Assignment 1

Hongyu He (2632195) & Bruno Hoevelaken (2645065)

Group CS 6

Theoretical exercises

Exercise 1.1

10. The sampling method used in this scenario is **flawed**. *The book clearly states* (p7) **that internet polls**, in which people online can decide whether to respond, **are by their very nature seriously flawed**, as they result in **voluntary response samples**. This also means **nonresponses** might have occured. This would result in **missing data** regarding individuals who were unwilling to participate in the online poll. Due to the missing data, we are unable extrapolate any meaningful conclusions out of this sample, regarding the wider populous.

11. It is difficult to fully conclude that a sampling method is sound, with limited context. The **sample size** seems to be **sufficient**, and the fact that the the subjects were **randomly selected** is also a **good sign**. There can still be **potential issues** regarding the sampling method. For example, we are *not told how many* of the 1018 adults *actually picked up*, and if they did pick up, how many wished to participate.

12. This exercise presents a scenario which suffers from the *same issue as exercise 10*, as it relies on an **internet poll**. Another issue is the fact that it is unclear what the other 59% of users said.

Exercise 1.2

22. The depths (km) of earthquakes measured relative to the earth's crust(s), which means there is a meaningful **natural zero starting point**, and so fall under the **ratio** level of measurement.

23. The lineup numbers exist as distinguishing features, each corresponding to a player. This means that they fall under the **nominal** level of measurement. So *calculating their mean is pointless.*

Exercise 1.3

6. The researchers do not merely observe the subjects, sincec they make the *subjects go through treatment prior to observation.* This means that it would be considered an **experiment**.

7. Since the researchers are collecting samples at every interval of 5 meters, they are making use of a systematic mehtod. It falls under a **systematic sampling** method because the collection happens at every kth elementh of the population. In this case, k is 5.

8. This clearly follows the **cluster sampling** method. The population has already been *divided into clusters* through the existence of polling stations. ABC then *randomly selects certain clusters*, and finally *surveyed all voters* (members) of the selected clusters.

Exercise 1.4

a.

The axis representing the amount of seats starts at 50, while the 3 bars fall under that threshold. This reduces clarity.

b.

1. This data would fall under nominal data. A Pareto bar chart would be the most useful method to display such data.

2. Since the numbers of previously done votes is **ratio** data, a histogram seems like the best graphical representation.

Exercise 1.5

a.

Histogram of SampleA