EXAMINING TRACY-WIDOM DISTRIBUTION AS AN APPROXIMATION FOR ROY'S LARGEST ROOT IN MULTIVARIATE ANALYSIS OF VARIANCE (MANOVA)

AGNES FIORELA CHRISNANDA[1], BRYAN FLORENTINO LEO[2],
DAUD MUHAMAD AZHARI[3], RADEN AJENG CALLYSTA SEVANDA CALYACETTA[4*]

Abstract. Multivariate analysis of variance (MANOVA) is a well-known statistical technique in numerous fields of science. Despite its popularity, drawbacks sourced from its assumptions and test statistics were discovered and proven in past researches. This article introduces Tracy-Widom distribution as a comparing approximation against the classical F-statistic to Roy's largest root, one of the existing test statistics in MANOVA. The quality of approximations is examined from datasets of two distinct construction methods, namely simulation and real observation.

1. Introduction

Multivariate analysis of variance (MANOVA) is a parametric statistical method utilized to analyze whether there exists a significant difference between vectors of means from groups of dependent variables. It is still a well-known technique in numerous fields of science, such as agriculture, economics, psychology, education, and so on. This happens due to the involvement of designs in experimental researches, specifically randomized block designs.

Halfway to two decades ago, Johnstone (2008) [4] argued how flaws had surrounded the popularity of MANOVA. To begin with, the procedures behind MANOVA is preceded by tests of assumptions, namely the fitting into multivariate normal distribution, homogeneity of variance-covariance matrices, and independence. This is disadvantageous since the majority of datasets available in the present world does not fulfil these assumptions. The remaining drawbacks came from the test statistics. Roy's largest root, one of the discovered, has a null hypergeometric distribution without a general and closed form. Moreover, it depends on three parameters that its table of critical values reaches up to 25 pages. While statistical softwares are developed through the growth of technology, the computation of F-statistic as an approximate test statistic had been found to be inaccurate for datasets with larger than two groups. To summarize, MANOVA would be highly complex and inappropriate concerning the presence of big data in the recent years.

Alongside these weaknesses, Johnstone (2008, 2009) [4, 5] suggested Tracy-Widom distribution as a limiting approximation in multivariate analyses involving Jacobi ensembles and largest root. This finding becomes the underlying motivation of this article to compare the quality of Tracy-Widom distribution against the classical F-statistic as approximations to Roy's largest root statistic in MANOVA.

2. Theoretical Framework

2.1. Multivariate Analysis of Variance (MANOVA). The multivariate analysis of variance (MANOVA) is the multivariate analog of the analysis of variance (ANOVA) procedure used for univariate data. We often measure several dependent variables on each experimental unit instead of just one variable. In the multivariate case, we assume that k independent random samples of size n are obtained from p-variate normal populations with equal covariance matrices. In practice, the observation vectors y_{ij} (observation from subject j in group i) would ordinarily be listed in row form, and sample 2 would appear below sample 1, and so on. This is illustrated in Table 1 below.

Table 1. The general form of randomized block design for MANOVA.

<table>
<tr><th colspan="6">Treatment</th></tr>
<tr><th colspan="2"></th><th>1</th><th>2</th><th>$\ldots$</th><th>g</th></tr>
<tr><th rowspan="4">Subject</th><th>1</th><td>$\mathbf{Y}_{11} = \begin{pmatrix} Y_{111} \\ Y_{112} \\ \vdots \\ Y_{11p} \end{pmatrix}$</td><td>$\mathbf{Y}_{21} = \begin{pmatrix} Y_{211} \\ Y_{212} \\ \vdots \\ Y_{21p} \end{pmatrix}$</td><td>$\ldots$</td><td>$\mathbf{Y}_{g1} = \begin{pmatrix} Y_{g11} \\ Y_{g12} \\ \vdots \\ Y_{g1p} \end{pmatrix}$</td></tr>
<tr><th>2</th><td>$\mathbf{Y}_{12} = \begin{pmatrix} Y_{121} \\ Y_{122} \\ \vdots \\ Y_{12p} \end{pmatrix}$</td><td>$\mathbf{Y}_{22} = \begin{pmatrix} Y_{221} \\ Y_{222} \\ \vdots \\ Y_{22p} \end{pmatrix}$</td><td>$\ldots$</td><td>$\mathbf{Y}_{g2} = \begin{pmatrix} Y_{g21} \\ Y_{g22} \\ \vdots \\ Y_{g2p} \end{pmatrix}$</td></tr>
<tr><td>$\vdots$</td><td>$\vdots$</td><td>$\vdots$</td><td>$\ddots$</td><td>$\vdots$</td></tr>
<tr><td>n_i</td><td>$\mathbf{Y}_{1n_1} = \begin{pmatrix} Y_{1n_11} \\ Y_{1n_12} \\ \vdots \\ Y_{1n_1p} \end{pmatrix}$</td><td>$\mathbf{Y}_{2n_2} = \begin{pmatrix} Y_{2n_21} \\ Y_{2n_22} \\ \vdots \\ Y_{2n_2p} \end{pmatrix}$</td><td>$\ldots$</td><td>$\mathbf{Y}_{gn_g} = \begin{pmatrix} Y_{gn_g1} \\ Y_{gn_g2} \\ \vdots \\ Y_{gn_gp} \end{pmatrix}$</td></tr>
</table>

The model for each observation vector is

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, 2, \ldots, k, j = 1, 2, \ldots, n.$$

In terms of the p variables in y_{ij}, the model for the r^{th} variable $(r = 1, 2, \ldots, p)$ in each vector y_{ij} is

$$y_{ijr} = \mu_r + \alpha_{ir} + \epsilon_{ijr} = \mu_{ir} + \epsilon_{ijr}, \quad i = 1, 2, \ldots, k, j = 1, 2, \ldots, n.$$

There are four assumptions before undergoing the hypothesis testing of MANOVA, which are:

(1) The data from group i has common mean vector $\boldsymbol{\mu}_i = \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{ip} \end{pmatrix}$.
(2) The data from all groups have a common variance-covariance matrix, $\boldsymbol{\Sigma}$.
(3) The subjects are independently sampled.
(4) The data are multivariate normally distributed.

The goal of MANOVA is to compare the mean vectors of k samples to find significant differences. The hypotheses are

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_k$$

(The k mean vectors are equal for each variable. That is, $\boldsymbol{\mu}_{1r} = \boldsymbol{\mu}_{2r} = ... = \boldsymbol{\mu}_{kr}$.)

vs

H_1: At least two $\boldsymbol{\mu}$'s are unequal.

In the multivariate case, we have "between" and "within" matrices $\mathbf{H}$ and $\mathbf{E}$, defined as

$$\mathbf{H} = n \sum_{i=1}^{k} (\bar{y_{i.}} - \bar{y_{..}})(\bar{y_{i.}} - \bar{y_{..}})'$$

and

$$\mathbf{E} = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y_{i.}})(y_{ij} - \bar{y_{i.}})'.$$

There are at least four discovered test statistics for MANOVA given as follows.

(1) **Wilks' Test Statistic**
The likelihood ratio test of $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \ldots = \boldsymbol{\mu}_k$ is given by

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|}.$$

H_0 is rejected if $\Lambda \leq \Lambda_{\alpha,p,v_H,v_E}$. Exact critical values $\Lambda_{\alpha,p,v_H,v_E}$ are found in Table A.9 in Rencher (2002) [10], where:
p = number of variables (dimension),
v_H = degrees of freedom for hypothesis,
v_E = degrees of freedom for error.

