Statistical methods for linguistic research: Foundational Ideas

Shravan Vasishth

Universität Potsdam
vasishth@uni-potsdam.de
http://www.ling.uni-potsdam.de/∼vasishth

August 7, 2015

Summary

1. We know how to do simple t-tests.
2. We know how to fit simple linear models.

Now we are ready to look at an important type of linear model: linear mixed models.

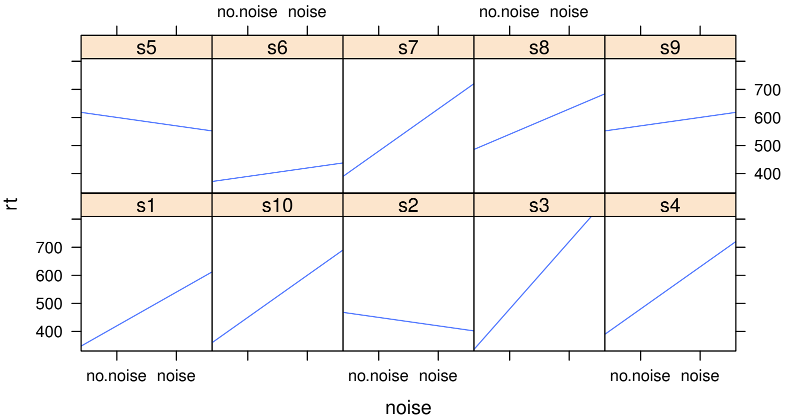
Linear models

Returning to our noise and deg data, one important point we've neglected is that different subjects have different effects of noise and deg. In the linear models we fit we were ignoring this.

```
noisedeg<-read.table("data/noisedeg.txt",
                     header=TRUE)

means.noise<-with(noisedeg,tapply(rt,list(subj,noise),
                                  mean))
means.deg<-with(noisedeg,tapply(rt,list(subj,
                                        deg),mean))
```

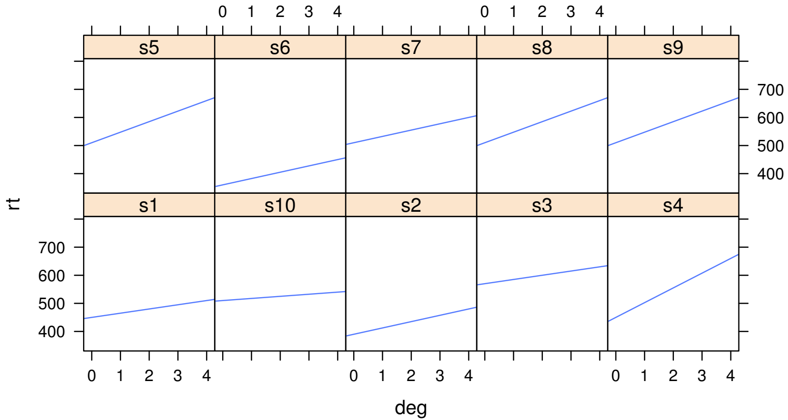
Linear models

We can visualize the different responses of subjects:



Linear models

We can do the same for degree:



Linear models

Given these differences between subjects, you could fit a separate linear model for each subject, collect together the intercepts and slopes for each subject, and then check if the intercepts and slopes are significantly different from zero.

Linear models

Fit a separate model for one subject (s1):

```
## fit a separate linear model for subject s1:
s1data<-subset(noisedeg,subj=="s1")
m<-lm(rt~noise,s1data)
round(summary(m)$coefficients,digits=2)

##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)      420      42.43     9.9     0.01
## noisenoise       120      60.00     2.0     0.18
```

Look at the means for s1 for noise and compare them to the coefficients above.

