Analysis of Variance from Multiply Imputed Data Sets

Trivellore Raghunathan
Department of Biostatistics, School of Public Health &
Institute for Social Research
University of Michigan
Ann Arbor, MI 48109.
E-mail: teraghu@umich.edu

Qi Dong
Survey Methodology Program
Survey Research Center
Insitute for Social Research
University of Michigan
Ann Arbor, MI 48106.
E-mail: qidong@umich.edu

Analysis of Variance from Multiply Imputed Data Sets

Abstract

The analysis of variance is a popular method used in many scientific applications. There are standard software for handling unbalanced data due to missing values in the outcome/dependent variable. The analysis becomes difficult when the missing values are in predictors. Multiple imputation is an increasingly popular method for handling such incomplete data. This approach involves replacing the missing set of values by more than one plausible set of values, preferably generated from their posterior predictive distribution given the observed data. Each plausible set of imputed values when combined with the observed set of values results in a completed data. Each completed data set is analyzed separately and the point estimates and their standard errors are combined to form a single inference. Many analysis of variance models may be formulated as regression models and then apply the standard multiple imputation combining rules. This is often not possible when the design is complex involving repeated measures and/or nested, random or interaction effects. It may be more convenient to directly combine the analysis of variance tables generated from each completed data to test appropriate hypotheses. This paper develops a combining rule for the completed data mean squares. Approximate F-tests are developed and evaluated using the actual and simulated data sets. The method is extended to comparison of regression models using partial F-tests in multiple linear regression analysis or the deviance statistics in fitting regression models using the Generalized Estimating Equations.

Key Words: Design of experiments, F-distribution, Nested effects, Random effects

1 Introduction

The analysis of variance (ANOVA) is a very useful analytical tool for drawing inferences from observational and experimental studies. There are standard software packages to carry out these analyses for a variety of designs involving repeated measures, nested and random effects etc. Though, these software packages can handle unbalanced data arising due to missing values in the outcome or dependent variable, they exclude subjects with missing predictor variables.

When the missing data is confined to the outcome variable, a selected set of ANOVA models can be re-formulated as regression models and then maximum likelihood approach can be used to fit them assuming that the data are Missing at Random (MAR) (Rubin, 1976). Validity of this approach is questionable when the sample size is small or when the missing data cannot assumed to be MAR. Regardless of the sample size, it is difficult to carry out a valid analysis when the data are missing for predictors or factors, unless the data are missing completely at random (MCAR).

Multiple imputation is increasingly becoming a popular method for analyzing such incomplete data. This method involves imputing M plausible set of values for the missing set of values to generate M completed data sets. Each completed data set is then analyzed separately using the standard complete data software and then the results are combined to form a single inference. Several combining rules have been developed for a variety of analytical methods. An earliest combining rule is for parameter estimates such as regression coefficients (Rubin and Schenker(1986), Rubin(1987)). Procedures for performing multivariate tests of hypotheses are considered in Li, Raghunathan and Rubin (1991); combining the p-values from completed data sets in Li et al (1992); the likelihood ratio tests from multiply imputed data set is considered in Meng and Rubin (1992).

Some of these combining rules could be used especially when the underlying model can be reformulated as a regression model and then the ANOVA tests can expressed as multivariate hypotheses tests involving a set of regression coefficients. This is often not possible when the design is complex involving nested, random or interaction effects. It is more convenient to directly combine the analysis of variance tables generated from each completed data set to test appropriate hypotheses, even if the formulation as a regression model is possible.

This paper develops and evaluates a procedure that combines the mean squares from each completed data set and then constructs an approximate F-statistics. It is assumed that the missing values in the data set have been multiply imputed by an imputer using an appropriate approach and the multiply imputed data set has been made available to the analyst. The task of the analyst is to test a particular hypothesis using these multiply imputed data sets. This paper does not consider alternative methods for generating imputations. However, it is assumed that the imputation procedure used by the imputer yields unbiased estimates of the components of variances for the particular tests being conducted. Some suggestions for imputing the missing values are discussed later.

The rest of the article is organized as follows. Section 2 describes the approach and illustrates its application using a numerical example. The technical justification is given in Section 3. Section 4 evaluates the approach from a repeated sampling perspective by computing the exact level of significance for the nominal 5% level tests through simulations. In the simulation study, the imputations are carried out using the sequential regression approach (Raghunathan et al (2001)) as implemented in the software IVEare (www.isr.umich.edu/src/smp/ive). Section 5 extends the methodology to partial F-tests in the multiple regression analysis and comparison of models using deviance statistics in the analysis using Generalized Estimating Equations. Finally, Section 6 concludes with the discussion and potential future work.

