TEST 3 NAME________KEY____________________________________
STAT 6604

Be clear. Be neat - no scratchouts accepted. Present your work in a folder. Make it easy to read and to follow.

1. Use the summary data (gas mileage per gallon) below to construct an One-Way ANOVA table and then answer the questions.

Truck	Average MPG	Variance of MPG	Sample size
Chevy	14.9	0.16	5
Dodge	14.4	0.07	5
Ford	14.5	0.20	5

$SSTR = 5[(14.9 - 14.6)^2 + (14.4 - 14.6)^2 + (14.5 - 14.6)^2] = .7$
$SSE = 4(.16 + .07 + .20) = 1.72$

Source of variation	**df**	**SS**	**MS**	**F***	**p-value**
Truck Make	**2**	**.7**	**.35**	**2.442**	**.129**
Error	**12**	**1.72**	**.143**		
Total	**14**	**2.42**			

(i) Does average gas mileage differ between the standard four-wheel drive pickup trucks made by Chevrolet, Dodge and Ford? Use $\alpha = .10$. State the hypotheses and provide a well-written conclusion.

Let μ_i denote the mean mpg for truck make i.

H_0: $\mu_1 = \mu_2 = \mu_3$
H_A: at least two means differ ($\mu_i \neq \mu_j$ for some $i \neq j$)

Since the p-value $= .129 > .10$, we do not have sufficient evidence to reject the H_0. We conclude that the mean mpg for the different truck makes could be the same.

(ii) Estimate the proportion of the overall variation in mpg that can be explained by the differing truck types?

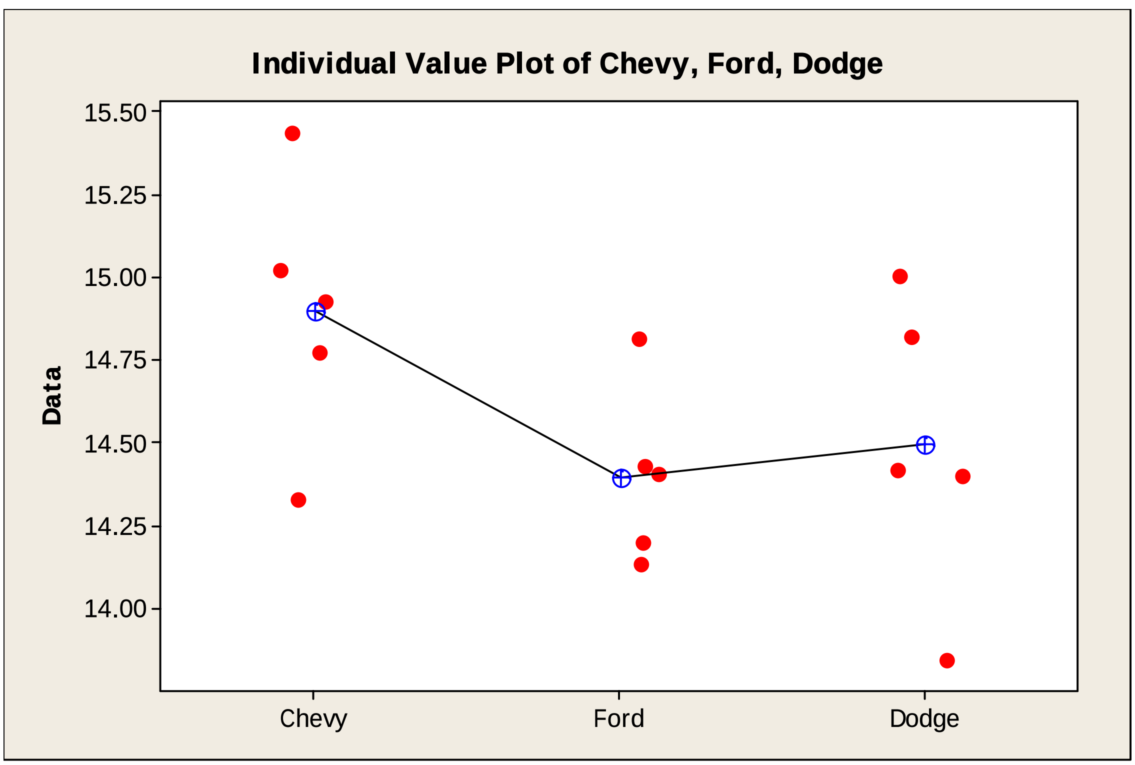
$R^2 = .7/2.42 = .289$

(iii) Create a dataset that satisfies the table above,
then use MINITAB to verify your earlier calculations.
For example, to create the Chevy data:

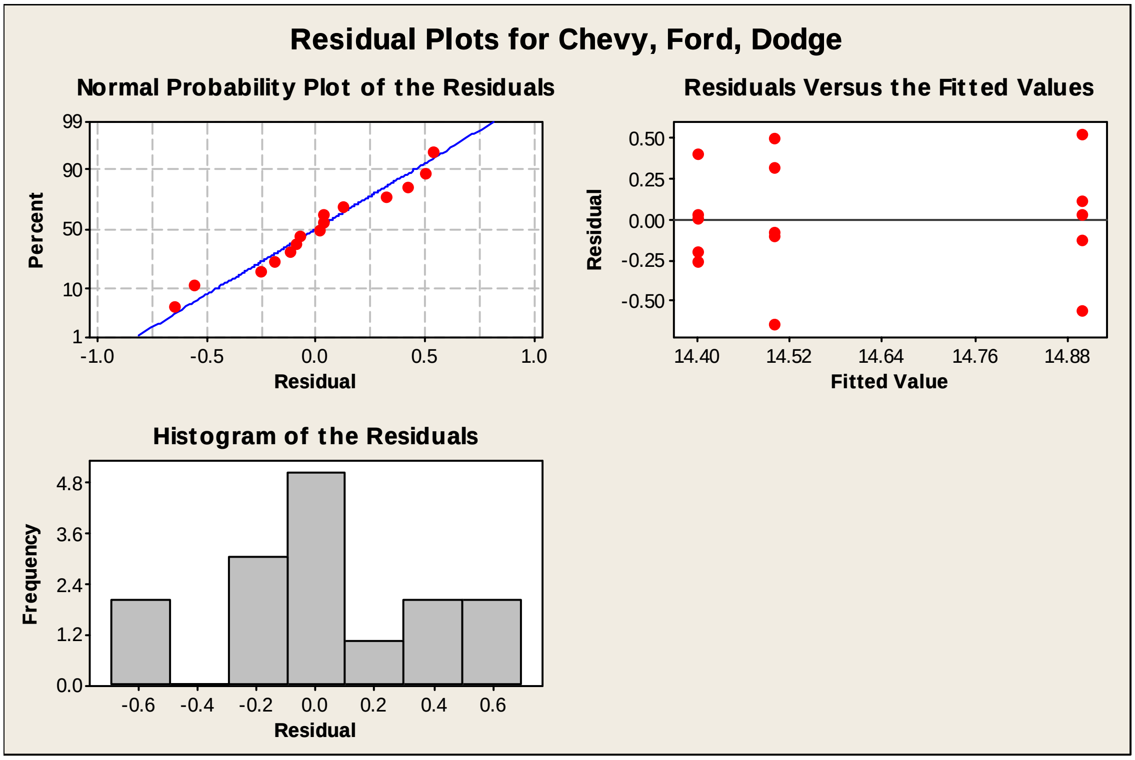
```
MTB> random 5 c91
MTB> let k94=stdev(c91)
MTB> let k93=(.16)**.5
MTB> let c92=(k93/k94)*c91
MTB> let k95=mean(c92)
MTB> let c93=c92-k95+14.9
MTB> print c93
MTB> describe c93
MTB> copy c93 c1
MTB>erase c91-c93
MTB> erase k92-k95
```

						sample mean	sample variance
Chevy	**14.333**	**14.932**	**15.437**	**15.021**	**14.777**	**14.9**	**0.16**
Ford	**14.432**	**14.139**	**14.204**	**14.815**	**14.41**	**14.4**	**0.07**
Dodge	**14.824**	**15.005**	**14.422**	**13.848**	**14.401**	**14.5**	**0.20**

(iv) Using the simulated data to provide appropriate (data and residual) plots to illustrate your conclusions and to check the validity of the one-way model assumptions.



