

變異數分析與實驗設計期末報告

壹、研究動機

首先，根據犯罪心理學家Simon Field 在1922發表的[論文](#)指出在英國，隨著氣溫增加，犯罪率也會跟著提升。再者，ROBERT J. SAMPSON與其他作者於2012年聯合發表的[論文](#)則認為婚姻乃人們腳色的轉變，以及重新建構與他人的關係的轉捩點，從而抑制犯罪。故個人想藉由變異數分析與實驗設計的方法分析上述潛在犯罪因子（氣溫、結婚率）並驗證台灣是否存在上述現象。

貳、資料來源

犯罪資料來自內[政部警政署資料開放平台](#)，各縣市平均氣溫則來自[觀測資料查詢系統](#)。

肆、因子設計

- 犯罪種類 (crime) : drug (毒品) / theft (竊盜) / robbery (搶劫) / rape (強姦)
- 各縣市月均溫 (temp) : 15~20 (低溫) / 20~25 (合宜溫度) / 25~30 (高溫)
- 各縣市結婚率 : 4.5 ~ 4.83 (低結婚傾向) / 4.83 ~ 5.16 (中等婚傾向) / 5.16 ~ 5.5 (高結婚傾向)

伍、實驗設計

設計3個變量，共3, 3, 4個 level

共計 $3 \times 3 \times 4 = 36$ 個cell

陸、模型假設(Fix Effect)

$$\log(\mu_{ijkl}) = \mu + \text{crime}_i + \text{marry}_j + \text{temp}_k + \text{crime} * \text{marry}_{ij} + \text{crime} * \text{temp}_{ik} + \text{marry} * \text{temp}_{kl} + \text{crime} * \text{marry} * \text{temp}_{ijk} + \epsilon_{ijkl}$$

$$\begin{cases} i = \text{drug, theft, robbery, rape} \\ j = 15 \sim 20, 20 \sim 25, 25 \sim 30 \\ k = 4.5 \sim 4.83, 4.83 \sim 5.16, 5.16 \sim 5.5 \\ l = \text{replications in each cell} \end{cases}$$

$$\epsilon_{ijkl} \sim N(0, \sigma^2)$$

柒、實驗結果

- 殘差分析：
資料點皆接近45度Q-Qline[見附錄圖1](#)，並無嚴重違反殘差的分配假設。
- 模型顯著性分析：
所有變數的main effect皆為顯著。此外，溫度與犯罪率之間有著二次方關係，但模型中存在交互作用（結婚率與犯罪種類）之影響[見附錄圖2](#)。因此分析重點將著重於main effect與interaction plot的搭配。
- 邊際顯著性：
所有的變數除了犯罪種類以外，其邊際檢定（結婚率 / 溫度）皆為不顯著[見附錄圖3](#)。由此可知，在探討與解讀變數的效應時必須同時考量其餘變數給定下的效果。

捌、結果詮釋

- 在先前的[研究](#)中提及氣溫與犯罪率的連結在於人們的活動程度。在英國，合宜的氣溫下，犯罪率將達致最高點。而[BBC事實和查](#)中則把焦點放在人們因溫度而起的生理反應使得暴力犯罪率上升（強姦、強盜）而竊盜犯罪率則會下降，因為高溫下會使人們待在家裡，使竊賊無機可乘。以下將藉由變異數分析與實驗設計的方法來驗證台灣的犯罪資料是否支持以上論點，以及毒品犯罪率是否與溫度有關係。
 - 竊盜案的發生是因為人們不在家而有機乘的論點只有在低結婚率的城市中（結婚率4.5~4.83）成立[見附錄圖7](#)(左一)。甚至在中等結婚率的城市中，理論上最多人在家的氣溫條件下，犯罪率最高，呈現一個U字形[見附錄圖7](#)(左二)。個人推測，先前研究不成立的可能原因為研究的緯度不同。因為台灣的極端冷熱溫度並沒有高緯度國家明顯，造成該英國研究的現象在台灣可能不成立。
 - 溫度使心情不穩定，從而提高犯毒品罪可能：

