Listy Biometryczne – Biometrical Letters
Vol. 31 (1994), No. 2, 79-100

On modelling of experiments in natural sciences

Stanisław Mejza

Department of Mathematical and Statistical Methods,
Agricultural University, Poznań, Poland

Summary

The paper deals with the problems connected with model building for popular types of designed experiments. The one- and two-factor experiments carried out in design with one or more blocking systems are taken into account only.

A block design, a nested block design, a row-column design, and a block design with nested rows and columns are considered for one-factor experiments. For two-factor experiments a classic two- (many-) factorial design, a split-plot and split-block design are considered only. The former designs include the incomplete, complete and over complete cases of that designs.

Model building is based on some assumptions connected with the experimental unit, its properties and scheme of randomization used in the experiment.

In particular, it is assumed that the observed response is a sum of three components: a *conceptual response* connected with an experimental unit, a *pure effect* due to treatment (combination) and a *technical effect* connected with measurements. It means that additivity among these three components is also assumed.

Special attention is paid to validity, with respect to the randomization point of view, of linear model assumptions adopted in many applications.

The paper is the survey paper where research done by the author in the considered area is reported.

1. Introduction

The experiment is an important tool of research in natural sciences. Hence, planning, modelling and inference problems are of fundamental importance for every experimenter using experiments in his research work.

There are, in practice, two main approaches to the model building of a linear model of observations. In the first approach we assume a'priori a form of the linear model, usually before performing the experiment. The linear model and

Key words: randomization, additivity, model building, designs with block structures, factorial experiments.

its dispersion structure are assumed to be independent of the type of experiment and the structure of the experimental material. Sometimes some additional assumptions concerning dispersion structure (correlation, auto correlation) are added. The problem is how to check these assumptions. Usually, the amount of data necessary to verify the assumptions, is beyond the ability of the experimenter to collect them.

In the second approach, the model is strictly connected with a given experiment, i.e. with the structure of its experimental material and with the method of assigning treatments to the units, the so-called scheme of randomization. This procedure should be worked out separately for every type of the experiment considered.

In this paper we present several different schemes of randomizations for the most commonly used types of designs, i.e. a block design, a nested block design, a row-column design, a classic two-factor design, a split-plot design and a split-block design.

The main purpose of the paper is to illustrate a more objective procedure for derivation (definition) of the model. The emphasis is directed to the experimental situation and to the design. In our considerations the process of randomization plays a central role.

At the beginning let us consider the factors which have an influence on the value of the observed data (also called observed response, observed yield). Let us consider more exactly these factors in some areas of natural science.

Let us begin with an agricultural field experiment. In this kind of experiment the plot usually constitutes an experimental unit. An observed datum is then often called an observed yield. Such terminology is very helpful for understanding the considered ideas.

Note that every unit possesses some kind of fertility which gives some yield in the case when treatments do not occur on a unit and in the case in which no treatments have an effect on the yield. This yield will be called zero yield (conceptual response). The increase (or decrease) in zero yield due to the treatment used on the experimental unit will be called pure effect (due to treatment). We have a similar situation in biological (and some agricultural) experiments in which an animal (or a set of animals) constitutes an experimental unit. Then, analogous to fertility is the so called vitality or resistance, or immunity (it depends on what kind of properties of a body are taken into account).

Suppose we are interested in investigating the influence of some additional levels of vitamins or drugs on some body characteristics. These characteristics are natural properties of the organism. The body characteristic resulting from these natural properties of the organism is called conceptual response. The additional levels of vitamin (treatments) can increase (or decrease) the conceptual response. Hence, the increase (or decrease) of the conceptual response caused by the treatment applied on a given unit (animal or a set of animals) will be called pure effect (due to treatment).

We have quite a similar situation in some medical experiments.

Usually the sum of zero yield (conceptual response) and pure effect due to treatment is called the pure yield (pure response) and is often the base of the statistical analysis.

Let us note that when observing the response of the unit in reality, any observation may be affected by a "technical error", (artifact) an error due to some technical inaccuracy in performing the experiment and due to some error connected with measurements of the response (data). This error is also called measurement error (cf. Neyman et al., 1935).

The properties of the random variable representing measurement error are strictly connected with the measurements. It means that the assumption that there is no systematic error implies 0 as the expected value. Moreover, independence of measurements - implies that all covariances are 0 while all variances are equal to σ^2, where σ^2 is the common variance for all units (resulting from using the same methods of obtaining data). Hence, we assume that the technical errors, denoted by ε, are independent variables, all with the same expected value equal to 0 and with the same variance σ^2. Also, we assume that technical errors are independent of the other random terms of a linear model.

In this paper, the so called comparative experiments are taken into account. The purpose of such an experiment is to discover whether the treatment has an effect and how great it is, as compared with some others treatments, often the standard (control) ones.

Let us note that in natural science experiments the most important thing is to eliminate, as far as possible, a heterogeneity of experimental material. In field experiments it may be possible to find or to prepare a homogeneous set of parcels (plots). However, in biological and medical experiments the situation is more complex. In medical experiments, where the patients are units, the patients are not all alike, nor are physicians, hospitals or communities. Moreover, under the most carefully controlled conditions, patients do as they please. Also, the psychological conditions which are practically impossible to eliminate are very important. Hence, as we will see later, the randomization of units is of great importance in such cases.

The starting point of our considerations is some theoretical (master) plan of the experiment. In this plan, say $\mathcal{D}$, we take into account all the experimenter's suggestions concerning the statistical properties of design and the experimental conditions. It means that plan $\mathcal{D}$ will not be chosen at random.

The basic problem worked out here, is the way of assigning plan $\mathcal{D}$ to a given experimental material. This is defined by the scheme of randomization which describes how to assign the theoretical units of plan $\mathcal{D}$ (with their treatments) to the experimental plots. In our considerations the treatments will not be randomized.

Randomization plays the main role in our considerations. We assume that we have no problems with performing it. For example, in medical experiments (clinical trials) there are many ethical and psychological problems connected with randomization of the patients (see for example, Meier, 1975; Maike and Stanley, 1982; Gehan, 1987).

Suppose that the randomization is performed as described by Nelder (1954) by randomly permuting, for example for block design, blocks within their total area and by randomly permuting units within blocks.

It will be assumed that the treatments under consideration are homogeneous (or additive) in the sense that the variation of the response among the available experimental units does not depend on the treatment received (cf. Kempthorne, 1952; Nelder, 1965, p.168; White, 1975, p.560; Bailey, 1981).

In model building the so called conceptual response of the unit will be used. This response may be obtained in an experiment in which all units receive the same treatment, no matter which one, that will be called the "null" treatment. This conceptual response of a unit will be denoted by m (with some indices). The dot convention will be used to denote the means. Throughout this paper the number of treatments will be denoted by v.

The property of the random variables δ_i, $i = 1,2,\ldots,n$, that they are mutually uncorrelated and independently distributed with expected value equal to a and variance equal to b, will be denoted as $\delta_i \sim (a,b)$.

2. Block designs

2.0. Introduction

Let us assume that the population of experimental units has a nested structure i.e. let it be divided into some number of blocks and then let blocks be divided into units. The number of blocks and block size will be characterized exactly in the cases considered.

Let m_{it} stand for a conceptual response of the t-th unit within the i-th block before randomization (it means that the first subscript denotes the number of block and the second one – the number of unit).

