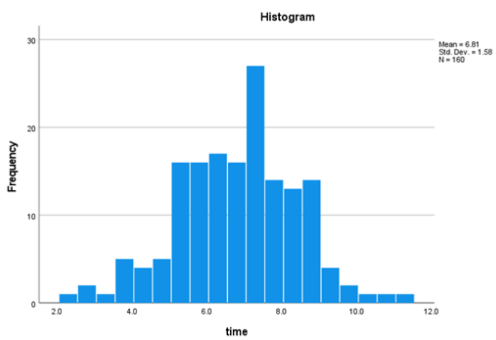
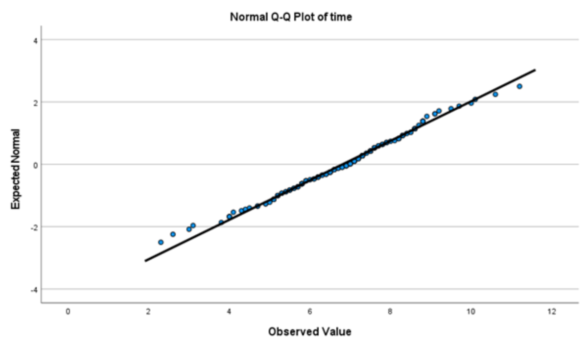
Assignment 8

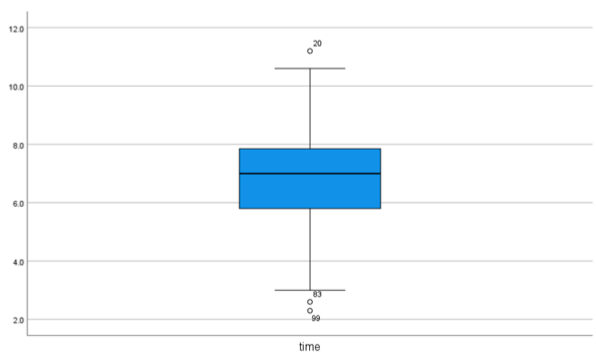
Sainivedhitha Arunajatesan

Exercise 1 (Cats & Flavors)

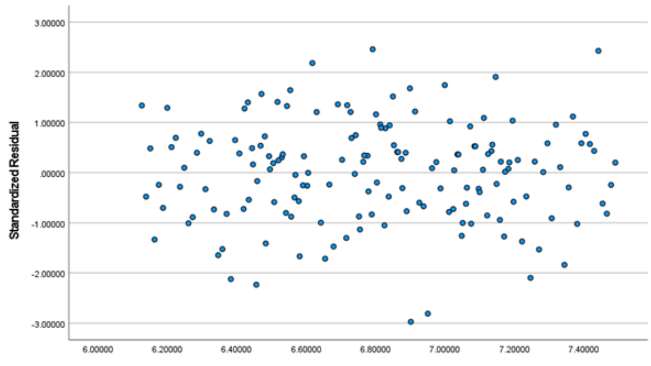
A. The data set was checked for usage in ANOVA test. Based on the plots, it can be seen that the data are normally distributed with a few outliers (20,83,99).







Now we proceed with the ANOVA test.



The random scatter plot of unstandardized predictive value and standardized residual proves that there is a normal distribution but no interaction between the dependent variable (time) and independent variables (cats and flavors).

ANOVA WITHOUT INTERACTION

Tests of Between-Subjects Effects

Dependent Variable: time

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	267.906[a]	42	6.379	5.789	.000
Intercept	7416.091	1	7416.091	6730.745	.000
flavor	19.479	3	6.493	5.893	.001
cat	248.427	39	6.370	5.781	.000
Error	128.913	117	1.102		
Total	7812.910	160			
Corrected Total	396.819	159			

a. R Squared = .675 (Adjusted R Squared = .559)

Flavor: The critical value was found to be between 2.60 and 2.70 from the F distribution table. We reject the null hypothesis (H_0) if F $\geq$ c = [2.60,2.70] (F^3_{117} & $\alpha = 5\%$). From the table of test between the subjects provided above, the value of F statistic for flavors is 5.893. Therefore, we **reject** the null hypothesis that the flavors has no effect on describing the time. The rejection of null hypothesis can also be proved by the p value (0.001) in the table.

