**Model Part**

We all agree with that if we work as a statistician or a data scientist, model is an important part in daily work. Thus, we prepared several questions about model for the interview. The questions and corresponding answers are as follow.

1. Q: Which part do you think is the most complex or most time-consuming in your models? (Proposed by Xiang Li)

A: I would bring in three challenges that I once met.

The first one is random effects and modeling correlation. Data that comes from genetics and typically populations always automatically brings random effects in models and cooperating information from patent tests requires us to model covariation and correlation. The best solution to solve both of the two problems is to use mixed model.

The next challenge is different type of traits make kinds of responses in our models. Besides the common continuous variable, we also have dummy variables, counting variables and even category as a response. We must spend time to build different models for each kind of responses.

Finally, we have to be fast and meanwhile let the solution as good and robust as possible, which causes us hardly trade off between a simple model and a elaborate model.

1.1 Follow-up Q: Could you give us more examples in mixed models? (Proposed by Jinrui Du)

A: Because the data is actually connected by pedigree, so pedigree information is important in the moment of analyzing the data. That all comes quite naturally in mixed model framework.

2. Q: Usually, statistical models can be used to predict future data, classify target objects, find effective factors and etc. So, what is the main purpose of statistical model in your research? (Proposed by Xiang Li)

A: The main purpose is about prediction and prediction with small data. Predicting which material or hybrid can be a commercially potential product helps us narrow down the number of experiments to a small amount. Therefore, we don’t need to waste couple of years and a lot of money cost on those invalid experiments.

2.1 Follow-up Q: What do you mean by small data? (Proposed by Jinrui Du)

A: Small data means the number of experiments you can run is small. Because time is limited and we cannot spend 10 years on trailing material to find out the proper one. We have to quickly make our products come to the market.