

# 如何研究資訊內容

要在毫無頭緒的資料中找出值得一提的「故事」，

就要先找出資料之間的隱藏關係。

如果想要培養這種洞察力，

就要從每一次的實作中累積——弄清資料的性質，

選擇最適合的處理方式。

以下以三個資料型態不同的例子來說明。

# 分析原始量化數據 以台灣各縣市薪資所得觀察為例

案例學習目的

一般取得的數據資料，會經過的基本處理

一份資料可能隱含了各種可能性，但我們只需要把握幾個大原則，就可以從資料中找到其中的特點，接下來我們就用大家都容易取得、也跟大家息息相關的資料：「薪資所得」，來說明一下該怎麼觀察資料。

假設我們今天想了解台灣各城市的薪資所得狀況，首先我們可以從政府開放資料平台上取得2012年所得稅的資料，並從中擷取薪資所得的數據來使用！擷取部份資料整理如下表：

第一頁，總共有二十二個縣市，不過，為了避免一開始太嚇人，我們先取薪資所得前三名的縣市出來解說好了。

把資料貼到Excel裡面，我們就可以開使用圖表工具來畫圖了！

縣市	納稅單位	營利所得	執行業務所得	薪資所得	利息所得	其他所得	稿費所得
台北市	855978	2124152	2342740	54755889	1059584	1974861	54118
新北市	683971	638389	851232	16321660	240179	393989	5152
桃園市	682256	665605	840398	14898259	261261	282914	7062
台中市	104683	52995	45952	2073975	15635	40836	779
台南市	...	...	...	...	...	...	...
高雄市	29423	18243	13618	683705	29423	9806	143
基隆市	3276	2233	796	95851	3276	550	0



### 排排站——總量的比較

將薪資所得的前五名取出來後，我們很快發現，台北市的薪資所得遙遙領先，新北次之，而高雄、台中則表現差不多。這代表什麼呢？台北第一、新北第二嗎？還有哪些資料值得我們繼續發掘？

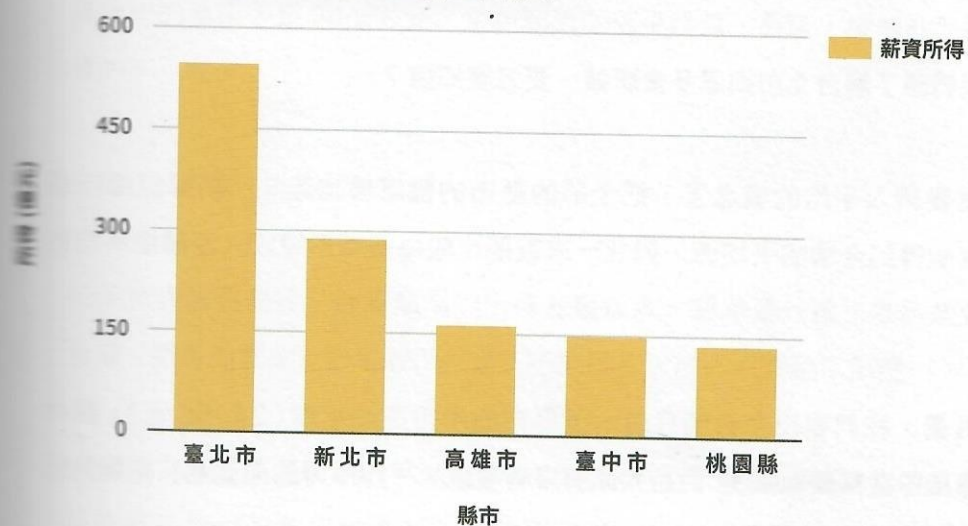
### 站在一樣的起跑點——一致的比較基準

反應快的讀者可能會發現，資料中每個縣市人口數是不一樣的，人口越多，收入總合就越多，因此直接比較是不公平的。所以，接下來我們必須要「標準化」這些數據：將各縣市薪資所得，除以各縣市的納稅單位（納稅人數），即可以得出平均所得。（平均所得 = 薪資所得 ÷ 納稅單位）

