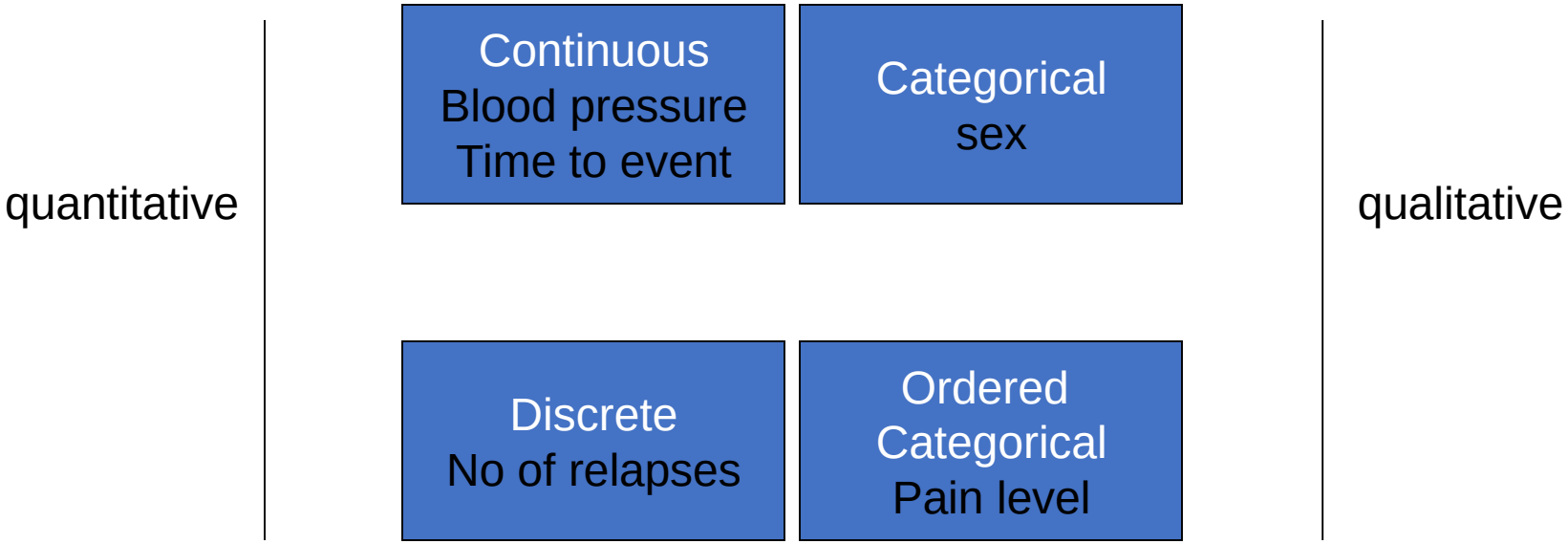
Lecture 8

Continuous Data

Types of Data



Types of data analysis (Inference)

Parametric
Vs
Non parametric

Model based
Vs
Data driven

Frequentist
Vs
Bayesian

One way analysis of variance (Anova1)

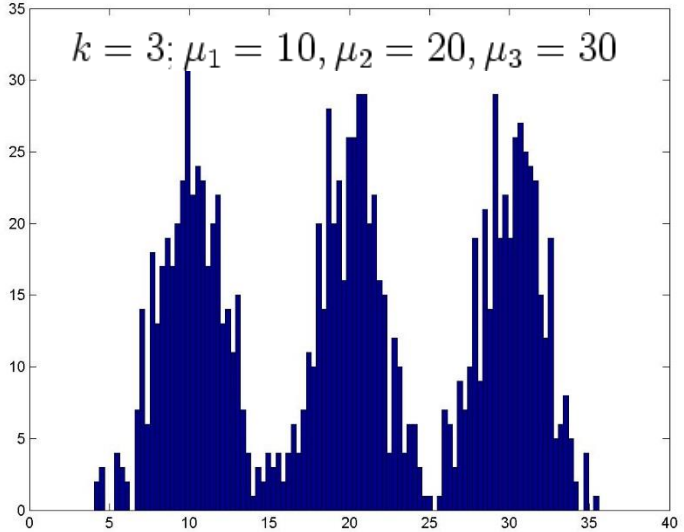
- Here we want to compare several (*k*) groups (doses) with respect to the average response in each group. The following model is assumed

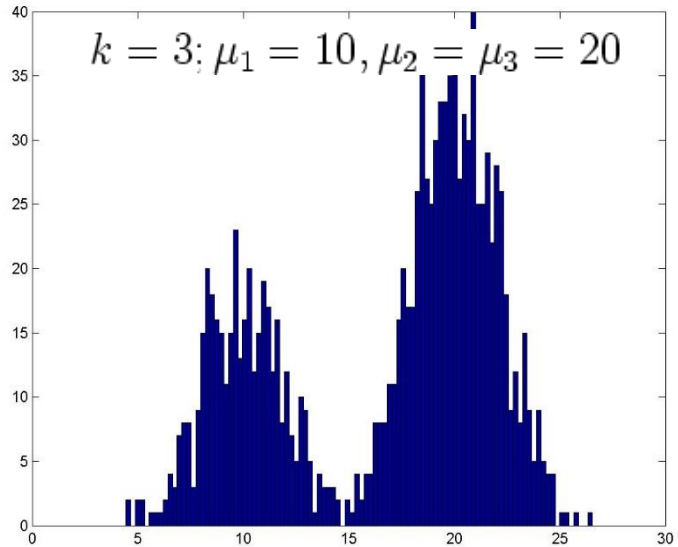
$$Y_{ij} = \mu_i + e_{ij} = \mu + \tau_i + e_{ij} \quad \text{and} \quad Var\left[e_{ij}\right] = \sigma^2$$

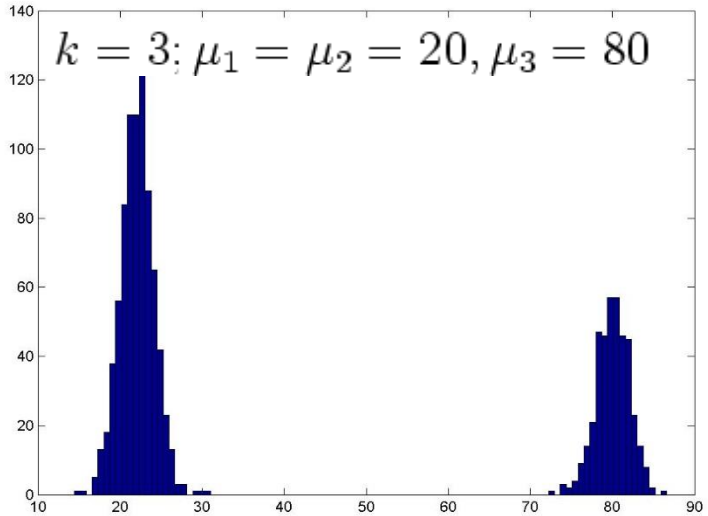
- Main point: individuals only differ with respect to one criterion: group belonging

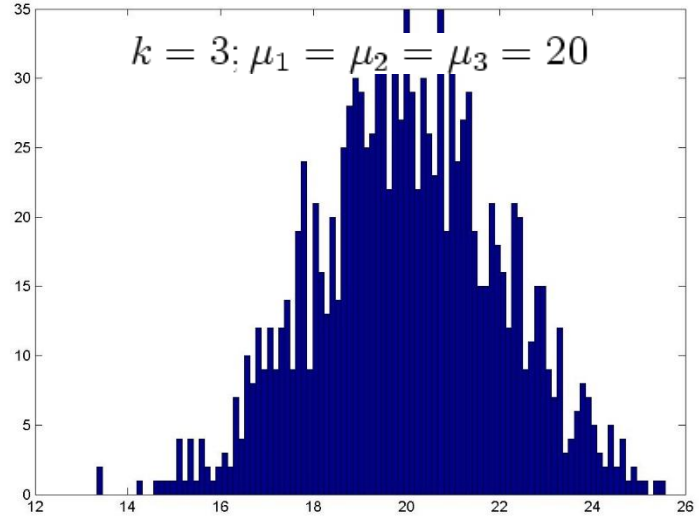
$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$
$$H_a : \mu_i \neq \mu_j, i \neq j$$

$$\begin{array}{cccc} Y_{11} & Y_{12} & \ldots & Y_{1n_1} \\ Y_{21} & Y_{22} & \ldots & Y_{2n_2} \\ \ldots, & \ldots & \ldots & \\ Y_{k1} & Y_{k2,} & \ldots & Y_{kn_k} \end{array}$$









Looking at the pictures we see that the variation says something about the difference between the cases where the groups are different and when they are similar.

Anova 1 continued

- The null hypothesis and the alternative hypothesis can be formulated as

$$H_0 : \mu_1 = \mu_2 = ... = \mu_k$$

$$H_a : \mu_i \neq \mu_j, i \neq j$$

- The deviation of an individual observation from the overall mean can be described by

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{..})^2 = \sum_{i=1}^{k} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i.})^2$$

SST/SST SSA/SSB SSE/SSW

Anova1 continued

- Manipulating these gives:

$$E[\frac{SSE}{n-k}] = \sigma^2, E[\frac{SSA}{k-1}] = \sigma^2 + \frac{1}{k-1}\sum_{i=1}^{k}(\mu_i - \mu)^2$$

MSE MSA

- When the null hypothesis is false we expect the ratio MSA/MSE to be large. To calculate exact significance levels we use the fact that it F-distributed with (k-1,N-k) d.f.

