Are Kenna and Berche unfair to Pure Mathematics? Let's look at the data.

by P.M.E.Altham, Statistical Laboratory, University of Cambridge. March 2013

Abstract
I investigate the 'segmented lines' fit of data used by Kenna and Berche (2012) to assess the relationship between the 2008 RAE 'score' and the Department size, for the UK Statistics and OR Departments, and then for the UK Pure Mathematics Departments.

The journal 'Significance' in December 2012 contained an interesting article by Ralph Kenna and Bertrand Berche called 'Statistics of Statisticians: Critical masses for research groups'. This looked at the large online dataset from the UK Research Assessment Exercise of 2008. Here subjects were grouped by 'unit of assessment', and below I have reproduced a subset of the data for Unit 22, which is 'Statistics and Operational Research'. According to Kenna and Berche, the RAE 'score' for a department is calculated as
score $= X4s + (3 * X3s + X2s)/7$
where for example $X4s$ is the percentage of FTEstaff in the $4*$ category (ie the top category).
It may be seen that, roughly speaking, this score increases as the number of FTEstaff increases, but the relationship is non-linear, and not really quadratic either. We download the R package 'segmented' due to Vito Muggeo in order to fit the following model
$score = a_1 + b_1 N$ for $N \leq N_{crit}$, and
$score = a_2 + b_2 N$ for $N > N_{crit}$.
Here $a_1, b_1, a_2, b_2, N_{crit}$ are all parameters to be estimated, with N_{crit} being of special interest: it is the 'critical mass' of a research group, also referred to as the Dunbar number.
Kenna and Berche find that for this dataset, $N_{crit} = 17$, with $se = 6$, and I agree, using the 'segmented' package. However, I think they are wrong in their analysis of the Pure Mathematics data.
(Note that Kenna and Berche give no details of their calculation.)
First, the first few rows of data for Statistics and OR departments, with the university names shortened by me, for ease of plotting.

```
             FTEstaff X4s X3s X2s X1s unclassified
Bath            15.00  20  40  35   5            0
Bristol         23.00  25  45  30   0            0
Brunel          10.00  15  35  40  10            0
Cambridge       16.00  30  45  25   0            0
Durham          11.60   5  45  45   5            0
```

and here's what I did with the data: can you do better?

```
x <- read.csv("Reduoa22.csv", header=T)
x[1,] # we have removed the Edinburgh/Heriot-Watt joint submission
attach(x)
score <- X4s + (3*X3s + X2s)/7 # This is the formula used for the score
N <- FTEstaff
first.lm <- lm(score ~ N) ; summary(first.lm)
plot(score~ N) ; abline(first.lm)
N <- N*N ; next.lm <- lm(score ~ N + NN) ; summary(next.lm)
N <- FTEstaff ; NN <- N*N ; next.lm <- lm(score ~ N + NN) ; summary(next.lm)
plot(score~ N) ; abline(first.lm)
install.packages("segmented")
set.seed(12)
library(segmented)
first.glm <- glm(score ~ N)
o <- segmented(first.glm, seg.Z = ~N, psi =list(N=17))
summary(o)
slope(o) # for N above 17, the graph is FLAT
```

So here is the key part of the corresponding output.

```
        ***Regression Model with Segmented Relationship(s)***

Call:
segmented.glm(obj = first.glm, seg.Z = ~N, psi = list(N = 16))

Estimated Break-Point(s):
   Est. St.Err
17.340  5.672
```

This is our estimate of N_{crit}.
A little more of the output gives us the estimates of b_1, b_2.

```
> slope(o) # for N above 17, the graph is FLAT

            Est. St.Err.  t value CI(95%).l CI(95%).u
slope1  1.91300  0.4023  4.75600     1.085     2.742
slope2 -0.08029  1.5000 -0.05354    -3.169     3.008
```

Now here are the first few lines of the corresponding dataset for Pure Mathematics, for which I think Kenna and Berche come to quite the wrong conclusion.

