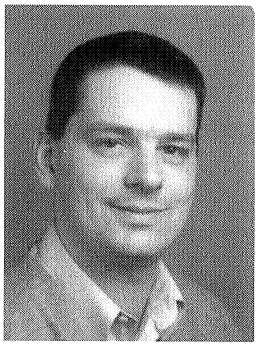
IN THIS ISSUE

VALUE-AT-RISK : AN APPLICATION OF QUANTILE ESTIMATION

Value-at-Risk (VaR) has become a standard for the assessement of market risks in the banking industry and represents the α-quantile of the profit-and-loss (P&L) distribution. Exact expressions of the P&L distributions are intractable in all but the trivial cases are generally obtained by the means of scenario-based methods. This paper reviews different sample-based quantile estimators and discusses the accuracy of the resulting estimates in cases of a small sample size.



Pierre-Yves Moix was born on December 8, 1965 in Sion, Switzerland. He received his degree in Business Administration from the University of St.Gallen in 1990 and his M.S. degree in Statistics from the University of Neuchâtel in 1998. Pierre-Yves Moix is currently writing a Ph.D. thesis about risk management at the University of St.Gallen and worked several years at the Union Bank of Switzerland and at Credit Suisse in the areas of portfolio management and derivatives pricing. He is holder of the CFA (Chartered Finacial Analyst) and FRM (Financial Risk Manager) designations.

EFFECT OF DESIGN ON ESTIMATES IN STUDIES USING A STRATIFIED, TWO-STAGE SAMPLE: COMPARISON OF DESIGN- AND MODEL-BASED ESTIMATORS

In this paper, Gerhard Gmel compares two types of estimator in a Survey of alcohol consumption in Switzerland. One estimator takes into account the design effects by including variables of the design in a statistical model, and the other is model-based and calculated by means of specialized sample survey software. Estimators of these two types are further compared with estimators that disregard the sampling design

Gerhard Gmel, Ph. D. is a senior research psychologist with experience in statistics and methodology, including survey methodology, test theory, and programme evaluation. He obtained his MS in Statistics from the Universsity of neuchâtel in 1999. Before joining the Swiss Institute for the Prevention of Alcohol and other Drug Problems, in 1994, he was associate professor of methodology and statistics at the Psychological Institute of the Technical University of Berlin. His interest is in large-scale general population surveys on alcohol, tobacco and drug use. He is currently principal investigator of the Swiss Health Surveys as regards licit and illicit drug use; the second of those Surveys was conducted in 1997.

MULTIPLE OUTLIER DETECTION IN LOGISTIC REGRESSION

In the third paper of this issue, Stéphan Munier suggests a robust method for the detection of atypical and influential observations in logistic regression. His technique is based on a forward search procedure which is a generalization of the method called Blocked Adaptative Computationnaly-efficient Outliers Nominators or BACON proposed by Hadi and Velleman (1998).

Stéphan Munier was born on May 5, 1971 in La Chaux-de-Fonds. He Obtained his Bachelor in Mathematics from the University of Neuchâtel in 1995 and his Master degree in statistics in 1999. Stéphan Munier is actually a Ph.D. Student at the Swiss Federal Institute of Technology, Lausanne. His current research is in the field of genetical statistics and in particular in the study of the human DNA sequences.

TED ANDERSON, THE MULTIVARIATE MAN

Professor Teodore W. Anderson was born in June 5, 1918. He dedicated over half century of his life to the statistical community and in particular in the area of multivariate data analysis. Dr. Anderson's An Introduction to Statistical Analysis which was published in 1954 became the reference to the field. The second edition was published in 1984 and currently he is working on the third edition. We are greatful to Ted who generously provided the readers od Student with many historical pictures, his hand writting and an article on the Stationarity of an Estimated Autoregressive process.

DATA & STATISTICS

M.M. Oliveira and J.T. Mexia provide the readers of Data & Statistics with a data set on AIDS in Portugal. The data set covers the period from 1984 to 1997 and gives per district different features related to AIDS. This data set can be found in its complete form on our web site : http://www.unine.ch/statistics/student/data/AIDS.htm

This issue of ***Student*** is dedicated to the memory of

Bernard Flury

(June 4, 1951 - July 6, 1999)

Bernard Flury, 48, died in a tragic accident on July 6, 1999 when he was hit by a falling boulder while hiking in the Dolomite Mountains near Trento, Italy. Born in Berne, Switzerland on June 4, 1951, Bernard began his career studying psychology, mathematics, and statistics at the University of Berne. In 1982 he earned a Ph.D. in statistics from the University of Berne under the direction of Hans Riedwyl. In 1987, he joined the Mathematics Department at Indiana University. He was appointed to Full Professor in 1995. For the academic year of 1989-90, Bernard was Professor at the Seminar für Statistik at the University of Fribourg, Switzerland.

Bernard had a wide area of research interests, primarily in the field of multivariate statistics. His book "Common Principal Components and Related Multivariate Models" (Wiley, 1988) gives a comprehensive review of principal components and the generalization to several groups. Dr. Flury worked on problems in many other areas such as discriminant analyses, finite mixtures, canonical correlation, principal points and self-consistency. He published over 70 research papers as well as two other books, "Multivariate Statistics : A Practical Approach" with Hans Riedwyl (1988 Chapman and Hall) and "A First Course in Multivariate Statistics" (1997 Springer). He was a Fellow of the American Statistical Association.

A warm and generous man, Bernard was a thoughtful and enthusiastic teacher who inspired many to pursue graduate work in statistics. He is survived by his wife Leah and his three sons Stefan, Andreas and Christoph.

In January 1999, Bernard joined the Statistics Group at the University of Neuchâtel, Switzerland as Professor of Applied Statistics, where he spent almost all his time with colleagues doing consulting for them.

We are dedicating to his memory this issue of Student for which he had a number of projects including a section which he named it *Fun with Statistics*, a sample of which is presented in this issue and we hope to continue his idea in future. We also include in this issue his article jointly written with his former student Beat Neuenschwander. Bernard will always be remembered for his wonderful sense of humor.

Student (1999), Vol. 3, No. 2, 117 - 126

Multiple outlier detection in Logistic Regression

Stéphan Munier

Swiss Federal Institute of Technology, Lausanne

Abstract : This article describes and illustrates a robust method for identifying multiple outliers in multivariate and regression data. The method called Blocked Adaptative Computationnaly-efficient Outlier Nominators or BACON serves as the basis for the investigation and we propose a generalisation of the BACON method to logistic regression data.

Key Words : Discrepancies, Logistic regression, Multivariate outliers, Outlier detection, Regression outliers, Residuals, Robust distance.

1 Introduction

Most statistical methods assume homogeneous data in which all observations satisfy the same model. However, real data often contain outliers. Robust methods relax the homogeneity assumption but they are often computationally infeasible for moderate to large size data. Outlier detection methods offer the advantage over robust methods of providing the analyst with a set of proposed outliers. These can be separated from the body of the data for additional analysis. The remaining data then more nearly satisfy homogeneity assumptions and can be safely analyzed with standard methods. Note that:

- Every robust fit generates an outlier detection method : the robust residuals of greatest magnitude are potential outliers.
- Every outlier nomination method generates a robust fitting technique: remove outliers and fit with standard methods.

There is a large litterature on outlier detection, see for example Cook and Weisberg (1982), Chatterjee and Hadi (1988) or Barnett and Lewis (1994). Al-

though the detection of isolated single outliers is now relatively standard. The more realistic situation in which there may be multiple outliers poses greater challenges. Blocked Adaptative Computationally-efficient Outlier Nominators or BACON is a promising outlier detection method proposed by Hadi and Velleman (1998). Section 2 presents the algorithm in a general form. It also describes the outline of the general algorithm as applied to multivariate and regression data proposed by Hadi and Velleman (1998). An illustrative example and results of a simulation-study are given in Section 3. Finally, Section 4 describes a generalized BACON algorithm as applied to logistic regression data.

