

Bayesian Stats Techniques and Models Capstone: Weight Loss Analysis

Simon Thornewill von Essen
Dates: 07/02/2020 - 11/02/2020

Abstract—One challenge with respect to with weight loss can sometimes be understanding if the change in weight over time is real and not due to random chance. In this report, bayesian ANOVA is used to determine whether there is significant weight loss between two months measured using a smart scale to automatically track progress over time. It was found that the means between the two months were very different, producing posterior distributions for the means of the two months that barely overlapped. As such, it has been concluded that significant weight loss was achieved.

I. INTRODUCTION

LATELY, I have been trying to lose weight. One question that I have is whether I have had any significant weight loss over the past two months. In order to evaluate this problem, I have decided to weigh myself using a smart scale to automatically log my progress over the course of Janurary and Feburary 2020. This scale not only measures my weight, but also tracks other metrics such as body fat percentage, bone density, blood protein percentage and so on.

What I did was weigh myself every day, once in the morning and once in the evening. Although your weight tends to be most consistent in the morning and you tend to gain “weight” by the evening, I weighed myself twice to get a better picture of what my average weight for the day was.

Luckily, since this data is generated by my self, there are no issues with the data such as missing values. The only issue was removing units such as “kg” from the values and renaming the columns so that they work better within a coding environment. As such, there were not a lot of challenges with aqiring the data itself.

II. DATA EXPLORATION

Before doing any modeling I wanted to do some quick plots to get a sense of the data that I was interested in. I was mostly interested in how my weight changed over time, which means that I was primarily concerned with this this metric instead of all of the columns that I was able to make for the dataset. In retrospect, it might have been a good idea to drop some of the columns that were estimated by the smart scale. (We’ll see why I think that in the next subsection.)

Note that these plots can be found in the appendix section so that they do not break the flow of the text.

A. Body Fat Percentage vs. Body Weight

One of the things I was most interested in was the change in body fat percentage as a function of Body Weight. We can see in this plot that the relationship between body weight and body fat percentage if roughly linear. This is something that we do not expect because as a person’s body weight decreases, the amount of weight they need to drop to reduce their fat further is smaller and *vice. versa*. The reason why the plot looks linear is because the weight range that I am looking at is much smaller than the range from being extremely under weight to being extremely obese.

B. Month over Month Change in Body Weight

Looking at the month over month change in body weight, we can see that in January there is a large variability in body weights whereas the second month is more consistent, with mean values being weound 81 kg for Janurary and roughly 80 kg in Feburary. It seems like there was meaningful weight loss, given that the boxes themselves do not overlap, but if you include the whiskers then there is still some overlap.

C. Week over Week Change in Body Weight

For the week over week change in body weight we can see the story unfold a little more clearly. We can see that in the 2nd week of 2020 that the variability is quite high, but the weight stabilises a little more from the 3rd week onwards. The difference between weeks 4 and 5 are quite profound, with week 6 increasing again in variability.

It should be noted that during the week with the most weight loss I was using intermittent fasting instead of simple caloric restriction. While I found that it was good at the start, I found that my weight loss plateaued quickly. I think this was because I was losing weight too quickly, causing my body to go into what is known as “starvation mode” to try and retain the body fat that I was losing.

III. MODELING

A. Model Postulation

In order to find out whether the weight loss was meaningful between the two months, I want to do an ANOVA test to see if this was really the case. ANOVAs test for proportions between the intra-group and inter-group variances to see if a categorisation is “meaningful”. This test is appropriate because it tests for data that have categorical and continuous

variables.

ANOVA models assume that data is normally distributed and that each group (in this case, we have two) and that each group has its own mean. Note that it is also possible for each group to have its own variance as well. To start, we will assume that each group has its own variance but we can already see from our exploratory data analysis that this assumption may not actually hold in this case. We are interested in seeing the differences between the posterior distributions of the means.

Thus, using the description above we can write the model likelihood where we assume the variances between groups are the same:

$$y_i | g_i, \mu, \sigma^2 \stackrel{\text{iid}}{\sim} N(\mu_g, \sigma^2)$$

Where the prior distributions for the means and variances are as follows:

$$\mu_j \stackrel{\text{iid}}{\sim} N\left(0.0, \frac{1.0}{10^6}\right)$$

$$\sigma^2 \stackrel{\text{iid}}{\sim} \frac{1}{\text{Gamma}\left(\frac{1.0}{2.0}, \frac{1.0 \times 1.0}{2.0}\right)}$$

For the prior distributions, I chose uninformative priors. Where the uninformative prior for the mean of a normal distribution is also normally distributed with a mean of zero and a very large variance. *Note: JAGS uses a precision parameter instead of variance, which is why the reciprocal was taken.*

Furthermore, the prior distribution for the σ^2 parameter is drawn from an inverse gamma distribution, with a prior sample size of 1 and a prior guess of 1 as well. Note that this is also meant to be uninformative.

B. Model Fitting

Now that the model has been defined, it needs to be fit in R with RJAGS. The code for the model string is as follows:

```
model{
  for(i in 1:length(y)){
    y[i] ~ dnorm(mu[month[i]], prec)
  }

  for(j in 1:max(month)){
    mu[j] ~ dnorm(0.0, 1/1.0e6)
  }

  prec ~ dgamma(1.0/2.0, 1.0*1.0/2.0)
  sig = sqrt(1.0 / prec)
}
```

Where y is equal to my measured weight in kilograms, and month is the month when the measurement was taken. The model was instantiated with three chains and updated for 1,000 iterations before sampling from each of the chains 1,000 times.

