Chapter 9
Hotelling's T^2 Test

The Hotelling's T^2 test is used to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ when there is one sample, and $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ when there are two samples. Other applications include the multivariate matched pairs test and a test in the repeated measurements setting. These tests are robust to nonnormality.

The one-sample Hotelling's T^2 test, multivariate matched pairs test, and two-sample Hotelling's T^2 test are analogs of the univariate one-sample t test, matched pairs t test, and two-sample t test, respectively. For the multivariate Hotelling's T^2 tests, there are $p > 1$ variables and their correlations are important.

9.1 One Sample

The one-sample Hotelling's T^2 test is used to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_A : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. The test rejects H_0 if

$$T_H^2 = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^T \boldsymbol{S}^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) > \frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha}$$

where $P(Y \leq F_{p,d,\alpha}) = \alpha$ if $Y \sim F_{p,d}$.

If a multivariate location estimator T satisfies

$$\sqrt{n}(T - \boldsymbol{\mu}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{D}),$$

then a competing test rejects H_0 if

$$T_C^2 = n(T - \boldsymbol{\mu}_0)^T \hat{\boldsymbol{D}}^{-1} (T - \boldsymbol{\mu}_0) > \frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha}$$

D. J. Olive, *Robust Multivariate Analysis*,
https://doi.org/10.1007/978-3-319-68253-2_9

where $\hat{\boldsymbol{D}}$ is a consistent estimator of $\boldsymbol{D}$. The scaled F cutoff can be used since $T_C^2 \xrightarrow{D} \chi_p^2$ if H_0 holds, and

$$\frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha} \rightarrow \chi^2_{p,1-\alpha}$$

as $n \rightarrow \infty$. This idea is used for small p by Srivastava and Mudholkar (2001) where T is the coordinatewise trimmed mean. The one-sample Hotelling's T^2 test uses $T = \overline{\boldsymbol{x}}$, $\boldsymbol{D} = \boldsymbol{\Sigma}_{\boldsymbol{x}}$, and $\hat{\boldsymbol{D}} = \boldsymbol{S}$.

The Hotelling's T^2 test is a large sample level α test in that if $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid from a distribution with mean $\boldsymbol{\mu}_0$ and nonsingular covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}$, then the type I error = P(reject H_0 when H_0 is true) $\rightarrow \alpha$ as $n \rightarrow \infty$. We want $n \geq 10p$ if the DD plot is linear through the origin and subplots in the scatterplot matrix all look ellipsoidal. For any n, there are distributions with nonsingular covariance matrix where the χ_p^2 approximation to T_H^2 is poor.

Let pval be an estimate of the pvalue. We typically use $T_C^2 = T_H^2$ in the following four-step test. i) State the hypotheses $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$.
ii) Find the test statistic $T_C^2 = n(T - \boldsymbol{\mu}_0)^T \hat{\boldsymbol{D}}^{-1} (T - \boldsymbol{\mu}_0)$.
iii) Find pval =

$$P\left(T_C^2 < \frac{(n-1)p}{n-p} F_{p,n-p}\right) = P\left(\frac{n-p}{(n-1)p} \ T_C^2 < F_{p,n-p}\right).$$

iv) State whether you fail to reject H_0 or reject H_0. If you reject H_0 then conclude that $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ while if you fail to reject H_0 conclude that the population mean $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ or that there is not enough evidence to conclude that $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. Reject H_0 if pval $\leq \alpha$ and fail to reject H_0 if pval $> \alpha$. As a benchmark for this text, use $\alpha = 0.05$ if α is not given.

If $\boldsymbol{W}$ is the data matrix, then $R(\boldsymbol{W})$ is a large sample $100(1-\alpha)\%$ confidence region for $\boldsymbol{\mu}$ if $P[\boldsymbol{\mu} \in R(\boldsymbol{W})] \rightarrow 1 - \alpha$ as $n \rightarrow \infty$. If $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid from a distribution with mean $\boldsymbol{\mu}$ and nonsingular covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}$, then

$$R(\boldsymbol{W}) = \{\boldsymbol{w} | n(\overline{\boldsymbol{x}} - \boldsymbol{w})^T \boldsymbol{S}^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{w}) \leq \frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha}\}$$

is a large sample $100(1-\alpha)\%$ confidence region for $\boldsymbol{\mu}$. This region is a hyperellipsoid centered at $\overline{\boldsymbol{x}}$. Note that the estimated covariance matrix for $\overline{\boldsymbol{x}}$ is $\boldsymbol{S}/n$ and $n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^T \boldsymbol{S}^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{\mu}) = D^2_{\boldsymbol{\mu}}(\overline{\boldsymbol{x}}, \boldsymbol{S}/n)$. If $\boldsymbol{\mu}$ is close to $\overline{\boldsymbol{x}}$ with respect to the Mahalanobis distance based on dispersion matrix $\boldsymbol{S}/n$, then $\boldsymbol{\mu}$ will be in the confidence region.

Recall from Theorem 1.1e that $\max_{\boldsymbol{a} \neq \boldsymbol{0}} \frac{\boldsymbol{a}^T (\overline{\boldsymbol{x}} - \boldsymbol{\mu})(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^T \boldsymbol{a}}{\boldsymbol{a}^T \boldsymbol{S} \boldsymbol{a}} =$ $n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^T \boldsymbol{S}^{-1} (\overline{\boldsymbol{x}} - \boldsymbol{\mu}) = T^2$. This fact can be used to derive large sample simultaneous confidence intervals for $\boldsymbol{a}^T \boldsymbol{\mu}$ in that separate confidence statements using different choices of $\boldsymbol{a}$ all hold simultaneously with probability tending to $1-\alpha$. Let $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ be iid with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma_x} > 0$. Then simultaneously for all $\boldsymbol{a} \neq \boldsymbol{0}$, $P(L_{\boldsymbol{a}} \leq \boldsymbol{a}^T \boldsymbol{\mu} \leq U_{\boldsymbol{a}}) \to 1-\alpha$ as $n \to \infty$ where

$$[L_{\boldsymbol{a}}, U_{\boldsymbol{a}}] = \boldsymbol{a}^T \overline{\boldsymbol{x}} \pm \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p,1-\alpha} \boldsymbol{a}^T \boldsymbol{S} \boldsymbol{a}}.$$

