

# Homework 1: Model Building and Model Selection/Fitting

Name: Yilun Xu

email: [yilunx@uchicago.edu](mailto:yilunx@uchicago.edu)

## Building Models

### Deviant aggressive behavior

Theory I shows that people's behavior will be influenced by other people's behavior and its result. According to this theory, I believe that society should establish and improve reward and punishment systems in various fields to lead social trends. For example, the government can promote the construction of credit reward and punishment mechanisms, broaden the application of credit information sharing, vigorously promote credit construction in key areas, continuously improve credit regulations and systems, strengthen publicity and education on integrity culture, and innovate the construction of a social credit system. People's social behaviors (such as economic transactions) are recorded in a unified social credit system. If a person's social credit is too low, he will be punished accordingly, such as being restricted from using transportation. If a person's social credit is good, he can get corresponding rewards, such as his credit card limit can be increased accordingly. Another case is that, the construction of laws and regulations should be improved, and certain behaviors should be rewarded in more details (for example, teachers with strong teaching ability can enjoy more opportunities for professional development), and certain behaviors should be punished (for example, after Drivers should revoke their licenses for a period of time).

Theory II shows that when people encounter setbacks, they tend to make deviant aggressive behavior to some people with the authority to vent their emotions. According to this theory, I believe that society should promote the protection of equality among people in including different regions, different social and economic statuses, different education levels and different genders. In addition, we must not only strengthen equality for all in the law, but also pay attention to equality in social welfare and resource allocation. For example, the government can invest more educational resources in areas with poor education to protect the local children's right to education. Furthermore, the protection of the right to life of individuals should be strengthened. To be specific, domestic violence, assaulting supervisors and public officials in the country should be recognized to be against the law. Those who commit crimes should be punished by law.

Theory III shows that people who do not have a dominant position in resource allocation are more likely to make deviant aggressive behavior. According to this theory, it is important to promote equality among people including different ages, classes, cultures and health statuses. To be specific, for example, the large gap between the rich and the poor is an important issue currently facing Chinese society, and this problem has caused many social contradictions. For instance, people living below the poverty line may be more likely to commit crimes due to a lack of life resources or an unbalanced mentality (for example, in recent years, many of the murderers in campus homicides in China have lived in extreme poverty). In order to reduce these phenomena, the government can improve the transfer payment system and allocate social resources from the developed regions to the less developed regions. The government can also adjust the tax structure and rely more on personal income for taxation.

Theory IV means that the deviant subculture produced in certain social environments will lead people who are exposed to this subculture to produce deviant aggressive behavior, and people in this subculture are regarded deviant. As a result, in areas with the highest levels of social chaos and poverty in cities, certain forms of crime may have become a cultural norm, passed from one generation to the next as part of a normal socialization model. I think society should give people of the unprivileged class more life-changing opportunities. For example, let children in slums or areas with higher crime rates have better access to education, and adults in these places have more job opportunities through re-education. When these people who should be exposed to deviant subcultures are integrated into the social environment, they are likely to change their views and abandon the subcultures of their original beliefs. In addition, the government should provide various opportunities for people affected by subcultures and those who are not affected by subcultures to communicate more and reduce isolation. For example, the government can consciously increase the diversity of public schools, allowing children from subcultural regions and non-subcultural regions to study together. This policy can be beneficial to break isolation, increase communication, and change the bad habits of children from subcultural regions.

## **Waiting until the last minute**

- a. I can figure out some explanations for this phenomenon.
  - a) Sometimes people set abstract goals for themselves and do not know where to begin. As a result, they choose to stop doing it. For example, when university students start to work on their graduation thesis, they just do not know how to write it from choosing the topic to collecting the data. Since the students have a lot of other things to do, for efficiency they may just put off the graduation thesis and do other things first, like finishing a problem set due soon. However, if they set more specific goals for themselves, like on Monday they should choose a topic and on Tuesday they should finish the

literature review, they may less likely to procrastinate writing the thesis.

