Random effects & repeated measurements

Anton Rask Lundborg
arl@math.ku.dk

Copenhagen Causality Lab
Department of Mathematical Sciences

December 20, 2023

Lecture outline

- Why is it called ANOVA?
- Experimental design and random effects.
    - Hypothetical data example: Comparison of two treatments.
- Data example 1: Color of pork.
    - Random effects in models with continuous response.
- Data example 2: Germination of Orobanche seeds.
    - Overdispersion in logistic regression.
- Data example 3: Growth of Baobab trees.
    - Repeated measurements.
    - General concepts about analysis of repeated measurements. We will also briefly talk about analysis using summary measures.

ANalysis Of VAriance: Why this name?

Growth of rats example:

antibio	vitamin	
	0	5
0	1.30, 1.19, 1.08	1.26, 1.21, 1.19
40	1.05, 1.00, 1.05	1.52, 1.56, 1.55

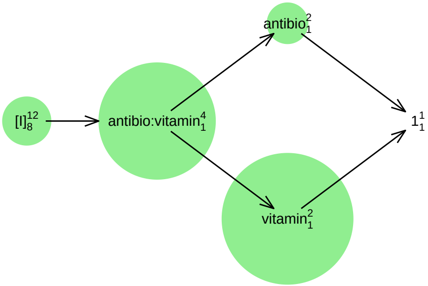
```
> anova(lm(growth ~ antibio + vitamin + antibio:vitamin, data = rats))
Analysis of Variance Table

Response: growth
                Df   Sum Sq  Mean Sq F value    Pr(>F)
antibio          1 0.020833 0.020833  5.6818  0.044292 *
vitamin          1 0.218700 0.218700 59.6455 5.622e-05 ***
antibio:vitamin  1 0.172800 0.172800 47.1273  0.000129 ***
Residuals        8 0.029333 0.003667
```

- The total variation in the response is decomposed using the right-hand side of $\sim$ from left to right(!):

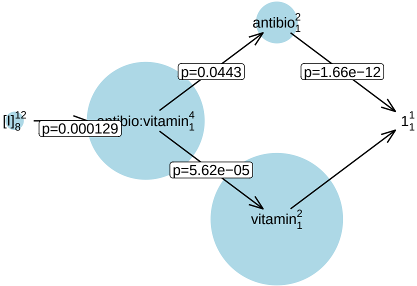
$$SS_{\text{total}} = \sum_{i=1}^{N} \left(\text{growth}_i - \mu_{\text{growth}}\right)^2 = SS_{\text{antibio}} + SS_{\text{vitamin}} + SS_{\text{antibio:vitamin}} + SS_{\text{error}}$$

Sum-of-Squares and Mean-Sum-of-Squares

F-test: Is the systematic variation large relative to the random variation?



SS_{antibio}, SS_{vitamin}, $SS_{\text{antibio:vitamin}}$, SS_{error} visualized by area of the green circles.



Variance estimates $\text{MSS} = \text{SS}/\text{df}$ visualized by area of the blue circles.

What does `drop1()` do for linear normal models?

The same tests but obeying the hierarchical principle!

```
> anova(lm(growth ~ antibio + vitamin + antibio:vitamin, data=rats))
Analysis of Variance Table

Response: growth
                Df   Sum Sq  Mean Sq F value    Pr(>F)
antibio          1 0.020833 0.020833  5.6818  0.044292 *
vitamin          1 0.218700 0.218700 59.6455 5.622e-05 ***
antibio:vitamin  1 0.172800 0.172800 47.1273  0.000129 ***
Residuals        8 0.029333 0.003667
```

```
> drop1(lm(growth ~ antibio * vitamin, data=rats), test="F")
Single term deletions

Model:
growth ~ antibio * vitamin
                Df Sum of Sq      RSS     AIC F value   Pr(>F)
<none>                       0.029333 -64.167
antibio:vitamin  1    0.1728 0.202133 -43.005  47.127 0.000129 ***
```

Summary of ANOVA

Analysis-of-variance for linear normal models:

- The total variation around the common mean is decomposed into the parts explained by the used fixed effects and the error term.
- Dividing by the corresponding degrees-of-freedom yields estimates for the variation "between groups" (the fixed effects) and "within groups" (the error term).
- Test for the null hypothesis of no effect is done on the ratio

  $$\frac{\text{between group variation}}{\text{within group variation}} = \frac{\text{MSS}_{\text{effect}}}{\text{MSS}_{\text{error}}}.$$

  Thus, we are comparing systematic variation to random variation, just as we do with a t-test!
- For non-orthogonal designs, this is no longer as simple since the order of tests can affect the p-values.

Random effects

- Every linear normal model has at least one random effect, namely the error term, which captures non-modelled biological variation, e.g.

$$\text{son}_i = \alpha + \beta \cdot \text{father}_i + \underbrace{\text{error}_i}_{\sim\mathcal{N}(0,\sigma^2)}$$

- Additional random effects may be used to capture common non-modelled biological variation, e.g.

$$\text{son}_i = \alpha + \beta \cdot \text{father}_i + \underbrace{A(\text{family}_i)}_{\sim\mathcal{N}(0,\sigma^2_{\text{family}})} + \underbrace{\text{error}_i}_{\sim\mathcal{N}(0,\sigma^2)}$$

$$\text{son}_j = \alpha + \beta \cdot \text{father}_j + \underbrace{A(\text{family}_j)}_{\sim\mathcal{N}(0,\sigma^2_{\text{family}})} + \underbrace{\text{error}_j}_{\sim\mathcal{N}(0,\sigma^2)}$$

  If $\text{family}_i = \text{family}_j = \text{Lundborg}$, say, the two sons (e.g. lecturer and his hypothetical brother) share a common unobserved component $A(\text{Lundborg})$ (inherited genes).

Hypothetical data example: 1-way ANOVA

Comparison of two treatments. Five animals per treatment

treat	Observations					Mean
1	13.8	14.2	11.2	11.4	13.9	12.90
2	13.1	12.3	11.0	14.0	10.7	12.22

```
> anova(lm(y ~ treat, data=rep1))
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value Pr(>F)
treat      1  1.156  1.1560  0.5643  0.474
Residuals  8 16.388  2.0485
```

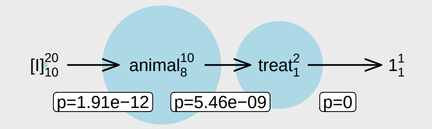
The *F*-test compares variation between treatments relative to variation within treatments. Here treatment is non-significant ($p = 0.474$).