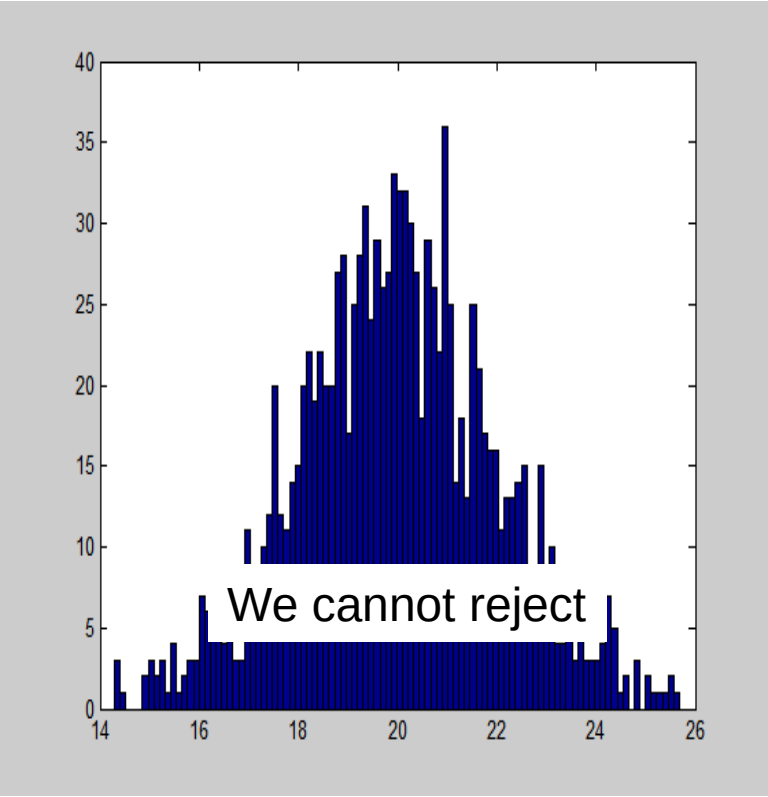
Anova table

Source of variation	**Sum of squares**	**d.f.**	**Mean squares**	**F/p-value**
Beween	SSA	K-1	MSA	F_0=MSA/MSE /p
Within	SSE	N-k	MSE	
Total	SST	N-1		

- Example: 3 groups with the same mean 20 and the same standard deviation 2. We reject for large values of F_0.

$$p = P(F > F_0) \quad \Big| \quad 0.29 = P(F > 1.22),$$

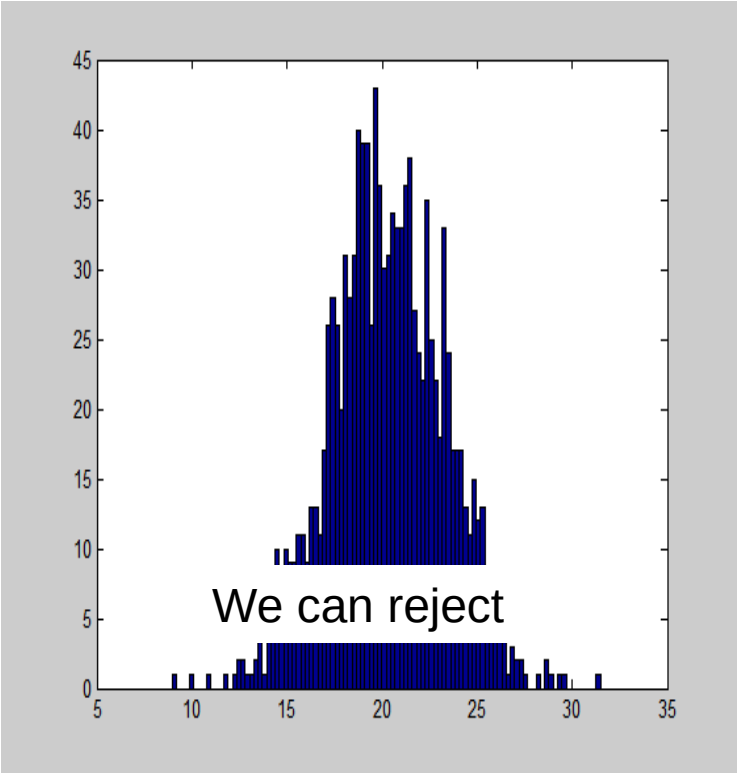
F is F-distributed and F_0 the value of MSA/MSE



Source of variation	Sum of squares	d.f.	Mean squares	F_0/p-value
Treatment	9.56	2	4.78	1.22/ 0.29
Error	4671.1	1197	3.9	
Total	4680.65	1199		

- Example: 3 groups with means 20, 20 and 20.5 and the same standard deviation 3.

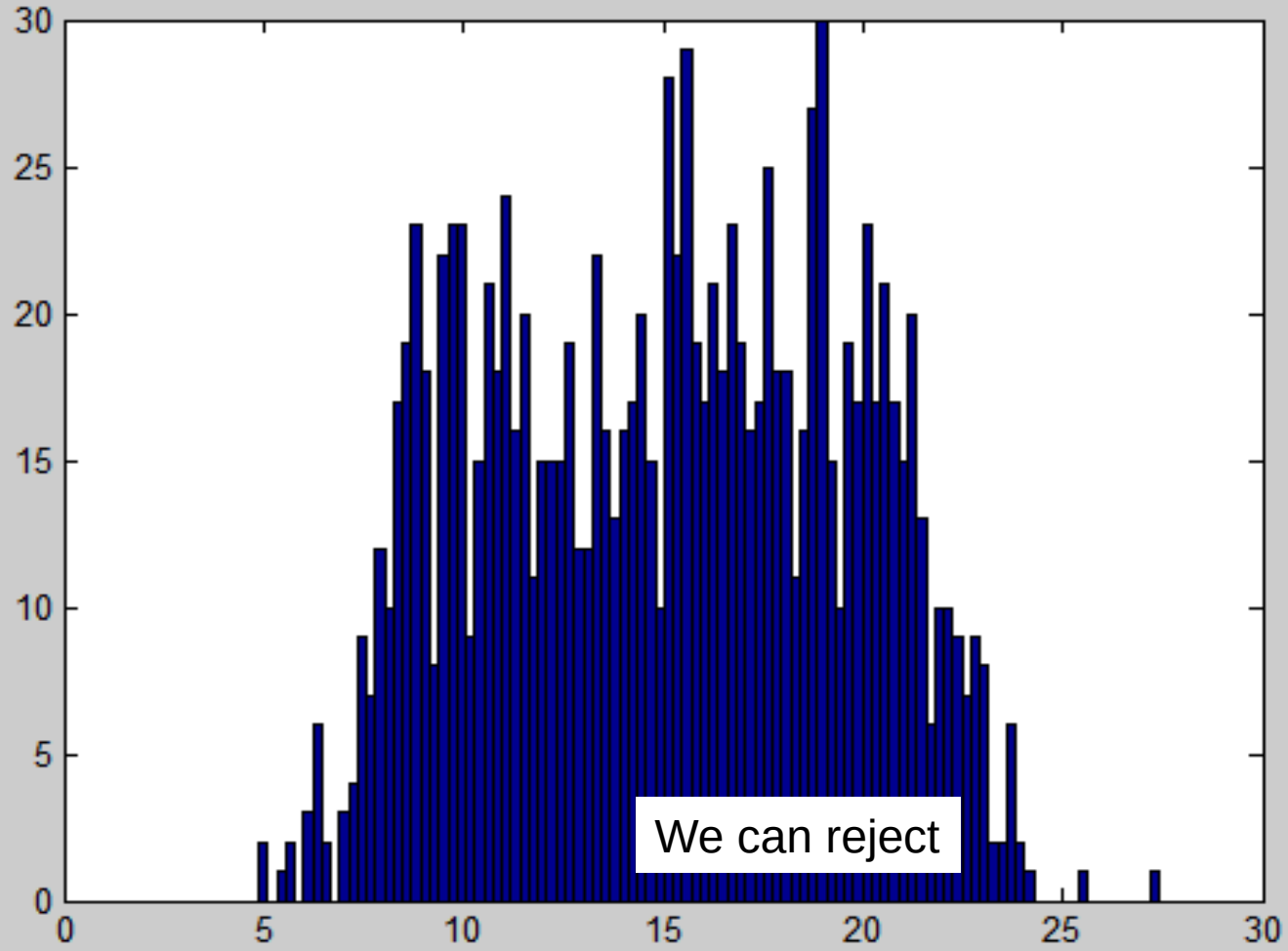
$$0.039 = P(F > 3.25),$$



Source of variation	Sum of squares	d.f.	Mean squares	F/p-value
Treatment	59.4	2	29.7	3.25/ 0.039
Error	10942.8	1197	9.14	
Total	11002.3	1199		

ANOVA Table

Source	SS	df	MS	F	Prob>F
Groups	19117.5	2	9558.74	2477.71	0
Error	4617.9	1197	3.86		
Total	23735.4	1199			



Multiple comparisons

- When we reject the null hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$ we only know that the groups are not equal but some of them might still be.
- To find out more, we have to consider all the pairwise comparisons between the groups. With k groups this gives m=k(k-1)/2 such comparisons.
- How do we do this and still have a reasonable overall significance level?
- The simplest way to deal with this is using Bonferroni's inequality. This implies that when performing m tests if each test is at the 1-α/m level then the tests taken simultaneously will be on the 1- α level.
- We will deal with problem in detail later.

Two way analysis of variance (anova 2)

- Here we assume that *individuals can differ with respect to two factors* (e.g. two drugs, treatment and centre). The following model is usually suitable for this situation

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + e_{ijk}$$

- As before the deviation of an individual observation from the overall average can be decomposed into terms related to the effects of the factors (e.g. treatment, center, treatment by center as well as to a pure random component.

Anova2

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk}-\bar{Y}_{...})^2 = bn\sum_{i=1}^{a} n_i(\bar{Y}_{i..}-\bar{Y}_{...})^2 +$$

$$an\sum_{j=1}^{b}(Y_{.j.}-\bar{Y}_{...})^2 + n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...})^2 +$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk}-\bar{Y}_{ij.})^2$$

- We use a similar notation as before

$$SST = SSA + SSC + SS(AC) + SSE$$

$$E[MSA] = E\frac{SSA}{a-1} = \sigma^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$$

$$E[MSC] = E\frac{SSC}{b-1} = \sigma^2 + \frac{an\sum_{j=1}^{b}\beta_j^2}{a-1}$$

$$Var[e_{ijk}] = \sigma^2$$

$$E[MS(AC)] = E\frac{SS(AC)}{(a-1)(b-1)} = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\tau\beta)_{ij}^2}{(a-1)(b-1)}$$

$$E[MSE] = E\frac{SSE}{ab(n-1)} = \sigma^2$$

- Tests of treatment effect, center effect or treatment by center interaction can be performed by using the appropriate ratio between mean square values.

