Micah Fadrigo (ID # 42923836)

Professor Yu

Stats 240P

15 May 2023

Effect of Covariance-Variance Matrix Structures in Balanced & Unbalanced Designs

Abstract

This study aims to investigate the impact of positive dependence in variance-covariance matrices on the type I error rate and power of the four MANOVA tests: Pillai, Wilks, Hotelling-Lawley, and Roy. With increasing strengths of dependence in the variance-covariance matrix, we examined the trend of type I error rate and power across all tests. Then we assessed how these trends changed across increasing sample sizes. We considered a **balanced** design and an **unbalanced** design with various scenarios. In the balanced design, when sample size increased, type I error rate fluctuated and had no distinct pattern for any of the tests. The Roy test consistently had the highest type I error rate, which fluctuated around 0.175. Similar patterns were shown in the unbalanced design. In the balanced design, we saw a decrease in power as the strength of dependence increased (across all tests), which suggested that the tests had a harder time detecting group differences in the presence of strong variable dependence. However, as sample size increased, this trend became less prominent. In the unbalanced design, all three tests except Roy had nearly identical power trends as strength of dependence increased. Roy consistently had the highest power in both designs.

Introduction

In the context of multivariate analysis research, dependent variables are commonly chosen with the expectation that they are related to one another, often driven by common underlying factors. To adequately account for the relationship between variables, it's essential to consider the structure of the dependent variance-covariance matrix when measuring the precision and accuracy of statistical tests. Pillai's Trace, Wilks' Lambda, Hotelling-Lawley's Trace, and Roy's Largest Root rely on the relationship between the dependent variables to evaluate the significance of group differences in multivariate data.

In this study, we examine the trends of type I error rate and power across all of the aforementioned tests as a function of a fixed value '***c***' corresponding to the covariance terms (off-diagonal elements) of the variance-covariance matrix. We also assume a fixed variance of one for all variables (refer to **Figure 1** in Appendix). Furthermore, we studied if these trends (or the lack thereof) would hold or change across varying sample sizes. As we know, sample size has a direct effect on power such that the larger the sample size, the stronger ability the tests have to detect group differences under different covariance structures. Also, higher (lower) power typically leads to a lower (higher) type I error rate and so sample size also has an indirect effect on type I error rate.

To explore the effect of sample size, we considered both balanced and unbalanced designs. When testing power under the balanced design, we focused on a specific scenario where there are three groups, two of which have a mean of (0 0 0) and one has a different mean of (1 1 1). This effect size was sufficient so that the power was neither too high nor too low across tests, and allowed for noticeable changes in trends as sample size increased. When testing power under the unbalanced design, we introduced more variability by lowering the effect size - mean of (0.5 0.5 0.5) - for the differing group. If the mean was (1 1 1), the power would've been too high across all tests and all combinations of unbalanced groups, making it challenging to distinguish between the tests.

By exploring how the structure of a dependent variance-covariance matrix affects the type I error rate and power of statistical tests, under varying conditions of sample size, we gain insight into the performance of the selected MANOVA tests.

Simulation Methods

When testing type I error rate, all group means had to be the same and were all set to (0 0 0). When testing power, only one group had a differing mean (not (0 0 0)). In the balanced design, the differing mean was set to (1 1 1) and in the unbalanced design the differing mean was set to (0.5 0.5 0.5). These were chosen for reasons explained in the Introduction section.

For the **balanced** design, we first iterated through the sample sizes (10, 15, 20, 30) and for each sample size, we iterated through each of the variance-covariance structures which differed by the ***'c'*** parameter. For each pair of sample size and Sigma, we performed 300 simulations, where each simulation generated data for the three groups (according to the current sample size) and calculated the type I error rate or power. Finally, the overall type I error rate or power was calculated by taking the average over the 300 simulations.

For the **unbalanced** design, we followed a similar process, except we iterated through different scenarios of unbalanced groups. The various scenarios were: **group three** having a much **smaller** sample size than the others (1000, 1000, 10), **group three** having a much **larger** sample size than the others (10, 10, 1000), **group two and group three** having a much **smaller** sample size than the remaining group (1000, 10, 10), and **group two and group three** having a much **larger** sample size than the remaining group (10, 1000, 1000).