Linear models

Now we can do this for every one of our 10 subjects. I don't print this result out for space reasons.

```
## do the same for each subject using a for-loop
subjects<-paste("s",rep(1:10),sep="")
for(i in subjects){
  sdata<-subset(noisedeg,subj==i)
        lm(rt~noise,sdata)
}
```

Linear models

There is a function in the package `lme4` that does the above for you: `lmList`.

```
## do the same as the above for-loop for each subject:
library(lme4)

## Loading required package:  Matrix

lmlist.fm1<-lmList(rt~noise|subj,noisedeg)
```

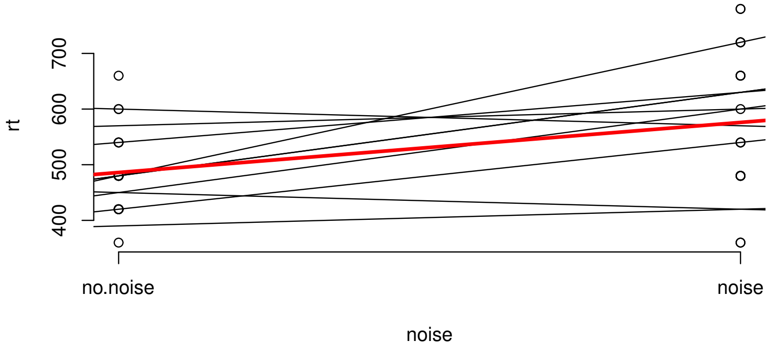
Linear models

```
print(lmlist.fm1$s1)

##
## Call:
## lm(formula = formula, data = data)
##
## Coefficients:
## (Intercept)   noisenoise
##         420          120
```

Linear models

One can plot the individual lines for each subject, as well as the linear model m0's line (this shows how each subject deviates in intercept and slope from the model m0's intercept and slopes).

Linear models



Linear models

To find out if there is an effect of noise, you can simply check whether the slopes of the individual subjects' fitted lines taken together are significantly different from zero.

Linear models

```
t.test(coef(lmlist.fm1)[2])

##
##  One Sample t-test
##
## data:  coef(lmlist.fm1)[2]
## t = 3.2225, df = 9, p-value = 0.01045
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   26.82139 153.17861
## sample estimates:
## mean of x
##        90
```

Linear models

The above is called **repeated measures regression**. We now transition to the next stage of multiple regression: the linear mixed model.

Linear mixed models

The **linear mixed model** does something related to the above by-subject fits, but with some crucial twists, as we see below.
In the model shown in the next slide, the statement
(1|subj)
means that the variance associated with subject intercepts should be estimated, and from that variance the intercepts for each subject should be predicted.

Linear mixed models

```
m0.lmer<-lmer(rt~noise+(1|subj),noisedeg)
```

Abbreviated output:

```
Random effects:
 Groups   Name        Variance Std.Dev.
 subj     (Intercept) 2491     49.91
 Residual             8876     94.21
Number of obs: 40, groups:  subj, 10

Fixed effects:
            Estimate Std. Error t value
(Intercept)   486.00      26.32  18.463
noisenoise     90.00      29.79   3.021
```

Linear mixed models

One thing to notice is that the coefficients of the fixed effects of the above model are identical to those in the linear model m0 above.
The **predicted** (not estimated!) varying intercepts for each subject can be viewed by typing:

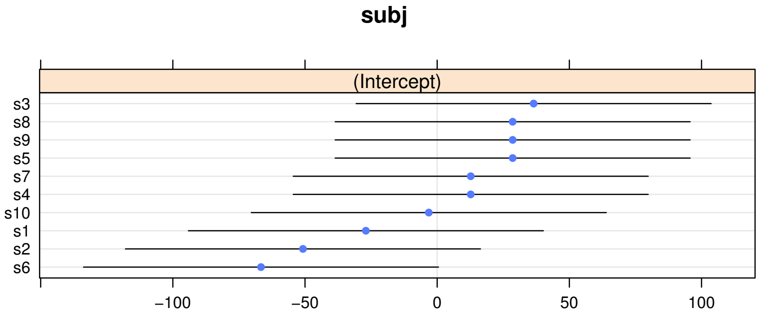
Linear mixed models

```
ranef(m0.lmer)

## $subj
##      (Intercept)
## s1    -26.972985
## s10    -3.173292
## s2    -50.772677
## s3     36.492862
## s4     12.693169
## s5     28.559631
## s6    -66.639139
## s7     12.693169
## s8     28.559631
## s9     28.559631
```