Incorrect analysis of 1-way ANOVA with 2 measurements

We measure each animal again with a precise measuring tool:

treat	Observations					Mean
1	13.8	14.2	11.2	11.4	13.9	12.90
	13.7	14.3	11.1	11.3	14.1	
2	13.1	12.3	11.0	14.0	10.7	12.22
	13.2	12.1	11.0	14.1	10.7	

```
> anova(lm(y ~ treat + animal, data=rbind(rep1, rep2)))
Analysis of Variance Table
Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
treat      1  2.312  2.3120  330.29 5.461e-09 ***
animal     8 34.466  4.3082  615.46 1.906e-12 ***
Residuals 10  0.070  0.0070
```



How can doubling the number of measurements on the same animals change significance from (NS) to (***) ??

What went wrong?

- In the first analysis (slide 8) residual variation (mainly) represents variation between animals.
- In the second analysis (slide 9) (with `animal` as a predictor) residual variation represents measurement error.
- Treatments are compared by comparing different animals. Therefore, the relevant variation for comparing treatments is variation between animals.
- Conclusion: First analysis is OK. Second analysis shows that the two groups of animals are different, but not necessarily that they are more different than animals are in general!

Factors with random effect

- The factor `animal` is not reproducible.

- The specific animals are
  - of no interest beyond the present experiment.
  - representatives of a population.

  These statements characterize factors which we want to include as random effects in the model.

- Estimates describe properties of the population, typically the standard deviation in the scale of the response.

- Typical factors with random effects: field, litter, replication, day, herd, and block factors in general.
  - But possibly(!) also: strain, species, and other factors of no particular interest in a given experiment.

Factors with fixed (also called “systematic”) effect

- The factor `treatment` is reproducible.

- The specific treatments
  - are of interest beyond the present experiment — that is why we did it!
  - contain reproducible effects.
  - only represent themselves.

  These statements are typical for factors which we want to include as fixed effects in the model.

- Estimates describe properties of the individual “treatments”, typically the mean of the response.

Fixed effects vs. Random effects

- Fixed effects:
    - Estimated mean for each level of the factor.
    - Require many parameters (which is a bad thing!).

- Random effects:
    - Estimated standard deviation between the levels.
    - Only require 1 parameter (which is a good thing!).
    - Requires iterative search for parameter estimates (which is a bad thing, although it does not really pose an actual problem).

Rough rule of thumb

A block factor with 4 or fewer levels is often used as a fixed effect (this is to avoid estimating variation from few points and since the normality assumption on the random effect is uncheckable).

Three models for the dataset on slide 9

(A) $Y_i = \alpha(\text{treat}_i) + \text{error}_i$

- $\alpha(1)$, $\alpha(2)$ are constants (fixed effects).
- Wrong: Effect of animal is ignored.

(B) $Y_i = \alpha(\text{treat}_i) + \beta(\text{animal}_i) + \text{error}_i$

- $\alpha(1)$, $\alpha(2)$ and $\beta(1),\ldots,\ \beta(10)$ are constants (fixed effects).
- Wrong: Meaningless model for testing treatment effect.

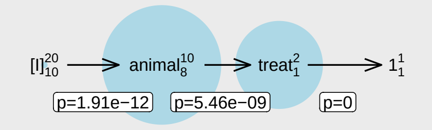
(C) $Y_i = \alpha(\text{treat}_i) + A(\text{animal}_i) + \text{error}_i$

- $\alpha(1)$, $\alpha(2)$ constants (fixed effects).
- $A(1),\ldots,\ A(10)$ independent $\mathcal{N}(0, \sigma_A^2)$ distributed random variables.
- Reasonable: Model with random effect of animal that allows treatment effect estimation.

Design diagrams for the two last models

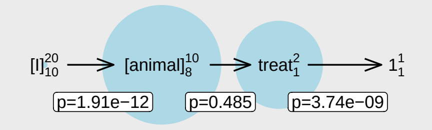
Random factors are designated by square brackets

(B) $Y_i = \alpha(\text{treat}_i) + \beta(\text{animal}_i) + \text{error}_i$ has diagram



Remark: `treat` cannot be tested since `animal` is nested within it.

(C) $Y_i = \alpha(\text{treat}_i) + A(\text{animal}_i) + \text{error}_i$ has diagram



Remark: `treat` is tested against the random factor `animal`.

Rule of thumb

When testing the effect of a factor (here `treat`) any factors nested within it (here `animal`) should usually be modelled as a random effect.

Random effects using R

- Two packages (developed by Bates et al.) available: nlme and lme4
  - nlme also analyses repeated measurements and non-linear models.
  - However, lme4 is maintained by the developers, allows for non-nested random effects, and may also be used for categorical responses.
  - For models without repeated measurements I recommend the lme4-package.

- See also lmerTest-package, which is an add-on to lme4.

- A methodological challenge is that random effects models for continuous responses may be estimated by two different methods.
  - ML (maximum likelihood) is needed if you want to do likelihood-ratio tests (which we will do).
  - REML (restricted maximum likelihood) is recommended for estimation of effects.
  - Luckily lme4 and drop1() automatically take care of this!

- Model validation: Needed for both fixed and random effects.

Checkpoint

- Questions?
- After the break we see how to use a random effect in a so-called split-plot design. This example also exemplifies "Where is the effect?" and shows how to estimate random effects models in R.
- After this, we discuss how to estimate parameters in categorical regressions with random effects.

Time for a break!

Data example 1: Color of pork

- 2 breeds $\begin{cases} \text{old: 10 pigs} \\ \text{new: 10 pigs} \end{cases}$
- After slaughter: 6 pork chops from each pig.
- Storage in light or darkness for 1, 4 or 6 days.
- Response = redness of meat (continuous measurement)

Storage	1 days	4 days	6 days
Dark	chop 1	chop 2	chop 3
Light	chop 4	chop 5	chop 6

- In total $2 \cdot 10 \cdot 6 = 120$ pork chops.