Balanced Design: Type I Error

# Groups (G)	Sample Size (N_1 = N_2 = N_3)	Group 1 Mean Vector (mu_1)	Group 2 Mean Vector (mu_2)	Group 2 Mean Vector (mu_3)	# Simulations (B)	Sigma (SEE FIG. 1 IN APPENDIX)
3	10	(0 0 0)	(0 0 0)	(0 0 0)	300	Fig. 1
3	15	(0 0 0)	(0 0 0)	(0 0 0)	300	Fig. 1
3	20	(0 0 0)	(0 0 0)	(0 0 0)	300	Fig. 1
3	30	(0 0 0)	(0 0 0)	(0 0 0)	300	Fig. 1

Balanced Design: Power

# Groups (G)	Sample Size (N_1 = N_2 = N_3)	Group 1 Mean Vector (mu_1)	Group 2 Mean Vector (mu_2)	Group 2 Mean Vector (mu_3)	# Simulations (B)	Sigma (SEE FIG. 1 IN APPENDIX)
3	10	(0 0 0)	(0 0 0)	(1, 1, 1)	300	Fig. 1
3	15	(0 0 0)	(0 0 0)	(1, 1, 1)	300	Fig. 1
3	20	(0 0 0)	(0 0 0)	(1, 1, 1)	300	Fig. 1
3	30	(0 0 0)	(0 0 0)	(1, 1, 1)	300	Fig. 1

Unbalanced Design: Type I Error

# Groups (G)	Sample Size (N_1, N_2 , N_3 respectively)	Group 1 Mean Vector (mu_1)	Group 2 Mean Vector (mu_2)	Group 2 Mean Vector (mu_3)	# Simulations (B)	Sigma (SEE FIG. 1 IN APPENDIX)
3	1000, 1000, 10	(0 0 0)	(0 0 0)	(0 0 0)	300	Fig. 1
3	1000, 10, 10	(0 0 0)	(0 0 0)	(0 0 0)	300	Fig. 1
3	10, 10, 1000	(0 0 0)	(0 0 0)	(0 0 0)	300	Fig. 1

3	10, 1000, 1000	(0 0 0)	(0 0 0)	(0 0 0)	300	Fig. 1

Unbalanced Design: Power

# Groups (G)	Sample Size (N_1, N_2 , N_3 respectively)	Group 1 Mean Vector (mu_1)	Group 2 Mean Vector (mu_2)	Group 2 Mean Vector (mu_3)	# Simulations (B)	Sigma (SEE FIG. 1 IN APPENDIX)
3	1000, 1000, 10	(0 0 0)	(0 0 0)	(0.5 0.5 0.5)	300	Fig. 1
3	1000, 10, 10	(0 0 0)	(0 0 0)	(0.5 0.5 0.5)	300	Fig. 1
3	10, 10, 1000	(0 0 0)	(0 0 0)	(0.5 0.5 0.5)	300	Fig. 1
3	10, 1000, 1000	(0 0 0)	(0 0 0)	(0.5 0.5 0.5)	300	Fig. 1

