Austral. J. Statist., **14 (2)**, 1972, 167–171

ON THE BEHRENS-FISHER PROBLEM[1]

D. A. S. Fraser and F. Streit
University of Toronto, Canada, and
University of Berne, Switzerland

Summary

In this article the Behrens-Fisher problem is reformulated in terms of a structural model of inference. For this version of the problem a solution is obtained which is valid for arbitrary absolutely continuous error distributions. These results are further discussed for the standard normal distribution and for some other special cases with not normally distributed populations.

1. Introduction

The Behrens-Fisher problem in its original simple version can be formulated as follows. Given are two samples

$$x_{11}, \ldots, x_{1n_1};$$
$$x_{21}, \quad \ldots \quad, x_{2n_2}.$$

It is assumed that the values of the first sample are generated from a normal distribution with mean μ_1 and variance σ_1^2 and that the values of the second sample come from a normal distribution with mean μ_2 and variance σ_2^2. The true values $\mu_{10}, \mu_{20}, \sigma_{10}^2, \sigma_{20}^2$ of the parameters are not known and the sample sizes and the variances are possibly not equal, i.e. $n_1 \neq n_2$ and $\sigma_{10}^2 \neq \sigma_{20}^2$. The problem consists in making inference about the actual value ζ_0 of the difference $\zeta = \mu_1 - \mu_2$ of the means.

We first give a short review of the possibility of application of various methods of inference to this problem. A good solution in terms of an estimator can easily be found. On the other hand an exact solution in form of a confidence interval for ζ which is based only on the sample means and variances does not exist as noted by Wilks (1940). Several other solutions are discussed in the literature (see Banerji (1960), Chand (1950), Chernoff (1949), Scheffé (1943, 1944), Wallace (1958)), but they do not seem to make optimal use of the information provided by the sample; the most prominent of them is probably the one due to Scheffé. Recently attempts have been made to obtain confidence intervals by sequential sampling; for instance Srivastava (1966, 1970) has proposed a procedure with some desirable asymptotic properties. Similarly no entirely satisfactory test for the Behrens-Fisher problem has been derived (for the most frequently used tests see Scheffé (1970)). Provided that a prior distribution for the unknown parameters is effectively known an appropriate approach is to use the Bayesian method of inference. Finally we mention that the solution for this problem originally presented by Fisher (1935) (see also Behrens (1929)) is given in form

[1] Manuscript received July 22, 1971; revised January 4, 1972.

of a probability-distribution for ζ which is conditioned on the observed data and which can be characterized by the relation

$$\zeta = \mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2 - \left(\frac{s_{x_1}}{\sqrt{n_1}} t_{n_1 - 1} - \frac{s_{x_2}}{\sqrt{n_2}} t_{n_2 - 1} \right). \tag{1}$$

Here and in the following $\bar{x}_j$ denotes the sample mean and s_{x_j} the standard deviation of the j-th sample and t_ν is a random variable with a t-distribution on ν degrees of freedom. Fisher obtained (1) in using the "method of fiducial inference". Tukey (1957) has shown in the meantime that this derivation is inconsistent, because the distribution obtained for ζ by this method is not unique.

In the attempts to solve the Behrens-Fisher problem which have been discussed so far, the observed values are considered to be the primary quantities and an assumption about their distribution is made. In contrast, the basic assumption of the structural approach which we shall now develop for this problem (see also Fraser and Streit (1969)) is that the random fluctuations apparent in the experiment are generated by a random variable with known distribution. It is assumed that there correspond to the above indicated observed values the error values

$$\begin{matrix} e_{11}, \ldots, e_{1n_1}; \\ e_{21}, \quad \ldots \quad, e_{2n_2}. \end{matrix}$$

e_{jl} measures the effect of the random fluctuations inherent in the experiment on x_{jl}. We may assume that the e_{jl}'s are standard normally distributed. Furthermore, we can suppose that the relation between the x_{jl}'s and the e_{jl}'s is

$$x_{jl} = \mu_j + \sigma_j e_{jl} = [\mu_j, \sigma_j] \; 0 \; e_{jl} \quad [l = 1, \ldots, n_j; \; j = 1, 2], \tag{2}$$

thus the observed values are obtained by multiplying the corresponding error values by the unknown standard deviation and adding the unknown mean of the population. Note that we can assign to $(\mu_1, \mu_2, \sigma_1, \sigma_2)$ a pair of linear transformations, namely $[\mu_1, \sigma_1]$ acting on $\mathscr{R}^{n_1}$ and $[\mu_2, \sigma_2]$ acting on $\mathscr{R}^{n_2}$.[1]

