A Repeated Measures Random Effects Model

Today we look at a repeated measures random effects model.

**Goals:**

1. Run a normal repeated measures random effects model with covariates.
2. Handling missing data in the outcome variable.
3. Calculating in JAGS the probability that a regression coefficient is greater than zero.
4. Intra-class correlation and other transformations of the parameters.
5. Transformations of the parameters after computing.
6. A t error distribution version of the same model.

# Basic Tasks

Please duplicate the repeated measures analysis of the pain data. When you are comfortable with that, please adapt the model to use t distributions instead of normals. Data files are at <https://faculty.biostat.ucla.edu/robweiss/biostat234-restricted#lab4>. Please note the various features mentioned above. Inspect the data, prior parameters, initial values and understand the structure of each variable. Then run the analysis and get out the necessary output.

# Pain data summary

Our responses are up to 4 repeated measures on 64 kids aged 8-10. The response is a proxy measurement of pain tolerance, measured in seconds. The kids put their arms in very cold water. When the arm hurts too much, they remove the arm, and the total time of immersion is recorded. The response is log transformed, as it is quite skewed on the original scale of measurement.

Several kids have fewer than the full 4 repeated measures. These kids have *missing data*.

Each kid has two covariates, coping style (CS) either 0 or 1. CS=0 are *attenders* (A) who pay attention to their arm, the feelings, the experimental apparatus. CS=1 are *distractors* (D), who think about other things such as homework or a vacation. Just before the last (fourth) measurement, an intervention is given, a counseling treatment (TMT) to either attend (A), distract (D), or None (N). The treatment is assumed to interact with CS. During the first three times, kids are only in the baseline (B) status, either CS=0 or 1. At the last time point, they are in one of 6=3\*2 treatment groups. Kids are also expected to have their own typical response time, this is handled by a kid-specific intercept, a random effect.

**The** **Repeated Measures Random Effects Model**

The model is

yij = xijt + i + ij,

where i runs from 1 to 64, j runs from 1 to 4, with j=4 indicating the response after treatment. The i's are modeled as normal(0, D) and the ij as independent and normal(0,2). The presence of the i's introduces correlation between the observations within each kid. It is convenient to think of the yij’s as belonging to a 64 by 4 matrix. The x’s belong to a 64 by 4 by 8 array, which was turned into a 64\*4 long by 8 across matrix. The x’s are all 0’s and 1’s.

Here are the first 16 rows of the x array.

x[,1] x[,2] x[,3] x[,4] x[,5] x[,6] x[,7] x[,8]

1 0 0 0 0 0 0 0

1 0 0 0 0 0 0 0

1 0 0 0 0 0 0 0

1 0 1 0 0 0 0 0

1 1 0 0 0 0 0 0

1 1 0 0 0 0 0 0

1 1 0 0 0 0 0 0

1 1 0 0 0 0 1 0

1 0 0 0 0 0 0 0

1 0 0 0 0 0 0 0

1 0 0 0 0 0 0 0

1 0 0 0 1 0 0 0

1 1 0 0 0 0 0 0

1 1 0 0 0 0 0 0

1 1 0 0 0 0 0 0

1 1 0 0 0 1 0 0

The first column is the intercept. The second column indicates if the kid is a attender (=0) or distractor (=1). The parameter alpha[1] is the intercept, and alpha[2] is the difference between distractors minus attenders at baseline. Each set of 4 rows has been separated by lines (which do not exist in the original file on my web site). The first 4 rows belong to kid 1, the second 4 to kid 2 and so on. The first and third kids are attenders, the second and fourth kids are distractors.

The last 6 columns indicate which coping style by treatment group the kid is in. We believe there is an interaction between coping style and treatment; hence there is a treatment effect for each coping style. The first three rows of the last six columns are always zeros, because the treatment has not yet been given. The last row, last 6 columns, has a single one indicating which of the 6 groups the kid is in. (Why are there 8 x variables? Because there are two coping style groups and treatment can be in one of four groups: baseline (B), or tmt A, D or N, and 2\*4=8.) The last six columns are AA, AD, AN, DA, DD, and DN respectively. So kid 1 is an attender taught to attend or AA, kid 2 is a distractor taught to distract or DD, kid three is an attender taught nothing or AN, and the fourth kid is a distractor taught to attend or DA. Thus there are 6 TMT effects in the model.

