ST 502 Final Project

For this project you will **work by yourself**. The project involves involves analyzing a data set using McNemar's test (a test for paired categorical data). Then you will show derivations of parts of the test. Finally, you'll conduct a Monte Carlo simulation to determine properties of the test in a situation similar to that of the original dataset. You'll create a final report to turn in.

Please make sure that your R file follows the guidelines on moodle. If these guidelines are not met, you will lose credit. You can submit a .R or a .Rmd file for the code portion. You should submit an HTML or PDF file for the report portion. (So you'll turn in two documents in total!)

To be clear, **you are not allowed to share code or derivations with others in the class**. You can discuss ideas and things you are stuck on, but you can't show someone your code or derivation. **You are not allowed to use ChatGPT or other such software to help you code on this project.** The coding portion is worth 40% of the project. **If I see any use of generative AI on the project, you will receive a 0 for this portion of the project**. Some things that are clear signs for me:

- Use of `cat()`
- Use of naming row elements in a matrix or data frame (`my_matrix['n = 20, p = 0.1']` for instance)
- Code structure that follows very closely with what chatGPT and other similar products produce. If you don't use these, you won't naturally have a program that does this so don't worry about it.
- Code that closely mirrors another student's submission. Again, if you code this yourself, this isn't anything to worry about.
- Other specific minor mistakes as compared to our instructions.

Analyzing a Dataset

Consider a dataset on 250 subjects in which acid reflux is treated by a drug. The response is either success (reflux stopped) or failure (reflux still present). Half of the subjects are randomly selected to use drug A the first day they have reflux and drug B the second day they have reflux. The others were assigned to use drug B first, then drug A.

We want to know if the drugs have a different probability of successful relief. The outcomes for the relief given by the drugs is given in the table below.

		Relief status		
		Success	Failure	Total
Drug	Drug A	100	150	250
	Drug B	125	125	250
	Total	225	275	500

We wouldn't want to analyze this data with a test of homogeneity because the observations in the table are not independent! We really only have 250 subjects so we can't have 500 independent observations. Instead, we can look at a table of concordant and discordant pairs.

		Drug B Relief status		
		Success	Failure	Total
Drug A Relief status	Success	85	15	100
	Failure	40	110	150
	Total	125	125	250

The diagonal now represents the observations that 'agreed' for an individual (concordant pairs). The off diagonals represent observations that didn't agree for individuals (discordant pairs).

To test whether or not the drugs have a different effect, we can now apply McNemar's test on the table of concordant/discordant pairs. Let's formally write this out (you'll derive parts of this in the next section).

We can consider the table of probabilities given **by the table of concordant/discordant pairs**:

		Drug B Relief status		
		Success	Failure	Total
Drug A Relief status	Success	π_{11}	π_{12}	$\pi_{1\bullet}$
	Failure	π_{21}	π_{22}	$\pi_{2\bullet}$
	Total	$\pi_{\bullet 1}$	$\pi_{\bullet 2}$	$\pi_{\bullet\bullet}$

We really want to test a restricted multinomial vs a free multinomial with this table!

$$H_0 : \text{No relationship between drug and relief, or, } \pi_{1\bullet} = \pi_{\bullet 1} \text{ and } \pi_{2\bullet} = \pi_{\bullet 2} \text{ equivalent to } \pi_{12} = \pi_{21}$$

$$H_A : \text{cell probabilities are 'free' (other than the sum to 1 constraint)}$$

If we use Pearson's Chi-Square test statistic (call it X^2), we have

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(Obs_{ij} - Exp_{ij})^2}{Exp_{ij}} = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

where n_{ij} represents the observed count in cell ij. (You'll show this later on.) Our reference distribution is a χ_1^2.

This test can be done in R using the `mcnemar.test()` function from the `stats` package (automatically loaded into R). We simply pass the function the matrix of concordant/discordant pairs.

Tasks - Part 1

First, find the value of the test statistic, the rejection region, and the p-value of the test using R (coding these parts yourself).

Next, create this data in R (`matrix()` function will work) and run McNemar's test to determine if we have evidence that the drugs have different probabilities of relief using the `mcnemar.test()` function. Use `correct = FALSE` in `mcnemar.test()` so that you get the same answer here!