2 The BACON algorithm

BACON starts by identifying a clean subset of the data than can be presumed free of outliers and then performs a "forward search". The principal improvements over other forward selection methods derive from allowing the subset of outlier-free points to grow rapidly, testing against a criterion and incorporating blocks of observations at each step. This saves computations both by reducing the number of covariance matrices computed and inverting.

2.1 The general BACON algorithm

Step 1 Identify an initial basic subset of size $m > p$ observations that we are confident does not contain outliers, where p is the dimension of the data.

Step 2 Fit an appropriate model to the basic subset, and from that model compute discrepancies for each of the observations.

Step 3 Expand the basic subset to hold observations known (by their discrepancies) to be homogeneous with the basic subset.

Step 4 Iterate to refine the basic subset using a stopping rule to determine when the basic subset can no longer grow safely.

Step 5 Nominate the observations excluded by the final basic subset as outliers.

Hadi and Velleman (1998) give two examples of data analysis situations in which BACON can be used for outlier detection in multivariate and regression data.

2.2 Outliers in multivariate data

For $\mathbf{X}$, a matrix of p variables and n observations, an initial basic subset $\mathbf{X}_b$ of $m = cp$ (c is a small integer chosen by the data analyst) is chosen. A set of observations nominated as outliers can be found with the discrepancies $d_i(\overline{\mathbf{x}}_b,\mathbf{S}_b) = \sqrt{(\mathbf{x}_i - \overline{\mathbf{x}}_b)'\mathbf{S}_b(\mathbf{x}_i - \overline{\mathbf{x}}_b)}$ $i = 1, ..., n$ ($\overline{\mathbf{x}}_b$ and $\mathbf{S}_b$ are the mean and the covariance matrix of the basic subset) and the stopping rule $c_{npr}\chi_{p,\alpha/n}$. The constant c_{npr} is a corrector factor and $\chi^2_{p,\alpha}$ is the $1 - \alpha$ percentile of the chi-square distribution with p degrees of freedom.

2.3 Outliers in regression data

For $\mathbf{X}$ an $n \times p$ matrix of covariate data and $\mathbf{y}$ an $n \times 1$ vector holding the response variable, an initial basic subset $\mathbf{X}_b, \mathbf{y}_b$ of size $m = cp$ is chosen. In this case, outliers can be nominated with the discrepancies $d_i(\mathbf{y}_b,\mathbf{X}_b) = \frac{\mathbf{y}_i - \mathbf{x}_i^t \widehat{\boldsymbol{\beta}}_b}{\widehat{\sigma}_b \sqrt{1 - \mathbf{x}_i^{t\prime}(\mathbf{X}_b\mathbf{X}_b)^{-1}\mathbf{x}_i}}$ if $\mathbf{x}_i \in \mathbf{X}_b$ and $d_i(\mathbf{y}_b,\mathbf{X}_b) = \frac{\mathbf{y}_i - \mathbf{x}_i^t \widehat{\boldsymbol{\beta}}_b}{\widehat{\sigma}_b \sqrt{1 - \mathbf{x}_i^{t\prime}(\mathbf{X}_b\mathbf{X}_b)^{-1}\mathbf{x}_i}}$ if $\mathbf{x}_i \notin \mathbf{X}_b$. In this formulas, $\widehat{\boldsymbol{\beta}}_b$ and $\widehat{\sigma}_b$ are the least squares estimates of β and σ based on the observations in the subset $\mathbf{y}_m$ and $\mathbf{X}_m$. For the stopping rule we can use $t_{(\alpha/2(r+1),r-p)}$, where $t_{(\alpha,r-p)}$ is the $1-\alpha$ percentile of the t-distribution with $r - p$ degrees of freedom, and r is the size of the current basic subset at each step.

3 Some notes about BACON

3.1 Example

Consider the following 12 pairs of numbers (Antille and El May 1992):

x	3	4	5	7	7	5	4	5	6	8	8	14
y	2	4	6	7	8	8	8	9	11	13	15	6

The LMS line for these pairs of points is given by $\widehat{y} = 9.1667 - 0.3333x$. Scatter plot of this data with the LMS line and the BACON line are shown in figure 1. LMS prefer a line fits with about 4 outliers whereas BACON identify observations 12 as an outlier and provide us a LS fit line without observation 12.

3.2 Simulations

Hadi and Velleman performed simulation experiments to compare BACON with other methods with regard to both performance and computational expense. To complement their study, we will examine BACON's behavior by varying the following parameters:

- n: number of observations
- p: number of variables
- c: a small integer that determines the size of the basic subset
- K: true number of outliers

Our experiment considers outlier detection in multivariate data, but the same conclusions apply to the regression situation. First, we check whether BACON is efficient for data that are free of outliers.

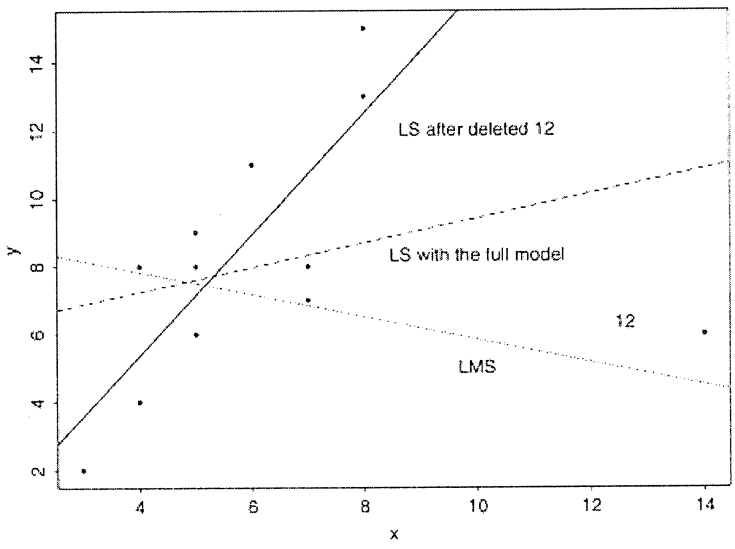


Figure 1 : Comparaison between LMS and BACON on an artificial dataset of size 12

3.2.1 Simulations with data free of outliers

We fix the number of variables $p = 5$ and take $n = 50, 100$ and 500 observations. For each value of n, we set the size of the basic subset $m = cp$ with $c = 3$. For each of these configurations, we generate 100 data sets each of size $n \times p$ from a $p-$dimensional standard Normal distribution. We chose $\alpha = 0.05$ for our outlier tests. We consider the following measure of performance:

$$Out = \text{Average proportion of observations detected as outliers.}$$

Thus, perfect performance corresponds to $Out = 0$. Table 1 gives the simulation results where the data are generated from $N(\mathbf{1}, \mathbf{I}_p)$ where $\mathbf{1}^t = (1, ..., 1)$ and $\mathbf{I}_p$ the identity matrix of order p.

As we can see in Table 1, for $n = 100$ and $p = 10, 20, 30$ BACON seems to be more and more efficient but this could also be a weakness of the algorithm.