Anova2

- As an example assume we want to test if there is a treatment effect: $H_0 : \tau_i = 0$
- This can be tested using the test statistic

$$F_1 = \frac{MSA}{MSE}$$

- Which has under the null hypothesis an F distribution with (a-1) and ab(n-1) degrees of freedom.
- The various tests are summarized in the following table

Two-way anova

$Variation$	SS	df	MS	F
$Treatment$	SSA	$a-1$	MSA	$F_1 = \frac{MSA}{MSE}$
$Center$	SSC	$b-1$	MSC	$F_2 = \frac{MSC}{MSE}$
$Treatment/Center$	$SS(AC)$	$(a-1)(b-1)$	$MS(AC)$	$F_3 = \frac{MS(AC)}{MSE}$
$Error$	SSE	$ab(n-1)$	MSE	
$Total$	SST	$abn-1$	MST	

Analysis of covariance

- Here we assume that individuals can differ with respect baseline values. It is sometimes desirable to adjust the model for the endpoint measures Y so baseline values X are taken into account. The following model can then be useful

$$Y_{ij} = \mu + \tau_i + \beta X_{ij} + e_{ij}$$

The analysis uses an F distribution based on Mean sums of squares (cf. Table 8.4.5)

Table 8.4.5 Analysis of Variance for the Two-Way Classification Fixed Model

Source of Variation	Sum of Squares	*df*	Mean Squares	*F*
Treatment	$\text{SSA} = bn\sum_{i=1}^{a}(\bar{Y}_{i..} - \bar{Y}_{...})^2$	$a - 1$	$\text{MSA} = \dfrac{\text{SSA}}{(a-1)}$	$F_1 = \dfrac{\text{MSA}}{\text{MSE}}$
Center	$\text{SSC} = an\sum_{j=1}^{b}(\bar{Y}_{.j.} - \bar{Y}_{...})^2$	$b - 1$	$\text{MSC} = \dfrac{\text{SSC}}{b-1}$	$F_2 = \dfrac{\text{MSC}}{\text{MSE}}$
Treatment *center	$\text{SS}(AC) = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$	$(a-1)(b-1)$	$\text{MS}(AC) = \dfrac{\text{SS}(AC)}{(a-1)(b-1)}$	$F_3 = \dfrac{\text{MS}(AC)}{\text{MSE}}$
Error	$\text{SSE} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \bar{Y}_{ij.})^2$	$ab(n-1)$	$\text{MSE} = \dfrac{\text{SSE}}{ab(n-1)}$	
Total	$\text{SST} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(Y_{ijk} - \bar{Y}_{...})^2$	$abn - 1$		

Example: Blood pressure

Testing a drug that lowers the blood pressure on patients with hypertension

- This study is a 8-week, multicenter, randomised, double blind, 3-arm, parallel group, efficacy and safety study, in patients with moderate to severe hypertension.

Primary objective:

- To compare sitting diastolic blood pressure (DBP) lowering effect of hypersartan 16 mg with that of Placebo

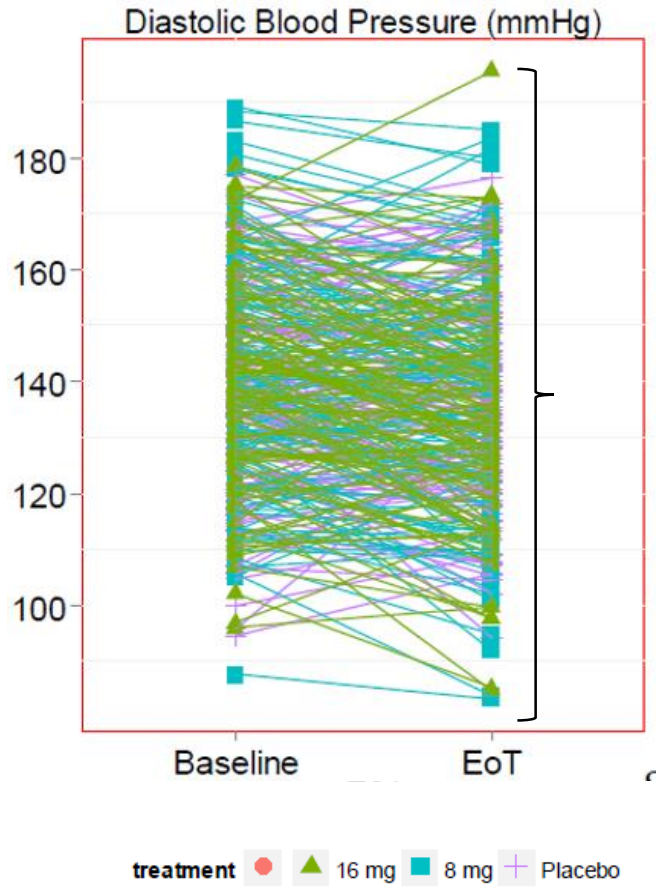
Primary variable:

- Sitting DBP

Example: Blood pressure

How to estimate and test treatment differences?

- Compare BP at End of Treatment (EoT) by ANOVA



$$EoT_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

overall mean

treatment effect
i = 1,2,3
{16 mg, 8 mg, Placebo}

$\varepsilon_{ij} \sim NID(0,\sigma^2)$

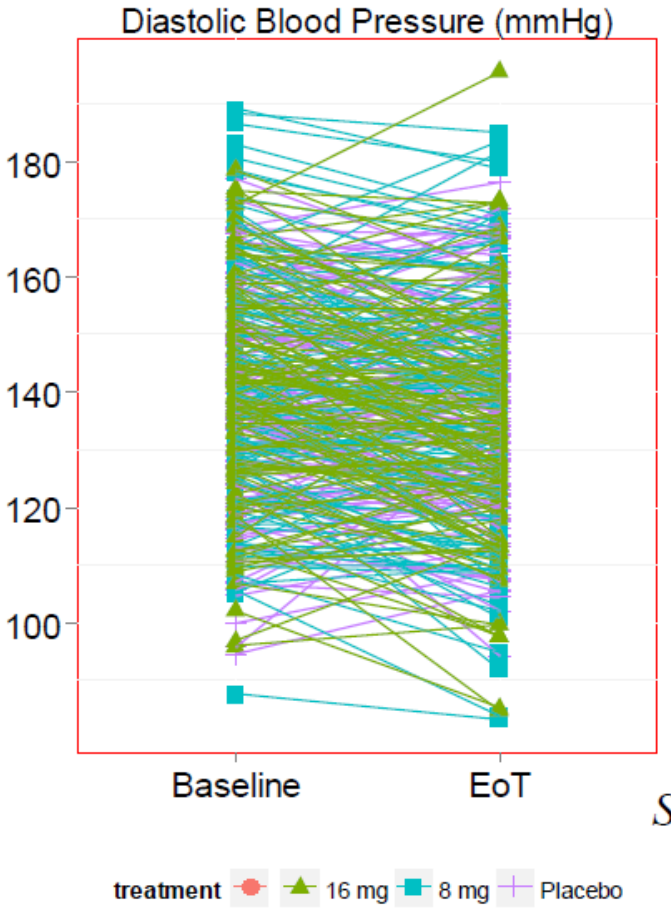
Model	F-test p-value	16mg – Placebo Mean (p-value)	8mg – Placebo Mean (p-value)	16mg – 8mg Mean (p-value)	Residual SD
ANOVA EoT	0.55	-2.7 (0.29)	-0.7 (0.79)	-2.0 (0.43)	19.7

$$SD_{EoT} = 19.7$$

Example: Blood pressure

How to estimate and test treatment differences?

- Compare BP change from baseline to EoT by ANOVA



$$(EoT\text{-}Baseline)_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

overall mean

treatment effect
i = 1,2,3
{16 mg, 8 mg, Placebo}

$\varepsilon_{ij} \sim NID(0,\sigma^2)$

Model	F-test p-value	16mg – Placebo Mean (p-value)	8mg – Placebo Mean (p-value)	16mg – 8mg Mean (p-value)	Residual SD
ANOVA Diff	0.03	-4.6 (0.01)	-3.5 (0.049)	-1.1 (0.54)	13.7

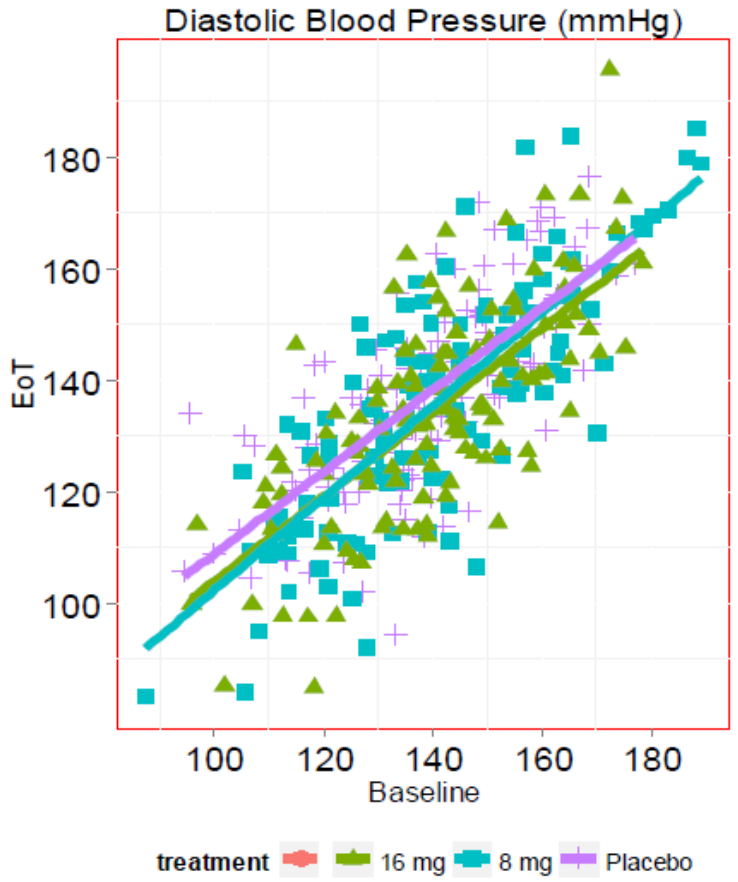
$$SD_{EoT-Baseline} = SD\sqrt{2\cdot(1-\rho)} = 19.7\sqrt{2\cdot(1-0.75)} \approx 13.7$$

Estimated correlation between EoT and baseline, ρ =0.75

Example: Blood pressure

How to estimate and test treatment differences?

