The Stata Journal (2009)
9, Number 4, pp. 593–598

Implementation of a new solution to the multivariate Behrens–Fisher problem

Ivan Žežula
Šafárik University
Institute of Mathematics
Slovak Republic
ivan.zezula@upjs.sk

Abstract. Krishnamoorthy and Yu (2004, *Statistics and Probability Letters* 66: 161–169) published a new approximate solution to the multivariate Behrens–Fisher problem. It is a modification of Nel and Van der Merwe's (1986, *Communications in Statistics, Theory and Methods* 15: 3719–3735) test. The test is invariant and identical to Welch's test for one-dimensional data. In this article, I describe an implementation of the test in Stata. The `hotelmnm` command allows you to perform the test easily and returns computed values for possible further computations.

Editors' note: The `mvtest means` command introduced in Stata 11 produces the same test as the `hotelmnm` command introduced here. Use the `heterogenous` option of `mvtest means` to obtain the test; see [MV] **mvtest means**. The `hotelmnm` command will still be of interest to those using prior versions of Stata.

Keywords: st0180, hotelmnm, multivariate Behrens–Fisher problem, Nel and Van der Merwe's test, Welch's test

1 Introduction

In its univariate form, the Behrens–Fisher problem is the test of the difference between the means of two normally distributed populations when the variances of the populations are not necessarily equal. Because an exact analytic solution is computationally intractable, different approximate solutions are used. The most popular is Welch's test. This test is provided in Stata through `ttest` using the `unequal` option.

Multivariate generalization of the t test, testing the equality of two vector means, is Hotelling's test. As in the univariate case, Hotelling's test assumes that the variance matrices of the two groups are equal. Stata provides the `hotelling` procedure for this case. In this article, I provide a modification of `hotelling`, called `hotelmnm`, that can be used when the variance–covariance matrices of the group-specific outcome means may be unequal.

Let's introduce some notation. We assume two independent p-variate random samples from normal distributions

$$X_1, \ldots, X_m \sim \mathcal{N}_p(\mu_1, \Sigma_1)$$

and

$$Y_1, \ldots, Y_n \sim \mathcal{N}_p(\mu_2, \Sigma_2)$$

Thus every X_i and Y_j is a vector of length p (there are p different characteristics measured on one object). Mean values of the two populations are μ_1 and μ_2, and their variance matrices are Σ_1 and Σ_2. Sample sizes m and n may be different. We want to test the null hypothesis $H_0: \ \mu_1 = \mu_2$ against $H_a: \ \mu_1 \neq \mu_2$ when $\Sigma_1 \neq \Sigma_2$.

Test of this hypothesis is called the multivariate Behrens–Fisher problem.

The situation in a general p-variate case is more complicated than in the univariate one. First, we have to realize that H_0 is equivalent to

$$H_A: \ A\mu_1 = A\mu_2$$

for any nonsingular matrix A. That is why it is reasonable to request that the test of H_0 should be independent of any data transformation by a nonsingular matrix A. This property, if present, is called the invariance of the test. The exact solution is again known but computationally intractable (see Nel, Van der Merwe, and Moser [1990]). Other published solutions include the following:

- Solution of Scheffé (1943) and Bennett (1950): uses adjusted paired differences. It is an exact solution but has little power because it does not use the information from the samples well. It is analogical to using a paired t test in place of a two-sample t test.
- Approximate solutions:
  - Kim (1992), Nel and Van der Merwe (1986): not invariant.
  - James (1954), Yao (1965), Johansen (1980): varying quality of the approximation. James' solution is not used any more. It is difficult to predict which one of the latter two will be better in a specific situation.

As a result, none of the solutions is commonly accepted. A new solution appeared recently: Krishnamoorthy and Yu (2004). It is a modified version of Nel and Van der Merwe's (1986) test. The merit of the authors is especially in the derivation of the correct number of degrees of freedom, when they noticed incorrectness in Nel and Van der Merwe's derivation. This solution is invariant, and it seems to have a stable test level close to the chosen α. It coincides with Welch's test for $p = 1$. In my opinion, it has a big chance to become the most popular solution of the problem.

2 The principle

Notice that for $\Sigma_1 = \Sigma_2$, we can write Hotelling's test statistic in the following way:

$$T^2 = \left(\overline{X} - \overline{Y}\right)' \left(\frac{1}{m}S + \frac{1}{n}S\right)^{-1} \left(\overline{X} - \overline{Y}\right)$$

where S is the pooled variance matrix estimator.

