Chapter 9
Multivariate Analysis of Variance

When samples are drawn from univariate normal populations, the test of hypotheses for mean parameters is conducted under various conditions. For example, the significance of the mean in one sample is determined using a t-test when the variance is unknown and estimated from the sample. The equality of means of two normal populations is tested using a t-test for independent or dependent samples when the variances are unknown. The equality of means for more than two normal populations is tested by analysis of variance in a univariate set-up, termed as ANOVA. The equality of mean vectors for more than two multivariate normal populations is tested by multivariate analysis of variance, briefly termed as MANOVA

In the context of the design of experiments, when only one variable is measured per plot, the design is analysed using analysis of variance. When more than one variable is measured per plot, the design is analysed using multivariate analysis of variance.

As in the univariate case, we have one-way ANOVA, two-way ANOVA, etc; similarly, in the multivariate case, we have one-way MANOVA, two-way MANOVA, etc. So we have a direct extension of every univariate design to a multivariate setup. The ANOVA tests differences in a single dependent variable, and MANOVA extends this to multiple dependent variables that are also correlated to test whether groups differ. Such analysis in MANOVA helps understand the interrelationships among groups and provides a more comprehensive picture than conducting two separate ANOVAs.

We follow here the ANOVA approach used in the analysis of the design of experiments.

Formulation of Multivariate One-Way Classification

Consider a formulation resulting from a design of experiment. Let k treatments be assigned in a completely random order to some experimental material, say, agricultural land.

Let

n_i : number of plots receiving i^{th} treatment, $i = 1, 2, \ldots, k$, and

$\underset{\sim}{x}_{ij}$: $p \times 1$ yield vector of the j^{th} plot receiving i^{th} treatment.

Assume $\underset{\sim}{x}_{ij}$ are generated from the model

$$\underset{\sim}{x}_{ij} = \underset{\sim}{\mu} + \underset{\sim}{\alpha}_i + \underset{\sim}{\epsilon}_{ij}, \qquad i = 1, 2, \ldots, k, \; j = 1, 2, \ldots, n_i$$

where $\underset{\sim}{\epsilon}_{ij}$ are independent and follow $N_p(0, \Sigma)$, $\underset{\sim}{\mu}$ represents the overall effect on yield vector and $\underset{\sim}{\alpha}_i$ represents the effect due to i^{th} treatment.

This design can be viewed as a multi-sample problem, i.e., we can consider $\underset{\sim}{x}_{ij}, j = 1, 2, \ldots, n_i$ as a random sample from $N_p(\underset{\sim}{\mu}_i, \Sigma), i = 1, 2, \ldots, k$ where

$$\underset{\sim}{\mu}_i = \underset{\sim}{\mu} + \underset{\sim}{\alpha}_i, \qquad i = 1, 2, \ldots, k.$$

Now

$$H_0 : \underset{\sim}{\mu}_1 = \underset{\sim}{\mu}_2 = \ldots. = \underset{\sim}{\mu}_k$$

or $\quad H_0$: There is no difference between the treatments $\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, \ldots, \underset{\sim}{\alpha}_k$.

Now we discuss the One-way MANOVA in detail, which is developed on similar lines to ANOVA for one-way classification in the univariate case.

Multivariate One-Way Analysis of Variance Model

Suppose there are several dependent variables on each experimental unit instead of just one variable. Assume that there are $k$independent random samples of size n which are obtained from p-variate multivariate normal distributed population with equal covariance matrices. The observations are presented in rows as observation vectors. Let $\underset{\sim}{x}_{ij}$ be a p-variate row vector from $N_p(\underset{\sim}{\mu}_i, \Sigma)$ which is a $p \times 1$ vector of j^{th} plot receiving i^{th} treatment.

The observations are arranged as follows:

	Sample 1	Sample 2	$\cdots$	Sample k
Observations	$\underset{\sim}{x}_{11}$ $\underset{\sim}{x}_{12}$ $\vdots$ $\underset{\sim}{x}_{1n}$	$\underset{\sim}{x}_{21}$ $\underset{\sim}{x}_{22}$ $\vdots$ $\underset{\sim}{x}_{2n}$	$\vdots$	$\underset{\sim}{x}_{k1}$ $\underset{\sim}{x}_{k2}$ $\vdots$ $\underset{\sim}{x}_{kn}$
Total	$\underset{\sim}{x}_{1.}$	$\underset{\sim}{x}_{2.}$	$\cdots$	$\underset{\sim}{x}_{k.}$
Mean	$\bar{\underset{\sim}{x}}_{1.}$	$\bar{\underset{\sim}{x}}_{2.}$	$\cdots$	$\bar{\underset{\sim}{x}}_{k.}$

Denote the

- Total of i^{th} sample $\underset{\sim}{x}_{i.} = \sum_{j=1}^{n} \underset{\sim}{x}_{ij}$
- Overall total $\underset{\sim}{x}_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n} \underset{\sim}{x}_{ij}$
- Mean of i^{th} sample $\bar{\underset{\sim}{x}}_{i.} = \frac{\underset{\sim}{x}_{i.}}{n}$
- Overall mean $\bar{\underset{\sim}{x}}_{..} = \frac{\underset{\sim}{x}_{..}}{kn}$.

