

## R Programming CIA 4

By Pranav Gopalkrishna, 1940223, 2CMS

2020-03-10

```
setwd("~/rAssignments/CIA 4")
mydata = read.csv("ChickWeight.csv")
head(mydata, 6)
```

```
##   X weight Time Chick Diet
## 1 1     42    0     1    1
## 2 2     51    2     1    1
## 3 3     59    4     1    1
## 4 4     64    6     1    1
## 5 5     76    8     1    1
## 6 6     93   10     1    1
```

```
attach(mydata)
#QUESTION 1
summary(mydata)
```

```
##           X           weight           Time           Chick
##  Min.   : 1.0   Min.   : 35.0   Min.   : 0.00   Min.   : 1.00
## 1st Qu.:145.2   1st Qu.: 63.0   1st Qu.: 4.00   1st Qu.:13.00
##  Median :289.5   Median :103.0   Median :10.00   Median :26.00
##  Mean   :289.5   Mean   :121.8   Mean   :10.72   Mean   :25.75
## 3rd Qu.:433.8   3rd Qu.:163.8   3rd Qu.:16.00   3rd Qu.:38.00
##  Max.   :578.0   Max.   :373.0   Max.   :21.00   Max.   :50.00
##           Diet
##  Min.   :1.000
## 1st Qu.:1.000
##  Median :2.000
##  Mean   :2.235
## 3rd Qu.:3.000
##  Max.   :4.000
```

*#Explanation:*

*#This is the serial number, and we see there are 578 observations*

*#the minimum weight of chicks is 35, and maximum is 373.*

*#The average weight is 121.8.*

*#(The most logical unit of measurement is grams)*

*#Measures the time elapsed from the start of the diet for each chick.*

*#We see that a diet for a chick was studied for maximum 21.*

*#Time always starts at 0 for every chick.*

*#(The most logical unit of measurement is days)*

```

#(Chick and Diet are factor variables)

#How many missing values?
sum(is.na(mydata))

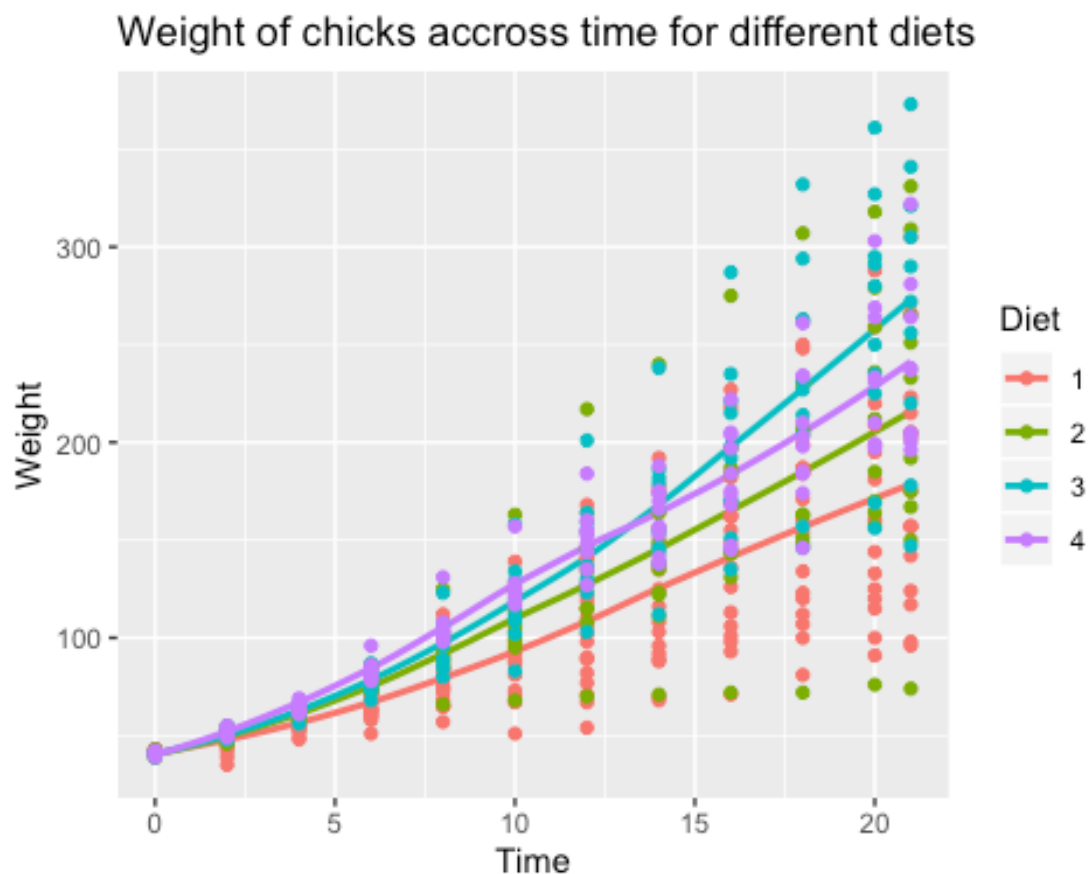
## [1] 0

#None

#QUESTION 2
library(ggplot2)
?geom_step
mydata$Diet = as.factor(Diet)
q2 = ggplot(mydata, aes(x = Time, y = weight, color = Diet)) + geom_smooth(se
= FALSE) + geom_point()
q2 + ggtitle("Weight of chicks accross time for different diets") +
ylab("Weight") + xlab("Time")

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'

