physical sciences data 39

statistical methods in applied chemistry

physical sciences data 39

statistical methods in applied chemistry

jurand czermiński
University of Gdańsk, Poland

andrzej iwasiewicz
The Cracow Academy of Economics, Poland

zbigniew paszek
The Cracow Academy of Economics, Poland

andrzej sikorski
The Cracow Academy of Economics, Poland

ELSEVIER
Amsterdam–Oxford–New York–Tokyo
PWN-Polish Scientific Publishers
Warszawa
1990

Translated by Elżbieta Krasodomska from the Polish revised and enlarged edition *Metody statystyczne dla chemików*, published by Państwowe Wydawnictwo Naukowe, Warszawa 1986

Distribution of this book is being handled by the following publishers:
For the USA and Canada
ELSEVIER SCIENCE PUBLISHING CO., INC.
655 Avenue of the Americas
New York, N. Y. 10010

For Albania, Bulgaria, Cuba, Czechoslovakia, German Democratic Republic, Hungary, Korean People's Democratic Republic, Mongolia, People's Republic of China, Poland, Romania, the USSR, Vietnam and Yugoslavia
ARS POLONA
Krakowskie Przedmieście 7, 00-068 Warszawa, Poland

For all remaining areas
ELSEVIER SCIENCE PUBLISHERS B. V.
Sara Burgerhartstraat 25
P. O. Box 211, 1000 AE Amsterdam, The Netherlands

ISBN 0-444-98862-9 (vol.39)
ISBN 0-444-41689-7 (series)

Printed in Poland by D.R.P.

"I can see it but I do not believe it"
GEORG CANTOR about a result

Instead of a foreword

A few words on the need and use of probabilistic thinking

Mathematical statistics is known to examine variation using calculus of probability methods in order to find some regularities which occur among random events. Random events are, in turn, characterized by random variables. It is their realizations that we encounter in experimentation, hence the possibility or even necessity (we do not hesitate to use this emphatic expression) to apply statistical methods.

This is particularly the case where a researcher wishes to assign to his decisions a numerically defined probability of their being accurate (or inaccurate, for that matter). It cannot be done without using calculus of probability.

Although such is the principal use of statistical methods, it is by no means the only one. We shall not proceed to enumerate the benefits which are derived from applying the methods or to discuss them.[1] We shall only mention some of them, i.e. the possibility of detecting the significance of differences between sets of observations, of answering questions concerning the real effect of variables, their correlation or interaction, etc.

We share the opinion of W. Hamilton and J. Hadamard[2] who state that expressing a thought by means of words makes it stable. In our opinion, the application of probability as a measure of the researcher's risk makes the results of his experiments stable. It can be easily seen that such evaluation of experi-

[1] In our book, we do not shun discussion of a number of advanced problems even though we know that many people regard discussion as a way of disagreement which only makes them more confirmed in their errors.

[2] J. Hadamard, *An Essay on the Psychology of Invention in the Mathematical Field*, Dover Publications, 1954.

ments is not only a necessary part of a research method but it also facilitates realization of the aims of science which are assumed to be the basic ones, i.e., explaining and predicting.

To detect a regularity in variation is our main task as experimenters. This is the "integrating spark" which enables us to see the subject of chemical research in probabilistic perspective.

Acknowledgements

We owe special gratitude to Professor A. K. Jonscher, University of London, who has continuously supported and encouraged this work. We are indebted to Professor T. M. Krygowski, A. G. Dickson, PhD, who made many valuable suggestions concerning Chapter 5. We are obliged to the staff of the Computer Centre at Gdańsk University for their kindness and assistance in preparing the material which comprises this chapter. It is our pleasure to acknowledge the contribution that Mrs J. Kosmulska, M. Sc. made to this text at its final stage of preparation.

