Table of Contents

Introduction

The purpose of this report is to learn about, make derivations about, and use the McNemar Test for testing discordant pairs in a 2x2 two-way table. The McNemar Test is used in situations where there is matched pairs categorical data, and each category can be thought of as success and failure. For example, if a person takes two different drugs at two different times, but for the same reason, then a success might be that the drug worked, whereas a failure could be that a drug failed to work or made no improvement on the patient. The McNemar test can also be applied to see changes in population opinion over time, where one preference of a certain candidate (Candidate A) could be measured as a success and preference of the other candidate (Candidate B) could be measured as a failure.

In these situations, we are interested in the changes when the success or failure has changed from the first measurement to the second measurement. If the distribution of success and failure is for each measurement is homogeneous, then the number of failure-success and success-failure should be approximately equivalent. The McNemar Test is used to test the likelihood that any difference in failure-success and success-failure is due to random chance or if there is a significant difference between the distribution of successes and failures for each measurement.

Analysis of Sample Data

To study the McNemar test, we used a data from 250 subjects in which acid reflux is treated by two drugs: drug A and drug B. In this experiment, success was counted if the reflux stopped, while failure was counted if reflux was still present. Following good experimental design for a Matched Pairs study, half of the subjects were randomly selected to use drug A on the first day, while the other half used drug B. The next day, the subjects switched and used the other drug. Results are in the tables below.

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

Figure 1: Two-Way Table of Drug Type vs. Success/Failure

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

Figure 2: Table of Concordant and Discordant Pairs

We cannot perform a Chi-Squared Test for Homogeneity using Figure 1 to check to see if the distribution of success and failure is the same for both drugs because the observations are not independent of one another. We must use Figure 2 to perform our tests.

		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	$\pi_{\bullet 1}$	$\pi_{\bullet 2}$	$\pi_{\bullet\bullet}$

Figure 3: Table of Concordant and Discordant Pair Probabilities

Using Figure 3, we want to test two hypotheses:

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

H_A: Cell probabilities are 'free' (other than sum to 1 constraint)

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

Using the test statistic above, we found that $\chi^2 = 11.36364$, with a rejection region of $\{X^2 : X^2 > 3.841459\}$. Our value of Chi-Squared fell in our rejection region, so we will reject the null hypothesis in favor of the alternative. Furthermore, our p-value associated with this test statistic is $P(X^2 > 11.36364) = 0.0007$ which is well below any reasonable alpha. We found significant evidence that the Cell Probabilities are 'free' and that there is a relationship between drug and relief.

We then used the mcnemar.test() function in R to confirm these results. (See Appendix A for code on Part 1).

Derivations

First, we need to show that $\pi_{1*} = \pi_{*1}$ and $\pi_{2*} = \pi_{*2}$ equivalent to $\pi_{12} = \pi_{21}$, from Figure 3. (Note: I could not get the circles to work, so I used asterisks instead.)

$$\pi_{11} + \pi_{12} = \pi_{1*} \text{ and } \pi_{11} + \pi_{21} = \pi_{*1}$$

$$\pi_{1*} = \pi_{*1}, \text{ under } H_0$$

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

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

Then, we want to derive the maximums for $\pi_{11}, \pi_{12}, \pi_{21}, and\ \pi_{22}$, with the constraint that $(1 - \pi_{11} - 2a - \pi_{22} = 0)$. Since $\pi_{12} = \pi_{21}$ under the null, we will define

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

$$L(\pi_{11}, a, \pi_{22}) \propto (\pi_{11}^{n_{11}})(a^{n_{12}})(a^{n_{21}})(\pi_{22}^{n_{22}})$$

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

With the Lagrange Multiplier becomes,

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

Partial derivatives:

$$\frac{\partial l}{\partial \pi_{11}} = \frac{n_{11}}{\pi_{11}} - \lambda, \quad \frac{\partial l}{\partial \pi_{22}} = \frac{n_{22}}{\pi_{22}} - \lambda, \quad \frac{\partial l}{\partial a} = \frac{n_{12}+n_{21}}{a} - 2\lambda, \quad \frac{\partial l}{\partial \lambda} = 1 - \pi_{11} - 2a - \pi_{22}$$