The plot above appears to indicate no significant differences among the means. Also, the model assumption of homoskedasticity appears reasonable.



The linearity of the normal probability plot of the residuals indicates that the model assumption of normal errors is a reasonable assumption. The plot of residuals versus fitted values shows no significant signs of heteroskedastacity for the errors.

(v) What sample sizes are needed for each truck type so that the power is at least .95 to detect a maximum difference in means of 0.50? Assume a commom standard deviation of 0.40.

MINITAB SAMPLE SIZE DETERMINATION...

```
Power and Sample Size
One-way ANOVA
Alpha = 0.05  Assumed standard deviation = 0.4  Number of Levels = 3

          Sample  Target                    Maximum
SS Means    Size   Power  Actual Power  Difference
   0.125      21    0.95      0.951923         0.5

The sample size is for each level.
```

2. A researcher conducted an experiment to compare the effects of three different insecticides on a variety of string beans. To obtain a sufficient amount of data, it was necessary to use four different plots of land. Since the plots had somewhat different soil fertility, drainage characteristics, and sheltering from winds, the researcher decided to conduct a randomized complete block design with the plots serving as blocks. Each plot was subdivided into three rows. A suitable distance was maintained between rows within a plot so that the insecticides could be confined to a particular row. Each row was planted with a 100 seeds and then maintained under the insecticide assigned to the row. The insecticides were randomly assigned to the rows within a plot so that each insecticide appeared in one row within all our plots. The response Y_{ij} of interest was the number of seedlings that emerged per row. The data and means are given in the table below.

		Plot			
Insecticide	1	2	3	4	Insecticide Mean
1	56	48	66	62	58
2	83	78	94	93	87
3	80	72	83	85	80
Plot Mean	73	66	81	80	75

(i) Fill in the blanks to complete the RCB model description:

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \epsilon_{ij} \qquad (1)$$

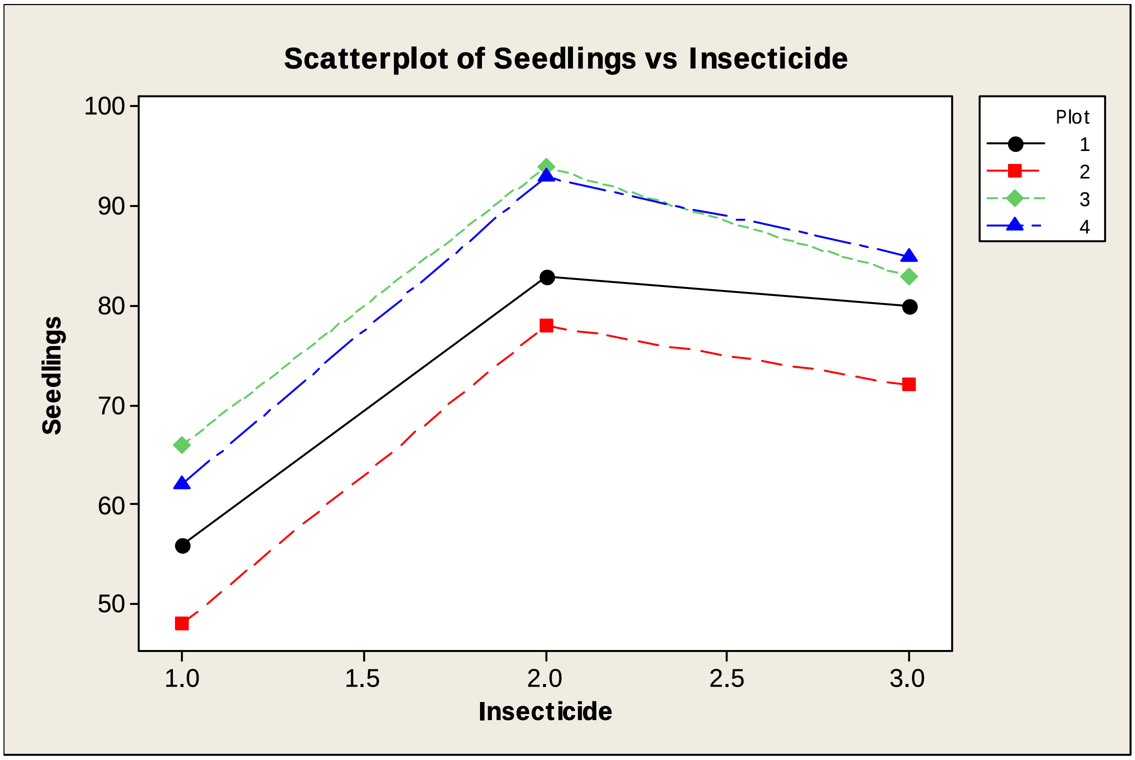
where $\mu_{..}$ is a constant,
ρ_i are constants for the block (plot) effects,
τ_j are constants for the treatment (insecticide) effects,
ϵ_{ij} are independent $N(0, \sigma^2)$ random variables,
$i = 1, ..., 4; \quad j = 1, ..., 3.$

