DATA ANALYSIS EVALUATION

Name: _____ **Matric no.:** _____

Instructions:

Please read the instructions carefully before you answer the questions.

Question	Possible points	Score
1	16	
2	12	
3	34	
4	30	
Total	**92**	

1. You can use a computer to solve the problems and refer to your notes to answer the questions. **NO DISCUSSION IS ALLOWED**. You can ask the instructor if you are stuck with the R codes.
2. The duration to answer the questions is 24 hours and you need to submit within this period. Late submission will be penalized.
3. **Please write all your answers below the questions in this documen**t and **solve the problem using R or Excel softwares**. **DO NOT PRINT** this question and **DO NOT SCAN** the printed question paper to submit your answer. For drawing, please use the drawing tool in this Word document or you can draw on a piece of paper, snap the picture, then insert the image in this document. Make sure everything is clear for grading. Blurry images may affect the marking process and also the marks given.
4. Answer each question (Question 1, 2, 3) on a new page. But you can continue answering the sub-questions (a, b, …, f) within the same page. The marks will be given based on the answer for each question. If the numbering is incorrect, then no marks will be given.
5. **Answer all questions within 12-13 pages. You need to include ONLY relevant R output and R codes to support your answers. If no R codes are provided in the answer that asks for R codes, no marks will be given although you attach your R files separately.**
6. **You need to submit two documents**, which are (i) data analysis (for grading) and (ii) R file that saved your codes and run the analysis (Name.R file) (for my reference).
7. For your answer paper, you must save your work in **pdf** with **your name and matric number (e.g.: Ahmad bin Hamid – Evaluation 2 (GS123456).pdf)**
8. **Similarly, the R codes file should also be saved with your name and matric number.**
9. The marks will be given based on correct answers as well as the originality of your work. Answers among students that are similar or seems the same will be penalized.
10. Please insert the full R codes in the R file as I may need to run the codes in my RStudio during grading. Make sure the codes given are functioning. You can start writing the codes from reading the data until the end of analysis (I will use the path of directory on my computer)
11. **Submit your answer (pdf file) and the R codes (. R file) files through the link given in Putrablast.**

Question 1 (10 marks)

a) What is the purpose of replication in an experiment? (3 marks)

b) In a factorial experiment, is that appropriate to only explain or focus on the main effects of factor A and B if the p-value for interaction between factor A and B is smaller than alpha? For example, the p-value is 0.016 and alpha is 0.05. (4 marks)

c) Explain the theory behind your answer above. Provide an example if needed (3 marks)

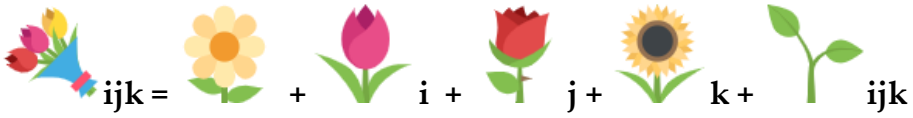
d) Determine the factors of which effect can be assessed and the number of interactions from the following experiments. (6 marks)

Experiments (Factor and levels)	Factors that can be assessed	Number of interactions	Number of treatments
Factor 1: Temperature (25 ^{o}C) Factor 2: Genotype (red and grey) Factor 3: Fertilizer rate (200 kg/ha) Factor 4: Weeding method (none and herbicide)			
Factor 1: Magnesium (0, 10 kg Mg/ha) Factor 2: Potassium (20 and 30 kg K /ha) Factor 3: Nitrogen (80 kg N/ha)			

Question 2 (12 marks)

Answer the following questions using the linear model (with symbols) below.

Equation 1



Where i = treatment, j=block 1, k=block 2

a) What is the possible experimental design based on the above linear additive model (2 marks)

b) Below is the ANOVA from an experiment with i = 4. The linear model for this experiment is given in Equation 1 above. Fill in the blanks (A to P). (8 marks)

Sources of variation	Degree of freedom	Sum of square	Mean square	F value	F table (5%)
Block 1	3	43.21	**G)**	**K)**	**N)**
Block 2	3	15.20	**H)**	**L)**	**O)**
Treatment	**C)**	**F)**	**I)**	**M)**	**P)**
A)	**D)**	15.70	**J)**		
B)	**E)**	200.00			

c) What does this result suggest regarding the blocks in this experiment? (2 marks)

Question 3 (34 marks)

The data below are from an experiment to test the effects of seed priming and types of foliar application on yield of melon (kg/plot). The experiment was conducted under a rain shelter. Other factors such as light interception, temperature, watering, fertilizer and growing media are considered uniform for the whole experiment.