The scheme of randomization can be characterized by some zero-one dummy variables (cf. Kempthorne, 1952). These variables are described exactly for block designs only. The generalization of these variables to a more complicated structure of experimental material and more complicated scheme of randomization is not especially difficult, whilst calculations connected with their statistical properties can be more complicated. In the paper these calculations are omitted but the meaning of the parameters will be given. This is very important in a further inference from an experiment.

2.1. Case A

Let us assume that a population of experimental units (potential population of units) comprises b blocks of sizes $K_1, K_2, \ldots, K_b$ units, respectively.

Let the chosen theoretical plan $\mathcal{D}$ utilize b blocks of sizes $k_1, k_2, \ldots, k_b$, respectively, $k_i \leq K_i$, $i = 1,2,\ldots,b$.

The treatments will be arranged on experimental material in the following way: i) blocks of experimental material are assigned to the blocks of plan $\mathcal{D}$ arbitrarily (not random), ii) units within each experimental block are assigned at random to the units of theoretical plan $\mathcal{D}$; all the b randomizations are independent.

The above described scheme of randomization can be characterized by the following dummy variable: d_{jt}^i is equal to 1, if within the i-th block the t-th unit of experimental material is assigned to the j-th unit of plan $\mathcal{D}$; otherwise d_{jt}^i is equal to 0, (note that blocks are not randomized and hence experimental blocks and theoretical blocks of plan $\mathcal{D}$ can be identified by the same subscript i).

The conceptual response of the j-th unit of the i-th block after randomization is equal to $Y_{ij} = \Sigma_t\, d_{jt}^i\, m_{it}$, while under identity $m_{it} = m_{i\cdot} + (m_{it} - m_{i\cdot})$ is equal to $Y_{ij} = \mu_i + \eta_{ij}$, $j = 1,2,\ldots,k_i$, where $\mu_i = m_{i\cdot}$ denotes the mean of the i-th block, η_{ij} denotes the random effect of the j-th unit in the i-th block i.e., $\eta_{ij} = \Sigma_t\, d_{jt}^i\,(m_{it} - m_{i\cdot})$.

According to the approach described previously, the linear model of observed response can be expressed as:

$$y_{ij(s)} = \mu + \beta_i + \gamma_{j(s)} + \eta_{ij} + \varepsilon_{ij}\,, \quad \mathrm{E}(\,y_{ij(s)}) = \mu + \beta_i + \gamma_{j(s)}, \tag{1}$$
$$i = 1,2,\ldots,b \quad j = 1,2,\ldots,k_i, \quad s = 1,2,\ldots,v,$$

where $\mu = m_{\cdot\cdot}$ denotes the general mean, $\beta_i = m_i - m_{\cdot\cdot}$ – the fixed effect of the i-th block, $\gamma_{j(s)}$ – the effect of the s-th treatment (assigned to the j-th unit). The covariance structure of model (1) has the form

$$\mathrm{Cov}(\,y_{ij(s)}, y_{i'j'(s)}) = \begin{cases} \sigma_{\eta_i}^2 + \sigma^2\,, & i=i',\ j=j', \\ -(K_i - 1)^{-1}\sigma_{\eta_i}^2\,, & i=i',\ j\neq j', \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $\sigma_{\eta_i}^2 = K_i^{-1}\Sigma_t\,(m_{it} - m_{i\cdot})^2$ denotes the unit variance within the i-th block, $i = 1,2,\ldots,b$.

Particular cases.

1° $k_i = K_i$, $i = 1,2,\ldots,b$.

This means that in a considered case the whole population of units within blocks takes part. This case is applicable in a very broad class of agricultural, medical and biological problems. However, there are some problems with statistical analysis of model (1) on (2). There are some problems with estimation and testing hypotheses. Some additional assumptions are needed.

2° $K_i \to \infty$, $\sigma^2_{\eta_i} = \sigma^2_{\eta_i}$, $i = 1,2,\ldots,b$.

This case leads to the following covariance structure of model (1): $\mathrm{Cov}(y_{ij(s)}, y_{i'j'(s)}) = \sigma^2_\eta + \sigma^2$ if $i=i'$, $j=j'$, and otherwise 0.

The obtained linear fixed model is a basic model of the so called intra-block analysis of experiments carried out in block designs (cf. Pearce, 1983).

Before applying this model, the assumptions under which this model is adequate, have to be checked, that is: i) experiment units within blocks compose a random sample drawn from an infinite population of such units, ii) the variances of units within all blocks are the same, iii) complete additivity between treatments and units holds.

Note that these assumptions are very restrictive and usually impossible to assert in practice.

2.2. Case B

Let us assume that a population of units is divided into B blocks of sizes K units.

Let the chosen theoretical plan $\mathcal{D}$ utilize only b blocks of sizes $k_1, k_2,\ldots,k_b$ units, $k_i \le K$, $i = 1,2,\ldots,b$.

The scheme of randomization is as follows: i) from the experimental material we draw at random one block for the i-th block of plan $\mathcal{D}$, ii) in the chosen experimental block we draw at random k_i units, (to the k_i theoretical units of plan $\mathcal{D}$), on which we allocate the treatments. This procedure is repeated for all blocks of plan $\mathcal{D}$.

The scheme of randomization may be characterized by the following random variables: l_{iw} is equal to 1, if i-th block after randomization receives the number w (or if the i-th block of plan $\mathcal{D}$ gets number w of the experimental material); otherwise l_{iw} is equal to 0, f_{jp}^{iw} is equal to 1, if the j-th unit within the i-th block receives the number p in the w-th block (or if the j-th unit of the i-th block of the plan $\mathcal{D}$ gets number p of the w-th block of the experimental material); otherwise f_{jp}^{iw} is equal to 0.

The conceptual response m_{wp} before randomization may be written as follows: $m_{wp} = m_{..} + (m_{w.} - m_{..}) + (m_{wp} - m_{w.})$.

Adopting the similar method of obtaining a model as in case A, the observed response can be expressed as:

$$y_{ij(s)} = \mu + \beta_i + \gamma_{j(s)} + \eta_{ij} + \varepsilon_{ij}\,, \qquad \mathrm{E}(y_{ij(s)}) = \mu + \gamma_{j(s)}\,, \tag{3}$$

$$i = 1,2,\ldots,b, \quad j = 1,2,\ldots,k_i, \quad s = 1,2,\ldots,v,$$

where now β_i denotes the random effect of the i-th block, such that $\mathrm{E}(\beta_i) = 0$, $\mathrm{Var}(\beta_i) = \sigma_\beta^2$, $\mathrm{Cov}(\beta_i, \beta_{i'}) = -(B-1)^{-1}\sigma_\beta^2$, $i,i' = 1,2,\ldots,b$, σ_β^2 – block variance $\sigma_\beta^2 = B^{-1}\Sigma(m_{w.} - m_{..})^2$.

The covariance structure of model (3) has the form

$$\mathrm{Cov}(y_{ij(s)}, y_{i'j'(s)}) = \begin{cases} \sigma_\beta^2 + \sigma_\eta^2 + \sigma^2, & i=i',\ j=j', \\ \sigma_\beta^2 - (K-1)^{-1}\sigma_\eta^2, & i=i',\ j\neq j', \\ -(B-1)^{-1}\sigma_\beta^2, & \text{otherwise}, \end{cases} \tag{4}$$

where σ_η^2 denotes variance of units, $\sigma_\eta^2 = (BK)^{-1}\Sigma\Sigma(m_{wp} - m_{w.})^2$.

Let us note that in the considered case we do not demand units to be uniform within all blocks. By randomization procedure we even expect, to some extent, a possible heterogeneity of these units.

Particular cases.

1° $B = b$, $k_i = k = K$.