在高結婚率與低結婚率[見附錄圖4](#)(左一/左三)之下，毒品犯罪率隨著溫度上升而急遽增加，呈現二次關係[見附錄圖4](#)與[附錄圖2](#)（temperature : quadratic）。該現象的可能成因為：溫度的上升會使人們感到幸福感與認知能力下降（[BBC事實和查](#)）。故能帶來快感的毒品在高結婚率與低結婚率之下[見附錄圖4](#)（左一/左三），對癮君子產生的幸福感的誘惑將會倍增，從而造成毒品犯罪率的上升。
 - 中等結婚率下的毒品犯罪率對溫度反應敏感：

在中等結婚率下，只要溫度過高或過低，毒品的犯罪率皆會飆升[見附錄圖4](#)(左二)。
 - 毒品犯案率與在控制溫度下，其趨勢與竊盜犯罪率相近[見附錄圖5](#)：

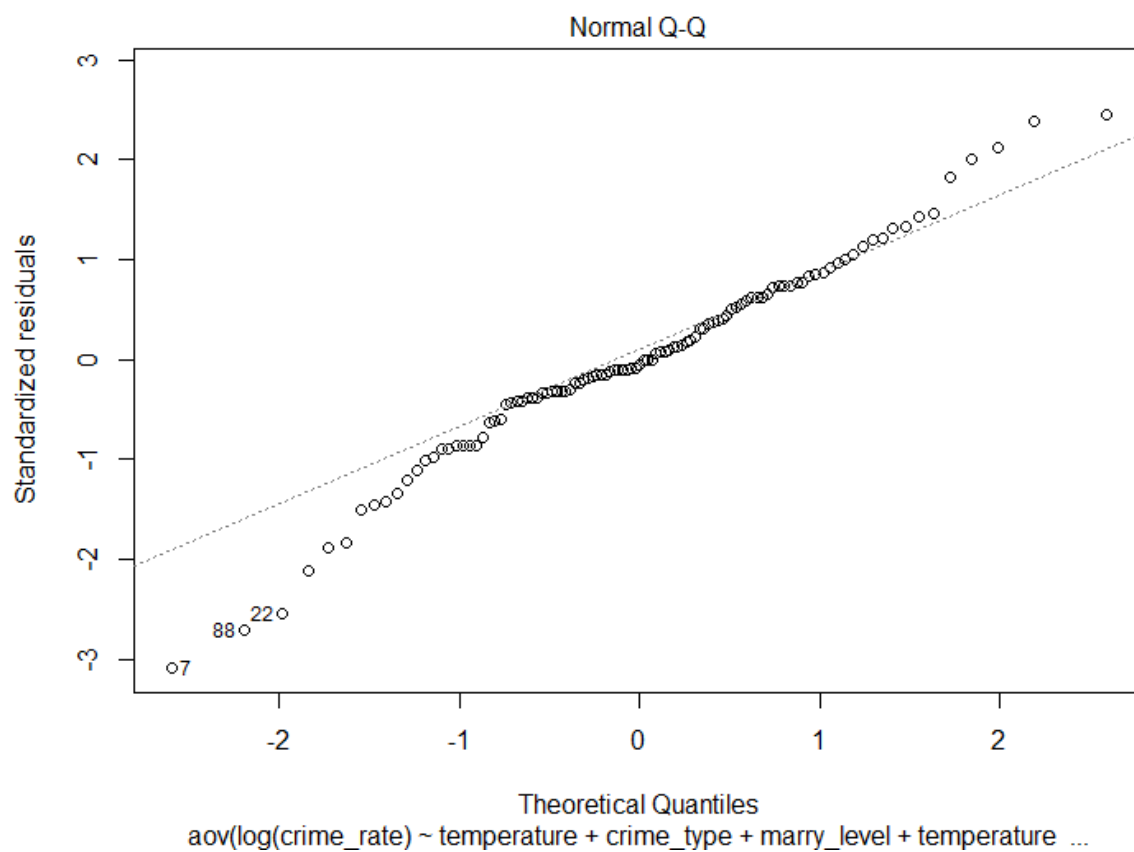
造成這種現象的可能的解釋原因為犯罪者為了購買毒品但金錢不足從而導致竊盜案的發生。
- 除了溫度之外，根據先前[研究](#)，結婚率與犯罪率呈現負向的關係。此外，實驗結果也顯示婚姻與犯罪種類之前具有交互作用。故以下將檢視實驗結果是否與上述論點一致，並額外討論不同犯罪類別與婚姻之間的交互作用。
 - 給定溫度之下，犯罪率並不會單純隨著結婚率上升而下降（[見附錄圖6與圖7](#)）：

