ST502 Fall Final Project Report

Saurabh Gupta

December 10, 2024

Introduction to McNemar's Test

McNemar's test is a statistical method used to analyze paired nominal data, particularly when evaluating the consistency or changes in classifications across two related samples or conditions. It is often applied in situations where the goal is to compare two correlated proportions, such as the performance of a model or system before and after an intervention.

The test specifically examines the discordant pairs—cases where the outcomes differ between the two conditions—and determines whether the difference in these outcomes is statistically significant.

McNemar's test is a statistical test used on paired nominal data. It is applied to a 2×2 contingency table with a dichotomous trait, with matched pairs of subjects, to determine whether the row and column marginal frequencies are equal.

	Test 2 positive	**Test 2 negative**	**Row total**
Test 1 positive	a	b	$a+b$
Test 1 negative	c	d	$c+d$
Column total	$a+c$	$b+d$	N

Table 1: Contingency table for McNemar's test

The null hypothesis of marginal homogeneity states that the two marginal probabilities for each outcome are the same, i.e.,

$$p_a + p_b = p_a + p_c \quad \text{and} \quad p_c + p_d = p_b + p_d.$$

Thus, the null and alternative hypotheses are:

$$H_0 : p_b = p_c \quad \text{and} \quad H_1 : p_b \neq p_c.$$

Here, p_a, p_b, p_c, and p_d denote the theoretical probability of occurrences in cells with the corresponding label.

The McNemar test statistic is:

$$\chi^2 = \frac{(b-c)^2}{b+c}.$$

Under the null hypothesis, with a sufficiently large number of discordants (cells b and c), χ^2 has a chi-squared distribution with 1 degree of freedom.

Where It Is Used

McNemar's test is commonly applied in fields such as:

- **Medical Research**: To compare diagnostic tests or treatments on the same subjects. For example, assessing the accuracy of two diagnostic tools or the effectiveness of a treatment over time.
- **Machine Learning and AI**: To compare the performance of classification models on the same dataset. For instance, determining whether a new model significantly outperforms an existing one in terms of misclassification rates.

- **Psychology and Behavioral Studies**: To analyze pre-test and post-test changes in categorical outcomes, such as attitudes, behaviors, or preferences, for the same subjects.
- **Survey Research**: To evaluate shifts in opinions or choices among the same respondents over time.

In summary, McNemar's test is a valuable tool when dealing with paired categorical data and provides insights into whether differences observed between two conditions are statistically significant.

Analysis of Acid Reflux Treatment the dataset

0.1 Observed Frequencies

The contingency table for paired responses is:

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

Consider a dataset consisting of 250 subjects, each treated for acid reflux with either Drug A or Drug B. The goal is to determine whether the two drugs have a different probability of success in relieving acid reflux. The response variable is binary, either *success* (reflux stopped) or *failure* (reflux still present).

The table above represents the number of successes and failures for each drug. However, because the study is a within-subject design (each subject uses both drugs), the observations are not independent. Thus, we cannot perform a test of homogeneity.

Statistical Test: McNemar's Test

To determine if there is a significant difference between the effects of Drug A and Drug B, we apply **McNemar's test** to the table of concordant and discordant pairs. McNemar's test is designed for paired nominal data.

$$\chi^2 = \frac{(b-c)^2}{b+c}$$

The null hypothesis (H_0) is that the two drugs have the same probability of success, meaning the discordant pairs should be balanced. The alternative hypothesis (H_1) is that the drugs have different probabilities of success, leading to an imbalance in the discordant pairs.

McNemar's test will help us determine if the observed imbalance in discordant pairs (15 and 40) is statistically significant. If the test statistic is significant, we conclude that Drug A and Drug B have different effects on the probability of reflux relief.

0.2 Hypotheses

- Null hypothesis (H_0): Drugs A and B have the same probability of success ($\pi_{12} = \pi_{21}$).
- Alternative hypothesis (H_A): The probabilities differ ($\pi_{12} \neq \pi_{21}$).

0.3 Test Statistic

The test statistic is:

$$X^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}},$$

where:

$$n_{12} = 15, \quad n_{21} = 40.$$

Computing:

$$X^2 = \frac{(15-40)^2}{15+40} = \frac{625}{55} = 11.36.$$

0.4 Rejection Region

The reference distribution is χ^2_1 (chi-square with 1 degree of freedom). For $\alpha = 0.05$, the critical value is:

$$\chi^2_{1,0.05} = 3.841.$$

Since $X^2 = 11.36 > 3.841$, we reject H_0. This provides strong evidence that the drugs differ in their probability of success.

