

## **ACTIVITY 1: Worksheet**

Often, statisticians are interested in examining how two quantitative variables relate. In this set of activities, the focus will be on exploring linear relationships. Two variables can be merely correlated meaning as one variable changes, the other changes in sync. Just because two variables are related does not mean changes in one variable will directly lead to changes in the other variable. Hence correlation does not imply causation. When causation can be established, it is important to distinguish between the independent and dependent variables. Data will be collected on countries to investigate whether there is a linear relationship that can be modeled using a least squares regression line.

A least squares regression line is a linear equation that models the relationship between two quantitative variables enabling prediction.

First, it is essential to grasp the distinction between independent and dependent variables.

We will start by looking at some variables for countries. For the variables below, match 5 of them together in such a way that the x-variable might result in a change in the y-variable. You may use any of the variables more than once.

Contraceptive use (% of women aged 15-49)	Sugar per person (grams per day)	Maternal mortality ratio (per 100,000 live births)
Personal computers (per 100 people)	Babies per woman	At least basic water source, overall (%)
Body mass index (men or women)	Corruption Perception index	Food supply (kilocalories per person per day)
Medical doctors per 1000	Fish and seafood consumption per capita	Adult literacy rate (%)
Bad teeth (weighted average 12 years)	Cars, trucks & buses per 1000 people	OWID education index (average years schooling)
Extreme poverty (% people below \$x per day)	Infant mortality (per 1000 live births)	Breast cancer new cases per 100,000 women
Government health spending per person	CO2 emissions per capita	Underweight children (%)
Happiness Score	Births attended by skilled health staff (% of total)	Human Development Index
Broadband subscribers (per 100 people)	Child mortality (0-5 years old dying per 1000 born)	Dead kids per woman (# of children on average that die before 35 <sup>th</sup> birthday)
Stillbirths (per 1000 births)	Life expectancy	At least basic sanitation overall (%)

**Example:** Contraceptive use (% of women aged 15-49) (x) and Babies per woman (y)

After the in-class discussion on the difference between the independent variable and dependent variable, write down the definitions below.

**Independent variable:**

**Dependent variable:**

Choose one of the relationships discussed, fill in your independent variable for x and dependent variable for y, and go to gapminder.com to find the data for the simple random sample of 25 countries below. Choose the most recent data available for each country. If data is not available for a country, write "NA" for both variables.

**Note: Make sure that you are using a rate (per 1000, per 100, or percentage rather than a total). Why would this be important?**

Country	x –	y –
Zimbabwe		
Turkey		
Australia		
China		
Georgia		
Peru		
Burkina Faso		
Greece		
Nepal		
Belgium		
Monaco		
Kuwait		
Philippines		
Malawi		
Rwanda		
South Korea		
Tajikistan		
Togo		
Denmark		
Haiti		
Qatar		
Uganda		
Vietnam		
Romania		
Guyana		

**You will be using this data for the remainder of the activities.**

At the end of class, you can start a PowerPoint with a detailed description of the two quantitative variables you chose. Make sure you label them as the independent and dependent variables. Explain how and why you believe your chosen independent variable will impact your chosen dependent variable.