Repeated measures, simple methods

Anders Stockmarr
Course developers: Anders Stockmarr, Elisabeth Wreford Andersen,
Per Bruun Brockhoff

DTU Department of Applied Mathematics and Computer Science
Section for Statistics and Data Analysis
Technical University of Denmark
anst@dtu.dk

January 10th 2025

DTU Compute
Department of Applied Mathematics and Computer Science

Plan for this week

Monday	Statistical inference, and the t-test
Tuesday	Simple and Multiple regression
Wednesday	ANOVA, ANCOVA, and linear models
Thursday	Categorical data, Writing statistical reports, Logistic regression
Friday	Repeated measurements, Principal Component Analysis

Overview of this module

1. The repeated measurements setup
   - Aspirin Example
   - Activity of rats
2. Separate analysis for each time–point
   - Example: rats data
3. Analysis of a summary statistic
   - Example: rats data
4. Random effects model - simple version
   - Example: rats data
5. Random effects model - advanced version
6. Pros and cons of simple approaches

Aim of this module

- Present simple methods for dealing with repeated measurements - and a few more advanced applications
- Easy to use, and a lot better than pretending to have independent observations
- Useful even after more advanced models are presented
- See how to specify these models in R

Overview

The repeated measurements setup

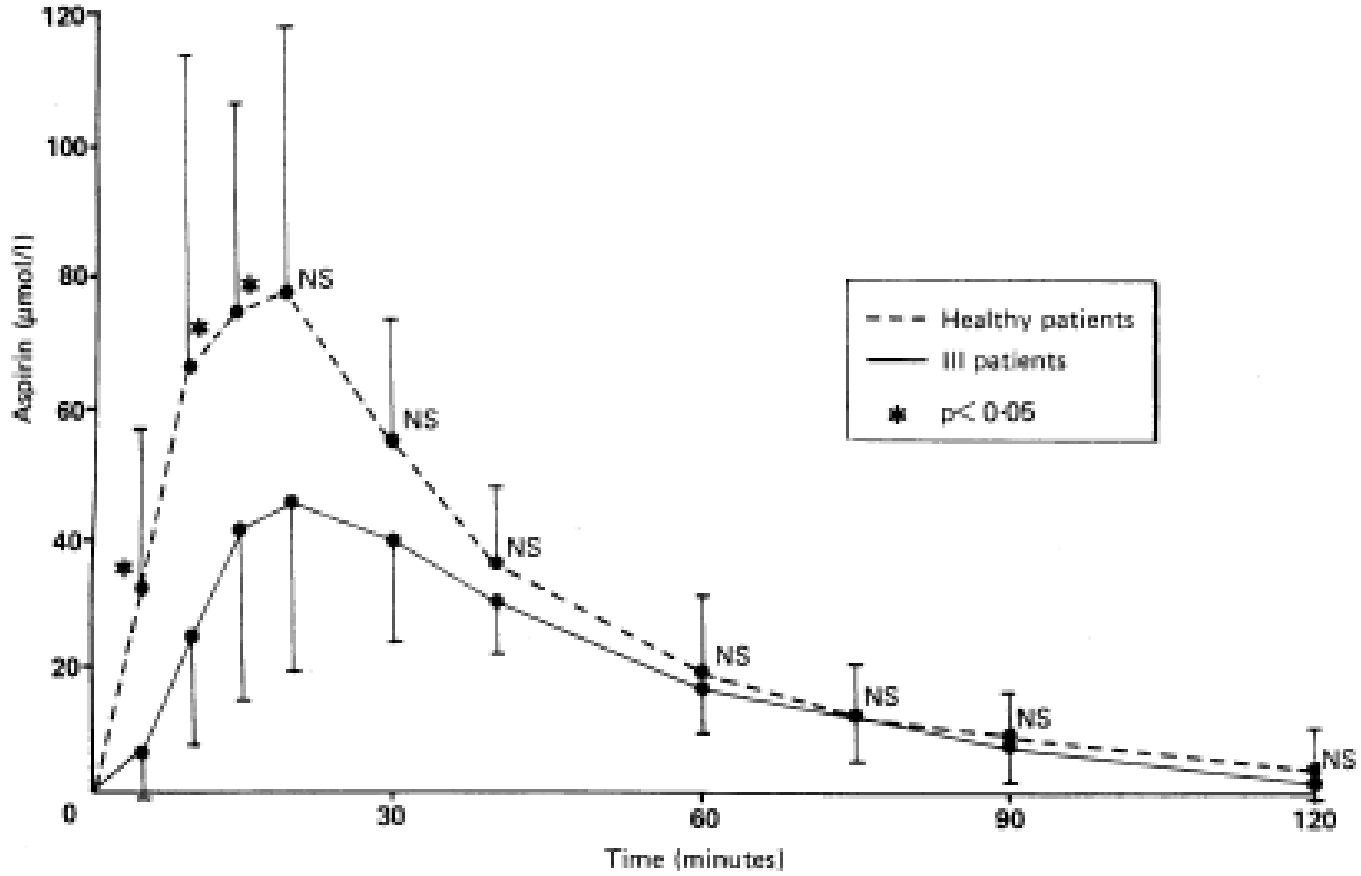
- Several ”individuals”
- Several measurements on each individual
- Two measurements on the same individual might be correlated
- Might even be highly correlated if “close” and less correlated if “far apart”

The repeated measurements setup

- Several "individuals"
- Several measurements on each individual
- Two measurements on the same individual might be correlated
- Might even be highly correlated if "close" and less correlated if "far apart"
- Typical example (Matthews et al 1990):
  - Two groups of patients (ill/healthy)
  - Treated with asprin
  - The asprin concentration is measured in the blood 10 times during 2 hours

To pretend all observations are independent can lead to wrong conclusions

Traditional figure of means at each time point



What is the aim of the study?

It is always important to consider the aim of the study!

1. To describe pattern over time
   - Do we have a change over time?
   - Linear or curve?
   - Is the pattern the same for all groups?
2. Can we see a difference between groups?
   - Is the difference the same at all time points?
   - Is the difference in the levels?
   - Is the difference in the trend?

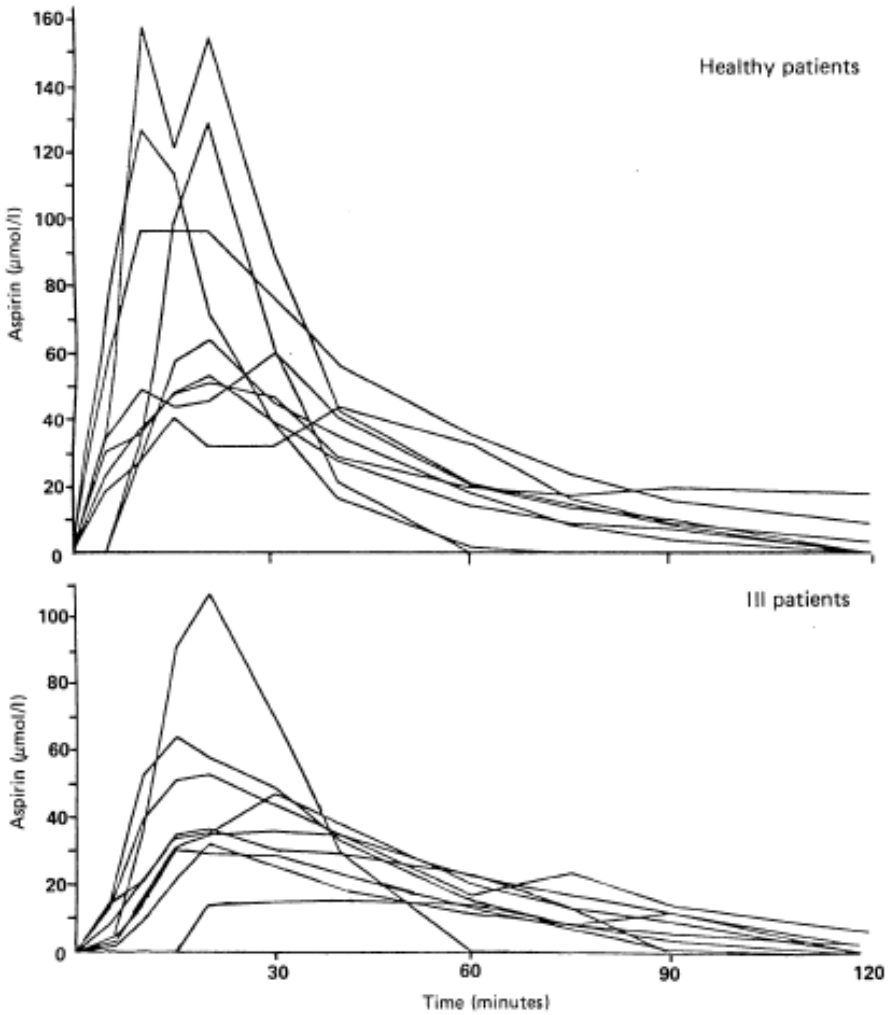
Problems using the traditional presentation of the data

- Comparing the groups at each time point.
  - Not effecient
  - Tests are not indenpendent, carried out on the same individuals
  - Difficult to interpret
- The pattern over time
  - This can only be studied if we use that each individual has several observations

Take care when using average curves

- Important structures can be overlooked if you start with the average curves.
- You cannot see the variation over time.
- Always make a plot of the individual observations.
- Only use the average if the patterns are similar, i.e. we only see shifts up and down.