The Authors

List of symbols

Chapters 1 and 2

Ω —set of elementary events
$A, B, C, \ldots$—random events
$\bar{A}$ —event opposite to A
$\mathscr{B}$ —Borel's body of events
$P(A)$ —probability of random event A
$P(A|B)$ —conditional probability of random event A
A_n —sequence of random events
$V_n(A)$ —relative frequency of random event A
$X, Y, Z, \ldots$ —random variables
$x, y, z, \ldots$ —realizations of random variables
$F(x)$ —distribution function of random variable
$f(x)$ —probability density function
$E(x), \mu_x$ —expected value of general population
D^2X, σ_x^2 —general population variance
$D(X), \sigma_x$ —standard deviation of general population
$\varphi(u)$ —probability density of standardized variable
$\Phi(u)$ —distribution function of standardized random variable
$\theta(u)$ —Laplace's function
$H(x)$ —error function
$\Gamma(p)$ —Euler's function
$(X, Y), (Y, T), (X_1, X_2) \ldots$—two-dimensional random variables
$F(x, y)$, —distribution function of two-dimensional random variable
$f(x, y)$ —probability density function of two-dimensional random variable
$f_1(x), f_2(y)$—marginal probabilities of random variables (X, Y)
$F_1(x), F_2(x)$—distribution functions of marginal distributions of random variables (X, Y)
ϱ —linear correlation coefficient of probability distribution of two-dimensional random variable (X, Y)
$\bar{x}$ —sample arithmetic mean
$\tilde{x}$ —median
$\hat{x}$ —modal value
r_n —sample range
s_x^2 —sample variance
s_x —standard deviation
v_x —variation coefficient

n —sample size
N —size of general population
$\mathrm{cov}(X, Y)$ —covariance of variables X, Y
χ^2 —chi-square random variable
t —random variable with Student's t distribution
F —random variable with F-distribution
Q —value of general population parameter being estimated
α —level of significance
γ_2 —confidence coefficient
$\sigma_{\bar{x}}^2$ —mean error of arithmetic mean

Chapter 3

H_0 —null hypothesis
H_1, H_{-1} —alternative hypotheses
α —probability of Type I error; level of significance of test
β —probability of Type II error
D —statistic of H_0 (general denotation)
D_k —critical region (general denotation)
u —statistic of H_0 in $N(0; 1)$ distribution test
t —statistic of H_0 in Student's test
λ —statistic of H_0 in chi-square test
F —statistic of H_0 in F-test
$F_{\max}$ —statistic of H_0 in Hartley's test
v —statistic of H_0 in Aspin–Welch test
U^2 —statistic of H_0 in Bartlett's test
r_n —statistic of H_0 in sign test of differences
T_n —statistic of H_0 in rank sign test of differences
$\bar{x}$ —sample arithmetic mean; also control chart used for controlling the expected value of random variable X with $N(\mu_x, \sigma_x)$ distribution
x_g, x_d — upper and lower control limits of chart $\bar{x}$
$\bar{x}-s$ —two-line control chart used for controlling the expected value and standard deviation of random variable X with $N(\mu_x, \sigma_x)$ distribution
s_g, s_d —upper and lower control limits for lines s of control chart $\bar{x}-s$
$\bar{x}-r$ —two-line control line for controlling the expected value and range of random variable $X-N(\mu_x, \sigma_x)$ by means of arithmetic mean $\bar{x}$ and sample range (r_n)
$r_{n,g}, r_{n,d}$ —upper and lower control limits on line r of control chart $\bar{x}-r$
np —control chart used for controlling the number of defectives in a sample with determined size
z_g, z_d —upper and lower control limits of chart np
p —control chart for controlling sample fraction defectives
w_g w_d, —upper and lower control limits of chart p
X, Y, Z, U—random variables
U —standardized random variable with $N(0; 1)$ distribution