The above interpretation of the problem differs from the usual one in so far as a random variable for the error is incorporated in the model, as the distribution of the observed value is only implicitly specified by this distribution and the so-called structural equation (2), and as use is made of the one-to-one correspondence between the parameter values and the elements of a group of transformations acting on the sample space. For this new version of the problem the observed values x_{jl} are also normally distributed with mean μ_j and variance σ_j^2; however, this formulation provides us with more information about the generation of the observed values, which can be statistically evaluated.

2. Structural Analysis of a k-Sample Generalization of the Behrens-Fisher Problem

We now consider the problem in a generalized form in assuming that we have k different samples of observations x_{jl} $[l = 1, \ldots, n_k;$ $j = 1, \ldots, k]$ and that the corresponding error values e_{jl} are generated from an arbitrary absolutely continuous distribution with density function f. In Fraser (1968) it is shown how one can derive for models

[1] $\mathscr{R}^\nu$ denotes the ν-dimensional Euclidean space and 0 in (2) designates the action of the group.

of this type an induced distribution for the unknown parameters, which is conditioned on the observed data. This so-called " structural-distribution of the parameters " reflects the knowledge which we have about these quantities after the performance of the experiment. In our case we find that the structural distribution of the unknown parameter-vectors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ is specified by the density function

(3) $$g^*_{\mu,\sigma}(\mu_1, \ldots, \mu_k, \sigma_1, \ldots, \sigma_k : X)d\mu_1, \ldots, d\mu_k, d\sigma_1 \ldots, d\sigma_k =$$

$$c(X) \prod_{j=1}^{k} \left\{ \prod_{l=1}^{n_j} f\left(\frac{x_{jl}-\mu_j}{\sigma_j}\right) \cdot \frac{s_{x_j}^{n_j-1}}{\sigma_j^{n_j+1}} d\mu_j d\sigma_j \right\}$$

$$[n_j \geq 2\ ;\ \ k=1, 2, \ldots].$$

Here X denotes the set $\{x_{jl}\}$ of all observed values and $c(X)$ is the normalizing constant defined by the requirement that g^* is a density function with respect to the parameters. From (3) the distribution of any function of the parameters may be determined by an appropriate integration. For a linear combination of the location parameters $\delta = \sum_{j=1}^{k} p_j \mu_j$ with $p_k \neq 0$ for instance we obtain

(4) $$g^*_\delta(\delta : X)d\delta =$$

$$\left[| p_k |^{-1} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{0}^{\infty} \cdots \int_{0}^{\infty} g^*_{\mu,\sigma}\left(\mu_1, \ldots, \mu_{k-1}, \frac{1}{p_k}\left(\delta - \sum_{j=1}^{k-1} p_j \mu_j\right),\right.\right.$$

$$\left.\sigma_1, \ldots, \sigma_k : X)d\mu_1 \ldots d\mu_{k-1} d\sigma_1 \ldots d\sigma_k \right] d\delta.$$

For the special case $k=2$, $p_1=1$, $p_2=-1$ relation (4) provides us with a solution for the Behrens-Fisher problem for error variables with an arbitrary absolutely continuous distribution.

3. Special Results

Next we discuss some special results, which are obtained from the general formulae (3) and (4) by a specific choice of the density f.