Simulation Results

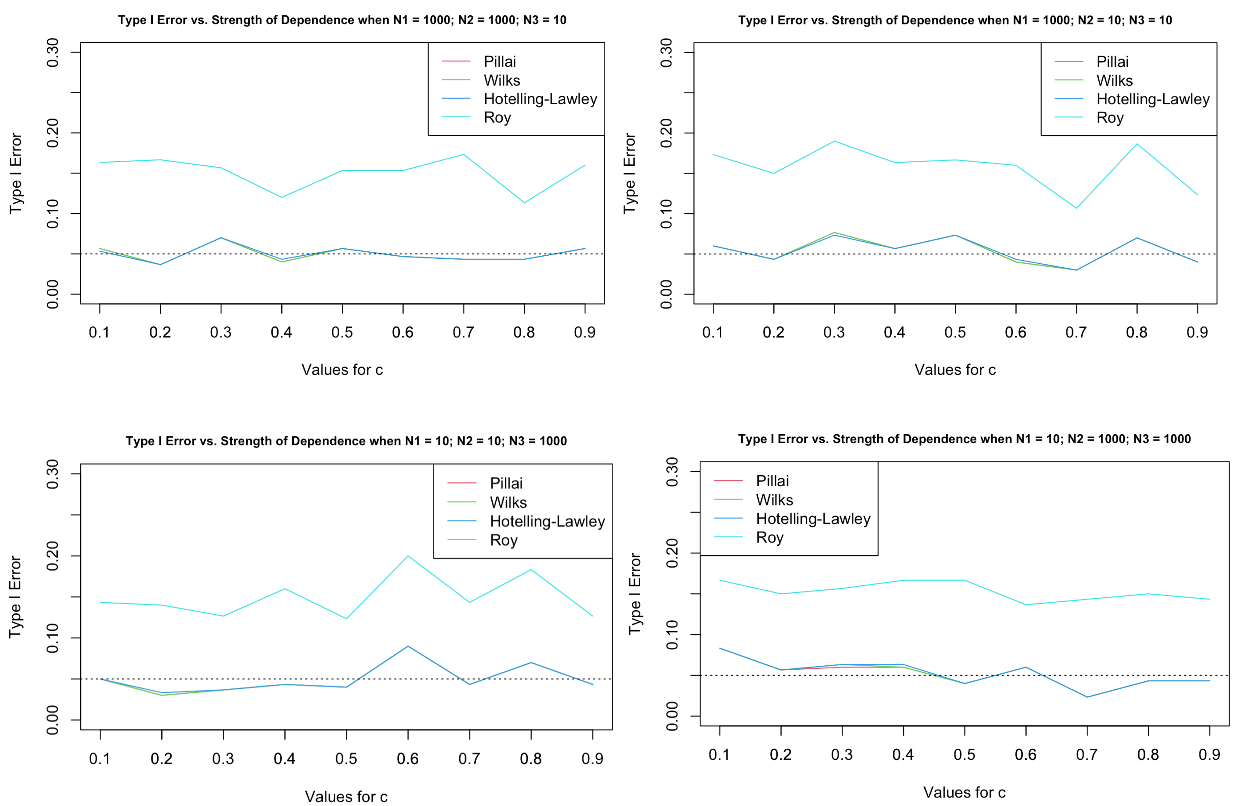
Type I Error: Balanced Design

Note: Plots for N = 15 and N = 20 were omitted because they show similar results.

When the sample size is small, there is no distinct pattern between strength of dependence and type I error, in which all tests fluctuate. Pillai, Wilks, and Hotelling-Lawley fluctuate around 0.05

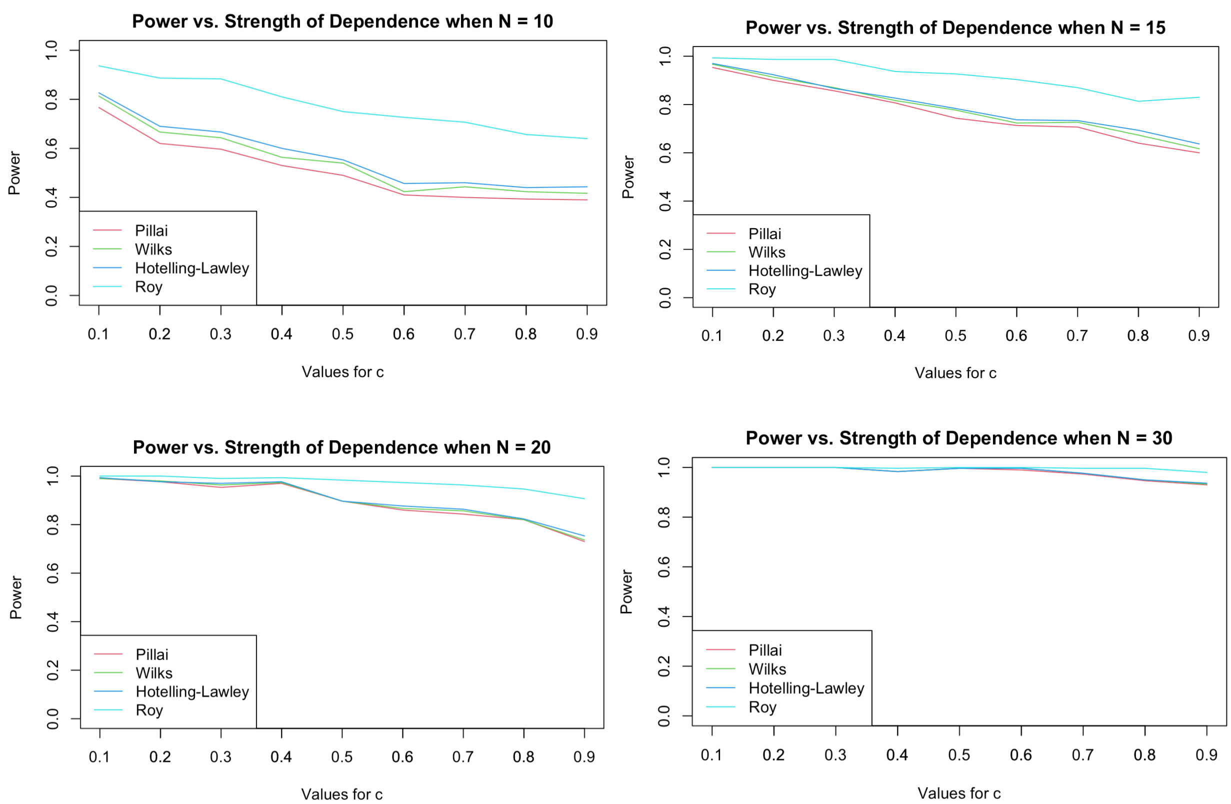
as the parameter **'*c*'** increases. This fluctuation stabilizes below 0.05 as sample size increases and their trends of type I error rates become nearly identical. Roy fluctuates around 0.175 as the parameter **'*c*'** increases. Although this fluctuation does not stabilize as sample size increases, the overall type I error rate decreases (fluctuates around 0.15).