- b) People may always want to get their work done before the deadline due to their excessive pursuit of perfectionism. For example, students always want a higher grade for their homework. Then before the deadline, they will use their time to modify their work as much as possible to make it look more perfect.
- c) People may have many different tasks over time, and they will tend to do the most urgent. Therefore, the closer the deadline is, the more pressure people have to do it. At this point, they may show high productivity under the influence of urgency. Therefore, to minimize working hours, people tend to complete this task before the deadline.
- d) People are unwilling to delay gratification. If people go to work, they have to take the burden of work and give up the ease of their current rest. Therefore, some people may be reluctant to work immediately. They are willing to postpone the task to get the current happiness.
- e) According to Temporal Motivation Theory (Steel et al., 2018), the higher the willingness of people to impulsiveness and delay, the lower the objective expectations of this matter, and the lower the value of this matter, the more likely they are to delay doing it.

b. Model 1: Procrastination = 
$$\frac{\text{Present Satisfaction} * \text{Perfectionism}}{\text{Working efficiency}}$$

In this model, we assume that the working efficiency and self-control ability will not be 0. That is to say, the person knows or can learn how to do this task, and he can control himself not to absolutely ignore this task. The dependent variable measures people's tendency to procrastinate. The higher the value of the dependent variable is, the person is more likely to finish the task at the last minute. There follow the explanations of the independent variables.

**Present Satisfaction** measures people's choice between delayed satisfaction and present satisfaction. A person with a strong ability to delay satisfaction instead of enjoying immediate happiness will be assigned a lower value on this variable and thus he is less likely to procrastinate.

**Perfectionism** measures how people want their work to be subjectively perfect. The more persistent a person is in pursuing perfectionism, the more likely he is to use all the time before the deadline to perfect his work, the more likely he is to delay.

**Working efficiency** measures how quickly the person will finish the task. If a person is very familiar with the task, he will be sure that he can finish it in a certain amount of time. Thus, he may not immediately take action to deal with the task and he can certainly wait until the task is very urgent.

c. Model 2: Procrastination = 
$$\frac{\text{left time} * \text{age}}{\text{how work planned} * \text{Willingness to do the task}}$$

In this model, we assume that the variables, how work planned and willingness to do the task, are not equal to 0. That is to say, the person has the ability to finish this task (so he knows how to make the plan) and he will do the task (even later or very closer to the deadline). The dependent variable measures people's tendency to procrastinate. The higher the value of the dependent variable is, the person is more likely to finish the task at the last minute. There follow the explanations of the independent variables.

**Left time** measures how much time from a time point to the deadline. We assume that the more urgent the task is, the person is more willing to do the task and will not procrastinate.

**Age** measures a person's age. We assume that the younger the person is, the person is more energetic. Thus, working will be a bigger burden for older people. Therefore, younger people will regard the same piece of work more easily and work more efficiently, and as a result, they can spend fewer days to finish the tasks which makes Procrastination lower.

**How work planned** measures how specific the task is divided into different tinier tasks. We assume that with clearer and more specific tasks, people better know what they should do, and they are more likely to take action without procrastination.

**Willingness to do the task** measures how the person likes the task and wants to finish it. If the person has a strong motivation to finish the task, he is less likely to procrastinate.

d. For model 1

① High Procrastination

Say Julia has a winter break (1 month) and 1 month at the beginning of the spring semester to finish her graduation thesis (she can use the 2 months to finish the thesis at most). As long as she finishes the thesis, she can get her degree. However, at the end of the last semester, she married John and they planned to spend a honeymoon. Julia thinks John is very important to her and she does not want to put off the honeymoon.

Therefore, in this case, Julia has a higher value of present satisfaction.

However, Julia takes her study very seriously, so when she comes back from the honeymoon, she started to work very hard on the thesis. She wants to write a high-quality graduation thesis to get the title of honorary graduate. Therefore, she has a high perfectionist tendency in this matter, and she hopes to use all available time to improve her thesis. As a result, she has a high value of perfectionism. Also because of her perfectionism, Julia is very careful about each step of the thesis, including choosing the

topic, reading the literature, collecting data, choosing statistical methods to analyze the data, and so on. She must think repeatedly and make an appointment with her mentor to discuss it before she can decide what she thinks. Therefore, her working efficiency is very low.

In this case, Julia is very likely to submit a thesis at the last minute.

## ② Low Procrastination

Say Bill is a senior data scientist with excellent professional skills at Google. His boss told him that he could take a month off and enjoy a vacation as long as he finished the project at hand. Bill really wants to use a vacation to travel to Japan because his girlfriend is studying in Japan. Therefore, he might sacrifice time to watch a movie in the last few days to work. In other words, his present satisfaction value is very low.

