

## Univariate Data Analysis Case II: Facebook Case

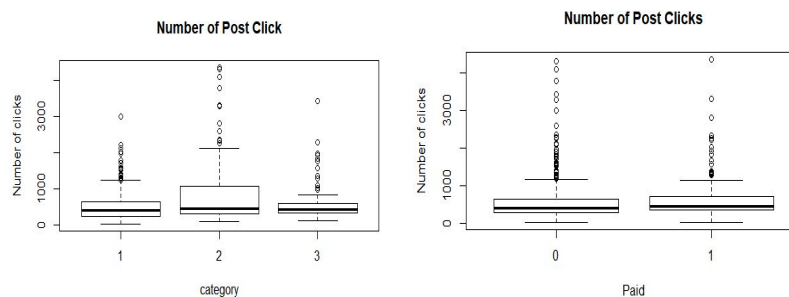
We will try and answer the following questions:

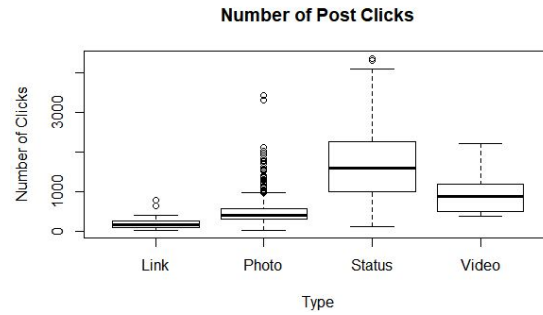
1. Who are the major contributors towards Post Clicks?
2. Is paying for getting post clicks effective?
3. How can you quantify the results obtained for paid effectiveness?

```
> names(facebook)
[1] "Nbr.Post_Clicks" "Category"
[3] "Type"           "Post_month"
[5] "Post_day"       "Post_time"
[7] "Paid"           "Page_total_likes"
> str(facebook)
'data.frame':   499 obs. of  8 variables:
 $ Nbr.Post_Clicks : int  119 1108 132 1386 396 1016 379 422 1250 199 ...
 $ Category        : int   2  2  3  2  2  2  3  3  2  3 ...
 $ Type            : Factor w/ 4 levels "Link","Photo",...: 2  3  2  2  2  3  2 ...
 $ Post_month      : int   12 12 12 12 12 12 12 12 12 12 ...
 $ Post_day        : int    4  3  3  2  2  1  1  7  7  6 ...
 $ Post_time       : Factor w/ 2 levels "A","B": 2  1  2  1  2  1  2  1  2  1 ...
 $ Paid            : int    0  0  0  1  0  0  1  1  0  0 ...
 $ Page_total_likes: int  139441 139441 139441 139441 139441 139441 ...
> unique(Type)
[1] Photo Status Link Video
Levels: Link Photo Status Video
> unique(Post_time)
[1] B A
Levels: A B
```

Looking into the box plot of the number of clicks based on category after converting them into factors.

```
> boxplot(Nbr.Post_Clicks ~ Category, main = "Number of Post Click",
+         xlab = "category", ylab = "Number of clicks")
> boxplot(Nbr.Post_Clicks ~ paid, main = "Number of Post Clicks", xlab =
"Paid", ylab = "Number of clicks")
> boxplot(Nbr.Post_Clicks ~ type, main = "Number of Post Clicks", xlab =
"Type", ylab = "Number of Clicks")
```





We can see that it is most spread out for category 2. It seems like there is not much use in paid ads. The most number of post clicks was received by Status with Photo being the most spread out. The following test quantifies that paying does not really help.

```
> t.test(Nbr.Post_Clicks~Paid)
Welch Two Sample t-test
```

```
data: Nbr.Post_Clicks by Paid
t = -1.1916, df = 243.52,
p-value = 0.2346
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -196.36190  48.33612
sample estimates:
mean in group 0 mean in group 1
    589.9583      663.9712
```