Part 3 - Multilevel model

Code

Importing Libraries and Reading in the data.

Hide

```
library(car)
```

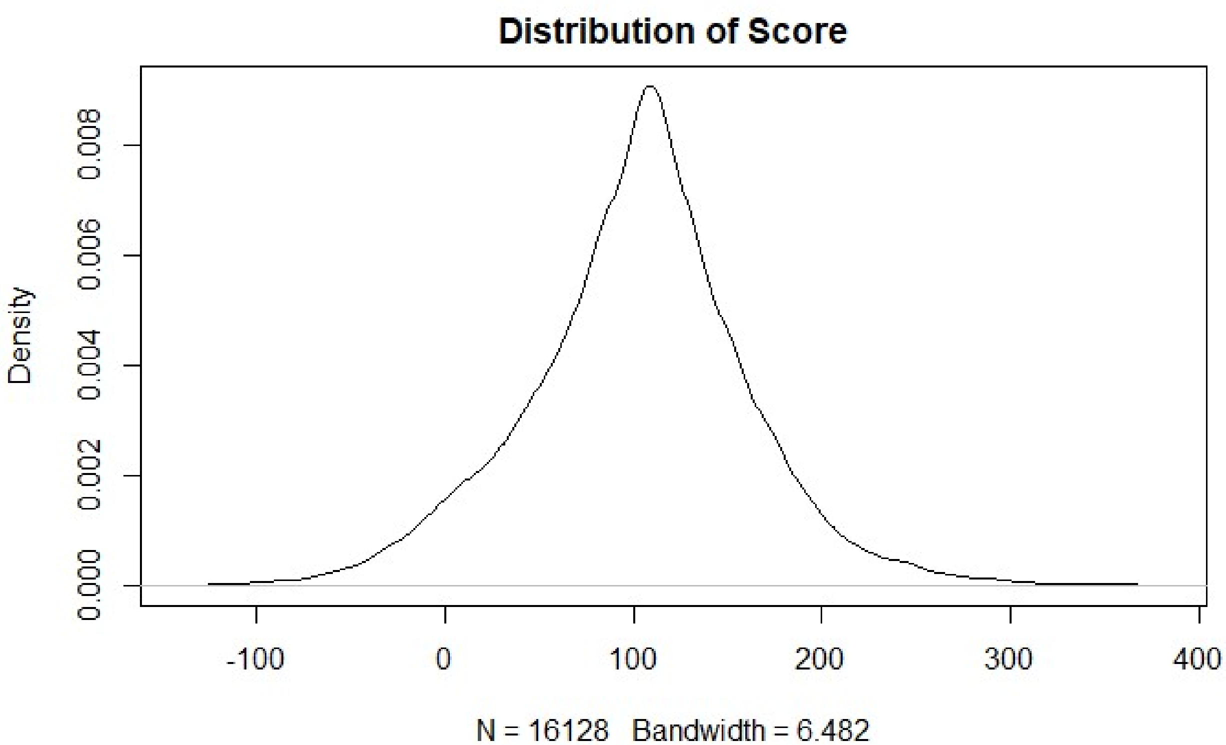
```
Loading required package: carData
Registered S3 method overwritten by 'data.table':
  method           from
  print.data.table
```

From the summary, we see we have 500 subjects. For these 500 subjects, we have data about their scores for different sessions. We see on average the mean score is 102.3