如給定高溫下（25~30），偷竊行為甚至會隨著結婚率而上升。該現象的可能解釋為婚姻會造成經濟壓力，迫使人們為了維持家計挺而走險。
 - 在高溫下，結婚率與強姦案和強盜案的交互作用最為強烈：

先前提到，溫度的升高可加強人們犯下強姦的作案動機(如心率增加、睪丸素激增)。而結婚率則可成為降低人們作案動機的力量(有穩定伴侶發生關係)造成強姦案件在高結婚率下的急遽下降。促使強姦犯罪與強盜犯罪的cell mean在高溫與高結婚率下產生極大的差距產生交互作用。[見附錄圖6](#)(左三)。

玖、附錄

圖一



圖二

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
temperature	2	3.3	1.67	3.532	0.0345	*
temperature: Linear	1	0.8	0.81	1.713	0.1948	
temperature: Quadratic	1	2.5	2.52	5.351	0.0236	*
crime_type	3	483.6	161.20	341.858	<2e-16	***
marry_level	2	3.5	1.77	3.764	0.0279	*
temperature:crime_type	6	3.6	0.59	1.258	0.2880	
temperature:crime_type: Linear	3	2.9	0.98	2.071	0.1118	
temperature:crime_type: Quadratic	3	0.6	0.21	0.445	0.7218	
temperature:marry_level	4	2.0	0.50	1.068	0.3788	
temperature:marry_level: Linear	2	0.0	0.01	0.030	0.9707	
temperature:marry_level: Quadratic	2	2.0	0.99	2.106	0.1292	
crime_type:marry_level	6	7.6	1.26	2.671	0.0215	*
temperature:crime_type:marry_level	12	6.5	0.54	1.143	0.3410	
temperature:crime_type:marry_level: Linear	6	2.3	0.39	0.820	0.5580	
temperature:crime_type:marry_level: Quadratic	6	4.1	0.69	1.465	0.2025	
Residuals	71	33.5	0.47			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

