American Journal of Theoretical and Applied Statistics
2022; 11(1): 45-57
http://www.sciencepublishinggroup.com/j/ajtas
doi: 10.11648/j.ajtas.2022451101.17
ISSN: 2326-8999 (Print); ISSN: 2326-9006 (Online)

Minimum Number of Replications for Tests in Four-Way ANOVA in Cross Classification and Split-Plot Design

Rob Verdooren[1, *], Dieter Rasch[2]

[1]Danone Nutricia Research, Division Data Sciences, Utrecht, the Netherlands

[2]Institute of Statistics, University of Natural Resources and Life Sciences, Vienna, Austria

Email address:

l.r.verdooren@outlook.com (R. Verdooren), dieter.rasch@boku.ac.at (D. Rasch)

[*]Corresponding author

To cite this article:

Rob Verdooren, Dieter Rasch. Minimum Number of Replications for Tests in Four-Way ANOVA in Cross Classification and Split-Plot Design. *American Journal of Theoretical and Applied Statistics*. Vol. 11, No. 1, 2022, pp. 45-57. doi: 10.11648/j.ajtas.20221101.17

Received: January 17, 2022; **Accepted**: February 9, 2022; **Published**: February 25, 2022

Abstract: In statistical books for the analysis of designed experiments one can finds sometimes also the computation of the number of replications for balanced one-factor and two-factors designs. Later there were papers published concerning the computation of the number of replications of at most three-factors crossed or nested balanced designs. In 2011 the book “Optimal experimental design with R” was published; further a special R- program OPDOE was made to do the computation for these designs and the OPDOE program was used in this book. In this paper an extension of the determination of the minimum number of replications for balanced designs is given for four-factor crossed designs. The balanced cross classification of the four-way analysis of variance of the following models are investigated: Model 1 The factors *A*, *B*, *C* and *D* are all fixed; Model 2 ***D*** is random *A*, *B* and *C* are fixed; Model 3 ***C*** and ***D*** are random, *A* and *B* are fixed; Model 4 ***B, C*** and ***D*** are random, *A* is fixed. For these models small R-programs are given to compute the minimal number of the replications for testing the fixed effects using the non-centrality parameter λ of the non-central *F*- distribution $F(df_1, df_2, \lambda)$. Further balanced Split-Plot design with one or two fixed factors in the main-plots are considered. The Blocks are denoted with *B*. The *F* statistics for testing the significance of the fixed factors are described and small R-programs for the determination of the minimal number of replications are given using the non-centrality parameter λ of the non-central *F*- distribution $F(df_1, df_2, \lambda)$.

Keywords: Balanced Four-way ANOVA, Cross Classification, Split-plot Designs, Non-centrality Parameter λ of the Non-central, *F*-distribution $F(df_1, df_2, \lambda)$, Minimal Number of Replications

1. Introduction

In experiments and surveys often the influence of several factors on a character y modelled by a random variable $\boldsymbol{y}$ are investigated. Each of the factors has at least two levels. There are several possibilities how factors can be combined. Let us consider two factors *A, B* with levels $A_1, \dots, A_a, a \geq 2$, and $B_1, \dots, B_b, b \geq 2$ respectively. If each level of *A* can occur together with each level of *B, A* and *B* are cross classified – symbol AxB and we may have ab sub-classes in the experiment. Such an experiment is called balanced (orthogonal) if in the cross classification all ab factor combinations occur in the experiment with equal sub-class numbers n.

The determination of the minimal number of replications in balanced designs for testing fixed effects has already a long history for one-factor or two-factor studies. This was made possible by the publication of Tang [16] with the distribution of the non-central *F*-distribution. The charts of Tang were later also published by Owen [5] and Pearson and Hartley [6]. Kuehl [1] gives in section 2.14 “How many replications for the F-test (of one-factor)” and in section 6.8 : “How many replications to test factor effects (for two-factors)” using the table on pp. 616-625 of charts of the power function of the F-test. Also Ott and Longnecker [4] described in section 14.6 “Determining the Number of Replications (of one-factor)” using the table on pp. 1123-1126 with charts of the Power of the analysis of variance test. In Kutner et al. [2] is given in section 16.10 “Planning sample sizes with power approach (for one-factor study)” and in section 19.11 “Planning of sample sizes for two-factor studies”, using the table on pp. 1337-1341 “Power Values for

Analysis of Variance (fixed effects)". In Wang et al. [17] and Rasch and Verdooren [13] the determination of the size for three-factors studies in a balanced experiment for mixed ANOVA is given. In Rasch et al. [11] an overview of the determination of the minimal number in balanced cross-classifications and nested-classifications of fixed and mixed models for two-factors and three-factors is given together with an R-program package OPDOE (Optimal Design of Experiments). See also Rasch et al. [12, 13] and Spangl et al. [15] for a balanced three-way ANOVA classification to determine the minimal number of replications.

In this paper is treated the four cross-classified factors *A, B, C, D* in balanced (orthogonal) experiments. In Example 7 below a clinical study is given : a factor *A* of *a*=3 Covid 19 vaccines Biontech (A_1), Astra Seneca (A_2) and Johnson&Johnson (A_3); a factor *B* of female (B_1) and male (B_2) patients and a factor *C* with an age below 60 years (C_1) and 60 years and older (C_2). Because as well the number of levels as also the levels are fixed independently from the experimenter, we call the factors fixed. The survey should be executed with the random factor $\boldsymbol{D}$ in *d*=4 hospitals selected randomly from a huge number of hospitals in a country.

Further are considered two balanced Split-Plot Designs. The first one is the fixed factor *A* used in the main-plots of the factor *B* of blocks; further the fixed factor *C* and *D* are used in the Split-Plots. The second one are the fixed factors *A* and *C* used in the Main-plots of the factor *B* of blocks; further the fixed factor *D* is used in the Split-Plots.

Tests for testing the significance of fixed main or interaction effects in models of the balanced four-way cross classified analysis of variance with at least one fixed factor is then treated. The four factors are denoted by *A, B, C*, and *D*. Random factors as well as random variables in the models are bold printed. The symbol *x* between factors means cross classification.

The four-way cross classification Model 1 is based on the model (1) (with all interactions) here written for the case where all factors are fixed.

$$y_{ijklm} = \mu + a_i + b_j + c_k + d_l + (ab)_{ij} + (ac)_{ik} + (ad)_{il} + (bc)_{jk} + (bd)_{jl} + (cd)_{kl} + (abc)_{ijk} + (abd)_{ijl} + (acd)_{ikl} + (bcd)_{jkl} + (abcd)_{ijkl} + \boldsymbol{e_{ijklm}} \tag{1}$$

i = 1, …, a, j= 1, …, b, k = 1, …, c, l= 1, …, d, m = 1,…, n ≥ 2.

Assumed are that all fixed effects in Model 1 sum up to zero, when summation is done over at least one of the superscripts of the effect. For instance $\sum_{k=1}^{b}(bcd)_{jkl}=0$. The random effects $\boldsymbol{e_{ijklm}}$ are all independent from each other and are normally distributed with expected value zero and variance σ^2.

2. The ANOVA – Table, Expected Mean Squares and F Statistics

Here are investigated four models with four, three, two factors fixed and also only one factor fixed.

Table 1. *Analysis of Variance Table of a Four-way Cross-Classification with equal Subclass Numbers n, SS is Sum of Squares and df are the degrees of freedom.*

