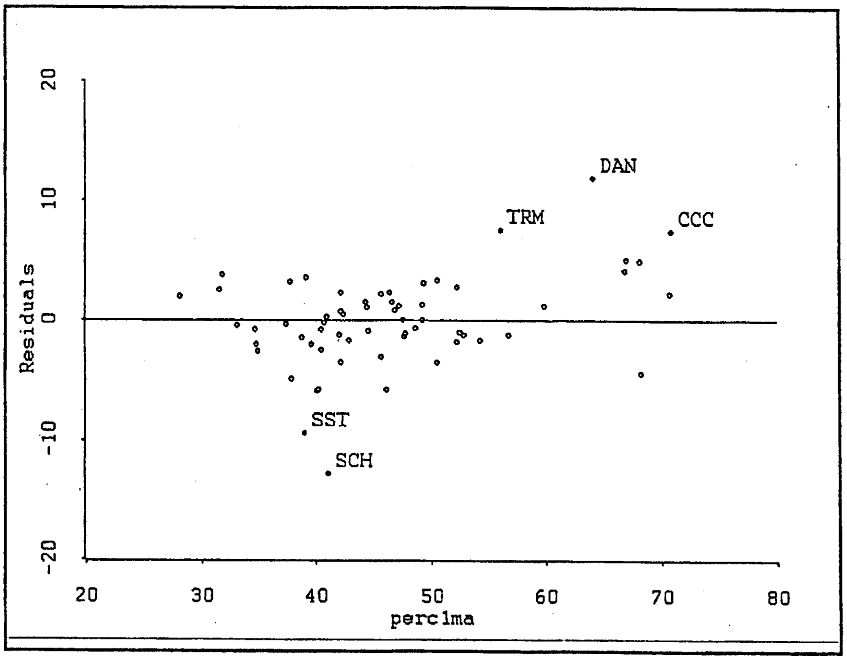


Graph 4. Scatterplot of Fiscal Year 1992 and Fiscal Year 1993 per capita inmate costs with Observations MAR, RCH, and SPR removed.

Graph 4 shows the scatterplot of fiscal year 1992 and 1993 with the three observations removed. The graph shows DAN slightly above the plane in which most institutions lie, and a small group of facilities that lie just below the plane. Two facilities, OAD and CAR, are the most extreme for fiscal year 1992 and they are among the most extreme for fiscal year 1993. These two institutions have special missions. One of the institutions is a detention facility and the other a medical facility, so their extreme values are not surprising.

Graph 5 displays a scatterplot of the observed values by the unstandardized residuals of the model. (The model residuals are the discrepancies between the observed values and the values predicted by the model and are an indication of the degree to which the model is inadequate to explain the behavior of the observations). The unstandardized residuals are used because we are analyzing the population, not a sample, and because we are interested in viewing the residuals in the metric of the response measure. The institutions with the largest residuals have been identified by their Bureau mnemonic in the graphs. The residual values are the dollar amount differences between the observed and predicted per capita cost. Observations DAN and TRM have an observed average daily cost that is about $12 and $8 more, respectively, than is predicted by our model. Conversely, SCH and SST have an observed average daily cost that is about $12 and $8 less, respectively, than is predicted by our model.

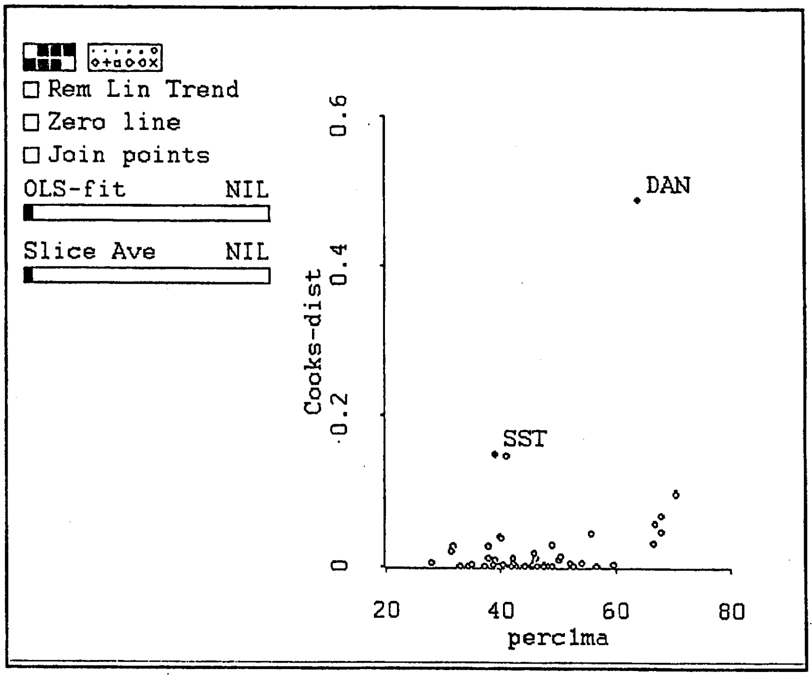
A comparison of Graphs 4 and 5 demonstrates that, with the possible exception of DAN, the institutions with the largest residuals in Graph 5 (those that the model is least able to accurately predict) would not have been identified or expected based on either the univariate or bivariate perspective depicted in Graph 4. Furthermore, the two institutions with the most extreme univariate and bivariate values depicted in Graph 4, do not have extreme residual values in Graph 5. Even though these two facilities had extreme values on the response variable and its lag, these response values are well described by the model (i.e., detention and medical facilities are expected to have consistently higher per capita costs).



Graph 5. Scatterplot of the Observed Values of Fiscal Year 1993 Per Capita Inmate cost and the Model Residuals.

