

# Course Overview

## Stat 151A: Linear Models

### Data science aphorisms

A big part of the class is developing the ability to move fluently between common sense and formal mathematical statements of data science truisms. Here are some that we cover in this class:

- Correlation does not imply causation.
- Random data makes inference uncertain.
- Unusual observations can be extremely misleading or extremely informative.
- Often we're interested in the effect of a variable "all else equal."
- Fitting your data perfectly doesn't mean your analysis is useful.

### Unit outline

- Unit 0
  - Linear algebra
    - \* Formal operations: Valid dimensions in matrix multiplication, transpose, trace
    - \* Algebra: As a representation of linear systems, invertibility, determinant
    - \* Geometry: orthogonal vectors, basis vectors, linear combinations of basis vectors
    - \* Eigenvalues and eigenvectors of square symmetric matrices
  - Probability
    - \* Means, covariance, variance, correlation
    - \* Conditional expectation, conditional variance, independence
    - \* Asymptotics: Law of large numbers, central limit theorem, continuous mapping theorem
    - \* The standard normal distribution
  - Tasks in data analysis

- \* Descriptive statistics
- \* Prediction
- \* Inference
- Unit 1
  - Simple linear regression
    - \* The basic formula
    - \* Sample means as a special case
    - \* Coefficient as a measure of association
    - \* Asymmetry between x and y
  - Matrix form
    - \* The regressor matrix and response and residual vectors
    - \* The derivation of the formula for vector-valued  $\hat{\beta}$
    - \* Being able to express familiar simpler regressions in matrix form
      - Including sample mean, simple linear regression
    - \* ESS, RSS, TSS, and R squared, and the limitations of these measures of fit
  - Interpretations
    - \* As a descriptive statistic
    - \* As a maximum likelihood estimator under normality
    - \* As a consistent estimator of a true coefficient
    - \* As a risk minimizer for prediction
  - Transformations
    - \* One-hot encodings and categorical variables
    - \* Linear transformations of regressors leave the fit unchanged
      - A one-hot encoding and a constant
      - Rescaling and centering continuous variables
    - \* Non-linear transformations of regressors give more expressive regressions
      - Polynomials, indicators, splines
    - \* Interactions between regressors and indicator variables
      - Example: different slopes in different categories
    - \* Transformations of the response
      - Changing units of the response
      - Choosing the scale that makes the most sense
      - Log transformations of heavy tailed responses
- Unit 2
  - The logic of testing confidence intervals in general
  - The selection of rejection regions
  - Level and power

- Nulls and alternatives
- Statistical tests
  - \* The multivariate normal and its linear transformations
  - \* How to construct tests of linear combinations of the OLS coefficients with the t-statistic
  - \* Testing single entries as linear combinations as t-statistics
  - \* Tests of multiple linear combinations of OLS coefficients via the F statistic
  - \* Tests that sets of coefficients are zero as F statistics
- Four classes of assumptions, when they might apply, and what they imply
  - \* Normal errors
    - Assumed “true” parameters
    - Rarely applies without good reason
    - Distribution of OLS given in closed form for all N for fixed regressors
    - Only assumption under which we have chi2, t distributions and F distributions
  - \* Homoskedastic errors
    - Assumed “true” parameters
    - Residuals have the same variance, which is more plausible than normality but often suspect
    - Similar to the normal, but only asymptotically via the CLT
  - \* Heteroskedastic errors
    - Assumed “true” parameters
    - Residual variance can depend on x
    - Similar to the normal asymptotically, but with different “sandwich” covariance
  - \* Machine learning assumption
    - IID pairs (x, y)
    - More appropriate for prediction problems, though IID assumption still needs to be taken critically
    - No “true” parameters, but there is a limiting value as N goes to infinity that plays the role of true parameters
    - Inference is the same as in heteroskedastic errors
- Unit 3
  - Expressive basis functions
    - \* What are some examples? (Splines, polynomials)
    - \* Why do we use them in a prediction context?
    - \* What are the tradeoffs determining how many we should use?
  - The bias-variance tradeoff
    - \* In the context of linear regression, what does this mean? (Squared error, conditional on the regressors)

- \* Why is it important for prediction problems?
- \* What are some ways to trade off bias for variance in regression problems.
- Ridge regression
  - \* What is it? (Formula, Bayesian interpretation)
  - \* How does the ridge parameter affect the bias and variance in the bias-variance tradeoff?
  - \* Why do we use ridge regression together with expressive basis functions?
- Unit 4
  - FWL theorem
    - \* Be able to apply it with guidance (as in the homework)
    - \* Understand what it means for constants (centering regressors)
  - Errors in regressors / regression to the mean
    - \* Recognizing that inference with OLS assumes that  $x$  is measured without noise
    - \* If you have noise in your regressors, you underestimate the true effect exactly as in ridge regression
    - \* Be able to recognize examples where you're actually trying to measure the effect of something you only observe with noise
  - Omitted variables
    - \* Why do we care for inference? Why don't we care for prediction?
    - \* When is omitted variable bias a problem? (When the omitted variable is associated with both the response and variable of interest — apply the FWL theorem)
    - \* Understand why you can't get around omitted variables by putting every possible regressor in the regression
  - Influence and outliers
    - \* Whether something is an outlier depends strongly on context
    - \* Understand the different kinds of influence of  $x$  outliers and  $y$  outliers
    - \* Know how to diagnose (residuals for  $y$ , leverage for  $x$ )
    - \* Know basic properties of leverage scores (between 0 and 1, sum to  $P$ )