UNIVERSITY OF GÖTEBORG
Department of Statistics

RESEARCH REPORT 1986:3

ISSN - 0349-8034

A TIPPETT-ADAPTIVE METHOD OF COMBINING INDEPENDENT STATISTICAL TESTS.

by

Margareta Westberg

Maj 1986

Statistiska institutionen
Göteborgs Universitet
Viktoriagatan 13
S 411 25 Göteborg
Sweden

Abstract

A new procedure is proposed in order to combine the information of P-values obtained from several independent tests in order to test an overall hypothesis. The test statistic of this new procedure is of the same type as Tippett's since in each step one of the P-values is compared with a constant. This new procedure is adaptive in the sense that the choice of P-value depends on the data. The procedure is very simple and in the performed examination this method is better than Tippett's in almost all situations. Thus this new "Tippett-adaptive" method is a good alternative to Tippett's procedure.

1. Introduction

$P_1, P_2, \ldots, P_k$ are independent tail-area probabilities arising from k continuous distributions of test statistics in k different experiments or tests. Since the test statistics have continuous distributions, then when the individual hypothesis H_{0i} is true, P_i is uniformly distributed over the interval $[0,1]$. Testing the combined null hypothesis

$$H_0 : H_{0i} \text{ is true for all } i$$

against

$$H_1 : H_{0i} \text{ is false for at least one } i$$

presents a problem in combination of tests, where only the P-values are to be used.

Many procedures have been proposed for combining the P-values arising from several independent tests in order to test whether all null hypotheses are true. Two commonly used methods of combining independent significance levels $P_1, \ldots, P_k$ are Fisher's procedure (1932) : H_0 is rejected if the product $P_1 \ldots P_k \leq c$. c is a constant which depends on the significance level for the overall test, and Tippett's procedure (1931) : H_0 is rejected if any of $P_1, P_2, \ldots, P_k \leq \alpha'$. If the overall significance level is α then $\alpha' = 1 - (1-\alpha)^{1/k}$.

In studies by Frisén (1974) and Westberg (1985 a) Fisher's method was compared with Tippett's according to the power. These studies show that the power functions have an intersection point and that neither of the methods is generally more powerful than the other.

In situations where a high power is desired when just one of the individual hypotheses is false and the deviation from H_0 is large Tippett's method is preferable. In other applications where it is more important to detect alternatives for which many of the individual hypotheses might be false, Fisher's method is likely to be preferable.

An indication of the number of false hypotheses could be used in an adaptive way to decide on an appropriate test statistic between Fisher's and Tippett's. This could be done in different ways.

In Westberg (1985 b) a test statistic which has the same structure as Fisher's, namely a product of P-values was proposed. The k ordered P-values are denoted by

$$P_{(1)} \geq P_{(2)} \geq \ldots \geq P_{(k-1)} \geq P_{(k)} .$$

Constants $a_i (i=1,\ldots,k)$ are chosen with

$$1 > a_1 > a_2 > \ldots > a_k > 0 .$$

The test is then based on the test statistic

$$z(n) = \prod_{i=k-n+1}^{k} (P_{(i)}/a_{k-n+1}) , \; n > 0 .$$

The stepwise procedure to determine the random variable n and the critical values are described in Westberg (1985 b). Here n is the greatest integer such that $P_{(k-n+1)} < a_{k-n+1}$.

It is suggested that the a_i's should be such that

$$a_i = a_1 \{1 - (i-1)/k\} .$$

This procedure seems to be a good procedure in the cases when Fisher's is a good procedure. In the case when n=k the test statistic is formally identical to Fisher's but the procedure is not the same as n is stochastic. In the case when n=1 the test statistic is identical to Tippett's but the procedure is not the same.

In the present paper another procedure is described. This procedure is more Tippett-like than that procedure proposed in Westberg (1985 b). The new procedure has a test statistic of the same type as Tippett's, namely a comparison of one P-value with a constant. This procedure is an adaptive one and has good power properties in the cases when Tippett's is a good procedure according to the power.

In the present study this Tippett-adaptive method is compared to Tippett's, Fisher's and the Fisher-adaptive procedures for normally distributed test statistics.

2. The Tippett-adaptive procedure

In situations when Tippett's method might be a good test it would be desirable with a procedure that is adaptive and Tippett-like in the sense that it has good properties when Tippett's is the best.

The "Tippett-adaptive" method is a test based on a comparison between just one P-value and a constant. The value of this constant depends both on the overall significance value and the number of individual hypotheses.

If the k ordered P-values are still denoted by

$$P_{(1)} \geq P_{(2)} \geq \ldots \geq P_{(k-1)} \geq P_k$$

and the constants $a_i (i=1,\ldots,k)$ are chosen with

$$1 > a_1 > a_2 > \ldots > a_k > 0 \ .$$

The test rejects the overall hypothesis H_0 if any of the attained significance levels $P_{(i)}$ is less than the corresponding constant a_i.

