Lab 4: Part 2

The aim of this part of the lab exercise is to give you practical experience in Correlation and Linear Regression using R Studio. Make sure you download the .csv file and it is recommended that you set up a new R Studio project in the same location as the file. Complete this worksheet within a .Rmd file in R Studio.

1 Linear Regression and Correlation

In this lab you will use data about birthweight to practice linear regression and correlation

In an investigation conducted by Secher et al., (1987), The birth weight (BW) in grams for 107 babies was ascertained. For all babies, both the abdominal (AD) and biparietal (BPD) diameters (in mm) were measured shortly before birth using ultrasound.

The purpose of this study was to describe the relationship between birthweight and these two ultrasound measurements in order to establish a way to predict birthweight (or fetus weight).

The data includes the following variables:

- bw: Birth weight of the baby in grams
- bpd: biparietal diameter (in mm), as determined by ultrasound
- ad: abdominal diameter (in mm), as determined by ultrasound
- id: identification of the mother (ignore this is not needed for the analysis)

1. Load the `birthweight.csv` data into R studio
2. Explore the data numerically and graphically
3. Are the three numerical columns correlated?
4. Build two linear regression models, each should have `bw` as the dependent variable and `bpd` or `ad` as the explanatory variable
5. Compare the diagnostics and fit of each of the two models Using only one predictor at a time, which is better at predicting AD or BD? and explain why?
6. (OPTIONAL) Try some transformations
7. An expectant mother has been told their bpd is 80 - what is the estimated birthweight? Explain your answer
8. An expectant mother has been told their ad is 105 - what is the estimated birthweight? Explain your answer

9. The mean birth weight in the UK populations is 3300 gr. Given the mean birthweight in this sample test a one-way hypothesis using this sample mean. Explain what H_0 and H_1 you are using and the conclusion you reach.

```r
# reading a csv file into R
my.data<-read.csv("name-of-my-file.csv")

#correlation coefficient r
cor(x,y)

#correlation coefficient and testing significance
cor.test(x,y)

#linear regression
lm(y~x)

#detailed output of the regression model
summary(lm(y~x))

#regression model plots
plot(lm(y~x))

#Hypothesis testing for a mean
t.test()
```