

# ChickWeight Data Set Report

Shadh Shanavas | Nada Saiyed | Neethu Mariya

2020-02-02

## 1. Data Summary

The ChickWeight data frame provides us with information on the effect of four different diets on different chicks. It has **578 observations** and it has **4 variables**. The variables are **Weight, Time, Chick** and **Diet**.

A summary of all the variables in ChickWeight Data Set is given below which gives us the details about all 4 columns:

dfSummary(ChickWeight)					
## -----					
##	No Variable	Stats / Values	Freqs (% of Valid)	Graph	Valid
##	d Missing				
## -----					
## 1	weight	Mean (sd) : 121.8 (71.1)	212 distinct values	:	578
## 0					
##	[numeric]	min < med < max:		:	(100
##	(0%)				
##		35 < 103 < 373		: : .	
##		IQR (CV) : 100.8 (0.6)		: : : :	
##				: : : : . .	
## 2	Time	Mean (sd) : 10.7 (6.8)	12 distinct values	:	578
## 0					
##	[numeric]	min < med < max:		:	(100
##	(0%)				
##		0 < 10 < 21		: . . . . . . . . . .	
##		IQR (CV) : 12 (0.6)		: : : : : : : : : :	
##				: : : : : : : : : :	
## 3	Chick	1. 18	2 ( 0.3%)		578
## 0					
##	[ordered, factor]	2. 16	7 ( 1.2%)		(100
##	(0%)				
##		3. 15	8 ( 1.4%)		
##		4. 13	12 ( 2.1%)		
##		5. 9	12 ( 2.1%)		
##		6. 20	12 ( 2.1%)		
##		7. 10	12 ( 2.1%)		
##		8. 8	11 ( 1.9%)		
##		9. 17	12 ( 2.1%)		
##		10. 19	12 ( 2.1%)		
##		[ 40 others ]	478 (82.7%)	IIIIIIIIIIIIIIIIIIII	
## 4	Diet	1. 1	220 (38.1%)	IIIIIII	578
## 0					
##	[factor]	2. 2	120 (20.8%)	IIIII	(100
##	(0%)				
##		3. 3	120 (20.8%)	IIII	
##		4. 4	118 (20.4%)	IIII	
##					
## -----					

**Time** is a numeric vector which gives the number of days elapsed since the birth of the chick. Time can take on the following values: 0,2,4,6,8,10,12,14,16,18,20 and 21.

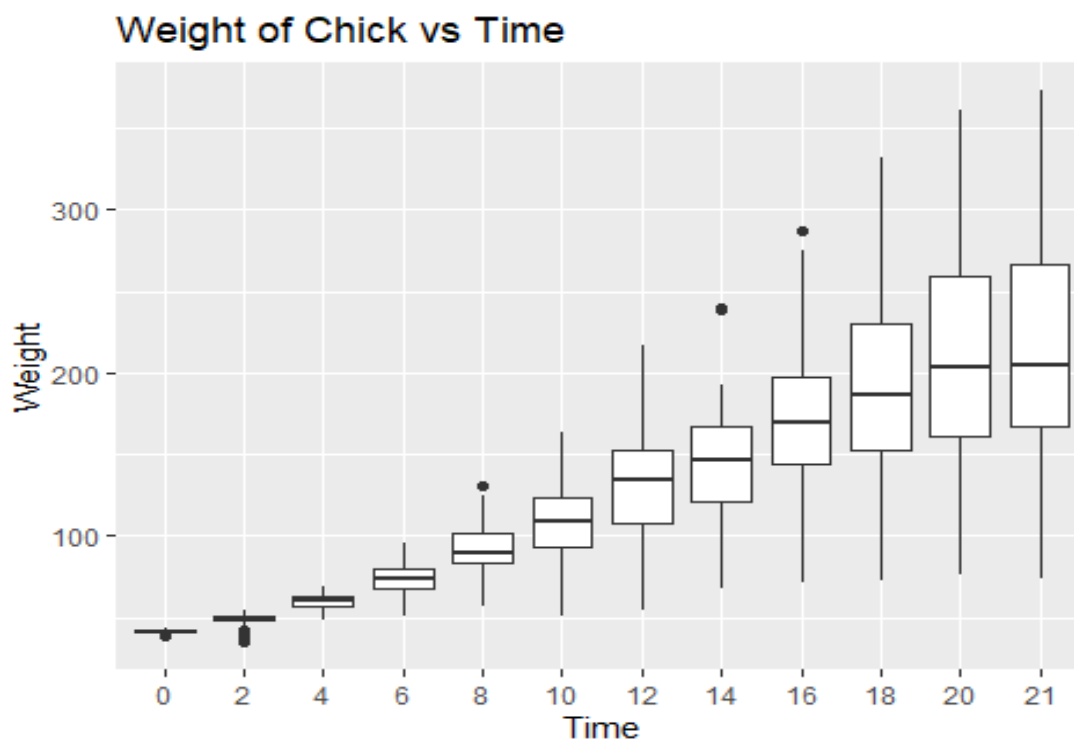
**Weight** is a numeric vector which gives the weight of the chick measured on each of these days, weight ranges from 35-373 and it has a **standard deviation of 71.07196**(high variability). From the rough histogram that we obtained and considering that **mean (121.8)** is more than **median (103)**, we realize that weight distribution has positive skewness and is right skewed.

