應變數PAY = (近三個月每個月平均繳費金額) – (近3期平均出帳金額)

大於0很多者為「高潛力用戶」、略大於0者為好用戶、小於0者為壞用戶

首先，將用戶屬性納入解釋變數當中，此部分選定的有CUR\_AGE(目前年齡)、TNR(累積租期)、MAIN\_MSISDN(主門號註記)、NP\_OUT\_GRP(新申裝或攜碼業者)等變數，其餘用戶屬性則因無特殊意義且屬性皆相同或納入後對模型影響甚微而剔除。

接著，在繳款資訊的部分，由於欠費、溢繳等問題可能會與我們的應變數有較高的共線性可能，因此主要選取的變數為同樣是以三個月尺度來看的：N3\_PAID\_CHANNEL\_CATE(近三個月主要繳費通路)

在帳單細項資訊方面，由於各個項目之間重複較多，最後決定以近三個月平均每月的各項消費作為解釋變數，其中，由於國際簡訊費納入模型會不顯著，因此最後選取的變數有：

N3\_VAS\_FEE\_PER\_MON(近三個月平均每個月加值服務消費金額)

N3\_MSG\_FEE\_PER\_MON(近三個月平均每個月的簡訊費)

在帳務資訊上，除了出帳金額與CNTRB\_LVL(出帳時用戶貢獻度)外，我們選擇試圖類別化出帳時經過隱碼的手機廠牌(HANDSET)，以評估其對應變數之影響；而月底資訊同樣考量到有多元共線性之可能，因此最後只選取RATE\_PLAN\_RNG(月底專案實收總資費級距)作為解釋變數。

而在DCB額度資訊上，主要有進線與使用額度兩大類的資料，因此我們選擇各採取一個指標納入模型中，分別為：

L3\_AVG\_MODI\_CNT(近三個月平均每月進線調整次數)

L3\_AVG\_USED\_QUOTA(近三個月每月平均使用額度)

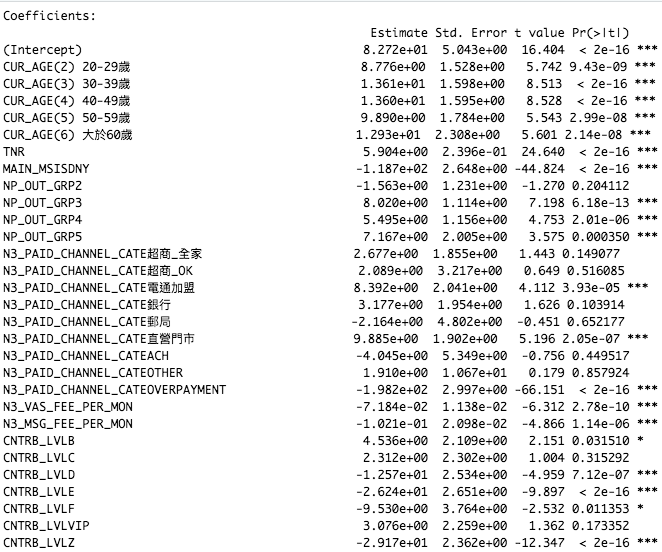
在DCB交易資訊上，則納入了其近三個月的購買狀況，包含了：

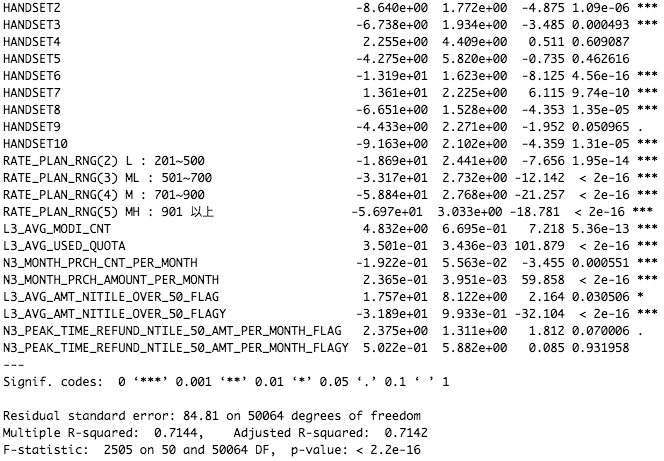
N3\_MONTH\_PRCH\_CNT\_PER\_MONTH(近三個月平均購買次數)

N3\_MONTH\_PRCH\_AMOUNT\_PER\_MONTH(近三個月平均購買金額)

另外，我們也想了解購買程度在全體排名當中的位置與我們所設定的應變數之間的關係，因此採用了L3\_AVG\_AMT\_NITILE\_OVER\_50\_FLAG(是否為近三個月整體DCB用戶消費前50%)，以N為reference level，探討其與NA、Y之間的截距差。

\*N3\_PEAK\_TIME\_REFUND\_NTILE\_50\_AMT\_PER\_MONTH\_FLAG





模型一總結與詮釋及需要改進之處：

在年齡層上，以20歲以下者作為reference level，各年齡層相對於我們所選定之指標皆有顯著差異；而累積租期越長，對指標PAY亦有正面影響；至於有主門號註記者反而使指標下降，推測可能與非主門號手機反而會是進行代收消費的客群有關。被隱碼的新申裝與攜碼客戶經類別化變數後，除了2以外，其餘皆與reference level 1有顯著差異。在繳費通路方面，以711為reference level，可發現電通加盟與直營門市繳費者相對而言有顯著差異，應該是需要積極經營的高潛力客戶。

此外，平均每個月的加值服務消費金額與簡訊消費金額則會使指標PAY下降，可能與消費者的預算限制有所關聯。至於出帳時用戶貢獻度以A作為reference level時與其他level間有顯著性不等的差異，因此可考慮改變參考基準後再次衡量此變數對於應變數的影響。另外，被隱碼的出帳時手機廠牌對指標PAY亦有顯著性不等之差異，建議可進一步針對品牌了解之。在月底專案實收總資費級距的方面，級距越高者會使指標PAY越低，亦是可進一步分析之議題。

在額度資訊上，無論是平均每月進線調整次數與平均每月使用額度對指標PAY而言皆有正面影響，至於購買次數對指標PAY有負面影響、購買金額則又對PAY有正面影響，則可能需再進一步調整模型或分析。在「是否為近三個月整體DCB用戶消費前50%」這個議題上，以N為reference level時，其與NA或Y之間的差距顯著，單就這個變數相比之下，資料為NA或N者或許是更具潛力之用戶。最後，N3\_PEAK\_TIME\_REFUND\_NTILE\_50\_AMT\_PER\_MONTH\_FLAG這個變數在此模型中則不甚顯著。

此模型之R2與 Adjusted R2皆約0.714左右，由於應變數Y之分佈本身有負偏態之情形，導致QQ-plot也有一定程度之偏差，視後續情況進行調整。