

Applied Data Analytics

Statistics — Basics & location

Statistics conditional on other variables

Hans-Martin von Gaudecker and Aapo Stenhammar

GDP p.c. and population over the years

year	gdpPercap	pop
1952	2662.96	37.24
1977	4865.34	67.17
2002	5791.72	119.79

- gapminder data via plotly express, `px.data.gapminder()`
- Egypt, Lesotho, South Africa

Conditioning on country

- Saw:

$$\frac{\text{GDP}}{\text{pop}} \Big|_{\text{country} \in \{\text{Egypt, Lesotho, South Africa}\}}$$

- Want:

$$\frac{\text{GDP}}{\text{pop}} \Big|_{\text{country} = \text{Egypt}}$$

etc.

Wide format

year	Egypt	Lesotho	South Africa
1952	1419	299	4725
1977	2785	745	8029
2002	4755	1275	7711

Long format

country	year	gdpPercap
Egypt	1952	1419
Egypt	1977	2785
Egypt	2002	4755
Lesotho	1952	299
Lesotho	1977	745
Lesotho	2002	1275
South Africa	1952	4725
South Africa	1977	8029
South Africa	2002	7711

Caveat

- Remember:

$$f(\text{mean}(\text{data})) \neq \text{mean}(f(\text{data}))$$

unless f is an affine transformation $f(x) = a + b \cdot x$

- Here: $f(x) = \frac{x_1}{x_2}$

Population in millions

year	Egypt	Lesotho	South Africa
1952	22.22	0.75	14.26
1977	38.78	1.25	27.13
2002	73.31	2.05	44.43

Solution

- Calculate total GDP
- Calculate total population
- Divide