

Model Assumptions and Model Selection

Shravan Vasishth

6/11/2019

Goals of this lecture

- learn to check for the normality of residuals
- learn to log transform the data
- learn to compare models to decide which one to use
- learn carry out your hypothesis test using the likelihood ratio test

Chinese relative clause data

```
dat<-read.table("data/gibsonwu2012datarepeat.txt",header=TRUE)
head(dat)
```

```
##      subj item condition pos   rt   region
## 9    1m1   15   obj-ext   8  832 head noun
## 20   1m1    8   subj-ext  8 2131 head noun
## 33   1m1   11   obj-ext   8  553 head noun
## 46   1m1   10   subj-ext   8 1091 head noun
## 62   1m1   16   subj-ext   8  598 head noun
## 75   1m1   14   subj-ext   8  645 head noun
```

```
dat$cond<-ifelse(dat$condition=="subj-ext",-1,1)
m0<-lmer(rt~cond + (1|subj),dat)
summary(m0)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: rt ~ cond + (1 | subj)
##      Data: dat
##
## REML criterion at convergence: 9423
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.765 -0.289 -0.170 -0.004 14.782
##
## Random effects:
##      Groups      Name      Variance Std.Dev.
##      subj      (Intercept) 13491    116
##      Residual              445327    667
## Number of obs: 595, groups:  subj, 40
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    500.7      33.0    15.2
## cond           -54.7      27.4     -2.0
##
```

```
## Correlation of Fixed Effects:
##      (Intr)
## cond 0.004

m1<-lmer(rt~cond + (1+cond||subj),dat)

## boundary (singular) fit: see ?isSingular

summary(m1)

## Linear mixed model fit by REML ['lmerMod']
## Formula: rt ~ cond + ((1 | subj) + (0 + cond | subj))
##      Data: dat
##
## REML criterion at convergence: 9423
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -0.765 -0.289 -0.170 -0.004  14.782
##
## Random effects:
##      Groups      Name      Variance Std.Dev.
##      subj      (Intercept)  13490    116
##      subj.1      cond           0         0
##      Residual                445328    667
## Number of obs: 595, groups:  subj, 40
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    500.7      33.0     15.2
## cond          -54.7      27.4     -2.0
##
## Correlation of Fixed Effects:
##      (Intr)
## cond 0.004
## convergence code: 0
## boundary (singular) fit: see ?isSingular

m2<-lmer(rt~cond + (1+cond|subj),dat)

## boundary (singular) fit: see ?isSingular

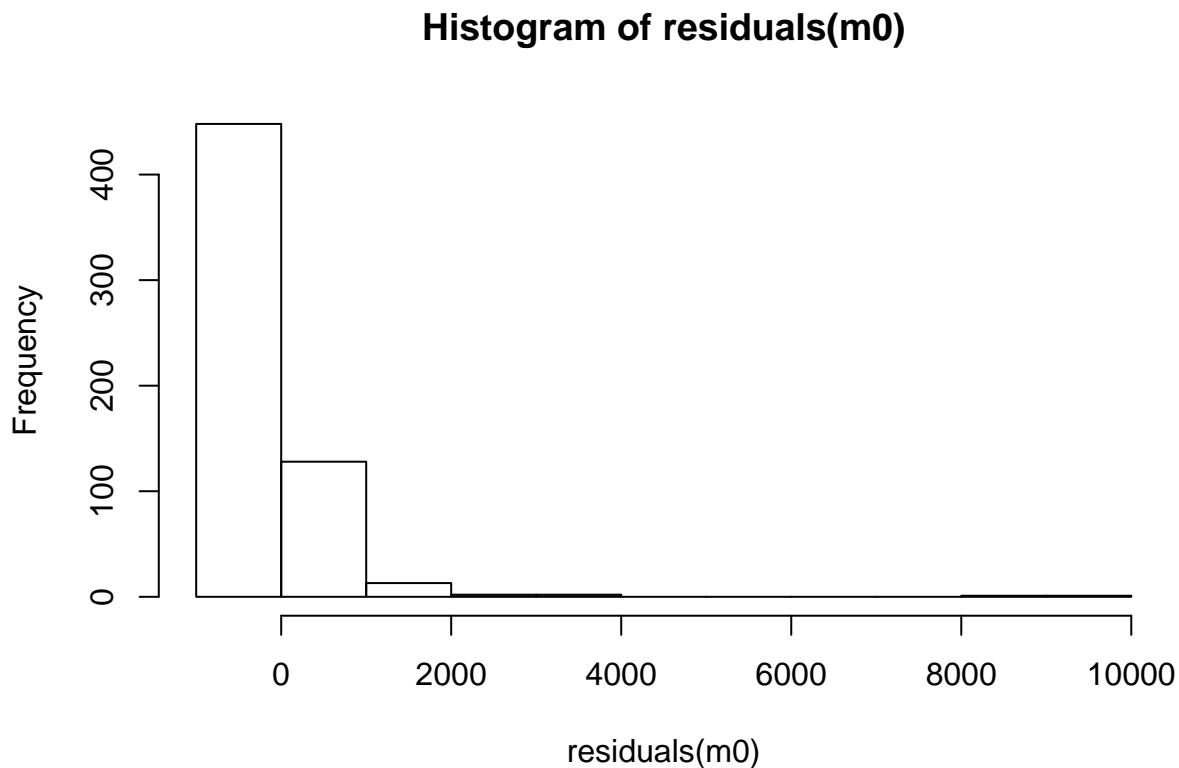
summary(m2)

## Linear mixed model fit by REML ['lmerMod']
## Formula: rt ~ cond + (1 + cond | subj)
##      Data: dat
##
## REML criterion at convergence: 9415.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.436 -0.259 -0.173  0.002  14.376
##
## Random effects:
##      Groups      Name      Variance Std.Dev. Corr
##      subj      (Intercept)  19439    139.4
```

```
##           cond           9134    95.6    -1.00
## Residual           430534    656.2
## Number of obs: 595, groups:  subj, 40
##
## Fixed effects:
##           Estimate Std. Error t value
## (Intercept)    501.5      34.8    14.42
## cond          -54.8      30.9     -1.77
##
## Correlation of Fixed Effects:
##      (Intr)
## cond -0.307
## convergence code: 0
## boundary (singular) fit: see ?isSingular
```

Model assumption: residuals are normal

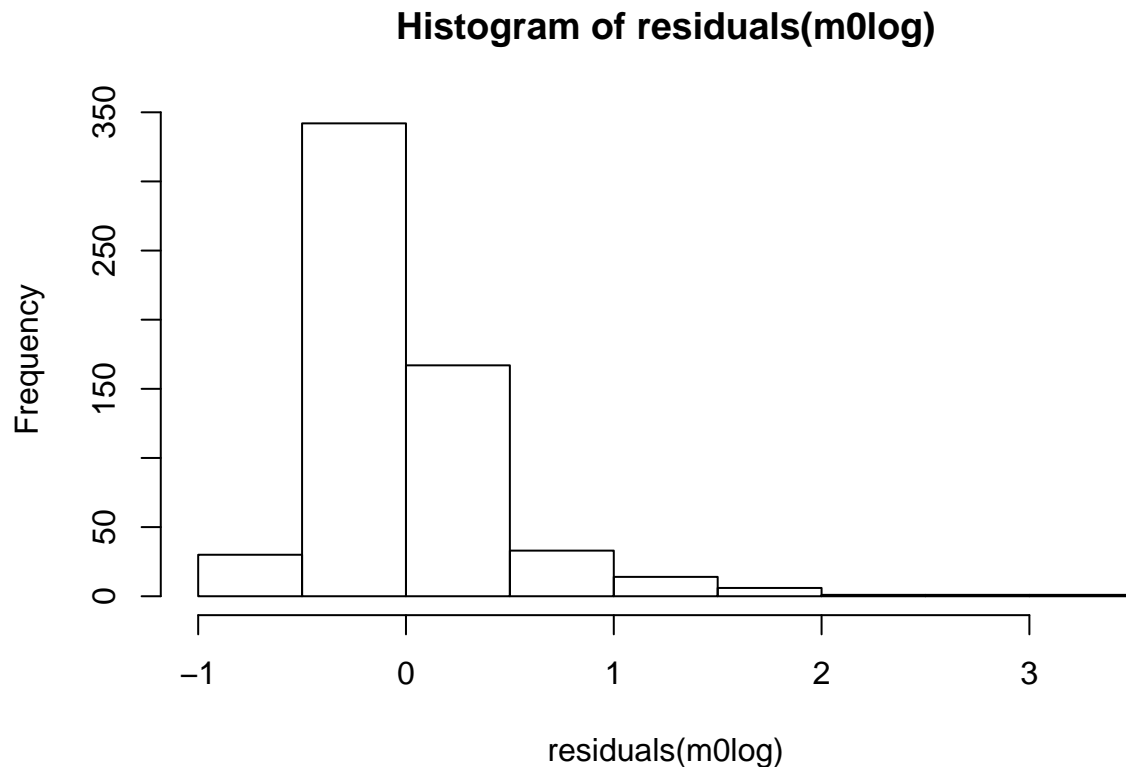
```
hist(residuals(m0))
```



This assumption is clearly violated. A log transform on the reading times will reduce the skew:

```
m0log<-lmer(log(rt)~cond + (1|subj),dat)
m1log<-lmer(log(rt)~cond + (1+cond||subj),dat)
m2log<-lmer(log(rt)~cond + (1+cond|subj),dat)
```

```
hist(residuals(m0log))
```



This is good enough for now.

Model selection

Ignoring model assumptions for a second and analyzing raw rt's, we have three models, m0, m1, m2. Which model is best? There are two schools of thought.

Barr et al 2013: Always fit the maximal model

Under this view, m2 is always the best model (as long as it converges). If it doesn't converge, then back to the most complex model (m1 or m0) that converges.

See: <http://idiom.ucsd.edu/~rlevy/papers/barr-et-al-2013-jml.pdf>

The likelihood ratio test (aka analysis of variance or anova)

We compare the ratios of the likelihoods of the two models of interest. For technical reasons, we have to set a value REML to FALSE in the lmer function when doing model comparison (the reason for this is too advanced for this course).

```
m0<-lmer(rt~cond + (1|subj),dat,REML=FALSE)
logLik(m0)
```

```
## 'log Lik.' -4720.2 (df=4)
```

```
m1<-lmer(rt~cond + (1+cond||subj),dat,REML=FALSE)
```

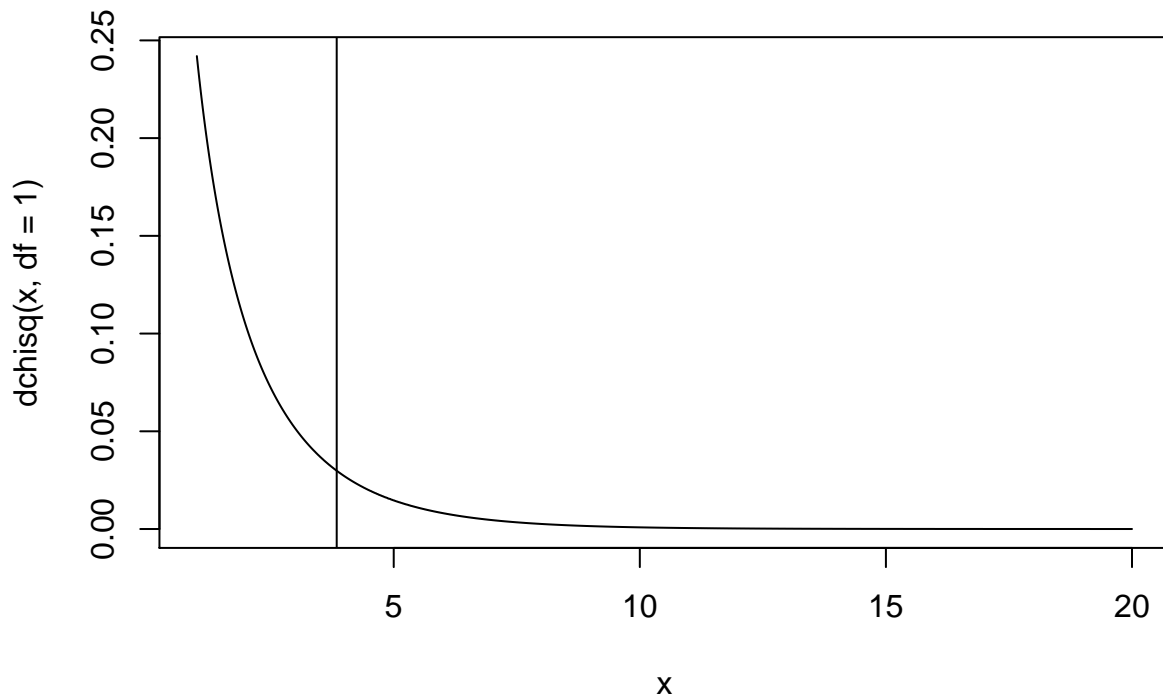
```
## boundary (singular) fit: see ?isSingular
```

```
logLik(m1)
```

```
## 'log Lik.' -4720.2 (df=5)
```

You can see that the ratio is 1. The ratio of the log likelihoods follows a chi-square distribution with the parameter degrees of freedom (df) being the difference in the number of parameters in the two models being compared. In model m0 there are 4 parameters, and in m1 there are 5, so the difference is $df=1$. So the relevant Chi-sq distribution is $\text{chisq}(df=1)$. Let's visualize this and draw the critical chi-sq value (just like the critical t-value in the t-distribution):

```
x<-seq(1,20,by=0.001)
plot(x,dchisq(x,df=1),type="l")
crit_chisq<-qchisq(0.95,df=1)
abline(v=crit_chisq)
```



If the ratio is bigger than the critical chi-squared value (for a given degree of freedom), then we reject the null hypothesis that the two models have the same log likelihoods. If there is no evidence for one model being better, then we choose the simpler model, on grounds of parsimony (Occam's razor). See Bates, Kliegl, Vasishth, Baayen, Parsimonious Mixed Models: <https://arxiv.org/abs/1506.04967>.

