

5 ex.
mit Hemelrijk

MATHEMATISCH CENTRUM

2e BOERHAAVESTRAAT 49

AMSTERDAM

STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig

Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

SP 20

Some remarks on the combination
of independent tests,

by

J. Hemelrijk.

1953.

Some remarks on the combination
of independent tests.

by J. Hennebrink.

1. ~~Introduction.~~ The use of the integral transformation for the combination of independent tests of ~~the~~ statistical hypotheses has first been introduced by R.A. Fisher (1932). The present note is an extension of a paper of E.S. Pearson (1930) on this subject and ~~immainly~~ contains some results about the combination of two-sided tests, which are not given in that paper. ~~In order to give a clear picture and which are of some practical importance.~~ The methods used are the same as Pearson's and ~~in~~ in order to give a clear picture of the situation, some of his results are summarized in the following section.

~~2. We consider m independent tests of m hypotheses $H_1, H_2, \dots, H_m$, which may, but need not, be identical. Let the test of H_i ($i=1, \dots, m$) be based on ~~the~~ a statistic $\underline{t}_i$.²⁾ ~~Under the hypotheses $H_1, \dots, H_m$ the statistics $\underline{t}_i$ must have to be~~ stochastically independent and we denote their distribution functions, under H_i , by F_i .~~

$$(1) \quad F_i(x | H_i) \stackrel{\text{def}}{=} P[\underline{t}_i \leq x | H_i]$$

and, on the other hand

$$(2) \quad R_i(x | H_i) \stackrel{\text{def}}{=} P[\underline{t}_i \geq x | H_i].$$

then defining

$$(3) \quad \left\{ \begin{array}{l} \underline{u}_i \stackrel{\text{def}}{=} F_i(\underline{t}_i | H_i) \\ \underline{v}_i \stackrel{\text{def}}{=} R_i(\underline{t}_i | H_i) \\ \underline{w} \stackrel{\text{def}}{=} 2 \min(\underline{u}_i, \underline{v}_i) \end{array} \right. \quad (i=1, \dots, m)$$

²⁾ Random variables ~~are~~ underlined in order to ~~give~~ distinguish them from ordinary variables ~~as~~ and from values assumed by random ordinary numbers (e.g. values assumed by random variables, which ~~are~~ are sometimes denoted by the same letter, ~~as the random variables~~ not underlined).

Experiment

2. Consider m hypotheses $H_1, H_2, \dots, H_m$, which may, but need not, be identical. ~~Suppose that, in independent experiments have been performed, to test these~~
~~order to test these hypotheses, an experiment has been performed~~
 Suppose that these hypotheses are to be tested on the evidence of m independent experiments, one for each of them. Let $\underline{t}_i$ ($i=1, \dots, m$) be the test criterion ~~for~~ used for testing H_i . The experiments being independent, the statistics $\underline{t}_i$ are stochastically independent. The problem we wish to consider is, how to test $H_1, \dots, H_m$ simultaneously on the available data, i.e. how to combine the tests based on $\underline{t}_1, \dots, \underline{t}_m$ into one test of the hypothesis H_0 , which is the conjunction of $H_1, \dots, H_m$:

$$(1) \quad H_0 \equiv H_1 \text{ and } H_2 \text{ and } \dots \text{ and } H_m.$$

Pearson derived a solution for this problem by means of the likelihood ratio principle²⁾ for several sets of admissible hypotheses.

If we denote the ^(cumulative) probability distribution function of $\underline{t}_i$, under hypothesis H_i , by

$$(2) \quad F_i(x | H_i) \stackrel{\text{def}}{=} P[\underline{t}_i \leq x | H_i]$$

and if we define

$$(3) \quad R_i(x | H_i) \stackrel{\text{def}}{=} P[\underline{t}_i \geq x | H_i],$$

then the variables

$$(4) \quad \begin{cases} \underline{u}_i \stackrel{\text{def}}{=} F_i(\underline{t}_i | H_i) \\ \underline{v}_i \stackrel{\text{def}}{=} R_i(\underline{t}_i | H_i) \\ \underline{w}_i \stackrel{\text{def}}{=} 2 \min(\underline{u}_i, \underline{v}_i) \end{cases} \quad (i=1, \dots, m)$$

are random variables, which satisfy the following inequalities

$$(5) \quad \begin{cases} P[\underline{u}_i \leq y] \leq y \\ P[\underline{v}_i \leq y] \leq y \\ P[\underline{w}_i \leq y] \leq y \end{cases} \quad (0 \leq y \leq 1; i=1, \dots, m).$$

The equality sign holds for every y between 0 and 1, ~~when~~ if the corresponding $\underline{t}_i$ is continuously distributed;

¹⁾ Random variables have been underlined are denoted by underlined symbols, to distinguish them from ordinary variables or numbers. Values ~~of the~~ assumed by random variables are sometimes denoted by the same symbol, not underlined.