Data example 1: Color of pork – Table of Variables

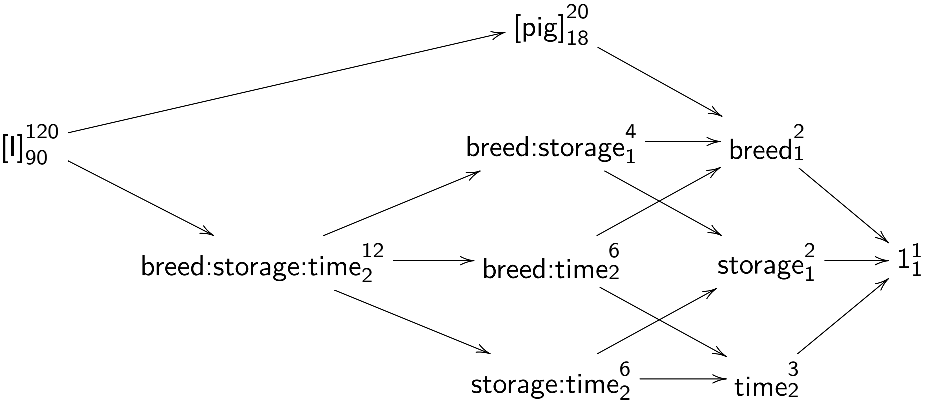
Variable	**Type**	**Range**	**Usage**
breed	nominal	old, new	fixed effect
pig	nominal	$1, \ldots, 20$	random effect
storage	nominal	dark, light	fixed effect
time	ordinal	$1 < 4 < 6$ (days)	fixed effect
redness	continuous	$[1.606; 14.123]$	response

- `pig` is nested within `breed`
  (breed is the “nest” and pigs are the “eggs”).
- `breed` is a between-pig factor.
- `storage` and `time` are within-pig factors.
- Useful factorial design for the fixed effects (include all interactions).

Example is a typical split-plot experiment

- Two types of experimental units:
  - Pigs (“whole-plots”) with the “whole-plot factor” `breed`.
  - Chops (“sub-plots”) with the “sub-plot factors” `storage` and `time`.

Design diagram:



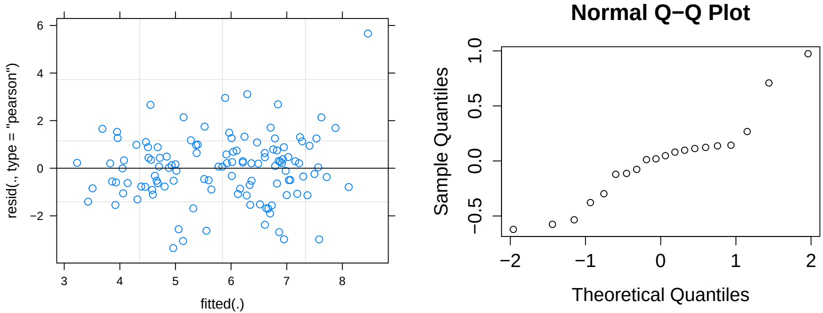
R code: Validation of random effects model – see `redness.R`

$$\text{redness}_i = \alpha(\text{breed}_i, \text{storage}_i, \text{time}_i) + A(\text{pig}_i) + \epsilon_i,\ A(j) \sim \mathcal{N}(0, \sigma_A^2),\ \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

```
# Read dataset: Recode some variables as factors
redness <- read.delim("redness.txt")
redness$pig <- factor(redness$pig)
redness$time <- factor(redness$time)

# Fit random effects model
m0 <- lmer(redness ~ breed * storage * time + (1 | pig), data = redness)

# Residual plot
plot(m0)

# Normal quantile plot for residuals
qqnorm(residuals(m0))

# Normal quantile plot for random effects
qqnorm(ranef(m0)$pig[, 1])
```

Two of the three validation plots

Only showing residual plot and QQ-plot for random effects:



- Residual plot (mean zero? variance homogeneity?): Standardized residuals vs. predicted values (including random effects).
  - Observation no. 44 appears to be an outlier.
- Normal quantile plot for predicted random effects.
  - Note that we only have 20 points, one for each pig.

R code: Backward model reduction – see `redness.R`

In the example using AIC, but we have also asked for *p*-values:

```
# Refit model without obs. no. 44
m1 <- lmer(redness ~ breed * storage * time + (1 | pig), data = redness[-44, ])

# Test and reduce model
drop1(m1, test = "Chisq")
m2 <- update(m1, . ~ . - breed:storage:time)
drop1(m2, test = "Chisq")
m3 <- update(m2, . ~ . - breed:time)
drop1(m3, test = "Chisq")
m4 <- update(m3, . ~ . - breed:storage)
drop1(m4, test = "Chisq")
```

- Syntax for lmer():
  - Random effects specified in terms $(1|\cdot)$.
  - Other things done as for lm().
- Technicality: Tests done as chi-squared test on likelihood ratio.
- Automated backward model selection using *p*-values via `step()` function in `lmerTest`-package.

Results for the final model

Final model:

$\text{redness}_i = \alpha(\text{breed}_i) + \beta(\text{storage}_i, \text{time}_i) + A(\text{pig}_i) + \epsilon_i,\ A(j) \sim \mathcal{N}(0, \sigma_A^2),\ \epsilon_i \sim \mathcal{N}(0, \sigma^2)$

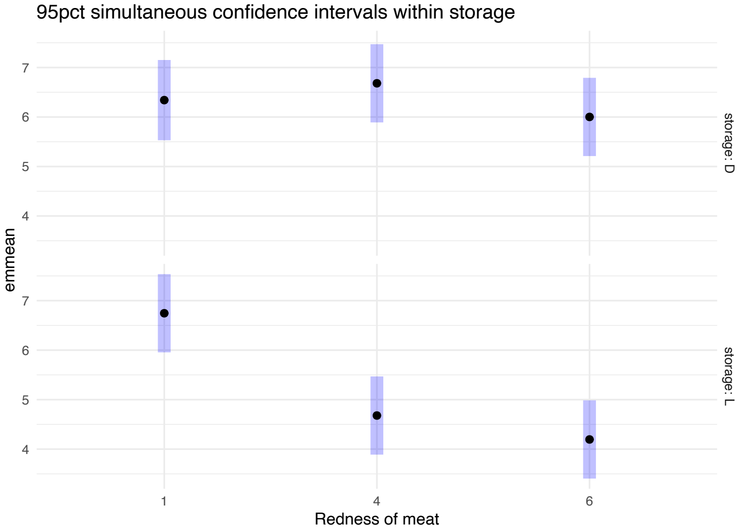
- Likelihood-ratio test in model ‘m4’:
  - for `breed`, $p = 0.0141$ and $\text{df} = 1$,
  - for `storage:time`, $p = 0.0001$ and $\text{df} = 2$.
- The old breed has more redness that the new breed:

$$\texttt{emmean(old)} - \texttt{emmean(new)} = 0.8116 \quad (95\%\text{CI} : [0.1317; 1.4915])$$

