IBM **Data Processing Application** C20-1645-0

Outlines of Statistical Techniques, Applications, and Programs for Industry, Engineering and Science

This manual outlines nine statistical techniques, giving simple definitions and examples, a summary of input and output, and references to numerous applications and computer programs. The techniques covered are: correlation, factor analysis, cluster analysis, regression, discriminant analysis, experimental analysis, evolutionary operation, Bayes formula, and time series analysis.

CONTENTS

Copies of this and other IBM publications can be obtained through IBM branch offices. Address comments concerning the contents of this publication to IBM, Technical Publications Department, 112 East Post Road, White Plains, N.Y. 10601

INTRODUCTION

We need the quantitative methods of statistics both to clarify and to solve problems in every science, engineering discipline, and industrial enterprise.

This manual spotlights nine statistical techniques. They are general-purpose tools and attack problems in which many variables or factors operate simultaneously. They thrive where data are highly variable and where no neat, determinate mathematical model is known. Widely divergent groups — behavioral scientists, electrical engineers, steel manufacturers — use these techniques.

Computers handle statistical techniques with great speed and accuracy. Computers process many measurements on a great number of factors; allow easy experimentation and analysis.

We cover the nine techniques in a short, simple manner. A thorough familiarity with statistics is not required to read this manual. The essential feature of a technique, the wide applicability, the perspective — these are made plain without cautions, hedgings, assumptions, or mathematical precision. Statistics is not taught here, or computer programming, or the subject matter implied in the illustrations (whether in geology, aeronautical engineering, or petroleum refining). However, for each technique area this manual:

- gives a simple definition
- illustrates its application
- tells the type of data you begin with (inputs) and what answers you end with (outputs)
- references numerous applications
- lists some available computer programs.

Remember in each of the nine sections to follow that the statistical technique is only loosely defined. Generally, a technique has to satisfy many conditions to be validly used, and is not infallible in effect. However, some techniques are now controlling huge industrial operations. Others are providing researchers new insights into what were once complicated or puzzling situations.

Three kinds of listings accompany the majority of the techniques:

- a file of applications with references
- a list of computer programs in the "IBM Catalog of Programs" series
- a list of references which cite the use of IBM Systems in carrying out a technique.

The first item above covers applications and examples, some tutorial, some in actual operation.

The second item cites computer programs for the techniques covered, as well as for related methods. Each computer system has a program library contributed to by IBM customers or IBM personnel. The so-called Type I and Type II programs are supported by documentation and test procedures. IBM serves solely as the distribution agent for Type III and Type IV programs. More details on each program referenced can be found in the "IBM Catalog of Programs" appropriate to a machine (available through a local IBM Branch Office). There are, of course, hundreds of statistical programs in use other than those in the IBM Catalog.

The third item provides a lead to other possible sources of programs.

References are to be found in alphabetic order in the back of this manual. A citation is often made to a later source rather than to the original one. At times a judgment about the use of a technique is based only on a title or an abstract. "Computer Abstracts" and "The H. W. Wilson Company Indexes" proved useful in gathering raw data on the use of a computer or a technique.

Readers are encouraged to recommend other techniques, and contribute new citations to applications or programs. Write: Technical Publications, IBM Corporation, Data Processing Division, 112 East Post Road, White Plains, New York, 10601. We welcome comments and criticisms, too.

CORRELATION

in essence Correlation analysis measures the strength of relationship between two or more variables.

example in paper-making Thirteen kinds of tests on each of a great number of rolls of paper reduce to four tests because of correlation analysis. Ten of the tests were so highly correlated (when one had high values, the others did too; when one had low values, so did the others) that nine of the original tests could be dispensed with. (Ref. Draper, p. 315)

example in nutrition Correlations among 184 nutritional, metabolic, biochemical, and physiological variables (vitamin intake, mineral intake, amounts absorbed, characteristics of subjects, etc.) revealed important interrelations. Significant insights came from observing how the correlations changed as the experiment progressed in time. (Ref. James)

summary From repeated measurements on a number of variables, correlation analysis measures the strength of association between every pair of variables. Degree of association runs from +1.0 (perfect), down through 0.0 (no correlation), to -1.0 (perfect inverse association). You search for high correlations (actually, either high direct or high inverse) in the paper-making example, to cut out superfluous tests. In the nutrition example, high correlations point to possible cause and effect relations (although high correlation need not, and usually does not, imply cause and effect situations). Many varieties of correlation measures exist.

	V1	V2	V3	V4
M1	-	-	-	-
M2	-	-	-	-
M3	-	-	-	-
⋮				

→ **CORRELATION ANALYSIS** →

	V1	V2	V3	V4
V1	1.0	.9	.1	.2
V2	-	1.0	.7	.1
V3	-	-	1.0	-.8
V4	-	-	-	1.0

Input:

repeated measures on a set of variables V1, V2, V3, V4. These four variables could be measures of height, weight, pulse, and blood pressure on individuals M1, M2, ...

Output:

correlations, a high 0.9 for V1 and V2; a low 0.1 for V2 and V4; a high inverse -0.8 for V3 and V4. Now highly correlated variables can be collected and/or some eliminated. Searches can be made for possible causes and effects, when appropriate.

applications

A few applications of correlation analysis:

Subject	Reference
earth sciences	Miller, ch. 13
psychotropic problems	Hall
paper testing	Draper, p. 315
nutrition, metabolism	James
geochronology	Martin
educational measurements	Cooley, p.21
heart pathology	Tolles

catalog of programs

Programs available from "IBM Catalog of Programs" (Ref. IBM) in correlation analysis and related areas:

Subject	Computer	Program Form Number
correlation	7070	7070-11.3.005, .008
correlation programs	1620	1620-06.0.013, .015, .021, .022, .038, .039, .040, .051, .064, .097, .104, .114, .121, .125, .162, .170, .175, .188, .205
correlation	1401	1401-06.0.005, .006
correlation	360	360A-CM 03X
	1130	1130-CM 02X

use of IBM systems

Citations in trade journals, periodicals and texts on the use of IBM systems in carrying out correlation analysis:

Subject	Computer	Reference
correlation pairing	709/0	Priest
tetrachoric correlation	709/0	Castellan
correlation, health of aviators	1620	Osborne
correlation	709	Sorenson
biserial correlation	709/0	Castellan (2)
BIMED programs (14)	7090	Massey
nutritional relationships	1620	James
geochronology	7094	Martin

FACTOR ANALYSIS

in essence

Factor analysis finds new, more fundamental quantities (the factors) underlying measured variables.