Deriving Parts of the Test

I'll break the things you need to derive into pieces below.

Tasks - Part 2

- First, in the statement of the null hypothesis we wrote:

$$\pi_{1\bullet} = \pi_{\bullet 1} \text{ and } \pi_{2\bullet} = \pi_{\bullet 2} \text{ equivalent to } \pi_{12} = \pi_{21}$$

  Show this is true.

- Second, under this null restriction on our multinomial, derive the maximum's for $\pi_{11}, \pi_{12}, \pi_{21}$, and π_{22}. This can be done using Lagrange multipliers or by substituting in carefully to include the restriction. No need to show your resulting critical values are maximums.

- Third, although we derived the form of the LRT generally for a restricted vs free multinomial, I'd like you to do this for this specific problem. You should show that

$$-2ln\left(\frac{L(\tilde{\pi_{11}}, \tilde{\pi_{12}}, \tilde{\pi_{21}}, \tilde{\pi_{22}})}{L(\hat{\pi_{11}}, \hat{\pi_{12}}, \hat{\pi_{21}}, \hat{\pi_{22}})}\right) = 2\sum_{i=1}^{2}\sum_{j=1}^{2} Obs_{ij} ln\left(\frac{Obs_{ij}}{Exp_{ij}}\right)$$

and then argue that the appropriate reference distribution has 1 degree of freedom. You don't need to derive the overal maximums for our π_{ij} values. You can just use the usual MLE result as we did in class.

- Lastly, we know we can use Pearson's chi-square test statistic instead of this LRT and they are asymptotically equivalent. In our above data example, we used

$$X^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(Obs_{ij} - Exp_{ij})^2}{Exp_{ij}} = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

Show that Pearson's chi-square test statistic can be simplified into the form on the right.

Simulation Study

We might like to know how well our test does at detecting certain alternatives (i.e. its power). While we can try to derive these things, using simulation based methods provides a nice method for finding approximate results.

The Pearson statistic can be derived as a Taylor series approximation to the LRT. For this last part of the project, we'll investigate the α control of the Pearson chi-square test and its power (so we don't have to worry about the `ln(0)` that can sometimes pop up for the LRT).

Goal of simulation study:

- Determine how well the asymptotic rejection region performs at controlling α
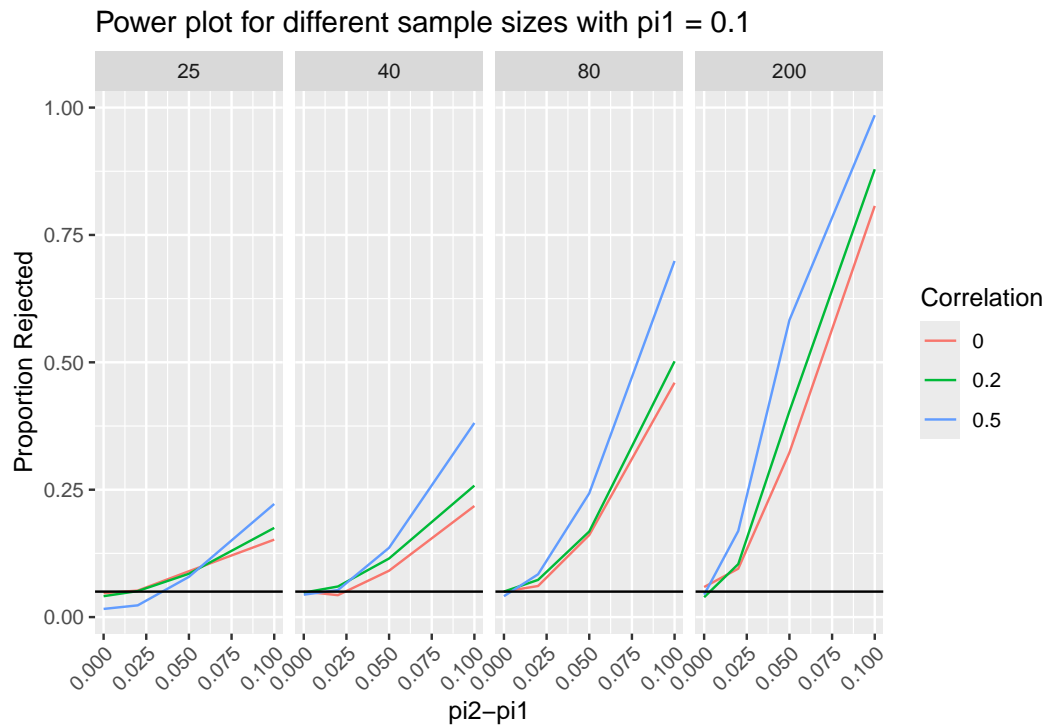- Determine the power of the asymptotic test when comparing certain alternatives

