# Chapter 8 and 9 Homework

## Due 3-18-2018

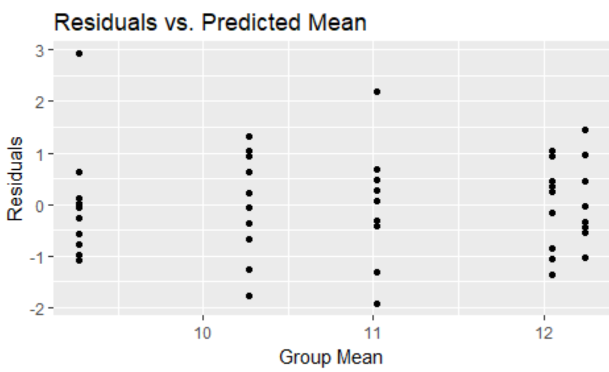
### Problem 9.13

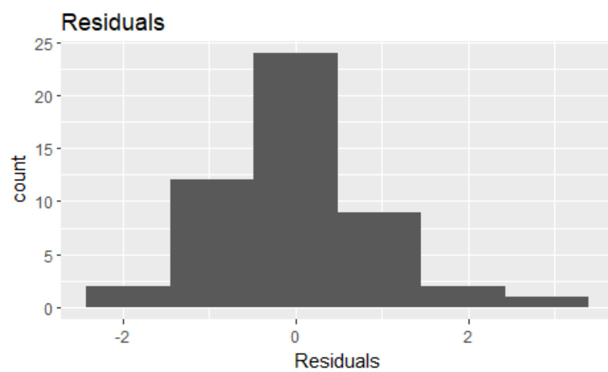
**a) Assess ANOVA assumptions using the graph from PROC MIXED.**

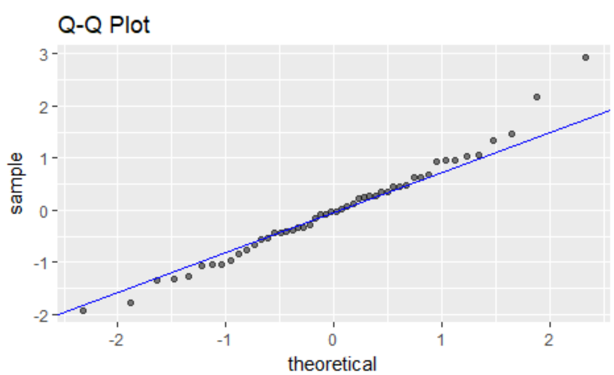
In order to proceed with ANOVA, we must check for the following pieces:

- No obvious pattern in our Residual Scatterplot
- Normal shape to the Residual Histogram.
- Residuals fit along the linear prediction of our Actual "Normality" to the Theoretical Normal Model.

Since there is no readily observable pattern in our residuals, our distribution is mound shaped, and our Residuals fit the along the Q-Q plot(although, there may be evidence of a right skew), we will proceed with ANOVA to see if at least one mean is different.



**b) Perform ANOVA to determine if there is a difference among the five weight-reducing agents, $\alpha = 0.05$.**

Table 1: ANOVA Table for Weight Loss Study

| Row Names | SS | df | MS | F | P-Value |
|---|---|---|---|---|---|
| Treatments (T) | 61.618 | 4 | 15.4045 | 15.6805 | $4.16\text{x}10^{-8}$ |
| Error (E) | 44.207 | 45 | 0.9824 | | |
| Total | 105.825 | 49 | | | |

Hypotheses:

$H_0$: $\mu_{a1} = \mu_{a2} = \mu_{a3} = \mu_{a4} = \mu_s$
$H_1$: At least one mean is different.

Test Statistic:

F = 15.6805

Rejection Region:

Reject $H_0$ if $F_0 > F_{\alpha,4,45}$
$F_{0.95,4,45} = 5.72$

Conclusion/Interpretation:

Since our $F_0 > 5.72$, there is strong enough evidence to support rejecting the null hypothesis that the means are the same. The data provided does suggest at least one mean is different from the rest.

**c) Determine significantly different pairs using Tukey's W with $\alpha = 0.05$**

To find which groups are different from each other, we will utilize Tukey's W to check which means are different from one another. We will verify this by checking which one's p-values are less than 0.05 after using the TukeyHSD test on our ANOVA Model in R.