C. Convergence Diagnostics

Before analysing the posterior sample to answer the question we set out with, it's a good idea to see whether the chain converged. This will be assessed using a number of diagnostic tools.

- Trace plots of the samples drawn from the posterior distribution.
- Gelman-Rubin diagnostics.
- Autocorrelation diagnostics. (Including visualisation.)
- Effective sample sizes of each parameter.

The trace plots for the analysis can be found in the appendix under fig. 4. As mentioned in the caption, the chains seem to be sampling from the posterior distribution well, without spending a lot of time in one area or wandering around the space.

This is confirmed by the Gelman-Rubin diagnostics, which are printed below:

Potential scale reduction factors:

	Point est.	Upper C.I.
mu[1]	1	1
mu[2]	1	1
sig	1	1

Multivariate psrf

1

From these diagnostics, we can conclude that there is no need to do any thinning of the chains, which is further evidence that the chains are sampling well from the posterior distribution.

Next, let's have a look at the autocorrelation diagnostics:

	mu[1]	mu[2]	sig
Lag 0	1.0000000000	1.0000000000	1.00000000
Lag 1	-0.0006874275	0.005679199	0.06212891
Lag 5	-0.0001819209	-0.012440970	0.02044134
Lag 10	-0.0136230695	0.020004604	0.01804867
Lag 50	0.0207523518	0.030407469	0.00853365

We can see there, as well as what is visualised in the appendix in fig. 5 that there is a lot of autocorrelation for the first iteration, but the chain quickly reduces its autocorrelation to zero. Once again, this is more evidence that the chain is sampling from the posterior distribution well. There are no concerns to be made here.

Finally, we want to have a look at the effective sample sizes:

```
mu[1]      mu[2]      sig
3000.000 3000.000 2648.149
```

We can see here that for the means of the groupings, have a 100% effective sample size. Note that we have three chains of 1,000 samples, which we combine into one large sample of 3,000 observations. However, we should note that the standard deviation sample size is roughly 88%. We already noted that it is unlikely that the standard deviations between the group distributions are the same based on our visualisations, and this confirms that hypothesis.

If we wanted to improve our model, we would consider using separate parameters for the standard deviation for each group. We could confirm that this is really an improvement by looking at the DIC diagnostics between the models to see which has the lowest penalty. However, this will not be done here because the effective sample size seems to be large enough to have some information over our driving question.

D. Tests for Assumptions

Plots for the residuals can be found in figures 7 and 8. We can see in both of these plots that the variance between the two groups are clearly not the same, which we already knew. To combat this, we will change our model to have two parameters for the standard deviation. The model gets changed below such that the precision parameter is in side the for loop for each group parameter along with the calculation for the standard deviation:

```
model{
  for(i in 1:length(y)){
    y[i] ~ dnorm(mu[month[i]], prec[month[i]])
  }

  for(j in 1:max(month)){
    mu[j] ~ dnorm(0.0, 1/1.0e6)
    prec[j] ~ dgamma(1.0/2.0, 1.0*1.0/2.0)
    sig[j] = sqrt(1.0 / prec[j])
  }
}
```

The trace plot for this model can be found at 6. We won't run convergence diagnostics because the trace plots look okay and the rest of the diagnostics looked good the first time around.

Next, we want to analyse which model is better using the DIC value from before. For model one we get the following results in table I.

We can see that although we have a higher penalty in our second model since it is more complex, the mean deviance is lower because it has explanatory power. Overall, the penalized deviance for the second model is lower than the first, so we will use this model when doing analysis on the posterior distribution.

statistic	model 1	model 2
mean deviance	158.8	152.7
penalty	3.154	4.197
penalised deviance	162	156.9

TABLE I
RESULTS OF DIC FOR MODELS 1 AND 2. MODEL 2 HAS THE LOWEST PENALISED DIC, WHICH IS EVIDENCE FOR IT BEING THE SUPERIOR MODEL TO NUMBER 1.

IV. RESULTS

When taking the means μ_1 and μ_2 , we can see that μ_2 is smaller 100% of the time. If you view the posterior distributions for the means in a visalisation, then you can see that there is no overlap between them. This means that meaningful weightloss has been achieved. We can also see from the posterior summary below that the weight in the second month becomes more consistent as well, exactly as we expect.

	Mean	SD
mu[1]	81.2049	0.1598
mu[2]	79.9153	0.1692
sig[1]	1.0511	0.1120
sig[2]	0.6237	0.1255

However, the real question is if this finding generalises into further weight loss. For this we would want to wait a longer time, perhaps four months or so and then do another analysis based on those results. Note as well that it would also be good to know what actions caused the weight loss as opposed to analysing whether it has taken place in a meaningful way.

V. CONCLUSION

We set out on this analysis to discover if there has been meaningful change of weight as a result of habit changes. It was found that there was a very high probability that the differences in means were not due to chance alone, with a mean loss of weight of roughly 1.29 kilograms over the time of measurement.

