BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Multivariate Analysis

Course Code: 23ST12CC

Unit-IV

Wishart Distribution and Discriminant Function

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

Ms. I. Angel Agnes Mary
Guest Faculty
Department of Statistics

UNIT – IV

WISHART DISTRIBUTION AND DISCRIMINANT FUNCTION

Wishart distribution

If x_1, x_2, . . ., x_n are independent observations from N(μ, σ^2), it is known as $(n-1)s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 \sim \sigma^2 \chi^2_{n-1}$. The multivariate analogue of (n -1) s^2 is the matrix A and is called Wishart matrix. In other words the Wishart matrix is defined as the $p \times p$ symmetric matrix of sums of squares and cross products (of deviations about the mean) of the sample observations, from a p-variate nonsingular normal distribution. The distribution of A when the multivariate distribution is assumed normal is called Wishart distribution and is a generalization of χ^2 distribution in the univariate case. By the definition of A, the joint distribution of the $\frac{p(p+1)}{2}$ distinct elements a_{ij}, (i, j =1, 2,. . ., p; i ≤ j) of the symmetric matrix A.

Sampling distribution of sample mean and covariance matrix

One-dimensional case

The sample mean of x_1, . . . , x_n, we define

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and the sample variance is

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

p-dimensional case

Suppose we have p-variates X_1, . . . , X_p. For the vector of variates

$$\bar{X} = \begin{bmatrix} X_1 \\ \vdots \\ \vdots \\ X_p \end{bmatrix}$$

we have a p-variate sample with size n: $\bar{x}_1, \cdots, \bar{x}_n \in R^p$.

This sample of n observations gives the following data matrix:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = \begin{bmatrix} \bar{x}_1^T \\ \bar{x}_2^T \\ \vdots \\ \bar{x}_n^T \end{bmatrix}$$

Sample covariance matrix

For each variate X_j , j = 1, . . . , p, define its sample variance as

$$s_{jj} = s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2,\ j = 1, \cdots, p$$

and sample covariance between X_j and X_k

$$s_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)\,(x_{ik} - \bar{x}_k), 1 \le k,\ j \le p,\ j \ne k.$$

The sample covariance matrix is defined as

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix}$$

Then

$$S = \begin{bmatrix} \frac{1}{n-1}\sum_{i-1}^{n}(x_{i1} - \bar{x}_1)^2 & \cdots & \frac{1}{n-1}\sum_{i-1}^{n}(x_{i1} - \bar{x}_1)(x_{ip} - \bar{x}_p) \\ \vdots & \ddots & \vdots \\ \frac{1}{n-1}\sum_{i-1}^{n}(x_{ip} - \bar{x}_p)(x_{i1} - \bar{x}_1) & \cdots & \frac{1}{n-1}\sum_{i-1}^{n}(x_{ip} - \bar{x}_p)^2 \end{bmatrix}$$

$$= \frac{1}{n-1}\sum_{i-1}^{n}\begin{bmatrix} (x_{i1} - \bar{x}_1)^2 & \cdots & (x_{i1} - \bar{x}_1)(x_{ip} - \bar{x}_p) \\ \vdots & \ddots & \vdots \\ (x_{ip} - \bar{x}_p)(x_{i1} - \bar{x}_1) & \cdots & (x_{ip} - \bar{x}_p)^2 \end{bmatrix}$$

$$= \frac{1}{n-1}\sum_{i-1}^{n}\begin{bmatrix} x_{i1} - \bar{x}_1 \\ \vdots \\ x_{ip} - \bar{x}_p \end{bmatrix}\begin{bmatrix} x_{i1} - \bar{x}_1 & \cdots & x_{ip} - \bar{x}_p \end{bmatrix}$$

$$= \frac{1}{n-1}\sum_{i-1}^{n}(x_i - \bar{x})\,(x_i - \bar{x})^T$$

Properties of Wishart distribution

Theorem - 1

Suppose A_i $(i = 1, 2)$ are distributed independently according to $W_p(u_i, \Sigma)$ respectively, then $A_1 + A_2 \sim W_p(u_1 + u_2, \Sigma)$.

Proof

We know that the characteristic function of A_1, if $A_1 \sim W_p(u_1, \Sigma)$, is

$$\phi_{A_1}(\Theta) = \left|I - 2i\Theta\Sigma\right|^{-u_1/2}$$

Similarly, the characteristic function of A_2 will be

$$\phi_{A_2}(\Theta) = \left|I - 2i\Theta\Sigma\right|^{-u_2/2}$$

Since A_1 and A_2 are independently distributed, so

$$\phi_{A_1+A_2}(\Theta) = \phi_{A_1}(\Theta)\phi_{A_2}(\Theta) = \left|I - 2i\Theta\Sigma\right|^{-(u_1+u_2)/2}$$

The characteristic function of $W_p(u_1 + u_2, \Sigma)$, therefore

$$A_1 + A_2 \sim W_p(u_1 + u_2, \Sigma).$$

Hence proved.

Theorem - 2

If $A \sim W_p(n-1, \Sigma)$, then the distribution of $l'Al \sim (l'\Sigma l)\chi^2_{n-1}$, where l is a known vector.

Proof

Given $A \sim W_p(n-1, \Sigma)$, then $A = \sum_{\alpha=1}^{n-1} Z_\alpha Z'_\alpha$, where $Z_\alpha \sim N_p(0, \Sigma)$, and

$l'Al = \sum_{\alpha=1}^{n-1} l'Z_\alpha Z'_\alpha l = \sum_{\alpha=1}^{n-1} (l'Z_\alpha)(Z'_\alpha l)' = \sum_{\alpha=1}^{n-1} V_\alpha^2$, where $V_\alpha = l'Z_\alpha$ is $N(0, l'\Sigma l)$.

Therefore,

$$l'Al \sim (l'\Sigma l)\chi^2_{n-1}$$

Hence proved.