This case is typical in agricultural, medical and biological experiments. It means that in our experiments we utilize the whole available population of units.

2° $B \to \infty$, $K \to \infty$.

The variance-covariance structure of model (3) has the form

$$\mathrm{Cov}(y_{ij(s)}) = \begin{cases} \sigma_\beta^2 + \sigma_\eta^2 + \sigma^2, & i=i',\ j=j', \\ \sigma_\beta^2, & i=i',\ j\neq j', \\ 0, & \text{otherwise}. \end{cases} \tag{5}$$

This is equivalent to the assumptions that $\beta_i \sim (0, \sigma_\beta^2)$ and $\eta_{ij} \sim (0, \sigma_\eta^2)$.

Model (3) with (5) is the linear model used in the so-called theory of recovery of inter-block information (cf. Mejza, 1985a).

The theory presented above implies the necessity of fulfilling some assumptions for model (3) with (5) to be adequate, namely: i) the population of potential blocks from which we randomly draw the blocks for an experiment should be infinite, ii) every block contains an infinite number of experimental units from which we draw k_i units, iii) additivity between treatment and unit holds.

For more details the reader is referred to the following articles: Mejza and Mejza, (1989); Caliński and Kageyama, (1991). In the latter paper the situation of Case B is considered where the population of blocks has unequal number of units.

3. Nested block designs

3.0. Introduction

Let us assume that an experimental material has a double nested block structure, i.e., let it be divided into one system of blocks, called superblocks, and then let each superblock be divided into blocks and at last blocks be divided into units. Such a situation is typical for the so-called α-resolvable block designs, i.e., designs in which treatments are assigned to the units so that each one of them is replicated exactly α-times in each superblock.

In our consideration the property of resolvability plays no special role. Naturally, the obtained linear model can be utilized in the analysis of the α-resolvable block designs.

Let m_{ijp} denote the conceptual response of the p-th unit within the j-th block of the i-th superblock (i.e., the first subscript denotes number of superblock, the second – block, and the third – unit).

3.1. Case A

Let us assume that the population of units is divided into R superblocks in such a way that the i-th superblock contains b_i blocks and let the size of the j-th block of the i-th superblock be equal to K_{ij}, $j = 1,2,\ldots,b_i$, $i = 1,2,\ldots,R$.

Let the master plan $\mathcal{D}$ be so that it utilizes k_{ij} from K_{ij} ($k_{ij} \le K_{ij}$) units. We assume also that because of some reasons the superblocks and blocks within superblocks of the plan $\mathcal{D}$ are assigned to the experimental ones in an arbitrary, non random way. The scheme of randomization assigns (randomly) only the theoretical units to the experimental units in every block of each superblock. We randomize separately and independently in every block.

According to the structure of the experimental material and the scheme of randomization the conceptual response can be expressed as follows: $m_{ijp} = m_{ij\cdot} + (m_{ijp} - m_{ij\cdot})$.

Then, the observed response of the (i, j, t) unit can be expressed as:

$$y_{ijt(s)} = \mu + \alpha_i + \beta_{ij} + \gamma_{t(s)} + \eta_{ijt} + \varepsilon_{ijt}\,, \qquad E(y_{ijt(s)}) = \mu + \alpha_i + \beta_{ij} + \gamma_{t(s)}, \tag{6}$$

$$i = 1,2,\ldots,R,\;\; j = 1,2,\ldots,b_i,\;\; t = 1,2,\ldots,k_{ij}\;,$$

where $\mu = \mu_{..}$, $(\mu_{ij} = m_{ij\cdot})$ denotes the mean of experiment, $\alpha_i = \mu_{i\cdot} - \mu_{..}$ denotes the effect of the i-th superblock, $\beta_{ij} = \mu_{ij} - \mu_{i\cdot}$ denotes the effect of the j-th block of the i-th superblock, η_{ijt} and ε_{ijt} denote the unit and measurement errors, respectively.

The dispersion structure of model (6) has a form

$$\mathrm{Cov}(y_{ijt(s)}, y_{i'j't'(s)}) = \begin{cases} \sigma^2_{\eta_{ij}} + \sigma^2, & i=i',\ j=j',\ t=t', \\ -(K_{ij}-1)^{-1}\sigma^2_{\eta_{ij}}, & i=i',\ j=j',\ t\neq t', \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where $\sigma^2_{\eta_{ij}} = K^{-1}_{ij}\Sigma(m_{ijp} - m_{ij\cdot})^2$.

Particular case.

1) $K_{ij} \to \infty$.

The observations can be considered uncorrelated. Some additional assumptions are needed for making analysis of variance possible (cf. particular case of § 3.1.). Many authors adopt in this case the classic fixed linear model, especially to the α-resolvable block designs.

3.2. Case B

Let us assume that experimental units are divided into R superblocks and the i-th superblock consists of B_i blocks of sizes K_i, $i = 1,2,\ldots,R$.

Let us also assume that the chosen master plan $\mathcal{D}$ utilizes R superblocks each having b_i blocks of sizes k_i ($\leq K_i$) respectively.

In the considered case we perform two-step randomization. The superblocks of plan $\mathcal{D}$ are assigned to the experimental ones by natural, not random way, whereas we assign in a random way the experimental blocks to the theoretical ones within each superblock and similarly we treat the units within blocks.

According to the structure of the experimental material and the scheme of randomization the conceptual response can be expressed as follows: $m_{ips} = m_{i\cdot\cdot} + (m_{ip\cdot} - m_{i\cdot\cdot}) + (m_{ips} - m_{ip\cdot})$.

Then the observed response of the (i,j,t) unit can be expressed as:

$$y_{ijt(s)} = \mu + \alpha_i + \beta_{ij} + \gamma_{t(s)} + \eta_{ijt} + \varepsilon_{ijt}\,, \qquad E(y_{ijt(s)}) = \mu + \alpha_i + \gamma_{t(s)}, \tag{8}$$

$$i = 1,2,\ldots,R,\;\; j = 1,2,\ldots,b_i,\;\; t = 1,2,\ldots,k_i\;,$$

where now β_{ij} denotes the random effect of the j-th block of the i-th superblock.

The covariance structure of model (8) has the form

$$\mathrm{Cov}(y_{ijt(s)}, y_{i'j't'(s)}) = \begin{cases} \sigma^2_{\beta_i} + \sigma^2_{\eta_i} + \sigma^2, & i=i',\ j=j',\ t=t', \\ \sigma^2_{\beta_i} - (K_i - 1)^{-1}\sigma^2_{\eta_i} & i=i',\ j=j',\ t\neq t', \\ -(B_i - 1)^{-1}\sigma^2_{\beta_i}, & i=i',\ j\neq j', \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

where $\sigma^2_{\beta_i} = B_i^{-1}\Sigma(m_{ip\cdot} - m_{i\cdot\cdot})^2$ denotes the block variance within the i-th superblock whereas $\sigma^2_{\eta_i} = (B_i K_i)^{-1}\Sigma\Sigma(m_{ips} - m_{ip\cdot})^2$ denotes the unit variances within the i-th superblocks.

Particular case

1) $B_i \to \infty$, $K_i \to \infty$, $\sigma^2_{\beta_i} = \sigma^2_{\beta}$, $\sigma^2_{\eta_i} = \sigma^2_{\eta}$.

Such a situation is often used in practical applications. It is worth noting here that assumptions about the equality of the variances are extremely restrictive.

3.3. Case C

Let us assume that experimental units are divided into R superblocks and each of them is divided into B blocks of K units.

