Návrhy experimentů / Experimental design (NMST436)

Z. Hlávka

Univerzita Karlova (v Praze)
Matematicko-fyzikální fakulta
Katedra pravděpodobnosti a matematické statistiky
www.karlin.mff.cuni.cz/~hlavka

Experimental design (NMST436)

- Repetition: linear model, ANOVA, multiple comparisons.
- Block designs, nested designs, repeated measurements.
- Factorial experiments, response surface methodology.
- Computer experiments.
- Regression experiments.

Basic references:

Milliken, Johnson (2009) *Analysis of Messy Data I: Designed Experiments*, 2nd ed, CRC Press.

Milliken, Johnson (1989) *Analysis of Messy Data II: Nonreplicated Experiments*, Chapman&Hall/CRC.

Likeš (1968) *Navrhování průmyslových experimentů*, SNTL.

Wu & Hamada (2011) *Experiments: planning, analysis, and optimization*, Wiley.

Pázman a kol. (1986) *Riešené situácie z navrhovania experimentov*, Alfa.

Further references:

Bailey (2008) *Design of Comparative Experiments*, Cambridge University Press.

Cochran, Cox (1957) *Experimental designs*, Wiley.

Scheffé (1959) *The analysis of variance*, Wiley.

Pázman (1986) *Foundations of optimum experimental design*, D. Reidel.

Myers, Montgomery (2002) *Response surface methodology*, Wiley.

Neter, Kutner, Nachtsheim, Wasserman (1996) *Applied linear statistical models*, Irwin.

Pinheiro, Bates (2000) *Mixed-effects models in S and S-PLUS*, Springer.

Fang, Wang (1993) *Number-theoretic methods in statistics*, CRC Press.

Further usable literature:

Yandell (1997) *Practical data analysis for designed experiments*, CRC Press.

Robinson (2000) *Practical strategies for experimenting*, Wiley.

Cox, Reid (2002) *The theory of the design of experiments*, CRC Press.

Christensen (2011) *Plane answers to complex questions: the theory of linear models*, Springer.

Week 1

Assumed prior knowledge:

- linear model,
- factors (contrasts),
- analysis of variance (ANOVA),
- ANOVA tables (Type I, II, III),
- random effects and mixed models.

Tukey, J. W. (1954). Unsolved Problems of Experimental Statistics. Journal of the American Statistical Association, 49(268), 706-731.

The four hypergeneral principles, which may seem harmless until we come to their consequences, run as follows:

(A) Different ends require different means and different logical structures.
(B) In each area, statistical method must and does evolve, mainly by adding *both* immediate ends *and* considerations.
(C) While techniques are important in experimental statistics, knowing when to use them and why to use them are more important.
(D) In the long run, it does not pay a statistician to fool either himself or his clients.

Experimental design

- Experimental design $\times$ data analysis.
- Bias $\times$ randomization?
- Statistician's participation on designing the experiment (statistician is often consulted only after the data set has been collected).
- Sample size calculations (in practice often *ex post*).
- Experimental design has to account for financial limitations, law requirements, and natural laws.

Experimental design needs a lot of practice (assuming that one is capable to think about the problems).

Aim of the lecture: basic experimental designs (including "standard terminology") and its analysis (linear regression, mixed models, GLMs).

The lecture

Prior knowledge: linear model with full rank, estimable parameter, analysis of variance model (interactions), ANOVA table (type I, II, III), contrasts, multiple comparisons (Bonferroni, Tukey).

Questions:

- Do we have to revisit the linear model and analysis of variance?
- Exercise in K4? Notebooks? Time of the lecture?

Example:

Wu & Hamada (2009), p. 57:

Laser-assisted manufacturing of a thermoplastic composite, the response is interply bond strength of the composite as measure by a short-beam-sher test (Mazumdar & Hoa, 1995).

Laser Power		
40 W	50 W	60 W
25.66	29.15	35.73
28.00	35.09	39.56
20.65	29.79	35.66

Two-sample t-test

The simplest and most effective approach is to use experimental designs allowing application of two-sample t-test.

Such designs are often using *randomization* (i.e., the random assignment of *treatment* (ošetření) to individuals).

The power of the two-sample t-test is a function of number of observations, the variance and the true difference of expected values. Using a realistic (expert) estimate of the variance, we can easily calculated approximate power of the test against "reasonable" alternatives.

Example: `power.t.test()`

Two-sample t-test vs. Wilcoxon (Mann-Whitney) test

"...in most practical cases the two-sample t-test is so robust that it can be recommended in nearly all applications." (Rasch et al, 1987)

References:

Rasch, Teuscher, Guiard, How robust are tests for two independent samples?, *Journal of Statistical Planning and Inference*, Volume 137, Issue 8, Pages 2706–2720.
`http://www.sciencedirect.com/science/article/pii/S0378375807000225`

Heeren T, D'Agostino R. Robustness of the two independent samples t-test when applied to ordinal scaled data. *Stat Med.* 1987 Jan–Feb;6(1):79–90.

Two-sample t-test (Welch approximation)

Without the assumption of homoskedasticity, the obvious test statistic is

$$T = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_y}}}.$$

Unfortunately, this test statistic does not have t-distribution. Welch (1938) used moment method in order to approximate the null distribution of $\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_y}$ by χ_f^2, where

$$f = \frac{(\sigma_X^2/n_X + \sigma_Y^2/n_Y)^2}{\sigma_X^4/(n_X^2(n_X-1)) + \sigma_Y^4/(n_Y^2(n_Y-1))}.$$

Therefore, t_f is usable as an approximation of the distribution of T.

Question: Sample size calculations for Welch test?

ANOVA: sample size calculations

See NKNW, Chapter 26:

- power for a chosen *noncentrality parameter*,
- power for a chosen *interesting difference* between factor levels,
- length of confidence intervals for *parameters of interest*,
- finding the best treatment.

ANOVA: linear combinations with unequal variance

It may happen that we reject the hypothesis of homoskedasticity. In that case, we can approximate the distribution of the estimator of the linear combination $\sum c_i \mu_i$.

Obviously

$$\sum c_i \hat{\mu}_i \sim N\left(\sum c_i \mu_i, \sum c_i^2 \sigma_i^2 / n_i\right)$$

and

$$\frac{\sum c_i \hat{\mu}_i - \sum c_i \mu_i}{\sqrt{\sum c_i^2 \sigma_i^2 / n_i}} \sim N(0,1).$$

For large values of n_i (at least for $c_i > 0$), we can use the asymptotic normality of the statistic:

$$Z = \frac{\sum c_i \hat{\mu}_i - \sum c_i \mu_i}{\sqrt{\sum c_i^2 \hat{\sigma}_i^2 / n_i}} \xrightarrow{\mathcal{D}} N(0,1).$$

ANOVA: Satterthwaite's approximation

If n_i's are too small, we can rewrite Z as a ratio of two independent random variables:

$$Z = \frac{\sum c_i \hat{\mu}_i - \sum c_i \mu_i}{\sqrt{\sum c_i^2 \hat{\sigma}_i^2 / n_i}} = \frac{\left(\sum c_i \hat{\mu}_i - \sum c_i \mu_i\right) / \sqrt{\sum c_i^2 \sigma_i^2 / n_i}}{\sqrt{\sum c_i^2 \hat{\sigma}_i^2 / n_i} / \sqrt{\sum c_i^2 \sigma_i^2 / n_i}}$$

The distribution of the numerator is $N(0,1)$ and the distribution of Z can be approximated by t_ν-distribution (since we can approximate $(\nu\times)$numerator by χ^2_ν-distribution—the only problem is the choice of the number of degrees of freedom ν).

Satterthwaite suggested $\nu = \frac{\left(\sum c_i^2 \sigma_i^2 / n_i\right)^2}{\sum\{c_i^4 \sigma_i^4 / n_i^2 (n_i - 1)\}}$, guaranteeing the equality of the first two moments. Plugging in $\hat{\sigma}^2$, we obtain Satterthwaite's $\hat{\nu}$ (and the distribution of Z is approximated by $t_{\hat{\nu}}$-distribution.)

Types of studies

Yandell (1997, p. 35-36) comments of the most common types of studies:

pure observational studies may lead to seriously biased results,

sample surveys monitor small samples of the population of interest,

designed experiments, where the scientist controls the study by either (i) randomizing subjects to study groups or by (ii) selecting subjects at random from the populations of interest (e.g., male vs. female).

Designed experiments

Designed experiments:

- randomize over extraneous factors while controlling assignment to groups,
- require careful attention to a *protocol* established before the experiment is run.

[Yandell (1997, p. 36)]

In the following, we describe basic designs that were developed mainly in agriculture and industry. Biostatistical experiments will be treated in NMST532: Design and Analysis of Medical Studies.

Publication bias and reproducibility

[Begley, Ioannidis (2015) Reproducibility in science: Improving the standard for basic and preclinical research, Circ Res. 116, 116–126] . . . inability to replicate the majority of findings presented in high-profile journals. The estimated for irreproduciblity based on these empirical observations range from 75% to 90%.

[The Economist (2013) Trouble at the lab]: . . . a lot of this priming research is poorly founded. Over the past few years various researchers have made systematic attempts to replicate some of the more widely cited priming experiments. Many of these repolications have failed.

Stages in a statistically designed experiment (Bailey, 2008, Section 1.1)

1. Consultation.
2. Statistical design.
3. Data collection.
4. Data scrutiny.
5. Analysis.
6. Interpretation.

The ideal and the reality (Bailey, 2008, Section 1.2)

1. Purpose of the experiment.
2. Replication.
3. Local control.
4. Constraints.
5. Choice.

Defining terms (Bailey, 2008, Section 1.4)

An experimental unit is the smallest unit to which a treatment can be applied.

A treatment is the entire description of what can be applied to an experimental unit.

An observational unit is the smallest unit on which a response will be measured.

Treatment structure means meaningful ways of dividing up the set of treatments.

Plot structure means meaningful ways of dividing up the set of plots, ignoring the treatments.

The design is the allocation of treatments to plots.

The plan or layout is the design translated into actual plots (some randomization is usually involved).

Week 2

Topic:

- Randomized block design / znáhodněné bloky.
- Balanced incomplete block design / vyvážené neúplné bloky.
- Latin and Graeco-Latin squares / latinské a řecko-latinské čtverce.

Randomized blocks

In order to investigate the effect of treatment (factor A) on the dependent variable, one usually applies the analysis of variance (ANOVA) model.

Often, the sample is not homogeneous but it can be split into b homogeneous groups (in agriculture: field, in industry: meltage/tavba, in education: class).

Randomized blocks (znáhodněné bloky) assign treatments levels (i.e., factor A levels) within each block randomly.

Complete randomized blocks (úplné znáhodněné bloky) assign all treatment levels within each block.

Example: Likeš (1968), example 4.1.1 (in Czech).

Byl proveden experiment, jehož účelem bylo vyšetřit účinek původní velikosti částic a pěchovacího tlaku na pevnost v tahu slinutého železného prášku. Slinování bylo provedeno při standardní teplotě ve vodíkové atmosféře na vzorcích materiálu šesti velikostí částic, tlak byl uvažován na dvou úrovních. Dvanáct kombinací úrovní si můžeme představit jako 12 úrovní faktoru A.

Pro experiment bylo použito tří podobných pecí; v každé peci bylo zkoumáno všech 12 úrovní faktoru A.

Example: Likeš (1968), example 4.1.1 (in English).

The aim of the experiment was to investigate the effect of the original particle size and ramming pressure on the tensile strength of iron power alloy. Alloying was performed at standard temperature in hydrogen atmosphere on experimental units with six particle sizes and two pressure levels. The resulting twelve level combinations were treated as 12 factor A levels.

The experiment was performed in three similar furnaces and all twelve treatment levels were investigated in each furnace.

Tab. 13. Hodnoty y_{ij}

Pec	Tlak	Velikost částic					
		V_1	V_2	V_3	V_4	V_5	V_6
1	T_1	11,3	12,2	12,9	12,1	16,9	14,3
	T_2	21,1	21,1	21,7	24,4	23,6	23,5
2	T_1	11,9	10,4	12,4	13,9	14,9	15,0
	T_2	21,3	21,4	22,0	24,1	25,5	22,1
3	T_1	10,0	9,9	11,3	13,3	12,4	13,8
	T_2	18,8	19,5	21,6	23,8	23,3	20,5

We will analyze the data set during exercises, the main aim is to investigate the treatment effect (the influence of factor A on tensile strength).

Questions:

- Should be treat factor B (block) as fixed or random effect?
- Can we split the factor A into some "subfactors"?
- Which treatments are significantly different?

Treatment structure × design structure

See Milliken & Johnson 2009, p. 77

The experiment design can be split into two parts (the treatment structure and the design structure) and it is useful to understand its distinction.

Treatment structure (struktura ošetření) consists of treatments, factors, or groups that the investigator wants to study or compare.

The treatment structure is determined by the factors of interest: these are typically various treatments (often including also control or standard treatment).

Treatment structure × design structure

Design structure (struktura návrhu, struktura rušivých faktorů) is determined by the structure of the experimental units into homogeneous groups or blocks.

The design structure contains factors determining (describing) the homogeneous groups of experimental units.

In the most simple setup, all units are homogeneous and, using only one block, one can randomly assign treatment levels to all experimental units. The resulting data set can by analyzed by using simple one-way analysis of variance.

Often, the size of block is smaller than number of treatments and complete randomized blocks are not applicable.

Example: It may happen that a furnace (in the previous example) can contain at most 6 samples. In such case, we cannot investigate effects of all twelve treatents within each block.

Tables of *balanced incomplete randomized blocks* can be found in relevant literature.

The description of the experimental design:

- t number of factor A levels,
- b number of blocks,
- k number of (homogeneous) units within block,
- r number of blocks containing each level of factor A,
- λ number of occurences of each pair (of factor A levels).

V tabulce 15 je uveden příklad [11] vyváženého neúplného bloku pro případ $t = 7$, $k = 3$, $r = 3$, $b = 7$, $\lambda = 1$.

Tab. 15. Příklad vyváženého neúplného bloku

Číslo bloku	Úrovně faktoru A
1	A_1, A_2, A_4
2	A_2, A_3, A_5
3	A_3, A_4, A_6
4	A_4, A_5, A_7
5	A_1, A_5, A_6
6	A_2, A_6, A_7
7	A_1, A_3, A_7

Unbalanced two-way ANOVA (model analýzy rozptylu, nevyvážené dvojné třídění) with 0 or 1 observations in each subclass.

Example: Likeš (1968), example 4.2.1 (in Czech).

U čtyř pneumatik byla zkoumána trvanlivost. Každá pneumatika byla rozdělena na tři části. Každá část byla vyrobena různým způsobem (faktor A). Byly uvažovány čtyři úrovně.

Pneumatiky lze považovat za bloky, každý blok obsahuje $k = 3$ jednotky.

Přiřazení úrovní faktorů A jednotlivým blokům (pneumatikám) lze snadno vyčíst z tabulky na následující straně; zřejmě se jedná o vyvážené neúplné bloky.

Example: Likeš (1968), example 4.2.1.

Durability of four tyres was investigated. Each tyre was split into three parts that were produced by a different technology (factor A with four levels).

The blocks are defined by tyres, i.e., each block contains $k = 3$ experimental units.

The factor A levels assigment within each block (tyre) is given from the table on the next page. By randomized the (order of) treatments within each block, we obtain balanced incomplete randomized blocks.

Tab. 17. Hodnoty relativní trvanlivosti

	1	2	3	4	$Y_{i.}$
A_1	238	196	254	—	688
A_2	238	213	—	312	763
A_3	279	—	334	421	1 034
A_4	—	308	367	412	1 087
$Y_{.j}$	755	717	955	1 145	3 572

Latin squares

Blocks allow to eliminate the effect of a single nuisance (rušivý) factor. *Latin squares* can be applied in order to account for two (or more) nuisance factors.

Latin squares are applicable, if the nuisance factors have the same number of factor levels as factor A.

Example: Four levels of factoru A, two nuisance factors with four levels, 16 observations.

	1	2	3	4
1	1	2	3	4
2	2	1	4	3
3	3	4	1	2
4	4	3	2	1

Example: Likeš (1968), example 4.3.1 (in Czech).

Byl proveden experiment, jehož účelem bylo vyšetřit, jak se projevuje variabilita materiálu na pevnost v tahu žíhaných měděných trubek. Variabilita materiálu byla vyšetřována tak, že bylo vybráno z osmi různých dnů v období třítýdenní výrobní periody vždy osm trubek. Těchto osm dní charakterizovalo změnu materiálu v tomto období. Skupiny trubek z jednotlivých dnů představují faktor A na osmi úrovních. Trubky byly žíhány v peci, přičemž byly upevněny do přípravku, který měl otvory v osmi řádcích a osmi sloupcích. Aby byl eliminován vliv horizontální a vertikální polohy trubky v žíhací peci, byl experiment uspořádán do latinského čtverce o straně $t = 8$.

Example: Likeš (1968), example 4.3.1.

The aim of the experiment was to investigate the effect of material variability on tensile strength of annealed copper tubes. The material variability was investigated by choosing eight tubes from each of eight days chosen within a given three-weeks production period. These eight days represented the change of the material during the production period. Hence, tubes from different days represent factor A with eight levels. During annealing, tubes were fastened in a template with holes in eight rows and eight columns. The experiment was arranged as latin squares with size $t = 8$ in order to eliminate the effect of the horizontal an vertial position of a tube in the annealing furnace.

Previous experience with the annealing furnace suggests that the temperature varies either vertically or horizontally and that there is not any interaction between rows and columns. In addition, there is not any reason to assume that there is an interaction of fator A (day) and the position of tubes in the furnace. This experiment was performed for varying temperature. For each temperature, the experiment was arranged as a Latin square. Results for the temperature 300 °C are given in the following table.

Tab. 22. Hodnoty pevnosti v tahu žíhaných trubek

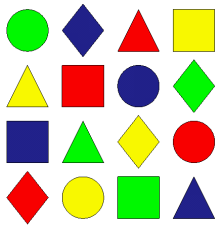
Řádky	Sloupce 1	2	3	4	5	6	7	8	Součty $Y_{i.}$
1	$A_4 = 25{,}7$	$A_8 = 26{,}2$	$A_3 = 27{,}0$	$A_2 = 27{,}0$	$A_5 = 24{,}4$	$A_1 = 28{,}2$	$A_7 = 24{,}3$	$A_6 = 24{,}5$	207,3
2	$A_6 = 24{,}6$	$A_5 = 25{,}4$	$A_7 = 24{,}5$	$A_1 = 29{,}4$	$A_8 = 27{,}3$	$A_2 = 27{,}6$	$A_3 = 29{,}3$	$A_4 = 26{,}5$	214,6
3	$A_2 = 26{,}5$	$A_3 = 26{,}0$	$A_8 = 29{,}8$	$A_4 = 25{,}7$	$A_7 = 24{,}5$	$A_6 = 27{,}6$	$A_5 = 28{,}5$	$A_1 = 28{,}4$	217,0
4	$A_1 = 27{,}4$	$A_7 = 24{,}6$	$A_5 = 25{,}3$	$A_6 = 24{,}8$	$A_3 = 27{,}3$	$A_4 = 27{,}6$	$A_8 = 28{,}1$	$A_2 = 25{,}6$	210,7
5	$A_3 = 27{,}0$	$A_2 = 26{,}4$	$A_4 = 26{,}0$	$A_8 = 29{,}8$	$A_1 = 31{,}5$	$A_5 = 28{,}5$	$A_6 = 24{,}6$	$A_7 = 24{,}3$	218,1
6	$A_5 = 25{,}6$	$A_6 = 24{,}8$	$A_1 = 26{,}2$	$A_7 = 24{,}6$	$A_4 = 26{,}5$	$A_3 = 27{,}1$	$A_2 = 27{,}0$	$A_8 = 30{,}4$	212,2
7	$A_7 = 24{,}5$	$A_1 = 26{,}2$	$A_6 = 24{,}6$	$A_5 = 25{,}6$	$A_2 = 27{,}3$	$A_8 = 30{,}1$	$A_4 = 26{,}5$	$A_3 = 28{,}4$	213,2
8	$A_8 = 28{,}8$	$A_4 = 27{,}0$	$A_2 = 27{,}0$	$A_3 = 29{,}8$	$A_6 = 26{,}0$	$A_7 = 24{,}3$	$A_1 = 27{,}0$	$A_5 = 28{,}5$	218,4
Součty $Y_{.j}$	210,1	206,6	210,4	216,7	214,8	221,0	215,3	216,6	1 711,5

Is it possible to test also the assumptions of the model?

Graeco-Latin squares

Graeco-Latin square is an overlay of two latin squares (it is known that such squares exist for $t \geq 3$ with the exception of $t = 6$).

Aα	Bγ	Cδ	Dβ
Bβ	Aδ	Dγ	Cα
Cγ	Dα	Aβ	Bδ
Dδ	Cβ	Bα	Aγ



Example: Tensile strength of annealed copper tubes: is it possible to consider also the production order of tubes within each day?

Hyper-Graeco-Latin squares

It is possible to design also an experiment with four nuisance (block) factors.

The true factor levels of all factor should be randomized (i.e., the true factor levels are randomly assigned to the number given in the following table).

	1	2	3	4	5
1	A11	B22	C33	D44	E55
2	D23	E34	A45	B51	C12
3	B35	C41	D52	E31	A24
4	E42	A53	B14	C25	D31
5	C54	D15	E21	A32	B43

Summary: design structure

Design structure (nuisance factors):

- Completely randomized design (assuming that all experimental units are homogeneous).
- Complete randomized blocks.
- Incomplete randomized blocks.
- Latin and Graeco-Latin squares.
- Various modifications (consequence of "collision with reality").

Summary: treatment structure

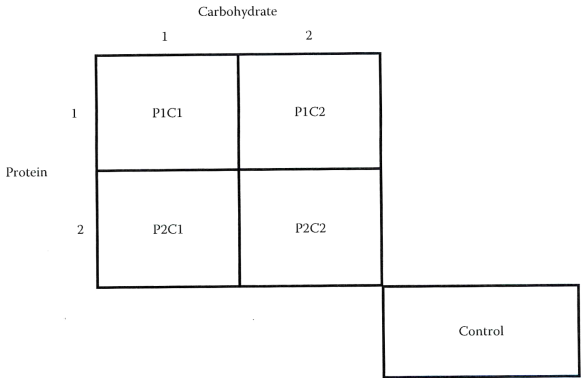
Treatment structure (investigated factors):

- Single treatment (one-way ANOVA, linear model with one factor variable).
- All combinations of two treatments (two-way ANOVA).
- All combinations of many treatments (factorial experiment).
- Selected combinations of many treatments (fractional factorial experiment / zkrácený faktoriální experiment).
- Optimal regression design (typically for numerical explanatory variables).
- Factorial experiment with controls.

Practical advice: **don't panic!**

Modifications

In practice, modifications of these "standard" designs are common. For example, MJ1 consider block design for "diet effects" including additionally a control group.



Factorial experiment 2^2 with controls

[MJ1, Section 4.3.2] say that "the diet treatment structure is a two-way factorial arrangement with a control that, when crossed with sex of person, generates a three-way treatment structure with two controls... The design structure is completely randomized where each treatment combination is to be assigned to two persons."

Analysis of Variance Table for a Treatment Structure Consisting of Three-Way Factorial Arrangement Combined with Two Controls in a Completely Randomized Design Structure

Source of Variation	*df*
Sex	1
Diet	4
Control vs 2^2	1
Protein	1
Carbohydrate	1
Protein × carbohydrate	1
Sex × Diet	4
Sex × control vs 2^2	1
Sex × protein	1
Sex × carbohydrate	1
Sex × protein × carbohydrate	1
Error	10

Week 3

Topic:

- Analysis of variance (ANOVA):
  - one-way ANOVA / jednoduché třídění,
  - two-way ANOVA / dvojné třídění,
  - fixed, random, and mixed effects / pevné, náhodné a smíšené efekty.
- Multilevel designs / návrhy experimentů s více úrovněmi.
  - cupcakes (see Milliken & Johnson, section 5.1),
  - meat (see Milliken & Johnson, example 5.4),
  - cheese (see Milliken & Johnson, example 5.5).

One-way ANOVA with fixed effects

ANOVA table:

Source of variation	Sum of squares	df	Mean square	EMS
Pramen měnlivosti	Součet čtverců	Stupně volnosti	Průměrný čtverec	EMS
A	$S_A = \sum_{i=1}^{I} r(y_{i.} - y_{..})^2$	$I-1$	$\frac{S_A}{I-1}$	$\sigma^2 + \frac{r}{I-1}\sum \alpha_i^2$
Residual	$S_e = \sum\sum(y_{iv} - y_{i.})^2$	$I(r-1)$	$s^2 = \frac{S_e}{I(r-1)}$	σ^2
Total	$S = \sum\sum(y_{iv} - y_{..})^2$	$Ir-1$		

Derivation of null distribution

Theorem:[Nelder, 1965] If $C_1, \ldots, C_k$ are $n \times n$ symmetric matrices with ranks $r_1, \ldots, r_k$ and $\sum_{i=1}^{k} C_i = \mathcal{I}$, then any one of the four following conditions implies the other three:

- the C_i are idempotent;
- $C_i C_j = 0$ for $i \neq j$;
- $\sum_{i=1}^{k} r_i = n$;
- there exists an orthogonal matrix H such that the non-zero parts of $H^\top C_i H$ are disjunct unit matrices.

Theorem:[Anděl, 1985, Věta 17] Assume that $X \sim N_n(\mu, V)$ and $A \geq 0$ is symmetric. If AV is nonzero and idempotent, then $X^\top AX \sim \chi^2_r$, where $r = \mathrm{tr}(AV)$.

Example: Assume that $Y \sim\sim N_n(\mu, \sigma^2\mathcal{I})$. The vector of means is $\overline{Y} = MY$, where $M = \mathbf{1}_n\mathbf{1}_n^\top/n$ and centered observations are $Y^c = Y - \overline{Y} = (\mathcal{I} - M)Y = HY$. Obviously, matrices H and M are idempotent, $\mathcal{I} = H + M$, and it follows that the asymptotic distribution of

$$(n-1)\frac{Y^\top MY}{Y^\top HY}$$

is $F_{1,n-1}$ if $\mu = 0$.

Example: Similarly, we can derive the null distribution in one-way ANOVA. Writing the vector of 'blockwise means' as $\overline{Y}_A = M^*Y$, we have that $\mathcal{I} = M + (M^* - M) + (\mathcal{I} - M^*)$, where matrices M, $(M^* - M)$, and $(\mathcal{I} - M^*)$ are obviously idempotent. It is easy to see that, under H_0,

$$\frac{\frac{Y^\top(M^*-M)Y}{I-1}}{\frac{Y^\top(\mathcal{I}-M^*)Y}{n-I}} \sim F_{I-1,n-I}.$$

One-way ANOVA with random effects

ANOVA table:

Source of variation	Sum of squares	df	Mean square	EMS
A	$S_A = \sum_{i=1}^{I} r(y_{i.} - y_{..})^2$	$I-1$	$\frac{S_A}{I-1}$	$\sigma_e^2 + r\sigma_a^2$
Residual	$S_e = \sum\sum(y_{iv} - y_{i.})^2$	$I(r-1)$	$s^2 = \frac{S_e}{I(r-1)}$	σ_e^2
Total	$S = \sum\sum(y_{iv} - y_{..})^2$	$Ir-1$		

Two-way ANOVA with fixed effects (and interaction)

Shortened ANOVA table:

Source of variation	Mean Square	EMS
A	$\frac{Jr\sum_{i=1}^{I}(y_{i..}-y_{...})^2}{(I-1)}$	$\sigma^2 + \frac{Jr}{I-1}\sum\alpha_i^2$
B	$\frac{Ir\sum_{j=1}^{J}(y_{.j.}-y_{...})^2}{(J-1)}$	$\sigma^2 + \frac{Ir}{J-1}\sum\beta_j^2$
AB	$\frac{r\sum\sum(y_{ij.}-y_{i..}-y_{.j.}+y_{...})^2}{(I-1)(J-1)}$	$\sigma^2 + \frac{r}{(I-1)(J-1)}\sum\sum(\alpha\beta)_{ij}^2$
Residual	$\frac{\sum\sum\sum(y_{ijv}-y_{ij.})^2}{IJ(r-1)}$	σ^2
Total	$\frac{\sum\sum\sum(y_{ijv}-y_{..})^2}{IJr-1}$	

Two-way ANOVA with random effects (and interaction)

Shortened ANOVA table:

Source of variation	Mean Square	EMS
A	$\frac{Jr\sum_{i=1}^{I}(y_{i..}-y_{...})^2}{(I-1)}$	$\sigma_e^2 + r\sigma_{ab}^2 + Jr\sigma_a^2$
B	$\frac{Ir\sum_{j=1}^{J}(y_{.j.}-y_{...})^2}{(J-1)}$	$\sigma_e^2 + r\sigma_{ab}^2 + Ir\sigma_b^2$
AB	$\frac{r\sum\sum(y_{ij.}-y_{i..}-y_{.j.}+y_{...})^2}{(I-1)(J-1)}$	$\sigma_e^2 + r\sigma_{ab}^2$
Residual	$\frac{\sum\sum\sum(y_{ijv}-y_{ij.})^2}{IJ(r-1)}$	σ_e^2
Total	$\frac{\sum\sum\sum(y_{ijv}-y_{..})^2}{IJr-1}$	

Two-way ANOVA with mixed effects (and interaction)

Shortened ANOVA table:

Source of variation	Mean Square	EMS
A	$\frac{Jr\sum_{i=1}^{I}(y_{i..}-y_{...})^2}{(I-1)}$	$\sigma_e^2 + \frac{Jr}{I-1}\sum\alpha_i^2 + r\sigma_{ab}^2$
B	$\frac{Ir\sum_{j=1}^{J}(y_{.j.}-y_{...})^2}{(J-1)}$	$\sigma_e^2 + Ir\sigma_b^2$
AB	$\frac{r\sum\sum(y_{ij.}-y_{i..}-y_{.j.}+y_{...})^2}{(I-1)(J-1)}$	$\sigma_e^2 + r\sigma_{ab}^2$
Residual	$\frac{\sum\sum\sum(y_{ijv}-y_{ij.})^2}{IJ(r-1)}$	σ_e^2
Total	$\frac{\sum\sum\sum(y_{ijv}-y_{..})^2}{IJr-1}$	

NNKW (page. 981): The derivations are tedious, but simple rules have been developed for finding the expected mean squares.

Milliken & Johnson (2009) Analysis of Messy Data, Volume 1, Designed Experiments, CRC Press, str. 101:

Consulting statisticians do not always get the chance to design the experiments for which they must help construct appropriate analyses. Instead, the statistician must first identify the type of designed experiment the researcher has employed. The first and most important step in the identification process is to determine if more than one size of experimental unit has been used, and if so, to identify each size of experimental unit.

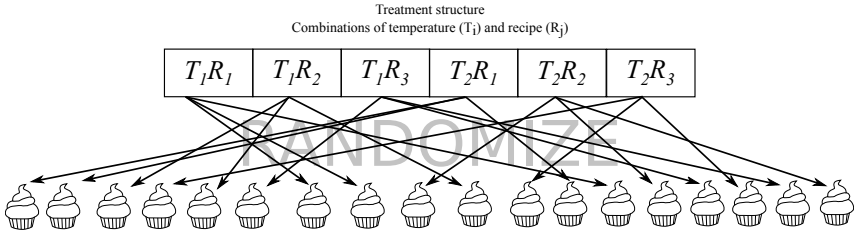
As will become evident in this section, each size of experimental unit will have an associated design structure and treatment structure. After the different sizes of the experimental units have been identified, the model for carrying out an appropriate analysis can be constructed by combining the models used to describe the design structure and treatment structure corresponding to each size of experimental unit.