Cats: The rejection of null hypothesis can be proved by the p value (0.000) in the table. The value of sig is less than 0.05 and so we **reject** the null hypothesis that the cats has no effect on describing the time.
In conclusion, both the flavor and cats have an effect on the time as independent factors.

ANOVA WITH INTERACTION

Tests of Between-Subjects Effects

Dependent Variable: time

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	396.819[a]	159	2.496	.	.
Intercept	7416.091	1	7416.091	.	.
flavor	19.479	3	6.493	.	.
cat	248.427	39	6.370	.	.
flavor * cat	128.913	117	1.102	.	.
Error	.000	0	.		
Total	7812.910	160			
Corrected Total	396.819	159			

a. R Squared = 1.000 (Adjusted R Squared = .)

An increased R squared value indicates increased interaction between the dependent variable and the independent variable. But in this case (flavor*cat), the R squared value of 1.000 makes it not possible to use.

B. A Tukey, Scheffe and Bonferroni Post-Hoc test results in the following table
Based on the results provided above, the significant combinations of flavors (i-j) are:

Multiple Comparisons

Dependent Variable: time

	(I) flavor	(J) flavor	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	1	2	-.582	.2347	.068	-1.194	.029
		3	-.750*	.2347	.010	-1.362	-.138
		4	-.930*	.2347	.001	-1.542	-.318
	2	1	.582	.2347	.068	-.029	1.194
		3	-.168	.2347	.892	-.779	.444
		4	-.348	.2347	.452	-.959	.264
	3	1	.750*	.2347	.010	.138	1.362
		2	.168	.2347	.892	-.444	.779
		4	-.180	.2347	.869	-.792	.432
	4	1	.930*	.2347	.001	.318	1.542
		2	.348	.2347	.452	-.264	.959
		3	.180	.2347	.869	-.432	.792
Scheffe	1	2	-.582	.2347	.110	-1.248	.083
		3	-.750*	.2347	.020	-1.416	-.084
		4	-.930*	.2347	.002	-1.596	-.264
	2	1	.582	.2347	.110	-.083	1.248
		3	-.168	.2347	.917	-.833	.498
		4	-.348	.2347	.536	-1.013	.318
	3	1	.750*	.2347	.020	.084	1.416
		2	.168	.2347	.917	-.498	.833
		4	-.180	.2347	.899	-.846	.486
	4	1	.930*	.2347	.002	.264	1.596
		2	.348	.2347	.536	-.318	1.013
		3	.180	.2347	.899	-.486	.846
Bonferroni	1	2	-.582	.2347	.087	-1.212	.047
		3	-.750*	.2347	.011	-1.380	-.120
		4	-.930*	.2347	.001	-1.560	-.300
	2	1	.582	.2347	.087	-.047	1.212
		3	-.168	.2347	1.000	-.797	.462
		4	-.348	.2347	.849	-.977	.282
	3	1	.750*	.2347	.011	.120	1.380
		2	.168	.2347	1.000	-.462	.797
		4	-.180	.2347	1.000	-.810	.450
	4	1	.930*	.2347	.001	.300	1.560
		2	.348	.2347	.849	-.282	.977
		3	.180	.2347	1.000	-.450	.810

Based on observed means.
The error term is Mean Square(Error) = 1.102.
*. The mean difference is significant at the .05 level.

- Tukey: 1-3; 1-4; 3-1; 4-1
- Scheffe: 1-3; 1-4; 3-1; 4-1
- Bonferroni: 1-3; 1-4; 3-1; 4-1

In conclusion, **flavor 2** has **no significant impact** as flavor i or j. Flavors 1,3 and 4 have a significance effect with respect to time.

C. Bonferroni Confidence Intervals for $\beta_1 - \beta_3$:

$$\begin{aligned}
\text{Confidence interval} &= 100 * (1 - (\text{significance threshold/no. of different possible variable combinations})) \\
&= 100 * (1 - \frac{0.05}{6}) = 99.17\% \\
t &= \text{flavors} = 4 \\
r &= \text{cats} = 40 \\
df &= (r-1) * (t-1) = 39 * 3 = 117 \\
\text{Probability} &= \frac{0.05}{6} = 0.0083
\end{aligned}$$

For determining c we use the program Excel, we found **c to be 2.685** (applying statistical function T.INV.2T with probability 0.05/6 and df=117).
For determining $\beta_1 - \beta_3$ we use the program SPSS, we found $\beta_1 - \beta_3$ **to be -0.750** (the value can be found in the table of multiple comparisons in the previous question).

