

ORIGIN := 0

Protocol(s) for Linear Mixed Models

Constructing an appropriate Linear Mixed Model using Maximum Likelihood (and REML) methodologies found in the {nlme} package in R is a multi-step process requiring a disciplined and purposeful method. By far the best explanation I've found of one approach is the 10-step backwards selection protocol found in Ch. 5 of Zuur et al. 2009 (Za), *Mixed Effects Models and Extensions in Ecology with R*. Additional suggestions along with more formal description of the software and underlying math are extensively covered in Pinheiro & Bates 2004 (PB), *Mixed-Effects Models in S and S-Plus*. The core of this protocol *in order* consists of:

- A - Fitting a Full Model (FM) containing as many fixed factors as practical followed by validation plots looking for potential patterns in residuals relative to each factor that indicate possible problems to be addressed in the fitting process. In addition, PB suggest an "inside-out" approach to initial validation that consists of contrasting a single linear model produced by `lm()` with multiple individual models for each group (or block) produced by `lmList()`. The ultimate Linear Mixed Model fit will be intermediate between these extremes in the final estimate of regression parameters.
- B - Fitting an appropriate Random Component to the Full Model (FMe) using Restricted Maximum Likelihood (REML) followed by comparison of alternatives using information criteria (AIC or BIC - smaller is better), formal tests (likelihood ratio test or F-test) and graphical validation.
- C - Fitting an appropriate Fixed Component, but also incorporating the "best fit" Random Component found in B, using Maximum Likelihood (ML). The strategy is to first remove interactions between fixed factors followed by the fixed factors themselves until all remaining independent variables are significant. This process of "backward selection" (i.e., successive removal of elements from the model) results in a "best fit" Reduced Model (RM) as judged by likelihood criteria or formal tests.
- D - Refitting the "best fit" Reduced Model (RM) with REML to obtain the best estimates of regression coefficients and variance components. This is followed by validation looking for potential improvements in the model, followed by retracing steps B-D as necessary.

The following example on Owl begging behavior is drawn from Za with additions to the analysis here based in part on their Ch. 4-7. The accompanying R script also contains the original numbered steps found in Ch. 4. It should be noted that their Step 10 consists of biological interpretation of the final model which is clearly important but generally beyond the scope of a formal protocol.

#LMM 060 PROTOCOL FOR LINEAR MIXED MODELS

```
library(nlme)
```

```
setwd("c:/DATA/Models")
```

```
Owls=read.table("owls.txt",header=T)
```

```
Owls
```

#LOG TRANSFORMATION OF RESPONSE VARIABLE TO MATCH ZUUR

```
Owls$LogNeg=log10(Owls$NegPerChick+1)
```

< LogNeg is the response variable

```
> Owls
```

	Nest	FoodTreatment	SexParent	ArrivalTime	SiblingNegotiation	BroodSize	NegPerChick	LogNeg
1	AutavauxTV	Deprived	Male	22.25	4	5	0.8000000	0.25527251
2	AutavauxTV	Satiated	Male	22.38	0	5	0.0000000	0.00000000
3	AutavauxTV	Deprived	Male	22.53	2	5	0.4000000	0.14612804
4	AutavauxTV	Deprived	Male	22.56	2	5	0.4000000	0.14612804
5	AutavauxTV	Deprived	Male	22.61	2	5	0.4000000	0.14612804
6	AutavauxTV	Deprived	Male	22.65	2	5	0.4000000	0.14612804
...								
595	Yvonnand	Deprived	Female	27.25	7	7	1.0000000	0.30103000
596	Yvonnand	Deprived	Male	28.45	5	7	0.7142857	0.23408320
597	Yvonnand	Deprived	Female	28.86	15	7	2.1428571	0.49732464
598	Yvonnand	Deprived	Male	29.21	10	7	1.4285714	0.38535088
599	Yvonnand	Satiated	Female	29.23	0	7	0.0000000	0.00000000

Step A - Fitting Full Model and Validation:

Step 1A - FITTING THE FULL LINEAR MODEL FIXED PART

```
FMform=formula(LogNeg~SexParent*FoodTreatment+SexParent*ArrivalTime)
```

```
FM=lm(FMform,data=Owls)
```

```
summary(FM)
```

< formula put into the variable
FMform for convenience
(saves typing in models below).

< FM = Full model fit with lm()

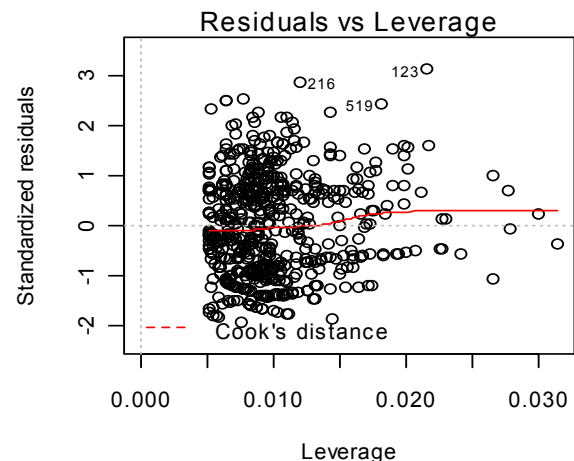
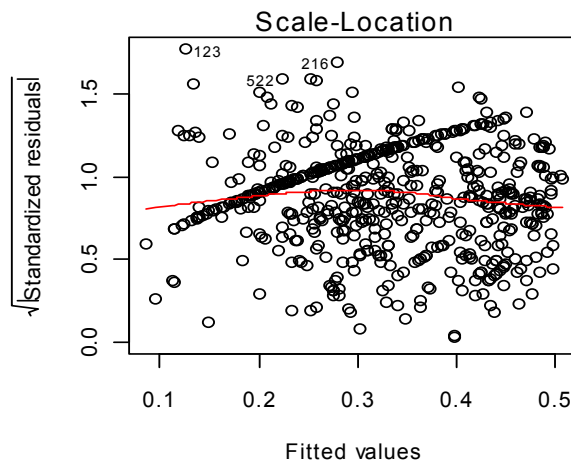
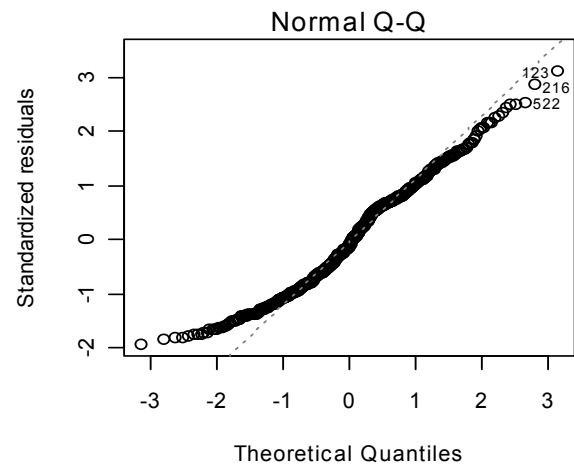
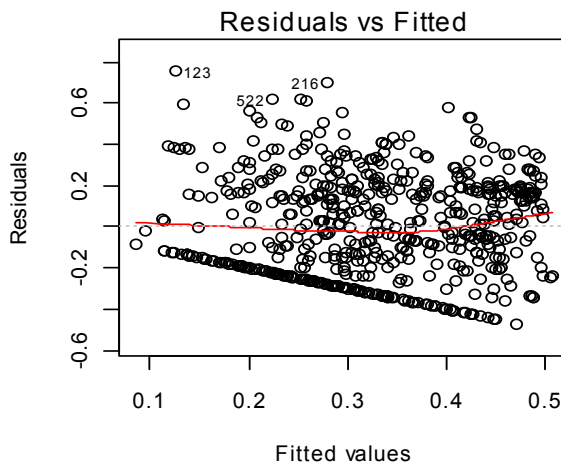
#STANDARD VALIDATION PLOTS

```
op=par(mfrow = c(2, 2))
```

```
plot(FM)
```

```
par(op)
```

Validation Plots:



#HOMEMADE PLOTS

```
op=par(mfrow = c(3,2))
```

```
plot(Owls$SexParent, resid(FM),pch=19,col='green',xlab="SexParent",ylab="Residuals")
```

```
plot(Owls$FoodTreatment, resid(FM),col='red',xlab="FoodTreatment",ylab="Residuals",pch=20)
```

