Topic 4: Confidence intervals and hypothesis tests for the mean vector

Hypothesis tests for the multivariate normal mean

In this topic, we will begin to apply the multivariate normal distribution to answer questions based on the population mean. Here, we will consider confidence intervals and hypothesis tests about μ, as well as how to compare means of several populations.

Hotelling's T^2

Suppose again that, Like in Topic 3 (slide "Estimation of the Mean Vector and Covariance Matrix of Multivariate Normal Distribution"), we have observed n independent realisations of p-dimensional random vectors from $N_p(\boldsymbol{\mu}, \Sigma)$. Suppose for simplicity that Σ is non-singular. The data matrix has the form

$$\boldsymbol{X} = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1j} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2j} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{i1} & X_{i2} & \cdots & X_{ij} & \cdots & X_{in} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pj} & \cdots & X_{pn} \end{pmatrix} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n]$$

Based on our knowledge from Topic 3 of this week (slide "Distributions of MLE of mean vector and covariance matrix of multivariate normal distribution"), we can claim that $\bar{\boldsymbol{X}} \sim N_p(\boldsymbol{\mu}, \frac{1}{n}\Sigma)$ and $n\hat{\boldsymbol{\Sigma}} \sim W_p(\Sigma, n-1)$.

Consequently, any linear combination $\boldsymbol{c}^\top \bar{\boldsymbol{X}}, \boldsymbol{c} \neq 0 \in \mathbb{R}^p$ follows $N(\boldsymbol{c}^\top \boldsymbol{\mu}, \frac{1}{n}\boldsymbol{c}^\top \Sigma \boldsymbol{c})$ and the quadratic form $n\boldsymbol{c}^\top \hat{\boldsymbol{\Sigma}} \boldsymbol{c} / \boldsymbol{c}^\top \Sigma \boldsymbol{c} \sim \chi^2_{n-1}$. Further, we have shown that $\bar{\boldsymbol{X}}$ and $\hat{\boldsymbol{\Sigma}}$ are independently distributed and hence

$$T = \sqrt{n}\boldsymbol{c}^\top (\bar{\boldsymbol{X}} - \boldsymbol{\mu}) / \sqrt{\boldsymbol{c}^\top \frac{n}{n-1} \hat{\boldsymbol{\Sigma}} \boldsymbol{c}} \sim t_{n-1},$$

i.e. follows the t distribution with $n-1$ degrees of freedom. This result has useful applications in testing for contrasts.

Indeed, if we would like to test $H_0 : \boldsymbol{c}^\top \boldsymbol{\mu} = \sum_{i=1}^p c_i \mu_i = 0$, we note that under H_0, T becomes simply

$$T = \sqrt{n}\boldsymbol{c}^\top \bar{\boldsymbol{X}} / \sqrt{\boldsymbol{c}^\top \boldsymbol{S}\boldsymbol{c}},$$

it does not involve the unknown $\boldsymbol{\mu}$ and can be used as a test-statistic whose distribution under H_0 is known. If $|T| > t_{1-\alpha/2,n-1}$ we should reject H_0 in favour of $H_1 : \boldsymbol{c}^\top \mu = \sum_{i=1}^p c_i\mu_i \neq 0.$

The formulation of the test for other (one-sided) alternatives is left for you as an exercise.

More often we are interested in testing the mean vector of a multivariate normal. First consider the case of *known* covariance matrix Σ (variance σ^2 in the univariate case). The standard univariate ($p = 1$) test for this purpose is the following: to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ at level of significance α, we look at $U = \sqrt{n}\frac{\bar{X}-\mu_0}{\sigma}$ and reject H_0 if $|U|$ exceeds the upper $\frac{\alpha}{2} \cdot 100\%$ point of the standard normal distribution. Checking if $|U|$ is large enough is equivalent to checking if $U^2 = n(\bar{X} - \mu_0)(\sigma^2)^{-1}(\bar{X} - \mu_0)$ is large enough. We can now easily generalise the above test statistic in a natural way for the multivariate ($p > 1$) case: calculate $U^2 = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^\top \Sigma^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)$ and reject the null hypothesis of $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ when U^2 is large enough. Similarly to the proof of *Property 5* of the multivariate normal distribution (Topic 2, slide "Properties of multivariate normal") and by using Theorem 1.4 of Topic 3 (slide "Distributions of MLE of mean vector and covariance matrix of multivariate normal distribution"), you can convince yourself that $U^2 \sim \chi^2_p$ under the null hypothesis. Hence, tables of the χ^2-distribution will suffice to perform the above test in the multivariate case.