```



```

#CONCLUSIONS
#The following conclusions are drawn from the preceding scatter and Linear
plot.
#In general, weight increases as time increases.

```

*#However, for different diets, the points scatter differently i.e. for different diets, the weights of chicks change differently.*  
*#Diet 2 seems to have the most range in its changes i.e. it shows small to negative weight change in some cases as well as high weight change in other cases.*  
*#Diet 3 seems to be the most successful diet overall, showing mid to high weight changes throughout the observations, and achieving the maximum weight changes for many cases.*  
*#Diet 4 seems to be the most consistent, since its points are the least scattered around the average line. It shows rising weight in all cases, even though it does not achieve the same maximum levels as diet 3.*  
*#Diet 1 seems to be a mid to low performing diet, since most of its points lower than the points of the other diets. It is the lowest performing diet of the four.*

### *#QUESTION 3*

```
cor(Time, weight)
```

```
## [1] 0.8371017
```

*#Time and weight show strong positive correlation i.e. as time increases, weight increases almost to the same extent.*

*#This makes sense, since chicks are likely to grow heavier as they grow older.*

```
?cor
```

```
cor(as.numeric(Diet), weight)
```

```
## [1] 0.2081957
```

*#Diet and weight show weak positive correlation i.e. a change in diet leads to a usually smaller change in weight.*

*#The positivity makes sense, since diet is an important contributor to growth, and thus, to weight.*

*#However, the weak correlation suggests diet is a factor, but not a strong factor, that is affecting chicks' weight.*

*#Hence, we can say that while diet must be considered, there may be other more important factors affecting weight.*

### *#QUESTION 4*

```
lm(weight~Time, mydata)
```

```
##
```

```
## Call:
```

```
## lm(formula = weight ~ Time, data = mydata)
```

```
##
```

```
## Coefficients:
```