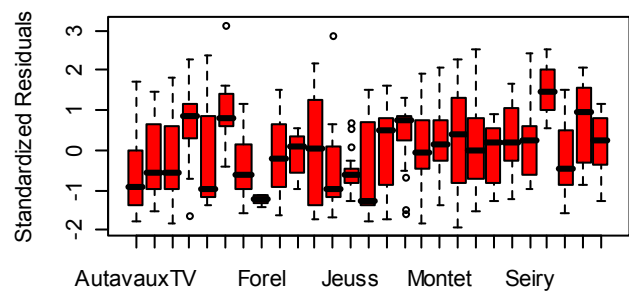
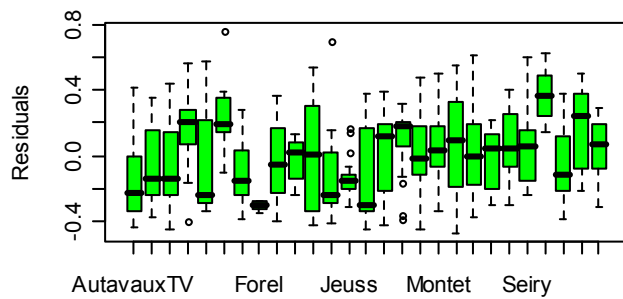
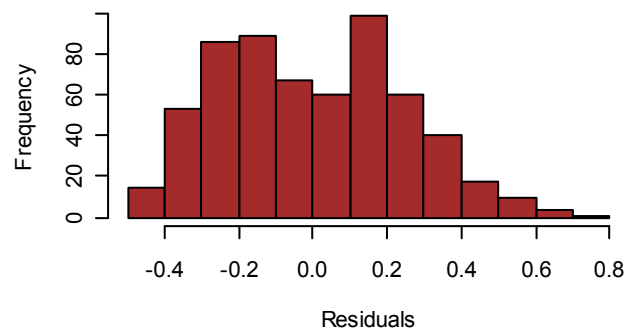
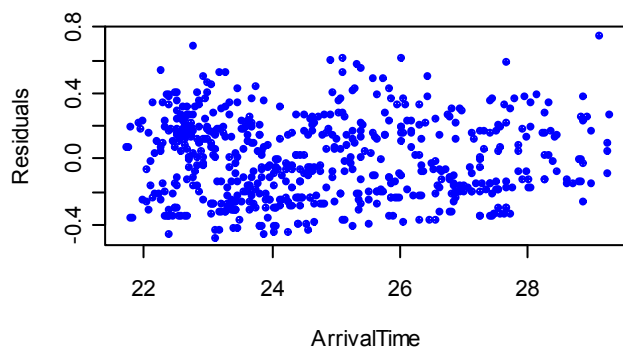
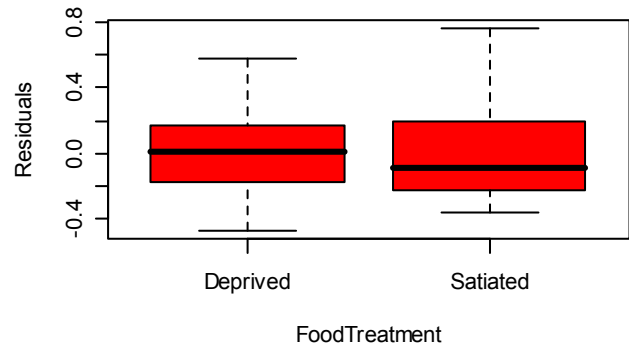
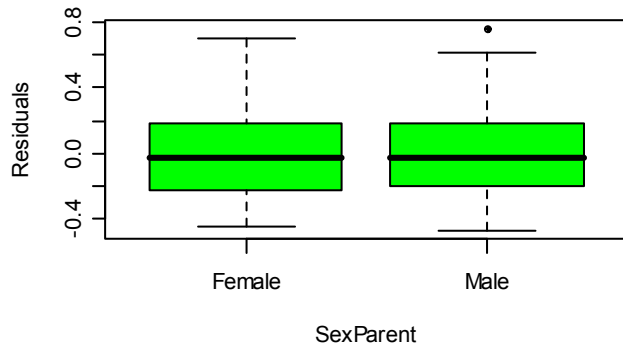
```
plot(Owls$ArrivalTime, resid(FM),col='blue',xlab="ArrivalTime",ylab="Residuals",pch=20)
```

```
hist(resid(FM),col="brown",main="",xlab="Residuals")
```

```
boxplot(resid(FM)~Nest,data=Owls,col='green',ylab='Residuals')
```

```
boxplot(rstandard(FM)~Nest,data=Owls,col='red',ylab='Standardized Residuals')
```

```
par(op)
```



```
#STEP 1B - SEPARATE LINEAR MODELS FOR EACH NEST
#NOTE: SexParent doesn't have two levels for each nest
# so must be dropped from formula here
FMLform=formula(LogNeg~FoodTreatment+ArrivalTime | Nest)
OG=groupedData(FMLform,data=Owls)
OG
FMG=lmList(OG)
plot(FMG)
plot(getGroups(FMG),resid(FMG),col="red",ylab='Residuals')
plot(intervals(FMG))
```

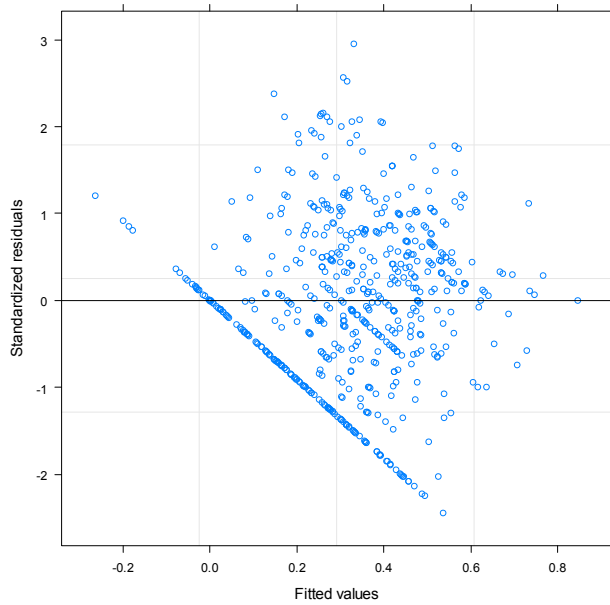
< groupedData object made for making linear models for each Nest and for graphing.

< FMG = Grouped Full Model by Nest

Separate linear Full Model regressions are made for each nest.

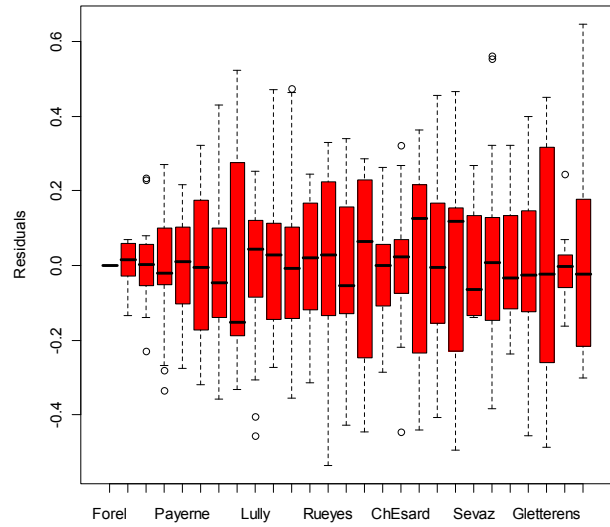
Validation plot of standardized residuals :

> plot(FMG)



