标红为进阶课程

共XX课时+XX课时。

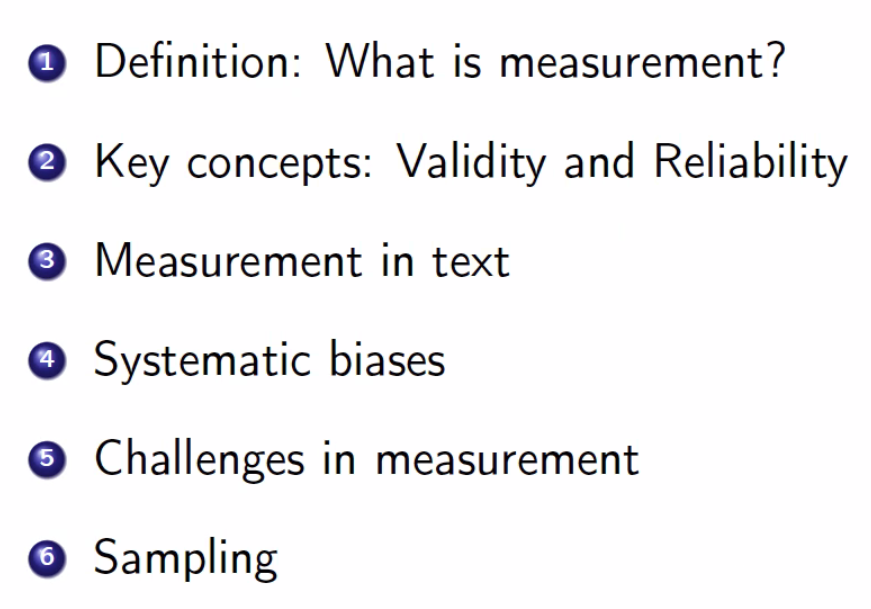
1. 研究类型：描述性、预测性、因果推断

描述数据：R的入门

* + 均值Mean, 方差Variance, 标准差Standard deviation
    - 数学解释
  + 箱型图 Box plot
  + 直方图 Histogram
  + 协方差Covariance, 相关性Correlation，R2
  + 散点图Scatter plot
    - 加回归线 regression line

1. 因果推断入门
   * 反事实 Counterfactual
   * 符号表达Notations
     + 潜在结果 potential outcome
     + 有条件的潜在结果 conditional potential outcome
   * 选择性偏差 selection bias
   * 随机对照实验RCT
2. RCT+二变量回归模型Bivariate OLS regression
   * 为啥叫“回归”
   * 中心极限定理 central limit theorem
   * 假说检验 Hypothesis testing
     + 单尾/双尾
   * 标准误差 Standard error
   * T检验 (T-test)
   * 显著性 significance level
   * P值 p-value
   * 置信区间 confidence interval
3. 多变量OLS的使用
   * 遗漏变量偏差 (omitted variable bias)
     + 数学推导
   * R2 vs adjusted R2
   * OLS假设
   * 多重共线性 multicollinearity
     + 导致的标准误差变大
     + 数学解释
     + 虚拟变量陷阱（Dummy variable trap）
   * R的线性回归使用
   * 复习T检验
   * 联合假设 – F检验
4. OLS/社科量化研究的问题

* 有效性 validity
  + 外部
  + 内部
* 可靠性 Reliability
* 社会偏好偏差 Social Desirability bias
* Hawthorne Effect
* 缺失偏差 Attrition bias
* 代表性 Representativeness
* 内生性Endogeneity
  + OVB
  + 错误形式 Misspecification
    - 异质影响 heterogeneous effect (交互项 interaction)
    - 非线性关系
      * 对数 logarithm
      * 多项式 polynomial
  + 测量误差 measurement error
  + 反向因果Reverse causality
  + 抽样误差
* 同方差性/异方差性 homoskedasticity/heteroskedasticity
  + 稳健标准差Robust standard errors
* 自相关 autocorrelation
* OLS vs 其他方法

(GV249 W7)

1. 面板数据Panel data+双重差分法DID
2. 合成控制Synthetic Control+合成双重差分SDID
3. 工具变量IV
4. 断点回归RDD
5. 研究纪律和研究道德

* 出版偏差 publication bias
  + Pre-analysis plan
  + Pre-registration
* 可复现性 replicability
* P-hacking/研究者自由度researcher degrees of freedom
* 代码可阅读性
* 涉及人类主体human subject参与
  + 匿名性
  + 道德审核