

# Statistical Analysis of MRI Data

Shelby Cummings

August 1, 2012

## Abstract

Every day, numerous people around the country go under medical testing with the use of MRI technology. Developed in the late twentieth century, this noninvasive procedure can be used to not only diagnose medical problems, but also to learn about the structure and function of the human body. This remarkable piece of machinery can collect and store a multitude of data in a very short time period, but what do we, as researchers, do with this data? One way of using this data is to determine the locations of voxels responsible for different motor functions. By looking how the activation level in a voxel changes over time from periods of rest to activity, we can determine whether or not the specific voxel is involved in the given motor function. By using linear regression and likelihood ratios, we can test to see whether or not this change in activation level over time is statistically significant, which then allows us to determine whether or not the voxel plays an important role in the motor function. Learning which voxels are involved with different motor functions, and eventually other human functions, can allow us to map out the brain and give us a better picture of how the human brain works.

## 1 Introduction

The creation of MRI technology has advanced our medical knowledge substantially in recent years. What began as physicists simply looking at how protons spin and interact with magnetic field quickly turned into them finding ways to quantify the effects the magnetic fields had on the protons. Eventually, physicists discovered that by exposing a proton to a magnetic

field, you get a signal from the proton. This signal can be split into frequency components which can be used to gain information about the location of the proton. From here, MRI technology was born. MRIs are a very useful procedure in the medical world because they are very non-invasive and do not put the patient in an extreme amount of pain or discomfort. In a very short period of time, a lot can be learned from this procedure. MRIs can be used to not only diagnose medical conditions, but also to learn more about the structure and function of the human body. An anatomical MRI can be used to diagnose medical conditions that are a product of abnormal structure. In the case of brain abnormalities, an MRI scan showing the anatomy of a patients brain could be used as a helpful diagnostic tool. By looking at a particular patients brain anatomy and comparing it to the structure of a normal human brain, a doctor can then determine what is structurally wrong with the patients brain. All of this can be done without any invasive, exploratory surgery, which significantly lowers the patients pain, discomfort, and risk of complications.

The other way in which MRI scans can be used is to determine the function of different human organs. Doctors know that everything we, as humans, do is somehow controlled by the brain. Decision making, seeing, hearing, walking, even something as simple as snapping your fingers is controlled by specific portions of the brain. MRIs allow researchers to determine which portion of the brain is responsible for, or active during, a certain brain activity. This can be helpful for many different reasons. For example, if a person is having difficulties with their speech, knowing the part of the brain responsible for controlling speech may allow doctors to pinpoint where in the brain a problem may be occurring. If doctors are contemplating brain surgery to correct a problem such as brain abnormalities or epilepsy, knowing the structure of the brain, being about to have a map of where different activities are controlled, can guide doctors in deciding whether or not the surgery is worth it, or even an option. If a brain operation involves removing a part of the brain, and doctors determine that that part is involved with speech, they may determine that the surgery is not worth the risk. When an accident victim comes in the ER, with head trauma, doctors may be able to predict the patients outcome simply by knowing what part of the brain was affected and what that part of the brain is responsible for.

For these reasons, and many others, it is very important for us to be able to locate what part of the brain is responsible for different functions. In order to figure this out, we must compare how the brain is working both during

periods of activity and during periods of rest. If an area of the brain becomes more active during the period of activity than during the period of rest, we are brought to believe that the area is involved in whatever activity is being performed. In order to test this, we use linear regression and likelihood ratio tests to determine whether or not the difference we are seeing is statistically significant: is it a real difference caused by the brain actually being involved in the activity, or is it simply caused by noise. Once we have determined which areas of the brain are active during a given motor function or activity, we can then create maps of the brain, showing where all the active areas are located. By piecing all of this information together, doctors and researchers can begin to get an idea of how the human brain works.

## 2 Data Collection

In order to begin to find out where different activities are controlled in our brain, we must first begin by looking at data. In this case, data comes in the form of brain scans. For each test subject, we must get a reading of every area in their brain both during periods of rest and periods of activity. In order to do this, we split the brain up into voxels, or volumetric pixels. Each one of these is a small area of the brain. When getting a reading, we get an activation level for each brain voxel. We can then create a matrix for that reading, with each entry in the matrix representing the activation level in the corresponding voxel. This matrix is then recreated with new activation levels for each reading taken on a subject. In most cases, readings are taken every one to two seconds for one hundred and twenty eight seconds. The subject is first instructed to lie at rest for a given number,  $n$ , seconds. Then the subject is instructed to perform the given motor function for  $n$  seconds. This process is repeated a few times before the subject is finished. Researchers then have a series of matrices showing how the entire brain is changing and working through periods of rest and activity. Figure 1 shows how we look at a brain image over time and its associated matrices to see how activation levels change.

