Name: Connor Raymond Stewart
ID: 101041126

STAT 2509 Assignment 4:

① Run the Procedures on the provided Regression Data:

a) Use the forward selection procedure using $F_0^* = 4.2$ (to-add-variable):

We must find the Max SSR for our one term models:

X_1 is higest at 14829

thus, we can see that:

$$F = \frac{SSR(X_1)}{MSE(X_1)} = \frac{SSR(X_1)}{MSE(X_1)} = \frac{14829}{119.45210} \doteq \boxed{124.14} > F_0 = \underline{4.2}$$

∴ We keep the X_1 term

We continue the same thing on the two term model:

- X_1 was kept, so we must use a model w/ X_1 in it:

X_1, X_4 has the highest SSR at 15543

thus, we can see that:

$$F_2 = \frac{MSR(X_4|X_1)}{MSE(X_1X_4)} = \frac{SSR(X_4|X_1)/(df_{SSR(X_1X_4)} - df_{SSR(X_1)})}{MSE(X_1X_4)}$$

$$= \frac{[SSR(X_1) - SSR(X_1X_4)]/(12-13)}{69.93149}$$

$$= -\frac{14829 - 15543}{69.93149} = -\frac{-714}{69.93149} \doteq \boxed{10.21} > F_0 = \underline{4.2}$$

∴ We keep the X_1, X_4 term

We continue for the three term model, keeping X_1, X_4:

X_1, X_2, X_4 has the highest SSR at 15848

thus, we can see that:

$$F_3 = \frac{MSR(X_2|X_1X_4)}{MSE(X_1X_2X_4)} = \frac{SSR(X_2|X_1X_4)/(df_{SSR(X_1X_2X_4)} - df_{SSR(X_1X_4)})}{MSE(X_1X_2X_4)}$$

$$= \frac{[SSR(X_1X_4) - SSR(X_1X_2X_4)]/(11-12)}{48.59743} = -\frac{15543 - 15848}{48.59743}$$

$$= \frac{305}{48.59743} \doteq 6.28 > F_0 = 4.2$$

∴ We keep the X_1, X_2, X_4 term

We Continue for the four term model, Keeping X_1, X_2, X_4:
X_1, X_2, X_3, X_4 is the only model with SSR at 15857.8
thus, we can see that:

$$F_4 = \frac{MSR(X_3 | X_1 X_2 X_4)}{MSE(X_1 X_2 X_3 X_4)} = \frac{SSR(X_3 | X_1 X_2 X_4) / (df_{SSR(X_1X_2X_3X_4)} - df_{SSR(X_1X_2X_4)})}{MSE(X_1 X_2 X_3 X_4)}$$

$$= \frac{[SSR(X_1X_2X_4) - SSR(X_1X_2X_3X_4)] / (10-11)}{52.47557}$$

$$= \frac{[15848 - 15857.8] / (-1)}{52.47557}$$

$$\doteq \boxed{0.18687} < F_0 = 4.2$$

$\therefore$ we dont add the X_3 term So the best model is $\boxed{X_1X_2X_4}$

b) use the Backward Elimination Procedure using $F_0^{**} = 4.1$ (to-delete-variable):
note that $(t_0^{**}) = F_0^{**}$ | let $P = 15857.8$, $q = 52.47557$

AS:
$$\frac{MSR(X_{n1} | X_{n2} X_{n3} X_{n4})}{MSE_f} = \frac{(SSR_f - SSR_r)/(df_{SSR_f} - df_{SSR_r})}{MSE_f}$$

$$F_1 = \frac{MSR(X_1 | X_2 X_3 X_4)}{MSE_f} = \frac{(12277 - P)/(10-11)}{q} \doteq 68.20 > F_0$$

$$F_2 = \frac{MSR(X_2 | X_1 X_3 X_4)}{MSE_f} = \frac{(15561 - P)/(10-11)}{q} \doteq 5.66 > F_0$$

$$F_3 = \frac{MSR(X_3 | X_1 X_2 X_4)}{MSE_f} = \frac{(15848 - P)/(10-11)}{q} \doteq \boxed{0.19} < F_0$$

$$F_4 = \frac{MSR(X_4 | X_1 X_2 X_3)}{MSE_f} = \frac{(14997 - P)/(10-11)}{q} \doteq 16.40 > F_0$$

$\hookrightarrow$ $F_3 = 0.19$ is the Smallest So eliminate X_3 from the model. let $q = 48.57743$

AS:
$$\frac{MSR(X_{n1} | X_{n2} X_{n3})}{MSE_f} = \frac{(SSR_f - SSE_r)/(df_{SSR_f} - df_{SSR_r})}{MSE_f}$$

$$F_1 = \frac{MSR(X_1 | X_2 X_4)}{MSE_f} = \frac{(11110 - 15848)/(11-12)}{q} \doteq 97.49 > F_0$$

$$F_2 = \frac{MSR(X_2 | X_1 X_4)}{MSE_f} = \frac{(15543 - 15848)/(11-12)}{q} \doteq 6.28 > F_0$$

$$F_4 = \frac{MSR(X_4 | X_1 X_2)}{MSE_f} = \frac{(14990 - 15848)/(11-12)}{q} \doteq 17.66 > F_0$$

All F values are greater then F_0 So no variables can be deleted.

$\hookrightarrow$ We remove only the X_3 term from the model thus the best model is $\boxed{X_1X_2X_4}$

Note: (For question 1b)

$$F = \frac{MSR_f}{MSE_f} = \frac{SSR_f / df_{SSR_f}}{MSE_f} = \frac{15657.8/4}{52.47559} \doteq 75.55$$

$$F_{(df_{MSR}, df_{MSE})} = F_{(1,13);0.01} = 9.074$$

$df_{MSR} = 10$

$df_{MSR} = \text{total df} - df_{MSE} = 14 - 10 = 4$

1 term model has $df_{MSE} = 13$ so $df_{MSR} = 1$

↳ Total df $= 14$

Since:

$F > R.R \Rightarrow 75.55 > 9.074$

the full model fits & is significant

c) Use the Stepwise regression Procedure using $F_0^* = 4.2$ (to-add) & $F^{**} = 4.1$ (to-delete):