2 Combining ANOVA Results

At the basic level, the analysis of variance involves comparing two independent or approximately independent mean squares (sums of squares divided by their degrees of freedom) that have the same expected values (the associated variance components) under the specific null hypothesis. The ratio of the mean squares to the corresponding variance components defines a pivotal quantity, a function of a statistic and the parameter having a sampling distribution independent of the parameter. To be concrete, consider the one-way analysis of variance setup with k treatments or groups and n_i subjects in group $i = 1, 2, \ldots, k$. Let y_{ij} be the response variable on subject $j = 1, 2, \ldots, n_i$ in group i; $\sum_j y_{ij}/n_i$ be the mean for group i and $\sum_i \sum_j y_{ij}/n = \sum_i n_i \bar{y}_i/n$ be the overall mean where $n = \sum_i n_i$ is the total sample size. To analyze the data from such a design, the following ANOVA model is posited: $y_{ij} = \mu + \alpha_i + \epsilon_{ij}$ where ϵ_{ij} iid $N(0, \sigma^2)$ and $\sum_i \alpha_i = 0$. Under this model, the between-group mean square $BMS = \sum_i n_i(\bar{y}_i - \bar{y})^2/(k-1)$ has the expected value, the variance component, $\sigma_b^2 = \sigma^2 + \sum_i n_i \alpha_i^2/(k-1)$ and the pivotal quantity is $P_N = BMS/\sigma_b^2$ which has a chi-square distribution with $k-1$ degrees of freedom. Similarly, the mean square error, $MSE = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2/(n-k)$ has the expectation σ^2 and $P_D = MSE/\sigma^2$ is also a pivotal quantity which has a Chisquare distribution with $n-k$ degrees of freedom where $n = \sum_i n_i$ is the total sample size. Given the independence of the two mean squares, BMS and MSE, the ratio P_N/P_D is also a pivotal quantity having an F-distribution with $k-1$ and $n-k$ degrees of freedoms. Under the null hypothesis, $H_o : \alpha_i = 0, i = 1, 2, \ldots, k$, the ratio $P_N/P_D = BMS/MSE$ becomes a statistic (independent of the parameters) with the same F- distribution.

The distributions of the pivotal quantities can be interpreted from the Bayesian perspectives as well. In the Bayesian framework, these pivotal quantities describe approximate posterior distributions of the variance components conditional on the observed data (See, for example, Box and Tiao (1973)). For example, defining $\rho = \sigma^2/\sigma_b^2$, the posterior distribution of ρ is a multiple MSE/BMS of an F-distribution with $k-1$ and $n-k$ degrees of freedom. Based on this posterior distribution, one could construct, say, a $(1-\gamma)100\%$ Bayesian confidence or credible interval for ρ. If the interval excludes 1 then we may reject the null hypothesis. This pragmatic view of interpreting the distribution of the pivotal quantities from both Bayesian and Frequentist perspectives will be used to justify the procedure described in this paper. That is, the combining rules will be developed from the Bayesian perspective for the ratio of the variance components and then reverted to a Frequentist test statistics under the specific null hypothesis.

In general, testing of hypothesis under any ANOVA model involves choosing the appropriate ratio of pivotal quantities (mean squares divided by the corresponding variance component) that become independent of parameters under the hypothesis being tested has an (or approximate) F- distribution. Suppose that, in the absence any missing data (i.e. based the complete data),s_N is the numerator mean square with ν_N degrees of freedom and the associated component of variance σ_N^2. Similarly, let s_D, ν_D and σ_D^2 be the corresponding quantities for the denominator mean square error, the degrees of freedom and the variance components, respectively. The two complete data pivotal quantities $P_N = \nu_N s_N/\sigma_N^2$ and $P_D = \nu_D s_D/\sigma_D^2$ have chi-square distributions with the degrees of freedom, ν_N and ν_D, respectively.

