Dependent data data: mixed effect Models

Jan van den Broek

January, 2024

Universiteit Utrecht

Introduction

Dependent data

From 79 cows the protein content in the milk was measured in the several weeks following calving.

cow i: the protein content from the first week y_{i1} and the protein content of the second week y_{i2}.

These measurements are continuous variables for which the normal distribution can be used. The models to be considered are linear models.

two observations are taken from the same cow which is the sampling unit. These observations are likely to be dependent

Dependent data

1. The first observation contains information about the observation in the second week, since both are from the same cow.
2. In the case that for each sampling unit two observations are obtained, the probability distribution must deal with two observations (variables) at the same time. This probability distribution is called a bi-variate probability distribution.

Dependent data

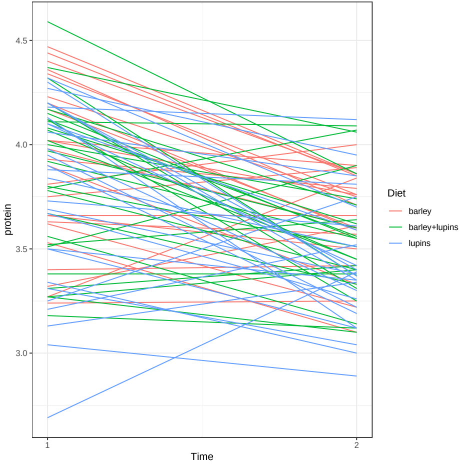
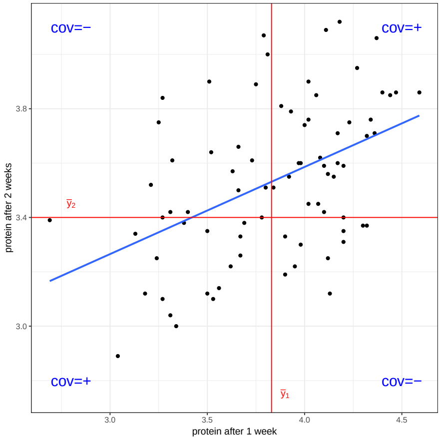


Figure 1: Protein content in the milk of cows in the 2 weeks following calving

Covariance

$$cov(y_1, y_2) = \frac{1}{n-1}\sum_{i=1}^{n}(y_{i1} - \bar{y}_1)(y_{i2} - \bar{y}_2)$$



Correlation

The size of the covariance has no meaning since it depends on the scale on which the variables are measured.

If for instance a variable is measured in kg and that is changed to gram, then the covariance will change, it is multiplied by 1000. This is a drawback of the covariance.

In order to get a measure that is scale independent, one can divide the covariance by the standard deviations of y_1 and y_2:

$$r = \frac{cov(y_1, y_2)}{s_{y_1} s_{y_2}}$$

The correlation between y_1 and y_2 is calculated as 0.46

t-tests

Compare the independent analysis case to the dependent one.

Variable y_1 is measured from cows from group number 1 and y_2 from cows in group number 2. In that case a two sample t-statistic could be calculated:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

It is often assumed that the population variances in both groups are the same.Both variances in the denominator are then replaced by s^2.

The outcome of this t-statistic is 5.4 based on $78 + 78 - 2 = 154$ degrees of freedom (One cow had missing values). ($\bar{y}_1 = 3.83, \bar{y}_2 = 3.53, sd_1 = .401, sd_2 = .284$)

t-test

If the data is considered dependent, then the paired t-statistic is calculated.

For each cow the values for y_2 and y_1 are subtracted to get the difference: $d_i = y_{1i} - y_{2i}$. The mean of this difference is calculated:

$$
\begin{aligned}
\bar{d} &= \frac{1}{n}\sum_{i=1}^{n} d_i \\
&= \frac{1}{n}\sum_{i=1}^{n}(y_{1i} - y_{2i}) \\
&= \frac{1}{n}\sum_{i=1}^{n} y_{1i} - \frac{1}{n}\sum_{i=1}^{n} y_{2i} \\
&= \bar{y}_1 - \bar{y}_2
\end{aligned}
$$

t-tests

and the paired t-statistic then is :

$$t = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{n}}} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{s_d^2}{n}}}$$

Comparing these t-statistics one can see that the difference is in the denominator, it is in the way the variances are calculated and thus also giving different degrees of freedom.

