5 Inferences about a Mean Vector

In this chapter we use the results from Chapter 2 through Chapter 4 to develop techniques for analyzing data. A large part of any analysis is concerned with inference – that is, reaching valid conclusions concerning a population on the basis of information from a sample.

At this point, we shall concentrate on inferences about a population mean vector and its component parts. One of the central messages of multivariate analysis is that p correlated variables must be analyzed jointly.

Example (Bern-Chur-Zürich, p. 1-4)**.** In Keller (1921) we find values for the annual mean temperature, annual precipitation and annual sunshine duration for Bern, Chur and Zürich (see Table 5.1). Note that the locations of the weather stations, where these values have been measured, have changed in the meantime as the different altitudes indicate. We use the values of Keller (1921) for Bern as reference (population mean) and compare them with the time series (sample mean) for Bern from MeteoSchweiz.

Table 5.1: Temperature, precipitation and sunshine duration for Bern, Zürich and Chur. Source: Keller (1921).

Station	Temperature	Precipitation	Sunshine duration
Bern (572 m)	7.8 °C	922 mm	1781 h
Chur (600 m)	NA	803 mm	NA
Zürich (493 m)	8.5 °C	1147 mm	1671 h

Remark. A hypothesis is a conjecture about the value of a parameter, in this section a population mean or means. Hypothesis testing assists in making a decision under uncertainty.

5.1 Plausibility of $\boldsymbol{\mu}_0$ as a Value for a Normal Population Mean

Example (Climate Time Series, p. 1-3)**.** Assume that we know the population mean average winter temperatures $\boldsymbol{\mu}$ for the six Swiss stations Bern, Davos, Genf, Grosser St. Bernhard, Säntis and Sils Maria. We want to answer the question, whether the mean average temperatures $\boldsymbol{\mu}_0$ of the years 1950-2002 differ from $\boldsymbol{\mu}$. Figure 5.1 shows the scatterplot matrix of these six stations. We see, that there is one very cold year. Figure 5.2 shows the boxplots of the data set with and without the very cold winter 1962 as well as the population mean and the corresponding chi-square plots.

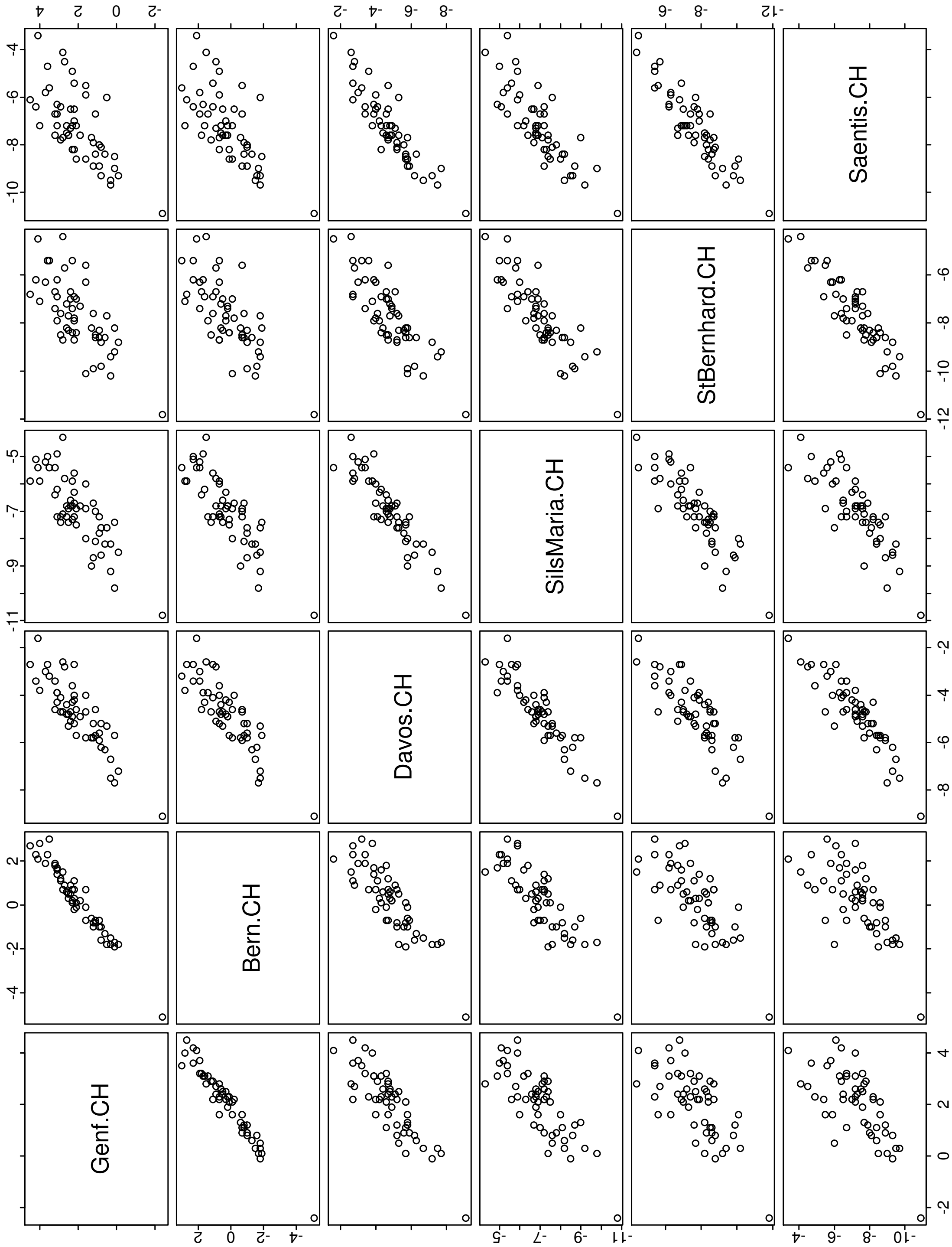

Figure 5.1: Scatterplot matrix of the mean average winter temperatures for six Swiss stations from 1950-2002. Data set: Climate Time Series, p. 1-3.

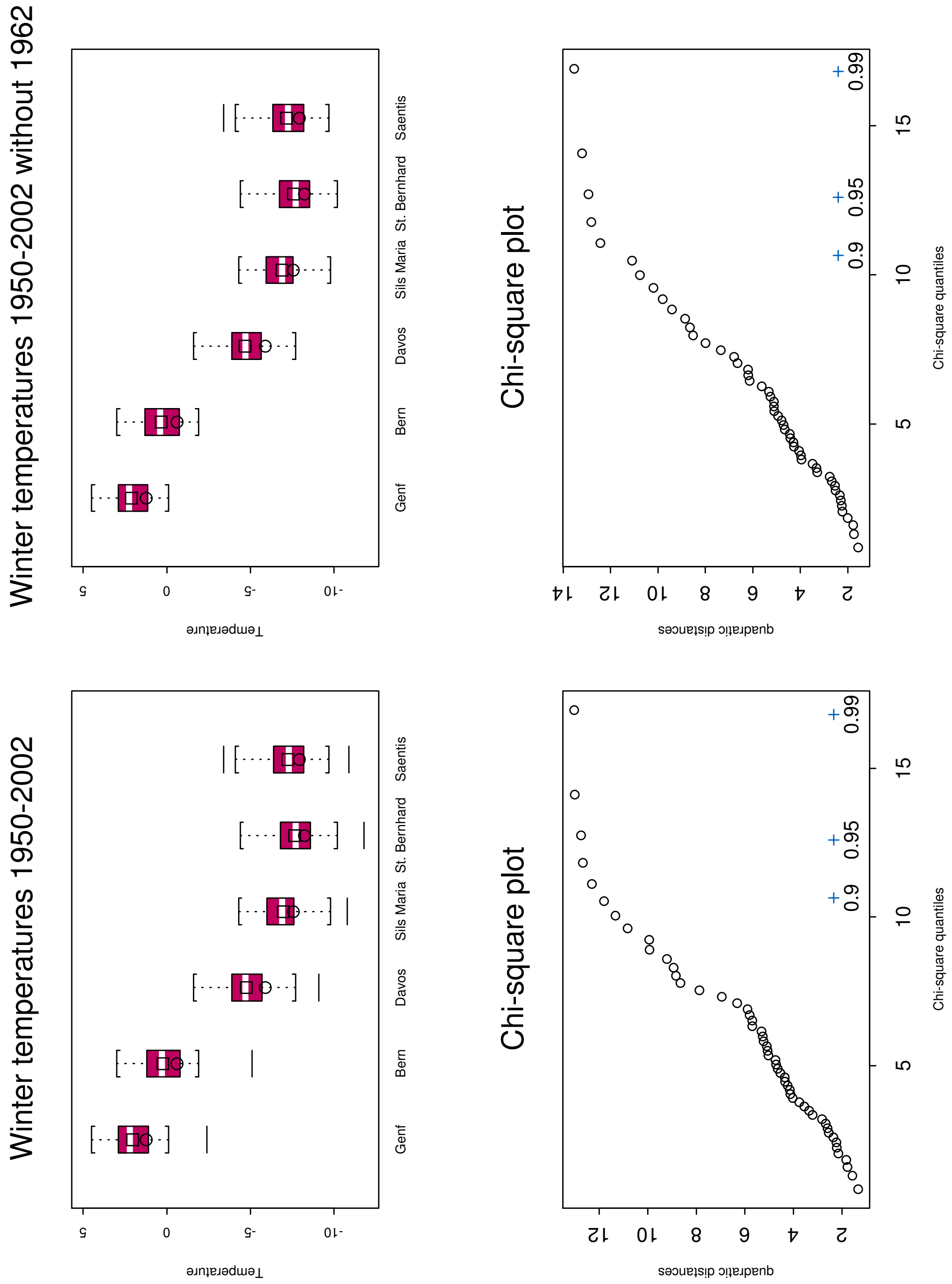

Figure 5.2: Boxplots and chi-square plots of the mean average winter temperature for the six Swiss stations from 1950-2002. The circles in the boxplots show the population mean $\boldsymbol{\mu}$, the squares the overall sample mean for the period 1950-2002 with and without 1962. Data set: Climate Time Series, p. 1-3.

5.1.1 Univariate case

Let us start with the univariate theory of determining whether a specific value μ_0 is a plausible value for the population mean μ. For the point of view of hypothesis testing, this problem can be formulated as a test of the competing hypotheses

$$H_0 : \mu = \mu_0 \text{ and } H_1 : \mu \neq \mu_0.$$

Here H_0 is the null hypothesis and H_1 is the two-sided alternative hypothesis. If $X_1, \ldots, X_n$ be a random sample from a normal population, the appropriate test statistic is

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}, \tag{5.1}$$

where $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$ and $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2$. This test statistic has a student's t-distribution with $n-1$ degrees of freedom (d.f.). We reject H_0, that μ_0 is a plausible value of μ, if the observed $|t|$ exceeds a specified percentage point of a t-distribution with $n-1$ d.f.

From (5.1) it follows that

$$t^2 = n(\overline{X} - \mu_0)(s^2)^{-1}(\overline{X} - \mu_0). \tag{5.2}$$

Reject H_0 in favor of H_1 at significance level α, if

$$n(\overline{x} - \mu_0)(s^2)^{-1}(\overline{x} - \mu_0) > t_{n-1}^2(\alpha/2), \tag{5.3}$$

where $t_{n-1}^2(\alpha/2)$ denotes the upper $100(\alpha/2)$th percentile of the t-distribution with $n-1$ d.f.