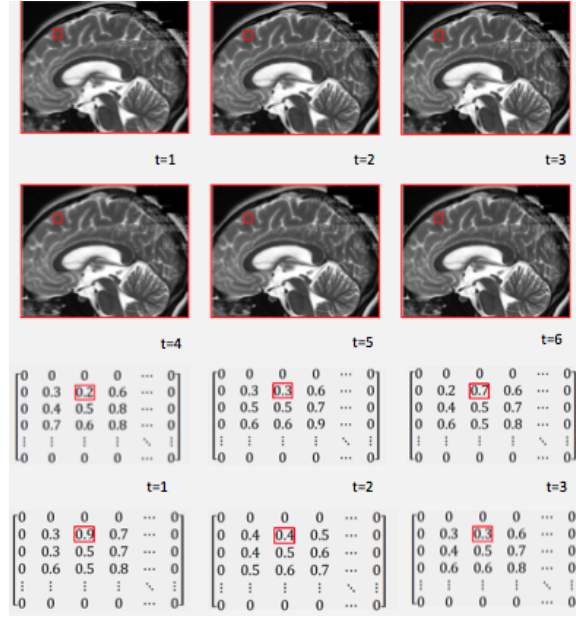


Figure 1: Brain Image Produced from MRI Scan

### 3 Data Manipulation

Now researchers need to manipulate and analyze the data they have collected. The first thing that needs to be done is the data needs to be separated by voxel. Since the end goal is to determine in which areas of the brain activity is going on, we must look at each individual area and see if a change exists in that area. Through careful indexing of these matrices, we can follow one voxel through the entire experiment, watching how its degree of activation changes over time. We then create a new matrix which consists of the activation levels for a single voxel over the span of the entire experiment. The code for doing so looks like such:

```
function SplitIntoVoxels(m1,m2,m3,m4,m5,m6,m7,m8,m9,m10)
s = size(m1);
m = s(1);
n = s(2);
for i = 1:m
    for j = 1:n
        voxel = zeros(10,1);
```

```

        voxel(1,1) = m1(i,j);
        voxel(2,1) = m2(i,j);
        voxel(3,1) = m3(i,j);
        voxel(4,1) = m4(i,j);
        voxel(5,1) = m5(i,j);
        voxel(6,1) = m6(i,j);
        voxel(7,1) = m7(i,j);
        voxel(8,1) = m8(i,j);
        voxel(9,1) = m9(i,j);
        voxel(10,1) = m10(i,j);
        strP = ['Voxel ', num2str(i), ', ', num2str(j)];
        disp(strP)
        disp(voxel)
    end
end

```

After we have created a matrix with the activation levels, we can look at how that specific voxel is changing over time. If the voxel is not responsible for or active during the motor function, we would expect there to be no change in activation energy over the course of the experiment, regardless of whether or not the subject is performing the given motor function or not. We would expect a graph of activation level over time to look as such:

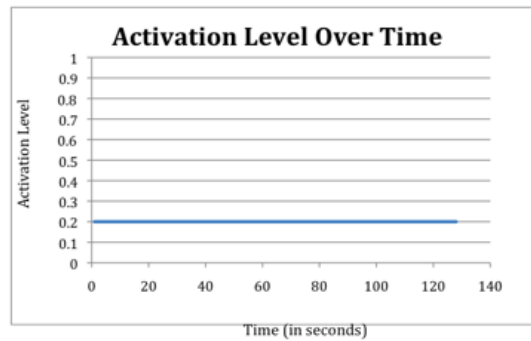


Figure 2: Graph of Activation Level Over Time With No Voxel Involvement: Ideal Data

In this graph, the activation level is consistent over time, and does not change at any time during the experiment, regardless of whether or not the subject is performing the given motor function. In reality, these graphs do

not look exactly like this, due to things such as experimental error, noise, etc. They tend to end up looking more like the graph displayed in Figure 3.