Simultaneous confidence intervals (CIs) can be made after collecting data and hence are useful for "data snooping." Following Johnson and Wichern (1988, pp. 184–5), the p confidence intervals (CIs) for μ_i and the $p(p-1)/2$ CIs for $\mu_i - \mu_k$ can be made such that for each of the two types of CI, they all hold simultaneously with confidence $\to 1-\alpha$. Hence if $\alpha = 0.05$, then in 100 samples, we expect all p CIs to contain μ_i about 95 times, and we expect all $p(p-1)/2$ CIs to contain $\mu_i - \mu_k$ about 95 times. For each of the two types of CI, about 5 times at least one of the CIs will fail to contain its parameter (μ_i or $\mu_i - \mu_k$). The simultaneous CIs for μ_i are

$$[L, U] = \overline{x}_i \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,1-\alpha}} \sqrt{\frac{S_{ii}}{n}}$$

while the simultaneous CIs for $\mu_i - \mu_k$ are

$$[L, U] = \overline{x}_i - \overline{x}_k \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,1-\alpha}} \sqrt{\frac{S_{ii} - 2S_{ik} + S_{kk}}{n}}.$$

Example 9.1. Following Mardia et al. (1979, p. 126), data for first and second adult sons had $n = 25$ and variables X_1 = head length of first son and X_2 = head length of second son. Suppose $\boldsymbol{\mu}_0 = (182, 182)^T$ and $T_C^2 = 1.28$. Perform the one-sample Hotelling's T^2 test.

Solution: i) $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$

ii) $T_C^2 = 1.28$

iii) $\frac{n-p}{(n-1)p} T_C^2 = \frac{25-2}{(24(2)} 1.28 = 0.6133$, and pval $= P(0.613 < F_{2,23}) > 0.05$

iv) Fail to reject H_0, so $\boldsymbol{\mu} = (182, 182)^T$.

9.1.1 A Diagnostic for the Hotelling's T^2 Test

Now the RMVN estimator is asymptotically equivalent to a scaled DGK estimator that uses $k = 5$ concentration steps and two "reweight for efficiency" steps. Lopuhaä (1999, pp. 1651–1652) showed that if (E1) holds, then the classical estimator applied to cases with $D_i(\overline{\boldsymbol{x}}, S) \leq h$ is asymptotically normal with

$$\sqrt{n}(T_{0,D} - \boldsymbol{\mu}) \xrightarrow{D} N_p(\mathbf{0}, \kappa_p \boldsymbol{\Sigma}).$$

Here h is some fixed positive number, such as $h = \chi^2_{p,0.975}$, so this estimator is not quite the DGK estimator after one concentration step.

We conjecture that a similar result holds after concentration:

$$\sqrt{n}(T_{RMVN} - \boldsymbol{\mu}) \xrightarrow{D} N_p(\mathbf{0}, \tau_p \boldsymbol{\Sigma})$$

for a wide variety of elliptically contoured distributions where τ_p depends on both p and the underlying distribution. Since the "test" is based on a conjecture, it is ad hoc and should be used as an outlier diagnostic rather than for inference.

For MVN data, simulations suggest that τ_p is close to 1. The ad hoc test that rejects H_0 if

$$\frac{T_R^2}{f_{n,p}} = n(T_{RMVN} - \boldsymbol{\mu}_0)^T \hat{\boldsymbol{C}}^{-1}_{RMVN}(T_{RMVN} - \boldsymbol{\mu}_0)/f_{n,p} > \frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha}$$

where $f_{n,p} = 1.04 + 0.12/p + (40 + p)/n$ gave fair results in the simulations described later in this subsection for $n \geq 15p$ and $2 \leq p \leq 100$.

Table 9.1 Hotelling simulation

p	n = 15p	hcv	rhcv	n = 20p	hcv	rhcv	n = 30p	hcv	rhcv
10	150	0.0476	0.0300	200	0.0516	0.0304	300	0.0498	0.0286
15	225	0.0474	0.0318	300	0.0506	0.0308	450	0.0492	0.0320
20	300	0.0540	0.0368	400	0.0548	0.0314	600	0.0520	0.0354
25	375	0.0444	0.0334	500	0.0462	0.0296	750	0.0456	0.0288
30	450	0.0472	0.0324	600	0.0516	0.0358	900	0.0484	0.0342
35	525	0.0490	0.0384	700	0.0522	0.0358	1050	0.0502	0.0374
40	600	0.0534	0.0440	800	0.0486	0.0354	1200	0.0526	0.0336
45	675	0.0406	0.0390	900	0.0544	0.0390	1350	0.0512	0.0366
50	750	0.0498	0.0430	1000	0.0522	0.0394	1500	0.0512	0.0364
55	825	0.0504	0.0502	1100	0.0496	0.0392	1650	0.0510	0.0374
60	900	0.0482	0.0514	1200	0.0488	0.0404	1800	0.0474	0.0376
65	975	0.0568	0.0602	1300	0.0524	0.0414	1950	0.0548	0.0410
70	1050	0.0462	0.0530	1400	0.0558	0.0432	2100	0.0522	0.0424
75	1125	0.0474	0.0632	1500	0.0502	0.0486	2250	0.0490	0.0370
80	1200	0.0524	0.0620	1600	0.0524	0.0432	2400	0.0468	0.0356
85	1275	0.0482	0.0758	1700	0.0496	0.0456	2550	0.0520	0.0404
90	1350	0.0504	0.0746	1800	0.0484	0.0454	2700	0.0484	0.0398
95	1425	0.0524	0.0892	1900	0.0472	0.0506	2850	0.0538	0.0424
100	1500	0.0554	0.0808	2000	0.0452	0.0506	3000	0.0488	0.0392

The correction factor $f_{n,p}$ was found by simulating the "robust" and classical test statistics for 100 runs, plotting the test statistics, then finding a correction factor so that the identity line passed through the data. The following *R* commands were used to make Figure 9.1, which shows that the plotted points of the scaled "robust" test statistic versus the classical test statistic scatter about the identity line.