0.5 P-Value

The p-value is:

$$P = P(\chi^2_1 \geq 11.36) = 0.00076.$$

R Code: McNemar's Test and Calculation of Test Statistic

```r
# Observed concordant/discordant pairs data
dataset <- matrix(c(85, 15, 40, 110), nrow = 2, byrow = TRUE,
                  dimnames = list("Drug A" = c("Success", "Failure"),
                                  "Drug B" = c("Success", "Failure")))
print(dataset)

# Extract discordant counts
n12 <- dataset[1, 2] # Drug A Success, Drug B Failure
n21 <- dataset[2, 1] # Drug A Failure, Drug B Success

# Calculate test statistic
X2_value <- (n12 - n21)^2 / (n12 + n21)

# p-value
p_value <- pchisq(X2_value, df = 1, lower.tail = FALSE)

# Critical value or rejection region for alpha = 0.05
alpha <- 0.05
critical_value <- qchisq(1 - alpha, df = 1)

# Decision
decision <- ifelse(X2_value > critical_value, "Reject H0", "Fail to Reject H0")

# Results
list(test_statistic = X2_value, p_value = p_value,
critical_value = critical_value, decision = decision)

# Perform McNemar's test without continuity correction
mcnemar_result <- mcnemar.test(dataset, correct = FALSE)
print(mcnemar_result)
```

Null Hypothesis proof

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	$\pi_{11}(concordant)$	$\pi_{12}(discordant)$	$\pi_{1\bullet}$
Failure	$\pi_{21}(discordant)$	$\pi_{22}(concordant)$	$\pi_{2\bullet}$
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!

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

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

1. Define Notations

From the contingency table of probabilities:

$$\pi_{ij} \quad \text{is the probability for the } (i,j)\text{-th cell in the table.}$$

The probabilities in the table of concordant/discordant pairs are:

$$\pi_{11} : \quad \text{Probability of success for both Drug A and Drug B.}$$
$$\pi_{12} : \quad \text{Probability of success for Drug A and failure for Drug B.}$$
$$\pi_{21} : \quad \text{Probability of failure for Drug A and success for Drug B.}$$
$$\pi_{22} : \quad \text{Probability of failure for both Drug A and Drug B.}$$

The marginal probabilities are defined as:

$$\pi_{1\cdot} = \pi_{11} + \pi_{12} : \quad \text{Probability that Drug A was successful (irrespective of Drug B).}$$
$$\pi_{\cdot 1} = \pi_{11} + \pi_{21} : \quad \text{Probability that Drug B was successful (irrespective of Drug A).}$$
$$\pi_{2\cdot} = \pi_{21} + \pi_{22} : \quad \text{Probability that Drug A failed (irrespective of Drug B).}$$
$$\pi_{\cdot 2} = \pi_{12} + \pi_{22} : \quad \text{Probability that Drug B failed (irrespective of Drug A).}$$

The null hypothesis states:

$$H_0 : \pi_{1\cdot} = \pi_{\cdot 1} \quad \text{and} \quad \pi_{2\cdot} = \pi_{\cdot 2}$$

2. Null Hypothesis Equivalence

The null hypothesis states that:

$$\pi_{1\cdot} = \pi_{\cdot 1}, \quad \pi_{2\cdot} = \pi_{\cdot 2}.$$

Definitions of Marginal Probabilities
The marginal probabilities for rows and columns are defined as:

$$\pi_{1\cdot} = \pi_{11} + \pi_{12}, \quad \pi_{\cdot 1} = \pi_{11} + \pi_{21}.$$

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

Equating Marginals Under H_0
Under the null hypothesis, we assume:

$$\pi_{1\cdot} = \pi_{\cdot 1}.$$

Substituting the definitions:

$$\pi_{11} + \pi_{12} = \pi_{11} + \pi_{21}.$$

Simplifying by canceling π_{11} from both sides:

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

For the second condition:

$$\pi_{2\cdot} = \pi_{\cdot 2}.$$

Substituting the definitions:

$$\pi_{21} + \pi_{22} = \pi_{12} + \pi_{22}.$$

Simplifying by canceling π_{22} from both sides:

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

3. Interpret the Result

From the above, both conditions reduce to:

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

Thus, under the null hypothesis, the equivalence of row and column marginals ($\pi_{1\cdot} = \pi_{\cdot 1}$ and $\pi_{2\cdot} = \pi_{\cdot 2}$) implies that the probabilities of discordant pairs are equal ($\pi_{12} = \pi_{21}$).

Deriving the Maximum Likelihood Estimates (MLEs)

Under the null hypothesis restriction $\pi_{12} = \pi_{21}$, derive the MLEs for $\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}$.

