Unit IV

Multivariate Analysis of Variance

Introduction

Multivariate Analysis of variance is simply an ANOVA with several dependent variables That is to say, ANOVA test for the difference in means between two or more groups, while MANOVA tests for difference in two or more vectors of means

"MANOVA" stands for "Multivariate Analysis of variance". MANOVA in statistics contain multiple dependent variable They help in determining the differences between either two or more than two independent variables It assists in determining this difference simultaneously.

The MANOVA method determines if the dependent variables get significantly affected by changes in the independent variables. ~~of~~ It assists in determining. It also determines the interactions taking place amoungst dependent variables MANOVA finally determines the interaction taking place amoungst independent variables to.

Null hypothesis for MANOVA:

We would test ~~No~~ to see if the vectors of means of the dependent variable is equal for multiple independent groups and our null hypothesis is

$$H_0 : \begin{bmatrix} \bar{x}_{11} \\ \bar{x}_{21} \\ \vdots \\ \bar{x}_{p1} \end{bmatrix} = \begin{bmatrix} \bar{x}_{12} \\ \bar{x}_{22} \\ \vdots \\ \bar{x}_{p2} \end{bmatrix} = \begin{bmatrix} \bar{x}_{13} \\ \bar{x}_{23} \\ \vdots \\ \bar{x}_{p3} \end{bmatrix} = \cdots = \begin{bmatrix} \bar{x}_{1k} \\ \bar{x}_{2k} \\ \vdots \\ \bar{x}_{pk} \end{bmatrix}$$

where p represents the total no of dependent variables for k levels

Test statistics : $\Lambda = \frac{|W|}{|T|} = \frac{|W|}{|B+W|}$

where W and T are determinants of the within and Total sum of squares and cross product matrices.

This means that if between effect B, is very large then A approaches A, if B is small (or) even 0 then A approches 1

Assumptions of MANOVA:

Normal data

The dependent variable should be normally distributed within groups. Overall, the F-test is robust to non-normality. if the non-normality is caused by skewness rather than by outliers. Tests for outliers should be run before performing a MANOVA and outliers should be transformed or removed.

Linearity

MANOVA assumes that there are linear relationship among all pairs of dependent variables, all pairs of covariates and all dependent variable covariate pairs in each all. Therefore when the relationship ~~the~~ deviates from linearity the power of analysis will be compromised.

Homogenity of variances'

Homogenity of variances assumes that the dependent variable exhibit equal level of variance accross the range of predictor variable remember that the error variable is computed (ss error) by adding up the sum of squares within each group. If the variances in the two groups are different from each other, then adding them together is not appropriate and will not yield an estimate of the common group variance. Homoscedastecity can be examined graphically or by means of a no of statistical tools

Homogeneity of variances and covariances

In multivariate designs, with multiple dependent measures, the homogenety of variances assumption described earlier also applies. However since there are multiple dependent variables, it is also required that their inter correlation (covariances) are homogeneous accross the cells of the design

Special case of MANOVA:

Two special cases earise in in MANOVA the inclusion of within-subjects important variables and unequal sample size in cell

Unequal sample size

As in ANOVA. when cells in a factorial MANOVA have different sample sizes, the sum of squares for effect plus error does not equal to total sum of squares. This causes tests of main effects and interaction to be correlated. SPP effects the adjustment for unequal sample size MANOVA.

Within subjects Design

Problem arises if the researcher measures several different dependent variables on different occasions. This situation can be viewed as a within subject independent variable with as many levels as occasions of it can be viewed as separate dependent variables for each occasion

Additional limitations:

Outliers

like ANOVA, MANOVA is extremely sensition to outliers. Outliers may produce either a type I or type II error and give no indication as to which type of ~~data~~ error is occuring in the analysis. There are several programs available to test for univariate and multivariate outliers.

Multicollinearity and singularity

When there is high correlation between dependent variables, one dependent variable~~s~~ becomes a near- linear combination of the other dependent variables. Under such sircumstances it would become statistecally reductant and sus~~res~~pect to includes both combinations.

Multivariate Hotelling T^2 speciall issues concerning hotelling T^2 MANOVA.