If H_0 is not rejected, we conclude that μ_0 is a plausible value for the normal population mean. From the correspondence between acceptance regions for tests of $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ and confidence intervals for μ, we have

$$\begin{aligned}
& \left|\frac{\overline{x} - \mu_0}{s/\sqrt{n}}\right| \leq t_{n-1}(\alpha/2) \\
\Longleftrightarrow \quad & \{\text{ do not reject } H_0 : \mu = \mu_0 \text{ at level } \alpha\} \\
\Longleftrightarrow \quad & \left\{ \mu_0 \text{ lies in the } (1-\alpha) \text{ confidence interval } \overline{x} \pm t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}} \right\} \\
\Longleftrightarrow \quad & \overline{x} - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}} \leq \mu_0 \leq \overline{x} + t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}.
\end{aligned}$$

Remark. Before the sample is selected, the $(1-\alpha)$ confidence interval is a random interval because the endpoints depend upon the random variables $\overline{X}$ and s. The probability that the interval contains μ is $1-\alpha$; among large numbers of such independent intervals, approximately $100(1-\alpha)\%$ of them will contain μ.

5.1.2 Multivariate Case

Example (Bern-Chur-Zürich, p. 1-4)**.** Keller (1921) cites the following climatological variables for Bern (572 m): Annual mean temperature 7.8 °C, annual precipitation 922 mm and annual sunshine duration 1781 h. Assume that these values form the reference vector (population mean vector). Now we compare this vector with the time series (sample mean vectors) of Bern given from MeteoSchweiz for the two time periods 1930-1960 and 1960-1990. We state in the null hypothesis that the sample mean vectors of temperature, precipitation, sunshine duration are the same as the reference vector of temperature, precipitation and sunshine duration.

Consider now the problem of determining whether a given $p \times 1$ vector $\boldsymbol{\mu}_0$ is a plausible value for the mean of a multivariate normal distribution. We have the hypotheses

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \text{ and } H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0.$$

A natural generalization of the squared distance in (5.2) is its multivariate analog

$$T^2 = (\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)' \left(\frac{1}{n} \mathbf{S} \right)^{-1} (\overline{\boldsymbol{X}} - \boldsymbol{\mu}_0)$$

where

$$\underset{(p\times1)}{\overline{\boldsymbol{X}}} = \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{X}_j$$

$$\underset{(p\times p)}{\mathbf{S}} = \frac{1}{n-1} \sum_{j=1}^{n} \underset{(p\times1)(1\times p)}{(\boldsymbol{X}_j - \overline{\boldsymbol{X}})(\boldsymbol{X}_j - \overline{\boldsymbol{X}})'}$$

$$\underset{(p\times1)}{\boldsymbol{\mu}_0} = (\mu_{10}, \ldots, \mu_{p0})'$$

The T^2-statistic is called Hotelling's T^2. We reject H_0, if the observed statistical distance T^2 is too large – that is, if $\overline{\boldsymbol{x}}$ is too far from $\boldsymbol{\mu}_0$. It can be shown that

$$T^2 \sim \frac{(n-1)p}{n-p} F_{p,n-p}, \tag{5.4}$$

where $F_{p,n-p}$ denotes a random variable with an F-distribution with p and $n-p$ d.f.

To summarize, we have the following proposition:

Proposition 5.1.1. *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Then*

$$\begin{aligned} \alpha &= P\left(T^2 > \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) \right) \\ &= P\left(n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\overline{\boldsymbol{X}} - \boldsymbol{\mu}) > \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) \right) \end{aligned} \tag{5.5}$$

whatever the true $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Here $F_{p,n-p}(\alpha)$ is the upper (100α)th percentile of the $F_{p,n-p}$ distribution.

Statement (5.5) leads immediately to a test of the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. At the α level of significance, we reject H_0 in favor of H_1 if the observed

$$T^2 = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) > \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha). \tag{5.6}$$

Remark. The T^2-statistic is invariant under changes in the units of measurements for $\boldsymbol{X}$ of the form

$$\underset{(p\times 1)}{\boldsymbol{Y}} = \underset{(p\times p)}{\mathbf{C}} \underset{(p\times 1)}{\boldsymbol{X}} + \underset{(p\times 1)}{\boldsymbol{d}}$$

with $\mathbf{C}$ nonsingular.

5.2 Confidence Regions and Simultaneous Comparisons of Component Means

To obtain our primary method for making inferences from a sample, we need to extend the concept of a univariate confidence interval to a multivariate confidence region. Let $\boldsymbol{\theta}$ be a vector of unknown population parameters and $\boldsymbol{\Theta}$ be the set of all possible values of $\boldsymbol{\theta}$. A confidence region is a region of likely $\boldsymbol{\theta}$ values. This region is determined by the data, and for the moment, we shall denote it by $R(\mathbf{X})$, where $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$ is the data matrix. The region $R(\mathbf{X})$ is said to be a $(1-\alpha)$ confidence region if, before the sample is selected,

$$P\left(R(\mathbf{X}) \text{ will cover the true } \boldsymbol{\theta}\right) = 1 - \alpha.$$

This probability is calculated under the true, but unknown, value of $\boldsymbol{\theta}$.

Proposition 5.2.1. *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population. Before the sample is selected, the confidence region for the mean $\boldsymbol{\mu}$ is given by*

$$P\left(n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \leq \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha)\right) = 1 - \alpha$$

whatever the values of the unknown $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

For a particular sample, $\overline{\boldsymbol{x}}$ and $\mathbf{S}$ can be computed, and we find a $(1-\alpha)$ confidence region for the mean of a p-dimensional normal distribution as the ellipsoid determined by all $\boldsymbol{\mu}$ such that

$$n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \leq \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) \tag{5.7}$$

where $\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{j=1}^n \boldsymbol{x}_j$, $\mathbf{S} = \frac{1}{n-1}\sum_{j=1}^n (\boldsymbol{x}_j - \overline{\boldsymbol{x}})(\boldsymbol{x}_j - \overline{\boldsymbol{x}})'$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are the sample observations.

Any $\boldsymbol{\mu}_0$ lies within the confidence region (is a plausible value for $\boldsymbol{\mu}$) if the distance

$$n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) \leq \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha). \tag{5.8}$$

Since this is analogous to testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, we see that the confidence region of (5.7) consists of all $\boldsymbol{\mu}_0$ vectors for which the T^2-test would not reject H_0 in favor of H_1 at significance level α.

Remark. We can calculate the axes of the confidence ellipsoid and their relative lengths. These are determined from the eigenvalues λ_i and eigenvectors $\boldsymbol{e}_i$ of $\mathbf{S}$. As in (4.1), the directions and lengths of the axes of

$$n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \leq c^2 = \frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) \tag{5.9}$$

are determined by going

$$\frac{\sqrt{\lambda_i}c}{\sqrt{n}} = \sqrt{\lambda_i}\sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)}$$

units along the eigenvectors $\boldsymbol{e}_i$. Beginning at the center $\overline{\boldsymbol{x}}$, the axes of the confidence ellipsoid are

$$\pm\sqrt{\lambda_i}\sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)}\, \boldsymbol{e}_i$$

where

$$\mathbf{S}\boldsymbol{e}_i = \lambda_i \boldsymbol{e}_i, \qquad i = 1, \ldots, p.$$

Example (Bern-Chur-Zürich, p. 1-4)**.** Figure 5.3 shows that for the period 1930-1960 the confidence regions with the sample mean (big circle) includes the population mean (large triangle). Therefore the null hypothesis can not be rejected and we conclude that the sample mean vector does not differ from the population mean vector.

Figure 5.4 corresponds to the time period 1960-1990 and here we see, that both 95% confidence regions do not include the population mean vectors (large triangles). So in this case the null hypothesis of equal means can be rejected. To see why the null hypothesis is rejected a univariate analysis can be done, comparing separately all components.

5.2.1 Simultaneous Confidence Statements

While the confidence region $n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \leq c^2$, for c a constant, correctly assesses the joint knowledge concerning plausible values of $\boldsymbol{\mu}$, any summary of conclusions ordinarily includes confidence statements about the individual component means. In so doing, we adopt the attitude that all of the separate confidence statements should hold simultaneously with a specified high probability. It is the guarantee of a specified

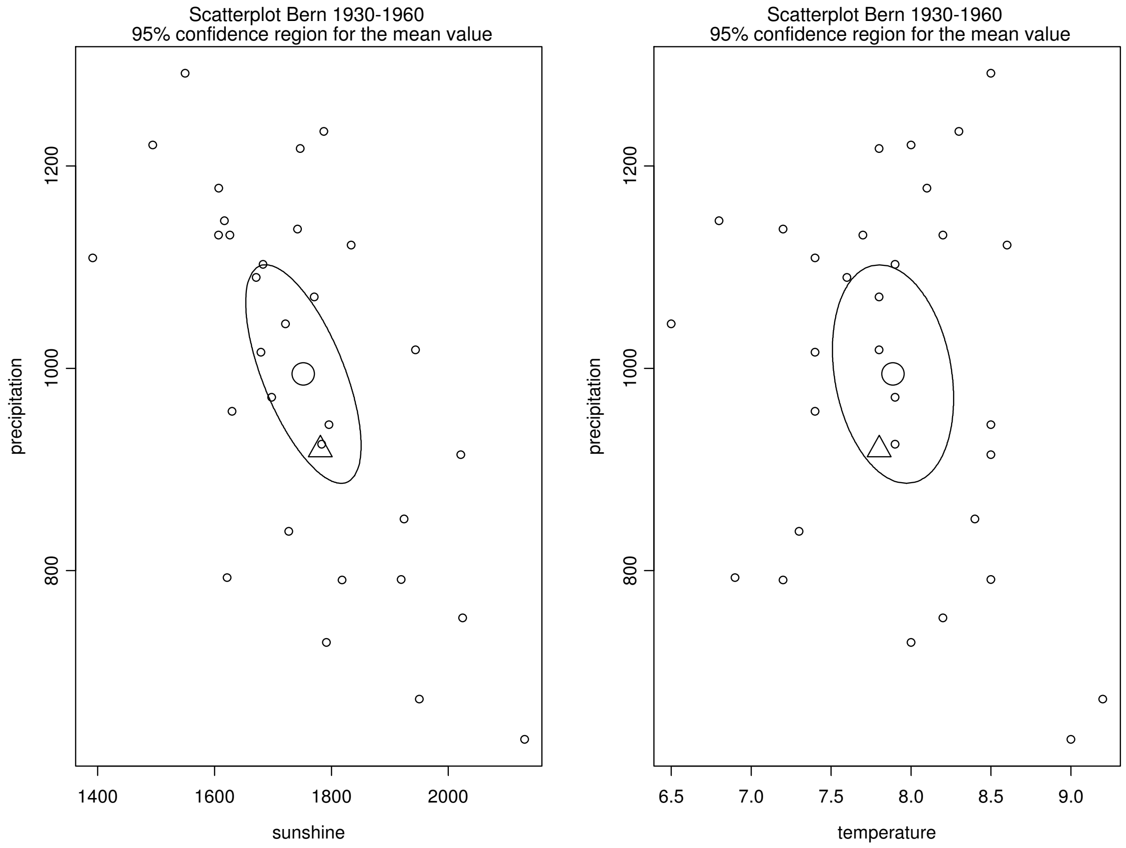

Figure 5.3: Confidence regions for the pairs sunshine-precipitation and temperature-precipitation with the corresponding population mean (large triangle). Data set: Bern-Chur-Zürich, p. 1-4.

probability against any statement being incorrect that motivates the term simultaneous confidence intervals.