Check if we really need X_1

fit all one term model: forward Selection

↳ we keep X_1

fit all 2-term models: $X_1 \to X_1 X_4$

is X_1 redundent when X_4 is added into the model

$$F_1 = \frac{MSR(X_1|X_4)}{MSE(X_1X_4)} = \frac{[SSR(X_4) - SSR(X_1X_4)]/(12-13)}{69.93149}$$

$$= \frac{[11085 - 15543][-1]}{69.93149} \doteq 63.75$$

Since $F_1 > F_0$, we keep both X_1, X_4

For 3 term models, from forward Selection, we also keep X_2 (X_1 & X_4)

IS X_1 redundent when X_2 & X_4 are in the model:

$$F_1 = \frac{MSR(X_1/X_2X_4)}{MSE(X_1X_2X_4)} = \frac{[SSR(X_2X_4) - SSR(X_1X_2X_4)]/(11-12)}{48.59743}$$

$$= \frac{[11100 - 15848]/(-1)}{48.59743} \doteq 97.70$$

Since $F_1 > F_0$ we keep X_1

IS X_4 redundent when X_1 & X_2 are in the model:

$$F_2 = \frac{MSR(X_4|X_1X_2)}{MSE(X_1X_2X_4)} = \frac{[SSR(X_1X_2) - SSR(X_1X_2X_4)]/(11-12)}{48.59743}$$

$$= \frac{[14830 - 15848]/(11-12)}{48.59743} \doteq 20.95$$

Since $F_1 > F_0$ we keep X_4

So far we have X_1, X_2, X_4

From forward Selection Procedure we already know X_3 is not in the model

∴ Don't test for X_3

∴ The best Set is $\boxed{X_1X_2X_3}$

② Run the tests on the Provided Site data:

CRD, three Steps:

Parametric test → must have Normal distribution

Steps: test all three Conditions

I) Harteley's Test → tests Assumption of Constant varience

II) Main test → tests differences between treatments

III) Tukeys test → Which groups are different

C.R.D Test:

Assumptions:

1) Plots are randomly assigned to 4 independent Swampy Sites
2) taken from 4 normally distributed populations
3) With **equal** varience, σ^2

I) **To check the** Assumption of equal varience use Hartley's test, we need s_i^2's for $i=1,2,3,4$ Where $n_1=n_2=n_3=n_4=6$, $K=4$, $\bar{n}=6$, $[\bar{n}]=6$, $n=24$:

$$S_1^2=\frac{\sum_{j=1}^{n_1} y_{1j}^2-\left(\sum_{j=1}^{n_1} y_{1j}\right)^2/n_1}{n-1} \quad \Big| \quad \sum_{j=1}^{n_1} y_{1j}^2=217.47,\ \sum_{j=1}^{n_1} y_{1j}=36.1=T_1=\bar{y}_1$$

$$=\frac{(5.7^2+6.3^2+6.1^2+6^2+5.8^2+6.2^2)-[5.7+6.3+6.1+6+5.8+6.2]^2/6}{6-1}$$

$$=\frac{217.47-(36.1)^2/6}{5}\doteq \boxed{0.053666\dot{6}66} \leftarrow \text{min}$$

$$S_2^2=\frac{\sum_{j=1}^{n_2} y_{2j}^2-\left(\sum_{j=1}^{n_2} y_{2j}\right)^2/n_2}{n-1} \quad \Big| \quad \sum_{j=1}^{n_2} y_{2j}^2=192.31,\ \sum_{j=1}^{n_2} y_{2j}=33.9=T_2=\bar{y}_2$$

$$=\frac{192.31-(33.9)^2/6}{6-1}=0.155$$

$$S_3^2=\frac{\sum_{j=1}^{n_3} y_{3j}^2-\left(\sum_{j=1}^{n_3} y_{3j}\right)^2/n_3}{n-1} \quad \Big| \quad \sum_{j=1}^{n_3} y_{3j}^2=172.57,\ \sum_{j=1}^{n_3} y_{3j}=32.1=T_3=\bar{y}_3$$

$$=\frac{172.57-(32.1)^2/6}{6-1}=\boxed{0.167} \leftarrow \text{Max}$$

$$S_4^2=\frac{\sum_{j=1}^{n_4} y_{4j}^2-\left(\sum_{j=1}^{n_4} y_{4j}\right)^2/n_4}{n-1} \quad \Big| \quad \sum_{j=1}^{n_4} y_{4j}^2=80.35,\ \sum_{j=1}^{n_4} y_{4j}=21.9=T_4=\bar{y}_4$$

$$=\frac{80.35-79.935}{6-1}=0.083$$

Claim: The variences are not all the Same

$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$

H_a: At least one of the σ^2 does not equal one of the others

$\alpha = 0.01$ OR 0.05

Test-Statistic:

$$F_{max} = \frac{S^2_{max}}{S^2_{min}} = \frac{0.167}{0.0536666\overline{6}} \doteq \boxed{3.1118}$$

Rejection-Region:

We reject H_0 if $F_{max} > F_{max(K, [\bar{n}]-1); \alpha} = \left< \begin{array}{l} F_{max(4,5); 0.01} = 28 \\ F_{max(4,5); 0.05} = 13.7 \end{array} \right]$ for $\alpha = 0.1$ we check both F_{max} tests

$\alpha = 0.1$ So find both 0.01 & 0.05 for the F_{max} test.

↳ $K = \#\text{Sites} = 4$, $[\bar{n}] = \#\text{Samples} = 6$

thus, $F_{max(4, 6-1); \alpha} = F_{max(4,5); \alpha}$

We Can now See:

$$F_{max} < F_{max(4,5); 0.05} < F_{max(4,5); 0.01}$$

↳ Thus, We Cannot Reject H_0 ⇒ the varience of each group is equal

∴ We Conclude at 90% Confidence our assumption of equal varience is not violated; Assumption is Valid.