Step 1: Define the Constraints

1. Probability Constraint:

$$\pi_{11} + \pi_{12} + \pi_{21} + \pi_{22} = 1$$

2. Null Hypothesis Restriction:

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

Step 2: Likelihood Function

$$L(\pi_{11}, \pi_{12}, \ldots, \pi_{IJ}) \propto \pi_{11}^{n_{11}} \pi_{12}^{n_{12}} \cdot \cdots \cdot \pi_{IJ}^{n_{IJ}}$$

Step 3: Log-Likelihood Function

The multinomial log-likelihood function is:

$$l(\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}) = c + n_{11} \ln(\pi_{11}) + n_{12} \ln(\pi_{12}) + n_{21} \ln(\pi_{21}) + n_{22} \ln(\pi_{22})$$

Using the null restriction ($\pi_{12} = \pi_{21}$):

$$l(\pi_{11}, \pi_{12}, \pi_{22}) = c + n_{11} \ln(\pi_{11}) + (n_{12} + n_{21}) \ln(\pi_{12}) + n_{22} \ln(\pi_{22})$$

Step 4: Apply Lagrange Multipliers

Define the Lagrange function: Given the sum-to-1 constraint, we know that:

$$\sum_{i=1}^{I} \pi_{i\cdot} = 1, \quad \sum_{j=1}^{J} \pi_{\cdot j} = 1$$

Either use the substitute method or use Lagrange multipliers.
Write the constraints equal to 0:

$$\sum_{i=1}^{I} \pi_{i\cdot} - 1 = 0, \quad \sum_{j=1}^{J} \pi_{\cdot j} - 1 = 0$$

Step 5: Derivatives

Taking partial derivatives of $\mathcal{L}$ with respect to $\pi_{i\cdot}, \pi_{\cdot j}, \lambda_1, \lambda_2$, and setting them to zero:

1. With respect to $\pi_{i\cdot}$:

$$\frac{\partial \mathcal{L}}{\partial \pi_{i\cdot}} = \frac{\sum_{j=1}^{J} n_{ij}}{\pi_{i\cdot}} - \lambda_1 = 0 \quad \Rightarrow \quad \pi_{i\cdot} = \frac{\sum_{j=1}^{J} n_{ij}}{\lambda_1}$$

2. With respect to $\pi_{\cdot j}$:

$$\frac{\partial \mathcal{L}}{\partial \pi_{\cdot j}} = \frac{\sum_{i=1}^{I} n_{ij}}{\pi_{\cdot j}} - \lambda_2 = 0 \quad \Rightarrow \quad \pi_{\cdot j} = \frac{\sum_{i=1}^{I} n_{ij}}{\lambda_2}$$

3. With respect to λ_1 (row sum constraint):

$$\frac{\partial \mathcal{L}}{\partial \lambda_1} = \sum_{i=1}^{I} \pi_{i\cdot} - 1 = 0$$

4. With respect to λ_2 (column sum constraint):

$$\frac{\partial \mathcal{L}}{\partial \lambda_2} = \sum_{j=1}^{J} \pi_{\cdot j} - 1 = 0$$

Step 6: Solving for λ_1 and λ_2

Using the row sum constraint:

$$\sum_{i=1}^{I} \pi_{i\cdot} = 1 \quad \Rightarrow \quad \sum_{i=1}^{I} \frac{\sum_{j=1}^{J} n_{ij}}{\lambda_1} = 1$$

Simplify:

$$\lambda_1 = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} = n$$

Similarly, using the column sum constraint:

$$\lambda_2 = n$$

Step 7: Final MLEs

Substituting $\lambda_1 = n$ and $\lambda_2 = n$ back into the equations for $\pi_{i\cdot}$ and $\pi_{\cdot j}$:

$$\pi_{i\cdot} = \frac{\sum_{j=1}^{J} n_{ij}}{n}, \quad \pi_{\cdot j} = \frac{\sum_{i=1}^{I} n_{ij}}{n}$$

Deriving the Form of the Likelihood Ratio Test (LRT)

Here's how to derive the form of the Likelihood Ratio Test (LRT) for the given problem and establish its equivalence to the specified formula.

The likelihood ratio statistic is:

$$\Lambda = \frac{L(\text{restricted})}{L(\text{unrestricted})},$$

where L is the likelihood function.

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

The test statistic is:

$$-2 \ln(\Lambda) = -2 \left(\ln L(\text{restricted}) - \ln L(\text{unrestricted}) \right).$$

1. Likelihood Ratio Test Framework

The likelihood ratio statistic is defined as:

$$\text{LRT} = -2 \ln \left(\frac{L(\tilde{\pi}_{ij})}{L(\hat{\pi}_{ij})} \right),$$

where:

- $\tilde{\pi}_{ij}$ are the Maximum Likelihood Estimates (MLEs) under the null hypothesis H_0,
- $\hat{\pi}_{ij}$ are the MLEs under the alternative hypothesis H_A.

