Decompositions

Mind the Gap

When faced with a gap in mean outcomes between two groups, researchers frequently examine how much of the gap can be explained by differences in observable characteristics.

The simple approach is to estimate the pooled regression including an indicator variable for group membership as well as the other observable characteristics, interpreting the coefficient on the group indicator as the unexplained component.

The Oaxaca-Blinder (O-B) decomposition represents an alternative approach.

Oaxaca-Blinder decomposition

Consider a categorical (or dummy) variable d that splits our dataset into two groups.

In this case we can run regressions of the form $y = X\beta + \epsilon$ to estimate the the mean difference between groups, as follows

$$
\begin{aligned}
\mathbb{E}\left[y^0\right] &= \mathbb{E}\left[X^{(0)}\right]\beta_0;\text{ group } d=0\\
\mathbb{E}\left[y^1\right] &= \mathbb{E}\left[X^{(1)}\right]\beta_1;\text{ group } d=1
\end{aligned}
$$

Alternatively

$$
\begin{aligned}
\bar{y}_0 &= \bar{X}_0\beta_0;\text{ group } d=0\\
\bar{y}_1 &= \bar{X}_1\beta_1;\text{ group } d=1
\end{aligned}
$$

Then the mean difference in outcomes is:

$$
\begin{aligned}
\mathbb{E}\left[y^1\right]-\mathbb{E}\left[y^0\right] &= \mathbb{E}\left[X^{(1)}\right]\beta_1-\mathbb{E}\left[X^{(0)}\right]\beta_0 \\
&= \left(\mathbb{E}\left[X^{(1)}\right]-\mathbb{E}\left[X^{(0)}\right]\right)\beta_1+\mathbb{E}\left[X^{(0)}\right](\beta_1-\beta_0) \\
&= \left(\mathbb{E}\left[X^{(1)}\right]-\mathbb{E}\left[X^{(0)}\right]\right)\beta_0-\mathbb{E}\left[X^{(1)}\right](\beta_0-\beta_1)
\end{aligned}
$$

Alternatively

$$
\begin{aligned}
\bar{y}_1-\bar{y}_0 &= \bar{X}_1\beta_1-\bar{X}_0\beta_0 \qquad (1)\\
&= \left(\bar{X}_1-\bar{X}_0\right)\beta_1+\bar{X}_0(\beta_1-\beta_0) \\
&= \left(\bar{X}_1-\bar{X}_0\right)\beta_0-\bar{X}_1(\beta_0-\beta_1)
\end{aligned}
$$

Where: $\left(\bar{X}_1-\bar{X}_0\right)\beta_1$ is the "explained" component (differences in characteristics), and $\bar{X}_0(\beta_1-\beta_0))$ is the "unexplained" component (differences in returns to characteristics), where $\bar{X}_0$ is the baseline (this decomposition is not unique, as you can see from the second and third lines of equation (1).

We define

$$
\begin{aligned}
\text{Gap}^1 &= \bar{X}_0(\beta_1-\beta_0)\,;\ \bar{X}_0 \text{ is the baseline} \\
\text{Gap}^0 &= \bar{X}_1(\beta_0-\beta_1)\,;\ \bar{X}_1 \text{ is the baseline} \\
\text{Gap}^{\text{OLS}} &= \delta_d;\ \text{where } y=\delta_0+d\delta_d+X\delta_1+\epsilon
\end{aligned}
$$

where, given the strong assumptions that the model is properly specified and that coefficients are equal across groups, a sensible definition of the population unexplained gap is δ_d.

We an define another measure Gap^p as

$$
\begin{aligned}
\bar{y}_1-\bar{y}_0 &= \left(\bar{X}_1-\bar{X}_0\right)\hat{\beta}_p+\text{Gap}^p \\
\text{Gap}^p &= \bar{X}_1\left(\hat{\beta}_1-\hat{\beta}_p\right)+\bar{X}_0\left(\hat{\beta}_p-\hat{\beta}_0\right)
\end{aligned}
$$

where $\hat{\beta}_p$ is the coefficient from the pooled regression of y on X (i.e. $y=X\beta_p+\epsilon$).

The Case in Which Coefficients Are Equal across Groups

Here we assume that the mean outcomes between groups 0 and 1 differ only by a constant and that the outcome is influenced by only the observable characteristic(s) x.

An O-B unexplained gap can always be written as the difference in overall mean outcomes minus the difference in predicted mean outcomes, and both of these differences can be denoted by linear projections.