Graph 6 displays Cook's distance for the Fiscal Year 1993 per capita inmate costs. This regression diagnostic measure is a function of the studentized (standardized) residual, the leverage (a quantity that depends on the predictors in the model, and is largest for observations farthest from a predictors mean and smallest for observations closest to a predictors mean), and the number of predictors. This measure further illuminates the extent to which DAN violates our expectations as they are dictated by the model used to produce this set of residual values.

From a management perspective this does not indicate that there is definitely a problem at this facility. However, it does suggest that a prudent manager look into why the per capita cost for this institution does not adhere to the same model (is not a function of the same predictors) as the other Bureau facilities. In fact, the behavior of DAN is readily explained by a mission change it experienced in FY 93. The facility changed from male to female inmates, and consequently the male inmate population was brought to zero before the female inmate population was introduced into the facility. Since the number of staff at the facility remained the same over the course of the mission change, and because staff salaries are the most costly portion of a facilities operating expenses, the per capita inmate cost was higher than one would have otherwise expected.

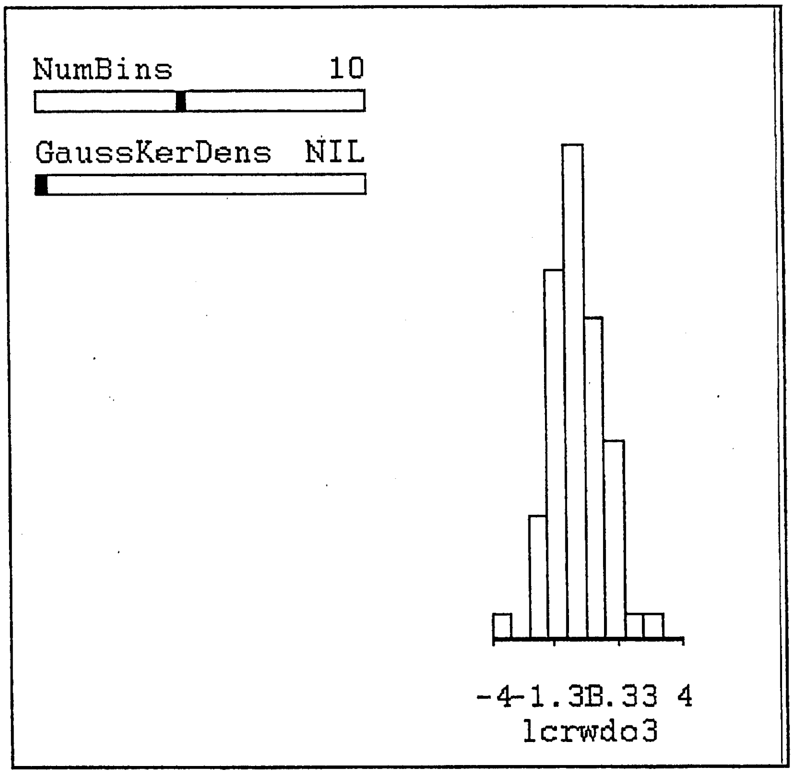


Graph 6. Cook's Distance for Final Model of 1993 Per Capita Inmate Costs.

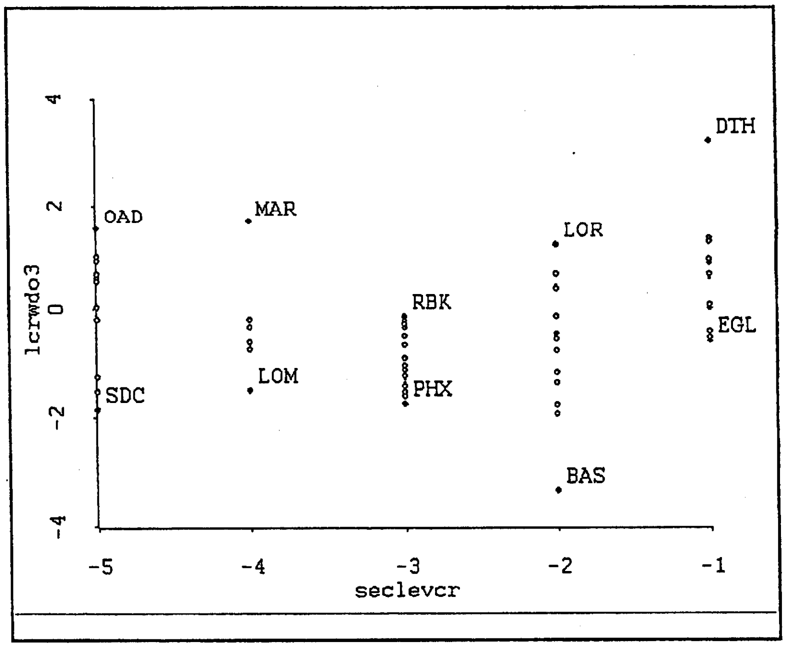
We turn now to the model for staff's perceptions of the level of crowding outside the housing areas. Graph 7 displays the histogram for the response measure. The metric is the logit of the proportion of staff responding favorably to their perception of institution crowding on the 1993 administration of the Prison Social Climate Survey. The distribution is fairly symmetric, although there are some observations on the left side of the scale that are separated from the rest.

Table 4 provides the fit of the model. The fit of the model is fairly good, with the predictors explaining about 73% of the institutional variance for staff perceptions of crowding.

