

Exploration of McNemar's Test

Silvia Ying
ST502

Introduction

McNemar's test is a statistical technique for comparing paired nominal data. It is particularly valuable when assessing changes or differences between two related conditions or treatments. The test is designed to examine matched pairs of subjects, including paired samples or before-and-after measurements. McNemar's test (also known as "marginal homogeneity testing") examines a 2×2 contingency table to assess if the row and column marginal frequencies are equal. Comparing how well two diagnostic tests work on the same patient population is a typical practice in medical diagnostics. The test statistic has a chi-squared distribution, making it possible to identify variations from marginal homogeneity. In our study, we will explore a dataset about two treatments for acid reflux with Monte Carlo simulation.

Dataset Description

1. Sample Size: The dataset contains information on 250 subjects who experienced acid reflux.
2. Treatment Assignment: Subjects are randomly assigned to one of two treatment sequences: half of the subjects use Drug A on the first day they have reflux and Drug B on the second day. The other half use Drug B first, then Drug A.
3. Response: shown with Table of Concordant and Discordant Pairs. This table is used to assess the level of agreement between Drug A and Drug B in providing relief outcomes. It helps in understanding how often the drugs provide consistent relief outcomes and where they may differ.
 - a. Concordant Pairs (green cells): These are pairs of observations where both Drug A and Drug B agree on the relief outcome. In this case, both drugs either provide relief success or relief failure for the patient.
 - b. Discordant Pairs (red cells): These are pairs of observations where Drug A and Drug B disagree on the relief outcome. One drug provides relief success while the other provides relief failure.
 - c. Table Structure: The table has two rows and two columns. The rows represent the relief outcomes for Drug A, while the columns represent the relief outcomes for Drug B. Each cell in the table represents the count of patients falling into that particular combination of relief outcomes.

Table of Concordant and Discordant Pairs

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

Task 1 - McNemar's Test in R

Code explanation: I first collect observed frequencies for each cell of a 2x2 contingency table, representing the outcomes of two variables Drug A and Drug B. Then, I input observed into a matrix named `data_matrix`. Using R's `mcnemar.test()` function, I compute McNemar's test statistic and p-value based on the provided data matrix, setting the correct argument to `FALSE` to avoid applying a continuity correction. Afterward, I interpret the results by printing the computed McNemar's test statistic and p-value to the console using the `cat()` function. These values allow me to determine whether there is a significant difference in the marginal proportions of the two categorical variables.

Test result:

- McNemar's Test Statistic: 11.36364
- P-value: 0.0007489604

The test statistic of 11.36364 indicates the degree of disagreement between the paired observations. A higher value of the test statistic suggests more disagreement between the treatments. The p-value of 0.0007489604 is the probability of obtaining a test statistic as extreme as, or more extreme than, the observed value under the null hypothesis that there is no difference between the two treatments.

Since the p-value is much smaller than 0.05, we reject the null hypothesis and conclude that there is a statistically significant difference between the effectiveness of Drug A and Drug B in treating acid reflux. We have evidence to suggest that there is a significant difference in the effectiveness of Drug A and Drug B in treating acid reflux.

Task Part 2.

1. Show $\pi_{1.} = \pi_{.1}$ and $\pi_{2.} = \pi_{.2}$ equivalent to $\pi_{12} = \pi_{21}$

$\because \pi_{1.} = \pi_{11} + \pi_{12}$, and $\pi_{.1} = \pi_{11} + \pi_{21}$

\therefore when $\pi_{1.} = \pi_{.1}$, $\pi_{11} + \pi_{12} = \pi_{11} + \pi_{21}$

subtract π_{11} from both sides

$$\pi_{12} = \pi_{21}$$

$\because \pi_{2.} = \pi_{12} + \pi_{22}$, and $\pi_{.2} = \pi_{21} + \pi_{22}$

when $\pi_{2.} = \pi_{.2}$, $\pi_{12} + \pi_{22} = \pi_{21} + \pi_{22}$

Subtract π_{22} from both sides

$$\pi_{12} = \pi_{21}$$

therefore, $\pi_{1.} = \pi_{.1}$ and $\pi_{2.} = \pi_{.2}$ equivalent to $\pi_{12} = \pi_{21}$

2. derive the maximum's for $\pi_{11}, \pi_{12}, \pi_{21}$ and π_{22}

Deriving Part of the Test:

Part 1:

General likelihood given by $L(\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}) \propto \pi_{11}^{n_{11}} \pi_{12}^{n_{12}} \pi_{21}^{n_{21}} \pi_{22}^{n_{22}}$

Under Ho: $\pi_{12} = \pi_{21}$, therefore $\pi_{22} = 1 - \pi_{11} - 2\pi_{12}$ and we can replace π_{21} with π_{12}

$$l(\pi_{11}, \pi_{12}) = C + n_{11} \ln(\pi_{11}) + n_{12} \ln(\pi_{12}) + n_{21} \ln(\pi_{12}) + n_{22} \ln(1 - \pi_{11} - 2\pi_{12})$$

$$\frac{d}{d\pi_{11}} = \frac{n_{11}}{\pi_{11}} - \frac{n_{22}}{1 - \pi_{11} - 2\pi_{12}}, \text{ set } = 0$$

$$\text{We get } \frac{n_{11}}{\pi_{11}} - \frac{n_{22}}{1 - \pi_{11} - 2\pi_{12}} = 0$$

$$n_{11}(1 - \pi_{11} - 2\pi_{12}) - n_{22}\pi_{11} = 0$$

$$n_{11} - n_{11}\pi_{11} - 2n_{11}\pi_{12} - n_{22}\pi_{11} = 0$$

$$\pi_{11}(n_{22} + n_{11}) = n_{11} - 2n_{11}\pi_{12}$$

$$\pi_{11} = \frac{n_{11} - 2n_{11}\pi_{12}}{n_{22} + n_{11}}$$

$$\frac{d}{d\pi_{12}} = \frac{n_{12}}{\pi_{12}} + \frac{n_{21}}{\pi_{12}} - \frac{2n_{22}}{1 - \pi_{11} - 2\pi_{12}}, \text{ set } = 0$$

$$\text{We get } n_{12}(1 - \pi_{11} - 2\pi_{12}) + n_{21}(1 - \pi_{11} - 2\pi_{12}) - 2n_{22}\pi_{12} = 0$$

$$-2\pi_{12}(n_{12} + n_{21} + n_{22}) = -n_{12} + n_{12}\pi_{11} - n_{21} + n_{21}\pi_{11}$$

$$\pi_{12} = \frac{-n_{12} + n_{12}\pi_{11} - n_{21} + n_{21}\pi_{11}}{2(n_{12} + n_{21} + n_{22})}$$

Then put $\pi_{12} = \frac{-n_{12} + n_{12}\pi_{11} - n_{21} + n_{21}\pi_{11}}{2(n_{12} + n_{21} + n_{22})}$ back to $\pi_{11} = \frac{n_{11} - 2n_{11}\pi_{12}}{n_{22} + n_{11}}$, solve for π_{11}

$$\text{Solve: } \pi_{11} = \frac{n_{11} - 2n_{11} \left(-\frac{-n_{12} + n_{12}\pi_{11} - n_{21} + n_{21}\pi_{11}}{2(n_{12} + n_{21} + n_{22})} \right)}{n_{22} + n_{11}}$$

$$\pi_{11} = \frac{\frac{2n_{11}(n_{12} + n_{21} + n_{22}) + 2n_{11}(-n_{12} + n_{12}\pi_{11} - n_{21} + n_{21}\pi_{11})}{2(n_{12} + n_{21} + n_{22})}}{n_{22} + n_{11}}$$

$$\pi_{11} = \frac{n_{11}(n_{12} + n_{21} + n_{22}) + n_{11}(-n_{12} + n_{12}\pi_{11} - n_{21} + n_{21}\pi_{11})}{(n_{12} + n_{21} + n_{22})(n_{22} + n_{11})}$$

$$\pi_{11} = \frac{n_{11}n_{12}\pi_{11} + n_{22}n_{11} + n_{11}n_{21}\pi_{11}}{n_{22}n_{12} + n_{11}n_{12} + n_{22}n_{21} + n_{11}n_{21} + n_{22}^2 + n_{22}n_{11}}$$

$$\pi_{11}n_{22}n_{12} + \pi_{11}n_{11}n_{12} + \pi_{11}n_{22}n_{21} + \pi_{11}n_{11}n_{21} + \pi_{11}n_{22}^2 + \pi_{11}n_{22}n_{11}$$

