



Research Report
No. 75

Combining multiple imputation t , χ^2 , and F inferences

Tim Hesterberg

Draft, Last revised
May 7, 1998

Acknowledgments: This work was supported by NIH 2R44CA65147-02.

MathSoft, Inc.
1700 Westlake Ave. N, Suite 500
Seattle, WA 98109-9891, USA
Tel: (206) 283-8802
FAX: (206) 283-6310

E-mail: timh@statsci.com

Combining multiple imputation t , χ^2 , and F inferences

Tim C. Hesterberg

August 26, 1998

Abstract

We discuss rules for combining inferences from multiple imputations when complete-data inferences would be based on t -distributions rather than normal distribution, or F -distributions rather than χ^2 distributions. Standard errors are obtained based on a distinction between the squared standard error and the actual variance of a t -distribution. Degrees of freedom are based on the coefficient of variation of a squared standard error, and combine the simulation error from using a finite number of imputations and the degrees of freedom in the original problem, adjusted for the estimated loss of information due to missing data.

We extend these ideas to situations where complete-data inferences would be based on χ^2 - and F -distributions, or are based on p -values or χ^2 (Wald) or F statistics.

We conclude with a discussion about appropriate calculations for regression summaries.

This is work in progress, and comments are welcomed.

Key Words: Missing data; Multiple imputations; Incomplete data.

1 Introduction

For an introduction to statistical analysis using multiple imputations, see Schafer (1997) (referred to as S97 in the sequel). We use notation from S97, in particular Sections 4.3.2 and 4.3.3 (S432 and S433 in the sequel) S432 is based on (Rubin (1987), Chapter 3).

S432 provides rules for combining inferences across multiple imputations which are appropriate if complete-data inferences (where no missing data is present) would be based on normal distributions; e.g. confidence intervals in the absence of missing data would be of the form

$$\hat{Q} \pm z_{\alpha/2} \sqrt{U}$$

where \hat{Q} is a parameter estimate, U is its variance (or a very accurate estimate), and z_a is the $1 - a$ quantile of a standard normal distribution. We focus on combining inferences when complete-data inferences would instead be based on t -distributions,

$$\hat{Q} \pm t_{\alpha/2, \nu} \sqrt{U}$$

where ν is the degrees of freedom and U is an estimate of the variance.

Similarly, S433 provides rules for combining inferences when complete-data inferences would be based on χ^2 distributions; we provide alternative versions of those rules, and extend those rules to situations where complete-data inferences would be based on F distributions.

We begin in Section 2 with a review of rules for normal-based problems, and discuss rules for t -based problems in Section 3. In Section 4 we review rules for χ^2 -based problems, and generalize those rules. In Section 5 we extend those rules to F -distributions. Section 6 relates to combining F -statistics, when only F -statistics (not the parameter estimates and covariances matrices that were used to obtain the statistics) are available. In Section 7 we discuss the construction of common summary statistics from a linear model.

2 Combining normal-based inferences

We begin by reviewing the rules from S432 for combining inferences in normal problems. The observed data are Y_{obs} , and the missing data are replaced by one of m sets of imputations $Y_{mis}^{(t)}$, $t = 1, \dots, m$. Let

$$\hat{Q}^{(t)} = \hat{Q}(Y_{obs}, Y_{mis}^{(t)})$$

and

$$U^{(t)} = U(Y_{obs}, Y_{mis}^{(t)})$$

be the point and variance estimates using the t th set of imputed data, $t = 1, \dots, m$.

The multiple-imputation point estimate for Q is the average of the complete-data point estimates,

$$\bar{Q} = \frac{1}{m} \sum_{t=1}^m \hat{Q}^{(t)}. \quad (1)$$

The variance estimate associated with \bar{Q} has two components. The *within-imputation variance* is the average of the complete-data variance estimates,

$$\bar{U} = \frac{1}{m} \sum_{t=1}^m U^{(t)}. \quad (2)$$

The *between-imputation variance* is the variance of the complete-data point estimates,

$$B = \frac{1}{m-1} \sum_{t=1}^m (\hat{Q}^{(t)} - \bar{Q})^2. \quad (3)$$

The *total variance* is defined as

$$T_{z1} = \bar{U} + (1 + m^{-1})B, \quad (4)$$

and inferences are based on the approximation

$$(Q - \bar{Q})/\sqrt{T_1} \sim t_{\nu_1} \quad (5)$$

where the degrees of freedom are

$$\nu_{z1} = (m - 1) \left[1 + \frac{\overline{U}}{(1 + m^{-1})B} \right]^2. \quad (6)$$

