A Comparison of Imputation Strategies for Ordinal Missing Data on Likert Scale Variables

Wei Wu and Fan Jia

University of Kansas

Craig Enders

Arizona State University

Supplementary Material

Appendix A

R code for the normal data model and logistic regression model approaches for the empirical example.

```
############# Multiple Imputation using Amelia/mice
##      install.packages("Amelia")
##      install.packages("mice")
##      install.packages("psych")
##      install.packages("lavaan")
library(Amelia)
library(mice)

## impMeth==1 observed continuous (Amelia)
## impMeth==2 naive rounding (Amelia)
## impMeth==3 multinomial logistic (mice)
## impMeth==4 proportional odds (mice)

datMiss <- read.table("qoc_for_R.dat", header=T)
impMeth <- 1
nImp <- 50


if (impMeth==1){
```

```
	##impute
	imp <- amelia(datMiss, m=nImp, p2s=0, empri =.01*nrow(datMiss))
	##extract imputed data
	datImpd <- list()
	for (i in 1:nImp){
		datImpd[[i]] <- imp$imputations[[i]]
	}

} else if (impMeth==2){
	##impute
	imp <- amelia(datMiss, m=nImp, p2s=0, empri =.01*nrow(datMiss))
	##extract imputed data
	b1 <- seq(1.5, (7-0.5), 1)
	b2 <- seq(1.5, (5-0.5), 1)
	datImpd <- list()
	for (i in 1:nImp){
		datImpd[[i]] <- imp$imputations[[i]]
		for (j in 2:9) {
			datImpd[[i]][,j] <- cut(datImpd[[i]][,j], breaks=c((min(datImpd[[i]][,j])-
999), b1, (max(datImpd[[i]][,j])+999)),
			include.lowest = TRUE, labels=FALSE)
		}
		for (j in 10:20)
			datImpd[[i]][,j] <- cut(datImpd[[i]][,j], breaks=c((min(datImpd[[i]][,j])-
999), b2, (max(datImpd[[i]][,j])+999)),
			include.lowest = TRUE, labels=FALSE)
		}

} else if (impMeth==3){
	##impute
	datMiss[,1:20] <- lapply(datMiss[,1:20] , as.factor)
	imp <- mice(datMiss, meth=c("logreg", rep("polyreg",19)), m=nImp ,print=FALSE)
	##extract imputed data
	imputations <- complete(imp, "long")
	datImpd <- list()
	for (i in 1:nImp){
		datImpd[[i]] <- imputations[imputations$.imp==i, -1:-2]
		nnd1 <- lapply(datImpd[[i]], as.character)
		nnd2 <- lapply(nnd1, as.numeric)
		datImpd[[i]] <- do.call("cbind", nnd2)
	}

} else if (impMeth==4){
	##impute
	datMiss[,1:20] <- lapply(datMiss[,1:20] , as.ordered)
	imp <- mice(datMiss, meth=c("logreg", rep("polr",19)), m=nImp ,print=FALSE)
```

```
        ##extract imputed data
        imputations <- complete(imp, "long")
        datImpd <- list()
        for (i in 1:nImp){
                datImpd[[i]] <- imputations[imputations$.imp==i, -1:-2]
                nnd1 <- lapply(datImpd[[i]], as.character)
                nnd2 <- lapply(nnd1, as.numeric)
                datImpd[[i]] <- do.call("cbind", nnd2)
        }
}

for (m in 1:nImp){
write.table(datImpd[[m]], file=paste0("impMeth",impMeth ,"_", m, ".dat"), row.names=F,
col.names=F)
}
```

Appendix B

Mplus code for the latent variable approach for the empirical example.

```
TITLE: multiple imputation
DATA: FILE IS qoc_for_mplus.dat;
VARIABLE: NAMES = gender a1-a8 b1-b11;
USEVARIABLES = gender a1-a8 b1-b11;

MISSING = ALL (-999);

DATA IMPUTATION:

IMPUTE = gender (c) a1-a8 (c) b1-b11 (c);

NDATASETS = 50;

SAVE = impMethMplus_*.dat;

ANALYSIS: TYPE = BASIC;
```