$$= n_{11}n_{12}\pi_{11} + n_{22}n_{11} + n_{11}n_{21}\pi_{11}$$

$$n_{22}\pi_{11}(n_{22} + n_{11} + n_{21} + n_{12}) = n_{22}n_{11}$$

$$\pi_{11} = \frac{n_{11}}{n_{22} + n_{12} + n_{11} + n_{21}}$$

Because $n_{22} + n_{12} + n_{11} + n_{21} = n$

therefore $\pi_{11} = \frac{n_{11}}{n}$

Replace $\pi_{11} = \frac{n_{11}}{n}$ to $\pi_{12} = -\frac{-n_{12} + n_{12}\pi_{11} - n_{21} + n_{21}\pi_{11}}{2(n_{12} + n_{21} + n_{22})}$

$$\pi_{12} = -\frac{-n_{12} + \frac{n_{12}n_{11}}{n} - n_{21} + \frac{n_{21}n_{11}}{n}}{2(n_{12} + n_{21} + n_{22})}$$

Because $n_{12} + n_{21} + n_{22} + n_{11} = n$, so $n - n_{11} = n_{12} + n_{21} + n_{22}$

$$\pi_{12} = -\frac{-n_{12}(n - n_{11})n_{21}(n - n_{11})}{2n(n - n_{11})}$$

$$\pi_{12} = \frac{n_{12} + n_{21}}{2n}$$

Therefore, $\pi_{12} = \frac{n_{12} + n_{21}}{2n}$

Since $\pi_{12} = \pi_{21}$

Therefore $\pi_{21} = \frac{n_{12}+n_{21}}{2n}$

Since $\pi_{22} = 1 - \pi_{11} - 2\pi_{12}$

Therefore $\pi_{22} = 1 - \frac{n_{11}}{n} - \frac{n_{12}+n_{21}}{2n} = \frac{n-n_{11}-n_{12}-n_{21}}{n} = \frac{n_{22}}{n}$

We obtain null maximum:

$$\tilde{\pi}_{11} = \frac{n_{11}}{n}$$

$$\tilde{\pi}_{12} = \frac{n_{12} + n_{21}}{2n}$$

$$\tilde{\pi}_{21} = \frac{n_{12} + n_{21}}{2n}$$

$$\tilde{\pi}_{22} = \frac{n_{22}}{n}$$

3.

$$\Omega = \frac{L(\text{null max})}{L(\text{overall max})}$$

Overall max: found by using MLE,

$$\text{we get } \hat{\pi}_{11} = \frac{n_{11}}{n}, \hat{\pi}_{12} = \frac{n_{12}}{n}, \hat{\pi}_{21} = \frac{n_{21}}{n}, \hat{\pi}_{22} = \frac{n_{22}}{n}$$

$$\Omega = \frac{L(\text{all max})}{L(\text{overall max})} = \frac{\tilde{\pi}_{11}^{n_{11}} \tilde{\pi}_{12}^{n_{12}} \tilde{\pi}_{21}^{n_{21}} \tilde{\pi}_{22}^{n_{22}}}{\hat{\pi}_{11}^{n_{11}} \hat{\pi}_{12}^{n_{12}} \hat{\pi}_{21}^{n_{21}} \hat{\pi}_{22}^{n_{22}}} = \prod_i^2 \prod_j^2 \left(\frac{\tilde{\pi}_{ij}}{\hat{\pi}_{ij}} \right)^{n_{ij}}$$

$$-2 \ln(\Omega) = -2 \sum_i^2 \sum_{ij}^2 n_{ij} \left[\ln \left(\frac{\tilde{\pi}_{ij}}{\hat{\pi}_{ij}} \right) \right] = 2 \sum_i^2 \sum_j^2 n_{ij} \left[\ln \left(\frac{\hat{\pi}_{ij}}{\tilde{\pi}_{ij}} \right) \right]$$

$$\text{known } \hat{\pi}_{ij} = \frac{n_{ij}}{n}$$

$$\therefore 2 \sum_i^2 \sum_j^2 n_{ij} \left[\ln \left(\frac{n_{ij}}{n \cdot \tilde{\pi}_{ij}} \right) \right]$$

$\therefore n\tilde{\pi}_{ij}$ is the expected # of observations under class ij under H_0

$$\therefore 2 \sum_{i=1}^2 \sum_{j=1}^2 Obs_{ij} \left[\ln \left(\frac{Obs_{ij}}{\text{Expected}_{ij}} \right) \right]$$

Based on Large sample theory

$$-2 \log(\Omega) \stackrel{H_0}{\sim} x_{\dim}^2(\Omega) - \dim(\omega_0)$$

$\dim(\Omega) : 3, 3$ out of 4 is free. Sum to 1.

