A Nonparametric Test of Interaction in the General Two-Way Layout

Genel İki-Yönlü Düzende Etkileşimin Nonparametrik Testi

John ÖHRVIK[a]

[a]Department of Medicine,
Karolinska Institutet, Sweden

Geliş Tarihi/*Received:* 24.06.2009
Kabul Tarihi/*Accepted:* 10.09.2009

Yazışma Adresi/*Correspondence:*
John ÖHRVIK
Karolinska Institutet,
Department of Medicine,
Cardiology Research Unit N5:00,
SE-17176 Stockholm, Sweden
john.ohrvik@ki.se

ABSTRACT The two-way layout is frequently occurring, e.g. blocking is used to reduce the between subject variability when comparing treatments or many medical centers are included in a clinical trial to recruit a sufficient number of patients. In epidemiological studies, it is common to study the interactions between genetic and environmental factors. This paper is concerned with the statistical analysis of data arising in these situations when assumptions like normality do not necessarily apply. The main objective of this paper is to propose a test for interactions for continuous data based on joint ranking of all observations after iteratively eliminating the two main effects. The validity of the significance levels of the test when using a finite sample version of the asymptotic distribution of the test statistic is manifested and the power against different alternatives illustrated by extensive simulation experiments. The proposed test is compared with competitors on published data sets. Data from the Survey of Adolescent Life in Vestmanland (SALVe) project are analyzed both with our test and Brunner and Puri's proposal. The test shows good asymptotic and small sample properties. It can be used in the general two-way layout and multiple comparisons can be performed in a straightforward way. The analyses of the SALVe project show that our proposal can be useful in such studies. It has the potential to pick-up interactions hidden in the noise of gross errors when using standard ANOVA. The test may become a valuable tool and alternative to other proposals in exploring interactions.

Key Words: Epistasis, genetic; statistics, nonparametric; computer simulation; research design

ÖZET Tedavileri karşılaştırırken ya da yeterli hasta sayısına ulaşmak için bir klinik denemeye pek çok medikal merkez dahil dildiğinde, denekler arasındaki değişkenliği azaltmak için genellikle iki-yönlü düzen (two-way layout), örneğin bloklama, kullanılır. Epidemiyolojik çalışmalarda, genetik ve çevresel faktörler arasındaki etkileşim üzerinde çalışmak yaygındır. Bu çalışmada, normallik gibi varsayımların mutlaka sağlanmış olmadığı, bu tip durumlarda elde edilen veri setinin istatistiksel analizi üzerinde durulmuştur. Bu çalışmanın ana amacı, 2 ana etki iteratif olarak elimine edildikten sonra, bütün gözlem değerlerinin bileşik sıralamasına dayanan sürekli verinin etkileşimleri için bir test önermektir. Test istatistiğinin asimptotik dağılımının sonlu örneklem versiyonunu kullanırken, testin anlamlılık düzeylerinin geçerliliği ortaya konulmuş ve kapsamlı simülasyon deneyleriyle farklı alternatiflere karşı gücü gösterilmiştir. Önerilen test, yayınlanmış veri setlerindeki alternatiflerle karşılaştırılmıştır. Vestmanland'da Adölesan Yaşamı Anketi (Survey of Adolescent Life in Vestmanland - SALVe) projesindeki veriler, hem bizim testimizle hem de Brunner ve Puri'nin testiyle analiz edilmiştir. Test, iyi asimptotik ve küçük örneklem özellikleri göstermektedir. Genel iki-yönlü düzende kullanılabilir ve çoklu karşılaştırmalar kolay bir şekilde yapılabilir. SALVe projesinin analizleri bizim önerimizin bu tip çalışmalarda kullanışlı olabileceğini göstermektedir. Standart Varyans analizi kullanılırken büyük hatalarda saklanan etkileşimleri yakalayabilme özelliğine sahiptir. Test, etkileşimlerin araştırılmasında değerli bir araç ve diğer önerilere bir alternatif olabilir.

Anahtar Kelimeler: Epistasi, genetik; istatistik, non-parametrik, bilgisayar simülasyonu, araştırma tasarımı

Turkiye Klinikleri J Biostat 2010;2(1):27-41

In practice the shapes of the underlying distributions are very seldom known or there may be distinct indications that the underlying distributions are non-normal. In such cases the use of the standard parametric methods assuming normality can be criticized regarding validity and optimality.

Nonparametric methods are valid for a broad family of underlying distributions and are yet reasonably efficient relative the best parametric

method under normality. For example, under the normal shift model the asymptotic relative efficiency (ARE) of the Wilcoxon rank sum test relative the *t*-test is 95.5%.

In medical data sets based on patient data the occurrence of gross errors, i.e. longtailedness, can reach 8-12%.

It has long been argued that with the exception of very simple designs there exists no nonparametric alternatives or they have low power due to separate ranking within each factor, e.g. the Friedman test.[1] A substantial improvement can be achieved if the observations are aligned with respect to (w.r.t.) one factor when studying the other factor. Since the aligned observations from different levels of an aligning factor are comparable when studying the other factor they can be ranked jointly. In general the asymptotic efficiency under the normal shift model of a nonparametric test using aligned ranks relative to the optimal parametric test is 95.5%.[2]

The use of ranks in the one-way layout, the so-called *H*-test, was first proposed by Kruskal and Wallis.[3] The limiting distribution of the *H*-statistic under the null hypothesis of equal distributions was derived by Kruskal[4] and later Andrews[5] derived the limiting distributions under the alternative hypothesis. Hodges and Lehmann[6] proposed the use of aligned ranks in a randomized block design to increase the efficiency of the rank based test of treatment effects, see also.[7]

Mehra and Sarangi[8] proposed a test of treatment effects based on aligned ranks in a randomized block design. The only limitation was that the number of subjects on each treatment did not vary between blocks. They derived the limiting distributions under the null and alternative hypotheses. Independently Sen[9] proposed the same test. This test is not only asymptotically distribution-free but also conditionally distribution-free. Later Mehra and Sen[10] proposed a test for interaction effects in the two-way layout using the same ideas. The test statistic was, however, difficult to compute in applied situations. Öhrvik[11] proposed a test for interactions based on aligned ranks for the randomized block design assuming symmetric distributions.

Multi-center trials with two treatments have been studied by Boos and Brownie[12] and Brunner, Puri and Sun[13] who have derived tests for interaction and treatment effects, the latter for both fixed and random center effects. The special case with only two centers was studied in.[14]

Sawilowsky[15] provided a very good review of nonparametric approaches to testing for interaction effects, including also the method called 'rank transform' (RT) – first suggested by Conover and Iman.[16,17] This for several years popular but controversial method had been recommended by among others SAS Institute.[18] However, Akritas[19] showed that for the two-way layout the RT method was not valid to test for main effects in the presence of interactions nor was it valid to test for interactions since it is a nonlinear transform of the data. He also showed that homoscedasticity of the error terms was not in general transformed to the ranks.

Hettmannsperger[2] and Hettmannsperger and McKean[20] discussed modified aligned rank tests including tests for interactions and main effects in the linear model. In recent years Akritas, Arnold and Brunner[21] and Brunner and Puri[22] presented a general framework for nonparametric analysis in factorial designs. These methods can be performed in SAS by applying Proc Mixed on the rank transformed data assuming a heterogeneous covariance structure.

Statistical interaction is usually defined as departure from additivity in a specific linear model describing the relationship between predictive factors. With this definition the choice of scale is important since factors, which are non-additive w.r.t. a response measured on one scale, may not manifest any interaction when a differently transformed scale is used. For example, a logarithmic transform of the response variable will often remove an interaction, which was present on the original scale.

A rapidly growing field of research is studies on gene environment interactions, especially in epidemiological surveys. Survey of Adolescent Life in Vestmanland (SALVe) is one of these studies with non-normally distributed outcomes. This un-

derlines the need for interaction tests which are robust w.r.t. deviations from the normal distribution e.g. longtailedness and/or skewness.

The main objective of this study is to propose a nonparametric test for interaction effects for continuous data based on joint ranking of all observations after iteratively eliminating the two main effects in the general two-way layout where factor β has b levels and factor τ has t levels with n_{ij} observations per cell. Pairwise comparisons of the cells will also be discussed.

