# Water Availability and Pollution

**What explains differences among nation-states in the percentage of their population that has access to clean water? (This is the "water" dependent variable.)**

As an important factor measuring nations' progress, Human Development Index include three important development dimensions: healthy life, access to education and living standard. Looking at the dataset, I think countries' Human Development Index (HDI) (variable: "hd") could be one important factor that could impact counties' access to clean water. Since some regions are much developed than others, I assume the region of a country (variable: "i.region") could also be revealing about the availability of clean water to a nation. Having below hypotheses, I run regression on the variables to examine the causation. I first run each independent variable individually, then together.

HA1: A country's Human Development Index affects the percentage of its population's access to clean water.
H01: A country's Human Development Index does not affect the percentage of its population's access to clean water.

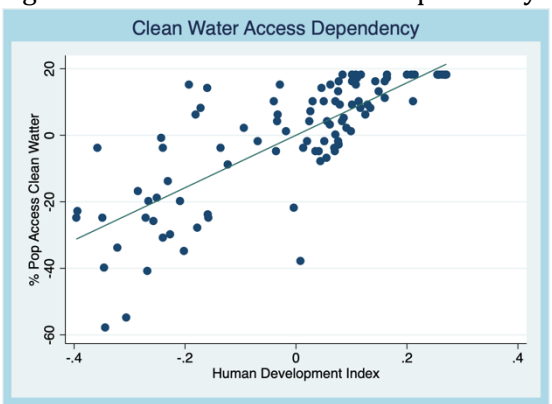
HA1: A country's region affects the percentage of its population's access to clean water.
H01: A country's region does not affect the percentage of its population's access to clean water.

Table 1: Percentage of Population with Access to Clean Water

| | (1)<br>Clean Water Access |
|---|---|
| Human<br>Development Index | 78.85*<br>(6.044) |
| _cons | 28.85*<br>(4.202) |
| *N* | 103 |
| adj. $R^2$ | 0.624 |

Standard errors in parentheses
$^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$

Figure 01: Clean Water Access Dependency




As assumed and shown by Figure 01, the regression model shows a strong dependency of clean water access to countries' human development index. The R-Squared also indicates a good explanatory power of the model. Here, I reject the null hypothesis in favor of my alternative hypothesis indicating dependency of access to clean water on human development.

As can be seen in Table 2, the regression for the region as independent variable shows causation only in two regions of Europe and Africa with opposite direction. However, the dependency of region on access to clean water cannot be seen in other regions.

Table 2: Percentage of Population with Access to Clean Water

| | (1)<br>Clean Access to Water |
|---|---|
| Europe | 8.709*<br>(3.854) |
| Africa | -27.39*<br>(3.542) |
| Mid-East & N. Africa | -1.238<br>(5.058) |
| Asia | -4.024<br>(4.200) |
| Australia, New Zealand | 11.76<br>(12.46) |
| _cons | 88.24*<br>(2.656) |
| *N* | 90 |
| adj. $R^2$ | 0.562 |

Standard errors in parentheses
$^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$

Similar conclusion can be draw when running model for human development index and controlling for region, however, with a comparatively much weaker dependency on HDI. Also, the model only shows a negative causation in Africa but weaker than running the model individually.

Table 3: Percentage of Population with Access to Clean Water

| | (1)<br>Clean Access to Water |
|---|---|
| Human Development Index | 49.81*<br>(11.58) |
| Europe | 4.649<br>(3.631) |
| Africa | -12.30*<br>(4.762) |
| Mid-East & N. Africa | 2.254<br>(4.672) |
| Asia | -0.256<br>(3.920) |
| Australia, New Zealand | 2.939<br>(11.52) |
| _cons | 50.29*<br>(9.145) |
| *N* | 90 |
| adj. $R^2$ | 0.638 |

Standard errors in parentheses
$^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$

Unlike model in Table 02, the dependency is not significant Europe which reveals the bias of omitted variables in the model, when rune individually. Figure 02 visualizes the findings.

