

Don't Panic!

A Guide to Statistical Methods
for Outbreak Investigation

Maria Quigley & Mark Myatt
Brixton Books

First Published in 1994 by Brixton Books, PO Box 4398, London SW9 9XJ

This edition published 1994

Copyright © 1994 by Maria Quigley & Mark Myatt

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, without the prior permission of the publisher, nor otherwise circulated in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

LOGISTIC is provided 'as is' without warranty of any kind. The entire risk as to the quality, performance, and fitness for intended purpose is with you. You assume responsibility for the selection of the program and for the use of results obtained from that program. LOGISTIC should not be modified in any way without permission from the author.

A CIP catalogue record for this book is available from the British Library

ISBN 1-873937-26-1

Trademarks mentioned in this book belong to the following organisations and persons:

Apple Computer: Macintosh
Borland International: dBase
International Business Machines: OS/2
Insignia Solutions: SoftPC
Microsoft: MSDOS, Microsoft Windows
Gerard Dallal: StatTools

Printed and bound by: St Edmundsbury Press, Bury St Edmunds, Suffolk, England

We would like to thank the participants of the courses held at the University of Manchester Medical School, London School of Hygiene and Tropical Medicine, and St. George's Hospital Medical School on whom the material for this book was piloted. Gerard Dallal kindly gave his permission to include LOGISTIC as part of the course software. Datasets were supplied by Jacques Pepin, Université de Sherbrooke (GUDHIV.REC) and David Baxter, University of Manchester and Stockport Health Authority (SALEX.REC). We would also like to thank Ursula Werneke who suggested the title for this book.

Workers with research experience would appreciate a book which, without entering into the mathematical theory of statistical methods, should embody the latest results of that theory, presenting them in the form of practical procedures appropriate to those types of data with which research workers are concerned. The practical application of general theorems is a different art from their establishment by mathematical proof. It requires fully as deep an understanding of their meaning, and is, moreover, useful to many to whom the other is unnecessary.

R.A. Fisher, 1941

Contents

Introduction	1
Typographical conventions	2
The course software	3
Sample datasets	4
Measures of disease and exposure effects: 2-by-2 tables	7
Absolute risk	8
Exposure-specific risk or attack rate	9
Relative risk or risk ratio	10
Odds ratio	11
Epi-Info ANALYSIS	12
Examining data	15
Using ANALYSIS to produce 2-by-2 tables	16
Confidence intervals	18
Testing a hypothesis about association	20
Relative risk, confidence intervals and tests with ANALYSIS	24
Epi-Info STATCALC	28
Analysing data with more than one level of exposure	31
The chi-square test of association	31
The chi-square test for linear trend	33
Non-linear data	36
Non-ordered data	36
Using STATCALC to calculate a chi-square test for trend	37
Analysing two exposure variables at the same time	38
The BATEMAN outbreak	38
Confounded associations	41
The Mantel-Haenszel relative risk and significance test	43
Testing for an interaction: Woolf's test	45
Stratified analysis with ANALYSIS	47
Stratified analysis with STATCALC	49
Analysing several exposures at the same time	52
The SALEX outbreak	52
The logistic model for a single exposure variable	54
Estimating the coefficients in the model	56
Fitting the model using LOGISTIC	57
Significance tests	58
Starting LOGISTIC	60
Using LOGISTIC to fit a logistic regression model	61
The logistic model for two exposure variables	63
Significance tests for the model with two exposure variables	64
Interactions	65
Using LOGISTIC to fit a model with two exposure variables	66
Fitting a model with an interaction term	68
The logistic model for several exposure variables	69
Using LOGISTIC to fit a model with several exposure variables	72
Enter, backward elimination and forward selection	77
An epidemiological approach	78
Analysing data with more than one level of exposure	79
Indicator variables	84
A strategy of analysis	87
Presentation of results	88

