Application and Comparison of Missing Data Methods for Factor Analysis and Multinomial Logistic Regression

Lia Homeister

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Education

University of Washington

2024

Committee:

Oscar L. Olvera Astivia

Elizabeth Sanders

Program Authorized to Offer Degree:

College of Education

©Copyright 2024

Lia Homeister

University of Washington

Abstract

Application and Comparison of Missing Data Methods for Factor Analysis and Multinomial Logistic Regression

Lia Homeister

Chair of the Supervisory Committee:
Oscar L. Olvera Astivia
Department of Measurement and Statistics

Higher education institutions collect a plethora of data on their students each year from admissions, to learning management systems, to satisfaction surveys. Computer software has made collecting data from participants ubiquitous in the college environment. Despite having the ability to collect a large amount of data, missing data values for certain fields remains a challenge for successful data analysis, particularly with surveys covering sensitive topics. Traditionally, missing data would lead to removing observations which makes analysis easier but could compromise estimates. This methodological paper will review and compare conventional and advanced missing data methods of listwise deletion, regression imputation, full information maximum likelihood (FIML) and multiple imputation (MI) in the context of confirmatory factor analysis and multinomial logistic regression analysis. The data set used for these analyses is a campus climate survey conducted at a 2-year college in Washington State. The confirmatory factor analysis will cover the construct of campus climate, while the multinomial logistic

regression will make use of the demographic variables. The results of the present study showed first that, in the confirmatory factor analysis model, the five missing data handling methods resulted in similar loading estimates and fit indices, except that the categorical multiple imputation technique performed better than the other methods in regard to precision of estimates and standard errors. With respect to the multinomial logistic regression model results, use of listwise deletion resulted in a larger confidence interval range for coefficients and higher standard errors than the imputation methods. In short, we recommend use of multiple imputation across all sample sizes, and for larger samples, use of FIML is also appropriate. Listwise deletion is not recommended.

Application and Comparison of Missing Data Methods for Factor Analysis and Multinomial Logistic Regression

Researchers and practitioners who work with survey data have likely experienced issues related to missing data. There are many causes for a dataset to have missing data. For example, a respondent may refuse to answer a question or may unintentionally overlook answering a question. In addition, in longitudinal studies, over time, participant attrition may occur, as people move or drop out of the study. Furthermore, survey design may promote missingness if items include response options, such as “Prefer not to respond” or “Don’t know” or allow skipping items altogether. There are philosophical reasons why a survey may be designed this way. And yet still, almost all statistical methods assume that there are no missing data on the variables included in the analysis (Allison, 2001, p. 1). This is problematic, as assuming more data on the sample is present can underestimate standard errors, artificially lower p-values, etc., which would lead to an increase in Type I error rate (Schafer & Olsen, 1998). For practitioners, this would bias the results of the analysis, which could lead to ineffective, and possibly harmful, interventions for problems.

When looking at missing data, it is important to determine the mechanisms in which the missingness occurred. Missingness can happen in differing ways, which impacts if and how it can be managed. Essentially, when can the way in which missing data occurred be ignored? Rubin (1976) has set up the framework that many researchers follow, defining missingness as Missing Completely at Random, Missing at Random, and Missing Not at Random.

1. Missing Completely at Random (MCAR) – The probability of missing data on a variable is unrelated to the value of the variable itself or to the values of any other variables in the

dataset. Unless it is part of the research design, this scenario is highly unlikely in real data collection.

2. Missing at Random (MAR) – The probability of missing data on a variable is unrelated to the value of the variable, after controlling for other variables in the analysis. Per Allison (2001), when the missingness is MAR, it is considered “ignorable”. The formal expression of MAR is:

$$Pr(Y\ missing|Y, X) = Pr(Y\ missing|X) \quad (1)$$

3. Missing Not at Random (MNAR) – If the missingness is due to the variable itself, then it does not fall into the MCAR or MAR categories. It is MNAR, which is difficult to model and manage (Schafer & Graham, 2002).

Conventional Missingness Handling Methods

Listwise deletion continues to be a common method for handling missing data. In this scenario, the researcher would simply remove the entire observation whenever there are missing data on any variable. This is appealing to researchers, as it can be used for any kind of statistical analysis and does not require special computation. Thus, it is easy and efficient. However, listwise deletion can be a problem, particularly when the sample size is small. As one may presume, removing entire observations with any missing data could greatly reduce the sample size and lower power (Roth, 1994).

Allison (2001) summarizes alternative techniques with their pros and cons, including pairwise deletion, mean imputation, and regression imputation. Pairwise deletion would be most beneficial when the missingness is MCAR and there is a large sample size. However, if not MCAR, the standard errors and tests statistics would be biased, and would have the potential for inconsistency of the estimations of the covariance matrix (Kim & Curry, 1977). For mean imputation, one would substitute for the missing values the mean of the other values. Although this method is easy and straightforward, it would lead to biased variances and covariances and should be avoided. Even in large samples, estimates tend to be biased if the correlation among predictors is strong (Donner, 1982). Regression imputation would use multiple regression to impute the missing values, based on the other values that are observed. This method would provide the best results when missingness is MCAR with a large sample size. However, issues include underestimated standard errors and overestimated test statistics.

Modern Missingness Handling Methods

With the increase in computational power, more advanced statistical techniques, such as full information maximum likelihood (FIML) and multiple imputation (MI), are becoming more popular and can address the issues found in the traditional methods for handling missing data.

FIML uses the maximum likelihood function, which finds the optimal values of estimates of a distribution based on observed data. When there are missing data, one takes the joint probability for the remaining variables. If the missing variables are discrete, this is the joint probability summed over all possible values of the missing variables. If the missing variables are continuous, integrals are used instead of summations. In general, the likelihood function can be written as follows:

$$L(\theta) = \prod_{i=1}^{n} f_i(y_{i1}, y_{i2}, \dots, y_{ik} | \theta) \tag{2}$$

where f_i is the joint probability function for observation *i,* and θ is a set of parameters to be estimated (Allison, 2012). In simulation studies, FIML was found to be superior to conventional methods across MCAR and MAR scenarios in structural equation modeling (Enders & Bandalos, 2001). This method, however, assumes multivariate normality and is not suitable for categorical data.

MI, on the other hand, is quite flexible and can be utilized with categorical data. MI produces approximately unbiased estimates in large samples, is almost efficient, and asymptotically normal (Allison, 2001). MI includes a random component by introducing random variation that inputs missing values, generates multiple data sets with slightly different imputed values. According to the rules set by Rubin (1987), statistical analysis can then be conducted, and the results are combined to get the parameter estimates. For the initial step of imputation, a popular method is to use regression imputation. To account for variance, one must take a random draw from the standard normal distribution as an error term. To prevent low standard errors, this process is iterated several times with different random errors drawn. According to Rubin (1987), five datasets are usually sufficient. The downside of MI is that one will get slightly different estimates, due to the random draws of the introduced error term, and requires pooling the estimates using both within- and between-imputation variances.

This paper will explore a real-world example of an institutional survey on campus climate where, due to the sensitivity of the topics contained, there is a substantial number of items where a response is not mandatory. It will apply various missing data methods and compare estimates using factor analysis and multinomial logistic regression.

Research Goals

Using a college campus climate survey, the present study focuses on achieving the following research goals:

1. Compare conventional and advanced missing data methods, including listwise deletion, regression imputation, full information maximum likelihood, and multiple imputation.

2. Validate the missing data methods with confirmatory factor analysis of the campus climate construct.
3. Validate the missing data methods with multinomial logistic regression for the demographic variables to predict religion.

Methods

Dataset

The data set was acquired from a Washington two-year college resulting from a campus climate survey conducted in 2023. The campus climate survey was a proprietary instrument from the Higher Education Data Sharing (HEDS) Consortium. The survey selected was the 2022-2023 version of the Diversity and Equity Campus Climate Survey. For the higher education context, campus climate is defined as the "current perceptions, attitudes, and expectations that define the institution and its members" (Peterson and Spencer, 1990). Researchers assess campus climate by the perceptions and attitudes of various groups on campus. Thus, it can be specified from different perspectives. More recently the term has been used to refer to racial climate. Hurtado et al. (1999) defines racial climate as a multidimensional construct with four factors: compositional or structural diversity, the psychological dimension of the climate, the behavioral dimension of the climate, and an institution's history and legacy of inclusion or exclusion.