Visual inspection

Hide

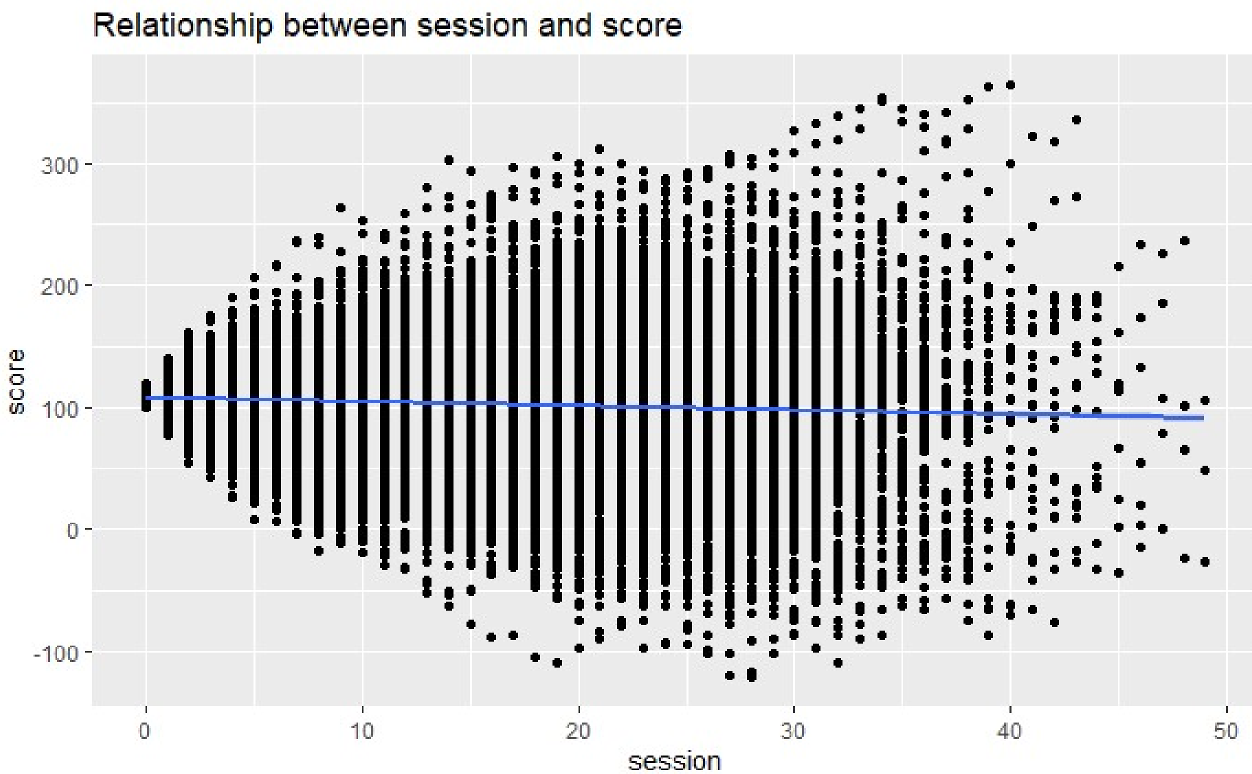
```
plot(density(mydata$score),main="Distribution of Score")
```



From the above figure, it seems as though the scores in our data follows a double exponential distribution.

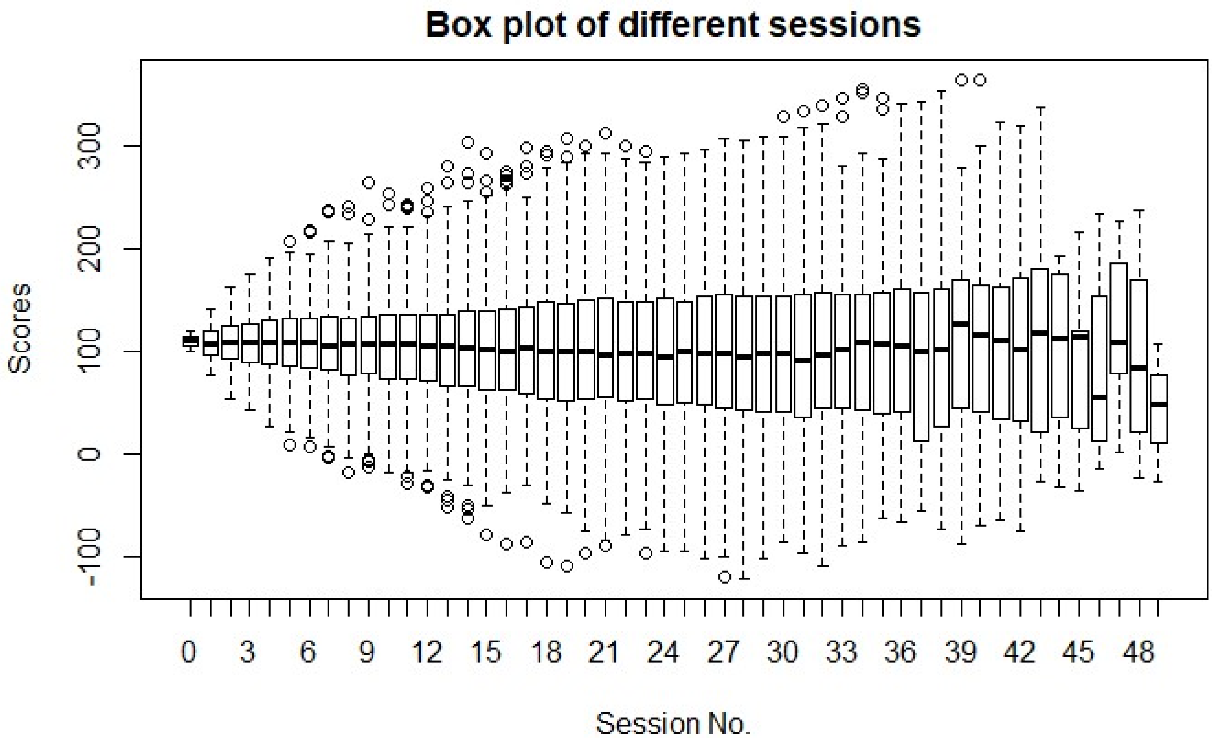
Hide

```
scatter <- ggplot(mydata, aes(session, score))
scatter+geom_point() + geom_smooth(method="lm") + ggtitle("Relationship between sessio
n and score")
```