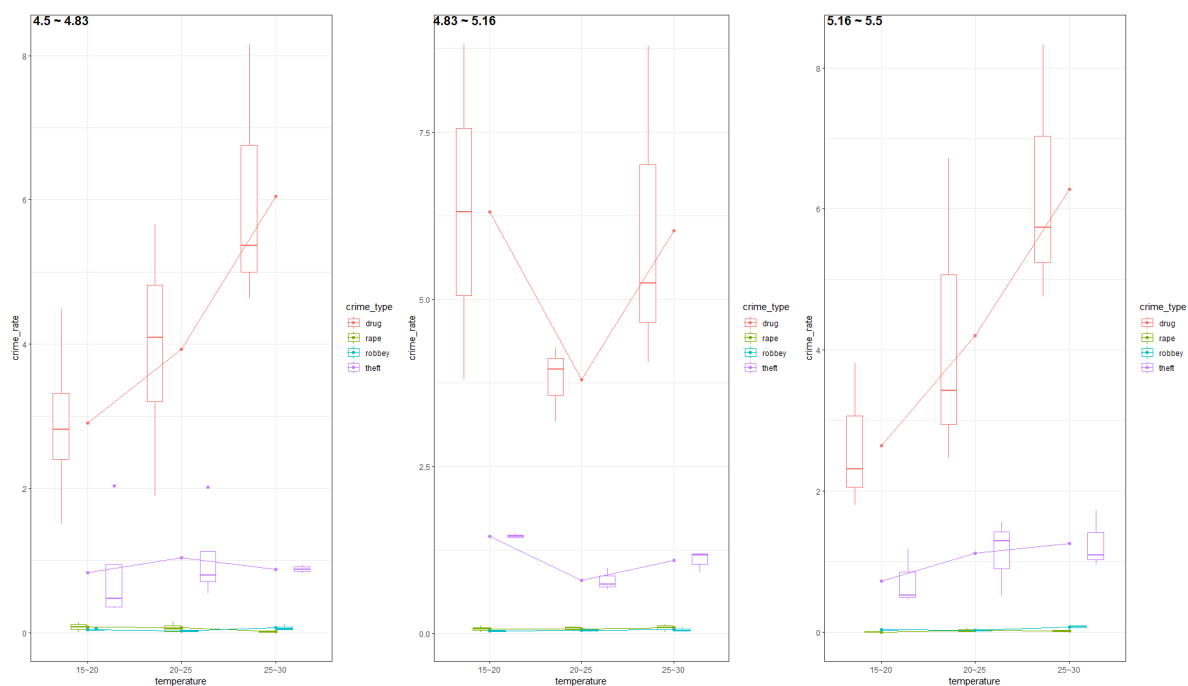
圖三

```

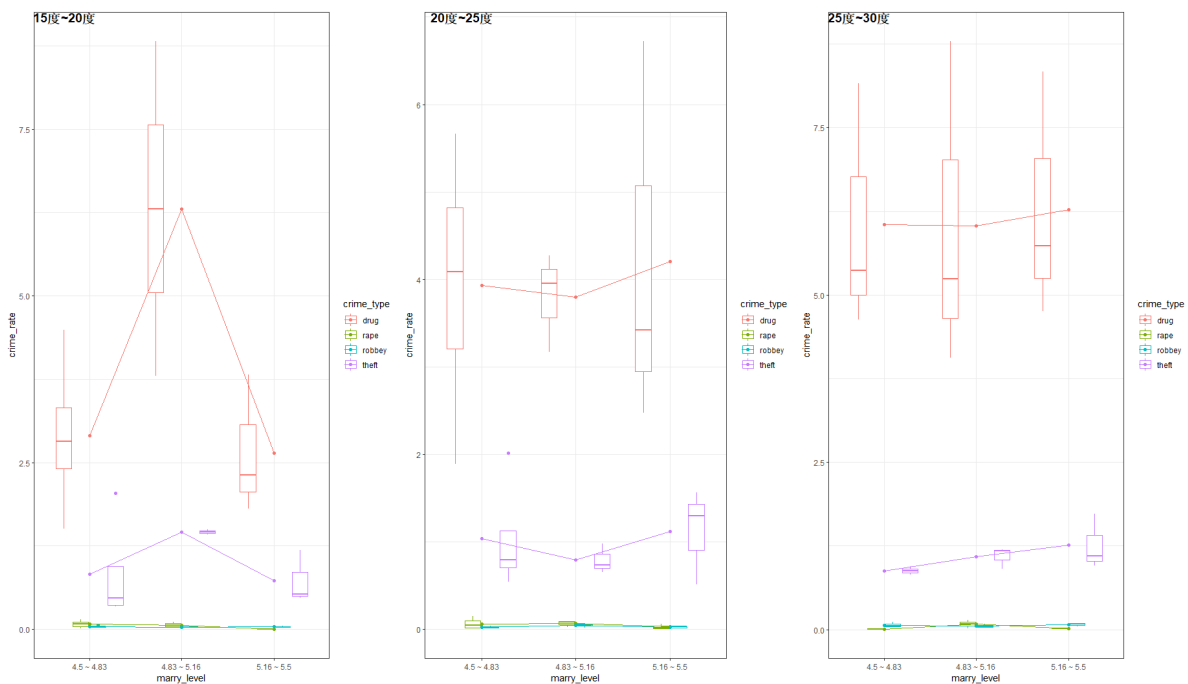
> summary(aov(log(crime_rate)~temperature, data = df))
            Df Sum Sq Mean Sq F value Pr(>F)
temperature  2    3.3   1.665    0.321  0.726
Residuals   104   540.2    5.195
> summary(aov(log(crime_rate)~crime_type, data = df))
            Df Sum Sq Mean Sq F value Pr(>F)
crime_type   3   484.2   161.39   279.9 <2e-16 ***
Residuals   103    59.4     0.58
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(aov(log(crime_rate)~marry_level, data = df))
            Df Sum Sq Mean Sq F value Pr(>F)
marry_level  2     2.0    0.990    0.19  0.827
Residuals   104   541.6    5.208