example in advertising

In an advertising effectiveness study, 20 variables (size of ad, colors, type sizes, copy blocks, product facts, benefits, pictures, readership, etc.) bunch into 6 groups or factors. One factor related to the pictorial and color variables, a second to ad size variables, a third to typography variables, a fourth to information variables, etc. (Ref. Ferber, p. 101)

example in credit

A factor analysis of 22 questions on a credit request revealed 6 basic factors: two connected with questions on the transaction, and four related to questions on personal history. (Ref. Myers)

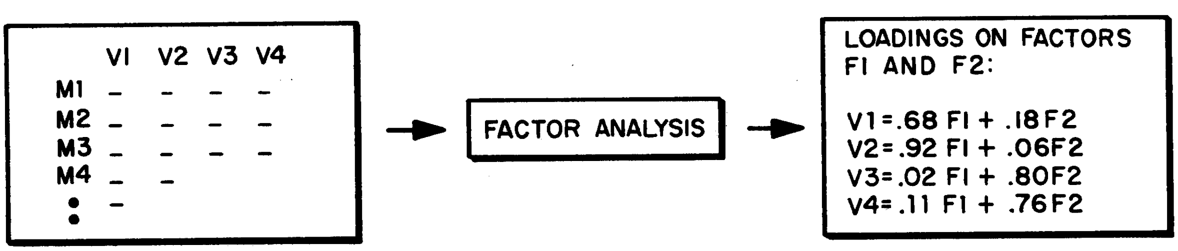
example in psychology

Some of the 48 scores (variables) on a Rorschach test were too much dependent on the number of responses given. Factor analysis into one general factor (interpreted as productivity) and 15 other factors, independent of the general factor, removed the difficulty. (Ref. Cooley, p. 164)

summary

With repeated measurements on a set of variables, factor analysis discerns underlying factors distinct from, and fewer in number than, the original variables. Some factor analytic methods seek out a general factor present in all original variables (the Rorschach example). In other methods, original variables are grouped so as to depend on a few distinct factors (advertising and credit examples).

Formulas which show each variable as a combination of the factors are important outputs of factor analysis. For example, in the advertising illustrations above, the variable representing "number of words" equals
$0.05\ F1 + 0.46\ F2 - 0.10\ F3 + 0.71\ F4 + 0.00\ F5 - 0.03\ F6$.
The 6 variables, which like this one had strong contribution from F4, all seem to concern the "information factor" in the ads.



Input:

repeated measurements of variables V1, V2, V3, V4,(e.g., size, width, color, words tallied for a series of ads M1, M2, etc.).

Output:

formulas for four variables in terms of fewer factors. V1 and V2 are based mainly on F1; V3 and V4 on F2. You now have fewer factors to deal with. Hopefully they are more meaningful than the original four variables.

applications

A few applications of factor analysis:

Subject	Reference
metallurgy, steels, blast furnace	Spurrell
psychology, multiple time series	Anderson
metropolitan economy	Carleton
industrial relations	Boehr
consumer attitudes	Adams
psychology	Overall
physical fitness tests	Falls
psychological data	Schönemann
tests on paper products	Draper, p. 316
refinery process	Thomas
botanical applications	Pearce
language ability	Weaver
listening tests	Bateman
retail credit	Myers
census data	Massey, W.F.
earth sciences	Miller, ch. 13
Rorschach test (48 dimensions)	Cooley, p. 164

catalog of programs

Programs available from "IBM Catalog of Programs" (Ref. IBM) in factor analysis and related areas:

Subject	Computer	Program Form Number
factor analysis	7070	7070-11.3.005, .008
factor analysis	1620	1620-06.0.053, .091, .094, .103, .145, .169
factor analysis, varimax	360	360A-CM 03X
	1130	1130-CM 02X

use of IBM systems

Citations in trade journals, periodicals, and texts of the use of IBM systems in carrying out factor analysis:

Subject	Computer	Reference
factor analysis, oblique	7094	Hendrickson
multivariate statistics	7090/4	Jones
nonlinear factor analysis	7090/4	McDonald
factor analysis, rotation	7090	Wolf
principal axes	650	Burket
principal components	7070	Bendig
BIMD programs (14)	7090	Massey, F.
convulsive disorders	704	Rodin
principal components	7090	Steidler
factor analysis, square root	7090	Lingoes
principal components	1620	Moore, D.W.
factor analysis (3 mode)	709	Walsh
principal axes	709	Burket
psychological data	7094	Schönemann
language ability	709	Weaver

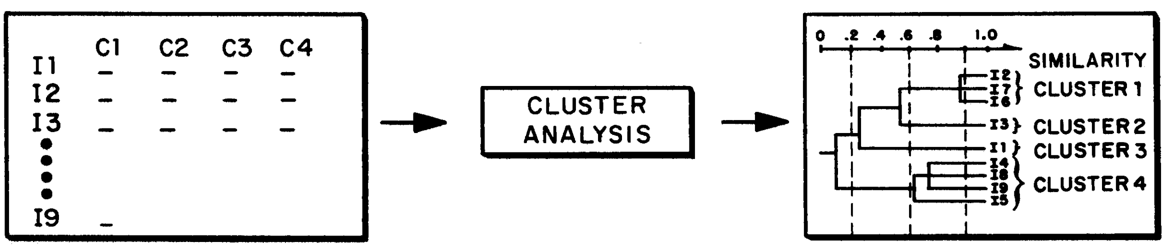
CLUSTER ANALYSIS

in essence — Cluster analysis groups items or individuals by means of their characteristics.

example in bacteriology — Cluster methods applied to bacteria have grouped them into their proper orders, or grouped strains within a species. Classification of an atypical plague bacillus later proved accurate. (Ref. Sokal, p. 259)

example in education — Attitudes toward mathematics were thought to have a significant relation to general personality variables. In the study which confirmed this hypothesis, 42 personality variables (dominance, sociability, tolerance, etc.) clustered as follows. Variables 1, 2, 3, 7, 9, 10 constituted a "extroversion" group. Variables 4, 5, 6, 7, 8, 10, a "conscientiousness" cluster. Variables 5, 7, 11, 14, a "self-control" cluster, etc. (Ref. Aiken)

summary — From characteristics of individuals in a large, unorganized collection, cluster analysis groups similar individuals together. Among the many variants of cluster methods, certain ones place the clusters into hierarchies according to degree of similarity (the bacteriology reference above). The technique also can cluster attributes (variables) and thus is similar in purpose to factor analysis.



Input:

four characteristics C1 to C4 measured on nine items. For example, length, width, wing span, and antenna length are measured on nine insects.

