Chapter 9

Hotelling's T^2 Test

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

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

if H_0 holds and $\hat{\boldsymbol{D}}$ 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

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

as $n \to \infty$. 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 $n \rightarrow \infty$. Want $n > 10p$ 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^2_p approximation to T^2_H is poor.

Let pval be an estimate of the pvalue. Typically use $T^2_C = T^2_H$ in the following 4 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 $T^2_C = n(T - \boldsymbol{\mu}_0)^T \hat{\boldsymbol{D}}^{-1}(T - \boldsymbol{\mu}_0)$.
iii) Find pval =

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

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 $< \alpha$ and fail to reject H_0 if pval $\geq \alpha$. 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

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

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 $n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})^T \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) = D^2_{\boldsymbol{\mu}}(\overline{\boldsymbol{x}}, \boldsymbol{S}/n)$. Hence $\boldsymbol{\mu}$ that are close to $\overline{\boldsymbol{x}}$ with respect to the Mahalanobis distance based on dispersion matrix $\boldsymbol{S}/n$ are 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 $\boldsymbol{a} \neq \boldsymbol{0}$, $P(L_{\boldsymbol{a}} < \boldsymbol{a}^T\boldsymbol{\mu} < U_{\boldsymbol{a}}) \rightarrow 1-\alpha$ as $n \rightarrow \infty$ 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, p. 184-5), the p confidence intervals (CIs) for μ_i and $p(p-1)/2$ CIs for $\mu_i - \mu_k$ can be made such that they all hold simultaneously with confidence $\rightarrow 1-\alpha$. Hence if $\alpha = 0.05$, then in 100 samples, expect all $p + p(p-1)/2$ CIs to contain μ_i and $\mu_i - \mu_k$ about 95 times while about 5 times at least one of the CIs will fail to contain its parameter. The 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 CIs for $\mu_i - \mu_k$ are

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

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, p. 1651-1652) shows that if (E1) holds, then the classical estimator applied to cases with $D_i(\overline{\boldsymbol{x}}, S) \leq h$ is asymptotically normal with

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

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:

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

for a wide variety of elliptically contoured distributions where τ_p depends on both p and the underlying distribution. 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

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

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

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.

```
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)
```

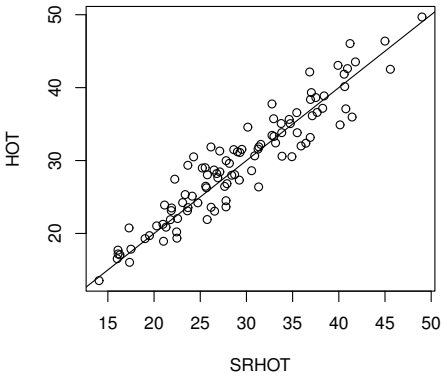

Figure 9.1: Scaled "Robust" Statistic Versus T_H^2 Statistic

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 (number of rejections is less than the nominal 0.05) except when $n = 15p$ and $75 \leq p \leq 100$. 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.

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

9.2 Matched Pairs

Assume that there are $k = 2$ treatments, and both treatments are given to the same n cases or units. For example, systolic and diastolic blood pressure

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

could be compared before and after the patient (case) receives blood pressure medication. Then $p = 2$. Alternatively use m 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 $\boldsymbol{y}_i = (Y_{i1}, Y_{i2}, ..., Y_{ip})^T$ denote the p measurements from the 1st treatment, and $\boldsymbol{z}_i = (Z_{i1}, Z_{i2}, ..., Z_{ip})^T$ denote 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 $\boldsymbol{\Sigma_x}$. 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 $P(F_{p,n} \leq F_{p,n,\delta}) = \delta$. Then the one sample Hotelling's T^2 inference is done on the differences $\boldsymbol{x}_i$ using m instead of n and using $\boldsymbol{\mu}_0 = \boldsymbol{0}$. If the p random variables are continuous, make 3 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 4 steps.
i) State the hypotheses $H_0 : \boldsymbol{\mu} = \boldsymbol{0} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{0}$.
ii) Find the test statistic $T^2 = n\overline{\boldsymbol{x}}^T \boldsymbol{S}^{-1} \overline{\boldsymbol{x}}$.
iii) Find pval =

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

iv) State whether you fail to reject H_0 or reject H_0. If you reject H_0 then conclude that $\boldsymbol{\mu} \neq \boldsymbol{0}$ 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 $< \alpha$ and fail to reject H_0 if pval $\geq \alpha$. 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

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

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 $\boldsymbol{S}$.