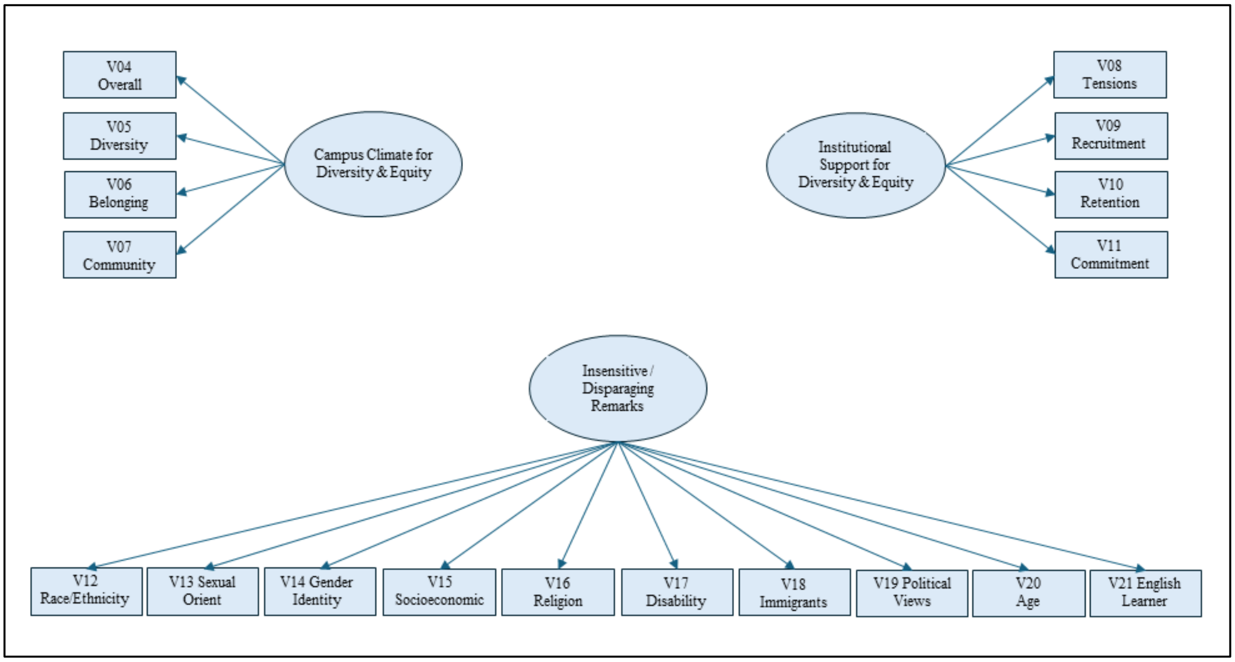
The instrument contains 45 items with 20 campus climate questions and 25 demographic questions. It was distributed broadly as a census using Qualtrics software. Links were sent out via email and mobile communications, while QR codes were included on physical advertisements (i.e., posters, flyers). A raffle to win one of eight $50 gift cards was used as an incentive. Due to the nature of the survey, questions did not require a response and skipping was allowed. Thus, for each item there was a different proportion of missing responses. The responses were collected and re-coded by HEDS and a clean data set was sent to the college via Excel file. Out of a potential 4,685 respondents, 307 people completed the survey.

Variables

The current study compares missing data methods to fulfill two data analyses: (1) confirmatory factor analysis with the campus climate variables and (2) multinomial logistic regression with the demographic variables. According to the third-party's survey codebook, this campus climate survey has four indicators: Campus Climate for Diversity and Equity, Institutional Support for Diversity and Equity, Insensitive/Disparaging Remarks, and Discrimination/Harassment. Due to the sensitivity and personal nature of the last indicator, this study focused on the three former indicators in confirmatory factor analysis. For the indicators of campus climate and institutional support, there were four items each. For the insensitive/disparaging remarks indicator there were 10 items, with each item being reverse coded. A parallel analysis and exploratory factor analysis supported the number of factors and items statistically. Although one item correlated higher to a different factor than proposed by HEDS, the confirmatory factor analysis was conducted based on HEDS' construct. For simplicity of testing missing data methods, three one-factor models were used. The path diagram for this can be found in Figure 2. Thus, 18 items total were used in the factor analysis. All 18 items were 5-point rating scale items. Table A1 contains a full list of the items in the appendix.

Figure 1

Campus Climate Path Diagram



Moreover, in response to an inquiry about religion on campus, demographic variables were analyzed. A question was asked about predicting the number of Muslim students on campus based on the demographic information collected. Religion had 16.61% missing data. According to Pew Research Center's Religious Landscape Study published in 2015, 81% of Muslims in the United States were between the ages of 18 and 49; 65% identify as men, 64% identify as immigrant, and 62% identify voting (or lean toward voting) as a Democrat (Pew Research Center, 2015). As such, the variables of age, gender, political affiliation, and international status were selected as predictors in the model. Table 1 summarizes the variables with the number of categories. Categories were collapsed for religion (from 19 to five), age (from six to three), and gender (from three to two). It should be noted that the data provided to the institution from the HEDS contained some collapsing and re-coding already. As a result, religion was a nominal variable with five categories, age was an ordinal variable with three categories, political affiliation was a nominal variable with three categories, and the remaining variables were nominal with two categories.

Table 1

Descriptive Statistics of Demographic Variables

Variable	*n*	*# Categories*	*% Missing*
Religion	245	5	17%
Age	275	3	6%
Gender	285	2	3%
Politics	175	3	40%
International	266	2	10%

Study 1: Campus Climate Factor Analysis

Descriptive Statistics. Overall, there were 307 participants with varying degrees of missingness in the variables, ranging from 0.00% to 3.26% for campus climate items. Amongst the observations, 277 observations had 0.00% missingness, while at the high end, one observation had 83.33% missing values. Missingness was seen more prevalently in the later items in the survey. All rating-scale items had skewness and their histograms showed ceiling effects. A summary of the descriptive statistics can be found in Table 2.

Table 2

Descriptive Statistics of Campus Climate Variables

Variable	*Factor*	*n*	*M*	*SD*	*Skewness*	*Kurtosis*	*SE*	*% Missing*
V04 Overall	Campus Climate	305	4.15	0.96	-1.29	1.48	0.05	1%
V05 Diversity	Campus Climate	300	4.19	0.92	-1.35	1.96	0.05	2%
V06 Belonging	Campus Climate	305	4.09	1.04	-1.14	0.73	0.06	1%
V07 Community	Campus Climate	304	3.91	1.05	-0.83	0.07	0.06	1%
V08 Tensions	Inst. Support	307	3.79	1.12	-0.80	-0.19	0.06	0%
V09 Recruitment	Inst. Support	305	4.16	0.86	-0.86	0.37	0.05	1%
V10 Retention	Inst. Support	304	3.94	1.00	-0.73	-0.02	0.06	1%
V11 Commitment	Inst. Support	302	4.13	1.03	-1.30	1.35	0.06	2%
V12 Race/Ethnicity	Remarks	299	4.47	0.84	-1.52	1.68	0.05	3%
V13 Sexual Orient	Remarks	300	4.60	0.75	-2.14	4.95	0.04	2%
V14 Gender Identity	Remarks	299	4.50	0.86	-1.93	3.71	0.05	3%
V15 Socioeconomic	Remarks	302	4.50	0.86	-1.72	2.43	0.05	2%
V16 Religion	Remarks	298	4.58	0.75	-2.02	4.60	0.04	3%
V17 Disability	Remarks	300	4.56	0.91	-2.32	5.10	0.05	2%
V18 Immigrants	Remarks	297	4.51	0.89	-1.85	2.86	0.05	3%
V19 Political View	Remarks	298	4.26	1.01	-1.17	0.50	0.06	3%
V20 Age	Remarks	297	4.37	0.93	-1.44	1.49	0.05	3%
V21 English Learner	Remarks	297	4.37	0.93	-1.47	1.67	0.05	3%

Conventional Missing Data Techniques. The current study employed various missing data techniques prior to conducting the confirmatory factor analysis. The intent was to compare the parameter estimates when different missing data methods were used. More specifically, this study explored possible differences between conventional techniques (listwise deletion and regression imputation) and advanced techniques (full information maximum likelihood and multiple imputation). The first missing data technique utilized was listwise deletion, which removes all cases with any amount of missing data across its variables. To do this, the na.omit() function was utilized in R, which dropped the number of observations from 307 to 277, a loss of 9.77% of the data. A new data set without missing values could then be used for the analysis.

The next missing data technique used was regression imputation, which utilized the mice package in R. The mice package, developed by van Buuren and Groothuis-Oudshoorn generates multivariate imputations by chained equations (MICE). It specifies the imputation model on a variable-by-variable basis by a set of conditional densities (van Buuren & Groothuis-Oudshoorn, 2011). To illustrate a traditional technique, within the mice function, the imputation method used was "norm.predict," which assumes a normal distribution. Treating the 5-category items as continuous was a choice in this situation based on research comparing robust continuous and categorical estimation methods (Rhemtulla et al., 2012), where it's supported to treat 5-point rating items as continuous. There were five imputations set per the recommendation of Rubin (Rubin, 1987). The variables' distributions were non-normal. Therefore, later during the confirmatory factor analysis, robust maximum likelihood (MLR) estimation was used to account for non-normality (Satorra & Bentler, 2001). MLR implies the use of Huber-White standard errors and the Yuan-Bentler test statistic.

