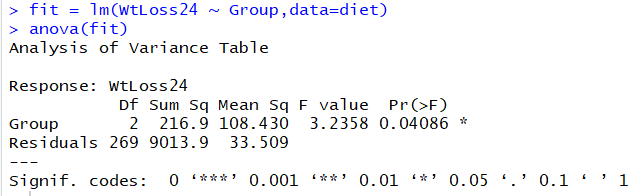
**Stats HW 6**

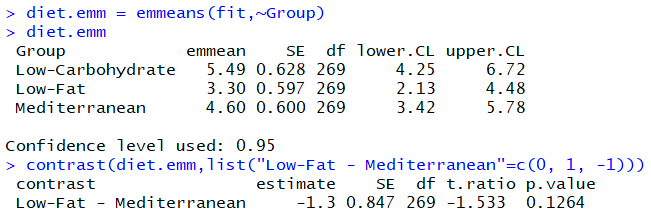
**Blake Mitchell**

1a.) H0: µ1 = µ2 = µ3 versus HA: µ1, µ2, µ3 are not all equal where µ1, µ2, µ3 correspond to the mean kg lost per subject after 24 months for low carb, low fat, and Mediterranean diets, respectively.



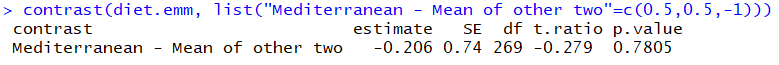
The F-value is 3.2358 and the p-value is 0.04086. The numerator df is 2 and the denominator df is 269. We have some evidence to suggest that not all diets resulted in the same mean kg lost per subject after 24 months.

1b.) H0: µ2 = µ3 versus HA: µ2 ≠ µ3 where µ2 and µ3 correspond to the mean kg lost per subject after 24 months for low fat and Mediterranean diets, respectively.



The estimate of the contrast is -1.3 with a standard error of 0.847. The r-ratio is -1.533 with 269 df. The p-value is 0.1264 so we have no evidence against the null hypothesis. We have no evidence to suggest that the mean kg lost per subject after 24 months differs between the low fat and Mediterranean diets.

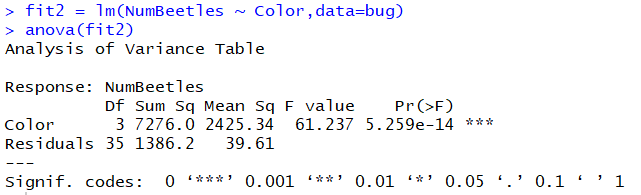
1c.) H0: µ3 = µ4 versus HA: µ3 ≠ µ4 where µ3 is the mean kg lost per subject after 24 months for the Mediterranean diets and µ4 is the mean kg lost per subject after 24 months for the average of the low carb and low fat diets.



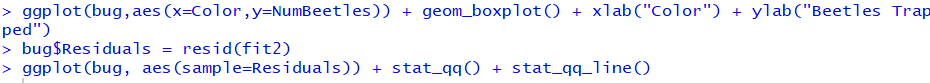
The estimate of the contrast is -0.206 with a standard error of 0.74. The r-ratio is -0.279 with 269 df. The p-value is 0.7805 so we have no evidence against the null hypothesis. We have no evidence to suggest that the mean kg lost per subject after 24 months differs between the Mediterranean diets and the average of the low fat and low carb diets.