VI. APPENDIX

A. Figures

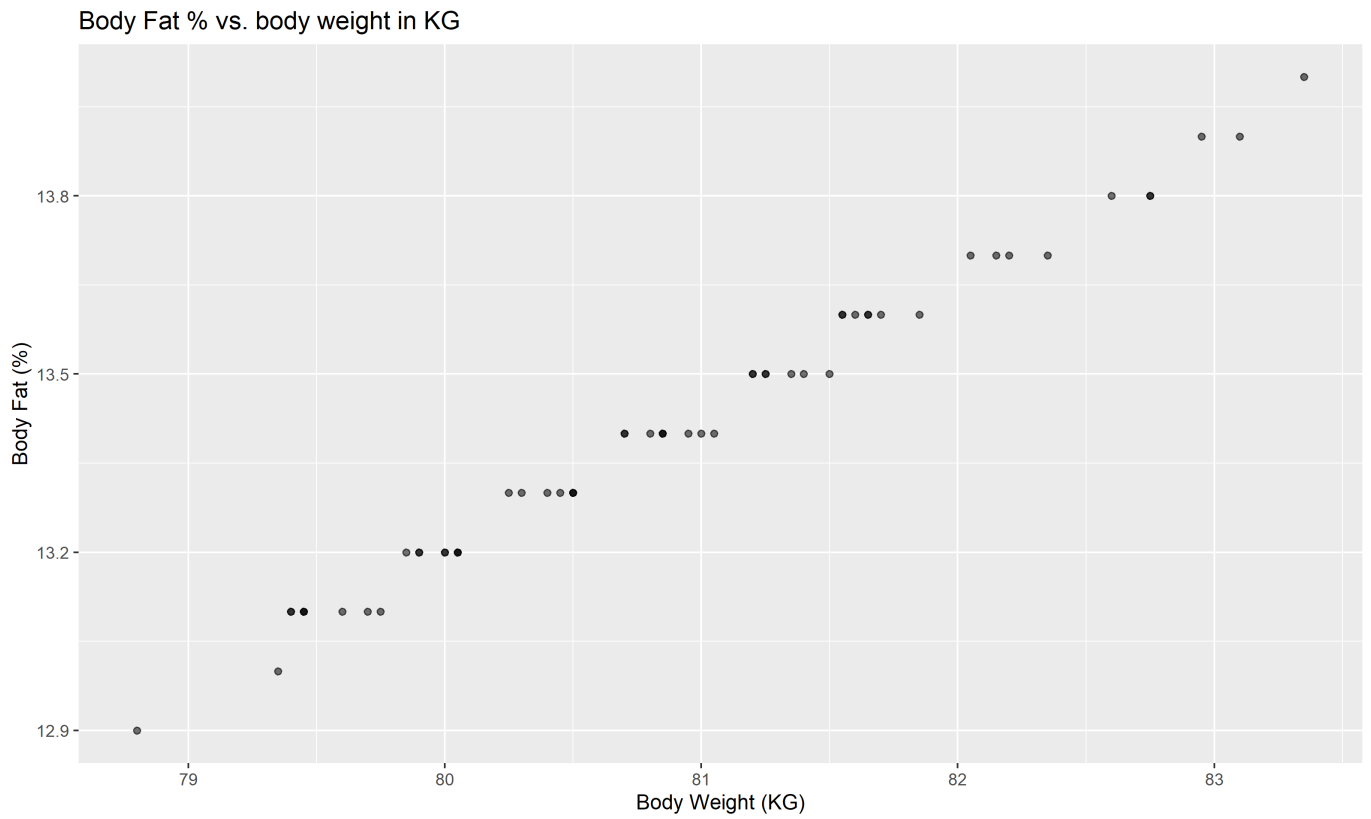


Fig. 1. Scatter plot of two numeric values, body fat percentage versus body weight. We can see a linear relationship between the two variables. However, it is expected that it only looks linear because of the relatively small body weight range, if I were to measure myself as being extremely under/over weight then we would expect the relationship to be non-linear.



Fig. 2. Boxplot showing change in body weight over the two months where it was measured. We can see that there is certainly a change in body weight, but is it meaningfully different?