t-test

In general the formula for calculating the variance of the difference of two dependent variables is :

$$var(y_1 - y_2) = var(y_1) + var(y_2) - 2 \cdot cov(y_1, y_2)$$

The variances of the differences $d_i = y_{1i} - y_{2i}$ can be estimated by using the formula above or by calculating the variance of the difference s_d^2 wich is 0.37 giving a t-statistic of 7.13 based on 77 degrees of freedom. where s_d^2 is the variance of the differences d_i.

conclusion

So the influence of the dependencies in the data on the data-analysis is shown in the variance: the means and the difference in means are calculated the same in the independent case as in the dependent one but the difference is in the variance.

conclusion

- Observations on the same sampling unit are very likely dependent
- This (linear) dependence in the case of continuous data can be measured with the correlation coefficient
- The dependence in the data is influencing the data-analysis through the variance. With dependent data the variances are different as compared to the independent case and thus are the degrees of freedom also different.

Shared random effects

Shared random effects

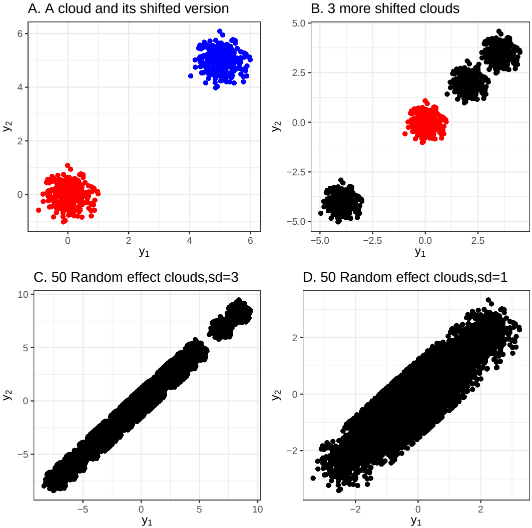


Figure 3: Fixed and random effects

Shared random effects

The coordinates of these shifted clouds have a number in common, one cloud the number 5, an other cloud the number -3 etc. This is the case for the linear models where there are different groups of observations.

All observations within a group have there group mean in common. In that case one can calculate a group effect for every group.

If two variables y_1 and y_2 share a fixed effect then that does not change there independence.

Shared random effects

What does it mean when the outcome of a variable is considered to random?

To get an idea of what random means, think of throwing a die.

Considering the outcome of a variable as random means, that all possible outcomes with there attached probabilities are taken into account .

Shared random effects

Adding a random number to both variables means we have to take all possible outcomes of that variable with there attached probabilities into account and average over it.

Shared random effects

If the same random number is added to two variables, we have to take all possible values of that random number with their attached probability into account, which means we have to take all possible clouds into account and thus have to consider the resulting stretched cloud which implies correlation between the two variables. Or differently, if two variables have a random effect in common, they will be correlated.

If two variables y_1 and y_2 share a random effect then they will be correlated, because then all possible values of the random effect have to be taken into account resulting in a stretched cloud.

Shared random effects

If the standard deviation of the distribution of the random effects is reasonable small as in part D, then these random effects will be concentrated around zero resulting in a condensed cloud.

In that case the correlation between the resulting variables will not be large.

If, however the standard deviation of the distribution of the random effects is reasonable large as in part C. then these random effects will be mostly around zero but large positive or large negative values are also possible resulting in a stretched cloud.

In that case the correlation of the resulting variables will be large.

Shared random effect

The variance of the resulting variable, the variable with the random effects added, will be the variance the original variable had plus the variance of the random effect.

This is how the dependence of the data is modeled, by using random effects and the variance of these random effects determine the height of the correlation.

Shared random effects

There are two observations for each cow.

A linear model could be used for this situation, treating the cow numbers as groups and thus have two observations within each group.

A linear model for this situation might be $y_{ij} = \mu_i + \epsilon_{ij}$, where μ_i is the mean of cow number i and ϵ_{ij} is the residual, the difference between observation y_{ij} and the group mean.

Shared random effects

Another formulation of this model is $y_{ij} = \mu + (\mu_i - \mu) + \epsilon_{ij}$, where μ is the overall mean and $(\mu_i - \mu)$ is the i th group (cow) effect.