The model for each observation vector is

$$\begin{aligned} \underset{\sim}{x}_{ij} &= \underset{\sim}{\mu} + \underset{\sim}{\alpha}_i + \underset{\sim}{\epsilon}_{ij} \\ &= \underset{\sim}{\mu}_i + \underset{\sim}{\epsilon}_{ij} \quad , \quad i = 1, 2, .., k; j = 1, 2, .., n \end{aligned}$$

which, in terms of p variables, can be written as

$$\begin{aligned} \begin{pmatrix} x_{ij1} \\ x_{ij2} \\ \vdots \\ x_{ijp} \end{pmatrix} &= \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} + \begin{pmatrix} \alpha_{i1} \\ \alpha_{i2} \\ \vdots \\ \alpha_{ip} \end{pmatrix} + \begin{pmatrix} \epsilon_{ij1} \\ \epsilon_{ij2} \\ \vdots \\ \epsilon_{ijp} \end{pmatrix} \\ &= \begin{pmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{ip} \end{pmatrix} + \begin{pmatrix} \epsilon_{ij1} \\ \epsilon_{ij2} \\ \vdots \\ \epsilon_{ijp} \end{pmatrix} \end{aligned}$$

so that the model for the r^{th} variable $(r = 1, 2, \ldots, p)$ in each vector $\underset{\sim}{x}_{ij}$ is

$$\begin{aligned} x_{ijr} &= \mu_r + \alpha_{ir} + \epsilon_{ijr} \\ &= \mu_{ir} + \epsilon_{ijr}. \end{aligned}$$

We wish to compare the mean vectors of k samples for significant differences as

$H_0 : \underset{\sim}{\mu}_1 = \underset{\sim}{\mu}_2 = \cdots = \underset{\sim}{\mu}_k$

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

Thus H_0 implies p sets of inequalities

$$H_0 : \begin{pmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{pmatrix} = \begin{pmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{pmatrix} = \cdots = \begin{pmatrix} \mu_{k1} \\ \mu_{k2} \\ \vdots \\ \mu_{kp} \end{pmatrix}$$

or

$$\begin{aligned} H_0 : \quad & \mu_{11} = \mu_{21} = \cdots = \mu_{k1} \\ & \mu_{12} = \mu_{22} = \cdots = \mu_{k2} \\ & \qquad\quad \vdots \\ & \mu_{1p} = \mu_{2p} = \cdots = \mu_{kp} \ . \end{aligned}$$

All $p(k-1)$ equalities must hold for H_0 to be true. Failure of only one equality will reject the hypothesis.

To better understand MANOVA before deriving the likelihood-ratio test, let us compare the different sums of squares (SS) in the univariate and multivariate settings.

Under a univariate set-up, the sum of squares between the groups and within the groups are given as

$$SSB = n \sum_{i=1}^{k} (\bar{x}_{i.} - \bar{x}_{..})^2 = \sum_{i=1}^{n} \frac{x_{i.}^2}{n} - \frac{x_{..}^2}{kn}$$

and

$$SSW = \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{x}_{i.})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij}^2 - \sum_{i=1}^{k} \frac{x_{i.}^2}{n},$$

respectively.

Under a multivariate set-up, the sum of squares between the groups and within the groups are given as a matrix as follows:

$$\begin{aligned} B &= n\sum_{i=1}^{k}(\bar{\underset{\sim}{x}}_{i.} - \bar{\underset{\sim}{x}}_{..})(\bar{\underset{\sim}{x}}_{i.} - \bar{\underset{\sim}{x}}_{..})' \\ &= \frac{1}{n}\sum_{i=1}^{k}\underset{\sim}{x}_{i.}\underset{\sim}{x}_{i.}' - \frac{1}{kn}\underset{\sim}{x}_{..}\underset{\sim}{x}_{..}' \end{aligned}$$

and

$$\begin{aligned} W &= \sum_{i=1}^{k}\sum_{j=1}^{n}(\underset{\sim}{x}_{ij} - \bar{\underset{\sim}{x}}_{i.})(\underset{\sim}{x}_{ij} - \bar{\underset{\sim}{x}}_{i.})' \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n}\underset{\sim}{x}_{ij}\underset{\sim}{x}_{ij}' - \frac{1}{n}\sum_{i=1}^{k}\underset{\sim}{x}_{i.}\underset{\sim}{x}_{i.}' \ , \end{aligned}$$

respectively.