II) Move on to Main Test:

$$TSS = \sum_{i=1}^{4}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{\left(\sum_{i=1}^{4}\sum_{j=1}^{n_i} y_{ij}\right)^2}{n}$$

$$= \sum_{j=1}^{n_1} y_{1j}^2 + \sum_{j=1}^{n_2} y_{2j}^2 + \sum_{j=1}^{n_3} y_{3j}^2 + \sum_{j=1}^{n_4} y_{4j}^2 - \frac{\left[\sum_{j=1}^{n_1} y_{1j} + \sum_{j=1}^{n_2} y_{2j} + \sum_{j=1}^{n_3} y_{3j} + \sum_{j=1}^{n_4} y_{4j}\right]^2}{24}$$

$$= 217.47 + 192.31 + 172.57 + 80.35 - \frac{[36.1 + 33.9 + 32.1 + 21.9]^2}{24}$$

$$= 662.7 - \frac{(124)^2}{24} = 662.7 - 640\tfrac{2}{3} = 22\tfrac{1}{30} \doteq \boxed{22.0\overline{333}}$$

$$SST_r = \sum_{i=1}^{4} \frac{T_i^2}{n_i} - \frac{\left(\sum_{i=1}^{4}\sum_{j=1}^{n_i} y_{ij}\right)^2}{n}$$

$$= \frac{36.1^2}{6} + \frac{33.9^2}{6} + \frac{32.1^2}{6} + \frac{21.9^2}{6} - \frac{\left[\sum_{j=1}^{n_1} y_{1j} + \sum_{j=1}^{n_2} y_{2j} + \sum_{j=1}^{n_3} y_{3j} + \sum_{j=1}^{n_4} y_{4j}\right]^2}{24}$$

$$= 660\tfrac{61}{150} - \frac{124^2}{24} = \boxed{19.74}$$

$$SSE = TSS - SST_r = 22.0\dot{3}\overline{33} - 19.74 \doteq \boxed{2.29\overline{3333}}$$

$$MST_r = \frac{SST_r}{K-1} = \frac{19.74}{4-1} = \frac{19.74}{3} = \boxed{6.58}$$

$$MSE = \frac{SSE}{n-K} = \frac{2.29\overline{3333}}{24-4} = \boxed{0.114\overline{666}}$$

$$F_T = \frac{MST_r}{MSE} = \frac{6.58}{0.11466\overline{666}} \doteq 57.3837 \doteq \boxed{57.38}$$

ANOVA Table:

Source	df	Sum of Squares	Mean Squares	F-Value	Pr>F
Model	3	19.74	6.58	57.38	<0.0001
Error	20	$2.29\overline{33333}$	$0.11466\overline{666}$		
Corrected Total	23	$22.0\overline{33333}$			

$df_T = K-1 = 4-1 = 3$

$df_E = n-K = 24-4 = 20$

Total $df = df_T + df_E = 20+3 = 23$

$$F_\alpha(K-1, n-K) = F_{0.10(4-1,24-4)} = F_{0.10(3,20)} = 2.38$$

$\hookrightarrow F_T > F_{0.10(3,20)}$ as $57.38 > 2.38$

$\hookrightarrow \therefore$ Since $F_T > F_\alpha \Rightarrow$ Reject H_0

Note: our Claim for the main test

Claim: there is a difference in the mean plant growth for the Swamp Sites

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \Rightarrow$ the means are the same

H_a: at least one of the μ's does not equal another μ

thus, As per above, we can conclude as per a 10% level of Significence there's a difference in the means of the plant growth of the Swamp Sites.

III) Turkeys Test $\rightarrow$ Which groups are different:

1) Calculate $\binom{K}{2} = \binom{4}{2} = 6$ pairs of $|\bar{y}_i - \bar{y}_j|$ for

$H_0: \mu_i = \mu_j \rightarrow$ Sites are the same

$H_a: \mu_i \neq \mu_j \rightarrow$ Sites are different

Claim: the means are different

For $i=1, j=2$: $|36.1 - 33.9|/6 = 2.2/6 \doteq 0.366667$

$i=1, j=3$: $|36.1 - 32.1|/6 = 4/6 \doteq 0.6666667$

$i=1, j=4$: $|36.1 - 21.9|/6 = 14.2/6 \doteq 2.366667$

$i=2, j=3$: $|33.9 - 32.1|/6 = 1.8/6 = 0.3$

$i=2, j=4$: $|33.9 - 21.9|/6 = 12/6 = 2$

$i=3, j=4$: $|32.1 - 21.9|/6 = 10.2/6 = 1.7$

2) $h.s.d. = q_\alpha(k, n-k)\sqrt{\frac{MSE}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$

$= q_{0.10(4,24-4)}\sqrt{\frac{0.1146666666\ldots6}{2}\left(\frac{1}{6}+\frac{1}{6}\right)}$

$= 3.462\left(\frac{43}{2250}\right)^{1/2} \doteq 3.462(0.138242742)$

$\doteq 0.478597066 \doteq \boxed{0.4786}$

thus:

$|\bar{y}_1-\bar{y}_2| \doteq 0.367 < 0.4786 \Rightarrow \mu_1 = \mu_2$

$|y_1-y_3| \doteq 0.67 > 0.4786 \Rightarrow \mu_1 \neq \mu_3$

$|y_1-y_4| \doteq 2.37 > 0.4786 \Rightarrow \mu_1 \neq \mu_4$

$|y_2-y_3| = 0.3 < 0.4786 \Rightarrow \mu_2 = \mu_3$

$|y_2-y_4| = 2 > 0.4786 \Rightarrow \mu_2 \neq \mu_4$

$|y_3-y_4| = 1.7 > 0.4786 \Rightarrow \mu_3 \neq \mu_4$

We thus Know that Sites:

I & III
I & IV
II & IV
III & IV

all differ from each other with a 90% Confidence

So we may reject Ho since $\mu_i \neq \mu_j$ in general

⇒ Accept Ha: Some Sites (namely those above) are different from each other

Since we No longer have a normal distribution, we can no longer use the Parametric test, thus we must use the non-parametric test:

Non-Parametric test (Kruskal-Wallis test):

Assume:

1) C.R.D. (4 independent Samples from 4 treat't populations) with
2) Approximatly the Same Shape & Spread

First, rank observations from Smallest to largest:

	Site1		Site2		Site3		Site4	
$T_{R_1} = 119$	5.7	15.5	6.2	22.5	5.4	12	3.7	4
$T_{R2} = 90.5$	6.3	24	5.3	11	5.0	8	3.2	1
$T_{R3} = 69.5$	6.1	21	5.7	15.5	6.0	19	3.9	5
$T_{R4} = 21$	6.0	19	6.0	19	5.6	14	4.0	6
	5.8	17	5.2	9.5	4.9	7	3.5	2
	6.2	22.5	5.5	13	5.2	9.5	3.6	3

Hypothesis test:

I) Claim: At least one median differs from the rest

$H_0: Md_1 = Md_2 = Md_3 = Md_4$

H_a: At least one of the Md's $\neq$

II) Test-Statistics:

$$H = \frac{12}{n(n+1)}\left[\sum_{i=1}^{4} \frac{T_{R_i}^2}{n_i}\right] - 3(n+1)$$

$$= \frac{12}{24(25)}\left[\frac{119^2}{6} + \frac{90.5^2}{6} + \frac{69.5^2}{6} + \frac{21^2}{6}\right] - 3(25)$$

$$= \frac{12}{600}\left[4603.75\right] - 75$$

$$= 90.075 - 75 = 17.075$$

III) Rejection-Region:

We reject H_0 if $H > \chi^2_{\alpha;(K-1)} = \chi^2_{0.10;(3)} = \boxed{6.25}$

$H = 17.075 > \chi^2_{0.10;(3)} = 6.25 \rightarrow 17.075 > 6.25$ So $H > \chi^2_{0.10;(3)}$

$\rightarrow$ We reject H_0 Since $H > \chi^2_{0.10;(3)}$

IV) Conclusion:

We have rejected H_0, So We accept H_a, that atleast one of the Md's $\neq$ another of the Md's.

The attached SAS Code verifies the above results.

Q2 SAS output

The SAS System

The ANOVA Procedure

Class Level Information		
Class	Levels	Values
site	4	siteI siteII siteIII siteIV

4 sites

Number of Observations Read	24
Number of Observations Used	24

data entries

Connor 101041125 ods graphics off

The SAS System

The ANOVA Procedure

Dependent Variable: growth

MSTr FT

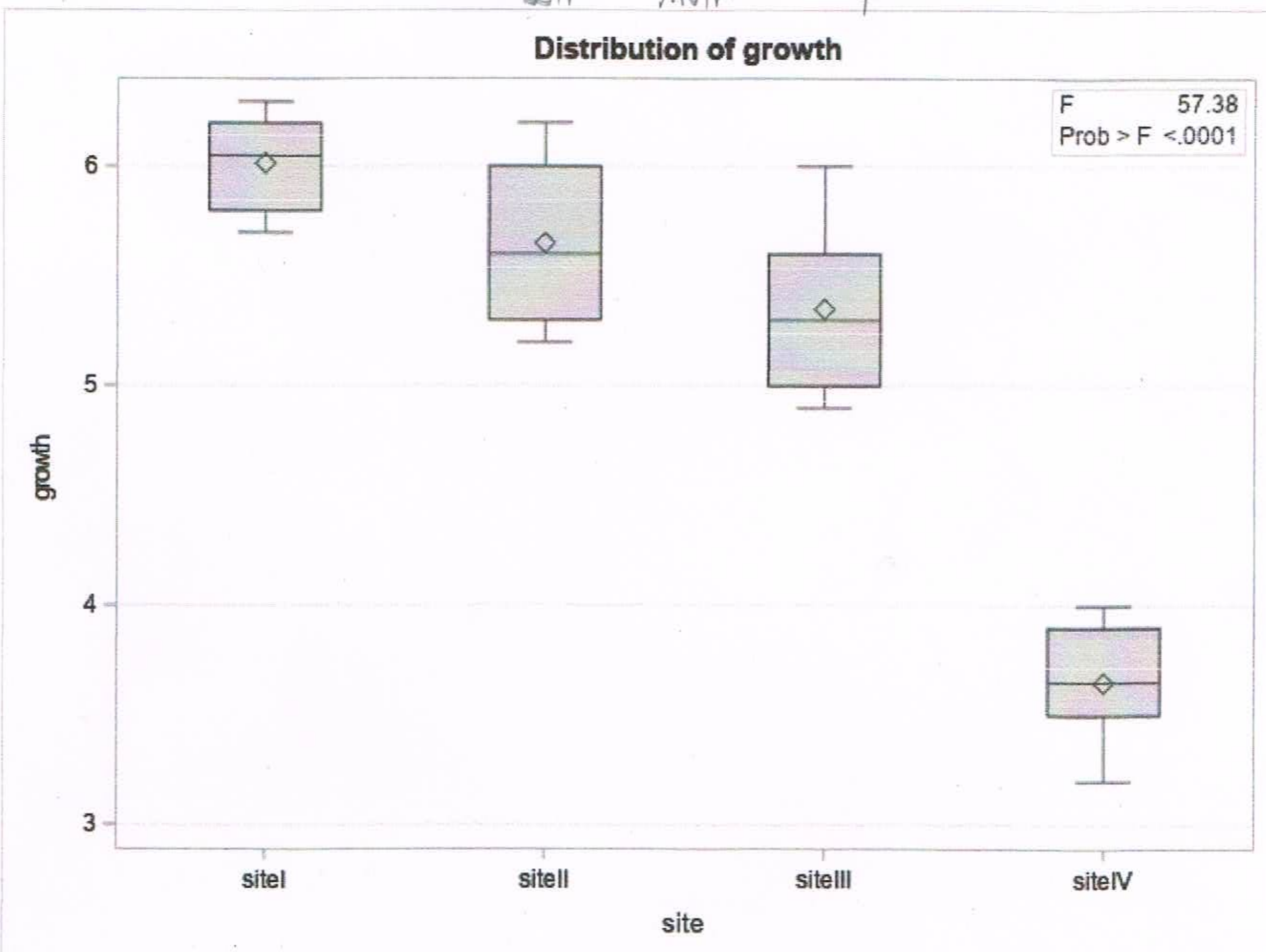
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model dfT	3 SSTr	19.74000000	6.58000000	57.38	<.0001
Error dfE	20 SSE	2.29333333	0.11466667 MSE		
Corrected Total	23	22.03333333			

Total df TSS

R-Square	Coeff Var	Root MSE	growth Mean
0.895915	6.554026	0.338625	5.166667

Source	DF	Anova SS	Mean Square	F Value	Pr > F
site	3	19.74000000	6.58000000	57.38	<.0001