We begin by considering simultaneous confidence statements which are related to the joint confidence region based on the T^2-statistic. Let $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and form the linear combination

$$Z = a_1 X_1 + \ldots + a_p X_p = \boldsymbol{a}'\boldsymbol{X}.$$

Then

$$\begin{aligned} \mu_Z &= \mathrm{E}(Z) = \boldsymbol{a}'\boldsymbol{\mu}, \\ \sigma_Z^2 &= \mathrm{Var}(Z) = \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} \quad \text{and} \\ Z &\sim N(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}). \end{aligned}$$

If a random sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from the $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population is available, a corresponding sample of Z's can be created by taking linear combinations. Thus

$$Z_j = \boldsymbol{a}'\boldsymbol{X}_j, \qquad j = 1, \ldots, n.$$

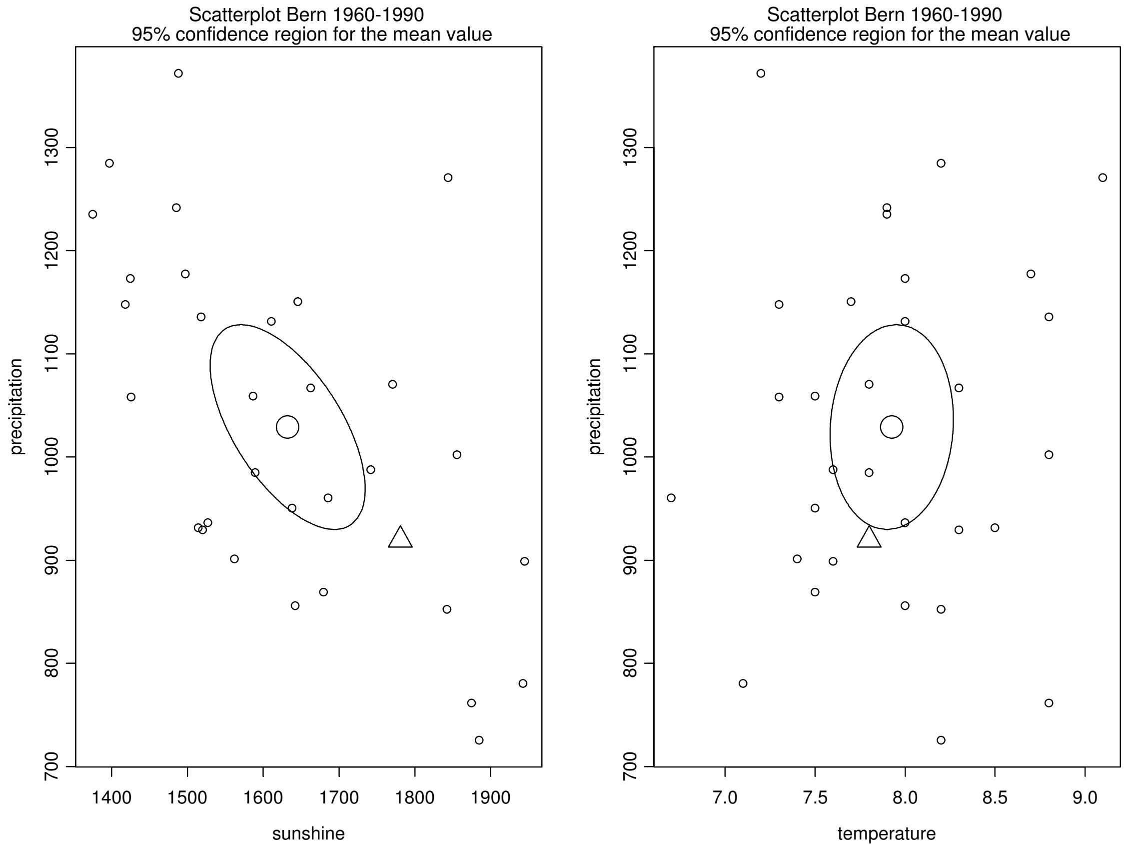

Figure 5.4: Confidence regions for the pairs sunshine-precipitation and temperature-precipitation with the corresponding population mean (large triangle). Data set: Bern-Chur-Zürich, p. 1-4.

The sample mean and variance of the observed values $z_1, \ldots, z_n$ are $\overline{z} = \boldsymbol{a}'\overline{\boldsymbol{x}}$ and $s_z^2 = \boldsymbol{a}'\mathbf{S}\boldsymbol{a}$, where $\overline{\boldsymbol{x}}$ and $\mathbf{S}$ are the sample mean vector and the covariance matrix of the $\boldsymbol{x}_j$'s, respectively.

Case i: $\boldsymbol{a}$ fixed

For $\boldsymbol{a}$ fixed and σ_z^2 unknown, a $(1-\alpha)$ confidence interval for $\mu_z = \boldsymbol{a}'\boldsymbol{\mu}$ is based on student's t-ratio

$$t = \frac{\overline{z} - \mu_Z}{s_z/\sqrt{n}} = \frac{\sqrt{n}(\boldsymbol{a}'\overline{\boldsymbol{x}} - \boldsymbol{a}'\boldsymbol{\mu})}{\sqrt{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}}}$$

and leads to the statement

$$\boldsymbol{a}'\overline{\boldsymbol{x}} - t_{n-1}(\alpha/2)\frac{\sqrt{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}}}{\sqrt{n}} \leq \boldsymbol{a}'\boldsymbol{\mu} \leq \boldsymbol{a}'\overline{\boldsymbol{x}} + t_{n-1}(\alpha/2)\frac{\sqrt{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}}}{\sqrt{n}}, \qquad (5.10)$$

where $t_{n-1}(\alpha/2)$ is the upper $100(\alpha/2)$th percentile of a t-distribution with $n-1$ d.f.

Example. For $\boldsymbol{a}' = (1, 0, \ldots, 0)$, $\boldsymbol{a}'\boldsymbol{\mu} = \mu_1$, and (5.10) becomes the usual confidence interval for a normal population mean. Note, in this case, $\boldsymbol{a}'\mathbf{S}\boldsymbol{a} = s_{11}$.

Remark. Of course we could make several confidence statements about the components of $\boldsymbol{\mu}$, each with associated confidence coefficient $1 - \alpha$, by choosing different coefficient vectors $\boldsymbol{a}$. However, the confidence associated with all of the statements taken together is not $1 - \alpha$.

Remark. Intuitively, it would be desirable to associate a "collective" confidence coefficient of $1 - \alpha$ with the confidence intervals that can be generated by all choices of $\boldsymbol{a}$. However, a price must be paid for the convenience of a large simultaneous confidence coefficient: intervals that are wider (less precise) than the interval of (5.10) for a specific choice of $\boldsymbol{a}$.

Case ii: $\boldsymbol{a}$ arbitrary

Given a data set $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ and a particular $\boldsymbol{a}$, the confidence interval in (5.10) is that set of $\boldsymbol{a}'\boldsymbol{\mu}$ values for which

$$t^2 = \frac{n(\boldsymbol{a}'(\overline{\boldsymbol{x}} - \boldsymbol{\mu}))^2}{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}} \leq t^2_{n-1}(\alpha/2). \tag{5.11}$$

A simultaneous confidence region is given by the set of $\boldsymbol{a}'\boldsymbol{\mu}$ values such that t^2 is relatively small for all choices of $\boldsymbol{a}$. It seems reasonable to expect that the constant $t^2_{n-1}(\alpha/2)$ in (5.11) will be replaced by a larger value, c^2, when statements are developed for many choices of $\boldsymbol{a}$.

Considering the values of $\boldsymbol{a}$ for which $t^2 \leq c^2$, we are naturally led to the determination of

$$\max_{\boldsymbol{a}} t^2 = \max_{\boldsymbol{a}} \frac{n(\boldsymbol{a}'(\overline{\boldsymbol{x}} - \boldsymbol{\mu}))^2}{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}}.$$

Using the maximization lemma (Johnson and Wichern (2007), p. 80) we get

$$\max_{\boldsymbol{a}} \frac{n(\boldsymbol{a}'(\overline{\boldsymbol{x}} - \boldsymbol{\mu}))^2}{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}} = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) = T^2 \tag{5.12}$$

with the maximum occurring for $\boldsymbol{a}$ proportional to $\mathbf{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu})$.

Example. Let $\boldsymbol{\mu}' = (0, 0)$, $\overline{\boldsymbol{x}}' = (1, 2)$ and $\mathbf{S} = \begin{pmatrix} 1 & 2 \\ 2 & 100 \end{pmatrix}$. Then

$$\frac{n(\boldsymbol{a}'\overline{\boldsymbol{x}})^2}{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}} = n\frac{(a_1 + 2a_2)^2}{(a_1 + 2a_2)^2 + 96a_2^2}$$

has its maximum at $\boldsymbol{a}' = (c, 0)$ with $c \neq 0$, which is proportional to $\mathbf{S}^{-1}\overline{\boldsymbol{x}} = (1, 0)'$.

Proposition 5.2.2. *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from an $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population with $\boldsymbol{\Sigma}$ positive definite. Then, simultaneously for all $\boldsymbol{a}$, the interval*

$$\left(\boldsymbol{a}'\overline{\boldsymbol{X}} - \sqrt{\frac{(n-1)p}{n(n-p)}F_{p,n-p}(\alpha)\boldsymbol{a}'\mathbf{S}\boldsymbol{a}}, \quad \boldsymbol{a}'\overline{\boldsymbol{X}} + \sqrt{\frac{(n-1)p}{n(n-p)}F_{p,n-p}(\alpha)\boldsymbol{a}'\mathbf{S}\boldsymbol{a}}\right) \tag{5.13}$$

will contain $\boldsymbol{a}'\boldsymbol{\mu}$ with probability $1-\alpha$.

Proof. From (5.12)

$$T^2 = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \leq c^2 \text{ implies } \frac{n(\boldsymbol{a}'(\overline{\boldsymbol{x}} - \boldsymbol{\mu}))^2}{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}} \leq c^2$$

for every $\boldsymbol{a}$, or

$$\boldsymbol{a}'\overline{\boldsymbol{x}} - c\sqrt{\frac{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}}{n}} \leq \boldsymbol{a}'\boldsymbol{\mu} \leq \boldsymbol{a}'\overline{\boldsymbol{x}} + c\sqrt{\frac{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}}{n}}$$

for every $\boldsymbol{a}$. Choosing $c^2 = \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$ (compare equation (5.4)) gives intervals that will contain $\boldsymbol{a}'\boldsymbol{\mu}$ for all $\boldsymbol{a}$, with probability $1-\alpha = P(T^2 \leq c^2)$. □

It is convenient to refer to the simultaneous intervals of (5.13) as T^2-intervals, since the coverage probability is determined by the distribution of T^2. The successive choices $\boldsymbol{a}' = (1, 0, \ldots, 0)$, $\boldsymbol{a}' = (0, 1, \ldots, 0)$, and so on through $\boldsymbol{a}' = (0, 0, \ldots, 1)$ for the T^2-intervals allow us to conclude that

$$\begin{gathered}
\overline{x}_1 - \sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}} \leq \mu_1 \leq \overline{x}_1 + \sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}} \\
\overline{x}_2 - \sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}} \leq \mu_2 \leq \overline{x}_2 + \sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}} \\
\vdots \\
\overline{x}_p - \sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{pp}}{n}} \leq \mu_p \leq \overline{x}_p + \sqrt{\frac{(n-1)p}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{pp}}{n}}
\end{gathered}$$

all hold simultaneously with confidence coefficient $1-\alpha$.

Example (Climate Time Series, p. 1-3)**.** Figure 5.5 shows the 95% confidence ellipse and the simultaneous T^2-intervals for the component means of the mean average winter temperatures for Bern and Davos.

Remark. Note that, without modifying the coefficient $1-\alpha$, we can make statements about the differences $\mu_i - \mu_k$ corresponding to $\boldsymbol{a}' = (0, \ldots, 0, a_i, 0, \ldots, 0, a_k, 0, \ldots, 0)$, where $a_i = 1$ and $a_k = -1$. In this case $\boldsymbol{a}'\mathbf{S}\boldsymbol{a} = s_{ii} - 2s_{ik} + s_{kk}$.

