STAT 614 - HW 5 – Solutions

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

   There are four hypothesis tests to be conduct. Let μ_i = population mean MAXFWT score of exposure population group *i* = 1, 2, 3 (control, current, previous). Then we are interested in testing:

   (i) Ho: $\mu_1 = \mu_2 = \mu_3$ vs. Ha: Not all population means are equal.

   (ii) Ho: $\mu_1 = \mu_2$ vs. Ha: $\mu_1 > \mu_2$ as the researchers hypothesize higher neurological function in the control group than the currently exposed group, on average.

   (iii) Ho: $\mu_1 = \mu_3$ vs. Ha: $\mu_1 > \mu_3$ as the researchers hypothesize higher neurological function in the control group than the previously exposed group, on average.

   (iv) Ho: $\mu_3 = \mu_2$ vs. Ha: $\mu_3 \neq \mu_2$ as the researchers did not hypothesize a directional effect.

2. Write the ANOVA model to be fit.

   We will consider the ANOVA model

   $$y_{ij} = \mu_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

   where y_{ij} = MAXFWT score for subject *j* in exposure group *i* and μ_i = population mean MAXFWT score of exposure group *i* = 1, 2, 3 (control, current, previous). (ε_{ij} is the unobserved random deviation/error.)

3. Conduct a brief exploratory analysis of the `MAXFT` variable by exposure group (`lead_typ`). Give supporting graphs, descriptive statistics, and interpret these results.

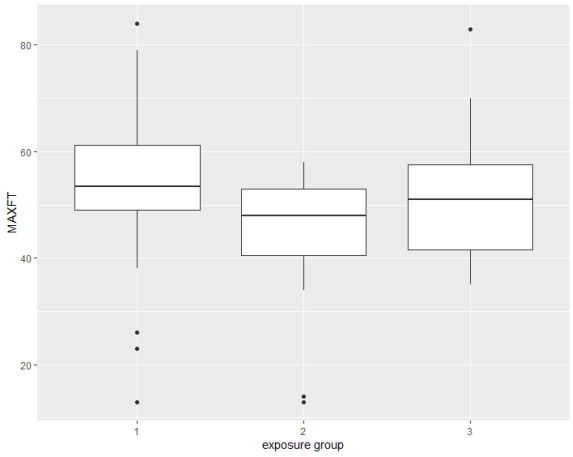
   Summary statistics and graphical displays are below. From these, the control exposure group had the largest sample size, with n = 64 participants having had the MAXFT score measured. The two exposure groups had much smaller sample sizes, 19 in the current exposed group and 16 in the previously exposed group. The current exposed group had the lowest average MAXFT scores, with mean of 44 taps and median of 48, followed by the previously exposed having a mean and median of 51, and the control group having a mean of 54.4 taps and median of 53.5 taps. The standard deviations are similar with the previous group having the largest standard deviation (12.9) and IQR (16). The histograms suggest some potential skewness in the current and previous exposure groups. These and the boxplots suggest there are outliers in the control group (at least 4 – one extreme high value and 3 extreme low values) and in the current group (2 extreme low outliers). The QQ Plots actual suggest there are only small deviations from normality in each of the groups – with outliers in all groups.

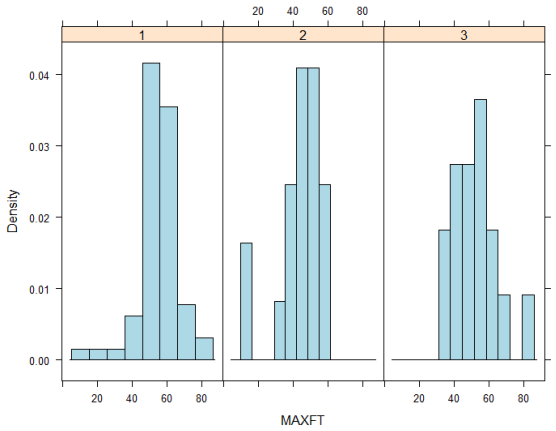
Output from Exploratory Data Analysis:

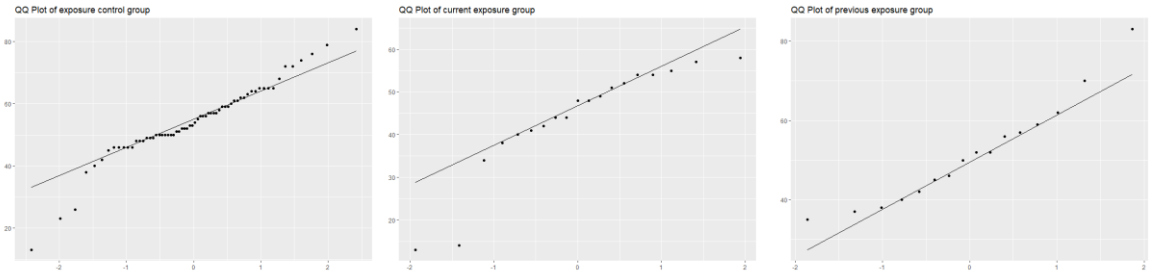
```
factor(lead_typ) min   Q1 median    Q3 max    mean       sd  n missing
control  1        13 49.0   53.5 61.25  84 54.4375 12.05658 64      14
current  2        13 40.5   48.0 53.00  58 44.0000 12.65350 19       5
previous 3        35 41.5   51.0 57.50  83 51.5000 12.94604 16       6
```

I also got the variances and IQRs by group:

```
                var    IQR
1 control    145.3611  12.25
2 current    160.1111  12.50
3 previous   167.6000  16.00
```







4. 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?

    (i) Normality: MAXFT scores follow a normal distribution within each exposure group or equivalently, the (unobserved) residuals follow a normal distribution.

    (ii) Equal variances: The (population) variance of MAXFT scores is the same in each of the three exposure groups. (Or, equivalently the unobserved residuals have equal variances.)

    (iii) Outliers: There are no influential outliers.

    (iv) Independence between groups: The exposure groups are independent of one another.

    (v) Independence within groups: Individuals within a given exposure group are independent of one another.

    The exploratory analysis below and a residual analysis are used to assess the assumptions of normality and equal variances, and to identify outliers. **We will hold off on the residual analysis for this assignment but if you included it – great! Just also include in on the next HW.**

As discussed in part (3), there are small deviations from normality in all groups, primarily through outliers, particularly on the lower end of the control and current exposure groups and the upper end of the previous exposed group. The sample standard deviations are similar (ranging from about 12 to about 13 with variances ranging from about 145 to 168) but these are sensitive to outliers. The IQRs, ranging from 12.25 to 16.00, are not too different. Independence of exposure groups is reasonably met based on the study design although as this is an observational study, I would hesitate to draw causal conclusions from these results. We aren't really given enough information to assess independence of individuals within groups, so we proceed as if they are independent.

I am most concerned about the outliers as the two exposed groups have small sample sizes.

e. Conduct the appropriate analysis (i.e. incorporate the need adjustments from (4) if you had them). Clearly and briefly state the conclusions of your analysis. Be sure you address the researcher's questions.

As there are clearly outliers, I will present the ANOVA based on the full data analysis and an analysis holding out outliers. You can also try a transformation, such as using a logarithmic or square root transformation. I considered these but there were still many outliers after transforming so I decided to pursue other options. If you found a transformation that is reasonable – well done!

I go into a lot of detail on the following two pages to explain some of the choices and concepts. I don't expect this level of explanation on your assignments!

Full data analysis

The ANOVA overall F-test of equal population means has a p-value of 0.007, providing strong evidence of at least one difference in population means. The pairwise comparisons (using the LSD method) suggest evidence of differences in the mean MAXFWT scores between the control (1) and current exposed groups (2), with one-sided p-value of 0.002/2 = 0.001. There is only marginal (but not "statistically significant" or conclusive) evidence of a difference between the currently (2) and previously (3) exposed groups with a two-sided p-value of 0.076. The one-sided test of group 1 (control) vs. group 3 (previously exposed) is p = 0.395/2 = 0.1975 suggests little evidence of differences in average neurological function between these populations.

A 95% confidence interval estimates that the mean finger wrist tap scores are 4.1 to 16.9 points higher in the control group than the currently exposed group. Estimates and confidence intervals for the other two comparisons are also provided, with max finger wrist tap scores in the control population estimated to be 2.9 points higher, on average, as compared to the previously exposed populations, but with 95% confidence we can at best say the difference in means is reasonably -3.9 points to 9.8 points between these two populations (hence the inconclusive hypothesis test!) The max finger wrist tap scores are estimated to be 7.5 points higher in the previously exposed population versus the currently exposed population, on average and from only 0.8 lower to 15.8 higher on average, with 95% confidence. While this confidence interval does contain zero (and hence we "fail to reject" the null of no difference in population means) it is suggestive of some improved neurological function, on average, in the previously exposed population.