Conclusion/Interpretation

With our Tukey W test being run, the differences lie in the following means:

- S differs with $A_1$, $A_2$, and $A_4$
- $A_3$ differs with $A_1$ and $A_4$

**d) Determine which, if any, of the new agents have significantly larger mean weight loss as compared to the standard agent; $\alpha = 0.05$**

Using s as our control, we will performs Dunnett's test to determine if there's a significantly larger difference with the new agents. The first step in checking out if any of the new agents are yielding better weight loss results is to find Dunnett's D.

$$D = d_\alpha(k, v)\sqrt{\frac{2s_W^2}{n}}$$
$$D = d_{0.05}(4, 45)\sqrt{\frac{2(0.982)}{10}}$$
$$D = 2.23\sqrt{\frac{1.964}{10}}$$
$$D = 2.23(0.4432)$$
$$D = 0.9888$$

Table 2: Dunnett's Control Test

| Treatment | $\bar{y}_i - \bar{y}_c$ | Comparison | Conclusion |
|---|---|---|---|
| a1 | 2.78 | > D | Greater Than Control |
| a2 | 1.75 | > D | Greater Than Control |
| a3 | 1.00 | > D | Greater Than Control |
| a4 | 2.97 | > D | Greater Than Control |

Conclusion/Interpretation:

All four of the agents are statistically significant. Thus, we can reject the null hypotheses that the difference in weight loss metrics between the original agent(s) and the new agents($a_1, a_2, a_3$, and $a_4$) is zero. Each of the new agents appear to have an increase in weight loss over the agent s.

**9.17 - construct the contrasts for the following:**

**a) Compute the mean for the standard agent to the average of the means for the four new agents**

$$\hat{l} = \bar{y}_S - \frac{1}{4}\bar{y}_{A_1} - \frac{1}{4}\bar{y}_{A_2} - \frac{1}{4}\bar{y}_{A_3} - \frac{1}{4}\bar{y}_{A_4}$$

**b) Compare the mean for the agents with counseling to those without counseling. (Ignore the standard)**

$$\hat{l} = \bar{y}_{A_1} + \bar{y}_{A_3} - \bar{y}_{A_2} - \bar{y}_{A_4}$$

**c) Compare the mean for the agents with exercise to those without exercise.**

$$\hat{l} = \bar{y}_{A_1} + \bar{y}_{A_2} - \bar{y}_{A_3} - \bar{y}_{A_4}$$

**d) Compare the mean for the agents with counseling to the standard.**

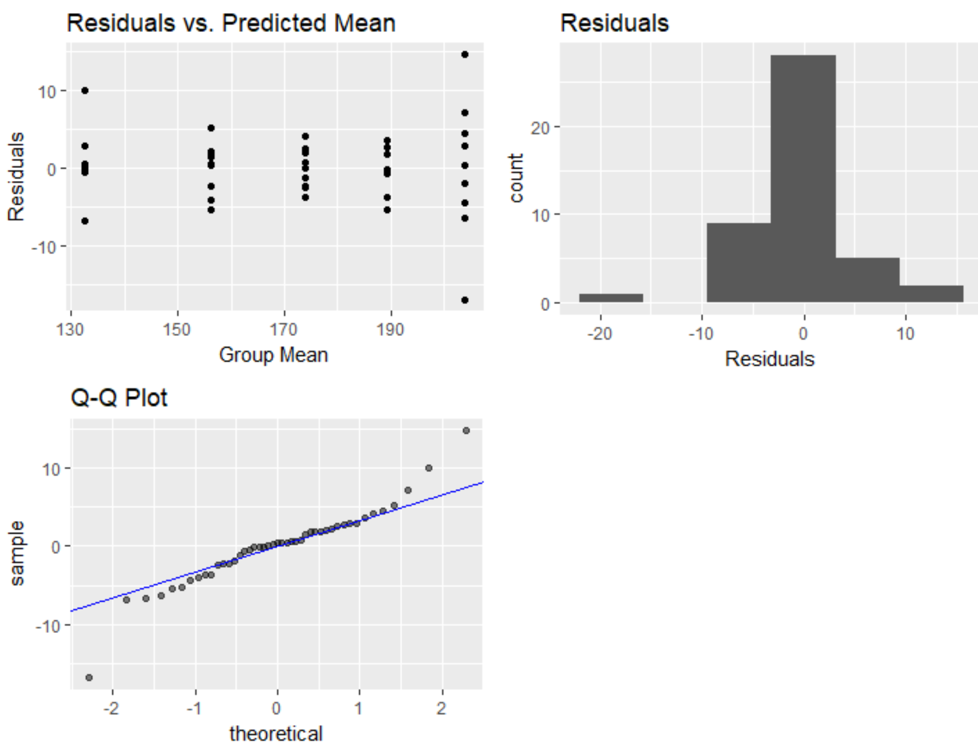
$$\hat{l} = \bar{y}_{A_1} + \bar{y}_{A_3} - 2\bar{y}_S$$

