COMMUN. STATIST.—THEORY METH., 22(5), 1257-1297 (1993)

On Optimal Tests for General Interval-Hypotheses

Günther H. Mehring

CIBA-GEIGY Ltd, Pharmaceuticals Division, Medical Department
CH-4002 Basle, Switzerland

KEY WORDS : testing for equivalence; testing for non-zero difference (or for non-unit); α-adjustment; two one-sided tests; confidence interval; interval hypotheses; uniform most powerful (unbiased) test; T-test; χ^2-test; F-test; distributions of Pólya Type; exponential family; noncentral F-distribution; noncentral t-distribution; 'stretched' χ^2-distribution; 'stretched' F-distribution

ABSTRACT

Optimal statistical tests, using the normality assumptions for general interval hypotheses including equivalence testing and testing for non-zero difference (or for non-unit) are presented. These tests are based on the decision theory for Pólya Type distributions and are compared with usual confidence tests and with 'two one-sided tests'- procedures. A formal relationship between some optimal tests and the Anderson and Hauck procedure as well as a procedure recommended by Patel and Gupta is given. A new procedure for a generalisation of Student's test as well as for equivalence testing for the t-statistics is shown.

INTRODUCTION

The interpretation of a significant result of a usual test for significance suffers from the problem of saying nothing about the difference between the true value and the assumed value of the parameter of interest.

Let us consider the following hypotheses : **H** : $\gamma = \gamma_0$ vs **K** : $\gamma \neq \gamma_0$. If a test for significance rejects the hypothesis **H** then we still have no information about the distance between γ and γ_0, which would be important for practical applications. Such a situation arises, when the available data are extensive, as the test becomes very powerful and can detect even small differences. The test then shows a significant result that may be less important with respect to the problem in question. The usual way to overcome the difficulty is to plan the experiment properly based on sample size calculations. Let us regard the very common situation where the variables are normally distributed. Sample size calculations require the specifications of the type I and type II errors, a 'difference to detect' for the means and the standard deviation. The latter is crucial because the standard deviation is usually not well known. An overestimate of this parameter leads to an overestimate of the sample size so that the trial has more power than expected and can lead to the situation described above. Another point has to be raised. Even if the planning was correct, one cannot conclude from a significant outcome that the difference between the means coincides with the assumed 'difference to detect', as a significant result also can be caused by (true) differences of means less than those the experimenter wished to demonstrate. The probability of this event will be less than the power the sample size calculations were based on, but greater than the type I error. Therefore, the 'difference to detect' should not only be included in the planning but also in the test procedure. The issue can be addressed by the following type of hypotheses :

D) **H** : $\gamma_1 \leq \gamma \leq \gamma_2$ vs **K** : $\gamma < \gamma_1$ or $\gamma > \gamma_2$,

where γ_1 and γ_2 are real values (of practical interest) with $\gamma_1 \leq \gamma_2$.

If the hypothesis **H** is rejected by a suitable test then we have precise information about the parameter γ (i.e., $\gamma < \gamma_1$ or $\gamma > \gamma_2$). Note that when $\gamma_1 = \gamma_2$ we have the usual two-sided hypotheses. The problem is usually solved by using the method of confidence estimation: If (c_l, c_u) denotes a $1\text{-}2\cdot\alpha$ confidence interval, $0 < \alpha < 1$, for the parameter γ then **H** will be rejected if (c_l, c_u) lies outside the interval $[\gamma_1, \gamma_2]$.

The opposite formulation, i.e., how to show that a parameter (or differences between parameters) lies inside a given range, is well known as the problem of equivalence testing. The structure of the hypotheses for an equivalence test is obviously related to the above hypotheses. We only have to change the meaning of **H** and **K** :

E) **H** : $\gamma \le \gamma_1$ or $\gamma \ge \gamma_2$ vs **K** : $\gamma_1 < \gamma < \gamma_2$.

One would expect that the test for equivalence is analogous to the test for a difference. This is clear for the confidence interval method : equivalence is accepted, i.e., **H** is rejected, when the interval (c_l, c_u) lies inside (γ_1, γ_2).

There may be situations where it is not clear whether one should test for equivalence or for a difference. As an example we regard the development of a new drug in medicine. The problem is to decide at a very early stage whether the new drug is different or equivalent to a known drug with respect to some parameters of interest. The case of equivalence could lead to a 'no go decision', i.e. there is no interest in further developing the new drug, whereas the demonstration of difference in the framework of **D)** would give a good argument for supporting a 'go decision'. Therefore we have to test the hypotheses **E)** and **D)** simulaneously so that the problem of adjustment of the a-error appears. Another way to solve the problem is a test for the combination of **E)** and **D)** as follows :

E+D) **H** : $\gamma_1 \le \gamma \le \gamma_2$ or $\gamma_3 \le \gamma \le \gamma_4$ vs **K** : $\gamma < \gamma_1$ or $\gamma > \gamma_4$ or $\gamma_2 < \gamma < \gamma_3$,

where $\gamma_1, \ldots, \gamma_4$ are given values satisfying the order $\gamma_1 \le \ldots \le \gamma_4$.

The intervals $\gamma_1 \le \gamma \le \gamma_2$ and $\gamma_3 \le \gamma \le \gamma_4$ of the hypothesis **H** can be regarded as a region of 'indifference', as the data support neither equivalence nor difference. The question what to do is open again. A further development of the new drug may then depend on other properties.

The problem of equivalence **E)** has been discussed extensively since it was introduced to applied statistics by Westlake (1972) (see O'Quigley and Baudoin (1988) for an overview and Anderson and Hauck (1983) for the rationale with respect to medical situations). The question **D)** is much older and was already discussed by Berkson (1938) and by Hodges and Lehmann (1954). All these problems are formally related, and it seems resonable that solutions can be derived within a general framework. This is the objective of the paper : optimal statistical tests are presented for the problems **D)**, **E)** and **E+D)** based on the theory of Pólya Type distributions for the common one- and two- sample problems concerning the normal distribution for the sample(s). The decision theory for Pólya Type distributions was first established by Karlin (1956) in order to answer some theoretical questions without regard for any applications. Therefore, it may be interesting to see how it can be used for problems that have arisen in the biological field and seem to be far removed from original theory. The properties of these optimal tests will be discussed and compared with the usual solutions, i.e., confidence tests and 'two one-sided tests'-procedures. Previous approaches to the problem **E)** for the mean of two normal distributions with equal but unknowns variances, i.e., the Anderson and Hauck procedure (1983) and the Gupta and Patel procedure (1984) are formally related to some optimal tests. This relationship is helpful in order to demonstrate easily the properties of these tests. As a result we will establish better procedures for equivalence testing and a new procedure for testing a non-zero difference between the means of two normal distributions, i.e., a generalisation of Student's test. Finally, a systematical overview of the field based on decision-theoretical arguments is attempted, although our main attention will be focused on **D)** and **E)** as these cover most applications. The problem of designing experiments for applying these tests meaningfully, is beyond the scope of this paper. Readers interested in this topic are referred to a paper published by S. Senn (1991).

I. MATHEMATICAL BACKGROUND

In this chapter we present an overview of some important aspects of the decision theory for Pólya Type distributions as the theory seems to be not well known, but without any proofs. Readers interested in further details are refered to Karlin (1957).

I.1) DEFINITION (Pólya Type distributions) :

Let $(P_\gamma)_{\gamma\in\Gamma}$, where $\Gamma\subseteq\mathbb{R}$, be a family of distributions and λ an σ-finite, dominate measure. The distributions P_γ may have the following λ-densities : $P_\gamma = p(\cdot,\gamma)d\lambda$ with continuous function $p : \mathbb{R}\times\Gamma \to \mathbb{R}$, $\gamma\in\Gamma$. We say, that $(P_\gamma)_{\gamma\in\Gamma}$ <u>belongs to Pólya Type n</u> , $n\in\mathbb{N}$, if

$$\begin{vmatrix} p(x_1,\gamma_1) & \dots & p(x_1,\gamma_m) \\ \vdots & & \vdots \\ p(x_m,\gamma_1) & \dots & p(x_m,\gamma_m) \end{vmatrix} > 0$$

for all $m \le n$ and for given numbers $x_1 <\ldots< x_m$ where $x_i\in\mathbb{R}$, $\gamma_1 <\ldots< \gamma_m$ and $\gamma_i\in\Gamma$. $(P_\gamma)_{\gamma\in\Gamma}$ <u>belongs to Pólya Type ∞</u> , if $(P_\gamma)_{\gamma\in\Gamma}$ is Pólya Type n for every $n\in\mathbb{N}$.

One can show that distributions belonging to the one-parameter exponential family (for example: the normal, the χ^2, the binomial, the Poisson distribution), the non-central t- and the non-central F-distribution all belong to Pólya Type ∞. The Cauchy distribution is not Pólya Type. Therefore the most important distributions occuring in statistical practice can be regarded as belonging to the Pólya Type.

I.2) The theory deals with the following 'two action' decision problem: Let $I_1,\ldots,I_m$ be some closed intervals (i.e., proper or unbounded intervals or single points) with $I_i \cap I_j = \emptyset$ for $i\neq j$ and $\bigcup_{i=1}^{m} I_i \neq \Gamma$.

We define the decision problem for <u>general interval hypotheses of type n</u> as follows:

$$\mathbf{H}^{(n)}: \gamma \in \bigcup_{i=1}^{m} I_i \quad \text{vs} \quad \mathbf{K}^{(n)}: \gamma \notin \bigcup_{i=1}^{m} I_i \ .$$

Here, n represents the number of boundary points of $\bigcup_{i=1}^{m} I_i$, where an interval consisting of a single point is regarded as having two boundary points. There is $n \le 2 \cdot m$.
The decision problems **D)** or **E)** correspond to a decision problem for hypotheses of type 2 and the problem **E+D)** to type 4.

It is clear that a test for the above hypotheses structure is not as simple as a test for the usual hypotheses $\mathbf{H} : \gamma = \gamma_0$ vs $\mathbf{K} : \gamma \ne \gamma_0$. Therefore, the following definition is helpful:

I.3) DEFINITION (monotone procedure):
A (randomized) decision function $\phi : \mathbb{R} \to [0,1]$ is called a monotone procedure, if ϕ is of the following form:

$$\phi(t) := \begin{cases} 1, & \text{for } c_{2\cdot i} < t < c_{2\cdot i+1}, \quad i = 0,1,\ldots,[\frac{n}{2}] \\ \tau_j, & \text{for } t = c_j,\ 0 \le \tau_j \le 1, \quad j = 1,2,\ldots,n \\ 0, & \text{elsewhere} \end{cases}$$

$[a]$ denotes the greatest integer $\le$ a and $c_0 = -\infty$.
$\phi(t)$ represents the probability of accepting the alternative hypothesis, and t represents the outcome of a (sufficient) statistic. All decision functions of this form will be summarized by the class $\mathfrak{M}_n$, if only two successive numbers c_i and c_{i+1} coincide.
(The term 'monotone procedure' has been taken from Karlin's paper.)

I.4) The link between a decision function and a test φ is given by a statistic $T : (\mathfrak{X},\mathbb{B},(\mathbb{P}_\omega)_{\omega\in\Omega}) \to \mathbb{R}$, $\varphi(X) := \phi(T(X))$, where $\mathfrak{X}$ denotes the sample space, $\mathbb{B}$ a σ-algebra on $\mathfrak{X}$ and $(\mathbb{P}_\omega)_{\omega\in\Omega}$ a family of distributions describing the data from the experiment. $X \in \mathfrak{X}$ denotes a random vector, and T is regarded as sufficient or invariant. It is assumed that for every $\omega\in\Omega$ there exists a parameter $\gamma\in\Gamma$ with $\mathbb{P}_\omega^T = P_\gamma$. Therefore, the properties of a test depend only on the decision function. Usually, we do not distinguish between a test and the corresponding desicion function. The relationship between Pólya Type distributions and monotone procedures is given by the following theorem :

I.5) THEOREM :

i) Let $H^{(n)}$ vs $K^{(n)}$ be a given decision problem with a number $0 < \alpha < 1$ and let us assume that $(P_\gamma)_{\gamma\in\Gamma}$ belongs to Pólya Type n+1. Then for any randomized decision function ϕ not in $\mathfrak{M}_n$ exists a unique ϕ^* in $\mathfrak{M}_n$ such that $R(\gamma,\phi^*) \le R(\gamma,\phi)$ with inequality everywhere except for $\gamma = \gamma_i$ with $i=1,\dots,n$. R stands for the risk function and is defined with respect to Neyman-Pearson loss functions:

$$R(\gamma,\phi) = \begin{cases} \int_{\mathbb{R}} \phi(t)\cdot p(t,\gamma)\, d\lambda(t) \ , & \text{if } \gamma \in H^{(n)} \\ \int_{\mathbb{R}} (1-\phi(t))\cdot p(t,\gamma)\, d\lambda(t) \ , & \text{if } \gamma \in K^{(n)} \end{cases}$$

ii) $\phi_\alpha(t) := \alpha$, $t\in \mathbb{R}$, is a randomized decision function for every hypotheses **H)** *vs* **K)**, hence a test $\varphi(X) = \phi^*(T(X))$, $\phi^* \in \mathfrak{M}_n$ is unbiased.

iii) A monotone procedure $\phi^* \in \mathfrak{M}_n$ is (λ-a.e.) uniquely determined by the numbers c_i^*, τ_i^* satisfying the conditions :

$$\int_{\mathbb{R}} \phi^*(t)\cdot p(t,\gamma_i)\, d\lambda(t) + \sum_{j=1}^{n} p(c_j^*,\gamma_i)\cdot\tau_j^* = \alpha, \text{ for } i=1,\dots,n$$

and (in addition) if $\gamma_i = \gamma_{i+1}$,

$$\frac{\partial}{\partial\gamma} \left(\int_{\mathbb{R}} \phi^*(t)\cdot p(t,\gamma)\, d\lambda(t) + \sum_{j=1}^{n} p(c_j^*,\gamma)\cdot\tau_j^* \right) \Bigg|_{\gamma=\gamma_i} = 0$$

(Note, that only two successive numbers γ_i, γ_{i+1} can coincide!)

Some remarks are necessary concerning the theorem.

The condition i) allows us to consider decision functions belonging to the class $\mathfrak{M}_n$ as they reduce uniformly the type I and type II errors. These functions are determined by iii). The trivial decision function $\phi_\alpha(t) = \alpha$, $t\in\mathbb{R}$ is improved in terms of risk by every $\phi \in \mathfrak{M}_n$, therefore all tests based on the decision function belonging to $\mathfrak{M}_n$ will be unbiased (ii). The type I error will only be reduced, if H contains a proper interval. The maximum of this error is reached at the boundary, i.e., the tests are 'similar on the boundary'.

