HW12

Animesh Sengupta

2022-12-07

Question 2

I have completed the tutorials

Question 6

The two definitions that I chose was the following :

(1) Fixed effects are constant across individuals, and random effects vary. For example, in a growth study, a model with random intercepts a_i and fixed slope b corresponds to parallel lines for different individuals i, or the model y_it = a_i + b t. Kreft and De Leeuw (1998) thus distinguish between fixed and random coefficients.

(2) Effects are fixed if they are interesting in themselves or random if there is interest in the underlying population. Searle, Casella, and McCulloch (1992, Section 1.4) explore this distinction in depth.

One of the major differences between the two definitions is how we intrinsically look at the random/fixed effects. The former definition relies more on the characteristics of the data while the latter definitions is more concerned about the way the problem statement/objective of analysis shaped. So accordingly , the first definition talks and proposes that the predictors effect is constant throughout the analysis cycle i.e either fixed or random. The second definition offers a bit more variability in the behaviour of predictors , i.e it can either be fixed/random solely based on the goal/objective of analysis.

Question 3

3a

```
library("ggplot2")
library("GGally")
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library("readr")

setwd("/Users/animeshsengupta/Work_Directory/DACSS/STAT625/Homeworks")
data<-read_csv("/Users/animeshsengupta/Work_Directory/DACSS/STAT625/Homeworks/HSB_data.csv",skip=0)
```

```
## Rows: 7185 Columns: 6

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (6): id, ses, mathach, size, sector, meanses
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

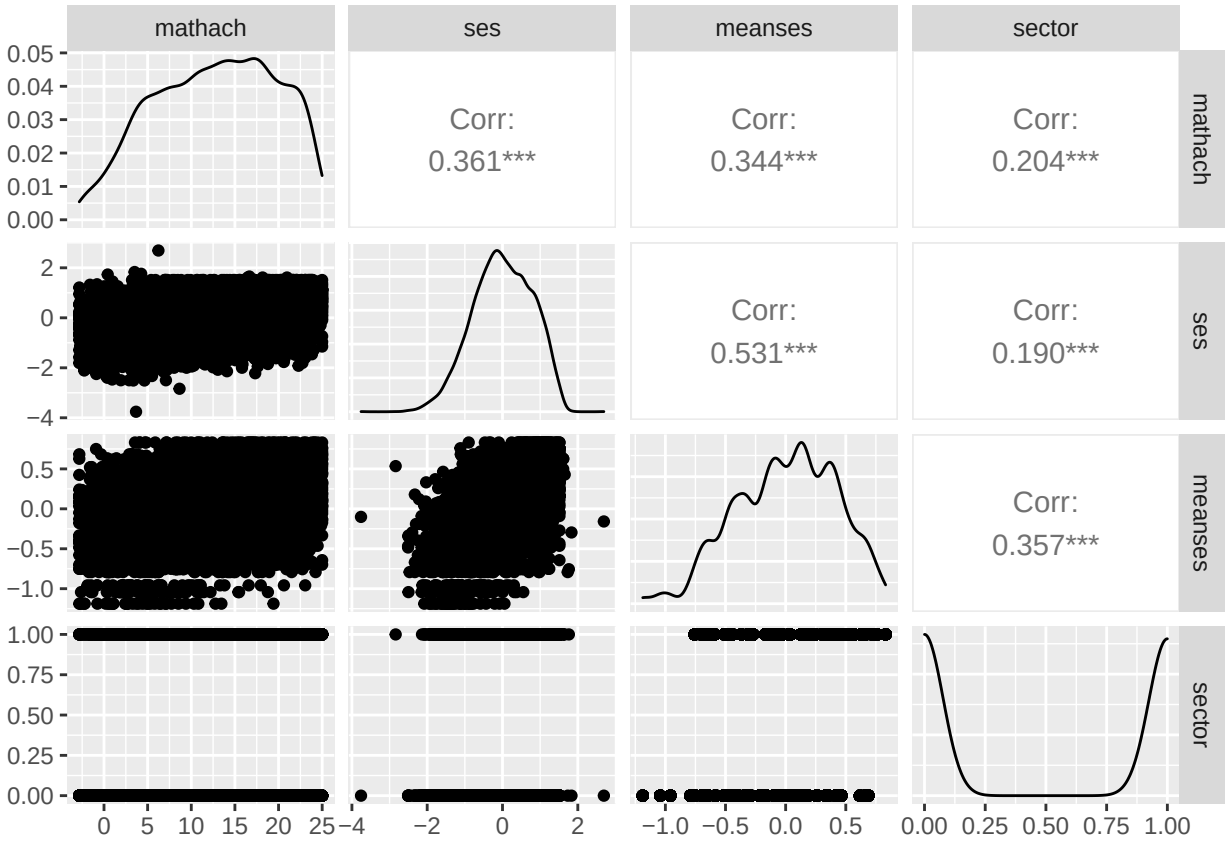
```
colnames(data)
```