Let the chosen master plan $\mathcal{D}$ utilize r superblocks where the i-th superblock contains b_i blocks of sizes $k_{i1}, k_{i2}, \ldots, k_{ib_i}$, $(\leq K)$, respectively.

We assume that in the considered case the three-step randomization is performed, i.e., the randomization of the superblocks, randomization of blocks within the superblocks and randomization of the units within blocks.

Now the conceptual response can be expressed as follows: $m_{hps} = m_{\cdots} + (m_{h\cdot\cdot} - m_{\cdots}) + (m_{hp\cdot} - m_{h\cdot\cdot}) + (m_{hps} - m_{hp\cdot})$.

Then the observed response of the (i,j,t) unit can be expressed as:

$$y_{ijt(s)} = \mu + \alpha_i + \beta_{ij} + \gamma_{t(s)} + \eta_{ijt} + \varepsilon_{ijt}, \quad \mathrm{E}(y_{ijt(s)}) = \mu + \gamma_{t(s)}, \tag{10}$$

$$i = 1,2,\ldots,r, \quad j = 1,2,\ldots,b_i, \quad t = 1,2,\ldots,k_{ij},$$

where now additionally α_i denotes the random effect of the i-th superblock.

The covariance structure of model (10) has the form

$$\mathrm{Cov}(y_{ijt(s)}, y_{i'j't'(s)}) = \begin{cases} \sigma^2_{\alpha} + \sigma^2_{\beta} + \sigma^2_{\eta} + \sigma^2, & i=i',\ j=j',\ t=t', \\ \sigma^2_{\alpha} + \sigma^2_{\beta} - (K-1)^{-1}\sigma^2_{\eta}, & i=i',\ j=j',\ t\neq t', \\ \sigma^2_{\alpha} - (B-1)^{-1}\sigma^2_{\beta}, & i=i',\ j\neq j', \\ -(R-1)^{-1}\sigma^2_{\alpha}, & i\neq i', \end{cases} \tag{11}$$

where:

$\sigma_\alpha^2 = R^{-1}\Sigma(m_{h\cdot\cdot} - m_{\cdot\cdot\cdot})^2$ denotes the superblock variance,

$\sigma_\beta^2 = (RB)^{-1}\Sigma\Sigma(m_{hp\cdot} - m_{h\cdot\cdot})^2$ denotes the block variance, whereas

$\sigma_\eta^2 = (RBK)^{-1}\Sigma\Sigma\Sigma(m_{hps} - m_{hp\cdot})^2$ denotes the unit variance.

Particular cases

1) $R\to\infty$, $B\to\infty$, $K\to\infty$.

Such a situation is often used in practical applications also. It is the classic linear mixed model applied for α-resolvable block design. An equivalent situation is that in which it is assumed that $\alpha_i \sim (0,\sigma_\alpha^2)$, $\beta_{ij} \sim (0,\sigma_\beta^2)$ and $\eta_{ijt} \sim (0,\sigma_\eta^2)$.

2) $r = R$, $b_i = b = B$, $k_{i1} = k_{i2} = \ldots = k_{ib_i} = k$.

The whole population of units takes part in an experiment.

For more details the reader is referred to the following papers: John (1987), Mejza S. (1989), Mejza and Mejza (1989).

4. Row-column designs

Let us assume that experimental material is divided into orthogonally disposed blocking systems, one called rows and the other called columns, each intersection of row and column constitutes a unit.

Let K_1 and K_2 denote the number of rows and the number of columns and let our master plan $\mathcal{D}$ utilize k_1 and k_2 from them respectively. By m_{pf} we denote a conceptual response of a unit being intersection of the p-th row and the f-th column.

The conceptual response can be expressed as:

$$m_{pf} = m_{\cdot\cdot} + (m_{p\cdot} - m_{\cdot\cdot}) + (m_{\cdot f} - m_{\cdot\cdot}) + (m_{pf} - m_{p\cdot} - m_{\cdot f} + m_{\cdot\cdot}).$$

Applying the above equality and using two independent randomizations, i.e., randomization of rows and randomization of columns, the observed response can be written as:

$$y_{ij(s)} = \mu + \gamma(s) + \alpha_i + \beta_j + \eta_{ij} + \varepsilon_{ij}, \quad E(y_{ij(s)}) = \mu + \gamma(s) \tag{12}$$

$$i = 1,2,\ldots,k_1, \quad j = 1,2,\ldots,k_2,$$

where α_i (β_j) denotes the random effect of the i-th (j-th) row (column), η_{ij} and ε_{ij} denote the unit errors and technical errors, respectively.

The dispersion structure of model (12) is of a form

$$\mathrm{Cov}(y_{ij(s)}, y_{i'j'(s)}) = \tag{13}$$

$$= \begin{cases} \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\eta^2 + \sigma^2, & i=i',\ j=j', \\ \sigma_\alpha^2 - (K_2-1)^{-1}\sigma_\beta^2 - (K_2-1)^{-1}\sigma_\eta^2, & i=i',\ j\neq j', \\ -(K_1-1)^{-1}\sigma_\alpha^2 + \sigma_\beta^2 - (K_1-1)^{-1}\sigma_\eta^2, & i\neq i',\ j=j', \\ -(K_1-1)^{-1}\sigma_\alpha^2 - (K_2-1)^{-1}\sigma_\beta^2 + (K_1-1)^{-1}(K_2-1)^{-1}\sigma_\eta^2, & i\neq i',\ j\neq j', \end{cases}$$

where $\sigma_\alpha^2 = K_1^{-1}\Sigma(m_{p\cdot} - m_{\cdot\cdot})^2$ denotes the row variance, $\sigma_\beta^2 = K_2^{-1}\Sigma(m_{\cdot f} - m_{\cdot\cdot})^2$ denotes the column variance, $\sigma_\eta^2 = K_1^{-1}K_2^{-1}\Sigma\Sigma(m_{pf} - m_{p\cdot} - m_{\cdot f} + m_{\cdot\cdot})^2$ denotes the unit variance.

Particular cases

Two particular cases are of great interest.

1) $K_1 \to \infty$, $K_2 \to \infty$,

In this case the observations can be treated as uncorrelated. The obtained linear model is classic when we consider linear mixed model for row-column designs. Often such assumptions are written in the form $\alpha_i \sim (0, \sigma_a^2)$, $\beta_j \sim (0, \sigma_b^2)$ and $\eta_{ij} \sim (0, \sigma_\eta^2)$.

2) $k_1 = K_1$, $k_2 = K_2$.

Such a situation we usually have in practical applications. Let us note that in many textbooks the above randomization is recommended but then it is not seriously taken into account. For more details the reader is referred to John (1987).

5. Block designs with nested rows and columns

Let us assume that a population of experimental units (set of potential units) is stratified into R superblocks so that each superblock is additionally divided into two orthogonally arranged blocking systems, one called rows and the other called columns, each intersection of a row and column constituting a unit.

Let K_1 denote the number of rows, and K_2 – the number of columns in our population of units. Let the chosen master plan $\mathcal{D}$ utilize $r\ (\leq R)$ superblocks, $k_1\ (\leq K_1)$ rows and $k_2\ (\leq K_2)$ columns and let both of the numbers k_1 and k_2 be the same in all superblocks.

In the paper we consider a three-step randomization, i.e., randomization of superblocks, randomization of rows (columns) and randomization of columns (rows) within each superblock.