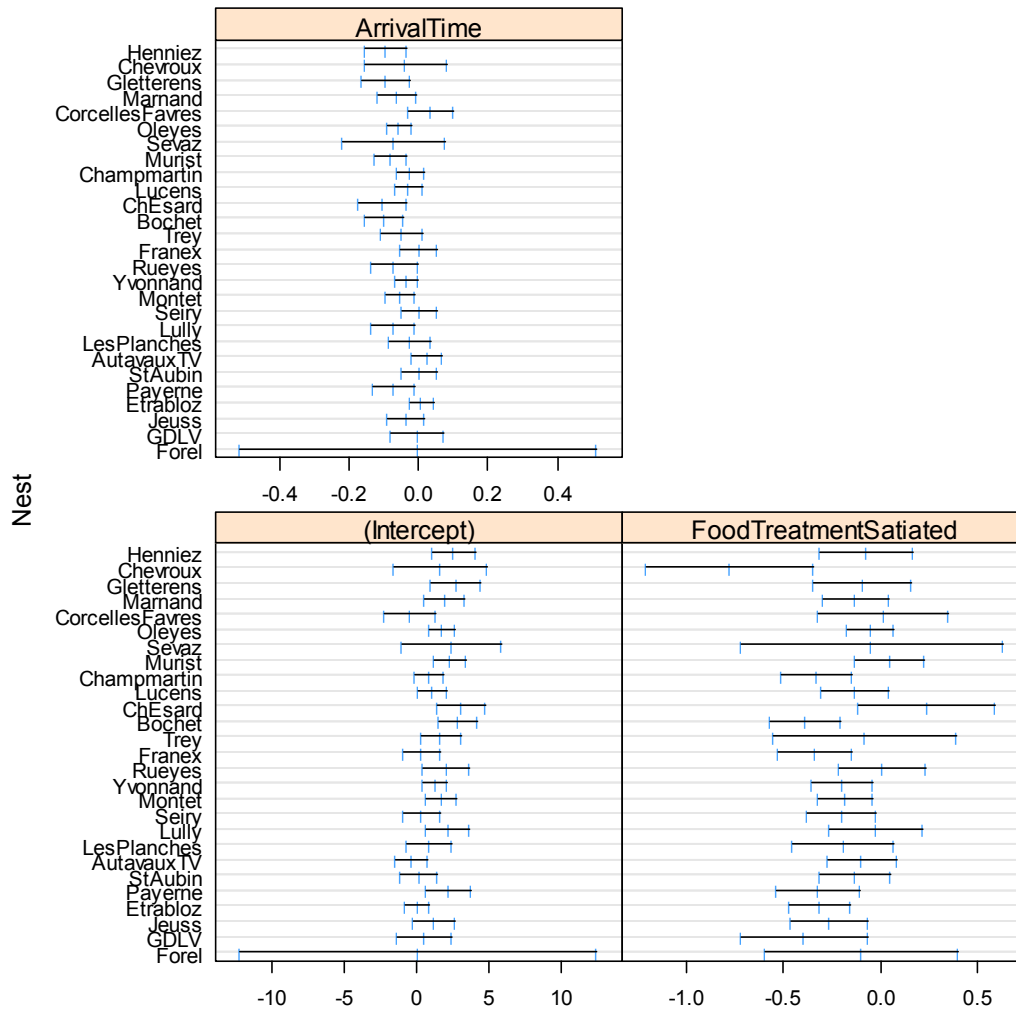
Plot of residuals for each nest:

> plot(getGroups(FMG), resid(FMG), col="red", ylab='Residuals')



> plot(intervals(FMG))

95 % Confidence Intervals of independent variables for each Nest:



#STEP 2 - FITTING THE FULL MODEL FIXED PART WITH gls()**FMgls=gls(FMform,data=Owls)****summary(FMgls)****< FMgls is the gls() fit that is fully equivalent to lm() fit.****> summary(FM)**

```
Call:
lm(formula = FMform, data = Owls)

Residuals:
    Min       1Q   Median       3Q      Max
-0.47057 -0.20195 -0.02452  0.18172  0.75759

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.118291   0.202485   5.523 5.00e-08
SexParentMale    0.076523   0.265572   0.288 0.773336
FoodTreatmentSatiated -0.158117  0.031370 -5.040 6.18e-07
ArrivalTime     -0.029930   0.008129 -3.682 0.000253
SexParentMale:FoodTreatmentSatiated 0.002990  0.040917  0.073 0.941767
SexParentMale:ArrivalTime -0.001422  0.010645 -0.134 0.893745

(Intercept)          ***
SexParentMale
FoodTreatmentSatiated ***
ArrivalTime          ***
SexParentMale:FoodTreatmentSatiated
SexParentMale:ArrivalTime
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2454 on 593 degrees of freedom
Multiple R-squared: 0.1402, Adjusted R-squared: 0.133
F-statistic: 19.35 on 5 and 593 DF, p-value: < 2.2e-16
```

> summary(FMgls)

```
Generalized least squares fit by REML
Model: FMform
Data: Owls
      AIC      BIC    logLik
64.37422 95.07058 -25.18711

Coefficients:
                Value Std. Error  t-value p-value
(Intercept)      1.1182915 0.20248539  5.522826 0.0000
SexParentMale    0.0765232 0.26557189  0.288145 0.7733
FoodTreatmentSatiated -0.1581174 0.03136955 -5.040473 0.0000
ArrivalTime     -0.0299302 0.00812918 -3.681822 0.0003
SexParentMale:FoodTreatmentSatiated 0.0029903 0.04091748  0.073080 0.9418
SexParentMale:ArrivalTime -0.0014224 0.01064493 -0.133625 0.8937

Correlation:
                (Intr) SxPrnM FdTrtS ArrvlT SPM:FT
SexParentMale   -0.762
FoodTreatmentSatiated -0.112 0.085
ArrivalTime    -0.994 0.758 0.035
SexParentMale:FoodTreatmentSatiated 0.086 -0.111 -0.767 -0.027
SexParentMale:ArrivalTime 0.759 -0.994 -0.027 -0.764 0.038

Standardized residuals:
      Min       Q1       Med       Q3      Max
-1.91794158 -0.82309063 -0.09992718  0.74065156  3.08779589

Residual standard error: 0.2453511
Degrees of freedom: 599 total; 593 residual
```

Note: gls() fits are generally preferred to lm() when intended for comparison with lme() fits.

Step B - Fitting Random Component to the Full Model:

#STEPS 3-5 - CHOOSE A VARIANCE STRUCTURE using REML

FMe=lme(FMform,random=~1 | Nest,data=Owls)

summary(FMe)

anova(FMgls,FMe)

> summary(FMe)

Linear mixed-effects model fit by REML

Data: Owls

	AIC	BIC	logLik
	37.71547	72.79702	-10.85773

Random effects:

Formula: ~1 | Nest

	(Intercept)	Residual
StdDev:	0.09311255	0.2321393

StdDev: 0.09311255 0.2321393

Fixed effects: list(FMform)

	Value	Std.Error	DF	t-value	p-value
(Intercept)	1.1236414	0.19522087	567	5.755744	0.0000
SexParentMale	0.1082138	0.25456854	567	0.425087	0.6709
FoodTreatmentSatiated	-0.1818952	0.03062840	567	-5.938776	0.0000
ArrivalTime	-0.0290079	0.00781832	567	-3.710251	0.0002
SexParentMale:FoodTreatmentSatiated	0.0140178	0.03971071	567	0.352998	0.7242
SexParentMale:ArrivalTime	-0.0038358	0.01019764	567	-0.376144	0.7070

Correlation:

	(Intr)	SxPrnM	FdTrtS	ArrvlT	SPM:FT
SexParentMale	-0.749				
FoodTreatmentSatiated	-0.102	0.080			
ArrivalTime	-0.989	0.746	0.022		
SexParentMale:FoodTreatmentSatiated	0.078	-0.113	-0.756	-0.019	
SexParentMale:ArrivalTime	0.747	-0.994	-0.019	-0.755	0.037

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-2.2182205	-0.7902406	-0.0778161	0.6991138	3.3273193

Number of Observations: 599

Number of Groups: 27

> anova(FMgls,FMe)

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
FMgls	1	7	64.37422	95.07058	-25.18711			
FMe	2	8	37.71547	72.79702	-10.85773	1 vs 2	28.65875	<.0001

^ Strong preference is shown for the Linear Mixed Full Model FMe over Linear Full Model FMgls. Za use this grouped (nested) design by Nest as their "best fit" Random Component. We'll consider other alternatives suggested by their Chapters 4 & 6 in a *separate set* of Steps B-D below. The following continues Za's analysis...

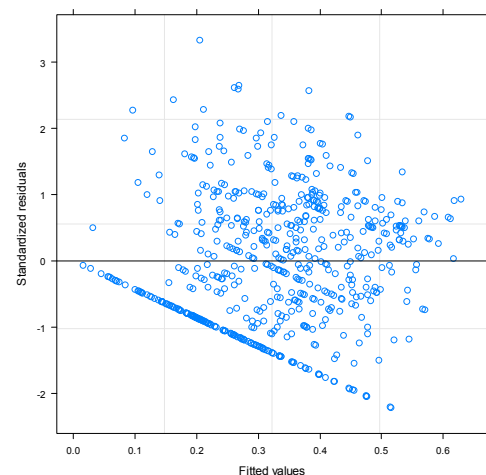
#STEP 6 - VALIDATE RANDOM PART OF FULL MODEL

#STANDARD VALIDATION PLOT

plot(FMe)

Validation Plots similar to those above but now with the Random Component of the Full Model included...