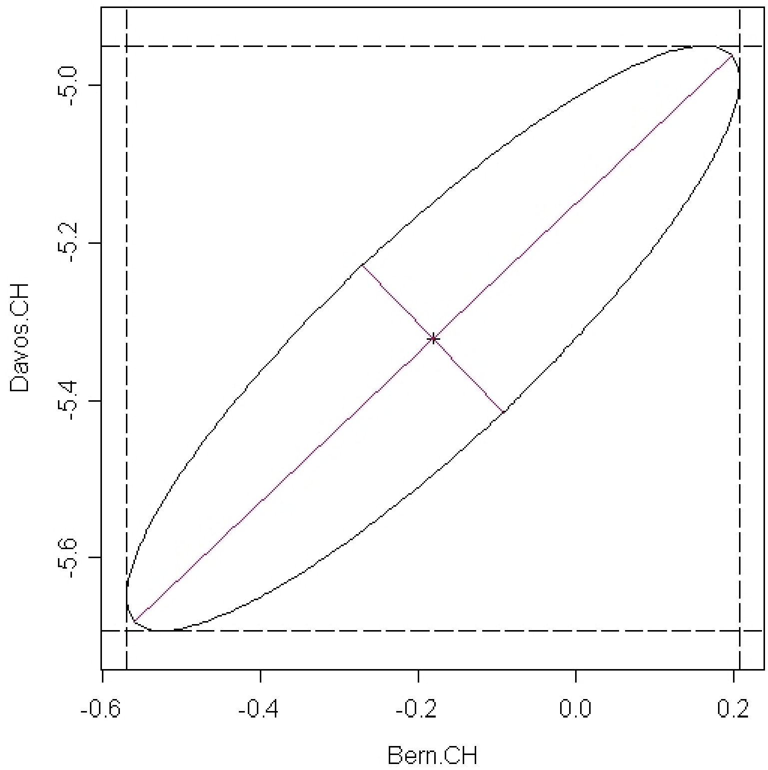

Figure 5.5: 95% confidence ellipse and the simultaneous T^2-intervals for the component means as projections of the confidence ellipse on the axes. Data set: Climate Time Series, p. 1-3.

Remark. The simultaneous T^2 confidence intervals are ideal for "data snooping". The confidence coefficient $1-\alpha$ remains unchanged for any choice of $\boldsymbol{a}$, so linear combinations of the components μ_i that merit inspection based upon an examination of the data can be estimated.

Remark. The simultaneous T^2 confidence intervals for the individual components of a mean vector are just the projections of the confidence ellipsoid on the component axes.

5.2.2 Comparison of Simultaneous Confidence Intervals with One-at-a-time Intervals

An alternative approach to the construction of confidence intervals is to consider the components μ_i one at a time, as suggested by (5.10) with $\boldsymbol{a}' = (0, \ldots, 0, a_i, 0, \ldots, 0)$ where $a_i = 1$. This approach ignores the covariance structure of the p variables and leads to the intervals

$$\overline{x}_1 - t_{n-1}(\alpha/2)\sqrt{\frac{s_{11}}{n}} \leq \mu_1 \leq \overline{x}_1 + t_{n-1}(\alpha/2)\sqrt{\frac{s_{11}}{n}}$$
$$\vdots$$
$$\overline{x}_p - t_{n-1}(\alpha/2)\sqrt{\frac{s_{pp}}{n}} \leq \mu_p \leq \overline{x}_p + t_{n-1}(\alpha/2)\sqrt{\frac{s_{pp}}{n}}.$$

Example (Climate Time Series, p. 1-3). Figure 5.6 shows the 95% confidence ellipse and the 95% one-at-a-time intervals of the mean average winter temperatures for Bern

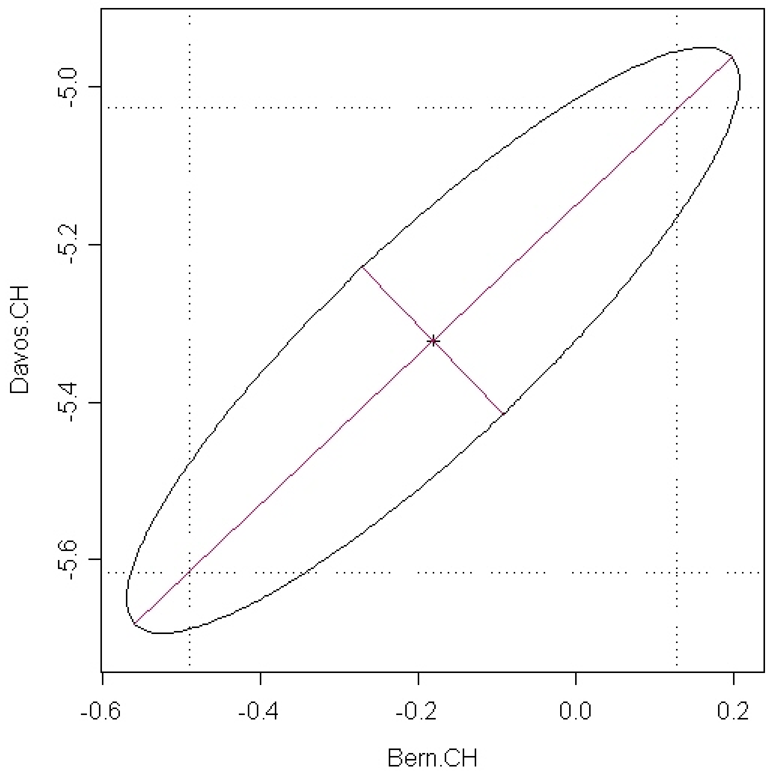

Figure 5.6: 95% confidence ellipse and the 95% one-at-a-time intervals. Data set: Climate Time Series, p. 1-3.

Although prior to sampling, the ith interval has probability $1-\alpha$ of covering μ_i, we do not know what to assert, in general, about the probability of all intervals containing their respective μ_i's. As we have pointed out, this probability is not $1-\alpha$. To guarantee a probability of $1-\alpha$ that all of the statements about the component means hold simultaneously, the individual intervals must be wider than the separate t-intervals; just how much wider depends on both p and n, as well as on $1-\alpha$.

Example. For $1-\alpha = 0.95$, $n = 15$, and $p = 4$, the multipliers of $\sqrt{s_{ii}/n}$ are

$$\sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(0.05)} = \sqrt{\frac{56}{11} 3.36} = 4.14$$

and $t_{n-1}(0.025) = 2.145$, respectively. Consequently, the simultaneous intervals are $100(4.14-2.145)/2.145 = 93\%$ wider than those derived from the one-at-a-time t method.

The T^2-intervals are too wide if they are applied to the p component means. To see why, consider the confidence ellipse and the simultaneous intervals shown in Figure 5.5. If μ_1 lies in its T^2-interval and μ_2 lies in its T^2-interval, then (μ_1, μ_2) lies in the rectangle formed by these two intervals. This rectangle contains the confidence ellipse and more. The confidence ellipse is smaller but has probability 0.95 of covering the mean vector $\boldsymbol{\mu}$ with its component means μ_1 and μ_2. Consequently, the probability of covering the two individual means μ_1 and μ_2 will be larger than 0.95 for the rectangle formed by the T^2-intervals. This result leads us to consider a second approach to making multiple comparisons known as the Bonferroni method.

Bonferroni Method of Multiple Comparisons

Often, attention is restricted to a small number of individual confidence statements. In these situations it is possible to do better than the simultaneous intervals of (5.13). If the number m of specified component means μ_i or linear combinations $\boldsymbol{a}'\boldsymbol{\mu} = a_1\mu_1 + \ldots + a_p\mu_p$ is small, simultaneous confidence intervals can be developed that are shorter (more precise) than the simultaneous T^2-intervals. The alternative method for multiple comparisons is called the Bonferroni method.

Suppose that confidence statements about m linear combinations $\boldsymbol{a}_1'\boldsymbol{\mu}, \ldots, \boldsymbol{a}_m'\boldsymbol{\mu}$ are required. Let C_i denote a confidence statement about the value of $\boldsymbol{a}_i'\boldsymbol{\mu}$ with

$$P(C_i \text{ true}) = 1 - \alpha_i, \quad i = 1, \ldots, m.$$

Then

$$\begin{aligned} P(\text{all } C_i \text{ true}) &= 1 - P(\text{at least one } C_i \text{ false}) \\ &\geq 1 - \sum_{i=1}^{m} P(C_i \text{ false}) = 1 - \sum_{i=1}^{m} (1 - P(C_i \text{ true})) \\ &= 1 - (\alpha_1 + \ldots + \alpha_m). \end{aligned} \tag{5.14}$$

Inequality (5.14) allows an investigator to control the overall error rate $\alpha_1 + \ldots + \alpha_m$, regardless of the correlation structure behind the confidence statements.

Let us develop simultaneous interval estimates for the restricted set consisting of the components μ_i of $\boldsymbol{\mu}$. Lacking information on the relative importance of these components, we consider the individual t-intervals

$$\overline{x}_i \pm t_{n-1}\left(\frac{\alpha_i}{2}\right)\sqrt{\frac{s_{ii}}{n}}, \qquad i = 1, \ldots, m$$

with $\alpha_i = \alpha/m$. Since

$$P\left(\overline{X}_i \pm t_{n-1}\left(\frac{\alpha}{2m}\right)\sqrt{\frac{s_{ii}}{n}} \text{ contains } \mu_i\right) = 1 - \frac{\alpha}{m}, \qquad i = 1, \ldots, m,$$

we have, from (5.14),

$$\begin{aligned} P\left(\overline{X}_i \pm t_{n-1}\left(\frac{\alpha}{2m}\right)\sqrt{\frac{s_{ii}}{n}} \text{ contains } \mu_i, \text{ all } i\right) &\geq 1 - \underbrace{(\alpha/m + \cdots + \alpha/m)}_{m \text{ terms}} \\ &= 1 - \alpha. \end{aligned}$$

Therefore, with an overall confidence level greater than or equal to $1 - \alpha$, we can make the following $m = p$ statements:

$$\begin{gathered} \overline{x}_1 - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{11}}{n}} \leq \mu_1 \leq \overline{x}_1 + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{11}}{n}} \\ \vdots \\ \overline{x}_p - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{pp}}{n}} \leq \mu_p \leq \overline{x}_p + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{pp}}{n}}. \end{gathered}$$

Example (Climate Series Europe, p. 1-3)**.** Figure 5.7 shows the 95% confidence ellipse as well as the 95% simultaneous T^2-intervals, one-at-a-time intervals and Bonferroni simultaneous intervals for the mean average winter temperatures for Bern and Davos.

5.3 Large Sample Inference about a Population Mean Vector

When the sample size is large, tests of hypotheses and confidence regions for $\boldsymbol{\mu}$ can be constructed without the assumption of a normal population. All large-sample inferences about $\boldsymbol{\mu}$ are based on a χ^2-distribution. From (4.3), we know that

$$n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu})$$

is approximately χ^2 with p d.f. and thus,

$$P\left(n(\overline{\boldsymbol{X}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\boldsymbol{X}} - \boldsymbol{\mu}) \leq \chi^2_p(\alpha)\right) = 1 - \alpha$$

where $\chi^2_p(\alpha)$ is the upper (100α)th percentile of the χ^2_p-distribution.