Generalized variance

The multivariate analogue of the variance σ^2 of a univariate distribution is the covariance matrix Σ, and the determinant of covariance matrix is termed as generalized variance of the multivariate distribution. Similarly, the generalized variance of a sample $x_1, x_2, \ldots, x_n$ is defined as

$$|S| = \left| \frac{1}{n-1} \sum_{\alpha=1}^{n} (x_\alpha - \bar{x})(x_\alpha - \bar{x})' \right|.$$

Simple correlation coefficient

A correlation coefficient is a measure of the strength of a linear relationship between two variables. In general, correlation coefficient values range from -1 to 1:

- 1 = a strong positive linear relationship. This means that for every positive increase in one variable, there is a proportional positive increase in the other variable. For instance, belt sizes increase almost perfectly in correlation with waist size.
- -1 = a strong negative linear relationship. In other words, for every positive increase in one variable, there is a proportional negative decrease in the other variable. For example, the amount of gas in a vehicle's tank decreases almost perfectly in correlation with speed.
- 0 = no linear relationship between the variables.

Formula for the sample correlation coefficient and the population correlation coefficient is

Sample correlation coefficient

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

where, S_x and S_y are the sample standard deviations, and S_{xy} is the sample covariance.

Population correlation coefficient

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

where, σ_x and σ_y as the population standard deviation, and σ_{xy} is the population covariance.

Estimation of simple correlation coefficient

Let $(X_1, Y_1), \ldots, (X_N, Y_N)$ be independent and identically distributed with bivariate normal distribution with means μ_X, μ_Y, variances σ_X^2, σ_Y^2, and correlation ρ $(|\rho| < 1)$. The sample correlation coefficient r is the sample covariance divided by the product of sample standard deviations

$$r_{XY} = \frac{S_{XY}}{S_X S_Y}$$

where,

$$S_{XY} = \frac{1}{n-1}\sum_{i-1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})$$

$$S_X^2 = \frac{1}{n-1}\sum_{i-1}^{N}(X_i - \overline{X})^2$$

$$S_Y^2 = \frac{1}{n-1}\sum_{i-1}^{N}(Y_i - \overline{Y})^2$$

Inferences about population correlation coefficient ρ are based on the famous Fisher's Z transformation, which has an approximately normal distribution irrespective of ρ and N is

$$Z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \sim N\left(\mu_Z, \sigma_Z^2\right)$$

where,

$$\mu_Z = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

$$\sigma_Z^2 = \frac{1}{N-3}$$

Multiple correlation coefficients

Multiple correlation was used in multiple linear regression to find the relationship between the dependent variable and the combined effect of independent variables on dependent variables in the model.

- It is nonnegative and varies from 0 to 1.
- When the multiple correlation coefficient was 1, then the relation was perfect, and regression residuals were zero.

- The multiple correlation coefficient was always greater than equal to any other combination variables 'simple correlation' in the model.
- If multiple correlations were zero, the dependent variable was uncorrelated with the variables in the model, and multiple regression failed to estimate the dependent variable when independent variables were known.

The formula and calculation procedure are as follows to calculate multiple correlation and R-square:

$$W = \begin{bmatrix} 1 & r_{xy} & r_{xz} \\ r_{yx} & 1 & r_{yz} \\ r_{zx} & r_{zy} & 1 \end{bmatrix}$$

$$W_{yy} = \begin{bmatrix} 1 & r_{xz} \\ r_{zx} & 1 \end{bmatrix}$$

$$R_{y.xz} = \sqrt{1 - \frac{W}{W_{yy}}}$$

$$R^2_{y.xz} = 1 - \frac{W}{W_{yy}}$$

Where, W is the determinant of the correlation matrix of all the factors (dependent and independent) in the model. W_{yy} is the cofactor of the dependent variable in the correlation matrix. y is the dependent variable, and x and z are independent variables in the model. R^2 is the coefficient of determination used as a goodness of fit of the model to explain the variance in the model, R is a multiple correlation coefficient.

Estimation of multiple correlation coefficients

The multiple correlation in the population is

$$\rho_{1(2,\cdots,p)} = \sqrt{\frac{\sigma_{12}' \Sigma_{22}^{-1} \sigma_{12}}{\sigma_{11}}} = \sqrt{\frac{\beta' \Sigma_{22} \beta}{\sigma_{11}}}$$

Given x_a $(a = 1, \ldots, n)$, $n > p$. We estimate Σ by $\hat{\Sigma} = \frac{A}{n} = \frac{n-1}{n} S$,

Where, $A = \sum_a (x_a - \bar{x})(x_a - \bar{x})'$.

Now A is partitioned as follows

$$\frac{A}{n} = \begin{pmatrix} \frac{a_{11}}{n} & \frac{a_{12}}{n}' \\ \frac{a_{12}}{n} & \frac{A_{22}}{n} \end{pmatrix}$$, and the estimate of β is $\hat{\beta}' = \sigma_{12}' \Sigma_{22}^{-1} = \frac{a_{12}'}{n} \left(\frac{A_{22}}{n} \right)^{-1} = a_{12}' A_{22}^{-1}$.

Using the above estimate, the sample multiple correlation coefficient of X_1 on $X_2, \ldots, X_p$ is

$$R_{1(2,\cdots,p)} = \sqrt{\frac{\hat{\sigma}_{12}' \hat{\Sigma}_{22}^{-1} \hat{\sigma}_{12}}{\hat{\sigma}_{11}}} = \sqrt{\frac{a_{12}' A_{22}^{-1} a_{12}}{a_{11}}}$$

And

$$1 - R_{1(2,\cdots,p)}^2 = \frac{a_{11} - a_{12}' A_{22}^{-1} a_{12}}{a_{11}} = \frac{\left| a_{11} - a_{12}' A_{22}^{-1} a_{12} \right| \left| A_{22} \right|}{a_{11} \left| A_{22} \right|} = \frac{|A|}{a_{11} |A_{22}|}.$$

Sampling distribution of multiple correlation coefficients in null case

The sample multiple correlation coefficient between X_1 and $X^{(2)}$ is defined by relation

$$R^2 = \frac{a_{12}' A_{22}^{-1} a_{12}}{a_{11}} \text{ and } 1 - R^2 = \frac{a_{11} - a_{12}' A_{22}^{-1} a_{12}}{a_{11}},$$

Where, $R^2 = R_{1(2,\cdots,p)}^2$ and $A = \begin{pmatrix} a_{11} & a_{12}' \\ a_{12} & A_{22} \end{pmatrix}$.

