Measurement of Policy Outcomes

Pre-Exam Summary

Jessica Mahoney, Lukas Puschnig, and Artur Obminski

January 10, 2019

You can find the latest version of this document here.

1 Tutorial Notes 3
- 1.1 Tutorial 1: Introduction — Design effect and sampling . . . 3
- 1.2 Tutorial 2: Health . . . 4
- 1.3 Tutorial 3: Income and poverty . . . 6
- 1.4 Tutorial 4: Inequality . . . 9
- 1.5 Tutorial 5: Employment and unemployment . . . 11
- 1.6 Tutorial 6: Education . . . 14

2 Definitions and Formulas 16

Introductory Note with Disclaimers

Dear classmates and PSE students,

This document is aimed to help you prepare for the exam in the Measurement of Policy Outcomes course. It is, however, due to time restrictions, limited in scope. The document focuses on summaries of the course tutorials, as well as definitions and formulas that we assume may be relevant for the exam. **The summary does not and will not cover all of the material.**

Because of time constraints, the summary was first shared in an incomplete state and is gradually built upon. **It may never reach completion.** You can find the latest version here. We hope that the summary may still prove useful.

Good luck with your studies!

Best,
Jessica, Lukas, and Artur

Change Log

January 10, 2019: Error corrected on Page 3. m is the number of students per PSU/-cluster. (Thanks, Lukas!)

January 8, 2019: Updated with summaries of remaining tutorials (3–6), on remaining homework assignments (3–6).

January 7, 2019: First version. Includes summaries of material assumed to be exam relevant from tutorials 1–3, on homework assignments 1–3.

1 Tutorial Notes

1.1 Tutorial 1: Introduction — Design effect and sampling

- Two-stage sampling
  - $Var(\bar{X}) = \text{deff} \times Var(X)/n$
    - Design effect (deff) $= 1 + (m-1) \times \rho$
    - m = Number of students per PSU/cluster; ρ = Correlation between observations within a
  - Two-stage sampling always decreases precision of estimators (deff ≥ 1)
  - On average, sample less likely to be representative with two-stage sampling
  - The more clusters are representative of overall population (lower correlation between observations within PSUs), the less problematic (smaller correlation coefficient)
  - Keeping total n constant, a larger m implies more respondents per PSU and getting further away from random sampling
    - Equal probability of schools being drawn $\longrightarrow$ Unequal probability of students drawn (if school size is different)

1.2 Tutorial 2: Health

- Weighted sampling
  - Advantages
    - Accounting for embedded unequal probabilities of being sampled
      - Weights inverse of probability of being sampled (overweighting low sampling probability observations)
    - Increasing precision on a particular group
      - e.g. Increasing precision within a certain sparsely populated region
    - Compensating for non-response
      - e.g. Rich people have especially low response rates, so overweighted based on proportion of rich in the population
  - Methods
    - Ex ante
      1. Stratification (use all strata, with proportions) or clustering (first randomly select strata)
      2. Extra cases in some strata to increase the precision on these groups
    - Ex post
      1. Classify the sample by socio-demographic categories
      2. Compare with the distribution in the true population (census data, administrative data...)
- Testing for different between coefficients
  - If confidence intervals don't overlap $\longrightarrow$ Statistically significant difference
  - If confidence intervals overlap, but neither estimate is within the confidence interval of the other estimate $\longrightarrow$ Need for t-test
  - If one estimate is within the confidence interval of the other estimate $\longrightarrow$ NOT statistically significant difference
- Measuring mortality
  - Measuring on 0–4-year-olds (directly on the group of interest)
    - Not observing deaths happening after time of survey $\longrightarrow$ Always underestimating true mortality rate

    - Bias corrected by assuming that death patterns observed today for each age group will hold for the future death patters of the younger cohorts of interest
- Measuring on 5–9-year-olds still alive today (lagged)
    - More accurate account of death toll
    - Assumes that lagged and current cohorts exposed to same environment
    - Noisier measure, with risk of bad recall
    - Delayed data