Using different notation this model can be written as: $y_{ij} = \beta_0 + b_{0i} + \epsilon_{ij}$, where β_0 is the overall mean, b_{0i} is the effect of group number i and $\beta_0 + b_{0i} = \mu + (\mu_i - \mu) = \mu_i$, the mean of group i.

This group effect depends on the overall mean and the group mean, so is the same for each observation y_{ij} with each group.

Or, to put it different each observation within a group share the same group effect. If this group effect is treated as fixed and we thus have 79 different cow groups, then this does not change the independence.

Shared random effects

If however, the group effects b_{0i} are random drawings from a normal distribution with mean zero and variance σ_b^2,

then we have to take all possible values (infinitely many) of that random effect with their attached probabilities into account and as a result we have to consider one big stretched cloud.

This means that the resulting variables are correlated.

Shared random effects

The random effects have mean zero. So most of the randomly drawn effects will be around zero.

The variance determines the range of possible random effects. If the variance is large then a stretched cloud is obtained, and there will be a high correlation, when the variance is low, the cloud will be more condensed and the correlation will be low.

Shared random effects

The random effect b_{0i} depends on i and thus on the cow.

This means that there is one random effect for all the observations on each cow, these random effects having the the same normal distribution.

This is also called a **shared random effects model**.

More observations per sampling unit

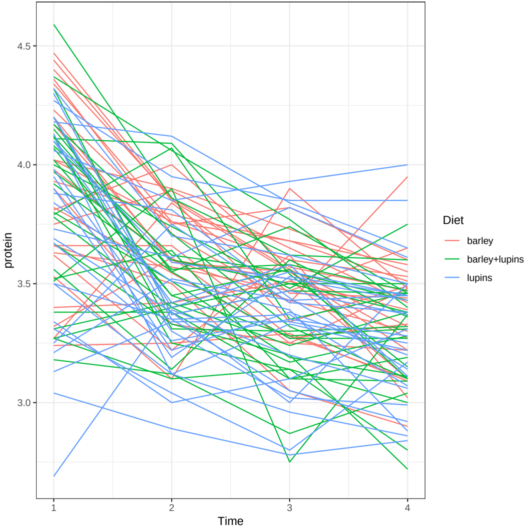


Figure 4: Protein content in the milk of cows in the 4 weeks following

More observations per sampling unit

The anova model is $y_{ij} = \beta_0 + b_{0i} + \epsilon_{ij}$, where y_{ij} is the protein level of cow i at week j, β_0 is the overall mean, b_{0i} is the effect of cow i and ϵ_{ij} is the residual of observation y_{ij}.

Use the linear model for the case where cow is the grouping variable and there are 4 observations for each cow to write the 4 observations for cow number i as:

$$\begin{array}{ll} y_{i1} = \beta_0 + b_{0i} + \epsilon_{i1} & y_{i2} = \beta_0 + b_{0i} + \epsilon_{i2} \\ y_{i3} = \beta_0 + b_{0i} + \epsilon_{i3} & y_{i4} = \beta_0 + b_{0i} + \epsilon_{i4} \end{array}$$

More observations per sampling unit

This is the linear model for the one way anova case.

In this case the b_{0i} are fixed numbers. It represents the deviation for cow i from the overall mean.

Or, the mean for cow i is $\beta_0 + b_{0i}$, just as in the case of two observations above.

Because the observations for cow i have a fixed effect in common they are modeled as independent.

Since in that case the cow effects are fixed, conclusions from the study are limited to the cows in the study.

More observations per sampling unit

When the b_{0i} are treated as being drawn from a $N(0, \sigma^2_{int})$-distribution, this is they are random numbers, then the observations share a random effect and as a consequence, by considering all possible outcomes of the random effect, the observations on the same cow are dependent.

The variance between the cows is σ^2_{int} where int stands for intercept.

More observations per sampling unit

The correlation between any two observations on the same cow is the same, because the drawings are from the same distribution, with the same variance, so the stretching is the same. This correlation structure is called the **exchangeable correlation structure** or **compound symmetry**.

Since the cow effects are random, one can regard the cows in the study as a random sample from a population of cows. This means that the conclusion from this study can be generalized to the population of cows from which the cows in the study were sampled.

The fixed effect part of this model, β_0, is the mean protein level of an average cow. An average cow is a cow with random effect zero.

