ALDA\_HW1\_EH.R

elizabethhawkey

Sat Sep 9 18:42:50 2017

#############################################  
#### Applied Longitudinal Data Analysis #####  
#### Multilevel modeling PDS dataset ########  
#### 9/5/2017; HW 1 Elizabeth Hawkey ########  
#############################################  
  
#load lme4 package: lme4 provides functions for fitting and analyzing mixed models: linear (lmer),   
#generalized linear (glmer) and nonlinear (nlmer.)  
library(lme4)

## Loading required package: Matrix

library(tidyverse)

## Loading tidyverse: ggplot2  
## Loading tidyverse: tibble  
## Loading tidyverse: tidyr  
## Loading tidyverse: readr  
## Loading tidyverse: purrr  
## Loading tidyverse: dplyr

## Conflicts with tidy packages ----------------------------------------------

## expand(): tidyr, Matrix  
## filter(): dplyr, stats  
## lag(): dplyr, stats

library(broom)  
library(dplyr)  
library(tidyr)  
  
#load in newest dataset   
setwd("~/Documents/PDS\_project/Longitudinal\_project")  
PDS\_data\_ALDA <- read.csv("~/Documents/PDS\_project/Longitudinal\_project/stats/PDS\_dataset\_stats\_HW1.csv")  
#View(PDS\_data\_ALDA)  
print(names(PDS\_data\_ALDA))

## [1] "row" "Subid" "ADHDsum\_1" "ADHDsum\_2" "ADHDsum\_3"  
## [6] "FPNGE15\_1" "FPNGE15\_2" "FPNGE15\_3" "DATE\_1" "DATE\_2"   
## [11] "DATE\_3" "age\_1" "age\_2" "age\_3"

#1 Move your data into a long format and a wide format.   
#Did you have any specific challenges that you encountered? If so, discuss them.  
##I had to rename all of my variables to use a common separator and had to change the   
##type of variable to date prior to importing.  
  
wide\_to\_long <- PDS\_data\_ALDA %>%  
 gather(ADHDsum\_1:age\_3, key = "time", value = "value")

## Warning: attributes are not identical across measure variables;  
## they will be dropped

#View(wide\_to\_long)  
  
wide\_to\_long2 <- wide\_to\_long %>%  
 separate(time, into = c("var", "scan", "omit"), sep = "\_", convert = TRUE) %>%  
 select(-omit) %>%  
 spread(key = var, value = value) %>%  
 arrange(Subid)

## Warning: Too few values at 2532 locations: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10,  
## 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...

#View(wide\_to\_long2)  
  
#this is another way to use separate using a specific character location  
#separate(time, into = c("var", "scan", "omit"), sep = c(8,9), convert = TRUE) %>%  
  
long\_to\_wide <- wide\_to\_long %>%  
 spread(time, value)  
  
  
#2 Create a wave variable and date variable (if applicable).  
##scan is the wave variable I will be using (created in the previous step),   
##date of assessment was included in my dataset.  
  
#3 What is your sample size for each wave of assessment?  
wide\_to\_long2$ADHDsum <- as.numeric(wide\_to\_long2$ADHDsum)  
wide\_to\_long2$FPNGE15 <- as.numeric(wide\_to\_long2$FPNGE15)  
wide\_to\_long2$age <- as.numeric(wide\_to\_long2$age)  
  
library(psych)

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

describeBy(wide\_to\_long2$ADHDsum, wide\_to\_long2$scan, mat=TRUE)

## item group1 vars n mean sd median trimmed mad min max  
## X11 1 1 1 207 2.743961 4.017818 0 1.9161677 0 0 17  
## X12 2 2 1 190 2.147368 3.518563 0 1.3552632 0 0 17  
## X13 3 3 1 157 1.452229 2.889960 0 0.7637795 0 0 17  
## range skew kurtosis se  
## X11 17 1.623347 2.002717 0.2792577  
## X12 17 2.098086 4.556833 0.2552634  
## X13 17 2.708806 8.466127 0.2306439

#scan1 = 207  
#scan2 = 190  
#scan3 = 157  
  
#4 Take the date variable and convert it to a different date format such as time in study   
#or age (if appropriate). What scale is most suitable for your analyses? (weeks/months/years?)  
#convert imported character date variables to date format  
DATE\_1 <- as.Date(long\_to\_wide$DATE\_1, format = "%m/%d/%y")  
DATE\_2 <- as.Date(long\_to\_wide$DATE\_2, format = "%m/%d/%y")  
DATE\_3 <- as.Date(long\_to\_wide$DATE\_3, format = "%m/%d/%y")  
##if captial Y imports the year as: "0008"; for "2008" use lowercase y for current century  
#strptime() for different classes  
  
#difference in weeks between each assessment  
datediff\_1 <- formatC(difftime(DATE\_1, DATE\_1, units = "weeks")/100)  
datediff\_2 <- formatC(difftime(DATE\_2, DATE\_1, units = "weeks")/100)  
datediff\_3 <- formatC(difftime(DATE\_3, DATE\_1, units = "weeks")/100)  
  
#add time variable back to wide data frame  
long\_to\_wide$datediff\_1 <- as.numeric(datediff\_1)  
long\_to\_wide$datediff\_2 <- as.numeric(datediff\_2)

## Warning: NAs introduced by coercion

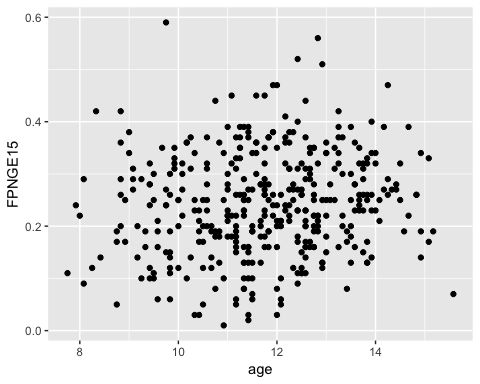
long\_to\_wide$datediff\_3 <- as.numeric(datediff\_3)