- Estimated marginal means for combination of storage (D, L) and time (1, 4, 6) are found and displayed (see next slide) using the R code:

```
plot(emmeans(m4, ~ time | storage), int.adjust = "tukey", horizontal = FALSE) +
ylab("Redness of meat") +
ggtitle("95pct simultaneous confidence intervals within storage")
```

Meat fades over time (when stored in light)



Quantification of sources of variation

In some analyses this quantification is the main objective:

```
> summary(m4)
Linear mixed model fit by REML
Formula: y ~ breed + storage*time + (1 | pig)
...
Random effects:
 Groups   Name        Variance Std.Dev.
 pig      (Intercept) 0.200    0.4473
 Residual             1.923    1.3866
Number of obs: 119, groups:  pig, 20
...
```

- Total variance $= 0.200 + 1.923 = 2.123$.
  - Of this 9% is due to variation between pigs (random effect of pig).
  - And 91% comes from other sources (residual).

Data example 2: Binary data

Number of Orobanche seeds germinating (yes/no) in extracts of bean and cucumber roots. In total 21 batches of two different orobanche varieties as shown below.

O. 75				O. 73			
Bean		Cucumber		Bean		Cucumber	
yes	no	yes	no	yes	no	yes	no
10	29	5	1	8	8	3	9
23	39	53	21	10	20	22	19
23	58	55	17	8	20	15	15
26	25	32	19	23	22	32	19
17	22	46	33	0	4	3	4
		10	3				

Orobanche purpurea (Yarrow Broomrape),

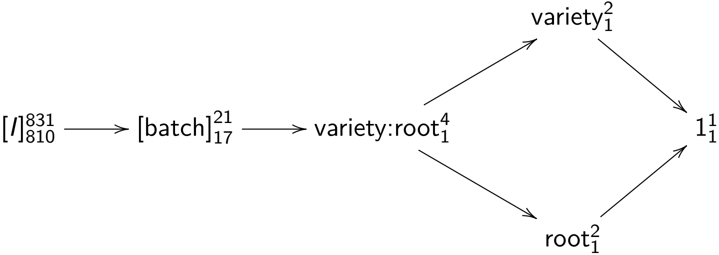
picture from Wikipedia

We see e.g. that the first batch consisted of 39 seeds of variety O.75 grown in bean roots, out of which 10 seeds germinated.

Data example 2: Germination of Orobanche seeds

Variable	**Type**	**Range**	**Usage**
`variety`	nominal	O.75, O.73	fixed effect
`root`	nominal	bean, cucumber	fixed effect
`batch`	nominal	1, . . . , 21	random effect
`germination`	binary	yes, no	response

Interested in the effect of `variety` and `root` on germination of the seeds. The particular `batch` is not of interest, but representative of a population.



Logistic regression with random effects

- Logistic regression models probability of germination via

  $$\log(\text{odds for germination of i'th seed}) = \alpha(\text{variety}_i, \text{root}_i)$$

  But what if the batches are different?

- A solution is to include `batch` as a random effect. In this model, for independent $B(1), \ldots, B(21) \sim \mathcal{N}(0, \sigma_B^2)$ we have

  $$\log(\text{odds for germination of i'th seed}) = \alpha(\text{variety}_i, \text{root}_i) + B(\text{batch}_i)$$

  The additional variability induced by the random factor is called overdispersion.

- Alternatives to random effects models:
  - Correct for overdispersion by rescaling the standard errors: done using quasibinomial-family in glm().
  - Estimate empirical correlation using the gee-approach.

Checkpoint

- Questions?
- After the break we continue to analyze the dataset using a logistic mixed model.
- After this we discuss models for repeated measurements.

Time for a break!

R code: Validation of random effects model – see `orobanche.R`

$\log(\frac{p_i}{1-p_i}) = \alpha(\text{variety}_i, \text{root}_i) + B(\text{batch}_i)$, where $B(j) \sim \mathcal{N}(0, \sigma_B^2)$

```
 # Fit GLMM (generalized linear mixed effects model)
 germination       <- read.delim("orobanche.txt")
 germination$batch <- factor(germination$batch)
 m1 <- glmer(cbind(yes, no) ~ variety * root + (1 | batch),
 family = binomial(link = "logit"), data = germination
)

 # Validate normality of random effects
 qqnorm(ranef(m1)$batch[, 1])
```

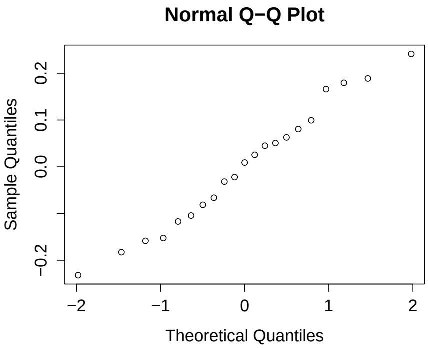
Remarks on R code from the preceding slide

- Model fitted by `glmer()`.
  - Random effects specified in terms $(1|\cdot)$.
  - As in glm() you should remember to specify distribution-family, possibly with link-function. Syntax as for `glm()`.
  - For non-normal responses there is no concept of REML.

- Normality of random effects validated by normal quantile plot.
- We did not validate model specification as the predictors are categorical, this model cannot be misspecified!

- Analysis proceed as usual:
  - Hypothesis tests using `drop1(.,test="Chisq")`.
  - Estimates, em-means, and post hoc test as usual. Remember, that these are on the scale of the link function (e.g. logit).
  - Estimates and confidence limits may be backtransformed.

Validation plot for the GLMM



- Normal quantile plot for predicted random effects.
- Plot looks good!

Data example 2 continued – Comparison with other approaches

Logistic regression model	**Test of variety:root**	**Overdispersion**
GLMM (what we did)	$p = 0.0413$	modelled
GEE	$p = 0.0374$	scale=1.8618
quasibinomial	$p = 0.0636$	scale=1.8618
Plain	$p = 0.0114$	not modelled

- In this example plain logistic regression is questionable, and $p = 0.0114$ is not trustworthy.
- In practice all the 3 other methods can be used.
    - While they disagree on the significance here, this is not a paradox, but simply different tests for the same hypothesis.
- To see if your data is overdispersed, check whether the scale parameter in the quasibinomial analysis is (significantly) larger than 1.
    - Alternatively, perform hypothesis testing on the random effect in the GLMM.