Linear mixed models

Or you can display them graphically.

```
print(dotplot(ranef(m0.lmer,condVar=TRUE)))

## $subj
```



Linear mixed models

The model m0.lmer above prints out the following type of linear model:

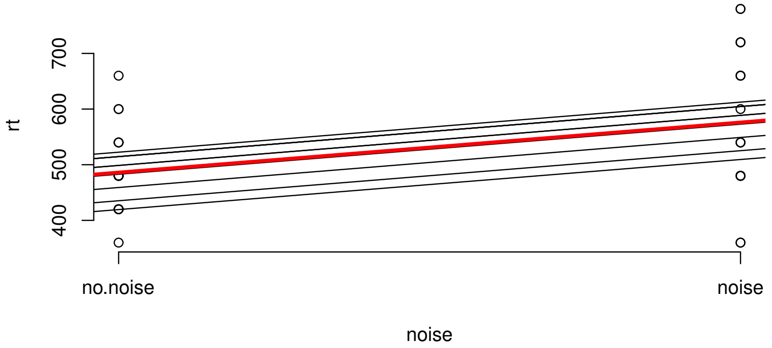
$$Y_{ijk} = \beta_j + b_i + \epsilon_{ijk} \tag{1}$$

$i = 1, \dots, 10$ is subject id, $j = 1, 2$ is the factor level (no noise or noise), k is the number of replicates (here 1).
$b_i \sim N(0, \sigma_b^2), \epsilon_{ijk} \sim N(0, \sigma^2)$.
It's just like our linear model except that there are different *predicted* (cf. the lmlist function above, where they are *estimated* for each subject) intercepts b_i for each subject.

Linear mixed models

Note that these b_i are assumed by lmer to come from a normal distribution centered around 0; see Gelman and Hill 2007 for more.
The ordinary linear model m0 has one intercept β_0 for all subjects, whereas the linear mixed model with varying intercepts m0.lmer has a different intercept $(\beta_0 + b_i)$ for each subject.
We can visualize these different intercepts for each subject as shown below.

Linear mixed models



Linear mixed models

Note that, unlike the figure associated with the lmlist.fm1 model above, which also involves fitting separate models for each subject, the model m0.lmer assumes different intercepts for each subject **but the same slope**.
We can have lmer fit different intercepts AND slopes for each subject.

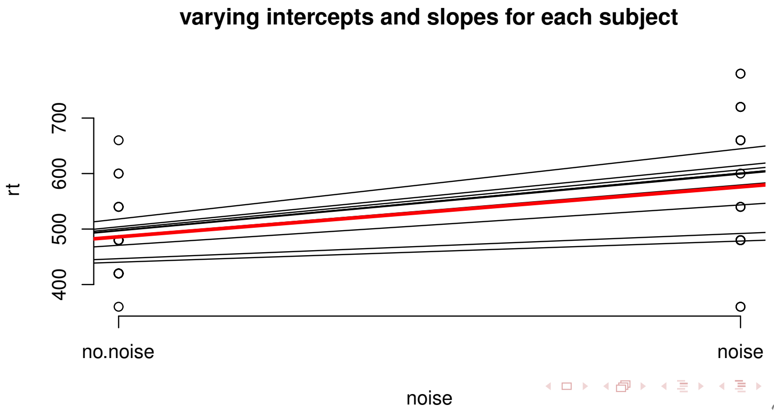
Linear mixed models

```
m1.lmer<-lmer(rt~noise+(1+noise|subj),noisedeg)
```

```
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 subj     (Intercept) 1093     33.05
          noisenoise  1408     37.52    1.00
 Residual             8359     91.43
Number of obs: 40, groups:  subj, 10

Fixed effects:
            Estimate Std. Error t value
(Intercept)   486.00      22.96   21.17
noisenoise     90.00      31.25    2.88
```