Hide

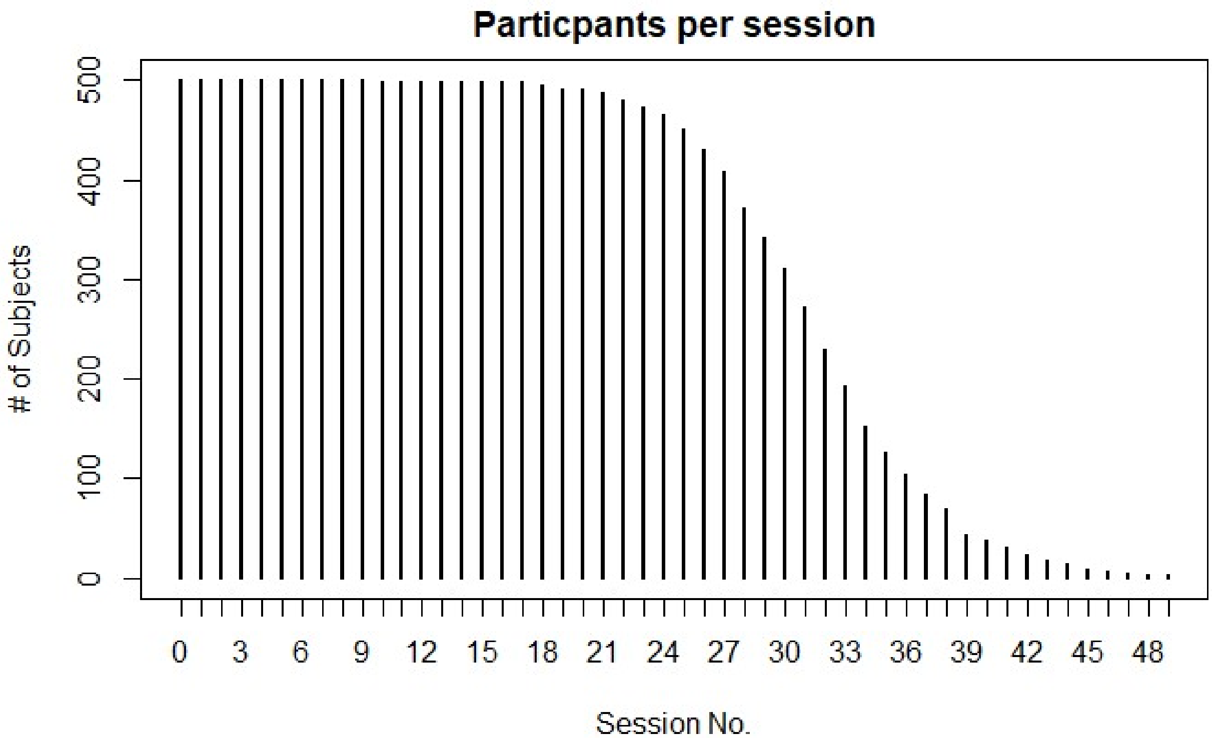
```
boxplot(score~session,data=mydata, main="Box plot of different sessions",
        xlab="Session No.", ylab="Scores")
```



From the above scatterplot, it can be seen that scores go down very slightly for the later sessions.From the box plot we see that the spread of our scores for different sessions is not homogenous.It increases with the session number and finally decreases for the final few sessions.