²⁾ Cf. J. Neyman and E. S. Pearson (1933).

(4) denotes the above mentioned integral transformation in three slightly different forms and u_i , e.g., is the random probability, that a second observation of t_i , taken independently of a first observation of this statistic, will not exceed this first one.

We shall mainly consider the case, ^{that} ~~where~~ $t_1, \dots, t_m$ are continuously distributed (some remarks about the discontinuous case are given in section 6); then u_i , v_i and w_i will, under H_0 , have a ~~be~~ rectangular distribution.

Pearson considered i.a. the following sets of ^(Ω_1, Ω_2 and Ω_3) admissible hypotheses.

$$(6) \Omega_1: f_i(u|H') = (a+1)u^a \quad (-1 < a \leq 0; i=1, \dots, m),$$

$$(7) \Omega_2: g_i(v|H'') = (b+1)v^b \quad (-1 < b \leq 0; i=1, \dots, m),$$

$$(8) \Omega_3: h_i(u|H''') = \frac{\Gamma(2c+2)}{[\Gamma(c+1)]^2} u^c (1-u)^c \quad (-1 < c \leq 0; i=1, \dots, m).$$

Here f_i ~~and~~ and h_i denote the probability density of u_i , ~~under~~ g_i the probability density of v_i ; a , b and c respectively characterise simple admissible hypotheses H' , H'' and H''' respectively, which coincide with H_0 if $a=0$, $b=0$ and $c=0$ respectively. The statistics $u_1, \dots, u_m$ are ^{again} independently distributed, corresponding to the independence of the m original tests of $H_1, \dots, H_m$. The same applies to $v_1, \dots, v_m$.

The type of the hypotheses of these three classes of admissible hypotheses has been sketched in figure 1.

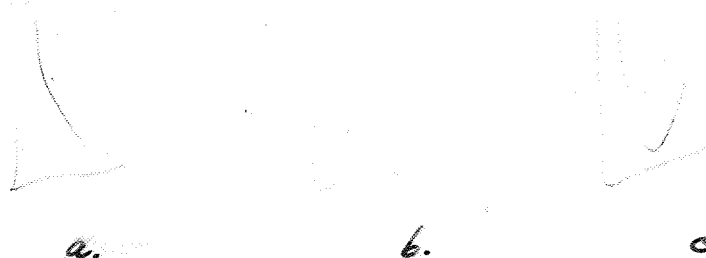


Fig. 1. Classes Ω_1, Ω_2 and Ω_3 of admissible hypotheses.

Using ~~the~~ the likelihood ratio principle, Pearson found, that the statistics

$$(9) \quad U \stackrel{\text{def}}{=} -2 \ln \prod_{i=1}^m u_i,$$

$$(10) \quad V \stackrel{\text{def}}{=} -2 \ln \prod_{i=1}^m v_i$$

and

$$(11) \quad W \stackrel{\text{def}}{=} -2 \ln \prod_{i=1}^m w_i,$$

and V respectively
 large values of U, V
 being taken as
 critical values.

where w_i is defined by the last equation (4), may be used as test criteria for H_0 with Ω_1, Ω_2 and Ω_3 respectively as the set of admissible hypotheses. It is not difficult to prove, that under H_0 each of these statistics has a χ^2 -distribution with $2m$ degrees of freedom (not approximately ~~this time~~ in this case, but exactly). These tests, being λ -tests, may be expected to have a satisfactory power function.

