Agricultural Field Trials – Today and Tomorrow

Hohenheim – October 8th to 10th, 2007

Wiesław Pilarczyk

The extent and prevailing shape of spatial relationships in Polish variety testing trials on cereals

Two sets of data:

I)

- series of 122 1-resolvable incomplete block trials conducted in years 1998 and 1999, four replicates, winter and spring wheat, number of varieties varied from 27 to 48

II)

- series of 58 incomplete split-block trials built on the basis of 1-resolvable incomplete block by longwise splitting of every superblock , two replicates, winter wheat

  period: 2002, 2004 and 2005

  year 2003 – removed (hard winter demages)

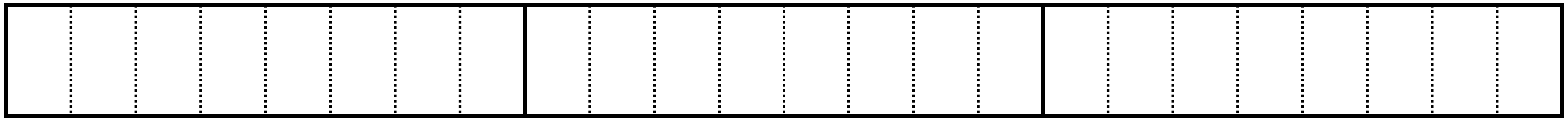
  varieties: 44, 42 and 50 respectively

  second factor - two levels of fertilizing (and protecting)

