# ChickWeight

## Group2 (Akshita, Khalid, Nilesh)

### 31/01/2020

# 1. Understanding the Data

This **ChickWeight** dataset describes the body weights of the chicks measured at birth and every second day thereafter until day 20. They were also measured on day 21. **The purpose of the report is to examine the effect of 4 diets on early growth of chicks.**. The dataset has **578 rows and 4 columns**, having **weight**, **Time**, **Chick** and **Diet** as four variables with no missing values (checked using *is.na(ChickWeight)* function).

**Structure of ChickWeight Dataset:**

Firstly, we will describe the general structure of ChickWeight Dataset describing the variables (their datatypes and levels)

```
##      weight           Time           Chick     Diet
##  Min.   : 35.0   Min.   : 0.00   13     : 12   1:220
##  1st Qu.: 63.0   1st Qu.: 4.00   9      : 12   2:120
##  Median :103.0   Median :10.00   20     : 12   3:120
##  Mean   :121.8   Mean   :10.72   10     : 12   4:118
##  3rd Qu.:163.8   3rd Qu.:16.00   17     : 12
##  Max.   :373.0   Max.   :21.00   19     : 12
##                                  (Other):506
```

Using *str(ChickWeight)* function, we get the following details:
**Weight**
*datatype:* numeric
*description:* weight of the chicken
**Time**
*datatype:* numeric
*description:* the day weight was recorded beginning with day 0 as the day it was born
**Chick**
*datatype:* an ordered factor with levels 50
*description:* a unique identifier for the chick. For instance, the first row in chick variable shows that the chick with unique ID# 13 has the lowest weight, followed by chick ID 9, 20 and so on.
**Diet**
*datatype:* factor with levels 1, 2, 3, 4
*description:* indicates which experimental diet the chicks receive

# 2. Analyzing the Data

**Summary of the ChickWeight dataset:**

Here, we will focus on **Weight, Time, Chick and Diet - giving a sense of their central tendency, dispersion, range, levels, etc.** There are 2 numeric variables (weight and Time) for which the measure of central tendency, dispersion and range are calculated.

**Measures of central tendency**

```
##     Measure |  Weight |   Time
## 1:     Mean | 121.818 | 10.718
## 2:   Median |     103 |     10
## 3:     Mode |      41 |    0 2
```

**Measures of Dispersion**

```
##                           Measure |   Weight |   Time
## 1:                      Variance | 5051.223 | 45.676
## 2:            Standard Deviation |   71.072 |  6.758
## 3: Inter-Quartile Range (IQR) |   100.75 |     12
## 4:                         Range |   35 373 |   0 21
```

**Summarizing the above measures grouped by diet**

Average weights for each diet is shown in the following graph:



From the above graph, the average weight of *Diet 3 > Diet 4 > Diet 2 > Diet 1*. In order to understand the distribution, skewness, and general idea of mean and median, the following graphs show the *overall weight vs Diet*, and *individual weight vs Diet.*
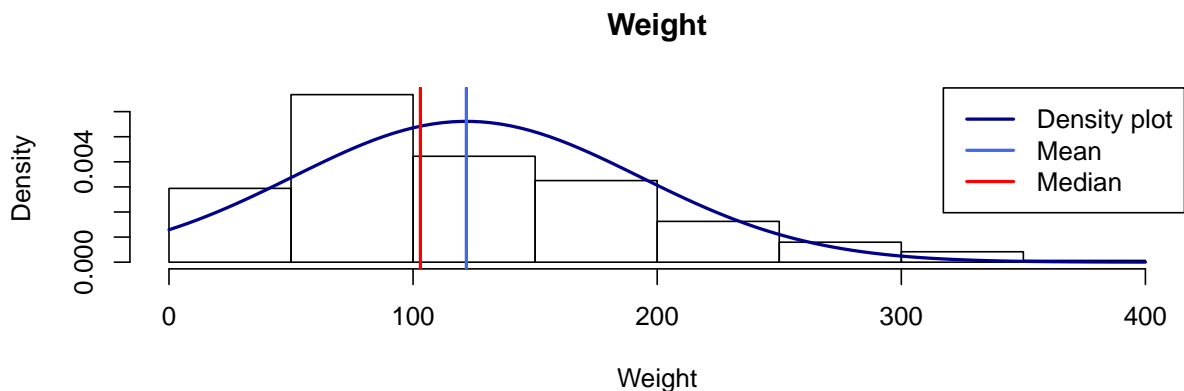


