STAT 2509 A
Assignment #4

SOLUTION // 65

1. **[19 marks]**

[3] (a)

If men's wear: $y = \beta_0 + \beta_1 x_1 + \beta_2(0) + \beta_3(0) + \beta_4 x_1(0) + \beta_5 x_1(0) + \varepsilon$

or $y = \beta_0 + \beta_1 x_1 + \varepsilon$ …………………… **(M)**

If children's wear: $y = \beta_0 + \beta_1 x_1 + \beta_2(1) + \beta_3(0) + \beta_4 x_1(1) + \beta_5 x_1(0) + \varepsilon$

or $y = (\beta_0 + \beta_2) + (\beta_1 + \beta_4)x_1 + \varepsilon$ …………………… **(C)** **[3]**

If women's wear: $y = \beta_0 + \beta_1 x_1 + \beta_2(0) + \beta_3(1) + \beta_4 x_1(0) + \beta_5 x_1(1) + \varepsilon$

or $y = (\beta_0 + \beta_3) + (\beta_1 + \beta_5)x_1 + \varepsilon$ …………………… **(W)**

[4.5] (b)

[1]

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	74.830	5	14.966	110.496	<.001[b]
	Residual	1.219	9	.135		
	Total	76.049	14			

a. Dependent Variable: sales
b. Predictors: (Constant), x1x3, adds_expenditure, x2, x3, x1x2

$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ $\alpha = 0.05$
H_a : *at least one of the* $\beta' s \neq 0$ **[1]**

test-statistics: F = 110.496

R.R. we reject H_0 if *p-value* < α **[1]** (or if $F > F_{\alpha;(k,n-(k+1))} = F_{0.05;(5,9)} = 3.48$)

Since *p-value* < 0.001 < 0.05 **[1/2]** (or F = 110.496 > 3.48), we reject H_0 **[1/2]** and conclude that at 5% level of significance we have enough evidence to conclude that the full model is useful, i.e. it can be used. **[1/2]**

[7] (c)

$H_0: \beta_4 = \beta_5 = 0$,
H_a : *at least one of* $\beta's \neq 0$ **[1]** $\alpha = 0.05$

Full model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1 x_3 + \varepsilon$

Reduced Model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

ANOVA for the full model (above in part b)) produced following: $SSR_f = 74.830$, df = 5
$SSE_f = 1.219$, df = 9

ANOVA for the reduced model is below:

ANOVA[a] **[1]**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	73.784	3	24.595	119.445	<.001[b]
	Residual	2.265	11	.206		
	Total	76.049	14			

a. Dependent Variable: sales
b. Predictors: (Constant), x3, adds_expenditure, x2

[1/2] if correct SPSS values were used in the calculation of test statistics

test-statistics: **[1/2]** $F_{drop} = \dfrac{[SSE_r - SSE_f]/[df_{SSE_r} - df_{SSE_f}]}{MSE_f} = \dfrac{[SSE_r - SSE_f]/[n-4-(n-6)]}{SSE_f / n-6}$

[1/2]

$$= \frac{(2.265-1.219)/(11-9)}{1.219/9} = \frac{1.046/2}{1.219/9} = \frac{0.523}{0.13544} = \underline{\underline{3.8615}}$$ **[1/2]**

[1/2]

(1/2 mark for each correct d.f.)

or equivalently $F_{part} = \dfrac{[SSR_f - SSR_r]/[df_{SSR_f} - df_{SSR_r}]}{MSE_f} = \dfrac{(74.830-73.784]/(5-3)}{1.219/9}$

$$= \frac{1.046/2}{1.219/9} = \frac{0.523}{0.13544} = \underline{\underline{3.8615}}$$

R.R. we reject H_0 **if** F_{drop} (or F_{part}) $> F_{\alpha;(2,n-6)} = F_{0.05;(2,9)} = 4.26$ **[1]**

- **Since** F_{drop} (or F_{part}) $= 3.8615 \not> 4.26$ **[1/2]** **, we** **<u>do not reject</u>** H_0 **[1/2]** **and conclude that at 5% level of significance there is evidence that the interaction terms are not needed.** **[1/2]**

[4.5] (d)

Coefficients[a] **[1] if they used reduced model for the coefficients**

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
	(Constant)	4.610	.321		14.367	<.001
	adds_expenditure	.870	.083	.546	10.501	<.001
	x2	2.240	.287	.469	7.805	<.001
	x3	4.520	.287	.946	15.750	<.001

a. Dependent Variable: sales

$$\left.\begin{array}{l} H_0 : \beta_1 = 0 \\ H_a : \beta_1 \neq 0 \end{array}\right\} \text{[1]} \qquad \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$$

test-statistics: t = 10.501

R.R. we reject H_0 if p-value < α **[1]** (or if $|t| > t_{\alpha/2;n-(k+1)} = t_{0.025;11} = 2.201$)

Since *p*-value < 0.001 < 0.05 **[1/2]** (or t = 10.501 > 2.201), we reject H_0 **[1/2]** and conclude that at 5% level of significance we have enough evidence to conclude that the 'advertising expenditure' is useful in predicting 'total weekly sales'. **[1/2]**

NOTE: students can also use F_{part} (or F_{drop}) with the full model from part c) and the reduced model without x1.

