STAT 7995 Final Report Genetic Mutation in Mice

Dominic Matriccino

2022-12-02

1 Executive Summary

This mouse experiment is interested in whether a mutation that causes inflammation has an effect on the tibia size of mice. Mice in this study either had the mutation or did not, and there were two groups with distinct genetic backgrounds from which we could compare. After running the analysis, it is clear that the mutation is highly significant. The presence of the mutation is highly associated with shorter tibial lengths. With respect to the two genetic backgrounds, there are significant differences between them. Mice from the A background are generally larger when controlling for the mutation. Gender was also determined to be insignificant in this analysis. Lastly, when conducting a pairwise t-test to determine if the means of the four unique strains are significantly different, we found that all strains are significantly different except for A1 and B2, when applying a Bonferroni adjustment. We can say with confidence that this mutation is highly associated with smaller tibias in mice. We cannot say that the mutation causes the tibia to not grow, but rather that the inflammation caused by the mutation somehow causes side effects that ultimately affect the development of the mice. Further research should be conducted to determine exactly what is causing the deficiency in development of the mice due to this mutation.

2 Introduction

2.1 General Background

Mice are the most common animal model for studying human disease. However, current mouse models of immune diseases are created using mice with endogenous Vitamin C production. It is unknown whether endogenous vitamin C has the same effect on the immune system as exogenous vitamin C. While differences exist between humans and mice, the genomes of the two species are very similar and many genes/pathways are well conserved. The data from the following analysis comes from a larger study regarding endogenous vitamin C production of mice, but our data comes from a subset of the study where this area of interest was not considered. The scientists created a genetic mutation within certain mice that required that the mice obtain vitamin C through their food. This portion of the study is highly confidential, but fortunately there was still some data that was collected that did not include this specific mutation that we were able to use. The data on the mice we have either had no mutation at all or had a mutation that causes excessive inflammation to occur. 4 month old mice were selected for this study to ensure that only mature and healthy mice are being experimented on. Both male and female mice of two distinct genetic backgrounds were selected and within these two backgrounds two strains of mice were created. One strain contained the mutation that causes inflammation to occur within the mice before they reach the age of four months. The other strain did not included this genetic mutation and thus can be considered as a control group. The mutation was created to determine how inflammation affects the overall health of the mice. In our experiment we can see that how this affects the mice by looking at the tibia length after the mice had been euthanized. The hind limbs were removed and stored in ethanol for preservation and then measurements were taken. Tissues and organs were removed and measured for subsequent analyses, but none of the tissue and organ information were available within our subset of the study.

2.2 Objectives

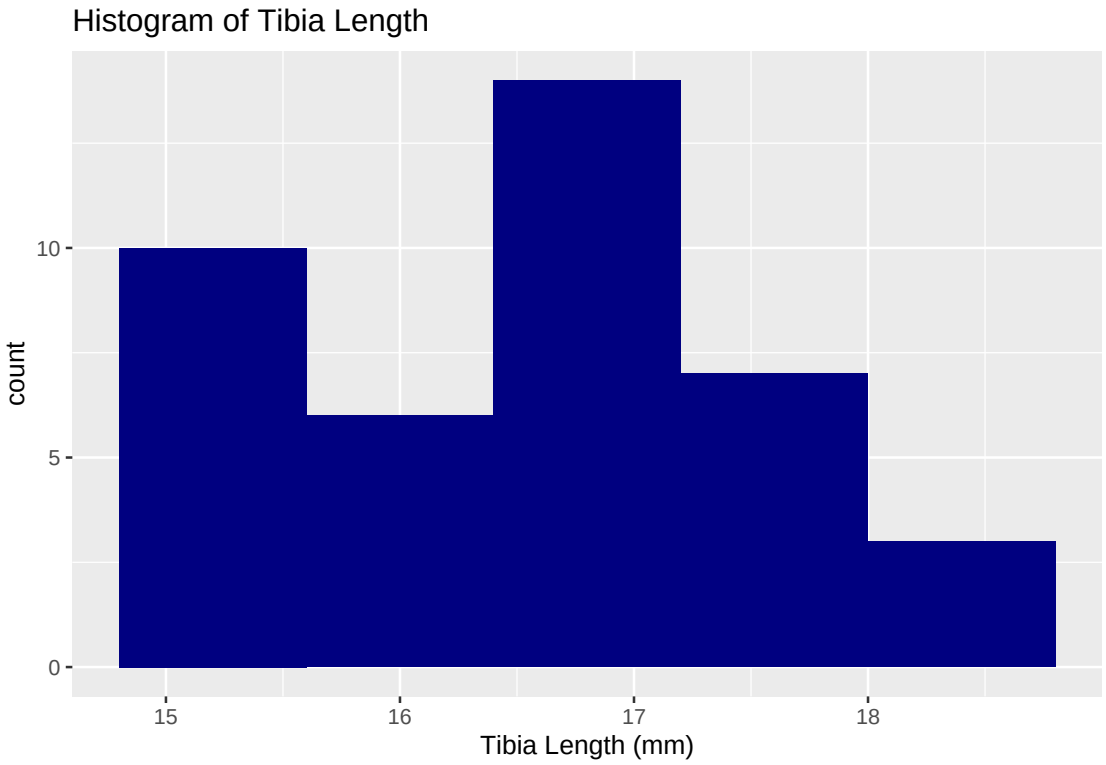
Our project revolves around understanding the nature of this mutation and gender on tibia length in mice. We can break this down into three parts.

