IV Lecture

Some important test statistics and their Distributions

4.1. Tests and confidence regions for the multivariate normal mean

4.1. Hotelling's T^2

Suppose again that, like in III Lecture, we have observed n independent realizations of p-dimensional random vectors from $N_p(\mu, \mathbf{\Sigma})$. Suppose for simplicity that $\mathbf{\Sigma}$ is non-singular. The data matrix has the form

$$\mathbf{x} = \begin{pmatrix} x_{11} & x_{12} & .. & x_{1j} & .. & x_{1n} \\ x_{21} & x_{22} & .. & x_{2j} & .. & x_{2n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ x_{i1} & x_{i2} & .. & x_{ij} & .. & x_{in} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ x_{p1} & x_{p2} & .. & x_{pj} & .. & x_{pn} \end{pmatrix} = [\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_n}]$$

Based on our knowledge from III Lecture we can claim that $\bar{\mathbf{X}} \sim N_p(\mu, \frac{\mathbf{1}}{\mathbf{n}}\mathbf{\Sigma})$ and $n\hat{\mathbf{\Sigma}} \sim \mathbf{W_p}(\mathbf{\Sigma}, \mathbf{n-1})$.

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

$$T = \sqrt{n}\mathbf{c}'(\bar{\mathbf{X}} - \mu)/\sqrt{\mathbf{c}'\frac{\mathbf{n}}{\mathbf{n-1}}\hat{\mathbf{\Sigma}}\mathbf{c}} \sim t_{n-1}$$

i.e. follows t-distribution with (n-1) degrees of freedom.

This result has useful applications in testing for contrasts.

Indeed, if we would like to test $H_0 : c'\mu = \sum_{i=1}^p c_i\mu_i = 0$, we note that under H_0, T becomes simply

$$T = \sqrt{n}\mathbf{c}'\bar{\mathbf{X}}/\sqrt{\mathbf{c}'\mathbf{S}\mathbf{c}},$$

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

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 $\mathbf{\Sigma}$ (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 $U = \sqrt{n}\frac{\bar{\mathbf{X}} - \mu_0}{\sigma}$ and reject H_0 if $|U|$ exceeds the upper $\frac{\alpha}{2}.100\%$ point of the standard normal distribution. Checking if $|U|$ is large enough is equivalent to checking if $U^2 = n(\bar{\mathbf{X}} - \mu_\mathbf{0})(\sigma^\mathbf{2})^{-\mathbf{1}}(\bar{\mathbf{X}} - \mu_\mathbf{0})$ is large enough.

We can now easily generalize the above test statistic in a natural way for the multivariate ($p > 1$) case: calculate $U^2 = n(\bar{\mathbf{X}} - \mu_{\mathbf{0}})'\mathbf{\Sigma}^{-\mathbf{1}}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})$ and reject the null hypothesis when U^2 is large enough. Similarly to the proof of *Property 5* of the multivariate normal distribution (II Lecture) and by using Theorem 3.3 of III Lecture you can convince yourself (do it (!)) 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 $\mathbf{\Sigma}$. 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

$$T = \sqrt{n}\frac{\bar{\mathbf{X}} - \mu_{\mathbf{0}}}{\mathbf{S}}, \mathbf{S^2} = \frac{\mathbf{1}}{\mathbf{n-1}}\sum_{\mathbf{i=1}}^{\mathbf{n}}(\mathbf{X_i} - \bar{\mathbf{X}})^{\mathbf{2}}$$

and reject H_0 if $|T|$ exceeds the upper $\frac{\alpha}{2}.100\%$ 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 $T^2 = n(\bar{\mathbf{X}} - \mu_{\mathbf{0}})(\mathbf{s^2})^{-\mathbf{1}}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})$ 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 generalize the above test statistic in a natural way for the multivariate ($p > 1$) case:

Definition 4.1.1 *Hotelling's* T^2. The statistic

$$T^2 = n(\bar{\mathbf{X}} - \mu_{\mathbf{0}})'\mathbf{S}^{-\mathbf{1}}(\bar{\mathbf{X}} - \mu_{\mathbf{0}}) \tag{4.1}$$

where $\bar{\mathbf{X}} = \frac{\mathbf{1}}{\mathbf{n}}\sum_{\mathbf{j=1}}^{\mathbf{n}}\mathbf{X_i}, \mathbf{S} = \frac{\mathbf{1}}{\mathbf{n-1}}\sum_{\mathbf{i=1}}^{\mathbf{n}}(\mathbf{X_i} - \bar{\mathbf{X}})(\mathbf{X_i} - \bar{\mathbf{X}})', \mu_{\mathbf{0}} \in \mathbf{R^p}, \mathbf{X_i} \in \mathbf{R^p}, i = 1, \dots, n$ is named after Harold Hotelling.

Obviously, the test procedure based on Hotelling's statistic will reject the zero hypothesis $H_0 : \mu = \mu_0$ if the value of T^2 is significantly large. 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 4.1.2. Under the null hypothesis $H_0 : \mu = \mu_0$, 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.

Proof. Indeed, we can write the T^2 statistic in the form:

$$T^2 = \frac{\mathbf{n}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})'\mathbf{S}^{-\mathbf{1}}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})}{\mathbf{n}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})'\mathbf{\Sigma}^{-\mathbf{1}}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})}.\mathbf{n}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})'\mathbf{\Sigma}^{-\mathbf{1}}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})$$

Denote by $\mathbf{c} = \sqrt{\mathbf{n}}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})$. Conditionally on $\mathbf{c}$ we have:

$$\frac{\mathbf{n}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})'\mathbf{S}^{-\mathbf{1}}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})}{\mathbf{n}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})'\mathbf{\Sigma}^{-\mathbf{1}}(\bar{\mathbf{X}} - \mu_{\mathbf{0}})} = \frac{\mathbf{c}'\mathbf{S}^{-\mathbf{1}}\mathbf{c}}{\mathbf{c}'\mathbf{\Sigma}^{-\mathbf{1}}\mathbf{c}}$$

has a distribution that only depends on the data through $\mathbf{S}^{-\mathbf{1}}$. Noting that $n\hat{\mathbf{\Sigma}} = (\mathbf{n-1})\mathbf{S}$ and having in mind the third property of Wishart distributions from III Lecture, we can claim that this distribution is the same as of $(n-1)/\chi^2_{n-p}$. Note also that the distribution does **not** depend on the particular $\mathbf{c}$. The second factor $n(\bar{\mathbf{X}} - \mu_{\mathbf{0}})\mathbf{\Sigma}^{-\mathbf{1}}(\bar{\mathbf{X}} - \mu_{\mathbf{0}}) \sim \chi^{\mathbf{2}}_{\mathbf{p}}$ and its distribution depends on the data through $\bar{\mathbf{X}}$ only. Because of the independence of the mean and covariance estimators, we have that the distribution of T^2 is the same as the distribution of $\frac{\chi^2_p(n-1)}{\chi^2_{n-p}}$ where the two chi-squares are **independent**. But this means that $\frac{T^2(n-p)}{p(n-1)} \sim F_{p,n-p}$ and hence $T^2 \sim \frac{p(n-1)}{n-p} F_{p,n-p}$.

