INTRODUCTION

1. In this thesis on "Certain Problems in Multivariate Analysis", are included five chapters divided in two parts, the first part containing four chapters on multicollinearity of means in multivariate normal populations and the second part containing only one chapter on multivariate power-series distributions.

2. Just as χ^2 distribution is important in the univariate analysis for various types of tests, so is the Wishart distribution in the multivariate analysis. In univariate analysis, Aitken (2,3), Cochran (14), Craig (15), Hotelling (30), Lancaster (46), Madow (48), Matérn (51), Ogawa (56,57) and George & Graybill (25) have established the conditions for the quadratic forms : $\mathbf{x}'A\mathbf{x}$ to be distributed independently and/or to obey χ^2 distribution and Laha (45) has obtained the conditions for the independence of two second-degree polynomials in x's. In the same manner, for the multivariate case, considering the forms of the type: $XAX' + (LX'+XL')/2+C$, we have established in this thesis the conditions for the forms to be distributed independently and/or to obey Wishart distribution. Splitting XX' (sum of squares and products matrix of observations) into $\sum_{i=1}^{k} XA_iX'$, we have established the sum of squares and products matrices (S. P. M.) due to the null hypothesis and due to the error. It can

be easily shown from the conditions established, that S.P.M. due to error (XBX', say) and that due to hypothesis (XAX', say) are independently distributed and also that they are separately distributed as Wishart under certain conditions. Hence this gives us the statistics as the proper functions of the elements of XAX' $(XBX')^{-1}$ required for testing the given null hypothesis. Also, we may note that various persons like Anderson (5), Bartlett (10,11), Bose (12,13), Fisher (19), Hotelling (28,29), Hsu (31,32), Kendall (35,36), Kullback (42,43,44), Mahalanobis (49), Narain (53,54), Pillai (61,62), Rao (68,69,70,71), Roy (75,76,77,79,80) and Wilks (85,86,87), from various points of view, arrive at various statistics for testing the same null hypothesis, but the important common property about all those statistics is that they all depend upon the proper functions of the elements of (XAX') $(XBX')^{-1}$. For example, we may consider the problem of testing the equality of mean vectors for k different populations. Let S_2 and S_1 be the respective S.P.M. due to error and that due to the null hypothesis. Then Wilks (85), by considering the likelihood function, arrives at the Λ-criterion, which is constant times the geometric mean of the characteristic roots (or simply roots) of $S_2(S_1 + S_2)^{-1}$. Hotelling (29), extending the idea of generalised student-t, Hotelling T^2 (28), arrives at the Hotelling T-criterion which is constant times the arithmetic mean of the roots of $S_1 S_2^{-1}$. But Pillai (62), visualising the presence of the above two means in those methods, gives the H-criterion which is the constant times the harmonic mean of the roots of $(I + S_1 S_2^{-1})$, while Roy (13), by consideration of heuristic method of test construction (75), arrives at the test - criterion of maximum root of $S_1 S_2^{-1}$. Here also we may note that all the four criteria mentioned above satisfy the condition that they are all proper - functions of the elements of $S_1 S_2^{-1}$.

In this thesis, we give the test-criteria for testing the multicollinearity of means (and departures ffom it). Multicollinearity of Means defined by S.N.Roy (79) is of two types; namely (i) the matrix of means of the first p-variates is a constant matrix times, the matrix of means of the remaining q-variates, the constant matrix factor being equal to the regression matrix of the p-set on q-set, whatever this regression matrix might be. We define this as the first kind of multicollinearity of means, and (ii) in the second kind of multicollinearity of means, the matrix of means of the first p-variates is <u>a constant (and given) matrix times</u> the matrix of means of the remaining q-variates. This also belongs to linear hypothesis in multivariate analysis of variance of means and the test-criteria can be given similar to the testing of means in k-multivariate normal populations. The first kind can be successfully handled by splitting S.P.M. into various components of which only two are considered in this thesis-one due to hypothesis and the other due to error. When there are two multivariate populations, Rao (68) has considered this case by naming it as the problem of increased distance. He gave R or U statistics for this purpose. We name them for the above general problem as generalised R (which is the harmonic mean of roots) and generalised U (which is the arithmetic mean of roots) statistics (38). Roy's criterion of maximum root is also considered. The first criterion, namely, the geometric mean of roots has already been considered by Bartlett (11), Hsu (33), Lawly (47), Rao (68) and Wilks(85). Narain (54) has studied the unbiasedness of R or U statistics given by Rao. Here we study the test-criteria: genefalised R, generalised U and maximum root λ_p in detail bringing out their various properties. Also the simultaneous confidence bounds on the various types are also considered.

3. <u>Chapter One</u> deals with the forms of the type: $XAX' + (LX' + XL')/2 + C$ with their conditions for independence

and for obeying Wishart distribution. Also their applications in testing the various statistical hypotheses have been given and the test criteria which are invariant under certain non-singular linear transformations of variates have been derived.

Chapter Two deals with the problem of obtaining the distributions of the statistics for the multicollinearity of means of first kind, and also a numerical example is given to illustrate the technique of computation of these statistics.

Chapter Three deals with the property of the monotonic lower bound of the test procedure (depending on the maximum root λ_p).

Chapter Four deals with the simultaneous confidence bounds on the parameters connected with multicollinearity of means and the regression like parameters.

4. Noack (55) has defined a univariate power-series distribution as $\Pr(\xi=x) = a_x z^x/f(z)$, for $x=0,1,2,\ldots$ where $a_x z^x \geqslant 0$, a_x is a function of x or constant, $f(z) = \sum_{0}^{\infty} a_x z^x$ is convergent for some $|z| \leq r$. We (41) extend this definition to multivariate power-series distribution as

$$\Pr(\xi_1=x_1,\ldots,\xi_k=x_k)= a_{x_1,\ldots x_k}\, z_1^{x_1}\ldots z_k^{x_k}/f(z_1,\ldots,z_k)$$

where $a_{x_1,\ldots x_k} z_1^{x_1} \ldots z_k^{x_k} \geqslant 0$, $a_{x_1,\ldots x_k}$ is a function of $x_1,\ldots, x_k$ or constant, and $f(z_1, \ldots, z_k) = \sum_{x_1,\ldots,x_k} a_{x_1,\ldots,x_k} z_1^{x_1} \ldots z_k^{x_k}$ is convergent.

Certain recurrence relations for the factorial-cumulants, cumulants in multivariate as well as univariate cases are given and it has been shown that any power-series distribution is uniquely determined from its first two moments (i.e. means, variances and covariances). For multivariate and univariate cases, certain illustrative examples (like multinomial, negative multinomial, multivariate Poisson, generalised Poisson etc) are given to illustrate these properties of power-series distribution.

All these properties are easily extended to truncated power-series distributions.

*****