Linear Mixed-Effects Models to Assess the Effect of Taking Phenacetin-Containing Analgesics On Kidney Function

Abu Saleh Mosa Faisal

MATH 582: Linear Statistical model with Application

Abstract

The goal of the analysis of Swiss Analgesic study was to see effect of taking phenacetin-containing analgesics on kidney function. We will see here how the slopes and intercepts differ among 3 groups (1=High NAPAP, 2=Low NAPAP, 3 = control). The assessment is done by using mixed effect modelling approach here to check the effect of different factors on response variable. This is an important aspect to check how the factors are creating impact on serum creatinine level over time. To check this, we tried to find here the significance of parameters and find out best model to fit the data set. With some diagnostic plots, model comparison, computing confidence intervals for parameters we tried to check the model validation.

Introduction

Serum-creatinine level which is an important index of kidney function are using here as an indicator or response variable of this modelling approach. It is important to identify the whether all of the explanatory variables(Age,Year,ID,Group) are random or some may be random and some fixed. Analysis of variance models for studies in which some factors are random and some fixed are called ANOVA models III. Here in data have some missing values which made treatment sample sizes unequal. So, to tackle this situation Maximum likelihood approach is adopted here.

We used one specific R package to perform the analysis part. This package is lme4.

Model selection process is performed by likelihood ratio test. As lme4 does not provide P-value for individual parameter.

We use Kenward-Roger appox and lmerTest R package to get P-values for model parameters.

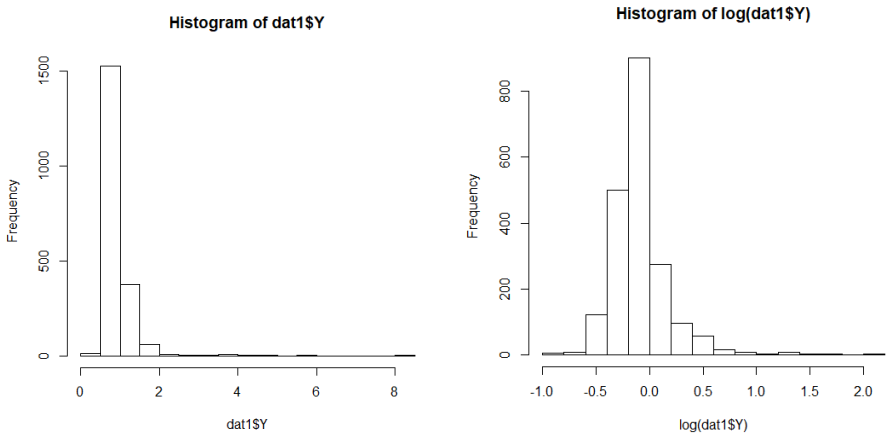
We used general profile zeta and some other related diagnostic plots.

Method

To assess the effect of taking phenacetin-containing analgesics on kidney function and other heath parameters we will use Mixed-effect models in R using the lme4 package.

Checking Data Pattern

To check pattern of data we create a histogram. It shows data are right skewed. To handle this problem, we need transform data. Here, we use log transformation of data. After transformation it shows approximately symmetrical pattern.



Linear Mixed-effect model

Most of the Ecological and biological data are often messy and complex. If we tried to fit complicated models with many parameters sample size could be an important aspect. On the above, collected data points might not be truly independent. This is why mixed effect models are using to deal with such messy data and to allow us to use all our data even when we have low sample sizes, structured data and many covariates to fit. On top of all that, mixed models allow us to save degrees of freedom compared to running standard linear models.