MODEL

Let the factors β and τ be indexed by i=1,2,...,b and j=1,2,...,t respectively. Suppose there are n_{ij} indexed by k replications per cell (i,j). Let Y_{ijk} be the kth observation in the cell (i,j), and assume that these n_{ij} independent observations have a common continuous distribution function

$$F_{ij}(y) = F(y + \mu_{ij}), \quad (2.1)$$

where μ_{ij} is the mean level in cell (i,j). For models with interaction effects, μ_{ij} can be decomposed as

$$\mu_{ij} = \mu + \beta_i + \tau_j + (\beta\tau)_{ij}, \quad (2.2)$$

where μ is the overall mean level, β_i is the effect of the ith level of factor β, $\Sigma_i\beta_i$=0, τ_j is the effect of the jth level of factor τ, $\Sigma_j\tau_j$=0 and $(\beta\tau)_{ij}$ is the interaction effect between the ith level of β and the jth level of τ, $\Sigma_i(\beta\tau)_{ij}$=$\Sigma_j(\beta\tau)_{ij}$=0.

TEST STATISTICS

TEST FOR INTERACTION

The hypothesis of interest, namely, that there are no interaction effects can be expressed as

$$H_0 : (\beta\tau)_{11}=(\beta\tau)_{12}=...=(\beta\tau)_{bt}. \quad (3.1)$$

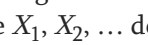
To isolate the parameter of interest, $(\beta\tau)_{ij}$, we suggest aligning by removing the factor effects. First subtract from the observations in level i of factor β an estimate of β_i, i=1,2,…,b. Then subtract from the resulting residual observations in level j of factor τ an estimate of τ_j, j=1,2,…,t. Unless the mean is used to estimate the factor effects this has to be done iteratively until the changes in the estimates from one iteration to the next are zero or negligible. In practice however, it is seldom necessary to use more than a few iterations, see.[23] These location estimates must be symmetric functions of the observations in each level of factor β and τ respectively, i.e. they must be invariant under permutations of the observations (this condition is satisfied by most location estimators) and for any constant c satisfy the conditions

$$\beta\,(Y_{1j1} + \alpha,\ldots,\ Y_{bjn(bj)} + c) = \beta\,(Y_{1j1},\ldots,\ Y_{bjn(bj)}) + c$$

and

$$\tau\,(Y_{i11} + \alpha,\ldots,\ Y_{itn(it)} + c) = \tau\,(Y_{i11},\ldots,\ Y_{itn(it)}) + c$$

where $\beta\,(.)$ and $\tau\,(.)$ denote the location estimates of factor β and τ.

We recommend estimating the factor effects by the Hodges-bservations, i.e. H/L =  , where X_1, X_2, … denote independent observations from a common distribution F. The iterative procedure for elimination of the main effects is given in the Appendix. Let Z_{ijk} denote the ijkth residuals after the final iteration and  its rank in the joint ranking of the entire $N = \Sigma_i\Sigma_j n_{ij}$ aligned observations. Mid-ranks are used to break ties.

The H/L estimator has very good robustness properties. Compared with the mean and the median it is neither sensitive to a few "wild" observations (gross error sensitivity) nor is it sensitive to gaps in the middle of the data set (local shift sensitivity). Properties like this can be studied using the influence function, $IC(x)$, which shows the asymptotic effect on an estimator of a point contamination at x.[24]

In Figure 1 the influence function of the H/L estimator together with the mean and the median are plotted for the normal distribution. As can be seen the H/L estimator has both bounded gross error sensitivity,

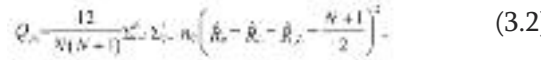
$$\sup_x |IC(x)|$$

and bounded local shift sensitivity,

$$\sup_{x\neq y} |IC(x) - IC(y)|/|x-y|$$

which neither the mean nor the median have.

To test the null hypothesis in expression (3.1) we propose the test statistic

 (3.2)

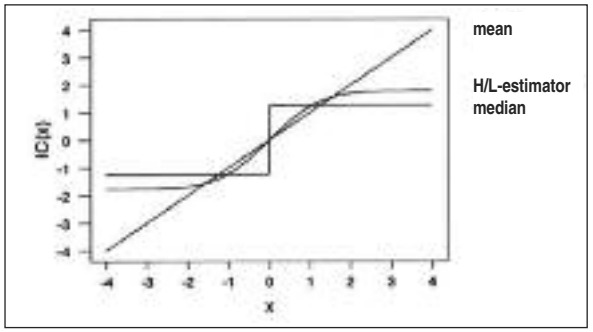


FIGURE 1: The influence function of some location estimators under the normal distribution.

where $\bar{R}_{i.}=\frac{1}{n_{i.}}\sum_j\sum_k R_{ijk}$, $\bar{R}_{.j}=\frac{1}{n_{.j}}\sum_i\sum_k R_{ijk}$, $\bar{R}_{ij}=\frac{1}{n_{ij}}\sum_k R_{ijk}$, $\frac{1}{N}\sum_i\sum_j\sum_k R_{ijk}=\frac{N+1}{2}$ $n_{i.}=\Sigma_j\, n_{ij}$ and $n_{.j}=\Sigma_i\, n_{ij}$. Under the null hypothesis the squared difference within the bracket is expected to be small. Thus the hypothesis of no interactions is rejected when $Q_{\beta\tau}$ is sufficiently large, say when $Q_{\beta\tau}\geq c$.

Remarks: (i) The test is conditional on the configuration i.e. the estimates of the factor effects.

(ii) Ties make the test conservative since the true variation is smaller than that given by expression (3.2) for $Q_{\beta\tau}$ which is based on continuous variables. To adjust for this

$$\sum_i\sum_j\sum_k\left(\bar{R}_{ij}-\bar{R}_{..}\right)^2/N \qquad (3.3)$$

can be used instead of $N(N+1)/12$ in expression (3.2). Straightforward calculations gives that in case of no ties expression (3.3) tends $N(N+1)/12$ if H_0 is true. Since there are usually fewer ties among the aligned observations compared to the original this adjustment is often unnecessary. Expression (3.2) can be used with reasonable confidence unless the number of ties is very large in the original dataset.

(iii) With many ties in the dataset, e.g. when the response variable takes only a small number of integer values it might be preferable to align using the mean instead of the H/L-estimator due to gaps in the distribution of the pairwise means.

PROPERTIES OF THE TEST STATISTIC

The following theorem justifies the weights in the test statistic

Theorem 3.1: When H_0 is true, $\lim E(Q_{\beta\tau})=(b-1)(t-1)$ if for each (i,j)

$\lim n_{ij}/N$ exists and is positive.

The proof is given in the Appendix.

Remark: If the distribution $F_{ij}(y)$ given in expression (2.1) is symmetric the 2nd and 3rd sum of squares on the 2nd line in expression (A.1) for $Q_{\beta\tau}$ in the Appendix i.e.

$\frac{12}{N(N+1)}\sum_i^{a-1}\sum_j^{b-1}n_{ij}\left(\bar{R}_{ij}-\frac{N-1}{2}\right)^2$ and $\frac{12}{N(N-1)}\sum_i^{a-1}\sum_j^{b-1}n_{ij}\left(\bar{R}_{ij}-\frac{N+1}{2}\right)^2$

will tend to zero as $N\to\infty$ and we get the following simpler form of $Q_{\beta\tau}$:

$$Q'_{\beta\tau}=\frac{12}{N(N+1)}\sum_i^{a-1}\sum_j^{b-1}n_{ij}\left(\bar{R}_{ij}-\frac{N+1}{2}\right)^2, \qquad (3.4)$$

which was studied in.[12]

The following theorems are stated without proofs.

Theorem 3.2: If H_0 is true, and for each (i,j) $\lim n_{ij}/N$ exists and is positive the statistic $Q_{\beta\tau}$, defined by (3.2) is asymptotically Chi-square distributed with $(b-1)(t-1)$ degrees of freedom (df).

For a proof see e.g. (3).