3. The classes Ω_1 and Ω_2 may be considered as a satisfactory model of situations, often encountered in the practice of statistical analysis, where one-sided tests are to be combined. Ω_3 was meant to represent the two-sided case. For most two-sided tests, however, - in fact for all two-sided tests, which the present author can call to mind - the alternative hypothesis would never take the form of ~~the~~ a hypothesis H^m of Ω_3 .
 Let e.g. t_i be the ~~test~~ Student's t -test for the hypothesis, that the mean of a normally distributed variable is equal to zero. If this is not the case, the mean will either be negative - and then the probability distribution of u_i will resemble ~~have a~~ resemble fig. 1, a) - or it will be positive - and then the distribution of u_i will be something like fig. 1, b) - but in no case will the distribution of u_i have the form of figure 1, c).

The relevant class of admissible hypotheses for the two-sided combination of two-sided tests seems to be

$$(12) \quad \Omega_4: \text{ for every } i=1, \dots, m \text{ separately either} \\
 f_i(u|H^*) = (a+1)u^a, \text{ or } g_i(v|H^*) = (a+1)v^a, \\
 \text{with } -1 < a \leq 0,$$

~~a again characterizing the simple hypothesis~~
~~hypothesis H^* of Ω_4 , which coincides with H_0~~

f_i and g_i again denoting the probability densities of u_i and v_i under the simple hypothesis H^* of Ω_4 , H^* being characterized by the value of a and coinciding with H_0 if $a=0$; $u_1, \dots, u_m$ are again stochastically independent.

(and also $v_1, \dots, v_m$)

It is to be expected, that the difference between the classes Ω_3 and Ω_4 will not influence the result of the application of the likelihood ratio principle with Ω_4 as class of admissible hypotheses will yield the same result as found for Ω_3 . This is indeed the case, as will be proved in section 4.

There is, however, another class of admissible hypotheses, which is worth considering and for which the likelihood ratio test does not coincide with the ~~one~~ ^{one} of the above mentioned tests. When using Ω_4 , we admit the possibility, that some of the u_i will have a distribution of the type of fig. 1, a and some others of the type of fig. 2, b. It may be, and as a matter of fact it is far from rare, that either all of these ~~to~~ probability distributions are of the first or all of them are of the second type mentioned. This will e.g. be true, if the m original hypotheses to be tested, $H_1, \dots, H_m$, are identical, with identical ~~or~~ ^{nearly identical} alternatives hypotheses. We may, as an example, think of $\bar{x}_i$ as Student's t , for every i ~~with~~ ^{for testing} the mean of normal distributions, which have the same mean for $i=1, \dots, m$ but possibly different standard deviations. In that case, the hypothesis, that this mean is equal to 0, cannot be tested by pooling the m sets of data, because of the different standard deviations, ~~and thus a method of combining the data is necessary, and as the result is the same for every i the class Ω_4 contains hypotheses, which should not~~ and thus another method of combining the data is necessary. The class Ω_4 ~~is~~ ^{is} The means (unknown) mean being the same for every i , the class Ω_4 contains hypotheses which should be excluded and we must look for another class of admissible hypotheses, which is relevant for two sided tests of this type. This class ~~seems~~ We may then consider the class

probability distributions of $u_1, \dots, u_m$

(12) Ω_5 : ~~for every $i=1, \dots, m$ we have either~~

(13) Ω_5 : either $f_i(u | H^*) = (a+1)u^a$ ~~($a < 0$)~~
for every $i=1, \dots, m$, or $g_i(v | H^*) = (a+1)v^a$ ~~($a < 0$)~~
for every $i=1, \dots, m$, with $-1 < a \leq 0$,

with f_i, g_i and a ~~on the~~ ^{having the} same meaning as before, ~~again~~ $u_1, \dots, u_m$ (and also $v_1, \dots, v_m$) being stochastically independent. The class Ω_5 may also be defined as the union of Ω_1 and Ω_2 .

