Chapter 7
Hypothesis Testing

In the preceding chapter, the theoretical basis of estimation theory was presented. Now we turn our interest towards testing issues: we want to test the hypothesis H_0 that the unknown parameter θ belongs to some subspace of $\mathbb{R}^q$. This subspace is called the *null set* and will be denoted by $\Omega_0 \subset \mathbb{R}^q$.

In many cases, this null set corresponds to restrictions which are imposed on the parameter space: H_0 corresponds to a "reduced model". As we have already seen in Chapter 3, the solution to a testing problem is in terms of a *rejection region R* which is a set of values in the sample space which leads to the decision of rejecting the null hypothesis H_0 in favour of an alternative H_1, which is called the "full model".

In general, we want to construct a rejection region R which controls the size of the type I error, i.e. the probability of rejecting the null hypothesis when it is true. More formally, a solution to a testing problem is of predetermined size α if:

$$\text{P(Rejecting } H_0 \mid H_0 \text{ is true)} = \alpha.$$

In fact, since H_0 is often a composite hypothesis, it is achieved by finding R such that

$$\sup_{\theta \in \Omega_0} \text{P}(\mathcal{X} \in R \mid \theta) = \alpha.$$

In this chapter we will introduce a tool which allows us to build a rejection region in general situations; it is based on the likelihood ratio principle. This is a very useful technique because it allows us to derive a rejection region with an asymptotically appropriate size α. The technique will be illustrated through various testing problems and examples. We concentrate on multinormal populations and linear models where the size of the test will often be exact even for finite sample sizes n.

Section 7.1 gives the basic ideas and Section 7.2 presents the general problem of testing linear restrictions. This allows us to propose solutions to frequent types of analyses (including comparisons of several means, repeated measurements and profile analysis). Each case can be viewed as a simple specific case of testing linear restrictions. Special attention is devoted to confidence intervals and confidence regions for means and for linear restrictions on means in a multinormal setup.

7.1 Likelihood Ratio Test

Suppose that the distribution of $\{x_i\}_{i=1}^n$, $x_i \in \mathbb{R}^p$, depends on a parameter vector θ. We will consider two hypotheses:

$$H_0 : \theta \in \Omega_0$$
$$H_1 : \theta \in \Omega_1.$$

The hypothesis H_0 corresponds to the "reduced model" and H_1 to the "full model". This notation was already used in Chapter 3.

Example 7.1 Consider a multinormal $N_p(\theta, \mathcal{I})$. To test if θ equals a certain fixed value θ_0 we construct the test problem:

$$H_0 : \theta = \theta_0$$
$$H_1 : \text{no constraints on } \theta$$

or, equivalently, $\Omega_0 = \{\theta_0\}$, $\Omega_1 = \mathbb{R}^p$.

Define $L_j^* = \max_{\theta \in \Omega_j} L(\mathcal{X}; \theta)$, the maxima of the likelihood for each of the hypotheses. Consider the *likelihood ratio* (LR)

$$\lambda(\mathcal{X}) = \frac{L_0^*}{L_1^*}. \tag{7.1}$$

One tends to favour H_0 if the LR is high and H_1 if the LR is low. The *likelihood ratio test* (LRT) tells us when exactly to favour H_0 over H_1. A likelihood ratio test of size α for testing H_0 against H_1 has the rejection region

$$R = \{\mathcal{X} : \lambda(\mathcal{X}) < c\}$$

where c is determined so that $\sup_{\theta \in \Omega_0} P_\theta(\mathcal{X} \in R) = \alpha$. The difficulty here is to express c as a function of α, because $\lambda(\mathcal{X})$ might be a complicated function of $\mathcal{X}$.

Instead of λ we may equivalently use the log-likelihood

$$-2\log\lambda = 2(\ell_1^* - \ell_0^*).$$

In this case the rejection region will be $R = \{\mathcal{X} : -2\log\lambda(\mathcal{X}) > k\}$. What is the distribution of λ or of $-2\log\lambda$ from which we need to compute c or k?

Theorem 7.1 (Wilks Theorem) *If $\Omega_1 \subset \mathbb{R}^q$ is a q-dimensional space and if $\Omega_0 \subset \Omega_1$ is an r-dimensional subspace, then under regularity conditions*

$$\forall \, \theta \in \Omega_0 : \quad -2\log\lambda \xrightarrow{\mathcal{L}} \chi^2_{q-r} \quad \text{as } n \to \infty.$$

An asymptotic rejection region can now be given by simply computing the $1 - \alpha$ quantile $k = \chi^2_{1-\alpha;q-r}$. The LRT rejection region is therefore

$$R = \{\mathcal{X} : -2\log\lambda(\mathcal{X}) > \chi^2_{1-\alpha;q-r}\}.$$

Theorem 7.1 is thus very helpful: it gives a general way of building rejection regions into many problems. Unfortunately, it is only an asymptotic result, meaning that the size of the test is only approximately equal to α, although the approximation becomes better when the sample size n increases. The question is "how large should n be?". There is no definite rule: we encounter here the same problem that was already discussed with respect to the Central Limit Theorem in Chapter 4.

Fortunately, in many standard circumstances, we can derive exact tests even for finite samples because the test statistic $-2 \log \lambda(\mathcal{X})$ or a simple transformation of it turns out to have a simple form. This is the case in most of the following standard testing problems. All of them can be viewed as an illustration of the likelihood ratio principle.

Test Problem 1 is an *amuse-bouche*: in testing the mean of a multinormal population with a known covariance matrix the likelihood ratio statistic has a very simple quadratic form with a known distribution under H_0.

Test Problem 1 Suppose that $X_1, \ldots, X_n$ is an i.i.d. random sample from a $N_p(\mu, \Sigma)$ population.

$$H_0 : \mu = \mu_0, \ \Sigma \text{ known versus } H_1 : \text{ no constraints.}$$

In this case H_0 is a simple hypothesis, i.e., $\Omega_0 = \{\mu_0\}$ and therefore the dimension r of Ω_0 equals 0. Since we have imposed no constraints in H_1, the space Ω_1 is the whole $\mathbb{R}^p$ which leads to $q = p$. From (6.6) we know that

$$\ell_0^* = \ell(\mu_0, \Sigma) = -\frac{n}{2} \log |2\pi \Sigma| - \frac{1}{2} n \operatorname{tr}(\Sigma^{-1} \mathcal{S}) - \frac{1}{2} n (\overline{x} - \mu_0)^\top \Sigma^{-1} (\overline{x} - \mu_0).$$

Under H_1 the maximum of $\ell(\mu, \Sigma)$ is

$$\ell_1^* = \ell(\overline{x}, \Sigma) = -\frac{n}{2} \log |2\pi \Sigma| - \frac{1}{2} n \operatorname{tr}(\Sigma^{-1} \mathcal{S}).$$

Therefore,

$$-2 \log \lambda = 2(\ell_1^* - \ell_0^*) = n (\overline{x} - \mu_0)^\top \Sigma^{-1} (\overline{x} - \mu_0) \tag{7.2}$$

which, by Theorem 4.7, has a χ_p^2-distribution under H_0.

Example 7.2 Consider the bank data again. Let us test whether the population mean of the forged bank notes is equal to

$$\mu_0 = (214.9, 129.9, 129.7, 8.3, 10.1, 141.5)^\top.$$

(This is in fact the sample mean of the genuine bank notes.) The sample mean of the forged bank notes is

$$\overline{x} = (214.8, 130.3, 130.2, 10.5, 11.1, 139.4)^\top.$$

Suppose for the moment that the estimated covariance matrix $\mathcal{S}_f$ given in (3.5) is the true covariance matrix Σ. We construct the likelihood ratio test and obtain

$$-2\log \lambda = 2(\ell_1^* - \ell_0^*) = n(\overline{x} - \mu_0)^\top \Sigma^{-1}(\overline{x} - \mu_0)$$
$$= 7362.32,$$

the quantile $k = \chi^2_{0.95;6}$ equals 12.592. The rejection consists of all values in the sample space which lead to values of the likelihood ratio test statistic larger than 12.592. Under H_0 the value of $-2\log \lambda$ is therefore highly significant. Hence, the true mean of the forged bank notes is significantly different from μ_0!

Test Problem 2 is the same as the preceding one but in a more realistic situation where the covariance matrix is unknown; here the Hotelling's T^2-distribution will be useful to determine an exact test and a confidence region for the unknown μ.

Test Problem 2 Suppose that $X_1, \ldots, X_n$ is an i.i.d. random sample from a $N_p(\mu, \Sigma)$ population.

$$H_0 : \mu = \mu_0, \ \Sigma \text{ unknown versus } H_1 : \text{ no constraints.}$$

Under H_0 it can be shown that

$$\mathcal{S}_0 = \frac{1}{n}\left[x - 1_n\mu_0^\top - 1_n\overline{x}^\top + 1_n\overline{x}^\top\right]^\top \left[x - 1_n\mu_0^\top - 1_n\overline{x}^\top + 1_n\overline{x}^\top\right]$$
$$= \mathcal{S} + (\overline{x} - \mu_0)(\overline{x} - \mu_0)^\top$$
$$\ell_0^* = \ell(\mu_0, \mathcal{S} + dd^\top), \quad d = (\overline{x} - \mu_0) \tag{7.3}$$

and under H_1 we have

$$\ell_1^* = \ell(\overline{x}, \mathcal{S}).$$

This leads after some calculation to

$$-2\log \lambda = 2(\ell_1^* - \ell_0^*)$$
$$= -n\log|\mathcal{S}| - n\operatorname{tr}(\mathcal{S}^{-1}\mathcal{S}) - n(\overline{x} - \overline{x})^\top \mathcal{S}^{-1}(\overline{x} - \overline{x}) + n\log|\mathcal{S} + dd^\top|$$
$$\quad + n\operatorname{tr}\left[(\mathcal{S} + dd^\top)^{-1}\mathcal{S}\right] + n(\overline{x} - \mu_0)^\top (\mathcal{S} + dd^\top)^{-1}(\overline{x} - \mu_0)$$
$$= n\log\left|\frac{\mathcal{S} + dd^\top}{\mathcal{S}}\right| + n\operatorname{tr}\left[(\mathcal{S} + dd^\top)^{-1}\mathcal{S}\right] + nd^\top(\mathcal{S} + dd^\top)^{-1}d - np$$
$$= n\log\left|\frac{\mathcal{S} + dd^\top}{\mathcal{S}}\right| + n\operatorname{tr}\left[(\mathcal{S} + dd^\top)^{-1}(dd^\top + \mathcal{S})\right] - np$$
$$= n\log\left|\frac{\mathcal{S} + dd^\top}{\mathcal{S}}\right|$$
$$= n\log|1 + \mathcal{S}^{-1/2}dd^\top \mathcal{S}^{-1/2}|.$$