1. Find the effects of the strains on tibia length and determine if they are significant
2. Find the effects of gender on tibia length and determine if they are significant
3. Determine if the strains are significantly different from each other.

With this information we can gain insight into whether the mutation that the scientists created affects the health of the mice.

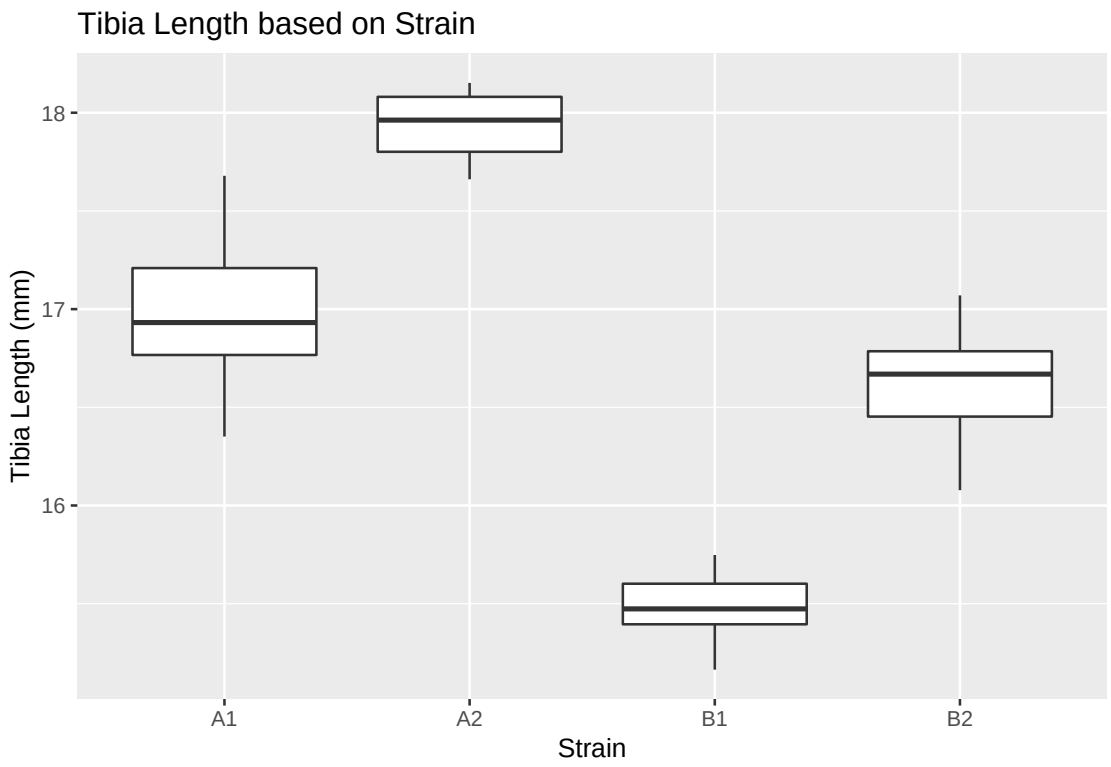
3 Approach to Project

To get a better understanding of our response variable. We can plot a basic histogram that will give us an idea of the distribution of this quantitative variable, tibia length.