Proposition 5.3.1. *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from a population with mean $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. When $n - p$ is large, the hypothesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ is rejected in favor of $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, at a level of significance approximately α, if the observed*

$$n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) > \chi^2_p(\alpha). \tag{5.15}$$

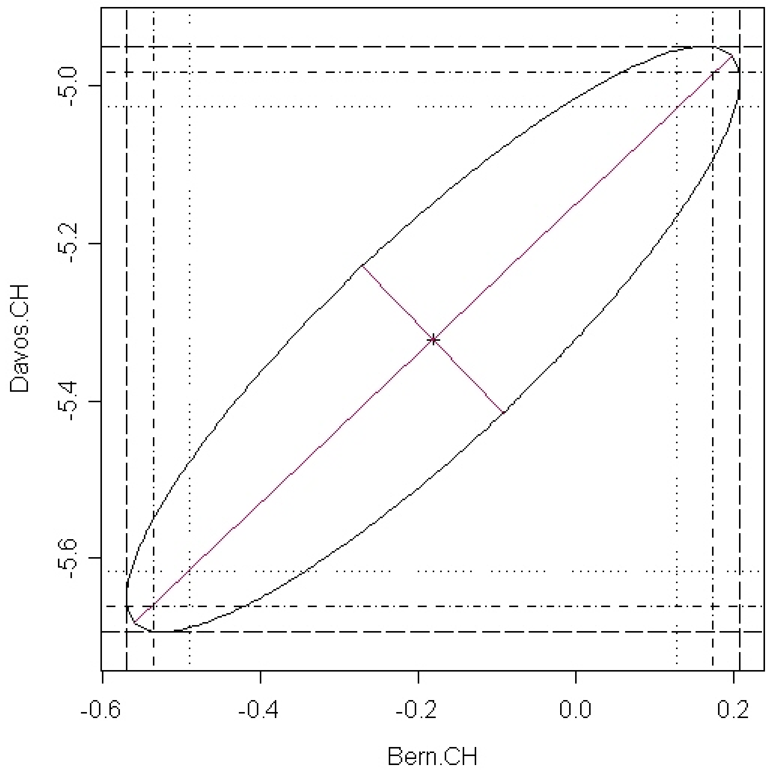

Figure 5.7: 95% confidence ellipse and different confidence intervals (T^2: dashed, one-at-a-time: dotted, Bonferroni: combined). Data set: Climate Time Series, p. 1-3.

Remark. Comparing the test in Proposition 5.3.1 with the corresponding normal theory test in (5.6), we see that the test statistics have the same structure, but the critical values are different. A closer examination reveals that both tests yield essentially the same result in situations where the χ^2-test of Proposition 5.3.1 is appropriate. This follows directly from the fact that $\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$ and $\chi^2_p(\alpha)$ are approximately equal for n large relative to p.

Proposition 5.3.2. *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be a random sample from a population with mean $\boldsymbol{\mu}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$. If $n-p$ is large,*

$$\boldsymbol{a}'\overline{\boldsymbol{X}} \pm \sqrt{\chi^2_p(\alpha)}\sqrt{\frac{\boldsymbol{a}'\mathbf{S}\boldsymbol{a}}{n}}$$

will contain $\boldsymbol{a}'\boldsymbol{\mu}$, for every $\boldsymbol{a}$, with probability approximately $1-\alpha$. Consequently, we *can make the* $(1-\alpha)$ *simultaneous confidence statements*

$$\bar{x}_1 \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{s_{11}}{n}} \quad \textit{contains } \mu_1$$
$$\vdots$$
$$\bar{x}_p \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{s_{pp}}{n}} \quad \textit{contains } \mu_p$$

and, in addition, for all pairs (μ_i, μ_k)*,* $i, k = 1, \ldots, p$*, the sample mean-centered ellipses*

$$n(\bar{x}_i - \mu_i, \bar{x}_k - \mu_k) \begin{pmatrix} s_{ii} & s_{ik} \\ s_{ik} & s_{kk} \end{pmatrix}^{-1} \begin{pmatrix} \bar{x}_i - \mu_i \\ \bar{x}_k - \mu_k \end{pmatrix} \leq \chi_p^2(\alpha) \ \textit{contain } (\mu_i, \mu_k).$$

6 Comparisons of Several Multivariate Means

The ideas developed in Chapter 5 can be extended to handle problems involving the comparison of several mean vectors. The theory is a little more complicated and rests on an assumption of multivariate normal distribution or large sample sizes.

We will first consider pairs of mean vectors, and then discuss several comparisons among mean vectors arranged according to treatment levels. The corresponding test statistics depend upon a partitioning of the total variation into pieces of variation attributable to the treatment sources and error. This partitioning is known as the multivariate analysis of variance (MANOVA).

1. Paired comparisons: Comparing measurements before the treatment with those after the treatment.

2. Repeated measures design for comparing treatments: q treatments are compared with respect to a single response variable. Each subject or experimental unit receives each treatment once over successive periods of time.

3. Comparing mean vectors from two population: Consider a random sample of size n_1 from population 1 and a random sample of size n_2 from population 2.

   For instance, we shall want to answer the question whether $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. Also, if $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \neq \mathbf{0}$, which component means are different?

4. Comparing several multivariate population means.

6.1 Paired Comparisons

Measurements are often recorded under different sets of experimental conditions to see whether the responses differ significantly over these sets. For example the efficacy of a new campaign may be determined by comparing measurements before the "treatment" with those after the treatment. In other situations, two or more treatments can be administered to the same or similar experimental units, and responses can be compared to assess the effects of the treatments.

One rational approach to comparing two treatments, is to assign both treatments to the same units. The paired responses may then be analyzed by computing their differences, thereby eliminating much of the influence of extraneous unit-to-unit variation.

6.1.1 Univariate Case

In the univariate (single response) case, let X_{j1} denote the response to treatment 1 (or the response before treatment), and let X_{j2} denote the response to treatment 2 (or the response after treatment) for the jth trial. That is, (X_{j1}, X_{j2}) are measurements recorded on the jth unit or jth pair of like units. The n differences

$$D_j := X_{j1} - X_{j2}, \quad j = 1, \ldots, n$$

should reflect only the differential effects of the treatments. Given that the differences D_j represent independent observations from an $N(\delta, \sigma_d^2)$ distribution, the variable

$$t := \frac{\overline{D} - \delta}{s_d/\sqrt{n}}$$

where

$$\overline{D} = \frac{1}{n}\sum_{j=1}^{n} D_j \quad \text{and} \quad s_d^2 = \frac{1}{n-1}\sum_{j=1}^{n}(D_j - \overline{D})^2$$

has a t-distribution with $n-1$ d.f. Then a $(1-\alpha)$ confidence interval for the mean difference $\delta = \mathrm{E}(X_{j1} - X_{j2})$ is given by the statement

$$\overline{d} - t_{n-1}\left(\frac{\alpha}{2}\right)\frac{s_d}{\sqrt{n}} \leq \delta \leq \overline{d} + t_{n-1}\left(\frac{\alpha}{2}\right)\frac{s_d}{\sqrt{n}}.$$

Remark. When uncertainty about the assumption of normality exists, a nonparametric alternative to ANOVA called the Kruskal-Wallis test is available. Instead of using observed values the Kruskal-Wallis procedure uses ranks and then compares the ranks among the treatment groups.

6.1.2 Multivariate Case

Example (Bern-Chur-Zürich, p. 1-4)**.** Consider the mean vectors for Bern and Zürich of the four variables pressure, temperature, precipitation and sunshine duration. In the paired case we take the differences of the annual values and the null hypothesis states, that the difference vector is the zero vector (see Figure 6.1). It seems to be reasonable to assume that all differences are statistically different from zero. This can be confirmed with Hotelling's T^2 test for paired samples.

Additional notation is required for the multivariate extension of the paired-comparison procedure. It is necessary to distinguish between p responses, two treatments, and n experimental units. We label the p responses within the jth unit as

$$\begin{aligned}
X_{1j1} &= \text{variable 1 under treatment 1}\\
X_{1j2} &= \text{variable 2 under treatment 1}\\
&\vdots\\
X_{1jp} &= \text{variable } p \text{ under treatment 1}\\
X_{2j1} &= \text{variable 1 under treatment 2}\\
&\vdots\\
X_{2jp} &= \text{variable } p \text{ under treatment 2}
\end{aligned}$$

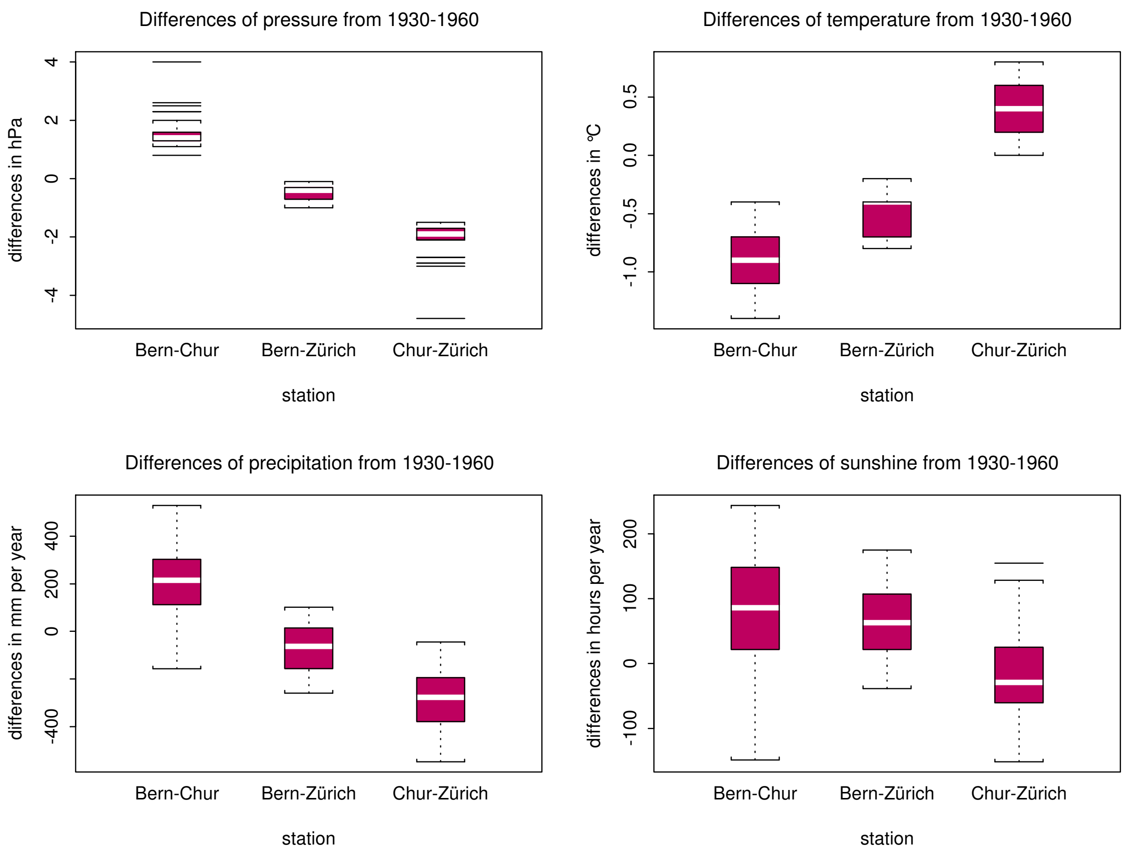

Figure 6.1: Boxplots of the paired differences of the four variables for Bern, Chur and Zürich for the time period 1930-1960. Data set: Bern-Chur-Zürich, p. 1-4.

and the p paired-difference random variables become

$$\begin{aligned} D_{j1} &= X_{1j1} - X_{2j1} \\ &\vdots \\ D_{jp} &= X_{1jp} - X_{2jp}. \end{aligned}$$

Let $\boldsymbol{D}_j' = (D_{j1}, \ldots, D_{jp})$, and assume, for $j = 1, \ldots, n$, that

$$\mathrm{E}(\boldsymbol{D}_j) = \boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)' \text{ and } \mathrm{Cov}(\boldsymbol{D}_j) = \boldsymbol{\Sigma}_d.$$

Proposition 6.1.1. *Let the differences $\boldsymbol{D}_1, \ldots, \boldsymbol{D}_n$ be a random sample from an $N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma}_d)$ population. Then*

$$T^2 = n(\overline{\boldsymbol{D}} - \boldsymbol{\delta})'\mathbf{S}_d^{-1}(\overline{\boldsymbol{D}} - \boldsymbol{\delta}) \sim \frac{p(n-1)}{n-p} F_{p,n-p},$$

where

$$\overline{\boldsymbol{D}} = \frac{1}{n}\sum_{j=1}^{n} \boldsymbol{D}_j \text{ and } \mathbf{S}_d = \frac{1}{n-1}\sum_{j=1}^{n}(\boldsymbol{D}_j - \overline{\boldsymbol{D}})(\boldsymbol{D}_j - \overline{\boldsymbol{D}})'. \qquad (6.1)$$

If n and $n-p$ are both large, T^2 is approximately distributed as a χ^2_p random variable, regardless of the form of the underlying population of differences.

