商業分析：SAS / R HW6

108208004 經濟三 白植允

1. 某網紅想分析他 Facebook 上寫的文章。他的文章分為兩種(condition)：建議(tips)和工具(tools)，利用 A/B Testing 課程所教的，畫圖及用檢定方法，幫助網紅分析他的粉絲喜歡哪種文章，以後該網紅應該多寫哪種文章來增加觸擊率。（可自行決定你要分析的面相，如按讚率或分享率等。）

```
  condition mean_click_article
  <fct>                  <dbl>
1 tips                   0.602
2 tools                  0.608
```

```
  condition mean_click_like
  <fct>               <dbl>
1 tips               0.166
2 tools              0.0691
```

```
  condition mean_click_share
  <fct>                <dbl>
1 tips                0.0329
2 tools               0.03
```

=>初步看起來，文章的種類對按讚的比率比較有影響，但就是差別顯不顯著要用額外的 anova 去檢驗。

```
	Welch Two Sample t-test

data:  data[data$condition == "tips", ]$clicked_article and data[data$condition == "tools", ]$clicked_article
t = -1.0867, df = 29998, p-value = 0.2772
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.017195512  0.004928845
sample estimates:
mean of x mean of y 
0.6023333 0.6084667 
```

=>p-value = 0.2772，由文章種類分出來的平均點擊率並無顯著差異 。

```
        Welch Two Sample t-test

data:  data[data$condition == "tips", ]$clicked_like and data[data$condition == "tools", ]$clicked_like
t = 26.426, df = 26449, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.08999062 0.10440938
sample estimates:
 mean of x  mean of y
0.16626667 0.06906667
```

=> p-value < 2.2e^-16，由文章種類分出來的平均按讚率有顯著差異。

```
        Welch Two Sample t-test

data:  data[data$condition == "tips", ]$clicked_share and data[data$condition == "tools", ]$clicked_share
t = 1.4228, df = 29940, p-value = 0.1548
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.001082403  0.006815737
sample estimates:
 mean of x  mean of y
0.03286667 0.03000000
```

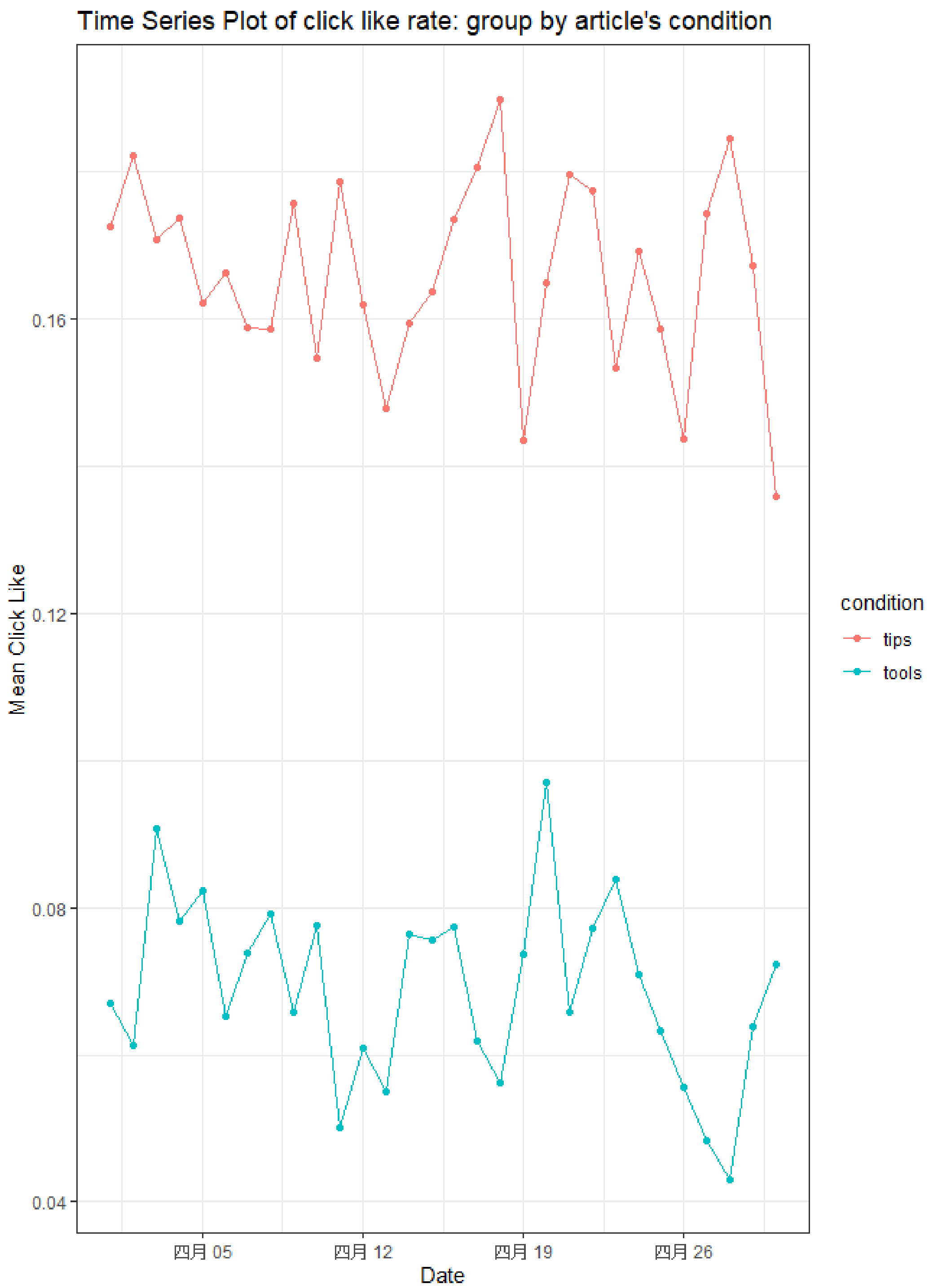
=>p-value = 0.1548，由文章種類分出來的平均分享率並無顯著差異 。

```
> aov.model <- aov(clicked_like ~ visit_date+condition+time_spent_homepage_sec+gender,data)
> summary(aov.model)
                           Df Sum Sq Mean Sq F value Pr(>F)    
visit_date                  1    0.2    0.24   2.403  0.121    
condition                   1   70.9   70.86 698.342 <2e-16 ***
time_spent_homepage_sec     1    0.0    0.00   0.014  0.907    
gender                      3    0.4    0.12   1.208  0.305    
Residuals               29993 3043.2    0.10                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

=> 看起來只有文章種類對按讚與否有顯著影響。

