**Tell me about yourself**

First, thank you for the opportunity to talk to your team. This month, I just graduated from the University of Minnesota, Twin Cities with a master degree in Statistics.

In my school life, I have taken many statistical analysis courses including both in theory and applied.

And, in school, I was also in a student group called Statistics in the Community (STATCOM) as a volunteer statistical consultant and also an officer of the group. In STATCOM, I have been working with some local non-profit organizations. What I did is analyzing their real-life data and then answering their questions based on the analysis results we made.

And, I also had experience working as Home-Based Model Consultant in DiDi which is a company similar to Uber but it is in China. In DiDi, I was responsible to propose predictive models for their real-time data by doing research on automated models.

And now, I kept self-learning knowledges of computer science, machine learning statistical algorithms and I am also interested in artificial intelligence.

Beside just learning by myself, I also try to apply what I learned to my personal projects including creating my personal website (using Jekyll, HTML, CSS and some javascript libraries and host it on Github Pages ) which is about my journey on learning data science, I also try to do automated web scraping to collect and analyze data from job searching web (like job searching websites).

I am the kind of person who is excited about seeing the problems and solving problems and am not afraid of learning new thing and I am excited about that. So, I believe that I can definitely get used to the agile environment of e-commerce industry and bring benefits to your team and company.

**Questions to ask employers/ interviewers**

1. When I was doing research about e-commerce. I saw a lot of people talking about how AI can change the world of e-commerce. What do you think about that? Do you agree that or do you think it is not so easy, it still have a lot of difficulties?
2. What are the big challenges that you are facing now? What kind of the data problems that you really want to solve now?
3. If I started to work in the company, what is your expectations on me?
4. How do you like working in your company?

**Experience**

1. **DiDi Chuxing (Home-based Model Consultant)**

Methods:

Exponential Smoothing (adding weight to the past data in an exponential way. 4th power of alpha)

🡪

TBATS model (add trend and seasonality, ARMA on the residuals part to de-correlate data, Box-Cox transformation to deal with non-linear feature and then have trigonometric transformation on seasonality terms to reduce number of parameters)

Situation

In this position, my tasks are doing

1. model improvement to increase the accuracy of the point estimation for a really short period of time like next minute.

2. data anomaly detection by using prediction interval

and all of these must be based on their current model Holt-Winters model (Triple Exponential Smoothing) and with strong theoretical support.

Task

The problems of their current model are that it is not reliable in certain situations. For example,

holiday and weekend. And their prediction interval is manually defined without any supports from theory.

Action

To solve these problems, I read many academic paper about automatic time series prediction methods (most of the resources are from Prof. Rob J Hyndman from Monash University, Australia) and summarized all the literatures I read for company’s future use. Then I proposed using TBATS model to replace the original model. With the help from my teammates, we have successfully tested the TBATS model online and see the model performances.

Result

In the end, the supervisor valued the model a lot and has invited me to work in Beijing for further discussion and testing. It is my great honor.

1. **Statistics in the Community (An Officer and a Statistical Consulting)**

* Statistical Consultant

Situation:

I had a project working with United Way of Washington County – East

This is an organization who gathers peoples’ donation and then distributes to people or other organizations in need.

Task:

In this project,

The data we got is yearly amount of donations for each donor about 4,000 from year 2005 to 2013. And the client know that their total donations were decreasing since 2006 and wanted to stop the decreasing by doing something on their following fundraising campaigns.

Action:

After working with friends in business school and taking suggestions from our alumni, our team decided to add the business knowledge, which is called 80/20 rule, in our **K-means Time Series clustering** model. In the end, we gave the client a list of donors categorized by their donation trend.

The main categories are target donors (who have ever donated $5,000 or more) and small donors. And in each category, we have subcategories like donors with increasing donation trend, donors with decreasing donation trend, donors with stable donation trend…etc.