Output:

at about .6 on a degree of similarity scale, four clusters emerge. Cluster 1 contains items I2, I7, I6. Cluster 2 has only I3, etc. At .2 on the similarity scale, only two clusters emerge: cluster 4 and a coalescence of clusters 1, 2, 3. You might be searching for a genus at similarity =.2; or for a species at similarity =.6; or strains at similarity =.9.

applications

A few applications of cluster analysis:

Subject	Reference
botany (rice, manioc, farinosae)	Sokal
paleontology (fish)	Sokal
bateria (soil, plage, actinomycetes)	Sokal
viruses, genes, protein patterns	Sokal
plant ecology, soil classification plankton	Sokal
physical anthropology	Sokal
medical diagnosis	Sokal
rock identification	Sokal
oil exploration	Sokal
legislative voting patterns	Sokal
language translation	Sokal
pattern recognition	Sokal
archeology	Sokal
library, document classification	Sokal
philology, authorship tests	Sokal
leukemia classification	Whitfield
information retrieval	Sokal
zoology (bees, mosquitos, man, cats, sponges)	Sokal
earth sciences	Miller
math and personality	Aiken

catalog of programs

Programs available from "IBM Catalog of Programs" (Ref. IBM) in cluster analysis and related areas:

Subject	Computer	Program Form Number
optimal clustering	7092	7092-G2 IBM0026
continuous variables	7090	7090-Z0 IBM0015
binary variables	7090	7090-Z0 IBM0002
nonnumeric	1620	1620-06.0.201

use of IBM systems

Citations in trade journals, periodicals and texts of the use of IBM systems in carrying out cluster analysis:

Subject	Computer	Reference
taxonomic classification	7090	Lingoes (4)
disease classification	7090	Bonner (2)
cluster analysis	1620	Hyvarinen
pattern recognition	7090	Bonner

REGRESSION

in essence

Regression analysis computes prediction formulas from data.

examples in metallurgy

Regression related the Charpy fracture temperature to 11 variables (percent carbon C, manganese Mn, phosphorus P, etc.) by the formula

$$(311.7)C + (133.6)Mn + (1194.7)P + (57.5)Si - (26.4)Ni + \ldots\ldots$$

(Ref. Pehlke, p. 58)

example in hospitals

Regression found equations to predict hospital beds needed in various case categories. For example, neurology cases per month = 6.65 + (0.635)X17 + (3.47)X18 where variables X17 is Indiana deaths and X18 Indiana births. One hundred seventeen variables (births, accident rates, number of specialists, etc.) were screened. (Ref. Beenhakker)

example in agriculture

The method of regression found the explicit formula for milk output as a function of X1 = concentrate feed, X2 = hand feed, X3 = grassland acre days, and X4 = average cows in herd. The decimal exponents on each of the variable X1 thru X4 can be handled by using logarithms.

$$\text{milk output} = 39.80(X1)^{0.24}\ (X2)^{0.15}\ (X3)^{0.02}\ (X4)^{0.05}$$

(Ref. Cowling)

summary

Starting with data on several variables and an idea of the general formula connecting the variables, regression analysis finds the specific numerical formula relating the variables. Statements can usually be made about how confident you are in the formula and in predictions using it. For example, in the agriculture example above, about 75 percent of the variation in the data is explained through the formula for milk output. Reduction of masses of data to simple equations makes regression important. Two other advantages are understanding the nature of a process and prediction of future events.

Inputs:

data on three variables V1, V2, V3. Also, the general form of one in terms of others.

Outputs:

the numeric coefficients in the regression formula: A=0.1, B=1.2, C=0.6. Also, certain quantities used to measure confidence in the formula or in the coefficients. R^2=0.76 means 76 percent of the variation in V3 is explained by the equation in V1 and V2. The formula has summarized much data; it can be used to predict a new V3 from V1 and V2.

applications

A few applications of regression analysis:

Subject	Reference
paper industry	Moore, P.G.
agriculture, milk production functions	Cowling
weather forecasting	Glahn
educational measurements	Cooley, p. 38
meteorology	Lund
bank debits as an indicator	Carleton
construction economics	Dilbeck
urban transportation	Kain
U.S. import demand	Reimer
wartime production	Rapping
urban refuse collection	Hirsch
open hearth production	Leckie
earth sciences	Miller, ch. 8, 9, 17
gasoline production	Ostle, p. 167, 183
psychological tests, canonical correlation	Meredith
psychological scores	Meredith (2)
forging, metallurgy, lattice parameters	Pehlke
economic model of United Kingdom (simultaneous)	Ball
inertial navigation, error sources	Eisner
aerodynamics, downwash	Fromme
dairy production	Jarrett
X-ray fluorescence	Alley
foundry steels	Sprinkle
hospital bed needs	Beenhakker

catalog of programs

Programs available from "IBM Catalog of Programs" (Ref. IBM) in regression analysis and related areas:

Subject	Computer	Program Form Number
regression	650	0650-00.0.056
regression	705	0705-11.3.001
regression	1410	1410-11.3.001, .002
regression	7070	7070-11.3.001, .007, .011
regression	1620	1620-06.0.001, .003, .006, .031, .042, .049, .057, .066, .077, .084, .101, .118, .120, .122, .142, .143, .154, .157, .159, .168, .173, .181, .187

Subject	Computer	Program Form Number
nonlinear regression	704	0704-G2 3226N11
iterative least squares	709	0709-E2 3024LSQ
orthogonal polynomials	709	0709-E2 3197CF
systems of equations	709	0709-F1 3090NORM
mortality curves	7040	7040-G1 3356MORT
nonlinear least squares	7040	7040-G2 3094LIN
stepwise regression	7040	7040-G2 3205RRG
polynomial fit	7090	7090-E2 3289PLYF
regression	7090	7090-G2 3104RGNL
stepwise regression	7090	7090-G2 3143MPR2, MPR3
differential equations	7090	7090-G2 3146NLR
constrained regression	7094	7094-E2 3363BJ06
polynomial least squares	7094	7094-E2 3372AM26
regression	1401	1401-06.0.002, .003, .004, .005, .007, .008
regression subroutines (multiple, polynomial, canonical)	360 1130	360A-CM 03X 1130-CM 02X

use of IBM systems

Citations in trade journals, periodicals, and texts of the use of IBM systems in carrying out regression analysis:

Subject	Computer	Reference
resonance spectra	1410	Nelson
multivariate programs (30)	7090/4	Jones
regression prediction	7090	Lingoes (2)
decay data	0650	Worsley
simultaneous regression	7090	Lingoes (3)
psychophysiology	1620	Williams
multiple regression	7090	Steidler
heart disease	0650	Ward
tomato factors	7070	Mittler
chemical analysis	0650	Winchell
BIMD programs (14)	7090	Massey, F.
generalized regression	0704	Eisenpress
metallurgy	0704	Pehlke
aerospace naviation	7090/4	Eisner
hospital bed prediction	7090	Beenhakker

