Phase 3

**Chi-Square Test**

* Load in Packages
  + Packages used are dpylr, tidyr, ggplot2, EnvStats, and stringr (used for string editing)
  + Load in Master Model data located in Phase 3
* Set parameters of the simulation.
* Num\_of\_q how many times you want the simulation to repeat.
* Creation of data frames.
  + Quantiles is a data frame with all a column for every n, theta, percentile combination we looked at with a length of the number of simulations we want to run.
  + When a column name ends in .1 it represents the bin of the 0 percentile to the 10th percentiles. .2 is the 10th to the 20th and so on.
    - This will be used to store the number of correlation coefficients that fall into each bin
  + Df\_nt is a data frame with a column for every combination of n and theta we studied with a length of the number of correlation coefficients we want to simulate.
    - This will be used to store the correlation coefficients as is does in when simulating the MLE\_Alpha’s
* Create Bin Boundaries
  + Found the values of the percentiles (.1, .2, .3, .4, .5, .6, .7, .8, .9) of each beta distribution.
  + This will make it easier to sort the data later.
* For Loop
  + Works exactly like the For loop from phase one.
    - Instead of finding MLE’s and storing them now. We find the length(how many) of the correlation coefficients fall into each bin by satisfying the inequalities in the code.
* Change to long data
  + I used the gather command to change to quantiles data frame to have the old column names in a column (string) next to its values(counts).
  + The following lines of code change the string variable to split to string up in 3 different columns n, theta, bin (labeled by percentiles).
* Chi-Square Values
  + The first batch of lines run takes my data frame and only takes the given lines, sorts the rows by n, theta and percentile (bin), creates a new column of the chi-square value for each bin, and stores all of this in the quantiles.long data frame.
* The next batch of lines turns the previous data frame into a new data frame where the bins are turned into columns names representing their chi-square values so you can sum the chi-squares across the row to find the total chi-square value for the distribution.
* The next batch of lines uses the old data frame and adds a new column of total chi-squares, and then creates a new data frame with the columns n, theta, and mean chi-squares that were grouped by n and theta.
  + Change the theta column name to Theta so it matches the df.Rho column names
  + Joins the data with df.Rho, to adds chi-square column to df.Rho
  + The rows with negative values of fita are removed.
  + Saved as quantiles.wide\_1, and saved as chisq gof\_1.Rdata (Number after \_ represents the rho value used).

**Relative Error**

* Same set up as and data as Chi-square
* Create data frames
  + Create a data frame with columns n, theta, and a column for each of the quantiles we examine with 56 rows (each combination of n and theta) (df\_Truth.)
  + Create a data frame for each combination of n and theta with the columns n, theta, and r (n and theta will be the same for each row).
* For Loop
  + Outer loop
    - Will be completed until df\_Turth data frame is full.
    - The if statement checks what the n and theta value in the df\_truth data frame to see what parameters to use in simulating |r|
      * Inner Loop, Works the same as inner loop is phase 1 only it only does the simulation for this one n theta combo instead of combination.
        + Instead of a column in df\_nt it is stored in separate df.
    - Back to outer loop 🡪 finds the given quantiles and stores them in df\_Truth.
* Create comparison data frame.
  + Creates a data frame with columns n, theta, quantile, average value, and Beta, where the length of the data frame is the number of different combinations of n, theta, quantile combinations observed. The average value average quantiles value grouped by n and theta. The Beta column quantile values of the beta distribution our models generate. Saved as compare.
* Compute Relative Error
  + Takes the compare data frame and adds a new column with the relative errors.
  + Change the theta column name to Theta so it matches the df.Rho column names
  + Joins the compare data with df.Rho, to adds df.Rho columns to compare
  + The rows with negative values of fita are removed.
  + Saved as compare\_1, and saved as data (Number after \_ represents the rho value used) in phase 3

**Combine Data**

* Same set up as chi-square
* Combine Chi-Square
  + Load in all of the chi square data form different rho values.
  + Lines in-between loading in data are making sure the data frame has removed negative values of fita, and the data has the same column names as every other data frame
  + Rbinds combines the data frame
  + Df.chi is the data frame with the columns n, theta, rho, and chi-square
  + Df.chi is saved as the Master ChiSq data in phase 3
* Combine Relative Error
  + Same as combining chis square
  + Saves df.rel as Master Relative Error in Phase 3.
* Graphs
  + Builds heat maps for chi-square data