Minimum Ellipsoids

By Donald D. Fisher

1. Introduction. A well-known statistical problem is to determine an ellipsoid R_S^e in E_n which contains a certain fraction of the points from a set S. Here we use the word "contains" to mean that a point $p_i \in S$ is either in the interior or on the surface of the ellipsoid R_S^e. Although the determination of R_S^e is easy computationally, the determination of the ellipsoid of minimum volume, say, which contains all the points of S is quite difficult. In this note we give a method for determining ellipsoids satisfying a certain minimum property and compare these with ones obtained by statistical methods.

2. Statistical Test Region. If the points of S tend to be correlated, an ellipsoid is an appropriate regular test region. One statistical test region is set up by making use of the F-distribution and multivariate analysis. We assume that S has a multivariate normal distribution with an unknown mean μ which we estimate by taking the mean of the points in S. Furthermore, we assume the unknown covariance matrix associated with S may be estimated by the (symmetric) matrix

$$(2.1)\qquad \mathcal{S} = \frac{1}{N-1}\begin{bmatrix} \sum (x_i - \bar{x})^2 & \sum (x_i - \bar{x})(y_i - \bar{y}) & \sum (x_i - \bar{x})(z_i - \bar{z}) & \cdots \\ \cdot & \sum (y_i - \bar{y})^2 & \sum (y_i - \bar{y})(z_i - \bar{z}) & \cdots \\ \cdot & \cdot & \sum (z_i - \bar{z})^2 & \cdots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{bmatrix}$$

where N is the number of points in S [3].

We assume a point q of unknown lineage has the same distribution as the points $p_i \in S$. Let α be the fraction of allowed false positives, i.e., the allowed fraction of unknowns which do not lie in R_S^e, but by some other test are found to be a member of S. The unknown point q belongs to S with probability $1 - \alpha$ if

$$(2.2)\qquad (q - \bar{w})^T \mathcal{S}^{-1} (q - \bar{w}) \leqq \frac{N_2(N_1 + 1)(N_1 - 1)}{N_1(N_1 - N_2)} F^{\alpha}_{N_2, N_1 - N_2},$$

where N_1, N_2 are the number of degrees of freedom of the denominator and numerator, respectively, associated with the F-distribution and $\bar{w}$ is the computed mean of the points $p_i \in S$. Inequality (2.2) derives from Hotelling's T^2 statistic in multivariate analysis [1]. Geometrically, (2.2) defines the interior and boundary of an ellipsoid with the mean $\bar{w}$ as center. For fixed N_1, N_2, $F^{\alpha}_{N_2,N_1-N_2}$ increases as α decreases, consequently, the size of the ellipsoid increases as α decreases. For a given F the volume of R_S^e is

Received August 23, 1963. This study was supported by the Office of Naval Research under contract Nonr-225(37) (NR-044-211) and by the National Science Foundation under contract NSF-GP 948.

(2.3) $$V_S^{\,e} = \pi \left\{N_2(N_1 + 1)(N_1 - 1)F^{\alpha}_{N_2,N_1-N_2}\left[N_1(N_1 - N_2)\prod_j \lambda_j\right]^{-1}\right\}^{1/2}.$$

In E_3 the eigenvalues λ_j of $\mathcal{S}$ may be computed directly by trigonometric relations [6]. In general a certain number of points from S will lie outside the above ellipsoid. As F increases more and more points of S will be contained in $R_S^{\,e}$.

3. "Minimum" Ellipsoid Region. C. Loewner has proved that there is a unique ellipsoid of minimum volume which contains all points of S. The problem may be stated as a minimum problem, i.e., given $p_i \in S$ with mean $\bar{w}$, determine the matrix $\mathfrak{I}$ for which

(3.1) $$\det\{\mathfrak{I}^{-1}\} \equiv \prod_j \lambda_j$$

is a minimum subject to the constraints

(3.2) $$(p_i - \bar{w})^T\mathfrak{I}(p_i - \bar{w}) \leqq 1.$$

This formulation possesses two disadvantages [4]: (i) The volume becomes insensitive to change in all λ_j if any $\lambda_l \to 0$; (ii) $\det\{\mathfrak{I}^{-1}\}$ is not a convex function of $\mathfrak{I}$ and consequently the numerical computations for the minimum are difficult to perform.