DISCRIMINATION

in essence — Discriminant analysis assigns individuals to known groups.

example in anthropology — Four measurements, X1, X2, X3, X4, taken on each of many human and chimpanzee fossil teeth yielded a discriminant function.

$$D = X1 - (7.49)X2 + (2.34)X3 + (4.70)X4$$

with D averaging -5.0 ± 2.45 for human teeth and averaging $+17.6 \pm 2.45$ for chimpanzee teeth. The Taungs skull, with a $D = -7.9$, was subsequently classified as probably human. (Ref. Keeping, p. 366)

example in linguistics — Three hundred words in each of three languages — English, Swedish, and Finnish — were treated by discriminant analysis. Quantitative measures (e.g., number of A's, of B's, ..., of syllables, etc.) were taken on each word. A first discriminate function separates Finnish from the other two; a second function distinguishes English and Swedish. (Ref. Mustonen)

summary — Starting from known groups of individuals, each individual with measured characteristics, discriminant analysis derives the so-called discriminant functions. They allot the originally given individuals to their proper group and new individuals to an appropriate group. The functions can't always cleanly distinguish the groups as they did in the linguistics example. A misclassification percent measures the effectiveness of the discriminant functions in allocating the original individuals into their known groups. Cluster analysis differs from discriminant analysis in that cluster analysis discovers groups, whereas discriminant analysis begins with recognized groups.

		C1	C2	C3
GROUP 1	I1	–	–	–
	I2	–	–	–
GROUP 2	I3	–	–	–
	I4	–	–	–
	I5	–	–	–

→ DISCRIMINANT ANALYSIS →

THE DISCRIMINANT FUNCTION FOUND IS
$F1 = 2.6 + 1.1(C1) + 0.9(C2) - 7.0(C3)$
WITH AVERAGE VALUE
FOR GROUP 1 $= 18.2 \pm .9$
GROUP 2 $= -1.3 \pm .9$

Inputs:

the five individuals I1 to I5 fall into two known groups. Characteristics C1, C2, C3 are measured on each individual.

Outputs:

the discriminant function serves to classify a new individual into his appropriate group on the basis of his characteristics.

applications — A few applications of discriminant analysis:

Subject	Reference
judgments in cooking	Baten
brand loyalty	Farley
coal analysis	Baten (2)
earth sciences	Miller, ch. 12
educational methods	Baten (3)
weather prediction	Glahn, p. 122
textile research	Baten (4)
medicine	Radhakrishna
soil differentiation	Cox
linguistic problems	Mustonen
biological taxonomy	Burnaby
U.S. National Health Survey	Fisher
physical anthropology	Ashton
meteorlogy	Lund
basalic lava discrimination	Chayes

catalog of programs — Programs available from "IBM Catalog of Programs" (Ref. IBM) in discriminant analysis and related areas:

Subject	Computer	Program Form Number
classification	1620	1620-06.0.076, .201, .208
discrimination	7094	7094-BMD 04M, 05M
discrimination routines	360	360A-CM 03X
	1130	1130-CM 02X

use of IBM systems — Citations in trade journals, periodicals and texts of the use of IBM systems in carrying out discrimination analysis:

Subject	Computer	Reference
classification	7094	Kossack
multivariate statistics	7090/4	Jones
health survey	704	Fisher

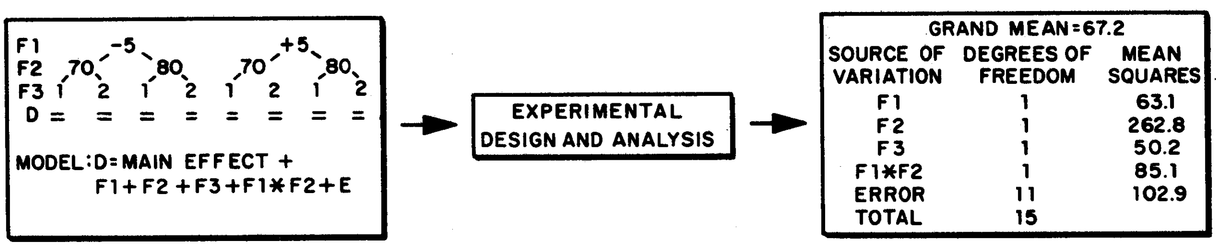
EXPERIMENTAL DESIGN AND ANALYSIS

in essence Experimental design and analysis methods decide whether various factors or combination of factors influence a result.

example in aluminum The impact extrusion of aluminum turned out to be influenced by 7 alloy variations, 4 process variables, and 3 annealing temperatures. (Ref. Pehlke)

example in ceramics The physical properties of a semivitreous body were shown by experimental design to depend on such factors as particle size, water, entrapped air, thickness, firing rate. The water, air, and thickness variables appeared linear; particle size was nonlinear. Combinations between levels of particle size and other factors caused important changes. (Ref. Conrad)

summary One starts with measured outcomes for various combinations of influencing factors (many at preselected values, some simple observed). Experimental design and analysis techniques sort out the important factors or combination of factors producing the outcome. A host of "designs" are used. Randomization and repetition of runs insure confidence in results.



Inputs:

three factors F1, F2, F3, each with two levels (F1 at -5 and +5 for example) are examined for their effect on outcome D, both separately and in F1*F2 combination. These separate contributions are over and above an average main effect and error, E. The outcome D is measured twice (replicated).

Outputs:

the so-called "mean squares" and "degrees of freedom" determine whether or not the factors or combinations really affect the outcome. Other models including new factors or triple combinations F1*F2*F3 can be tried.

applications

A few applications of experimental design and analysis:

Subject	Reference
testing rocket engines	Wood
geology, paleontology, etc.	Miller, ch. 7
electric machinery electrodes	Hicks
psychological experiments	Chan
machinery metals	Hamaker
psychotropic problems	Hall
transistor industry	Hamaker
aluminum impact extrusions	Pehlke
castings	Brownlee
textiles, cotton spinning	Peake
general review	Hunter
steels, Ausforming process	Duckworth
chemical tests of textiles	Bainbridge
paper pack seals	Moore, p. 307
soaps, R/D, manufacturing, marketing	Michaels
U.S. Patent Office, transistor circuits	Bryant, p. 204
automobile purchasing	Jung
ceramics	Conrad
hardening of steels	Hopkins, A.D.
etc.	