4.1.3. Remark It is possible to extend the definition of the Wishart distribution in 3.2.2 by allowing the random vectors $\mathbf{Y_i}, \mathbf{i} = \mathbf{1}, \ldots, \mathbf{n}$ there to be independent with $N_p(\mu_\mathbf{i}, \mathbf{\Sigma})$ (instead of just having all $\mu_\mathbf{i} = \mathbf{0}$). One arrives at the *noncentral* Wishart distribution with parameters $\mathbf{\Sigma}, \mathbf{p}, \mathbf{n}-\mathbf{1}, \mathbf{\Gamma}$ that way (denoted also as $\mathbf{W_p}(\mathbf{\Sigma}, \mathbf{n}-\mathbf{1}, \mathbf{\Gamma})$. Here $\mathbf{\Gamma} = \mathbf{M}\mathbf{M}' \in \mathscr{M}_{\mathbf{p},\mathbf{p}}, \mathbf{M} = [\mu_\mathbf{1}, \mu_\mathbf{2}, \ldots, \mu_\mathbf{n}]$ is called a *noncentrality parameter*. When all columns of $\mathbf{M} \in \mathscr{M}_{\mathbf{p},\mathbf{n}}$ are zero, this is the usual (*central*) Wishart distribution. Theorem 4.1.2 can be extended to derive the distribution of the T^2 statistic under alternatives, i.e. the distribution of $T^2 = n(\mathbf{\bar{X}} - \mu)'\mathbf{S}^{-\mathbf{1}}(\mathbf{\bar{X}} - \mu)$ for $\mu \neq \mu_0$. This distribution turns out to be related to *noncentral F-distribution*. It is helpful in studying power of the test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. We shall spare the details here.

4.1.4. 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 $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. This safeguards the asymptotic optimality of the test suggested in 4.1.1-4.1.2. To see this , first recall the likelihood function (3.2). Its unconstrained maximization gives as a maximum value:

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

On the other hand, under H_0:

$$\max_{\Sigma} \mathbf{L}(\mathbf{x}; \mu_\mathbf{0}, \mathbf{\Sigma}) = \max_{\Sigma} \frac{1}{(2\pi)^{\frac{np}{2}} |\mathbf{\Sigma}|^{\frac{\mathbf{n}}{\mathbf{2}}}} e^{(-\frac{1}{2}\sum_{i=1}^n (x_i-\mu_0)'\Sigma^{-1}(x_i-\mu_0))}$$

Since $logL(\mathbf{x}; \mu_\mathbf{0}, \mathbf{\Sigma}) = -\frac{\mathbf{np}}{\mathbf{2}} \log(\mathbf{2}\pi) - \frac{\mathbf{n}}{\mathbf{2}} \log|\mathbf{\Sigma}| - \frac{\mathbf{1}}{\mathbf{2}} \mathbf{tr}[\mathbf{\Sigma}^{-\mathbf{1}}(\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}}(\mathbf{x_i} - \mu_\mathbf{0})(\mathbf{x_i} - \mu_\mathbf{0})')]$, on applying Anderson's lemma (see Theorem 3.1 in III lecture) we find that maximum of $logL(\mathbf{x}; \mu_\mathbf{0}, \mathbf{\Sigma})$ (whence also of $L(\mathbf{x}; \mu_\mathbf{0}, \mathbf{\Sigma})$) is obtained when $\hat{\mathbf{\Sigma}}_\mathbf{0} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} (\mathbf{x_i} - \mu_\mathbf{0})(\mathbf{x_i} - \mu_\mathbf{0})'$ and the maximal value is

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

Hence the likelihood ratio is:

$$\mathbf{\Lambda} = \frac{\max_{\mathbf{\Sigma}} \mathbf{L}(\mathbf{x}; \mu_\mathbf{0}, \mathbf{\Sigma})}{\max_{\mu, \mathbf{\Sigma}} \mathbf{L}(\mathbf{x}; \mu, \mathbf{\Sigma})} = \left(\frac{|\hat{\mathbf{\Sigma}}|}{|\hat{\mathbf{\Sigma}}_\mathbf{0}|}\right)^{\frac{\mathbf{n}}{\mathbf{2}}} \tag{4.2}$$

The equivalent statistic $\mathbf{\Lambda}^{\frac{\mathbf{2}}{\mathbf{n}}} = \frac{|\hat{\mathbf{\Sigma}}|}{|\hat{\mathbf{\Sigma}}_\mathbf{0}|}$ is called **Wilks' lambda**. Small values of Wilks' lambda lead to rejecting $H_0 : \mu = \mu_0$.

