Two-way ANOVA
=============

This R markdown document provides an example of performing a regression using the lm() function in R and compares the output with the jmv::ANOVA() function in the jmv (Jamovi) package.

Package management in R

```r
# keep a list of the packages used in this script
packages <- c("tidyverse","rio","jmv")
```

This next code block has eval=FALSE because you don’t want to run it when knitting the file. Installing packages when knitting an R notebook can be problematic.

```r
# check each of the packages in the list and install them if they're not installed already
for (i in packages){
  if(! i %in% installed.packages()){
    install.packages(i,dependencies = TRUE)
  }
  # show each package that is checked
  print(i)
}
```

```r
# load each package into memory so it can be used in the script
for (i in packages){
  library(i,character.only=TRUE)
  # show each package that is loaded
  print(i)
}
```

```
    ## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

    ## v ggplot2 3.3.3     v purrr   0.3.4
    ## v tibble  3.0.6     v dplyr   1.0.4
    ## v tidyr   1.1.2     v stringr 1.4.0
    ## v readr   1.4.0     v forcats 0.5.1

    ## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
    ## x dplyr::filter() masks stats::filter()
    ## x dplyr::lag()    masks stats::lag()

    ## [1] "tidyverse"
    ## [1] "rio"
    ## [1] "jmv"
```

ANOVA is a linear model

The ANOVA is a type of linear model. We're going to compare the output from the lm() function in R with ANOVA output.To use a categorical variable in a linear model it needs to be dummy coded. One group needs to be coded as 0 and the other group needs to be coded as 1.If you compare the values for F from lm() and t from the t-test you'll see that t^2 = F. You should also notice that the associated p values are equal.

Nice example:
https://sites.utexas.edu/sos/guided/inferential/numeric/glm/

Open data file

The rio package works for importing several different types of data files. We're going to use it in this class. There are other packages which can be used to open datasets in R. You can see several options by clicking on the Import Dataset menu under the Environment tab in RStudio. (For a csv file like we have this week we'd use either From Text(base) or From Text (readr). Try it out to see the menu dialog.)

```r
# Using the file.choose() command allows you to select a file to import from another folder.
dataset <- rio::import(file.choose())
# This command will allow us to import a file included in our project folder.
# dataset <- rio::import("Album Sales.sav")
```

Get R code from Jamovi output

You can get the R code for most of the analyses you do in Jamovi.

1. Click on the three vertical dots at the top right of the Jamovi window.
2. Click on the Syndax mode check box at the bottom of the Results section.
3. Close the Settings window by clicking on the Hide Settings arrow at the top right of the settings menu.
4. you should now see the R code for each of the analyses you just ran.

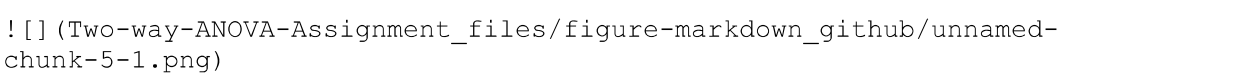
lm() function in R

Many linear models are calculated in R using the lm() function. We'll look at how to perform a regression using the lm() function since it's so common.

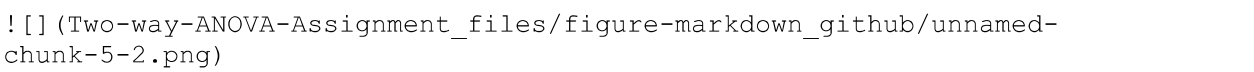
Visualization

```r
```

```r
# plots for outcome split by groups
ggplot(dataset, aes(x = Attractiveness))+
  geom_histogram(binwidth = 1, color = "black", fill = "white")+
  facet_grid(FaceType ~ .)
```



```r
ggplot(dataset, aes(x = Attractiveness))+
  geom_histogram(binwidth = 1, color = "black", fill = "white")+
  facet_grid(Alcohol ~ .)
```



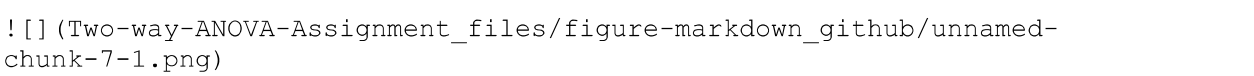
```r
# Make a factor for the box plot
dataset <- dataset %>% mutate(FaceType_f = as.factor(FaceType))
levels(dataset$FaceType_f)
```

```
## [1] "0" "1"
```

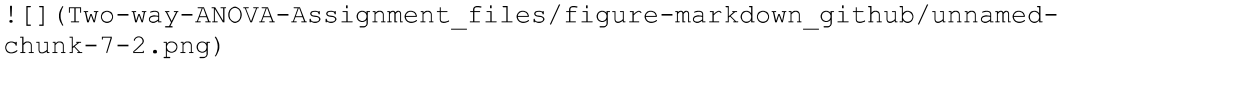
```r
dataset <- dataset %>% mutate(Alcohol_f = as.factor(Alcohol))
levels(dataset$Alcohol_f)
```

```
## [1] "0" "1" "2"
```

```r
ggplot(dataset, aes(x = FaceType_f, y = Attractiveness)) +
  geom_boxplot()
```



```r
ggplot(dataset, aes(x = Alcohol_f, y = Attractiveness)) +
  geom_boxplot()
```



Dummy codes

If a categorical variable is designated as a factor in R, the lm() function will dummy code it according to alphabetical order of the factor levels. The reference level will be the first category when the categories are put in alphabetical order. Since we already made factor variables from our categorical variables, we'll use those in the linear

model.

Computation

If we include independent variables in the model using the plus (+) sign, each variable in the equation will be included in the model. If we include independent variables in the model using the multiplication (*) sign, each variable will be included in the model, but interaction terms between the variables will also be included.

```r
model <- lm(formula = Attractiveness ~ FaceType_f*Alcohol_f, data = dataset)
model
```

```
##
## Call:
## lm(formula = Attractiveness ~ FaceType_f * Alcohol_f, data = dataset)
##
## Coefficients:
##             (Intercept)              FaceType_f1               Alcohol_f1
##                   3.500                    2.875                    1.375
##              Alcohol_f2  FaceType_f1:Alcohol_f1  FaceType_f1:Alcohol_f2
##                   3.125                   -1.250                   -3.375
```

Model assessment

```r
summary(model)
```

```
##
## Call:
## lm(formula = Attractiveness ~ FaceType_f * Alcohol_f, data = dataset)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.500 -0.625 -0.125  0.625  2.500
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)                3.5000     0.4137   8.461 1.29e-10 ***
## FaceType_f1                2.8750     0.5850   4.914 1.41e-05 ***
## Alcohol_f1                 1.3750     0.5850   2.350 0.023531 *
## Alcohol_f2                 3.1250     0.5850   5.342 3.49e-06 ***
## FaceType_f1:Alcohol_f1    -1.2500     0.8274  -1.511 0.138319
## FaceType_f1:Alcohol_f2    -3.3750     0.8274  -4.079 0.000197 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.17 on 42 degrees of freedom
## Multiple R-squared:  0.5154, Adjusted R-squared:  0.4578
## F-statistic: 8.936 on 5 and 42 DF,  p-value: 7.677e-06
```

function in Jamovi

Compare the output from the lm() function with the output from the
function in the jmv package.

```r
jmv::ANOVA(
    formula = Attractiveness ~ FaceType + Alcohol + FaceType:Alcohol,
    data = dataset,
    effectSize = "omega",
    homo = TRUE,
    norm = TRUE,
    qq = TRUE,
    postHoc = ~ FaceType + Alcohol,
    postHocES = "d",
    emMeans = ~ FaceType:Alcohol + Alcohol:FaceType)
```

```
    ## NOTE: Results may be misleading due to involvement in interactions
    ## NOTE: Results may be misleading due to involvement in interactions

    ##
    ##  ANOVA
    ##
    ##  ANOVA - Attractiveness
    ##
-------------------------------------------------------------------------------------------
    ##                          Sum of Squares    df    Mean Square    F
p            <U+03C9>²
    ##
-------------------------------------------------------------------------------------------
    ##    FaceType                  21.33333     1      21.333333    15.582609
0.0002952    0.1663195
    ##    Alcohol                   16.54167     2       8.270833     6.041304
0.0049434    0.1149955
    ##    FaceType:Alcohol          23.29167     2      11.645833     8.506522
0.0007913    0.1712288
    ##    Residuals                 57.50000    42       1.369048
    ##
-------------------------------------------------------------------------------------------
    ##
    ##
    ##  ASSUMPTION CHECKS
    ##
    ##  Homogeneity of Variances Test (Levene's)
    ##  ----------------------------------------
    ##    F            df1    df2    p
    ##  ----------------------------------------
    ##    0.7015989      5     42    0.6253432
    ##  ----------------------------------------
    ##
    ##
    ##  Normality Test (Shapiro-Wilk)
    ##  -----------------------------
    ##    Statistic    p
    ##  -----------------------------
    ##    0.9872998    0.8779276
```

```
    ##  ------------------------------
    ##
    ##
    ##  POST HOC TESTS
    ##
    ##  Post Hoc Comparisons - FaceType
    ##
-------------------------------------------------------------------------------------------
    ##    FaceType          FaceType    Mean Difference    SE           df
t            p-tukey      Cohen's d
    ##
-------------------------------------------------------------------------------------------
    ##    0           -    1                  -1.333333    0.3377681
42.00000    -3.947481    0.0002952    0.5697698
    ##
-------------------------------------------------------------------------------------------
    ##
    ##
    ##  Post Hoc Comparisons - Alcohol
    ##
-------------------------------------------------------------------------------------------
    ##    Alcohol          Alcohol    Mean Difference    SE           df
t            p-tukey      Cohen's d
    ##
-------------------------------------------------------------------------------------------
    ##    0           -    1               -0.7500000    0.4136798    42.00000
-1.812997    0.1777260    0.2616835
    ##                -    2               -1.4375000    0.4136798    42.00000
-3.474910    0.0033704    0.5015601
    ##    1           -    2               -0.6875000    0.4136798    42.00000
-1.661914    0.2317125    0.2398766
    ##
-------------------------------------------------------------------------------------------
```