Table 1. Melon yield data (kg/plot) for different foliar fertilizer and seed priming treatment.

Seed priming treatment	Rep	Types of foliar fertilizer		
		Liquid fertilizer	Organic fertilizer	Growth promoter
Control	r1	1.85	1.4	2.33
	r2	1.59	1.96	2.02
	r3	1.83	1.44	2.14
	r4	1.43	1.89	2.52
	r5	1.42	1.56	2.23
Primed	r1	3.15	1.68	1.55
	r2	3.24	1.37	1.33
	r3	3.32	1.88	1.61
	r4	3.76	1.79	1.22
	r5	3.47	1.68	1.78

i. Determine the treatment or factors involved in this experiment and list the number of levels for each factor. What is the possible experimental design used in this experiment? (3 marks)

ii. Write 5 important steps required during data analysis in Rstudio to test the hypothesis regarding the effects of treatments on yield of melon in this experiment. Explain why you need to have the steps that are listed during data analysis. To get full marks, explain step by step how the analysis should be done from reading the data until making a decision and conclusion based on ANOVA from R output. (5 marks)

Step	Description
1	
2	
3	
4	

5	

iii. Use R to show the structure of data required for this analysis. Explain the types of variables in the data and the level shown in the output. (2 marks)

iv. Use R software to determine the effect of priming and foliar on melon yield following the steps you listed above. Show your R codes and outputs relevant to answering the question in order to get full marks. An output without R codes will receive partial marks (10 marks)

v. Do the data meet the assumption for normality and homogeneity of variance? Show your answer and explain. (4 marks)

vi. Present your results of means in a table following a publication format. Include the appropriate title and footnotes to receive full credit. (6 marks)

vii. Interpret your results. (4 marks)

Question 4 (30 marks)

A researcher conducted an experiment to study the effects of potassium (K) rates and planting distance (cm) on yield of tomato. The seedlings were planted in a **1.5 x 0.5 m** plot. The field is located near a pond on the left side (west) of the field thus, the experimental units were arranged in blocks due to slope. The experimental unit was replicated 4 times in each treatment. At the end of growing season, the yield was determined by harvesting the whole plot. Answer the following questions regarding this experiment by analyzing the data provided in Table 1.

Table 1. Yield (kg/plot) of tomato at different K fertilizer rates and planting density.

K fertilizer (kg/ha)	Distance (cm)	Replication			
		R1	R2	R3	R4
25	20	6.52	6.25	6.29	6.26
	40	6.50	6.27	6.25	6.21
	60	6.66	6.49	6.34	6.37
50	20	7.70	7.47	7.33	7.45
	40	7.66	7.39	7.27	7.24
	60	7.67	7.40	7.39	7.37
100	20	8.06	7.75	7.74	7.73
	40	8.14	7.97	7.85	7.88
	60	7.98	7.76	7.69	7.66

a) The table given above is considered a 'wide data' format. To analyze the data in the R software, you need to convert this data to a 'long data' format. Explain how to convert the format using R and show the codes you used and the result table after conversion. (3 marks)

b) What is the potential experimental design used in this experiment and explain the reason for your answer. Sketch the layout of this experimental design (8 marks)

c) Run an analysis of variance to test the hypothesis that there is no significant interaction between K fertilizer and planting distance. Assign density as factor 1 and density as factor 2. State the hypothesis for the interaction. Write your ANOVA in the table below.

i. Hypothesis for interaction: (2 marks)

Ho:

Ha:

ii) ANOVA R codes (3 marks)

(Include the codes in this section here to get marks)

iii) ANOVA table (4 marks)

*(**Notes: When writing the values in the table, use only 2-4 decimal places only or use standard form (e.g. 1 x 10^{-9}))*

Source of variation	Degree of freedom	Mean square	F value	Significance stars

d) Based on ANOVA result, conduct a mean comparison using DMRT for the treatment in this experiment. Determine the most appropriate way to present the result. Include the appropriate title and footnotes to receive full credit. (5 marks)

e) Interpret your results above and determine which K fertilizer rate and density that you recommend to farmers to get the highest yield? Also, which one should be avoided? (5 marks)