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
	Regression	51.077	2	25.539	12.272	.001[b]
	Residual	24.972	12	2.081		
	Total	76.049	14			

a. Dependent Variable: sales

b. Predictors: (Constant), x3, x2

[1/2] if correct SPSS values were used in the calculation of test statistics

[1/2] if correct d.f. were used in the calculation of test statistics

[1/2] $$F_{part} = \frac{[SSR_f - SSR_r]/[df_{SSR_f} - df_{SSR_r}]}{MSE_f} = \frac{(73.784 - 51.077)/(3-2)}{2.265/11}$$

$$= \frac{22.707/1}{2.265/11} = \frac{22.707}{0.205909} = \underline{\underline{110.2768}} \quad \textbf{[1/2]}$$

R.R. **we reject** H_0 **if** F_{drop} (or F_{part}) $> F_{\alpha;(1,11)} = F_{0.05;(1,11)} = 4.84$ **[1]**

Since F_{part} = 110.2768 > 4.84 **[1/2]**, we reject H_0 **[1/2]** and conclude that at 5% level of significance we have enough evidence to conclude that the 'advertising expenditure' is useful in predicting 'total weekly sales'. **[1/2]**

2. [6 marks]

[2] (a)

Variables Entered/Removed[a] **[1]**.

Model	Variables Entered	Variables Removed	Method
Step 1 → 1	x3	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
Step 2 → 2	adds_expenditure	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
Step 3 → 3	x2	.	Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: sales

Hence, the best model is with variables **x1, x2 and x3** **[1]**.

[1.5] (b)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	73.784	3	24.595	119.445	<.001[b]
	Residual	2.265	11	.206		
	Total	76.049	14			

a. Dependent Variable: sales

b. Predictors: (Constant), adds_expenditure, x3, x2

p-value < α
∴ full model can be used

[1/2]

no variables were removed
[1/2]

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	adds_expenditure, x3, x2[b]	.	Enter

a. Dependent Variable: sales
b. All requested variables entered.

Hence, the best model is with variables **x1, x2 and x3** [1/2]

[2.5] (c)

Variables Entered/Removed[a] [1/2].

Model	Variables Entered	Variables Removed	Method
1	x3	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	adds_expenditure	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	x2	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: sales

Excluded Variables[a] [1/2].

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
1	x2	.469[b]	2.455	.030	.578	.750
	adds_expenditure	.546[b]	4.290	.001	.778	1.000
2	x2	.469[c]	7.805	<.001	.920	.750

a. Dependent Variable: sales
b. Predictors in the Model: (Constant), x3
c. Predictors in the Model: (Constant), x3, adds_expenditure

all 3 p-values are <0.05, hence we keep all 3 x's

Hence, the best model is with variables **x1, x2 and x3** [1/2]

Hence, the best model is with variables **x1, x2 and x3** [1/2], which confirms the results in Q1 [1/2].

3. **[40 marks]**

C.R.D.

Assume: 1) 4 independent random samples of swamp plants (given) **[1/2]**
2) 4 normally distributed swamp plants populations **[1/2]**
3) with equal variance, σ^2 (?) **[1/2]**

- to check the assumption of equal variance using **Hartley's test**, we need s_i^2's for i = 1, 2, 3, 4, where $n_1 = n_2 = n_3 = n_4 = 6$

$$k = 4, \bar{n} = 6, [\bar{n}] = 6, n = 24$$

i.e. $$s_1^2 = \frac{\sum_{j=1}^{n_1} y_{1j}^2 - \frac{\left(\sum_{j=1}^{n_1} y_{1j}\right)^2}{n_1}}{n_1 - 1} = \frac{217.47 - \frac{(36.1)^2}{6}}{5} = \underline{\underline{0.053667}}$$ **[1/2]** ← min