> plot(FMe)



#HOMEMADE PLOTS

```
op=par(mfrow = c(3,2))
```

```
plot(Owls$SexParent, resid(FMe),pch=19,col='green',xlab="SexParent",ylab="Residuals")
```

```
plot(Owls$FoodTreatment, resid(FMe),col='red',xlab="FoodTreatment",ylab="Residuals",pch=20)
```

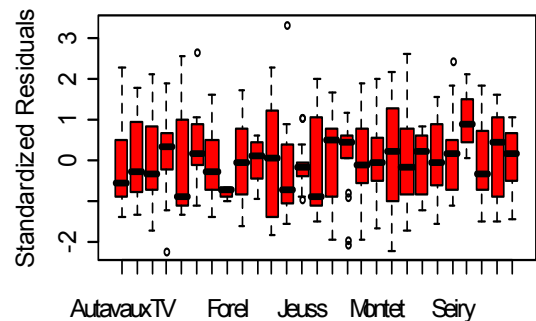
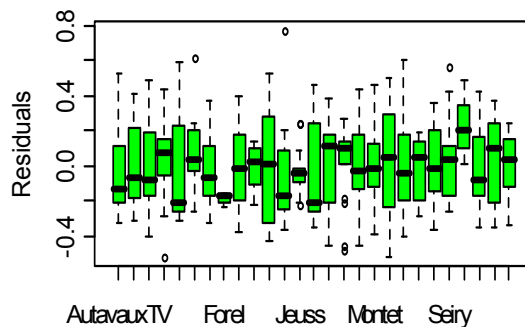
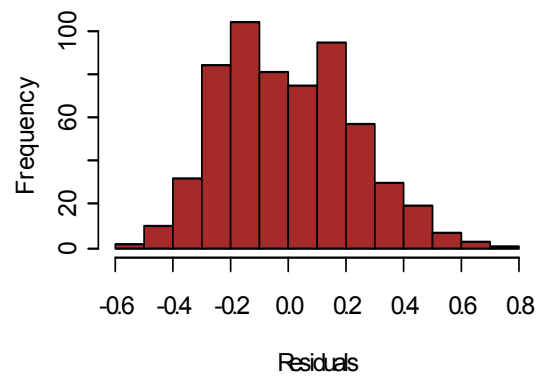
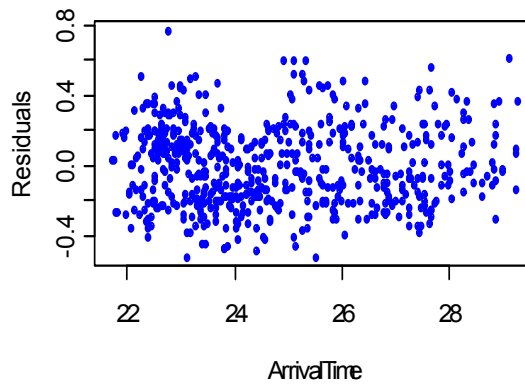
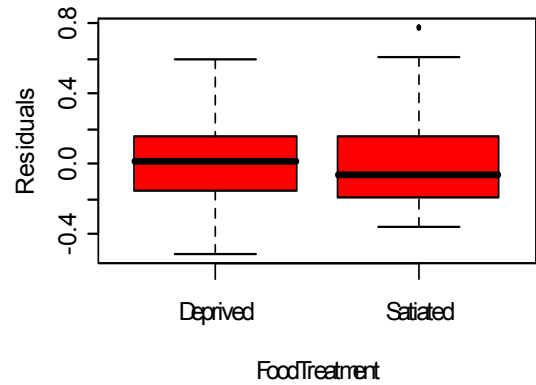
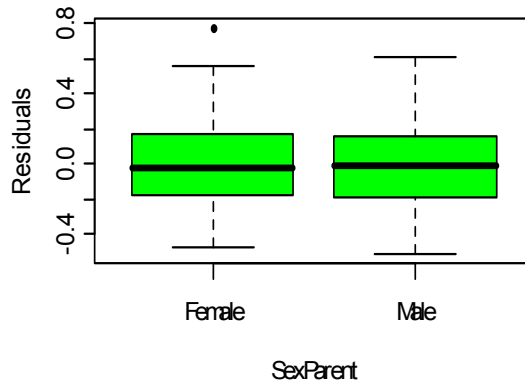
```
plot(Owls$ArrivalTime, resid(FMe),col='blue',xlab="ArrivalTime",ylab="Residuals",pch=20)
```

```
hist(resid(FMe),col="brown",main="",xlab="Residuals")
```

```
boxplot(resid(FMe)~Nest,data=Owls,col='green',ylab='Residuals')
```

```
boxplot(residuals(FMe,type='pearson')~Nest,data=Owls,col='red',ylab='Standardized Residuals')
```

```
par(op)
```



Step C - Finding the "best fit" Fixed Component using ML:

#STEPS 7-8 - FIND OPTIMAL REDUCED MODEL FIXED STRUCTURE using ML

FMe.ml=update(FMe,method="ML")

summary(FMe.ml)

anova(FMe.ml,type="marginal")

< FMe.ml = Mixed Full Model refitted with ML

```

> anova(FMe.ml,type="marginal")
              numDF denDF  F-value p-value
(Intercept)          1   567  33.08198 <.0001
SexParent            1   567   0.17693  0.6742
FoodTreatment        1   567  35.08614 <.0001
ArrivalTime          1   567  13.75610  0.0002
SexParent:FoodTreatment  1   567  0.12256  0.7264
SexParent:ArrivalTime  1   567   0.13697  0.7115

```

highest P value >

RMe1.ml=update(FMe.ml,~.-SexParent:FoodTreatment,data=Owls)

summary(RMe1.ml)

anova(RMe1.ml,type="marginal")

anova(FMe.ml,RMe1.ml)

< removing one Interaction

```

> anova(RMe1.ml,type="marginal")
              numDF denDF  F-value p-value
(Intercept)          1   568  33.01585 <.0001
SexParent            1   568   0.21456  0.6434
FoodTreatment        1   568  74.88418 <.0001
ArrivalTime          1   568  13.73157  0.0002
SexParent:ArrivalTime  1   568  0.14682  0.7017

```

highest P value >

> anova(FMe.ml,RMe1.ml)

Reduced Model RMe1.ml is preferred >

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
FMe.ml	1 8	-0.7484292	34.41366	8.374215			
RMe1.ml	2 7	-2.6246932	28.14214	8.312347	1 vs 2	0.123736	0.725

RMe2.ml=update(RMe1.ml,~.-SexParent:ArrivalTime,data=Owls)

summary(RMe2.ml)

anova(RMe2.ml,type="marginal")

anova(RMe1.ml,RMe2.ml)

< removing another Interaction

```

> anova(RMe2.ml,type="marginal")
              numDF denDF  F-value p-value
(Intercept)          1   569  82.58361 <.0001
SexParent            1   569   0.93872  0.333
FoodTreatment        1   569  74.88190 <.0001
ArrivalTime          1   569  37.14489 <.0001

```

highest P value >

> anova(RMe1.ml,RMe2.ml)

Reduced Model RMe2.ml is preferred >

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
RMe1.ml	1 7	-2.624693	28.14214	8.312347			
RMe2.ml	2 6	-4.476920	21.89465	8.238460	1 vs 2	0.1477732	0.7007

RMe3.ml=update(RMe2.ml,~.-SexParent,data=Owls)

summary(RMe3.ml)

anova(RMe3.ml,type="marginal")

anova(RMe2.ml,RMe3.ml)

< removing fixed factor

all P values now significant >

```

> anova(RMe3.ml,type="marginal")
              numDF denDF  F-value p-value
(Intercept)          1   570  83.90695 <.0001
FoodTreatment        1   570  76.60180 <.0001
ArrivalTime          1   570  36.72317 <.0001

```

> anova(RMe2.ml,RMe3.ml)

Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
RMe2.ml	1 6	-4.476920	21.89465	8.238460			
RMe3.ml	2 5	-5.545145	16.43116	7.772572	1 vs 2	0.9317755	0.3344

Modeling proceeds by backwards deletion of Fixed Components with high P-values looking first to remove Interactions followed by factors as each may perhaps be unnecessary in a "best fit" Reduced Model. Significant (low P-value) Interactions require us maintaining the factors involved within the model regardless of their significance. At the end of the selection process, all remaining elements in a Reduced Model should be significant. Here, RMe3.ml is chosen as "best fit". The quotes indicate that this is the best model so far found, nothing more.