In statistical T^2 MANOVA is multivariate distn proportional to the F- distn and arises importantly as a distn of a set of statistics which are natural generalisation of the statistics

Underlying student t-distn, Hotelling T^2 statistics is a generalisation of student t statistic that is used in multivariate hypothesis testing

If a r.v 'x' has hotelling T^2 distn, $x \sim T^2_{p,m}$ then $\frac{m-p+1}{mp}$, $x \sim F_{p,m-p+1}$ where, $F_{p,m-p+1}$ is the F distn with parameter p and $m-p+1$

Test versions:- One sample: The multivariate vector means for a group equals a hypothetical vector of means

Two sample The multivariate vector means for a two group are equal

Two sample T^2 Hotellings:-

If you know how to run a two sample t-test, then you know how to run a two sample

Hotelling T^2:

The basic steps are the same although you will use a different formula to calculate the T^2 value and you will use a different table (F table) to find the critical value

Hotelling T^2 has several advantages over the t-test:

The type I error rate is well controlled

The relationship b/w multivariate variables in taken into account.

It can generate an overall conclusion even if multiple (single) t-test are inconsistent while a t test will tell you which variable differ b/w groups. Hotelling T^2 summarizes the b/w group difference.

Test hypothesis: H_0: The two sample are from popln with the same multivariate means

H_1: The two samples are from popln with different multivariate means

Like the t-test you want to find a value for T^2 (in this case for T^2) and compare it with F table value

If the calculated values is greater than the table values You can reject the null hypothesis (H_0)

For ease of this calculation Hotelling T^2 is first transformed into an F-statistics.

$$F = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)} T^2 \sim F_{p,\, n_1+n_2-p-1}$$

where n_1 and n_2 sample sizes

p : number of variable measured

n_1+n_2-p-1 = degrees of freedom

Reject the null hypothesis (at a choosen significant level). If the calculated value is greater than the F table critical value

Rejecting the null hypothesis means that atleast one of the parameters is significantly different.

Note:

Testing the multiple dependent variables accomplist by creating new dependent variables that maximize group differences these artificial dependent variable are linear combination of the measured dependent variable.

MANOVA comparing two groups:-

A T^2 statistic for testing the equality of vecto means from two multivariate popln can be developed by analogy with the univariate procedure. This T^2 statistic i appropriate for compareing responses from one set of experimental ~~select~~ setting (population 1) with independent responses I/p from ~~one~~ another set of experimental setting (population 2)

The comparison can be without explicity controlling for unit to unit variability as in the paired comparison case.

If possible, the experimental units should be randomly assigned to the sets of experimental conditions. Randomisation to some extent mitigate the effect at unit-to-unit variability is a sub-sequent comparison of treatments. Although some precision is lost relative to paired comparisons the inferences in the two population case are ordinarily, applicable to a more general collection of experimental units simply because unit homogeneity is not required.

Consider a random sample of size n_1 from population 1 and a sample of size n_2 from population 2

Principal compo~~und~~nent analysis.

A PCA is concerned with explaining the variance-covariance structure of a set of variables. Its general objectives are i) Data reduction ii) Interpretation

Although p components are required to reproduc the total system variability often much of this variability can be accounted for by a small number k of the principal component. If so there is as information in the k components as there is in the original pvariables The k principal component can then replace the initial P variables and original data set. Consisting of n measurements on p variables is reduced to a data set consisting of n measurements on k principal components.

An analysis of Principal Component often reveals relationship that were not previously suspected and there by allows interpretation that would not ordinarily result Good example of this is provided by stock market

Analysis of pc are more of a means to an end rather then themselves because they frequently serve as intermediate steps in much larger investigations pc may be input to a multiple regression (or) cluster analysis moreover pc are factoring of the covariance matrix for the factor analysis

Uses:-

PCA is a technique that is useful for the Compression and classification of data The purpose is to reduce the dimensionality of a data set (sample) by finding a new set of variable smaller than the original set of variables, that nonetheless returns most of the sample information.

Computation of Principal Components in the population:-

Algebrically, principal components are particular linear combinations of the p random variables $x_1, x_2 \dots x_p$

Geometrically, these linear combinations represent the selection of a new co-ordinate system obtained by relating the original system with $x_1, x_2, \dots x_p$ as the co-ordinates axes. The new axes represent the directions with maximum variability and a simpler and more persimonious ~~b~~ description of the covariance structure

Principal components depends solely on the covariance matrix Σ (on the correlation matrix p) $x_1, x_2 \dots x_p$ Their development does not require a multivariate normal assumption

Let the random vector $x' = [x_1, x_2 \quad x_p]$ have the covariance matrix Σ with eigen values $\lambda_1 > \lambda_2 > \lambda_3 \dots > \lambda_p \geq 0$

Consider the linear combination

$$\left.\begin{aligned} Y_1 &= a_1' X = a_{11} X_1 + a_{12} X_2 + \dots + a_{1p} X_p \\ Y_2 &= a_2' X = a_{21} X_1 + a_{22} X_2 + \dots + a_{2p} X_p \\ Y_p &= a_p' X = a_{p1} X_1 + a_{p2} X_2 + \dots + a_{pp} X_p \end{aligned}\right\} \text{— (1)}$$

Then we obtain $V(Y_1) = a_i' \Sigma a_i$ $i = 1, 2 \dots p$ —— ②

$cov(Y_i, Y_k) = a_i' \Sigma a_{k}$, $i = 1, 2 \dots p$ —— ③

This principal components are these uncorrelated linea combinations $Y_1, Y_2 \dots Y_p$ whose variance in ② are as large as possible.

