

# 1 Authorship of Federalist Papers

## 1.1 problem description

The Federalist Papers are a series of 85 pamphlets written to promote the adoption of the US Constitution during 1787-1788. Their detailed arguments are still cited today for their insight on the intended meaning of the Constitution. Written under the pseudonym 'Publius,' authorship has been attributed to Alexander Hamilton, James Madison, and John Jay. The consensus among scholars is that Jay wrote five: 2-5 and 64, Madison wrote fourteen: 10, 14, 37-48, and Hamilton wrote most of the remainder (fifty-one). Three are thought to be joint (18-20), and the twelve (49-58, 62, 63) comprise the list whose disputed authorship is the source of interest. In 1964, Mosteller and Wallace published a landmark statistical analysis of word frequencies to attribute authorship to the disputed twelve.

The original text is available on the Web in two forms:

1. as a single large (~1 Mb) ascii file at Project Gutenberg ([www.gutenberg.org](http://www.gutenberg.org))
2. as a series of html pages at two sites: Library of Congress ([thomas.loc.gov/home/histdox/fedpapers.htm](http://thomas.loc.gov/home/histdox/fedpapers.htm)) and Founding Fathers ([www.foundingfathers.info/federalistpapers/fedi.htm](http://www.foundingfathers.info/federalistpapers/fedi.htm))

The analysis can be divided into five steps:

1. Determine the beginning and end of each paper
2. Strip out punctuation, footnotes, and in the case of the websites, HTML code
3. Compute word counts for key words (Mosteller & Wallace used a list of thirty words)
4. Put word frequencies together across papers, forming an array of 30 frequencies across 85 papers
5. Perform discriminant analysis or some other statistical analysis to determine authorship

The target audience is advanced undergraduates or masters students.

## 1.2 learning objectives

- parsing
- breaking problem into subtasks
- handling progressively more complicated tasks
- compiling results from subtasks

### **1.3 computational tools**

In general: character manipulation, stripping unnecessary data, computing frequencies, merging frequency counts

In SAS:

1. character functions, macros, merge by
2. procs freq, sort, transpose, append, discrim

In R: (I can't do it in R.)

John Monahan, 18 July 2009