

# Overview of Data Collection Principles Applications

YOUR NAME

20 August, 2020

## Exercises

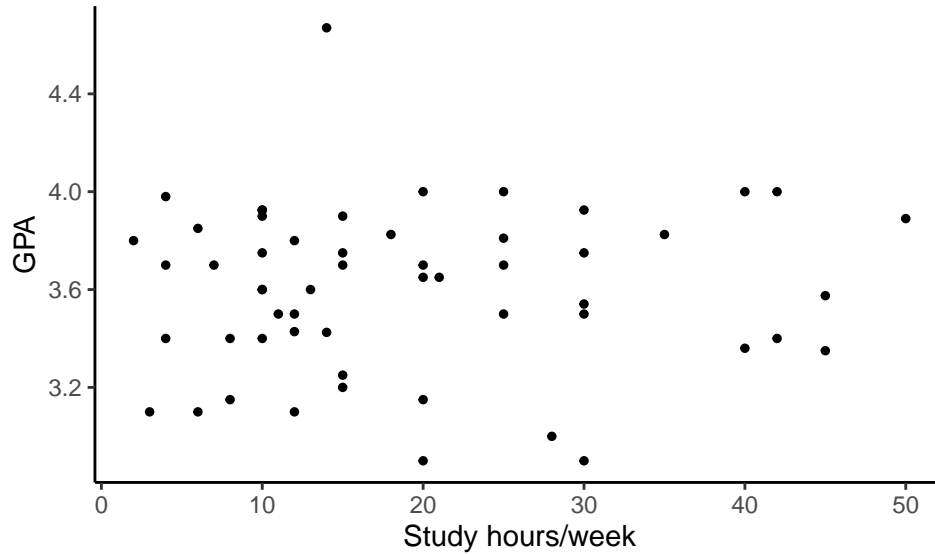
1. **Generalizability and causality.** Identify the population of interest and the sample in the studies described below, these are the same studies from the previous lesson. Also comment on whether or not the results of the study can be generalized to the population and if the findings of the study can be used to establish causal relationships.
  - a. Researchers collected data to examine the relationship between pollutants and preterm births in Southern California. During the study air pollution levels were measured by air quality monitoring stations. Specifically, levels of carbon monoxide were recorded in parts per million, nitrogen dioxide and ozone in parts per hundred million, and coarse particulate matter (PM<sub>10</sub>) in  $\mu g/m^3$ . Length of gestation data were collected on 143,196 births between the years 1989 and 1993, and air pollution exposure during gestation was calculated for each birth. The analysis suggested that increased ambient PM<sub>10</sub> and, to a lesser degree, CO concentrations may be associated with the occurrence of preterm births.<sup>1</sup>
  - b. The Buteyko method is a shallow breathing technique developed by Konstantin Buteyko, a Russian doctor, in 1952. Anecdotal evidence suggests that the Buteyko method can reduce asthma symptoms and improve quality of life. In a scientific study to determine the effectiveness of this method, researchers recruited 600 asthma patients aged 18-69 who relied on medication for asthma treatment. These patients were split into two research groups: one practiced the Buteyko method and the other did not. Patients were scored on quality of life, activity, asthma symptoms, and medication reduction on a scale from 0 to 10. On average, the participants in the Buteyko group experienced a significant reduction in asthma symptoms and an improvement in quality of life.<sup>2</sup>

---

<sup>1</sup>B. Ritz et al. "Effect of air pollution on preterm birth among children born in Southern California between 1989 and 1993". In: Epidemiology 11.5 (2000), pp. 502-511.

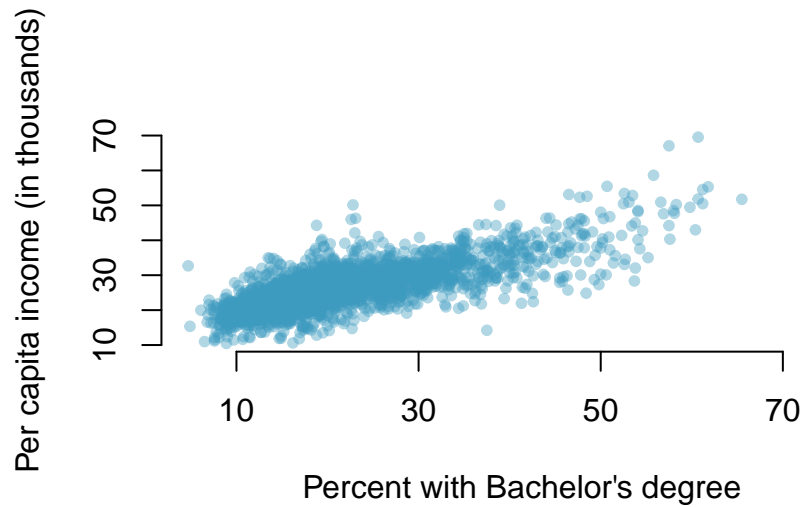
<sup>2</sup>J. McGowan. "Health Education: Does the Buteyko Institute Method make a difference?" In: Thorax 58 (2003).

2. **GPA and study time.** A survey was conducted on 55 undergraduates from Duke University who took an introductory statistics course in Spring 2012. Among many other questions, this survey asked them about their GPA and the number of hours they spent studying per week. The scatterplot below displays the relationship between these two variables.



- What is the explanatory variable and what is the response variable?
- Describe the relationship between the two variables. Make sure to discuss unusual observations, if any.
- Is this an experiment or an observational study?
- Can we conclude that studying longer hours leads to higher GPAs?

3. **Income and education** The scatterplot below shows the relationship between per capita income (in thousands of dollars) and percent of population with a bachelor's degree in 3,143 counties in the US in 2010.



- What are the explanatory and response variables?
- Describe the relationship between the two variables. Make sure to discuss unusual observations, if any.
- Can we conclude that having a bachelor's degree increases one's income?