Step D - Refitting "best fit" Reduced Model with REML with estimates and validation:

#STEP 9 REFIT OPTIMAL MODEL with REML AND VALIDATE

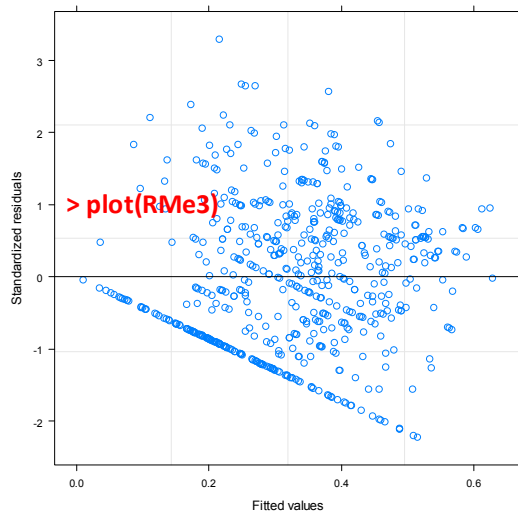
```
RMe3=update(RMe3.ml,method="REML")
```

```
summary(RMe3)
```

```
anova(RMe3)
```

```
#STANDARD VALIDATION PLOT
```

```
plot(RMe3)
```



> summary(RMe3)

Linear mixed-effects model fit by REML

Data: Owls

AIC	BIC	logLik
15.07383	37.02503	-2.536915

Random effects:

Formula: ~1 | Nest

(Intercept) Residual

StdDev: 0.09468769 0.2316398

Fixed effects: LogNeg ~ FoodTreatment + ArrivalTime

	Value	Std.Error	DF	t-value	p-value
(Intercept)	1.1821386	0.12897491	570	9.165648	0
FoodTreatmentSatiated	-0.1750754	0.01996606	570	-8.768650	0
ArrivalTime	-0.0310214	0.00511232	570	-6.067954	0

Correlation:

	(Intr)	FdTrtS
FoodTreatmentSatiated	-0.112	
ArrivalTime	-0.984	0.039

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.22283609	-0.78307304	-0.07461892	0.68690000	3.29183331

Number of Observations: 599

Number of Groups: 27

> anova(RMe3)

	numDF	denDF	F-value	p-value
(Intercept)	1	570	246.52327	<.0001
FoodTreatment	1	570	72.93930	<.0001
ArrivalTime	1	570	36.82006	<.0001

Final estimates of Variance for the random components (displayed as standard deviations) are:

Nest (Intercept): (0.09468769)²

Residual : (0.2316398)²

Final regression coefficients for independent variables are also shown in blue above. For ANOVA factors, results need to be interpreted according to R's default system of "treatment" contrasts.

#HOMEMADE PLOTS

```
op=par(mfrow = c(3,2))
```

```
plot(Owls$SexParent, resid(RMe3),pch=19,col='green',xlab="SexParent",ylab="Residuals")
```

```
plot(Owls$FoodTreatment, resid(RMe3),col='red',xlab="FoodTreatment",ylab="Residuals",pch=20)
```

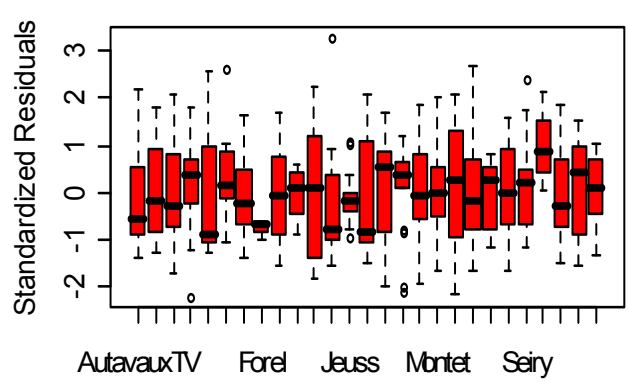
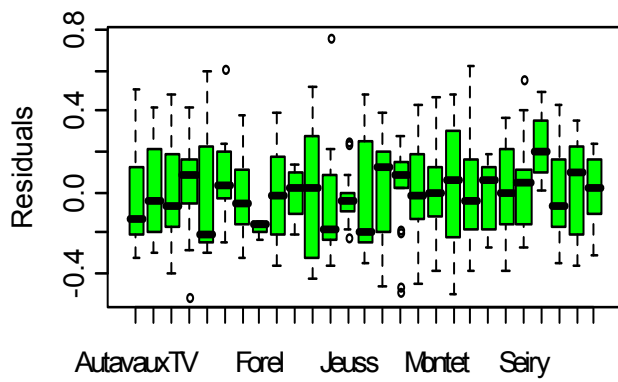
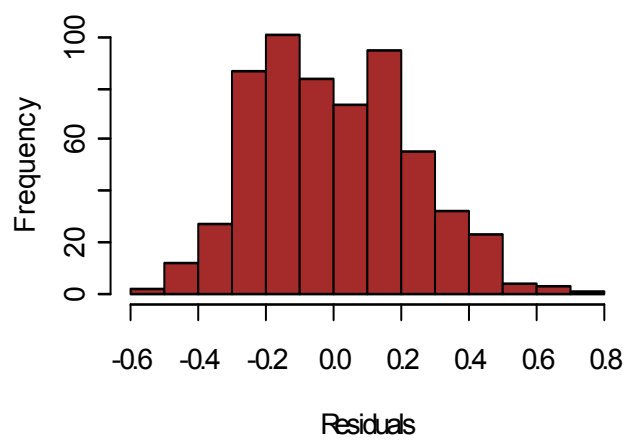
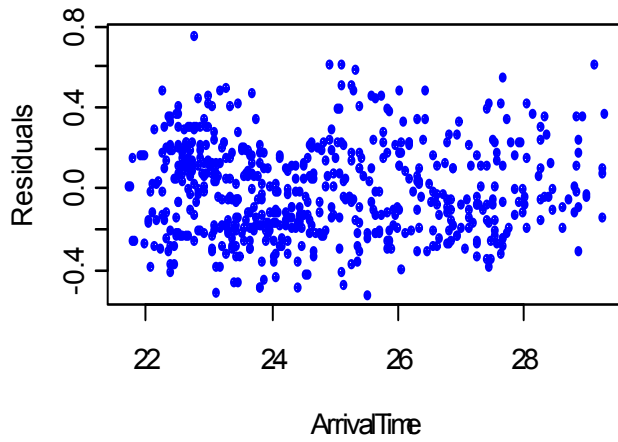
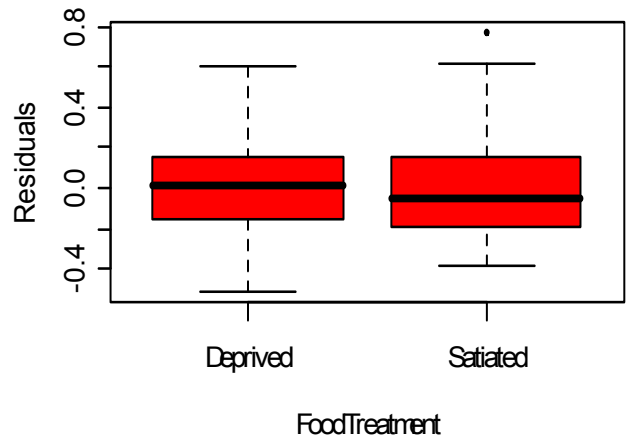
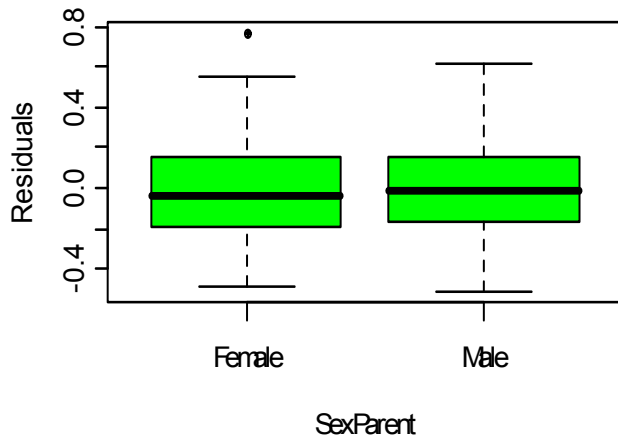
```
plot(Owls$ArrivalTime, resid(RMe3),col='blue',xlab="ArrivalTime",ylab="Residuals",pch=20)
```

```
hist(resid(RMe3),col="brown",main="",xlab="Residuals")
```

```
boxplot(resid(RMe3)~Nest,data=Owls,col='green',ylab='Residuals')
```

```
boxplot(residuals(RMe3,type='pearson')~Nest,data=Owls,col='red',ylab='Standardized Residuals')
```

```
par(op)
```



Validation plots of "best fit" Reduced Model looking for additional things to fix...

