a. Using baseline constraints, write the linear model for this two factor experiment in vector-matrix notation form.

The form of the linear model is $\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The vector $\boldsymbol{Y}$ contains the 36 values in the table above, in order down each cell and then across the rows.

The parameters in the vector $\boldsymbol{\beta}$ will be $(\mu, \alpha_1, \alpha_2, \tau_1, \tau_2, \alpha\tau_{11}, \alpha\tau_{12}, \alpha\tau_{21}, \alpha\tau_{22})^T$

*The matrix **X** will have 36 rows. Only 9 rows, one for each treatment will be unique. Each of these 9 unique rows will appear in the X matrix 4 times. For the baseline constraints, these 9 unique rows will be:*

μ	α_1	α_2	τ_1	τ_2	$\alpha\tau_{11}$	$\alpha\tau_{12}$	$\alpha\tau_{21}$	$\alpha\tau_{22}$
1	1	0	1	0	1	0	0	0
1	1	0	0	1	0	1	0	0
1	1	0	0	0	0	0	0	0
1	0	1	1	0	0	0	1	0
1	0	1	0	1	0	0	0	1
1	0	1	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0
1	0	0	0	1	0	0	0	0
1	0	0	0	0	0	0	0	0

b. Using sum to zero constraints, write the linear model for this two factor experiment in vector-matrix notation form.

Everything will be the same in this statement of the linear model, except for the X matrix.

The parameters in the vector $\boldsymbol{\beta}$ will be $(\mu, \alpha_1, \alpha_2, \tau_1, \tau_2, \alpha\tau_{11}, \alpha\tau_{12}, \alpha\tau_{21}, \alpha\tau_{22})^T$

*The matrix **X** will have 36 rows. Only 9 rows, one for each treatment will be unique. Each of these 9 unique rows will appear in the X matrix 4 times. For the sum to zero constraints, these 9 unique rows will be:*

μ	α_1	α_2	τ_1	τ_2	$\alpha\tau_{11}$	$\alpha\tau_{12}$	$\alpha\tau_{21}$	$\alpha\tau_{22}$
1	1	0	1	0	1	0	0	0
1	1	0	0	1	0	1	0	0
1	1	0	-1	-1	-1	-1	0	0
1	0	1	1	0	0	0	1	0
1	0	1	0	1	0	0	0	1
1	0	1	-1	-1	0	0	-1	-1
1	-1	-1	1	0	-1	0	-1	0
1	-1	-1	0	1	0	-1	0	-1
1	-1	-1	-1	-1	1	1	1	1

c. Use SAS to fit the two-factor model with interactions. Which type of constraints is SAS using to fit the model – baseline constraints or sum to zero constraints? Explain how you can tell from the SAS output.

It is fitting the parameters using baseline constraints. The values of $\alpha_3, \tau_3,$ and any interaction term $(\alpha\tau)_{ij}$ with either $i = 3$ or $j = 3$ are set to 0.

d. Use the estimated parameters from the SAS output to calculate the nine sample treatment means. Hint: Start with the level 3 treatments for both factors first.

Here is the edited table from the SAS output with the 9 parameter estimates.

Parameter	Estimate
Intercept	57.9
variety 1	18.05
variety 2	12.525
nitrogen 150	10.7
nitrogen 210	6.625
variety*nitrogen 1 150	-20.125
variety*nitrogen 1 210	-13.6
variety*nitrogen 2 150	-19.675
variety*nitrogen 2 210	-14.5

Here is the table of treatments, with the calculation of the values of the means.

Sugar Cane Variety	Nitrogen Level = 150 lbs/acre	Nitrogen Level = 210 lbs/acre	Nitrogen Level = 270 lbs/acre
1	57.9 + 18.05 + 10.7 – 20.125 = **66.525**	57.9 + 18.05 + 6.625 – 13.6 = **68.975**	57.9 + 18.05 = **75.95**
2	57.9 + 12.525 + 10.7 – 19.675 = **61.45**	57.9 + 12.525 + 6.625 – 14.5 = **62.55**	57.9 + 12.525 = **70.425**
3	57.9 + 10.7 = **68.6**	57.9 + 6.625 = **64.525**	**57.9**

e. Use the ANOVA table from the SAS output to test for the significance of the nine treatments on the sugar cane yield. State your conclusions based on this analysis.

The null hypothesis is that all 9 treatment means are the same and the alternative hypothesis is that at least one of them is different. Here is the ANOVA table from the SAS output.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	935.702222	116.962778	2.52	0.0346
Error	27	1254.460000	46.461481		
Corrected Total	35	2190.162222			

The F-test statistic value is 2.52 with a p-value of 0.0346. Since this p-value is smaller than 0.05, we will reject the null hypothesis and conclude at least one of the 9 treatment means is different.

f. Use the ANOVA table from the SAS output to test for the main effect of variety on the sugar cane yield, averaging over the three levels of nitrogen. State your conclusions based on this analysis.

Here is the ANOVA Table from the SAS output.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
variety	2	319.3738889	159.6869444	3.44	0.0468
nitrogen	2	56.5405556	28.2702778	0.61	0.5515
variety*nitrogen	4	559.7877778	139.9469444	3.01	0.0355
Error	27	1254.460000	46.461481		
Corrected Total	35	2190.162222			

The null hypothesis is that the three means for sugar cane varieties are the same and the alternative hypothesis is that at least one of these means is different. The F-test statistic from the ANOVA Table is 3.44 with a p-value of 0.0468. Since this p-value is less than 0.05, we will reject the null hypothesis and conclude at least one of the three means for sugar cane varieties is different.

g. Use the ANOVA table from the SAS output to test the main effect of nitrogen. State your conclusions based on this analysis.