In the sequel we write T for a generic estimate of total variation, and use subscripts such as T_{z1} to denote specific estimates, where the z indicates an estimate for normal-based problems; t , c , and f will indicate t , χ^2 , and F -based problems, respectively. Similarly, ν_T will indicate a generic estimate of the final degrees of freedom, and specific estimates are written e.g. ν_{z1} .

3 Combining Inferences in t -based problems

In this section we discuss how to estimate total variation in t -based problems. We discuss two ways to combine t -based inferences, one based on adding variances, and the other based on adding variance parameters. This distinction does not arise with normal-based problems, because the variance of a normal distribution is its variance parameter. But with t -shaped distributions there is a difference. For example, suppose that S is the standard error for an estimate \hat{Q} , such that the posterior distribution of $(Q - \hat{Q})/S$ has a t -distribution with ν degrees of freedom, then the posterior variance for Q has variance parameter (squared standard error) S^2 but variance $\nu/(\nu - 2)S^2$. This distinction may be important when combining inferences.

We begin by extending the notation in S432. Let $U^{(t)}$ be the complete-data squared standard error for the t th complete data set, and

$$U^{*(t)} = \nu/(\nu - 2)U^{(t)}$$

the variance of the corresponding scaled t distribution. Similarly, write T and $T^* = \nu_T/(\nu_T - 2)T$ for the squared standard error and corresponding variance for the combined analysis, where ν_T is the final degrees of freedom for the inference, discussed below.

Note that from the point of view of a user of software it is most convenient to work with squared standard errors, rather than the (posterior) variances. However, it may be appropriate to work with variances internally within software that combines inferences.

One way to combine t -based inferences works solely with the squared standard errors, and uses (1, 2, 3) from Section 2, yielding the final squared standard error estimate as

$$T_{t1} = \overline{U} + (1 + m^{-1})B, \quad (7)$$

(this is distinct from (4) because $U^{(t)}$ now a squared standard error rather than a variance).

The second way adds variances of posterior distributions. Here

$$T^* = \overline{U}^* + (1 + m^{-1})B,$$

where $\bar{U}^* = m^{-1} \sum_{(t)} U^{*(t)}$, and the final squared standard error is

$$T_{t2} = \frac{\nu_T - 2}{\nu_T} T^* = \frac{\nu_T - 2}{\nu_T} \left(\left(\frac{\nu}{\nu - 2} \right) \bar{U} + (1 + m^{-1})B \right) \quad (8)$$

This is undefined if $\nu_T \leq 2$ ($\nu_T \leq \nu$), and has a small factor on B if ν is slightly greater than 2.

T_{t1} is the simpler estimate, and is more conservative (i.e. larger) than T_{t2} , which is generally preferred. However, we give here one example in which only T_{t2} gives correct inferences.

Example 1 This is Example 1, page 73, of S97. Suppose that the observed data are a sample of n_1 observations from a univariate normal distribution with unknown mean μ and variance ψ , there are $n_0 = n - n_1$ missing observations from the same distribution, and that multiple imputations would be based on data augmentation using the diffuse prior $\pi(\mu, \psi) \propto \psi^{-1}$. Using this prior for a Bayesian analysis based solely on the observed data matches the standard frequentist confidence interval for μ of the form

$$\bar{y}_1 \pm t_{\alpha/2, n_1 - 1} s_1$$

where \bar{y}_1 and s_1 are the mean and sample standard deviation of the observed data. It would be desirable for multiple-imputation based inference to yield the same interval, in the limit as $m \rightarrow \infty$ (for finite m there is simulation variability); the interval should not be different because in this univariate example multiple imputations do not add information to that contained in the observed data.

