Logistic Regression

Logistic regression allows us to predict membership in a dichotomous outcome using continuous and categorical predictor variables.

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"
```

RMANOVA is a linear model

Logistic regression is a type of generalization of the linear model. This assignment will allow us to compare the results of fitting a model using the glm() function with the output from the jmv package from Jamovi. Some explanation is available in Field chapter 20.3.

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("eel.sav")
```

glm() function in R

Most of the analyses we’ve looked at this semester have been forms of the linear model. Logistic regression is used with a dichotomous categorica outcom variable instead of a continuous variable. A more general version of a linear model called a generalized linear model can be used with outcome variables which are not continuous. Those can be analyzed in RStudio using a function called glm(). Some explanation can be found at
https://daviddalpiaz.github.io/appliedstats/logistic-regression.html

Change the categorical variables to factors

``` r
# Make factors
dataset <- dataset %>% mutate(Cured_f = as.factor(Cured))
levels(dataset$Cured_f)
```

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

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

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

``` r
modelLog <- glm(Cured_f ~ Intervention_f + Duration, data = dataset, family = binomial)
```

``` r
round(coef(modelLog), 1)
```

```
##     (Intercept) Intervention_f1        Duration 
##            -0.2             1.2             0.0
```

``` r
summary(modelLog)
```

```
## 
## Call:
## glm(formula = Cured_f ~ Intervention_f + Duration, family = binomial, 
##     data = dataset)
## 
## Deviance Residuals: 
##     Min       1Q   Median       3Q      Max  
## -1.6025  -1.0572   0.8107   0.8161   1.3095  
## 
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)   
## (Intercept)     -0.234660   1.220563  -0.192  0.84754   
## Intervention_f1  1.233532   0.414565   2.975  0.00293 **
## Duration        -0.007835   0.175913  -0.045  0.96447   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 154.08  on 112  degrees of freedom
## Residual deviance: 144.16  on 110  degrees of freedom
## AIC: 150.16
## 
## Number of Fisher Scoring iterations: 4
```

``` r
confint(modelLog)
```

```
## Waiting for profiling to be done...
```

```
##                      2.5 %   97.5 %
## (Intercept)     -2.6668137 2.163228
## Intervention_f1  0.4360018 2.068325
## Duration        -0.3546155 0.341240
```

``` r
deviance(modelLog)
```

```
## [1] 144.1558
```

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.

Logistic regression in jmv package

A note about some of the changes I had to make to this syntax to run it in RStudio.

change data = data to data = dataset (or whatever you name your dataframe) change ref = “Not Cured” to ref = “0” change ref = “No Treatment” to ref = “0”

``` r
output = jmv::logRegBin(
    data = dataset,
    dep = Cured,
    covs = Duration,
    factors = Intervention,
    blocks = list(
        list(
            "Intervention",
            "Duration")),
    refLevels = list(
        list(
            var="Cured",
            ref="0"),
        list(
            var="Intervention",
            ref="0")),
    modelTest = TRUE,
    bic = TRUE,
    pseudoR2 = c("r2mf", "r2cs", "r2n"),
    omni = TRUE,
    OR = TRUE,
    ciOR = TRUE,
    class = TRUE,
    acc = TRUE,
    spec = TRUE,
    sens = TRUE,
    auc = TRUE,
    rocPlot = TRUE,
    collin = TRUE)
```

output

```
    ##
    ##  BINOMIAL LOGISTIC REGRESSION
    ##
    ##  Model Fit Measures
    ##
------------------------------------------------------------------------------------------
    ##    Model    Deviance    AIC         BIC         R²-McF        R²-CS
R²-N         <U+03C7>²          df    p
    ##
------------------------------------------------------------------------------------------
    ##        1    144.1558    150.1558    158.3380    0.06443358
0.08411095    0.1130136    9.928185     2    0.0069843
    ##
------------------------------------------------------------------------------------------
    ##
    ##
    ##  MODEL SPECIFIC RESULTS
    ##
    ##  MODEL 1
    ##
    ##  Omnibus Likelihood Ratio Tests
    ##  --------------------------------------------------
    ##    Predictor       <U+03C7>²            df    p
    ##  --------------------------------------------------
    ##    Intervention    9.317005172     1    0.0022704
    ##    Duration        0.001983528     1    0.9644765
    ##  --------------------------------------------------
    ##
    ##
    ##  Model Coefficients - Cured
    ##
------------------------------------------------------------------------------------------
    ##    Predictor        Estimate        SE           Z              p
Odds ratio    Lower         Upper
    ##
------------------------------------------------------------------------------------------
    ##    Intercept        -0.234659530    1.2205629    -0.19225518
0.8475423     0.7908401    0.07230090    8.650349
    ##    Intervention:
    ##    1 - 0             1.233532057    0.4145650     2.97548547
0.0029253     3.4333349    1.52348373    7.737390
    ##    Duration         -0.007835250    0.1759132    -0.04454042
0.9644736     0.9921954    0.70284503    1.400667
    ##
------------------------------------------------------------------------------------------
    ##    Note. Estimates represent the log odds of "Cured = 1" vs. "Cured =
0"
    ##
    ##
    ##  ASSUMPTION CHECKS
    ##
    ##  Collinearity Statistics
```

```
## ----------------------------------------
##                   VIF        Tolerance
## ----------------------------------------
##   Intervention    1.075431    0.9298601
##   Duration        1.075431    0.9298601
## ----------------------------------------
##
##
## PREDICTION
##
## Classification Table - Cured
## ------------------------------------
##   Observed    0     1     % Correct
## ------------------------------------
##          0    32    16     66.66667
##          1    24    41     63.07692
## ------------------------------------
##   Note. The cut-off value is set
##   to 0.5
##
##
## Predictive Measures
## --------------------------------------------------------
##   Accuracy     Specificity    Sensitivity    AUC
## --------------------------------------------------------
##   0.6460177      0.6666667      0.6307692    0.6575321
## --------------------------------------------------------
##   Note. The cut-off value is set to 0.5
```

```
![](Week-14-Assignment-Logistic-Regression-RNotebook_files/figure-
markdown_github/unnamed-chunk-8-1.png)
```