The restricted mixed ANOVA model for Mixed factor effects studies, where factor A is fixed and factor B is random, can now be stated as follows:

$$Y_{ijk} = \mu.. + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

where:

$\mu..$ is a constant

α_i are constants subject to the restriction

β_j are independent

$(\alpha\beta)_{ij}$ are $N(0, \frac{a-1}{a}\sigma^2_{\sigma\beta})$

ε_{ijk} are independent $N(0,\sigma^2)$

i=1…a j=1…b, k=1….. n

Likelihood Ratio Test

Model comparison with likelihood ratio tests is a better way of testing whether a parameter is significant. That is, if adding the parameter to your model significantly improves model fit, then that parameter should be included in the model. The likelihood ratio test essentially tells us how much more likely the data is under a more complex model than under the simpler model (these models need to be nested!):

D=-2*ln(likelihood for simple model /likelihood for complex model)=-2*ln(likelihood for simple model)+2*ln(likelihood for complex model)

The distribution of D is approximately chi-2 with df2-df1 degrees of freedom. We do this by using the anova() function.

Profiling and Profile Zeta Function

The deviance (or likelihood) profile, −2Lp(), for a focal model parameter P is the minimum value of the deviance conditioned on a particular value of P. For each parameter of interest, our goal is to evaluate the deviance profile for many points – optimizing over all of the non-focal parameters each time – over a wide enough range and with high enough resolution to evaluate the shape of the profile and to find the values of P, which represent the profile confidence intervals. The profile method systematically varies the parameters in a model, assessing the best possible fit that can be obtained with one parameter fixed at a specific value and comparing this fit to the globally optimal fit, which is the original model fit that allowed all the parameters to vary. The models are compared according to the change in the deviance, which is the likelihood ratio test statistic. We apply a signed square root transformation to this statistic and plot the resulting function, which we call the profile zeta function or ζ, versus the parameter value. The signed aspect of this transformation means that ζ is positive where the deviation from the parameter estimate is positive and negative otherwise, leading to a monotonically increasing function which is zero at the global optimum.

General profile zeta and related Plots

The profile zeta plot is simply a plot of the profile zeta function for each model parameters. The profile density plot displays an approximation of the probability density function of the sampling distribution for each parameter. The profile pairs plot gives an approximation of the two dimensional profiles of pairs of parameters, interpolated from the univariate profiles. The profile pairs plot shows two-dimensional 50%,80%,90%,95%,99% marginal confidence regions based on the likelihood ratios, as well as the profile traces, which indicate the conditional estimates of each parameter for fixed values of the other parameters.

Data Used

A group of 624 women were identified from workplaces near Basel, Switzerland, with high intake of phenacetine-containing analgesics. This constitute the study group. In addition, a control group of 626 women were identified form the same workplaces and with normal N-acetyl-P-aminophenyl (NAPAP) levels. The urine NAPAP level was used as a marker of recent phenacetine intake. The women were examined at baseline during 1967-1972 and 1975,1978. Data set contain variables ID, age of individuals, group, creatinine level in each year.

Result and Discussion

Fitted model

According to the AIC and Chi-square Ratio test for comparison of models

lmr <- lmer(log(Y) ~ age+ yr +group+ (1 | Id), data = dat1, REML=F)

is our best model for SWISS data.

Calculating ANOVA with P values and Confidence Intervals

ANOVA Table
Type III Analysis of Variance Table with Kenward-Roger's method

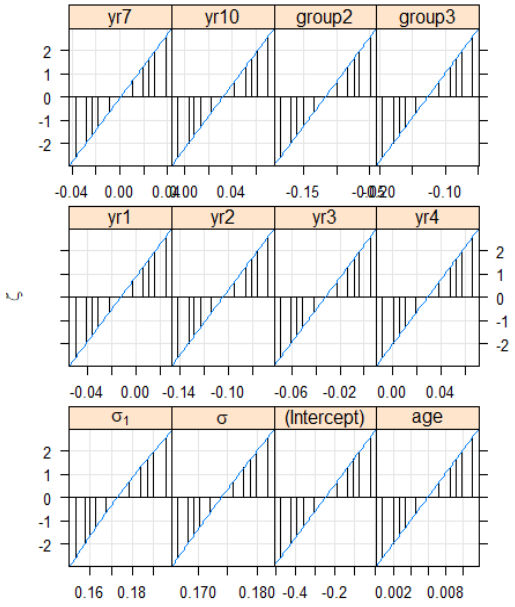
	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
age	0.2697	0.26970	1	296.42	8.8864	0.003111 **
yr	3.6566	0.60943	6	1694.97	20.0804	< 2.2e-16 ***
group	0.8515	0.42575	2	295.20	14.0283	1.513e-06 ***