Table 1: Simulations with data free of outliers

n	c	p	m	Out
50	3	5	15	0.04
100	3	5	15	0.03
500	3	5	15	0.03
100	3	10	30	0.03
100	3	20	60	0.01
100	3	30	90	0

3.2.2 Simulations with data contaminated with outliers

We fix the number of variables $p = 5$ and take $n = 50, 100$ and 500 observations. For each value of n, we set the size of the basic subset $m = cp$ with $c = 2, 3, 4, 5$. For each of these configurations, we generate 100 data sets each of size $n \times p$ from a $p-$dimensional standard Normal distribution. We then perturb 5% of the data points so that they are outliers in the sense of originating from a model with a different location. Thus $K = 3, 5, 25$. Thus, for each data set, $n - K$ observations are generated from the multivariate normal distribution $N(\mathbf{0}, \mathbf{I}_p)$. The K outliers are generated from $N(4 \times \mathbf{1}, \mathbf{I}_p)$. We chose $\alpha = 0.05$ for our outlier tests and consider the following measure of performance:

Out = Average proportion of observations detected as outliers
$TrueOut$ = Average proportion of added outliers correctly detected

Thus, perfect performance corresponds to $Out = TrueOut = 1$, indicated that all of the true outliers have been found, and no non-outlying observations have been declared to be outliers. Table 2 gives the simulation results where the outliers are generated from $N(4 \times \mathbf{1}, \mathbf{I}_p)$.

Table 2: The simulation results for the case where K outliers generated from $N(4 \times \mathbf{1}, \mathbf{I}_p)$ are added.

n	p	K	c	m	Out	$TrueOut$
50	5	3	2	10	1.850	1.000
			3	15	1.116	1.000
			4	20	0.996	0.980
			5	25	0.990	0.986
100	5	5	2	10	1.5	1.000
			3	15	1.004	1.000
			4	20	0.994	0.990
			5	25	0.994	0.982
500	5	25	2	10	1.0008	1.000
			3	15	1.0004	1.000
			4	20	1.002	1.000
			5	25	1.0012	1.000

Some notes about Table 2:

i. BACON seems to be more efficent for $c = 3$.

ii. The number of false positive identifications of non-outlying points as outliers remains small.

iii. BACON identifies all of the added outliers for $c = 2, 3$.

iv. BACON seems to be more efficient when $n \gg p$.

We performed additional simulations to check point (iv). We fix the number of observation $n = 100$ and take $p = 5, 10, 20, 50$ variables. For each value of p, the size of the basic subset is $m = cp$ with $c = 3$. For each of these configurations we generate 100 data sets each of size $n \times p$ from a $p-$dimensional standard Normal distribution. We then perturb $5, 10, 20, 30, 40\%$ of the data points then $K = 5, 10, 20, 30, 40$. Thus, for each data set, $n - K$ observations are generated from the multivariate normal distribution $N(\mathbf{0}, \mathbf{I}_p)$. The K planted outliers are generated from $N(4 \times \mathbf{1}, \mathbf{I}_p)$. We chose $\alpha = 0.05$ for our outlier tests and the previous measures of performance. Table 3 gives the simulation results.

Table 3: The simulation results for the case where K and p vary.

K	p	n	c	m	Out	$TrueOut$
5	5	100	3	15	1.016	1.08
10					1.001	1.02
20					0.981	1
30					0.951	1.03
40					0.93	1.04
5	10	100	3	30	0.988	1.04
10					0.983	1.03
20					0.953	1.06
30					0.900	1.01
40					0.921	1.07
5	20	100	3	60	0.926	1
10					0.907	1
20					0.821	1.04
30					0.818	1.01
40					0.779	1
5	30	100	3	90	0.588	1
10					0.44	1
20					0	0
30					0	0
40					0	0

Some notes about Table 3:

i. BACON rarely detects false outliers.

ii. For large values of p and K, BACON is less powerful.

iii. BACON is an efficient method for data that do not contain more than 20% of outliers.

4 BACON for logistic regression data

Consider the logistic model $\text{logit}(y_i) = \log\left(\frac{y_i}{1-y_i}\right) = \mathbf{x}_i^t\boldsymbol{\beta}+\epsilon_i$ with $\mathbf{y} = (y_1, ..., y_n)^t$ an $n \times 1$ binary response vector and $\mathbf{X}$ an $n \times p$ matrix of covariate data of rank $p < n$, $\boldsymbol{\beta}$ an $n \times 1$ vector of unknown parameters and $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)^t$ an $n \times 1$ vector of errors. We propose to generalized BACON to this model.

4.1 Algorithm for selecting an initial basic subset for logistic regression data

Step 0 Apply the multivariate algorithm to the $\mathbf{X}$ data (after removing the constant column, if any). Let $\mathbf{y}_m$ and $\mathbf{X}_m$ the set of m observations with the smallest values of the distances $d_i(\overline{\mathbf{x}}_b,\mathbf{S}_b)$. If $\mathbf{X}_m$ is not of full rank, increase the basic subset by adding observations with smallest values of $d_i(\overline{\mathbf{x}}_b,\mathbf{S}_b)$ until it has full rank. For $i = 1, ..., n$, compute

$$d_i^*(\mathbf{y}_m,\mathbf{X}_m) = \begin{cases} \dfrac{d_i}{\sqrt{1 - \mathbf{W}_m^{1/2}\mathbf{x}_i^t(\mathbf{X}_m^t\mathbf{W}_m\mathbf{X}_m)^{-1}\mathbf{W}_m^{1/2}\mathbf{x}_i}} & \text{if} \quad \mathbf{x}_i \in \mathbf{X}_m \\ \dfrac{d_i}{\sqrt{1 + \mathbf{W}_m^{1/2}\mathbf{x}_i^t(\mathbf{X}_m^t\mathbf{W}_m\mathbf{X}_m)^{-1}\mathbf{W}_m^{1/2}\mathbf{x}_i}} & \text{if} \quad \mathbf{x}_i \notin \mathbf{X}_m \end{cases}$$

with $d_i = \text{sign}(y_i - \widehat{y}_i)\sqrt{2}\left[y_i \log\left(\frac{y_i}{\widehat{y}_i}\right) + (1-y_i)\log\left(\frac{1-y_i}{1-\widehat{y}_i}\right)\right]^{1/2}$ the deviance residuals, $\widehat{y}_i = \exp(\mathbf{x}_i^t\widehat{\boldsymbol{\beta}}_m)/\left[1 + \exp(\mathbf{x}_i^t\widehat{\boldsymbol{\beta}}_m)\right]$, $\mathbf{W}_m = \text{diag}(\widehat{y}_i\,(1-\widehat{y}_i))$ (See Atkinson and Riani (1998) and Pregibon (1981)) and $\widehat{\boldsymbol{\beta}}_m$ the estimate of $\boldsymbol{\beta}_m$ based on the observations in the subset $\mathbf{y}_m$ and $\mathbf{X}_m$. Identify the $p+1$ observations smallest $|d_i^*(\mathbf{y}_b, \mathbf{X}_b)|$ and declare them to be the initial basic subset $\mathbf{y}_b$ et $\mathbf{X}_b$.

Step 1 If the new $\mathbf{X}_b$ is not of full rank, increase the subset by as many observa-

tions as needed for it to become full rank, adding observations with smallest $d_i^*(\mathbf{y}_b, \mathbf{X}_b)$ first, and increase m by the number of observations added to make the subset full rank. For $i = 1, ..., n$ compute

$$d_i^*(\mathbf{y}_b,\mathbf{X}_b) = \begin{cases} \dfrac{d_i}{\sqrt{1 - \mathbf{W}_b^{1/2}\mathbf{x}_i^t(\mathbf{X}_b^t\mathbf{W}_b\mathbf{X}_b)^{-1}\mathbf{W}_b^{1/2}\mathbf{x}_i}} & \text{if} \quad \mathbf{x}_i \in \mathbf{X}_b \\ \dfrac{d_i}{\sqrt{1 + \mathbf{W}_b^{1/2}\mathbf{x}_i^t(\mathbf{X}_b^t\mathbf{W}_b\mathbf{X}_b)^{-1}\mathbf{W}_b^{1/2}\mathbf{x}_i}} & \text{if} \quad \mathbf{x}_i \notin \mathbf{X}_b \end{cases}$$

Step 2 Let r be the size of the current basic subset. Identify the $r + 1$ observations with smallest $|d_i^*(\mathbf{y}_b, \mathbf{X}_b)|$ and declare them to be the new basic subset. (Note that these observations need not include all of the previous basic subset observations.)