(2) **Roy's Largest Root**
The test statistic is given by

$$\theta = \frac{\lambda_1}{1 + \lambda_i},$$

Where λ_1 is the largest eigenvalue of $\mathbf{E}^{-1}\mathbf{H}$. We reject H_0 for $\theta \geq \theta_{\alpha,s,m,N}$, where
$s = \min(v_H, p)$,

TABLE 2. Transformations of Wilks' Λ to Exact Upper Tail F-Tests.

Parameters (p, v_H)	Statistic Having F-Distribution	Degrees of Freedom
Any p, $v_H = 1$	$\frac{1-\Lambda}{\Lambda}\frac{v_E - p + 1}{p}$	$p, v_E - p + 1$
Any p, $v_H = 2$	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\frac{v_E - p + 1}{p}$	$2p, 2(v_E - p + 1)$
$p = 1$, any v_H	$\frac{1-\Lambda}{\Lambda}\frac{v_E}{v_H}$	v_H, v_E
$p = 2$, any v_H	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\frac{v_E - 1}{v_H}$	$2v_H, 2(v_E - 1)$

$m = \frac{1}{2}(|v_H - p| - 1)$,
$N = \frac{1}{2}(v_E - p - 1)$.
We do not have a satisfactory F-approximation for θ or λ_1, but upper bound on F which is provided in some software programs is given by

$$F = \frac{(v_E - d - 1)\lambda_1}{d}$$

with degrees of freedom d and $v_E - d - 1$, where $d = \max(p, v_H)$. The term upper bound indicates that the F is greater than the "true F"; that is, $F > F_{d, v_E - d - 1}$. Therefore, we feel safe if H_0 is accepted.

(3) **Pillai Test**
The Pillai statistic is given by

$$V = tr[(\mathbf{E} + \mathbf{H})^{-1}\mathbf{H}] = \sum_{i=1}^{s} \frac{\lambda_i}{1 + \lambda_i}.$$

We reject H_0 for $V \geq V_\alpha$. The upper percentage points, V_α, are given in Table A.11 in Rencher (2002) [10], indexed by s, m, and N, where
$s = \min(v_H, p)$,
$m = \frac{1}{2}(|v_H - p| - 1)$,
$N = \frac{1}{2}(v_E - p - 1)$.

(4) **Lawley–Hotelling Test**
The Lawley–Hotelling statistic (Lawley, 1938, and Hotelling, 1951) is defined as

$$U = tr(\mathbf{E}^{-1}\mathbf{H}) = \sum_{i=1}^{s} \lambda_i.$$

We reject H_0 for large values of the test statistic.

2.2. **Wishart Distribution.** According to Mardia, Kent, and Bibby (1979) [6], the Wishart distribution is a distribution related to multivariate normal or multinormal distribution. A matrix $\mathbf{M}$ following a Wishart distribution is denoted by

$\mathbf{M} \sim W_p(\boldsymbol{\Sigma}, m)$, where for $\boldsymbol{\Sigma} > 0$ and $m \geq p$, its probability density function (pdf) is given by

$$f(\mathbf{M}) = \frac{|\mathbf{M}|^{\frac{m-p-1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\mathbf{M}\right)\right)}{2^{\frac{mp}{2}} \pi^{\frac{p(p-1)}{4}} |\boldsymbol{\Sigma}|^{\frac{m}{2}} \prod_{i=1}^{p} \Gamma\left(\frac{1}{2}(m+1-i)\right)},$$

restricted to $\mathbf{M} > 0$.

2.3. **Double Wishart matrices.** The definition of double Wishart matrices begins with understanding Hermitian matrices as follows.

Definition 2.1. A square matrix $\mathbf{A} = [a_{ij}]$ of the dimension $n \times n$ is called a Hermitian matrix if and only if the conjugate transpose of $\mathbf{A}$ is equal to itself (denoted as $\mathbf{A}^* = \mathbf{A}$), that is, for every $a_{ij} \in \mathbf{A}$, $\bar{a}_{ij} = a_{ij}$, $1 \leq i, j \leq n$. Additionally, if $\mathbf{A}^* = -\mathbf{A}$, a square matrix.

Paul and Aue [8] defined double Wishart problems as follows.

Definition 2.2. Double Wishart problems are problems in multivariate analysis which could be addressed through the analysis of two random and independent Hermitian matrices of the same dimension.

Some common double Wishart problems are multivariate analysis of variance (MANOVA), canonical correlation analysis (CCA), tests for equality of covariance matrices, and tests for linear hypotheses in multivariate linear regression problems.

The two Hermitian matrices involve in double Wishart problems are referred to as double Wishart matrices. Beforehand, the definition of a Wishart matrix is given below.

Definition 2.3. Let $\mathbf{X} = [x_{ij}]$ be a matrix of size $p \times n$ such that the columns, $X_j = [X_{ij}]_{i=1}^{p}$, are i.i.d $N_p(0, \boldsymbol{\Sigma})$. The sample covariance matrix is a matrix of size $p \times p$ given by

$$\mathbf{S} = n^{-1}\mathbf{X}\mathbf{X}^T. \tag{2.1}$$

The unnormalized version,

$$n\mathbf{S} = \mathbf{X}\mathbf{X}^T, \tag{2.2}$$

is called a Wishart matrix, for $\mathbf{X}\mathbf{X}^T$ follows a Wishart distribution, $W_p(n, \boldsymbol{\Sigma})$.

Definition 2.4. Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be two independent matrices of sizes $p \times n_1$ and $p \times n_2$, respectively, with entries having means of 0 and finite variances. A double Wishart matrix ensemble is given by the general form

$$\mathbf{X}_1\mathbf{X}_1^*(\mathbf{X}_1\mathbf{X}_1^* + \mathbf{X}_2\mathbf{X}_2^*)^{-1}. \tag{2.3}$$

If the entries of the data matrix are i.i.d real or complex Gaussian, double Wishart matrices are referred to as Jacobi Orthogonal Ensemble (JOE) and Jacobi Unitary Ensemble (JUE), respectively.

2.4. **Jacobi Orthogonal Polynomials.** The origin of involving the word "Jacobi" in naming the special cases of random double Wishart matrix ensemble is due to Jacobi orthogonal polynomials.

Definition 2.5. A set of orthogonal polynomials is an infinite sequence of polynomials, $p_0(x), p_1(x), p_2(x), \ldots$, where $p_n(x)$ has a degree of n and any two polynomials in the set are orthogonal to each other, that is, for $i \neq j$,

$$\int_a^b p_i(x)p_j(x)\,dx = 0. \tag{2.4}$$

The interval $[a, b]$ is called the interval of orthogonality and may be infinite at either or both ends.

All sets of orthogonal polynomials have a number of properties.

