# Install & Load required packages   
if(!require('lme4')){install.packages('lme4')}

## Loading required package: lme4

## Loading required package: Matrix

library(lme4)  
load("D:/Clouds/Dropbox/Dropbox/ARP - statistics in R/R/AssignmentWorkspace.rdata") # Change directory according to your device  
  
# These assignments are about a study on the popularity of highschool students. A total of 515 students from 25 different   
# classes took part in this study in which we determined how extraversion, gender, and teacher experience influenced a   
# student's popularity. The data of this study show a clear hierarchy - or nesting - , with students nested in classes.   
# Since observations of students from the same class are most likely not independent, we can't just analyse these data using   
# normal lineair regression. In addition, we are dealing with variables on different levels. We have pupil level variables like  
# extraversion and gender, but we also have class characteristics like teacher experience. Only with multilevel analysis can we   
# estimate both types of variables in a single analysis. In this assignment we'll start with some regular regression analyses  
# on the data of each separate class (even though the independence assumption is violated!). Because this will help you get the  
# hang of R, but also because it will make the multilevel analyses more insightful. Then, given the hierarchical nature of the   
# data, we run a multilevel regression analyis on the complete data.  
  
# Assignment 1:  
# The data for the 25 different classes are stored in de the dataframes Class1 - Class25. The dataframe Total   
# contains the combined data of all classes. Inspect the dataframes (maybe make a plot using the plot command. For help, simply  
# type ?plot).  
 head(Class1)

## Pupil Class Extraversion Gender teacherExp Popular  
## 1 1 1 5 1 24 6.3  
## 2 2 1 7 0 24 4.9  
## 3 3 1 4 1 24 5.3  
## 4 4 1 3 1 24 4.7  
## 5 5 1 5 1 24 6.0  
## 6 6 1 4 0 24 4.7

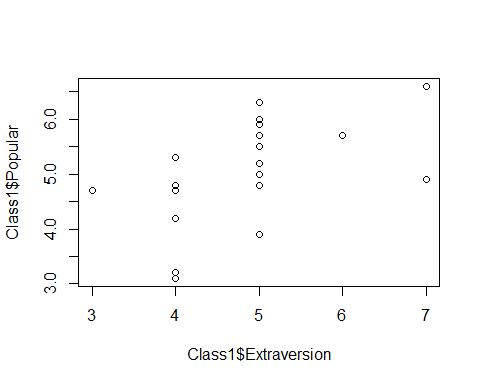
head(Total)

## Pupil Class Extraversion Gender teacherExp Popular  
## 1 1 1 5 1 24 6.3  
## 2 2 1 7 0 24 4.9  
## 3 3 1 4 1 24 5.3  
## 4 4 1 3 1 24 4.7  
## 5 5 1 5 1 24 6.0  
## 6 6 1 4 0 24 4.7

tail(Total)

## Pupil Class Extraversion Gender teacherExp Popular  
## 510 15 25 5 1 19 3.3  
## 511 16 25 4 0 19 4.0  
## 512 17 25 5 0 19 2.9  
## 513 18 25 5 0 19 3.8  
## 514 19 25 5 0 19 3.7  
## 515 20 25 5 0 19 4.3

plot(Class1$Extraversion, Class1$Popular)



#List of all the variables:  
# - pupil:pupil identification variable, not needed in the analysis  
# - class:class identification variable, the linking variable to define the 2 - level structure  
# - student-level independent variables: extraversion (continuous; higher scores mean higher extraversion) and gender (dichotomous; 0=male, 1 =female)  
# - class-level independent variables: teacher experience (in years)  
# - outcome variable: popular (continuous outcome variable at the student-level, higher scores indicate higher popularity)  
  
# Assignment 2:  
# Run regression analyses for each of the 25 classes separately using the lm command, and save the intercepts and regression   
# coefficients of each of these analyses. In these analyses, use popularity as the dependent variable and extraversion and   
# gender as independent variables. If needed as for instructions on using the lm command by typing ?lm in R.   
  
