World Survey of Climatology Volume 1B

General Climatology, 1B

Elements of Statistical Analysis

by O. M. ESSENWANGER

610 Mountain Gap Drive,
Huntsville, AL 35803 (U.S.A.)

ELSEVIER Amsterdam-London-New York-Tokyo 1986

Contents

FOREWORD VII

ACKNOWLEDGEMENTS IX

Chapter 1. INTRODUCTION 1

Probability and expectancy 1
Probability density, cumulative distribution, and classes. 3
Contingency tables 4
Primitive and elementary characteristics 7
Mathematical characteristics, moments 9
Moment generating function and cumulants 12
Homogeneity 14
Persistence 16
Simple tools 16
Autocorrelation and repetitions 18
Areal persistence 19
Significance and confidence 20
Estimation 21
Definitions 21
Maximum-likelihood estimation 23
Moments estimation 25
Least square estimation. 25
Bayesian estimation 25
Other estimations. 26
Order statistics 26
Mathematical and empirical symbols 27

Chapter 2. FREQUENCY DISTRIBUTIONS. 29

The binomial distribution 29
The Poisson distribution 32
The negative binomial distribution 33
The Gaussian distribution. 36
The incomplete gamma function 42

Maximum-likelihood estimators . . . 45
Straight-line fitting . . . 46
The hypergamma function . . . 48
The Weibull distribution . . . 50
Moments estimators . . . 51
Maximum-likelihood estimation . . . 52
Estimation by utilization of Gumbel's distribution . . . 53
Straight-line fit estimators . . . 53
Estimation by order statistics . . . 55
Miscellaneous estimators . . . 55
The kappa distribution . . . 57
The logarithmic series distribution . . . 60
The logarithmic negative mixture distribution . . . 61
Moments estimators . . . 62
The circular normal distribution . . . 69
Gumbel's model . . . 69
The generalized circular distribution . . . 70
Estimation by the method of steepest descent . . . 72
Extreme value distributions . . . 75
The Fisher-Tippett Type I or Gumbel's law . . . 76
The return period . . . 79
Risk of occurrence, 80 — Return period from time series, 80
Fisher-Tippett Type II . . . 83
Distribution, moments and estimators, 83 — Test for suitability of Type I or Type II model, 87
The *m*-th value distribution . . . 89
Gumbel's *m*-th distribution . . . 90
Cramer's *m*-th distribution . . . 92
Other extreme value distributions . . . 94
The time interval distribution . . . 95
The generalized extreme value distribution . . . 96
Jenkinson's model . . . 97
Maximum-likelihood estimates . . . 98
Initial estimators . . . 99
The Pareto distribution . . . 100

Chapter 3. CORRELATIONS AND LINEAR REGRESSION . . . 103

General regression model . . . 103
Residual variance and percentage reduction . . . 107
Correlations . . . 108
Linear correlation . . . 108
Autocorrelation . . . 111
Spurious correlation . . . 112
Rank correlation, correlation of attributes . . . 113

The correlation coefficient as a measure of collinearity . . . 113
Linear relationship between climatological stations . . . 114
Multiple correlation . . . 116
Three variables . . . 117
Four variables . . . 118
Conditional correlations and partial correlations . . . 120
The canonical correlation . . . 120
Definition . . . 120
Canonical correlation of four variates . . . 121
Vector correlation . . . 123
Correlations between two vectors . . . 123
Measure of dependence . . . 124
Correlation ratio . . . 128
Polynomial analysis and curve fitting . . . 132
Polynomial series . . . 132
Orthogonal polynomials . . . 134
Residual variance and orthogonal polynomials . . . 135
Correlation and orthogonal polynomials . . . 136
Empirical polynomials and factor analysis . . . 137
Areal orthogonal polynomials . . . 141
Special functions . . . 141

Chapter 4. FOURIER AND SPECTRUM ANALYSIS . . . 143

Harmonic or Fourier analysis . . . 143
Fast Fourier transform (FFT) . . . 147
Periodogram analysis . . . 151
The spectrum . . . 151
Periodogram analysis . . . 151
Significance test by analysis of variance . . . 154
Power spectrum analysis . . . 156
Power spectrum and autocorrelation . . . 156
The cross-spectrum . . . 158
Examples and comparison with Fourier analysis . . . 160
Power spectrum and turbulence spectrum . . . 167
Testing of spectra, red noise . . . 170
Red noise . . . 170
Testing of spectra . . . 175
Miscellaneous Fourier and power spectrum analyses . . . 176
Modulation of amplitude . . . 176
Separation of waves and aliasing effect . . . 176