(1) Any polynomial $f(x)$ of degree n can be expanded in terms of $p_0, p_1, \ldots, p_n$, that is, there exist coefficients a_i such that

$$f(x) = \sum_{i=0}^{n} a_i p_i(x).$$

(2) Each polynomial $p_k(x)$ in the set is orthogonal to any polynomial of degree less than k.
(3) Any set has a recurrence that relates any three consecutive polynomials in the sequence defined by

$$p_{n+1} = (a_n x + b_n)p_n - c_n p_{n-1},$$

where a, b, and c depend on n.
(4) Each polynomial in the set has all n of its roots real, distinct, and strictly within the interval of orthogonality.
(5) The roots of p_n lie strictly inside the roots of p_{n+1}.

Referring to Paul and Aue [8], the joint distributions of eigenvalues of Wigner, Wishart, and double Wishart random matrix models are given by the closed general form:

$$f_{w,\beta}(x_1, \ldots, x_N) = c_{w,\beta} \prod_{j=1}^{N} w(x_j)^{\beta/2} \prod_{1 \leq j < k \leq N} |x_j - x_k|^{\beta}, \tag{2.5}$$

where w is a nonnegative weight function on $\mathbb{R}$, $\beta = 1$ for real entries and $\beta = 2$ for complex entries, and $c_{w,\beta}$ is a normalizing constant. For double Wishart matrix ensembles, w determines Jacobi orthogonal polynomial family given by

$$w(x) = (1 - x)^a (1 + x)^b, \quad a, b > 0 \tag{2.6}$$

with a domain of $(0, 1)$.

2.5. **Tracy-Widom Distribution.** Recall the definition of convergence in distribution as given by Bain and Engelhardt (1992) [1] below.

Definition 2.6. Let $Y_1, Y_2, \dots$ be a sequence of random variables with a corresponding sequence of CDF's $G_1(y), G_2(y), \dots$ such that for each $n = 1, 2, \dots$,

$$G_n(y) = P[Y_n \leq y].$$

The sequence is said to converge in distribution to $Y \sim G(y)$, denoted $Y_n \xrightarrow{d} Y$, if $Y_n \sim G_n(y)$ for each $n = 1, 2, \dots$ and for some CDF $G(y)$,

$$\lim_{n \to \infty} G_n(y) = G(y). \tag{2.7}$$

The distribution corresponding to the CDF $G(y)$ is called the limiting distribution of Y_n.

Johnstone (2001) [3] defined the Tracy-Widom law for real entries as follows.

Theorem 2.7. *Let $\mathbf{X} = [x_{ij}]$ be a matrix of dimension $p \times n$ with real and i.i.d $N(0,1)$ entries. Also, let $l_1 > \ldots > l_p$ be the sample eigenvalues of the corresponding Wishart matrix $\mathbf{X}^T\mathbf{X}$. Define the center and scaling constants as follows:*

$$\mu_{n,p} = \left(\sqrt{n-1} + \sqrt{p}\right)^2, \tag{2.8}$$

$$\sigma_{n,p} = \left(\sqrt{n-1} + \sqrt{p}\right)\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{\frac{1}{3}}. \tag{2.9}$$

The Tracy-Widom law of order 1 has a cumulative distribution function given by

$$F_1(s) = \exp\left(-\frac{1}{2}\int_s^\infty q(x) + (x-s)q^2(x)\,dx\right), \qquad s \in \mathbb{R}, \tag{2.10}$$

where $q(x)$ is the solution to the Painlevé II differential equation,

$$q''(x) = xq(x) + 2q^3(x). \tag{2.11}$$

Under the assumption that $n = n(p) \to \infty$ such that $n/p \to \gamma \geq 1$,

$$\frac{l_1 - \mu_{n,p}}{\sigma_{n,p}} \xrightarrow{d} W_1 \sim F_1. \tag{2.12}$$

On the other hand, Johansson (2000) [2] discovered the analogous theorem for complex entries by doing a slight modification given as follows.

Theorem 2.8. *Let $\mathbf{X} = [x_{ij}]$ be a matrix of dimension $p \times n$ with complex and i.i.d $N(0,1)$ entries. Also, let $l_1 > \ldots > l_p$ be the sample eigenvalues of the corresponding Wishart matrix $\mathbf{X}^T\mathbf{X}$. Define the center and scaling constants as follows:*

$$\mu_{n,p} = \left(\sqrt{n} + \sqrt{p}\right)^2, \tag{2.13}$$

$$\sigma_{n,p} = \left(\sqrt{n} + \sqrt{p}\right)\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p}}\right)^{\frac{1}{3}}. \tag{2.14}$$

The Tracy-Widom law of order 2 has a cumulative distribution function given by

$$F_2(s) = \exp\left(-\int_s^\infty (x-s)q^2(x)\,dx\right), \qquad s \in \mathbb{R}, \tag{2.15}$$

where $q(x)$ is the solution to the Painlevé II differential equation. Under the assumption that $n = n(p) \to \infty$ such that $n/p \to \gamma \geq 1$,

$$\frac{l_1 - \mu_{n,p}}{\sigma_{n,p}} \xrightarrow{d} W_2 \sim F_2. \tag{2.16}$$

2.6. **Tracy-Widom Distribution in MANOVA.** Mardia, Kent, and Bibby (1979) [6] defined the greatest root statistic as follows.

Definition 2.9. Let $\mathbf{U} \sim W_p(\mathbf{I}, m)$ be independent of $\mathbf{V} \sim W_p(\mathbf{I}, n)$, where $m \geq p$. The greatest root statistic is the largest eigenvalue of $(\mathbf{U}+\mathbf{V})^{-1}\mathbf{V}$ and its distribution is denoted $\theta(p,m,n)$. Alternatively, $\theta(p,m,n)$ could also be defined as the largest root of the determinantal equation

$$\det[\mathbf{V} - \theta(\mathbf{U}+\mathbf{V})] = 0.$$

Since $\mathbf{U}$ is positive definite, $0 < \theta < 1$. As a double Wishart problem, the Jacobi orthogonal polynomial is represented in the joint density function of $\min(n,p)$ nonzero roots of the determinantal equation above, which is

$$p(\theta) = C \prod_{i=1}^{\min(n,p)} \theta_i^{\frac{|n-p|-1}{2}} (1-\theta_i)^{\frac{m-p-1}{2}} \prod_{i \neq j} |\theta_i - \theta_j|. \tag{2.17}$$

The standardized logit of $\theta(p,m,n)$ is approximately Tracy-Widom distributed, that is,

$$\frac{\log\left(\frac{\theta(p,m,n)}{1-\theta(p,m,n)}\right) - \mu(p,m,n)}{\sigma(p,m,n)} \overset{d}{\Rightarrow} F_1, \tag{2.18}$$

where:

$$\begin{aligned}
\mu(p,m,n) &= 2\log\tan\left(\frac{\phi+\gamma}{2}\right); \\
\sigma^3(p,m,n) &= \frac{16}{(m+n-1)^2\sin^2(\phi+\gamma)\sin\phi\sin\gamma}; \\
\sin^2\left(\frac{\phi}{2}\right) &= \frac{\max(p,n)-\frac{1}{2}}{m+n-1}; \text{ and} \\
\sin^2\left(\frac{\gamma}{2}\right) &= \frac{\min(p,n)-\frac{1}{2}}{m+n-1}.
\end{aligned}$$