Because of the structure of the experimental units, the conceptual response m_{hps} of the unit at the intersection of the p-th row and the s-th column within the h-th superblock can be suitably expressed in the form

$$m_{hps} = m_{\cdot\cdot\cdot} + (m_{h\cdot\cdot} - m_{\cdot\cdot\cdot}) + (m_{hp\cdot} - m_{h\cdot\cdot}) + (m_{h\cdot s} - m_{h\cdot\cdot}) + (m_{hps} - m_{hp\cdot} - m_{h\cdot s} + m_{h\cdot\cdot}),$$
$$h = 1,2,\ldots,R, \quad p = 1,2,\ldots,K_1, \quad s = 1,2,\ldots,K_2.$$

Then, the observed response of the (j,t)-th unit of the i-th superblock, can be written as:

$$y_{ijt(s)} = \mu + \gamma(s) + \rho_i + \eta_{ij} + \theta_{it} + \varphi_{ijt} + \varepsilon_{ijt}, \quad \mathrm{E}(y_{ijt(s)}) = \mu + \gamma(s) \tag{14}$$
$$i = 1,2,\ldots,r, \quad j = 1,2,\ldots,k_1, \quad t = 1,2,\ldots,k_2,$$

where $\mu = m_{\cdot\cdot\cdot}$ denotes the mean, ρ_i – denotes the effect of the i-th superblock, η_{ij} and θ_{it} stand for the effect of the j-th row and the t-th column within the i-th superblock, respectively, and where φ_{ijt} stands for the effect (error) of the (i,j,t)-th unit.

The dispersion structure of model (14) is as follows:

$$\mathrm{Cov}(y_{ijt(s)}, y_{i'j't'(s)}) = \tag{15}$$
$$= \begin{cases} \sigma_\rho^2 + \sigma_\eta^2 + \sigma_\theta^2 + \sigma_\varphi^2 + \sigma^2, & i=i',\ j=j',\ t=t', \\ \sigma_\rho^2 + \sigma_\eta^2 - (K_2-1)^{-1}\sigma_\theta^2 - (K_2-1)^{-1}\sigma_\varphi^2, & i=i',\ j=j',\ t\neq t', \\ \sigma_\rho^2 - (K_1-1)^{-1}\sigma_\eta^2 + \sigma_\theta^2 - (K_1-1)^{-1}\sigma_\varphi^2, & i=i',\ j\neq j',\ t=t', \\ \sigma_\rho^2 - (K_1-1)^{-1}\sigma_\eta^2 - (K_2-1)^{-1}\sigma_\theta^2 + (K_1-1)^{-1}(K_2-1)^{-1}\sigma_\varphi^2, & i=i',\ j\neq j',\ t\neq t', \\ -(R-1)^{-1}\sigma_\rho^2, & \text{otherwise,} \end{cases}$$

where variance components denote respectively:

$\sigma_\rho^2 = R^{-1}\Sigma(m_{h\cdot\cdot} - m_{\cdot\cdot\cdot})^2$ – the superblock variance,

$\sigma_\eta^2 = (RK_1)^{-1}\Sigma\Sigma(m_{hp\cdot} - m_{h\cdot\cdot})^2$ – the row variance,

$\sigma_\theta^2 = (RK_2)^{-1}\Sigma\Sigma(m_{h\cdot s} - m_{h\cdot\cdot})^2$ – the column variance,

$\sigma_\varphi^2 = (RK_1K_2)^{-1}\Sigma\Sigma\Sigma(m_{hps} - m_{hp\cdot} - m_{h\cdot s} + m_{h\cdot\cdot})^2$ – the unit (error) variance.

Particular cases

Two particular cases, as usual, are of great interest.

1) $R \to \infty$, $K_1 \to \infty$, $K_2 \to \infty$.

In this case the observations can be treated as uncorrelated. The obtained linear model is classic when we consider a linear mixed model for a block design with nested rows and columns. So the above assumptions are written in the form $\rho_i \sim (0, \sigma_a^2)$, $\eta_{ij} \sim (0, \sigma_\eta^2)$, $\theta_{it} \sim (0, \sigma_\theta^2)$ and $\varphi_{ijt} \sim (0, \sigma_\varphi^2)$;

2) $r = R$, $k_1 = K_1$, $k_2 = K_2$.

In this situation the whole population of experimental units takes part in an experiment.

For more details the reader is referred to Mejza and Mejza (1994).

6. Two-factor designs

6.0. Introduction

Let us consider a two-factor experiment in which factor A (or set of factors) occurs on s levels $A_1, A_2, \ldots, A_s$ and second factor, B, occurs on t levels $B_1, B_2, \ldots, B_t$. Moreover, by treatment we will mean the treatment combination $A_f B_g$, $f = 1,2,\ldots,s$, $g = 1,2,\ldots,t$, while by the effect of the i-th treatment we will mean

$$\tau_i = \alpha_f + \beta_g + (\alpha\beta)_{fg}, \quad i = (f-1)t + g, \; f = 1,2,\ldots,s, \; g = 1,2,\ldots,t, \tag{16}$$

where α_f denotes the effect of the f-th level of the factor A, β_g denotes the effect of the g-th level of the factor B and $(\alpha\beta)_{fg}$ stands for the interaction effect.

Let us note that the structure of random terms of the linear model does not depend on the occurrence of treatments on the units. It means that our considerations cover the three possible occurrence cases of the factor levels on the units, i.e., incompleteness, completeness, and over completeness. Distinguishing the above situations is not necessary in model building for observations. It is however very important in the further analysis of data. It means that the problem should be taken into account in the master plan $\mathcal{D}$ choice.

In this section of the paper only the situation in which a whole population of units takes part in an experiment will be considered. It seems that the proposed scheme of randomization is the most useful in applications.

Let $v=st$ denote the number of treatments (treatment combinations) and let b denote the number of blocks in the design.

6.1. Classic factorial designs

By the classic factorial design we mean design in which all factor levels combinations are treated as typical treatments and no special distinguishing of some combination is taken into account during the randomization. Then applying the scheme of randomization as described in sections 2 and 3, it is easy to obtain the linear model for the observed response.

6.2. Split-plot designs

Let us assume that a population of units is divided into b blocks in such a way that each block contains k whole-units and each whole-unit contains m sub-units. It means that in our design there are $n = bkm$ units.

Let levels of the factor A be the whole-unit treatments while levels of the factor B be the sub-unit treatments.

A traditional split-plot design is such that the levels of factor A are arranged on whole-units of a randomized complete block design and the levels of factor B are arranged on the sub-units of a different randomized complete block design within each level of A, provided the whole-units are treated as blocks. It means that the number of the whole-units within each block should be equal to s (i.e. $k=s$) and the number of sub-units within each whole-unit should be equal to t (i.e. $m=t$).

In this paper we consider a general situation of the treatment combination occurrence on the units. It means that the design can be called (with respect to occurrence of treatments on the units) either incomplete, complete or over complete.

Our model building is based on the three step randomization, i.e. randomization of blocks, randomization of whole-units within each block, randomization of sub-units within each whole-unit of each block.

Let m_{wsq} denote the conceptual response of the q-th sub-unit in the s-th whole-unit of the w-th block.