Now let us turn to the (practically more relevant) case of unknown covariance matrix Σ. The standard univariate ($p = 1$) test for this purpose is the t-test. Let us recall it: to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ at level of significance α, we look at

$$T = \sqrt{n}\frac{\bar{X} - \mu_0}{S}, \; S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

and reject H_0 if $|T|$ exceeds the upper $\frac{\alpha}{2} \cdot 100\%$ point of the t-distribution with $n - 1$ degrees of freedom. We note that checking if $|T|$ is large enough is equivalent to checking if $T^2 = n(\bar{X} - \mu_0)(s^2)^{-1}(\bar{X} - \mu_0)$ is large enough. Of course, under H_0, the statistic T^2 is F-distributed: $T^2 \sim F_{1,n-1}$ which means that H_0 would be rejected at level α when $T^2 > F_{\alpha;1,n-1}$. We can now easily generalise the above test statistic in a natural way for the multivariate ($p > 1$) case:

Definition 1.3 (Hotelling's T^2). The statistic

$$T^2 = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^\top \boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0) \tag{1.14}$$

where $\bar{\boldsymbol{X}} = \frac{1}{n}\sum_{j=1}^n \boldsymbol{X}_i$, $\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^n(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top$, $\boldsymbol{\mu}_0 \in \mathbb{R}^p$, $\boldsymbol{X}_i \in \mathbb{R}^p, i = 1, \ldots, n$ is named after Harold Hotelling.

Sampling distribution of T^2

Obviously, the test procedure based on Hotelling's statistic will reject the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ if the value of T^2 is sufficiently high. It turns out we do not need special tables for the distribution of T^2 under the null hypothesis because of the following basic result (that represents a true generalisation of the univariate ($p = 1$) case:

Theorem 1.5. *Under the null hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, Hotelling's T^2 is distributed as $\frac{(n-1)p}{n-p} F_{p,n-p}$ where $F_{p,n-p}$ denotes the F-distribution with p and $n-p$ degrees of freedom.*

Noncentral Wishart

It is possible to extend the definition of the Wishart distribution (see Definition 1.2) by allowing the random vectors $\boldsymbol{Y}_i, i = 1, \ldots, n$ there to be independent with $N_p(\boldsymbol{\mu}_i, \Sigma)$ (instead of just having all $\boldsymbol{\mu}_i = \mathbf{0}$). One arrives at the *noncentral* Wishart distribution with parameters $\Sigma, p, n-1, \Gamma$ in that way (denoted also as $W_p(\Sigma, n-1, \Gamma)$. Here $\Gamma = MM^\top \in \mathcal{M}_{p,p}$, $M = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_n]$ is called a *noncentrality parameter*. When all columns of $M \in \mathcal{M}_{p,n}$ are zero, this is the usual (*central*) Wishart distribution. Theorem 1.5 can be extended to derive the distribution of the T^2 statistic under alternatives, i.e. the distribution of $T^2 = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu})^\top \boldsymbol{S}^{-1}(\bar{\boldsymbol{X}} - \boldsymbol{\mu})$ for $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. This distribution turns out to be related to *noncentral F-distribution*. It is helpful in studying power of the test of $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$.

This section is continued on the next slide.

