Class Concepts - Mapping to Final Project

Friday, February 10, 2023 10:34 AM

High Level Course Progression

Your data exists in different types

Is it continuous 1, 1.2, 1.87, 5
Is it discrete 1, 2, 3, 4, 5
Is it categorical dog, cat, pig
Is it a date/datetime 2023-02-10 (12:00:00)

**note sometimes discrete numerical is actually categorical. Students in group 1,2,3 othertimes discrete is discrete, number of children 1,2,3. The difference is whether the numbers have a rank-order meaning*

Ways of using data to derive meaning

-grouping and summing, pivot tables

Conducting surveys

- interrogating the gathering of data (may or may not be applicable)

Statistical tests (null hypotheses and p-values)

Student's T-Test

(chances that the differences between two groups numerical data can be explained by random chance)

Chi Square Test

(chances that the categorical choices can be explained by random chance rather than a difference in the groups)

ANOVA

(chances that among all your categories, there exists a chi square test (or t-test) that is significant)

Error analysis

for how representative your survey should be for the wider population.

Graphs and Plots

are quick ways to see relationships - start by graphing a bunch of stuff, see if there appear to be patterns.

Different types of plots and how to make them

Statistically verify patterns found in graphs

Correlation

When one number moves up or down, does the other number move up or down (or down or up for negative correlation)

linear regression

What is the best line to predict this data (presumes a roughly normal distribution, i.e. bell curv)

mulit-linear regression

What is the best sum of lines to model this data (multiple linear regressions added together)

logistic regression/classifier

(turning things into a yes/no prediction) Can I use data in my dataset to predict things into classes. Live/Die, Purchase/Don't purchase, Win/Lose