Then applying the equality $m_{wsq} = m_{\cdots} + (m_{w\cdot\cdot} - m_{\cdots}) + (m_{ws\cdot} - m_{w\cdot\cdot}) + (m_{wsq} - m_{ws\cdot})$ and adopting the same approach as in previous sections we obtain the linear model for the observed response in the form

$$y_{rhj(i)} = \mu + \tau(i) + \rho_r + \eta_{rh} + \varphi_{rhj} + \varepsilon_{rhj}, \qquad \mathrm{E}(y_{rhj(i)}) = \mu + \tau(i), \tag{17}$$

$$r = 1,2,\ldots,b,\ h = 1,2,\ldots,k,\ j = 1,2,\ldots,m,\ i = 1,2,\ldots,v,$$

where $\mu = m_{\cdots}$ denotes the mean, $\tau(i)$ denotes the effect of the i-th treatment combination occurring on the (r,h,j)-th unit, ρ_r denotes the effect of the r-th block, η_{rh} and φ_{rhj} stand for the effect (error) of the (r,h)-th whole-unit and (r,h,j)-th sub-unit, respectively.

The dispersion structure of model (17) is as follows:

$$\mathrm{Cov}(y_{rhj(i)}, y_{r'h'j'(i)}) = \begin{cases} \sigma_\rho^2 + \sigma_\eta^2 + \sigma_\varphi^2 + \sigma^2, & r=r',\ h=h',\ j=j', \\ \sigma_\rho^2 + \sigma_\eta^2 - (m-1)^{-1}\sigma_\varphi^2, & r=r',\ h=h',\ j\neq j', \\ \sigma_\rho^2 - (k-1)^{-1}\sigma_\eta^2, & r=r',\ h\neq h', \\ -(b-1)^{-1}\sigma_\rho^2, & r\neq r', \end{cases} \tag{18}$$

where variance components denote respectively:

$\sigma_\rho^2 = b^{-1}\Sigma(m_{w\cdot\cdot} - m_{\dots})^2$ – the block variance,

$\sigma_\eta^2 = (bk)^{-1}\Sigma\Sigma(m_{ws\cdot} - m_{w\cdot\cdot})^2$ – the whole-unit (error) variance,

$\sigma_\varphi^2 = (bkm)^{-1}\Sigma\Sigma\Sigma(m_{wsq} - m_{ws\cdot})^2$ – the sub-unit (error) variance.

For more details the reader is referred to the following articles: Mejza and Mejza (1984), Mejza (1985a, 1987).

6.3. Split-block designs

Let the experimental material be divided into b blocks and let each block be additionally divided into two orthogonally disposed blocking systems, one called rows and the other called columns. Each intersection of a row and a column constitutes an (experimental) unit. Let k_1 denote the number of rows, k_2 the number of columns and let both of them be the same in all blocks. Furthermore, let us assume that the levels of factor A occur on the rows whereas the levels of factor B occur on the columns. Let as usual $\mathcal{D}$ denote the plan of our experiment.

In a classic (i.e. complete) case we have $k_1 = s$ and $k_2 = t$ and each combination of factors occurs one time in each block. Let m_{wsq} denote the conceptual response of the unit being the intersection of the s-th row and q-th column in the w-th block. Because of the experimental material structure we have equality

$$m_{wsq} = m_{\dots} + (m_{w\cdot\cdot} - m_{\dots}) + (m_{ws\cdot} - m_{w\cdot\cdot}) + (m_{w\cdot q} - m_{w\cdot\cdot}) + (m_{wsq} - m_{ws\cdot} - m_{w\cdot q} + m_{w\cdot\cdot}),$$

which implies the linear model for the observed response as:

$$y_{rhj(i)} = \mu + \tau(i) + \rho_r + \eta_{rh} + \theta_{rj} + \varphi_{rhj} + \varepsilon_{rhj}, \quad \mathrm{E}(y_{rhj(i)}) = \mu + \tau(i), \tag{19}$$

$$r = 1,2,\dots,b,\ h = 1,2,\dots,k_1,\ j = 1,2,\dots,k_2,\ i = 1,2,\dots,v,$$

where $\mu = m_{\dots}$ denotes the mean, ρ_r denotes the effect of the r-th block, η_{rh} (θ_{rj}) denote the random effects of the h-th row (j-th column) within the r-th block and φ_{rhj} stands for the effect (error) of the (r,h,j)-th sub-unit.

The dispersion structure of the of model (19) is as follows:

$$\mathrm{Cov}(y_{rhj(i)}, y_{r'h'j'(i)}) = \tag{20}$$

$$\begin{cases} \sigma_\rho^2 + \sigma_\eta^2 + \sigma_\theta^2 + \sigma_\varphi^2 + \sigma^2, & r=r',\ h=h',\ j=j', \\ \sigma_\rho^2 + \sigma_\eta^2 - (k_2-1)^{-1}\sigma_\theta^2 - (k_2-1)^{-1}\sigma_\varphi^2, & r=r',\ h=h',\ j\neq j', \\ \sigma_\rho^2 - (k_1-1)^{-1}\sigma_\eta^2 + \sigma_\theta^2 - (k_1-1)^{-1}\sigma_\varphi^2, & r=r',\ h\neq h',\ j=j', \\ \sigma_\rho^2 - (k_1-1)^{-1}\sigma_\eta^2 - (k_2-1)^{-1}\sigma_\theta^2 + (k_1-1)^{-1}(k_2-1)^{-1}\sigma_\varphi^2, & r=r',\ h\neq h',\ j\neq j', \\ -(b-1)^{-1}\sigma_\rho^2, & \text{otherwise,} \end{cases}$$

where variance components denote, respectively:

$\sigma_\rho^2 = b^{-1}\Sigma(m_{w\cdot\cdot} - m_{\dots})^2$ – the block variance,

$\sigma_\eta^2 = (bk_1)^{-1}\Sigma\Sigma(m_{ws\cdot} - m_{w\cdot\cdot})^2$ – the row variance,

$\sigma_\theta^2 = (bk_2)^{-1}\Sigma\Sigma(m_{w\cdot q} - m_{w\cdot\cdot})^2$ – the column variance,

$\sigma_\varphi^2 = (bk_1k_2)^{-1}\Sigma\Sigma\Sigma(m_{wsq} - m_{ws\cdot} - m_{w\cdot q} + m_{w\cdot\cdot})^2$ – the unit (error) variance.

For more details the reader is referred to the following articles: Mejza I. (1989), Bhargava and Shah (1975).

6.4. Split-plot designs – whole-unit treatments in a row-column design

Let an experimental material be divided into k_1 rows and k_2 columns as in row-column designs. Further, let all first order units (whole-units) formed at the crosses of rows and columns be divided into k_3 secondary units (sub-units). A two-factor experiment is designed on this experimental material in such a way that levels of the first factor (A) and second factor (B) are distributed on whole-units and sub-units, respectively. This experimental design is referred to as a row-column design with split units. There are practically no restrictions on the distribution of factor A and B levels on the respective units in the considered design.

The assignment of the master plan $\mathcal{D}$ to a given experimental material is as follows: to each row and each column of the plan $\mathcal{D}$ a row and column of the experimental material are randomly attributed. According to the plan, it is known which one out of s levels of the factor A occurs on the whole-unit lying at the crossing of the rows and columns. Likewise, from the plan $\mathcal{D}$ it is known which k_3 out of t levels of factor B occur on a given whole-unit. These levels are now randomly allocated among its k_3 sub-units. All of the random assignments are independent.

