

Big Data and Compiler

Homework #1

張家豪 周詮儒
組別：滴滴打__

首先，在這篇報告中，我們將最佳預算分配定義為「最高銷售金額」。以行銷業的角度來看，行銷的目的即為刺激銷售、提高銷售額，而測量行銷成果的方法應為銷售狀況。

一開始先觀察一下原始數據，我們認為對於「客戶參訪」這項，由於客戶的來訪並非是我們能夠控制的，而且也沒有直接的證據顯示「客戶參訪」對於「銷售收入」有影響，所以我們選擇不把這項列入我們的考量；而對於宣傳來看，分類方式有媒體和地區兩種，但是如果兩種都採用的話會導致我們model的independent variable太多，所以我們決定不考慮地區的因素，直接對媒體的種類做回歸分析。首先，我們先把媒體做分類，一共分成書面廣告、戶外看板、網路行銷、手機行銷、電視廣告、現場活動、其他(找不到適當分類)等七類，接著把時間以一個月為一段期間，統計每個月的銷售總額以及花費在各類媒體上的宣傳支出，制定出以下的model：

$$\hat{y} = 0 + 0x_0 - 13.03x_1 - 46.23x_2 + 30.08x_3 + 46.89x_4 - 53.69x_5 + 232.5x_6 + 65.88x_7$$

其中 \hat{y} 代表該月份的銷售額

x_0 為時間序列(單位為月)

x_1 是前一個月在第一類媒體(書面廣告)上的宣傳支出

x_2 是前一個月在第二類媒體(戶外看板)上的宣傳支出

x_3 是前一個月在第三類媒體(網路行銷)上的宣傳支出

x_4 是前一個月在第四類媒體(手機行銷)上的宣傳支出

x_5 是前一個月在第五類媒體(電視廣告)上的宣傳支出

x_6 是前一個月在第六類媒體(現場活動)上的宣傳支出

x_7 是前一個月在第七類媒體(其他)上的宣傳支出

結果如下：

	係數	標準誤	t 統計	P-值
截距	6.05E+09	5.02E+09	1.204945	0.351457
time	-5.6E+07	2.86E+08	-0.19496	0.863432
mt1	-13.0303	19.96354	-0.65271	0.580946
mt2	-46.2319	394.9775	-0.11705	0.917516
mt3	30.07728	280.1851	0.107348	0.924311
mt4	46.89342	80.56577	0.582051	0.619402
mt5	-53.6916	98.05258	-0.54758	0.638924
mt6	232.5023	218.8627	1.06232	0.399399
mt7	65.88029	465.6217	0.141489	0.900449

迴歸統計	
R 的倍數	0.800228
R 平方	0.640365
調整的 R	-0.79818
標準誤	1.29E+09
觀察值個數	11

R平方為64.04%，因此這個model大約可以解釋64.04%的值

根據結果，我們應該要在第六類媒體(現場活動，係數232.5)下最多的宣傳效果會是最好的，另外第二類(戶外看板，係數-13.03)、第三類(戶外看板，係數-46)和第五類(電視廣告，係數-53)在這個模型中對我們的銷售有負面的影響。因此對於明年的預算分配，我們應將現場活動的預算提高，並將戶外看板、網路行銷和電視廣告的預算降低，甚至降為零。此外，我們發現房地產市場在這一年中受到時間序列的影響極小（-5.6E+07），幾乎可以忽略。換句話說，房地產市場在一年中熱季淡季分別不明顯。

對此，我們的解釋是：

1. 全面性放送的行銷活動（廣告）會使房地產的消費者族群卻步、降低購買慾望，使銷售金額下降。因為消費者覺得現在可能是房地產的銷售熱季，應該等到進入淡季後，房屋售價下降後才消費。然而，我們發現沒有實際的熱季、淡季的差別。廣告多少的數量可能是消費者判斷熱季或者淡季的方法之一。
2. 全面性放送的行銷活動不會讓非潛在消費者轉為潛在消費者（成為潛在消費者的門檻高）。
3. 全面性放送的行銷活動在房地產中，難以做到對於個別產品的行銷案。
4. 小地區、地方性的現場行銷活動，較容易吸引潛在消費者，並且較容易對於個別產品推出行銷方法。此外，我們認為產品銷售同樣也會受到地域性的影響，消費者可能偏好在原本居家附近尋找新房子。
5. 手機、網路行銷較容易對於個別廣告接收者（手機、網路使用者）產生個人化廣告、再行銷（Remarketing）或者關鍵字（Keywords）廣告，這類廣告通常能較直接明確的行銷預算用在Target Audience（或潛在消費者）上面。