

2.)

Recall we discussed the role that random variables play in applied data science. We also discussed the distinction between discrete and continuous random variables. For the random variables below, indicate the more appropriate random variable (discrete or continuous) and why you believe this to be the case. [7 Points]

The number of taxi rides taken in a month by a NYC resident.

Discrete Variable, as they do not show rapid change and the samples can be present at defined intervals.

The speed of a bicyclist on Jay Street.

Continuous because this will be measured by averaging the speed of all the people in dataset by no of people and it can be defined at small intervals.

The luminosity of light emitted by street lamps in Brooklyn.

Continuous because it needs to be measured and defined within certain limit,can defined at minute intervals.

The income of bankers working on Wall Street.

Discrete because certain highest salaries are known and average salary don't show rapidly change.

The number of hotels in Manhattan.

Discrete would be more appropriate because no of hotels don't change frequently and their number remains constant over a certain period.

The ambient sound generated by trash trucks picking up trash at midnight.

Discrete as it can be categorized into harsh,mild,soulful,etc

The quality of coffee served in the student lounge.

Discrete, because it can categorized into bad,good,etc.

3.)

Consider the salaries of bankers on Wall Street. One argument that could be advanced is that Wall Street bankers have high salaries because they have attained high levels of education. Another argument could be advanced that Wall Street banker have high salaries because they have high Intelligence Quotients (IQ). If you had a dataset that provided you with the salaries of a sample of Wall Street bankers, together with their education levels and IQs, discuss how you might explore these arguments. Are there other methods beyond relying on this sample of bankers that could allow you to explore these arguments? (As with most applied data analytics, there is no right or wrong answer to this question. You may find it helpful to consider what we discussed in class to address it.) [6 points]

Basically we are struck at a situation when we have a dataset comprising salaries, IQ and Education level of wall street bankers. And with not much information, the questions and facts that arise are:

- 1) Are IQ and Education sufficient for making an accurate prediction of wall street bankers
- 2) Are IQ and Education showing multi collinearity, means in most of the cases more educated people are the ones with highest IQ's, will that not create an identical dependence on salaries, so both will denote same correlation with the dependent variable hence we need to include other factors for measuring variation in salary.
- 3) Salaries of wall street bankers are continuous variables.
- 4) Other factors that highly influence the salaries are like the background of the family, qualifications of parents, standard of living, no of bankers in the family, social status have not been included in the dataset and their potential influence on the dependent variable (salaries).

Lets see a sample dataset as below

Education	IQ(Scale On 10)	Salary(\$)
1	4	100,000
2	5	120000
3	6	130000
1	4	110000
2	5	120000
3	6	130000
3	6	130000
2	5	120000

As we can see from above as IQ decreases salary decreases. and as educational qualification drops salary drops so also IQ and Education seem to be affecting the

salary in the same way. Hence this becomes a situation of multi collinearity where both variables (IQ and Education) have equivalent dependence which implies that we need to include some external factors so as to make an accurate prediction of salary and able to see the variations in the same.

In order to be more explicit what needs to be done in this case of dataset, the question is very subjective in terms of what possibilities we have with the data. Hence in order to be certain about right measuring of variation of salaries, it would be advised to include other factors and try figuring out whether IQ and Education affect the salary in the same way or not.