Suppose that under a certain null hypothesis, $\sigma_N^2 = \sigma_D^2$, then the ratio of the two pivotal quantities P_N/P_D is a statistic and its sampling distribution follows an F-distribution with degrees of freedom (ν_N, ν_D). More generally, the pivotal quantities can be used to construct frequentist confidence intervals or Bayesian credible intervals for the variance components or for the ratio of the variance components. These will be numerically identical but differ in interpretations.

Our goal is to use the same framework to develop approximate pivotal quantities when the missing values in the data sets have been multiply imputed. Suppose that based on the completed data, $l = 1, 2, \ldots, M$, the mean squares are $s_N^{(l)}$ and $s_D^{(l)}$ with the respective degrees of freedoms $\nu_N^{(l)}$ and $\nu_D^{(l)}$. Typically, the degrees of freedom across the completed data sets will be constant (i.e. $\nu_N^{(l)} = \nu_N$ and $\nu_D^{(l)} = \nu_D$), but, occasionally, the degrees of freedom may differ across the completed data sets, for example, when Satterthwaite approximation is used to constructed tests from each completed data set under some nested random effects models.

Define $A_N = \sum_l (1/s_N^{(l)})/M$, $B_N = \sum_l (1/(\nu_N^{(l)} \times s_N^{2(l)}))/M$ and $C_N = \sum_l (1/s_N^{2(l)} - A_N)^2/(M-1)$. Similarly, A_D, B_D and C_D are defined for the denominator mean square and its degrees of freedom. The proposed procedure is to use $F_{MI} = A_D/A_N$ as the multiple imputation

F-statistics with the degrees of freedom (r_N, r_D) where $r_N = 2A_N^2(2B_N + (M+1)C_N/M)$ and $r_D = 2A_D^2(2B_D + (M+1)C_D/M)$. Note that the proposed F-statistic is the ratio of the harmonic means of the completed data numerator and denominator mean square errors.

To illustrate the methodology, we use the data from Wave 1 of the National Longitudinal Study of Adolescent Health (Add Health)(Harris and Udry (1994)). The dependent variable (Y) is the vocabulary score and the two factors are the Household income (X_1)(four categories corresponding to quartiles) and the Attitude towards sexual behavior (X_2) (four categories based on quartiles of an index constructed from 8 items). The total sample size is $n = 6,224$ and if one were to perform the complete-case analysis (that is, remove all subjects with missing values in the two predictors)the sample size would have been reduced to $n_{complete} = 2,853$.

The missing values were imputed using the sequential regression imputation approach which included main effects and all two-factor interactions as predictors in each regression model. Specifically, in imputing Y the predictors were X_1, X_2 and $X_1 \times X_2$, while imputing X_1 the predictors were Y, X_2 and $Y \times X_2$ and for imputing X_2 the predictors were Y, X_1 and $Y \times X_1$. The normal linear regression model was used to impute Y, and multinomial logit models were used for imputing X_1 and X_2.

We illustrate the test for the main effect of X_2 using($M = 5$)five imputed data sets. For the complete data, the numerator degrees of freedom is $\nu_N = 3$ and the denominator degrees of freedom is $\nu_D = 6207$. The five mean squares for X_2 are: 2521.85, 2534.79, 3021.39, 3873.45, and 3246.94. The corresponding five mean square errors are: 201.98, 201.11, 200.77, 199.68, and 199.46. These numbers result in $A_N = (1/2521.85 + 1/2534.79 + \ldots + 1/3246.94)/5 = 0.000338$ and $A_D = (1/201.98 + \ldots + 1/199.46)/5 = 0.004985$ resulting in $F_{MI} = 0.004985/0.000338 = 14.77$. In contrast, the complete-case analysis based on 2,853 subjects results in $F_{CC} = 6.85$.

To compute the degrees of freedom, note that $B_N = (1/3)(1/2521.85^2 + \ldots + 1/3246.94^2)/5 = 3.89 \times 10^{-8}$ and $B_D = 4 \times 10^{-9}$. The variance of the reciprocals of the mean squares are

$C_N = 3.49 \times 10^{-9}$ and $C_D = 6.71 \times 10^{-10}$. These numbers result in $r_N = 2 \times 0.000338^2/(2 \times 3.89 \times 10^{-8} + (5+1) \times 3.49 \times 10^{-9}/5) = 2.78$ and $r_D = 5639.97$. The p-value based on this test statistic is 4×10^{-7} in contrast to the complete-case analysis p-value 14×10^{-5}. Though the substantive conclusions remain the same under both the analysis, the multiple imputation analysis indicates much stronger evidence of the association between the attitude towards the sexual behavior and vocabulary score.