Once again we rely on the t-values to indicate the relative importance of the measures. The most important is the lag of the institution's value on the crowding measure, followed by the inmate to active staff ratio for 1993. It is also interesting to note that management apparently does make a difference, as the measure of staff's perceptions of the efficiency and effectiveness of institutional operations has a positive influence on the institutional level of perceived crowding. The remaining predictors are all rather intuitive: the percent over or under the facilities rated capacity, the rated capacity (the number of inmates the facility is rated to house), and the effects vectors for minimum and hi security level facilities.



Graph 7. Histogram for the Logit of the proportion of staff responding favorably to the crowding measure in 1993.



Graph 8. Scatterplot of Facility Security Level and the logit of the Proportion of Staff who responded Favorably to their Perception of Crowding in 1993.

Graph 8 displays the crowding measure by institution security level. The -1 corresponds to minimum, -2 to low, -3 to medium, -4 to hi, and

Model name = Crowd9321, Response = lcrwdo3
Deleted cases are
(ALF ALM EST FLF FTD GUA MAN MIA)
Coefficient Estimates

Label	Estimate	Std. Error	t-value
Constant	1.6332	0.427574	3.81968
lcrwdo2	0.421103	0.0935937	4.49927
linsto3	0.334575	0.129437	2.58484
actv_xmx	-0.306894	0.0938677	-3.26943
rcp_fxmx	-0.00646204	0.00320585	-2.015703
ratcap93	-0.00112437	0.000428711	-2.62267
{F}seclevcr[-1]	0.922314	0.290925	3.17028
{F}seclevcr[-4]	0.565014	0.288479	1.9586

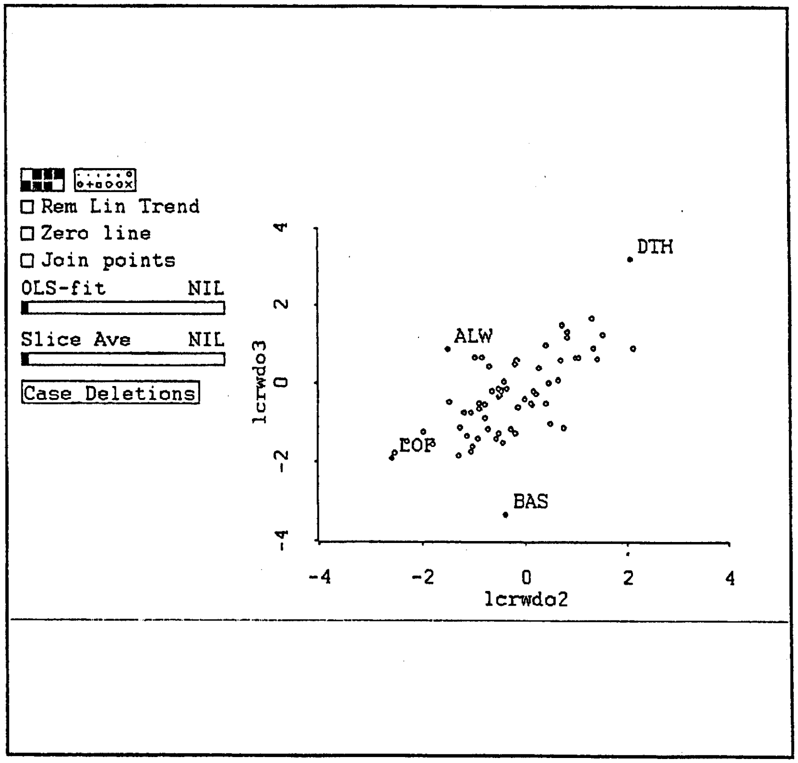
R Squared:	0.731413
Sigma hat:	0.605877
Number of cases:	72
Number of cases used:	64
Degrees of freedom:	56
Cp (rel to full model):	6.279

Model name = Crowd9321, Response = lcrwdo3
Deleted cases are
(ALF ALM EST FLF FTD GUA MAN MIA)
Summary Analysis of Variance Table

Source	df	SS	MS	F	p-value
Regression	7	55.9802	7.99717	21.79	0.0000
Residual	56	20.5569	0.367087		

Table 4. The Logit of the Proportion of Staff Responding Favorably to their Perceptions of Crowding Outside the Housing Area, as a Function the Lag of That Perception & Other Population Characteristics.

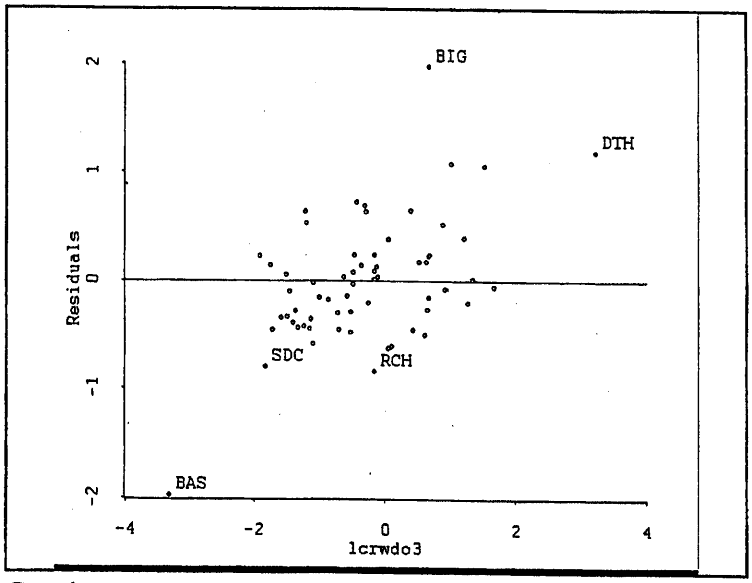
-5 to administrative facilities. The larger the crowding value the more favorable (less crowded) are staff's perceptions. Graph 8 shows that minimum and low facilities have a wider range of values than the other security levels, and that medium security level facilities have the smallest range of values. BAS staff have by far the lowest favorable perception of crowding, that is, a larger portion of the staff responding to the questionnaire in 1993 believed the facility was crowded.