Type I Error: Unbalanced Design



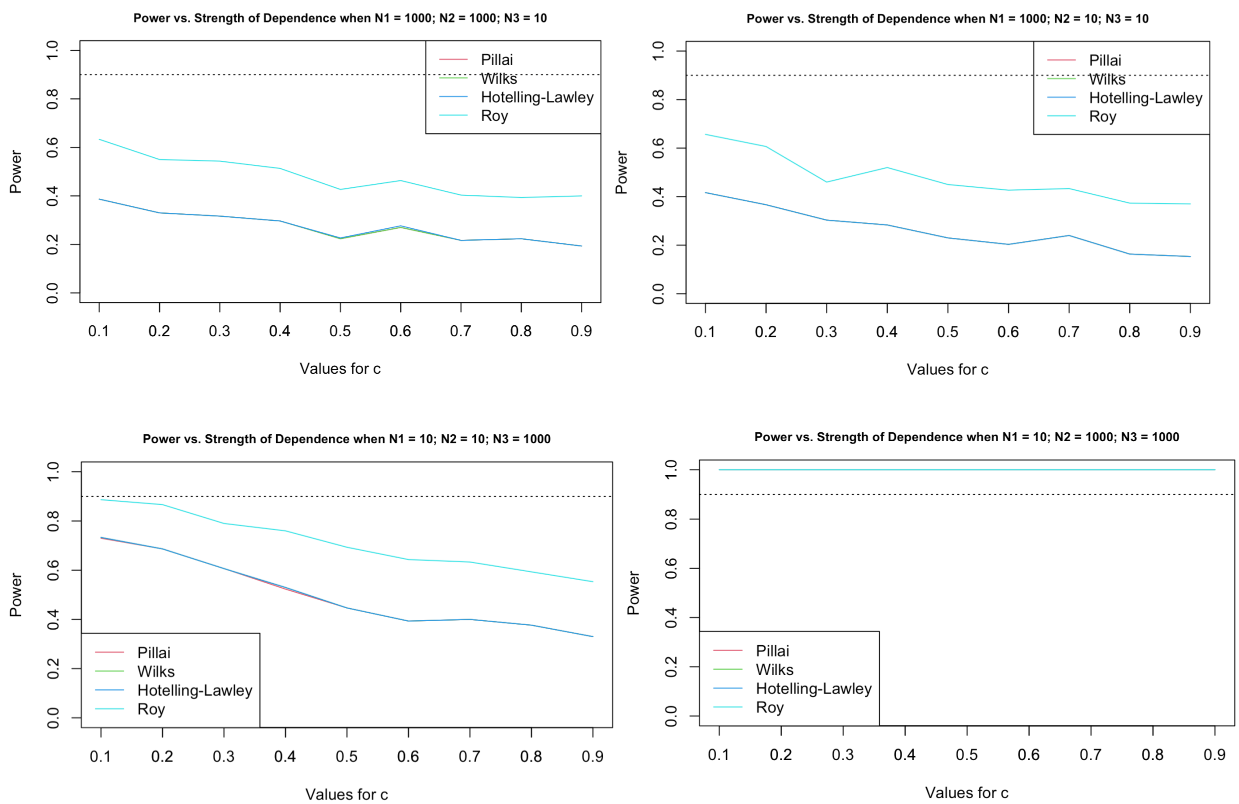
As for the unbalanced design, all tests still tend to fluctuate in type I error rate (around the same values mentioned in the balanced design) and there is no evident relationship between the parameter **'*c*'** and type I error rate. The Roy test remains to have the highest type I error rate compared to the other tests, across all values of c and all sample size combinations. Pillai, Wilks, and Hotelling-Lawley have nearly identical trends of type I error rate in each plot of a different sample size. Trends of type I error rate seem to stabilize across all tests when there are a large number of observations for each group in a pair of groups that truly differ. For example, in the last plot, although groups two and three differ, they each have a large number of observations, which allows for more evidence to distinguish between the groups. As a result of having more evidence, the tests become more reliable in detecting true group differences (higher power) and type I error rate tends to stabilize. This suggests that even with variations in the dependence within the variance-covariance matrix, the tests can still identify true group differences if the sample size is large enough.