Results from the full data analysis should be considered tenuous until a fuller analysis of the impact of the outliers is examined (and until a better assessment of all assumptions is addressed on the next HW using residuals!).

```
> # ANOVA on full data
> model.fit <- aov(MAXFT ~ factor(lead_typ), data = sublead)
> anova(model.fit)
Analysis of Variance Table

Response: MAXFT
                Df  Sum Sq Mean Sq F value   Pr(>F)
factor(lead_typ)  2  1600.1  800.04  5.2773 0.006692 **
Residuals        96 14553.8  151.60
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> PostHocTest(model.fit, method = "lsd")  # lsd = "Least Significant
Difference" / "F protected"

  Posthoc multiple comparisons of means : Fisher LSD
    95% family-wise confidence level

$`factor(lead_typ)`
        diff      lwr.ci    upr.ci   pval
2-1 -10.4375 -16.8227981 -4.052202 0.0016 **
3-1  -2.9375  -9.7688090  3.893809 0.3955
3-2   7.5000  -0.7928946 15.792895 0.0758 .
```

Outlier analysis:
The plots in (3) suggest up to 7 observations as being outliers in the boxplots. These correspond to the largest MAXFT scores (83 and 84) and the smallest (ranging from 13-26). Seven observations which is a lot to investigate. With just one or two, I would hold each out separately and then together and re-examine the assumptions and model fits – which corresponds to THREE more analyses to consider. Imagine what we'd have to do with 7 outliers! (Don't image - I'll tell you: if you we held out each one at a time, then two at a time, then three, etc… , and all the way up to all of them then that would be 127 additional analyses to consider. WOWZA! We will be looking at *better* ways to assess the impact of outliers in later chapters - thankfully.)

So, what should we do here? For this HW, I'm going to hold out these seven outliers all out once to assess their impact (i.e. see if they *influence* my conclusions). The outliers are:

Observation #	Exposure Group	MAXFWT		Observation #	Exposure Group	MAXFWT
17	1 (Control)	84		73	2 (Current)	14
30	1 (Control)	23		75	2 (Current)	13
34	1 (Control)	26		94	3 (Previous)	83
49	1 (Control)	13				

Summary statistics and graphical displays from an exploratory analysis are presented on the following pages. From these we see that all of the prior concerns with the assumptions (normality and outliers) have been addressed. There is still one outliers but it is not that extreme (and, there will *always* be outliers – we could remove, refit, reassess, etc. and continue to have outliers!) The sample variances do range from about 50 to just over 100 (a ratio of 2) but in a formal test of equal variances we fail to reject the hypothesis of equal variances.

The full results of the ANOVA model fit are given after the summary plots and stats. The tables below give a direct comparison of the two analyses (full data vs. without outliers). Removing the outliers had little impact on the overall F-test of equal means (there is still convincing evidence of differences in mean max finger wrist tap scores between the three exposure populations) as displayed in Table 1. Nor

was there a major impact on the comparison between the control and currently exposed populations (1 vs. 3) as displayed in the first row of Table 2, with strong evidence in either analysis of higher average scores in the control populations versus previously exposed populations. The estimated difference in mean scores is 10.4 (full data) vs. 8.0 (without outliers) and the 95% confidence interval is "more upward" when using the full data (4.1 to 16.8 vs. 3.2 to 12.9). That is, holding the outliers out "attenuates" the estimates a bit.

The impact of the outliers is more visible when looking at the last two comparisons. The estimated difference in means increases from 2.9 to 6.9 for the control vs the currently exposed population and is significant when using the data without outliers (p = 0.1975 vs. p = 0.008). Note that the confidence interval estimates are *somewhat* consistent in that, while inconclusive when using the full data, we have comparable upper bounds (9.8 vs. 11.3). Finally, there is no substantial evidence of a difference in mean MAXFT between the previous and current exposed populations in the analysis without outliers. The p-value is 0.564 when previously it was 0.076 (not "significant" but certain providing some evidence). The confidence interval bounds also change, from -0.8 to 15.8 using the full data to -4.4 to 8.0 using the data without outliers. The outliers are fairly *influential* for these last two comparisons.

So, what can we conclude from all this? The comparison of the control to the previously exposed population (1 vs. 2) was pretty stable between the two analyses. Comparisons to the currently exposed group are sensitive to outliers and thus inconclusive.I would prefer to use an *outlier resistant* method, such as a nonparametric technique. I would also prefer to more formally address the assumptions using the residuals. All for the next HW!