```

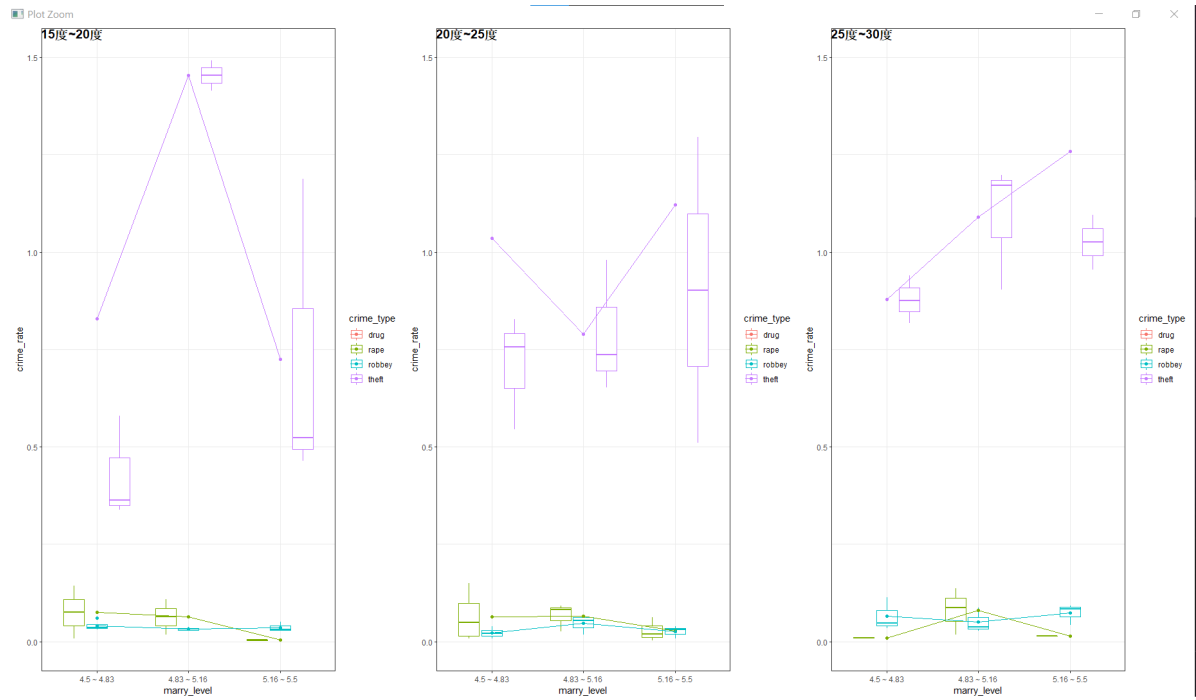
圖四



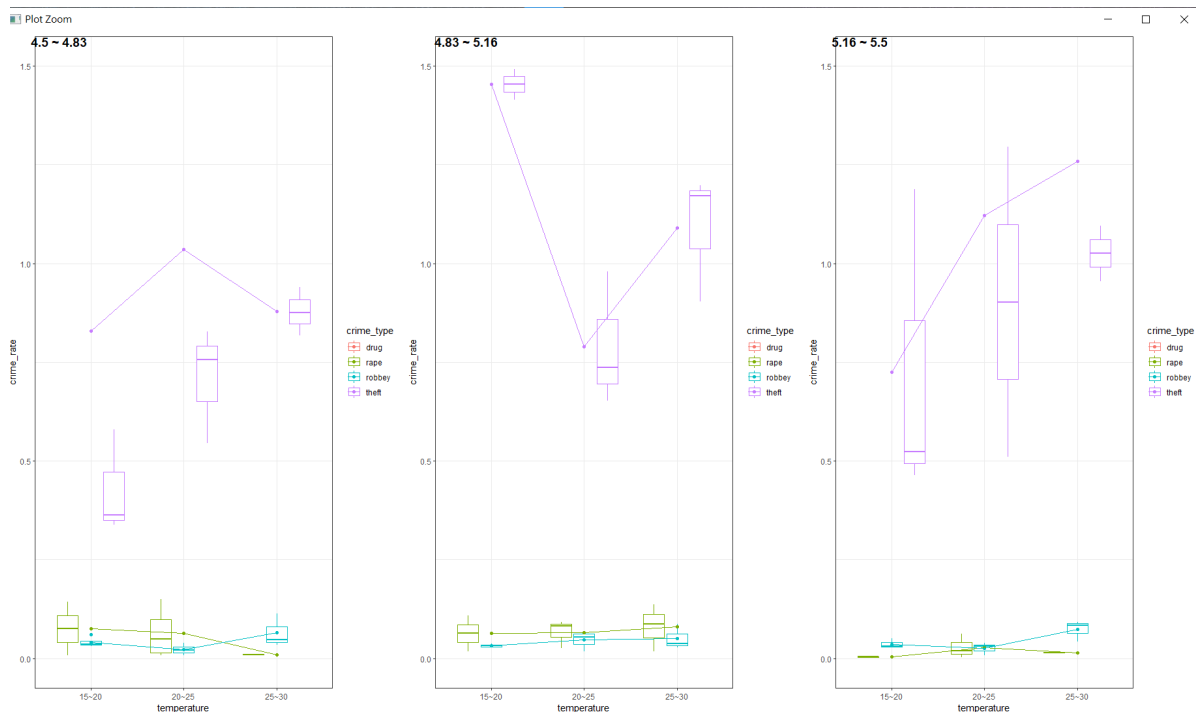
圖五



圖六



圖七



圖八

	<code>marry_level</code>	<code>crime_type</code>	<code>cellmean_temp_15_20</code>	<code>replication</code>	<code>standatd_error</code>
1	4.5 ~ 4.83	drug	2.906209577	4	1.493090e+00
2	4.5 ~ 4.83	rape	0.075162235	3	4.551515e-03
3	4.5 ~ 4.83	robbery	0.041219019	4	1.820632e-04
4	4.5 ~ 4.83	theft	0.829770760	4	6.599228e-01
5	4.83 ~ 5.16	drug	6.305782242	2	1.259194e+01
6	4.83 ~ 5.16	rape	0.063628390	2	4.085767e-03
7	4.83 ~ 5.16	robbery	0.032033454	2	4.659580e-05
8	4.83 ~ 5.16	theft	1.453793227	2	3.049285e-03
9	5.16 ~ 5.5	drug	2.645381944	3	1.098398e+00
10	5.16 ~ 5.5	rape	0.003693845	2	4.424430e-08
11	5.16 ~ 5.5	robbery	0.036381075	3	1.396816e-04
12	5.16 ~ 5.5	theft	0.725285441	3	1.610232e-01

圖九

	marry_level	crime_type	cellmean_temp_20_25	replication	standatd_error
1	4.5 ~ 4.83	drug	3.93287056	4	2.5420607778
2	4.5 ~ 4.83	rape	0.06383550	4	0.0044206254
3	4.5 ~ 4.83	robbery	0.02223136	3	0.0002356254
4	4.5 ~ 4.83	theft	1.03643981	4	0.4417251035
5	4.83 ~ 5.16	drug	3.79889817	3	0.3216449063
6	4.83 ~ 5.16	rape	0.06635951	3	0.0012916479
7	4.83 ~ 5.16	robbery	0.04717880	3	0.0006708415
8	4.83 ~ 5.16	theft	0.78998103	3	0.0287915403
9	5.16 ~ 5.5	drug	4.20571885	3	4.9681891358
10	5.16 ~ 5.5	rape	0.02813854	3	0.0009040157
11	5.16 ~ 5.5	robbery	0.02544827	3	0.0002671922
12	5.16 ~ 5.5	theft	1.12179769	3	0.2978605675