Step 3 Repeat Steps 1 and 2 until the basic subset contains m observations.

This algorithm results in an initial basic subset of at least m observations that is free of outliers.

4.2 The BACON algorithm for logistic regression

Step 1 Use the previous algorithm to select an initial basic subset of size $m = cp$. The value of c is selected by the data analyst (usually 3, 4 or 5).

Step 2 Compute the discrepancies $d_i^*(\mathbf{y}_b, \mathbf{X}_b)$.

Step 3 The new basic subset consists of all points with discrepancies less than $t_{(\alpha/2(r+1),r-p)}$ where r is the size of the current subset.

Step 4 Stopping rule: Iterate Steps 2 et 3 until the size of the basic subset no longer changes.

Step 5 Nominate the observations excluded by the final basic subset as outliers

This algorithm results in a set of observations identified as outliers and the complementary set of homogenous observations.

4.3 Example

Consider the following data on vaso-constriction in the skin of the digits. The binary response y indicates the occurence (1) or nonoccurence (0) of vaso-constriction.

Table 4: Finney's Data on vaso-constriction.

Volume	Rate	Response	Volume	Rate	Response
3.7	0.825	1	1.8	1.8	1
3.5	1.09	1	0.4	2	0
1.25	2.5	1	0.95	1.36	0
0.75	1.5	1	1.35	1.35	0
0.8	3.2	1	1.5	1.36	0
0.7	3.5	1	1.6	1.78	1
0.6	0.75	0	0.6	1.5	0
1.1	1.7	0	1.8	1.5	1
0.9	0.75	0	0.95	1.9	0
0.9	0.45	0	1.9	0.95	1
0.8	0.57	0	1.6	0.4	0
0.55	2.75	0	2.7	0.75	1
0.6	3	0	2.35	0.03	0
1.4	2.33	1	1.1	1.83	0
0.75	3.75	1	1.1	2.2	1
2.3	1.64	1	1.2	2.0	1
3.2	1.6	1	0.8	3.33	1
0.85	1.415	1	0.95	1.9	0
1.7	1.06	0	0.75	1.9	0
			1.3	1.625	1

Source: Cox and Snell (1989)

We fit the model:

$$\text{logit}(p) = \beta_1 + \beta_2 \log(\text{RATE}) + \beta_3 \log(\text{VOLUME})$$

The estimated coefficients and their standard errors are:

$$\begin{aligned} \widehat{\beta}_1 &= -2.875 \ (1.319) \\ \widehat{\beta}_2 &= 5.179 \ (1.862) \\ \widehat{\beta}_3 &= 4.562 \ (1.835). \end{aligned}$$

The deviance for the fit is 29.23 on 36 degrees of freedom, and the corresponding chi-squared statistic is 34.15. Both are less than their asymptotic expectation of 36, indicating no gross inadequacies with the model. The odered components of χ^2 are plotted against standard normal quantiles in Figure 2. Evidently, two observations, the 4th and the 18th, are not well explained by the model. The logistic version of BACON nominates these observations as outliers.

Résumé : Cet article décrit et illustre un méthode robuste de détection de valeurs aberrantes pour des données multivariées et dans le cas du modèle de régression. Cette méthode appelée Blocked Adaptative Computationnally-efficient Outlier Nominator ou BACON dont les avantages sont un coût trés faible en terme calculs par rapport autres méthodes robustes et une résistance à l'effet de masque,

a été proposée par Hadi et Velleman (1998). Nous proposons de la généraliser au cas du modèle de régression logistique.

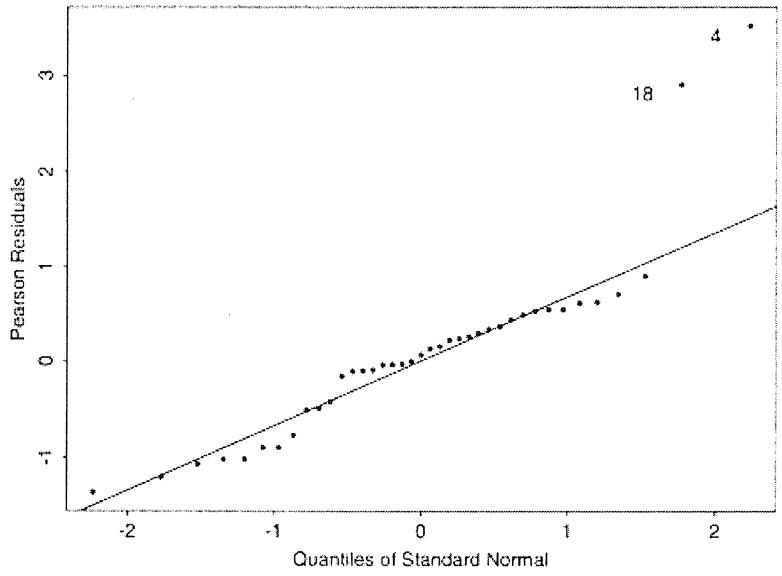


Figure 2 : Gaussian probability plot of the Pearson residuals calculated from Finney's data.

References

[1] Antille, G. and El May, H. (1992) The Use of Slices in the LMS and the Method od Density Slices: Foundation and Comparison. In Dodge, ed. *Computational Statistics*, Physica-Verlag, Heidelberg.

[2] Atkinson, A. and Riani, M. (1998) Regression Diagnostics for Binomial Data from the Forward Search. *Submitted for publication.*

[3] Barnett, V. and Lewis, T (1994) *Outliers in Statistical Data.* 3rd ed. John Wiley, New York.

[4] Chatterjee, S. and Hadi, A. S. (1988) *Sensitivity Analysis in Linear Regression*, John Wiley & Sons, New York.

[5] Cook, R. D. and Weisberg, S. (1982) *Residuals and Influence in Regression*, Chapman and Hall, London.

[6] Cox, D. R. and Snell, E. J. (1989) *Analysis of binary data*, 2nd edn, Chapman and Hall, London

[7] Hadi, A. D. and Velleman, F. (1996) BACON: Blocked Adaptative Computationally-efficient Outlier Nominators. *Submitted for publication.*

[8] Pregibon, D. (1981) Logistic regression Diagnostics. *The Annals of Statistics.* **9**, 705-24.

Student (1999), Vol. 3, No. 2, 128 - 130

Fun with statistics: Intransitive Dice

Bernard D. Flury

Statistics Group, University of Neuchâtel, Switzerland

The statistician Bradley Efron from Stanford University is most famous for bootstrap methods, but he has made many other important contributions to statistics and probability as well. Efron proposed the following game with four dice to highlight a seemingly paradoxical result from probability theory.

Take four regular fair dice and relabel their faces as follows, using the numbers 1 through 24:

die 1:	1	2	16	17	18	19
die 2:	3	4	5	20	21	22
die 3:	6	7	8	9	23	24
die 4:	10	11	12	13	14	15

Offer your friend the following game. After looking at all four dice she gets to choose one, and then you choose one from the remaining three. Both of you roll the selected dice simultaneously, and whoever has the larger number gets 1 Euro from the opponent. Clearly, your friend has an advantage as she is allowed to choose first, and therefore (hopefully) identifies the best of the four dice before the game starts. That means, you offer her an advantage.

Well, not quite. Suppose that your friend chose die # 1. You will then select # 2 for yourself. As both dice are fair and independent, there are 36 equally likely outcomes, among which 24 are in your favor, i.e.,

$$\Pr[\text{die 2 beats die 1}] = 2/3.$$

Of course after a few experiments your friend will figure out that she is really at a disadvantage, and so you generously offer her to take die # 2. For yourself,

you now select die # 3. A little calculation shows that

$$\Pr[\text{die 3 beats die 2}] = 2/3.$$

Similarly,

$$\Pr[\text{die 4 beats die 3}] = 2/3,$$

and

$$\Pr[\text{die 1 beats die 4}] = 2/3.$$

Thus, whatever die your friend has chosen, you can always select one to beat her with probability 2/3 !