When doing data augmentation with the above prior (conditional on the observed data), the steady-state distribution for $\psi^{(t)}$ is

$$\psi^{(t)} \sim (n_1 - 1) s_1^2 / \chi_{n_1 - 1}^2,$$

which has expected value

$$E(\psi^{(t)}) = \left(\frac{n_1 - 1}{n_1 - 3} \right) s_1^2.$$

Conditional on $\psi^{(t)}$, the steady-state distribution for $\mu^{(t)}$ is

$$\mu^{(t)} \sim N(\bar{y}_1, n_1^{-1} \psi^{(t)}),$$

which has variance

$$\text{var}(\mu^{(t)}) = \frac{1}{n_1} E(\psi^{(t)}) = \frac{1}{n_1} \left(\frac{n_1 - 1}{n_1 - 3} \right) s_1^2.$$

The unconditional variance of \bar{y}_{mis} is

$$\text{var}(\bar{Y}_{mis}) = \left(\frac{1}{n_0} + \frac{1}{n_1} \right) \left(\frac{n_1 - 1}{n_1 - 3} \right) s_1^2,$$

so that

$$\text{var}(\bar{Y}) = \left(\frac{n_0}{n}\right)^2 \left(\frac{1}{n_0} + \frac{1}{n_1}\right) \left(\frac{n_1 - 1}{n_1 - 3}\right) s_1^2.$$

The complete-data sample variance has expected value

$$E(S^{2(t)}) = \left(\frac{n_1 - 1}{n_1 - 3}\right) \left(\frac{n - 3}{n - 1}\right) s_1^2.$$

Let $\hat{Q}^{(t)} = \bar{Y}^{(t)}$ and $U^{(t)} = S^{2(t)}/n$. Then the following limits hold as $m \rightarrow \infty$:

$$\begin{aligned} \bar{Q} &\rightarrow \bar{y}_1 \\ \bar{U} &\rightarrow E(S^{2(t)})/n \\ &= \frac{1}{n} \left(\frac{n_1 - 1}{n_1 - 3}\right) \left(\frac{n - 3}{n - 1}\right) s_1^2 \\ B &\rightarrow \text{var}(\bar{Y}^{(t)}) \\ &= \frac{n_0}{nn_1} \left(\frac{n_1 - 1}{n_1 - 3}\right) s_1^2 \\ T_{t1} &\rightarrow \frac{s_1^2}{n_1} \left(\frac{n_1 - 1}{n_1 - 3}\right) \left(1 - \frac{2n_1}{(n - 1)n}\right) \end{aligned} \tag{9}$$

Note that T_{t1} does not approach the desired s_1^2/n_1 .

Using the second way of combining inferences yields the same limits for \bar{Q} and B , but

$$\begin{aligned} \bar{U}^* &\rightarrow E(S^{2(t)})/n \\ &= \frac{1}{n} \left(\frac{n_1 - 1}{n_1 - 3}\right) s_1^2 \\ T^* &\rightarrow \frac{s_1^2}{n} \frac{n_1 - 1}{n_1 - 3} (1 + n_0/n_1) \\ &= \frac{s_1^2}{n_1} \frac{n_1 - 1}{n_1 - 3} \\ T_{t2} &= T^* \frac{\nu_T - 2}{\nu_T} \\ &\rightarrow \frac{s_1^2}{n_1} \text{ if } \nu_T \rightarrow n_1 - 1 \end{aligned}$$

Note that T_{t2} has the desired limiting value, if ν_T approaches the correct limiting degrees of freedom $n_1 - 1$.

3.1 Degrees of freedom for combining normal inferences

S97 provides a heuristic Bayesian justification for the procedure in S432 and Section 2, and indicates that the degrees of freedom ν_{z1} “are obtained by approximately matching the first

two moments of the reduced-information posterior to those of a t -distribution.” We provide an alternate interpretation, which suggests alternatives to ν_{z1} for normal-based problems, and which provides a way to combine ν and ν_{z*} for t -based problems.

Recall that if Z has a normal distribution with mean 0 and variance σ^2 , and if T is an estimate of σ^2 which has mean σ^2 , is proportional to a χ^2 variable and is independent of Z , then

$$\frac{Z}{\sqrt{T}} \sim t_{\nu_T} \quad (10)$$

where ν_T is the degrees of freedom for T . The reference distribution is t rather than normal because of variation in the denominator, causing the ratio to have a wider distribution.

