BINGHAM McCUTCHEN LLP
DONN P. PICKETT (SBN 72257)
GEOFFREY M. HOWARD (SBN 157468)
HOLLY A. HOUSE (SBN 136045)
ZACHARY J. ALINDER (SBN 209009)
BREE HANN (SBN 215695)
Three Embarcadero Center
San Francisco, CA 94111-4067
Telephone: (415) 393-2000
Facsimile: (415) 393-2286
donn.pickett@bingham.com
geoff.howard@bingham.com
holly.house@bingham.com
zachary.alinder@bingham.com
bree.hann@bingham.com

BOIES, SCHILLER & FLEXNER LLP
DAVID BOIES (Admitted *Pro Hac Vice*)
333 Main Street
Armonk, NY 10504
Telephone: (914) 749-8200
Facsimile: (914) 749-8300
dboies@bsfllp.com
STEVEN C. HOLTZMAN (SBN 144177)
FRED NORTON (SBN 224725)
1999 Harrison St., Suite 900
Oakland, CA 94612
Telephone: (510) 874-1000
Facsimile: (510) 874-1460
sholtzman@bsfllp.com
fnorton@bsfllp.com

DORIAN DALEY (SBN 129049)
JENNIFER GLOSS (SBN 154227)
500 Oracle Parkway, M/S 5op7
Redwood City, CA 94070
Telephone: (650) 506-4846
Facsimile: (650) 506-7114
dorian.daley@oracle.com
jennifer.gloss@oracle.com

Attorneys for Plaintiffs Oracle USA, Inc., *et al.*

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
OAKLAND DIVISION

ORACLE USA, INC., *et al.*, Plaintiffs, v. SAP AG, *et al.*, Defendants.	CASE NO. 07-CV-01658 PJH (EDL) **EXHIBIT 5 TO THE DECLARATION OF DANIEL S. LEVY, PH.D. IN SUPPORT OF MOTION NO. 1: TO EXCLUDE TESTIMONY OF DEFENDANTS' EXPERT STEPHEN CLARKE** **FILED PURSUANT TO DKT. NO. 915**

EXHIBIT 5

☆ ☆ ☆ ☆ ☆

Sixth Edition

ME

GEORGE W. SNEDECOR

Professor Emeritus of Statistics
and Former Director, Statistical Laboratory
Iowa State University

WILLIAM G. COCHRAN

Professor of Statistics
Harvard University

HE IOWA STATE UNIVERSITY PRESS

AMES, IOWA, U.S.A.

George W. Snedecor is professor emeritus of statistics, Iowa State University, where he taught from 1913 to 1958 and where he was for fourteen years director of the statistical laboratory. His writings include a body of scientific journal articles, research bulletins, and books, including *Correlation and Machine Calculation* (with H. A. Wallace), *Calculation and Interpretation of Analysis of Variance and Covariance*, and *Statistical Methods*. He holds a master of science degree from the University of Michigan, and honorary doctor of science degrees from North Carolina State University and Iowa State University. He is a member of the International Statistical Institute, past president of the American Statistical Association, and an honorary Fellow of the British Royal Statistical Society. He has served also as consultant, Human Factors Division, U.S. Navy Electronics Laboratory, San Diego, California, where he now lives.

William G. Cochran is professor of statistics, Harvard University. He has served formerly on the faculties of Johns Hopkins University, North Carolina State University, and Iowa State University. He holds master of arts degrees from Glasgow University and Cambridge University and an honorary master of arts degree from Harvard University. He is past president of the American Statistical Association, the Institute of Mathematical Statistics, and the Biometric Society. His writings include many research papers in the professional journals of his field; *Sampling Techniques*, 2nd ed., 1963; and *Experimental Designs* (with Gertrude M. Cox), 2nd ed., 1957.

Printed in the U.S.A.

First edition, 1937
Second edition, 1938
Third edition, 1940

Fourth edition, 1946
Second printing, 1946
Third printing, 1948
Fourth printing, 1950
Fifth printing, 1953
Sixth printing, 1955
Seventh printing, 1955

Fifth edition, 1956
Second printing, 1957
Third printing, 1959
Fourth printing, 1961
Fifth printing, 1962
Sixth printing, 1964
Seventh printing, 1965
Eighth printing, 1966

Sixth edition, 1967
Second printing, 1968
Third printing, 1969
Fourth printing, 1971
Fifth printing, 1972
Sixth printing, 1973
Seventh printing, 1974
Eighth printing, 1976
Ninth printing, 1978
Tenth printing, 1979

International Standard Book Number: 0 8138 1560 6
Library of Congress Catalog Card Number: 67-21577

Prefac

In preparing the sixth edition we have kept in mind the two purpo this book has served during the past thirty years. Prior editions have b used extensively both as texts for introductory courses in statistics and reference sources of statistical techniques helpful to research workers the interpretation of their data.

As a text, the book contains ample material for a course extend throughout the academic year. For a one-term course, a suggested of topics is given on the page preceding the Table of Contents. As past editions, the mathematical level required involves little more tl elementary algebra. Dependence on mathematical symbols has b kept to a minimum. We realize, however, that it is hard for the reader use a formula with full confidence until he has been given proof of formula or its derivation. Consequently, we have tried to help the read understanding of important formulas either by giving an algebraic pr where this is feasible or by explaining on common-sense grounds the r played by different parts of the formula.

This edition retains also one of the characteristic features of book – the extensive use of experimental sampling to familiarize the rea with the basic sampling distributions that underlie modern statisti practice. Indeed, with the advent of electronic computers, experimer sampling in its own right has become much more widely recognized a research weapon for solving problems beyond the current skills of mathematician.