Table 9.2 Hotelling power simulation

p	n	hcv	rhcv	δ	n	hcv	rhcv	δ	n	hcv	rhcv	δ
5	75	0.459	0.245	0.20	100	0.366	0.184	0.15	150	0.333	0.208	0.12
5	75	0.682	0.416	0.25	100	0.599	0.368	0.20	150	0.577	0.394	0.16
5	75	0.840	0.588	0.30	100	0.816	0.587	0.30	150	0.860	0.708	0.40
10	150	0.221	0.113	0.10	200	0.312	0.182	0.10	300	0.469	0.340	0.10
10	150	0.621	0.400	0.17	200	0.655	0.467	0.15	300	0.647	0.504	0.12
10	150	0.888	0.729	0.22	200	0.848	0.692	0.18	300	0.872	0.767	0.15
15	225	0.314	0.188	0.10	300	0.442	0.294	0.10	450	0.317	0.228	0.07
15	225	0.714	0.543	0.15	300	0.623	0.449	0.12	450	0.648	0.522	0.10
15	225	0.881	0.738	0.18	300	0.858	0.755	0.15	450	0.853	0.762	0.12
20	300	0.408	0.276	0.10	400	0.341	0.230	0.08	600	0.291	0.216	0.06
20	300	0.691	0.525	0.13	400	0.674	0.534	0.11	600	0.554	0.433	0.08
20	300	0.935	0.852	0.17	400	0.858	0.742	0.13	600	0.790	0.701	0.10
25	375	0.304	0.214	0.08	500	0.434	0.319	0.08	750	0.354	0.266	0.06
25	375	0.728	0.580	0.12	500	0.676	0.531	0.10	750	0.660	0.556	0.08
25	375	0.926	0.837	0.15	500	0.868	0.771	0.12	750	0.887	0.815	0.10
30	450	0.374	0.264	0.08	600	0.395	0.290	0.07	900	0.290	0.217	0.05
30	450	0.602	0.467	0.10	600	0.639	0.517	0.09	900	0.743	0.642	0.08
30	450	0.883	0.763	0.13	600	0.867	0.770	0.11	900	0.876	0.808	0.09

```
n<-4000; p <- 30 #May take a few minutes.
zout <- rhotsim(n=4000,p=30)
SRHOT <- zout$rhot/(1.04 + 0.12/p + (40+p)/n)
HOT <- zout$hot
plot(SRHOT,HOT)
abline(0,1)
```

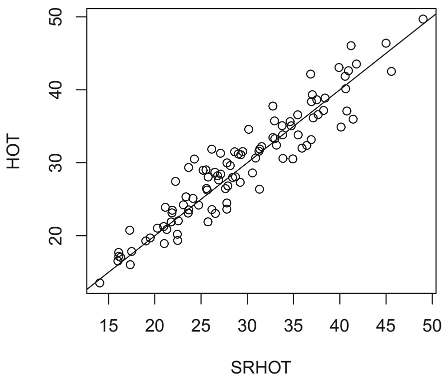

Fig. 9.1 Scaled "Robust" Statistic Versus T_H^2 Statistic

For the Hotelling's T_H^2 simulation, the data is $N_p(\delta \mathbf{1}, diag(1, 2, ..., p))$ where $H_0 : \boldsymbol{\mu} = \mathbf{0}$ is being tested with 5000 runs at a nominal level of 0.05. In Table 9.1, $\delta = 0$ so H_0 is true, while hcv and rhcv are the proportion of rejections by the T_H^2 test and by the ad hoc robust test. Sample sizes are $n = 15p, 20p$, and $30p$. The robust test is not recommended for $n < 15p$ and appears to be conservative (the proportion of rejections is less than the nominal 0.05) except when $n = 15p$ and $75 \le p \le 100$. See Zhang (2011).

If $\delta > 0$, then H_0 is false and the proportion of rejections estimates the power of the test. Table 9.2 shows that T_H^2 has more power than the robust test, but suggests that the power of both tests rapidly increases to one as δ increases.

9.1.2 Bootstrapping Hotelling's T^2 Type Tests

The prediction region method of Section 5.3 is useful for bootstrapping the test $H_0 : \boldsymbol{\mu}_T = \boldsymbol{\mu}_0$ versus $H_A : \boldsymbol{\mu}_T \neq \boldsymbol{\mu}_0$ where the test statistic T estimates the parameter $\boldsymbol{\mu}_T$. Take a sample of size n with replacement from the cases $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ to make the bootstrap statistic T_1^*. Repeat to get the bootstrap sample $T_1^*, ..., T_B^*$. Apply the nonparametric prediction region to the bootstrap sample and see if $\boldsymbol{\mu}_0$ is in the region. Equivalently, apply the nonparametric prediction region to $\boldsymbol{w}_i = T_i^* - \boldsymbol{\mu}_0$, $i = 1, ..., B$, and fail to reject H_0 if $\mathbf{0}$ is in the region, otherwise reject H_0.

The *mpack* function `rhotboot` bootstraps T where T is the coordinatewise median or T is the RMVN location estimator. The function `medhotsim` simulates the test with $\boldsymbol{\mu}_0 = \mathbf{0}$ when T is the coordinatewise median. The simulated data are as in Section 6.3, with $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{z}$, except that $\boldsymbol{z} = \boldsymbol{u} - \mathbf{1}$ was used for the multivariate lognormal distribution with $u_i = \exp(w_i)$ and

$w_i \sim N(0,1)$, so that the population coordinatewise median of $\boldsymbol{x}$ and $\boldsymbol{z}$ was $\boldsymbol{0}$ when H_0 is true. When H_0 was false, $\boldsymbol{\mu}_0 = \delta\boldsymbol{1}$ with $\delta > 0$.