A general expression for an O-B unexplained gap is

$$\begin{aligned}
\text{Gap} &= [\bar{y}_1 - \bar{y}_0] - \left[\hat{\beta}\left(\bar{x}_1 - \bar{x}_0\right)\right] \\
&= \text{b}(y|d) - \text{b}(\hat{\beta}x|d) \qquad (2)
\end{aligned}$$

where $\hat{\beta}$ is a coefficient computed from sample data and $\text{b}(z|w)$ is the slope from a regression of z on w and an intercept. The choice of $\hat{\beta}$ is what identifies the different O-B decompositions.

So equation (3) resolves to Gap^1 when $\hat{\beta}$ is the slope coefficient of a regression of y on x using the data from group 1, and gives Gap^0 when $\hat{\beta}$ is the slope coefficient using data from group 0[1].

$$\begin{aligned}
\text{Gap} &= \text{b}(y|d) - \text{b}(\hat{\beta}x|d) \\
&= \frac{\text{cov}\left(d, y\right)}{\text{var}(d)} - \hat{\beta}\frac{\text{cov}\left(x, d\right)}{\text{var}(d)} \qquad (3) \\
&= \frac{\text{cov}\left(d, \beta_0 + \beta_d d + \beta_x x\right) + \epsilon}{\text{var}(d)} - \hat{\beta}\frac{\text{cov}\left(d, x\right)}{\text{var}(d)} \\
&= \beta_d + \beta_x\frac{\text{cov}\left(d, x\right)}{\text{var}(d)} - \hat{\beta}\frac{\text{cov}\left(d, x\right)}{\text{var}(d)} \qquad (4) \\
&\to \beta_d, \text{ when } \beta_x = \hat{\beta}
\end{aligned}$$

given our assumption that the mean outcomes between groups 0 and 1 differ only by a constant and that the outcome is influenced by only the observable characteristic(s) x (and so $\hat{\beta} = \delta_x$ in both group-specific regressions), then all the gap expressions are equivalent, except for Gap^p.

Gap^p is different because in the pooled regression used to compute Gap^p, d is a omitted variable. In particular:

[1] under the data generation process described by $y = \delta_0 + \delta_d d + \delta_1 X + \epsilon$

$$
\begin{aligned}
\mathrm{Gap}^p &= \mathrm{b}(y|d) - \mathrm{b}(x\mathrm{b}(y|x)|d) \\
&= \frac{\operatorname{cov}(d, y)}{\operatorname{var}(d)} - \frac{\operatorname{cov}(d, [\operatorname{cov}(x, y)/\operatorname{var}(x)])}{\operatorname{var}(d)} \\
&= \frac{\operatorname{cov}(d, y)}{\operatorname{var}(d)} - \frac{\operatorname{cov}(d, x)}{\operatorname{var}(d)} \times \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)} \\
&= \frac{1}{\operatorname{var}(d)} \left(\operatorname{cov}(d, y) - \frac{\operatorname{cov}(d, x)\operatorname{cov}(x, y)}{\operatorname{var}(x)} \right)
\end{aligned} \tag{5}
$$

Compare this to $\mathrm{Gap}^{\mathrm{OLS}}$, defining $\tilde{z}(w)$ as the component of z that is orthogonal to w in the population (so that $\tilde{z}(w) = z - w\mathrm{b}(z|w)$, alternatively $\tilde{z}(w) = z - w\frac{\operatorname{cov}(z,w)}{\operatorname{var}(w)}$)[2]

$$
\begin{aligned}
\mathrm{Gap}^{\mathrm{OLS}} &= \delta_d \\
&= \frac{\operatorname{cov}(\tilde{d}(x), \tilde{y}(x))}{\operatorname{var}(\tilde{d}(x))};\ \text{per FWL} \\
&= \frac{\operatorname{cov}(d, \tilde{y}(x))}{\operatorname{var}(\tilde{d}(x))} - \frac{\operatorname{cov}(x, \tilde{y}(x))}{\operatorname{var}(\tilde{d}(x))} \times \frac{\operatorname{cov}(d, x)}{\operatorname{var}(x)};\ \text{per definition of } \tilde{d}(x) \\
&= \frac{\operatorname{cov}(d, \tilde{y}(x))}{\operatorname{var}(\tilde{d}(x))};\ \text{per definition of } \tilde{y}(x),\ \operatorname{cov}(x, \tilde{y}(x)) = 0 \\
&= \frac{1}{\operatorname{var}(\tilde{d}(x))} \left(\operatorname{cov}(d, y) - \operatorname{cov}(d, x)\frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)} \right)
\end{aligned} \tag{6}
$$

and so $\mathrm{Gap}^p = \frac{\operatorname{var}(\tilde{d}(x))}{\operatorname{var}(d)}\mathrm{Gap}^{\mathrm{OLS}}$

The Case in Which Coefficients Vary across Groups

The relationship between Gap^p and GapOLS is exact and general.