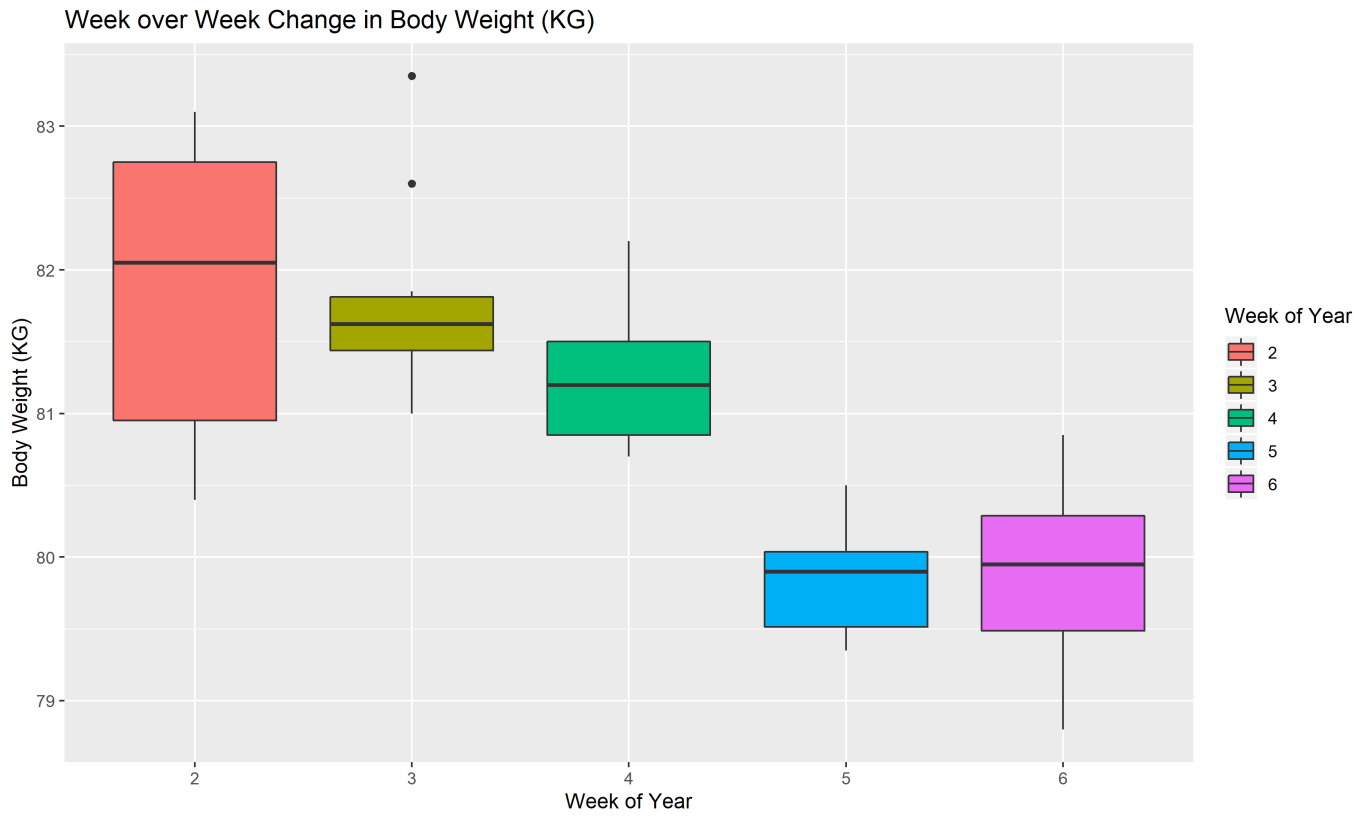


Fig. 3. Boxplot showing change in body weight over the weeks where it was measured. In the first week we can see that there were large variations in body weight, during weeks 5 and 6 the weight is very different from the start at weights 2 and 3.