By using the result for the determinant of a partitioned matrix, it equals to

$$n \log \begin{vmatrix} 1 & -d^\top S^{-1/2} \\ S^{-1/2}d & I \end{vmatrix}$$

$$= n \log \begin{vmatrix} 1 & -d^\top S^{-1/2}{}_1 & -d^\top S^{-1/2}{}_2 & \cdots & -d^\top S^{-1/2}{}_p \\ S^{-1/2}d_1 & 1 & 0 & \cdots & 0 \\ S^{-1/2}d_2 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \\ S^{-1/2}d_p & 0 & 0 & \cdots & 1 \end{vmatrix}$$

$$= n \log 1 + n \log \sum_{i=1}^{p} -d^\top S^{-1/2}{}_i (-1)^{1+(i+1)} \begin{vmatrix} S^{-1/2}d_1 & 1 & 0 & \cdots & 0 \\ S^{-1/2}d_2 & 0 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ S^{-1/2}d_i & 0 & 0 & \cdots & 0 \\ \vdots & & & & \\ S^{-1/2}d_p & 0 & 0 & \cdots & 1 \end{vmatrix}$$

$$= n \log 1 + \sum_{i=1}^{p} -d^\top S^{-1/2}{}_i (-1)^{2+i} S^{-1/2}d_i (-1)^{i+1}$$

$$= n \log(1 + d^\top S^{-1}d). \tag{7.4}$$

This statistic is a monotone function of $(n-1)d^\top S^{-1}d$. This means that $-2 \log \lambda > k$ if and only if $(n-1)d^\top S^{-1}d > k'$. The latter statistic has by Corollary 5.3, under H_0, a Hotelling's T^2-distribution. Therefore,

$$(n-1)(\bar{x} - \mu_0)^\top S^{-1}(\bar{x} - \mu_0) \sim T^2(p, n-1), \tag{7.5}$$

or equivalently

$$\left(\frac{n-p}{p}\right)(\bar{x} - \mu_0)^\top S^{-1}(\bar{x} - \mu_0) \sim F_{p,n-p}. \tag{7.6}$$

In this case an exact rejection region may be defined as

$$\left(\frac{n-p}{p}\right)(\bar{x} - \mu_0)^\top S^{-1}(\bar{x} - \mu_0) > F_{1-\alpha;p,n-p}.$$

Alternatively, we have from Theorem 7.1 that under H_0 the asymptotic distribution of the test statistic is

$$-2 \log \lambda \xrightarrow{\mathcal{L}} \chi_p^2, \quad \text{as } n \to \infty$$

which leads to the (asymptotically valid) rejection region

$$n \log\{1 + (\bar{x} - \mu_0)^\top S^{-1}(\bar{x} - \mu_0)\} > \chi_{1-\alpha;p}^2,$$

but of course, in this case, we would prefer to use the exact F-test provided just above.

Example 7.3 Consider the problem of Example 7.2 again. We know that $\mathcal{S}_f$ is the empirical analogue for Σ_f, the covariance matrix for the forged banknotes. The test statistic (7.5) has the value 1153.4 or its equivalent for the F distribution in (7.6) is 182.5 which is highly significant ($F_{0.95;6,94} = 2.1966$) so that we conclude that $\mu_f \neq \mu_0$.

Confidence Region for *μ*

When estimating a multidimensional parameter $\theta \in \mathbb{R}^k$ from a sample, we saw in Chapter 6 how to determine the estimator $\widehat{\theta} = \widehat{\theta}(\mathcal{X})$. For the observed data we end up with a point estimate, which is the corresponding observed value of $\widehat{\theta}$. We know $\widehat{\theta}(\mathcal{X})$ is a random variable and we often prefer to determine a *confidence region* for θ. A confidence region (CR) is a random subset of $\mathbb{R}^k$ (determined by appropriate statistics) such that we are "confident", at a certain given level $1 - \alpha$, that this region contains θ:

$$P(\theta \in CR) = 1 - \alpha.$$

This is just a multidimensional generalisation of the basic univariate confidence interval. Confidence regions are particularly useful when a hypothesis H_0 on θ is rejected, because they eventually help in identifying which component of θ is responsible for the rejection.

There are only a few cases where confidence regions can be easily assessed, and include most of the testing problems on mean presented in this section.

Corollary 5.3 provides a pivotal quantity which allows confidence regions for μ to be constructed. Since $(\frac{n-p}{p})(\bar{x} - \mu)^\top \mathcal{S}^{-1}(\bar{x} - \mu) \sim F_{p,n-p}$, we have

$$P\left\{\left(\frac{n - p}{p}\right)(\mu - \bar{x})^\top \mathcal{S}^{-1}(\mu - \bar{x}) < F_{1-\alpha;p,n-p}\right\} = 1 - \alpha.$$

Then,

$$CR = \left\{\mu \in \mathbb{R}^p \mid (\mu - \bar{x})^\top \mathcal{S}^{-1}(\mu - \bar{x}) \leq \frac{p}{n - p} F_{1-\alpha;p,n-p}\right\}$$

is a confidence region at level $(1 - \alpha)$ for μ. It is the interior of an iso-distance ellipsoid in $\mathbb{R}^p$ centred at $\bar{x}$, with a scaling matrix $\mathcal{S}^{-1}$ and a distance constant $(\frac{p}{n-p})F_{1-\alpha;p,n-p}$. When p is large, ellipsoids are not easy to handle for practical purposes. One is thus interested in finding confidence intervals for $\mu_1, \mu_2, \ldots, \mu_p$ so that simultaneous confidence on all the intervals reaches the desired level of say, $1 - \alpha$.

Below, we consider a more general problem. We construct *simultaneous confidence intervals* for all possible linear combinations $a^\top \mu$, $a \in \mathbb{R}^p$ of the elements of μ.

Suppose for a moment that we fix a particular projection vector a. We are back to a standard univariate problem of finding a confidence interval for the mean $a^\top \mu$

of a univariate random variable $a^\top X$. We can use the t-statistics and an obvious confidence interval for $a^\top \mu$ is given by the values $a^\top \mu$ such that

$$\left| \frac{\sqrt{n-1}(a^\top \mu - a^\top \bar{x})}{\sqrt{a^\top S a}} \right| \leq t_{1-\frac{\alpha}{2};n-1}$$

or equivalently

$$t^2(a) = \frac{(n-1)\{a^\top(\mu - \bar{x})\}^2}{a^\top S a} \leq F_{1-\alpha;1,n-1}.$$

This provides the $(1-\alpha)$ confidence interval for $a^\top \mu$:

$$\left(a^\top \bar{x} - \sqrt{F_{1-\alpha;1,n-1} \frac{a^\top S a}{n-1}} \leq a^\top \mu \leq a^\top \bar{x} + \sqrt{F_{1-\alpha;1,n-1} \frac{a^\top S a}{n-1}} \right).$$

Now it is easy to prove (using Theorem 2.5) that:

$$\max_a t^2(a) = (n-1)(\bar{x} - \mu)^\top S^{-1}(\bar{x} - \mu) \sim T^2(p, n-1).$$

Therefore, simultaneously for all $a \in \mathbb{R}^p$, the interval

$$\left(a^\top \bar{x} - \sqrt{K_\alpha a^\top S a}, \ a^\top \bar{x} + \sqrt{K_\alpha a^\top S a} \right) \tag{7.7}$$

where $K_\alpha = \frac{p}{n-p} F_{1-\alpha;p,n-p}$, will contain $a^\top \mu$ with probability $(1-\alpha)$.

A particular choice of a are the columns of the identity matrix $\mathcal{I}_p$, providing simultaneous confidence intervals for $\mu_1, \ldots, \mu_p$. We therefore have with probability $(1-\alpha)$ for $j = 1, \ldots, p$

$$\bar{x}_j - \sqrt{\frac{p}{n-p} F_{1-\alpha;p,n-p} s_{jj}} \leq \mu_j \leq \bar{x}_j + \sqrt{\frac{p}{n-p} F_{1-\alpha;p,n-p} s_{jj}}. \tag{7.8}$$

It should be noted that these intervals define a rectangle inscribing the confidence ellipsoid for μ given above. They are particularly useful when a null hypothesis H_0 of the type described above is rejected and one would like to see which component(s) are mainly responsible for the rejection.

Example 7.4 The 95% confidence region for μ_f, the mean of the forged banknotes, is given by the ellipsoid:

$$\left\{ \mu \in \mathbb{R}^6 \mid (\mu - \bar{x}_f)^\top S_f^{-1}(\mu - \bar{x}_f) \leq \frac{6}{94} F_{0.95;6,94} \right\}.$$

The 95% simultaneous confidence intervals are given by (we use $F_{0.95;6,94} = 2.1966$)

$$214.692 \leq \mu_1 \leq 214.954$$
$$130.205 \leq \mu_2 \leq 130.395$$
$$130.082 \leq \mu_3 \leq 130.304$$
$$10.108 \leq \mu_4 \leq 10.952$$
$$10.896 \leq \mu_5 \leq 11.370$$
$$139.242 \leq \mu_6 \leq 139.658.$$

Comparing the inequalities with $\mu_0 = (214.9, 129.9, 129.7, 8.3, 10.1, 141.5)^\top$ shows that almost all components (except the first one) are responsible for the rejection of μ_0 in Examples 7.2 and 7.3.

In addition, the method can provide other confidence intervals. We have at the same level of confidence (choosing $a^\top = (0, 0, 0, 1, -1, 0)$)

$$-1.211 \leq \mu_4 - \mu_5 \leq 0.005$$

showing that for the forged bills, the lower border is essentially smaller than the upper border.

Remark 7.1 It should be noted that the confidence region is an ellipsoid whose characteristics depend on the whole matrix S. In particular, the slope of the axis depends on the eigenvectors of S and therefore on the covariances s_{ij}. However, the rectangle inscribing the confidence ellipsoid provides the simultaneous confidence intervals for μ_j, $j = 1, \ldots, p$. They do not depend on the covariances s_{ij}, but only on the variances s_{jj} (see (7.8)). In particular, it may happen that a tested value μ_0 is covered by the confidence ellipsoid but not covered by the intervals (7.8). In this case, μ_0 is rejected by a test based on the simultaneous confidence intervals but not rejected by a test based on the confidence ellipsoid. The simultaneous confidence intervals are easier to handle than the full ellipsoid but we have lost some information, namely the covariance between the components (see Exercise 7.14).

The following problem concerns the covariance matrix in a multinormal population: in this situation the test statistic has a slightly more complicated distribution. We will therefore invoke the approximation of Theorem 7.1 in order to derive a test of approximate size α.

Test Problem 3 Suppose that $X_1, \ldots, X_n$ is an i.i.d. random sample from a $N_p(\mu, \Sigma)$ population.

$$H_0 : \Sigma = \Sigma_0, \ \mu \text{ unknown versus } H_1 : \text{ no constraints.}$$

Under H_0 we have $\widehat{\mu} = \overline{x}$, and $\Sigma = \Sigma_0$, whereas under H_1 we have $\widehat{\mu} = \overline{x}$, and $\widehat{\Sigma} = S$. Hence

$$\ell_0^* = \ell(\overline{x}, \Sigma_0) = -\frac{1}{2}n \log |2\pi \Sigma_0| - \frac{1}{2}n \operatorname{tr}(\Sigma_0^{-1}\mathcal{S})$$

$$\ell_1^* = \ell(\overline{x}, \mathcal{S}) = -\frac{1}{2}n \log |2\pi \mathcal{S}| - \frac{1}{2}np$$

and thus

$$-2 \log \lambda = 2(\ell_1^* - \ell_0^*)$$
$$= n \operatorname{tr}(\Sigma_0^{-1}\mathcal{S}) - n \log |\Sigma_0^{-1}\mathcal{S}| - np.$$

Note that this statistic is a function of the eigenvalues of $\Sigma_0^{-1}\mathcal{S}$. Unfortunately, the exact finite sample distribution of $-2 \log \lambda$ is very complicated. Asymptotically, we have under H_0

$$-2 \log \lambda \xrightarrow{\mathcal{L}} \chi_m^2 \quad \text{as } n \to \infty$$

with $m = \frac{1}{2}\{p(p+1)\}$, since a $(p \times p)$ covariance matrix has only these m parameters as a consequence of its symmetry.