Here we assume that the coefficient on x varies between the two groups, and that x is a scalar:

$$y = \delta_0 + d\delta_d + \delta_x x + \delta_{dx} dx + \epsilon$$

Based on equation (4) we have

[2] Note: $\tilde{d}(x) = d - P(x) = d - x\frac{\operatorname{cov}(x,d)}{\operatorname{var}(x)}$, the linear propensity, and $\tilde{y}(x) = y - x\frac{\operatorname{cov}(y,x)}{\operatorname{var}(x)}$, so the coefficient δ_d is $\frac{\operatorname{cov}(\tilde{d}(x), \tilde{y}(x))}{\operatorname{var}(\tilde{d}(x))}$.

$$\text{Gap}^1 = \frac{\text{cov}\,(d,y)}{\text{var}(d)} - \frac{\text{cov}\,(x,d)}{\text{var}(d)} \times \frac{\text{cov}\,(x,y|d=1)}{\text{var}(x|d=1)} \tag{7}$$

and

$$\text{Gap}^0 = \frac{\text{cov}\,(d,y)}{\text{var}(d)} - \frac{\text{cov}\,(x,d)}{\text{var}(d)} \times \frac{\text{cov}\,(x,y|d=0)}{\text{var}(x|d=0)} \tag{8}$$

and finally,

$$\text{Gap}^{\text{OLS}} = w_1 \text{Gap}^1 + w_0 \text{Gap}^0 \tag{9}$$

where the weights are:

$$w_1 = \frac{\mathbb{P}\,(d=1)\,\text{var}\,(x|d=1)}{\mathbb{P}\,(d=1)\,\text{var}\,(x|d=1) + \mathbb{P}\,(d=0)\,\text{var}\,(x|d=0)} \tag{10}$$

$$w_0 = \frac{\mathbb{P}\,(d=0)\,\text{var}\,(x|d=0)}{\mathbb{P}\,(d=1)\,\text{var}\,(x|d=1) + \mathbb{P}\,(d=0)\,\text{var}\,(x|d=0)} \tag{11}$$

Now writing $\text{Gap}^{\text{OLS}} = w_1 \text{Gap}^1 + w_0 \text{Gap}^0$ with $w_1 + w_0 = 1$ (equation (9)):

$$\begin{aligned} w_1 \text{Gap}^1 + w_0 \text{Gap}^0 &= w_1 \left[\frac{\text{cov}\,(d,y)}{\text{var}(d)} - \frac{\text{cov}\,(x,d)}{\text{var}(d)} \times \frac{\text{cov}\,(x,y|d=1)}{\text{var}(x|d=1)} \right] \\ &+ w_0 \left[\frac{\text{cov}\,(d,y)}{\text{var}(d)} - \frac{\text{cov}\,(x,d)}{\text{var}(d)} \times \frac{\text{cov}\,(x,y|d=0)}{\text{var}(x|d=0)} \right] \\ &= \frac{\text{cov}\,(d,y)}{\text{var}(d)} - \frac{\text{cov}\,(x,d)}{\text{var}(d)} \Pi \end{aligned}$$

where

$$\begin{aligned} \Pi = & \frac{\mathbb{P}\,(d=1)\,\text{var}\,(x|d=1)}{\mathbb{P}\,(d=1)\,\text{var}\,(x|d=1) + \mathbb{P}\,(d=0)\,\text{var}\,(x|d=0)} \frac{\text{cov}\,(x,y|d=1)}{\text{var}(x|d=1)} \\ + & \frac{\mathbb{P}\,(d=0)\,\text{var}\,(x|d=0)}{\mathbb{P}\,(d=1)\,\text{var}\,(x|d=1) + \mathbb{P}\,(d=0)\,\text{var}\,(x|d=0)} \frac{\text{cov}\,(x,y|d=0)}{\text{var}(x|d=0)} \end{aligned}$$

simplifying Π:

$$\begin{aligned}
\Pi &= \frac{\pi \text{var}\left(x|d=1\right)}{\pi \text{var}\left(x|d=1\right)+\left(1-\pi\right)\text{var}\left(x|d=0\right)}\frac{\text{cov}\left(x,y|d=1\right)}{\text{var}(x|d=1)} \\
&+ \frac{\left(1-\pi\right)\text{var}\left(x|d=0\right)}{\pi \text{var}\left(x|d=1\right)+\left(1-\pi\right)\text{var}\left(x|d=0\right)}\frac{\text{cov}\left(x,y|d=0\right)}{\text{var}(x|d=0)} \\
&= \frac{\pi \text{cov}\left(x,y|d=1\right)+\left(1-\pi\right)\text{cov}\left(x,y|d=0\right)}{\pi \text{var}\left(x|d=1\right)+\left(1-\pi\right)\text{var}\left(x|d=0\right)} \\
&= \frac{\text{cov}(x,y)-\frac{\text{cov}(d,x)\text{cov}(d,y)}{\text{var}(d)}}{\text{var}(x)-\frac{\text{cov}(d,x)^2}{\text{var}(d)}}
\end{aligned}$$

and so

$$\begin{aligned}
w_1\text{Gap}^1+w_0\text{Gap}^0 &= \frac{\text{cov}\left(d,y\right)}{\text{var}(d)}-\frac{\text{cov}\left(x,d\right)}{\text{var}(d)}\Pi \\
&= \frac{\text{cov}\left(d,y\right)}{\text{var}(d)}-\frac{\text{cov}\left(x,d\right)}{\text{var}(d)}\left(\frac{\text{cov}(x,y)-\frac{\text{cov}(d,x)\text{cov}(d,y)}{\text{var}(d)}}{\text{var}(x)-\frac{\text{cov}(d,x)^2}{\text{var}(d)}}\right) \\
&= \frac{\text{cov}\left(d,y\right)\text{var}(x)-\text{cov}\left(d,x\right)\text{cov}(x,y)}{\text{var}(x)\text{var}(d)-\text{cov}(d,x)^2}
\end{aligned}$$

Now we have

$$\text{Gap}^{\text{OLS}} = \frac{1}{\text{var}(\tilde{d}(x))}\left(\text{cov}(d,y)-\text{cov}(d,x)\frac{\text{cov}(x,y)}{\text{var}(x)}\right)$$

and since $\tilde{d}(x)$ represents the residuals from a population regression of d on x,

$$\begin{aligned}
\text{var}\left(\tilde{d}(x)\right) &= \text{var}\left(d-x\frac{\text{cov}\left(d,x\right)}{\text{var}\left(x\right)}\right) \\
&= \text{var}\left(d\right)-\frac{\text{cov}\left(d,x\right)^2}{\text{var}\left(x\right)}
\end{aligned}$$

and so

$$
\begin{aligned}
\text{Gap}^{\text{OLS}} &= \frac{1}{\text{var}(\tilde{d}(x))}\left(\text{cov}(d,y) - \text{cov}(d,x)\frac{\text{cov}(x,y)}{\text{var}(x)}\right) \\
&= \frac{\text{var}(x)}{\text{var}(x)\text{var}(d) - \text{cov}(d,x)^2}\left(\text{cov}(d,y) - \text{cov}(d,x)\frac{\text{cov}(x,y)}{\text{var}(x)}\right) \\
&= \frac{\text{cov}\left(d,y\right)\text{var}(x) - \text{cov}\left(d,x\right)\text{cov}(x,y)}{\text{var}(x)\text{var}(d) - \text{cov}(d,x)^2} \\
&= w_1 \text{Gap}^1 + w_0 \text{Gap}^0
\end{aligned}
$$

Appendix

The law of total variance, or variance decomposition formula is:

$$
\text{var}(X) = \mathbb{E}[\text{var}(X|Y)] + \text{var}(\mathbb{E}[X|Y])
$$

where the first term is the between-group variance and the second term is the within-group variance.

In our problem, the corresponding variance decomposition is

$$
\begin{aligned}
\text{var}(x) = (1-\pi)\left(\text{var}(x|d=0) + \left(\mathbb{E}\left[x|d=0\right] - \mathbb{E}\left[x\right]\right)^2\right) \\
+ \pi\left(\text{var}(x|d=1) + \left(\mathbb{E}\left[x|d=1\right] - \mathbb{E}\left[x\right]\right)^2\right)
\end{aligned}
$$

but we have

$$
\mathbb{E}\left[x\right] = \mathbb{E}\left[\mathbb{E}\left[x|y\right]\right] = \pi\mathbb{E}\left[x|d=1\right] + (1-\pi)\mathbb{E}\left[x|d=0\right]
$$