The responses Y_{ij} are independent and normally distributed
with mean $E\{Y_{ij}\} = \mu_{..} + \rho_i + \tau_j$
and variance $\sigma^2\{Y_{ij}\} = \sigma^2$.

(ii) Provide point estimators for all nine parameters (unknown constants) in model 1.

Parameter	**Estimate**	**Parameter**	**Estimate**	**Parameter**	**Estimate**
ρ_1	$\overline{Y}_{1.} - \overline{Y}_{..} = -2$	ρ_4	**5**	τ_3	**5**
ρ_2	$\overline{Y}_{2.} - \overline{Y}_{..} = -9$	τ_1	-17	$\mu_{..}$	**75**
ρ_3	$\overline{Y}_{3.} - \overline{Y}_{..} = 6$	τ_2	**12**	σ^2	**MSE = 4.333**

(iii) Provide a plot of Number of Seedlings versus Insecticide Type (line-connected by plots). Discuss the model assumption of no interactions based on evidence from the plot.



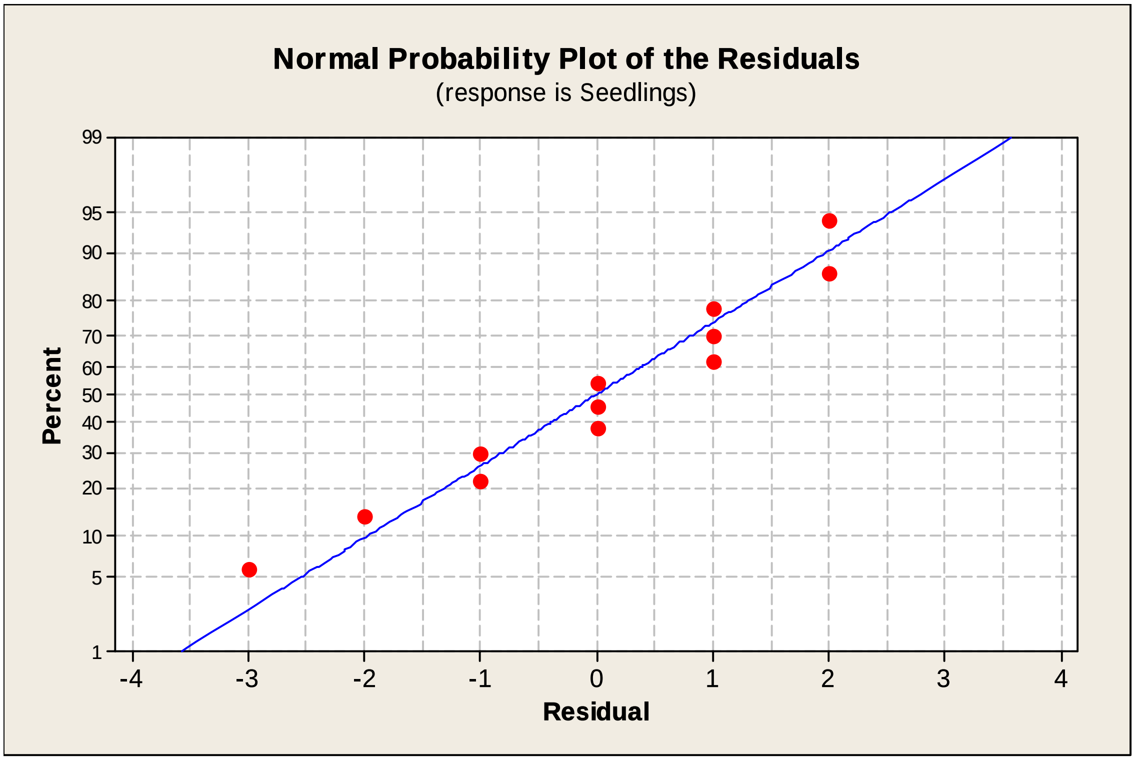
The parallelism in the plot above indicates that there are no significant interactions between insecticides and plots.