Example 7.5 Consider the US companies data set (Table B.5) and suppose we are interested in the companies of the energy sector, analysing their assets (X_1) and sales (X_2). The sample is of size 15 and provides the value of $\mathcal{S} = 10^7 \times \begin{bmatrix} 1.6635 & 1.2410 \\ 1.2410 & 1.3747 \end{bmatrix}$. We want to test if $\operatorname{Var}\binom{X_1}{X_2} = 10^7 \times \begin{bmatrix} 1.2248 & 1.1425 \\ 1.1425 & 1.5112 \end{bmatrix} = \Sigma_0$. ($\Sigma_0$ is in fact the empirical variance matrix for X_1 and X_2 for the manufacturing sector.) The test statistic (**Q** MVAusenergy) turns out to be $-2 \log \lambda = 5.4046$ which is not significant for χ_3^2 (p-value $= 0.1445$). So we cannot conclude that $\Sigma \neq \Sigma_0$.

In the next testing problem, we address a question that was already stated in Chapter 3, Section 3.6: testing a particular value of the coefficients β in a linear model. The presentation is carried out in general terms so that it can be built on in the next section where we will test linear restrictions on β.

Test Problem 4 Suppose that $Y_1, \ldots, Y_n$ are independent r.v.'s with $Y_i \sim N_1(\beta^\top x_i, \sigma^2)$, $x_i \in \mathbb{R}^p$.

$$H_0 : \beta = \beta_0, \ \sigma^2 \text{ unknown versus } H_1 : \text{ no constraints.}$$

Under H_0 we have $\beta = \beta_0, \widehat{\sigma}_0^2 = \frac{1}{n}||y - \mathcal{X}\beta_0||^2$ and under H_1 we have $\hat{\beta} = (\mathcal{X}^\top\mathcal{X})^{-1}\mathcal{X}^\top y, \hat{\sigma}^2 = \frac{1}{n}||y - \mathcal{X}\beta||^2$ (see Example 6.3). Hence by Theorem 7.1

$$\begin{aligned} -2 \log \lambda &= 2(\ell_1^* - \ell_0^*) \\ &= n \log \left(\frac{||y - \mathcal{X}\beta_0||^2}{||y - \mathcal{X}\hat{\beta}||^2} \right) \\ &\xrightarrow{\mathcal{L}} \chi_p^2. \end{aligned}$$

We draw upon the result (3.45) which gives us

$$F = \frac{(n-p)}{p} \left(\frac{||y - \mathcal{X}\beta_0||^2}{||y - \mathcal{X}\hat{\beta}||^2} - 1 \right) \sim F_{p,n-p},$$

so that in this case we again have an exact distribution.

Example 7.6 Let us consider our "classic blue" pullovers again. In Example 3.11 we tried to model the dependency of sales on prices. As we have seen in Figure 3.5 the slope of the regression curve is rather small, hence we might ask if $\binom{\alpha}{\beta} = \binom{211}{0}$. Here

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_{10} \end{pmatrix} = \begin{pmatrix} x_{1,1} \\ \vdots \\ x_{10,1} \end{pmatrix}, \qquad \mathcal{X} = \begin{pmatrix} 1 & x_{1,2} \\ \vdots & \vdots \\ 1 & x_{10,2} \end{pmatrix}.$$

The test statistic for the LR test is

$$-2 \log \lambda = 9.10$$

which under the χ_2^2 distribution is significant. The exact F-test statistic

$$F = 5.93$$

is also significant under the $F_{2,8}$ distribution ($F_{2,8;0.95} = 4.46$).

Summary

→ The hypotheses $H_0 : \theta \in \Omega_0$ against $H_1 : \theta \in \Omega_1$ can be tested using the likelihood ratio test (LRT). The likelihood ratio (LR) is the quotient $\lambda(\mathcal{X}) = L_0^*/L_1^*$ where the L_j^* are the maxima of the likelihood for each of the hypotheses.

→ The test statistic in the LRT is $\lambda(\mathcal{X})$ or equivalently its logarithm $\log \lambda(\mathcal{X})$. If Ω_1 is q-dimensional and $\Omega_0 \subset \Omega_1$ r-dimensional, then the asymptotic distribution of $-2 \log \lambda$ is χ_{q-r}^2. This allows H_0 to be tested against H_1 by calculating the test statistic $-2 \log \lambda = 2(\ell_1^* - \ell_0^*)$ where $\ell_j^* = \log L_j^*$.

→ The hypothesis $H_0 : \mu = \mu_0$ for $X \sim N_p(\mu, \Sigma)$, where Σ is known, leads to $-2 \log \lambda = n(\overline{x} - \mu_0)^\top \Sigma^{-1} (\overline{x} - \mu_0) \sim \chi_p^2$.

→ The hypothesis $H_0 : \mu = \mu_0$ for $X \sim N_p(\mu, \Sigma)$, where Σ is unknown, leads to $-2 \log \lambda = n \log\{1 + (\overline{x} - \mu_0)^\top S^{-1} (\overline{x} - \mu_0)\} \longrightarrow \chi_p^2$, and $(n-1)(\overline{x} - \mu_0)^\top S^{-1} (\overline{x} - \mu_0) \sim T^2(p, n-1)$.

Summary (continued)
$\hookrightarrow$ The hypothesis $H_0 : \Sigma = \Sigma_0$ for $X \sim N_p(\mu, \Sigma)$, where μ is unknown, leads to $-2 \log \lambda = n \operatorname{tr}(\Sigma_0^{-1} \mathcal{S}) - n \log |\Sigma_0^{-1} \mathcal{S}| - np \longrightarrow \chi_m^2$, $m = \frac{1}{2} p(p+1)$.
$\hookrightarrow$ The hypothesis $H_0 : \beta = \beta_0$ for $Y_i \sim N_1(\beta^\top x_i, \sigma^2)$, where σ^2 is unknown, leads to $-2 \log \lambda = n \log \left(\frac{\|y - \mathcal{X}\beta_0\|^2}{\|y - \mathcal{X}\hat{\beta}\|^2} \right) \longrightarrow \chi_p^2$.

7.2 Linear Hypothesis

In this section, we present a very general procedure which allows a linear hypothesis to be tested, i.e., a linear restriction, either on a vector mean μ or on the coefficient β of a linear model. The presented technique covers many of the practical testing problems on means or regression coefficients.

Linear hypotheses are of the form $\mathcal{A}\mu = a$ with known matrices $\mathcal{A}(q \times p)$ and $a(q \times 1)$ with $q \leq p$.

Example 7.7 Let $\mu = (\mu_1, \mu_2)^\top$. The hypothesis that $\mu_1 = \mu_2$ can be equivalently written as:

$$\mathcal{A}\mu = \begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = 0 = a.$$

The general idea is to test a normal population $H_0 : \mathcal{A}\mu = a$ (restricted model) against the full model H_1 where no restrictions are put on μ. Due to the properties of the multinormal, we can easily adapt the Test Problems 1 and 2 to this new situation. Indeed we know, from Theorem 5.2, that $y_i = \mathcal{A}x_i \sim N_q(\mu_y, \Sigma_y)$, where $\mu_y = \mathcal{A}\mu$ and $\Sigma_y = \mathcal{A}\Sigma\mathcal{A}^\top$.

Testing the null $H_0 : \mathcal{A}\mu = a$, is the same as testing $H_0 : \mu_y = a$. The appropriate statistics are $\bar{y}$ and $\mathcal{S}_y$ which can be derived from the original statistics $\bar{x}$ and $\mathcal{S}$ available from $\mathcal{X}$:

$$\bar{y} = \mathcal{A}\bar{x}, \qquad \mathcal{S}_y = \mathcal{A}\mathcal{S}\mathcal{A}^\top.$$

Here the difference between the translated sample mean and the tested value is $d = \mathcal{A}\bar{x} - a$. We are now in the situation to proceed to Test Problems 5 and 6.

Test Problem 5 Suppose $X_1, \ldots, X_n$ is an i.i.d. random sample from a $N_p(\mu, \Sigma)$ population.

$$H_0 : \mathcal{A}\mu = a, \ \Sigma \text{ known versus } H_1 : \text{ no constraints.}$$

By (7.2) we have that, under H_0:

$$n(\mathcal{A}\bar{x} - a)^\top (\mathcal{A}\Sigma\mathcal{A}^\top)^{-1}(\mathcal{A}\bar{x} - a) \sim \mathcal{X}_q^2,$$

and we reject H_0 if this test statistic is too large at the desired significance level.

Example 7.8 We consider hypotheses on partitioned mean vectors $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$. Let us first look at

$$H_0 : \mu_1 = \mu_2, \text{ versus } H_1 : \text{no constraints,}$$

for $N_{2p}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix}\right)$ with known Σ. This is equivalent to $\mathcal{A} = (\mathcal{I}, -\mathcal{I})$, $a = (0, \ldots, 0)^\top \in \mathbb{R}^p$ and leads to

$$-2\log\lambda = n(\bar{x}_1 - \bar{x}_2)(2\Sigma)^{-1}(\bar{x}_1 - \bar{x}_2) \sim \chi_p^2.$$

Another example is the test whether $\mu_1 = 0$, i.e.,

$$H_0 : \mu_1 = 0, \text{ versus } H_1 : \text{no constraints,}$$

for $N_{2p}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix}\right)$ with known Σ. This is equivalent to $\mathcal{A}\mu = a$ with $\mathcal{A} = (\mathcal{I}, 0)$, and $a = (0, \ldots, 0)^\top \in \mathbb{R}^p$. Hence

$$-2\log\lambda = n\bar{x}_1\Sigma^{-1}\bar{x}_1 \sim \chi_p^2.$$

Test Problem 6 Suppose $X_1, \ldots, X_n$ is an i.i.d. random sample from a $N_p(\mu, \Sigma)$ population.

$$H_0 : \mathcal{A}\mu = a, \ \Sigma \text{ unknown versus } H_1 : \ \text{no constraints.}$$

From Corollary (5.4) and under H_0 it follows immediately that

$$(n-1)(\mathcal{A}\bar{x} - a)^\top (\mathcal{A}\mathcal{S}\mathcal{A}^\top)^{-1}(\mathcal{A}\bar{x} - a) \sim T^2(q, n-1) \qquad (7.9)$$

since indeed under H_0,

$$\mathcal{A}\bar{x} \sim N_q(a, n^{-1}\mathcal{A}\Sigma\mathcal{A}^\top)$$

is independent of

$$n\mathcal{A}\mathcal{S}\mathcal{A}^\top \sim W_q(\mathcal{A}\Sigma\mathcal{A}^\top, n-1).$$

Example 7.9 Let's come back again to the bank data set and suppose that we want to test if $\mu_4 = \mu_5$, i.e., the hypothesis that the lower border mean equals the larger border mean for the forged bills. In this case:

$$\mathcal{A} = (0 \quad 0 \quad 0 \quad 1 \quad -1 \quad 0)$$
$$a = 0.$$

The test statistic is:

$$99(\mathcal{A}\bar{x})^{\top}(\mathcal{A}S_f\mathcal{A}^{\top})^{-1}(\mathcal{A}\bar{x}) \sim T^2(1, 99) = F_{1,99}.$$

The observed value is 13.638 which is significant at the 5% level.