This implies that asymptotically the null hypothesis of no interaction effects is rejected at the significance level a if

$$Q_{\beta\tau}>\chi^2_a((b-1)(t-1)), \qquad (3.5)$$

where $\chi^2_a(\nu)$ denotes the upper a-quantile of a Chi-square distribution with ν df. For small n_{ij} a better approximation of the actual significance level is given by a F-distribution with $(b-1)(t-1)$ df in the numerator and N df in the denominator i.e. reject H_0 if

$$Q_{\beta\tau}/[(b-1)(t-1)]>F_\alpha[(b-1)(t-1),N]. \qquad (3.6)$$

Theorem 3.3: If for each (i,j) $\lim n_{ij}/N$ exists and is positive then under the shift model $G_{ij}(z)=G[y+(\beta\tau)_{ij}/\sqrt{N}]$ for all (i,j), where $G_{ij}(z)$ is the distribution of Z_{ijk}, the statistic $Q_{\beta\tau}$ defined by (3.2) is as-

ymptotically non-central Chi-square distributed with $(b-1)(t-1)$ df and non-centrality parameter

$$\Delta^2_{Q_{\beta\tau}} = \frac{12}{N}\left(\int g^2(z)dz\right)\sum_i^t\sum_j^b n_{ij}\left[(\beta r)_{ij}-(\overline{\beta r})\right]^2, \quad (3.7)$$

where $(\beta r)_{ij}=\frac{1}{N}\sum_i^t\sum_j^b n_{ij}(\beta r)_{ij}$.

Under mild regularity conditions on $G(z)$ Andrews[6] proved this theorem for an asymptotically equivalent test statistic.

Under the normal shift model we have

$g(z) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(z-\mu)^2}{2\sigma^2}}$

and thus

$$\int g^2(z)dz = \frac{1}{2\sigma\sqrt{\pi}}\int e^{-\frac{(z-\mu)^2}{\sigma^2}}dz = \frac{1}{2\sigma\sqrt{\pi}}. \quad (3.8)$$

This leads to

$$\Delta^2_{Q_{\beta\tau}} = \frac{3}{N\pi\sigma^2}\sum_i^t\sum_j^b n_{ij}\left[(\beta r)_{ij}-(\overline{\beta r})\right]^2. \quad (3.9)$$

The corresponding F-test will under the normal shift model be asymptotically non-central Chi-square distributed with $(b-1)(t-1)$ df and non-centrality parameter

$$\Delta^2_{F_{bt}} = \frac{1}{N\sigma^2}\sum_i^t\sum_j^b n_{ij}\left[(\beta r)_{ij}-(\overline{\beta r})\right]^2. \quad (3.10)$$

The asymptotic efficiency of the $Q_{\beta\tau}$-test relative to the F_{bt}-test is given by the ratio of the two non-centrality parameters

$$e_{Q_{\beta\tau},F_{bt}} = \Delta^2_{Q_{\beta\tau}}/\Delta^2_{F_{bt}} = 3/\pi \quad (3.11)$$

Under the normal shift model the ARE of the $Q_{\beta\tau}$-test to its corresponding parametric test is asymptotically equivalent to that of the Kruskal-Wallis test to the F-test in the one-way analysis of variance. However, as in the standard ANOVA case the power of the test for interactions is lower than that for the main effects. This can for example lead to the ignorance of the variability in the treatment effects in a multi-center study. In practice it is therefore advisable to use a higher significance level for the interaction test, say 10% instead of 5%.

Comment: If the distribution $F_{ij}(y)$ given in expression (2.1) is symmetric theorem 3.2 and 3.3 holds also for $Q^*_{\beta\tau}$

MULTIPLE COMPARISONS

When the overall test of no interaction effects is rejected it is of interest to see which cells differ significantly. This can be done by multiple comparisons. We propose the modified sequentially rejective Bonferroni correction to control the simultaneous significance level, since it is superior to the classical Bonferroni correction w.r.t. power, see.[25]

The method can be described as follows. Say that we want to perform all $C=\binom{tb}{2}$ significance tests on the cell differences $\mu_{ij}-\mu_{kl}$, with a simultaneous significance level of at most α. The individual tests $H_0: \mu_{ij}-\mu_{kl}=0$ can be performed using the techniques of the one-way layout, where the cells are the groups. Let R_{ijk} denote the rank of Y_{ijk} in the joint ranking of the entire $N=\Sigma_i\Sigma_j n_{ij}$ observations. Consider the quantity

$$\frac{1}{\sqrt{N}}(R_{ij.}-R_{kl.}), \quad (3.12)$$

where as before $R_{ij.}=\frac{1}{n_{ij}}\sum_k R_{ijk}$. Under the null hypotheses it follows from the proof of theorem 3.1 that

$E(R_{ij.}) = (N+1)/2,$

$Var(R_{ij.}) = (N-n_{ij})(N+1)/(12n_{ij})$

and

$Cov(R_{ij},R_{kl.}) = -(N+1)/12.$

Hence for the test statistic in (3.12) we get

$$E\left[\frac{1}{\sqrt{N}}(R_{ij.}-R_{kl.})\right]=0 \quad (3.13)$$

and

$$Var\left[\frac{1}{\sqrt{N}}(R_{ij.}-R_{kl.})\right] = \frac{(N+1)}{12}\left(\frac{1}{n_{ij}}+\frac{1}{n_{kl}}\right) \quad (3.14)$$

In conformity with the results in[5] we have under the null hypothesis that the test statistic $(R_{ij.}-R_{kl.})/\sqrt{N}$ is asymptotically normal distributed with mean and variance given by expressions (3.13) and (3.14) respectively.

The asymptotic two-sided p-value for the difference between cell (i, j) and (k,l) is

$$p_{ij,kl} = 2\left[1-\Phi\left(\frac{1}{\sqrt{N}}(R_{ij.}-R_{kl.})\Big/\sqrt{\frac{(N+1)}{12}\left(\frac{1}{n_{ij}}+\frac{1}{n_{kl}}\right)}\right)\right] \quad (3.15)$$

where $\Phi(\cdot)$ denotes the standard Normal distribution. The classical Bonferroni correction implies that each $p_{ij,kl}$ is compared with α/C. The modified sequentially rejective Bonferroni correction is also based on the p-values. Let $p_{(1)}\le p_{(2)}\le\ldots\le p_{(C)}$ denote the ordered $p_{ij,\,kl}$:s. The procedure can be described as: Let s_c denote the maximum number of possibly true hypotheses, given that at least $c-1$

hypotheses are false. If $p_c \leq \alpha/s_c$ we reject the null hypothesis and continue by comparing $p_{(c+1)}$ with $\alpha/s_{(c+1)}$, else we stop and declare the actual and all remaining comparisons nonsignificant at level α, where $c=1,2,\ldots,C-1$. The possible values of s_c can be obtained from the recursion formula, see.[25]

$$A(bt) = \quad , \qquad (3.16)$$

where $A(bt)$ is the set of possible numbers of true hypotheses with bt cells, $bt \geq 2$ and $A(0)=A(1)=\{0\}$.

Comment: Above we have used the joint ranking in the pairwise comparisons.

Alternatively we could have used separate ranking for each comparison. For a discussion concerning the two ranking methods in this context see.[8]

SAMPLING EXPERIMENTS

To study the small sample behavior of the tests based on the statistic $Q_{\beta\tau}$ a sampling experiment was performed.

To achieve a reasonable precision 10000 replicates were used in all situations. The same random numbers were used for each parameter combination, to reduce the uncertainty when comparing results from different parameter combinations.

Uniformly distributed random numbers (U_j) were generated using the following multiplicative congruence generator

$I_j = 7^5\ I_{j-1} \bmod (2^{31}-1)$ where $I_0 \in [1, 2^{31}-2]$ and $U_j = I_j / 2^{31}$.

Finally the U_j's were randomly permuted in batches of 500 to improve their random properties. The standard normal random deviates (Z_j) were calculated using the following algorithm (polar method): Given that $S = (2U_j-1)^2 + (2U_{j+1}-1)^2 \leq 1$ will

$Z_j = (2U_j-1)/\sqrt{-2\log(S)/S}$ and $Z_{j+1} = (2U_{j+1}-1)\sqrt{-2\log(S)/S}$

be independent and normally distributed with $\mu = 0$ and $\sigma^2 = 1$ (log denotes the natural logarithm). Chi-square random variables with 2 df were generated using the inverse method $X_j = F^{-1}(U_j)$ where $F(x) = 1-\exp(-x/2)$. Since a sum of independent Chi-square distributed variables is Chi-square distributed with the df equal to the sum of the df of the included variables $X_j + X_{j+1}$ will be Chi-square distributed with 4 df.