It is worth mentioning that results for the formulations **D)**, **E)** were already given by Lehmann but for the one-parameter exponential family only. Usually, this does not lead to any problems as the common two-sample statistical tests in a multiparameter-exponential family can be reduced to a one-sample problem by conditioning (see Lehmann (1986) for a recent presentation). So, the theory of Pólya Type distributions can be regarded as a generalisation of Lehmann's theorem with respect to the geometrical structure of the hypotheses and the class of distributions. Now, we are in a position to be able to apply the theory by addressing the above introduced problems **D)** and **E)** to hypotheses of the form $\mathbf{H}^{(2)}$ vs $\mathbf{K}^{(2)}$ and **E+D)** to $\mathbf{H}^{(4)}$ vs $\mathbf{K}^{(4)}$.

II. MONOTONE PROCEDURES FOR THE MEAN OF THE NORMAL DISTRIBUTION

Optimal decision functions for the normal distribution are an excellent example for demonstrating the above discussed ideas as computations are very easy, although in practice tests are usually based on the t-distribution (i.e variance(s) unknown) rather than the normal distribution (i.e variance(s) known). However, normal tests make good approximations. In what follows we only need to regard normal distributions $N(\gamma,\sigma)$ with a standard deviation $\sigma=1$, as the test statistics to be used later on are standardised with respect to σ. The family $(N(\gamma,1))_{\gamma\in\mathbb{R}}$ of normal distributions belong to the exponential family, hence it is of Pólya Type ∞. The application of theorem I.5,iii) leads to the following equations for computing the critical values $c_i^* = c_i^*(\gamma_1,\gamma_2)$ (or $c_i^* = c_i^*(\gamma_1,\gamma_2,\gamma_3,\gamma_4)$) (note that the number τ_i^* can be chosen to be 0 because λ is the Lebesque measure here) :

D) for $\gamma_1 < \gamma_2$:

$$\int_{-\infty}^{c_1^*} g(t-\gamma_1)dt + \int_{c_2^*}^{\infty} g(t-\gamma_1)dt = \int_{-\infty}^{c_1^*} g(t-\gamma_2)dt + \int_{c_2^*}^{\infty} g(t-\gamma_2)dt = \alpha \qquad \textbf{(2.1)}$$

$$\text{E) for } \gamma_1 < \gamma_2 : \quad \int_{c_1^*}^{c_2^*} g(t-\gamma_1)dt = \int_{c_1^*}^{c_2^*} g(t-\gamma_2)dt = \alpha \tag{2.2}$$

E+D) for $\gamma_1 < \gamma_2 < \gamma_3 < \gamma_4$:

$$\int_{-\infty}^{c_1^*} g(t-\gamma_i)dt + \int_{c_4^*}^{\infty} g(t-\gamma_i)dt + \int_{c_2^*}^{c_3^*} g(t-\gamma_i)dt = \alpha, \ i=1,\dots,4 \tag{2.3}$$

Here, $g(t-\gamma)$ denotes the density $(2\cdot\pi)^{-1/2}\cdot\exp(-\frac{1}{2}\cdot(t-\gamma)^2)$, $t,\gamma \in \mathbb{R}$, of the normal distribution. The case $\gamma_1 = \gamma_2$ ($\gamma_1 = \gamma_2$ and/or $\gamma_3 = \gamma_4$) can be handled by applying the second equation of theorem I.5,iii) or by simply allowing $\gamma_1 \to \gamma_2$ ($\gamma_1 \to \gamma_2$ and/or $\gamma_3 \to \gamma_4$). The figures 1 and 2 show graphically the relationship between γ_i and the critical values c_i^* for equations (2.1) and (2.2).

With the help of I.4) we present the t-test with known variance(s).

We make the usual assumptions : Let $X_1,\dots,X_n$, $Y_1,\dots,Y_m$ be stochastically independent random variables, which are distributed $N(\mu_1,\sigma_1^2)$, $N(\mu_2,\sigma_2^2)$, with known variances $\sigma_1^2 > 0$ and $\sigma_2^2 > 0$, respectively. Differences between the means are of interest, therefore the decision problems are of the following form :

$$\begin{aligned}
&\text{D) :} && H : \mu \in [\delta_1,\delta_2] \quad \text{vs} \quad K : \mu \notin [\delta_1,\delta_2] \\
&\text{E) :} && H : \mu \notin (\delta_1,\delta_2) \quad \text{vs} \quad K : \mu \in (\delta_1,\delta_2) \\
&\text{E+D) :} && H : \mu \in [\delta_1,\delta_2] \cup [\delta_3,\delta_4] \quad \text{vs} \quad K : \mu \notin [\delta_1,\delta_2] \cup [\delta_3,\delta_4] ,
\end{aligned} \tag{2.4}$$

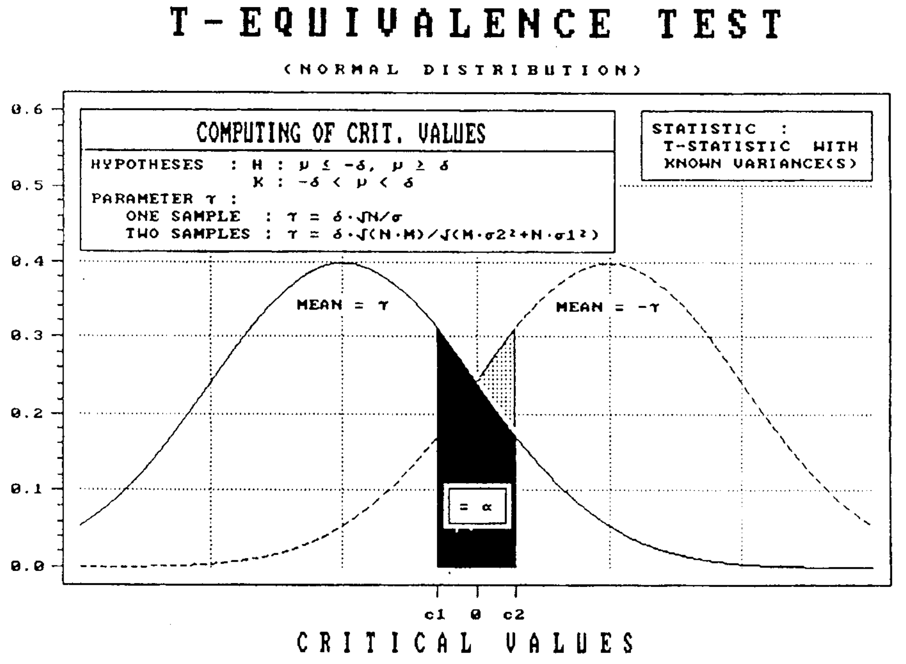
with $\mu = \mu_1$ for the one sample, $\mu=\mu_1-\mu_2$ for the two-sample problem and $\delta_1 \le \delta_2$ (or $\delta_1 \le\dots\le \delta_4$).

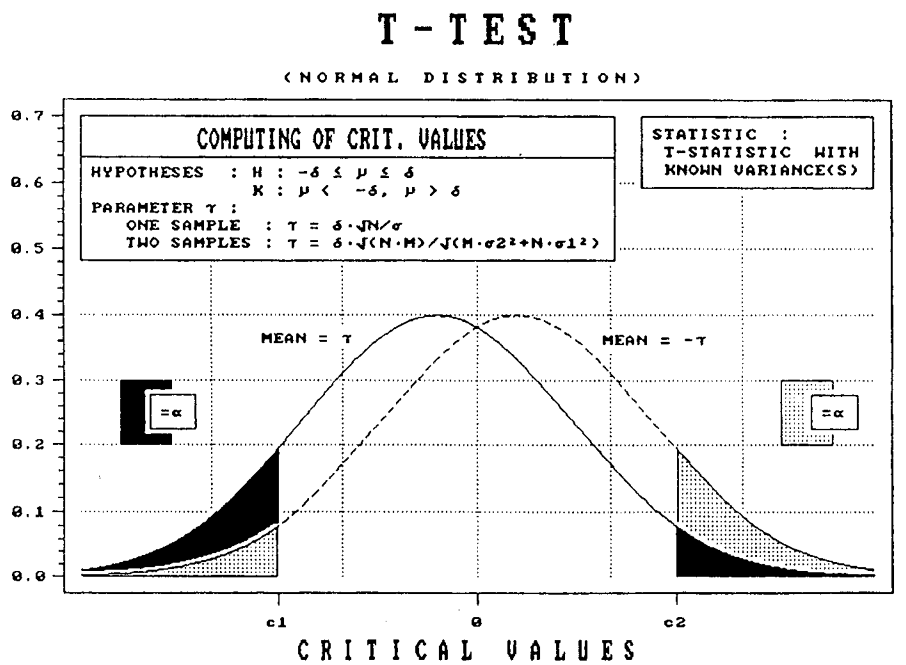
Hence, the parameters γ_i for computing the critical values are

$$\gamma_i = \frac{\delta_i}{\sigma_1}\cdot\sqrt{n} \quad \text{or} \quad \gamma_i = \frac{\delta_i}{\sqrt{n\cdot\sigma_2^2 + m\cdot\sigma_1^2}}\cdot\sqrt{n\cdot m} \quad , \tag{2.5}$$

The statistic $T=T(X)$ or $T=T(X,Y)$ is given by

$$T(X) = \frac{\bar{X}}{\sigma_1}\cdot\sqrt{n} \quad \text{or} \quad T(X,Y) = \frac{\bar{X} - \bar{Y}}{\sqrt{n\cdot\sigma_2^2 + m\cdot\sigma_1^2}}\cdot\sqrt{n\cdot m}$$





FIGURES 1 AND 2

Now, we are able to establish the test procedure :

$$\left.\begin{array}{ll}
1.) & \text{Define numbers } \delta_1,\ \delta_2 \ (\text{or } \delta_1,\dots,\delta_4) \text{ on the basis} \\
 & \text{of practical considerations} \\
2.) & \text{Calculate the parameters } \gamma_1,\ \gamma_2 \ (\text{or } \gamma_1,\dots,\gamma_4). \\
3.) & \text{Calculate the critical values } c_1^*,\ c_2^* \ (\text{or } c_1^*,\dots,c_4^*). \\
4.) & \text{Calculate the t-value.} \\
5.) & \text{Accept K (reject H), if T lies inside the critical} \\
 & \text{region i.e. } \mathbf{D)}:\ T<c_1^* \text{ or } T>c_2^*,\quad \mathbf{E)}:\ c_1^*<T<c_2^*, \\
 & \qquad \mathbf{E+D)}:\ T<c_1^* \text{ or } T>c_4^* \text{ or } c_2^*<T<c_3^*.
\end{array}\right\} \quad (2.6)$$

It is interesting to compare the above solution with standard approaches.

II.1) Firstly, we focus on the equivalence test problem; the others can be handled in a very similar manner. There are two procedures in use: the confidence test ϕ_E^{conf}, which was aready outlined and a 'two one-sided tests'-procedure $\phi_E^{(2)}$ (see Hodges and Lehmann (1954)). This means we have to test each of the following two one-sided hypotheses at the level of α :

$$H_1 : \gamma \le \gamma_1 \text{ vs } K_1 : \gamma > \gamma_1 \quad \text{and} \quad H_2 : \gamma \ge \gamma_2 \text{ vs } K_2 : \gamma < \gamma_2$$

If K_1 and K_2 are accepted then equivalence will be accepted (i.e., $\gamma \in [\gamma_1,\gamma_2]$). Hence, the test $\phi_E^{(2)}$ has the form :

$\phi_E^{(2)}(t) = 1_{(u_{1-\alpha}+\gamma_1,u_\alpha+\gamma_2)}(t)$, $t\in\mathbb{R}$, where γ_i are the numbers of (2.5)

and u_α denotes the α-fractile of Ψ, where Ψ stands for the probability function of the normal distribution $N(0,1)$. (By 1_S we denote the characteristic function of the set S, i.e., $1_S(t) = 1$ if $t\in S$ and $1_S(t) = 0$ if $t\notin S$.) The size is always lower than the nominal level because the critical region of $\phi_E^{(2)}$ is the intersection of the critical regions of the one-sided tests. (By the size of a test we mean as usual the upper bound of its probability of first kind error.). If the confidence test ϕ_E^{conf} is based on a $1\text{-}2\cdot\alpha$ confidence interval then $\phi_E^{(2)} = \phi_E^{conf}$, as the density of the normal distribution is symmetric. This is why we have to compare the optimal test with the confidence test only.

Without loss of generality we assume $\gamma_1 = -\gamma_2$ where $\gamma_2 \ge 0$, hence $c_1^*(\gamma_2) = -c_2^*(\gamma_2)$, $c_2^*(\gamma_2) > 0$. This can be seen from the following considerations: If $H : \mu \notin (\delta_1, \delta_2)$ vs $K : \mu \in (\delta_1, \delta_2)$ are any hypotheses to test with $\delta_1 \neq -\delta_2$ then we can define $\delta_i' = \delta_i - \bar{\delta}$, $\bar{\delta} = \frac{1}{2}\cdot(\delta_1 + \delta_2)$ and have to test the hypotheses $H' : \mu - \bar{\delta} \notin (\delta_1', \delta_2')$ vs $K' : \mu - \bar{\delta} \in (\delta_1', \delta_2')$ that are symmetric around zero.

