



## 一、網頁抽樣

在網路時代，以「網頁內容」為研究對象，已經成為有意義、且有發展潛力的研究方法。其研究樣本單位為「網頁」，而合格的網頁，通常由「內容關鍵字」決定。

網頁抽樣與基本抽樣原理相同，也包括3道程序，只是應用技術不同。

### (一)隨機性/等機率性與母群清冊

網頁內容研究在母群清冊完整性、取得性上，有極大的優勢、也具備與傳統清冊極大的不同特性，就是內容關鍵字的「Google 排行榜」提供了一個「完整」「排序後」、且具備「排序動態性」的清冊。

「Google 排行榜」是一個已經由「點閱變項」比較後的清冊，而「點閱變項」就已反映資料的「質」，從而可以有以下不同的抽樣策略與抽樣方法：系統抽樣、主質量抽樣、分層抽樣。同時應注意「質量分割點」的選擇。

### (二)樣本數

原則上，人類行為研究的樣本數，最低建議仍為200。

但由於「Google 排行榜」已經等於作過「前測過濾」的特性，如果研究的題目類似「三十大、百大網站」等，可以依據研究主題調降樣本數，並不會嚴重損害代表性。

### (三)抽樣方法

「Google 排行榜」作為母群清冊，有2個層級：

關鍵字搜尋結果頁面 [SERP \(Search engine results page\)](#)，可簡稱「搜尋頁」，預設是每搜尋頁包括10網頁。

網頁才是真正的抽樣單位。

#### 1、系統抽樣

「Google 排行榜」會列出與關鍵字有關的搜尋頁數，從而可以使用標準2階段系統抽樣法：先隨機抽「搜尋頁」，再隨機抽網頁。

但考慮「Google 排行榜」是已經排序的清冊，同時，研究發現排在第10「搜尋頁」後的網頁，被點閱率非常低，而研究者通常都是要研究讀者的行為，所以後段網頁的可研究質量，相對很低。

從樣本質量的策略上考量，系統抽樣法是不被鼓勵的。



## 2、主質量抽樣

「主質量」抽樣即策略上要抽出最具質量的樣本，根據前述點閱行為，就是由第一頁開始抽。

如果是抽一般最少樣本200頁，就是抽前20「搜尋頁」。

若要增加代表性，就繼續按最適樣本數條件加抽。

就「主質量」、微小誤差不計的觀點而言，幾乎接近「普查」的效果。

## 3、質量分割

「Google 排行榜」還有一個經驗法則的特色，就是總頁數很多，但後段有高比例網頁只是搭邊，亦即質量並不高。

所以可先以大間距，隨機抽幾頁研究，評估選擇「質量分割點」，捨棄後段的網頁。

譬如原訂抽800個樣本，即80個搜尋頁，則每先10搜尋頁抽1頁。

如果發現第50搜尋頁的各網頁的質量，已經降低近無關了，就在40-50頁中，折半選第45搜尋頁來研究。同理，繼續探索，找到質量尚可的搜尋頁，譬如第42頁，而以該頁作為「質量分割點」。

從而可以調整樣本數為420個，而其代表性仍應在優等以上。

## 4、分層抽樣

「Google 排行榜」的各搜尋頁之間，自然形成「頁內變異最小、頁間變異最大」的「分層」條件。

而第10搜尋頁以後，「頁間變異」也變小了。

如果作「三十大網站」等主題，抽樣策略可為：

第1搜尋頁之10網頁全部抽選。

其他20網頁，就「質量分割點」前之網頁，分層抽出。

視個案主題與條件，直接就前2-10搜尋頁，分層抽出，亦非不可。

### (四)多關鍵字抽樣

如果研究關照較寬闊，可以採用「聯集式」多關鍵字抽樣。

如果研究關照較收斂，可以採用「交集式」多關鍵字抽樣。

後續實施方法與單關鍵字抽樣方法相同。