Therefore,

$$\frac{R^2}{1 - R^2} = \frac{a_{12}' A_{22}^{-1} a_{12}}{a_{11.2}}.$$

We know that, if A is partitioned as

$A = \begin{array}{c} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{array}{l} q \\ p-q \end{array} \end{array}$ and $A \sim W_p(n-1, \Sigma)$, then, $A_{11} \sim W_p(n-1, \Sigma_{11})$ and $A_{11} - A_{12}A_{22}^{-1}A_{21} \sim W_q(n-1-(p-q), \Sigma_{11.2})$.

In this case,

$$A_{11} \sim W_1(n-1, \sigma_{11}), \Rightarrow \frac{a_{11}}{\sigma_{11}} \sim \chi_{n-1}^2.$$

In null case $\rho_{1(2, 3 \ldots, p)} = 0$

$$\Sigma_{11.2} = \sigma_{11} - \sigma_{12}'\Sigma_{22}^{-1}\sigma_{21} = \sigma_{11}, \text{ since } \sigma_{12}' = 0, \text{ so that}$$

$$a_{11} - a_{12}'\Sigma_{22}^{-1}a_{21} \sim W_1(n-1-(p-1), \sigma_{11})$$

$$\Rightarrow \frac{a_{11} - a_{12}' A_{22}^{-1} a_{12}}{\sigma_{11}} \sim \chi_{n-p}^2$$

Consider

$$\frac{a_{11}}{\sigma_{11}} = \frac{a_{11} - a_{12}' A_{22}^{-1} a_{12}}{\sigma_{11}} + \frac{a_{12}' A_{22}^{-1} a_{12}}{\sigma_{11}}$$

$Q = Q_1 + Q_2$

Where, $Q \sim \chi_{n-1}^2$ and $Q_1 \sim \chi_{n-p}^2$.

From Fisher Cochran theorem Q_2 is independently distributed as $\chi_{n-1-(n-p)}^2$ i.e., $Q_2 \sim \chi_{p-1}^2$ and is independent of Q_1, hence,

$$F = \frac{R^2}{1-R^2} \times \frac{n-p}{p-1} = \frac{a_{12}' A_{22}^{-1} a_{12} / \sigma_{11}}{a_{11.2} / \sigma_{11}} \times \frac{n-p}{p-1} = \frac{\chi_{p-1}^2 / p-1}{\chi_{n-p}^2 / n-p} \sim F_{p-1, n-p}.$$

The distribution of the statistic F is,

$$df(F) = \frac{\left(\frac{v_1}{v_2}\right)^{v_1/2} F^{\frac{v_1}{2}-1}}{B\left(\frac{v_1}{2}, \frac{v_2}{2}\right)\left(1 + \frac{v_1}{v_2} F\right)^{(v_1+v_2)/2}} dF,$$

Where, $v_1 = p-1$, $v_2 = n-p$.

Put, $F = \frac{R^2}{1-R^2}\frac{v_2}{v_1}$, Then $dF = \frac{dR^2}{(1-R^2)}\frac{v_2}{v_1}$

$$df(R^2) = \frac{\left(\frac{v_1}{v_2}\right)^{v_1/2}\left(\frac{R^2}{1-R^2}\frac{v_2}{v_1}\right)^{\frac{v_1}{2}-1}}{B\left(\frac{v_1}{2}, \frac{v_2}{2}\right)\left(1 + \frac{R^2}{1-R^2}\right)^{(v_1+v_2)/2}} \frac{v_2}{v_1} \frac{dR^2}{(1-R^2)^2}$$

$$= \frac{\left(\frac{v_1}{v_2}\right)^{\frac{v_1}{2}-\frac{v_1}{2}+1-1}\left(\frac{R^2}{1-R^2}\right)^{\frac{v_1}{2}-1}}{B\left(\frac{v_1}{2},\frac{v_2}{2}\right)\left(\frac{R^2}{1-R^2}\right)^{(v_1+v_2)/2}}\frac{dR^2}{(1-R^2)^2}$$

$$= \frac{1}{B\left(\frac{v_1}{2},\frac{v_2}{2}\right)}(R^2)^{\frac{v_1}{2}-1}(1-R^2)^{\frac{v_1+v_2}{2}-\frac{v_1}{2}+1-2}dR^2$$

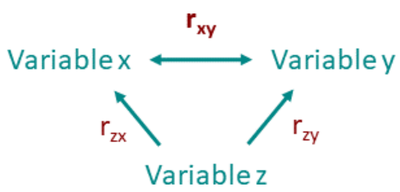
$$= \frac{1}{B\left(\frac{v_1}{2},\frac{v_2}{2}\right)}(R^2)^{\frac{v_1}{2}-1}(1-R^2)^{\frac{v_2}{2}-1}dR^2, \text{ put } dR^2 = 2R\,dR,$$

Thus the distribution of R,

$$df(R) = \frac{2R^{(v_1-1)}(1-R^2)^{\frac{v_2}{2}-1}}{B\left(\frac{v_1}{2},\frac{v_2}{2}\right)}dR = \frac{2R^{p-2}(1-R^2)^{\frac{n-p}{2}-1}}{B\left(\frac{p-1}{2},\frac{n-p}{2}\right)}dR, \; 0 < R < 1.$$

Partial correlation coefficient

Partial correlation, calculates the correlation between two variables to the exclusion of a third variable. This makes it possible to find out whether the correlation r_{xy} between variables x and y is generated by the variable z.



The partial correlation $r_{xy,z}$ tells how strongly the variable x correlates with the variable y, if the correlation of both variables with the variable z is calculated out.

