Final Project

Statistical Analysis on

Carbon Dioxide Uptake in Grass Plants Dataset

BANA 6043 Statistical Computing

Group Members:

Pavithra, Narayanan

M13028352

Varun Raj, Rayabarapu

M12467636

1. Describing the data: Part-A

The CO2 uptake of six plants from Quebec and six plants from Mississippi was measured at different levels of ambient CO2 Concentration. Half the plants of each type were chilled overnight before the experiment was conducted. It consists of 84 rows and 5 columns of data from an experiment on the cold tolerance of the grass species Echinochloa crus-galli.
First, 10 observations from the dataset are shown below Fig 1.1:

	Plant	Type	Treatment	conc	uptake
1	Qn1	Quebec	nonchilled	95	16.0
2	Qn1	Quebec	nonchilled	175	30.4
3	Qn1	Quebec	nonchilled	250	34.8
4	Qn1	Quebec	nonchilled	350	37.2
5	Qn1	Quebec	nonchilled	500	35.3
6	Qn1	Quebec	nonchilled	675	39.2
7	Qn1	Quebec	nonchilled	1000	39.7
8	Qn2	Quebec	nonchilled	95	13.6
9	Qn2	Quebec	nonchilled	175	27.3
10	Qn2	Quebec	nonchilled	250	37.1

Fig 1.1 First 10 Observations of the CO2 data

Obs	VAR1	Plant	Type_new	Treatment_new	conc_new	uptake
1	1	Qn1	Q	NC	1	16
2	2	Qn1	Q	NC	2	30.4
3	3	Qn1	Q	NC	3	34.8
4	4	Qn1	Q	NC	4	37.2
5	5	Qn1	Q	NC	5	35.3
6	6	Qn1	Q	NC	6	39.2
7	7	Qn1	Q	NC	7	39.7
8	8	Qn2	Q	NC	1	13.6
9	9	Qn2	Q	NC	2	27.3
10	10	Qn2	Q	NC	3	37.1

Fig 1.2 First 10 Observations of the CO2 data

Source: datasets in R   # of Rows(observations): 84   # of columns: 5

Names and type of Columns:
1. Plant: A categorical variable ordered with levels Qn1 < Qn2 < Qn3 < ... < Mc1 giving a unique identifier for each plant. This is also an Independent variable.
2. Type: This Independent variable represents the Origin of the Quebec or Mississippi and is a categorical variable.
3. Treatment: The plant is or not chilled before subject to different ambient concentration levels of CO2 is provided by this categorical column. And, this is an Independent Variable.
4. Concentration: This provides the information of different levels of the ambient CO2 concentration in (mL/L): 97, 175, 250, 350, 500, 675 and 1000. This is an Independent variable.
5. Uptake: This column will represent the CO2 uptake of the plant in (umol/m2 sec). And, this is a dependent variable.

A little data manipulation is done in R as follows:
Representing concentration levels as "1-7":

conc	conc_new
95	1
175	2
250	3
350	4
500	5
675	6
1000	7

Representing treatment as:

treatment	treatment_new
Chilled	C
nonchilled	NC

Representing type as:

type	type_new
Quebec	Q
Mississippi	M

First, 10 observations from the dataset are shown in Fig 1.2 as SAS dataset after the above manipulations.

Part-B:

Hypothesis:

1. There exists significant difference between the plant's mean CO2 uptake for the two origins (Quebec and Mississippi).

$$\mu_{\text{CO2 uptake of Quebec}} \neq \mu_{\text{CO2 uptake of Mississippi}}$$

2. There exits significant difference between the plant's mean CO2 uptake for the two treatments (Chilled and Non-chilled).

$$\mu_{\text{CO2 uptake of Chilled}} \neq \mu_{\text{CO2 uptake of Nonchilled}}$$

3. There exists significant difference between the plant's mean CO2 uptake for the 7 concentration levels (1-7).

$$\mu_{\text{CO2 uptake of L\_1}} \neq \mu_{\text{CO2 uptake of L\_2}} \neq \mu_{\text{CO2 uptake of L\_3}} \neq \mu_{\text{CO2 uptake of L\_4}} \neq \mu_{\text{CO2 uptake of L\_5}} \neq \mu_{\text{CO2 uptake of L\_6}} \neq \mu_{\text{CO2 uptake of L\_7}}$$

4. There exists a relation between how the differences between the plant's mean CO2 uptake against interaction between type and treatment.

2. Data Summary: Part-A:

The SAS dataset that is created with CO2 uptake data consists of 12 unique identifiers for each plant Qn1 < Qn2 < Qn3 < ... < Mc1, these labelled in such a way that it reflects whether the plant is chilled or non-chilled as well as its origin.
For example: Qn1 denotes the plant sample is from Quebec and is non-chilled before the experiment.
And, each plant is subject to 7 different concentration levels (12x7=84 => Total # of observations).