The $p \times p$ matrix B has a between sum of squares on the diagonals for each of the p variables and sums of products for each pair of variables on the off-diagonals. The matrix B has the following form:

$$B = \begin{pmatrix} SSB_{11} & SPB_{12} & \cdots & SPB_{1p} \\ SPB_{12} & SSB_{22} & \cdots & SPB_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ SPB_{1p} & SPB_{2p} & \cdots & SSB_{pp} \end{pmatrix}$$

where, for example,

$$\begin{aligned} SSB_{22} &= n\sum_{i=1}^{k}(\bar{x}_{i.2} - \bar{x}_{..2})^2 \\ &= \frac{1}{n}\sum_{i=1}^{k}x_{i.2}^2 - \frac{x_{..2}^2}{kn} \\ SPB_{12} &= n\sum_{i=1}^{k}(\bar{x}_{i.1} - \bar{x}_{..1})(\bar{x}_{i.2} - \bar{x}_{..2}) \\ &= \frac{1}{n}\sum_{i=1}^{k}x_{i.1}x_{i.2} - \frac{x_{..1}x_{..2}}{kn}. \end{aligned}$$

The subscripts 1 and 2 indicate the first and second variables, respectively. For example, $\bar{x}_{i.2}$ is the second element in $\underset{\sim}{\bar{x}}_{i.} = (\bar{x}_{i.1}, \bar{x}_{i.2} \ldots \bar{x}_{i.p})'$!

Assume that there is no linear dependence in the variables, so

$$Rank(B) = min(p, \nu_H)$$

where $\nu_H = k - 1$ is the degrees of freedom of the hypothesis.

Similarly, the $p \times p$ matrix W has within sum of squares for each variable on the diagonal and sums of products for each pair of variables on the off-diagonal as follows:

$$W = \begin{pmatrix} SSW_{11} & SPW_{12} & \cdots & SPW_{1p} \\ SPW_{12} & SSW_{22} & \cdots & SPW_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ SPW_{1p} & SPW_{2p} & \cdots & SPW_{pp} \end{pmatrix}$$

where for example,

$$\begin{aligned} SSW_{22} &= \sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij2} - \bar{x}_{i.2})^2 \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij2}^2 - \sum_{i=1}^{k} \frac{x_{i.2}^2}{n} \\ SPW_{12} &= \sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij1} - \bar{x}_{i.1})(x_{ij2} - \bar{x}_{i.2}) \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij1}x_{ij2} - \frac{\sum_{i=1}^{k} x_{i.1}x_{i.2}}{n}. \end{aligned}$$

$$Rank(W) = min(p, \nu_E) = p \text{ if } \nu_E \geq p$$

where $\nu_E = p$ is the error degrees of freedom.

In the model

$$\underset{\sim}{x}_{ij} = \underset{\sim}{\mu} + \underset{\sim}{\alpha}_i + \underset{\sim}{\epsilon}_{ij}, \quad i = 1, 2, .., k; j = 1, 2, .., n_i.$$

$\underset{\sim}{\mu}$ represents the overall effect on yield vector and $\underset{\sim}{\alpha}_i$ represents the effect due to i^{th} treatment. We assume $\underset{\sim}{\epsilon}_{ij} \sim N_p(0, \Sigma)$ and $\underset{\sim}{\epsilon}_{ij}$'s are independently distributed, so $\underset{\sim}{x}_{ij} \sim N_p(\underset{\sim}{\mu}_i, \Sigma), i = 1, 2, .., k$.

Next, we develop the likelihood ratio test, which is also known as Wilk's Λ test.

Likelihood Ratio Test (Wilk's Λ Test)

First, we consider the likelihood function under H_1 and the respective maximum likelihood estimators.

The likelihood function under H_1 is

$$L = \prod_{i=1}^{k}\left(\frac{1}{2\pi|\Sigma|}\right)^{\frac{n_i}{2}} \exp\left[-\frac{n_i}{2} tr\Sigma^{-1}S_i - \frac{n_i}{2}(\underset{\sim}{\bar{x}}_{i.} - \underset{\sim}{\mu}_i)'\Sigma^{-1}(\underset{\sim}{\bar{x}}_{i.} - \underset{\sim}{\mu}_i)\right]$$

$$\log L = -\frac{1}{2}\sum_{i=1}^{k}\left[n_i \log|2\pi\Sigma| + n_i tr\Sigma^{-1}\{S_i + (\underset{\sim}{\bar{x}}_{i.} - \underset{\sim}{\mu}_i)(\underset{\sim}{\bar{x}}_{i.} - \underset{\sim}{\mu}_i)'\}\right].$$

There are no restrictions on the population mean, so the maximum likelihood estimator of $\underset{\sim}{\mu}_i$ is $\underset{\sim}{\bar{x}}_{i.}$. Let

$$n = \sum_{i=1}^{k} n_i$$

$$W = \sum_{i=1}^{k} n_i S_i = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\underset{\sim}{x}_{ij} - \underset{\sim}{\bar{x}}_{i.})(\underset{\sim}{x}_{ij} - \underset{\sim}{\bar{x}}_{i.})'$$

$$\underset{\sim}{\bar{x}}_{i.} = \frac{1}{n_i}\sum_{j=1}^{n_i}\underset{\sim}{x}_{ij}$$

$$\text{and} \quad \underset{\sim}{\bar{x}}_{..} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\underset{\sim}{x}_{ij}.$$