1.3 Tutorial 3: Income and poverty

- Types of poverty measures
  - Objective v.s. subjective
  - Monetary v.s. non-monetary (e.g. access to services)
  - Absolute v.s. relative
  - Poor households v.s. poor individuals v.s. poor countries
- U.S. defined poverty
  - Includes money income (earnings and cash benefits) — not non-cash benefits (e.g. food stamps)
  - Gross income, before taxes
  - Poverty thresholds:
    - Resources necessary to afford the cost of basic needs (food and non-food component)
    - Food component: Computed separately and anchored to nutritional requirements
    - Non-food component: Food share estimated on all families of more than 3 people based on a survey
    - Poverty threshold: Food basket $\times$ (1 / Average food share in household expenditures)
      - Food component is absolute (nutritional needs)
      - Non-food factor is relative (it depends on the consumption of all)
      - If families of a certain composition are better off, on average, than families of another composition, brings the food share down because of income effect $\longrightarrow$ Poverty threshold up
- EU/Eurostat defined poverty
  - 60% of median
    - Median more robust measure than mean
  - Left unchanged by a multiplication of all incomes, sensitive to absolute increases

  - Used to compare EU countries, but new EU countries on average much poorer (in absolute terms)

- U.S. v.s. EU/Eurostat definition
  - Absolute v.s. relative poverty
  - Gross v.s. disposable income (and definitions)
  - Differences in accounting for household size

- Equivalized household size
  - Take into account decreasing marginal cost of additional family units, lower marginal costs of children
  - Problems
    * Not context dependent — e.g. whether education is free or paid for affects the cost of children
    * Assume that budget is common to all household members — in practice often not true
    * Age cutoffs sometimes somewhat arbitrary
  - Oxford (1/0.7/0.5) vs. OECD (1/0.5/0.3) scale:
    * Large households relatively poorer (lower economies of scale) with Oxford
    * Poverty rates affected differently depending on countries' demographic distribution — difficult to predict

- Imputed rents: Rents that would have been paid if not owning one's house/flat
  - Make countries with large ownership v.s. large tenant base comparable
    * To make taxation more fair, avoiding disincentivizing geographical mobility though less taxation living in own house than receiving taxed rent payments from tenant of owned house while one lives somewhere else

- Subjective poverty
  - Welfare state may decrease subjective poverty, giving citizens ability to face unexpected expenses
  - Difficult to compare continuous subjective poverty scales, e.g. 1–6, to binary poverty measures

  - Measures used may be culturally/societally dependent — e.g. vacations culturally more valued in some countries

- Poverty gap
  - Measuring individuals' distance from (below) the poverty threshold

1.4 Tutorial 4: Inequality

- Importance of question of inequality
  - Social cohesion
  - Growth consequences
  - Social mobility, opportunities
  - Tax policy, redistribution
- Conflicting interests with regards to inequality
  - Self-interest, loss aversion
  - Values/objectives
  - Belief systems (e.g. efficiencies, causal mechanism like meritocracy)
- Economists agnostic on choice of welfare function to maximize (e.g. Utilitarian, Rawls' Maximin)
- Inequality measurement problems
  - Top and bottom consumption/income levels less likely to be observed
  - Inequality increasing with measurement errors — even if errors are random
  - Inequality measures sensitive to extreme values
    - If measurement errors, should be excluded
    - If true values, depends on interest (e.g. if social cohesion is the aim, maybe extreme values are irrelevant)
  - Level measure typically positively (rightward) skewed (mass shifted to the left, long tail to the right)
    - Logarithm typically more resembling normal distribution
- Axiomatics of welfare function
  - Anonymity: Changing the position of persons doe not change what can say about welfare
  - Pareto principle: If person can be better off with a change without lowering the welfare of any other person, this new distribution leads to strictly higher welfare and is preferred to the old distribution
- Axiomatics of Lorenz Curve

- Anonymity
- Pigou-Dalton Principle: Allocation making the distribution more equitable preferred — transferring from rich to poor should decrease inequality
- Relativity: Multiplying all observations by a constant does not change the inequality measure