An example of 1-resolvable incomplete block design with three replicates

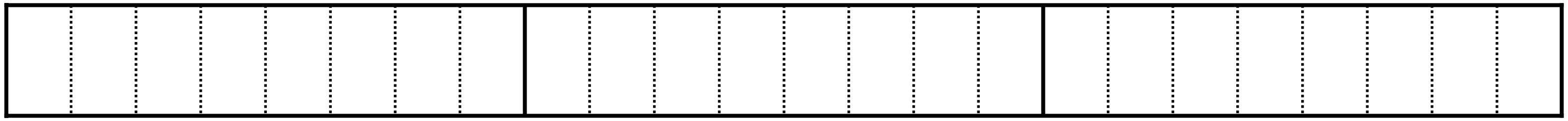
Replicate (superblock) 1

block 1 block 2 block 3



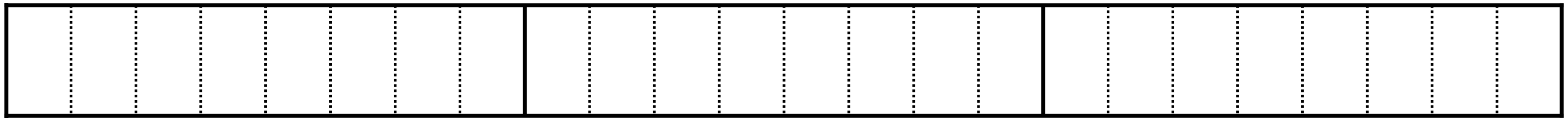
Replicate (superblock) 2

block 4 block 5 block 6



Replicate (superblock) 3

block 7 block 8 block 9



An example of incomplete split-block design in two replicates

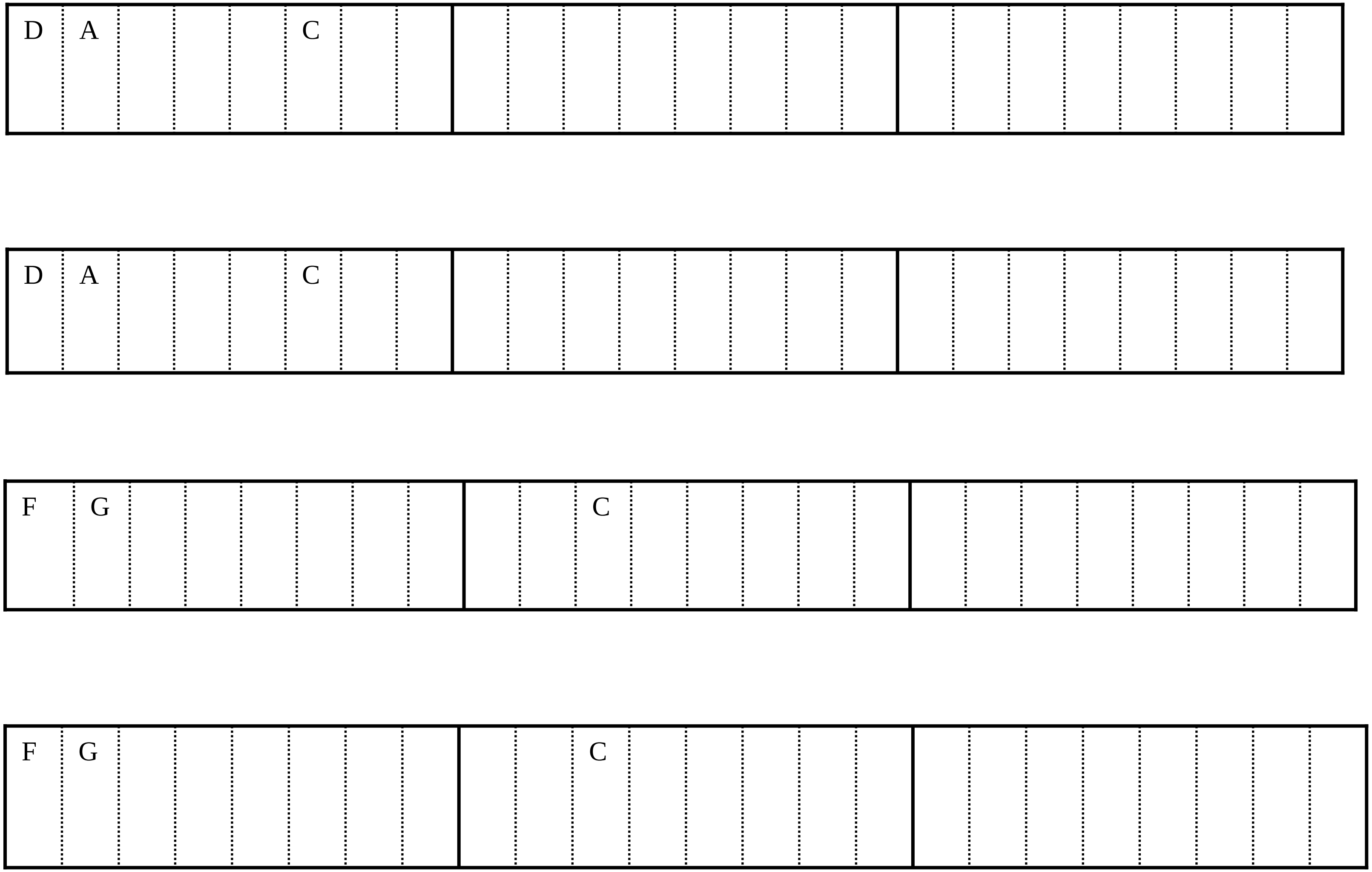
Replicate 1

block 1								block 2								block 3							
D	A				C																		
D	A				C																		

Replicate 2

block 4								block 5								block 6							
F	G									C													
F	G									C													

- Field scheme of incomplete split-block



- Application of 1-resolvable designs in variety testing trials in Poland
- Application of incomplete split-block designs
- The assumptions of analysis of variance
- The alternative approach: so-called spatial methods and their assumptions
- Geostatistical methods

Semivariance

$$\gamma(h) = \frac{1}{2}\operatorname{var}\left[z(x) - z(x+h)\right]$$

$$\hat{\gamma}(h) = \frac{1}{2m(h)}\sum_{i=1}^{m(h)}\left[z(x_i) - z(x_i + h)\right]^2$$