Then the maximum likelihood estimator of Σ is

$$\hat{\Sigma} = \frac{W}{n}.$$

Next, we consider the likelihood function under H_0 and the respective maximum likelihood estimators as follows: Since all the observations can be viewed under H_0 as constituting a single random sample, the maximum likelihood estimators of $\underset{\sim}{\mu}$ and Σ are

$$\hat{\underset{\sim}{\mu}} = \underset{\sim}{\bar{x}}_{..}$$

$$\text{and} \qquad \hat{\Sigma} = S = \frac{T}{n},$$

respectively, where

$$T = \sum_{i=1}^{k}\sum_{j=1}^{n}(\underset{\sim}{x}_{ij} - \underset{\sim}{\bar{x}}_{..})(\underset{\sim}{x}_{ij} - \underset{\sim}{\bar{x}}_{..})'.$$

Then the likelihood ratio test statistic is obtained as

$$\lambda = \left[\frac{|W|}{|nS|}\right]^{\frac{n}{2}} = |T^{-1}W|^{\frac{n}{2}},$$

where $T = nS$ is the matrix of the total sum of squares and products. The matrix T is derived by regarding all the data matrices as if they constituted a single sample, and

$$T = W + B$$

where W and B are the within and between group matrices of the sum of squares and products, and

$$B = T - W = \sum_{i=1}^{k} n_i(\underset{\sim}{\bar{x}}_{i.} - \underset{\sim}{\bar{x}}_{..})(\underset{\sim}{\bar{x}}_{i.} - \underset{\sim}{\bar{x}}_{..})'.$$

Then

$$\lambda^{\frac{2}{n}} = \frac{|W|}{|B + W|} = |I + W^{-1}B|^{-1}.$$

We need to find the probability distribution of the likelihood ratio test statistic.

Note that $W^{-1}B$ is a generalization of the univariate variance ratio. It tends to zero if H_0 is true.

To derive the distribution of $\lambda^{\frac{2}{n}}$, write k samples as a single data matrix X of order $n \times p$ as

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix}$$

where X_i is an $n_i \times p$ matrix representing the observations from the i^{th} sample, $i = 1, 2, .., k$.

Let E_i be an $n \times 1$ vector with the elements unity in places corresponding to i^{th} sample and 0 elsewhere.

Set $F_i = diag(E_i)$. Then $F = \sum_{i=1}^{k} F_i$ and $E = \sum_{i=1}^{k} E_i$. Let the centering matrix for i^{th} sample be

$$H_i = F_i - \frac{E_i E_i'}{n_i}$$

so that $n_i S_i = X' H_i X$.

Wilk's Lambda

When $A \sim W_p(I, m)$ and $B \sim W_p(I, n)$ are independent, $m \geq p$, then

$$|\Lambda| = \frac{|A|}{|A + B|} = |I + A^{-1}B|^{-1}$$

has a Wilk's lambda distribution with parameters p, m and n and denoted as $\Lambda(p, m, n)$.

The Λ family of distributions occurs frequently in the context of the likelihood ratio test. The parameters m and n usually represent the error degrees of freedom and the hypothesis degrees of freedom, respectively. So $(m + n)$ represents the total degrees of freedom.

Distribution of Λ

The Wilk's lambda is distributed as the distribution of the product of independent Beta-distributed random variables as follows:

$$\Lambda(p, m, n) \sim \prod_{i=1}^{n} u_i$$

where $u_1, u_2, \ldots, u_n$ are n independent variables and

$$u_i \sim Beta\left(\frac{m + i - p}{2}, \frac{p}{2}\right), i = 1, 2, .., n.$$

Proof: Write

$$B = X'X$$

where rows of X are identically and independently distributed following $N_p(0, I)$.

Let X_i be an $i \times p$ matrix consisting of the first i rows of X and let $M_i = A + X_i' X_i, i = 1, 2, \ldots, n$.

Note that

$$\begin{aligned} M_0 &= A \\ M_n &= A + B \\ \text{and} \quad M_i &= M_{i-1} + \underset{\sim}{x}_i \underset{\sim}{x}_i'. \end{aligned}$$

Now write

$$\begin{aligned} \Lambda(p, m, n) &= \frac{|A|}{|A+B|} = \frac{|M_0|}{|M_n|} \\ &= \frac{|M_0|}{|M_1|} \cdot \frac{|M_1|}{|M_2|} \cdot \ldots \cdot \frac{|M_{n-1}|}{|M_n|} \\ &= u_1 \cdot u_2 \cdot \ldots \cdot u_n \end{aligned}$$

where $u_i = \frac{|M_{i-1}|}{|M_i|},\ i = 1, 2, \ldots, n$.