Setup:

- We'll generate correlated binary data using the `draw.correlated.binary()` function from the `MultiRNG` package. You'll likely need to install this package.

    - The first argument, `no.row`, is the sample size
    - The second argument, `d`, is the number of variables to create (we want 2)
    - The third argument, `prop.vec`, is the vector of true proportions of success (of length two, corresponding to our two variables)
    - The fourth argument, `corr.mat`, is a (2x2 for our purposes) correlation matrix. This should have 1's along the diagonal and the value of the correlation in the off-diagonals.

- We'll generate data under all combinations of the following:

    - $n = 25, 40, 80$, and 200
    - $\pi_1 = 0.1, 0.4, 0.8$ ('drug A' variable's success probability)
    - $\pi_2 = \pi_1$, $\pi_2 = \pi_1 + 0.02$, $\pi_2 = \pi_1 + 0.05$, $\pi_2 = \pi_1 + 0.1$ ('drug B' variable's success probability)
    - $\rho = 0, 0.2, 0.5$ (this is the correlation with which we generate our data)

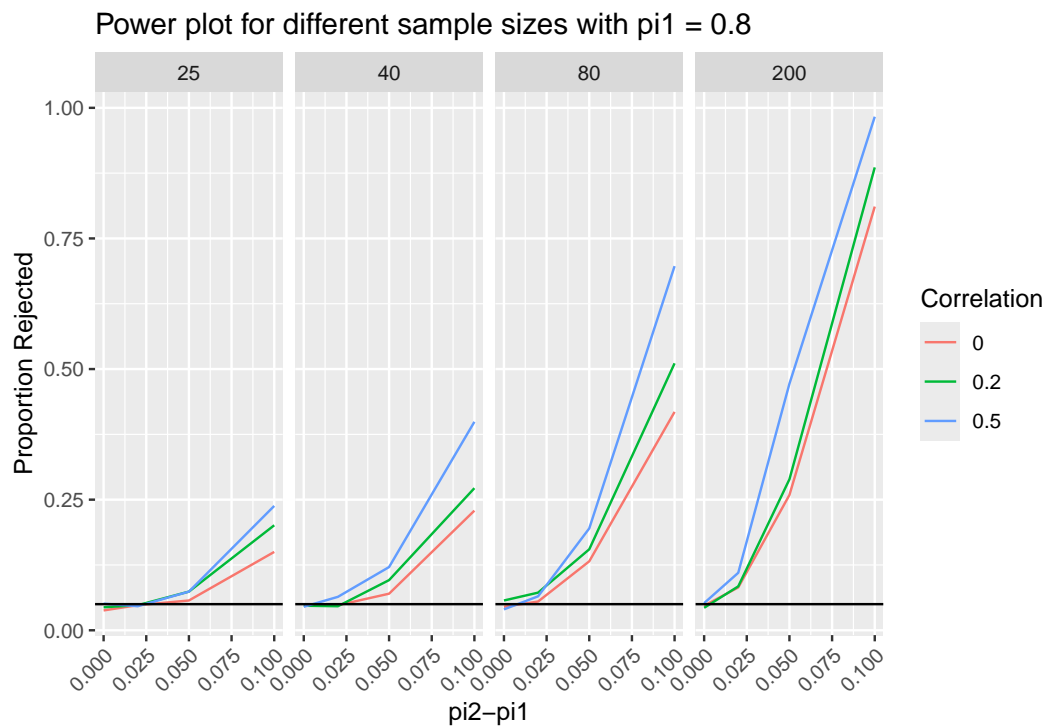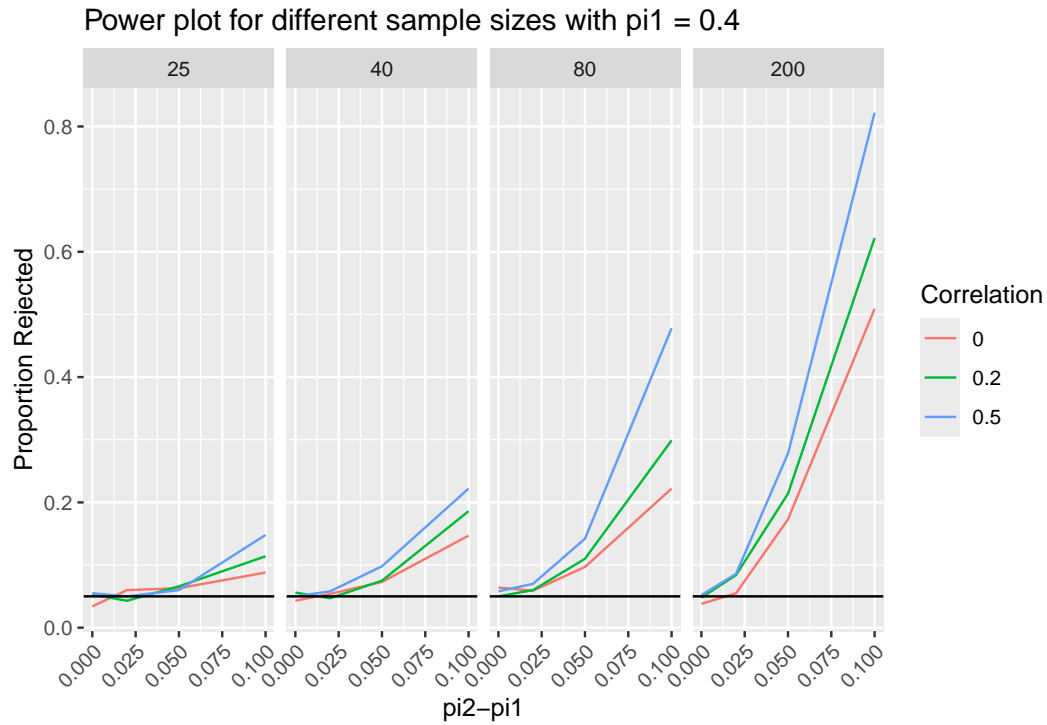- We'll generate $N = 1000$ datasets under each of these settings.

We'll be able to determine α control by looking at the case when $\pi_1 = \pi_2$. All of the other cases will allow us to investigate power under the alternative created by the difference in π's and correlation.

You should create plots similar to those that appear below.

Some coding hints:

- I wouldn't try to generate all of the data and keep it somewhere. (You can if you want to.). I'd do the following:
    - Write a function to conduct McNemar's test given a sample size, value of correlation, π_1, and π_2 values. The output of the function would just be whether you reject or fail to reject (a `TRUE/FALSE`)
    - You can then use the `replicate()` function to repeatedly run tests under a given combination of the above. You can find the mean of the results from `replicate()` and that would represent the probability of rejecting for a given situation.
    - I don't really care about the efficiency of your code!

Power plot for different sample sizes with pi1 = 0.1

4

Power plot for different sample sizes with pi1 = 0.4

Power plot for different sample sizes with pi1 = 0.8

Report

You should then write up all of the above into a coherent report with the following pieces:

- Introduce the general idea of McNemar's test and where it is used in an introduction (**10 points**)

- Analyze the dataset given and briefly discuss the results **(10 points)**
- Derive the test pieces as described (you should use math type, latex, markdown, etc. to typeset your math symbols) **(40 points)**
- Describe the simulation you will do. Your report should include the R code in the text or in an appendix. The plots should be within the text with a brief discussion of the results. **(40 points)**

That's it! You've then finished ST 502 - woot