9.3 Repeated Measurements

Repeated measurements = longitudinal data analysis. Take p measurements on the same unit, often the same measurement, eg 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. The $E(\boldsymbol{x}) = (\mu_1, ..., \mu_p)^T = (\mu + \tau_1, ..., \mu + \tau_p)^T$. Let $y_{ij} = x_{ij} - x_{i,j+1}$ for $i = 1, ..., n$ and $j = 1, ..., p-1$. 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 4 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 $< \alpha$ and fail to reject H_0 if pval $\geq \alpha$. Give a nontechnical sentence, if possible.

9.4 Two Samples

Suppose there are two independent random samples $X_{1,1}, ..., X_{n_1,1}$ and $X_{1,2}, ..., X_{n_2,2}$ 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,

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

or

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

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

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

If the sequence of positive integer $d_n \to \infty$ 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 $P(Y_n \leq \chi^2_{p,\alpha}) = \alpha$ if Y_n has a χ^2_p distribution, and $P(Y_n \leq F_{p,d_n,\alpha}) = \alpha$ if Y_n has an F_{p,d_n} distribution.

Let pval denote the estimated pvalue. The 4 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 $< \alpha$ and fail to reject H_0 if pval $\geq \alpha$. Give a nontechnical sentence if possible. As a benchmark for this text, use $\alpha = 0.05$ if α is not given.

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 $T_H^2 = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^T \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) > \frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha}$ 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 $T_C^2 = n(T-\boldsymbol{\mu}_0)^T \hat{\boldsymbol{D}}^{-1}(T-\boldsymbol{\mu}_0) > \frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha}$ if H_0 holds and $\hat{\boldsymbol{D}}$ 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 $\frac{(n-1)p}{n-p} F_{p,n-p,1-\alpha} \to \chi^2_{p,1-\alpha}$ as $n \to \infty$.

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 $T_C^2 = T_H^2$ in the following 4 step **one sample Hotelling's T_C^2 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 =

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

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 $< \alpha$ and fail to reject H_0 if pval $\geq \alpha$.

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 $\boldsymbol{\mu} = E(\boldsymbol{x}) = E(\boldsymbol{y}) - E(\boldsymbol{z})$. Want to test if $\boldsymbol{\mu} = \boldsymbol{0}$, so $E(\boldsymbol{y}) = E(\boldsymbol{z})$. The test can also be used if $(\boldsymbol{x}_i, \boldsymbol{y}_i)$ are matched (highly dependent) in some way. For example if identical twins are in the study, $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ could be the

measurements on each twin. Let $(\overline{\boldsymbol{x}}, \boldsymbol{S}_x)$ be the sample mean and covariance matrix of the $\boldsymbol{x}_i$.

5) The **large sample multivariate matched pairs test** has 4 steps.
i) State the hypotheses $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ $\quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{0}$.
ii) Find the test statistic $T_M^2 = n\overline{\boldsymbol{x}}^T \boldsymbol{S}_x^{-1} \overline{\boldsymbol{x}}$.
iii) Find pval =

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

iv) State whether you fail to reject H_0 or reject H_0. If you reject H_0 then conclude that $\boldsymbol{\mu} \neq \boldsymbol{0}$ 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 $< \alpha$ and fail to reject H_0 if pval $\geq \alpha$. Give a nontechnical sentence if possible.