Source of Variation	***SS***	***df***
Between *A*-levels	$SS_A = \frac{1}{bcdn}\sum Y^2_{i....} - \frac{1}{N}Y^2_{......}$	$a-1$
Between *B*- levels	$SS_B = \frac{1}{acdn}\sum Y^2_{.j...} - \frac{1}{N}Y^2_{......}$	$b-1$
Between *C*- levels	$SS_C = \frac{1}{abdn}\sum Y^2_{..k..} - \frac{1}{N}Y^2_{......}$	$c-1$
Between *D*- levels	$SS_D = \frac{1}{abcn}\sum Y^2_{...l.} - \frac{1}{N}Y^2_{......}$	$d-1$
Interaction $A \times B$	$SS_{AB} = \frac{1}{cdn}\sum Y^2_{ij...} - \frac{1}{bcdn}\sum Y^2_{i....} - \frac{1}{acdn}\sum Y^2_{.j...} + \frac{1}{N}Y^2_{......}$	$(a-1)(b-1)$
Interaction $A \times C$	$SS_{AC} = \frac{1}{bdn}\sum Y^2_{i.k..} - \frac{1}{bcdn}\sum Y^2_{i....} - \frac{1}{abdn}\sum Y^2_{..k..} + \frac{1}{N}Y^2_{......}$	$(a-1)(c-1)$
Interaction $A \times D$	$SS_{AD} = \frac{1}{bcn}\sum Y^2_{i..l.} - \frac{1}{bcdn}\sum Y^2_{i....} - \frac{1}{abcn}\sum Y^2_{...l.} + \frac{1}{N}Y^2_{......}$	$(a-1)(d-1)$
Interaction $B \times C$	$SS_{BC} = \frac{1}{adn}\sum Y^2_{.jk..} - \frac{1}{acdn}\sum Y^2_{.j...} - \frac{1}{abdn}\sum Y^2_{..k..} + \frac{1}{N}Y^2_{......}$	$(b-1)(c-1)$
Interaction $B \times D$	$SS_{BD} = \frac{1}{acn}\sum Y^2_{.j.l.} - \frac{1}{acdn}\sum Y^2_{.j...} - \frac{1}{abcn}\sum Y^2_{...l.} + \frac{1}{N}Y^2_{......}$	$(b-1)(d-1)$
Interaction $C \times D$	$SS_{CD} = \frac{1}{abn}\sum Y^2_{..kl.} - \frac{1}{abdn}\sum Y^2_{..k..} - \frac{1}{abcn}\sum Y^2_{...l.} + \frac{1}{N}Y^2_{......}$	$(c-1)(d-1)$
Interaction $A \times B \times C$	$SS_{ABC} = \frac{1}{dn}\sum_{i,j,k}\boldsymbol{Y}^2_{ijk..} + \frac{1}{bcdn}\sum_{i}\boldsymbol{Y}^2_{i....} + \frac{1}{acdn}\sum_{j}\boldsymbol{Y}^2_{.j...} + \frac{1}{abdn}\sum_{k}\boldsymbol{Y}^2_{..k..} - \frac{1}{cdn}\sum_{i,j}\boldsymbol{Y}^2_{ij...}$ $-\frac{1}{adn}\sum_{j,k}\boldsymbol{Y}^2_{..jk.} - \frac{1}{bdn}\sum_{i,k}\boldsymbol{Y}^2_{i.k..} - \frac{\boldsymbol{Y}^2_{......}}{N}$	$(a-1)(b-1)(c-1)$
Interaction $A \times B \times D$	$SS_{ABD} = \frac{1}{cn}\sum_{i,j,l} Y^2_{ij.l.} + \frac{1}{bcdn}\sum_{i}\boldsymbol{Y}^2_{i....} + \frac{1}{acdn}\sum_{j}\boldsymbol{Y}^2_{.j...} + \frac{1}{abcn}\sum_{l}\boldsymbol{Y}^2_{...l.} - \frac{1}{cdn}\sum_{i,j}\boldsymbol{Y}^2_{ij...}$ $-\frac{1}{acn}\sum_{j,l}\boldsymbol{Y}^2_{..jl.} - \frac{1}{bcn}\sum_{i,l}\boldsymbol{Y}^2_{i..l.} - \frac{\boldsymbol{Y}^2_{......}}{N}$	$(a-1)(b-1)(d-1)$

Source of Variation	*SS*	*df*
Interaction $A \times C \times D$	$SS_{ACD} = \frac{1}{bn}\sum_{i,k,l} \boldsymbol{Y}^2_{i.kl.} + \frac{1}{bcdn}\sum_{i} \boldsymbol{Y}^2_{i....} + \frac{1}{abdn}\sum_{k} \boldsymbol{Y}^2_{..k..} + \frac{1}{abcn}\sum_{l} \boldsymbol{Y}^2_{...l.} - \frac{1}{bdn}\sum_{i,k} \boldsymbol{Y}^2_{i.k..}$ $-\frac{1}{abn}\sum_{k,l} \boldsymbol{Y}^2_{..k,l.} - \frac{1}{bcn}\sum_{i,l} \boldsymbol{Y}^2_{i..l.} - \frac{\boldsymbol{Y}^2_{......}}{N}$	$(a-1)(c-1)(d-1)$
Interaction $B \times C \times D$	$SS_{BCD} = \frac{1}{an}\sum_{j,k,l} \boldsymbol{Y}^2_{.jkl.} + \frac{1}{acdn}\sum_{j} \boldsymbol{Y}^2_{.j...} + \frac{1}{abdn}\sum_{k} \boldsymbol{Y}^2_{..k..} + \frac{1}{abcn}\sum_{l} \boldsymbol{Y}^2_{...l.} - \frac{1}{adn}\sum_{j,k} \boldsymbol{Y}^2_{.jk..}$ $-\frac{1}{acn}\sum_{j,l} \boldsymbol{Y}^2_{..jl.} - \frac{1}{abn}\sum_{k,l} \boldsymbol{Y}^2_{.kl..} - \frac{\boldsymbol{Y}^2_{......}}{N}$	$(b-1)(c-1)(d-1)$
Interaction $A \times B \times Cx\ D$	$SS_T - SS_A - SS_B - SS_C - SS_D - SS_{AB} - SS_{AC} - SS_{AD} - SS_{BC} - SS_{BD} - SS_{CD} - SS_{ABC} - SS_{ABD} - SS_{BCD} - SS_{res}$-	$(a-1)(b-1)(c-1)(d-1)$
Within the Classes (residual)	$SS_{res} = \sum y^2_{ijklm} - \frac{\sum Y^2_{ijkl.}}{n}$	$abcd(n-1)$
Corrected Total	$SS_T = \sum Y^2_{ijkl.} - \frac{Y^2_{.....}}{N}$	*N*-1

$N = abcdn$ with $n \geq 2$.

Table 2. *Mean Squares, Expectations for Model 1 and F –Statistics.*

Mean Squares	Expected Mean Squares	*F* -Statistic
$MS_A = \frac{SS_A}{a-1}$	$\sigma^2 + \frac{bcdn}{a-1}\sum a_i^2$	$F_A = \frac{abcd(n-1)}{a-1}\frac{SS_A}{SS_{res}}$
$MS_B = \frac{SS_B}{b-1}$	$\sigma^2 + \frac{acdn}{b-1}\sum b_j^2$	$F_B = \frac{abcd(n-1)}{b-1}\frac{SS_B}{SS_{res}}$
$MS_C = \frac{SS_C}{c-1}$	$\sigma^2 + \frac{abdn}{c-1}\sum c_k^2$	$F_C = \frac{abcd(n-1)}{c-1}\frac{SS_C}{SS_{res}}$
$MS_D = \frac{SS_D}{d-1}$	$\sigma^2 + \frac{abcn}{d-1}\sum d_l^2$	$F_D = \frac{abcd(n-1)}{d-1}\frac{SS_D}{SS_{res}}$
$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$\sigma^2 + \frac{cdn}{(a-1)(b-1)}\sum(ab)^2_{ij}$	$F_{AB} = \frac{abcd(n-1)}{(a-1)(b-1)}\frac{SS_{AB}}{SS_{res}}$
$MS_{AC} = \frac{SS_{AC}}{(a-1)(c-1)}$	$\sigma^2 + \frac{bdn}{(a-1)(c-1)}\sum(ac)^2_{ik}$	$F_{AC} = \frac{abcd(n-1)}{(a-1)(c-1)}\frac{SS_{AC}}{SS_{res}}$
$MS_{AD} = \frac{SS_{AD}}{(a-1)(d-1)}$	$\sigma^2 + \frac{bcn}{(a-1)(d-1)}\sum(ad)^2_{il}$	$F_{AD} = \frac{abcd(n-1)}{(a-1)(d-1)}\frac{SS_{DA}}{SS_{res}}$
$MS_{BC} = \frac{SS_{BC}}{(b-1)(c-1)}$	$\sigma^2 + \frac{adn}{(b-1)(c-1)}\sum(bc)^2_{jk}$	$F_{BC} = \frac{abcd(n-1)}{(b-1)(c-1)}\frac{SS_{BC}}{SS_{res}}$
$MS_{BD} = \frac{SS_{BD}}{(b-1)(d-1)}$	$\sigma^2 + \frac{acn}{(b-1)(d-1)}\sum(bd)^2_{jl}$	$F_{BD} = \frac{abcd(n-1)}{(b-1)(d-1)}\frac{SS_{BD}}{SS_{res}}$
$MS_{CD} = \frac{SS_{CD}}{(c-1)(d-1)}$	$\sigma^2 + \frac{abn}{(c-1)(d-1)}\sum(cd)^2_{kl}$	$F_{CD} = \frac{abcd(n-1)}{(c-1)(d-1)}\frac{SS_{CD}}{SS_{res}}$
$MS_{ABC} = \frac{SS_{ABC}}{(a-1)(b-1)(c-1)}$	$\sigma^2 + \frac{dn}{(a-1)(b-1)(c-1)}\sum(abc)^2_{ijk}$	$F_{ABC} = \frac{abcd(n-1)}{(a-1)(b-1)(c-1)}\frac{SS_{ABC}}{SS_{res}}$
$MS_{ABD} = \frac{SS_{ABD}}{(a-1)(b-1)(d-1)}$	$\sigma^2 + \frac{cn}{(a-1)(b-1)(d-1)}\sum(abd)^2_{ijl}$	$F_{ABD} = \frac{abcd(n-1)}{(a-1)(b-1)(d-1)}\frac{SS_{ABD}}{SS_{res}}$
$MS_{ACD} = \frac{SS_{ACD}}{(a-1)(c-1)(d-1)}$	$\sigma^2 + \frac{bn}{(a-1)(c-1)(d-1)}\sum(acd)^2_{ikl}$	$F_{ACD} = \frac{abcd(n-1)}{(a-1)(c-1)(d-1)}\frac{SS_{ACD}}{SS_{res}}$
$MS_{BCD} = \frac{SS_{BCD}}{(b-1)(c-1)(d-1)}$	$\sigma^2 + \frac{an}{(b-1)(c-1)(d-1)}\sum(bcd)^2_{jkl}$	$F_{BCD} = \frac{abcd(n-1)}{(b-1)(c-1)(d-1)}\frac{SS_{BCD}}{SS_{res}}$
$MS_{ABCD} = \frac{SS_{ABCD}}{(a-1)(b-1)(c-1)(d-1)}$	$\sigma^2 + \frac{n}{(a-1)(b-1)(c-1)(d-1)}\sum(abcd)^2_{ijkl}$	$F_{ABCD} = \frac{abcd(n-1)}{(a-1)(b-1)(c-1)(d-1)}\frac{SS_{ABCD}}{SS_{res}}$
$MS_{res} = \boldsymbol{s}^2 = \frac{SS_{res}}{abcd(n-1)}$	σ^2	