It is suggested that the a_i's should be such that

$$a_i = a_1 \{1 - (i-1)/k\} \ .$$

If a_1 is chosen to be α this test rejects H_0 at the level α since

$$1 - \alpha = \Pr \text{ (not reject } H_0) =$$

$$= \Pr(P_{(1)} \geq a_1 \cap \ldots \cap P_{(k-1)} \geq 2a_1/k \cap P_k \geq a_1/k) \ .$$

The latter expression is proved in a different context in Westberg (1985 b) to be $1 - a_1$.

The following example will illustrate this procedure: the attained significance levels of three tests are $P_{(1)}=0.08$, $P_{(2)}=0.03$ and $P_{(3)}=0.02$. The overall significance level α is 0.05 and k=3. Then $a_1=0.05$, $a_2=0.033$ and $a_3=0.0166$. Since $P_2=0.03$ is less than a_2 the overall hypothesis H_0 is rejected.

This procedure is extremely simple to use and may not be mistaken for Wilkinson's method (1951): H_0 is rejected if and only if $P_i \leq c$ for r or more of the P-values, where r is a predetermined integer, $1 \leq r \leq k$, and c is constant corresponding to the desired significance level. The k possible choices of r give k different procedures which are referred to as case 1 (r=1), case 2 (r=2) etc. For example, if k=2 and a test at level $\alpha=0.05$ is desired, the case 2 procedure is : reject H_0 if both P_1 and P_2 equal or exceed $c = 1 - (0.05)^{1/2} = 0.776$. Case 1 is the same procedure as the procedure proposed by Tippett.

3. Evaluations

The Tippett-adaptive procedure is compared to Fisher's, Tippett's and the Fisher-adaptive method according to their power.

The power function of Tippett's method is computed exactly, while the simulation technique was used to compute the power of the three other procedures.

The non-centrality parameter of the alternative normal distributions are m_i, $i=1,\ldots,k$ and m_i was assigned the values 0(0,5) 6.

The simulation was performed by generating at least 2000 standard normal random numbers for each of the k populations. The same set of numbers was used in all tests. 2000 replications will ensure a 95% confidence interval that is ±0.01 when the power is 0.05. When the power is 0.50 the corresponding interval is ±0.02. Sometimes the number of replications is 10000 which will be seen in the following.

The power is computed for the cases when the results from two and fifteen tests, respectively, are combined. In most of the calculations the value of the non-centrality parameter is the same for all alternatives, that is $m_i=m$. In the case when two tests are combined and both hypotheses are false the power is also computed for $m_1=m$ and $m_2=0.5m$. This result is presented in figure 3. AS can be seen the power curves are of the same shape as when $m_1=m_2=m$.

The results are displayed in figures 1-7. As can be seen from the power-graphs none of the methods is generally more powerful than the others.

It can be seen from the figures that in the cases examined this new method is rather "Tippett-like" since it has good properties when Tippett is the best method according to the power. The differences between Tippett and this new method are in these situations, almost negligible. In the other cases this Tippett-adaptive method is always better than Tippett's but not as good as Fisher's and the Fisher-adaptive method.

In order to establish that the Tippett-adaptive method really can be the best one of the four methods when two of the fifteen hypotheses are equally false the number of replications was chosen to be 10000 from each of the 15 populations. The power of the Tippett-adaptive method is 0.8685 and is 0.8575 of Tippett's method when the non-centrality parameter m=3.0. This difference is significant at the 5% level when normal approximation is used.

4. Conclusions

The procedure proposed in this paper is an adaptive procedure and more similar to Tippett's than the method proposed in Westberg (1985 b), which procedure is more similar to Fisher's.

In the cases examined the Tippett-adaptive method is better than Tippett's in almost all situations and nearly as good as Tippett's when Tippett's is the best one according to the power. This procedure is similar to Tippett's because just one P-value is compared to a constant. It can be any of the P-values but which one depends on the data.

The properties of this method depend on k, the number of the individual hypotheses and the choice of a_1 depends on the overall significance value α.

This Tippett-adaptive method is a good alternative to Tippett's since it is almost as simple as Tippett's and nearly always better than Tippett's procedure.

REFERENCES

Fisher, R.A. (1932). Statistical Methods for Research Workers, 4th ed. (Edingburgh, Olivier & Boyd).

Frisén, M. (1974). Stochastic Deviation from Elliptical Shape (Stockholm, Almqvist & Wiksell).

Tippett, L.M.C. (1931). The methods of Statistics (London, Williams & Norgate).