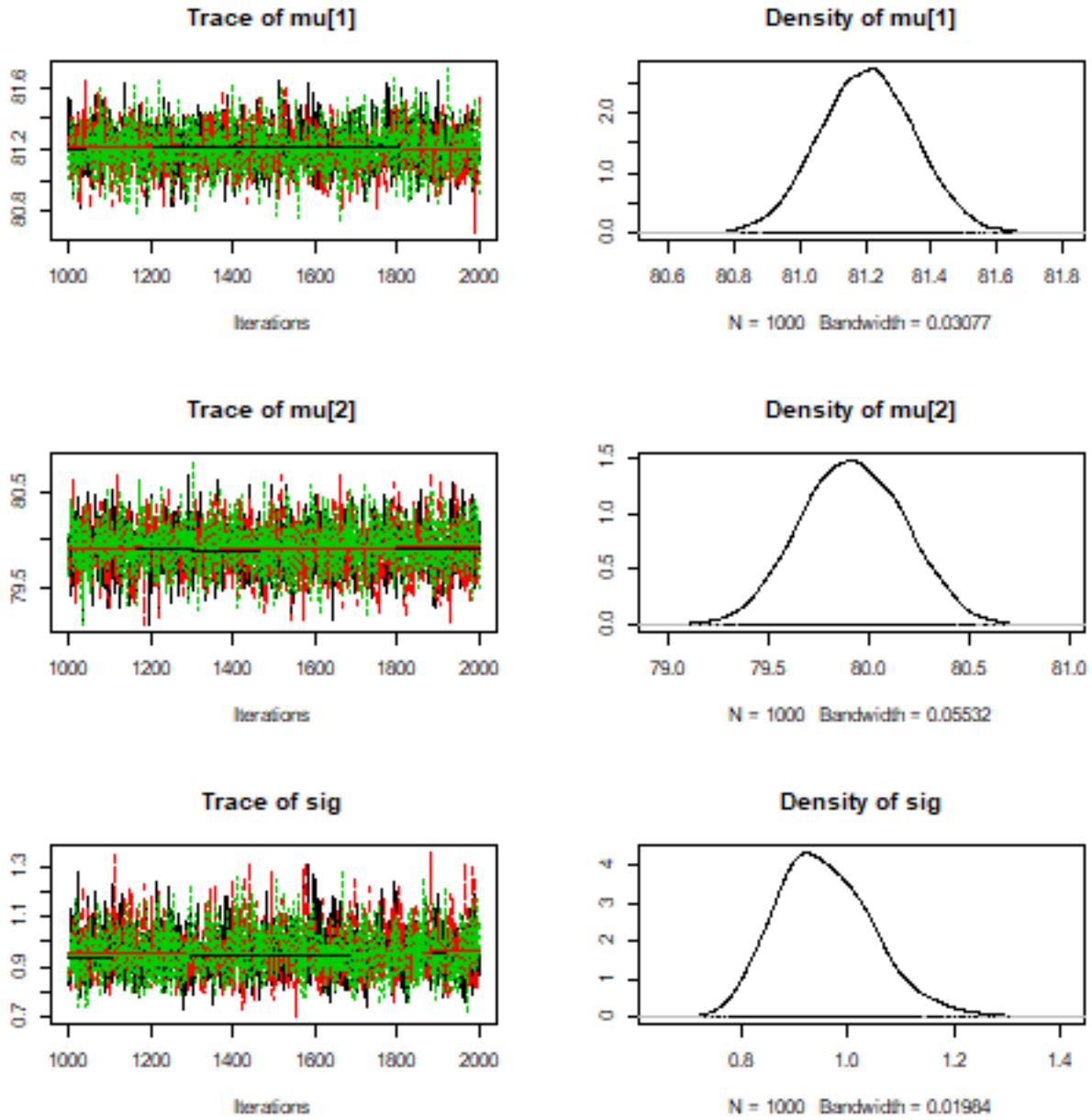


Fig. 4. Visualisation of the MCMC chains for the parameters of the ANOVA test. We can see from the trace plots that the sampler is exploring the space appropriately, without lingering too long around one area while still converging on one value.

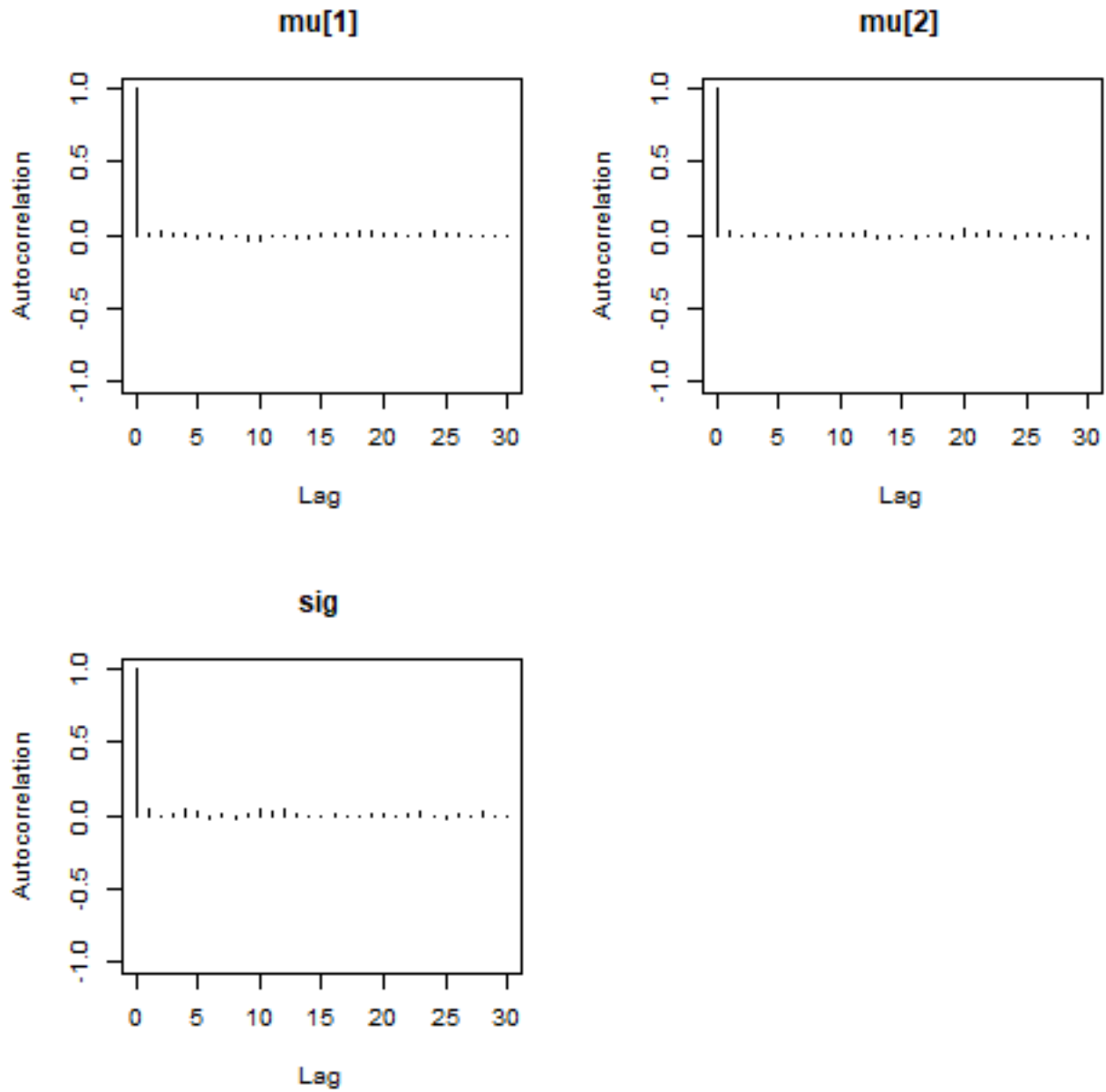


Fig. 5. Visualisation of the MCMC chain autocorrelations. We can see that there is large autocorrelation at the start but it very quickly dies out. There is no cause for concern here.