For the calculation of the partial correlation, the three correlations between the individual variables are required. The formula for partial correlation is

$$r_{xy,z} = \frac{r_{xy} - r_{xz} \times r_{yz}}{\sqrt{\left(1-r_{xz}^2\right)\times\left(1-r_{yz}^2\right)}}$$

- r_{xy} is correlation between variable x and y
- r_{xz} is Correlation of the third variable z with the variable x
- r_{yz} is Correlation of the third variable z with the variable y

Estimation of partial correlation coefficient

The population partial correlation coefficient between X_i and X_j holding the components of $X^{(2)}$ fixed, is given by

$$\rho_{ij.q+1\cdots p} = \frac{\sigma_{ij.q+1\cdots p}}{\sqrt{(\sigma_{ii.q+1\cdots p})(\sigma_{jj.q+1\cdots p})}}.$$

Given a sample x_α $(\alpha = 1, 2, \ldots, n > p)$ from $N(\mu, \Sigma)$, the maximum likelihood estimate of $\rho_{ij.q+1\ldots p}$ is

$$\hat{\rho}_{ij.q+1\cdots p} = \frac{\hat{\sigma}_{ij.q+1\cdots p}}{\sqrt{(\hat{\sigma}_{ii.q+1\cdots p})(\hat{\sigma}_{jj.q+1\cdots p})}}$$

i.e., $r_{ij.q+1\cdots p} = \frac{a_{ij.q+1\cdots p}}{\sqrt{(a_{ii.q+1\cdots p})(a_{jj.q+1\cdots p})}}$ is called the sample partial correlation coefficient between X_i and X_j holding $X_{q+1}, \ldots, X_p$ fixed.

where, $a_{ij.q+1\cdots p} = A_{11} - A_{12}A_{22}^{-1}A_{21} = A_{11.2}$, $a_{ij.q+1\ldots p}$ is the i^{th} and j^{th} element of $A_{11.2}$, and $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$.

Sampling distribution of partial correlation coefficient in null case

The sample partial correlation coefficient between X_i and X_j holding $X_{q+1}, \ldots, X_p$ fixed is defined as

$$r_{ij.q+1\cdots p} = \frac{a_{ij.q+1\cdots p}}{\sqrt{(a_{ii.q+1\cdots p})(a_{jj.q+1\cdots p})}}$$

where, $a_{ij.q+1\cdots p} = A_{11} - A_{12}A_{22}^{-1}A_{21} = A_{11.2}$, i.e., $a_{ij.q+1\ldots p}$ is the i^{th} and j^{th} element of $A_{11.2}$, and $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$.

Let, $A = \sum_\alpha (x_\alpha - \bar{x})(x_\alpha - \bar{x})' \sim W_p(n-1, \Sigma)$, then

$$A_{11} - A_{12}A_{22}^{-1}A_{21} \sim W_q(n-1-(p-q), \Sigma_{11.2}).$$

The distribution of the sample partial correlation $r_{ij.q+1...p}$ based on a sample of size n from a distribution with population correlation $\rho_{ij.q+1...p}$ is same as the distribution of ordinary correlation coefficient r_{ij} based on a sample of size n-(p-q) from a distribution with the corresponding population partial correlation $\rho_{ij.q+1...p} = \rho$. Thus,

$$\sqrt{n-(p-q)-2}\,\frac{r}{\sqrt{1-r^2}} \sim t_{n-(p-q)-2}.$$

Discriminant function

Discriminant analysis is a technique for analyzing data when the criterion or dependent variable is categorical and the predictor or independent variables are interval in nature.

For example, the dependent variable may be the choice of the make of a new car (A, B or C) and the independent variables may be ratings of attributes of PCs on a seven-point Likert scale.

Discriminant analysis techniques are described by the number of categories possessed by the criterion variable. When the criterion variable has two categories, the technique is known as two-group discriminant analysis. When three or more categories are involved, the technique is referred to as multiple discriminant analysis. The main distinction is that in the twogroup case it is possible to derive only one discriminant function, but in multiple discriminant analysis more than one function may be computed.

Discriminant analysis can be used in medicine to assess the severity of a patient's condition and the prognosis of their disease outcome.

Objectives of discriminant function

- Development of discriminant functions, or linear combinations of the predictor or independent variables, that best discriminate between the categories of the criterion or dependent variable (groups).
- Examination of whether significant differences exist among the groups, in terms of the predictor variables.
- Determination of which predictor variables contribute to most of the intergroup differences.

- Classification of cases to one of the groups based on the values of the predictor variables.
- Evaluation of the accuracy of classification.

Assumptions of discriminant function

- **Multivariate normality:** The data for the variables represent a sample from a multivariate normal distribution.
- **Homogeneity of variances/covariances:** The variance/covariance matrices of variables are homogeneous across groups.
- **Independence:** Participants are assumed to be randomly sampled, and a participant's score on one variable is assumed to be independent of scores on that variable for all other participants.
- **No multicollinearity:** The variables that are used to discriminate between groups are not completely redundant.
- **Tolerance values:** The tolerance value for each variable is constantly checked.
- **Size of the smallest group:** The size of the smallest group must be larger than the number of predictor variables.