Westberg, M. (1985a). Combining independent statistical tests, The Statistician, 34, pp 287-296.

Westberg, M. (1985b). An adaptive method of combining independent statistical tests. Research Report 1985:06, Statistiska institutionen, Göteborgs universitet.

Wilkinson, B. (1951). A Statistical Consideration in Psychological Research, Psychological Bulletin, 48, pp 156-157.

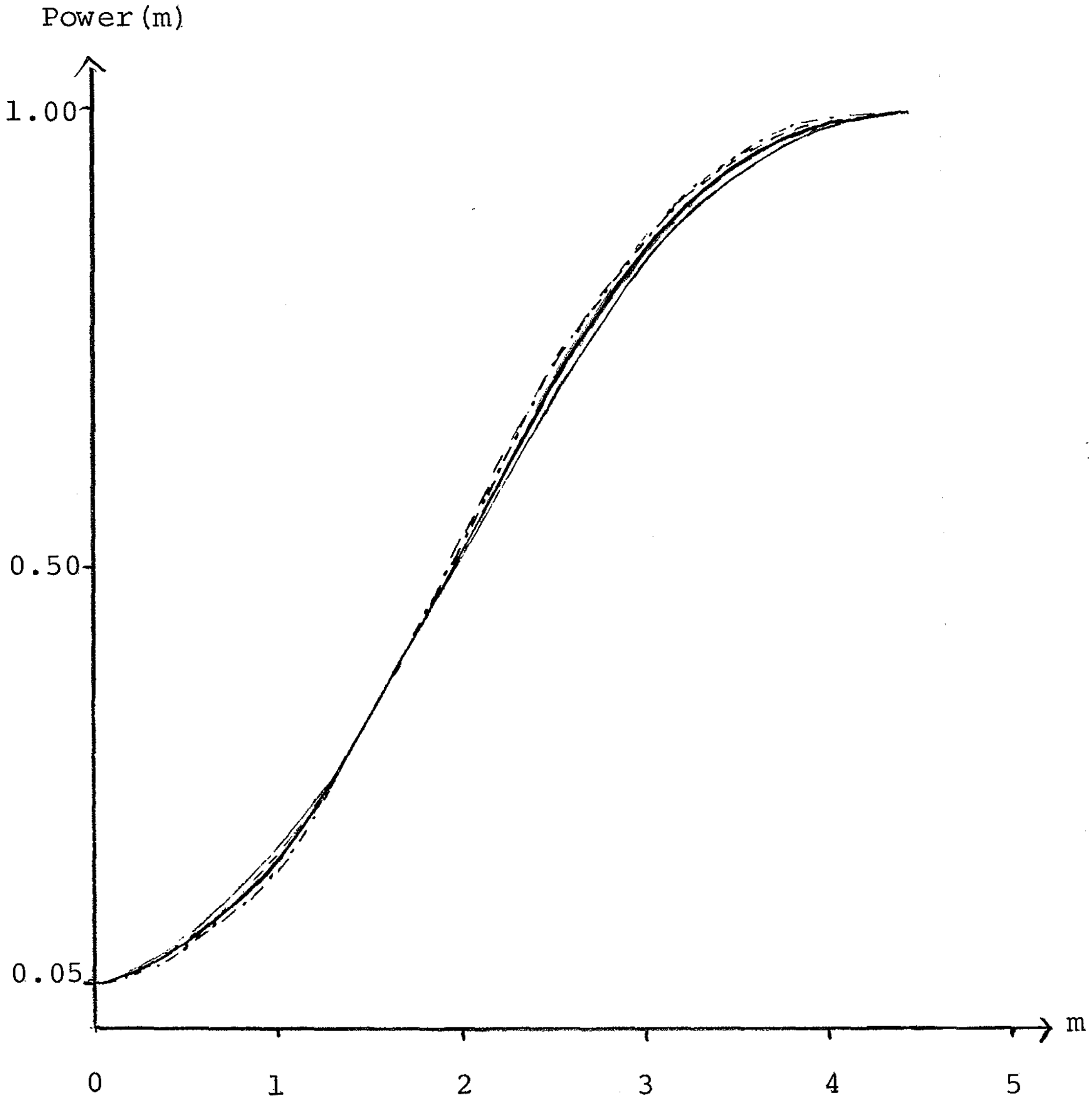

Fig 1

The power graphs when results from two tests are combined. One of the hypotheses is false with the non-centrality parameter m.