Repeated Measurements

In many situations, n independent sampling units are observed at p different times or under p different experimental conditions (different treatments, ...). So here we repeat p one-dimensional measurements on n different subjects. For instance, we observe the results from n students taking p different exams. We end up with a $(n \times p)$ matrix. We can thus consider the situation where we have $X_1, \ldots, X_n$ i.i.d. from a normal distribution $N_p(\mu, \Sigma)$ when there are p repeated measurements. The hypothesis of interest in this case is that there are no treatment effects, $H_0 : \mu_1 = \mu_2 = \cdots = \mu_p$. This hypothesis is a direct application of Test Problem 6. Indeed, introducing an appropriate matrix transform on μ we have

$$H_0 : \mathcal{C}\mu = 0 \quad \text{where } \mathcal{C}((p-1) \times p) = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & -1 \end{pmatrix}. \quad (7.10)$$

Note that in many cases one of the experimental conditions is the "control" (a placebo, standard drug or reference condition). Suppose it is the first component. In that case one is interested in studying differences to the control variable. The matrix $\mathcal{C}$ has therefore a different form

$$\mathcal{C}((p-1) \times p) = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}.$$

By (7.9) the null hypothesis will be rejected if:

$$\frac{(n-p+1)}{p-1}\bar{x}^{\top}\mathcal{C}^{\top}(\mathcal{C}S\mathcal{C}^{\top})^{-1}\mathcal{C}\bar{x} > F_{1-\alpha; p-1, n-p+1}.$$

As a matter of fact, $\mathcal{C}\mu$ is the mean of the random variable $y_i = \mathcal{C}x_i$

$$y_i \sim N_{p-1}(\mathcal{C}\mu, \mathcal{C}\Sigma\mathcal{C}^{\top}).$$

Simultaneous confidence intervals for linear combinations of the mean of y_i have been derived above in (7.7). For all $a \in \mathbb{R}^{p-1}$, with probability $(1 - \alpha)$ we have

$$a^\top C\mu \in a^\top C\bar{x} \pm \sqrt{\frac{(p-1)}{n-p+1} F_{1-\alpha;p-1,n-p+1} a^\top CSC^\top a}.$$

Due to the nature of the problem here, the row sums of the elements in C are zero: $C 1_p = 0$, therefore $a^\top C$ is a vector having sum of elements equals to 0. This is called a *contrast*. Let $b = C^\top a$. We have $b^\top 1_p = \sum_{j=1}^{p} b_j = 0$. The result above thus provides for all contrasts of μ, and $b^\top \mu$ simultaneous confidence intervals at level $(1 - \alpha)$

$$b^\top \mu \in b^\top \bar{x} \pm \sqrt{\frac{(p-1)}{n-p+1} F_{1-\alpha;p-1,n-p+1} b^\top Sb}.$$

Examples of contrasts for $p = 4$ are $b^\top = (1 \ -1 \ 0 \ 0)$ or $(1 \ 0 \ 0 \ -1)$ or even $(1 \ -\frac{1}{3} \ -\frac{1}{3} \ -\frac{1}{3})$ when the control is to be compared with the mean of 3 different treatments.

Example 7.10 Bock (1975) considers the evolution of the vocabulary of children from the eighth through eleventh grade. The data set contains the scores of a vocabulary test of 40 randomly chosen children. This is a repeated measurement situation, $(n = 40, p = 4)$, since the same children were observed from grades 8 to 11. The statistics of interest are:

$$\bar{x} = (1.086, 2.544, 2.851, 3.420)^\top$$
$$S = \begin{pmatrix} 2.902 & 2.438 & 2.963 & 2.183 \\ 2.438 & 3.049 & 2.775 & 2.319 \\ 2.963 & 2.775 & 4.281 & 2.939 \\ 2.183 & 2.319 & 2.939 & 3.162 \end{pmatrix}.$$

Suppose we are interested in the yearly evolution of the children. Then the matrix C providing successive differences of μ_j is:

$$C = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

The value of the test statistic is $F_{obs} = 53.134$ which is highly significant for $F_{3.37}$. There are significant differences between the successive means. However, the analysis of the contrasts shows the following simultaneous 95% confidence intervals

$$-1.958 \leq \mu_1 - \mu_2 \leq -0.959$$
$$-0.949 \leq \mu_2 - \mu_3 \leq 0.335$$
$$-1.171 \leq \mu_3 - \mu_4 \leq 0.036.$$

Thus, the rejection of H_0 is mainly due to the difference between the childrens' performances in the first and second year. The confidence intervals for the following contrasts may also be of interest:

$$-2.283 \le \mu_1 - \frac{1}{3}(\mu_2 + \mu_3 + \mu_4) \le -1.423$$

$$-1.777 \le \frac{1}{3}(\mu_1 + \mu_2 + \mu_3) - \mu_4 \le -0.742$$

$$-1.479 \le \qquad \mu_2 - \mu_4 \qquad \le -0.272.$$

They show that μ_1 is different from the average of the 3 other years (the same being true for μ_4) and μ_4 turns out to be higher than μ_2 (and of course higher than μ_1).

Test Problem 7 illustrates how the likelihood ratio can be applied to testing a linear restriction on the coefficient β of a linear model. It is also shown how a transformation of the test statistic leads to an exact F test as presented in Chapter 3.

Test Problem 7 Suppose $Y_1, \ldots, Y_n$, are independent with $Y_i \sim N_1(\beta^\top x_i, \sigma^2)$, and $x_i \in \mathbb{R}^p$.

$$H_0 : \mathcal{A}\beta = a, \ \sigma^2 \text{ unknown versus } H_1 : \text{ no constraints.}$$

To get the constrained maximum likelihood estimators under H_0, let $f(\beta, \lambda) = (y - x\beta)^\top (y - x\beta) - \lambda^\top (\mathcal{A}\beta - a)$ where $\lambda \in \mathbb{R}^q$ and solve $\frac{\partial f(\beta, \lambda)}{\partial \beta} = 0$ and $\frac{\partial f(\beta, \lambda)}{\partial \lambda} = 0$ (Exercise 3.24), thus we obtain:

$$\tilde{\beta} = \hat{\beta} - (\mathcal{X}^\top \mathcal{X})^{-1} \mathcal{A}^\top \{\mathcal{A}(\mathcal{X}^\top \mathcal{X})^{-1} \mathcal{A}^\top\}^{-1} (\mathcal{A}\hat{\beta} - a)$$

for β and $\tilde{\sigma}^2 = \frac{1}{n}(y - \mathcal{X}\tilde{\beta})^\top (y - \mathcal{X}\tilde{\beta})$. The estimate $\hat{\beta}$ denotes the unconstrained MLE as before. Hence, the LR statistic is

$$\begin{aligned}
-2\log\lambda &= 2(\ell_1^* - \ell_0^*) \\
&= n\log\left(\frac{\|y - \mathcal{X}\tilde{\beta}\|^2}{\|y - \mathcal{X}\hat{\beta}\|^2}\right) \\
&\xrightarrow{\mathcal{L}} \chi_q^2
\end{aligned}$$

where q is the number of elements of a. This problem also has an exact F-test since

$$\begin{aligned}
\frac{n-p}{q}&\left(\frac{\|y - \mathcal{X}\tilde{\beta}\|^2}{\|y - \mathcal{X}\hat{\beta}\|^2} - 1\right) \\
&= \frac{n-p}{q} \frac{(\mathcal{A}\hat{\beta} - a)^\top \{\mathcal{A}(\mathcal{X}^\top \mathcal{X})^{-1} \mathcal{A}^\top\}^{-1} (\mathcal{A}\hat{\beta} - a)}{(y - \mathcal{X}\hat{\beta})^\top (y - \mathcal{X}\hat{\beta})} \sim F_{q, n-p}.
\end{aligned}$$

Example 7.11 Let us continue with the "classic blue" pullovers. We can once more test if $\beta = 0$ in the regression of sales on prices. It holds that

$$\beta = 0 \quad \text{iff} \ (0 \quad 1) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = 0.$$

The LR statistic here is

$$-2 \log \lambda = 0.284$$

which is not significant for the χ_1^2 distribution. The F-test statistic

$$F = 0.231$$

is also not significant. Hence, we can assume independence of sales and prices (alone). Recall that this conclusion has to be revised if we consider the prices together with advertising costs and hours sales manager hours.

Recall the different conclusion that was made in Example 7.6 when we rejected $H_0 : \alpha = 211$ and $\beta = 0$. The rejection there came from the fact that the *pair of values* was rejected. Indeed, if $\beta = 0$ the estimator of α would be $\bar{y} = 172.70$ and this is too far from 211.

Example 7.12 Let us now consider the multivariate regression in the "classic blue" pullovers example. From Example 3.15 we know that the estimated parameters in the model

$$X_1 = \alpha + \beta_1 X_2 + \beta_2 X_3 + \beta_3 X_4 + \varepsilon$$

are

$$\hat{\alpha} = 65.670, \qquad \hat{\beta}_1 = -0.216, \qquad \hat{\beta}_2 = 0.485, \qquad \hat{\beta}_3 = 0.844.$$

Hence, we could postulate the approximate relation:

$$\beta_1 \approx -\frac{1}{2}\beta_2,$$

which means in practice that augmenting the price by 20 EUR requires the advertising costs to increase by 10 EUR in order to keep the number of pullovers sold constant. Vice versa, reducing the price by 20 EUR yields the same result as before if we reduced the advertising costs by 10 EUR. Let us now test whether the hypothesis

$$H_0 : \beta_1 = -\frac{1}{2}\beta_2$$

is valid. This is equivalent to

$$\begin{pmatrix} 0 & 1 & \frac{1}{2} & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = 0.$$

The LR statistic in this case is equal to (MVAlrtest)

$$-2\log\lambda = 0.012,$$

the F statistic is

$$F = 0.007.$$

Hence, in both cases we will not reject the null hypothesis.

Comparison of Two Mean Vectors

In many situations, we want to compare two groups of individuals for whom a set of p characteristics has been observed. We have two random samples $\{x_{i1}\}_{i=1}^{n_1}$ and $\{x_{j2}\}_{j=1}^{n_2}$ from two distinct p-variate normal populations. Several testing issues can be addressed in this framework. In Test Problem 8 we will first test the hypothesis of equal mean vectors in the two groups under the assumption of equality of the two covariance matrices. This task can be solved by adapting Test Problem 2.

In Test Problem 9 a procedure for testing the equality of the two covariance matrices is presented. If the covariance matrices differ, the procedure of Test Problem 8 is no longer valid. If the equality of the covariance matrices is rejected, an easy rule for comparing two means with no restrictions on the covariance matrices is provided in Test Problem 10.

