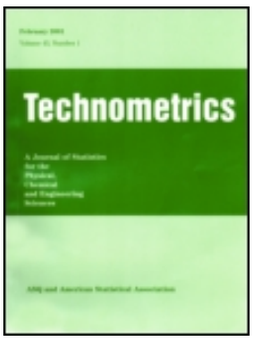

Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: http://amstat.tandfonline.com/loi/utch20

Control Procedures for Residuals Associated With Principal Component Analysis

J. Edward Jackson & Govind S. Mudholkar

To cite this article: J. Edward Jackson & Govind S. Mudholkar (1979) Control Procedures for Residuals Associated With Principal Component Analysis, Technometrics, 21:3, 341-349

To link to this article: http://dx.doi.org/10.1080/00401706.1979.10489779

TECHNOMETRICS©, VOL. 21, NO. 3, AUGUST 1979

Control Procedures for Residuals Associated With Principal Component Analysis

J. Edward Jackson
Eastman Kodak Company
Kodak Park Division
Rochester, NY 14650

Govind S. Mudholkar
Statistics Department
University of Rochester
Rochester, NY 14620

This paper is concerned with the treatment of residuals associated with principal component analysis. These residuals are the difference between the original observations and the predictions of them using less than a full set of principal components. Specifically, procedures are proposed for testing the residuals associated with a single observation vector and for an overall test for a group of observations. In this development, it is assumed that the underlying covariance matrix is known; this is reasonable for many quality control applications where the proposed procedures may be quite useful in detecting outliers in the data. A numerical example is included.

KEY WORDS

Principal components
Residual analysis
Multivariate quality control

1. INTRODUCTION

In the early days of principal component analysis, most of the attention was devoted to ways of obtaining characteristic roots and vectors from a covariance or correlation matrix and to interpretation of the roots and vectors. In more recent times, the effort has shifted to the problems of inference, such as estimation and tests of hypotheses concerning these parameters, to deeper examination of the optimal properties of principal components and to the development of new tools for the application of these techniques. The improved understanding of principal components as a data reduction tool, their role in applications such as regression analysis and multivariate quality control, and the availability of high speed computers and convenient software packages have made the incorporation of principal component techniques in routine data analysis not only feasible but common. This increased use implies imperatives such as model fit questions in general and examination of residuals resulting from using a subset of principal components in the model in particular. The problems are similar to those in regression analysis and require a test for outliers.

Whether principal components are used as a data reduction technique, a diagnostic tool or a control device, the residuals associated with them are useful for checking the fit and testing for outliers. In this paper we are concerned with the control situation, which together with its associated residual analysis, has been considered previously [7], [8], [9], [11]. Specifically, our objective is to examine and extend the use of a statistic suggested by Jackson and Morris for control purposes. In Section 3, we propose a Gaussian approximation for the distribution of their statistic and later compare it with the distribution conjectured by them. Later, in Section 6, we consider the use of principal components for control when a group of observations are available. After reviewing the use of Hotelling's statistics in this context, we propose some omnibus alternatives for the analysis of the residuals.

Received May 1977; revised November 1978

2. PRELIMINARY RELATIONSHIPS

Let $\mathbf{x}' = [x_1, x_2, \cdots, x_p]$ be a set of random variables and let us assume that $\mathbf{x}$ has a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. Further, assume that $\boldsymbol{\Sigma}$ has full rank p.

It will be convenient, in the development to follow, to define the characteristic vectors by the relationship:

$$\mathbf{U}'\boldsymbol{\Sigma}\mathbf{U} = \boldsymbol{\Lambda}, \tag{2.1}$$

where Λ is a $p \times p$ diagonal matrix of distinct characteristic roots $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$ and $\mathbf{U}$ is an orthogonal matrix such that $\mathbf{U'U} = \mathbf{I}$. Because U-vectors are scaled to unit length, they are useful for diagnostic purposes as well as establishing mathematical relationships. Some confusion often occurs in the literature because characteristic vectors are often scaled differently. One in particular is:

$$\mathbf{W} = \mathbf{U}\Lambda^{-1/2}, \tag{2.2}$$

where $\Lambda^{1/2}$ is the diagonal matrix made up of the square roots of the characteristic roots. The main utility of W-vectors is that the principal components obtained from them by the relationship:

$$\mathbf{y} = \mathbf{W'x} = \Lambda^{-1/2}\mathbf{U'x} \tag{2.3}$$

have a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{W'\Sigma W} = \mathbf{I}$. Because the principal components all have unit variances and T^2 reduces to $T^2 = \mathbf{y'y}$, this normalization is quite popular as a control device.