Graph 9. Scatterplot of the Logit of Favorable Views of Crowding Levels for the 1992 and 1993 administrations of the PSCS.

Graph 9 shows the relationship between the institutional measures for 1992 and 1993. The facilities at the extremes have been identified by their Bureau Mnemonics. The institution in the upper right corner, DTH, is a minimum security facility, as is ALW. The facilities BAS and LOF are low security.

Graph 10 shows the residuals for the model plotted against the observed values. We see that DTH and BAS are outliers, as we would expect based on what we have already observed. However, we also see that BIG has an even larger positive residual. In this instant the positive residuals are favorable, indicating that these facilities have a larger portion of staff who see the crowding level favorably than we would expect based on our model. We also see that SDC and RCH have a smaller segment of their staff who view the crowding level favorably than our model predicts. The residual values are in the logit metric of the response measure. Therefore, the approximate proportions of staff who view the crowding level more or less favorably than our model predicts can be obtained from the graph by taking the antilog of the residual. Since the logit is based on the natural log this is the exponential. Taking the exponential we see that BIG had about 7 staff who had a favorable view to each member with an unfavorable view, while DTH had about 3 staff with a favorable view to each member with an unfavorable view. Conversely, SDC and RCH had about 4 staff in every 10 who expressed a favorable view, while BAS had only about 1 in every 10 express a favorable view. The Cook's distance diagnostic statistic was evaluated for this second example but yielded no additional understanding over what was observed in the plot of the unstandardized residuals.



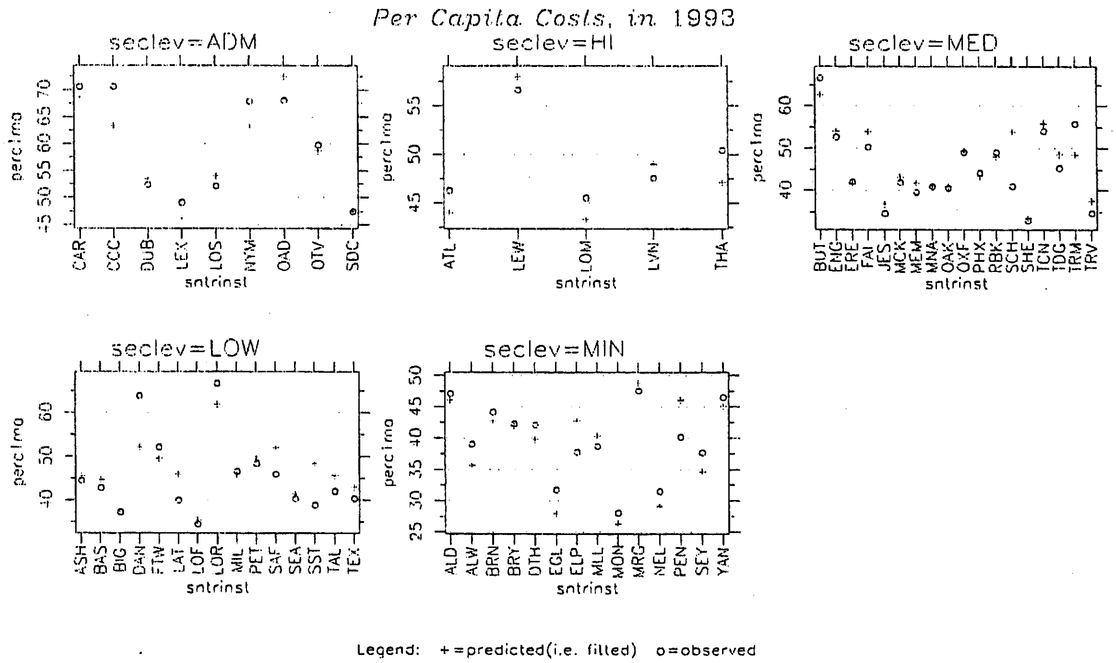
Graph 10. Scatterplot of the Observed Values of the Perceptions of Crowding and the Unstandardized Model Residuals.

Further Graphical Musing

The graphical methods demonstrated thus far exhibit interesting possibilities for establishing the adequacy of our view of organizational processes, as they are expressed by our models, and they enhance our understanding of the performance of units within our organization. However, they are more effective in the interactive mode permitted via a computer than they are in the static mode depicted on paper. At least at this point in time it is more likely that the Bureau's executive staff members would pursue the strategies for organizational performance monitoring suggested here on paper rather than interactively via computer. In consideration of that, the static versions of the graphics demonstrated thus far are limited in that they often obscure the identity of some observations and therefore limit one's view of the full constellation of organizational units within the context of their performance throughout some organizational process.