From (2.2) it follows : $\alpha = \int_{-\infty}^{u_\alpha} g(t)dt = \int_{c_1^*(\gamma_2)-\gamma_2}^{c_2^*(\gamma_2)-\gamma_2} g(t)dt < \int_{-\infty}^{c_2^*(\gamma_2)-\gamma_2} g(t)dt$,

therefore $c_2^*(\gamma_2) > u_\alpha + \gamma_2$ and $c_1^*(\gamma_2) < u_{1-\alpha} + \gamma_1$. These inequalities demonstrate again the well known 'conservatism' of the confidence test :

$$\phi_E^{conf}(t) = 1_{(u_{1-\alpha}+\gamma_1, u_\alpha+\gamma_2)}(t) < 1_{(c_1^*(\gamma_2), c_2^*(\gamma_2))}(t) = \phi_E^*(t), \ t \in \mathbb{R} \ \text{and}$$

$$\alpha = \int_{c_1^*(\gamma_2)}^{c_2^*(\gamma_2)} g(t-\gamma_2)dt > \int_{u_{1-\alpha}+\gamma_1}^{u_\alpha+\gamma_2} g(t-\gamma_2)dt = \int_{u_{1-\alpha}-2\cdot\gamma_2}^{u_\alpha} g(t)dt = \alpha' \underset{\gamma_2 \to \infty}{\to} \alpha .$$

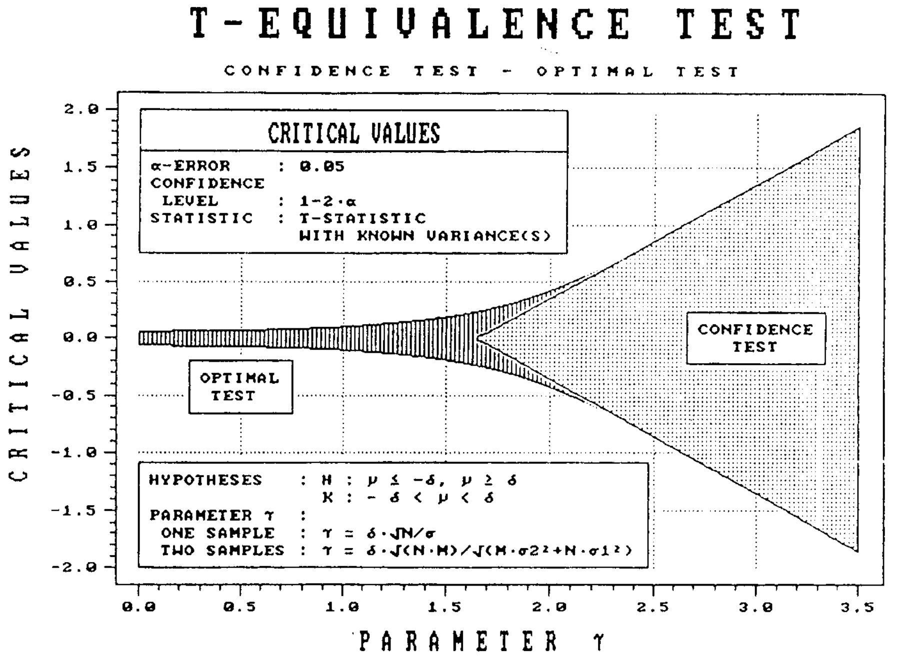
Here we have denoted the decision function for the optimal equivalence test by $\phi_E^*(t)$. The critical values $u_\alpha + \gamma_2$ and $c_2^*(\gamma_2)$ are asymptotically equal. This is a result of the inequality for $0 < \alpha < 1$ (appendix I):

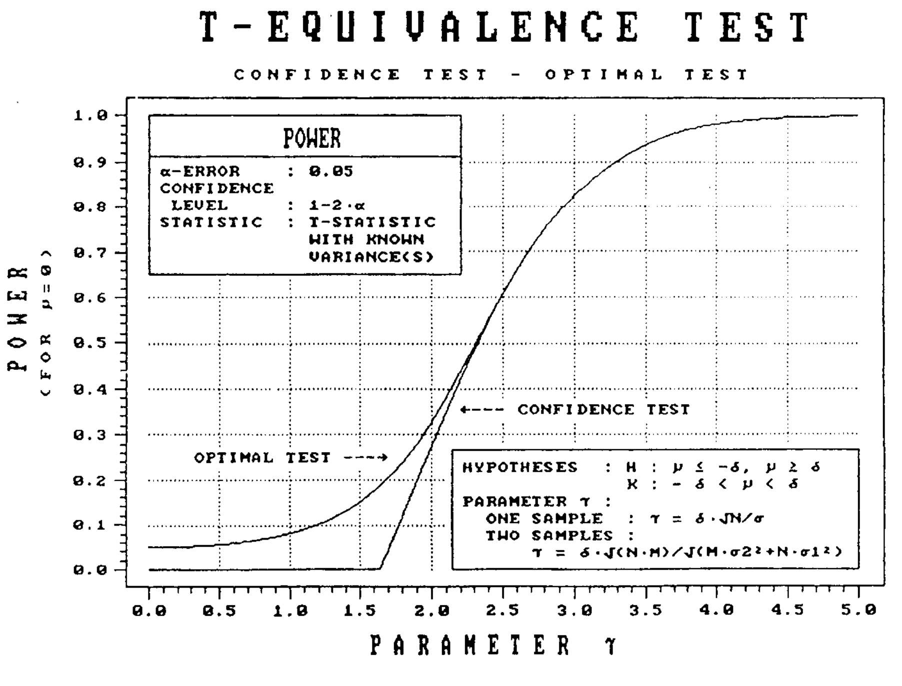
$$c_2^*(\gamma_2) - (u_\alpha + \gamma_2) < \frac{\Psi(-2\cdot\gamma_2 + u_{1-\alpha})}{\min(g(u_\alpha), g(u_{(1+\alpha)/2}))} \quad \text{for } \gamma_2 \geq \min(0, u_{1-\alpha}) \tag{2.7}$$

(It is $\min(g(u_\alpha), g(u_{(1+\alpha)/2})) = g(u_\alpha)$ for the usual choice of α.)

Clearly the confidence test cannot be used for parameters $\gamma_1, \gamma_2 \in \mathbb{R}$ with $\frac{1}{2}\cdot|\gamma_1 - \gamma_2| \leq \min(0, u_{1-\alpha})$, because $\alpha' = 0$. However, the power of the optimal test ϕ_E^* is limited for these parameters by the number $\Psi(c_2^*(\min(0, u_{1-\alpha}))) - \Psi(c_1^*(\min(0, u_{1-\alpha})))$.

A remark should be added about the meaning of these results. Although the theory of Pólya Type distributions improves the confidence test method, one is inclined to regard the improvement as unimportant for practical purposes. The power of the optimal test is very low for those parameters where the confidence test fails. It ranges from 0.05 to $\cong 0.18$ for $\alpha = 0.05$ (see figures 3 and 4). One can see that the confidence test is a very good approximation to the optimal test if the





FIGURES 3 AND 4

power is above a certain level (> 50 %). The optimal test cannot only be applied by computing the critical values $c_1^*(\gamma_2)$ because the symmetry of g allows the following alternative formulation in p-values :

$$c_1^*(\gamma_2) < T < c_2^*(\gamma_2) \iff |T| < c_2^*(\gamma_2) \iff$$
$$\Psi(|T|-\gamma_2)-\Psi(-|T|-\gamma_2) = \int_{-|T|}^{|T|} g(t-\gamma_2)dt < \int_{c_1^*(\gamma_2)}^{c_2^*(\gamma_2)} g(t-\gamma_2)dt = \alpha \tag{2.8}$$

Following from this, however, both tests can easily be applied.

What can one say about the problems **D)** and **E+D)** ?

II.2) An optimal test for the problem **D)** can be obtained from ϕ_E^*, because of the identity $\phi_D^* = 1-\phi_E^*$ (λ-a.e.), where ϕ_E^* is based on a level of $1-\alpha$. The analogous identity holds for the confidence test, which corresponds to the 'two one-sided tests'-procedure and is therefore of the form

$\phi_D^{conf}(t) = 1_{(-\infty,u_\alpha+\gamma_1)\cup(u_{1-\alpha}+\gamma_2,\infty)}(t)$, $t\in\mathbb{R}$, $\gamma_1 = -\gamma_2$, $\gamma_2 \geq 0$.

We learn from the inequality $\phi_D^{conf}(t) > 1_{(-\infty,c_1^*(\gamma_2))\cup(c_2^*(\gamma_2),\infty)}(t) = \phi_D^*(t)$, $t\in\mathbb{R}$, that the size always exceeds the nominal level, but the optimal test can be approximated by the confidence test even when the parameters tend to small values, i.e., $\gamma_2 > 1$. This is the meaning of the following formula, which follows from (2.7) by changing α into $1-\alpha$:

$$c_2^*(\gamma_2) - (u_{1-\alpha}+\gamma_2) < \frac{\Psi(-2\cdot\gamma_2-u_{1-\alpha})}{g(u_{1-\alpha/2})}, \quad \text{for } \gamma_2 > 0 \text{ and } \alpha < \frac{1}{2} \tag{2.9}$$

But the reader should be aware of the fact that the confidence test becomes very 'liberal' in the neigbourhood of $\gamma_2 = 0$. The size of the test reaches the maximum of $2\cdot\alpha$ for $\gamma_2 = 0$. This can be overcome by using $1-\alpha$ instead of $1-2\cdot\alpha$ for the confidence level, but the size then tends to $\alpha/2$ for larger values of γ_2 and testing for difference is only of interest for larger parameters γ_2. The test can also be used in terms of p-values :

Accept **K**, if $1-\Psi(|T|-\gamma_2)+\Psi(-|T|-\gamma_2) < \alpha$.

II.3) Recall that the hypotheses of a test for **E+D**) are of the form

$$H : \gamma_1 \le \gamma \le \gamma_2 \text{ or } \gamma_3 \le \gamma \le \gamma_4 \quad \text{vs} \quad K = K_E \cup K_D$$

with $K_D : \gamma < \gamma_1$ or $\gamma > \gamma_4$ and $K_E : \gamma_2 < \gamma < \gamma_3$ where $\gamma_1,\ldots,\gamma_4$ are real numbers satisfying the order $\gamma_1 \le \ldots \le \gamma_4$ (only two successive numbers γ_i, γ_{i+1} may be equal).

The critical values $c_i^* = c_i^*(\gamma_1,\ldots,\gamma_4)$ are determined by the equation (2.3). The test is of the form $\phi^*_{E+D}(t) = 1_{(-\infty,c_1^*)\cup(c_4^*,\infty)\cup(c_2^*,c_3^*)}(t) = 1_{(-\infty,c_1^*)\cup(c_4^*,\infty)}(t)+1_{(c_2^*,c_3^*)}(t)$, $t\in\mathbb{R}$ and can be regarded as the addition of a test for equivalence and for a difference. In what follows, we consider the case of symmetric hypotheses only, i.e., $\gamma_1 = -\gamma_4$, $\gamma_2 = -\gamma_3$, $\gamma_3 > 0$ and $\gamma_4 = \gamma_3+d$ with $d \ge 0$. Using this procedure, we are not only interested in accepting the global alternative hypothesis K, but we wish to decide whether the data support equivalence or a difference. Therefore another type of error is to consider: the probability of accepting equivalence (i.e., $\gamma_2 < \gamma < \gamma_3$) when there is a true difference (i.e., $\gamma < \gamma_1$ or $\gamma > \gamma_4$) and vice versa.

If the first probability is denoted by $P(K_E|K_D)$ and the second one by $P(K_D|K_E)$ then we obtain the following lower and upper bounds :

$$P(K_E|K_D) \le \int_{c_2^*-\gamma_3-d}^{c_3^*-\gamma_3-d} g(t)dt \le \alpha \text{ and } \left(\int_{-\infty}^{c_1^*} + \int_{c_4^*}^{\infty}\right) g(t)dt \le P(K_D|K_E) < \alpha.$$

The formulae show that these errors depend basically on the distance between the hypotheses K_E and K_D and this fact can be used for adjusting the α-error by applying it twice to the confidence test: once for equivalence and once for difference, respectively. This is possible, as ϕ^*_{E+D} can be approximated by a test $\phi^{conf}_{E+D}(t) = 1_{(-\infty,u_1-\gamma_3-d)\cup(u_2+\gamma_3+d,\infty)}(t)+1_{(u_1-\gamma_3,u_2+\gamma_3)}(t)$, $t\in\mathbb{R}$ for large values of γ_3 (i.e., $\gamma_3 \ge 2.5$), where the numbers $u_1,u_2\in\mathbb{R}$ are the (uniquely determined) solution to the equation (appendix II):

$$\left(\int_{-\infty}^{u_1} + \int_{u_2}^{\infty}\right) g(t)dt = \left(\int_{-\infty}^{u_1} + \int_{u_2}^{\infty}\right) g(t-d)dt = \alpha \quad \text{for } \alpha < \tfrac{1}{2} \tag{2.10}$$

Usually, one is unwilling to apply the confidence test twice by solving the above equation first. Therefore, we compute the size for the confidence test method based on a given confidence level. We let $\gamma_3 > u_{1-\alpha}$ and obtain the following expressions for the size depending on γ_3 or γ_4, respectively :

$$E_{\gamma_2}\phi^{conf}_{E+D} = E_{\gamma_3}\phi^{conf}_{E+D} = \alpha + \left(\int_{u_{1-\alpha}+d}^{\infty} - \int_{-u_{1-\alpha}-2\cdot\gamma_3-d}^{u_{1-\alpha}-2\cdot\gamma_3} \right) g(t)dt \quad \text{and}$$

$$E_{\gamma_1}\phi^{conf}_{E+D} = E_{\gamma_4}\phi^{conf}_{E+D} = \alpha + \left(\int_{u_{1-\alpha}+d}^{\infty} - \int_{-u_{1-\alpha}-2\cdot\gamma_3-2\cdot d}^{u_{1-\alpha}-2\cdot\gamma_3-d} \right) g(t)dt \ .$$

It is $E_{\gamma_3}\phi^{conf}_{E+D} = E_{\gamma_4}\phi^{conf}_{E+D} \to 2\cdot\alpha$ for $d = 0$ as $\gamma_3 \to \infty$. This means that the confidence level has simply to be halved for a hypothesis $H = \{\gamma_1=\gamma_2,\ \gamma_3=\gamma_4\}$, which consists of two single points only. This result can also be obtained by (2.10) for letting $d \to 0$. The left sides of the formulae differ from zero for $d > 0$, but they become negligible for the usual choice of α. The size of the test can be calculated with the help of $E_{\gamma_i}\phi^{conf}_{E+D} \cong \alpha+\Psi(u_\alpha-d)$, $i=1,\dots4$.