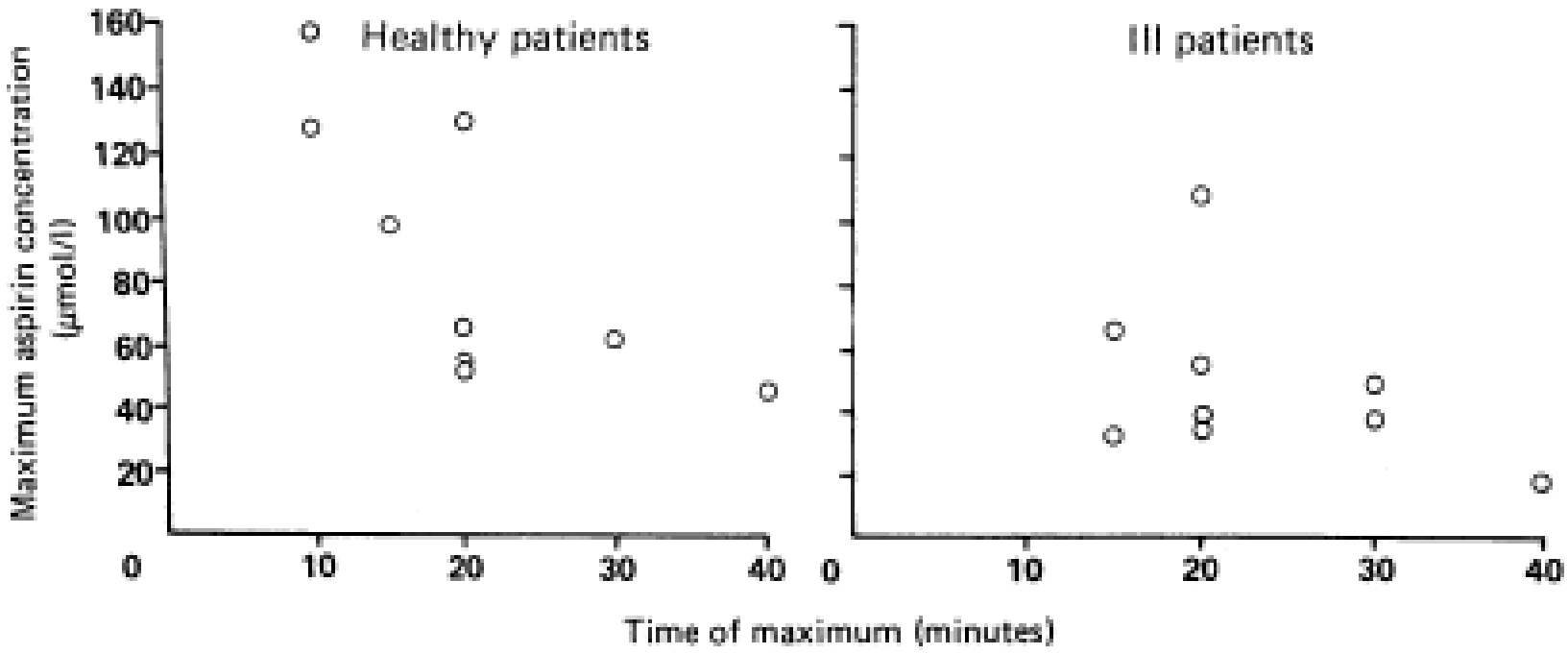
Individual curves



Analysis of summary statistics

- Choose a single measure to summarize the individual curves
- This again reduces the data set to independent observations
- Popular choices of summary measures - "relevant"feature extraction:
  - Total increase (last point minus first point)
  - Area under curve (AUC)
  - Maximum or minimum point
  - Average over time
  - Slope in regression with time
  - (or higher order polynomial coefficients)
- Good method with few and easily checked assumptions
- Information may be lost
- Important to choose a good summary measure

Summary Statistics



Analysis of aspirin study

TABLE I — *Analysis of data from aspirin study*

	Area under curve	Maximum concentration
Healthy patients (n = 9)		
Arithmetic mean (SD) (μmol/l)	26·5 (8·8)	86·0 (41·5)
Geometric mean (μmol/l)	25·4	77·8
Ill patients (n = 9)		
Arithmetic mean (SD) (μmol/l)	17·5 (5·0)	46·7 (26·3)
Geometric mean (μmol/l)	16·8	41·2
Ratio of geometric means	1·52	1·89
95% Confidence interval	1·11 to 2·08	1·14 to 3·13
t Test	2·83 (df = 16)	2·66 (df = 16)
p Value	0·01	0·02

Example: Activity of rats

Summary of experiment:

- 3 treatments: 1, 2, 3 (concentrations)
- 10 cages per treatment
- 10 months
- The response is activity ($\log$(count) of intersections of light beam during 57 hours)

The rats data

```
rats <- read.csv("rats.txt")

# make treatment and cage factors
rats$treatm <- factor(rats$treatm)
rats$cage <- factor(rats$cage)

# make two versions of the time variable
 # - one quantitative and one factor
rats$monthQ <- rats$month
rats$month <- factor(rats$month)

summary(rats)
str(rats)
```

In this example we have repeated measurements from each cage.

Data in two formats, long

This is the format we want. One response (here lnc) in each line

```
##    treatm cage month    lnc
## 1       1    1     1 9.9323
## 2       1    1     2 9.6447
## 3       1    1     3 9.7628
## 4       1    1     4 9.6014
## 5       1    1     5 9.3227
## 6       1    1     6 9.2463
## 7       1    1     7 9.0739
## 8       1    1     8 9.2077
## 9       1    1     9 9.1670
## 10      1    1    10 8.8319
```

Data in two formats, wide

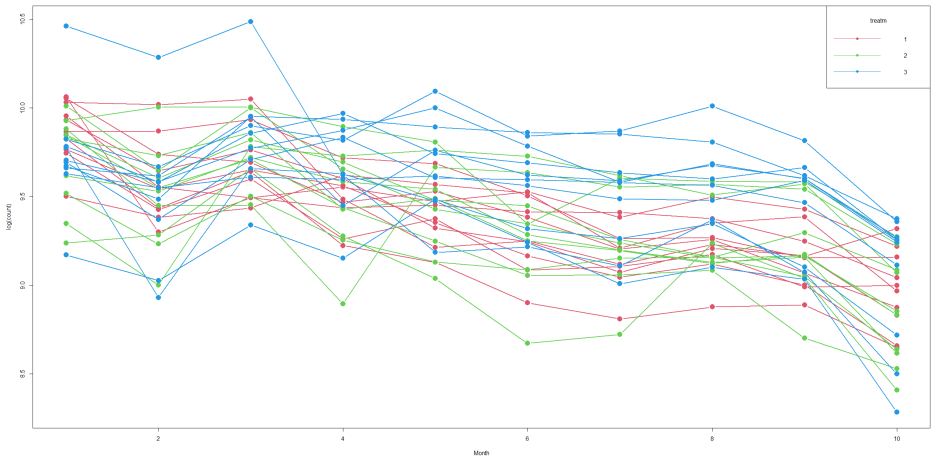
One line for each cage.

```
##    treatm cage   lnc.1  lnc.2  lnc.3  lnc.4  lnc.5  lnc.6  lnc.7  lnc.8
## 1       1    1  9.9323 9.6447 9.7628 9.6014 9.3227 9.2463 9.0739 9.2077
## 11      1    3 10.0547 9.7384 9.6928 9.4676 9.5297 9.3837 9.2052 9.2582
## 21      1    5  9.7448 9.4278 9.5995 9.2614 9.3744 9.0857 9.1068 9.1732
## 31      1    7  9.8660 9.8687 9.9329 9.7183 9.6861 9.5047 9.2622 9.2699
## 41      1    9  9.9552 9.5536 9.4930 9.4369 9.4520 9.4147 9.4107 9.3751
##     lnc.9 lnc.10
## 1  9.1670 8.8319
## 11 9.0681 8.8753
## 21 8.9907 8.9995
## 31 9.1555 9.1584
## 41 9.2475 9.0435
```

Not usually the format wanted for analysis but adapted to plots and group averages.

The rats data - plotting the individual patterns

```
plot(rats$monthQ,rats$lnc,xlab="Months",
  ylab="log(count)",type ="n")
for(i in 1:dim(rats2)[1]){
  lines(1:10,rats2[i,-(1:2)],type="b",col=rats2$treatm[i]+1,
  pch=16,lwd=2,cex=2)
}
legend("topright",legend=1:3,col=2:4,pch=16,lty=2,title="treatm")
```