```
## [1] "id"      "ses"     "mathach" "size"    "sector"  "meanses"
```

```
head(data)
```

```
## # A tibble: 6 x 6
##      id    ses mathach  size sector meanses
##   <dbl>  <dbl>   <dbl> <dbl>  <dbl>   <dbl>
## 1  1224 -1.53     5.88   842      0  -0.428
## 2  1224 -0.588   19.7    842      0  -0.428
## 3  1224 -0.528   20.3    842      0  -0.428
## 4  1224 -0.668    8.78   842      0  -0.428
## 5  1224 -0.158   17.9    842      0  -0.428
## 6  1224  0.022    4.58   842      0  -0.428
```

```
ggpairs(columns=c("mathach","ses","meanses","sector"),data=data)
```



3.b

as evident from the pairwise plot, we have some outliers in both ses and messures. These outliers are particulalry visible because majority of datapoints are concentrated at a particular locality.Hence it would be important to analyse the effect of these outliers. Due to extraordinary large datapoints , the relationship between ses and measures against mathach would have to be carefully examined and we need to identify inderlying groups and take into account the random effects based on those groups. Probably sector would be one of those grouping variables. Sector looks more like a categorical data with 2 values, although it can be used for analysing the random effect, but according to the tutorial it was mentioned that it is a good practice for the grouping variable to have more than 3 categories. Lastly, all the predictors seemed significantly important and we should try to take them into account for further regrerssion analtysis.

question 4

4a

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
hsblmer<-lmer(mathach~(1|sector)+meanses+ses,data=data)
summary(hsblmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: mathach ~ (1 | sector) + meanses + ses
##    Data: data
##
## REML criterion at convergence: 46773.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3066 -0.7491  0.0137  0.7741  2.7630
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  sector   (Intercept)  0.8072  0.8984
##  Residual             39.2846  6.2677
## Number of obs: 7185, groups:  sector, 2
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  12.7379     0.6396   19.92
## meanses       2.9811     0.2216   13.45
## ses           2.1912     0.1119   19.57
##
## Correlation of Fixed Effects:
##         (Intr) meanss
## meanses -0.003
## ses      0.001 -0.505
```

4b

```
hsblmer2<-lmer(mathach~(1+meanses:ses|sector)+meanses+ses,data=data)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
summary(hsblmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: mathach ~ (1 | sector) + meanses + ses
##    Data: data
##
## REML criterion at convergence: 46773.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3066 -0.7491  0.0137  0.7741  2.7630
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  sector   (Intercept)  0.8072  0.8984
##  Residual             39.2846  6.2677
## Number of obs: 7185, groups:  sector, 2
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  12.7379     0.6396   19.92
## meanses       2.9811     0.2216   13.45
## ses           2.1912     0.1119   19.57
##
## Correlation of Fixed Effects:
##         (Intr) meanss
## meanses -0.003
## ses      0.001 -0.505
```