(Generally, 80% of incomes/resources or money is from 20% of population. In our case, 80% of their total yearly donation were generally from donators who donated $5,000 or more in a year)

Result

This list not only gave them a good understanding to their donors but also provide them some ideas about what to do in their following fundraising campaigns.

When we tried to follow up the fundraising campaign, our client told us that they were in personnel transition and he left the organization and had no idea about how the campaign went. So, we stopped tracking this project.

* Officer

Situation:

It is a totally student-run organization. The time I took charge of this organization in 2015 is only its second year. It is really new.

Task:

At that time, not many students or local non-profit organizations know our group. So, the first thing we did in our year was to “build the foundation” and “advertise STATCOM”.

Action:

* I started to construct official website and publish it on the school page.
* Also, to make our group be more organized, we established the project process flow and set up organization rules for statistical consultant and clients.
* My team and I actively recruited students as volunteer statistical consultant by giving talks in student orientation and seminar.
* And develop new clients by actively sending email to introduce who we are, giving presentation on Minnesota Councils of Nonprofit.

ex:

It’s like before projects start, statistical consultants have to sign the Statistician Contract and clients have to sign the Client Pledge to make sure that both party are on the same pages.

Result:

Then, we did a good job on establishing positive group image and received an interview from the College of Liberal Arts (in Nov. 2015). They wrote an article about us and published on school website. So far, we have been working with over 10 local nonprofit organizations in Minnesota.

1. **Statistical Consulting Clinic (Statistical Consulting Intern)**

Situation:

This clinic provides statistical support to improve research in the University of Minnesota.

Task:

Under the supervision of senior statistical consultant, I worked with students and professors to do the data analysis part in their research paper.

Action:

I was responsible to do data quality checks, data cleaning and statistical analysis in R and attended regular meetings to present weekly summary with clients and senior statistical consultant. And in the end of the projects, I wrote the final data analysis report and explain in detail to the clients.

Chemistry

“Mitochondria study in a Muscle Fiber and Mass Cytometry Study in Mouse

Proteins.“

I used permutation test to check our hypothesis and execute

bootstrapping techniques to find the confidence interval of mean difference.

H0: mitochondria size in left end, middle and right end are the same.

H1: they are not all the same

Nutrition

“Liking of food textures and relationship with oral physiological parameters”

This project was about reproducing results from the published paper in their field by recollecting data and redoing the data analysis. I help the student on data cleaning and executing the non-supurvised machine learning technique, Principle Component Analysis (PCA) in R.

Result:

I not only provided professional statistical suggestions to my clients but also ease their pain on doing statistical programming and data analysis.

1. **My master thesis (Analysis of size-biased mitochondria data)**

* Sized-biased data always happened when you do sampling on image.
* Imagine that you have a picture with thousands of cells in there. Some of them are large, some of them are small. And you are interested in knowing the mean size of all the cells.
* The easiest way for you to do sampling is generating a list of random coordinates and the cells will be chosen if their area on the image includes more than one coordinates.
* As the result, the cells with larger size are more easily to be chosen compared to the small ones. In this situation, if we use sample mean as our estimator for the population mean, it will be overestimated.

Situation:

My master thesis is the extended study of one of my projects in Statistical Consulting Clinic. The one about biochemistry.

Task:

I found that the sampling process in the summer project violated the assumption of the statistical method we used and then may make the estimation become biased.

Action:

So, in my thesis,

I started doing research on finding a better estimator under this kind of sampling process. And did data simulation in R to verify the result and the hypothesis.

Result:

In the end, I found the appropriate estimate for this project and provide a strong foundation for further

research.

**Technical Questions:**

**Central Limited Theory (CLT):**

* The sampling distribution of the sample mean will approximate to normal distribution when the sample size is large.
* For example, suppose you have a class of 30 students, and then you randomly picked 20 of them and then calculated their sample mean score. And keep repeating sampling and calculating. In the end, you will find that the histograms of all the sample means will be like normal distribution.