Random variables from the Cauchy, t and contaminated normal distributions were generated using the normal/independent method i. e. they can be expressed as $X_j = Z_j / Y_{j+1}$ where $Z_j \sim N(0, 1)$ and $Y_{j+1} > 0$ are independent random variables. If $Y_{j+1} =$ X_j will be Cauchy distributed and if ν $= \chi^2(\nu)$, where $\chi^2(\nu)$ denotes a Chi-square distributed random variable with n df, X_j will be t distributed with ν df ($t(\nu)$). Finally if $Y_{j+1} = \sigma\ U_{j+1} \in (0, \kappa]$ and 1 otherwise X_j will be contaminated normally distributed, i. e. follow a standard normal distribution contaminated by on the average 100κ % observations coming from a normal distribution with variance σ^2 ($CN(\kappa,\sigma)$). A thorough description of the random number generation process is given in.[26]

The small sample results are divided into two parts. First we investigated how well the suggested finite sample null distribution of the $Q_{\beta\tau}$ statistic approximates the actual null distribution for some chosen levels α. As a control the level of the corresponding F-test was estimated for the same data sets at $\alpha = 5\%$.

The following model was used

$$Y_{ijk} = \beta_i + \tau_j + \varepsilon_{ijk},$$

where the error term ε_{ijk} had symmetry point equal to 0 and scale equal to the Fisher information for the symmetric error distributions and $E(\varepsilon_{ijk}) = 0$ and $Var(\varepsilon_{ijk}) = 1$ for the skewed distributions. Without loss of generality we can set $\beta_i \equiv 0$ all i and $\tau_j \equiv 0$ all j.

The results are summarized in Table 1 to 3 for different number of levels of the factors β and τ.

The tables show that the asymptotic nominal level corresponds well with the actual level. The general impression being that the 1% level tends to be too conservative and the 10% level too liberal independent of error distribution but more accentuated for smaller cell sizes. Approximate confidence intervals for the results based on the normal

TABLE 1: Actual significance levels (in per cent) for the test of interaction effects. Factor effects estimated by the H/L-estimator. Number of levels in 1st and 2nd factor (4, 2).

Error distribution	Cell sizes n_{ij}	Asymptotic nominal level α in percent; statistic $Q_{\beta\tau}$		F-test nominal level	
		10%	5%	1%	5%
Cauchy[2]	4	11.2	5.3	0.7	1.7
	8	10.5	5.4	0.9	1.4
	16	10.2	4.6	0.9	1.6
	32	9.6	4.8	1.0	1.6
	n^1	11.3	5.6	0.7	-
	$2*n^1$	10.8	5.4	1.0	-
t(2)[2]	4	11.0	4.5	0.5	3.1
	8	10.8	4.9	0.7	3.5
	16	10.7	5.1	0.9	3.9
	32	10.0	4.9	0.9	3.8
	n^1	11.1	5.4	0.9	-
	$2*n^1$	10.2	5.1	1.0	-
t(4)[2]	4	10.6	4.6	0.5	4.3
	8	10.4	4.9	0.9	4.6
	16	10.1	4.6	0.7	4.6
	32	9.5	4.7	1.0	4.7
	n^1	11.3	5.4	0.7	-
	$2*n^1$	10.4	4.9	0.7	-
CN(.1,3)[3]	4	11.0	4.6	0.4	4.6
	8	10.7	4.8	0.6	4.7
	16	9.7	4.6	0.8	4.6
	32	10.1	5.2	1.0	5.0
	n^1	10.4	5.0	0.7	-
	$2*n^1$	9.9	4.9	0.9	-
N(0,1)[2]	4	10.9	4.7	0.4	5.1
	8	10.6	5.0	0.6	4.9
	16	9.8	4.5	0.8	4.8
	32	10.1	5.1	1.0	4.9
	n^1	10.5	4.9	0.7	-
	$2*n^1$	10.0	4.7	0.9	-
$\sqrt{(\chi^2(2)/2)}$[4]	4	10.7	4.6	0.4	4.6
	8	10.2	4.6	0.7	5.0
	16	10.4	5.0	0.8	5.0
	32	9.6	4.9	0.9	4.9
	n^1	10.5	5.0	0.9	-
	$2*n^1$	9.8	4.8	0.8	-
$\sqrt{(\chi^2(4)/4)}$[4]	4	10.4	4.9	0.5	5.0
	8	10.4	5.0	0.6	5.0
	16	10.4	5.0	1.0	4.9
	32	10.2	4.7	0.8	5.0
	n^1	10.3	4.8	0.6	-
	$2*n^1$	10.0	4.9	0.7	-

[1] n = [4 4; 16 4; 4 8; 16 8]

[2] Fisher information $[I(F)]^{-1} = (\nu + 3)/(\nu + 1)$, where ν is the degrees of freedom

[3] $[I(F)]^{-1} \approx 1.256201$

[4] $E(\chi) \approx (1 - 1/2\nu)^{1/2}\{1 + [16\nu(\nu - 1)]^{-1}\}$ and $Var(\chi) \approx (2\nu)^{-1/2}(1 - 1/4\nu - 1/8\nu^2 + 5/64\nu^3)$

TABLE 2: Actual significance levels (in per cent) for the test of interaction effects. Factor effects estimated by the H/L-estimator. Number of levels in 1st and 2nd factor (4, 3).

Error distribution	Cell sizes n_{ij}	Asymptotic nominal level α in percent; statistic $Q_{\beta\tau}$		F-test nominal level	
		10%	5%	1%	5%
Cauchy[2]	4	10.9	4.5	0.3	1.7
	8	10.7	4.9	0.7	1.8
	16	10.0	4.6	0.7	1.6
	32	10.1	4.9	0.8	1.4
	n^1	10.9	5.3	0.9	-
	$2*n^1$	10.1	5.0	0.8	-
t(2)[2]	4	10.1	4.1	0.3	3.4
	8	10.3	4.8	0.7	4.0
	16	10.7	5.4	0.9	4.0
	32	9.5	4.6	0.9	3.6
	n^1	10.7	4.9	0.7	-
	$2*n^1$	10.7	5.1	1.0	-
t(4)[2]	4	10.6	4.4	0.4	4.6
	8	10.7	4.9	0.8	5.1
	16	9.9	4.7	0.7	4.6
	32	10.0	4.8	0.8	4.7
	n^1	10.9	5.0	0.7	-
	$2*n^1$	10.0	4.7	0.8	-
CN(.1,3)[3]	4	10.2	4.0	0.3	4.4
	8	9.6	4.3	0.7	4.4
	16	10.2	5.0	0.8	5.0
	32	10.0	4.8	0.9	4.9
	n^1	9.8	4.4	0.5	-
	$2*n^1$	10.4	4.7	0.7	-
N(0,1)[2]	4	10.2	4.2	0.3	5.1
	8	9.7	4.3	0.6	4.6
	16	10.3	5.0	0.8	5.2
	32	10.0	4.9	1.0	4.9
	n^1	9.8	4.4	0.5	-
	$2*n^1$	10.4	4.9	0.7	-
$\sqrt{(\chi^2(2)/2)}$[4]	4	10.7	4.6	0.5	5.3
	8	10.2	4.7	0.6	5.0
	16	10.2	5.1	0.9	5.0
	32	10.4	5.0	1.0	5.0
	n^1	10.3	4.4	0.7	-
	$2*n^1$	10.4	4.8	0.8	-
$\sqrt{(\chi^2(4)/4)}$[4]	4	10.8	4.7	0.4	5.2
	8	9.7	4.3	0.6	4.7
	16	9.8	4.7	0.8	4.6
	32	10.2	5.0	0.8	5.3
	n^1	10.3	4.1	0.5	-
	$2*n^1$	9.5	4.7	0.7	-

[1] n = [4 4 4; 16 4 8; 4 4 8; 16 16 8]

[2] Fisher information $[I(F)]^{-1} = (\nu + 3)/(\nu + 1)$, where ν is the degrees of freedom

[3] $[I(F)]^{-1} \approx 1.256201$

[4] $E(\chi) \approx (1 - 1/2\nu)^{1/2}\{1 + [16\nu(\nu - 1)]^{-1}\}$ and $Var(\chi) \approx (2\nu)^{-1/2}(1 - 1/4\nu - 1/8\nu^2 + 5/64\nu^3)$

TABLE 3: Actual significance levels (in per cent) for the test of interaction effects. Factor effects estimated by the H/L-estimator. Number of levels in 1st and 2nd factor (6, 3).