The term *hotcov* was the proportion of times the bootstrap test rejected H_0 with a nominal level of 0.05. With $n = 100$ and $p = 2$, *hotcov* was near 0.05 when H_0 was true. The test usually had good power if $\boldsymbol{\mu} = (0.5, 0.5)^T$. See output below where 1000 runs were used.

```
medhotsim(xtype=1,nruns=1000)
0.046            #MVN((0,0)^T, diag(1,2)) data
medhotsim(xtype=1,nruns=1000,delta=0.5)
0.995            #MVN((0.5,0.5)^T, diag(1,2)) data
```

9.2 Matched Pairs

Assume that there are $k = 2$ treatments, and both treatments are given to the same n cases or units. Then p measurements are taken for both treatments. For example, systolic and diastolic blood pressure could be compared before and after the patient (case) receives blood pressure medication. Then $p = 2$. Alternatively use n correlated pairs, for example, pairs of animals from the same litter or neighboring farm fields. Then use randomization to decide whether the first member of the pair gets treatment 1 or treatment 2. Let $n_1 = n_2 = n$ and assume $n - p$ is large.

Let $\boldsymbol{y}_i = (Y_{i1}, Y_{i2}, ..., Y_{ip})^T$ denotes the p measurements from the 1st treatment, and $\boldsymbol{z}_i = (Z_{i1}, Z_{i2}, ..., Z_{ip})^T$ denotes the p measurements from the 2nd treatment. Let $\boldsymbol{d}_i \equiv \boldsymbol{x}_i = \boldsymbol{y}_i - \boldsymbol{z}_i$ for $i = 1, ..., n$. Assume that the $\boldsymbol{x}_i$ are iid with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}$. Let $T^2 = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^T \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu})$. Then $T^2 \xrightarrow{P} \chi^2_p$ and $pF_{p,n-p} \xrightarrow{P} \chi^2_p$. Let $P(F_{p,n} \le F_{p,n,\delta}) = \delta$. Then the one-sample Hotelling's T^2 inference is done on the differences $\boldsymbol{x}_i$ using $\boldsymbol{\mu}_0 = \boldsymbol{0}$. If the p random variables are continuous, make three DD plots: one for the $\boldsymbol{x}_i$, one for the $\boldsymbol{y}_i$, and one for the $\boldsymbol{z}_i$ to detect outliers.

Let pval be an estimate of the pvalue. The **large sample multivariate matched pairs test** has four steps.
i) State the hypotheses $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ $\quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{0}$.
ii) Find the test statistic $T^2_M = n\overline{\boldsymbol{x}}^T \boldsymbol{S}^{-1}\overline{\boldsymbol{x}}$.
iii) Find pval =

$$P\left(T^2_M < \frac{(n-1)p}{n-p} F_{p,n-p}\right) = P\left(\frac{n-p}{(n-1)p}\, T^2_M < F_{p,n-p}\right).$$

iv) State whether you fail to reject H_0 or reject H_0. If you reject H_0 then conclude that $\boldsymbol{\mu} \neq \boldsymbol{0}$ while if you fail to reject H_0 conclude that the population mean $\boldsymbol{\mu} = \boldsymbol{0}$ or that there is not enough evidence to conclude that $\boldsymbol{\mu} \neq \boldsymbol{0}$.

Reject H_0 if pval $\leq \alpha$ and fail to reject H_0 if pval $> \alpha$. As a benchmark for this text, use $\alpha = 0.05$ if α is not given.

A large sample $100(1-\alpha)\%$ confidence region for $\boldsymbol{\mu}$ is

$$\{\boldsymbol{w}| \ n(\overline{\boldsymbol{x}} - \boldsymbol{w})^T \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{w}) \leq \frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha}\},$$

and the p large sample simultaneous confidence intervals (CIs) for μ_i are

$$[L, U] = \overline{x}_i \pm \sqrt{\frac{p(n-1)}{(n-p)} F_{p,n-p,1-\alpha}} \sqrt{\frac{S_{ii}}{n}}$$

where $S_{ii} = S_i^2$ is the ith diagonal element of $\boldsymbol{S}$.

Example 9.2. Following Johnson and Wichern (1988, pp. 213–214), wastewater from a sewage treatment plant was sent to two labs for measurements of biochemical demand (BOD) and suspended solids (SS). Suppose $n = 11$, $p = 2$, and $T_M^2 = 13.6$. Perform the appropriate test.

Solution: i) $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ $\quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{0}$

ii) $T_M^2 = 13.6$

iii) $\dfrac{n-p}{(n-1)p} T_M^2 = \dfrac{11-2}{(11-1)2} 13.6 = 6.12$, and pval $= P(6.12 < F_{2,9}) < 0.05$

iv) Reject H_0. Hence $\boldsymbol{\mu} \neq (0,0)^T$, and the two labs are giving different mean measurements for $(\mu_{BOD}, \mu_{SS})^T$.

To get a bootstrap analog of this test, bootstrap the $\boldsymbol{d}_i = \boldsymbol{x}_i$ as in Section 9.1.2 where usually $H_0 : \boldsymbol{\mu} \equiv \boldsymbol{\mu}_T = \boldsymbol{0}$. Again robust location estimators, such as the coordinatewise median or RMVN location estimator T_{RMVN}, could be used on the $\boldsymbol{x}_i$.

9.3 Repeated Measurements

Repeated measurements = longitudinal data analysis. Take p measurements on the same unit, often the same measurement, e.g., blood pressure, at several time periods. Hence each unit or individual is measured repeatedly over time. The variables are X_1, ..., X_p where often X_k is the measurement at the kth time period. Then $E(\boldsymbol{x}) = (\mu_1, ..., \mu_p)^T = (\mu + \tau_1, ..., \mu + \tau_p)^T$. Let the $(p-1) \times 1$ vector $\boldsymbol{y}_j = (x_{1j} - x_{2j}, x_{2j} - x_{3j}, ..., x_{p-1,j} - x_{pj})^T$ for $j = 1, ..., n$. Hence $y_{ij} = x_{ij} - x_{i+1,j}$ for $j = 1, ..., n$ and $i = 1, ..., p-1$. Then $\overline{\boldsymbol{y}} = (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2, \overline{\boldsymbol{x}}_2 - \overline{\boldsymbol{x}}_3, ..., \overline{\boldsymbol{x}}_{p-1} - \overline{\boldsymbol{x}}_p)^T$. If $\boldsymbol{\mu_y} = E(\boldsymbol{y}_i)$, then $\boldsymbol{\mu_y} = \boldsymbol{0}$ is equivalent to $\mu_1 = \cdots = \mu_p$ where $E(X_k) = \mu_k$. Let $\boldsymbol{S_y}$ be the sample covariance matrix of the $\boldsymbol{y}_i$.