Summary of random effects

- A factor should be used as a random effect if it is...
  - non-reproducible,
  - of no interest beyond the present experiment,
  - representative of a broader population.
- Often block factors are used as random effects.
- Rule of thumb: To test a fixed effect any factor nested within it should usually be modelled as a random effect.
- In these lectures we only discussed models with random intercepts. However, we may also have models with random slopes of some continuous covariate. For instance, if `redness$time` is continuous:

  ```
  lmer(y ~ breed * storage * time + (1 + time | pig), data = redness)
  ```
- Random effects also possible for categorical regression models:
  - Related to overdispersion.
  - GLMM (generalized linear mixed effects models) used in these lectures.

Repeated measurement models: why?

- Data example 1: Color of pig meat.
- 10 pigs from both old and new breed: 7 chops from each pig.

Storage	Time			
	0 days	1 days	4 days	6 days
Dark	chop 1: data not used	chop 2	chop 3	chop 4
Light		chop 5	chop 6	chop 7

Random effect model for chops 2 to 7 (in total $2 \cdot 10 \cdot 6 = 120$ observations):

$$\text{redness}_i = \alpha(\text{storage}_i, \text{time}_i, \text{breed}_i) + \underbrace{A(\text{pig}_i)}_{\sim\mathcal{N}(0,\sigma_A^2)} + \underbrace{\epsilon_i}_{\sim\mathcal{N}(0,\sigma^2)}$$

- But what if the experimental design had been like this?

Storage	Time			
	0 days	1 days	4 days	6 days
Dark	chop 1	chop 1	chop 1	chop 1
Light	chop 2	chop 2	chop 2	chop 2

Example continued

Alternative design has 4 measurements for each pork chop

Storage	Time			
	0 days	1 days	4 days	6 days
Dark	chop 1	chop 1	chop 1	chop 1
Light	chop 2	chop 2	chop 2	chop 2

Random effect model for the alternative design:

$$\text{redness}_i = \alpha(\text{storage}_i, \text{time}_i, \text{breed}_i) + \underbrace{A(\text{pig}_i)}_{\sim\mathcal{N}(0,\sigma_A^2)} + \underbrace{B(\text{pig}_i, \text{chop}_i)}_{\sim\mathcal{N}(0,\sigma_B^2)} + \underbrace{\epsilon_i}_{\sim\mathcal{N}(0,\sigma^2)}$$

Here the 4 measurements on the same pork chop (from the same pig) share an additional random effect B(pig,chop).

- But perhaps measurements taken close in time are more correlated than measurements taken far apart in time! How to model that?

General remarks on repeated measurements

- Repeated measurements originate from study designs where the experimental units have been measured several times (typically at different time points or at different spatial positions):
  - "Economic" necessity, e.g. when experimental units are expensive.
  - Experimental units may serve as their own controls.
  - Response profile (i.e. response over time) is of scientific interest.
  - Repeated measurements are analyzed either to gain power or to investigate the response profile.

- A summary measure is a single number capturing the important feature of the response profile.
  - Examples: AUC (area under the curve), mean, maximum, minimum, range between max and min, time under a pre-specified level, slope, curvature, halving time, slope after the minimum, . . .
  - Summary measure preferably suggested from the scientific study, not from the statistical analysis.
  - Summary measures reduce the repeated measurements to a single observation $\Longrightarrow$ statistical analysis without repeated measurements.

Summary measures vs. Random effects vs. Repeated measurements

- Summary measures are always an option.
  - Unless you have particular interest in the response profile analysis of summary measures is recommended (if it has sufficient power).

- With few repeated measurements per subject, say 4 or less, it does not make sense to estimate the serial correlation structure.
  - Use a random effect model.

- With many repeated measurements per subject, say 5 or more, you have enough information to estimate the serial correlation structure.
  - This is necessary to have trustworthy *p*-values and confidence intervals.
  - However, sometimes it is possible to model serial correlation via random slopes. But we will not investigate this further in this lecture.

Case study: Growth of Baobab trees under water stress

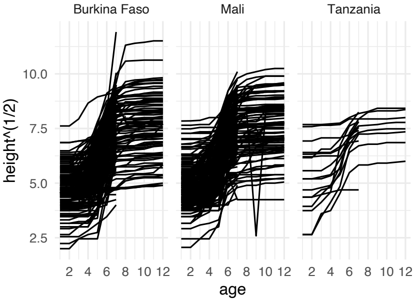
Data kindly provided by Henri-Noël Bouda

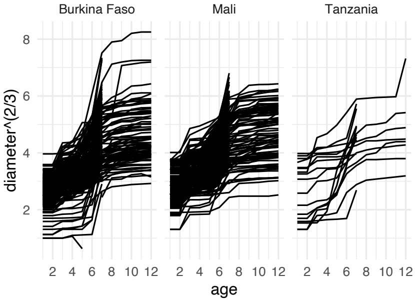
- Baobab seeds from 7 locations in 3 countries sown in the beginning of 2009.
- Plants grown under 3 water regimes (100%, 75% and 50% field capacity).
- Diameter and height of plants measured monthly.
- To measure root weight etc. some plants were harvested in August 2009, some plants in February 2010.
- Purpose of experiment:
  - How does water drought effect growth of trees?
  - Is there an interaction with country and/or origin?

Individual response profiles (subject profiles)

362 baobab trees that survived until harvest

- A helpful plot to construct. Gives overview of the data and provides an impression of the "typical" time-response relationship.

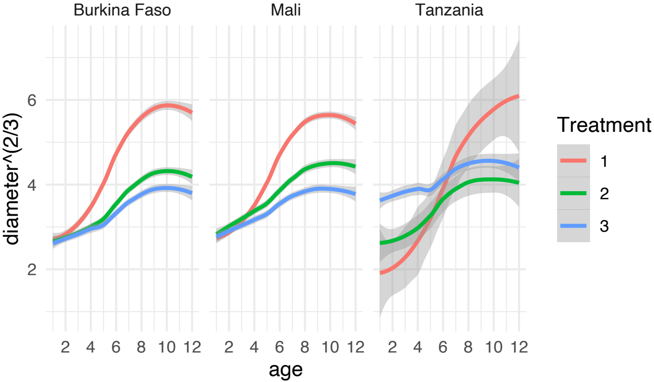




- Here the power transformations **height** $\rightsquigarrow$ **height**$^{1/2}$ and **diameter** $\rightsquigarrow$ **diameter**$^{2/3}$ were chosen from a Box-Cox analysis.

