

統計期末報告

B06705002 黃啟宏

B06705019 黃勛哲

B06705020 郭又嘉

B06705021 許亦佑

B06705027 黃柏叡

B06705034 吳禹辰

B06705049 王松億

B06705058 劉品桷

報告主題：

8+9？人口？氣溫？想犯罪了

與台灣犯罪率相關數據之調查

研究動機：

我們日常中常常提到的「8+9」與「流氓」、「黑幫」等犯罪份子，造成社會動盪。8+9是台灣宮廟文化中八家將的簡稱，也常常在看電影時知道許多黑道分子和宮廟有部的牽扯。因此，我們十分好奇究竟我們想要調查的8+9是否為台灣犯罪的原因。另外，在小時候曾經聽長輩們閒聊時說過擁擠定律，在人口密度越高的地方，社會也會比較髒，治安會比較亂。而我們曾經在一篇新聞報導中得知一條令人驚訝的報導--溫度越高的地方犯罪率也會越高。我們希望能將溫度也作為犯罪率的一項參考。另外，我們也想知道坐擁較多資源的直轄市在治安上是否有較好的表現。因此，我們希望能夠從宮廟密度、人口密度、溫度和是否為直轄市對刑事犯罪率的分析中，找出各自因素與刑事犯罪率的關係。

## 統計分析：

### ● 資料蒐集

- 使用搜尋引擎，查詢各地區警察分局轄區人數與面積
- 警政署、部分警察分局皆有網路公開資訊，可供民眾查詢刑事案件數量
- 透過全國宗教資訊網查詢各轄區廟宇數量
- 透過氣象局網站查詢各轄區年均溫度

### ● 資料處理

- 將人口數除以面積得到人口密度（人/平方公里）

	全縣警察/總計			縣區/計			縣區/新埔縣區			縣區/新大縣區			縣區/汽車縣區			縣區/機車縣區			
	發生數 (件)	總數 (件)	總數 (人)	發生數 (件)	總數 (件)	總數 (人)	發生數 (件)	總數 (件)	總數 (人)	發生數 (件)	總數 (件)	總數 (人)	發生數 (件)	總數 (件)	總數 (人)	發生數 (件)	總數 (件)	總數 (人)	
1996 年																			
總計	41 071	39 782	40 990	6 062	5 064	3 673	4 931	4 102	3 572	0	10	11	90	88	52	1 032	864	238	29
大 湖 分 局	2 910	2 565	2 507	420	305	257	325	236	228	0	0	0	11	5	5	84	64	24	3
萬 里 分 局	3 106	4 902	4 904	811	680	557	816	538	505	0	0	0	21	18	10	174	146	42	3
中 山 分 局	3 096	4 898	4 935	638	553	395	450	433	338	1	1	1	4	4	2	104	113	14	5
大 安 分 局	4 559	4 321	4 341	659	485	383	551	424	373	2	2	2	1	2	0	105	57	8	3
中 正 分 局	1 907	1 923	1 910	400	316	244	316	251	220	0	0	0	1	1	1	83	64	23	1

- 各地氣溫(攝氏溫度)

	20.4	30.0/12	16.2/3	674.0	10.7/350.0/20	19.4/300.0/18	88	59/5	900.0
日月潭	20.4	30.0/12	16.2/3	674.0	10.7/350.0/20	19.4/300.0/18	88	59/5	900.0
臺中	25.4	33.7/15	18.9/3	524.0	5.1/170.0/27	14.2/10.0/20	82	52/13	999.8
臺南	26.9	33.9/14	20.2/8	168.0	8.3/340.0/20	21.2/320.0/20	75	53/14	1007.1
臺北	25.0	35.1/19	17.9/7	335.8	7.3/80.0/11	12.7/60.0/3	77	33/21	1006.9
臺東	25.6	31.9/20	19.6/4	165.7	5.7/40.0/7	13.3/30.0/1	80	57/22	1009.1
淡水	24.2	33.4/12	17.4/2	488.0	5.3/20.0/20	12.0/10.0/20	82	39/21	1008.2
梧棲	24.7	31.1/19	17.8/2	272.0	11.9/350.0/3	20.0/360.0/3	81	51/22	1007.0
新豐	24.1	31.8/25	17.4/7	500.7	12.2/230.0/19	19.8/240.0/19	81	51/21	1007.5
宜蘭	23.9	32.5/27	18.0/2	360.2	5.9/110.0/11	11.1/360.0/1	84	55/21	1009.9