The **large sample repeated measurements test** has four steps.
i) State the hypotheses $H_0 : \boldsymbol{\mu_y} = \boldsymbol{0}$ $\quad H_1 : \boldsymbol{\mu_y} \neq \boldsymbol{0}$.
ii) Find the test statistic $T_R^2 = n\overline{\boldsymbol{y}}^T \boldsymbol{S}_{\boldsymbol{y}}^{-1} \overline{\boldsymbol{y}}$.
iii) Find pval =

$$P\left(\frac{n-p+1}{(n-1)(p-1)}\, T_R^2 < F_{p-1,n-p+1}\right).$$

iv) State whether you fail to reject H_0 or reject H_0. If you reject H_0 then conclude that $\boldsymbol{\mu_y} \neq \boldsymbol{0}$ so not all p of the μ_i are equal, while if you fail to reject H_0 conclude that the population mean $\boldsymbol{\mu_y} = \boldsymbol{0}$ or that there is not enough evidence to conclude that $\boldsymbol{\mu_y} \neq \boldsymbol{0}$. Reject H_0 if pval $\leq \alpha$ and fail to reject H_0 if pval $> \alpha$. Give a nontechnical sentence, if possible.

Example 9.3. Following Morrison (1967, pp. 139–141), reaction times to visual stimuli were obtained for $n = 20$ normal young men under conditions A, B, and C of stimulus display. Let $\overline{x}_A = 21.05, \overline{x}_B = 21.65$, and $\overline{x}_C = 28.95$. Test whether $\mu_A = \mu_B = \mu_C$ if $T_R^2 = 882.8$.

Solution: i) $H_0 : \boldsymbol{\mu_y} = \boldsymbol{0}$ $\quad H_1 : \boldsymbol{\mu_y} \neq \boldsymbol{0}$
ii) $T_R^2 = 882.8$
iii) $\dfrac{n-p+1}{(n-1)(p-1)} T_R^2 = \dfrac{20-3+1}{(20-1)(3-1)} 882.8 = 418.168$, and
pval $= P(418.168 < F_{2,18}) \approx 0$
iv) Reject H_0. The three mean reaction times are different.

An alternative test would use a statistic T, such as the coordinatewise median or RMVN location estimator, on the $\boldsymbol{y}_j$, and the bootstrap method of Section 9.1.2 can be applied with $\boldsymbol{\mu_y} = \boldsymbol{0}$. This test is equivalent to $H_0 : \mu_1 = \cdots = \mu_p$ where μ_k is a population location parameter for the kth measurement. Hence if the coordiatewise median is being used, then μ_k is the population median of the kth measurement.

9.4 Two Samples

Suppose there are two independent random samples $\boldsymbol{x}_{1,1}, ..., \boldsymbol{x}_{n_1,1}$ and $\boldsymbol{x}_{1,2}, ..., \boldsymbol{x}_{n_2,2}$ from populations with mean and covariance matrices $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{\boldsymbol{x}_i})$ for $i = 1, 2$. Assume the $\boldsymbol{\Sigma}_{\boldsymbol{x}_i}$ are positive definite and that it is desired to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ versus $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ where the $\boldsymbol{\mu}_i$ are $p \times 1$ vectors. To simplify large sample theory, assume $n_1 = kn_2$ for some positive real number k.

By the multivariate central limit theorem,

$$\begin{pmatrix} \sqrt{n_1}\,(\overline{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1) \\ \sqrt{n_2}\,(\overline{\boldsymbol{x}}_2 - \boldsymbol{\mu}_2) \end{pmatrix} \xrightarrow{D} N_{2p}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{x}_1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\boldsymbol{x}_2} \end{pmatrix}\right],$$

or

$$\begin{pmatrix} \sqrt{n_2}\,(\overline{\boldsymbol{x}}_1 - \boldsymbol{\mu}_1) \\ \sqrt{n_2}\,(\overline{\boldsymbol{x}}_2 - \boldsymbol{\mu}_2) \end{pmatrix} \xrightarrow{D} N_{2p}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \frac{\boldsymbol{\Sigma}_{\boldsymbol{x}_1}}{k} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\boldsymbol{x}_2} \end{pmatrix}\right].$$

Hence

$$\sqrt{n_2}\,[(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] \xrightarrow{D} N_p(\mathbf{0}, \frac{\boldsymbol{\Sigma}_{\boldsymbol{x}_1}}{k} + \boldsymbol{\Sigma}_{\boldsymbol{x}_2}).$$

Using $n\boldsymbol{B}^{-1} = \left(\frac{\boldsymbol{B}}{n}\right)^{-1}$ and $n_2 k = n_1$, if $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$, then

$$n_2(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)^T \left(\frac{\boldsymbol{\Sigma}_{\boldsymbol{x}_1}}{k} + \boldsymbol{\Sigma}_{\boldsymbol{x}_2}\right)^{-1} (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2) =$$

$$(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)^T \left(\frac{\boldsymbol{\Sigma}_{\boldsymbol{x}_1}}{n_1} + \frac{\boldsymbol{\Sigma}_{\boldsymbol{x}_2}}{n_2}\right)^{-1} (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2) \xrightarrow{D} \chi^2_p.$$

Hence

$$T_0^2 = (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)^T \left(\frac{\boldsymbol{S}_1}{n_1} + \frac{\boldsymbol{S}_2}{n_2}\right)^{-1} (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2) \xrightarrow{D} \chi^2_p.$$

The above result is easily generalized to other statistics. See Rupasinghe Arachchige Don and Pelawa Watagoda (2017). If the sequence of positive integers $d_n \to \infty$ and $Y_n \sim F_{p,d_n}$, then $Y_n \xrightarrow{D} \chi^2_p/p$. Using an F_{p,d_n} distribution instead of a χ^2_p distribution is similar to using a t_{d_n} distribution instead of a standard normal $N(0,1)$ distribution for inference. Instead of rejecting H_0 when $T_0^2 > \chi^2_{p,1-\alpha}$, reject H_0 when

$$T_0^2 > pF_{p,d_n,1-\alpha} = \frac{pF_{p,d_n,1-\alpha}}{\chi^2_{p,1-\alpha}}\chi^2_{p,1-\alpha}.$$

The term $\frac{pF_{p,d_n,1-\alpha}}{\chi^2_{p,1-\alpha}}$ can be regarded as a small sample correction factor that improves the test's performance for small samples. We will use $d_n = \min(n_1 - p, n_2 - p)$. Here $P(Y_n \le \chi^2_{p,\alpha}) = \alpha$ if Y_n has a χ^2_p distribution, and $P(Y_n \le F_{p,d_n,\alpha}) = \alpha$ if Y_n has an F_{p,d_n} distribution.

