Chicken Weight Dataset

Looking at the dataset

A help search on this dataset in R returns the following summary of data :

The **chickWeight** data is a built in data “Weight versus age of chicks on different diets”, this dataframe has **578 rows and 4 columns** from an experiment on the effect of diet on early growth of chicks.It includes the following data-features :

Weight: a numeric vector giving the body weight of the chick (gm).

Time: a numeric vector giving the number of days since birth when the measurement was made.

Chick: an ordered factor with levels 18 < ... < 48 giving a unique identifier for the chick. The ordering of the levels groups chicks on the same diet together and orders them according to their final weight (lightest to heaviest) within diet.

Diet: a factor with levels 1,2,3, 4 indicating which experimental diet the chick received.

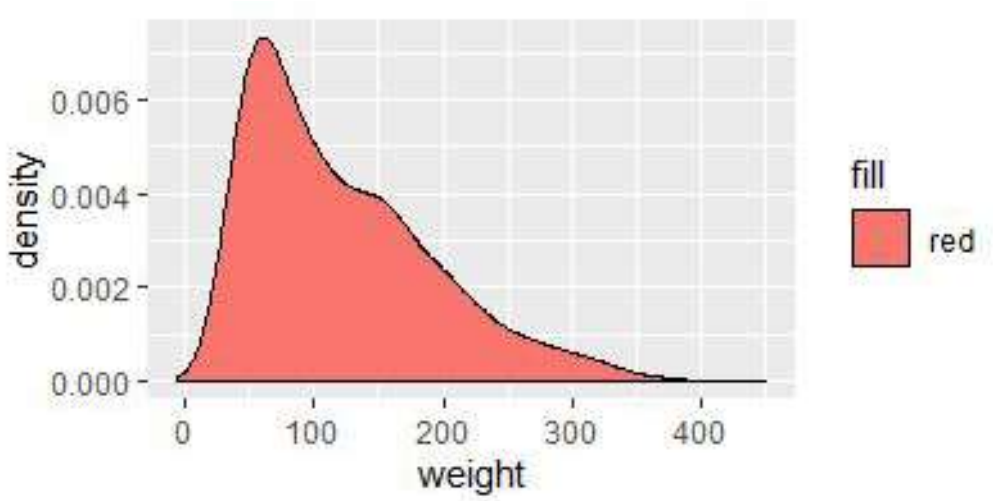
Let's look at a few samples from the data set

weight	Time	Chick	Diet
163	10	21	2
181	20	11	1
101	16	10	1
152	12	49	4
66	4	40	3

By looking at the above data, we can see we have chicken weight for various times(vary from day=0 to day =21), and this data is for various chick which are labelled (from 1 to 50) and are kept on different diets(diets vary in group of chicks and are labelled as 1,2,3 and 4).

Let's dive deep into each parameter’s data we collected for our experiment :

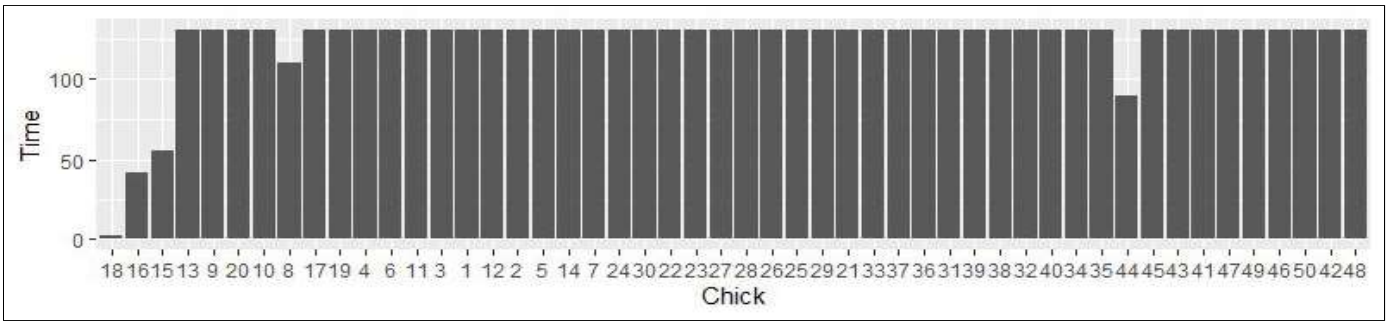
1. Weight: It's a numerical- ratio variable, for the given data it varies from 35 gms to 373 gms with a mean of 121.82 gms, median of 103 gms and standard deviation of 71.07.Lets see the distribution of weight for the given data in the density plot, depicted below. By looking at the plot we can see, the weight distribution is positively skewed, which means there is asymmetry in data, which can be because of the following :



- 1. Different diets affect the weight in different ways and there is a non- uniform distribution of the number of chicken for each diet.
- 2. Data for some chicks is not taken for all of the days.