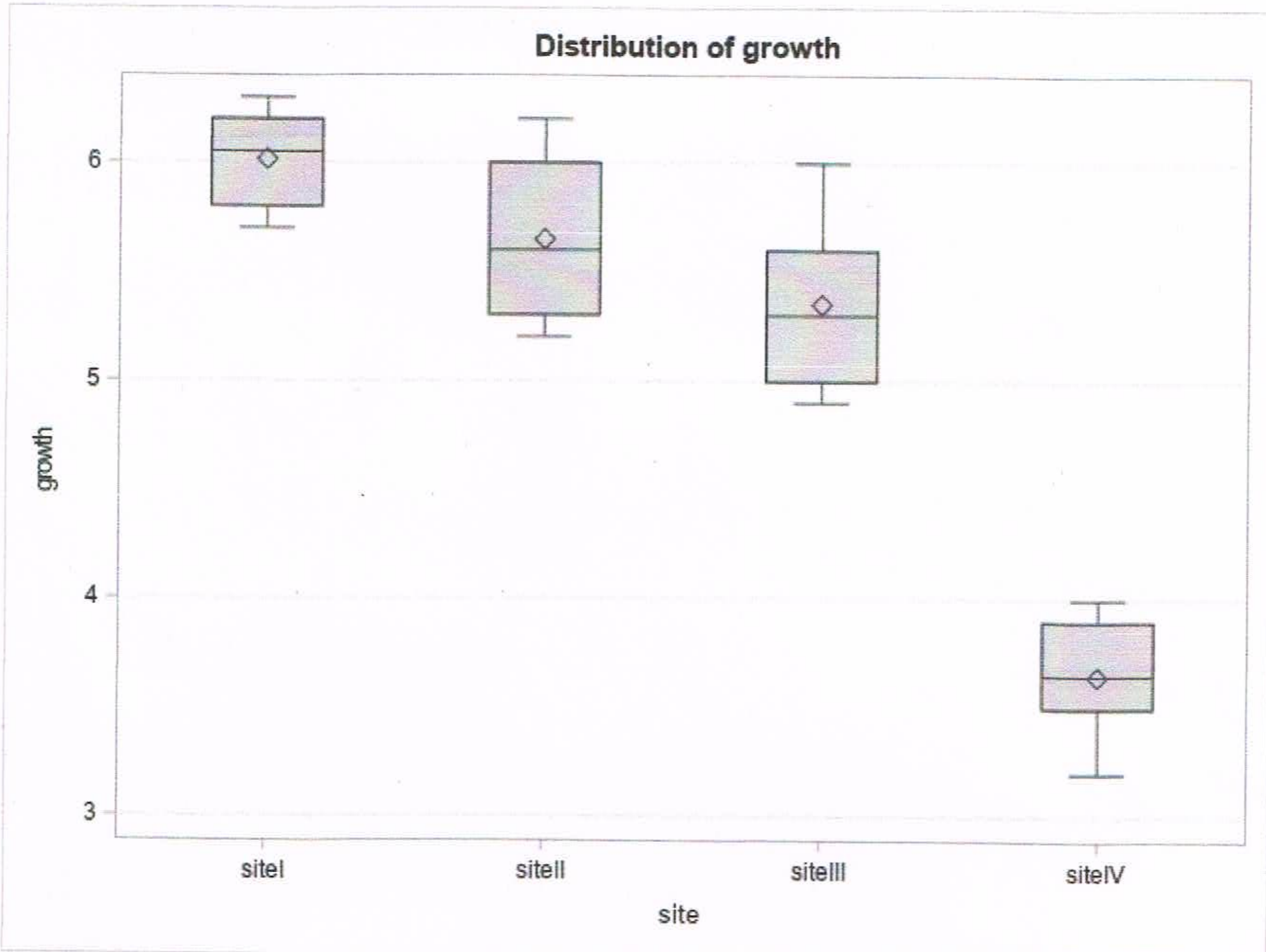
SSTr MSTr F



Connor 101041125 ods graphics off

The SAS System

The ANOVA Procedure



Connor 101041125 ods graphics off

The SAS System

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for growth

Note: This test controls the Type I experimentwise error rate.

Alpha	0.1	→ $\alpha = 0.1$
Error Degrees of Freedom	20	→ $n-K$
Error Mean Square	0.114667	→ MSE
Critical Value of Studentized Range	3.46154	→ $q_{\alpha}(K, n-K)$
Minimum Significant Difference	0.4785	→ h.s.d.

Comparisons significant at the 0.1 level are indicated by *.**					
site Comparison	**Difference Between Means**	**Simultaneous 90% Confidence Limits**			
siteI - siteII	$\lvert\bar{y}_1-\bar{y}_2\rvert$ 0.3667	-0.1119	0.8452		$\mu_1=\mu_2$
siteI - siteIII	$\lvert\bar{y}_1-\bar{y}_3\rvert$ 0.6667	0.1881	1.1452	***	$\mu_1\neq\mu_3$
siteI - siteIV	$\lvert\bar{y}_1-\bar{y}_4\rvert$ 2.3667	1.8881	2.8452	***	$\mu_1\neq\mu_4$
siteII - siteI	$\lvert\bar{y}_2-\bar{y}_1\rvert$ -0.3667	-0.8452	0.1119		$\mu_2=\mu_1$
siteII - siteIII	$\lvert\bar{y}_2-\bar{y}_3\rvert$ 0.3000	-0.1785	0.7785		$\mu_2=\mu_3$
siteII - siteIV	$\lvert\bar{y}_2-\bar{y}_4\rvert$ 2.0000	1.5215	2.4785	***	$\mu_2\neq\mu_4$
siteIII - siteI	$\lvert\bar{y}_3-\bar{y}_1\rvert$ -0.6667	-1.1452	-0.1881	***	$\mu_3\neq\mu_1$
siteIII - siteII	$\lvert\bar{y}_3-\bar{y}_2\rvert$ -0.3000	-0.7785	0.1785		$\mu_3=\mu_2$
siteIII - siteIV	$\lvert\bar{y}_3-\bar{y}_4\rvert$ 1.7000	1.2215	2.1785	***	$\mu_3\neq\mu_4$
siteIV - siteI	$\lvert\bar{y}_4-\bar{y}_1\rvert$ -2.3667	-2.8452	-1.8881	***	$\mu_4\neq\mu_1$
siteIV - siteII	$\lvert\bar{y}_4-\bar{y}_2\rvert$ -2.0000	-2.4785	-1.5215	***	$\mu_4\neq\mu_2$
siteIV - siteIII	$\lvert\bar{y}_4-\bar{y}_3\rvert$ -1.7000	-2.1785	-1.2215	***	$\mu_4\neq\mu_3$