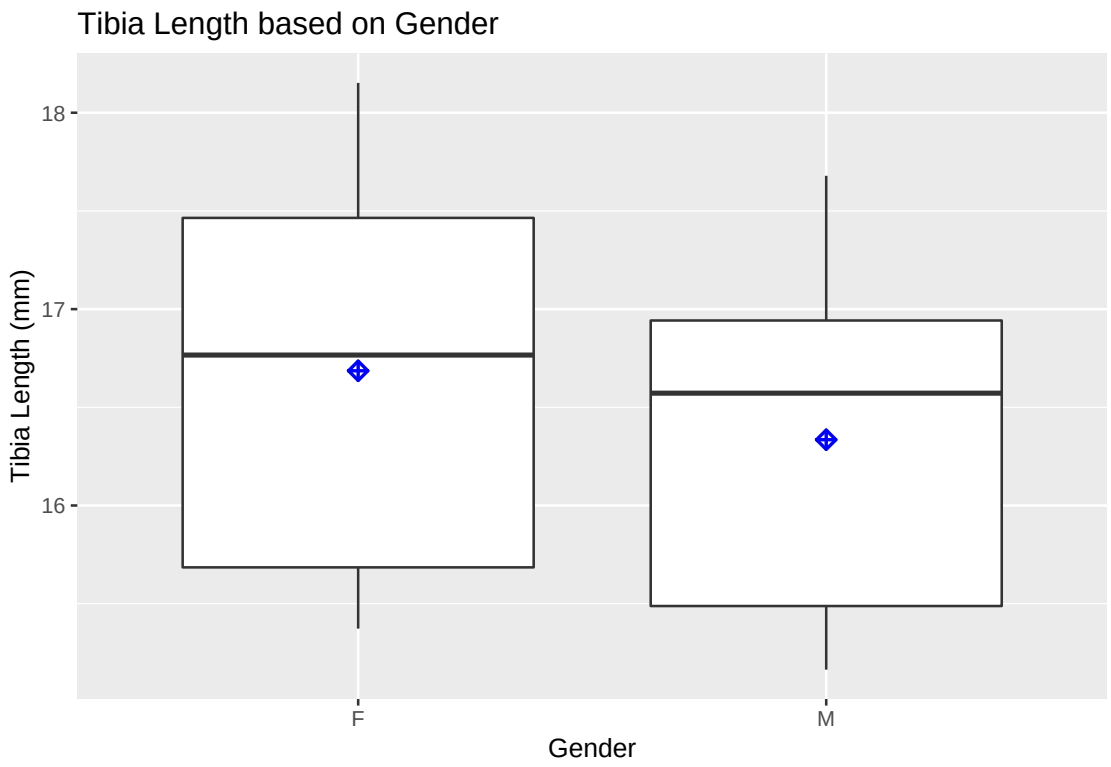
The data does not appear to be normal, although with only 40 observations, it is difficult to obtain a histogram that depicts the true distribution of the data. This data could theoretically still be considered normal, but we cannot say with certainty what distribution the data follows based solely on the current histogram. The observations fall between 15 and 18 millimeters, with the median value being 16.685. Parametric models require that observations come from a normal distribution, so this may be problematic depending on the type of analysis selected.

We can understand tibia length even further by plotting boxplots of the length grouped by the four unique strains of mice.



