7

Tangent Space Inference

In this chapter we outline some simple approaches to inference based on tangent space approximations, which are valid in datasets with small variability in shape. In particular, we discuss one and two sample Hotelling's T^2 tests for mean shape. We consider the isotropic model as a special case, where more powerful procedures can be obtained (if the model is appropriate). Finally, we consider other multivariate techniques which work directly on shape coordinates, with Euclidean approximations to the non-Euclidean nature of the shape space.

7.1 Tangent Space Inference for Shapes

The tangent space to shape space is a linear approximation to shape space. A practical approach to analysis is to use the Procrustes tangent space coordinates if the data are concentrated and then perform standard multivariate analysis in this linear space. One choice of tangent co-ordinates are the Procrustes residuals (discussed in Chapter 3) or full/partial Procrustes tangent co-ordinates (Chapter 4 of Dryden and Mardia, 1998). This tangent based approach is an approximation to inference using a general multivariate normal model for the landmarks.

7.1.1 One sample Hotelling's T^2 test

Consider carrying out a test on a mean shape of a single population and whether or not the mean shape has a particular special shape $[\mu_0]$, i.e. test between

Let $X_1, \ldots, X_n$ be a random sample of configurations with partial Procrustes tangent coordinates (with pole $\hat{\mu}$ from the full Procrustes mean with unit size) given from Equation (4.35) in Dryden and Mardia (1998) by $v_1, \ldots, v_n$ where

$$v_i = (I_{km-m} - \mathrm{vec}(\hat{\mu})\mathrm{vec}(\hat{\mu})^{\mathrm{T}})\mathrm{vec}(X_i^P/\|X_i^P\|). \quad (7.1)$$

Let the tangent coordinates of μ_0 be γ_0 where

$$\gamma_0 = (I_{km-m} - \mathrm{vec}(\hat{\mu})\mathrm{vec}(\hat{\mu})^{\mathrm{T}})\mathrm{vec}(\mu_0^P/\|\mu_0^P\|)$$

and μ_0^P is the Procrustes fit of μ_0 onto $\hat{\mu}$. Since the dimension of the tangent space is $M = km - m - m(m-1)/2 - 1$ and the length of each vector v_i is $(k-1)m > M$ we have a singular covariance matrix and so we could use generalized inverses.

Definition 7.1 *A* **generalized inverse** *of a symmetric square matrix A is denoted by A^- and satisfies*

$$A^- A A^- = A.$$

The **Moore–Penrose generalized inverse** *of A is*

where γ_j are the eigenvectors of A corresponding to the p non-zero eigenvalues $\lambda_j, \quad j = 1, \ldots, p$.

To obtain a one sample test a standard multivariate analysis approach is carried out on v_i, where a multivariate normal model for v_i is assumed,

independently for $i = 1, \ldots, n$. The one sample Hotelling's T^2 test could be used (e.g. Mardia et al., 1979, p.125). We write $\bar{v} = \frac{1}{n} \sum v_i$ for the sample mean and we write $S_v = \frac{1}{n} \sum (v_i - \bar{v})(v_i - \bar{v})^{\mathrm{T}}$, for the sample covariance matrix (with divisor n). The Mahalanobis squared distance between v_i and γ_0 is

(7.2)

where S_v^- is the Moore–Penrose generalized inverse of S_v. The rank of S_v is $\min(M, n-1)$ and **we assume that the rank of our sample covariance matrices is M in this chapter**.

Important point: The test statistic is taken as

where the $s_j = \gamma_j^{\mathrm{T}}(\bar{v} - \gamma_0)$ is the jth principal component (PC) score for $(\bar{v} - \gamma_0), \quad j = 1, \ldots, M$. The test statistic F has an $F_{M,n-M}$ distribution under H_0. Hence, we reject

H_0 for large values of F.

Example 7.1 Consider the digit 3 data, described in Section 1.2.4, with $k = 13$ points in $m = 2$ dimensions on $n = 30$ objects. We might wish to examine whether the population mean shape could be an idealized template, such as that displayed in Figure 69, with equal sized loops, with 12 of the landmarks lying equally spaced on two regular octagons (apart from landmark 7 in the middle). The μ_0 is taken as the template and the data are projected into the tangent plane with the pole at the Procrustes mean $\hat{\mu}$. The $M = 22$ PC scores are retained and the squared Mahalanobis distance from $\bar{v}$ to the pole in the tangent space is $\sum_{j=1}^{M} s_j^2/\lambda_j = 47.727$ and hence $F = 17.356$. Since $P(F_{22,8} > 17.356) \approx 0.0002$ we have very strong evidence that the population mean shape does not have the shape of this template. □

This example is used as an illustration of the one sample test, although one could question why one should really be interested in this particular μ_0 for the digit 3. However, the procedure can of course be used to obtain confidence regions for shape, where the confidence region is the set of all μ_0 where H_0 is not rejected.

