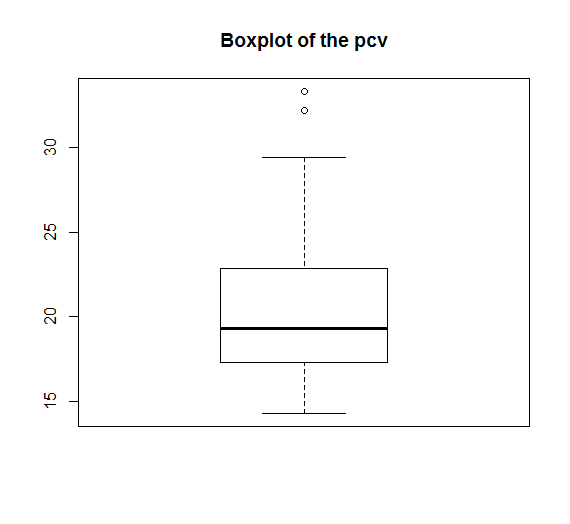
Summary statistics

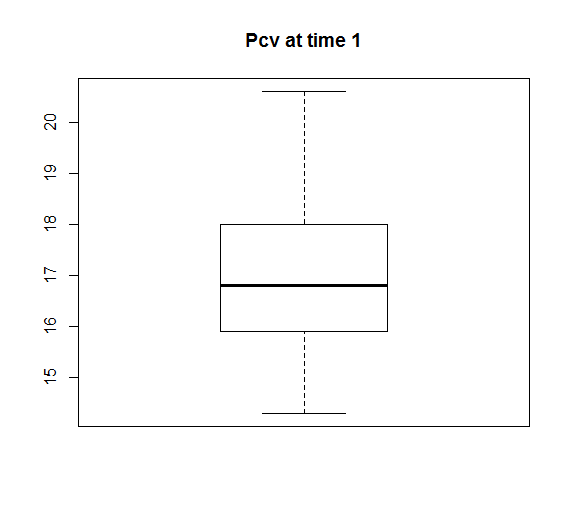
In this part the aim is to have some insights about our data. By performing a basic statistical analysis some relations between the explaining variables and the response one will emerge. This way we will have a first idea on how we can explain the response variable and also some answers when it comes to knowing the most efficient dose.

We will try so see the effect of each variable on the response variable in order to have ideas on the future results.

First we need to know basic information about the data. The response variable, the pcv, can be summarized by the following boxplot.



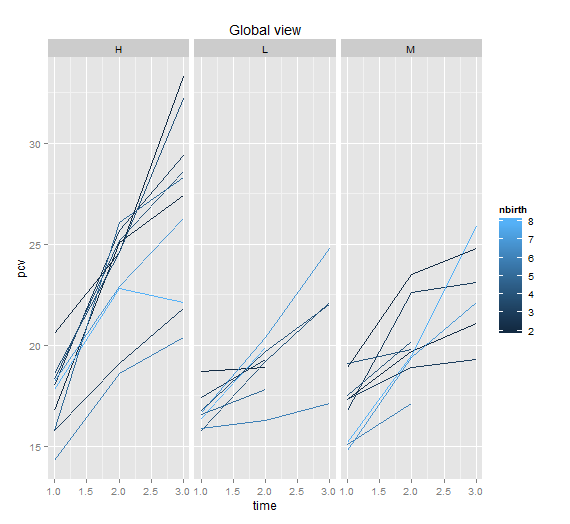
In our dataset we only have sick cows therefore the value of the Packed Cell Volume are low. A “regular” cow should have a pcv between 24 and 46%. Here the mean is slightly above 20% (20.44%) but the median is below as we can see on the boxplot. Pcv values after treatment were included in this boxplot; this is why the upper part is wider. If we only take into account the pcv values before treatment then we have a different distribution.



Here the mean drops to 17.6% and the median to 16.8. We already show all the positive impact of the treatment. This result was of course expected.

Now we will try to have some insights based on a global approach.

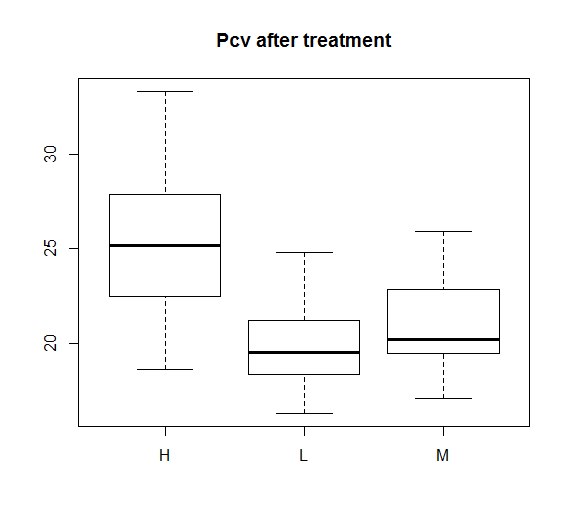
We can have a global idea on how interact the response variable and the other one in the following graphic.



This graphic gives us much information. The first one is the main difference between the types of dose. It seems that cows treated with the high dose tend to have their pcv value higher than the other ones. One could also think that time has an influence on the pcv in the way that the effect of the second dose seems to be less important than the first one. When it comes to the effect of nbirth this graph does not provide us with useful information or at least an idea on whether this variable has a positive effect on the response variable.

Influence of the kinds of dose.

We expect to find that the high dose is more efficient than the medium and the low one. In order to verify if this is true we first plot the boxplot of the pcv depending on the dose for the values after treatment, that is to say for time 1 and 2.



Whereas there is not a big difference between the low and the medium dose, the pcv values obtained after the injection of a high dose are quite superior to the other ones. The previous boxplot summarize the values taken by the response variable pcv based on which dose was used.

The means for each dose, high, medium and low are respectively 25.23, 21.13 and 19.79. We can clearly see that pcv values are higher when the cows are treated with the high dose.

A linear regression model on the dose reinforces the idea that the high dose has a positive effect on the pcv when compared to the other doses. Indeed the estimated coefficients for the medium and the low dose are both negative which means that these doses are less efficient when it comes to increase the pcv value.

Influence of the time.