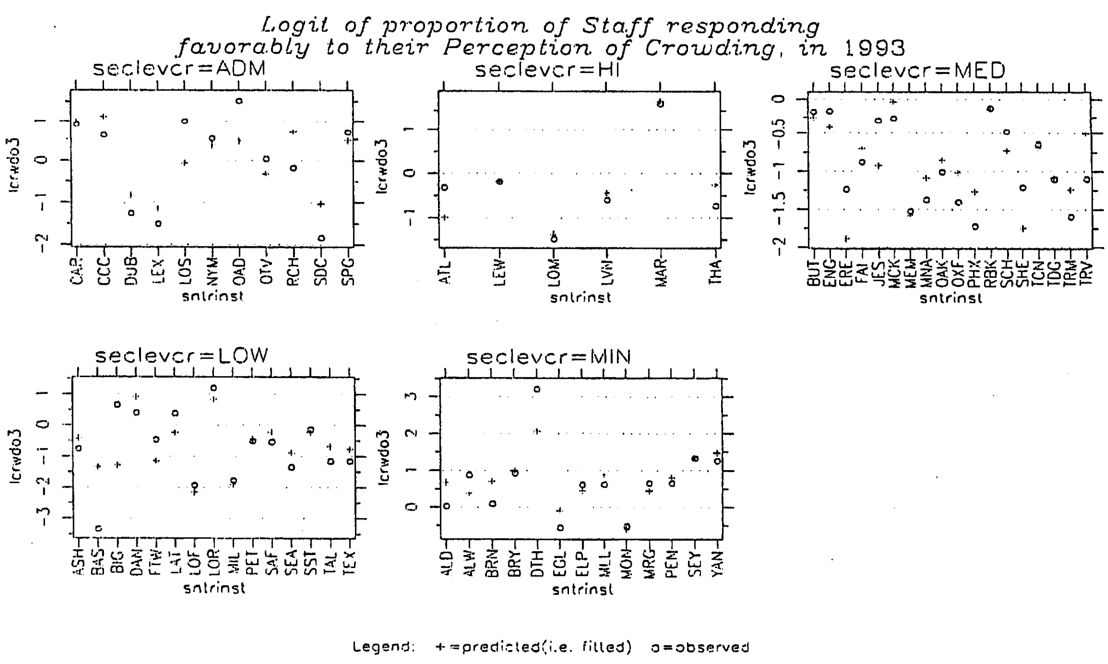
Another possibility that can provide a more adequate static mode of application, but also has some interesting possibilities in an interactive mode as well, is the multiway dot plot suggested by Cleveland (1993). Graphs 11 through 16 illustrate several variations of this type of graphic display using the per capita cost and perceptions of crowding measures. The "o" in the graph represents each institution's observed value, that is, the value produced by our measurement device, our accounting procedures in the per capita costs example and the PSCS in the perceptions of crowding example. The "+" represents each institution's predicted value, or the value we expect the facility to have if it adheres to our view of the organizational process, as it is captured by our model.

One could interpret an institution's predicted value as an institutional statistical composite - one composed of the expected contributions that each of the relevant operational characteristics of the institution (e.g., inmate missions, rated capacities, security levels, and so forth) make to the value of its performance measure. The predicted value yields the performance measure that we expect of an identically equipped facility, when the behavior of each of the operational components that make up our target institution are consistent with the performance behavior of these same components in other institutions, even though the same combination of components may not actually exist in any other institution. In other words, the model provides a replica of what a particular facility's performance would be if an identically configured institution existed, and it was performing per the expectations established by our model. Stated differently, the predicted value for each institution is an excellent performance benchmark because it allows one to easily determine whether an institution is performing above, at, or below what would be expected of a similarly configured institution that was operating in a typical or average capacity.



KI/SSS 5/30/1996 20:11 BOP-0

Graph 11. Multiway dot plot of observed and predicted Per Capita costs for observations ordered alphabetically.



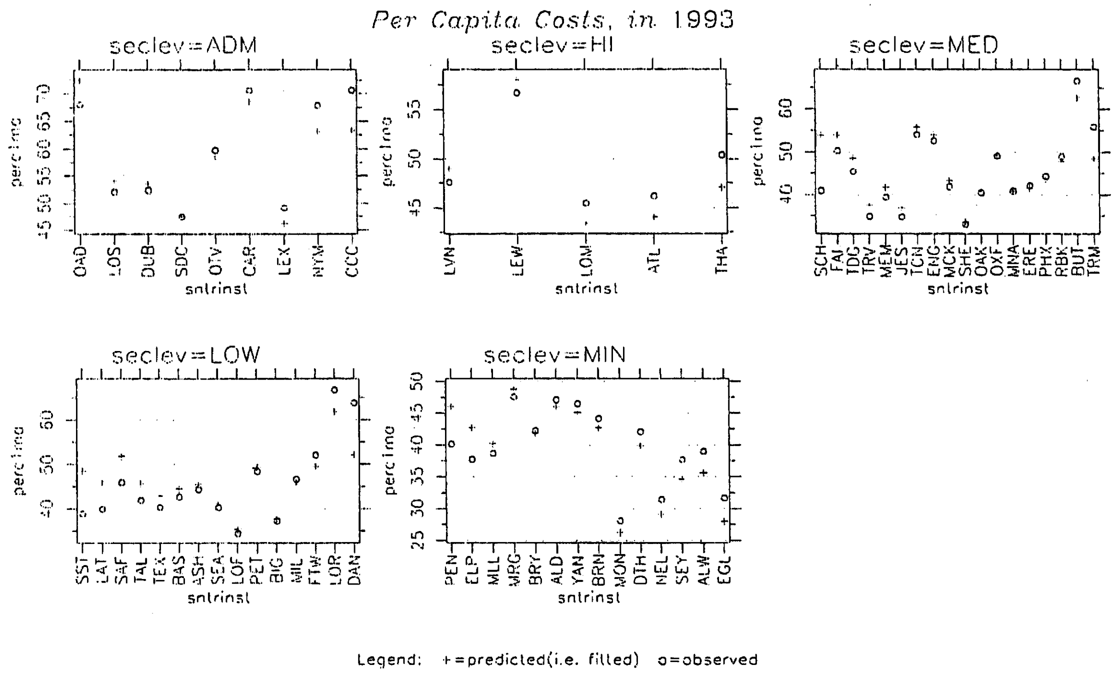
KI/SSS 5/30/1996 20:19 BOP-0

Graph 12. Multiway dot plot of observed and predicted perceptions of crowding for observations ordered alphabetically.

