

Introduction to Quantitative Methods
Homework 4
Due: 17 February 2010

Note: This exercise must be done in R.

For this exercise we will use data on the politics departments in Britain from the 2008 Research Assessment Exercise (see <http://www.rae.ac.uk/> for details). For ease of access, I have saved a CSV-formatted data file with the politics results that you can access using this command (assuming you are connected to the Internet):

```
read.csv(url("http://www.kenbenoit.net/courses/iqrm/rae2008politics.csv"))
```

This “staff” variable provides the number of staff presented for the exercise, while the “zero”, ..., “four” variables are the percentages of staff that met the none through four-star ratings respectively.

1. Discuss the variability of the staff variable, using the:
 - a. Range
 - b. Inter-quartile range
 - c. Difference between 90th and 10th percentiles
 - d. Standard deviation

2. Plot the distribution of the staff variable using
 - a. Histogram
 - b. Density

3. Create a new variable to summarize the overall quality of departments, based on an index of their proportions of staff in each category. This can be created by multiplying zero by 0, one by 1, etc. up to four by 4, summing, and dividing the result by 4. (This will provide a measure from 0-100 where a 100 would mean that 100% of staff in the exercise were of four-star quality.)
 - a. Plot the density of the new overall score
 - b. Assuming the data was normally distributed, what are the two values between which approximately 68% of the data would fall?
 - c. Without assuming anything about the shape of the data, what are the values between which 68% of the data actually does fall? (Hint: This would be the middle 68% of the distribution, or between the 50+/(68/2) percentiles.) How close are these values to those predicted by the normal curve from b)?
 - d. *Extra credit:* using R, list the top five departments on the overall score. (Hint: This involves the `order()` command on the RAE data as a data frame.)