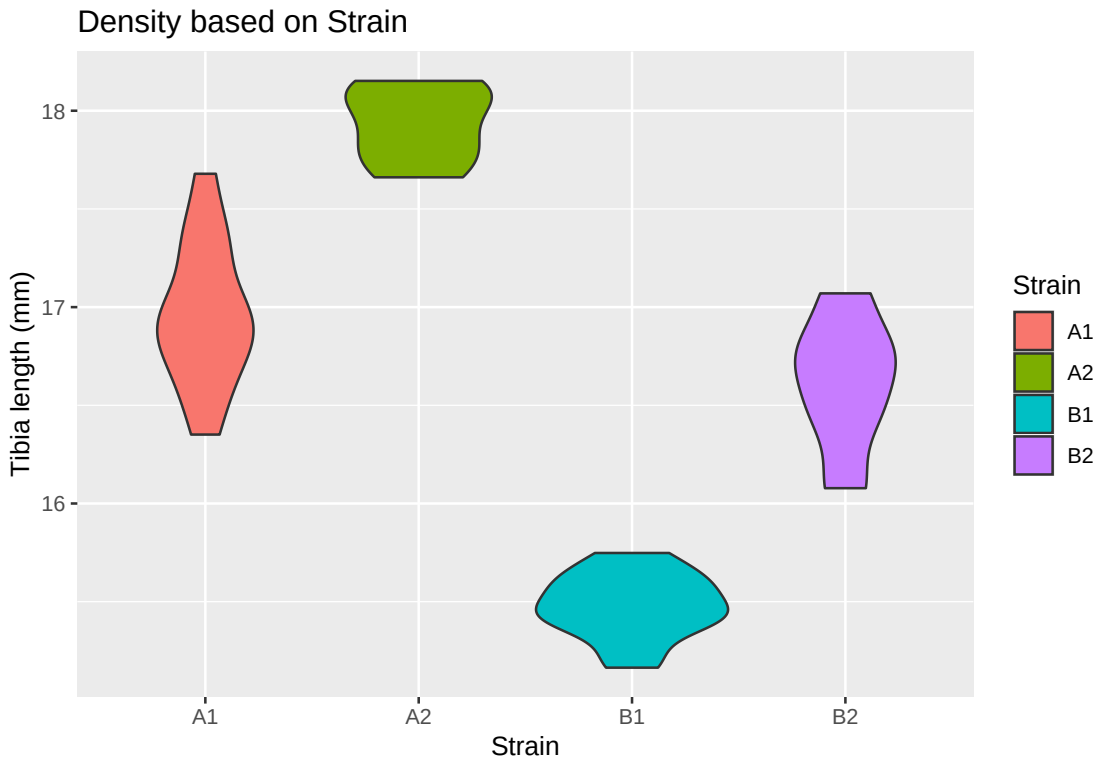
Strain A mice are on average larger than mice from strain B, regardless of whether they have the mutation or not. With respect to the mutation, tibia length is smaller for mice that contain this mutation. This can be seen by looking at the differences between A1 and A2, and B1 and B2. The tibial lengths of A1 mice are all shorter than those of A2. This pattern is also true for B1 and B2. This gives us evidence that the mutation is associated with a smaller tibia.

It may also be useful to examine how tibia length changes depending on the gender of the mouse.



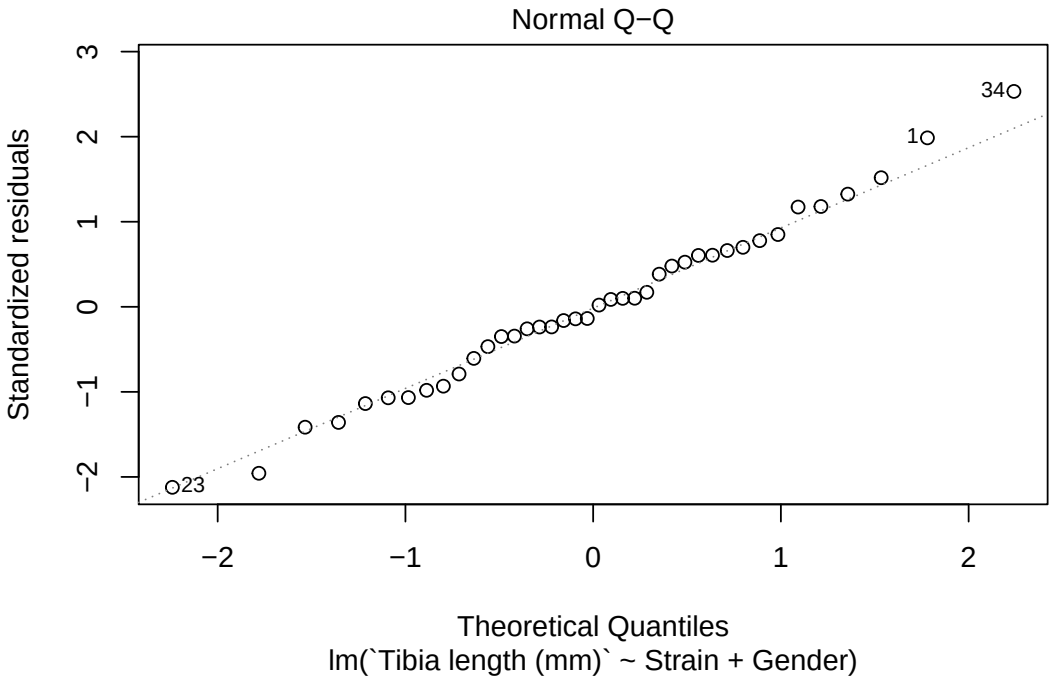
Looking at the boxplots, the difference between male and female mice appears to be minimal. The medians are off by approximately a tenth of a millimeter, and the means are similarly close. We have evidence to show that the gender of mice might not be significant in any type of analysis.

