DS 1.1 Study Guide

Final Exam Format: Your final exam will be on March 6th. It will be a written exam that will take 1 hour.

Overview: Your final exam will focus on the learning outcomes for DS 1.1 course. Review slides or Jupyter Notebooks and summarize your understanding of each slide in high-level. There would be no-formula or computations or coding on Final Exam but will ask about the general understanding of different topics. No question about Time Series would be on Final Exam.

Sample Questions:

Q1: Explain what is null-hypothesis by mentioning an example

Answer:Imagine we have design a website with template A. The frontend developer applies changes and now the template is B.

Statistically we know the users interaction with template A, for example we know the users click-through rate (CTR)

The new template B is exposed to N new users. We gather the CTR for these new N users

The null-hypothesis says that the average CTR for template B is the same as average CTR for template A, or we say nothing has changed for CTR

The alternative-hypothesis says, with confidence of 95%, the average CTR for template B is greater than template A

Q2: What are the steps to reject or accept null-hypothesis

Answer:
1- If N > 30 or the standard-deviation (std) for template A is known, we calculate z-score

2- Obtain the survival function of Z-distribution (standard Normal) at z-score. This would be p-value

3- Compare the p-value with significant level (5% for example). If p-value is less than significant level then reject the null-hypothesis

4- if N< 30 we do the same steps with t-score and T-distribution

Q3: What is confidence interval for mean, by mentioning an example

Answer: We have N (50) samples of sepal length for setosa, we want to obtain an interval for the mean value of sepal length and say we are 95% confident that if we gather much more samples for sepal length the mean of these population with be in [L, U] range. We obtain L and U

Q4: What is an unpaired t-test, by an example

Answer: In this case, we have not applied changes, but we are comparing two groups, for example weights of female versus male. We want to say we are 95% confident that the weights of female are different than weights of male

Q5: A list for weights of men at Make School is given, how we do calculate the CDF for men's weight?

Q6: The cumulative-density-function (CDF) of temperatures for two cities in last winter has been plotted, which city is generally colder than the other?

Q7: For two list a and b, they have similar histogram (or pdf), for example both are have Normal distribution. Can we say the sample in a and b are the same?