Now

$$M_i = M_{i-1} + \underset{\sim}{x}_i \underset{\sim}{x}_i'$$

and therefore using the result:

Result 1: $\frac{|M|}{|M + \underset{\sim}{d}\underset{\sim}{d}'|} \sim Beta(\frac{m-p+1}{2}, \frac{p}{2})$ where $\underset{\sim}{d} \sim N_p(0, I)$ and $M \sim W_p(m, I)$ and $\underset{\sim}{d}$ and M are independent, we consider M_{i-1} corresponding to M and $\underset{\sim}{x}_i$ to d, and have

$$u_i \sim Beta\left(\frac{m+i-p}{2}, \frac{p}{2}\right), i = 1, 2, \ldots, n.$$

It remains to show that u_i are statistically independent. From the following theorem:

Theorem 2: If $\underset{\sim}{d} \sim N_p(0, I)$, $M \sim W_p(I, m)$ and d and M are independently distributed, then $\underset{\sim}{d}' M^{-1} \underset{\sim}{d}$ is independent of $M + \underset{\sim}{d}\underset{\sim}{d}'$.

We observe that M_i is independent of

$$1 + \underset{\sim}{x}_i' M_{i-1}^{-1} \underset{\sim}{x}_i = \frac{|M_i|}{|M_{i-1}|} = u_i^{-1}.$$

Since u_i is independent of $\underset{\sim}{x}_{i+1}, \underset{\sim}{x}_{i+2}, \ldots, \underset{\sim}{x}_n$ and $M_{i+j} = M_i + \sum_{k=1}^{j} \underset{\sim}{x}_{i+k} \underset{\sim}{x}_{i+k}'$, it follows that u_i is also independent of $M_{i+1}, M_{i+2}, \ldots, M_n$ and hence independent of $u_{i+1}, u_{i+2}, \ldots, u_n$. The result follows.

Some Properties of Wilk's Λ

The moments of Wilk's Λ are

$$E(\Lambda^h) = \prod_{i=1}^{n} \frac{\Gamma(\frac{m+n+1-i}{2})\Gamma(h+\frac{m+1-i}{2})}{\Gamma(h+\frac{m+n+1-i}{2})\Gamma(\frac{m+1-i}{2})}.$$

Also

$$\begin{aligned}
\frac{1-\Lambda(p,m,1)}{\Lambda(p,m,1)} &\sim \frac{p}{m-p+1} F_{p,m-p+1} \\
\frac{1-\Lambda(1,m,n)}{\Lambda(1,m,n)} &\sim \frac{n}{m} F_{n,m} \\
\frac{1-\sqrt{\Lambda(p,m,2)}}{\sqrt{\Lambda(p,m,2)}} &\sim \frac{p}{m-p+1} F_{2p,2(m-p+1)} \\
\frac{1-\sqrt{\Lambda(2,m,n)}}{\sqrt{\Lambda(2,m,n)}} &\sim \frac{n}{m-1} F_{2n,2(m-1)}.
\end{aligned}$$

Theorem 3: The $\Lambda(p,m,n)$ and $\Lambda(n, m+n-p, p)$ distributions are the same.

Proof: Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ denote the characteristic roots of $A^{-1}B$ and let $k = \min(n,p)$ denote the number of non-zero such characteristic roots.

Using the following result,

Result 4: For a $p \times n$ matrix B and $n \times p$ matrix C,

$$|I_p + BC| = |I_n + CB|,$$

we can write

$$\begin{aligned}
\Lambda(p,m,n) &= |I + A^{-1}B| \\
&= \prod_{i=1}^{p} (1+\lambda_i)^{-1}.
\end{aligned}$$

Thus Λ is a function of the nonzero characteristic roots of $A^{-1}B$, so the result follows from

$$\Lambda = |I + A^{-1}B|^{-1} \sim \Lambda(p,m,n).$$

Some more results about Wilk's lambda are

1. Parameters p and m can be interchanged. The distribution of $\Lambda(p, m, n)$ is the same as that of $\Lambda(m, p, m + n - p)$.

2. If $\lambda_1, \lambda_2, \ldots, \lambda_s$ are the characteristic roots of $A^{-1}B$, then
$$\Lambda = \prod_{i=1}^{s} \frac{1}{1 + \lambda_i}.$$
The number of nonzero characteristic roots of $A^{-1}B$ is $s = \min(p, m) = rank(B)$. The matrix AB^{-1} has the same characteristic roots as of $B^{-1}A$.