Fisher's discriminant function

Let μ denote the mean vector of the combined populations and B_μ the between groups sums of cross products, so that

$$B_\mu = \sum_{i=1}^{g} (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})' \qquad \text{where } \bar{\mu} = \frac{1}{g}\sum_{i=1}^{g} \mu_i$$

We consider the linear combination

$$Y = a'X$$

which has expected value

$$E(Y) = a'E(X \mid \pi_i) = a'\mu_i \qquad \text{for population } \pi_i$$

and variance

$$V(Y) = a'Cov(X)a = a'\Sigma a \qquad \text{for population}$$

Consequently, the expected value $\mu_{iY} = a'\,\mu_i$ changes as the population from which X is selected changes and define the overall mean

$$\bar{\mu}_Y = \frac{1}{g}\sum_{i=1}^{g}\mu_i Y = \frac{1}{g}\sum_{i=1}^{g}a'\mu_i = a'\left(\frac{1}{g}\sum_{i=1}^{g}\mu_i\right)$$

$$= a'\bar{\mu}$$

and form the ratio

$$\frac{\begin{pmatrix} sum\,of\;\; squared\; dis\tan ces\;\; form \\ populations\,to\,overall\; mean\,of\,Y \end{pmatrix}}{Variance\,of\;Y} = \frac{\sum_{i=1}^{g}(\mu_{iY}-\bar{\mu}_Y)^2}{\sigma_Y^2} = \frac{\sum_{i=1}^{g}(a'\mu_i - a'\bar{\mu})^2}{a'\Sigma a}$$

$$= \frac{a'\left(\sum_{i=1}^{g}(\mu_i-\bar{\mu})(\mu_i-\bar{\mu})'\right)a}{a'\Sigma a}$$

$$\frac{\sum_{i=1}^{g}(\mu_{iY}-\bar{\mu}_Y)^2}{\sigma_Y^2} = \frac{a'B_\mu a}{a'\Sigma a} \qquad \text{- - - - - - - - (1)}$$

The ratio in (1) measures the variability between the groups of Y-values relative to the common variability within groups. We can then select *a* to maximize this ratio.

Ordinarily, Σ and the μ_i are unavailable, but we have a training set consisting of correctly classified observations. Suppose the training set consists of a random sample of size n_i from population π_i, i = 1,2, ... , g. Denote the $n_i \times p$ data set, from population π_i, by X_i and its jth row by X'_{ij}. After first constructing the sample mean vectors

$$\bar{x}_j = \frac{1}{n_i}\sum_{j=1}^{n_i}X_{ij}$$

and the covariance matrices S_i, i = 1, 2, ... , g, we define the "overall average" vector

$$\bar{x} = \frac{1}{g}\sum_{i=1}^{g}\bar{x}_i$$

which is the $p\times 1$ vector average of the individual sample averages.

Next, analagous to B_μ we define the *sample between groups* matrix B. Let

$$B = \sum_{i=1}^{g}(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})'$$

Also, an estimate of Σ is based on the sample within groups matrix.

$$W=\sum_{i=1}^{g}(n_i-1)\,S_i=\sum_{i=1}^{g}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)\,(x_{ij}-\bar{x}_i)'$$

Consequently, $W/(n_1 + n_2 + ... + n_g - g) = S_{pooled}$ is the estimate of Σ. Before presenting the sample discriminants, we note that W is the constant $(n_1 + n_2 + ... + n_g - g)$ times S_{pooled}, so the same a that maximizes $\frac{\hat{a}'B\hat{a}}{\hat{a}'S_{pooled}\hat{a}}$ also maximizes $\frac{\hat{a}'B\hat{a}}{\hat{a}'W\hat{a}}$. Moreover, we can present the optimizing $\hat{a}$ in the more customary form as eigenvectors $\hat{e}_i$ of $W^{-1}B$, because if $W^{-1}B\hat{e}=\hat{\lambda}\hat{e}$ then $S_{pooled}^{-1}B\hat{e}=\hat{\lambda}(n_1+n_2+\cdots+n_g-g)\hat{e}$.

Wilk's criterion

Wilk's lambda (Λ) is a test statistic that's reported in results from MANOVA, discriminant analysis and other multivariate procedures. Wilk's Λ criterion is used to test the equality of the mean vectors of *k*, *p*-variate normal distributions. The distribution of the test statistic can be expressed as a function of several independent beta distributions.

$$\Lambda^*=\frac{|E|}{|H+E|}$$

Here, the determinant of the error sums of squares and cross-products matrix *E* is divided by the determinant of the total sum of squares and cross-products matrix $T = H + E$. If *H* is large relative to *E*, then $|H + E|$ will be large relative to $|E|$. Thus, we will reject the null hypothesis if Wilks lambda is small (close to zero).

$$\Lambda^*=\prod_{j=1}^{p}(1-\theta_j)\text{, where } e\geq p.$$

The following approximation based on the F-distribution is used to determine significance levels:

$$F_{ph,\,ft-g}=\frac{(ft-g)(1-\Lambda^{1/t})}{ph\,\Lambda^{1/t}}$$

where,

$$f=e-\frac{1}{2}(p-h+1)$$

$$g = \frac{ph-2}{2}$$

$$t = \sqrt{\frac{p^2h^2-4}{p^2+h^2-5}}, \qquad \text{if} \quad p^2+h^2-5>0$$

This approximation is exact if *p or h* ≥ 2.

Interpretation of Wilks' lambda

- In MANOVA, Wilks' lambda tests if there are differences between group means for a particular combination of dependent variables. It is similar to the F-test statistic in ANOVA. Lambda is a measure of the percent variance in dependent variables not explained by differences in levels of the independent variable. A value of zero means that there isn't any variance not explained by the independent variable (which is ideal). In other words, the closer to zero the statistic is, the more the variable in question contributes to the model. We would reject the null hypothesis when Wilk's lambda is close to zero, although this should be done in combination with a small p-value.
- In discriminant analysis, Wilk's lambda tests how well each level of independent variable contributes to the model. The scale ranges from 0 to 1, where 0 means total discrimination, and 1 means no discrimination. Each independent variable is tested by putting it into the model and then taking it out generating a Λ statistic. The significance of the change in Λ is measured with an F-test; if the F-value is greater than the critical value, the variable is kept in the model.

Interpretation of Correlation

The strength of the correlation can be interpreted as follows:

- **Very strong:** r is between 0.8 and 1.0 for a positive correlation, and between -0.8 and -1.0 for a negative correlation.
- **Strong:** r is between 0.6 and 0.8 for a positive correlation, and between -0.6 and -0.8 for a negative correlation
- **Moderate:** r is between 0.4 and 0.6 for a positive correlation, and between -0.4 and -0.6 for a negative correlation.
- **Weak:** r is between 0.2 and 0.4 for a positive correlation, and between -0.2 and -0.4 for a negative correlation.
- **Very weak:** r is between 0.0 and 0.2 for a positive correlation, and between 0.0 and -0.2 for a negative correlation.
- **No correlation:** r is 0.