The ANOVA table is the same as in problem f.

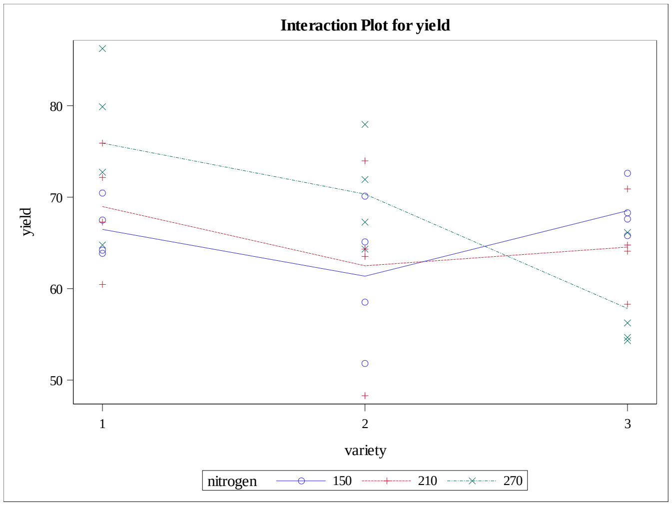
The null hypothesis is that the three means for nitrogen levels are the same and the alternative hypothesis is that at least one of these means is different. The F-test statistic from the ANOVA Table is 0.61 with a p-value of 0.5515. Since this p-value is greater than 0.05, we will fail to reject the null hypothesis and conclude the three nitrogen means are the same.

h. Use the ANOVA table from the SAS output to answer research question on the significance of the interaction between the sugar cane variety and the nitrogen levels. State your conclusions based on this analysis.

The ANOVA table is the same as in problem f.

The null hypothesis is that there is no interaction between the two factors and the alternative hypothesis is that there is an interaction between the two factors. The F-test statistic from the ANOVA Table is 3.01 with a p-value of 0.0355. Since this p-value is less than 0.05, we will reject the null hypothesis and conclude there is a significant interaction *between the two factors. So, the mean of sugar cane variety depends on the nitrogen level.*

i. Study the interaction plot in the SAS output. Explain why, based on this plot, the interaction is significant in the model.



Here is the interaction plot from SAS. The plot shows differences in the pattern of the means of the sugar cane varieties over the nitrogen levels. The pattern for nitrogen levels 150 and 210 are somewhat similar. However, the pattern for nitrogen level 270 is very different. Noticeable differences in these patterns indicate a significant interaction between the two factors.

j. Use Tukey's HSD method to make pairwise comparisons of the marginal means for the three sugar cane varieties. State your conclusions based on this analysis.

The Tukey HSD output from SAS indicates all marginal means for the three sugar cane varieties are not significantly different from each other.

k. Use Tukey's HSD method to make pairwise comparisons of the marginal means for the three nitrogen values. State your conclusions based on this analysis.

The Tukey HSD output from SAS indicates all marginal means for the three nitrogen levels are not significantly different from each other.

l. Use orthogonal contrasts to compare (1) the means of varieties 1 and 2 with variety 3 and (2) varieties 1 and 2. Give the values of the vector ***c*** for both contrasts and explain why these two contrasts are orthogonal. State your conclusions based on this analysis.

*The values of **c** for each contrast are:*

Varieties 1 and 2 – (1, -1, 0)
Means of Varieties 1 and 2 versus variety 3 – (0.5, 0.5, -1)

These two contrasts are orthogonal since 1(0.5) + (-1)(0.5) + 0(-1) = 0

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
v1-v2	1	193.2337500	193.2337500	4.16	0.0513
(v1+v2)-v3	1	126.1401389	126.1401389	2.71	0.1110

Neither of the contrasts are statistically significant at the 5% level, however, the first contrast is very close to this significance level.

m. Use orthogonal contrasts to test for a (1) linear and (2) quadratic relationship between the three levels of nitrogen and sugar cane yield. Give the values of the vector $\boldsymbol{c}$ for both contrasts and explain why these two contrasts are orthogonal. State your conclusions based on this analysis.

The values of $\boldsymbol{c}$ *for each contrast are:*

Linear – (-1, 0, 1)
Quadratic – (-1, 2, -1)

These two contrasts are orthogonal since (-1)(-1) + 0(2) + 1(-1) = 0

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
linear	1	39.5266667	39.5266667	0.85	0.3645
quadratic	1	17.0138889	17.0138889	0.37	0.5501

Neither of the contrasts are statistically significant at the 5% level, indicating there is not a linear or quadratic relationship between the three levels of nitrogen and sugar cane yield.

n. Based on your analysis, which variety would you recommend planting on a field with a low nitrogen level? What about a high nitrogen level? Explain your answer.

The significance in the model is coming from the interaction terms. So, we should consider planting different sugar cane varieties depending on the level of nitrogen in the field. Using the interaction plot, we can see that for a low level of nitrogen, variety 3 would produce the largest yield. For a high level of nitrogen, variety 1 would produce the largest yield.

o. Study the residuals versus predicted value plot. Is there anything of concern in this plot?

The points in the plot do not appear to have any patterns, so there is nothing of concern in this plot.

p. Study the residuals by varieties plot and the residuals by nitrogen levels plot. Is there anything of concern in this plot?

The variation in the residuals by varieties and by nitrogen levels do not show any large differences. The residuals have a slightly smaller variation for nitrogen level 270, but the difference isn't enough to affect our results.

q. Study the normal probability plot for the residuals. Is there anything of concern in this plot?

No, the points fall in a straight line pattern, indicating the normal distribution assumption is met.