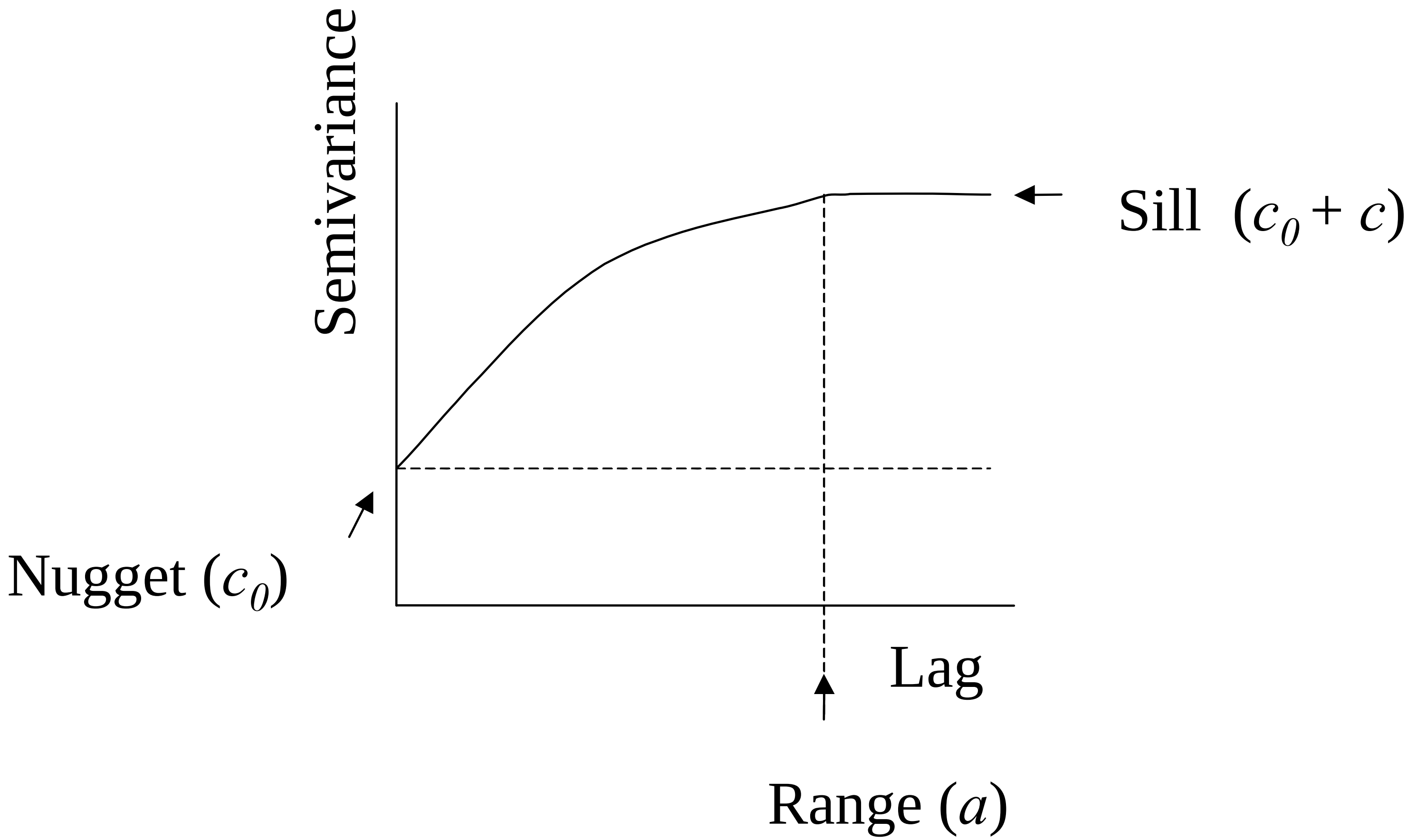


Fig 2. General form of semivariogram

The formula used in practice

$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{j=1}^{r} \sum_{i=1}^{v-h} \left(e_{ij} - e_{i+h,j}\right)^2,$$

where m(h) denotes the number of cases that difference $(e_{ij} - e_{i+h,j})$ exists, and e_{ij} denotes the residual received for i-th plots within j-th superblock from randomized complete block analysis of variance model.