Further Analysis:

Za do not look into other kinds of Random Components to their Full Model in their Chapter 5. However, in other parts of the book they mention other possibilities including adding `varIdent()` variance structure to control for difference variances between Nests (Ch. 4) and `corAR1` Temporal Autocorrelation (Ch. 6). The following analysis looks at these two ways to change variance structure, but by no means exhausts the range of possibilities that may be employed. After a refined variance structure is found in a new Step B below, new Steps C-D complete the analysis.

Step B - Fitting Random Component to the Full Model:

```
#####
```

```
#ADDITIONAL STEPS 3-5 - CHOOSE A VARIANCE STRUCTURE using REML
```

```
FMe=lme(FMform,random=~1|Nest,data=Owls)
```

< original Linear Mixed Full Model from above

```
summary(FMe)
```

```
anova(FMe)
```

Adding varIdent correlation structure:

```
#ADDING CORRELATION STRUCTURES FOR Nest USING weights OPTION IN gls()
```

```
FMgls1=gls(FMform,data=Owls,weights=varIdent(form=~1|Nest))
```

```
summary(FMgls1)
```

< weights option provides means to control for differences in variances between Nests.

```
anova(FMgls1)
```

```
anova(FMe,FMgls1) #FMe strongly preferred over FMgls1
```

```
anova(FMgls,FMgls1) #FMgls1 with varIdent variance structure slightly preferred by test & AIC but not BIC
```

```
#ADDING varIdent CORRELATION STRUCTURE FOR Nest USING weights OPTION IN gls() & lme()
```

```
FMe2=lme(FMform,random=~1|Nest,data=Owls,weights=varIdent(form=~1|Nest),control=lmeControl(tolerance=1-e-3))
```

```
#method fails to converge here
```

```
> anova(FMe,FMgls1)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
FMe	1	8	37.71547	72.79702	-10.857734			
FMgls1	2	33	63.66514	208.37655	1.167432	1 vs 2	24.05033	0.5165

Linear Mixed Model FMe is greatly preferred (AIC) over Linear Model FMgls1 with varIdent variance structure.

```
> anova(FMgls,FMgls1)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
FMgls	1	7	64.37422	95.07058	-25.187109			
FMgls1	2	33	63.66514	208.37655	1.167432	1 vs 2	52.70908	0.0015

FMgls1 with varIdent variance structure is slightly preferred over Linear Model FMgls by test & AIC but not BIC because the latter more strongly penalizes model complexity (compare degrees of freedom). Although the more complex model FMgls1 (df = 33) is slightly significant via the Likelihood Ratio test, *the varIdent variance option is nevertheless rejected here in favor of much simpler FMe.*

Adding Compound Symmetric correlation structure:

```
#ADDING COMPOUND SYMMETRIC (NESTED) CORRELATION STRUCTURE USING weights OPTION IN gls()
```

```
FMgls2=gls(FMform,data=Owls,correlation=corCompSymm(form=~1|Nest))
```

```
summary(FMgls2) #IDENTICAL RESULTS TO FMe
```

```
anova(FMgls2)
```

^ The correlation option in `gls()` is a way of specifying many different kinds of correlation structures. The Compound Symmetry `CompSymm()` function specifies the same Random component correlation structures as `lme()`. Results on the next page may be compared with `summary(FMe)` and `anova(FMe)` in the first Step B above to see that the reports are mostly equivalent.

> summary(FMgls2)

Generalized least squares fit by REML

```

Model: FMform
Data: Owls
      AIC      BIC    logLik
37.71547 72.79702 -10.85773

```

Correlation Structure: Compound symmetry

```

Formula: ~1 | Nest
Parameter estimate(s):
  Rho
0.138589

```

Coefficients:

	Value	Std.Error	t-value	p-value
(Intercept)	1.1236414	0.19522087	5.755744	0.0000
SexParentMale	0.1082138	0.25456855	0.425087	0.6709
FoodTreatmentSatiated	-0.1818952	0.03062840	-5.938776	0.0000
ArrivalTime	-0.0290079	0.00781832	-3.710251	0.0002
SexParentMale:FoodTreatmentSatiated	0.0140178	0.03971072	0.352998	0.7242
SexParentMale:ArrivalTime	-0.0038358	0.01019764	-0.376143	0.7069

Correlation:

	(Intr)	SxPrnM	FdTrtS	ArrvlT	SPM:FT
SexParentMale	-0.749				
FoodTreatmentSatiated	-0.102	0.080			
ArrivalTime	-0.989	0.746	0.022		
SexParentMale:FoodTreatmentSatiated	0.078	-0.113	-0.756	-0.019	
SexParentMale:ArrivalTime	0.747	-0.994	-0.019	-0.755	0.037

Standardized residuals:

	Min	Q1	Med	Q3	Max
	-1.8992070	-0.8201118	-0.1110308	0.6947445	3.1053805

Residual standard error: 0.2501171

Degrees of freedom: 599 total; 593 residual

> anova(FMgls2)

Denom. DF: 593

	numDF	F-value	p-value
(Intercept)	1	252.64652	<.0001
SexParent	1	1.52859	0.2168
FoodTreatment	1	71.43971	<.0001
ArrivalTime	1	37.13833	<.0001
SexParent:FoodTreatment	1	0.13472	0.7137
SexParent:ArrivalTime	1	0.14148	0.7069

Adding AR-1 Autocorrelation structure:**#ADDING corAR1 AUTOCORRELATION FOR FoodTreatment WITHIN Nest TO glm(FM)****FMgls3=glm(FMform,data=Owls,correlation=corAR1(form=~1 | FoodTreatment))****summary(FMgls3)****anova(FMgls3)****anova(FMe,FMgls3) #FMgls3 STRONGLY PREFERRED****< FMgls3 Linear Full Model with AR1 Autocorrelation****> anova(FMgls3)**

Denom. DF: 593

	numDF	F-value	p-value
(Intercept)	1	490.9237	<.0001
SexParent	1	1.2449	0.2650
FoodTreatment	1	26.7421	<.0001
ArrivalTime	1	24.0311	<.0001
SexParent:FoodTreatment	1	0.3793	0.5382
SexParent:ArrivalTime	1	0.2553	0.6136

> anova(FMe,FMgls3)

Model	df	AIC	BIC	logLik	
FMe	1	8	37.71547	72.79702	-10.85773
FMgls3	2	8	-19.70481	15.37675	17.85241

Za point out that because times for two nights in each nest are recorded under conditions of different FoodTreatments, the latter variable is the correct one to use to correctly model autocorrelation or response for times *within* each night for each nest.