A third normalization is:

$$\mathbf{V} = \mathbf{U}\Lambda^{1/2} \tag{2.4}$$

resulting in the relationships:

$$\mathbf{V'\Sigma V} = \Lambda^2 \tag{2.5}$$

$$\mathbf{V'V} = \Lambda \tag{2.6}$$

(i.e. V-vectors are scaled to their roots) and

$$\mathbf{V V'} = \mathbf{\Sigma}. \tag{2.7}$$

The resultant principal components, $\mathbf{V'x}$, are in the same units as the original variables.

If we do not use the full set of vectors, but only the first k of them, we shall denote this $p \times k$ matrix as $\mathbf{U}_k$. Hence:

$$\mathbf{U}_k'\mathbf{\Sigma U}_k = \Lambda_k \tag{2.8}$$

where Λ_k is a $k \times k$ matrix with only the first k characteristic roots dispalyed. The residual covariance matrix after k vectors have been extracted is $\mathbf{\Sigma} - \mathbf{U}_k\Lambda_k\mathbf{U}_k'(= \mathbf{\Sigma} - \mathbf{V}_k\mathbf{V}_k')$. For uniformity, we shall use U-vectors for the rest of this paper although the use of the other normalizations would sometimes result in simpler expressions.

3. TREATMENT OF RESIDUALS

Once the principal components have been obtained from an observation vector $\mathbf{x}$ by (2.3), the adequacy of the model may be checked by obtaining a predicted observation vector $\hat{\mathbf{x}}$ by the relationship:

$$\hat{\mathbf{x}} = \mathbf{U}_k\Lambda_k^{1/2}\mathbf{y}. \tag{3.1}$$

From this we obtain the sums of squares of the residuals:

$$Q = (\mathbf{x} - \hat{\mathbf{x}})'(\mathbf{x} - \hat{\mathbf{x}}) \tag{3.2}$$

which will produce an overall measure of the fit of an observation to the model (3.1).

Let $\theta_i = \sum_{j=k+1}^{p}\lambda_j^i$, $i = 1, 2, 3$ and $h_0 = 1 - (2\theta_1\theta_3)/3\theta_2^2$. It will be shown in Appendix A that the quantity:

$$c = \frac{\theta_1[(Q/\theta_1)^{h_0} - 1 - \theta_2 h_0(h_0 - 1)/\theta_1^2]}{\sqrt{2\theta_2 h_0^2}} \tag{3.3}$$

is approximately normally distributed with zero mean and unit variance. The control limit for Q becomes

$$Q_\alpha = \theta_1\left[\frac{c_\alpha\sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0 - 1)}{\theta_1^2}\right]^{\frac{1}{h_0}}, \tag{3.4}$$

where c_α is the normal deviate corresponding to the upper $(1 - \alpha)$ percentile.

The functions of the last $(p\text{-}k)$ characteristic roots of $\mathbf{\Sigma}$ appearing in this approximation can be obtained with greater accuracy by the relationships:

$$\theta_i = T_r\,\mathbf{\Sigma}^i - \sum_{j=1}^{k}\lambda_j^i, \qquad i = 1, 2, 3 \tag{3.5}$$

especially when the $(p\text{-}k)$ remaining roots are small and numerous.

4. NUMERICAL EXAMPLE

For an illustration, we shall return to the work of Jackson and Morris [11]. Their example was a nine-variable problem where the variables represented red, green, and blue optical density measurements at three exposure levels for a photographic process. The details are given in the original article. The original data for this example are no longer available, so we shall start with the covariance matrix which was given in the article. This is displayed in Table 1. Since only three digits are given for this matrix, the characteristic vectors shown in Table 2 will differ slightly from the original article. The characteristic roots associated with these vectors are also given in Table 2. The last four roots are: .00038, .00007, .00006, and .00004. The quantity $h_0 = .152$.

TABLE 1—*Within-week covariance matrix ($\times 10^5$) in photographic example.*

Shoulder			Middle-Tone			Toe		
Red	Green	Blue	Red	Green	Blue	Red	Green	Blue
177	179	95	96	53	32	-7	-4	-3
	419	245	131	181	127	-2	1	4
		302	60	109	142	4	4	11
			158	102	42	4	3	2
				137	96	4	5	6
					128	2	2	8
						34	31	33
							39	39
								48

Table 3 shows the calculations for a sample data vector given as difference from standard. The principal components, $y_1, y_2, \cdots, y_5$, are scaled such that their standard deviations are all equal to unity, hence the overall measure of departure from the standard, $T^2 = \mathbf{y}'\mathbf{y}$, has a χ^2-distribution with 5 degrees of freedom. T^2 and Q, the sum of squares of the residuals, are also included.

