Inferences for Ratios of Normal Means

by Gemechis Dilba, Frank Schaarschmidt, Ludwig A. Hothorn

Introduction

Inferences concerning ratios of means of normally distributed random variables or ratios of regression coefficients arise in a variety of problems in biomedical research. For example, in tests for non-inferiority of one or more experimental treatments against a positive control, it is often easier to define and also to interpret the non-inferiority margin as percentage changes (or fraction retained compared to the mean of the control group). In bioassay problems, one is also interested in ratios of regression coefficients, for instance in parallel line or slope ratio assays. Our aim here is to introduce an R extension package called **mratios** which can perform inferences about one or more such ratio parameters in the general linear model. For two-sample problems, the package is capable of constructing Fieller confidence intervals and performing the related tests when the group variances are assumed homogeneous or heterogeneous. In simultaneous inferences for multiple ratios, the package can (i) perform multiple tests, (ii) construct simultaneous confidence intervals using a variety of techniques, and (iii) calculate the sample sizes required for many-to-one comparisons in simultaneous tests for non-inferiority (or superiority) based on relative margins. We demonstrate the functionality of the package by using several data examples.

Two-sample Problem

The two-sample problem is one of the standard methods routinely used in practice. Here the interest is in comparing the means of two independent normally distributed random variables in terms of the ratio of their means. This can be accomplished by using the `t.test.ratio` function. If the variances are homogeneous, this function performs a ratio formatted t-test (also known as Sasabuchi test) and computes Fieller's confidence interval. If variance homogeneity is not tenable (the default), the test proposed by Tamhane and Logan (2004) is performed using Satterthwaite adjusted degrees of freedom. For confidence interval estimation under variance heterogeneity, Satterthwaite degrees of freedom depends on the unknown ratio. To circumvent this problem, we plug in the maximum likelihood estimate of the ratio (i.e., ratio of sample means) in the approximate expression for the number of degrees of freedom.

Example 1. Consider the mutagenicity assay data described in the **mratios** package. A first step in the analysis of the data could be to test whether the active control (cyclophosphamide at dose 25mg/kg) results in a significantly higher number of mutations than the vehicle control. The data appear to be heteroscedastic, and therefore we use the unequal variances option (the default) to compare the two treatments.

```
> library("mratios")
> data("Mutagenicity")
> muta2 <- subset(Mutagenicity, Treatment ==
+     "Vehicle" | Treatment == "Cyclo25")
> t.test.ratio(MN ~ Treatment, data = muta2,
+     alternative = "greater")

    Ratio t-test for unequal variances

data:  Cyclo25 and Vehicle
t = 5.0071, df = 3.07, p-value = 0.0073
 alternative hypothesis: true ratio of means
    is greater than 1
95 percent confidence interval:
 5.110079      Inf
sample estimates:
   mean Cyclo25    mean Vehicle
      25.000000        2.571429
Cyclo25/Vehicle
       9.722222
```

Note that when testing a ratio of means against 1, the p-value computed by the `t.test.ratio` function is exactly the same as that computed by `t.test` when testing the difference of means against 0.

Simultaneous Inferences

In this section we consider inferential problems involving one or more ratio parameters. The basic distribution underlying the analyses is the multivariate t-distribution. Under the assumption of normality and homogeneous variance for the error terms, the joint distribution of the test statistics associated with the various contrasts of interest follows a multivariate t-distribution. For the computation of the related multivariate t probabilities and equi-coordinate critical points, we refer to Hothorn et al. (2001).

Multiple Tests

Assume a normal one-way ANOVA model with homogeneous variances. The interest is in simultaneous tests for several ratios of linear combinations of the treatment means. Such tests for ratio hypotheses (ratios of normal means) appear, for example, in tests for non-inferiority (or superiority) of several experimental treatments compared to a control

(placebo). These are so called many-to-one comparisons. In the R-function `simtest.ratio`, most of the routinely used multiple comparison procedures [e.g., many-to-one (Dunnett type), all pairs (Tukey type), sequence (successive comparisons of ordered treatment effects)] are implemented in the context of ratio hypotheses. In general, the function also allows for any user-defined contrast matrices.