Part-B: Continuous variable (uptake)

The normal histogram plot of the uptake variable in Fig 2.1 (it is left skewed):

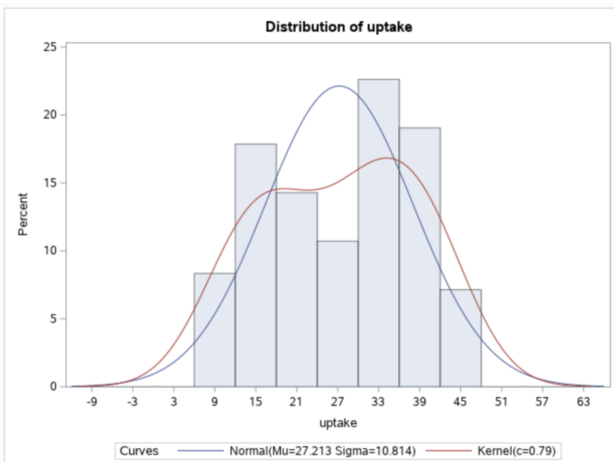

Fig 2.1 Normal Histogram plot of uptake variable

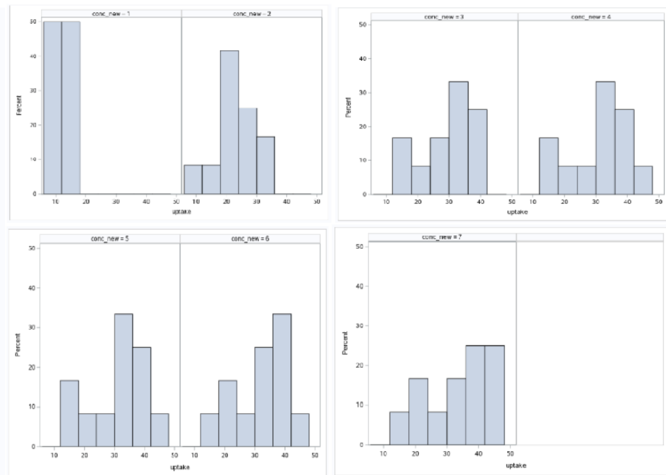

Fig 2.2 Histogram of uptake by conc_new level (1-7 -> L-R)

However, the individual plots of histograms of uptake by concentration level is little skewed as shown in the Fig 2.2. Additionally, the histogram plots for uptake by type and treatment are shown below in Fig 2.3 and Fig 2.4 respectively.
From these plots it is observed that all histograms are skewed. Furthermore, by conducting the normality tests it is revealed that their skewness varied around -1 to 1 but not huge numbers as well as their probability plots are considerably close to the linear line.

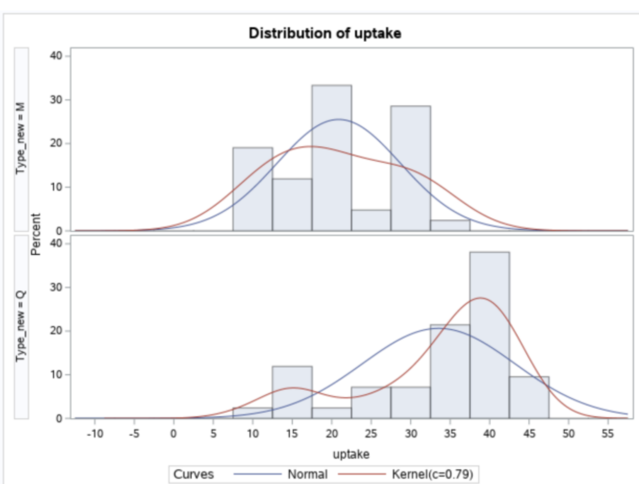

Fig 2.3 Histogram of uptake by type_new (M & Q)

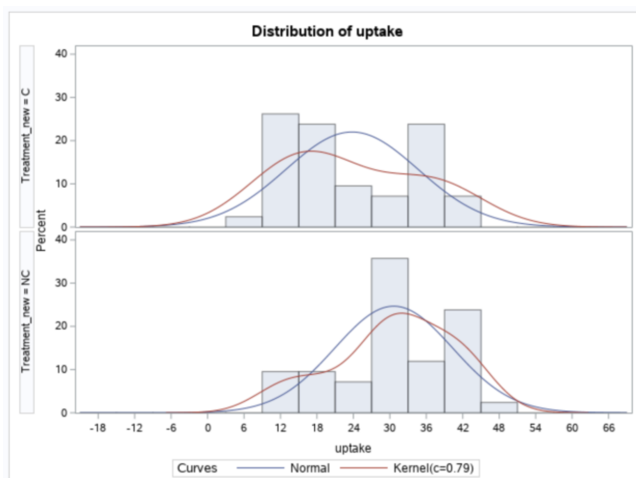

Fig 2.4 Histogram of uptake by treatment_new (C & NC)

Part-C: Categorical variable (type_new, treatment_new and conc_new)

The two-way tables are shown in the Fig 2.5 and Fig 2.6 for the categorical variables type_new, treatment_new and conc_new.