圖十

	marry_level	crime_type	cellmean_temp_25_30	replication	standatd_error
1	4.5 ~ 4.83	drug	6.050709872	3	3.467246e+00
2	4.5 ~ 4.83	rape	0.009711228	2	6.599434e-06
3	4.5 ~ 4.83	robbery	0.064910451	3	1.759473e-03
4	4.5 ~ 4.83	theft	0.878205011	3	3.684889e-03
5	4.83 ~ 5.16	drug	6.028245151	3	6.063836e+00
6	4.83 ~ 5.16	rape	0.080411158	3	3.487955e-03
7	4.83 ~ 5.16	robbery	0.050278979	3	1.021995e-03
8	4.83 ~ 5.16	theft	1.090521540	3	2.651697e-02
9	5.16 ~ 5.5	drug	6.275417053	3	3.418488e+00
10	5.16 ~ 5.5	rape	0.014775380	2	7.079087e-07
11	5.16 ~ 5.5	robbery	0.073188163	3	7.202210e-04
12	5.16 ~ 5.5	theft	1.258816782	3	1.681002e-01

程式碼

```

# import data####
library(dplyr)
library(ggplot2)
library(plyr)
setwd("C:/Users/user/Desktop/R/DOE/Data")
df = readxl::read_xlsx("aov_data.xlsx")

# function####
marry_level_dum = function(x){
  if(x == "4.5 ~ 4.83"){
    return(1)
  }
  else if(x == "4.83 ~ 5.16"){
    return(2)
  }
  else{return(3)}
}

inte_plot_temp = function(temper, df){
  df_test = df %>% filter(temperature == temper)
  df_int = ddply(df_test,.(marry_level,crime_type),summarise, val =
mean(crime_rate)) # cell mean

  res = ggplot(df_test, aes(x = marry_level, y = crime_rate, colour =
crime_type)) +
    geom_boxplot() +
    geom_point(data = df_int, aes(y = val)) +
    geom_line(data = df_int, aes(y = val, group = crime_type)) +
    theme_bw()

```

```

    return(res)
}
inte_plot_temp_test = function(temper, df){
  df_test = df %>% filter(temperature == temper) # cell mean
  df_int = ddply(df_test,.(marry_level,crime_type),summarise, val =
mean(crime_rate)) # cell mean

  res = ggplot(df_test, aes(x = marry_level, y = crime_rate, colour =
crime_type)) +
    geom_boxplot() +
    geom_point(data = df_int, aes(y = val)) +
    geom_line(data = df_int, aes(y = val, group = crime_type)) +
    theme_bw()+ scale_y_continuous(limits = c(0, 1.5))
  return(res)
}
inter_plot_marr = function(marr, df){
  df_test = df %>% filter(marry_level == marr)
  df_int = ddply(df_test,.(temperature,crime_type),summarise, val =
mean(crime_rate)) # cell mean

  res2 = ggplot(df_test, aes(x = temperature, y = crime_rate, colour =
crime_type)) +
    geom_boxplot() +
    geom_point(data = df_int, aes(y = val)) +
    geom_line(data = df_int, aes(y = val, group = crime_type)) +
    theme_bw()
  return(res2)
}
inter_plot_marr_test = function(marr, df){
  df_test = df %>% filter(marry_level == marr)
  df_int = ddply(df_test,.(temperature,crime_type),summarise, val =
mean(crime_rate)) # cell mean

  res2 = ggplot(df_test, aes(x = temperature, y = crime_rate, colour =
crime_type)) +
    geom_boxplot() +
    geom_point(data = df_int, aes(y = val)) +
    geom_line(data = df_int, aes(y = val, group = crime_type)) +
    theme_bw() + scale_y_continuous(limits = c(0, 1.5))
  return(res2)
}
df$marr_dummy = sapply(df$marry_level, marry_level_dum) %>% as.vector()
df$temperature = as.factor(df$temperature)
df$crime_type = as.factor(df$crime_type)
df$marry_level = as.factor(df$marry_level)

# perform anova####
m = aov(log(crime_rate) ~ temperature + crime_type +
  marry_level + temperature*crime_type +
  temperature*marry_level + marry_level*crime_type +
  temperature * crime_type * marry_level
, data = df)

summary(m)

summary(aov(log(crime_rate)~temperature, data = df))
summary(aov(log(crime_rate)~crime_type, data = df))
summary(aov(log(crime_rate)~marry_level, data = df))

