Chapter 2
The Linear Hypothesis

2.1 Linear Regression

Ini this chapter we consider a number of linear hypotheses before giving a general definition. Our first example is found in regression analysis.

Example 2.1 Suppose we have a random variable y with mean θ and we assume that θ is a linear function of p non-random variables $x_0, x_1, \ldots, x_{p-1}$ called regressors or explanatory variables, namely,

$$\theta = \beta_0 x_0 + \beta_1 x_1 + \cdots + \beta_{p-1} x_{p-1},$$

where the β's are unknown constants (parameters). For n values of the x's, we get n observations on y, giving the model G

$$\begin{aligned} y_i &= \theta_i + \varepsilon_i \\ &= x_{i0}\beta_0 + x_{i1}\beta_1 + \cdots + x_{i,p-1}\beta_{p-1} + \varepsilon_i, \quad (i = 1, 2, \ldots, n), \end{aligned}$$

where $E[\varepsilon_i] = 0$; generally $x_{i0} = 1$, which we shall assume unless stated otherwise. This is known as a multiple linear regression model with p parameters, and by putting $x_{ij} = x_i^j$ we see that the polynomial regression model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_{p-1} x_i^{p-1} + \varepsilon_i,$$

of degree $p-1$ for a single variable x is included as a special case. We can also have a mixture of both models. The linearity resides in the parameters.

Two further assumptions about the errors ε_i are generally made: (i) the errors are uncorrelated, or $\mathrm{cov}[\varepsilon_i, \varepsilon_j] = 0$ for all $i \neq j$ and (ii) the errors have the same variance σ^2. If we wish to test the null hypothesis $H : \beta_r = \beta_{r+1} = \cdots = \beta_{p-1} = 0$, then we

G.A.F. Seber, *The Linear Model and Hypothesis*, Springer Series in Statistics,
DOI 10.1007/978-3-319-21930-1_2

need to add a further assumption that the errors are normally distributed. If we define $\mathbf{X} = (x_{ij})$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$, and let $\mathbf{X}_r$ represent the matrix consisting of the first r columns of $\mathbf{X}$, then the model, assumptions, and hypothesis can be written in the form $\mathbf{y} = \boldsymbol{\theta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N_n[\mathbf{0}, \sigma^2\mathbf{I}_n]$, $G : \boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta}$ and $H : \boldsymbol{\theta} = \mathbf{X}_r\boldsymbol{\beta}_r$, where $\boldsymbol{\beta}_r$ is the vector of the first r elements of $\boldsymbol{\beta}$. In this situation $\mathbf{X}$ usually has full rank, that is the rank of $\mathbf{X}$ is p. If we define the two column spaces $\Omega = \mathcal{C}[\mathbf{X}]$ and $\omega = \mathcal{C}[\mathbf{X}_r]$, then it follows from Sect. 1.2 that Ω and ω are vector subspaces of $\mathbb{R}^n$ and $\omega \subset \Omega$. Thus H is the linear hypothesis that $\boldsymbol{\theta}$ belongs to a vector space ω given the assumption G that it belongs to a vector space Ω. We also have that $\mathrm{Var}[\mathbf{y}] = \mathrm{Var}[\mathbf{y} - \boldsymbol{\theta}] = \mathrm{Var}[\boldsymbol{\varepsilon}] = \boldsymbol{\sigma}^2\mathbf{I}_n$ (Theorem 1.5(v)) so that $\mathbf{y} \sim N_n[\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n]$.

2.2 Analysis of Variance

Example 2.2 We note that some of the x-variables in our regression model can also be so-called *indicator variables*, that is variables taking the values of 0 or 1. For example consider n observations from the straight-line model

$$E[y_i] = \beta_0 + \beta_1 x_i, \quad i = 1, 2, \ldots, n,$$

where $x_i = 0$ for $i = 1, 2, \ldots n_1$ and $x_i = 1$ for $i = n_1 + 1, n_1 + 2, \ldots, n$. If $n - n_1 = n_2$, then $\mathbf{X}\boldsymbol{\beta}$ takes the form

$$\mathbf{X}\boldsymbol{\beta} = \begin{pmatrix} \mathbf{1}_{n_1} & \mathbf{0} \\ \mathbf{1}_{n_2} & \mathbf{1}_{n_2} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}.$$

This model splits into two models or samples, namely $E[y_i] = \beta_0$ for $i = 1, 2, \ldots, n_1$ and $E[y_i] = \beta_0 + \beta_1$ for $i = 1, 2, \ldots, n_2$. This would give us a model for comparing the means $\mu_1 (= \beta_0)$ and $\mu_2 (= \beta_0 + \beta_1)$ of two samples of sizes n_1 and n_2 respectively. Testing if $\mu_1 = \mu_2$ is equivalent to testing $\beta_1 = 0$. This type of model where variables enter qualitatively is sometimes referred to as an analysis of variance (ANOVA) model.

