**STAT 614 - HW 5**

By Sihyuan Han

**Due**: Thursday, October 29, 2020 in Blackboard by 11:59pm.

**Instructions**: Please type your solutions in a separate document and upload the document in Blackboard. Include supporting work (plots, etc.) when appropriate, but do not copy all computer output. Select only relevant output. I will not be collecting syntax for this assignment.

**Notes**:

* For this HW you will need some concepts from chapter 5 on the ANOVA model.
* HW 6 will finish out the ANOVA section.

The effects of exposure to lead on the psychological and neurological well-being of children were studied by Landrigan et al. (1975). Complete raw data for this study are in the data set lead.csv in Blackboard. The data describe a group of children who lived near a lead smelter in El Paso, Texas. Two exposed groups of children were identified who had blood-lead levels > 40 g/ml in 1972 or in 1973. Because neurological and psychological tests were performed in 1973, researchers argued that it would be better to define an exposure group based on blood-lead levels in 1973 only. For this purpose, the variable lead\_typ in the data file gives three exposure groups:

If lead\_typ = 1, then the child had normal blood-lead levels (<40 μg/ 100 mL) in both 1972 and 1973 (control group).

If lead\_typ = 2, then the child had elevated blood-lead levels (>40 μg/100 mL) in 1973 (the currently exposed group).

If lead\_typ = 3, then the child had elevated blood-lead levels in 1972 and normal blood-lead levels in 1973 (the previously exposed group).

One important measure of neurological function studied was MAXFT = the number of finger-wrist taps in the dominant hand. Researchers are interested in whether there is evidence of differences in neurological function, as measured by MAXFT, on average, between the three exposure populations. They would also like to test and estimate the average difference in MAXFT between each pair of exposure populations, **with the expectation that populations with** **normal blood-lead levels will have higher average MAXFT scores**. **It is unclear if previously exposed populations will have “recovered” any function as compared to a currently exposed population**. Address these research questions by answering the following questions.

1. State the hypotheses of interest to be tested. Include the overall test of group differences in addition to all possible pairwise comparisons of interest.

**Ans**:

RQ1- Normal blood-lead levels will have higher average MAXFT scores

H0: group1=group3

Ha: group1>group3

RQ2- Previously exposed populations will have “recovered” compared to a currently exposed population