Cupcakes

Milliken & Johnson (2009) describe several experiments comparing the influence of three recipes (factor R) and two temperatures (factor T) on cupcakes (end-point is not specified).

These designs are substantially different even though we are always going to bake 18 cupcakes.

The most simple experimental design is to assume that we can bake all 18 cupcakes 18 times under identical (homogeneous) conditions.



Analysis: two-way ANOVA with interactions (treatment design), 18 measurements, 5 parameters.

Crucial assumption is the homogeneity of "all baking conditions".

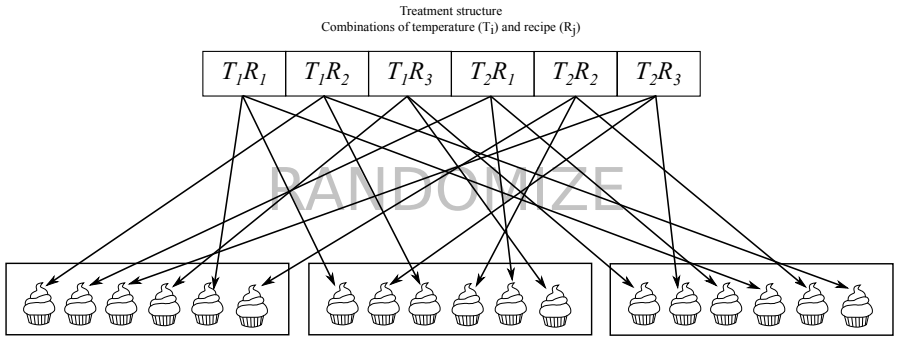
Structure of ANOVA table:

	df	**EMS**
Temperature	1	$\sigma_e^2 + \Phi^2(\tau)$
Recipe	2	$\sigma_e^2 + \Phi^2(\beta)$
Temperature $\times$ Recipe	2	$\sigma_e^2 + \Phi^2(\tau\beta)$
Error	12	σ_e^2

In practice, baking experiments encounter "real-life difficulties".

Mixing dough and baking can be time consuming and it is possible to bake only six homogeneous cupcakes in a day (and cupcakes from different days can be heterogeneous). In such situation, we can use complete randomized blocks (with days defining blocks).

The treatment structure remains the same but we have one nuisance factor (day) leading to three blocks with six measurements.



Analysis: three-way ANOVA (including day as a blocking factor).

We have to bake six homogeneous cupcakes each day.

Structure of ANOVA table:

	df	**EMS**
Day	2	$\sigma_e^2 + 6\sigma_d^2$
Temperature	1	$\sigma_e^2 + \Phi^2(\tau)$
Recipe	2	$\sigma_e^2 + \Phi^2(\beta)$
Temperature $\times$ Recipe	2	$\sigma_e^2 + \Phi^2(\tau\beta)$
Error	10	σ_e^2

Note: complete randomized blocks.

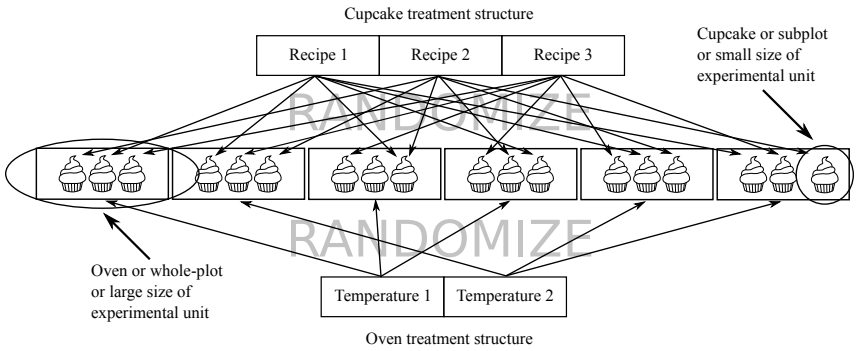
Split-plot design/ rozdělené dílce

Assume that three cupcakes can be baked in one oven at the same time. Then, it would be inefficient to bake only one cupcake in each batch.

It is important to realize that, within one batch, it is not a problem to compare three recipes but we can not compare different temperatures.

Factor R levels (recipes) can be assigned randomly to three cupcakes in given oven but these three cupcakes must be baked at the same temperature.

This experiment contains two treatment levels: recipes are assigned to "cupcakes within oven" while temperatures are assigned to ovens (containing three cupcakes).



Cupcake point of view: six homogeneous blocks (three times with lower and three times with higher temperature), treatment effect = "recipe".

Oven point of view: six measurements, one-way anova (three measurements with lower and three measurements with higher temperature).

The model has altogether nine parameters: temperature (1 par), oven (4 pars), recipe (2 pars), recipe and temperature interaction (2 pars).

Note 1: factor "oven" has six levels but we need only four parameters because of linear dependencies with the temperature and the intercept.

Note 2: factor "day" is not present because we can bake six times in a day (anyway, the effect of day would be confounded with "oven").

Structure of ANOVA table:

	df	**EMS**
Temperature	1	$\sigma_e^2 + 3\sigma_t^2 + \Phi^2(\tau)$
Error (oven)	4	$\sigma_e^2 + 3\sigma_t^2$
Recipe	2	$\sigma_e^2 + \Phi^2(\beta)$
Temperature $\times$ Recipe	2	$\sigma_e^2 + \Phi^2(\tau\beta)$
Error	8	σ_e^2

MJ: "split-plot design with a completely randomized whole-plot design structure"

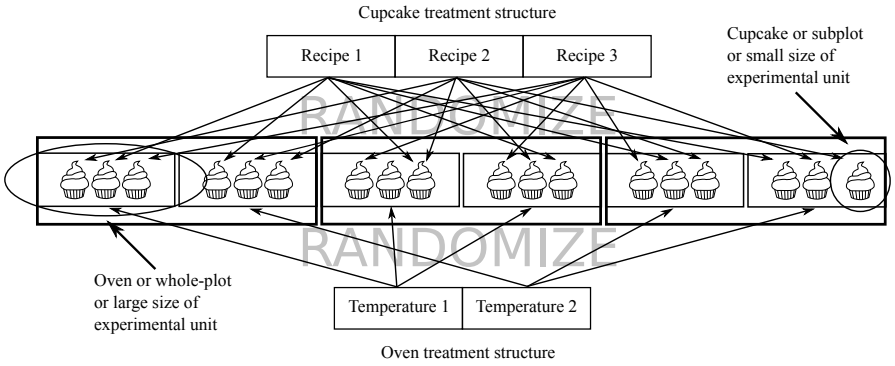
Note: whole-plot = oven, subplot = cupcake.

Assuming that we can bake only two batches per day, the design can be easily extended by including the factor day (naturally confounded with oven).

From the point of view of a cupcake, the model does not change.

From the oven point of view, we have to consider blocks defined by days. This leads to complete randomized blocks, where two temperatures are randomized within each day.

Altogether, we are working with three levels: day, oven (within day) and cupcake (within oven). We investigate two treatments, where recipes are applied on "cupcakes within oven" and temperatures on "ovens within day".



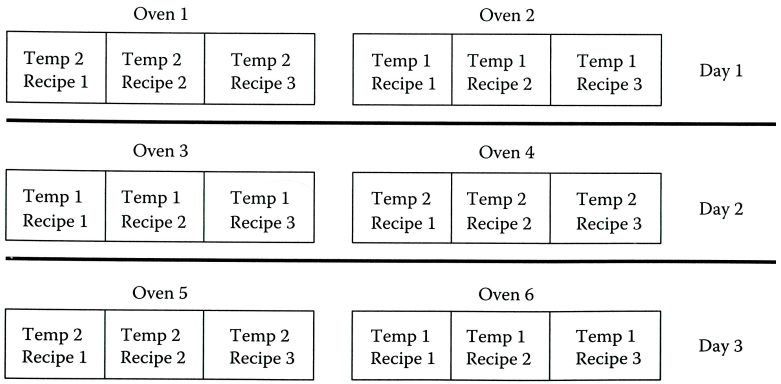
The analysis basically does not change. Adding the factor day leads to a different parameterization but the number of parameters is the same.

Structure of ANOVA table:

	df	EMS
Day	2	$\sigma_e^2 + 3\sigma_t^2 + 6\sigma_d^2$
Temperature	1	$\sigma_e^2 + 3\sigma_t^2 + \Phi^2(\tau)$
Error (oven)	2	$\sigma_e^2 + 3\sigma_t^2$
Recipe	2	$\sigma_e^2 + \Phi^2(\beta)$
Temperature $\times$ Recipe	2	$\sigma_e^2 + \Phi^2(\tau\beta)$
Error	8	σ_e^2

MJ: "split-plot design with a randomized complete block whole-plot design structure"

Split-plot design can be interpreted also as an incomplete randomized blocks design: the nuisance factor is "oven within day" and we investigate 6 levels of a single treatment by using blocks of size 3.



Strip-plot design (provázané?? dílce)

strip:

- pruh, proužek, pás(ek) látky ap.
- bulvár, třída s restauracemi a obchody
- kreslený seriál v časopise ap.
- seřvat, sprdnout, sjet koho
- svléknout se, (vy)svléct se
- obnažit, odkrýt, svléknout co
- rozebrat, rozmontovat co stroj ap.
- zbavit koho čeho
- odejmout, odebrat komu co

Source: slovnik.seznam.cz

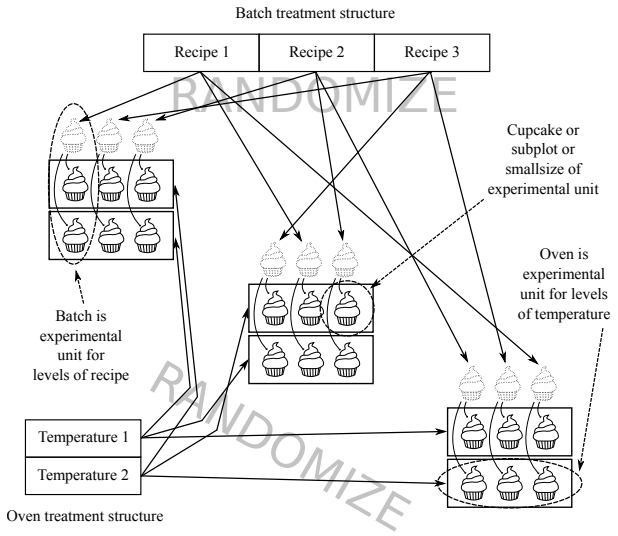
přepásané, propletené, propojené, svázané, spletené?

Strip-plot design

When we want to bake two cupcakes using the same recipe within a single day, we can reorganize the experiment so that these two cupcakes are prepared from a single batch of dough.

This simplifies the realization of the experiment (and reduces its price) but we have to consider another factor variable (factor "batch of dough").

We proceed as follows: measurement planned for each day are arranged in rectangles: values of the first and second factor (temperature of the oven and recipe used to prepare the dough batch) are assigned randomly to rows and columns, respectively.



Strip-plot design model

Analysis: mixed model with random effects of oven and batch nested within the random effect of day (the rectangle).

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + d_k + t_{ik} + b_{jk} + e_{ijk}$$

d_k efect of k-th day ("rectangle"),
t_{ik} effect of oven within day,
b_{jk} effect of batch within day.

Structure of ANOVA table (strip-plot design):

	df	**EMS**
Day	2	$\sigma_e^2 + 3\sigma_t^2 + 3\sigma_b^2 + 6\sigma_d^2$
Temperature	1	$\sigma_e^2 + 3\sigma_t^2 + \Phi^2(\tau)$
Error (oven)	2	$\sigma_e^2 + 3\sigma_t^2$
Recipe	2	$\sigma_e^2 + 2\sigma_b^2 + \Phi^2(\beta)$
Error (batch)	4	$\sigma_e^2 + 2\sigma_b^2$
Temperature $\times$ Recipe	2	$\sigma_e^2 + \Phi^2(\tau\beta)$
Error	4	σ_e^2

Meat (MJ: example 5.4, page 125, in Czech)

Klient si přeje vyhodnotit vliv faktorů:

- teplota (T: 3 úrovně),
- obal (O: 2 úrovně),
- typ světla (S: 4 úrovně),
- intenzita světla (I: 4 úrovně)

na barvu masa, které je 7 dní uložené v chladničce.

K dispozici máme šest chladniček. Každá chladnička je rozdělená na 16 přihrádek, které tvoří mřížku 4×4. Intenzita světla je určena vzdáleností od boku ledničky a všechny přihrádky v jednom sloupci tedy mají stejnou intenzitu světla. Těmto čtyřem přihrádkám (se stejnou intenzitou světla) náhodně přiřadíme různé typy osvětlení. Do každé přihrádky pak uložíme dva steaky (ve dvou různých obalech).

Meat (MJ: example 5.4, page 125)

A client wants to investigate the influence of:

- temperature (T: 3 levels),
- packaging (P: 2 levels),
- light type (L: 4 levels),
- Light intensity (I: 4 levels)

on the color of meat stored in refrigerator for 7 days.

We can use six fridges. Each fridge is split into 16 compartments organized in a 4×4 layout. The light intensity is determined by the distance from the side of the fridge. Hence, all compartments in a single column have the same light intensity and the four different types of lighting are randomly assigned to these four compartments. Finally, two steaks are stored in each compartment (in two different packagings).

Cheese (MJ: example 5.5, page 131, in Czech)

Mlékárna porovnává vlastnosti sýra vyrobeného a skladovaného za různých podmínek. Zkoumané faktory jsou: obsah tuku (2 úrovně), typ sýra (3 úrovně), teplota při uskladnění (2 úrovně), vlhkost vzduchu při uskladnění (2 úrovně). Experiment se skládá ze dvou kroků: výroba (6 kombinací) a uskladnění (4 možnosti).

Při výrobě se vyrobí dávka každého ze šesti typů sýra. Každá dávka je pak rozdělena na čtyři části, které jsou skladovány po dobu čtyř týdnů za různých podmínek (mlékárna má čtyři komory, kde může nastavit teplotu a vlhkost vzduchu). Mlékárna si přeje mít čtyři měření pro každou kombinaci faktorů a celý experiment je tedy proveden čtyřikrát (to trvá dohromady čtyři měsíce). Každý měsíc tedy získáme celkem 24 měření odpovídajících čtyřem různým ošetřením (na úrovni "komory") a šesti typům (dávkám) sýra.

Otázky: O jaký typ návrhu experimentu se jedná? Jak přesně má mlékárna postupovat při randomizaci? Jaký model použijeme k vyhodnocení experimentu?

Cheese (MJ: example 5.5, page 131)

A dairy company compares cheese produced and stored in different conditions.

The factors of interest are:

- fat content (2 levels),
- type of cheese (3 levels),
- storage temperature (2 levels),
- storage humidity (2 levels).

The experiment consists of two steps: production (6 combinations of factor levels) and storage (4 factor levels combinations).

The dairy company produces one batch of each type of cheese. Each batch is split into four parts stored for four weeks in different conditions (the dairy company has four chambers with climate control).

The dairy company wants to have four measurements for each combination of factors and, therefore, this experiment is repeated four times (this takes altogether four months).

Each month, we obtain 24 measurements corresponding to four treatment levels (in different chambers) and six types (batches) of cheese.

Question: What is the type of the proposed experiment design? How do we randomize the experiment? Which model can be used to evaluate the experiment?

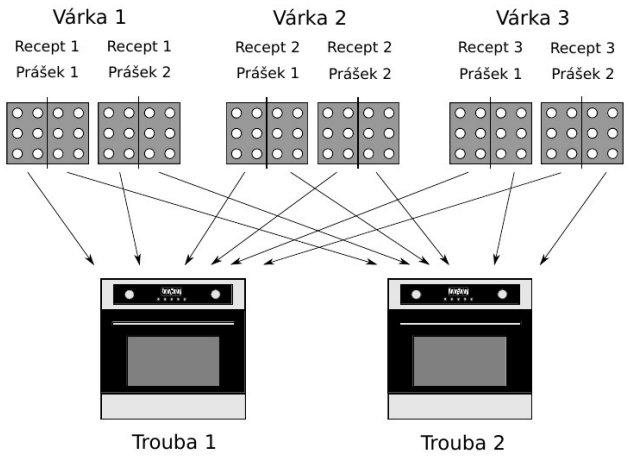
Baking powder (prdopeč)

In 2016, four students (TM, TCh, KJ, MH) have investigated the effect of baking powder on cupcakes. The experiment was designed using two ovens (located in student dormitory in Troja) with baking trays allowing to bake twelve cupcakes at the same time.

The main aim of the experiment was to investigate the amount of waste depending on two types of baking powder (with and without phosphates) and three recipes. Second aim was to find out which baking powder results in a better tasting cupcakes.

See
`www.karlin.mff.cuni.cz/~hlavka/vyuka/planex/folie/prasek.pdf`
for detailed information.

Baking powder experiment



Nesting and crossing (for blocking/nuisance factors)

Nelder (1965) notes that:

'All simple block structures can be built up using two basic operations, *nesting* (denoted by $\rightarrow$) and *crossing* (denoted by $\times$). The two simplest structures are written as $B_1 \rightarrow B_2$ and $B_1 \times B_2$. $B_1 \rightarrow B_2$ is the structure of the randomized block design...

A *simple block structure* may then be defined to be any formula involving $\rightarrow$ and $\times$ signs with suitable brackets to indicate the order of combination and with the n_i, one for each category, giving the number of units involved in the category'.

Simple block structures with 3 or 4 categories

Some design structures (for blocking factors) are given in Nelder (1965), page 150:

$(B_1 \to B_2) \to B_3 = B_1 \to (B_2 \to B_3)$ replicated split-plot design
$B_1 \to (B_2 \times B_3)$ mutliple criss-cross (Latin squares)
$(B_1 \times B_2) \to B_3$ criss-cross with split-plots
$(B_1 \to B_2) \times B_3$ criss-cross with split rows
$B_1 \times (B_2 \to B_3)$ criss-cross with split columns
$(B_1 \to B_2) \times (B_3 \to B_4)$ criss-cross with split rows and columns
$B_1 \to \{B_2 \times (B_3 \to B_4)\}$ repeated criss-cross with split columns
$B_1 \to (B_2 \times B_3) \to B_4$ repeated criss-cross with split plots
$(B_1 \times B_2) \to (B_3 \times B_4)$ criss-cross with split criss-crosses

Week 4

Topic:

- nested & hierarchical designs / hierarchicky uspořádané experimenty,
- repeated measures / opakovaná pozorování,
- cross-over design,
- analysis of repeated measurements in R.

Nested design

with random effects:

$$y_{ijr} = \mu + a_i + b_{j(i)} + e_{ijr}$$

with fixed effects:

$$y_{ijr} = \mu + \alpha_i + \beta_{j(i)} + e_{ijr}$$

ANOVA table for nested design

	Sum of squares	df	Mean square	EMS fixed effects	EMS random effects
A	$Jr\sum(y_{i..}-y_{...})^2$	$I-1$	$\frac{S_A}{I-1}$	$\sigma^2+\frac{Jr}{I-1}\sum\alpha_i^2$	$\sigma_e^2+r\sigma_b^2+Jr\sigma_a^2$
B(A)	$r\sum\sum(y_{ij.}-y_{i..})^2$	$I(J-1)$	$\frac{S_{B(A)}}{I(J-1)}$	$\sigma^2+\frac{r}{I(J-1)}\sum\sum\beta_{j(i)}^2$	$\sigma_e^2+r\sigma_b^2$
Res.	$\sum\sum\sum(y_{ijv}-y_{ij.})^2$	$IJ(r-1)$	$\frac{S_e}{IJ(r-1)}$	σ^2	σ_e^2
Tot.	$\sum\sum\sum(y_{ijv}-y_{...})^2$	$IJr-1$			

The test statistic (for testing factor A) is not the same for fixed and random effects model.

Example: training centres (NKNW: p. 1121)

A big manufacturing company has three regional training centres. Two instructors are working in each centre. Each instructor trains a group of approximately 15 employees for three weeks.

The company would like to evaluate the effects of training centre (factor A) and the instructor (factor B) on the "outcome" evaluated by knowledge tests delivered after the end of the course. The experiment was carried out in a six weeks period: the training groups were created as usual and the instructors were assigned randomly.

		factor B	
		instructor 1	instructor 2
factor A (center)	Atlanta	25	14
		29	11
	Chicago	11	22
		6	18
	San Francisco	17	5
		20	2

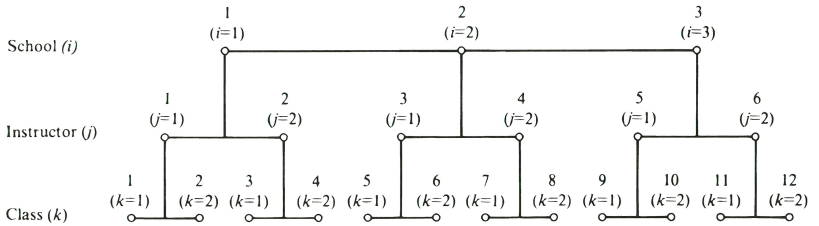
Instructor 1 in Atlanta $\neq$ instructor 1 in Chicago and, therefore, main effects of factor B do not have any meaning (instead, "averages of local instructors' effects" are confounded with centers).

(a) Crossed Factors

	Instructor (factor B)					
School (factor A)	1	2	3	4	5	6
Atlanta						
Chicago						
San Francisco						

(b) Nested Factors

	Instructor (factor B)					
School (factor A)	1	2	3	4	5	6
Atlanta						
Chicago						
San Francisco						



Nested design = hierarchicky uspořádaný experiment

Important consequence: j-th level of factor B (instructor) in one training centre does not correspond to the same level of the same factor in another center.

The model will contain the main effect of factor A and the interaction A:B (the effect of instructors "within" each center).

Airplanes (MJ: p. 145)

Airplane manufacturer wants to compare seven engines (A, B, ..., G) and three types of airplanes (1, 2, 3). Technical requirements imply that the first airplane can be used only with engines A, B, or C; the second airplane can be used with engines D or E and the third airplane only with engines F or G.

Three airplanes were produced for all possible combinations of type of airplane and engines. Testing pilot then evaluated resulting twenty one airplanes (in random order) and recorded the requested information.

In this case, we should use fixed effects model (in order to compare "treatment effects"):

$$y_{ijk} = \mu + L_i + M_{j(i)} + \varepsilon_{ijk}.$$

ANOVA table:

	df	**EMS**
Type of airplane	2	$\sigma_e^2 + \Phi^2(L)$
Engine (type of airplane)	4	$\sigma_e^2 + \Phi^2(M(L))$
Error (airplane)	14	σ_e^2

Coffee (MJ: p. 629)

Consumer magazine in USA investigated the variability of coffee price in towns with more than 20.000 inhabitants. The main aim was to evaluated the influence of state, town (within state) and store (within town within state).

The investigator chose r states randomly. Within each state, t_i towns with more than 20.000 were selected in random. Finally, n_{ij} stores were chosen in ij-th town, where the price of a certain type of coffee was established.

The variability of coffee price can be described using a random effects model:

$$y_{ijk} = \mu + s_i + m_{j(i)} + o_{k(ji)},$$

where $s_i \sim N(0, \sigma_s^2)$ is the effect of state, $m_{j(i)} \sim N(0, \sigma_m^2)$ effect of town, and $o_{k(ji)} \sim N(0, \sigma_o^2)$ effect of store.

The main aim is to estimate variances of random effects.

It can be useful to calculate:

$$Var(y_{ijk}) = \sigma_s^2 + \sigma_m^2 + \sigma_o^2$$

$$Cov(y_{ij1}, y_{ij2}) = \sigma_s^2 + \sigma_m^2$$

The correlation coefficient for coffee prices in two stores within a single town:

$$\rho_{y_{ij1}, y_{ij2}} = \frac{\sigma_s^2 + \sigma_m^2}{\sigma_s^2 + \sigma_m^2 + \sigma_o^2}$$

Similarly:

$$\rho_{y_{i11}, y_{i22}} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_m^2 + \sigma_o^2}$$

Question: how can be obtain confidence intervals for these correlation coefficients?

Repeated measures (opakovaná měření)

Repeated measures are not independent measurements in homogeneous conditions but measurements taken repeatedly (e.g. at different times) on the same subjects. Consequently, measurements taken on one subject are correlated.

Repeated measures can be seen as hierarchial experiments (nested design) or hierarchical experiments with more treatment effects (such as "split-plot design").

Layout for a Simple Repeated Measures Experiment

		TIME				
TRT	Subject	1	2	3		p
1	1	—	—	—	...	—
	2	—	—	—	...	—
	.	.	.	.	...	.
	.	.	.	.	...	.
	.	.	.	.	...	.
	n_1	—	—	—	...	—
2	1	—	—	—	...	—
	2	—	—	—	...	—
	.	.	.	.	...	.
	.	.	.	.	...	.
	.	.	.	.	...	.
	n_2	—	—	—	...	—
.	.	.	.	.	...	.
	.	.	.	.	...	.
	.	.	.	.	...	.
t	1	—	—	—	...	—
	2	—	—	—	...	—
	.	.	.	.	...	.
	.	.	.	.	...	.
	.	.	.	.	...	.
	n_t	—	—	—	...	—

Model with repeated measurements can be analyzed as:

- split-plot design (under certain assumptions),
- multivariate dependent variable (dimension is the number of "times"),
- correlated data (e.g., mixed models or GEE).

Horses (MJ, p. 136, in Czech)

Veterinární klinika zkoumá dva způsoby léčby zlomeniny kloubu u koní. Experiment je naplánován tak, že se vezme několik koní a zlomí se jim zvolený kloub, který se ošetří jednou nebo druhou technikou. Po čtyřech měsících se pak vyhodnotí vhodná míra uzdravení (tentokrát pomocí rentgenu, tj. bez opětovného lámání kostí).

Kromě vlivu způsobu léčby chce klinika vyhodnotit i to, jestli oba způsoby léčby fungují stejně dobře na přední i zadní nohu.

Experiment je potřeba rozumně naplánovat, protože koně jsou drazí a pro tento experiment jsou k dispozici pouze v omezeném množství (4 kusy). Každému koni se tedy bude muset zlomit přední i zadní noha.

MJ definují model s opakovanými pozorováními jako situaci, kdy úrovně jistého faktoru nelze blokům (zde koním) přiřazovat náhodně. Zde je určeno, že každému koni se zlomí přední i zadní noha...

Horses (MJ, p. 136)

Animal hospital investigates two treatments of joint fractures in horses. The investigator takes several horses, breaks its joints, and treats the fractures by one of two techniques. Suitable measure of recovery is evaluated after four months (using x-rays, i.e., without fracturing the joint).

Apart of the treatment effect, the hospital wants to establish whether the treatment effect is the same for front and hind legs.

Careful planning of the experiment is crucial because horses are expensive and there is only a limited supply. In this case, we can use four horses. Therefore, we have to break more joints on each horse.

MJ define repeated measures model as a situation, when levels of certain factor (typically time) cannot be assigned to blocks (horses) randomly. Here, we cannot assign time and front/hind legs.

Mixed model for repeated measures:

$$y_{ijk} = \mu_{ijk} + h_m + \varepsilon_{ijkm},$$

where μ_{ijk} is the expectation of the response for i-th treatment, j-th leg and k-th time, h_m is random effect of horse and ε_{ijkm} is random error (effect of "leg").

MJ compare two designs.

The first proposal is:

Horse			
1	2	3	4
$F_1P_1T_1$	$F_1P_1T_2$	$F_2P_1T_1$	$F_2P_1T_2$
$F_1P_2T_2$	$F_1P_2T_1$	$F_2P_2T_2$	$F_2P_2T_1$

(F_1 and F_2 are two treatments, P_1 and P_2 denote front and hind leg and T_1 and T_2 are times of the experiment).

The properties of a given design can be relatively easily assessed using the R function `alias()`

```
> lm1= lm(y~horse+f+p+t)
> alias(lm1)
Model :
y ~ horse + f + p + t

Complete :
   (Intercept) kun2 kun3 kun4 p2 t2
f2 0           0    1    1    0  0
```

Obviously, the treatment effect (F) cannot be distinguished from the difference between two pairs of horses. Hence, this experiment design should not be used to evaluate the treatment effect.

The second design is:

Horse			
1	2	3	4
$F_1P_1T_1$	$F_2P_1T_1$	$F_1P_1T_2$	$F_2P_1T_2$
$F_2P_2T_2$	$F_1P_2T_2$	$F_2P_2T_1$	$F_1P_2T_1$

```
> lm2= lm(y~kun+F*P+T)
> alias(lm2)
Model :
y ~ kun + F * P + T

Complete :
      (Intercept) kun2 kun3 kun4 F2   P2   T2
F2:P2    0         -1/2    0 -1/2  1/2  1/2    0
```

Attitudes / postoje (MJ: p. 583)

Sociologists have repeatedly (thrice in half-year intervals) investigated attitudes of families in towns and in countryside. 10 town families and 7 country families were recruited. Each family consists of two parents and one son.

In this dataset, we have two sets of repeated measurements: three repeated measurements (for each participant) and three members of each family (because levels of this factor—father/mother/son —cannot be randomly assigned to individual subjects).

	Family Member								
	Son			Father			Mother		
Family	T1	T2	T3	T1	T2	T3	T1	T2	T3
Urban									
1	17	17	19	18	19	21	16	16	18
2	12	14	15	19	19	21	16	16	18
3	8	10	11	16	18	19	11	12	12
4	5	7	7	12	12	13	13	14	14
5	2	5	6	12	14	14	14	16	18
6	9	11	11	16	17	18	14	15	16
7	8	9	9	19	20	20	15	16	18
8	13	14	16	16	17	18	18	18	20
9	11	12	13	13	16	17	7	8	10
10	19	20	20	13	15	15	11	12	12
Rural									
1	12	11	14	18	19	22	16	16	19
2	13	13	17	18	19	22	16	16	19
3	12	13	16	19	18	22	17	16	20
4	18	18	21	23	23	26	23	22	26
5	15	14	16	15	15	19	17	17	20
6	6	6	10	15	16	19	18	19	21
7	16	17	18	17	17	21	18	20	23

Assuming independence of random effects, the attitudes data set can be analyzed similarly as split-plot design.

MJ (page 585) investigate the structure of the correlation matrix of the nine-dimensional random error for each family using SAS PROC MIXED (AIC, AICC, BIC) and reach the conclusion that the Kronecker product of a unstructured correlation matrix for family members and compound symmetry for measurement times is appropriate for this data set.

This correlation structure is used to assess the significance of fixed effects (factors town/country, family member father/mother/son, times T1/T2/T3 and their interactions).

Cross-over design

When we want to compare two (or more) treatments in a repeated measurement setup, it may be helpful to expose each subject successively to different treatments.

I.e., a subject is first exposed to treatment A. After measuring the effect of treatment A and after disappearence of all physiological effects, the same subject is exposed to treatment B, et cetera. Hence, each subject serves as its own control (and the estimate of treatment effect will not be influenced by between-subject variability).

The randomization proceeds by creating sequences of treatments (here AB or BA) and these sequences are then randomly assigned to subjects.

The resulting experiment has two levels:

- treatment sequences AB or BA are randomly assigned to subjects (typically complete randomization),
- individual measurements (within subjects) create random blocks.