Graphs 11 and 12 provide the institutions ordered alphabetically, and would be convenient if one's interest is in locating a particular correctional facility. This might be necessary if, for example, one is interested in comparing the expected and observed values of a model for the same facility or set of facilities at several points in time. Or, possibly, if one is interested in comparing the behavior of a facility or set of facilities on some number of models.

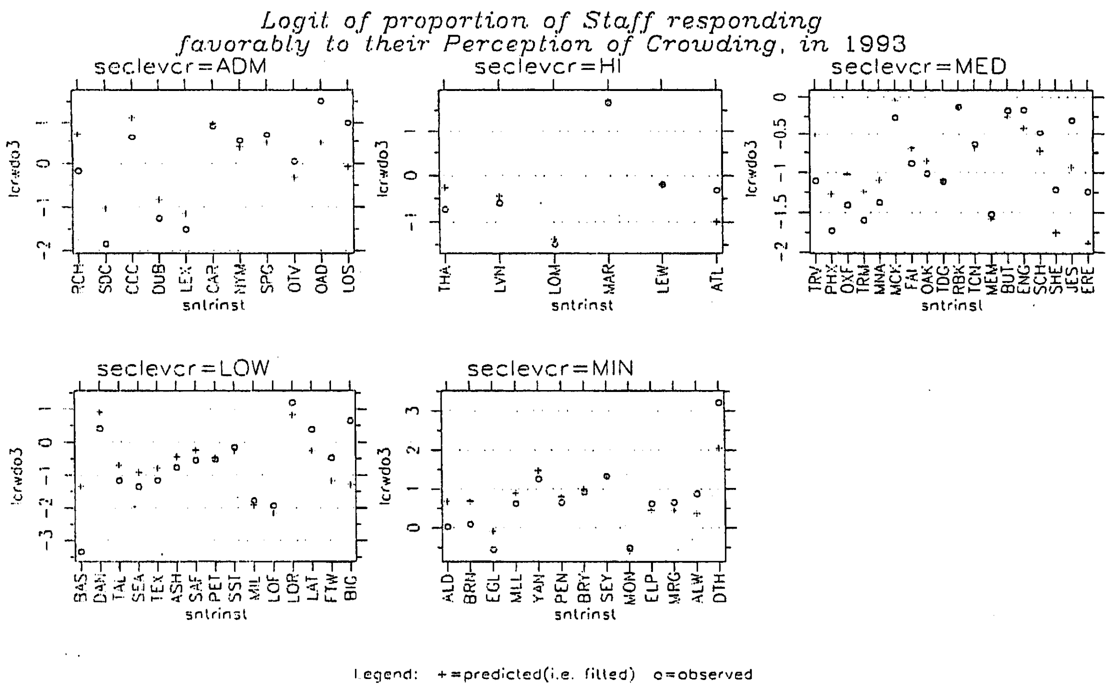
Alternatively, graphs 13 and 14 display the institutions ordered by the size of their residual, that is, by the magnitude of the facility's deviation from our expectation, as that expectation is conveyed by our model of an organizational process. This set of graphs allows us to quickly locate those facilities that are farthest below or above their expected performance by looking at the left and right ends of the graphical spectrum, respectively, and those that performed closest to our expectations by viewing the facilities in the center of the graphical spectrum. For example, in the per capita cost for administrative types of facilities (ADM), in the upper left hand corner of graph 13, facility OAD was the least expensive relative to what we expected (the predicted value "+" is greater than the observed value "o"), while OTV's observed cost was closest to its expected cost (the "o" and "+" are close together), and CCC was the most expensive relative to its expected cost (the "o" value is greater than the "+" value).

These residual values can be viewed as a proxy measure of management's effectiveness with respect to the process that has been modeled. That is, since the model adjusts the outcome measure for all of the influential or confounding factors that should explain the behavior of an institution's outcome measure, the difference between each facility's observed value and its predicted value is, arguably, due to the influence of local management. Presumably, then, good management practices will result in residuals that are large in the desirable direction and less effective management will result in residuals that are large in the undesirable direction.