The random intercept model

The random intercept model

Let's extend the model with the random effects above to include a linear time effect:

$$y_{ij} = \beta_0 + b_{0i} + \beta_1 Time + \epsilon_{ij}$$

Just as in a regression model, time is treated as a continuous variable and β_1 is the regression coefficient. In this model there is only one regression coefficient so the decline is modeled the same for every cow. The intercepts, however, are different for every cow.

The random intercept model

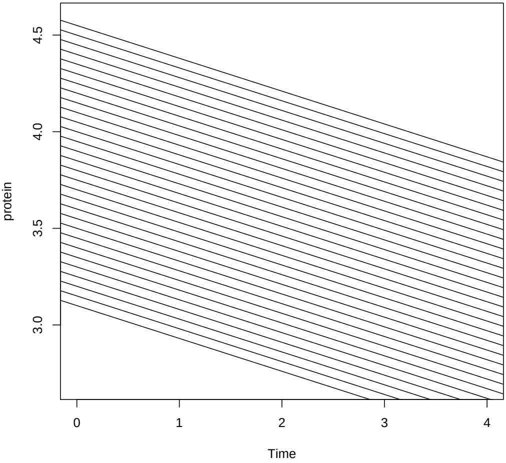


Figure 5: Different intercepts, same slope

The random intercept model

Time is treated as a continuous variable and β_1 is the slope.

In this model there is only one regression coefficient so the decline is modeled the same for every cow.

The intercepts, however, are different for every cow.

Recall the model for the analysis of covariance :
$y_{ij} = \alpha + (\alpha_i - \alpha) + \beta Time + \epsilon_{ij}$.

The general intercept is α and $(\alpha_i - \alpha)$ is the effect of group number i and $\alpha + (\alpha_i - \alpha) = \alpha_i$ is the mean of group number i.

The random intercept model

Now replace α by β_0 and $(\alpha_i - \alpha)$ by b_{0i} and the model $y_{ij} = \beta_0 + b_{0i} + \beta_1 Time + \epsilon_{ij}$ is obtained where the cows are the groups.

In this model the intercept for cow i is $\beta_0 + b_{0i}$ where β_0 is the general mean of the intercepts and b_{0i} is the deviation from the general intercept for cow i.

So this model has different intercepts for the cows and for every cow the same slope. So for every cow different lines are fitted but the lines are parallel.

The random intercept model

If the b_{0i} are treated as random, then the observations on the same cow have a random number in common and are thus correlated.

The correlation is the same for every pair of observations from the same cow (exchangeable correlation structure).

This model is called **a random intercept model**.

The random intercept model

the model contains a random part and a fixed part:

Random part $b_{0i} + \epsilon_{ij}$, the random cow intercepts and the random residuals.

fixed part $\beta_0 + \beta_1 Time$, the general intercept and the regression on time.

One recognizes the fixed part as an ordinary regression model.

This fixed part is a description for the average cow, for a cow with random effect zero.

The random intercept model

There were 3 diet-groups: add a grouping variable (2 indicator variables) to the linear part of the model.

- The model contains an intercept and time as a continuous variable. So the fixed part is a regression model.
- One can add the grouping variable diet to the regression model above. The fixed part is then an analysis of covariance model.

The random intercept model

- To the analysis of covariance model the interaction between time and diet can be added. The fixed part is then a model with different intercepts and different slopes.
- Instead of treating time as a continuous variable, one can treat time as a grouping variable. One then has two grouping variables in the model: diet and time. This is a two-way anova model or a two way factorial model.
- In the model with the two grouping variables one can add the interaction between diet and time. This is a full two way factorial model.