Full Information Maximum Likelihood (FIML). Full information maximum likelihood (FIML) was used during the confirmatory factor analysis using the lavaan package, specifying the missing argument as "FIML" within the cfa() function. The MLR estimator was used for this as well. According to Little et al. (2013), FIML estimation uses "casewise log-likelihoods to achieve an analogous effect when used to fit a statistical model to incomplete data." Basically, it can use the available information in the dataset based on what is known from the observed data. It is important to note that FIML estimation does not impute missing values, but discreetly estimates model parameters and standard errors using data that are available (Enders, 2001).

Multiple Imputation (MI). Two versions of multiple imputation were tested, one using the MLR estimator, which assumed continuous distribution, and one using the WLSMV estimator. WLSMV stands for weighted least square mean and variance adjusted, which is best to use with categorical data with skewed distributions and small samples (Finney & DiStefano, 2006). For both scenarios, the cfa.mi() function from the semTools package was used with the number of imputations specified as five. There are three basic steps for multiple imputation that are repeated as many times as desired (Allison, 2012):

1. Introduce random variation into the imputation process and generate several data sets.
2. Perform analysis on each of the data sets.
3. Combine the results into a single set of parameter estimates, standard errors, and test statistics.

Confirmatory Factor Analysis. Confirmatory factor analysis was conducted to test the model fit of the three latent dimensions and their items from the survey. Once the missing data techniques were established, three one-factor models were fitted using the lavaan package in R. The three factors and their corresponding items were specified in the survey codebook. The code of the models can be found in the appendix. The factor loadings, standard errors, confidence intervals, R^2, and *p*-values were compared across the five different missing data techniques. Note that although fit indices were initially assessed, due to the focus of this study, determining the effects of the differing missing data methods were emphasized versus attempting to get the best-fit CFA model.

Study 2: Religion Prediction

Descriptive Statistics. Of the five variables, missingness ranged from 4.56% for gender to 39.41% for political affiliation. Of the 307 observations, five had 100% missing data on the demographic variables, and thus were removed from the dataset, with 302 remaining. Furthermore, observations that selected non-binary for gender were removed due to category sparseness (below 3%), which left 294 observations remaining. In addition, due to data sparseness in this study, the eight categories for religion were collapsed to five and the six categories for age were collapsed to three. Table 1 displays the descriptive statistics of the demographic variables, while Tables 3 and 4 show the category proportions for the variables of religion and age after they were collapsed. In addition, contingency tables for each demographic variable can be found in Figure 3.

Table 3

Category Proportions for Religion Variable following Category Collapse

Religion	*n*	%
Atheist	38	13%
Christian	113	38%
Muslim	24	8%
Spiritual	33	11%
Other	37	13%
NA's	49	17%
Total	294	100%

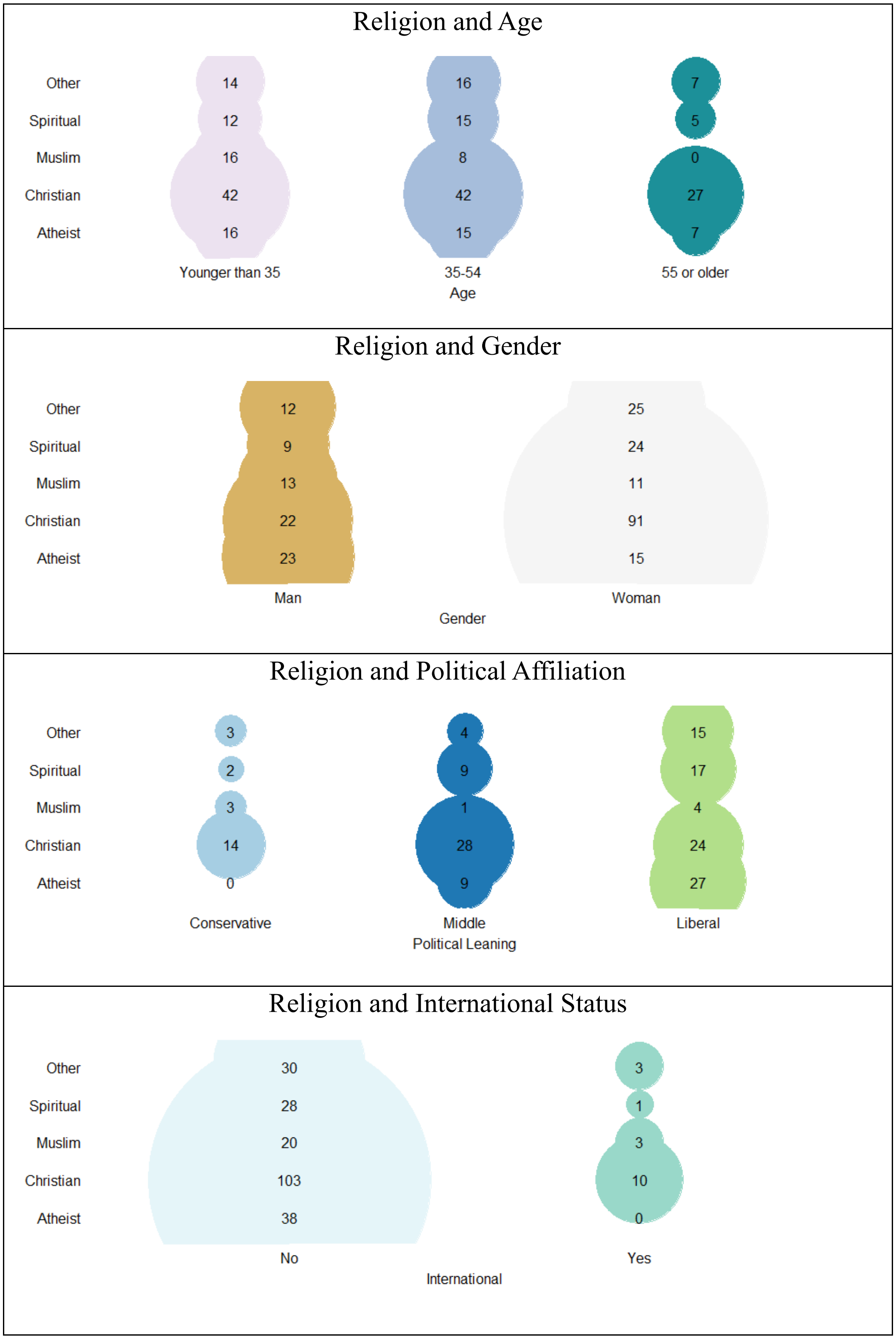
Table 4

Category Proportions for Age Variable following Category Collapse

Age	*n*	*%*
Younger than 35	114	39%
35-54	107	36%
55 or older	54	18%
NA's	19	6%
Total	294	100%

Figure 2

Contingency Tables for Demographic Variables



Missing Data Techniques. Three different missing data methods were tested. One utilized listwise deletion while two utilized multiple imputation. For listwise deletion, the na.omit() function was utilized in R, which dropped the number of observations from 294 to 154, a loss of 47.62% of the data. (Note that these numbers are different than the previous listwise deletion because different variables were used in this study, which contained different levels of missingness.) Two applications of the mice function in R were used—one for imputing the predictor variables only and one for imputing predictor variables and the dependent variable. In the case when only predictor variables were imputed, the data frame was combined with the dependent variable column that contained missing data. The imputation method selected for each variable was specified with the mice function based on variable type. The ordered logit model (polr) was used for age and logistic regression (logreg) for gender and international status. The multinomial logit model (polyreg) was specified for the remaining variables. The methods chosen were based on the variable type and the built-in univariate imputation techniques in the mice package (van Buuren & Groothuis-Oudshoorn, 2011).

Multinomial Logistic Regression. According to Agresti (2019), logistic regression is a special case of generalized linear model where the random component for the outcomes has a binomial distribution (Agresti, 2019, p. 69). The dependent variable is binary, meaning there are two levels or categories. With multinomial logistic regression, one is dealing with multiple categories greater than two. The multicategory generalization of the binomial distribution is the multinomial distribution. The multinomial logit model for nominal response variables simultaneously uses all pairs of categories by specifying the odds of one outcome over the other (Agresti, 2019, p. 159).

Because this study utilizes categorical data in regression analysis, re-coding was necessary so that the variables can be added to the regression models. For gender, political affiliation, and international status variables, treatment (or dummy) coding was applied. This coding method compares each category level to a reference category. For this study, the largest categories were used as the reference categories. For the age variable, orthogonal polynomial coding was applied, due to its ordered nature. This coding scheme looks for linear, quadratic, and cubic trends in the categorical variable (UCLA: Statistical Consulting Group, 2021). For the dependent variable, religion, Christian was set as the reference category due to it being the category with the highest frequency.

