R Basic Data Analysis Primer and Cheat Sheet

Coding Variables

Make sure your variables are of the correct type before doing analysis.

Categorical variables should be set as *factors* and continuous variables set as *numerics*

```r
#Set a variable as a factor:
dataFrame$variable<-as.factor(dataFrame$variable)

#Set a variable as a numeric:
dataFrame$variable<-as.numeric(dataFrame$variable)
```

Summary Statistics

Mean:

```r
mean(dataFrame$x,na.rm=T)
```

Median:

```r
median(dataFrame$x,na.rm=T)
```

Standard Deviation:

```r
sd(dataFrame$x,na.rm=T)
```

Summary:

```r
#This function outputs a number of useful summary values
summary(dataFrame$x)
```

Difference Between Two Means

Between-Subjects *t*-Test

To test the difference between the means of two different groups of people, do a between-subjects *t*-test:

```r
tTestName<-t.test(outcome~predictor,
                  data=dataFrame)
#Outcome = your outcome variable
#Predictor = a two-level predictor variable
#dataFrame = the dataframe containing your data

#To get the output of your t-test, execute:
tTestName
```

Within-Subjects *t*-Test

To test the difference between the means of two measurements of the same people at two different times, do a within-subjects *t*-test:

```
tTestName <- t.test(dataFrame$measure2,
                    dataFrame$measure1,
                    paired = T)

#measure2 = your second measure of each person
#measure1 = your first measure of each person
#"paired = T" = do a within-subjects t.test
#dataFrame = the dataframe containing your data

#To get the output of your t-test, execute:
tTestName
```

Difference Across Multiple Means

To test the difference between the means of 3 or more groups of people that differ along a single variable, do a one-way ANOVA:

```
anovaName<-lm(outcome~predictor,
              data=dataFrame)

#outcome = your outcome variable
#predictor = a categorical predictor variable with at least 3 levels
#dataFrame = the dataframe containing your data

#To get the output of your ANOVA, execute:
anova(anovaName)

#To view means as a function of group, do:

tapply(dataFrame$outcome,
       dataFrame$predictor,
       function(x)
         mean(x,na.rm=T))
```

To test the difference between group means as a function of multiple variables, do an *n*-way ANOVA:

```
anovaName<-lm(outcome~predictor1*predictor2,
              data=dataFrame)

#outcome = your outcome variable
#predictor1 = your first categorical predictor variable
#predictor2 = your second categorical predictor variable
#dataFrame = the dataframe containing your data

#To get the output of your ANOVA, execute:
anova(anovaName)
```

```
#To view means as a function of the predictors, do:

tapply(dataFrame$outcome,
       list(dataFrame$predictor1,
            dataFrame$predictor2),
       function(x)
         mean(x,na.rm=T))
```

To test the difference between multiple (>=3) means measured on the same people at different points in time, do a within-subjects ANOVA.

To do this, you first must "melt" your data from "wide format" to "long format."

Often we store data like this:

Participant	Observation 1	Observation 2
1	4	3
2	7	1

This is "wide" data. Melting data reorganizes it to "long" format like this:

Participant	Observation #	Score
1	1	4
1	2	3
2	1	7
2	2	1

One easy way to do this is the *melt()* function inside the **reshape2** package:

```
#Make sure you run library(reshape2) before you try this

dataMelted<-melt(dataFrame,
                 measure.vars=c("measure1","measure2"),
                 variable.name="observationNumber",
                 value.name="score")

#dataFrame = the dataframe you want melted
#"measure1" = your first measure of the outcome variable
#"measure2" = your second measure of the outcome variable
#variable.name = what you want your observation # variable to be called
#value.name = what you wnat your score variable to be called
```

You can then do a a within-subjects ANOVA on the melted data:

```
anovaName<-aov(outcome~predictor1*predictor2+Error(PIN),
               data=dataFrame)

#outcome = your outcome variable
#predictor1 = your first predictor variable
#predictor 2 = your second predictor variable (optional)
#PIN = a unique participant identifier
#dataFrame = the dataframe containing your data
```

Associations Between Continuous Variables

If you want to see if two continuous variables are associated with one another, do a correlation test:

```
corrName<-cor.test(dataFrame$x,
                   dataFrame$y)

#x = your first continuous variable
#y = your second continuous varaible
#dataFrame = the dataframe containing your data

#To get the output of your correlation, execute:
corrName
```

If you want to see if one continuous outcome variable is associated with multiple predictors (continuous or categorical), do a regression:

```
#For just main effects:
regName<-lm(outcome~predictor1+predictor2,
            data=dataFrame)

#For interactions:
regName<-lm(outcome~predictor1*predictor2,
            data=dataFrame)

#outcome = your outcome variable
#predictor1 = your first predictor variable, categorical or continuous
#predictor2 = your second predictor variable, categorical or continuous
#dataFrame = the dataframe containing your data

#To get the output of your regression, execute:
summary(r1)
```

Basic Data Cheat Sheet

Test	Code
Mean	mean(dataFrame$x, na.rm=T), colMeans(dataFrame[,n_1:n_2], na.rm=T)
Median	median(dataFrame$x, na.rm=T)
Standard deviation	sd(dataFrame$x, na.rm=T)
Summary	summary(dataFrame$x)
Between-subjects t-test	tTestName<-t.test(y~x, data=dataFrame)
Within-subjects t-test	tTestName<-t.test(dataFrame$measure1, dataFrame$measure2, paired=T)
One-way ANOVA	anovaName<-lm(y~x, data=dataFrame)
Two-way ANOVA	anovaName<-lm(y~x1*x2, data=dataFrame)
Melt data (reshape2)	dataMelt<-melt(dataFrame, measure.vars=c(“y1”,“y2”), variable.name=“x”, value.name=“y”)
Within-subjects ANOVA	anovaName<-aov(y~x+Error(PIN), data=dataMelt)
Correlation	cor.test(dataFrame$x, dataFrame$y)
Regression	regName<-lm(y~x1+x2, data=dataFrame)