

Tsinghua Mathcamp Summer 2015
Probability & Statistics
Wei Zhu, Ph.D.
Professor & Deputy Chair
Department of Applied Mathematics & Statistics (<http://www.ams.sunysb.edu/>)
State University of New York at Stony Brook (Stony Brook University)

Dear Students:

When I was your age, I was wondering what my future college major should be. My father, an accountant himself, suggested accounting. However a younger me felt that dealing with those numbers must be incredibly boring. So, I chose mathematics as my undergraduate major instead. As I matured up, I realized that numbers are indeed, not boring at all – it is absolutely fascinating to analyze all sorts of data and let the data speak for themselves on nearly every facet of our lives. Those stories are not just about the past, or the present – but, perhaps most interesting of all – also our future. Imagine that through this summer camp, you will gain the double credentials of (1) a ‘data’ detective, and (2) a ‘scientific’ fortune teller. I think you must be motivated or at least curious by now.

Now, what shall we really learn through this summer camp? We will start by learning the fundamental probability and mathematical statistics theories. As you may have heard, statistics is more like a sibling of math. If you like math, you will enjoy statistics – as I have experienced this myself. The difference between mathematics and statistics can be described by the following simple example. Suppose we are curious about which of the following two soda brands are more popular, ‘Coca Cola’ or ‘Pepsi’? If we have a way to learn the opinion of everybody in the population of interest, say China – then all we need to do is to simply count how many prefers Coca Cola, and how many prefers Pepsi. The dispute can be easily resolved by basic math. However, as you are intelligent, you must have realized that this is a mission impossible. We have neither the manpower nor the funding to ask the opinion of every eligible Chinese citizen. So instead, the best we can do is to select a good, representative sample (of modest size) from the entire population, and then inquire the preference of everyone in the sample. However, one such sample may yield one conclusion – say more prefers Coca Cola, while another may yield a contradictory one – say, more prefers Pepsi. So, how can we make a unanimous (and convincing) conclusion about the entire huge population based on one modest sized sample? This is the job of a statistician. And this is one of the first things you will learn in this summer camp.

Statistics is applicable to nearly every walk of life, research and business. Some of us statisticians focus on helping medical researchers unlock the mysteries of genes and proteins and disease pathways – to develop better medicines and to save more lives. Some of us focus on the quality control of various industries and business. Some of us help various retailers examine customer credentials and preferences to ensure best sales and services. Some of us develop stochastic models to understand how our economy works together, to predict future financial crisis, and to devise various socially responsibly investment strategies. With the rapid development of information technology, we have increasingly more data to analyze – this is the age of big data – and bigger roles for current and future statisticians. So, welcome!