H0: group2=group3

Ha: group2>group3

Overall test of equal population means

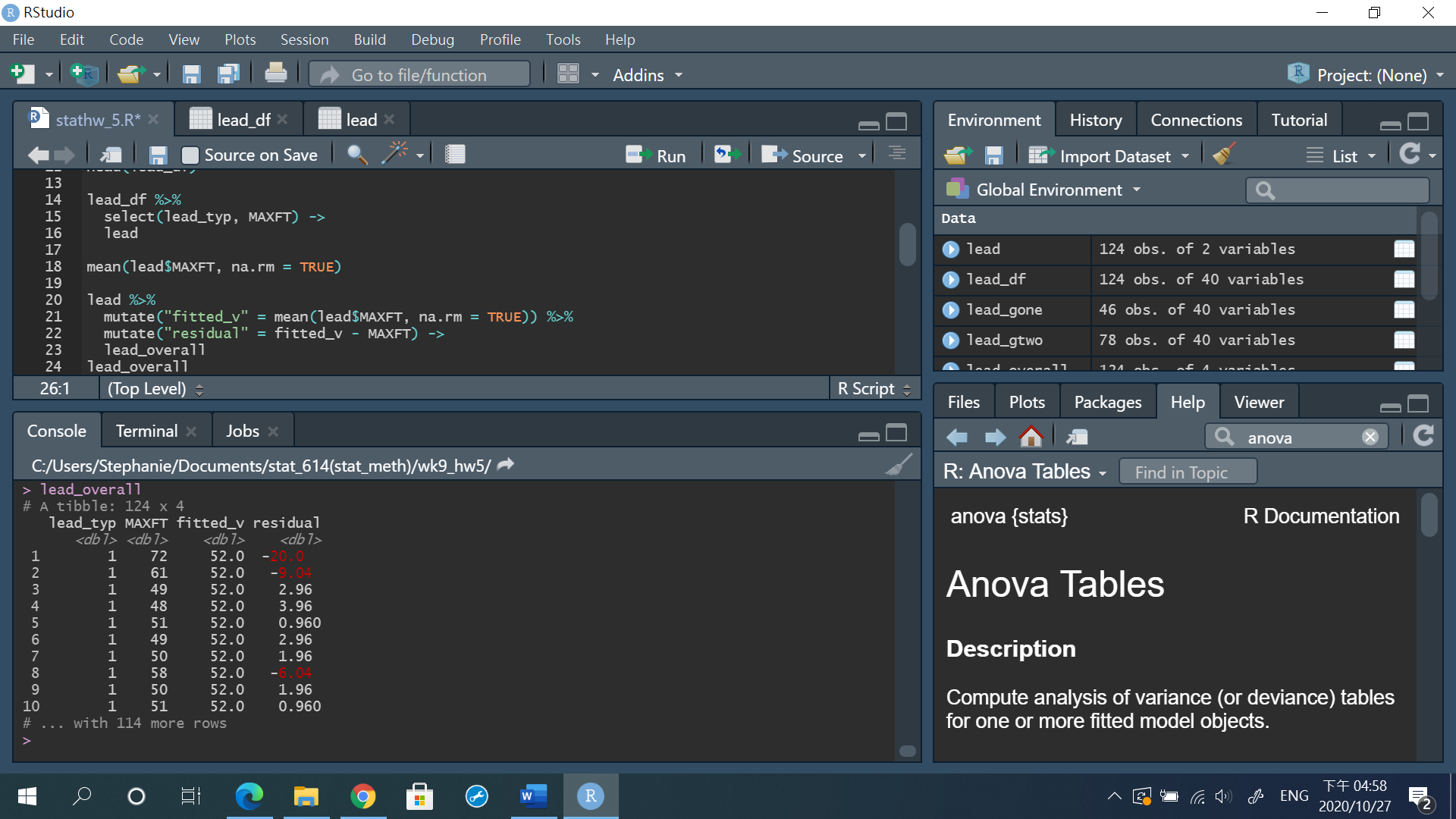
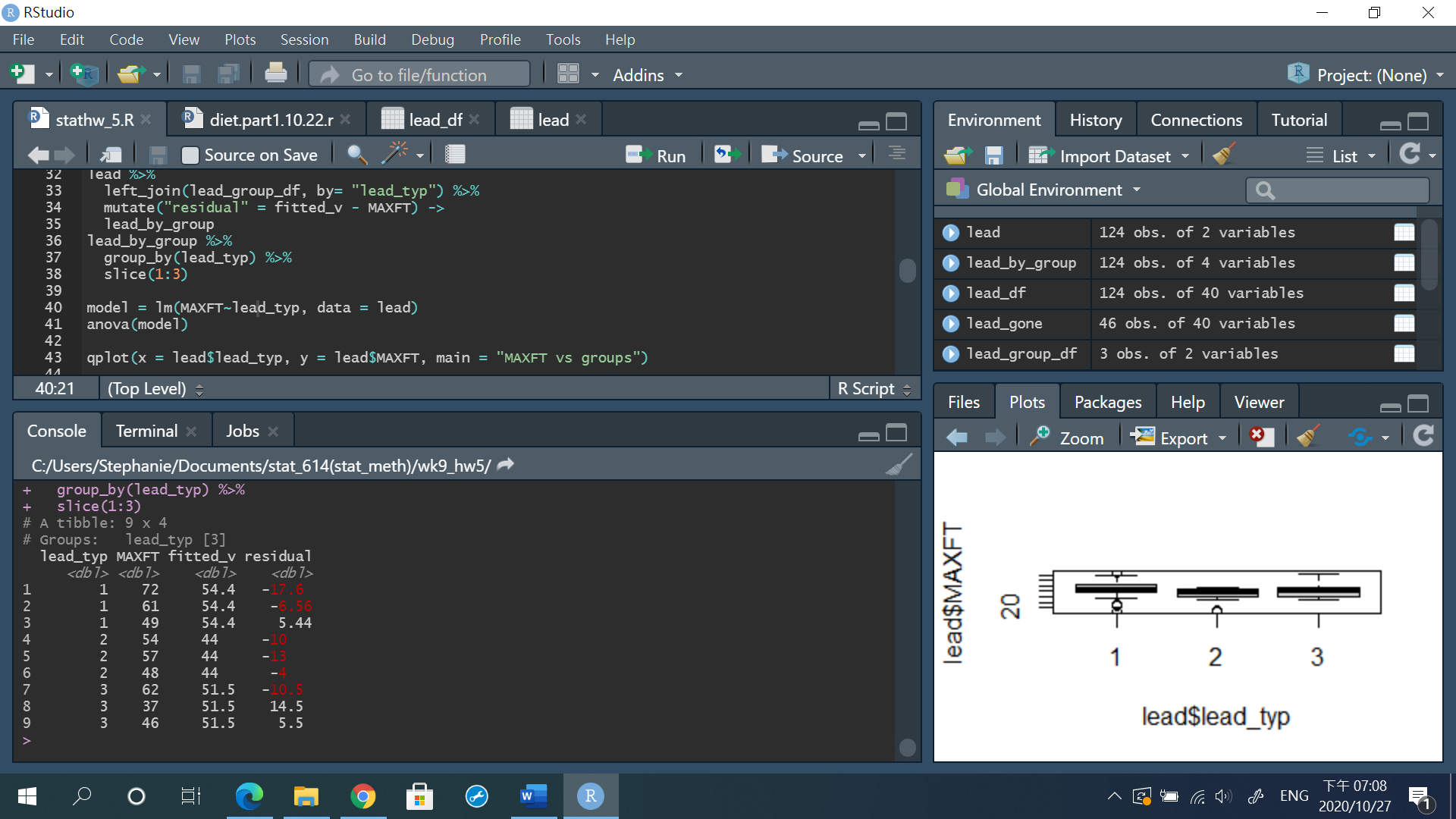
H0: group1 = group2 = group3