Johnstone (2009) [4] stated that the parameters p, m, and n defined by Mardia, Kent, and Bibby are related to the statistics table parameter $\mathtt{s}$, $\mathtt{m}$, and $\mathtt{n}$ by

$$\begin{aligned}
\mathtt{s} &= \min(n,p);\\
\mathtt{m} &= \frac{|n-p|-1}{2};\\
\mathtt{n} &= \frac{m-p-1}{2};\ \text{and}\\
\mathtt{N} &= 2(\mathtt{s}+\mathtt{m}+\mathtt{n})+1
\end{aligned}$$

such that

$$\begin{aligned}
p(\theta) &= C\prod_{i=1}^{\mathtt{s}}\theta_i^{\mathtt{m}}(1-\theta_i)^{\mathtt{n}}\Delta(\theta);\\
\sigma^3 &= \frac{16}{\mathtt{N}^2\sin^2(\phi+\gamma)\sin\phi\sin\gamma};\\
\sin^2\left(\frac{\phi}{2}\right) &= \frac{\mathtt{s}+2\mathtt{m}+\frac{1}{2}}{\mathtt{N}};\ \text{and}\\
\sin^2\left(\frac{\gamma}{2}\right) &= \frac{\mathtt{s}-\frac{1}{2}}{\mathtt{N}}.
\end{aligned}$$

The characterization of θ is

$$\theta = \lambda_{\max}(\mathbf{U}^{-1}\mathbf{V}) = \max_{|\mathbf{u}|=1}\frac{\mathbf{u}^T\mathbf{V}\mathbf{u}}{\mathbf{u}^T\mathbf{U}\mathbf{u}}. \tag{2.19}$$

The quadratic forms of $\mathbf{U}$ and $\mathbf{V}$ consecutively follow $\chi^2_{(n)}$ and $\chi^2_{(m)}$, implying the ratio to follow an $F_{n,m}$ distribution. However, Johnstone (2009) [4] said the usage of the $F_{n,m}$ distribution is anticonservative as the dimension p gets higher. Therefore, the aim of this journal is to compare the performance of p-values from the Tracy-Widom and F distributions, which are defined as follows.

Definition 2.10. Let $\theta(p,m,n)=\theta_\alpha$ be the greatest root statistic corresponding to an exact p-value, α. The p-value obtained from the Tracy-Widom approximation is given by

$$P_{TW}(\theta_\alpha) = 1 - F_1\left(\frac{\operatorname{logit}(\theta_\alpha)-\mu}{\sigma}\right) \tag{2.20}$$

and the F bound on the p-value is denoted as

$$P_F(\theta_\alpha) = 1 - F_{\nu_1,\nu_2}\left(\frac{\nu_2\theta_\alpha}{\nu_1(1-\theta_\alpha)}\right) \tag{2.21}$$

where $\nu_1 = \mathtt{s}+2\mathtt{m}+1$ and $\nu_2=\mathtt{s}+2\mathtt{n}+1$. (Johnstone, 2009) [5]

3. Methods

3.1. **Data.** This article compares three Roy's largest root approximations in the test of common mean vectors among groups by using a simulated data set. The three approximations are F approximations, Tracy-Widom approximation, and F approximation from Tracy-Widom.

This research used two datasets. The first dataset is a randomly generated data from a multivariate normal-distributed population. The data consists of four groups which have a specific mean vector and a uniform variance-covariance matrix. Dimensions for each group are a combination of the p variables of 10, 50, and 250 matched with the n observations of 10, 20, 50, 100, 250, and 1000. Hence, there are 18 unique combinations of p and n that will produce a p/n ratio in the range of 0.01 to 25. On the other hand, each unique combination will be simulated 200 times for two cases. The first case is by simulating all four groups using common mean vectors and the second case by simulating each of four groups using different mean vectors.

The second dataset is Mice Protein Expression, which was taken from the Kaggle Datasets. The dataset consists of the expression levels of 77 proteins/protein modifications that produced detectable signals in the nuclear fraction of the cortex. There are 38 control mice and 34 trisomic mice (down syndrome), yielding a total of 72 mice. In the experiments, fifteen measurements were registered of each protein per sample/mouse. Each measurement can be considered as an independent sample/mouse. The eight classes of mice are described based on features such as genotype, behavior and treatment. According to behavior, some mice have been stimulated to learn (context-shock) and others have not (shock-context) and in order to assess the effect of the drug memantine in recovering the ability to learn in trisomic mice, some rats have been injected with the drug and others have not.

In this research, the focus is constrained on trisomic mice (down syndrome), so there are four classes based on the combination of behavior and treatment. Furthermore, since there are several missing values, only response variables (77 proteins) with missing values no more than ten were selected. Then, any observations that still contain missing values were removed, producing 71 variables and 507 observations combined from the four classes. Subsequently, any variables that are strongly correlated with other variables were removed by taking a correlation threshold of 0.9 as a criterion in discarding the variable, obtaining a final p-dimensional data of 58 with n-observation of 135, 105, 135 and 132 for each class.

3.2. **Analysis Method.** Firsthandedly, an explorative analysis is conducted to determine the real properties of the datasets in terms of variance and mean. Next, it is necessary to check each dataset in terms of MANOVA assumptions: (1) possession of both common mean vector and (2) common variance-covariance matrix for group i (determining each variance homogeneity), (3) independence of sampling the observations, and (4) multivariate normality of the observations. The simulated dataset is not evaluated based on the assumptions as it is randomly generated from a multivariate normal-distributed population with a common mean vector and a uniform variance-covariance matrix.

Thereupon, comparisons are applied on both simulated and real datasets in the hypothesis tests of common mean vectors among the groups through three Roy's largest root approximations: F approximations, Tracy-Widom approximation, and F approximation from Tracy-Widom. This research is intrigued to discover the overall accuracy of hypothesis testing results through simulated datasets (each with varying p and n of one another) and real dataset. Whilst simulated datasets use proportion of rejected null hypotheses on the equality of vector means across groups under a null and alternative condition, application to real dataset intends to further prove conclusion from the simulated datasets.

4. Results and Discussion

4.1. Simulated Data.

The outcomes of the simulated data are given on the table below.

Table 3: The Outcomes Under Null Condition.

No.	p	n	F app.	TW	F app. TW
1	10.00	10.00	32.00	65.00	53.50
2	10.00	20.00	41.50	33.50	33.50
3	10.00	50.00	36.00	12.00	12.00
4	10.00	100.00	30.50	4.50	4.50
5	10.00	250.00	36.00	5.50	5.50
6	10.00	1000.00	26.50	4.50	4.50
7	50.00	10.00	41.00	85.50	85.50
8	50.00	20.00	41.00	85.50	85.50
9	50.00	50.00	41.00	85.50	85.50
10	50.00	100.00	37.00	36.00	36.00
11	50.00	250.00	39.00	12.00	12.00
12	50.00	1000.00	40.00	4.50	4.50
13	250.00	10.00	40.00	100.00	100.00
14	250.00	20.00	40.00	100.00	100.00
15	250.00	50.00	40.00	100.00	100.00
16	250.00	100.00	40.00	100.00	100.00

Continued on the next page...

Table 3: The Outcomes Under Null Condition (continued)

No.	p	n	F app.	TW	F app. TW
17	250.00	250.00	40.00	100.00	100.00
18	250.00	1000.00	39.00	31.00	31.00

Table 3 shows the proportion of rejected null hypotheses on the equality of vector means across groups. Utilizing a null condition assuming equal vector means across the groups and setting $\alpha = 0.05$, theoretically, 5% of the simulated data would be rejected by the test, constituting a Type I error for 5% of the simulated dataset. The table summarizes the proportion of Type I errors resulting from 200 simulated datasets for each combination of p and n using three approximations of Roy's largest root statistical test.