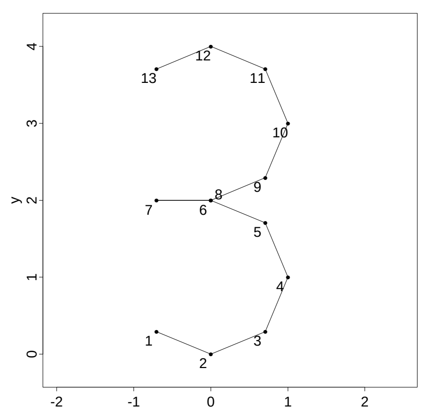

Figure 69 A template number 3 digit, with two equal sized arcs, and with 12 landmarks lying on two regular octagons.

7.1.2 Two independent sample Hotelling's T^2 test

Consider two independent random samples $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ from independent populations with mean shapes $[\mu_1]$ and $[\mu_2]$. To test between

we could carry out a Hotelling's T^2 two sample test in the Procrustes tangent space, where the pole corresponds to the overall pooled full Procrustes mean shape $\hat{\mu}$ (i.e. the full Procrustes mean shape calculated by GPA on all $n_1 + n_2$ individuals). Let $v_1, \ldots, v_{n_1}$ and $w_1, \ldots, w_{n_2}$ be the partial Procrustes tangent coordinates (with pole $\hat{\mu}$). The multivariate normal model is proposed in the tangent space, where and the v_i and w_j are all mutually independent, and common covariance matrices are assumed. We write $\bar{v}, \bar{w}$ and S_v, S_w for the sample means and sample covariance matrices (with divisors n_1 and n_2) in each group. The Mahalanobis distance squared between $\bar{v}$ and $\bar{w}$ is

(7.3)

where $S_u = (n_1 S_v + n_2 S_w)/(n_1 + n_2 - 2)$, and S_u^- is the Moore–Penrose generalized inverse of S_u (see Definition 7.1). Under H_0 we have $\xi_1 = \xi_2$, and we use the test statistic

(7.4)

The test statistic has an F_{M,n_1+n_2-M-1} distribution under H_0. Hence, we reject H_0 for large values of F.

Example 7.2 Consider the gorilla skull data described in

Section 1.2.2. There are $n_1 = 30$ female gorilla skulls and $n_2 = 29$ male gorilla skulls, with $k = 8$ landmarks in two dimensions, and so there are $M = 2k - 4 = 12$ shape dimensions. To examine the assumptions of the model we see whether equal covariance matrices in each group are reasonable. The first three PCs in each group explain 34.8%, 22.9%, 11.2% (females) and 42.0%, 18.0%, 12.4% (males) of the variability in each group. A plot of the first PC for each group is given in Figures 70 and 71 showing fairly similar structures (a contrast with the front of the face and the back of the braincase (pr, na, l) versus the rest). So we take the equal covariance matrix assumption as reasonable. A formal test such as Box's M test (cf. Mardia et al., 1979, p.140) could also be carried out, and here Box's M statistic is 67.18, which has an approximate chi-squared distribution with 78 degrees of freedom under the null hypothesis of equal covariance matrices. Since $P(\chi^2_{78} > 67.18) = 0.8$ we cannot reject the null hypothesis of equal covariance matrices. The percentages of variability explained by the first three within group PCs are 37.3%, 16.0%, 14.7% and the first three PCs are included in Figure 72. In addition, we have no reason to doubt multivariate normality from the pairwise scatters of the standardized PC scores of the data (some of the PC scores are shown in Figure 73).

The observed test statistic is $F = 26.470$ and since $P(F_{12,46} > 4.47) = 0.0001$ we have very strong evidence that the mean shapes are different. So our conclusion would be that there is a significant difference in mean shape between the female and male gorilla skulls in the