Chapter 5. SMOOTHING, FILTERING, INTER- AND EXTRAPOLATION . . . 179

General concepts, filtering . . . 179
Smoothing . . . 180

Moving averages, inter- and extrapolation, weights 180
Exponential smoothing 183
Smoothing by differences 184
Smoothing by functions 188
Smoothing by overlapping polynomials (filtering) 190
Filters 194
General filters 194
Low, high, and band pass filters 196
Transverse filters, 196 — Low pass filters, 196 — High pass filters, 198 — Band pass filters, 198
Special filters 200
Martin-Graham filter, 201 — Kertz and Olberg system, 203 — A symmetric linear filter, 212
Filtering and persistence 213
Kalman filtering 214
One-dimensional model 214
Multidimensional model 215
Linear dynamical systems 216
Kalman filters 217
Special applications 220
Spline functions 221
General interpolation rules 221
Spline fitting 222
Two-dimensional spline fitting 226
Smoothing and graduation in two dimensions 229
Minimum of E_1 230
Minimum of E_2 231
Second condition for E_2 231
Third condition for E_2 232

Chapter 6. TESTS 233

Statistical inference 233
General concepts 233
Hypothesis testing 234
Data characteristics 236
Orthogonality and independence 236
Error sources 237
Sampling effect, 237 — Persistence, 238 — Spurious correlation, 238 — Selection effect, 239 — Trend, 240
Significance and power of the test 240
Significance and Type I error, 240 — Type II error, 242
Confidence limits 243
Degrees of freedom 247
Standard error 247

Standard error of moments . . . 247
Statistical test for a Gaussian distribution . . . 249
Standard error of percentiles . . . 250
Standard error of miscellaneous statistical parameters . . . 251
Persistence and the standard error . . . 254
Non-parameter tests and homogeneity . . . 254
General concept . . . 254
Sign tests (differences) . . . 255
Variance test of differences . . . 258
Wilcoxon rank test of differences . . . 258
Wilcoxon's two-sample rank test . . . 260
Wilcoxon's stratified rank test . . . 262
Mann-Whitney test . . . 262
Kruskal-Wallis test . . . 264
Run tests, median . . . 265
Other non-parametric tests . . . 268
Student's *t*-test . . . 269
Test of sample means . . . 270
Test of the correlation coefficient . . . 273
Test for regression line coefficients . . . 274
The non-central *t*-distribution . . . 275
Fisher's *F*-test . . . 276
Relationship to χ^2 and *t* . . . 277
Multiple variances . . . 278
Analysis of variance . . . 279
Fisher's *z*-transformation . . . 279
Testing of the linear correlation coefficient . . . 279
Chi-square distribution and test . . . 281
General . . . 281
Tables and approximations . . . 283
Expansion of the chi-square application . . . 284
Modified chi-square test with equal and optimum classes . . . 285
Other modifications of the chi-square test . . . 287
Non-central chi-square, non-central *F* . . . 287
Approximation by chi-square, 290 — Gaussian approximation, 290 — Moments fit, 290
Likelihood ratio tests, *G*-test . . . 292
Likelihood ratio . . . 292
Log-likelihood ratio . . . 293
Power and sample size for approximate chi-square tests . . . 294
Test of independence . . . 295
Test of homogeneity . . . 296
Contingency tables . . . 296
The 2 × 2 contingency table . . . 297
Significance check for Poisson sampling in contingency tables . . . 301
Goodness-of-fit tests . . . 302

General 302
Kolmogorov-Smirnov test 303
Two-sided test, 304 — One-sided test, 306
Comparison of two empirical distributions 307
Other goodness-of-fit tests based on the cumulative distribution 307
Cramer-von Mises, 307 — Anderson-Darling, 308 — Johnson's distribution, 309 — Watson's statistic, 310 — Kuiper's test statistic, 311 — Interrelationships, 311
Order statistics 312
Uniform distribution, 312 — Test criteria for Gaussian distribution, 313
Quality control and assurance of observations 314
Individual data checking procedures 315
Trivial error checks, 315 — Error checking by adjacent data, 315
Frequency distribution checks 316
Gross errors and outliers 318
General quality control 320
Neyman's $C(\alpha)$ test 320
Concluding remarks 321

Chapter 7. CHARACTERISTICS OF METEOROLOGICAL ELEMENTS 323

The wind vector 323
The wind speed 325
Duration of runs 327
The wind direction 327
Statistical background for the definition of a mean wind direction 328
Calculation of the mean wind direction by desk calculator 329
Mean wind direction calculations by runs 331
The cumulative distribution 335
The wind shear 337
The distribution of cloud cover 337
Statistical models of precipitation frequencies 342
The negative binomial and Poisson distribution for precipitation frequencies . . 343
Logarithmic scale and other transformations 345
The incomplete gamma function 347
Beta distribution for precipitation data 349
Markov chain models 349
Dry–wet runs 352
Green's model, 1964, 352 — Green's model, 1970, 358 — Other models for wet and dry runs, 359 — Conditional and unconditional probabilities, 362
Miscellaneous models 363
Duration of rainstorms, 363 — A model of n-day precipitation, 364 — Dropsize distributions, spectra and total rainfall, 367 — Areal precipitation and patterns, 370 — Average annual precipitation and daily amounts, 372 — Total rainfall and number of days, 373 — Rainfall intensity and rainfall rates, 374 — Precipitable water, 375

Weather modification 375
Concluding remarks. 377

Chapter 8. REFERENCES 379

REFERENCE INDEX 405

SUBJECT INDEX 413

LIST OF SYMBOLS 423