Average response profiles

Another helpful plot. Gives overview of the treatment effects.



Diameter response profiles averaged within the 9 combinations of treatments and countries

Data organization in Excel sheet

Wide form (also called "horizontal organization") of responses: diameter, height

Variable	**Levels**	**Description**
Country	3 (Burkina Faso, Mali, Tanzania)	Country
Origin	7 (Kolangal, . . . , Samé)	3 locations in Burkina Faso, 3 in Mali, 1 in Tanzania
Plant	362 (BKol-04, . . . , TNku-76)	Plant ID
Block	3 (1, 2, 3)	Field blocks
Treatment	3 (1, 2, 3)	Water regime
HarvestDate	3 (aug-09, feb-10, missing)	Day of harvest
Dia0209	continuous (or missing)	Diameter, February 2009
⋮	⋮	⋮
Dia0110	continuous (or missing)	Diameter, January 2010
Hei0209	continuous (or missing)	Height, February 2009
⋮	⋮	⋮
Hei0110	continuous (or missing)	Height, January 2010

Long form: diameter, height for each month

Variable	**Levels**	**Description**
Country	3 (Burkina, Mali, Tanzania)	Country
Origin	7 (Kolangal, . . . , Samé)	3 locations in Burkina Faso, 3 in Mali, 1 in Tanzania
Plant	362 (BKol-04, . . . , TNku-76)	Plant ID
Block	3 (1, 2, 3)	Field blocks
Treatment	3 (1, 2, 3)	Water regime
HarvestDate	3 (aug-09, feb-10, missing)	Day of harvest
month	12 (2, . . . , 12, 1)	Month of year
year	2 (09, 10)	Year
diameter	continuous	Diameter
height	continuous	Height
age	continuous (1 to 12)	Months since January'09

Long form as needed for the statistical analysis

```
  Country      Origin   Plant   Block Treatment HarvestDate         month year  diameter height   age
   <chr>        <chr>    <fct>   <fct> <fct>     <dttm>              <chr> <chr>    <dbl>  <dbl> <dbl>
 1 Burkina Faso Kolangal BKol-11 1     1         2009-08-01 00:00:00 02    09        4.33     33     1
 2 Burkina Faso Kolangal BKol-11 1     1         2009-08-01 00:00:00 03    09        4.33     33     2
 3 Burkina Faso Kolangal BKol-11 1     1         2009-08-01 00:00:00 04    09        5.62     40     3
 4 Burkina Faso Kolangal BKol-11 1     1         2009-08-01 00:00:00 05    09        5.65     42     4
 5 Burkina Faso Kolangal BKol-11 1     1         2009-08-01 00:00:00 06    09        7.1      52     5
 6 Burkina Faso Kolangal BKol-11 1     1         2009-08-01 00:00:00 07    09       10        54     6
 7 Burkina Faso Kolangal BKol-11 1     1         2009-08-01 00:00:00 08    09       13.4      69     7
 8 Burkina Faso Kolangal BKol-78 1     1         2009-08-01 00:00:00 02    09        4.65     39     1
 9 Burkina Faso Kolangal BKol-78 1     1         2009-08-01 00:00:00 03    09        4.65     39     2
10 Burkina Faso Kolangal BKol-78 1     1         2009-08-01 00:00:00 04    09        6.43     39     3
11 Burkina Faso Kolangal BKol-78 1     1         2009-08-01 00:00:00 05    09        6.5      43     4
12 Burkina Faso Kolangal BKol-78 1     1         2009-08-01 00:00:00 06    09        8        49     5
13 Burkina Faso Kolangal BKol-78 1     1         2009-08-01 00:00:00 07    09        8.6      62     6
14 Burkina Faso Kolangal BKol-78 1     1         2009-08-01 00:00:00 08    09       13.7      75     7
15 Burkina Faso Kolangal BKol-86 1     1         2009-08-01 00:00:00 02    09        6.25     29     1
16 Burkina Faso Kolangal BKol-86 1     1         2009-08-01 00:00:00 03    09        6.25     29     2
17 Burkina Faso Kolangal BKol-86 1     1         2009-08-01 00:00:00 04    09        8.57     31     3
18 Burkina Faso Kolangal BKol-86 1     1         2009-08-01 00:00:00 05    09        8.61     35     4
19 Burkina Faso Kolangal BKol-86 1     1         2009-08-01 00:00:00 06    09        9.1      43     5
...
...
...
```

Table of variables

Variable	**Type**	**Range**	**Usage**
Country	nominal	Burkina Faso, Mali, Tanzania	fixed effect
Origin	nominal	Kolangal, . . . , Samé Remark: nested in Country	fixed effect
Plant	nominal	BKol-04, . . . , TNku-76	random effect subject ID
Block	nominal	1, 2, 3	fixed effect
Treatment	ordinal	$1 < 2 < 3$	fixed effect
age	nominal continuous	1, 2, . . . , 12 $[1; 12]$	fixed effect correlation effect
diameter	continuous	$[0; 23.7]$	response
height	continuous	$[0; 142]$	response

- The two responses (diameter, height) analyzed separately.
- R code via `lme()` in nlme-package:
  - fixed effects are specified in model formula.
  - random effects are specified in random option.
  - correlations are specified in corr option.

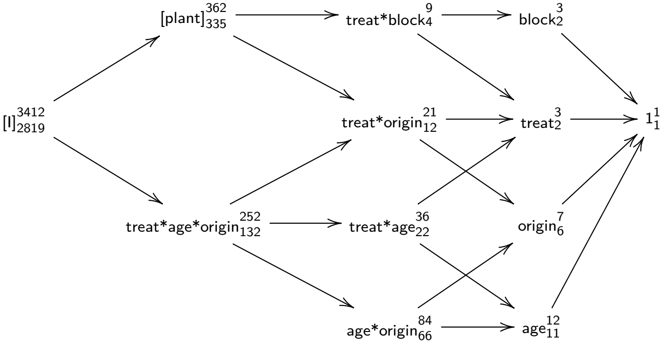
Checkpoint

- Questions?
- After the break, we analyze the baobab dataset using a repeated measurements model.
- After this we discuss analysis of summary measures as an “easy” alternative to repeated measurements models.