The variables were placed in three multinomial logistic regression models using the different missing data applications, with religion as the dependent variable and age, gender, political affiliation, and international status as the predictors. The multinom() function was used from the nnet package in R. The log-odds of each multinomial logistic regression were reported and checked for statistical significance. In addition, the log-odds were transformed into odd-ratios for interpretation. Then probabilities were calculated using the predict() function from the stats package. Those results were placed in a confusion matrix to evaluate the model prediction. Evaluation statistics, such as accuracy, were assessed. Accuracy is defined as the fraction of cases the model correctly predicted (Erickson & Kitamura, 2021).

Results

Study 1: Campus Climate Factor Analysis

Factor models 1 and 2 have only two degrees of freedom, so their ability to detect misfit is limited. Factor model 3 only showed appropriate fit as per the SRMR metric and the robust CFI metric in the categorical multiple imputation model. For this model, the $X^2(35) = 181.86$, $p <$

.001. Although model fit was an issue, the purpose of this study was to review the differences between key statistics when different missing data methods were applied before running analyses. A summary of model fit statistics can be found in Table 5. The next section will discuss the results of each model and compare statistics using the traditional and advanced missing data techniques.

Table 5

Summary of Model Fit for Missing Data Methods Models

Model	Model Fit Statistics					
	X^2	*df*	*p*	Robust CFI	Robust RMSEA	SRMR
Listwise Deletion						
Factor Model 1	5.64	2	.059	0.99	0.15	0.02
Factor Model 2	10.57	2	.005	0.95	0.21	0.06
Factor Model 3	82.48	35	<.001	0.94	0.12	0.05
Regression Imputation						
Factor Model 1	8.67	2	.013	0.97	0.20	0.03
Factor Model 2	9.89	2	.007	0.95	0.19	0.06
Factor Model 3	102.05	35	<.001	0.93	0.13	0.05
FIML						
Factor Model 1	7.93	2	.019	0.98	0.18	0.02
Factor Model 2	9.45	2	.009	0.95	0.19	0.05
Factor Model 3	96.15	35	<.001	0.93	0.12	0.04
MI						
Factor Model 1	8.57	2	.014	0.97	0.19	0.03
Factor Model 2	9.90	2	.007	0.95	0.19	0.06
Factor Model 3	93.60	35	<.001	0.93	0.12	0.05
Categorical MI						
Factor Model 1	24.48	2	<.001	0.99	0.19	0.03
Factor Model 2	29.73	2	<.001	0.98	0.21	0.07
Factor Model 3	181.86	35	<.001	0.96	0.12	0.04

For Factor model 1, listwise deletion produced factor loadings from 0.77 to 0.90. Regression imputation, FIML, and multiple imputation produced similar results. However, categorical multiple imputation showed an increase in factor loadings from 0.88 to 0.92, showing the highest loadings in Factor model 1 of the different missing data methods. Categorical multiple imputation also showed lower standard errors than the other methods, up to 0.05 below the other approaches. This method also demonstrated the highest R^2 values for Factor model 1, from .78 to .85.

Again, for Factor model 2, categorical multiple imputation outperformed the other models. Listwise deletion did a fair job, performing similar to the other missing data methods. The categorical multiple imputation produced factor loadings from 0.62 to 0.90, compared to listwise deletion that produced factor loadings from 0.61 to 0.91. Both advanced missing data methods produced slightly lower standard errors than listwise deletion. The categorical multiple imputation method produced the lowest standard errors from 0.02 to 0.04. Like in Factor model 1, the first four methods produced similar R^2 values, while categorical multiple imputation method produced the highest R^2 values.

For Factor model 3, regression imputation, FIML, and multiple imputation were similar or slightly higher than listwise deletion regarding factor loadings of the 10 items. Like the other factors, categorical multiple imputation had the highest factor loadings for all 10 items. Similarly, the categorical multiple imputation method had the lowest standard errors. Likewise, the R^2 values for the advanced methods tended to be higher than the values for the listwise deletion method, with the categorical multiple imputation method having the highest R^2 values.

A summary comparing different missing data methods can be found in Table 6. All statistics can be found in Tables A2 to A6 in the appendix.

Table 6

Missing Data Methods Comparison of Factor Loadings

Variable	Listwise Deletion			Regression Imputation			FIML			MI			Categorical MI		
	Estimate	*SE*	R^2	*Estimate*	*SE*	R^2	*Estimate*	*SE*	R^2	*Estimate*	*SE*	R^2	*Estimate*	*SE*	R^2
Factor 1 Campus Climate for Diversity & Equity															
V04 Overall	0.87	0.06	.81	0.84	0.06	.77	0.83	0.06	.75	0.84	0.06	.77	0.92	0.01	.85
V05 Diversity	0.77	0.07	.70	0.76	0.06	.70	0.78	0.06	.71	0.76	0.07	.69	0.88	0.02	.78
V06 Belonging	0.90	0.06	.75	0.89	0.06	.75	0.90	0.06	.76	0.89	0.06	.74	0.91	0.01	.82
V07 Community	0.90	0.05	.74	0.90	0.05	.74	0.91	0.05	.74	0.90	0.05	.74	0.91	0.02	.82
Factor 2 Institutional Support for Diversity & Equity															
V08 Tensions	0.61	0.08	.29	0.58	0.07	.27	0.58	0.07	.26	0.58	0.07	.27	0.62	0.04	.39
V09 Recruitment	0.71	0.06	.66	0.71	0.06	.67	0.71	0.06	.68	0.71	0.06	.67	0.88	0.02	.77
V10 Retention	0.91	0.06	.80	0.88	0.05	.78	0.88	0.05	.78	0.89	0.05	.78	0.90	0.02	.82
V11 Commitment	0.70	0.07	.46	0.70	0.07	.46	0.70	0.07	.46	0.70	0.07	.46	0.78	0.03	.60
Factor 3 Insensitive/Disparaging Remarks															
V12 Race/Ethnicity	0.71	0.05	.70	0.69	0.05	.68	0.69	0.05	.67	0.69	0.05	.67	0.88	0.02	.78
V13 Sexual Orient	0.62	0.06	.70	0.62	0.07	.69	0.63	0.07	.70	0.63	0.07	.69	0.93	0.02	.86
V14 Gender Identity	0.70	0.06	.67	0.71	0.06	.68	0.71	0.06	.68	0.71	0.06	.68	0.92	0.02	.84
V15 Socioeconomic	0.68	0.05	.71	0.73	0.06	.73	0.73	0.06	.73	0.73	0.06	.73	0.91	0.02	.83
V16 Religion	0.62	0.06	.74	0.64	0.06	.75	0.65	0.06	.75	0.64	0.06	.75	0.92	0.02	.84
V17 Disability	0.61	0.06	.52	0.66	0.06	.54	0.67	0.07	.54	0.67	0.07	.54	0.85	0.03	.72
V18 Immigrants	0.72	0.06	.65	0.73	0.05	.66	0.74	0.06	.67	0.73	0.05	.66	0.90	0.02	.81
V19 Political View	0.67	0.06	.46	0.69	0.06	.48	0.70	0.06	.48	0.70	0.06	.47	0.78	0.03	.60
V20 Age	0.66	0.05	.56	0.70	0.05	.58	0.70	0.06	.57	0.70	0.06	.58	0.84	0.02	.71
V21 English Learner	0.74	0.06	.65	0.75	0.06	.65	0.75	0.06	.64	0.75	0.06	.65	0.89	0.02	.79

Study 2: Religion Prediction

Three multinomial logistic regressions were run utilizing different missing data techniques. Before each regression can be run, missing values must be removed or imputed. As such, the three techniques used were listwise deletion, multiple imputation of the predictors only, and multiple imputation of all variables. A summary of the classification evaluation statistics can be found in Table 7. Of note, the evaluation statistics for each model were not optimal. However, for the purposes of this study, the effects of the missing data methods were compared.

Table 7

Classification Evaluation Statistics

Method	*Accuracy*	*95% CI*		*NIR*	*p*	*K*
		LL	*UL*			
Method 1: Listwise Deletion	0.54	0.46	0.62	0.42	.001	0.28
Method 2: Predictor Imputation	0.51	0.45	0.58	0.46	.055	0.21
Method 3: All Variable Imputation	0.48	0.42	0.54	0.44	.089	0.17

Method 1: Listwise Deletion. The regression that utilized listwise deletion had an accuracy score of 0.54 (n = 154), p = .001, 95% CI [0.46 to 0.62], the highest of the models. Cohen's Kappa was 0.28, which according to Landis & Koch (1977), would roughly indicate a fair strength of agreement. Looking at the confusion matrix, this regression was able to predict correctly 91% of Christian, 64% of Atheist, 25% of Muslim, and 0% of Spiritual or Other. Table A7 shows the full confusion matrix.

Method 2: Multiple Imputation for Predictors Only. For the other two regression models that utilized multiple imputation, the evaluation metrics vary slightly. For method 2, the accuracy metric was 0.51 (n = 245), p = .055, 95% CI [0.45 to 0.58], with a Cohen's Kappa statistic of 0.21, indicating a fair strength of agreement. The regression model was able to predict correctly 90% of Christian, 61% of Atheist, 4% of Muslim, 0% Spiritual and Other. Table A8 shows the full confusion matrix for this model.