Fisher's method ———— Fisher-adaptive ————

Tippett's method — — — — Tippett-adaptive —·—·—·—

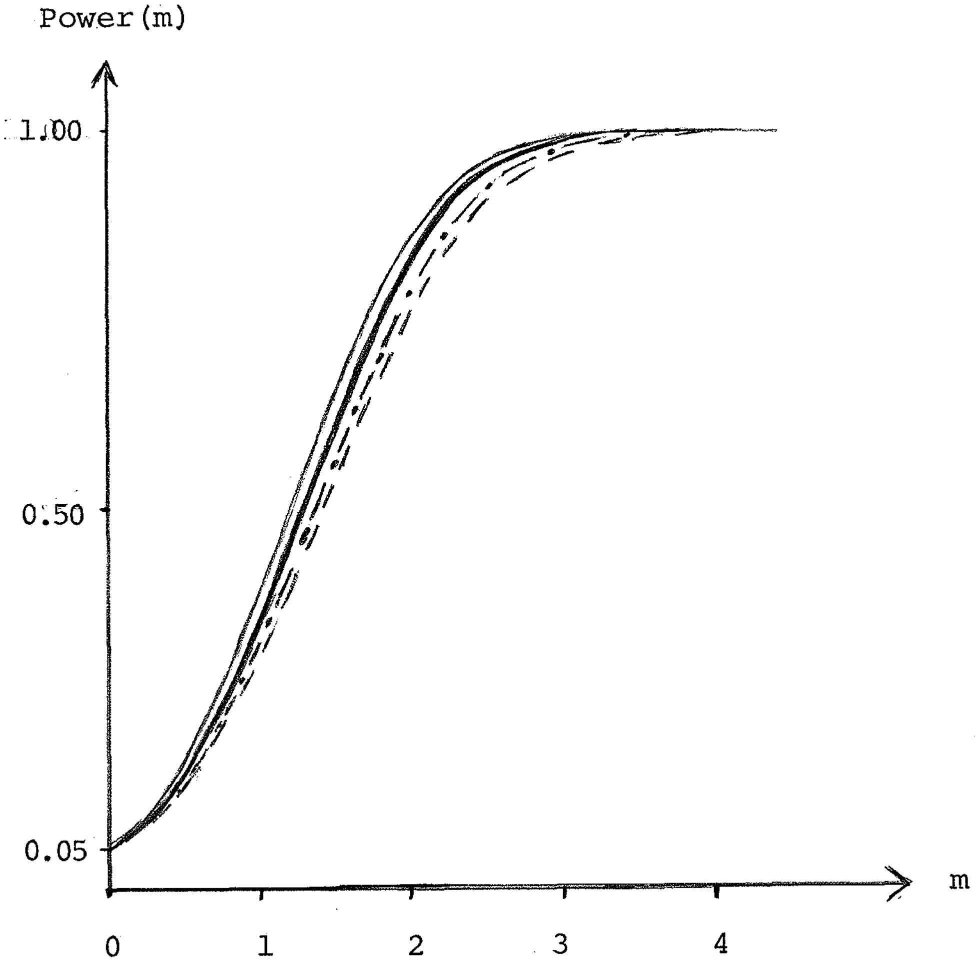

Fig 2

The power graphs when results from two tests are combined. Both hypotheses are equally false with the non-centrality parameters m.

