商業分析 R/SAS 應用 Homework6

107304023 統計四 張煜均

#不同類型文章的點擊、按讚、分享數、網頁停留時間直方圖

兩種文章各項點擊次數比較圖

```
# A tibble: 2 x 2
  condition mean_time_spent_homepage_sec
  <fct>                            <dbl>
1 tips                              50.0
2 tools                             50.0
```
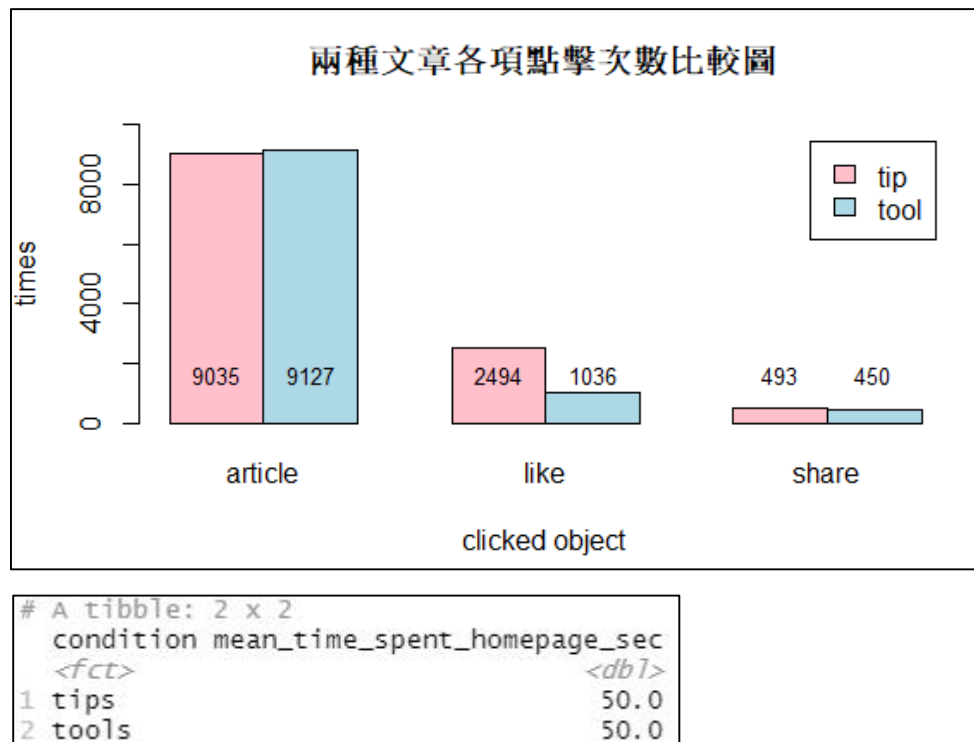
由上圖顯示兩種文章的點擊次數、網頁停留時間皆差不多，tips 類型的文章按讚數較高，文章分享數也稍微多一些，但兩種文章的分享次數差異不大。

#以 condition 分組

```
> table(condition,gender)
         gender
condition female male neutral others
    tips    3781 3833    3716   3670
    tools   3776 3803    3725   3696
```

將兩種文章的粉絲以性別分組，粉絲的性別分布很相近，因此這兩種文章的客群沒有特定性別，為大眾皆會閱讀的文章。

```
> fb_tip %>%
+   group_by(gender) %>%
+   summarise(clicked_article_rate=sum(clicked_article)/n(),
+             clicked_like_rate=sum(clicked_like)/n(),
+             clicked_share_rate=sum(clicked_share)/n())
# A tibble: 4 x 4
  gender  clicked_article_rate clicked_like_rate clicked_share_rate
  <fct>                  <dbl>             <dbl>              <dbl>
1 female                 0.607             0.162             0.0325
2 male                   0.591             0.176             0.0290
3 neutral                0.611             0.159             0.0353
4 others                 0.601             0.168             0.0349
```

```
> #tools
> fb_tool %>%
+   group_by(gender) %>%
+   summarise(clicked_article_rate=sum(clicked_article)/n(),
+             clicked_like_rate=sum(clicked_like)/n(),
+             clicked_share_rate=sum(clicked_share/n()))
# A tibble: 4 x 4
  gender  clicked_article_rate clicked_like_rate clicked_share_rate
  <fct>                  <dbl>             <dbl>              <dbl>
1 female                 0.607            0.0691             0.0302
2 male                   0.616            0.0684             0.0260
3 neutral                0.610            0.0677             0.0301
4 others                 0.601            0.0712             0.0338
```