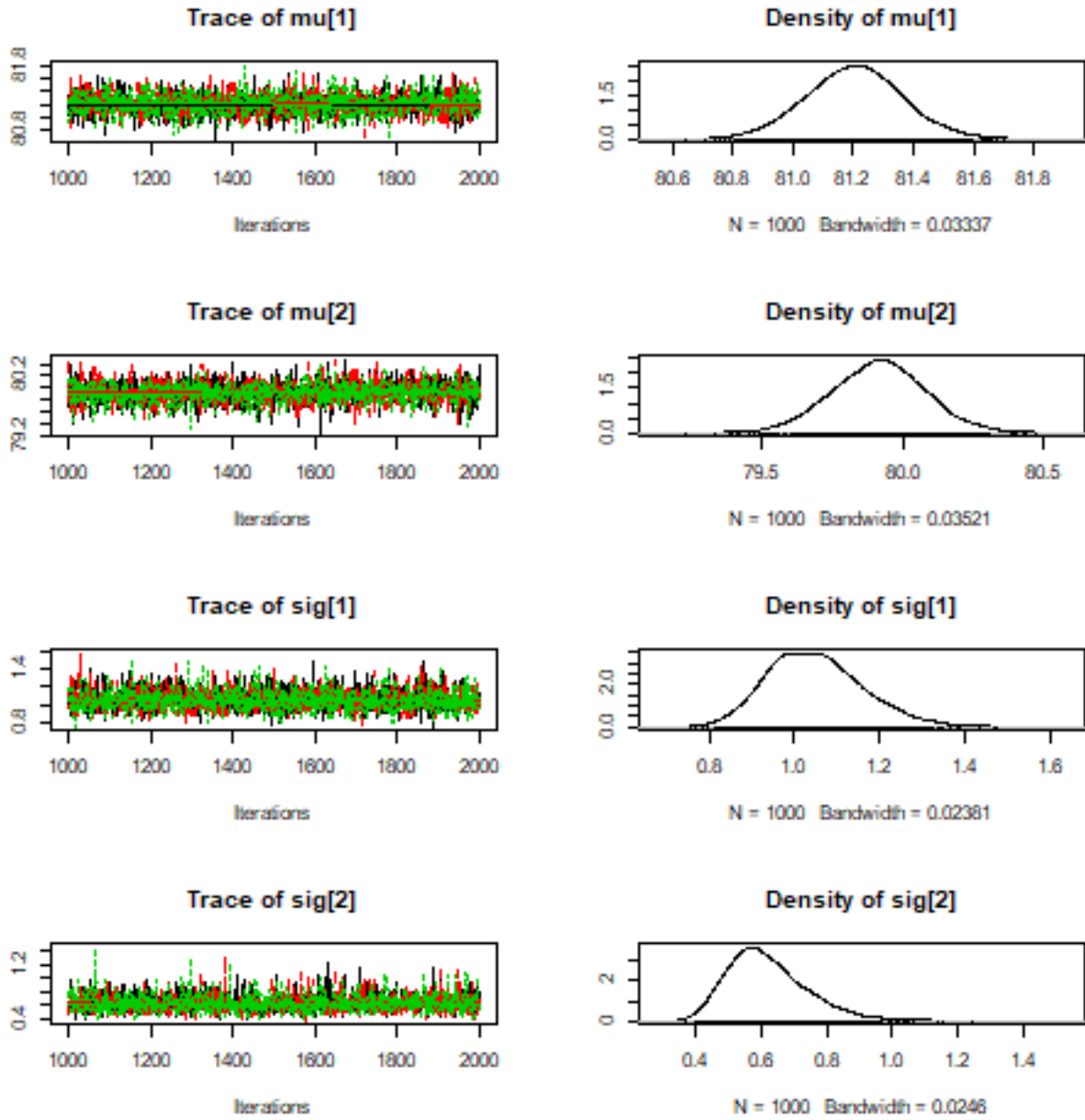


Fig. 6. Visualisation of the MCMC chains for the parameters of the ANOVA test with separate parameters for standard deviation. We can see from the trace plots that the sampler is exploring the space appropriately, without lingering too long around one area while still converging on one value.