(iv) Complete the ANOVA table.

Source of Variation	SS	df	MS	F^*	p-value
Blocks (plots)	438	3	146.00	33.69	0.000
Treatments (insecticides)	1832	2	916.00	211.38	0.000
Error (unexplained)	26	6	4.333		
Total	2296	11			

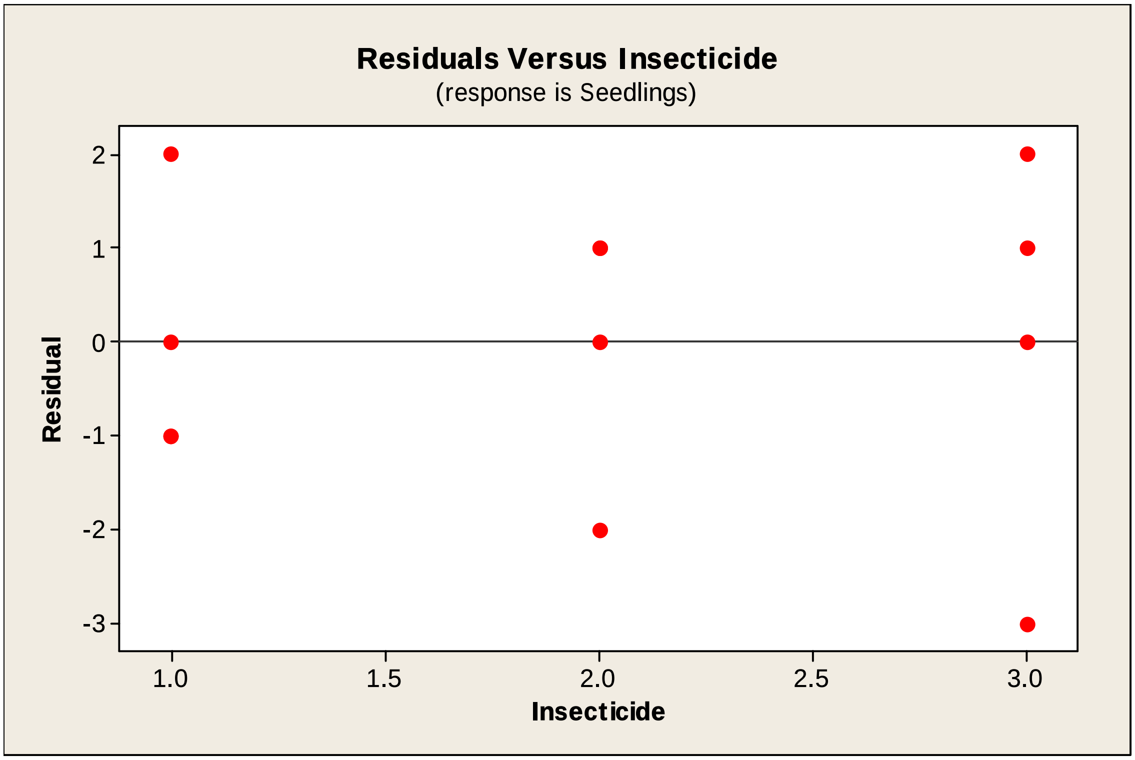
```
Two-way ANOVA: Seedlings versus Plot, Insecticide

Source        DF    SS       MS       F      P
Plot           3   438  146.000   33.69  0.000
Insecticide    2  1832  916.000  211.38  0.000
Error          6    26    4.333
Total         11  2296

S = 2.082   R-Sq = 98.87%   R-Sq(adj) = 97.92%
```