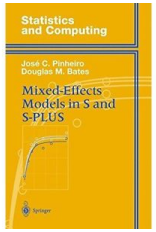
	Subject (animal)			
	1	2	...	n_i
Sequence 1 (AB)				
A (time 1)	y_{11A}	y_{12A}	...	y_{1n_1A}
B (time 2)	y_{11B}	y_{12B}	...	y_{1n_1B}
Sequence 2 (BA)				
B (time 1)	y_{21B}	y_{22B}	...	y_{2n_2B}
A (time 2)	y_{21A}	y_{22A}	...	y_{2n_2A}

$$y_{ijk} = \mu_{ik} + s_{ij} + \varepsilon_{ijk},$$

where μ_{ik} is the effect of k-th treatment in i-th sequence, s_{ij} is random effect of j-th subject (in i-th sequence) and ε_{ijk} is random error.

Mixed models in R and SAS

Pinheiro, J. C. & Bates, D. M. (2000). Mixed-effects models in S and S-PLUS. Springer.



(AKA the 'yellow book')

R: function `lme` in library `nlme` & function `lmer` in library `lme4`.

SAS: PROC MIXED

Dogs (PB, page 40), in Czech

Na univerzitě ve Wisconsinu u 10 psů pomocí počítačové tomografie opakovaně měřili střední intenzitu pixelů v obou podpažích po dobu 14 dní po aplikaci kontrastní látky.

Výsledná data Pixel jsou k dispozici v knihovně nlme.

Tato data jsou příkladem na hierarchické uspořádání náhodných efektů (nested design) a zároveň na opakovaná měření (opakovaná měření na jednotlivých psech).

Cílem experimentu je popsat chování pozorovaných růstových křivek (k tomu účelu musíme do R zapsat vhodný model).

Dogs (PB, page 40)

Mean pixel intensities were measured by CT scans in lymph nodes in the anxillary region (on left and right side) of 10 dogs repeatedly during 14 days after injecting a dye contrast.

The resulting data set Pixel is available in library nlme.

This is an example both for nested design (random effects model) and repeated measurements (subjects = dogs).

The aim of the experiment is to describe the observed growth curves. Therefore, we need a suitable mixed effects model in R.

Experts expect that the results measured on left and right side (of a dog) will be different but one cannot say in advance whether higher values will be observed on the left or the right side.

The plot `plot(Pixel)` shows that, at first, the pixel intensity increases and, after some time, starts to decrease. This suggests quadratic regression model. The left and the right side of each dog usually looks similar (up to a shift). PB are using mixed effects model with random intercept and slope (for the dependency of intensity on time) in order to describe the individual variability (between dogs).

```
library(nlme)
data(Pixel)
plot(Pixel)
fm1Pixel <- lme( pixel ~ day + I(day^2), data = Pixel,
                 random = list (Dog = ~ day, Side = ~ 1 ))
intervals(fm1Pixel)
plot(augPred(fm1Pixel))
```



```
Approximate 95% confidence intervals
 Fixed effects:    lower         est.        upper
(Intercept) 1053.0968388 1073.3391382 1093.5814376
day            4.3796925    6.1295971    7.8795016
I(day^2)      -0.4349038   -0.3673503   -0.2997967

 Random Effects:
  Level: Dog       lower       est.      upper
sd((Intercept))    15.9296203 28.3699038 50.5254631
sd(day)             1.0815006  1.8437505  3.1432398
cor((Intercept),day) -0.8943486 -0.5547222  0.1905316
  Level: Side      lower     est.    upper
sd((Intercept)) 10.41733 16.82431 27.17176

 Within-group standard error:
   lower      est.     upper
 7.634529  8.989606 10.585199
```



```
fm1Pixel.lme4=lmer(pixel~day+I(day^2)+(day|Dog)+(1|Dog:Side),
                   data=Pixel)
summary(fm1Pixel.lme4)

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 Dog:Side (Intercept) 283.055  16.824
 Dog      (Intercept) 804.854  28.370
          day           3.399   1.844   -0.55
 Residual              80.813   8.990
Number of obs: 102, groups: Dog:Side, 20; Dog, 10

Fixed effects:
             Estimate Std. Error t value
(Intercept) 1073.33914   10.17169  105.52
day            6.12960    0.87932    6.97
I(day^2)      -0.36735    0.03395  -10.82
```


Model building (according to PB)

Test for the random effect of side:

```
fm2Pixel=update(fm1Pixel, random=~ day |Dog)
anova(fm1Pixel, fm2Pixel)

Model df      AIC      BIC    logLik   Test L.Ratio p-value
    1  8 841.2102 861.9712 -412.6051
    2  7 884.5196 902.6854 -435.2598 1 vs 2 45.3094  <.0001
```

Test for the random slope:

```
fm3Pixel=update(fm1Pixel, random=~ 1 |Dog/Side)
anova(fm1Pixel, fm3Pixel)

Model df      AIC      BIC    logLik   Test  L.Ratio p-value
    1  8 841.2102 861.9712 -412.6051
    2  6 876.8390 892.4098 -432.4195 1 vs 2 39.62885  <.0001
```

Expert does not expect that the difference between left and right side won't be the same for all dogs. This assumption can be verified.

```
fm4Pixel=update(fm1Pixel, pixel ~ day + I(day^2) + Side)

...
Fixed effects: pixel ~ day + I(day^2) + Side
               Value Std.Error DF   t-value p-value
(Intercept) 1077.9484 10.862705 80  99.23388  0.0000
day            6.1296  0.879023 80   6.97323  0.0000
I(day^2)      -0.3674  0.033923 80 -10.82914  0.0000
SideR         -9.2175  7.626768  9  -1.20858  0.2576
...
```


Summary

The functions `lme()` and `lmer()` implement mixed models but, in certain circumstance, the choice matters: plotting is easier for `lme()` but the syntax of `lmer()` is more general.

Examples of some wrong models:

```
fm5Pixel=update(fm1Pixel, pixel ~ day + day^2 + Side)
fm6Pixel=update(fm1Pixel, pixel ~ day + Side)
fm7Pixel=update(fm1Pixel, pixel ~ dog )
```


Exercise

Analysis of simulated cupcakes (strip-plot design) from previous week:

Recommended steps:

- proper coding of variables day, oven (within day) and cupcake (within oven),
- randomization, i.e., the assignment of treatments to all levels in the experiment (i.e., temperatures should be assigned to ovens and recipes to batches of dough).,
- simulation of the response,
- analysis using either `lme()` or `lmer()`,
- comparison of the resulting estimates with the chosen (hence: known) parameters.

Week 5

Topic:

- Multiple comparisons / metody mnohonásobného porovnávání: Scheffé, Tukey, ...

See also *The Economist: Trouble in the Lab*

https://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble

Error rate

Comparisonwise error rate: the number of incorrect conclusions divided by the total number of tests in all analyzed experiments (i.e., the level of significance).

Experimentwise error rate: the number of experiments with at least one incorrect conclusion divided by the total number of experiments.

Familywise error rate: the probability that one or more conclusions are incorrect within a prespecified family of k tests or confidence intervals.

False discovery rate: the expectation of the ratio of the number of incorrectly rejected hypotheses and the number of rejected hypotheses.

Error rates in practice

EER & FWER:

EER is used for testing equality of expected values (ANOVA).

Knowing in advance that some expectations are different, the power of the test may be increased by applying FWER on a smaller number of null hypotheses (e.g. by using Bonferroni inequality).

FDR:

FDR is typically applied in genomics (with huge numbers of comparisons).

One-way ANOVA and contrasts

In one-way ANOVA model $y_{ij} = \mu_i + \varepsilon_{ij}$, we can investigate differences (contrasts) $\mu_i - \mu_j$. Obviously, $\widehat{\mu}_i - \widehat{\mu}_j \sim N(\mu_i - \mu_j, \sigma^2(1/n_i + 1/n_j))$.

Obviously, we reject $H_0 : \mu_i = \mu_j$ if

$$|\widehat{\mu}_i - \widehat{\mu}_j| \geq t_\nu(1-\alpha/2)\widehat{\sigma}\sqrt{\frac{1}{n_i}+\frac{1}{n_j}},$$

(with *comparisonwise error rate* $= \alpha$).

MJ (p. 47) calculated EER (experimentwise error rate) for various numbers of treatments:

I	2	3	4	5	6	8	10	20
CER	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
EER	0.05	0.12	0.20	0.28	0.36	0.47	0.59	0.90

LSD and Fisher's LSD

The test from previous slide is sometimes called *Least Significant Difference* (LSD).

LSD controls only the *comparisonwise error rate* and, therefore, it is not a usable method for multiple comparisons.

Fisher proposed to use LSD only after rejecting the null hypothesis of equality of expected values (by ANOVA). On one hand, this "modification" allows to control EER but, on the other hand, other properties are not good; for example, resulting confidence intervals are too short etc.

Bonferroni correction

Theorem: For random events E_i:

$$P\left(\bigcup_{i=1}^{k} E_i\right) \leq \sum_{i=1}^{k} P(E_i)$$

The application is straightforward: if $P(E_i)$ is the *comparisonwise error rate*, then $P(\bigcup_{i=1}^{n} E_i)$ is FWER for k tests. In order to guarantee that FWER $\leq \alpha$, it is enough to set $P(E_i) = \alpha/k$.

In practice, all p-values (obtained from k tests) are usually adjusted as $p^{bonferroni} = \min(1, kp)$, see `p.adjust()`.

Šidák's method, using $1-(1-\alpha)^{1/k}$ instead of α/k is a bit more powerful.

Monte Carlo approach

Monte Carlo method can be easily applied in order to control FWER for p contrasts $\sum c_{iq}\mu_i$, $q = 1, \ldots, p$.

1. Generated randomly data sets with the given structure (i.e., with given number of observations in each class).
2. Calculate t-statistics for given contrasts, i.e., $t_q = \sum c_{iq}\widehat{\mu}_i / \sqrt{\widehat{\sigma}^2 \sum c_{iq}^2/n_i}$.
3. Calculate the maximum of absolute values of t_q, $T_s = \max_{q=1,\ldots,p}(|t_q|)$.
4. Repeat steps 1, 2, and 3 and calculate $1-\alpha$ empirical quantile $T_{1-\alpha}$ of T_s.
5. Calculate t_q for the original data set and reject q-th null hypothesis if $|t_q| > T_{1-\alpha}$. We obtain confidence intervals $\sum c_{iq}\widehat{\mu}_i \pm T_{1-\alpha}\sqrt{\widehat{\sigma}^2 \sum c_{iq}^2/n_i}$.

Tasks and puls rate (MJ, p. 5 and 51), in Czech

Cílem experimentu bylo zjistit, jak různé pracovní úkoly ovlivňují tepovou frekvenci zaměstnanců. Celkem 78 zaměstnanců bylo náhodně zařazeno do šesti skupin tak, aby v každé skupině bylo 13 zaměstnanců. Každá skupina pak byla vycvičena a zařazena k vykonávání jistého pracovního úkolu.

Tepová frekvence pak byla u všech zaměstnanců změřena ve vybraný den jednu hodinu po začátku práce. Několik zaměstnanců však bohužel podalo výpověď ještě v průběhu výcviku a tak získaná data nejsou vyvážená.

Tasks and puls rate (MJ, p. 5 and 51)

The aim of the experiment was to investigate the effect of various tasks on puls rate. 78 employees were randomized into six groups, where each group contained 13 people. Each group was trained and assigned to a certain task.

The puls rates were measured one hour after the start of work on a chosen day. Unfortunately, some employees quit already during the training and, therefore, the resulting data set is not balanced.

Pulsation Data and Summary Information for Six Tasks

	Task					
	1	2	3	4	5	6
	27	29	34	34	28	28
	31	28	36	34	28	26
	26	37	34	43	26	29
	32	24	41	44	35	25
	39	35	30	40	31	35
	37	40	44	47	30	34
	38	40	44	34	34	37
	39	31	32	31	34	28
	30	30	32	45	26	21
	28	25	31	28	20	28
	27	29			41	26
	27	25			21	
	34					
$y_{i\cdot}$	415	373	358	380	354	317
n_i	13	12	10	10	12	11
$\bar{y}_{i\cdot}$	31.9231	31.0833	35.8000	38.0000	29.5000	28.8182
SS_i	294.9231	352.9167	253.6000	392.0000	397.0000	225.6364

Percentage Points Used for All Pairwise Comparisons of the Six Task Means

Simulation Results

Method	95% Quantile	Estimated α	99% Confidence Limits	
Simulate	2.932480	0.0500	0.0450	0.0550
Tukey–Kramer	2.940710	0.0486	0.0436	0.0535
Bonferroni	3.053188	0.0359	0.0316	0.0401
Šidák	3.044940	0.0370	0.0326	0.0413
Scheffé	3.437389	0.0131	0.0105	0.0157
t	1.998972	0.3556	0.3446	0.3666

T-statistics for a difference of two expectations are compared to these critical values.

Critical Differences Used to Compare the Differences between Pairs of Means for the Unadjusted t and Several Multiple Comparison Procedures

TASK	_TASK	Estimate	Standard Error	t	Bonferroni	Tukey–Kramer	Scheffé	Šidák	Simulate
1	2	0.840	2.225	4.449	6.795	6.544	7.650	6.776	6.526
1	3	–3.877	2.338	4.674	7.139	6.876	8.038	7.120	6.857
1	4	–6.077	2.338	4.674	7.139	6.876	8.038	7.120	6.857
1	5	2.423	2.225	4.449	6.795	6.544	7.650	6.776	6.526
1	6	3.105	2.277	4.553	6.953	6.697	7.828	6.935	6.679
2	3	–4.717	2.380	4.758	7.267	7.000	8.182	7.248	6.980
2	4	–6.917	2.380	4.758	7.267	7.000	8.182	7.248	6.980
2	5	1.583	2.270	4.537	6.929	6.674	7.801	6.911	6.655
2	6	2.265	2.321	4.639	7.085	6.824	7.977	7.066	6.805
3	4	–2.200	2.486	4.970	7.591	7.311	8.546	7.570	7.291
3	5	6.300	2.380	4.758	7.267	7.000	8.182	7.248	6.980
3	6	6.982	2.429	4.855	7.416	7.143	8.349	7.396	7.123
4	5	8.500	2.380	4.758	7.267	7.000	8.182	7.248	6.980
4	6	9.182	2.429	4.855	7.416	7.143	8.349	7.396	7.123
5	6	0.682	2.321	4.639	7.085	6.824	7.977	7.066	6.805

Adjusted Significance Levels to Test the Equality of All Pairwise Comparisons of TASK Minus _TASK Obtained from Six Procedures Where t Corresponds to the Unadjusted t

TASK	_TASK	t	Bonferroni	Tukey–Kramer	Scheffé	Šidák	Simulate
1	2	0.7072	1.0000	0.9990	0.9996	1.0000	0.9990
1	3	0.1024	1.0000	0.5642	0.7378	0.8021	0.5646
1	4	0.0117	0.1751	0.1129	0.2552	0.1615	0.1111
1	5	0.2805	1.0000	0.8840	0.9446	0.9928	0.8804
1	6	0.1777	1.0000	0.7484	0.8661	0.9469	0.7501
2	3	0.0520	0.7795	0.3645	0.5642	0.5509	0.3657
2	4	0.0051	0.0761	0.0546	0.1506	0.0735	0.0545
2	5	0.4880	1.0000	0.9815	0.9923	1.0000	0.9813
2	6	0.3328	1.0000	0.9238	0.9651	0.9977	0.9234
3	4	0.3796	1.0000	0.9488	0.9772	0.9992	0.9474
3	5	0.0103	0.1543	0.1014	0.2364	0.1437	0.0985
3	6	0.0055	0.0831	0.0590	0.1596	0.0799	0.0584
4	5	0.0007	0.0104	0.0087	0.0366	0.0104	0.0090
4	6	0.0004	0.0053	0.0046	0.0219	0.0053	0.0052
5	6	0.7699	1.0000	0.9997	0.9999	1.0000	0.9998

Scheffé (Anděl 1985, theorem 2, page 147)

Theorem: Assume that $X \sim N_m(\mu, \sigma^2 V)$, where V is a known matrix and $\sigma^2 > 0$ is unknown parameter. Let A denote some t-dimensional subspace of $\mathbb{R}^m$. Let s^2 be an independent estimator of σ^2 with ν degrees of freedom (i.e., $\nu s^2/\sigma^2 \sim \chi^2_\nu$ and s^2 and X are independent). Then, the probability that the inequality

$$|a^\top X - a^\top \mu| \leq \sqrt{t s^2 F_{t,\nu}(1-\alpha) a^\top V a}$$

holds, for all $a \in A$ simultaneously, is equal to $1 - \alpha$.

Testing the equality of several expected values

Let $\mu = (\mu_1, \ldots, \mu_m)$ and $V = (v_{ij})$. The null hypothesis $H_0 : \mu_1 = \cdots = \mu_m$ holds if and only if $a^\top \mu = 0$ for all vectors a such that $a^\top \mathbf{1}_m = 0$. These vectors a define a subspace of dimension $m - 1$ in $\mathbb{R}^m$.

In order to test that $\mu_i = \mu_j$ for all $i \neq j$, we define vectors a_{ij} with i-th element equal to 1, j-th element equal to -1 and all other elements equal to zero. Obviously, $\mu_i = \mu_j$ holds if and only if $a_{ij}^\top \mu = 0$.

Clearly, $a_{ij}^\top V a_{ij} = v_{ii} + v_{jj} - 2v_{ij}$ and, by Scheffé' theorem, we obtain that the inequality

$$|X_i - X_j| \leq \sqrt{(v_{ii} + v_{jj} - 2v_{ij})(m-1) s^2 F_{m-1,\nu}(1-\alpha)}$$

holds for all pairs (i, j) with probability greater or equal to $1 - \alpha$.

Example: one-way ANOVA

Example: In the unbalanced one-way ANOVA model

$$y_{ij} = \mu_i + \varepsilon_{ij} = \mu + \alpha_i + \varepsilon_{ij},$$

the sample means $y_{i\cdot} = \sum_j y_{ij}/n_j$ estimate the parameters μ_i and the variance matrix of the random vector $(y_{1\cdot}, \ldots, y_{I\cdot})$ is $\sigma^2 V$, where $V = \mathrm{diag}(n_1^{-1}, \ldots, n_I^{-1})$.

Applying Scheffé's method, we reject the equality of i-th and k-th expectation if

$$|y_{i\cdot} - y_{k\cdot}| > \sqrt{(n_i^{-1} + n_j^{-1})(I-1) s^2 F_{I-1,n-I}(1-\alpha)}.$$

Notice that Scheffé's method also shows *which pairs* of expectations are significantly different.

Tukey (Anděl 1985, page 150)

Theorem: Assume that $X_1, \ldots, X_m$ is a random sample from $N(\mu, \sigma^2)$, $\sigma^2 > 0$. Denote by

$$R = \max_i X_i - \min_i X_i$$

the so-called *range*. Let s^2 denote an independent estimator of the variance σ^2 with ν degrees of freedom; this means that $\nu s^2/\sigma^2 \sim \chi^2_\nu$ and s^2 and $X = (X_1, \ldots, X_m)^\top$ are independent. Denote by

$$Q = R/s$$

the so-called *studentized range*. Then, the distribution of the random variable Q, denoted by the symbol $q_{m,\nu}$, does not depend on μ and σ^2.

In the following, $q_{m,\nu}(1-\alpha)$ will denote the $1-\alpha$ quantile of the studentized range distribution $q_{m,\nu}$.

These quantiles are implemented in the function `Tukey()` in R.

Tukey: application to 'pair differences'

Theorem: Assume that $X_1, \ldots, X_m$ are independent random variables such that $X_i \sim N(\mu_i, b^2\sigma^2)$, $i = 1, \ldots, m$, where b is a known positive constant. Let s^2 denote an independent estimator of the variance σ^2 with ν degrees of freedom (i.e., $\nu s^2/\sigma^2 \sim \chi^2_\nu$). Define $T = bq_{m,\nu}(1-\alpha)$. Then, the probability that inequalities

$$X_i - X_j - Ts \leq \mu_i - \mu_j \leq X_i - X_j + Ts$$

hold for all pairs simultaneously, is $1 - \alpha$.

Proof: Denoting $Z_i = X_i - \mu_i$, the previous theorem implies that

$$P\left\{\frac{\max Z_i - \min Z_i}{bs} \leq q_{m,\nu}(1-\alpha)\right\} = 1 - \alpha.$$

Example: balanced one-way ANOVA

Example: Balanced one-way layout guarantees that all random variables $y_{1.}, \ldots, y_{I.}$ have the same variance. Denoting by r denote the number of observations in each subclass, obviously $y_{i.} \sim N(\mu_i, \sigma^2/r)$ and, choosing $b = 1/r$ in the previous theorem, we reject equality of expectations in i-th and j-th group if

$$|y_{i.} - y_{j.}| > sq_{I,n-I}(1-\alpha)/\sqrt{r}.$$

This method is usable only for balanced one-way setup but it is obviously stronger than Scheffé's method if
$q_{I,n-I}(1-\alpha)/\sqrt{r} < \sqrt{(n_i^{-1} + n_j^{-1})(I-1)F_{I-1,n-I}(1-\alpha)}$.

Example: Recall the laser example from the first lecture:

```
> lm.laser=lm(strength~ordered(power),data=laser)
> summary(aov.laser <- aov(lm.laser))
               Df Sum Sq Mean Sq F value Pr(>F)
ordered(power)  2 224.18   112.1   11.32 0.0092 **
Residuals       6  59.42     9.9
```

Sample averages for each laser power are:

```
> tapply(laser$strength,laser$power,mean)
      40       50       60
24.77000 31.34333 36.98333
```

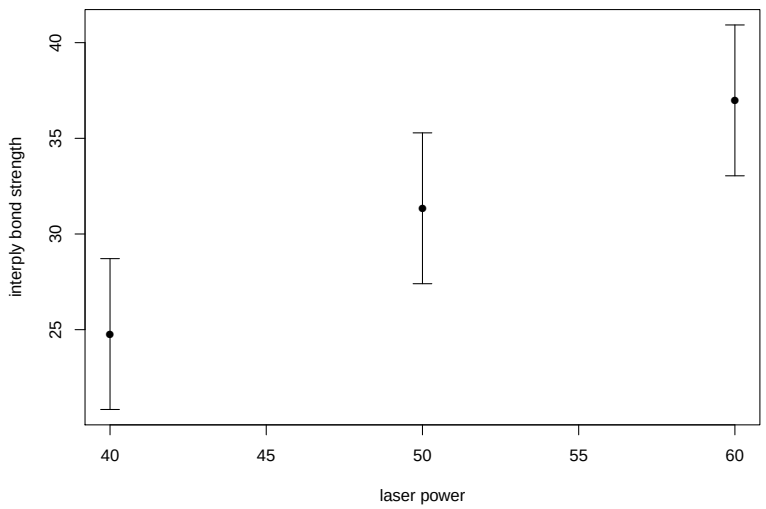

Critical values for the studentized range:

```
> summary(lm.laser)$sigma*qtukey(0.95,nmeans=3,df=6)/sqrt(3)
[1] 7.883982
```

Our conclusion should agree with the R function TukeyHSD:

```
> TukeyHSD(aov.laser,"ordered(power)")
           diff       lwr      upr     p adj
50-40  6.573333 -1.310648 14.45731 0.0947896
60-40 12.213333  4.329352 20.09731 0.0075171
60-50  5.640000 -2.243982 13.52398 0.1506523
```

We can mark down different groups more easily after sorting the groups in increasing order (according to observed sample means).



Modification of Tukey's method

Tukey's method was designed for comparing *pairs of expected values* in a homoskedastic situation with independent observations.

Therefore, Tukey's method can be applied directly also for comparisons of treatment effects in a balanced two-way ANOVA without interactions (interactions complicate the interpretation of treatment effects).

Let us have a look at modifications for:

- set of all contrasts (like Scheffé's method),
- heteroskedasticity (e.g., for unbalanced ANOVA models),
- correlated observations.

Theorem: Assume that $X_1, \ldots, X_m$ are independent random variables such that $X_i \sim N(\mu_i, b^2\sigma^2)$, $i = 1, \ldots, m$, where b is a known positive constant and that s^2 is an independent estimator of the variance σ^2 with ν degrees of freedom.

Denoting $T = bq_{m,\nu}(1-\alpha)$, the probability that

$$\sum_i c_i X_i - Ts \sum_i |c_i|/2 \leq \sum_i c_i \mu_i \leq \sum_i c_i X_i + Ts \sum_i |c_i|/2$$

holds simultaneously for all contrasts ($\sum_i c_i = 0$) is equal to $1-\alpha$.

Proof (Scheffé, Analysis of Variance, p. 74):

The statement follows from the inequality $|\sum c_i u_i| \leq h(\sum |c_i|/2)$ that holds for all contrasts (i.e. $\sum_i c_i = 0$) if $|u_i - u_j| \leq h$ (it suffices to plug in $u_i = X_i - \mu_i$ and $h = Ts$).

Tukey-Kramer

Tukey-Kramer method rejects $H_0 : \mu_i = \mu_j$ if

$$|\mu_i - \mu_j| > q_{I,\nu}(1-\alpha)\sqrt{\frac{s^2}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)},$$

where n_i is the number of observations in i-th subclass.

Tukey's method is obtained for $n_i \equiv r$.

Example: incomplete blocks

Example: In the first lecture, we have analyzed tyres using an incomplete randomized block design. In this case, the treatment effect estimators are not independent:

```
trv = c (238,196,254,238,213,312,279,334,421,308,367,412)
pneu= factor(c (1,2,3,1,2,4,1,3,4,2,3,4))
typ = factor(rep(1:4,each=3))
drzi=data.frame(trv,pneu,typ)

lm.drzi1=lm(trv~typ + pneu,data=drzi)
anova(lm.drzi1)
mm1=model.matrix(lm.drzi1)
v1=solve(t(mm1)%*%mm1)
v1
```


Correlated observations

Theorem: Assume that $X_1, \ldots, X_m$ are random variables such that $X_i \sim N(\mu_i, b^2\sigma^2)$, $i = 1, \ldots, m$, where b is a known positive constant and that $Cov(X_i, X_j) = c\sigma^2$ for all $i \neq j$, where constants b and c satisfy $-b^2 \leq (m-1)c \leq 0$. Let s^2 denote an independent estimator of the variance σ^2 with ν degrees of freedom.

Setting $T = \sqrt{b^2 - c}\, q_{m,\nu}(1-\alpha)$, the probability that

$$\sum_i c_i X_i - Ts \sum_i |c_i|/2 \leq \sum_i c_i \mu_i \leq \sum_i c_i X_i + Ts \sum_i |c_i|/2$$

holds simultaneously for all contrasts ($\sum_i c_i = 0$) is equal to $1 - \alpha$.

Note: The covariance structure in this theorem looks similar to "compound symmetry" from the previous week (coffee or attitudes example) but the correlations have different sign.

Proof: Let us find $X_0 \sim N(0, \sigma_0^2)$ such that $\tilde{X}_i = X_0 + X_i$, $i = 1, \ldots, m$ are independent. This is quite simple because $Cov(\tilde{X}_i, \tilde{X}_j) = \cdots = c\sigma^2 + \sigma_0^2$ and it suffices to choose $\sigma_0^2 = -c\sigma^2$.

Next,

$$Var\, \tilde{X}_i = Var(X_i + X_0) = (b^2 - c)\sigma^2$$

and we can apply Tukey's theorem on the independent random variables $Var\, \tilde{X}_i$.

Question: Can we modify the above proof for positive correlations?

Example: Do we have constant correlation in the tyre examples?

```
> lm.drzi2=lm(trv~ -1 +  typ + pneu,data=drzi)
> anova(lm.drzi2)
> mm2=model.matrix(lm.drzi2)
> v2=solve(t(mm2)%*%mm2)
> v2[1:4,1:4]
          typ1      typ2      typ3      typ4
typ1 0.5833333 0.2083333 0.2083333 0.3333333
typ2 0.2083333 0.5833333 0.2083333 0.3333333
typ3 0.2083333 0.2083333 0.5833333 0.3333333
typ4 0.3333333 0.3333333 0.3333333 0.8333333
> coef(lm.drzi2)
   typ1    typ2    typ3    typ4   pneu2   pneu3   pneu4
224.792 229.167 301.042 325.667 -20.875  34.500  96.375
```

Example: How comes that the variances and correlations are constant in the following output?

```
> lm.drzi3=lm(trv~ -1 +  typ + ordered(pneu),data=drzi)
> anova(lm.drzi3)
> mm3=model.matrix(lm.drzi3)
> v3=solve(t(mm3)%*%mm3)
> v3[1:4,1:4]
            typ1        typ2        typ3        typ4
typ1  0.36458333 -0.01041667 -0.01041667 -0.01041667
typ2 -0.01041667  0.36458333 -0.01041667 -0.01041667
typ3 -0.01041667 -0.01041667  0.36458333 -0.01041667
typ4 -0.01041667 -0.01041667 -0.01041667  0.36458333
> round(coef(lm.drzi3),3)
       typ1        typ2        typ3        typ4
    252.292     256.667     328.542     353.167
ordered(pneu).L ordered(pneu).Q ordered(pneu).C
         77.033          41.375         -15.597
```


Further useful methods

Dunnet Variant of Tukey's method for comparisons of $I-1$ treatment effects to the control group (instead of all pairwise comparisons).

Multivariate t If we are interested in p linearly independent contrasts, the situation can be "reduced" to finding a quantile of maximum of absolute values of p independent t-distributions.

Holm P-values are sorted (in increasing order) and compared sequentially to $\alpha/(p-k+1)$, where p denotes the number of comparisons and k the ranks of p-values. If the null hypothesis is not rejected for some k_0, we stop testing and null hypotheses are not rejected for any $k > k_0$. This method is obviously stronger than Bonferroni (apart of the first comparison).

Benjamini-Hochberg Sequential procedure controlling only False Discovery Rate = FDR (instead of Familywise Error Rate = FWER).

Practical recommendations (MJ, p. 45), part I

- Multiple comparison methods may not be applied in exploratory studies (typically followed by a carefully planned *confirmatory* experiment).
- We should use the most powerful test (see the table with comparison of critical values for studentized differences).
- Dunnet's test is used to compare treatment effects to a control group (there is both a one- and both-sided version).
- Tukey's method is recommended for all pairwise comparisons (for constant sample sizes). For unequal sample sizes, critical values can be obtained by Monte Carlo methods.
- "Multivariate t" is recommended for independent contrasts. For correlated contrasts, more precise critical value can be obtained by Monte Carlo methods.

Practical recommendations (MJ, p. 45), part II

- Confidence intervals for a small number of linear combinations of parameters can be obtained by Bonferroni method. Scheffé method is better for more linear combinations (20). Šidák and Holm methods work well for testing several uncorrelated linear combinations. Bootstrap or Monte Carlo methods can be used for correlated linear combinations.
- Scheffé's method is recommended for 'data snooping', i.e., when hypotheses are based on observed data (it holds for *all* contrasts).
- Only FDR is controlled in genomics and in studies with many tests (1000), e.g., Benjamini-Hochberg method.
- Multiple comparison methods should not be used in a *safety study*, i.e., when dangers or side effects of some treatments are compared to a control group (because multiple comparisons would increase probability of type II error).