$\dim(\omega_0) = 2 \quad \because \pi_{12} = \pi_{21}$ (one is free), \therefore sum to 1, the rest π_{11}, π_{22}

One of them is free

$$\therefore \dim(\Omega) - \dim(\omega_0) = 3 - 2 = 1$$

$$\therefore X_{df}^2 = 1$$

$$\therefore \dim(\Omega) - \dim(\omega_0) = 3 - 2 = 1$$

4.

$$Obs_{11} = n_{11}$$

$$Exp_{11} = n \cdot \frac{n_{11}}{n} = n_{11}$$

$$Obs_{21} = n_{21}$$

$$Exp_{12} = n \cdot \frac{n_2 + n_{21}}{2n} = \frac{n_2 + n_{21}}{2}$$

$$Obs_{12} = n_{12}$$

$$Exp_{21} = n \cdot \frac{n_{12} + n_{21}}{2n} = \frac{n_{12} + n_{21}}{2}$$

$$Obs_{22} = n_{22}$$

$$Exp_{22} = \frac{n_{22}}{n} \times n = n_{22}$$

$$\begin{aligned} X^2 &= \sum_{i=1}^2 \sum_{j=1}^2 \frac{(Obs_{ij} - Exp_{ij})^2}{Exp_{ij}} \\ &= \frac{(n_{11} - n_{11})^2}{n_{11}} + \frac{[n_{12} - \frac{(n_{12} + n_{21})}{2}]^2}{\frac{(n_{12} + n_{21})}{2}} + \frac{[n_{21} - \frac{(n_{12} + n_{21})}{2}]^2}{\frac{(n_{12} + n_{21})}{2}} + \frac{(n_{22} - n_{22})^2}{n_{22}} \\ &= \left(\frac{2n_{12} - n_{12} - n_{21}}{2} \right)^2 \times \frac{2}{n_2 + n_{21}} + \left(\frac{2n_{21} - n_{12} - n_{21}}{2} \right)^2 \times \frac{2}{n_{21} + n_{12}} \\ &= \frac{(n_{12}^2 - n_{21})^2}{2(n_{12} + n_{21})} + \frac{(n_{21} - n_{12})^2}{2(n_{21} + n_{12})} \\ &\because (n_{12} - n_{21})^2 = (n_{21} - n_{12})^2 \\ &\therefore = \frac{2(n_{12} - n_{21})^2}{2(n_{12} + n_{21})} = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}} \end{aligned}$$

Simulation

In this simulation study, I begin by setting the seed to ensure reproducibility of random numbers using ``set.seed(100)``. Following that, I define ``N = 1000`` to specify the number of simulations to conduct.

Next, I compute the critical value for the chi-square distribution. To do this, I set ``alpha`` to ``0.05`` as the significance level and ``df`` to ``1`` to represent the degrees of freedom which we derived earlier. The critical value ``cv`` is computed using ``qchisq(1 - alpha, df)``.

Moving on to the function ``generate_data()``, it is designed to simulate data based on specified parameters. This function takes four arguments: ``sample_sizes``, ``pi1``, ``add_on``, and ``corr_coef``, representing the number of observations, the proportion of successes for the first variable, an additional proportion for the second variable, and the correlation coefficient between the two variables, respectively. Within this function:

- I construct a correlation matrix ``corr_matrix`` using the provided correlation coefficient.
- I generate a dataset ``mydata`` using ``draw.correlated.binary()`` from the ``psych`` package.
- Counts of cases where the first variable is 1 and the second is 0 (``sim_n12``) and vice versa (``sim_n21``) are calculated.
- The chi-square test statistic ``sim_ts`` is computed.
- Finally, I check if ``sim_ts`` is greater than or equal to the critical value ``cv``, and return a logical result. If the test statistics is equal or larger than critical value, I reject the H_0 and it returns ``False``, otherwise, it returns ``True``.

