

Peter Lewis

# **Maps and Statistics**

Peter Lewis

# **Maps and Statistics**

Methuen & Co Ltd

*First published 1977 by Methuen & Co Ltd*  
*11 New Fetter Lane, London EC4P 4EE*  
© 1977 Peter Lewis

*Printed in Great Britain at the*  
*University Printing House, Cambridge*

ISBN 0 416 65370 7 *hardback*  
ISBN 0 416 65380 4 *paperback*

This book is available in both hardbound and paperback editions. The paperback edition is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, resold, hired out or otherwise circulated without the publisher's prior consent 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.

**Library of Congress Cataloging in Publication Data**

Lewis, Peter, 1938—  
Maps and statistics.

Bibliography: p. 310  
Includes index.

I. Maps, Statistical. I. Title.

GA109.8.L48 1977      001.4'226      77-1184  
ISBN 0-470-99094-5

# The Field of Geography

Progress in modern geography has brought rapid changes in course work. At the same time the considerable increase in students at colleges and universities has brought a heavy and sometimes intolerable demand on library resources. The need for cheap textbooks introducing techniques, concepts and principles in the many divisions of the subject is growing and is likely to continue to do so. Much post-school teaching is hierarchical, treating the subject at progressively more specialized levels. This series provides textbooks to serve the hierarchy and to provide therefore for a variety of needs. In consequence some of the books may appear to overlap, treating in part of similar principles or problems, but at different levels of generalization. However, it is not our intention to produce a series of exclusive works, the collection of which will provide the reader with a 'complete geography', but rather to serve the needs of today's geography students who mostly require some common general basis together with a selection of specialized studies.

Between the 'old' and the 'new' geographies there is no clear division. There is instead a wide spectrum of ideas and opinions concerning the development of teaching in geography. We hope to show something of that spectrum in the series, but necessarily its existence must create differences of treatment as between authors. There is no general series view or theme. Each book is the product of its author's opinions and must stand on its own merits.

*University of London,  
King's College  
August 1971*

W. B. Morgan  
J. C. Pugh

# Categorization of test procedures

PROPOSITION	MEASUREMENT	
	NOMINAL <i>Class or category</i>	DISCRETE QUANTITY <i>Position in sequence, ranked</i>
RANDOMNESS	Probability distributions Binomial (1.9, 2.2.1) Multinomial (2.2.2) Hypergeometric (2.2.3, 2.3.1) Poisson (2.2.4) Number of runs (2) (3.1.1) Number of runs (1) (3.1.2) Number of runs, linked pairs (3.1.3) Wald-Wolfowitz (3.4.1)	Cox and Stuart (3.3.1) Kolmogorov (3.5.1)
INDEPENDENCE DEPENDENCE ASSOCIATION	Fisher's exact (2.3.1) Chi-square (2.3.2, 2.5, 2.6, 3.5.2) Cramer's coefficient (2.4) Tschuprow's coefficient (2.4) Pearson's contingency coefficient (2.4) Kendall's phi coefficient (2.4) Cross-product ratio (2.4) Chi-square ( $r \times c$ ) (2.5) Chi-square ( $r \times c \times I$ ) (2.6) Log-linear models (2.6)	Chi-square ( $r \times 2$ ) subsets (2.5) Chi-square ( $r \times 2$ ) regression (2.5) Chi-square ( $r \times c \times I$ ) (2.6) Log-linear models (2.6)
EQUALITY OF LOCATION IDENTITY OF DISTRIBUTION GOODNESS-OF-FIT	Sign test (1.9) Wald-Wolfowitz (3.4.1) Westenb�rg-Mood median test (3.4.2) Chi-square (goodness-of-fit) (3.5.2) McNemar's test (Ex. 3.7)	Wilcoxon rank-sum (2.7.1) Mann-Whitney, $U$ (2.7.2) Kruskal-Wallis (2.7.3) Friedman (2.7.4) Matched-pairs sign test (3.4.2) Wilcoxon signed-rank (3.4.3)

Point symbol maps are discussed in section 2. Line symbol maps and angular measurements are discussed in section 3. Numbers in brackets indicate the article in which the test statistic is discussed in detail.