### 8.27 a) Assess ANOVA assumptions using the graph from PROC MIXED

In order to proceed with ANOVA, we must check for the following pieces:

- No obvious pattern in our Residual Scatterplot
- Normal shape to the Residual Histogram.
- Residuals fit along the linear prediction of our Actual "Normality" to the Theoretical Normal Model.

There appears to be a pattern in the residuals. The best way to describe the pattern is a increasing fan shape at the edges. There is an increase in the residuals as you are on the extremes of the means of each group. Finally, a pattern in the Q-Q plot has this sample data straying from the Normal model by mismatching the Normal model outside of the -1 and + 1.5 standard deviations.




### b) Run an ANOVA on the data to see if there is a difference in the suppliers

Table 3: ANOVA Table for Supplier's Variability in Power

| Row Names | SS | df | MS | F | P-Value |
|---|---|---|---|---|---|
| Suppliers (S) | 28024 | 4 | 7006 | 865.9 | $< 2\text{x}10^{-16}$ |
| Error (E) | 1054 | 40 | 26 | | |
| Total | 29078 | 44 | | | |

Hypotheses:

$H_0 : \mu_{S1} = \mu_{S2} = \mu_{S3} = \mu_{S4} = \mu_{S5}$
$H_1$ : At least one mean is different.

Test Statistic:

F = 865.9

Rejection Region:

Reject $H_0$ if $F_0 > F_{\alpha,4,20}$ $F_{0.05,4,25} = 5.80$

Conclusion/Interpretation:

With a test statistic greater than the rejection bound, there does seem to extreme enough evidence to reject the null hypothesis that the means are equal. It does appear that at least one mean is different.

**c) Perform a Kruskal-Wallis test to determine if there is a difference among the suppliers; $\alpha$ = 0.05**

Hypothesis:

$H_0$: There is no difference among the five suppliers with respect to median Variability in Power.
$H_1$: At least one of the five suppliers' medians differs from the others.

Test Statistic:

H = 41.596 with 4 degrees of freedom and a p-value of 2.023x$10^{-8}$

Rejection Region:

We will reject if H > $\chi^2_{0.05,4}$
$\chi^2_{0.05,4} = 9.488$

Conclusion/Interpretation

Since our Test Statistic is greater than 9.488, we have significant evidence to reject the Null Hypothesis that the all the medians are the same. There does appear to be at least one median that is different than the rest.

**d) Compare the results from b) and c) - which analysis would you say is correct?**

Reasoning:

In determining which analysis to agree with, I am hesitant to base decisions on the ANOVA approach due to its pattern in its residuals. Additionally, such a small sample set of 9 observations per Supplier has me raising an eye brow at the thought of trusting an ANOVA test's results. Finally, the Q-Q plot's lack of fidelity to the theoretical Normal distribution, particularly at the extreme ends of the linear model, has me doubting even more. Thus, the non-parametric approach with Kruskal-Wallis is what I would base my decision making on.

**e) Determine significantly different pairs using Tukey's W with $\alpha$ = 0.05.**

To find which groups are different from each other, we will utilize Tukey's W to check which means are different from one another. We will verify this by checking which one's p-values are less than 0.05 after using the TukeyHSD test on our ANOVA Model in R.
Conclusion/Interpretation:

With our Tukey W test being run, the differences lie in the following means:

- All the Suppliers are different with each other. They all have a p-value less than 0.05.