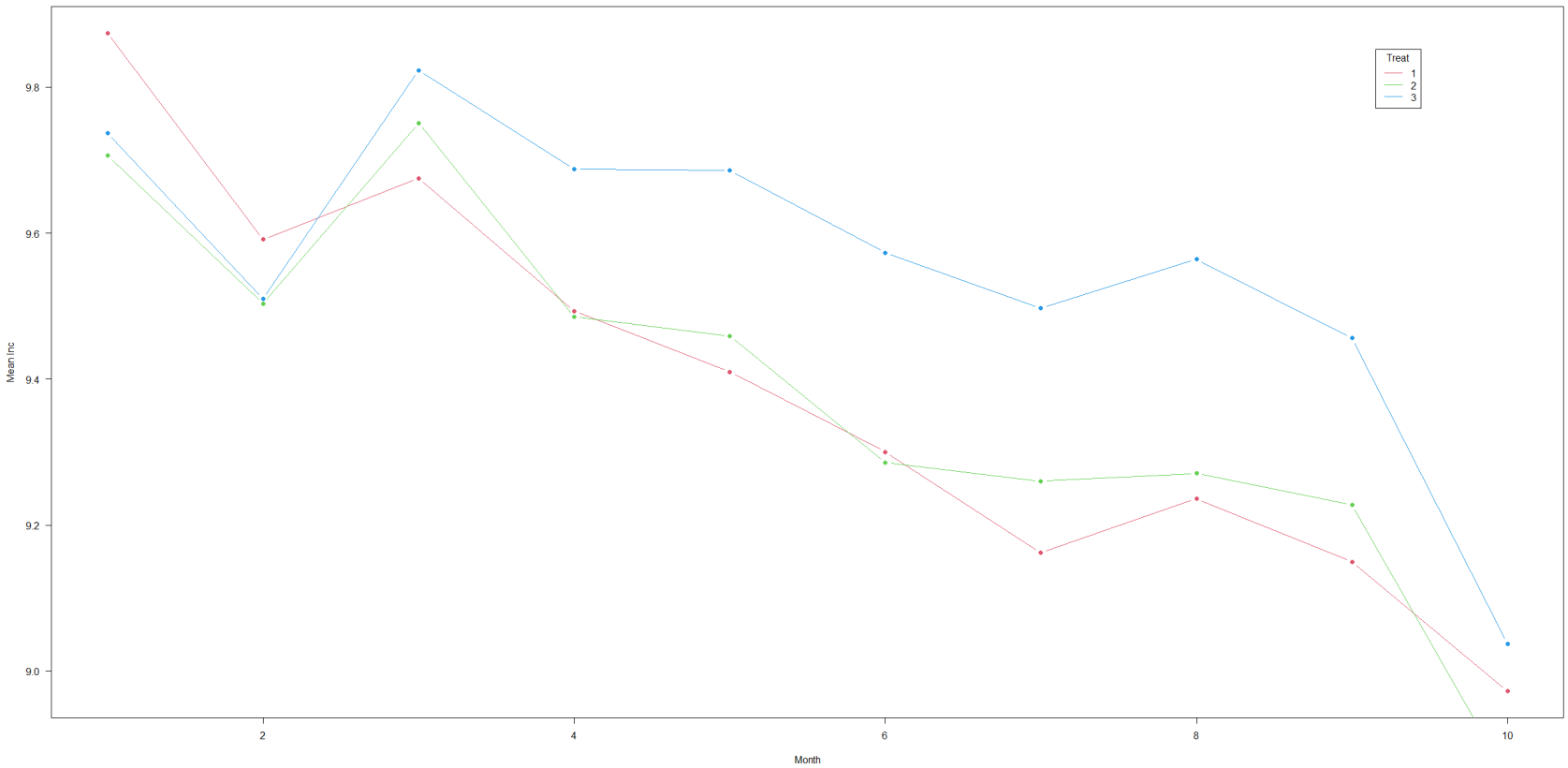


The rats data - plotting the average patterns

```
Mean_data <- aggregate(rats$lnc, by = list(rats$month, rats$treatm)
                       mean)
names(Mean_data) <- c("month","treatm","Meanlnc" )
#Plot the means
Grp1<-subset(Mean_data,treatm==1)
Grp2<-subset(Mean_data,treatm==2)
Grp3<-subset(Mean_data,treatm==3)

plot(as.numeric(Grp1$month), Grp1$Meanlnc, type = "b", pch=16,
  xlab = "Month",ylab = "Mean lnc", las = 1, col= 2)
lines(as.numeric(Grp2$month), Grp2$Meanlnc, type = "b",
      col = 3, pch=16)
lines(as.numeric(Grp3$month), Grp3$Meanlnc, type = "b",
      col = 4, pch=16)
legend(locator(1),              # we will place it with a mouse click
       legend = c("1","2","3"),
       title = "Treat",lty = c(2,2,2),col= 2:4)
```

The rats data - plotting the average patterns



Overview

Separate analysis for each time–point (continued)

- Select a fixed time point
- The observations at that time (one from each individual) are independent
- Do a separate analysis for the observations at that time
- This is not wrong, but (possibly) a lot of information is wasted
- This can be done for several time–points, but
  - Difficult to reach a coherent conclusion
  - Sub–tests are not independent
  - Tempting to select time–points supporting our preference
  - Mass significance: If many tests are carried out at 5% level some might be significant by chance. (Bonferroni correction: Use significance level $0.05/n$ instead of 0.05)

Separate analysis of rats data

```
#use the by function to make 10 tests
byMonth <- by(rats, rats$monthQ,
              function(x) anova(lm(lnc ~ treatm, data = x)))
#The largest effect at month 8
byMonth[[8]]

## Analysis of Variance Table
##
## Response: lnc
##           Df Sum Sq Mean Sq F value Pr(>F)
## treatm     2  0.649   0.324    7.29 0.0029 **
## Residuals 27  1.201   0.044
```

Separate analysis of rats data

- The model at each month is:

$$\texttt{lnc}_i = \mu + \alpha(\texttt{treatm}_i) + \varepsilon_i \ , \quad \varepsilon_i \sim \text{i.i.d. } N(0, \sigma^2), \quad i = 1 \ldots 30$$

- The result of the ten tests for no treatment effect:

Month	1	2	3	4	5	6	7	8	9	10
F–value	1.22	0.27	1.02	2.30	3.87	4.10	4.70	7.29	4.09	0.88

Compare with $F_{95\%;2,27} = 3.35$ or $F_{99.5\%;2,27} = 6.49$ if Bonferroni correction is used

Overview

Rats data analyzed via summary measure

- The log of the total activity in each cage is chosen as summary measure $\texttt{lnTot} = \log(\text{Total count})$
- The one way ANOVA model becomes:

$$\texttt{lnTot}_i = \mu + \alpha(\texttt{treatm}_i) + \varepsilon_i, \quad \varepsilon_i \sim \text{ i.i.d. } N(0, \sigma^2), \quad i = 1 \ldots 30$$

- Notice the simplicity of the model and the relative few assumptions

Rats data analyzed via a summary measure – log total activity

```
rats$count <- exp(rats$lnc)
DataTot <- aggregate(rats$count, by = list(rats$cage,
                                           rats$treatm), sum)
names(DataTot) <- c("cage","treatm", "Tot")
head(DataTot)
```

```
##   cage treatm    Tot
## 1    1      1 124864
## 2    3      1 131277
## 3    5      1 110166
## 4    7      1 145418
## 5    9      1 128819
## 6   11      1 136214
```

log total activity (contd.)

```
reg1 <- lm(log(Tot) ~ treatm, data = DataTot); summary (reg1)

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 11.73354    0.05409  216.91   <2e-16 ***
## treatm2     -0.00567    0.07650   -0.07    0.941
## treatm3      0.16728    0.07650    2.19    0.038 *
##
## Residual standard error: 0.171 on 27 degrees of freedom
## Multiple R-squared:  0.196,Adjusted R-squared:  0.137
## F-statistic:  3.3 on 2 and 27 DF,  p-value: 0.0522

anova(reg1)

## Analysis of Variance Table
##
## Response: log(Tot)
##           Df Sum Sq Mean Sq F value Pr(>F)
## treatm     2  0.193  0.0965     3.3  0.052 .
## Residuals 27  0.790  0.0293
```

Rats data analyzed via a summary measure - slopes

```
# byCage is a regression with different slope and intercept
# for each cage. Saves the coefficients
byCage <- coef(lm(lnc ~ -1 + cage + monthQ:cage, data = rats))
slope1 <- data.frame(matrix(byCage, nrow=30, byrow=F))

names(slope1) <- c("Intercept", "Slope")

#We also need info about cage and treat
SlopeData <- cbind(DataMean[ , 1:2], slope1)
head(SlopeData)
```

```
##   cage treatm Intercept     Slope
## 1    1      1    9.9685 -0.107163
## 2    3      1   10.0411 -0.111593
## 3    5      1    9.6981 -0.076675
## 4    7      1   10.0831 -0.098336
## 5    9      1    9.8067 -0.066990
## 6   11      1    9.7310 -0.043147
```