Test Problem 8 Assume that $X_{i1} \sim N_p(\mu_1, \Sigma)$, with $i = 1, \ldots, n_1$ and $X_{j2} \sim N_p(\mu_2, \Sigma)$, with $j = 1, \ldots, n_2$, where all the variables are independent.

$$H_0 : \mu_1 = \mu_2, \text{ versus } H_1 : \text{ no constraints.}$$

Both samples provide the statistics $\bar{x}_k$ and $\mathcal{S}_k$, $k = 1, 2$. Let $\delta = \mu_1 - \mu_2$. We have

$$(\bar{x}_1 - \bar{x}_2) \sim N_p \left(\delta, \frac{n_1 + n_2}{n_1 n_2} \Sigma \right) \tag{7.11}$$

$$n_1 S_1 + n_2 S_2 \sim W_p(\Sigma, n_1 + n_2 - 2). \tag{7.12}$$

Let $\mathcal{S} = (n_1 + n_2)^{-1}(n_1 S_1 + n_2 S_2)$ be the weighted mean of $\mathcal{S}_1$ and $\mathcal{S}_2$. Since the two samples are independent and since $\mathcal{S}_k$ is independent of $\bar{x}_k$ (for $k = 1, 2$) it follows that $\mathcal{S}$ is independent of $(\bar{x}_1 - \bar{x}_2)$. Hence, Theorem 5.8 applies and leads to a T^2-distribution:

$$\frac{n_1 n_2 (n_1 + n_2 - 2)}{(n_1 + n_2)^2} \{(\bar{x}_1 - \bar{x}_2) - \delta\}^{\top} \mathcal{S}^{-1} \{(\bar{x}_1 - \bar{x}_2) - \delta\}) \sim T^2(p, n_1 + n_2 - 2) \tag{7.13}$$

or

$$\{(\bar{x}_1 - \bar{x}_2) - \delta\}^\top \mathcal{S}^{-1} \{(\bar{x}_1 - \bar{x}_2) - \delta\} \sim \frac{p(n_1 + n_2)^2}{(n_1 + n_2 - p - 1)n_1 n_2} F_{p, n_1 + n_2 - p - 1}.$$

This result, as in Test Problem 2, can be used to test H_0: $\delta = 0$ or to construct a confidence region for $\delta \in \mathbb{R}^p$. The rejection region is given by:

$$\frac{n_1 n_2 (n_1 + n_2 - p - 1)}{p(n_1 + n_2)^2} (\bar{x}_1 - \bar{x}_2)^\top \mathcal{S}^{-1} (\bar{x}_1 - \bar{x}_2) \geq F_{1-\alpha; p, n_1+n_2-p-1}. \qquad (7.14)$$

A $(1 - \alpha)$ confidence region for δ is given by the ellipsoid centred at $(\bar{x}_1 - \bar{x}_2)$

$$\{\delta - (\bar{x}_1 - \bar{x}_2)\}^\top \mathcal{S}^{-1} \{\delta - (\bar{x}_1 - \bar{x}_2)\} \leq \frac{p(n_1 + n_2)^2}{(n_1 + n_2 - p - 1)(n_1 n_2)} F_{1-\alpha; p, n_1+n_2-p-1},$$

and the simultaneous confidence intervals for all linear combinations $a^\top \delta$ of the elements of δ are given by

$$a^\top \delta \in a^\top (\bar{x}_1 - \bar{x}_2) \pm \sqrt{\frac{p(n_1 + n_2)^2}{(n_1 + n_2 - p - 1)(n_1 n_2)} F_{1-\alpha; p, n_1+n_2-p-1} a^\top \mathcal{S} a}.$$

In particular we have at the $(1 - \alpha)$ level, for $j = 1, \ldots, p$,

$$\delta_j \in (\bar{x}_{1j} - \bar{x}_{2j}) \pm \sqrt{\frac{p(n_1 + n_2)^2}{(n_1 + n_2 - p - 1)(n_1 n_2)} F_{1-\alpha; p, n_1+n_2-p-1} s_{jj}}. \qquad (7.15)$$

Example 7.13 Let us come back to the questions raised in Example 7.5. We compare the means of assets (X_1) and of sales (X_2) for two sectors, energy (group 1) and manufacturing (group 2). With $n_1 = 15$, $n_2 = 10$, and $p = 2$ we obtain the statistics:

$$\bar{x}_1 = \begin{pmatrix} 4084.0 \\ 2580.5 \end{pmatrix}, \qquad \bar{x}_2 = \begin{pmatrix} 4307.2 \\ 4925.2 \end{pmatrix}$$

and

$$\mathcal{S}_1 = 10^7 \begin{pmatrix} 1.6635 & 1.2410 \\ 1.2410 & 1.3747 \end{pmatrix}, \qquad \mathcal{S}_2 = 10^7 \begin{pmatrix} 1.2248 & 1.1425 \\ 1.1425 & 1.5112 \end{pmatrix},$$

so that

$$\mathcal{S} = 10^7 \begin{pmatrix} 1.4880 & 1.2016 \\ 1.2016 & 1.4293 \end{pmatrix}.$$

The observed value of the test statistic (7.14) is $F = 2.7036$. Since $F_{0.95; 2, 22} = 3.4434$ the hypothesis of equal means of the two groups is not rejected although it would be rejected at a less severe level $(F > F_{0.90; 2, 22} = 2.5613)$. By directly applying (7.15), the 95% simultaneous confidence intervals for the differences (🔍 MVAsimcidif) are obtained as:

$$-4628.6 \leq \mu_{1a} - \mu_{2a} \leq 4182.2$$
$$-6662.4 \leq \mu_{1s} - \mu_{2s} \leq 1973.0.$$

Example 7.14 In order to illustrate the presented test procedures it is interesting to analyse some simulated data. This simulation will point out the importance of the covariances in testing means. We created 2 independent normal samples in $\mathbb{R}^4$ of sizes $n_1 = 30$ and $n_2 = 20$ with:

$$\mu_1 = (8, 6, 10, 10)^\top$$
$$\mu_2 = (6, 6, 10, 13)^\top.$$

One may consider this as an example of $X = (X_1, \ldots, X_n)^\top$ being the students' scores from 4 tests, where the 2 groups of students were subjected to two different methods of teaching. First we simulate the two samples with $\Sigma = \mathcal{I}_4$ and obtain the statistics:

$$\bar{x}_1 = (7.607, 5.945, 10.213, 9.635)^\top$$
$$\bar{x}_2 = (6.222, 6.444, 9.560, 13.041)^\top$$
$$\mathcal{S}_1 = \begin{pmatrix} 0.812 & -0.229 & -0.034 & 0.073 \\ -0.229 & 1.001 & 0.010 & -0.059 \\ -0.034 & 0.010 & 1.078 & -0.098 \\ 0.073 & -0.059 & -0.098 & 0.823 \end{pmatrix}$$
$$\mathcal{S}_2 = \begin{pmatrix} 0.559 & -0.057 & -0.271 & 0.306 \\ -0.057 & 1.237 & 0.181 & 0.021 \\ -0.271 & 0.181 & 1.159 & -0.130 \\ 0.306 & 0.021 & -0.130 & 0.683 \end{pmatrix}.$$

The test statistic (7.14) takes the value $F = 60.65$ which is highly significant: the small variance allows the difference to be detected even with these relatively moderate sample sizes. We conclude (at the 95% level) that:

$$0.6213 \leq \delta_1 \leq 2.2691$$
$$-1.5217 \leq \delta_2 \leq 0.5241$$
$$-0.3766 \leq \delta_3 \leq 1.6830$$
$$-4.2614 \leq \delta_4 \leq -2.5494$$

which confirms that the means for X_1 and X_4 are different.

Consider now a different simulation scenario where the standard deviations are 4 times larger: $\Sigma = 16\mathcal{I}_4$. Here we obtain:

$$\bar{x}_1 = (7.312, 6.304, 10.840, 10.902)^\top$$
$$\bar{x}_2 = (6.353, 5.890, 8.604, 11.283)^\top$$
$$\mathcal{S}_1 = \begin{pmatrix} 21.907 & 1.415 & -2.050 & 2.379 \\ 1.415 & 11.853 & 2.104 & -1.864 \\ -2.050 & 2.104 & 17.230 & 0.905 \\ 2.379 & -1.864 & 0.905 & 9.037 \end{pmatrix}$$

$$\mathcal{S}_2 = \begin{pmatrix} 20.349 & -9.463 & 0.958 & -6.507 \\ -9.463 & 15.502 & -3.383 & -2.551 \\ 0.958 & -3.383 & 14.470 & -0.323 \\ -6.507 & -2.551 & -0.323 & 10.311 \end{pmatrix}.$$

Now the test statistic takes the value 1.54 which is no longer significant ($F_{0.95,4,45} = 2.58$). Now we cannot reject the null hypothesis (which we know to be false!) since the increase in variances prohibits the detection of differences of such magnitude.

The following situation illustrates once more the role of the covariances between covariates. Suppose that $\Sigma = 16\mathcal{I}_4$ as above but with $\sigma_{14} = \sigma_{41} = -3.999$ (this corresponds to a negative correlation $r_{41} = -0.9997$). We have:

$$\bar{x}_1 = (8.484, 5.908, 9.024, 10.459)^\top$$
$$\bar{x}_2 = (4.959, 7.307, 9.057, 13.803)^\top$$
$$\mathcal{S}_1 = \begin{pmatrix} 14.649 & -0.024 & 1.248 & -3.961 \\ -0.024 & 15.825 & 0.746 & 4.301 \\ 1.248 & 0.746 & 9.446 & 1.241 \\ -3.961 & 4.301 & 1.241 & 20.002 \end{pmatrix}$$
$$\mathcal{S}_2 = \begin{pmatrix} 14.035 & -2.372 & 5.596 & -1.601 \\ -2.372 & 9.173 & -2.027 & -2.954 \\ 5.596 & -2.027 & 9.021 & -1.301 \\ -1.601 & -2.954 & -1.301 & 9.593 \end{pmatrix}.$$

The value of F is 3.853 which is significant at the 5% level (p-value $= 0.0089$). So the null hypothesis $\delta = \mu_1 - \mu_2 = 0$ is outside the 95% confidence ellipsoid. However, the simultaneous confidence intervals, which do not take the covariances into account are given by:

$$-0.1837 \le \delta_1 \le 7.2343$$
$$-4.9452 \le \delta_2 \le 2.1466$$
$$-3.0091 \le \delta_3 \le 2.9438$$
$$-7.2336 \le \delta_4 \le 0.5450.$$

They contain the null value (see Remark 7.1 above) although they are very asymmetric for δ_1 and δ_4.

Example 7.15 Let us compare the vectors of means of the forged and the genuine bank notes. The matrices $\mathcal{S}_f$ and $\mathcal{S}_g$ were given in Example 3.1 and since here $n_f = n_g = 100$, $\mathcal{S}$ is the simple average of $\mathcal{S}_f$ and $\mathcal{S}_g : \mathcal{S} = \frac{1}{2}(\mathcal{S}_f + \mathcal{S}_g)$.

$$\bar{x}_g = (214.97, 129.94, 129.72, 8.305, 10.168, 141.52)^\top$$
$$\bar{x}_f = (214.82, 130.3, 130.19, 10.53, 11.133, 139.45)^\top.$$

The test statistic is given by (7.14) and turns out to be $F = 391.92$ which is highly significant for $F_{6,193}$. The 95% simultaneous confidence intervals for the differences $\delta_j = \mu_{gj} - \mu_{fj}$, $j = 1, \ldots, p$ are:

$$-0.0443 \leq \delta_1 \leq 0.3363$$
$$-0.5186 \leq \delta_2 \leq -0.1954$$
$$-0.6416 \leq \delta_3 \leq -0.3044$$
$$-2.6981 \leq \delta_4 \leq -1.7519$$
$$-1.2952 \leq \delta_5 \leq -0.6348$$
$$1.8072 \leq \delta_6 \leq 2.3268.$$

All of the components (except for the first one) show significant differences in the means. The main effects are taken by the lower border (X_4) and the diagonal (X_6).

