Nonparametric Location Tests Against Restricted Alternatives

Wolfgang Kössler

Berichte aus der Mathematik

Wolfgang Kössler

Nonparametric Location Tests Against Restricted Alternatives

Shaker Verlag
Aachen 2006

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the internet at http://dnb.ddb.de.

Zugl.: Berlin, Humboldt-Univ., Habil.-Schr., 2003

Printed in Germany.

ISBN-10: 3-8322-5004-2
ISBN-13: 978-3-8322-5004-1
ISSN 0945-0882

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen
Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9
Internet: www.shaker.de • eMail: info@shaker.de

To My Parents

Preface

The problem of testing the equality of several location parameters on the basis of independent random samples appears very often, especially in economical, social, medical and biological applications. In contrast to most of the (text)books we assume that we have some information about the alternative. Perhaps it is already known that the alternative is ordered or that it has the form of an umbrella. This information is efficiently used to construct tests with good power properties. Moreover, we do not rely on the assumption that the underlying populations are normally distributed. Arbitrary continuous distributions are allowed.

This book contains a review of rank-based procedures and related test procedures for restrictive alternatives (ordered alternatives, umbrella alternatives with known peak and with unknown peak) together with their properties and comparisons.

For understanding the theory some background in the theory of linear rank tests may be useful. But also the pure applicant may profit from this book since many suggestions for the application of a suitable procedure are made.

This work is based on my habilitation thesis which I prepared at the Humboldt-University of Berlin. The material is slightly modified and it is translated from German. The own results are incorporated in already existing ones, asymptotic results are accompanied by illustrative examples and quite extensive simulation studies.

I am especially grateful to Professor Egmar Rödel at Humboldt-University for giving me the opportunity to work on this subject. Special thanks go to Professor Büning from the Free University of Berlin with whom I wrote many papers which became the basis for this book. Many important hints I got from Professor Marie Hušková from the Prague University and from Professor Olaf Bunke from the Humboldt-University. I also thank Petra Kämpfer for supporting me in solving computer problems.

Berlin, March 14, 2006 Wolfgang Kössler

Contents

List of the most important symbols

c	Number of treatments or populations
l	Umbrella peak
Hypotheses	
H_0	Null hypothesis
H_A	Two-sided alternative hypothesis
H_1, H_{1A}, H_{1B}	Ordered alternative hypotheses
H_2, H_{2A}, H_{2B}	Umbrella alternative hypotheses (known peak)
H_3	Umbrella alternative hypothesis (unknown peak, $l \geq 2$)
H_4	Umbrella alternative hypothesis (unknown peak)
Sample sizes and alternatives	
n_i	Sample size of the ith treatment
N_i	Sample size of the first ith treatments
N	Total sample size
L	Sample size (general)
$\lambda_i = n_i/N$	Sample size fraction of the ith treatment
$\Lambda_i = N_i/N$	Sample size fraction of the first ith treatments
$\mathbf{n} = (n_1, \ldots, n_c)$	Sample size configuration
$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_c) = \boldsymbol{\lambda}_N$	Sample size configuration
$\vartheta_i, \vartheta_{i,N}$, $\theta_i = \sqrt{N}\vartheta_{i,N}/\Delta$	Location parameter of the ith treatment
$k_{iN} = \vartheta_{i,N}/\sigma_F$	Used in the definition of the simulated alternatives
σ_F^2	Variance of the ‘standard’ distribution of F
Δ, b', b''	Arguments of the power function
δ_i, δ	Spacings of the alternative
$\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_c)$	Alternative configuration
$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_c)$	Alternative configuration
$\mathbf{k}_N = (k_{1,N}, \ldots, k_{c,N})$	Alternative configuration (for simulation)