SCL PROBLEMS

1. Multiple Correlation

Find Multiple Correlation from the following information:

Sample No.	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
a	24	27	28	23	25	32	33	24	23	25
B	41	44	44	40	43	49	52	47	44	45
C	55	52	56	50	51	56	48	55	56	55
D	36	39	38	30	37	34	31	32	38	35

Sample No.	***11***	***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***
A	24	23	26	23	27	21	34	24	26	32
B	41	47	46	41	43	37	42	56	53	44
C	56	48	56	55	43	57	56	56	53	48
D	38	32	36	36	39	28	33	38	31	32

Procedure

- Mean Vector = $\begin{bmatrix} \bar{a} \\ \bar{b} \\ \bar{c} \\ \bar{d} \end{bmatrix}$

- $\sigma_{ij} = \sigma_i^2 = \dfrac{\sum X_i^2}{n} - \bar{X}_i^2$

- $\sigma_{ij} = \dfrac{\sum X_i X_j}{n} - \bar{X}_i \bar{X}_j$

- $r_{ij} = \dfrac{\sigma_{ij}}{\sigma_i \sigma_j}$

- **Multiple Correlation:** Given variables x, y, and z, we define the multiple correlation coefficient

$$R_{z,xy} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2r_{xz}r_{yz}r_{xy}}{1 - r_{xy}^2}}$$

Calculation

Calculating mean vector and covariance matrix

Sample No.	***a***	***b***	***C***	***d***	***a²***	***b²***	***c²***	***d²***	***ab***	***ac***	***ad***	***bc***	***bd***	***cd***
1	24	41	55	36	576	1681	3025	1296	984	1320	864	2255	1476	1980
2	27	44	52	39	729	1936	2704	1521	1188	1404	1053	2288	1716	2028
3	28	44	56	38	784	1936	3136	1444	1232	1568	1064	2464	1672	2128
4	23	40	50	30	529	1600	2500	900	920	1150	690	2000	1200	1500
5	25	43	51	37	625	1849	2601	1369	1075	1275	925	2193	1591	1887
6	32	49	56	34	1024	2401	3136	1156	1568	1792	1088	2744	1666	1904
7	33	52	48	31	1089	2704	2304	961	1716	1584	1023	2496	1612	1488
8	24	47	55	32	576	2209	3025	1024	1128	1320	768	2585	1504	1760
9	23	44	56	38	529	1936	3136	1444	1012	1288	874	2464	1672	2128
10	25	45	55	35	625	2025	3025	1225	1125	1375	875	2475	1575	1925
11	24	41	56	38	576	1681	3136	1444	984	1344	912	2296	1558	2128
12	23	47	48	32	529	2209	2304	1024	1081	1104	736	2256	1504	1536
13	26	46	56	36	676	2116	3136	1296	1196	1456	936	2576	1656	2016
14	23	41	55	36	529	1681	3025	1296	943	1265	828	2255	1476	1980
15	27	43	43	39	729	1849	1849	1521	1161	1161	1053	1849	1677	1677
16	21	37	57	28	441	1369	3249	784	777	1197	588	2109	1036	1596
17	34	42	56	33	1156	1764	3136	1089	1428	1904	1122	2352	1386	1848
18	24	56	56	38	576	3136	3136	1444	1344	1344	912	3136	2128	2128
19	26	53	53	31	676	2809	2809	961	1378	1378	806	2809	1643	1643
20	32	44	48	32	1024	1936	2304	1024	1408	1536	1024	2112	1408	1536
Total	**524**	**899**	**1062**	**693**	**13998**	**40827**	**56676**	**24223**	**23648**	**27765**	**18141**	**47714**	**31156**	**36816**

Mean Vector

$$\bar{a} = \frac{524}{20} = 26.2 \qquad \bar{b} = \frac{899}{20} = 44.95$$

$$\bar{c} = \frac{1062}{20} = 53.1 \qquad \bar{d} = \frac{693}{20} = 34.65$$

$$\text{Mean Vector} = \begin{bmatrix} 26.2 \\ 44.95 \\ 53.1 \\ 34.65 \end{bmatrix}$$

Covariance Matrix

$$\sigma_{aa} = \sigma_a^2 = \frac{13998}{20} - (26.2)^2 = 13.46$$

$$\sigma_{bb} = \sigma_b^2 = \frac{40827}{20} - (44.95)^2 = 20.85$$

$$\sigma_{cc} = \sigma_c^2 = \frac{56676}{20} - (53.1)^2 = 14.19$$

$$\sigma_{dd} = \sigma_d^2 = \frac{24223}{20} - (34.65)^2 = 10.53$$

$$\sigma_{ab} = \frac{23648}{20} - (26.2 \times 44.95) = 4.71$$

$$\sigma_{ac} = \frac{27765}{20} - (26.2 \times 53.1) = -2.97$$

$$\sigma_{ad} = \frac{18141}{20} - (26.2 \times 34.65) = -0.78$$

$$\sigma_{bc} = \frac{47714}{20} - (44.95 \times 53.1) = -1.145$$

$$\sigma_{bd} = \frac{31156}{20} - (44.95 \times 34.65) = 0.2825$$

$$\sigma_{cd} = \frac{36816}{20} - (53.1 \times 34.65) = 0.885$$

$$\text{Covariance Matrix} = \begin{bmatrix} 13.46 & 4.71 & -2.97 & -0.78 \\ 4.71 & 20.85 & -1.145 & 0.2825 \\ -2.97 & -1.145 & 14.19 & 0.885 \\ -0.78 & 0.2825 & 0.885 & 10.53 \end{bmatrix}$$

