

# Log 708 - Chapter 1 Solutions

Halvard Arntzen

## 1.1

- a. Categorical, nominal
- b. Numerical, continuous
- c. Categorical, nominal
- d. Numerical, discrete
- e. Categorical, ordinal
- f. Numerical, continuous

## 1.2

By the empirical rule (the sample is fairly large) we consider the interval  $\bar{x} \pm 2S$  to be a fair approximation of an interval containing 95% of durations. Clearly this could be valuable for planning. E.g.  $56 + 2 \cdot 11 = 78$  minutes would be an upper limit for almost all operations.

## 1.3

- a. Considering the points with around 50 years of life expectancy, a point guess of about 20 seems reasonable. An interval from 0 - 50% appears likely to contain the number.
- b. At 80 years, we might guess 70% urbanization. An interval guess could be 40 - 95%.
- c. No. Clearly, a developed and industrialized country will normally have more urbanization as well as a more developed health care system. Thus, some hidden variables may well cause the correlation. This means that the level of development in a country affects both of the observed variables, which then become correlated. However, there is no reason to believe in a direct *causal* effect from high urbanization to high expected lifelength. The causation lies in the level of development, which is *hidden* when we only consider the two variables in the question isolated. This “phenomenon” is very common in statistics and *quite often* leads to false interpretation of correlation as causation. We should always think about the possibility for hidden variables when someone shows a correlation result. The effect can be illustrated with the figure below, where blue arrows are causal relations, and the red arrow is suspected to be a correlation relationship with little or no causal explanation.

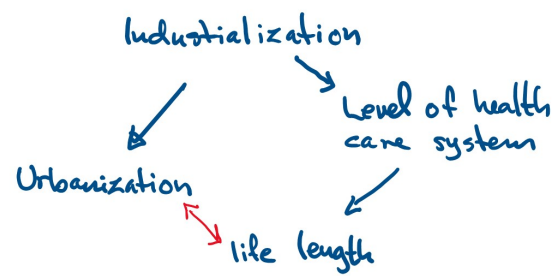


Figure 1: Visualizing hidden variables effect