

# 演算法的時代

## 章節 1/2 - 資料歸屬 ( Data Ascription )

文 | Andrew Green, Global Head of Audience Solutions at Ipsos Connect

Mario Paic, Head of Data Science at Ipsos Connect

編譯 | June Chen 益普索公關行銷專員

演算法是用來計算或解決問題的一套規則或過程，其在數位世界中無所不在，例如，銀行是否願意放貸，雇主是否會錄用面試者，都有可能使用到演算法。

Google 的佩奇排名 ( PageRank ) 演算法決定了其搜尋引擎結果中的網站如何排名。同時，Facebook 的資訊流 ( News Feed ) 演算法也控制著使用者在社群網站中所看到的內容。

演算法在受眾測量 ( Audience Measurement ) 的領域中，一般使用於進度分析軟體，這種軟體讓媒體策劃、購買及銷售人員能估算可能接觸到廣告資訊的受眾數量。現今也越來越多受眾測量的進階統計工具，包括資料歸屬、資料融合和受眾建模。

僅有調查資料本身，已不足以捕捉快速變化的媒體環境，越來越多的傳播平台以及資訊轟炸，讓受眾已經不再有耐心參加漫長乏味的調查。而統計技術讓行銷人員不用蒐集太多資料就能預測消費者行為，並且對資料精確度也有較高的容忍度。

不過這些統計方法的設計和執行，是需要需要高水準的專業知識和技巧。而為了跟上不斷變化的消費者特徵和習慣，統計方法變得越來越關鍵。

### 情境

想要確定一個人的血型，不需要抽取他的全部血液。同樣，透過抽樣調查，並對人群進行詢問或追蹤他們的行為，就能獲得很多資訊。良好的市場調查，其有效樣本應盡可能地反映被測量人群的變化性。

我們知道性別、年齡、文化程度、職業、家庭人數、收入水準以及人們居住的區域都會影響到人們閱讀報紙或雜誌、觀看某些節目或收聽廣播的可能性。因此，這些特徵必須忠實地反映在用來測量媒體使用的有效樣本中。

非人口因素也很重要。例如，當詢問數位設備使用行為時，正確地反映一個家庭中擁有的媒體接收設備 ( 桌上型電腦、平板電腦、智慧手機等 ) 的數量和範圍就很重要，因為這些可能都會影響使用頻率。

本文將介紹，我們是如何盡可能地在受訪者能接受的時間內，獲得更多受訪者的資訊，且考量到很少人有完美的記憶，也必須避免詢問一些人們不太可能記得的資訊。

## 統計調整有助於以下方面

為了能將調查研究的投資報酬最大化，行銷人員希望能知道目標消費者的一切，：

- 他們是誰（人口統計學、地理人口統計學、心理學等方面）？
- 他們對於被問及的品牌看法是什麼？
- 他們的消費行為如何（購買水準、品牌選擇等）？
- 他們未來會想要購買什麼？

而策劃及廣告宣傳活動購買者，則需要找出觸及目標受眾的最佳方法：

- 他們在不同的時間，接觸到哪些媒體（電視節目、報紙、雜誌、廣播電臺、網站、應用程式、海報板.....）？
- 在一天中的不同時間，他們會關注或觸及到哪些媒體？
- 何時是向目標消費者傳達廣告資訊的最佳時機（對資訊的接受度，他們什麼時候會在市場上購買.....）？

但事實上，沒有一個受訪者會願意回答這麼多的問題。而且很多問題都無法準確地回答。因此，有兩個密切相關的統計技術，可以用來幫助解決這個問題，即：資料歸屬（data ascription）和資料融合（data fusion）。

## 資料歸屬（data ascription）

若調查問卷的回答有缺失或不完整，我們可以透過查看類似的調查受訪者給出的回答來推斷這些回答有可能是什麼。

在這種情況下，我們可以設計兩份（或更多）調查問卷，每份問卷都有一些共同的核心問題，但也有些關於其他主題的單獨問題。這些問卷可以同時分開用於不同但相似的人群樣本，也可以在不同的時間進行（例如：問卷 A 可以進行一個月，然後問卷 B 再進行一個月，等等）。