Standard Error:

$$se(\beta_1 - \beta_3) = S * \sqrt{(2/r)}$$

From table of tests of between subjects effects without interaction, $S^2 = 1.102$

$$\begin{aligned}
S &= \sqrt{1.102} = 1.05 \\
se(\beta_1 - \beta_3) &= 1.05 * \sqrt{(2/40)} = 0.235
\end{aligned}$$

The Bonferroni confidence interval for $\beta_1 - \beta_3$ is thus,

$$\begin{aligned}
CI &= (\beta_1 - \beta_3) \pm c * se(\beta_1 - \beta_3) \\
&= (-0.750 - 2.685 * 0.235, -0.750 + 2.685 * 0.235) \\
&= (-1.38, -0.12)
\end{aligned}$$

These values for the confidence interval are equal to the calculated values using SPSS which are: **(-1.380, -0.120)**.

Tukey Confidence Intervals for $\beta_1 - \beta_3$:

$$\begin{aligned}
q &= \text{c-value from studentized range table} \\
k &= flavors = 4 \\
r &= cats = 40 \\
df &= v = 117 \\
S &= 1.05 \\
q(v = 60) &= 3.74 \\
q(v = 120) &= 3.68 \\
q(v = 117) &= 3.683 (\text{Using Linear Interpolation}) \\
q * (S/\sqrt{r}) &= 3.683 * (1.05/\sqrt{40}) = 0.612
\end{aligned}$$

The Tukey confidence interval for $\beta_1 - \beta_3$ is thus,

$$\begin{aligned}
CI &= (\beta_1 - \beta_3) \pm q * (S/\sqrt{r}) \\
&= (-0.750 - 0.612, -0.750 + 0.612) \\
&= (-1.362, -0.138)
\end{aligned}$$

These values for the confidence interval are equal to the calculated values using SPSS which are: **(-1.362, -0.138)**.

Scheffe Confidence Intervals for $\beta_1 - \beta_3$:

$$
\begin{aligned}
c &= \text{c-value from F distribution table} \\
t &= flavors = 4; df = 3 \\
r &= cats = 40; df = 39 \\
df(error) &= 117 \\
S &= 1.05 \\
c(F^3_{100}) &= 2.70 \\
c(F^3_{\infty}) &= 2.60 \\
c(F^3_{117}) &= 2.69 \text{(Using Linear Interpolation)} \\
S * \sqrt{2c(t-1)/r} &= 1.05 * \sqrt{2 * 2.69(3)/40} = 0.667
\end{aligned}
$$

The Scheffe confidence interval for $\beta_1 - \beta_3$ is thus,

$$
\begin{aligned}
CI &= (\beta_1 - \beta_3) \pm S * \sqrt{2c(t-1)/r} \\
&= (-0.750 - 0.667, -0.750 + 0.667) \\
&= (-1.417, -0.083)
\end{aligned}
$$

These values for the confidence interval are equal to the calculated values using SPSS which are: **(-1.416, -0.084)**.

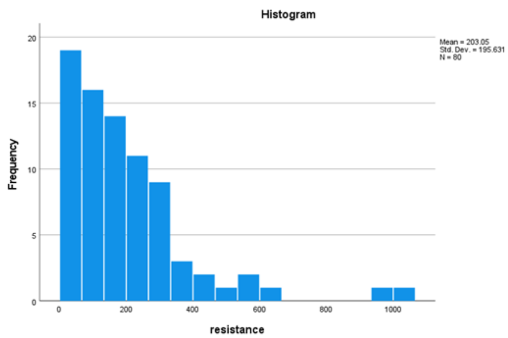
	Manual Calculation	Δint	SPSS	Δint
Bonferroni	(-1.38,-0.12)	1.26	(-1.38,-0.12)	1.26
Tukey	(-1.362,-0.138)	1.224	(-1.362,-0.138)	1.224
Scheffe	(-1.417,-0.083)	1.334	(-1.416,-0.084)	1.334

