Decomposition methods in the social sciences

Bamberg Graduate School of Social Sciences, June 7–8, 2018

Ben Jann

University of Bern, Institut of Sociology

Introduction and course overview

Contents

Introduction

- Decomposition methods are used to analyze **distributional differences** in an outcome variable **between groups** or time points.
- In particular, the methods decompose the observed difference into a components that is due to **compositional differences** between the groups, and a component that is due to **differential mechanisms**.
- Example question: How can the difference in average wages between men and women (the gender wage gap) be explained? Is the difference due to ...
  - ...group differences in wage determinants (i.e. in characteristics that are relevant for wages, such as education)? (compositional differences)
  - ...differential compensation for these determinants (e.g. different returns to education for men an women, or wage discrimination against women)? (differential mechanisms)

Introduction

- Similar questions can be asked in **other contexts** (for different groups, for different outcome variables) or also when analyzing changes **over time**.
- Example question: How can the increase or earnings inequality (e.g. measured by the Gini coefficient or the D9/D1 ratio) be explained? Is the increase due to ...
    - ...changes in the distribution of characteristics that determine earnings? (changes in composition)
    - ...changes in how these characteristics affect earnings? (changes in mechanisms)

Introduction

- Conceptually, decomposition methods are closely related to the **counterfactual** model of causality (although results from decomposition analyses are often not interpreted causally).
- Hence, estimation techniques from the causal inference literature (e.g. matching or inverse probability weighting) can sometimes be useful for decomposition analyses.
- In essence, decomposition methods work by creating counterfactuals such as "How would the outcome distribution in group A look like if it had the same distribution of determinants as group B?"
- Example questions:
  - How high would the mortality rate in country A be if it had the demographic composition of country B?
  - How do average test scores between different schools compare after taking into account the socio-economic composition of the schools' pupils?

Historic development

- Decomposition methods have their origins in the seminal works of Oaxaca (1973) and Blinder (1973), who analyzed mean wage differences between groups (males vs. females, whites vs. blacks).
  - An even earlier reference is Winsborough and Dickenson (1971).
  - Similar methods have also been developed in other disciplines (for example, direct and indirect standardization in demography/epidemiology; see Kitagawa 1955, Das Gupta 1978).
- Pronounced increase in US earnings inequality since the end of the 1970s fostered various methodological innovations in labor economics since the mid 1990s.
- These more recent developments focus on topics such as ...
  - ... providing methods for distributional measures other than the mean
  - ... taking account of non-linearities and providing methods for categorical and other types of variables
  - ... clarifying the basic assumptions made by these procedures
  - ... solving statistical inference

This workshop

- This workshop will give a detailed introduction to the basic Oaxaca-Blinder (OB) decomposition and also covers some of the newer developments.
- We will focus on **counterfactual decompositions** in the sense described above. There are also other types of decompositions, such as the factor decomposition of inequality measures, that will not be covered.

Readings

- A comprehensive (but for applied researcher sometimes not very accessible) review is provided by:
    - Fortin, Nicole, Thomas Lemieux, Sergio Firpo (2011). Decomposition Methods in Economics. Pp. 1–102 in: O. Ashenfelter and D. Card (eds.). Handbook of Labor Economics. Amsterdam: Elsevier.
- For a concise and easy to understand overview of the basic Oaxaca-Blinder decomposition, see:
    - Jann, Ben (2008). The Blinder-Oaxaca decomposition for linear regression models. The Stata Journal 8(4):453–479.
- Apart from that: read the specialized literature on the various decomposition procedures. References will be provided throughout the slides.

Course overview

- Conceptually, the course is divided into two thematic blocks.
- Block I: Oaxaca-Blinder decomposition
  - Basics; post-estimation, tables and graphs; problems such as the index problem, arbitrary transformations, and the base category problem; extensions to non-linear models; difference-in-difference decompositions
- Block II: Beyond the mean
  - Reweighting, Juhn-Murphy-Pierce 1993, Machado-Mata, distributional regression, RIF-Regression
- Each block contains several cycles of theoretical inputs, examples, and exercises (using Stata).

