Topic 6 problems: Model fitting

1. Use cross-validation to choose between the three models (linear, quadratic, and quadratic-sinusoidal) that you fit to the Keeling curve in the problems for lecture topic 4 (the general linear model). You can follow the same general approach as in the lecture code scripts `xval1.R` and `xval2.R`. That is, for each of the three models, do the following. (a) Divide the Keeling data points randomly into ten groups. (b) Fit the model to the data in nine of the groups (the training data). (c) Find the model's error (e.g., the sum-of-squares error) in predicting the data in the tenth group (the validation data). (d) Repeat the previous two steps, finding the model's prediction error with each of the ten groups of data as a validation set. See which of the three models has the lowest average prediction error on the validation data.

   As an exercise in good coding practice, try to avoid repeating code as much as possible. For example, you could write a cross-validation function that fits an arbirary function $f(x)$ to the training data and finds its average error on the validation data. Then you can just call that cross-validation function three times, once with a function that makes a linear fit, once with a function that makes a quadratic fit, and once with a function that makes a quadratic-sinusoidal fit. One approach to writing code like this would be to first write the cross-validation code to test the linear fit, and then when the code is written and working, modify it so that it can test an arbitrary function $f(x)$ instead of just a linear function.

2. Wichmann and Hill (2001) show that it is important to include a *lapse rate* parameter when fitting a psychometric function. This is a parameter that gives the psychometric function a minimum other than zero, and a maximum other than one. This models the fact that on some trials an observer's attention wanders, and the observer simply guesses the correct response.

   For example, if we add a lapse rate parameter λ to the cumulative normal psychometric function that we used in the lecture code, we get:

   $$\psi(x; \mu, \sigma, \lambda) = \lambda + (1 - 2\lambda)\Phi(x; \mu, \sigma)$$

   This function has a minimum value λ and a maximum value $1 - \lambda$.

   Write an R script that fits this function to the made-up data we used in the lecture code.