- Compare BP change from baseline to EoT by ANCOVA



$$(EoT)_{ij} = \mu + BL + \tau_i + + \varepsilon_{ij}$$

Baseline covariate

overall mean

treatment effect
i = 1,2,3
{16 mg, 8 mg, Placebo}

$\varepsilon_{ij} \sim NID(0,\sigma^2)$

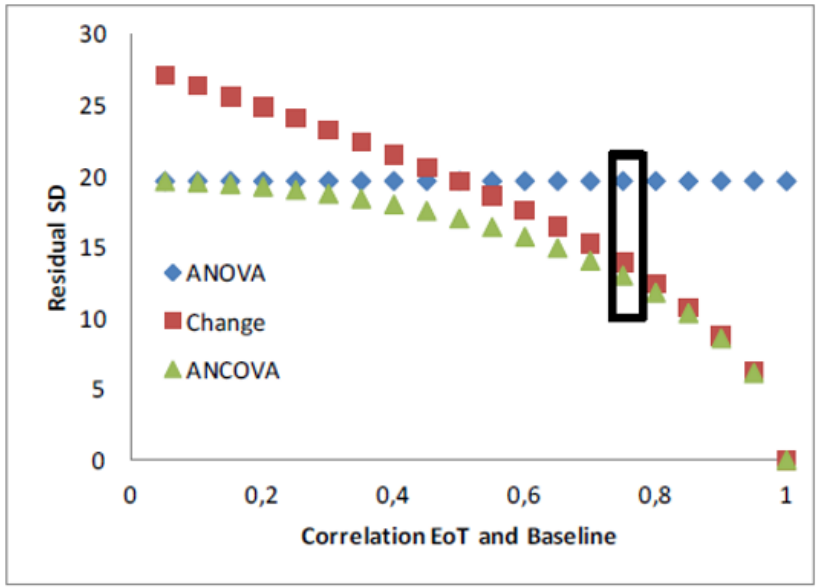
Model	F-test p-value	16mg – Placebo Mean (p-value)	8mg – Placebo Mean (p-value)	16mg – 8mg Mean (p-value)	Residual SD
ANCOVA EoT	0.04 (F-test Baseline: 0.0001)	-4.2 (0.01)	-2.9 (0.09)	-1.3 (0.45)	13.1

$$SD_{EoT-Baseline} = SD\sqrt{1-\rho^2} = 19.7\sqrt{1-0.75^2} \approx 13.1$$

Estimated correlation between EoT and baseline, ρ =0.75

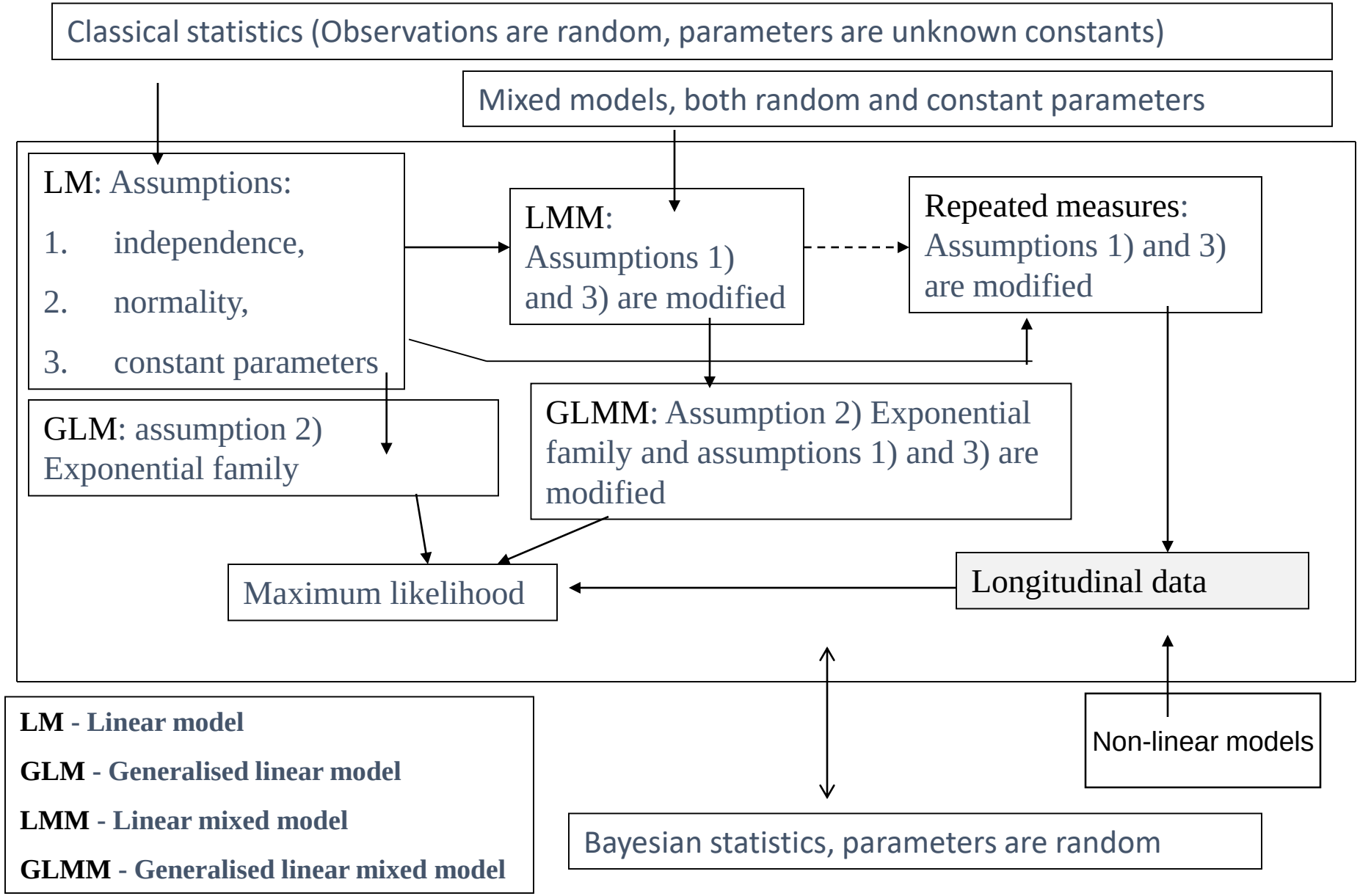
Example: Blood pressure

Testing a drug that lowers the blood pressure



- Certain assumptions required for the models
  - Normality, homoscedasticity, relation baseline-EoT same between treatments, …
- If baseline differ between groups then you´ll get biased estimates. These can be reduced by ANCOVA

Various forms of models and relation between them



Simple linear regression

$$Y_{ij} = \beta_{1i} + \beta_{2i} t_{ij} + \varepsilon_{ij},\ j = 1,..,n_i$$

- Matrix notation:

$$\boldsymbol{Y_i} = Z_i \boldsymbol{\beta_i} + \boldsymbol{\varepsilon_i}$$

- Z_i matrix:

$$Z_i = \begin{pmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ \vdots & \vdots \\ 1 & t_{in_i} \end{pmatrix} \qquad \beta_i = \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \end{bmatrix}$$

The Linear Mixed-effects Model

- The linear mixed effects model is quite flexible and does not need balance, independence etc. Usually some version of maximum l likelihood is used for the inference

$$\boldsymbol{Y}_i = Z_i\boldsymbol{\beta}_i + \boldsymbol{\varepsilon_i}$$

Linear model

$$\boldsymbol{\beta}_i = K_i\boldsymbol{\beta} + \boldsymbol{b_i}$$

Random parameter

$$\Longrightarrow \boldsymbol{Y}_i = Z_iK_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i$$

$$= X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i$$

Average evolution

Subject specific

The general mixed effects models can be summarized by:

$$\begin{cases} \boldsymbol{Y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \\ \boldsymbol{b}_i \sim N(\boldsymbol{0}, D), \\ \boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \Sigma_i), \\ \boldsymbol{b}_1, \ldots, \boldsymbol{b}_N, \boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_N \text{ independent}, \end{cases}$$

Convenient using multivariate normal.
Very difficult with other distributions

Terminology:

- Fixed ffects: β
- Random effects: b_i
- Variance components: elements in D and Σ_i

Remarks

1. It is occasionally unclear if we should treat an effect as a fixed or a mixed effect. For example in clinical trials with treatment and clinic as “factors” ? should we consider clinics as random?

2. Considering the general form of a mixed effects model

$$Y_i = X_i\beta + Z_i b_i + \varepsilon_i$$

notice that the fixed effects are involved only in mean values (just like in ordinary linear models) while random effects modify the covariance matrix of the observations.

The hierarchical versus the marginal Model

The general mixed model is given by

$$\begin{cases} \boldsymbol{Y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \\ \boldsymbol{b}_i \sim N(\boldsymbol{0}, D), \\ \boldsymbol{\varepsilon}_i \sim N(\boldsymbol{0}, \Sigma_i), \\ \boldsymbol{b}_1, \ldots, \boldsymbol{b}_N, \boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_N \text{ independent}, \end{cases}$$