^ The Models do not interest, so no formal tests are reported. However, AIC & BIC preference is shown for *some* model incorporating Autocorrelation, even though FMgls3 doesn't include a Random Component for Nest to also be preferred... The following will put AR1 autocorrelation into a FMe model.

```
#ADDING corAR1 AUTOCORRELATION FOR Nest/FoodTreatment TO 1-level lme(FM)
FMe1=lme(FMform,random=~1 | Nest,data=Owls,correlation=corAR1(form=~1 | Nest/FoodTreatment))
summary(FMe1)
anova(FMe1)
anova(FMe1,FMe) #FMe1 STRONGLY SUPPORTED
anova(FMe1,FMgls3)
```

< *1-level Mixed Model with 2-level autocorrelation*

```
> anova(FMe1,FMe)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
FMe1	1	9	-34.27260	5.19415	26.13630			
FMe	2	8	37.71547	72.79702	-10.85773	1 vs 2	73.98807	<.0001

^ *1-level Mixed Model with 2-level autocorrelation FMe1 is strongly supported versus 1-level Mixed Model FMe.*

```
> anova(FMe1,FMgls3)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
FMe1	1	9	-34.27260	5.194147	26.13630			
FMgls3	2	8	-19.70481	15.376746	17.85241	1 vs 2	16.56779	<.0001

^ **Mixed Model with 2-level autocorrelation FMe1 supported versus Linear Model with 2nd level autocorrelation FMgls3**

```
#ADDING corAR1 AUTOCORRELATION FOR Nest/FoodTreatment TO 2-level lme(FM)
FMe2=lme(FMform,random=~1 | Nest/FoodTreatment,data=Owls,
correlation=corAR1(form=~1 | Nest/FoodTreatment))
```

```
summary(FMe2)
```

< *2-level Mixed Model with 2-level autocorrelation*

```
anova(FMe2)
```

```
anova(FMe2,FMe1)
```

```
> anova(FMe2,FMe1)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
FMe2	1	10	-32.2726	11.579342	26.1363			
FMe1	2	9	-34.2726	5.194147	26.1363	1 vs 2	1.861108e-07	0.9997

^ **Models are nearly equivalent, Mixed 1-level Model with 2-level autocorrelation FMe1 supported versus Mixed 2-level Model with 2-level autocorrelation by AIC & BIC**

```
#ADDING Nest/FoodTreatment 2-level lme(FM) ONLY
```

```
FMe3=lme(FMform,random=~1 | Nest/FoodTreatment,data=Owls)
```

```
summary(FMe3)
```

< *2-level Mixed Model with no Autocorrelation*

```
anova(FMe3)
```

```
anova(FMe3,FMe) #COMPARING ONE VS TWO LEVEL NESTING
```

```
anova(FMe3,FMe1) #COMPARING TWO LEVEL
```

```
anova(FMe3,FMe2)
```

```
> anova(FMe3,FMe)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
FMe3	1	9	31.77999	71.24674	-6.889997			
FMe	2	8	37.71547	72.79702	-10.857734	1 vs 2	7.935474	0.0048

^ *2-level Mixed Model FMe3 preferred versus 1-level Mixed model FMe, neither with autocorrelation*

```
> anova(FMe3,FMe1)
```

	Model	df	AIC	BIC	logLik
FMe3	1	9	31.77999	71.24674	-6.889997
FMe1	2	9	-34.27260	5.19415	26.136301

^ **1-level Mixed Model with 2-level autocorrelation versus 2-level Mixed Model. Models do not interest.**

```
> anova(FMe3,FMe2)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
FMe3	1	9	31.77999	71.24674	-6.889997			
FMe2	2	10	-32.27260	11.57934	26.136301	1 vs 2	66.0526	<.0001

^ **Mixed 2-level Model with 2-level autocorrelation FMe2 preferred versus 2-level Mixed Model FMe3**

AIC(FM,FMgls3,FMe1,FMe2,FMe3)

FMe1 appears preferred by AIC >

> AIC(FM,FMgls3,FMe1,FMe2,FMe3)

	df	AIC
FM	7	24.59000
FMgls3	8	-19.70481
FMe1	9	-34.27260
FMe2	10	-32.27260
FMe3	9	31.77999

Warning message:

In AIC.default(FM, FMgls3, FMe1, FMe2, FMe3) :
models are not all fitted to the same number of observations

All of the above looks into the several possible Random Component variance structures that might be incorporated into the Full Model to serve the modeling process perhaps better than Za's original choice above. In the end, it appears that FMe1 may be a "best fit" and is therefore chosen as the Random Component of the Mixed model below

#STEP 6 - VALIDATE RANDOM PART OF FULL MODEL

#STANDARD VALIDATION PLOT

plot(FMe1)

#HOMEMADE PLOTS

op=par(mfrow = c(3,2))

plot(Owls\$SexParent, resid(FMe1),pch=19,col='green',xlab="SexParent",ylab="Residuals")

plot(Owls\$FoodTreatment, resid(FMe1),col='red',xlab="FoodTreatment",ylab="Residuals",pch=20)

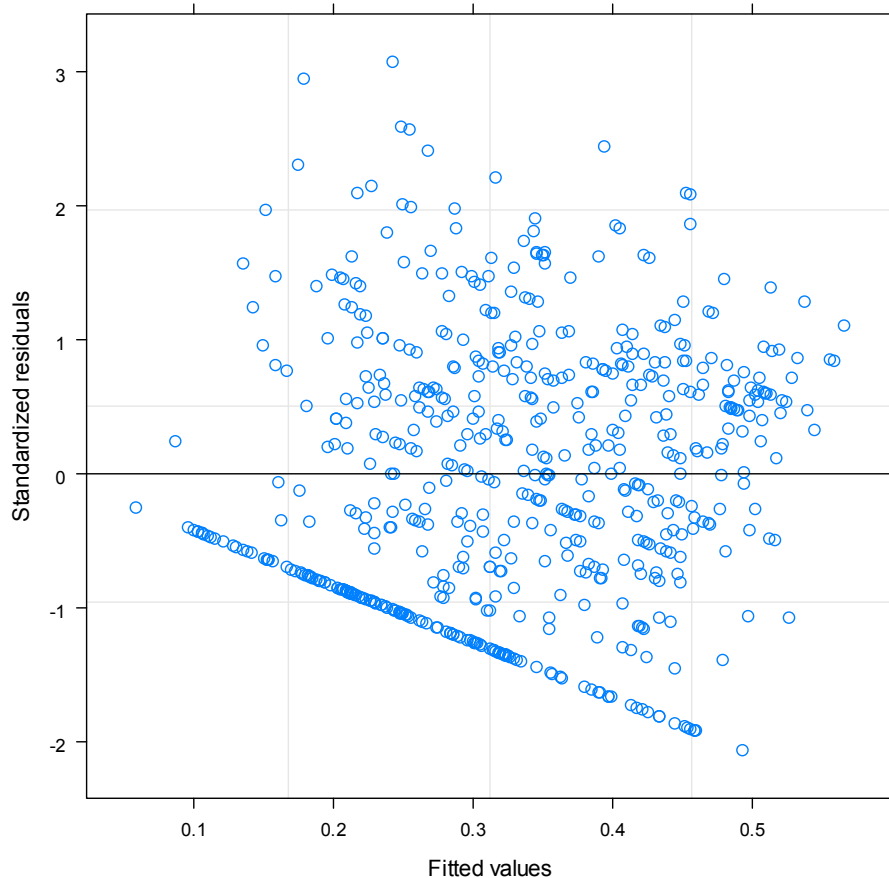
plot(Owls\$ArrivalTime, resid(FMe1),col='blue',xlab="ArrivalTime",ylab="Residuals",pch=20)

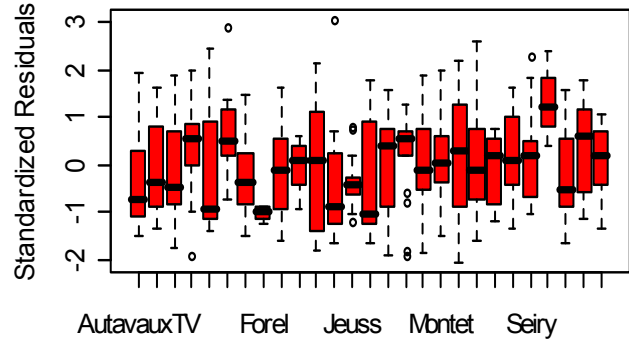
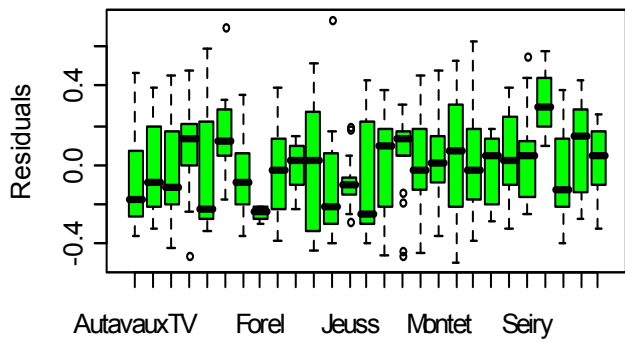
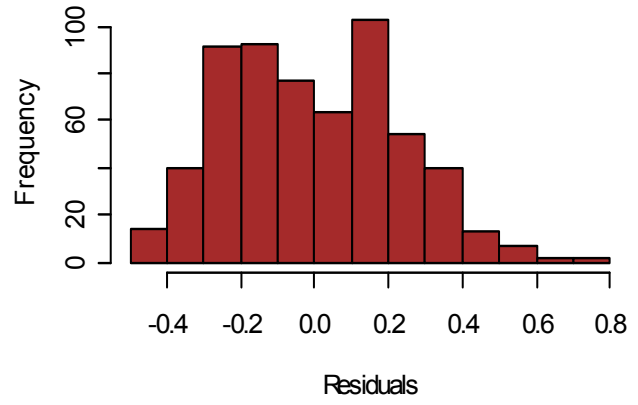
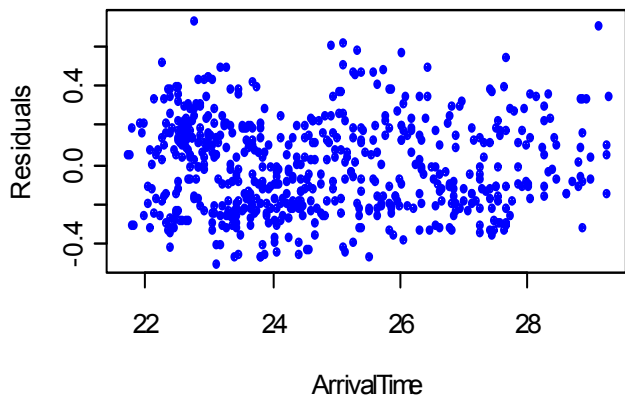
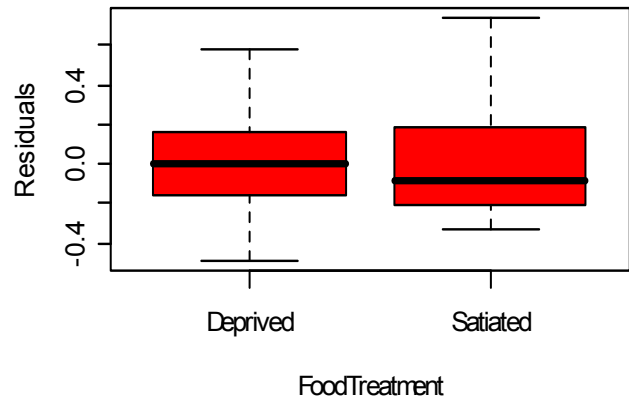
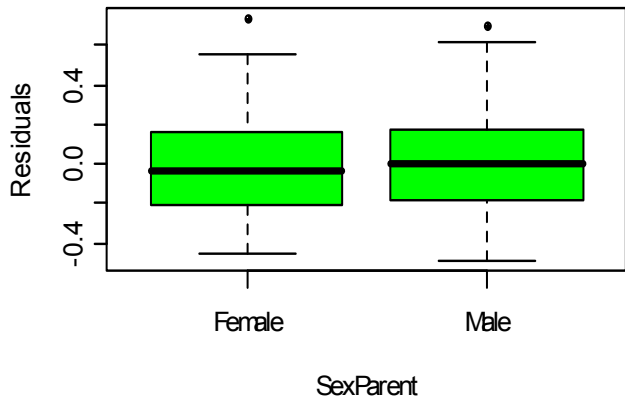
hist(resid(FMe1),col="brown",main="",xlab="Residuals")

boxplot(resid(FMe1)~Nest,data=Owls,col='green',ylab='Residuals')

boxplot(residuals(FMe1,type='pearson')~Nest,data=Owls,col='red',ylab='Standardized Residuals')

par(op)





Step C - Finding the "best fit" Fixed Component using ML:

#STEPS 7-8 - FIND OPTIMAL REDUCED MODEL FIXED STRUCTURE using ML