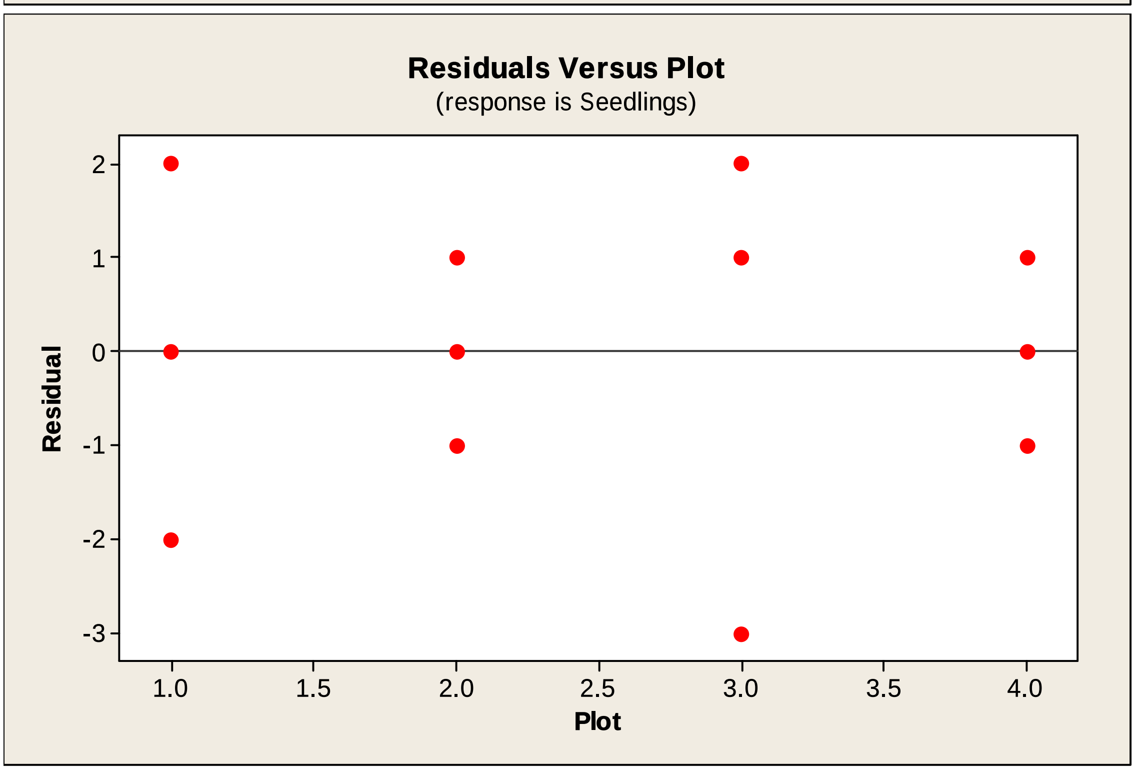
(v) Provide a residual plot that addresses the model assumption of error normality. Does the plot provide convincing evidence against this model assumptions? Discuss.



The linearity in the normal probability plot of residuals above indicates no significant depatures from the model assumption of normall distributed errors.

(vi) Provide two residual plots that address the model assumption of homoskedasticity. Do the plots provide convincing evidence against this model assumption? Discuss.





Both plots above indicate no significant heteroskedasticity of errors.

(vii) Perform a Tukey Test for Additivity (use α=.05).

Hypotheses **H_0: There are no treatment-block interactions.**
H_A: There are some treatment-block interactions.

Obtain the Tukey test statistic and p-value

Since $\sum_i \sum_j (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})Y_{ij} = -247$, **we have**

$$\textbf{SSBL.TR}^* = \frac{(\sum_i \sum_j (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})Y_{ij})^2}{\textbf{(SSBL/r)(SSTR/n}_b\textbf{)}} = \frac{(-247)^2}{(438/3)(1832/4)} = .9124$$

and $\textbf{SSRem}^* = \textbf{SSTO} - \textbf{SSBL} - \textbf{SSTR} - \textbf{SSBL.TR}^*$

$= \textbf{2296} - \textbf{438} - \textbf{1832} - (.9124)$

$= 25.0876$.