As depicted in the table, the proportion of Type I errors for the F-approximation method appears to remain relatively constant across different combinations of p and n. This proportion ranges from 25% to 42%, significantly surpassing the $\alpha = 0.05$. Specifically, for $n \leq p$, the obtained proportion of Type I error is approximately 40%, with the exception of $p = 10$ and $n = 10$, resulting in a 32% Type I error. Conversely, for each value of p, the Type I error tends to be smaller for $p/n \leq 1$ compared to $p/n > 1$. Overall, under the true null condition, the performance of the F-approximation tends to produce a Type I error that is much larger than the $\alpha = 0.05$.

In contrast, both the Tracy-Widom method and the F-approximation of the Tracy-Widom consistently yield the same proportion of Type I error (except for $p = 10$ and $n = 10$). Specifically, for $p/n \geq 1$, there is a tendency to reject null hypotheses, with proportions ranging from 50% to 100%. Conversely, in the case of $p/n < 1$, the proportion of Type I error is lower than using the F-approximation method. Furthermore, for $p/n < 0.1$, the proportion of Type I error does not significantly differ from $\alpha = 0.05$.

Table 4: The Outcomes Under Alternative Condition.

No.	p	n	F app.	TW	F app. TW
1	10.00	10.00	40.00	70.50	52.00
2	10.00	20.00	34.50	28.50	28.50
3	10.00	50.00	52.50	17.00	17.00
4	10.00	100.00	82.50	50.50	50.50
5	10.00	250.00	100.00	100.00	100.00
6	10.00	1000.00	100.00	100.00	100.00

Continued on the next page...

Table 4: The Outcomes Under Alternative Condition (continued)

No.	p	n	F app.	TW	F app. TW
7	50.00	10.00	61.00	96.50	96.50
8	50.00	20.00	61.00	96.50	96.50
9	50.00	50.00	61.00	96.50	96.50
10	50.00	100.00	100.00	100.00	100.00
11	50.00	250.00	100.00	100.00	100.00
12	50.00	1000.00	100.00	100.00	25.00
13	250.00	10.00	100.00	100.00	0.00
14	250.00	20.00	100.00	100.00	0.00
15	250.00	50.00	100.00	100.00	0.00
16	250.00	100.00	100.00	100.00	0.00
17	250.00	250.00	100.00	100.00	0.00
18	250.00	1000.00	100.00	100.00	0.00

Table 4 shows the proportions of rejected null hypotheses on the equality of vector means across groups. Under alternative condition, we can say that those proportions are the power of statistic test. The table summarizes the proportion of power resulting from 200 simulated datasets for each combination of p and n using three approximations of Roy's largest root statistical test. The table illustrates a consistent superiority of the Tracy-Widom method over the F-approximation method of Tracy-Widom. Consequently, a detailed comparison will be conducted only between the Tracy-Widom method and the F-approximation.

In the case of large-dimensional samples ($p = 250$), both methods exhibit a power of 100% across 200 samples. Similarly, for samples with dimension $p = 50$, it is clear that for $p/n > 1$ results in a power of 100% for both methods. Conversely, when the $p/n \leq 1$, the F-approximation method achieves a power of 61%, while the Tracy-Widom method achieves a higher power of 96.5%.

In the case of small-dimensional samples ($p = 10$), a power of 100% is obtained when the $n > 100$. However, for $n \leq 100$, the power tends to be proportional to the value of n. In this scenario, the F-approximation method demonstrates a slight superiority than Tracy-Widom method.

Based on the information provided in the two preceding tables, it can be inferred that when the $p/n \geq 1$, the F-approximation method is preferable to both the Tracy-Widom method and the F-approximation of Tracy-Widom. This preference arises due to its smaller Type I error. Conversely, for a $p/n < 1$, considering the Type I error and power for each method, both the F-approximation and Tracy-Widom methods can be concurrently employed, particularly for small-dimensional data ($p = 10$). Meanwhile, the Tracy-Widom method is deemed suitable for large-dimensional data ($p = 50, 250$).

4.2. **Real Data Application.**

4.2.1. *Exploratory Data Analysis (EDA).* Through exploratory analysis, insights of data characteristics in terms of certain indicators can be obtained: uniformity of the mean vector for each group, the distribution and presence of outliers for each group and variables, and the uniformity of the variance-covariance matrix between groups.

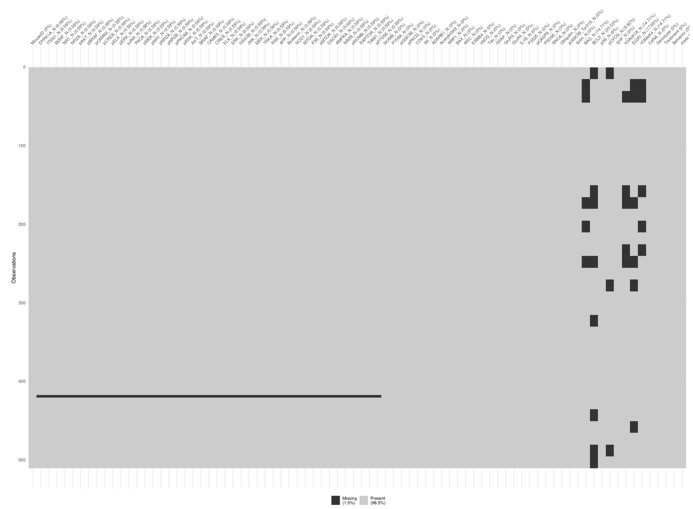


FIGURE 1. Variables expelled from the dataset.

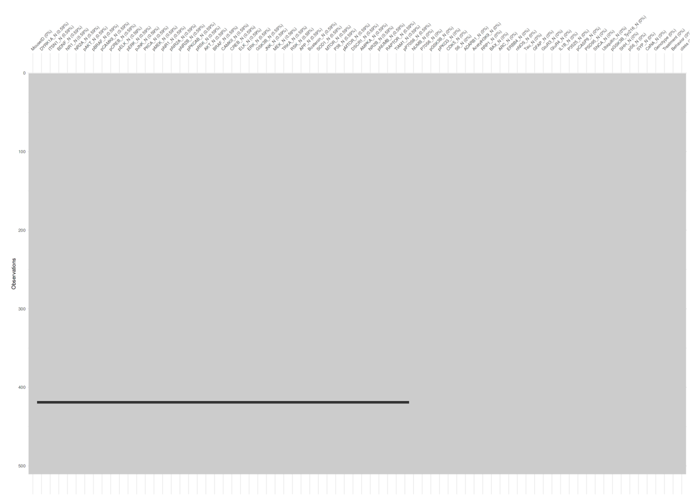


FIGURE 2. Detection of missing values in the dataset.

As a precedent step, removal of variables and observations with missing values are conducted. This process leads by Figure 1, for there are seven variables expelled from further research; alongside certain observations containing missing values shown in Figure 2 are dismissed. Thus, a final data is obtained with $p = 58$ and $n = 507$.