We can consider one final plot before deciding on an analysis. In this plot we can expand on the information within our boxplots broken down by strain to get an even better depiction of the data. Each density plot has a distinct shape across both genetic background and mutation. This again gives us evidence that the mice from each of the four strains provide significant variation.



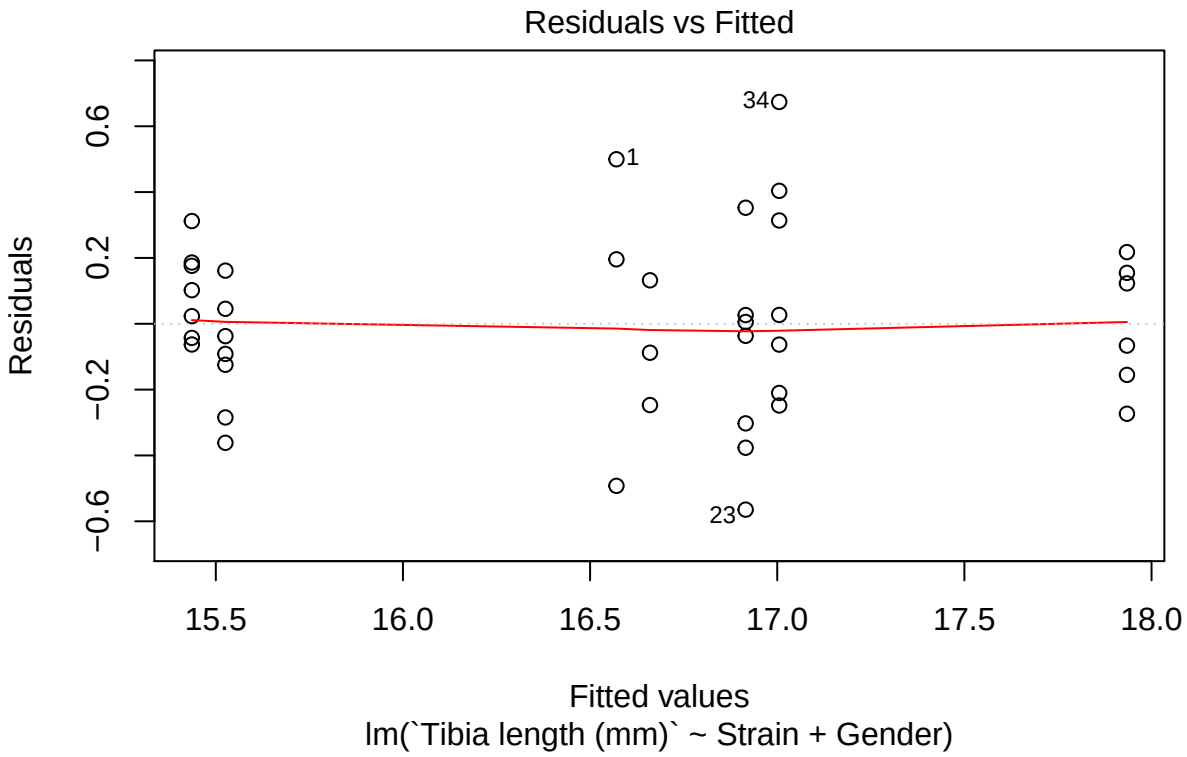
Since we have only two categorical variables and we are interested in finding out if they provide significant variation in our response variable, we can answer this using a linear regression/ANOVA model. In this analysis we can easily estimate the effects of the strains and gender on tibia length. In order to proceed with this analysis we need to check the assumptions of ANOVA. These include normality, equality of variance and independence. Our independence assumption is met because the health of one mouse has no effect on the health of another.

To test for normality we can look at a simple Q-Q plot.



We can see that the observations generally follow the theoretical quantiles of a true normal distribution. There is some deviation in this line especially at the extremes, but given our limited sample size I would say that this reasonably satisfies the normality assumption.