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

Theorem 4.1.5. The likelihood ratio test is equivalent to the test based on T^2 since $\mathbf{\Lambda}^{\frac{\mathbf{2}}{\mathbf{n}}} = (\mathbf{1} + \frac{\mathbf{T^2}}{\mathbf{n}-\mathbf{1}})^{-\mathbf{1}}$ holds.

Proof. Consider the matrix $A \in \mathscr{M}_{p+1,p+1}$:

$$A = \begin{pmatrix} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})' & \sqrt{n}(\bar{x} - \mu_0) \\ \sqrt{n}(\bar{x} - \mu_0)' & -1 \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

It is easy to check that $|A| = |A_{22}||A_{11} - A_{12}A_{22}^{-1}A_{21}| = |A_{11}||A_{22} - A_{21}A_{11}^{-1}A_{12}|$ holds from which we get:

$$(-1)|\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})' + n(\bar{x} - \mu_0)(\bar{x} - \mu_0)'| =$$

$$|\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'|| - 1 - n(\bar{x} - \mu_0)'(\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})')^{-1}(\bar{x} - \mu_0)|$$

Hence $(-1)|\sum_{i=1}^{n}(x_i - \mu_0)(x_i - \mu_0)'| = |\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})'|(-1)(1 + \frac{T^2}{n-1})$.Thus $|\hat{\Sigma}_0| = |\hat{\Sigma}|(1 + \frac{T^2}{n-1})$, i.e.

$$\Lambda^{\frac{2}{n}} = (1 + \frac{T^2}{n-1})^{-1} \tag{4.3}$$

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 4.1.2.

4.1.6. Note about numerical calculation. Relation (4.3) can be used to calculate T^2 from $\Lambda^{\frac{2}{n}} = \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|}$ thus **avoiding the need to invert the matrix S when calculating T^2!**

4.1.7. Asymptotic distribution of T^2. Since $\mathbf{S^{-1}}$ is a consistent estimator of $\mathbf{\Sigma^{-1}}$, the limiting distribution of T^2 will coincide with the one of $n(\bar{\mathbf{x}} - \mu)'\mathbf{\Sigma^{-1}}(\bar{\mathbf{x}} - \mu)$ 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).

4.2. Confidence regions for the mean vector and for its components

4.2.1. Confidence region for the mean vector

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

$$\{\mu|\mathbf{n}(\bar{\mathbf{x}} - \mu)'\mathbf{S^{-1}}(\bar{\mathbf{x}} - \mu) \le F_{p,n-p}(\alpha)\frac{p}{n-p}(n-1)\}$$

where $F_{p,n-p}(\alpha)$ is the upper $\alpha.100\%$ percentage point of the F distribution with $(p, n-p)$ df. This confidence region has the form of an *ellipsoid* in R^p centered at $\bar{\mathbf{x}}$. The axes of this *confidence ellipsoid* are directed along the eigenvectors $\mathbf{e_i}$ of the matrix $\mathbf{S} = \frac{\mathbf{1}}{\mathbf{n-1}}\sum_{\mathbf{i=1}}^{\mathbf{n}}(\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})'$. The lengths of the axes are determined by the expression $\sqrt{\lambda_i}\sqrt{\frac{p(n-1)F_{p,n-p}(\alpha)}{n(n-p)}}, \lambda_i, i = 1, \ldots, p$ being the corresponding eigenvectors, i.e.

$$\mathbf{Se_i} = \lambda_\mathbf{i}\mathbf{e_i}, \mathbf{i} = \mathbf{1}, \ldots, \mathbf{p}$$