Let pval denote the estimated pvalue. The four-step test is
i) State the hypotheses $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$.
ii) Find the test statistic $t_0 = T_0^2/p$.
iii) Find pval $= P(t_0 < F_{p,d_n})$.
iv) State whether you fail to reject H_0 or reject H_0. If you reject H_0 then conclude that the population means are not equal while if you fail to reject H_0 conclude that the population means are equal or that there is not enough evidence to conclude that the population means differ. Reject H_0 if pval $\leq \alpha$ and fail to reject H_0 if pval $> \alpha$. Give a nontechnical sentence if possible. As a benchmark for this text, use $\alpha = 0.05$ if α is not given.

Example 9.4. Following Mardia et al. (1979, p. 153), cranial length and breadth (X_1 and X_2) were measured on $n_1 = 35$ female frogs and $n_2 = 14$ male frogs with $\overline{\boldsymbol{x}}_1 = (22.86, 24.397)^T$ and $\overline{\boldsymbol{x}}_2 = (21.821, 22.442)^T$. Test $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ if $T_0^2 = 2.550$.

Solution: i) $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$
ii) $t_0 = T_0^2/p = 2.550/2 = 1.275$
iii) pval $= P(1.275 < F_{2,14-2}) > 0.05$
iv) Fail to reject H_0. There is not enough evidence to conclude that the mean lengths and breadths differ for the male and female frogs.

The plots for the one way MANOVA model in Section 10.2 are also useful for the two-sample Hotelling's T^2 test. An alternative to the above test is to used the pooled covariance matrix. This Hotelling's T^2 test is a special case of the one way MANOVA model with two groups covered in Section 10.3.

9.4.1 Bootstrapping Two-Sample Tests

Bootstrapping the two-sample test is similar to bootstrapping discriminant analysis and one way MANOVA models. Take a sample of size n_i with replacement from random sample i for $i = 1, 2$, and compute $T_{11}^* - T_{21}^*$. Repeat B times to get the bootstrap sample $\boldsymbol{w}_1 = T_{11}^* - T_{21}^*, ..., \boldsymbol{w}_B = T_{1B}^* - T_{2B}^*$. Apply the nonparametric prediction region on the $\boldsymbol{w}_i$, and fail to reject $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ if $\mathbf{0}$ is in the prediction region, and reject H_0, otherwise. See Rupasinghe Arachchige Don and Pelawa Watagoda (2017).

Some *R* output is below for the Gladstone (1905) data where several infants are outliers. We first tested the first 133 cases versus the last 134 cases. It turned out that the first group was younger and had all of the infants, so H_0 was rejected. Then a random sample of 133 was used as the first group and the remaining 134 as the second group. Then the test failed to reject H_0. Using the nominal level $\alpha = 0.05$ of the large sample bootstrap test, reject H_0 if the test statistic is larger than the cutoff, where 4.102 was the cutoff for the first test which used RMVN.

```
zz <- cbrainx[,c(1,3,5,6,7,8,9,11)]
#get rid of qualitative variables
zx <- zz[1:133,]
zy <- zz[134:267,]
out<-rhot2boot(zx,zy,med=F) #RMVN takes a while.
tem<-predreg(out$mus)
> tem$cuplim
   95.4%
4.101788
> tem$D0
[1] 7.529998 #> 4.102 so reject Ho
out<-rhot2boot(zx,zy,med=T) #coord. median is fast
tem<-predreg(out$mus)
> tem$cuplim
   95.4%
4.046958
> tem$D0
[1] 12.87506 #> 4.05 so reject Ho
plot(zx[,1],zy[-134,1])
#zx people tend to be older, infants are in zy
indx <- sample(1:267,133)#random sample for zx and zy
zx <- zz[indx,]
zy <- zz[-indx,]
out<-rhot2boot(zx,zy,med=F)
tem<-predreg(out$mus) #RMVN
> tem$cuplim
   95.4%
4.065357
> tem$D0
[1] 2.94968 #< 4.07 so fail to reject Ho
out<-rhot2boot(zx,zy,med=T)
tem<-predreg(out$mus) #coord. median
> tem$cuplim
   95.4%
3.915687
> tem$D0
[1] 2.802046 #< 3.92 so fail to reject Ho
```

9.5 Summary

1) The one-sample Hotelling's T^2 test is used to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_A : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. The test rejects H_0 if $T^2_H = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^T \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) > \frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha}$ where $P(Y \leq F_{p,d,\alpha}) = \alpha$ if $Y \sim F_{p,d}$.

If a multivariate location estimator T satisfies $\sqrt{n}(T - \boldsymbol{\mu}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{D})$, then a competing test rejects H_0 if $T^2_C = n(T - \boldsymbol{\mu}_0)^T \hat{\boldsymbol{D}}^{-1}(T - \boldsymbol{\mu}_0) > \frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha}$ where $\hat{\boldsymbol{D}}$ is a consistent estimator of $\boldsymbol{D}$. The scaled F cutoff can be used since $T^2_C \xrightarrow{D} \chi^2_p$ if H_0 holds, and $\frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha} \rightarrow \chi^2_{p,1-\alpha}$ as $n \rightarrow \infty$.

2) Let pval be an estimate of the pvalue. As a benchmark for hypothesis testing, use $\alpha = 0.05$ if α is not given.