It can be written as

$$\boldsymbol{Y}_i|\boldsymbol{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i, \Sigma_i)$$

$$\boldsymbol{b}_i \sim N(\boldsymbol{0}, D)$$

It is therefore also called a hierarchical model

- A model for Y_i given b_i
- A model for b_i

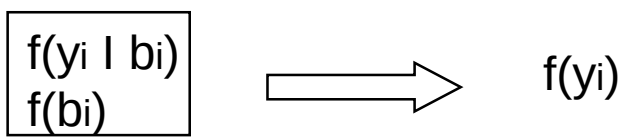
Marginally, we have that $\boldsymbol{Y}_i$ is distributed as:

$$\boldsymbol{Y_i} \sim N(X_i\boldsymbol{\beta}, Z_iDZ_i' + \Sigma_i)$$

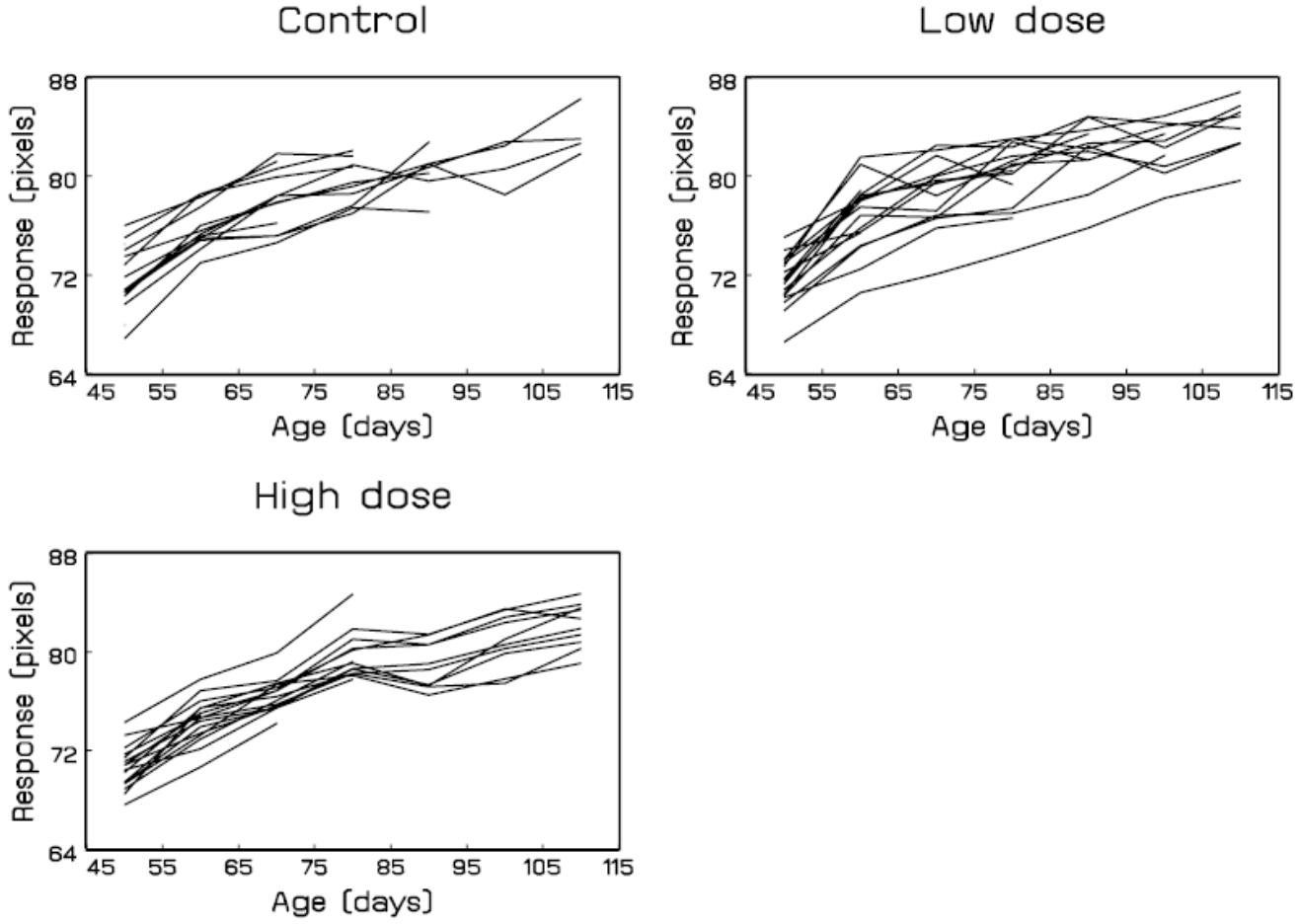
Hence, very specific assumptions are made about the dependence of mean and covariance on the covariates X_i and Z_i:

- **Implied mean :** $X_i\boldsymbol{\beta}$
- **Implied covariance :** $V_i = Z_iDZ_i' + \Sigma_i$

Note that the hierarchical model implies the marginal one, **not** vice versa



Example (Verbecke et al – mixed models course)



Example

- **Stage 1 model:**

$$Y_{ij} = \beta_{1i} + \beta_{2i} t_{ij} + \varepsilon_{ij}, \quad j = 1, \ldots, n_i,$$

- **Stage 2 model:**

$$\begin{cases} \beta_{1i} = \beta_0 + b_{1i}, \\ \beta_{2i} = \beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i}, \end{cases}$$

Can be negative or positive reflecting individual deviation from average

Linear model where each Subject has own intercept and own slope

$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i}) t_{ij} + \varepsilon_{ij}$$

$$= \begin{cases} \beta_0 + b_{1i} + (\beta_1 + b_{2i}) t_{ij} + \varepsilon_{ij}, & \text{if low dose} \\ \beta_0 + b_{1i} + (\beta_2 + b_{2i}) t_{ij} + \varepsilon_{ij}, & \text{if high dose} \\ \beta_0 + b_{1i} + (\beta_3 + b_{2i}) t_{ij} + \varepsilon_{ij}, & \text{if control.} \end{cases}$$

Stage 2 model:

- In the second stage, the subject-specific intercepts and time effects are related to the treatment of the rats

$$\begin{cases} \beta_{1i} = \beta_0 + b_{1i}, \\ \\ \beta_{2i} = \beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i}, \end{cases}$$

- L_i, H_i, and C_i are indicator variables:

$$L_i = \begin{cases} 1 & \text{if low dose} \\ 0 & \text{otherwise} \end{cases}$$

$$H_i = \begin{cases} 1 & \text{if high dose} \\ 0 & \text{otherwise} \end{cases}$$

$$C_i = \begin{cases} 1 & \text{if control} \\ 0 & \text{otherwise} \end{cases}$$

Parameter interpretation:

▷ β_0: average response at the start of the treatment (independent of treatment)

▷ β_1, β_2, and β_3: average time effect for each treatment group

$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i + b_{2i}) t_{ij} + \varepsilon_{ij}$$

$$= \begin{cases} \beta_0 + b_{1i} + (\beta_1 + b_{2i}) t_{ij} + \varepsilon_{ij}, & \text{if low dose} \\ \beta_0 + b_{1i} + (\beta_2 + b_{2i}) t_{ij} + \varepsilon_{ij}, & \text{if high dose} \\ \beta_0 + b_{1i} + (\beta_3 + b_{2i}) t_{ij} + \varepsilon_{ij}, & \text{if control.} \end{cases}$$

A model which assumes that all variability in subject-specific slopes can be ascribed to treatment differences can be obtained by omitting the random slopes b_{2i} from the above model:

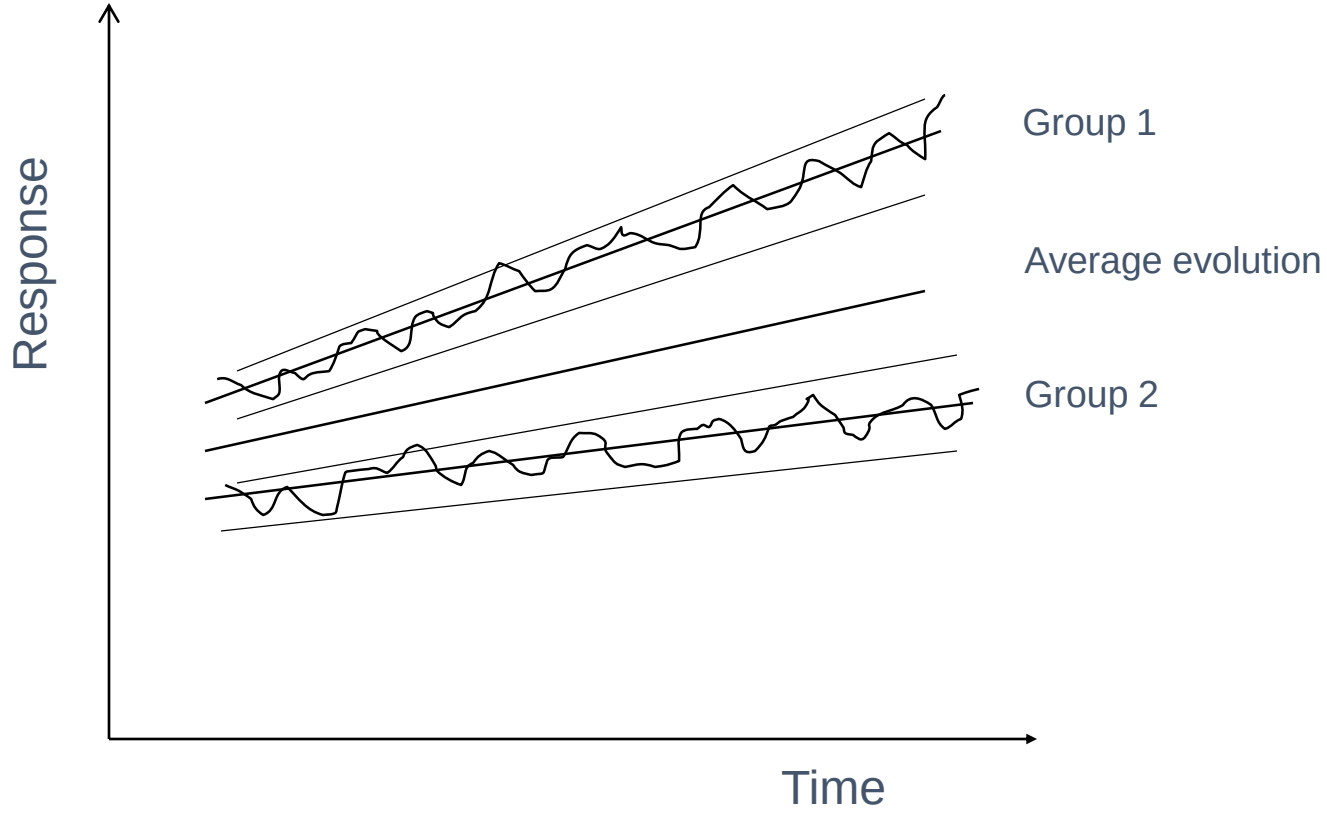
$$Y_{ij} = (\beta_0 + b_{1i}) + (\beta_1 L_i + \beta_2 H_i + \beta_3 C_i) t_{ij} + \varepsilon_{ij}$$

$$= \begin{cases} \beta_0 + b_{1i} + \beta_1 t_{ij} + \varepsilon_{ij}, & \text{if low dose} \\ \beta_0 + b_{1i} + \beta_2 t_{ij} + \varepsilon_{ij}, & \text{if high dose} \\ \beta_0 + b_{1i} + \beta_3 t_{ij} + \varepsilon_{ij}, & \text{if control.} \end{cases}$$

This is the so-called random-intercepts model

The same marginal mean structure is obtained as under the model with random slopes

