Mathematics and Problem Solving

Lecture 10.1

Statistics

Statistics

- A **statistic** is a number that provides a summary of a set of other numbers
  - Statistics are key to modern science
  - Most modern AI is essentially sophisticated statistics

A simple statistic

- Lets say you have some text that has been encrypted using a simple cipher. It looks something like this:
  - `rhy iuksq plcaz bcd nuwfo cxyl rhy tmjg vce`

- We can **count** each letter
  - In English, 'e' is more common than 'q'.

Numeric Data

- Often our data takes the form of numbers
  - 1, 4, 2.2, 5.3, 8.2, 2.1, … etc.

- i.e. **numeric** or **ratio** or **scalar data**
  - Temperature measurements
  - Game scores

- **Can be ordered**, and the **differences between values** is meaningful

Ordinal Data

- **Ordinal data** can be **ordered**, but the **ratio** between the values is not meaningful

- e.g. we ask people to rate how much they enjoyed using our app, on a scale of 1-5

  - 1, 2, 5, 2, 3, 4, 1, 2

  - In this case 2 does not necessarily mean they enjoyed it twice as much as 1

Nominal Data

- **Nominal data** has no natural ordering

- e.g. we ask 6 people for their favourite animal
  - Cat, Dog, Squirrel, Dog, Dog, Rabbit

Statistics

- There are two types of statistics
  - **Descriptive Statistics**
    - Describe properties of your data (e.g. counting occurances)
  - **Inferrential Statistics**
    - Used to infer things about the world from data. (e.g. hypothesis testing)

Applying Descriptive Statistics

- Lets say you're designing a game that involves probability (e.g. rolling dice)

- There are three ways to analyse it

  - Playtesting

  - Analytically

  - Statistically – through simulation

Applying Inferrential Statistics

- **A/B testing** is widespread online
  - Deliver a new version of a website/game to a subset of users
  - Measure something (click throughs, conversions, etc.)
  - Compare the test population against a control group to determine whether the change improved engagement (**hypothesis testing**)

Variables

- **Independant Variable** (IV)

  – The thing you control

- **Dependant Variable** (DV)

  – The thing that you observe

Summary

- Statistics are useful
  - Describe what things are like
  - Test links between things
- What type of data do you have?
  - Nominal
  - Ordinal
  - Numeric