Statistical Methods for Discrete Response, Time Series, and Panel Data (W271)

Professor Jeffrey Yau
UC Berkeley

datascience@berkeley

Learning Objective

Be able to conduct rigorous statistical modeling for various response data types to answer "data science" questions

Discrete Response Modeling	Time Series Analysis and Forecasting	Panel Data Modeling

- Pros and cons of each of the models
- "Applied statistics" but not shy away from the mathematical formulation
- Form a solid foundation to continue to learn more advanced stat models

Learning Objective in the first 5 lectures
(Section 1 of the Course)

Be able to conduct rigorous statistical analysis for discrete data and develop discrete response statistical models to answer "data science" questions

Discrete Response Variable

- Binary
- Unordered multiclass
- Ordered multiclass
- Count variable

Binomial Probability Model: 1 Variable	Multinomial Probability Model: 1 Variable	Poisson Probability Model: 1 Variable
Binomial Probability Model: 2 Variables	Multinomial Probability Model: 2 Variables	Poisson Probability Model: 2 Variables
Binomial Probability Model: N Variables	Multinomial Probability Model: N Variables	Poisson Probability Model: N Variables
Binary Logistic Regression	Multinomial Logistic Regression	Poisson Regression
	Ordinal Logistic Regression	
Coin tossing	Dice casting and Ratings	Counts within a time period

- Parameter estimation
- Statistical inference

- Regression parameters
- Odds of a class being observed
- Probability of a class being observed

- Wald-based Approach
- Likelihood Ratio Based Approach

- A Key Assumption:
  - Independence across observations

Learning Objective in lectures 6 - 20
(Section 2 of the Course)

Be able to conduct rigorous statistical analysis for time series data and develop time series statistical models to answer "data science" questions

Time Series Data

- Univariate time series
- Multivariate time series

Time Series EDA

Basic Concepts and Intro to TSA

- Time series
- Stochastic process
- Stationarity
- TS Forecasting Formulation

Modeling

- Trend
- Seasonality
- Both trend and seasonality using simple techniques

Modeling

- AR
- MA
- ARMA
- ARIMA
- SARIMA

VAR, cointegration

Learning Objective in the last 3 lectures
(Section 3 of the Course)

Be able to conduct rigorous statistical analysis on panel data and develop statistical panel data models to answer "data science" questions

Panel Data

Data with both the temporal and cross-sectional dimensions

EDA on panel data

Modeling

- OLS (ignoring the panel structure)
- OLS (for independent cross-sections)
- Pooled OLS
- First Difference

Modeling

- Fixed Effect Models
- Random Effect Models

Modeling

- Mixed Effect Models