The random intercept model

get the data

```
milk <- nlme::Milk[nlme::Milk$Time<5,]
```

For the t test use the function t.test() and use ?t.test for help. \
To make the plot with two time points:

```
library(ggplot2)
ggplot(milk,aes(x=Time,y=protein,group=Cow))+
  geom_line(aes(color=Diet))+
  scale_x_continuous(breaks=c(1,2,3,4))+
  theme_bw()

fit <- lme4::lmer(protein~factor(Time)+
      factor(Diet)+factor(Diet):factor(Time)+
      (1|Cow),REML=FALSE,data = milk)
```

The random intercept model


```
summary(fit,correlation=FALSE)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: protein ~ factor(Time) + factor(Diet) + factor(Diet):factor(Time) +
##     (1 | Cow)
##    Data: milk
##
##      AIC      BIC   logLik deviance df.resid
##     89.8    142.3    -30.9     61.8      301
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4965 -0.5184  0.0440  0.5236  2.3195
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Cow      (Intercept) 0.03841  0.1960
##  Residual             0.05015  0.2239
## Number of obs: 315, groups:  Cow, 79
##
## Fixed effects:
##                                          Estimate Std. Error t value
## (Intercept)                              3.886800   0.059517  65.306
## factor(Time)2                           -0.247602   0.064147  -3.860
## factor(Time)3                           -0.388800   0.063339  -6.138
## factor(Time)4                           -0.510400   0.063339  -8.058
## factor(Diet)barley+lupins               -0.025689   0.082596  -0.311
## factor(Diet)lupins                      -0.128652   0.082596  -1.558
## factor(Time)2:factor(Diet)barley+lupins -0.073509   0.088485  -0.831
## factor(Time)3:factor(Diet)barley+lupins -0.126756   0.087901  -1.442
## factor(Time)4:factor(Diet)barley+lupins -0.072933   0.087901  -0.830
## factor(Time)2:factor(Diet)lupins        -0.082769   0.088485  -0.935
```

Random coefficients model

Random coefficients model

Instead of taking different intercepts for every cow, one could take the slopes different per cow:

$$y_{ij} = \beta_0 + (\beta_1 + b_{1i})\,Time + \epsilon_{ij} = \beta_0 + \beta_1\,Time + b_{1i}\,Time + \epsilon_{ij}$$

Now there is only one intercept but every cow has her own slope: $\beta_1 + b_{1i}$.

The 4 observations for cow i can now be described as:

$$\begin{array}{ll} y_{i1} = \beta_0 + (\beta_1 + b_{1i})\,Time + \epsilon_{i1} & y_{i2} = \beta_0 + (\beta_1 + b_{1i})\,Time + \epsilon_{i2} \\ y_{i3} = \beta_0 + (\beta_1 + b_{1i})\,Time + \epsilon_{i3} & y_{i4} = \beta_0 + (\beta_1 + b_{1i})\,Time + \epsilon_{i4} \end{array}$$

Random coefficients model

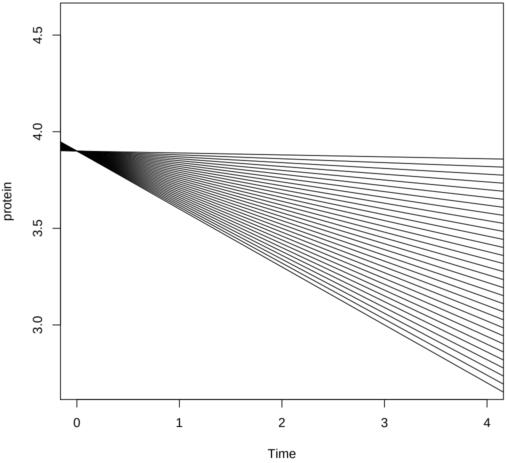


Figure 6: The same intercepts, different slopes

Random coefficients model

The b_{1i} can be treated as random numbers from a $N(0, \sigma^2_{slope})$-distribution,

where σ^2_{slope} is the variance between the slopes of the cows.

In that case the observations from the same cow have a random component in common and thus they are correlated.

Random coefficients model

But the random effects b_{1i} do not get the same weight, as can be seen from the model.

How heavily b_{1i} is weighted depends on time. Suppose the time points are 1, 2, 3 and 4. Then:

y_{i1} depends on $1 \times b_{1i}$

y_{i2} depends on $2 \times b_{1i}$

y_{i3} depends on $3 \times b_{1i}$

y_{i4} depends on $4 \times b_{1i}$

Random coefficients model

The b_{1i} can be treated as random numbers from a $N(0, \sigma^2_{slope})$-distribution, where σ^2_{slope} is the variance between the slopes of the cows.

In that case the observations from the same cow have a random component in common and thus they are correlated.

The weight of the random b_{1i} on the observation gets larger as the time progresses. This means that the observations on the same cow are correlated but this correlation depends on time.