catalog of programs

Programs available from "IBM Catalog of Programs" (Ref. IBM) in experimental design and related areas:

Subject	Computer	Program Form Number
variance analysis	7070	7070-11.5.002
covariance analysis	1620	1620-06.0.023, .024, .025, .032, .080, .092, .107, .109, .129
analysis of variance	1620	1620-06.0.026, .027, .028, .029, .030, .033, .041, .043, .060, .061, .062, .065, .069, .070, .081, .083, .086, .087, .088, .089, .102, .105, .113, .123, .132, .139, .140, .152, .161, .174, .176, .202, .207, .210, .213
variance, covariance	7040	7040-G2 3365 ANOV
analysis of variance	7090	7090-G4 3027 ANAWZ
factorial analysis	7094	7094-G4 3337 ANV
factorial analysis	1401	1401-06.0.012, .014
factorial design	360	360A-CM-03X
	1130	1130-CM-02X

use of IBM systems

Citations in trade journals, periodicals, and texts of the use of IBM systems in carrying out experimental design and analysis:

Subject	Computer	Reference
variance, mean difference	7072	Turk
paired comparisons	7090	Gulliksen
electroencephalogram, anova	7090	Sorkin
Latin squares	7090	Gilbert, E.N.
incomplete factorial	7094	Webb
variance, covariance	7090	Finn
multivariate statistics	7090/4	Jones
analysis of variance	7074	Hemmerle
analysis of variance	7070	Bendig
BIMD programs	7090	Massey, F.
factor analysis	1410	Cientat
analysis of variance	709/90	Hopkins
variance, covariance (7)	1401	Sterling

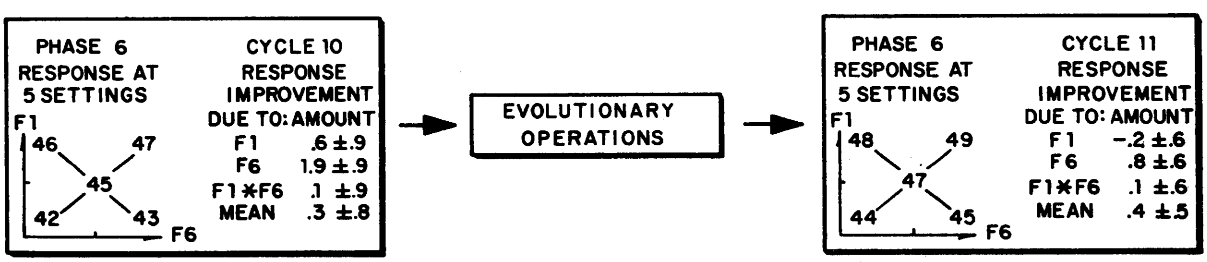
EVOLUTIONARY OPERATION

in essence — Evolutionary operation resets variables step by step to make a process better.

example in chemicals — A high-cost organic chemical process has yields depending on two important factors: reaction time and weight of additives. Starting from time = 75 minutes and weight = 120 pounds, evolutionary operation gradually improved yield from 562 to 588 pounds by adjusting time to 25 minutes and weight to 110 pounds. (Ref. De Busk)

example in petroleum — In a catalytic cracking application four variables — feed rate, reactor temperature, recycle, and space velocity — were adjusted to maximize catalytic distillate yield and catalytic light gas oil yield. The levels and order of successive runs were imposed by operating personnel. (Ref. Klingel)

summary — Given a process whose yield (or some other response) is a function of a number of factors, evolutionary operation tries small changes in two or three controlling factors to improve the yield gradually. The slight, deliberate changes in process variables must cause some improved response over and above mere random fluctuation.



Inputs:

a process has been operating during previous phases (with other settings for factors, or other factors, or another response variable). Factor F6 at its high level causes the response to exceed mere "noise" (±.9). The process was then shifted permanently to this high value of F6 and evolutionary operations begun.

Outputs:

response is measured at settings of F1 and F6, around the central operating point. Again the high setting of F6 improved the response. F1 did not affect the response; however, next cycle around, F1's operating limits were spread wider to discover its influence.

applications A few applications of evolutionary operation:

Subject	Reference
geological facies maps	Miller, p. 394
alloying process, quality control	Bingham
textiles, spinning machinery	Dudley
chemical process	De Busk
catalytic cracking	Klingel
response surfaces, general	Bradley
cutting-tool equation	Wu
chemical process	Kochler
chemical plants	Box
quality in plastic processes	Harrington

BAYES FORMULA

in essence — Bayes formula finds probable causes from measured effects.

example in medicine — Medical knowledge is available in the form of probabilities of a complex of symptoms arising from a given disease complex. Also known are the frequencies of occurrence of the disease complexes. Bayesian techniques discover the most probable disease complex from a newly presented set of symptoms. (Ref. Ledley, Sect. 12-4)

example in quality control — In sample and quality control there are subjective estimates for the chances that different fractions of a lot are defective. On the basis of new information from samples, Bayesian analysis revises these subjective estimates. (Ref. Locke)

summary — Given the probabilities that certain effects follow certain causes (based on historical data); and given the frequency of occurrence (or probabilities) of the causes (based on historical or subjective knowledge). Use of Bayes formula finds the probability of a cause behind a newly presented effect. Strictly speaking, Bayesian statistics or Bayesian decision theory is a quite wide field. Here we narrow attention to using Bayes formula to deal with what can be looked upon as causes (C) and effects (E). The probability, P (Ci|Ej), of a cause Ci given the presence of an effect Ej is

$$P\,(Ci \mid Ej) = \frac{P(Ci)\ P(Ej \mid Ci)}{\underset{\substack{\text{over all} \\ \text{diseases, } i}}{\text{Sum}}\ (P(Ci)\ P(Ej \mid Ci))}$$

In the medical example, symptoms and diseases take the place of effects and causes. In the quality control example, sample information (effect) revises subjective estimates of the fraction (cause) of lot defectives.

Inputs:

P(E1/C2) means the probability of effect, E1 given existence of cause, C2. These kinds of probabilities are known, together with probabilities of causes. P(C2) is that of cause, C2.