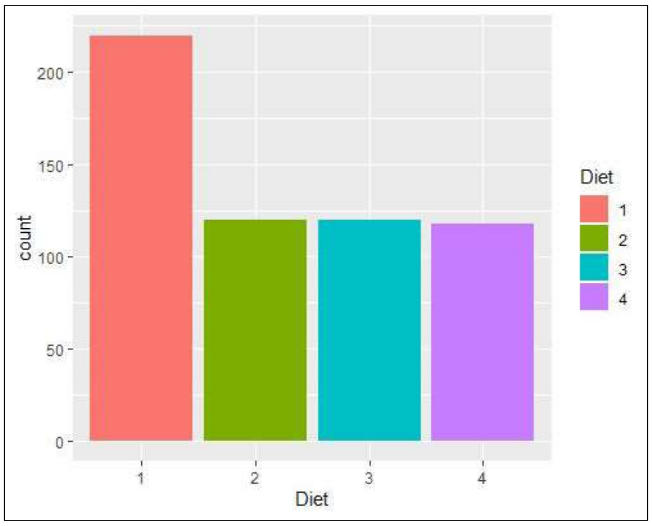
2. Time: It is a numerical ratio variable, measured in number of days.It represents number of days from the birth of the chicken.If we calculate the statistic for this data, we find out it has a mean count of 10.72, median count of 10 days and a range of 0 to 21 days.
By looking at these statistics we can say that we do not have access for chicken data for all of the chickens for all 0 to 21 days.

3. Chick: Chicks column represents the labels used to identify the 50 chickens kept under observation, it is stored as ordinal data ranging from 1 to 50.Let's plot the number of observations we have for various chicks.



By looking at the above plot we can see, except for chick specified as 18,16,15,8 and 44 we have similar day (time) data for remaining chicks, thus verifying the point(2) made above.

4.Diet: It's a nominal parameter, which varies as numerical values as 1,2,3,4.Lets see the bar plot for each type of Diet.

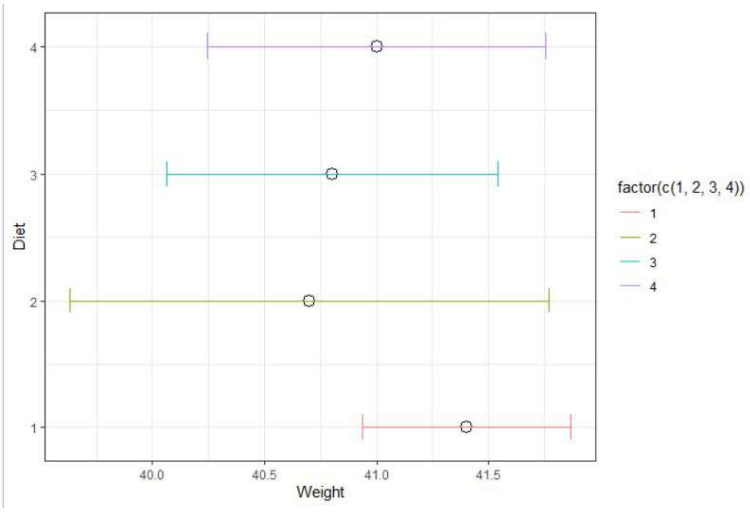


From the Plot and counts performed in R file, we can see, we have 220 observations for Diet 1, 120 observations for Diet 2, 120 observations for Diet 3 and 118 observations for Diet 4. Moreover, if we measure observations taken at day 0, we would find that there were 20,10,10 and 10 observations for diet 1,2,3 and 4 respectively.

Analysing the data

Since the experiment is about understanding the effect of four diets (1,2,3 and 4) lets see on chicken weight, the weight of chicken at day 0.

By looking at the graph plotted on the left and calculations performed in R, we can say that at **day = 0**, we had following observations for chicks on these four diets.