Table of Treatment_new by conc_new								
	conc_new							
Treatment_new	1	2	3	4	5	6	7	Total
C	6 7.14 14.29 50.00	6 7.14 14.29 50.00	6 7.14 14.29 50.00	6 7.14 14.29 50.00	6 7.14 14.29 50.00	6 7.14 14.29 50.00	6 7.14 14.29 50.00	42 50.00
NC	6 7.14 14.29 50.00	6 7.14 14.29 50.00	6 7.14 14.29 50.00	6 7.14 14.29 50.00	6 7.14 14.29 50.00	6 7.14 14.29 50.00	6 7.14 14.29 50.00	42 50.00
Total	12 14.29	12 14.29	12 14.29	12 14.29	12 14.29	12 14.29	12 14.29	84 100.00

Fig 2.5 Two-way table between treatment_new and conc_new

Table of Treatment_new by Type_new			
	Type_new		
Treatment_new	M	Q	Total
C	21 25.00 50.00 50.00	21 25.00 50.00 50.00	42 50.00
NC	21 25.00 50.00 50.00	21 25.00 50.00 50.00	42 50.00
Total	42 50.00	42 50.00	84 100.00

Fig 2.6 Two-way table between type_new and conc_new

From treatment_new and conc_new two-way frequency table it shows that each treatment (chilled and non-chilled) observations are equal for every concentration (6). The same values show up with the type_new vs conc_new variable. And, for type_new and conc_new two-way frequency table it shows that each treatment (chilled and non-chilled) observations are equal for every type (21).

3. Analysis: Part-A

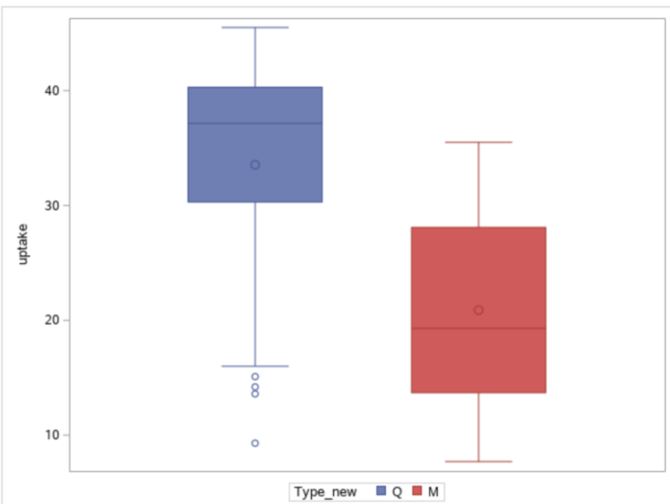

Fig 3.1 Box plots of uptake by type

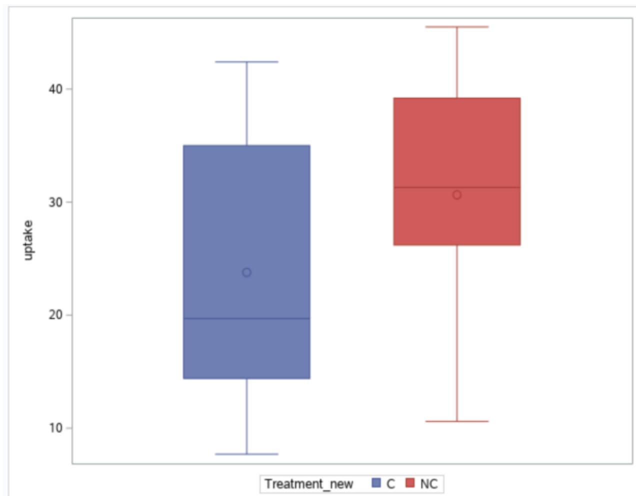

Fig 3.2 Box plots of uptake by treatment

Examining the Hypothesis 1:

There exists significant difference between the plant's mean CO2 uptake for the two origins (Quebec and Mississippi).

$$\mu_{\text{CO2 uptake of Quebec}} \neq \mu_{\text{CO2 uptake of Mississippi}}$$

From the Fig 3.1 box plots of uptake by type, it is evident that mean CO2 uptake of plants vary significantly with origin even though the four outliers are trying to pull the mean of type: Quebec(Q) down. So, the hypothesis seems to be valid. Furthermore, t-test will clarify the doubt.

Examining the Hypothesis 2:

There exits significant difference between the plant's mean CO2 uptake for the two treatments (Chilled and Non-chilled).

$$\mu_{\text{CO2 uptake of Chilled}} \neq \mu_{\text{CO2 uptake of Nonchilled}}$$

From the Fig 3.2 box plots if uptake by treatment, it is not very sure of whether the mean CO2 uptake of plants vary significantly whether chilled or non-chilled. So, the hypothesis claim is at question which should be resolved with the t-test results.

