Stat 602

Group 3: Sungrim Lee, Kyungjin Cho

**Preliminary Report**

**1. Introduction**

For the final project, our group has chosen to experiment with the difference in the driving time it takes from Hilldale to Capitol, Madison, WI. As students in UW-Madison experiencing the traffic every weekday, we were genuinely curious about the time it takes to drive to work in Madison. Since a lot of people with personal vehicles in Madison reside in Hilldale, we set the starting point as Hilldale. We set the destination point as the State Capitol due to the fact that a lot of firms are located there. Our question of interest would be “Will there be a significant difference in the time it takes to drive to the State Capitol from Hilldale depending on the specific time during the weekday?”

**2. Experimental Design**

Our group will set the driving time taken to Wisconsin State Capitol from Hilldale as a response variable (in minutes), and a treatment variable will be a time slot separated by three different times (9 AM, 6 PM, 10 PM). However, on the weekday, the driving time taken can be affected by different days. Hence, in order to remove this controllable nuisance factor, we will block this factor by measuring the driving time every three days (Monday, Wednesday, Friday, but not weekends) per week. To make this experiment feasible, Randomized Complete Block Design will be used.

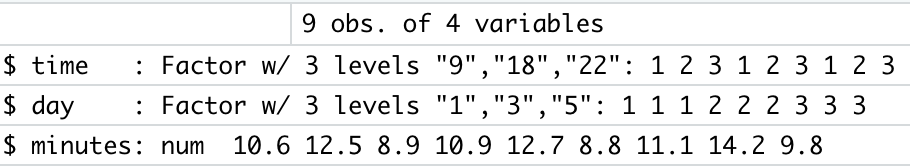
**3. Justification for Chosen Experimental Design**

We decided to use Randomized Complete Block Design since our nuisance factor is controllable, as mentioned above, by blocking this factor by driving on different weekdays. This way, we are able to eliminate the effect on our treatment of interest. Our response variable, driving time, is also quantitative and is possible to measure. Finally, we decided to replicate our experiment so that the precision of the experiment would increase.

**4. Measurement, Model and Analysis**

Measurements were repeated under almost the same conditions. The same person in our group drove the same car with the same route from the same parking lot in front of Target, Hilldale to the State Capitol, not speeded. Fortunately, it wasn’t snowing or raining while driving, which if it was, might have affected the measurements.

The dataset consists of three columns: time, day, and minutes. ‘Time’ is a feature of interest where it indicates three different times (24-hour basis), and ‘day’ is a potential nuisance factor in which each number means a categorical value of three days (1: Monday, 3: Wednesday, 5: Friday). Driving time is measured in minutes and rounded up to one decimal place before implementing it in the ‘minutes’ section. From the observations, we have Randomized Complete Block Design with 3 treatments, 3 blocks, and 1 observation per treatment for each block.



**Model : yij = μ + τi + βj + eij**

*where μ is the grand mean, τ is the treatment effect, β is the block effect*

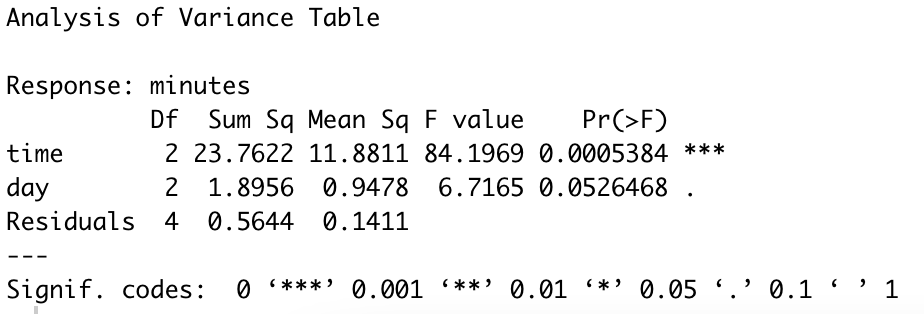
Based on this design, our model is constructed like the above. Because the τ and β indicate deviations from the grand mean, summations of all the elements from τ and β are 0, respectively. Our effect of interest is whether different times affect the driving time taken to the State Capitol from Hilldale, so our hypotheses are:

**H0 : τ1 = τ2 = ···= τa = 0**

**HA : τi is not equal to 0 for any i**

To test the hypotheses, variance decomposition is used to calculate F-statistics.