Fisher's method ———— Fisher-adaptive ━━━━

Tippett's mthod — — — — Tippett-adaptive —·—·—·—

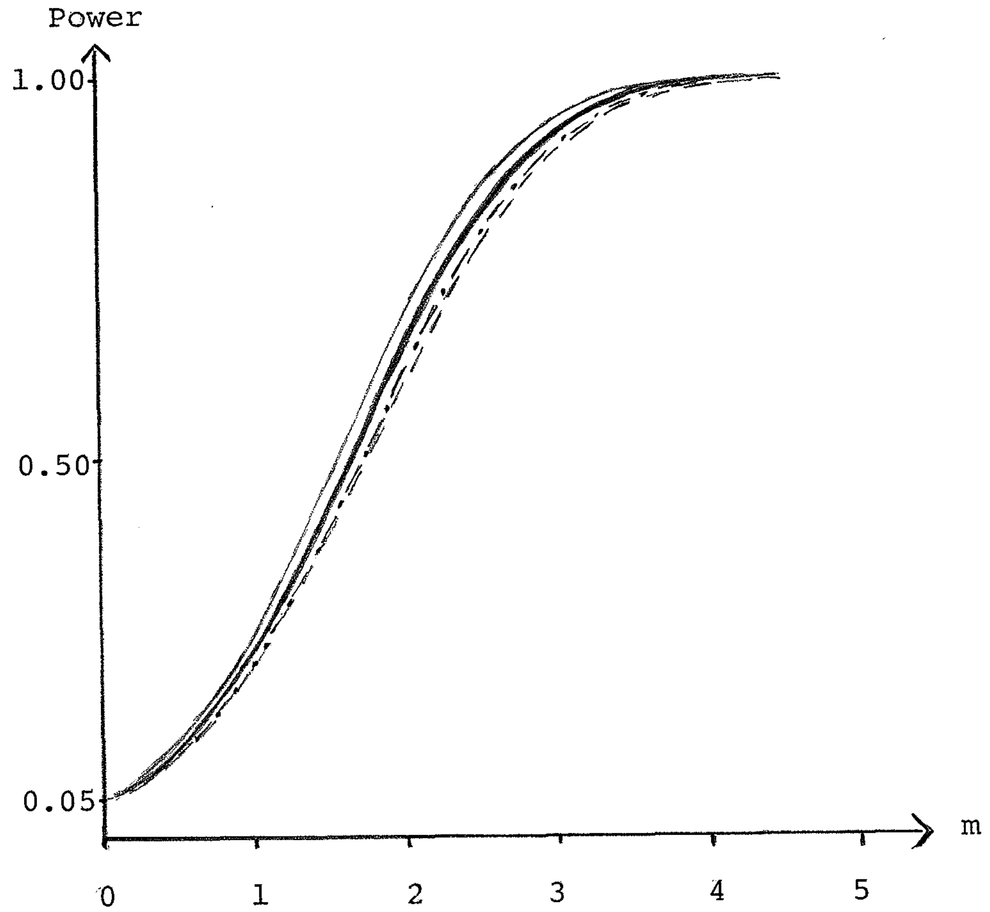

Fig 3

The power graphs when results from two tests are combined. Both hypotheses are false with the non-centrality parameters $m_1=m$ and $m_2=0.5$ m.

Fisher's method ——————   Fisher-adaptive ——————

Tippett's method — — — —   Tippett-adaptive —·—·—·—

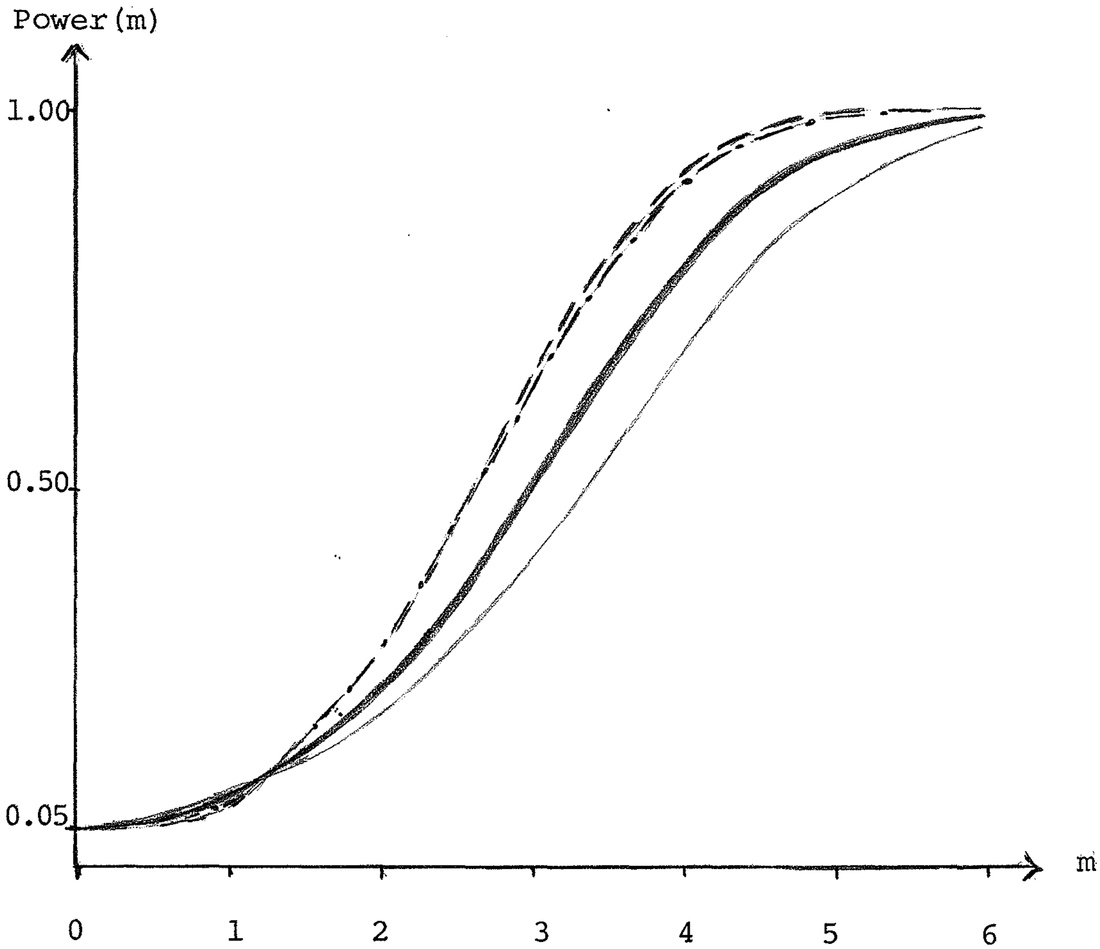

Fig 4

The power graphs when results from fifteen tests are combined. One of the hypotheses is false with the non-centrality parameters m.