To construct a comprehensive dataset for analysis, I utilized the ``expand.grid()`` function. This function facilitated the creation of a structured data frame named ``x``. This dataframe incorporates various combinations of parameters pivotal for our analysis. These parameters include:

- Sample sizes $n = 25, 40, 80, 200$
- π_{i1} with values $= 0.1, 0.4, 0.8$
- Add_on ($\pi_{i2} - \pi_{i1}$) $= 0, 0.02, 0.05, 0.1$
- Correlation Coefficient $= 0, 0.2, 0.5$

Essentially, ``x`` contains 48 distinct configurations, each representing a unique scenario for our simulation study.

Subsequently, I proceeded to conduct simulations and ascertain the proportion of true values across the diverse parameter combinations. Leveraging the capabilities of the ``dplyr`` package, I employed ``rowwise()`` in conjunction with ``mutate()`` to implement transformations iteratively across each row of the ``x`` dataframe. Within this iterative process, I employed the ``replicate()`` function to execute $N=1000$ simulations for each configuration. This involved invoking the ``generate_data()`` function, which is specifically designed to simulate data based on the provided parameters. It then computes a test statistic and evaluates whether it falls within the rejection region, determined by comparing it against the critical value derived earlier. By using

the mean function, I obtain 'proportion of successfully rejecting H_0 ' for each parameter combination.

Through this meticulous approach, I obtained a comprehensive dataset 'dataf' enriched with the proportion of true values for each parameter combination.

