

Dear Penny,

I received an email that you are having problems with understanding the multiple testing or penalized regression. I hope my explanations can help you understand those concepts.

Multiple testing is a statistical method used when multiple hypothesis tests are involved. Statisticians do multiple testing to avoid incorrect null hypothesis rejections (this is also called a Type 1 error). For example, assume we do a hypothesis test with a 5% confidence level when the null hypothesis is true. In other words, with the chance of 5%, we may reject the null hypothesis even if it is true. So, if we do the 10,000 hypothesis test like your fMRI data, we will reject about 500 null hypotheses even if all the null hypotheses are true, which is critical. So, to reduce the Type 1 error of the analysis, we should use multiple testing in this data. Usually, statisticians use Bonferroni, Fisher Combination, Simes Test, or Benjamini-Hochberg for multiple testing. If you want to know more about each test, please let me know.

Also, we cannot use the standard regression here. Unlike the usual datasets, fMRI data has more predictors than its sample size. In this case, if we do the standard regression, the data must be overfitted. This overfitted standard regression will have zero bias and high variance, which is a statistically unfavorable situation. To reduce the variance, or make our regression smoother, we do the penalized regression. The word penalized in the penalized regression means we will penalize the using of predictors in the model. To penalize the regression, unlike the standard regression, we often minimize the

$$E \left[\|y - X\beta\|^2 \right] + \lambda \|\beta\|, \text{ or } E \left[\|y - X\beta\|^2 \right] + \lambda \|\beta\|_2^2$$

The first penalization is ridge regression, and the second penalization is lasso regression. As λ increases, the new term $\lambda \|\beta\|$ and $\lambda \|\beta\|_2^2$ would also increase, so we would choose smaller β s for our model. As a result, we would have smaller β s, which would lead our model to be smoother.

In conclusion, multiple testing is involved the hypothesis tests. If we do multiple hypothesis test, our Type 1 error would increase, so we need special analyses called multiple testing. Penalized regression is involved with a variance of the regression. If our model is overfitted, we want to reduce the variance of our regression, so we add penalized terms to penalize the excessive use of predictors.

I hope this email can help you to understand the statistical concepts you want to know. If you have any more questions, please let me know.

Sincerely,

... ..