Correlation Matrix

$$r_{ab} = \frac{\sigma_{ab}}{\sigma_a \sigma_b} = \frac{4.71}{\sqrt{13.46} \times \sqrt{20.85}} = 0.2812$$

$$r_{ac} = \frac{\sigma_{ac}}{\sigma_a \sigma_c} = \frac{-2.97}{\sqrt{13.46} \times \sqrt{14.19}} = -0.2149$$

$$r_{ad} = \frac{\sigma_{ad}}{\sigma_a \sigma_d} = \frac{-0.781}{\sqrt{13.46} \times \sqrt{10.53}} = -0.0655$$

$$r_{bc} = \frac{\sigma_{ad}}{\sigma_a \sigma_d} = \frac{-1.145}{\sqrt{20.85} \times \sqrt{14.19}} = -0.0666$$

$$r_{bd} = \frac{\sigma_{bd}}{\sigma_b \sigma_d} = \frac{0.2825}{\sqrt{20.85} \times \sqrt{10.53}} = 0.0191$$

$$r_{cd} = \frac{\sigma_{cd}}{\sigma_c \sigma_d} = \frac{0.885}{\sqrt{14.19} \times \sqrt{10.53}} = 0.0724$$

$$\text{Correlation Matrix} = \begin{bmatrix} 1 & 0.2812 & -0.2149 & -0.0655 \\ 0.2812 & 1 & -0.0666 & 0.0191 \\ -0.2149 & -0.0666 & 1 & 0.0724 \\ -0.0655 & 0.0191 & 0.0724 & 1 \end{bmatrix}$$

Multiple Correlation

Multiple correlation coefficient between a, b and c

$$R_{c,ab} = \sqrt{\frac{r_{ac}^2 + r_{bc}^2 - 2r_{ac}r_{bc}r_{ab}}{1 - r_{ab}^2}}$$

Here, r_{ac} = -0.2149, r_{ab} = 0.2812 and r_{bc} = -0.0666

$$R_{c,ab} = \sqrt{\frac{(-0.2149)^2 + (-0.0666)^2 - 2(-0.2149 \times 0.2812 \times (-0.0666))}{1 - (0.2812)^2}}$$

$$= 0.21$$

Since R is 0.21, so the variables relationship is weak. Hence we conclude that there is weak positive correlation between a, b and c.

Result

- Mean Vector = $\begin{bmatrix} 26.2 \\ 44.95 \\ 53.1 \\ 34.65 \end{bmatrix}$

- Covariance Matrix = $\begin{bmatrix} 13.46 & 4.71 & -2.97 & -0.78 \\ 4.71 & 20.85 & -1.145 & 0.2825 \\ -2.97 & -1.145 & 14.19 & 0.885 \\ -0.78 & 0.2825 & 0.885 & 10.53 \end{bmatrix}$

- Correlation Matrix = $\begin{bmatrix} 1 & 0.2812 & -0.2149 & -0.0655 \\ 0.2812 & 1 & -0.0666 & 0.0191 \\ -0.2149 & -0.0666 & 1 & 0.0724 \\ -0.0655 & 0.0191 & 0.0724 & 1 \end{bmatrix}$

- **Multiple Correlation:** Since R is 0.21, so the variables relationship is weak. Hence we conclude that there is weak positive correlation between a, b and c.

2. Partial Correlation

Find Partial Correlation from the following information:

Sample No.	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
a	24	27	28	23	25	32	33	24	23	25
b	41	44	44	40	43	49	52	47	44	45
c	55	52	56	50	51	56	48	55	56	55
d	36	39	38	30	37	34	31	32	38	35

Sample No.	***11***	***12***	***13***	***14***	***15***	***16***	***17***	***18***	***19***	***20***
a	24	23	26	23	27	21	34	24	26	32
b	41	47	46	41	43	37	42	56	53	44
c	56	48	56	55	43	57	56	56	53	48
d	38	32	36	36	39	28	33	38	31	32

Procedure

- Mean Vector = $\begin{bmatrix} \bar{a} \\ \bar{b} \\ \bar{c} \\ \bar{d} \end{bmatrix}$

- $\sigma_{ij} = \sigma_i^2 = \dfrac{\sum X_i^2}{n} - \bar{X}_i^2$

- $\sigma_{ij} = \dfrac{\sum X_i X_j}{n} - \bar{X}_i \bar{X}_j$

- $r_{ij} = \dfrac{\sigma_{ij}}{\sigma_i \sigma_j}$

- **Partial Correlation:** the partial correlation of x and z holding y constant is defined as follows:

$$R_{zx,y} = \frac{r_{zx} - r_{zy}r_{xy}}{\sqrt{1-r_{zy}^2}\sqrt{1-r_{xy}^2}}$$

Calculation

Calculating mean vector and covariance matrix

Sample No.	*a*	*b*	*C*	*d*	*a²*	*b²*	*c²*	*d²*	*ab*	*ac*	*ad*	*bc*	*bd*	*cd*
1	24	41	55	36	576	1681	3025	1296	984	1320	864	2255	1476	1980
2	27	44	52	39	729	1936	2704	1521	1188	1404	1053	2288	1716	2028
3	28	44	56	38	784	1936	3136	1444	1232	1568	1064	2464	1672	2128
4	23	40	50	30	529	1600	2500	900	920	1150	690	2000	1200	1500
5	25	43	51	37	625	1849	2601	1369	1075	1275	925	2193	1591	1887
6	32	49	56	34	1024	2401	3136	1156	1568	1792	1088	2744	1666	1904
7	33	52	48	31	1089	2704	2304	961	1716	1584	1023	2496	1612	1488
8	24	47	55	32	576	2209	3025	1024	1128	1320	768	2585	1504	1760
9	23	44	56	38	529	1936	3136	1444	1012	1288	874	2464	1672	2128
10	25	45	55	35	625	2025	3025	1225	1125	1375	875	2475	1575	1925
11	24	41	56	38	576	1681	3136	1444	984	1344	912	2296	1558	2128
12	23	47	48	32	529	2209	2304	1024	1081	1104	736	2256	1504	1536
13	26	46	56	36	676	2116	3136	1296	1196	1456	936	2576	1656	2016
14	23	41	55	36	529	1681	3025	1296	943	1265	828	2255	1476	1980
15	27	43	43	39	729	1849	1849	1521	1161	1161	1053	1849	1677	1677
16	21	37	57	28	441	1369	3249	784	777	1197	588	2109	1036	1596
17	34	42	56	33	1156	1764	3136	1089	1428	1904	1122	2352	1386	1848
18	24	56	56	38	576	3136	3136	1444	1344	1344	912	3136	2128	2128
19	26	53	53	31	676	2809	2809	961	1378	1378	806	2809	1643	1643
20	32	44	48	32	1024	1936	2304	1024	1408	1536	1024	2112	1408	1536
Total	**524**	**899**	**1062**	**693**	**13998**	**40827**	**56676**	**24223**	**23648**	**27765**	**18141**	**47714**	**31156**	**36816**