2.1. Model 1 Cross Classification with All Factors Fixed

A model with four fixed factors *A, B*, *C*, *D*: *AxBxCxD* is called Model 1. The ANOVA – table given above is independent of the model and will be used for all four models. The expected Mean Squares E(*MS*) depend on the models.

2.2. Model 2, 3 and 4 for Cross Classification with Three, Two and One Fixed Factor(s)

A model with three fixed factors *A*, *B*, *C*: *AxBxCx**D*** is called Model 2. If another factor in place of ***D*** is random, the factors are rearranged without loss of generality.

The Table with Mean Squares and their expectations is given in Table 2. How to derive the Expected Mean Squares for all four and other models is described in Rasch and Schott ([8], [9, Section 7.2]).

The model equation for Model 2 is given by

$$y_{ijklm} = \mu + a_i + b_j + c_k + \boldsymbol{d}_l + (ab)_{ij} + (ac)_{ik} + (\boldsymbol{ad})_{il} + (bc)_{jk} + (\boldsymbol{bd})_{jl} + (\boldsymbol{cd})_{jl} + (abc)_{ijk} + (\boldsymbol{abd})_{ijl} + (\boldsymbol{acd})_{ikl} + (\boldsymbol{bcd})_{jkl} + (\boldsymbol{abcd})_{ijkl} + \boldsymbol{e}_{ijklm} \quad (2)$$

i = 1, …, a, j= 1, …, b, k = 1, …, c, l= 1, …, d, m = 1,…, n ≥ 2.

The side conditions for the fixed effects are as in Model 1. If some of the factors or factor combinations are random, assumed is that the corresponding main and interaction effects are independent, normally distributed and have expectation zero and variances with small subscripts equal to the capital letters of the factors.

A model with two fixed factors *A* and *B*: *AxBx**CxD*** is called Model 3.

The model equation for Model 3 is given by

$$y_{ijklm} = \mu + a_i + b_j + \boldsymbol{c_k} + \boldsymbol{d}_l + (ab)_{ij} + (\boldsymbol{ac})_{ik} + (\boldsymbol{ad})_{ik} + (\boldsymbol{bc})_{jk} + (\boldsymbol{bd})_{jl} + (\boldsymbol{abc})_{ijk} + (\boldsymbol{abd})_{ijl} + (\boldsymbol{acd})_{ikl} + (\boldsymbol{bcd})_{jkl} + (\boldsymbol{abcd})_{ijkl} + \boldsymbol{e}_{ijklm} \quad (3)$$

i = 1, …, a, j= 1, …, b, k = 1, …, c, l= 1, …, d, m = 1,…, n ≥ 2.

The side conditions for the fixed effects are as in Model 1, the random effects are assumed to be independent with expectation zero, normally distributed and have expectation zero and variances with small subscripts equal to the capital letters of the factors.

A model with one fixed factor *A*: *Ax**BxCxD*** is called Model 4.

The model equation for Model 4 is given by

$$y_{ijklm} = \mu + a_i + \boldsymbol{b_j} + \boldsymbol{c_k} + \boldsymbol{d}_l + (\boldsymbol{ab})_{ij} + (\boldsymbol{ac})_{ik} + (\boldsymbol{ad})_{il} + (\boldsymbol{bc})_{jk} + (\boldsymbol{bd})_{jl} + (\boldsymbol{abc})_{ijk} + (\boldsymbol{abd})_{ijl} + (\boldsymbol{acd})_{ikl} + (\boldsymbol{bcd})_{jkl} + (\boldsymbol{abcd})_{ijkl} + \boldsymbol{e}_{ijklm} \quad (4)$$

i = 1, …, *a*, *j*= 1, …, *b*, *k* = 1, …, *c*, *l*= 1, …, *d*, *m* = 1,…, *n* ≥ 2.

The side conditions for the fixed effects are as in Model 1, the random effects are assumed to be independent with expectation zero, normally distributed and have expectation zero and variances with small subscripts equal to the capital letters of the factors.

Because the *F* statistics for fixed effects of Model 2, 3 and 4 are identical with those in Table 2, a Table of Expected Mean Squares and *F* –Statistics for Model 2, 3, and 4 is not presented. As already mentioned the reader may derive them by using the algorithm described in Rasch and Schott [8, 9].

3. Determination of the Minimum Number of Replications for Testing Fixed Effects

In the four models considered in this paper the minimum number of replications for testing fixed effects depends not on the model.

The random error term $\boldsymbol{e}_{ijklm}$, containing independent observational errors with $E(\boldsymbol{e}_{ijklm}) = 0$, $var(\boldsymbol{e}_{ijklm}) = \sigma^2$, is the same for all observations. For testing null-hypothesis of the effects we assume that the $\boldsymbol{e}_{ijklm}$ are normally distributed.

In the ANOVA tables the *E*(*MS*) (Expected *MS*) is given. The non-centrality parameter λ can be obtained by the general formula given in Lindman ([3], p. 151),

$$\lambda = \frac{df_1[E(MS_1) - E(MS_2)]}{E(MS_2)} \quad (5)$$

where $E(MS_1)$ and $E(MS_2) = \sigma^2$, are the expected mean sum of squares of the numerator and denominator of the *F*-test statistic with degrees of freedom df_1 and df_2, respectively.

If the fixed factor has at least three levels the minimum experimental size depends on the values of its factor levels. As described in Rasch and Verdooren [13] and Rasch et al. [10-12] we calculate the minimum experimental size for the least favourable (maximin size) and the most favourable (minimin size) case of the location of the values of the factor levels of *A*. We first describe what we mean by the minimum of the minimal experimental size and the maximum of the minimal experimental size as given by Rasch et al. [7] and Rasch et al. [14]. The relation

$$F(f_1, f_2, 0 \mid 1 - \alpha) = F(f_1, f_2, \lambda \mid \beta), \quad (6)$$

is used, where f_1 and f_2 are the degrees of freedom of the numerator and the and the denominator, respectively. Further α and β are the two risks of the first and second kind of the corresponding *F*-test respectively, and λ is the non-centrality parameter of the non-central *F*-distribution. Equation (6) plays an important role in all other sections of this paper. Beside f_1, f_2, α and β the difference δ between the largest and the smallest effect (main effect or in the following sections also interaction effect) of the fixed factor *A*, to be tested against zero, belongs to the precision requirements. The solution λ in (6) we denote by

$$\lambda = \lambda(\alpha, \beta, f_1, f_2).$$

Let E_{min} and E_{max} be the minimum and the maximum of *q* real effects $E_1, E_2, \ldots, E_q$ of a fixed factor *E* or of interaction effects.

The minimal size of the experiment depends on λ according to the exact position of all q effects. But this is unknown before the experiment starts. We consider two extreme cases, the most favourable (resulting in the smallest minimal size $n_{\min}$) and the least favourable (resulting in the largest minimal size $n_{\max}$) case. The least favourable case leads to the smallest non-centrality parameter λ_{min} and by this to the so-called maximin size $n_{\max}$. This occurs if the effects $E_{min} = E_1 = -E$ and $E_{max} = E_q = E$ and the $q - 2$ non-extreme effects are equal to $M = \frac{E_{\max} + E_{\min}}{2}$. For $\bar{E} = 0$, $\sum_{i=1}^{q}(E_i - \bar{E})^2 = 2E^2$ this is shown in the following scheme.