Time for a break!

Diagram of fixed and random factors

Not all details included in the diagram, otherwise the plot would be too complex



- In the model reduction `origin` is nested within `country`.
- Are the residuals ϵ_i, $i = 1, \ldots, 3407$, independent?

Repeated measurements model

Remember: Box-Cox analysis suggested to use response $=$ diameter$^{2/3}$

$$\text{diameter}_i^{2/3} = \alpha(\text{treat}_i, \text{age}_i, \text{origin}_i, \text{block}_i) + \underbrace{A(\text{plant}_i)}_{\sim\mathcal{N}(0,\sigma_A^2)} + \underbrace{\underbrace{B(\text{plant}_i, \text{age}_i)}_{\sim\mathcal{N}(0,\sigma_B^2)} + \underbrace{\epsilon_i}_{\sim\mathcal{N}(0,\sigma^2)}}_{\text{[I]-term in the design diagram}}$$

- **Random effect A:** Some plants are bigger than others.
- **Correlated effect B:** Correlated within plants ($\sim$ subject ID). Correlation typically decreases with increasing time distance. Uncorrelated between plants.
  - Possible interpretation is variation between time position of growth period.
- **Error term ϵ:** Possible interpretation is measurement error.

Three examples of correlation structures for B

Model: $\text{diameter}_i^{2/3} = \alpha(\text{treat}_i, \text{age}_i, \text{origin}_i, \text{block}_i) + A(\text{plant}_i) + B(\text{plant}_i, \text{age}_i) + \epsilon_i$

(A) The model without the serial correlated effect B.

- In this case we have a random effect model. This model is also referred to as the random intercept model or the compound symmetry model.

(B) $\text{Var}\Big(B(\text{plant}, \text{age}_i), B(\text{plant}, \text{age}_j)\Big) = \sigma_B^2 \exp\left(-\frac{|\text{age}_i - \text{age}_j|}{d}\right)$

- Correlation has exponential decrease.

(C) $\text{Var}\Big(B(\text{plant}, \text{age}_i), B(\text{plant}, \text{age}_j)\Big) = \sigma_B^2 \exp\left(-\frac{|\text{age}_i - \text{age}_j|^2}{\rho^2}\right)$

- Correlation has Gaussian decrease.
- When a random effect A and an error term ϵ are present, this model is sometimes referred to as the Diggle model after Peter Diggle.

R code – see baobab.R

```
# Transformation of wide-format into long-format
long <- baobab %>%
  pivot_longer(
    cols = Dia0209:Hei0110,
    names_to = c(".value", "month", "year"),
    names_sep = c(3, 5),
    values_drop_na = TRUE
  ) %>%
  mutate(age = (as.numeric(month) - 1) + 12 * (as.numeric(year) - 09)) %>%
  mutate(Treatment = factor(Treatment)) %>%
  rename(diameter = Dia, height = Hei) %>%
  filter(diameter > 0) %>% filter(height > 0)
# Diggle model: two other models in R script
mGauss <- lme(
  diameter^(2 / 3) ~ Treatment * Block + Treatment * factor(age) * Origin,
  random = ~ 1 | Plant,
  corr = corGaus(form = ~ age | Plant, nugget = TRUE),
  data = long, na.action = na.omit
)
```

Which repeated measurements model to use?

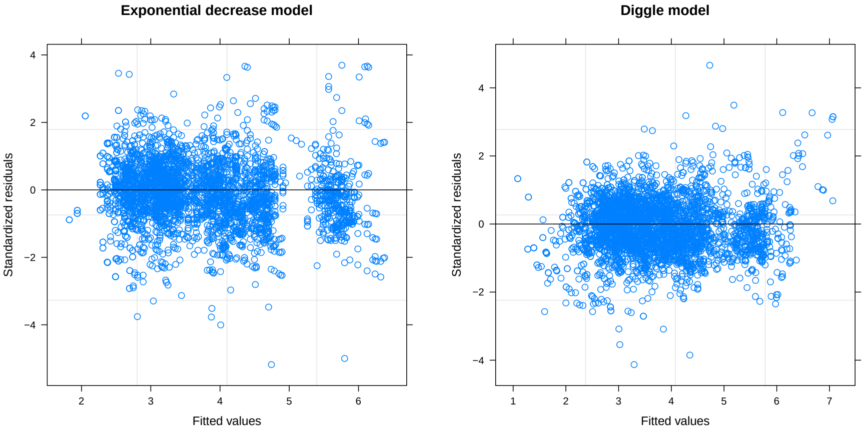
There exist many other models than those listed on slide 50!

- Is the model valid?
  - Residual plot: Non-random scatter suggests that the explanatory variables have not been used appropriately, e.g. an interaction or a quadratic term might be missing.
  - Normal quantile plots: Residuals not on a straight line suggests that the response variable should perhaps be transformed.
  - Semi-variogram: Compares empirical correlation structure (the dots) to the fitted theoretical correlation structure (the line).

- Interpretation?
  - Random effects have a simple interpretation, which speak in favor of the compound symmetry model. The exponential decrease and the Diggle model have similar interpretations.

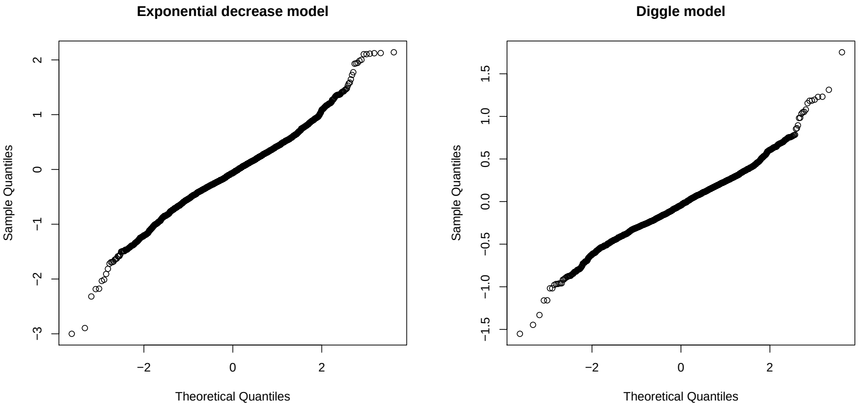
- Akaike Information Criterion (AIC): "The smaller the better".
  - For the Baobab dataset the compound symmetry model is clearly rejected by the AIC.