$$s_2^2 = \frac{\sum_{j=1}^{n_2} y_{2j}^2 - \frac{\left(\sum_{j=1}^{n_2} y_{2j}\right)^2}{n_2}}{n_2 - 1} = \frac{192.31 - \frac{(33.9)^2}{6}}{5} = \underline{\underline{0.155}}$$ **[1/2]**

$$s_3^2 = \frac{\sum_{j=1}^{n_3} y_{3j}^2 - \frac{\left(\sum_{j=1}^{n_3} y_{3j}\right)^2}{n_3}}{n_3 - 1} = \frac{172.57 - \frac{(32.1)^2}{6}}{5} = \underline{\underline{0.167}}$$ **[1/2]** ← max

$$s_4^2 = \frac{\sum_{j=1}^{n_4} y_{4j}^2 - \frac{\left(\sum_{j=1}^{n_4} y_{4j}\right)^2}{n_4}}{n_4 - 1} = \frac{80.35 - \frac{(21.9)^2}{6}}{5} = \underline{\underline{0.083}}$$ **[1/2]**

$$\left.\begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 \\ H_a : \textit{at least one of the } \sigma^2\text{'s} \neq \ ; \end{array}\right\} \alpha = 0.05$$ **[1]**

test-stattistic: **[1/2]** $F_{\max} = \frac{s_{\max}^2}{s_{\min}^2} = \frac{0.167}{0.053667} = \underline{\underline{\mathbf{3.111801}}}$ **[1/2]**

R.R.: we reject H_0 if $F_{\max} > F_{\max(k,[\bar{n}]-1);\alpha} = F_{\max(4,5);0.05} = 13.7$ **[1]**

Since F_{max} = 3.11 $\ngtr$ 13.7 [1/2], we do not reject H_0 [1/2] and conclude that at 5% level of significance there is no evidence to say that the variances are not equal (i.e. we have equal variance). [1/2]

$\therefore$ **we may proceed with the main test:**

[1/2] $TSS = \sum_{i=1}^{4}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{\left(\sum_{i=1}^{4}\sum_{j=1}^{n_i} y_{ij}\right)^2}{n} = 662.7 - \frac{(124)^2}{24}$ = 662.7 – 640.6667 = 22.03333 [1/2]

[1/2] $SST_r = \sum_{i=1}^{4}\frac{T_i^2}{n_i} - \frac{\left(\sum_{i=1}^{4}\sum_{j=1}^{n_i} y_{ij}\right)^2}{n} = \left[\frac{(36.1)^2}{6} + \frac{(33.9)^2}{6} + \frac{(32.1)^2}{6} + \frac{(21.9)^2}{6}\right] - \frac{(124)^2}{24} =$

= 660.4067 – 640.6667 = 19.74 [1/2]

[1/2] $SSE = TSS - SST_r =$ 2.293333 [1/2]

[1/2] $MST_r = \frac{SST_r}{k-1} = \frac{19.74}{3}$ = 6.58 [1/2]

[1/2] $MSE = \frac{SSE}{n-k} = \frac{2.293333}{20} =$ 0.114667 [1/2]

[1/2] $F_T = \frac{MST_r}{MSE} =$ 57.38372 [1/2]

Source	*d.f.*	*SS*	MS	*F*
Treatments	3	19.74	6.58	57.38372
Error	20	2.293333	0.114667	
Total	23	22.03333		
	[1/2]	[1/2]	[1/2]	[1/2]

(1/2 mark for each column, if values are entered correctly)

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$; $\alpha = 0.05$

H_a : *at least one of the* μ's $\neq$ [1]

test-statistics: $F_T = \frac{MST_r}{MSE} =$ **57.38372**

R.R: we reject H_0 if $F_T > F_{\alpha(k-1,n-k)} = F_{0.05(3,20)} =$ 3.10 [1]

Since F_T = 57.38372 > 3.10 [1/2], we reject H_0 [1/2] and conclude that at 5% level of significance there is an evidence to say that the mean leaf length of swamp plants differ between the 4 swamp locations. [1/2]

Which treatments (i.e. swamp locations) differ? Tukey's h.s.d.