## SCALE

CONTINUOUS QUANTITY	
<i>Distribution-free or not-normal</i>	<i>Von Mises (angular) Normal</i>
Runs up and down, Edgington (3.2.1) Periodicity in runs, Noether (3.2.2) Kolmogorov (3.5.1)	
Cox and Stuart (3.3.1) Spearman's rho (3.3.2) Daniel's test for trend (3.3.2) Kendall's tau (3.3.2) Kendall's concordance, $W$ (3.3.2)	$t$ for $b$ coefficient $t$ for Pearson's $r$ coefficient
Smirnov test (3.4.4) Kolmogorov test (3.5.1) Lilliefors test (3.5.1) Kuiper's test (angular data) (3.6.1) Mardia's uniform scores test (angular data) (3.6.2)	$t$ in 1 and 2 sample cases $F$ in multi-sample cases Goodness-of-fit not needed by assumption Watson and Williams, Rayleigh for angles

# Acknowledgements

I should like to thank the following for permission to reproduce the tables in appendix 2: Addison-Wesley Publishing Company for table 9; Biometrika Trustees for tables 4, 6, 7, 12 and 17; Department of Statistics, Florida State University, for table 14; Institute of Mathematical Statistics, Hayward, California, for tables 10 and 15; *Journal of the American Statistical Association* for tables 3(b), 8, 11 and 16; *Journal of the Royal Statistical Society* for table 18; *Statistica Neerlandica* for table 13; and Stanley Thornes (Publishers) Ltd for tables 1, 2, 3(a) and 5.

Implicit in any book is the debt owed to the people whose work appears in the references. It is a pleasure to offer my thanks to those people who have contributed to the development of this book. Like the other authors in this series I have benefited from the suggestions made so pleasantly by Bill Morgan, the series editor, and by Janice Price, general editor for Methuen. I should like to thank Toby Lewis of Hull University and Joyce Snell of Imperial College, London, for reading the manuscript and making valuable recommendations, and I hope my response to those suggestions has been adequate. My thanks, too, to the technical and secretarial staff of the Department of Geography, Birkbeck College, for producing the fair copies of the illustrations and for typing a difficult manuscript. My deepest thanks happily go to Rosemary and Simon and Catherine for providing the environment for writing this book, for giving so much encouragement during its preparation, and for helping on so many of the jobs that arose at all stages of its formation.

# Contents

Categorization of test procedures	x
List of figures, tables and maps	xii
Acknowledgements	xviii
<b>Section 1 Maps, measurement and probability</b>	<b>1</b>
1.1 The concept of location class	4
1.2 Using location as a quantity	7
1.3 Scale variations in location measurement	10
1.4 Establishing conventions of map inference	14
1.4.1 <i>The proposition</i>	15
1.4.2 <i>The attributes</i>	16
1.4.3 <i>The measurement</i>	16
1.5 Measurement and probability judgements	17
1.5.1 <i>Distribution-free and non-parametric test statistics</i>	17
1.6 Counting rules for probability calculations	19
1.7 Counting and the probability of an event	23
1.8 Random variables	25
1.8.1 <i>Probability distributions</i>	26
1.8.2 <i>Characteristics of random variables</i>	30
1.8.3 <i>Population and sample</i>	35
1.9 Hypothesis testing	37
Exercises	44
<b>Section 2 Analysing point symbol maps</b>	<b>51</b>
2.1 The construction of point symbol maps	53
2.1.1 <i>Bijjective mappings</i>	54
2.1.2 <i>Surjective mappings</i>	55
2.2 Propositions of randomness	57
2.2.1 <i>Randomness and a binomial random variable</i>	59
2.2.2 <i>Randomness and a multinomial random variable</i>	64
2.2.3 <i>Randomness and a hypergeometric random variable</i>	67
2.2.4 <i>Randomness and the Poisson distribution</i>	69



