Homework1

Rohan Sadale

Jan 28, 2016

```
library(alr4)
```

```
## Warning: package 'alr4' was built under R version 3.2.3
```

```
## Warning: package 'car' was built under R version 3.2.3
```

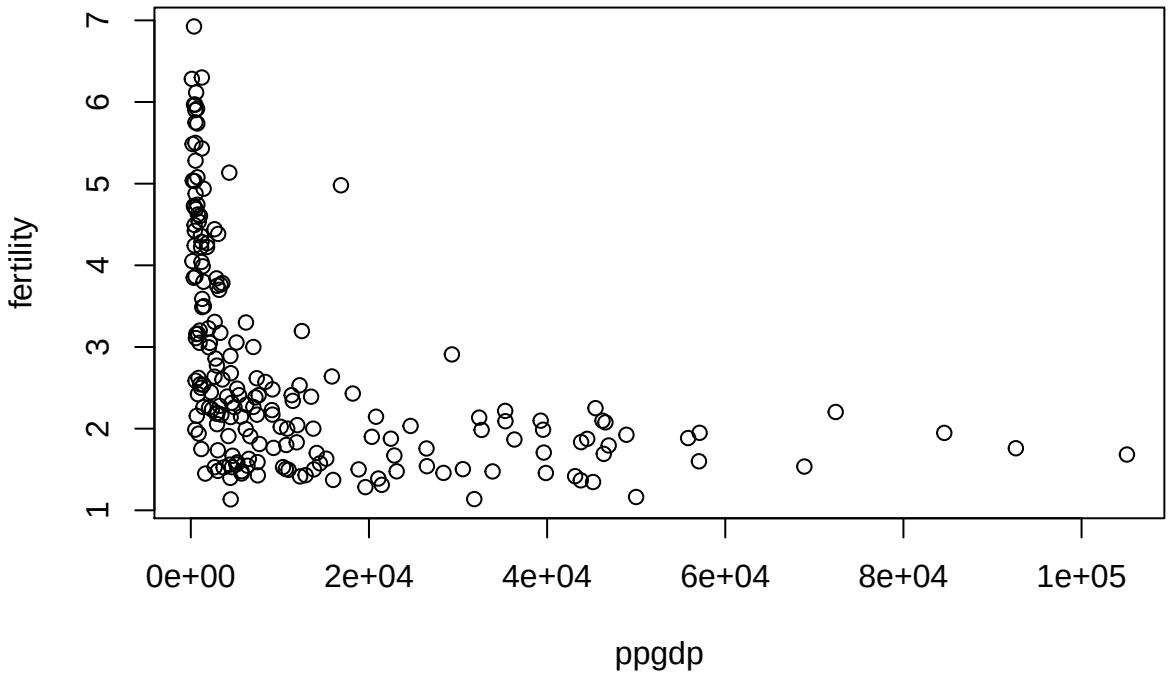
```
## Warning: package 'effects' was built under R version 3.2.3
```

1.1. 1.1.1

- Predictor: ppgdp
- Response: fertility

1.1.2

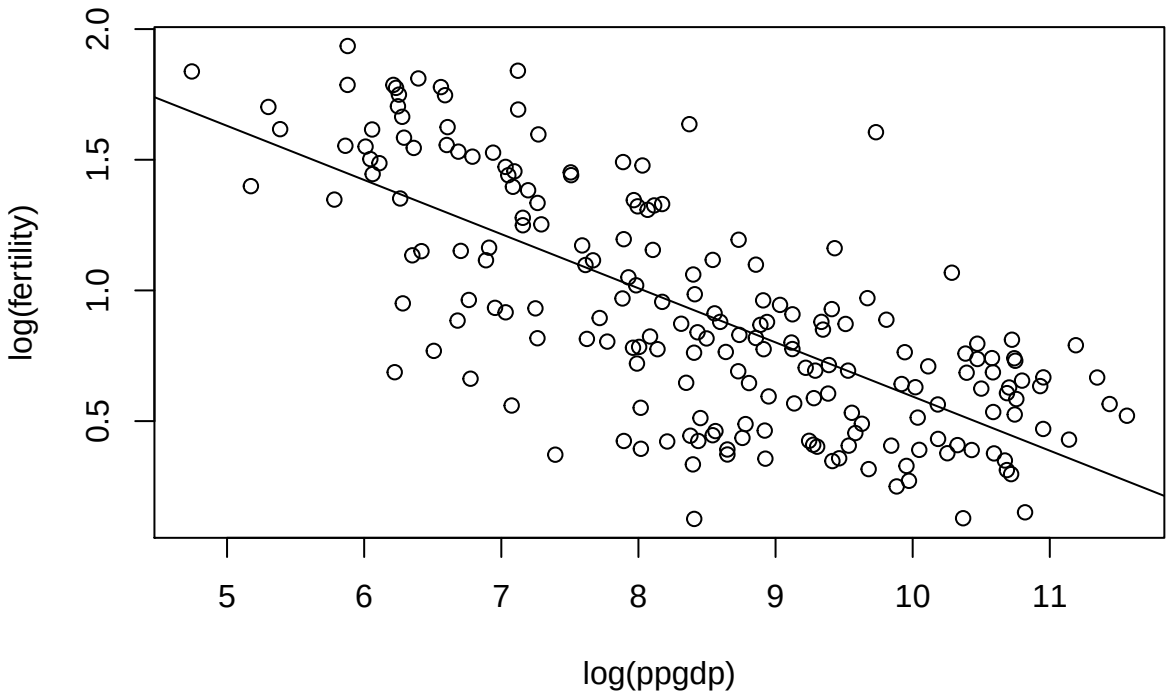
```
with(UN11, plot(fertility ~ ppgdp))
```