Stochastic components in general linear mixed model



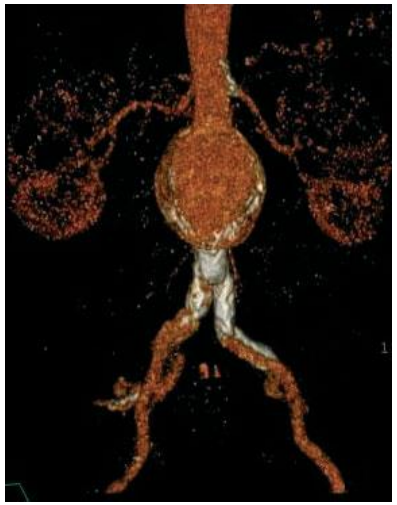
Comments:

- Linear average evolution in each group
- Equal average intercepts
- Different average slopes

Abdominal Aortic Aneurysm (AAA)

- Summary of Growth Data

- Endpoint AD with range 25-70.
- AD0: baseline value with range 25-70
- Number of subjects 211
- Number of screens 2-18
- Time 0-13
- Other variables: Age (5-87) - Diabetes -



Possible NLS models

mod1=nls(AD~(ADO+k*TIME), start=list(k=2), data=G)

mod2=nls(AD~(ADO+k*TIME*TIME), start=list(k=2), data=G)

mod3=nls(AD~(ADO*exp(k*TIME)), start=list(k=0.2), data=G)

mod4=nls(AD~(ADO*exp(k*TIME*TIME)), start=list(k=0.2), data=G)

Summary NLS mod1-mod4

> summary(mod1)

Formula: AD ~ (ADO + k * TIME)

Parameters:

	Estimate		Std. Error	t value	Pr(>|t|)
k	**1.85931**	**0.03666**	**50.71**	**<2e-16 *****	

> summary(mod2)

Formula: AD ~ (ADO + k * TIME * TIME)

Parameters:

	Estimate		Std. Error	t value	Pr(>|t|)
k	**0.232701**	**0.006618**	**35.16**	**<2e-16 *****	

> summary(mod3)

Formula: AD ~ (ADO * exp(k * TIME))

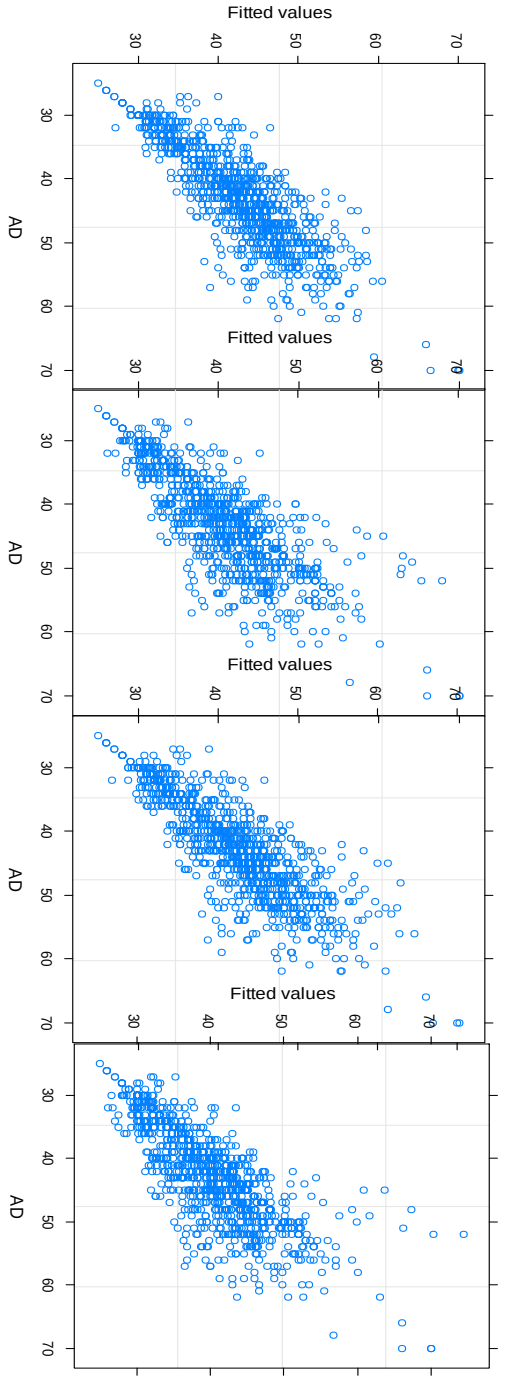
Parameters:

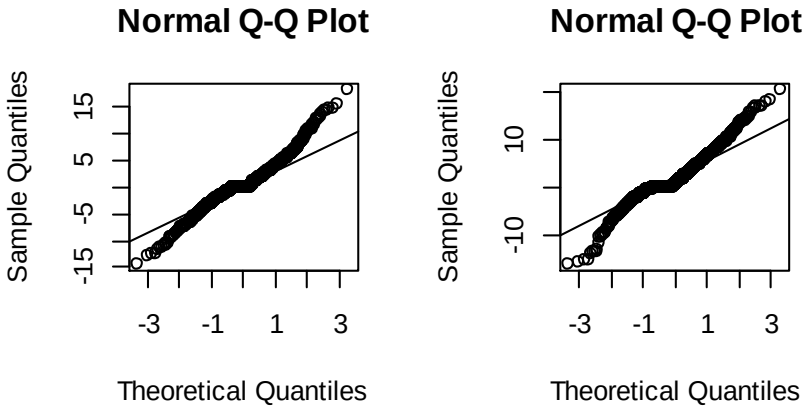
	Estimate		Std. Error	t value	Pr(>|t|)
k	**0.0454132**	**0.0008365**	**54.29**	**<2e-16 *****	

> summary(mod4)

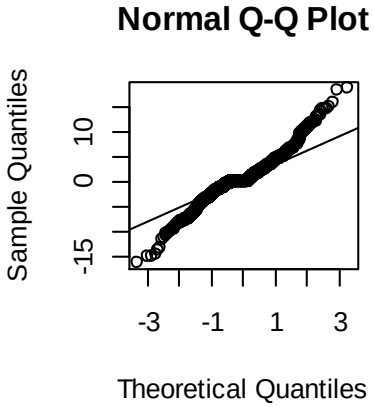
Formula: AD ~ (ADO * exp(k * TIME * TIME))

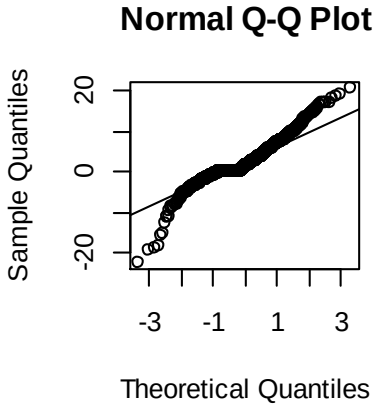
Parameters:

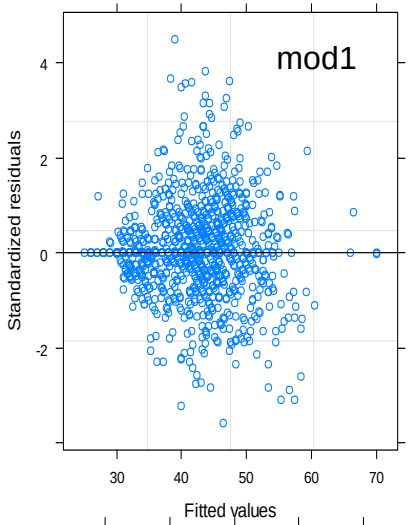


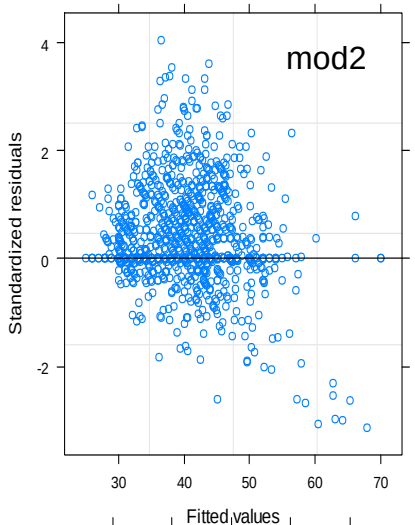

Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-15
-5
5
15
-3
-1
1
3

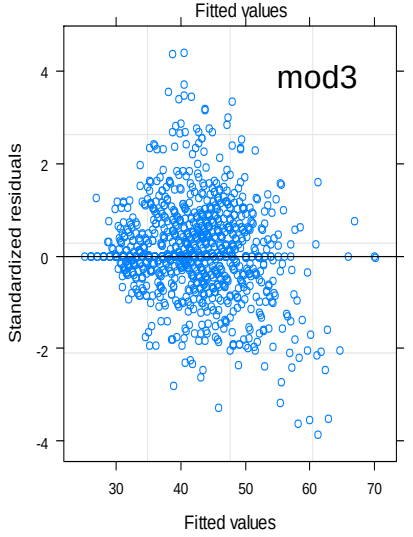
Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-10
10
-3
-1
1
3