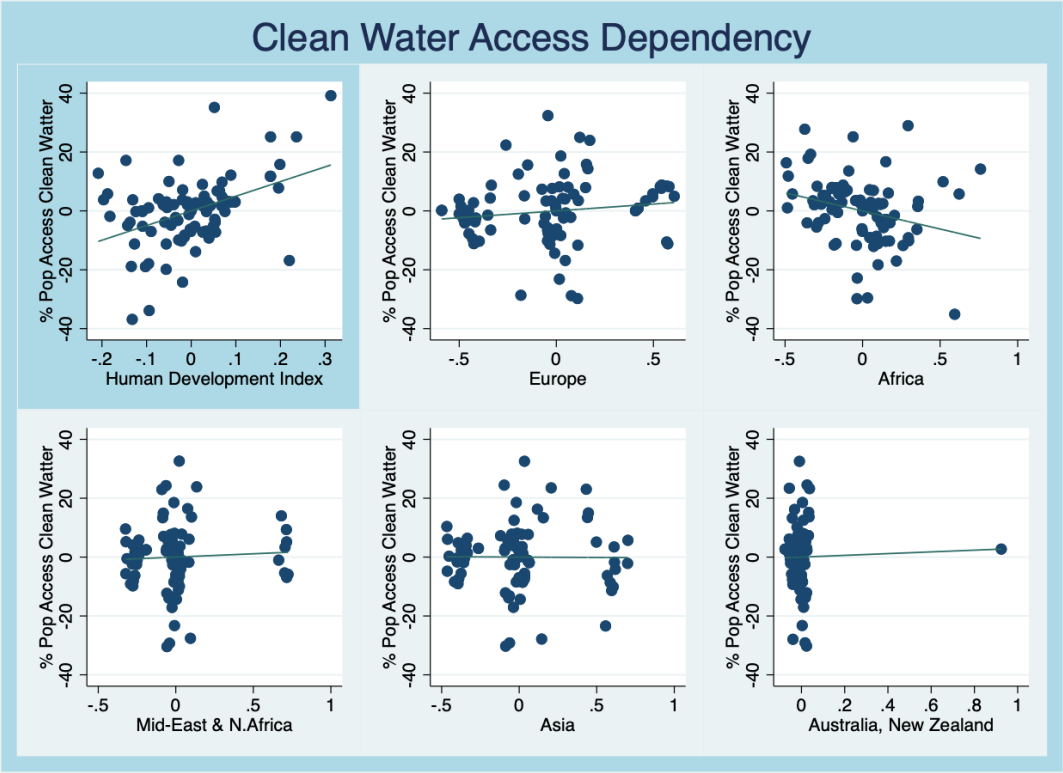



Figure 2: Clean Water Access Dependency on HDI based on region.

I run model multiple times controlling for other independent variables to see the impact of omitted variables. However, I got similar results to Table 02 for the model with some minor changes which do not alter result of model significantly. Therefore, the model running risk of omitted variable bias is low.

**What explains differences among nation-states in their output of water-borne pollutants? (H2OPOLU2)**

My assumption is that countries' population and their level of carbon dioxide in metric ton would have significant impact on their output of water-borne pollutants. Below, I test my assumptions by running regression on the dataset.

Table 4: Water-borne Pollutants Dependency Population and CO2 of Countries

| | (1)<br>Clean Access to Water |
|---|---|
| Population | 3.731*<br>(0.298) |
| CO2 | 26.78*<br>(12.96) |
| _cons | -79802.9<br>(106266.4) |
| *N* | 41 |
| adj. $R^2$ | 0.795 |

Standard errors in parentheses
$^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$

As can be seen in Table 4, there is a strong causal relation between independent variables (population and CO2) and the outcome of water-borne pollutants. The R-Squared is also indicative of a strong explanatory power of the model. However, as visualized by Figure 3, the strong causal power of population could be because of outliers in data. Accounting for those outliers may weaken the causality.

Also, I am curious to know if a country's level of political competition has any impact on the output of water-borne pollutants. Hence, I run a regression model to examine it.

**Figure 3: Water-borne Pollutants Dependency Population and CO2 of Countries**

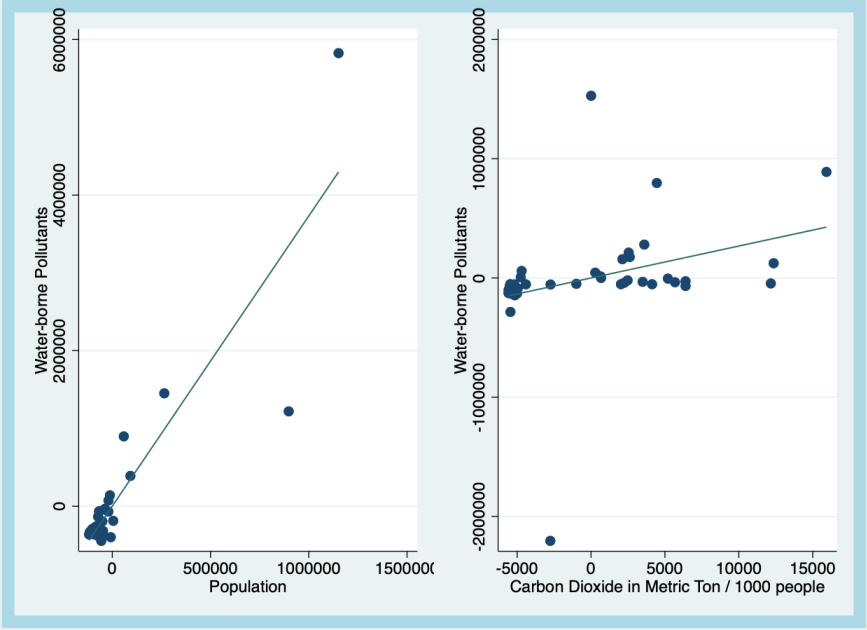


Table 5: Water-borne Pollutants Dependency Population and CO2 of Countries

| | (1)<br>Water-borne Pollutants |
|---|---|
| pop | 1.999*<br>(0.294) |
| CO2 | 31.30*<br>(13.07) |
| **Political Competition Level** | |
| 2 | -3632412.0*<br>(550624.6) |
| 3 | -3590547.1*<br>(545572.4) |
| 6 | -3658066.0*<br>(498064.7) |
| 7 | -3610172.6*<br>(487934.3) |
| 8 | -3504253.0*<br>(480081.2) |
| 9 | -3665021.7*<br>(448350.7) |
| 10 | -3684129.0*<br>(482130.6) |
| _cons | 3618329.5*<br>(469020.3) |
| *N* | 40 |
| adj. $R^2$ | 0.922 |

Standard errors in parentheses
+ $p < 0.10$, * $p < 0.05$

As can be seen in Table 4, controlling the mode for level of political competition weakens the causal of CO2 and increases the coefficient of population impact. However, it seems that political competition in general has a very strong negative impact on countries output of water-borne pollutants. But it does not differ based on the level of competition. Therefore, I think that controlling the model for political competition reveal the multicollinearity of the model and hence should not be included.

While doing the regression for two above purposes, I noticed there were much fewer independent variables with significant impact on percentage of people's access to clean water than on countries' output of water-borne pollutants.

Given the above difference between two model, a policy implication could be that to increase percentage of people who have access to clean water requires a better performance in countries human development index. However, output of water-borne pollutants depends on a number of factors and controlling water pollution require a much comprehensive multi-dimension plan.