Hide

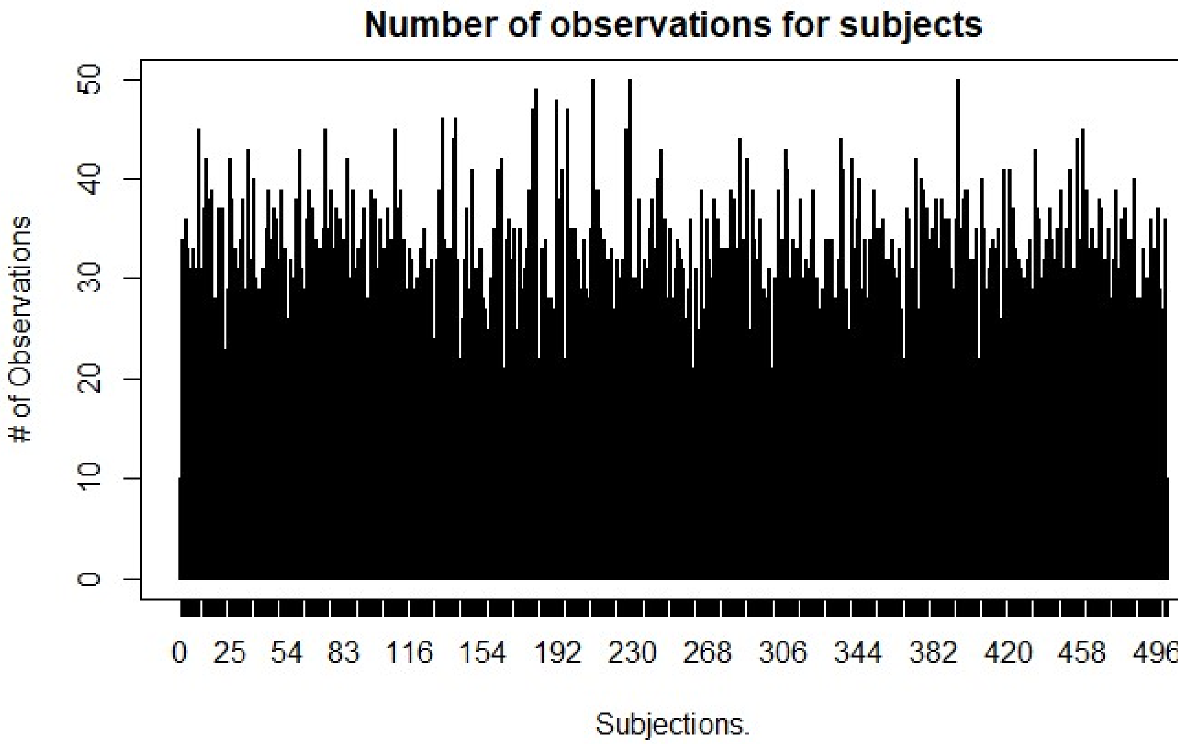
```
plot(table(mydata$session),ylab="# of Subjects",xlab="Session No.", main="Particpants per session")
```



Here we can see that the number of participants start going down after the 20th session i.e not all sessions have the same number of participants.