Setting each equal to 0 yields:

$$\widetilde{\pi_{11}} = \frac{n_{11}}{\lambda}$$

$$\widetilde{n_{22}} = \frac{n_{22}}{\lambda}$$

$$\widetilde{\pi_{12}} = \widetilde{\pi_{21}} = \frac{n_{12} + n_{21}}{\lambda}$$

$$1 - \pi_{11} - 2a - \pi_{22} = 0$$

$$1 - \frac{n_{11}}{\lambda} - 2\frac{n_{12} + n_{21}}{2\lambda} - \frac{n_{22}}{\lambda} = 0$$

$$\lambda - n_{11} - n_{12} - n_{21} - n_{22} = 0$$

$$\lambda - n = 0$$

$$\lambda = n$$

Therefore:

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

$$\widetilde{n_{22}} = \frac{n_{22}}{n}$$

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

Next, we wanted to derive the form of the LRT.

$$\Lambda = \prod_{i=1}^{2} \prod_{j=2}^{2} \left(\frac{\widetilde{\pi_{ij}}}{\widehat{\pi_{ij}}} \right)^{n_{ij}}$$

$$-2\ln(\Lambda) = -2 \sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij} \ln\left(\frac{\widehat{\pi_{ij}}}{\widetilde{\pi_{ij}}} \right)$$

We know that $\widehat{\pi_{ij}} = \frac{n_{ij}}{n}$, n_{ij} is the observed counts, and $n\widetilde{\pi_{ij}}$ is the expected counts.

$$-2\ln(\Lambda) = -2 \sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij} \ln\left(\frac{n_{ij}}{n\widetilde{\pi_{ij}}} \right) = -2 \sum_{i=1}^{2} \sum_{j=1}^{2} Obs_{ij} \ln\left(\frac{Obs_{ij}}{Exp_{ij}} \right)$$

Under the null hypothesis, this would follow a Chi-Squared distribution with 1 degree of freedom. There are two free parameters in Ω and one free parameter in ω_0, so the degrees of freedom is 2 -1 = 1.

Lastly we want to show that Pearson's chi-square test statistic can be simplified into the test statistic for the McNemar Test.

$$X^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{\left(Obs_{ij} - Exp_{ij}\right)^2}{Exp_{ij}}$$

$$X^2 = \frac{(n_{11} - n_{11})^2}{n_{11}} + \frac{\left(n_{12} - \frac{n_{12} + n_{21}}{2}\right)^2}{\frac{n_{12} + n_{21}}{2}} + \frac{\left(n_{21} - \frac{n_{12} + n_{21}}{2}\right)^2}{\frac{n_{12} + n_{21}}{2}} + \frac{(n_{22} - n_{22})^2}{n_{22}}$$