Examining the Hypothesis 3:

There exists significant difference between the plant's mean CO2 uptake for the 7 concentration levels (1-7).

$$\mu_{\text{CO2 uptake of L\_1}} \neq \mu_{\text{CO2 uptake of L\_2}} \neq \mu_{\text{CO2 uptake of L\_3}} \neq \mu_{\text{CO2 uptake of L\_4}} \neq \mu_{\text{CO2 uptake of L\_5}} \neq \mu_{\text{CO2 uptake of L\_6}} \neq \mu_{\text{CO2 uptake of L\_7}}$$

From the Fig 3.3 box plots of uptake by concentration levels, it is not certain that the mean CO2 uptake of plants vary significantly with the ambient CO2 concentration as you can see the level increases after 3 the plants seem to become saturated. So, the hypothesis is to be verified and concluded with the ANOVA test results.

Examining the Hypothesis 4:

There exists a relation between how the differences between the plant's mean CO2 uptake and both type & treatment interaction. This claim is verified by ANOVA two-factor model test. Cannot infer interaction from the boxplots of Fig 3.4.

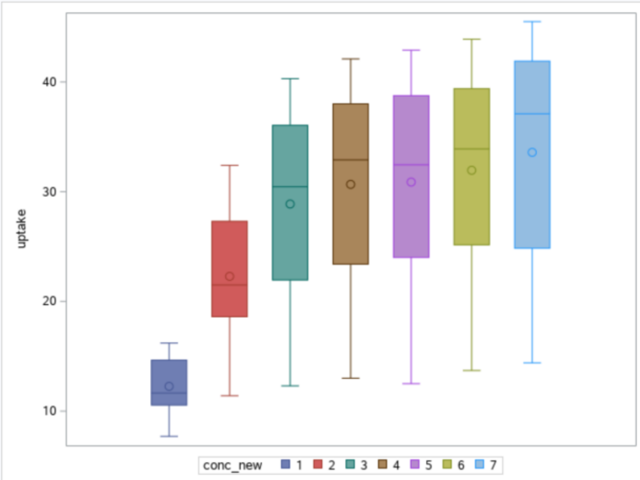

Fig 3.3 Box plots of uptake by concentration levels

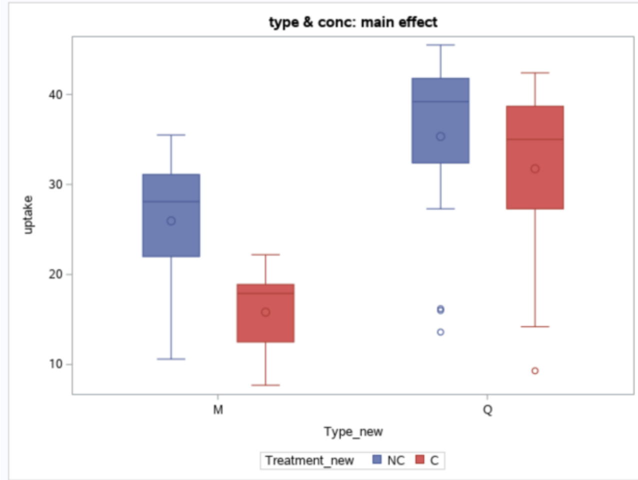

Fig 3.4 Box plots of uptake by type and treatment

Part-B: Test the Hypothesis claims

Testing the Hypothesis 1: t-test on uptake against type

α: 0.05 **corresponding critical value: 1.99*(for DF: 85)**

Null Hypothesis:

$\mu_{\text{CO2 uptake of Quebec}} = \mu_{\text{CO2 uptake of Mississippi}}$

Alternate Hypothesis:

$\mu_{\text{CO2 uptake of Quebec}} \neq \mu_{\text{CO2 uptake of Mississippi}}$

Results of t-test on uptake against type:

The TTEST Procedure

Variable: uptake

Type_new	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
M		42	20.8833	7.8158	1.2060	7.7000	35.5000
Q		42	33.5429	9.6738	1.4927	9.3000	45.5000
Diff (1-2)	Pooled		-12.6595	8.7940	1.9190		
Diff (1-2)	Satterthwaite		-12.6595		1.9190		

Type_new	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
M		20.8833	18.4478	23.3189	7.8158	6.4309	9.9664
Q		33.5429	30.5283	36.5574	9.6738	7.9597	12.3357
Diff (1-2)	Pooled	-12.6595	-16.4770	-8.8420	8.7940	7.6297	10.3810
Diff (1-2)	Satterthwaite	-12.6595	-16.4796	-8.8395			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	82	-6.60	<.0001
Satterthwaite	Unequal	78.533	-6.60	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	41	41	1.53	0.1763

Fig 3.5 t-test results of uptake against type

From the above t-test results, the following can be inferred:

1. Since, the p-value is <0.0001 less **α** than for both methods (Pooled—assumes equal variances and Satterthwaite— assumes unequal variances), a statistically significant result that the **null hypothesis is rejected in favor of alternative hypothesis**.
2. Furthermore, the t-value -6.60 (Pooled) and -6.60 (Satterthwaite) which is <-1.99 (two-sided critical value), which also signifies that this difference in mean uptake is practically significant.
3. The mean difference is also given in the results as around -12.65 with 95% CL mean as -16.47 and -8.84. (NOTE: The negative sign signifies that mean uptake of Q type is more than M type).