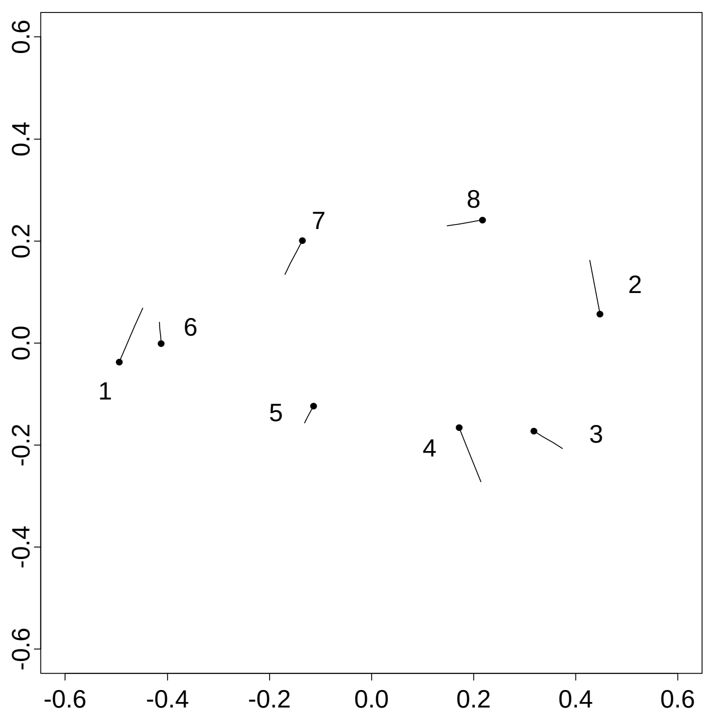

Figure 70 The first PC for the gorilla females. The mean shape is drawn with vectors to an icon +3 (——) standard deviations along the first PC away from the mean.

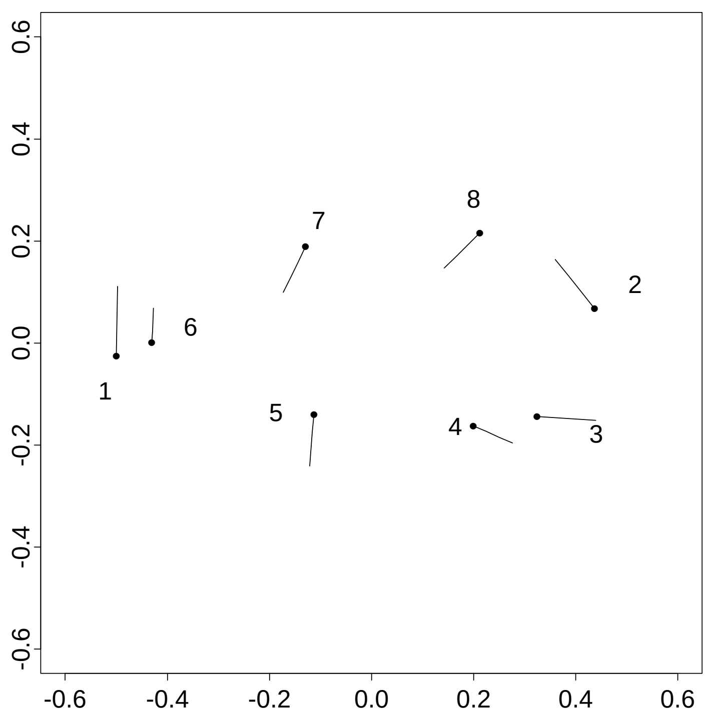

Figure 71 The first PC for the gorilla males. The mean shape is drawn with vectors to an icon +3 (——) standard deviations along the first PC away from the mean.

midline.

□

7.1.3 *Permutation test*

The assumptions of the Hotelling T^2 test may be doubted in certain applications. An alternative procedure is to consider a **permutation test** (Good, 1994; Dryden and Mardia, 1993; Bookstein, 1997b), with the null hypothesis that the groups have equal mean shapes. For a two sample permutation test the data are permuted into two groups of the same size as the groups in the data, and the test statistic is evaluated for all possible permutations $T_1, \ldots, T_P$. The ranking r of the observed test statistic T_{obs} is then used to give the p-value of the permutation test:

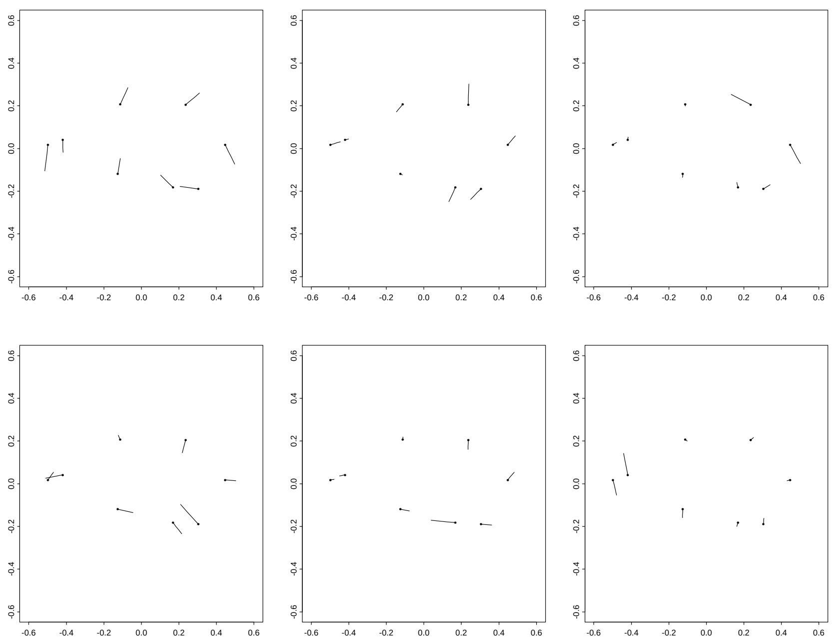

Figure 72 Principal components for the gorilla data using a pooled within group covariance matrix. The first row displays PCs 1, 2 and 3 (from left to right) and the second row displays PCs 9, 11 and 12 (from left to right) . In each plot the mean shape is drawn with vectors to an icon +3 standard deviations along the PC away from the mean. The vectors on the top row are magnified 3 times and the vectors on the bottom row are magnified 10 times.

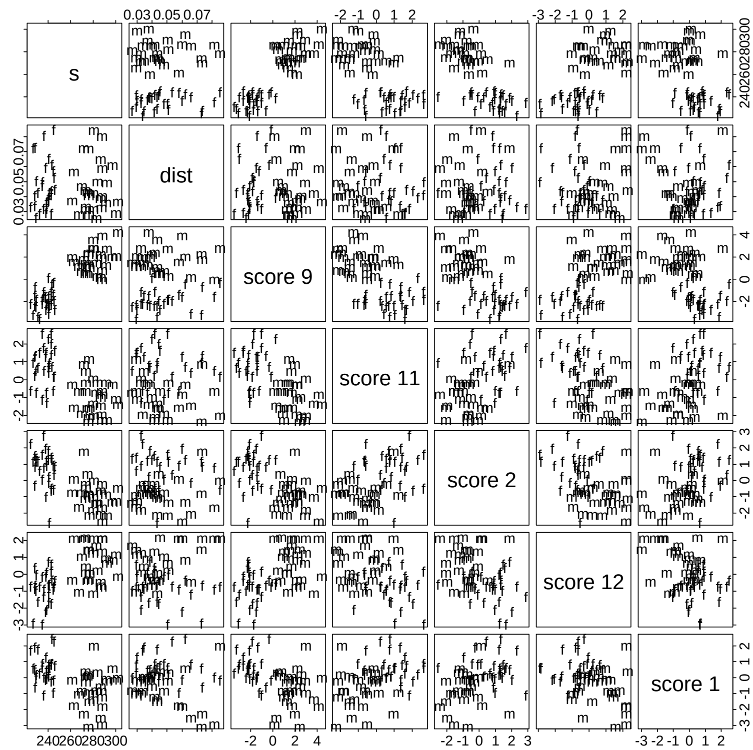

Figure 73 Pairwise scatter plots of the centroid sizes, the full Procrustes distances to the pooled mean and PC scores 9, 11, 2, 12, 1, $(s_i, d_{Fi}, c_{i9}, c_{i11}, c_{i2}, c_{i12}, c_{i1})^{\mathrm{T}}$ for the gorilla data: males (m) and females (f). These particular PC scores c_{ij} have the highest correlation with the observed group shape difference, and we can see a clear separation between the groups in terms of shape, using score 9 and any other score.

An alternative to evaluating all T_i is to consider a number B (say 100 or 1000) of random permutations, and the procedure is called a **Monte Carlo test**. The ranking r of the observed test statistic from B random permutations gives a p-value of

Example 7.3 Consider the 3D macaque data of Section 1.2.8. There are $n_1 = 9$ males and $n_2 = 9$ females, each with $k = 7$ landmarks in $m = 3$ dimensions. We wish to test whether the mean population shapes for both sexes are equal. After performing full Procrustes GPA on the pooled dataset we transform to the tangent space coordinates of Equation (7.1). The dimension of the shape space is $M = 14$ [7×3 (coordinates) -3 (location) -3 (rotation) -1

(size)]. Examining the shape variability in the two groups it seems doubtful that both groups have the same covariance matrix – the male group is more variable in shape, as seen in Example 5.1. In spite of this evidence, proceeding with the Hotelling's T^2 test the squared Mahalanobis distance D^2 is 28.90 between the groups and so $F = 1.74$ and $P(F_{14,3} > 1.74) = 0.36$, and so this is not a significant difference in mean shape.

However, as the assumption of equal variances is doubtful a permutation test could be carried out. The data are randomly split into two groups each of size 9. Out of 99 such permutations the observed F statistic of Equation (7.4) had rank 61, giving a p-value of 0.4, and so there is no evidence for a difference in mean shape.

