Abstract

- ***Often we can come across a scenario where someone might be interested to know which treatment be it different drug-doses to prevent certain disease, different combination of fertilizers to get maximum yield or in this case which diet will perform better for chickens over time, then this case study may come in handy.***
- I had started with ***exploratory data analysis*** which gave an insight of the data then continued with fitting ***regression model***. Here one could see the use of ***non-parametric bootstrap*** to generate samples for regression coefficients. After that using ***boxplot, confidence interval, testing of hypothesis*** I reached to a conclusion.
- ***One can refer this file to understand the findings.***

Description of Dataset

- The ‘**ChickWeight**’ data frame has 578 rows and 4 columns from an experiment on effect of diet on growth of chicks.
- 4 columns are –

weight : body weight of the chick (gm)

Time : number of days since birth

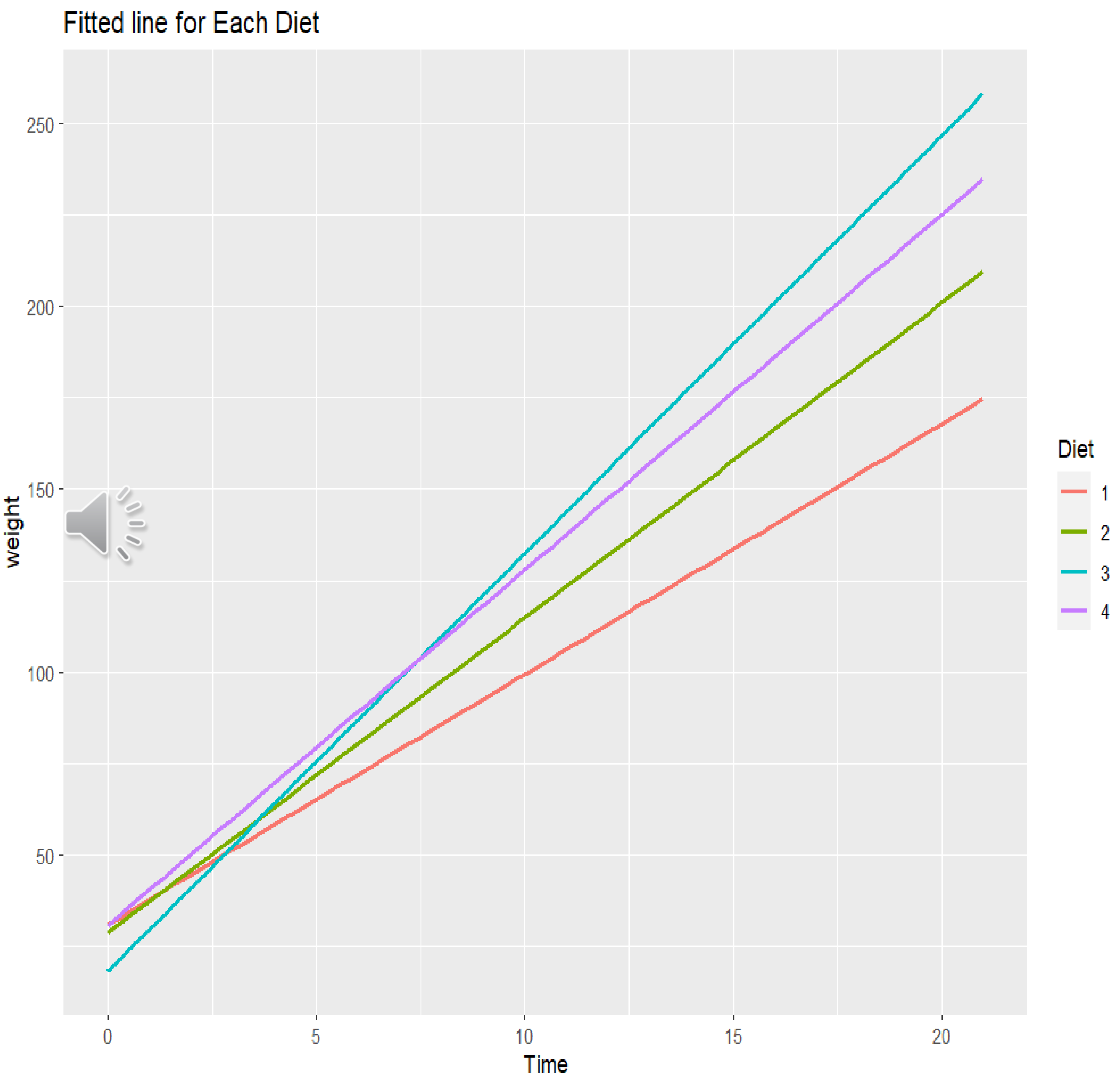
Chick : unique identifier for the chick with levels 1<2<….<50