```

```

plot(m)

plot(TukeyHSD(m, "temperature"))
plot(TukeyHSD(m, "marry_level"))

# contrasts(df$rate) = contr.poly(3)
summary.aov(m, split = list(temperature = list("Linear"=1, "Quadratic" = 2)))
# summary.lm(m, split = list(temperature = list("Linear"=1, "Quadratic" = 2)))

lm(log(crime_rate) ~ temperature + crime_type +
    marry_level + temperature*crime_type +
    temperature*marry_level + marry_level*crime_type +
    temperature * crime_type * marry_level

    ,data = df
    ,contrasts = list(temperature = contr.treatment(3)))

# interaction plot####

low_temp_plot = inte_plot_temp("15~20", df)
mid_temp_plot = inte_plot_temp("20~25", df)
high_temp_plot = inte_plot_temp("25~30", df)

plot_grid(low_temp_plot, mid_temp_plot, high_temp_plot ,
    labels = c("15度~20度", "20度~25度", "25度~30度"),
    ncol =3, nrow = 1)

low_marr_plot = inter_plot_marr("4.5 ~ 4.83", df)
mid_marr_plot = inter_plot_marr("4.83 ~ 5.16", df)
high_marr_plot = inter_plot_marr("5.16 ~ 5.5", df)

plot_grid(low_marr_plot, mid_marr_plot, high_marr_plot ,
    labels = c("4.5 ~ 4.83", "4.83 ~ 5.16", "5.16 ~ 5.5"),
    ncol =3, nrow = 1)

low_marr_plot_test = inter_plot_marr_test("4.5 ~ 4.83", df)
mid_marr_plot_test = inter_plot_marr_test("4.83 ~ 5.16", df)
high_marr_plot_test = inter_plot_marr_test("5.16 ~ 5.5", df)

plot_grid(low_marr_plot_test, mid_marr_plot_test, high_marr_plot_test ,
    labels = c("4.5 ~ 4.83", "4.83 ~ 5.16", "5.16 ~ 5.5"),
    ncol =3, nrow = 1)

low_temp_plot_test = inte_plot_temp_test("15~20", df)
mid_temp_plot_test = inte_plot_temp_test("20~25", df)
high_temp_plot_test = inte_plot_temp_test("25~30", df)

plot_grid(low_temp_plot_test, mid_temp_plot_test, high_temp_plot_test ,
    labels = c("15度~20度", "20度~25度", "25度~30度"),
    ncol =3, nrow = 1)

# cell mean####
df_temp_low = df %>% filter(temperature=="15~20")
df_temp_mid = df %>% filter(temperature=="20~25")
df_temp_high = df %>% filter(temperature=="25~30")

```



```
ddply(df_temp_low,.(marry_level,crime_type),summarise,  
      cellmean_temp_15_20 = mean(crime_rate),  
      replication = length(crime_rate),  
      standatd_error = var(crime_rate)) # 温度15~20  
  
ddply(df_temp_mid,.(marry_level,crime_type),summarise, cellmean_temp_20_25 =  
mean(crime_rate),  
      replication = length(crime_rate),  
      standatd_error = var(crime_rate)) # 温度20~25  
ddply(df_temp_high,.(marry_level,crime_type),summarise, cellmean_temp_25_30 =  
mean(crime_rate),  
      replication = length(crime_rate),  
      standatd_error = var(crime_rate)) # 温度25~30
```