It is observed that the distribution is rightly skewed, as Mean > Median.

| Weight for Diet 1 | Weight for Diet 2 | Weight for Diet 3 | Weight for Diet 4 |
|---|---|---|---|
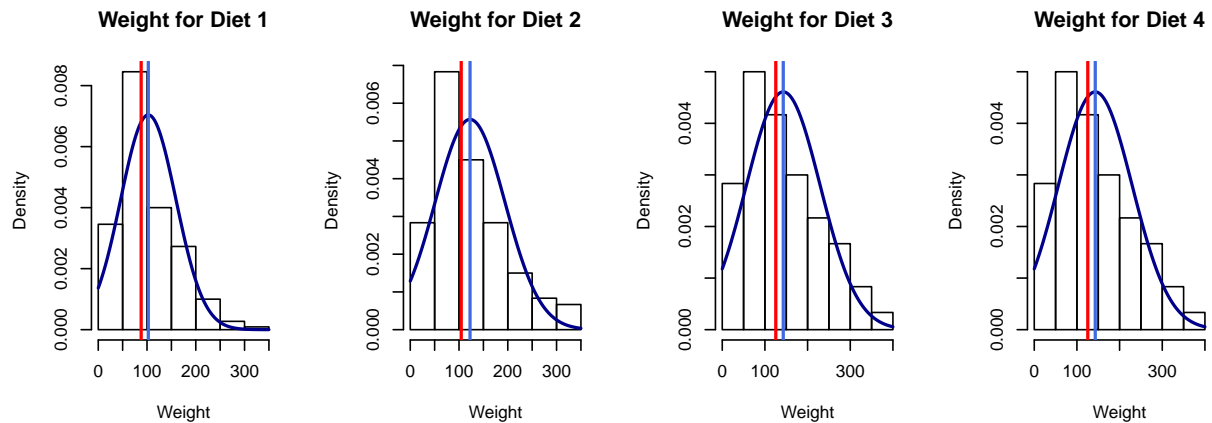


Overall Co-relation for Weight~Time for the given dataset

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

**Confidence intervals for mean weight at day 20 for each diet**
*This is the analysis for part b of Assignment 1*
Here we are anlaysing the mean weight observed with each of the 4 diets with 95% confidence intervals to further confirm the weight~Diet relationship that was observed above.

```
##      Diets | Sample_size | Avg_Weight | Lower_Value | Upper_Value | Range_of_CI
## 1: Diet 1 |          17 |   170.4118 |    141.9093 |    198.9142 |    57.00496
## 2: Diet 2 |          10 |   205.6000 |    155.3446 |    255.8554 |   100.51086
## 3: Diet 3 |          10 |   258.9000 |    212.2273 |    305.5727 |    93.34535
## 4: Diet 4 |           9 |   233.8889 |    205.0115 |    262.7663 |    57.75477
```
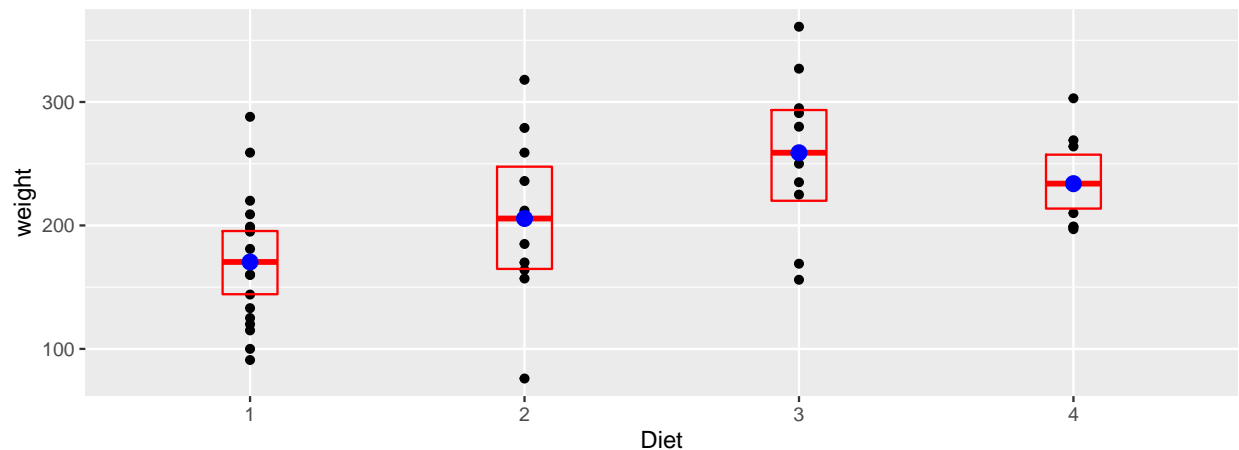