再利用每一樣本的已知人口統計資料、其他特徵，及其他一般問題的回答，來與回答不同問卷的受訪者作「配對」。接著將 A 問卷的回答，歸屬為回答 B 問卷的相匹配的受訪者，反之亦然。如此一來，我們就能得到一個更大的資料庫。

## 數據歸屬的研究例子

在巴西，IpsosEGM 在各個主要地區進行了一項包含許多問題的面對面問卷調查。該問卷有兩個版本，媒體和人口統計學方面的問題都相同，但有關品牌和產品的使用情況以及態度的問題則不相同。兩個版本的問卷每半年交替執行，這樣我們就可以把所有蒐集到的回答合併成一個為期 12 個月的資料庫。

在澳洲，益普索 EMMA（Enhanced Media Metrics Australia）調查，用於了解媒體消費和產品使用情況。該調查涵蓋了大量的產品類別，我們向一對匹配的樣本中的其中一個受訪者，詢問關於約一半產品類別的詳細品牌問題，再將其答案歸屬為另一個受訪者，生成一個最終綜合資料，進而為整個樣本提供詳細的品牌資訊。

在下一個章節中我們將繼續介紹資料融合（data fusion）及其他統計技術及其應用，

#### Ipsos Connet 益普索媒體與傳播研究

Ipsos Connect 結合 Ipsos 全球產品服務，專精於品牌溝通、廣告測試及傳播媒體的相關研究。現今品牌溝通由於廣告及傳播媒體變得日益複雜、零碎及數位化，而更顯複雜困難。Ipsos 擁抱這複雜且現在的環境，整合品牌表現調查及媒體環境內容，並提供我們客戶更創新、真實、即時且數據和科技驅動的研究發現洞察。

更多的資訊，請參考 [www.ipsos.com/connect](http://www.ipsos.com/connect)

或洽詢我們的研究團隊

研究總監 Darren Freeman 費愷毅

02 2701-7278 ext.155 [darren.freeman@ipsos.com](mailto:darren.freeman@ipsos.com)

# The Age of the Algorithm

## Part 1 – Data Ascription

Andrew Green, Global Head of Audience Solutions at Ipsos Connect

Mario Paic, Head of Data Science at Ipsos Connect

---

### Introduction

An algorithm is a process or set of rules followed in calculations or other problem-solving operations, especially by a computer. They are all-pervasive in the digital world, determining, for example, whether banks will lend or employers will pick people for interview.

In the online world, Google's PageRank algorithm determines how websites in the company's search engine results are ranked. Meanwhile, Facebook's News Feed algorithm controls who sees what within the social network.

In the Audience Measurement area, algorithms are commonly used in schedule analysis software, employed by media planners, buyers and sellers to estimate the number of people likely to see or hear an advertising message.

They are also central to more advanced statistical techniques being used increasingly in audience measurement including data ascription, data fusion and audience modelling.

Survey data alone is no longer enough to capture the complexities of a fast-changing media environment, with more content choices and more distribution platforms from which to access them. We are demanding more and more from our respondents at a time when many are becoming less willing to participate in long or boring surveys. Employing statistical techniques allows us to collect less data from individuals and to tolerate less precision in their answers.

The design and execution of these statistical methods demands a high level of expertise and skill. For users of research data, they have become increasingly critical in the race to keep abreast of changing audience characteristics and habits.

### The Context

We don't need to extract all the blood from somebody in order to determine their blood type. In the same way, we can learn a lot from taking a sample of a population and questioning them or tracking their behaviour. Good market research practice demands that a sample should represent as much of the variability in the population being measured as possible to be effective.

In the audience measurement domain, we know that gender, age, education level, occupation, household size, income levels and the region where people live will all affect the probability that they read certain newspapers or magazines, watch given programmes or listen to the radio. So these characteristics must be

faithfully reflected in the composition of any sample purporting to measure media usage.