The sample sizes are very small here and we might expect the Hotelling's T^2 test to be not very powerful. □

Bookstein (1991, p.282) and Bookstein and Sampson (1990) also describe Hotelling's T^2 tests for shape difference using Bookstein coordinates (see Section 7.3), and they also consider testing for affine or uniform shape changes (see Section 10.6.3).

7.1.4 Extensions

Further inference, such as testing the equality of the mean shapes in several groups, proceeds in a similar manner. An overall pooled full Procrustes mean is taken as the pole and multivariate analysis of variance (MANOVA) (see, for example, Mardia et al., 1979, p.333) is carried out on the Procrustes tangent coordinates. General linear models could be proposed in the tangent space and the full armoury of multivariate data analysis can be used to analyse shape data, provided variations are small.

7.1.5 Dimension reduction in inference

In some datasets there are few observations and possibly many landmarks on each individual. Although inference can be carried out in a suitable tangent space there is often a problem with the space being over-dimensioned. For example, a Hotelling's T^2 test may not be very powerful unless there are a large number of observations available. A practical solution is to perform a PCA on the pooled datasets and retain the first few PC scores, although there are obvious dangers, particularly if a true group difference is orthogonal to the first few PCs. An alternative approach is to delete excess elements in each vector v_i.

7.2 Inference Using Procrustes Statistics Under Isotropy

Another simple approach to statistical inference is to work with statistics based on squared Procrustes distances. Goodall (1991) has considered such an approach using approximate chi-squared distributions, following from the work of Sibson (1978, 1979) and Langron and Collins (1985). The underlying model is that configurations are isotropic normal perturbations from mean configurations, and the distributions of the squared Procrustes distances are approximately chi-squared distributions. The procedures require a much more restrictive isotropic model than the previous section, but when the model is appropriate more powerful tests result.

7.2.1 One sample Goodall's F test

We consider first the case when a random sample of n observations $X_1, \ldots, X_n$ (each a $k \times m$ matrix) is taken from an isotropic normal model with mean μ and transformed by an additional location, rotation and scale, i.e.

$$X_i = \beta_i(\mu + E_i)\Gamma_i + 1_k\gamma_i^{\mathrm{T}}, \quad \mathrm{vec}(E_i) \sim N(0, \sigma^2 I_{km}), \tag{7.5}$$

where $\beta_i > 0$ (scale), $\Gamma_i \in SO(m)$ (rotation) and $\gamma_i \in \mathbb{R}^m$ (translation), and σ is small.

The following approximate analysis of variance (ANOVA) identity holds for $\hat{\mu} \approx \mu$ and small σ:

$$\sum_{i=1}^{n} d_F^2(X_i, \mu) \approx \sum_{i=1}^{n} d_F^2(X_i, \hat{\mu}) + n d_F^2(\mu, \hat{\mu}),$$

where $\hat{\mu}$ is the full Procrustes mean and d_F is the full

Procrustes distance of Equation (4.15). The proof can be seen using Taylor series expansions. Note the similarities with analysis of variance in classical regression analysis – the left-hand side of the equation is like a total sum of squares and the right-hand side is like the residual sum of squares plus the explained (regression) sum of squares.

Consider testing between $H_0 : [\mu] = [\mu_0]$ and $H_1 : [\mu] \neq [\mu_0]$. Under the null model it can be shown that approximately (to second order terms in E_i) that

$$d_F^2(X_i, \mu_0) \sim \tau_0^2 \chi_M^2 \tag{7.6}$$

independently for $i = 1, \ldots, n$, where $M = km - m - m(m-1)/2 - 1$ is the dimension of the shape space, $\tau_0 = \sigma/\delta_0$, and $\delta_0 = S(\mu_0) = \|C\mu_0\|$ is the centroid size of μ_0. The proof can be obtained by Taylor series expansions, after Sibson (1978), and the proof for the

$m = 2$ dimensional case is seen from Equation (6.16), when discussing the complex Watson distribution.

From the additive property of independent chi-squared distributions,

$$\sum_{i=1}^{n} d_F^2(X_i, \mu_0) \sim \tau_0^2 \chi_{nM}^2.$$

In addition, since M parameters are estimated in $\hat{\mu}$ we have

$$\sum_{i=1}^{n} d_F^2(X_i, \hat{\mu}) \sim \tau_0^2 \chi_{(n-1)M}^2$$

and $d_F^2(\mu_0, \hat{\mu})$ is approximately independent of $\sum d_F^2(X_i, \hat{\mu})$. Hence, approximately

$$n d_F^2(\mu_0, \hat{\mu}) \sim \tau_0^2 \chi_M^2, \tag{7.7}$$

again using the additive property of independent chi-squared distributions. So, under H_0 we have the approximate result

$$F = (n-1)n \; \frac{d_F^2(\mu_0, \hat{\mu})}{\sum_{i=1}^{n} d_F^2(X_i, \hat{\mu})} \sim F_{M,(n-1)M}. \tag{7.8}$$

This is valid for small σ and μ_0 close to $\hat{\mu}$, and so we reject H_0 for large values of this test statistic. We call the test the **one sample Goodall's F test**, after Goodall (1991).