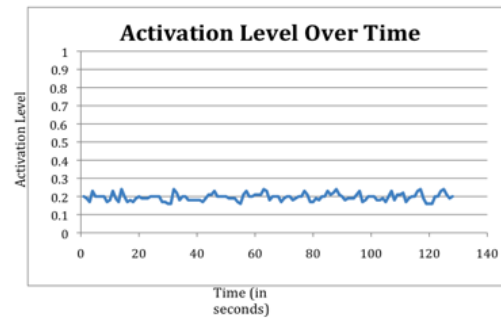
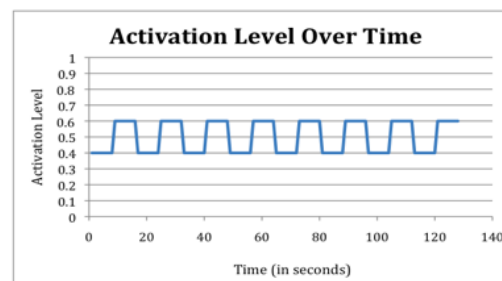


Figure 3: Graph of Activation Level Over Time With No Voxel Involvement: Real World Data

While these graphs do show some variation in the activation levels over time, it can also be noted that they are fairly consistent over time. They look nothing like the graphs we would expect of a voxel that is involved in the given motor function to look like. For these voxels, we would expect the activation levels to spike up during periods of activity, and then lower during periods of rest, like this:



Ideal Data

Figure 4: Graph of Activation Level Over Time With Voxel Involvement: Ideal Data

This graph shows the up and down pattern we would expect with activation levels given that we are alternating between periods of rest and periods

of activity. Of course, once we account for experimental error and noise, the graphs end up looking more like this:

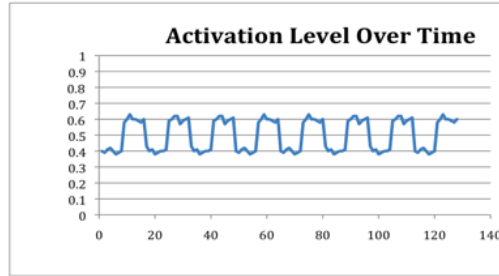


Figure 5: Graph of Activation Level Over Time With Voxel Involvement: Real World Data

With this graph, we can still see the apparent up and down pattern in the data, even though there is some definite noise. But what would happen if we were presented with a graph like displayed in Figure 6.

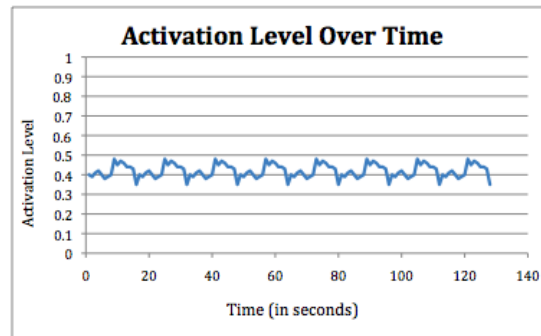


Figure 6: Graph of Activation Level Over Time Within a Voxel

When we are presented with data from a voxel that looks like this, how are we to tell whether the differences in activation level we are see is due to the voxel actually being involved with the motor function, or simple due to noise? In order to determine this, we must turn to statistical tests to figure out what is truly going on with our data set.

## 4 Statistical Analysis

In order to determine whether or not a specific voxel is involved in a given motor function, we must use statistical analysis. We first start by finding both the mean and the standard deviation of each voxels activation level over time. We can do so using the following code:

```
function MeanAndStdDev (m1,m2,m3,m4,m5,m6,m7,m8,m9,m10)
s = size(m1);
m = s(1);
n = s(2);
for i = 1:m
    for j = 1:n
        mean1 = (m1(i,j) + m2(i,j) + m3(i,j) + m4(i,j)
+ m5(i,j) + m6(i,j) + m7(i,j) + m8(i,j) +
m9(i,j) + m10(i,j))/10;
        stdDev1 = sqrt((((m1(i,j)-mean1)^2) +
((m2(i,j)-mean1)^2) + ((m3(i,j)-mean1)^2)
+ ((m4(i,j)-mean1)^2) + ((m5(i,j)-mean1)^2)
+ ((m6(i,j)-mean1)^2) + ((m7(i,j)-mean1)^2)
+ ((m8(i,j)-mean1)^2) + ((m9(i,j)-mean1)^2)
+ ((m10(i,j)-mean1)^2))/10);
        str1 = ['Position ', num2str(i),',', num2str(j)];
        str2 = ['Mean = ', num2str(mean1)];
        str3 = ['Standard Deviation= ', num2str(stdDev1)];
        disp('')
        disp(str1)
        disp(str2)
        disp(str3)
    end
end
```