**Diet** is a factor with levels 1 to 4 and it indicates which experimental diet the chicks have received. From the frequency distribution graph we see that Diet 1 has been served approximately 100% more times compared to rest of the 3 diets.

**Chick** is an ordered factor which takes on values from 1 to 50 and it serves as a unique identifier for the chick. A peculiar thing that we notice here, when we look at the mode of Chick column, is that chick 18 has no records after 2 days, chick 16 has records for just 7 days and chick 15 for 8 days. While for chick 44 and 8 it is 10 and 11, rest of them have complete 12 records each.

Since weight is the variable that we are going to be analyzing, we can get an understanding of how weight varies across the time period by looking at the boxplot below.

```
ChickWeight %>% ggplot(aes(x=as.factor(Time),y=weight))+geom_boxplot()+labs(x="Time",y="Weight",title="Weight of Chick vs Time")
```



We can easily understand from this that with Time, the median weight of the Chicks increase

## 2. Analysis

We can calculate the confidence intervals for mean weight at day 20 for each Diet by first finding subsets of the data for each diet and then using the CI function “Rmisc” package.

##	upper	mean	lower	Diet
## 1	198.9142	170.4118	141.9093	1
## 2	255.8554	205.6000	155.3446	2
## 3	305.5727	258.9000	212.2273	3
## 4	262.7663	233.8889	205.0115	4



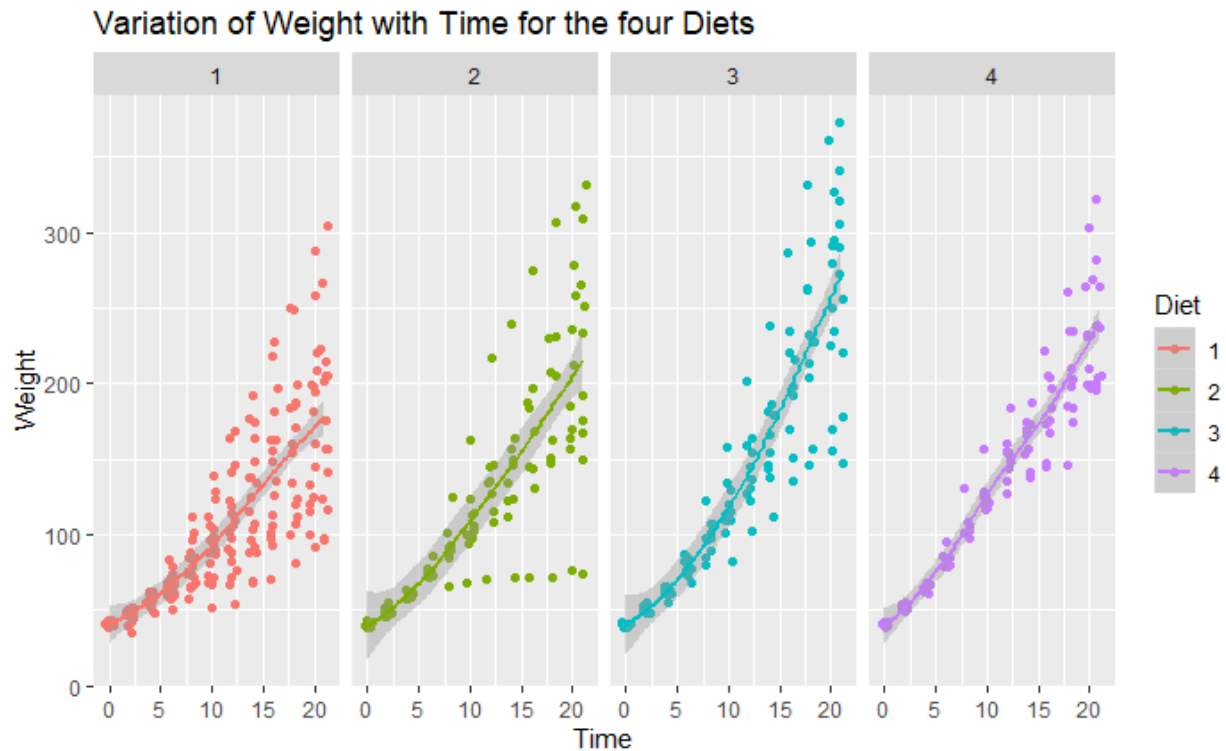
From the above graph it is evident that Diet 3 has the highest upper limit, lower limit and mean. This is followed by Diet 4. which has the 2nd highest upper limit, lower limit and mean value. Diet 2 has comparatively almost equal upper limit as of Diet 4 while the mean is approximately equal to Diet 4's lower limit. Diet 1 has the lowest lower limit, upper limit and mean values overall.

We can also infer that variability in weight for Diet 2 is the highest, followed by Diet 3, whereas Diet 1 and Diet 4 has less variability for weight.

## Conclusion

Analyzing the previous C.I. graph, we can claim that Diet 3 gives the maximum increase in weight of chicks in 20 days, compared to rest of the 3 diets. Assuming that the cost of each diet is same, our friend should avoid using Diet 1 as it gives the least growth and prefer Diet 3 for his chicks.

We plot weight vs time graph, grouped by Diets, to provide further evidence to our claim.



From the plot, we can clearly see that Diet 3 is the best for most weight gain, while Diet 4 has most linear relationship. Diet 4 has least variability as well, while Diet 1 and Diet 3 has most variability.