The preceding test implicitly uses the fact that the two samples are extracted from two different populations with common variance Σ. In this case, the test statistic (7.14) measures the distance between the two centres of gravity of the two groups w.r.t. the common metric given by the pooled variance matrix $\mathcal{S}$. If $\Sigma_1 \neq \Sigma_2$ no such matrix exists. There are no satisfactory test procedures for testing the equality of variance matrices which are robust with respect to normality assumptions of the populations. The following test extends Bartlett's test for equality of variances in the univariate case. But this test is known to be very sensitive to departures from normality.

Test Problem 9 (Comparison of Covariance Matrices) Let $X_{ih} \sim N_p(\mu_h, \Sigma_h)$, $i = 1, \ldots, n_h, h = 1, \ldots, k$ be independent random variables,

$$H_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_k \text{ versus } H_1 : \text{ no constraints.}$$

Each sub-sample provides $\mathcal{S}_h$, an estimator of Σ_h, with

$$n_h \mathcal{S}_h \sim W_p(\Sigma_h, n_h - 1).$$

Under H_0, $\sum_{h=1}^{k} n_h \mathcal{S}_h \sim W_p(\Sigma, n - k)$ (Section 5.2), where Σ is the common covariance matrix X_{ih} and $n = \sum_{h=1}^{k} n_h$. Let $\mathcal{S} = \frac{n_1 \mathcal{S}_1 + \cdots + n_k \mathcal{S}_k}{n}$ be the weighted average of the $\mathcal{S}_h$ (this is in fact the MLE of Σ when H_0 is true). The likelihood ratio test leads to the statistic

$$-2 \log \lambda = n \log |\mathcal{S}| - \sum_{h=1}^{k} n_h \log |\mathcal{S}_h| \tag{7.16}$$

which under H_0 is approximately distributed as a $\mathcal{X}_m^2$ where $m = \frac{1}{2}(k-1)p(p+1)$.

Example 7.16 Let's come back to Example 7.13, where the mean of assets and sales have been compared for companies from the energy and manufacturing sector assuming that $\Sigma_1 = \Sigma_2$. The test of $\Sigma_1 = \Sigma_2$ leads to the value of the test statistic

$$-2 \log \lambda = 0.9076 \tag{7.17}$$

which is not significant (p-value for a $\chi_3^2 = 0.82$). We cannot reject H_0 and the comparison of the means performed above is valid.

Example 7.17 Let us compare the covariance matrices of the forged and the genuine bank notes (the matrices S_f and S_g are shown in Example 3.1). A first look seems to suggest that $\Sigma_1 \neq \Sigma_2$. The pooled variance $\mathcal{S}$ is given by $\mathcal{S} = \frac{1}{2}(\mathcal{S}_f + \mathcal{S}_g)$ since here $n_f = n_g$. The test statistic here is $-2 \log \lambda = 127.21$, which is highly significant χ^2 with 21 degrees of freedom. As expected, we reject the hypothesis of equal covariance matrices, and as a result the procedure for comparing the two means in Example 7.15 is not valid.

What can we do with unequal covariance matrices? When both n_1 and n_2 are large, we have a simple solution:

Test Problem 10 (Comparison of two means, unequal covariance matrices, large samples) Assume that $X_{i1} \sim N_p(\mu_1, \Sigma_1)$, with $i = 1, \ldots, n_1$ and $X_{j2} \sim N_p(\mu_2, \Sigma_2)$, with $j = 1, \ldots, n_2$ are independent random variables.

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \text{ no constraints.}$$

Letting $\delta = \mu_1 - \mu_2$, we have

$$(\bar{x}_1 - \bar{x}_2) \sim N_p\left(\delta, \frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2}\right).$$

Therefore, by (5.4)

$$(\bar{x}_1 - \bar{x}_2)^\top \left(\frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2}\right)^{-1} (\bar{x}_1 - \bar{x}_2) \sim \chi_p^2.$$

Since $\mathcal{S}_i$ is a consistent estimator of Σ_i for $i = 1, 2$, we have

$$(\bar{x}_1 - \bar{x}_2)^\top \left(\frac{\mathcal{S}_1}{n_1} + \frac{\mathcal{S}_2}{n_2}\right)^{-1} (\bar{x}_1 - \bar{x}_2) \xrightarrow{\mathcal{L}} \chi_p^2. \tag{7.18}$$

This can be used in place of (7.13) for testing H_0, defining a confidence region for δ or constructing simultaneous confidence intervals for δ_j, $j = 1, \ldots, p$.

For instance, the rejection region at the level α will be

$$(\bar{x}_1 - \bar{x}_2)^\top \left(\frac{\mathcal{S}_1}{n_1} + \frac{\mathcal{S}_2}{n_2}\right)^{-1} (\bar{x}_1 - \bar{x}_2) > \chi_{1-\alpha; p}^2 \tag{7.19}$$

and the $(1 - \alpha)$ simultaneous confidence intervals for δ_j, $j = 1, \ldots, p$ are:

$$\delta_j \in (\bar{x}_1 - \bar{x}_2) \pm \sqrt{\chi^2_{1-\alpha;p} \left(\frac{s_{jj}^{(1)}}{n_1} + \frac{s_{jj}^{(2)}}{n_2} \right)} \tag{7.20}$$

where $s_{jj}^{(i)}$ is the (j, j) element of the matrix $\mathcal{S}_i$. This may be compared to (7.15) where the pooled variance was used.

Remark 7.2 We see, by comparing the statistics (7.19) with (7.14), that we measure here the distance between $\bar{x}_1$ and $\bar{x}_2$ using the metric $(\frac{\mathcal{S}_1}{n_1} + \frac{\mathcal{S}_2}{n_2})$. It should be noted that when $n_1 = n_2$, the two methods are essentially the same since then $\mathcal{S} = \frac{1}{2}(\mathcal{S}_1 + \mathcal{S}_2)$. If the covariances are different but have the same eigenvectors (different eigenvalues), one can apply the common principal component (CPC) technique, see Chapter 10.

Example 7.18 Let us use the last test to compare the forged and the genuine bank notes again (n_1 and n_2 are both large). The test statistic (7.19) turns out to be 2436.8 which is again highly significant. The 95% simultaneous confidence intervals are:

$$-0.0389 \leq \delta_1 \leq 0.3309$$
$$-0.5140 \leq \delta_2 \leq -0.2000$$
$$-0.6368 \leq \delta_3 \leq -0.3092$$
$$-2.6846 \leq \delta_4 \leq -1.7654$$
$$-1.2858 \leq \delta_5 \leq -0.6442$$
$$1.8146 \leq \delta_6 \leq 2.3194$$

showing that all the components except the first are different from zero, the largest difference coming from X_6 (length of the diagonal) and X_4 (lower border). The results are very similar to those obtained in Example 7.15. This is due to the fact that here $n_1 = n_2$ as we already mentioned in the remark above.

Profile Analysis

Another useful application of Test Problem 6 is the repeated measurements problem applied to two independent groups. This problem arises in practice when we observe repeated measurements of characteristics (or measures of the same type under different experimental conditions) on the different groups which have to be compared. It is important that the p measures (the "profile") are comparable, and, in particular, are reported in the same units. For instance, they may be measures of blood pressure at p different points in time, one group being the control group and the other the group receiving a new treatment. The observations may be the scores obtained from p different tests of two different experimental groups. One is then interested in
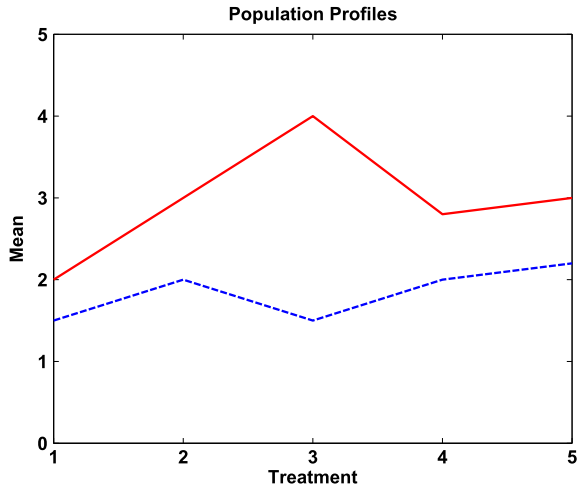
comparing the profiles of each group: the profile being just the vectors of the means
of the p responses (the comparison may be visualised in a two dimensional graph
using the parallel coordinate plot introduced in Section 1.7).

We are thus in the same statistical situation as for the comparison of two means:

$$X_{i1} \sim N_p(\mu_1, \Sigma) \quad i = 1, \ldots, n_1$$
$$X_{i2} \sim N_p(\mu_2, \Sigma) \quad i = 1, \ldots, n_2$$

where all variables are independent. Suppose the two population profiles look like
in Figure 7.1.

The following questions are of interest:

1. Are the profiles similar in the sense of being parallel (which means no interaction
   between the treatments and the groups)?
2. If the profiles are parallel, are they at the same level?
3. If the profiles are parallel, is there any treatment effect, i.e., are the profiles hori-
   zontal (profiles remain the same no matter which treatment received)?

The above questions are easily translated into linear constraints on the means and a
test statistic can be obtained accordingly.

Parallel Profiles

Let $\mathcal{C}$ be a $(p-1) \times p$ matrix defined as

$$\mathcal{C} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & -1 \end{pmatrix}.$$

The hypothesis to be tested is

$$H_0^{(1)} : \mathcal{C}(\mu_1 - \mu_2) = 0.$$

From (7.11), (7.12) and Corollary 5.4 we know that under H_0:

$$\frac{n_1 n_2}{(n_1 + n_2)^2}(n_1 + n_2 - 2)\{\mathcal{C}(\bar{x}_1 - \bar{x}_2)\}^\top (\mathcal{C}\mathcal{S}\mathcal{C}^\top)^{-1}\mathcal{C}(\bar{x}_1 - \bar{x}_2)$$

$$\sim T^2(p - 1, n_1 + n_2 - 2) \tag{7.21}$$

where $\mathcal{S}$ is the pooled covariance matrix. The hypothesis is rejected if

$$\frac{n_1 n_2 (n_1 + n_1 - p)}{(n_1 + n_2)^2 (p - 1)}(\mathcal{C}\bar{x})^\top (\mathcal{C}\mathcal{S}\mathcal{C}^\top)^{-1}\mathcal{C}\bar{x} > F_{1-\alpha; p-1, n_1+n_2-p}.$$

Equality of Two Levels

The question of equality of the two levels is meaningful only if the two profiles are parallel. In the case of interactions (rejection of $H_0^{(1)}$), the two populations react differently to the treatments and the question of the level has no meaning.