Thus 1st principal component is the linear combination with max variance (ie) it maximizes $Var(Y_i) = a_i' \Sigma a_i$ It is clear that $var(Y_i) = a_i' \Sigma a_i$ can be increased by multiplying any a_i by some constant We therefore define

First principal component = linear combination $a_1' x$ that maxim $V(a_1' x)$ subject to $a_1' a_1 = 1$ & $cov(a_1' X, a_1' X) = 1$

Second principal component = linear combination $a_2' x$ that maximiz $V(a_2' x)$ subject to $a_2' a_2 = 1$ & $Covar(a_1' x, a_2' x) = 0$

at the ith step ith principal component = linear Combinatio $a_i' x$ that maximizes $V(a_i' x)$ subject to $a_i' a_i = 1$ and $cov(a_k' x_k, a_i' x_i) = 0$ for $k < i$

Extraction of principal components.

There is always the question of how many components to retain There is no definite answer to this ques

Things to consider include:

1) The amount of total sample variance explained

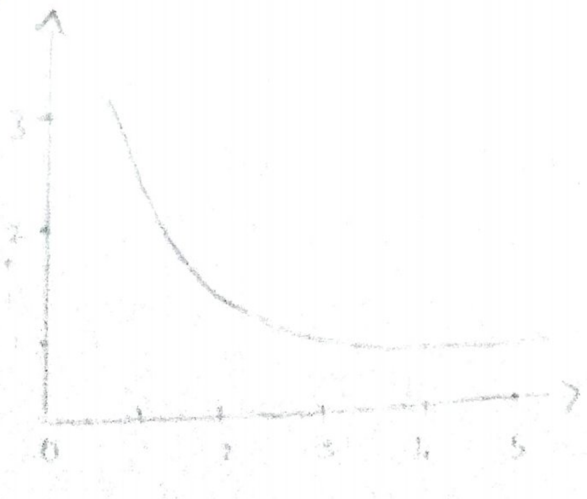
2) The relative sizes of the eigen values (the variance of the sample components) and

3) The subject matter interpretations of the components

In addition, a component associated with an eigen value near zero, deemed to be unimportant, may indicate an unsuspected linear dependency in the data.

A useful visual aid to determine an appropriate no of principal components is a "scree plot" with the eigen values ordered from larger to smaler, a scree plot is a plot of λ_i, the magnitude of eigen values versus i, its number

To determine the appropriate number of components, we look for an the scree plot. The no. of components is taken to be the point at which the remaining eigen values are relatively small and all about the same size. The following figure shows a scree plot for a situation with prune



An elbow occurs in the plot in the figure ~~af~~ at about i=3. That is the eigen values after $\hat{\lambda}_2$ are all relatively small and about the same size. In this case, it appears without any other evidence that two sample PC ~~offar~~ effectively summarise the total sample variance

Interpretation of principal components

The sample PC have several interpretations. First suppose the underlying distn of X is nearly $N_p(\mu,\Sigma)$. Then the sample principal components $\hat{y}_i = \hat{e}_i (x-\bar{x})$ are realizations of popln PC $y_i = e_i'(x-\mu)$, which have an $N_p(0,\Lambda)$ distn. The diagonal matrix A has entries $\lambda_1, \lambda_2 \dots \lambda_p$ and (λ_i, e_i) are the eigen value - eigen vector pairs of Σ.

Also from the sample values x_i. we can approximate μ by $\bar{x}$ and Σ by s. If s is +ve definite. the contour consisting of all px vectors x satisfying

$$(x-\bar{x})'s^{-1}(x-\bar{x}) = c^2 \quad (17)$$

estimates the constant density contours $(x-\mu)'\Sigma^{-1}(x-\mu) = c^2$ of the underlying normal density. The apporoximate countours can be drawn on the scatter plot to indicate the normal distn that generate the data. The normality assumption is useful for the inference procedures discussed but it is not required for the development of the properties of estimates the constant density cont

the sample pc summarized.