Hypothesis tests for the multivariate normal mean continued

T^2 as a likelihood ratio statistic

It is worth mentioning that Hotelling's T^2 that we introduced by *analogy* with the univariate squared t statistic can in fact also be derived as the *likelihood ratio test statistic* for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. This safeguards the asymptotic optimality of the test suggested above. To see this, first recall the likelihood function (1.9)

$$L(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}i-\boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{x}i-\boldsymbol{\mu})}$$

Its unconstrained maximisation gives as a maximum value:

$$L(\boldsymbol{x}; \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \frac{1}{(2\pi)^{\frac{np}{2}} |\hat{\boldsymbol{\Sigma}}|^{\frac{n}{2}}} \mathrm{e}^{-\frac{np}{2}}$$

On the other hand, under H_0 :

$$\max_{\Sigma} L(x; \boldsymbol{\mu}_0, \Sigma) = \max_{\Sigma} \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu}_0)^\top \Sigma^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_0)}$$

Since $\log L(\boldsymbol{x}; \boldsymbol{\mu}_0, \Sigma) = -\frac{np}{2}\log(2\pi) - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\operatorname{tr}[\Sigma^{-1}(\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}_0)(\boldsymbol{x}_i - \boldsymbol{\mu}_0)^\top)]$, on applying Anderson's lemma (see Theorem 1.2) we find that maximum of $\log L(\boldsymbol{x}; \boldsymbol{\mu}_0, \Sigma)$ (whence also of $L(\boldsymbol{x}; \boldsymbol{\mu}_0, \Sigma)$) is obtained when $\hat{\boldsymbol{\Sigma}}_0 = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu}_0)(\boldsymbol{x}_i - \boldsymbol{\mu}_0)^\top$ and the maximal value is

$$\frac{1}{(2\pi)^{\frac{np}{2}} |\hat{\boldsymbol{\Sigma}}_0|^{\frac{n}{2}}} \mathrm{e}^{-\frac{np}{2}} .$$

Hence the likelihood ratio is:

$$\Lambda = \frac{\max_{\Sigma} L(x; \boldsymbol{\mu}_0, \Sigma)}{\max_{\boldsymbol{\mu},\Sigma} L(x; \boldsymbol{\mu}, \Sigma)} = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{\frac{n}{2}} \tag{1.15}$$

The equivalent statistic $\Lambda^{\frac{2}{n}} = \frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}$ is called *Wilks' lambda*. Small values of Wilks' lambda lead to rejecting $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$.

Wilks' lambda and T^2

The following theorem shows the relation between Wilks' lambda and T^2:

Theorem 1.6. *The likelihood ratio test is equivalent to the test based on T^2 since $\Lambda^{\frac{2}{n}} = (1 + \frac{T^2}{n-1})^{-1}$*

holds.

Numerical calculation of T^2

Hence H_0 is rejected for small values of $\Lambda^{\frac{2}{n}}$ or equivalently, for large values of T^2. The critical values for T^2 are determined from Theorem 1.5. Relation in Theorem 1.6 can be used to calculate T^2 from $\Lambda^{\frac{2}{n}} = \frac{|\hat{\mathbf{\Sigma}}|}{|\hat{\mathbf{\Sigma}}_0|}$ thus avoiding the need to invert the matrix $\boldsymbol{S}$ when calculating T^2!

Asymptotic distribution of T^2

Since $\boldsymbol{S}^{-1}$ is a consistent estimator of Σ^{-1}, the limiting distribution of T^2 will coincide with the one of $n(\bar{\boldsymbol{x}} - \boldsymbol{\mu})^\top \Sigma^{-1}(\bar{\boldsymbol{x}} - \boldsymbol{\mu})$ which, as we know already, is χ^2_p. This coincides with a general claim of asymptotic theory which states that $-2\log\Lambda$ is asymptotically distributed as χ^2_p. Indeed:

$$-2\log\Lambda = n\log(1 + \frac{T^2}{n-1}) \approx \frac{n}{n-1}T^2 \approx T^2$$

(by using the fact that $\log(1+x) \approx x$ for small x).