```
FMe1.ml=update(FMe1,method="ML")
summary(FMe1.ml)
anova(FMe1.ml,type="marginal")
RMe1.ml=update(FMe1.ml,~.-SexParent:FoodTreatment,data=Owls)
summary(RMe1.ml)
anova(RMe1.ml,type="marginal")
anova(FMe1.ml,RMe1.ml)
RMe2.ml=update(RMe1.ml,~.-SexParent:ArrivalTime,data=Owls)
summary(RMe2.ml)
anova(RMe2.ml,type="marginal")
anova(RMe1.ml,RMe2.ml)
RMe3.ml=update(RMe2.ml,~.-SexParent,data=Owls)
summary(RMe3.ml)
anova(RMe3.ml,type="marginal")
anova(RMe2.ml,RMe3.ml)
```

```
AIC(FMe1.ml,RMe1.ml,RMe2.ml,RMe3.ml)
```

< Exactly the same sequence of steps were required to find "best fit" Reduced Model....

```
> AIC(FMe1.ml,RMe1.ml,RMe2.ml,RMe3.ml)
```

	df	AIC
FMe1.ml	9	-72.22589
RMe1.ml	8	-74.02650
RMe2.ml	7	-75.10373
RMe3.ml	6	-76.70135

^ RMe3.ml is chosen...

Step D - Refitting "best fit" Reduced Model with REML with estimates and validation:

#STEP 9 REFIT OPTIMAL MODEL with REML AND VALIDATE

```
RMe3=update(RMe3.ml,method="REML")
summary(RMe3)
anova(RMe3)
#STANDARD VALIDATION PLOT
plot(RMe3)
```

< compare with Step D results above

```
> summary(RMe3)
```

```
Linear mixed-effects model fit by REML
Data: Owls
      AIC      BIC    logLik
-57.09017 -30.74873  34.54509
```

Random effects:

```
Formula: ~1 | Nest
      (Intercept) Residual
StdDev:  0.06474152 0.2389247
```

Correlation Structure: AR(1)

```
Formula: ~1 | Nest/FoodTreatment
Parameter estimate(s):
      Phi
0.3820835
```

Fixed effects: LogNeg ~ FoodTreatment + ArrivalTime

	Value	Std.Error	DF	t-value	p-value
(Intercept)	1.1469447	0.16576041	570	6.919292	0
FoodTreatmentSatiated	-0.1709077	0.02838117	570	-6.021870	0
ArrivalTime	-0.0297112	0.00660596	570	-4.497641	0

Correlation:

	(Intr)	FdTrtS
FoodTreatmentSatiated	-0.11	
ArrivalTime	-0.99	0.03

Standardized Within-Group Residuals:

	Min	Q1	Med	Q3	Max
	-2.02994661	-0.80443914	-0.07286733	0.70675002	3.00326221

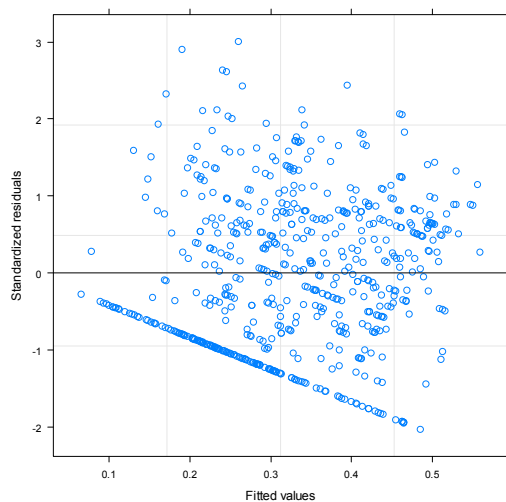
Number of Observations: 599

Number of Groups: 27

```
> anova(RMe3)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	570	298.93440	<.0001
FoodTreatment	1	570	34.70394	<.0001
ArrivalTime	1	570	20.22878	<.0001

```
> plot(RMe3)
```



#HOMEMADE PLOTS

```
op=par(mfrow = c(3,2))
```

```
plot(Owls$SexParent, resid(RMe3),pch=19,col='green',xlab="SexParent",ylab="Residuals")
```

```
plot(Owls$FoodTreatment, resid(RMe3),col='red',xlab="FoodTreatment",ylab="Residuals",pch=20)
```

```
plot(Owls$ArrivalTime, resid(RMe3),col='blue',xlab="ArrivalTime",ylab="Residuals",pch=20)
```

```
hist(resid(RMe3),col="brown",main="",xlab="Residuals")
```

```
boxplot(resid(RMe3)~Nest,data=Owls,col='green',ylab='Residuals')
```

```
boxplot(residuals(RMe3,type='pearson')~Nest,data=Owls,col='red',ylab='Standardized Residuals')
```

```
par(op)
```