Outputs:

probabilities of causes from a knowledge of effects. There is a probability of .4 that cause C2 underlies effect EO. Most probable causes for given effects are evident from a scan of a column of probabilities.

applications A few applications of Bayesian analysis:

Subject	Reference
reliability, quality control	Schafer
sequential life testing	Ginsburg
quality control	Locke
competitive bidding, equipment selection	Peterson
clinical trials	Novick
statistical decision	Greyson
quality control	Hamburg
marketing research costs	Bass
medical diagnosis	Ledley, Sect. 12-4
hospital engineering	Aitchison
psychiatric classification	Birnbaum
control theory	Ho
criminalistics	Kingston
military information processing	Kaplan

TIME SERIES ANALYSIS

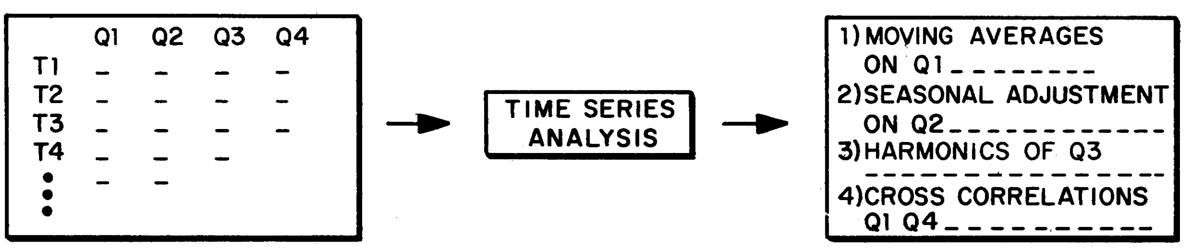
in essence — Time series methods analyze, compare, and predict quantities which change in time.

example in agriculture — Ten years of monthly data on hog production reveal, through time series methods, a major 12-month cycle, a medium 6-month cycle, and a minor 4-month cycle. The model accounts for a shifting of peaks and troughs over the ten-year period. (Ref. Abel)

example in economics — Of 800 time series examined for their relationship to a general business cycle, 21 series were selected as statistical indicators. Eight led the general cycle, 5 lagged behind it, and 8 coincided with it. The predictive ability of the indicators was satisfactory. (Ref. Chou, P. 566)

example in geophysics — To discover underlying geologic structures in inaccessible or covered regions, aeromagnetic maps are taken. Autocorrelation and spectral analysis techniques discover dominant trends, faults, and lithologic periodicities. (Ref. Horton)

summary — With measurements over time on one or several quantities, time series techniques (1) smooth out the quantity (weighted averages); (2) extract important trend-cycle-seasonal ingredients (seasonal adjustment); (3) discover pure cyclic ingredients (harmonic analysis); or (4) compare pairs of series (cross-correlation). Prediction is a major objective. The agriculture example above is a use of harmonic analysis. The other two examples relate to cross- and auto-correlation. The geophysics illustration substitutes variables as functions of distance for variables depending on time.



Inputs:

one or several quantities, Q1, Q2, Q3, Q4, are known over time. A variety of approaches in time series allow prediction of the quantities at future times.

Outputs:

(1) 20 day moving averages on Q1 might be 97.8, 96.2, 95.9, ...
(2) Q2 without seasonals might be 15.8, 17.2, 18.3, 16.1, ...
(3) the cycle size of Q3 might be .2 for 7 day, .5 for 14 day, ...
(4) for lags of 0, 1, 2, ... the cross correlation of Q1 and Q4 might be .98, .80, .67, .20, ...

applications

A few applications of time series:

Subject	Reference
chemical process control, cum sum	Truax
biological systems	Attinger
tracking (radar, sonar), exp. smoothing	Helms
mobile radios reception, power spectra	Gilbert, E. N.
new-car sales forecasting	Dyckman
stock prices, random walks	Fama
consumer attitudes, regressions	Adams
earth sciences	Miller, Ch. 15
electroencephalography, cross-correlation	Ledley, Sect. 10-2
X-ray patterns, Fourier series	Pehlke
hog production, harmonic analysis	Abel
textiles, periodograms correlograms	Foster
meteorology, power spectra	Craddock
inventory, production control, exponential	Wagle
economic cycles, spectral analysis	Adelman
marine profiles, spectral analysis	Neidell
labor force, employment, seasonal adjustment	Shiskin
closing stock prices	Brown
radar data	Jenkins
aeromagnetic maps	Horton

catalog of programs

Programs available from "IBM Catalog of Programs" (Ref. IBM) in time series analysis and related areas:

Subject	Computer	Program Form Number
seasonal adjustment	650	0650-06.0.041
autocovariance	7070	7070-11.2.001, .002
power spectrum	1620	1620-06.0.005, .056, .126 .133, .147, .166
seasonal adjustment	1620	1620-06.0.054
business cycles	709	0709-G1 3103 BCA
econometric forecasts	7090	7090-GO 32419 FES
series decomposition	7090	7090-C1 3144 TSDA
seasonal adjustment	1401	1401-CA 04X and -06.0.015
time series	7094	7094-BMD 01S, 03S
time series routines	360	360A-CM 03X
(averaging exponential smoothing, Fourier analysis, auto- and cross-correlation)	1130	1130-CM 02X

use of IBM systems

Citations in trade journals, periodicals, and texts of the use of IBM systems in carrying out time series analysis:

Subject	Computer	Reference
upper winds, smoothing	7090	Reiter
autocorrelation	7090	Schmid
geophysics	7090	Simpson
economic time series	7090	Karreman
power systems spectra	0650	Cooke
filtering	7094	Whittlesey
sales, exponential smooth	0704	Whelan
X-ray patterns	0709	Pehlke

REFERENCES

AC	Analytical Chemistry
ACSB	American Ceramic Society Bulletin
AD	(Government Report)
AER	American Economics Review
AIAA	American Institute of Aeronautics & Astronautics
AJS	American Journal of Science
AJST	Australian Journal of Statics
AMI	American Mineralogist
AMS	American Statistician
ARS	Annual Review of Psychology
AS	Applied Statistics (JRSS-C)
ASAJ	American Statistical Association Journal
ASAP	American Statistical Association, Proceedings of the Social Statistics Section
BIS	Biometrics
BIT	Nordisk Tidskrift for Informations-Behandling
BJ	Biophysical Journal
BS	Behavioral Science
BSTJ	Bell System Technical Journal
C	Control
CACM	Communications of the Association for Computing Machines
CEP	Chemical Engineering Progress
EPM	Educational & Psychological Measurements
FT	Food Technology
G	Geophysics
HPPR	Hydrocarbon Processing & Petroleum Refiner
IBMJ	IBM Journal of Research & Development
IEC	Industrial & Engineering Chemistry
IEEE	Institute of Electrical & Electronics Engineers
IQC	Industrial Quality Control
ISIJ	Iron & Steel Institute Journal
JACM	Journal of the Association of Computing Machinery
JAP	Journal of Applied Psychology
JB	Journal of Business
JC	Journal of Communications
JEE	Journal of Experimental Education
JER	Journal of Educational Research
JM	Journal of Metals
JPT	Journal of Petroleum Technology
JRSS	Journal of the Royal Statistical Society
JS	Journal of Science (Iowa University)
JTI	Journal of the Textile Institute
NYAS	New York Academy of Sciences, Transactions
NYAS	New York Academy of Sciences, Transactions
OR	Journal of the Operations Research Society of America
PRS	Proceedings of the Royal Society
PS	Psychometrika
RER	Review of Educational Research
RES	Review of Economics and Statistics
S	The Statistician
SML	Statistical Methods in Linguistics
SIAM	Society for Industrial & Applied Mathematics, Journal of the
TRJ	Textile Research Journal

Note: + indicates more than one author.
The twelve months are abbreviated Ja, F, Mr, Ap, My, Je, Jy, Ag, S, O, N, D.