In practice, the t reference distribution is often used when the numerator is only approximately normal, the denominator is only approximately proportional to a χ^2 variable, and the numerator and denominator have small covariance. What is important for our purposes is that variation in the denominator gives rise to a t -distribution. Furthermore, the relationship between the variance and mean of the denominator determines the degrees of freedom. If T is proportional to a χ^2 variate, then

$$\frac{\text{var}(T)}{E(T)^2} = \frac{2}{\nu_T}. \quad (11)$$

In other words, the (squared) coefficient of variation of T is inversely proportional to the degrees of freedom. Rearranging (11) yields an expression for the degrees of freedom,

$$\nu_T = \frac{2E(T)^2}{\text{var}(T)}, \quad (12)$$

which may be used whether or not T has a χ^2 distribution.

In S432, T_{z1} (4) is the squared standard error of \bar{Q} , so the appropriate degrees of freedom depends on $\text{var}(T_{z1})$, which in turn involves the variance of \bar{U} , variance of B , and the covariance of \bar{U} and B . We now make two assumptions:

A1: $\text{var}(\bar{U}) = 0$ (this implies that the covariance is also zero), and

A2: B is proportional to a χ^2 variate with $m - 1$ degrees of freedom.

Then

$$\text{var}(T_{z1}) = (1 + m^{-1})^2 \text{var}(B) = (1 + m^{-1})^2 (2/(m - 1)) E(B)^2. \quad (13)$$

Substituting into (12), and replacing $E(T_{z1})$ and $E(B)$ with T_{z1} and B , respectively, we obtain:

$$\begin{aligned} \hat{\nu}_T &= \frac{2(\hat{E}(T_{z1}))^2}{\hat{\text{var}}(T_{z1})} \\ &= \frac{2T_{z1}^2}{(1 + m^{-1})^2 (2/(m - 1)) B^2} \end{aligned}$$

which simplifies to (6). In other words, assumptions A1 and A2 lead to the degrees of freedom formula (6).

In normal-based problems, it may be possible to improve on (6) by avoiding assumptions A1 and A2. The variances and covariances of \bar{U} and B may be estimated from multiple imputations; note that T_{z1} can be written as a sample average,

$$T_{z1} = m^{-1} \sum_{t=1}^m T^{(t)} \quad (14)$$

where

$$T^{(t)} = U^{(t)} + (m/(m-1))(\hat{Q}^{(t)} - \bar{Q})^2, \quad (15)$$

so that $\text{var}(T)$ may be estimated by the usual formulas for the variance of a sample average,

$$\hat{\text{var}}(T_{z1}) = \frac{1}{m(m-1)} \sum_{t=1}^m (T^{(t)} - T)^2 \quad (16)$$

and used in (12) to estimate the final degrees of freedom,

$$\nu_{z2} = (2T^2) \left(\frac{1}{m(m-1)} \sum_{t=1}^m (T^{(t)} - T)^2 \right)^{-1} \quad (17)$$

However, doing this accurately requires that m be relatively large, but in practice it is usually small, say 3 to 5. Furthermore, the improvement over (6) is likely to be relatively small in normal problems, where the presumption is that U is exact so that variability in \bar{U} should be small.

3.2 Degrees of freedom in t -based problems

In this section we derive formulas for combining ν_{z1} (or any replacement that avoids assumptions A1 and A2) and ν . In this section T indicates one of T_{t1} or T_{t2} .

We begin with a frequentist derivation, decomposing the variance of T by conditioning on the observed data Y_{obs} :

$$\begin{aligned} \text{var}(T) &= E(\text{var}(T|Y_{obs})) + \text{var}(E(T|Y_{obs})) \\ &= E(\text{simulation variance}) + \text{other variance} \end{aligned} \quad (18)$$

In this derivation the underlying parameter Q is fixed, so there are two sources of variation: one due to random observed data Y_{obs} , and the simulation variance in choosing m random sets of imputations $Y_{mis}^{(t)}$.

The first term in (18) involves $\text{var}(T|Y_{obs})$, the simulation variation in T due to using a finite number m of random imputations, after conditioning on the observed data. This is

the variance that was estimated in (13), implicitly conditioned on Y_{obs} . We estimate this term as:

$$\begin{aligned}\hat{E}(\text{simulation variance}) &= \hat{E}(\text{var}(T|Y_{obs})) = \hat{\text{var}}(T|Y_{obs}) \\ &= \frac{2}{\nu_{sim}}(\hat{E}(T))^2 = \frac{2}{\nu_{sim}}T^2\end{aligned}\quad (19)$$

where ν_{sim} is an estimate of degrees of freedom due to simulation error in estimating T ; in particular, we may use $\nu_{sim} = \nu_{z1}$.