Let $\gamma_j = \boldsymbol{c}_j'\boldsymbol{\mu}/\boldsymbol{d}_j'\boldsymbol{\mu},\ j = 1,\ldots,r$ denote the ratios of interest, where $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_k)'$ is a vector of the treatment means, $\boldsymbol{c}_j$ and $\boldsymbol{d}_j$ are known vectors of real constants each of dimension $k \times 1$, and r is the number of ratios. To specify the ratios, we define two contrast matrices, namely, numerator and denominator contrast matrices. The numerator contrast matrix is a matrix whose row vectors are $\boldsymbol{c}_1',\ldots,\boldsymbol{c}_r'$, and the denominator contrast matrix is a matrix whose row vectors are $\boldsymbol{d}_1',\ldots,\boldsymbol{d}_r'$. Therefore, the dimensions of both the numerator and denominator contrast matrices are each $r \times k$. Further, let $(\psi_1,\ldots,\psi_r)'$ denote the set of margins against which we test the r ratios. Then, for example, for one-sided upper-tailed alternative hypotheses, the hypotheses of interest are $H_{0j} : \gamma_j \leq \psi_j$ versus $H_{1j} : \gamma_j > \psi_j$, $j = 1,\ldots,r$.

Given a data frame containing the observations, the contrast matrices, the vector of margins, and the family-wise type I error rate, the function `simtest.ratio` calculates the point estimates of the ratios, the test statistics, the raw p-values and the multiplicity adjusted p-values. The adjusted p-values are computed by adapting the results of Westfall et al. (1999) for ratio hypotheses and general contrasts.

In general, note that the function `simtest.ratio` allows for varying margins for the set of comparisons. This can be quite appealing, for example, in test problems involving a mixture of non-inferiority and superiority hypotheses.

Example 2. Bauer et al. (1998) analyzed data from a multi-dose experiment including a positive control and placebo. In the experiment, patients with chronic stable angina pectoris were randomized to five treatment arms (placebo, three doses of a new compound, and an active control). The response variable is the difference in the duration of an exercise test before and after treatment. Now, due to the unavailability of the original data values, we randomly generated independent samples (from a normal distribution) that satisfy the summary statistics given in Table II of Bauer et al. (1998). This data set is available in the **mratios** package. The interest is in simultaneous tests for non-inferiority of the three doses versus the active control by including the placebo. Following Pigeot et al. (2003), the hypotheses can succinctly be formulated as $H_{0i} : (\mu_j - \mu_2)/(\mu_1 - \mu_2) \leq 0.9$ versus $H_{1i} : (\mu_j - \mu_2)/(\mu_1 - \mu_2) > 0.9$, $j = 3,4,5$, where μ_i, $i = 1,\ldots,5$ denote the means for the active control, placebo, dose 50, dose 100 and dose 150, consecutively. In this example, the non-inferiority margins are all set to 0.9.

```
> data("AP")
> NC <- rbind(N1 = c(0, -1, 1, 0, 0),
+             N2 = c(0, -1, 0, 1, 0),
+             N3 = c(0, -1, 0, 0, 1))
> DC <- rbind(D1 = c(1, -1, 0, 0, 0),
+             D2 = c(1, -1, 0, 0, 0),
+             D3 = c(1, -1, 0, 0, 0))
> ap.test <- simtest.ratio(pre_post ~
+     treatment, data = AP, Num.Contrast = NC,
+     Den.Contrast = DC, Margin.vec = 0.9,
+     alternative = "greater")
> ap.test

Alternative hypotheses: Ratios greater
 than margins

      margin estimate statistic
N1/D1    0.9    5.306    2.9812
N2/D2    0.9    4.878    2.7152
N3/D3    0.9    1.969    0.7236
      p.value.raw p.value.adj
N1/D1    0.001554    0.004429
N2/D2    0.003505    0.009799
N3/D3    0.234952    0.451045
```

By using the command `summary(ap.test)`, one can get further information — for example, the correlation matrix under the null hypotheses and the critical point (equi-coordinate percentage point of the multivariate t-distribution).

Simultaneous Confidence Intervals

Unlike in multiple testing, in simultaneous estimation of the ratios $\gamma_j = \boldsymbol{c}_j'\boldsymbol{\mu}/\boldsymbol{d}_j'\boldsymbol{\mu},\ j = 1,\ldots,r$, the joint distribution of the associated t-statistics follows a multivariate t-distribution with a correlation matrix that depends on the unknown ratios. This means that the critical points that are required for confidence interval construction depend on these unknown parameters. There are various methods of dealing with this problem. They are (i) using the unadjusted intervals (Fieller confidence intervals without multiplicity adjustments); (ii) Bonferroni (Fieller intervals with simple Bonferroni adjustments); (iii) a method called MtI which consists of replacing the unknown correlation matrix of the multivariate t-distribution by an identity matrix of the same dimension according to Sidak and Slepian inequalities (Hochberg and Tamhane, 1987) for two- and one-sided confidence intervals, respectively; and (iv) plug-in (plugging the maximum likelihood estimates of the ratios into the unknown correlation matrix). The latter method is known to have good simultaneous coverage probabilities and hence it is set as a default method in the R functions to be introduced. For details regarding these methodologies, we refer to Dilba et al. (2006a).