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

4.2.2. Simultaneous confidence statements

For a given confidence level $(1 - \alpha)$ the confidence ellipsoids in 4.2.1. correctly reflect the joint (multivariate) knowledge about plausible values of $\mu \in R^p$ 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 $X \sim N_p(\mu, \Sigma)$ then for any $l \in R^p : l'X \sim N_1(l'\mu, l'\Sigma l)$ and, hence, for any fixed l we can construct an $(1-\alpha).100\%$ confidence interval of $l'\mu$ in the following simple way:

$$l'\bar{x} - t_{n-1}(\alpha/2)\frac{\sqrt{l'Sl}}{\sqrt{n}}, l'\bar{x} + t_{n-1}(\alpha/2)\frac{\sqrt{l'Sl}}{\sqrt{n}} \tag{4.4}$$

By taking $l' = [1, 0, \ldots, 0]$ or $l' = [0, 1, 0, \ldots, 0]$ etc. we obtain from (4.4) 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 $t_{n-1}(\alpha/2)$ in the right hand side of the inequality $|\frac{\sqrt{n}(l'\bar{x}-l'\mu)}{\sqrt{l'Sl}}| \leq t_{n-1}(\alpha/2)$ (or equivalently $\frac{n(l'\bar{x}-l'\mu)^2}{l'Sl} \leq t^2_{n-1}(\alpha/2)$).

Theorem 4.2.3. Simultaneously for all $l \in R^p$, the interval

$$(l'\bar{x} - \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)l'Sl}, l'\bar{x} + \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)l'Sl})$$

will contain $l'\mu$ with a probability at least $(1-\alpha)$.

Proof. Note that according to Cauchy-Bunyakovski-Schwartz Inequality:

$$[l'(\bar{x}-\mu)]^2 = [(S^{1/2}l)'S^{-1/2}(\bar{x}-\mu)]^2 \leq \| S^{1/2}l \|^2 . \| S^{-1/2}(\bar{x}-\mu) \|^2 = l'Sl.(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu)$$

Therefore

$$\max_l \frac{n(l'(\bar{x}-\mu))^2}{l'Sl} \leq n(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu) = T^2 \tag{4.5}$$

Inequality (4.5) helps us to claim that whenever a constant c has been such that $T^2 \leq c^2$ then also $\frac{n(l'\bar{x}-l'\mu)^2}{l'Sl} \leq c^2$ holds for any $l \in R^p, l \neq 0$. Equivalently,

$$l'\bar{x} - c\sqrt{\frac{l'Sl}{n}} \leq l'\mu \leq l'\bar{x} + c\sqrt{\frac{l'Sl}{n}} \tag{4.6}$$

for every l. Now it remains to choose $c^2 = p(n-1)F_{p,n-p}(\alpha)/(n-p)$ to make sure that $1-\alpha = P(T^2 \leq c^2)$ holds and this will automatically ensure that (4.6) will contain $l'\mu$ with probability $1-\alpha$.

Illustration: Example 5.4, p. 226 in Johnson and Wichern.

The simultaneous confidence intervals when applied for the vectors $l' = [1, 0, \ldots, 0], l' = [0, 1, 0, \ldots, 0]$ etc. are much more reliable at a given confidence level than the one-at-a-time intervals. Note that the former also utilize 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 **Bonferonni method** may prove more efficient instead. The idea of the

Bonferonni approach is based on a simple probabilistic inequality. Assume that simultenous confidence statements about m linear combinations $l_1'\mu, l_2'\mu, \dots, l_m'\mu$ are required. If $C_i, i = 1, 2, \dots, m$ denotes the ith confidence statement and $P(C_i \text{ true }) = 1 - \alpha_i$ then

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

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

Hence, if we choose $\alpha_i = \frac{\alpha}{m}, i = 1, 2, \dots, m$ (that is, if we test *each separate* hypothesis at a level $\frac{\alpha}{m}$ instead of α) then the *overall* level will not exceed α.

4.2.4. Remark. 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 important practically relevant problem. It is discussed in the masters course on Longitudinal data analysis.