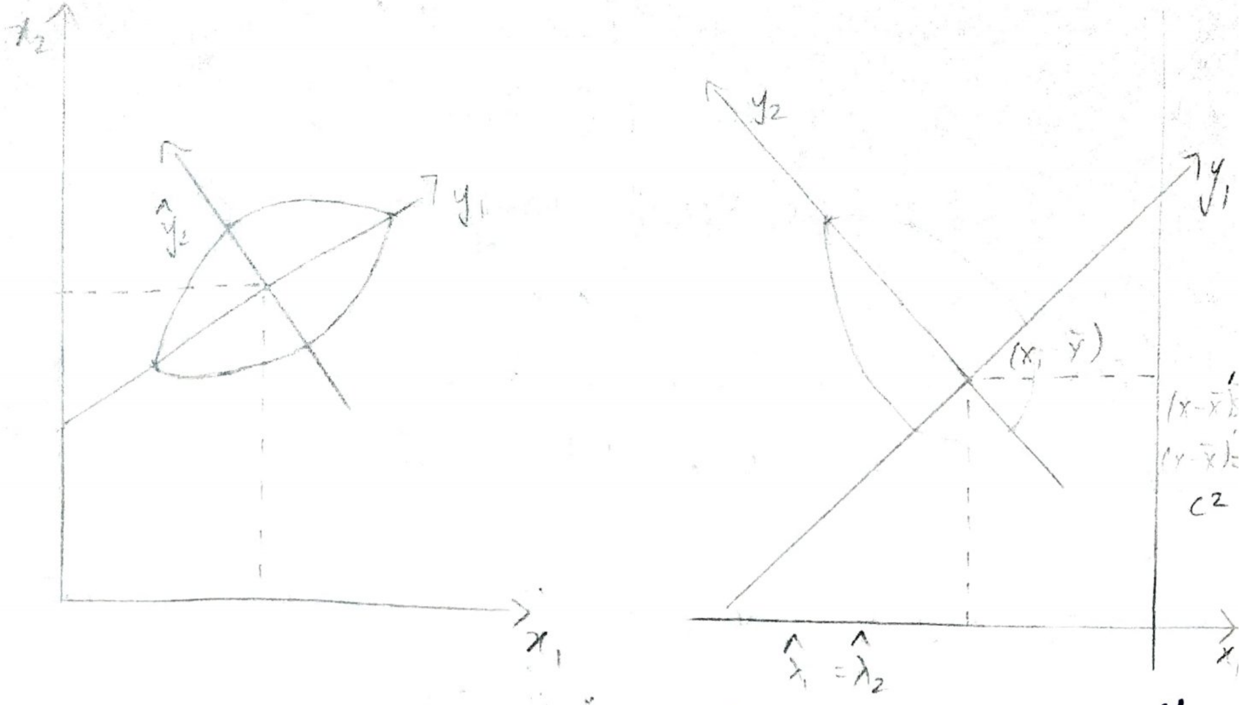
Even when the normal assumption is suspect and the scatter plot may depart somewhat from an elliptical pattern we can still extract the eigen values from s and obtain the sample pc Geometrically, the data may be plotted as n points in p.space. The data can then be expressed in the new co-ordinates; which coincide with the axes of the contour ①. Now ① defines a hyper~~bola~~ elliptic said that is centered at $\bar{x}$ and whose axes are given by the eigen vectors of s^{-1} or, equivalently of s. The length of these hyper~~bola~~ elliptic said axes are proportional to $\sqrt{\hat{\lambda}_i}$ $i=1,2..p$ where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p \geq 0$ are the ~~given~~ eigen values of s

Because $\hat{e}_i$ has length i, the absolute value of the ith principal component, $|\hat{y}_i| = |\hat{e}_i'(x - \bar{x})|$, gives the length of the projection of the vector $(x - \bar{x})$ on the unit vector $\hat{e}_i$. Thus the sample pcs, $\hat{y}_i = \hat{e}_i'(x - \bar{x})$ $i=1,2..p$ lie along the axes of the hyper~~bola~~ elliptic said and their absolute values are the length of the projection of $x - \bar{x}$ in the directions of the axes $\hat{e}_i$. Consequently, the sample pcs can be views as the results of translating the origin of the original coordinate

System to $\bar{x}$ and then rotating the coordinate axes untill they pass through the scatter in the directions of maximum variance.