The first column of data represents the fourth observation in the original article. None of the principal components, T^2 nor Q are significant. The second column is exactly the same as the first except that the Blue Toe has a change in sign, representing a type of clerical error that could easily appear from time to time. Although the principal components and T^2 are still quite small, Q is now significantly large, indicating that the principal component model does not adequately represent the data. The third column illustrates another typical error, that of transposition for the Blue Toe. In this case, one of the principal components, y_4, is quite large and T^2 is significant. However, Q is also very large and should point out that the data be carefully examined first before attempting to explain the significance of y_4 and T^2.

The individual residuals are shown in Table 4 and point out that the residuals are not independent of each other and that the larger the aberration, the worse things get. It is not obvious from an examination of the residuals that the Blue Toe is the cause of the problem with either the second or third observation, although it is at least a candidate. This is not surprising since parallel situations are well known in regression, particularly when dealing with models of less than full rank such as design models where a single outlier can affect several residuals.

Remark. In proposing the use of residuals to check the adequacy of the model and to detect outliers, Jackson and Morris conjectured that the quantity

$$\text{Res. } SS = \frac{(p-k)Q}{\theta_1} \tag{4.1}$$

was asymptotically distributed as chi-square with $p-k$ degrees of freedom. It turns out that this conjecture will hold only when the principal components associated with *all* of the significant roots are used in the model or if more than that number are employed. There are many times when some of the significant roots are explaining some of the inherent variability of a system and are not relevant to a control model. If any of these components are dropped, (4.1) will not hold and will result in limits on Q which are too tight thus confounding the effect of outliers with the fact that an incomplete model is used, probably deliberately. Since the distribution for Q derived in this paper will hold no matter how many significant principal components are omitted or how many nonsignificant components are included, the use of (4.1) can no longer be recommended.

5. ALTERNATIVE CONTROL MEASURES

There are a number of possible alternatives to Q as a control measure. For example, one may use $(\mathbf{x} - \hat{\mathbf{x}})'\{\text{Diag}(\boldsymbol{\Sigma} - \mathbf{U}_k\boldsymbol{\Lambda}_k\mathbf{U}_k')\}^{-1}(\mathbf{x} - \hat{\mathbf{x}})$ which is a χ^2_{p-k} variable under the hypothesis of fit. However, the advantage of a simple null distribution is outweighed by its instability due to near-zero residual variances.

Rao [18] suggests:

$$\sum_{l=k+1}^{p} y_l^2 = \mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x} - \sum_{l=1}^{k} y_l^2 \tag{5.1}$$

the sum of squares of the last $(p - k)$ components as a measure of the inadequacy of the k-component model. Gnanadesikan and Kettenring [3] illustrate its use in a diagnostic situation. (Hawkins [4] also discusses this measure, one involving the maximum of the last $(p - k)$ principal components and a third involving its counterpart after varimax rotation.) The measure $\sum_{l=k+1}^{p} y_l^2$ also has a χ^2_{p-k} null distribution. However, its computation involves either $\boldsymbol{\Sigma}^{-1}$ or a large number of uninterpretable principal components; both of these could have some numerical

TABLE 2—*Vectors for photographic example.*

		$\underset{\sim}{u}_1\lambda_1$	$\underset{\sim}{u}_2\lambda_2$	$\underset{\sim}{u}_3\lambda_3$	$\underset{\sim}{u}_4\lambda_4$	$\underset{\sim}{u}_5\lambda_5$
Shoulder	Red	3.25	-10.96	11.48	8.92	12.55
	Green	6.97	- 3.41	5.09	1.29	-22.75
	Blue	5.16	13.27	6.57	3.58	14.89
Middle-Tone	Red	2.79	-11.09	-12.74	-3.37	16.18
	Green	3.45	- .85	-13.80	-8.70	- 5.59
	Blue	2.89	8.41	- 7.49	-7.27	5.41
Toe	Red	.03	1.29	- 7.13	13.62	- 1.44
	Green	.06	1.21	- 7.41	15.69	- 2.41
	Blue	.15	1.99	- 7.87	17.36	- 1.93
	λ	.00879	.00196	.00129	.00103	.00081

TABLE 3—*Calculation of principal components and residual analysis.*

		Obs. 1	Obs. 2	Obs. 3
Shoulder	Red	.01	.01	.01
	Green	.02	.02	.02
	Blue	.01	.01	.01
Middle-Tone	Red	-.01	-.01	-.01
	Green	0	0	0
	Blue	.01	.01	.01
Toe	Red	.04	.04	.04
	Green	.02	.02	.02
	Blue	.02	-.02	.20
Principal Components	y_1	.23	.22	.26
	y_2	.27	.19	.62
	y_3	-.26	.06	-1.68
	y_4	1.32	.62	4.44
	y_5	-.43	-.36	-.78
$T^2 = y' \Lambda y$		2.12	.60	23.60
Q		.00056	.00218	.01696

Data are deviations from standard.