A similar calculations for testing the interaction effect$X_1 \times X_2$ yields $F_{MI} = 0.7312, r_N = 3.1706$ and the p-value of 0.533. The corresponding complete case analysis yields $F_{CC} = 1.077$ and the p-value 0.376. The multiple imputation analysis indicates somewhat weaker evidence of interaction than the complete-case analysis.

3 Technical Justification

The most straightforward justification of the multiple imputation inferences is from the Bayesian perspective although such procedures have been shown to have desirable repeated sampling properties. We adopt the same standard multiple imputation Bayesian framework to justify the approximation of the pivotal quantities that leads the test statistics described in the previous section. Writing the l^{th} completed data set as $D_l = (D_{obs}, D_{mis}^{(l)})$, where the D_{obs} is the observed data and $D_{mis}^{(l)}$ is the l^{th} imputations of the missing data D_{mis}, we have

$$\frac{\nu_N^{(l)} s_N^{(l)}}{\sigma_N^2} | D_l \sim \chi^2_{\nu_N^{(l)}}.$$

The completed data posterior mean of σ_N^{-2} is $1/s_N^l$ and the posterior variance is $2/(\nu_N^{(l)} s_N^{2(l)})$. Using the standard multiple imputation theory from Rubin (1987), the Multiple imputation posterior mean of σ_N^{-2}, conditional on the observed data is $A_N = \sum_l 1/s_N^{(l)}/M$ and its posterior variance is

$$T_N = 2B_N + (1 + M^{-1})C_N$$

We approximate the posterior distribution of σ_N^{-2} by a multiple of a chi-square distribution, $a\chi_b^2$, and a and b are determined by matching the posterior mean and variance. Specifically, we get $ab = A_N$ and $2a^2b = T_N$. Solving, for a and b, we get $b = r_N$ defined in the previous section and $a = A_N/r_N$. Thus, we have $\sigma_N^{-2}|D_{obs} \approx a\chi_b^2$ or equivalently $A_N^{-1}\sigma_N^{-2}|D_{obs} \approx \chi_{r_N}^2/r_N$, a pivotal quantity. Using the same argument for the denominator quantities, we obtain $A_D^{-1}\sigma_D^{-2}|D_{obs} \approx \chi_{r_D}^2/r_D$.

Thus, the ratio $A_D\sigma_D^2/A_N\sigma_N^2$ has an approximate F- distribution with the degrees of freedom r_N and r_D. From the frequentist perspective, if a particular null hypothesis implies that $\sigma_N^2 = \sigma_D^2$, then the test statistic A_D/A_N can be referred to an F distribution with r_N and r_D degrees of freedom. Alternatively, one could construct a Bayesian credible interval using the approximation for the posterior distribution of the ratio $\sigma_D^2/\sigma_N^2|D_{obs} \approx (A_D/A_N)F_{r_N,r_D}$ and rejecting the hypothesis if 1 is contained in the interval. See Box and Tiao (1973) for constructing the highest posterior density credible interval for the ratio of the variance components.

4 Simulation Study

Obviously, any single data analysis cannot determine which of the two approaches used in the ADD-HEALTH example, (Complete-case versus Multiple Imputation Analysis) is correct or valid. Therefore, we conducted a simulation study to evaluate the exact level of the nominal 5% tests using the above method. The complete data consisted of 3 variables X_1 generated from a multinomial distribution with 4 categories with probabilities $Pr(X_1 = i) = 0.25, i = 1, 2, 3, 4$ and X_2 generated from an independent multinomial distribution with 3 categories with probabilities $Pr(X_2 = j) = 1/3, j = 1, 2, 3$. For each of the 12 combinations of X_1 and X_2 (cells), the outcome, Y was generated from 4 different models given below. The cell sample size was fixed at 10. One thousand data sets were generated from each of the following four models:

1. $y_{ijk} \sim N(0, 1), k = 1, 2, \ldots, 10$. For this model, all three null hypotheses, no interaction ($X_1 \times X_2$) and no main effects (X_1 and X_2) are true.

2. $y_{ijk} \sim N(\mu + \alpha_i + \beta_j, 1)$ where $\mu = 0$, $\alpha = (-1.5, -0.5, 0.5, 1.5)$ and $\beta = (-1, 0, 1)$. The null hypothesis of no interaction effect is true for this model.