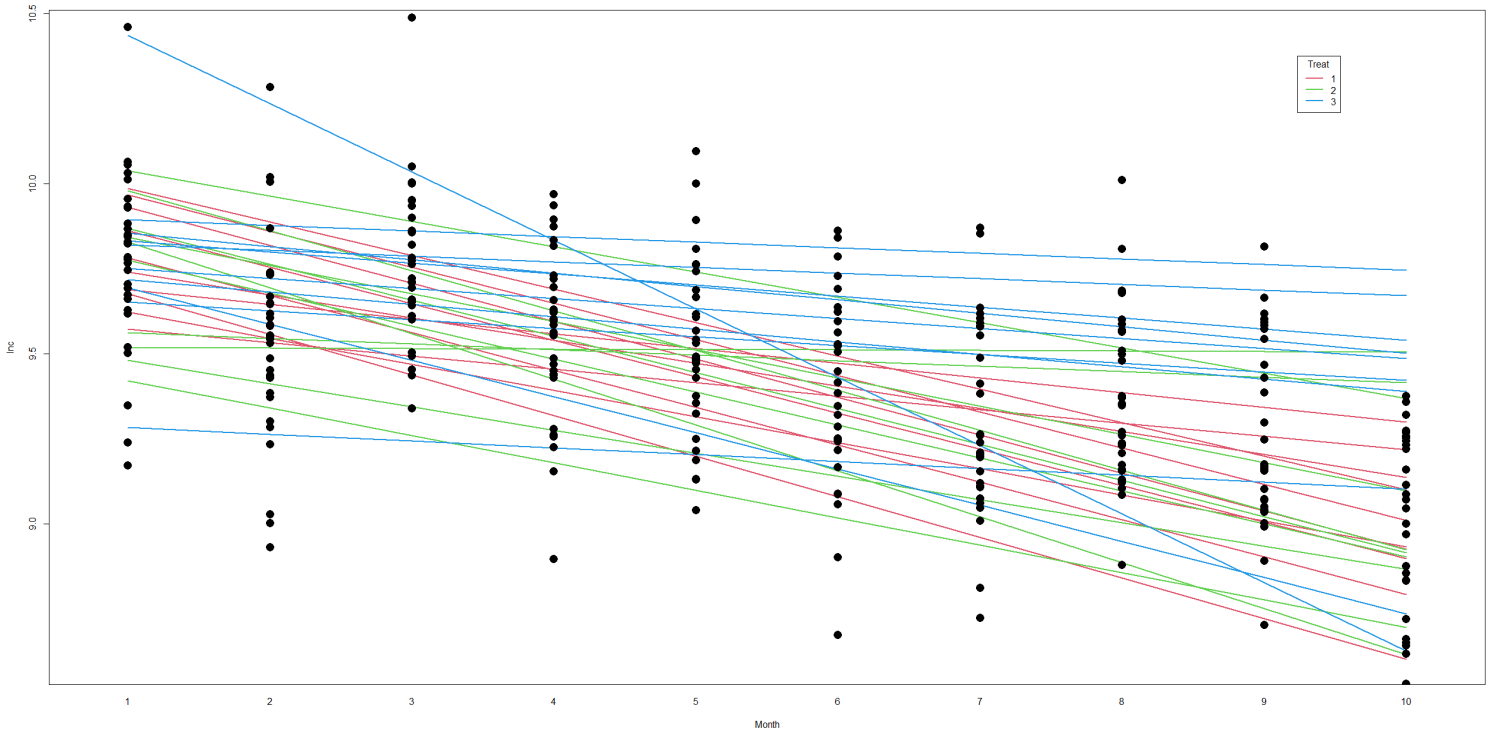
Rats data analyzed via a summary measure - plotting slopes

```
# Individual slopes:
fit <- unlist(by(rats, rats$cage,
          function(x) fitted.values(lm(lnc ~ monthQ, data=x))))
names(fit) <- NULL

#plotting the linear fit by cage
interaction.plot(rats$monthQ, rats$cage, fit,
  xlab="Month", ylab="lnc",legend=F,
  col=as.numeric(SlopeData$treatm)+1,lty=1,lwd=2)

lines(rats$monthQ,rats$lnc, type="p",pch=16,cex=2)
legend(locator(1),  # we will place it with a mouse click
       legend = c("1","2","3"), title = "Treat",
       lty = c(1,1,1),lwd=2,col= 2:4)
```

Rats data analyzed via a summary measure - plotting slopes



Rats data analyzed via a summary measure - slopes (contd)

```
reg3 <- lm(Slope ~ treatm, data = SlopeData)
summary(reg3)

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.0879     0.0142   -6.20  1.2e-06 ***
## treatm2       0.0100     0.0200    0.50    0.620
## treatm3       0.0355     0.0200    1.77    0.088 .
##
## Residual standard error: 0.0448 on 27 degrees of freedom
## Multiple R-squared:  0.11,Adjusted R-squared:  0.0439
## F-statistic: 1.67 on 2 and 27 DF,  p-value: 0.208
```

Rats data analyzed via a summary measure - slopes (contd)

```
reg3 <- lm(Slope ~ treatm, data = SlopeData)
summary(reg3)

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.0879     0.0142   -6.20  1.2e-06 ***
## treatm2       0.0100     0.0200    0.50    0.620
## treatm3       0.0355     0.0200    1.77    0.088 .
##
## Residual standard error: 0.0448 on 27 degrees of freedom
## Multiple R-squared:  0.11,Adjusted R-squared:  0.0439
## F-statistic: 1.67 on 2 and 27 DF,  p-value: 0.208
```

- Treat 1 has been chosen as reference with estimated slope -0.09.
- Treat 2 is very similar to Treat 1 with slope $-0.09 + 0.01 = -0.08$
- Treat 3 has slope closer to 0: $-0.09 + 0.04 = -0.05$

Rats data analyzed via a summary measure - slopes (contd)

```
anova(reg3)

## Analysis of Variance Table
##
## Response: Slope
##           Df Sum Sq Mean Sq F value Pr(>F)
## treatm     2 0.0067 0.00334    1.67   0.21
## Residuals 27 0.0542 0.00201
```

From the one-way analysis of variance we do not see a significant difference in the slopes between the three groups.

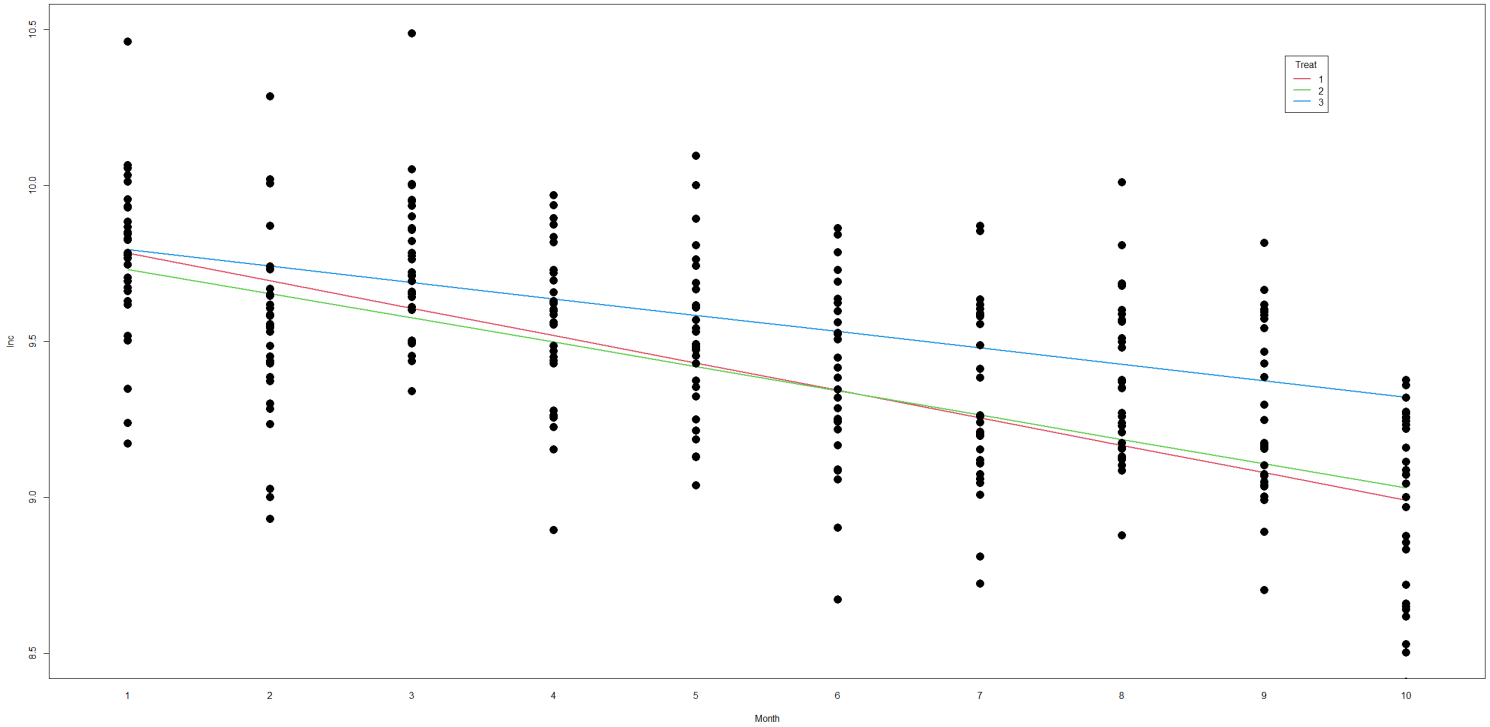
Rats data analyzed via a summary measure - plotting slopes

```
# Treatment average slopes:
fit2 <- unlist(by(rats, rats$treatm,
          function(x) fitted.values(lm(lnc ~ monthQ, data=x))))
names(fit2) <- NULL

#plotting the linear fit by treatment
interaction.plot(rats$monthQ, rats$treatm, fit2,
  xlab="Month", ylab="lnc",legend=F,
  col=2:4,lty=1,lwd=2,ylim=c(8.5,10.5))

lines(rats$monthQ,rats$lnc, type="p",pch=16,cex=2)
legend(locator(1),
       legend = c("1","2","3"), title = "Treat",
       lty = c(1,1,1),lwd=2,col= 2:4)
```

