

## Measures of Dispersion

Note: The dataset used in this tutorial and the R Script are on Moodle:

### Loading the 2016 CCES dataset

```
install.packages("foreign", dependencies=TRUE)
library(foreign)

dat <- read.dta(file.choose(), convert.factors=FALSE)
```

### Range of Variables

#### 1. Range

Displaying the range of a interval/ratio level variable is quite easy. You can use the same command as you have used in the past to see the mean and median.

**Example:** Here, we recode the ideology, age, and income variables and display the ranges.

```
table(dat$CC16_340a)
dat$ideology <- recode(dat$CC16_340a, "8=NA")
summary(dat$ideology)

table(dat$birthyr)
dat$age <- 2016 - dat$birthyr
table(dat$age)
summary(dat$age)

table(dat$faminc)
dat$income <- recode(dat$faminc, "31=NA; 97=NA")
table(dat$income)
summary(dat$income)
```

### Standard Deviation of Variables

#### 1. Calculating standard deviation

It is possible to calculate the standard deviation of variables quite easily too. However, you need to be aware of NA's (i.e. missing data). The standard deviation command can only calculate the standard deviations when you remove NA's from the calculation.

**Example:** Here we calculate standard deviations for the ideology, age, and income variables.

```
sd(na.omit(dat$ideology))
sd(na.omit(dat$age))
sd(na.omit(dat$income))
```

## Box and Whiskers Plot

### 1. Creating a box and whisker plot

**Example:** We can create a box and whisker plot in order to view one variable. Here, we do this with age.

```
boxplot(dat$age, main="Box Plot", ylab="Age of Respondent")
```

**Example:** Or, we could create a box and whisker plot in order to view one variable by a category of another variable. Here, we do this with ideology by gender (after recoding the gender variable.

```
table(dat$gender)
dat$gender1 <- recode(dat$gender, "1='Man'; 2='Woman'")
dat$gender1 <- as.factor(dat$gender1)
table(dat$gender1)
```

```
boxplot(dat$age ~ dat$gender1, main="Box Plot", ylab="Age of Respondent", xlab="Gender")
```

## Standard deviations by grouping

We can also acquire standard deviations by groupings. For example, here we calculate the standard deviations for the three variables by gender.

```
tapply(dat$ideology, dat$gender1, sd, na.rm=TRUE)
tapply(dat$age, dat$gender1, sd, na.rm=TRUE)
tapply(dat$income, dat$gender1, sd, na.rm=TRUE)
```

## Lab Activity

In the 2020 Finland European Social Survey dataset, you are to find the variable labels for the three variables provided below assessing the state of social trust in Finland. Then, explore the three variables and recode them so that they are in a usable format. Next, calculate the mean and standard deviation for each variable. Explain what the results conveys substantively for each variable. Finally, provide a box and whisker plot for each of these variables by gender.

1. Most people can be trusted or you can't be too careful?
2. Most people try to take advantage of you, or try to be fair?
3. Most of the time people try to be helpful or that they are mostly looking out for themselves?