This is quite contradictory to what one would naively expect. Writing "$i < j$" if the probability of die j to beat die i is larger than 1/2, we have just seen that

$$1 < 2 < 3 < 4 < 1,$$

that is, the relation "$<$" is not transitive. There is simply no such thing as a "best die" among the four.

When you inspect the four dice carefully, you will notice that, for instance, the exact number shown by die # 4 is totally irrelevant as the six numbers on the faces of die # 4 are consecutive. Similarly, it is irrelevant if die # 1 shows a 1 or a 2. Thus, the following set of dice will have exactly the same properties:

die 1:	1	1	5	5	5	5
die 2:	2	2	2	6	6	6
die 3:	3	3	3	3	7	7
die 4:	4	4	4	4	4	4

In this version it is quite obvious that the result of die # 4 is irrelevant, and it is also easier to see upon short inspection that, for instance, die # 1 is better than die # 4 but worse than die # 2. Thus it is better for you to use the original version as it hides the secret more carefully. Of course many other ways to label the faces can be found, maintaining the same winning probability of 2/3.

What about three dice only? It is possible to construct a set of three intransitive six–sided dice, but then at least one of the winning probabilities is at most 5/9, which is not quite as drastic, and may take a large number of experiments to show up. An example is given by the following three dice:

die 1:	1	2	13	14	15	16
die 2:	3	4	5	6	17	18
die 3:	7	8	9	10	11	12

In this example, die 2 beats die 1 with probability 5/9, die 3 beats die 2 with probability 2/3, and die 1 beats die 3 with probability 2/3.

With a winning probability of 2/3 it takes usually only a few rolls for your friend to realize that she is at a disadvantage. For instance, after just 5 rolls

the probability that you are ahead of your opponent is $192/241 \approx 4/5$, and more than 10 experiments are hardly ever needed to convince her completely that she has not chosen well.

This game with dice has a popular non–stochastic analog: the "scissors–paper–rock" game. It is played as follows. Two children secretly choose one of the three objects scissors, paper, and rock, symbolyzing the chosen object by two fingers spread apart (scissors), a flat hand (paper), or a fist (rock), hidden behind their backs. They say simultaneously "scissors, rock, paper", and reveal their respective choice by showing their hands. If both children chose the same object, it's a tie. Otherwise, scissors win against paper (cutting), paper wins against rock (covering), and rock wins against scissors (beating). If one of the children happens to know the other child's choice ahead of time, then this is obviously an unfair game. Just as in our game with dice, it is the second player (the one who has the information about the first player's choice) who is better off.

Further reading

Finnell, G. (1971), "A Dice Paradox," *Epilogue*, July 1971, 2–3.

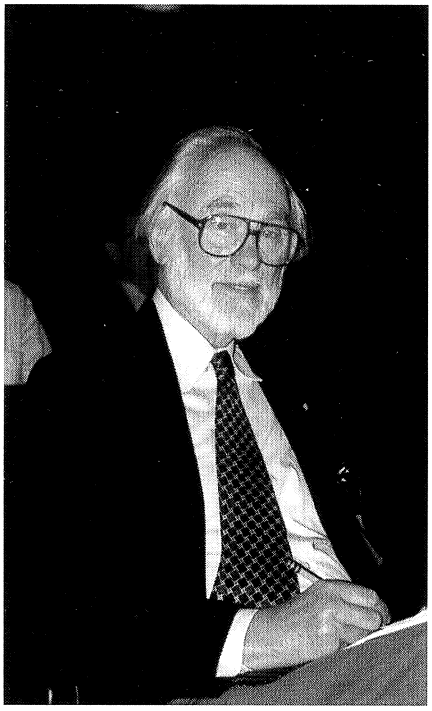
Gardner, M. (1974), "Mathematical Games," *Scientific American*, October 1974.

Gardner, M. (1983), *Wheels, Life, and Other Mathematical Amusements*, W.H. Freeman and Company, New York.

Tenney, R., and Foster, C. (1976), "Nontransitive Dominance," *Mathematics Magazine*, **49**, May–June 1976, 115–120.

Student (1999), Vol.3, No.2, 133 - 140

Ted W. Anderson, The Multivariate Man



Ted W. Anderson was born in Minneapolis, Minnesota on June, 5 1918. He recieved his Ph.D in Mathematics from Princeton University in 1945. In 1945-46 he worked as Research Associate in the Cowles Commission for Research in Economics at the University of Chicago. From 1946 he was member of the department of Mathematical Statistics at Columbia Universisty where he became professor in 1956. In 1967 he was named Professor of Statistics and Economics at Stanford University. He was a fellow of Guggenheim, in 1947-1948, editor of the *Annals of Mathematical Statistics* from 1950 to 1952, president of the *Institute of Mathematical Statistics* in 1963 and vice president of the *American Statistical Association* from 1971 to 1973.

He is fellow of the *American Academy of Arts and Sciences*, of the *National Academy of Sciences,* of the *Institute of Mathematical Statistics* and of the *Royal Statistical Society.* Among other, T.W. Anderson is the author of *Statistical Analysis of Time Series*, *A Bibliography of Multivariate Statistical Analysis* and *An Introduction to the Statistical Analysis of Data.*

T.W. Anderson published several papers in the field of discriminant functions and tests of rank. In 1954, he and Donald Darling developed a test for goodness of fit that bears his name: the Anderson-Darling goodness of fit test.

In the field of econometrics, he developed the limited information maximum likelihood method for estimating coefficients of a single equation. This led to the development of the two-stage least-squares procedure. Also in the field of econometrics, he wrote on distributions of estimators in simultaneous-

equation models and asymptotic expansions of them and published 1971 *The Statistical Analysis of Time Series.*

But Anderson's best known contributions to statistics are in the field of multivariate analysis. In 1958, he published *An Introduction to Multivariate Statistical Analysis*, that was translated into Russian. This book has been a reference in this area of statistics for more than forty years and T.W. Anderson published the second edition in 1984.

Since the beginning of the 80's, he wrote books and articles on the statistical analysis of data. He was influenced by the development of the computer tools and wrote a guide to the Statistical Analysis of Data with MINITAB.

Dorothy and Ted Anderson, Tampere, Finland, 1999.

T.W. Anderson and his family

CANONICAL ANALYSIS AND REDUCED RANK REGRESSION FOR AUTOREGRESSIVE PROCESSE

T. W. ANDERSON

STANFORD UNIVERSITY

INTRODUCTION

Canonical Correlation Analysis

Y, X random vectors
$k\times 1$, $q\times 1$

$EY = 0$, $EX = 0$

$$E\begin{bmatrix} Y \\ X \end{bmatrix}[Y', X'] = \begin{bmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{bmatrix}$$

Canonical Correlations, Variable

$$\begin{bmatrix} -\rho\Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & -\rho\Sigma_{XX} \end{bmatrix}\begin{bmatrix} \alpha \\ \omega \end{bmatrix} = 0$$

$$\alpha'\Sigma_{YY}\alpha = 1 = \omega'\Sigma_{XX}\omega$$

Solutions:

$\rho_1 \geq \ldots \geq \rho_k \geq 0 \geq -\rho_k \geq \ldots \geq -\rho_1 \quad (k \leq q)$

$\alpha_1, \ldots, \alpha_k, \quad \omega_1, \ldots, \omega_q$

Derived equations:

$$\Sigma_{YX} \Sigma_{XX}^{-1} \Sigma_{XY} \alpha = \rho^2 \Sigma_{YY} \alpha$$

$$\Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{YX} \omega = \rho^2 \Sigma_{XX} \omega$$

$U_1 = \alpha_1' Y, \quad V_1 = \omega' X$

$\rho_1 = \rho(U_1, V_1)$

$= \max \rho(\omega' Y, \omega' X)$

Hotelling (1936)