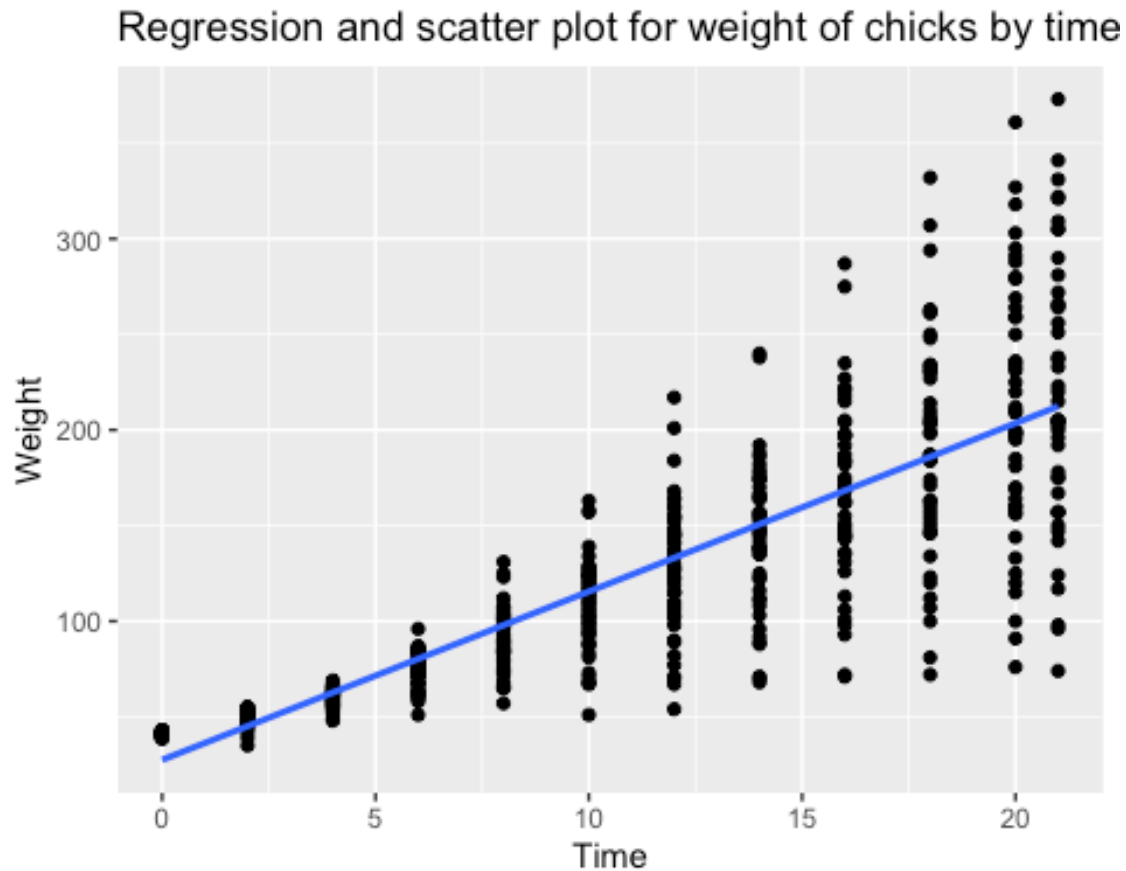
```
## (Intercept)      Time
```

```
##      27.467      8.803
```

*#Obtained equation:*

*#weight = 27.467 + 8.803\*Time*

```
q4pa = ggplot(mydata, aes(x = Time, y = weight, colours = "blue")) +
geom_point() + geom_smooth(method = "lm", se = FALSE)
q4pa + ggtitle("Regression and scatter plot for weight of chicks by time") +
xlab("Time") + ylab("Weight")
```



```
#From obtained equation, the weight of chicks at time = 11 is
27.467 + 8.803*11
```

```
## [1] 124.3
```

*#i.e. after 11 days, the chicks' weights are around 124.3g*

```
#From the obtained equation, when the weight of chick is 185, we have
```

```
#185 = 27.467 + 8.803*Time
```

```
#Hence, Time = (185 - 27.467) / 8.803
```

```
(185 - 27.467) / 8.803
```

```
## [1] 17.89538
```

*#i.e. it takes around 17.89538 i.e. around 18 days for chicks to achieve the weight of 185g.*