substituting the RHS for $\mathbb{E}\left[x\right]$ we have

$$
\begin{aligned}
\text{var}(x) = (1-\pi)\left(\text{var}(x|d=0) + \left(\mathbb{E}\left[x|d=1\right] - \mathbb{E}\left[x|d=0\right]\right)^2\pi^2\right) \\
+ \pi\left(\text{var}(x|d=1) + \left(\mathbb{E}\left[x|d=1\right] - \mathbb{E}\left[x|d=0\right]\right)^2(1-\pi)^2\right)
\end{aligned}
$$

but we have $\mathbb{E}\left[x|d=1\right] - \mathbb{E}\left[x|d=0\right] = \text{cov}(x,d)/\text{var}(d)$, and $\text{var}(d) = \pi(1-\pi)$, so

$$
\begin{aligned}
\operatorname{var}(x) &= (1-\pi)\operatorname{var}(x|d=0) + \pi\operatorname{var}(x|d=1) + \frac{\operatorname{cov}(d,x)^2}{\operatorname{var}(d)^2} \times \left[(1-\pi)\pi^2 + \pi(1-\pi)^2\right] \\
&= (1-\pi)\operatorname{var}(x|d=0) + \pi\operatorname{var}(x|d=1) + \frac{\operatorname{cov}(d,x)^2}{\operatorname{var}(d)^2} \times \left[(1-\pi)\operatorname{var}(d) + \pi\operatorname{var}(d)\right] \\
&= (1-\pi)\operatorname{var}(x|d=0) + \pi\operatorname{var}(x|d=1) + \frac{\operatorname{cov}(d,x)^2}{\operatorname{var}(d)}
\end{aligned}
$$

by the same logic

$$
\operatorname{cov}(x,y) = (1-\pi)\operatorname{cov}(x,y|d=0) + \pi\operatorname{cov}(x,y|d=1) + \frac{\operatorname{cov}(d,x)\operatorname{cov}(d,y)}{\operatorname{var}(d)}
$$

Gap weights

Example

data:

```r
# https://cran.r-project.org/web/packages/oaxaca/vignettes/oaxaca.pdf
# https://www.worldbank.org/content/dam/Worldbank/document/HDN/Health/HealthEquityCh12.pdf
# https://pmc.ncbi.nlm.nih.gov/articles/PMC8343972/
# https://www.sciencedirect.com/science/article/abs/pii/S0169721811004072
# https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2528391
# https://link.springer.com/article/10.1186/s12982-021-00100-9
# https://giacomovagni.com/blog/2023/oaxaca/
# https://journals.sagepub.com/doi/pdf/10.1177/1536867X0800800401
# https://ocw.mit.edu/courses/14-662-labor-economics-ii-spring-2015/resources/mit14_662s15_lecnotes1/

# Load HMDA data
data('PSID1982', package = "AER")
hmda <- PSID1982

# data("nswcps", package = "hettreatreg")

# Prepare data for analysis
# Looking at income differences between racial groups
lending_data <- hmda |>
  dplyr::mutate(
    minority = factor(race != "white"),
    log_income = log(income)
  ) |>
  dplyr::select(log_income, minority, education, hrat, ccred, mcred, pubrec)
```

```r
# Fit separate regressions
model_a <-
  lm(wage ~ education + experience, data = dat |> dplyr::filter(union=='yes'))
model_b <-
  lm(wage ~ education + experience, data = dat |> dplyr::filter(union=='no'))

# Get mean characteristics (including intercept)
```

```r
X_mean_a <-
  c(1,
    colMeans(
    dat |>
      dplyr::filter(union=='yes') |>
      dplyr::select(education, experience)
    )
  )
X_mean_b <-
  c(1,
    colMeans(
      dat |>
        dplyr::filter(union=='no') |>
        dplyr::select(education, experience)
    )
  )

# Get coefficients
beta_a <- coef(model_a)
beta_b <- coef(model_b)

# Calculate decomposition
tibble::tibble(
  explained = sum((X_mean_a - X_mean_b) * beta_a)
  , unexplained = sum(X_mean_b * (beta_a - beta_b))
  , total_gap <- explained + unexplained
)

# Perform Oaxaca-Blinder decomposition
decomp <-
  oaxaca::oaxaca(
    wage ~ education + experience | union
    , data =
      dat |>
      dplyr::mutate(union = dplyr::case_when(union=='yes'~0, TRUE ~1))
  )
```