3. $y_{ijk} \sim N(\mu + \alpha_i, 1)$. For this model, the two true null hypotheses are no interaction effect $X_1 \times X_2$ and no main effect X_2; Finally,

4. $y_{ijk} \sim N(\mu + \beta_j, 1)$. For this model, the two true null hypotheses are no interaction effect $X_1 \times X_2$ and no main effect X_1.

Some values of X_1 and X_2 were deleted using two different mechanisms. The first was Missing Completely at Random (MCAR)where roughly 15% of X_1 and 15% of X_2 were set to missing at random. The second mechansism was Missing at Random where

$$logit Pr(X_1 = missing) = \theta_o + \theta_1 Y.$$

The values of X_2 were set to missing with probability given by

$$logit Pr(X_2 = missing) = \theta_2 + \theta_3 Y,$$

if X_1 is missing and

$$logit Pr(X_2 = misssing) = \theta_4 + \theta_5 Y + \theta_6 X_1 + \theta_7 X_1 \times Y,$$

if X_1 is not missing. The coefficients in these logistic models were all non-zero and were determined to yield about 70% complete cases. In addition, the outcome was set to missing for about 10% of subjects completely at random for both mechanisms. It is possible that some subjects had all three variables with missing values (unit nonresponse) in some data sets.

All data sets with missing values were multiply imputed ($M = 10$) using the sequential regression approach as implemented in IVEware (2001). Regardless of the true model or mechanism, the imputation of missing values were draws from the posterior predictive distribution corresponding to regression model with Y as the outcome and three dummy variables for X_1, two for X_2, and their product as predictors. The intercept was also included. For imputing X_1, we used multinomial logit model with Y, two dummy variables for X_2 and their product as predictors. Similarly, for imputing X_2, we used multinomial logit model with Y, three dummy variables for X_1 and their product as predictors. The intercept term was also included in both multinomial logit models.

Three hypothesis tests, one for no interaction and two for no main effects were carried out for "Before Deletion", "Complete Cases" and "Multiply Imputed" data sets. The nominal level used were 5% and the proportion of 1000 data sets for which the null hypotheses were rejected is tabulated in Table 1. The results are provided only for true null hypotheses under each model.

Table 1: The exact levels (in %) of various null hypotheses based on Before deletion (BD), complete-cases (CC) and multiply imputed (MI) data sets under four models and two mechanisms

Model	Source	MCAR			MAR		
		BD	CC	MI	BD	CC	MI
1	X_1	5.2	5.4	5.3	5.8	3.8	5.1
	X_2	4.6	4.7	6.1	4.9	4.0	4.4
	$X_1 \times X_2$	4.9	5.4	4.4	5.3	3.1	4.1
2	$X_1 \times X_2$	4.6	5.0	4.5	4.6	46.7	3.5
3	X_1	4.4	6.1	5.5	4.8	48.6	3.8
	$X_1 \times X_2$	4.8	5.6	3.7	4.4	37.4	4.1
4	X_2	4.4	5.2	5.1	4.7	46.0	4.9
	$X_1 \times X_2$	4.3	5.5	4.6	4.9	47.0	3.6

Based on the table, the proposed test has desirable levels across all simulation conditions and perhaps slightly conservative for the interaction effects. The complete case analysis is considerably biased under MAR.

5 Extensions

Partial F-test, which is a ratio of two chi-square statistics is used in comparing models in the multiple regression analysis. Suppose that the residual sum of squares from the reduced model from the l^{th} completed data set is $R_o^{(l)}$ and for the full model it is $R_1^{(l)}$. Define $s_N^{(l)} = (R_o^{(l)} - R_1^{(l)})/(p_1 - p_o)$ where $p_1(p_o)$ is the number of parameters (regression coefficients) in the full (reduced) model, $s_D^{(l)} = R_1^{(l)}/(n - p_1)$ where n is the sample size, $\nu_N = p_1 - p_o$ and $\nu_D = n - p_1$. We can now apply the procedure described in this paper to perform a multiple imputation partial F-test to compare the fit of reduced and full models.