After we have found the mean and standard deviation of each voxels activation level over time, we can begin to use linear regression to see if changes are occurring within the voxel. Linear Regression is a way of fitting a line, sometimes referred to as the line of best fit to a series of data points. For voxels not involved in the given motor function, we would expect a line fit to the data points to be a horizontal line with a slope of zero and a y-intercept of the mean activation level of the voxel. For voxels involved in the



given motor function, we would not expect this to be true. For these voxels, we would expect the line to have some sort of non-zero slope. So, in order to determine whether or not the voxel is significantly involved in the given motor function, we must see if the slope coefficient of the line fit to the data is non-zero, and if it is so, is it statistically significant.

We first begin by finding the slope, or  $\beta$ , of the regression line for each voxel. For each voxel, we need to create two column matrices, each with as many rows as samples taken on the subject. The first matrix, denoted  $X$ , represents time, while the second matrix,  $Y$ , represents the activation level of the brain. We can then use the following formula to estimate  $\beta$ :

$$\hat{\beta} = (X'X)^{-1}X'Y$$

This provides us with a good estimator of  $\beta$  because the expected value of  $\hat{\beta}$  is  $\beta$ , which makes  $\hat{\beta}$  an unbiased estimator of  $\beta$ . The following code shows how to find the  $\hat{\beta}$  estimator for every voxel.

```
function RunLinReg(m1,m2,m3,m4,m5,m6,m7,m8,m9,m10)
s = size(m1);
m = s(1);
n = s(2);
t = [1;2;3;4;5;6;7;8;9;10];
tt = t';
ttt = tt * t;
for i = 1:m
    for j = 1:n
        voxel = zeros(10,1);
        voxel(1,1) = m1(i,j);
        voxel(2,1) = m2(i,j);
        voxel(3,1) = m3(i,j);
        voxel(4,1) = m4(i,j);
        voxel(5,1) = m5(i,j);
        voxel(6,1) = m6(i,j);
        voxel(7,1) = m7(i,j);
        voxel(8,1) = m8(i,j);
        voxel(9,1) = m9(i,j);
        voxel(10,1) = m10(i,j);
        BetaHat = (ttt)\(voxel);
        strP = ['Voxel Number ', num2str(i), ', ', num2str(j)];
```

```

        strP2 = [ 'Beta    Hat = ', num2str(BetaHat(1)) ];
        disp(strP)
        disp(strP2)
    end
end

```

Any voxel that has a  $\hat{\beta}$  value other than zero should be tested for significance. If these voxels have a  $\hat{\beta}$  that is significantly non-zero, then it is involved in the given motor function. To determine whether or not a particular  $\hat{\beta}$  is significant, we use likelihood ratio tests. Likelihood ratio tests are a way of generating test statistics of common distributions. In order to do so, we must first start with the first step of all hypothesis tests, creating our null and alternative hypotheses. In this case, our null and alternative hypotheses are:

$$H_0 : \beta = 0, \sigma^2 > 0$$

$$H_1 : \beta \neq 0, \sigma^2 > 0$$

We then begin by writing out the likelihood function, which looks as such:

$$L(\beta, \sigma^2) = (2\pi\sigma^2)^{\frac{n}{2}} e^{(-0.5\sigma^{-2})(y-X\beta)'(y-X\beta)}$$

We then take the log of this function, to create the log likelihood function, as this is an easier function to work with. We then take the partial derivatives of this function with respect to both  $\beta$  and  $\sigma^2$  under the conditions in both the null and the alternative hypothesis. After that, we take the ratio of the likelihood function assuming the null and the alternative hypotheses. We can use algebra to transform this into a variable which follows the t-distribution.

