

Reweighting

1. Repeat the reweighting decomposition from the slides for the *mean* and the *variance* of log wages between the private sector and the public sector (see section “Example analysis”), but use the “opposite” counterfactual. That is, compute a decomposition in which the distribution of X in the public sector is adjusted to the distribution observed in the private sector (on the slides, the distribution of X in the private sector was adjusted to the distribution observed in the public sector). In addition to schooling and experience, include occupational status as measured by the international socio-economic index (*isei*) as a predictor.

Try to find a good specification for the `logit` model including the ISEI variable. In the example on the slides, schooling, experience, experience squared, and all interactions were included in the model. That is, check whether interaction terms are also needed with respect to ISEI or whether a simple additive term suffices.

2. Optional: Try to replicate the decomposition using command `dstat`.

Hint: See slide 33 for an example. Command `dstat` does not directly compute a decomposition, but it can be used to compute reweighted statistics (i.e. the counterfactuals) by applying the `balance()` option (it is called “balance” because its purpose is to “balance” the covariates). Within `balance()` use `reference()` to specify the reference group (i.e. the group that provides the “reference” distribution of the covariates; the other group will be reweighted such that its covariate distribution is adjusted to the one observed in the reference group).

The unexplained part of the decomposition can be obtained directly by `dstat` by applying sub-option `contrast` within `over()`. In this case the difference between the reweighted result in one group and the raw result in the other group is reported (i.e., this answers the question of how large the difference would be if both groups had the same covariate distribution). Instead of using sub-option `contrast` you can also apply command `lincom` or `nlcom` after running `dstat`.

For the explained part you need to compare results from two calls to `dstat`, since the explained part is about comparing the raw result in a group with the results you get for that group if the group is reweighted. You can determine the explained part manually by copying the relevant results from the outputs of the two calls to `dstat`, but this will not provide you a standard error. To obtain the standard error use the approach based on RIFs (recentered influence functions) as in the example on slide 33. We will hear more about RIFs in the next session.

3. Optional: Wrap the analysis from task 1 into a small program and apply the bootstrap to compute standard errors and confidence intervals. How do the bootstrap standard errors compare to the analytic standard errors provided by `dstat`?

Hint: See solutions to previous exercises on how to write a program that can be bootstrapped.

4. Optional: Evaluate how results change if you use entropy balancing to compute the weights instead of IPW based on a `logit` model.

Hint: You can either use command `kmatch eb` or `dstat`. For `kmatch eb` see the example on the slides. Options `targets(2)` and `covariances` request that means, variances and covariances of the predictors are balanced. This is similar to using a logit model that includes all squares and interactions.

For `dstat` you can use `balance(eb:...)` rather than `balance(ipw:)`. `dstat` does not provide options such as `targets(2)` and `covariances`, but you can use factor-variable notation when

specifying the predictors (i.e., you can specify the variables in the same way as you specified them in the logit model; for example, including a squared variable is equivalent to balancing the variance, including an interaction is equivalent to balancing the covariance between two variables).

5. Optional: Compute a reweighted Oaxaca-Blinder decomposition to explain the public-private wage gap and interpret the results. Include schooling, experience, and international socio-economic index (`isei`) as predictors. Do the analysis in both directions (i.e. once reweighting the private sector and once reweighting the public sector).