Ha: group1>group3

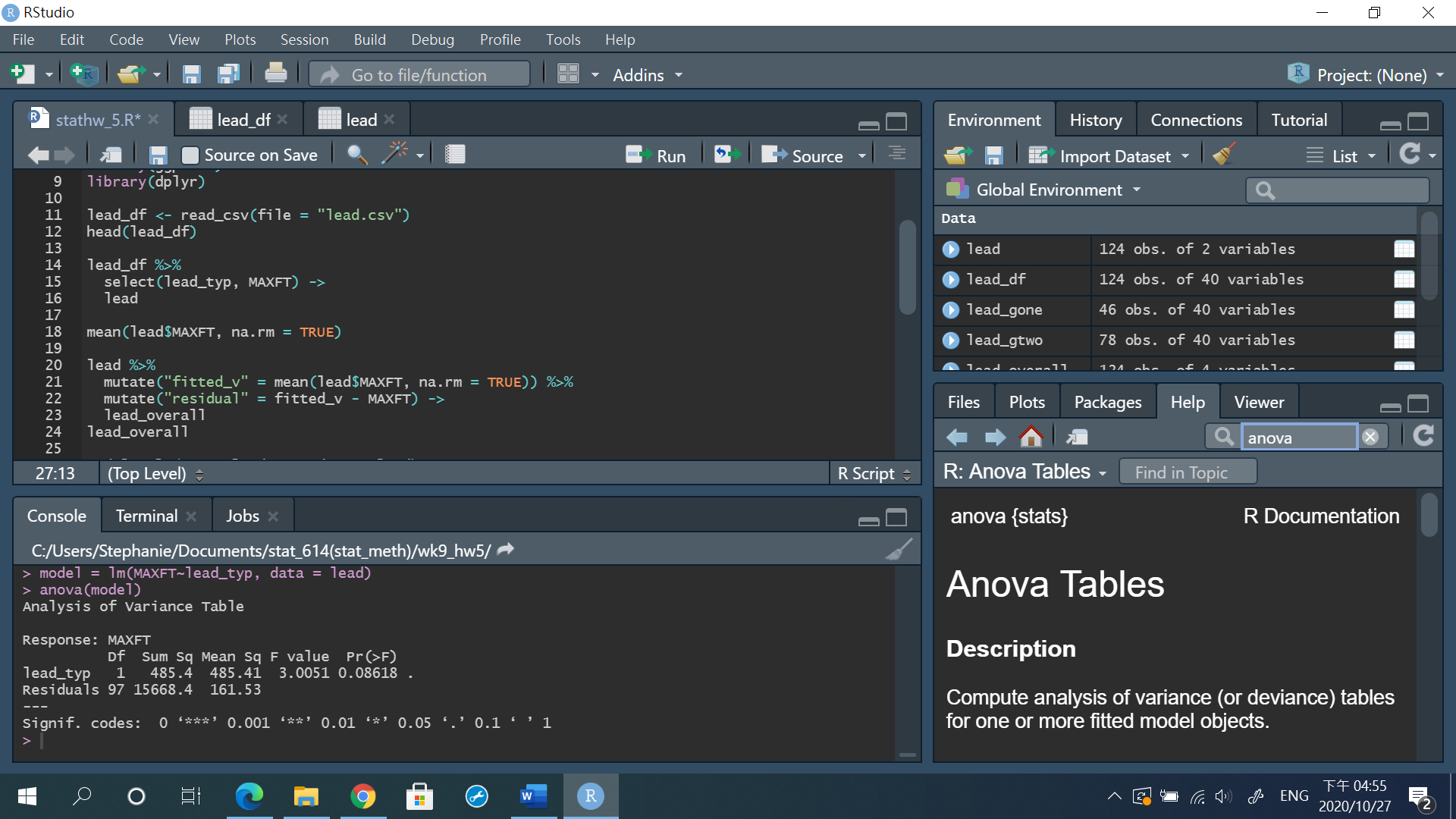
1. Write the ANOVA model to be fit.

