MATH550/SCC461: Statistics in Practice
Lab 1
Assessment

Tom Palmer

Before you start

- Each weekly coursework involves you submitting an R script via the module page on moodle.
- The coursework is marked out of 10.
- You must upload a pdf version of the knitted version of your .R script – do not upload your .R script itself.
- Include your library card number at the top of your document.
- Each mark is effectively worth 0.5% of your final grade for this module. Do not scoff at these, they add up fast. But equally, do not panic if you miss a few marks, there is plenty of room for trial and error.
- **Deadline** for coursework submission is **9am Monday** 9^{th} **October 2017**.
- I will let you know how well you have done before the following lab.
- I will upload my solution to Moodle. You should download this and compare it to your own code.
  Warning Since I will upload my solutions on Monday late work will score 0.

Assessment

There are six parts to the coursework. Before you start, work through the content below. Then complete the tasks.

Accumulation of pollutant

Toxins are often not dangerous in small quantities. However many toxins can accumulate in an individual and eventually build up to dangerous levels. Polychlorinated biphenyl (PCB) is such a toxin. PCBs were widely used as coolants in electrical apparatus until they were recognised as a persistent organic pollutant. Data were collected on the concentration of PCB residues in lake trout from Cayuga Lake, NY and reported in Bache *et al.* (1972). The data consist of the age (years) and the PCB residues in parts per million (ppm) for each fish.

In Bates and Watts (1988) a linear model is used to predict log(PCB) concentrations as a function of the age of a fish:

$$l = a + b \times age^{\frac{1}{3}},$$

where l is the log(PCB) concentration and a and b are unknown constants.

The (least squares) estimates of the parameters a and b are $a = -2.3907$ and $b = 2.300$. We can then use the equation for l (log(PCB)) along with a and b to add a line to the graph to show how we expect l to vary with age. To do this we need to generate a vector of ages covering the range that we observed in the table earlier:

```
>     ages <- seq(from=0, to=12, by=0.1)
>     a <- -2.3907
>     b <- 2.300
>     l <- a + b*ages^(1/3)
```

The number of elements in `l` should be the same as `ages`:

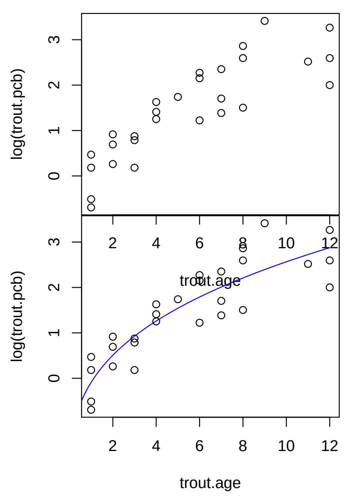
```
>     length(ages)
>     length(l)
```

To add the `ages` and `l` data to the graph which we have already created, we use the `points()` function like so:

```
>     plot(x=trout.age, y=log(trout.pcb))
>     points(x=ages, y=l, type="l", col="blue")
```

Table 1: Observations of age of lake trout with PCB concentrations (ppm) taken from (Bache *et al.*, 1972).

Age	PCB (ppm)
1	0.6
1	1.6
1	0.5
1	1.2
2	2
2	1.3
2	2.5
3	2.2
3	2.4
3	1.2
4	3.5
4	4.1
4	5.1
5	5.7
6	3.4
6	9.7
6	8.6
7	4
7	5.5
7	10.5
8	17.5
8	13.4
8	4.5
9	30.4
11	12.4
12	13.4
12	26.2
12	7.4



References

Bache, C. A., Serum, J. W., Youngs, W. D., and Lisk, D. J. (1972). Polychlorinated biphenyl residues: Accumulation in Cayuga Lake trout with age. *Science* **117**, 1192–1193.

Bates, D. M., and Watts, D. G. (1988). Nonlinear Regression Analysis and Its Applications. Wiley, New York.

Tasks

1. Write a script for the log(PCB) against age, reproducing the final plot containing both the equation line and the data points.

2. Rewrite the log(PCB) equation as a function which has arguments; `a`, `b` and `age`, and returns the predicted log(PCB). Print this function.

   [3 marks]

3. By extending the range of age considered, produce a plot which shows the curve for the expected log(PCB) concentration for lake trout up to 20 years old. Print this plot.

   [2 marks]

4. Now extract the maximum expected log(PCB) from the values used to draw the equation line.

   [1 mark]

5. It can be shown that a non-linear model of the form

   $$l = a + b \times age^{c},$$

   where a, b and c are constants provides a slightly better fit to the data. The optimal choices are $a = -4.865$, $b = 4.7016$ and $c = 0.1969$.

   (a) Rewrite the log(PCB) equation as a function which has arguments; `a`, `b`, `c` and `age`, and returns the predicted log(PCB). Print this function. [1 marks]

   (b) Compare the Bates-Watts estimator and the new estimator for the expected log(PCB) concentration of a 10 year old lake trout. [1 marks]

   (c) Create a new plot which has both the old line and new line, allowing a comparison of the differences. Print this. [2 marks]