Overview

5M

This introduction provides a general idea of what credibility is.

Consider the scenario below:

All-Care Insurance Co. plans to sell a new type of car insurance in Iowa, and they are interested in the expected annual total claim size per insured. The company gets data from two sources:

- All-Care's internal observations, where each Iowa insured has an annual claim size of 100 on average.
- Industry-wide data, where each Iowa driver has an annual claim size of 200 on average.

What is the expected annual total claim size for each Iowa driver that All-Care should use in their analysis? 100, 200, or something in between?

To answer this question, we can apply ***credibility theory***. This theory tells us to what degree we should rely on a particular experience when predicting future outcomes.

Take the scenario above as an example. One intuitive way to determine the expected aggregate claim is to "combine" the two numbers by taking their weighted average, with weights depending on how much we trust each source.

- The internal observations may address the specific need better, but the sample size is likely limited.
- The industry-wide data has a larger sample size, but it might be too general.

Thus, the real question is - how much weight should be placed on All-Care's internal data as opposed to the industry-wide data?

To answer this, we use the ***credibility factor***, or ***credibility*** for short, which is denoted as Z. In general,

- Z represents the weight assigned to the sample/observations
- $1 - Z$ represents the weight assigned to the general group

Since Z is a weight, it must be that $0 \le Z \le 1$.

For instance, if $Z = 0.3$, it means 30% credibility is assigned to All-Care's internal observations, and 70% credibility is assigned to the industry-wide data. Therefore, the annual aggregate claim per insured is estimated to be:

$$\begin{aligned}\text{Estimate} &= 0.3(100) + 0.7(200) \\ &= 170\end{aligned}$$

The larger the value of Z, the more we trust the sample of observations. This, along with other characteristics of Z, are reflected in the following three properties, where n denotes the number of observations in the sample:

I. $0 \le Z \le 1$. The worst that can happen is that we discard the data, i.e., $Z = 0$. On the other hand, $Z = 1$ means we rely entirely on the data. Regardless, Z must be in between these bounds.

II. $\dfrac{\mathrm{d}Z}{\mathrm{d}n} > 0$. As the number of observations increase, the change in the credibility factor is always positive. This means that increasing the number of observations also increases the credibility factor.

III. $\dfrac{\mathrm{d}\left(\frac{Z}{n}\right)}{\mathrm{d}n} < 0$. As the number of observations increase, the ratio of Z/n decreases. In other words, a credibility factor increases by a larger amount when the number of observations is smaller. For instance, adding 500 observations to a dataset of 250 observations will have a greater impact on Z than adding 500 observations to a sample of 1,000,000 observations.

Coach's Remarks

The official textbook uses E instead of n for the three properties above. We use n to be consistent with the notation we use in this course.

This section covers ***classical credibility***, which is also known as ***limited fluctuation credibility***. It suggests how large the sample has to be to assign 100% credibility to the observations.