To test for equality of variance we can look at a plot of the residuals.



The residuals in this plot are clearly centered around zero, and there is no clear pattern/trend. It is concerning that the data is not more randomly spread out across this plot. Ideally, there would be no large gaps between the residuals, but since we are working with a limited amount of data, the chance of obtaining the ideal residual plot is low. Since we have some concerns about the model assumptions, we can use a robust method to ensure that our conclusions are still valid.

4 Results

Running our regression model we received the following summary output.

```
##
## Call: rlm(formula = mouse1$`Tibia length (mm)` ~ Strain + Gender, data = mouse1)
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.569915 -0.156293 -0.009116  0.154799  0.724829
##
## Coefficients:
##             Value    Std. Error t value
## (Intercept)  16.9209   0.0906   186.8026
## StrainA2      1.0134   0.1475     6.8691
## StrainB1     -1.4549   0.1078   -13.4948
## StrainB2     -0.3101   0.1392    -2.2280
## GenderM       0.0333   0.0978     0.3399
##
## Residual standard error: 0.2302 on 35 degrees of freedom
```

The summary output shows that strains A2 and B1 are the most significant, with t-values well above the critical value. For strain A2 mice, we would expect that these mice will have larger tibias by about 1.01 millimeters, all else being constant. The B1 strain has the largest magnitude, with an average decrease in tibia length of 1.45 millimeters, all else being constant.

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  mouse1$`Tibia length (mm)` and mouse1$Strain
##
##    A1      A2      B1
## A2 1.3e-07 -       -
## B1 2.4e-15 < 2e-16 -
## B2 0.096   6.3e-09 4.2e-09
##
## P value adjustment method: bonferroni
```

We can perform a pairwise t-test to test if the theoretical population means of each unique strain are significantly different from each other. The results of the test tell us that all strains are significantly different except for A1 and B2. This can be clearly understood by looking at the boxplot of both strains. Since there is a portion of the boxplots that overlap we expect not to receive a significant result when doing a pairwise t-test.

5 Conclusions and Appropriate Recommendations

Our regression/ANOVA model gave us evidence that the strains do provide significant variation in the tibia sizes of the mice. Not only is each unique strain significant, but the genetic background of the mice also proves to be significant. The population means of A1 and B1 mice are significantly different from each other. This comparison allows us to look only at the effects of the genetic background, while holding the effect of whether the mice had the mutation constant. The results suggest that the genetic background of the mice is what is causing some of the variation when holding the mutation constant. On the other hand, when we hold the genetic background constant, we also get a significant result. Strains A1 and A2 are significantly different with a p-value ≈ 0 and strains B1 and B2 are significantly different with a p-value ≈ 0.

The mutation did in fact affect the health of the mice in such a way that their tibias did not develop normally. Correlation does not equal causation, though, so we cannot say that this mutation is entirely responsible for the decrease in tibia length. Since the mutation causes inflammation to develop significantly more than in the mice without the mutation, it likely negatively impacts their overall development, such that they do not grow to be as big as they could be if they were not carrying the genetic mutation.

As mentioned before, we only have 40 observations, which is challenging to work with, especially when conducting analysis that requires certain assumptions. More data could give us more opportunities to conduct a more powerful analysis as well as more confidence in the validity of our findings. This is, of course, expensive and time consuming. When conducting an experiment like this, many resources are depleted. The potential benefits of the findings of our data are also limited. Since we do not have data directly related to the purpose of the original experiment, we cannot fully engage with the topic of the study. As such, our conclusions are most likely not going to result in any sort of advancement in medical understanding. However, it is still important to understand the results and allow them to inform future research and experiments related to this specific mutation.

6 Appendix

These are the linear regression coefficients for a normal linear regression model as well as with a robust linear regression model.

```
##      (Intercept) StrainA2 StrainB1 StrainB2 GenderM
## lmc       16.916    1.019   -1.480   -0.345   0.090
## rlmc      16.921    1.013   -1.455   -0.310   0.033
```

The mean, median, standard deviation and sample size of the four strains are printed below.

```
##                         A1     A2     B1     B2
## Mean                16.960 17.934 15.481 16.615
## Median              16.931 17.962 15.473 16.669
## Standard Deviation  0.361  0.195  0.164  0.344
## n                   14.000  6.000 14.000  6.000
```