

1. Each year the U.S. Naval Postgraduate School sets aside a “Discovery Day” during which the general public is invited into their laboratories. Our data come from 21 October 1995, when visitors could test their reaction times and hand-eye coordination in the Human Systems Integration Laboratory. The variable of interest, “anticipatory timing,” was measured by a Bassin timer, which measures a person’s ability to estimate the speed of a moving light and its arrival at a designated point.

The Timer consists of a 10 foot row of lights which is controlled by a variable speed potentiometer. The lights are switched on sequentially from one end to the other so that light “travels” at 5 miles per hour down the Timer. Each visitor was instructed to anticipate the “arrival” of the light at one end of the Timer and at that time to swing a plastic bat across a light beam at the same end of the Timer. An automatic timing device measured the difference between the breaking of the beam and the actual arrival of the light. A negative value of a trial variable indicated the bat broke the beam before the light actually arrived, while a positive value indicated that the bat broke the beam after the light arrived.

Each of 113 visitors completed the trial five times; there are no missing data. Age and gender were also recorded, since the researchers were interested in age and gender differences in reaction times. Visitors tended to come in family groups, but that information was not recorded.

You can find these data from the file `timetrial.dat` off the course web page, or in `/home/merganser/course_data/correlated.data/timetrial.dat`. These data are sorted with one row per person and one column per each of the five trials. Depending on what software you use for graphing, they may need to be re-arranged to have one row per person per trial instead. At the end of this Homework, you will find SAS, R, and S-Plus code to do this. Make sure you understand how this re-arrangement is done; we’ll do it many times before the end of the semester.

- (a) Suppose we are not interested in whether a participant broke the beam too early or too late; we are only interested in the magnitude of how far off the participant was in timing. **Take an appropriate transformation of the outcome variable.** Then explore both the mean and covariance structures using eight of the plotting techniques discussed in class. For each plot, comment in *one or two sentences* on the type of information the plot tells you. For example, describe any trends (or lack thereof) that you observe. In particular, your eight plots should *at least* address the following: Are visitors in general improving across the five trials? Does the mean trend in the outcome across trials differ by gender? By age? Do your plots indicate that an interaction between age and gender might be needed? How does the variability in the outcome differ by gender or age? What type of correlation structure do the data suggest?
- (b) We are told that visitors tended to come in family groups, but that information was not recorded. How might that affect the correlation structure we observe in these data?

More on the back...

2. From the BioMed Library, find the article “Near vision impairment predicts cognitive decline: Data from the Hispanic Established Populations for Epidemiologic Studies of the Elderly” by Reyes-Ortiz, et al. from *Journal of the American Geriatrics Society*, 2005; 53:681-686. Read the Methods and the Results sections.
 - (a) This study had repeated measures taken on each participant. What was measured repeatedly, and when? How was the primary outcome variable defined (data shown in Figure 1)?
 - (b) What type of model was fit to the primary outcome variable (model results shown in Table 2)?
 - (c) Did the authors describe any attempts to check the model’s assumptions? If so, what did they check?
 - (d) What is the source of correlation among the observations in this study’s data?
 - (e) What is the overall conclusion of the study?

Code to re-arrange the data set for Problem 1: Note that this is far simpler to do in SAS than R or S-Plus, so you may want to re-arrange in SAS and then read that data set into R, S-Plus, or whatever graphing package you prefer. The SAS code below also shows how to write the re-arranged data to a space-delimited file that is easily readable by R, S-Plus, or many other packages.

SAS code:

```
data trial;
  infile '/home/merganser/course_data/correlated.data/timetrial.dat' firstobs=2;
  input Sex $ Age Trial1-Trial5;
run;
data repdat;
  set trial;
  time=trial1; trial=1; subj=_N_; output;
  time=trial2; trial=2; subj=_N_; output;
  time=trial3; trial=3; subj=_N_; output;
  time=trial4; trial=4; subj=_N_; output;
  time=trial5; trial=5; subj=_N_; output;
run;
data _NULL_;
  set repdat;
  file 'new.timetrial.dat';
  put time sex age trial subj;
run;
```

To read this data set `repdat` into R or S-Plus:

```
data = read.table('new.timetrial.dat',header=F,col.names=
  c('time','sex','age','trial','subj'))
```

R code:

```
data = read.table(file="/home/muskie/correlated.data/timetrial.dat",
                  header=T)
nsub = length(data$Age)
ntimes = 5
nobs = ntimes*nsub

##### create a vector containing a new variable for subject number #####
##### which will look like 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, etc. #####
##### for the five observations per person #####
##### the second command sorts the numbers into the correct order #####
tmp = rep(1:nsub,ntimes)
subj = tmp[order(tmp)]

##### create vector containing the gender and age #####
##### note that R and S-Plus ARE case sensitive #####
sex = rep(-99,length(subj))
age = rep(-99,length(subj))
for(j in 0:(nsub-1)){
  sex[(ntimes*j+1):(ntimes*(j+1))] = t(t(rep(data$Sex[j+1],ntimes)))
  age[(ntimes*j+1):(ntimes*(j+1))] = t(t(rep(data$Age[j+1],ntimes)))
}

##### create a vector containing the trial values #####
trial = rep(c(1,2,3,4,5),nsub)

##### stack each subject's data into one vector #####
time = rep(-99,nsub*4)
for(i in 0:(nsub-1)){
  time[ntimes*i+1] = data[i+1,3]
  time[ntimes*i+2] = data[i+1,4]
  time[ntimes*i+3] = data[i+1,5]
  time[ntimes*i+4] = data[i+1,6]
  time[ntimes*i+5] = data[i+1,7]
}

##### combine all the data columns into one matrix #####
repmat = cbind(time,sex,age,trial,subj)
```

S-Plus code: The R code will work in S-Plus as well, EXCEPT you want to make a numeric (not character) variable for `sex` before combining it with the other vectors into `repmat`.

```
female = ifelse(sex=='M',0,1)
repmat = cbind(time,female,age,trial,subj)
```