The geometrical interpretation of the sample pcs for $p=2$ shows an ellipse of constant distance, centered at $\bar{x}$ with $\hat{\lambda}_1 > \hat{\lambda}_2$. The sample pcs are well determined. They lie along the axes of the ellipe in two perpendicular directions, including the direction of the original coordinate axes. similarly, the sample pcs can lie in 'any two perpendicular directions Including those of the original coordinate axes when the contours of constant distances are not unequally determined and can lie in any two perpendicular directions, including the direction of the original co-ordinate axes Similarly the sample pcs can lie in the any two perpendicular directions including those of the original co-ordinate axes when the contours of constant distance are nearly circular or equivalently, when the eigen values of s are nearly equal. The sample variation is homogeneous in all directions It is then not possible to represent data well in fewer than p dimensions

Sample pcs and ellipse of constant distance.



If the last few eigen values $\hat{\lambda}_i$ are sufficiently small such that the variation in the corresponding $\hat{e}_i$ directions is negligible, the last few sample pcs can often be ignored and the data can be adequately approximated by their representations in the space of the retained components.

Exactly. ~~supplement~~

Computation of the MLE of the principal components

There are several ways of computing the characteristic roots and characteristic vector [pc] of a matrix Σ or $\hat{\Sigma}$. We shall indicate some of them.

One method ~~for~~ involves expanding the ~~dimens~~ determinantal eqn, $0 = |\Sigma - \lambda I|$ — ① and solving the resulting pth degree eqn in λ.

Example, [By Newton's method or the second method] for the roots $\lambda_1 > \lambda_2 > \lambda_3 \cdots \lambda_p$. Then $\Sigma - \lambda I$ is of rank p-1 and a solution of $(\Sigma - \lambda_i I)\beta = 0$ can be obtained by taking

$B_j^{(1)}$ as the cofactor of the element in the 1st (or any other fixed) column and j row of $\Sigma - \lambda_i I$.

The 2nd method iterates using the eqn for a characteristic vector $\Sigma . x = \lambda x$ —— ②

We have writen the eqn for the popln. Let $x_{(0)}$ be any vector not orthogonal to the 1st characteristic vector, and define

$$x_{(i)} = \Sigma y_{i-1}, \quad y_{(i)} = \frac{x_{(i)}}{\sqrt{x'_{(i)} x_{(i)}}} . \; i = 0, 1 \ldots \quad \text{—— ③}$$

It can be shown that

$$\lim_{i \to \infty} y_{(i)} = \pm \beta^{(1)} \quad \lim_{i \to \infty} x'_{(i)} = \lambda_1^2 \quad \text{—— ④}$$

The rate of convergence depends on the ratio λ_2 / λ_1; the closer this ratio is to 1, the slower the convergence.

To find the 2nd root and vector define

$$\Sigma_2 = \Sigma - \lambda_1 \beta^{(1)} \beta^{(1)'} \quad \text{—— ⑤}$$

Then $\Sigma_2 \beta^{(i)} = \Sigma \beta^{(i)} - \lambda_1 B^{(1)'} \beta^{(i)} = \Sigma \beta^{(i)} - \lambda_i \beta^{(i)}$ —— ⑥

if $i \neq 1$ and $\Sigma_2 \beta^{(1)} = 0$ —— ⑦

Thus, λ_2 is ~~true~~ the largest root of Σ_2 and $\beta^{(2)}$ is the corresponding vector. The iteration process is now applied to Σ_2 to find λ_2 and $\beta^{(2)}$. Defining $\Sigma_3 = \Sigma_2 - \lambda_2 \beta^{(2)} \beta^{(2)'}$, we can find λ_3 and $\beta^{(3)}$ and so forth.

There are several ways in which the labour of the iteration procedure may be reduced. One is to raise Σ

a power before proceeding with the iteration. thus one can use Σ^2 defining,

$$x_{(i)} = \Sigma^2 y_{(i-1)}, \quad y_{(i)} = \frac{x_{(i)}}{\sqrt{x'_{(i)} x_{(i)}}} \quad , i = 0, 1, \ldots \text{ — } \textcircled{8}$$

This procedure will give twice as rapid convergence as the use of (3). Using $\Sigma^4 = \Sigma^2 \Sigma^2$ will lead to convergence four times as rapid, and so on. It should be noted that since, Σ^2 is symmetric, there are only $p(p+1)/2$ elements to be found.

Efficient computation, however, uses other methods. One method is the QR or QL algorithm. Let $\Sigma_0 = \Sigma$, define recuratatively the orthogonal Q_i and lower triangular L_i by $\Sigma_i = Q_i L_i$ and $\Sigma_{i+1} = L_i Q_i \; (= Q_i' \Sigma_i Q_i)$ $i = 1, 2 \ldots$ (The vam -schmidt orthogonalization is a way of finding Q_i and L_i. the QR method replaces a lower triangular matrix.