6) Repeated measurements = longitudinal data analysis. Take p measurements on the same unit, often the same measurement, eg 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. The $E(\boldsymbol{x}) = (\mu_1, ..., \mu_p)^T = (\mu + \tau_1, ..., \mu + \tau_p)^T$. Let $y_{ij} = x_{ij} - x_{i,j+1}$ for $i = 1, ..., n$ and $j = 1, ..., p-1$. 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 4 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 $< \alpha$ and fail to reject H_0 if pval $\geq \alpha$. Give a nontechnical sentence, if possible.

8) The F tables give left tail area and the pval is a right tail area. Table 15.5 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 $n > 10p$, the tests in 89), 91) and 93) 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 $X_{1,1}, ..., X_{n_1,1}$ and $X_{1,2}, ..., X_{n_2,2}$ 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 4 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 $< \alpha$ and fail to reject H_0 if pval $\geq \alpha$. Give a nontechnical sentence if possible.

11) Tests for covariance matrices are very nonrobust to nonnormality. Let a plot of x versus y have x on the horizontal axis and y on the vertical axis. A good diagnostic is to use the DD plot. So a diagnostic for $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{\Sigma}_0$ is to plot $D_i(\overline{\boldsymbol{x}}, \boldsymbol{S})$ versus $D_i(\overline{\boldsymbol{x}}, \boldsymbol{\Sigma}_0)$ for $i = 1, ..., n$. If $n > 10p$ and H_0 is true, then the plotted points in the DD plot should cluster tightly about the identity line.

12) A test for sphericity is a test of $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}} = d\boldsymbol{I}_p$ for some unknown constant $d > 0$. As a diagnostic, make a "DD plot" of $D_i^2(\overline{\boldsymbol{x}}, \boldsymbol{S})$ versus $D_i^2(\overline{\boldsymbol{x}}, \boldsymbol{I}_p)$. If $n > 10p$ and H_0 is true, then the plotted points in the "DD plot" should cluster tightly about the line through the origin with slope d.

13) Now suppose there are k samples, and want to test $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}_1} = \cdots = \boldsymbol{\Sigma}_{\boldsymbol{x}_k}$, that is, all k populations have the same covariance matrix. As a diagnostic, make a DD plot of $D_i(\overline{\boldsymbol{x}}_j, \boldsymbol{S}_j)$ versus $D_i(\overline{\boldsymbol{x}}_j, \boldsymbol{S}_{pool})$ for $j = 1, ..., k$ and $i = 1, ..., n_i$.

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, Pison, Rousseeuw, and Van Aelst (2002) use similar reasoning to present a diagnostic based on the FMCD estimator.

Yao (1965) suggests 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, p. 55-57) provides randomization tests as competitors for the two sample Hotelling's T^2 test.

9.7 Problems

PROBLEMS WITH AN ASTERISK * ARE ESPECIALLY USEFUL.

R/Splus 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, eg *ddplot*, will display the code for the function. Use the `args` command, eg *args(ddplot)*, to display the needed arguments for the function.

9.1*. Use the *R* commands in Subsection 1.1.1 to make a plot similar to Figure 9.1.

9.2. Conjecture:

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

for a wide variety of elliptically contoured distributions where τ_p depends on both p and the underlying distribution. 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

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

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 $\boldsymbol{x}_i \sim N_p(\mathbf{0}, diag(1, 2, ..., p))$ where $H_0 : \boldsymbol{\mu} = \mathbf{0}$ 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. Want hcv and rhcv near 0.05. THIS SIMULATION WILL TAKE ABOUT 5 MINUTES. Record hcv and rhcv. Were hcv and rhcv near 0.05?

b) The R commands for this part use simulated data is $\boldsymbol{x}_i \sim N_p(\delta\mathbf{1}, diag(1, 2, ..., p))$ where $H_0 : \boldsymbol{\mu} = \mathbf{0}$ is being tested with 5000 runs at a nominal level of 0.05. In 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. Want hcv and rhcv near 1 so that the power is high. Paste the output into *Word*. THIS SIMULATION WILL TAKE ABOUT 5 MINUTES. Record hcv and rhcv. Were hcv and rhcv near 1?