Table 4: Tukey HSD Test for Variability in Power per Supplier

| Supplier | diff | lower bound | upper bound | p-value |
|---|---|---|---|---|
| S3-S4 | 23.6667 | 16.7516 | 30.5772 | $3.6567x10^{-11}$ |
| S2-S4 | 41.3889 | 34.4783 | 48.2994 | $4.6363x10^{-13}$ |
| S1-S4 | 56.5333 | 49.6228 | 63.4439 | $4.6363x10^{-13}$ |
| S5-S4 | 71.3333 | 64.4228 | 78.2439 | $4.6363x10^{-13}$ |
| S2-S3 | 17.7222 | 10.8117 | 24.6328 | $6.5207x10^{-13}$ |
| S1-S3 | 32.8667 | 25.9561 | 39.7772 | $4.6829x10^{-13}$ |
| S5-S3 | 47.6667 | 40.7561 | 54.5772 | $4.6363x10^{-13}$ |
| S1-S2 | 15.1444 | 8.2339 | 22.0550 | $1.9784x10^{-06}$ |
| S5-S2 | 29.9444 | 23.0339 | 36.8550 | $5.0204x10^{-13}$ |
| S5-S1 | 14.8000 | 7.895 | 21.7106 | $3.1290x10^{-06}$ |

**f) Determine significantly different pairs using the Kruskal-Wallis approach using $\alpha = 0.05$.**

Table 5: Table of Suppliers' Mean Ranks

| Supplier 1 | Supplier 2 | Supplier 3 | Supplier 4 | Supplier 5 |
|---|---|---|---|---|
| 33.00 | 23.00 | 14.00 | 5.00 | 40.22 |

Rejection Region:

With the means of the ranks calculated for each group, we will calculate the Rejection bound for the KW test.

$$KW \approx \frac{q_\alpha(t,\infty)}{\sqrt{2}}\sqrt{\frac{n_T(n_T+1)}{12}\Big(\frac{1}{n_i}+\frac{1}{n_j}\Big)} \tag{1}$$

$$KW \approx \frac{3.86}{\sqrt{2}}\sqrt{\frac{45(46)}{12}\Big(\frac{1}{9}+\frac{1}{9}\Big)}$$

$$KW \approx 2.7294(6.1914)$$

$$KW \approx 16.8988$$

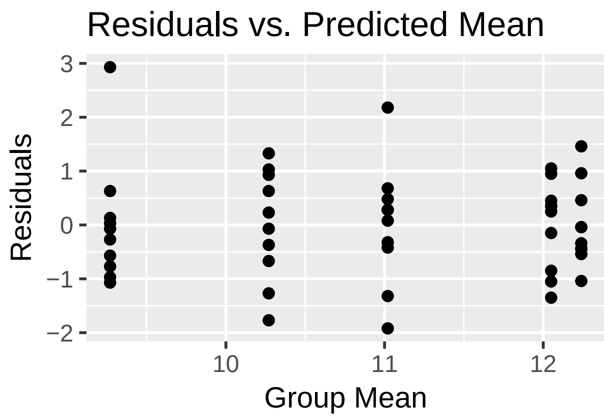
Table 6: Tukey HSD Test for Variability in Power per Supplier