The multinomial likelihood is:

$$L(\pi) = \prod_{i=1}^{2} \prod_{j=1}^{2} \pi_{ij}^{n_{ij}}.$$

Taking the log of the likelihood:

$$\ln L(\pi) = \sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij} \ln \pi_{ij}.$$

2. Substituting MLEs

(a) Under the Null Hypothesis (H_0): The null hypothesis assumes symmetry ($\pi_{12} = \pi_{21}$), Substituting into the total probability constraint:

$$\pi_{11} + x + x + \pi_{22} = 1 \quad \Rightarrow \quad \pi_{11} + \pi_{22} + 2x = 1.$$

Using the property of MLE for multinomial probabilities:

$$\pi_{ij} = \frac{\text{Expected count in cell } (i,j)}{\text{Total number of observations}}.$$

leading to:

$$\tilde{\pi}_{11} = \frac{n_{11}}{n}, \quad \tilde{\pi}_{22} = \frac{n_{22}}{n}, \quad \tilde{\pi}_{12} = \tilde{\pi}_{21} = \frac{n_{12} + n_{21}}{2n}.$$

The log-likelihood under H_0 is:

$$\ln L(\tilde{\pi}_{ij}) = n_{11} \ln \tilde{\pi}_{11} + n_{12} \ln \tilde{\pi}_{12} + n_{21} \ln \tilde{\pi}_{21} + n_{22} \ln \tilde{\pi}_{22}.$$

Substituting the MLEs:

$$\ln L(\tilde{\pi}_{ij}) = n_{11} \ln \frac{n_{11}}{n} + n_{12} \ln \frac{n_{12} + n_{21}}{2n} + n_{21} \ln \frac{n_{12} + n_{21}}{2n} + n_{22} \ln \frac{n_{22}}{n}.$$

(b) Under the Alternative Hypothesis (H_A): The alternative hypothesis assumes no restrictions, so:

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

The log-likelihood under H_A is:

$$\ln L(\hat{\pi}_{ij}) = n_{11} \ln \frac{n_{11}}{n} + n_{12} \ln \frac{n_{12}}{n} + n_{21} \ln \frac{n_{21}}{n} + n_{22} \ln \frac{n_{22}}{n}.$$

3. Likelihood Ratio Test Statistic

$$-2\ln\left(\frac{L(\tilde{\pi}_{11},\tilde{\pi}_{12},\tilde{\pi}_{21},\tilde{\pi}_{22})}{L(\hat{\pi}_{11},\hat{\pi}_{12},\hat{\pi}_{21},\hat{\pi}_{22})}\right) = -2\ln\left(\frac{L(\tilde{\pi})}{L(\hat{\pi})}\right)$$

The LRT is:

$$\text{LRT} = -2\left(\ln L(\tilde{\pi}_{ij}) - \ln L(\hat{\pi}_{ij})\right).$$

Substituting $\ln L(\tilde{\pi}_{ij})$ and $\ln L(\hat{\pi}_{ij})$:

$$\text{LRT} = -2\Bigg[\left(n_{11}\ln\frac{n_{11}}{n} + n_{12}\ln\frac{n_{12}+n_{21}}{2n} + n_{21}\ln\frac{n_{12}+n_{21}}{2n} + n_{22}\ln\frac{n_{22}}{n}\right)$$

$$-\left(n_{11}\ln\frac{n_{11}}{n} + n_{12}\ln\frac{n_{12}}{n} + n_{21}\ln\frac{n_{21}}{n} + n_{22}\ln\frac{n_{22}}{n}\right)\Bigg].$$

Simplify term by term:

- The terms $n_{11}\ln\frac{n_{11}}{n}$ and $n_{22}\ln\frac{n_{22}}{n}$ cancel out.
- For n_{12} and n_{21}:

$$n_{12}\ln\frac{\frac{n_{12}+n_{21}}{2n}}{\frac{n_{12}}{n}} + n_{21}\ln\frac{\frac{n_{12}+n_{21}}{2n}}{\frac{n_{21}}{n}}.$$

Combine:

$$\text{LRT} = 2\left[n_{12}\ln\frac{n_{12}}{\frac{n_{12}+n_{21}}{2}} + n_{21}\ln\frac{n_{21}}{\frac{n_{12}+n_{21}}{2}}\right].$$

Using logarithmic properties:

$$\text{LRT} = 2\sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij}\ln\frac{\text{Obs}_{ij}}{\text{Exp}_{ij}}.$$

Thus, degrees of freedom:

$$\text{df} = 3 - 2 = 1.$$

5. Final Result

The LRT statistic is:

$$\text{LRT} = 2\sum_{i=1}^{2}\sum_{j=1}^{2} \text{Obs}_{ij}\ln\frac{\text{Obs}_{ij}}{\text{Exp}_{ij}}.$$

It follows a χ_1^2-distribution, aligning with the Pearson's chi-square statistic, demonstrating their asymptotic equivalence.