3.1. Standard Normally Distributed Errors

If the error variables are standard normally distributed the structural distribution of the unknown parameters is given by

(5) $$g^*_{\mu,\sigma}(\boldsymbol{\mu}, \boldsymbol{\sigma} : X)d\boldsymbol{\mu} d\boldsymbol{\sigma} =$$

$$\prod_{j=1}^{k} \left(c_j \exp\left[-\frac{1}{2\sigma_j^2}\left(\sum_{l=1}^{n_j} (x_{jl}-\mu_j)\right)^2 \right] \frac{s_{x_j}^{n_j-1}}{\sigma_j^{n_j+1}} d\sigma_j d\mu_j \right)$$

with $$c_j = \sqrt{\frac{n_j}{2\pi}} \frac{(n_j-1)^{\frac{n_j-1}{2}}}{\Gamma\left(\frac{n_j-1}{2}\right) 2^{\frac{n_j-3}{2}}} \qquad [j=1, \ldots, k]$$

where Γ denotes the gamma-function.

This implies that the distribution of any linear combination $\delta = \sum_{j=1}^{k} p_j \mu_j$ of the means of the populations is characterized by the expression

$$(6) \qquad \delta = \sum_{j=1}^{k} p_j \mu_j = \sum_{j=1}^{k} p_j \bar{x}_j - \sum_{j=1}^{k} \frac{p_j s_{x_j}}{\sqrt{n_j}} t_{n_j - 1} \qquad [n_j \geq 2].$$

δ is thus the sum of relocated and rescaled t-variables. If in particular we choose $p_1 = 1$ and $p_2 = -1$ the question under investigation reduces to the original Behrens-Fisher problem and (6) becomes equivalent to (1), the solution proposed by R. A. Fisher. The method of structural inference thus leads to a new justification of this relation and to a derivation which is free of the inconsistencies of the fiducial mode of inference.

3.2. Errors with a Rectangular Distribution

If we want to compare the means μ_j of different rectangularly distributed populations with density function

$$(7) \qquad r(x : \mu_j, \sigma_j) = \frac{1}{2\sigma_j} \qquad [| x - \mu_j | \leq \sigma_j],$$

$$[\mu_j \varepsilon \mathscr{R}, \ \sigma_j > 0]$$

$$= 0 \qquad [| x - \mu_j | > \sigma_j],$$

where μ_j and σ_j are the unknown parameters we choose as density function for the error values

$$f(e) = r(e : 0, 1).$$

Substituting (7) into (3) and (4) we obtain

$$(8) \qquad g^*_{\mu,\sigma}(\boldsymbol{\mu}, \boldsymbol{\sigma} : X) d\boldsymbol{\mu} d\boldsymbol{\sigma} = \prod_{j=1}^{k} \left(\frac{n_j(n_j - 1)(x_{j(n_j)} - x_{j(1)})^{n_j - 1}}{2^{n_j} \sigma_j^{n_j + 1}} d\sigma_j d\mu_j \right)$$

$$[-\infty < \mu_j < \infty, \ \sigma_j \geq \max_{l=1}^{n_j} | x_{jl} - \mu_j | ; \ j = 1, \ldots, k]$$

$$= 0 \quad [\mu_j, \sigma_j \text{ otherwise}]$$

and

$$(9) \quad \delta = \sum_{j=1}^{k} p_j \mu_j = \sum_{j=1}^{k} \frac{p_j (x_{j(n_j)} + x_{j(1)})}{2} + \sum_{j=1}^{k} \frac{p_j (x_{j(n_j)} - x_{j(1)})(1 - | v_{n_j - 1} |)}{2 v_{n_j - 1}}$$

$$[n_j \geq 2]$$

where v_ν denotes a random variable with density function

$$(10) \qquad \psi(v_\nu) = \frac{\nu | v_\nu |^{\nu - 1}}{2} \qquad [-1 \leq v_\nu \leq 1, \ v_\nu \neq 0]$$

$$[\nu = 1, 2, 3, \ldots]$$

$$= 0 \qquad [v_\nu \text{ otherwise}].$$

As we see the structural density of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and of δ depends on the observed data only in terms of the largest value $x_{j(n_j)}$ and the smallest value $x_{j(1)}$ of each sample.