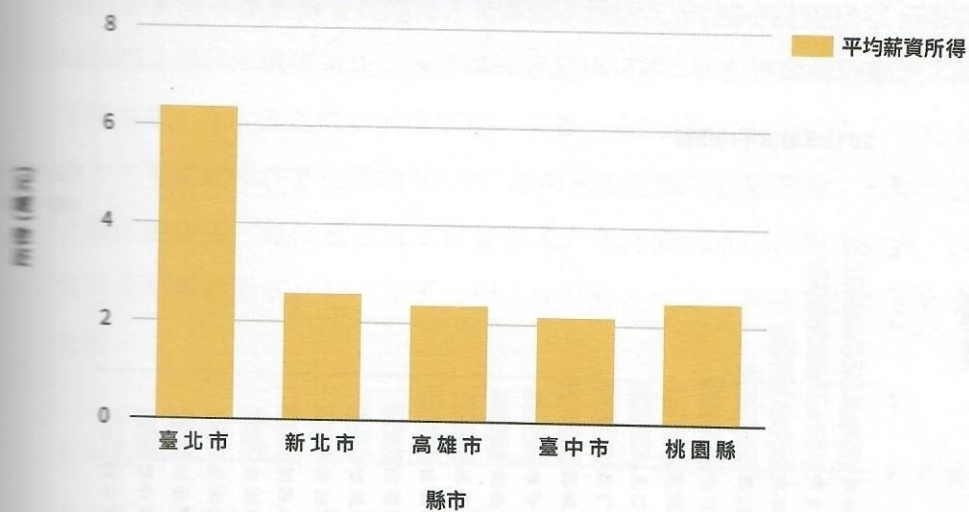
根據平均所得的數據，將會發現台北市還是遙遙領先，此時我們已經可以信心滿滿地宣布：台北市是這五個縣市平均薪資所得最高的城市。另外，從這張圖也可以看出，在總量部份，原本第二名的新北市領先第三名的高雄市幾乎快要一倍，但平均下來看之後，新北市其實和其他縣市的所得水準也是差不多，可見他就是屬於因為人口多，而把總量衝高的類型。

所以，看完上面的例子，請記得「一致的比較基準」很重要，因為這樣的比較才是真的有意義。不過你或許會問：這份資料這樣就用完了嗎？怎麼可能！只要仔細觀察，從資料中是可以看出很多故事的！

### 2012年薪資所得總額前五大縣市



### 2012年薪資所得總額前五大縣市(平均)





## 以中央伍為準：平均值

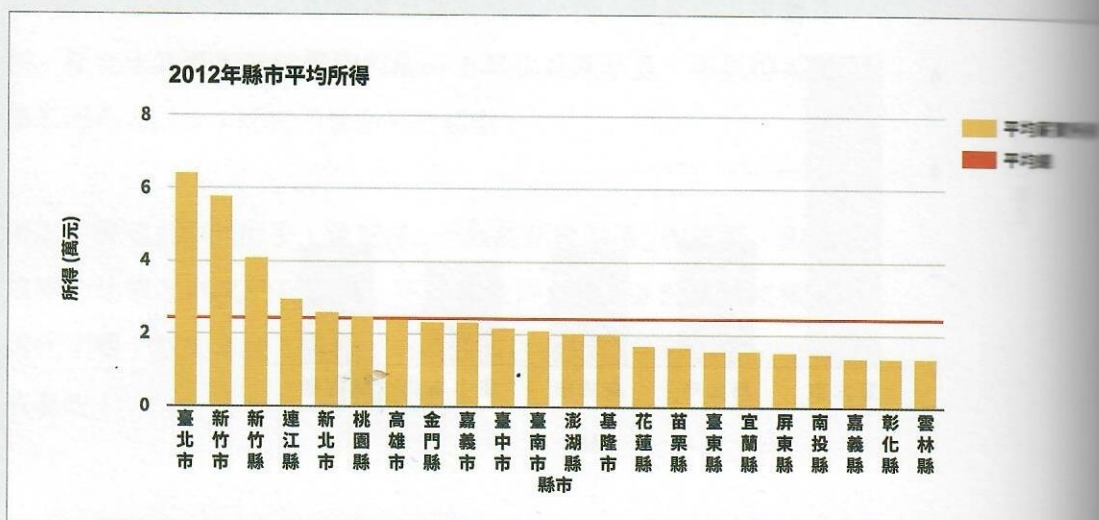
如果我們想了解台北市到底多會賺錢，要怎麼知道？

這時就要帶入平均的概念了，把全部的縣市的數字都加起來，再除以縣市個數，就能得到全國的平均值，如此一來就能比較每個縣市總收入距離總平均的距離。

從這張圖，我們看出全台灣只有六個縣市的所得高於平均（24,156元），顯示出所得集中在某些地區。（而且大家有沒有發現，平均所得的前五名和剛剛的所得總額前五名是不一樣的！）