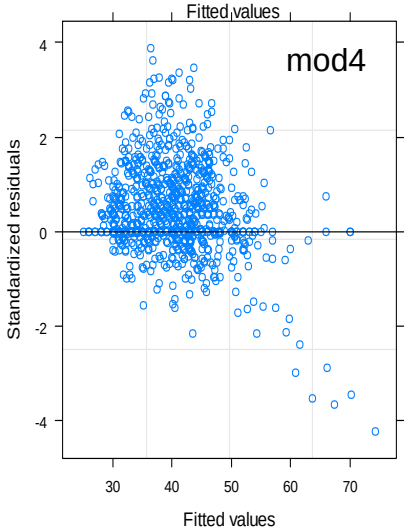

Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-15
0
10
-3
-1
1
3


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
-20
0
20
-3
-1
1
3


mod1
Standardized residuals
Fitted values
-2
0
2
4
30
40
50
60
70

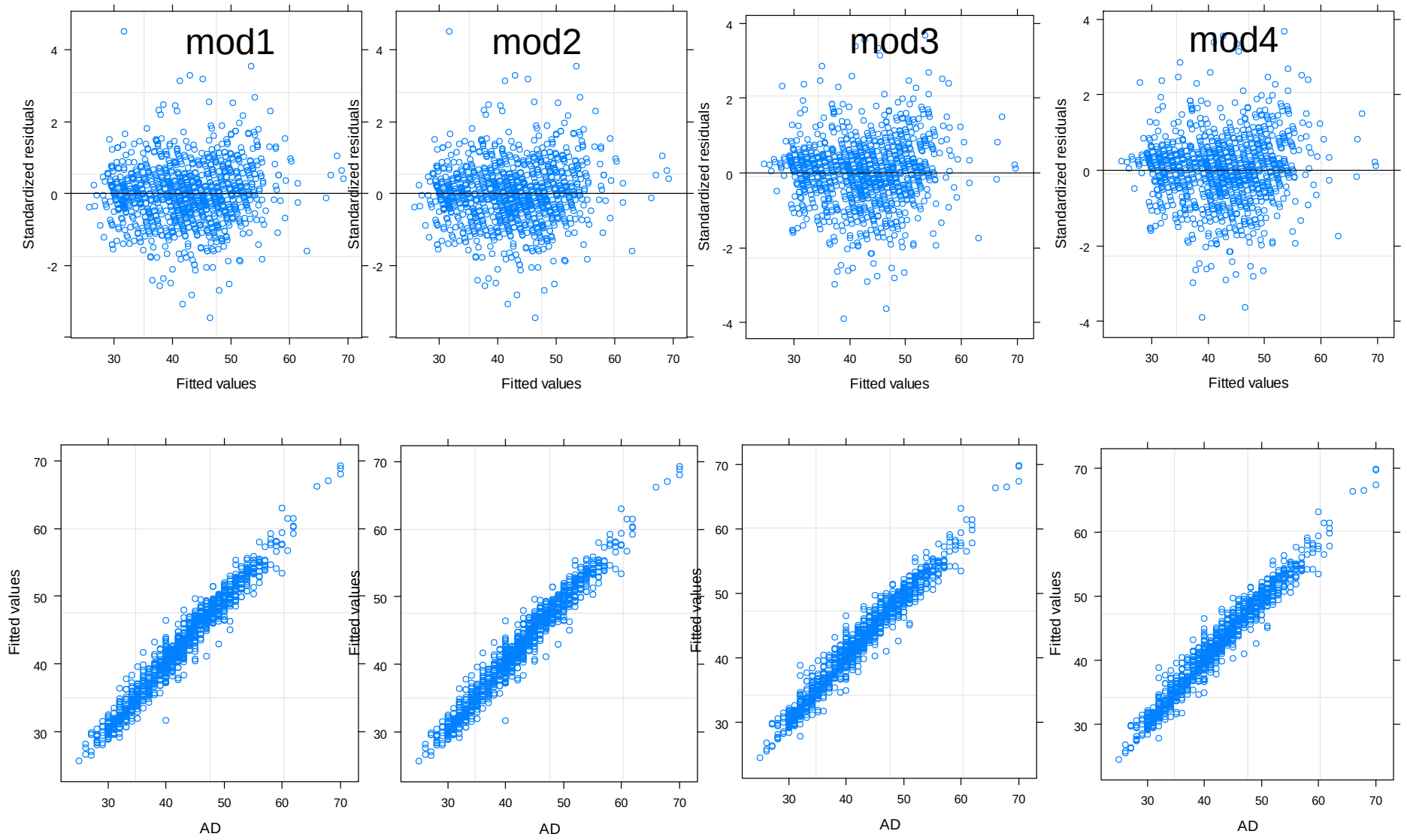

mod2
Standardized residuals
Fitted values
-2
0
2
4
30
40
50
60
70


mod3
Standardized residuals
Fitted values
-4
-2
0
2
4
30
40
50
60
70

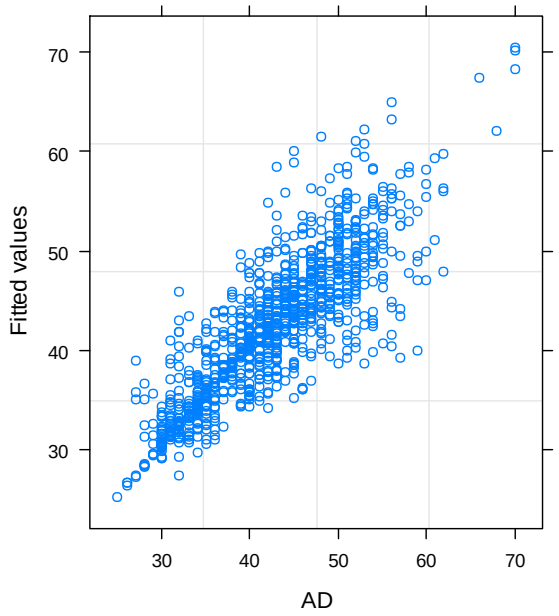

mod4
Standardized residuals
Fitted values
-4
-2
0
2
4
30
40
50
60
70

Linear mixed effects (lme) models

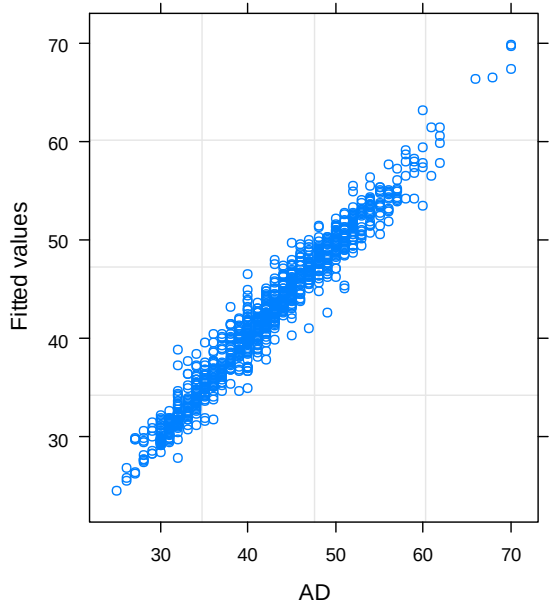
- **md1<- lme(AD~TIME, data=Growth, random=~ TIME|SUBJECT)**
- **md2<- lme(AD~TIME+TIME*TIME, data=Growth, random=~ TIME|SUBJECT)**
- **md3<- lme(AD~TIME+AGE+ADO, data=Growth, random=~ TIME|SUBJECT)**
- **md4<- lme(AD~TIME+TIME*TIME+AGE+ADO, data=Growth, random=~ TIME|SUBJECT)**


mod1
mod2
mod3
mod4
Standardized residuals
Fitted values
Fitted values
AD

mod3= nls(AD~(ADO*exp(k*TIME+b*AGE+h*SUBJECT)),
start = list(k=0.2,b=0.1,h=0.1), data=G)

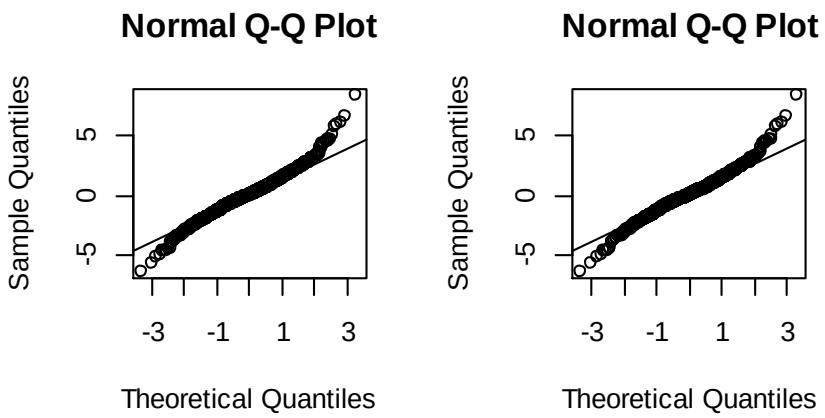


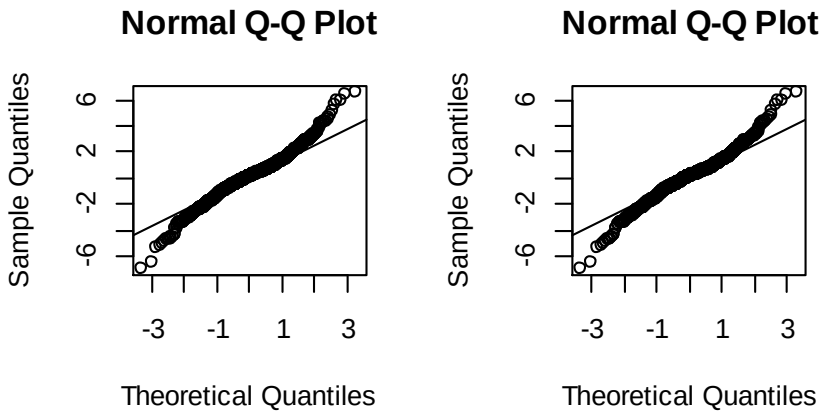
lme mod1



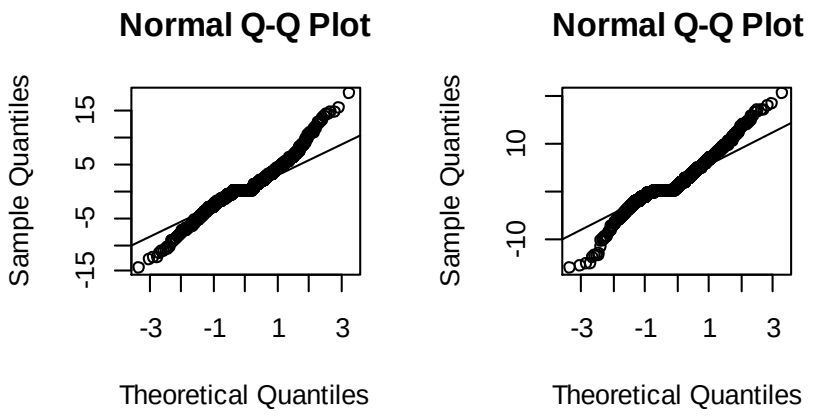
Without random effects: all variability is in the error term

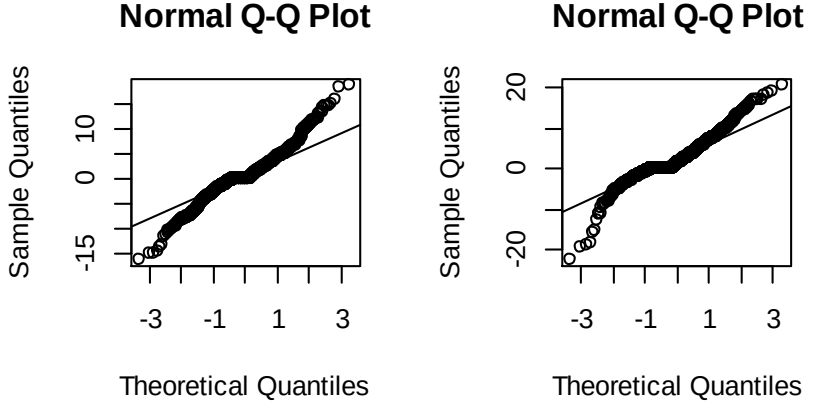
lme md1-md4





nls mod1-mod4





Bayesian Inference

Simplified scientific Process

1. A hypothesis about the efficacy of a drug needs to be tested
2. We perform a clinical trial and obtain some data
3. Are our data in agreement with our hypothesis?
4. If the answer to the above question is no, we reject the hypothesis. Otherwise we cannot reject it!

