**Tech Questions**

**1.What’s the difference between R and Python in doing data analysis?**

1. Python has more API. For example, API with AWS etc.

2. Normally, Python is for CS background. Instead, R is for statistician. Prof. said in terms of data analysis, 40% data scientist use R, 60% ones use Python.

3. Python has proficient packages for Deep Learning

**2. what’s the difference between Histogram and Bar graph?**

<https://keydifferences.com/difference-between-histogram-and-bar-graph.html>

**3. what’s the difference between Probability Distributions: Discrete vs. Continuous?**

https://stattrek.com/probability-distributions/discrete-continuous.aspx

**4. Sample and Population**

https://www.mathsisfun.com/data/standard-deviation.html

**5. what is bird view of Statistics, Probability, Mathematics and CS**

Wikipedia

**6. What’s the difference between Covariance and Correlation?**

<http://ci.columbia.edu/ci/premba_test/c0331/s7/s7_5.html>

**7. The application of F-Dsitribution**

https://stepupanalytics.com/application-of-f-distribution/

**8. Why do we use Logarithm for cost function of Logistic regression?**

<https://stackoverflow.com/questions/32986123/why-the-cost-function-of-logistic-regression-> has-a-logarithmic-expression

**Data Science Concept & Definition**

Statistics and Probability

**Random variable**

**Probability distribution:**

F-Distribution

t-distribution

chi-distribution

**Skewness:** https://365datascience.com/explainer-video/skewness-example/

**Explanatory variable**

**Response variable**

**Degree of freedom**

**Residual**

**Coefficients**

**Parameter**

**Hyperparameter**

**Significant level**

**Confidence interval**

**Mean**

**Median**

**Medium**

**Average**

**Quantiles, Q-Q plot**

**Quartile is a type of Quantiles, Q1= 1st quartile, Q2= Median, Q3=3rd quartiles**

**Variance**

**Covariance**

**Correlation coefficient**

**Coefficient of variance**

**All kinds of distribution: Discrete Variable, Continuous Variable**

**Hypothesis test: P-value**

**Moment**

**Moment-generating function**

**R2: coefficient of determination**

**Standard error: https://stattrek.com/estimation/standard-error.aspx**

Scatterplot matrix: https://www.r-bloggers.com/scatterplot-matrices/

Mathematics

Computer Science

**API**

Database & Big Data

ML & DL

**Linear Regression**

**Logistic Regression**

**Regularization:**

L1 Regularization

L2 Regularization

**Lasso Regression**

**Decision Tree**

**Random Forest**

**Naïve Bayes:**

Independent variable

**SVM:**

Error Function: C\* Classification Error + Margin Error

Hyperparameter C

Ensemble Methods

Model Evaluation