The `sci.ratio` function is used to construct simultaneous CIs for ratios of linear combinations of treatment means in a one-way ANOVA model. Several standard contrast types (e.g., Dunnett, Tukey, sequence, and many others) are implemented in this function. The default contrast is many-to-one comparisons (Dunnett type) with the mean of the first level of the factor (in alpha-numeric order) taken as the denominator of the ratios. In addition, this function has an option for user-defined contrast matrices.

Example 3. Recall the data from the multi-dose experiment in Example 2 above. Now, suppose that the interest is to calculate simultaneous lower 95% confidence limits for the ratios of the three doses and the active control to the placebo. Noting that placebo is the second level in the alpha-numeric order of the treatments, we use the following R code to calculate the limits.

```
> ap.sci <- sci.ratio(pre_post ~
+     treatment, data = AP, type = "Dunnett",
+     base = 2, alternative = "greater",
+     method = "MtI")
```

The graph of the confidence intervals can be obtained by applying the plot function to the object in which the confidence interval estimates are stored, see Figure 1.

```
> plot(ap.sci)
```

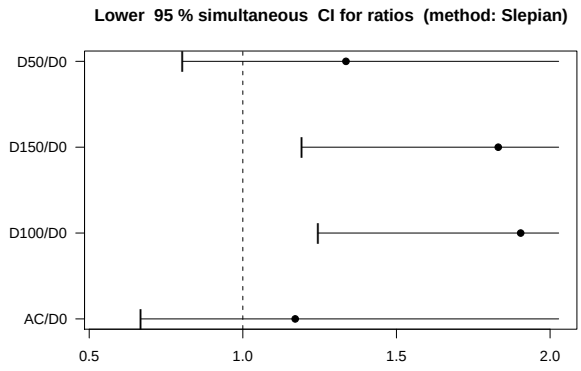

Figure 1: Graphical visualization of the `ap.sci` object.

The `sci.ratio.gen` function is a more general function that can construct simultaneous confidence intervals for ratios of linear combinations of coefficients in the general linear model. For this function, it is necessary to specify the vector of responses, the design matrix, and the numerator and denominator contrast matrices.

Example 4. Consider the problem of simultaneously estimating relative potencies in a multiple slope ratio assay. Jensen (1989) describes an experiment in which three preparations are compared to a control. The response variable (Y) is pantothenic acid content of plant tissues. The model is $Y_{ij} = \alpha + \beta_i X_{ij} + \epsilon_{ij}$, $i = 0,1,2,3$; $j = 1,\ldots,n_i$, where the X_{ij}s are the dose levels and $i = 0$ refers to the control group. The vector of regression coefficients is $(\alpha, \beta_0, \beta_1, \beta_2, \beta_3)'$. Now using the data in Table 5 of Jensen (1989), the interest is to construct simultaneous CIs for β_i/β_0, $i = 1,2,3$. The function `sci.ratio.gen` needs the response vector Y and the design matrix X as an input.

```
> data(SRAssay)
> Y <- SRAssay[, "Response"]
> X <- model.matrix(Response ~ Treatment:Dose,
+     data = SRAssay)
> NC <- matrix(c(0, 0, 1, 0, 0,
+                0, 0, 0, 1, 0,
+                0, 0, 0, 0, 1),
+       nrow = 3, byrow = TRUE)
> DC <- matrix(c(0, 1, 0, 0, 0,
+                0, 1, 0, 0, 0,
+                0, 1, 0, 0, 0),
+       nrow = 3, byrow = TRUE)
> s.ratio <- sci.ratio.gen(Y, X,
+     Num.Contrast = NC, Den.Contrast = DC)
> s.ratio

Two-sided 95 % simultaneous confidence
 intervals for ratios:


   estimate  lower  upper
C1   1.1217 1.0526 1.1964
C2   0.7193 0.6603 0.7805
C3   0.7537 0.6942 0.8157
```

Using the command `summary(s.ratio)`, one can also get further details regarding the fitted regression model, the contrast matrices and an estimate of the correlation matrix (when the plug-in method is used). The estimate of the correlation matrix used for critical point calculation can also be obtained as

```
> s.ratio[["CorrMat.est"]]

          [,1]      [,2]      [,3]
[1,] 1.0000000 0.4083451 0.4260802
[2,] 0.4083451 1.0000000 0.3767098
[3,] 0.4260802 0.3767098 1.0000000
```

Note that by choosing the option for method as 'Sidak', one gets the results reported by Jensen (1989).

