Problem 1

A multivariate data set contains 42 measurements on air-pollution variables in Los Angeles. The columns are x_1 = wind, x_2 = solar radiation, x_3 = CO, x_4 = NO, x_5 = NO_2, x_6 = O_3, and x_7 = HC.

1. Using notation consistent with our lessons, explain briefly what X_{ij} and $\overline{X}_j$ represent in terms of the variables here. Similarly, explain briefly what μ_j and σ_{jk} represent. What does it mean if $\sigma_{jk} = 0$?

2. Explain briefly what it would mean if the correlation between solar radiation and CO is close to 1, but the partial correlation between solar radiation and CO is close to 0, given NO.

3. Consider only solar radiation and CO (carbon monoxide) for these parts.

   (a) Draw by hand or describe in words what you expect a QQ plot to look like if these variables have a multivariate normal distribution.

   (b) Repeat part (a) if these variables do not have a multivariate normal distribution.

   (c) With sample mean vector $[74.0, 5.0]'$ and sample covariance matrix

   $$\begin{bmatrix} 300.0 & 4.0 \\ 4.0 & 1.5 \end{bmatrix}$$

   compute a 95% confidence interval for the population mean of solar radiation and a 95% confidence interval for the population mean of CO. Show your work for this.

   (d) Explain why one may *not* say that "with 95% confidence, both of the intervals above in (c) cover their respective parameters simultaneously."

Problem 2

Given $\mathbf{X} = [X_1, X_2, X_3]' \sim N(\mu_{\mathbf{X}}, \mathbf{\Sigma}_{\mathbf{X}})$ with

$$\mu_{\mathbf{X}} = \begin{bmatrix} -3 \\ 1 \\ 4 \end{bmatrix} \quad \text{and} \quad \mathbf{\Sigma}_{\mathbf{X}} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

what is the conditional distribution of $X_1 + X_2 + X_3$ given that $2X_1 - X_2 - X_3 = c$ and $X_3 - X_2 = d$?

Problem 3

Consider the SAS code `https://onlinecourses.science.psu.edu/stat505/sites/onlinecourses.science.psu.edu.stat505/files/sas/ellplot.sas` from our lessons.

1. How would this code need to be altered to produce 90% prediction ellipses for a bivariate normal distribution with mean vector and covariance matrix

   $$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2.0 & -.5 \\ -.5 & 1.0 \end{bmatrix}$$

2. Include both the original and modified ellipses, or explain in words what qualities you would see with these plots. Comment on their differences.