Test statistics for the two-sided alternative

F	F-test
KW	Kruskal-Wallis test
KWT	Kruskal-Wallis type test

Test statistics for ordered alternatives

J	Jonckheere-Terpstra test
JT	Jonckheere-type test (JT-test)
PT	Puri-type test (PT-test)
TT	Tryon-Hettmansperger type test (TT-test)
HT	Hettmansperger-Norton type test (HT-test)
RM	RM-test of Fairley and Fligner
AM, J^*	Statistics of Fairley and Fligner
$CS1$, $CS2$	Statistics of Chakraborti and Schaafsma
$S(O)$, $S'(O)$	Statistics of Shirahata
GJT	Generalized Jonckheere-type test (GJT-test)
GTT	Generalized Tryon-Hettmansperger type test (GTT-test)
GHT	Generalized Hettmansperger-Norton type test (GHT-test)
$\tilde{J}T$	JT-test based on U-statistics
$\tilde{P}T$	PT-test based on U-statistics
$\tilde{T}T$	TT-test based on U-statistics
$\tilde{G}\tilde{J}T$	GJT-test based on U-statistics
$\tilde{G}\tilde{T}T$	GTTT-test based on U-statistics

Test statistics for umbrella alternatives with known peak

MW	Mack-Wolfe test
MWT	Mack-Wolfe type test (MWT-test)
PT	Puri-type test (PT-test)
THT	Tryon-Hettmansperger type test (THT-test)
HNT	Hettmansperger-Norton type test (HNT-test)
ST	Shi-type test (ST-test)
$GMWT$	Generalized Mack-Wolfe type test (GMWT-test)
$GTHT$	Generalized Tryon-Hettmansperger type test (GTHT-test)
$GHNT$	Generalized Hettmansperger-Norton type test (GHNT-test)

Test statistics for umbrella alternatives with unknown peak

MW_{max}	Chen-Wolfe test
CWT	Chen-Wolfe type test (CWT-test)
THT_{max}	Tryon-Hettmansperger type test (THT_{max}-test)
HNT_{max}	Hettmansperger-Norton type test (HNT_{max}-test)
ST_{max}	Shi-type test (ST_{max}-test)
$SM(q), SM(\frac{1}{2})$	Simpson-Margolin procedure, SM(q)-test

Adaptive tests, various measures

$A(\hat{S})$, $B(\hat{S})$,$C(\hat{S})$	Adaptive tests A, B and C based on $\hat{S}$
$A(\hat{S}')$, $B(\hat{S}')$,$C(\hat{S}')$	Adaptive tests A, B and C based on $\hat{S}'$
$\hat{S} = (\hat{t}, \hat{s})$	Selector statistic based on the pooled sample
$\hat{S}' = (\hat{t}', \hat{s}')$	Selector statistic based on the single samples
$t, t_{0.05,0.15}$	Measures for tailweight
$s, s_{0.05}$	Measures for skewness
$\hat{t}, \hat{t}'$	Estimates for tailweight
$\hat{s}, \hat{s}'$	Estimates for skewness
D_i, $\widetilde{D}_i$, $D_i, D_{i,A}, D_{i,B}$	Subsets of the range of the selector statistics

Special tests based on U-statistics

KGM	Kumar-Gill-Mehta test
SGB	Shetty-Govindarajulu-Bhat test
AK	Amita-Kochar test
BK	Bhapkar-test
DP	Deshpande-test
$\widetilde{PT}^C$	Compagnone-Denker-Puri test

Parametric tests

BB	Barlow-Bartholomew-Bremner-Brunk test
HT_{par}	Modification of BB
AT	Abelson-Tukey test (AT-test)
S_{01}	Modification of the LQ-test
$\widetilde{S}_{01}$	Pincus-test