Simplification of Pearson's Chi-Square Test Statistic

To show that Pearson's chi-square test statistic simplifies to:

$$X^2 = \frac{(n_{12}-n_{21})^2}{n_{12}+n_{21}},$$

we proceed as follows:

1. Pearson's Chi-Square Formula

The general form of Pearson's chi-square test statistic is:

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(\text{Obs}_{ij} - \text{Exp}_{ij})^2}{\text{Exp}_{ij}},$$

where:

- Obs_{ij} are the observed counts in each cell,
- Exp_{ij} are the expected counts under the null hypothesis.

For the 2×2 table of concordant and discordant pairs:

	Drug B: Success	Drug B: Failure	Total
Drug A: Success	n_{11}	n_{12}	$n_{1\cdot}$
Drug A: Failure	n_{21}	n_{22}	$n_{2\cdot}$
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	n

Here:

- n_{11} and n_{22}: Concordant pairs,
- n_{12} and n_{21}: Discordant pairs.

2. Focus on Discordant Pairs

McNemar's test specifically focuses on paired data (discordant pairs), and it does not assume independence between rows and columns. Instead, McNemar's test evaluates the difference in probabilities between discordant pairs. Under the null hypothesis, the expected counts for the discordant pairs (Exp_{12} and Exp_{21}) are:

$$\text{Exp}_{12} = \text{Exp}_{21} = \frac{n_{12} + n_{21}}{2}.$$

The chi-square statistic for the discordant pairs is:

$$X^2 = \frac{(n_{12} - \text{Exp}_{12})^2}{\text{Exp}_{12}} + \frac{(n_{21} - \text{Exp}_{21})^2}{\text{Exp}_{21}}.$$

Since $\text{Exp}_{12} = \text{Exp}_{21} = \frac{n_{12}+n_{21}}{2}$, substitute this into the formula:

$$X^2 = \frac{(n_{12} - \frac{n_{12}+n_{21}}{2})^2}{\frac{n_{12}+n_{21}}{2}} + \frac{(n_{21} - \frac{n_{12}+n_{21}}{2})^2}{\frac{n_{12}+n_{21}}{2}}.$$

3. Simplify Each Term

For n_{12}:

$$n_{12} - \frac{n_{12} + n_{21}}{2} = \frac{2n_{12} - n_{12} - n_{21}}{2} = \frac{n_{12} - n_{21}}{2}.$$

Squaring this term:

$$\left(n_{12} - \frac{n_{12} + n_{21}}{2}\right)^2 = \left(\frac{n_{12} - n_{21}}{2}\right)^2 = \frac{(n_{12} - n_{21})^2}{4}.$$

Divide by $\frac{n_{12}+n_{21}}{2}$:

$$\frac{\frac{(n_{12}-n_{21})^2}{4}}{\frac{n_{12}+n_{21}}{2}} = \frac{(n_{12} - n_{21})^2}{2(n_{12} + n_{21})}.$$

Similarly, for n_{21}, the result is the same:

$$\frac{(n_{12} - n_{21})^2}{2(n_{12} + n_{21})}.$$

4. Combine the Terms

Add the two terms together:

$$X^2 = \frac{(n_{12} - n_{21})^2}{2(n_{12} + n_{21})} + \frac{(n_{12} - n_{21})^2}{2(n_{12} + n_{21})}.$$

$$X^2 = \frac{2(n_{12} - n_{21})^2}{2(n_{12} + n_{21})}.$$

$$X^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}.$$

5. Final Result

Thus, Pearson's chi-square statistic simplifies to:

$$X^2 = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}.$$

Simulation Study

The goal of this simulation study is to assess:

- How well the Pearson chi-square test controls the type I error rate (denoted as α)
- The power of the Pearson chi-square test when comparing certain alternatives

Study Setup

We will generate correlated binary data using the `draw.correlated.binary()` function from the `MultiRNG` package. This function requires the following arguments:

- `no.row`: Sample size.
- `d`: Number of variables to create (we are generating 2 variables).
- `prop.vec`: A vector of true proportions of success for the two variables.
- `corr.mat`: A 2x2 correlation matrix, where diagonal elements are 1, and off-diagonal elements represent the correlation between the two variables.

Parameters for Data Generation

We will generate data under all combinations of the following parameters:

- `n` = 25, 40, 80, 200 (with `n` = 40 replaced by `n` = 50).
- π_1 (success probability for Drug A) = 0.1, 0.4, 0.8.
- π_2 (success probability for Drug B) = π_1, $\pi_1 + 0.02$, $\pi_1 + 0.05$, $\pi_1 + 0.1$.
- ρ (correlation between the two drugs) = 0, 0.2, 0.5.