Method 3: Multiple Imputation for All Variables. For method 3, the accuracy score was 0.48 (n = 294), p = .089, 95% CI [0.42 to 0.54], with a Cohen's Kappa statistic of 0.17, indicating slight strength of agreement. The regression model was able to correctly predict 86% Christian, 50% Atheist, 15% Muslim, 0% Spiritual, and 4% Other. Table A9 shows the full confusion matrix. Below is Table 8 with a summary of the sensitivity statistics of the three confusion matrices.

Table 8

Method Comparison of Correct Prediction of Categories

Category	Method 1: Listwise Deletion	Method 2: Predictor Imputation	Method 3: All Variable Imputation
Christian	91%	90%	86%
Atheist	64%	61%	50%
Muslim	25%	4%	15%
Spiritual	0%	0%	0%
Other	0%	0%	4%

Parameter of Interest: Muslim Religion Category. For this study, there was particular interest in predicting Muslim respondents, and thus, all the models were analyzed looking at the Muslim category. When looking at the methods of each missing data technique, the percentage of respondents who identified as Muslim went from 5% of the respondents when missing data is completely removed, to 9-10% when data is imputed from both imputation methods. As the sample size increased from 154 for method 1, to 245 for method 2, and to 294 for method 3, the 95% confidence interval ranges also narrowed, and the standard errors got smaller as the sample size increased.

Reviewing all three methods, for the older age categories relative to those younger than 35-years-old, the odds of moving from Christian to Muslim increased, on average, by a factor close to 0.00, 95% CI [0.00 to 0.00]. For method 1, for men relative to women, the odds of moving from Christian to Muslim increased, on average, by a factor of 6.58, 95% CI [1.16 to 37.52], compared to 4.98, 95% CI [1.87 to 13.29] for method 2 and 3.28, 95% CI [1.28 to 8.38] for method 3. In addition, for method 3, for those who politically identified as "Middle of the Road" relative to being liberal, the odds of moving from Christian to Muslim decreased, on average, by a factor of 0.28, 95% CI [0.09 to 0.90]. The rest of the coefficients were insignificant. The full results can be seen in the Appendix in Tables A10 to A12.

Discussion

Study 1: Campus Climate Factor Analysis

Study 1 applied five different missing data techniques in order to conduct a confirmatory factor analysis based on data gathered from a college campus climate survey. Traditional methods of listwise deletion and regression imputation performed similarly to each other. More surprisingly, some of the advanced methods also performed similarly to the traditional methods. One advanced method, categorical multiple imputation, stood out providing better results for the CFA analysis.

The items in this data set were 5-point rating scales. In the literature, a 5-point rating scale can be treated as continuous or categorical (Rhemtulla et al., 2012). From this study, it appears that treating the data as categorical saw great advantages when utilizing missing data techniques, as opposed to treating it as continuous and linear. It produced higher estimates, R^2 values, lower standard errors, and more narrow confidence intervals. Whether this advantage holds up with longer rating scales is yet to be determined and should be studied further.

On the other hand, the additional computation needs of applying the other methods, may be no better than using listwise deletion. In which case, listwise deletion with its benefits of flexibility and low computational requirements, performed better than expected.

Figure 3

Factor Loadings Estimates for each Variable by Missing Data Method

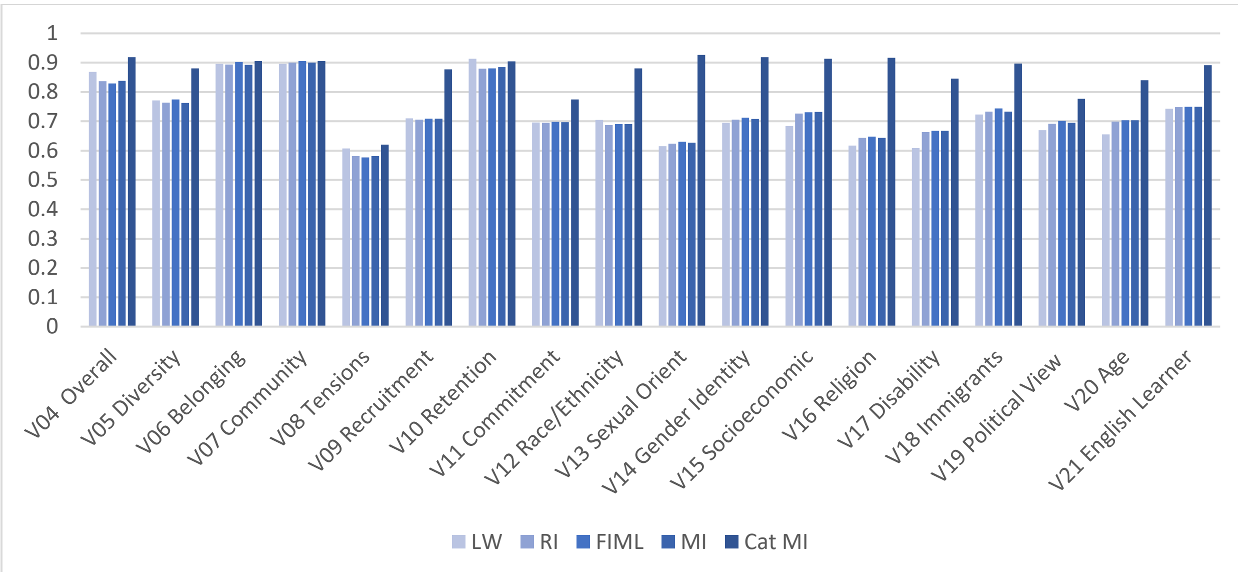
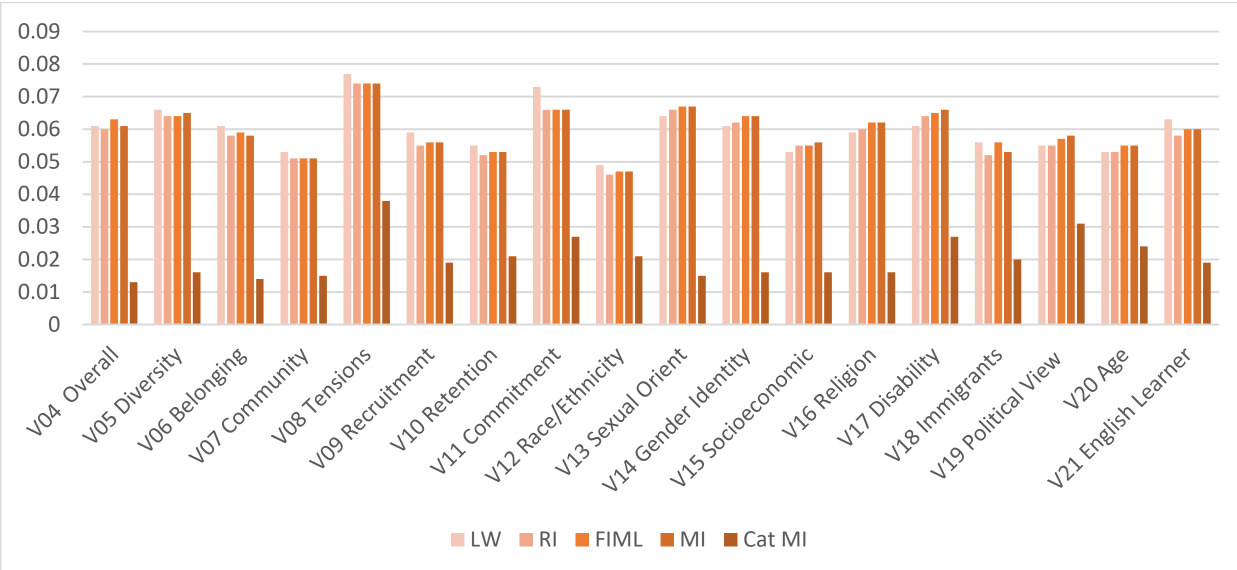


Figure 4

Factor Loadings Standard Errors for each Variable by Missing Data Method



Study 2: Religion Prediction

It is difficult to determine the advantages of using an advanced missing data technique with multinomial logistic regression in this study. Particularly, with the small sample size and the non-fixed nature of multiple imputed values, the statistical results of the parameter of interest were too unstable to make solid conclusions on the benefits, as opposed to simply using listwise deletion. That said, theoretically, the limitations of listwise deletion, such as reducing the sample size and potentially biasing estimates and standard errors, should be considered, and certainly was observed in this study. Additional research on the topic may provide more evidence for missing data solutions with multinomial logistic regression.