Error distribution	Cell sizes n_{ij}	Asymptotic nominal level α in percent; statistic $Q_{\beta\tau}$		F-test nominal level	
		10%	5%	1%	5%
Cauchy[2]	4	11.0	4.5	0.3	1.9
	8	10.7	5.0	0.7	1.6
	16	9.6	4.8	0.7	1.7
	32	9.5	4.5	0.8	1.7
	n[1]	11.2	5.1	0.7	-
	2*n[1]	10.4	5.1	0.8	-
t(2)[2]	4	11.1	5.0	0.5	3.6
	8	11.0	5.3	0.9	3.9
	16	10.5	5.2	1.0	4.0
	32	9.6	4.4	0.9	3.8
	n[1]	10.3	4.6	0.5	-
	2*n[1]	10.3	4.9	0.8	-
t(4)[2]	4	10.7	4.4	0.4	4.2
	8	11.0	5.2	0.8	4.8
	16	9.8	4.7	0.8	4.9
	32	9.8	4.8	0.8	4.9
	n[1]	10.9	5.1	0.6	-
	2*n[1]	10.2	4.9	0.8	-
CN(.1,3)[3]	4	10.8	4.6	0.5	4.7
	8	10.2	4.6	0.8	4.7
	16	10.2	4.7	0.9	5.0
	32	10.4	5.2	0.9	5.0
	n[1]	10.6	4.7	0.6	-
	2*n[1]	10.3	4.7	0.9	-
N(0,1)[2]	4	10.8	4.4	0.5	5.1
	8	10.4	4.8	0.8	5.0
	16	10.5	4.9	0.9	4.9
	32	10.4	5.0	0.9	5.1
	n[1]	10.4	4.6	0.7	-
	2*n[1]	10.4	4.9	0.9	-
$\sqrt{(\chi^2(2)/2)}$[4]	4	10.9	4.2	0.5	5.0
	8	10.0	4.5	0.7	4.8
	16	10.1	4.9	1.1	5.0
	32	10.7	5.1	0.8	5.2
	n[1]	10.4	4.8	0.7	-
	2*n[1]	10.1	4.6	0.8	-
$\sqrt{(\chi^2(4)/4)}$[4]	4	11.0	4.6	0.5	5.0
	8	10.2	4.4	0.5	4.6
	16	10.0	4.8	0.8	4.8
	32	10.2	5.0	1.0	5.2
	n[1]	10.4	4.7	0.5	-
	2*n[1]	10.2	4.9	0.7	-

[1] n = [4 4 4; 16 4 8; 4 4 8; 16 16 8; 8 8 4; 4 4 16]

[2] Fisher information $[I(F)]^{-1} = (\nu + 3)/(\nu + 1)$, where ν is the degrees of freedom

[3] $[I(F)]^{-1} \approx 1.256201$

[4] $E(\chi) \approx (1 - 1/2\nu)^{1/2}\{1 + [16\nu(\nu - 1)]^{-1}\}$ and $Var(\chi) \approx (2\nu)^{-1/2}(1 - 1/4\nu - 1/8\nu^2 + 5/64\nu^3)$

approximation are given by the observed actual level ±0.6%, ±0.4% and ±0.2% for α = 10%, 5% and 1% respectively.

To study the power (Π) of the tests against different alternatives the following model was used

$$Y_{ijk} = \beta_i + \tau_j + (\beta\tau)_{ij} + \varepsilon_{ijk},$$

where the error term ε_{ijk} had symmetry point equal to 0 and scale equal to the Fisher information for the symmetric error distributions and $E(\varepsilon_{ijk}) = 0$ and $Var(\varepsilon_{ijk}) = 1$ for the skewed distributions.

The following interaction effects were used

$$(\boldsymbol{\beta\tau}) = 0.25\, W_{bt},$$

where

$$W_{bt} = \begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{pmatrix}$$

$$W_{bt} = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

and

$$W_{bt} = \begin{pmatrix} 1 & -1 & -1 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix}.$$

As above we can set $\beta_i \equiv 0$, all i and $\tau_j \equiv 0$, all j without loss of generality.

The results are summarized in Figure 2 to 4 which show the logit transform of the power ($\log[\Pi/(1-\Pi)]$) in relation to the total sample size N.

The graphs show that under the normal shift model the power of the Q-test (Π_Q) is close to the F-test (Π_F) in all cases. This is consistent with expression (3.11). For the longer tailed distributions the Q-test outperforms the F-test.

NUMERICAL COMPARISONS WITH OTHER TEST STATISTICS

Numerical comparisons between the $Q_{\beta\tau}$ test and other proposals are presented in Table 4. The datasets are briefly described in the table, where also references to the papers describing them are given. The different test statistics and references to them are given in the footnotes to the table.

As can be seen the $Q_{\beta\tau}$ test behaves well in most situations, especially when aligning on the H/L-estimator or the mean. The results of the

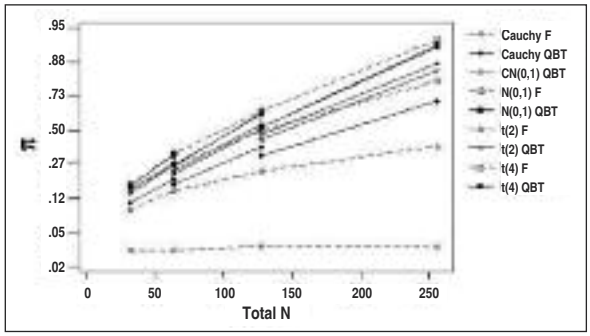


FIGURE 2: Power of the Qβτ -test and corresponding F-test for interactions (logit scale). Factor effects estimated by the H/L-estimator. Number of levels in 1st and 2nd factor (4, 2).

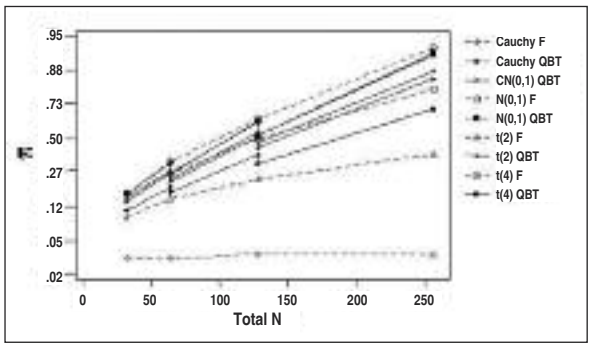


FIGURE 3: Power of the Qβτ -test and corresponding F-test for interactions (logit scale). Factor effects estimated by the H/L-estimator. Number of levels in 1st and 2nd factor (4, 3).

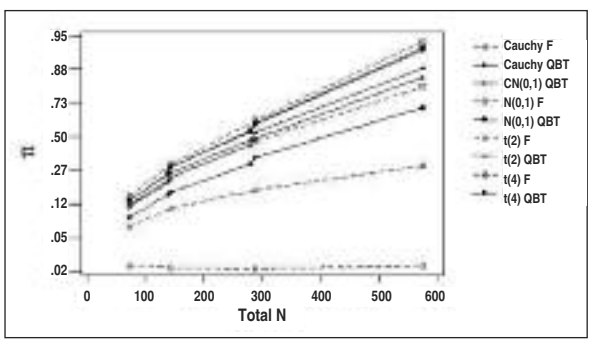


FIGURE 4: Power of the Qβτ -test and corresponding F-test for interactions (logit scale). Factor effects estimated by the H/L-estimator. Number of levels in 1st and 2nd factor (6, 3).

Brunner and Puri's small sample test $F_N(\boldsymbol{M})$ and the standard ANOVA test are close in most cases. The model-fitting criterion rank test D^* and the aligned rank test S^* suggested by[3] behave like our proposals for the survival data, the D^* test being liberal.[15]

Finally some simulation results comparing our proposal $Q_{\beta\tau}$ with Brunner and Puri's large and small sample statistics $Q_N(\boldsymbol{C})$ and $F_N(\boldsymbol{M})$ are given in Table 5. The results for $Q_N(\boldsymbol{C})$ and $F_N(\boldsymbol{M})$ are from Table 3 in.[27] As can be seen from the table the actual significance levels are close to the nominal levels for the $Q_{\beta\tau}$ and the $F_N(\boldsymbol{M})$ statistics, $F_N(\boldsymbol{M})$ being a little more conservative, while for the $Q_N(\boldsymbol{C})$ statistic the actual significance levels heavily exceeds the nominal levels. In additional simulations (data not presented in their paper) Brunner, Dette and Munk[27] studied the validity of the significance levels of the $F_N(\boldsymbol{M})$ statistic for other distributions (both longtailed and skew), unequal cell sizes and other experimental designs. They concluded that they got similar results, which corresponds to our results for the $Q_{\beta\tau}$ statistic (see Table 1 to 3).