Regression Model

$Y = BX + Z$, Z unobserved

$\Sigma_{XZ} = 0, \quad \Sigma_{ZZ}$

Then

$$\Sigma_{YY} = B \Sigma_{XX} B' + \Sigma_{ZZ}$$

$$\Sigma_{YX} = B \Sigma_{XX}$$

Alternative

$$B \Sigma_{XX} B' \phi = \theta \Sigma_{ZZ} \phi$$

$$\phi' \Sigma_{ZZ} \phi = 1$$

Then

$\theta = \frac{\rho^2}{1-\rho^2}, \quad \phi = \alpha (1-\rho^2)^{-\frac{1}{2}}$

Sample:

$$y_1, x_1, y_2, x_2, \ldots, y_N, x_N$$

Define

$$S_{YY} = \frac{1}{N} \sum_{t=1}^{N} y_t y_t'$$

$$S_{YX}, S_{XX}, \text{etc}$$

$$\begin{bmatrix} -r S_{YY} & S_{YX} \\ S_{XY} & -r S_{XX} \end{bmatrix} \begin{bmatrix} a \\ w \end{bmatrix} = 0$$

$$a' S_{YY} a = 1 = w' S_{XX} w$$

Solutions:

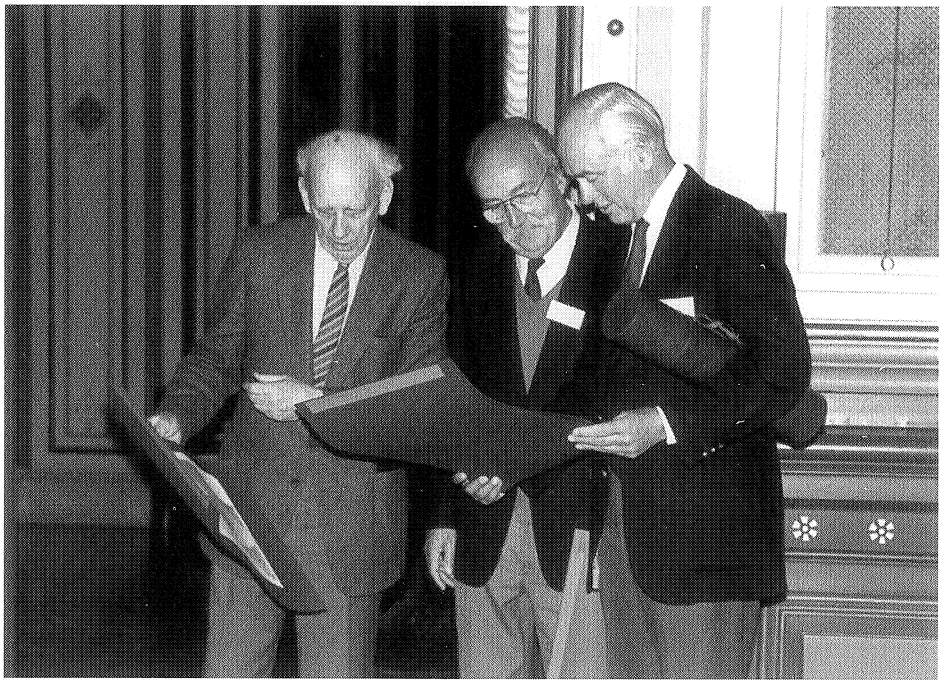
$$r_1 > \ldots > r_k > -r_k > \ldots > -r_1$$

$$a_1, \ldots, a_k \qquad w_1, \ldots, w_q$$

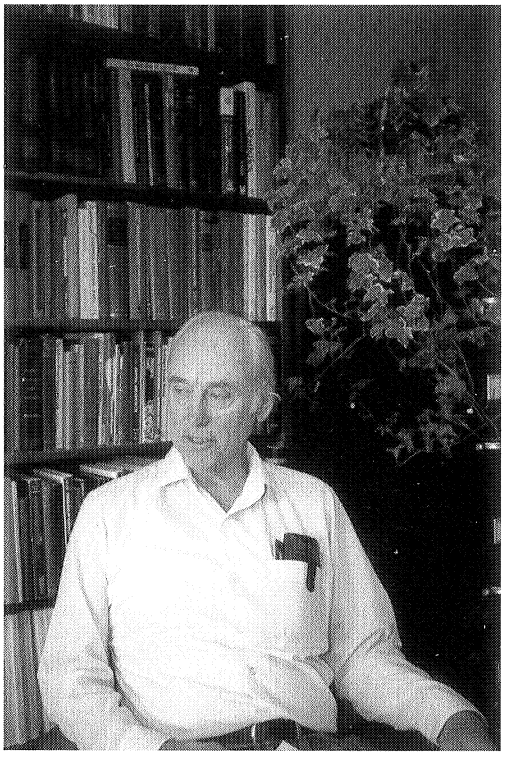
Harald Cramér and T.W. Anderson, Stockholm, 1947

Ted Anderson with David Blackwell, Churchill Eisenhart and Erich Lehmann

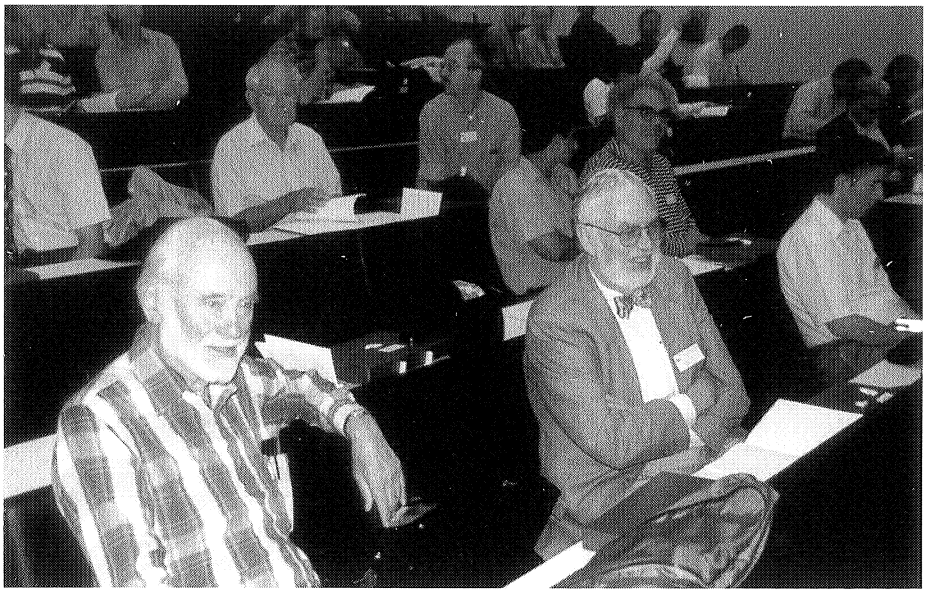
T.W Anderson with his wife Dorothy and Maurice Kendall, Athens 1972



E.J. Hannan, G.E.P. Box and T.W. Anderson, Tampere, 1987



Ted Anderson in his office



ISI satellite meeting, Tampere 1999

Student (1999), Vol.3, No.1, 141 - 143

The Stationarity of an Estimated Autoregressive Process

T.W. Anderson

Departement of Statistics, Stanford University, California

Abstract : It is shown that the estimators of the coefficients of an autoregressive process define a stationary process if the estimators constitute the solution to the sample Yule-Walker equations.

Key Words : autoregressive process, Yule-Walker equations, stationarity.