The plot is right skewed. Hence Simple Linear Regression cannot be a good fit for this graph. The straight line mean function is not possible for this graph.

1.1.3

```
with(UN11, plot(log(fertility) ~ log(ppgdp)))
model <- lm(log(fertility) ~ log(ppgdp), data = UN11)
abline(model)
```



Yes, as we can see from the plot, Simple Linear Regression seem plausible for a summary of this graph.

```
meanData <- with(wblake, tapply(Length, Age, mean))
print(meanData)
```

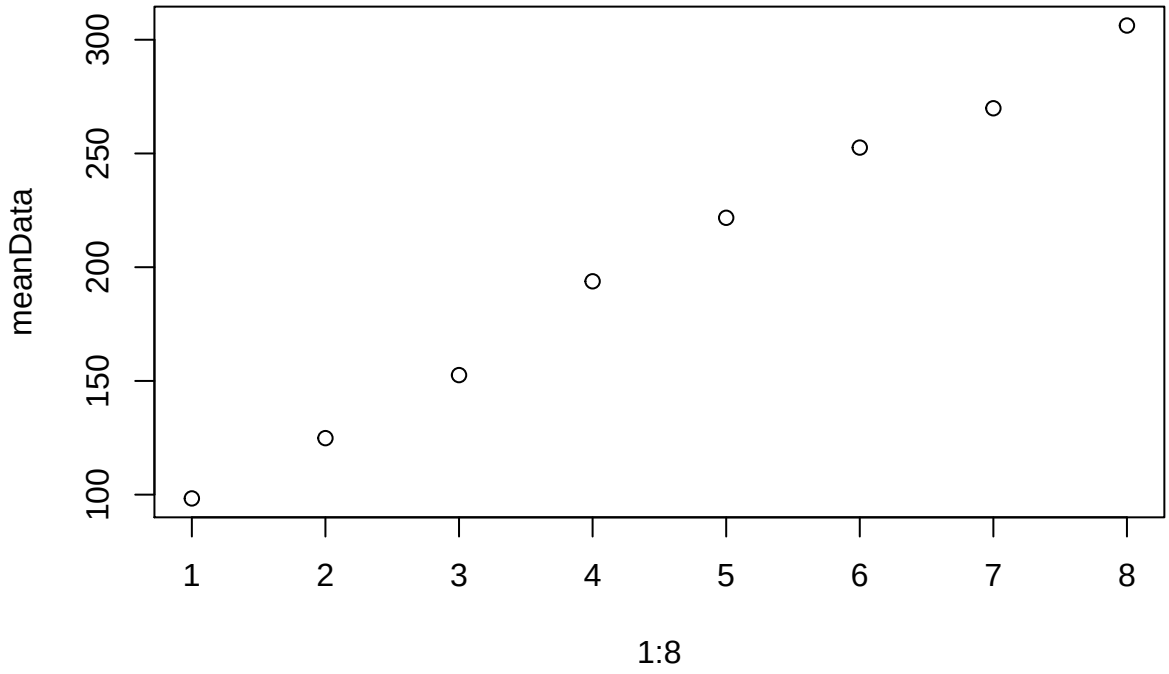
1.2

```
##         1         2         3         4         5         6         7
##  98.34211 124.84722 152.56383 193.80000 221.72059 252.59770 269.86885
##         8
## 306.25000
```

```
varianceData <-with(wblake, tapply(Length, Age, var))
print(varianceData)
```

