Midterm Assignment (Polsci 733)

Jaewon Chung

March 17, 2015

1 Discussion about the coeeficient estimates

I created a function that will conduct an OLS regression. To discuss the difference about coefficient estimates from models, I run three models by using the function I made. The first model is called original model. Also, the second model and third model are named listwise deletion model and imputation model respectively.

According to Table 1, the form of original model is

$$gini_net_std = -0.220 + 2.010 * ELF_ethnic0.096 * polity2.$$

This equation means that if ELF_ethnic increases 1unit, then gini_net_std increases 2.010 and if polity2 increases 1 unit, then gini_net_std decreases 0.096. Also, according to Table 1, p-value of polity2 is 0.235 that is greater than alpha = 0.05, so we consider whether we delete this variable.

According to Table 2, the form of listwise model is

$$gini_net_std = 0.669 + 0.805 * ELF_ethnic0.165 * polity2.$$

This equation means that if ELF_ethnic increases 1unit, then gini_net_std increases 0.805 and if polity2 increases 1 unit, then gini_net_std decreases 0.165. Also, according to Table 2, p value of ELF_ethnic is 0.078 that is greater than alpha = 0.05, so we consider whether we delete this variable.

Table 1: Coefficients of Original Model

	Estimate	Std.Error	T-Statistic	P-Value	Lower 95% CI	Upper 95% CI
"(Intercept)"	-0.220	0.854	-0.258	0.798	-8.808	8.368
Ethnic Frac.	2.010	0.631	3.184	0.003	-6.577	10.598
Demo. Score	-0.096	0.080	-1.204	0.235	-8.684	8.492

According to Table 3, the form of listwise model is

$$gini_net_std = 1.938 + 0.334 * ELF_ethnic 0.287 * polity2.$$

This equation means that if ELF_ethnic increases 1unit, then gini_net_std increases 0.334 and if polity2 increases 1 unit, then gini_net_std decreases 0.287. Also, according to Table 3, p value of ELF_ethnic is 0.464 that is greater than alpha = 0.05, so we consider whether we delete this variable.

Comparing the original model and imputation model, even though polity2 variable is not significant at the .05 level in the original model, polity2 variable in imputation model is significant at the .05 level, but ELF_ethnic variable in imputation model is not significant at the .05 level. Also, the length of confidence interval of any variables in the original model is larger than that in imputation model. Comparing listwise deletion model and imputation model, ELF_ethnic variable is not significant at the .05 level in both models. The length of confidence interval of any variables in listwise deletion model is smaller than that in imputation model. Comparing the original model and listwise deletion model, even though polity2 variable in the original model is not significant at the .05 level, polity2 variable in listwise deletion model is significant at the .05 level, but ELF_ethnic variable is not significant at the .05 level. Also, the length of confidence interval of any variables in listwise deletion model is smaller than that in the original model.

Table 2: Coefficients of Listwise Deletion Model

	Estimate	Std.Error	T-Statistic	P-Value	Lower 95% CI	Upper 95% CI
"(Intercept)"	0.669	0.569	1.175	0.247	-2.359	3.698
Ethnic Frac.	0.805	0.444	1.816	0.078	-2.223	3.834
Demo. Score	-0.165	0.052	-3.157	0.003	-3.193	2.864

Table 3: Coefficients of Imputation Model

	Estimate	Std.Error	T-Statistic	P-Value	Lower 95% CI	Upper 95% CI
"(Intercept)"	1.938	0.460	4.215	0.0001	-1.912	5.788
Ethnic Frac.	0.334	0.452	0.739	0.464	-3.516	4.185
Demo. Score	-0.287	0.038	-7.493	0	-4.137	3.563

Appendix: R codes for Midterm Assignment

0. Setting Up the Workspace

```
# Set up workspace
rm(list=ls())
setwd('C:/Users/Jaewon Chung/Desktop/Duke University/15 Spring/MLE/HWMID/')
set.seed(6886)

# 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('foreign', 'lmtest', 'sandwich', 'Amelia', 'sbgcop')
loadPkg(packs)
```