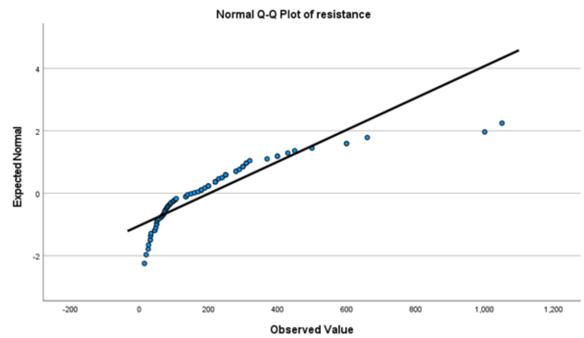
The manual calculations of confidence intervals are almost the same as the calculations from SPSS. Among these values, the smallest interval (Δint = 1.224) was calculated through the **Tukey** method.

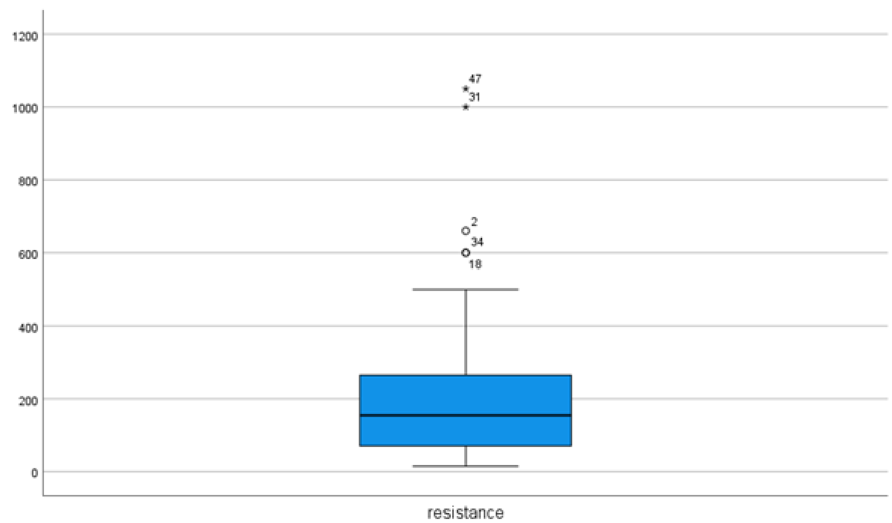
The smallest interval can be predicted using the **significance value** in the multiple comparisons table. The lowest significance value gives the smallest interval. In this case, Tukey HSD has the lowest significance value (0.010) which in turn has the lowest interval.

Exercise 2 (Resistance)

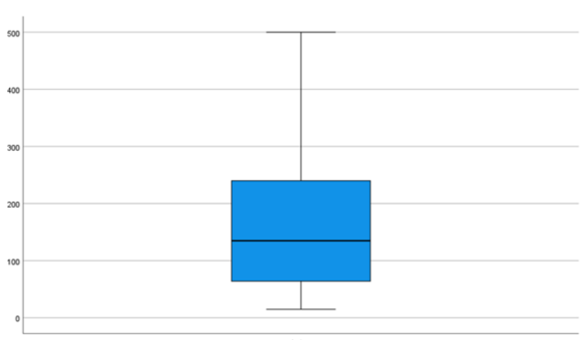
A. The data set was checked for usage in ANOVA test. Based on the plots, it can be seen that the data are normally distributed with a few outliers (2,18,31,34,47).