2.3	Propositions of independence (1)	80
2.3.1	<i>Fisher's exact test</i>	81
2.3.2	<i>Chi-square and <math>2 \times 2</math> contingency tables</i>	88
2.4	Measures of association	91
2.4.1	<i>Contingency coefficients</i>	93
2.4.2	<i>Cross-product ratio</i>	96
2.5	Propositions of independence (2)	100
2.5.1	<i>Chi-square and <math>r \times c</math> contingency tables</i>	101
2.5.2	<i>Partitioning the degrees of freedom</i>	108
2.5.3	<i>Independence in incomplete contingency tables</i>	112
2.5.4	<i>Chi-square and ordered categories</i>	118
2.6	Propositions of independence (3)	128
2.6.1	<i>The log-linear model, <math>2 \times 2</math> case</i>	129
2.6.2	<i>The log-linear model, <math>r \times c \times l</math> case.</i>	134
2.7	Propositions of equality of location and identity of distribution	147
2.7.1	<i>Wilcoxon rank-sum test</i>	147
2.7.2	<i>Mann-Whitney U</i>	152
2.7.3	<i>Kruskal-Wallis test</i>	153
2.7.4	<i>Friedman's related samples test</i>	156
	Exercises	160
	<b>Section 3 Analysing line symbol maps</b>	175
3.1	Propositions of randomness (1)	177
3.1.1	<i>Number of runs tests, two attributes</i>	178
3.1.2	<i>Number of runs tests, one attribute</i>	184
3.1.3	<i>Number of paired attributes</i>	186
3.2	Propositions of randomness (2)	188
3.2.1	<i>Edgington's number of runs up and down test</i>	189
3.2.2	<i>Noether's test for cyclic change</i>	193
3.3	Propositions of dependence	195
3.3.1	<i>Cox and Stuart's test for trend</i>	195
3.3.2	<i>Rank correlation tests</i>	198
3.4	Propositions of equality of location and identity of distribution	203
3.4.1	<i>Wald-Wolfowitz number of runs test</i>	204
3.4.2	<i>The matched-pairs sign test</i>	205
3.4.3	<i>Wilcoxon signed-rank test</i>	206
3.4.4	<i>Kolmogorov-Smirnov test statistics: Smirnov's two-sample test</i>	212
3.5	Goodness-of-fit statistics	218
3.5.1	<i>Kolmogorov's test, Lilliefors's modification</i>	220
3.5.2	<i>Chi-square test</i>	226
3.6	Propositions and angular measurements	228
3.6.1	<i>Kuiper's test, a Kolmogorov analogue</i>	229
3.6.2	<i>Uniform scores test and propositions of identity or equality</i>	231
	Exercises	235

Appendix 1: Using the standard normal probability distribution	239
Appendix 2: Tables for distribution-free statistics	243
Answers to exercises	303
Bibliography and further reading	310
Index	315

# Figures, tables and maps

1.1	Objects and location as a class	6
1.2	Objects and location as a quantity	8
1.3	Location as position in a sequence	10
1.4	Location as a quantity on different scales	11
1.5	Counting rule 1	20
1.6	Counting rule 2, case 1	20
1.7	Counting rule 2, case 2	21
1.8	The probability of points in a sample space	23
1.9	Probability distribution function of a discrete random variable	27
1.10	Cumulative distribution function of a discrete random variable	28
1.11	The expected value of a binomial random variable, showing the algebra and the arithmetic of its calculation	31
1.12	Mean pollution levels in a stream before and after legislation	37
1.13	Hypothesis testing and type I, type II errors	40
1.14	Values of $T$ and a rejection region	42
1.15	Probability distribution and a rejection region	43
2.1	Mapping relationships	54
2.2	Industries of Humberside using graphical rational pattern symbols	56
2.3	Changes in density	57
2.4	Changes in pattern	58
2.5	Characterizing pattern	58
2.6(a)	Region of interest and a binomial random variable	60
2.6(b)	Probability distribution of the event in fig. 2.6(a)	61
2.7(a)	Region of interest of arbitrary size	62
2.7(b)	Probability distribution of the event in fig. 2.7(a)	62
2.8	Unconnected region of interest	63
2.9	Region of interest and a multinomial random variable	66
2.10	Exact probabilities and $\chi^2$ approximations for the event in fig. 2.9	66
2.11	Region of interest and a hypergeometric random variable	68