First four rows of the response data

list(y=structure(.Data=

c(3.022860 , 3.564166 , 2.659559 , 2.460443 ,

3.336836 , 3.187178 , 2.763800 , 3.010620 ,

2.484906 , 2.302585 , 2.505525 , 2.112634 ,

2.742773 , 3.121483 , NA , NA ,

Now those are on the log scale. Here is the unlogged original data in seconds.

list(y=structure(.Data=

c(20.55 , 35.31 , 14.29 , 11.71 ,

28.13 , 24.22 , 15.86 , 20.30 ,

12.00 , 10.00 , 12.25 , 08.27 ,

15.53 , 22.68 , NA , NA ,

These values are in seconds, recorded to the nearest 100th of a second. You will use the logged data for this analysis. You’ll notice that kid one has measurements ranging from 11 seconds up to 35 seconds. Kid two has times ranging from 15 to 28 seconds, kid three has times that are definitely somewhat shorter than kids 1 or 2. Kid four has two missing observations for times 3 and 4. These times will not contribute to the analysis, as they have not been observed. Bayesian analysis treats the missing y’s as random variables, and JAGS samples from the posterior of all the parameters and all the missing data also. These distributions for the missing data are *predictions* of what might have occurred, had the observations actually been observed. The predictions take into account the covariates, and the kid level intercept i.

The parameters in this analysis are alpha[1:8], tau.e = 1/sigma2, tau.b=1/D, and the beta[1:64]’s. The beta’s are random effects giving each kids average level above or below the typical level of an attender alpha[1] or distractor alpha[1] + alpha[2]. The beta[i]’s have mean zero and the kid intercept is alpha[1] + beta[i] for attenders and alpha[1] + alpha[2] + beta[i] if the kid is a distractor.

**Model.**

model

{

for( i in 1 : 64 ) {

for( j in 1 : 4 ) {

s[i, j]<-4\*(i-1)+j

y[i, j] ~ dnorm(mu[i , j],tau.e)

mu[i , j] <- inprod(x[s[i,j],],alpha[])+beta[i]

}

beta[i]~dnorm(0, tau.b)

}

for( k in 1:8) {

alpha[k]~dnorm(m[k],varinv[k])

alphasign[k] <- step(alpha[k])

}

tau.e ~ dgamma(ea,eb)

tau.b~dgamma(ba,bb)

sigma <- 1 /sqrt( tau.e)

sqrtD <- 1 /sqrt( tau.b)

rho <- sqrtD\*sqrtD/(sigma\*sigma + sqrtD \*sqrtD)

}

The likelihood is given at the top, inside the doubly nested loop. The outer loop on i is over cases (kids), and the inner loop j is over observations within kid. The calculation of s[i,j], is to find the correct row in the 64\*4 by 8 X matrix. Please check that s[i,j] runs from 1 to 64\*4. In particular, for kid 1 it runs from 1 to 4 as j increases from 1 to 4. Then for kid 2, it runs from 5 to 8, and those are the rows corresponding to kid #2’s covariate data. The line for mu[i,j] calculates the sampling mean for kid i at time j. The sampling variance is 2. Inside the outer loop only is the prior for the beta[i]’s or i’s (what happens if you put it in the inner loop? oops). The i’s have variance D = 1/tau.b.

The next loop on k from 1 to 8 specifies a normal prior for alpha[k]. The sign of the means m[k] was given by the researcher who created the data set, I set the magnitudes. If you take the exponent of the m[k] values you will find that they are logs of .9, 1.2, or 1.0, indicating whether there was expected to be any difference between the baseline groups m[2], or whether the various treatments were expected to have any effect. The null treatment was expected to have no effect, while the same treatment (AA or DD) as the child’s natural coping style was expected to help kids (by 20% a priori) while the opposite treatment (AD or DA) was expected to hurt ( by 10% a priori). The alphasign[k] variable is the sign of the underlying alpha[k] parameter.

The variances in the prior data were chosen ad hocly. The original Bayesian analysis of this data (Weiss, Wang and Ibrahim 1997, Biometrics) had a rather complex construction for the variances. However, these are the original prior means that were used there.

The priors for the variances are Inverse Gammas, and used 4 degrees of freedom or a=2 and an elicited point estimate to set the other parameter of the Inverse Gamma.

Finally, the *intra-class correlation coefficient* describes the correlation between observations within kid. If it is high, it means that repeated observations are highly correlated; if low, repeated measures within kid are not well correlated. If rho=1, then the repeated observations are really only a single observation as they are perfectly correlated. (This would happen if I asked you your monthly rent cost four times in a row on a single day.) If rho=0, then each repeated measure is of the equivalent value of a whole new kid’s observation. The intraclass correlation coefficient  = D/(D+2) is a function of the *within-subject variance* 2 and the *between subject variance* D.