1. Creating an OLS function

```
# Load data
load('midTermData.rda')

# Creating a function
ols <- function(formula, data, impute=FALSE){


if(impute){
  dataAmelia = amelia(x=data, m=1)
  names(dataAmelia$imp)
  data=dataAmelia$imp$imp1
  miss=FALSE
}

data=na.omit(data)

# Retrieve vars from formula input
dv = all.vars(form)[1]
ivs = all.vars(form)[ 2:length(all.vars(form)) ]

y = data[,dv]
x = data.matrix(cbind('(Intercept)', data[,ivs]))
i = diag(1, nrow=nrow(x), ncol=ncol(x))


# General parameters
n = nrow(x)                   # Number of observations
```

```r
p = length(ivs)              # Number of parameters


# Calculate coefficient estimates
xy = t(x) %*% y              # x`y
xxi = solve(t(x) %*% x)      # (x`x)^-1
h = x %*% xxi %*% t(x)       # hat mtrix of x
i = diag(1, nrow=n, ncol=n)
b = as.vector(xxi %*% xy)    # estimated coefficients

# Calculating standard errors

yhat = as.vector(x %*% b)    # predicted values for y
res = y-yhat   # model residuals

sst = sum((y-mean(y))^2)     # total sum of squres
ssr = sum(res^2)             # residual sum of squares
ssm = sst-ssr                # sum of squares for model

df.e = (n-p-1)               # dgrees of freedom for error
df.t = (n-1)                 # total degrees of freedom
df.m = df.t-df.e             # degrees of freedom for model

s2 = as.vector(ssr/df.e)     # (sigma hat)^2
sigma2 = ssr/(n-p)

# Variance-covariance
varcov = s2 * xxi
serrors = sqrt(diag(varcov)) # coefficients standard errors

# R-squared
Rsq = 1-(ssr/sst)

# F-statisics
f = (ssm/df.m)/(ssr/df.e)
fpvalue = pf(f, df.m, df.e, lower.tail =FALSE)

# Calculate confidence intervals around regression estimates
## upper 95% CI and lower 95% CI

up95=b+qnorm(0.975)*ssr/sqrt(n)
lo95=b-qnorm(0.975)*ssr/sqrt(n)

# t-statistics for st.errors
tstats = b/serrors

# p-values
pval = 2*pt(abs(tstats), df.e, lower.tail = F)

# Coefficients matrix
b.table = cbind(Estimate = b, Std.Error = serrors, 'T-Statistic' = tstats, 'P-Value' = pval,
                'Lower 95% CI' = lo95, 'Upper 95% CI' = up95)
```

```r
return(list(coefficients=b.table, varcov=varcov, Rsq=Rsq,
            Fstat=paste("F-statistic:", f=round(f,3), "on", df.m,"and", df.e, "DF,",
                        "p-value:", fpvalue=round(fpvalue,3) ))
)
}
```

2. Running models

```r
# Set up the model formula
form = formula(gini_net_std ~ ELF_ethnic + polity2)

# Run the various models by using the function
model = ols(formula=form, data=data)
modelListDel = ols(formula = form, data=dataMiss)
modelAmelia = ols(formula = form, data=dataMiss, impute=TRUE)
```

```
## -- Imputation 1 --
##
##   1  2  3  4
```

	Estimate	Std.Error	T-Statistic	P-Value	Lower 95% CI	Upper 95% CI
"(Intercept)"	-0.220	0.854	-0.258	0.798	-8.808	8.368
ELF_ethnic	2.010	0.631	3.184	0.003	-6.577	10.598
polity2	-0.096	0.080	-1.204	0.235	-8.684	8.492

	Estimate	Std.Error	T-Statistic	P-Value	Lower 95% CI	Upper 95% CI
"(Intercept)"	0.669	0.569	1.175	0.247	-2.359	3.698
ELF_ethnic	0.805	0.444	1.816	0.078	-2.223	3.834
polity2	-0.165	0.052	-3.157	0.003	-3.193	2.864

	Estimate	Std.Error	T-Statistic	P-Value	Lower 95% CI	Upper 95% CI
"(Intercept)"	1.938	0.460	4.215	0.0001	-1.912	5.788
ELF_ethnic	0.334	0.452	0.739	0.464	-3.516	4.185
polity2	-0.287	0.038	-7.493	0	-4.137	3.563