Hide

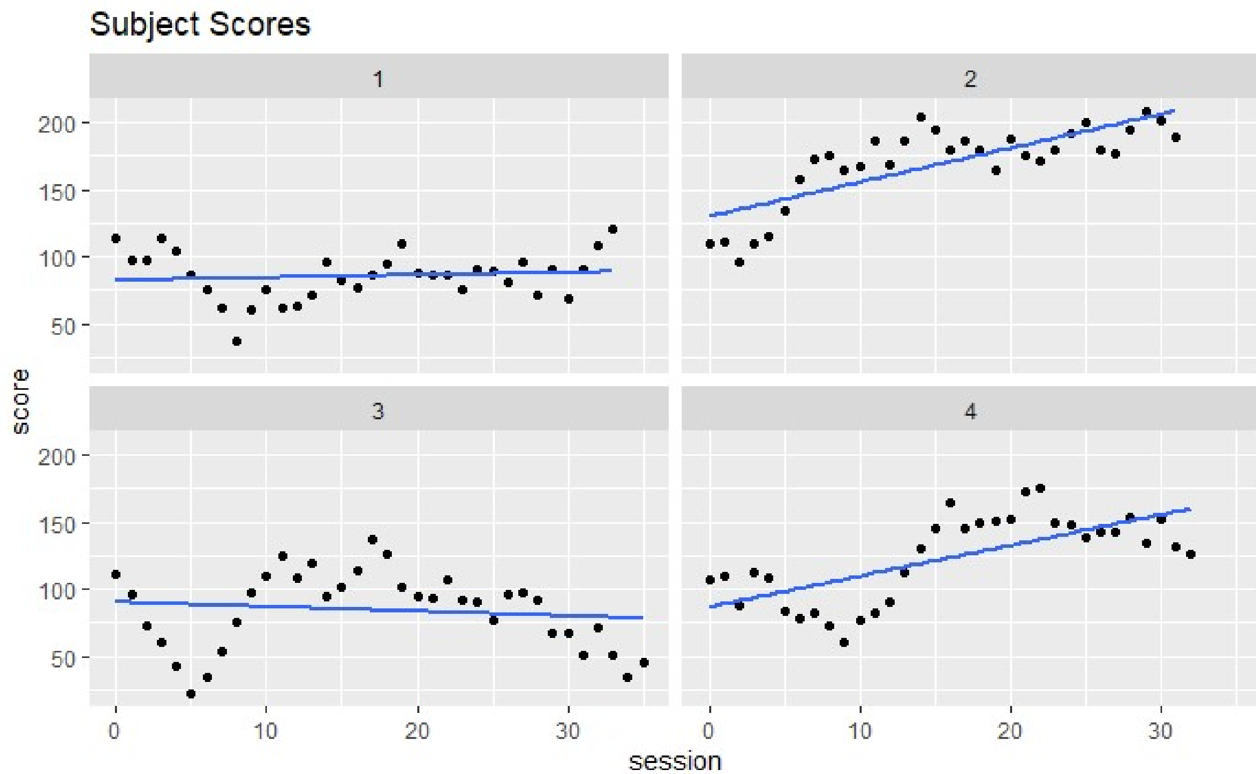
```
plot(table(mydata$Subject),ylab="# of Observations",xlab="Subjections.", main="Number of observations for subjects")
```



From this we can see that unlike other studies, in longitudinal studies we don't assume the same number of observations for each subject.

Hide

```
subdata <- subset(mydata, Subject>0&Subject < 5)
scatter <- ggplot(subdata, aes(session, score)) + labs(title="Subject Scores")
scatter +geom_point() + geom_smooth(method="lm", se= F)+
  facet_wrap(~Subject,5,2)
```



From the above scatter plots for 4 different subjects, we see that the subject's scores are quite varied with each session. For subjects 2 and 4, we see an upward trend in the scores, for subject 3 we see it's slightly downward. And lastly, for subject 1 we see that the slope is very close to 0. Therefore we can already see that the fitted lines have different slopes.

Multilevel analysis

Model Comparisons

Hide

```
randomInterceptOnly <- lme(score ~ 1, data = mydata,
                           random = ~1|Subject, method = "ML")

randomInterceptSession <- lme(score ~ session, data = mydata,
                              random = ~1|Subject, method = "ML")

anova(randomInterceptOnly,randomInterceptSession)
```

	Mo… <int>	df <dbl>	AIC <chr>	BIC <chr>	logLik <chr>	Test <fctr>	L.Ratio <chr>
randomInterceptOnly	1	3	162676.1	162699.1	-81335.04		
randomInterceptSession	2	4	162453.4	162484.1	-81222.70	1 vs 2	224.6805

2 rows

Here we have compared our random intercept(random = ~1|Subject) model with no fixed effects(score ~ 1) with respect to the random intercept(random = ~1|Subject) model where Session is also included as a fixed effect(score ~ session). We see from the comparison of our models using anova that including session(p.value <.001) as a fixed effect improves the fit our model. There we conclude that session does indeed impact people's scores.