We shall prove, in section 4, that ^{in this case} The likelihood ratio principle yields a test based on the statistic

$$(14) \quad X \stackrel{\text{def}}{=} \text{Max}(\underline{U}, \underline{V}),$$

with $\underline{U}$ and $\underline{V}$ defined by (9) and (10), large values of X being critical. This criterion has been proposed by K. Pearson (1934), who, however, did not take into account that X does not have a χ^2 -distribution with $2m$ degrees of freedom. Consequently he underrated the probability of an error of the first kind. We shall show, that under H_0 we have

$$(15) \quad P[X \geq d(\alpha) | H_0] < 2\alpha,$$

if $d(\alpha)$ is the $(1-\alpha)$ -quantile of a χ^2 -distribution with $2m$ degrees of freedom. ^{then} Using (15) X may be used to test H_0 against Ω_5 and, being a likelihood ratio test, this test may be expected to give better results than $\underline{U}, \underline{V}$ or $\underline{W}$ for classes of admissible hypotheses resembling Ω_5 .³⁾

4. Following the proofs of E. S. Pearson (1938) we consider the function

$$\text{def } \varphi(a, p) = (a+1)^m p^a \quad (-1 \leq a \leq 0; 0 \leq p \leq 1),$$

where m is a given positive integer. It is easy to prove, that for given p this function has a maximum

$$(17) \quad \psi(p) \stackrel{\text{def}}{=} \text{Max}_{-1 \leq a \leq 0} \varphi(a, p) = \begin{cases} 1 & \text{if } p > e^{-m} \\ \frac{m}{p} \left(\frac{e^{-1}}{-\ln p} \right)^m & \text{if } p \leq e^{-m}, \end{cases}$$

and that

$$(18) \quad \psi(p_1) \leq \psi(p_2) \quad \text{if } p_1 \geq p_2.$$

Denoting the simultaneous probability density of $\underline{u}_1, \dots, \underline{u}_m$ under hypothesis H^* by $f(\underline{u}_1, \dots, \underline{u}_m | H^*)$, we have, when H^* is a hypothesis of Ω_4 :

$$(19) \quad f(\underline{u}_1, \dots, \underline{u}_m | H^*) = (a+1)^m \left\{ \prod_{i=1}^m z_i \right\}^a \quad (-1 \leq a \leq 0),$$

³⁾ The test in this form was proposed to the author by Mr. J. ^{K.E.} ~~van~~ ^{van} in ~~his~~ personal communication, which formed the inducement to this study.

where z_i , for every i separately, is either equal to u_i or to v_i .
The likelihood ratio is

$$(10) \quad \lambda = \frac{f(u_1, \dots, u_m | H_0)}{\max_{H^* \in \Omega_4} f(u_1, \dots, u_m | H^*)}$$

The numerator of this ratio is equal to 1 (cf. (5)) and the denominator ~~must~~ be computed by choosing a suitable value of α in (19) and by taking z_i equal to u_i or v_i in such a way, that the right hand ~~side~~ ^{member} of (19) is maximized. According to (17) and (18) this is accomplished by minimizing

$$\prod_{i=1}^m z_i$$

by means of the choice of z_i . According to definition (4), ~~this minimum is equal to 2^{-m} .~~

$$2^{-m} \prod_{i=1}^m w_i = 2^{-m} W,$$

~~and (cf. (11)) this is equal to W defined by (1).~~
We thus have

$$(11) \quad \lambda = 2^m \cdot \max_{-1 \leq \alpha \leq 0} \{(\alpha+1)^m W^\alpha\}$$

~~and the λ_0 smallest values of λ are critical, the statistic we must take large the critical region of a test likelihood ratio test is always of the form~~

$$(12) \quad \lambda \leq \lambda_0.$$

~~we must take large values of W as critical values. This completes the proof for the test with Ω_4 as class of admissible hypotheses.~~

Now consider Ω_5 . We then have

$$(13) \quad f(u_1, \dots, u_m | H^{**}) \text{ is either equal to } (\alpha+1)^m \left\{ \prod_{i=1}^m u_i \right\}^\alpha \text{ or to } (\alpha+1)^m \left\{ \prod_{i=1}^m v_i \right\}^\alpha.$$

The numerator of λ is again equal to 1 and we must now maximize $f(u_1, \dots, u_m | H^{**})$ by choosing H^{**} suitably from Ω_1 or Ω_2 or Ω_5 , i.e. from Ω_1 or Ω_2 . It follows from (17) and (18), that we should take the smallest of the two products $\prod_{i=1}^m u_i$ and $\prod_{i=1}^m v_i$. According to (14) this means, that

According to (10)
 λ does not increase
when W does.