Testing the Hypothesis 2: t-test on uptake against treatment

α: 0.05 **corresponding critical value: 1.99*(for DF: 85)**

Null Hypothesis:

$\mu_{\text{CO2 uptake of Chilled}} = \mu_{\text{CO2 uptake of Nonchilled}}$

Alternate Hypothesis:

$\mu_{\text{CO2 uptake of Chilled}} \neq \mu_{\text{CO2 uptake of Nonchilled}}$

Results of t-test on uptake against treatment:

The TTEST Procedure

Variable: uptake

Treatment_new	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
C		42	23.7833	10.8843	1.6795	7.7000	42.4000
NC		42	30.6429	9.7050	1.4975	10.6000	45.5000
Diff (1-2)	Pooled		-6.8595	10.3115	2.2502		
Diff (1-2)	Satterthwaite		-6.8595		2.2502		

Treatment_new	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
C		23.7833	20.3915	27.1751	10.8843	8.9557	13.8793
NC		30.6429	27.6186	33.6671	9.7050	7.9853	12.3755
Diff (1-2)	Pooled	-6.8595	-11.3358	-2.3832	10.3115	8.9463	12.1724
Diff (1-2)	Satterthwaite	-6.8595	-11.3367	-2.3824			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	82	-3.05	0.0031
Satterthwaite	Unequal	80.945	-3.05	0.0031

Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	41	41	1.26	0.4660

Fig 3.6 t-test results of uptake against treatment

From the above t-test results, the following can be inferred:

1. Since, the p-value is 0.0031 less **α** than for both methods (Pooled—assumes equal variances and Satterthwaite— assumes unequal variances), a statistically significant result that the **null hypothesis is rejected in favor of alternative hypothesis**.
2. Furthermore, the t-value -3.05 (Pooled) and -3.05 (Satterthwaite) which is <-1.99 (two-sided critical value), which also signifies that this difference in mean uptake is practically significant.
3. The mean difference is also given in the results as around -6.85 with 95% CL mean as -11.33 and -2.38. (NOTE: The negative sign signifies that mean uptake of NC type is more than C type).

Testing the Hypothesis 3: ANOVA-test on uptake against concentration levels

α: 0.05 **corresponding critical value: 2.22*(for DF: (6,77))**

Null Hypothesis:

$$\mu_{\text{CO2 uptake of L\_1}} = \mu_{\text{CO2 uptake of L\_2}} = \mu_{\text{CO2 uptake of L\_3}} = \mu_{\text{CO2 uptake of L\_4}} = \mu_{\text{CO2 uptake of L\_5}} = \mu_{\text{CO2 uptake of L\_6}} = \mu_{\text{CO2 uptake of L\_7}}$$

Alternative Hypothesis:

$$\mu_{\text{CO2 uptake of L\_1}} \neq \mu_{\text{CO2 uptake of L\_2}} \neq \mu_{\text{CO2 uptake of L\_3}} \neq \mu_{\text{CO2 uptake of L\_4}} \neq \mu_{\text{CO2 uptake of L\_5}} \neq \mu_{\text{CO2 uptake of L\_6}} \neq \mu_{\text{CO2 uptake of L\_7}}$$

Results of ANOVA-test on uptake against concentration levels:

The ANOVA Procedure

Dependent Variable: uptake

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	4068.771429	678.128571	9.26	<.0001
Error	77	5638.204167	73.223431		
Corrected Total	83	9706.975595			

R-Square	Coeff Var	Root MSE	uptake Mean
0.419160	31.44467	8.557069	27.21310

Source	DF	Anova SS	Mean Square	F Value	Pr > F
conc_new	6	4068.771429	678.128571	9.26	<.0001

Fig 3.7 ANOVA-test results of uptake against concentration levels 1-7

From the above ANOVA-test results, the following can be inferred:

1. Since, the p-value is <0.0001 less **α**, a statistically significant result that the **null hypothesis is rejected in favor of alternative hypothesis**.
2. Furthermore, the f-value -9.26 which is >2.22, which also signifies that this difference is practically significant.
3. The R-Square, which is 0.4191, explains the variance about 41.91%.

NOTE: The above test only proves that at least one mean uptake is different from the other but not all are different from each other so, furthermore t-tests must be done by grouping two concentrations at once or from box plot it is evident that uptake reaches saturation after Level 3. So, conducting another ANOVA for the levels 3-7 will also serve the purpose.