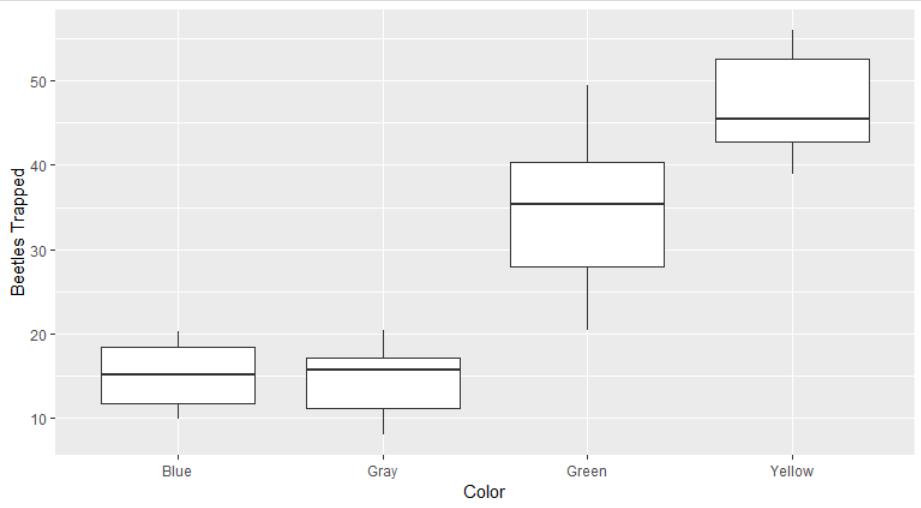
1d.) Dropout would be a concern because it would increase the likelihood of a false positive for the p-value. Those that don’t think the diet is working would drop out and a majority of the people who are left think the diet works, thus lowering the p-value and increasing the likelihood of getting a significant p-value.

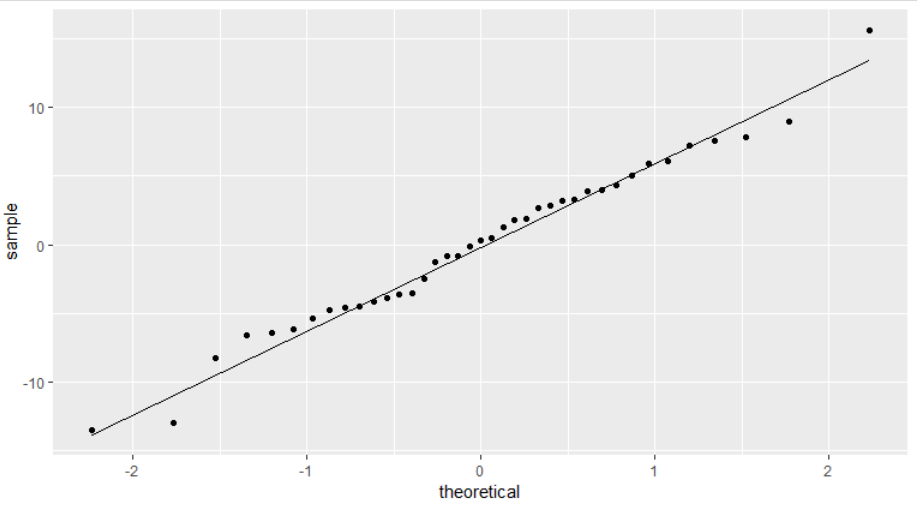
2a.) H0: µ1 = µ2 = µ3 = µ4 versus HA: µ1, µ2, µ3, µ4 are not all equal where µ1, µ2, µ3, and µ4 correspond to the mean number of trapped beetles for blue, gray, green, and yellow boards.



The F-ratio is 61.237 with 3 and 35 degrees of freedom. The p-value is 5.259e-14 so we have very strong evidence against the null hypothesis. We have very strong evidence to suggest that the mean number of trapped beetles is not the same for blue, gray, green, and yellow boards.

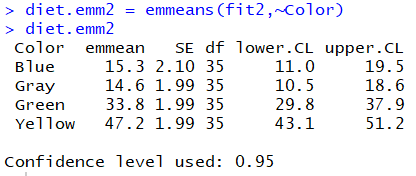
2b.) 

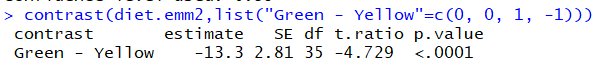




The observations within and between groups are independent so that assumption is met. The homoscedasticity assumption is not met because the widths of the boxes and lengths of the whiskers are not the same across the four groups. The data appear to fall along the straight line so the normality assumption is met.

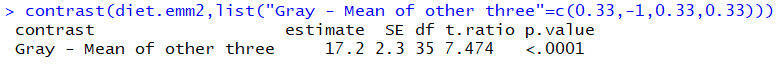
2c.) H0: µ3 = µ4 versus HA: µ3 ≠ µ4 where µ3 and µ4 correspond to the mean number of trapped beetles for green and yellow boards, respectively.