Figure 5

Estimates Range in Confidence Intervals for Non-Zero Categories by Missing Data Method

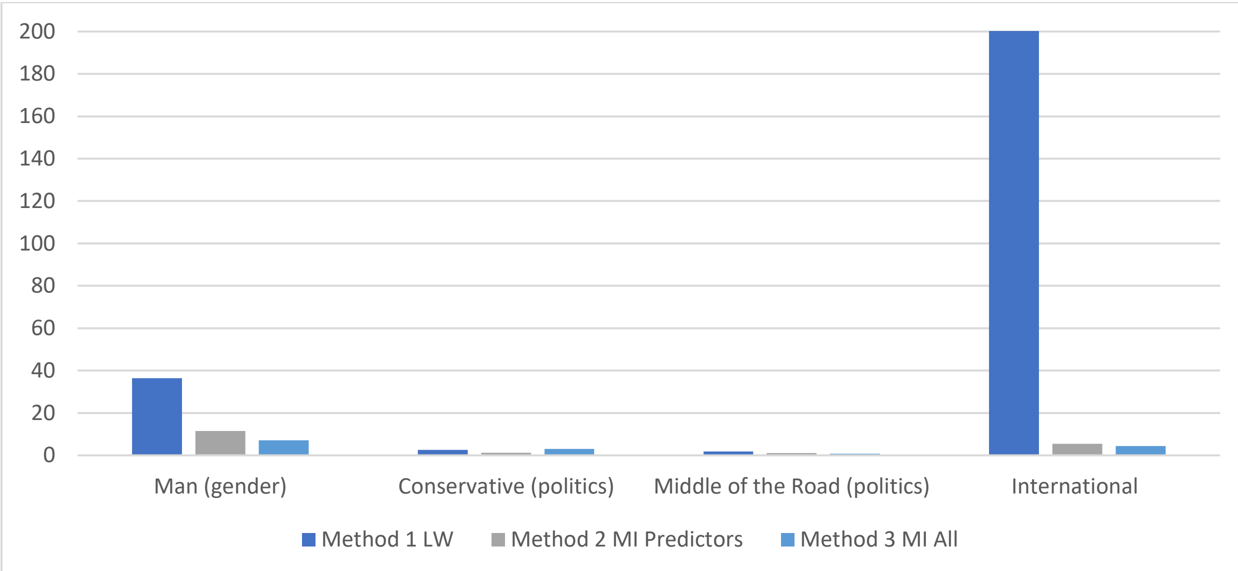
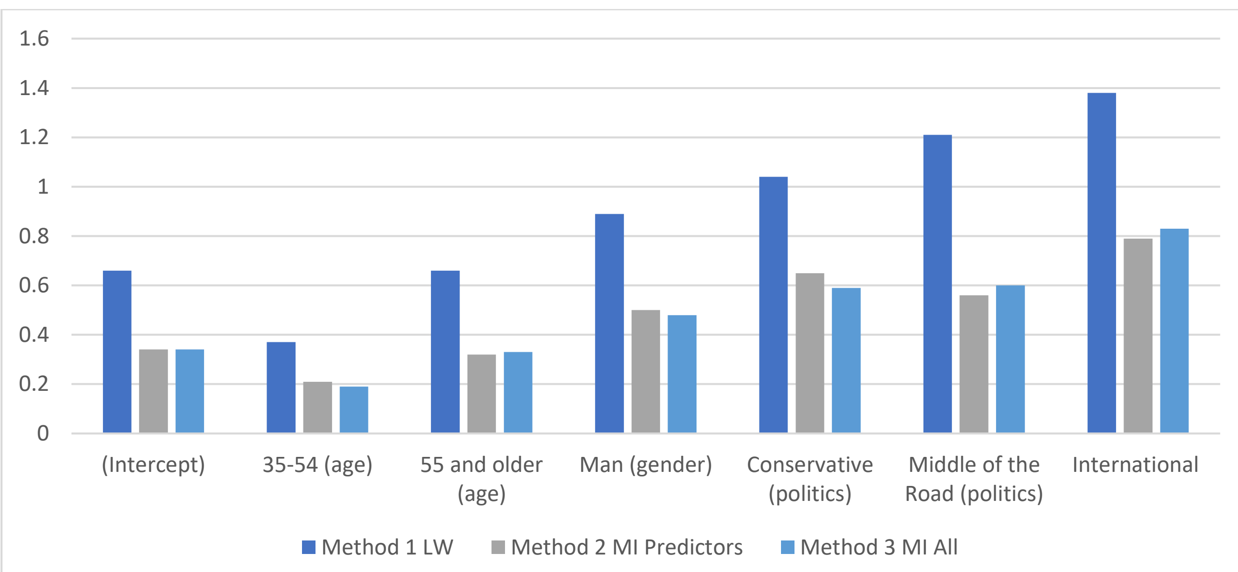


Figure 6

Logistic Regression Standard Errors for Categories by Missing Data Method



Limitations and Future Directions

One major limitation with Study 1 was the skewness of the items. All the items were highly skewed left and displayed ceiling effects. Although advanced missing data methods employed robust maximum likelihood, alternative statistical corrections, such as item response theory (Schoenmakers et al., 2024) or Bayesian joint models (Teimourian et al., 2015), may be explored in the future to improve results.

One major limitation with Study 2 was the overall small sample size (n = 294) and uneven distribution of the dependent variable's categories. The dependent variable, religion, was collapsed to five categories, with 38% identified as Christian and only 8% Muslim, this study's parameter of interest. Future studies should investigate how Firth's penalized logistic regression method may replace traditional maximum likelihood estimation to account for small sample size and small categories (Firth, 1993; Kosmidis & Firth, 2011). Additionally, because of category sparseness, categories needed to be collapsed for three of the variables, religion, age, and gender.

Finally, model fit for either study was not the primary focus of this paper, but should be investigated further to garner more conclusions. It would be beneficial to compare the missing data methods, once proper model fit for both CFA and multinomial logistic regression can be determined.

Conclusion

The contribution of this study is three-fold. First, it provides an empirical example of how to use multiple missing data techniques with two common psychometric analyses in a real-world data set with all its flaws. Second, it validates the weaknesses of the conventional listwise deletion method and provides evidence for using more advanced missing data techniques. It not only provides one way of addressing missing data, but multiple ways. Finally, it provides further evidence of the threats of category sparseness and skewness of 5-point rating items. Overall, education data is ubiquitous, especially in higher education. Data are collected on so many parts of the student journey, and with online learning tools becoming more commonplace, more insights will be available for researchers. However, missing data is a problem experienced heavily in the education setting, particularly around personal or sensitive topics. If researchers want to learn more about these topics and apply more statistical analyses, managing missing data accurately and efficiently is crucial to do effective work.

References

Agresti, A. (1983). A Survey of Strategies for Modeling Cross-Classifications Having Ordinal Variables. *Journal of the American Statistical Association*, *78*(381), 184–198. https://doi.org/10.2307/2287128

Agresti, A. (2019). *An introduction to categorical data analysis*. John Wiley & Sons.

Allison, P. D. (2001). *Missing Data*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07–136. Thousand Oaks, CA: Sage.

Allison, P. D. (2012). Handling Missing Data by Maximum Likelihood. SAS Global Forum https://statisticalhorizons.com/wp-content/uploads/MissingDataByML.pdf

Begg, C. B., & Gray, R. (1984). Calculation of Polychotomous Logistic Regression Parameters Using Individualized Regressions. *Biometrika*, *71*(1), 11–18. https://doi.org/10.2307/2336391

Donner, A. (1982). The Relative Effectiveness of Procedures Commonly Used in Multiple Regression Analysis for Dealing with Missing Values. *The American Statistician, 36*(4), 378–381. https://doi.org/10.1080/00031305.1982.10483055

Enders, C. K. (2001). The Performance of the Full Information Maximum Likelihood Estimator in Multiple Regression Models with Missing Data. *Educational and Psychological Measurement, 61*(5), 713–740. https://doi.org/10.1177/00131640121971482

Enders, C. K., & Bandalos, D. L. (2001). The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Engel, J. (1988). Polytomous logistic regression. *Statistica Neerlandica, 42*(4), 233–252. https://doi.org/10.1111/j.1467-9574.1988.tb01238.x

Erickson, B. J., & Kitamura, F. (2021). Magician's Corner: 9. Performance Metrics for Machine Learning Models. *Radiology. Artificial intelligence*, *3*(3), e200126. https://doi.org/10.1148/ryai.2021200126

Finney, S. J., & DiStefano, C. (2006). Non-Normal and Categorical Data in Structural Equation Modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–314). IAP Information Age Publishing.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika, 80*(1), 27–38. https://doi.org/10.1093/biomet/80.1.27

Hurtado, S., Milem, J. F., Clayton-Pedersen, A. R., & Allen, W. R. (1999). *Enacting diverse learning environments: Improving the campus climate for racial/ethnic diversity.* ASHE/ERIC Higher Education Reports Series.