Before closing this section, we give two important remarks.

i) According to the Slepian inequality (Hochberg and Tamhane, 1987), it is appropriate to use the MtI method for estimating one-sided simultaneous confidence limits only when all the elements of the correlation matrix are non-negative. Therefore, if some of the (estimated) correlations are negative, `sci.ratio` and `sci.ratio.gen` functions report a warning message about the inappropriateness of the MtI method.

ii) In simultaneous CI estimation (using either `sci.ratio` or `simtest.ratio.gen`), one may encounter the case where some of the contrasts in the

denominators of the ratios are not significantly different from zero. In this situation, NSD (standing for "non-significant denominator") will be printed. For instance, in Example 2 above, since there is no significant difference between the placebo and the active control, one gets NSD in constructing the related simultaneous CIs for the three ratios.

Sample Size Calculation

Consider the design of a special problem in simultaneous comparison of $m \geq 2$ treatments with a control for non-inferiority (or superiority), where the margins are expressed as a percentage of the mean of the control group. For sample size calculation, we implement a method based on normal approximation to the exact method which involves inversion of a univariate (multivariate) non-central t-distribution (see Dilba et al. (2006b) for details on the exact method). Given the number of comparisons (`m`), the non-inferiority (superiority) margin (`rho`), the power (`Power`), the coefficient of variation of the control group (`CV0`), the percentage (of the mean of the control group) to be detected (`rho.star`), the family-wise type-I error rate (`alpha`), and the kind of power to be controlled (by default minimal power), the function `n.ratio` calculates the sample size required in a balanced design.

Example 5. Suppose that we have a response variable where large response values indicate better treatment benefits. The following R code calculates the sample size required per treatment in designing a non-inferiority trial with four treatment arms (including the control).

```
> n.ratio(m = 3, rho = 0.7, Power = 0.8,
+     CV0 = 0.5, rho.star = 0.95,
+     alpha = 0.05, Min.power = TRUE)

 Number of observations per treatment =  52
         Total number of observations = 208
```

If the aim is to control the complete power, we set `Min.power` to `FALSE`.

For the two-sample design ($m = 1$), the sample sizes required in the non-inferiority trials discussed by Laster and Johnson (2003) can be calculated as a special case.

Remarks

We conclude by giving some general remarks regarding the four basic functions in the **mratios** package.

- In two-sample ratio problems with homogeneous variances, `t.test.ratio` is a special case of `simtest.ratio` and `sci.ratio`.
- The `simtest.ratio` function with all the elements of the vector of margins equal to 1 gives the same result as the analysis based on the difference of treatment means. Thus, the difference-based test is a special case of the ratio-based test with the thresholds set to 1.
- The `sci.ratio` function is a special case of `sci.ratio.gen` for the one-way layout.

Bibliography

P. Bauer, J. Röhmel, W. Maurer and L.A. Hothorn. Testing strategies in multi-dose experiments including active control. *Statistics in Medicine*, 17: 2133-2146, 1998. 21

G. Dilba, F. Bretz and V. Guiard. Simultaneous confidence sets and confidence intervals for multiple ratios. *Journal of Statistical Planning and Inference*, 136:2640–2658, 2006a. 21

G. Dilba, F. Bretz, L.A. Hothorn and V. Guiard. Power and sample size computations in simultaneous tests for non-inferiority based on relative margins. *Statistics in Medicine*, 25:1131–1147, 2006b. 23

J. Hochberg and A. Tamhane. *Multiple comparison procedures*. Wiley, New York, 1987, p366. 21, 22

T. Hothorn, F. Bretz and A. Genz. On Multivariate t and Gauss Probabilities. *R News*, 1(2):27–29, 2001. 20

G. R. Jensen. Joint confidence sets in multiple dilution assays. *Biometrical Journal*, 31:841–853, 1989. 22

L. L. Laster and M. F. Johnson. Non-inferiority trials: the 'at least as good as' criterion. *Statistics in Medicine*, 22:187–200, 2003.

I. Pigeot, J. Schäfer, J. Röhmel and D. Hauschke. Assessing non-inferiority of a new treatment in a three-arm clinical trial including a placebo. *Statistics in Medicine*, 22:883–899, 2003. 21

A. C. Tamhane and B. R. Logan. Finding the maximum safe dose level for heteroscedastic data. *Journal of Biopharmaceutical Statistics*, 14:843–856, 2004. 20

P. H. Westfall, R. D. Tobias, D. Rom, R. D. Wolfinger and Y. Hochberg. Multiple comparisons and multiple tests using the SAS system. Cary, NC, SAS Institute Inc,1999, 65–81.

21

G. Dilba, F. Schaarschmidt, L. A. Hothorn
Institute of Biostatistics
Faculty of Natural Sciences
Leibniz University of Hannover, Germany

`dilba@biostat.uni-hannover.de`
`schaarschmidt@biostat.uni-hannover.de`
`hothorn@biostat.uni-hannover.de`