```
> interaction.model <- aov(clicked_like ~ condition*visit_date +condition*time_spent_homepage_sec+condition*gender,data)
> summary(interaction.model)
                                     Df Sum Sq Mean Sq F value Pr(>F)    
condition                             1   70.9   70.86 698.339 <2e-16 ***
visit_date                            1    0.2    0.24   2.372  0.124    
time_spent_homepage_sec               1    0.0    0.00   0.014  0.907    
gender                                3    0.4    0.12   1.207  0.305    
condition:visit_date                  1    0.0    0.01   0.096  0.756    
condition:time_spent_homepage_sec     1    0.0    0.03   0.332  0.564    
condition:gender                      3    0.3    0.10   1.024  0.381    
Residuals                         29988 3042.8    0.10                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

=>考量文章種類與其他變數的交互作用，發現交互作用的影響皆不顯著，得出結論，若要增加按讚比率，多寫建議類的文章是有效的。



=>另外看一下時間序列圖，的確 tips 類的文章按讚率比較高，但值得注意的是，tips 類的文章變異有越來越大的趨勢，也許時間對 tips 類的文章的按讚率有影響。

附錄:R 程式碼

```
library(tidyverse)
library(pwr)
```

```
data <- read.csv("hw6-fb.csv")
str(data)
data$visit_date = as.Date(data$visit_date)
data$condition = as.factor(data$condition)
#data$clicked_article = as.factor(data$clicked_article)
#data$clicked_like = as.factor(data$clicked_like)
#data$clicked_share = as.factor(data$clicked_share)
data$gender = as.factor(data$gender)

data %>%
  group_by(condition) %>%
  summarise(mean_click_article = mean(clicked_article))

data %>%
  group_by(condition) %>%
  summarise(mean_click_like = mean(clicked_like))

data %>%
  group_by(condition) %>%
  summarise(mean_click_share = mean(clicked_share))

t.test(data[data$condition == "tips",]$clicked_article,
      data[data$condition == "tools",]$clicked_article,
      alternative = "two.sided")



t.test(data[data$condition == "tips",]$clicked_like,
      data[data$condition == "tools",]$clicked_like,
      alternative = "two.sided")

t.test(data[data$condition == "tips",]$clicked_share,
      data[data$condition == "tools",]$clicked_share,
      alternative = "two.sided")

aov.model <- aov(clicked_like ~
visit_date+condition+time_spent_homepage_sec+gender,data)
summary(aov.model)
```

```
interaction.model <- aov(clicked_like ~ condition*visit_date
+condition*time_spent_homepage_sec+condition*gender,data)
summary(interaction.model)

daily.clicked_like <- data %>%
  group_by(visit_date,condition) %>%
  summarise(mean_click_like = mean(clicked_like))

ggplot(daily.clicked_like,aes(x=visit_date,y=mean_click_like,c
olor = condition))+
  geom_point()+geom_line()+
  xlab("Date")+ylab("Mean Click Like")+
  ggtitle("Time Series Plot of click like rate: group by
article's condition")+
  theme_bw()
```