Some changes have been made in the structure of the chapters, mai at the suggestion of teachers who have used the book as a text. The forn chapter 8 (Large Sample Methods) has disappeared, the retained mate being placed in earlier chapters. The new chapter 8 opens with an int duction to probability, followed by the binomial and Poisson distributi (formerly in chapter 16). The discussion of multiple regression (chap 13) now precedes that of covariance and multiple covariance (chapter 1 termined levels X_1', X_2', . . . on a voltmeter. Owing to errors in the voltmeter or other defects in the apparatus, the true voltages X_1, X_2, . . . differ from the set voltages.

In this situation we still have $Y = \alpha + \beta X + \varepsilon$, $X' = X + e$. In both our original case (X normal) and in Berkson's case (X' fixed) it follows that

$$Y = \alpha + \beta X' + (\varepsilon - \beta e) \qquad (6.17.3)$$

The difference is this. In our case, e and X' are correlated because of the relation $X' = X + e$. Consequently, the residual $(\varepsilon - \beta e)$ is correlated with X' and does not have a mean zero for fixed X'. This vitiates Assumption 2 of the basic model (section 6.4). With X' fixed, however, e is correlated with X but not with X', and the model (6.17.3) satisfies the assumptions for a linear regression. The important practical conclusion is that b', the regression of Y on X', remains an unbiased estimate of β.

6.18—Fitting a straight line through the origin. From some data the nature of the variable Y and X makes it clear that when $X = 0$, Y must be 0. If a straight line regression appears to be a satisfactory fit, we have the relation

$$Y = \beta X + \varepsilon$$

where, in the simplest situations, the residual ε follows $\mathcal{N}(0, \sigma^2)$. The least squares estimate of β is $b = \Sigma XY/\Sigma X^2$. The residual mean square is

$$s_{y \cdot x}^2 = \{\Sigma Y^2 - (\Sigma XY)^2/\Sigma X^2\}/(n - 1)$$

with $(n - 1)$ $d.f.$ Confidence limits for β are

$$b \pm t s_b$$

where t is read from the t-table with $(n - 1)$ $d.f.$ and the appropriate probability.

This model should not be adopted without careful inspection of the data, since complications can arise. If the sample values of X are all some distance from zero, plotting may show that a straight line through the origin is a poor fit, although a straight line that is not forced to go through the origin seems adequate. The explanation may be that the population relation between Y and X is curved, the curvature being marked near zero but slight in the range within which X has been measured. A straight line of the form $(a + bx)$ will then be a good approximation within the sample range, though untrustworthy for extrapolation. If the mathematical form of the curved relation is known, it may be fitted by methods outlined in chapter 15.

It is sometimes useful to test the null hypothesis that the line, assumed straight, goes through the origin. The first step is to fit the usual two-parameter line $(\alpha + \beta x)$, i.e., $\alpha + \beta(X - \bar{X})$, by the methods given earlier in this chapter. The condition that the population line goes

through the origin is $\alpha - \beta\bar{X} = 0$. The sample estimate of this quantity is $\bar{Y} - b\bar{X}$, with estimated variance

$$s_{y \cdot x}^2(1/n + \bar{X}^2/\Sigma x^2)$$

Hence, the value of t for the test of significance is

$$t = \frac{\bar{Y} - b\bar{X}}{s_{y \cdot x}\sqrt{\{1/n + \bar{X}^2/\Sigma x^2\}}} \qquad (6.18.1)$$

with $(n - 2)$ $d.f.$ This test is a particular case of the technique presented in section 6.11 for finding confidence limits for the population mean value of Y corresponding to a given value of X.

The following example comes from a study (9) of the forces necessary to draw plows at the speeds commonly attained by tractors. Those results of the regression calculations that are needed are shown under table 6.18.1.

TABLE 6.18.1

DRAFT AND SPEED OF PLOWS DRAWN BY TRACTORS

Draft (lbs.)	Y	425	420	480	495	540	530	590	610	690	680
Speed (m.p.h.)	X	0.9	1.3	2.0	2.7	3.4	3.4	4.1	5.2	5.5	6.0

$\bar{X} = 3.45$ m.p.h. $\quad \bar{Y} = 546$ lbs. $\quad n = 10$

$\Sigma x^2 = 27.985 \quad \Sigma y^2 = 82{,}490 \quad \Sigma xy = 1{,}492.0$

$b = 53.31$ lbs. per mile

$s_{y \cdot x}^2 = 368.1$ with 8 $d.f.$

One might suggest that the line should go through the origin, since when the plow is not moving there is no draft. However, inspection of table 6.18.1, or a plot of the points, makes it clear that when the line is extrapolated to $X = 0$, the predicted Y is well above 0, as would be expected since inertia must be overcome to get the plow moving. From (6.18.1) we have

$$t = \frac{546 - (53.31)(3.45)}{\sqrt{\left[(368.1)\left\{\frac{1}{10} + \frac{(3.45)^2}{27.985}\right\}\right]}} = \frac{362.1}{13.90} = 26.0$$

with 8 $d.f.$, confirming that the line does not go through the origin.

When the line is straight and passes through (0, 0), the variance of the residual ε is sometimes not constant, but increases as X moves away from zero. On plotting, the points lie close to the line when X is small but diverge from it as X increases. The extension of the method of least squares to this case gives the estimate $b = \Sigma w_x XY/\Sigma w_x X^2$, where w_x is the reciprocal of the variance of ε at the value of X in question.

If numerous observations of Y have been made at each selected X the variance of ε can be estimated directly for each X and the form of the