Thus the test statistic value is given by

$$F^* = \frac{\textbf{SSBL.TR}^*/1}{\textbf{SSRem}^*/(rn_b - r - n_b)}$$

$$= \frac{.9124}{25.0876/(12-3-4)}$$

$$= 0.1818$$

The p-value $= P(F > .1818 | F_{1,5}) =$ **Fcdf(.1818, 10^99, 1, 5) = .688.**

Conclusion Since the p-value is .688, we do not have sufficient evidence to reject the null hypothesis of no interactions.

(viii). Use $\alpha = .05$ and test for significant insecticide effects.

H_0: $\tau_1 = \tau_2 = \tau_3 = 0$
H_1: at least one $\tau_j \neq 0$

Test statistic: $F^* = \frac{MSTR}{MSE} = 211.38$ **with p-value $< .0005$.**

Conclusion. We have sufficient evidence to conclude that there some statistically significant insecticide effects.

(ix) Estimate the proportion of the total response variability that is explained by the different insecticides used.

$$R^2(insecticide) = \frac{1832}{2296} = .798$$

(x). Perform all 3 pairwise comparisons of the $\mu_{\cdot j}$ (insecticide means).
To lessen the number of negative signs, let $D_1 = \mu_{\cdot 2} - \mu_{\cdot 1},\ D_2 = \mu_{\cdot 3} - \mu_{\cdot 1},$ and $D_3 = \mu_{\cdot 2} - \mu_{\cdot 3}$. Use the Tukey procedure with a family confidence coefficient of .95. After constructing all intervals, note which means differ significantly.

Tukey multiple $T = \frac{1}{\sqrt{2}} q(1-\alpha,\ r,\ (n_b - 1)(r-2))$

$$= \frac{1}{\sqrt{2}}\, q(.95,\ 3, 6) = \frac{4.34}{\sqrt{2}} = 3.0688$$

and $s_{\hat{D}} = \sqrt{MSE(\frac{1}{n_b} + \frac{1}{n_b})} = \sqrt{\frac{4.333}{2}} = 1.472$

Thus margin of error $E = T\ s_{\hat{D}} = \mathbf{3.0688(1.472)} = \mathbf{4.517}$

All pairwise comparisons (Tukey with family confidence .95)

Comparison	Point Estimate	Margin of error	Lower bound	Upper bound	
$D_1 = \mu_{\cdot 2} - \mu_{\cdot 1}$	29	4.52	24.48	33.52	
$D_2 = \mu_{\cdot 3} - \mu_{\cdot 1}$	22	4.52	17.48	26.52	
$D_3 = \mu_{\cdot 2} - \mu_{\cdot 3}$	7	4.52	2.48	11.52	

Therefore, on average, insecticide 2 beats insecticide 3 (at least 2 more seedlings on average) and beats insecticide 1 (at least 24 more seedlings on average). Furthermore, on average, insecticide 3 beats insecticde 1 (at least 17 more seedlings on average).

(xi). Use $\alpha = .05$ and test for significant plot effects.

H$_0$: $\rho_1 = \rho_2 = \rho_3 = \rho_4 = 0$
H$_1$: at least one $\rho_i \neq 0$

Test statistic: $F^* = \frac{MS\mathbf{BL}}{MSE} = \mathbf{33.69}$ **with p-value $<$ .0005.**

Conclusion. We have sufficient evidence to conclude that there some statistically significant plot (block) effects.

(xii) Estimate the proportion of the total response variability that is explained by the different plots used in the experiment.

$$R^2\textbf{(plots)} = \frac{SSBL}{SSTO} = \frac{438}{2296} = .191$$

(xiii) Explain why a randomized complete block design was appropriate for this experiment.

Since the different plots of land have a significant effect on the number of seedlings that emerge, using the plots as blocks, the researcher reduced a considerable amount of the unexplained response variability. Although the effects of the different insecticdes would have been detected without blocking (to see this, run a one-way ANOVA), the pairwise comparisons are more precise with blocking since the error mean square is reduced. Often, if blocks are not used, otherwise significant treatment effects are camouflaged by the resulting unexplained variation.