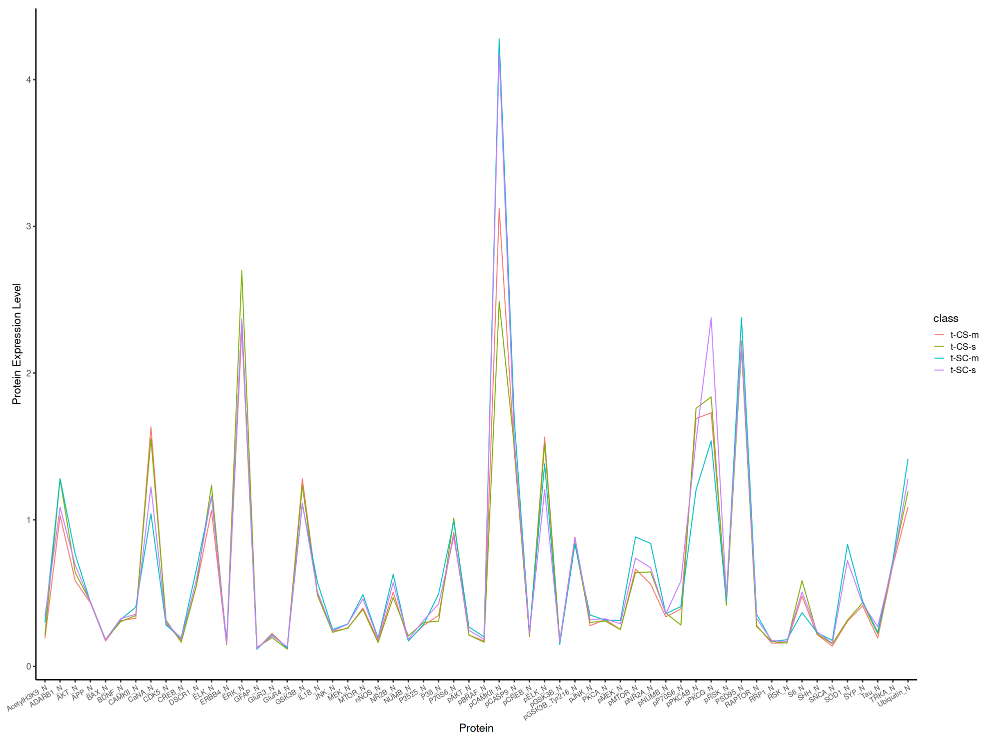
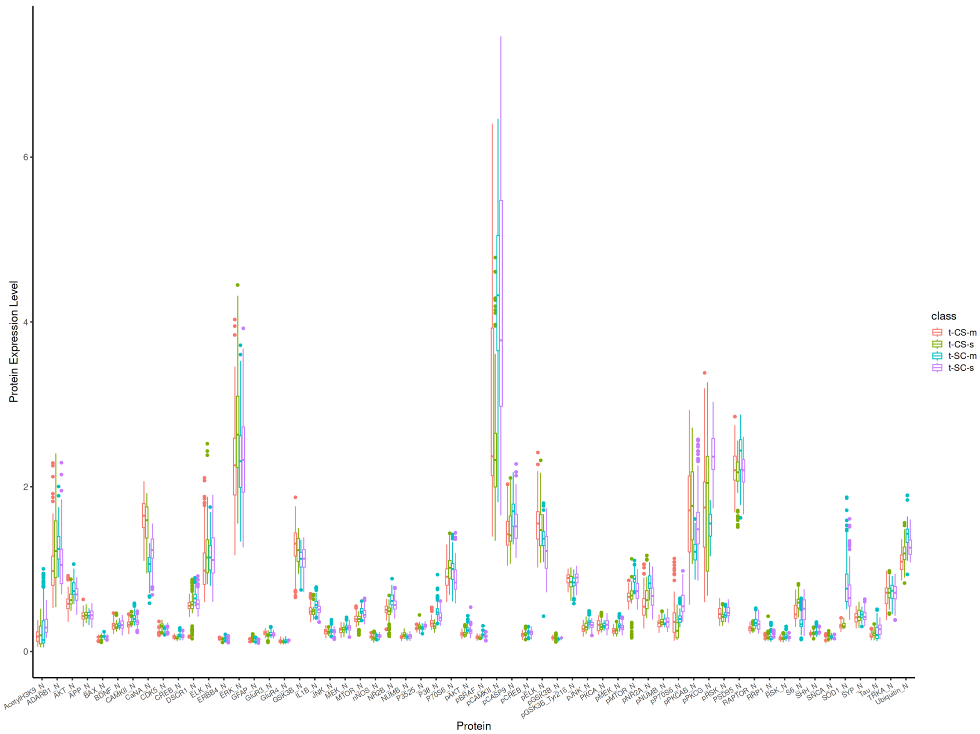
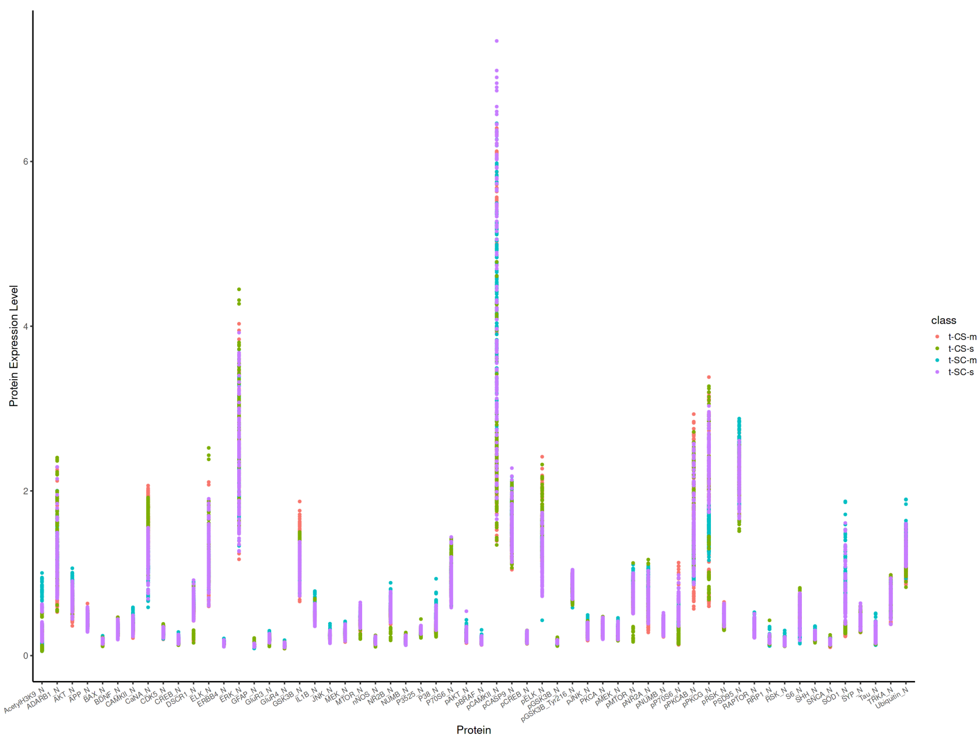


FIGURE 3. Protein comparison level.