2.12	Arbitrariness of position and random variables considered in this article	69
2.13(a)	Comparison of exact binomial probabilities and their Poisson approximation	74
2.13(b)	Calculating $e^{-\lambda}$ using logarithms	75
2.14	Distribution of homes of Birkbeck undergraduates	76
2.15	Probability distribution of the event in fig. 2.14 as a Poisson random variable	77
2.16	Schematic comparison of different scales	79
2.17	Notation for Fisher's exact test, $2 \times 2$ case	82
2.18	Objects and attributes as a hypergeometric random variable	83
2.19	Summary table for fig. 2.18	83
2.20	Summary tables of all possible $2 \times 2$ values for fig. 2.18 showing their exact probabilities and their $X^2$ approximation	84
2.21	Probability distribution for fig. 2.20	85
2.22	Expected values under $H_0$ for fig. 2.19	89
2.23	Details of $X^2$ calculation for fig. 2.19	90
2.24	Comparison of exact probabilities and $X^2$ values for fig. 2.19	92
2.25	Comparison of <i>cdf</i> and $X^2$ values for fig. 2.19	92
2.26	Summary table for fig. 2.18, repeated	93
2.27	Comparison of $X^2$ , $D_C$ , $CPR$ for fig. 2.26	97
2.28	Steps in reducing a $2 \times 2$ table to standard form	99
2.29	The general $r \times c$ contingency table with alternative notations for marginal totals	102
2.30	Contingency table showing location class and decay class of buildings in part of London	102
2.31	Expected values under $H_0$ for fig. 2.30	104
2.32	$X^2$ contributions of each cell for fig. 2.30	104
2.33	Capacity and distribution of papermaking mills in the United Kingdom, 1965	105
2.34	Contingency table showing capacity class and distance-to-market class of the mills in fig. 2.33	106
2.35	Calculating the degrees of freedom in an $r \times c$ contingency table	107
2.36	Expected values under $H_0$ for fig. 2.34	107
2.37	$X^2$ contributions of each cell for fig. 2.34	107
2.38	The general $r \times c$ contingency table in a notation for partitioning the degrees of freedom	108
2.39	Fig. 2.34 in the format of fig. 2.38	109
2.40	Four $2 \times 2$ subtables of fig. 2.39 corresponding to the four degrees of freedom	109
2.41	Calculating the $X^2$ contribution of subtable (i) in fig. 2.40	110
2.42	Calculating the $X^2$ contribution of subtable (ii) in fig. 2.40	111
2.43	Calculating the $X^2$ contribution of subtable (iii) in fig. 2.40	111
2.44	Calculating the $X^2$ contribution of subtable (iv) in fig. 2.40	111
2.45	Contingency table of papermaking mills according to capacity class in 1915 and 1965	113