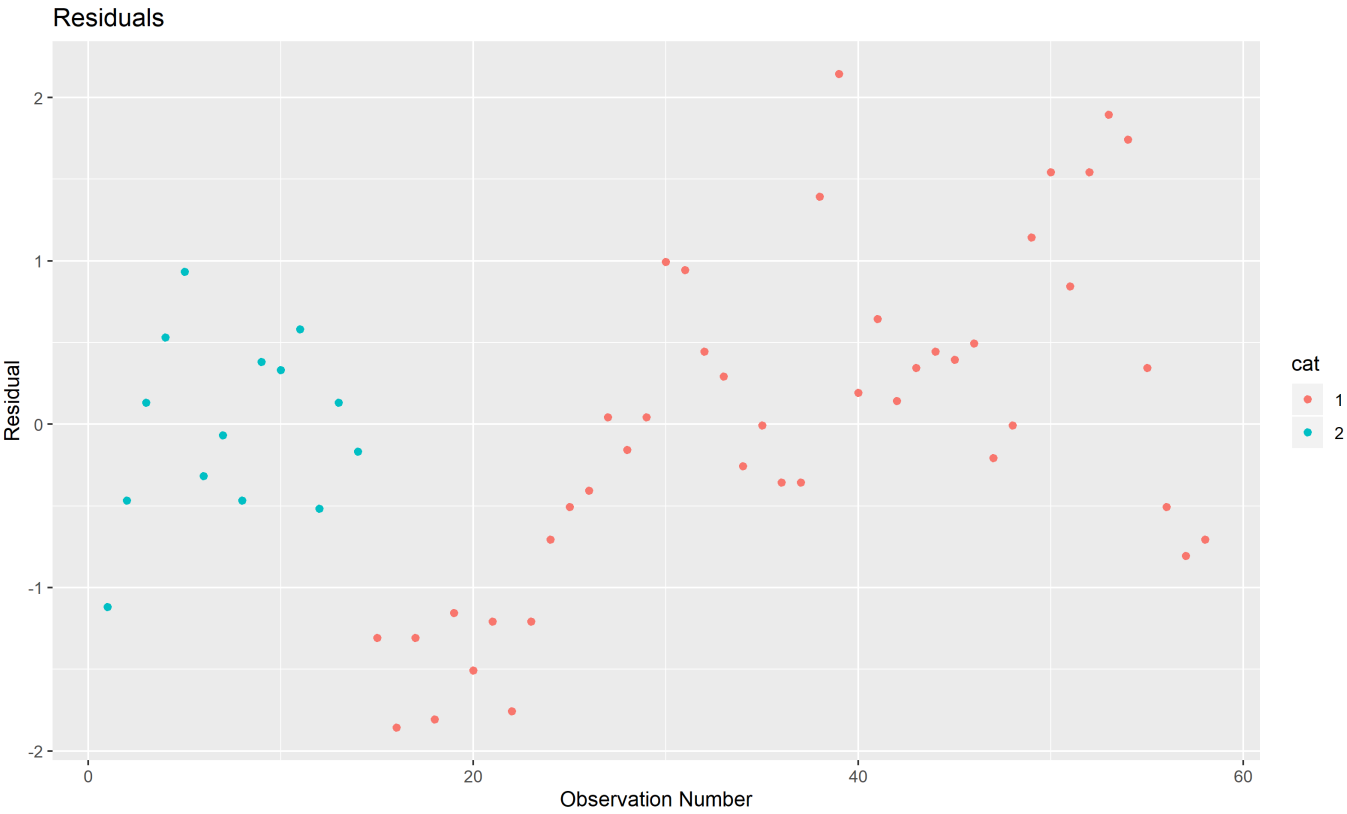


Fig. 7. Plot of residuals, we can see that the two groups show a pattern where the first month has higher variance than the second.