The problem may be recast in a form which avoids these difficulties. We determine the matrix $\mathfrak{I}$ for which

(3.3) $$\varphi(\mathfrak{I}) \equiv \text{trace}\{\mathfrak{I}^{-1}\} = \sum(\lambda_j)^{-1} = \sum r_j^2$$

is a minimum subject to the constraints

(3.4) $$(p_i - \bar{w})^T\mathfrak{I}(p_i - \bar{w}) \leqq 1, \qquad t_{ii} \geqq \sum_{j \neq i} | t_{ij} |,$$

where r_j is the length of the jth semiaxis and t_{ij} is the i, j element of $\mathfrak{I}$. In [4] it is shown that trace$\{\mathfrak{I}^{-1}\}$ is a strictly convex function of $\mathfrak{I}$ and that $\mathfrak{I}^{-1}$ is a convex function of $\mathfrak{I}$. This measure of size is unique and, furthermore, there is a method (gradient projection (GP) method) for determining the associated ellipsoid $R_S^{\,v}$ numerically [5], [7].

Let $\mathfrak{I}^{-1} = \mathfrak{U} = (u_{ij})$. The gradient of $\varphi(\mathfrak{I})$ is given by

(3.5) $$\frac{\partial \varphi}{\partial t_{ij}} = -2u_i^{\,T}u_j, \qquad i \neq j,$$ $$\frac{\partial \varphi}{\partial t_{ii}} = -u_i^{\,T}u_i,$$

where u_j is the jth column of $\mathfrak{U}$.

4. Numerical Results. The ellipsoids $R_S^{\,e}$ and $R_S^{\,v}$ (indicated in the tables by R^v_{Sk} and R^e_{Sk}, $k = 1, \cdots, 11$) were determined for 11 sets of points in E_3 with 150 points in each set (denoted by S^k_{150}) and for 11 subsets of the above sets with 25 points in each subset (denoted by S^k_{25}). The data originated from vectorcardiographic studies [2] of subjects which had been assigned to specific sets based on independent tests. The ellipsoids R_S^v were determined by 6 variables and either 153 constraints or 28 constraints, respectively. Most of the constraints turn out to be inactive since a unique ellipsoid in E_3 is determined by 3 "independent" points.

The relation of the size of R_S^e to the size of R_S^v highlights a characteristic of the F test. The larger the sample size the better the estimate of the various statistical quantities, hence the sharper the F test. For a sample size of 25 the F test allows for considerably more scatter than for a sample size of 150. Note that R_S^v contains all points of S whereas, by assumption, R_S^e contains only the fraction $1 - \alpha$ of the points of S. The actual number of points exterior to R_S^e is given in Table 4. Comparisons in Tables 1, $\cdots$, 4 are based on an F value for $\alpha = 0.05$. For the smaller sample size both the volume and sum of squares of the semiaxes for R_S^v are smaller

TABLE 1

$r_{\max}^k / r_{\min}^k$

k	S_{25}^k		S_{150}^k	
	R_{Sk}^v	R_{Sk}^e, $\alpha = 0.05$	R_{Sk}^v	R_{Sk}^e, $\alpha = 0.05$
1	1.36	2.46	1.52	2.37
2	1.78	2.03	2.23	2.14
3	2.49	2.37	3.31	2.02
4	2.70	1.83	2.59	2.09
5	1.31	1.74	1.82	2.09
6	1.84	1.57	1.60	1.64
7	1.35	1.52	1.50	1.32
8	1.68	1.53	1.31	1.40
9	1.64	2.43	1.84	1.71
10	1.64	2.13	1.58	1.66
11	2.52	3.07	1.72	2.04