1) Calculate $\binom{k}{2} = \binom{4}{2} = 6$ 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, 4
$i \neq j$

2) [1/2] h.s.d. = $q_\alpha(k, n-k)\sqrt{\frac{MSE}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} = q_{0.05}(4,20)\sqrt{\frac{0.114667}{2}\left(\frac{2}{6}\right)} =$

$= 3.96\sqrt{0.019111} =$ **0.547444** [1/2]
[1/2]

$$\bar{y}_1 = \frac{T_1}{n_1} = \frac{36.1}{6} = 6.01666, \quad \bar{y}_2 = \frac{T_2}{n_2} = \frac{33.9}{6} = 5.65$$

$$\bar{y}_3 = \frac{T_3}{n_3} = \frac{32.1}{6} = 5.35, \quad \bar{y}_4 = \frac{T_4}{n_4} = \frac{21.9}{6} = 3.65$$

3)

$|\bar{y}_1 - \bar{y}_2| = 0.36666 \not> 0.547444 \Rightarrow \mu_1 = \mu_2$

$|\bar{y}_1 - \bar{y}_3| = 0.66666 > 0.547444 \Rightarrow \underline{\underline{\mu_1 \neq \mu_3}}$ [1/2]

$|\bar{y}_1 - \bar{y}_4| = 2.36666 > 0.547444 \Rightarrow \underline{\underline{\mu_1 \neq \mu_4}}$ [1/2]

$|\bar{y}_2 - \bar{y}_3| = 0.3 \not> 0.547444 \Rightarrow \mu_2 = \mu_3$

$|\bar{y}_2 - \bar{y}_4| = 2 > 0.547444 \Rightarrow \underline{\underline{\mu_2 \neq \mu_4}}$ [1/2]

$|\bar{y}_3 - \bar{y}_4| = 1.7 > 0.547444 \Rightarrow \underline{\underline{\mu_3 \neq \mu_4}}$ [1/2]

i.e. [1/2] there are differences between swamp locations (I & III), (I & IV), (II & IV) and (III & IV).

Non-parametric Analysis (Kruskal-Wallis test)

Assume: 1) C.R.D. [1/2] (4 independent random samples from 4 treat't populations) with
2) approximately the same shape[1/2] and spread[1/2]

First we need to rank the observations from smallest to the largest:

Site I		**Site II**		**Site III**		**Site IV**	
5.7	**(15.5)**	6.2	**(22.5)**	5.4	**(12)**	3.7	**(4)**
6.3	**(24)**	5.3	**(11)**	5.0	**(8)**	3.2	**(1)**
6.1	**(21)**	5.7	**(15.5)**	6.0	**(19)**	3.9	**(5)**
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)**
	$T_{R_1}=119$ **[1/2]**		$T_{R_2}=90.5$ **[1/2]**		$T_{R_3}=69.5$ **[1/2]**		$T_{R_4}=21$ **[1/2]**

(***Check:*** $\dfrac{n(n+1)}{2}=\dfrac{24(25)}{2}=300$ **and** $\displaystyle\sum_{i=1}^{4}T_{R_i}=119+90.5+69.5+21=300$)

$$\left.\begin{array}{l} H_0: Md_1 = Md_2 = Md_3 = Md_4 \\ H_a: \text{at least one of the } Md\text{'s} \neq \end{array}\right\} \quad ; \qquad \alpha=0.05. \quad \textbf{[1]}$$

test-statistics:

[1/2] $$H=\frac{12}{n(n+1)}\left[\sum_{i=1}^{4}\frac{T_{R_i}^2}{n_i}\right]-3(n+1)=\frac{12}{24(25)}\left[\frac{(119)^2}{6}+\frac{(90.5)^2}{6}+\frac{(69.5)^2}{6}+\frac{(21)^2}{6}\right]-3(25)=$$

= 0.02(4 603.75) – 75 = 92.075 – 75 = 17.075 **[1/2]**

R.R: **we reject** H_0 **if** $H>\chi^2_{\alpha;(k-1)}=\chi^2_{0.05;(3)}$ = **7.815** **[1]**

Since *H* = 17.075 > 7.815 [1/2], we reject H_0 **[1/2] and conclude that at 5% level of significance there is an evidence to say that the medians of leaf length of swamp plants differ between the 4 swamp locations. [1/2]**

➢ **Which treatments differ? Dunn's procedure**

1) Calculate $\binom{k}{2}=\binom{4}{2}=6$ **pairs of** $|\bar{R}_i-\bar{R}_j|$ **for** $H_0: Md_i = Md_j$ **vs** $H_a: Md_i \neq Md_j$,

for *i, j* = 1, 2, 3, 4
i $\neq$ ***j***

2) Critical range =[1/2] $z_{\frac{\alpha}{k(k-1)}}\sqrt{\dfrac{n(n+1)}{12}\left(\dfrac{1}{n_i}+\dfrac{1}{n_j}\right)}=z_{\frac{0.05}{4(3)}}\sqrt{\dfrac{24(25)}{12}\left(\dfrac{2}{6}\right)}=$