95% Limits for { Principal Components: ± 1.96; T^2 = 11.1; Q: .0017

problems. If this measure were to be significantly large, one would probably have to go back to the individual residuals anyhow to ascertain the nature of the problem. We prefer Q on account of its intuitive appeal as a sum of squares of residuals and also because it requires only the easily computed sums θ_1, θ_2 and θ_3.

6. OPERATIONS WITH SUBGROUPS

We have, so far, been concerned with the residuals associated with a single observation vector. We will now extend this procedure to subgroups of n observation vectors. In using conventional control chart methodology, subgroups are preferable to single observations because of the obvious advantage of the mean over a single observation and also because a measure of dispersion can be charted. Multivariate analogues of these quantities have been developed by Hotelling [5] and have been modified for use with principal components [8], [9], [11].

TABLE 4—*Residuals for data in Table 3.*

		Obs. 1	Obs. 2	Obs. 3
Shoulder	Red	.005	.005	.008
	Green	0	0	-.001
	Blue	-.005	-.003	-.011
Middle-Tone	Red	-.004	-.004	-.004
	Green	-.001	-.001	0
	Blue	-.009	.008	.014
Toe	Red	.018	.031	-.040
	Green	-.005	.009	-.071
	Blue	-.008	-.032	.099

Suppose, now, we have a sample of n vector observations $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$. Let $T_i^2 = \mathbf{y}'\mathbf{y}$ be the sum of squares of the first k principal components for the ith observation. Then:

$$\chi^2_0 = \sum_{i=1}^{n} T_i^2 \tag{6.1}$$

which presents the generalized distance of the sample from the standard, has a χ^2-distribution with nk degrees of freedom.

Restating (2.3) in terms of the sample average we obtain:

$$\bar{\mathbf{y}} = \Lambda^{-1/2}\mathbf{U}'\bar{\mathbf{x}}. \tag{6.2}$$

(Since this is a linear operation, $\bar{\mathbf{y}}$ can also be obtained by averaging the individual principal component vectors, i.e. $\bar{\mathbf{y}} = \sum_{i=1}^{n}\mathbf{y}_i/n$.) Then:

$$\chi^2_M = n\bar{\mathbf{y}}'\bar{\mathbf{y}} \tag{6.3}$$

which represents the distance between the sample mean and the standard, is distributed as χ^2 with k degrees of freedom. The difference:

$$\chi^2_D = \chi^2_0 - \chi^2_M \tag{6.4}$$

which is used as a measure of the sample variability about its own mean, is distributed according to a χ^2-distribution with $(n-1)k$ degrees of freedom. χ^2_M and χ^2_D are multivariate analogues of the univariate mean and range (or standard deviation) charts.

To handle outliers, we propose two additional statistics, Q_M and Q_0. If we restate (3.1) in terms of averages:

$$\hat{\bar{\mathbf{x}}} = \mathbf{U}_k\Lambda_k^{1/2}\bar{\mathbf{y}} \tag{6.5}$$

then

$$Q_M = n(\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}})'(\bar{\mathbf{x}} - \hat{\bar{\mathbf{x}}}) \tag{6.6}$$

which has the same general form as (3.2) and has the same distribution. This represents the inability of the principal components of the averages (or the average of the principal components) to predict the original vector averages.

While Q_M tells us whether or not the average is consistent with the principal component model, it is not enough. We also need to examine the residuals jointly for the information in them regarding the fit of the model to the corresponding observations, since it is possible for an unstable process to produce a satisfactory average. We propose to do this by using methodology currently known as the *combination of independent tests of significance.* An account of these methods may be found in Oosterhoff [17] and George [2]. We propose the use of the well-known omnibus procedure due to Fisher [1] or the more recent logit