Confidence regions for the mean vector and for its components

Confidence region for the mean vector

For a given confidence level $(1-\alpha)$ it can be constructed in the form

$$\left\{\boldsymbol{\mu}|n(\bar{\boldsymbol{x}}-\boldsymbol{\mu})^\top \boldsymbol{S}^{-1}(\bar{\boldsymbol{x}}-\boldsymbol{\mu}) \leq F_{1-\alpha,p,n-p}\frac{p}{n-p}(n-1)\right\}$$

where $F_{1-\alpha,p,n-p}$ is the upper $\alpha \cdot 100\%$ percentage point of the F distribution with $(p, n-p)$ df. This confidence region has the form of an *ellipsoid* in $\mathbb{R}^p$ centred at $\bar{\boldsymbol{x}}$. The axes of this *confidence ellipsoid* are directed along the eigenvectors $\boldsymbol{e}_i$ of the matrix $S = \frac{1}{n-1}\sum_{i=1}^n(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})^\top$. The half-lengths of the axes are given by the expression $\sqrt{\lambda_i}\sqrt{\frac{p(n-1)F_{1-\alpha,p,n-p}}{n(n-p)}}$, with $\lambda_i, i=1,\dots,p$ being the corresponding eigenvalues, i.e.

$$S\boldsymbol{e}_i = \lambda_i \boldsymbol{e}_i, i=1,\dots,p$$

For illustration, see numerical example 5.3, pages 221–223, Johnson and Wichern.

Simultaneous confidence statements

For a given confidence level $(1-\alpha)$ the confidence ellipsoids in the section above correctly reflect the joint (multivariate) knowledge about plausible values of $\boldsymbol{\mu} \in \mathbb{R}^p$ but nevertheless one is often interested in confidence intervals for means of each individual component. We would like to formulate these statements in such a form that *all of the separate confidence statements should hold* simultaneously with a prespecified probability. This is why we are speaking about *simultaneous confidence intervals*.

First, note that if the vector $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ then for any $\boldsymbol{l} \in \mathbb{R}^p : \boldsymbol{l}^\top \boldsymbol{X} \sim N_1(\boldsymbol{l}^\top\boldsymbol{\mu}, \boldsymbol{l}^\top\Sigma\boldsymbol{l})$ and, hence, for any fixed $\boldsymbol{l}$ we can construct an $(1-\alpha)\cdot 100\%$ confidence interval of $\boldsymbol{l}^\top\boldsymbol{\mu}$ in the following simple way:

$$\left(\boldsymbol{l}^\top\bar{\boldsymbol{x}} - t_{1-\alpha/2,n-1}\frac{\sqrt{\boldsymbol{l}^\top\boldsymbol{S}\boldsymbol{l}}}{\sqrt{n}}, \boldsymbol{l}^\top\bar{\boldsymbol{x}} + t_{1-\alpha/2,n-1}\frac{\sqrt{\boldsymbol{l}^\top\boldsymbol{S}\boldsymbol{l}}}{\sqrt{n}}\right) \tag{1.16}$$

By taking $\boldsymbol{l}^\top = [1,0,\dots,0]$ or $\boldsymbol{l}^\top = [0,1,0,\dots,0]$ etc. we obtain from (1.16) the usual confidence interval for each separate component of the mean. Note however that the confidence level for all these statements taken together is not $(1-\alpha)$. To make it $(1-\alpha)$ for all possible choices simultaneously we need to take a larger constant than $t_{1-\alpha/2,n-1}$ in the right hand side of the inequality $|\frac{\sqrt{n}(\boldsymbol{l}^\top\bar{\boldsymbol{x}}-\boldsymbol{l}^\top\bar{\boldsymbol{\mu}})}{\sqrt{\boldsymbol{l}^\top\boldsymbol{S}\boldsymbol{l}}}| \leq t_{1-\alpha/2,n-1}$ (or equivalently $\frac{n(\boldsymbol{l}^\top\bar{\boldsymbol{x}}-\boldsymbol{l}^\top\bar{\boldsymbol{\mu}})^2}{\boldsymbol{l}^\top\boldsymbol{S}\boldsymbol{l}} \leq t^2_{1-\alpha/2,n-1}$).

