Attribution Modelling

R Markdown

If you are in the digital marketing industry, you know that everyone talks about campaign optimization. But before campaign optimization, you need to assess the value of your campaigns. Two main questions that arise when you are trying to understand the conversion journeys are:

- How to attribute conversion success to marketing spend?
- Where to spend the next marketing dollar?

I have explianed different approachs for solving attiubtion problems in the below posts:

Here I'm going to wite about survival analysis and how it can be used in defining the right credit for each marketing channel.

Survival analysis is a popular approach among biostatisticians for analyzing the expected duration of time until one or more events happen, such as death in biological organisms. However, this statistical technique attracted scientists and technicians in other domains. Survival analysis is called reliability theory or reliability analysis in engineering, duration analysis in economics, and event history analysis in sociology. Two example questions that survival analysis attempts to answer are listed as follows:

What is the proportion of a population which will survive past a certain time? How do particular events or treatments increase or decrease the probability of survival? In survival analysis, death or failure is considered an 'event'; traditionally only a single event occurs for each subject, after which the organism or mechanism is dead or broken. Chandler-Pepelnjak (2010) proposed to use survival analysis for marketing attribution analysis. He suggested that each event in a customer's' journey gives an indication of whether the journey is still alive. And when the customer converts to purchase, the journey then 'dies'.

If you are into marketing do you want to know how long your customers are likely to stay with you; or whether you customers fall into a certain demographic profile, or whether those who enter your website via a certain channel or campaign tend to convert more quickly?

We found this survival analysis technique extremely valuable for attribution modelling. Using survival analysis, we measure time of conversion and probability of conversion based on the channels that customers used on their journey. The probability of being alive (not converting) and being dead (conversion) is used for the following example:

What is the visitors' probability to convert after 50 days?

Before going further, I have breifly summarised the math behind survival analysis. Let's say T is representing the waiting time until an event occurs (for example the probability of conversion after 50 days). The probability density function of (p.d.f.) is f(t) and the comulative distribution function (c.d.f) is indicated as below:

$F(t) = Pr[T < t] = \int_{-\infty}^{t} f(u)du$

Let's take the example of marketing channel and assume that the event of interest is the conversion after exposing to a particular marketing channel. As a result, the probability that this event is not occured (no conversion) can be indicated by the below formula:

$S(t) = 1 - F(t) = Pr[T \geq t]$

Also, we need to calculate the conditional probability that event will occure within $[t, t + dt)$ We call it hazard function that has been expresses as below:

$h(t) = \lim_{dt \to 0} \frac{\Pr[t \leq T < t+dt \ | T \geq t]}{dt} = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t)$

And the relationship between hazard and survival functions has been shown in the below formula:

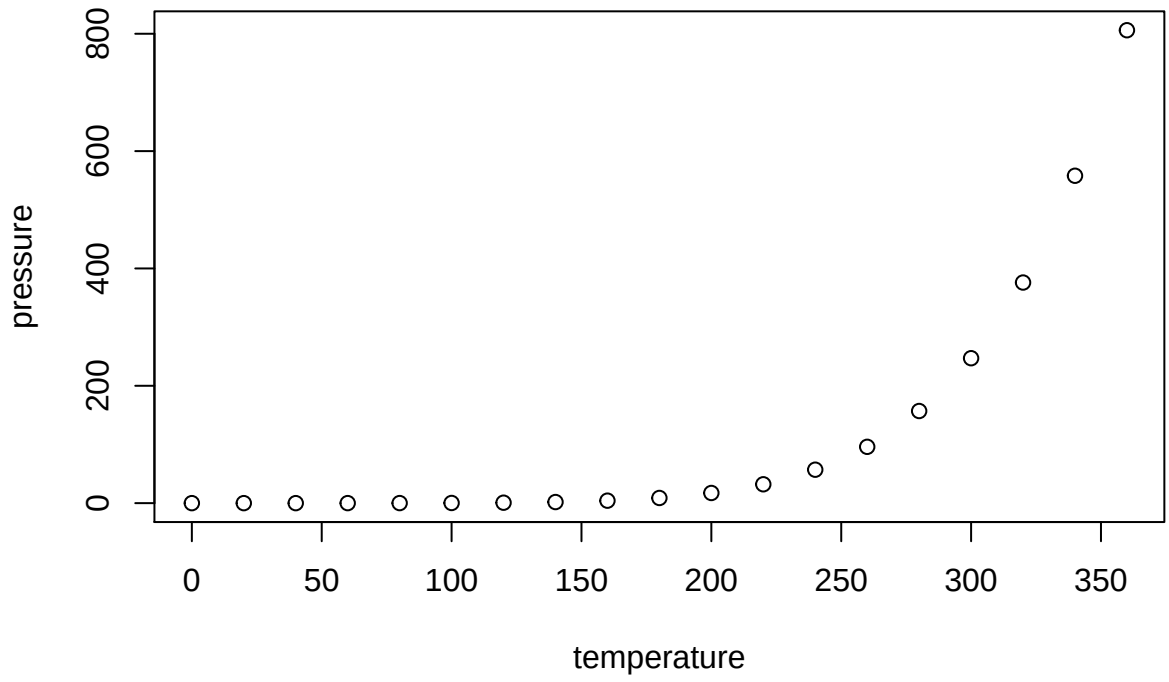
$S(t) = \exp\left(-\int_0^t h(u)du\right)$

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.