Let's denote

$$S_1 = \frac{1}{m-1}\sum_{i=1}^{m}\left(X_i - \overline{X}\right)\left(X_i - \overline{X}\right)' \quad \text{and} \quad S_2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)\left(Y_i - \overline{Y}\right)'$$

both sample variance matrices, and

$$\Sigma = \frac{1}{m}\Sigma_1 + \frac{1}{n}\Sigma_2 \quad \text{and} \quad S = \frac{1}{m}S_1 + \frac{1}{n}S_2$$

For $\Sigma_1 \neq \Sigma_2$ it is natural to define

$$T^2 = \left(\overline{X} - \overline{Y}\right)'\left(\frac{1}{m}S_1 + \frac{1}{n}S_2\right)^{-1}\left(\overline{X} - \overline{Y}\right)$$

This test statistic has approximately Hotelling's distribution, as it is shown in Krishnamoorthy and Yu (2004). The corresponding number of degrees of freedom is

$$f^* = \frac{\operatorname{Tr}\left(I_p^2\right) + \left(\operatorname{Tr} I_p\right)^2}{\frac{1}{m-1}\left\{\operatorname{Tr}\left(V_1^2\right) + \left(\operatorname{Tr} V_1\right)^2\right\} + \frac{1}{n-1}\left\{\operatorname{Tr}\left(V_2^2\right) + \left(\operatorname{Tr} V_2\right)^2\right\}}$$

where

$$V_1 = \Sigma^{-\frac{1}{2}}\frac{1}{m}\Sigma_1\Sigma^{-\frac{1}{2}} \quad \text{and} \quad V_2 = \Sigma^{-\frac{1}{2}}\frac{1}{n}\Sigma_2\Sigma^{-\frac{1}{2}}$$

Reasonable estimators of V_1 and V_2 are

$$\widehat{V}_1 = S^{-\frac{1}{2}}\frac{1}{m}S_1S^{-\frac{1}{2}} \quad \text{and} \quad \widehat{V}_2 = S^{-\frac{1}{2}}\frac{1}{n}S_2S^{-\frac{1}{2}}$$

As a consequence, a reasonable estimator of f^* is

$$d = \frac{p\,(p+1)}{\frac{1}{m^2(m-1)}\left\{\operatorname{Tr}\left(S_1S^{-1}\right)^2 + \left(\operatorname{Tr} S_1S^{-1}\right)^2\right\} + \frac{1}{n^2(n-1)}\left\{\operatorname{Tr}\left(S_2S^{-1}\right)^2 + \left(\operatorname{Tr} S_2S^{-1}\right)^2\right\}}$$

or

$$\frac{1}{d} = \frac{1}{p\,(p+1)}\left[\frac{1}{m^2(m-1)}\left\{\operatorname{Tr}\left(S_1S^{-1}\right)^2 + \left(\operatorname{Tr} S_1S^{-1}\right)^2\right\} + \right. \\ \left. + \frac{1}{n^2(n-1)}\left\{\operatorname{Tr}\left(S_2S^{-1}\right)^2 + \left(\operatorname{Tr} S_2S^{-1}\right)^2\right\}\right]$$

Thus

$$T^2 \overset{H_0}{\approx} T^2_{d,p}$$

or

$$\frac{d-p+1}{dp}\,T^2 \overset{H_0}{\approx} F_{p,d-p+1}$$

It is easy to see that for $p = 1$, d is equal to Welch's number of approximate degrees of freedom. Moreover, Krishnamoorthy and Yu (2004) showed that even for $p > 1$, d is bound in the same way as in the one-dimensional case:

$$\min(m-1, n-1) \leq d \leq m+n-2$$

d being close to the upper bound tells us that the two variance matrices are (almost) equal. The closer d is to the lower bound, the bigger the discrepancy is between them. The lower bound is attained only if one of S_1, S_2 is a zero matrix.

3 The hotelmnm command

3.1 Syntax

The syntax of the `hotelmnm` command is

`hotelmnm` *varlist* [*if*] [*in*] `, by(`*groupvar*`)` [`notable`]

The `if` or `in` condition can restrict input data (observations).

3.2 Options

`by(`*groupvar*`)` is required. It specifies the name of the grouping variable. *groupvar* must contain exactly two different values.