Measurement of jump height and speed: 796 children

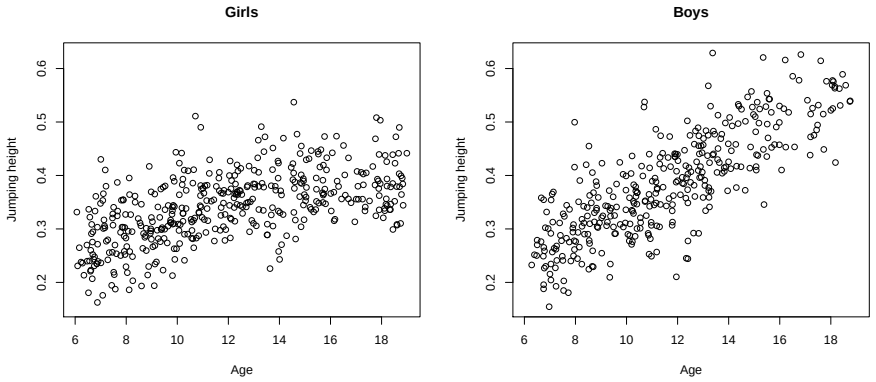
432 girls, 364 boys (6–19 years)

Single two-legged jump
–aims to achieve *maximum jump height*.

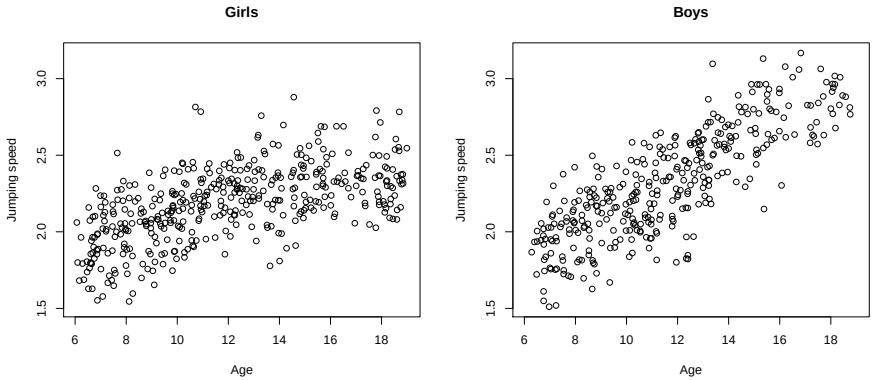
Multiple one-legged hopping
–aims to achieve maximum voluntary forefoot ground reaction force during landing. One possible application of this test is to evaluate the maximal force to which the tibia is exposed, and thus it might serve to evaluate the muscle-bone unit.

Šumník, Z., Matysková, J., Hlávka, Z., Durdilová, L., Souček, O., & Zemková, D. (2013). Reference data for jumping mechanography in healthy children and adolescents aged 6-18 years. *Journal of musculoskeletal & neuronal interactions*, 13(3), 259-273.

Jump height



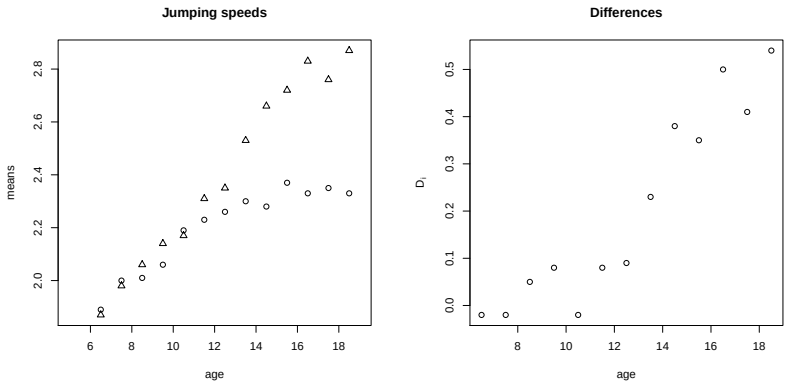
Jumping speed



	girls		boys		
Age	$\overline{Y}_1\ (\widehat{\sigma}_1)$	n_1	$\overline{Y}_2\ (\widehat{\sigma}_2)$	n_2	t-test
6	1.89(0.17)	33	1.87(0.18)	19	0.7799
7	2.00(0.21)	43	1.98(0.20)	38	0.6459
8	2.01(0.21)	33	2.06(0.21)	38	0.3688
9	2.06(0.18)	42	2.14(0.18)	29	0.0811.
10	2.19(0.22)	42	2.17(0.19)	45	0.7131
11	2.23(0.15)	30	2.31(0.23)	37	0.0616.
12	2.26(0.13)	41	2.35(0.23)	40	0.0473*
13	2.30(0.22)	32	2.53(0.21)	36	0.0001***
14	2.28(0.23)	31	2.66(0.19)	20	0.0000***
15	2.37(0.17)	29	2.72(0.22)	26	0.0000***
16	2.33(0.19)	17	2.83(0.28)	9	0.0005***
17	2.35(0.18)	25	2.76(0.16)	13	0.0000***
18	2.33(0.17)	34	2.87(0.10)	14	0.0000***

Standard multiple testing adjustment

	p-value	Holm	Hoch.	Hommel	Bonf.	B-H	B-Y
6	0.780	1.000	0.780	0.780	1.000	0.780	1.000
7	0.646	1.000	0.780	0.780	1.000	0.763	1.000
8	0.369	1.000	0.780	0.780	1.000	0.479	1.000
9	0.081.	0.405	0.41	0.405	1.000	0.117	0.372
10	0.713	1.000	0.780	0.780	1.000	0.773	1.000
11	0.062.	0.369	0.370	0.308	0.800	0.100	0.318
12	0.047*	0.331	0.331	0.243	0.615	0.088.	0.279
13	0.000***	0.001***	0.001***	0.006***	0.001***	0.000***	0.001***
14	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
15	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
16	0.001***	0.004**	0.004**	0.004***	0.006**	0.001**	0.003**
17	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***
18	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***	0.000***



Observed sample means of jumping speed for boys (△) and girls (◯) in thirteen age categories. The right plot shows the observed differences D_i.

Traditional one-sample change-point analysis

$Y_1, \ldots, Y_n$ are independent observations with distribution functions $F_1, \ldots, F_n$

One-sample change-point problem concerns the test of

$$H_0 : F_1 = F_2 = \cdots = F_n \quad \text{against} \quad H_1 : \exists k_0 \text{ such that } F_{k_0} \neq F_{k_0+1}.$$

Statistical inference is complicated because the location of the possible change-point (k_0) is not known and there could be more than one change-point.

A lot of results concerning changes in mean (shifts) and variance, change in regression coefficient, abrupt and gradual change, changes occurring in time series, off-line and on-line testing (monitoring), nonparametric approach, etc.

General two-sample change-point problem

Two independent random samples $Y_{1,1}, \ldots, Y_{1,N_1}$ and $Y_{2,1}, \ldots, Y_{2,N_2}$ are collected at the ordered time points $t_1 < \cdots < t_n$.

At each time t_i, we observe $n_{1,i}$ observations from the first sample and $n_{2,i}$ observations from the second sample.

With $F_{j,i}$ denoting the distribution function of $Y_{j,i}$, we introduce the general two-sample change-point problem:

$$\begin{aligned} H_0 &: \quad F_{1,i} = F_{2,i} \text{ for all } i = 1, \ldots, n \\ H_1 &: \quad \exists k_0 \text{ such that } F_{1,k} \neq F_{2,k}, \text{ for } k \geq k_0. \end{aligned}$$

Apart of testing H_0, we are interested in estimating k_0.

Additional assumption of **gradual** change leads to *two-sample gradual change-point analysis*.

Model of gradual change

1. Observations Y_{jik} $(j = 1, 2; k = 1 \ldots, n_{ji})$ are obtained at time i $(i = 1 \ldots, n)$.
2. All observations are independent.
3. $E(\overline{Y}_{1i} - \overline{Y}_{2i}) = 0 + \delta_n (i - k_0)_+ / n$ $(i = 1, \ldots, n)$, where δ_n and k_0 are unknown parameters and $k_0 = n\theta_0$ for some $\theta_0 \in (0, 1)$.
4. $Var(Y_{jik}) = \sigma^2_{ji} > 0$ $(j = 1, 2; i = 1, \ldots, n; k = 1 \ldots, n_{ji})$.

We use the notation $\overline{Y}_{ji} = \sum_{k=1}^{n_{ji}} Y_{jik}/n_{ji}$, $a_+ = \max(a, 0)$ with k_0 denoting the unknown location of the change point, μ the unknown expectation of difference before the change, and δ_n the slope (speed) of the gradual change after k_0.

See also [Hinkley (1971). Inference in two-phase regression, JASA 66(336): 736–743], [Hušková, M. (1999). Gradual changes versus abrupt changes. JSPI 76(1), 109-125].

Least squares estimators

The least squares estimator of k_0 (under homoskedasticity) is:

$$\widehat{k} = \arg \max_{k \in (1,n)} \left[\frac{\left\{ \sum_{i=1}^{n} x_{ik} (\overline{Y}_{1i} - \overline{Y}_{2i}) \right\}^2}{\sum_{i=1}^{n} x_{ik}^2} \right].$$

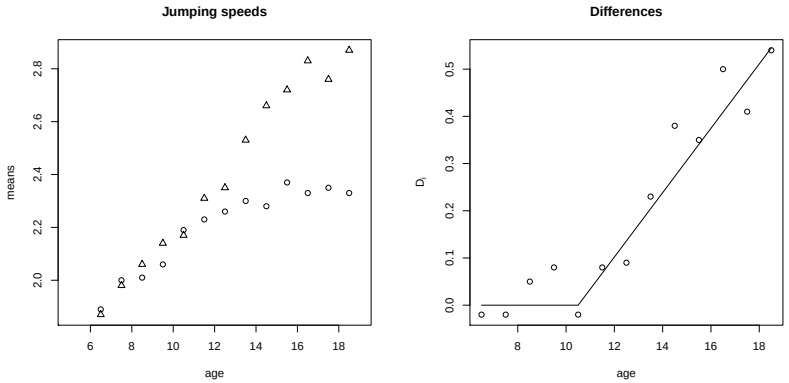
Denoting by $\sigma_i^2 = Var(\overline{Y}_{1i} - \overline{Y}_{2i})$, we define:

$$\widehat{k}_{\sigma^2} = \arg \max_{k \in (1,n)} \left[\frac{\left\{ \sum_{i=1}^{n} x_{ik} (\overline{Y}_{1i} - \overline{Y}_{2i})/\sigma_i^2 \right\}^2}{\sum_{i=1}^{n} x_{ik}^2/\sigma_i^2} \right].$$

Using sample variances $\hat{\sigma}^2_{ji}$ observed in each (age/gender) category, we arrive to:

$$\widehat{k}_{\hat{\sigma}^2} = \arg \max_{k \in (1,n)} \left[\frac{\left\{ \sum_{i=1}^{n} x_{ik} (\overline{Y}_{1i} - \overline{Y}_{2i})/\hat{\sigma}_i^2 \right\}^2}{\sum_{i=1}^{n} x_{ik}^2/\hat{\sigma}_i^2} \right],$$

where $\hat{\sigma}_i^2 = \hat{\sigma}^2_{1i}/n_{1i} + \hat{\sigma}^2_{2i}/n_{2i}$.

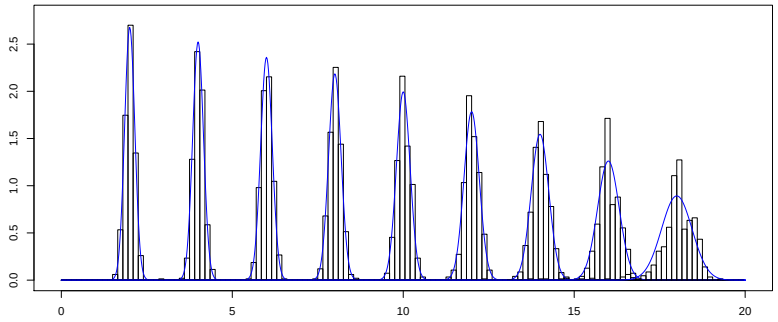


Observed sample means of jumping speed for boys (△) and girls (◯) in thirteen age categories. The right plot shows the observed differences D_i and the least squares fit.

Asymptotic distribution of $\widehat{k}$ (homoscedastic case)

Following [*Hušková M. (1998) Estimators in the location model with gradual changes, Comment.Math.Univ.Carolin. 39,1: 147–157*], it can be shown that:

$$\widehat{k} - k_0 \sim N\left(0, \frac{\sigma^2 n^2 4}{\delta^2 (n - k_0)}\right).$$



Bootstrap for $\widehat{k}_{\widehat{\sigma}^2}$ (heteroscedastic case)

We use wild bootstrap and simulate bootstrap replicates from normal distribution with the same parameters as the observed sample mean differences $D_i = \overline{Y}_{1i} - \overline{Y}_{2i}$ with standard errors $\hat{\sigma}_i = \sqrt{\hat{\sigma}_{1i}^2/n_{1i} + \hat{\sigma}_{2i}^2/n_{2i}}$.

Bootstrap algorithm:

1. Estimate parameters δ and k_0 and calculate *fitted values* $\widehat{D}_i = \hat{\delta}(i - \widehat{k}_{\widehat{\sigma}^2})_+/n$.
2. Generate bootstrap sample $D_i^* = \widehat{D}_i + \hat{\sigma}_i \varepsilon_i^*$, where $\varepsilon_i^* \sim N(0,1)$, $i = 1, \ldots, n$ and calculate the change-point estimator $\widehat{k}^*_{\widehat{\sigma}^2}$.
3. Repeat the previous step B times in order to obtain a random sample from $\widehat{k}^*_{\widehat{\sigma}^2} - \widehat{k}_{\widehat{\sigma}^2}$.
4. Finally, the quantiles q_α^* of $\widehat{k}^*_{\widehat{\sigma}^2} - \widehat{k}_{\widehat{\sigma}^2}$ are used as an approximation of the quantiles q_α of $\widehat{k}_{\widehat{\sigma}^2} - k_0$.

One-sided confidence intervals and p-values

The bootstrap approximation of q_α by q_α^* leads to:

$$1 - \alpha = P(\widehat{k}_{\widehat{\sigma}^2} - k_0 > q_\alpha) \doteq P(\widehat{k}_{\widehat{\sigma}^2} - k_0 > q_\alpha^*) = P(k_0 < \widehat{k}_{\widehat{\sigma}^2} - q_\alpha^*).$$

The resulting one-sided confidence interval is $(-\infty, \widehat{k}_{\widehat{\sigma}^2} - q_\alpha^*)$.

Similarly, for fixed k_1, wild bootstrap can be used to calculate p-values for the test of the null hypothesis

$$H_0 : k_0 \geq k_1 \text{ against } H_1 : k_0 < k_1$$

$$\text{p-value} = P^*(\widehat{k}^*_{\widehat{\sigma}^2} - \widehat{k}_{\widehat{\sigma}^2} < \widehat{k}_{\widehat{\sigma}^2} - k_1) \doteq \sum_{b=1}^{B} I(\widehat{k}^*_{\widehat{\sigma}^2,b} - \widehat{k}_{\widehat{\sigma}^2} < \widehat{k}_{\widehat{\sigma}^2} - k_1)/B$$

Coverage of 95% confidence intervals

	k_0	$\widehat{k}$	$\widehat{k}_{\widehat{\sigma}^2}$
$\sigma^2 = 1$	2	96.0	95.3
	4	94.5	95.1
	8	95.8	94.8
	12	97.8	97.3
	16	98.5	98.4
	18	98.7	97.8
$\sigma^2 = 2$	2	95.7	94.3
	4	96.8	96.2
	8	97.0	96.5
	12	97.1	96.2
	16	98.8	98.0
	18	90.8	89.3

1000 simulations, $n = 20$, $n_{ij} = 20$, $B = 2000$.

Application to jumping speeds

For the real data analysis, we have to correctly understand the observed sample means (and its differences):

Index	Label	Meaning		$\overline{Y}_1$ ($\widehat{\sigma}_1$)	$\overline{Y}_2$ ($\widehat{\sigma}_2$)
1	6	6–7 years	∼6.5 years	1.89(0.17)	1.87(0.18)
⋮	⋮	⋮	⋮	⋮	⋮
13	18	18–19 years	∼18.5 years	2.33(0.17)	2.87(0.10)

The estimator $\widehat{k}_{\widehat{\sigma}^2} = 5$ corresponds to the estimated change point $5 + 5.5 = 10.5$ years and the upper limit of one-sided 95% confidence interval is $6.14 + 5.5 = 11.64$ years.

In our data set, the observed sample means are based on continuous explanatory variable: this may lead to rounding bias. Therefore, we consider also a bias corrected estimator $\widehat{k}^{\text{bc}}_{\widehat{\sigma}^2}$.

Testing for the change-point location

Bootstrap is used in order to calculate one-sided confidence intervals and p-values for $H_0 : k_0 \geq k_1$ against $H_1 : k_0 < k_1$ for chosen $k_1 = 1, \ldots, 13$ (i.e., $k_1^{\text{age}} = 6, \ldots, 18$).

For correct interpretation, we have to understand the test (the interpretation depends on k_1 and it is **not** the same as the interpretation of the two-sample t-test).

	$\overline{Y}_1$ $(\widehat{\sigma}_1)$	n_1	$\overline{Y}_2$ $(\widehat{\sigma}_2)$	n_2	t-test	CPT
⋮	⋮	⋮	⋮	⋮	⋮	⋮
12–13	2.26(0.13)	41	2.35(0.23)	40	0.047*	0.003*
⋮	⋮	⋮	⋮	⋮	⋮	⋮

	t-test	Bonferroni	BH	$\widehat{k}_{\widehat{\sigma}^2}$	$\widehat{k}^{\text{bc}}_{\widehat{\sigma}^2}$	Age
6–7	0.780	1.000	0.780	1.000	1.000	6
7–8	0.646	1.000	0.763	1.000	1.000	7
8–9	0.369	1.000	0.479	1.000	1.000	8
9–10	0.081.	1.000	0.117	0.999	0.997	9
10–11	0.713	1.000	0.773	0.861	0.846	10
11–12	0.062.	0.800	0.100	0.113	0.117	11
12–13	0.047*	0.615	0.088.	0.003**	0.003**	12
13–14	0.000***	0.001***	0.000***	0.000***	0.000***	13
14–15	0.000***	0.000***	0.000***	0.000***	0.000***	14
15–16	0.000***	0.000***	0.000***	0.000***	0.000***	15
16–17	0.001***	0.006**	0.001**	0.000***	0.000***	16
17–18	0.000***	0.000***	0.000***	0.000***	0.000***	17
18–19	0.000***	0.000***	0.000***	0.000***	0.000***	18

Note: $\widehat{k}^{\text{age}} = 10.5$ years and the upper limit of one-sided 95% conf. int. is 11.64 years.

Conclusion

Summary:

1. Two-sample gradual change-point test can replace "multiple testing" by a single test concerning only the change-point location.
2. The change-point approach is applicable to any "table of estimates with standard errors".
3. The resulting p-values are decreasing and correspond to right-sided confidence intervals.
4. Left-sided confidence intervals may be useful for growth curve estimation.

Reference: Hlávka, Hušková (2017) Two-sample gradual change analysis, REVSTAT, 15(3), 355–372.

Week 6–7

Topic:

- Factorial experiments (experiments with at most one observation per subclass).
- Tukey's and Mandel's test of additivity.
- Multiplicative interactions.

Many factors

Often, one needs to investigate many factors using a limited number of observations. In such situations, we usually have **at most one observation** in each subclass (defined by a unique combination of factor levels).

MJ2 (p. 2) claim that such data set (with only one observation per subclass) can be obtained by mistake:

Example: Biologist investigates the influence of 20 combinations of temperature and humidity on sorghum growth. He can use 20 greenhouses allowing to set all possible combinations of temperature and humidity. 10 sorghum plants are grown in each greenhouse but the biologist does not realize that both treatments are applied on greenhouse level and that, disregarding the number of plants, we will always have only 20 observations (although higher number of plants should decrease residual variance).

	Humidity (%)			
Temperature (F)	20	40	60	80
50	12.3	19.6	25.7	30.4
60	13.7	16.9	27.0	31.5
70	17.8	20.0	26.3	35.9
80	12.1	17.4	36.9	43.4
90	6.9	18.8	35.0	53.0

Measurements of 200 sorghum plants lead to 20 observations (with one observation in each subclass). Moreover, the interaction of temperature and humidity is of interest and cannot be omitted.

An important consequence of having only one observation in each subclass is the nonexistence of "repetitions of the experiment in identical conditions" (that is usually needed to estimate the variance σ^2).

The biologist has following possibilities:

1. Reduce the number of factor levels (with six combinations of treatment levels, he could have three repetitions in each subclass).
2. Repeat the entire experiment several times (including new randomizations). This is possible only for fast growing plants.
3. Use more greenhouses (this may not be possible in practice).
4. Keep the original design and apply methods from MJ2: *[Milliken & Johnson (1989) Analysis of Messy Data, Vol. 2, Nonreplicated Experiments, Chapman & Hall/CRC, Boca Raton.]*

Two-way layout with one observation per subclass

The usual variance estimators cannot be used in the model

$$y_{ij} = \mu_{ij} + \varepsilon_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij} + \varepsilon_{ij},$$

because $\hat{\mu}_{ij} = y_{ij}$ and all residuals are equal to zero.

The simplest solution is to assume the the interaction term can be omitted, i.e., to use the model:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij},$$

leading directly the estimator
$\hat{\sigma}^2 = \sum\sum(y_{ij} - y_{i.} - y_{.j} + y_{..})^2/\{(I-1)(J-1)\}$.

Before removing the interaction term, we should test it (but, obviously, the usual F-test cannot be used).

Milliken & Rasmuson

Let us assume that:

$$y_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij} + \varepsilon_{ij}.$$

1. The observations are split into I groups according to levels of the factor T.
2. In each group, we calculate the sample variance $v_i = \sum_{j=1}^{J}(y_{ij} - y_{i.})/(J-1)$. This is, in the model with interaction, an unbiased estimator of $\sigma^2 + \sum_j(\beta_j - \beta_. + \gamma_{ij} - \gamma_{i.})^2/(J-1) = \delta_i^2$ (the point is that $\delta_1^2 = \cdots = \delta_I^2$ in the model without interaction).
3. Milliken and Rasmuson proposed to test the hypothesis $H_0 : \delta_1^2 = \cdots = \delta_I^2$ using standard tests of homogeneity of variances (Bartlett, Levene).

Milliken & Rasmuson

Weaknesses:

- assumptions for variance homogeneity tests are satisfied only approximately (sample variances v_i have noncentral χ^2 distribution even under the null hypothesis)
- with certain interaction patterns (e.g., latin squares), it may happen that all row and column variances are identical...

Tukey's nonadditivity test

Tukey's nonadditivity test (AKA Tukey's single df test) can be described as a test of the hypothesis $H_0 : \lambda = 0$ (against two-sided alternative) in the model $\mu_{ij} = \mu + \tau_i + \beta_j + \lambda\tau_i\beta_j$ assuming that the interaction γ_{ij} is proportional to the product of row and columns effects.

Tukey's test statistic is based on

$$SSN = \frac{\{\sum_i \sum_j (y_{i.} - y_{..})(y_{.j} - y_{..})(y_{ij} - y_{i.} - y_{.j} + y_{..})\}^2}{\sum_i (y_{i.} - y_{..})^2 \sum_j (y_{.j} - y_{..})^2}$$

and

$$SSR = \sum_i \sum_j (y_{ij} - y_{i.} - y_{.j} + y_{..})^2 - SSN.$$

The null hypothesis $H_0 : \lambda = 0$ is rejected if $SSN/\{SSR/(bt - b - t)\} > F_{1,bt-b-t}(1-\alpha)$.

Interaction plots

Two-way ANOVA with interactions:

$$y_{ij} = \mu_{ij} + \varepsilon_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij} + \varepsilon_{ij},$$

symbols τ_i and β_j denote row and column effects.

Type I interaction plot shows μ_{ij} against i (or j) for all j's (or i's),
Type II interaction plot shows μ_{ij} against τ_i (or β_j) for all j's (or i's).

These plots thus consist of J (or I) polygonal lines connecting I (or J) points.

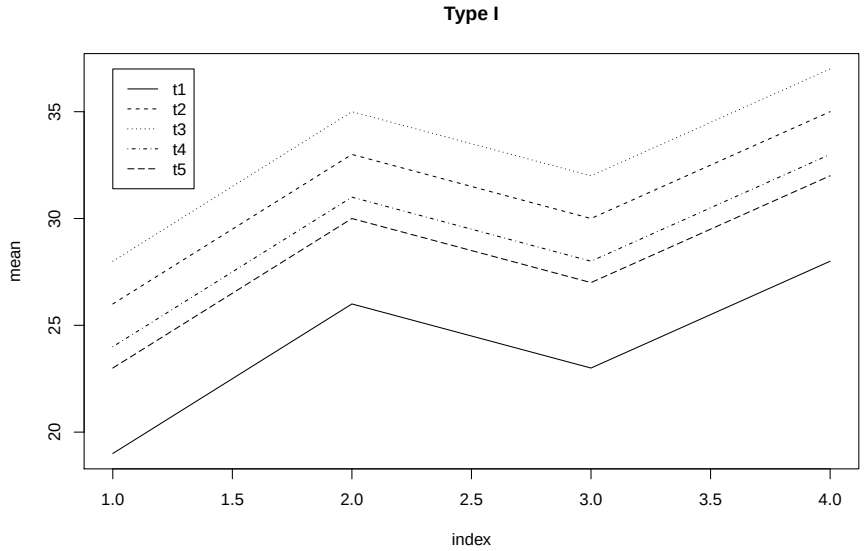
Example: Let's assume that the true expectations μ_{ij} satisfy the model $\mu_{ij} = \mu + \tau_i + \beta_j$ with $\mu = 29$, $\tau = (-5, 2, 4, 0, -1)^\top$, $\beta = (-5, 2, -1, 4)^\top$. The expectations (cell means) are:

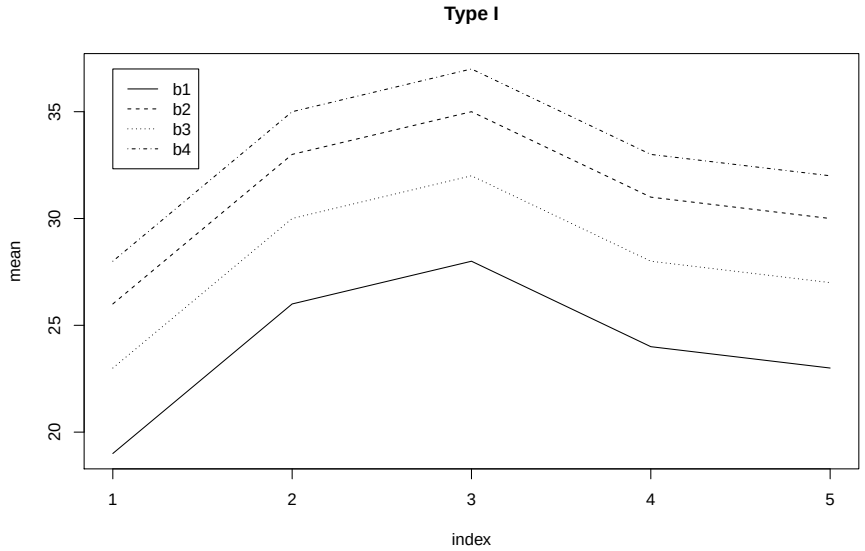
```
mu=29
tau=c(-5,2,4,0,-1)
beta=c(-5,2,-1,4)
cm.add=mu+outer(tau,beta,"+")
rownames(cm.add)=paste("t",1:5,sep="")
colnames(cm.add)=paste("b",1:4,sep="")
```

```
   b1 b2 b3 b4
t1 19 26 23 28
t2 26 33 30 35
t3 28 35 32 37
t4 24 31 28 33
t5 23 30 27 32
```


Type I interaction plot:

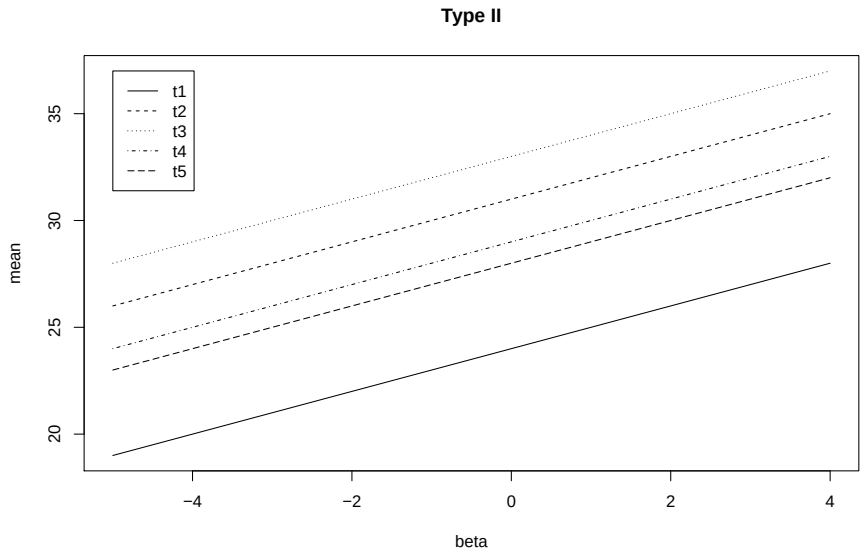
```
# type I - rows
plot(c(1,4),c(max(cm.add),min(cm.add)),type="n")
for (radek in 1:nrow(cm.add)) {
    lines(1:4,cm.add[radek,],lty=radek)
}
legend(1,max(cm.add),legend=row.names(cm.add),lty=1:5)

# type I - columns
plot(c(1,5),c(max(cm.add),min(cm.add)),type="n")
for (sloup in 1:ncol(cm.add)) {
    lines(1:5,cm.add[,sloup],lty=sloup)
}
legend(1,max(cm.add),legend=colnames(cm.add),lty=1:4)
```

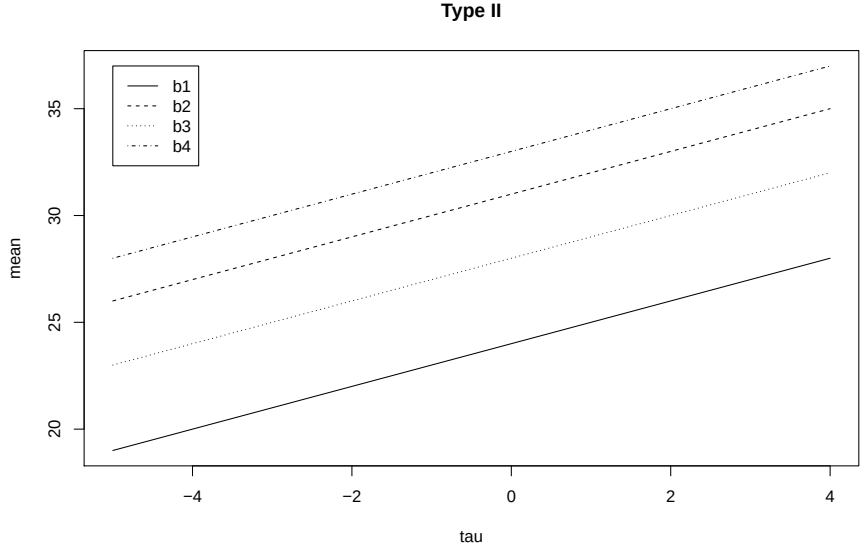





Type II interaction plot:

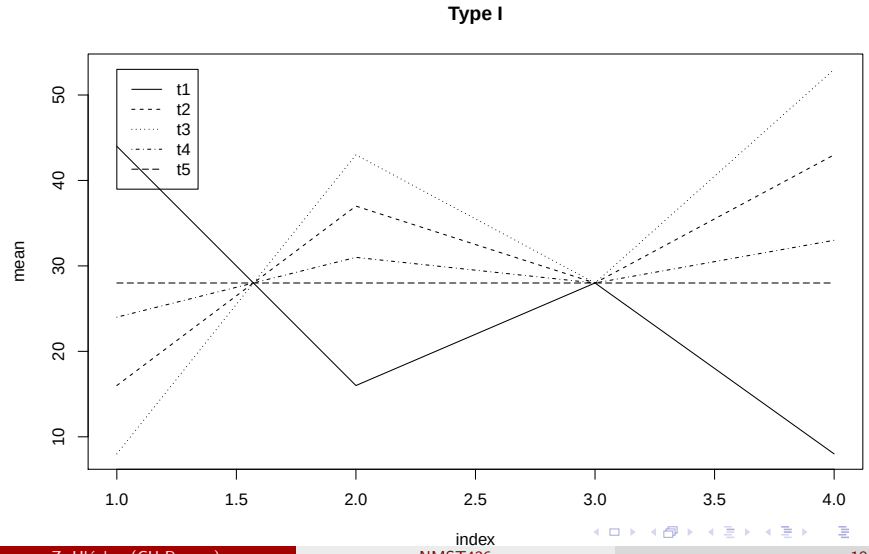
```
# type II - rows
plot(c(max(beta),min(beta)),c(max(cm.add),min(cm.add)),type="n")
ob=order(beta)
for (radek in 1:nrow(cm.add)) {
    lines(beta[ob],cm.add[radek,][ob],lty=radek)
}
legend(min(beta),max(cm.add),legend=row.names(cm.add),lty=1:5)

# type II - columns
plot(c(max(tau),min(tau)),c(max(cm.add),min(cm.add)),type="n")
ot=order(tau)
for (sloup in 1:ncol(cm.add)) {
    lines(tau[ot],cm.add[,sloup][ot],lty=sloup)
}
legend(min(tau),max(cm.add),legend=colnames(cm.add),lty=1:4)
```

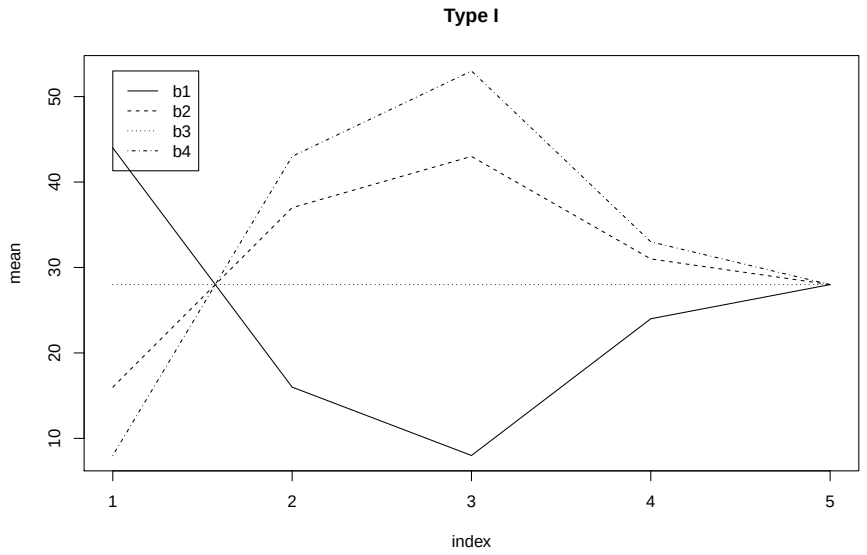



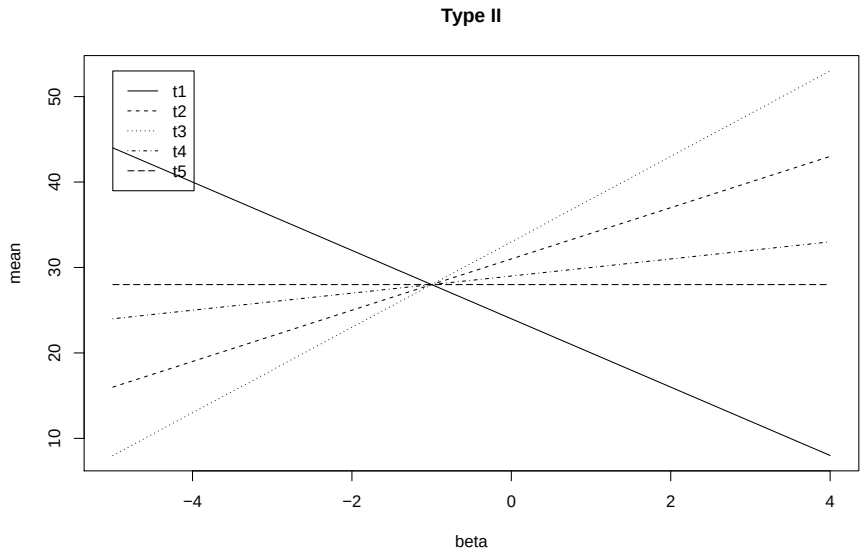


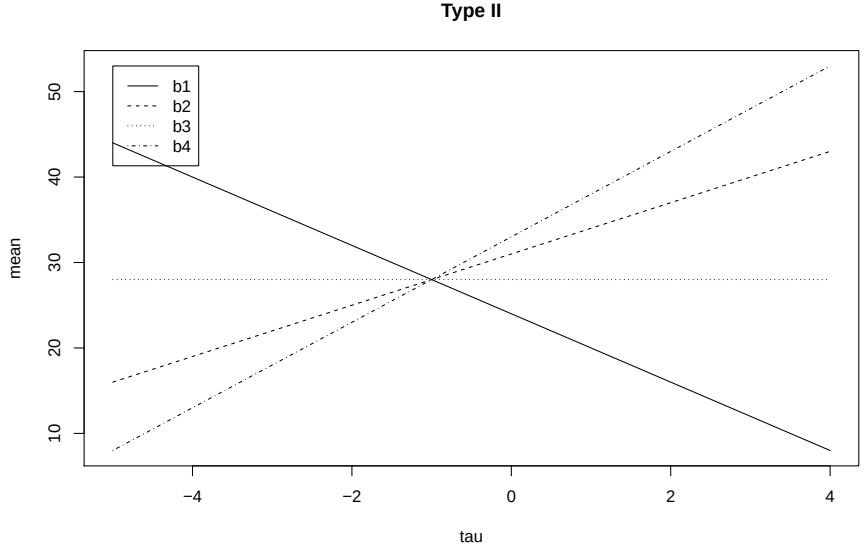
Interaction plots for Tukey's model look somewhat different:

`cm.tuk=cm.add+tau%*%t(beta)`









Mandel's test

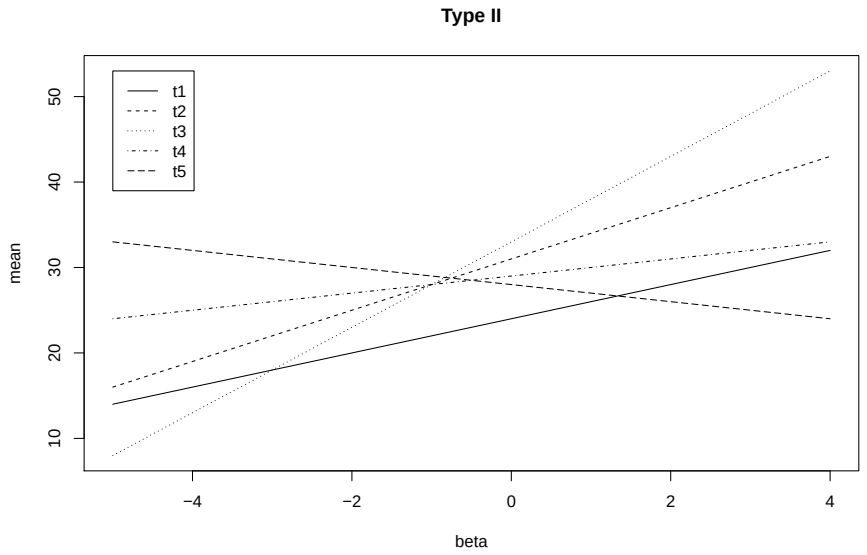
Mandel generalized Tukey's model as

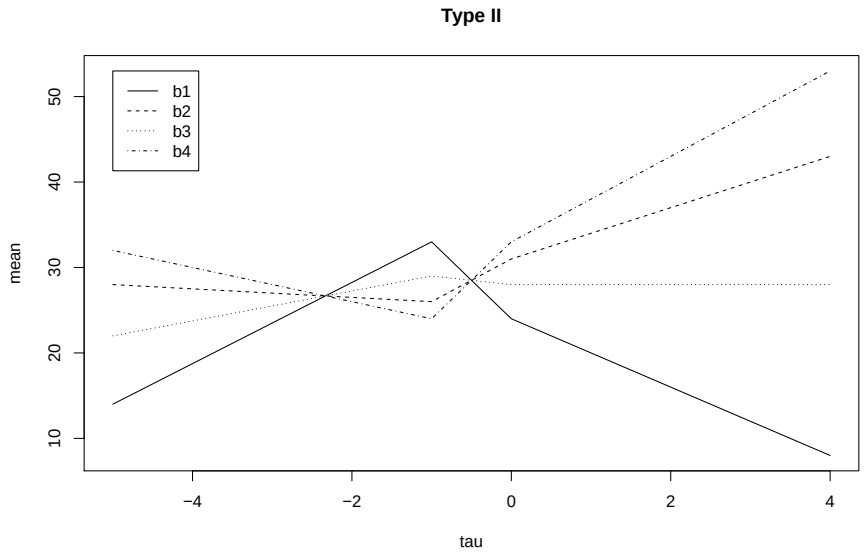
$$\mu_{ij} = \mu + \tau_i + \beta_j + \lambda\alpha_i\beta_j$$

with the null hypothesis of additivity formulated as $H_0 : \alpha_1 = \cdots = \alpha_I$.

Example: Interaction plots (type II) for $\alpha = (1, 2, 4, 0, -2)^\top$

One of two type II interaction plots looks like a "bundle of straight lines".





Some remarks

Tukey's and Mandel's test are implemented in standard statistical software. If not, we can perform these tests by first estimating the main effects and then testing the appropriate hypotheses concerning the interaction term.

Tukey's single df/Mandel's model can be further modified (for example, MJ2, p. 31 describe interaction terms defined by $\gamma_{ij} = \lambda\tau_i \exp(-\beta_j)$).

Main purpose of these tests is verification of absence of interactions (so that we can justify the model without interactions) in two-way ANOVA with one observation in each subclass.

Some remarks

This approach does not really solve anything in presence of significant interactions. In some situations, transformations may help (e.g., logarithmic transformation for Tukey's model) but MJ2 do not recommend transformations because:

- assumption of additive random errors is not transferred into the transformed model,
- interpretation is usually simpler on the original scale,
- transforms are unnecessary if we can model our data (e.g., GLM), interpretation follows from the applicable model.

Model with multiplicative interactions

Another natural extension of Tukey's

$$\mu_{ij} = \mu + \tau_i + \beta_j + \lambda\tau_i\beta_j$$

and Mandel's model

$$\mu_{ij} = \mu + \tau_i + \beta_j + \lambda\alpha_i\beta_j$$

is the model with multiplicative interactions

$$\mu_{ij} = \mu + \tau_i + \beta_j + \lambda\alpha_i\gamma_j,$$

where $0 = \sum \tau_i = \sum \beta_j = \sum \alpha_i = \sum \gamma_j$ and $1 = \sum \alpha_i^2 = \sum \gamma_j^2$.

Multiplicative interactions model

$$\mu_{ij} = \mu + \tau_i + \beta_j + \lambda\alpha_i\gamma_j,$$

does not assume that interactions depend on main effects.

Interaction of two rows (i and i') with arbitrary two columns (j and j') can be expressed as

$$\begin{aligned} & \mu_{ij} - \mu_{ij'} - \mu_{i'j} + \mu_{i'j'} \\ = \; & \ldots \\ = \; & \lambda(\alpha_i - \alpha_{i'})(\gamma_j - \gamma_{j'}) \end{aligned}$$

Therefore, $\alpha_i = \alpha_{i'}$ implies that there isn't any interaction of i-th and i'-th row with columns effects.

The multiplicative interactions model is usually more appropriate (than Tukey's or Mandel's model) if:

1. The model contains interactions although all row and column effects are equal to zero.
2. The interaction appears only for one combination of row and column effects. Such situation can appear for outlying observations or if some combnation of row and column treatments behaves *strange* (e.g., in presence of control group that does not allow to use the additive model)
3. The interactions appears only in one row or column (corresponding, e.g., to a control group that *behaves differently*).

LS estimators in the model

$$y_{ij} = \mu + \tau_i + \beta_j + \lambda\alpha_i\gamma_j + \varepsilon_{ij},$$

are:

$$\begin{aligned} \hat{\mu} &= y_{..} \\ \hat{\tau}_i &= y_{i.} - y_{..} \\ \hat{\beta}_j &= y_{.j} - y_{..} \end{aligned}$$

Let $Z = (z_{ij}) = (y_{ij} - y_{i.} - y_{.j} + y_{..})$ denote the matrix of residuals from the additive model. Then

$$\begin{aligned} \hat{\lambda}^2 &= \text{largest eivenvalue of } Z^\top Z \text{ (or } ZZ^\top) \\ \hat{\alpha} &= \text{eigenvector } ZZ^\top \text{ corresponding to } \hat{\lambda}^2 \\ \hat{\gamma} &= \text{eigenvector } Z^\top Z \text{ corresponding to } \hat{\lambda}^2 \end{aligned}$$

In other words: estimators of λ, α and β can be obtained by the SVD decomposition of the matrix Z.

Test of additivity

In the multiplicative interactions model

$$y_{ij} = \mu + \tau_i + \beta_j + \lambda\alpha_i\gamma_j + \varepsilon_{ij},$$

the null hypothesis $H_0 : \lambda = 0$ can be tested using tabulated critical values.

If $l_1 > \cdots > l_p$ are nonzero eigenvalues of $Z^\top Z$ (or $ZZ^\top$), the critical values for the test statistic

$$U_1 = l_1/RSS = l_1/\sum l_i$$

(where $RSS = \sum z_{ij}^2$ is the residual sum of squares from the additive model), can be found, e.g., in Table A.1 in MJ2 for selected values of $p = \min(I-1, J-1)$ and $n = \max(I-1, J-1)$.

Table A.1 Critical Points for $\ell_1/(\ell_1 + \ell_2 + \cdots + \ell_p)$

	p									
n	2	3	4	5	6	7	9	11	15	19
	UPPER 5%									
2	.999*									
3	.987*	.857								
4	.965*	.882*	.841							
5	.941*	.851	.759*	.668						
6	.917*	.801*	.682	.657*	.596					
7	.897*	.781	.683*	.599	.575*	.535				
9	.863*	.704	.632*	.590	.522*	.469	.414			
11	.836*	.694	.595*	.524	.484*	.450	.402	.367		
15	.795*	.630	.543*	.476	.434*	.399	.339	.306	.266	
19	.766*	.629	.509*	.454	.400*	.365	.314	.288	.241	.212
31	.713*	.535	.450*	.389	.344*	.311	.264	.234	.194	.170
49	.671*	.508	.407*	.348	.300	.271	.230	.202	.164	.142
99	.622*	.461	.358	.302	.260	.231	.189	.163	.131	.112

Further modifications can be found in MJ2

- Several multiplicative interactions (estimated by SVD).
- Choice of the number of multiplicative interactions (sequential testing of eigenvalues).
- Variance estimation (pseudo degrees of freedom $(I-1)(J-1) - \eta_1$, where η_1 is the expectation of l_1/σ^2 if $\lambda = 0$) in the model with one multiplicative interaction (or assuming that $H\alpha = 0$ for some matrix H).
- More involved analysis of interactions (equality of some coefficients implies non-existence of interactions in some rows or columns: testing the null hypothesis $H\alpha = 0$ [Marasinghe & Johnson (1981). Testing subhypotheses in the multiplicative interaction model. Technometrics, 23(4), 385-393.] — alternatively, Tukey's additivity test can be used on all row and column pairs.

Horses

Example: Let us recall the example with horses.

Assume that the experiment is designed in a way that allows estimation of all regression parameters in the linear model

$$y_{it} = \text{horse}_i + \text{time}_t + F_{it} * \text{leg}_{it} + \varepsilon_{it},$$

where $i = 1, \ldots, 4$ and $t = 1, 2$.

Q: Can we apply some additivity test on this data set?

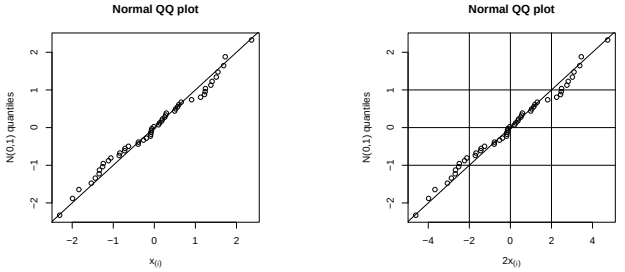
Week 7–8

Topic:

- Factorial experiments 2^N.
- Residual variance estimators:
  - without higher-order interactions,
  - half-normal plot.
- Blocking / slučování.
- Fractional factorial experiments / zkrácené faktoriální experimenty.
- Nonregular designs (Plackett-Burman).

Some repetition: QQ plot

In order to display goodness-of-fit graphically, we often plot sorted observations against theoretical quantiles.



Clearly, the (theoretical) slope for observations from $N(\mu, \sigma^2)$ in a QQ-plot with theoretical quantiles of $N(0,1)$ is $1/\sigma$.

Factorial experiment 2^N

In order to investigate as many factors as possible (including also their interactions) with limited number of observations, one usually considers only two levels for each factor. Typically, one works with only one observation in each subclass (cell).

The aim of a factorial experiment 2^N (or 3^N or mixed experiment $2^N 3^M$ etc.) is to identify factors that may have significant influence on the response variable.

The identified important factors are then investigated using further experiments (of different type).

Yates notation

The most simple situation is two-way ANOVA model (with interactions) and one observation in each subclass: $y_{ij} = \mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$, where $i, j \in \{0, 1\}$. *The effect of factor A* measures the influence of higher level compared to the lower level of factor A (i.e., we have $A = 2\alpha_1$ with the `contr.sum` contrasts satisfying $\sum \alpha_i = 0$).

The so-called *Yates notation* may seem unusual but it has its advantages: the lower level of each factor is denoted by 1, higher level is denoted by lower case letters $(a, b, \dots)$. Combinations of factor levels are then denoted by 'products' of these symbols.

In the factorial experiment 2^2, we obtain four possible factor level combinations: (1), a, b, ab.

The model: $y_{ij} = \mu_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij}$, where $i, j \in \{1, 2\}$

A	B	značení	stř. hodnota
lower	lower	(1)	$\mu_{--} = \mu_{00} - A/2 - B/2 + (AB)/2$
higher	lower	a	$\mu_{+-} = \mu_{00} + A/2 - B/2 - (AB)/2$
lower	higher	b	$\mu_{-+} = \mu_{00} - A/2 + B/2 - (AB)/2$
higher	higher	ab	$\mu_{++} = \mu_{00} + A/2 + B/2 + (AB)/2$

Estimators:

$A = \overline{y}_{A+} - \overline{y}_{A-} = \{-(1) + a - b + ab\}/2$

$B = \overline{y}_{B+} - \overline{y}_{B-} = \{-(1) - a + b + ab\}/2$

The effect A with factor B on higher level $(-b + ab)/2$, the effect A with B on lower level $(-(1) + a)/2$. The interaction AB is the difference of these two (conditional) effecs, i.e.,

$$AB = \{(1) - a - b + ab\}/2$$

Some simple algebra

$2A = -(1) + a - b + ab = (a-1) + b(a-1) = (a-1)(b+1)$

$2B = -(1) - a + b + ab = (b-1) + a(b-1) = (a+1)(b-1)$

$2AB = (1) - a - b + ab = b(a-1) - (a-1) = (a-1)(b-1)$

Similar rules hold also in higher order models. For example, the interaction $ABCD$ in a factorial experiment 2^6 can be estimated by the contrast:
$ABCD = (a-1)(b-1)(c-1)(d-1)(e+1)(f+1)/2^5$

The signs for estimating interactions can be also obtained by multiplying the signs of the corresponding main effects, see table [L68] on the following slide.

Tab. 37. Vyjádření hlavních efektů a interakcí v experimentech typu 2^2, 2^3 a 2^4

	(1)	a	b	ab	c	ac	bc	abc	d	ad	bd	abd	cd	acd	bcd	$abcd$
A	—	+	—	+	—	+	—	+	—	+	—	+	—	+	—	+
B	—	—	+	+	—	—	+	+	—	—	+	+	—	—	+	+
AB	+	—	—	+	+	—	—	+	+	—	—	+	+	—	—	+
C	—	—	—	—	+	+	+	+	—	—	—	—	+	+	+	+
AC	+	—	+	—	—	+	—	+	+	—	+	—	—	+	—	+
BC	+	+	—	—	—	—	+	+	+	+	—	—	—	—	+	+
ABC	—	+	+	—	+	—	—	+	—	+	+	—	+	—	—	+
D	—	—	—	—	—	—	—	—	+	+	+	+	+	+	+	+
AD	+	—	+	—	+	—	+	—	—	+	—	+	—	+	—	+
BD	+	+	—	—	+	+	—	—	—	—	+	+	—	—	+	+
ABD	—	+	+	—	—	+	+	—	+	—	—	+	+	—	—	+
CD	+	+	+	+	—	—	—	—	—	—	—	—	+	+	+	+
ACD	—	+	—	+	+	—	+	—	+	—	+	—	—	+	—	+
BCD	—	—	+	+	+	+	—	—	+	+	—	—	—	—	+	+
$ABCD$	+	—	—	+	—	+	+	—	—	+	+	—	+	—	—	+

Buns (housky) (MJ2, p. 98)

Example: A bakery has used factorial experiment 2^5 to investigate the influence of water (W), mixing time (M), temperature (T) and fat type (C) on the quality of buns. Two types of mixers (P) were used to prepare dough.

The resulting data set is given on next slide.

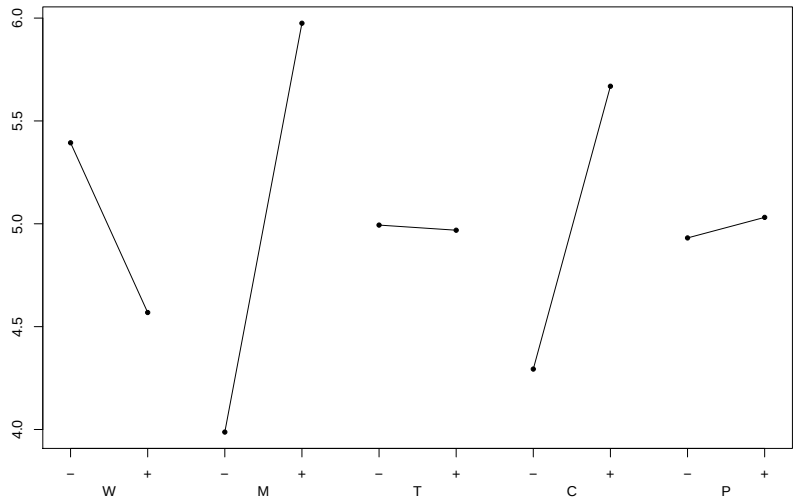
All effects can be easily obtained from `lm(Quality~(W+M+T+C+P)^5)` but one has to be careful about the choice (and meaning) of contrasts.

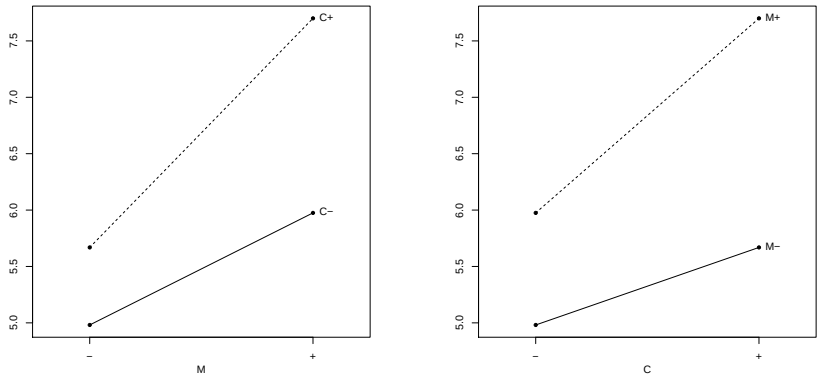
Unfortunately, the variance of ε_{ijklm} cannot be estimated from the above model. In this situation, we can choose from two possible approaches:

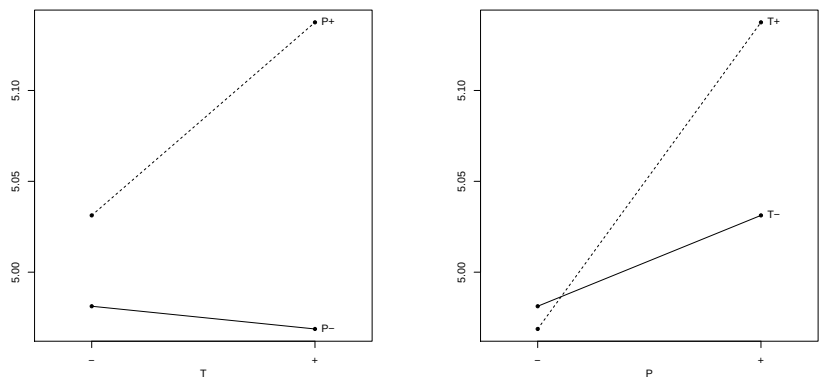
1. usual ANOVA with higher-order interactions,
2. half-normal plot (see MJ2).

Run	*W*	*M*	*T*	*C*	*P*	*Quality*
27	0	0	0	0	0	4.8
3	1	0	0	0	0	3.9
11	0	1	0	0	0	5.0
19	1	1	0	0	0	2.2
22	0	0	1	0	0	3.9
15	1	0	1	0	0	4.2
5	0	1	1	0	0	3.0
23	1	1	1	0	0	2.2
21	0	0	0	1	0	5.7
14	1	0	0	1	0	2.2
2	0	1	0	1	0	8.4
12	1	1	0	1	0	8.3
17	0	0	1	1	0	5.3
7	1	0	1	1	0	2.3
29	0	1	1	1	0	8.6
25	1	1	1	1	0	8.9
32	0	0	0	0	1	4.2
1	1	0	0	0	1	5.0
16	0	1	0	0	1	5.8
24	1	1	0	0	1	5.2
8	0	0	1	0	1	4.6
10	1	0	1	0	1	4.1
26	0	1	1	0	1	5.4
4	1	1	1	0	1	5.2
30	0	0	0	1	1	2.9
6	1	0	0	1	1	3.0
9	0	1	0	1	1	6.7
28	1	1	0	1	1	6.6
20	0	0	1	1	1	5.0
18	1	0	1	1	1	2.7
31	0	1	1	1	1	7.0
13	1	1	1	1	1	7.1


```
library(AlgDesign)
buns=gen.factorial(levels=2,nVars=5,varNames=c("W","M","T",
                                               "C","P"))
buns$Quality=c(4.8,3.9,5.0,2.2,3.9,4.2,3.0,2.2,5.7,2.2,8.4,
   8.3,5.3,2.3,8.6,8.9,4.2,5.0,5.8,5.2,4.6,4.1,5.4,5.2,2.9,
   3.0,6.7,6.6,5.0,2.7,7.0,7.1)
lm.buns3=lm(Quality~(W+M+T+C+P)^3,data=buns)
lm.buns5=lm(Quality~(W+M+T+C+P)^5,data=buns)
a3.1=aov(lm.buns3)
a3.2=anova(lm.buns3)
library(car)
a3.3=Anova(lm.buns3)
# Residual standard error: 0.6086187
a5.2=anova(lm.buns5)
ss5.2=a5.2$"Sum Sq"[1:31]
cf5.2=lm.buns5$coefficients[-1]
32*cf5.2^2
```





Half-normal plot / polonormální graf

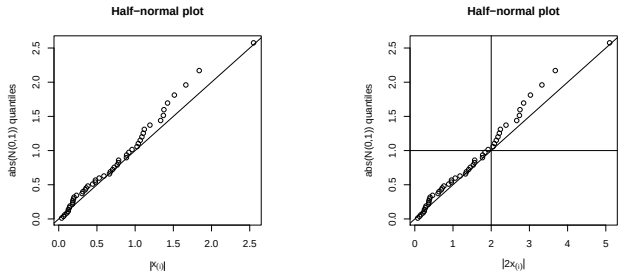
Sum of squares S_i corresponding to each factor are part of standard output in all statistical packages. In balanced experiments, these sums of squares are independent random variables with χ^2_1 distribution (multiplied by σ^2) under the null hypothesis.

Hence, random variables $\sqrt{S_i}$ have the same distribution as the *absolute value of* $N(0, \sigma^2)$.

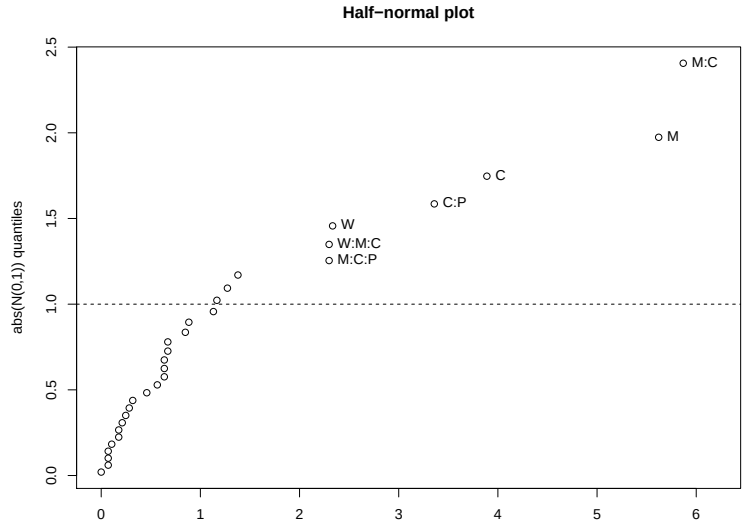
Half-normal plot is QQ plot for the verification of goodness-of-fit of the observed values $\sqrt{S_i}$ with the 'positive half of $N(0, 1)$'.

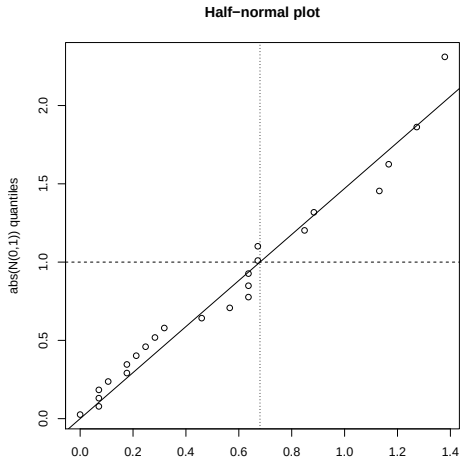
In principle, QQ plots could be constructed also for the estimated effects but these may not be defined uniquely and the interpretation would be more complicated.

Interpretation of the half-normal plot



Similarly as with the normal QQ plot, the slope depends on the parameter σ. Therefore, an estimator of σ can be obtained as the inverse value of the slope of a fitted regression line passing through the origin (in practice, we have to omit significant effects that may be identified as outliers).