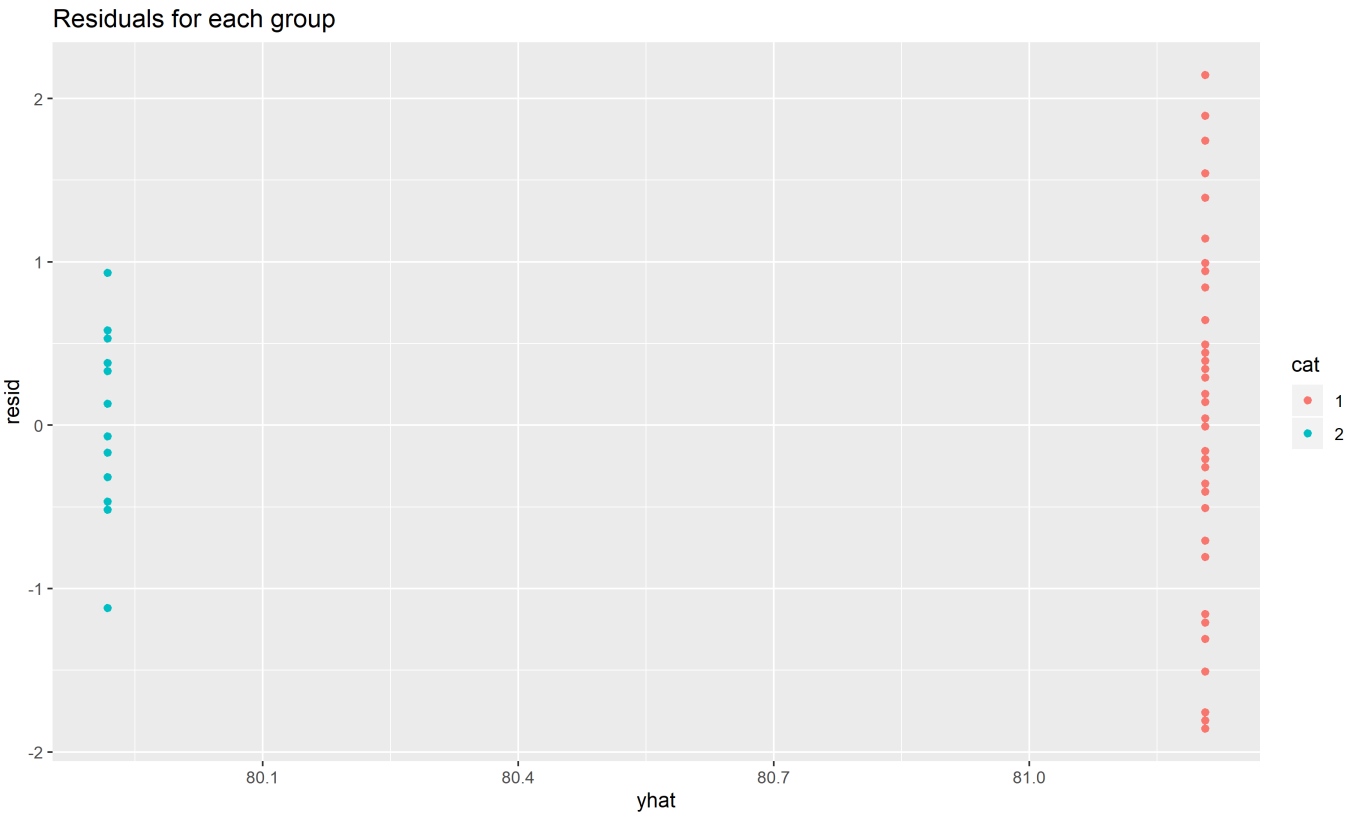


Fig. 8. Plot of residuals, we can see that the two groups show a pattern where the first month has higher variance than the second.

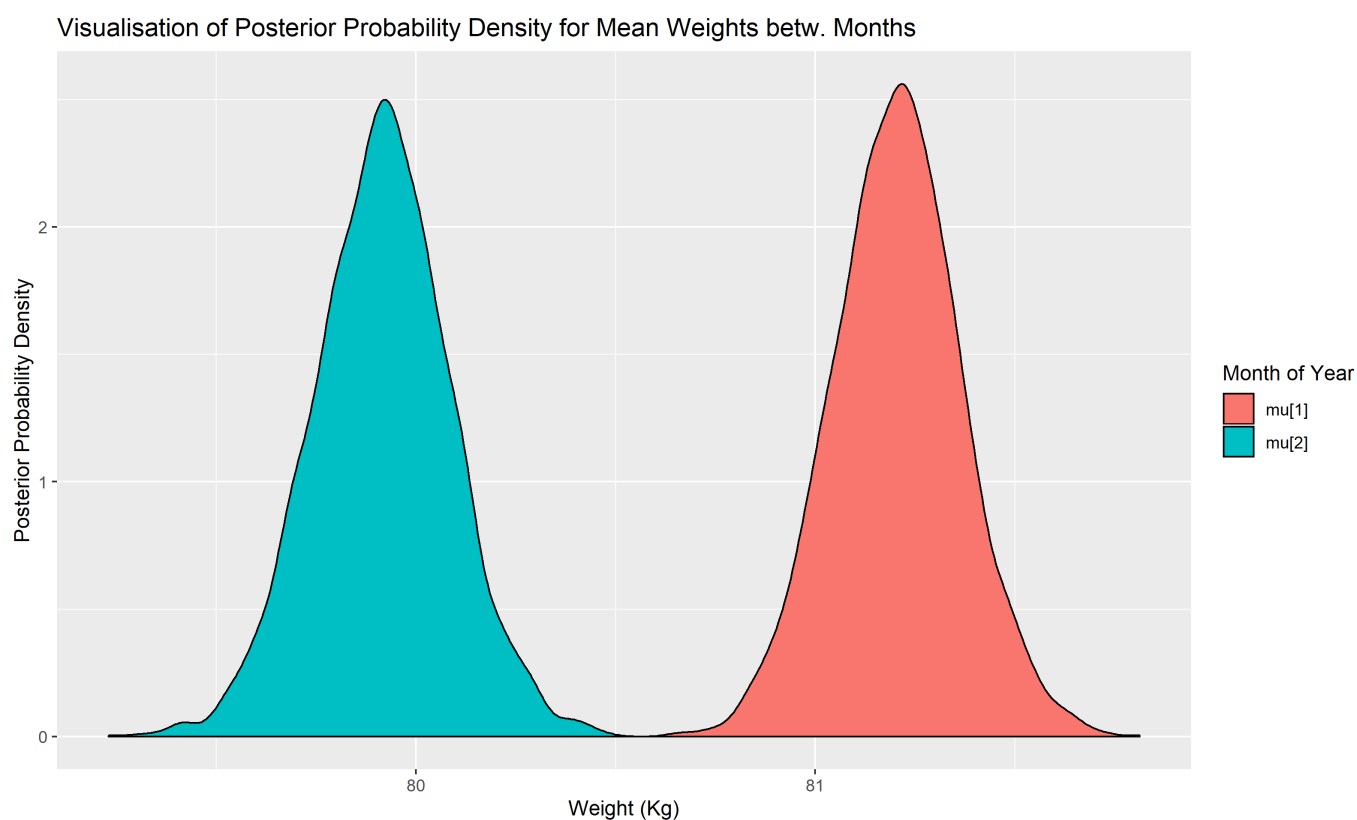


Fig. 9. Density plots for the posterior samples for the first and second means. We can see very little overlap between the two distributions which strongly suggests that the weights between the two months are meaningful.