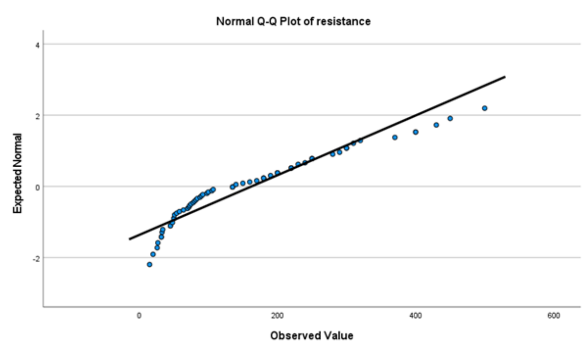






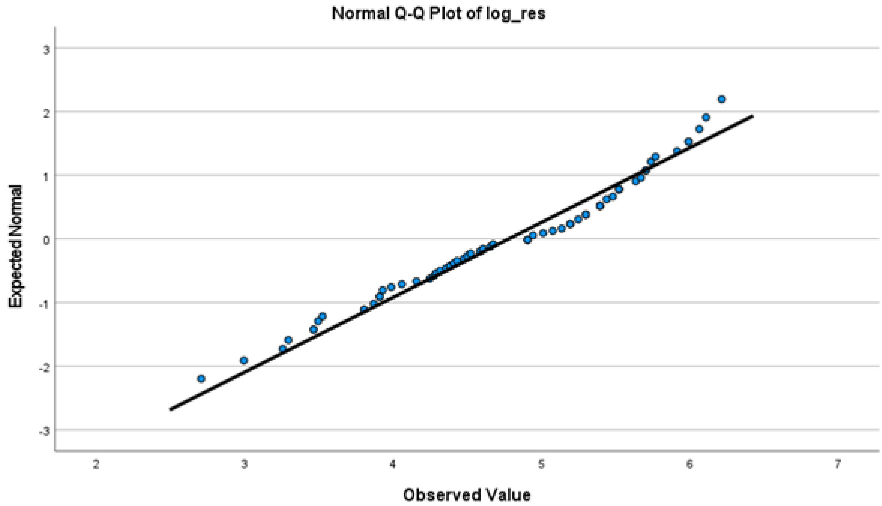
These results prove the presence of 5 outliers (2 from subject 15 and 3 from subject 2). So, I decided to delete the measurements of both the subjects. The box plot and QQ-plot after removing the outliers is given below.



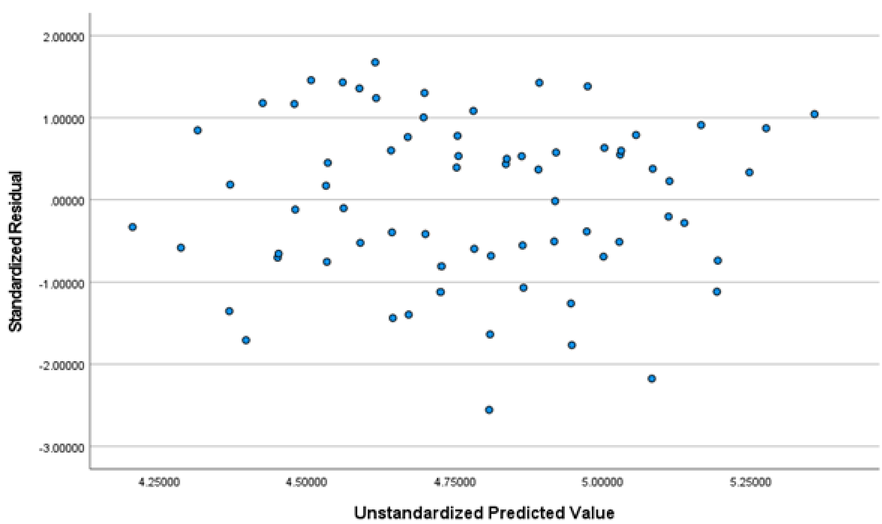


Even after removing the outliers, the QQ-plot has not changed. So, we proceed with the log transformation of the resistance value to distribute normally.

The QQ-plot after log transformation seems to be normal.



Now we proceed with the ANOVA test.



The random scatter plot of unstandardized predictive value and standardized residual proves that there is a normal distribution.

ANOVA WITHOUT INTERACTION

Electrode Type: The critical value was found to be between 2.53 and 2.61 from the F distribution table. We reject the null hypothesis (H_0) if F $\geq$ c = [2.53,2.61] (F^4_{52} & $\alpha = 5\%$). From the table of test between the subjects provided below, the value of F statistic for electrode type is 1.500. Therefore, we **accept** the null hypothesis that the electrode type has no effect on describing the resistance. The acceptance of null hypothesis can also be proved by the p value (0.216) in the table.

