4.5: Red Herrings in Brain Imaging

Neuroscientists do massive numbers of comparisons regularly. They often perform fMRI studies, where a three-dimensional image of the brain is taken before and after the subject performs some task. The images show blood flow in the brain, revealing which parts of the brain are most active when a person performs different tasks.

But how do you decide which regions of the brain are active during the task? A simple method is to divide the brain image into small cubes called voxels. A voxel in the "before" image is compared to the voxel in the "after" image, and if the difference in blood flow is significant, you conclude that part of the brain was involved in the task. Trouble is, there are thousands of voxels to compare and many opportunities for false positives.

One study, for instance, tested the effects of an "open-ended mentalizing task" on participants. Subjects were shown "a series of photographs depicting human individuals in social situations with a specified emotional valence," and asked to "determine what emotion the individual in the photo must have been experiencing." You can imagine how various emotional and logical centers of the brain would light up during this test.

The data was analyzed, and certain brain regions found to change activity during the task. Comparison of images made before and after the mentalizing task showed a $p = 0.001$ difference in a 81mm^3 cluster in the brain.

The study participants? Not college undergraduates paid $\$10$ for their time, as is usual. No, the test subject was one 3.8-pound Atlantic salmon, which "was not alive at the time of scanning."[8]

Of course, most neuroscience studies are more sophisticated than this; there are methods of looking for clusters of voxels which all change together, along with techniques for controlling the rate of false positives even when thousands of statistical tests are made. These methods are now widespread in the neuroscience literature, and few papers make such simple errors as I described. Unfortunately, almost every paper tackles the problem differently; a review of 241 fMRI studies found that they performed 223 unique analysis strategies, which, as we will discuss later, gives the researchers great flexibility to achieve statistically significant results.[13]