## 第一排和最後一排的差距：觀察離散程度

再讓我們觀察一下，台北市平均所得最高（63,968元），雲林縣則最低（13,575元），1個台北人的收入接近5個雲林人的收入，哇！是不是滿驚人的。



平均所得與平均線

另外，我們也可以計算出全距是 50,393 元，而全台灣的平均所得，根據剛剛算出來的結果，是 24,156 元，全距超過兩個平均值，看來，城鄉的所得差距真的不小呢。

### A 班 VS. B 班：跨群體的比較

如果我們想要表達，台灣縣市的所得差距很大，標準是什麼？是不是要走出台灣，跟其他國家比較看看？但是，不同國家，收入水準也不同吧……。

沒關係，套用前面學到的，站在同一個起跑點上，我們只要把每個值和平均值的差距除以平均值的大小，就能用百分比的方式來觀察這些差距了！不然，我們借用日本的資料來比較看看好了：

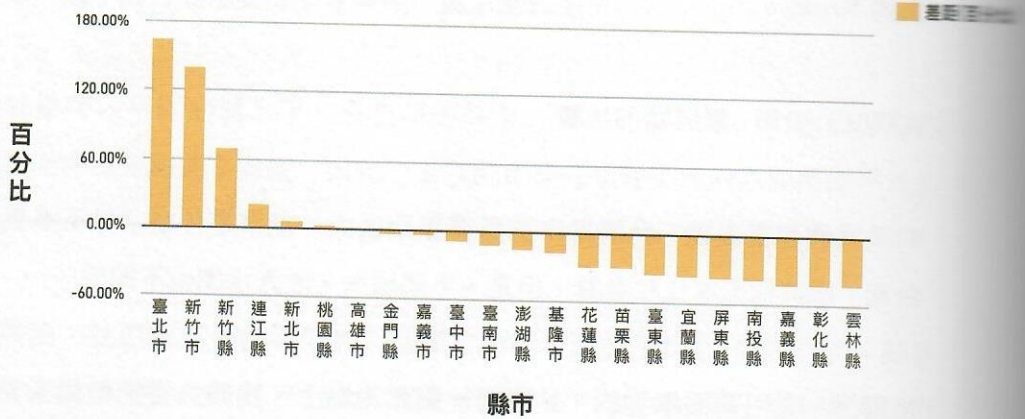
### 日本內政部統計人民經濟狀況數據 [www.esri.cao.go.jp/index.html](http://www.esri.cao.go.jp/index.html)

因為日本的縣市數量達到四十七縣，我們就不一把名稱列出了，不過大家應該都猜得到，遙遙領先、獲得第一名是東京都：從左邊縱軸的單位，就可以了解大概：日本收入最多的東京都，其實也只比平均多出 60%，但台北市的收入，可是比平均多出了約 160%；從所得差距的中心點來看，日本也比較接近圖形的中央，可以看出日本各縣市收入差距變化不如台灣來的大。當然，這只是兩個地區的比較，如果我們去探索更多地區，有機會找到更多有趣的發現。