Rats data analyzed via a summary measure - plotting slopes



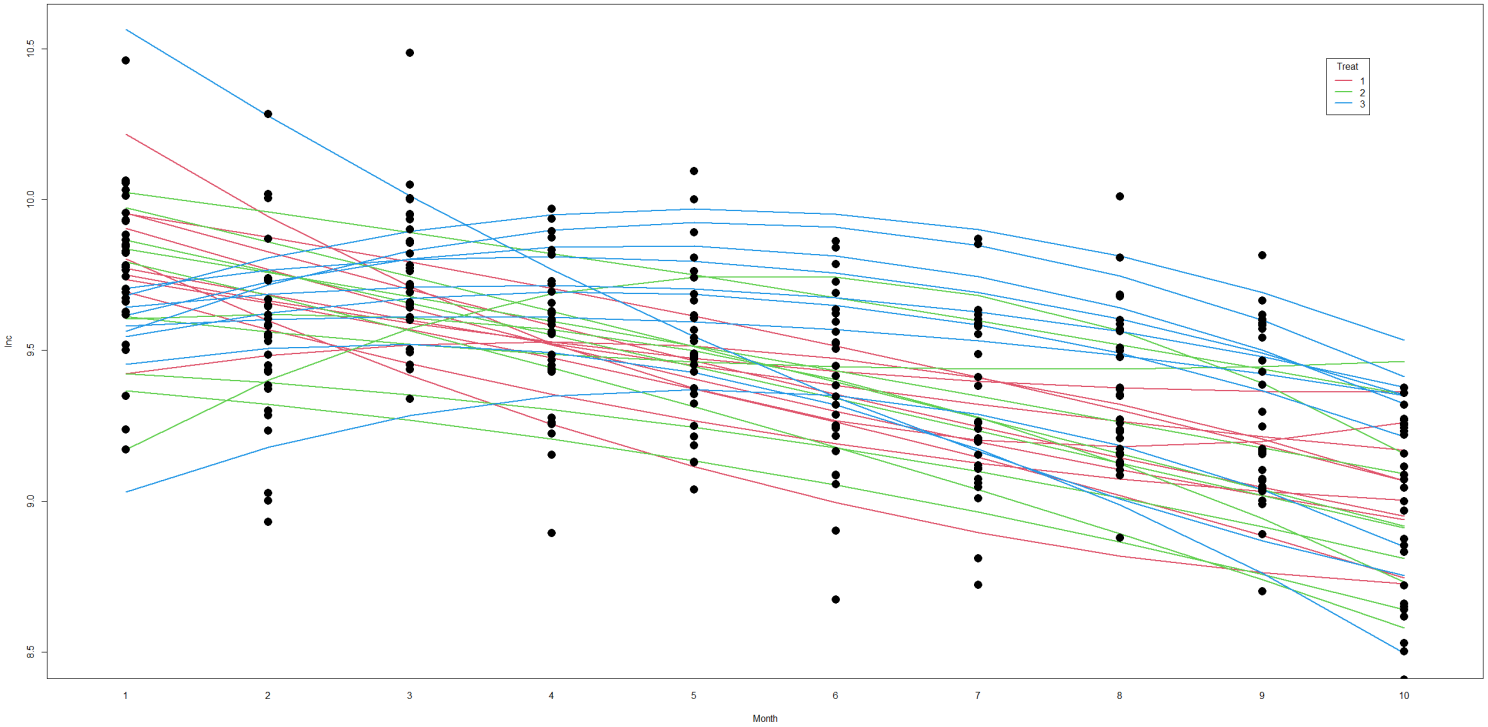
Rats data analyzed via a summary measure - plotting curves

```
# Individual curves: Look similar apart from one of the blue
rats$monthQ2 <- rats$monthQ^2
fit3 <- unlist(by(rats, rats$cage,
   function(x) fitted.values(lm(lnc ~ monthQ+monthQ2, data=x))))
names(fit3) <- NULL


#plotting the fit by cage
interaction.plot(rats$monthQ, rats$cage, fit3,
  xlab="Month", ylab="lnc",legend=F,
  col=as.numeric(SlopeData$treatm)+1,lty=1,lwd=2,)


lines(rats$monthQ,rats$lnc, type="p",pch=16,cex=2)
legend(locator(1),
       legend = c("1","2","3"), title = "Treat",
       lty = c(1,1,1),lwd=2,col= 2:4)
```

Rats data analyzed via a summary measure - plotting curves



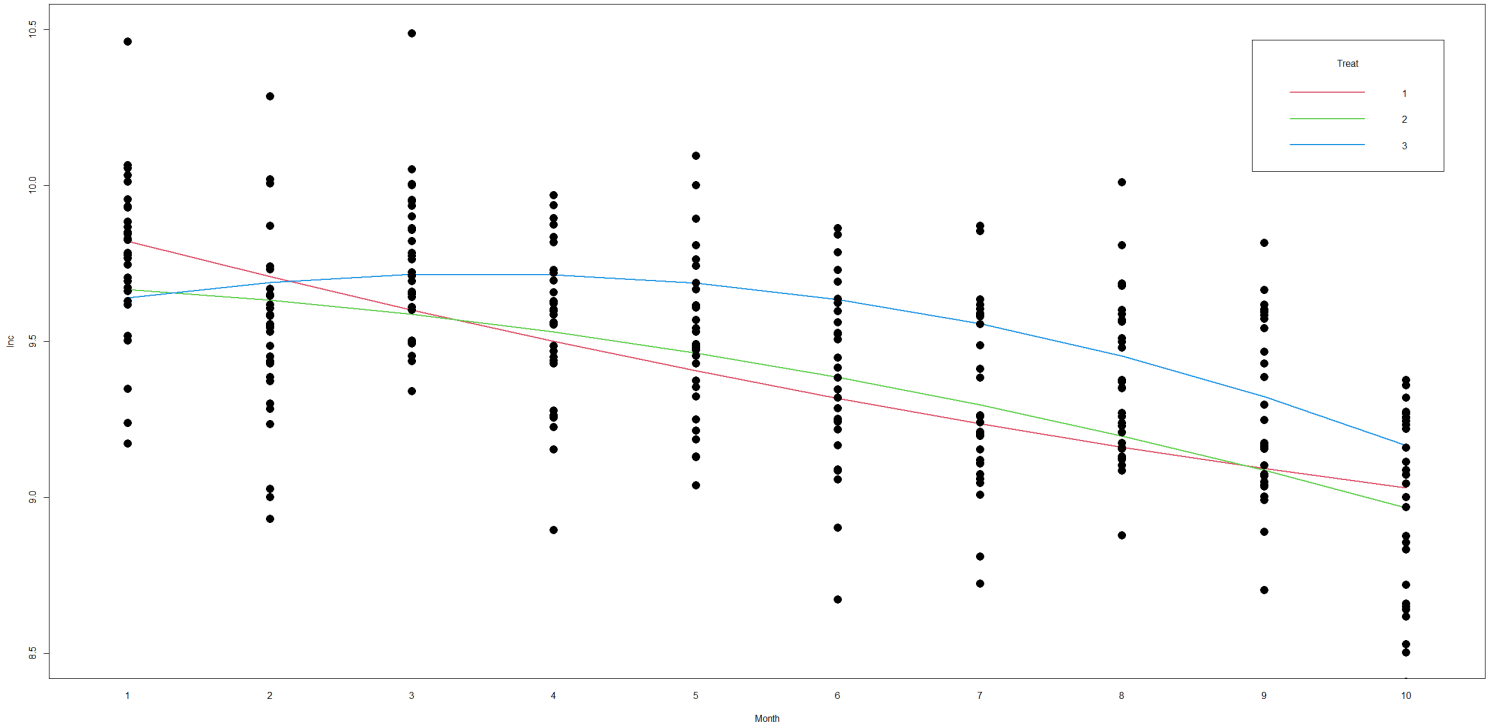
Rats data analyzed via a summary measure - plotting curves

```
# Treatment average curves:
fit4 <- unlist(by(rats, rats$treatm,
  function(x) fitted.values(lm(lnc ~ monthQ+monthQ2, data=x))))
names(fit4) <- NULL

#plotting the fit by treatment
interaction.plot(rats$monthQ, rats$treatm, fit4,
  xlab="Month", ylab="lnc",legend=F,
  col=2:4,lty=1,lwd=2,ylim=c(8.5,10.5))

lines(rats$monthQ,rats$lnc, type="p",pch=16,cex=2)
legend(locator(1),
       legend = c("1","2","3"), title = "Treat",
       lty = c(1,1,1),lwd=2,col= 2:4)
```

Rats data analyzed via a summary measure - plotting curves



Rats data analyzed via a summary measure - curvatures

```
byCage2 <- coef(lm(lnc ~ -1 + cage + monthQ:cage +
                     monthQ2:cage, data = rats))
#I want a data frame, one row for each cage
curve <- data.frame(matrix(byCage2, nrow=30, byrow=F))
names(curve) <- c("Intercept", "monthQ", "monthQ2")
#We also need info about cage and treat
CurveData <- cbind(DataMean[ , 1:2], curve)
head(CurveData)

##   cage treatm Intercept    monthQ    monthQ2
## 1    1      1   10.0465 -0.146171  0.0035462
## 2    3      1   10.0886 -0.135343  0.0021591
## 3    5      1    9.8302 -0.142721  0.0060042
## 4    7      1   10.0237 -0.068641 -0.0026996
## 5    9      1    9.8664 -0.096853  0.0027148
## 6   11      1    9.8445 -0.099881  0.0051576
```

Rats data analyzed via a summary measure - curvatures (contd)

Each estimate of the curve is used as the response

```
reg4 <- lm(monthQ2 ~ treatm, data = CurveData)
summary(reg4)

## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.00320    0.00298    1.08  0.29185
## treatm2     -0.00854    0.00421   -2.03  0.05237 .
## treatm3     -0.01608    0.00421   -3.82  0.00071 ***
##
## Residual standard error: 0.00941 on 27 degrees of freedom
## Multiple R-squared:  0.351,Adjusted R-squared:  0.303
## F-statistic:  7.3 on 2 and 27 DF,  p-value: 0.00291
```

- For Treat 1 the estimated curve is positive (nearly zero 0.003)
- For Treat 2 the estimated curve is negative (0.003-0.009=-0.006)
- For Treat 2 the estimated curve is also negative (0.003-0.016=-0.013)