$E_1 = -E \qquad 0 = E_2 = \cdots = E_{q-1} \qquad E_q = E$

For the general case of this configuration with $\bar{E} = M$ we have

$$\bar{E} = M \text{ and } E_q = (E_{\max} - M) = \frac{E_{max} - E_{min}}{2},\ E_1 = (E_{\min} - M) = -\frac{E_{max} - E_{min}}{2} \text{ and } \sum_{i=1}^{q}(E_i - \bar{E})^2 = 2\left(\frac{\delta}{2}\right)^2 = \frac{\delta^2}{2}$$

with $\delta = E_{max} - E_{min}$.

The most favourable case for even $q=2m$ occurs if m of the E_i equal $E_{\min} = -E$ and the m other E_i equal $E_{\max} = E$. For odd $q=2m + 1$ again m of the E_i should equal $E_{\min}$ and m other E_i should equal $E_{\max}$, and the remaining effect should be equal to one of the two extremes $E_{\min}$ or $E_{\max}$. For $\bar{E} = 0$, $\sum_{i=1}^{q}(E_i - \bar{E})^2 = qE^2$ this is shown for even q in the following scheme.

$E_1 = E_2 = \cdots = E_m = -E \qquad 0 \qquad E_{m+1} = E_{m+2} = \cdots = E_q = E$

(This corresponds with a D-optimal design in linear regression).

Hence for the general case of this configuration we have $\bar{E} = M$,

$$E_q = (E_{max} - M) = \frac{E_{max} - E_{min}}{2} \text{ and } E_1 = (E_{min} - M) = -(E_{max} - E_{min})/2$$

and

$$\sum_{i=1}^{q}(E_i - \bar{E})^2 = q(\delta/2)^2 = q\delta^2/4.$$

and

$$\lambda_{min} = \delta^2/(2\sigma^2) \text{ and } \lambda_{max} = q\delta^2/(4\sigma^2) \text{ with } \delta = (E_{\max} - E_{\min}) \tag{7}$$

This derivation is also given in Rasch and Verdooren [13].

In the following sections the formulae and R-programs for n_{min} and n_{max} for the Models 1, 2, 3 and 4 are given.

3.1. Experimental Size for Model 1

The calculation of the sample size minimin and maximin for the test of the factor A of the Null hypothesis: H_0: $a_1 = a_2 = \cdots . = a_a$ is demonstrated.

Under the side condition $\sum_i a_i = 0$ the null hypothesis can also be formulated as H_0: $a_1 = a_2 = \cdots . = a_a = 0$.

F –statistic: $F = \frac{MS_A}{MS_{res}}$, with $df(A) = (a-1)$ and $df(res) = abcd(n-1)$.

Non-centrality parameter $\lambda_A = \frac{bcdn\sum_{i=1}^{a} a_i^2}{\sigma^2}$.

For $\lambda_A = \frac{bcdn\sum_{i=1}^{a} a_i^2}{\sigma^2}$ we have $\lambda_{A\,max} = \frac{abcdn\delta^2}{4\sigma^2}$ and $\lambda_{A\,min} = \frac{bcdn\delta^2}{2\sigma^2}$ with $a_{\max} - a_{\min} \geq \delta$.

For the non-centrality parameter λ_A we use in the R–program delta = δ/σ.

A program in R which gives the solution of formula for the sample size maximin of the fixed effect A is:

```
> maximinA = function( p, a, b, c, d, delta , beta)
{
f = function(n, p, a, b, c, d, delta , beta )
{
A = qf(p, a-1, a*b*c*d*(n-1), 0)
B = qf( beta, a-1, a*b*c*d*(n-1), delta*delta*n *a*b*c*d/4)
C = A-B
}
k = uniroot(f, c(2, 10000 ), p=p, a=a, b=b, c=c, d=d, delta=delta, beta=beta) $root
k0 = ceiling(k)
print (paste(" maximinA sample number: n = ", k0), quote=F)
}
```

Example 1

where $p = 1 - \alpha$:

```
> maximinA( p=0.95, a=3, b = 2, c = 2, d= 0.5, delta = 1, beta = 0.05)
[1] maximinA sample number: n = 11
```

A program in R which gives the solution of formula (6) for the sample size minimin of the fixed effect *A* is:

```
> miniminA = function( p, a, b, c, d, delta , beta)
{
 f = function(n, p, a, b, c, d, delta , beta )
 {
 A = qf(p, a-1, a*b*c*d*(n-1), 0)
 B = qf( beta, a-1, a*b*c*d*(n-1), 0.5* delta*delta*n*b*c*d)
 C = A-B
 }
 k = uniroot(f, c(2, 10000), p=p, a=a, b=b, c= c, d=d, delta=delta, beta=beta) $root
 k0 = ceiling(k)
 print (paste(" miniminA sample number: n = ", k0), quote=F)
 }
```

Example 2

where $p = 1 - \alpha$:

```
> miniminA( p=0.95, a=3, b = 2, c = 2, d= 2, delta = 0.5, beta = 0.05)
[1] miniminA sample number: n = 16
```

To give an impression over the variation of the sizes we give minimin and maximin sizes for $\alpha = 0.05$; $\beta = 0.1$ and $\delta = 0.2\ \sigma$ for some values of *a, b, c, d* in Table 3. Note that when *a* is larger than 2 minimin and maximin are different.

Table 3. *Minimum sub-class numbers (upper entry minimin; lower entry maximin) for test of fixed factors A in the four-way cross classification Model 1 with α = 0.05, β = 0.1 and δ = 0.2 σ.*

a	b	c=2		c=3	
		d=3	d=10	d=3	d=10
2	2	44	14	30	9
		44	14	30	9
	3	30	9	20	6
		30	9	20	6
3	2	53	16	36	11
		36	11	24	8
	3	36	11	24	8
		24	8	16	5
5	2	65	20	43	13
		26	8	18	6
	3	43	13	29	9
		18	6	14	4

Remark

If the values of *a, b, c, d* are such that the total number of observations $N = abcdn$ for $n = 2$ is already quite large, the R-program gives no value for the minimin and maximin. In this case we can use the value $n = 2$ for the minimal number of replications. This is demonstrated in the Examples 3 and 4.

Example 3

where $p = 1 - \alpha$:

```
> miniminA( p=0.95, a=10, b = 3, c = 4, d= 5, delta = 1, beta = 0.05)
Error in uniroot(f, c(2, 10000), p = p, a = a, b = b, c = c, d = d, delta = delta, :
f() values at end points not of opposite sign
```

We calculate now the power directly for miniminA with *n* =2:

```
> Fpvalue = qf(0.95, 10-1, 10*3*4*5*(2-1), 0)
> Fpvalue
[1] 1.895472
> powerminiminA = 1- pf(Fpvalue, 10-1,10*3*4*5*(2-1), 1*1*2*3*4*5/2)
> powerminiminA
[1] 0.9999908
```

Example 4

where $p = 1 - \alpha$:

```
> maximinA( p=0.95, a=10, b= 3, c= 4, d=5, delta = 1, beta = 0.05)
Error in uniroot(f, c(2, 10000), p = p, a = a, b = b, c = c, d = d, delta = delta, :
f() values at end points not of opposite sign
In addition: There were 14 warnings (use warnings() to see them)
```

We calculate now the power directly for maximinA with *n* =2:

```
> Fpvalue = qf(0.95, 10-1, 10*3*4*5*(2-1), 0)
> Fpvalue
[1] 1.895472
> powermaximinA = 1- pf(Fpvalue, 10-1,10*3*4*5*(2-1), 1*1*2*10*3*4*5/4)
> powermaximinA
[1] 1
```

For the other tests of the fixed effects a change must be done analogously in the R program according to the non-centrality parameter λ derived from the E(*MS*) column of the ANOVA-table and in the R-program; used is delta = δ/σ.

For example now the calculation is demonstrated of the sample size minimin and maximin for the test of the fixed interaction effect *AB* of the effects *A* and *B* with the side condition in (1) as :

Null hypothesis: H_0: $(ab)_{ij} = 0$, for all i,j.

F –statistic: $F = MS_{AB} / MS_{Res}$ with $df\ (AB) = (a-1)(b-1)$ and $df(\text{res}) = abcd(n-1)$.

Non-centrality parameter $\lambda_{AB} = \frac{cdn\ \Sigma \sum_{ij=1}^{ab} (ab)_{ij}^2}{\sigma^2}$.

For $\lambda_{AB} = \frac{cdn\ \Sigma \sum_{ij=1}^{ab} (ab)_{ij}^2}{\sigma^2}$ we have $\lambda_{AB\ max} = \frac{abcdn\delta^2}{4\sigma^2}$ and $\lambda_{AB\ min} = \frac{cdn\delta^2}{2\sigma^2}$

with $a_{\max} - a_{\min} \geq \delta$.

For the non-centrality parameter λ_{AB} in the R –program delta = δ/σ is used.

