Special Topics

Some complex model-building problems can be handled using the linear regression approach covered up to this point.

- For example, piecewise regression, including piecewise linear regression and spline regression.

Some require more general nonlinear approaches.

- For example, logistic and probit regression for binary responses.
Piecewise Regression

Consider the compressive strength of concrete:

cement <- read.table("Text/Exercises&Examples/CEMENT.txt",
                        header = T)
with(cement, plot(RATIO, STRENGTH))

RATIO is the ratio of water to cement (by mass). The compressive strength decreases as the ratio increases.

We could try a quadratic model:

lq <- lm(STRENGTH ~ RATIO + I(RATIO^2), cement)
summary(lq)
curve(predict(lq, data.frame(RATIO = x)), add = TRUE)
Sometimes, theory may suggest that a relationship is a straight line, but with different slopes in different parts of the data.

For example,

\[
E(Y) = \begin{cases} 
\beta_0 + \beta_1x & x \leq \xi, \\
\beta^*_0 + \beta^*_1x & x > \xi.
\end{cases}
\]

Usually we want the model to be continuous at \( x = \xi \):

\[
\beta_0 + \beta_1\xi = \beta^*_0 + \beta^*_1\xi
\]

Instead of imposing this as a constraint, the model is usually reparametrized as

\[
E(Y) = \beta_0 + \beta_1x + \beta_2 \max(x - \xi, 0).
\]
In the example, suppose the theory suggests that $\xi = 70$ is the critical ratio:

```r
xi <- 70
lp70 <- lm(STRENGTH ~ RATIO + pmax(RATIO - xi, 0), cement)
summary(lp70)
curve(predict(lp70, data.frame(RATIO = x)), add = TRUE,
     col = "red")
```

Note the slightly higher $R^2$ and $R_a^2$. 
In fact, $\xi = 65$ gives even higher $R_a^2$, but is not suggested by theory.

We can treat $\xi$ as an additional parameter, to be estimated, but the model is no longer a linear regression model.

Nonlinear fitting shows that $\xi$ is not significantly different from 70.
Sometimes the lines are *not* constrained to be continuous.

For example, reading scores:

```r
rScores <- read.table("Text/Exercises&Examples/READSCORES.txt", header = T)
with(rScores, plot(Age, ReadScore))
```

To fit separate lines for $\text{Age} \leq 14$ and $\text{Age} > 14$, include the interaction of $\text{Age}$ and the indicator variable for $\text{Age} > 14$:

```r
l14 <- lm(ReadScore ~ Age * (Age > 14), rScores)
summary(l14)
curve(predict(l14, data.frame(Age = x)), to = 14, add = TRUE, col = "blue")
curve(predict(l14, data.frame(Age = x)), from = 15, add = TRUE, col = "blue")
```
Spline Functions

The continuous piece-wise linear model that was used for the cement example is a simple example of a **spline function**.

The break-point $\xi$ is a *knot*.

Several knots could be used, if an argument could be made for them.

For each knot $\xi_j$, include $\max(x - \xi_j, 0)$ as an additional term in the model.
More generally, we could use a piecewise \textit{polynomial} model.

For example, a \textit{cubic} spline:

\[
E(Y|x) = s(x),
\]

where:
- between knots, \( s(x) \) is a cubic polynomial;
- at each knot, \( s, s', \) and \( s'' \) are continuous.

Imposing the continuity constraints looks difficult; instead, reparametrize (well, not really...) as

\[
E(Y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{j=1}^{J} \beta_{3+j} \max(x - \xi_j, 0)^3.
\]