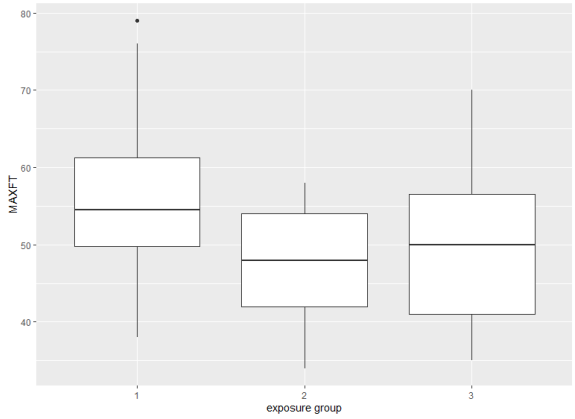
Table 1. ANOVA F test of equal means			
Full data		Without outliers	
F Stat	p-value	F Stat	p-value
5.277	0.007	7.095	0.001

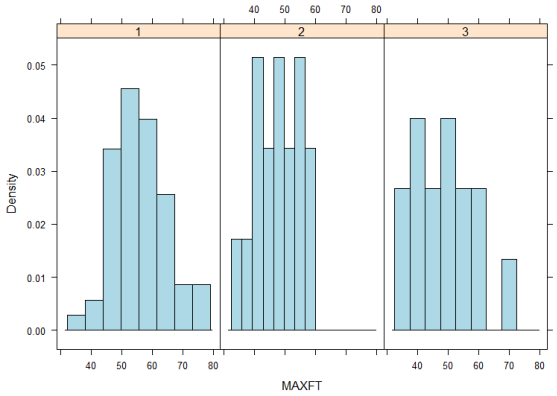
Table 2. Pairwise Comparison confidence interval bounds and p-values								
	Full data				Without outliers			
	Estimated difference in means	p-value	Lower Bound	Upper Bound	Estimated difference in means	p-value	Lower Bound	Upper Bound
1 vs. 2	10.4	0.001	4.1	16.8	8.0	0.0005	3.2	12.9
1 vs. 3	2.9	0.1975	-3.9	9.8	6.2	0.008	1.2	11.3
3 vs. 2	7.5	0.076	-0.8	15.8	1.8	0.564	-4.4	8.0

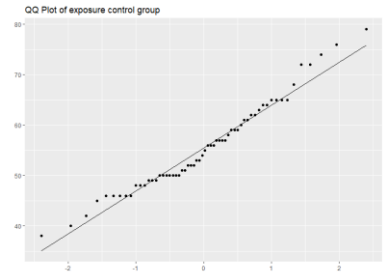
Exploratory Data Analysis on data without seven outliers:

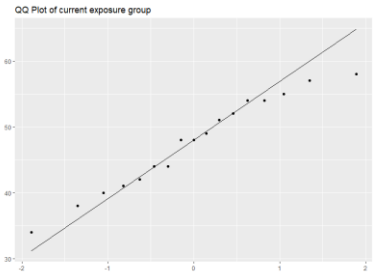
```
> favstats(MAXFT ~ factor(lead_typ), data = sublead.o)
  factor(lead_typ) min    Q1 median    Q3 max     mean        sd  n missing
control          1  38 49.75   54.5 61.25  79 55.63333  8.910034 60       0
current          2  34 42.00   48.0 54.00  58 47.58824  7.080420 17       0
previous         3  35 41.00   50.0 56.50  70 49.40000 10.196638 15       0

                 var   IQR
control    79.38870   11.5
current    50.13235   12.0
previous  103.97143   15.5
```

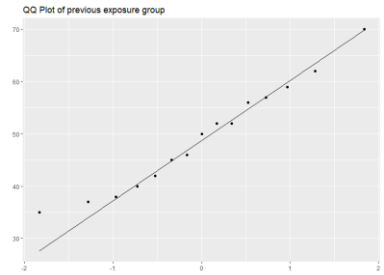
```
> leveneTest(MAXFT ~ factor(lead_typ), data = sublead.o)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  2  0.9709 0.3827
      89
```












```
> anova(model.fit)
Analysis of Variance Table

Response: MAXFT
                 Df Sum Sq Mean Sq F value   Pr(>F)
factor(lead_typ)  2 1106.8  553.39  7.0951 0.001383 **
Residuals        89 6941.7   78.00
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> PostHocTest(model.fit, method = "lsd")  # lsd = "Least Significant
Difference" / "F protected"

  Posthoc multiple comparisons of means : Fisher LSD
    95% family-wise confidence level

$`factor(lead_typ)`
         diff     lwr.ci    upr.ci   pval
2-1 -8.045098 -12.866517 -3.223679 0.0013 **
3-1 -6.233333 -11.299028 -1.167639 0.0165 *
3-2  1.811765  -4.404571  8.028101 0.5640

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```