TABLE 5—*Analysis of subgroup data: 95% limits for n = 3.*

Principal Components: ± 1.96
Principal Component Average ± .653

T^2 = y'y: 11.06
X_0^2: 24.99
X_M^2: 11.06
X_D^2: 18.31

Q and Q_M: .0017
Q_0: 12.6
Q_L: 1.73

Data in Terms of Deviation from Standard

		Set #1				Set #2				Set #3	
		X_1	X_2	X_3	$\bar{X}$	X_1	X_2	X_3	$\bar{X}$	X_3	$\bar{X}$
Shoulder	Red	.02	-.03	-.08	-.030	-.11	-.09	-.08	-.093	.22	.007
	Green	.02	.01	-.01	.007	-.28	-.29	-.31	-.293	.50	-.023
	Blue	.01	-.03	-.10	-.040	-.12	-.12	-.09	-.110	.20	-.013
Middle Tone	Red	-.01	-.03	-.05	-.030	-.11	-.06	-.09	-.087	.15	.007
	Green	.00	-.01	-.01	-.007	-.08	-.08	-.08	-.080	.15	-.003
	Blue	.01	-.03	-.06	-.027	-.02	-.03	-.01	-.020	.05	.000
Toe	Red	.03	.00	.02	.017	.02	.02	.02	.020	.02	.020
	Green	.02	.01	.05	.027	.02	.02	.02	.020	.02	.020
	Blue	.02	-.01	.01	.007	.02	.01	.03	.020	.01	.013
Principal Components	y_1	.26	-.39	-1.19	-.440	-3.56*	-3.46*	-3.44*	-3.487*	6.32*	-.233
	y_2	.14	-.02	-.25	-.043	1.78	.93	1.83	1.513*	-2.76*	-.017
	y_3	-.07	.26	-.99	-.267	-1.27	-1.58	-1.29	-1.380*	1.66	-.397
	y_4	1.27	.03	.84	.713*	.37	.27	.74	.460	1.91	.850*
	y_5	-.29	-1.65	-3.51*	-1.817*	1.65	2.90*	3.51*	2.687*	-3.87*	.227
Summary Statistics for Individuals	T^2	1.79	2.93	15.50*		20.32*	23.79*	29.68*		68.90	
	Q	.00041	.00038	.00116		.00280*	.00145	.00397*		.00191*	
	p_i	.434	.462	.111		.012	.070	.003		.037	
Summary Statistics for Sub-Group	X_0^2				20.22				73.80*		113.02*
	X_M^2				12.22*				71.29*		2.96
	X_D^2				8.00				2.50		110.06*
	Q_M				.00090				.00684*		.00093
	Q_D				7.61				25.54*		20.76*
	Q_L				.840				4.307*		3.450*

* Starred quantities are significant at the 5% level.

procedure due to George and Mudholkar studied in [2], [16].

Under the null hypothesis that the principal component model holds, c defined by (3.3) is normally distributed with zero mean and unit standard deviation. Let the one-tail probability associated with c_i for the ith observation be denoted p_i. Then Fisher's statistic becomes:

$$Q_0 = -2 \sum_{i=1}^{n} ln\, p_i \qquad (6.7)$$

which under the null hypothesis, has a χ^2-distribution with $2n$ degrees of freedom. The logit statistic is:

$$Q_L = \frac{1}{\pi} \sqrt{\frac{3(5n+4)}{n(5n+2)}} \sum_{i=1}^{n} \ln\left(\frac{1-p_i}{p_i}\right) \qquad (6.8)$$

which under the null hypothesis is adequately approximated by Student's t-distribution with $(5n+4)$ degrees of freedom for any n [2], [15]. Although the two procedures have comparable power properties [2], Fisher's method is sensitive when the departure from the null hypothesis is concentrated unevenly over the subgroup. The logit statistic, on the other hand, is more suitable for detecting departures distributed somewhat evenly over the subgroup. Another alternative, $\sum_{i=1}^{n} c_i^2$, while having the advantage of being extremely easy to use, having a χ^2-distribution with n degrees of freedom, does not have the power of the other alternatives.

If one were to use either of the alternatives to Q discussed in Section 5, the equivalent forms of (6.6) and (6.7) or (6.8) could be obtained in a similar manner.

7. NUMERICAL EXAMPLE FOR SUBGROUPS

The techniques discussed in Section 6 will be illustrated with three examples, shown in Table 5. In each case the sample size is $n = 3$. For each observation vector is given the first five principal components, $T^2 = \mathbf{y}'\mathbf{y}$, Q and the probability p_i associated with Q. In addition, for each group of observations are given the three generalized χ^2-statistics, Q_M and Q_0. The 5% levels for each of these statistics are also given. Set #1 is a case where the means have shifted but that is about all since the only significant quantities are χ^2_M and two of the principal component averages. For set #2, χ^2_M is again significant and χ^2_0 is not, which might lead one to believe that the process had shifted level but was still stable. However, both Q_M and Q_0 are also significant, which indicates that there are some consistent problems with the residuals in this sample and that the principal component model did not adequately characterize the process at the time this sample was obtained. Note, in particular, that this is not a general deterioration of the process because χ^2_D is still small indicating that whatever has happened to the process, it is a real shift, not the result of a single outlying observation or excessive variability within the sample.

By contrast, consider set #3. The first two observations are the same as those in set #2 but $\mathbf{x}_3$ has been contrived to bring the average of the sample close to standard. In doing so, both T^2 and Q for this third observation are extremely large. In combining these three observations, χ^2_M is not significant as we had intended and Q_M is also not significant which says that the principal component averages adequately describe the sample average of the original observations. Both χ^2_D and Q_0 are highly significant indicating that the process is unstable, the principal component model not holding for the individual observations and the within-sample variability being excessive besides. The residuals for these examples are shown in Table 6. Note that for these examples the results of analysis using Q_0 and Q_L are quite similar.