A stationary stochastic process $\{y_t\}$ with mean $Ey_t = 0$ satisfies a stochastic difference equation if there exist constants $\beta_0 = 1, \beta_1, ..., \beta_k$ such that $\{u_t\}$ defined by

$$\sum_{r=0}^{k} \beta_r y_{t-r} = u_t, \qquad t = ..., -1, 0, 1, ..., \tag{1}$$

consists of independently identically distributed random variables. The process $\{y_t\}$ is stationary and y_t is independent of $u_{t+1}, u_{t+2}, ...$ if and only if the roots of the associated polynomial equation

$$\sum_{r=0}^{p} \beta_r x^{p-r} = 0 \tag{2}$$

are less than 1 in absolute value. The process is autoregressive of order p. We assume $Eu_t = 0$ and $Eu_t^2 = \sigma^2$ with $0 < o^2 < \infty$.

Let $y_1, ..., y_T$ be T successive observations on the process. To estimate the coefficients $\beta_1, ..., \beta_p$ one can solve the linear equations

$$\sum_{j=1}^{p} c_{i-j} b_j = -c_i \qquad i = 1, 2, ..., p, \tag{3}$$

where

$$c_i = c_{-i} = \frac{1}{T} \sum_{t=1}^{T-i} y_{t+i} y_t, \qquad i = 0, 1, ..., p. \tag{4}$$

These are the sample Yule-Walker equations. See, for example, Section 5.4 of T. W. Anderson (1971). We assume that there are at least p different nonzero values of y_t observed. The purpose of this note is to show that the solution of (3) yields coefficients corresponding to a stationary process; that is, the roots of

$$\sum_{r=0}^{p} b_r x^{p-r} = 0 \tag{5}$$

($b_0 = 1$) are less than 1 in absolute value. Pagano (1971) has shown this result by a different method.

Let $y_{-p+1} = y_{-p+2} = \ldots = y_0 = 0$ and $y_{T+1} = y_{T+2} = \ldots = y_{T+p} = 0$. Define the vectors

$$\mathbf{y}_t = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix}, \qquad t = 0, 1, \ldots, T+p. \tag{6}$$

The equations (3) can be written

$$\mathbf{b}' \sum_{t=1}^{T+p} \mathbf{y}_t \mathbf{y}_t' = -\sum_{t=1}^{T+p} \mathbf{y}_t \mathbf{y}_{t-1}'. \tag{7}$$

The equation (7) is the first row of

$$\mathbf{B} \sum_{t=1}^{T+p} \mathbf{y}_t \mathbf{y}_t' = -\sum_{t=1}^{T+p} \mathbf{y}_t \mathbf{y}_{t-1}' \tag{8}$$

and $\mathbf{b}'$ is the first row of $\mathbf{B}$. The other $p-1$ rows of $\mathbf{B}$ constitute the matrix $(\ -\mathbf{I}\ \ \mathbf{0}\ )$.

Theorem 1 *The matrix* $\mathbf{B}$ *defined by (8) has characteristic roots less than 1 in absolute value.*

Proof. If $\mathbf{u}$ is a left-sided characteristic vector of $\mathbf{B}$ corresponding to a characteristic root λ (that is, $\mathbf{u}'\mathbf{B} = \lambda\mathbf{u}'$),

$$\lambda \mathbf{u}' \sum_{t=1}^{T+p} \mathbf{y}_t \mathbf{y}_t' = -\mathbf{u}' \sum_{t=1}^{T+p} \mathbf{y}_t \mathbf{y}_{t-1}'. \tag{9}$$

Normalize $\mathbf{u}$ so that

$$\begin{aligned} 1 &= \mathbf{u}' \sum_{t=1}^{T+p} \mathbf{y}_t \mathbf{y}_t' \overline{\mathbf{u}} = \sum_{t=1}^{T+p} (\mathbf{u}'\mathbf{y}_t)\,\overline{(\mathbf{u}'\mathbf{y}_t)} \\ &= \mathbf{u}' \sum_{t=1}^{T+p} \mathbf{y}_{t-1} \mathbf{y}_{t-1}' \overline{\mathbf{u}} = \sum_{t=1}^{T+p} (\mathbf{u}'\mathbf{y}_{t-1})\,\overline{(\mathbf{u}'\mathbf{y}_{t-1})}, \end{aligned} \tag{10}$$

where $\overline{\mathbf{u}}$ is the complex conjugate of $\mathbf{u}$. Then multiplication of (9) on the right by $\overline{\mathbf{u}}$ gives

$$\lambda = -\mathbf{u}' \sum_{t=1}^{T+p} \mathbf{y}_t \mathbf{y}'_{t-1} \overline{\mathbf{u}} = - \sum_{t=1}^{T+p} (\mathbf{u}'\mathbf{y}_t) \overline{(\mathbf{u}'\mathbf{y}_{t-1})} \,. \tag{11}$$

By the Cauchy-Schwarz Inequality $|\lambda| \leq 1$. We can have equality only if $\mathbf{u}'\mathbf{y}_t = \mathbf{u}'\mathbf{y}_{t-1}$, $t = 1, ..., T+p$, which is impossible because of (10). Q.E.D.

Since the characteristic roots of $\mathbf{B}$ are the roots of (5), the desired result has been proved. (See Section 5.4. of T. W. Anderson (1971)).

Theorem 2 *The roots of (5), where $b_1, ..., b_p$ are the solution to (3), are less than 1 in absolute value.*

The algebra used to prove Theorems 1 and 2 can be used to prove one of the statements in the first paragraph. Suppose $\{y_t\}$ is a stochastic process stationary in the wide sense; that is $Ey_{t+i}y_t = Ey_{s+i}y_s$, $\quad t, s = ..., -1, 0, 1, ...$.
Define

$$\gamma_i = Ey_{t+i}y_t, \qquad i = ..., -1, 0, 1, ... \,. \tag{12}$$

Define $\boldsymbol{\beta}^{(p)} = (\beta_1 \, , ..., \beta_p)'$,where $\beta_{1,...,}\beta_p$ is the solution to

$$\sum_{j=1}^{p} \gamma_{i-j}\beta_j = -\gamma_{i,} \qquad i = 1, ..., p. \tag{13}$$

Suppose that for some p, u_t defined by (1) with $\beta_0 = 1$ is independent of y_{t-1}, y_{t-2} , ... Then the roots of (2) are less than 1 in absolute value.

The proof of this statement is left to the reader.

References

[1] Anderson, T.W. (1971), The Statistical Analysis of Time Series, John Wiley & Sons, Inc.

[2] Pagano, Marcello (1971), "When is an Autoregressive Scheme Stationary ?", Research Report No. 53, Department of Statistics, State University of New York at Buffalo.

Data & Statistics in *Student (1999), Vol.3. No.2, 147 - 150*

AIDS in Portugal

M.M. Oliveira

Departamento de Matemática, Universidade de Évora, Portugal

J.T. Mexia

Departamento de Matemática, Universidade Nova de Lisboa, Portugal

Abstract: The data set covers the period from 1984 to 1997 and gives, per district, the relevant features of AIDS in Portugal. The STATIS[1] method is particulary useful to analyse this kind of data since it is organised in tables, one for each year. Moreover F tests may be used to detect a significant common structure between the studies in the series. This common structure may be expressed by a structure vector whose components constitute a time series which was analysed though the hidden periods techniques.

Key words: STATIS, F tests, Common Structure, Hidden Periods.

1 Introduction

Given a sequence of objects x variables matrices taken at sucessive times it may be important to condense the information into a time series. We now show how to use STATIS in order to perform such condensation for the data on AIDS in Portugal (1984 to 1997). This data was provided by the "Center for Epidemiological Vigilance of Transmissive Deseases".