III. MONOTONE PROCEDURES FOR THE NON-CENTRAL AND SHIFTED T-DISTRIBUTION

We obtain optimal decision functions for the noncentrality parameter $\gamma\in\mathbb{R}$ of the non-central t-distribution in a completely analogous manner as we did for the mean of the normal distribution; in the formulae (2.1)-(2.3) we change only $g(t-\gamma)$ to the density $p_{k;\gamma}(t)$ of the non-central t-distribution. The usual application is the t-test (with unknown variance(s)) :

The assumptions are the same as in the previous section, but we regard the variance σ or the variances $\sigma_1^2 = \sigma_2^2 =: \sigma^2$ to be unknown. The non-centrality parameter is of the form

$$\gamma = \frac{\mu}{\sigma}\cdot\sqrt{n} \quad \text{or} \quad \gamma = \frac{\mu_1-\mu_2}{\sigma\cdot\sqrt{n+m}}\cdot\sqrt{n\cdot m} \quad \text{and the statistic } \mathbf{T} \text{ is given}$$

by $T = \frac{\bar{X}}{\hat{\sigma}(X)} \cdot \sqrt{n}$ or $T = \frac{(\bar{X}-\bar{Y}) \cdot \sqrt{n \cdot m}}{\hat{\sigma}(X,Y) \cdot \sqrt{n+m}}$, with

$\hat{\sigma}(X,Y) = \frac{\sqrt{\hat{\sigma}(Y)^2 \cdot (n-1) + \hat{\sigma}(X)^2 \cdot (m-1)}}{\sqrt{n+m-2}}$, $\hat{\sigma}(X) = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2}$,

$\hat{\sigma}(Y)$ respectively.

The test procedure is the same as (2.6) but starts with the definition of the values γ_i, $i = 1,2$ (or $i = 1,\ldots,4$), as σ is now considered to be unknown. In the case of a symmetric equivalence interval i.e., $\gamma_1 = -\gamma_2$, $\gamma_2 \geq 0$, the critical values $c_1^*(\gamma_2) = -c_2^*(\gamma_2)$, $c_2^*(\gamma_2) > 0$ can be computed by the non-central F-distribution (see Johnson and Kotz (1970), p. 205) :

$$\int_{-c_2^*(\gamma_2)}^{c_2^*(\gamma_2)} p_{k;\gamma}(t)dt = \int_{0}^{c_2^*(\gamma_2)^2} p_{1,k;\gamma_2^2}(t)dt = \alpha \text{ with and } k=n-1 \text{ or } k=n+m-2. \quad (3.1)$$

Here $p_{1,k;\gamma}(t)$ stands for the density of the non-central F-distribution The result is that we have found an optimal solution for the problems **E)**, **D)** and **E+D)** but for the noncentrality parameter γ and not for the mean. This is consequence of the invariance of the t-statistic, hence the result is not surprising.

Unfortunately, a test for the noncentrality parameter says nothing about the parameters of the distribution of the original data as a ratio of the form $\frac{\mu}{\sigma}$ is considered and not the parameters themselves. This is clear for the one-sample problem.

However, a test concerning the decision problems **E)**, **D)** and **E+D)** for the difference of means can be obtained by approximation, for example by the confidence test method. There are two other approaches described for equivalence testing : a procedure suggested by Anderson and Hauck (1983) and another one recommended by Patel and Gupta (1984). In all procedures one has to estimate the unknown parameters γ_i by estimating the unknown variance σ^2. Although the properties of these procedures have been studied intensively during the last years (e.g. see Frick 1987, 1991 and 1992) it seems to be worth mentioning how they can be obtained from the approach discussed above. The relative merits of this approach is to provide a detailed look into their features from a systematical viewpoint. As a result, we are able to present 'average' tests that improve the Anderson and Hauck procedure and the confidence test as well.

III.1) Anderson and Hauck's procedure deals with the decision problem of the form (2.4,E)), but $\delta_1 = -\delta_2$, $\delta_2 := \delta > 0$. The alternative hypothesis is accepted if $F_k(|T|-\hat{\gamma}_2)-F_k(-|T|-\hat{\gamma}_2) < \alpha$. Here, T is the usual t-statistic, F_k denotes the central t-distribution with k degrees of freedom (k = n-1 or K = n+m-2) and $\hat{\gamma}_2$ is an estimate of the parameters (2.5), i.e.,

$$\gamma_2 = \frac{\delta}{\sigma}\cdot\sqrt{n} \quad \text{or} \quad \gamma_2 = \frac{\delta}{\sigma}\cdot\sqrt{n\cdot m\ /\ (m+n)} \quad , \quad \gamma_1 = -\gamma_2 \tag{3.2}$$

by substituting $\hat{\sigma}$ for σ (i.e., $\hat{\sigma} = \hat{\sigma}(X)$ or $\hat{\sigma} = \hat{\sigma}(X,Y)$). Let $p_k(t) := p_{k;0}(t)$, $t\in\mathbb{R}$ be the desity of the t-distribution and $c_{AH}(\hat{\gamma}_2) > 0$ a (uniquely determined) solution to the equation $\int_{-x}^{x} p_k(t-\hat{\gamma}_2)dt = \alpha$, with $x > 0$, then equivalence will be accepted if $|T| < c_{AH}(\hat{\gamma}_2)$ ('AH' stands for 'Anderson and Hauck'). This is a result of the inequality :

$F_k(|T|-\hat{\gamma}_2)-F_k(-|T|-\hat{\gamma}_2) = \int_{-|T|}^{|T|} p_k(t-\hat{\gamma}_2)dt < \int_{-c_{AH}(\hat{\gamma}_2)}^{c_{AH}(\hat{\gamma}_2)} p_k(t-\hat{\gamma}_2)dt = \alpha$. The symmetry of p_k implies

$$\int_{-c_{AH}(\hat{\gamma}_2)}^{c_{AH}(\hat{\gamma}_2)} p_k(t-\hat{\gamma}_1)dt = \int_{-c_{AH}(\hat{\gamma}_2)}^{c_{AH}(\hat{\gamma}_2)} p_k(t-\hat{\gamma}_2)dt \quad \text{with } \hat{\gamma}_1 = -\hat{\gamma}_2 . \tag{3.3}$$

Therefore the corresponding decision function $\phi_E^{AH}(t) = 1_{(-c_{AH}(\hat{\gamma}_2),c_{AH}(\hat{\gamma}_2))}(t)$, $t\in\mathbb{R}$ can also be obtained by solving the equation (2.2), replacing the density $g(t-\gamma_2)$ by the shifted density of the central t-distribution $p_k(t-\hat{\gamma}_2)$. The family of shifted t-distributions is not Pólya Type ! Thus the theory of Pólya Type distributions cannot be applied. However, the formula (3.3) coincide with (2.8) when the variance is known and the t-distribution is replaced by the normal distribution. This formal relationship between the Anderson and Hauck procedure with unknown variance and the optimal test ϕ^*_E for the mean with known variance(s) is quite helpful for a further discussion of the procedure. In chapter II.1) we found that the confidence test ϕ^{conf}_E is a very good approximation to the optimal test ϕ^*_E with regard to practical applications. It seems reasonable that a similar result will be obtained for the Anderson and Hauck procedure. Let us write $\phi^{conf}_E(t) = 1_{(t_{k,1-\alpha}-\hat{\gamma}_2,\, t_{k,\alpha}+\hat{\gamma}_2)}(t)$, $t \in \mathbb{R}$ for the confidence test based on the t-distribution (with k degrees of freedom). Repeating the considerations of (II.1) and appendix I) leads to the result for $0 < \alpha < 1$:

$$c_{AH}(\hat{\gamma}_2) > t_{k,\alpha}+\hat{\gamma}_2\ , \quad -c_{AH}(\hat{\gamma}_2) < t_{k,1-\alpha}-\hat{\gamma}_2\ , \quad \text{and} \tag{3.4}$$

$$c_{AH}(\hat{\gamma}_2)-(t_{k;\alpha}+\hat{\gamma}_2) < \frac{F_k(-2\cdot\hat{\gamma}_2 + t_{k;1-\alpha})}{\min(p_k(t_{k;\alpha}),p_k(t_{k;(1+\alpha)/2}))} =: R(\hat{\gamma}_2) \tag{3.5}$$

$$\text{for } \hat{\gamma}_2 > \max(0,t_{k,1-\alpha})$$

We obtain for the power function the expression (see appendix III) : $E_\gamma\phi^{AH}_E(\gamma_2) = I_\gamma(c_{AH}(\gamma_2))$, where $I_\gamma(c(\gamma_2))$ denotes the integral

$$I_\gamma(c(\gamma_2)) = \int_0^\infty \left(\Psi(\frac{u}{\sqrt{k}}\cdot c(\frac{\sqrt{k}}{u}\cdot\gamma_2)-\gamma)-\Psi(-\frac{u}{\sqrt{k}}\cdot c(\frac{\sqrt{k}}{u}\cdot\gamma_2)-\gamma)\right)\eta_k(u)du$$

$\eta_k(x)$ stands for the density (3.6)

$$\eta_k(x) = \begin{cases} \dfrac{\sqrt{2\cdot\pi}}{2^{(k-2)/2}\cdot\Gamma(\frac{k}{2})}\cdot x^{k-1}\cdot g(x) & ,\ \text{for } x \geq 0 \\ 0\ , & \text{for } x < 0 \end{cases} \qquad k \in \mathbb{N}\ ,$$

where g stands for the normal density.

It should be noted that the power function of the confidence test can be obtained by changing the critical values $c_{AH}(\hat{\gamma}_2)$ into $t_{k,\alpha}+\hat{\gamma}_2$ in (3.6).

The power function of the tests are related to the distribution function of the noncentral t that would appear, if the critical values would not depend on u (see appendix III, (III.1)).
A word regarding the notation is needed. The parameter $\gamma_2 = \gamma_2(k,\frac{\delta}{\sigma})$ (or $\gamma = \gamma(k,\frac{\mu}{\sigma})$) is a function of the degrees of freedom k and the noncentrality paramter $\frac{\delta}{\sigma}$ (or $\frac{\mu}{\sigma}$). In what follows, we denote the dependence of γ_2 upon $\tau = \frac{\delta}{\sigma}$ (or γ upon μ) for given sample size(s) by $\gamma_2(\tau)$ (or by $\gamma(\mu)$). We write $\gamma_2(k)$ if γ_2 is regarded as a function of the sample size(s) for given ratio $\frac{\delta}{\sigma}$.

Differentation of the power function to μ shows immediately that both tests are unbiased, as we get the expression :

$$\frac{\partial}{\partial\mu} E_{\gamma(\mu)}\phi_E^{AH}(\gamma_2),\ \frac{\partial}{\partial\mu} E_{\gamma(\mu)}\phi_E^{conf}(\gamma_2) = \begin{cases} 0, & \text{for } \mu=0 \\ <0, & \text{for } \mu\neq 0 \end{cases} .$$

The power reaches the maximum for $\mu = 0$ for both tests and decreases strictly to zero if $|\mu| \to \infty$. Hence, the size can be obtained by setting $\mu = \delta$, i.e., $\gamma = \gamma_2$.
The asymptotic behaviour of the size $E_{\gamma_2}\phi_E^{AH}(\gamma_2)$ can be studied easily with the help of the lower and upper bounds given below. This is a direct application of (3.4) and (3.5) on (3.6).
The following inequalities hold for all $\gamma_2 = \gamma_2(\tau)$ with $\tau \geq \tau_0 > 0$ (appendix IV):

$$E_{\gamma_2}\phi_E^{AH}(\gamma_2) < \int_0^{z(\tau)} \Psi\left(\frac{u}{\sqrt{k}}\cdot t_{k,\alpha} + \frac{u}{\sqrt{k}}\cdot R\left(\frac{\sqrt{k}}{u}\cdot\gamma_2(\tau_0)\right)\right) \eta_k(u)du + \int_{z(\tau)}^{\infty} \eta_k(u)du \tag{3.7}$$

and

$$E_{\gamma_2}\phi_E^{AH}(\gamma_2) > E_{\gamma_2(\tau_0)}\phi_E^{conf}(\gamma_2(\tau_0)) =$$
$$\alpha - \int_{z(\tau)}^{\infty} \Psi\left(\frac{u}{\sqrt{k}}\cdot t_{k,\alpha}\right) \eta_k(u)du - F_{k,2\cdot\gamma_2(\tau_0)}(t_{k,\alpha}) . \tag{3.8}$$

Here it is $z(\tau) = \begin{cases} \sqrt{k}\cdot\gamma_2(\tau)/|t_{k,\alpha}| & , \text{ if } \alpha \neq \frac{1}{2} \\ \infty , & \text{ if } \alpha = \frac{1}{2} \end{cases}$

The unequality (3.8) proves a statement mentioned by Schuirmann (1987) that the size of the confidence test is always dominated by the size of the Anderson and Hauck procedure.

Application of Lebesgue's theorem ('dominated convergence') on the right side of these formulae ($\tau_0 \to \infty$) and on (3.6) leads to :

$$\lim_{\tau\to\infty} E_{\gamma_2(\tau)}\phi_E^{AH}(\gamma_2(\tau)) = \lim_{\tau\to\infty} E_{\gamma_2(\tau)}\phi_E^{conf}(\gamma_2(\tau)) = F_k(t_{k,\alpha}) = \alpha \text{ and}$$
$$\lim_{\tau\to 0} E_{\gamma_2(\tau)}\phi_E^{AH}(\gamma_2(\tau)) = F_k(t_{k,(1+\alpha)/2})-F_k(-t_{k,(1+\alpha)/2}) = \alpha. \tag{3.9}$$

The first formula says that the size of the confidence test coincides approximately with Anderson and Hauck's procedure. Both tests maintain the level for large values of τ as well as the latter one if we let τ tend to zero (second equation). We learn from (3.8) that the size of ϕ_E^{conf} is always lower than α and converges to zero as $\tau \to 0$. From numerical computation shown in figure 5, it follows that the size of Anderson and Hauck's test is always higher than the given level for $\gamma_2 > 0$. (The reader should be aware of the fact that this statement has not the same strength as the others as no mathematical proof is presented. The evaluation of (3.6) is rather arduous, but ought to be done sometime for completition.)

However, the asymptotic behaviour for $\tau \geq \tau_0 > 0$ and increasing degrees of freedom is the same for both tests :

The confidence test as well as Anderson and Hauck's procedure converge uniformly to α for $\tau \geq \tau_0 > 0$.