A program in R which gives the solution of formula (6) for the sample size maximin for the interaction *AB* is:

```
> maximinAB = function( p, a, b, c, d, delta , beta)
{
 f = function(n, p, a, b, c, d, delta , beta )
 {
 A = qf(p, (a-1)*(b-1), a*b*c*d*(n-1), 0)
 B = qf( beta, (a-1)*(b-1), a*b*c*d*(n-1), delta*delta*n
```

```
*a*b*c*d/4)
  C = A-B
  }
  k = uniroot(f, c(2, 10000 ), p=p, a=a, b=b, c=c, d=d, delta=delta, beta=beta) $root
  k0 = ceiling(k)
  print (paste(" maximinAB sample number: n = ", k0), quote=F)
  }
```

Example 5

where $p = 1 - \alpha$:

```
> maximinAB( p=0.95, a=3, b = 2, c = 2, d= 2, delta = 1, beta = 0.05)
[1] maximinAB sample number: n = 3
```

A program in R which gives the solution of formula (6) for the sample size minimin for *AB* is:

```
> miniminAB = function( p, a, b, c, d, delta , beta)
  {
  f = function(n, p, a, b, c, d, delta , beta )
  {
  A = qf(p, (a-1)*(b-1), a*b*c*d*(n-1), 0)
  B = qf( beta, (a-1)*(b-1), a*b*c*d*(n-1), 0.5* delta*delta*n*c*d)
  C = A-B
  }
  k = uniroot(f, c(2, 10000), p=p, a=a, b=b, c= c, d=d, delta=delta, beta=beta) $root
  k0 = ceiling(k)
  print (paste(" miniminAB sample number: n = ", k0), quote=F)
  }
```

Example 6

where $p = 1 - \alpha$:

```
> miniminAB( p=0.95, a=3, b= 2, c= 2, d=2, delta = 1, beta = 0.05)
[1] miniminAB sample number: n = 8
```

3.2. Experimental Size for the Models 2, 3 and 4

For the fixed effects the determination of the minimin and maximin one can use the same R programs as given in section 3.1.

Example 7

Problem: We consider a fixed factor *A* of COVID 19 survey including three vaccines Biontech (A_1), Astra Seneca (A_2) and Johnson (A_3) (levels of *A*), a fixed factor *B* of female (B_1) and male (B_2) patients and a fixed factor *C* with an age below 60 years (C_1) and 60 years and older (C_2). The survey should be executed with the random factor ***D*** in four hospitals randomly selected from a huge number of hospitals in a country. This is a Model 2 with factors *A, B, C* fixed and factor ***D*** random.

How many patients of each of the 48 groups ($3 * 2 * 2 * 4 = 48$) are needed, when we like to test the null hypothesis that all three vaccines have the same effectiveness?

Use $\alpha = 0.01$, $\beta = 0.05$ and $\delta = 0.5\ \sigma$ for calculating the maximin size and the minimin size of the patients.

.

Solution

where $p = 1 - \alpha$

```
> maximinA( p=0.99, a=3, b= 2, c= 2, d=4, delta = 0.5, beta = 0.05)
[1] maximinA sample number: n = 7
> miniminA( p=0.99, a=3, b= 2, c= 2, d=4, delta = 0.5, beta = 0.05)
[1] miniminA sample number: n = 11
```

4. Minimum Number of Replications for Testing in Split-Plot Designs with *A* at the Main-plots

Consider the design that the fixed factor *A* has been randomized laid down with *an* Main-plots in a balanced randomized block design with the factor *B* as Blocks with *an* Main-plots per Block. Each Main-plot is split into *cdn* Split-plots. The fixed factors *C* and *D* are randomized laid down on the *cdn* Split-plots in the Main-plots. The fixed factors *A*, *C* and *D* have equal sizes of $n \geq 2$ replications.

The four-way cross classification is based on the model (with all interactions) here written for the case where the factor *B* denotes the Blocks and the factors *A*, *C* and *D* are fixed; the Main-plots have a random Main-plot error ***p***. The factor *A* with *n* replications is randomized laid down on the Main-plots of the Blocks *B*, otherwise stated *A* is nested in *B,* the notation $B \succ A$ denotes that *A* is nested in *B*; the factor combinations of *C*x*D* with *n* replications are randomized laid down on the Split-plots of the Main-plots, otherwise stated *C*x*D* is nested in *A*, $A \succ CxD$; the Split-plots have a random Split-plot error ***e***. Such a design is called a Split-Plot design with Factor *A* on Main-plots in Blocks and factor *C* and *D* on Split-plots.

$$y_{ijklm} = \mu + b_j + a_{i(j)} + \boldsymbol{p}_{m(ij)} + c_{k(ij)} + d_{l(ij)} + (ac)_{ik(ij)} + (ad)_{il(ij)} + (acd)_{ikl(ij)} + \boldsymbol{e}_{\boldsymbol{m(ijkl)}} \quad (8)$$

$$i = 1, \ldots, a, j = 1, \ldots, b, k = 1, \ldots, c, l = 1, \ldots, d, m = 1, \ldots, n \geq 2.$$

The first part of (8) are the Main-plot effects and the second part of (8) are the Split-plot effects.

Assumed is that all fixed effects in sum up to zero, when summation is done over at least one of the superscripts of the effect. For instance $\sum_{k=1}^{b}(acd)_{ikl} = 0$. The random effects $\boldsymbol{p}_{m(ij)}$ and $\boldsymbol{e}_{m(ijkl)}$ are assumed to be independent and normally distributed with expectation zero and with variances respectively σ^2_1 and σ^2.

The ANOVA – table and the expected Mean Squares E(*MS*) are given below in Table 4 and Table 5 respectively.

Table 4. *Analysis of Variance Table of a Split-Plot design with factor A at the Main-plots in Blocks B and equal Split-plot numbers n ≥ 2.*

Source of Variation	***SS***	***df***
Between B- levels	$SS_B = \frac{1}{acdn}\sum Y_{.j...}^2 - \frac{1}{N}Y_{......}^2$	$b-1$
Between A-levels	$SS_A = \frac{1}{bcdn}\sum Y_{i....}^2 - \frac{1}{N}Y_{......}^2$	$a-1$
Residual p of Main-Plots = Res1	$SS_{Res1} = \frac{1}{cdn}\sum Y_{ij...}^2 - \frac{1}{N}Y_{......}^2 - SS_B - SS_A$	$df\,Res1 = abn - b - a + 1$.
Between C- levels	$SS_C = \frac{1}{abdn}\sum Y_{..k..}^2 - \frac{1}{N}Y_{......}^2$	c - 1
Between D- levels	$SS_D = \frac{1}{abcn}\sum Y_{...l.}^2 - \frac{1}{N}Y_{......}^2$	d - 1
Interaction $A\times C$	$SS_{AC} = \frac{1}{bdn}\sum Y_{i.k..}^2 - \frac{1}{bcdn}\sum Y_{i....}^2 - \frac{1}{abdn}\sum Y_{..k..}^2 + \frac{1}{N}Y_{......}^2$	$(a-1)(c-1)$
Interaction $A\times D$	$SS_{AD} = \frac{1}{bcn}\sum Y_{i..l.}^2 - \frac{1}{bcdn}\sum Y_{i....}^2 - \frac{1}{abcn}\sum Y_{...l.}^2 + \frac{1}{N}Y_{......}^2$	$(a$ - $1)(d$ - $1)$
Interaction $C\times D$	$SS_{CD} = \frac{1}{abn}\sum Y_{..kl.}^2 - \frac{1}{abdn}\sum Y_{..k..}^2 - \frac{1}{abcn}\sum Y_{...l.}^2 + \frac{1}{N}Y_{......}^2$	$(c-1)(d-1)$
Interaction $A\times C\times D$	$SS_{ACD} = \frac{1}{bn}\sum Y_{i.kl.}^2 - \frac{1}{bdn}Y_{i.k..}^2 - \frac{1}{bcn}Y_{i..l.}^2 - \frac{1}{abn}Y_{..kl.}^2 + \frac{1}{bcdn}\sum Y_{i....}^2 + \frac{1}{abdn}\sum Y_{..k..}^2 + \frac{1}{abcn}\sum Y_{...l.}^2 - \frac{1}{N}Y_{.....}^2$	$(a-1)(c-1)(d-1)$
Residual of split-plots = Res	$SS_{Res} = SS_T - SS_A - SS_B - SS_{Res1} - SS_C - SS_D - SS_{AC} - SS_{AD} - SS_{CD} - SS_{ACD}$	$dfRes = N$-1 $-abn - acd + a + 1$
Corrected Total	$SS_T = \sum Y_{ijkl.}^2 - \frac{Y_{.....}^2}{N}$	N-1

N = abcdn with n ≥ 2.