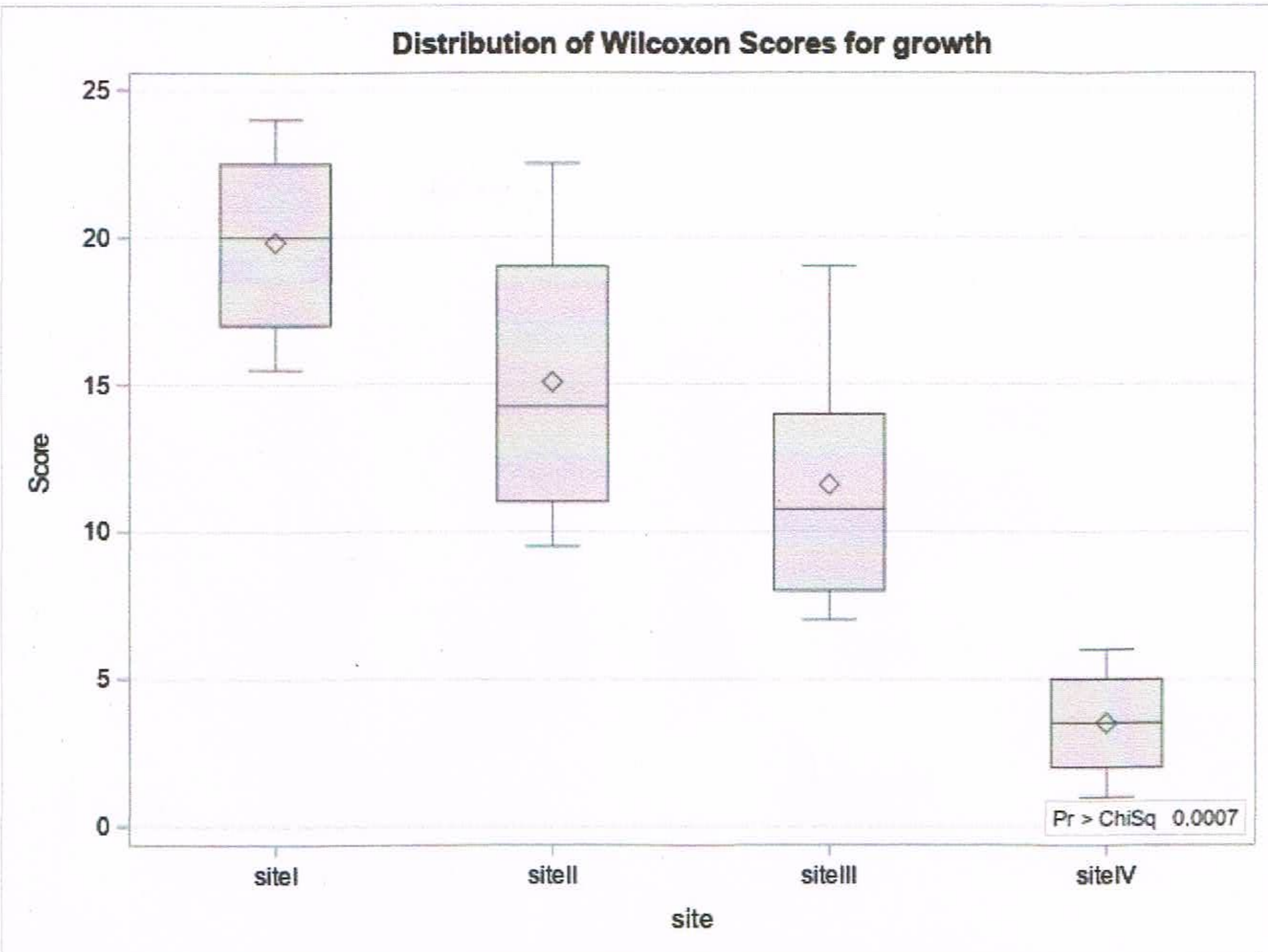
Connor 101041125 ods graphics off

The SAS System

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable growth Classified by Variable site					
site	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
siteI	6	119.00 T_{R1}	75.0	14.977157	19.833333
siteII	6	90.50 T_{R2}	75.0	14.977157	15.083333
siteIII	6	69.50 T_{R3}	75.0	14.977157	11.583333
siteIV	6	21.00 T_{R4}	75.0	14.977157	3.500000
Average scores were used for ties.					

Kruskal-Wallis Test	
Chi-Square	17.1271 → H
DF	3 → K−1
Pr > Chi-Square	0.0007



Connor 101041125 ods graphics off

③ Run the tests on the provided plot-treatment data:

RBD, three Steps:

I) Harteley's Test

II) Main Test

III) Turkeys Test

Parametric test

RBD Test:

Assumptions:

1) three independent insecticides assigned randomly to four plots
2) Populations Correspond to each Combination of insecticide - Plot are normally distributed
3) With equal varience, σ^2
4) No interactions between treatment-group Combination on the insecticide-Plot Combination

I) To check the assumption of equal varience use Hartley's test, we need S_i^2's for $i=1,2,3$, where $n_1=n_2=n_3=4$, $K=3$, $b=4$, $\bar{n}=b=4$, $[\bar{n}]=4$, $n=bK=12$:

$$S_1^2=\frac{\sum_{j=1}^{b} y_{1j}^2-\frac{\left(\sum_{j=1}^{b} y_{1j}\right)^2}{b}}{b-1}=\frac{13362-(230)^2/4}{4-1}\doteq 45.\dot{6}\dot{6}$$

$$S_2^2=\frac{\sum_{j=1}^{b} y_{2j}^2-\frac{\left(\sum_{j=1}^{b} y_{2j}\right)^2}{b}}{b-1}=\frac{30625-(349)^2/4}{4-1}\doteq 58.25\,\Delta^{max.}$$

$$S_3^2=\frac{\sum_{j=1}^{b} y_{3j}^2-\frac{\left(\sum_{j=1}^{b} y_{3j}\right)^2}{b}}{b-1}=\frac{25698-(320)^2/4}{4-1}\doteq 32.\dot{6}\dot{6}\,\Delta^{min.}$$

$$\sum_{j=1}^{b} y_{1j}\overset{=T_1^2}{=}230,\ \sum_{j=1}^{b} y_{2j}\overset{=T_2^2}{=}349,\ \sum_{j=1}^{b} y_{3j}\overset{=T_3^2}{=}320$$

$$\sum_{j=1}^{b} y_{1j}^2=13362,\ \sum_{j=1}^{b} y_{2j}^2=30625,\ \sum_{j=1}^{b} y_{3j}^2=25698$$

