Quantitative and Qualitative Analysis

Devan Becker

Section 1

Introduction

Land Acknowledgement

UWO exists on the traditional territories of:

- Anishinaabek
- Haudenosaunee
- Lūnaapéewak
- Chonnonton

These lands are connected with the London Township and Sombra Treaties of 1796 and the Dish with One Spoon Covenant Wampum.

About Me

- B.Sc. Math (Laurier)
- M.Sc. Stats/Biostats
- Ph.D. Stats
- Postdoc - bioinformatics and data science
- Music, reading, outdoorsy stuff

Outline

- **Quantitative:** Dealing with numbers
  - Any number in a range
  - Only 0's and 1's (maybe a 2)
  - Things we can turn into numbers
- **Qualitative:** Dealing with descriptions
  - Using your brain
  - Using your computer
- **Meditative:** Dealing with everything
  - How to get started
  - Accessing resources
  - Not being afraid of coding

Before we begin

- Interrupt at any time
- All notes/links/resources are on GitHub
- I have allowed myself **ONE** equation.

`The GitHub version also has my (approximate) script inside :::notes::: tags, which show up as text in the pdf.`

Section 2

Quantitative

Regression

- The **Target** could be any number in a range.
  - A.k.a. dependent variable or response.
- The **Features** could be any data type
  - A.k.a. explantory or independent variables

illustration

Most things are linear models

illustration: y v. x1 and y v. x2, mostly linear-looking, with slopes and intercepts labelled

Linear models: intercepts

- Exists to make the model fit better
- Not always interpretable

Mean-centering

illustration: not mean-centered -> mean centered

Linear models: slopes

illustration: one unit increase in x1

For every one unit increase in x1, y goes up by blah.

Binary Features

Suppose x1 is labelled 0 or 1.

What does the slope represent?

The story so far

- Intercept is a mathematical necessity
- Slopes answer our questions

But how good is our model?

The most important part!

illustration: error

Residual plots

illustration: the subtle sine curve that I sometimes show

Residual plots

illustration: some interesting relationship between features

Putting it all together

1. Get data
   - Data cleaning is the hardest part.
2. Plot data
   - If you haven't plotted it, you're doing it wrong.
3. Fit model
4. Check model
   - If you haven't plotted it, you're doing it wrong.

Non-linear models?

illustration: sine, polynomial, spline smoothing

It's all just linear!

Feature Selection

This is hard

Section 3

Machine Learning

Lasso Regression

- It's like linear regression, but it automatically removes features.

xgBoost

Remember the residual plots?

What if we fit a regression to those residuals?

Neural Nets