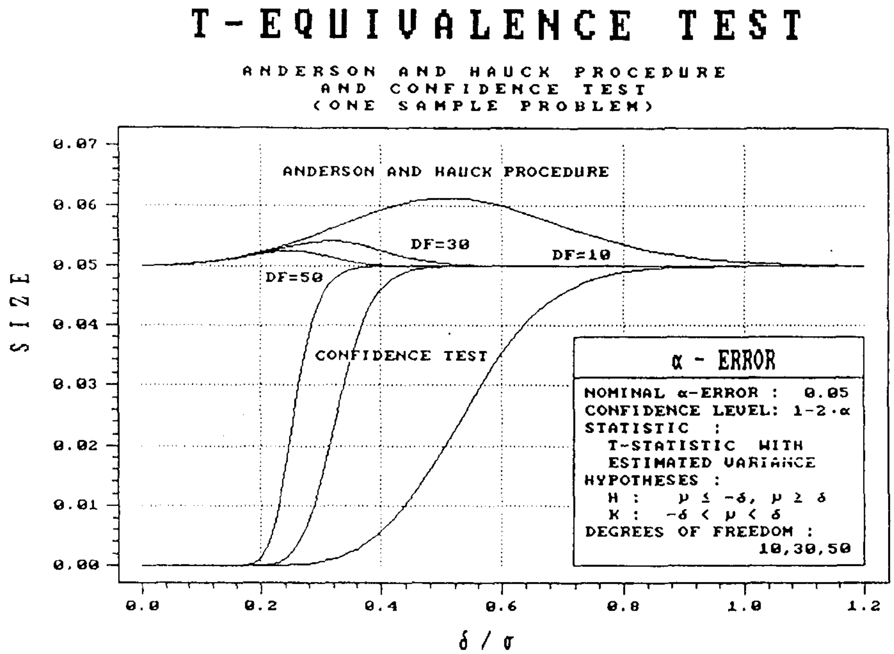
This follows from (3.7), (3.8) and (III.2) (see appendix III).

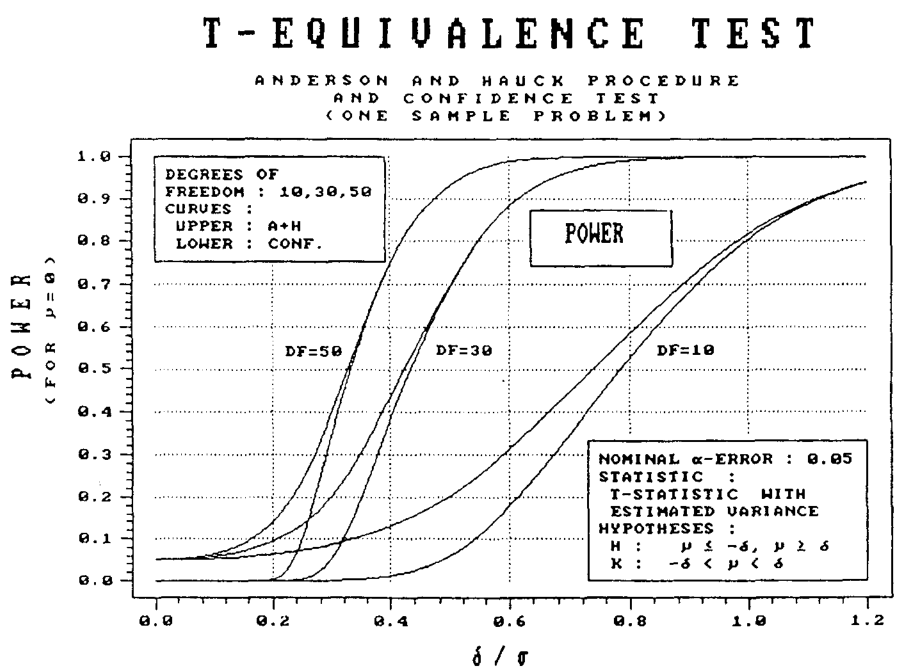
Finally, we note that the Anderson and Hauck procedure is asymptotically equal to the optimal test ϕ_E^* as $k \to \infty$:

$$\lim_{k\to\infty} (c_{AH}(\gamma_2(k))-c_2^*(\gamma_2(k))) = 0, \text{ because } |c_{AH}(\gamma_2(k))-c_2^*(\gamma_2(k))| \leq$$
$$|c_{AH}(\gamma_2(k))-(t_{k,\alpha}+\gamma_2(k))| + |u_\alpha+\gamma_2(k)-c_2^*(\gamma_2(k))| + |t_{k,\alpha}-u_\alpha| ,$$

where $c_2^*(\gamma_2(k))$ denotes the critical value of the optimal test ϕ_E^* .

The Anderson and Hauck procedure is always 'liberal', whereas the con-





FIGURES 5 AND 6

fidence test is 'conservative', particularly for low sample sizes. This fact has been noted rather often. A better procedure can be established by 'averaging' and will be described in the next section.

III.2). The method of 'averaging' and its rationale can be explained with a few words. The decision problem and the test statistics are the same as before, but the critical values are determined differently.

Therefor, we note that the integral (3.6) is an increasing functional in $c(\cdot)$, i.e.

if $0 \le c_1(\gamma_2) \le c_2(\gamma_2)$ for $\gamma_2 \ge 0$ then $I_\gamma(c_1(\gamma_2)) \le I_\gamma(c_2(\gamma_2))$ for $\gamma, \gamma_2 \ge 0$. This means that any function of critical values $c(\gamma_2)$, $\gamma_2 \ge 0$ that dominate those of the confidence test increases the size of the confidence test. We consider the following example :

$$c_{LB}(\gamma_2) = \begin{cases} t_{k,(1+\alpha)/2}, & \text{for } \gamma_2 \le t_{k,(1+\alpha)/2} + t_{k,1-\alpha} \\ t_{k,\alpha} + \gamma_2, & \text{for } \gamma_2 > t_{k,(1+\alpha)/2} + t_{k,1-\alpha} \end{cases}, \quad \gamma_2 \ge 0$$

('LB' stands for 'lower bound'). The 'lower bound test', i.e. the test belonging to the critical values $c_{LB}(\gamma_2)$ improves already the confidence test and maintains the level for $\gamma_2 = 0$ and $\gamma_2 \to \infty$ (see (3.9)). The test is also strictly 'conservative' (verified by numerical computations.) We have $c_{LB}(\gamma_2) \le c_{AH}(\gamma_2)$ for $\gamma_2 \ge 0$.

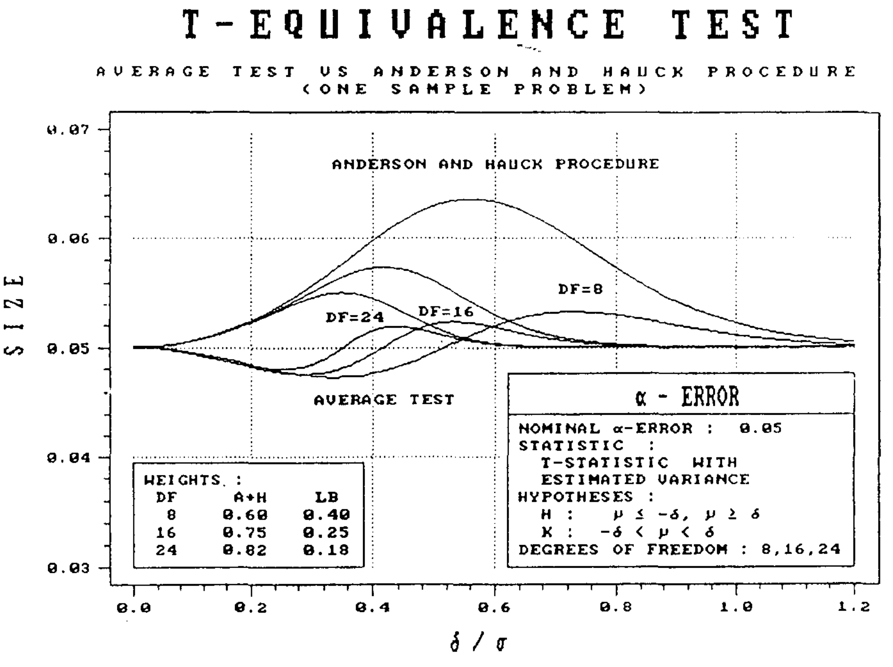
Let us define 'average' critical values $c_{AV}(\gamma_2)$ with weight functions $w_i(\gamma_2)$ as follows :

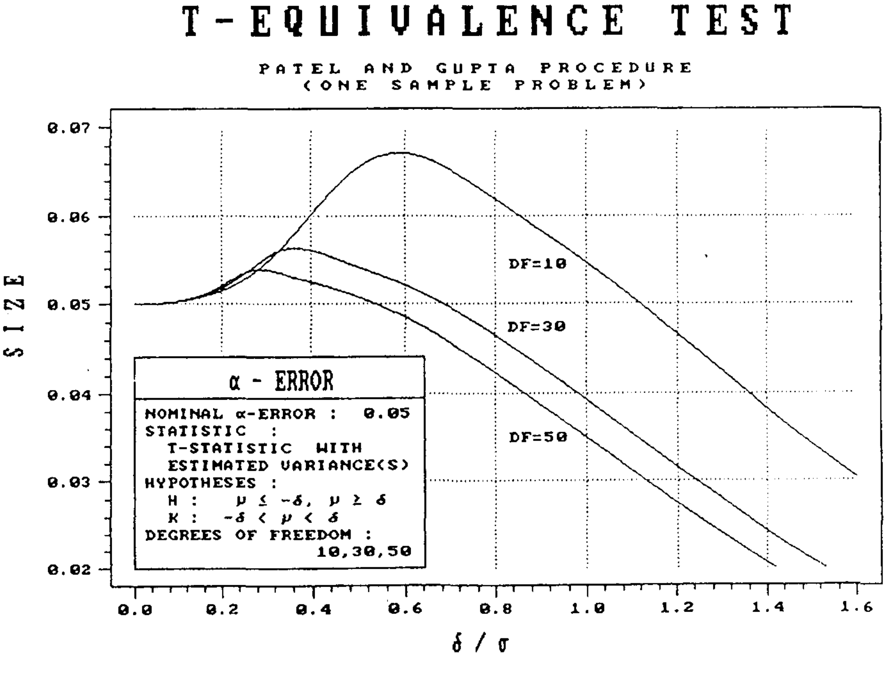
$c_{AV}(\gamma_2) = w_1(\gamma_2)\cdot c_{LB}(\gamma_2) + w_2(\gamma_2)\cdot c_{AH}(\gamma_2)$, $w_1(\gamma_2), w_2(\gamma_2) \ge 0$ and

$w_1(\gamma_2) + w_2(\gamma_2) = 1$, for $\gamma_2 \ge 0$.

There is $c_{LB}(\gamma_2) \le c_{AV}(\gamma_2) \le c_{AH}(\gamma_2)$, $\gamma_2 \ge 0$ for all weight functions.

Hence, any test $\phi_E^{AV}(t) = 1_{(-c_{AV},c_{AV})}(t)$, $t \in \mathbb{R}$, with $c_{AV} = c_{AV}(\gamma_2)$ improves the confidence test as well as the Anderson and Hauck procedure. The problem to find appropriate weight functions is as difficult as to find an optimal test. In what follows, we set the weight functions to constants. In that case the weight function should depend at least on the degrees of freedom to take into account the fact that Anderson and Hauck's procedure is asymptotically equal to the optimal test ϕ^*_E as $k \to \infty$. A meaningful and easily to handle criterion to determine the weight functions could be to ensure an 'average' level of α. From the first equation in (3.9) we know that for smaller parameters $0 \le \tau \le \tau_1$ (τ_1 depends on $k \in \mathbb{N}$) the deviation of the size from the given level becomes important. We are able to compute the weights in a way that the equation $\frac{1}{\tau_1} \cdot \int_0^{\tau_1} E_{\gamma_2(\tau)} \phi_E^{AV}(\gamma_2(\tau)) \, d\tau = \alpha$ is fullfilled, i.e., the 'average' level is provided for $0 \le \tau \le \tau_1$. Therefore, we have only to note that test becomes strictly 'conservative' when w_1 tends to 1 and strictly 'liberal' when w_2 tends to 1 and that the true level $E_{\gamma_2} \phi_E^{AV}(\gamma_2)$ depends continuously on w_1, w_2. The examples given in figure 7 show that 'average' tests provide quite accurate results even for low degrees of freedom k, i.e., $k \le 20$. Clearly, an 'average' test is also unbiased and converges uniformly to α, for $\tau \ge \tau_0 > 0$. The test cannot be expressed in terms of p-values. The equation (3.3) has always to be solved. It needs only to be applied when the confidence test shows an insignificant result but the Anderson and Hauck procedure a significant result. 'Average' tests increase the power of the confidence test, but the improvement seems to be of minor importance as this is the case with any improvement of the confidence test. This can be seen from figures 5 and 6 that summarize the above considerations graphically. The power of any test that maintains the α-level is bounded by the power of Anderson and Hauck's procedure and the confidence test (the 'lower bound test' gives a slighly sharper lower bound) and will lie inside these bounds. The point with equivalence testing is that the power is limited by the length of the alternative hypothesis and can only be increased by increasing the sample size so that asymptotic features become relevant. If the size of the test is regarded as a function of the power and if we assume that a meaningful experiment has to have at least 50 % power in the middle of the equivalence interval $(-\delta,\delta)$,





FIGURES 7 AND 8

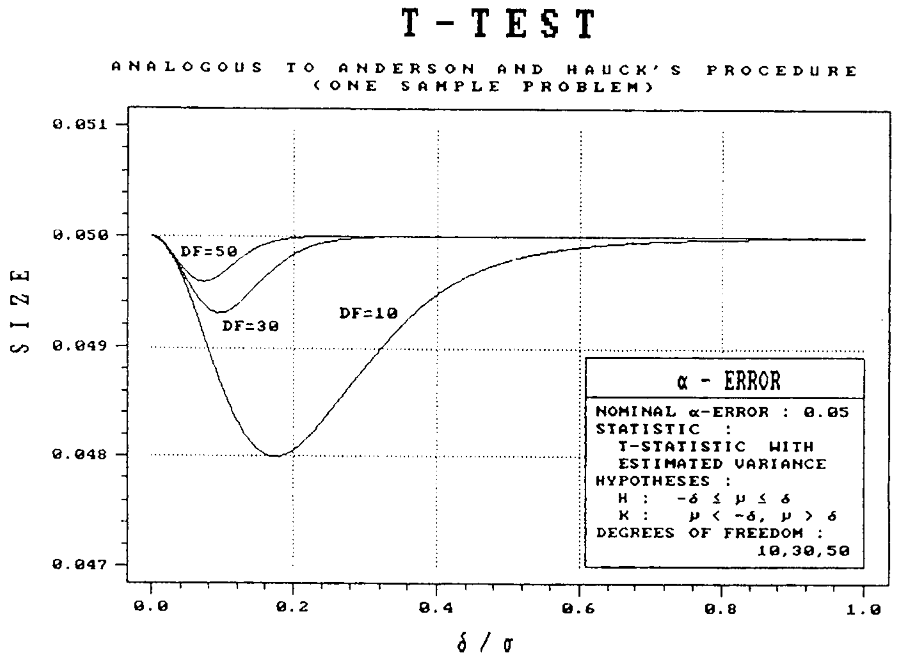
which is clearly a week condition, then the size of the confidence test is about 4,9 % (k=10) when α is set to 5 % !