Diet : indicates which diet the chick has received. It has 4 categories – 1,2,3,4

Exploratory Data Analysis

- At the time of birth, chicks’ weight corresponding to Diet 1 and 4 are alike, for Diet 3 it is a little less. But for Diet 3 initial weight is the minimum of all.
- As time increases slope of the fitted line for Diet 3 is maximum, followed by Diet 4, 2, 1 respectively.

Which essentially indicates that even though weight at time of birth was not that high for the chicks having Diet 3 as compared to the other chicks, but over time increase in expected growth of those chicks is much higher than that of Diet 4,3,1.



Model Fitting (just for reference)

matrix of indicator variables for diet :

	X1	X2	X3
Diet-1	0	0	0
Diet-2	1	0	0
Diet-3	0	1	0
Diet-4	0	0	1

Model :

w = weight ; t = Time

w=b0+b11*X1+b12*X2+b13*X3+b21*X1*t+b22*X2*t+b23*X3*t+b3*t

Diet-1 : w=b0+b3*t

Diet-2 : w=(b0+b11)+(b3+b21)*t ; b21=diff between the slopes of Diet-2 and Diet-1

Diet-3 : w=(b0+b12)+(b3+b22)*t ; b22=diff between the slopes of Diet-3 and Diet-1

Diet-4 : w=(b0+b13)+(b3+b23)*t ; b23=diff between the slopes of Diet-4 and Diet-1

Estimates of Regression Coefficients and Interpretation

```
                Estimate
• (Intercept)   30.9310
• Time           6.8418
• Diet2         -2.2974
• Diet3        -12.6807
• Diet4         -0.1389
• Time:Diet2     1.7673   # b21
• Time:Diet3     4.5811   # b22
• Time:Diet4     2.8726   # b23
```

Interpretation :

- b21 will represent the difference between the increase in expected weight of chicks having Diet-2 and the chicks having Diet-1 if time increases by 1 unit.
- b22 will represent the difference between the increase in expected weight of chicks having Diet-3 and the chicks having Diet-1 if time increases by 1 unit.
- b23 will represent the difference between the increase in expected weight of chicks having Diet-4 and the chicks having Diet-1 if time increases by 1 unit.

Here b21, b22, b23 > 0 which essentially indicates **expected increase in weight by unit value increase in time for Diet-2, 3, 4 is more as compared to Diet-1. Also Diet-3 seems to have more impact on weight gain than that of others.** But our data is just a sample from the whole population. So let’s test our assumptions.

Let’s Verify!

Methods to be used :