Abel, M. E.	ASAJ 57: 655-67 (S'62)
Adams, F. G.	RES 47: 367-78 (N'65)
Adelman, I.	AER 55: 444-63 (Je'65)
Aiken, L. R.	JER 56: 476-80 (My'63)
Aitchison, J.	S 15: 313-18 ('65)
Alley, B. J. +	AC 37: 1685-90 (D'65)
Anderson, T. W.	PS 28: 1-25 (Mr'63)
Ashton, E. H. +	PRS 146B: 552-72 ('57)
Attinger, E. O. +	BJ 6: 291-304 (My'66)
Baehr, M. E.	PS 28: 199-209 (Je'63)
Bainbridge, T. R.	IQC 22: 12-20 (Jy'65)
Ball, R. T.	AS 12: 14-25 (Mr'63)
Bass, F. M.	JB 36: 77-90 (Ja'63)
Bateman, D. +	JC 14: 183-89 (S'64)
Baten, W. D. +	IQC 14: 6-10 (Ja'58)
Baten, (2), W. D. +	IEC 16: 32 - ('44)
Baten, (3), W. D. +	JEE 12: 184-6 ('44)
Baten, (4), W. D. +	TRJ 20: 869-72 ('50)
Beenhakker, H. L.	OR 11: 824-39 ('63)
Bendig, A. W.	BS9: 85-6 (Ja'64)
Bingham, R. S.	IQC 20: 17-23 (S'63)
Birnbaum, A. +	AS 9: 152-69 (N'60)
Bonner, R. E.	IBMJ 6: 353-60 ('62)
Bonner, (2), R. E.	IBMJ 8: 22-32 (Ja'64)
Box, G. E.	AS 6: 81-101 (Je'57)
Bradley, R. A.	IQC 15: 16-20 (Jy'58)
Brown, R. G. +	OR 9: 678-85 ('61)
Brownlee, K. A.	IQC 13: 12-20 (F'57)
Bryant, E. C.	ICIREPAT, 3rd Annual Meeting, Spartan Books ('64)

Burket, G. R.	AD - 419370 (F'60)
Burket, (2), G. R.	AD - 269318 (O'61)
Burnaby, T. P.	BIS 22: 96-110 (Mr'66)
Carleton, W. T. +	JB 38: 409-15 (O'65)
Castellan, (2), N. J. +	BS 8: 368-9 (O'63)
Castellan, N. J. +	BS 9: 292 (Jy'64)
Chan, M.	PS 29: 233-40 (S'64)
Chayes, F. +	AJS 263: 206-22 ('65)
Chou, Y.	"Applied Business & Economics Statistics" Holt, R., and W. ('63)
Cientat, V. J. +	BS 9: 197-98 (Ap'64)
Conrad, K. L. +	ACSB 44: 616-19 (Ag'65)
Cooke, J. L.	IEEE P-83: 34-41 (Ja'64)
Cooley, W. W. +	"Multivariate Procedure for the Behavioral Sciences" Wiley ('62)
Cowling, K. +	S 14: 255-66 ('64)
Cox, G. M. +	JS11: 323-31
Craddock, J. M.	S 15: 167-190 ('65)
De Busk, R. E.	IQC 19: 15-21 (O'62)
Dhrymes	RES 46: 76-81 (F'64)
Dilbeck, H. R.	JB 37: 295-308 (Jy'64)
Draper, J.	S 14: 311 ('64)
Duckworth, W. E.	S 15: 7-30 ('65)
Dudley, A.	JTI 49: T317 ('58)
Dyckman, T. R.	JB 38: 252-66 (Jy'65)
Eisenpress, H.	Western Data Processing Center, Proc. UCLA: 143-59 (Ja'59)
Eisner, W. +	A1AA2: 722-27 (Ap'64)
Falls, H. B. +	BIS 19: 658 (D'63)
Fama, E. F.	JB 38: 34-105 (Ja'65)
Farley, J. U.	JB 37: 370-81 (O'64)
Finn, J.	BS 10: 323 (Jy'65)
Fisher, G.	AS11: 148-163 (N'62)
Foster, G.A.R.	JRSS, Supp. 8, 42, (46)
Fromme, J. A.	A1AA 2: 1349-50 (Jy'64)
Gilbert, E. N.	SIAM 7: 189-98 (Ap'65)
Gilbert, E. N.	BSTJ 44: 1779-1803 (O'65)
Ginsburg, H. +	IQC 21: 186-91 (O'64)
Glahn, H. R.	S 15: 111-142 ('65)
Green, P. E.	IEC 54: 30-34 ('62)
Greyson, C. J.	JPT 14: 603-7 (Je'62)
Gulliksen, H. +	AD - 605691 (Jy'64)
Hall, C. E.	PS 31: 316-21 (Ap'66)
Hamaker, H. C.	IQC 16: 12-14 (Jy'59)
Hamaker, (2), H. C.	IQC 17: 16-20 (Mr'61)
Hamburg, M.	IQC 19: 10-14 (D'62)
Harrington, E. C.	IQC 21: 494-98 (Ap'65)
Helms, H. D.	BSTJ 44: 2337-62 (D'65)
Hemmerle, W. J.	JACM 11: 234-39 (Ap'64)
Hendrickson, A. E.	BS 9: 288-89 (Jy'64)
Hicks, C. R.	IQC 17: 17-20 (D'60)
Hirsch, W. Z.	RES 47: 87-92 (F'65)

