36-463/663: Multilevel and Hierarchical Models
Project, Part II
Due: FRIDAY DECEMBER 16.

Please submit as a single pdf on Blackboard.

Instructions & General Information

- DO NOT WORK WITH OTHERS. Please do this project entirely on your own. Acceptable resources are:
  - The Gelman & Hill text, and the Lynch text.
  - R, and anything you can install in R as a library package.
  - Any incidental computing aid such as a calculator, excel spreadsheet, etc. Please cite any such resources in your final submission.
  - Class notes or anything else I have handed out in class, or posted on either of these websites:
    - `http://www.stat.cmu.edu/~brian/463-663/`
    - `http://www.cmu.edu/blackboard/`
  - Talking with or emailing the instructor (BJ) or TA (Nicholas Kim).
  - *Static* resources on the www or in the library, For example:
    - **These are OK**: books, webpages and pdf's; *Cite these as references in your final submission.*
    - **These are NOT OK**: email (except with BJ or the TA), chats, IM, social networking sites, etc.; *Don't use them.*[1]
  - Anything not in the above list? Check with BJ *first*.
  - Anything you used? *Cite it* **whether it is on the OK list or the not-OK list**.

  Submitting your work for this part of the project constitutes your personal guarantee that you worked on your own, following strictly the resource guidelines above. Violation of your guarantee will result in a grade of 0 (zero) on this part of the project.
- Please assemble your work for this part of the project into a single pdf document and submit on Blackboard. Late submissions will not be accepted[2].
  - Use the filename "project-part-two-junker-brian.pdf" (with your last name and first name, not mine!).
  - Organize your work so that it is easy to read and easy to find the answers to each part of each question below. Clearly label all sections / subparts / tables / graphs / etc. **Answers that are not easy to find or read will receive no credit!**
- This part of the project will be worth 10% of your final grade.

[1] Except for the class Q&A Discussion Board on Blackboard, please don't use discussion boards either. You may post questions on the Q&A board, but let me or Nick answer them.

[2] If there is some good reason you can't turn it in on time, please talk to me about *well in advance*.

Problem

The **Setting** and the **Data** are the same as in Project Part I. You can again find the data set `dec-data.csv` in the same directory as these instructions.

In this part of the project we will be interested in building hierarchical linear models for the *reaction time* for each player's attempt at each question, to determine what factors make a player take longer or shorter to answer a question in *Battleship Numberline* (`http://playpowerlabs.org/bsnl/v21/brainpop/BSNL.html`).

Once again, since answers to questions from the same player are likely to be more similar to each other than answers from different players, players are a natural choice for a random effect. It's also possible that part or all of the IP number, or some other grouping, might be a second random effect.

As before, you can learn about the design of Battleship Numberline and some of Dr. Lomas' research questions from the short research report Lomas et al. (2013); a copy of the article is provided in the same subdirectory as this problem statement. Refer to the instructions for Project Part I for more details.

References

Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* NY: Cambridge Univ Press.

Lomas, D., Patel, K., Forlizzi, J.L., and Koedinger, K.R. (2013). Optimizing Challenge in an Educational Game Using Large-Scale Design Experiments. Paper presented at CHI 2013, Paris, France. Obtained online at
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.480.2493&rep=rep1&type=pdf

Lynch, Scott M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists.* New York: Springer.

Schofield, L. S., Junker, B., Taylor, L. J., & Black, D. A. (2014). Predictive inference using latent variables with covariates. *Psychometrika,* 283–314.

van der Linden W.J. (2009). Conceptual Issues in Response-Time Modeling. *Journal of Educational Measurement, 46(3),* 247–272.

The Goal

Analyze this data to determine what factors affect the length of time it takes players to respond to questions, and whether there are other factors besides the individual question, that affect the response time. Account appropriately for any important grouping effects, and any other variables of interest.

This part of the project has has five sections/problems, and is worth a total of 100 points.

Please adhere to the page limits suggested for each part of the project.

Also, please attach two appendices to your hw paper:

- **Appendix I** (mandatory) a list of all materials you referred to, formatted in the same style as the list of references on pp. 575 ff. of Gelman & Hill. *Please adhere closely to the style of references in G&H. Failure to do so may result in misunderstanding about your work and loss of points on this takehome.*

- **Appendix II** (optional) any material (R code, computer output, figures, tables, etc.) that you think is important for your homework but would not fit in the page limits for each part of the project above. *The more stuff that is in here, the less likely that I will consider it when grading your project, so please choose wisely what (if anything) to put here.*

1. *Thinking about the response variable reacTime.* [20 pts]

   Our main response variable for this part of the project will be reacTime.

   (a) Explain why lrt = log(reacTime) might make more sense to use as a response variable, than reacTime itself, if we are going to build standard hierarchical linear models with normal errors and normal random effects. Make sure to write complete, clear sentences; feel free to use figures or calculations, if it will help.