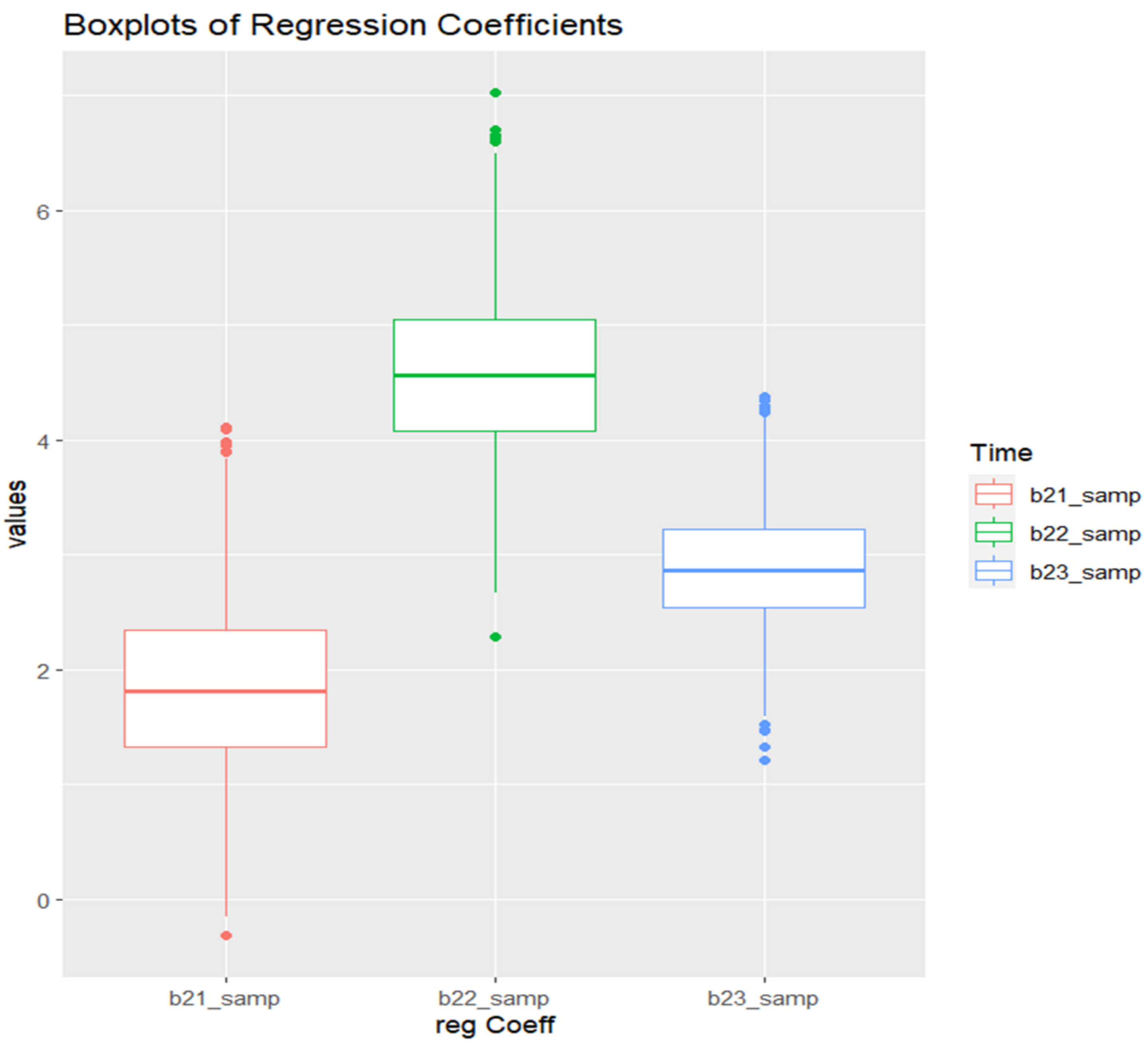
- Observing the distributions of b21, b22, b23 by Boxplot.
- Plotting Confidence Interval for b21 , b22, b23.
- Testing of Hypothesis

Non-parametric Bootstrap : Here we will not make any distributional assumption for the regression coefficients, we will use non-parametric bootstrap to get samples where each time we will reshuffle the rows i.e observations and will estimate the coefficients for every resampling.

In this analysis 1000 bootstrap samples were taken.

Boxplot

We will plot the samples of difference of increase in weight between the groups for Diet-1 and 2,3,4 respectively.



Summary of the plot :