The condition $\boldsymbol{\delta} = \mathbf{0}$ is equivalent to "no average difference between the two treatments." For the ith variable, $\delta_i > 0$ implies that treatment 1 is larger, on average, than treatment 2.

Proposition 6.1.2. *Given the observed differences $\boldsymbol{d}'_j = (d_{j1}, \ldots, d_{jp})$, $j = 1, , \ldots, n$, corresponding to the random variables $D_{j1}, \ldots, D_{jp}$, an α-level test of $H_0 : \boldsymbol{\delta} = \mathbf{0}$ versus $H_1 : \boldsymbol{\delta} \neq \mathbf{0}$ for an $N_p(\boldsymbol{\delta}, \boldsymbol{\Sigma}_d)$ population rejects H_0 if the observed*

$$T^2 = n\overline{\boldsymbol{d}}'\mathbf{S}_d^{-1}\overline{\boldsymbol{d}} > \frac{p(n-1)}{n-p} F_{p,n-p}(\alpha).$$

Here $\overline{\boldsymbol{d}}$ and $\mathbf{S}_d$ are given by (6.1).

A $(1-\alpha)$ confidence region for $\boldsymbol{\delta}$ consists of all $\boldsymbol{\delta}$ such that

$$(\overline{\boldsymbol{d}} - \boldsymbol{\delta})'\mathbf{S}_d^{-1}(\overline{\boldsymbol{d}} - \boldsymbol{\delta}) \leq \frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha).$$

Also, $(1-\alpha)$ simultaneous confidence intervals for the individual mean differences δ_i are given by

$$\delta_i : \quad \overline{d}_i \pm \sqrt{\frac{p(n-1)}{n-p} F_{p,n-p}(\alpha)} \sqrt{\frac{s^2_{d_i}}{n}}$$

where $\overline{d}_i$ is the ith element of $\overline{\boldsymbol{d}}$ and $s^2_{d_i}$ is the ith diagonal element of $\mathbf{S}_d$.

For $(n-p)$ large

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

and normality need not be assumed.

The Bonferroni $(1-\alpha)$ simultaneous confidence intervals for the individual mean differences are

$$\delta_i : \quad \overline{d}_i \pm t_{n-1}\left(\frac{\alpha}{2p}\right) \sqrt{\frac{s^2_{d_i}}{n}}.$$

6.2 Comparing Mean Vectors from Two Populations

Example (Bern-Chur-Zürich, p. 1-4)**.** Consider the mean vectors for Bern and Zürich of the four variables pressure, temperature, precipitation and sunshine duration. In the two populations case (unpaired case) the null hypothesis states, that the mean vectors are the same (see Figure 6.2). Calculating Hotelling's T^2 for unpaired observations shows that the null hypothesis can be rejected. Considering the univariate tests we see that only the mean temperatures differ whereas for pressure, precipitation and sunshine duration no differences of the means can be found.

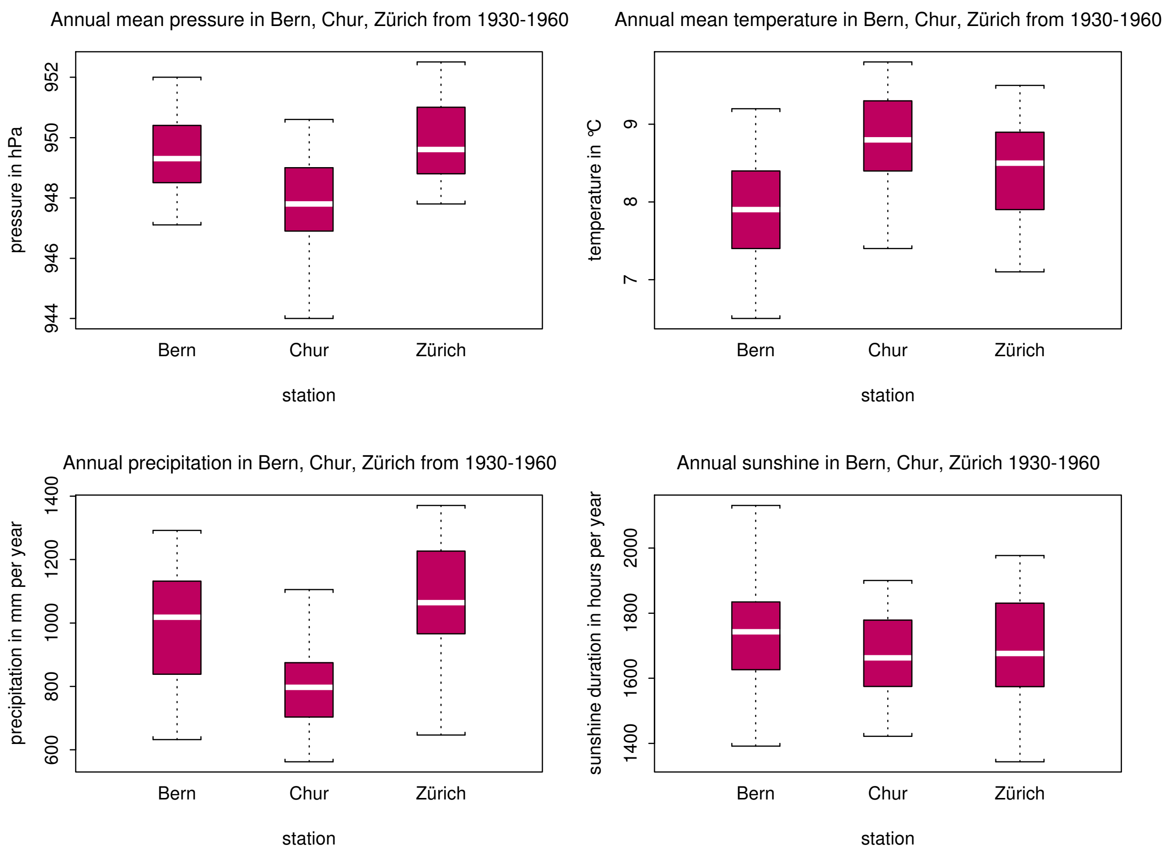

Figure 6.2: Boxplots of the four variables for Bern, Chur and Zürich for the time period 1930-1960. Data set: Bern-Chur-Zürich, p. 1-4.

A T^2-statistic for testing the equality of vector means from two multivariate populations can be developed by analogy with the univariate procedure. This T^2-statistic is appropriate for comparing responses from one set of experimental settings (population 1) with independent responses from another set of experimental settings (population 2). The comparison can be made without explicitly controlling for unit-to-unit variability, as in the paired-comparison case.

Consider a random sample of size n_1 from population 1 and a sample of size n_2 from population 2. The observations on p variables can be arranged as follows:

sample	mean	covariance matrix
Population 1:		
$\boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1n_1}$	$\overline{\boldsymbol{x}}_1 = \dfrac{1}{n_1}\sum_{j=1}^{n_1} \boldsymbol{x}_{1j}$	$\mathbf{S}_1 = \dfrac{1}{n_1 - 1}\sum_{j=1}^{n_1}(\boldsymbol{x}_{1j} - \overline{\boldsymbol{x}}_1)(\boldsymbol{x}_{1j} - \overline{\boldsymbol{x}}_1)'$
Population 2:		
$\boldsymbol{x}_{21}, \ldots, \boldsymbol{x}_{2n_2}$	$\overline{\boldsymbol{x}}_2 = \dfrac{1}{n_2}\sum_{j=1}^{n_2} \boldsymbol{x}_{2j}$	$\mathbf{S}_2 = \dfrac{1}{n_2 - 1}\sum_{j=1}^{n_2}(\boldsymbol{x}_{2j} - \overline{\boldsymbol{x}}_2)(\boldsymbol{x}_{2j} - \overline{\boldsymbol{x}}_2)'.$

We want to make inferences about $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$: is $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ and if $\boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$, which component means are different? With a few tentative assumptions, we are able to provide answers to these questions.

Assumptions concerning the Structure of the Data

1. The sample $\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n_1}$, is a random sample of size n_1 from a p-variate population with mean vector $\boldsymbol{\mu}_1$ and covariance matrix $\boldsymbol{\Sigma}_1$.
2. The sample $\boldsymbol{X}_{21}, \ldots, \boldsymbol{X}_{2n_2}$, is a random sample of size n_2 from a p-variate population with mean vector $\boldsymbol{\mu}_2$ and covariance matrix $\boldsymbol{\Sigma}_2$.
3. Independence assumption: $\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n_1}$, are independent of $\boldsymbol{X}_{21}, \ldots, \boldsymbol{X}_{2n_2}$.

We shall see later that, for large samples, this structure is sufficient for making inferences about the $p \times 1$ vector $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. However, when the sample sizes n_1 and n_2 are small, more assumptions are needed.

Further Assumptions when n_1 and n_2 are small

1. Both populations are multivariate normal.
2. Both samples have the same covariance matrix: $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$.

Remark. The second assumption, that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, is much stronger than its univariate counterpart. Here we are assuming that several pairs of variances and covariances are nearly equal.

When $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$ we find that

$$\sum_{j=1}^{n_1}(\boldsymbol{x}_{1j} - \overline{\boldsymbol{x}}_1)(\boldsymbol{x}_{1j} - \overline{\boldsymbol{x}}_1)' \text{ is an estimate of } (n_1 - 1)\boldsymbol{\Sigma} \quad \text{and}$$
$$\sum_{j=1}^{n_2}(\boldsymbol{x}_{2j} - \overline{\boldsymbol{x}}_2)(\boldsymbol{x}_{2j} - \overline{\boldsymbol{x}}_2)' \text{ is an estimate of } (n_2 - 1)\boldsymbol{\Sigma}.$$

Consequently, we can pool the information in both samples in order to estimate the common covariance $\mathbf{\Sigma}$. We set

$$\mathbf{S}_{\text{pooled}} = \frac{\sum_{j=1}^{n_1}(\boldsymbol{x}_{1j} - \overline{\boldsymbol{x}}_1)(\boldsymbol{x}_{1j} - \overline{\boldsymbol{x}}_1)' + \sum_{j=1}^{n_2}(\boldsymbol{x}_{2j} - \overline{\boldsymbol{x}}_2)(\boldsymbol{x}_{2j} - \overline{\boldsymbol{x}}_2)'}{n_1 + n_2 - 2}$$

$$= \frac{n_1 - 1}{n_1 + n_2 - 2}\mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2}\mathbf{S}_2.$$

Since the independence assumption on p. 6-6 implies that $\overline{\boldsymbol{X}}_1$ and $\overline{\boldsymbol{X}}_2$ are independent and thus $\text{Cov}(\overline{\boldsymbol{X}}_1, \overline{\boldsymbol{X}}_2) = \mathbf{0}$, it follows that

$$\text{Cov}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) = \text{Cov}(\overline{\boldsymbol{X}}_1) + \text{Cov}(\overline{\boldsymbol{X}}_2) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{\Sigma}.$$

Because $\mathbf{S}_{\text{pooled}}$ estimates $\mathbf{\Sigma}$, we see that $\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}$ is an estimator of $\text{Cov}(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2)$.