2.46	Contingency table of fig. 2.45 reduced by one cell entry	113
2.47(a)	Format for analysis of reduced contingency tables	114
2.47(b)	Fig. 2.46 in the format of fig. 2.47(a)	114
2.48(a)	Expected values under $H_0$ in the reduced table of fig. 2.46	115
2.48(b)	$X^2$ contributions of each cell for fig. 2.46	115
2.49	Fig. 2.45 modified to show papermaking mills changing capacity class between 1915 and 1965	116
2.50	$r \times 2$ contingency table with $r$ ordered classes	119
2.51	Summary format for analysis of $r \times 2$ contingency tables with ordered categories	122
2.52	Initial details required for analysing tables like fig. 2.50	122
2.53	Partitioning fig. 2.50 into two subsets giving $X^2$ contribution for appropriate degrees of freedom	123
2.54	Fig. 2.50 in the format of fig. 2.51	124
2.55	Regression analysis of ordered categories in $r \times 2$ contingency tables; details shown for fig. 2.50	125
2.56	Results of regression analysis of fig. 2.50	126
2.57	Graphical comparison of the observed and expected values under the regression hypothesis for fig. 2.50	126
2.58	Objects with two attributes each at two levels, satisfying $H_0$ of independence; no two-factor interaction in which the two levels of location (attribute $A$ ) are equally probable given the level of response (attribute $B$ )	129
2.59	Objects with two attributes each at two levels, satisfying $H_0$ of independence; a uniform distribution in which the joint variable $AB$ , defined as in fig. 2.58, is equally probable	129
2.60	Objects with two attributes each at two levels, satisfying $H_0$ of independence; no two-factor interaction in which the two levels of response ( $B$ ) are equally probable given the level of location ( $A$ )	129
2.61(a)	Contingency table corresponding to fig. 2.58	130
2.61(b)	Contingency table corresponding to fig. 2.59	130
2.61(c)	Contingency table corresponding to fig. 2.60	130
2.61(d)	General contingency table for such maps	130
2.62	Reappraisal of the data in figs. 2.18, 2.19	133
2.63	General notation for $r \times c$ contingency table analysis in terms of hypotheses and marginal totals implied (after Goodman)	133
2.64	General format for a $2 \times 2 \times 2$ contingency table	134
2.65	General notation for $r \times c \times l$ contingency table analysis in terms of hypothesis and marginal totals implied (after Goodman 1970)	135
2.66	Map showing objects in terms of three properties each at two levels corresponding to simplest $r \times c \times l$ contingency table	136
2.67	Contingency table corresponding to fig. 2.66	136
2.68(a)	General format for hypothesis description, fitted marginals and their expected values for three-variable map contingency table analysis	138

2.68(b)	Results of $H_0 : A \otimes B   C$ for fig. 2.67	138
2.68(c)	Results of $H_0 : B \otimes C   A$ for fig. 2.67	138
2.68(d)	Results of $H_0 : A \otimes C   B$ for fig. 2.67	138
2.68(e)	Results of $H_0 : BA \otimes C$ for fig. 2.67	139
2.68(f)	Results of $H_0 : C = \Phi   AB$ for fig. 2.67	139
2.68(g)	Results of $H_0 : A \otimes B \otimes C$ for fig. 2.67	139
2.68(h)	Results of $H_0 : B \otimes C \cap A = \Phi   BC$ for fig. 2.67	139
2.68(i)	Results of $H_0 : CA = \Phi   B$ for fig. 2.67	140
2.68(j)	Results of $H_0 : ABC = \Phi$ for fig. 2.67	140
2.69	Modification of fig. 2.66 to satisfy $H_0 : BA \otimes C$ exactly	140
2.70	Modification of fig. 2.66 to satisfy $H_0 : A \otimes B \otimes C$ exactly	141
2.71	Modification of fig. 2.66 to satisfy $H_0 : CA = \Phi   B$ exactly, but marginal totals are no longer all the same as in fig. 2.66	141
2.72	Contingency tables (three-way and implied two-way) summarizing a map of land-use, age, decay of buildings in East London, 1973	143
2.73	Acceptable and unacceptable hypotheses from fig. 2.72	144
2.74	Expected values under $H_0 : B \otimes C   A$ for fig. 2.72	146
2.75	Increment in farm income	149
2.76	Wilcoxon rank-sum calculations for fig. 2.75	150
2.77	Wilcoxon rank-sum calculations for a subset of fig. 2.75	150
2.78	Probability distribution of Wilcoxon rank-sum for $n=3$ , $m=6$	151
2.79	<i>cdf</i> for Wilcoxon rank-sum for $n=3$ , $m=6$	152
2.80	Number-line format for Mann-Whitney $U$ statistic applied to the subset of fig. 2.77	153
2.81	Organic content of soil at 6" under three woodland types in format for the Kruskal-Wallis test statistic	155
2.82	General format for Friedman's related samples test statistic	156
2.83	Divided symbol representation of paper-types made in a sample of mills in Kent, 1860-5	158
2.84	Details of fig. 2.83 in tabular form	159
3.1	Arrangement of objects of two types to give the maximum number of runs	178
3.2	Arrangement of objects of two types to give the minimum number of runs	178
3.3	Counting rule 4	179
3.4	Probability of the number of runs in fig. 3.1, showing <i>cdf</i> and normal approximation	181
3.5	Arrangements of runs of $n_1 = 8$ , $n_2 = 7$ objects to compare with the frequencies in fig. 3.4	182
3.6	Probability distribution of $r$ in fig. 3.4 to illustrate asymmetry	183
3.7	Probability distribution for $r$ treated as a hypergeometric random variable	185
3.8	Occurrence of shop type in Market Street, Manchester, 1967	185
3.9	Data in fig. 3.8 reduced to two types, $R$ = retail, $S$ = not retail	186