 # The code for each separate analysis is given by,  
 ResultsClass1 <- with(data = Class1,  
 lm(Popular ~ 1 + Extraversion + Gender)  
 )  
 # where you obvioulsy change the name of the class you're interested in.  
 # The coefficients are stored in a separate part of the ResultsClass1 variable called "Coefficients". To acces them   
 # separately ask for ResultsClass1$coefficients,  
 ResultsClass1$coefficients

## (Intercept) Extraversion Gender   
## 2.5502762 0.4127072 1.0461878

# The intercept is the first value sotred under coefficients and can be asked for by typing,  
 ResultsClass1$coefficients[1]

## (Intercept)   
## 2.550276

# Similarly, the regression coefficients for Extraversion and Gender can be obtained by typing  
 ResultsClass1$coefficients[2] # and,

## Extraversion   
## 0.4127072

ResultsClass1$coefficients[3]

## Gender   
## 1.046188

# The above can be done for each class separately, or you can save all intercepts and regression coefficients using a  
 # for-loop.  
 Coefficients <- matrix(NA, ncol = 3, nrow = 25) # First we create a matrix in which we store the inetercepts and coefficients  
 colnames(Coefficients) <- c("Intercept", "Extraversion", "Gender") # And then we specify the columnnames  
 # Then we automatically run lm on all the 25 classes and store the results  
 for (i in 1:25) {  
 Coefficients[i,] <- lm(Popular ~ 1+Extraversion+Gender, data = eval(parse(text = paste("Class", i, sep = ""))))$coefficients  
 }  
  
   
# Assignment 3:  
# Why can't we include teacher experience as a predictor in the separate regression analyses above?  
# "Because the scores on teacher experience don't vary within classes"  
  
# Assignment 4:  
# For each class, save the mean popularity score and the teacher experience score.  
# Run a regression in which you predict the mean popularity usinf teacher experience and save the intercept and regression  
# coefficient. What is the sample size in this analysis?  
  
 # Getting the means can be done for each dataset by typing,  
 mean(Class1$Popular) # and,

## [1] 5.075

mean(Class1$teacherExp)

## [1] 24

# and then saving the values in two separate vectors  
 # Alternatively we can again do this automatically with  
 Mean <- matrix(NA, ncol = 2, nrow = 25) # First we create a matrix in which we store the inetercepts and coefficients  
 colnames(Mean) <- c("Popular", "TeacherExperiece") # And then we specify the columnnames  
 # Then we automatically run lm on all the 25 classes and store the results  
 for (i in 1:25) {  
 Mean[i, 1] <- mean(eval(parse(text = paste("Class", i, sep = "")))$Popular)  
 Mean[i, 2] <- mean(eval(parse(text = paste("Class", i, sep = "")))$teacherExp)  
 }  
 # The regression analysis can subsequently be run using  
 lm(Mean[,1]~1+Mean[,2])

##   
## Call:  
## lm(formula = Mean[, 1] ~ 1 + Mean[, 2])  
##   
## Coefficients:  
## (Intercept) Mean[, 2]   
## 4.21554 0.05206

# Assignment 5:  
# Run a multilevel regression analysis on the data of all classes simulataneously (use the "Total" dataframe for this) using the   
# lme command. Start with the intercept only model in which you use Popular as the dependent variable, and calculate the ICC.   
# If needed ask for instruction using ?lme  
  
 lmer(Popular ~ 1 + (1 | Class), Total)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: Popular ~ 1 + (1 | Class)  
## Data: Total  
## REML criterion at convergence: 1581.243  
## Random effects:  
## Groups Name Std.Dev.  
## Class (Intercept) 0.6514   
## Residual 1.0647   
## Number of obs: 515, groups: Class, 25  
## Fixed Effects:  
## (Intercept)   
## 4.928

# The ICC is equal to .6514/(1.0647+.6514) = .380  
  
# Assignment 5b:  
# Now add the first level variables extraversion and gender to the model as fixed effects.  
  
 lmer(Popular ~ 1 + Gender + Extraversion + (1 | Class), Total)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: Popular ~ 1 + Gender + Extraversion + (1 | Class)  
## Data: Total  
## REML criterion at convergence: 1235.949  
## Random effects:  
## Groups Name Std.Dev.  
## Class (Intercept) 0.7262   
## Residual 0.7409   
## Number of obs: 515, groups: Class, 25  
## Fixed Effects:  
## (Intercept) Gender Extraversion   
## 2.033 1.146 0.450