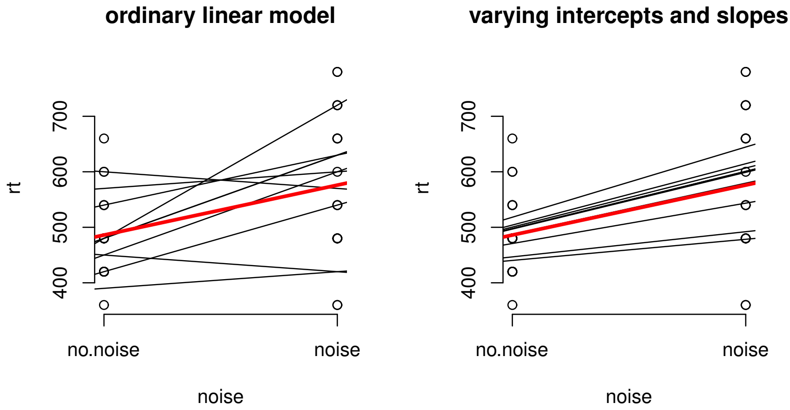
Linear mixed models

These fits for each subject are visualized below (the red line shows the model with a single intercept and slope, i.e., our old model m0):



Linear mixed models

Compare this model with the lmlist.fm1 model we fitted earlier:



Linear mixed models

1. The above graphic shows some crucial difference between the lmlist (repeated measures) model and the lmer model.
2. Note that the fitted line for each subject in the lmer model is much closer to the m0 model's fitted (red) line.
3. This is because lmlist uses each subject's data separately (resulting in possibly wildly different models, depending on the variability between subjects), whereas lmer "borrows strength from the mean" and pushes (or "shrinks") the estimated intercepts and slopes of each subject closer to the mean intercepts and slopes (the model m0's intercepts and slopes).
4. Because it shrinks the coefficients towards the means, this is called shrinkage. This is particularly useful when several data points are missing in a particular condition for a particular subject: in an ordinary linear model, estimating coefficients using lmList would lead to very poor estimates for that subject; by contrast, lmer assumes that the estimates for such a subject are not reliable and therefore shrinks that subject's estimate to the mean values.

Linear mixed models

To see an example of shrinkage, consider the case where we remove three of the data points from subject s8, resulting in exaggeratedly high means for that subject.
First, we read in a data frame which is just the same as noisedeg, except that subject 8 (s8) has only three data points, not six (I took out three of s8's low measures). This skews the subject's estimates for intercept and slope in the lmlist model fit.
I am now using the full noisedeg dataset (this is a 2x3 design, with three levels of degree, 0, 4, 8).

Linear mixed models

```
noisedeg2<-read.table("data/noisedegfull.txt",header=T)
```

Next, let's confirm that the new data frame has extreme means for s8.

Linear mixed models

```
with(noisedeg,tapply(rt,list(subj,noise),mean,na.rm=TRUE))

##      no.noise noise
## s1        420   540
## s10       450   600
## s2        450   420
## s3        480   720
## s4        480   630
## s5        600   570
## s6        390   420
## s7        480   630
## s8        540   630
## s9        570   600
```

Linear mixed models

```
with(noisedeg2,tapply(rt,list(subj,noise),mean,na.rm=TRUE))

##      no.noise noise
## s1        440   620
## s10       480   660
## s2        460   480
## s3        500   740
## s4        500   720
## s5        580   620
## s6        380   460
## s7        520   700
## s8        660   810
## s9        560   660
```

Linear mixed models

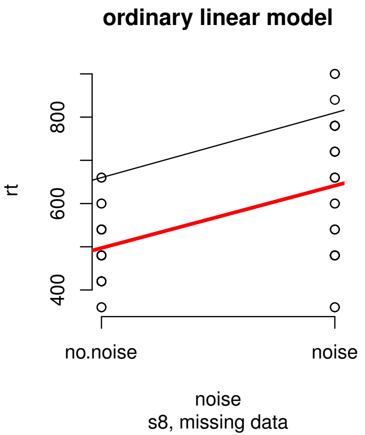
We now fit the lmlist model and the linear mixed model.

```
lmlist.fm2<-lmList(rt~noise|subj,noisedeg2)

m2.lmer<-lmer(rt~noise+(1+noise|subj),noisedeg2)
```