　　將兩種文章用性別分組，以肉眼觀察其文章點擊數、按讚數、分享數，發現數據皆沒有很大的差異，之後將以 A/B test 來檢測各組數據是否有差異。

```
> aov1 <- aov(time_spent_homepage_sec ~ ., fb) #皆不顯著
> summary(aov1)
                 Df Sum Sq Mean Sq F value Pr(>F)
X                 1      0  0.1360   0.135  0.713
visit_date        1      0  0.3079   0.306  0.580
condition         1      0  0.0609   0.060  0.806
clicked_article   1      0  0.0061   0.006  0.938
clicked_like      1      0  0.0129   0.013  0.910
clicked_share     1      1  0.6701   0.665  0.415
gender            3      0  0.0491   0.049  0.986
Residuals     29990  30203  1.0071
> aov2 <- aov(clicked_article~ ., fb) #皆不顯著
> summary(aov2)
                        Df Sum Sq Mean Sq F value Pr(>F)
X                        1      0  0.3155   1.320  0.251
visit_date               1      0  0.4756   1.990  0.158
condition                1      0  0.0247   0.103  0.748
time_spent_homepage_sec  1      0  0.0014   0.006  0.938
clicked_like             1      0  0.1975   0.827  0.363
clicked_share            1      0  0.0054   0.023  0.880
gender                   3      0  0.1372   0.574  0.632
Residuals            29990   7165  0.2389
```

　　上述檢定結果顯示，頁面停留時間、文章點擊數皆和其他變數不顯著。

```
> aov3 <- aov(clicked_like ~ ., fb) #與visit_date, condition顯著
> summary(aov3)
                          Df Sum Sq Mean Sq F value Pr(>F)
X                          1    0.5    0.53   5.251 0.0219 *
visit_date                 1   49.6   49.55 488.423 <2e-16 ***
condition                  1   21.2   21.17 208.648 <2e-16 ***
time_spent_homepage_sec    1    0.0    0.00   0.013 0.9096
clicked_article            1    0.1    0.08   0.827 0.3632
clicked_share              1    0.3    0.31   3.074 0.0796 .
gender                     3    0.4    0.12   1.203 0.3069
Residuals              29990 3042.6    0.10
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> aov4 <- aov(clicked_share ~ ., fb) #與condition顯著
> summary(aov4)
                          Df Sum Sq Mean Sq F value Pr(>F)
X                          1    0.0 0.00090   0.029 0.8638
visit_date                 1    0.0 0.00693   0.228 0.6334
condition                  1    0.1 0.12276   4.033 0.0446 *
time_spent_homepage_sec    1    0.0 0.02021   0.664 0.4152
clicked_article            1    0.0 0.00061   0.020 0.8876
clicked_like               1    0.1 0.09359   3.074 0.0795 .
gender                     3    0.2 0.06526   2.144 0.0924 .
Residuals              29990  912.9 0.03044
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

檢定結果顯示，文章按讚數和造訪日、文章類型有顯著的關聯；而文章分享數則和文章類型有顯著相關。

```
> #chi squared test 獨立性檢定
> chisq.test(fb$visit_date,fb$condition)

        Pearson's Chi-squared test

data:  fb$visit_date and fb$condition
X-squared = 38.346, df = 29, p-value = 0.1148
```

由卡方獨立性檢定，檢測變數之間是否互相獨立，造訪日和文章類型檢定結果不顯著，兩者之間無顯著關聯性。

```
> chisq.test(fb$clicked_like,fb$condition)