- 刑事犯罪數除以轄區人口得到刑事犯罪率（犯罪數/十萬人）

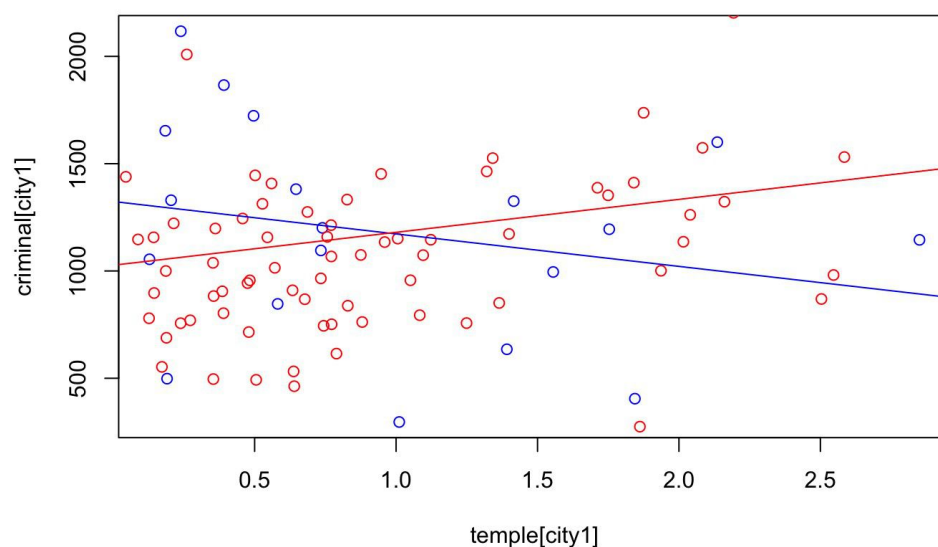
注：「刑事犯罪」指的是，觸犯普通刑法與特別刑法的案件，包括殺人、強盜、擄人勒贖、性侵害等重大暴力案件，以及竊盜、詐欺、毒品、妨害風化、偽造文書等各類案件也屬之。