If τ_0 is small, E_i is isotropic (but not necessarily normal) and nM is large, then approximately

$$\sum_{i=1}^{n} d_F^2(X_i, \mu) \sim N(\tau_0^2 nM, 2\tau_0^4 nM)$$

by applying the central limit theorem.

The test based on the isotropic model can be seen as a special case of the Hotelling's T^2 procedure of Section 7.1.1. If we replace S_v with $s_v^2 I_{2k-2}$, where s_v^2 is the unbiased estimate of variance, then the Mahalanobis distance of Equation (7.2) becomes

$$D^2 = s_v^{-2} \|\gamma_0 - \bar{v}\|^2 = d_F^2(\mu_0, \hat{\mu})/s_v^2,$$

from Equation (4.30). Now

$$s_v^2 = \frac{1}{n-1}\sum_{i=1}^{n}\|v_i - \bar{v}\|^2 = \frac{1}{n-1}\sum_{i=1}^{n} d_F^2(X_i, \hat{\mu})$$

and hence the test statistic for the one sample Hotelling's T^2 test statistic would be proportional to

$$d_F^2(\mu_0, \hat{\mu}) \Big/ \sum_{i=1}^{n} d_F^2(X_i, \hat{\mu}).$$

Hence the one sample Hotelling's T^2 test under the isotropic model would be identical to using the F statistic of Equation (7.8).

7.2.2 *Two independent sample Goodall's F test*

Consider independent random samples $X_1, \ldots, X_{n_1}$ from a population modelled by Equation (7.5) with mean μ_1, and $Y_1, \ldots, Y_{n_2}$ from Equation (7.5) with mean μ_2. Both populations are assumed to have a common variance for each coordinate σ^2. We wish to test $H_0 : [\mu_1] = [\mu_2](=$

$[\mu_0]$), say, against $H_1 : [\mu_1] \neq [\mu_2]$. Let $[\hat{\mu}_1]$ and $[\hat{\mu}_2]$ be the full Procrustes means of each sample, with icons $\hat{\mu}_1$ and $\hat{\mu}_2$. Under H_0, with σ small, the Procrustes distances are approximately distributed as

$$\sum_{i=1}^{n_1} d_F^2(X_i, \hat{\mu}_1) \sim \tau_0^2 \chi^2_{(n_1-1)M},$$

$$\sum_{i=1}^{n_2} d_F^2(Y_i, \hat{\mu}_2) \sim \tau_0^2 \chi^2_{(n_2-1)M},$$

$$d_F^2(\hat{\mu}_1, \hat{\mu}_2) \sim \tau_0^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \chi^2_M,$$

where $\tau_0 = \sigma/\delta_0$ and $\delta_0 = S(\mu_0)$. Again, proofs of the results can be obtained using Taylor series expansions. In addition these statistics are approximately mutually independent (exactly in the case of the first two expressions). Hence, under H_0 we have the approximate distribution

$$F = \frac{n_1 + n_2 - 2}{{n_1}^{-1} + {n_2}^{-1}} \quad \frac{d_F^2(\hat{\mu}_1, \hat{\mu}_2)}{\sum_{i=1}^{n_1} d_F^2(X_i, \hat{\mu}_1) + \sum_{i=1}^{n_2} d_F^2(Y_i, \hat{\mu}_2)} \sim F_{M,(n_1+n_2-2)M}, \tag{7.9}$$

and again this result is valid for small σ. We reject H_0 for large values of this test statistic. We call the test the **two independent sample Goodall's F test** after Goodall (1991).

Example 7.4 Consider the schizophrenia data described in Section 1.2.3. We wish to test whether the mean shapes of brain landmarks are different in the two groups of control subjects and schizophrenic patients. There are $k = 13$ landmarks in $m = 2$ dimensions. The Procrustes rotated data for the groups are displayed in Figures 74 and 75. We see that there are generally circular scatters of points at each landmark in each group (as required for the isotropic