        Pearson's Chi-squared test with Yates' continuity correction

data:  fb$clicked_like and fb$condition
X-squared = 681.57, df = 1, p-value < 2.2e-16
```
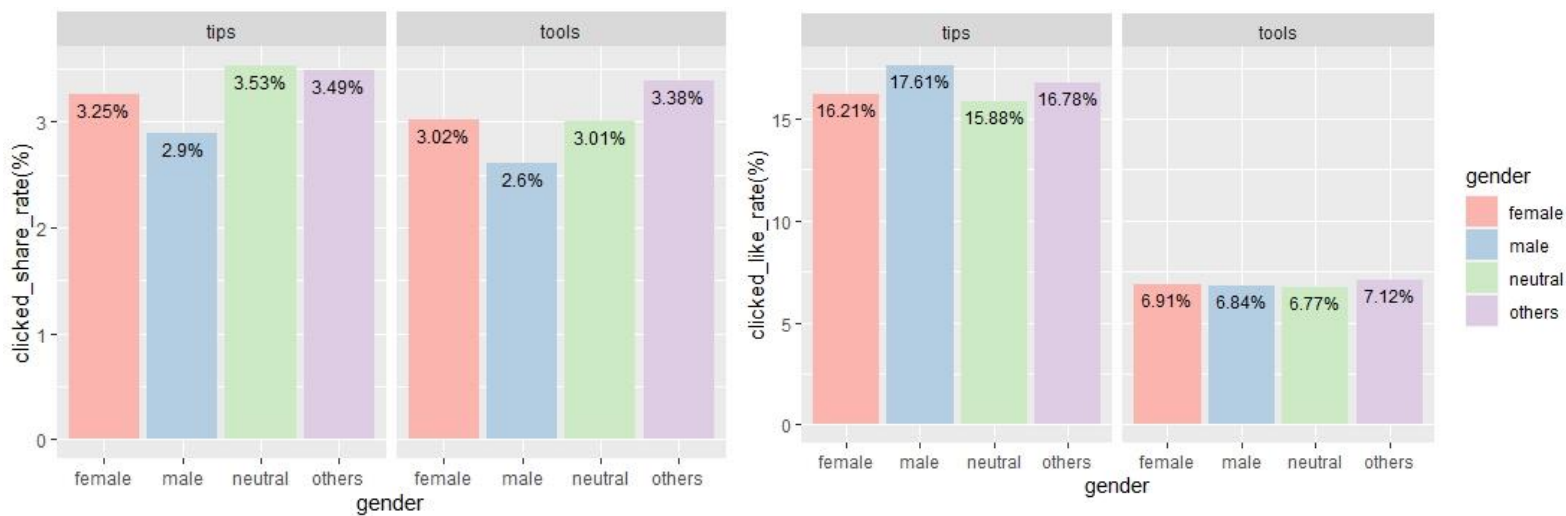
而文章按讚數跟文章類型顯著，兩者的確有關聯。

#依文章和性別分組，計算按讚比例與分享比例並畫圖

#tips 類型的各性別按讚比例與分享比例皆無顯著差異

```
        4-sample test for equality of proportions without continuity
        correction

data:  tip.like.n out of tip.count
X-squared = 4.714, df = 3, p-value = 0.194
alternative hypothesis: two.sided
sample estimates:
    prop 1     prop 2      prop 3      prop 4
0.1621264 0.1761023 0.1587729 0.1678474
```

```
        4-sample test for equality of proportions without continuity
        correction

data:  tip.share.n out of tip.count
X-squared = 2.9872, df = 3, p-value = 0.3936
alternative hypothesis: two.sided
sample estimates:
    prop 1     prop 2      prop 3      prop 4
0.03253108 0.02895904 0.03525296 0.03487738
```

#tools 類型的各性別按讚比例與分享比例皆無顯著差異

```
        4-sample test for equality of proportions without continuity
        correction

data:  tool.like.n out of tool.count
X-squared = 0.39664, df = 3, p-value = 0.9409
alternative hypothesis: two.sided
sample estimates:
    prop 1     prop 2      prop 3      prop 4
0.06912076 0.06836708 0.06765101 0.07115801
```

```
        4-sample test for equality of proportions without continuity
        correction

data:  tool.share.n out of tool.count
X-squared = 3.9166, df = 3, p-value = 0.2706
alternative hypothesis: two.sided
sample estimates:
    prop 1     prop 2      prop 3      prop 4
0.03019068 0.02603208 0.03006711 0.03382035
```