Chapter 4

η_i —effect of the ith object on the value of observation x_{ij} or x_{ipj}
$\hat{\eta}_i$ —estimate (e.g. $\hat{\eta}_i$ estimate of effect η_i)
e_{ij}, e_{ipj} —random error (deviation)
k —number of columns
r —number of rows
n_i —number of observations in the ith object
n —number of observations when $n_i = n$
S —total sum of squares
S_{ij} —sum of squares within objects
S_i —sum of squares among objects (columns)
σ_{η}^2 —denotes contribution of systematic variation of means to variation among objects
μ_i —mean (real) value of the ith object
μ —total real value
$\bar{x}_i$ —arithmetic mean of the ith object (column)
$\bar{\bar{x}}$ —total arithmetic mean
σ_i —variance of the ith general population
s_i^2 —estimate of variance σ_i^2
F_0 —computed value of F-test
F_φ —tabulated value of F (boundary value)
$s_{\bar{x}_i}^2$ —estimate of mean variance
$\bar{x}_p$ —group mean
d_L —difference $\bar{x}_i - \bar{x}_p$
μ_0 —determined value
ϑ_p —effect of the pth object on the value of observation x_{ipj}
λ_{ip} —measure of interaction of factors A and B, λ_{ip} effect
$\bar{x}_p$ —mean of the pth object (row); group mean in Tukey's method
$\bar{x}_{ip}$ —mean of subclass $a_i b_p$
a_i —levels of factor A
b_p —levels of factor B
σ_v^2 —represents contribution of systematic variation of means μ_p to variation among objects (rows)
σ_λ —interaction variance
S_p —sum of squares among objects (rows)
S_{ip} —sum of squares for interaction
S_{ipj} —sum of squares for subclasses $a_i b_p$
w_i —weight of the ith object
$\bar{x}_w$ —weighted total mean
f_1, f_2 —degrees of freedom

Chapter 5

B —matrix constructed from table of values of explanatory variable expanded by vector composed of ones only
C —matrix of system of normal equations
D —dielectric permeability
E —extinction
E_S —steric component of free energy excess
$E(X)$ —expected value of random variable X
F_D —free energy of ion transmission between environments with different permittivity
$F(X, \beta)$ —function defining exact model of experiment
H_0 —null hypothesis
H —Hessian matrix
J —Jacobian matrix
K —dissociation constant
K_s —association constant
L —likelihood function
L —matrix of sample central moments of H order (estimate of matrix)
N —number of measuring points
P —correlation matrix
Q —non-negative measure which is a superposition of several r types errors
Q'_{b_i} —first derivative of function Q in relation to b_i
$Q''_{b_i b_j}$ —second derivative of function Q in relation to b_i and b_j
R —matrix of estimates of sample correlation coefficients
R —positive definite matrix generating direction vector in gradient methods
V —covariance matrix
b —vector of parameter estimates β_i
$f(x)$, $g(x)$, $h(x)$—functions of one real variable
$f_+(x)$ —upper confidence curve
$f_-(x)$ —lower confidence curve
e —vector of residuals
k —reaction-velocity constant
h —hyperconjugation constant of free energy excess
h_{ij} —elements of Hessian matrix
l —vector of right sides of system of normal equations
l_{ij} —elements of **L** matrix
n —number of parameters under estimation
Δn —difference between numbers of hydrogen atoms involved in hyperconjugation
n_i —mean ligand number
$\mathbf{q}_i$ —gradient calculated at the ith iterative step
r —sample correlation coefficient
t —Student's test statistic
w —vector of statistical weights
v_i —direction vector in gradient methods at the ith iterative step
x —vector of explanatory variables of a model

$\bar{x}$ – sample mean value (arithmetic mean)
v —vector of explanatory variables
g_{ik} —elements of matrix Λ^{-1}
α —level of significance
β —vector of model parameters under estimation
β_i —complex constants in the Rosottis equation
δ —proportionality coefficient of steric component
$\delta(X, Y)$ —measure of dependence of random variables X and Y
ε —error vector
λ —multiplier controlling convergence of Marquardt's method
λ —equivalent conductivity
λ_{ij} —elements of matrix Λ
Λ —matrix of II order central moments
$\varphi_i(x)$ —orthogonal polynomial of the ith degree
Ψ —component of free energy excess due to π-electron substructure coupling with reaction centre
θ_{yx} —correlation ratio (formula (5.3))
Σ —diagonal matrix composed of standard deviations
ξ —random variable whose realizations are errors ε_i
dim(x) —dimension of space, number of components of vector x
sign(x) —signum function (sign)
$[L]$ —equilibrium concentration of ligand in the Rosotti's equation

There are no new symbols in Chapters 6 and 7.

Contents