Claim: The variences are not all the Same

$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2$

H_a: At least one of the $\sigma^2 \neq$ the others

$\alpha = 0.01$

Test-Statistic:

$$F_{max} = \frac{S^2_{max}}{S^2_{min}} = \frac{58.25}{32.66} = 1\tfrac{307}{392} \doteq 1.783163265 \doteq \boxed{1.7832}$$

Rejection-Region:

We reject H_0 if $F_{max} > F_{max(K,[n]-1);\alpha}$ $\left\langle \begin{array}{l} F_{max(3,3);0.01} = 85 \\ F_{max(3,3);0.05} = 27.8 \end{array}\right.$

thus, $F_{max(3,4-1);\alpha} = F_{max(3,3);\alpha}$

$\alpha = 0.1$ So find both 0.01 & 0.05 for the f_{max} test:

$$F_{max} < F_{max(3,3);0.05} < F_{max(3,3);0.01}$$

↳ Thus, We Cannot Reject $H_0 \Rightarrow$ the Varience of each group is equal

∴ We Conclude at 99% Confidence our assumption of equal Varience is not Violated; Assumption is Valid.

II) Move on to Main Test:

$$TSS = \sum_{i=1}^{3}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{\left(\sum_{i=1}^{3}\sum_{j=1}^{n_i} y_{ij}\right)^2}{bK}$$

$$= 13362 + 30625 + 25698 - \frac{[230+349+320]^2}{12}$$

$$= 69685 - \frac{899^2}{12} = 2334\tfrac{11}{12} \doteq \boxed{2334.91\dot{6}\dot{6}}$$

$$SST_r = \sum_{j=1}^{3} \frac{T_i^2}{n_i} - \frac{\left(\sum_{i=1}^{3}\sum_{j=1}^{n_i} y_{ij}\right)^2}{bK}$$

$$= \frac{230^2}{4} + \frac{349^2}{4} + \frac{320^2}{4} - \frac{(230+349+320)^2}{12}$$

$$= 69275.25 - \frac{808201}{12} = \boxed{1925.1\dot{6}\dot{6}}$$

$$SSB = \sum_{j=1}^{n_i} \frac{B_i^2}{K} - \frac{\left(\sum_{i=1}^{3}\sum_{j=1}^{n_i} y_{ij}\right)^2}{bK}$$

$$= \frac{220^2}{3} + \frac{199^2}{3} + \frac{242^2}{3} + \frac{238^2}{3} - \frac{(230+349+320)^2}{12}$$

$$= 67736\tfrac{1}{3} - 67350\tfrac{1}{12} = \boxed{386.25}$$

$$SSE = TSS - SST_r - SSB = 2334.91\dot{6}6 - 1925.1\dot{6}6 - 386.25$$

$$\boxed{SSE = 23.5} \quad \Big| \quad SST = TSS - SSE = 2334.91\dot{6}6 - 23.5 = 2311.41\dot{6}6$$

$$MST_r = \frac{SST_r}{K-1} = \frac{1925.1\dot{6}6}{3-1} = 962.58\dot{3}3$$

$$MSB = \frac{SSB}{b-1} = \frac{386.25}{4-1} = 128.75$$

$$MSE = \frac{SSE}{df_E} = \frac{23.5}{(K-1)(b-1)} = \frac{23.5}{2 \cdot 3} \doteq 3.91\dot{6}6$$

$$MST = \frac{SST}{df_T} = \frac{2311.41\dot{6}6}{5} \doteq 462.28\dot{3}3$$

Total $df = Kb - 1 = 4 \cdot 3 - 1 = 12 - 1 = 11$, $df_E = (K-1)(b-1) = 2 \cdot 3 = 6$

$df_T = \text{Total } df - df_E = 11 - 6 = 5$

$$F = \frac{MST}{MSE} = \frac{462.28\dot{3}3}{3.91\dot{6}6} \doteq 118.03$$

ANOVA Table:

Source	df	Sum of Squares	Mean Square	F-Value	Pr > F
Model	5	$2311.41\dot{6}6$	$462.28\dot{3}3$	118.03	<0.0001
Error	6	23.5	$3.91\dot{6}6$		
Corrected Total	11	$2334.91\dot{6}6$			

Source	df	ANOVA SS	MS	F-Value	Pr > F
Plot	3	386.25	128.75	32.87	0.0004
insect	2	$1925.1\dot{6}6$	$962.58\dot{3}3$	245.77	<0.0001

$$F_T = \frac{MST_r}{MSE} = \frac{962.58\dot{3}3}{3.91\dot{6}6} \doteq 245.77$$

$$F_B = \frac{MSB}{MSE} = \frac{128.75}{3.91\dot{6}6} \doteq 32.87$$

$$F_{\alpha(K-t,\, bK)} = F_{0.10(3,12)} = 2.61$$

$\hookrightarrow F > F_{0.10(3,12)}$ as $118.03 > 2.61$ $\therefore$ Reject H_0

Note: Our Claim for the main test

Claim: there's a difference in the mean effectiveness of the insecticides on the Plots

$H_0: \mu_1 = \mu_2 = \mu_3 \Rightarrow$ the means are the Same

H_a: At least one μ's doesn't equal another μ

thus, as per above, we Can Conclude as per a 1% level of Significence there's a difference in the means of the insecticide effectiveness on the Plots.

III) Turkeys Test:

1) Calculate $\binom{k}{2} = \binom{3}{2} = 3$ Pairs of $|\bar{y}_i - \bar{y}_j|$ for $H_0: \mu_i = \mu_j$ vs. $H_a: \mu_i \neq \mu_j$ for $i,j = 1,2,3$; $i \neq j$

H_0: Sites are Same

H_a: Sites are different

Claim: the treatments all are the Same

For $i=1, j=2: |230-349|/4 = 29.75$

$i=1, j=3: |230-320|/4 = 22.5$

$i=2, j=3: |349-320|/4 = 7.25$

2) $h.s.d. = q_\alpha(k, (b-1)(k-1)) \sqrt{\frac{MSE}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$

$= q(3,6) (47/48)^{1/2}$

$= 3.558 (47/48)^{1/2} \doteq 3.5207$

thus:

$|\bar{y}_1 - \bar{y}_2| = 29.75 > 3.5207 \Rightarrow \mu_1 \neq \mu_2$

$|\bar{y}_1 - \bar{y}_3| = 22.5 > 3.5207 \Rightarrow \mu_1 \neq \mu_3$

$|\bar{y}_2 - \bar{y}_3| = 7.25 > 3.5207 \Rightarrow \mu_2 \neq \mu_3$

So we Know that all the Sites are the Same within limits to a 99% Confidence

So, we Cannot reject H_0

↳ Accept H_0: the Sites are all the Same means for treatment

Non-Parametric Test:

Assume:

1) R.B.D.
2) In each insecticide-Plot Combination we have population with approximatly the Same Shape & Spread
3) No interactions between Plots & insecticides

Firstly, rank observations from Smallest to largest:

Insecticide	Plot 1	2	3	4	
1	56 ①	41 ①	65 ①	60 ①	$T_{R1} = 4$
2	84 ③	78 ③	94 ③	93 ③	$T_{R2} = 12$
3	80 ②	72 ②	83 ②	85 ②	$T_{R3} = 8$

Hypothesis Test:

I) Claim: At least one median differs from the rest

$H_0: Md_1 = Md_1 = Md_2 = Md_3 = Md_4$

H_a: At least one of the Md's $\neq$ the others

II) Test-Statistics:

$$H = \frac{12}{bK(K+1)}\left[\sum_{i=1}^{4} T_{R_i}^2\right] - 3b(K+1)$$

$$= \frac{\cancel{12}}{\cancel{12}(4)}\left[4^2+12^2+8^2\right] - 3(4)(3+1)$$

$$= \frac{1}{4}\left[224\right] - 48 = \boxed{8}$$

III) Rejection-Region:

We reject Ho if $H > \chi^2_{\alpha;(K-1)} = \chi^2_{0.10;(2)} = 11.34$

$H = 8 < \chi^2_{\alpha;(2)} = 11.34 \Rightarrow 8 < 11.34$ So $H < \chi^2_{\alpha;(2)}$

$\Rightarrow$ We Cannot Reject Ho Since $H < \chi^2_{\alpha;(2)}$

IV) Conclusion:

We don't reject Ho & Conclude at 1% level of Significence there's an evidence to Say that the medians of the insecticides does not differ.

Q3 SAS Code

The SAS System

The ANOVA Procedure

Class Level Information		
Class	**Levels**	**Values**
plot	4	1 2 3 4
insect	3	1 2 3

Columns

Rows

Number of Observations Read	12
Number of Observations Used	12

#data Samples

Connor 101041125

The SAS System

The ANOVA Procedure

Dependent Variable: seedlings

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model df_T	5	SST 2311.416667	MST 462.283333	118.03 ↳ F	<.0001
Error df_E	6	SSE 23.500000	MSE 3.916667		
Corrected Total	11 ↳ Total df	2334.916667 TSS			

R-Square	Coeff Var	Root MSE	seedlings Mean
0.989935	2.641678	1.979057	74.91667

Source	DF	SSB Anova SS	MSB Mean Square	F Value	F_B Pr > F
plot	3	386.250000	128.750000	32.87	0.0004
insect	2	1925.166667 ↳ SST_r	962.583333 ↳ MST_r	245.77 ↳ F_T	<.0001

Connor 101041125

The SAS System

The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for seedlings

Note: This test controls the Type I experimentwise error rate.

Alpha	0.01
Error Degrees of Freedom	6
Error Mean Square	3.916667
Critical Value of Studentized Range	6.33032
Minimum Significant Difference	6.264

$\to \alpha = 0.01$

$\to K-1$

Comparisons significant at the 0.01 level are indicated by *.**

	insect Comparison	Difference Between Means	Simultaneous 99% Confidence Limits		
$i=2, j=3$	2 - 3	7.250	0.986	13.514	***
$i=2, j=1$	2 - 1	29.750	23.486	36.014	***
$i=3, j=2$	3 - 2	-7.250	-13.514	-0.986	***
$i=3, j=1$	3 - 1	22.500	16.236	28.764	***
$i=1, j=2$	1 - 2	-29.750	-36.014	-23.486	***
$i=1, j=3$	1 - 3	-22.500	-28.764	-16.236	***

Same as Worksheet Q3, Turkeys Test Results

$\mu_i \neq \mu_j$ for all listed Comparisons

Connor 101041125

The SAS System

The FREQ Procedure

Summary Statistics for insect by seedlings
Controlling for plot

Cochran-Mantel-Haenszel Statistics (Based on Rank Scores)				
Statistic	**Alternative Hypothesis**	**DF**	**Value**	K-1 **Prob**
1	Nonzero Correlation	1	2.0000	0.1573
2	Row Mean Scores Differ	2	8.0000	0.0183

DH

Total Sample Size = 12

Connor 101041125

Q2 SAS Code

Name: Connor Raymond Stewart
ID: 101041125

```
footnote 'Connor 101041125'
ods graphics off;
data ecology;
input site$ growth @@;
cards;
siteI 5.7 siteI 6.3 siteI 6.1 siteI 6.0 siteI 5.8 siteI 6.2
siteII 6.2 siteII 5.3 siteII 5.7 siteII 6.0 siteII 5.2 siteII 5.5
siteIII 5.4 siteIII 5.0 siteIII 6.0 siteIII 5.6 siteIII 4.9 siteIII 5.2
siteIV 3.7 siteIV 3.2 siteIV 3.9 siteIV 4.0 siteIV 3.5 siteIV 3.6

run;
proc anova;
class site;
model growth=site;
means site/tukey cldiff alpha=0.10;
run;
proc NPAR1WAY WILCOXON;
class site;
run;
```

Q3 SAS Code

Name: Connor Raymond Stewart

ID: 101041125

```
footntoe 'Connor 101041125';
ods graphics off;
data beans;
input insect plot seedlings @@;
cards;
1 1 56 1 2 49 1 3 65 1 4 60
2 1 84 2 2 78 2 3 94 2 4 93
3 1 80 3 2 72 3 3 83 3 4 85


run;
proc anova;
class plot insect;
model seedlings=plot insect;
means insect/tukey cldiff alpha=0.01;
run;
proc freq;
tables plot*insect*seedlings/CMH2 scores=rank noprint;
run;
```