So, with this model the correlation between y_{i1} and y_{i2} and the correlation between y_{i1} and y_{i3} are different because the times at which y_{i2} and y_{i3} are measured are different.

Usually the correlation between two observations decreases if they are further apart in time.

Random coefficients model

So with the random part one models the correlation structure: if only random intercepts are taken the correlation between the observations is constant (exchangeable), with random coefficients the correlations depend on time.

Random coefficients model

Here the model also consists of a random and a fixed part:

Random part $b_{1i}\,Time + \epsilon_{ij}$, the random cow slopes and the random residuals.

Fixed part $\beta_0 + \beta_1\,Time$, the general intercept and the regression on time.

Random coefficients model

In a mixed effect model the random part models the correlation structure and the fixed part models the patterns in the data.

random intercepts + random coefficients

Model with random intercepts and random coefficients

A model with a different intercept and different slope for each cow is:

$$y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{1i})Time + \epsilon_{ij} = (\beta_0 + b_{0i}) + \beta_1 Time + b_{1i} Time + \epsilon_{ij}$$

Now every cow has her own intercept $\beta_0 + b_{0i}$ and her own slope: $\beta_1 + b_{1i}$. The 4 observations for cow i can now be described as:

$$\begin{aligned} y_{i1} &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})Time + \epsilon_{i1} \\ y_{i2} &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})Time + \epsilon_{i2} \\ y_{i3} &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})Time + \epsilon_{i3} \\ y_{i4} &= \beta_0 + b_{0i} + (\beta_1 + b_{1i})Time + \epsilon_{i4} \end{aligned}$$

models with random intercepts and random coefficients

The b_{0i} are treated as a draw from a $N(0, \sigma^2_{int})$-distribution, where σ^2_{int} is the variance between the intercepts per cow.

The b_{1i} can be treated as a random number from a $N(0, \sigma^2_{slope})$-distribution, where σ^2_{slope} is the variance between the slopes of the cows.

In that case the observations from the same cow have two random components in common, one of which is weighted by time.

The observations from the same cow are thus correlated and the correlation structure again depends on time.

The correlation structure in this model is more complex as compared to the random coefficient model.

Model with random intercepts and random coefficients

Again the model consists of a random and a fixed part:

Random part $b_{0i} + b_{1i}\,Time + \epsilon_{ij}$, the random cow intercepts and slopes and the random residuals.

fixed part $\beta_0 + \beta_1\,Time$, the general intercept and the regression on time.

The fixed part in this model is again just a regression model. Also here a linear model with diet can be used.

And time can be taken as a grouping variable and interactions can be added, just as with the other models.

Model with random intercepts and random coefficients

In a mixed effect model the random part models the correlation structure and the fixed part models the patterns in the data.

Fitting the models

Fitting the models

Write down the likelihood for each model discussed above, and maximize this likelihood with respect to all the parameters in the model in order to obtain the maximum likelihood estimates.

If these estimates are plugged into the likelihood then the maximum value of the likelihood is obtained for the model used.

One can then fit another model (by leaving out one of the terms) and obtain the maximum of the likelihood for that model.

Then these two models can be compared using the AIC or the likelihood ratio test.

This is just the standard likelihood procedure. This procedure works fine for the fixed effect part.

Fitting the models

Using the standard likelihood procedure here gives estimates for the different variance components that can be biased. For this reason the likelihood procedure is modified.

The likelihood is transformed in such a way that the transformed likelihood does no longer depend on the fixed part of the model.

Attention is restricted to the random effects part.

For that reason one calls this the restricted likelihood. This restricted likelihood is maximized to obtain the maximum likelihood estimates, now called restricted maximum likelihood estimates.

This procedure is known as restricted maximum likelihood estimation: REML

Fitting the models

Because the transformation used in REML depends on the fixed part of the model, a change in this fixed part gives another transformation and thus also a different random part.

If one then compares the models not only the change in the fixed part is measured but also a change in the random part. For this reason REML is not appropriate for testing the fixed part of the model.

Fitting the models

A reasonable procedure is to first fit the complete model and test the random part using REML.

After that, refit this model using maximum likelihood and test the fixed part. This procedure is illustrated below.