8. DISCUSSION AND SUMMARY

We have described a number of statistics to use as control tools. Some of these such as T^2 are special cases of general multivariate control situations and may be employed either with or without the use of principal components. The Q-statistics, on the other hand, are developed expressly to deal with residuals related to principal component analysis.

When one is making significance tests, certain assumptions are made. It is desirable to test these assumptions as well, since any indication of invalid assumptions will affect the credibility of the test. For this reason, it is better to test the assumptions first. For principal components, we appear to be well-equipped to do this, and we propose the following "priority list" for looking at these statistics:

For the case of an individual observation vector, one should first test Q. If Q is not significant, this indicates that the principal component model holds and T^2 should then be tested. If T^2 is not significant, the procedure is complete: the process may be assumed to be in control. If T^2 is significant, then the individual principal components should be tested to determine the nature of the disturbance (not the other way around; T^2 is the overall test and as long as this is in control we need not be concerned with the individual principal components with the exception of trends, etc. alluded to later.) If, on the other hand, Q is significant, the residuals should be investigated immediately since the principal component model does not hold. Our experience over the years has been that when Q is significant for an individual observation vector, it is generally an indication of a clerical error, bad test results, or similar type of error and Q is generally regarded as a "bad data" detector. For whatever reason, if Q is significant, doubt will be cast

on any of the other statistics as the examples in Section 3 pointed out.

For a sample of n observations, the order of testing should be: Q_0 or Q_L, χ^2_D, Q_M, and χ^2_M. Q_0 or Q_L and χ^2_D are primarily measures of process stability and should be examined before the tests for mean level are carried out, again, as a check on the assumptions. If Q_0 or Q_L are significant, it indicates that the data set does not fit the principal component model and one can go back to the individual values of Q to find whether this is a general condition or the result of one or more outliers. If the process passes this hurdle but fails on χ^2_D, it indicates that the individual observations fit the principal component model but the variability as excessive. An examination of the individual T^2's will indicate, again, whether or not this is the result of one or more outliers. Both of these tests are useful in screening individual data vectors. If Q_0 or Q_L and χ^2_D are not significant, Q_M is a possible indicator of the inadequacy of the principal component model overall. If all three of these are not significant, then χ^2_M can indicate the presence of a level shift in the process and if χ^2_M is significant, the principal component averages may then indicate the nature of the difficulty. We have never found the quantity χ^2_0 to be of much use. If this quantity is significant, it is an indicator that either χ^2_M or χ^2_D is significant if not both. On the other hand, it is a common occurrence for one of these measures to be significant without putting χ^2_0 out of control. The only time χ^2_0 will be significant by itself is if both χ^2_M and χ^2_D are close to being significant and this would seem to be its only utility.

For the control situation, it would be useful to check the individual principal components for trends and other cases of nonrandomness as an additional control tool in the same manner as is currently performed in standard univariate control chart procedures as tests for runs, cumulative sum charts, or geometric moving average charts.

The examples used here have related to control situations where we may reasonably assume that the covariance matrix is known so that the theory developed in this paper is applicable. For diagnostic work, most of these measures may still be used as descriptive devices even though their sampling distributions are unknown since the results given in this paper can serve as asymptotic results for the case Σ unknown. However, for diagnostic work, one should keep in mind that an outlier will affect not only one or more residuals but the characteristic vectors as well. When the sample size is small, in fact, the main effect of an outlier may be in the vectors and not in the residuals so the outlier may go undetected.

9. ACKNOWLEDGMENTS

We wish to acknowledge the assistance of Mr. F. Terry Hearne of the Eastman Kodak Company who performed many of the computations in this paper and who is often represented in the pronouns "we" and "our" when used in this paper in connection with remarks about experience.

Research partially sponsored by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-77-3360.

APPENDIX—THE DISTRIBUTION OF Q

In this appendix we describe the distribution of Q, the sums of squares of the residuals for a single vector observation after being fitted by the first k principal components. We first consider the case where all of the roots of Σ are positive and distinct and later consider the case where some roots are multiple or zero. Finally, we present an approximation to this distribution.

a. *Roots of Σ are positive and distinct.*

From (3.1) and (2.3) we obtain

$$\hat{\mathbf{x}} = \mathbf{U}_k\Lambda_k^{1/2}\mathbf{y} = \mathbf{U}_k\mathbf{U}_k'\mathbf{x}. \qquad (A.1)$$

This latter form allows us to rewrite (3.2) as

$$Q = (\mathbf{x} - \hat{\mathbf{x}})'(\mathbf{x} - \hat{\mathbf{x}}) = (\mathbf{x} - \mathbf{U}_k\mathbf{U}_k'\mathbf{x})'(\mathbf{x} - \mathbf{U}_k\mathbf{U}_k'\mathbf{x})$$
$$= \mathbf{x}'(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k')\mathbf{x} = \mathbf{x}'\mathbf{A}\mathbf{x}. \qquad (A.2)$$