# Prior Data

list(m=c(2.77,.182,.182,-.105,0,-.105,.182,0),

varinv=c(.5, 1, 1, 1, 1, 1, 1, 1),

ea=2, eb=.2178,

ba=2, bb=1.7)

**Starting Values: Random Effects and Missing Data**

Two features of the starting values.

One is the need for starting values for the beta parameters. I used mostly zeros which are the prior means for these parameters. This is a case where one can use JAGS’ built in initial values (by omitting the beta initial values), but I find this sometimes causes serious problems which are avoided when I set the initial values myself. The reason to not have all exact zeros is that if JAGS tries to calculate the variance of the beta[i]s, the variance won’t be zero in the beginning.

Second, there were missing y values, and the starting values also have starting values for some of the y’s. If you look at the y or the log(y) matrix, you will find that if there is a data point in the original data matrix, then there is an “NA” in the start values. If there is an “NA” in the data values, then that parameter needs a start value. Since there are 11 missing observations in this data set (broken bones, kids absent for the last two trials), there are 11 non-NA values in the initial values for y.

**TODO 1:** Please run this model, and get the output.

See the script file “lab4JAGS.R”. How the data were created is in the file “lab4 data setup script.R”. Graphs are in the file “lab4JAGSGraphs.R”.

**TODO 2:** T model.Change the normal densities for the data and the random effects to t densities. In previous analyses of this data set with the normal model I had noticed that there were potentially a number of outliers, particularly in the ij residuals. **Modify** the normal model to run with t errors instead of normal errors. The normal model is given in the R document. You can modify the original normal model program to get to the t model. The changes are that the two dnorm’s for the y’s and for the beta’s become dt’s with degrees of freedom (df) dfy and dfbeta. Possible priors include

Each df will need a distribution. Perhaps use

y[i,j] ~ dt( mu[i,j] , tau.e, df1)

df1 <- 1/invdf1

invdf1 ~ dunif(0,.5)

This keeps the degrees of freedom to be at least 2. A t with 1df is the Cauchy distribution, a t with more than 2 df has a mean and a variance. Do the same for the t-error distribution for the random beta[i] effects.

1. Turn in results from the t-model. Be sure to run sufficient iterations.
   1. How is the convergence? Show an illustrative autocorrelation function and time-series plot for two parameters of interest.
   2. Turn in a table of results for the fixed effects, the two standard deviations sqrtD and , and the two degrees of freedom parameters. Label rows appropriately and format the table carefully.
2. Compare the results from the normal model to the results from the t model: What changes are there? In particular, what scientific conclusions change?
3. Reproduce figures 1-5 (see below for the normal model figures) for your t model. **Label** your figures appropriately.
4. Invent another prior for the df, and in one sentence explain its properties (ie support, mean, sd or other characteristics) and why it is better than the above prior.
5. Extra Credit: (i) Which model fits better t or normal? You can judge this from the posterior for the two df parameters: do the posteriors provide support for values of the df parameters that are at all close to the normal model? We can also use DIC to compare models. The model with the lowest DIC value fits the data better. Which model has the better DIC?

Figures from the normal random effects model.

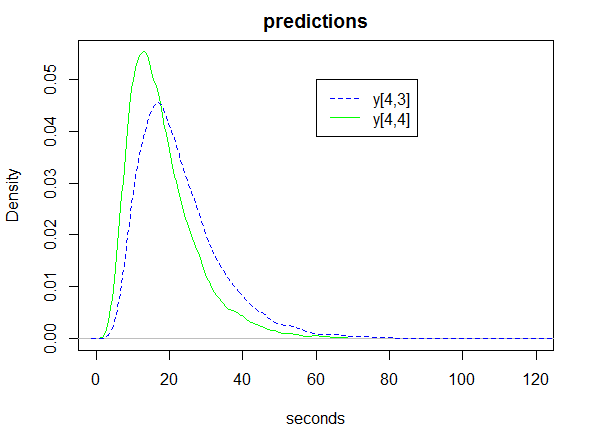


Figure 1. Predictions for subject 4’s missing 3rd and 4th observations.