Half-normal plot without significant effects, $\widehat{\sigma} \doteq 0.68$.

Significant effects are W, M, C, C:P, M:C, W:M:C, M:C:P. The main effect of mixer is not significant (P) but the model contains significant interactions of mixer with type of fat and mixing time (C:P a M:C:P).

MJ2 propose (as the most simple approach) to omit temperature and analyzed the data set separately for both mixer types.

```
anova(lm.p0<-lm(Quality~(W+M+C)^3,subset=(P<0),data=housky))

          Df Sum Sq Mean Sq F value   Pr(>F)
W          1  6.891   6.891 20.1554 0.002030 **
M          1 12.781  12.781 37.3839 0.000285 ***
C          1 26.266  26.266 76.8282 2.25e-05 ***
W:M        1  0.856   0.856  2.5027 0.152305
W:C        1  0.276   0.276  0.8062 0.395460
M:C        1 33.351  33.351 97.5521 9.31e-06 ***
W:M:C      1  5.881   5.881 17.2011 0.003220 **
Residuals  8  2.735   0.342
```



```
anova(lm.p1<-lm(Quality~(W+M+C)^3,subset=(P>0),data=housky))
          Df  Sum Sq Mean Sq F value    Pr(>F)
W          1  0.4556  0.4556  1.2211  0.301267
M          1 19.1406 19.1406 51.2982 9.593e-05 ***
C          1  0.1406  0.1406  0.3769  0.556326
W:M        1  0.0756  0.0756  0.2027  0.664516
W:C        1  0.1806  0.1806  0.4841  0.506282
M:C        1  6.3756  6.3756 17.0871  0.003282 **
W:M:C      1  0.6806  0.6806  1.8241  0.213782
Residuals  8  2.9850  0.3731

anova(lm.p1mc<-lm(Quality~(M+C)^3,subset=(P>0),data=housky))
          Df  Sum Sq Mean Sq F value    Pr(>F)
M          1 19.1406 19.1406 52.4700 1.024e-05 ***
C          1  0.1406  0.1406  0.3855  0.546290
M:C        1  6.3756  6.3756 17.4774  0.001275 **
Residuals 12  4.3775  0.3648
```



```
nd0=gen.factorial(levels=2,nVars=3,varNames=c("W","M","C"))
data.frame(nd0,Estimate=predict(lm.p0,newdata=nd0))

   W  M  C Estimate
1 -1 -1 -1     4.35
2  1 -1 -1     4.05
3 -1  1 -1     4.00
4  1  1 -1     2.20
5 -1 -1  1     5.50
6  1 -1  1     2.25
7 -1  1  1     8.50
8  1  1  1     8.60
```

Water interactions are interesting: more water decreases the quality of dough if the mixing time is lower and fat higher or if the mixing time higher and fat lower.

The second mixer type leads to worse results.

```
nd1=gen.factorial(levels=2,nVars=2,varNames=c("M","C"))
data.frame(nd1,Estimate=predict(lm.p1mc,newdata=nd1))
```

```
   M  C Estimate
1 -1 -1    4.475
2  1 -1    5.400
3 -1  1    3.400
4  1  1    6.850
```

Best dough is obtained with the first mixer type, longer mixing time, and with better type of fat. In this circumstances, the amount of water does not seem to be very important (although the quality seems to be slightly better for higher level of the factor water).

Practical recommendations

MJ2 recommend to use half-normal plot for small number of factors (where omitting some important interaction could lead to information loss).

For higher number of factors (> 6), one can usually neglect interactions from 5th order and use standard ANOVA methodology.

In practice, one can use the model without higher order interactions also for smaller number of factors — half-normal plot can be then used to verify the assumption that the omitted interactions do not have any influence on the response.

More formal statistical analysis of the half-normal plot can be based on Lenth's method based on a robust variance estimator (based on a suitable product of the median of the absolute values, i.e., the so-called pseudo standard error) and tabulated critical values for the resulting pseudo t-statistics.

Some general principles (Wu & Hamada)

Basic principles for factorial experiments according to [Wu & Hamada (2011). Experiments: planning, analysis, and optimization. Wiley]:

Effect hierarchy principle One can expect that lower order effects and interactions will be important more often than higher order interactions.

Effect sparsity principle The number of really important effects in factorial experiments usually tends to be small.

Effect heredity principle For each significant interaction, at least one of the corresponding main effects should be also significant.

Example: horses II

Example: Veterinary hospital compares two treatments of joint fracture in horses. In this experiment, one takes several horses and breaks their joints in a precisely defined way. These injuries are treated by one of the two treatments (factor A) and with one of two antiseptics (factor B). Suitable measure of recovery is evaluated after two months. Apart of the treatment effect, the hospital wants to establish whether the treatment effect is the same for front and hind legs (factor C).

Careful planning is crucial because horses are expensive and there is only a limited supply.

In this case, we investigate effects of three factors on two levels (factorial experiment 2^3). Interestingly, two horses have altogether $2^3 = 8$ feet. Unfortunately, we do not have 8 independent repetitions, because horse is an additional blocking factor (say D).

Example: horses II

Considering also the horse effect, we obtain factorial experiment 2^4 that cannot be carried out because of the limited number of horse feet.

Altogher, we will have only eight observations. The final design can be described as a factorial experiment 2^3 with one nuisance (blocking) factor.

The so-called *blocking in factorial experiments* technique allows to estimate the possibly interesting effects and interactions in the presence of nuisance factors (blocks).

Blocking / slučování

Factorial experiments often need many observations. If constant conditions cannot be guaranteed for the entire experiment, the measurements have to be split into homogenenous blocks.

The assignment of observations into blocks can be based on confounding the blocks with some higher order interaction (looking at the table with "expressions for main effects and interactions", we find the row with the chosen interaction and use it to split the observations into two blocks).

In a factorial experiment 2^3 with blocking defined by 3rd order interaction ABC, we split the experiment to $\{(1), ab, ac, bc\}$ and $\{a, b, c, abc\}$.

The same can be done with higher number of blocks but one cannot estimate also the so-called *generalized interactions* of the interactions that has been used to define the blocks.

Tab. 37. Vyjádření hlavních efektů a interakcí v experimentech typu 2^2, 2^3 a 2^4

	(1)	a	b	ab	c	ac	bc	abc	d	ad	bd	abd	cd	acd	bcd	$abcd$
A	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+
B	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+
AB	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+
C	−	−	−	−	+	+	+	+	−	−	−	−	+	+	+	+
AC	+	−	+	−	−	+	−	+	+	−	+	−	−	+	−	+
BC	+	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+
ABC	−	+	+	−	+	−	−	+	−	+	+	−	+	−	−	+
D	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+
AD	+	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+
BD	+	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+
ABD	−	+	+	−	−	+	+	−	+	−	−	+	+	−	−	+
CD	+	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+
ACD	−	+	−	+	+	−	+	−	+	−	+	−	−	+	−	+
BCD	−	−	+	+	+	+	−	−	+	+	−	−	−	−	+	+
$ABCD$	+	−	−	+	−	+	+	−	−	+	+	−	+	−	−	+

Fractional replications / zkrácené faktoriální experimenty

Fractioning (krácení) is similar to blocking but its purpose is to decrease the number of observations.

The method is based on collecting only observations defined by having the same sign in a row for chosen higher order interaction. The chosen interaction cannot be estimated (just like in blocking) but we cannot directly estimate also some other effects.

Example: Fractioning the experiment 2^3 by the interaction ABC, we can investigate only four treatment level combinations: (1), ab, ac, bc (this is fractional factorial experiment 2^{3-1}). Consequently, some effects cannot be distinguished from the corresponding generalized interactions (between the effect and the interaction that has been chosen for fractioning). The effect A is thus confounded with the interaction BC, the effect B with the interaction AC and the effect C with the interaction AB.

Exercise

Re-analyze the buns example after fractioning to one half (e.g., using the interaction WMTCP).

You may proceed as follows:

1. Determine the signs for estimating the fifth order interaction (this can be done by looking at the table).
2. Choose one half of the data set and use the function `alias()` to find out what can be estimated.
3. Plot a half-normal plot for sums of squares.
4. Find a reasonable ANOVA model (i.e., a model without higher order interactions that can be neglected).

Try to fraction the same experiment to one fourth. What can you estimate from only eight observations?

Plackett-Burman

An advantage of factorial experiments (and fractional factorial experiments) are uncorrelated estimators. (In a so-called *regular design*, all estimators have correlation either 0 or 1.)

However, factorial experiments 2^N may not be usable in practice because sample size has to be equal to a power of 2 (i.e., 8, 16, 32, 64, 128, ...).

(Nonregular) *Plackett-Burman designs* were proposed for sample sizes equal to multiples of 4 (that are not power of 2) and are suitable mainly for estimation of main effects (estimation of interactions is more complicated due to correlations).

Plackett-Burman designs are also (in some way) balanced.

Orthogonal arrays

Orthogonal array $OA(N, s_1^{m_1} \dots s_\gamma^{m_\gamma}, t)$ of power t is a matrix $N \times m$, where $m = m_1 + \dots + m_\gamma$ such that m_i columns contain $s_i \geq 2$ different symbols or levels so that all combinations of symbols occur with the same frequency in each t columns.

Orthogonal arrays of power 2 are most popular (and their power is usually omitted).

Having m two-level factors and N repetitions, the experiment can be planned as an orthogonal array $OA(N, 2^m)$ (clearly, the number of observations N has to be multiple of 4 because all possible combinations of two symbols have to be repeated with the same frequency in every two columns).

It can be shown that $OA(N, 2^{N-1})$ is equivalent to the so-called *Hadamard matrix* ($N \times N$ orthogonal matrix containing $+1$ a -1).

Plackett-Burman

For some sample sizes, Plackett-Burman designs can be produced by sequential shifts of generating vectors (WH, p. 376) by rows or columns:

Table 8.5 Generating Row Vectors for Plackett-Bruman Designs of Run Size N

N	Vector
12	+ + − + + + − − − + −
20	+ + − − + + + + − + − + − − − − + + −
24	+ + + + + − + − + + − − + + − − + − + − − − −
36	− + − + + + − − − + + + + + − + + − − + − − − − + − + − + + − − + −
44	+ + − − + − + − − + + + − + + + + − − − + − + + + − − − − − + − − − + + − + − + + −

12 observations for 11 factors (WH)

Example:

Table 8.2 Design Matrix and Lifetime Data, Cast Fatigue Experiment

	Factor											Logged
Run	*A*	*B*	*C*	*D*	*E*	*F*	*G*	8	9	10	11	Lifetime
1	+	+	−	+	+	+	−	−	−	+	−	6.058
2	+	−	+	+	+	−	−	−	+	−	+	4.733
3	−	+	+	+	−	−	−	+	−	+	+	4.625
4	+	+	+	−	−	−	+	−	+	+	−	5.899
5	+	+	−	−	−	+	−	+	+	−	+	7.000
6	+	−	−	−	+	−	+	+	−	+	+	5.752
7	−	−	−	+	−	+	+	−	+	+	+	5.682
8	−	−	+	−	+	+	−	+	+	+	−	6.607
9	−	+	−	+	+	−	+	+	+	−	−	5.818
10	+	−	+	+	−	+	+	+	−	−	−	5.917
11	−	+	+	−	+	+	+	−	−	−	+	5.863
12	−	−	−	−	−	−	−	−	−	−	−	4.809

Exercise

Re-analyze the buns example using 12 observations chosen according to Plackett-Burman design.

Try to answer these questions:

1. Is it possible to estimate all main effects?
2. Calculate the correlation matrix of the estimators and compare it to the fractional factorial experiment.

Summary

Advantages of nonregular design:

1. we can use less observations,
2. more flexibility for the choice of number of observations (especially compared to factorial experiments 3^N),
3. more flexibility for mixed designs (e.g., for mixed two- and three-level factors).

An important property of nonregular designs are correlated estimators. Therefore, both the evaluation of the proposed design and data analyses are more involved. On the other way, it may be possible to estimate some effects that are not estimable in a fractional factorial experiment.

WH (p. 427–433) recommend to find appropriate mode using stepwise regression, information criteria or Bayesian approach.

Week 8–9

Topic:

- Response surface methodology.
- Taguchi approach.

References:

Wu & Hamada (2011). *Experiments: planning, analysis, and optimization*. Wiley.

Myers, Montgomery (2002) *Response surface methodology*, Wiley.

Display of a factorial experiment 2^3

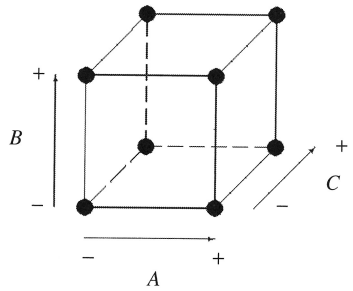


Figure 4.2. Cuboidal representation of a 2^3 design.

Response surface / výpovědní plocha

Fractional factorial experiments 2^{N-k} were originally developed in industrial applications and later applied also in agriculture.

"Response surface" methodology was developed in chemical industry (Imperial Chemical Industries) and its aim is to find precise values of important factors maximizing, e.g., amount of a product obtained by chemical reaction.

The usual sequential algorithm consists of two steps:

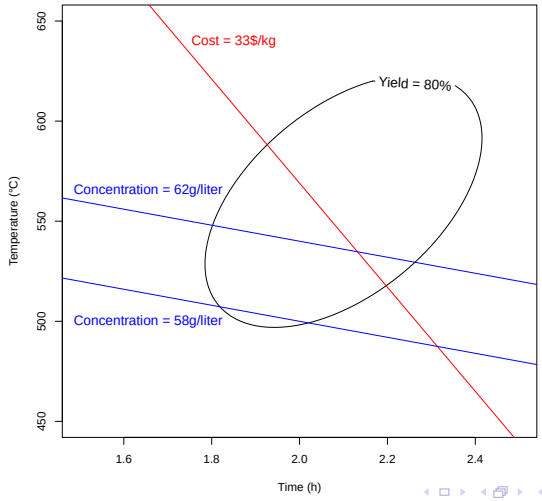
1. try to 'move close' to optimal values (if we are 'far away'),
2. determine the optimal values as precisely as possible (if we are 'close').

Applications

Apart of searching for the optimal values of some factors, RSM can be used to:

1. Construction of models for the expectation of a response in some region: this allows to predict changes that will happen in case of some (forced) change in the production process' settings.
2. Searching for input conditions guaranteeing certain output parameters: using models for more responses, results can be summarized in a single graph (see the graph with contours for three response variables on the following slide).

Using models for concentration andyield, it is easy to plot the region where Yield>80%, Cost<33, and Concentration$\in (58, 62)$.



Response surface methodology (RSM)

Let us assume that a response is an unknown function of input factors:

$$Y = f(X_1, \ldots, X_p) + \varepsilon$$

In practice, explanatory variables are standardized so that their values are $-\alpha$, -1, 0, 1, α (note that only values -1 a $+1$ are used in factorial experiments).

The dependency of (expected value of) Y on input factors can be plotted as a surface in $\mathbb{R}^p$—this suggests the name *response surface methodology* (RSM).

In practice, polynomial models for the entire response surface are not usable. On the other hand, local approximations using first or second order polynomials usually work very well.

⋆ Farther from the optimal values, one uses linear approximation:

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \varepsilon$$

and we can apply a fractional factorial experiment 2^{N-k}.

⋆ Close to the optimal values, one applies quadratic approximation:

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \sum_{i<j}^{p} \beta_{ij} x_i x_j + \sum_{i=1}^{p} \beta_{ii} x_i^2 + \varepsilon,$$

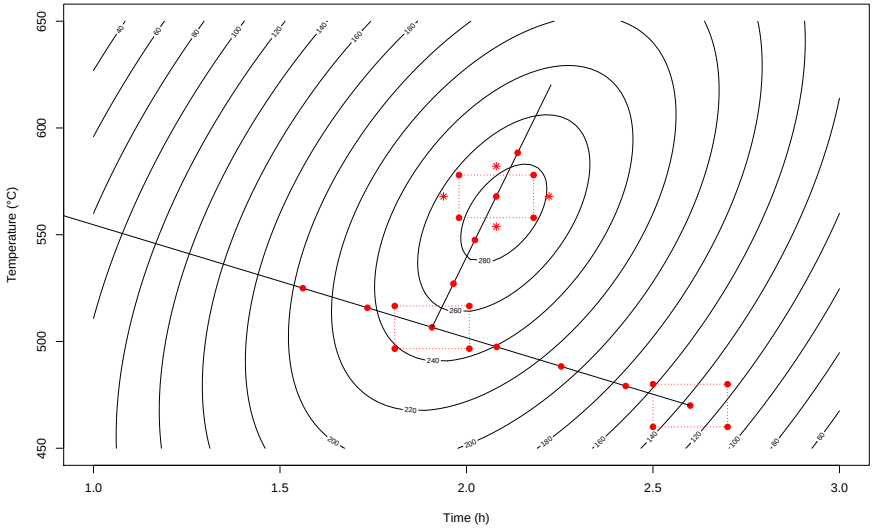
allowing the estimation of input factors' optimal values (in this case, fractional factorial experiments do not suffice).

Typical experimenting process

Wu, C. J., & Hamada, M. S. (2011). *Experiments: planning, analysis, and optimization*. John Wiley & Sons.

Typical experimental process in practice:

1. screening: choice of possibly important factors (typically fractional factorial experiment, Plackett-Burman design, etc.)
2. chosen (possibly important) factors are investigated in more detail:
   1. first, we have to verify whether the used values of chosen factors are 'close' to optimal values (e.g., factorial experiment 2^N with additional central point),
   2. in the neighborhood of optimal values, we use quadratic model (with central composite design) as an approximation of the 'response surface'.



Tests of nonlinearity

Typically, we already have some data from a factorial experiment (that have been collected when we were trying to identify important factors).

Let's assume that we have n_f observations from the factorial experiment and that we additionally collect n_c observations in the center (point 0): $\bar{Y}_c$ denotes the sample mean of central values and $\bar{Y}_f$ the sample mean of non-central values.

Under the quadratic model, we have $E\bar{Y}_c = \beta_0$ a $E\bar{Y}_f = \beta_0 + \sum \beta_{ii}$. The difference of sample means $\bar{Y}_f - \bar{Y}_c$ estimates $\sum \beta_{ii}$ and linearity can be tested using the test statistic:

$$\frac{\bar{Y}_f - \bar{Y}_c}{S_c\sqrt{n_f^{-1} + n_c^{-1}}} \overset{H_0}{\sim} t_{n_c-1},$$

where S_c^2 is the sample variance calculated from n_c central measurements.

Example: Wu & Hamada illustrate this method with an example, aimed at maximizing the amoung of a product obtained from a chemical reaction. The factors are time and temperature. The estimators obtained from a factorial experiment 2^2 and 2 central (the regression model $EY = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + (\beta_{11} + \beta_{22}) x_1^2$) are:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   135.875      2.571  52.851    0.012 *
x             -28.191      1.818 -15.508    0.041 *
y              11.845      1.818   6.516    0.097 .
I(x^2 + y^2)   -2.421      1.574  -1.538    0.367
x:y            -3.812      1.818  -2.097    0.283

Residual standard error: 3.636 on 1 degrees of freedom
Multiple R-squared:  0.9966,Adjusted R-squared:  0.9828
F-statistic: 72.43 on 4 and 1 DF,  p-value: 0.08788
```

Hence, linearity is not rejected.

Steepest ascent search

If we would reject linearity, we could use quadratic regression model and estimate the optimal values of input factors (this will be discussed few minutes later).

If linearity is not rejected, we can estimate the direction in which we can find *better* values of the input factors.

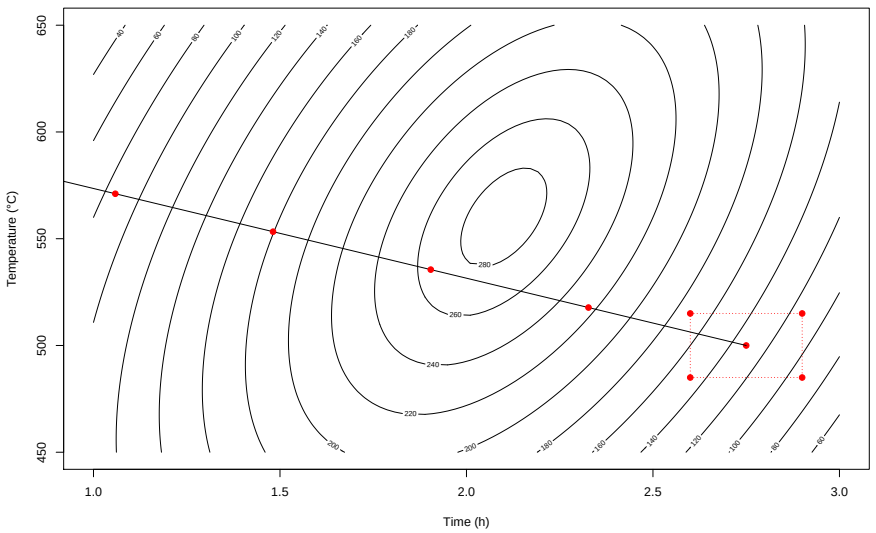
Assuming the linear model $Y = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \varepsilon = \beta_0 + x^\top \beta$, the steepest ascent direction is the gradient $\partial(\beta_0 + x^\top \beta)/\partial x = \beta$).

Hint: $x^* = \arg\max_{x \in \mathbb{R}^p} x^\top \beta / (\|x\| \|\beta\|) = \ldots ?$

Instead of "steepest ascent", one can also use "rectangular grid search", where the initial values of factors are chosen wide apart and the resolution is sequentially refined in further steps (similarly as in the half-interval search).

In the chemical product example, we have $\hat{\beta}_1 = -28.191$ a $\hat{\beta}_2 = 11.845$ and the steepest ascent direction is, e.g., $(-1, -\hat{\beta}_2/\hat{\beta}_1)^\top \doteq (-1, 0.420)^\top$

```
       time     temp          x         y     yield
1  2.600000 485.0000  -1.00000 -1.000000 143.56824
2  2.600000 515.0000  -1.00000  1.000000 174.88111
3  2.900000 485.0000   1.00000 -1.000000  94.80920
4  2.900000 515.0000   1.00000  1.000000 110.87455
5  2.750000 500.0000   0.00000  0.000000 133.30422
6  2.750000 500.0000   0.00000  0.000000 138.44603
7  2.327129 517.7668  -2.81914  1.184456 225.59176
8  1.904258 535.5337  -5.63828  2.368911 266.83921
9  1.481387 553.3005  -8.45742  3.553367 181.19676
10 1.058516 571.0673 -11.27656  4.737822  92.19063
```



Second order model / kvadratický model

Near the optimal values, one usually uses a second order model:

$$\hat{Y}(x) = \hat{\beta}_0 + \sum_{i=1}^{p} \hat{\beta}_i x_i + \sum_{i<j}^{p} \hat{\beta}_{ij} x_i x_j + \sum_{i=1}^{p} \hat{\beta}_{ii} x_i^2 = \hat{\beta}_0 + \hat{\beta}^\top x + x^\top \hat{\boldsymbol{\beta}} x,$$

where

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/2 & \dots & \hat{\beta}_{1p}/2 \\ \hat{\beta}_{12}/2 & \hat{\beta}_{22} & \dots & \hat{\beta}_{2p}/2 \\ \vdots & \vdots & & \vdots \\ \hat{\beta}_{1p}/2 & \hat{\beta}_{2p}/2 & \dots & \hat{\beta}_{pp} \end{pmatrix}$$

The parameter estimates are typically obtained from a central composite design (centrální složený experiment).

Central composite design / centrální složený experiment

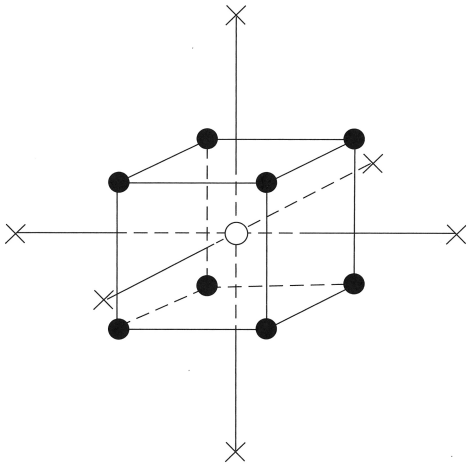


Figure 10.1. A central composite design in three dimensions [cube point (dot), star point (cross), center point (circle)].

Central composite design

Central composite design consists of:

- n_f cube (or corner) points with $x_i = -1, 1$ for $i = 1, \dots, p$. These points correspond to a (fractional) factorial experiment 2^p.
- n_c central points ($x_i = 0$ for $i = 1, \dots, p$).
- $2p$ star (or axial) points defined by $(0, \dots, x_i, \dots, 0)^\top$, where $x_i = -\alpha, \alpha$ for $i = 1, \dots, p$.

The measurements can be collected at the same time or we can add central and axial points sequentially (after rejecting linearity).

Practical remarks (from WH)

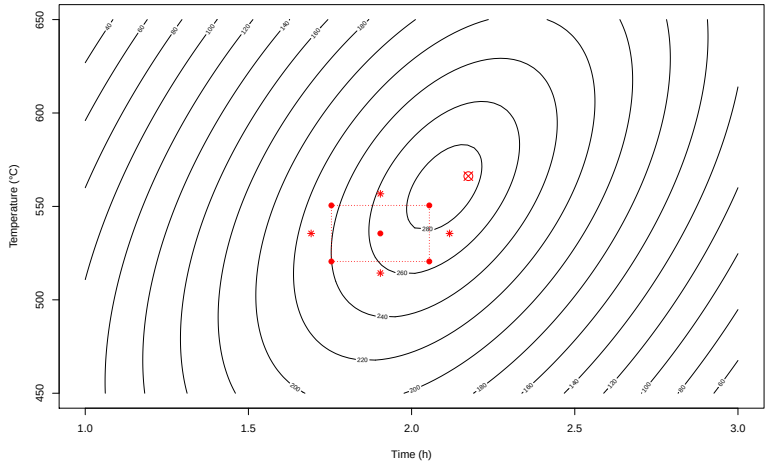
For central composite design, we have to:

- Choose the "factorial" part of the experiment (the number of distinct design points $n_f + 1 + 2p$ should be greater than the number of parameters $(p+1)(p+2)/2$. More detailed discussion and recommendations for various dimensions are given in WH (Section 10.7, p. 485).
- Choose α for axial points (usually between 1 and $\sqrt{p}$). Value 1 leads to the "face center design" (with axial points on the surface of unit cube), $\sqrt{p}$ leads to a spherical experiment (with axial points on unit sphere).
- Choose the number of central points n_c: WH recommend 3–5 for α close to $\sqrt{p}$, 1–2 for α close to 1, and 2–4 for α somewhere in between.

Box-Hunter wire model of central composite design

[Barton R. R. (2012). Graphical methods for the design of experiments, Lecture Notes in Statistics 143, Springer, page 167]

Example: Central composite design for the amount of product obtained by a chemical reaction.



	time	temp	x	y	yield
1	1.754258	520.5337	-1.000000	-1.000000	232.1077
2	1.754258	550.5337	-1.000000	1.000000	228.8667
3	2.054258	520.5337	1.000000	-1.000000	262.8238
4	2.054258	550.5337	1.000000	1.000000	287.7537
5	1.904258	535.5337	0.000000	0.000000	268.1906
6	1.904258	535.5337	0.000000	0.000000	269.6552
7	1.904258	535.5337	0.000000	0.000000	267.8820
8	1.904258	514.3205	0.000000	-1.414214	256.7880
9	1.904258	556.7469	0.000000	1.414214	267.4575
10	1.692126	535.5337	-1.414214	0.000000	227.5922
11	2.116390	535.5337	1.414214	0.000000	275.6034

These values suggest that this is a spherical central composite design.

We have already noted that, close to the optimal values, we usually use quadratic regression:

$$\hat{Y}(x) = \hat{\beta}_0 + \hat{\beta}^\top x + x^\top \hat{\boldsymbol{\beta}} x.$$

The maximum of the function $\hat{Y}(x)$ can be found by differentiating:

$$\frac{\partial \hat{Y}(x)}{\partial x} = \hat{\beta} + 2\hat{\boldsymbol{\beta}} x = 0,$$

defining the so-callled *stationary point*:

$$x_s = -\frac{1}{2}\hat{\boldsymbol{\beta}}^{-1}\hat{\beta}.$$

The character of the fitted quadratic surface (in the neighborhood of the stationary point x_s) can be desribed using an alternative system of coordinates.

Let the columns of the matrix P be the standardized eigenvectors of $\hat{\boldsymbol{\beta}}$. Then

$$P^\top \hat{\boldsymbol{\beta}} P = \boldsymbol{\Lambda},$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues.

In a new coordinate system (centered in x_s with rotation determined by columns of the matrix P, i.e., $v = P^\top(x - x_s) = P^\top z$) we have

$$\hat{Y} = \cdots = \hat{Y}_s + \sum \lambda_i v_i^2.$$

The type of the quadratic surface in the neighborhood of x_s can be classified according to the eigenvalues.

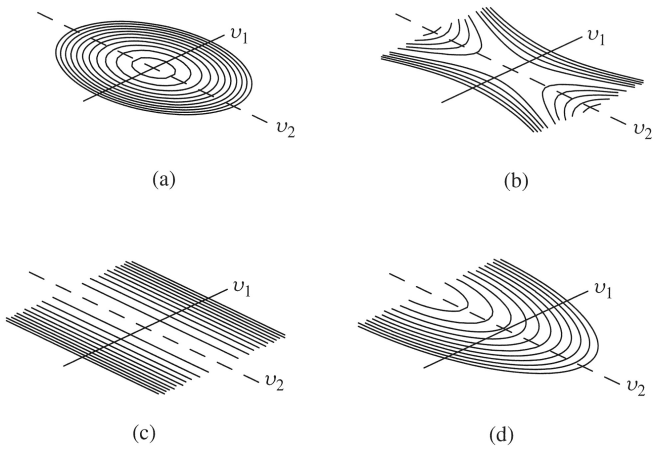


Figure 10.7. Classification of second-order responses surfaces: (a) Elliptic, (b) hyperbolic, (c) stationary ridge, (d) rising ridge.

Example: Investigation of the stationary point.

	time	temp	x	y	yield
1	1.754258	520.5337	-1.000000	-1.000000	232.1077
2	1.754258	550.5337	-1.000000	1.000000	228.8667
3	2.054258	520.5337	1.000000	-1.000000	262.8238
4	2.054258	550.5337	1.000000	1.000000	287.7537
5	1.904258	535.5337	0.000000	0.000000	268.1906
6	1.904258	535.5337	0.000000	0.000000	269.6552
7	1.904258	535.5337	0.000000	0.000000	267.8820
8	1.904258	514.3205	0.000000	-1.414214	256.7880
9	1.904258	556.7469	0.000000	1.414214	267.4575
10	1.692126	535.5337	-1.414214	0.000000	227.5922
11	2.116390	535.5337	1.414214	0.000000	275.6034

```
> Beta

          [,1]     [,2]
[1,] -9.482132  3.52136
[2,]  3.521360 -4.21964

> eigen(Beta)