Index

2-by-2 tables	7	Mantel-Haenszel, relative risk & significance	43,44
Absolute risk	8,24	Maximum likelihood	63
add , command in LOGISTIC	62	model , command in LOGISTIC	61
Alternative hypothesis	20	Multi-factorial, disease	41
Analysis, stratified	47	Multiple logistic regression	53
ANALYSIS, dBase files	14	Non-linear, data	36,83
ANALYSIS, described	13	Non-linear, trend	35,81
ANALYSIS, entering commands	13	Non-ordered, data	36,84
ANALYSIS, freq command	15	Null hypothesis	20
ANALYSIS, help system	13	Observed values	20
ANALYSIS, output	15	Odds	11,54
ANALYSIS, quit command	17	Odds ratio	11
ANALYSIS, read command	14	Odds ratio, adjusted	44,63
ANALYSIS, retrieving a data file	14	Odds ratio, and coefficients	56
ANALYSIS, set lines command	24	Odds ratio, confidence interval	19
ANALYSIS, set percents command	24	Odds, logarithm of	54,55
ANALYSIS, set statistics command	25	Odds, of disease	54
ANALYSIS, stratified analysis	47	P-values	22
ANALYSIS, variables command	15	P-values, one and two-tailed tests	26
Attack rate	9,24	Percentages	15
Backward elimination, method in logistic	77	Percentages, in tables	24
Case-control study	8,11	Population	18
Chi-square statistic, calculation	21,22	Protective effect, food-borne outbreak	10,26
Chi-square, distributions	22,23	quit , command in ANALYSIS	17
Chi-square, Mantel-Haenszel	25,43	quit , command in LOGISTIC	62
Chi-square, Pearson's	22	Random variation	18
Chi-square, test for trend	31,33-37	read , command in ANALYSIS	14
Chi-square, Wald's test	59	read , command in LOGISTIC	61
Chi-square, Woolf's test	45	Relative Risk	10,11
Chi-square, Yates' corrected	22	Relative risk, adjusted	44
Coefficients, and odds ratio	56	Relative risk, confidence intervals	18,19
Coefficients, estimating	56,57	Relative risk, crude	44
Cohort study	8,11	Relative risk, Mantel-Haenszel	44
Confidence interval, interpretation of	18	Relative risk, point estimate	18
Confidence intervals, defined	18	Relative risk, pooled	43,44
Confidence intervals, exact	51	Relative risk, population	18
Confidence level, defined	18	Relative risk, summary	44
Confidence limits, defined	18	Relative risk, weighted average	43,44
Confidence interval, significance	27	Removing terms from models	73-78
Confounding	41-44	Risk, absolute	8,24
Course software, starting & quitting	3	Risk, excess	10,11,56
Cross-sectional study	8,11	Risk, exposure specific	9,24
Data, non-linear	36,83	Risk, ratio	10,11
Data, non-ordered	36,84	Risk, relative	10,11
dBase files, retrieving in ANALYSIS	14	Sample size	29-30
Degrees of freedom	22	Significance testing	20-22,58,71
delete , command in LOGISTIC	73	STATCALC, 2-by-2 tables	29,30
Effect modification	41,46	STATCALC, chi-square for trend	32,37
Enter, method of logistic regression	77	STATCALC, described	28
Estimate, command in LOGISTIC	61	STATCALC, starting	28
Estimating, coefficients	56	STATCALC, stratified analysis	49-51
Estimation, approach	18, 27	Stratified analysis	43,47-51
Excess risk	10,11,56	Susceptibility	18
Expected values, calculation	20	tables , command in ANALYSIS	16
Exposure specific risk	9,24	tables , format of command	16,25,47
Fisher's exact test	22	Tables, notation defined	7
Forward selection, method in logistic regression	77	Tables, separating lines	24
freq , command in ANALYSIS	15,38	Tables, showing percentages	24
Hypothesis, alternative	20	Tables, showing statistics	25
Hypothesis, null	20	Terms, adding to models	62
Hypothesis, statistical testing	20	Terms, removing from models	73
Indicator variables	84-86	Tests, association	31
Interaction	41,45-46,65-68	Tests, chi-square	21,22
Likelihood ratio test	58	Tests, dose-response	31,33-37
Log Odds	55	Tests, exact	22
Logistic regression & logistic models	54-56	Tests, interaction	58
LOGISTIC, add Command	62	Tests, likelihood ratio	58
LOGISTIC, delete command	73	Tests, linear trend	31,33-37
LOGISTIC, estimate command	61	Tests, significance	20
LOGISTIC, interaction terms	65,68	Tests, Wald's	59
LOGISTIC, lr command	62	Trend, test for in STATCALC	32
LOGISTIC, model command	61	Trends, linear & non-linear	34,35,81-83
LOGISTIC, quit command	62	variables , command in ANALYSIS	15
LOGISTIC, read command	61	Variation, random	18,20
LOGISTIC, starting	60	Wald's test	59
lr , command in LOGISTIC	62	Woolf's chi-square	45,46