The purpose of lab today is to work through the data analysis using SAS for the two-factor experiment.

1. Data set for the protein yield example involving factors copper concentration and zinc concentration. Corresponding SAS code is **protein.sas**.

2. Practice what we have learned about two-factor factorial treatment design in a CRD using a second example. The data come from an experiment to determine the best combination of sugar cane variety and amount of nitrogen on the yield of sugar cane. The variables for the experiment were:

   Factor A: Three varieties of sugar cane (1, 2, 3)
   Factor B: Three levels of nitrogen (150, 210, 270), all in lbs/acre
   Response Variable: Yield of sugar cane (tons/acre)

   This is a full factorial treatment design including nine treatments. The nine treatments were randomly applied to 4 plots of land each, giving a total of 36 plots (experimental units). Hence, the experimental design is CRD. The data are given in the table below. Corresponding SAS code is **cane.sas**

Sugar Cane Variety	Nitrogen Level = 150 lbs/acre	Nitrogen Level = 210 lbs/acre	Nitrogen Level = 270 lbs/acre
1	70.5 67.5 63.9 64.2	67.3 75.9 72.2 60.5	79.9 72.8 64.8 86.3
2	58.6 65.2 70.2 51.8	64.3 48.3 74.0 63.6	64.4 67.3 78.0 72.0
3	65.8 68.3 72.7 67.6	64.1 64.8 70.9 58.3	56.3 54.7 66.2 54.4

   a. Using baseline constraints, give the nine unique rows in the design matrix X and give the parameter vector $\boldsymbol{\beta}$ for this two factor experiment.
   b. Using sum to zero constraints, give the nine unique rows in the design matrix X and give the parameter vector $\boldsymbol{\beta}$ for this two factor experiment.
   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.
   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.
   e. Use the ANOVA table from the SAS output for the overall F-test for the effects of the nine treatments on the sugar cane yield. State your conclusions based on this analysis.
   f. Use the ANOVA table from the SAS output to test for the main effect of variety on the sugar cane yield. State your conclusions based on this analysis.
   g. Use the ANOVA table from the SAS output to test for the main effect of nitrogen on the sugar cane yield. State your conclusions based on this analysis.

h. Use the ANOVA table from the SAS output to test for the interaction effect between variety and nitrogen on the sugar cane yield. State your conclusions based on this analysis.
i. Study the interaction plot in the SAS output. Explain why, based on this plot, the interaction is significant in the model.
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.
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.
l. Use orthogonal contrasts to compare (1) the average of varieties 1 and 2 with variety 3 and (2) varieties 1 and 2. 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.
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.
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.
o. Study the residuals versus predicted value plot. Is there anything 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?
q. Study the normal probability plot for the residuals. Is there anything of concern in this plot?