Tutorial: Correlation in Stata

The World Bank (http://data.worldbank.org/) is a great source of free public data on trends in health and economics around the world. In this example, we use public data from the World Bank to examine trends in immunization coverage for measles and DPT over time in low income countries. Open the dataset WorldBank.dta.

- Calculate the pairwise correlations between measles vaccination coverage, DPT vaccination coverage, and time.

```
pwcorr measles dpt year
```

- Make a scatterplot including both measles and immunization coverage on the plot. Does the plot explain the results above?

```
twoway (scatter dpt year) (scatter measles year)
```

Yes, there is a very strong positive relationship between time and immunization coverage. Further, it seems evident that trends in scaling up in immunization were similar for measles and DPT.

- Test whether there is a linear relationship between time and measles vaccination coverage. What are the null and alternative hypotheses? What is your conclusion?

```
pwcorr measles year, sig
```

Statistics > Summaries Tables and Tests > Summaries and Descriptive Statistics > Pairwise Correlations

- Test for a monotonic relationship between time and measles vaccination coverage. What are the null and alternative hypotheses? What is your conclusion?

```
spearman measles year
```

Statistics > Summaries, tables, and tests > Nonparametric tests of hypotheses > Spearman's rank correlation

- Why do you think the correlations are so high in this example? Should you always have such high aspirations regarding the magnitude of your correlation coefficients when analyzing public health data?

Source: Created from: World Bank, World Development Indicators and Global Development Finance. Vaccination coverage from WHO and UNICEF