Distributions of random variables

- Random variables X, Y, Z ...
- The variables can be:
  - continuous: any value within an given interval is possible
  - discrete: only a fixed set of values are possible
- Discrete variables are often categorical in the sense that distances between values have no meaning (e.g. group membership).
- Distribution function $F_Y(y)$ (CDF)
  - displays the probability that Y is smaller than or equal to some given value y

$$F_Y(y) = \Pr(Y \leq y)$$

- Density function $f_Y(y)$ (PDF)
  - displays how the probability mass is distributed along the values of Y
  - for continuous variables, the PDF is defined as

$$f_Y(y) = F'_Y(y) \quad \text{such that} \quad F_Y(y) = \int_{-\infty}^{y} f_Y(z)\, dz$$

Distributions of random variables

- the integral of the PDF between two points is equal to the probability mass between these two points

$$\Pr(a \leq Y \leq b) = \int_a^b f_Y(y)\,dy$$

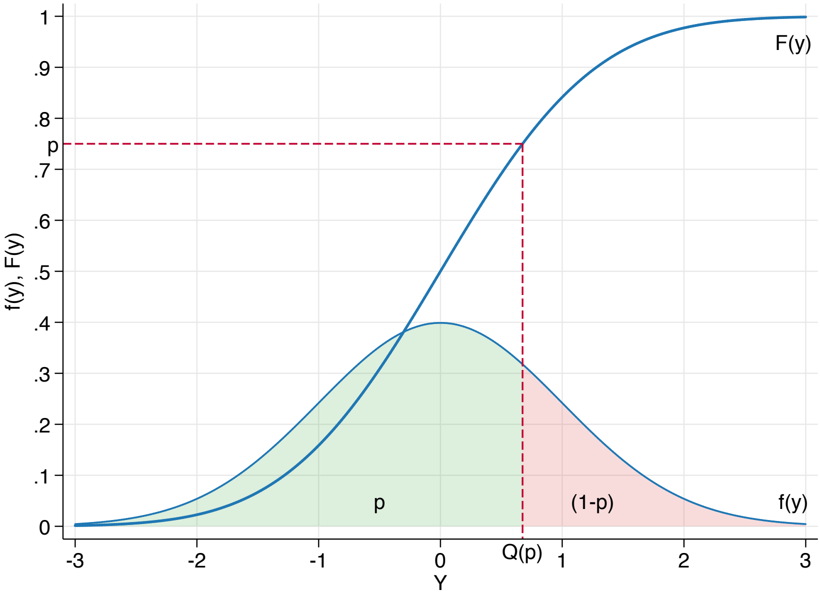
- for discrete variables with possible values $y_1, y_2, \ldots, f_Y(y)$ is a probability mass function

$$f_Y(y) = \Pr(Y = y) \quad \text{such that} \quad F_Y(y) = \sum_{y_i \leq y} \Pr(Y = y_i)$$

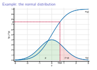
- below I will sometimes use integrals of $f_Y(y)$ even if Y is a discrete variable; this is an abuse of notation for sake of simplicity; think of the integral being a sum in these cases

- Quantile function $Q_Y(p)$
  - is equal to the value of Y for which $\Pr(Y \leq y)$ is equal to p
  - $Q_Y(p)$ is the inverse of the distribution function: $Q_Y(p) = F_Y^{-1}(p)$