Non-demographic factors can be important too. For example, when asking about digital behaviour, it is important to properly represent the number and range of media-receiving devices (PCs, tablets, Smartphones etc.) owned by households, which are likely to be associated with usage levels.

But while surveys have proven very valuable to companies and governments over the years, they have their limitations. One limitation is that they can never be perfectly representative of a population, unless we talk to everybody and unless they all answer with total honesty and perfect recall.

Statistical weighting is often used in market research to correct these sorts of imbalances.

The more important limitation for the purposes of this paper is how we extract as much information as possible from our respondents. There are natural limits to how much time people are prepared to spend answering questions. Given that few are blessed with perfect recall, we also have to be careful to word questions clearly and not to ask for information it is unreasonable to expect people to remember.

The challenge today is that clients want us to collect more information, not less.

### **Where Statistical Adjustment Can Help**

Marketers want to know everything they can about their target consumer in order to maximise the return on their research investment:

- Who they are (demographics, geo-demographics, psychographics etc.)
- What they think about brands in the category they are asking about
- How they behave (purchasing levels, brand choice etc.)
- What they intend to purchase in the future

Those planning and buying advertising campaigns need to uncover the best ways of reaching and influencing their target audiences:

- Which media do they come into contact with at different times of day (TV programmes, newspapers, magazines, radio stations, web sites, apps, poster panels...)?
- Which media are they more or less attentive to or engaged with at different times of day?
- When is the best time to reach people with an advertising message (message receptiveness, when are they in the market to buy...)?

But no individual respondent will agree to answer such a large number of questions. And many of the questions will be impossible to answer accurately. There are two closely related statistical techniques used to help address this: data ascription and data fusion.

## Data Ascription

Where answers to a survey are missing or incomplete, it is possible to infer what those answers would be by looking at answers given by similar survey respondents. These responses may be missing by accident (people forgot or omitted to answer them) or by design (where we had too many questions to ask, so split the questionnaire between different sub-samples).

In this case, we design two (or more) questionnaires, each sharing certain core questions in common, but with separate sets of questions on other topics. These questionnaires can either be served simultaneously to separate but similar samples of people or they could be asked at different times (e.g. Questionnaire A can be asked for a month, then Questionnaire B for a month and so on).

The assumption is that we can then 'match' people answering the different questionnaires using the known demographic and other characteristics of each sample member, as well as the answers they give to other common questions. We then take answers to the first set of questions and ascribe them to matching respondents who answered the second set of questions and vice versa. This gives us a larger database of answers than we could have had with a single sample of people.

### Examples of Data Ascription:

In Brazil, the Ipsos EGM survey has a lengthy questionnaire administered face-to-face in all the major regions of the country. There are two versions of the questionnaire, with media and demographic questions identical on both, but different questions asked about usage of various brands and products, as well as about attitudes. Each version of the questionnaire is served in alternate half years, allowing us to merge answers to all the questions into a single database for 12 month periods.

In Australia, our emma (Enhanced Media Metrics Australia) study measures both media consumption and product usage. Because there are a large number of product categories covered, all participants in the study are asked about top line product category usage (e.g. do they drive a car? what kinds of food and beverages do they consume?). Two matched samples are each then asked detailed brand questions about half of the product categories. The responses from each half of the sample are then ascribed back to the other half of the sample, producing a final integrated dataset where we have detailed brand information for the entire sample.

## Ipsos Connect

Ipsos Connect, a global specialized business to coordinate Ipsos services in the domains of Brand Communications, Advertising and Media.

As the worlds of brand communications, advertising and media become increasingly complex, fragmented and digitalized, Ipsos helps its clients embrace this modern complexity by bringing together brand expression, media and content, & via consolidating our expertise to bring new, data-driven, real-life, real-time, tech-driven insights to our clients.

For more information, visit us online at: [www.ipsos.com/connect](http://www.ipsos.com/connect)

or contact our research team:

Darren Freeman, Research Director

02 2701-7278 ext.155 [darren.freeman@ipsos.com](mailto:darren.freeman@ipsos.com)