**Ans**:

Equal means: Separate means:

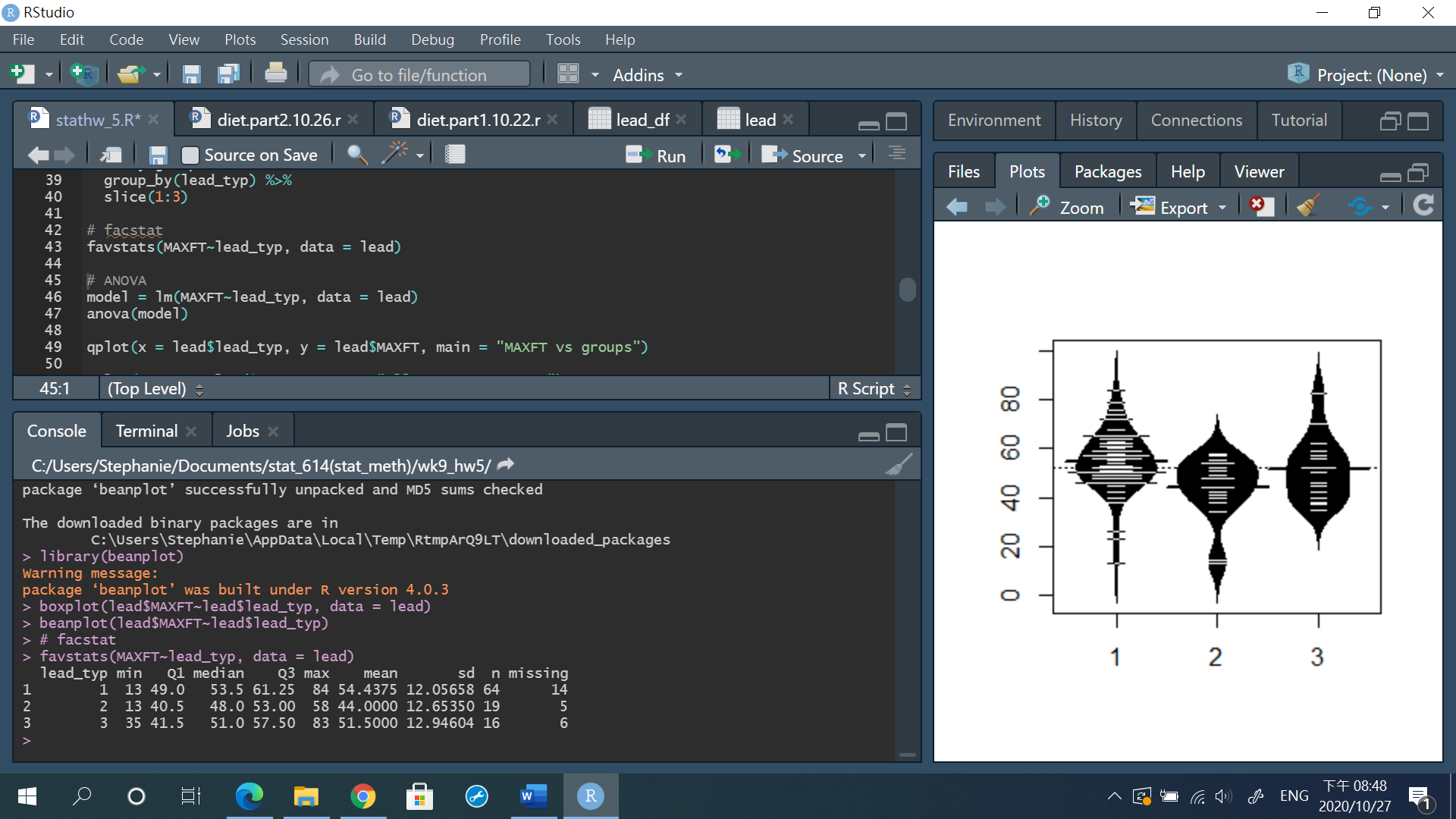
F-test for equal means

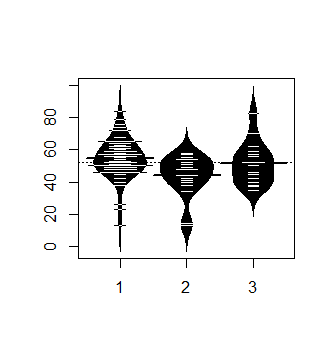
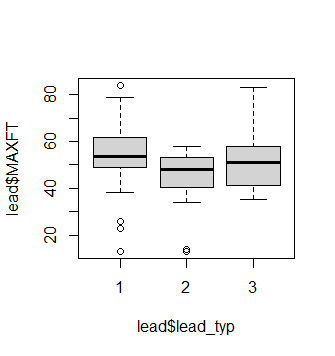


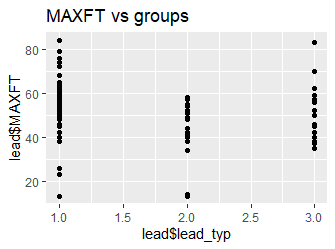
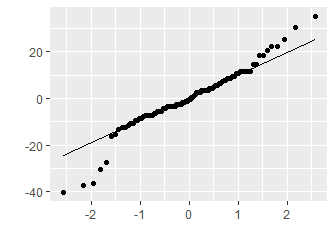
Total (SST)- sum of squares: 16153.8/ df: 98

1. Conduct a brief exploratory analysis of the MAXFT variable by exposure group (lead\_typ). Give supporting graphs, descriptive statistics, and interpret these results.

**Ans**:





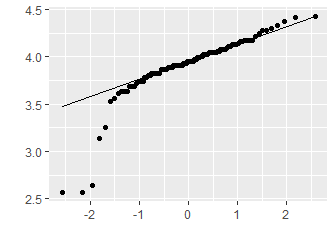


1. What are the assumptions of the model (and corresponding hypothesis tests)? Based on the exploratory analysis in (3), are the assumptions reasonably met for this data? If not, what adjustments should you make in your analysis? (You don’t need to use residuals here – you’ll do that on HW 6 and then for the rest of the semester!!!)

**Ans**:

1. Independent observations with groups (study design)
2. Independent samples (between groups) (study design)
3. Normally distributed – based on the QQ plot, it is not linear so denied.
4. Equal variances- yes, as it shows in the boxplot above.
5. Influential outliers assessed– as show in the plot, we can see there are many outliers!