```
         2-sample test for equality of proportions with continuity correction

data:  like.n out of count
X-squared = 681.57, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
 0.08992453 0.10447547
sample estimates:
    prop 1      prop 2
0.16626667 0.06906667

> fbf$condition
[1] tips  tools
Levels: tips tools
```

　　總結上述結果，文章的**按讚數僅和文章類型**有顯著相關，且兩文章的按讚比例顯著不同，即 **tips 類型文章的按讚率顯著高於 tools 類的文章**。而文章的**分享數和文章類型**也有顯著相關，由前方的直方圖可知，**tips 文章的分享數也高於 tools 類的文章。**

#結論

daily viewing time
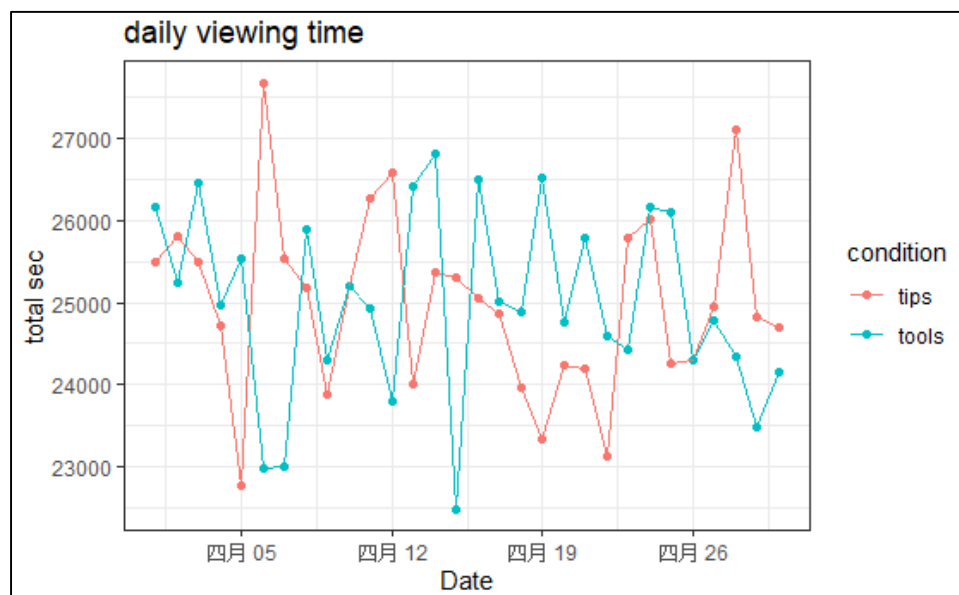
　　上圖為兩種文章每日的總停留時間，tips 類的文章在 4/6 及 4/28 有較高的觀看時間，而這兩天 tools 類的文章觀看時間較短，仔細觀察可發現兩種文章的觀看時長大部分呈現負相關，可能是某種文章曝光時，另一種文章容易被忽略，綜合上述分析，推測該網紅的粉絲**較喜歡 tips 類型文章**，故建議網紅可專注在此類文章，以增加文章的整體曝光度、按讚數及分享數。

```
#程式碼
setwd("C:/Users/USER/Downloads")
fb <- read.csv("hw6-fb.csv")
library(tidyverse)
library(ggplot2)
attach(fb)

names(fb)
str(fb)
fb$gender <- as.factor(gender)
fb$visit_date <- as.Date(visit_date)
fb$condition <- as.factor(condition)

#資料分析
summary(fb)
fb1 <- fb %>%
    group_by(condition) %>%
    summarise(mean_spent_sec=mean(time_spent_homepage_sec),
              article.n=sum(clicked_article),
              like.n=sum(clicked_like),
              share.n=sum(clicked_share))

fb1 = as.matrix(fb1[,3:5])
b1 <- barplot(fb1, names.arg=c('article','like','share'),
col=c("pink","lightblue"),
              xlab="clicked object", ylab="times",
ylim=c(0,10000),beside=T,
              legend.text=c('tip','tool'),main="兩種文章各項點擊次數比
較圖")
text(b1,labels=fb1,y=2,pos=3,offset=1.2,cex=0.8)
```