From Figure 3, a common mean vector for each group might exist throughout every variables considered. The graph further implies mean differences amongst groups at least on one variable (also implies an alternative condition). Nonetheless, abundance of outliers might present in certain group and variable combinations as shown from the Figure 4(a) and Figure 4(b), respectfully.



(a) Outliers (1).

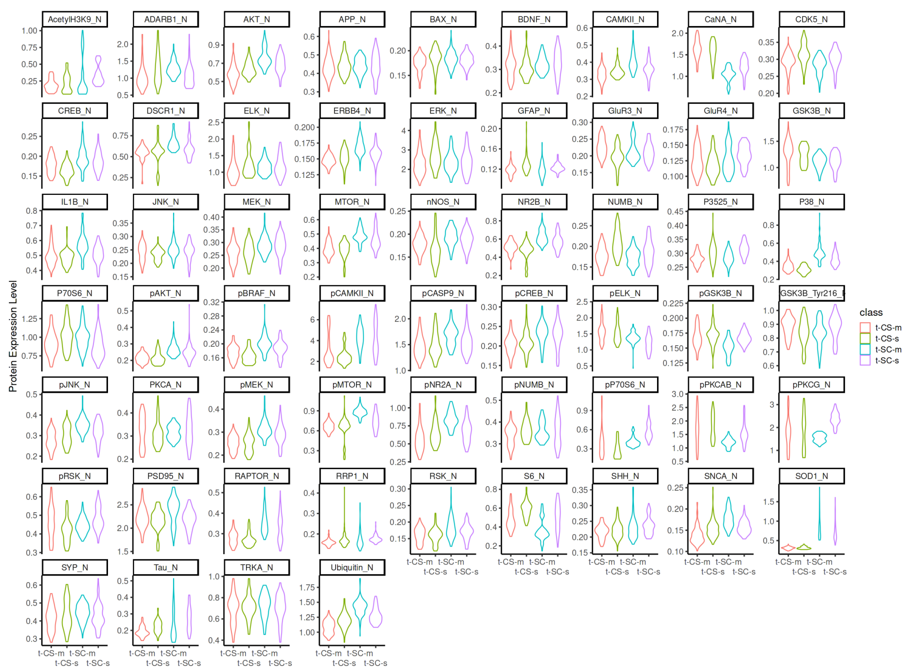


(b) Outliers (2).

FIGURE 4. Outliers present in group and variable combinations.

In more itemized manner, each variable's distribution alongside its outliers prone to shown both from its representative boxplots (Figure 5(a)) and violinplots (Figure 5(b)).

(a) Boxplots.



(b) Violinplots.

FIGURE 5. Outliers based on boxplots and violinplots.

Generally, variance-covariance matrix of the data straightforwardly shown in Figure 6. However, variance-covariance matrices could also be prevailed for groups from Figure 7. It is visible and concluded in a broad manner, each group does not possess similar matrices throughout the variables considered, nor it is similar to the grand variance-covariance matrix properties.

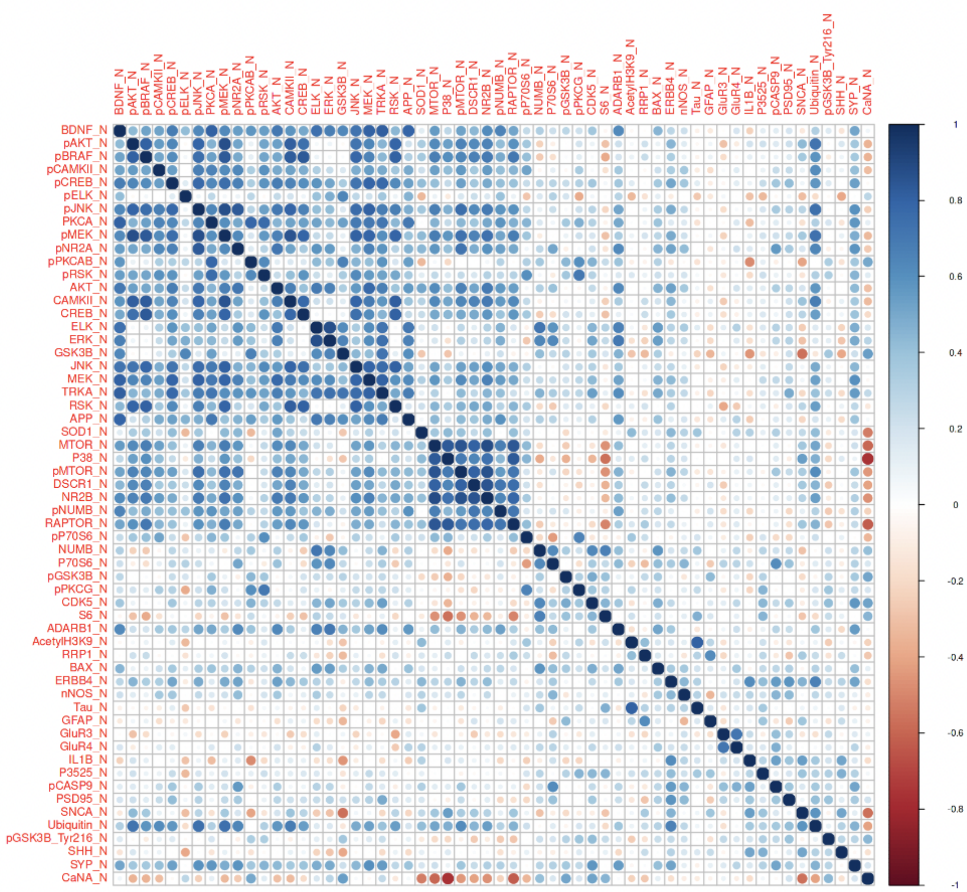


FIGURE 6. General variance-covariance matrix of the data.