Fisher's method ——————— Fisher-adaptive

Tippett's method — — — — Tippett-adaptive

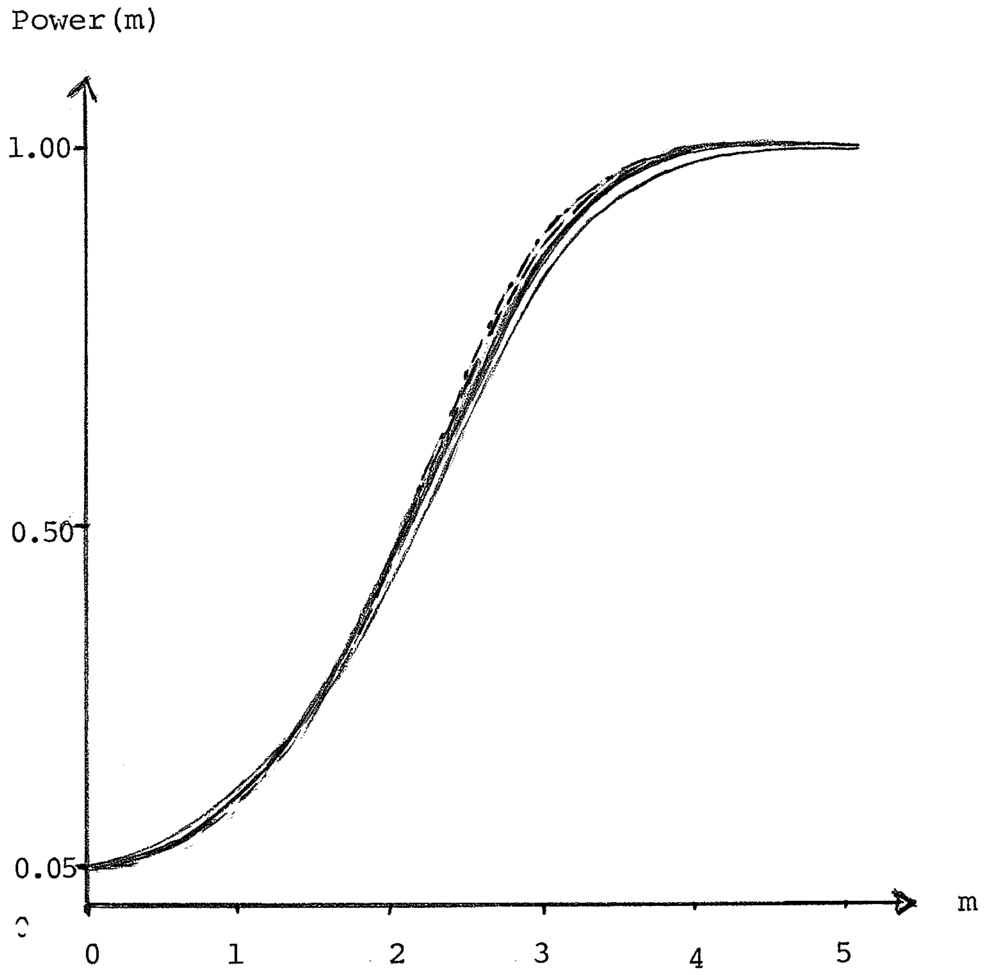

Fig 5

The power graphs when results from fifteen tests are combined. Two of the hypotheses are equally false with the noncentrality parameters m.