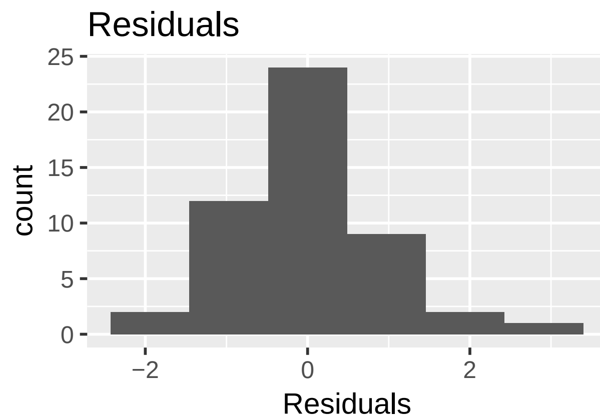
| Supplier Difference | $\lvert R_i - R_j \rvert$ | Conclusion |
|---|---|---|
| Supplier1-Supplier2 | $\lvert 33.00 - 23.00 \rvert = 10$ | Not Significantly Different |
| Supplier1-Supplier3 | $\lvert 33.00 - 14.00 \rvert = 19$ | Significantly Different |
| Supplier1-Supplier4 | $\lvert 33.00 - 5.00 \rvert = 28.00$ | Significantly Different |
| Supplier1-Supplier5 | $\lvert 33.00 - 40.22 \rvert = 7.22$ | Not Significantly Different |
| Supplier2-Supplier3 | $\lvert 23.00 - 14.00 \rvert = 9$ | Not Significantly Different |
| Supplier2-Supplier4 | $\lvert 23.00 - 5.00 \rvert = 18.00$ | Significantly Different |
| Supplier2-Supplier5 | $\lvert 23.00 - 40.22 \rvert = 17.22$ | Significantly Different |
| Supplier3-Supplier4 | $\lvert 14.00 - 5.00 \rvert = 9.00$ | Not Significantly Different |
| Supplier3-Supplier5 | $\lvert 14.00 - 40.22 \rvert = 26.22$ | Significantly Different |
| Supplier4-Supplier5 | $\lvert 5.00 - 40.22 \rvert = 35.22$ | Significantly Different |

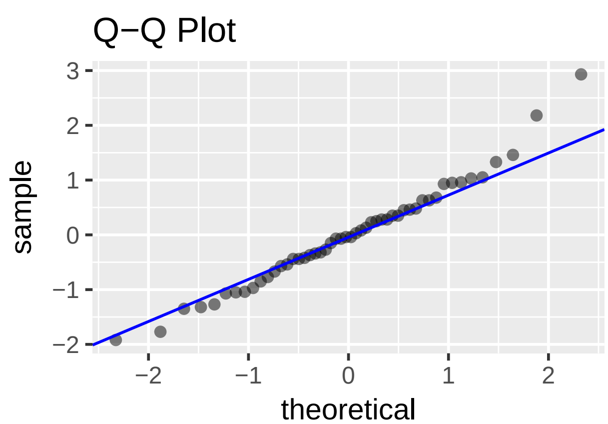
# Chapter 8 And 9

*Kyle Ligon*

9.13 a) Checking the results from Proc Mixed in order to do ANOVA




9.13 b) Perform ANOVA test on the data: Show ANOVA Table First, then Run the Test

```
anova_mod
```

```
## Call:
##    aov(formula = wt_loss ~ treatment, data = gather_frame)
##
## Terms:
##                 treatment Residuals
## Sum of Squares     61.618    44.207
## Deg. of Freedom         4        45
##
## Residual standard error: 0.9911497
## Estimated effects may be unbalanced
```

```
summary(anova_mod)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## treatment    4  61.62  15.404   15.68 4.16e-08 ***
## Residuals   45  44.21   0.982
## ---
```