In practice, we can use the `anova` function (this is literally the likelihood ratio test I showed above) for model comparison:

```
anova(m0,m1)
```

```
## Data: dat
## Models:
## m0: rt ~ cond + (1 | subj)
## m1: rt ~ cond + ((1 | subj) + (0 + cond | subj))
##      Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m0   4 9448 9466  -4720     9440      0   1      1
## m1   5 9450 9472  -4720     9440      0   1      1
```

```
anova(m1,m2)
```

```
## refitting model(s) with ML (instead of REML)

## Data: dat
## Models:
## m1: rt ~ cond + ((1 | subj) + (0 + cond | subj))
## m2: rt ~ cond + (1 + cond | subj)
##      Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m1   5 9450 9472  -4720     9440
## m2   6 9445 9471  -4716     9433  7.43      1    0.0064

anova(m0,m2)
```

```
## refitting model(s) with ML (instead of REML)

## Data: dat
## Models:
## m0: rt ~ cond + (1 | subj)
## m2: rt ~ cond + (1 + cond | subj)
##      Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m0   4 9448 9466  -4720     9440
## m2   6 9445 9471  -4716     9433  7.43      2    0.024
```

Here, m2 is the best model under the likelihood ratio test.

On the log scale, the conclusion is very different!

```
anova(m0log,m1log)

## refitting model(s) with ML (instead of REML)

## Data: dat
## Models:
## m0log: log(rt) ~ cond + (1 | subj)
## m1log: log(rt) ~ cond + ((1 | subj) + (0 + cond | subj))
##      Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m0log  4 840 857  -416      832
## m1log  5 842 863  -416      832  0.24      1    0.62

anova(m1log,m2log)
```

```
## refitting model(s) with ML (instead of REML)

## Data: dat
## Models:
## m1log: log(rt) ~ cond + ((1 | subj) + (0 + cond | subj))
## m2log: log(rt) ~ cond + (1 + cond | subj)
##      Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m1log  5 842 863  -416      832
## m2log  6 843 870  -416      831  0.02      1    0.88

anova(m0log,m2log)
```

```
## refitting model(s) with ML (instead of REML)

## Data: dat
## Models:
## m0log: log(rt) ~ cond + (1 | subj)
## m2log: log(rt) ~ cond + (1 + cond | subj)
##      Df  AIC  BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m0log  4 840 857  -416      832
```

```
## m2log 6 843 870 -416 831 0.26 2 0.88
```

m0 is good enough. Since the model assumptions are severely violated in the raw reading time analyses, I would only trust the log rt analyses.

Checking if a predictor is significant

After you have decided on which model you want to choose as the final one, you can now do a significance test to test whether a predictor is statistical significant, using the likelihood ratio test.

Suppose we decide on the m0log model. Then, we can check if relative clauses have an effect as follows. The null hypothesis is that $\beta_1 = 0$.

```
## Null model:
m0logNULL<-lmer(log(rt)~ 1 + (1|subj),dat)
## Alternative model:
m0log<-lmer(log(rt)~1 + cond + (1|subj),dat)
anova(m0logNULL,m0log)

## refitting model(s) with ML (instead of REML)

## Data: dat
## Models:
## m0logNULL: log(rt) ~ 1 + (1 | subj)
## m0log: log(rt) ~ 1 + cond + (1 | subj)
##           Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m0logNULL 3 843 856 -419      837
## m0log      4 840 857 -416      832  5.3    1    0.021
```

If you had gone the Barr et al route, then you would do:

```
m2logNULL<-lmer(log(rt)~1 + (1+cond|subj),dat)
m2log<-lmer(log(rt)~1+cond + (1+cond|subj),dat)
anova(m2logNULL,m2log)

## refitting model(s) with ML (instead of REML)

## Data: dat
## Models:
## m2logNULL: log(rt) ~ 1 + (1 + cond | subj)
## m2log: log(rt) ~ 1 + cond + (1 + cond | subj)
##           Df AIC BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m2logNULL 5 846 868 -418      836
## m2log      6 843 870 -416      831  4.54    1    0.033
```

In this example, the conclusion is the same. But we will see later that the conclusion can change depending on whether you fit a maximal model or not.

Notice that if I forget to write REML=FALSE when doing model comparison with anova, the software does it automatically.

How to report your results in a paper.

“A linear mixed model was fit with {varying intercepts/varying intercepts and slopes with no correlation/varying intercepts and slopes, the maximal model}, with SR coded as -1 and OR as +1. The dependent variable

(reading time in milliseconds) was log-transformed. The results show that the object relative clause was read faster than the subject relative clause $\chi^2(1) = 4.54, p = 0.033$.”