The second term involves variance due to random Y_{obs} . Let $U_\infty = E(U^{(t)}|Y_{obs})$ and $B_\infty = E(B|Y_{obs})$. Note that these are functions of Y_{obs} , but not of the random imputations. Then

$$\text{other variance} = \text{var}(E(T|Y_{obs})) = \text{var}(U_\infty + (1 + m^{-1})B_\infty). \quad (20)$$

In the absence of missing data, $B_\infty = 0$ and $U_\infty = U$, the complete-data squared standard error for \hat{Q} . In this case, the degrees of freedom ν from the complete-data problem implies by (11) that $\hat{\text{var}}(U) = (2/\nu)U^2$.

In the presence of missing data, we begin with a simple estimate for the non-simulation variance of T , then propose adjustments. The simple estimate supposes that $(aU_\infty + bB_\infty)$ is proportional to a χ^2 variable with ν degrees of freedom for any positive a and b (particular a and b correspond to T_{t1} and T_{t2}), yielding the estimate

$$\hat{\text{var}}(E(T|Y_{obs})) = (2/\nu)T^2. \quad (21)$$

This choice in combination with (19) leads to a simple estimate for the final degrees of freedom,

$$\begin{aligned}\nu_{t1} &= \frac{2T^2}{\hat{\text{var}}(T)} = \frac{2T^2}{(2/\nu_{sim})T^2 + (2/\nu)T^2} \\ &= \left(\frac{1}{\nu_{sim}} + \frac{1}{\nu}\right)^{-1}\end{aligned}\quad (22)$$

Note that this is never greater than ν ; this is desirable in statistical software, where the degrees of freedom for an analysis with multiple imputations should be at least as small as would obtain with complete data. However, note that as $m \rightarrow \infty$ that (22) approaches ν , but the actual degrees of freedom should be smaller because some data are missing.

We propose two ways to adjust the relatively simple estimate (22). The basic idea is to estimate the fraction of nonmissing data as \bar{U}/T , and adjust the original degrees of freedom by this quantity, yielding

$$\nu_{t2} = \left(\frac{1}{\nu_{sim}} + \frac{1}{(\bar{U}/T)\nu}\right)^{-1} \quad (23)$$

A minor variation on the previous adjustment is based on maintaining a distinction between sample size and degrees of freedom. For example, in Example 1, the observed and

complete-data sample sizes were n_1 and n , respectively, while the appropriate degrees of freedom are $n_1 - 1$ and $n - 1$; the degrees of freedom are offset from the sample size by 1. In other problems the offset is different, e.g. $p + 1$ in linear regression with p coefficients and an intercept. If n is known and the appropriate offset is $(n - \nu)$; then the adjusted non-simulation degrees of freedom would be $(\bar{U}/T)n - (n - \nu)$, yielding

$$\nu_{t3} = \left(\frac{1}{\nu_{sim}} + \frac{1}{(\bar{U}/T)n - (n - \nu)} \right)^{-1} \quad (24)$$

Both (23) and (24) may be calculated using T_{t1} or T_{t2} ; using T_{t2} requires solving a system of two equations in two unknowns. The combination of T_{t2} and (24) and yields the desired answer as $m \rightarrow \infty$ in Example 1.

3.3 Summary for univariate estimates

The single estimate for total variation T in normal-based problems is T_{z1} (4).

Estimates for degrees of freedom ν_T in normal-based problems are ν_{z1} and ν_{z2} (6, 17). As long as m is relatively small, we suggest using ν_{z1} because ν_{z2} would be highly variable.

Estimates for total variation T in t -based problems are: T_{t1} and T_{t2} (7, 8), both linear combinations of \bar{U} and B . T_{t1} is simplest. T_{t2} is more accurate in Example 1 as $m \rightarrow \infty$, but T_{t1} is more conservative, which would generally be preferred in more complicated situations (where some of the assumptions underlying the methods may not hold) or where m is small. Differences between these will be small if the original degrees of freedom ν are large, or if the fraction of missing information is smaller. We suggest using T_{t1} , which is simpler and more conservative.