1.5 Tutorial 5: Employment and unemployment

- Moments
  - First moment: Mean, μ
    - Estimated with $\hat{\mu} = \frac{1}{N}\sum X_i$
  - Second moment: Variance, σ^2
    - Estimated with $\hat{\sigma}^2 = \frac{1}{N-1}\sum(X_i - \bar{X})^2$
    - Variance of sum and sum of variance same if independently distributed
- Notation
  - P: Population
  - W: Working age population
  - L: Labor force
  - U: Unemployed
  - E: Employed
  - I: Inactive
  - u: Unemployment rate
  - e: Employment rate
  - Y: GDP / National income
- Relations and definitions
  - $W = L + I$
  - $L = U + E$
  - $u = U/L = U/(U + E)$
  - $e = E/W = E/(U + E + I)$
  - $$\underbrace{\frac{Y}{P}}_{\text{GDP per capita}} = \underbrace{\frac{W}{P}}_{\text{Support ratio}} \times \underbrace{\frac{L}{W}}_{\text{Labor force participation}} \times \underbrace{\frac{E}{L}}_{1-u} \times \underbrace{\frac{Y}{E}}_{\text{Per-worker productivity}}$$
- Employment
  - Employed if 15–74 years old who:
    - Has done at least one hour of paid work during reference week, or of unpaid work for a relative's business

    - OR has not worked during reference week but...
        - Has job and is on paid vacation, parental leave, part-time job, training course paid by the employed, technical unemployment, strike
        - Sick leave for less than 1 year
        - Parental leave or other leave, unpaid by the employer for less than three months
    - OR 75+ years old with a job

- Unemployment
  - NOT employed AND...
    - Has used at least one active method of job search
    - OR is waiting to start a job in less than 3 months
    - OR is temporarily away from work but will start working again in less than 3 months
  - AND: Is able to start within 2 weeks

- Interpreting ILO definitions
  - Problem of officiality: Formal attachment to a job? (e.g. go to same workplace each day)
  - Problem of arbitrariness: Duration of maternity leave, sick leave, holidays, strike
  - Problem of participation quality: Nature of internships — similar to normal employment or not?

- Other problems with definitions of unemployment
  - "Unemployment" while looking for a suitable job — maybe better whan working in a poorly fitted job
  - Active choice of leisure rather than work
  - Quality of employment
    - Precarious working conditions
    - Overqualification
    - Underemployment
  - Inconsistencies in classifications

    - Underemployed by force: Employed
    - Unemployed by force: Inactive
  - Actively looking for job criterion
    - Discouraged non-workers should be counted as unemployed, not inactive
    - Only one active means of job search could be considered insufficient
  - Availability to work criterion
    - Sickness
    - Disability — excluded from labor force? what about disabled who do work?
  - Working age population
    - Incarcerated people, people in institutions not counted and surveyed
- Measures of depth of unemployment
  - Underemployment
  - Long term unemployment
  - Spells of unemployment: Duration for different categories
  - Risk of unemployment

1.6 Tutorial 6: Education

- Transition matrices and relative changes
  - Using z-scores, average change = 0 by definition, unless differential attrition (e.g. weak students missing in later survey)
- Rash model
  - Ability measure continuous, answers discrete
  - Assumes logistic relationship between probability of answering correctly and ability
  - $P(X_ij = 1 \mid \beta_i, \delta_j) = \frac{exp(\beta_i - \delta_j)}{1+exp(\beta_i - \delta_j)}$
    - ...where β_i is ability of student i and δ_j is difficulty of question j
  - Steps
    - Identify item parameter by comparing rate of correct answers to other items
    - Model probability of any response pattern
    - Derive from the response pattern an estimate of student ability (maximum likelihood value in Rasch's initial approach, 5 randomly drawn plausible values in OECD approach)
- OECD PISA
  - Measures *abilities* rather than system-specific skills
  - Sampling done at age 15
  - Sampling design
    - Two-stage sampling
    - Schools drawn proportionately to size
    - Minimum of 150 schools, each school draws 30 students
    - Probability of student being drawn same across schools, except for students in schools with fewer than 35 15-year-olds, then all students take the test
  - Using Rash model, because not all students asked the same questions
  - Problems

* No missing plausible values, substantial number of missing values for raw scores — PVs may be unreliable
* Sampling procedures may not be comparable between countries (excluding up to 5% of schools in a country allowed)
* Age 15 may imply different schooling levels in different countries
* Schooling participation rates at age 15 may differ between countries, in a non-random fashion
* Definition of "school" can vary — high schools with regular and vocational tracks can be one or two schools
* Differences between countries and over time often statistically insignificant
* Maybe other measures than means, e.g. spreads/inequality, more/also important
* Maybe other subjects than those measured also important

2 Definitions and Formulas

TBC.