	FTEstaff	X4s	X3s	X2s	X1s	unclassified
Bath	10.00	25	35	40	0	0
Birmingham	18.00	15	40	35	5	5
Bristol	34.53	30	40	25	5	0
Cambridge	55.00	30	45	25	0	0
Durham	15.00	20	40	35	0	5

The corresponding analysis for the Pure Maths figures suggests that a 'broken line' model is NOT a good fit, and Vito Muggeo confirms this.
Compare the two plots given in Figure 1. Kenna and Birche find $N_{crit} \leq 4$. (They make quite a big deal of the 'fact' that the optimum size for a Pure Maths department is ≤ 4.) But their estimate is unreliable since the 'broken line' model is such a poor fit, and so maximising the log-likelihood is bound to be fraught with difficulties. Try it for yourself: you will find that you'll get quite different estimates according to where you take the 'starting values' of the iterative estimation process. A tell-tale sign is that the se of your estimate of N_{crit} will be very large. Figure 1 shows just one of the possible fits for the Pure Mathematics data.
Incidentally, the 'star' performer in Pure Mathematics is Imperial College London, with data

	FTEstaff	X4s	X3s	X2s	X1s	unclassified
ImpCollLondon	21.80	40	45	15	0	0

Acknowledgement Correspondence with Vito Muggeo of the University of Palermo was very helpful.
The two datasets were downloaded from `http://www.rae.ac.uk/results/outstore/uoa22.xls` for Statistics and OR, `http://www.rae.ac.uk/results/outstore/uoa20.xls` for Pure Mathematics. (I converted these .xls files into .csv files in order to read them into R.)

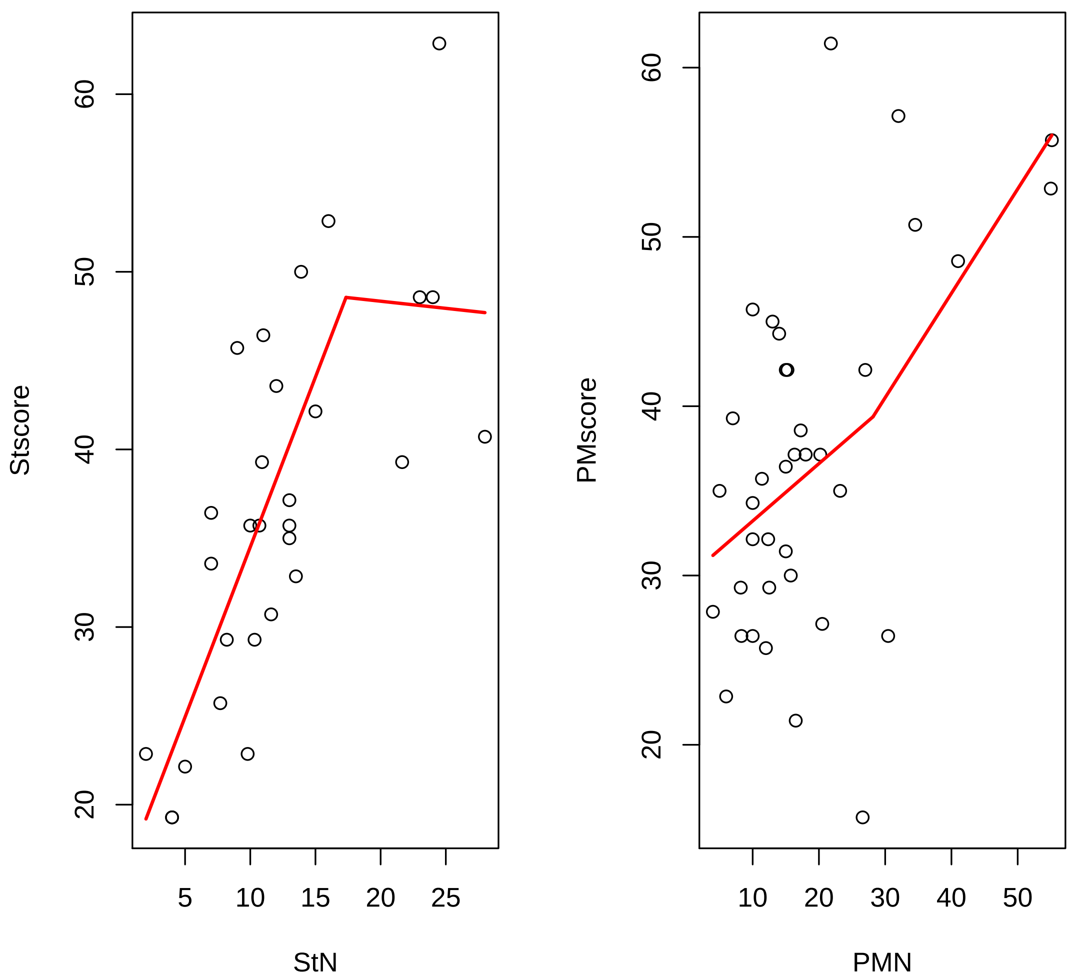

Figure 1: The 2008 RAE scores against Department size for Statistics and OR (left) and Pure Mathematics (right)