Power: Balanced Design



A decrease in power across all tests as the strength of dependence increases, suggests that the tests have a harder time detecting true group differences when variables are highly correlated, and therefore the tests become less powerful. However, we see that when the sample size is sufficiently large enough, the trend of decreasing power (with increasing strength of dependence in the variance-covariance matrix) becomes less pronounced. For sample sizes of 10, 15, and 20, the Roy test had the highest power across all values of **‘c’**. This could be because the Roy test also had the highest type I error rate. For a sample size of 30, all tests had a power near one across all values of **‘c’**.

Power: Unbalanced Design



As for the unbalanced design, all tests still tend to exhibit a negative relationship between the strength of the dependence of the covariance term and power, as seen in the unbalanced design. The Roy test remains to have the highest power compared to the other tests, across all values of **‘c’**. Pillai, Wilks, and Hotelling-Lawley have nearly identical trends of power in each plot of a different sample size. In the last plot, although groups two and three differ, they each have a large number of observations, which allows for more evidence to distinguish between the groups.

As a result of having more evidence, the tests become more reliable in detecting true group differences, which is why they all have a power of 1 (regardless of the dependence structure in the variance-covariance matrix). This is consistent with the findings for type I error rate in the unbalanced design.

Discussion

Both the dependence structure and sample size impact the type I error rate and power of the four MANOVA tests. For the balanced and unbalanced designs, when strength of the dependence increased, type I error rate showed no distinct pattern, and power tended to decrease (across all tests). However, when sample size increased (for all groups or majority of the groups), overall type I error rate decreased and overall power increased. The Roy test consistently had the highest power and type I error rate among all the tests. Pillai, Wilks, and Hotelling-Lawley tests performed similarly on both measures in both the balanced and unbalanced designs.

One limitation is that the same variance-covariance matrix was used across all groups in a single simulation. The results could be different from mine if there was slight variation in the matrices between groups (while still satisfying the MANOVA assumption that the matrices are sufficiently similar). The variance of the variables was also assumed to be the same, in which these values could've differed, potentially yielding different results. The trends of type I error rate and power could be studied furthermore by using a wider range for the parameter **'c'** or the sample size. Another limitation is that only group three had a different mean. Having all three groups differ in mean could yield different results.