The software used:

Procedure **ADOJ** – for analysis of variance and calculating residuals

Procedures: **FVARIOGRAM** and **MVARIOGRAM** of GENSTAT package – for calculating semivariances and estimating semivariograms (by weighted least squares method)

Semivariogram mathematical models

1) Spherical model

$$\gamma(h)=\begin{cases}0 & \text{for } h=0\\ c_0+c\left[\frac{3}{2}\frac{h}{a}-\frac{1}{2}\left(\frac{h}{a}\right)^3\right] & \text{for } 0<h\le a.\\ c_0+c & \text{for } h>a\end{cases}$$

2) Boundedlinear model

$$\gamma(h)=\begin{cases}0 & \text{dla } h=0\\ c_0+c\left(\frac{h}{a}\right) & \text{dla } 0<h\le a.\\ c_0+c & \text{dla } h>a\end{cases}$$

3) Pentaspherical model

$$\gamma(h)=\begin{cases}0 & \text{for } h=0\\ c_0+c\left[\frac{15}{8}\frac{h}{a}-\frac{5}{4}\left(\frac{h}{a}\right)^3+\frac{3}{8}\left(\frac{h}{a}\right)^5\right] & \text{for } 0<h\le a.\\ c_0+c & \text{for } h>a\end{cases}$$

4) Circular model

$$\gamma(h)=\begin{cases}0 & \text{for } h=0\\ c_0+c\left(1-\frac{2}{\pi}\arccos\left(\frac{h}{a}\right)+\frac{2h}{\pi a}\sqrt{1-\left(\frac{h}{a}\right)^2}\right) & \text{for } 0<h\le a.\\ c_0+c & \text{for } h>a\end{cases}$$

5) Doublespherical model

$$\gamma(h)=\begin{cases} 0 & \text{for } h=0 \\ c_0+c_1\left[\frac{3}{2}\frac{h}{a_1}-\frac{1}{2}\left(\frac{h}{a_1}\right)^3\right]+c_2\left[\frac{3}{2}\frac{h}{a_2}-\frac{1}{2}\left(\frac{h}{a_2}\right)^3\right] & \text{for } 0<h\le a_1 \\ c_0+c_1+c_2\left[\frac{3}{2}\frac{h}{a_2}-\frac{1}{2}\left(\frac{h}{a_2}\right)^3\right] & \text{for } a_1<h\le a_2 \\ c_0+c_1+c_2 & \text{for h > a} \end{cases}$$

6) Exponential model

$$\gamma(h) = \begin{cases} 0 & \text{for } h = 0 \\ c_0 + c\left[1 - \exp\left(-\frac{h}{r}\right)\right] & \text{for } 0 < h \end{cases}$$

7) Gaussian model

$$\gamma(h) = \begin{cases} 0 & \text{for } h = 0 \\ c_0 + c\left[1 - \exp\left(-\frac{h}{r}\right)^2\right] & \text{for } 0 < h \end{cases}$$

8) Besselk$_1$ model

$$\gamma(h) = \begin{cases} 0 & \text{for } h = 0 \\ c_0 + c\left[1 - \frac{h}{r} K_1\left(\frac{h}{r}\right)\right] & \text{for } 0 < h \end{cases}.$$

where K_1 is a modified Bessel function of the second kind

9) Power model

$$\gamma(h) = \begin{cases} 0 & \text{for } h = 0 \\ c_0 + gh^{\alpha} & \text{for } 0 < h \end{cases}.$$

where $0 < \alpha < 2$

10) Linear model

$$\gamma(h) = \begin{cases} 0 & \text{for } h = 0 \\ c_0 + gh & \text{for } 0 < h \end{cases}.$$

Table 1. Number of successfully ended estimations of 10 models in 122 trials and number of cases that the model was the best among all considered (first data set)