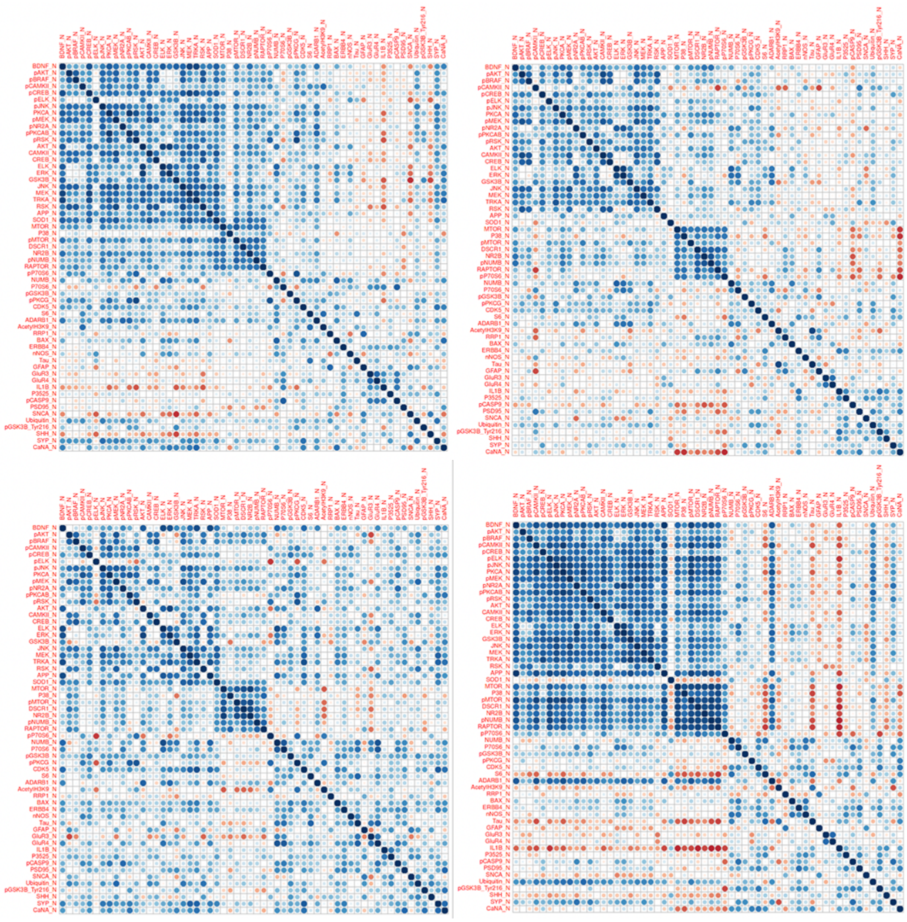


FIGURE 7. Variance-covariance matrices defined for each group.

4.2.2. *MANOVA Assumptions.* The first assumption stated that there are no subpopulations with different mean vectors for each of the four classes. Here, this assumption might be violated if mice with the same treatment and behavior had inconsistencies. By how the sample was collected, it might satisfy this assumption. However, from the exploratory analysis, there are lots of outliers in some of the variables, which might indicate the violation of this assumption.

The second assumption stated that the data from all four classes have a common variance-covariance matrix. From the exploratory analysis, there is a slight difference in the variance-covariance matrices of the four sample classes. Furthermore, we can check this assumption by using Box's M test of equality of variance-covariance matrices for all four classes. We find statistically significant evidence against the null hypothesis that the variance-covariance matrices are homogeneous ($L' = 22722.27$; $d.f. = 5133$; $p < 0.001$). We can conclude that there is a difference in the variance-covariance matrices of the four classes.

The third assumption stated that the subjects are independently sampled. This assumption really depends on how the sample was collected and could not be tested using any statistical tests. From the metadata of the dataset, each measurement or row can be considered as independent sample/mouse. However, this assumption is likely to be violated because some of the observations (rows) were collected from the same mouse.

The last assumption stated that the data are multivariate normally distributed. From the exploratory analysis, there are lots of outliers in some of the variables, indicating that the sample might be collected from heavy-tailed distribution other than normal distribution. Furthermore, we can check this assumption by using the Energy test for each of the four classes. We find a statistically significant evidence against the null hypothesis that the data from each of the four classes are sampled from multivariate normal distribution ($E = 4.24, 4.41, 4.12, 4.25$; $p < 0.001$). Therefore, we conclude that the data are not multivariate normally distributed.

4.2.3. *Roy's Largest Root for MANOVA.* Below is the outcome of Roy's largest root test arranged from preceding dataset.

TABLE 5. The Outcome of Roy's Largest Root Test.

Roy's largest root	P-value (F app.)	θ_α	P-value (F app. TW)
62.68911	0	0.177306	1

H_0 rejected (F app.)	H_0 rejected (TW)	H_0 rejected (F app. TW)
TRUE	TRUE	FALSE

Table 5 show that both traditional approach of F-approximation and Tracy-Widom approach gave identical conclusion of H_0 (on the equality of vector means across group) rejection under alternative condition, as the traditional F-approximation's p-value $< \alpha$ and Roy's largest root $> \theta_\alpha$. Obversely, the F-approximation from Tracy-Widom concluded a preference towards H_0 as its p-value $> \alpha$. This further prove superiority of both traditional F-approximation and Tracy-Widom approach as stated from the test towards simulated data with Tracy-Widom approach (0.05 critical value θ_α) by using Roy's largest root directly as a test statistic is preferable for this case ($p/n = 58/507 < 1$).

5. Conclusion

Based on the 200 simulated datasets of various combination of p and n, can be inferred that when the $p/n \geq 1$, the F-approximation method is preferable to both the Tracy-Widom method and the F-approximation of Tracy-Widom. This preference arises due to its smaller Type I error. Conversely, for a $p/n < 1$, considering the Type I error and power for each method, both the F-approximation and Tracy-Widom methods can be concurrently used, particularly for small-dimensional data ($p = 10$). Meanwhile, the Tracy-Widom method is deemed suitable for large-dimensional data ($p = 50, 250$). This further proven through the application towards the preprocessed Mice Protein Expression dataset; while it can be concurrently used with the classical F-approximation from Roy's largest root, Tracy-Widom approach directly using Roy's largest root test statistics convey a preferable result for the given dimensions of the dataset ($p/n = 58/507 < 1$).

References

1. Bain, L. J. and Engelhardt, M. (1992). *INTRODUCTION TO PROBABILITY AND MATHEMATICAL STATISTICS*. Duxbury Thomson Learning.
2. Johansson, K. (2000). Shape fluctuations and random matrices. *Communications in Mathematical Physics*, *209*, 437–476. https://doi.org/10.48550/arXiv.math/9903134
3. Johnstone, I. M. (2001). ON THE DISTRIBUTION OF THE LARGEST EIGENVALUE IN PRINCIPAL COMPONENTS ANALYSIS. *The Annals of Statistics*, *29*(2), 295—327. https://doi.org/10.1214/aos/1009210544
4. Johnstone, I. M. (2008). MULTIVARIATE ANALYSIS AND JACOBI ENSEMBLES: LARGEST EIGENVALUE, TRACY–WIDOM LIMITS AND RATES OF CONVERGENCE. Institute of Mathematical Statistics. *The Annals of Statistics*, *36*(6), 2638—2716. https://doi.org/10.1214/08-AOS605
5. Johnstone, I. M. (2009). APPROXIMATE NULL DISTRIBUTION OF THE LARGEST ROOT IN MULTIVARIATE ANALYSIS. Institute of Mathematical Statistics. *The Annals of Applied Statistics*, *3*(4), 1616–1633. https://doi.org/10.1214/08-AOAS220
6. Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press.
7. Orthogonal Polynomials. (n.d.). University of Sydney.
8. Paul, D. and Aue, A. (2014). Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference*, 150, 1-29. https://doi.org/10.1016/j.jspi.2013.09.005
9. Pennsylvania State University. *Lesson 8: Multivariate Analysis of Variance (MANOVA)*. https://online.stat.psu.edu/stat505/book/export/html/762
10. Rencher, A. C. (2002). *Methods of Multivariate Analysis*. A John Wiley Sons, Inc.

[1] Department of Mathematics, Universitas Gadjah Mada, Sleman, Indonesia.
Email address: agnes.fiorela.chrisnanda@mail.ugm.ac.id

[2] Department of Mathematics, Universitas Gadjah Mada, Sleman, Indonesia.
Email address: bryan.florentino.leo@mail.ugm.ac.id

[3] Department of Mathematics, Universitas Gadjah Mada, Sleman, Indonesia.
Email address: daud.muh2002@mail.ugm.ac.id

[4] Department of Mathematics, Universitas Gadjah Mada, Sleman, Indonesia.
Email address: raden.ajeng0903@mail.ugm.ac.id