Simultaneous confidence ellipsoid

Theorem 1.7. *Simultaneously for all* $\boldsymbol{l} \in \mathbb{R}^p$*, the interval*

$$\left(\boldsymbol{l}^\top \bar{\boldsymbol{x}} - \sqrt{\frac{p(n-1)}{n(n-p)} F_{1-\alpha,p,n-p} \boldsymbol{l}^\top S \boldsymbol{l}},\ \boldsymbol{l}^\top \bar{\boldsymbol{x}} + \sqrt{\frac{p(n-1)}{n(n-p)} F_{1-\alpha,p,n-p} \boldsymbol{l}^\top S \boldsymbol{l}}\right)$$

will contain $\boldsymbol{l}^\top \bar{\boldsymbol{\mu}}$ *with a probability at least* $(1-\alpha)$*.*

For Illustration, see example 5.4, p. 226 in Johnson and Wichern.

Bonferroni Method

The simultaneous confidence intervals when applied for the vectors $\boldsymbol{l}^\top = [1, 0, \ldots, 0], \boldsymbol{l}^\top = [0, 1, 0, \ldots, 0]$ etc. are much more reliable at a given confidence level than the one-at-a-time intervals. Note that the former also utilise the covariance structure of all p variables in their construction. However, sometimes we can do better in cases where one is interested in a small number of individual confidence statements.

In this latter case, the simultaneous confidence intervals may give too large a region and the Bonferroni method may prove more efficient instead. The idea of the Bonferroni approach is based on a simple probabilistic inequality. Assume that simultaneous confidence statements about m linear combinations $\boldsymbol{l}_1^\top \boldsymbol{\mu}, \boldsymbol{l}_2^\top \boldsymbol{\mu}, \ldots, \boldsymbol{l}_m^\top \boldsymbol{\mu}$ are required. If $C_i, i = 1, 2, \ldots, m$ denotes the ith confidence statement and $P(C_i \text{ true}) = 1 - \alpha_i$ then

$$P\text{ (all } C_i \text{ true) } = 1 - P\text{ (at least one } C_i \text{ false) } \geq$$

$$1 - \sum_{i=1}^m P\left(C_i \text{ false }\right) = 1 - \sum_{i=1}^m \left(1 - P\left(C_i \text{ true }\right)\right) = 1 - \left(\alpha_1 + \alpha_2 + \cdots + \alpha_m\right)$$

Hence, if we choose $\alpha_i = \frac{\alpha}{m}, i = 1, 2, \ldots, m$ (that is, if calculate each statement at confidence level $(1 - \frac{\alpha}{m}) \cdot 100\%$ instead of $(1-\alpha) \cdot 100\%$) then the probability of any statement being false will not exceed α.

Comparison of two or more mean vectors

Finally, let us note that comparison of the mean vectors of two or more than two different multivariate populations when there are independent observations from each of the populations is an important, practically relevant problem. For the purposes of this section, suppose that we observe two samples, $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_{n\boldsymbol{X}} \in \mathbb{R}^p$ and $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_{n\boldsymbol{Y}} \in \mathbb{R}^p$, with means $\boldsymbol{\mu}_{\boldsymbol{X}} \in \mathbb{R}^p$ and $\boldsymbol{\mu}_{\boldsymbol{Y}} \in \mathbb{R}^p$ respectively and variances $\Sigma_{\boldsymbol{X}} \in \mathcal{M}_{p,p}$ and $\Sigma_{\boldsymbol{Y}} \in \mathcal{M}_{p,p}$, respectively. Typically, we wish to test $H_0 : \boldsymbol{\mu}_{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{Y}} = \boldsymbol{\delta}_0$.

Multivariate ANOVA for comparing more than two populations is discussed in Topic 2 of Week 3.

