MATH 4773 Laboratory 2

2-Way Categorical Data Analysis Using R

In this lab you will learn about contingency tables and how to perform tests for independence. Generally you will want to decide whether the data supports differing underlying row profiles using NULL hypothesis testing. Once you have decided there is sufficient evidence to support the contention of different row (or column) profiles you will look for interval estimates for different combinations of categories.

# Objectives

In this lab you will learn how to:

* Create tables from data
* Create appropriate investigative plots
* Create and interpret hypothesis tests
  + P values
  + Confidence Intervals
  + Test statistic
  + Distributional assumptions
* Create statistics by hand

### Tasks

We will use Rstudio for all of the course.

All output should be made via Rmarkdown. Knit into html and upload both rmd and html documents

* Task 1
  + Download from CANVAS the zipped data files, “Dataxls”
  + Unzip the contents into a directory on your desktop (call it LAB)
  + Download the file “lab2.r”
  + Place this file with the others in LAB2.
  + Start Rstudio
  + Open “lab2.r” from within Rstudio (this is an exemplar file).
  + Using hash commenting and at the top of Lab 2 place the task number eg #Task 1
  + Go to the “session” menu within Rstudio and “set working directory” to where the source files are located.
  + Copy and paste the working directory by issuing the command getwd(): under #Task 1
* Task 2
  + Make sure you install the readxl package.
  + Locate the file MILE3.xls
  + Use read\_excel() to read the data into R, this function will already be available within the script lab1.r which you have opened in Rstudio.
  + Copy and paste the first six lines of the data using “head()” (use “courier new” font):
* Task 3
  + Using the library “vcdExtra” and the function xtabs() create a table of Evac Vs Distance. Place here!
  + Make a mosaic plot of the table above using the function mosaic() from the vcd package. Place here!
  + Make a function called “mychsq” which will take in any 2 way table of counts and produce the following output:
    - A mosaic plot
    - A list of :
      * expected frequencies
      * individual chisquare values
      * summed chi square value
      * P value
      * Degrees of freedom
  + Use mychisq() to test the MILE3 data for independence – what do you conclude?
  + Now use rowdistr() from the 20x package and compare within differences – record plots and command line output here! What do you conclude?
  + Now use rowdistr() from the 20x package and compare between differences – record plots and command line output here! What do you conclude?