Substatistics

$T_{i,N}$	Substatistic used for the definition of HT and HNT
$U_{i,j}$	Two-sample statistic based on the ith and jth sample
$S_{(1\ldots i-1)i}$	Two-sample statistic based on the first $i-1$ and the ith sample
$S_{(c\ldots i+1)i}$	Two-sample statistic based on the last $c-i$ and the ith sample
$\widetilde{U}_{i,j}$	U-statistics corresponding to $U_{i,j}$
$\widetilde{S}_{(1\ldots i-1)i}$	U-statistics corresponding to $S_{(1\ldots i-1)i}$
$\widetilde{S}_{(c\ldots i+1)i}$	U-statistics corresponding to $S_{(c\ldots i+1)i}$
$MWT_l, THT_l,$ HNT_l, ST_l	Umbrella statistics MWT, THT, HNT, $ST$with peak l
$\mathbf{MWT}^*, \mathbf{THT}^*,$ $\mathbf{HNT}^*, \mathbf{ST}^*$	Vectors of the standardized umbrella statistics
JT_i	Jonckheere-type statistics for the first i treatments

Ranks

R_{ij}	Ranks over all c samples
R_{ij}^k	Ranks over the ith and jth samples only
$R_{1\dots i}^k$	Ranks over the first i samples
$R_{c\dots i}^k$	Ranks over the last $c-i+1$ samples

Scores

$a_L(i)$	Scores, general
GA	Gastwirth scores
WI	Wilcoxon scores (the original ranks)
LT	Long tail scores
HFR	Hogg-Fisher-Randles scores
VW	van der Waerden scores
SA	Savage scores
-HFR	HFR-scores applied on the reflected sample ($X_i \to -X_i$)
-SA	Savage-scores applied on the reflected sample ($X_i \to -X_i$)

Functions and related measures

f, g	Density functions
F	Cumulative distribution function (cdf)
F^{-1}	Quantile function
Φ	Cdf of the standard normal distribution
$z_{1-\alpha}$	(1-α)-quantile of the standard normal distribution
$\phi, \phi(u), \phi(u,f), \phi(u,g)$	Score functions
$I(f), I(g)$	Fisher informations
$d(f,g)$, $J(f)$	Auxilary measures
$C(f,g)$	Cofactor in the formulas for the asymptotic efficacies
$C_U(f)$	Cofactor in the formulas for the asymptotic efficacies (U-statistics)
ϕ, ϕ^A, ϕ^C, $\phi^B_{(i)}, \phi^D_{(i)}$	Kernel functions

Special densities

U-L	Uniform-logistic
Lo	Logistic
L-D	Logistic-Doubleexponential
L-E	Logistic-Exponential
N	Normal
nGu	“negative” Gumbel ($f_{nGu}(x) = f_{Gu}(-x)$)
Gu	Gumbel
DE	Doubleexponential
Cau	Cauchy
CN1	Contaminated normal (variant 1)
CN2	Contaminated normal (variant 2)
Uni	Uniform
Ex	Exponential

Expectations and variances

$\mathbf{E}$, μ, μ_N	Expectation (general)
η_i	Expectation of substatistics (general)
$\boldsymbol{\eta}$	Vector of expectations (general)
σ^2, σ_N^2	Variance (general)
σ_U^2	Variance of U-statistics (general)
μ_i, $\mu_{(1\ldots i-1)i}$, $\mu_{(c\ldots i+1)i}$	Expectations of substatistics
σ_i^2, $\sigma^2_{(1\ldots i-1)i}$, $\sigma^2_{(c\ldots i+1)i}$	Variances of substatistics
$\mu_{JT}, \mu_{PT}, \mu_{TT}, \mu_{HT}$	Expectations of JT, PT, TT, HT
$\sigma_{JT}, \sigma_{PT}, \sigma_{TT}, \sigma_{HT}$	Standard deviations of JT, PT, TT, HT
$\mu_{\widetilde{JT}}$, $\mu_{\widetilde{PT}}$, $\mu_{\widetilde{TT}}$	Expectations of $\widetilde{JT}, \widetilde{PT}, \widetilde{TT}$
$\sigma_{\widetilde{JT}}$, $\sigma_{\widetilde{PT}}$, $\sigma_{\widetilde{TT}}$	Variances of $\widetilde{JT}, \widetilde{PT}, \widetilde{TT}$
$\mu_M, \mu_P, \mu_T, \mu_H, \mu_S$	Expectations of MWT, PT, THT, HNT, ST
$\sigma_M, \sigma_P, \sigma_T, \sigma_H, \sigma_S$	Standard deviations of MWT, PT, THT, HNT, ST
$\boldsymbol{\mu}_M, \boldsymbol{\mu}_T, \boldsymbol{\mu}_H, \boldsymbol{\mu}_S$	Expectation vectors of $\mathbf{MWT}^*, \mathbf{THT}^*, \mathbf{HNT}^*, \mathbf{ST}^*$