Rats data analyzed via a summary measure - curvatures (contd)

```
anova(reg4)

## Analysis of Variance Table
##
## Response: monthQ2
##           Df  Sum Sq  Mean Sq F value Pr(>F)
## treatm     2 0.00129 0.000647     7.3 0.0029 **
## Residuals 27 0.00239 0.000089
```

When using the curvature as the response we get a significant difference between the three groups ($p = 0.003 < 0.05$)

Overview

Random effects model - simple version

- This model uses all observations instead of reducing to one observation per individual
- We can test the time-by-treatment interaction
- Add "individual" as a random effect
- Makes measurements on same individual correlated
- Unfortunately equally correlated no matter if they are "close" or "far apart"
- Can be considered first step in modeling the actual covariance structure
- Usually only good for short series

Rats data analyzed via random effects approach

- The model can now be enhanced to:

$$\texttt{lnc}_i = \mu + \alpha(\texttt{treatm}_i) + \beta(\texttt{month}_i) + \gamma(\texttt{treatm}_i, \texttt{month}_i) + d(\texttt{cage}_i) + \varepsilon_i \ ,$$

- The covariance structure of this model is:

$$\text{cov}(y_{i_1}, y_{i_2}) = \begin{cases} 0 & , \text{ if } \text{cage}_{i_1} \neq \text{cage}_{i_2} \text{ and } i_1 \neq i_2 \\ \sigma_d^2 & , \text{ if } \text{cage}_{i_1} = \text{cage}_{i_2} \text{ and } i_1 \neq i_2 \\ \sigma_d^2 + \sigma^2 & , \text{ if } i_1 = i_2 \end{cases}$$

Rats data analyzed via random effects approach

- The model can now be enhanced to:

$$\texttt{lnc}_i = \mu + \alpha(\texttt{treatm}_i) + \beta(\texttt{month}_i) + \gamma(\texttt{treatm}_i, \texttt{month}_i) + d(\texttt{cage}_i) + \varepsilon_i \ ,$$

- The covariance structure of this model is:

$$\text{cov}(y_{i_1}, y_{i_2}) = \begin{cases} 0 & , \text{ if cage}_{i_1} \neq \text{cage}_{i_2} \text{ and } i_1 \neq i_2 \\ \sigma_d^2 & , \text{ if cage}_{i_1} = \text{cage}_{i_2} \text{ and } i_1 \neq i_2 \\ \sigma_d^2 + \sigma^2 & , \text{ if } i_1 = i_2 \end{cases}$$

One can think of the model as:
log(count)="mean"+"between cage variation" + "within cage variation"

Multilevel model structure

- This model is also called a two-level model.

Level	1	2
Unit	Each separate obs.	Cages
Variation	Within cage σ^2	Between cage σ_d^2
Covariates	month month:treatm	treatm

Model synonyms

- Two-level model
- Mixed model with random subject level
- Mixed model with random intercept
- Model with compound symmetry correlation structure
- Model with exchangable correlation structure

Compound symmetry

The model implies that all observations from the same cage are correlated with the intra-class correlation:

$$Corr(Y_{i_1}, Y_{i_2}) = \frac{cov(Y_{i_1}, Y_{i_2})}{\sqrt{var(Y_{i_1})}\sqrt{var(Y_{2_1})}} = \frac{\sigma_d^2}{\sigma_d^2 + \sigma^2}$$

Here $cage_{i_1} = cage_{i_2}$ and $i_1 \neq i_2$.
We are not taking the distance between observations into account.
Maybe too simple? Perhaps observations close in time are more similar than observations far apart.
This correlation structure is called exchangable or compound symmetry.

Potential mistakes when leaving the random effect out

Level	Unit	Covariates
1	Each separate obs.	month month:treatm
2	Cages	treatm

If a random effect is present:

- Potential bias in the mean (ignoring that observations go together)
- Estimates on Level 1 may have too much variation. Ignoring pairs $\rightarrow$ too high p-values. Effects can be overlooked
- Estimates on Level 2 may have too little variation $\rightarrow$ too small p-values. "Noise" can become an effect

Rats data analyzed via random effects approach in **R**

```
library(nlme)
model1 <- lme(lnc ~ month + treatm + month:treatm,
              random = ~1 | cage, data = rats)
anova(model1)

##                numDF denDF F-value p-value
## (Intercept)        1   243   85525  <.0001
## month              9   243      46  <.0001
## treatm             2    27       3  0.0557
## month:treatm      18   243       2  0.0059
```

Rats data analyzed via random effects approach in **R**

```
library(nlme)
model1 <- lme(lnc ~ month + treatm + month:treatm,
              random = ~1 | cage, data = rats)
anova(model1)

##                numDF denDF F-value p-value
## (Intercept)        1   243   85525  <.0001
## month              9   243      46  <.0001
## treatm             2    27       3  0.0557
## month:treatm      18   243       2  0.0059
```

So we have a significant interaction between treatment and month.

Table of fixed effects

```
#This time summary gives a lot of output!
#summary(model1)

#A table of fixed effects estimates
summary(model1)$tTable
```

```
##                     Value Std.Error  DF   t-value     p-value
## (Intercept)      9.874280  0.080856 243 122.12192 4.1653e-220
## month2          -0.282870  0.087062 243  -3.24908  1.3213e-03
## month3          -0.199010  0.087062 243  -2.28585  2.3124e-02
## month4          -0.381285  0.087062 243  -4.37948  1.7692e-05
## month5          -0.464289  0.087062 243  -5.33288  2.2101e-07
......
## month7:treatm3   0.472180  0.123124 243   3.83500  1.6005e-04
## month8:treatm3   0.465124  0.123124 243   3.77770  1.9913e-04
## month9:treatm3   0.443817  0.123124 243   3.60464  3.7919e-04
## month10:treatm3  0.202159  0.123124 243   1.64192  1.0190e-01
```

Table of random effects

```
#The estimates random effects
VarCorr(model1)

## cage = pdLogChol(1)
##             Variance StdDev
## (Intercept) 0.027478 0.16577
## Residual    0.037899 0.19468

#intra class correlation
0.027478/(0.027478+0.037899)

## [1] 0.4203
```

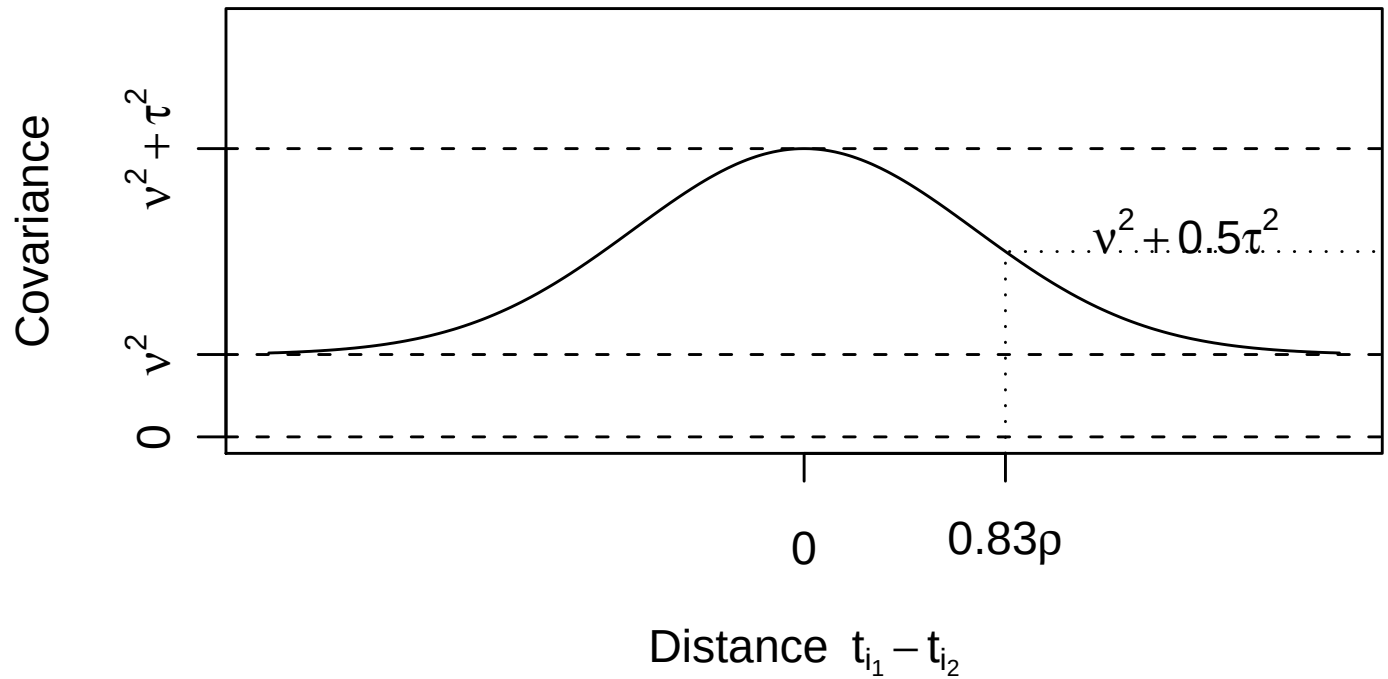
$$\hat{\sigma}_d^2 = 0.03 \text{ and } \hat{\sigma}^2 = 0.04$$

