Midterm

Jordan Roberts

Sunday, March 15, 2015

Table 1: Table of OLS Function Output

VARIABLE	No Missing	Listwise Deletion	Amelia
ELF	2.01	0.81	0.33
	(0.63)	(0.44)	(0.45)
Democracy	-0.10	-0.16	-0.29
	(0.08)	(0.05)	(0.038)
(intercept)	-0.22	0.67	1.94
	(0.85)	(0.57)	(0.46)
R-Squared	0.66	0.55	0.30

As can be seen from *Table 1*, the results differ considerably across the model with no missing data, the model with listwise deletion of missing data, and the model with imputed values for missing data. The ELF variable's magnitude is substantially larger in the model with no missing data, and this is the only model in which it is significant at conventional degrees of confidence. This effect is reversed for the Democracy variable, which takes on a coefficient farther from zero in the two models with missing data and is only significant for those models. The intercept is larger for the imputed data, and is only significant in that model. The Amelia estimates are farther from the "true" estimates than the listwise deletion estimates in the case of each variable (could this partially be because the artificial way missingness was introduced to the data was random, meaning the data is missing completely at random?). The proportion of variance in the dependent variable explained is highest in the full model, and higher in the listwise deletion model than the Amelia model.

```
### HERE IS MY CODE

# start with a clean workspace
rm(list=ls())

# function to load packages
loadPkg=function(toLoad){
  for(lib in toLoad){
    if(! lib %in% installed.packages()[,1])
    { install.packages(lib, repos='http://cran.rstudio.com/') }
    suppressMessages( library(lib, character.only=TRUE) ) }
}

# load libraries
packs=c('lmtest', 'Amelia', 'xtable')
loadPkg(packs)

# the function
ols = function(formula, data, impute=FALSE){
  # load var names
    dv = all.vars(formula)[1]
    ivs = all.vars(formula)[ 2:length(all.vars(form)) ]
  # impute missing if impute==TRUE
    if(impute==TRUE){
      data = amelia(x=data[,c(dv, ivs)], m=1)
```

```
    data = data$imp$imp1
  }
# listwise delete missing if impute==FALSE
  if(impute==FALSE){
    data = na.omit(data)
  }
# obtain data and general parameters from inputs
  y = data[,dv]
  x = data.matrix(cbind(1, data[,ivs]))
  n = nrow(x)
  k = length(ivs)
  df = n-k-1
# estimate model parameters
  # coefficients
    xTx = t(x) %*% x
    xTy = t(x) %*% y
    beta = solve(xTx) %*% xTy
  # standard errors
    yHat = x %*% beta
    sigma2 = sum((y - (yHat))^2)/df
    varcov = (sigma2 * solve(xTx))
    serrors = sqrt(diag(varcov))
  # confidence intervals
    upper95 = beta + (1.96*serrors)
    lower95 = beta - (1.96*serrors)
  # t-statistics and p-values
    tstats = beta/serrors
    pvals = 1.96*pt(abs(tstats), df, lower.tail=FALSE)
  # model performance metrics
    R2 = 1 - sum((yHat - mean(y))^2)/sum((y - mean(y))^2)
    MSreg = sum(yHat^2)/(k-1)
    MSres = sum((y - (yHat))^2)/(n-k)
    Fstat = (MSreg)/(MSres)
    FF = pf(Fstat, k-1, n-k, lower.tail=FALSE)
# create output
  # coefficients
    coefficients = cbind(beta, serrors, tstats, pvals, upper95, lower95)
    coefficients = as.matrix(coefficients)
    rownames(coefficients)[1] = "(Intercept)"
    colnames(coefficients) = c("Estimate", "Std. Error", "T-Statistic", "P-Value", "Lower 95% CI", "U
  # varcov
    varcov = as.matrix(varcov)
    rownames(varcov)[1] = "(Intercept)"
    colnames(varcov)[1] = "(Intercept)"
  # Rsq
    Rsq = as.numeric(R2)
  # Fstat
    Fstatistic = as.character(round(Fstat, digits = 3))
    Fpv = as.character(round(FF, digits = 3))
    df1 = as.character(k-1)
    df2 = as.character(n-k)
    Fstat = paste0(c("F-statistic: ", Fstatistic, " on ", df1, " and ",
            df2, " DF, p-value: ", Fpv), collapse="")
```

```
    # Keep them
      output = list(coefficients, varcov, Rsq, Fstat)
      names(output) = c("coefficients", "varcov", "Rsq", "Fstat")
      return(output)
}

# bring in data
load("D:/Dropbox/MLE Homeworks/Midterm/midTermData.rda")

# set a seed
set.seed(6886)

# set up the model formula
form = formula(gini_net_std ~ ELF_ethnic + polity2)

# run models
model = ols(formula=form, data=data)
modelListDel = ols(formula = form, data=dataMiss)
modelAmelia = ols(formula = form, data=dataMiss, impute=TRUE)

# create matrix for table
r0 = cbind("VARIABLE", "No Missing", "Listwise Deletion", "Amelia")

r1 = cbind(rownames(model$coefficients)[2], model$coefficients[2,1], modelListDel$coefficients[2,1], mo
r2 = cbind( "", model$coefficients[2,2], modelListDel$coefficients[2,2], modelAmelia$coefficients[2,2])

r3 = cbind(rownames(model$coefficients)[3], model$coefficients[3,1], modelListDel$coefficients[3,1], mo
r4 = cbind( "", model$coefficients[3,2], modelListDel$coefficients[3,2], modelAmelia$coefficients[3,2])

r5 = cbind(rownames(model$coefficients)[1], model$coefficients[1,1], modelListDel$coefficients[1,1], mo
r6 = cbind( "", model$coefficients[1,2], modelListDel$coefficients[1,2], modelAmelia$coefficients[1,2])

r7 = cbind("R-Squared", model$Rsq, modelListDel$Rsq, modelAmelia$Rsq)

table = rbind(r0, r1, r2, r3, r4, r5, r6, r7)
table = as.matrix(table)
xtable(table)
```