$values
[1]  -2.455044 -11.246728

$vectors
           [,1]       [,2]
[1,] -0.4480089 -0.8940291
[2,] -0.8940291  0.4480089
```


Notes

Other possible designs (apart of the clearly most popular central composite design) are:

Box-Behnken design: e.g., for three factors with three levels, it leads to observations in central point and "edge centers".

Equiradial design: regular polygon (pentagram, hexagon, heptagon, ...) with central point.

Small composite experiment: based on a fractional factorial experiment.

Koshal design: investigates factors *by one* (or by pairs if we want to estimate some interaction).

Hybrid design: modification of a central composite design satisfying certain moment (or other) conditions.

Design comparison

In practice, we should try to choose a good design that will provide reliable results under various circumstances. Designs, that are optimal only under ideal conditions (that may not be satisfied in reality), can not be recommended.

Central composite or Box-Behnken designs are recommended for RSM (quadratic model). Hybrid design can be used if we have to reduce the number of observations.

In special cases that do not allow application of standard approach, we can generate the experiment design numerically (using computer). In this situation, alphabetic optimality criteria (D, A, G, ...) can be used to assess the quality of the generated design.

Computer algorithms

Using computer, it is not difficult to compare quality of standard designs in given situation.

In some situations, computer can be used to design the experiment:

1. Restriction on values of explanatory variables (standard models assume cuboidal or spherical regions).
2. Nonlinear or high order polynomial model.
3. Limited number of observations: typically, computer can choose the best little composite or hybrid design.
4. Restrictions on block size that do not allow application of standard designs.

Taguchi design (robust parameter design)

Taguchi classified the explanatory variables (called parameters) as control variables (kontrolované proměnné) and noise variables (nekontrolované proměnné).

The idea of Taguchi approach is robust parameter design guaranteeing small sensitivity to possible changes in the noise variables.

robust parameter design = process robustness study

The influence of noise variables can be evaluated because the noise variables can be often controlled in an experiment (the control variables change into noise variables in the production phase).

Table 11.1 Some Examples of Control Variables and Noise Variables

Application	Control Variables	Noise Variables
Development of a cake mix	Amount of sugar, starch, and other ingredients	Oven temperature, baking time, amount of milk added
Development of a gasoline	Ingredients in the blend; other processing conditions	Type of driver, driving conditions, changes in engine type
Development of a tobacco product	Ingredient types and concentrations; other processing conditions	Moisture conditions; storage conditions on tobacco
Large-scale chemical process	Processing conditions, including nominal ambiant temperature	Deviations from nominal ambient temperature; deviations from other processing conditions
Production of a box-filling machine for filling boxes of detergent	Surface area; geometry of the machine (rectangular, circular)	Particle size of detergent
Manufacturing a dry detergent	Chemical formulation, processing variables	Temperature and relative humidity during manufacture

Example

Example:

During baking, temperature can be both controlled and noise variable.

The controlled variable is the nominal temperature set on the oven.

Noise variable could correspond to random deviations from the nominal temperature (due, e.g., to the thermoregulator switching on and off or to some other unpredictable effects)

Taguchi approach

Taguchi suggested to investigate all combinations of values of controlled and noise variables.

For example, a factorial experiment 2^2 for controlled variables is displayed as the *inner array* on the following slide. For each combination of values of controlled variables, we use factorial experiment 2^2 to investigate all combinations of values of noise variables (these are displayed as the *outer arrays*).

Observations in each outer array are summarized into a suitable descriptive statistics (comprising both mean and variance of these observations), the so-called *signal-to-noise ratio* (SNR) that is then used as the response.

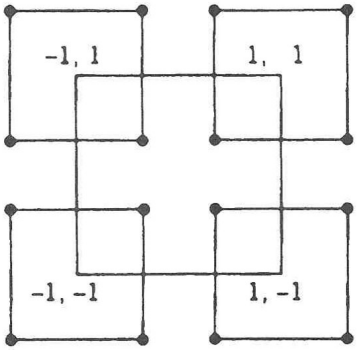


Figure 11.1 The $2^2 \times 2^2$ crossed array.

The essence of Taguchi approach is utilisation of noise variables jointly with minimization of variance.

In practice, one can:

1. either set values of controlled variables and carry out the experiment in the corresponding outer array,
2. or set the values of noise variables and, for these values, carry out the experiment in the inner array.

Both possible approaches can be described as split-plot design (that was not used by Taguchi). The second approach is usually recommended in practice.

In past, Taguchi design was criticized but these methods are popular (and useful) in practice.

Week 9

Topic:

- Consultation.
- Design.
- Protocol.
- Example (ACCEPT).

References: Bailey (2008). *Design of Comparative Experiments*. Cambridge University Press.

The ideal course of a designed experiment

1. Consultation (investigator & statistician).
2. Experiment design.
3. Data collection.
4. Data checks and corrections.
5. Data analysis.
6. Interpretation.

Consultation

Ideally, the investigator comes ahead of time, without sufficient time left for planning. In such case, the statistician should 1/ ask many questions, 2/ take time for thinking, 3/ compare several possible designs. In more complicated cases, one should consult the problem with other statisticians (specialists in respective fields).

In practice, the following two situations are common:

1. The investigator arrives one day before the start of the experiment. Then, his only aim is to check the field *experiment was consulted with a statistican* on a form and it would be pointless to pay much attention to his problems.
2. More often, the investigator has no idea that some time is needed for designing an experiment. Then, you should try to understand at least the main points of the experiment and propose as simple design as possible. You should stress that this may not be the optimal design.

Ideal and reality (Bailey, section 1.2)

Some tension between the client and statistican may arise mainly in these points:

1. clear definition of the aim of the experiment,
2. number of observations (repetitions): power $\times$ money,
3. blocks?
4. restrictions (financial, ethical, legal, ...),
5. choice of treatments (is placebo necessary?)
6. what are "experimental units" (e.g., field or part of field) for application of treatments?

Protocol

Protocol is written by statistician jointly with the investigator:

1. What is the aim of the experiment? I+s
2. Which treatments are to be compared? I
3. Methods: B
4. What exactly are experimental units?
5. Number of observations on each experimental unit.
6. Which measurements will be recorded. Description of data checks.
7. Applied experiment design.
8. Justification of the proposed design (number of observations, blocks, assumptions).
9. Randomization.
10. Plan of the experiment (which 'coded' treatments will be used at every experimental unit). This part of the protocol may be potentially hidden from those who will carry out the experiment,
11. Description of statistical analysis.

Example: project ACCEPT

Example:

Project ACCEPT: A Cluster-Randomized Trial of Community Mobilization, Mobile HIV Testing, Post-Test Support Services, and Real-Time Performance Feedback

Protocol can be downloaded at `http://www.cbvct.med.ucla.edu/`

Week 9–10

Topic:

- Design of computer experiments (uniform design).

References: Fang & Wang (1993). *Number-theoretic methods in statistics*. CRC Press.

Space-filling design

Today's topic is related to:

- calculation of expectation (over noise variables),
- numerical integration,
- computer experiments.

Often, we want to calculate the expected value of some characteristics, e.g., $Eh(X)$. For example, in order to estimate the expected output (of some industrial process), we can fix the optimal value of controlled variables (identified, e.g., by RSM) and average (integrate) the output over possible values of the noise variables.

We have to measure (or simulate) the output on a *grid* of values in the p-dimensional region of interest. In the following, we discuss some classical methods for choosing such "grid" (space-filling designs).

Design of computer experiments

Increasingly often, complex analytical problems are solved by sophisticated computer simulations.

Let us assume that behavior of certain device (or process) depends on random vector $\mathbf{X} = (X_1, \ldots, X_s)^\top$; FW94 describe an example of electric circuit depending on certain characteristics that may (randomly) vary. Other examples: artificial creatures (in computer games), guided missiles.

Usually, one develops appropriate mathematical model (e.g., a system of differential equations) and implements the resulting computer simulation. Note that, in this setup, the result with given starting conditions is usually nonrandom and, therefore, it is not reasonable to run the same simulation more than once.

Space-filling design are used to choose the set of input values for the computer simulation so that the estimate is as precise as possible.

Expected value estimation

A simple estimate, based on a sample mean of simulated values $h(\mathbf{X})$, can be obtained by Monte Carlo:

1. generate vectors $\mathbf{X}_i$ from the distribution of X,
2. $Eh(x)$ can be estimated by the sample mean $\bar{h} = \sum_{i=1}^{n} h(\mathbf{X}_i)/n$.

Remark 1: similar methods can be applied towards numeric integration (likelihood function for mixed models) or, e.g., in forest inventory (estimate of mean or total in more dimensional space).

Remark 2: Monte Carlo estimator is consistent but it is not efficient—therefore, some authors proposed alternative methods of generating $\mathbf{X}_i$ leading to estimators of $\bar{h}$ with smaller variance (these methods are usually trying to generate the values $\mathbf{X}_i$ more "uniformly").

Integration in rectangle

Let C^s denote the s-dimensional unit rectangle $[0,1]^s$ and consider the integral

$$I(f) = \int_{C^s} f(x)d(x).$$

There are interesting results concerning numerical evaluation of $I(f)$.

Theorem: (Koksma, Hlawka) Let f have bounded variation $V(f)$ on C^s. Then, for any $x_i \in [0,1)^s$, $i = 1, \ldots, n$,

$$\left| \int_{C^s} f(x)dx - \frac{1}{n}\sum_{i=1}^{n} f(x_i) \right| \leq V(f) D_N^*(x_1, \ldots, x_n),$$

where $D_N^*(x_1, \ldots, x_n)$ denotes the so-called star-discrepancy of the set $\{x_1, \ldots, x_n\}$.

Integration in rectangle

Proof: this was proved by [J. F. Koksma (1942–43) Een elgemeene stelling uit de theorie de gelijkmatige Verdeeling modulo 1, *Math.* B (Zutphen), **11**, 7–11] for $s = 1$ and [E. Hlawka (1961) Funktionen vom beschränkter Variation in der Theorie Gleichverteilung, *Ann. Mat. pure Appl.*, **54**, 325–333] for $s > 1$.

Application: better approximation can be obtained by choosing a set of points with smaller discrepancy.

For example [Hua, Kang: Applications of Number Theory to Numerical Analysis, Springer, 1981] show (on page 110) that:

$$\left| \int_{C^s} f(x)dx - \frac{1}{n}\sum_{i_1=0}^{m-1} \cdots \sum_{i_s=0}^{m-1} f(i_1/m, \ldots, i_s/m) \right| \leq V(f) 2^s n^{-1/s},$$

where $n = m^s$ is the number of quadrature points (on a regular grid).

Monte Carlo / simple random sampling

Assuming that the random vector X has uniform distribution on C^s, we obtain:

$$Ef(X) = \int_{C^s} f(x)dx = I(f).$$

An estimator of the integral $I(f)$ can be obtained by generating random sample $X_1, \ldots, X_n$ and by calculating the sample mean

$$I(f, n) = \frac{1}{n}\sum_{i=1}^{n} f(X_i).$$

For $s > 2$, CLT implies that the Monte Carlo estimator has a better rate of convergence ($O_P(n^{-1/2})$) than the quadrature rule on the previous slide (although the rate $O(n^{-1/s})$ can be improved for 'smoother' functions).

Rate of convergence

The rate of convergence of the Monte Carlo estimator is $O_P(n^{-1/2})$, where P denotes convergence in probability. Is it possible to obtain another type of convergence?

Law of iterated logarithm implies that:

$$\lim_{n\to\infty} \sup \sqrt{\frac{n}{2\log(\log n)}} \left| \int_{C^s} f(x)d(x) - \frac{1}{n}\sum_{i=1}^{n} f(X_i) \right| = \sigma^2(f)$$

with probability one (i.e. along almost all sequences x_i).

Therefore, the rate of convergence cannot be worse than $O(\sqrt{\log(\log n)/n})$ (almost surely).

Latin hypercube sampling (LHS) / latinské hyperkostky

Latin hypercube sampling is similar to latin squares. The aim is to guarantee that the final sample $\mathbf{X}_i$ covers uniformly all marginal distributions of the random vector $\mathbf{X} = (X_1, \ldots, X_s)^\top$.

Let's assume that $F(x) = \prod_{i=1}^{s} F_k(x_k)$ is the distribution function of $\mathbf{X}$.

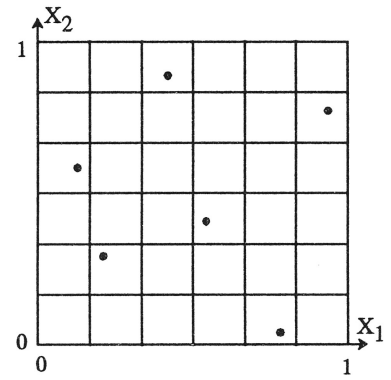
One possible LHS algorithm is:

1. Generate matrix $P\,(n \times s)$ such that its columns are independent random permutations of $\{0, 1, \ldots, n-1\}$.
2. Generate matrix $U\,(n \times s)$ containing iid random variables from $U(0,1)$ distribution (independent of P).
3. Observations $\{\mathbf{x}_j = (x_{j1}, \ldots, x_{js})^\top, j = 1, \ldots, n\}$, where

$$x_{jk} = F_k^{-1}\{(p_{jk} + u_{jk})/n\}$$

form a sample from the distribution $F(x)$ by LHS.

Example: LHS for $s = 2$ and $n = 6$.



Properties of LHS

Estimators $\bar{h}$ obtained by LHS are unbiased and, under some assumptions, have small variance (see FW94, p. 239–241). Further variance reduction can be achieved, e.g., by using "orthogonal arrays LHS".

According to FW94, p. 241, variances of $\bar{h}$ obtained from different samples $\mathbf{X}_i$ are:

$$\begin{aligned} \text{SRS:} &\quad \tfrac{1}{n}\, Var(h(\mathbf{X})), \\ \text{LHS:} &\quad \tfrac{1}{n}\, Var(h(\mathbf{X})) - \tfrac{c}{n} + O(n^{-1}), \\ \text{OALHS:} &\quad \tfrac{1}{n}\, Var(h(\mathbf{X})) - \tfrac{d}{n} + O(n^{-3/2}). \end{aligned}$$

IMPORTANT QUESTION: *Can we obtain better approximation by using sequences with smaller discrepancy?*

Uniform random design (URD)

Fang & Zhu (1993) proposed the glp (good lattice points) method. They proposed the following algorithm (for generating uniformly distributed design points within unit cube):

1. Generate glp set $\{a_k \in [0,1]^s, k = 1, \ldots, n\}$ using the generating vector $(n; h_1, \ldots h_s)$.
2. Generate n random vectors $u_i \in \mathbb{R}^s$ from the uniform distribution on $(-1, 1)^s$,
3. The URD sample is $\{x_k, k = 1, \ldots, n\}$, where

$$x_k = a_k + u_k/2n.$$

FW94 [Theorem 5.3–5.4] show that the estimator $\bar{h}_n$ is asymptotically unbiased ($|E\bar{h}_n - E(h(\mathbf{X}))| = O(n^{-1}\log^s n)$) and $Var(\bar{h}_n) = O(n^{-2}\log^{2s} n)$.

Good lattice points (glp)

Definition: Assume that $\mathcal{P} = \{x_k, k = 1, \ldots, n\}$ is a set of points on $C^s = [0,1]^s$, $N(\gamma, \mathcal{P})$ is the number of points such that $x_k \leq \gamma$, and $v([0,\gamma])$ denotes the volume of the rectangle $[0,\gamma]$. Then

$$D(n,\mathcal{P}) = \sup_{\gamma \in C^s} \left| \frac{N(\gamma,\mathcal{P})}{n} - v([0,\gamma]) \right|$$

is the *discrepancy* of the set $\mathcal{P}$.

Remark 1: an uniformly distributed set of points (on C^s) should have small discrepancy.

Remark 2: $D(n,\mathcal{P})$ is actually Kolmogorov-Smirnow distance for s-dimensional uniform distribution.

Good lattice points (glp)

A set with small discrepancy on C^s can be found by glp.

Definition: Let $(n; h_1, \ldots, h_s)$ be a vector of natural numbers such that $1 \leq h_i < n$, $h_i \neq h_j$ $(i \neq j)$, $s < n$ and the higherst common divisor $(n, h_i) = 1$, $i = 1, \ldots, s$. Let

$$\begin{cases} q_{ki} = kh_i (\text{mod } n) \\ x_{ki} = (2q_{ki} - 1)/2n, \end{cases} \qquad k = 1, \ldots, n, \ i = 1, \ldots, s,$$

where the multiplication modulo n is defined so that $1 \leq q_{ki} \leq n$. Then, $\mathcal{P}_n = \{x_k = (x_{k1}, \ldots, x_{ks})^\top, k = 1, \ldots, n\}$ is the *lattice point set* (množina síťových bodů) of the generating vector $(n; h_1, \ldots, h_s)$. If $\mathcal{P}_n$ has minimal discrepancy (among all possible generating vectors), it is the *glp* set.

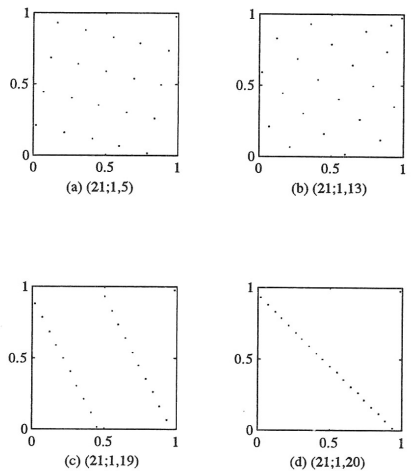


Figure 1.6 Comparisons between different generating vectors

Good lattice points (glp)

If $\{x\}$ denotes the fractional part of x, then x_{ki} can be easily calculated as follows:

$$x_{ki} = \left\{ \frac{2kh_i - 1}{2n} \right\}.$$

Theorem: For each prime number p, there exists a vector of natural number $h_p = (h_1, \ldots, h_s)$ such that the lattice point set of the generating vector $(p; h_1, \ldots, h_s)$ has discrepancy $D(p) < c(s)p^{-1}(\log p)^s$.

Proof: see, e.g., [Hlawka, E. (1962). Zur angenäherten berechnung mehrfacher integrale. Monatshefte für Mathematik, 66(2), 140–151].

Good lattice points (glp)

Tables of generating vectors can be found, e.g., in FW94[Appendix A] both for large n (suitable for numerical integration) and small n (convenient for computer design).

[Fang & Wang (1993). *Number-theoretic methods in statistics*. CRC Press] further discuss:

- other discrepancy measures,
- further methods for constructing small discrepancy point sets on C^s,
- methods for constructing small discrepancy sets on a simplex, unit ball, or unit sphere,
- multivariate normality or *sphericity* tests [Henze, Hlávka, Meintanis (2014) Testing for spherical symmetry via the empirical characteristic function, Statistics, 48(6), 1282–1296].

Generating vectors for small n and $s \in \{2, 3, 4\}$.

Table A.13 $s = 2, h_1 = 1$

n	5	7	9	11	13	15	17	19	21	23	25	27	29	31
h_2	2	3	4	7	5	11	5	14	13	9	7	16	23	21

Table A.14 $s = 3, h_1 = 1$

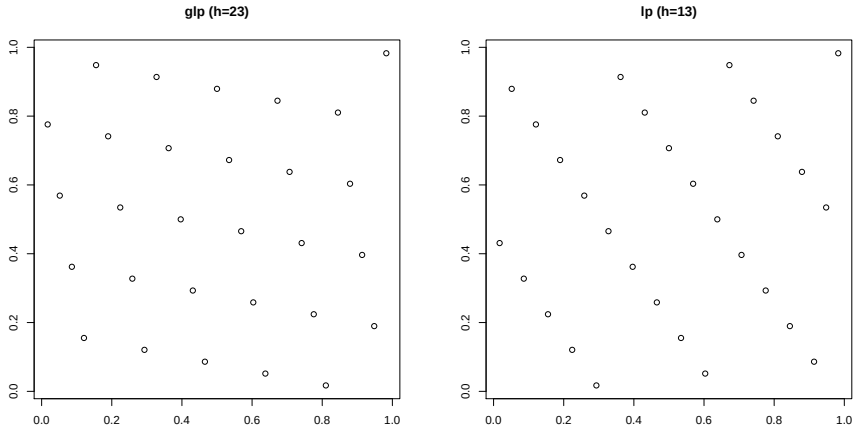
n	5	7	9	11	13	15	17	19	21	23	25	27	29	31
h_2	2	2	2	3	3	2	3	3	4	15	8	20	16	11
h_3	4	4	4	5	9	7	9	9	10	18	14	22	24	28

Table A.15 $s = 4, h_1 = 1$

n	7	9	11	13	15	17	19	21	23	25	27	29	31
h_2	2	2	2	6	2	2	5	2	2	4	5	4	15
h_3	3	4	5	8	4	4	7	10	5	6	17	6	19
h_4	6	7	7	10	8	8	9	17	10	9	25	16	22

Sample R code

