

Important concepts for the Statistics II Final Exam – Spring 2019

This document lists the most relevant topics, concepts and ideas covered in the course. It might help you to identify gaps in your knowledge as you revise for the final exam. You are strongly advised to revise all the slides and assignments!

Basic or general concepts

- Causal inference
- Potential outcomes framework (POF)
- Individual treatment effect (ITE)
- Counterfactuals
- Directed Acyclical Graphs (DAGs)
- Mediators, colliders and confounders
- DAGs: Back-door paths, Back-door criterion (how to satisfy it)
- Average Treatment Effect (ATE)
- Naïve Average Treatment Effect (NATE)
- Average Treatment on the Treated (ATT)
- Decomposing the bias of NATE: baseline bias and differential treatment effect bias
- Stable Unit Treatment Value Assumption (SUTVA)
- Independence/Ignorability of treatment assignment
- Conditional Independence/Conditional of treatment assignment
- Selection bias
- Randomized treatment assignment
- Observational vs experimental studies
- Natural experiments
- Exogeneity vs Endogeneity

Regression

- Mechanics of OLS
- Slope of the line of best fit: $\text{cov}(x,y)/\text{var}(x)$
- Interpretation of the coefficients
- Omitted Variable Bias (OVV)
- Estimation of the bias of a coefficient when there is OVB
- Conditions under which regression can be used to estimate causal effects
- Criteria for selecting relevant covariates
- Post-treatment bias

Matching

- Conditional vs unconditional randomization
- Stratification
- Exact matching
- Common support
- Propensity score matching
- Balance tests
- Selection on observables
- Interpreting matching output in R

Instrumental variables

- Intention to treat (ITT) effect
- Instrumental variable (Z) versus treatment receipt indicator (D)
- IV assumptions:
 - Relevance
 - Exogeneity/ignorability of the instrument
 - Exclusion restriction
 - Monotonicity
- Principal strata (or compliance types): Compliers, Always-takers, never-takers, defiers
- Estimation of LATE when Z, D and Y are binary variables ($\text{Cov}(Y, Z)/\text{Cov}(D, Z)$)
- Estimation of LATE with Two-Stage Least-Squares
- Causal graphs perspective on IV
- Which of the assumptions are testable?

Differences-in-Differences

- Assumption of parallel trends – why is it important?
- Visual approach to DD
- Algebraic approach to DD
- Implementing DD in regression: what is the estimator of the treatment effect?
- Potential sources of parallel trends violations

Fixed Effects

- Panel data vs pooled cross section
- Error term decomposition in panel setup: unit fixed effects, time fixed effects, idiosyncratic error
- Time demeaning
- Within vs between variation
- Least Squares Dummy Variables (LSDV) estimation
- Two-way FE model
- How to include time-invariant variables in a FE model
- Limitations of FE models

Regression Discontinuity

- Meaning of forcing (or running) variable X and treatment variable D
- Treatment threshold (cut-off point)
- Local randomization
- Sharp RD (SRD) vs fuzzy RD (FRD)
- Assumption of continuity of average potential outcomes
- LATE at the threshold
- Bandwidth choice (width of window in which treatment assignment can be regarded as quasi-randomized) and the trade-off between bias and variance
- Coefficient interpretation in different types of model:
 - linear with common slope
 - linear with different slopes
 - non-linear

- Falsification checks: sensitivity, balance checks, placebo thresholds, sorting
- Fuzzy RD estimation: What it is and how to interpret threshold
- Interrupted time-series (ITS): how does it compare to sharp RD?
- Internal and external validity of RD

Mediators and moderators

- Distinction between mediators and moderators
- Conceptual decomposition of total causal effect into:
 - Direct effects (varying Treatment but holding mediator constant)
 - Indirect effects (holding Treatment constant, varying mediator)

The toolbox is yours, but which tool to use?

- Which methods allow you to estimate causal effects when there is selection on the observables?
- Which methods do not require selection on the observables?
- Which methods can be used for observational data?