Report for statisticians

Becky Yuen

Introduction

This is the overall report for the analysis on the European Value Study (EVS) from 2017 which is a survey research program on how Europeans think about family, work, religion, politics, and society. We are mainly interested in Europeans thoughts on two questions:

1. When a mother works for pay, do Europeans think the children suffer?
2. When jobs are scarce, do Europeans think employers should give priority to local people over immigrants?

```
library(haven)
EVS = read_sav("../data/EVS_data_cleaned.sav")
```

Descriptives of variables

In the following table, the variables are:

1. `v72` represents the first question of interest (1-strongly agree, 2-agree, 3-disagree, or 4-strongly disagree)
2. `v80` represents the second question of interest (1-strongly agree, 2-agree, 3-neither agree nor disagree, 4-disagree, or 5-strongly disagree)
3. `sex` (1-male or 2-female)
4. `age` (years)
5. `education` (1-lower, 2-medium, or 3-higher)

```
library(memisc)
library(pander)

pander(summary(EVS[,-which(names(EVS)%in%c("country","sex","education"))]),
       caption="Descriptive table for continuous variables")
```

Table 1: Descriptive table for continuous variables

v72	v80	age
Min. :1.000	Min. :1.000	Min. :18.00
1st Qu.:2.000	1st Qu.:1.000	1st Qu.:35.00
Median :3.000	Median :2.000	Median :50.00
Mean :2.713	Mean :2.313	Mean :49.57
3rd Qu.:3.000	3rd Qu.:3.000	3rd Qu.:64.00
Max. :4.000	Max. :5.000	Max. :82.00

```r
library(insight)

cat_table = table(EVS$education, EVS$sex, dnn = c("Education","Sex"))
rownames(cat_table) = c("Lower","Medium","Higher")
colnames(cat_table) = c("M","F")

export_table(format_table(cat_table), format = "md",
             caption = "Descriptive table for categorical variables")
```

Table 2: Descriptive table for categorical variables

Education	Sex	Freq
Lower	M	4727.00
Medium	M	11992.00
Higher	M	8351.00
Lower	F	6802.00
Medium	F	13835.00
Higher	F	11048.00

Graphs

```r
library(ggplot2)

ggplot(EVS, aes(as.factor(v72), age)) +
  geom_boxplot() +
  labs(x = "When a mother works for pay, the children suffer", y = "Age (Years)") +
  scale_x_discrete(labels = c("strongly agree", "agree", "disagree", "strongly disagree"))
```

Figure 1: Boxplot for first question of interest (v72)

We can see that the distributions of age among categories of opinion are quite similar.

```r
ggplot(EVS, aes(as.factor(v80), age)) +
  geom_boxplot() +
  labs(x = "When jobs are scarce, give priority to local people over immigrants",
       y = "Age (Years)") +
  scale_x_discrete(labels = c("strongly agree", "agree", "neither agree nor disagree",
                              "disagree", "strongly disagree"))
```

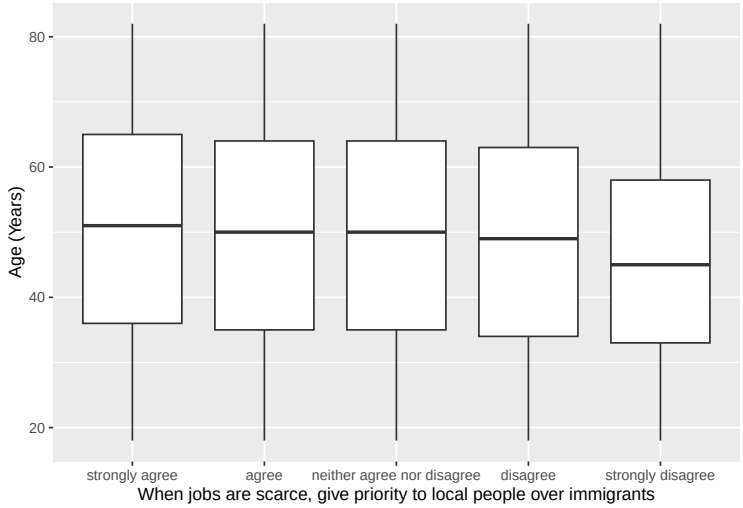

Figure 2: Boxplot for second question of interest (v80)

Same as the previous plot, we can see that the distributions of age among categories of opinion are quite similar.

Regression Analysis

Model: v72 ~ age + $\sqrt{\text{age}}$ + sex + education

```r
EVS$sex = factor(EVS$sex, levels=c(1,2), labels=c("-male","-female"))
EVS$education = factor(EVS$education, levels=c(1,2,3), labels=c("-lower","-medium","-higher"))
```

```r
model_v72 = lm(v72 ~ age + sqrt(age) + sex + education, data = EVS)
pander(summary(model_v72))
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.728	0.09723	28.06	4.655e-172
age	-0.004774	0.002203	-2.167	0.03023
sqrt(age)	-0.001149	0.02976	-0.03861	0.9692
sex-female	0.06448	0.007257	8.886	6.537e-19
education-medium	0.1233	0.009852	12.51	7.325e-36
education-higher	0.4012	0.01046	38.36	7.574e-318

Table 4: Fitting linear model: v72 ~ age + sqrt(age) + sex + education

Observations	Residual Std. Error	R^2	Adjusted R^2
56755	0.8576	0.04769	0.04761

The coefficient estimate for `sex` is 0.0644834 which means that the effect of a female respondent compared to a male is positive. The corresponding p-value is $6.5368574 \times 10^{-19}$ which is smaller than 0.05. Thus, `sex` is significant in the model.

Model: v80 ~ age + $\sqrt{\text{age}}$ + sex + education

```
model_v80 = lm(v80 ~ age + sqrt(age) + sex + education, data = EVS)
pander(summary(model_v80))
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.344	0.1427	16.43	1.646e-60
age	-0.003823	0.003232	-1.183	0.2369
sqrt(age)	0.006788	0.04367	0.1554	0.8765
sex-female	-0.03151	0.01065	-2.959	0.003084
education-medium	-0.03504	0.01446	-2.424	0.01536
education-higher	0.4238	0.01535	27.61	9.812e-167

Table 6: Fitting linear model: v80 ~ age + sqrt(age) + sex + education

Observations	Residual Std. Error	R^2	Adjusted R^2
56755	1.258	0.03124	0.03115

The coefficient estimate for `sex` is -0.0315131 which means that the effect of a female respondent compared to a male is negative. The corresponding p-value is 0.003084 which is smaller than 0.05. Thus, `sex` is significant in the model.