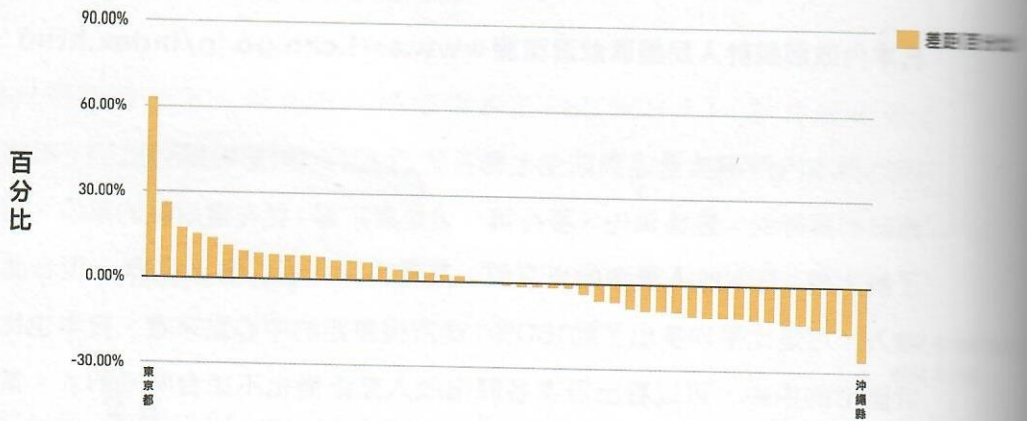
另外說明一下，在這次的案例中，我們用來將資料標準化的是比較直覺的方法，不過在統計上，其實要使用標準差和變異係數的概念來比較，才是較為嚴謹的詮釋方法。



### 2012年年台灣各縣市平均所得差距

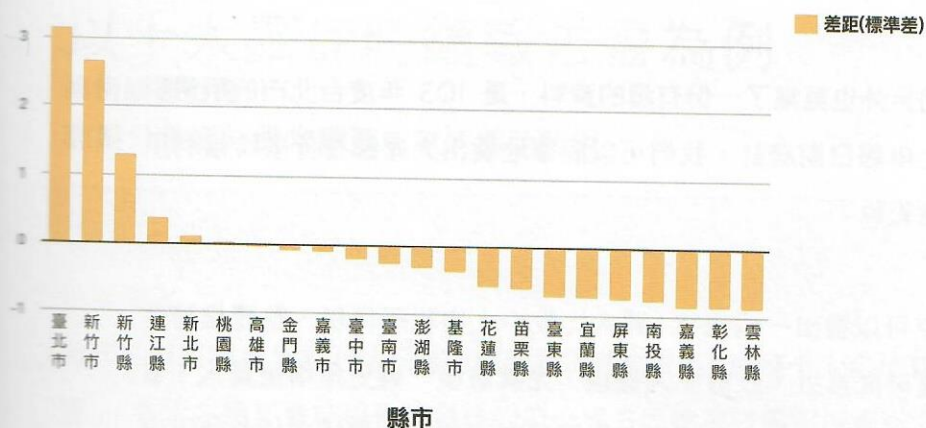


### 2012年日本各縣平均所得差距

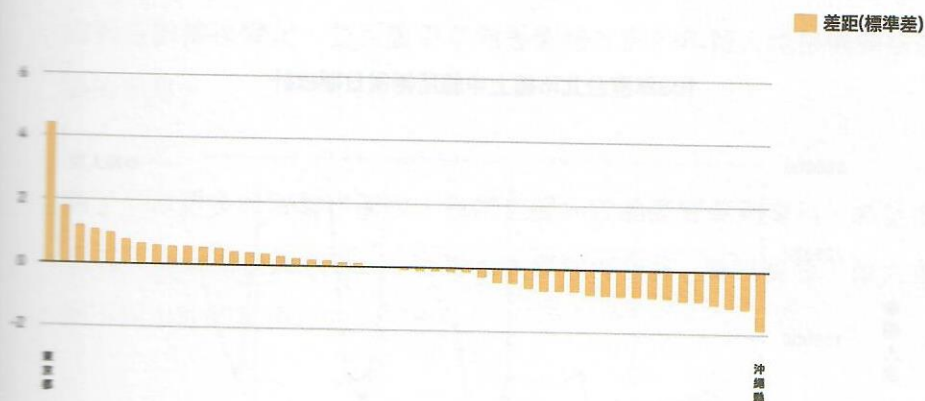


台日平均所得比較

### 2012年年台灣各縣市平均所得差距



### 2012年日本各縣平均所得差距



#### 小教室

#### 什麼是變異係數？

變異係數是一種相對差異量數，用以比較不同單位的資料分散情形。

變異係數 = 標準差 / 平均數 × 100%

台灣平均所得的變異係數是 56.01% 日本平均所得的變異係數 14.19%。

這樣也可以清楚的看出，台灣跟日本的縣市收入離散程度的差距。