Model	Year		In total	The number of cases that the model was the best
	1998	1999		
Spherical	34	22	56	5
Bounded-linear	43	29	72	25
Circular	35	23	58	8
Pentaspherical	34	23	57	3
Double-spherical	7	5	12	1
Exponential	33	25	58	4
Gauss	32	24	56	13
Bessel	44	34	78	4
Power	31	24	55	7
Linear	51	41	92	33

Linear model was the best in 33 trials

Bounded linear one – in 25 trials

At least one of two linear models was successfully fitted 92 out of 98 trials and was the best in 58 trials

Coefficients of determination (for the best model) were as follows:

$R^2 \le 30\%$ - in 18 trials

$30 < R^2 \le 70\%$ - in 39 trials

$R^2 > 70\%$ - in 41 trials

Table 2. Number of successfully ended estimations of 10 models in 58 trials and number of cases that the model was the best among all considered (second data set)

Model	Year			In total	The number of cases that the model was the best
	2002	2004	2005		
Spherical	8	10	7	25	2
Bounded-linear	4	6	3	13	7
Circular	7	9	7	23	9
Pentaspherical	7	8	5	20	1
Double-spherical	1	0	0	1	0
Exponential	8	10	8	26	1
Gauss	7	8	9	24	5
Bessel	11	12	9	32	1
Power	6	9	5	20	2
Linear	15	16	14	45	21

In the second data set, two linear models were the best in 28 trial out of 49 (57%), in the first set it was 63%

The Bessel model fitted in 32 trials, was the best only ones

Coefficients of determination (for the best fitted model) in the second set were:

$R^2 \leq 30\%$ in 20 trials

$30 < R^2 \leq 70\%$ in 19 trials

$R^2 > 70\%$ in 8 trials

Mixed model

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z\gamma} + \boldsymbol{\varepsilon}$$

y – denotes the $n \times 1$ dimensional vector of observations

β – is $p \times 1$ vector of unknown fixed effects

X – denotes known $n \times p$ design matrix for fixed effects

ε – is $n \times 1$ vector of errors

γ – denotes $b \times 1$ vector of unknown random effects

Z – means $n \times b$ design matrix associated with γ

γ i **ε** – have normal distribution with expectation

$$\mathrm{E}\begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

and variance-covariance matrix

$$\mathrm{Var}\begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}$$

The covariance matrix of **y** is **V**

$$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$$

where **G** has a form

$$\mathbf{G} = \mathbf{I}_{\mathrm{b}}\sigma^2_{\mathrm{B}}$$

where $\mathbf{I}_{\mathrm{b}}$ denotes b x b unit matrix

and σ^2_{B} is variance component for blocks.

The matrix **R** has linear spatial structure, where elements r_{ij} of it

have a form

$$r_{ij} = \begin{cases} \sigma^2\left(1 - \rho d_{ij}\right) & \text{dla } \rho d_{ij} \leq 1 \\ 0 & \text{dla } \rho d_{ij} > 1 \end{cases}$$

where d_{ij} denotes the distance between *i-th and* i *j-th* observations and

ρ is a parameter of spatial range of covariance

In analyzed trials

β – denotes the vector of variety means,

X – the design matrix for varieties,

Z – is the design matrix for blocks (if treated as random) (in reverse situation the columns of **Z** added to columns of matrix **X**),

γ – denotes the vector of random effects of blocks, with normal, independent distributions $N\left(0,\sigma_B^2\right)$ (when block effects are treated as fixed they are added to the vector **β**), All the trials belonging to the first set were re-analysed using five following models (using MIXED procedure of SAS system):

Model**1** – RCB model (blocks assumed fixed) with linear structure of spatial relationship added,

Model**2** – 1-resolvable incomplete block model (blocks fixed) with linear structure (within incomplete blocks) of spatial relationship added,

Model**3** – RCB model (blocks random) with linear structure of spatial relationship added,

Model**4** – 1-resolvable incomplete block model (blocks random) with linear structure of spatial relationships added,

Model**5** – 1-resolvable incomplete block model.