TABLE 6—*Residuals for data in Table 5.*

		Set #1			Set #2			Set #3
		$\hat{x}_1$	$\hat{x}_2$	$\hat{x}_3$	$\hat{x}_1$	$\hat{x}_2$	$\hat{x}_3$	$\hat{x}_3$
Shoulder	Red	-.008	-.007	-.009	.028	.020	.034	-.022
	Green	-.002	.001	.002	-.012	-.008	-.015	.010
	Blue	-.006	.006	.008	-.015	-.010	-.016	.012
Middle-Tone	Red	-.006	.005	.006	-.025	-.018	-.031	.020
	Green	.001	-.001	0	.019	.014	.027	-.017
	Blue	.011	-.010	-.013	.025	.017	.025	-.017
Toe	Red	.011	0	-.004	.001	.004	-.002	.009
	Green	-.002	.009	.021	.003	.006	0	0
	Blue	-.005	-.010	-.018	.001	-.005	.007	-.011

$\mathbf{U}_k\mathbf{U}_k'$ is idempotent implying that $(\mathbf{I} - \mathbf{U}_k\mathbf{U}_k')\mathbf{U}_k\mathbf{U}_k' = 0$. Hence $\mathbf{A} = \mathbf{I} - \mathbf{U}_k\mathbf{U}_k'$ is an idempotent matrix of rank $(p - k)$ (see Rao [19], sec. 11.7) and will have $(p - k)$ characteristic roots equal to unity and k roots equal to zero. If we define

$$\mathbf{x}^* = \boldsymbol{\Sigma}^{-1/2}\mathbf{x} \tag{A.3}$$

where $\boldsymbol{\Sigma}^{-1/2} = \mathbf{U}\boldsymbol{\Lambda}^{-1/2}\mathbf{U}'$ then $\mathbf{x}^*$ has a p-variable normal distribution with mean zero and covariance matrix $\mathbf{I}$. Thus

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}^{*\prime}\mathbf{B}\mathbf{x}^* \tag{A.4}$$

where $\mathbf{B} = \boldsymbol{\Sigma}^{1/2}\mathbf{A}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma} - \mathbf{U}_k\boldsymbol{\Lambda}_k\mathbf{U}_k'$ has the same rank as $\mathbf{A}$, $(p - k)$. A simple orthogonal transformation then reduces Q to:

$$Q = \sum_{i=1}^{p-k} \lambda_{k+i} z_i^2 \tag{A.5}$$

which is a linear combination of independent χ_1^2-variables, z_1^2; hence, λ_{k+1} are the $(p - k)$ smallest roots of $\boldsymbol{\Sigma}$.

b. *Multiple and zero characteristic roots.*

The material in Sections 1 and 2 assume that the covariance matrix $\boldsymbol{\Sigma}$ was of full rank and that the characteristic roots were distinct—no multiple roots. We will investigate the two situations in which this assumption does not hold.

(1) *Multiple roots.* If two or more characteristic roots are equal, the vectors associated with them are not uniquely defined other than that they must form part of the orthogonal set and hence the results in both sections still hold. This situation is rather unlikely to happen in practice but this is fortunate since otherwise it would result in some of the principal components having little utility. (Not so rare is the case where two or more of the larger roots are close in size. Although the vectors associated with them are unique, the standard errors of their coefficients may be so large as to render these components of little use also [10].)

(2) *Zero roots.* It is possible, in practice, to have a situation where one or more of the characteristic roots are identically zero. This can happen when one or more of the original variables are linear combinations of some of the others. Suppose the rank of $\boldsymbol{\Sigma}$ was $m < p$ so that $(p - m)$ roots are equal to zero. So long as $k < m$ we may define:

$$\boldsymbol{\Lambda}^{-a} = \begin{bmatrix} \lambda_1^{-a} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2^{-a} & & 0 & 0 & \cdots & 0 \\ & & \ddots & \lambda_m^{-a} & 0 & \cdots & 0 \\ & & & & 0 & \cdots & 0 \\ & & & & & & 0 \end{bmatrix} \tag{A.6}$$

since $\boldsymbol{\Lambda}$ is, in fact, singular it can be shown that all of the relationships in Sections 1 and 2 hold, that $Q = \sum_{i=1}^{m-k} \lambda_{k+i} z_i^2$ and that the rank of $\mathbf{A}$ is $(m - k)$.

c. *Approximation for distribution of Q.*

There is a wealth of material related to approximation to the distribution of a sum of chi-square variates. However, for most practical situations, the following [12] should suffice:

Let $\theta_1 = \sum \lambda_i$, $\theta_2 = \sum \lambda_i^2$, $\theta_3 = \sum \lambda_i^3$, with all summations going from $k + 1$ to p (or $k + 1$ to m for $\sum$ singular), and let $h_0 = 1 - (2\theta_1\theta_3/3\theta_2^2)$. Then

$$(Q/\theta_1)^{h_0} \sim N\left[1 + \frac{\theta_2 h_0(h_0 - 1)}{\theta_1^2}, \frac{2\theta_2 h_0^2}{\theta_1^2}\right]. \tag{A.7}$$

This means that the probability distribution for Q can be approximated by:

$$\int_0^Q f(q)dq \approx \int_{-\infty}^{c} g(x)dx \tag{A.8}$$

where $g(x)$ is the normal density function and

$$c = \frac{\theta_1[(Q/\theta_1)^{h_0} - 1 - \theta_2 h_0(h_0 - 1)/\theta_1^2]}{\sqrt{2\theta_2 h_0^2}}. \tag{A.9}$$

Conversely, for a fixed type I error α, the upper limit for Q may be approximated by

$$Q_\alpha = \theta_1\left[\frac{c_\alpha\sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0 - 1)}{\theta_1^2}\right]^{\frac{1}{h_0}}. \tag{A.10}$$

The results in [12] indicate that this approximation is adequate for our purposes. For the exact distribution of Q, the reader should consult [6], [12], [14], [20] and [13], the last containing a computer program.

REFERENCES

[1] FISHER, R. A. (1932). *Statistical Methods for Research Workers,* Fourth Edition. Edinburgh: Oliver and Boyd.

[2] GEORGE, E. O. (1977). Combining independent one-sided and two-sided statistical tests; Some theory and applications. Ph.D. Thesis, University of Rochester.

[3] GNANADESIKAN, R. and KETTENRING, J. R. (1972). Robust estimates, residuals and outlier detection with multiresponse data. *Biometrics, 28,* 81-124.

[4] HAWKINS, D. M. (1974). The detection of errors in multivariate data using principal components. *J. Amer. Stat. Assoc., 69,* 340-344.

[5] HOTELLING, H. (1947). Multivariate quality control. *Techniques of Statistical Analysis* (ed. by Eisenhart, Hastay, and Wallis), pp. 111-184. New York: McGraw-Hill.

[6] IMHOF, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika, 48,* 419-426.

[7] JACKSON, J. E. (1959). Quality control methods for several related variables. *Technometrics, 1,* 359-377.

[8] JACKSON, J. E. (1979). Principal components and factor analysis. Submitted for publication.

[9] JACKSON, J. E. and BRADLEY, R. A. (1966). Sequential multivariate procedures for means with quality control applications. *Multivariate Analysis* (P. R. Krishnaiah, ed.), pp. 507-519. New York: Academic Press.

[10] JACKSON, J. E. and HEARNE, F. T. (1973). Relationships among coefficients of vectors used in principal components. *Technometrics, 15,* 601-610.

[11] JACKSON, J. E. and MORRIS, R. H. (1957). An application of multivariate quality control to photographic processing. *J. Amer. Stat. Assoc., 52,* 186-199.

[12] JENSEN, D. R. and SOLOMON, H. (1972). A Gaussian

approximation for the distribution of definite quadratic forms. *J. Amer. Stat. Assoc., 67,* 898–902.

[13] KOERTS, J. and ABRAHAMSE, A. P. J. (1969). *On the Theory and Application of the General Linear Model.* Rotterdam Univ. Press.

[14] KOTZ, S., JOHNSON, N. L. and BOYD, D. W. (1967). Series representation of distribution of quadratic forms in normal variables: I. Central case; II. Non-central case. *Ann. Math. Stat., 38,* 823–848.

[15] MUDHOLKAR, G. S. and CHAUBEY, Y. P. (1975). Use of logistic distribution for approximating probabilities and percentiles of student's distribution. *J. Statistical Research, 9,* 1–9.

[16] MUDHOLKAR, G. S. and GEORGE, E. O. (1978). The logit statistic for combining probabilities. Proc. International Symp. Optimization and Statistics (J. S. Rustagi, ed.). New York: Academic Press.

[17] OOSTERHOFF, J. (1969). *Combination of One-Sided Statistical Tests.* Mathematics Centre Tracts #28, Mathematisch Centrum, Amsterdam.

[18] RAO, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhya, A 26,* 329–358.

[19] RAO, C. R. (1973). *Linear Statistical Inference and its Applications,* Second edition. New York: John Wiley and Sons, Inc.

[20] ROBBINS, H. E. and PITMAN, E. J. G. (1949). Application of the method of mixtures to quadratic forms in normal variates. *Ann. Math. Stat., 20,* 552–560.