Kim, J.-O., & Curry, J. (1977). The Treatment of Missing Data in Multivariate Analysis. *Sociological Methods & Research, 6*(2), 215–240. https://doi.org/10.1177/004912417700600206

Kim, R. Y. (2011). Religion and Ethnicity: Theoretical Connections. *Religions, 2*(3), 312–329. https://doi.org/10.3390/rel2030312

Kosmidis, I., & Firth, D. (2011). Multinomial logit bias reduction via the Poisson log-linear model. *Biometrika, 98*(3), 755–759. https://doi.org/10.1093/biomet/asr026

Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2013). On the Joys of Missing Data. *Journal of Pediatric Psychology, 39*(2), 151–162. https://doi.org/10.1093/jpepsy/jst048

Peterson, M. W., & Spencer, M. G. (1990). Understanding academic culture and climate. *New Directions for Institutional Research, 1990(*68), 3–18. https://doi.org/10.1002/ir.37019906803

Pew Research Center. (2015, May 12). *Religious Landscape Study*. Pew Research Center's Religion & Public Life Project. https://www.pewresearch.org/religion/religious-landscape-study/party-affiliation/#party-affiliation

Pew Research Center. (2015, May 12). *Religious Landscape Study*. Pew Research Center's Religion & Public Life Project. https://www.pewresearch.org/religion/religious-landscape-study/religious-tradition/muslim/#demographic-information

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. https://doi.org/10.1037/a0029315

Roth, P. L. (1994). Missing Data: A Conceptual Review for Applied Psychologists. *Personnel Psychology, 47*(3), 537–560. https://doi.org/10.1111/j.1744-6570.1994.tb01736.x

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470316696

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507–514. https://doi.org/10.1007/bf02296192

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147–177. https://doi.org/10.1037//1082-989x.7.2.147

Schafer, J. L., & Olsen, M. K. (1998). Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective. *Multivariate Behavioral Research, 33*(4), 545–571. https://doi.org/10.1207/s15327906mbr3304_5

Schoenmakers, M., Tijmstra, J., Vermunt, J., & Bolsinova, M. (2024). Correcting for Extreme Response Style: Model Choice Matters. *Educational and Psychological Measurement*, *84*(1), 145–170. https://doi.org/10.1177/00131644231155838

Teimourian, M., Baghfalaki, T., Ganjali, M., & Berridge, D. (2015). Joint modeling of mixed skewed continuous and ordinal longitudinal responses: a Bayesian approach. *Journal of Applied Statistics, 42*(10), 2233–2256. https://doi.org/10.1080/02664763.2015.1023557

UCLA: Statistical Consulting Group. (2021, August 22). *R Library Contrast Coding Systems for Categorical Variables*. https://stats.oarc.ucla.edu/r/library/r-library-contrast-coding-systems-for-categorical-variables/

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://doi.org/10.18637/jss.v045.i03

Appendix

Table A1

Campus Climate Items

Item	Response Options				
Please indicate your level of satisfaction with the following at [Institution Name]:					
V04. Overall campus climate	Very satisfied	Generally satisfied	Neither satisfied nor dissatisfied	Generally dissatisfied	Very dissatisfied
V05. The campus experience/environment regarding diversity at [Institution Name]	Very satisfied	Generally satisfied	Neither satisfied nor dissatisfied	Generally dissatisfied	Very dissatisfied
V06. The extent to which you experience a sense of belonging or community at [Institution Name]	Very satisfied	Generally satisfied	Neither satisfied nor dissatisfied	Generally dissatisfied	Very dissatisfied
V07. The extent to which you feel all community members experience a sense of belonging or community at [Institution Name]	Very satisfied	Generally satisfied	Neither satisfied nor dissatisfied	Generally dissatisfied	Very dissatisfied
Please indicate your level of agreement with each of the following statements about [Institution Name].					
V08. The campus environment is free from tensions related to individual or group differences.	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
V09. Recruitment of historically marginalized students, faculty, and staff is an institution priority.	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
V10. Retention of historically marginalized students, faculty, and staff is an institution priority.	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
V11. Senior leadership demonstrates a commitment to diversity and equity on this campus.	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
During your time at [Institution Name], about how often have you heard someone make an insensitive or disparaging remark about:					
V12. People who have a particular racial and/or ethnic identity	Never	Rarely	Sometimes	Often	Very Often
V13. People of a particular sexual orientation	Never	Rarely	Sometimes	Often	Very Often
V14. People of a particular gender or gender identity	Never	Rarely	Sometimes	Often	Very Often
V15. People from a particular socioeconomic background	Never	Rarely	Sometimes	Often	Very Often
V16. People from a particular religious background	Never	Rarely	Sometimes	Often	Very Often
V17. People with a particular disability	Never	Rarely	Sometimes	Often	Very Often
V18. People who are immigrants	Never	Rarely	Sometimes	Often	Very Often
V19. People with a particular political affiliation/view	Never	Rarely	Sometimes	Often	Very Often
V20. People of a particular age or generation	Never	Rarely	Sometimes	Often	Very Often
V21. People for whom English is not their native language	Never	Rarely	Sometimes	Often	Very Often

Table A2

Factor Loadings using Listwise Deletion

Variable	*Estimate*	*SE*	*95% CI* [LL, UL]	R^2	*p*
Factor 1					
V04 Overall	0.87	0.06	[0.75, 0.99]	.81	<.001
V05 Diversity	0.77	0.07	[0.64, 0.90]	.70	<.001
V06 Belonging	0.90	0.06	[0.78, 1.02]	.75	<.001
V07 Community	0.90	0.05	[0.79, 1.00]	.74	<.001
Factor 2					
V08 Tensions	0.61	0.08	[0.46, 0.76]	.29	<.001
V09 Recruitment	0.71	0.06	[0.60, 0.83]	.66	<.001
V10 Retention	0.91	0.06	[0.81, 1.02]	.80	<.001
V11 Commitment	0.70	0.07	[0.55, 0.84]	.46	<.001
Factor 3					
V12 Race/Ethnicity	0.71	0.05	[0.61, 0,80]	.70	<.001
V13 Sexual Orient	0.62	0.06	[0.49, 0.74]	.70	<.001
V14 Gender Identity	0.70	0.06	[0.58, 0.82]	.67	<.001
V15 Socioeconomic	0.68	0.05	[0.58, 0.79]	.71	<.001
V16 Religion	0.62	0.06	[0.50, 0.73]	.74	<.001
V17 Disability	0.61	0.06	[0.49, 0.73]	.52	<.001
V18 Immigrants	0.72	0.06	[0.61, 0.83]	.65	<.001
V19 Political View	0.67	0.06	[0.56, 0.78]	.46	<.001
V20 Age	0.66	0.05	[0.55, 0.76]	.56	<.001
V21 English Learner	0.74	0.06	[0.62, 0.87]	.65	<.001

Table A3

Factor Loadings using Regression Imputation

Variable	*Estimate*	*SE*	*95% CI* [LL, UL]	R^2	*p*
Factor 1					
V04 Overall	0.84	0.06	[0.72, 0.96]	.77	<.001
V05 Diversity	0.76	0.06	[0.64, 0.89]	.70	<.001
V06 Belonging	0.89	0.06	[0.78, 1.01]	.75	<.001
V07 Community	0.90	0.05	[0.80, 1.00]	.74	<.001
Factor 2					
V08 Tensions	0.58	0.07	[0.44, 0.73]	.27	<.001
V09 Recruitment	0.71	0.06	[0.60, 0.81]	.67	<.001
V10 Retention	0.88	0.05	[0.78, 0.98]	.78	<.001
V11 Commitment	0.70	0.07	[0.57, 0.82]	.46	<.001
Factor 3					
V12 Race/Ethnicity	0.69	0.05	[0.60, 0.78]	.68	<.001
V13 Sexual Orient	0.62	0.07	[0.50, 0.75]	.69	<.001
V14 Gender Identity	0.71	0.06	[0.58, 0.83]	.68	<.001
V15 Socioeconomic	0.73	0.06	[0.62, 0.84]	.73	<.001
V16 Religion	0.64	0.06	[0.53, 0.76]	.75	<.001
V17 Disability	0.66	0.06	[0.54, 0.79]	.54	<.001
V18 Immigrants	0.73	0.05	[0.63, 0.83]	.66	<.001
V19 Political View	0.69	0.06	[0.59, 0.80]	.48	<.001
V20 Age	0.70	0.05	[0.60, 0.80]	.58	<.001
V21 English Learner	0.75	0.06	[0.63, 0.86]	.65	<.001

