

ACTIVITY 1: Worksheet Answers

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 th 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)

Some possible relationships students could choose:

Sugar per person (grams per day) (x) and BMI (y)

Births attended by skilled health staff (% of total) (x) and Infant mortality (per 1000 live births) (y)

Human Development Index (x) and Happiness Score (y)

Personal computers (per 100 people) (x) and Broadband subscribers (per 100 people) (y)

Cars, trucks & buses per 1000 people (x) and CO2 emissions per capita (y)

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

Independent variable: The independent variable, x, causes changes in the response (dependent) variable.

Dependent variable: The dependent variable, y, measures an outcome in a study and is impacted by the independent variable (it depends on the independent 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? If a total is chosen, the population of the country will greatly impact that variable, thus masking the relationship.

Country	x – Contraceptive use (% women ages 15-49)	y – Babies per woman
Zimbabwe	66.8	3.85
Turkey	69.8	2.05
Australia	66.9	1.8
China	84.5	1.81
Georgia	39.3	2.11
Peru	74.4	2.22
Burkina Faso	35.2	4.77
Greece	61.3	1.32
Nepal	46.7	2.08
Belgium	66.7	2.03
Monaco	NA	NA
Kuwait	52	2.79
Philippines	54.3	2.87
Malawi	65.6	4
Rwanda	64.1	3.87
South Korea	82.3	.97
Tajikistan	29.3	3.29
Togo	23.8	4.5
Denmark	76.5	1.76
Haiti	34.3	3.02
Qatar	37.5	1.97
Uganda	50.2	4.59
Vietnam	72.8	1.94
Romania	69.8	1.39
Guyana	29.9	2.42

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.

Example for the variables used in the previous example.

Slide 1:

Independent Variable:

The contraceptive prevalence rate is measured as the percentage of married women or women in a union aged 15-49 who are practicing, or whose sexual partners are practicing any form of contraception.

Dependent Variable:

The birth rate is the average number of children per woman.

We believe that as the percentage of contraceptive use increases in countries, the birth rate will decrease. Birth control prevents pregnancy, so with a higher level of use within a country, the birth rate will naturally decrease for the country.