λ provides a likelihood ratio test, again with large values as critical values, for

According to definition (4) this minimum is equal to

$$2^{-m} \prod_{i=1}^m w_i = 2^{-m} e^{-\frac{1}{2} W}$$

as may be seen from (11). We thus have

$$(11) \quad \lambda = 2^m \left\{ \max_{-1 < a \leq 0} (a+1)^m e^{-\frac{1}{2} a W} \right\}$$

and, according to (18) this function is a non-increasing function of W . As the critical region of a likelihood ratio test is always of the form $\lambda \leq \lambda_0$ we must take large values of W as critical values. This completes the proof for the test with Ω_4 as class of admissible hypotheses.

Now consider Ω_5 . The numerator of λ is again equal to 1 and we must now maximize $f(u_1, \dots, u_m) / h(x)$ by choosing H^* suitably from Ω_5 . This probability density is equal either equal to

$$(a+1)^m \left\{ \prod_{i=1}^m u_i \right\}^a$$

or to

$$(a+1)^m \left\{ \prod_{i=1}^m v_i \right\}^a$$

and it follows again consulting (17) and (18) we see, that we must take the smallest of the two products $\prod u_i$ and $\prod v_i$ to reach the maximum. ~~Now~~ The definitions (9) and (10) may also be written as

$$\prod_{i=1}^m u_i = e^{-\frac{1}{2} U} \quad ; \quad \prod_{i=1}^m v_i = e^{-\frac{1}{2} V},$$

and this gives

$$\min \{ \prod u_i, \prod v_i \} = \min \{ e^{-\frac{1}{2} U}, e^{-\frac{1}{2} V} \} = e^{-\frac{1}{2} X}$$

according to (14). Therefore we find

$$(12) \quad \lambda = \max_{-1 < a \leq 0} \{ (a+1)^m e^{-\frac{1}{2} a X} \},$$

and which means, that X , with large values critical, yields a likelihood ratio test for H_0 against Ω_5 .

4. We shall now prove that $\underline{X}$ (cf (12)) provides a likelihood ratio test for H_0 , when S_0 is the class of admissible hypotheses.

In this case, we must maximize $f(u_1, \dots, u_m | H)$ by choosing a suitable H from Ω_1 or Ω_2 . ~~Restricting ourselves to Ω_1~~ This means, that $f(u_1, \dots, u_m | H)$ is either equal to

$$(a+1)^m \left\{ \prod_{i=1}^m u_i \right\}^a$$

or to

$$(a+1)^m \left\{ \prod_{i=1}^m u_i \right\}^a$$

It follows from ~~that~~ ~~According to (16)~~ the maximum will be reached by choosing an alternative H from Ω_1 or Ω_2 , ~~respectively~~, according to $\prod u_i < \prod u_i$ or $\prod u_i > \prod u_i$ respectively. This means, however, that $\underline{X}$, ~~with~~ defined by (12), provides a likelihood ratio test with large values as critical values.

~~The proof of (15) is now complete. We now have to prove (16), to make the test complete. This is very easy. The probability distribution of $\underline{X}$, under H_0 , satisfies (15); ~~this~~ may be seen as follows:~~

$$P[\underline{X} \geq d(\alpha) | H_0] = P[\underline{U} \geq d(\alpha) \text{ or } \underline{V} \geq d(\alpha) | H_0] \leq$$

$$\leq P[\underline{U} \geq d(\alpha) | H_0] + P[\underline{V} \geq d(\alpha) | H_0] = 2\alpha,$$

if H_0 is true.

$\underline{U}$ and $\underline{V}$ having a χ^2 distribution with $2m$ degrees of freedom. As a matter of fact the ~~same~~ probability, that $\underline{U} \geq d(\alpha)$ and $\underline{V} \geq d(\alpha)$ is positive and we may replace $\leq$ by $<$. The difference between the true level of significance and 2α is very small, when d is not large, ~~may be~~ This is illustrated by fig. 2, where the critical region ~~is~~ has been sketched for the case $m=2$.