Proposition 6.2.1. *If $\boldsymbol{X}_{11}, \ldots, \boldsymbol{X}_{1n_1}$ is a random sample of size n_1 from $N_p(\boldsymbol{\mu}_1, \mathbf{\Sigma})$ and $\boldsymbol{X}_{21}, \ldots, \boldsymbol{X}_{2n_2}$ is an independent random sample of size n_2 from $N_p(\boldsymbol{\mu}_2, \mathbf{\Sigma})$, then*

$$T^2 = (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{pooled}\right]^{-1} (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))$$

is distributed as

$$\frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p, n_1 + n_2 - p - 1}.$$

Consequently

$$P\left((\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{\text{pooled}}\right]^{-1} (\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \le c^2\right) = 1 - \alpha$$

where

$$c^2 = \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p, n_1 + n_2 - p - 1}(\alpha).$$

6.2.1 Simultaneous Confidence Intervals

It is possible to derive simultaneous confidence intervals for the components of the vector $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$. These confidence intervals are developed from a consideration of all possible linear combinations of the differences in the mean vectors. It is assumed that the parent multivariate populations are normal with a common covariance $\mathbf{\Sigma}$.

Proposition 6.2.2. *Let* $c^2 = \frac{(n_1+n_2-2)p}{n_1+n_2-p-1} F_{p,n_1+n_2-p-1}(\alpha)$*. With probability* $1-\alpha$

$$\boldsymbol{a}'(\overline{\boldsymbol{X}}_1 - \overline{\boldsymbol{X}}_2) \pm c\sqrt{\boldsymbol{a}'\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{S}_{pooled}\boldsymbol{a}}$$

will cover $\boldsymbol{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ *for all* $\boldsymbol{a}$*. In particular* $\mu_{1i} - \mu_{2i}$ *will be covered by*

$$(\overline{X}_{1i} - \overline{X}_{2i}) \pm c\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)s_{ii,pooled}} \quad \text{for } i = 1, \ldots, p.$$

6.2.2 Two-sample Situation when $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$

When $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, we are unable to find a "distance" measure like T^2, whose distribution does not depend on the unknown $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$. However, for n_1 and n_2 large, we can avoid the complexities due to unequal covariance matrices.

Proposition 6.2.3. *Let the sample sizes be such that* $n_1 - p$ *and* $n_2 - p$ *are large. Then, an approximate* $(1-\alpha)$ *confidence ellipsoid for* $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ *is given by all* $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ *satisfying*

$$(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))'\left(\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right)^{-1}(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \leq \chi^2_p(\alpha),$$

where $\chi^2_p(\alpha)$ *is the upper* (100α)*th percentile of a chi-square distribution with* p *d.f.*

Also, $(1-\alpha)$ *simultaneous confidence intervals for all linear combinations* $\boldsymbol{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$ *are provided by*

$$\boldsymbol{a}'(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \quad \text{belongs to} \quad \boldsymbol{a}'(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2) \pm \sqrt{\chi^2_p(\alpha)}\sqrt{\boldsymbol{a}'\left(\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right)\boldsymbol{a}}.$$

6.3 Comparing Several Multivariate Population Means (One-way MANOVA)

Often, more than two populations need to be compared. Multivariate Analysis of Variance (MANOVA) is used to investigate whether the population mean vectors are the same and, if not, which mean components differ significantly.

We start with random samples, collected from each of g populations:

$$\begin{aligned}
&\text{Population 1: } \boldsymbol{X}_{11}, \boldsymbol{X}_{12}, \ldots, \boldsymbol{X}_{1n_1}\\
&\text{Population 2: } \boldsymbol{X}_{21}, \boldsymbol{X}_{22}, \ldots, \boldsymbol{X}_{2n_2}\\
&\qquad\vdots\\
&\text{Population } g\text{: } \boldsymbol{X}_{g1}, \boldsymbol{X}_{g2}, \ldots, \boldsymbol{X}_{gn_g}
\end{aligned}$$

Assumptions about the Structure of the Data for One-way MANOVA

1. $\boldsymbol{X}_{l1}, \boldsymbol{X}_{l2}, \ldots, \boldsymbol{X}_{ln_l}$, is a random sample of size n_l from a population with mean $\boldsymbol{\mu}_l$, $l = 1, 2, \ldots, g$. The random samples from different populations are independent.
2. All populations have a common covariance matrix $\boldsymbol{\Sigma}$.
3. Each population is multivariate normal.

Condition 3 can be relaxed by appealing to the central limit theorem when the sample sizes n_l are large.

A review of the univariate analysis of variance (ANOVA) will facilitate our discussion of the multivariate assumptions and solution methods.

6.3.1 Summary of Univariate ANOVA

Example (Bern-Chur-Zürich, p. 1-4)**.** In Figure 6.2 we get an overview on the annual sunshine duration for Bern, Chur and Zürich for the time period 1930-1960. We observe that the sample median is higher for Bern than for Chur and Zürich, but the differences do not seem to be large. The ANalysis Of VAriance (ANOVA) is the statistical tool to check whether the population means of the sunshine duration are the same for all stations or not.

Assume that $X_{l1}, \ldots, X_{ln_l}$ is a random sample from an $N(\mu_l, \sigma^2)$ population, $l = 1, \ldots, g$, and that the random samples are independent. Although the null hypothesis of equality of means could be formulated as $\mu_1 = \ldots = \mu_g$, it is customary to regard μ_l as the sum of an overall mean component, such as μ, and a component due to the specific population. For instance, we can write

$$\mu_l = \mu + \underbrace{(\mu_l - \mu)}_{=\tau_l}.$$

Populations usually correspond to different sets of experimental conditions, and therefore, it is convenient to investigate the deviations τ_l associated with the lth population. The notation

$$\underset{\substack{l\text{th population}\\ \text{mean}}}{\mu_l} \quad = \quad \underset{\substack{\text{overall}\\ \text{mean}}}{\mu} \quad + \quad \underset{\substack{l\text{th population}\\ \text{treatment effect}}}{\tau_l} \tag{6.2}$$

leads to a restatement of the hypotheses of equality of means. The null hypothesis becomes

$$H_0 : \tau_1 = \ldots = \tau_g = 0.$$

The response X_{lj}, distributed as $N(\mu + \tau_l, \sigma^2)$, can be expressed in the form

$$\begin{array}{ccccccc} X_{lj} & = & \mu & + & \tau_l & + & e_{lj} \\ & & \text{overall} & & \text{treatment} & & \text{random} \\ & & \text{mean} & & \text{effect} & & \text{error} \end{array} \tag{6.3}$$

where the e_{lj} are independent $N(0, \sigma^2)$ random variables.

Motivated by the decomposition in (6.3), the analysis of variance is based upon an analogous decomposition of the observations

$$\begin{array}{ccccccc} x_{lj} & = & \overline{x} & + & (\overline{x}_l - \overline{x}) & + & (x_{lj} - \overline{x}_l) \\ \text{observation} & & \text{overall} & & \text{estimated} & & \text{residual} \\ & & \text{sample mean} & & \text{treatment effect} & & \end{array} \tag{6.4}$$

where $\overline{x}$ is an estimate of μ, $\hat{\tau}_l = (\overline{x}_l - \overline{x})$ is an estimate of τ_l, and $(\overline{x}_{lj} - \overline{x}_l)$ is an estimate of the error e_{lj}.

From (6.4) we calculate $(x_{lj} - \overline{x})^2$ and find

$$\begin{aligned} & (x_{lj} - \overline{x})^2 = (\overline{x}_l - \overline{x})^2 + (x_{lj} - \overline{x}_l)^2 + 2(\overline{x}_l - \overline{x})(x_{lj} - \overline{x}_l) \\ \sum_j : \quad & \sum_{j=1}^{n_l} (x_{lj} - \overline{x})^2 = n_l(\overline{x}_l - \overline{x})^2 + \sum_{j=1}^{n_l} (x_{lj} - \overline{x}_l)^2 \\ \sum_l : \quad & \underbrace{\sum_{l=1}^{g} \sum_{j=1}^{n_l} (x_{lj} - \overline{x})^2}_{\text{total (corrected) SS}} = \underbrace{\sum_{l=1}^{g} n_l(\overline{x}_l - \overline{x})^2}_{\text{between (samples) SS}} + \underbrace{\sum_{l=1}^{g} \sum_{j=1}^{n_l} (x_{lj} - \overline{x}_l)^2}_{\text{within (samples) SS}} \end{aligned}$$

The calculations of the sums of squares and the associated degrees of freedom are conveniently summarized by an ANOVA table (see Table 6.1).

Table 6.1: ANOVA table for comparing univariate population means

Sources of variation	Sum of squares (SS)	Degrees of freedom
Treatments	$SS_{tr} = \sum_{l=1}^{g} n_l(\overline{x}_l - \overline{x})^2$	$g - 1$
Residual (error)	$SS_{res} = \sum_{l=1}^{g} \sum_{j=1}^{n_l} (x_{lj} - \overline{x}_l)^2$	$\sum_{l=1}^{g} n_l - g$
Total (corrected for the mean)	$SS_{cor} = \sum_{l=1}^{g} \sum_{j=1}^{n_l} (x_{lj} - \overline{x})^2$	$\sum_{l=1}^{g} n_l - 1$

The F-test rejects $H_0 : \tau_1 = \ldots = \tau_g = 0$ at level α if

$$F = \frac{SS_{tr}/(g-1)}{SS_{res}/(\sum_{l=1}^{g} n_l - g)} > F_{g-1, \sum n_l - g}(\alpha)$$

where $F_{g-1, \sum n_l - g}(\alpha)$ is the upper (100α)th percentile of the F-distribution with $g-1$ and $\sum n_l - g$ d.f. This is equivalent to rejecting H_0 for large values of SS_{tr}/SS_{res} or for large values of $1+SS_{tr}/SS_{res}$. This statistic appropriate for a multivariate generalization rejects H_0 for small values of the reciprocal

$$\frac{1}{1 + SS_{tr}/SS_{res}} = \frac{SS_{res}}{SS_{res} + SS_{tr}}.$$

Multiple Comparisons of Means

Source: Schuenemeyer and Drew (2011), pp. 87-90.

After determining that there is a statistically significant difference between population means, the investigator needs to determine where the difference occur. The concept of a two-sample t-test was introduced earlier. At this point it may be reasonable to ask: Why not use it? The problem is that the probability of rejecting the null hypothesis simply by chance (where real differences between population means fail to exist) increases as the number of pairwise tests increase. It is difficult to determine what level of confidence will be achieved for claiming that all statements are correct. To overcome this dilemma, procedures have been developed for several confidence intervals to be constructed in such a manner that the joint probability that all the statements are true is guaranteed not to fall below a predetermined level. Such intervals are called multiple confidence intervals or simultaneous confidence intervals. Three methods are often discussed in literature, which will be summarized here.

Bonferronis Method Bonferronis method is a simple procedure that can be applied to equal and unequal sample sizes. If a decision is made to make m pairwise comparisons, selected in advance, the Bonferroni method requires that the significance level on each test be α/m. This ensures that the overall (experiment wide) probability of making an error is less than or equal to α. Comparisons can be made on means and specified linear combinations of means (contrasts). Bonferronis method is sometimes called the Dunn method. A variant on the Bonferroni approach is the Sidak method, which yields slightly tighter confidence bounds. If the treatments are to be compared against a control group, Dunnetts test should be used. Clearly, the penalty that is paid for using Bonferronis method is the increased difficulty of rejecting the null hypothesis on a single comparison. The advantage is protection against an error when making multiple comparisons.

Tukeys Method Tukey's method provides $(1-\alpha)$ simultaneous confidence intervals for all pairwise comparisons. Tukeys method is exact when sample sizes are equal and is conservative when they are not. As in the Bonferroni method, the Tukey method makes it more difficult to reject H_0 on a single comparison, thereby preserving the a level chosen for the entire experiment.