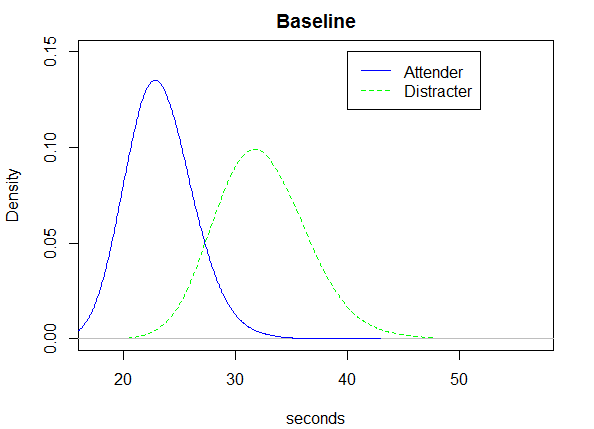


Figure 2. Baseline predicted median pain tolerance (seconds) for attenders and distracters.

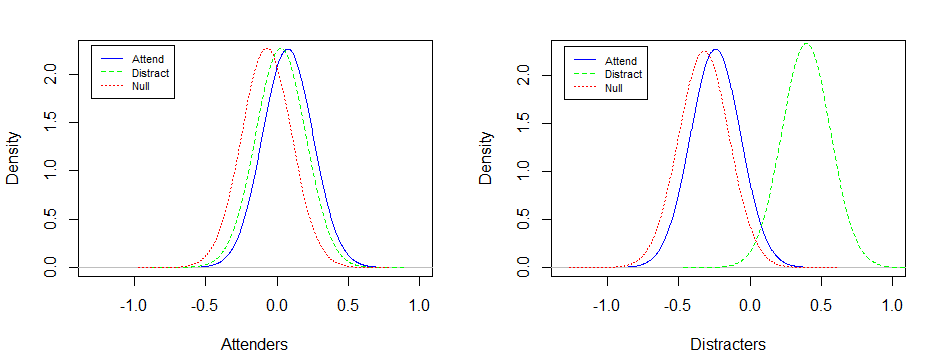


Figure 3. Treatment effects for attenders (left) and distracters (right), log scale.

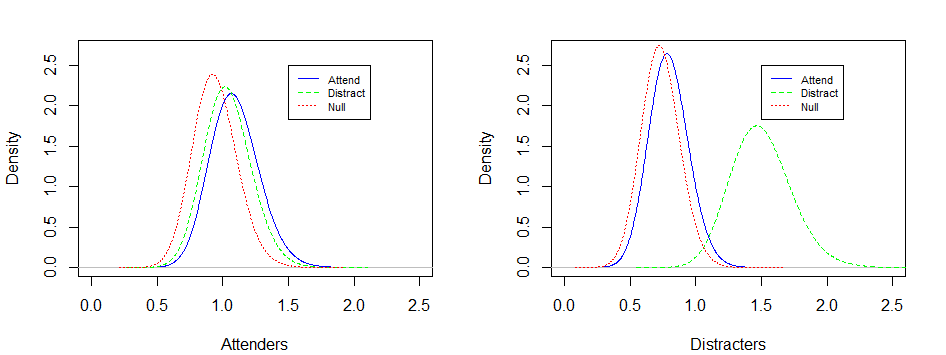


Figure 4. Treatment effects for attenders (left) and distracters (right), multiplicative scale.

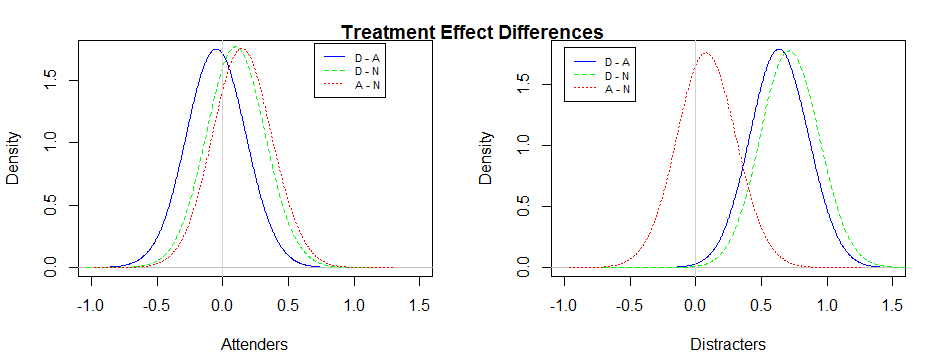


Figure 5. Pairwise differences between treatments, attenders (left), distracters (right).