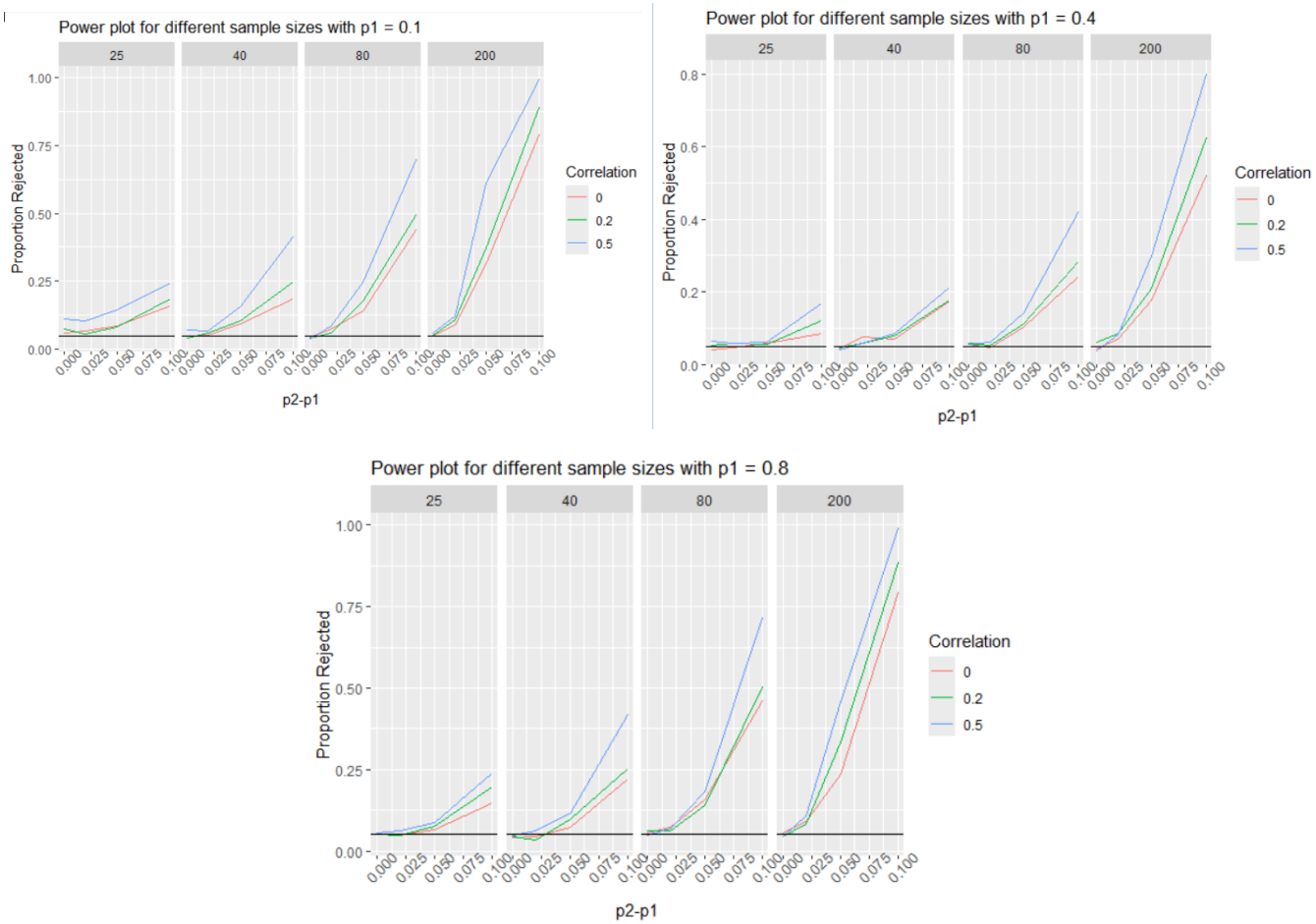
$$X^2 = \frac{n_{12}^2 - 2n_{21}n_{12} + n_{21}^2}{n_{12} + n_{21}} = \frac{(n_{12} - n_{21})^2}{n_{12} + n_{21}}$$

Simulations

In order to understand the power of the McNemar Test, we used R to draw many different samples of various sample size, correlation, Drug A success rate, and the difference between the success rate of Drug A and Drug B.

We were able to draw a binary sample (1 for success, 0 for failure), and determine the count of times that the both Drug A and Drug B were successful, when one of the Drugs was successful, and when neither of them were successful. These counts were stored in a Table of Concordant and Discordant counts. The McNemar Test was performed using the Table, and using the p-value from the test, we made a decision to either reject the null hypothesis or fail to reject it ($\alpha = 0.05$).

The proportion of the time the hypothesis was rejected is given in the graphs below.



In all cases of Drug A Success (p1), the power of the test increased as the sample size increased and the difference between Drug A and Drug B success increased, as expected. The first value of each graph at p2-p1 = 0 is the null hypothesis. Rejecting at p2-p1=0 would result in Type 1 Error. This should occur around 0.05, our chosen significance level. However, at low values of p1 and sample size, it appears that our test falsely rejected the null at a much higher rate than alpha = 0.05. The larger the correlation, the more likely we are to reject the null hypothesis.

