MATH 3215 Individual Project - Introduction to R

Introduction guide and formal documentation: An Introduction to R

To install and get started with R:

1. Download and follow the installer guide for your operating system
    a. https://cran.r-project.org/
2. Download and install RStudio, which will interpret and execute R commands that we write.
    a. https://posit.co/products/open-source/rstudio/
3. Download and install Rtools, which will help us use additional R packages
    a. https://cran.rstudio.com/bin/windows/Rtools/

Basics of R - An Introduction to R

- R is a dynamic programming language, which means R automatically interprets your code as you run it
    - No need for having to compile code
    - Run code each line at a time (similar to command line)
- Basic math operations: +, -, *, /
- Comments: Begin line with #
- Vectors: array, list, etc
    - Creating with a range: start:end, inclusive - similar to range() in python
    - Creating with individual objects: c(item1, item2, item3)
- Variables: name <- value
- round(), factorial(), mean(vector)
- sample(x = vector, size = k, replace = TRUE) → P(n, k)
- args(functionname) - gives detail about arguments of function
- help(functonname) - gives detailed information about any method/function
- my_function <- function(parameters) {content}
- for (i in vector) {content}
- Best way to figure out any method you're confused on is to use help(method), args(method) or google the method name

Sets - Set Documentation

- Initiate: set(item1, item2, item3) OR as.set(vector)
- Useful functions: is.set(x), set_is_empty(x)
- If a and b are sets:
    - Union: a | b
    - Difference: a - b
    - Intersection: a & b
    - Cartesian product: a * b
    - Power set: 2^a
    - Subsets/supersets: a < b, a <= b, a > b, a >= b
        - = means proper subset
    - Equality: a == b, a != b
    - Cardinality: length(a)
    - Set of all subsets with specific length: set_combn(a, n)
- Can iterate through sets with for loops to look at all elements of a set

Probability Distributions - Probability Distribution Documentation

- Helpful plotting guide
- R is built for probability and statistics, so it has a lot of built in functions for probability distributions (and statistics that go along with them)
- Format: [data type][distribution type](range vector, other parameters)
    - Eg: dbinom(x, size =, prob =)
- Prefixes: d (density/PMF), p (CMF), r (simulation / random draws)
- Distribution types:
    - Binomial: binom(x, size, prob)
    - Geometric: geom(x, prob)
    - Normal: norm(x, mean, sd)
    - Poisson: pois(x, lambda)
    - Exponential: exp(x, rate)
- Plotting:
    - Generate distribution, then use R functions to plot it
    - plot(x, y, main="title") takes vectors for x axis points and for y axis points
    - PMF graph: plot(0:n, binom(0:n, size=n, prob=p), main=sprintf("PMF Binomial Distribution, n = %s p = %s", n, p))

- Random data in a histogram to show distribution: hist(rbinom(0:10, size=10, prob=p), breaks = (0:(n+1)) - 0.5, main = “title here”)

Combinations & Permutations - combinat package documentation

- install.packages(“combinat”) and library(combinat) required before using these functions!!
    - Install R packages 1 time, declare them with library(packagename) in every file that uses that package
- combn(x, m) - combinations of all elements of vector x taken m at a time
    - Can also take combn(n, m) with n as an integer
    - n choose m, C(n, m)
    - dim(combn(x, m))[2] gives #possible combinations
- permn(x) - permutations of all elements of vector x
    - Can also take permn(n) with n as an integer
    - n choose n, P(n, n)
    - length(permn(n)) gives #possible permutations
- Can also find the # through factorial() method
- Helpful stackoverflow question on this

Extra: Statistical Models in R

- Linear regression, least squares, etc
- If interested, functions here: Statistical Models Documentation

Practice Problems:

1) Create sets A, B, and C with $B \cap C = \phi$ and $A \cup B = B$.
    a) Create corresponding variables in R of A, B, and C. Check that the assumptions are correct using R.
    b) Find the power set of B and check your work using R.
        i) Print out each element of the power set of B using a for loop
    c) Test the subset/superset functions in R with A and B. What functions return true?
2) Plot the PMF and CMF of the binomial distribution representing 7 coin flips of a fair coin and coining the number of heads.

3) Plot the PMF and CMF of the binomial distribution representing 10 rolls of a fair 6 sided and counting the number of 5s.
4) Generate 100 random points of adult men in the US's heights, assuming those heights follow a normal distribution with a mean of 70 in and a standard deviation of 3 in.
    a) Visualize this simulated data in a histogram
    b) Plot the normal distribution curve on the same plot. Does your data follow the same pattern as the curve?