Reducing to a single population

As with the univariate t-test, under some scenarios the test of a difference between two populations in fact reduces to a one-sample test. For example, if the samples are paired and $n_{\boldsymbol{X}} = n_{\boldsymbol{Y}} = n$, we may proceed analogously to the paired t-test: we take $\boldsymbol{D}_i = \boldsymbol{X}_i - \boldsymbol{Y}_i$ for $i = 1, \ldots, n$ and proceed as if with a 1-sample T^2 test:

$$T^2 = n(\bar{\boldsymbol{D}} - \boldsymbol{\delta}_0)^\top \boldsymbol{S}_{\boldsymbol{D}}^{-1} (\bar{\boldsymbol{D}} - \boldsymbol{\delta}_0) \sim \frac{(n-1)p}{n-p} F_{p,n-p}, \tag{1.17}$$

where $\bar{\boldsymbol{D}} \in \mathbb{R}^p$ and $\boldsymbol{S}_{\boldsymbol{D}} \in \mathcal{M}_{p,p}$ are the sample mean and variance of $\boldsymbol{D}_1, \ldots, \boldsymbol{D}_n$, respectively, assuming $\boldsymbol{D}_i$ are normally distributed. (It is important to note that any diagnostics for this test should be performed on the differences, not on the original values.)

We can also formulate this is in a "multivariate" form: let the *contrast matrix* $C \in \mathcal{M}_{p,p+p}$ be

$$C = \begin{pmatrix} +1 & & & -1 & & \\ & +1 & & & -1 & \\ & & +1 & & & -1 \end{pmatrix}.$$

Then, we can express $\boldsymbol{D}_i = C \begin{pmatrix} \boldsymbol{X}_i \\ \boldsymbol{Y}_i \end{pmatrix}$ and the test as $H_0 : C \begin{pmatrix} \boldsymbol{\mu}_{\boldsymbol{X}} \\ \boldsymbol{\mu}_{\boldsymbol{Y}} \end{pmatrix} = \boldsymbol{\delta}_0$. It is easy to show that the test statistic reduces to (1.17).

C can have more complex forms. For example, in a *repeated measures design*, we may measure the results of a series of p treatment outcomes on each sampling unit. If we then collect each individual i 's measurements into a vector $\boldsymbol{X}_i$, we may test whether all outcomes are the same by forming

$$C = \begin{pmatrix} 1 & -1 & & \\ \vdots & & \ddots & \\ 1 & & & -1 \end{pmatrix} \in \mathcal{M}_{p-1,p}$$

and testing $H_0 : C\boldsymbol{\mu}_{\boldsymbol{X}} = \mathbf{0}_{p-1}$. It is easy to show that $C\boldsymbol{\mu}_{\boldsymbol{X}} = \mathbf{0}_{p-1}$ holds if and only if all elements of $\boldsymbol{\mu}_{\boldsymbol{X}}$ are equal.

The two-sample T^2-test

We now turn to the scenario where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are, in fact, independent samples. As with the univariate test, we must decide whether we are prepared to assume that $\Sigma_{\boldsymbol{X}} = \Sigma_{\boldsymbol{Y}} = \Sigma$ in the population and therefore use the pooled test. If so—and necessarily if the sample sizes are small—we evaluate

$$\boldsymbol{S}_{\text{pooled}} = \frac{(n_{\boldsymbol{X}} - 1)\boldsymbol{S}_{\boldsymbol{X}} + (n_{\boldsymbol{Y}} - 1)\boldsymbol{S}_{\boldsymbol{Y}}}{n_{\boldsymbol{X}} + n_{\boldsymbol{Y}} - 2}.$$

Since $\boldsymbol{S}_{\text{pooled}}$ estimates Σ,

$$\text{Var}(\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}}) = \frac{\Sigma}{n_{\boldsymbol{X}}} + \frac{\Sigma}{n_{\boldsymbol{Y}}} \approx \frac{\boldsymbol{S}_{\text{pooled}}}{n_{\boldsymbol{X}}} + \frac{\boldsymbol{S}_{\text{pooled}}}{n_{\boldsymbol{Y}}} = \boldsymbol{S}_{\text{pooled}}\left(\frac{1}{n_{\boldsymbol{X}}} + \frac{1}{n_{\boldsymbol{Y}}}\right).$$