   (b) The variable timeLimit is equal to 10 for all observations in the data set; this is intended to mean that all players have a maximum of 10 seconds to answer, or their answer will be coded as "time's up!!", which in our analysis is a wrong answer (resp=0).

      i. Look for, and report, any evidence about how well or poorly this time limit is adhered to, in the data.

      ii. How might this time limit affect fitting standard linear regression models, hierarchical linear models with normal random effects, etc.? How might it affect residual plots?

      iii. Is there anything else about the reacTime variable that might lead you to recode or remove some values? If so, say what you found, and what you will do about it.

   Page limit for section 1: 2 pages.

2. *Ordinary linear regression for question effects.* [18 pts]

   (a) Fit a linear model for lrt, using currentQuestion *as a factor*, and omitting the intercept. Summarize the fit, and comment on why taking the intercept out might be a useful idea.

   (b) Plot the coefficients in this model against each of the following:

      - Item easiness (the proportion of players attemping the item, who got it right (resp=1))
      - The mean of all reaction times on this question
      - The log of the mean reaction time on this question
      - The target value for this question (currentQuestion as a numeric variable)

      Summarize the results, and provide and comment on any interesting plots (you do not have to provide plots that aren't interesting!).

   (c) Use an automatic variable selection procedure to find the best model that *does* include currentQuestion as a factor, but *does not include* SID in any form.

      i. Show or tell what variable selection method(s) you used, and what were the smallest and largest models you considered.

      ii. Show the final model and write a short paragraph interpreting the fitted model and parameter estimates etc. in the fitted model. *Remember, you are writing for Dr. Lomas, so provide a clear interpretation that will be useful to him in understanding the relationships between predictors and outcome in the model.*

      iii. if resp, or easiness, or CurrentQuestion as a numeric variable, is in the final model, does eacgh of their relationships with lrt make sense? Explain why or why not.

   Page limits for section 2: one page for each of the three subparts.

3. *Ordinary linear regression for player effects.* [20 pts]

   (a) Fit a linear model for lrt, using SID *as a factor*, and omitting the intercept. Summarize and interpret the fit.

   (b) Plot the estimated coefficients of this model agaist each of the following:

   - proportion correct: The proportion of items attempted by that player, which the player got right.
   - The player random effects from the model `glmer.resp <- glmer(resp ~ cq - 1 + (1|SID),family=binomial)`, where `cq <- factor(currentQuestion)`.
   - The mean reaction time of each player, over all the questions that player tried
   - The log of the mean reaction time

   Summarize the results, and provide and comment on any interesting plots.

   Page limits for section 3: 1 page for each of the 2 subparts.

4. *Mixed Effects Models* [20 pts]

   Our primary interest is in what factors affect the reaction time of each player on each question It is quite likely that questions answered by the same person will be dependent on each other through that player's ability. For these reasons, we will focus on questions as fixed effects and players as random effects.

   (a) Fit a standard mixed effects linear regression model predicting lrt, using currentQuestion as a fixed effect *factor*, omitting the intercept, and with a random intercept grouped by SID (that is, all these should be implemented in your model). Summarize and interpret the fit.

   (b) Use an automatic variable selection method to try to improve this model.

   i. Show or tell what variable selection method(s) you used, and what the smallest and largest models (both fixed and random effects) you considered were.

   ii. Show the final model and write a short paragraph interpreting the fitted model and parameter estimates etc. in the fitted model. *Remember, as before, you are writing for Dr. Lomas.*

   Page limits for section 4: 1 page for each of the 2 subparts.

5. *Models combining reaction time and correctness of response.* [22 pts]

   Many researchers believe that reaction time and correctness of response should be related in some way. For example, perhaps, as the question gets harder to answer correctly, the reaction time should go up. Or perhaps reaction time is not so much related to correctness of response as it is to how "engaging" the player finds the game to be.

   (a) Review your work from Section 1 (and possibly some of your work from Project Part I;) and try to determine whether one of these theories (or some other theory), about the connection between reaction time and correctness of response, is suggested. Write a short paragraph, with figures or other displays if needed, giving evidence for your conclusion(s).

(b) van der Linden (2009, p. 254)[3] suggests, essentially, that we make the logit of the probability of a correct response one of the predictors for the response time model. We can get the logits of the success probability for each question encountered by each player with `lp <- predict(glmer.resp)`, where `glmer.resp` is the model referred to in section 3 above.

   i. Using graphs and an ordinary linear regression model, explore the relationship between lrt and lp. Is lp a good predictor of lrt? Is the direction of the relationship what you would expect?

   ii. Use a variable selection procedure to try to improve the linear model you fitted in part (i). Summarize and interpret the fit of your final model.

(c) A basic problem with the linear regression in part (b) (i) here, is that it assumes that lp is fixed and known, but in fact lp is only estimated (through estimates of fixed and random effects in the model `glmer.resp`). The multilevel Bayesian model in Figure 1 (below) tries to correct this, by modeling lrt and resp at the same time, using the same components as the models in part (b) (i).

   i. Write, using the most precise mathematical language that you can use, the model represented in Figure 1. It is not necessary to write out the prior distributions, just get the structure of the model at levels 1 and 2 as clear and correct as possible. Feel free to use the code in Figure 1 as well as the modeling components in part (b) (i), to figure this out.

   ii. Unfortunately the mcmc fit does not run very fast—it takes about two hours for a typical useful run. Rather than make you wait for the runs, I have produced three fitted model objects, `joint.mcmc.test`, `joint.mcmc.0` and `joint.mcmc.1`, using the code in Figure 2. You can load these fitted models into your R session by copying the file fitted-mcmc.RData from the Project Part 2 directory into your working directory on your laptop, and issuing the command[4] `load("fitted-mcmc.RData")` in R.