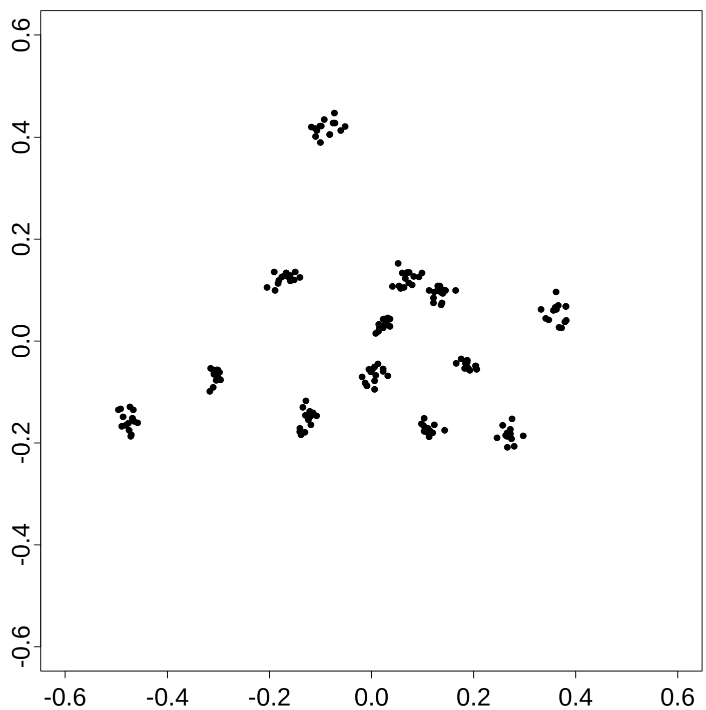

Figure 74 The Procrustes rotated brain landmark data for the 14 control subjects.

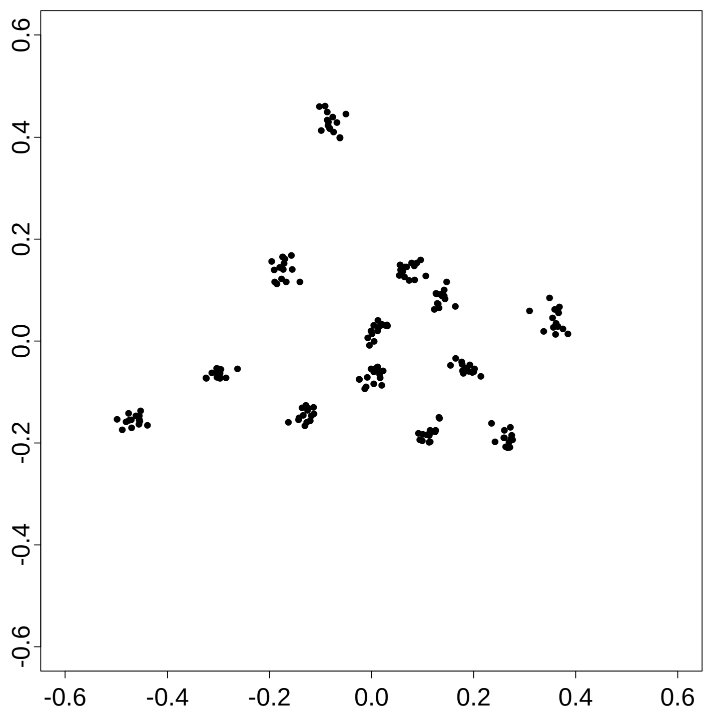

Figure 75 The Procrustes rotated brain landmark data for the 14 schizophrenic patients.

model).

The percentages of variability explained by the first three PCs are $31.6\%, 21.4\%, 13.2\%$ for the controls and $27.1\%, 21.7\%, 4.8\%$ for the schizophrenic patients. Box's M test was carried out and there is some evidence against equal covariance matrices. The root mean square of d_F in each group is 0.068 in the controls and 0.073 in the schizophrenia group. Carrying out a likelihood ratio test for isotropy in each group (e.g Mardia et al., 1979, p.235) on the 13 non-zero unstandardized scores, we cannot reject the null hypothesis of isotropy. Of course we only have small samples here, and so with more data we may reject isotropy – this topic merits further investigation. Despite Box's M test suggesting a difference in covariance matrices, for illustrative purposes we shall proceed with

with the Goodall F test, which assumes isotropy and equal variances.

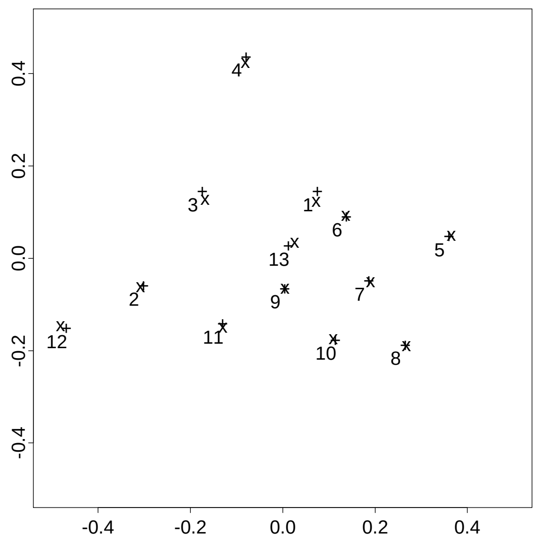

Figure 76 The full Procrustes mean shapes of the normal subjects (x) and schizophrenic patients (+) for the brain landmark data, rotated to each other by GPA.

