Unit 8: Analysis

Statistical Analysis

Lesson 56

Derek Ruths

Overview of unit

Objectives:
- Understand the philosophy of analysis
- Learn strategies for approaching data analysis
- Know major areas of data analysis and what questions they answer

1. What is analysis?
2. What is a result?
3. Most frequent word analysis
4. Statistical analysis   } *tools*

natrual language processing
netwrok analysis

need tools that are reproducible and well understood, widely used

Lesson overview

Objectives

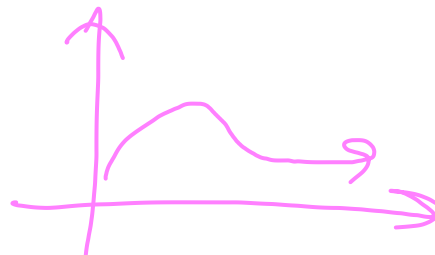- Understand various statistical tools and what they're used for

Outline

- Statistical distributions
- Significance testing
- Distribution comparison
- Time series analysis
- Survival analysis

Statistical distributions

~~fundamentally~~
capturing the probability
of observing events in specific ways

make measurements + plot a distribution

Probability
z-score
x^2

Significance testing
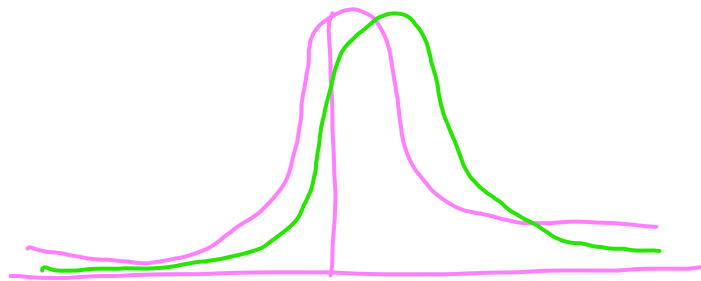
taking an observation
and asking how usual it is

tells us if a result is
unusual/significant

↑ how rare is that?
↳ use significance testing

comparing 2 distributions to each other

Distribution comparison

is dist 1 a different distribution
from dist 2 ?

need a large sample to
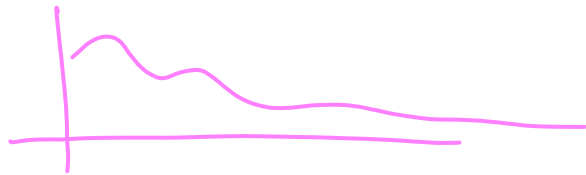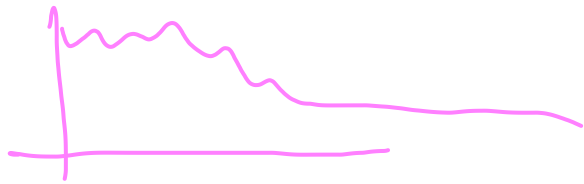get an accurate measure

Survival Analysis

Subanalysis of time-series analysis

when do things die?

- eg how long before the last comment?

- given that we see... how much
  longer will the trend last

measuring things over time

Time-Series Analysis

number of questions

↳ are the distribution/trends the same

↳ how long does a trend last

↳ will a trend resurge

Lesson wrap-up

Takeaways

- <mark>Statistics provides a wide array of tools we might use in analysis</mark>
- Get familiar with statistics!