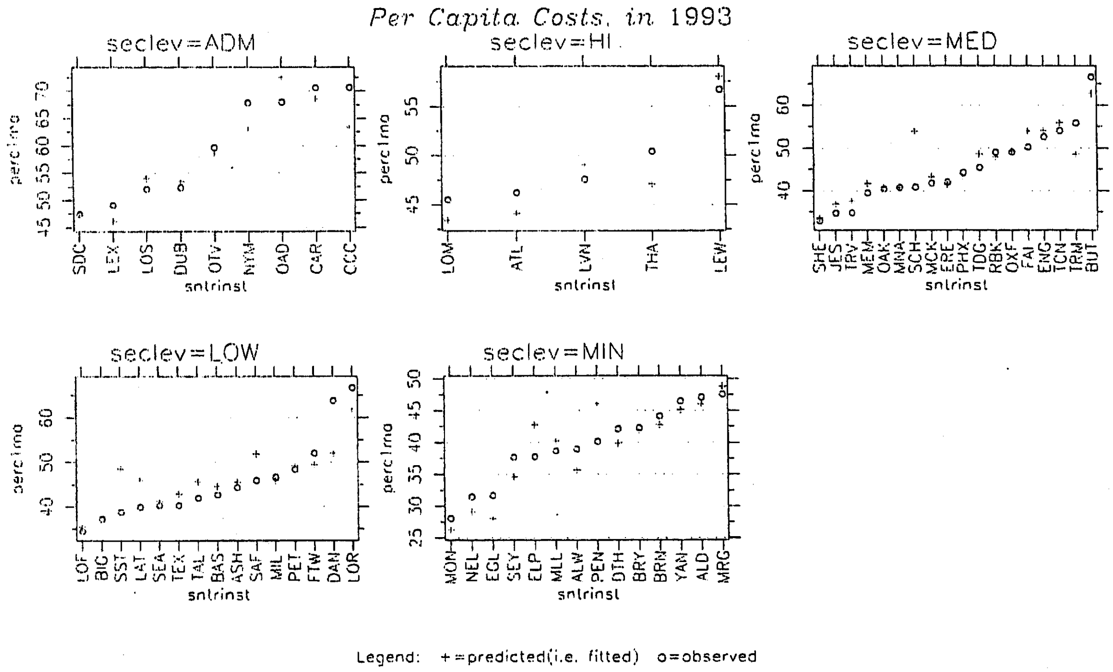
KI/SSS 5/30/1996 20:15 BOP-0

Graph 13. Multiway dot plot of Per Capita costs with observations ordered by the size of their residual.



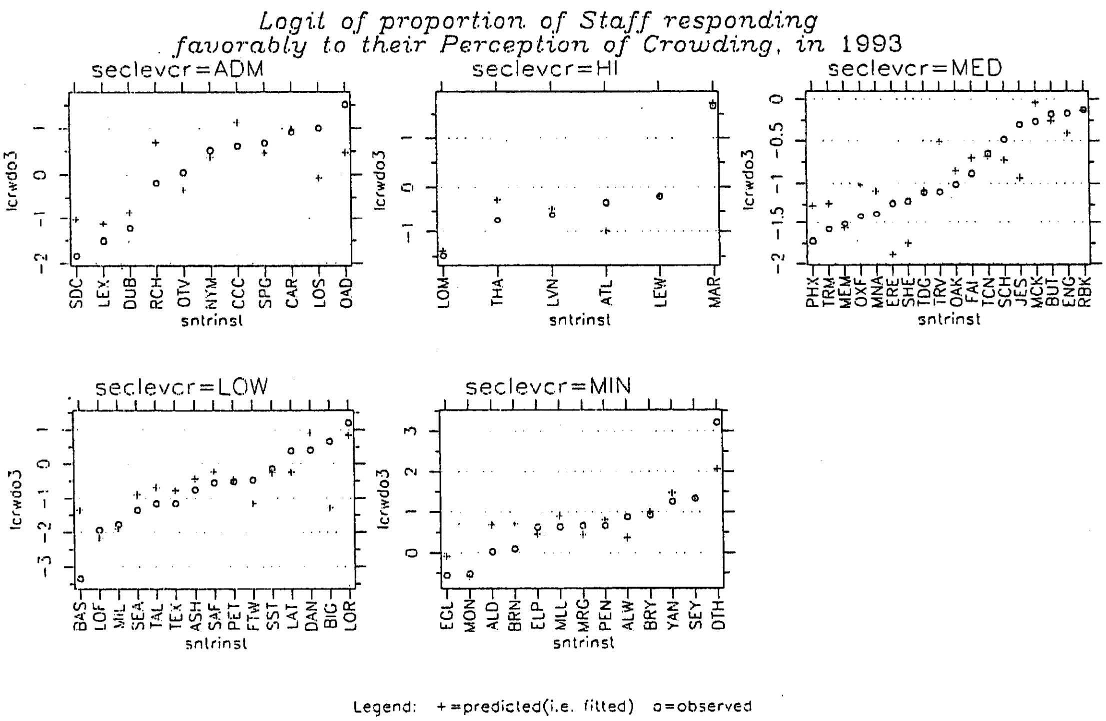
KI/SSS 5/30/1996 20:21 BOP-0

Graph 14. Multiway dot plot of perceptions of crowding with observations ordered by the size of their residual.



KI/SSS 5/30/1996 20:14 BOP-0

Graph 15. Multiway dot plot of Per Capita costs with observations ordered by their observed values.



KI/SSS 5/30/1996 20:20 BOP-0

Graph 16. Multiway dot plot of perceptions of crowding with observations ordered by their observed values.

The third pair of graphs, 15 and 16, display the facilities ordered by their observed values and, therefore, provides quick access to the location of a facility based on its observed value, as well as how the facility performed relative to our expectation. This depiction is also useful in observing whether facilities at any particular location on the continuum tend to be systematically over or under our expectation, and may suggest a problem with the adequacy of the fit of our model.

Each of the three variations of the dot plot could be useful. The one which is most suitable depends on the question being asked. This suggests that an interactive implementation of the dot plot, which would allow the user to click on a button to obtain one of the three, or more, views would be the most compact method for this application.

Conclusion

The methods demonstrated can provide management with an efficient means by which to filter institutions so as to quickly identify those that require additional observation and analysis. Although the methods will illuminate facilities that are performing differently from the rest (in either a positive or negative manner), in many if not most instances the methods will not yield an explanation as to why a facility's performance measure is different from what is expected. An explanation will generally require a closer analysis, one that brings additional information to bear on our understanding of why the model was insufficient to explain an institution's performance measure. This additional information could come from the executive staff's individual or collective experience or could be acquired by them through a search of, say, the executive staff management indicators module or KI/SSS in general. The method will inevitably lead to a better understanding of the process and therefore to decisions about corrective actions, policy development or other innovations that improve the process.