The mean configurations are displayed in Figure 76. The full Procrustes distance d_F between the mean shapes is 0.038. The sum of squared full Procrustes distances from each configuration to its mean shape is 0.140 and so the

F statistic is 1.89. Since $P(F_{22,572} \geq 1.89) \approx 0.01$ we have evidence for a significant difference in shape. So, we conclude that the subjects with schizophrenia have different shaped mean landmark configurations from the control subjects.

Following Bookstein (1997b) we also consider a Monte Carlo test, as described in Section 7.1.3, based on 999 random permutations. The configurations are randomly assigned into each of the two groups, the F statistic is calculated and the proportion of times that the resulting F statistic exceeds the observed value of 1.89 is the p-value for the test. From 999 random permutations we obtained a p-value of 0.04. Hence, we have some evidence that the mean configurations are different in shape, but with a larger p-value than for the isotropic based tests.

If we carry out a Hotelling's T^2 test in the tangent space we have $F = 0.834$ which is near the centre of the null distribution ($P(F_{22,5} > 0.834) = 0.66$). So, the Hotelling's T^2 provides no evidence for a shape difference, illustrating that the Hotelling's T^2 procedure is less powerful than Goodall's F test, when the isotropic normal model holds. Power is lost because many degrees of freedom are used in estimating the covariance matrix in the Hotelling's T^2 test.

□

Note that the test based on the isotropic model can be seen as a special case of the two sample Hotelling's T^2 procedure of Section 7.1.2. If we replace S_u with $s_u^2 I_{2k-2}$, where s_u^2 is the unbiased estimate of variance, then the Mahalanobis distance of Equation (7.3) becomes

$$D^2 = s_u^{-2} \|\bar{v} - \bar{w}\|^2 \approx d_F^2(\hat{\mu}_1, \hat{\mu}_2)/s_u^2,$$

from Equation (4.31). Now

$$
\begin{aligned}
s_u^2 &= \frac{1}{n_1+n_2-2}\left\{\sum_{i=1}^{n_1}\|v_i-\bar{v}\|^2+\sum_{j=1}^{n_2}\|w_j-\bar{w}\|^2\right\}\\
&\approx \frac{1}{n_1+n_2-2}\left\{\sum_{i=1}^{n_1}d_F^2(X_i,\hat{\mu}_1)+\sum_{j=1}^{n_2}d_F^2(Y_j,\hat{\mu}_2)\right\} \qquad (7.10)
\end{aligned}
$$

and so the test statistic for the two sample Hotelling's T^2 test statistic would be proportional to

$$
\frac{d_F^2(\hat{\mu}_1,\hat{\mu}_2)}{\sum_{i=1}^{n_1}d_F^2(X_i,\hat{\mu}_1)+\sum_{j=1}^{n_2}d_F^2(Y_j,\hat{\mu}_2)}.
$$

Hence the Hotelling's T^2 test under the isotropic model would be identical to using the F statistic of Equation (7.9).

7.2.3 One way analysis of variance

Consider a balanced analysis of variance with independent random samples $(X_{i1},...,X_{in})^{\mathrm{T}},\ i=1,...,n_G$, from n_G groups, each of size n. Let $[\hat{\mu}_j]$ be the group full Procrustes means and $[\hat{\mu}]$ is the overall pooled full Procrustes mean shape. A suitable test statistic (Goodall, 1991) is

$$F = n(n-1)n_G \ \frac{\sum_{j=1}^{n_G} d_F^2(\hat{\mu}_j, \hat{\mu})}{(n_G - 1)\sum_{j=1}^{n_G}\sum_{i=1}^{n} d_F^2(X_{ji}, \hat{\mu}_j)}.$$

Under the null hypothesis of equal means the approximate distribution of F is $F_{(n_G-1)M, n_G(n-1)M}$ and the null hypothesis is rejected for large values of the statistic. Since

$$d_F^2(\hat{\mu}_1, \hat{\mu}_2) = 2\left(d_F^2(\hat{\mu}_1, \hat{\mu}) + d_F^2(\hat{\mu}_2, \hat{\mu})\right),$$

the two sample test of the previous section (with $n_1 = n_2$) is a special case.