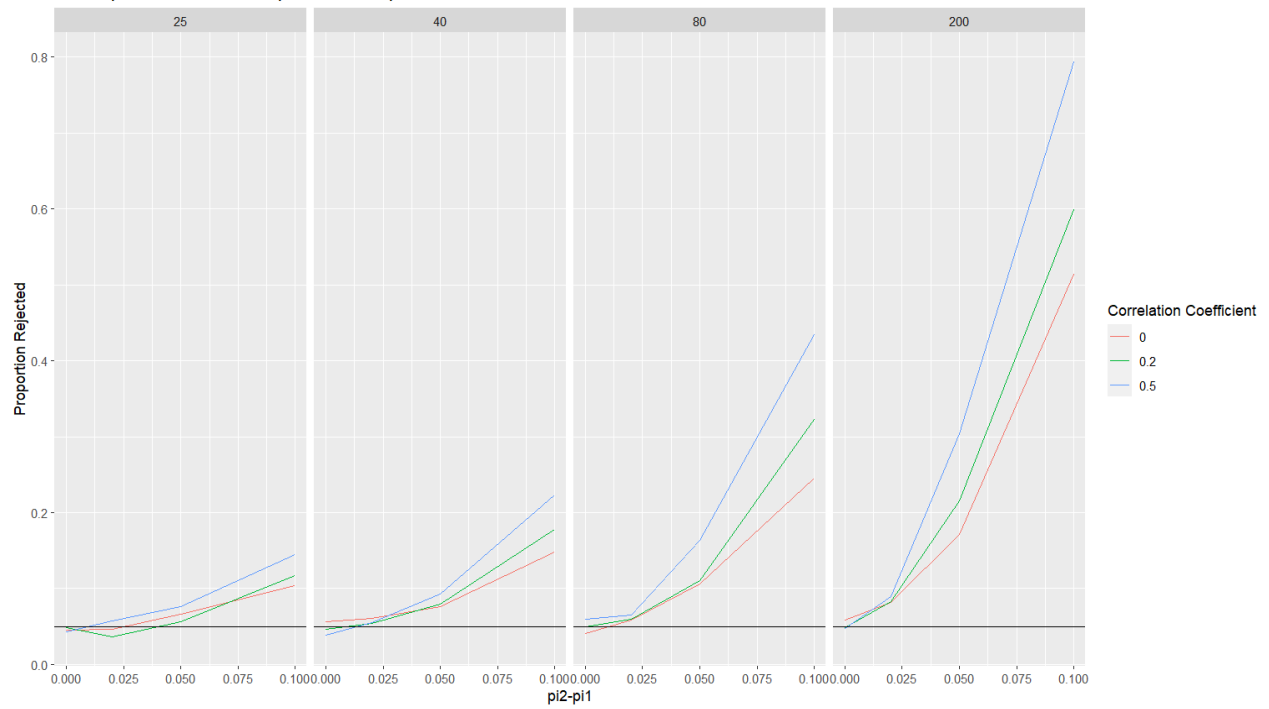
Result

Analyzing the data reveals a consistent trend: as the difference between the proportions ($\pi_2 - \pi_1$) increases, irrespective of π_1 values, sample size, or correlation coefficient, there is a noticeable escalation in power. This effect is amplified with larger sample sizes, indicating a more pronounced increase in power. For instance, when $\pi_1 = 0.1$ and $n = 25$, power remains negligible (near 0) for $\pi_2 - \pi_1 = 0$ but rises above 0.1 as the difference increases to 1, regardless of the correlation coefficient's influence depicted in Table 1. This pattern becomes more evident as the sample size expands.

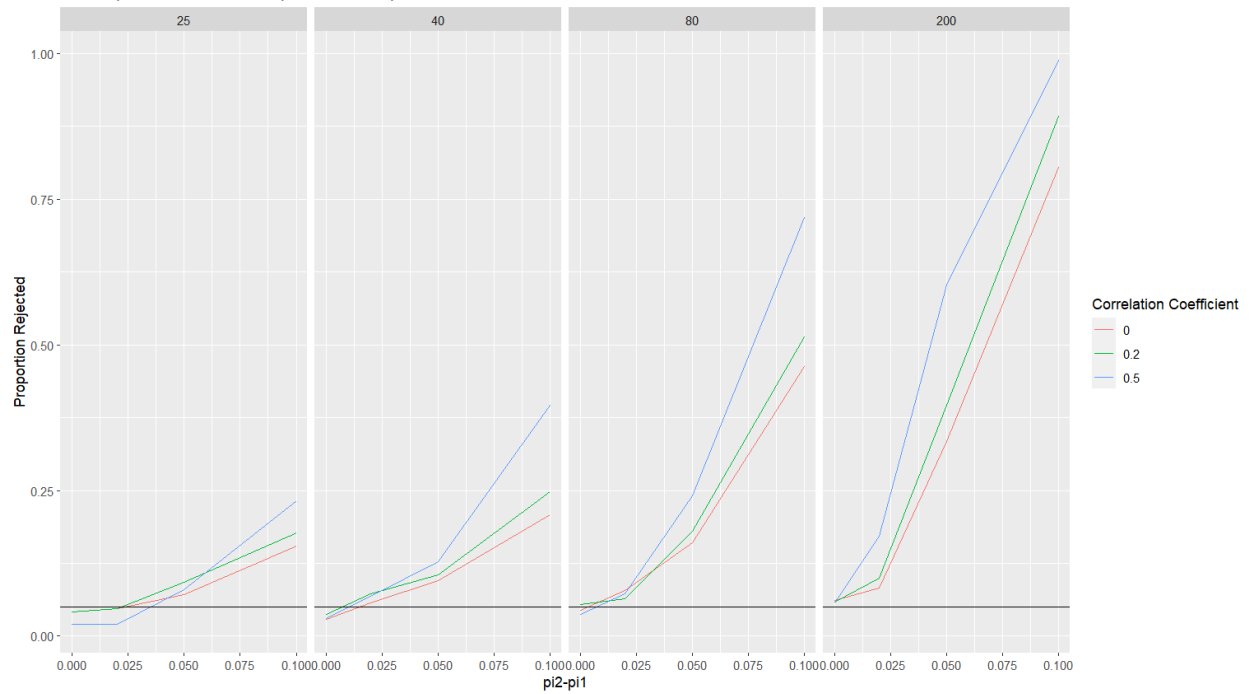
Moreover, observing Table 1, when $\pi_1 = 0.1$ and $n = 25$, all three lines converge near the critical value of 0.05 when $\pi_2 - \pi_1 = 0$. However, as this difference widens to 1, the power exceeds 0.75 across all correlation coefficient values, illustrating a substantial increase in power regardless of the correlation's impact. This consistent trend persists across various π_1 values in our simulation study.