3.3. Exponentially Distributed Errors

If we want to compare the location parameters μ_j of different populations with exponential density

$$(11) \qquad u(x : \mu_j, \sigma_j) = \frac{1}{\sigma_j} \exp \left[-\frac{x - \mu_j}{\sigma_j} \right] \qquad [x \geq \mu_j]$$

$$[\mu_j \varepsilon \mathscr{R}, \ \sigma_j > 0]$$

$$= 0 \qquad [x < \mu_j]$$

where the μ_j's and σ_j's are the unknown parameters we choose as density function for the error values

$$f(e)=u(e{:}0,1).$$

Substituting (11) into (3), we find

(12) $g^*_{\mu,\sigma}(\mu,\sigma{:}X)d\boldsymbol{\mu}d\boldsymbol{\sigma}=$

$$\prod_{j=1}^{k}\left(\frac{(n_j)^{n_j}(\bar{x}_j-x_{j(1)})^{n_j-1}}{\Gamma(n_j-1)}\exp\left[\frac{-n_j(\bar{x}_j-\mu_j)}{\sigma_j}\right]\frac{d\sigma_j d\mu_j}{\sigma_j^{n_j+1}}\right)$$

$$[n_j\geq 2,\ \sigma_j>0,\ -\infty<\mu_j\leq x_{j(1)}]$$

and

$$\delta=\sum_{j=1}^{k}p_j\mu_j=\sum_{j=1}^{k}p_j\bar{x}_j-\sum_{j=1}^{k}p_j(\bar{x}_j-x_{j(1)})h_{n_j} \qquad [n_j\geq 2] \tag{13}$$

where h_ν denotes a random variable with density function

$$\Gamma(h_\nu)=\frac{\nu-1}{h_\nu^\nu} \qquad [h_\nu\geq 1] \tag{14}$$

$$[\nu=2,3,\ldots]$$

$$=0 \qquad [h_\nu \text{ otherwise}].$$

We note that the structural density of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and of δ depends only on the observed data in terms of the smallest value $x_{j(1)}$ and of the mean $\bar{x}_j$ of each sample.

Acknowledgement

This research has been supported in part by NRC Contract A-3011 (National Research Council of Canada). The authors thank the referee for his valuable comments.

References

Banerji, S. K. (1960). "Approximate confidence interval for linear functions of means of K populations, when the population variances are not equal." *Sankhyā*, 22, 357–358.

Behrens, W. V. (1929). "Ein Beitrag zur Fehlerberechnung bei wenigen Beobachtungen." *Landwirtschaftliche Jahresberichte*, 68, 807–837.

Chand, U. (1950). "Distributions related to comparison of two means and two regression coefficients." *Ann. Math. Statist.*, 21, 507–522.

Chernoff, H. (1949). "Asymptotic studentization in testing of hypotheses." *Ann. Math. Statist.*, 20, 268–278.

Fisher, R. A. (1935). "The fiducial argument in statistical inference." *Ann. of Eugenics*, 6, 391–398.

Fraser, D. A. S. (1968). *The Structure of Inference.* J. Wiley, New York.

Fraser, D. A. S., and Streit, F. (1969). "Some applications of the method of structural inference involving several unknown parameters." Abstract, *Ann. Math. Statist.*, 40, 1871.

Scheffé, H. (1943). "On solutions of the Behrens-Fisher problem, based on the t-distribution." *Ann. Math. Statist.*, 14, 35–44.

Scheffé, H. (1944). "A note on the Behrens-Fisher problem." *Ann. Math. Statist.*, 15, 430–432.

Scheffé, H. (1970). "Practical solutions of the Behrens-Fisher problem." *J. Amer. Statist. Assoc.*, 65, 1501–1508.

Srivastava, M. S. (1966). "On the asymptotic theory of sequential confidence intervals." Abstract. *Ann. Math. Statist.*, 37, 313.

Srivastava, M. S. (1970). "On a sequential analogue of the Behrens-Fisher problem." *J. Roy. Statist. Soc., Ser. B*, 32, 144–148.

Tukey, J. W. (1957). "Some examples with fiducial relevance." *Ann. Math. Statist.*, 28, 687–695.

Wallace, D. L. (1958). "Asymptotic approximations to distributions." *Ann. Math. Statist.*, 29, 635–654.

Wilks, S. (1940). "On the problem of two samples from normal populations with unequal variances." *Ann. Math. Statist.*, 11, 475–476.

AUSTRALASIAN MEDICAL PUBLISHING CO. LTD.
71-79 ARUNDEL ST., GLEBE, N.S.W., 2037