Example 2.3 We now consider generalizing the above example to comparing I different samples with J_i observations in the ith sample. Let y_{ij} ($i = 1, 2, \ldots, I$ and $j = 1, 2, \ldots J_i$) be the jth observation from the ith sample, so that we have the model $y_{ij} = \mu_i + \varepsilon_{ij}$. Setting $\mathbf{y} = \boldsymbol{\theta} + \boldsymbol{\varepsilon}$, where

$$\mathbf{y}' = (y_{11}, y_{12}, \ldots y_{1J_1}, y_{21}, y_{22}, \ldots, y_{2J_2}, \ldots, y_{I1}, y_{I2} \ldots, y_{IJ_I}),$$

and $\boldsymbol{\theta}$ is similarly defined, we get $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\mu}$, where

$$\mathbf{X} = \begin{pmatrix} \mathbf{1}_{J_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \cdots & \mathbf{1}_{J_2} & \cdots & \mathbf{0} \\ \cdot & \cdot & \ddots & \cdot \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{J_I} \end{pmatrix}, \tag{2.1}$$

and $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_I)'$. Suppose we wish to test the hypothesis $H : \mu_1 = \mu_2 = \cdots = \mu_I$ (= μ, say), or $\boldsymbol{\theta} = \mathbf{1}_n\mu$, where $\mathbf{1}_n$ is obtained by adding the columns of $\mathbf{X}$ together. Then, from the previous section, $\Omega = \mathcal{C}[\mathbf{X}]$ and $\omega = \mathcal{C}[\mathbf{1}_n]$.

Alternatively, we can express H in the form

$$\mu_1 - \mu_2 = \mu_2 - \mu_3 = \cdots = \mu_{I-1} - \mu_I = 0,$$

which can be written in matrix form $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$, where

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

Since $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\mu}$ and $\mathbf{X}$ has full rank p, the $p \times p$ matrix $\mathbf{X}'\mathbf{X}$ has rank p and is therefore nonsingular (cf. A.4(ii)). From $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\mu}$ we can then multiply on the left by $\mathbf{X}'$ and get $\boldsymbol{\mu} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\theta}$. Hence H takes the form

$$\mathbf{0} = \mathbf{C}\boldsymbol{\mu} = \mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\theta} = \mathbf{B}\boldsymbol{\theta}, \tag{2.2}$$

say, or $\boldsymbol{\theta} \in \omega$, where $\omega = \mathcal{C}[\mathbf{X}] \cap \mathcal{N}[\mathbf{B}]$.

An alternative parametrization can be used for the above example that is more typical of analysis of variance models. Let $\mu = \sum_{i=1}^{I} \mu_i/I$ and define $\alpha_i = \mu_i - \mu$ so that $\mu_i = \mu + \alpha_i$. Then $\sum_{i=1}^{I} \alpha_i = 0$ is an "identifiability condition" (see Sect. 3.4) giving us $I + 1$ parameters or I free parameters still. We now have

$$\mathbf{X}\boldsymbol{\beta} = \begin{pmatrix} \mathbf{1}_{J_1} & \mathbf{1}_{J_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{1}_{J_2} & \mathbf{0} & \mathbf{1}_{J_2} & \cdots & \mathbf{0} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \mathbf{1}_{J_I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{1}_{J_I} \end{pmatrix} \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_I \end{pmatrix}, \tag{2.3}$$

where the first column of $\mathbf{X}$, namely $\mathbf{1}_n$, is the sum of the other columns, and the matrix $\mathbf{X}$ is no longer of full rank.

Example 2.4 We consider one other ANOVA model, the randomized block design where there are J blocks and I treatments randomized in each block. Let y_{ij} with mean θ_{ij} be the observation from the ith treatment in the jth block and, for $i = 1, 2, \ldots, I$, let $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iJ})'$ and $\boldsymbol{\theta}_i = (\theta_{i1}, \theta_{i2}, \ldots \theta_{iJ})'$. Let $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots \mathbf{y}_I')'$ with $\boldsymbol{\theta}$ and ε similarly defined. We assume the model

$$y_{ij} = \theta_{ij} + \varepsilon_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad (i = 1, 2, \ldots, I : j = 1, 2, \ldots, J),$$

or $\mathbf{y} = \boldsymbol{\theta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\delta}$, namely

$$\begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \cdot \\ \boldsymbol{\theta}_I \end{pmatrix} = \begin{pmatrix} \mathbf{1}_J & | & \mathbf{1}_J & \mathbf{0} & \mathbf{0} \cdots & \mathbf{0} & | & \mathbf{I}_J \\ \mathbf{1}_J & | & \mathbf{0} & \mathbf{1}_J & \mathbf{0} \cdots & \mathbf{0} & | & \mathbf{I}_J \\ \cdot & | & \cdot & \cdot & \cdot \cdots & \cdot & | & \cdot \\ \mathbf{1}_J & | & \mathbf{0} & \mathbf{0} & \mathbf{0} \cdots & \mathbf{1}_J & | & \mathbf{I}_J \end{pmatrix} \begin{pmatrix} \mu \\ \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix},$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_I)'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_J)'$.

