# Initial Posts

|  |
| --- |
| **ThinkStats Takeaways**  Allen Downey is super awesome at taking complex topics and breaking them down into a simple understandable topic.  I didn't like all the custom function stuff...  I enjoyed his take on simulation and resampling for estimating population statistics. There are however many analytical methods that are much faster than simulation but the caveat is that you have to have a deeper understanding of the mathematics behind them. |
| **Cumulative Distribution Function (CDF)**  One concept that I really enjoyed is the importance of CDF's in data analysis. They are incredibly simple to calculate and give a ton of insight as to how the probabilities density is distributed for a random variable all without any bias or density estimation. |
| **Classes & Methods**  During this course, although I didn't like how Allen Downey used a bunch of custom code, I did learn a lot about using Classes/Methods/Properties. I wanted to learn how he built his custom stuff which forced me to learn more about coding Classes in Python. After creating a couple of my own custom Classes I actually really enjoy them and see why they are so useful. |
| **Probability Density (Function vs. Estimation)**  I feel like this topic really switched my brain into thinking 'prediction' vs. 'inference'. At first I was like, "cool, yeah you're just taking the frequency of a value divided by total number of values". But that is just explaining the current data. You can't 'generalize' that onto a population. Then came the concept of Probability Density Estimation and then I got that 'ah ha!' moment. I really enjoyed trying to figure out how to test and simulate in order to find a distribution that fit the data the best so that you could then predict the population values. |
| **DSC 530 Takeaway**   * I really enjoyed this class and really feel like a took a big step forward in my understanding of predictive analytics and Python. * Professor Shankar Parajulee, I appreciate how responsive you were throughout the course and I thought you gave great feedback on our assignments. * If I had to pick 1 thing I could change it would probably be the copying of our posts in Teams, then opening Word, pasting the posts there, then at the end of the week going back into Teams and then uploading our posts that are already on Teams. It just seems sort of redundant to me. Other than that, this class was definitely my favorite so far. Very challenging and fun at the same time and also learned a lot of EDA, Python, and Statistics. |

# Replies

|  |
| --- |
| I'm taking 1 class per semester and took DSC 520 last semester but I totally agree that they pair nicely. I definitely enjoyed using Python over R though. I mean, R isn't bad or anything but I just prefer Python. It is much more intuitive to me and there seems to be much more documentation and a bigger community online. |
| That is one thing I really did not like about Thinkstats by Allen Downey. He had a bunch of custom Classes/Functions/Methods that made me curious how it is actually built. For the majority of the assignments I created my own functions because I wanted to learn how. On a positive note though I think he does a fantastic job at taking complex topics and breaking them down in a simplified way that is not all 'Statistic Gibberish'. I think too many statisticians make topics overly complex like their trying to write a Wikipedia article or something.... |
| I totally agree, I think visualizing data compliments statistical testing and data analysis very nicely. |
| I agree Robert, this class in particular has really made my brain start thinking more about 'predictive' analytics vs. 'inferential' analytics. |
| Yes, same here Robert, I feel like that is also something I could learn more about. I was actually just thinking about this the other day regarding density estimation. Like, how do we determine the function that will fit the probability distribution?  I know this is cheating and I really do want to find out more about fitting a formula to the distribution but check out the FITTER package below. You can pass a variable and it will run through over 80 known distribution functions and rank which one fits your data the best.  <https://medium.com/the-researchers-guide/finding-the-best-distribution-that-fits-your-data-using-pythons-fitter-library-319a5a0972e9> |