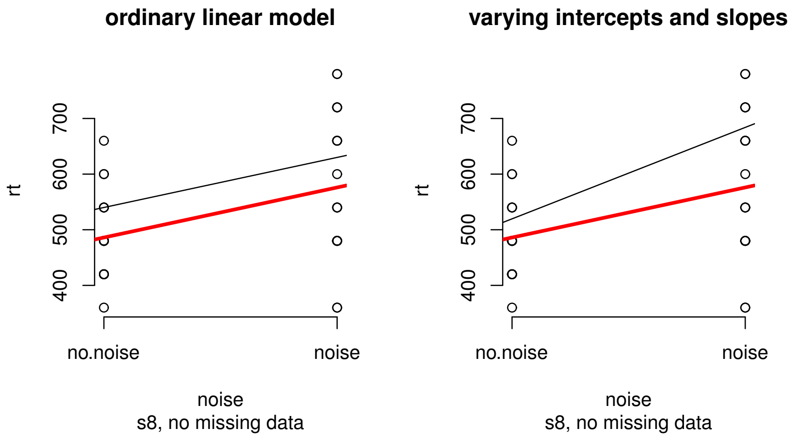
Linear mixed models

Now if we plot the model for s8, we find that the lmlist model indeed estimates **pretty extreme intercepts for s8**.
But the linear mixed model predicts an intercept that's **much closer to the mean** (the red line). Let's just plot s8's fitted line in both models relative to the linear model fitted line.

Linear mixed models



Linear mixed models



Linear mixed models

One crucial difference between the lmlist model and the lmer model is that **lmList estimates the parameters for each subject separately**.
By contrast, **lmer estimates the variance associated with subjects' intercepts** (and slopes, if you specify in the model that one should do that) and then *predicts* each subjects intercepts and slopes based on that variance.
I will formalize this on the last day if there is time.

Gibson and Wu 2013

Processing Chinese relative clauses in context

Research question: Are subject relatives harder than object relatives in Chinese?

```
data<-read.table("data/gibsonwu2012data.txt",header=TRUE)
headnoun<-subset(data,region=="headnoun")
head(headnoun[,c(1,2,3,7)])

##     subj item     type   rt
## 94     1   13  obj-ext 1561
## 221    1    6 subj-ext  959
## 341    1    5  obj-ext  582
## 461    1    9  obj-ext  294
## 621    1   14 subj-ext  438
## 753    1    4 subj-ext  286
```

Gibson and Wu 2013

Processing Chinese relative clauses in context

Published result was approximately the following (statistically significant):

```
contrasts(data$type)<-contr.sum(2)
contrasts(data$type)

##            [,1]
## obj-ext       1
## subj-ext     -1
```

Gibson and Wu 2013

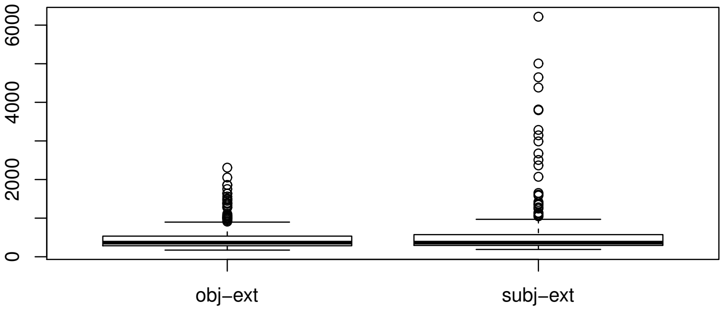
Processing Chinese relative clauses in context

```
m0<-lmer(rt~type+(1+type|subj)+(1|item),headnoun)
summary(m0)$coefficients

##                 Estimate Std. Error  t value
## (Intercept)   487.6670   52.37682 9.310741
## typesubj-ext 120.3888   56.86962 2.116926
```