Scheffés Method Scheffé's method applies to all possible contrasts of the form

$$C = \sum_{i=1}^{k} c_i \mu_i$$

where $\sum_{i=1}^{k} c_i = 0$ and k is the number of treatment groups. Thus, in theory an infinite number of contrasts can be defined. H_0 is rejected if and only if at least one interval for a contrast does not contain zero.

Discussion of Multiple Comparison Procedures If H_0 in an ANOVA is rejected, it is proper to go to a multiple comparison procedure to determine specific differences. On the other side Mendenhall and Sincich (2012) recommends to avoid conducting multiple comparisons of a small number of treatment means when the corresponding ANOVA F-test is nonsignificant; otherwise, confusing and contradictory results may occur.

An obvious question is, "Which procedure should I pick?" The basic idea is to have the narrowest confidence bounds for the entire experiment (set of comparisons), consistent with the contrasts of interest. Equivalently, in hypothesis-testing parlance, it is desirable to make it as easy as possible to reject H_0 on a single comparison while preserving a predetermined significance level for the experiment. If all pairwise comparisons are of interest, the Tukey method is recommended; if only a subset is of interest, Bonferronis method is a better choice. If all possible contrasts are desired, Scheffés method should be used. However, because all possible contrasts must be considered in Scheffés method, rejecting the null hypothesis is extremely difficult. As with most statistical procedures, no single method works best in all situations.

Further reading. Mendenhall and Sincich (2012), pp. 671–692.

6.3.2 Multivariate Analysis of Variance (MANOVA)

Example (Bern-Chur-Zürich, p. 1-4)**.** In addition to ANOVA we extend the analysis to more than one variable, to the Multivariate ANalysis Of VAriance (MANOVA). The null hypothesis states that the population mean vectors including the variables pressure, temperature, precipitation and sunshine duration for the three stations are the same. Figure 6.2, p. 6-5, gives a visual indication of possible differences of the population means.

Paralleling the univariate reparameterization, we specify the MANOVA model:

Definition 6.3.1. MANOVA model for comparing g population mean vectors:

$$\boldsymbol{X}_{lj} = \boldsymbol{\mu} + \boldsymbol{\tau}_l + \boldsymbol{e}_{lj}, \qquad j = 1, \ldots, n_l, \quad l = 1, \ldots, g, \tag{6.5}$$

where $\boldsymbol{e}_{lj}$ are independent $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ variables. Here the parameter vector $\boldsymbol{\mu}$ is an overall mean (level), and $\boldsymbol{\tau}_l$ represents the lth treatment effect with

$$\sum_{l=1}^{g} n_l \boldsymbol{\tau}_l = \mathbf{0}.$$

According to the model in (6.5), each component of the observation vector $\boldsymbol{X}_{lj}$ satisfies the univariate model (6.3). The errors for the components of $\boldsymbol{X}_{lj}$ are correlated, but the covariance matrix $\boldsymbol{\Sigma}$ is the same for all populations.

A vector of observations may be decomposed as suggested by the model. Thus,

$$\begin{array}{ccccccc} \boldsymbol{x}_{lj} & = & \overline{\boldsymbol{x}} & + & (\overline{\boldsymbol{x}}_l - \overline{\boldsymbol{x}}) & + & (\boldsymbol{x}_{lj} - \overline{\boldsymbol{x}}_l) \\ \text{observation} & & \text{overall sample} & & \text{estimated treatment} & & \text{residual} \\ & & \text{mean } \hat{\boldsymbol{\mu}} & & \text{effect } \hat{\boldsymbol{\tau}}_l & & \hat{\boldsymbol{e}}_{lj} \end{array} \tag{6.6}$$

The decomposition in (6.6) leads to the multivariate analog of the univariate sum of squares breakup in (6.5):

$$\begin{array}{ccccc} \sum_{l=1}^{g}\sum_{j=1}^{n_l}(\boldsymbol{x}_{lj} - \overline{\boldsymbol{x}})(\boldsymbol{x}_{lj} - \overline{\boldsymbol{x}})' & = & \sum_{l=1}^{g} n_l(\overline{\boldsymbol{x}}_l - \overline{\boldsymbol{x}})(\overline{\boldsymbol{x}}_l - \overline{\boldsymbol{x}})' & + & \sum_{l=1}^{g}\sum_{j=1}^{n_l}(\boldsymbol{x}_{lj} - \overline{\boldsymbol{x}}_l)(\boldsymbol{x}_{lj} - \overline{\boldsymbol{x}}_l)' \\ \text{total (corrected) SS} & & \text{between (samples) SS} & & \text{within (samples) SS} \end{array} \tag{6.7}$$

The within sum of squares and cross products matrix can be expressed as

$$\begin{aligned} \mathbf{W} &:= \sum_{l=1}^{g}\sum_{j=1}^{n_l}(\boldsymbol{x}_{lj} - \overline{\boldsymbol{x}}_l)(\boldsymbol{x}_{lj} - \overline{\boldsymbol{x}}_l)' \\ &= \sum_{l=1}^{g}(n_l - 1)\mathbf{S}_l, \end{aligned}$$

where $\mathbf{S}_l$ is the sample covariance matrix for the lth sample. This matrix is a generalization of the $(n_1 + n_2 - 2)\mathbf{S}_{\text{pooled}}$ matrix encountered in the two-sample case.

Analogous to the univariate result, the hypothesis of no treatment effects,

$$H_0 : \boldsymbol{\tau}_1 = \ldots = \boldsymbol{\tau}_g = \mathbf{0}$$

is tested by considering the relative sizes of the treatment and residual sums of squares and cross products. Equivalently, we may consider the relative sizes of the residual and total (corrected) sum of squares and cross products. Formally, we summarize the calculations leading to the test statistic in a MANOVA table.

This table is exactly of the same form, component by component, as in the ANOVA table, expect that squares are replaced by their vector counterparts.

One test of $H_0 : \boldsymbol{\tau}_1 = \ldots = \boldsymbol{\tau}_g = \mathbf{0}$ involves generalized variances. We reject H_0 if the ratio of generalized variances

$$\Lambda^\star = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} \qquad \text{Wilks' lambda}$$

is too small. The exact distribution of $\Lambda^\star$ can be derived for the special cases listed in Table 6.3. For other cases and large sample sizes, a modification of $\Lambda^\star$ due to Bartlett can be used to test H_0.

Table 6.2: MANOVA table for comparing population mean vectors

Sources of variation	Matrix of Sum of squares (SS) and cross products	Degrees of freedom
Treatment	$\mathbf{B} = \sum_{l=1}^{g} n_l(\overline{\boldsymbol{x}}_l - \overline{\boldsymbol{x}})(\overline{\boldsymbol{x}}_l - \overline{\boldsymbol{x}})'$	$g-1$
Residual (Error)	$\mathbf{W} = \sum_{l=1}^{g}\sum_{j=1}^{n_l}(\boldsymbol{x}_{lj} - \overline{\boldsymbol{x}}_l)(\boldsymbol{x}_{lj} - \overline{\boldsymbol{x}}_l)'$	$\sum_{l=1}^{g} n_l - g$
Total	$\mathbf{B} + \mathbf{W} = \sum_{l=1}^{g}\sum_{j=1}^{n_l}(\boldsymbol{x}_{lj} - \overline{\boldsymbol{x}})(\boldsymbol{x}_{lj} - \overline{\boldsymbol{x}})'$	$\sum_{l=1}^{g} n_l - 1$

Remark. There are other statistics for checking the equality of several multivariate means, such as Pillai's statistic, Lawley-Hotelling and Roy's largest root.

Remark. Bartlett has shown that if H_0 is true and $\sum n_l = n$ is large,

$$-\left(n - 1 - \frac{p+g}{2}\right)\ln \Lambda^\star$$

has approximately a chi-square distribution with $p(g-1)$ d.f. Consequently, for $\sum n_l = n$ large, we reject H_0 at significance level α if

$$-\left(n - 1 - \frac{p+g}{2}\right)\ln \Lambda^\star > \chi^2_{p(g-1)}(\alpha),$$

where $\chi^2_{p(g-1)}(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with $p(g-1)$ d.f.

6.3.3 Simultaneous Confidence Intervals for Treatment Effects

When the hypothesis of equal treatment effects is rejected, those effects that led to the rejection of the hypothesis are of interest. For pairwise comparisons the Bonferroni approach can be used to construct simultaneous confidence intervals for the components of the differences

$$\boldsymbol{\tau}_k - \boldsymbol{\tau}_l \quad \text{or} \quad \boldsymbol{\mu}_k - \boldsymbol{\mu}_l.$$

These intervals are shorter than those obtained for all contrasts, and they require critical values only for the univariate t-statistic.

Proposition 6.3.2. *Let $n = \sum_{k=1}^{g} n_k$. For the model in* (6.5), *with confidence at least* $(1-\alpha)$, $\tau_{ki} - \tau_{li}$ *belongs to*

$$\overline{x}_{ki} - \overline{x}_{li} \pm t_{n-g}\left(\frac{\alpha}{pg(g-1)}\right)\sqrt{\frac{w_{ii}}{n-g}\left(\frac{1}{n_k} + \frac{1}{n_l}\right)}$$

for all components $i = 1, \ldots, p$ and all differences $l < k = 1, \ldots, g$. Here w_{ii} is the ith diagonal element of $\mathbf{W}$.

Table 6.3: Distribution of Wilks' Lambda $\Lambda^\star$. Source: Johnson and Wichern (2007).

No. of variables	No. of groups	Sampling distribution for multivariate normal data
$p = 1$	$g \geq 2$	$\left(\frac{\Sigma n_\ell - g}{g - 1}\right)\left(\frac{1 - \Lambda^*}{\Lambda^*}\right) \sim F_{g-1, \Sigma n_\ell - g}$
$p = 2$	$g \geq 2$	$\left(\frac{\Sigma n_\ell - g - 1}{g - 1}\right)\left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1), 2(\Sigma n_\ell - g - 1)}$
$p \geq 1$	$g = 2$	$\left(\frac{\Sigma n_\ell - p - 1}{p}\right)\left(\frac{1 - \Lambda^*}{\Lambda^*}\right) \sim F_{p, \Sigma n_\ell - p - 1}$
$p \geq 1$	$g = 3$	$\left(\frac{\Sigma n_\ell - p - 2}{p}\right)\left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2p, 2(\Sigma n_\ell - p - 2)}$

6.3.4 Testing for Equality of Covariance Matrices

One of the assumptions made when comparing two or more multivariate mean vectors is that the covariance matrices of the potentially different populations are the same. Before pooling the variation across samples to form a pooled covariance matrix when comparing mean vectors, it can be worthwhile to test the equality of the population covariance matrices. One commonly employed test for equal covariance matrices is Box's M-test (see Johnson and Wichern (2007) p. 311).

With g populations, the null hypothesis is

$$H_0 : \mathbf{\Sigma}_1 = \ldots = \mathbf{\Sigma}_g = \mathbf{\Sigma},$$

where $\mathbf{\Sigma}_l$ is the covariance matrix for the lth population, $l = 1, \ldots, g$, and $\mathbf{\Sigma}$ is the presumed common covariance matrix. The alternative hypothesis H_1 is that at least two of the covariance matrices are not equal.

Remark. Box's χ^2 approximation works well if for each group l, $l = 1, \ldots, g$, the sample size n_l exceeds 20 and if p and g do not exceed 5.

Remark. Box's M-test is routinely calculated in many statistical computer packages that do MANOVA and other procedures requiring equal covariance matrices. It is known that the M-test is sensitive to some forms of non-normality. However, with reasonably large samples, the MANOVA tests of means or treatment effects are rather robust to non-normality. Thus the M-test may reject H_0 in some non-normal cases where it is not damaging to the MANOVA tests. Moreover, with equal sample sizes, some differences in covariance matrices have little effect on the MANOVA test. To summarize, we may decide to continue with the usual MANOVA tests even though the M-test leads to rejection of H_0.