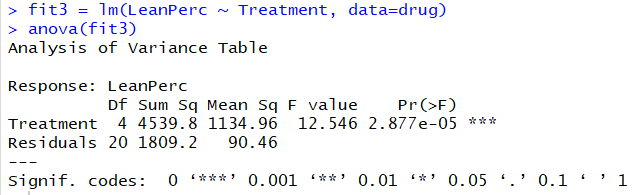
The estimate of the contrast is -13.3 with a standard error of 2.81. The r-ratio is -4.729 with 35 df. The p-value is less than 0.0001 so we have very strong evidence against the null hypothesis. We have very strong evidence to suggest that the green board captures fewer bugs on average than yellow boards.

2d.) H0: µ2 = µ5 versus HA: µ2 ≠ µ5 where µ2 and µ5 correspond to the mean number of trapped beetles for gray boards and the average of the other three colors, respectively.



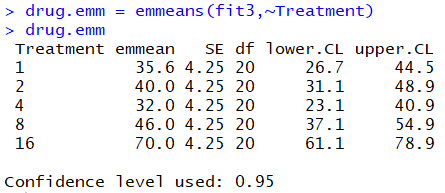
The estimate of the contrast is 17.2 with a standard error of 2.3. The r-ratio is 7.474 with 35 df. The p-value is less than 0.0001 so we have very strong evidence against the null hypothesis. We have very strong evidence to suggest that the gray board captures more bugs on average than the average of the other three colored boards.

3a.) H0: µ1 = µ2 = µ3 = µ4 = µ5 versus HA: µ1, µ2, µ3, µ4, µ5 are not all equal where µ1, µ2, µ3, µ4, µ5 correspond to the mean amount of drug contained in the solution for treatments 1, 2, 3, 4, and 5, respectively.

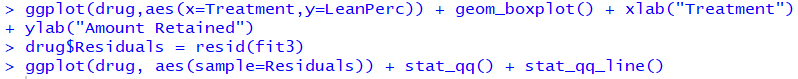


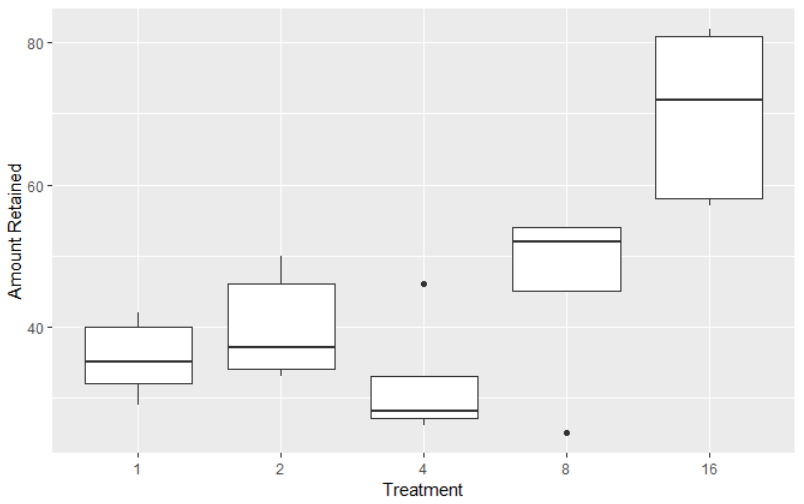
The F-ratio is 12.546 with 4 and 20 degrees of freedom. The p-value is 2.877e-05 so we have very strong evidence against the null hypothesis. We have very strong evidence to suggest that the mean amount of drug contained in the solution for treatments 1, 2, 3, 4, and 5 are not all equal.