Table A4

Factor Loadings using FIML

Variable	*Estimate*	*SE*	*95% CI* [LL, UL]	R^2	*p*
Factor 1					
V04 Overall	0.83	0.06	[0.71, 0.95]	.75	<.001
V05 Diversity	0.78	0.06	[0.65, 0.90]	.71	<.001
V06 Belonging	0.90	0.06	[0.79, 1.02]	.76	<.001
V07 Community	0.91	0.05	[0.81, 1.01]	.74	<.001
Factor 2					
V08 Tensions	0.58	0.07	[0.43, 0.72]	.26	<.001
V09 Recruitment	0.71	0.06	[0.60, 0.82]	.68	<.001
V10 Retention	0.88	0.05	[0.78, 0.98]	.78	<.001
V11 Commitment	0.70	0.07	[0.57, 0.83]	.46	<.001
Factor 3					
V12 Race/Ethnicity	0.69	0.05	[0.60, 0.78]	.67	<.001
V13 Sexual Orient	0.63	0.07	[0.50, 0.76]	.70	<.001
V14 Gender Identity	0.71	0.06	[0.59, 0.84]	.68	<.001
V15 Socioeconomic	0.73	0.06	[0.62, 0.84]	.73	<.001
V16 Religion	0.65	0.06	[0.53, 0.77]	.75	<.001
V17 Disability	0.67	0.07	[0.54, 0.80]	.54	<.001
V18 Immigrants	0.74	0.06	[0.64, 0.85]	.67	<.001
V19 Political View	0.70	0.06	[0.59, 0.81]	.48	<.001
V20 Age	0.70	0.06	[0.60, 0.81]	.57	<.001
V21 English Learner	0.75	0.06	[0.63, 0.87]	.64	<.001

Table A5

Factor Loadings using Multiple Imputation

Variable	*Estimate*	*SE*	*95% CI* [LL, UL]	R^2	*p*
Factor 1					
V04 Overall	0.84	0.06	[0.72, 0.96]	.77	<.001
V05 Diversity	0.76	0.07	[0.63, 0.89]	.69	<.001
V06 Belonging	0.89	0.06	[0.78, 1.01]	.74	<.001
V07 Community	0.90	0.05	[0.80, 1.00]	.74	<.001
Factor 2					
V08 Tensions	0.58	0.07	[0.44, 0.73]	.27	<.001
V09 Recruitment	0.71	0.06	[0.60, 0.82]	.67	<.001
V10 Retention	0.89	0.05	[0.78, 0.99]	.78	<.001
V11 Commitment	0.70	0.07	[0.57, 0.83]	.46	<.001
Factor 3					
V12 Race/Ethnicity	0.69	0.05	[0.60, 0.78]	.67	<.001
V13 Sexual Orient	0.63	0.07	[0.50, 0.76]	.69	<.001
V14 Gender Identity	0.71	0.06	[0.58, 0.83]	.68	<.001
V15 Socioeconomic	0.73	0.06	[0.62, 0.84]	.73	<.001
V16 Religion	0.64	0.06	[0.52, 0.77]	.75	<.001
V17 Disability	0.67	0.07	[0.54, 0.80]	.54	<.001
V18 Immigrants	0.73	0.05	[0.63, 0.84]	.66	<.001
V19 Political View	0.70	0.06	[0.58, 0.81]	.47	<.001
V20 Age	0.70	0.06	[0.60, 0.81]	.58	<.001
V21 English Learner	0.75	0.06	[0.63, 0.87]	.65	<.001

Table A6

Factor Loadings using Categorical Multiple Imputation

Variable	*Estimate*	*SE*	*95% CI* [LL, UL]	R^2	*p*
Factor 1					
V04 Overall	0.92	0.01	[0.89, 0.94]	.85	<.001
V05 Diversity	0.88	0.02	[0.85, 0.91]	.78	<.001
V06 Belonging	0.91	0.01	[0.88, 0.93]	.82	<.001
V07 Community	0.91	0.02	[0.88, 0.94]	.82	<.001
Factor 2					
V08 Tensions	0.62	0.04	[0.55, 0.70]	.39	<.001
V09 Recruitment	0.88	0.02	[0.84, 0.91]	.77	<.001
V10 Retention	0.90	0.02	[0.86, 0.95]	.82	<.001
V11 Commitment	0.78	0.03	[0.72, 0.83]	.60	<.001
Factor 3					
V12 Race/Ethnicity	0.88	0.02	[0.84, 0.92]	.78	<.001
V13 Sexual Orient	0.93	0.02	[0.90, 0.96]	.86	<.001
V14 Gender Identity	0.92	0.02	[0.89, 0.95]	.84	<.001
V15 Socioeconomic	0.91	0.02	[0.88, 0.94]	.83	<.001
V16 Religion	0.92	0.02	[0.88, 0.95]	.84	<.001
V17 Disability	0.85	0.03	[0.79, 0.90]	.72	<.001
V18 Immigrants	0.90	0.02	[0.86, 0.94]	.81	<.001
V19 Political View	0.78	0.03	[0.72, 0.84]	.60	<.001
V20 Age	0.84	0.02	[0.79, 0.89]	.71	<.001
V21 English Learner	0.89	0.02	[0.85, 0.93]	.79	<.001

Table A7

Confusion Matrix for Method 1: Listwise Deletion

Predicted	Actual				
	Christian	Atheist	Muslim	Spiritual	Other
Christian	**91%**	36%	50%	80%	62%
Atheist	9%	**64%**	25%	20%	33%
Muslim	0%	0%	**25%**	0%	0%
Spiritual	0%	0%	0%	**0%**	5%
Other	0%	0%	0%	0%	**0%**

Table A8

Confusion Matrix for Method 2: Predictor Imputation

Predicted	Actual				
	Christian	Atheist	Muslim	Spiritual	Other
Christian	**90%**	39%	58%	79%	76%
Atheist	9%	**61%**	33%	21%	24%
Muslim	1%	0%	**4%**	0%	0%
Spiritual	0%	0%	0%	**0%**	0%
Other	0%	0%	4%	0%	**0%**

Table A9

Confusion Matrix for Method 3: All Variable Imputation

Predicted	Actual				
	Christian	Atheist	Muslim	Spiritual	Other
Christian	**86%**	50%	70%	88%	79%
Atheist	9%	**50%**	11%	12%	13%
Muslim	4%	0%	**15%**	0%	4%
Spiritual	0%	0%	0%	**0%**	0%
Other	1%	0%	4%	0%	**4%**

Table A10

Multinomial Logistic Regression – Muslim Outcome for Method 1: Listwise Deletion

	Odds Ratio Estimate	Odds Ratio 95% CI		Log-Odds Estimate	*SE*	*p*
		LL	*UL*			
(Intercept)	<0.01	<0.01	<0.01	-8.71	0.66	<.001
35-54 (age)	<0.01	<0.01	<0.01	-14.30	0.37	<.001
55 and older (age)	<0.01	<0.01	<0.01	-7.27	0.66	<.001
Man (gender)	6.58	1.16	37.52	1.88	0.89	.034
Conservative (politics)	0.33	0.04	2.55	-1.11	1.04	.288
Middle of the Road (politics)	0.18	0.02	1.88	-1.74	1.21	.151
International	13.63	0.92	201.92	2.61	1.38	.058

Note. For the Religion variable, Christian is the reference category. For Age, Gender, Politics, and International variables, Younger than 35, Woman, Liberal, and non-international are the reference categories.

Table A11

Multinomial Logistic Regression – Muslim Outcome for Method 2: Predictor Imputation

	Odds Ratio Estimate	Odds Ratio 95% CI		Log-Odds Estimate	*SE*	*p*
		LL	*UL*			
(Intercept)	<0.01	<0.01	<0.01	-8.16	0.34	<.001
35-54 (age)	<0.01	<0.01	<0.01	-14.52	0.21	<.001
55 and older (age)	<0.01	<0.01	<0.01	-7.89	0.32	<.001
Man (gender)	4.98	1.87	13.29	1.61	0.50	.001
Conservative (politics)	0.39	0.11	1.39	-0.94	0.65	.148
Middle of the Road (politics)	0.41	0.14	1.24	-0.89	0.56	.116
International	1.20	0.26	5.65	0.18	0.79	.817

Note. For the Religion variable, Christian is the reference category. For Age, Gender, Politics, and International variables, Younger than 35, Woman, Liberal, and non-international are the reference categories.

Table A12

Multinomial Logistic Regression – Muslim Outcome for Method 3: All Variable Imputation

	Odds Ratio Estimate	Odds Ratio 95% CI		Log-Odds Estimate	*SE*	*p*
		LL	*UL*			
(Intercept)	<0.01	<0.01	<0.01	-9.87	0.34	<.001
35-54 (age)	<0.01	<0.01	<0.01	-18.24	0.19	<.001
55 and older (age)	<0.01	<0.01	<0.01	-9.71	0.33	<.001
Man (gender)	3.28	1.28	8.38	1.19	0.48	.013
Conservative (politics)	1.04	0.33	3.30	0.04	0.59	.945
Middle of the Road (politics)	0.28	0.09	0.90	-1.28	0.60	.033
International	0.90	0.18	4.59	-0.10	0.83	.903

Note. For the Religion variable, Christian is the reference category. For Age, Gender, Politics, and International variables, Younger than 35, Woman, Liberal, and non-international are the reference categories.