3. $0 \leq \Lambda \leq 1$

4. When $m = 1$ or 2 or when $p = 1$ or 2, Wilk's Λ transforms to an exact F-statistic.

Such transformations for those special cases are given as follows:

Parameters p, m	Statistic having F-distribution	Degrees of freedom
Any p, $m = 1$	$\frac{1-\Lambda}{\Lambda} \cdot \frac{n-p+1}{p}$	$p, n - p + 1$
Any p, $m = 2$	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \cdot \frac{n-p+1}{p}$	$2p, 2(n - p + 1)$
Any m, $p = 1$	$\frac{1-\Lambda}{\Lambda} \cdot \frac{n}{m}$	m, n
Any n $p = 2$	$\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}} \cdot \frac{n-1}{m}$	$2m, 2(n - 1)$

For other values of p and m other than these, an approximate F-statistic is
$$F = \frac{1 - \Lambda^{\frac{1}{t}}}{\Lambda^{\frac{1}{t}}} \cdot \frac{df_2}{df_1}$$

where

$$df_1 = pm$$
$$df_2 = wt - \frac{1}{2}(pm-2)$$
$$w = m+n-\frac{1}{2}(p+m+1)$$
$$t = \sqrt{\frac{p^2m^2-4}{p^2+m^2-5}}.$$

A less accurate approximate test is $\chi^2 = -[n - \frac{1}{2}(p-m+1)] \ln \Lambda \sim \chi^2_{pm}$.

Distribution of Likelihood Ratio Test Statistic

After understanding the Wilk's Λ distribution, we consider the distribution of $\lambda^{\frac{2}{n}} = \frac{|W|}{|B+W|}$.

Set

$$C_1 = \sum_{i=1}^{k} H_i$$
$$\text{and} \quad C_2 = \sum_{i=1}^{k} \frac{E_i E_i'}{n_i} - \frac{EE'}{n}.$$

It can be verified that

$$W = X'C_1X$$
$$\text{and} \quad B = X'C_2X$$

where C_1 and C_2 are idempotent matrices with

$$Rank(C_1) = n-k$$
$$Rank(C_2) = k-1$$
$$\text{and} \qquad C_1C_2 = 0.$$

Now under H_0, $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$. So using Cochran's theorem and Craig's theorem,

$$W = X'C_1X \sim W_p(\Sigma, n-k)$$
$$\text{and} \quad B = X'C_2X \sim W_p(\Sigma, k-1).$$

Furthermore, W and B are independent. Note that the statement of Craig's theorem is as follows:

Theorem 5 (Craig theorem): If the rows of X are identically and independently distributed following $N_p(\underset{\sim}{\mu}, \Sigma)$ and if $C_1, C_2, \ldots, C_k$ are symmetric matrices, then $X'C_1X$, $X'C_2X, \ldots, X'C_kX$ are jointly independent if $C_rC_s = 0$ for all $r \neq s$.

Therefore $|I + W^{-1}B|^{-1} \sim \Lambda(p, n-k, k-1)$ provided $n \geq p+k$.

Under H_0,

$$W \sim W_p(\Sigma, n-k)$$

and

$$B \sim W_p(\Sigma, k-1)$$

where W and B are independent.

If $n \geq p+k$, then

$$\Lambda = \frac{|W|}{|B+W|} \sim \Lambda(p, n-k, k-1).$$

The decision rule is to reject H_0 for small values of Λ.

For calculations of Λ, the following result is useful:

$$\Lambda = \prod_{j=1}^{p} (1+\lambda_j)^{-1}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the characteristic root of $W^{-1}B$. The result follows on noting that if $\lambda_1, \lambda_2, \ldots, \lambda_p$ are the characteristic roots of $W^{-1}B$ then$(\lambda_i + 1), i = 1, 2, \ldots, p$ are the characteristic roots of $W^{-1}(B+W)$ and

$$\frac{(m-p+1)[1-\sqrt{\Lambda(p,m,2)}]}{p\sqrt{\Lambda(p,m,2)}} \sim F_{2p, 2(m-p+1)}.$$

As in the case of ANOVA, the results in multivariate one-way analysis can be presented in a MANOVA table as follows:

One-Way MANOVA Table

Source	Degrees of freedom	SSP Matrix	Wilk's Λ Criterion
Between samples	$k-1$	B	$\frac{\lvert W\rvert}{\lvert B+W\rvert}$
Within samples	$n-k$	$W(=T-B)$	$\sim \Lambda(p, n-k, k-1)$
Total	$n-1$	T	

where $B = \sum_{i=1}^{k} n_i(\underset{\sim}{\bar{x}}_{i.} - \underset{\sim}{\bar{x}}_{..})(\underset{\sim}{\bar{x}}_{i.} - \underset{\sim}{\bar{x}}_{..})'$ and $T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\underset{\sim}{x}_{ij} - \underset{\sim}{\bar{x}}_{..})(\underset{\sim}{x}_{ij} - \underset{\sim}{\bar{x}}_{..})'$.

