1. Variable Selection:
   * Main Y Under 5 Mortality 2018: 195 countries available
   * Main X: Out-of-school rate Female Primary 154
     1. Options:
        1. Pre primary – Primary – Lower sec – Upper Sec
        2. Possibilities:
           1. Just Primary:

+ Highest sample size 154

+ In Most countries Secondary depends on Primary, lower

– Doesn’t contain strength of secondary education

* + - * 1. Primary + Secondary:

Controlling?

Cumulative or Separate?

Qualitative:

Low if all below 50%

Low-Secondary if Primary above 50% but secondary

Middle

High

High Secondary

+ More info

– Lower Sample

* + 1. Chose not completion because:
       1. Less missing than completion
       2. Easier to compare, completion may differ more across countries
    2. Possible robustness check: Completion rate
  + Controls
    1. General:
       1. Try to choose one that wouldn’t correlate with main X and other Xs
       2. Try to choose one that has some correlation with Y (in theory all should be above 0 so not concern)
       3. Not bad conditioning control variables
    2. Out-of-school rate Male Primary
    3. State of Healthcare:
       1. Possible mechanistic bad controls:
          1. Institutional Delviery vs life expectancy
          2. https://data.unicef.org/resources/data\_explorer/unicef\_f/?ag=UNICEF&df=GLOBAL\_DATAFLOW&ver=1.0&dq=.DM\_LIFE\_EXP..&startPeriod=2015&endPeriod=2018
    4. State of economy
       1. Possible mechanistic bad controls:
          1. Vs GSP per capita
          2. Maybe both: strength of state social net

Education options

|  |  |  |  |
| --- | --- | --- | --- |
|  | Secondary 50 + | Secondary 25-50 | Secondary 0-25 |
| Primary 50 + | High-High | High-Mid | - |
| Primary 25-50 | Mid-High | Mid-Mid | - |
| Primary 0-25 | - | Low-Mid | Low-Low |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Y | X1 | X2 | X3 | X4 | X5 | X6 |
| Under-5-Mortality (%, 2018) | Out-of-school Female (Primary, %) | Out-of-school Male (Primary, %) | life expectancy (Years) | Institutional Delivery (%) | GPPP (USD K) | Mother’s Benefits (%) |

Possible multicoll:

1. X1-X2
2. X1-X5
3. X1-X6

Questions:

1. X1:
   1. Primary vs Qualitative vs Cofounders
2. Bad conditioning:
   1. Health life exp vs institutional del
   2. GDPP interaction
3. Multicollinearity: Chose less correlation?
4. Try out all possible models (possibly robustness)
5. Don’t change GDPP to qual
6. Healthcare: number of beds possible
7. Weight by population
8. Normal Distribution is always preferred
9. If data is weird, check why
10. Check variation via histogram + table summary