Example: the normal distribution

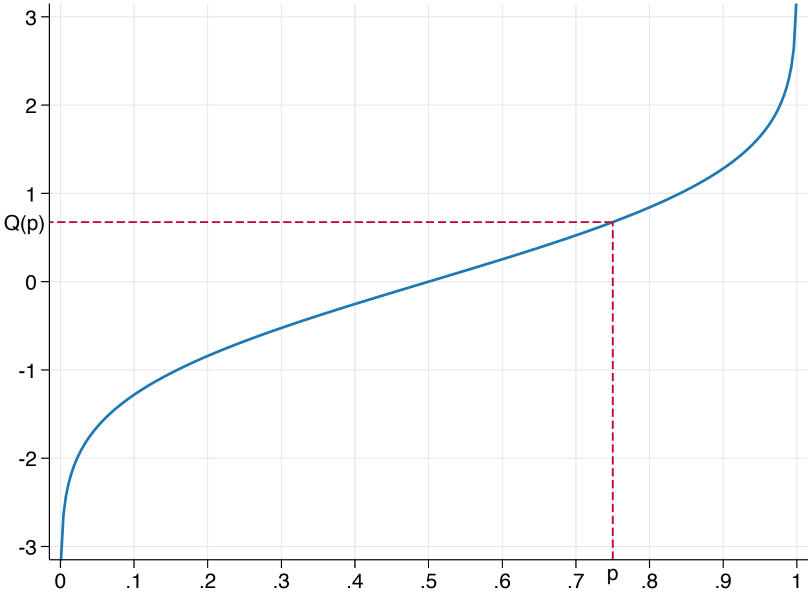


Example: the normal distribution

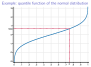


```
local p = .75
local Q75 = invnormal(`p')
twoway (function normalden(x), range(-3 `Q75') psty(p3) recast(area) lcolor(%0) fcolor(%20)) //
       (function normalden(x), range(`Q75' 3) psty(p4) recast(area) lcolor(%0) fcolor(%20)) ///
       (function normalden(x), range(-3 3) psty(p1)) ///
       (function normal(x), range(-3 3) psty(p1) lw(*1.5)) ///
       (pci `p' -3.1 `p' `Q75'  `p' `Q75' -.025 `Q75', lsty(xyline) lp(-)) ///
    , xlabel(#10) ylabel(#10) yti("f(y), F(y)") legend(off) plotr(margin(zero)) ///
    xti(Y) xscale(range(-3.1 3.1)) yscale(range(-.025 1.025)) ///
    ymlabel(`p' "p", add notick nogrid labsize(medsmall)) ///
    xmlabel(`Q75' "Q(p)", add notick nogrid labsize(medsmall)) ///
    text(.05 2.9 "f(y)" .95 2.9 "F(y)" 0.05 -0.5 "p" 0.05 1.25 "(1-p)")
```

Example: quantile function of the normal distribution



Example: quantile function of the normal distribution




```
local p = .75
local Q75 = invnormal(`p')
local ll = 1-normal(3.15)
local ul = normal(3.15)
twoway (function invnormal(x), range(`ll' `ul') psty(p1) lw(*1.5)) ///
       (pci -3.15 `p' `Q75' `p'  `Q75' `p' `Q75' -.015, lsty(xyline) lp(-)) ///
    , xlabel(#10) ylabel(#10) yti("") legend(off) plotr(margin(zero)) ///
    xti("") yscale(range(-3.15 3.15)) xscale(range(-.015 1.015)) ///
    xmlabel(`p' "p", add notick nogrid labsize(medsmall)) ///
    ymlabel(`Q75' "Q(p)", add notick nogrid labsize(medsmall))
```

Expected value and variance

- Expected value of Y ("the mean")

$$\mathrm{E}(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy$$

- Variance of Y

$$\mathrm{Var}(Y) = \mathrm{E}((Y - \mathrm{E}(Y))^2) = \mathrm{E}(Y^2) - \mathrm{E}(Y)^2$$

- Some useful relations (see, e.g., Mood et al. 1974)
  - $\mathrm{E}(a + bY) = a + b\,\mathrm{E}(Y)$
  - $\mathrm{Var}(a + bY) = b^2\,\mathrm{Var}(Y)$
  - $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$
  - $\mathrm{Var}(\sum_i a_i Y_i) = \sum_i \sum_j a_i a_j\,\mathrm{Cov}(Y_i, Y_j)$
    Special cases:
    - $\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y)$
    - $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X, Y)$
    - $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\,\mathrm{Cov}(X, Y)$

    If X and Y are independent:
    - $\mathrm{Var}(X + Y) = \mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

Conditional distributions

- Conditional distribution $F_{Y|X}(y|x)$
  - the distribution of Y given that $X = x$ (e.g. distribution in a subpopulation)
  - $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$ where $f_{X,Y}(x,y)$ is the joint density of X and Y and $f_X(x)$ is the marginal density of X
  - hence: $f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x)$

- Relationship between marginal and conditional distributions

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)\,dx$$

$$\begin{aligned} F_Y(y) &= \int_{-\infty}^{y} f_Y(z)\,dz = \int_{-\infty}^{y}\left\{\int_{-\infty}^{\infty} f_{X,Y}(x,z)\,dx\right\}dz \\ &= \int_{-\infty}^{y}\left\{\int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)\,dx\right\}dz \\ &= \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{y} f_{Y|X}(y|x)\,dz\right\}f_X(x)\,dx \end{aligned}$$

Conditional distribution

- Law of total/iterated expectation (follows from the relation above)

$$\mathsf{E}(Y) = \int_{-\infty}^{\infty} \mathsf{E}(Y|X = x) f_X(x)\, dx = \mathsf{E}(\mathsf{E}(Y|X = x))$$

- But note: the law does not hold for quantiles

$$Q_Y(p) \neq \int_{-\infty}^{\infty} Q_{Y|X}(p|x) f_X(x)\, dx = \mathsf{E}(Q_{Y|X}(p|x))$$

That is, the marginal quantile is not equal to the expectation of corresponding conditional quantiles.

General idea of decomposition methods

- There are two groups, $g = \{0, 1\}$, e.g. males and females. Variable G denotes group membership, e.g. $G = 0$ if male and $G = 1$ if female.
- Of interest is the overall difference between the groups with respect to some functional ν of Y (e.g. the mean, variance, or Gini coefficient):

$$\Delta^{\nu} = \nu\big(F_{Y|G=0}\big) - \nu\big(F_{Y|G=1}\big)$$

  where $F_{Y|G=g}$ is shorthand notation for $F_{Y|G}(y|g)$.
- Assume that Y is determined by covariates X and an error term ϵ:

$$Y = m^g(X, \epsilon)$$

  where $m^g()$ is some group-specific structural function. For example, $m()$ can be a linear function as in linear regression:

$$m(X, \epsilon) = \beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K + \epsilon$$

  However, $m()$ can also be something much more complicated.