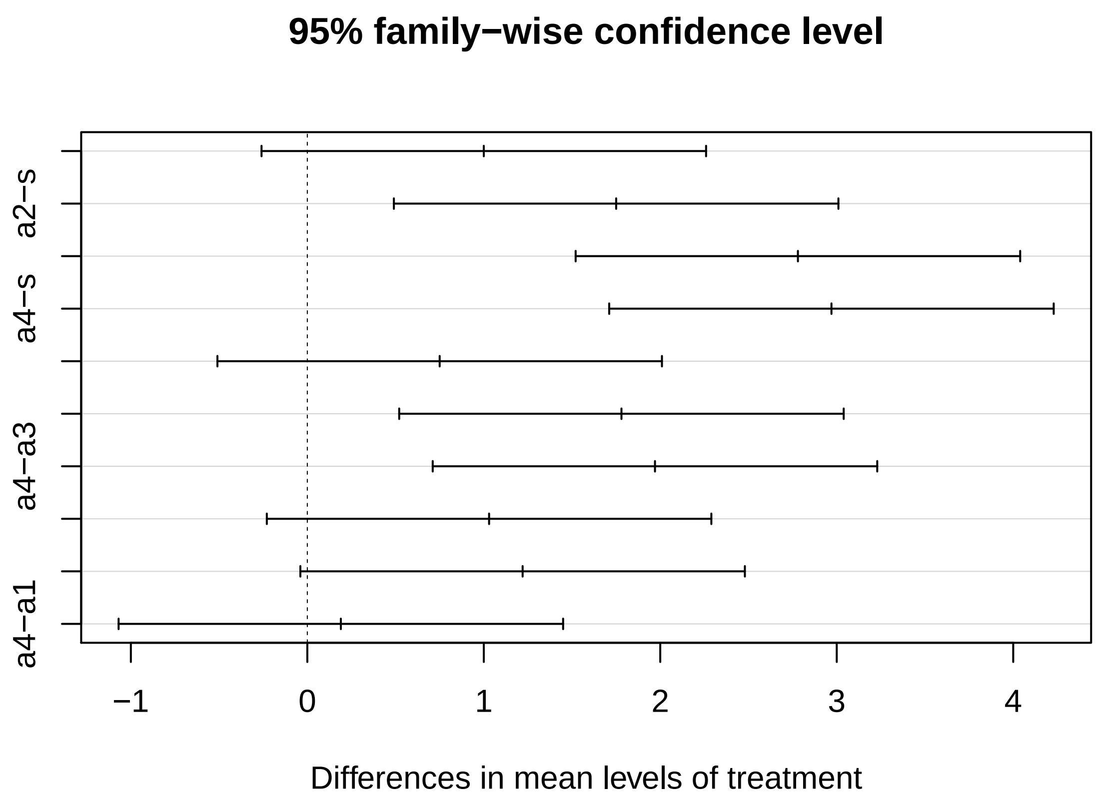
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

9.13 c) Perform Tukey's W on the significant pairs

```r
real_w <- TukeyHSD(anova_mod, ordered = TRUE)
real_w$treatment
```

```
##        diff        lwr      upr        p adj
## a3-s   1.00 -0.2594887 2.259489 1.784060e-01
## a2-s   1.75  0.4905113 3.009489 2.428628e-03
## a1-s   2.78  1.5205113 4.039489 1.200843e-06
## a4-s   2.97  1.7105113 4.229489 2.780828e-07
## a2-a3 0.75 -0.5094887 2.009489 4.490082e-01
## a1-a3 1.78  0.5205113 3.039489 1.980323e-03
## a4-a3 1.97  0.7105113 3.229489 5.243121e-04
## a1-a2 1.03 -0.2294887 2.289489 1.563263e-01
## a4-a2 1.22 -0.0394887 2.479489 6.176067e-02
## a4-a1 0.19 -1.0694887 1.449489 9.927171e-01
```