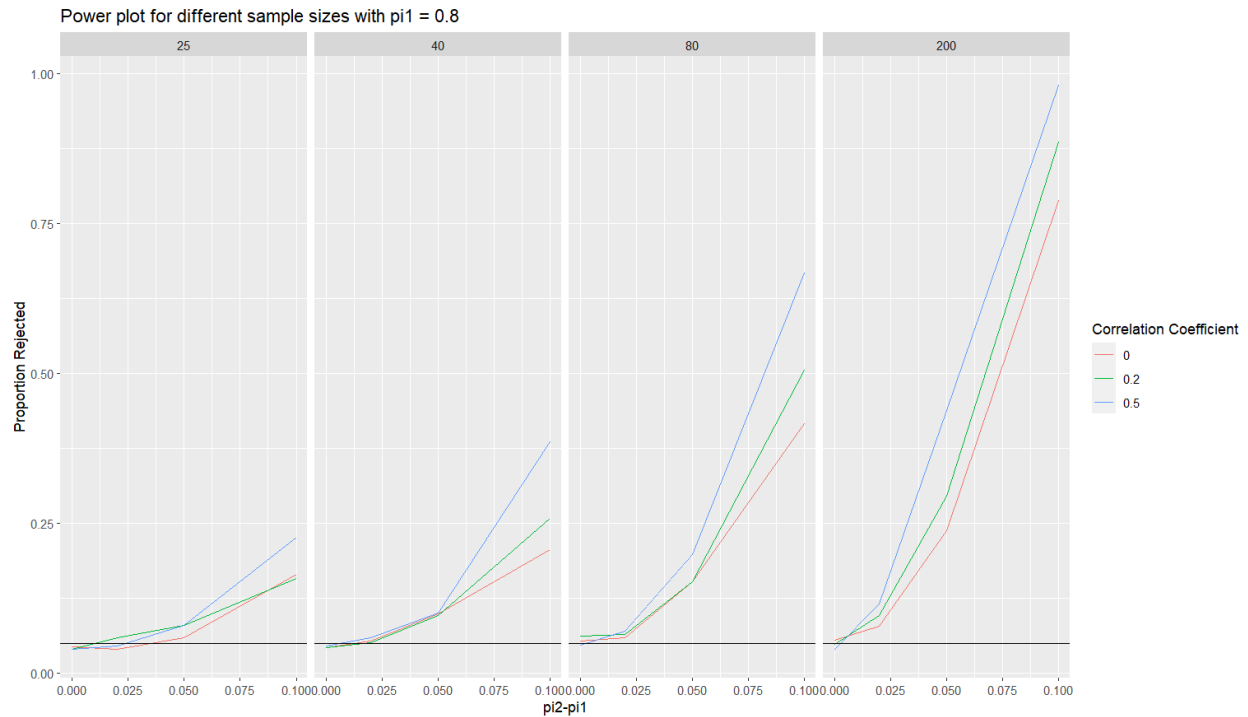
Additionally, a higher correlation coefficient between the data corresponds to heightened power when other variables remain constant. Furthermore, as the disparity between π_2 and π_1 ($i_2 - i_1$) grows larger, the discrepancy in power becomes more conspicuous among the three lines representing different correlation coefficients. These observations underscore the interplay between sample size, proportion differences, and correlation coefficients in influencing the statistical power of the analysis.

Power plot for different sample sizes with $\pi_1 = 0.4$



Power plot for different sample sizes with $\pi_1 = 0.1$





Analysis relate to McNemar's Test

The McNemar test, a non-parametric test for paired nominal data, exhibits notable trends in test characteristics. Similar to the observed pattern in the simulated data, the test's power is influenced by various factors, including the magnitude of differences between proportions, correlations between data, and sample size.

When applied to paired nominal data, such as before-and-after measurements or matched samples, the McNemar test's power tends to increase as the discrepancy between the proportions of the two categories ($\pi_2 - \pi_1$) becomes more significant. This aligns with the notion that larger differences between categories result in more detectable effects, leading to increased power to reject the null hypothesis. Furthermore, as sample size expands, the test's power experiences a more pronounced escalation. This effect is consistent with the statistical principle that larger sample sizes provide more reliable estimates and stronger evidence to detect true differences, thereby enhancing the test's power.

Moreover, the influence of correlation coefficients in the McNemar test, albeit less prominent compared to other factors, is notable. Higher correlation between paired observations may lead to increased power, all else being equal. This is because correlated observations tend to exhibit similar responses, making deviations from the null hypothesis more discernible.