Summary of the best model

Hide

```
summary(randomInterceptSession)
```

```
Linear mixed-effects model fit by maximum likelihood
 Data: mydata
```

	AIC <dbl>	BIC <dbl>	logLik <dbl>
	162453.4	162484.1	-81222.7

1 row

```


Random effects:
 Formula: ~1 | Subject
        (Intercept) Residual
StdDev:      46.443 34.96826

Fixed effects: score ~ session
```

	Value <chr>	Std.Error <chr>	DF <chr>	t-value <chr>	p-value <chr>
(Intercept)	108.66574	2.1398997	15626	50.78076	0
session	-0.42533	0.0282743	15626	-15.04304	0

2 rows

```
 Correlation:
        (Intr)
session -0.206

Standardized Within-Group Residuals:
         Min           Q1          Med           Q3          Max
-4.114815156 -0.647994395  0.007899928  0.638213193  4.160353413

Number of Observations: 16128
Number of Groups: 501
```

Based on the results shown above, we can see that there is a standard deviation of 46.4 between the intercepts that have been fitted on the scores for different subjects and standard deviation of 34.9 for the residuals. We also see that there is a negative coefficient for the session variable suggesting that scores go down as session number increases. Indeed based on our visual plot of the relationship between session and score we did see a slightly downward slope as well. Lastly, the value of the fixed intercept in this case is 108.66.

Hide

```
intervals(randomInterceptSession, 0.95)
```

```
Approximate 95% confidence intervals

 Fixed effects:
                  lower        est.       upper
(Intercept) 104.4715439 108.6657350 112.8599262
session      -0.4807492  -0.4253317  -0.3699142
attr(,"label")
[1] "Fixed effects:"

 Random Effects:
  Level: Subject
                   lower   est.   upper
sd((Intercept)) 43.60687 46.443 49.4636

 Within-group standard error:
   lower     est.    upper
34.58273 34.96826 35.35809
```

Based on the 95% confidence intervals, we see that indeed 95% percent of the time, the standard deviation between the intercepts for the different subjects is between 43.6 and 49.46. This doesn't include 0 so we can confidently say that there is a significant deviation between the different subjects. We also observed this in our plots for the different subject scores.

Conclusion.

We believe that different subjects do infact have a significant variance between the scores. This is because we see a significant deviation between the different intercepts fitted on our scores for the different subjects. We also saw that the session variable does have a significant impact the scores in our data.

Interclass correlation And R square at level 1.

With the interclass correlation, we can examine which proportion of the total variability in the scores in our data is attributable to the difference between the subjects.

Hide

```
cat("The ICC is",ICC)
```

```
The ICC is 0.6382023
```

We see that the subjects account for roughly 64 percent of the total variation in the scores of our data.

Hide

```
cat("The R square value at level 1 is",R_sq)
```

```
The R square value at level 1 is 0.003619935
```

We see that explained variance at the level 1 by including session has a fixed effect in model 1 as compared to model 0 is not very high.

Report section for a scientific publication

Our goal in this analysis was to see if session had a significant impact on the scores of our model and whether there was significant variation in the scores for different subjects. To see if session had a significant impact, we performed a multi-level analysis and compared the null model where we took subjects as the random intercept and a fixed intercept with an extended version of this model which also included session as a fixed effect. We saw that includcing session (df=4, AIC = 162453.4, Loglik=81222.70, L.Ratio= 224.6805, P.value<.0001) does indeed have a significant impact on the scores of our model. Furthermore, we found the estimate of the session variable to be negative with a value of -0.4253317 (95% CI, -0.4807492 to -0.3699142). In addition, we found that there was a significant variation among the scores for different subjects with a standard deviation of 46.443 (95%CI, 43.60687 to 49.4636) between the fitted intercepts for different subjects and similarly a standard deviation of 34.96826 (95% CI,34.58273 to 35.35809) for the residuals. The fixed intercept for the scores was found to be 108.6657350 (95% CI, 104.4715439 112.8599262). Lastly, the intercorrelation coefficient or the amount of variation in scores explained by the difference in subjects was found to be 64%.