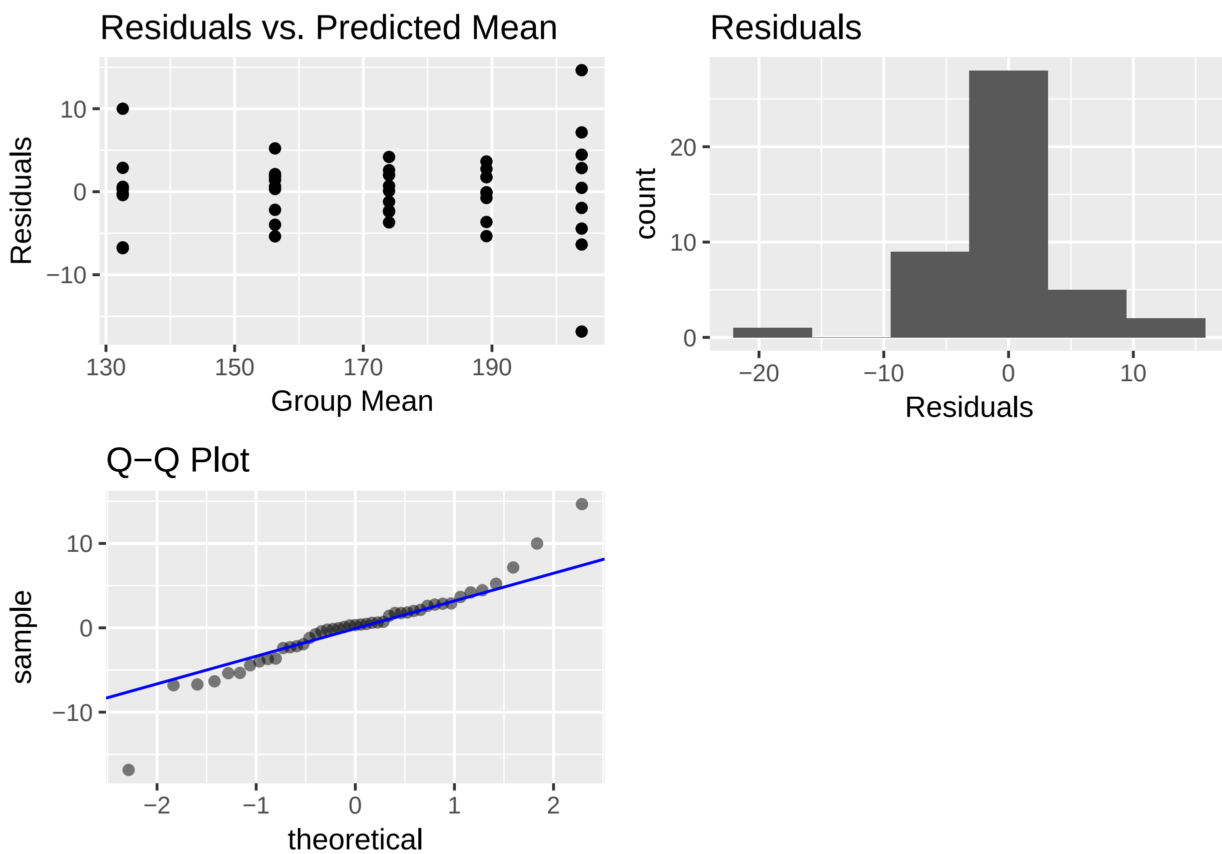
```r
plot(real_w)
```




9.13 d) Use Dunnett's to see if any of the new agents have significantly larger mean weights loss as compared to the standard agent. alpha = 0.05

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: User-defined Contrasts
##
```

```
##
## Fit: aov(formula = wt_loss ~ treatment, data = gather_frame)
##
## Linear Hypotheses:
##              Estimate Std. Error t value Pr(>|t|)
## a1 - s == 0    2.7800     0.4433   6.272   <0.001 ***
## a2 - s == 0    1.7500     0.4433   3.948   <0.001 ***
## a3 - s == 0    1.0000     0.4433   2.256    0.093 .
## a4 - s == 0    2.9700     0.4433   6.700   <0.001 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```




b) Perform an Anova

```
anova_lenses
```

```
## Call:
##    aov(formula = lov ~ Supplier, data = gather_lenses)
##
## Terms:
##                 Supplier Residuals
## Sum of Squares  28024.350  1053.789
## Deg. of Freedom         4        40
##
## Residual standard error: 5.132711
## Estimated effects may be unbalanced
```

```r
summary(anova_lenses)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Supplier     4  28024    7006   265.9 <2e-16 ***
## Residuals   40   1054      26
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c) Run a Kruskal-Wallis

```r
kw_lenses <- kruskal.test(lov ~ Supplier, data = gather_lenses)
kw_lenses
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  lov by Supplier
## Kruskal-Wallis chi-squared = 41.596, df = 4, p-value = 2.023e-08
```