Fisher's method ——————   Fisher-adaptive ━━━━━━━━

Tippett-s method — — — —   Tippett-adaptive — · — · — · —

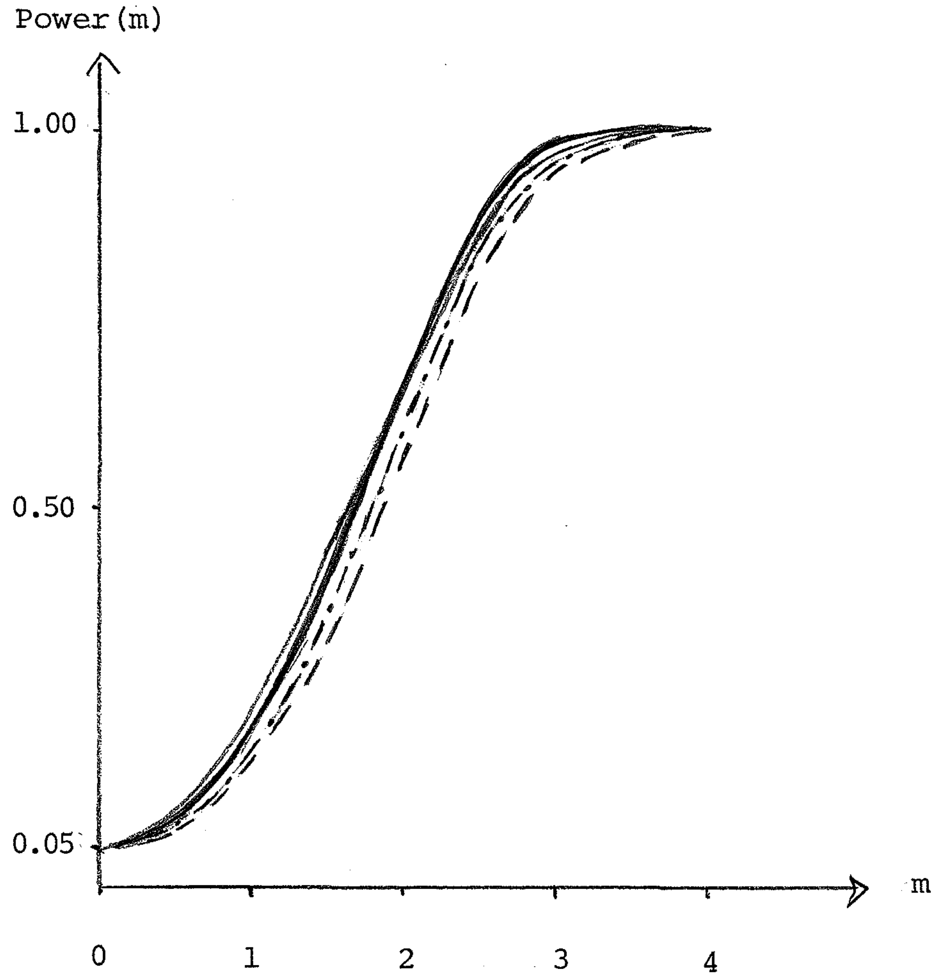

Fig 6

The power graphs when results from fifteen tests are combined. Three of the hypotheses are equally false with the non-centrality parameters m.

Fisher's method ———— Fisher-adaptive ━━━━

Tippett's method — — — — Tippett-adaptive —·—·—·—

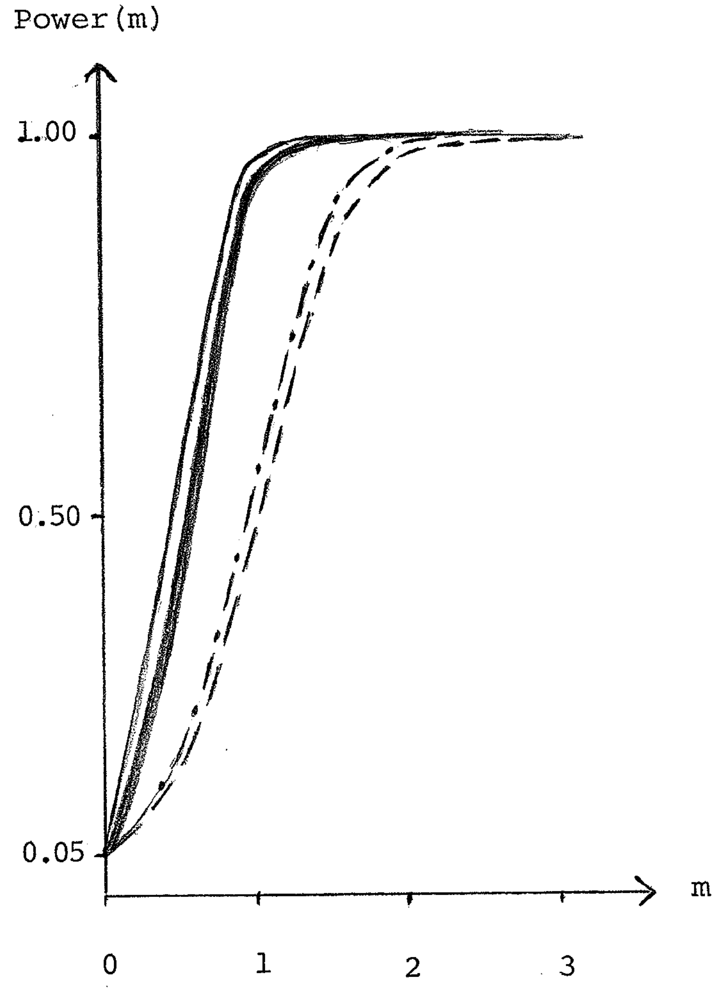

Fig 7

The power graphs when results from fifteen tests are combined. All hypotheses are equally false with the noncentrality parameters m.