Gibson and Wu 2013

Processing Chinese relative clauses in context

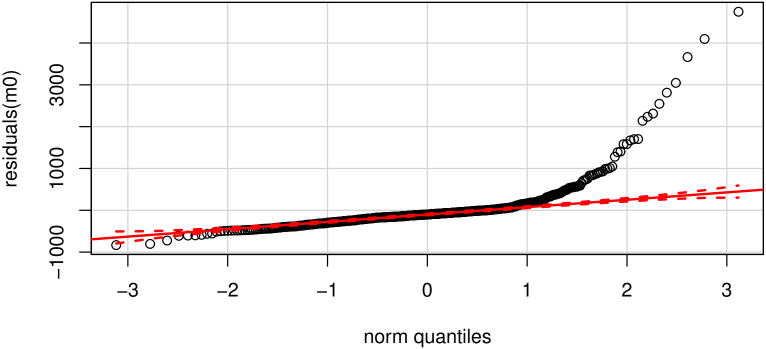
Result is driven by 1% of extreme values:



Gibson and Wu 2013

Processing Chinese relative clauses in context

Residuals are severely non-normal:



Gibson and Wu 2013

Processing Chinese relative clauses in context

With a log transform of reading time (to reduce influence of extreme values; see Box and Cox 1964), we see no evidence for an effect (inconclusive result):

```
m1<-lmer(log(rt)~type+(1+type|subj)+(1|item),headnoun)
summary(m1)$coefficients

##                Estimate Std. Error  t value
## (Intercept)  6.02554797 0.06404223 94.08710
## typesubj-ext 0.07249477 0.04830504  1.50077
```

Summary of LMMs

Varying intercepts model

The model for a categorical predictor is:

$$Y_{ijk} = \beta_j + b_i + \epsilon_{ijk} \tag{2}$$

$i = 1, \dots, 10$ is subject id, $j = 1, 2$ is the factor level, k is the number of replicates (here 1). $b_i \sim N(0, \sigma_b^2), \epsilon_{ijk} \sim N(0, \sigma^2)$.

Summary of LMMs

Varying intercepts model

For a continuous predictor:

$$Y_{ijk} = \beta_0 + \beta_1 t_{ijk} + b_{ij} + \epsilon_{ijk} \tag{3}$$

Summary of LMMs

Varying intercepts and varying slopes model

The model for a categorical predictor is:

$$Y_{ij} = \beta_1 + b_{1i} + (\beta_2 + b_{2i})x_{ij} + \epsilon_{ij} \quad i = 1, ..., M, j = 1, ..., n_i \quad (4)$$

with $b_{1i} \sim N(0, \sigma_1^2)$, $b_{2i} \sim N(0, \sigma_2^2)$, and $\epsilon_{ij} \sim N(0, \sigma^2)$.

Summary of LMMs

Varying intercepts and varying slopes model

Another way to write such models is:

$$Y_{ijk} = \beta_j + b_{ij} + \epsilon_{ijk} \tag{5}$$

$b_{ij} \sim N(0, \sigma_b)$. The variance σ_b must be a 2×2 matrix:

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{6}$$

See my linear modeling notes for much more.

Summary of LMMs

Further reading

1. The authoritative and classic textbook on LMMs is: Gelman and Hill 2007, Data analysis using regression and multilevel/hierarchical models.
2. Please also read: Barr et al 2013, Random effects structure in mixed-effects models: Keep it maximal, Journal of Memory and Language.
3. Also read a response to Barr et al: Bates et al, Parsimonious mixed models, ArXiv preprint http://arxiv.org/abs/1506.04967.

Learning Statistics

Further reading

A comprehensive education in statistical theory and applications can be obtained by doing

1. The one-year graduate certificate in Statistics offered by distance education at Sheffield's School of Mathematics and Statistics.
2. The two- or three-year MSc in Statistics also offered by Sheffield.

Learning Statistics

Further reading

For those willing to work on their own:

1. A comprehensive introduction using calculus: Kerns, 2010. Introduction to Probability and Statistics Using R
2. Miller and Miller. 2004. John E. Freund's Mathematical Statistics with Applications.

MSc Cognitive Systems

Taught in English

A unique combination of three research groups covering, language, computation, and cognition:

1. Computational Linguistics (Alexander Koller, Manfred Stede)
2. Computer Science (Tobias Scheffler, Torsten Schaub)
3. Cognitive Modeling (Shravan Vasishth)