Confidence Intervals

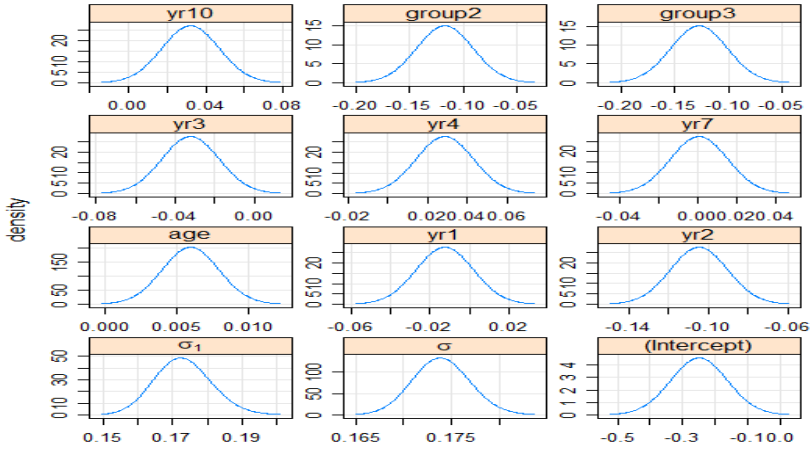
Confidence Interval for parameters by using likelihood profiling

	2.5 %	97.5%
sig01	0.1576	0.1899
sigma	0.1682	0.1799
Intercept	-0.4212	-0.0774
age	0.0020	0.0098
yr1	-0.0409	0.01563
yr2	-0.1331	-0.0764
yr3	-0.0606	-0.0039
yr4	0.0005	0.0570
yr7	-0.0279	0.0294
yr10	0.0032	0.0611
group2	-0.1696	-0.0653
group3	-0.1793	-0.0752

The Profile Plots

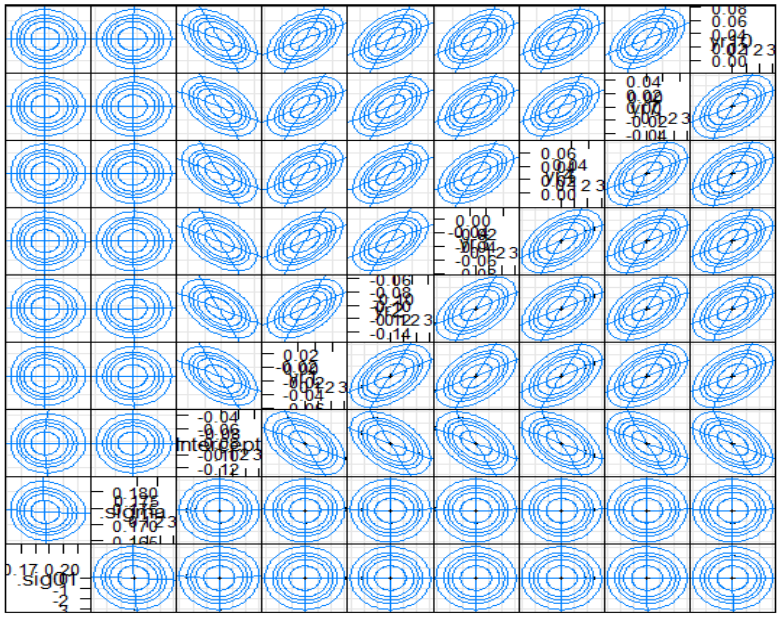


As it shows linearity approximately for all parameters so likelihood profile is quadratic and thus That likelihood profiling for parameters is reasonably accurate.