And, since $\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}} \sim N_p(\boldsymbol{\mu}_{\boldsymbol{X}} - \boldsymbol{\mu}_{\boldsymbol{Y}}, \Sigma(n_{\boldsymbol{X}}^{-1} + n_{\boldsymbol{Y}}^{-1}))$, we write

$$\begin{aligned} T^2 &= (\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}} - \boldsymbol{\delta}_0)^\top \left\{\boldsymbol{S}_{\text{pooled}}\left(\tfrac{1}{n_{\boldsymbol{X}}} + \tfrac{1}{n_{\boldsymbol{Y}}}\right)\right\}^{-1} (\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}} - \boldsymbol{\delta}_0) \\ &\sim \frac{(n_{\boldsymbol{X}} + n_{\boldsymbol{Y}} - 2)p}{n_{\boldsymbol{X}} + n_{\boldsymbol{Y}} - p - 1} F_{p, n_{\boldsymbol{X}} + n_{\boldsymbol{Y}} - p - 1}. \end{aligned} \tag{1.18}$$

We would thus reject H_0 if T^2 falls above the F critical value in (1.18), construct a confidence region based on

$$\left\{\boldsymbol{\delta} | (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}} - \boldsymbol{\delta})^\top \left\{S_{\text{pooled}}\left(\frac{1}{n_{\boldsymbol{X}}} + \frac{1}{n_{\boldsymbol{Y}}}\right)\right\}^{-1} (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}} - \boldsymbol{\delta}) \le \frac{(n_{\boldsymbol{X}} + n_{\boldsymbol{Y}} - 2)p}{n_{\boldsymbol{X}} + n_{\boldsymbol{Y}} - p - 1} F_{1-\alpha, p, n_{\boldsymbol{X}} + n_{\boldsymbol{Y}} -}$$

and simultaneous contrast confidence intervals

$$\boldsymbol{l}^\top (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}) \pm \sqrt{\frac{(n_{\boldsymbol{X}} + n_{\boldsymbol{Y}} - 2)p}{n_{\boldsymbol{X}} + n_{\boldsymbol{Y}} - p - 1} F_{1-\alpha, p, n_{\boldsymbol{X}} + n_{\boldsymbol{Y}} - p - 1} \boldsymbol{l}^\top S_{\text{pooled}} \left(\frac{1}{n_{\boldsymbol{X}}} + \frac{1}{n_{\boldsymbol{Y}}}\right) \boldsymbol{l}}.$$

If we are not prepared to make the pooling assumption, our test statistic is instead

$$T^2 = (\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}} - \boldsymbol{\delta}_0)^\top \left(\frac{\boldsymbol{S}_{\boldsymbol{X}}}{n_{\boldsymbol{X}}} + \frac{\boldsymbol{S}_{\boldsymbol{Y}}}{n_{\boldsymbol{Y}}}\right)^{-1} (\bar{\boldsymbol{X}} - \bar{\boldsymbol{Y}} - \boldsymbol{\delta}_0).$$

Even for modest sample sizes, under multivariate normality, the distribution of this T^2 is reasonably well approximated by $\frac{\nu p}{\nu - p + 1} F_{p, \nu - p + 1}$, where

$$\nu = \frac{p+p^2}{\sum_{i=1}^{2} \frac{1}{n_i} \left(\mathrm{tr}\left[\left\{ \frac{1}{n_i} \boldsymbol{S}_i \left(\frac{1}{n_1} \boldsymbol{S}_1 + \frac{1}{n_2} \boldsymbol{S}_2 \right)^{-1} \right\}^2 \right] + \left[\mathrm{tr} \left\{ \frac{1}{n_i} \boldsymbol{S}_i \left(\frac{1}{n_1} \boldsymbol{S}_1 + \frac{1}{n_2} \boldsymbol{S}_2 \right)^{-1} \right\} \right]^2 \right)},$$

where $\boldsymbol{S}_1 = \boldsymbol{S}_{\boldsymbol{X}}$ and $\boldsymbol{S}_2 = \boldsymbol{S}_{\boldsymbol{Y}}$, and analogously for n_i.