For each combination of the above parameters, we will generate 1000 datasets. We will use the case where $\pi_1 = \pi_2$ to assess the α control of the Pearson chi-square test. The other combinations (where $\pi_1 \neq \pi_2$) will be used to investigate the power of the test.

Test Setup

A function will be written to perform *McNemar's test* based on the sample size, correlation, and success probabilities π_1 and π_2. The output of this function will simply be whether we reject or fail to reject the null hypothesis (TRUE/FALSE).

R Code: Parameter setup and sample dataset

```r
# Define parameters
pi1_values <- c(0.1, 0.4, 0.8)          # 1 values
pi2_offsets <- c(0, 0.02, 0.05, 0.1)    # Offsets for 2
n_values <- c(25, 50, 80, 200)          # Sample sizes
rho_values <- c(0, 0.2, 0.5)            # Correlations

# Initialize an empty data frame to store the results
final_dataset <- data.frame()

# Create the base grid for n, rho, and pi1
sample_dataset <- expand.grid(
  n = n_values,
  rho = rho_values,
  pi1 = pi1_values
)

# Loop over each row in the base grid
for (i in seq_len(nrow(sample_dataset))) {
  # Extract the current row
  row <- sample_dataset[i, ]

  # Compute the 2 values for the current 1 using offsets
  pi2_values <- row$pi1 + pi2_offsets

  # Create rows for each 2 value
  temp <- data.frame(
    n = row$n,
    rho = row$rho,
    pi1 = row$pi1,
    pi2 = pi2_values
  )

  # Append to the final results
  final_dataset <- rbind(final_dataset, temp)
}

# View the final results
print(final_dataset)
```

R Code: McNemar Simulation and execution

```r

# Function to run a single McNemar's test simulation
McNemarSimulation <- function(n, pi1, pi2, rho) {
  # Generate correlated binary data
  data <- draw.correlated.binary(
    no.row = n,
    d = 2,
    prop.vec = c(pi1, pi2),
    corr.mat = matrix(c(1, rho, rho, 1), nrow = 2)
  )

```

```r
  # Create 2x2 contingency table
  n11 <- sum(data[, 1] == 1 & data[, 2] == 1) # Concordant
  n12 <- sum(data[, 1] == 1 & data[, 2] == 0) # Discordant
  n21 <- sum(data[, 1] == 0 & data[, 2] == 1) # Discordant
  n22 <- sum(data[, 1] == 0 & data[, 2] == 0) # Concordant

  # Calculate McNemar's test statistic
  if (n12 + n21 == 0) return(FALSE) # Prevent division by zero
  X2 <- (n12 - n21)^2 / (n12 + n21)
  p_value <- 1 - pchisq(X2, df = 1)

  # Determine whether to reject the null hypothesis (H0)
  reject_null <- p_value < 0.05

  # Return the result (TRUE if H0 is rejected, FALSE otherwise)
  reject_null
}

# Function to run multiple McNemar's test simulations and  rejection proportion
RunMcNemarSimulation <- function(n, pi1, pi2, rho, N = 1000) {
  results <- replicate(N, McNemarSimulation(n, pi1, pi2, rho))
  rejection_proportion <- mean(results) # Compute proportion of rejections
  return(rejection_proportion)
}

# Add a column to store rejection rates
final_dataset$rejection_rate <- NA

# Run simulations
for (i in seq_len(nrow(final_dataset))) {
  n <- final_dataset$n[i]
  pi1 <- final_dataset$pi1[i]
  pi2 <- final_dataset$pi2[i]
  rho <- final_dataset$rho[i]

  final_dataset$rejection_rate[i] <- RunMcNemarSimulation(n, pi1, pi2, rho)
  #print(final_dataset£rejection_rate[i])
}

# Check summary of rejection rates
summary(final_dataset$rejection_rate)
```

R Code: Graphing and Plotting

```r

library(ggplot2)

# Loop over each pi1 value
for (pi_1 in pi1_values) {

  # Filter the data for the current pi1 value
  filtered_df <- final_dataset[final_dataset$pi1 == pi_1, ]
```

```
  # Create the plot
  pt <- ggplot(filtered_df, aes(x = pi2 - pi1, y = rejection_rate,
  color = factor(rho), group = rho)) +
    geom_line() +
    facet_grid(. ~ n) +  # Facets for different sample sizes
    labs(
      title = sprintf("Rejection Rate for different sample sizes and
      correlations with pi1 = %.2f", pi_1),
      x = expression(pi[2] - pi[1]),
      y = "Rejection Rate",
      color = "Correlation ()"
    ) +
    theme_minimal()

  # Print the plot
  print(pt)
}
```

Power Analysis for Different Values of π_1

We conducted simulations to explore the rejection rates of McNemar's test for different values of π_1 and sample sizes. The following plots illustrate the rejection rates (y-axis) for various values of π_1 and correlation values (ρ).

Key Observations

- As the sample size increases (from $n = 25$ to $n = 200$), the rejection rates for McNemar's test consistently increase for all values of correlation (ρ). Larger sample sizes provide more power, allowing the test to detect differences between π_1 and π_2 more effectively.
- For a given sample size, higher correlation values (e.g., $\rho = 0.5$) lead to higher rejection rates compared to lower correlations (e.g., $\rho = 0$ or $\rho = 0.25$). This indictes that the dependency between paired observations influences the test's power.
- At smaller sample sizes ($n = 25$), rejection rates are low across all conditions, even for larger differences in $\pi_2 - \pi_1$. At larger sample sizes ($n = 200$), rejection rates approach 1 for large $\pi_2 - \pi_1$ values, particularly when the correlation is high ($\rho = 0.5$).

Summary for Report

- The power of McNemar's test increases with larger sample sizes. Larger sample sizes improve the ability of the test to detect differences in paired categorical data.
- The correlation between the two treatments plays a crucial role. Higher correlations increase the test's power, especially in the context of larger sample sizes and more significant differences between the success probabilities.
- The test is more effective at detecting differences as the magnitude of $\pi_2 - \pi_1$ increases. However, it is less sensitive to small differences, particularly with smaller sample sizes or lower correlations.

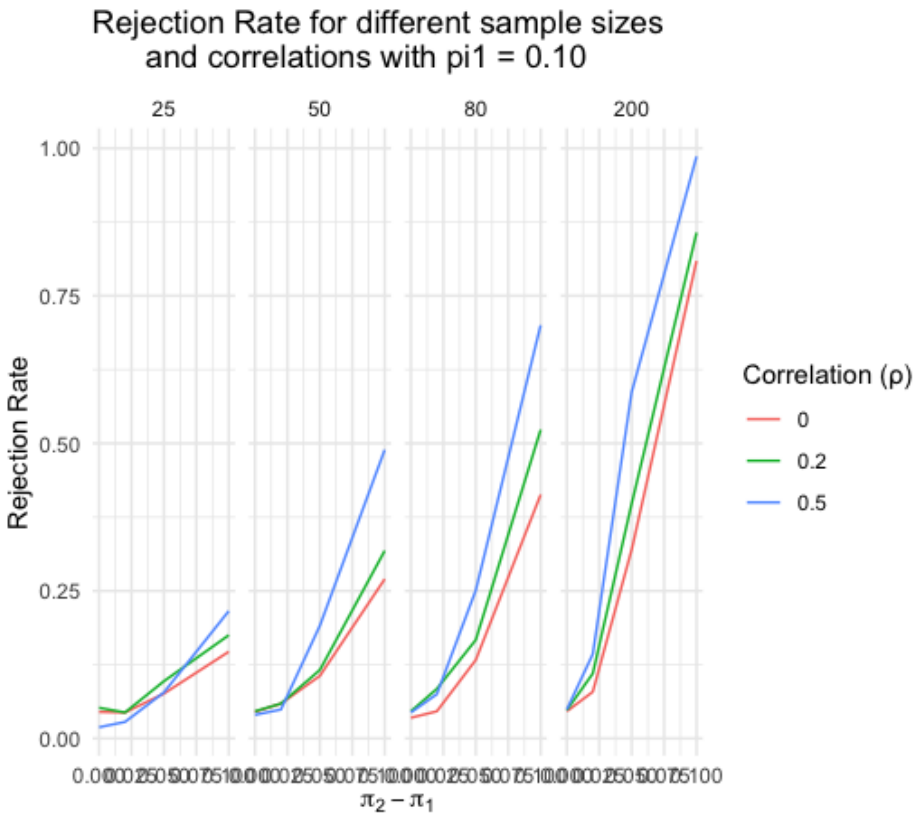

Figure 1: Rejection Rate for McNemar's Test with $\pi_1 = 0.1$.

- For effective use of McNemar's test, it's crucial to have larger sample sizes and consider the correlation structure between treatments. Small sample sizes and small differences in success probabilities ($\pi_2 - \pi_1 \leq 0.05$) may lead to insufficient power, making the test less effective.

In summary, the charts indicate that having larger sample sizes, higher initial probabilities, and stronger relationships between measurements enhance the power of a test. Additionally, even with smaller sample sizes, significant differences can still result in satisfactory power, especially when the correlations are stronger.

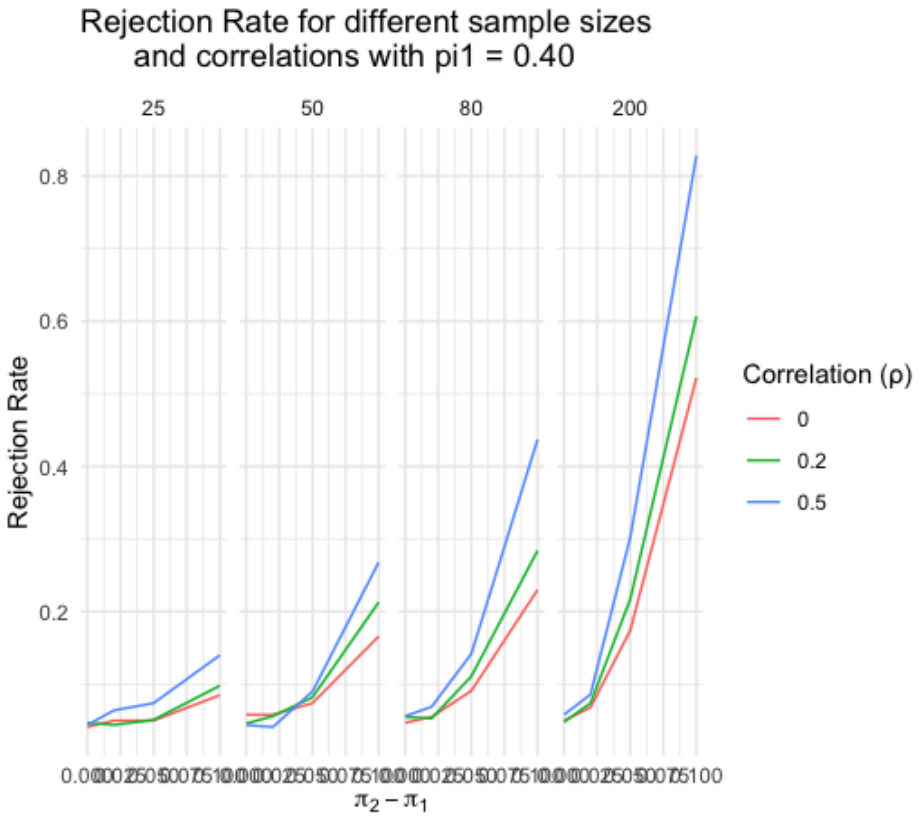

Figure 2: Rejection Rate for McNemar's Test with $\pi_1 = 0.4$.

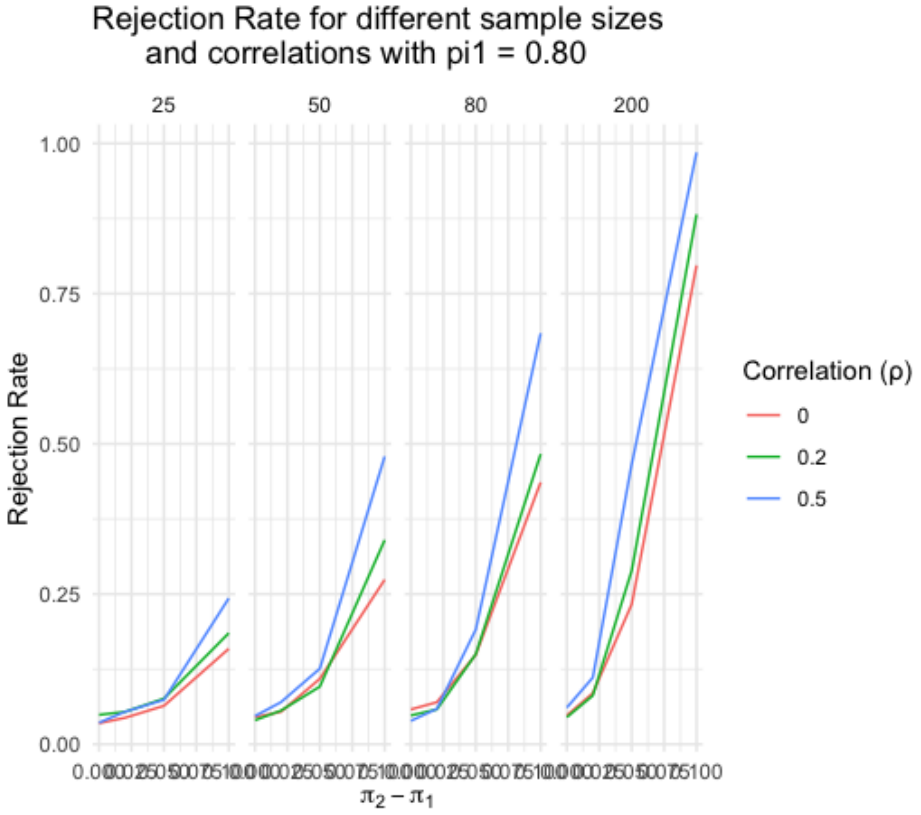

Figure 3: Rejection Rate for McNemar's Test with $\pi_1 = 0.8$.