The above graph depicts confidence intervals for mean weight at day 20 for each diet. It is observed that the Range of CI for Diet 2 > Diet 3 > Diet 4 ~ Diet 1. However, the average weight of Diet 3 > Diet 4 > Diet 2 > Diet 1.

**General trend of Weight observed with each diet over a period of 21 days (using GGplot and Face grid functions)**

3
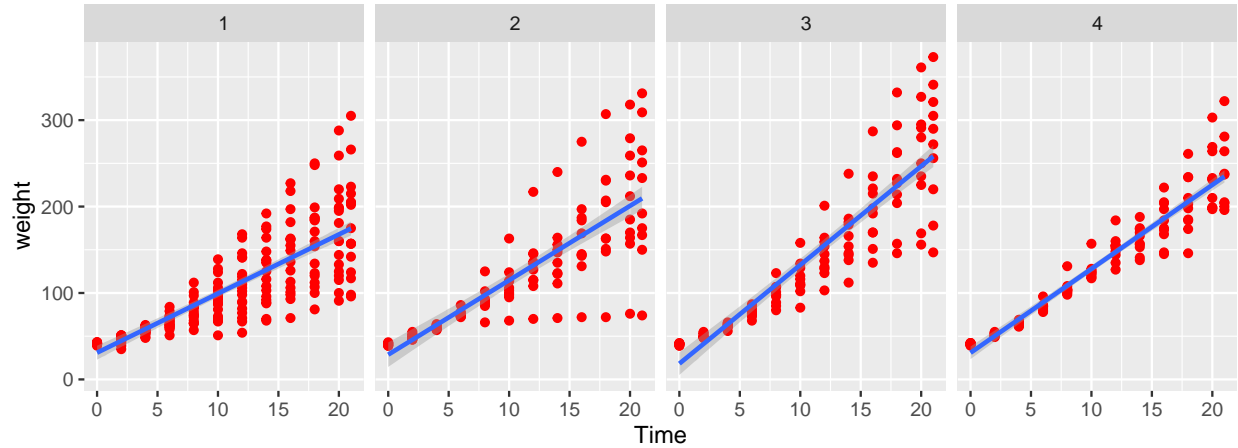
We observe that there is an increase in weight for all four Diets over 21 days, with *highest increase for Diet 3* (as we can see that it has the *maximum positive slope*).

# 3. Conclusion

**Conclusion**

1. The average weights for all four Diets are positively correlated with Time i.e. there is strong linear relationship and both the variables move in the same direction, but there is maximum correlation for Diet 3 and minimum for Diet 1 w.r.t time for 20 days period.

2. The average weight of *Diet 3 > Diet 4 > Diet 2 > Diet 1*

3. There is an increase in weight for all four Diets over 21 days, with *highest increase for Diet 3*

**Future Scope**

1. Overall co-relation for Weight~Time for the given dataset indicates a strong linear relationship i.e both the variables are positively corelated with one another, so if we have data for more days and if we go beyond 20 days, the weight of the chicken will further increase in a linearly.