In some cases, the distribution of the product reduces to an exact variance ratio after a transformation

$$\begin{aligned}
\frac{1-\Lambda(p,m,1)}{\Lambda(p,m,1)} &\sim \frac{p}{m-p+1}F_{p,m-p+1} \\
\frac{1-\Lambda(1,m,n)}{\Lambda(1,m,n)} &\sim \frac{n}{m}F(n,m) \\
\frac{1-\sqrt{\Lambda(p,m,2)}}{\sqrt{\Lambda(p,m,2)}} &\sim \frac{p}{m-p+1}F_{2p,2(m-p+1)} \\
\frac{1-\sqrt{\Lambda(p,m,2)}}{\sqrt{\Lambda(2,m,n)}} &\sim \frac{n}{m-1}F_{2n,2(m-1)}.
\end{aligned}$$

For other values of n and p, provided m is large, we may use Bartlett's approximation:

$$-\left[m - \frac{p-n+1}{2}\right]\ln\Lambda(p,m,n) \sim \chi^2_{np}$$

asymptotically as $m \to \infty$.

Also, the statistics $\Lambda(p,m,1)$ and $\Lambda(1, m+1-p, p)$ are equivalent, and each corresponds to a single Beta-distributed variable $Beta(\frac{m-p+1}{2}, \frac{p}{2})$.

Further, $\Lambda(p,m,2)$ and $\Lambda(2, m+2-p, p)$ are equivalent and correspond to the product of a $Beta(\frac{m-p+1}{2}, \frac{p}{2})$ statistic with an independent $Beta(\frac{m-p+2}{2}, \frac{p}{2})$ statistic.

Rao's Approximation

If $\Lambda \sim \Lambda(p, t-q, q)$, then

$$R = \frac{gs - 2\lambda}{pq} \cdot \frac{1-\Lambda^{\frac{1}{s}}}{\Lambda^{\frac{1}{s}}}$$

where $g = t - \frac{p+q+1}{2}$, $\lambda = \frac{pq-2}{4}$, and $s^2 = \frac{p^2q^2-4}{p^2+q^2-5}$, then R has an asymptotic F distribution with pq and $(gs - 2\lambda)$ degrees of freedom.

Note that $(gs - 2\lambda)$ may not be an integer. But this does not cause any difficulty in consulting a table of the significant values of the variance ratio. The appropriate value lies between the significant values of $[gs - 2\lambda]$ and $[gs - 2\lambda] + 1$.

For practical purposes, it is safer to use the greatest integer values of $[gs - 2\lambda]$.

The decision rule is to reject H_0 if the calculated value of F is greater than the tabulated value of F at α level of significance.

Two-Way Multivariate Classification

Suppose we have nrc independent observations generated by the model

$$\underset{\sim}{x}_{ijk} = \underset{\sim}{\mu} + \underset{\sim}{\alpha}_i + \underset{\sim}{\tau}_j + \underset{\sim}{\eta}_{ij} + \underset{\sim}{\epsilon}_{ijk}, \qquad i = 1, 2, \ldots, r; j = 1, 2, \ldots, c; k = 1, 2, \ldots, n$$

where $\underset{\sim}{\alpha}_i$ is the i^{th} row effect, $\underset{\sim}{\tau}_j$ is the j^{th} column effect, $\underset{\sim}{\eta}_{ij}$ is the interaction effect between the i^{th} row and j^{th} column and $\underset{\sim}{\epsilon}_{ijk}$ is the random error term which is independently distributed following $N_p(0, \Sigma)$ for all i, j, k.

The number of observations in each $(i, j)^{th}$ cell is the same, so that the total sum of squares and product matrix can be suitably decomposed.

The null hypotheses are

$$H_{01} : \underset{\sim}{\alpha}_1 = \underset{\sim}{\alpha}_2 = \ldots = \underset{\sim}{\alpha}_r$$

$$H_{02} : \underset{\sim}{\tau}_1 = \underset{\sim}{\tau}_2 = \ldots = \underset{\sim}{\tau}_c$$

$$H_{03} : \text{All } \underset{\sim}{\eta}_{ij}\text{'s are equal.}$$

Let

T : matrix of total sum of squares and products,

R : matrix of rows sum of squares and products, and

E : matrix of errors sum of squares and products.

As in the case of univariate ANOVA, we can show that the following identity holds in the case of MANOVA,

$$T = R + C + E$$

where

$$\begin{aligned}
T &= \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}(\underset{\sim}{x}_{ijk} - \underset{\sim}{\bar{x}}_{...})(\underset{\sim}{x}_{ijk} - \underset{\sim}{\bar{x}}_{...})' \\
R &= cn\sum_{i=1}^{r}(\underset{\sim}{\bar{x}}_{i..} - \underset{\sim}{\bar{x}}_{...})(\underset{\sim}{\bar{x}}_{i..} - \underset{\sim}{\bar{x}}_{...})' \\
C &= rn\sum_{j=1}^{c}(\underset{\sim}{\bar{x}}_{.j.} - \underset{\sim}{\bar{x}}_{...})(\underset{\sim}{\bar{x}}_{.j.} - \underset{\sim}{\bar{x}}_{...})' \\
E &= \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}(\underset{\sim}{x}_{ijk} - \underset{\sim}{\bar{x}}_{i..} - \underset{\sim}{\bar{x}}_{.j.} + \underset{\sim}{\bar{x}}_{...})(\underset{\sim}{x}_{ijk} - \underset{\sim}{\bar{x}}_{i..} - \underset{\sim}{\bar{x}}_{.j.} + \underset{\sim}{\bar{x}}_{...})' \\
\underset{\sim}{\bar{x}}_{i..} &= \frac{1}{cn}\sum_{j=1}^{c}\sum_{k=1}^{n}\underset{\sim}{x}_{ijk} \\
\underset{\sim}{\bar{x}}_{.j.} &= \frac{1}{rn}\sum_{i=1}^{r}\sum_{k=1}^{n}\underset{\sim}{x}_{ijk} \\
\underset{\sim}{\bar{x}}_{...} &= \frac{1}{rcn}\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}\underset{\sim}{x}_{ijk}.
\end{aligned}$$