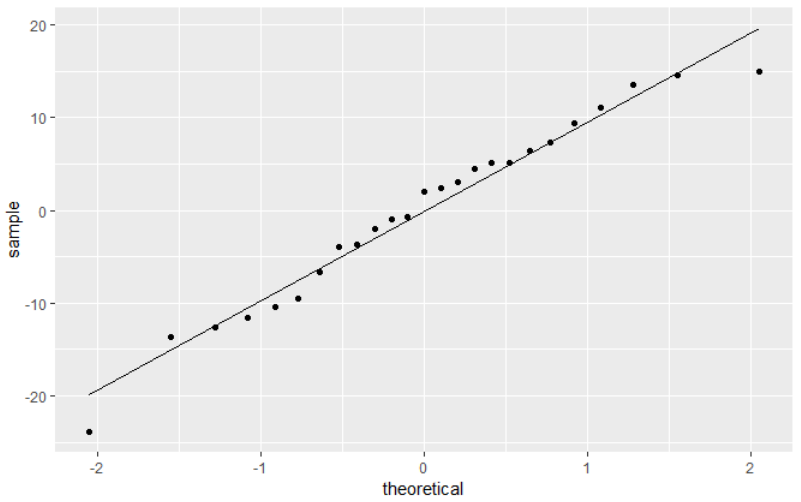
3b.)



There are no instances where the ratio of SD’s is greater than 3 so the standard deviation assumption is met. The sample size is the same across all 5 groups so it is expected that the SD’s will be the same.

3c.) 





The widths of the boxes and lengths of the whiskers are not the same across all groups so the equal variance assumption is violated. There is also some curvature in the QQ plot so I would say that the normality assumption is also violated.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | DF | SS | MS | F | p-value |
| Between | 3 | 5.42 | 1.81 | 5.97 | 0.0006 |
| Within | 226 | 68.57 | 0.3034 | – | – |
| Total | 229 | 73.99 | 0.3231 | – | – |

4a.)

229 = N – 1, N = 230

g = 4

between DF = g – 1 = 4 -1 = 3

within DF = N - g = 230 – 4 = 226

Sp = =

Sp = 0.3034 = MSE

SSRes = MSE(N-g) = 0.3034(230-4) = 68.57

SSBet = SSTtot-SSRes = 73.99 – 68.57 = 5.42

MSBet = SSBet/(g-1) = 5.42/(4-1) = 1.81

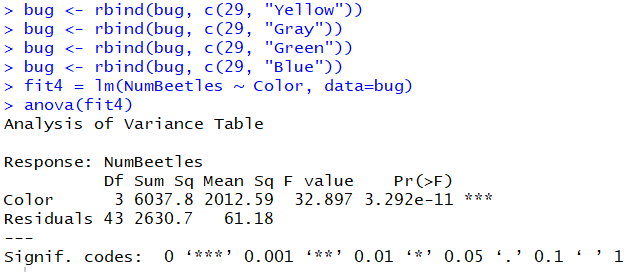
F = MSBet/MSE = 1.81/0.3034 = 5.97

MSTotal = SSTotal/df Total = 73.99/229 = 0.3231

4b.) H0: µ1 = µ2 = µ3 = µ4 versus HA: µ1, µ2, µ3, µ4 are not all equal where µ1, µ2, µ3, and µ4 correspond to the mean minimal luminal diameter for the placebo, probucol, multivitamin, and probucol and multivitamin groups, respectively.

The F-statistic is 5.97 with 3 and 226 df. The p-value is 0.0006 so we have very strong evidence against the null hypothesis. We have very strong evidence to suggest that the mean minimal luminal diameters are not all equal for the placebo, probucol, multivitamin, and probucol and multivitamin groups.

5a.) The p-value would likely increase and the statistically significant difference between groups would become smaller since they share more similar data (i.e. 29). In other words, there is less of a chance that the differences between groups would be statistically significant since they share more similar data points (i.e. 29). As you can see below, my hypothesis is supported and the p-value becomes larger (3.292e-11 instead of 5.259e-14 from 2a).



5b.) I would expect the p-value to be smaller since there is more data in each group that was already different from other groups in 2a. We already had evidence against the null so more of the same data will only reduce the p-value. More data exacerbates the already statistically significant difference, which is why the p-value is lowered. Note that I extracted the data into Excel and duplicated it there since I wasn’t sure how to do it in R. The duplicate dataset is bug2.