3.10	Probability distribution of incidence of paired types in fig. 3.8	187
3.11	Effect of a new measurement on three existing measurements	190
3.12	Mean pollution-level measurements at sample stations along the River Trent in 1959–61 and 1965–7	191
3.13	A standard line-symbol representation of the data in fig. 3.12	192
3.14	Exact probability distribution of a binomial random variable with $p=\frac{1}{3}$ for $r \leq 7$	194
3.15	A graphical interpretation of Cox and Stuart's test	196
3.16	Cox and Stuart's test results for the data in fig. 3.12, case 1	197
3.17	Cox and Stuart's test results for the data in fig. 3.12, case 2	197
3.18	Calculations for Spearman's coefficient of rank correlation, $\rho$	200
3.19	Calculations for Kendall's coefficient of rank correlation, $\tau$ , illustrating the use of concordant and discordant values	202
3.20	Westenberg-Mood median test table for data in fig. 3.12	206
3.21	Probability calculations for the events implied by the Wilcoxon signed-rank test for $n=4$ , full, $n=10$ , partial	208
3.22	Calculating the Wilcoxon signed-rank test statistic for the data of fig. 3.12	210
3.23	Confidence-interval for the median difference in the Wilcoxon signed-rank test	211
3.24	Possible paths in the Smirnov two-sample empirical distribution function statistic for $m=4$ , $n=3$	213
3.25	Forming the 35 ( $X$ , $Y$ ) arrangements for $m=4$ , $n=3$ in fig. 3.24	214
3.26	The 35 arrangements of fig. 3.24 showing the stepwise deviations	214
3.27(a)	Fixing a rejection region in Smirnov's test	215
3.27(b)	Rejection region in relation to step-wise deviations in Smirnov's test	216
3.27(c)	Possible arrangements and their paths within a given rejection region in Smirnov's test	216
3.28	Smirnov's test applied to the data of fig. 2.75	217
3.29	Total possible arrangements for the data in fig. 2.75	217
3.30	Total possible arrangements for a given rejection region for the data in fig. 2.75	218
3.31	Standard tabulation of values for calculating Smirnov's test statistic	219
3.32	Location of homes of 16 visitors to a National Park	221
3.33(a)	Graphical illustration of the Kolmogorov one-sample test applied to data of fig. 3.32	221
3.33(b)	Arrangement of symbols in fig. 3.32 along the unit radius	222
3.34	Standard format for calculating the value of the one-sample Kolmogorov test applied to the data of fig. 3.33	222
3.35	Random samples, taken sequentially, of British Parliamentary constituency results for the Labour Party, 1966, at sample sizes $n=5$ , $n=10$ , $n=15$	223

3.36	Random samples, taken sequentially, of British Parliamentary constituency results for the Conservative party, 1966, at sample sizes $n=5$ , $n=10$	225
3.37	Chi-square as a goodness-of-fit test statistic under $H_0$ of a normal population of measurements for grouped data and population parameters	227
3.38	Calculating Kuiper's test of goodness-of-fit for angular measurements using data from fig. 3.33	229
3.39	Graphical illustration of Kuiper's test as a Kolmogorov analogue	231
3.40	Angular measurements of the macrofabric of a till	232
3.41	The unit circle and the resultant	233
3.42	Calculating the resultant for the uniform scores test for angular measurements	234



# **Section 1/ Maps, measurement and probability**