Appendix – R Code

```
######################################################################
# Author: Evan Whitfield
# Last Editted: 12/9/24
# Purpose: Project 2 for ST502
######################################################################

#--Part 1--#

#Set alpha for RR
alpha <- 0.05

#Calc Test Statistic
test_stat <- (15-40)^2/(15+40)
print(test_stat)

#Calculate P Value
p_value <- pchisq(test_stat, df = 1, lower.tail = FALSE)
print(p_value)

#Determine Rejection Region for given alpha
reject_region <- qchisq(alpha, df = 1, lower.tail = FALSE)
print(reject_region)

#Making a matrix of our sample data values
raw_values <- c(85, 40, 15, 110)
data <- matrix(raw_values, nrow = 2, ncol = 2)

#Running the McNemar Test
mcnemar.test(data, correct = FALSE)

######################################################################
#--Part 2--#
library(MultiRNG)
library(ggplot2)

#Set values for N and alpha, and store values we want to change in vectors
N <- 1000
alpha <- 0.05
sample_sizes <- c(25, 40, 80, 200)
drug_A_success <- c(0.1, 0.4, 0.8)
drug_B_success <- c(0, 0.02, 0.05, 0.1)
corr <- c(0, 0.2, 0.5)

#Function that draws the sample data based on the current values we are interested in
draws <- function(n, drug_success, correlation){
    corr_matrix <- matrix(c(1,correlation,correlation,1), nrow = 2, ncol = 2)
        draw.correlated.binary(no.row = n, d = 2, prop.vec = c(drug_success[1], drug_success[2]), corr.mat =
corr_matrix)
}
```

```
#returns true if you should reject, false if you should not
trial <- function(data, alpha = 0.05){
  two_way_table = matrix(c(0,0,0,0), nrow = 2, ncol = 2)
  for(i in 1:nrow(data)){
    if(data[i,1] == 1 & data[i,2] == 1){
      two_way_table[1,1] <- two_way_table[1,1] + 1
    }
    else if(data[i,1] == 0 & data[i,2] == 1){
      two_way_table[1,2] <- two_way_table[1,2] + 1
    }
    else if(data[i,1] == 1 & data[i,2] == 0){
      two_way_table[2,1] <- two_way_table[2,1] + 1
    }
    else if(data[i,1] == 0 & data[i,2] == 0){
      two_way_table[2,2] <- two_way_table[2,2] + 1
    }
  }
  McNemar <- mcnemar.test(two_way_table, correct = FALSE)
  if(is.na(McNemar$p.value)){
    p_value <- 0
  }
  else {
    p_value <- McNemar$p.value
  }
  return(p_value < alpha)
}

#Create a Data Frame to store results from each simulation
results_df <- data.frame(Sample_Size = numeric(),
                         Drug_A_Success = numeric(),
                         DrugB_Minus_DrugA = numeric(),
                         Correlation = numeric(),
                         Proportion_Rejected = numeric()
                         )

#Run through all the values we are interested in and store results in a dataframe
for(m in 1:length(corr)){
  for(k in 1:length(drug_B_success)){
    for(j in 1:length(drug_A_success)){
      for(i in 1:length(sample_sizes)){
        results <- replicate(N, trial(draws(n = sample_sizes[i],
                                    drug_success = c(drug_A_success[j], drug_A_success[j]+drug_B_success[k]),
                                    correlation = corr[m]))
                             )
        new_row <- c(sample_sizes[i],
                     drug_A_success[j],
                     drug_B_success[k],
                     corr[m],
                     sum(results)/N
                     )
```

```
      results_df <- rbind(results_df, new_row)
    }
   }
  }
}

#I realized that the names of the columns were lost when adding rows.
#We used colnames in our last project, so I figured it would be okay to use here.
colnames(results_df) <- c("n","p1","p2_minus_p1","Corr","Prop")

#Create the appropriate plots. I had to use the ggplot2 website to understand different aspects of the graph.
#Website is in the bibliography.
#Creates graphs for p1 = 0.8
ggplot(subset(results_df,p1 == 0.8), aes(x = p2_minus_p1, y = Prop, color = as.factor(Corr))) +
  geom_line() +
  facet_wrap(~n, ncol = 4) +
  geom_hline(yintercept = alpha) +
  ggtitle("Power plot for different sample sizes with p1 = 0.8") +
  labs(color = "Correlation", x = "p2-p1",y = "Proportion Rejected") +
  theme(
    axis.text.x = element_text(angle = 45)
  )

#Create a new window for the next graph if running on a windows device.
windows()

#Creates 2nd group of graphs for p1 = 0.1
ggplot(subset(results_df,p1 == 0.1), aes(x = p2_minus_p1, y = Prop, color = as.factor(Corr))) +
  geom_line() +
  facet_wrap(~n, ncol = 4) +
  geom_hline(yintercept = alpha) +
  ggtitle("Power plot for different sample sizes with p1 = 0.1") +
  labs(color = "Correlation", x = "p2-p1",y = "Proportion Rejected") +
  theme(
    axis.text.x = element_text(angle = 45)
  )

#new window on windows device
windows()

#Creates 3rd group of graphs for p1 = 0.4
ggplot(subset(results_df,p1 == 0.4), aes(x = p2_minus_p1, y = Prop, color = as.factor(Corr))) +
  geom_line() +
  facet_wrap(~n, ncol = 4) +
  geom_hline(yintercept = alpha) +
  ggtitle("Power plot for different sample sizes with p1 = 0.4") +
  labs(color = "Correlation", x = "p2-p1",y = "Proportion Rejected") +
  theme(
    axis.text.x = element_text(angle = 45)
  )
```

Sources

https://ggplot2.tidyverse.org/ (visited on 12/8/2024)