Because of the experimental material structure we have

$$m_{wsq} = m_{\dots} + (m_{w\dots} - m_{\dots}) + (m_{.s.} - m_{\dots}) + (m_{ws.} - m_{.s.} - m_{w\dots} + m_{\dots}) + (m_{wsq} - m_{ws.})$$

which adopted in our approach generates the linear model for the observed response in the form

$$y_{rhj(i)} = \mu + \tau(i) + \rho_r + \eta_h + \theta_{rh} + \varphi_{rhj} + \varepsilon_{rhj}, \quad \mathrm{E}(y_{rhj(i)}) = \mu + \tau(i), \tag{21}$$

$$r = 1,2,\dots,k_1, \; h = 1,2,\dots,k_2, \; j = 1,2,\dots,k_3, \; i = 1,2,\dots,v,$$

where $\mu = m_{\dots}$ denotes the mean, ρ_r denotes the effect of the r-th row, η_h denotes the effect of the h-th column, θ_{rh} denotes the effect (error) of the (r,h) whole-unit and φ_{rhj} stands for the effect (error) of the (r,h,j)-th sub-unit.

The dispersion structure of model (21) is as follows:

$$\mathrm{Cov}(y_{rhj(i)}, y_{r'h'j'(i)}) = \tag{22}$$

$$\begin{cases} \sigma_\rho^2 + \sigma_\eta^2 + \sigma_\theta^2 + \sigma_\varphi^2 + \sigma^2, & r=r',\ h=h',\ j=j', \\ \sigma_\rho^2 + \sigma_\eta^2 + \sigma_\theta^2 - (k_3-1)^{-1}\sigma_\varphi^2, & r=r',\ h=h',\ j\neq j', \\ \sigma_\rho^2 - (k_2-1)^{-1}\sigma_\eta^2 - (k_2-1)^{-1}\sigma_\theta^2, & r=r',\ h\neq h', \\ -(k_1-1)^{-1}\sigma_\rho^2 + \sigma_\eta^2 - (k_1-1)^{-1}\sigma_\theta^2, & r\neq r',\ h=h', \\ -(k_1-1)^{-1}\sigma_\rho^2 - (k_2-1)^{-1}\sigma_\eta^2 + (k_1-1)^{-1}(k_2-1)^{-1}\sigma_\theta^2, & r\neq r',\ h\neq h', \end{cases}$$

where variance components denote respectively:

$\sigma_\rho^2 = k_1^{-1}\Sigma(m_{w..} - m_{...})^2$ – the row variance,

$\sigma_\eta^2 = k_2^{-1}\Sigma\Sigma(m_{.s.} - m_{...})^2$ – the column variance,

$\sigma_\theta^2 = (k_1k_2)^{-1}\Sigma\Sigma(m_{ws.} - m_{w..} - m_{.s.} + m_{...})^2$ – the whole-unit variance,

$\sigma_\varphi^2 = (k_1k_2k_3)^{-1}\Sigma\Sigma\Sigma(m_{wsq} - m_{ws.})^2$ – the unit (error) variance.

For more details the reader is referred to Kachlicka and Mejza (1990).

6.5. *Repeated row-column designs with split-units*

Let, in this paragraph, the experimental material with the structure described in § 6.4 be called a superblock. We assume here that the experimental material contains k_0 superblocks with k_1 rows, k_2 columns and k_3 sub-units within each whole-unit. Hence, an equal structure of the experimental material in each superblock is required. Otherwise, no constraints are laid upon the distribution of factors A and B levels in superblocks, which may be either equal or not. The experimental design obtained in this way will be called a repeated row-column design with split units. The allocation of factors A and B levels is performed according to plan $\mathcal{D}$. The way of assigning the theoretical plan $\mathcal{D}$ to a given experimental material is defined by the randomization scheme of the experiment. In this paragraph the randomization proceeds in four stages. In the first stage superblocks are randomized, in the second and third – rows and columns, respectively. The distribution of factor A levels on the whole-plots is obtained as a result of these randomizations. Thereafter the distribution of factor B levels on sub-units (independently on each whole-plot) is obtained in the fourth stage of randomization.

Let m_{wsqp} denote conceptual response of the p-th sub-unit in the (s,q) whole-unit being intersection of the s-th row and q-th column in the w-th superblock. Then, the conceptual response can be expressed as:

$$m_{wsqp} = m_{....} + (m_{w...} - m_{....}) + (m_{ws..} - m_{w...}) + (m_{w.q.} - m_{w...}) +$$
$$(m_{wsq.} - m_{ws..} - m_{w.q.} + m_{w...}) + (m_{wspq} - m_{wsq.}).$$

Applying the approach considered in this paper we obtain the linear model for the observed response in the form

$$y_{rhjl(i)} = \mu + \tau(i) + \rho_r + \eta_{rh} + \theta_{rj} + \varphi_{rhj} + \omega_{rhjl} + \varepsilon_{rhjl}, \qquad E(y_{rhjl(i)}) = \mu + \tau(i),$$

$$r=1,2,\ldots,k_0, \quad h=1,2,\ldots,k_1, \quad j=1,2,\ldots,k_2, \quad l=1,2,\ldots,k_3, \quad i=1,2,\ldots,v, \tag{23}$$

where $\mu = m_{...}$ denotes the mean, ρ_r denotes the effect of the r-th superblock, η_{rh} (θ_{rj}) denote the effect of the h-th row (j-th column) within the r-th superblock, φ_{rhj} stands for the effect (error) of the (h,j)-th whole-unit within the r-th superblocks and ω_{rhjl} denotes the sub-unit effect (error).

The dispersion structure of the of model (23) is as follows:

$$\mathrm{Cov}(y_{rhjl(i)}, y_{r'h'j'l'(i)}) = \tag{24}$$

$$\begin{cases} \sigma_\rho^2 + \sigma_\eta^2 + \sigma_\theta^2 + \sigma_\varphi^2 + \sigma_\omega^2 + \sigma^2, & r=r',\ h=h',\ j=j',\ l=l', \\ \sigma_\rho^2 + \sigma_\eta^2 + \sigma_\theta^2 + \sigma_\varphi^2 - (k_4-1)^{-1}\sigma_\omega^2, & r=r',\ h=h',\ j=j',\ l\neq l', \\ \sigma_\rho^2 + \sigma_\eta^2 - (k_2-1)^{-1}\sigma_\theta^2 - (k_2-1)^{-1}\sigma_\varphi^2, & r=r',\ h=h',\ j\neq j', \\ \sigma_\rho^2 - (k_1-1)^{-1}\sigma_\eta^2 + \sigma_\theta^2 + (k_1-1)^{-1}\sigma_\varphi^2, & r=r',\ h\neq h',\ j=j', \\ \sigma_\rho^2 - (k_1-1)^{-1}\sigma_\eta^2 - (k_2-1)^{-1}\sigma_\theta^2 + (k_1-1)^{-1}(k_2-1)^{-1}\sigma_\varphi^2, & r=r',\ h\neq h',\ j\neq j', \\ -(k_0-1)^{-1}\sigma_\rho^2, & r\neq r', \end{cases}$$

where variance components denote respectively:

$\sigma_\rho^2 = k_0^{-1}\Sigma(m_{w\cdots} - m_{\cdots\cdot})^2$ – the superblock variance,

$\sigma_\eta^2 = (k_0k_1)^{-1}\Sigma\Sigma(m_{ws\cdot\cdot} - m_{w\cdots})^2$ – the rows within superblocks variance,

$\sigma_\theta^2 = (k_0k_2)^{-1}\Sigma\Sigma(m_{w\cdot q\cdot} - m_{w\cdots})^2$ – the columns within superblocks variance,

$\sigma_\varphi^2 = (k_0k_1k_2)^{-1}\Sigma\Sigma\Sigma(m_{wsq\cdot} - m_{ws\cdot\cdot} - m_{w\cdot q\cdot} + m_{\cdots\cdot})^2$ – the whole-unit (error) variance and $\sigma_\omega^2 = (k_0k_1k_2k_3)^{-1}\Sigma\Sigma\Sigma\Sigma(m_{wsqp} - m_{wsq\cdot})^2$ – the sub-unit (error) variance.