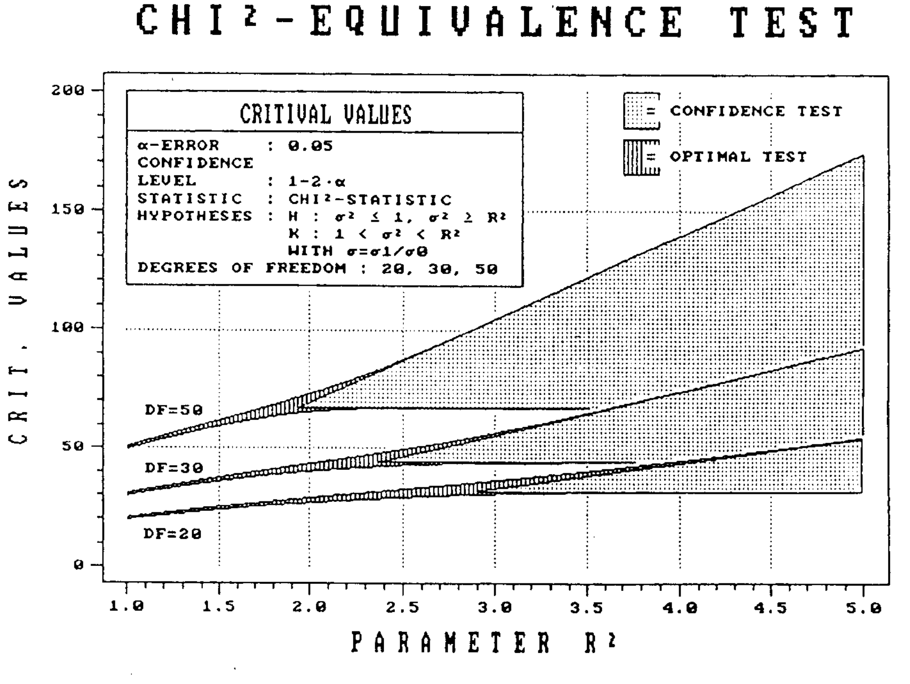
III.3) What about the procedure recommended by Patel and Gupta ? This procedure also leads to a decision on hypotheses (2.4,**E**)) with $\delta_1 = -\delta$, $\delta_2 = \delta > 0$. The alternative hypothesis is accepted if the statistic $T^2 \le c_{GP}(n,m,\hat{\gamma}_2^2,\alpha)^2 =: c_{GP}(\hat{\gamma}_2)^2$. The critical value $c_{GP}(\hat{\gamma}_2)^2$ is determined by the equation $\int_0^{c_{GP}(\hat{\gamma}_2)^2} p_{1,k;\hat{\gamma}_2^2}(x)dx = \alpha$, where $p_{1,k;\hat{\gamma}_2^2}$ denotes the density of the non-central F-distribution with 1 and k degrees of freedom and noncentrality parameter $\hat{\gamma}_2^2$. The relationship to the first-mentioned optimal test for the noncentrality parameter is as follows :
If $c_2^*(\hat{\gamma}_2)$ is the critical value of the optimal test computed (by setting $\gamma_2 = \hat{\gamma}_2$) then $c_{GP}(\hat{\gamma}_2) = c_2^*(\hat{\gamma}_2)$. This follows from the identity (3.1). Therefore the procedure of Patel and Gupta coincides with the optimal test for the noncentrality parameter if the unknown standard deviation σ in (3.2) is replaced by its estimator and the usual t-statistic T is used instead of T^2. The power function can be obtained by changing the number $c(\gamma_2)$ in (3.6) into $c_{GP}(\gamma_2)$. This test is also unbiased. The true level can be computed by setting $\gamma = \gamma_2$ and it is shown as a function of $\tau = \frac{\delta}{\sigma}$ in the figure 8. The curves in the graph support the conjecture that the size tends to zero for every fixed degree of freedom if the parameter τ tends to infinity and converges uniformly to zero for $\tau \ge \tau_0 > 0$. A short proof is given in appendix V. These properties are rather strange as convergence to 'normality' is expected at least for increasing degrees of freedom. We know from the lemma in appendix V that $c(\gamma_2) = d_0 + d_k \cdot \gamma_2$ is an upper bound for $c_{GP}(\gamma_2)$ as $\gamma \to \infty$. The number d_0 does not play a role for the asymtotic behaviour and may differ for every k (as long as the set of these numbers is bounded). Therefore we set $d_0 = t_{k,\alpha}$ to be able to compare the critical values $c_{GP}(\gamma_2)$ with those of the confidence test. For sufficiently large $\gamma_2 > 0$ we obtain :

$$c_{GP}(\gamma_2) < c(\gamma_2) = t_{k,\alpha} + d_k \cdot \gamma_2 < t_{k,\alpha} + \gamma_2$$

Now, it is easy to see that the factor $d_k < 1$ is to blame for the convergence of the size to zero as $\tau \to \infty$ when the degrees of freedom are kept constant. The (uniform) convergence of the size depends on the asymptotic properties of the sequence $(d_k)_{k\in\mathbb{N}}$. If $(d_k)_{k\in\mathbb{N}}$ tends 'slowly' to 1 then the size of the test converges to zero (see appendix V) although $c(\gamma_2)$ tends to the critical value of the confidence test. Any higher 'rate' of convergence of $(d_k)_{k\in\mathbb{N}}$ to 1 leads to a value of the size that lies between zero and α ! Therefore, the Gupta and Patel procedure is really worth studying as a nice example that some slight changes in the definition of a procedure may lead to completely unexpected results, but it cannot be recommended for equivalence testing.

III.4) A word should be said about the problem of testing a non-zero difference of means, i.e., a generalisation of Student's t-test. The problem is to test $\mathbf{H} : -\delta \le \mu \le \delta$ vs $\mathbf{K} : \mu < -\delta$ or $\mu > \delta$ with $\delta \ge 0$. This question was already covered by of a paper published by Hodges and Lehmann (1954). They presented a geometrical construction and proved when there is one degree of feedom that the construction leads to a conditional test which is 'similar on the boundary' and improves the power of the 'two one-sided tests'-method based on a level of α. The problem for higher degrees of freedom seems still to be unsolved.
The hypotheses can be tested with the 'two one-sided tests' procedure which coincides with the confidence test as we described in II.2) and we have the same delimma with the size : the test exceeds the given α-level in a neigbourhood of $\hat{\gamma}_2 = 0$, if it is based on a $1-2\cdot\alpha$ confidence interval, i.e. $\phi_D^{conf}(t) = 1_{(-\infty,t_{k,\alpha}-\hat{\gamma}_2)\cup(t_{k,1-\alpha}+\hat{\gamma}_2,\infty)}(t)$, $t\in\mathbb{R}$.
The size is kept for $\hat{\gamma}_2 = 0$, if $1-\alpha$ is used for the confidence level but tends to $\alpha/2$ for larger values of $\hat{\gamma}_2$.
Therefore we ask the following : What will happen, if we replace in the formula (2.1) the density of the shifted normal distribution $g(t-\hat{\gamma})$ by the density of the shifted t-distribution $p_k(t-\hat{\gamma})$ as we did for the Anderson and Hauck procedure ?
The notation may be the same as in III.1), but we denote by $c_D(\hat{\gamma}_2)$ a solution to the equation





FIGURES 9 AND 10

$$\int_{-\infty}^{-c_D(\hat{\gamma}_2)} p_k(t-\hat{\gamma}_2)dt + \int_{c_D(\hat{\gamma}_2)}^{\infty} p_k(t-\hat{\gamma}_2)dt = \int_{-\infty}^{-c_D(\hat{\gamma}_2)} p_k(t+\hat{\gamma}_2)dt + \int_{c_D(\hat{\gamma}_2)}^{\infty} p_k(t+\hat{\gamma}_2)dt = \alpha$$

and by ϕ_D^{AH} the following decision function

$\phi_D^{AH}(t) = 1_{(-\infty,-c_D(\hat{\gamma}_2))\cup(c_D(\hat{\gamma}_2),\infty)}(t)$, $t\in\mathbb{R}$. ϕ_D^{AH} is linked to Anderson and Hauck's procedure. If ϕ_D^{AH} is based on a $1-\alpha$ level then $\phi_D^{AH} = 1-\phi_E^{AH}$ (λ-a.e.). Therefore, the properties of ϕ_D^{AH} can be obtained from ϕ_E^{AH} easily: The power function of the test is given by the formula $E_\gamma\phi_D^{AH}(\gamma_2) = 1-E_\gamma\phi_E^{AH}(\gamma_2)$. The test is unbiased and coincides asymptotically with the confidence test (for $1-2\cdot\alpha$) because of the inequality

$$0 < c_D(\hat{\gamma}_2)-(t_{k;1-\alpha}+\hat{\gamma}_2) < \frac{F_k(-2\cdot\hat{\gamma}_2+t_{k;1-\alpha})}{p_k(t_{k;1-\alpha/2})} \quad \text{for } \hat{\gamma}_2 > 0,\ \alpha < \frac{1}{2} \tag{3.10}$$

The size can be computed by setting $\gamma = \gamma_2$ and converges uniformly to α for all $\tau \geq \tau_0 > 0$. Figure 9 shows that the test is slightly 'conservative', as expected, even for small degrees of freedom. Therefore, this procedure may be recommended as an acceptable generalisation of Student's test. The test can also be handled in terms of p-values as follows :

Accept **K**, i.e., $\mu < -\delta$ or $\mu > \delta$, if $1-F_k(|\mathbf{T}|-\hat{\gamma}_2)+F_k(-|\mathbf{T}|-\hat{\gamma}_2) < \alpha$.

IV. MONOTONE PROCEDURES FOR THE ('STRETCHED') χ^2- AND ('STRETCHED') F-DISTRIBUTION

Monotone procedures for the χ^2-distribution appear in various situations. The classical problem is to compare the (unknown) variance of a normal distribution with a given value. Another application may consist of an approximate approach to the goodness of fit testing.

We concentrate here on the equivalence and 'difference' testing only. Theorem I.5) easily allows us to set up appropriate tests.

The assumptions are the same as in II). We regard the random variables $X_1,\dots,X_n \propto N(\mu_1,\sigma_1^2)$, $Y_1,\dots,Y_m \propto N(\mu_2,\sigma_2^2)$ as stochastically independent and the variances $\sigma_1^2 > 0$, $\sigma_2^2 > 0$ and means μ_1, μ_2 to be unknown. The hypotheses are the following:

E) $H : (\sigma_1/\sigma_i)^2 \notin (r_1^2, r_2^2)$ $\quad K : (\sigma_1/\sigma_i)^2 \in (r_1^2, r_2^2)$, $\quad i = 0,2$

D) $H : (\sigma_1/\sigma_i)^2 \in [r_1^2, r_2^2]$ $\quad K : (\sigma_1/\sigma_i)^2 \notin [r_1^2, r_2^2]$, $\quad i = 0,2$

Here, $\sigma_i^2 > 0$ is a given constant for $i = 0$, the variance of the second distribution is given when $i = 2$, and $0 < r_1^2 \le r_2^2$ represent real numbers.

A remark should clarify the notation. The expression 'difference' is not appropriate in this context as a ratio and not a difference of parameters is considered. In order to avoid new notation, we also use this term for the ratio. The reader should keep in mind this distinction.

In what follows, we express the above decision problems in a standardized form by multiplying by $1/r_1$. The above hypotheses are then equivalent to the hypotheses below :

E') $H' : (\sigma_1/\sigma_i \cdot r_1)^2 \notin (1, r^2)$ $\quad K' : (\sigma_1/\sigma_i \cdot r_1)^2 \in (1, r^2)$, $\quad i = 0,2$

D') $H' : (\sigma_1/\sigma_i \cdot r_1)^2 \in [1, r^2]$ $\quad K' : (\sigma_1/\sigma_i \cdot r_1)^2 \notin [1, r^2]$, $\quad i = 0,2,$

where $r = r_2/r_1$.

We define $S(X) = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \bar{X})^2$, $S(Y)$ respectively. The statistic

$\chi^2 = \frac{(n-1) \cdot S(X)}{\sigma_0^2 \cdot r_1^2}$ is as 'stretched' χ^2-distributed with n-1 degrees of freedom and the parameter $\gamma^2 = (\frac{\sigma_1}{\sigma_0 \cdot r_1})^2$. The density $\chi^2_{\gamma^2;n}$ can be expressed by the density χ^2_n of the central χ^2-distribution:

$\chi^2_{\gamma^2;n}(x) = \frac{1}{\gamma^2} \cdot \chi^2_n(\frac{x}{\gamma^2})$, $x \in \mathbb{R}$, $n \in \mathbb{N}$, $\gamma^2 > 0$. The statistic for the two sample problem is $F = \frac{S(X)}{r_1^2 \cdot S(Y)}$, which is as 'stretched'-F distributed with n-1, m-1 degrees of freedom and the parameter $\gamma^2 = (\frac{\sigma_1}{\sigma_2 \cdot r_1})^2$.

The density $f_{\gamma^2;n,m}$ of this distribution can also be given using the density $f_{n,m}$ of the central F-distribution :

$f_{\gamma^2;n,m}(x) = \frac{1}{\gamma^2} \cdot f_{n,m}(\frac{x}{\gamma^2})$, $x \in \mathbb{R}$, $n,m \in \mathbb{N}$, $\gamma^2 > 0$ (Witting (1985), p. 217).