THE SALVe PROJECT

MATERIAL AND METHODS

All 16-year-old (ninth grade students, compulsory school) and 19-year-old (third grade students, secondary school) in Västmanland, a medium-sized county of Sweden, a total of 2987 ninth graders and 2186 third graders, comprised the target population. The students were asked to complete a mental and psycho-social health-screening questionnaire, "Survey of Adolescent Life in Vestmanland (SALVe)" during a 1-h session under the supervision of a specially trained research assistant. In total there were 2611 (mean age 16.0 year) and 1649 (mean age 19.2 year) students, 87 and 75 percent respectively, who completed the questionnaire. All students had an opportunity to give their informed consent to participate in an in-depth-interview and the drawing of a blood sample, by giving their full personal security number at the form front page. Informed consent was received from 785 students who could be traced with valid names. All students were classified with a risk index, depending on their risk behaviors (Alcohol, Narcotic, Sexual, Property offence, and Violent offence) reported in the questionnaire, and divided into quartiles according to their respective risk index. A random sample of 400 students, one-third from the lower,

TABLE 4: Comparisons between the $Q_{\beta\tau}$ test of interaction effects and other proposals.

	$Q_{\beta\tau}$ /NDF statistic, p-value Factor effects estimator			
Dataset	**H/L**	**mean**	**median**	**Test statistic, p-value**
Hettmannsperger's example on survival data a 3x4 factorial with 4 observations per cell, p.270-5 (2)	2.62 0.028	2.51 0.034	2.66 0.026	1.31, 0.296[1] 3.35, 0.010[2] 14.77, 0.022[3] 1.87, 0.113[4]
Hettmansperger & McKean,s example on life time of motors a 3x3 factorial with cell sizes ranging from 3 to 6 and one cell with zero variation, pp 255-8 (20)	2.21 0.085	2.07 0.103	2.43 0.063	1.42, 0.233[1] 2.69, 0.040[5] 1.30, 0.292[4]
Sawilowsky's fabricated data a 2x2 factorial with 10 observa-tions per cell, pp 109-15 (15)	4.34 0.042	4.23 0.044	4.57 0.037	4.54, 0.040[1] 4.45, 0.042[6] 4.34, 0.037[7] 3.98, 0.054[4]
Brunner & Puri's example on Dr. Beusher's fertility trial a 2x3 factorial with cell sizes ranging from 8 to 13, pp 24-5 (22)	1.13 0.331	1.45 0.242	1.09 0.346	0.92, 0.403[1] 1.78, 0.412[8] 0.83, 0.441[4]

1) Brunner & Puri's small sample test F_N (*M*) approx F [(b-1)(t-1),N-bt]
2) The model-fitting criterium rank test D* approx F [(b-1)(t-1),N-bt]
3) Aligned rank test S* approx χ^2 [(b-1)(t-1)]
4) Standard ANOVA F ~ F [(b-1)(t-1),N-bt]
5) R-analysis based on Wilcoxon scores approx F [(b-1)(t-1),N-bt]
6) Adjusted RT approx F [(b-1)(t-1),N-bt]
7) Puri & Sen's L approx F [(b-1)(t-1),N-bt]
8) Brunner & Puri's large sample test Q_N (C) approx χ^2 [(b-1)(t-1)]

TABLE 5: Actual significance levels (in per cent). Comparisons between the Qbt test for interaction effects (factor effects estimated by the H/L-estimator) and Brunner & Puri's large and small sample statistics QN(C) and FN(M). Number of levels in 1st and 2nd factor (6, 3).

		Asymptotic nominal level α								
		$Q_{\beta\tau}$ statistic			Q_N (C) statistic			F_N (M) statistic		
Error distribution	**Cell sizes n_{ij}**	**10%**	**5%**	**1%**	**10%**	**5%**	**1%**	**10%**	**5%**	**1%**
N(0,1)	4	10.8	4.4	0.5						
	7				24.1	16.8	7.3	9.2	4.2	0.6
	8	10.4	4.8	0.8						

one-third from the two middle and one-third from the upper quartile balanced w.r.t. age and gender, were drawn from the volunteers. The procedure with an initial risk survey was used to ensure that we should have enough participants from both ends of the deviant behavior continuum. There were no explicit exclusion criteria. Written informed consent was obtained from 81 boys and 119 girls who agreed to give blood samples and to take part in an interview when asked to participate a second time. The study was approved by the research ethics committee at Uppsala University, Sweden and complies with the Declaration of Helsinki.

The risk index showed no significant differences between the group interviewed and those responding to the initial questionnaire. For a more comprehensive description of the participants, see.[28]

Venous blood was drawn from all interviewed students for molecular genetic analyses. One boy and one girl were excluded due to hepatitis infection. DNA was extracted from venous blood using standard methods. PCR-based genotyping of the AP-2ß intron 2 polymorphism was performed as described in.[29] With regard to the 5-HTTLPR polymorphism, PCR-based genotyping was performed according to a modified protocol by Collier.[30] Finally for the MAO-A polymorphism PCR-based genotyping was performed as described in.[31]

To confirm that the correct regions of the AP-2ß gene, the 5-HTTLPR gene and the MAO-A gene were amplified, PCR products representing all genotypes were sequenced using BigDye®Terminator chemistry (Applied Biosystems) and analyzed by an automated ABI PRISM 310 Genetic Analyzer (ABI PRISM™ Perkin Elmer, Foster city, CA, USA). The DNA fragments were analyzed using the Squencer™ 3.1.1 software (Gene Codes Corporation, Ann Arbor, MI, USA).

RESULTS

Results from the following three SALVe study projects will be presented:

1. 5-HTTLPR Genotype (short/short, long/short and long/long alleles) and Quality of Family Relations (categorized into bad, equal and good) in relation to Alcohol Consumption, see.[28] Cell sizes: 6, 5, 33; 20, 9, 61; 9, 7, 46.
2. MAO-A Genotype (short present and long/long alleles) and Psychosocial Risk (defined as either living in multi-family house and/or having experience of violent victimization) in relation to Total Criminality Activity among Boys, see.[31] Cell sizes: 21, 11; 29, 17.
3. 5-HTTLPR and AP-2ß genotypes (short present and long/long alleles) in relation to Cloninger's Temperament and Character Inventory (TCI) dimensions especially Self Transcendence and its sub-scale Spiritual Acceptance among Boys and Girls, see.[32] Cell sizes: 24, 24; 22, 9 and 51, 34; 20, 11 for boys and girls respectively.

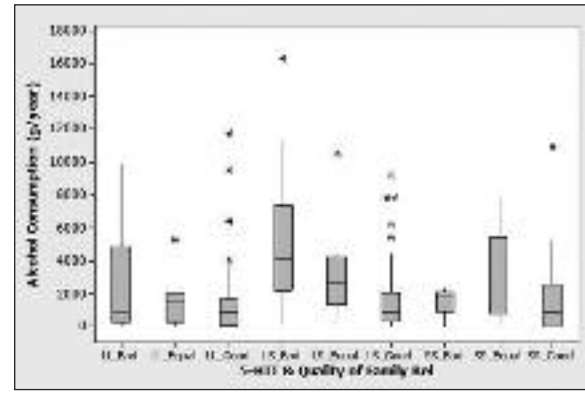


FIGURE 5: Box plots of the total alcohol consumption per year for the subgroups 5-HTTLPR gene expression combined with family relations.

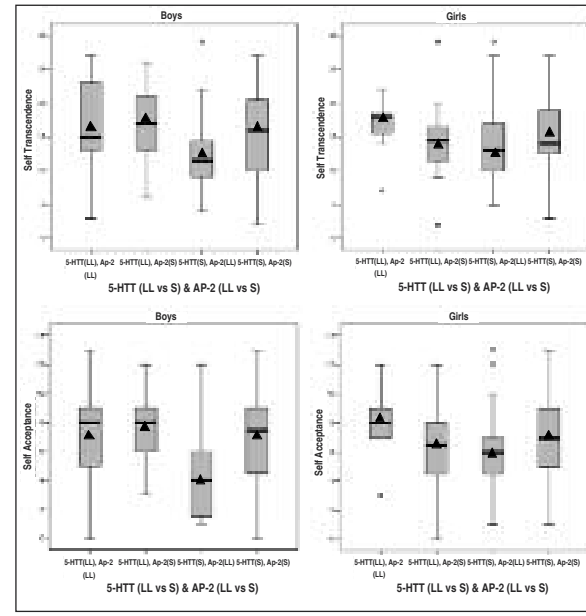


FIGURE 6: Box plots of genotype combination subgroups for Self-Transcendence and Spiritual Acceptance for boys and girls separately. H/L estimates marked as black triangles in the box plots.

