Week_6

Kalyani Cauwenberghs

2/14/2020

Part 1

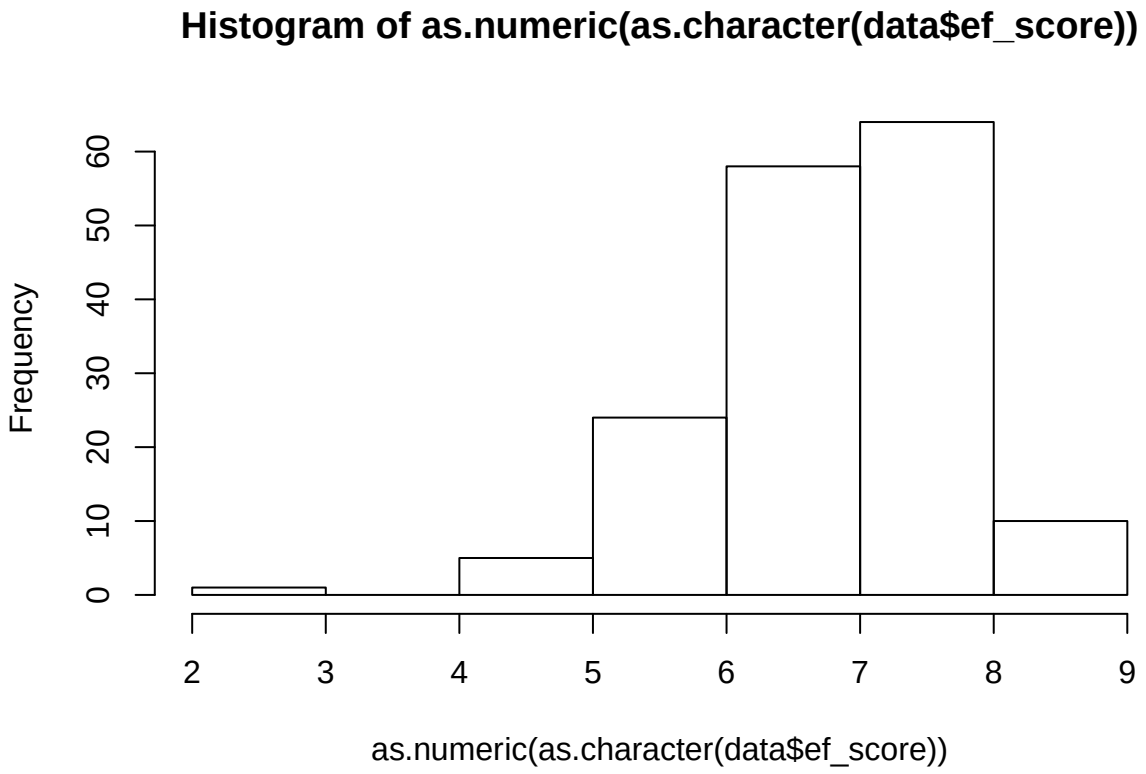
Part 2

Part 3

a.) Describe the distribution of the outcome variable, identify a main predictor that you're interested in studying its effect on the outcome