Exponential vs. Gaussian decrease: Residual plot



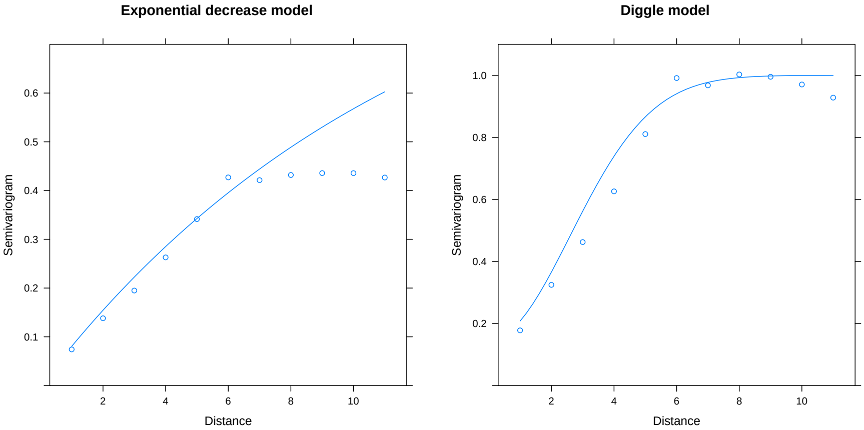
```
plot(mExp, main = "Exponential decrease model")
plot(mGauss, main = "Diggle model")
```

Exponential vs. Gaussian decrease: Normal quantile plot



```
qqnorm(resid(mExp), main = "Exponential decrease model")
qqnorm(resid(mGauss), main = "Diggle model")
```

Semi-variogram: $\gamma(h) = \frac{1}{2}\mathrm{var}\big(X(t+h) - X(t)\big)$



```
plot(Variogram(mExp), ylim = c(0, 0.7))
plot(Variogram(mGauss), ylim = c(0, 1.1))
```

Choice of repeated measurements model

Residual and normal quantile plots acceptable for all models

Model	Compound symmetry	Exponential decrease	Diggle
AIC	3172	1397	1514

- Akaike Information Criterion (AIC) clearly prefers the exponential decrease model, which we describe as:
  "An ANOVA with random effect of plant and residual errors correlated within plants. The errors consist of an independent component and a component with exponential decreasing correlation. For the fixed effects we used the concatenation of the full factorial design of (treatment, age, origin) and the full factorial design of (treatment, block)".

Overview of steps in a repeated measurements analysis

- List and classify the variables in the design.
  - Done (see slide 46).

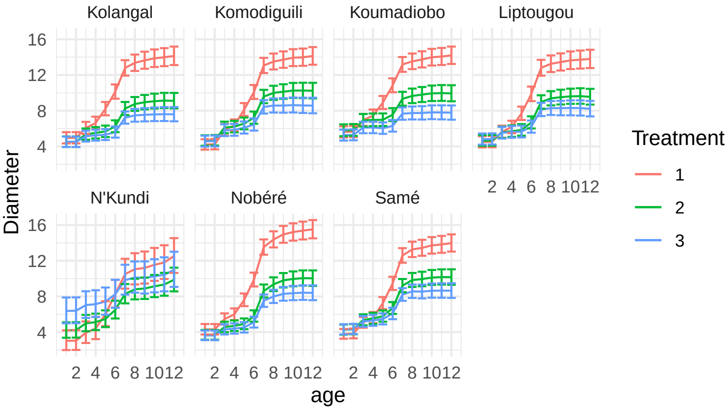
- Make plots of individual, and perhaps averaged, response profiles.
  - Done (see slides 41 and 42).

- Choose and validate a correlation structure.
  - Done (see slides 53 to 56).

- Test on fixed effects: Done as usual for ANOVA and ANCOVA models.
  - Remember to refit models using maximum likelihood (`method="ML"`).
  - Unfortunately the `drop1()` function does not work for lme-objects so this has to be done by hand (see Section 9.5.3 of the book).
  - Automatic model selection based on AIC may be done using `stepAIC()` from the MASS-package.

- Report estimates and conclusions from the final model (as usual, e.g. using the `emmeans`-package as seen on next slide).

Visualization of estimated marginal means



See `baobab.R` for code.

Analysis of Summary measures – alternative to repeated measurements

- Idea:
  - Reduce the curve for each subject to a single value.
  - Analyze this summary measure as usual (ANOVA, regression, . . .).

- As summary measures we could for example use:
  - Average response over time.
  - Area under curve (AUC, often used in medicine).
  - Slope of curve (rate of increase).
  - Maximal response.
  - Position (e.g. time) of maximal response.
  - Halving time since maximal response.
  - Curvature: fit $\alpha + \beta \cdot \text{time} + \gamma \cdot \text{time}^2$ for each individual and use $\hat{\gamma}$.

Note: The summary measures should be computed for each subject — not on the average profiles!

Principles for choosing summary measures

- Select a measure that addresses the problem under investigation.
- Do not choose summary measures on the basis of visual inspection of the treatment differences — this is cheating.
  - But it is OK to plot all profiles in one graph and select “typical features” of the curves for further investigation.
- You may analyze more than one summary measure. If so, then choose some that reflect different aspects of the curves. For example:
  - AUC and average response is NOT a good combination.
  - AUC and rate of increase might be a good combination.

  But be aware of the associated multiple testing problem.

Analysis of summary measures: pros and cons

- **Advantages:**
  - Simple analysis, which is more easily communicated.
  - Often powerful analysis if the summary measure is chosen appropriately.
  - Model validation easier and more transparent.

- **Disadvantages:**
  - Each curve is reduced to a single value — loss of information?
  - Which summary measure should be used?
  - No investigation of the “temporal” structure, which might be important for the problem under investigation.

See `baobab.R` for an example of such an analysis.

Summary

- We saw how dependence between individuals in a regression setting can be accounted for my including random effects.
- We also saw how random effects can be used to model overdispersion.
- Finally, we saw how to analyze repeated measures either by trying to model response curves or by looking at summary measures.
- Design diagrams may be a useful tool to understand the structure of the explanatory variables in a design and in particular for determining which hypotheses are testable.

Merry Christmas and happy new year!