General idea of decomposition methods

- Furthermore, assume (for now) that there are no distributional differences in ϵ between the groups. The overall difference Δ^ν can then be due to:
  - group differences in the distribution of X
  - group differences in structural function $m()$
- The goal of decomposition methods now is to partition the overall difference into these components:

$$\Delta^\nu = \Delta^\nu_X + \Delta^\nu_S$$

where Δ^ν_X denotes the component due to differences in the distribution of X and Δ^ν_S denotes the component due to differences in $m()$.

General idea of decomposition methods

- Potential outcomes (note the close relation to the counterfactual model of causality):
  - given are the *potential* outcomes $Y^0 = m^0(X, \epsilon)$ and $Y^1 = m^1(X, \epsilon)$
  - for the observed Y we have

$$Y = \begin{cases} Y^0 & \text{if } G = 0 \\ Y^1 & \text{if } G = 1 \end{cases}$$

  - that is, potential outcome Y^0 is observed in group 0 and potential outcome Y^1 is observed in group 1
  - for decompositions, however, we are also interested in the (unobserved) counterfactuals
    - ★ How would Y look like in group 0 if it would be generated by $m^1()$ instead of $m^0()$ (and vice versa)?
    - ★ Example: How much would a man with given characteristics earn if he were payed like a women?

General idea of decomposition methods

- That is, to identify Δ^ν_X and Δ^ν_S, we need a counterfactual distribution $F_{Y^g|G\neq g}$. I use red color to emphasize counterfactuals.
- For example, let $F_{Y^0|G=1}$ be the counterfactual distribution of Y in group 1 if we assume that Y is determined in group 1 according to group 0's structural function $m^0()$.
- By adding and subtracting $\nu\big(F_{Y^0|G=1}\big)$, the decomposition can then be written as:

$$\begin{aligned}
\Delta^\nu &= \nu\big(F_{Y|G=0}\big) - \nu\big(F_{Y|G=1}\big) \\
&= \big\{\nu\big(F_{Y|G=0}\big) - \nu\big(F_{Y^0|G=1}\big)\big\} + \big\{\nu\big(F_{Y^0|G=1}\big) - \nu\big(F_{Y|G=1}\big)\big\} \\
&= \Delta^\nu_X + \Delta^\nu_S
\end{aligned}$$

General idea of decomposition methods

- The main goal of different decomposition methods is to find good ways to determine $\nu\big(F_{Y^0|G=1}\big)$ or, likewise, $\nu\big(F_{Y^1|G=0}\big)$, so that Δ_X^ν and Δ_S^ν can be estimated.
- Depending on context, further distinctions are made, e.g. by allowing the distribution of ϵ to differ by groups (in which case the decomposition has a third component Δ_ϵ^ν) or by partitioning $m()$ into a part related to X and a part related to ϵ (to subdivide Δ_S^ν into a component related to observables and a component related to unobservables).
- Similarly, detailed decompositions that separate the contributions of the different X variables are often of interest.
  - Example: how much of the gender wage gap can be explained by differences in education, how much by differences in work experience?

References

- Blinder, Alan S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. The Journal of Human Resources 8(4):436–455.
- Das Gupta, Prithwis (1978). A General Method of Decomposing a Difference Between Two Rates into Several Components. Demography 15(1):99–112.
- Fortin, Nicole, Thomas Lemieux, Sergio Firpo (2011). Decomposition Methods in Economics. Pp. 1–102 in: O. Ashenfelter and D. Card (eds.). Handbook of Labor Economics. Amsterdam: Elsevier.
- Jann, Ben (2008). The Blinder-Oaxaca decomposition for linear regression models. The Stata Journal 8(4):453–479.
- Kitagawa, Evelyn M. (1955). Components of a Difference Between Two Rates. Journal of the American Statistical Association 50(272): 1168–1194.
- Mood, Alexander M., Franklin A. Graybill, Duane C. Boes (1974). Introduction to the Theory of Statistics. New York: McGraw-Hill.
- Oaxaca, Ronald (1973). Male-Female Wage Differentials in Urban Labor Markets. International Economic Review 14(3):693–709.
- Winsborough, H. H., Peter Dickinson (1971). Components of Negro-White Income Differences. Pp. 6–8 in: Proceedings of the Social Statistics Section. Washington, DC: American Statistical Association.