Classical Paradigm
vs
The Bayesian paradigm

- The classical paradigm is based on the consideration of

  P[Data | Hypothesis] (1)

- How likely is the data if the hypothesis was to be true?

Classical vs Bayesian (cont'd)

- The Bayesian paradigm is based on the consideration of

$$P[\text{ Hypothesis } | \text{ Data}] \qquad (2)$$

- How much support or belief (likelihood?) is there in the hypothesis given the data?

Bayes' Formula

States that

$$P[\,H \mid D] = \frac{P[\,D \mid H]\,P[\,H\,]}{P[\,D]} \qquad (3)$$

Where

H = Hypothesis (Theory)

and

D = Data

Simple formula with many interesting implications.

Implications of Bayes' Formula

1. (1) and (2) are not equivalent.

2. To work out (2) we have to estimate P[H] i.e. we need to put a probability on our belief in the hypothesis we are testing.

3. We cannot make P[H] disappear. (Similar to the uncertainty principle in quantum mechanics?)

Example: p-values

- Some toxin is associated with certain symptoms. μ denotes the toxin level in patients having the symptoms and μ_0 the toxin level in healthy individuals. We want to test if there is a difference. A test can be based on a sample of size n through

$$Z = \frac{\bar{Y} - \mu_0}{s/\sqrt{n}}$$

Under the null hypothesis

$$H_0: \mu = \mu_0,$$

z is an observation from a t-distribution with n-1 degrees of freedom.

If, moreover, the p-value is less than 0.05 it is then customary to consider the result as significant, i.e. we reject the null hypothesis.

We return to this in moment!

Preliminaries

- Consider two Hypothesis H_0 and H_1

$$P[H_0 | D] = \frac{P[D | H_0] P[H_0]}{P[D]}$$

$$P[H_1 | D] = \frac{P[D | H_1] P[H_1]}{P[D]}$$

- We consider the odds ratio between the two

$$\frac{P[H_0|D]}{P[H_1|D]} = \frac{P[H_0]}{P[H_1]} \times \frac{P[D|H_0]}{P[D|H_1]}$$

Posterior ratio = prior odds x likelihood ratio

New = old Bayes factor

Result

$$P[H_0 \mid D] = \left(1 + \frac{1 - P[H_0]}{P[H_0] \times BF}\right)^{-1}$$

where

$$BF = \frac{P[D \mid H_0]}{P[D \mid H_1]}$$

Back to the t-test

P. M. Lee Bayesian Statistics: An Introduction
2nd Ed. London: Arnold, p. 131 (1997)

shows that in the case of the t-est and under quite general conditions:

$$\mathrm{BF} \geq e^{\frac{-z^2}{2}}$$

Consequences

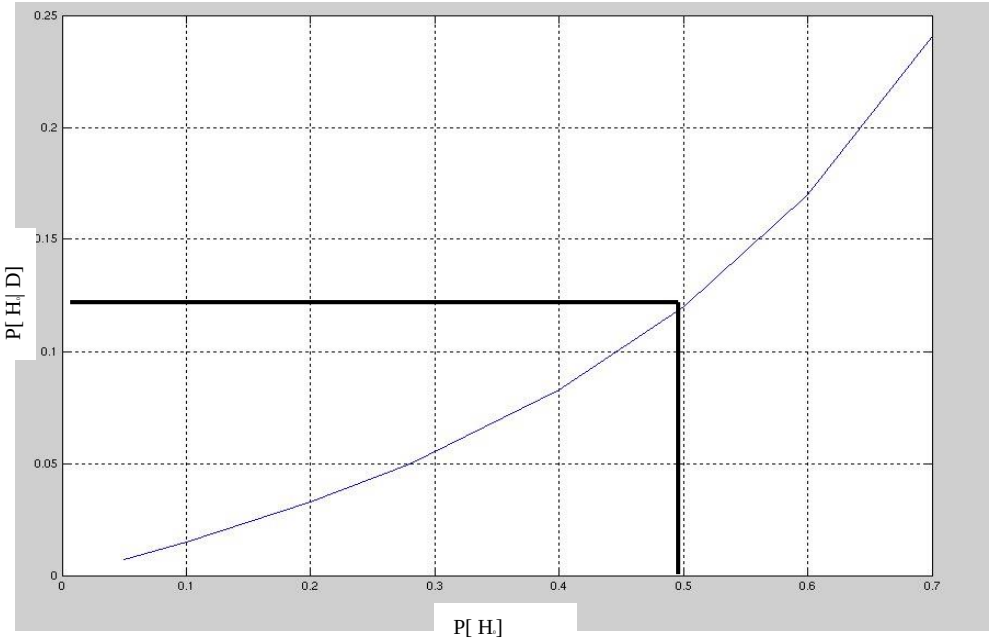
- Assume that there is no agreement on the effect of the toxin so we take P(H_0)=0.5. Then:

$$P[H_0 \mid D] \geq \left(1 + \frac{1}{e^{\frac{-z^2}{2}}}\right)^{-1} = \left(1 + e^{\frac{-z^2}{2}}\right)^{-1}$$

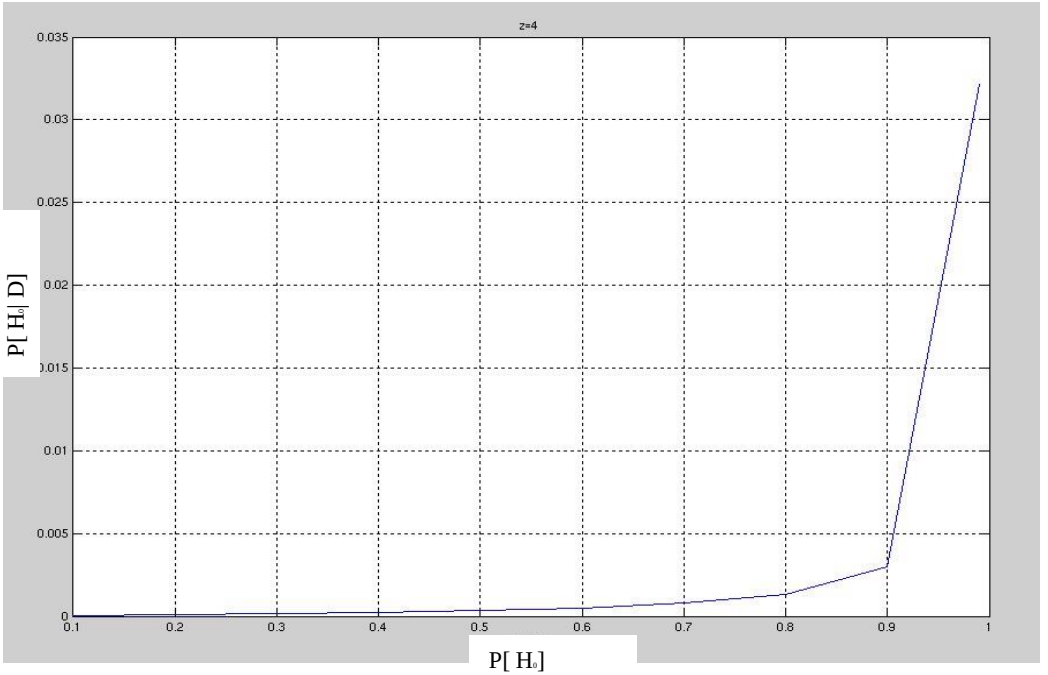
- Assume further that the value of z in the experiment turns out to be 2. Since this leads to a p-value of 0.044 we conclude that the result is significant at the 0.05 level. Setting z=2 in the formula above leads to:

$$P[H_0 \mid D] \geq \left(1 + e^{\frac{-2^2}{2}}\right)^{-1} = 0.12$$

Z=2


0.25
0.2
0.15
0.1
0.05
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
P[H.| D]
P[H.]

Z=4


z=4
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
P[H₀| D]
P[H₀]

Conclusions

- By introducing the element of degree of belief about a theory, we arrive at conclusions that do not agree with those obtained using the frequentist approach, i.e. prior knowledge matters

- Prior knowledge is part of the Bayesian approach.

Bayesian Analysis in Clinical Trials

Bayesian Framework

- ***The Advantages:***
  - Formal system for incorporating existing information
  - Natural approach to inference
  - Generally more efficient
  - Well suited for decision making
- ***The Challenges:***
  - Determining appropriate prior probabilities
  - Computational complexity
  - Lack of familiarity
  - Lack of software tools

Proof

$$\frac{P[H_0|D]}{P[H_1|D]} = \frac{P[H_0]}{P[H_1]} \times \frac{P[D|H_0]}{P[D|H_1]}$$

$$\Rightarrow P[H_0|D] = \frac{P[H_0]}{P[H_1]} \times \frac{P[D|H_0]}{P[D|H_1]} \times P[H_1|D]$$

$$= C \times (1 - P[H_0|D])$$

$$\Rightarrow P[H_0|D] + C \times P[H_0|D] = C$$

$$\Leftrightarrow P[H_0|D] = \left(\frac{C}{1+C}\right) = \left(\frac{1+C}{C}\right)^{-1} = \left(1 + \frac{1}{C}\right)^{-1}$$

$$= \left(1 + \frac{1 - P[H_0]}{P[H_0] \times \frac{P[D|H_0]}{P[D|H_1]}}\right)^{-1}$$

$$\Rightarrow P[H_0|D] = \left(1 + \frac{1 - P[H_0]}{P[H_0] \times BF}\right)^{-1}$$