Fitting the models


```
mlk <- nlme::Milk[nlme::Milk$Time<5,]
fit <- lme4::lmer(protein~factor(Time)+factor(Diet)+
       factor(Diet):factor(Time)+(1+Time|Cow),data = mlk)
```

The fixed part of the model is stated in the usual way.

The random part of the model is the part between brackets: (1+Time|Cow). The part 1+Time shows the random part of the model.

The |Cow part indicates that the random effects are per cow, so cow is treated as factor. So the 1 in this model represents the random cow intercepts.

This is a model with random (cow) intercepts and random (cow) slopes.

Note that the 1 in (1+Time|Cow) can be left out so (Time|Cow) represents the same model.

Fitting the models

```
summary(fit,correlation=FALSE )
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: protein ~ factor(Time) + factor(Diet) + factor(Diet):factor(Time) +
##     (1 + Time | Cow)
##    Data: mlk
##
## REML criterion at convergence: 90.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.67360 -0.53186  0.03346  0.46926  2.39156
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Cow      (Intercept) 0.134877 0.36726
##           Time        0.007439 0.08625  -0.87
##  Residual             0.039697 0.19924
## Number of obs: 315, groups:  Cow, 79
##
## Fixed effects:
##                                     Estimate Std. Error t value
## (Intercept)                         3.886800   0.071183  54.603
## factor(Time)2                      -0.247943   0.059672  -4.155
## factor(Time)3                      -0.388800   0.066076  -5.884
## factor(Time)4                      -0.510400   0.076511  -6.671
## factor(Diet)barley+lupins          -0.025689   0.098786  -0.260
## factor(Diet)lupins                 -0.128652   0.098786  -1.302
## factor(Time)2:factor(Diet)barley+lupins -0.073169   0.082321  -0.889
## factor(Time)3:factor(Diet)barley+lupins -0.126756   0.091699  -1.382
## factor(Time)4:factor(Diet)barley+lupins -0.072933   0.106180  -0.687
## factor(Time)2:factor(Diet)lupins   -0.082428   0.082321  -1.001
```

Fitting the models

From the part Random effects it can be seen that the standard deviation between the intercepts is 0.367 (variance is 0.135) and between the slopes 0.086.

The correlation between these two estimates is -0.87.

This means that the two random effects are not modeled as independent. The random intercepts and the random slope have a bi-variate distribution instead of each having an independent normal distribution.

So this is a model with correlated random intercept and random slope. This is default in the lme4 library.

To specify a model in which the random intercept and slope are not correlated one can use (1+Time||Cow).

Fitting the models

Sometimes one finds that the correlation between the estimate for the random intercept and slope is -1.

This means that they cannot be distinguished. In that case the model cannot be estimated.

A remedy for this is to change the time scale in the random effects, by dividing it by an appropriate number. If that does not work one can try to fit a model with only random slopes or only random intercepts.

Fitting the models

correlation matrix:

corr.	time 1	time 2	time 3	time4
time 1	1.000000	0.612828	0.475646	0.274812
time 2	0.612828	1.000000	0.489545	0.344990
time 3	0.475646	0.489545	1.000000	0.412471
time 4	0.274811	0.344990	0.412471	1.000000

As can be seen the correlation between the protein values at time 1 and time 2 is 0.61, between times 1 and 3 it is 0.48 and between times 1 and 4 0.27.

This illustrates something seen more often in time series: the further away the time points are, the lower their correlation.

Fitting the models

The model is fitted using the REML procedure by default (REML=TRUE).

This is used to determine what model for the random part is best according to the data.

Next two other models are fitted with the same fixed effects but with different random effects.

Fitting the models

```
fit2 <- lme4::lmer(protein~factor(Time)+factor(Diet)+
        factor(Diet):factor(Time)+(1|Cow),data = mlk)
fit3 <- lme4::lmer(protein~factor(Time)+factor(Diet)+
        factor(Diet):factor(Time)+(-1+Time|Cow),
        data = mlk)
AIC(fit,fit2,fit3)
```

```
##      df      AIC
## fit  16 122.5821
## fit2 14 138.3831
## fit3 14 182.8353
```