```r
fb2 <- fb %>%
  group_by(visit_date,condition) %>%
  mutate(total_sec=sum(time_spent_homepage_sec))

ggplot(fb2, aes(x = visit_date, y = total_sec, colour = condition)) +
  geom_point() + geom_line() +
  xlab("Date") + ylab("total sec") +
  ggtitle(label="daily viewing time") +
  theme_bw()

#分組
table(condition,gender)
fb_tip <- fb %>% filter(fb$condition=="tips")
fb_tool <- fb %>% filter(fb$condition=="tools")

#tips
fb_tip %>%
  group_by(gender) %>%
  summarise(clicked_article_rate=sum(clicked_article)/n(),
            clicked_like_rate=sum(clicked_like)/n(),
            clicked_share_rate=sum(clicked_share)/n())
fb_tip %>%
  group_by(gender) %>%

summarise(mean_time_spent_homepage_sec=mean(time_spent_home
page_sec))

#tools
fb_tool %>%
```

```r
  group_by(gender) %>%
  summarise(clicked_article_rate=sum(clicked_article)/n(),
            clicked_like_rate=sum(clicked_like)/n(),
            clicked_share_rate=sum(clicked_share/n()))
fb_tool %>%
  group_by(gender) %>%

summarise(mean_time_spent_homepage_sec=mean(time_spent_homepage_sec))

#anova
#檢定網頁停留時間與什麼變數顯著
aov1 <- aov(time_spent_homepage_sec ~ ., fb) #皆不顯著
aov2 <- aov(clicked_article~ ., fb) #皆不顯著
aov3 <- aov(clicked_like ~ ., fb) #與 visit_date, condition 顯著
aov4 <- aov(clicked_share ~ ., fb) #與 condition 顯著

summary(aov1)
summary(aov2)
summary(aov3)
summary(aov4)

#chi squared test  獨立性檢定
chisq.test(fb$visit_date,fb$condition)
chisq.test(fb$clicked_like,fb$condition)

#依文章和性別分組，計算案讚比例與分享比例
fb1 <- fb %>%
  group_by(condition,gender) %>%
  summarise(clicked_article_rate=sum(clicked_article)/n()*100,
```

```r
                clicked_like_rate=sum(clicked_like)/n()*100,
                clicked_share_rate=sum(clicked_share)/n()*100)

ggplot(fb1, aes(x=gender,y=clicked_like_rate,fill=gender))+
    geom_col()+

geom_text(aes(label=paste0(round(clicked_like_rate,2),"%")),vjust=1.5,size=3.1)+
    ylab("clicked_like_rate(%)")+
    facet_wrap(~condition)+
    scale_fill_brewer(palette = "Pastel1")

ggplot(fb1, aes(x=gender,y=clicked_share_rate,fill=gender))+
    geom_col()+

geom_text(aes(label=paste0(round(clicked_share_rate,2),"%")),vjust=1.5,size=3.1)+
    ylab("clicked_share_rate(%)")+
    facet_wrap(~condition)+
    scale_fill_brewer(palette = "Pastel1")

#分組中各性別按讚比例與分享比例是否有顯著差異
fbb <- fb %>%
    group_by(condition,gender) %>%
    summarise(like_n=sum(clicked_like),
              share_n=sum(clicked_share), count=n())

tip.like.n = as.numeric(fbb[1:4,3]$like_n)
tip.count = as.numeric(fbb[1:4,5]$count)
prop.test(tip.like.n,tip.count)
```

```r
tip.share.n = as.numeric(fbb[1:4,4]$share_n)
tip.count = as.numeric(fbb[1:4,5]$count)
prop.test(tip.share.n,tip.count)


tool.like.n = as.numeric(fbb[5:8,3]$like_n)
tool.count = as.numeric(fbb[5:8,5]$count)
prop.test(tool.like.n,tool.count)


tool.share.n = as.numeric(fbb[5:8,4]$share_n)
tool.count = as.numeric(fbb[5:8,5]$count)
prop.test(tool.share.n,tool.count)

#兩種文章按讚比例是否顯著不同
fbf <- fb %>%
    group_by(condition) %>%
    summarise(like_n=sum(clicked_like), count=n())

like.n = as.numeric(fbf$like_n)
count = as.numeric(fbf$count)
prop.test(like.n,count)
fbf$condition
```