

## Multivariate Least Squares Regression Line – Teacher Notes

1. After the first **lm** command the student gets the result:  $y = 2.26510 + 0.38789x_1 - 0.03106x_2$ . The y-intercept is 2.2651. The contribution to the model by an incremental increase by the  $x_1$  variable is 0.38789 and the incremental decrease by the  $x_2$  variable is 0.03106. The correlation coefficient is low (0.1978), so this isn't an ideal combination of variables to predict the risk of infection.
2. The correlation coefficient is 0.5697 for the second model using Culture (or Lab) and X.Ray. This is certainly a much more relevant model for predicting infection than length of stay and patient age.
3. In the forward selection procedure the resulting line of best fit does not include the  $x_2$  variable. This variable does not make a statistically significant contribution to the correlation coefficient for the equation so it is eliminated.
4. In the backward selection procedure the students will get the same results as the backward selection model. Again, the  $x_2$  variable will be eliminated from the equation.
5. When the student runs the stepwise selection model, it stops after two iterations because the method has determined that the best model is the one with all of the variables except  $x_2$ .