3) Typically, use $T^2_C = T^2_H$ in the following four-step **one-sample Hotelling's T^2_C test**. i) State the hypotheses $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$.
ii) Find the test statistic $T^2_C = n(T - \boldsymbol{\mu}_0)^T \hat{\boldsymbol{D}}^{-1}(T - \boldsymbol{\mu}_0)$.
iii) Find pval =

$$P\left(\frac{n-p}{(n-1)p} \ T^2_C < F_{p,n-p}\right).$$

iv) State whether you fail to reject H_0 or reject H_0. If you reject H_0 then conclude that $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ while if you fail to reject H_0 conclude that the population mean $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ or that there is not enough evidence to conclude that $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. Reject H_0 if pval $\leq \alpha$ and fail to reject H_0 if pval $> \alpha$.

4) The multivariate matched pairs test is used when there are $k = 2$ treatments applied to the same n cases with the same p variables used for each treatment. Let $\boldsymbol{y}_i$ be the p variables measured for treatment 1 and $\boldsymbol{z}_i$ be the p variables measured for treatment 2. Let $\boldsymbol{x}_i = \boldsymbol{y}_i - \boldsymbol{z}_i$. Let $\boldsymbol{\mu} = E(\boldsymbol{x}) = E(\boldsymbol{y}) - E(\boldsymbol{z})$. We want to test if $\boldsymbol{\mu} = \boldsymbol{0}$, so $E(\boldsymbol{y}) = E(\boldsymbol{z})$. The test can also be used if $(\boldsymbol{y}_i, \boldsymbol{z}_i)$ are matched (highly dependent) in some way. For example, if identical twins are in the study, $\boldsymbol{y}_i$ and $\boldsymbol{z}_i$ could be the measurements on each twin. Let $(\overline{\boldsymbol{x}}, \boldsymbol{S}_{\boldsymbol{x}})$ be the sample mean and covariance matrix of the $\boldsymbol{x}_i$.

5) The **large sample multivariate matched pairs test** has four steps.
i) State the hypotheses $H_0 : \boldsymbol{\mu} = \boldsymbol{0} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{0}$.
ii) Find the test statistic $T^2_M = n\overline{\boldsymbol{x}}^T \boldsymbol{S}^{-1}_{\boldsymbol{x}} \overline{\boldsymbol{x}}$.
iii) Find pval =

$$P\left(\frac{n-p}{(n-1)p} \ T^2_M < F_{p,n-p}\right).$$

iv) State whether you fail to reject H_0 or reject H_0. If you reject H_0 then conclude that $\boldsymbol{\mu} \neq \boldsymbol{0}$ while if you fail to reject H_0 conclude that the population mean $\boldsymbol{\mu} = \boldsymbol{0}$ or that there is not enough evidence to conclude that $\boldsymbol{\mu} \neq \boldsymbol{0}$.

Reject H_0 if pval $\leq \alpha$ and fail to reject H_0 if pval $> \alpha$. Give a nontechnical sentence if possible.

6) Repeated measurements = longitudinal data analysis. Take p measurements on the same unit, often the same measurement, e.g., blood pressure, at several time periods. The variables are X_1, ..., X_p where often X_k is the measurement at the kth time period. Then $E(\boldsymbol{x}) = (\mu_1, ..., \mu_p)^T = (\mu + \tau_1, ..., \mu + \tau_p)^T$. Let $\boldsymbol{y}_j = (x_{1j} - x_{2j}, x_{2j} - x_{3j}, ..., x_{p-1,j} - x_{pj})^T$ for $j = 1, ..., n$. Then $\overline{\boldsymbol{y}} = (\overline{x}_1 - \overline{x}_2, \overline{x}_2 - \overline{x}_3, ..., \overline{x}_{p-1} - \overline{x}_p)^T$. If $\boldsymbol{\mu_y} = E(\boldsymbol{y}_i)$, then $\boldsymbol{\mu}_Y = \boldsymbol{0}$ is equivalent to $\mu_1 = \cdots = \mu_p$ where $E(X_k) = \mu_k$. Let $\boldsymbol{S_y}$ be the sample covariance matrix of the $\boldsymbol{y}_i$.

7) The **large sample repeated measurements test** has four steps.
i) State the hypotheses $H_0 : \boldsymbol{\mu}_y = \boldsymbol{0} \quad H_1 : \boldsymbol{\mu}_y \neq \boldsymbol{0}$.
ii) Find the test statistic $T_R^2 = n\overline{\boldsymbol{y}}^T \boldsymbol{S}_y^{-1} \overline{\boldsymbol{y}}$.
iii) Find pval =

$$P\left(\frac{n-p+1}{(n-1)(p-1)}\; T_R^2 < F_{p-1,n-p+1}\right).$$

iv) State whether you fail to reject H_0 or reject H_0. If you reject H_0 then conclude that $\boldsymbol{\mu}_y \neq \boldsymbol{0}$ while if you fail to reject H_0 conclude that the population mean $\boldsymbol{\mu}_y = \boldsymbol{0}$ or that there is not enough evidence to conclude that $\boldsymbol{\mu}_y \neq \boldsymbol{0}$. Reject H_0 if pval $\leq \alpha$ and fail to reject H_0 if pval $> \alpha$. Give a nontechnical sentence, if possible.

8) The F tables give left tail area and the pval is a right tail area. The Section 15.5 table gives $F_{k,d,0.95}$. If $\alpha = 0.05$ and $\dfrac{n-p}{(n-1)p}\; T_C^2 < F_{k,d,0.95}$, then fail to reject H_0. If $\dfrac{n-p}{(n-1)p}\; T_C^2 \geq F_{k,d,0.95}$, then reject H_0.

a) For the one-sample Hotelling's T_C^2 test and the matched pairs T_M^2 test, $k = p$ and $d = n - p$.

b) For the repeated measures T_R^2 test, $k = p - 1$ and $d = n - p + 1$.

9) If $n \geq 10p$, the tests in 3), 5), and 7) are robust to nonnormality. For the one-sample Hotelling's T_C^2 test and the repeated measurements test, make a DD plot. For the multivariate matched pairs test, make a DD plot of the $\boldsymbol{x}_i$, of the $\boldsymbol{y}_i$, and of the $\boldsymbol{z}_i$.