Statistical process control management philosophies have demonstrated good performance in both industrial and service organizations. With the appropriate vehicle for producing continuous performance measures and a knowledge of the organization's operational processes, management can achieve their organization's goals and objectives. We have demonstrated some of the methods we are exploring to provide the Bureau of Prisons' management with graphic tools that will allow them to efficiently and interactively evaluate the Bureau's performance, and identify units that might be contributing unwanted variance to a process.

Nevertheless, some may still wonder why there is a need to model organizational processes in order to make comparisons. Indeed, some may remain unconvinced that it is even necessary to engage in comparisons of performance measures at all. Or, whether this methodology isn't making the task of comparing organizational performance measures more complex than is warranted. Why should Bureau of Prisons' managers, or other corrections managers, invest the additional effort and time to think in terms of models of organizational processes rather than forging ahead by comparing more readily available sets of singular performance measures in a tabular or graphic mode. How far off can a manager's inferences be if he or she compares sets of singular measures for facilities with the same security level or facilities that appear comparable in some other respect?

With respect to the need for performance monitoring (or at least the computerization of such processes), the justification appears to be based on the movement of our society further into the information age. With the increasing complexity of contemporary society and the organizations it is composed of, it makes little sense for an organization to labor with the complexity while ignoring the growth in technology that has been developed to help cope with that complexity. (Some may reasonably argue that society and organizations are more complex because of the advance of technology, and many may also agree that the increase in complexity and the increase in technology are part of a cyclical process. The end result, however, is that there is a level of complexity which contemporary managers of large organizations must contend with that did not exist for their earlier counterparts.) Automation of such processes can provide more uniformity and detail than could be achieved otherwise. Information management and information analysis tools become more valuable as organizations increase in size and complexity, and as individuals move further up the organizational hierarchy (and therefore more distant from the day to day operational details of the organization). It might be argued that the essence of good management is attention to details, and that the reason for information technology is to help manage the large number of details that characterize large organizations.

With respect to the issue of complexity, the modeling methods are not making the organizational processes complex. The organizational processes are complex and the methods are doing no more than to acknowledge that complexity and offer some means with which to deal with that complexity and promote sound decision making. Bureau of Prisons' institutions, like the correctional facilities of other jurisdictions, are multifaceted operations composed of many different activities and functions. In correctional institutions many of the activities and functions are dictated by the nature of the inmate population or populations confined at that facility (e.g., general, detention, medical, high security and so forth) which are determined by the mission or missions that are established for that facility. While there are certainly many communalities among correctional institutions, the permutations of missions that compose each one must be acknowledged in any valid comparison of two or more institutions. This becomes difficult, if not impossible, to accomplish if one makes comparisons of singular pieces of data.

However, comparisons need to be made, and will be made, irrespective of the institutional complexities that might challenge or even confound the validity of such comparisons. In reality, there are no two facilities in the Bureau of Prisons which lend themselves to legitimate and meaningful comparison on any set of performance measures without some explicit or implicit adjustments that account for differences in the operations of the facilities being compared. Such adjustments are frequently employed implicitly by individual managers based on their own experiences and observations. This is useful provided that the beliefs, which form the bases for the adjustments, are shared with and agreed upon by other managers in the organization. If managers don't have commonly held beliefs about their organization's processes and how they function this may lead to vagaries in our understanding of these organizational performance measures and how they can be optimized by management. The use of models provides a means by which to establish collective agreement of what is to be optimized and what the process is that leads to that optimization. This is true for comparisons of any of our facilities, but takes on additional meaning in the context of making comparisons for the purpose of evaluating the efficiency and effectiveness of special facilities or new conceptualizations of facilities such as the Federal Correctional Complexes or privatized facilities that add one more dimension to the already unique composites of federal institutions.

March 8, 1995

Revised June 13, 1997

References

Bring, J. 1994. "How to Standardize Regression Coefficients." The American Statistician 48:209-213.

Cleveland, William S. 1993. Visualizing Data. Summit, New Jersey: Hobart Press.

Cook, R. D., and S. Weisberg. 1994. An Introduction to Regression Graphics. New York:Wiley.

Costner, H. L. 1969. "Theory, Deduction, and Rules of Correspondence." American Journal of Sociology 75:245-263.

Einstein, A. 1934. The World as I See It. New York (cited in Laszo, 1972a)

Kiemele, M. J. and S. R. Schmidt. 1993. Basic Statistics: Tools for Continuous Improvement. Colorado Springs, CO:Air Academy Press.

Laszo, E. 1972a. Introduction to Systems Philosophy. New York:Harper&Row.

Laszo, E. 1972b. The Systems View of the World: The Natural Philosophy of the New Developments in the Sciences. New York:George Braziller.

Margenau, H. 1950. The Nature of Physical Reality. New York (cited in Laszo, 1972a)

Saylor, W. G. 1983. "Surveying Prison Environments." Unpublished manuscript, Federal Bureau of Prisons, Office of Research and Evaluation, Washington D.C.

Saylor, W. G. 1988. "Developing a Strategic Support System: Putting Social Science Research into Practice to Improve Prison Management." Paper presented at the American

Society of Criminology, Chicago, 1988.

Saylor, W. G. 1994. "The Design, Development and Maintenance of Strategic Support Systems for Correctional Agencies." Presented at the 1994 International Symposium on Criminal Justice Information Systems and Technology, Washington D.C., 1994

Suppe, F. 1974. "Theories and Phenomena." Pp. 41-91 in Developments in the Methodology of Social Science, edited by Leinfellner and Kohler. Dordrecht-Holland:D. Reidel Publishing Co.

Tierney, L. 1990. LISP-STAT. New York:Wiley.

Tierney, L. 1995. "Data Analysis Using Lisp-Stat." Sociological Methods and Research 23:329-351.