Table 5. *Mean Squares, Expectations for the Split-plot Model with factor A on the Main-plots in Blocks B and F –Statistics.*

Mean Squares	**Expected Mean Squares**	***F*** **-Statistic**
$MS_A = \frac{SS_A}{a-1}$	$\sigma^2 + cdn\,\sigma_1^2 + \frac{bcdn}{a-1}\sum a_i^2$	$F_A = \frac{(a-1)(b-1)}{(a-1)}\frac{SS_A}{SS_{Res1}}$
$MS_B = \frac{SS_B}{b-1}$		
$MS_{Res1} = \frac{SSRes1}{dfRes1}$	$\sigma^2 + cdn\,\sigma^2{}_1$	
$MS_C = \frac{SS_C}{c-1}$	$\sigma^2 + \frac{abdn}{c-1}\sum c_k^2$	$F_C = \frac{dfRes}{(c-1)}\frac{SS_c}{SS_{res}}$
$MS_D = \frac{SS_D}{d-1}$	$\sigma^2 + \frac{abcn}{d-1}\sum d_l^2$	$F_D = \frac{dfRes}{(d-1)}\frac{SS_D}{SS_{res}}$
$MS_{AC} = \frac{SS_{AC}}{(a-1)(c-1)}$	$\sigma^2 + \frac{bdn}{(a-1)(c-1)}\sum (ac)_{ik}^2$	$F_{AC} = \frac{dfRes}{(a-1)(c-1)}\frac{SS_{AC}}{SS_{res}}$
$MS_{AD} = \frac{SS_{AD}}{(a-1)(d-1)}$	$\sigma^2 + \frac{bcn}{(a-1)(d-1)}\sum (ad)_{il}^2$	$F_{AD} = \frac{dfRes}{(a-1)(d-1)}\frac{SS_{AD}}{SS_{res}}$
$MS_{CD} = \frac{SS_{CD}}{(c-1)(d-1)}$	$\sigma^2 + \frac{abn}{(c-1)(d-1)}\sum (cd)_{kl}^2$	$F_{CD} = \frac{dfRes}{(a-1)(d-1)}\frac{SS_{CD}}{SS_{res}}$
$MS_{ACD} = \frac{SS_{ACD}}{(a-1)(c-1)(d-1)}$	$\sigma^2 + \frac{bn}{(a-1)(c-1)(d-1)}\sum (acd)_{ikl}^2$	$F_{ACD} = \frac{dfRes}{(a-1)(d-1)}\frac{SS_{ACD}}{SS_{res}}$
$MS_{Res} = SS_{Res}$ / dfRes	σ^2	

The calculation is demonstrated of the sample size minimin and maximin for the test of the factor A on the Main-plots of the Null hypothesis:

$$H_0: a_1 = a_2 = \cdots . = a_a .$$

Under the side condition $\sum_i a_i = 0$ the null hypothesis can also be formulated as

$$H_0: a_1 = a_2 = \cdots . = a_a = 0.$$

F–statistic: $F_A = \frac{dfRes1}{(a-1)} \frac{SS_A}{SS_{Res1}}$, with df ($A$) = ($a$-1) and $dfRes1 = abn - b - a + 1$.

Non-centrality parameter $\lambda_A = \frac{bcdn \sum_{i=1}^{a} a_i^2}{\sigma^2 + cdn\,\sigma^2 1}$.

For $\lambda_A = \frac{bcdn \sum_{i=1}^{a} a_i^2}{\sigma^2 + cdn\,\sigma^2 1}$ we have $\lambda_{A\,max} = \frac{bcdn\delta^2}{4(\sigma^2 + cdn\,\sigma^2 1)}$ and $\lambda_{A\,min} = \frac{bcd\delta^2}{2(\sigma^2 + cdn\,\sigma^2 1)}$ with $a_{\max} - a_{\min} \geq \delta$.

For the non-centrality parameter λ_A we use in the R–program

$$\text{delta} = \delta/\sqrt{(\sigma^2 + cdn\sigma^2_1)}.$$

A program in R which gives the solution of formula for the sample size minimin of the fixed effect A on the Main-plots is:

```
> miniminMP1A = function( p, a, b, c, d, delta , beta)
{
f = function(n, p, a, b, c, d, delta , beta )
{
A = qf(p, a-1, a*b*n-b-a+1, 0)
B = qf( beta, a-1, a*b*n-b-a+1, delta*delta*b*c*d/2)
C = A-B
}
k = uniroot(f, c(2, 10000 ), p=p, a=a, b=b, c=c, d=d, delta=delta, beta=beta) $root
k0 = ceiling(k)
print (paste(" miniminMP1A sample number: n = ", k0), quote=F)
}
```

Example 8

where p = 1 – α :

```
> miniminMP1A( p=0.95, a=3, b = 2, c = 2, d= 2, delta = 2, beta = 0.05)
[1] miniminMP1A sample number: n = 15
```

A program in R which gives the solution of formula (6) for the sample size maximin of the fixed effect A on the Main-plots is:

```
> maximinMP1A = function( p, a, b, c, d, delta , beta)
{
f = function(n, p, a, b, c, d, delta , beta )
{
A = qf(p, a-1, a*b*n-b-a+1, 0)
B = qf( beta, a-1, a*b*n-b-a+1, delta*delta*n*b*c*d/4)
C = A-B
}
k = uniroot(f, c(2, 10000 ), p=p, a=a, b=b, c=c, d=d, delta=delta, beta=beta) $root
k0 = ceiling(k)
print (paste(" maximinMP1A sample number: n = ", k0), quote=F)
}
```

Example 9

where $p = 1 - \alpha$:

```
> maximinMP1A( p=0.95, a=3, b= 2, c= 2, d=2, delta = 2, beta = 0.05)
[1] maximinMP1A sample number: n = 3
```

The calculation is demonstrated of the sample size minimin and maximin for the test of the factor C on the Split-plots of the Null hypothesis:

$$H_0: c_1 = c_2 = \cdots . = c_c .$$

Under the side condition $\sum_i c_i = 0$ the null hypothesis can also be formulated as

$$H_0: c_1 = c_2 = \cdots . = c_c = 0.$$

F –statistic: $F_C = \frac{dfRes}{(c-1\;)} \frac{SS_C}{SS_{Res}}$ with df (C) = (c-1) and

$$dfRes = abcdn\text{-}1\text{-}abn\text{-}acd\text{+}a\text{+}1 .$$

Non-centrality parameter $\lambda_C = \frac{abdn \sum_{i=1}^{c} c_i^2}{\sigma^2}$.

For $\lambda_C = \frac{abdn \sum_{i=1}^{c} c_i^2}{\sigma^2}$ we have $\lambda_{C\,max} = \frac{abdn\delta^2}{4\sigma^2}$ and $\lambda_{C\,min} = \frac{abd\delta^2}{2\sigma^2}$ with

$$c_{max} - c_{min} \geq \delta .$$

For the non-centrality parameter λ_C in the R–program is used delta = δ/σ.

A program in R which gives the solution of formula for the sample size minimin of the fixed effect C on the Split-plots is:

```
> miniminSP1C = function( p, a, b, c, d, delta , beta)
{
f = function(n, p, a, b, c, d, delta , beta )
{
dfRes = a*b*c*d*n-1- a*b*n-a*c*d+a+1
A = qf(p, c-1, dfRes, 0)
B = qf( beta, c-1, dfRes, delta*delta *a*b*d/2)
C = A-B
}
k = uniroot(f, c(2, 10000 ), p=p, a=a, b=b, c=c, d=d, delta=delta, beta=beta) $root
k0 = ceiling(k)
print (paste(" miniminSP1C sample number: n = ", k0), quote=F)
}
```

Example 10

where $p = 1 - \alpha$:

```
> miniminSP1C( p=0.95, a=3, b = 2, c = 2, d= 2, delta = 1.5, beta = 0.05)
[1] miniminSP1C sample number: n = 4
```

A program in R which gives the solution of formula (6) for the sample size maximin of the fixed effect C on the Split-plots is:

```
> maximinSP1C = function( p, a, b, c, d, delta , beta)
{
f = function(n, p, a, b, c, d, delta , beta )
```

```
{
dfRes = a*b*c*d*n-1-a*b*n-a*c*d+a+1
A = qf(p, c-1, dfRes, 0)
B = qf( beta, c-1, dfRes, delta*delta*n*a*b*d/4)
C = A-B
}
k = uniroot(f, c(2, 10000), p=p, a=a, b=b, c= c, d=d, delta=delta, beta=beta) $root
k0 = ceiling(k)
print (paste(" maximinSP1C sample number: n = ", k0), quote=F)
}
```

Example 11

where $p = 1 - \alpha$:

```
> maximinSP1C( p=0.95, a=3, b= 2, c= 2, d=2, delta = 1, beta = 0.05)
[1] maximinSP1C sample number: n = 5
```

For the test of the factor D on the Split-plots a change must done analogously in the R program according to the non-centrality parameter λ derived from the E(MS) column of the ANOVA-table and in the R-program is used delta = δ/σ.