Estimates for degrees of freedom ν_T in t -based problems are: ν_{t1} , ν_{t2} , and ν_{t3} (22, 23, 24). We suggest using ν_{t3} , which should be the most accurate.

Combining estimate T_{t2} for total variation and degrees of freedom (ν_{t2} , ν_{t3}) require solving a system of two nonlinear equations in two unknowns. We suggest using the approximations obtained by first computing T_{t1} , using it to compute degrees of freedom, then using those degrees of freedom in T_{t2} .

4 Combining inferences in for multidimensional estimates, χ^2 situations

We begin by reviewing the rules from S433 for combining inferences in χ^2 problems; these largely parallel the rules in Section 2 from S432.

Let \hat{Q} be a complete-data point estimate of a k -dimensional parameter Q , and let U be its covariance matrix (or an very accurate estimate), and assume that \hat{Q} is approximately

distributed as $N(Q, U)$, so that complete-data inferences would be based on

$$(\hat{Q} - Q)^T U^{-1} (\hat{Q} - Q) \sim \chi_k^2$$

The multivariate analogs of (1-4) are

$$\bar{Q} = \frac{1}{m} \sum_{t=1}^m \hat{Q}^{(t)}, \quad (25)$$

$$\bar{U} = \frac{1}{m} \sum_{t=1}^m U^{(t)}, \quad (26)$$

$$B = \frac{1}{m-1} \sum_{t=1}^m (\hat{Q}^{(t)} - \bar{Q})(\hat{Q}^{(t)} - \bar{Q})^T, \quad (27)$$

$$T_{c1} = \bar{U} + (1 + m^{-1})B \quad (28)$$

Inferences are based on the test statistic

$$(Q - \bar{Q})^T T^{-1} (Q - \bar{Q}) \sim F_{k, \nu_T} \quad (29)$$

for some T and associated degrees of freedom ν_T .

S433 notes that B is a noisy estimate of $\text{var}(\hat{Q}|Y_{obs})$, and does not even have full rank of $m \leq k$, and implies that using T_{z1} for T in (29) yields a test statistic which may not be approximately F -distributed. S433 suggests assuming that

$$\mathbf{A3:} \quad \text{var}(\hat{Q}|Y_{obs}) \propto E(U|Y_{obs}),$$

and letting

$$T_{c2} = (1 + r_1)\bar{U} \quad (30)$$

where

$$r_1 = (1 + m^{-1})\text{tr}(B\bar{U}^{-1})/k$$

with degrees of freedom

$$\nu_{c1} = \begin{cases} k^*(1 + k^{-1})(1 + r_1^{-1})^2/2 & \text{if } k^* = k(m-1) \leq 4 \\ 4 + (k^* - 4)[1 + (1 - 2/k^*)r_1^{-1}]^2 & \text{otherwise} \end{cases} \quad (31)$$

S433 indicates that assumption A3 is equivalent to assuming that the fractions of missing information for all components of Q are equal. We believe that the assumption is actually stronger, that it implies that the fractions of missing information for all linear combinations of components of Q are equal.

We suggest the weaker assumption

A4: the correlation matrices corresponding to $\text{var}(\hat{Q}|Y_{obs})$ and $E(U|Y_{obs})$ are equal.

Then let

$$B' = \text{diag}(B)^{1/2} \text{diag}(\bar{U})^{-1/2} \bar{U} \text{diag}(\bar{U})^{-1/2} \text{diag}(B)^{1/2} \quad (32)$$

be the adjusted estimate of $\text{var}(\hat{Q}|Y_{obs})$ and, where $\text{diag}(M)$ for a square matrix M is the matrix with the same diagonal elements and zero elsewhere, and

$$T_{c3} = \bar{U} + (1 + m^{-1})B' \quad (33)$$

be the estimate of total variation. Note that the diagonal elements of this matrix are the same as if total variation were estimated individually for components of the multivariate parameter using (4).

It is easy to construct examples for which even the weaker assumption A4 is violated. For example, if X and Y are jointly gaussian with largely disjoint sets of missing values and the parameters of interest are the means of the variables, then the off-diagonal element of $\text{var}(\hat{Q}|Y_{obs})$ is really larger than implied by assumption A4.