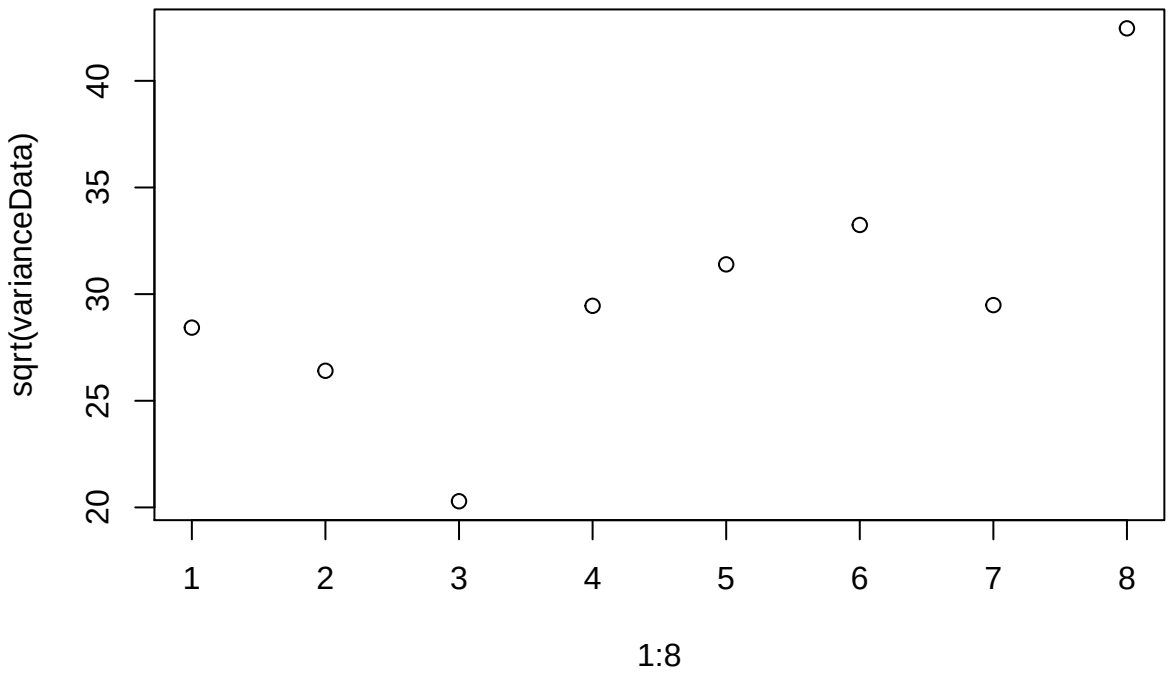
```
##         1         2         3         4         5         6         7
##  808.2312  697.2862  411.6679  867.4571  985.6969 1105.0805  869.3825
##         8
## 1802.9167
```

```
plot(1:8, meanData)
```



The graph shows a rising trend in average length of fish as it grows older, and a simple linear regression seems plausible for this graph. The graph has same pattern as compared to fig 1.5

```
plot(1:8, sqrt(varianceData))
```

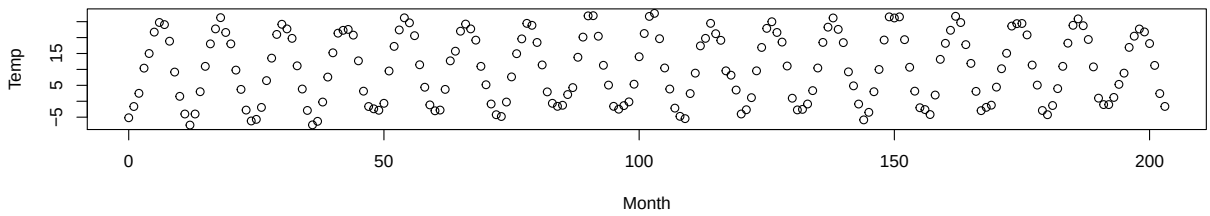


Though a straight line parallel to Age axis is possible on the plot, but the variance and standard deviation of the fishes at age 3 and age 8 will deviate significantly from a straight line. Thus, the plot is not a null plot.

1.3 1.3.1 There appears to be no dependence of average soil temperature on the number of months. The plot appears to be null plot as it doesn't show any particular important characteristics.

