# Data

59 subjects

for each subject, successive experiments stimulating in one or more locations (Stim\_loc)

for each stimulation observation:

* pre-stimulation biomarker value
* stimulation parameters (frequency, amplitude, duration)
* post-stimulation change in biomarker

# Hypothesis

There is a relationship between (change in biomarker) and (pre-stim biomarker + stim parameters). For simplicity, evaluate whether there is a linear relationship.

# Methods

Linear least-squares regression

X = [pre-stim biomarker, frequency, duration, amplitude]

Y = [post-stim change in biomarker]

No leave-one-out cross-validation (LOOCV) in this analysis. It seems like the standard regression work flow is to just fit and examine hypothesis. If we were concerned about overfitting, then we would use cross-validation, but since the performance is already low there is little worry. Any cross-validation would lower the prediction. So this first stage is just to see how good prediction is in best case.

I’ve been reading through these excellent tutorials on linear regression: <http://www.biostathandbook.com/linearregression.html>

<http://blog.minitab.com/blog/adventures-in-statistics/how-to-predict-with-minitab-using-bmi-to-predict-the-body-fat-percentage-part-1>

Turns out there is a “predicted r^2” that is similar to LOOCV where each sample is left out and the model used to predict. It is also referred to as the PRESS statistic. I haven’t gone through calculating this because I don’t expect over-fitting with linear, but good to know there is solid grounding in the approach. <https://en.wikipedia.org/wiki/PRESS_statistic>

Look at p-value to tell us “is there a significant relationship?”

Look at r^2 telling us how much of output variation is captured in model (how good at predicting).

For each subject, for each location of subject, regress. Only include stimulation parameters in model that changed for the experiment, ie. if amplitude was held constant for all of experiment, leave that out of model (if all the same, then standard error would be zero anyway).

For each location, combine results from all subjects tested at that location and regress.

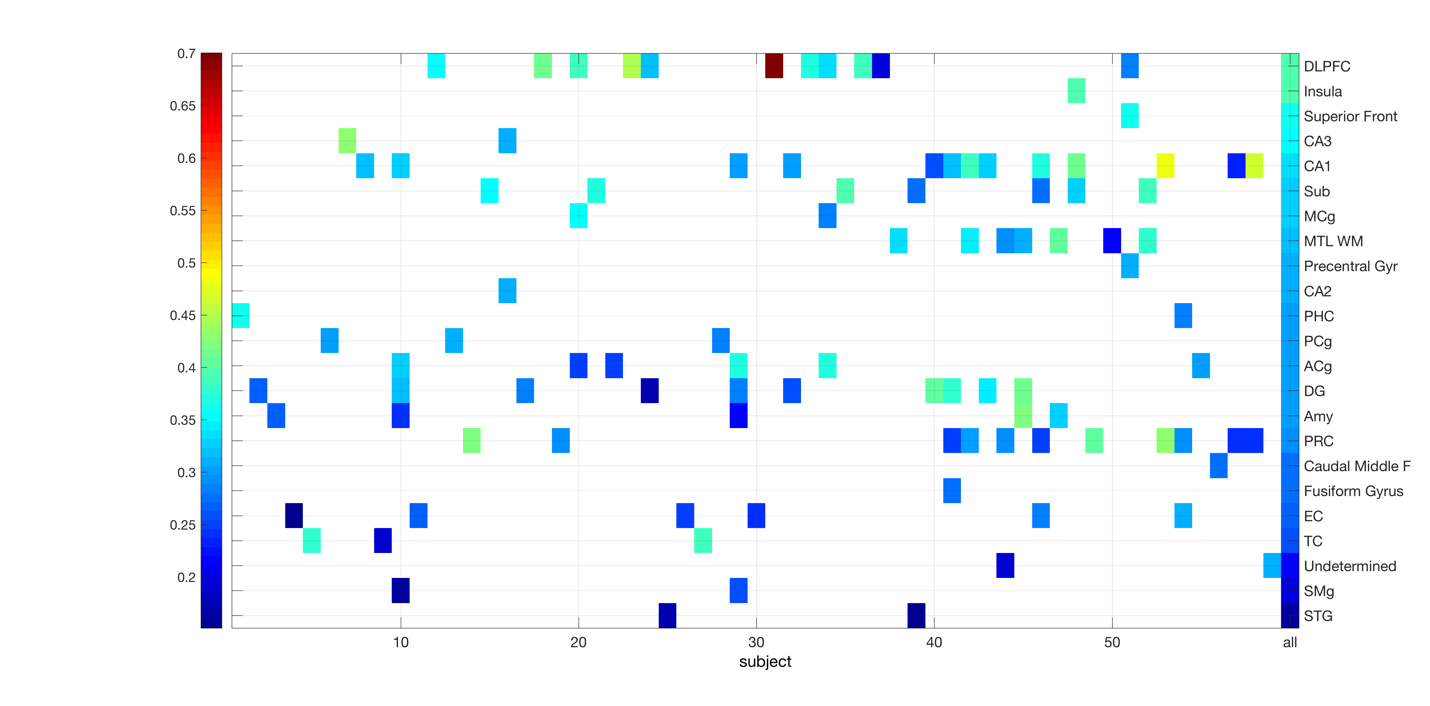
To explore the possibility of nonlinear relationships, instead of using linear, also re-run with quadratic and interaction terms.

# Results

Columns: subjects. Last column is across all subjects.

Rows: locations

Color squares indicate where an experiment was performed. White indicates no experiment.



The rows are ordered by r^2:

* DLPFC and Insula had highest prediction, but that was only about r^2=0.40.
* Superior Front and CA3 were next highest.
* Then CA1, Sub, MCg.

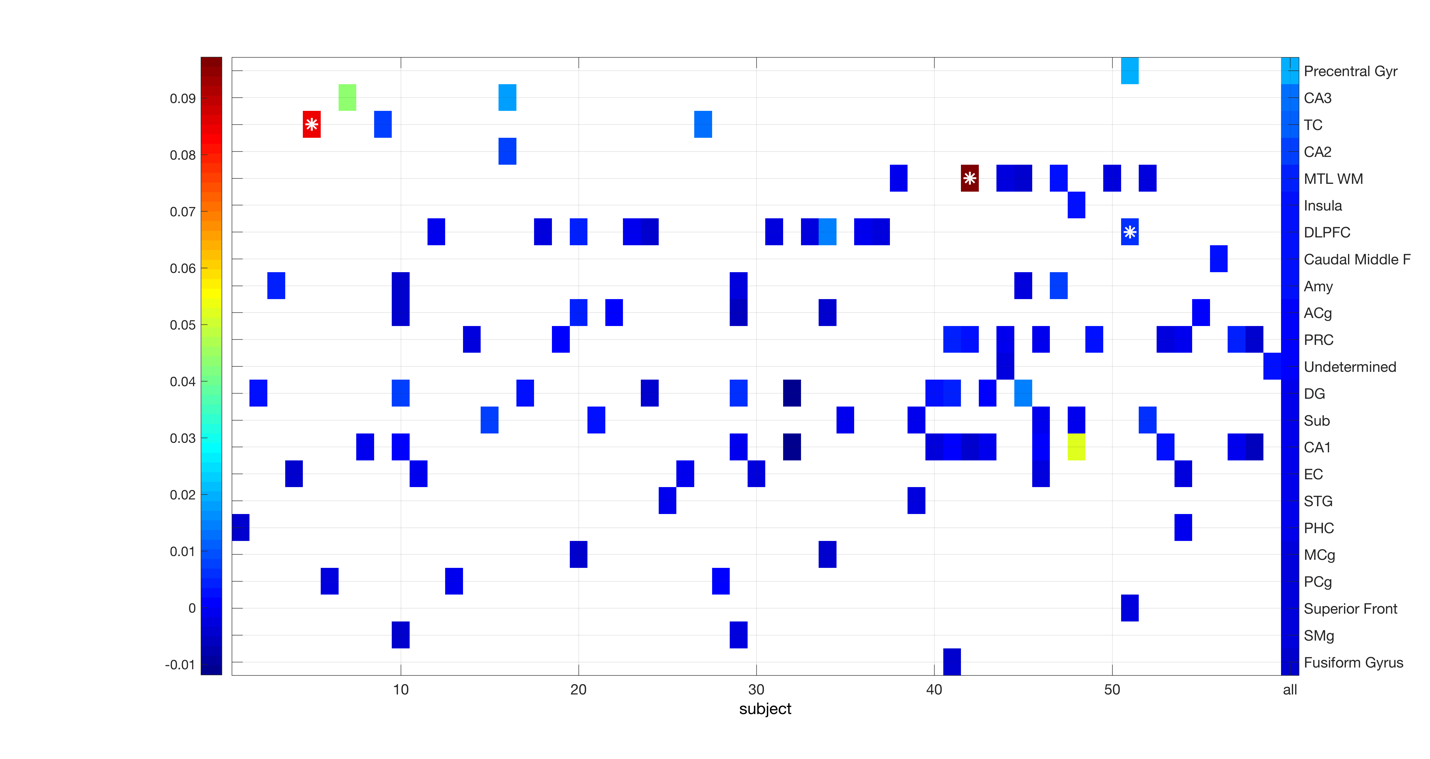
One notable outlier, Subject #31 had one experiment stimulating only DLPFC and achieved r^2=0.7

All locations had a very significant linear relationship (p-value), even when conservatively correcting for multiple tests across all subjects and locations.

This is not surprising. Recall from earlier experiments that we were finding rho=-0.66 between pre-stim biomarker and post-stim change, indicating that there is a very negative correlation: any stimulation seems to decrease the biomarker. It’s also not surprising because the dependent variable (change in biomarker) is related to the independent variable (pre-stim biomarker).

## Stimulation parameters alone have no significant effect

For all regressions, the biggest factor (coefficient) was for the pre-stim biomarker. The stimulation parameters contributed very little in comparison. As a check, I performed this same regression omitting the pre-stim biomarker, ie. regressing post-stim change against only stimulation parameters. This showed terrible results: r^2 was between 0-9% (mean 0.2%, median 0) with no significant p-values for any location (right-most column below). In other words, stimulation parameters have no significant effect on the change in biomarker.



## Nonlinear models make no difference

To move beyond linear, results of trying quadratic and interaction equations. This includes the linear terms as well as squares (x1\*x1) and interactions (x1\*x2). Looking at the ‘all’ column, this resulted in no real change in the r^2. Below are the descriptive stats on the last column using both types of models: linear, quadratic/interaction.

**R^2 using linear model (all the experiments above):**

mean 0.299643

std 0.0545365

median 0.304933

min 0.168033

max 0.39578

**R^2 using quadratic and interaction terms:**

mean 0.308794

std 0.0576119

median 0.307222

min 0.183249

max 0.42355

# Conclusions

Every experiment indicates a significant relationship between (pre-stim biomarker and stimulation parameters) and (post-stim biomarker change) (p-values <0.0001), although knowing the pre-stim biomarker or stimulation parameters tells you little about the post-stim change overall but gives modest information in DLPFC and Insula (r^2).

The actual value of the stimulation parameters affects has no relationship to the change. No matter what stimulation parameters are tried, the biomarker always goes down. But knowing the stimulation parameters, you cannot predict what that change will be.

# Extra readings

Biostats Handbook: <http://www.biostathandbook.com/linearregression.html>

Minitab regression tutorial: <http://blog.minitab.com/blog/adventures-in-statistics/regression-analysis-tutorial-and-examples> and <http://blog.minitab.com/blog/adventures-in-statistics/how-to-predict-with-minitab-using-bmi-to-predict-the-body-fat-percentage-part-1>

Evaluating r^2, adjusted r^2, and predicted r^2: <http://blog.minitab.com/blog/adventures-in-statistics/multiple-regession-analysis-use-adjusted-r-squared-and-predicted-r-squared-to-include-the-correct-number-of-variables>

Nonlinear regression: <http://blog.minitab.com/blog/adventures-in-statistics/linear-or-nonlinear-regression-that-is-the-question>

Residual analysis: <http://blog.minitab.com/blog/adventures-in-statistics/why-you-need-to-check-your-residual-plots-for-regression-analysis>

MATLAB linear regression workflow: <http://www.mathworks.com/help/stats/linear-regression-model-workflow.html>