

Bayesian Multilevel Models COURSE: Homework for Missing Data.

1. This problem deals with missing data in a linear model. First download the data from the syllabus, load it into R, and graph the data using:

```
star98.missing <- read.table("star98.missing.dat.txt",header=TRUE)
par(mfrow=c(1,2),mar=c(3,3,3,3))
plot(star98.missing$SUBSIDIZED.LUNCH,star98.missing$READING.ABOVE.50,pch="+",col="blue")
abline(lm(star98.missing$READING.ABOVE.50~star98.missing$SUBSIDIZED.LUNCH),lwd=3)
mtext(side=1,cex=1.3,line=2.5,"District Percent Receiving Subsidized Lunch")
mtext(side=2,cex=1.3,line=2.5,"District Percent Above National Reading Median")
plot(star98.missing$PTRATIO,star98.missing$READING.ABOVE.50,pch="+",col="blue")
abline(lm(star98.missing$READING.ABOVE.50~star98.missing$PTRATIO),lwd=3)
mtext(side=1,cex=1.3,line=2.5,"District Pupil/Teacher Ratio")
mtext(side=2,cex=1.3,line=2.5,"District Percent Above National Reading Median")
mtext(side=3,cex=1.5,outer=TRUE,line=-1,"California 9th Grade by District, 1998")
```

Determine how much missing data there is and if there is a discernable pattern. Now use mice to run a new model. Also run a model *omitting* cases with missing data. What differences do you observe? Which is better?

2. Explain what the following R does and why you would not want to do this.

```
mi <- function(data.mat) {
  for (i in 1:ncol(data.mat)) {
    if (sum(is.na(data.mat[,i])) > 0) {
      print(paste("column",i,"has missing data"))
      mean.col <- mean(data.mat[,i],na.rm=TRUE)
      for (j in 1:nrow(data.mat)) {
        if (is.na(data.mat[j,i]) ==TRUE) data.mat[j,i] <- mean.col
      }
    }
  }
  return(data.mat)
}
```

3. Find an article in your literature that uses case-wise deletion. Discuss how you might replicate the model and improve the work.