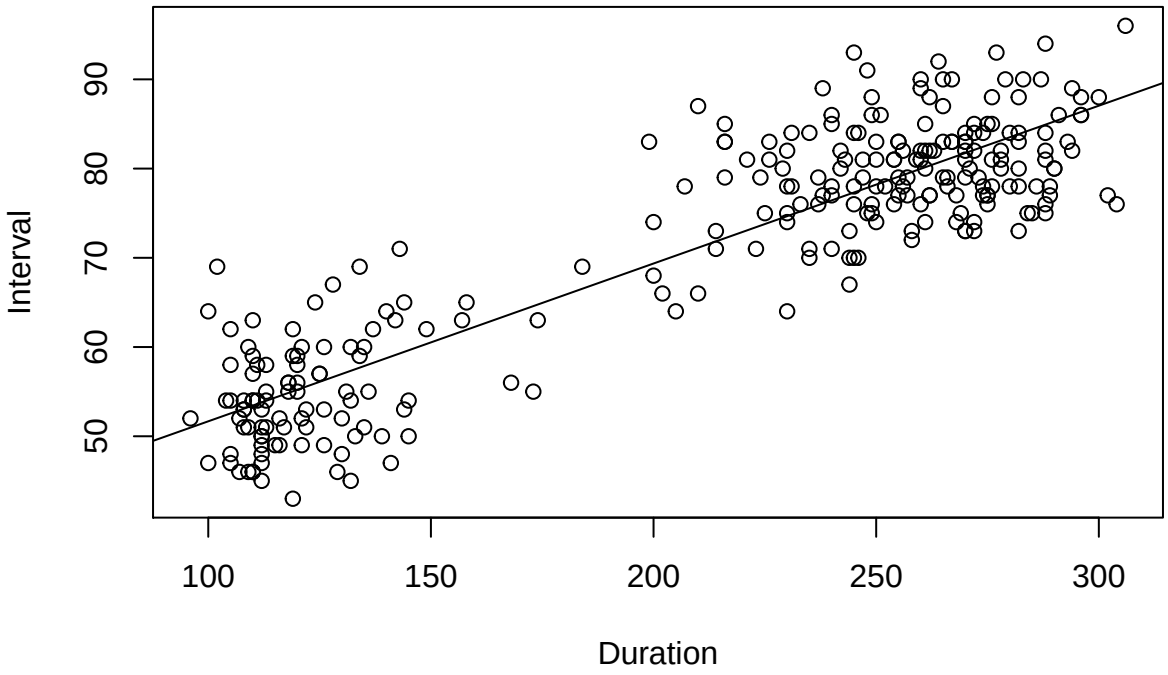
1.3.2

```
with(Mitchell, plot(Temp~Month))
```



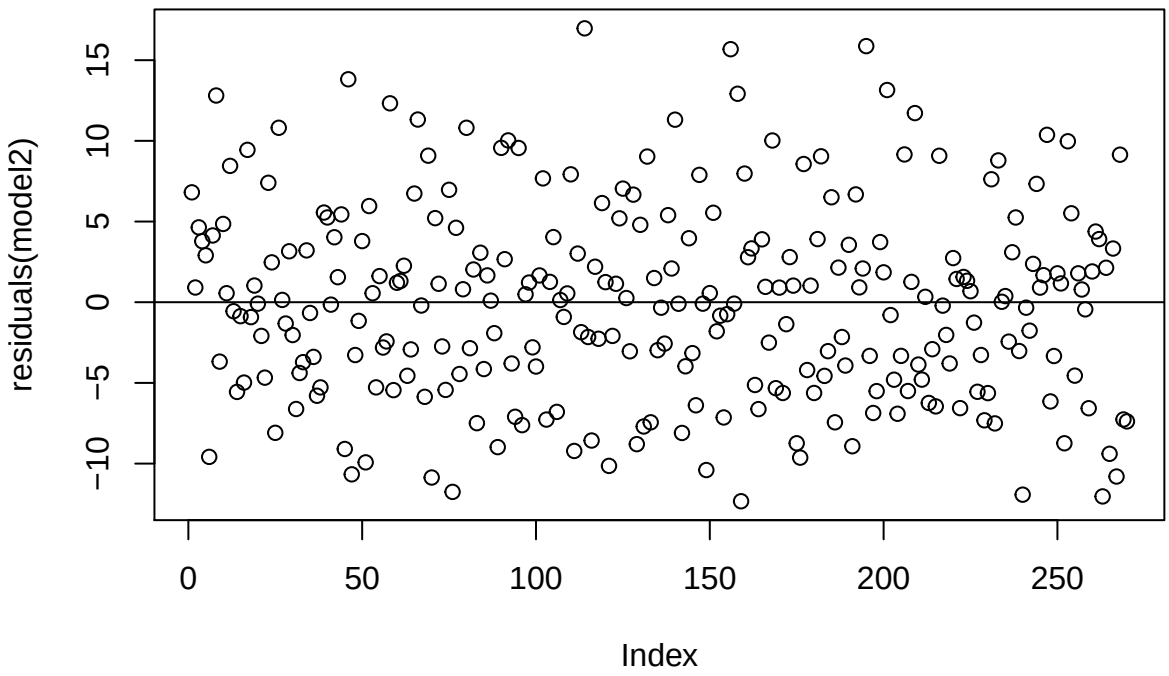
As the width changes, we are able to see a temporal pattern on data. We can say that the soil temperature falls in Winter and rises in summer.

```
with(oldfaith, plot(Interval~Duration))
model2 <- lm(Interval~Duration, data = oldfaith)
abline(model2)
```