## Warning: NAs introduced by coercion

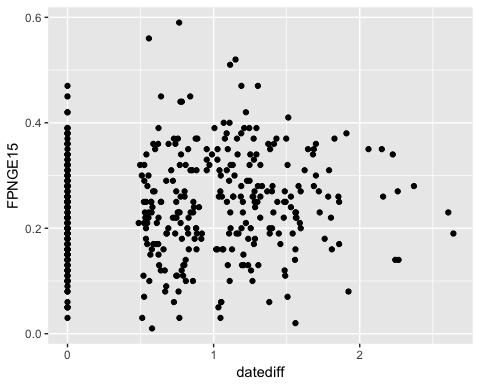
#View(long\_to\_wide)  
  
#return to long format for plotting  
plot\_long <- long\_to\_wide %>%  
 gather(ADHDsum\_1:datediff\_3, key = "time", value = "value")   
#View(plot\_long)  
  
plot\_long2 <- plot\_long %>%  
 separate(time, into = c("var", "scan", "omit"), sep = "\_", convert = TRUE) %>%  
 select(-omit) %>%  
 spread(key = var, value = value) %>%  
 arrange(Subid)

## Warning: Too few values at 3165 locations: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10,  
## 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...

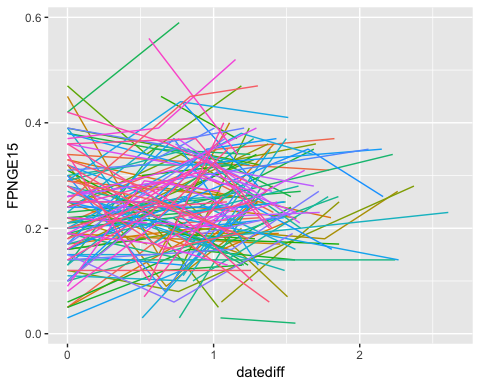
#View(plot\_long2)  
  
#4 Graph your data using the different time metrics, fitting individual curves for each person.  
library(ggplot2)  
#subset of participants to view graphs  
plot\_long3 <- subset(plot\_long2, FPNGE15 > 0)  
plot\_long3$datediff <- as.numeric(plot\_long3$datediff)  
plot\_long3$age <- as.numeric(plot\_long3$age)  
plot\_long3$FPNGE15 <- as.numeric(plot\_long3$FPNGE15)  
#View(plot\_long3)  
  
gg1 <- ggplot(data = plot\_long3,   
 aes(x = age, y = FPNGE15, group = Subid)) + geom\_point()  
gg1



gg1x <- ggplot(data = plot\_long3,   
 aes(x = datediff, y = FPNGE15, group = Subid)) + geom\_point()  
gg1x

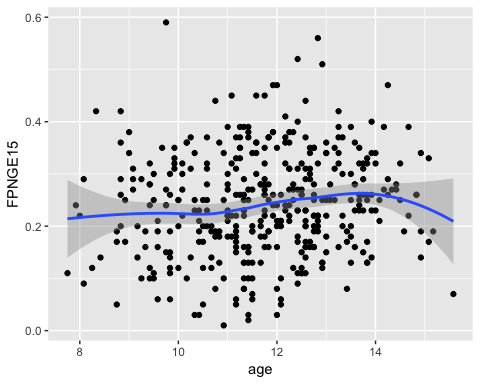


gg2 <- ggplot(data = plot\_long3,  
 aes(x = datediff, y = FPNGE15, group = Subid)) + geom\_line() +  
 aes(colour = factor(Subid)) + guides(colour=FALSE)   
gg2

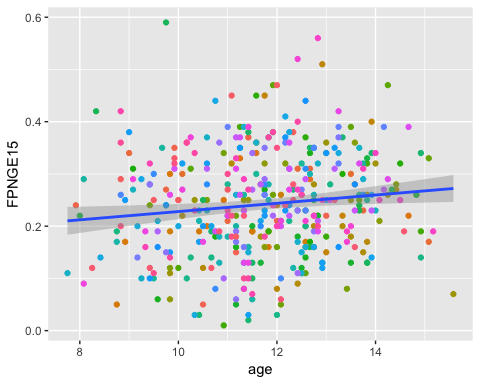


#5 Create an overall average trend of your data (split up into groups if appropriate).   
#Attempt to color your individual data points and/or shade different lines (highlight some   
#particiapnts,highlight the average trend line but not the individual level lines)  
gg3 <- ggplot(plot\_long3, aes(x = age, y = FPNGE15)) + geom\_point() + stat\_smooth()  
gg3

## `geom\_smooth()` using method = 'loess'



gg4 <- ggplot(plot\_long3, aes(x = age, y = FPNGE15)) + geom\_point() + stat\_smooth(aes(colour = Subid), method = "lm", se = TRUE) +  
 aes(colour = factor(Subid)) + guides(colour=FALSE)   
gg4



#6 Look at the correlations of your DV across time  
cor.dv <- subset(PDS\_data\_ALDA, select = c(FPNGE15\_1, FPNGE15\_2, FPNGE15\_3))  
cor(cor.dv, use = "pairwise.complete.obs")

## FPNGE15\_1 FPNGE15\_2 FPNGE15\_3  
## FPNGE15\_1 1.0000000 0.3249068 0.3712788  
## FPNGE15\_2 0.3249068 1.0000000 0.3340950  
## FPNGE15\_3 0.3712788 0.3340950 1.0000000