The equality of the two levels can be formalised as

$$H_0^{(2)} : 1_p^\top (\mu_1 - \mu_2) = 0$$

since

$$1_p^\top (\bar{x}_1 - \bar{x}_2) \sim N_1 \left(1_p^\top (\mu_1 - \mu_2), \frac{n_1 + n_2}{n_1 n_2} 1_p^\top \Sigma 1_p\right)$$

and

$$(n_1 + n_2)1_p^\top \mathcal{S}1_p \sim W_1(1_p^\top \Sigma 1_p, n_1 + n_2 - 2).$$

Using Corollary 5.4 we have that:

$$\frac{n_1 n_2}{(n_1 + n_2)^2}(n_1 + n_2 - 2)\frac{\{1_p^\top (\bar{x}_1 - \bar{x}_2)\}^2}{1_p^\top \mathcal{S}1_p} \sim T^2(1, n_1 + n_2 - 2) \tag{7.22}$$

$$= F_{1, n_1+n_2-2}.$$

The rejection region is

$$\frac{n_1 n_2 (n_1 + n_2 - 2)}{(n_1 + n_2)^2}\frac{\{1_p^\top (\bar{x}_1 - \bar{x}_2)\}^2}{1_p^\top \mathcal{S}1_p} > F_{1-\alpha; 1, n_1+n_2-2}.$$

Treatment Effect

If it is rejected that the profiles are parallel, then two independent analyses should be done on the two groups using the repeated measurement approach. But if it is accepted that they are parallel, then we can exploit the information contained in both groups (possibly at different levels) to test a treatment effect, i.e., if the two profiles are horizontal. This may be written as:

$$H_0^{(3)} : \mathcal{C}(\mu_1 + \mu_2) = 0.$$

Consider the average profile $\bar{x}$

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}.$$

Clearly,

$$\bar{x} \sim N_p\left(\frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}, \frac{1}{n_1 + n_2}\Sigma\right).$$

Now it is not hard to prove that $H_0^{(3)}$ with $H_0^{(1)}$ implies that

$$\mathcal{C}\left(\frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2}\right) = 0.$$

So under parallel, horizontal profiles we have

$$\sqrt{n_1 + n_2}\mathcal{C}\bar{x} \sim N_p(0, \mathcal{C}\Sigma\mathcal{C}^\top).$$

From Corollary 5.4 we again obtain

$$(n_1 + n_2 - 2)(\mathcal{C}\bar{x})^\top (\mathcal{C}\mathcal{S}\mathcal{C}^\top)^{-1}\mathcal{C}\bar{x} \sim T^2(p - 1, n_1 + n_2 - 2). \qquad (7.23)$$

This leads to the rejection region of $H_0^{(3)}$, namely

$$\frac{n_1 + n_2 - p}{p - 1}(\mathcal{C}\bar{x})^\top (\mathcal{C}\mathcal{S}\mathcal{C}^\top)^{-1}\mathcal{C}\bar{x} > F_{1-\alpha;p-1,n_1+n_2-p}.$$

Example 7.19 Morrison (1990b) proposed a test in which the results of 4 sub-tests of the Wechsler Adult Intelligence Scale (WAIS) are compared for 2 categories of people: group 1 contains $n_1 = 37$ people who do not have a senile factor and group 2 contains $n_2 = 12$ people who have a senile factor. The four WAIS sub-tests are X_1 (information), X_2 (similarities), X_3 (arithmetic) and X_4 (picture completion). The relevant statistics are

$$\bar{x}_1 = (12.57, 9.57, 11.49, 7.97)^\top$$
$$\bar{x}_2 = (8.75, 5.33, 8.50, 4.75)^\top$$

$$S_1 = \begin{pmatrix} 11.164 & 8.840 & 6.210 & 2.020 \\ 8.840 & 11.759 & 5.778 & 0.529 \\ 6.210 & 5.778 & 10.790 & 1.743 \\ 2.020 & 0.529 & 1.743 & 3.594 \end{pmatrix}$$

$$S_2 = \begin{pmatrix} 9.688 & 9.583 & 8.875 & 7.021 \\ 9.583 & 16.722 & 11.083 & 8.167 \\ 8.875 & 11.083 & 12.083 & 4.875 \\ 7.021 & 8.167 & 4.875 & 11.688 \end{pmatrix}.$$

The test statistic for testing if the two profiles are parallel is $F = 0.4634$, which is not significant (p-value $= 0.71$). Thus it is accepted that the two are parallel. The second test statistic (testing the equality of the levels of the 2 profiles) is $F = 17.21$, which is highly significant (p-value $\approx 10^{-4}$). The global level of the test for the non-senile people is superior to the senile group. The final test (testing the horizontality of the average profile) has the test statistic $F = 53.32$, which is also highly significant (p-value $\approx 10^{-14}$). This implies that there are substantial differences among the means of the different subtests.

Summary

$\hookrightarrow$ Hypotheses about μ can often be written as $\mathcal{A}\mu = a$, with matrix $\mathcal{A}$, and vector a.

$\hookrightarrow$ The hypothesis $H_0 : \mathcal{A}\mu = a$ for $X \sim N_p(\mu, \Sigma)$ with Σ known leads to $-2\log\lambda = n(\mathcal{A}\bar{x} - a)^\top (\mathcal{A}\Sigma\mathcal{A}^\top)^{-1}(\mathcal{A}\bar{x} - a) \sim \chi_q^2$, where q is the number of elements in a.

$\hookrightarrow$ The hypothesis $H_0 : \mathcal{A}\mu = a$ for $X \sim N_p(\mu, \Sigma)$ with Σ unknown leads to $-2\log\lambda = n\log\{1 + (\mathcal{A}\bar{x} - a)^\top (\mathcal{A}\mathcal{S}\mathcal{A}^\top)^{-1}(\mathcal{A}\bar{x} - a)\} \longrightarrow \chi_q^2$, where q is the number of elements in a and we have an exact test $(n-1)(\mathcal{A}\bar{x} - a)^\top (\mathcal{A}\mathcal{S}\mathcal{A}^\top)^{-1}(\mathcal{A}\bar{x} - a) \sim T^2(q, n-1)$.

$\hookrightarrow$ The hypothesis $H_0 : \mathcal{A}\beta = a$ for $Y_i \sim N_1(\beta^\top x_i, \sigma^2)$ with σ^2 unknown leads to $-2\log\lambda = \frac{n}{2}\log\left(\frac{\|y - \mathcal{X}\tilde{\beta}\|^2}{\|y - \mathcal{X}\hat{\beta}\|^2} - 1\right) \longrightarrow \chi_q^2$, with q being the length of a and with

$$\frac{n-p}{q} \frac{(\mathcal{A}\hat{\beta} - a)\{\mathcal{A}(\mathcal{X}^\top \mathcal{X})^{-1}\mathcal{A}^\top\}^{-1}(\mathcal{A}\hat{\beta} - a)}{(y - \mathcal{X}\hat{\beta})^\top (y - \mathcal{X}\hat{\beta})} \sim F_{q, n-p}.$$

7.3 Boston Housing

Returning to the Boston housing data set, we are now in a position to test if the means of the variables vary according to their location, for example, when they are located in a district with high valued houses. In Chapter 1, we built 2 groups of observations according to the value of X_{14} being less than or equal to the median of X_{14} (a group of 256 districts) and greater than the median (a group of 250 districts). In what follows, we use the transformed variables motivated in Section 1.9.

   Testing the equality of the means from the two groups was proposed in a multivariate setup, so we restrict the analysis to the variables X_1, X_5, X_8, X_{11}, and X_{13} to see if the differences between the two groups that were identified in Chapter 1 can be confirmed by a formal test. As in Test Problem 8, the hypothesis to be tested is

$$H_0: \ \mu_1 = \mu_2, \quad \text{where } \mu_1 \in \mathbb{R}^5, n_1 = 256, \text{ and } n_2 = 250.$$

Σ is not known. The F-statistic given in (7.13) is equal to 126.30, which is much higher than the critical value $F_{0.95;5,500} = 2.23$. Therefore, we reject the hypothesis of equal means.

   To see which component, X_1, X_5, X_8, X_{11}, or X_{13}, is responsible for this rejection, take a look at the simultaneous confidence intervals defined in (7.14):

$$\delta_1 \in (\ \ 1.4020, \ \ \ 2.5499)$$
$$\delta_5 \in (\ \ 0.1315, \ \ \ 0.2383)$$
$$\delta_8 \in (-0.5344, -0.2222)$$
$$\delta_{11} \in (\ \ 1.0375, \ \ \ 1.7384)$$
$$\delta_{13} \in (\ \ 1.1577, \ \ \ 1.5818).$$

These confidence intervals confirm that all of the δ_j are significantly different from zero (note there is a negative effect for X_8: weighted distances to employment centres) **Q** MVAsimcibh.

   We could also check if the factor "being bounded by the river" (variable X_4) has some effect on the other variables. To do this compare the means of $(X_5, X_8, X_9, X_{12}, X_{13}, X_{14})^\top$. There are two groups: $n_1 = 35$ districts bounded by the river and $n_2 = 471$ districts not bounded by the river. Test Problem 8 ($H_0: \mu_1 = \mu_2$) is applied again with $p = 6$. The resulting test statistic, $F = 5.81$, is highly significant ($F_{0.95;6,499} = 2.12$). The simultaneous confidence intervals indicate that only X_{14} (the value of the houses) is responsible for the hypothesis being rejected. At a significance level of 0.95

$$\delta_5 \in (-0.0603, 0.1919)$$
$$\delta_8 \in (-0.5225, 0.1527)$$
$$\delta_9 \in (-0.5051, 0.5938)$$
$$\delta_{12} \in (-0.3974, 0.7481)$$
$$\delta_{13} \in (-0.8595, 0.3782)$$
$$\delta_{14} \in (\ \ 0.0014, 0.5084).$$

Testing Linear Restrictions

In Chapter 3 a linear model was proposed that explained the variations of the price X_{14} by the variations of the other variables. Using the same procedure that was shown in Testing Problem 7, we are in a position to test a set of linear restrictions on the vector of regression coefficients β.

The model we estimated in Section 3.7 provides the following (MVAlin-regbh):

Variable	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	t	p-value
constant	4.1769	0.3790	11.020	0.0000
X_1	−0.0146	0.0117	−1.254	0.2105
X_2	0.0014	0.0056	0.247	0.8051
X_3	−0.0127	0.0223	−0.570	0.5692
X_4	0.1100	0.0366	3.002	0.0028
X_5	−0.2831	0.1053	−2.688	0.0074
X_6	0.4211	0.1102	3.822	0.0001
X_7	0.0064	0.0049	1.317	0.1885
X_8	−0.1832	0.0368	−4.977	0.0000
X_9	0.0684	0.0225	3.042	0.0025
X_{10}	−0.2018	0.0484	−4.167	0.0000
X_{11}	−0.0400	0.0081	−4.946	0.0000
X_{12}	0.0445	0.0115	3.882	0.0001
X_{13}	−0.2626	0.0161	−16.320	0.0000

Recall that the estimated residuals $Y - \mathcal{X}\widehat{\beta}$ did not show a big departure from normality, which means that the testing procedure developed above can be used.