# Assignment 5c:  
# Now add the second level variable teacher experience to the model.  
  
 lmer(Popular ~ 1 + Gender + Extraversion + teacherExp + (1 | Class), Total)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: Popular ~ 1 + Gender + Extraversion + teacherExp + (1 | Class)  
## Data: Total  
## REML criterion at convergence: 1232.748  
## Random effects:  
## Groups Name Std.Dev.  
## Class (Intercept) 0.6111   
## Residual 0.7408   
## Number of obs: 515, groups: Class, 25  
## Fixed Effects:  
## (Intercept) Gender Extraversion teacherExp   
## 0.98756 1.14285 0.45538 0.07466

# Assignment 5d:  
# Now check if the relation between the first level predictors and popularity is the same across classes.  
# What type of effect do you need to add for test this hypotheses?  
# Compare the intercept and regression coefficients of gender and extraversion from the multilevel analysis to the   
# average estimates across the 25 separate analyses. Are they the same? Why/Why not?  
# What about the multilevel regression coefficient for teacherExp? Is it the same as in the regression analysis you ran on   
# the mean popularity and mean teacherExp scores? Why/Why not?  
  
 lmer(Popular ~ 1 + Gender + Extraversion + teacherExp + (1 + Gender + Extraversion | Class), Total)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: Popular ~ 1 + Gender + Extraversion + teacherExp + (1 + Gender +   
## Extraversion | Class)  
## Data: Total  
## REML criterion at convergence: 1225.292  
## Random effects:  
## Groups Name Std.Dev. Corr   
## Class (Intercept) 1.1531   
## Gender 0.2420 -0.54   
## Extraversion 0.1288 -0.83 0.28  
## Residual 0.7191   
## Number of obs: 515, groups: Class, 25  
## Fixed Effects:  
## (Intercept) Gender Extraversion teacherExp   
## 1.22683 1.17164 0.45629 0.05356

# You need to add random slopes  
  
 # You can calculate the mean intercept and regression coefficients by hand,but if you stored them in a matrix you can  
 # use the following code  
 apply(Coefficients, 2, mean, na.rm = TRUE)

## Intercept Extraversion Gender   
## 1.9120545 0.4491835 1.2021451

# You see that the mean estimates across the separate analyses are not identical to the multilevel estimates.  
 # This is because multilevel analysis assumes that the inter-class differences in the intercept and coefficients are  
 # normally distributed whereas the estimates from the separate analyses don't impose a distribution on the individual  
 # estimates. As a result the multilevel estimates are "pulled" towards the mean parameter value across classes.  
  
 # No, the estimate for the effect of teacher Exp is not exactly the same. This is because the classes are not all the same size. Some have more pupils  
 # than others. The multilevel estimates are weighted for these class-size differences while the regresison analysis on the  
 # mean scores was not.  
  
# Assignment 5e:  
# Finally check if you can explain the variance in your random slope(s) with the second level predictor teacherExp  
  
 lmer(Popular ~ 1 + Gender + Extraversion + teacherExp + Gender \* teacherExp + Extraversion \* teacherExp +  
 (1 + Gender + Extraversion | Class), Total)

## Linear mixed model fit by REML ['lmerMod']  
## Formula:   
## Popular ~ 1 + Gender + Extraversion + teacherExp + Gender \* teacherExp +   
## Extraversion \* teacherExp + (1 + Gender + Extraversion | Class)  
## Data: Total  
## REML criterion at convergence: 1224.809  
## Random effects:  
## Groups Name Std.Dev. Corr   
## Class (Intercept) 0.87316   
## Gender 0.25719 -0.51   
## Extraversion 0.04215 -0.83 -0.06  
## Residual 0.71745   
## Number of obs: 515, groups: Class, 25  
## Fixed Effects:  
## (Intercept) Gender Extraversion   
## -1.04216 1.46446 0.79302   
## teacherExp Gender:teacherExp Extraversion:teacherExp   
## 0.21780 -0.02142 -0.02459

NA