1. Physical fitness testing is an important aspect of athletic training. A commonly used cardiovascular fitness metric is the maximum volume of oxygen uptake during strenuous exercise. A study was undertaken on 24 men aged 40-50 to examine relationship between time taken to run two miles (\(X\), measured in seconds) and maximum oxygen uptake (\(Y\), measured in liters per minute). Oxygen uptake was measured using standard laboratory methods as the subjects ran on a motor-driven treadmill. The data set is available from the BIOSTAT 523 Course Tools page in the “Data Sets” folder (oxygen1.txt), with the fields being (left to right) (i) subject ID number (ii) oxygen uptake (iii) time.

- Note: Include your SAS code at the back of your homework. The SAS output should be interleaved with your responses to the questions below. Only submit the portions of the output specifically requested.

(a) Using SAS, plot \(O_2\)-uptake (vertical axis) versus time (horizontal axis). Submit your plot and SAS code.

(b) Using SAS, fit a straight line model with plot \(O_2\)-uptake as the response and time as the covariate. Submit the PROC REG output (1 page).

(c) Plot (i) \(\hat{O}_i\) versus \(T_i\) and (ii) \(O_i\) versus \(T_i\) on the same graph. Does the SLR model appear to fit the data?
2. Using only PROC MEANS and/or PROC CORR, and some hand calculations where necessary, compute \( \hat{\beta}_1 \) and \( \hat{\beta}_0 \) for the oxygen data from Q1. Show your relevant SAS output (1 page) along with your calculations.

3. For the study described in Q1, the PI indicates that her 22-year old daughter recently completed a two-mile race in 1,000 seconds.

   (a) Using the regression function estimated in Q1, estimate her daughter’s \( O_2 \) uptake. Do you have any reason to believe that your estimate may be inaccurate?

   (b) Suppose that all the \( X_i \) values were replaced by \( (X_i - \bar{X}) \). It can be shown that such centering of the covariate has no impact on \( \hat{\beta}_1 \) for the straight line model. What impact would this have on the interpretation of \( \hat{\beta}_0 \)?

   (c) Using PROC REG (following some data manipulations), fit a SLR model such that the estimated mean oxygen uptake for someone with a two-mile run time of 1000 seconds can be read directly from the parameter estimates in the SAS output; i.e., without creating an OUTPUT data set or any MODEL statement options.

4. We now consider interval estimation. In particular, consider individuals who run 2 miles in 700 seconds. How much wider is the confidence interval than the prediction interval? Briefly describe the reason for the discrepancy.
5. A future study will be conducted which will further assess the correlation between two-mile run time and oxygen uptake. The investigators wish to obtain a slope estimate with a lower variance.

(a) If you had to suggest only one modification to the previous study, what would it be?

(b) Suppose that, due to budgetary constraints, the next study will also contain exactly \( n = 24 \) participants. If you had to offer one suggestion to increase the precision (lower the variance) of \( \hat{\beta}_1 \), what would it be?

6. For a straight line regression model, prove that the fitted regression line always passes through the point \((\overline{X}, \overline{Y})\).