Fisher's method ―――――― Fisher-adaptive ━━━━━━

Tippett's method — — — — Tippett-adaptive —·—·—·—

GRÖNA SERIEN RESEARCH REPORT

1975:1	Högberg, Per	Estimation of parameters in models for traffic prediction - A new approach
1975:2	Frisén, Marianne	The use of conditional inference in the analysis of a correlated contingency table
1975:3	Högberg, Per	Planning of traffic counts
1975:4	Jonsson, Robert	A branching poisson process
1975:5	Wold, Herman	Modelling in complex situations with soft information
1975:6	Areskoug, B., Lyttkens, E and Wold, H.	Six models with two blocks of observables as indicators for one or two latent variables
1976:1	Blomqvist, Nils och Svärdsudd, Kurt	Om sambandet mellan blodtryckets tillväxthastighet och nivå.
1976:2	Blomqvist, Nils	On the relation between change and initial value
1976:3	Wold, Herman	On the transition from pattern cognition to model building
1976:4	Blomqvist, Nils	Skattning av imprecision vid samtidig jämförelse av flera mätmetoder
1977:1	Klevmarken, N. A.	A comparative study of complete systems of demand functions
1977:2	Eriksson, Bo	An approximation of the variance of counts for a stationary stochastic point process

1978:1	Eriksson, Bo	Approximation of the variance for the estimated mean in a stationary stochastic process
1979:1	Klevmarken, Anders	Utjämning av lönekurvor
1979:2	Klevmarken, Anders	On the complete systems approach to demand analysis
1979:3	Jonsson, Robert	A branching poisson process model for the occurrence of miniature endplate potentials
1980:1	Flood, L. och Klevmarken, A.	Prognosmodeller för fördelning av den totala privata konsumtionen på 65 varugrupper
1980:2	Creedy, J., Hart, P.E., Jonsson, A and Klevmarken, A.	The distribution of cohort incomes in Sweden 1960-1973
1980:3	Klevmarken, A.	Age, qualification and promotion supplements. A study of salary formation for salaried employees in Swedish Industry
1980:4	Jonsson, A.	A general linear model approach for separating age, cohort and time effects
1980:5	Westberg, Margareta	Kombination av oberoende statistiska test
1981:1	Arvidsen, Nils och Johnson, T.	Variance reduction through negative correlation, a simulation study
1981:2	Eriksson, Sven	Kommunurval, väljarurval och analysansatser
1981:3	Westberg, Margareta	The combination of independent statistical test. A comparison between two combination methods when the test statistics either are normally or chi-square distributed.
1981:4	Frisén, Marianne	Evaluation of a stochastic model for visual capacity by two observational studies.

1981:5	Arvidsen, N. och Johnsson, T.	Sampling. An interactive computer program for survey sampling estimation.
1982:1	Klevmarken, A.	Age, Period and Cohort analysis: A survey.
1982:2	Johnsson, T.	Household market and non-market activities - design issues for a pilot study.
1982:3	Klevmarken, N.A.	Household market an non-market activities.
1982:4	Klevmarken, N.A.	Pooling incomplete data sets.
1983:1	Flood, L.	Time allocation to market and non-market activities in Swedish households.
1983:2	Eriksson, S.	Analys av kategoriska data. En metodstudie i anslutning till statsvetenskaplig forskning.
1983:3	Klevmarken, N.A.	Asymptotic properties of a least-squares estimator using incomplete data.
1984:1	Klevmarken, N.A.	Econometric inference from survey data.
1985:2	Guilbaud, Olivier	Stochastic order relations for one-sample statistics of the Kolmogorov-Smirnov type.
1985:3	Frisén, M.	Unimodal regression.
1985:4	Frisén, M. och Holm, S.	Nonparametric regression with simple curve characteristics.
1985:5	Jonsson, R.	Methods for discriminating betwwen children with the fetal alcohol syndrome and control children on the basis of measurements of ocular fundi. - Some procedures for explorative analysis, tests and individual discrimination.
1985:6	Westberg, M.	An adaptive method of combining independent statistical tests.
1986:1	Johnsson, T.	Multiple comparison tests based on the bootstrap.
1986:2	Dahlbom, U. Holm. S.	Parametric and nonparametric tests for bioequivalence trials.

1986:3	Westberg, M.	A Tippett-adaptive method of combining independent statistical test.