We have IJ observations and $1 + I + J$ unknown parameters. Setting $\overline{\theta}_{i\cdot} = \sum_j \theta_{ij}/J$ and $\overline{\theta}_{\cdot\cdot} = \sum_i \sum_j \theta_{ij}/IJ$ etc., we assume from the randomization process that the so-called interactions $\gamma_{ij} = \theta_{ij} - \overline{\theta}_{i\cdot} - \overline{\theta}_{\cdot j} + \overline{\theta}_{\cdot\cdot}$ are all zero, i.e., $\mathbf{C}\boldsymbol{\theta} = \mathbf{0}$ for some matrix $\mathbf{C}$. Since we have $\sum_i \gamma_{ij} = 0$ for $j = 1, 2, \ldots, J$, $\sum_j \gamma_{ij} = 0$ for $i = 1, 2, \ldots, I$, and both sets include $\sum_i \sum_j \gamma_{ij} = 0$, we have $IJ - I - J + 1 = (I-1)(J-1)$ independent constraints so that $\mathbf{C}$ will be $(I-1)(J-1) \times IJ$. The number of parameters that can be estimated is $IJ - (I-1)(J-1) = I + J - 1$, which means we have 2 too many parameters in $\boldsymbol{\delta}$. We need to add two identifiability constraints such as $\sum_i \alpha_i = 0$ and $\sum_j \beta_j = 0$, or $\alpha_I = 0$ and $\beta_J = 0$, for example. By summing columns, we see that the matrix $\mathbf{X}$ above has two linearly dependent columns so that it is $IJ \times (1 + I + J)$ of rank $I + J - 1$. If we set $\alpha_I = 0$ and $\beta_J = 0$ then $\mathbf{X}$ is reduced to $\mathbf{X}_1$, say, with full rank and the same column space as that of $\mathbf{X}$, and $\boldsymbol{\delta}$ is reduced by two elements to $\boldsymbol{\delta}_1$, say. We are usually interested in testing H that there are no differences in the treatments. Then $H : \alpha_1 = \alpha_2 = \cdots = \alpha_{I-1} = 0$ or $\mathbf{C}_1\boldsymbol{\delta}_1 = \mathbf{0}$, say. Using (2.2) with $\boldsymbol{\delta}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\theta}$, we now have $\Omega = \mathcal{C}[\mathbf{X}] \cap \mathcal{N}[\mathbf{C}]$ and $\omega = \Omega \cap \mathcal{N}[\mathbf{C}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1']$.

2.3 Analysis of Covariance

When we have a mixture of quantitive and qualitative explanatory variables we have a so-called analysis of covariance model. For example

$$y_{ij} = \mu_i + \gamma_i z_{ij} + \varepsilon_{ij} \quad (i = 1, 2, \ldots, I : j = 1, 2, \ldots, J_i)$$

represents observations from I straight-line models. Two hypotheses are of interest, namely H_1 that the lines are parallel (i.e. equal γ_i) and H_2 that the lines have the same intercept on the x-axis (i.e. equal μ_i). If both hypotheses are true, the lines are identical. This model G can usually be regarded as the "sum" of two models with $\Omega = \mathcal{C}[\mathbf{X}] \oplus \mathcal{C}[\mathbf{Z}]$, where $\mathbf{Z} = (z_{ij})$, $\mathbf{X}$ is given by Eq. (2.1) in the previous section, and $\mathcal{C}[\mathbf{X}] \cap \mathcal{C}[\mathbf{Z}] = \mathbf{0}$. Such "augmented" models are discussed in Chap. 7.

2.4 General Definition and Extensions

The above examples illustrate what we mean by a linear hypothesis, and we now give a formal definition. Let $\mathbf{y} = \boldsymbol{\theta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\theta}$ is known to belong to a vector space Ω, then a linear hypothesis H is a hypothesis which states that $\boldsymbol{\theta} \in \omega$, a linear subspace of Ω. The assumption that $\boldsymbol{\theta} \in \Omega$ we denote by G. For purposes of estimation we add the assumptions $\mathrm{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$ and $\mathrm{Var}[\mathbf{y}] = \mathrm{Var}[\boldsymbol{\varepsilon}] = \sigma^2\mathbf{I}_n$, and for testing H we add the further assumption that $\boldsymbol{\varepsilon}$ has the multivariate normal distribution. We now consider three extensions.