$$t = (C\hat{\beta} - \gamma)\sqrt{F}$$

We now have our t-statistic and can determine if our value for beta is statistically significant. The first step in doing so is to choose an alpha level. The most common alpha levels are 0.05, 0.01, and 0.001. For the purposes of this analysis, the alpha level of 0.05 was chosen. What this means is that we are taking a five percent chance of calling a slope significant when it is really just an effect of noise. What happens is, when we find a mean value of activation level for a voxel, this mean is based on a sample of values. We do not have data for the activation level at every single time, so we must do the next best thing and use a sample. Since the sample mean is an unbiased

estimator of the population mean, the expected value of the sample mean is the population mean. This is not to say that every sample mean we get is the same as the unknown population mean. The sample means come from a t-distribution centered around the unknown population mean. If we took a large number of samples, the majority of the means would fall around the population mean, and the further away you move from the population mean, the less sample means you would find. This works the same way with  $\beta$  and  $\hat{\beta}$  values. Since  $\hat{\beta}$  is simply an estimator of  $\beta$ , we do not always expect  $\hat{\beta}$  to equal  $\beta$ . If we operate under the assumption that the null hypothesis is true, then we would expect  $\hat{\beta}$  to be zero. Since not all values of  $\hat{\beta}$  will equal  $\beta$  though, some variation is expected. What happens is we need to determine whether or not the value of  $\hat{\beta}$  is significantly different than zero, that is, it is large enough for us to believe that it is not simple a varied form of  $\beta$  being zero, that it actually comes from a distribution of  $\beta$  that is not centered around zero. If this is true, then we are able to reject the null hypothesis in favor of the alternative hypothesis, and we believe that  $\beta$  does not equal zero, and the voxel is involved in the given motor function. We go through this process for each and every voxel in order to determine whether or not it is significantly involved in the motor function.

Once we have determined which voxels are significantly involved in a given motor function, we can begin to draw maps of the brain based on what voxels are working. The way this is done is by overlaying an anatomical picture of the brain with a color map of the voxels which are active during the given motor function. Those without a statistically significant  $\beta$  value are not colored with the color map. Those with statistically significant  $\beta$  values are assigned a color based upon the level of significance. Once those two maps are overlaid, you can see both the anatomical structure of the brain, as well as how the brain is working during the given motor function, as in Figure 7.

All of these steps can be used for any given motor function that can be completed while undergoing MRI scans. This is how we get pictures and maps of how the brain works.

## 5 Future Work

While the research completed thus far is a substantial step towards understanding how to analyze MRI data, there is still much that could be done with this research. First, it would be necessary to consolidate and revise the



Figure 7: Color Map of Active Areas in the Brain During Motor Function

MATLAB coding so that the coding was more versatile and could be used by others in this field. Another interesting thing to look at with this data would be to fit other models to the data. One last thing that should be considered is the correlation between different voxels. There is definitely spatial correlation involved in this data, so future work would involve looking into that to see how it affects the analyses and finding a better way to deal with this kind of data.

## 6 Summary

The purpose of this research was to discover ways to analyze the data output of MRI scans. MRI scans provide doctors and researchers with matrices full of data. Each data point represents a voxel, or one tiny area of the brain. This sheer amount of data can be overwhelming, but the process one goes about to analyze is really quite straightforward. By looking at how these activation levels change between periods of rest and periods of activity for a given motor function, it is possible to determine whether or not the voxel is involved in the given motor function. By running linear regression and using likelihood ratio functions, we can see if any changes in activation levels are statistically significant, and can then use this information to create maps of the brain. We can then use this information to figure out what is going on in different parts of the brain.

## 7 References

Haacke, E. Mark., Robert W. Brown, Michael R. Thompson, and Ramesh Venkatesan. Magnetic Resonance Imaging: Physical Principles and Sequence Design. New York: Wiley, 1999. Print.

Hanselman, Duane C., and Bruce Littlefield. The Student Edition of MATLAB: Version 5 : User's Guide. Upper Saddle River, NJ: Prentice-Hall, 1997. Print.

MATLAB: The Language of Technical Computing. Print.

Poldrack, Russell A. "Region of Interest Analysis for FMRI." Social Cognitive and Affective Neuroscience 2.1 (2007): 67-70. Web.

Rencher, Alvin C. Linear Models in Statistics. New York: Wiley, 2000. Print.

Rowe, Daniel B., and Brent R. Logan. "A Complex Way to Compute FMRI Activation." NeuroImage 23 (2004): 1078-092. Print.

Smith, S. M. "Overview of FMRI Analysis." British Journal of Radiology 77 (2004): 167-75. Web.