For more details the reader is referred to Kachlicka and Mejza (1991).

7. Concluding remarks

The paper presents one of the possible approaches to model building. In this approach the main role is played by the structure of the experimental units and the scheme of the applied randomization. We do not discuss the importance and the need of randomization. These problems are considered in many papers, for example, Greenberg (1951), Harville (1975), Kempthorne (1977), Bailey (1981), Thornet (1982), Folks (1984), Gehan (1987), Caliński and Kageyama (1991), Kala (1989, 1990, 1991).

One problem is a way of obtaining the linear model but another problem is its statistical analysis. For some linear models obtained there are no appropriate statistical methods. Hence, in the modelling of experiments the problem of further statistical analysis (estimation and inference) should be taken into account.

Acknowledgements

The paper was prepared while I was the C.N.R. research fellow at the Perugia University. Thanks are due to these institutions for their support.

REFERENCES

Bailey, R.A. (1981). A unified approach to design of experiments. *J. R. Statist. Soc. A.*, **144**, 214-223.

Bhargava, R.P., Shah, K.R. (1975). Analysis of some mixed models for block and split-plot designs. *Ann. Inst. Statist. Math.* **27**, 365-375.

Caliński, T. and Kageyama, S. (1991). A randomization theory of intra-block and inter-block estimation. *Listy Biometryczne - Biometrical Letters* **28**, 97-122.

Folks, J.L. (1984). Use of randomization in experimental research. In: *Experimental Design, Statistical Models, and Genetic Statistics. Essays in Honor of O.Kempthorne* (Ed. H. Hinkelmann). Marcel Dekker, Inc., New York

Gehan, E.A. (1987). The pros and cons of randomized cancer clinical trials. *Rivista di Statistica Applicata.* **20**, 71-82.

Greenberg, B.G. (1961). Why randomize ? *Biometrics* **7**, 309-322.

Harville, D.A. (1975). Experimental randomization: Who needs it? *The Amer. Statist.* **29**, 27-31.

John, J.A. (1987). *Cyclic designs.* Chapman and Hill, London.

Kachlicka, D., Mejza, S. (1990). Incomplete split-plot experiments - whole-plot treatments in a row-column design. *Comput. Statistics Data Anal.* **9**, 135-146.

Kachlicka, D., Mejza, S. (1991). Repeated row-column designs with split-units. Technical Report, Agricultural University, Poznań, Poland.

Kala, R. (1989). Elementy teorii randomizacji. I. Próba zrandomizowana. *Listy Biometryczne - Biometrical Letters* **26**, 41-55.

Kala, R. (1990). Elementy teorii randomizacji. II. Modelowanie doświadczeń prostych. *Listy Biometryczne - Biometrical Letters* **27**, 31-45.

Kala, R. (1991). Elementy teorii randomizacji. III. Randomizacja w układach blokowych. *Listy Biometryczne - Biometrical Letters* **28**, 3-23.

Kempthorne, O. (1952). *The Design and Analysis of Experiments.* Wiley, New York.

Kempthorne, O. (1977). Why randomize ? *J. Statist. Plan. and Infer.* **1**, 1-25.

Maike, V. and Stanley, K.S. (1982). *Statistics in Medical Research.* Wiley, New York.

Meier, P. (1975). Statistics and medical experimentation. *Biometrics* **31**, 511-529.

Mejza, I. (1989). A note on the analysis of split-block designs. In: *Proc. of the "Statistical Data Analysis -87"*, Bulgarian Academy of Science, 139-144.

Mejza, S. (1985a). On testing hypotheses in a mixed linear model for incomplete block designs. *Scand. J.Statist.* **12**, 241-247.

Mejza, S. (1985b). A split-plot design with wholeplot treatments in an incomplete block design. In: *Lecture Notes in Statistics. Vol. 35. Linear Statistical Inference.* Caliński, T. and Klonecki, W., Eds., Springer Verlag, New York, 211-222.

Mejza, S. (1987). Experiments in incomplete split-plot designs. *Proc. Second International Tampere Conference in Statistics.*, Pukkila, T. and Puntanen, S., Eds., University of Tampere, Tampere, 575-584.

Mejza, S. (1989). Experiments in resolvable block designs. In: *Proc. of the "Statistical Data Analysis-87"*, Bulgarian Academy of Science, 133-138.

Mejza, I., Mejza, S. (1984). Incomplete split-plot designs. *Statist. Probab. Lett.* **2**., 327-332.

Mejza, I., Mejza, S. (1989). A note on model building in block designs. *Biuletyn Oceny Odmian* **21-22**, 187-201.

Mejza, I., Mejza, S. (1994). Model builing and analysis for block designs with nested rows and columns. *Biom. J.* **36**, 327-340.

Nelder, J.N. (1954). The interpretation of negative components of variance. *Biometrika* **41**, 544-548.

Nelder, J.N. (1965). The analysis of experiment with orthogonal block structure. *Proc. R.Soc. Lond. A.* **283**, 147-178.

Neyman, J., Iwaszkiewicz, K. and Kołodziejczyk, St. (1935). Statistical problems in agricultural experimentation. *J.Roy. Statist. Soc. Suppl.* **2**, 107-154.

Pearce, S.C. (1983). *The Agricultural Field Experiment.* Wiley, New York.

Thornett, M.L. (1982). The role of randomization in model-based inference. *Australian J. Statist.* **24**, 137-145.

White, R.F. (1975). Randomization and the analysis of variance. *Biometrics* **31**, 552-571.

Received 26 July 1993; revised 9 September 1994

O modelowaniu doświadczeń w naukach przyrodniczych

Streszczenie

W pracy zajmujemy się problemami modelowania doświadczeń jednoczynnikowych i dwuczynnikowych zakładanych w układach blokowych. W przypadku układów jednoczynnikowych o strukturze blokowej, rozważania dotyczą takich układów jak: układ o blokach niekompletnych, zagnieżdżony układ o blokach niekompletnych, układ kolumnowo-wierszowy, układ o blokach niekompletnych z zagnieżdżonymi wierszami i kolumnami. W przypadku doświadczeń dwuczynnikowych rozważania ograniczamy do klasycznego układu dwuczynnikowego w blokach niekompletnych, układu split-plot i układu split-block. W ostatnim przypadku rozważane modele obejmują zarówno klasyczne układy kompletne jak i układy niekompletne ze względu na jeden lub oba czynniki doświadczalne. Model obserwacji uzyskany w pracy wynika bezpośrednio z zastosowanego w doświadczeniu schematu randomizacyjnego, dokładnie opisanego w pracy, oraz z pewnych zwykle przyjmowanych założeń. Podstawowe założenie orzeka, że obserwowana reakcja cechy jest sumą trzech składników, to jest "plonu zerowego", związanego z jednostką doświadczalną, "czystego efektu" pochodzącego od obiektu i "błędu technicznego", związanego z pewnymi nieścisłościami w przeprowadzaniu doświadczenia jak i nieścisłościami pomiarowymi. Praca ta ma charakter pracy przeglądowej. Stanowi ona pewne podsumowanie badań w tym zakresie. Oznacza to, że niektóre fragmenty tej pracy zostały częściowo opublikowane w innych, wcześniejszych pracach autora.

Słowa kluczowe: randomizacja, addytywność, budowanie modelu, układy o strukturze blokowej, doświadczenia czynnikowe.