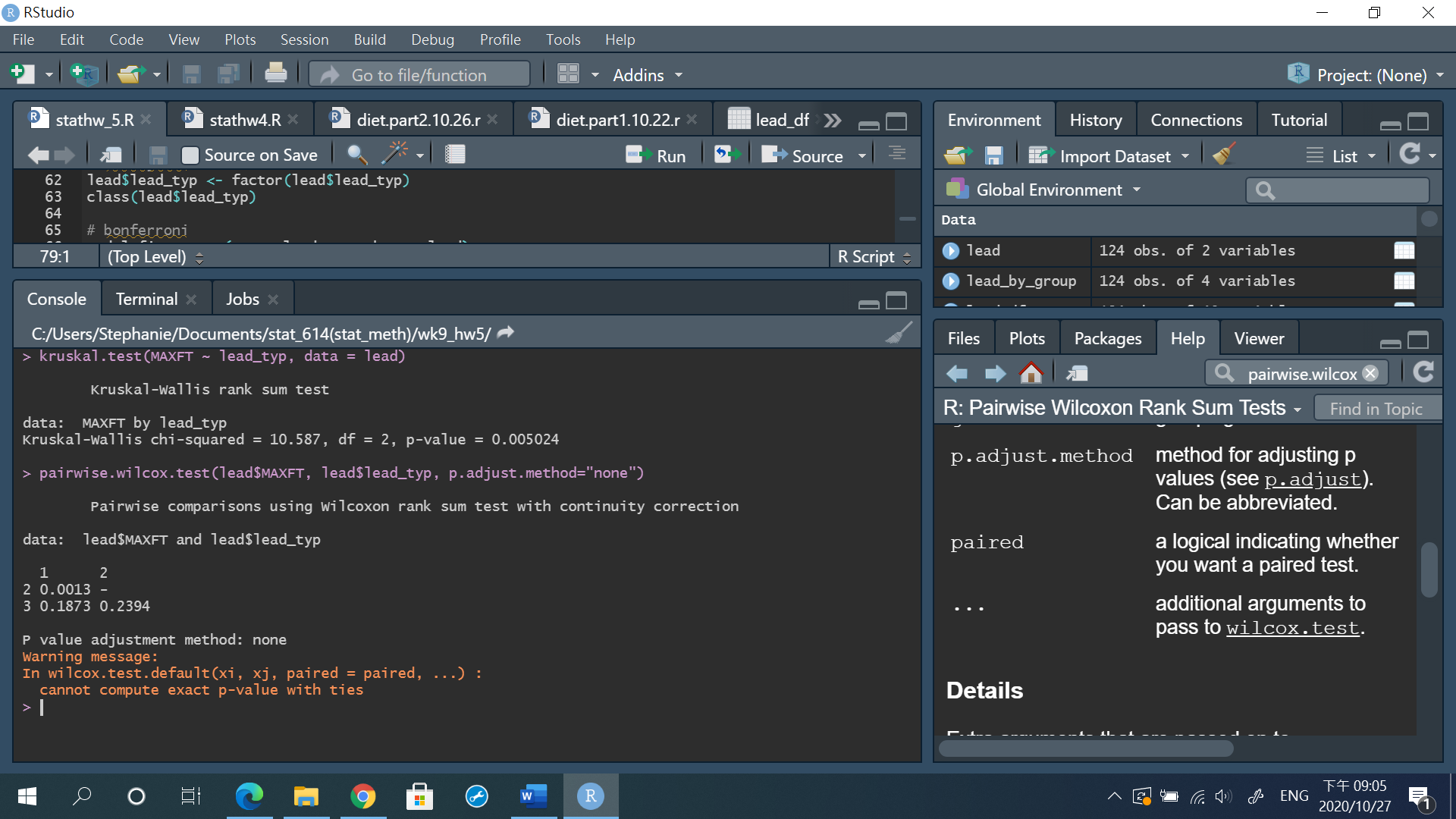
Try log the data and get the QQ plot, still not linear! Not normally distributed, so I’ll use nonparametric method.



1. Conduct the appropriate analysis (i.e. incorporate any recommended adjustments from (d) if you had them). Clearly and briefly state the conclusions of your analysis. Be sure you address the researcher’s questions.

**Ans**:

Nonparametric ANOVA method



For RQ1,