___2022-midterm_notesR.R

georges

2022-03-06

```
#
# 4939: 2022 midterm
#
```

```
library(spida2)
```

Using the 'hs' data set in 'spida2'

Write a function in R that takes a numerical vector as an input and returns a vector that is the difference between each value and the median of the values in the vector.

Write a function in R that takes two inputs, a numerical vector and a second vector whose values are ids for subjects. The functions return 1 if the first vector is a level 1 variable with respect to the ids, and 2 if it's a level 2 variable.

Mixed model theory: In a normal linear mixed model to fit data with two levels, a response variable Y and a single level-1 predictor X, k clusters of size $n_i, i = 1, ..., k$.
Suppose you use the R command `lme(Y ~ 1 + X, data, random = ~ 1 + X | id)`. Using the notation used in class for such a model:

- Derive $\text{Var}(\hat{\beta}_i)$ and $\text{Var}(\hat{\beta}_i - \beta_i)$ where β_i is the 'true' vector of coefficients in the ith cluster and $\hat{\beta}_i$ is the BLUE for β_i based on the data in cluster i.
- Discuss which variance is relevant if one is using $\hat{\beta}_i$ to make inferences about cluster i or to make inferences about the population from which cluster i is viewed as a sample.

In a normal linear mixed model to fit data with two levels, a response variable

Y and a single level-1 predictor X, k clusters of size $n_i, i = 1, ..., k$.
Suppose you use the R command `lme(Y ~ 1 + X, data, random = ~ 1 + X | id)`. Using the notation used in class for hierarchical models:

- Derive $\text{Var}(Y_i)$ and $\text{Var}(\bar{Y}_i)$ where Y_i is the vector of observations in the ith cluster.

Using the 'hs' data set in 'spida2', fit and discuss an appropriate model to explore whether the relationship between 'mathach' and 'ses' differs in the two sectors, 'Catholic' versus 'Public'.

Suppose you are analyzing the 'hs' data we have considered in class.
Reminder:

```
car::some(hs, 6)
```

```
##      school mathach     ses    Sex Minority Size    Sector PRACAD DISCLIM
## 179    2458  21.451   1.082 Female       No  545  Catholic   0.89  -1.484
## 195    2458   9.490   0.772 Female       No  545  Catholic   0.89  -1.484
## 415    2771   6.324   0.132 Female       No  415    Public   0.24   1.048
## 463    3013  20.633   0.222   Male       No  760    Public   0.56  -0.213
## 526    3610  21.034   1.012   Male       No 1431  Catholic   0.80  -0.621
## 938    5640  16.509  -0.118 Female       No 1152    Public   0.41   0.256
```

You want to study how the relationship between 'mathach' and 'ses' differs between the two Sectors: Public and Catholic. Comment on the strengths and pitfalls of using the following models to do so:

1. `lm(mathach ~ Sector * ses)`
2. `lm(mathach ~ Sector * ses + school)`
3. `lm(mathach ~ Sector + ses)`
4. `lme(mathach ~ Sector * ses, random = ~ 1 + ses | school)`
5. `lme(mathach ~ Sector * (ses + cvar(ses, school)), random = ~ 1 + ses | school)`
6. `lme(mathach ~ Sector * (ses + cvar(ses, school)), random = ~ 1 + cvar(ses, school) | school)`

Write a function in R that takes a numeric vector, y, as an argument and returns a logical vector that has the value TRUE if the corresponding element of y is more than 3 standard deviations of y away from the mean of y.