      A. Inspect the fitted model objects `joint.mcmc.0` and `joint.mcmc.1` with the print() and p3() commands, and any other tools if they seem useful. Does it look like, for either fit, the Markov chain has converged to its stationary distribution? Why or why not? Are the estimates from either run useful and reliable? Why or why not? Provide evidence for your conclusions.

      B. The setup for fitting `joint.mcmc.1` was slightly different from the setup for fitting `joint.mcmc.0`. Describe the difference, and explain why this might have contributed to a different level of success for one fit vs the other.

   iii. Choose whichever of `joint.mcmc.0` or `joint.mcmc.1` was the more successful and reliable fit, and interpret the fitted model. Compare it with the fitted model in part (b) (i). Is the conclusion about the relationship between lrt and lp the same[5]?

Page limits for section 5: 1 page each for parts (a) and (b); up to 3 pages for part (c).

---

[3]See references on page 2 of this document.

[4]If this does not work for you, you will have to re-run the models yourself, using the code in Figures 1 and 2. This takes about five hours.

[5]I have reason to trust the result of the Bayesian model more; see for example Schofield et al., (2014), listed in the references above, for a similar situation.

```
joint.model <-
"
data{
    for (n in 1:Nresp) {
        lrt[n] <- log(reacTime[n])

        # this is a data transformation, which
        # is why it's in its own section in the
        # jags model
    }
}

model {

    # LEVEL 1
    for (n in 1:Nresp) {
        resp[n] ~ dbern(p[n])
        logit(p[n]) <- lp[n]
        lrt[n] ~ dnorm(rho*lp[n],invsig2)
        lp[n] <- b[question[n]] + alpha[player[n]]
    }

    # FIXED EFFECTS PRIORS
    for (j in 1:J) {
        bb[j] ~ dnorm(0,1e-10)          # the question parameters get
    }                                   # flat priors...
    for (j in 1:J) {
        b[j] <- bb[j] - mean(bb[])      # ...except that we make them
    }                                   # sum to zero...
    rho ~ dnorm(0,1e-10)

    # LEVEL 2
    for (i in 1:N) {
        alpha[i] ~ dnorm(b0,invtau2)    # ...so that we can have an
    }                                   # overall intercept at level 2

    # LEVEL 2 PRIORS
    invtau2 <- pow(tau,-2)
    invsig2 <- pow(sig,-2)
    tau ~ dunif(0,10)
    sig ~ dunif(0,10)
    b0 ~ dnorm(0,1e-10)
}"
```

Figure 1: JAGS code for jointly modeling response time and correctness of response. This code is also contained in the file “joint-model.r”.

```
joint.data <- list(resp=resp,reacTime=reacTime,question=as.numeric(cq),
                   player=as.numeric(sID),Nresp=length(resp),
                   N=length(unique(sID)),J=length(unique(cq)))

joint.inits <- function() {
    return(list(tau=runif(1),sig=runif(1),bb=rnorm(20),rho=rnorm(1),b0=rnorm(1)))
}

rube(joint.model,data=joint.data,inits=joint.inits)

joint.mcmc.test <- rube(joint.model,data=joint.data,inits=joint.inits,
                   n.chains=1,n.iter=1000,
                   parameters.to.save=c("b","sig","tau","alpha","rho","b0"))

joint.mcmc.0 <- rube(joint.model,data=joint.data,inits=joint.inits,
#                   parallel=T,
                   parameters.to.save=c("b","sig","tau","alpha","rho","b0"))

b.mean <- joint.mcmc.0$median$b
b.sd <- joint.mcmc.0$sd$b

tau.mean <- joint.mcmc.0$median$tau
tau.sd <- joint.mcmc.0$sd$tau
sig.mean <- joint.mcmc.0$median$sig
sig.sd <- joint.mcmc.0$sd$sig

rho.mean <- joint.mcmc.0$median$rho
rho.sd <- joint.mcmc.0$sd$rho
b0.mean <- joint.mcmc.0$median$b0
b0.sd <- joint.mcmc.0$sd$b0

joint.inits <- function() {
    return(list(tau=runif(1,tau.mean-2*tau.sd,tau.mean+2*tau.sd),
                sig=runif(1,sig.mean-2*sig.sd,sig.mean+2*sig.sd),
                bb=rnorm(20,b.mean,b.sd),
                rho=rnorm(1,rho.mean,rho.sd),
                b0=rnorm(1,b0.mean,b0.sd)))
}

joint.mcmc.1 <- rube(joint.model,data=joint.data,inits=joint.inits,
#                   parallel=T,
                   parameters.to.save=c("b","sig","tau","alpha","rho","b0"))
```

Figure 2: R code for running the model in Fig 1. This code is also contained in the file “joint-model.r”.