5. Minimum Number of Replications for Testing in Split-plot Designs with *A* and *C* at the Main-plots

Consider the case that the fixed factors A and C has been laid down on Main-plots in a balanced randomized block design with the factor B as Blocks with acn Main-plots. The fixed factor D sizes $n \geq 2$ is laid down in the Split-plots of the Main-plots.

The four-way cross classification is based on the model (with all interactions) here written for the case where the factors A, C and D are fixed and the factor B as Blocks; the factors A and C are laid down on the Main-plots of the Blocks B: otherwise stated AxC is nested in B, hence $B \succ A$xC; the factor D is laid down on the Split-plots of the Main-plots of the Split-Plot design, hence Ax$C \succ D$.

The fixed factors A, C and D have equal sizes of $n \geq 2$ replications.

$$y_{ijklm} = \mu + b_j + a_{i(j)} + c_{k(j)} + (ac)_{ik(j)} + \boldsymbol{p}_{\boldsymbol{m(ijk)}} + d_{l(ijk)} + (ad)_{il(ijk)} + (cd)_{kl(ijk)} + (acd)_{ikl(ijk)} + \boldsymbol{e}_{\boldsymbol{m(ijkl)}} \quad (9)$$

$$i = 1, \ldots, a, j = 1, \ldots, b, k = 1, \ldots, c, l = 1, \ldots, d, m = 1, \ldots, n \geq 2.$$

The first part of (9) are the Main-plot effects and the second part of (9) are the Split-plot effects.

Assumed is that all fixed effects in model 1 sum up to zero, when summation is done over at least one of the superscripts of the effect. For instance $\sum_{k=1}^{b}(acd)_{ikl}=0$. The random effects for Main-plot error $\boldsymbol{p}_{\boldsymbol{m(ijk)}}$ and random effects for Split $-$ plot error $\boldsymbol{e}_{m(ijkl)}$ are assumed to be independent and normally distributed with expectation zero and with variances respectively σ^2_1 and σ^2.

The ANOVA – table and the expected Mean Squares E(MS) are given below in Table 6 and Table 7 respectively.

Table 6. *Analysis of Variance Table of a Split-plot design with factor A and C on Main-plots in Blocks B and equal Split-plot Numbers n ≥ 2.*

Source of Variation	***SS***	***df***
Between B- levels	$SS_B = \frac{1}{acdn}\sum Y^2_{.j....} - \frac{1}{N}Y^2_{......}$	$b-1$
Between A-levels	$SS_A = \frac{1}{bcdn}\sum Y^2_{i.....} - \frac{1}{N}Y^2_{......}$	$a-1$
Between C- levels	$SS_C = \frac{1}{abdn}\sum Y^2_{..k..} - \frac{1}{N}Y^2_{......}$	$c-1$
Interaction AxC	$SS_{AC} = \frac{1}{bdn}\sum Y^2_{i.k..} - \frac{1}{bcdn}\sum Y^2_{i.....} - \frac{1}{abdn}\sum Y^2_{..k..} + \frac{1}{N}Y^2_{......}$	$(a-1)(c-1)$
Residual p of Main-Plots = Res1	$SS_{Res1} = \frac{1}{acn}\sum Y^2_{ij...} - \frac{1}{N}Y^2_{......} - SS_B\text{-}SS_A\text{-}SS_C\text{-}SS_{AC}$	*dfRes1= abcn-ac-b+1*
Between D- levels	$SS_D = \frac{1}{abcn}\sum Y^2_{...m.} - \frac{1}{N}Y^2_{......}$	$d-1$
Interaction AxD	$SS_{AD} = \frac{1}{bcn}\sum Y^2_{i..l.} - \frac{1}{bcdn}\sum Y^2_{i.....} - \frac{1}{abcn}\sum Y^2_{...l..} + \frac{1}{N}Y^2_{......}$	$(a-1)(d-1)$
Interaction CxD	$SS_{CD} = \frac{1}{abn}\sum Y^2_{..kj.l.} - \frac{1}{abdn}\sum Y^2_{..k..} - \frac{1}{abcn}\sum Y^2_{...l.} + \frac{1}{N}Y^2_{......}$	$(c-1)(d-1)$
Interaction $AxCxD$	$SS_{ACD} = \frac{1}{bn}\sum Y^2_{i.kl.} - \frac{1}{bdn}Y^2_{i.k..} - \frac{1}{bcn}Y^2_{i..l.} - \frac{1}{abn}Y^2_{..kl.}$	$(a-1)(c-1)(d-1)$

Source of Variation	SS	df
	$+\frac{1}{bcdn}\sum Y_{i....}^2+\frac{1}{abdn}\sum Y_{..k..}^2+\frac{1}{abcn}\sum Y_{...l.}^2-\frac{1}{N}Y_{.....}^2$	
Residual of split-plots = Res	$SS_{Res} = SS_T - SS_A - SS_B - SS_C - SS_{AC} - SS_{Res1} - SS_D - SS_{AD} - SS_{CD} - SS_{ACD}$	*dfRes = N- 1 - abcn - acd + ac+1*
Corrected Total	$SS_T = \sum Y_{ijkl.}^2 - \frac{Y_{.....}^2}{N}$	*N*-1

N = *abcdn* with *n* ≥ 2.

Table 7. *Mean Squares, Expectations for the Split-Plot Model with factor A and C on the Main-Plots and F –Statistics.*

Mean Squares	Expected Mean Squares	*F* -Statistic
$MS_A = \frac{SS_A}{a-1}$	$\sigma^2 + cdn\,\sigma_1^2 + \frac{bcdn}{a-1}\sum a_i^2$	$F_A = \frac{dfRes1}{a-1}\frac{SS_A}{SS_{Res1}}$
$MS_B = \frac{SS_B}{b-1}$		
$MS_C = \frac{SS_C}{c-1}$	$\sigma^2 + cdn\,\sigma_1^2 + \frac{abdn}{c-1}\sum c_k^2$	$F_C = \frac{dfRes1}{c-1}\frac{SS_C}{SS_{Res1}}$
$MS_{Res1} = \frac{SS_{es1}}{dfRes1}$	$\sigma^2 + cdn\,\sigma_1^2$	
$MS_D = \frac{SS_D}{d-1}$	$\sigma^2 + \frac{abcn}{d-1}\sum d_l^2$	$F_D = \frac{dfRes}{d-1}\frac{SS_D}{SS_{res}}$
$MS_{AC} = \frac{SS_{AC}}{(a-1)(c-1)}$	$\sigma^2 + \frac{bdn}{(a-1)(c-1)}\sum (ac)_{ik}^2$	$F_{AC} = \frac{dfRes}{(a-1)(c-1)}\frac{SS_{AC}}{SS_{res}}$
$lMS_{AD} = \frac{SS_{AD}}{(a-1)(d-1)}$	$\sigma^2 + \frac{bcn}{(a-1)(d-1)}\sum (ad)_{il}^2$	$F_{AD} = \frac{dfRes}{(a-1)(d-1)}\frac{SS_{DA}}{SS_{res}}$
$MS_{CD} = \frac{SS_{CD}}{(c-1)(d-1)}$	$\sigma^2 + \frac{abn}{(c-1)(d-1)}\sum (cd)_{kl}^2$	$F_{CD} = \frac{dfRes}{(c-1)(d-1)}\frac{SS_{CD}}{SS_{res}}$
$MS_{ACD} = \frac{SS_{ACD}}{(a-1)(c-1)(d-1)}$	$\sigma^2 + \frac{bn}{(a-1)(c-1)(d-1)}\sum (acd)_{ikl}^2$	$F_{ACD} = \frac{dfRes}{(a-1)(c-1)(d-1)}\frac{SS_{ACD}}{SS_{res}}$
$MS_{Res} = SS_{Res}$ / dfRes	σ^2	

The calculation is demonstrated of the sample size minimin and maximin for the test of the factor *A* on the Main-plots of the Null hypothesis: $H_0: a_1 = a_2 = \cdots . = a_a$.

Under the side condition $\sum_i a_i = 0$ the null hypothesis can also be formulated as $H_0: a_1 = a_2 = \cdots . = a_a = 0$.

F–statistic: $F_A = \frac{dfRes1}{(a-1)}\frac{\boldsymbol{SS_A}}{\boldsymbol{SS_{Res1}}}$, with df (*A*) = (*a*-1) and *dfRes1= abcn-ac-b+1.*

Non-centrality parameter $\lambda_A = \frac{bcdn\sum_{i=1}^{a} a_i^2}{\sigma^2 + cdn\,\sigma^2 1}$.

For $\lambda_A = \frac{bcdn\sum_{i=1}^{a} a_i^2}{\sigma^2 + cdn\,\sigma^2 1}$ we have $\lambda_{A\,max} = \frac{bcdn\delta^2}{4(\sigma^2 + cdn\,\sigma^2 1)}$ and $\lambda_{A\,min} = \frac{bcd\delta^2}{2(\sigma^2 + cdn\,\sigma^2 1)}$ with $a_{\max} - a_{\min} \geq \delta$.