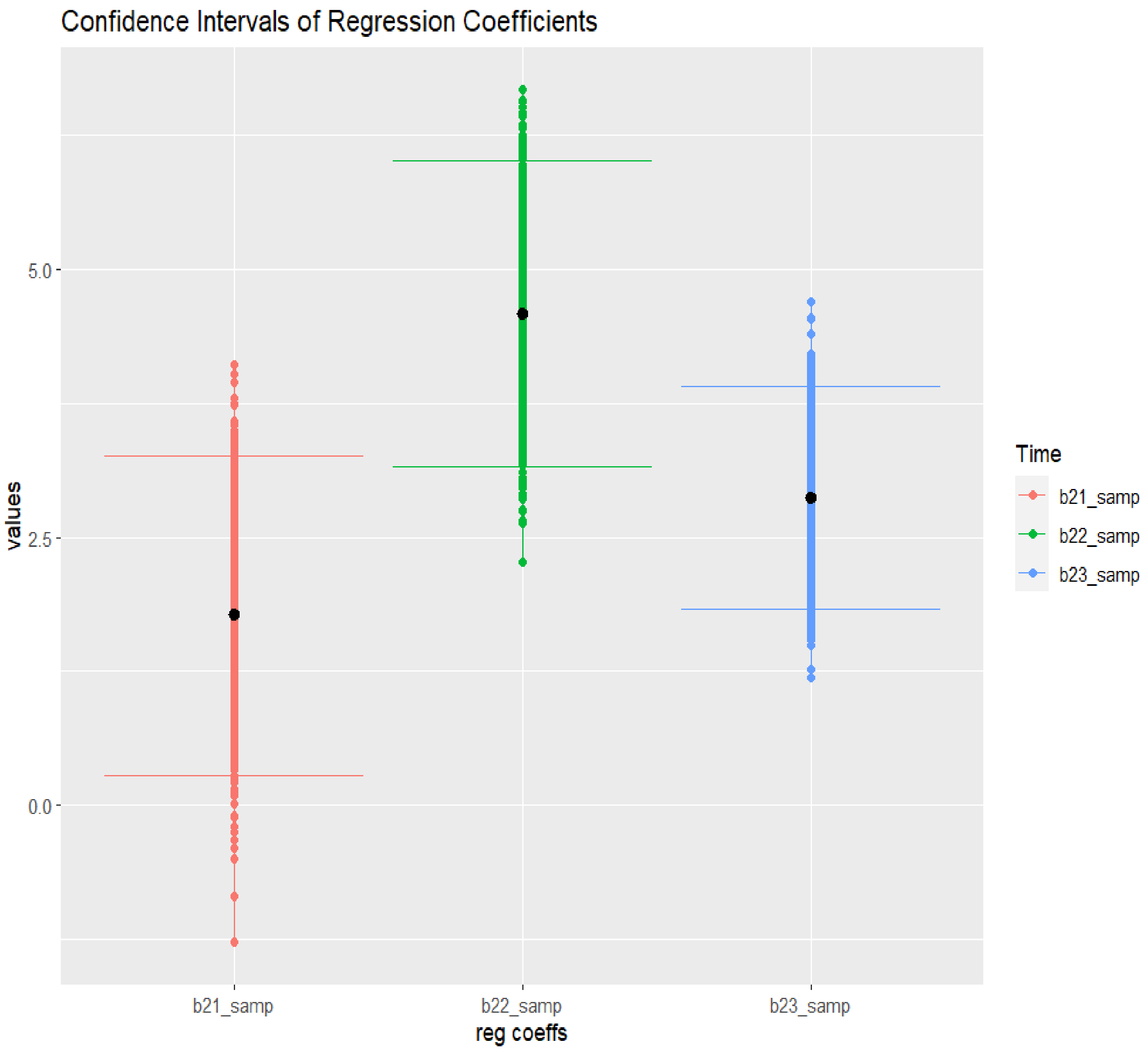
- All of them have equal spread around their means.
- The box represents the spread of the middle 50% of the sample and there is no overlapping region among the distributions.
- The line that divides the box into 2 parts represents Median or the middle most value. Being a measure of central tendency, it indicates on an averageDiet-3 is performing better over time.

95 % Confidence Interval

Summary of the plot :

If we run the experiment of getting samples 100 times then 95 of the times,

- we can say that their mean values are different from each other though there will be small overlapping regions among the values indicating moderately significant difference.
- mean value of the difference of increase in weight between the groups for Diet-1 and Diet-3 will be significantly larger than that of others.



Testing of Hypothesis

H0(null hypothesis) : Increase in weight of the chicks having Diet-2 is same as that of the chicks getting Diet-3 ; b21 = b22

H1 (alternative hypothesis) : Increase in weight of the chicks having Diet-2 is less than that of the chicks getting Diet-3 ; b21 < b22

Output : Welch Two Sample t-test

t = -84.467, df = 1993.7, p-value < 2.2e-16

H0(null hypothesis) : Increase in weight of the chicks having Diet-4 is same as that of the chicks getting Diet-3 ; b23 = b22

H1 (alternative hypothesis) : Increase in weight of the chicks having Diet-4 is less than that of the chicks getting Diet-3 ; b23 < b22

Output : Welch Two Sample t-test

t = -60.6, df = 1823.6, p-value < 2.2e-16

In both the cases p-value < 0.05 hence H0 get rejected , i.e over time weight increase in case of Diet-3 is more than other diets.

Conclusion

Each one of the methods is indicating the fact that Diet-3 has significantly better impact on gaining weight of the chickens hence in growth.

So based on this analysis we can conclude that our data supports the fact that Diet-3 would be right for chicken.

One can consider ANOVA Model as an alternative approach.