In order to apply the theory of Pólya Type distributions, we have to check whether these families are of Pólya Type. The 'stretched' χ^2-distribution $\left(\chi^2_{\gamma^2;n}\, d\lambda\right)_{\gamma^2 \in \mathbb{R}_+}$ is of exponential type, hence of Pólya Type ∞. The 'stretched'-F distribution $\left(f_{\gamma^2;n,m}\, d\lambda\right)_{\gamma^2 \in \mathbb{R}_+}$ also belongs to Pólya Type ∞. (This can be proved applying complete induction on definition I.1) and with the help of lemma 1 of Karlin's paper (1957), p.285). All assumptions of theorem I.5) are fillfulled so that part iii) can be applied to compute the critical $c^*_1 = c^*_1(r)$ values for $1 < r^2$:

$$\text{D')} \quad \int_0^{c^*_1} q_1(t)dt + \int_{c^*_2}^{\infty} q_1(t)dt = \int_0^{c^*_1} q_{r^2}(t)dt + \int_{c^*_2}^{\infty} q_{r^2}(t)dt = \alpha \tag{4.1}$$

$$\text{E')} \quad \int_{c^*_1}^{c^*_2} q_1(t)dt = \int_{c^*_1}^{c^*_2} q_{r^2}(t)dt = \alpha \tag{4.2}$$

Here, $q_r(t)$ denotes the 'stretched' χ^2-distribution for the one sample problem or the 'stretched' F-distribution for the two sample problem, respectively. These equations can be solved using the central χ^2- or F-distribution because of the identity $\int_{c^*_1}^{c^*_2} q_{r^2}(t)dt = \int_{c^*_1/r^2}^{c^*_2/r^2} q_1(t)dt$.

The test procedures are analogous to (2.6):

1.) Define numbers $0 < r_1 < r_2$ depending on the material problem.

2.) Standardize the hypotheses by multiplying r_1 and r_2 by $1/r_1$.

3.) Calculate the critical values c^*_1, c^*_2.

4.) Calculate the (standardized) χ^2- or F-statistic.

5.) Accept **K** (reject **H**) if χ^2 or F lies inside the critical region, i.e., **D)**: χ^2 (or F) $< c^*_1$ or χ^2 (or F) $> c^*_2$, **E)**: $c^*_1 < \chi^2$ (or F) $< c^*_2$.

Finally, we want to compare the optimal tests with the 'two one-sided tests'-procedure as we did in the previous section. (It is known that a confidence test does not coincide with the 'two one-sided tests'-procedure. This is caused be the asymmetry of the χ^2- and F-distribution, which means that the critical values do not correspond to $\alpha/2$- or $1-\alpha/2$ fractiles. See Lehmann (1986) p. 217 or Witting (1985) p. 263, 378) In order to simplify the notation, we denote by z_α the α-fractile of the central χ^2-distribution with n degrees of freedom or the α-fractile of the central F-distribution with n and m degrees of freedom, respectively.

The decision function of the procedure for equivalence is of the form $\phi_E^{(2)}(s) = 1_{(z_{1-\alpha}, r^2 \cdot z_\alpha)}(s)$ and it is $\phi_E^{(2)}(s) < \phi_E^*(s)$ with

$\phi_E^*(s) = 1_{(c_1^*(r), c_2^*(r))}(s)$ with $s \geq 0$. The test is 'conservative' and can only be used for a ratio of parameters $r^2 > z_{1-\alpha}/z_\alpha$. The χ^2- and F-distributions are not symmetric, therefore we need different inequalities for the lower and upper critical values.

We obtain the formulae for $\alpha < \frac{1}{2}$ and $r^2 > \frac{z_{1-\alpha}}{z_\alpha}$:

$$0 < z_{1-\alpha} - c_1^*(r) < \frac{1}{\min(q(z_{1-2\cdot\alpha}), q(z_{1-\alpha}))} \cdot \int_{c_0(r)}^{\infty} q(x)dx$$

$$0 < c_2^*(r) - r^2 \cdot z_\alpha < \frac{r^2}{\min(q(z_\alpha), q(z_{2\cdot\alpha}))} \cdot \int_0^{z_{1-\alpha}/r^2} q(x)dx \tag{4.3}$$

The value $c_0(r)$ is the (uniquely determined) solution to the equation $p_1(x) = p_{r^2}(x)$, $x > 0$ with $c_0(r) = 2 \cdot n \cdot \frac{r^2 \cdot \ln r}{r^2 - 1}$ for the χ^2-distribution with n degrees of freedom and $c_0(r) = \left(\frac{m}{n} \cdot (r^2)^{\frac{n}{n+m}} - 1 \right) \cdot \frac{r^2}{r^2 - (r^2)^{\frac{n}{n+m}}}$ for the F-distribution with n and m degrees of freedom. Finally, we can gain power using an optimal test procedure. The optimal χ^2-equivalence test increases the power by about 0.21 ($\alpha = 0.05$, 10 degrees of freedom). Therefore, the results are quite similar to those that are discussed at the end of II.1).

The 'two one-sided tests'-procedure for the problem **D)** is $\phi_D^{(2)}(s) = 1_{(0,z_\alpha)\cup(z_{1-\alpha}\cdot r^2,\infty)}(s)$, $s \geq 0$ and exceeds the nominal level α, as the relation $\phi_D^{(2)} > \phi_D^*$ holds, where $\phi_D^*(s) = 1_{(0,c_1^*(r))\cup(c_2^*(r),\infty)}(s)$, $s \geq 0$ denotes the optimal test. This fact seems to be less important, as a test for 'difference' is usually of interest for larger parameters only. *The following expressions tell us that this procedure approximates* to the optimal test for $\alpha < \frac{1}{2}$:

$$0 < z_\alpha - c_1^*(r) < \frac{1}{\min(q(z_\alpha),q(z_{\alpha/2}))} \cdot \int_{r^2\cdot z_{1-\alpha}}^{\infty} q(x)dx \quad \text{for } r^2 > \frac{z_{1-\alpha/2}}{z_{1-\alpha}} \tag{4.4}$$

$$0 < c_2^*(r) - r^2\cdot z_{1-\alpha} < \frac{r^2}{q(z_{1-\alpha/2})} \cdot \int_0^{z_\alpha/r^2} q(x)dx \qquad \text{for } r^2 > \frac{z_\alpha}{z_{\alpha/2}}$$

APPENDIX I

Proof of (2.7), (2.9), (3.5), (3.10), (4.3) and (4.4):

The inequality (2.7) can be seen by partitioning of $\int_{-\infty}^{c_2^*(\gamma_2)} g(t-\gamma_2)dt$ into:

$$\left(\int_{-\infty}^{u_\alpha+\gamma_2} + \int_{u_\alpha+\gamma_2}^{c_2^*(\gamma_2)}\right) g(t-\gamma_2)dt = \left(\int_{-\infty}^{-c_2^*(\gamma_2)} + \int_{-c_2^*(\gamma_2)}^{c_2^*(\gamma_2)}\right) g(t-\gamma_2)dt \,. \quad \text{It is}$$

$$\int_{-\infty}^{u_\alpha+\gamma_2} g(t-\gamma_2)dt = \int_{-c_2^*(\gamma_2)}^{c_2^*(\gamma_2)} g(t-\gamma_2)dt = \alpha \,, \text{ because of the equation (2.2).}$$

Hence : $\int_{u_\alpha+\gamma_2}^{c_2^*(\gamma_2)} g(t-\gamma_2)dt = \int_{-\infty}^{-c_2^*(\gamma_2)-\gamma_2} g(t)dt$. A lower bound of $\int_{u_\alpha+\gamma_2}^{c_2^*(\gamma_2)} g(t-\gamma_2)dt$ is given by $(c_2^*(\gamma_2)-(u_\alpha+\gamma_2))\cdot\min(g(u_\alpha),g(u_{(1+\alpha)/2}))$ because $u_\alpha < c_2^*(\gamma_2)-\gamma_2$

$< u_{(1+\alpha)/2}$ for $\gamma_2 > 0$. Otherwise we would obtain the contradiction:

$$\alpha = \int_{-c_2^*(\gamma_2)}^{c_2^*(\gamma_2)} g(t-\gamma_2)dt > \int_{-u_{(1+\alpha)/2}-2\cdot\gamma_2}^{u_{(1+\alpha)/2}} g(t)dt \geq 2 \cdot \int_0^{u_{(1+\alpha)/2}} g(t)dt = \alpha.$$

The inequality $-c_2^*(\gamma_2) < -(u_\alpha+\gamma_2)$ leads to the result:

$$(c_2^*(\gamma_2)-(u_\alpha+\gamma_2))\cdot\min(g(u_\alpha),g(u_{(1+\alpha)/2})) < \int_{u_\alpha+\gamma_2}^{c_2^*(\gamma_2)} g(t-\gamma_2)dt = \int_{-\infty}^{-c_2^*(\gamma_2)-\gamma_2} g(t)dt <$$

$\Psi(-2\cdot\gamma_2+u_{1-\alpha})$.

The analogous formula (2.9) for the problem **D)** can be obtained from (2.7) by changing α into $1-\alpha$. The same method works for the shifted t-distribution (see (3.5) and (3.10)) as well as for the χ^2- and F-distributions. We only demonstrate the first formula in (4.3). The others can be shown in a very similar manner.

$$\int_{c_1^*(r)}^{\infty} q(t)dt = \left(\int_{c_1^*(r)}^{z_{1-\alpha}} + \int_{z_{1-\alpha}}^{\infty}\right) q(t)dt = \left(\int_{c_1^*(r)}^{c_2^*(r)} + \int_{c_2^*(r)}^{\infty}\right) q(t)dt .$$

Therefore, it follows that $(z_{1-\alpha}-c_1^*(r))\cdot\min(q(z_{1-2\cdot\alpha}),q(z_{1-\alpha})) <$

$\int_{c_2^*(r)}^{\infty} q(t)dt$, because $z_{1-2\cdot\alpha} \leq c_1^*(r) < z_{1-\alpha} < z_\alpha\cdot r^2 < c_2^*(r)$. This is true as the contrary assumption $c_1^*(r) < z_{1-2\cdot\alpha}$ would lead to :

$\alpha = \int_{c_1^*(r)}^{c_2^*(r)} q(t)dt > \int_{z_{1-2\cdot\alpha}}^{z_{1-\alpha}} q(t)dt = \alpha$. In order to get an upper bound, where the value $c_2^*(r)$ does not appear, we see that :

$c_1^*(r) < c_0(r) < c_2^*(r)$ for all $r \geq 1$.

APPENDIX II

Proof of (2.10) : We demonstrate the property for the usual choice of α: let $0 < \alpha < \frac{1}{2}$ and $d > 0$. The test ϕ_{E+D}^* is 'similar on the boundary', therefore we get for $\gamma = \gamma_4 = \gamma_3+d$:

$$\alpha = \left(\int_{-\infty}^{c_1^*(\gamma_3)-\gamma_3} + \int_{c_2^*(\gamma_3)-\gamma_3}^{c_3^*(\gamma_3)-\gamma_3} + \int_{c_4^*(\gamma_3)-\gamma_3}^{\infty} \right) g(t-d)dt$$ and this is also true for

$\gamma = \gamma_3$. The differences $c_1^*(\gamma_3)-\gamma_3$ and $c_2^*(\gamma_3)-\gamma_3$ tend to $-\infty$ as $\gamma_3 \to \infty$, because $c_1^*(\gamma_3)$, $c_2^*(\gamma_3) < 0$. The difference $c_1^*(\gamma_3)-\gamma_3$ converges to a real number, say u_1. Hence, $\lim_{\gamma_3\to\infty} c_4^*(\gamma_3)-\gamma_3 = u_2$ and u_1, u_2 is a solution to (2.10). This is ensured since :

i) $c_3^*(0) > 0$.

ii) There exists a number s, so that $c_3^*(\gamma_3) < \gamma_3$ for $\gamma_3 \geq s$.

iii) The function $\gamma_3 \to c_3^*(\gamma_3)-\gamma_3$ has a lower bound.

iv) The function $\gamma_3 \to c_3^*(\gamma_3)-\gamma_3$ decreases.

i) follows from equation (2.3). If ii) is false we would obtain the contradiction : $\int_{c_2^*(\gamma_3)}^{c_3^*(\gamma_3)} g(t-\gamma_3)dt \geq \int_{-2\cdot\gamma_3}^{0} g(t)dt \to \frac{1}{2} > \alpha$ for $\gamma_3 \to \infty$, because $c_2^*(\gamma_3) = -c_3^*(\gamma_3)$.

ad iii) : The assumptions $\lim_{\gamma_3\to\infty} c_3^*(\gamma_3)-\gamma_3 = -\infty$ would lead to :

$\lim_{\gamma_3\to\infty} c_4^*(\gamma_3)-\gamma_3 =: u'$ with $\int_{u'}^{\infty} g(t-d)dt = \int_{u'}^{\infty} g(t)dt = \alpha$ for $d > 0$, as the first identity holds also for $\gamma = \gamma_3$.

Proof of iv): The function $\gamma_3 \to c_3^*(\gamma_3)-\gamma_3$ cannot be increasing. Otherwise $c_3^*(\gamma_3)-\gamma_3$ would be strictly positive, because of i). But this contradicts ii).

The case 'd = 0' can be obtained from the above one by letting $d \to 0$.

APPENDIX III

Proof of (3.6) : Let $c(\gamma_2)$, $\gamma_2 \geq 0$ be any critical value, the power function $\mathbb{P}_\omega(-c(\hat{\gamma}_2) < T < c(\hat{\gamma}_2))$, $\omega = (\mu,\sigma^2) \in \mathbb{R}\times\mathbb{R}_+$, with $T = T(X)$ or

$T = T(X,Y)$ then can be obtained by Fubini's theorem ('conditioning on s'). This yields :

$\mathbb{P}_\omega(-c(\hat{\gamma}_2) < T < c(\hat{\gamma}_2)) \; = \; E_\omega \mathbb{P}_\omega(-c(\gamma_2(s)) < T < c(\gamma_2(s)) \mid \hat{\sigma}=s)$, where $\gamma_2(s)$ is of the form (3.2), but σ replaced by s and γ denotes the non-centrality parameter. The estimation of the mean and standard deviation is stochastically independent, hence T is distributed as $N(\gamma\cdot\frac{\sigma}{s},(\frac{\sigma}{s})^2)$. So, the identity

$$\mathbb{P}_\omega(-c(\gamma_2(s)) < T < c(\gamma_2(s)) \mid \hat{\sigma}=s) = \Psi(\tfrac{s}{\sigma}\cdot c(\gamma_2(s))-\gamma)-\Psi(-\tfrac{s}{\sigma}\cdot c(\gamma_2(s))-\gamma),$$

holds. It is $\gamma_2(s) = \gamma_2\cdot\frac{\sqrt{k}}{u}$, with $u = \frac{s}{\sigma}\cdot\sqrt{k}$. $\frac{k\cdot s^2}{\sigma^2}$ follows a χ^2- distribution with k degrees of freedom, hence u is $\eta_k(x)d\lambda$ distributed, where $\eta_k(x)$ stands for the following density :

$$\eta_k(x) = \begin{cases} \dfrac{\sqrt{2\cdot\pi}}{2^{(k-2)/2}\cdot\Gamma(\frac{k}{2})}\cdot x^{k-1}\cdot g(x)\ , & \text{for } x \geq 0 \\ 0\ , & \text{for } x < 0 \end{cases} \qquad k \in \mathbb{N}\ ,$$

where g denotes the normal density. This gives the identity below when $c(\gamma_2)$ is replaced by $c_{AH}(\gamma_2)$:

$$E_\gamma\phi_E^{AH}(\gamma_2) = \int_0^\infty \left(\Psi(\frac{u}{\sqrt{k}}\cdot c_{AH}(\frac{\sqrt{k}}{u}\cdot\gamma_2)-\gamma)-\Psi(-\frac{u}{\sqrt{k}}\cdot c_{AH}(\frac{\sqrt{k}}{u}\cdot\gamma_2)-\gamma)\right)\ \eta_k(u)du$$

for the power function that shows a formal relationship to the distribution function of the noncentral t :

$$F_{k,\gamma}(t) = \int_0^\infty \Psi(\frac{u\cdot t}{\sqrt{k}}-\gamma)\ \eta_k(u)du. \tag{III.1}$$

There is $F_{k,0}(t) = F_k(t)$ and $F_{k,\gamma}(0) = \Psi(-\gamma)$, $t,\gamma \in \mathbb{R}$.