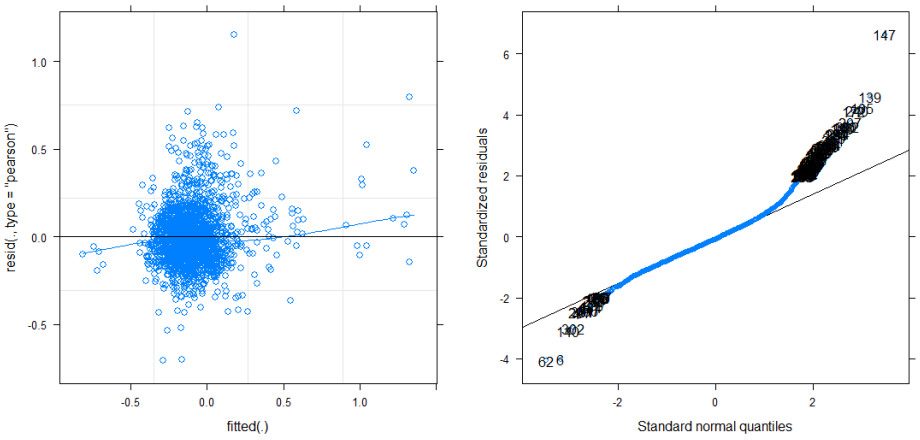


As The Profile Zeta plot is linear, the profile density plot is Gaussian. Here each parameter has approximately bell shaped distribution plot.

The profile Pairs plot

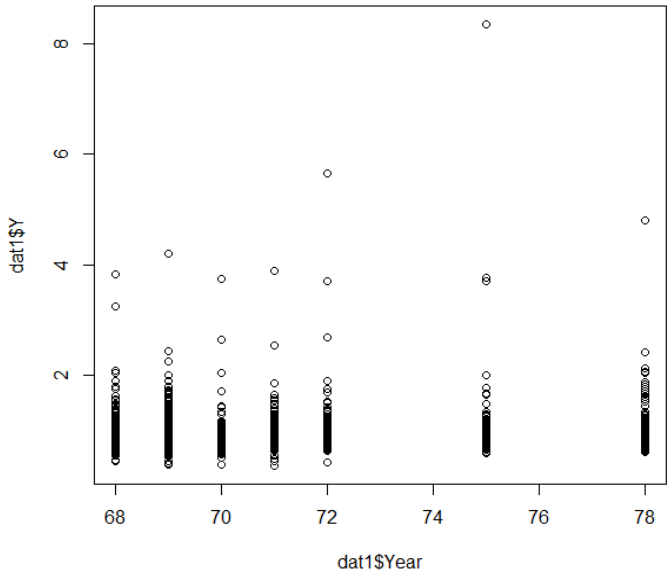


This plot provides visual assessment of the precision of parameter estimates. Any distortions from elliptical shape is due to nonlinearity of the traces.



The points in this residual plot are randomly dispersed around the horizontal axis.

The points seem to fall approximately in a straight line.



Data shows nonlinearity within Year. From this plot of Year and serum creatinine level we can see this nonlinear trend in data. So nonlinear modelling approach would be more appropriate than linear approach.

Woman with high phenacetin taking has shown greater change in serum-creatinine level compared with woman with low phenacetin intake as estimated parameter is -1.174e-01. So, high phenacetin taker has increased creatinine level over the year, which indicates poor kidney function.

References

1. Journal of Statistical Software, on Oct. 2015, with DOI 10.18637/jss.v067.i01.
2. Applied Linear Statistical Models
by Kutner, Michael H., Nachtsheim, Christopher J., Neter, John
3.web.stanford.edu/class/psych252/section/Mixed_models_tutorial.html