Example 2.5 There is one hypothesis that is basically linear, but does not satisfy the definition. For example, suppose $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ is $n \times p$ of full column rank p, say, and we wish to test $\mathbf{H} : \mathbf{A}\boldsymbol{\beta} = \mathbf{a}$, where $\mathbf{A}$ and $\mathbf{a}$ are known and $\mathbf{a} \neq \mathbf{0}$. Now $(\boldsymbol{\beta} = \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\theta}$, so that $\omega = \{\boldsymbol{\theta} : \mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\theta} = \mathbf{a}\}$ is not a linear vector space (technically a linear manifold) when $\mathbf{a} \neq \mathbf{0}$. However, if we choose any vector $\mathbf{c}$ such that $\mathbf{Ac} = \mathbf{a}$ (which is possible if the linear equations $\mathbf{A}\boldsymbol{\beta} = \mathbf{a}$ are consistent) and put

$$\mathbf{z} = \mathbf{y} - \mathbf{Xc}, \quad \boldsymbol{\phi} = \boldsymbol{\theta} - \mathbf{Xc} = \mathbf{X}(\boldsymbol{\beta} - \mathbf{c}), \quad \text{and} \quad \boldsymbol{\gamma} = \boldsymbol{\beta} - \mathbf{c},$$

we have

$$\mathbf{z} = \boldsymbol{\phi} + \boldsymbol{\varepsilon}, \quad G : \boldsymbol{\phi} = \mathbf{X}\boldsymbol{\gamma},$$

and $H : \mathbf{A}\boldsymbol{\gamma} = \mathbf{A}(\boldsymbol{\beta} - \mathbf{c}) = \mathbf{0}$ or $\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\phi} = \mathbf{A}_1\boldsymbol{\phi} = \mathbf{0}$ is now a linear hypothesis with $\omega = \mathcal{N}[\mathbf{A}_1] \cap \Omega$ and $\Omega = \mathcal{C}[\mathbf{X}]$.

Example 2.6 In some examples the underlying model takes the form $\mathbf{y} = \boldsymbol{\theta} + \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is $N_n[\mathbf{0}, \sigma^2\mathbf{B}]$ and $\mathbf{B}$ is a known positive-definite matrix. This implies that there exists a nonsingular matrix $\mathbf{V}$ such that $\mathbf{B} = \mathbf{VV}'$ (by A.9(iii)). Using the transformations $\mathbf{z} = \mathbf{V}^{-1}\mathbf{y}$, $\boldsymbol{\phi} = \mathbf{V}^{-1}\boldsymbol{\theta}$, and $\boldsymbol{\varepsilon} = \mathbf{V}^{-1}\boldsymbol{\eta}$ we can transform the model to $\mathbf{z} = \boldsymbol{\phi} + \boldsymbol{\varepsilon}$, where by Theorem 1.5(iii) in Sect. 1.6,

$$\begin{aligned}
\mathrm{Var}[\boldsymbol{\varepsilon}] &= \mathrm{Var}[\mathbf{V}^{-1}\boldsymbol{\eta}] \\
&= \mathbf{V}^{-1}\mathrm{Var}[\boldsymbol{\eta}](\mathbf{V}^{-1})' \\
&= \sigma^2\mathbf{V}^{-1}(\mathbf{VV}')(\mathbf{V}')^{-1} = \sigma^2\mathbf{I}_n,
\end{aligned}$$

as before. To see that linear hypotheses remain linear, let the columns of $\mathbf{W}$ be any basis of Ω. Then

$$\begin{aligned}\Omega &= \{\boldsymbol{\theta} : \boldsymbol{\theta} = \mathbf{W}\beta\} \\ &= \{\phi : \phi = \mathbf{V}^{-1}\mathbf{W}\beta\} \\ &= \mathcal{C}[\mathbf{V}^{-1}\mathbf{W}].\end{aligned}$$

To test $\mathbf{A}\boldsymbol{\beta} = \mathbf{0}$ we note from above that $\boldsymbol{\beta} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\boldsymbol{\theta}$ so that we have $H : \mathbf{A}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{V}\phi = \mathbf{0}$ or $\omega = \Omega \cap \mathcal{N}[\mathbf{A}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{V}]$.

Example 2.7 One model of interest is $\mathbf{y} = \boldsymbol{\theta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N_n[\mathbf{0}, \mathbf{I}_n]$, $\Omega = \mathbb{R}^n$, and ω is a subspace of $\mathbb{R}^n$. Although this model appears to be impractical, it does arise in the large sample theory used in the last three chapters of this monograph. Large sample models and hypotheses are shown there to be asymptotically equivalent to this simple situation.

Springer

http://www.springer.com/978-3-319-21929-5

The Linear Model and Hypothesis
A General Unifying Theory
Seber, G.
2015, IX, 205 p., Hardcover
ISBN: 978-3-319-21929-5