The relative efficiency RE of a given version of analysis was then calculated as a ratio of average standard deviation of treatment comparisons in randomised complete block model to the same average standard deviation in a model under consideration

Table 4

Relative efficiency of analysis under different models

Trial code	RE for model 1	RE for model 2	RE for model 3	RE for model 4	RE for model 5
ba1739	2,09	1,96	2,13	1,98	1,46
aa1017	2,03	1,90	2,07	1,87	1,71
ba1113	1,93	1,74	1,97	1,81	1,27
ba0830	1,87	1,75	1,90	1,73	1,26
aa1146	1,83	1,75	1,86	1,74	1,44
aa1239	1,77	1,75	1,80	1,66	1,66
aa1188	1,76	1,80	1,78	1,76	1,50
ba1268	1,67	1,62	1,67	1,62	1,24
ba1146	1,64	1,54	1,65	1,56	1,31
aa1282	1,60	1,14	1,62	1,44	1,48
aa1012	1,59	1,52	1,60	1,54	1,18
aa1088	1,57	1,70	1,58	1,65	1,44
aa1100	1,51	1,27	1,52	1,48	1,23
ba1013	1,50	1,30	1,51	1,37	1,08
aa1134	1,47	1,30	1,50	1,32	1,27
aa0961	1,47	1,36	1,49	---	1,12
aa0796	1,46	1,30	1,47	1,17	1,10
. . .	. . .	. . .	. . .	. . .	. . .
On average	**1,31**	**1,21**	**1,32**	**1,23**	**1,11**

Conlusions

- spatial relationships exist in majority of wheat trial,

- this relationship is satisfactorily approximated by linear form of underlying semivariogram,

- further analyses are needed to confirm (or to deny) these statements,

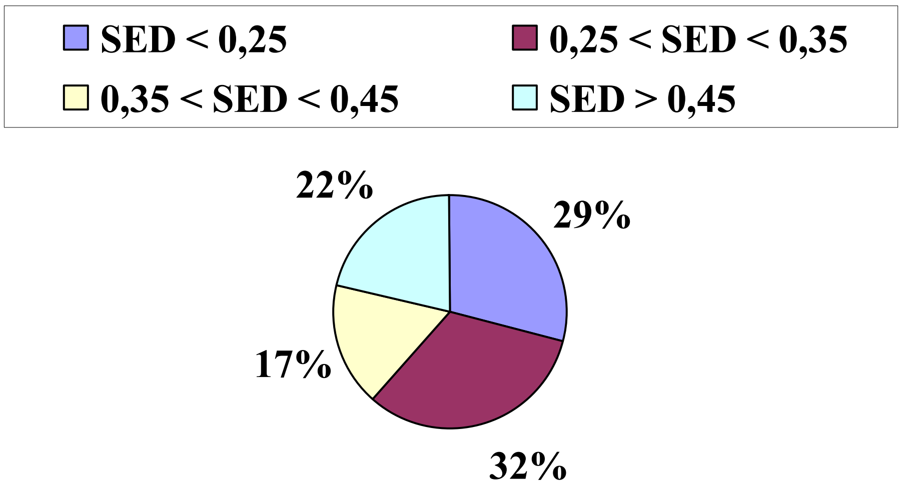
- incorporating revealed relationships into mathematical model of observation can lead to more efficient analysis of trial results (as shown also by Tomaszyk (2005)),

- the linear structure of spatial dependence added to incomplete block model has not resulted with significant increase the efficiency of the analysis. The average efficiency was smaller than average efficiency of randomized complete block with spatial dependence added.