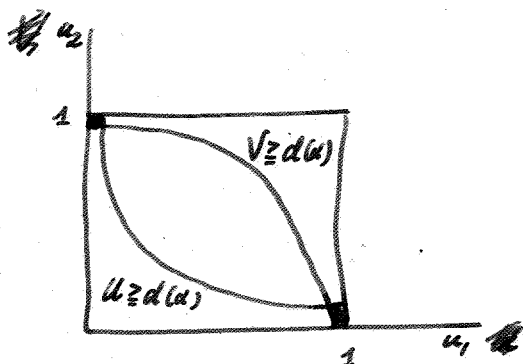


Fig. 2. Critical region $\underline{X} \geq d(\alpha)$.

$$\underline{U} = -2 \ln u_1, u_2; \underline{V} = -2 \ln u_1, u_2.$$

5. It is of interest to compare both tests for the two-sided case, based on $\underline{W}$ and $\underline{X}$ respectively. ~~for some numerical cases~~ This has been done for some numerical cases Ω_5 being a subclass of Ω_4 , this comparison is only of interest for hypotheses belonging to Ω_5 . For alternative hypotheses, which do not belong to Ω_5 , $\underline{X}$ should not be used. Restricting ourselves to Ω_5 , we are especially interested in cases, where all values $u_1, \dots, u_m$ found in an experiment are smaller than $\frac{1}{2}$ (or all larger; this comes to the same thing). Then (4) yields

$$w_i = 2u_i$$

and (11) gives

$$W = -2 \ln \prod_{i=1}^m w_i = -2 \ln \prod_{i=1}^m u_i - 2m \ln 2 = X - 2m \ln 2$$

because of (14). This relation

$$(23) \quad W = X - 2m \ln 2$$

is not generally valid, of course; in fact it has been derived under the supposition, that $u_1, \dots, u_m$ are all smaller than $\frac{1}{2}$. Thus it applies only to those cases values of $u_1, \dots, u_m$ satisfying that condition and the random variables $\underline{W}$ and $\underline{X}$ are not connected by a relation like (23).

With this restriction ~~Restricting ourselves to cases where the values of $u_1, \dots, u_m$, found in the experiments, are all smaller than $\frac{1}{2}$, we may compare the two~~ [tail probabilities

$$(24) \quad p_X \stackrel{\text{def}}{=} 2P[X \geq x | H_0]$$

and

$$(25) \quad p_W \stackrel{\text{def}}{=} P[W \geq x - 2m \ln 2 | H_0],$$

where $x = -2 \ln \prod_{i=1}^m u_i$

denotes the value of X found in the experiment ^{deriving} from a table of the χ^2 -distribution with $2m$ degrees of freedom. This gives e.g. the following numerical results.

m	$x - 2m \ln 2$	p_W	p_X
2	6	0.2	0.13
2	9	0.06	0.04
2	12	0.016	0.01
5	16	0.10	0.024
5	20	0.03	0.006
10	30	0.07	0.003

[two sided]

It is clear, that in such cases χ yields a far better result than W . It must be remarked, however, that this has not been proved for experiments, where ~~not~~ the values $u_1, \dots, u_m$ do not all lie at the same side of $\frac{1}{2}$.

[then
when applied
in the original
form,

6. The case of discontinuously distributed statistics $\underline{t}_i$ (cf. section 1) has been treated by W.A. Wallis (1942), who proved, that [the tests given by Pearson] result in an overestimation of the size of the critical region. This overestimation may be considerable, ~~which~~ ^{also} ~~which~~ means, that the power of the tests is much ~~smaller~~ ^{diminished} for the discontinuous $\underline{t}_i$ than for the continuous case. This result of Wallis applies to the test based on χ too. Methods to avoid this ~~disadvantage~~ disadvantage have been developed by H.O. Lancaster (1949) and ~~by~~ E.S. Pearson (1950).

References

- ✓ Fisher, R.A. (1932). Statistical methods for Research Workers, 4th ed.
- ✓ Lancaster, H.O. (1949). The combination of probabilities arising from data in discrete distributions. Biometrika, 36, 370.
- ✓ Neyman, J. and E.S. Pearson (1933). On the problem of the most efficient tests of statistical hypotheses. Phil. Trans., A, 231, 333.
- ✓ Pearson, E.S. (1938). The probability integral transformation for testing goodness of fit and combining independent tests of significance. Biometrika, 30, 134.
- ✓ Pearson, E.S. (1950). On questions raised by the combination of tests based on discontinuous distributions. Biometrika, 37, 303.
- ✓ Pearson, K. (1934). On a new method of determining goodness of fit. Biometrika, 26, 425.
- ✓ Wallis, W.A. (1942). Combining probabilities from independent significance tests. Biometrika, 40, 239.