Testing the Hypothesis 3: ANOVA-test on uptake against concentration levels 3-7

α: 0.05 **corresponding critical value: 2.54*(for DF: (4,55))**

Null Hypothesis:

$\mu_{CO2\ uptake\ of\ L\_3} = \mu_{CO2\ uptake\ of\ L\_4} = \mu_{CO2\ uptake\ of\ L\_5} = \mu_{CO2\ uptake\ of\ L\_6} = \mu_{CO2\ uptake\ of\ L\_7}$

Alternative Hypothesis:

$\mu_{CO2\ uptake\ of\ L\_3} \neq \mu_{CO2\ uptake\ of\ L\_4} \neq \mu_{CO2\ uptake\ of\ L\_5} \neq \mu_{CO2\ uptake\ of\ L\_6} \neq \mu_{CO2\ uptake\ of\ L\_7}$

Results of ANOVA-test on uptake against concentration levels:

The ANOVA Procedure

Dependent Variable: uptake

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	144.455667	36.113917	0.39	0.8165
Error	55	5123.198333	93.149061		
Corrected Total	59	5267.654000			

R-Square	Coeff Var	Root MSE	uptake Mean
0.027423	30.94382	9.651376	31.19000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
conc_new	4	144.4556667	36.1139167	0.39	0.8165

Fig 3.8 ANOVA-test results of uptake against concentration levels 3-7

From the above ANOVA-test results, the following can be inferred:

1. Since, the p-value is >0.05 greater **α**, a statistically insignificant result that the **null hypothesis is failed to be rejected and needs more evidence**.
2. Furthermore, the f-value 0.39 which is <2.54.
3. The R-Square, which is 0.4191, explains the variance about 2.74%.

Testing the Hypothesis 3: t-tests on uptake against concentration levels 1-2 and 2-3

α: 0.05 **corresponding critical value: 2.22*(for DF: (6,77))**

Null Hypothesis 1:

$\mu_{CO2\ uptake\ of\ L\_1} \neq \mu_{CO2\ uptake\ of\ L\_2}$

Alternative Hypothesis 1:

$\mu_{CO2\ uptake\ of\ L\_1} = \mu_{CO2\ uptake\ of\ L\_2}$

Null Hypothesis 2:

$\mu_{CO2\ uptake\ of\ L\_2} \neq \mu_{CO2\ uptake\ of\ L\_3}$

Alternative Hypothesis 2:

$\mu_{CO2\ uptake\ of\ L\_2} = \mu_{CO2\ uptake\ of\ L\_3}$

Results of t-test on uptake against concentration levels 1-2 and 2-3:

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	22	-5.08	<.0001
Satterthwaite	Unequal	15.038	-5.08	0.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	11	11	5.26	0.0105

Fig 3.9 t-tests results for Hypothesis 1

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	22	-2.08	0.0491
Satterthwaite	Unequal	19.658	-2.08	0.0505

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	11	11	2.05	0.2481

Fig 3.10 t-tests results for Hypothesis 2

From the above ANOVA-test results, the following can be inferred:

1. Since, the p-value is <0.05 greater **α**, that the **null hypothesis 1 is rejected in favor of alternative hypothesis**.
2. Since, the p-value is >0.05 greater **α**, that the **null hypothesis 2 is failed to be rejected and needs more evidence**.

Testing the Hypothesis 4: ANOVA-test on uptake against two-factor main effects model with interaction of type and treatment

α: 0.05 **corresponding critical value: 2.708***(for DF: (3,80))

Null Hypothesis:

$\mu_{\text{CO2 uptake of type and treatment interaction}} = \mu_{\text{CO2 uptake of type and treatment interaction}}$

Alternative Hypothesis:

$\mu_{\text{CO2 uptake of type and treatment interaction}} \neq \mu_{\text{CO2 uptake of type and treatment interaction}}$

Results of ANOVA-test:

The ANOVA Procedure

Dependent Variable: uptake

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	4579.378452	1526.459484	23.82	<.0001
Error	80	5127.597143	64.094964		
Corrected Total	83	9706.975595			

R-Square	Coeff Var	Root MSE	uptake Mean
0.471762	29.41941	8.005933	27.21310

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Type_new	1	3365.534405	3365.534405	52.51	<.0001
Treatment_new	1	988.114405	988.114405	15.42	0.0002
Type_new*Treatment_n	1	225.729643	225.729643	3.52	0.0642

Fig 3.8 ANOVA-test results of uptake against

From the above ANOVA-test results, the following can be inferred:

1. Since, the p-value is 0.0642 little greater than **α**, a statistically insignificant result that the **null hypothesis is failed to be rejected and needs more evidence**.