**5. Interpretation**



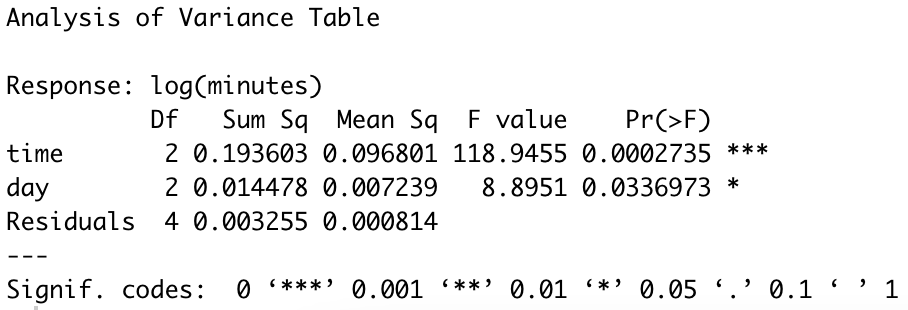
*ANOVA table of the Plain Model*

*(R code details can be found in the last section)*

The screenshot above shows an ANOVA table created from R. Since the p-value for the time variable is 0.0005 which is smaller than 0.05, we reject the null hypothesis and conclude that the driving time from the capital to Hilldale is significantly different from the time of the day one drives a car.

We did not regard the p-value for the blocking factor, ‘day’, due to several reasons. First, we already suspect that the blocks are different. We think that the day of the week will affect our experiment since there tends to be a higher volume of traffic on Friday. Also, since the blocks are not subject to randomization, we would not get a valid result to test the p-value for the ‘day’ factor. However, we still checked if our blocking factor was a good choice. The block is considered to be efficient and effective at reducing noise when . Since our , and , we conclude that the is bigger and concludes that the day of week is an efficient factor reducing the noise for our experiment.

One more thing to consider in RCBD is additivity. In our RCBD model, the effects of should be the same regardless of what we are using, and vice versa. Since this model might not be the best model in a situation where the interaction of ‘days of the week’ and ‘time of the day’ might be present, we included the log transformation model to validate the additivity of RCBD. The model below implements log transformation:



*ANOVA table of log transformation model*

Since the p-value for the time variable is 0.0002 which is less than 0.05, we can conclude again that the driving time from the capital to Hilldale is significantly different from the time of the day one drives a car. From this analysis with log transformation on y, some concerns about non-additivity were eliminated.

**6. Conclusion**

Our group analyzed the RCBD method to test the effect of a difference in the time of the day one drives a car from Hilldale to the Capitol. With the significant p-value of our ANOVA analysis, we conclude that there exists a significant difference in the time of the day one drives a car to work. We also found that our blocking factor is efficient and effective in reducing noise for our experiment. Furthermore, we analyzed the log transformation model to eliminate the non-additivity and got a small p-value, concluding with the same result as before. For future improvement in our analysis, we are trying to replicate the experiment once or maybe twice, if the time allows. In that way, we will make the experiment more feasible and precise.

**7. R code**

#reading the csv data

df<-read.csv('drivingtime.csv')

#setting time and day variable as a factor

df$time <- as.factor(df$time)

df$day <- as.factor(df$day)

#creating the ANOVA table

mod <- lm(minutes ~ time + day, data = df)

anova(mod)

#creating the log transformation ANOVA table

log\_mod <- lm(log(minutes) ~ time + day, data = df)

anova(log\_mod)