```r
kw_lenses$statistic
```

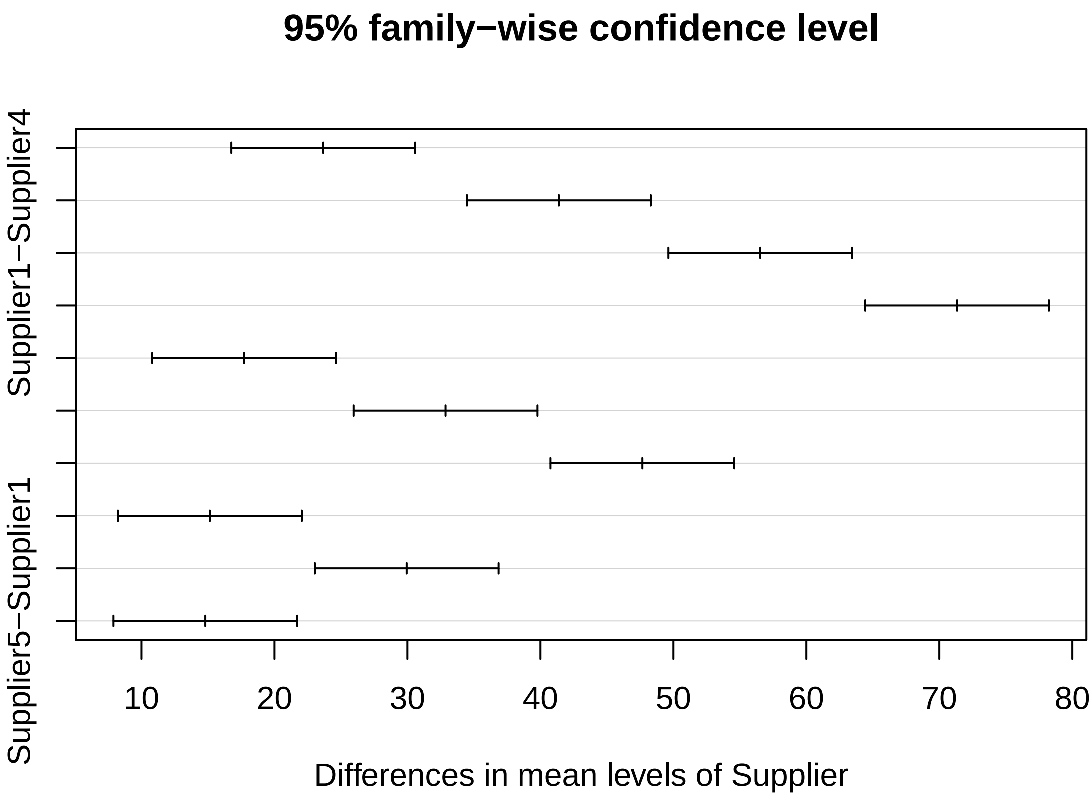
```
## Kruskal-Wallis chi-squared
##                    41.5963
```

d) Use Tukey's W to find out which pairs are significantly different.

```r
real_lenses <- TukeyHSD(anova_lenses, ordered = TRUE)
real_lenses$treatment
```

```
## NULL
```

```r
plot(real_lenses)
```




```
real_lenses$Supplier
```

```
##                          diff       lwr      upr        p adj
## Supplier3-Supplier4 23.66667 16.756116 30.57722 3.656697e-11
## Supplier2-Supplier4 41.38889 34.478338 48.29944 4.636291e-13
## Supplier1-Supplier4 56.53333 49.622783 63.44388 4.636291e-13
## Supplier5-Supplier4 71.33333 64.422783 78.24388 4.636291e-13
## Supplier2-Supplier3 17.72222 10.811672 24.63277 6.520741e-08
## Supplier1-Supplier3 32.86667 25.956116 39.77722 4.682921e-13
## Supplier5-Supplier3 47.66667 40.756116 54.57722 4.636291e-13
## Supplier1-Supplier2 15.14444  8.233894 22.05499 1.978429e-06
## Supplier5-Supplier2 29.94444 23.033894 36.85499 5.020429e-13
## Supplier5-Supplier1 14.80000  7.889450 21.71055 3.129018e-06
```

f) Run a KW pairwise comparison on the different suppliers to see if there's a difference in variability of power

```
ranks_1 <- c(39,32, 33, 28, 29, 34, 35, 31, 36)
ranks_2 <- c(19, 20, 21, 22, 23, 24, 25, 26, 27)
ranks_3 <- c(10, 11, 12, 13, 14, 15, 16, 17, 18)
ranks_4 <- c(1, 2, 3, 4, 5, 6, 7, 8, 9)
ranks_5 <- c(38, 39, 30, 40, 41, 42, 43, 44, 45)
ranks_table = data.frame(cbind(ranks_1, ranks_2, ranks_3, ranks_4, ranks_5))
colnames(ranks_table) <- c("Ranks_Group1", "Ranks_Group2", "Ranks_Group3", "Ranks_Group4", "Ranks_Group5
```