Covariances

$\boldsymbol{\Sigma}$	Covariance matrix (general)
$\boldsymbol{\Sigma}_M, \boldsymbol{\Sigma}_T, \boldsymbol{\Sigma}_H, \boldsymbol{\Sigma}_S$	Covariance matrix of $\mathbf{MWT}, \mathbf{THT}, \mathbf{HNT}, \mathbf{ST}$
$\boldsymbol{\Sigma}_M^*, \boldsymbol{\Sigma}_T^*, \boldsymbol{\Sigma}_H^*, \boldsymbol{\Sigma}_S^*$	Covariance matrix of $\mathbf{MWT}^*, \mathbf{THT}^*, \mathbf{HNT}^*, \mathbf{ST}^*$
$\rho_M, \rho_T, \rho_H, \rho_S$	Covariance between the statistics MWT_l^*, THT_l^*, HNT_l^* and ST_l^*, respectively (cf. Chapter 4)
ρ_l	Covariance between $S_{(1\ldots l-1)l}$ and $S_{(c\ldots l+1)l}$ (Chapter 3)

Efficacies and Efficiencies

$K_{\mathbf{n}}$	Efficacy
K, $K(\boldsymbol{\theta})$	Asymptotic efficacies (general)
$K_{JT}, K_{PT}, K_{TT}, K_{HT}, K_{RM}$	Asymptotic efficacies of JT, PT, TT, HT, RM
$K_{\widetilde{JT}}, K_{\widetilde{PT}}, K_{\widetilde{TT}}$	Asymptotic efficacies of $\widetilde{JT}, \widetilde{PT}, \widetilde{TT}, \widetilde{HT}$
K_M, K_P, K_T, K_H	Asymptotic efficacies of MWT, PT, THT, HNT
ARE	Asymptotic relative efficiency
$A^2(\boldsymbol{\lambda}, \boldsymbol{\theta})$, B, A_t^*	Factors in the formulas for the asymptotic efficacies
$A^2_{JT}, A^2_{TT}, A^2_{HT}$	Factors for JT (and PT), TT, HT
$A^2_M, A^2_T, A^2_H, A^2_S$	Factors for MWT (and PT), THT, HNT, ST
$\beta_{\boldsymbol{\lambda},\boldsymbol{\theta}}(\Delta)$	Asymptotic power function

Weights

$\boldsymbol{\omega}$	Vector of weights (general)
$\boldsymbol{\omega}_J, \boldsymbol{\omega}_T$	Vector of weights for GJT, GTT
$\boldsymbol{\omega}_M, \boldsymbol{\omega}_T$	Vector of weights for $GMWT$, $GTHT$
$\boldsymbol{\omega}_{J,i}, \boldsymbol{\omega}_{T,i}$ and $\boldsymbol{\omega}_{M,i}, \boldsymbol{\omega}_{T,i}$	Components of these vectors
w_i, v_i	Weights for the substatistics of HT and HNT
$\omega_{H,i}$	Weights for the substatistics of GHT
$\omega_{\widetilde{J}T,i}$, $\omega_{\widetilde{T}T,i}$	Weights for the substatistics of $G\widetilde{\widetilde{J}}T$ and $G\widetilde{\widetilde{T}}T$
ω_i^*	Weights used for the definition of RM and ST

Abbreviations

ARE	Asymptotic relative efficiency
cdf	Cummulative distribution function