The response variable in the 1st study is a positively skewed continuous variable, see Figure 5. In the 2nd and 3rd studies the response variables are tied in integer values but commonly this variables are treated as they were metric variables see e.g.[33] Box plots of the response variables in 3rd study are shown in Figure 6.

The results are presented in Table 6. As can be seen in the table the results are similar in most

TABLE 6: Gene – Environment and Gene – Gene interactions from the SALVe project. Comparisons between the $Q_{\beta\tau}$ test for interaction effects and Brunner & Puri's small sample statistic FN(M).

Dataset	$Q_{\beta\tau}$ /NDF statistic, p-value Factor effects estimator			FN (M) statistic, p-value
	H/L	mean	median	
5-HTTLPR Genotype and	3.79	3.96	3.53	0.68
Quality of Family Relations in	0.006	0.004	0.009	0.583
in relation to Alcohol Consumption				
in Adolescents (28)				
MAO-A Genotype and Psychosocial	7.15	6.93	6.41	8.60
Risk in relation to Total	0.009	0.010	0.013	0.005
Criminality among Boys (31)				
5-HTTLPR and AP-2ß genotypes				
in relation to the TCI				
character Self Transcendence				
and its sub-scale Spiritual				
Acceptance among boys and girls (32)				
Boys: Self Transcendance	2.251	2.412	2.05	1.94
	0.133	0.120	0.152	0.179
Boys: Spiritual Acceptance	0.69	0.74	0.64	0.49
	0.415	0.396	0.431	0.491
Girls: Self Transcendance	5.51	6.38	4.83	5.97
	0.020	0.012	0.028	0.018
Girls: Spiritual Acceptance	5.06	4.97	5.12	5.39
	0.025	0.026	0.024	0.021

[1] Adjusting the variance for ties using expression (3.3) gives $Q_{\beta\tau}$ /NDF =2.35 with p-value = 0.129

[2] Adjusting the variance for ties using expression (3.3) gives $Q_{\beta\tau}$ /NDF =2.53 with p-value = 0.116

cases, especially aligning using the H/L-estimator and the mean gives very similar results. Finally, when comparing with Brunner and Puri's small sample test $F_N(\boldsymbol{M})$ the only large difference occurs in the alcohol consumption study where the $Q_{\beta\tau}$ statistic give a significant interaction while $F_N(\boldsymbol{M})$ doesn't. Looking at the box plots in Figure 5 we can see that 5-HTTLPR heterozygous (LS) adolescents are sensitive to the quality of the family relations – increasing alcohol consumption as the family relations get worse. In homozygous (SS or LL) adolescents the quality of family relations doesn't affect the alcohol consumption. This is in consistence with the outcome of the $Q_{\beta\tau}$ statistic. In Figure 6 the box plots of the genotype combinations for Self-Transcendence and Spiritual Acceptance are given for boys and girls separately. Their appearance is also in consistence with the outcome of the test statistics.

Finally for the subscale Self-Transcendence also expression (3.3) was used instead of $N(N+1)/12$ in the $Q_{\beta\tau}$ statistic to adjust the variance for ties. The results using expression (3.3) are given in the footnotes to Table 6 and show that the differences between the adjusted and unadjusted values are negligible.

CONCLUSION

The proposed test for interaction effects shows good asymptotic and small sample properties. It can be used in the general two-way layout and multiple comparisons can be performed in a straightforward way applying the modified sequentially rejective Bonferroni correction to control the total experimental error. Mid-ranks can be used to break ties and as long as the data are not tied in a few values there is no need for an adjustment of the variance in the expression for the $Q_{\beta\tau}$ statistic as was

seen from the result in Table 4 and 6. Further the test remains valid w.r.t. type I error even when there are great differences between the cell sizes.

The comparisons with other proposals in Table 4 are advantageous for our proposal. The simulation comparisons between our proposal $Q_{\beta\tau}$ and Brunner and Puri's small sample statistics $F_N(\boldsymbol{M})$ in Table 5 show that the actual significance levels are close to the nominal levels for both statistics, $F_N(\boldsymbol{M})$ being a little more conservative. Although the $Q_{\beta\tau}$ statistic was evaluated for cell sizes 4 and 8 and $F_N(\boldsymbol{M})$ for cell size 7 the comparison is reliable since the behavior of the $Q_{\beta\tau}$ statistic for a cell size equal 7 could be interpolated to lie between those of cell size 4 and 8. The analyses of the three sub studies in the SALVe project show that the $Q_{\beta\tau}$ statistic can be useful in studies like these and its overall performance is at least as good as Brunner and Puri's small sample test $F_N(\boldsymbol{M})$.

The proposed test has the potential to pick up interaction effects hidden in the noise of the gross errors when using standard ANOVA. The test may become a valuable tool and alternative to other proposals in exploring interaction effects.

A computer program for the interaction test written in Fortran is available from the author upon request.

Appendix

Iterative elimination of the main effects

We will use the indices $h(i,k) = k + \sum_{s=1}^{i-1} n_{sj}$ and $l(j,k) = k + \sum_{s=1}^{j-1} n_{is}$ when there is no need to separate between factor β and subjects within factor β and factor τ and subjects within factor τ respectively, i.e. $y_{h(i,k)j} = y_{ijk}$ and $y_{il(j,k)} = y_{ijk}$. To further simplify the notation we will write y_{hj} and y_{il}.

If we let $\Delta\hat{\beta}_i^{(s)}$ and $\Delta\hat{\tau}_j^{(s)}$ denote the change in the estimates of β_i and τ_j in the sth iteration the aligning procedure can be described as follows:

Initial conditions $\hat{\beta}_i^{(0)}=0,\ i = 1...,b$ and $\hat{\tau}_j^{(0)}=0,\ j = 1,...t$.

First iteration:

Calculate the H/L estimate of factor β's effects including the overall mean level

$$\Delta\hat{\beta}_i^{(1)}=\underset{l\le l'}{\text{med}}\left\{\frac{1}{2}(Y_{il}+Y_{il'})\right\}.\quad i = 1,...,b,$$

Then set

$$T_{ijk}^{(1)}=Y_{ijk}-\Delta\hat{\beta}_i^{(1)}$$

and calculate the H/L estimate of factor τ's effects

$$\Delta\hat{\tau}_j^{(1)}=\underset{h\le h'}{\text{med}}\left\{\frac{1}{2}(T_{hj}^{(1)}+T_{h'j}^{(1)})\right\}.\quad j = 1,...,t.$$

The residuals after the first iteration are given by

$$Z_{ijk}^{(1)}= T_{ijk}^{(1)}-\Delta\hat{\tau}_j^{(1)}.$$

At the mth iteration we have:

$$\Delta\hat{\beta}_i^{(m)}=\underset{l\le l'}{\text{med}}\left\{\frac{1}{2}(Z_{il}^{(m-1)}+Z_{il'}^{(m-1)})\right\},\quad i = 1,...,b,$$

$$T_{ijk}^{(m)}=Z_{ijk}^{(m-1)}-\Delta\hat{\beta}_i^{(m)}$$

$$\Delta\hat{\tau}_j^{(m)}=\underset{h\le h'}{\text{med}}\left\{\frac{1}{2}(T_{hj}^{(m)}+T_{h'j}^{(m)})\right\},\qquad j = 1,...,t,$$

$$Z_{ijk}^{(m)}= T_{ijk}^{(m)}-\Delta\hat{\tau}_j^{(m)},$$

$$\hat{\beta}_i^{(m)}=\sum_{s=1}^{m}\Delta\hat{\beta}_i^{(s)},\quad i = 1,...,b,$$

and

$$\hat{\tau}_j^{(m)}=\sum_{s=1}^{m}\Delta\hat{\tau}_j^{(s)}, \quad j = 1,..., t.$$

The iteration process is stopped when all $\Delta\beta_i^{(m)}$'s and $\Delta\tau_j^{(m)}$'s are small in magnitude compared to the $Z_{ijk}^{(m)}$'s. Let Z_{ijk} denote the *ijk*th residuals after the final iteration and $\hat{R}_{ijk}$ its rank in the joint ranking of the entire $N = \Sigma_i\Sigma_j n_{ij}$ aligned observations.