1. First a global test of significance for the regression coefficients is performed,

$$H_0: (\beta_1, \ldots, \beta_{13}) = 0.$$

This is obtained by defining $\mathcal{A} = (0_{13}, \mathcal{I}_{13})$ and $a = 0_{13}$ so that H_0 is equivalent to $\mathcal{A}\beta = a$ where $\beta = (\beta_0, \beta_1, \ldots, \beta_{13})^\top$. Based on the observed values $F = 123.20$. This is highly significant ($F_{0.95; 13, 492} = 1.7401$), thus we reject H_0. Note that under H_0 $\widehat{\beta}_{H_0} = (3.0345, 0, \ldots, 0)$ where $3.0345 = \bar{y}$.
2. Since we are interested in the effect that being located close to the river has on the value of the houses, the second test is $H_0: \beta_4 = 0$. This is done by fixing

$$\mathcal{A} = (0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0)^\top$$

and $a = 0$ to obtain the equivalent hypothesis $H_0: \mathcal{A}\beta = a$. The result is again significant: $F = 9.0125$ ($F_{0.95; 1, 492} = 3.8604$) with a p-value of 0.0028. Note

that this is the same p-value obtained in the individual test $\beta_4 = 0$ in Chapter 3, computed using a different setup.

3. A third test notices the fact that some of the regressors in the full model (3.57) appear to be insignificant (that is they have high individual p-values). It can be confirmed from a joint test if the corresponding reduced model, formulated by deleting the insignificant variables, is rejected by the data. We want to test $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_7 = 0$. Hence,

$$
\mathcal{A} = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
$$

and $a = 0_4$. The test statistic is 0.9344, which is not significant for $F_{4,492}$. Given that the p-value is equal to 0.44, we cannot reject the null hypothesis nor the corresponding reduced model. The value of $\widehat{\beta}$ under the null hypothesis is

$$
\begin{aligned}
\widehat{\beta}_{H_0} = (&4.16, 0, 0, 0, 0.11, -0.31, 0.47, 0, -0.19, 0.05, -0.20, -0.04, \\
&0.05, -0.26)^\top.
\end{aligned}
$$

A possible reduced model is

$$
X_{14} = \beta_0 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_8 X_8 + \cdots + \beta_{13} X_{13} + \varepsilon.
$$

Estimating this reduced model using OLS, as was done in Chapter 3, provides the results shown in Table 7.1.

Note that the reduced model has $r^2 = 0.763$ which is very close to $r^2 = 0.765$ obtained from the full model. Clearly, including variables X_1, X_2, X_3, and X_7 does not provide valuable information in explaining the variation of X_{14}, the price of the houses.

Table 7.1 Linear regression for Boston housing data set
MVAlinreg2bh

Variable	$\widehat{\beta}_j$	SE	t	p-value
const	4.1582	0.3628	11.462	0.0000
X_4	0.1087	0.0362	2.999	0.0028
X_5	-0.3055	0.0973	-3.140	0.0018
X_6	0.4668	0.1059	4.407	0.0000
X_8	-0.1855	0.0327	-5.679	0.0000
X_9	0.0492	0.0183	2.690	0.0074
X_{10}	-0.2096	0.0446	-4.705	0.0000
X_{11}	-0.0410	0.0078	-5.280	0.0000
X_{12}	0.0481	0.0112	4.306	0.0000
X_{13}	-0.2588	0.0149	-17.396	0.0000

7.4 Exercises

Exercise 7.1 Use Theorem 7.1 to derive a test for testing the hypothesis that a dice is balanced, based on n tosses of that dice. (Hint: use the multinomial probability function.)

Exercise 7.2 Consider $N_3(\mu, \Sigma)$. Formulate the hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ in terms of $\mathcal{A}\mu = a$.

Exercise 7.3 Simulate a normal sample with $\mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 2 \end{pmatrix}$ and test $H_0 : 2\mu_1 - \mu_2 = 0.2$ first with Σ known and then with Σ unknown. Compare the results.

Exercise 7.4 Derive expression (7.3) for the likelihood ratio test statistic in Test Problem 2.

Exercise 7.5 With the simulated data set of Example 7.14, test the hypothesis of equality of the covariance matrices.

Exercise 7.6 In the U.S. companies data set, test the equality of means between the energy and manufacturing sectors, taking the full vector of observations X_1 to X_6. Derive the simultaneous confidence intervals for the differences.

Exercise 7.7 Let $X \sim N_2(\mu, \Sigma)$ where Σ is known to be $\Sigma = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$. We have an i.i.d. sample of size $n = 6$ providing $\bar{x}^\top = (1 \ \frac{1}{2})$. Solve the following test problems ($\alpha = 0.05$):

$$
\begin{aligned}
&\text{a)} \quad H_0 : \mu = \left(2, \tfrac{2}{3}\right)^\top && H_1 : \mu \neq \left(2, \tfrac{2}{3}\right)^\top \\
&\text{b)} \quad H_0 : \mu_1 + \mu_2 = \tfrac{7}{2} && H_1 : \mu_1 + \mu_2 \neq \tfrac{7}{2} \\
&\text{c)} \quad H_0 : \mu_1 - \mu_2 = \tfrac{1}{2} && H_1 : \mu_1 - \mu_2 \neq \tfrac{1}{2} \\
&\text{d)} \quad H_0 : \mu_1 = 2 && H_1 : \mu_1 \neq 2.
\end{aligned}
$$

For each case, represent the rejection region graphically (comment).

Exercise 7.8 Repeat the preceeding exercise with Σ unknown and $S = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$. Compare the results.

Exercise 7.9 Consider $X \sim N_3(\mu, \Sigma)$. An i.i.d. sample of size $n = 10$ provides:

$$
\bar{x} = (1, 0, 2)^\top
$$

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

a) Knowing that the eigenvalues of S are integers, describe a 95% confidence region for μ. (Hint: to compute eigenvalues use $|S| = \prod_{j=1}^{3} \lambda_j$ and $\text{tr}(S) = \sum_{j=1}^{3} \lambda_j$.)
b) Calculate the simultaneous confidence intervals for μ_1, μ_2 and μ_3.
c) Can we assert that μ_1 is an average of μ_2 and μ_3?

Exercise 7.10 Consider two independent i.i.d. samples, each of size 10, from two bivariate normal populations. The results are summarised below:

$$\bar{x}_1 = (3, 1)^\top; \qquad \bar{x}_2 = (1, 1)^\top$$

$$S_1 = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}; \qquad S_2 = \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix}.$$

Provide a solution to the following tests:

a) $H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$
b) $H_0 : \mu_{11} = \mu_{21}$ $H_1 : \mu_{11} \neq \mu_{21}$
c) $H_0 : \mu_{12} = \mu_{22}$ $H_1 : \mu_{12} \neq \mu_{22}$.

Compare the solutions and comment.

Exercise 7.11 Prove expression (7.4) in the Test Problem 2 with log-likelihoods ℓ_0^* and ℓ_1^*. (Hint: use (2.29).)

Exercise 7.12 Assume that $X \sim N_p(\mu, \Sigma)$ where Σ is unknown.

a) Derive the log likelihood ratio test for testing the independence of the p components, that is $H_0 : \Sigma$ is a diagonal matrix. (Solution: $-2 \log \lambda = -n \log |\mathcal{R}|$ where $\mathcal{R}$ is the correlation matrix, which is asymptotically a $\chi^2_{\frac{1}{2} p(p-1)}$ under H_0.)
b) Assume that Σ is a diagonal matrix (all the variables are independent). Can an asymptotic test for $H_0 : \mu = \mu_o$ against $H_1 : \mu \neq \mu_o$ be derived? How would this compare to p independent univariate t-tests on each μ_j?
c) Show an easy derivation of an asymptotic test for testing the equality of the p means (Hint: use $(C\bar{X})^\top (CSC^\top)^{-1} C\bar{X} \to \chi^2_{p-1}$ where $S = \text{diag}(s_{11}, \ldots, s_{pp})$ and C is defined as in (7.10).) Compare this to the simple ANOVA procedure used in Section 3.5.

Exercise 7.13 The yields of wheat have been measured in 30 parcels that have been randomly attributed to 3 lots prepared by one of 3 different fertilisers A, B and C. The data are

Fertilizer yield	A	B	C
1	4	6	2
2	3	7	1
3	2	7	1
4	5	5	1
5	4	5	3
6	4	5	4
7	3	8	3
8	3	9	3
9	3	9	2
10	1	6	2

Using Exercise 7.12,

a) test the independence between the 3 variables.
b) test whether $\mu^\top = [2\ 6\ 4]$ and compare this to the 3 univariate t-tests.
c) test whether $\mu_1 = \mu_2 = \mu_3$ using simple ANOVA and the χ^2 approximation.

Exercise 7.14 Consider an i.i.d. sample of size $n = 5$ from a bivariate normal distribution

$$X \sim N_2\left(\mu, \begin{pmatrix} 3 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

where ρ is a known parameter. Suppose $\bar{x}^\top = (1\ 0)$. For what value of ρ would the hypothesis $H_0 : \mu^\top = (0\ 0)$ be rejected in favour of $H_1 : \mu^\top \neq (0\ 0)$ (at the 5% level)?

Exercise 7.15 Using Example 7.14, test the last two cases described there and test the sample number one ($n_1 = 30$), to see if they are from a normal population with $\Sigma = 4\mathcal{I}_4$ (the sample covariance matrix to be used is given by S_1).

Exercise 7.16 Consider the bank data set. For the counterfeit bank notes, we want to know if the length of the diagonal (X_6) can be predicted by a linear model in X_1 to X_5. Estimate the linear model and test if the coefficients are significantly different from zero.

Exercise 7.17 In Example 7.10, can you predict the vocabulary score of the children in eleventh grade, by knowing the results from grades 8–9 and 10? Estimate a linear model and test its significance.

Exercise 7.18 Test the equality of the covariance matrices from the two groups in the WAIS subtest (Example 7.19).

Exercise 7.19 Prove expressions (7.21), (7.22) and (7.23).

Exercise 7.20 Using Theorem 6.3 and expression (7.16), construct an asymptotic rejection region of size α for testing, in a general model $f(x, \theta)$, with $\theta \in \mathbb{R}^k$, $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.

Exercise 7.21 Exercise 6.5 considered the pdf $f(x_1, x_2) = \frac{1}{\theta_1^2 \theta_2^2 x_2} e^{-(\frac{x_1}{\theta_1 x_2} + \frac{x_2}{\theta_1 \theta_2})}$, $x_1, x_2 > 0$. Solve the problem of testing $H_0 : \theta^\top = (\theta_{01}, \theta_{02})$ from an iid sample of size n on $x = (x_1, x_2)^\top$, where n is large.

Exercise 7.22 In Olkin and Veath (1980), the evolution of citrate concentrations in plasma is observed at 3 different times of day, X_1 (8 am), X_2 (11 am) and X_3 (3 pm), for two groups of patients who follow different diets. (The patients were randomly attributed to each group under a balanced design $n_1 = n_2 = 5$.) The data are:

Group	X_1 (8 am)	X_2 (11 am)	X_3 (3 pm)
I	125	137	121
	144	173	147
	105	119	125
	151	149	128
	137	139	109
II	93	121	107
	116	135	106
	109	83	100
	89	95	83
	116	128	100

Test if the profiles of the groups are parallel, if they are at the same level and if they are horizontal.