Mean Vector

$$\bar{a} = \frac{524}{20} = 26.2 \qquad \bar{b} = \frac{899}{20} = 44.95$$

$$\bar{c} = \frac{1062}{20} = 53.1 \qquad \bar{d} = \frac{693}{20} = 34.65$$

$$\text{Mean Vector} = \begin{bmatrix} 26.2 \\ 44.95 \\ 53.1 \\ 34.65 \end{bmatrix}$$

Covariance Matrix

$$\sigma_{aa} = \sigma_a^2 = \frac{13998}{20} - (26.2)^2 = 13.46$$

$$\sigma_{bb} = \sigma_b^2 = \frac{40827}{20} - (44.95)^2 = 20.85$$

$$\sigma_{cc} = \sigma_c^2 = \frac{56676}{20} - (53.1)^2 = 14.19$$

$$\sigma_{dd} = \sigma_d^2 = \frac{24223}{20} - (34.65)^2 = 10.53$$

$$\sigma_{ab} = \frac{23648}{20} - (26.2 \times 44.95) = 4.71$$

$$\sigma_{ac} = \frac{27765}{20} - (26.2 \times 53.1) = -2.97$$

$$\sigma_{ad} = \frac{18141}{20} - (26.2 \times 34.65) = -0.78$$

$$\sigma_{bc} = \frac{47714}{20} - (44.95 \times 53.1) = -1.145$$

$$\sigma_{bd} = \frac{31156}{20} - (44.95 \times 34.65) = 0.2825$$

$$\sigma_{cd} = \frac{36816}{20} - (53.1 \times 34.65) = 0.885$$

$$\text{Covariance Matrix} = \begin{bmatrix} 13.46 & 4.71 & -2.97 & -0.78 \\ 4.71 & 20.85 & -1.145 & 0.2825 \\ -2.97 & -1.145 & 14.19 & 0.885 \\ -0.78 & 0.2825 & 0.885 & 10.53 \end{bmatrix}$$

Correlation Matrix

$$r_{ab} = \frac{\sigma_{ab}}{\sigma_a \sigma_b} = \frac{4.71}{\sqrt{13.46} \times \sqrt{20.85}} = 0.2812$$

$$r_{ac} = \frac{\sigma_{ac}}{\sigma_a \sigma_c} = \frac{-2.97}{\sqrt{13.46} \times \sqrt{14.19}} = -0.2149$$

$$r_{ad} = \frac{\sigma_{ad}}{\sigma_a \sigma_d} = \frac{-0.781}{\sqrt{13.46} \times \sqrt{10.53}} = -0.0655$$

$$r_{bc} = \frac{\sigma_{ad}}{\sigma_a \sigma_d} = \frac{-1.145}{\sqrt{20.85} \times \sqrt{14.19}} = -0.0666$$

$$r_{bd} = \frac{\sigma_{bd}}{\sigma_b \sigma_d} = \frac{0.2825}{\sqrt{20.85} \times \sqrt{10.53}} = 0.0191$$

$$r_{cd} = \frac{\sigma_{cd}}{\sigma_c \sigma_d} = \frac{0.885}{\sqrt{14.19} \times \sqrt{10.53}} = 0.0724$$

$$\text{Correlation Matrix} = \begin{bmatrix} 1 & 0.2812 & -0.2149 & -0.0655 \\ 0.2812 & 1 & -0.0666 & 0.0191 \\ -0.2149 & -0.0666 & 1 & 0.0724 \\ -0.0655 & 0.0191 & 0.0724 & 1 \end{bmatrix}$$

Partial Correlation

Partial correlation coefficient between a and c holding b,

$$R_{ac,b} = \frac{r_{ac} - r_{bc} r_{ab}}{\sqrt{1 - r_{bc}^2}\sqrt{1 - r_{ab}^2}}$$

Here, r_{ac} = -0.2149, r_{ab} = 0.2812 and r_{bc} = -0.0666

$$R_{ca,b} = \frac{-0.2149 - (-0.0666 \times 0.2812)}{\sqrt{1 - (-0.0666)^2}\sqrt{1 - (0.2812)^2}}$$

= -0.20

Since R is -0.20, so the variables relationship is weak. Hence we conclude that there is weak negative correlation of *b* between *a* and *c*.

Result

- Mean Vector = $\begin{bmatrix} 26.2 \\ 44.95 \\ 53.1 \\ 34.65 \end{bmatrix}$

- Covariance Matrix = $\begin{bmatrix} 13.46 & 4.71 & -2.97 & -0.78 \\ 4.71 & 20.85 & -1.145 & 0.2825 \\ -2.97 & -1.145 & 14.19 & 0.885 \\ -0.78 & 0.2825 & 0.885 & 10.53 \end{bmatrix}$

- Correlation Matrix = $\begin{bmatrix} 1 & 0.2812 & -0.2149 & -0.0655 \\ 0.2812 & 1 & -0.0666 & 0.0191 \\ -0.2149 & -0.0666 & 1 & 0.0724 \\ -0.0655 & 0.0191 & 0.0724 & 1 \end{bmatrix}$

- **Partial Correlation:** Since R is -0.20, so the variables relationship is weak. Hence we conclude that there is weak negative correlation of b between a and c.