TABLE 2

$\prod_i r_i^k$

k	S_{25}^k		S_{150}^k	
	R_{Sk}^v	R_{Sk}^e, $\alpha = 0.05$	R_{Sk}^v	R_{Sk}^e, $\alpha = 0.05$
1	0.00225	0.00199	0.00212	0.00088
2	0.00864	0.01575	0.02220	0.00957
3	0.03387	0.04217	0.09718	0.03374
4	0.06468	0.07890	0.16194	0.07966
5	0.38915	0.52527	0.91020	0.59686
6	0.78645	1.37867	1.66261	1.22425
7	0.92816	1.64883	2.56710	1.89610
8	1.46047	2.40659	3.76361	2.52306
9	0.61088	0.98923	3.61434	1.47622
10	0.30282	0.51189	2.49975	0.81512
11	0.08671	0.20482	1.59846	0.47684

TABLE 3

$\sum_i (r_i^k)^2$

k	S_{25}^k		S_{150}^k	
	R_{Sk}^v	R_{Sk}^e, $\alpha = 0.05$	R_{Sk}^v	R_{Sk}^e, $\alpha = 0.05$
1	0.05347	0.06169	0.05248	0.03516
2	0.14026	0.22163	0.29139	0.16590
3	0.41744	0.45623	0.96584	0.36674
4	0.65938	0.61913	1.16601	0.65657
5	1.64423	2.17246	3.15794	2.51408
6	2.88333	3.96950	4.52434	3.72471
7	2.94184	4.43341	5.94049	4.71656
8	4.21288	5.73125	7.45132	5.76568
9	2.34140	3.78881	7.96869	4.27550
10	1.46843	2.29726	5.93542	2.84975
11	0.75918	1.48931	4.51029	2.14763

TABLE 4

Number of Points Exterior to R_s^e

k	S_{25}^k	S_{150}^k
1	1	10
2	0	6
3	1	8
4	1	12
5	0	6
6	0	8
7	0	5
8	0	9
9	1	11
10	0	9
11	0	10

in most cases than those for R_S^e. However, the volume and sum of squares of the semiaxes for R_S^v are greater than those for R_S^e for the larger sample size.

Table 1 summarizes the ratios of the maximum semiaxis $r_{\max}$ to the minimum semiaxis $r_{\min}$. Table 2 gives the volumes (apart from a multiplicative constant) of the ellipsoids and Table 3 gives the sum of squares of the semiaxes. Note that even though the volume of R_S^v is greater than the corresponding volume R_S^e for the larger sample, the ratio $r_{\max}/r_{\min}$ for R_S^v is not always greater than the ratio $r_{\max}/r_{\min}$ for R_S^e. This is accounted for by the fact that R_S^v must orient itself differently from R_S^e in order to include an extreme point, whereas an extreme point does not influence R_S^e significantly.

GP computing time on the 7090 for 8 cases with 6 bounds and 28 active constraints was 0.8 minutes. For 12 cases with 6 bounds and 153 active constraints the total 7090 time was 2.7 minutes.

5. Acknowledgment. The author is indebted to Professors R. Miller and J. B. Rosen for many helpful ideas.

Stanford Computation Center
Stanford University
Stanford, California

1. T. W. ANDERSON, *An Introduction to Statistical Analysis*, John Wiley and Sons, Inc., New York, 1957, p. 123, problem 5.
2. G. E. FORSYTHE, J. VON DER GROEBEN & J. G. TOOLE, "Vectorcardiographic diagnosis with the aid of Algol", *Comm. ACM*, v. 5, no. 2, 1962, p. 118–123.
3. D. B. OWEN, *Handbook of Statistical Tables*, Addison-Wesley, Reading, Mass., 1962.
4. J. B. ROSEN, *Pattern Separation by Convex Programming*, Stanford University Applied Mathematics and Statistics Laboratories Report No. 30, 1963.
J. B. ROSEN, "The gradient projection method for nonlinear programming", Part I, *J. Soc. Indust. Appl. Math.*, v. 8, 1960, p. 181–2 Part II, ibid. v. 9, 1961, p. 514–532.
6. O. K. SMITH, "Eigenvalues of a symmetric 3 × 3 matrix", *Comm. ACM*, v. 4, no. 4, 1961, p. 168.
7. Gradient Projection 7090 program. SHARE distribution No. 1399.