We may further decompose E into

$$E = I + W$$

where I and W are the matrices of the sum of squares and products due to interaction and residuals, respectively, where

$$\begin{aligned}
J &= n\sum_{i=1}^{r}\sum_{j=1}^{c}(\underset{\sim}{\bar{x}}_{ij.} - \underset{\sim}{\bar{x}}_{i..} - \underset{\sim}{\bar{x}}_{.j.} + \underset{\sim}{\bar{x}}_{...})(\underset{\sim}{\bar{x}}_{ij.} - \underset{\sim}{\bar{x}}_{i..} - \underset{\sim}{\bar{x}}_{.j.} + \underset{\sim}{\bar{x}}_{...})' \\
W &= \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}(\underset{\sim}{x}_{ijk} - \underset{\sim}{\bar{x}}_{ij.})(\underset{\sim}{x}_{ijk} - \underset{\sim}{\bar{x}}_{ij.})' \\
\underset{\sim}{\bar{x}}_{ij.} &= \frac{1}{n}\sum_{k=1}^{n}\underset{\sim}{x}_{ijk}.
\end{aligned}$$

Now we proceed to develop the tests for interaction and main effects on the same lines as in the case of univariate two-way ANOVA.

Test for Interaction

Under H_0,

$$T = R + C + I + W,$$

all $\underset{\sim}{\alpha}_i$'s, $\underset{\sim}{\tau}_j$'s and $\underset{\sim}{\eta}_{ij}$'s are zero, so T must have $W_p(\Sigma, rcn-1)$ distribution. Also, we can write

$$W = \sum_{i=1}^{r} \sum_{j=1}^{c} A_{ij}$$

whether or not H_0 holds, A_{ij}'s are identically and independently distributed following $W_p(\Sigma, n-1)$.

In one way MANOVA, we had

$$W \sim W_p(\Sigma, rc(n-1)).$$

In the same spirit, as in the case of univariate ANOVA, whether or not the $\underset{\sim}{\alpha}$'s and $\underset{\sim}{\tau}$'s vanish,

$$I \sim W_p(\Sigma, (r-1)(c-1)),$$

if the $\underset{\sim}{\eta}_{ij}$'s are equal.

It can be shown that the matrices R, C, I and W are distributed independently of one another, and the likelihood ratio test statistic for testing the equality of interaction terms is

$$\frac{|W|}{|W+I|} = \frac{|W|}{|E|} \sim \Lambda(p, \gamma_1, \gamma_2)$$

where $\gamma_1 = rc(n-1)$ and $\gamma_2 = (r-1)(c-1)$. The decision rule is to reject H_{03} of no interaction for low values of Λ.

Tests for Main Effect

If the column effect vanishes, then $C \sim W_p(\Sigma, c-1)$ distribution. The likelihood ratio test statistic to test equality of $\underset{\sim}{\tau}_j$ irrespective of $\underset{\sim}{\alpha}_i$ and $\underset{\sim}{\eta}_{ij}$ can be shown to be

$$\frac{|W|}{|W+C|} \sim \Lambda(p, \gamma_1, \gamma_2)$$

where $\gamma_1 = rc(n-1)$ and $\gamma_2 = c-1$.

The decision rule is to reject H_{02} for equality of $\underset{\sim}{\tau}_j$ for low values of Λ. Alternatively, look at the largest characteristic root θ of $C(W+C)^{-1}$.

Similarly, to test for equality of r row effect $\underset{\sim}{\alpha}_i$ irrespective of the column effect $\underset{\sim}{\tau}_j$, replace C by R, and interchange r and c in above.

Note that if significant interactions are present, then it does not make much sense to test for row and column effects. One possibility in this situation is to make tests separately on each of the rc rows and columns categories.

Alternatively, decide to ignore interaction effects completely, either because we have tested for them and shown them to be nonsignificant, or because $n = 1$, or because of various other reasons, we may have no $\underset{\sim}{\eta}_{ij}$ terms in our model. In such cases, we work on the error matrix E instead of W, and the test statistic for column effects is

$$\frac{|E|}{|E+C|} \sim \Lambda(p, \gamma_1, \gamma_2)$$

where $\gamma_1 = rcn - r - c + 1$, $\gamma_2 = c - 1$.