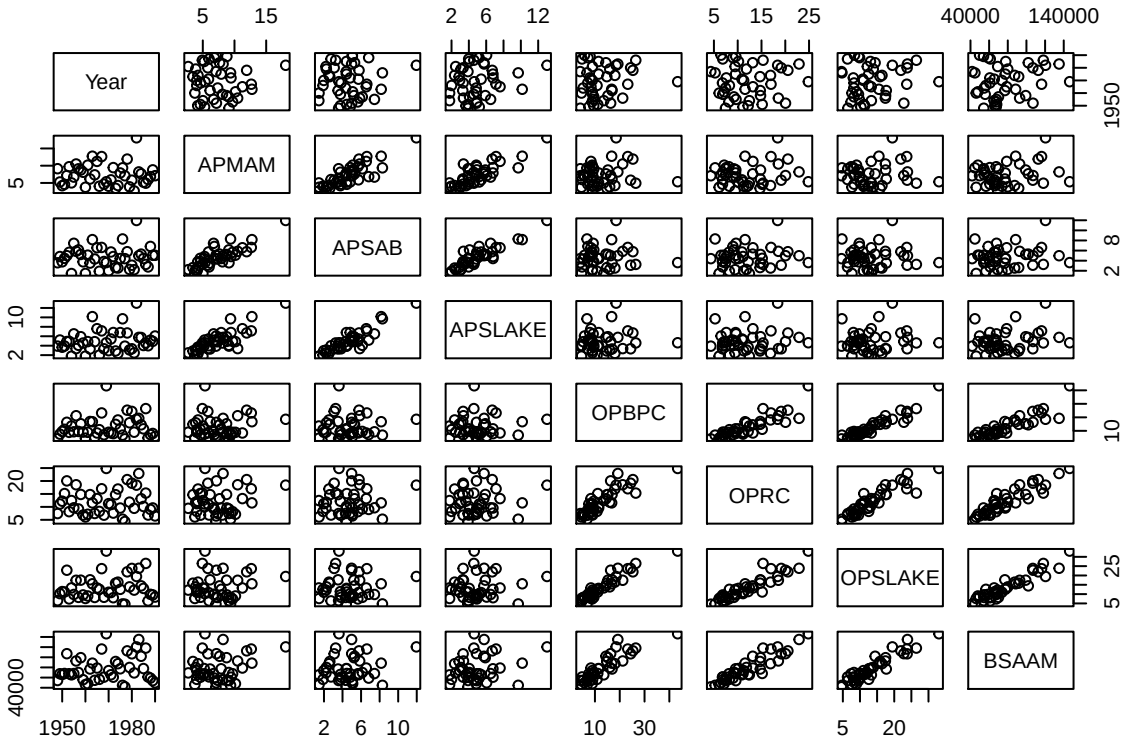
1.4

```
plot(residuals(model2))
abline(0,0)
```



The mean is almost linear, variance is almost constant and the residual plot can be a null plot. Thus, we can summarize the graph by simple linear regression.

```
plot(water)
```



1.5

* From the plot, we can say that Year is highly uncorelated to all other variables. * We can also notice that the variables OPBPC, OPRC, OPSLAKE and BSAAM are corelated with each other. Similarly, the variables - APMAM, APSAB and APSLAKE appear to be corelated with each other. * BSAAM is more corelated with OPBPC, OPRC and OPSLAK, while less corelated with APMAM, APSAB and APSLAKE. * We can also see that there is atleast one seperated point (outlier or leverage point) in each plot.

1.6

- We can see that quality and clarity, quality and helpfulness are highly corelated to each other. Clarity and helpfulness are a bit less corelated to each other as compared to former two.
- Easiness is very loosly corelated to helpfulness, clarity and quality.
- There seems no corelation between raterInterest and all other variables.