As mentioned above, Bill is an excellent data scientist. His working efficiency is high. At the same time, he can make sure that as long as he finishes a task, he can not go wrong on this project and his boss will be very satisfied with his work. Then, it is impossible for him to sacrifice too much time to study the project repeatedly due to perfectionism. His perfectionism value is low.

In this case, Bill is very unlikely to submit the project at the last minute.

## For model 2

### ① High Procrastination

Sam is a retired manager of a hotel. One day, his previous boss, Lucy, found him, hoping that he could share his work experience at the hotel's annual conference six months later without compensation. That is to say, Sam has a long time left to do the job (compared to the Low Procrastination example). At the same time, Sam is already an elderly person, his physical fitness is not particularly good, and his brain may not run very fast like young people. Because of his age, Sam may spend a greater number of days doing the work.

In addition, Lucy only told Sam that he needed to share his work experience. But Lucy didn't tell him who the audience was, what content they wanted to listen to, etc. As a result, Sam wasn't quite sure how to organize his speech. This task is more abstract for Sam. And Lucy, or the hotel, doesn't pay Sam for his speech. Therefore, Sam is not very happy to speak about the hotel.

In this case, Bill is very likely to finish the speech preparation at the last minute.

### ② Low Procrastination

Brenda has just graduated from Harvard with an undergraduate degree and is a new employee at Prada. She hopes to become a leader in the fashion industry in the future. She admires Melissa, a fashion editor at Prada. One day, Sarah, Melissa's secretary, became ill, but Melissa needed to attend a new product launch the next morning. She needed a PowerPoint file to support her speech. Sarah recommended Brenda to Melissa for the job. Melissa hopes to get this document as soon as possible to prepare in advance.

In this case, Brenda only has 24 hours of preparation from being informed about the task to Melissa's deadline for submission. In order to make Melissa more satisfied with her work, Brenda hopes to complete it as soon as possible. So, she might finish the task in less than 24 hours. Therefore, her left time value is extremely low. Brenda is only 22 years old this year, and she is very young (that is, the age value is very low). Sarah had already told Brenda very clearly the content and order of the PowerPoint file, so Brenda knew exactly what she needed to do the job. In other words, the work is well planned (big value). Brenda hopes to make a good impression on Melissa, and she thinks it is a good opportunity to improve her ability to work. So, she is very willing to do this job.

In this case, Brenda is very unlikely to submit the PowerPoint file at the last minute.

## Selecting and fitting a model

1.

a.

The performance of a flexible statistical learning method is better.

With a large number of samples, we can better fit the parameters. In addition, fewer predictors can make the model variance less.

b.

The performance of a flexible statistical learning method is better.

With too many predictors and only a few samples, a flexible statistical learning method can cause the problem of overfitting.

c.

The performance of a flexible statistical learning method is better.

The reason is that we can expect more possibilities for the shape of the model.

d.

The performance of a flexible statistical learning method is worse.

When the variance of the error terms is too high, it is possible that the flexible statistical learning model overfits.

2.

a.

**Bias** will decrease, because more factors will not be judged subjectively by people (whether they are important or not). People impose fewer assumptions on the model before analyzing the data.

**Variance** will increase, because a certain sample can exert a bigger influence on the prediction results.

**Training error** will decrease, because when a model causes the problem of overfitting, it at the same time can lower train error.

**Test error** will decrease at first and then goes up. It has the lowest point. Test error is the error when people run the model got from the train data on the test data. It is about the relationship between bias and variance. With models of different flexibilities, we must have one that makes the error lowest, that is why we get the lowest point. At first, when we increase the flexibility of the model, there must be one model that can be more accurate than others. However, after this flexibility degree, the increasing flexibility increases the test error.

**Irreducible error** does not change. Irreducible error is a measurement of the amount of noise in the data. No matter the quality of the model, the data includes noise.

## Reference

Piers Steel, Frode Svartdal, Tomas Thundiyil, & Thomas Brothen. (2018). Examining Procrastination Across Multiple Goal Stages: A Longitudinal Study of Temporal Motivation Theory. *Frontiers in Psychology*. <https://doi-org.proxy.uchicago.edu/10.3389/fpsyg.2018.00327>