4.1 Degrees of freedom in multivariate χ^2 situations

Degrees of freedom may also be computed individually for components of the multivariate parameter using (6) or (17). These may be combined to obtain the denominator degrees of freedom using the conservative choice of the smallest degree of freedom

$$\nu_{c2} = \min_{j=1,\dots,k} \nu_{1,j} \quad (34)$$

the reciprocal average

$$\nu_{c3} = \left(k^{-1} \sum_{j=1}^k \nu_{1,j}^{-1} \right)^{-1} \quad (35)$$

or the directionally-weighted reciprocal average

$$\nu_{c4} = \left((\sum w_j)^{-1} \sum_{j=1}^k w_j \nu_{1,j}^{-1} \right)^{-1} \quad (36)$$

where $w_j = (T_{5,j,j})^{-1/2}(\bar{Q}_j - Q_j)$ is proportional to the normalized value of $\bar{Q} - Q$ in the j th direction. We use reciprocal (weighted) averages because of the reciprocal relationship (12) between degrees of freedom and squared coefficient of variation of T .

4.2 Summary for χ^2 situations

Note that assumption A4 is much weaker than A3, and affects only the correlation structure of the total variation estimates. The correlation structure is where the lack of full rank in B would occur, and is also presumably where the greatest noise would occur.

We suggest the use of Assumption A4 and resulting estimate of total variation T_{c3} . The big advantage of this over A3 and T_{c2} is that overall results are consistent with component-wise results.

We suggest using the conservative choice of the smallest degrees of freedom, ν_{c2} ; in fact even this choice is not overly conservative, as it is easy to construct examples where the fraction missing information for a combination of parameters is higher than for any parameter individually. This conservative choice is sensitive to random variation with large k and small m , in that a single one of a large number of parameters may have a high individual estimated degrees of freedom. The choice ν_{c4} might be the most accurate in the majority of situations, but requires specification of Q in order to compute directional weights, leading to the unusual situation for F statistics that the denominator degrees of freedom depends on the particular null hypothesis value being tested.

5 Combining inferences in for multidimensional estimates, F situations

Inferences in F situations would involve the test statistic

$$k^{-1}(\hat{Q} - Q)^T U^{-1}(\hat{Q} - Q) \sim F_{k,\nu}$$

for complete data problems, where Q and \hat{Q} is as in Section 4 but U is an estimate of the covariance matrix for \hat{Q} (which is smaller than the covariance matrix U^* for the posterior distribution by a factor $(\nu - 2)/\nu$).

Combining multiple-imputations inferences in F situations involves a combination of the ideas from t and χ^2 situations. We summarize here the results that would obtain using the particular methods recommended in earlier sections.

We begin by extending the notation used in χ^2 situations, Let $U^{(t)}$ be the ‘‘covariance-error’’ matrix for the t th complete data set, with average \bar{U} , and let T denote the total covariance-error matrix. Let B' be as in (32). The final estimate of total variation based on (7) and (33) is

$$T_{f1} = (\bar{U} + (1 + m^{-1})B') \tag{37}$$

The test statistic is the F -statistic (29).

The degrees of freedom for individual components of T may be computed as for t -based problems, and the overall degrees of freedom (the denominator degrees of freedom for the F -statistic) computed using the conservative choice (34).

6 Combining F -statistics

In the previous section we discussed combining F -inferences, when the parameters and covariance estimates used to obtain F -statistics were available. Here we combine inferences

based solely on the statistics. This incorporates ideas from previous sections and the material on page 115 in S97, which is based on (Li et al. (1991)). We begin by reviewing that material, which relates to combining χ^2 statistics.

The complete-data Wald (χ^2) statistics are

$$d_W^{(t)} = (Q^{(t)} - Q_0)^T (U^{(t)})^{-1} (Q^{(t)} - Q_0).$$

The combined statistic is

$$D_2 = \frac{\bar{d}_W k^{-1} - (m+1)(m-1)^{-1} r_2}{1 + r_2}$$

where

$$\bar{d}_W = \frac{1}{m} \sum_{t=1}^m d_W^{(t)}$$

is the average of the Wald statistics, and

$$r_2 = (1 + m^{-1}) \left[\frac{1}{m-1} \sum_{t=1}^m \left(\sqrt{d_W^{(t)}} - \sqrt{\bar{d}_W} \right)^2 \right]$$