The same approach can be used to perform the overall F-test in a regression analysis. Suppose that the regression model under consideration is $y = \beta_o + \sum_j^p \beta_j x_j + \epsilon$. Let $M^{(l)}$ be the model sum of squares from the l^{th} completed data set and $R^{(l)}$ be the corresponding residual sum of squares. Define $\nu_N = p$, $\nu_D = n - p - 1$, $s_N^{(l)} = M^{(l)}/p$ and $s_D^{(l)} = R^{(l)}/(n - p - 1)$. The ratio $F_{MI} = A_D/A_N$ is the multiple imputation test for the null hypothesis $H_o : \beta_1 = \beta_2 = \ldots, \beta_p = 0$ and is an alternative to procedure described in Li, Raghunathan and Rubin (1991) which requires point estimates and the covariance matrix from each completed data. This may be more efficient or powerful than the procedure described in Li et al (1992) which only requires p-values from each completed data.

An interesting by-product of above analysis is the multiple imputation combined R-square. Using the well-known relationship between R^2 and F in the least square analysis using multiple linear regression model, we can define the multiple imputation R-square as

$$R_{MI}^2 = r_N F_{MI}/(r_N F_{MI} + r_D).$$

The adjusted R_{MI}^2 may be defined as $R_{a,MI}^2 = 1 - (1 - R_{MI}^2)(n - 1)/(n - p - 1)$.

When fitting regression models using Generalized Estimating Equations approach, the deviance is often used to compare the models. These can be viewed as partial F-test using Deviance statistic as residual sum of squares. Let $D_o^{(l)}$ and $D_1^{(l)}$ be the deviance statistics (or -2 log-likelihood) for the reduced and full model. Defining $s_N^{(l)} = (D_o^{(l)} - D_1^{(l)})/(p_1 - p_o)$, $s_D^{(l)} = D_1^{(l)}/p_1$, $\nu_N = p_1 - p_o$ and $\nu_D = p_1$. When applied to the likelihood ratio statistic, this procedure can be viewed as an alternative to a more complicated procedure discussed in Meng and Rubin (1992).

6 Discussion

We have proposed a simple procedure for combining mean square errors when an ANOVA model is used in the analysis of multiply imputed data sets. The standard multiple imputation theory is used to derive the approximate F test which can be used in a variety of contexts and is shown through a simulation study to have desirable repeated sampling properties.

The imputations in the example (and also in the simulations) were carried out using the sequential regression framework and the simulation study is also a testament to the validity of this imputation approach in this particular context. There are other methods for imputing the missing values such as General Location model (Shafer (1997))which can be used in this particular context. The combining procedure should be applicable as long as the imputation procedure results in the consistent estimates of the underlying variance components involved in the F-statistics.

The procedure could be used for many other tests involving pivotal quantities having chi-square distributions. Many such extensions have been outlined in Section 5 such as multiple linear regression models, Generalized Estimating Equations and in general any likelihood ratio tests. These procedure need to be evaluated using simulation studies and compared with other alternatives. The simulation study can be expanded in the future to consider these other situations and include other alternatives in the evaluation.

The combining procedure is easy to implement requiring only the relevant sufficient statistics, the mean squares from each completed data analysis. The manipulation of these mean squares can be performed using, for example, a spreadsheet program. Alternatively, a simple macro program, for example, in SAS can be written to implement the combining rule.

References

1. Box, G. E. P. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley.

2. Harris, K. M. and Udry, J. R. (1994). National Longitudinal Study of Adolescent Health (Add Health), 1994-2002[Computer File]. ICPSR21600v4. Ann Arbor, MI: Inter-University Consortium of Political and Social Research [distributor], 2010-12-08.DOI:10.3886/ICPSR21600.

3. Li, K.H., Raghunathan, T. E. and Rubin, D. B. (1991). Large-sample significance levels from multiply imputed data using moment-based statistic and an F-reference distribution, *J. Amer. Statist. Assoc.*, 86, 1065-73.

4. Li, K. H., Meng, X. L., Raghunathan, T. E. and Rubin, D. B. (1992). Significance levels from repeated p-values with multiply imputed data. *Statistica Sinica*, 1, 65-92.

5. Meng, X. L. and Rubin, D. B. (1992). Performing likelihood ratio tests with multiply imputed data set. *Biometrika*, 79, 103-111.

6. Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J. and Solenberger, P. W. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, 27,89-95.

7. Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.

8. Rubin, D. B. and Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J. Amer. Statist. Assoc.*, 81, 366-374.

9. Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

10. Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.