The confidence regions are then produced by

$$\left\{ \boldsymbol{\delta} | (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}} - \boldsymbol{\delta})^\top \left(\frac{\boldsymbol{S}_{\boldsymbol{X}}}{n_{\boldsymbol{X}}} + \frac{\boldsymbol{S}_{\boldsymbol{Y}}}{n_{\boldsymbol{Y}}} \right)^{-1} (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}} - \boldsymbol{\delta}) \leq \frac{\nu p}{\nu - p + 1} F_{p,\nu-p+1} \right\}$$

and simultaneous contrast confidence intervals

$$\boldsymbol{l}^\top (\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}) \pm \sqrt{\frac{\nu p}{\nu - p + 1} F_{p,\nu-p+1} \boldsymbol{l}^\top \left(\frac{\boldsymbol{S}_{\boldsymbol{X}}}{n_{\boldsymbol{X}}} + \frac{\boldsymbol{S}_{\boldsymbol{Y}}}{n_{\boldsymbol{Y}}} \right) \boldsymbol{l}}.$$

Check your understanding

Complete the below exercises to check your understanding of concepts presented so far.

1. Suppose $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are independent $N_p(\boldsymbol{\mu}, \Sigma)$ random vectors with sample mean vector $\bar{\boldsymbol{X}}, S$ and sample covariance matrix S. We wish to test the hypothesis

$$H_0 : \mu_2 - \mu_1 = \mu_3 - \mu_2 = \cdots = \mu_p - \mu_{p-1} = 1$$

where $\mu_1, \mu_2, \ldots, \mu_p$ are the elements of μ.

a) Determine a $(p-1) \times p$ matrix C so that H_0 may be written equivalently as $H_0 : C\mu = \mathbf{1}$ where $\mathbf{1}$ is a $(p-1) \times 1$ vector of ones.

b) Make an appropriate transformation of the vectors $\boldsymbol{X}_i, i = 1, 2, \ldots, n$ and hence find the rejection region of a size α *test of* H_0 in terms of $\bar{\boldsymbol{X}}, S$ and C.

2. A sample of 50 vector observations , each containing three components, is drawn from a normal distribution having covariance matrix

$$\Sigma = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

The components of the sample mean are 0.8, 1.1 and 0.6. Can you reject the null hypothesis of zero distribution mean against a general alternative?

3. Evaluate Hotelling's statistic T^2 for testing hypothesis $H_0 : \mu = \begin{pmatrix} 7 \\ 11 \end{pmatrix}$ using the data matrix $\boldsymbol{X} = \begin{pmatrix} 2 & 8 & 6 & 8 \\ 12 & 9 & 9 & 10 \end{pmatrix}$. Test the hypothesis H_0 at level $\alpha = 0.05$. What conclusion is reached?

Demonstration: Hotelling's

Please start by watching the following demonstration by Dr Krivitsky.

Transcript

This demonstration can be completed using the provided RStudio or your own RStudio.

To complete the demonstration select the 'Hotelling_CI_Examples.demo.Rmd' in the 'Files' section of RStudio. Follow the demonstration contained within the RMD file.

If you choose to complete the example in your own RStudio, upload the following file:

Hotelling_CI_Examples.demo.Rmd

The output of the RMD file is also displayed below:

Challenge: Hotelling's

If you choose to complete this task in your own RStudio, upload the following file:

Hotelling_CI_Examples.challenge.Rmd

Click on the 'Hotelling_CI_Examples.challenge.Rmd' in the 'Files' section to begin. Enter your response to the tasks in the 'Enter your code here' section.

This activity and the solution will be discussed at the Collaborate session this week. In the meantime, share and discuss your results in the 'Tutorials' discussion forum.

The output of the RMD file is also displayed below: