**Frequently Asked Questions:**

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

**General Workshop Questions:**

**Can you help me with my statistics questions after the workshop?**

While we are willing to answer a quick question by email, detailed requests about specific data sets can be difficult to answer, and so we may not have the time to answer these types of questions. We are, however, open to collaborations and would be able to dedicate more time to a specific project as collaborators.

**Where and how can I learn more about sparse modeling?**

Head over to our post-workshop resources page for some advice on where to go next.

**General statistics questions:**

**What is sparse modelling?**

Sparse modelling is a statistical technique used for fitting models when there is a large number of predictors, and we believe that only a few have strong effects. There are many different sparse modeling approaches, with both frequentist and Bayesian methods available. In this workshop, we present just two methods, but there are many more to learn about!

**What is an inference?**

An inference is a conclusion made based on logical reasoning. In statistics, inference refers to how selected predictor variables are related to a response variable in a model. Think about it like this- we use the sparse modeling process to infer that the predictor variables selected are important to the response variable.

**What is in-sample prediction?**

In-sample prediction refers to a prediction about the sampled population. If we use some data points to fit a model, and them make a prediction for those same data points to test how well our model fits the data, then we are making an in-sample prediction.

**What is out-of-sample prediction?**

Out-of-sample prediction refers to a prediction about data points outside the sampled population. If we use some data points to fit a model, and then use new data points to test our model, then we are making an out-of-sample prediction.

**What is regularization?**

Regularization is the addition of a penalty against large coefficient values of the predictors, in order to avoid overfitting. In sparse models, penalties can be added in different ways. Frequentist methods like lasso may regularize by adding a penalty term to the sum of squared residuals and then attempt to minimize that whole quantity. Bayesian methods may regularize by using certain prior distributions.

**Specific lasso and SuSiE questions:**

**Is there a way to have categorical variables in SuSiE or lasso?**

Not very well in the packages that we used in this workshop. In glmnet (lasso) you can make your categorical variables into "dummy variables" and get predictions, but the interpretation may be confusing. Check out the group lasso method (gglasso) if you work with a lot of categorical variables.

**Can I have random effects in my lasso and SuSiE models?**

This can get complicated- and wouldn't be suited to the packages we present here. If you want to get into it though, you can check out the glmmLasso package. In general, lasso is a more flexible method than SuSiE, and there are many alternative packages that implement lasso regression.

**What is the mechanism underlying lasso regression?**

Lasso regression uses lambda penalty to shrink the slope (coefficients) of each parameter in a model, forcing less important variables to have zero slope.

**Why does Lasso shrink coefficients to zero?**

This is the mechanism used in lasso to eliminate less important variables in a model to achieve better prediction and simplicity

**Is Lasso better than SuSiE?**

There are pluses and minuses to each method that are explained in more detail in the R-script. SuSiE is very fast even for huge datasets like those found in genomics, while lasso is more flexible to different peculiarities of data. Suffice to say, deciding which method to use will depend on your dataset.

**What is the Lasso penalty?**

The lasso penalty pushes the coefficients of variables to zero or towards zero so that less important variables have zero or near-zero coefficients.

**Is lasso L1 or L2?**

Lasso is an L1 regularization technique, while Ridge regression is an L2 regularization technique. The difference between them is how their penalty terms are computed.

**Is lasso supervised or unsupervised?**

Lasso is an L1 regularization method that is supervised

**Why is the SuSiE method so limited?**

The SuSiE method is a newer method from the field of genomics, and so it hasn’t yet developed as much as older methods like the lasso method. The algorithm itself can be a bit complicated, so it’s easiest to use it by using the functions in available in the packages. However, development is still in progress, and we should expect to see more flexible versions of SuSiE in the future.