10) Suppose there are two independent random samples $\boldsymbol{x}_{1,1}, ..., \boldsymbol{x}_{n_1,1}$ and $\boldsymbol{x}_{1,2}, ..., \boldsymbol{x}_{n_2,2}$ from populations with mean and covariance matrices $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{\boldsymbol{x}_i})$ for $i = 1, 2$ where the $\boldsymbol{\mu}_i$ are $p \times 1$ vectors. Let $d_n = \min(n_1 - p, n_2 - p)$. The **large sample two-sample Hotelling's T_0^2 test** is a four-step test:
i) State the hypotheses $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$.
ii) Find the test statistic $t_0 = T_0^2/p$.
iii) Find pval = $P(t_0 < F_{p,d_n})$.
iv) State whether you fail to reject H_0 or reject H_0. If you reject H_0 then conclude that the population means are not equal while if you fail to reject

H_0 conclude that the population means are equal or that there is not enough evidence to conclude that the population means differ. Reject H_0 if pval $\leq \alpha$ and fail to reject H_0 if pval $> \alpha$. Give a nontechnical sentence if possible.

9.6 Complements

The *mpack* function `rhotsim` is useful for simulating the robust diagnostic for the one-sample Hotelling's T^2 test. See Zhang (2011) for more simulations. Willems et al. (2002) used similar reasoning to present a diagnostic based on the FMCD estimator.

Yao (1965) suggested a more complicated denominator degrees of freedom than $d_n = \min(n_1 - p, n_2 - p)$ for the two-sample Hotelling's T^2 test. Good (2012, pp. 55–57), which provides randomization tests as competitors for the two-sample Hotelling's T^2 test. Bootstrapping the tests with robust estimators seems to be effective. For bootstrapping the two-sample Hotelling's T^2 test, see Rupasinghe Arachchige Don and Pelawa Watagoda (2017). Gregory et al. (2015) and Feng and Sun (2015) considered the two-sample test when $p \geq n$.

9.7 Problems

PROBLEMS WITH AN ASTERISK * ARE ESPECIALLY USEFUL.

9.1. Following Morrison (1967, pp. 122–123), the Wechsler Adult Intelligence Scale scores of $n = 101$ subjects aged 60 to 64 were recorded, giving a verbal score (X_1) and performance score (X_2) for each subject. Suppose $\boldsymbol{\mu}_0 = (60, 50)^T$ and $T_C^2 = 357.43$. Perform the one-sample Hotelling's T^2 test.

9.2. Following Morrison (1967, pp. 137–138), the levels of free fatty acid (FFA) in the blood were measured in $n = 15$ hypnotized normal volunteers who had been asked to experience fear, depression, and anger effects while in the hypnotic state. The mean FFA changes were $\overline{x}_1 = 2.669$, $\overline{x}_2 = 2.178$, and $\overline{x}_3 = 2.558$. Let $\mu_F = \mu + \tau_1$, $\mu_D = \mu + \tau_2$, and $\mu_A = \mu + \tau_3$. We want to know if the mean stress FFA changes were equal. So test whether $\mu_F = \mu_D = \mu_F$ if $T_R^2 = 2.68$.

9.3. Data is taken or modified from Johnson and Wichern (1988, pp. 185, 224).

a) Suppose $S_2^2 = S_{22} = 126.05$, $\overline{x}_2 = 54.69$, $n = 87$, and $p = 3$. Find a large sample simultaneous 95% CI for μ_2.

b) Suppose a random sample of 50 bars of soap from method 1 and a random sample of 50 bars of soap from method 2 are obtained. Let X_1 = lather and X_2 = mildness with $\overline{\boldsymbol{x}}_1 = (8.4, 4.1)^T$ and $\overline{\boldsymbol{x}}_2 = (10.2, 3.9)^T$. Test $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ if $T_0^2 = 52.4722$.

R Problems

Warning: Use the command *source("G:/mpack.txt")* **to download the programs. See Preface or Section 15.2** Typing the name of the `mpack` function, e.g., *rhotsim*, will display the code for the function. Use the `args` command, e.g., *args(rhotsim)*, to display the needed arguments for the function. For some of the following problems, the R commands can be copied and pasted from (http://lagrange.math.siu.edu/Olive/mrsashw.txt) into R.

9.4*. Use the R commands in Subsection 9.1.1 to make a plot similar to Figure 9.1. The program may take a minute to run.

9.5. Conjecture:

$$\sqrt{n}(T_{RMVN} - \boldsymbol{\mu}) \xrightarrow{D} N_p(\boldsymbol{0}, \tau_p \boldsymbol{\Sigma})$$

for a wide variety of elliptically contoured distributions where τ_p depends on both p and the underlying distribution. The following "test" is based on a conjecture and should be used as an outlier diagnostic rather than for inference. The ad hoc "test" that rejects H_0 if

$$\frac{T_R^2}{f_{n,p}} = n(T_{RMVN} - \boldsymbol{\mu}_0)^T \hat{\boldsymbol{C}}_{RMVN}^{-1}(T_{RMVN} - \boldsymbol{\mu}_0)/f_{n,p} > \frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha}$$

where $f_{n,p} = 1.04 + 0.12/p + (40 + p)/n$. The simulations use $n = 150$ and $p = 10$.

a) The R commands for this part use simulated data is

$$\boldsymbol{x}_i \sim N_p(\boldsymbol{0}, diag(1, 2, ..., p))$$

where $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ is being tested with 5000 runs at a nominal level of 0.05. So H_0 is true, and hcv and rhcv are the proportion of rejections by the T_H^2 test and by the ad hoc robust test. We want hcv and rhcv near 0.05. THIS SIMULATION MAY TAKE A FEW MINUTES. Record hcv and rhcv. Were hcv and rhcv near 0.05?

b) The R commands for this part use simulated data

$$\boldsymbol{x}_i \sim N_p(\delta\boldsymbol{1}, diag(1, 2, ..., p))$$

where $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ is being tested with 5000 runs at a nominal level of 0.05. In the simulation, $\delta = 0.2$, so H_0 is false, and hcv and rhcv are the proportion of rejections by the T_H^2 test and by the ad hoc robust test. We want hcv and rhcv near 1 so that the power is high. Paste the output into *Word*. THIS SIMULATION MAY TAKE A FEW MINUTES. Record hcv and rhcv. Were hcv and rhcv near 1?