Proof of theorem 3.1

Let $\theta = (N+1)/2$. $Q_{\beta\tau}$ can be rewritten as

$$Q_{\beta\tau}=\frac{12}{N(N+1)}\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(\left(\hat{R}_{ij.}-\theta\right)-\left(\hat{R}_{i..}-\theta\right)-\left(\hat{R}_{.j.}-\theta\right)\right)^2$$

$$=\frac{12}{N(N+1)}\left(\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(\hat{R}_{ij.}-\theta\right)^2-\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(\hat{R}_{i..}-\theta\right)^2-\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(\hat{R}_{.j.}-\theta\right)^2\right)$$

$$+\frac{12}{N(N+1)}2\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\ \Sigma_{k=1}^{b}\Sigma_{l=1}^{t}n_{ij}\frac{n_{kj}n_{il}}{n_{i.}n_{.j}}\left(\left(\hat{R}_{kj.}-\theta\right)\left(\hat{R}_{il.}-\theta\right)\right), \qquad (A.1)$$

since

$$\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(\left(\hat{R}_{ij.}-\theta\right)\left(\hat{R}_{i..}-\theta\right)\right)=\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(\hat{R}_{i..}-\theta\right)^2 \text{ and}$$

$$\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(\left(\hat{R}_{ij.}-\theta\right)\left(\hat{R}_{.j.}-\theta\right)\right)=\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(\hat{R}_{.j.}-\theta\right)^2.$$

The term $\frac{12}{N(N+1)}2\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\ \Sigma_{k=1}^{b}\Sigma_{l=1}^{t}n_{ij}\frac{n_{kj}n_{il}}{n_{i.}n_{.j}}\left(\left(\hat{R}_{kj.}-\theta\right)\left(\hat{R}_{il.}-\theta\right)\right)$

equals zero if all $n_{ij} = n$ otherwise it can be shown that it is of size $O(N^{-1})$ if for each (i,j) $\lim_{N\to\infty} n_{ij}/N$ exists and is positive.

When H_0 is true $E(\hat{R}_{ij.}) = \theta$. Given that the factor effects β and τ are equal to zero the expression

$$\frac{12}{N(N+1)}\left(\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(\hat{R}_{ij.}-\theta\right)^2-\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(\hat{R}_{i..}-\theta\right)^2-\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(\hat{R}_{.j.}-\theta\right)^2\right)$$

is asymptotically equivalent with

$$\frac{12}{N(N+1)}\left(\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(R_{ij.}-\theta\right)^2-\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(R_{i..}-\theta\right)^2-\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(R_{.j.}-\theta\right)^2\right),$$

where $R_{ij.}$, $R_{i..}$ and $R_{.j.}$ denote the sums of the unaligned ranks . Applying the results in Corollary 10 (8) page 395-396 lead to

$$E\left\{\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(R_{ij.}-\theta\right)^2\right\}=\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}Var\left(R_{ij.}\right)=\frac{N(N+1)}{12}(bt\text{ -1}),$$

$$E\left\{\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(R_{i..}-\theta\right)^2\right\}=\Sigma_{i=1}^{b}n_{i.}Var\left(R_{i..}\right)=\frac{N(N+1)}{12}(b\text{ -1}) \text{ and}$$

$$E\left\{\Sigma_{i=1}^{b}\Sigma_{j=1}^{t}n_{ij}\left(R_{.j.}-\theta\right)^2\right\}=\Sigma_{j=1}^{t}n_{.j}Var\left(R_{.j.}\right)=\frac{N(N+1)}{12}(t\text{ -1}).$$

The result follows by substituting these values in the expression for $E(Q_{\beta\tau})$.■

REFERENCES

1. Friedmann M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Assoc 1937;32(200): 675-701
2. Hettmannsperger TP. Statistical Inference Based on Ranks. 1st ed. New York: John Wiley & Sons. 1984.
3. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952;47(260):583-621.
4. Kruskal WH. A nonparametric test for several sample problem. Ann Math Statist 1952; 23(4):525-40.
5. Andrews F. Asymptotic behavior of some rank tests for analysis of variance. Ann Math Statist 1954;25(4):724-36.
6. Hodges JL, Jr., Lehmann EL. Rank methods for combination of independent experiments in the analysis of variance. Ann Math Statist 1962;33(2):482-97.
7. Lehmann EL, D'Abrera HJM Nonparametrics: statistical methods based on ranks. In: D'Abrera HJM, ed. 1st ed. Holden-Day, 1975.
8. Mehra KL, Sarangi J. Asymptotic efficiency of certain rank tests for comparative experiments. Ann Math Statist 1967;38(1):90-107.
9. Sen PK. On a class of aligned rank order tests in two-way layouts. Ann Math Statist 1968; 39(4):1115-24.
10. Mehra KL, Sen PK. On a class of conditionally distribution-free tests for interactions in factorial experiments. Ann Math Statist 1969;40(2):658-64.
11. Öhrvik J. Nonparametric methods in the two-way layout. Chiang Mai J Sci 2002;29(2):103-15.
12. Boos DD, Brownie C. A rank-based mixed model approach to multisite clinical trials. Biometrics 1992;48(1):61-72.
13. Brunner E, Puri ML, Sun S. Nonparametric methods for Stratified two-sample designs with applications to multiclinical trials. J Am Stat Assoc 1995;90(431):1004-14.
14. Patel KM, Hoel DG. A nonparametric test for interactions in factorial experiments. J Am Stat Assoc 1973;68:615-20.
15. Sawilowsky SS. Nonparametric tests of interaction in experimental design. Rev Educ Res 1990;60(1):91-126.
16. Conover WJ, Iman RL. On some alternative procedures using ranks for the analysis of experimental designs. Commun Stat Theory Methods 1976;5(14):1349-68.
17. Conover WJ, Iman RL. Rank transformations as a bridge between parametric and nonparametric statistics. Am Stat 1981;35(3):124-9.
18. SAS Institute. SAS procedures guide: version 6. 3rded. SAS Institute, 1990.
19. Akritas MG. The rank transform method in some two-factor design. J Am Stat Assoc 1990;85:73-8.
20. Hettmansperger TP, McKean JW. Robust Nonparametric Statistical Methods. 1st ed. London: Arnold; 1998.
21. Akritas MG, Arnold SF, Brunner E. Nonparametric hypotheses and rank statistics for unbalanced factorial design. J Am Stat Assoc 1997;92(437):258-65.
22. Brunner E, Puri ML. Nonparametric methods in factorial designs. Statistical Papers 2001; 42(1):1-52.
23. Emerson JD, Hoaglin DC. Analysis of two-way tables by medians. Understanding Robust and Exploratory Data Analysis. 1st ed. In: Hoaglin DC, Mosteller F, Tukey JW, eds. New York: John Wiley & Sons; 1983.
24. Hampel FR. The influence curve and it's role in robust estimation. J Am Stat Assoc 1974;69(346):383-93.
25. Shaffer JP. Modified sequentially rejective multiple test procedures. J Am Stat Assoc 1986;81(395):826-31.
26. Öhrvik J. Robust efficient estimation of location. (Thesis). Department of statistics, Stockholm university Stockholm. 1982.
27. Brunner E, Dette H, Munk A. Box-type approximations in nonparametric factorial designs. J Am Stat Assoc 1997;92(440): 1494-502.
28. Nilsson KW, Sjöberg RL, Damberg M, Alm PO, Ohrvik J, Leppert J, et al. Role of the serotonin transporter gene and family function in adolescent alcohol consumption. Alcohol Clin Exp Res 2005;29(4):564-70.
29. Damberg M, Eller M, Tõnissaar M, Oreland L, Harro J. Levels of transcription factors AP-2alpha and AP-2beta in the brainstem are correlated to monoamine turnover in the rat forebrain. Neurosci Lett 2001;313(1-2):102-4.
30. Collier DA, Stöber G, Li T, Heils A, Catalano M, Di Bella D, et al. A novel functional polymorphism within the promoter of the serotonin transporter gene: possible role in susceptibility to affective disorders. Mol Psychiatry 1996;1(6):453-60.
31. Nilsson KW, Sjöberg RL, Damberg M, Leppert J, Ohrvik J, Alm PO, et al. Role of monoamine oxidase A genotype and psychosocial factors in male adolescent criminal activity. Biol Psychiatry 2006;59(2):121-7.
32. Nilsson KW, Damberg M, Ohrvik J, Leppert J, Lindström L, Anckarsäter H, et al. Genes encoding for AP-2beta and the Serotonin Transporter are associated with the Personality Character Spiritual Acceptance. Neurosci Lett 2007;411(3):233-7.
33. Brandstrom S, Schlette P, Przybeck TR, Lundberg M, Forsgren T, Sigvardsson S, et al. Swedish normative data on personality using the Temperament and Character Inventory. Compr Psychiatry 1998;39(3):122-8.