$= z_{0.004167}\sqrt{16.66667}$ = **(2.635)*4.082483 = 10.75734** **[1/2]**
[1/2]

$$\bar{R}_1 = \frac{T_{R_1}}{n_1} = \frac{119}{6} = 19.8333\,, \qquad \bar{R}_2 = \frac{T_{R_2}}{n_2} = \frac{90.5}{6} = 15.0833$$

$$\bar{R}_3 = \frac{T_{R_3}}{n_3} = \frac{69.5}{6} = 11.5833\,, \qquad \bar{R}_4 = \frac{T_{R_4}}{n_4} = \frac{21}{6} = 3.5$$

3) $|\bar{R}_1 - \bar{R}_2| = 4.75 < 10.75734 \Rightarrow Md_1 = Md_2$

$|\bar{R}_1 - \bar{R}_3| = 8.25 < 10.75734 \Rightarrow Md_1 = Md_3$

$|\bar{R}_1 - \bar{R}_4| = 16.3333 > 10.75734 \Rightarrow \underline{\underline{Md_1 \neq Md_4}}$ **[1/2]**

$|\bar{R}_2 - \bar{R}_3| = 3.5 < 10.75734 \Rightarrow Md_2 = Md_3$

$|\bar{R}_2 - \bar{R}_4| = 11.5833 > 10.75734 \Rightarrow \underline{\underline{Md_2 \neq Md_4}}$ **[1/2]**

$|\bar{R}_3 - \bar{R}_4| = 8.0833 < 10.75734 \Rightarrow Md_3 = Md_4$

i.e. [1/2] there is a difference in medians of swamp locations (I & IV) and (II & IV).

<u>SPSS outputs</u>: (1 mark for each output table and **<u>½ mark if the highlighted/verified the SPSS values with those calculated by hand</u>**)

<u>Note:</u> Total 5.5 marks for SPSS part

LeafLength **[1]**

SwampSite	Mean	N	Std. Deviation	Median	Variance
Site I	6.017 $\bar{y}_1$	6	.2317	6.050	.054 s_1^2
Site II	5.650 $\bar{y}_2$	6	.3937	5.600	.155 s_2^2
Site III	5.350 $\bar{y}_3$	6	.4087	5.300	.167 s_3^2
Site IV	3.650 $\bar{y}_4$	6	.2881	3.650	.083 s_4^2
Total	5.167	24	.9788	5.450	.958

ANOVA [1]

LeafLength

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	19.740 SSTr	3	6.580 MSTr	57.384 F_T	<.001
Within Groups	2.293 SSE	20	.115 MSE		
Total	22.033 TSS	23			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: LeafLength

Tukey HSD [1]

(I) SwampSite	(J) SwampSite	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Site I	Site II	.3667	.1955	.270	-.181	.914
	Site III	.6667*	.1955	.014	.119	1.214
	Site IV	2.3667*	.1955	<.001	1.819	2.914
Site II	Site I	-.3667	.1955	.270	-.914	.181
	Site III	.3000	.1955	.437	-.247	.847
	Site IV	2.0000*	.1955	<.001	1.453	2.547
Site III	Site I	-.6667*	.1955	.014	-1.214	-.119
	Site II	-.3000	.1955	.437	-.847	.247
	Site IV	1.7000*	.1955	<.001	1.153	2.247
Site IV	Site I	-2.3667*	.1955	<.001	-2.914	-1.819
	Site II	-2.0000*	.1955	<.001	-2.547	-1.453
	Site III	-1.7000*	.1955	<.001	-2.247	-1.153

*. The mean difference is significant at the 0.05 level.

$\mu_1 \neq \mu_3$

$\mu_1 \neq \mu_4$

$\mu_2 \neq \mu_4$

$\mu_3 \neq \mu_4$

Homogeneous Subsets

LeafLength

Tukey HSD[a]

SwampSite	N	Subset for alpha = 0.05: 1	2	3
Site IV	6	3.650		
Site III	6		5.350	
Site II	6		5.650	5.650
Site I	6			6.017
Sig.		1.000	.437	.270

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 6.000.

optional

NPar Tests

Kruskal-Wallis Test

Ranks **[1]**

	SwampSite	N	Mean Rank
LeafLength	Site I	6	19.83 $\bar{R}_1$
	Site II	6	15.08 $\bar{R}_2$
	Site III	6	11.58 $\bar{R}_3$
	Site IV	6	3.50 $\bar{R}_4$
	Total	24	

Test Statistics[a,b] **[1]**

	LeafLength
Kruskal-Wallis H	17.127 $\leftarrow H$
df	3
Asymp. Sig.	<.001

a. Kruskal Wallis Test

b. Grouping Variable: SwampSite

p-value < 0.05

$\therefore$ reject H_0