Appendix

$$\Sigma = \begin{pmatrix} 1 & c & c \\ c & 1 & c \\ c & c & 1 \end{pmatrix}$$

Figure 1: Power & type I error rate as a function of ***c*** *s.t.* ***c*** *= {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}*

```
generate_data <- function(n, G, mu2, mu1 = c(0,0,0),
                          Sigma){
  Y <- c()

  for (g in 1:ceiling(G/2)){
    Y <- rbind(Y, mvrnorm(n, mu1, Sigma))
    }
  for (g in 1:floor(G/2)){
    Y <- rbind(Y, mvrnorm(n, mu2, Sigma))
    }
  Y
}
```

```
generate_data_unbalanced_small <- function(n, G, mu2, mu1 = c(0,0,0),
                                           Sigma){
  Y <- c()

  for (g in 1:ceiling(G/2)){
    Y <- rbind(Y, mvrnorm(n, mu1, Sigma)) # Group 1 & 2: 1000
    }
  for (g in 1:floor(G/2)){
    Y <- rbind(Y, mvrnorm(n/100, mu2, Sigma)) # Group 3: 10
    }
  Y
}
```

```
generate_data_unbalanced_large <- function(n, G, mu2, mu1 = c(0,0,0),
                                           Sigma){
  Y <- c()

  for (g in 1:ceiling(G/2)){
    Y <- rbind(Y, mvrnorm(n, mu1, Sigma)) # Group 1 & 2: 10
    }
  for (g in 1:floor(G/2)){
    Y <- rbind(Y, mvrnorm(n*100, mu2, Sigma)) # Group 3: 1000
    }
  Y
}
```

```
generate_data_unbalanced_small_two <- function(n, G, mu2, mu1 = c(0,0,0),
                                               Sigma){
  Y <- c()

  Y <- rbind(Y, mvrnorm(n, mu1, Sigma)) # Group 1: 1000
  Y <- rbind(Y, mvrnorm(n/100, mu1, Sigma)) # Group 2: 10
  Y <- rbind(Y, mvrnorm(n/100, mu2, Sigma)) # Group 3: 10

  Y
}
```

```
generate_data_unbalanced_large_two <- function(n, G, mu2, mu1 = c(0,0,0),
                                               Sigma){
  Y <- c()

  Y <- rbind(Y, mvrnorm(n, mu1, Sigma)) # Group 1: 10
  Y <- rbind(Y, mvrnorm(n*100, mu1, Sigma)) # Group 2: 1000
  Y <- rbind(Y, mvrnorm(n*100, mu2, Sigma)) # Group 3: 1000

  Y
}
```

```
test <- function(n, G, Y){
  groups <- rep(c(paste("Group", 1:G)), each=n)
  obj <- manova(Y ~ groups)
  tests <- c("Pillai", "Wilks", "Hotelling-Lawley", "Roy")
  reject <- rep(0, 4)
  for (t in 1:length(tests)){
    reject[t] <- summary(obj,
                         test = tests[t])$stats[1,6]<0.05
  }
  reject # return vector size of 4, binary: no reject/reject
}
```

```
simulate <- function(B, n, G, mu2, Sigma){
  results <- rep(0, 4)
  for (b in 1:B){
    Y <- generate_data(n, G, mu2, Sigma = Sigma)
    results <- results + test(n, G, Y)
    }
  results/B # returns proportion of rejections/#simulations per test
}
```

```
# Repeat Chunk for sample size: {10, 15, 20, 30}
n <- 10 # Change to test each sample size
mu2 <- c(0,0,0) # Tests type I error rate
# mu2 <- c(1,1,1) # Tests power (example)
S <- seq(0.1, 0.9, 0.1) # Values of 'c'
results <- list() # Store results of each simulation
alpha_matrix <- matrix(0, nrow = 4, ncol = length(S))

for (i in 1:length(S)) {
  tests <- c("Pillai", "Wilks", "Hotelling-Lawley", "Roy")
  alpha <- simulate(B = 300,
                    n = n,
                    G = 3,
                    mu2 = mu2,
                    Sigma = matrix(c(1, S[i], S[i],
                                     S[i], 1, S[i],
                                     S[i], S[i], 1), nrow = 3, byrow =
TRUE))
  alpha_matrix[,i] <- alpha
  names(alpha) <- tests
}
```

I made several functions that extended the basic ***generate_data*** *function to account for the various scenarios of unbalanced design. The* ***test*** *function returns the results of the four different tests in a specific scenario. The* ***simulate*** *function simulates a scenario a specified number of times and utilizes one of the data generation functions and the test function. A different* ***simulate*** *function was created for each of the data generation functions. I utilized a for-loop to consider the different variance-covariance matrices.*