`notable` suppresses the table of basic descriptive statistics in the output.

3.3 Saved results

`hotelmnm` saves the following in `r()`:

Scalars	
`r(k)`	number of variables
`r(N1)`	number of observations in the first group
`r(N2)`	number of observations in the second group
`r(df)`	number of approximate degrees of freedom
`r(T2)`	value of T^2 statistic
Matrices	
`r(X)`	averages of both groups
`r(S1)`	sample variance matrix of the first group
`r(S2)`	sample variance matrix of the second group

All these values can be used for further computations.

4 Example 1

```
. sysuse auto
(1978 Automobile Data)

. hotelmnm mpg headroom, by(foreign)
--------------------------------------------------------------------------------
-> foreign = Domestic

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         mpg |        52    19.82692    4.743297         12         34
    headroom |        52    3.153846    .9157578        1.5          5
--------------------------------------------------------------------------------
-> foreign = Foreign

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         mpg |        22    24.77273    6.611187         14         41
    headroom |        22    2.613636    .4862837        1.5        3.5

2-group approximate Hotelling^s T-squared with unequal variances = 18.402703
F test statistic: ((44.102882-2+1)/(44.102882)(2)) x 18.402703 = 8.9927177

H0: Vectors of means are equal for the two groups
              F(2,43.102882) =    8.9927
       Prob(F > F(2,43.102882)) =  0.000544
```

5 Example 2

```
. hotelmnm mpg headroom trunk if price<5000, by(foreign) notable

2-group approximate Hotelling^s T-squared with unequal variances = 13.209688
F test statistic: ((12.765519-3+1)/(12.765519)(3)) x 13.209688 = 3.7133664

H0: Vectors of means are equal for the two groups
              F(3,10.765519) =    3.7134
       Prob(F > F(3,10.765519)) =  0.046656

. display " n1 = " r(N1) ", n2 = " r(N2) ", dimension = " r(k)
 n1 = 29, n2 = 8, dimension = 3

. display " degrees of freedom = " r(df) ", T^2 = " r(T2)
 degrees of freedom = 12.765519, T^2 = 13.209688

. matrix list r(X)

r(X)[2,3]
              mpg   headroom      trunk
group1  22.137931  3.0689655  12.517241
group2     28.875       2.75     10.625

. matrix list r(S1)

symmetric r(S1)[3,3]
                 mpg    headroom       trunk
     mpg   19.051724
headroom  -2.2777094   .94150246
   trunk  -7.8953202    2.945197   13.544335

. matrix list r(S2)

symmetric r(S2)[3,3]
                 mpg    headroom       trunk
     mpg   23.839286
headroom  -.60714286   .21428571
   trunk  -9.9107143   .39285714   12.839286
```

6 References

Bennett, B. M. 1950. Note on a solution of the generalized Behrens–Fisher problem. *Annals of the Institute of Statistical Mathematics* 2: 87–90.

James, G. S. 1954. Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika* 41: 19–43.

Johansen, S. 1980. The Welch–James approximation to the distribution of the residual sum of squares in a weighted linear regression. *Biometrika* 67: 85–92.

Kim, S.-J. 1992. A practical solution to the multivariate Behrens–Fisher problem. *Biometrika* 79: 171–176.

Krishnamoorthy, K., and J. Yu. 2004. Modified Nel and Van der Merwe test for the multivariate Behrens–Fisher problem. *Statistics and Probability Letters* 66: 161–169.

Nel, D. G., and C. A. Van der Merwe. 1986. A solution to the multivariate Behrens–Fisher problem. *Communications in Statistics, Theory and Methods* 15: 3719–3735.

Nel, D. G., C. A. Van der Merwe, and B. K. Moser. 1990. The exact distributions of the univariate and multivariate Behrens–Fisher statistics with a comparison of several solutions in the univariate case. *Communications in Statistics, Theory and Methods* 19: 279–298.

Scheffé, H. 1943. On solutions of the Behrens–Fisher problem, based on the *t*-distribution. *Annals of Mathematical Statistics* 14: 35–44.

Yao, Y. 1965. An approximate degrees of freedom solution to the multivariate Behrens–Fisher problem. *Biometrika* 52: 139–147.

About the author

Ivan Žežula is a statistician at Šafárik University in Košice, Slovakia. He is involved in many applications of statistics, especially in medicine. He considers Stata to be a handy tool for both teaching and research, and he occasionally writes his own procedures.