Appendex

R-code:

```
#####;
#Author: Silvia Ying ;
#Purpose: ST502 Final Project ;
#Date: 4/22/2024 ;
#####;
#Package:
library("MultiRNG")
library('dplyr')
library('ggplot2')
#Task: Part 1-----
# Step 1: Collect Data
# 2x2 contingency table:
#           Drug B
#           success | failure |
# Drug A :success|  Obs_11=85  |  Obs_12=15 |
#       :failure|  Obs_21=40   |  Obs_22=110 |
Obs_11<-85
Obs_12<-15
Obs_21<-40
Obs_22<-110
# Step 2: Input Data
# Create a matrix to hold data
data_matrix <- matrix(c(Obs_11, Obs_12, Obs_21, Obs_22), nrow = 2, byrow = TRUE)

# Step 3: Compute McNemar's Test
result <- mcnemar.test(data_matrix, correct = FALSE)

# Step 4: Interpret Results
cat("McNemar's Test Statistic:", result$statistic, "\n")
cat("P-value:", result$p.value, "\n")

#-----
#Simulation Study
set.seed(100)
N=1000
# Computes chi-sq. critical value
alpha <- 0.05
df <- 1
cv <- qchisq(1-alpha, df)
```

```

# Function that extracts datapoint cases strictly from the rejection region:
generate_data <- function(sample_sizes, pi1, add_on, corr_coef){

  corr_matrix <- matrix(c(1, corr_coef,corr_coef,1), nrow = 2, ncol = 2)

  # dataset
  mydata <- draw.correlated.binary(no.row = sample_sizes, d = 2, prop.vec = c(pi1,
pi1+add_on), corr.mat = corr_matrix)

  # n12 n21 calculation
  sim_n12 <- sum(mydata[, 1] == 1 & mydata[, 2] == 0)
  sim_n21 <- sum(mydata[, 1] == 0 & mydata[, 2] == 1)

  # chi-sq. test statistic calculation using only sim_n12 and sim_n21
  sim_ts <- (sim_n12 - sim_n21)^2/(sim_n12 + sim_n21)

  # Check if both cv and sim_ts are valid numbers
  result<-(sim_ts >= cv)
  return(result)

}

#call the function
x<-expand.grid(
  sample_sizes =c(25,40,80,200), pi1 = c(0.1,0.4,0.8), add_on = c(0,0.02,0.05,0.1),
  corr_coef=c(0,0.2,0.5)
)
#Simulate and calculate prop_true for each parameter combination.
dataf<-x %>% rowwise()%>% mutate(prop_true = round(mean(replicate(N,
generate_data(sample_sizes, pi1, add_on, corr_coef)), na.rm = TRUE), digits = 3))

dataf
# Filter the data for pi1 == 0.1

# Plot for pi ==0.1
ggplot(dataf %>% filter(pi1 == 0.1), aes(x = add_on, y = prop_true, color = as.factor(corr_coef)))
+
  geom_line() +
  labs(
    title = "Power plot for different sample sizes with pi1 = 0.1",
    x = "pi2-pi1",

```

```

y = "Proportion Rejected",
color = "Correlation Coefficient"
) + facet_grid(rows = NULL, cols = vars(sample_sizes)) + # Arrange facets side by side
geom_hline(yintercept = 0.05, color = "black", linetype = "solid") # Add black horizontal line

# Plot for pi ==0.4
ggplot(dataf %>% filter(pi1 == 0.4), aes(x = add_on, y = prop_true, color = as.factor(corr_coef)))
+
geom_line() +
labs(
  title = "Power plot for different sample sizes with pi1 = 0.4",
  x = "pi2-pi1",
  y = "Proportion Rejected",
  color = "Correlation Coefficient"
) + facet_grid(rows = NULL, cols = vars(sample_sizes)) + # Arrange facets side by side
geom_hline(yintercept = 0.05, color = "black", linetype = "solid") # Add black horizontal line

# Plot for pi ==0.8
ggplot(dataf %>% filter(pi1 == 0.8), aes(x = add_on, y = prop_true, color = as.factor(corr_coef)))
+
geom_line() +
labs(
  title = "Power plot for different sample sizes with pi1 = 0.8",
  x = "pi2-pi1",
  y = "Proportion Rejected",
  color = "Correlation Coefficient"
) + facet_grid(rows = NULL, cols = vars(sample_sizes)) + # Arrange facets side by side
geom_hline(yintercept = 0.05, color = "black", linetype = "solid") # Add black horizontal line

```