**p-value:**

* use example to illustrate

Suppose there is one person say the population mean is >80. To see if his statement is correct or not, I do sampling and then get a sample mean equal to 20 with p-value 0.01.

That means that if what he said is correct, then the probability for me to get the sample mean equal or smaller than 20 is 0.01 which is so low that I may have hard time to believe what he said is correct.

**Type I and Type II error:**

* Use example to illustrate

The null hypothesis is that this woman is pregnant

To see if the null statement is true, I may use machine to check.

And then the type I error is that the true is woman is pregnant but the machine say she is not.

The type II error is that true is the woman is not pregnant but the machine say she is!

**Confidence interval:**

* This is the interval that I have confidence that the true parameter lies in.
* For example, if you asked me to guess what will be the mean score of our class, I can give you a point estimates but it is hard to be correct. So, I decide to give you an interval that I think the true parameter will lie in.

**Behavior Questions**

**Examples: STATCOM, Plan B, Personal Website, Web scraping project**

1. **Resolving Conflict Effectively**

* When I was an officer in STATCOM, other officers and I had some conflicts when setting up the organization rules.
* At that time, what we did were keeping communicating, talking about pros and cons, listing out all the possible bad consequences for each proposal and then choose the proposal that we think we can bear its all bad consequences.

1. **Strength**

* I am a person with self-motivation and strong persistence to reach goals. I would like to take my personal website as an example.
* Even though it is not easy for me who do not have any computer science background to set up a website, I did not give up but teach myself how to do it. I kept googling, took many open courses, did many experiment.
* Finally, in this January, I successfully put my personal website online. I tried to publish posts regularly. And, the part I am proud of most is that I am able to connect Rstudio with my website, so all the posts on my website can be generated through R Markdown which is the tool I am really familiar with.
* So, I think my passions to learn new technologies and new things can make me be adaptive to the consulting industry which is a fast-paced and dynamic environment.

And, now after having some basic knowledge about the website, I have a new goal which is learning how to do web scraping. And then start to do the project I am interested in by collecting data from websites.

1. **Why do you like statistics?**

I think the short answer for this is I want to solve real-life problems in a logical way. And, statistics is the subject that can combine real life questions with strong mathematical theory which is full of logic.

1. **Name three things that you want to get from this job?**

* I think the first one is experience. I really want to know what kind of real life problems are really needed to be solve and how can I apply my skills on that to solve it.
* The second thing is the chance to work with people from different background. Because only when many different kind of people work together, can things become big and influential.
* Finally, I love technology. So I am looking for a company who can always be excited to embrace new technology and new methods.

1. **What is the most pressured time you have encountered, and how do you handle it?**

* I think the most pressured time I have encountered is now, the time for job searching. I know that I really want to get a job, but I do not enjoy the process of searching job. Search the same kind of jobs but in different job websites.
* To make all the process becomes more interesting, I started to learn how to do web scraping and application development and expect in the near future I can have an application which can help me integrate the information from all the job websites. So, I do not need to do repeating work anymore.

1. **Do you have any team work experience, presentation experience?**

* Yes, I have many team work experience and presentation experience.
* When I was an officer in STATCOM, I had worked with other officers, alumni and school professors to discuss how to make our group become more prosperous and reputable.
* And to recruit more volunteer statisticians and develop new clients, we did many presentations on student orientations and the meeting of Minnesota Council of Nonprofit.
* And I also have experience on doing presentation to our STATCOM project clients and clients from Statistical Consulting Clinic.

1. **Disadvantage of yourself?**

* Well, I have a bad memory. And I found that the more I learned, the more I forget.
* I tried to record all my learning process and knowledges in my personal website. And, be honest, this is the main reason why I want to set up my personal website.
* And I expect that, if next time people ask me some questions that I could not remember the answers. I can just check out my website and all the memory will be fresh again and then I do not need to learn from the very beginning again.