4. Conclusion

Hypothesis:

1. There exists significant difference between the plant's mean CO2 uptake for the two origins (Quebec and Mississippi).

$\mu_{\text{CO2 uptake of Quebec}} \neq \mu_{\text{CO2 uptake of Mississippi}}$

Conclusions:

Yes, the hypothesis is **true** which is verified with the t-test results.

2. There exits significant difference between the plant's mean CO2 uptake for the two treatments (Chilled and Non-chilled).

$\mu_{\text{CO2 uptake of Chilled}} \neq \mu_{\text{CO2 uptake of Nonchilled}}$

Conclusions:

Yes, the hypothesis is **true** which is verified with the t-test results.

3. There exists significant difference between the plant's mean CO2 uptake for the 7 concentration levels (1-7).

$\mu_{\text{CO2 uptake of L\_1}} \neq \mu_{\text{CO2 uptake of L\_2}} \neq \mu_{\text{CO2 uptake of L\_3}} \neq \mu_{\text{CO2 uptake of L\_4}} \neq \mu_{\text{CO2 uptake of L\_5}} \neq \mu_{\text{CO2 uptake of L\_6}} \neq \mu_{\text{CO2 uptake of L\_7}}$

Conclusions:

Yes, the hypothesis is partially **true**, that is: $\mu_{\text{CO2 uptake of L\_2}} \neq \mu_{\text{CO2 uptake of L\_3}} \neq \mu_{\text{CO2 uptake of L\_4}} \neq \mu_{\text{CO2 uptake of L\_5}} \neq \mu_{\text{CO2 uptake of L\_6}} \neq \mu_{\text{CO2 uptake of L\_7}}$ is not supported by the test results, though the results support $\mu_{\text{CO2 uptake of L\_1}} \neq \mu_{\text{CO2 uptake of L\_2}}$.

4. There exists a relation between how the differences between the plant's mean CO2 uptake against interaction between type and treatment.

Conclusions:

No, the hypothesis is **not true** which is verified with the ANOVA-test two-factor main effects model with interaction of type and treatment.

5. Appendix:

Rcode

```
# loading package
library(dplyr)

# Getting to data to df
df<-CO2

# Changing the concentration levels as described in 1. Part-A
df2 <- df %>% mutate (conc_new=case_when(conc==95 ~ "1", conc==175 ~ "2",conc==250 ~ "3",
conc==350 ~ "4",conc==500 ~ "5", conc==675 ~ "6",conc==1000 ~ "7"))

# Getting relevant data columns
df3<-df2[, c(1:3,6,5)]

# Changing the type and treatment as described in 1. Part-A
df4 <- df3 %>% mutate (Type_new=case_when(Type=="Quebec" ~ "Q", Type=="Mississippi" ~ "M"), Treament_new =
case_when(Treatment=="nonchilled"~"NC",Treatment=="chilled"~"C"))

# Getting relevant data columns
df5<-df4[, c(1,6,7,4,5)]

# Writing df to a file
write.csv (x=df5,file="C:/Users/Varun/Desktop/Spring 2019/Statistical Computing/co2_n.csv")
```