Ho, Y. C. +	IEEE 9AC: 333-39 (O'64)
Hopkins, A. D. +	ISIJ 200: 308-16 (Ap'62)
Hopkins, C. E.	BS 8: 85 (Ja'63)
Horton, C. W. +	G 29: 582-601 (Ag'64)
Hughes, M. C.	IQC 22: 348-54 (Ja'66)
Hunter, J. S.	IQC 22: 391-96 (Fe'66)
Hyvarinen, L.	BIT 2: 83-89 ('62)
IBM Corp.	
IBM 650	Form C20-1600
1401/40/60	C20-1601
1410/705/70/74/80	C20-1602
1620	C20-1603
704/40/44/9/90/94	C20-1604
360	C20-1619
1130	C20-1630
James, W. H.	AC 36: 1848-49 (Ag'64)
Jarrett, F. G.	AJST 1: 3-11 (Ap'59)
Jenkins, G. M.	AS 14: 2-32 ('65)
Jones, K. L.	BS 10: 326-7 (Jy'65)
Jung, A. F.	AS 10: 93-97 (Je'61)
Kain, J. F.	RES 46: 55-64 (F'64)
Kaplan, R. J. +	IEEE 7 HFE: 49-63 (Mr'66)
Karreman, H. F.	AD - 418321 (Jy'63)
Keeping, E. S.	"Introduction to Statistical Inference" Van Nostrand ('62)
Kingston, C. R.	ASAJ 60: 1028-34 (D'65)
Klingel, A. R. +	AS 11: 79-92 (Je'62)
Koehler, T. L.	CEP 55: 76-9 (O'59)
Kossack, C. F.	IBMJ 2: 136-51 (Je'63)
Leckie, D. S. +	IQC 11: 3-8 (Je'55)
Ledley, R. S.	"Use of Computers in Biology and Medicine" McGraw-Hill ('65)
Levy, S. +	JB 35: 335-42 (O'62)
Lingoes, J. C.	BS 8: 267 (Jy'63)
Lingoes, (2), J. C.	BS 8: 368, 370 (O'63)
Lingoes, (3), J. C.	BS 9: 87-88 (Ja'64)
Locke, L. G.	IQC 20: 18-22 (Ap'64)
Lubin, A.	ARS 13: 345-70 ('62)
Lund, I. A.	AMS 17: 12-13 (Ap'63)
Martin, P. S. +	AJS 263: 313-58 ('65)
Massey, F. +	BS 9: 88-89 (Ja'64)
Massey, W. F.	ASAJ 60: 234-56 (Mr'65)
McDonald, R. P.	AD-623205 (Ag'65)
McElrath, D.	ASAP: 137-42 ('64)
Meredith, W.	PS 28: 239-58 (S'63)
Meredith, (2), W.	PS 29: 55-65 (Mr'64)
Michael, W. B. +	RER 33: 451-586 (D'63)
Michaels, S. E.	AS 13: 221-35 ('64)
Miller, R. L. +	"Statistical Analysis in the Geological Sciences" Wiley ('62)
Mittler, A.	FT 16: 22-27 (Mr'62)
Moore, D. W. +	EPM 24: 675-6 (Fall'64)

Moore, P. G.	S 14: 291-309 ('64)
Mustonen, S.	SML 4: 37-44 ('65)
Myers, J. H.	JAP 48: 168-70 (Je'64)
Nelson, J. E.	CACM 8: 313-4 (My'65)
Neidell, N. S.	G 31 122-34 (F'66)
Novick, M. R. +	ASAJ 60: 81-91 (Mr'65)
Osborne, R. K. +	AD - 419073 (Jy'63)
Ostle, B.	"Statistics in Research" Iowa State U. Press ('63)
Overall, J. E. +	PS 28: 415-22 (D'63)
Peake, R. E.	AS 2: 184-92 (N'53)
Pearce, S. C. +	AS 9: 1-7 (Mr'60)
Pehlke, R. D. +	"Computer Applications in Metallurgical Engineering" Am. Soc. for Metals ('64)
Peterson, R. W.	HPPR 44: 261-4 (My'65)
Priest, R. F.	BS 9: 291 (Jy'64)
Radhakrishna, S.	S 14: 147-167 ('64)
Rapping, L.	RES 47: 81-86 (F'65)
Reimer, R.	RES 46: 65-75 (F'64)
Reiter, E. R. +	AD - 604436 (Jy'64)
Rodin, E. A. +	IEEE 11BME: 19-23 (Ap'64)
Schafer, R. E.	IQC 21: 118-22 (S'64)
Schmid, L. P.	CACM 8: 115 (F'65)
Schonemann, P. H.	PS 31: 235-48 (Je'66)
Shiskin, J. +	U. S. Government Printing Office Series P-50 No. 82
Simpson, S. M.	AD-299069 (D'62)
Sobotka, S. P. +	JB 34: 10-30 (Ja'61)
Sokal, R. R. +	"Principles of Numerical Taxonomy" W. H. Freeman ('63)
Sorenson, R. C. +	AD-413481 (Ja'63)
Sorkin, A. +	BS 9: 377-8 (O'64)
Spicer, C. C.	S:14: 137-45 ('64)
Sprinkle, J. K. +	JM 17: 313-17 (Mr'65)
Spurrell, D. J.	AS 12: 180-88 ('63)
Steidler, F. E.	BS 8: 173-5 (Ap'63)
Sterling, T. +	"Medcomp, Part 1" U of Cincinatti ('63)
Sterling, (2), T. D. +	"Computers and the Life Sciences" Columbia U. Press ('65)
Thomas, B. A. M.	AS 10:1-8 (Mr'61)
Truax, H. M.	IQC 18: 18-25 (D'61)
Turk, H. +	BS 9: 291-2 (Jy'64)
Wagle,	S 15: 191-96 ('65)
Walsh, J. A.	EPM 24: 659-62 (Fall'64)
Ward, J. H. +	BS 7: 120-26 (Ja'62)
Weaver, W. W. +	JC 13: 252-61 (D'63)
Webb, S. R.	AD - 618518 (Je'65)
Whelan, J. N.	BS 8: 80-82 (Ja'65)
Whitfield, H.	BIS 19: 368 (Je'63)
Whittlesey, J. R. B.	SIAM Review 6: 260-64 (Jy'64)
Williams, B. J. +	BS 8: 82 (Ja'63)
Winchell, H.	AMI 47: 411-13 (Mr '62)
Winters, W. K.	JACM 12: 356-63 (Jy'65)
Wolf, R.	BS 8: 369-70 (O'63)
Wood, S. R. +	IQC 20: 14-18 (Je'64)
Worsley, B. H.	CACM 7:39-44 (Ja'64)
Wu, S. M. +	ASME 86B: 150-56 (My'64)

Printed in U.S.A. C20-1645

IBM®

International Business Machines Corporation
Data Processing Division
112 East Post Road, White Plains, N.Y. 10601
(USA Only)

IBM World Trade Corporation
821 United Nations Plaza, New York, New York 10017
(International)