is an estimate of the average relative increase in variance. The numerator degrees of freedom are k and the denominator degrees of freedom are

$$\nu_2 = k^{-3/m} (m-1) (1 + r_2^{-1})^2$$

To extend this methodology to the case where F statistics are available, we note that dividing a χ^2 variate by its degrees of freedom k yields an F variate with k and ∞ degrees of freedom. Given F -statistics, we propose to convert them into approximate χ^2 statistics by multiplying by k and applying the above methodology; however the denominator degrees of freedom combine the degrees of freedom ν_2 due to finite m with the original denominator degrees of freedom using (23), except that “fraction of information not lost to missing data” will be estimated by $(1 - 2/(\nu_2 + 3))/(r_2 + 1)$ instead of \bar{U}/T ; the former is based on (4.30) in S432.

In particular, if the individual F -statistics are $d_F^{(t)}$ with k and ν_{denom} let

$$d_W^{(t)} = D_F^{(t)} * k,$$

calculate D_2 , \bar{d}_W , r_2 , and ν_2 as above, let the overall statistic be

$$D_F = D_2/k$$

with numerator degrees of freedom k and denominator degrees of freedom

$$\nu_2' = \left(\frac{1}{\nu_2} + \frac{1}{\frac{1-2/(\nu_2+3)}{r_2+1} \nu_{\text{denom}}} \right)^{-1}$$

7 Linear Model Summary Statistics

There are well-known relationships that apply between common summary statistics for a linear model, in particular including those in an analysis of variance (anova) table of the form:

	SS	df	MS	F
model	SS_m	ν_m	MS_m	F
error	SS_e	ν_e	MS_e	
total	SS_T	ν_T		

as well as $R^2 = SS_m/SS_T$ and residual standard deviation $s = \sqrt{MS_e}$. In this section we discuss the computation of these quantities in a multiple-imputation setting.

Ideally, these summary statistics could be computed in the multiple-imputation context in a way that is consistent with their uses in the non-imputation context for both

- descriptive summaries and
- inference.

However, this does not appear possible. In particular, the F -statistic has two uses:

- it is a simple descriptive statistic measuring the quality of the model; it estimates the ratio between the actual reduction in residual sums of squares due to the model and what the reduction would be under the null hypothesis, and
- it is used in computing a p -value for determining statistical significance

The appropriate values of F for these uses differ in the multiple-imputation context.

We propose to maintain a distinction between inferential and non-inferential statistics. For inferential purposes, only F , ν_m and ν_e are needed; these were discussed in Section 5, albeit with different notation: $\nu_m = k$, and $\nu_e = \nu_{e2}$.

For computing descriptive statistics, we presume that all quantities in the anova table are available for each complete data set, and are denoted with a superscript^(t), e.g. $SS_m^{(t)}$. Then we propose the following definitions:

$$\begin{aligned}
 SS_m &= \overline{SS_m^{(t)}} & \nu_m &= \nu_m^{(t)} & MS_m &= \overline{MS_m^{(t)}} = SS_m/\nu_m & F &= MS_m/MS_e \\
 SS_e &= \overline{SS_e^{(t)}} & \nu_e &= \nu_e^{(t)} & MS_e &= \overline{MS_e^{(t)}} = SS_e/\nu_e \\
 SS_T &= \overline{SS_T^{(t)}} & \nu_T &= \nu_T^{(t)}
 \end{aligned}$$

The sum of squares and mean squares terms are all simple averages across the imputations. The degree of freedom terms are identical across imputations. Other quantities, including R^2 and s , are calculated as in the non-imputation setting. The values for ν_e and F are different than the values used for inference.

Note that these definitions are consistent with the descriptive use of the F ratio described above, the use of R^2 as a description of “the fraction of variance explained by the model,” and the use of $MS_e = s^2$ as an estimate of the conditional variance of the response given the explanatory variables (assuming homoskedasticity).

References

- Li, H. H., Meng, X. L., Raghunathan, T. E., and Rubin, D. B. (1991). Significance Levels from Repeated p -Values with Multiply-imputed Data. *Statistica Sinica*, 1:65–92.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley, New York.
- Schafer, J. (1997). *Analysis of incomplete multivariate data*. Chapman and Hall.