Overview

Advanced Covariance Structures: Gaussian spatial correlation

- *Spatial* covariance structures, depending on "how far" observations are apart.
- Gaussian decline of dependency:

$$Cov(Y_{i_1}, Y_{i_2}) = \begin{cases} 0 & , \text{ if } \texttt{indiv}_{i_1} \neq \texttt{indiv}_{i_2} \text{ and } i_1 \neq i_2 \\ \nu^2 + \tau^2 \exp\left\{\frac{-(t_{i_1} - t_{i_2})^2}{\rho^2}\right\} & , \text{ if } \texttt{indiv}_{i_1} = \texttt{indiv}_{i_2} \text{ and } i_1 \neq i_2 \\ \nu^2 + \tau^2 + \sigma^2 & , \text{ if } i_1 = i_2 \end{cases}$$



Rats data via spatial Gaussian correlation model

- Model for the entire observational vector:

$$
\begin{aligned}
\mathbf{Y} &\sim N(\mu, \mathbf{V}), \quad \text{where} \\
\mu_i &= \mu + \alpha(\texttt{treatm}_i) + \beta(\texttt{month}_i) + \gamma(\texttt{treatm}_i, \texttt{month}_i), \quad \text{and} \\
V_{i_1,i_2} &= \begin{cases} 0 & , \text{ if } \texttt{cage}_{i_1} \neq \texttt{cage}_{i_2} \text{ and } i_1 \neq i_2 \\ \nu^2 + \tau^2 \exp\left\{ \frac{-(\texttt{month}_{i_1} - \texttt{month}_{i_2})^2}{\rho^2} \right\} & , \text{ if } \texttt{cage}_{i_1} = \texttt{cage}_{i_2} \text{ and } i_1 \neq i_2 \\ \nu^2 + \tau^2 + \sigma^2 & , \text{ if } i_1 = i_2 \end{cases}
\end{aligned}
$$

- R code for model implementation:

```
analysis<-lme(lnc~month+treatm+month:treatm,
 random=~1|cage,
 correlation=corGaus(form=~as.numeric(month)|cage,nugget=T),
 data=rats)
```

- Partial R output:

```
Random effects:
 Formula: ~1 | cage
        (Intercept)  Residual
StdDev:   0.1404056 (= ν̂) 0.2171559 (= √(σ̂² + τ̂²))
Correlation Structure: Gaussian spatial correlation
Formula: ~as.numeric(month) | cage
 Parameter estimate(s):
    range    nugget
2.3863954 (= ρ̂²) 0.2186744   (= σ̂²/(σ̂² + τ̂²))
Number of Observations: 300
Number of Groups: 30
```

- Notice the R parametrization of the variance parameters

Other spatial correlation structures

- R has a number of build–in correlation structures. A few examples:

correlation=	Name	Correlation term
`sp(gau)(t)`	Gaussian	$\tau^2 \exp\{\frac{-(t_{i_1}-t_{i_2})^2}{\rho^2}\}$
`sp(exp)(t)`	exponential	$\tau^2 \exp\{\frac{-\lvert t_{i_1}-t_{i_2}\rvert}{\rho}\}$
`ar(1)`	autoregressive(1)	$\tau^2 \rho^{\lvert i_1-i_2\rvert}$
`un`	unstructured	$\tau^2_{i_1,i_2}$

How to select the correlation structure -Stationarity

- For all the listed covariance structures, the value of $Cov(Y_t, Y_{t+u})$ **does not depend on the time t**; only on the time difference u.
- We say that the error proces ε_t is weakly stationary of order 2;
- weakly because the stationarity is defined from moments, and not distributions (for Gaussian processes this is the same though);
- of order 2 because the stationarity is defined through 2nd order moments (variance, covariance).

The semi-variogram

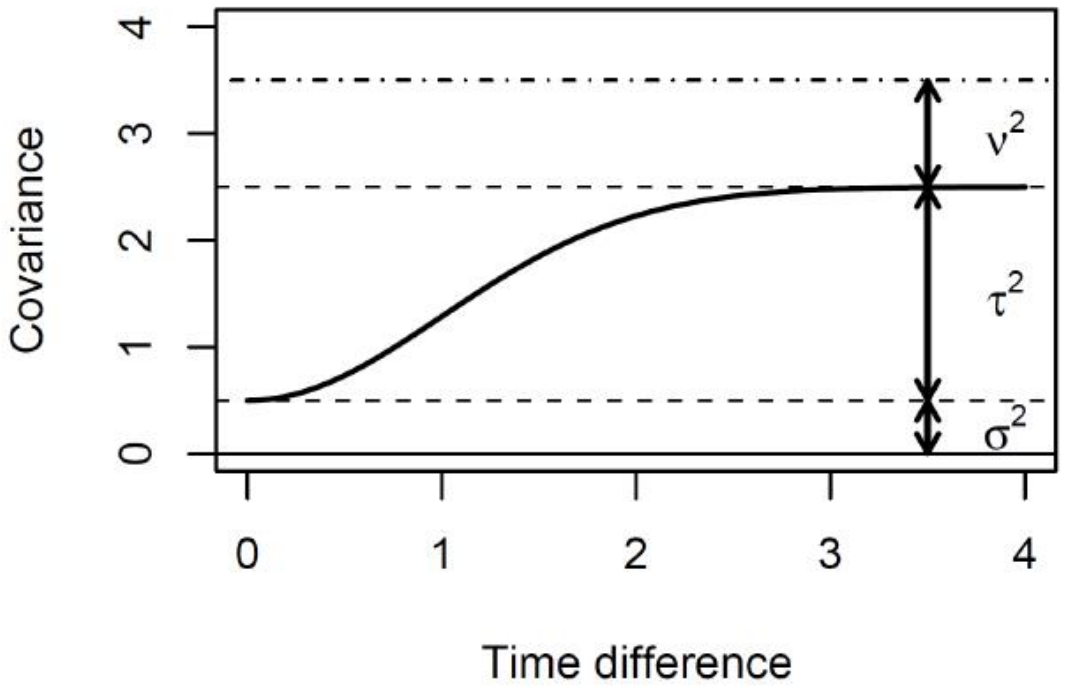
- The semi-variogram plots $\gamma(u) = \frac{1}{2}V(Y_t - Y_{t+u})$, $u > 0$:

$$\begin{aligned}
\gamma(u) =& \frac{1}{2}V(Y_t - Y_{t+u}) \\
=& \frac{1}{2}\left(V(Y_t) + V(Y_{t+u}) - 2Cov(Y_t, Y_{t+u})\right) \\
=& V(Y_t) - Cov(Y_t, Y_{t+u}) \\
=& \nu^2 + \tau^2 + \sigma^2 - \nu^2 - \tau^2\lambda(u) \\
=& \sigma^2 + \tau^2(1 - \lambda(u)),
\end{aligned}$$

where $\lambda(u) = exp(-u^2/\rho^2)$, $exp(-u/\rho)$, ρ^u for spatial Gaussian, exponential and AR1 correlation structures, respectively.

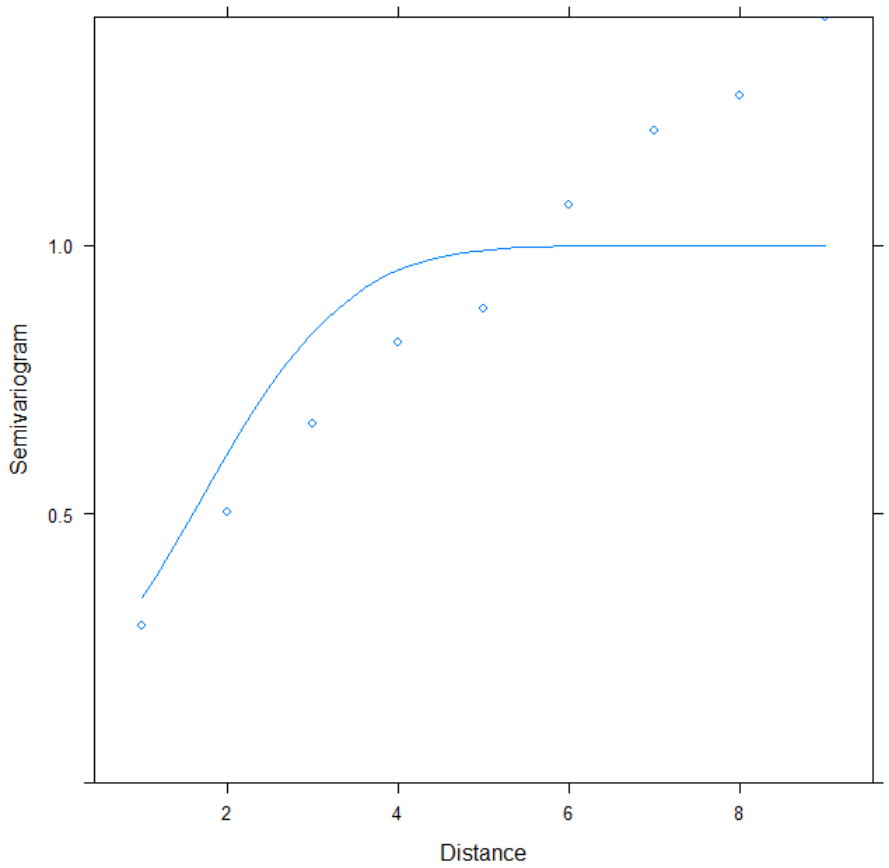
The semi-variogram

- Theoretical semi-variogram, Gaussian correlation structure:
- Plot of $\sigma^2 + \tau^2(1 - \lambda(u))$, where $\lambda(u) = exp(-u^2/\rho^2)$:



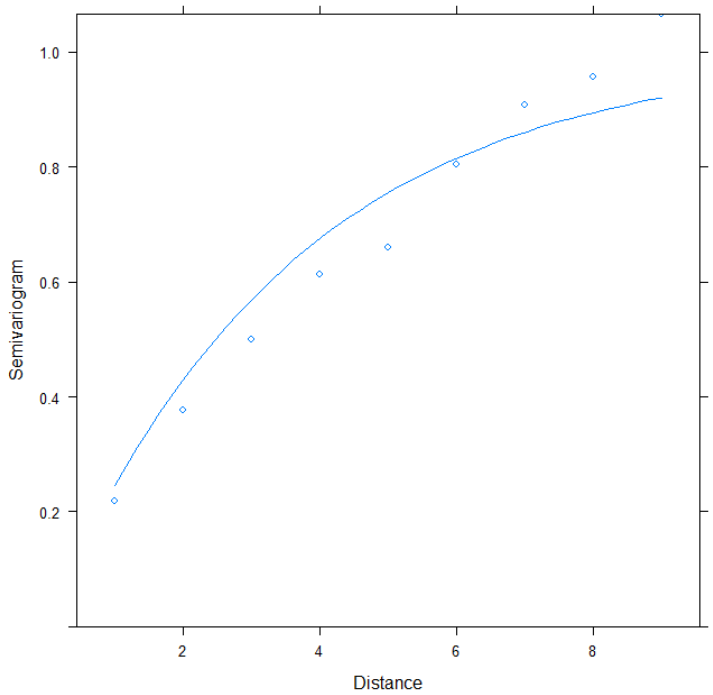
The semi-variogram

```
plot(Variogram(analysis,form=~monthQ|cage,data=rats))
```



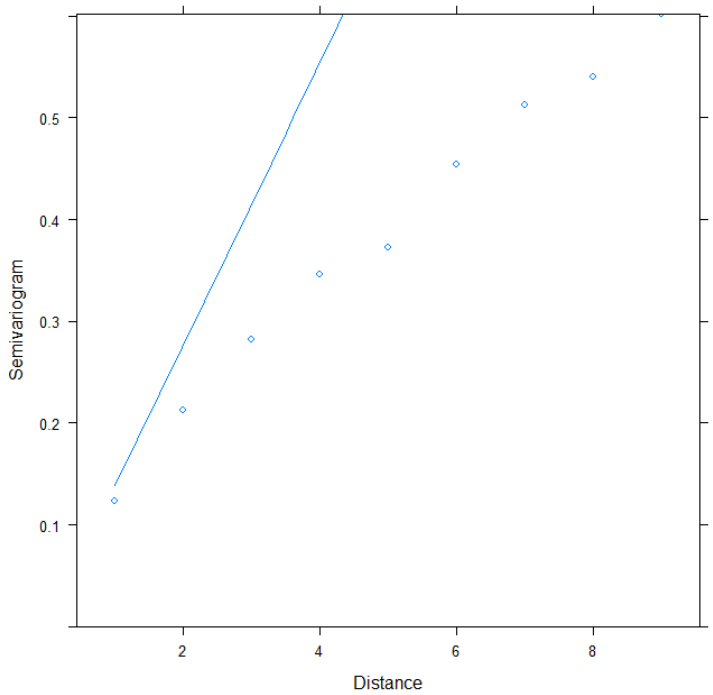
The semi-variogram, exponetially decreasing correlation

```
analysis2<-lme(lnc~month+treatm+month:treatm, random=~1|cage,
               correlation=corExp(form=~monthQ|cage,nugget=T),
               data=rats)
plot(Variogram(analysis2,form=~monthQ|cage,data=rats))
```



The semi-variogram, linearly decreasing correlation

```
analysis3<-lme(lnc~month+treatm+month:treatm, random=~1|cage,
                        correlation=corLin(form=~monthQ|cage),
                        data=rats)
plot(Variogram(analysis3,form=~monthQ|cage,data=rats))
```



The semi-variogram, summing up

- The semi-variogram **compares a theoretical and empircal correlation structure**. Similarity indicates a good model.
- For the *rats data*, the semi-variogram indicates that the optimal correlation structure seems to be an exponetially decreasing spatial correlation structure, rather than the standard spatial Gaussian structure.

The rats data, partial final model output

- Random effects:
  Formula: ~1 | cage
  (Intercept) Residual
  StdDev: 0.07828533 $(= \hat{\nu})$ 0.2510369 $(= \sqrt{\hat{\sigma}^2 + \hat{\tau}^2})$
  Correlation Structure: Exponential spatial correlation
  Formula: ~as.numeric(month) | cage
  Parameter estimate(s):
  range nugget
  3.556503e+00 $(= \hat{\rho})$ 3.341370e-08 $(= \hat{\sigma}^2/(\hat{\sigma}^2 + \hat{\tau}^2))$
  Number of Observations: 300
  Number of Groups: 30

Overview

Pros and cons of approaches

Separate analysis for each time–point

+ Not wrong
– Can be confusing
– Difficult to reach coherent conclusion
– In general not very informative

Analysis of summary statistic

+ Good method with few and easily checked assumptions
– Important to choose good summary measure(s)

Random effects approach - simple version

+ Good method for short series
+ Uses all observations
– Usually not good for long series

Random effects approach - qadvanced version

+ Works for short and long series
+ Uses all observations
– Requires appropriate choice of covariance decay function

Overview of this module

1. The repeated measurements setup
   - Aspirin Example
   - Activity of rats
2. Separate analysis for each time–point
   - Example: rats data
3. Analysis of a summary statistic
   - Example: rats data
4. Random effects model - simple version
   - Example: rats data
5. Random effects model - advanced version
6. Pros and cons of simple approaches

Exercise: Histamin in dogs

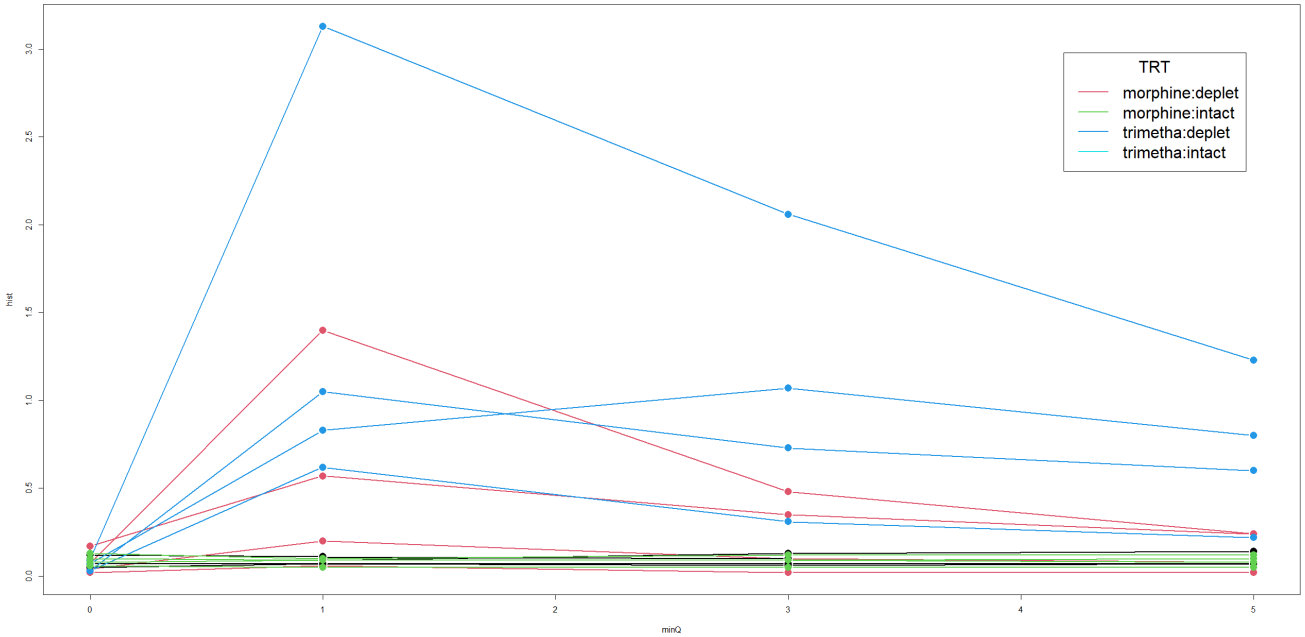
In an experiment with 16 dogs the blood histamine concentration was measured 0, 1, 3, and 5 minutes after injection of morphine or trimethaphane. Before injection the dogs were classified into two groups according to their level of histamine (intact or depleted).

```
histamin <- read.table("histamin.txt", header=T, sep=",", dec=".")
histamin$dog <- factor(histamin$dog)
histamin$minQ <- histamin$min
histamin$min <- factor(histamin$min)
histamin$TRT <- factor(paste(histamin$treatm,":",histamin$level,
                             sep=""))
#summary(histamin)
```

Histamin in dogs

```
plot(histamin$minQ,histamin$hist,xlab="minQ",ylab="hist",pch="")
for(i in 1:64){
  temp<-histamin[histamin$dog==i,]
lines(temp$minQ,temp$hist,col=(1:4)[temp$TRT[1]],lwd=2,type="b",
  pch=16,cex=2)
  }
legend(locator(1),
       legend = levels(histamin$TRT), title = "TRT",
       lty = c(1,1,1,1),lwd=2,col= 2:5,cex=2)
```

Histamin in dogs



Histamin in dogs - questions

First of all use the `TRT` factor for the analysis (defining 4 groups of dogs).

1. Make some plots of the data, for instance one line for each dog (maybe colored differently in each TRT group).
2. Analyze these data using one or more of "the simple methods".
3. How would you approach making a conclusion for the "real" treatment: morphine vs. trimetha?