SASCode:

```
/*Contents of the SAS Dataset in library lproject.co2_n*/
proc contents data=lproject.co2_n;
run;

/*Printing the first 10 observations of the SAS Dataset in library lproject.co2_n*/
proc print data=lproject.co2_n(obs=10); run;

/*Creating type_new column Sorted SAS datasets */
proc sort data=lproject.co2_n out=lproject.co2_n_t;
by type_new;
run;

/*Creating treatment_new column Sorted SAS datasets */
proc sort data=lproject.co2_n out=lproject.co2_n_tr;
by Treatment_new;
run;

/*Creating conc_new column Sorted SAS datasets */
proc sort data=lproject.co2_n out=lproject.co2_n_conc;
by conc_new;
run;

/*Normal Histogram Plot of Uptake as well as normal test*/
```

```
ods graphics on;
proc univariate data=lproject.co2_n normaltest;
   histogram uptake /normal kernel;
run;
ods graphics off;

/*Histogram plot of uptake by treatment_new as well as normal test*/
ods graphics on;
proc univariate data=lproject.co2_n normaltest;
   histogram uptake /normal kernel;
   class treatment_new;
run;
ods graphics off;

/*Histogram plot of uptake by type_new as well as normal test*/
ods graphics on;
proc univariate data=lproject.co2_n normaltest;
   histogram uptake /normal kernel;
   class type_new;
run;
ods graphics off;

/*Histogram plot of uptake by conc_new in one panel*/
ods graphics on;
proc sgpanel data=lproject.co2_n;
   panelby conc_new;
   histogram uptake;
run;
ods graphics off;

/*Histogram plot of uptake by type_new in one panel*/
ods graphics on;
proc sgpanel data=lproject.co2_n;
   panelby type_new;
   histogram uptake;
run;
ods graphics off;

/*Histogram plot of uptake by treatment_new in one panel*/
ods graphics on;
proc sgpanel data=lproject.co2_n;
   panelby treatment_new;
   histogram uptake;
run;
ods graphics off;

/*Normality test of uptake var by type_new var with probplot*/
ods graphics on;
ods select Moments TestsForNormality ProbPlot;
```

```
proc univariate data=lproject.co2_n_t normaltest;
  var uptake;
  by type_new;
  probplot uptake / normal (mu=est sigma=est)
                    square;
  label type_new = "type_new" uptake="Uptake";
  inset  mean std / format=6.4;
run;

/*Normality test of uptake var by treatment_new var with probplot*/
ods graphics on;
ods select Moments TestsForNormality ProbPlot;
proc univariate data=lproject.co2_n_tr normaltest;
  var uptake;
  by treatment_new;
  probplot uptake / normal (mu=est sigma=est)
                    square;
  label treatment_new = "treatment_new" uptake="Uptake";
  inset  mean std / format=6.4;
run;

/*All relevant Frequency tables*/
proc freq data=lproject.co2_n;
tables (treatment_new type_new)*(conc_new type_new);
run;

/* Examining the Hypotheses: type ~ uptake graphically*/

ods graphics on;
proc sgplot data=lproject.co2_n ;
vbox uptake/group=type_new;
run;
ods graphics off;

/* Examining the Hypotheses: treatment ~ uptake graphically*/

ods graphics on;
proc sgplot data=lproject.co2_n_tr ;
vbox uptake/group=treatment_new;
run;
ods graphics off;

/* Examining the Hypotheses: concentration ~ uptake graphically*/

ods graphics on;
proc sgplot data=lproject.co2_n ;
vbox uptake/group=conc_new;
run;
ods graphics off;
```

```
/* Examining the Hypotheses: type*concentration ~ uptake graphically*/

ods graphics on;

proc sgplot data=lproject.co2_n;
   vbox uptake / category=type_new
             group=treatment_new
             groupdisplay=clustered;
   title "type & conc: main effect";
run;

ods graphics off;

/* Testing the Hypotheses: type ~ uptake using t-test*/

ods graphics on;

proc ttest data=lproject.co2_n_t plots(only)=(histogram boxplot) alpha=0.05;
   var uptake;
   class type_new;
run;

ods graphics off;

/* Testing the Hypotheses: treatment ~ uptake using t-test*/

ods graphics on;

proc ttest data=lproject.co2_n_tr plots(only)=(histogram boxplot) alpha=0.05;
   var uptake;
   class treatment_new;
run;

ods graphics off;

/* simple ANOVA test: concentration ~ uptake*/
ods graphics on;

proc anova data=lproject.co2_n;
   class conc_new;
   model uptake = conc_new;
   label conc_new = "Concentration";
run;

ods graphics off;

/* From the above anova test we want to find if the differrence in meansof uptake of which concentrations
matters most*/
```

```
/* 1 and 2 conc levels*/

ods graphics on;

proc ttest data=lproject.co2_n_conc(obs=24) plots(only)=(histogram boxplot) alpha=0.05;
   var uptake;
   class conc_new;
run;

ods graphics off;

/* 2 and 3 conc levels*/

ods graphics on;

proc ttest data=lproject.co2_n_conc(firstobs=13 obs=36) plots(only)=(histogram boxplot) alpha=0.05;
   var uptake;
   class conc_new;
run;

ods graphics off;

/* 3 and 4 conc levels*/

ods graphics on;

proc ttest data=lproject.co2_n_conc(firstobs=25 obs=48) plots(only)=(histogram boxplot) alpha=0.05;
   var uptake;
   class conc_new;
run;

ods graphics off;

/* 4 and 5 conc levels*/

ods graphics on;

proc ttest data=lproject.co2_n_conc(firstobs=37 obs=60) plots(only)=(histogram boxplot) alpha=0.05;
   var uptake;
   class conc_new;
run;

ods graphics off;

/* 5 and 6 conc levels*/

ods graphics on;
```

```
proc ttest data=lproject.co2_n_conc(firstobs=49 obs=72) plots(only)=(histogram boxplot) alpha=0.05;
   var uptake;
   class conc_new;
run;

ods graphics off;

/* 6 and 7 conc levels*/

ods graphics on;

proc ttest data=lproject.co2_n_conc(firstobs=61 obs=84) plots(only)=(histogram boxplot) alpha=0.05;
   var uptake;
   class conc_new;
run;

ods graphics off;

/* ANOVA to find is there any difference in means of 2-7 concentration levels*/

ods graphics on;

proc anova data=lproject.co2_n_conc(firstobs=13 obs=84);
   class conc_new;
   model uptake = conc_new;
   label conc_new = "Concentration";
run;

ods graphics off;

/* ANOVA test two-factor model: type+treatment ~ uptake*/

proc anova data=lproject.co2_n;
   class type_new treatment_new;
   model uptake = type_new|treatment_new;
run;
```