Fitting the models

Now the fixed effects can be tested, beginning with the interaction term. But first the model has to be fitted again using the maximum likelihood method instead of REML. After that a drop1() command is used to see if the interaction is needed:

```
fitml <- lme4::lmer(protein~factor(Time)+factor(Diet)+
          factor(Diet):factor(Time)+(1+Time|Cow),
          REML=FALSE, data = mlk)
```

```
drop1(fitml)
```

```
## Single term deletions
##
## Model:
## protein ~ factor(Time) + factor(Diet) + factor(Diet):factor(Time) +
##     (1 + Time | Cow)
##                           npar    AIC
## <none>                          73.170
## factor(Time):factor(Diet)    6 66.792
```

Fitting the models

The model where the interaction is deleted has the lowest AIC so we can remove the interaction. Let's fit that model:

```
fitml2 <- lme4::lmer(protein~factor(Time)+factor(Diet)+
          (1+Time|Cow), REML=FALSE, data = mlk)
```

```
drop1(fitml2)
```

```
## Single term deletions
##
## Model:
## protein ~ factor(Time) + factor(Diet) + (1 + Time | Cow)
##              npar     AIC
## <none>              66.792
## factor(Time)    3 166.985
## factor(Diet)    2  67.739
```

There is not much difference between the first and third AIC's so we prefer the simpler model, that is the model with Diet deleted, so only Time in it. There is not enough evidence from the data that the diets are different w.r.t. the protein level. One can use confint() to obtain profile likelihood confidence intervals.

random effects Logistic regression

Logistic regression with random effects

Suppose that, in the cow example, at every time point it was measured whether or not the protein level was high.

One then measures binary data. This data should be modeled with logistic regression.

To deal with the dependencies in the data one can add a random intercept and random slopes to the linear part of the model, just as with an ordinary linear model.

That is, with a logistic regression model the dependencies are modeled on the logit scale.

Logistic regression with random effects

The logistic model becomes:

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + b_{0i} + (\beta_1 + b_{1i})\,Time$$

β_0 are the general log-odds at Time 0, and b_{0i} are the deviations from these general log-odds for every cow.

These b_{0i} are taken as draws from a normal distribution. In this way the data are modeled as being dependent, but on the logit scale. β_1 is the general coefficient for time and the b_{1i} are the deviations from this general time effect, now taken as random draws.

Logistic regression with random effects

The glmer function from the lme4 library is now used. To fit for example a logistic regression model with random intercepts, and with diet and time as grouping variables plus their interaction for the fixed part, one uses:

```
mlk$highprotein <- 1*(mlk$protein>3.5) #arbitrary choice
fith <- lme4::glmer(highprotein~factor(Time)+factor(Diet)+
        factor(Diet):factor(Time)+(1|Cow),family=binomial,
        data=mlk)
```

Logistic regression with random effects


```
summary(fith,correlation=FALSE)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## highprotein ~ factor(Time) + factor(Diet) + factor(Diet):factor(Time) +
##     (1 | Cow)
##    Data: mlk
##
##      AIC      BIC   logLik deviance df.resid
##    339.8    388.6   -156.9    313.8      302
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.7141 -0.4008 -0.0865  0.3727  2.8154
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  Cow    (Intercept) 4.261    2.064
## Number of obs: 315, groups:  Cow, 79
##
## Fixed effects:
##                                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)                            2.65965    0.83435   3.188  0.00143
## factor(Time)2                         -0.90172    0.89179  -1.011  0.31195
## factor(Time)3                         -2.22688    0.87507  -2.545  0.01093
## factor(Time)4                         -4.21687    0.99792  -4.226 2.38e-05
## factor(Diet)barley+lupins             -0.18307    1.11129  -0.165  0.86915
## factor(Diet)lupins                    -1.38568    1.05503  -1.313  0.18905
## factor(Time)2:factor(Diet)barley+lupins -1.47117  1.23299  -1.193  0.23280
## factor(Time)3:factor(Diet)barley+lupins -1.76819  1.26317  -1.400  0.16157
```

Logistic regression with random effects


```
drop1(fith)
```

```
## Single term deletions
##
## Model:
## highprotein ~ factor(Time) + factor(Diet) + factor(Diet):factor(Time) +
##     (1 | Cow)
##                            npar    AIC
## <none>                           339.83
## factor(Time):factor(Diet)    6 333.06
```

Logistic regression with random effects

So there is no evidence in the data that the interaction is needed and thus it can be left out.

The drop1() can then be used on the main effects model and if no terms can be left out.

confint() can be used to obtain profile likelihood confidence intervals.