For the non-centrality parameter λ_A is used in the R–program delta = $\delta/\sqrt{(\sigma^2 + cdn\sigma^2_1)}$.

A program in R which gives the solution of formula for the sample size minimin of the fixed effect *A* on the Main-plots is:

```
> miniminMP2A = function( p, a, b, c, d, delta , beta)
{
f = function(n, p, a, b, c, d, delta , beta )
{ A = qf(p, a-1, a*b*c*n-a*c-b+1 , 0)
B = qf( beta, a-1, a*b*c*n-a*c-b+1, delta*delta*b*c*d/2)
C = A-B
}
k = uniroot(f, c(2, 10000 ), p=p, a=a, b=b, c=c, d=d, delta=delta, beta=beta) $root
k0 = ceiling(k)
print (paste(" miniminMP2A sample number: n = ", k0), quote=F)
}
```

Example 12

where $p = 1 - \alpha$:

```
> miniminMP2A( p=0.95, a=3, b = 2, c = 2, d= 2, delta = 2, beta = 0.05)
[1] miniminMP2A sample number: n = 8
```

A program in R which gives the solution of formula (6) for the sample size maximin of the fixed effect *A* on the Main-plots is:

```
> maximinMP2A = function( p, a, b, c, d, delta , beta)
{
f = function(n, p, a, b, c, d, delta , beta )
{
A = qf(p, a-1, a*b*c*n-a*c-b+1, 0)
B = qf( beta, a-1, a*b*c*n-a*c-b+1, delta*delta*n*b*c*d/4)
C = A-B
}
```

```
k = uniroot(f, c(2, 10000), p=p, a=a, b=b, c= c, d=d,
delta=delta, beta=beta) $root
k0 = ceiling(k)
print (paste(" maximinMP2A sample number: n = ", k0),
quote=F)
}
```

Example 13

where $p = 1 - \alpha$:

```
> maximinMP2A( p=0.95, a=3, b= 2, c= 2, d=2, delta = 2,
beta = 0.05)
[1] maximinMP2A sample number: n = 3
```

For the test of the factor C on the Main-plots we must change analogously the R program according to the non-centrality parameter λ derived from the E(MS) column of the ANOVA-table and in the R-program we use *delta* = $\delta/\sqrt{(\sigma^2 + cdn\sigma^2_1)}$.

Now the calculation is demonstrated of the sample size minimin and maximin for the test of the factor D on the Split-plots of the Null hypothesis:

H_0: $d_1 = d_2 = \cdots . = d_d$.

Under the side condition $\sum_i d_i = 0$ the null hypothesis can also be formulated as H_0: $d_1 = d_2 = \cdots . = d_d = 0$.

F –statistic: $F_D = \frac{dfRes}{(d-1\ )}\frac{SS_D}{SS_{res}}$ with df (D) = (d-1) and df (Res) = abcdn -1-abcn-acd+ac+1.

Non-centrality parameter $\lambda_D = \frac{abcn\sum_{i=1}^{d} d_i^2}{\sigma^2}$.

For $\lambda_D = \frac{abcn\sum_{i=1}^{D} d_i^2}{\sigma^2}$ we have $\lambda_{D\ max} = \frac{abcn\delta^2}{4\sigma^2}$ and $\lambda_{D\ min} = \frac{abcn\delta^2}{2\sigma^2}$ with $d_{max} - d_{min} \geq \delta$.

For the non-centrality parameter λ_D is used in the R–program *delta* = δ/σ.

A program in R which gives the solution of formula for the sample size minimin of the fixed effect D on the Split-plots is:

```
> miniminSP2D = function( p, a, b, c, d, delta , beta)
{
f = function(n, p, a, b, c, d, delta , beta )
{
A = qf(p, d-1, a*b*c*d*n-1-a*b*c*n-a*c*d + a*c +1, 0)
B = qf( beta, d-1, a*b*c*d*n-1-a*b*c*n-a*c*d + a*c +1 ,
delta*delta*n *a*b*d/4)
C = A-B
}
k = uniroot(f, c(2, 10000 ), p=p, a=a, b=b, c=c, d=d,
delta=delta, beta=beta) $root
k0 = ceiling(k)
print (paste(" miniminSP2D sample number: n = ", k0),
quote=F)
}
```

Example 14

where $p = 1 - \alpha$:

```
> miniminSP2D( p=0.95, a=3, b = 2, c = 2, d= 2, delta = 1,
beta = 0.05)
[1] miniminSP2D sample number: n = 5
```

A program in R which gives the solution of formula (6) for the sample size maximin of the fixed effect D on the Split-plots is:

```
> maximinSP2D = function( p, a, b, c, d, delta , beta)
{
f = function(n, p, a, b, c, d, delta , beta )
{
A = qf(p, d-1, a*b*c*d*n-1-a*b*c*n-a*c*d + a*c +1 , 0)
B = qf( beta, d-1, a*b*c*d*n-1-a*b*c*n-a*c*d + a*c +1,
0.5* delta*delta*n*a*b*d)
C = A-B
}
k = uniroot(f, c(2, 10000), p=p, a=a, b=b, c= c, d=d,
delta=delta, beta=beta) $root
k0 = ceiling(k)
print (paste(" maximinSP2D sample number: n = ", k0),
quote=F)
}
```

Example 15

where $p = 1 - \alpha$:

```
> maximinSP2D( p=0.95, a=3, b= 2, c= 2, d=2, delta = 1,
beta = 0.05)
[1] maximinSP2D sample number: n = 3
```

6. Conclusion

This paper gives an extension of the literature about balanced cross-classification. Till now the maximum of three factors were considered in balanced designs to calculate the minimum number of replications for a fixed factor. Now for the four factor balanced crossed design and split-plot design the minimum number of replications can be calculated with provided small R-programs. This means that there is now an extension of the R-package OPDOE for calculating the minimum size of a fixed factor for testing with a certain power.

References

[1] Kuehl, R. O. (1994), Statistical Principles of Research Design and Analysis", Duxburry Press, Belmont, CA.

[2] Kutner, M. H., Nachtsheim. C. J., Neter, J. and Li, W. (2005), Applied Linear Statistical Models, 5th edition, McGraw-Hill Irwin, New York.

[3] Lindman, H. R. (1992), Analysis of Variance in Experimental Design, Springer-Verlag, New York.

[4] Ott, R. L. and Longnecker, M. (2001), An Introduction to Statistcal Methods and Data Analyis, 5th edition, Duxbury, Pacific Grove, CA.

[5] Owen, D. B. (1962), Handbook of Statistical Tables, Addison-Wesley Publishing Company, Reading, MA.

[6] Pearson, E. S. and Hartley, H. O. (1972), Biometrika Tables for Statisticians, Volume II, Cambridge, The University Press.

[7] Rasch, D., Herrendörfer, G., Bock, J., Victor, N. and Guiard, V. (2008), Verfahrensbibliothek Versuchsplanung und - auswertung, 2. verbesserte Auflage in einem Band mit CD, R. Oldenbourg Verlag, München, Wien.

[8] Rasch, D. und Schott, D. (2016), Mathematische Statistik, Wiley – VCH, Weinheim.

[9] Rasch, D. and Schott, D. (2018), Mathematical Statistics, Wiley, Oxford.

[10] Rasch, D., Kubinger, K.D. and Yanagida, T. (2011), Statistics in Psychology using R and SPSS, Wiley, New York.

[11] Rasch, D., Pilz, J., Verdooren, R. and Gebhardt, A. (2011), Optimal experimental design with R, Chapman and Hall, London.

[12] Rasch, D., Spangl, B. and Wang, M. (2012), Minimal experimental size in the three way ANOVA cross classification model with approximate F-tests, Communications in Statistics- Simulation and Computation 41 (7): 1120–1130.

[13] Rasch, D. and Verdooren, R. (2020), Determination of Minimum and Maximum Experimental Size in One-, Two- and Three-way ANOVA with Fixed and Mixed Models by R. *Journal of Statistical Theory and Practice* 14, 57 (2020). https://doi.org/10.1007/s42519-020-00088-6.

[14] Rasch, D, Verdooren, R. and Pilz, R. (2019), Applied Statistics, Wiley, Oxford.

[15] Spangl, B., Kaiblinger, N., Ruckdeschel, P. and Rasch, D. (2021), Minimum sample size in balanced ANOVA models of crossed, nested and mixed classifications, Communications in Statistics – Theory and Methods, https://doi.org/10.1080/03610926.2021.1938126.

[16] Tang, P.C. (1938), The power function of the analysis of variance tests with tables and illustrations of their use, *Statistical Research Memoirs,* 2, 126-149.

[17] Wang, M., Rasch, D. and Verdooren, R. (2005), Determination of the size of a balanced experiment in mixed ANOVA models using the modified approximate F-test. Journal of Statistical Planning and Inference 132 (1-2): 183-201.