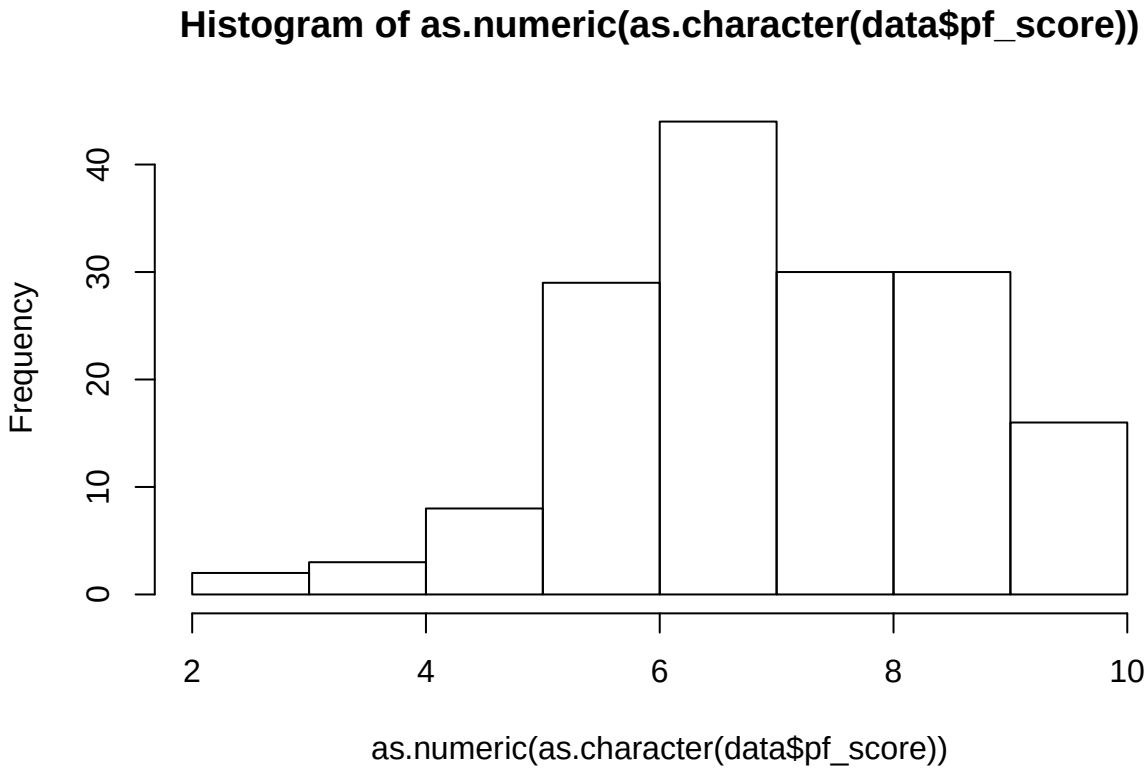
```
#read in dataset
data<-read.csv("hfi_cc_2019.csv")
data<-data[data$year=="2017",]
#predictor: political freedom
#outcome: economic freedom
#make sure both columns have no missing data
#sum(as.character(data$pf_score)=="-")==0
#sum(as.character(data$ef_score)=="-")==0

#distribution of outcome aka EF
hist(as.numeric(as.character(data$ef_score)))
```



Our outcome (EF score) is normal but left skewed, mean around 6.5, median around 7. Our main predictor is PF score.

```
hist(as.numeric(as.character(data$pf_score)))
```



b.) Identify other variables (i.e. predictors, often called covariates) that might be related to the outcome or the main predictor discuss these variables in the context of part 2 above of this assignment.

- The variables that determine political freedom can be our predictors for economic freedom because they are not directly used in the calculation of economic freedom (pf_rol, pf_ss, pf_movement, pf_expression, pf_identity).
- TODO: discuss these variables in the context of part 2 above of this assignment

c.) Carry out univariate logistic regression of the outcome on each of the predictors including the main predictor, interpret the results in terms of odds ratio etc.

```
#make the columns into numerics
pf_rol<-as.numeric(as.character(data$pf_rol))
pf_ss<-as.numeric(as.character(data$pf_ss))
pf_movement<-as.numeric(as.character(data$pf_movement))
pf_expression<-as.numeric(as.character(data$pf_expression))
pf_identity<-as.numeric(as.character(data$pf_identity))
ef_score<-as.numeric(as.character(data$ef_score))
#predictor: pf_rol, outcome: ef_score
rol<-glm(ef_score~pf_rol)
rol$coefficients
```

```
## (Intercept)      pf_rol
##   4.5847292   0.4335556
```

```
#predictor: pf_ss, outcome: ef_score
ss<-glm(ef_score~pf_ss)
ss$coefficients
```

```
## (Intercept)       pf_ss
##   3.7070908   0.3788679
```

```
#predictor: pf_movement, outcome: ef_score
movement<-glm(ef_score~pf_movement)
movement$coefficients
```

```
## (Intercept) pf_movement
##   5.4572220   0.1718419
```

```
#predictor: pf_expression, outcome: ef_score
expression<-glm(ef_score~pf_expression)
expression$coefficients
```

```
##   (Intercept) pf_expression
##     4.1799600     0.3379771
```

```
#predictor: pf_identity, outcome: ef_score
identity<-glm(ef_score~pf_identity)
identity$coefficients
```

```
## (Intercept) pf_identity
```

```
##    6.004151    0.122354
```

TODO: interpret the results in terms of conditional odds ratio etc.

d.) Fit a multiple logistic regression model by including more than one predictors, interpret the results in terms of conditional odds ratio etc.

```
reg<-glm(ef_score~pf_rol+pf_ss+pf_movement+
           pf_expression+pf_identity)
```

Coefficients:

```
reg$coefficients
```

```
##   (Intercept)        pf_rol         pf_ss   pf_movement pf_expression
##   3.791555627   0.334983860   0.082915167   0.048016569   0.032704874
##   pf_identity
##  -0.001300198
```

TODO: interpret the results in terms of conditional odds ratio etc.