We note for later purposes that the application of the substitution $u = \sqrt{x}$ on $\int_0^z \eta_k(u)du$ shows the identity : $\int_0^z \eta_k(u)du = P(\chi^2 \leq z^2)$ where $P(\chi^2\leq t)$ denotes the distribution function of χ^2 with k degrees of freedom. If k is sufficiently large, we have

$$\int_0^z \eta_k(u)du = P(\chi^2 \le z^2) \cong \Psi(\sqrt{2}\cdot z - \sqrt{2\cdot k - 1}) . \qquad \text{(III.2)}$$

APPENDIX IV

Proof of (3.7) and (3.8) : Some lower and upper bounds for the true level $E_{\gamma_2}\phi_E^{AH}(\gamma_2)$ can be established as follows : We set $\gamma_2 = \gamma_2(\tau)$ with $\tau \ge \tau_0 > 0$ and use the partition :

$$E_{\gamma_2}\phi_E^{AH}(\gamma_2) < \int_0^\infty \Psi(\frac{u}{\sqrt{k}}\cdot c_{AH}(\frac{\sqrt{k}}{u}\cdot\gamma_2)-\gamma_2)\ \eta_k(u)du =$$

$$(\int_0^{z(\tau)} + \int_{z(\tau)}^\infty)\ \Psi\Big(\frac{u}{\sqrt{k}}\cdot(c_{AH}(\frac{\sqrt{k}}{u}\cdot\gamma_2)-\frac{\sqrt{k}}{u}\cdot\gamma_2)\Big)\ \eta_k(u)du .$$

The application of (3.4) leads to

$$c_{AH}(\frac{\sqrt{k}}{u}\cdot\gamma_2)-\frac{\sqrt{k}}{u}\cdot\gamma_2 < t_{k,\alpha} + R(\frac{\sqrt{k}}{u}\cdot\gamma_2(\tau_0)),\ \text{for } u \le z(\tau),\ \text{where } R(\gamma_2)$$

stands for the function on the right in (3.5), which decreases monotonously in γ_2. Therefore, the integrant of the first term is dominated by $\Psi\Big(\frac{u}{\sqrt{k}}\cdot t_{k,\alpha} + \frac{\cdot u}{\sqrt{k}}\cdot R(\frac{\sqrt{k}}{u}\cdot\gamma_2(\tau_0)\Big)$ and the second by 1. This proves (3.7).

A lower bound can be given using the same partition :

$$E_{\gamma_2}\phi_E^{AH}(\gamma_2) > \int_0^{z(\tau)} \Big(\Psi(\frac{u}{\sqrt{k}}\cdot c_{AH}(\frac{\sqrt{k}}{u}\cdot\gamma_2)-\gamma_2)-\Psi(-\frac{u}{\sqrt{k}}\cdot c_{AH}(\frac{\sqrt{k}}{u}\cdot\gamma_2)-\gamma_2)\Big)\ \eta_k(u)du.$$

From (3.4) it follows that $c_{AH}(\frac{\sqrt{k}}{u}\cdot\gamma_2)-\frac{\sqrt{k}}{u}\cdot\gamma_2 > t_{k,\alpha}$ and

$$-c_{AH}(\frac{\sqrt{k}}{u}\cdot\gamma_2)-\frac{\sqrt{k}}{u}\cdot\gamma_2 < -t_{k,\alpha} - 2\cdot\frac{\sqrt{k}}{u}\cdot\gamma_2(\tau)\ \text{for } 0 < u \le z.$$

Finally, we get the expression:

$$E_{\gamma_2}\phi_E^{AH}(\gamma_2) > \int_0^{z(\tau)} \Big(\Psi(\frac{u}{\sqrt{k}}\cdot t_{k,\alpha})-\Psi(-\frac{u}{\sqrt{k}}\cdot t_{k,\alpha} - 2\cdot\gamma_2(\tau_0))\Big)\ \eta_k(u)du =$$

$$\alpha - \int_{z(\tau)}^{\infty} \Psi(\frac{u}{\sqrt{k}} \cdot t_{k,\alpha})\ \eta_k(u)du - \int_{0}^{z(\tau)} \Psi(-\frac{u}{\sqrt{k}} \cdot t_{k,\alpha} - 2\cdot\gamma_2(\tau_0))\ \eta_k(u)du >$$

$$\alpha - \int_{z(\tau)}^{\infty} \Psi(\frac{u}{\sqrt{k}} \cdot t_{k,\alpha})\ \eta_k(u)du - F_{k,2\cdot\gamma_2(\tau_0)}(t_{k,\alpha}) .$$

APPENDIX V

Some bounds for the critical values $c_{GP}(\gamma)$ of the Gupta and Patel procedure can easily be established with the help of (3.1),(III.1) and (III.2).

LEMMA : Let $\alpha < \frac{1}{3}$ and define $c(\gamma) = d_0 + d_k\cdot\gamma$, where d_0 stands for any real number and d_1 for a positive real number with $d_1 < 1$, $d_1 \cong 1$ then the unequality $c_{GP}(\gamma) < c(\gamma)$ holds for sufficiently large parameter $\gamma > 0$.

Proof : From (3.1) we learn that we have to demonstrate the unequation $\int_{-c(\gamma)}^{c(\gamma)} p_{k;\gamma}(t)dt = F_{k,\gamma}(c(\gamma))-F_{k,\gamma}(-c(\gamma)) > \alpha$. There is $-c(\gamma) < 0$ and

$F_{k,\gamma}(0) = \Psi(-\gamma)$. This leads to

$F_{k,\gamma}(c(\gamma))-F_{k,\gamma}(-c(\gamma)) > F_{k,\gamma}(c(\gamma))-\Psi(-\gamma)$. $F_{k,\gamma}(c(\gamma))$ can be treated with the help of formula (III.1) using the partition $\int_0^\infty = \int_0^z + \int_z^\infty$ with $z = \frac{\sqrt{k}}{d_1} + \varepsilon$ for $\varepsilon > 0$ and ε sufficiently small :

$$F_{k,\gamma}(c(\gamma)) > \int_z^\infty \Psi(\frac{u\cdot d_0}{\sqrt{k}} + (u\cdot \frac{d_1}{\sqrt{k}} - 1)\cdot\gamma)\ \eta_k(u)du >$$

$\Psi(d_0+\rho\cdot\gamma)\cdot\int_z^\infty \eta_k(u)du$, because there is $\frac{u}{\sqrt{k}} \geq 1$ and $u\cdot \frac{d_1}{\sqrt{k}} - 1 \geq \rho > 0$ for $u \geq z$. From (III.2) it follows : $\int_z^\infty \eta_k(u)du = 1 - P(\chi^2 \leq (\frac{\sqrt{k}}{d_k}+\varepsilon)^2)$. We have $P(\chi^2 \le (\frac{\sqrt{k}}{d_1}+\varepsilon)^2) \cong \frac{1}{2}$, as the expected value for χ^2 is k. Hence, the final result is obtained :

$$F_{k,\gamma}(c(\gamma))-F_{k,\gamma}(-c(\gamma)) > \Psi(d_0+\rho\cdot\gamma)\cdot\left(1-P(\chi^2 \le (\frac{\sqrt{k}}{d_1}+\varepsilon)^2)\right)-\Psi(-\gamma) > \alpha \quad , \text{ if}$$

$\gamma > 0$ is sufficiently large. -

In order to prove the statement of the Patel and Gupta procedure we define for every degree of freedom $k \ge 1$ and $\tau \ge \tau_0 > 0$ an upper bound of the critical values as follows :

$c(\gamma_2) = c(\tau,k) = d_0 + d_k\cdot\gamma_2$, with, $d_k = \frac{1}{1+f^{-1/4}}$, where f = n for the one sample problem and k = n-1 or $f = n\cdot m/(n+m)$ and k = n+m-2 for the two samples problem, respectively.

The above lemma tells us that there exists a number $\gamma_2^0 > 0$ so that $c(\gamma_2)$ dominates $c_{GP}(\gamma_2)$ when $\gamma_2 = \gamma_2(\tau,k) \ge \gamma_2^0$. From (III.1) and the above used partition with $z = z(\tau,k) = \sqrt{k}\cdot\gamma_2/\gamma_2^0$, we obtain the un-

equalities : $E_{\gamma_2}\phi_E^{GP}(\gamma_2) < \int_0^\infty \left(\Psi(\frac{u}{\sqrt{k}}\cdot c_{GP}(\frac{\sqrt{k}}{u}\cdot\gamma_2)-\gamma_2)\right)\,\eta_k(u)du <$

$$\int_0^z \left(\Psi(\frac{u}{\sqrt{k}}\cdot(d_0+ d_k\cdot\frac{\sqrt{k}}{u}\cdot\gamma_2)-\gamma_2)\right)\,\eta_k(u)du + \int_z^\infty \eta_k(u)du <$$

$$\int_0^\infty \left(\Psi(\frac{u}{\sqrt{k}}\cdot d_0- (1-d_k)\cdot\gamma_2)\right)\,\eta_k(u)du + \int_z^\infty \eta_k(u)du = F_{k,\xi}(d_0) + \int_z^\infty \eta_k(u)du$$

with $\xi = (1-d_k)\cdot\gamma_2(\tau,k)$.

Using the notation given in III.2), we have $\lim_{\gamma_2\to\infty}(1-d_k)\cdot\gamma_2 = \infty$ and $\lim_{\tau\to\infty} F_{k,\xi(\tau)}(d_0) = \lim_{k\to\infty} F_{k,\xi(k)}(d_0) = 0$.

There is $\lim_{\tau\to\infty}\int_{z(\tau)}^\infty \eta_k(u)du = 0$ for every degree of freedom k. The integral also converges to zero with respect to k when τ is kept fixed. This follows from (III.2) :

$\int_{z}^{\infty} \eta_k(u)du = 1 - P(\chi^2 \le z^2) \cong 1 - \Psi(\sqrt{2\cdot k\cdot f}\cdot\tau/\gamma_2^0 - \sqrt{2\cdot k-1})$. Finally, we have : $\lim_{\tau\to\infty} E_{\gamma_2(\tau)}\phi_E^{GP}(\gamma_2(\tau)) = \lim_{k\to\infty} E_{\gamma_2(k)}\phi_E^{GP}(\gamma_2(k)) = 0$. The latter convergence is uniform on $\tau \ge \tau_0 > 0$, as $\gamma_2(k,\tau)$ increases monotonously in τ. -

ACKNOWLEDGMENT

I would like to thank my colleague Stephen Senn for many helpful discussions on equivalence testing and Margitta Koch and Roger Heath for their valuable assistance with the English. The errors which remain are my own.

REFERENCES

Anderson, S. and Hauck, WW. (1983). *A new procedure for testing equivalence in comparative bioavailability and other clinical trials*, Comm. Stat. A 12, 2663-2692

Berkson, J. (1938). *Some difficulties of interpretation encountered in the application of the χ^2-test*, J. Amer. Statis. Ass., 33, 526-542

Frick, H. (1987). *On level and power of Anderson and Hauck's procedure for testing equivalence in comparative bioavailability*, Commun. Statis.-Theory Meth., 16(9), 2771-2778

Frick, H. (1991). *On a special feature of the Patel-Gupta equivalence test*, Biom. J. 33 (2), 221-224

Frick, H. (1992). *On some properties and asymptotic results of Anderson-Hauck's procedure and Patel-Gupta's noncentral t-method for testing equivalence*, Vollmar, J. (editor) : Biometrie in der chemisch-pharmazeutischen Industrie, Gustav Fischer Verlag, Stuttgart-New York 1992

Hauck, WW. and Anderson, S. (1992). *Types of bioequivalence and related statistical considerations*, Intern. J. of Clinical Pharmacology (30), 181-187

Hodges, J.L. Jr. and Lehmann, E.L. (1954). *Testing the approximate validity of statistical hypotheses*, J. Roy Statis. Soc. (B), 16, 261-268

Johnson, N.L. and Kotz, S. (1970). *Continuous univariate distributions - 2*, John Wiley & Sons, Inc.

Karlin, S. (1956). *Decision theory for Pólya Type distributions. Case of two actions, I*, Third Berkeley Symposium on Probability and Statistics, Vol. 1, 115-129

Karlin, S. (1957). *Pólya Type distributions, II*, Ann. Math. Statis., 28, 281-308

Lehmann, E.L. (1986). *Testing statistical hypotheses*, John Wiley & Sons, Inc., second edition

O'Quigley, J. and Baudoin, C. (1988). *General approaches to the problem of bioequivalence*, The Statistician, 37, 51-58

Patel, H.I. and Gupta, G.D. (1984). *A problem of equivalence in clinical trials*, Biom. J. 26 (5), 471-474

Schuirmann, D.J. (1987). *A comparision of the two one-sided test procedure and the power approach for assessing the equivalence of average biovialability*, J. Pharmacokinetics and Biopharmaceutics (15), 657-680

Senn, S.J. (1991). *Falsificationism and clinical trials*, Statistics in Medicine, 10, 1672-1692

Westlake, J.W. (1972). *The use of confidence intervals in the analysis of comparative bioavialability trials*, J. Pharmaceutical Science (61), 1340-1341

Witting, H. (1985). *Mathematische Statistik, I* B.G. Teubner, Stuttgart

Received May 1992; Revised January 1993