```
n=29      # nr.obs
h1=1
h2=23     # generating vector of glp set

k=1:n
fract=function(x){x-floor(x)}
xk1=fract((2*k*h1-1)/(2*n))
xk2=fract((2*k*h2-1)/(2*n))
```

Week 10

Topic:

- Forest inventory: one-phase sampling scheme.
- Optimal sampling schemes.

References: Mandallaz, D. (2007). *Sampling techniques for forest inventories*. CRC Press.

Sampling finite populations

Assuming a particular population $\mathcal{P}$ of N individuals, we are interested in p response variables $Y_i^{(m)}$, $m = 1, \ldots, p$, $i = 1, \ldots, N$. Usually, we want to estimate population totals, means and variances:

$$\begin{aligned} Y^{(m)} &= \sum_{i=1}^{N} Y_i^{(m)} \\ \overline{Y}^{(m)} &= \frac{1}{N} \sum_{i=1}^{N} Y_i^{(m)} \\ S^2_{\overline{Y}^{(m)}} &= \frac{1}{N-1} \sum_{i=1}^{N} (Y_i^{(m)} - \overline{Y}^{(m)})^2. \end{aligned}$$

Often, we need to estimate also ratios $Y^{(m)}/Y^{(l)}$, covariances, and correlations. In practice, N is very large and one can investigate only a subset $s \subset \mathcal{P}$. Apart of calculating the estimates, the problem is also the choice of s.

Survey sampling: Horvitz-Thompson estimator

Defining $I_i = I(i \in s)$, we have that $\pi_i = P(I_i = 1)$ is the probability that individual i will be included in the sample. Interestingly,

$$\pi_i = E(I_1) = \sum_{s \ni i} p(s),$$

where $p(s)$ denotes the probability of selecting sample s.

The famous Horvitz-Thompson estimator

$$\hat{Y}_\pi^{(m)} = \sum_{i \in s} \frac{Y_i^{(m)}}{\pi_i} = \sum_{i \in \mathcal{P}} \frac{I_i Y_i^{(m)}}{\pi_i}$$

is unbiased if $\pi_i > 0$, $i = 1, \ldots, N$. This is a so-called design-based approach (the response variables $Y_i^{(m)}$ are fixed). More information: Survey Sampling (NMST 438).

Forest inventory: sampling schemes

According to Mandallaz (2008, Chapter 4), forest inventory is usually split into two phases:

- first phase: auxilliary information (maps, aerial or satellite photographs), very large sample size.
- second phase: terrestrial information from a sub-sample of the first phase sample:
  - first-stage: approximation to response variables (cheap measurement),
  - second-stage: exact response.

The trees can be chosen by simple random sampling (this is not used in practice) or by cluster random sampling (the information is gathered in some fixed-shape regions with randomly chosen origins.)

Sampling (inclusion probabilities and local density)

Let $B \subseteq \mathbb{R}^2$ denote some region: drawing a random point X uniformly in F, the probability of X falling within B is $\lambda(B \cap F)/\lambda(F)$. Trees are selected if they are within the circle $K_r(X)$ (with radius r and center X). Let I_i denote the indicator of i-the tree being within the (random) circle $K_r(X)$. Obviously

$$I_i = 1 \quad \Leftrightarrow \quad X \in K_r(u_i)$$

and the inclusion probabilities are

$$\begin{aligned} \pi_i &= P(I_i(X) = 1) = P(X \in K_r(u_i)) = \frac{\lambda(K_r(u_i) \cap F)}{\lambda(F)}, \\ \pi_{ij} &= \lambda(K_r(u_i) \cap K_r(u_j) \cap F)/\lambda(F). \end{aligned}$$

Finally, the (observed) local density is $Y(X) = \frac{1}{\lambda(F)} \sum_{i=1}^{N} \frac{I_i(X) Y_i}{\pi_i}$.

Local density

The density of X is $1/\lambda(F)$ and the local density

$$Y(X) = \frac{1}{\lambda(F)} \sum_{i=1}^{N} \frac{I_i(X) Y_i}{\pi_i}$$

thus satisfies

$$\begin{aligned} EY(X) &= \frac{1}{\lambda(F)} \int_F Y(x) dx = \frac{1}{\lambda(F)} \int_F \sum_{i=1}^{N} \frac{I_i(x) Y_i}{\lambda(K_r(u_i) \cap F)} dx \\ &= \frac{1}{\lambda(F)} \sum_{i=1}^{N} \frac{\int_F I_i(x) dx Y_i}{\lambda(K_r(u_i) \cap F)} = \frac{1}{\lambda(F)} \sum_{i=1}^{N} Y_i = \overline{Y}_F. \end{aligned}$$

In other words, we need to estimate the expectation $EY(X)$ that can also be represented as an integral of $Y(x)$ over the forest area F.

One-phase (terrestrial) sampling schemes

Let $F \subset \mathbb{R}^2$ denote the forest area. We are interested in a well-defined population $\mathcal{P}$ of N trees lying in F (the tree coordinates are denoted by $u_i \in F \subset \mathbb{R}^2$, $i = 1, \ldots, N$).

Again, the error-free responses are $Y_i^{(m)}$. Given any set $G \subseteq F$, the objective is to estimate spatial means (densities), totals or ratios:

$$\begin{aligned} \overline{Y}_G^{(m)} &= \frac{1}{\lambda(G)} \sum_{i \in G} Y_i^{(m)}, \\ Y_G^{(m)} &= \sum_{i \in G} Y_i^{(m)}, \\ R_{l,m} &= \frac{\overline{Y}_G^{(l)}}{\overline{Y}_G^{(m)}} = \frac{Y_G^{(l)}}{Y_G^{(m)}}, \end{aligned}$$

where $\lambda(G)$ denotes the surface area of the set G.

Assume that n points (circle centers X_i) are drawn uniformly and independently in the forest area G. The estimator $\hat{Y} = \frac{1}{n} \sum_{i=1}^{n} Y(X_i)$ has the design based variance:

$$\frac{1}{n\lambda^2(G)} \left(\sum_{i=1}^{N} \frac{Y_i^2(1-\pi_i)}{\pi_i} + \sum_{i \neq j}^{N} \frac{Y_i Y_j (\pi_{ij} - \pi_i \pi_j)}{\pi_i \pi_j} \right)$$

$$= \frac{1}{n\lambda(G)} \int_G \{Y(x) - \overline{Y}\}^2 dx$$

leading to

$$\widehat{Var}(\hat{Y}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \{Y(X_i) - \hat{Y}\}^2.$$

For details see Mandallaz (2007).

National forest inventory (NFI = NIL) in Czechia

http://nil.uhul.cz/uvodni-informace/projekt-nil

Optimal sampling schemes: anticipated variance

Mandallaz, p. 155: Meaningful optimality criteria must rely somehow on a super-population model. The actual population to be surveyed is viewed as one realization of many similar ones. Design-based variance, i.e. under hypothetical repetition of the samples, is fixed for the given realization at hand. The average of that variance under the super-population model is called the *anticipated variance*. Optimal sampling schemes are those which minimize the anticipated variance for given costs or, conversely, minimize the costs for a given anticipated variance. This concept has been used successfully for many standard problems when sampling finite populations.

In other words: we should consider a suitable super-population model (e.g., local Poisson model for location of trees) and minimize the anticipated variance (of chosen sampling scheme) either theoretically or by computer simulation.

Anticipated variance under the local Poisson model

Theorem:[Mandallaz 2007, Theorem 9.2.1, p. 158] Assuming local Poisson model with negligible boundary effects, the anticipated variance under simple random sampling of the one-p[hase one-stage estimator $\hat{Y}$ is approximately given by

$$E_\omega V(\hat{Y}) = .$$

The anticipated variance under cluster random sampling is given by

$$E_\omega V(\hat{Y}) =$$

where

$$(1+\theta_1)\beta_1^2 =$$

is the inflation factor for cluster-sampling and $M_{1,k}(x) = \sum_{l=1}^{M} I_{F_k}(x_l)$ is the number of points in a cluster with its origin at x falling into the Poisson stratum $F_{1,k}$.

Week 11–12

Topic:

- Design of regression experiments:
  - design matrix / matice experimentu,
  - confidence ellipsoids,
  - optimality criteria / kritéria optimality,
  - software (briefly).

References: Pázman et al (1986) *Riešené situácie z navrhovania experimentov*, ALFA, Bratislava.

Simple linear regression

Example:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \qquad \varepsilon_i \text{ jsou iid}, E\varepsilon_i = 0, Var\,\varepsilon_i = \sigma^2.$$

We know that $Var\,\hat{\beta} = \sigma^2(\mathcal{X}^\top\mathcal{X})^{-1} = \ldots$ and it follows that

- $\hat{\beta}_0$ and $\hat{\beta}_1$ are correlated (the correlation can be 'removed' by centering),
- $$Var\,\hat{\beta}_1 = \frac{\sigma^2}{n}\frac{n}{\sum(x_i - \bar{x}_n)^2}.$$

In order to obtain more precise estimator $\hat{\beta}_1$, we have to increase the value $\sum(x_i - \bar{x}_n)^2/n$ (assuming that x_i can be chosen in the interval $[a, b]$, the optimal design is given by taking one half of measurements at the point a and the other half at b).

Example: The weight of three objects (θ_1, θ_2, θ_3) should be determined as precisely as possible (θ_1, θ_2, θ_3). The objects can be weighted both by one-by-one or in groups.

We assume that:

- the laboratory weighing scale is not calibrated and all measurements are influenced by a systematic bias θ_4.
- all remaining effects are considered to be random (with constant variance).

Individual measurements can be written as:

$$y_i = \{F\}_{i1}\theta_1 + \{F\}_{i2}\theta_2 + \{F\}_{i3}\theta_3 + \theta_4 + \varepsilon_i,$$

where $E\varepsilon_i = 0$ and $Var\,\varepsilon_i = \sigma^2$.

The first design is based on these measurements:

$$\begin{aligned} y_1 &= \theta_4 + \varepsilon_1 \\ y_2 &= \theta_1 + \theta_4 + \varepsilon_2 \\ y_3 &= \theta_2 + \theta_4 + \varepsilon_3 \\ y_4 &= \theta_3 + \theta_4 + \varepsilon_4 \end{aligned}$$

The design matrix is:

$$F = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad F^{-1} = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Normal equations imply that $\hat{\theta} = F^{-1}y$ and, therefore, $Var\,\hat{\theta}_i = 2\sigma^2$ for $i = 1, 2, 3$.

The second design uses these four measurements:

$$\begin{aligned} y_1 &= \theta_1 + \theta_2 + \theta_3 + \theta_4 + \varepsilon_1 \\ y_2 &= \theta_1 + \theta_4 + \varepsilon_2 \\ y_3 &= \theta_2 + \theta_4 + \varepsilon_3 \\ y_4 &= \theta_3 + \theta_4 + \varepsilon_4 \end{aligned}$$

The design matrix is:

$$F = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad F^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 1 \end{pmatrix}.$$

We obtain $\hat{\theta}_1 = (y_1 + y_2 - y_3 - y_4)/2$, $\ldots$, $\hat{\theta}_3 = (y_1 - y_2 - y_3 + y_4)/2$ and $Var\,\hat{\theta}_i = \sigma^2$ for $i = 1, 2, 3$.

Linear model (more formally)

Let us assume that the output variable $y(x)$ in the experiment $x \in \mathcal{H}$ (where $\mathcal{H}$ is the set of all possible design points) satisfies:

$$y(x) = \theta_1 f_1(x) + \ldots \theta_m f_m(x) + \varepsilon(x),$$

i.e.,

$$y(x) = \theta^\top f(x) + \varepsilon(x),$$

where $E\varepsilon(x) = 0$ and $Var\,\varepsilon(x) = \sigma^2(x)$. We assume that variances $\sigma^2(x)$ are known or that $\sigma^2(x) = kw(x)$, where $w(x)$ is known.

Let N denote the prescribed (maximum possible) number of trials. Then, the *experiment design with prescribed size* N is each N-tuplet of points $x^{(1)}, \ldots, x^{(N)}$ from the set $\mathcal{H}$. The design points can be used repeatedly. We assume that trials are repeated independently (this also implies that the order of trials does not matter).

Information matrix / informační matice

If $x^{(1)}, \ldots, x^{(N)}$ is experimental design with size N, then we have for the vector of measurements $y = (y(x^{(1)}), \ldots, y(x^{(N)}))^\top$ that

$$y = F\theta + \varepsilon,$$

where $\{F\}_{ij} = f_j(x^{(i)})$, for $i = 1, \ldots, N$ and $j = 1, \ldots, m$.

The *information matrix* is

$$M = F^\top \Sigma^{-1} F = \sum_{i=1}^{N} f(x^{(i)}) f^\top(x^{(i)}) \sigma^{-2}(x^{(i)}),$$

where Σ^2 is a diagonal matrix with $\sigma^2(x^{(i)})$ on the diagonal.

Clearly, the matrix M^{-1} (or $h^\top M^- h$) is the covariance matrix of the random vector $\hat{\theta}$ (or the estimator of the estimable function $h^\top\theta$). It follows that the information matrix M can be used for comparing alternative experimantal designs.

Normalized design measure and information matrix

Let $x^{(1)}, \ldots, x^{(N)}$ be an experimental design with size N. The *normalized design measure* ξ of $x^{(1)}, \ldots, x^{(N)}$ is defined as:

$$\xi(x) = N(x)/N; \quad (x \in \mathcal{H}),$$

where $N(x)$ denotes the number of repetitions of x in the design $x^{(1)}, \ldots, x^{(N)}$.

Obviously, $M = NM(\xi)$, where

$$M(\xi) = \sum_{x \in \mathcal{H}} f(x) f^\top(x) \sigma^{-2}(x) \xi(x)$$

is the so-called *normalized information matrix*.

Normalized design measure and information matrix

For a normalized design measure ξ (or, more concisely *design* ξ) it holds:

1. $\xi(x) \geq 0$ for $x \in \mathcal{H}$,
2. $\sum_{x \in \mathcal{H}} \xi(x) = 1$,
3. the set $\{x; x \in \mathcal{H}, \xi(x) > 0\}$ is finite.

In the following, each function ξ defined on $\mathcal{H}$ satisfying the above three conditions will be considered as a normalized design measure.

The interpretation of the design ξ is that measurements are realized only in trials with $\xi(x) > 0$ and the number $\xi(x)$ is proportional to the number of independent repetitions of x.

Some remarks

Remark 1: For most methods, it is important to assume that the set $\mathcal{H}$ is finite (e.g., the weighing example) or that the set $\{f(x) : x \in \mathcal{H}\}$ is a bounded and closed subset of $\mathbb{R}^m$ (e.g., the simple linear regression example).

Remark 2: Sometimes, we have to fulfill restrictions on total cost C (instead of a the maximal number of observations N). Denoting by $c(x)$ the price of one run of the experiment with x, we define the design measure:

$$\xi(x) = N(x)c(x)/N; \quad (x \in \mathcal{H}),$$

and the corresponding information matrix:

$$M^c(\xi) = \sum_{x \in \mathcal{H}} f(x)f^\top(x)\sigma^{-2}(x)c^{-1}(x)\xi(x).$$

Formally, this can be achieved by replacing the variance $\sigma^2(x)$ by $\sigma^2(x)c(x)$.

Comparison of regression designs

Two designs ξ and η are *equivalent* if $M(\xi) = M(\eta)$, i.e., if $Var_\xi(h^\top\hat{\theta}) = Var_\eta(h^\top\hat{\theta})$, $\forall h \in \mathbb{R}^m$.

Design ξ is *uniformly better* than design η if

$$Var_\xi(h^\top\hat{\theta}) \leq Var_\eta(h^\top\hat{\theta}), \quad \forall h \in \mathbb{R}^m.$$

This holds if and only the matrix $M(\xi) - M(\eta)$ is positive semidefinite ($u^\top[M(\xi) - M(\eta)]u \geq 0$, $\forall u \in \mathbb{R}^m$). If a parameter $h^\top\theta$ is not estimable with design ξ, we define $Var_\xi(h^\top\hat{\theta}) = \infty$.

The uniformly best design ξ does not generally exist. Instead, one usually maximizes suitable *optimality criteria*.

Optimality criteria

Optimality criteria are usually defined as a real function Φ defined on the set of all information matrices. The function Φ is chosen so that better design ξ leads to smaller values of $\Phi\{M(\xi)\}$.

We will say that a design μ is Φ-*optimal* if $\Phi\{M(\mu)\} = \min_{\xi \in \Xi} \Phi\{M(\xi)\}$.

Optimality criterial can be divided to total and partial (this depends on whether we estimate all parameters or only their subset; eventually only some linear combinations).

Further, optimality criteria can be divided to minimax and average (assuming full rank of M, average criteria have gradient simplifying the search for the Φ-optimal design).

Interpretation

The interpretation of most common optimality criteria (D-, E-, A-) is related to confidence ellipsoids.

Assuming that the estimator has $\hat{\theta}$ (approximately) normal distribution $N(\theta, M^{-1}(\xi)/n)$ then

$$\sqrt{n}M^{1/2}(\xi)(\hat{\theta} - \theta) \sim N(0, \mathcal{I}_m)$$

implying that

$$n(\hat{\theta} - \theta)^\top M(\xi)(\hat{\theta} - \theta) \sim \chi^2_m.$$

Therefore,

$$P\left((\theta - \hat{\theta})^\top M(\xi)(\theta - \hat{\theta}) \leq \frac{\chi^2_m(1-\alpha)}{n}\right) = 1 - \alpha$$

defines a $1 - \alpha$ confidence ellipsoid for the parameter θ.

Confidence ellipsoid

The confidence ellipsoid can be described as:

1. the center is $\hat{\theta}$,
2. the size is given by the constant $\chi^2_m(1-\alpha)/n$,
3. the shape depends on the matrix $M(\xi)$: main half-axes of length $\sqrt{\chi^2_m(1-\alpha)/(n\lambda_i)}$ are in the direction of eigenvectors γ_i of $M(\xi)$.

D-optimality and volume of the confidence ellipsoid

The volume $V(M(\xi), d)$ of the confidence ellipsoid $\{\theta : \theta^\top M(\xi)\theta \le d^2\}$ can be found by using the spectral decomposition $M(\xi) = \Gamma\Lambda\Gamma^\top$ and the transformation $z = \Lambda^{1/2}\Gamma^\top\theta$:

$$\begin{aligned} V(M(\xi), d) &= \int_{\{\theta:\theta^\top M(\xi)\theta \le d^2\}} 1 d\theta = \int_{\{z:z^\top z \le d^2\}} 1 |\Gamma\Lambda^{-1/2}| dz \\ &= V(\mathcal{I}_m, d)|\Gamma\Lambda^{-1/2}| = V(\mathcal{I}_m, d)\sqrt{|\Gamma\Lambda^{-1/2}||\Lambda^{-1/2}\Gamma^\top|} \\ &= V(\mathcal{I}_m, d)|M^{-1}(\xi)|^{1/2} = V(\mathcal{I}_m, d)\prod_{i=1}^m \lambda_i^{-1/2}, \end{aligned}$$

where $V(\mathcal{I}_m, d)$ denotes the volume of m-dimensional unit ball.

Hence, the volume of a confidence ellipsoid is inversely proportion to the square root of determinant of the matrix $M(\xi)$. The so-called *D-optimality* corresponds to the minimization of $-\log|M(\xi)|$.

E-optimality

D-optimality can lead to long ellipsoids with small volume (and some estimators may have large variance). Therefore, other criteria try to minimize some measure of *overall elongation*.

The length of the longest half-axis of the confidence ellipsoid is

$$\max_{\|h\|=1} \{\chi^2_m(1-\alpha)n^{-1}h^\top M^{-1}(\xi)h\}^{1/2},$$

and its minimization (i.e., minimization of $\max_{\|h\|=1} h^\top M^{-1}(\xi)h$) leads to the *E-optimal* design.

Obviously

$$\max_{\|h\|=1} h^\top M^{-1}(\xi)h = \cdots = \max_i \lambda_i^{-1} = (\min \lambda_i)^{-1},$$

where λ_i are the eigenvalues of the information matrix $M(\xi)$.

A-optimality

Clearly, projections of the confidence ellipsoid into coordinate axes are proportional to the (marginal) standard deviations $\hat{\theta}_i$.

The criterion of *A-optimality*, $\sum(M^{-1}(\xi))_{ii}$, leads to the minimization of the sum of marginal variances of $\hat{\theta}_i$. Geometrically, this corresponds to the minimization of the diagonal of a rectangle circumscribed to the confidence ellipsoid.

The criterion of A-optimality can be easily rewritten in terms of eigenvalues of $M(\xi)$:

$$\begin{aligned} \sum Var\,\hat{\theta}_i &= \text{tr}\, Var(\hat{\theta}) = \frac{1}{n}\text{tr}\{M^{-1}(\xi)\} \\ &= \frac{1}{n}\sum \lambda_i^{-1}. \end{aligned}$$

G-optimality

G-optimal design minimizes the criterion

$$\sup_{x \in \mathcal{H}} f^\top(x) M^{-1}(\xi) f(x),$$

that can be interpreted as *the maximal width of a confidence band* for $y(x)$ because

$$Var\, \hat{y}(x) = Var\, f^\top(x)\hat{\theta} = Var\, f^\top(x) M^{-1}(\xi) f(x)/n.$$

Remark: D-optimality is equivalent to G-optimality in a homoscedastic setup (P86, Proposition IV.6, str. 88).

Overview of other optimality criteria: see, e.g., P86, p. 150–151.

Kritérium		Súhrnné			Čiastkové		
Názov		$\Phi(\mathbf{M}) =$	Podmienky	Interpretácia	$\Phi(\mathbf{M}) =$	Podmienky	Interpretácia
Minimaxné kritériá	Všeobecné minimaxné	$\sup_{h \in \mathcal{H}} \boldsymbol{h}^\top \mathbf{M}^{-1} \boldsymbol{h}$	$\mathcal{H} \subset R^m$, $\dim \mathcal{H} = m$		$\sup_{h \in \mathcal{H}} \boldsymbol{h}^\top \mathbf{M}^{-1} \boldsymbol{h}$	$\mathcal{H} \subset R^m$, $\dim \mathcal{H} =$ ľubovoľná $< m$	
	G-optimality	$\sup_{x \in \mathcal{X}} \boldsymbol{f}^\top(x) \mathbf{M}^{-1} \boldsymbol{f}(x)$		minimalizácia rozptylu odhadu funkcie odozvy	čiastkové kritérium sa nepoužíva		
	E-optimality	$\max_{\substack{\|\boldsymbol{h}\|=1 \\ \boldsymbol{h} \in \mathcal{H}}} \boldsymbol{h}^\top \mathbf{M}^{-1} \boldsymbol{h}$		minimalizácia najväčšieho rozptylu odhadu funkcie parametrov	$\max_{\substack{\|\boldsymbol{h}\|=1 \\ \boldsymbol{h} \in \mathcal{L}}} \boldsymbol{h}^\top \mathbf{M}^{-1} \boldsymbol{h}$	$\mathcal{L} \subset R^m$ je s-rozmerný podpriestor $(s < m)$	
Priemerové kritériá	Všeobecné L_p-kritérium	$\left[\frac{\sum_{i=1}^{m} (\{\mathbf{H}\}_{i\cdot} \mathbf{M}^{-1} \{\mathbf{H}\}_{i\cdot})^p}{m}\right]^{1/p}$	$p > 0$, $\det \mathbf{H} \neq 0$	iný názov: stopové kritérium, je univerzálne. Ak $\mathbf{H} = \mathbf{I}$, $p \to 0$, tak D-optimalita, ak $p \to \infty$, tak E-optimalita	$\left[\frac{\sum_{i=1}^{s} (\{\mathbf{H}\}_{i\cdot} \mathbf{M}^{-1} \{\mathbf{H}\}_{i\cdot})^p}{s}\right]^{1/p}$	$s < m$, $p > 0$, hodnosť $(\mathbf{H}) = s$, $\exists\, \mathbf{H} = \mathbf{MQ}$	univerzálna trieda kritérií
	Lineárne kritériá: Všeobecné lineárne kritérium	$\sum_{i,j=1}^{m} \{\mathbf{W}\}_{ij} \{\mathbf{M}^{-1}\}_{ij}$	$\mathbf{W}$ je pozitívne definitná	vážený súčet rozptylov a kovariancií	$\sum_{i,j=1}^{s} \{\mathbf{W}\}_{ij} \{\mathbf{M}^{-}\}_{ij}$	$\mathbf{W}$ typu $s \times s$, pozitívne definitná, $s < m$	
	d-optimality	súhrnné neexistuje			$\boldsymbol{h}^\top \mathbf{M}^{-} \boldsymbol{h}$	ak $\boldsymbol{h}^\top \boldsymbol{\theta}$ je odhadnuteľné	rozptyl odhadu $\boldsymbol{h}^\top \boldsymbol{\theta}$
	A-optimality	$\sum_{i=1}^{m} \{\mathbf{M}^{-1}\}_{ii}$		súčet rozptylov odhadov $\hat{\theta}_1, \ldots, \hat{\theta}_m$	$\sum_{i=1}^{s} \{\mathbf{M}^{-1}\}_{ii}$	$s < m$, $\theta_1, \ldots, \theta_s$ sú odhadnuteľné	súčet rozptylov $\hat{\theta}_1, \ldots, \hat{\theta}_s$
	Váženej A-optimality	$\sum_{i=1}^{m} w_i \{\mathbf{M}^{-1}\}_{ii}$	$w_i > 0$ $i = 1, \ldots, m$	vážený súčet rozptylov odhadov $\hat{\theta}_1, \ldots, \hat{\theta}_m$	$\sum_{i=1}^{s} w_i \{\mathbf{M}^{-}\}_{ii}$	$s < m$, $\theta_1, \ldots, \theta_s$ sú odhadnuteľné	vážený súčet rozptylov odhadov $\hat{\theta}_1, \ldots, \hat{\theta}_s$
	D-optimality	$-\ln \det \mathbf{M}$		minimalizácia objemu elipsoidu spoľahlivosti pre $\hat{\theta}_1, \ldots, \hat{\theta}_m$	$\ln \det (\{\mathbf{M}^{-}\}_{i,j=1}^{s})$	$s < m$, $\theta_1, \ldots, \theta_s$ sú odhadnuteľné	minimalizácia objemu elipsoidu spoľahlivosti pre $\hat{\theta}_1, \ldots, \hat{\theta}_s$
Súčinové kritérium		súhrnné neexistuje			$\sum_{i_1, \ldots, i_s=1}^{m} h_{i_1, \ldots, i_s} \times \prod_{k=1}^{s} \{\mathbf{M}^{-}\}_{i_k i_k} h_{i_1, \ldots, i_s}$		minimalizácia rozptylu odhadu polynómu $p(\boldsymbol{\theta}) := \sum_{i_1, \ldots, i_s=1}^{m} h_{i_1, \ldots, i_s} \theta_{i_1} \ldots \theta_{i_s}$

Optimization methods

Most popular approaches to optimization are:

1. Numeric optimization (using gradient).
2. Orthogonal designs.
3. A catalogue of optimal designs.
4. Elfving's method (based on some interpretations related to a point in which certain line intersects the surface of a certain m-dimensional polyhedron).
5. Exclusion of under-informative observations (followed by application of another optimization method).

Assessment of the degree of optimality

A direct comparison of the value $\Phi\{M(\xi)\}$ to the optimal design $\Phi\{M(\xi^*)\}$ is not possible. Assuming the existence of the gradient $\nabla\Phi\{M(\mu)\}$, a convenient measure of quality (distance from the optimal design) can be defined as

$$d(\mu) = \sum_{i,j=1}^{m} \{\nabla\Phi\{M(\mu)\}\}_{ij} \{M(\mu)\}_{ij} - \min_x \sigma^{-2}(x) f^\top(x) \nabla\Phi\{M(\mu)\} f(x).$$

It holds that:

- $d(\mu) \geq 0$ for all designs μ,
- $d(\mu) = 0$ if and only if μ is Φ-optimal,
- $|\Phi\{M(\mu)\} - \min_\xi \Phi\{M(\xi)\}| \leq d(\mu)$.

Proof: Proposition IV.27 and IV.28 in Pázman (1986, Foundations).

Assessment of the degree of optimality

Example: For D-optimality, we have $\Phi\{M(\mu)\} = -\log|M(\mu)|$ and $\nabla\Phi\{M(\mu)\} = -M^{-1}(\mu)$.

$$d(\mu) = \max_x \sigma^{-2}(x) f^\top(x) M^{-1}(\mu) f(x) - m$$

The design μ is approximately D-optimal if

$$\sigma^{-2}(x) f^\top(x) M^{-1}(\mu) f(x) \le m + \varepsilon$$

for each x.

Linear dependency on factors

Let us assume the first-order regression model

$$E[Y(t)] = \theta_0 + \theta_1 t_1 + \ldots \theta_m t_{m-1}$$

where $t = (t_1, \ldots, t_{m-1})^\top \in X \subset \mathbb{R}^{m-1}$ and $\sigma^2(t) \equiv \sigma^2$. [P86]: (Such model) ... is often used when there is no theoretical knowledge of the observed object, and the model is the simplest approximation to reality.

Denote by Ξ_M a fixed size (N) design such that all explanatory variables are centred ($\sum_{t\in X} t_i \xi(t) = 0$) and normed ($\sum_{t\in X} t_i^2 \xi(t) = 1$).

Theorem: Suppose that X is a compact subset of $\mathbb{R}^{m-1}$ and that there is a design $\xi^* \in \Xi_M$ having a diagonal information matrix. Then $1 = Var_{\xi^*} \hat\theta_i \le Var_\xi \hat\theta_i$ and the design ξ^* is D-optimal (i.e., also G-optimal) and A-optimal within the set Ξ_M.

Proof: P86, Proposition VI.1, p. 171.

Elfving's method

Elfving's method minimizes the variance of the estimator $h^\top\hat\theta$ (the so-called *d-optimality*) and it proceeds as follows:

We define the set

$$T = \left\{ \frac{f(x)}{\sigma(x)} : x \in \mathcal{H} \right\} \bigcup \left\{ -\frac{f(x)}{\sigma(x)} : x \in \mathcal{H} \right\}$$

and its convex hull S.

Let p denote a line passing through the origin and parallel to the vector h. Let P denote the intersection of the line p and the border of S. The point P can be written as a convex combination of points from the set T, i.e. $P = \sum_{i=1}^n \lambda_i f(x^{(i)})$, defining the d-optimal design $\xi^*(x^{(i)}) = |\lambda_i|$.

Example: Motor thrust during acceleration is investigated on a test track. The movement of the car on the test track can be approximated by the function

$$s(x) = vx + zx^2/2,$$

where $s(x)$ is the position in time x, v is the speed in time 0 and z is the acceleration. Parameters v and z are not known. The measuring device allows to measure the car's location $10\times$ within the time interval 0–10 s (and it is possible to measure the position more than once in a single moment).

We use the linear model

$$Ey(x_i) = vx_i + \frac{z}{2}x_i^2 = \theta_1 x_i + \theta_2 x_i^2,$$

where $Var\, y(x_i) = \sigma^2$ (and $f(x_i) = (x_i, x_i^2)^\top$) with the information matrix

$$M = \sum_{k=1}^{10} f(x_k) f^\top(x_k) \sigma^{-2} = \frac{1}{\sigma^2} \begin{pmatrix} \sum x_k^2 & \sum x_k^3 \\ \sum x_k^3 & \sum x_k^4 \end{pmatrix}.$$

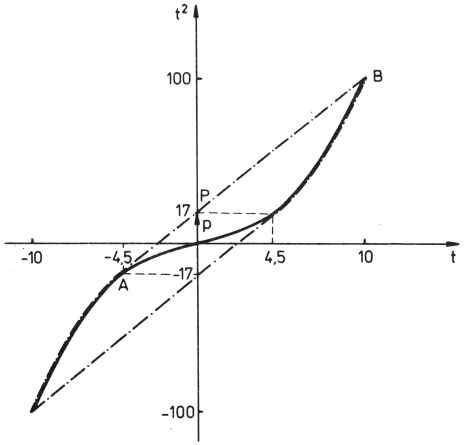
At first, we want to find the optimal experiment for estimating the accelaration. We choose the d-optimality criterion with $h = (0,1)^\top$, i.e., we want to minimize the function

$$\Phi_1(M) = (0,1)M^{-1}\begin{pmatrix}0\\1\end{pmatrix}.$$

The d-optimal design can be found by Elfving's method.

The set T is denoted by a full line, its convex hull (set S) is denoted by the dash-dotted line (*čerchovaně*, originally *bodkočiarkovane*).

The point P is the intersection of the border of S and the vector h. Finally, we express the point P as a linear combination of points A and B.



The point P is the intersection of the border of S and the vector h and it can be expressed as a linear combination of points A and B.

$$P = \begin{pmatrix}p_1\\p_2\end{pmatrix} = \begin{pmatrix}0\\17\end{pmatrix} = 0.7\begin{pmatrix}-4.5\\-4.5^2\end{pmatrix} + 0.3\begin{pmatrix}10\\10^2\end{pmatrix}$$

d-optimal design is:

$$\begin{aligned}\xi^*(4.5) &= 0.7\\ \xi^*(10) &= 0.3\end{aligned}$$

The information matrix of the design is:

$$M(\xi^*) = \sigma^{-2}\left[\begin{pmatrix}(4.5)^2 & (4.5)^3\\(4.5)^3 & (4.5)^4\end{pmatrix}0.7 + \begin{pmatrix}10^2 & 10^3\\10^3 & 10^4\end{pmatrix}0.3\right] = \dots$$

The variance matrix of the estimator $\hat\theta$ is:

$$M^{-1}(\xi^*) = \sigma^2\begin{pmatrix}0.2555 & 0.0283\\0.0283 & 0.0001\end{pmatrix}$$

and the variance of the estimator of θ^2 is $10^{-4}\sigma^2$.

Recommended work-flow

- **Determination of a model:** what can be measured; the relation between observations and unknown parameters; determination of measurement's precision (up-to a multiplicative constant); what should be estimated (choice of optimality criteria).
- **Determination of the optimal design:** using literature; comparison to some simple (but reasonable) design.
- **Verification of the optimal design:** is the design *reasonably good* also according to other optimality criteria?; consider small modifications that simplify the design and do not significantly worsen its properties of the design; check feasibility of the proposed design.

WARNING: the optimal design may not allow verificaiton (testing) of model's assumptions (e.g., linearity of the regression function)!

Software and exercise

Library `AlgDesign` in R: functions `optFederov()` and `optMonteCarlo()`.

Exercise 1: Try to calculate the d-optimal design for car's acceleration (eventually also other parameters) using library `AlgDesign`.

Exercise 2: Find the optimal design for weighing on a balance scale (i.e., if the objects can be put on both weighing pans).

Week 11

Topic:

- Elfving's method.
- Examples: bridge length measurements.

References: Pázman (1986) *Foundations of optimum experimental design*, D. Reidel. Pázman et al (1986) *Riešené situácie z navrhovania experimentov*, ALFA, Bratislava.

Ordering of regression experiments

Pázman (1986, kap. III) investigates the set of all information matrices $\mathfrak{M} \equiv \{M(\xi); \xi \in \Xi\}$, where Ξ is the set of all designs (normalized design measures).

Last week, we have used information matrices to define *partial ordering* on the set Ξ (designs are equivalent if their information matrices are equal; design ξ is better η if $M(\xi) - M(\eta)$ is positive definite).

Definition: Design ξ is acceptable if there is not any better design η.

Recall that $M = \sum_x f(x)f(x)^\top \xi(x)$. It is easy to see that both Ξ and $\mathfrak{M}$ are convex.

Some consequences of convexity

Theorem:[Caratheodory] Let $T \subset \mathbb{R}^k$. Then

$$\text{hull}(T) = \{z \in \mathbb{R}^k : z = \sum_{i=1}^{k+1} \alpha_i x_i, x_i \in T, \alpha_i \geq 0, \sum_{i=1}^{k+1} \alpha_i = 1\}.$$

Proof: Optimization (Lachout).

By Caratheodory theorem, each point in the convex hull of T can be written as a convex combination of at most $k+1$ points from T.

Some consequences of convexity

Theorem:[Pázman 1986, Proposition III.11] For each design ξ there exists equivalent design η such that its support $X_\eta = \{x; \eta(x) > 0\}$ contains at most $m(m+1)/2+1$ points.

Proof: see Pázman (1986), p. 60–61 (it follows from Caratheodory Theorem for convex hull: any point $x \in \mathbb{R}^d$ lying in the convex hull of set P can be expressed as a convex combination of $d+1$ or less points from P).

Elfving's set

Let S denote the smallest convex set in $\mathbb{R}^m$ containing the set $T \equiv \{f(x); x \in X\} \cup \{-f(x); x \in X\}$ (S is the convex hull of T).

Elfving's set S is important for design comparisons.

Theorem:[Pázman 1986, Proposition III.7] For each design ξ there exists a design η that is not worse and such that $\{f(x); \eta(x) > 0\}$ is a subset of the border of the set S.

Proof: see Pázman (1986), p. 56–57 (based on a construction of the design η).

This theorem simplifies the search for optimal design (because we can exclude points x that do not lie on the border of the convex hull S).

d-optimality

Let us denote $Var_\xi g = Var_\xi(g^\top \hat{\theta})$. Pázman (1986, sekce III.3) shows that the function $M(\xi) \longrightarrow Var_\xi g$ is lower semicontinuous on the compact set $\mathfrak{M}$ and, therefore, it achieves its minimum, i.e., there exists a design ξ^* such that $Var^*_\xi g = \min_{\xi \in \Xi} Var_\xi g$.

Theorem:[Pázman 1986, Proposition III.16] Let us assume that $Var_\xi g < \infty$ and vectors $\{f(x); \xi(x) > 0\}$ are linearly dependent. Then there exists design η such that $X_\eta \subsetneq X_\xi$ and $Var_\eta g \leq Var_\xi g$. The design η can be chosen so that it contains at most m points.

Proof: see Pázman (1986), p. 69–71 (based on a construction of the better design η).

This implies that the d-optimal design ξ^* (minimizing variance of the estimator $g^\top \hat{\theta}$) can be chosen as 'at most m points on the border of the set S'.

Elfving's Theorem

Theorem:[Pázman (1986, Proposition III.17), Elfving (1952)] The design ξ^* satisfies $Var^*_{\xi^*} g = \min_{\xi \in \Xi} Var_\xi g$ if and only if there exists set $Y \subset X_{\xi^*}$ and a constant c such that:

1. cg lies on the border of the set S,
2. $cg = \sum_{x \in Y} f(x)\xi^*(x) - \sum_{x \in X-Y} f(x)\xi^*(x)$.

Then

$$\underset{\xi^*}{Var}\, g = c^{-2} = \inf\{\lambda; \lambda > 0, \lambda^{-1} g \in S\}$$

Proof: see Pázman (1986), p. 72–74.

Example: bridge centre line / osa mostu (in Slovak)

Príklad 1.6. *Vytýčenie osi mosta*. V rámci prípravných prác stavby nového mosta ponad rieku je potrebné vytýčiť os mosta, t. j. určiť dva body P_1 a P_2 na brehoch rieky a presne zmerať ich vzdialenosť (pozri *obr. 1.4*). Pretože používané veľmi presné optické diaľkomery nemožno v niektorých prípadoch používať pri meraní ponad vodnú plochu (vznikajú reflexie od vodnej hladiny, ktoré narúšajú presnosť merania), treba vzdialenosť bodov P_1, P_2 zamerať nepriamo. Robí sa to tak, že na každom brehu rieky sa zvolí niekoľko stanovíšť. Tieto stanovištia vytvárajú geodetickú sieť, v ktorej je dovolené merať vzdialenosť medzi ľubovoľnými dvoma stanovišťami ležiacimi na tom istom brehu rieky a uhly definované ľubovoľnou trojicou stanovíšť (aj takých, ktoré ležia na rôznych brehoch rieky). Z údajov získaných meraním v geodetickej sieti sa na základe trigonometrických vzťahov vypočíta hľadaná dĺžka osi mosta.

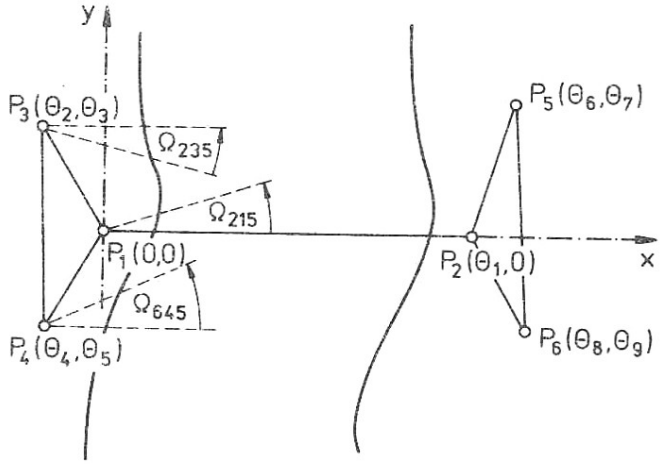
Example: bridge centre line / osa mostu

Example: Before the start of construction works of a new bridge, it is necessary to find its centre line, i.e., determine points P_1 and P_2 on both river banks and measure ther distance as precisely as possible (see the picture on the following slide). Unfortunately, standard and very precise optical distance-meters cannot be used for measuring distances above water (because water reflections disturb the instrument). Therefore, the distance between P_1 and P_2 has to be measured indirectly. This can be achieved by choosing several locations on each river bank. These locations define a geodetic net in which one can measure:

1. distances between any points on the same river bank,
2. angles defined by arbitrary three locations (including those on the other bank).

Finally, the distance between P_1 and P_2 is calculated from basic trigonometric identities.

Example: bridge centre line



Week 12

Topic:

- Summary (nonparametric regression models).
- Exam.

Summary

Main aim: introduction to standard notions in experimental design (so course graduates can, e.g., understand the description of libraries in `CRAN Task View: Design of Experiments`).

Further topics:

Nonparametric regression "Uniform design" is usually optimal, more information can be found in [Titterington: Optimal design in flexible models, including feed-forward networks and nonparametric regression, in: Optimum Design 2000 (A. Atkinson, B. Bogacka and A. Zhigljavsky, eds.), Nonconvex Optim. Appl., Vol. 51, Kluwer, Dordrecht, 2001, pp. 261–272.]

Applications in medicine NMST532: Plánování a analýza lékařských studií / planning and analysis of medical studies (in summer semester).