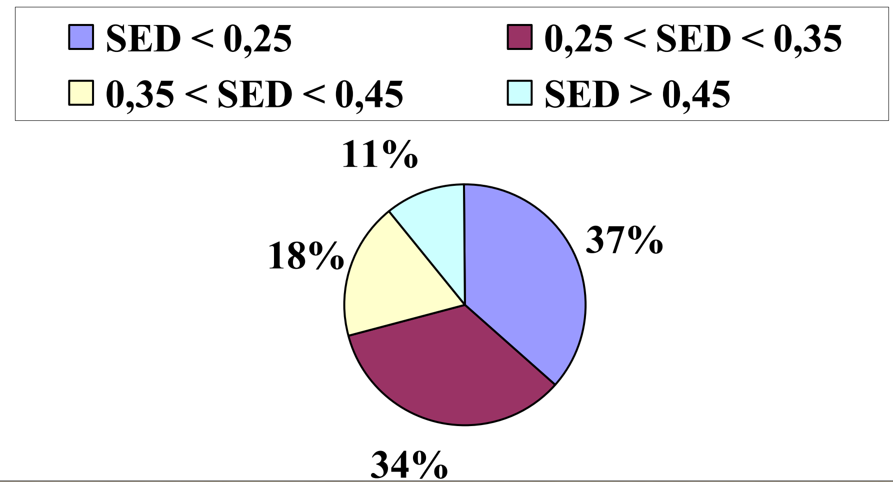
Thank you for your attention

Fractions of trials with different average values of SED (standard error of difference) depending on used mathematical model of observations

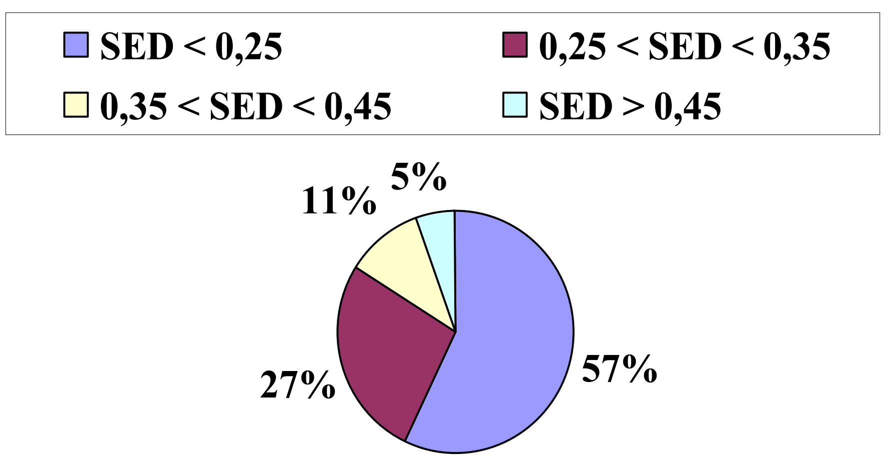
The randomized complete block model



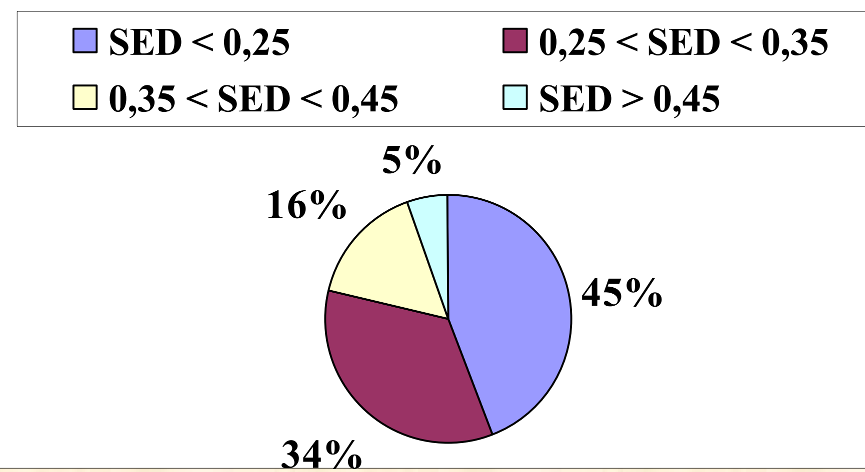
The 1-resolvable incomplete block model



The randomized complete block model with linear structure of spatial relationship added



The 1-resolvable incomplete block model with linear structure of spatial relationship added



Fractions of trials of different average values of SED (standard error of difference) depending on used mathematical model of observations