2 The Scientific Context

For each year we have a study, this is a two-way table. In these tables the lines correspond to the Portuguese districts and the columns to the variables:

[1] Structuration des Tableaux à trois Indices de la Statistique

AIDS INDICATOR DISEASE		VIRUS TYPE	
Opportunist Infection	OPIN	HIV1	HIV1
Kaposi's Sarcoma	KAPS	HIV2	HIV2
Opportunist Infection + Kaposi's Sarcoma	OIKS	**AGE GROUPE**	
Lymphoma	LYPH	0 - 11 months	A
Encephalopathy	ENCE	1 - 4 years	B
Emaciation Syndrome	EMAS	5 - 9 years	C
PILL	PILL	10 - 12 years	D
CICU	CICU	13 - 14 years	E
TRANSMISSION CATEGORY		15 - 19 yeras	F
Homosexual or Bisexual	HOBI	20 - 24 years	G
Drug Addiction	DRAD	25 - 29 years	H
Hemophilia	HEMO	30 - 34 years	I
Heterosexual	HTRO	35 - 39 years	J
Transfusions	TRNF	40 - 44 years	L
Mother/Child Transmission	MCTR	45 - 49 yeras	M
Unknown	UNKN	55 - 59 years	O
SEX		60 - 64 years	P
Female	FEMA	65 - 90 years	Q
Male	MALE	Unknown	R

To jointly analyze these studies the STATIS method, see Lavit (1988), may be used. In this method a linear operator is associated to each study. The relations between the studies are expressed by the HILBERT-SCHMIDT products between those operators. When there is a common structure between the studies the matrix of these products is the sum of a rank one $\lambda \overrightarrow{\alpha}\,\overrightarrow{\alpha}^t$ matrix with an error matrix. F tests for the existence of this common structure have been derived, see Oliveira & Mexia (1998). When a common structure exists it may be expressed by the structure vector $\lambda \overrightarrow{\alpha}$. The components of this vector correspond to the matrices and, if these refer to sucessive years, they constitute a time-series, which may be worthwhile to analyze.

3 The Data

The data are presented on the internet on page:
http://www.unine.ch/statistics/student/data/AIDS.htm

4 Main Results

In table one we present the HILBERT-SCHMIDT products

220.20													
163.58	246.88												
217.62	294.99	408.04											
158.88	219.29	280.50	240.89										
103.28	125.83	158.70	139.12	151.68									
125.78	138.40	170.04	140.78	113.75	142.35								
99.56	102.89	114.66	98.14	31.23	89.73	127.30							
73.77	96.57	99.05	92.09	80.60	82.11	89.27	116.39						
123.75	139.35	173.25	138.60	96.82	110.28	82.37	73.09	139.93					
54.78	54.04	62.39	72.19	74.93	63.43	62.09	64.69	84.24	177.62				
154.37	180.85	231.72	176.47	111.16	123.45	89.45	78.12	120.88	53.91	182.63			
152.15	161.68	203.83	160.64	115.93	120.40	98.38	83.38	120.44	70.09	140.87	171.53		
170.09	229.81	303.99	221.18	126.64	137.48	101.21	89.33	138.53	58.13	188.21	166.89	271.95	
62.02	61.32	54.50	57.79	55.68	51.27	53.66	84.79	45.00	42.86	50.67	58.73	55.37	245.96

Using the F tests we quoted alove a significant common structure was found to exist. The components of the structure vector were:

Year	1984	1985	1986	1987	1988	1989	1990
Coord. First Axis	12.15	14.17	18.76	14.47	9.60	10.17	7.84
Year	1991	1992	1993	1994	1995	1996	1997
Coord. First Axis	7.06	10.03	5.48	12.35	11.67	15.09	5.08

And the chronomgram was:

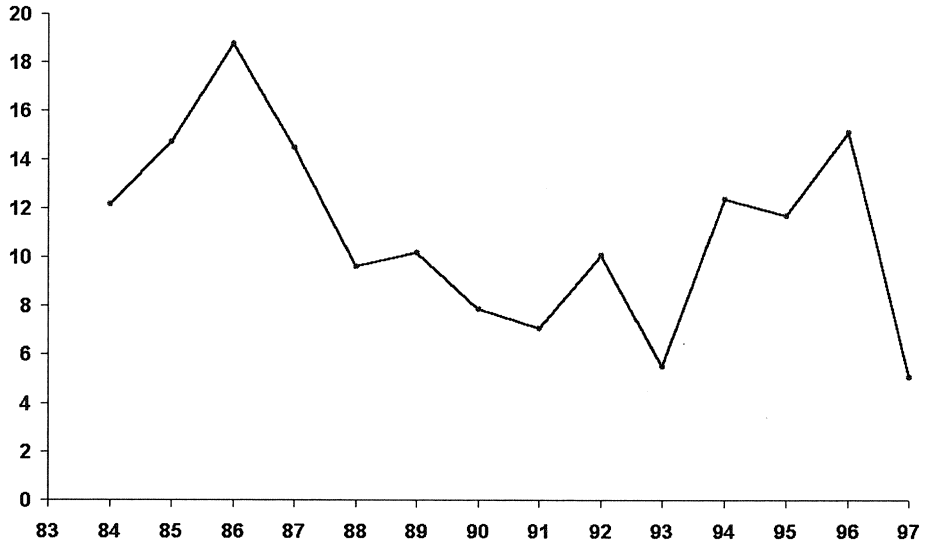


Figure 1: Chronogram

This chronogram suggested using the hidden periodicity model, see Grenander & Rosenblatt (1975, pg. 91 to 94).

We thus used the last 13 years to adjust the model:

$$\begin{aligned} Y_t &= 10.94571 - 3.34127\cos\left(\frac{2\pi t}{13}\right) - 2.366508\sin\left(\frac{2\pi t}{13}\right) \\ &\quad -3.340611\cos\left(\frac{6\pi t}{13}\right) + 0.322584\sin\left(\frac{6\pi t}{13}\right) \\ &\quad +3.735936\cos\left(\frac{12\pi t}{13}\right) + 1.585178\sin\left(\frac{12\pi t}{13}\right). \end{aligned}$$

In the following graph we placed the adjusted values as abscises and the observed ones ordinates:

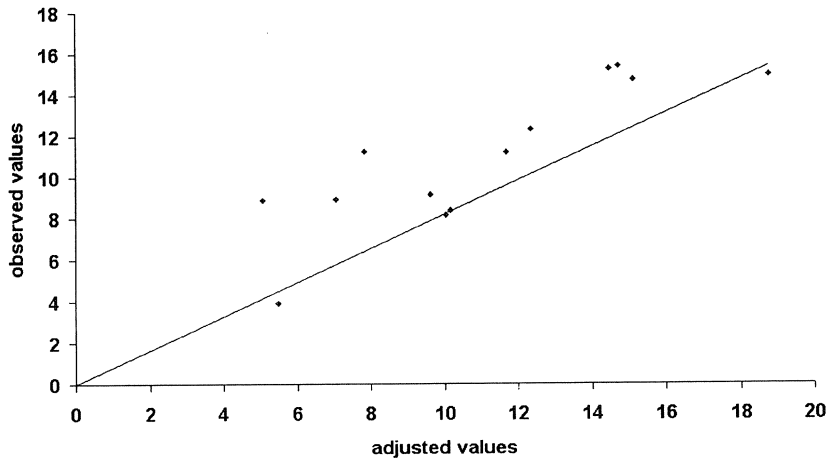


Figure 2: Adjusted and observed values

This graph shows an acceptable adjustment. We are thus led to assume the existence of cyclic components with periods 13.0, 4.33 and 2.16 years. The existence of these cycles points towards a certain co-existence between Man and AIDS in Portugal.

References

[1] Grenander and Rosenblatt (1975), *Statistical Analysis of Stationary Time Series*, John Wiley, New York

[2] Lavit C. (1988), *Analyse Conjointe de Tableaux Quantitatifs*, Collection Méthodes et Programmes, Masson, Paris

[3] Oliveira M.M. and J.T. Mexia (1998) *Tests for the rank of Hilbert-Schmidt product matrices*, Advances in Data Science and Classifications. (Rizzi, Vichi and Bock, Ed.), pp 619-625, Springer