Subject: The rejection of null hypothesis can be proved by the p value (0.000) in the table. The value of sig is less than 0.05 and so we **reject** the null hypothesis that the subject has no effect on describing the strength.

Tests of Between-Subjects Effects

Dependent Variable: log_res

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	26.161[a]	17	1.539	3.358	.000
Intercept	1600.172	1	1600.172	3491.534	.000
subject	23.411	13	1.801	3.929	.000
electrodetype	2.749	4	.687	1.500	.216
Error	23.832	52	.458		
Total	1650.164	70			
Corrected Total	49.992	69			

a. R Squared = .523 (Adjusted R Squared = .367)

In conclusion, only the subject has an effect on the resistance.

ANOVA WITH INTERACTION

Tests of Between-Subjects Effects

Dependent Variable: log_res

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	49.992[a]	69	.725	.	.
Intercept	1600.172	1	1600.172	.	.
subject	23.411	13	1.801	.	.
electrodetype	2.749	4	.687	.	.
subject * electrodetype	23.832	52	.458	.	.
Error	.000	0	.		
Total	1650.164	70			
Corrected Total	49.992	69			

a. R Squared = 1.000 (Adjusted R Squared = .)

An increased R squared value indicates increased interaction between the dependent variable and the independent variable. But in this case (electrode type*subject), the R squared value of 1.000 makes it not possible to use.

Still we are interested in finding the post hoc tests for the electrode type.

A Tukey, Scheffe and Bonferroni Post-Hoc test results in the following table

Multiple Comparisons

Dependent Variable: log_res

	(I) electrodetype	(J) electrodetype	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	1	2	-.3763	.25587	.586	-1.0994	.3467
		3	-.1252	.25587	.988	-.8483	.5978
		4	.1678	.25587	.965	-.5552	.8909
		5	.1366	.25587	.983	-.5864	.8596
	2	1	.3763	.25587	.586	-.3467	1.0994
		3	.2511	.25587	.862	-.4719	.9742
		4	.5442	.25587	.224	-.1789	1.2672
		5	.5129	.25587	.278	-.2101	1.2360
	3	1	.1252	.25587	.988	-.5978	.8483
		2	-.2511	.25587	.862	-.9742	.4719
		4	.2931	.25587	.782	-.4300	1.0161
		5	.2618	.25587	.843	-.4612	.9849
	4	1	-.1678	.25587	.965	-.8909	.5552
		2	-.5442	.25587	.224	-1.2672	.1789
		3	-.2931	.25587	.782	-1.0161	.4300
		5	-.0312	.25587	1.000	-.7543	.6918
	5	1	-.1366	.25587	.983	-.8596	.5864
		2	-.5129	.25587	.278	-1.2360	.2101
		3	-.2618	.25587	.843	-.9849	.4612
		4	.0312	.25587	1.000	-.6918	.7543
Scheffe	1	2	-.3763	.25587	.706	-1.1935	.4408
		3	-.1252	.25587	.993	-.9424	.6919
		4	.1678	.25587	.979	-.6493	.9850
		5	.1366	.25587	.990	-.6806	.9538
	2	1	.3763	.25587	.706	-.4408	1.1935
		3	.2511	.25587	.914	-.5660	1.0683
		4	.5442	.25587	.352	-.2730	1.3613
		5	.5129	.25587	.414	-.3042	1.3301
	3	1	.1252	.25587	.993	-.6919	.9424
		2	-.2511	.25587	.914	-1.0683	.5660
		4	.2931	.25587	.858	-.5241	1.1102
		5	.2618	.25587	.901	-.5553	1.0790
	4	1	-.1678	.25587	.979	-.9850	.6493
		2	-.5442	.25587	.352	-1.3613	.2730
		3	-.2931	.25587	.858	-1.1102	.5241
		5	-.0312	.25587	1.000	-.8484	.7859
	5	1	-.1366	.25587	.990	-.9538	.6806
		2	-.5129	.25587	.414	-1.3301	.3042
		3	-.2618	.25587	.901	-1.0790	.5553
		4	.0312	.25587	1.000	-.7859	.8484
Bonferroni	1	2	-.3763	.25587	1.000	-1.1265	.3738
		3	-.1252	.25587	1.000	-.8754	.6249
		4	.1678	.25587	1.000	-.5823	.9180
		5	.1366	.25587	1.000	-.6136	.8868
	2	1	.3763	.25587	1.000	-.3738	1.1265
		3	.2511	.25587	1.000	-.4990	1.0013
		4	.5442	.25587	.382	-.2060	1.2943
		5	.5129	.25587	.502	-.2372	1.2631
	3	1	.1252	.25587	1.000	-.6249	.8754
		2	-.2511	.25587	1.000	-1.0013	.4990
		4	.2931	.25587	1.000	-.4571	1.0432
		5	.2618	.25587	1.000	-.4883	1.0120
	4	1	-.1678	.25587	1.000	-.9180	.5823
		2	-.5442	.25587	.382	-1.2943	.2060
		3	-.2931	.25587	1.000	-1.0432	.4571
		5	-.0312	.25587	1.000	-.7814	.7189
	5	1	-.1366	.25587	1.000	-.8868	.6136
		2	-.5129	.25587	.502	-1.2631	.2372
		3	-.2618	.25587	1.000	-1.0120	.4883
		4	.0312	.25587	1.000	-.7189	.7814