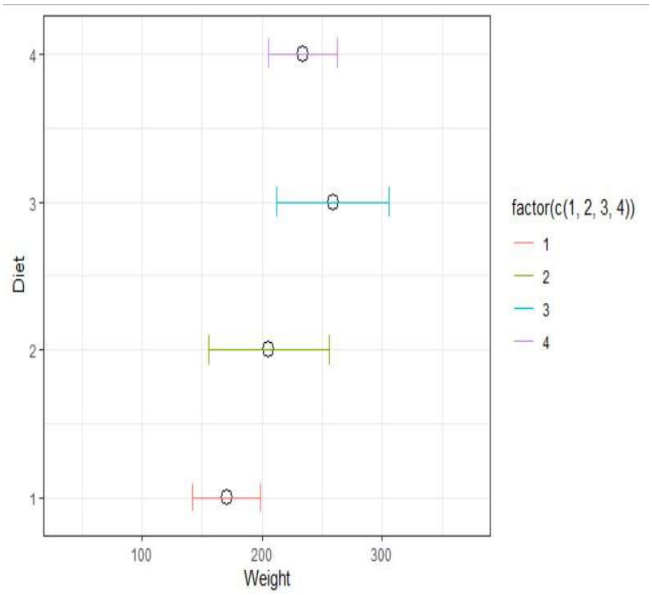
Diet 1 : has a mean weight of 41.4 gms can be found from 40.93 gms to 41.46 gms with 95 percent confidence interval.

Diet 2: has a mean weight of 41.76 gms can be found among 39.63 gms to 41.76 gms with 95 percent confidence interval.

Diet 3: has a mean weight of 40.80 gms can be found among 40.06 gms to 41.54 gms with 95 percent confidence interval.

Diet 4: has a mean weight of 41.0 gms can be found among 40.24 gms to 41.75 gms with 95 percent confidence interval. The experiment is conducted for 21 days lets see how the weight has varied on day 20 with various diets. Almost all of chicks at day =0 have weight varying from approximately 39 gms to 43 gms.

By looking at the graph plotted on the right and calculations performed in R, we can say that at **day = 20**, we had following observations for chicks on these four diets.



Diet 1 : has a mean weight of 170.41 gms can be found from 141.91 gms to 198.91 gms with 95 percent confidence interval.

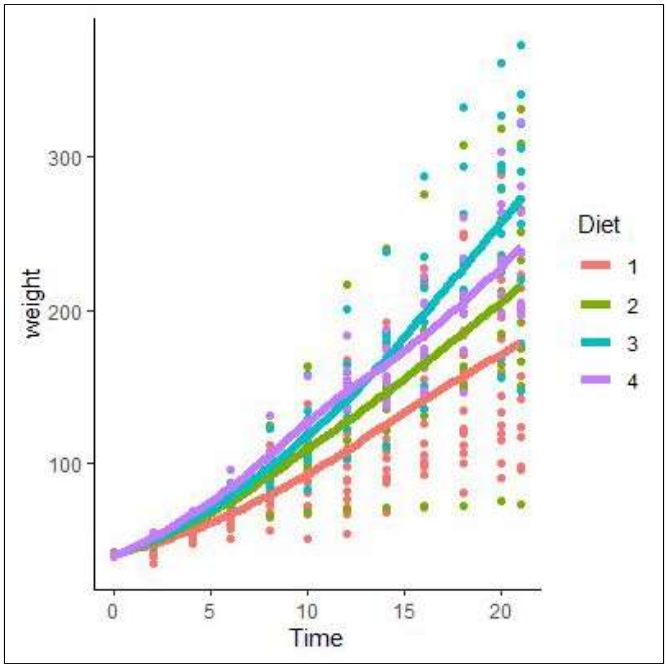
Diet 2: has a mean weight of 205.60 gms can be found among 155.34 gms to 255.85 gms with 95 percent confidence interval.

Diet 3: has a mean weight of 258.90 gms can be found among 212.23 gms to 305.57 gms with 95 percent confidence interval.

Diet 4: has a mean weight of 233.89 gms can be found among 205.01 gms to 262.76 gms with 95 percent confidence interval.

Thus, a simple observation leads to an inference that diet 3 had the most weight gaining effect on the chicken and diet 1 has been least effective in gaining weight compared to the other three diets.

Concluding the analysis:



From the plot on the left we can see how each diet affected the weight of chicken from day =0 to day = 21. The fitting curves explains that the order of weight growth for chicken kept on diet 3 was highest and the order of weight growth for chickens kept on diet 1 was lowest. The relative effect on weight of chicken for the four diets can be written as :

Diet 1 < Diet 2 < Diet 4 < Diet 3.

Thus we can conclude, it is preferable to keep the chicken on diet 3, if the goal is to maximize its weight by the end of day 20, and to keep chicken on diet 1, if the goal is not to increase its weight too much by the end of day 20.

This report is put together by Akshay Kapoor and Rishi Shah. Akshay has worked on plotting on graphs and assessing the dataset, while Rishi has focussed on analysis and statistical calculations.