Based on observed means.
The error term is Mean Square(Error) = .458.

Based on the results provided above, there are significant combinations of electrode types (i-j) available.

B. Confidence Intervals for new subjects and electrodes

Usually the confidence intervals will be calculated for a value of $\beta_j - \beta_k$. In this case, the value of $\beta_j - \beta_k$ are not known but it is known that they are similar for all the three methods. So, we omit the term $\beta_j - \beta_k$ in all the formula and proceed with the second term.

Bonferroni Confidence Intervals:

$$\begin{aligned}
\text{Confidence interval} &= 100 * (1 - (\text{significance threshold/no. of different possible variable combinations})) \\
&= 100 * (1 - \frac{0.05}{10}) = 99.5\% \\
t &= \text{electrode type} = 6 \\
r &= \text{subjects} = 20 \\
df &= (r-1) * (t-1) = 19 * 5 = 95 \\
\text{Probability} &= \frac{0.05}{10} = 0.005
\end{aligned}$$

For determining c we use the program Excel, we found **c to be 2.874** (applying statistical function T.INV.2T with probability 0.05/10 and df=95).

The Bonferroni confidence interval for $\beta_j - \beta_k$ is thus,

$$CI = (\beta_j - \beta_k) \pm c * \sqrt{2}$$

As we omit the first term,

$$c * \sqrt{2} = 2.874 * \sqrt{2} = 4.063$$

Tukey Confidence Intervals:

$$\begin{aligned}
k &= \text{electrode type} = 6 \\
r &= \text{subjects} = 20 \\
df = v &= 117
\end{aligned}$$

The Tukey confidence interval for $\beta_j - \beta_k$ is thus,

$$CI = (\beta_j - \beta_k) \pm q$$

As we omit the first term,

$$\begin{aligned}
q &= \text{c-value from studentized range table} \\
q(v = 60) &= 4.16 \\
q(v = 120) &= 4.10 \\
q(v = 95) &= 4.125 (\text{Using Linear Interpolation})
\end{aligned}$$

Scheffe Confidence Intervals:

$$\begin{aligned}
c &= \text{c-value from F distribution table} \\
t &= electrodetype = 6; df = 5 \\
r &= subjects = 20; df = 19 \\
df &= 95
\end{aligned}$$

The Scheffe confidence interval for $\beta_j - \beta_k$ is thus,

$$CI = (\beta_j - \beta_k) \pm \sqrt{2c(t-1)}$$

As we omit the first term,

$$\begin{aligned} c(F^5_{80}) &= 2.33 \\ c(F^5_{100}) &= 2.31 \\ c(F^5_{95}) &= 2.315 \text{(Using Linear Interpolation)} \\ \sqrt{2c(t-1)} &= \sqrt{2 * 2.315(5)} = 4.811 \end{aligned}$$

The manual calculations of confidence intervals reveal that the smallest interval (4.063) was calculated through the **Bonferroni** method.

$$\textbf{Bonferroni} < \textbf{Tukey} < \textbf{Scheffe}$$

Therefore, Bonferroni gives the accurate estimation of 95% confidence interval