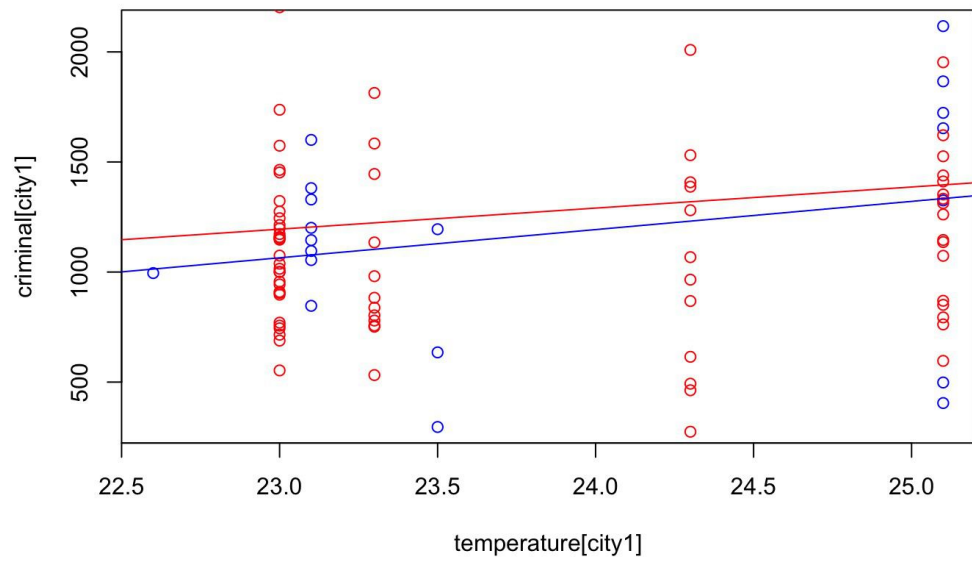
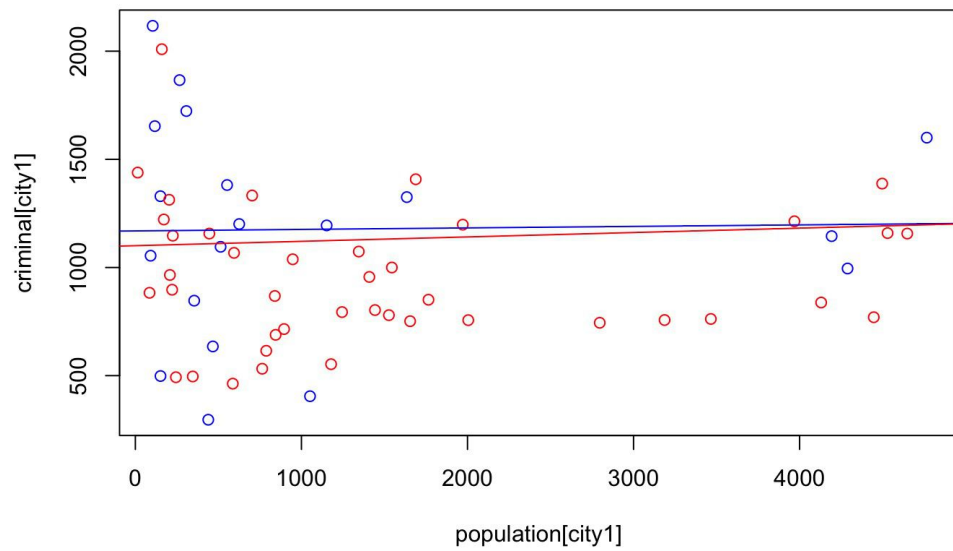
分局	廟宇數	面積	廟宇密度	總人口	犯罪件數	犯罪率	人口密度	是否為直轄市(1:直轄)	溫度
板橋+泰山分局	50	23.14	2.160761	551,557	7295	1322.619421	23835.65255	1	23
中和分局	39	20.14	1.936445	413,563	4540	1001.056671	20534.40914	1	23
永和分局	6	5.71	1.050788	221,731	2121	956.5644858	38832.04904	1	23
三重分局	34	16.32	2.083333	387,651	6101	1573.838324	23753.125	1	23
新莊分局	59	93.05	0.634968	601,224	5470	909.8106529	6461.300376	1	23
土城分局	28	29.56	0.947226	238,179	3459	1452.269092	8057.476319	1	23
新店分局	111	777.13	0.142833	340,211	4005	1156.808998	445.499466	1	23
蘆洲分局	63	81.79	0.770265	324,615	3939	1213.437457	3968.883727	1	23
三鶯分局	75	212.57	0.352825	201,465	2091	1037.897402	947.7583855	1	23
樹林分局	29	33.13	0.87534	184,170	1979	1074.550687	5559.009951	1	23
汐止分局	53	71.24	0.743964	199,319	1483	744.406909	2796.448624	1	23
淡水分局	66	136.65	0.482986	192,528	1841	956.2245492	1408.913282	1	22.2
瑞芳分局	83	388.29	0.213756	66,540	813	1221.821461	171.366762	1	22.2
金山分局	58	163.85	0.353982	56,638	281	496.133338	345.66982	1	22.2
大岡分局	29	5.68	5.105634	129,943	2910	2239.443448	22877.28873	1	23
萬華分局	23	8.85	2.58887	193,415	5106	2639.619344	21854.80226	1	23
中山分局	30	13.68	2.192982	231,332	5096	2202.894541	16910.23392	1	23
大安分局	15	11.36	1.320423	311,404	4559	1464.014592	27412.32394	1	23
中正一分局+中正	13	7.6	1.710526	160,484	3596	2240.721817	21116.31579	1	23
松山分局	13	9.29	1.399354	208,289	2442	1172.409489	22420.77503	1	23
信義分局	21	11.2	1.875	227,666	3955	1737.19396	20327.32143	1	23
士林分局	34	62.37	0.545134	289,916	3354	1156.898822	4648.324515	1	23
北投分局	43	56.82	0.756776	257,384	2983	1158.9687	4529.813446	1	23
內山一分局+內山	18	31.51	0.571247	275,253	2793	1014.776571	8734.782609	1	23

## ● 回歸分析

- 畫散佈圖

(藍色為直轄市，紅色微非直轄市)





○ 提出統計模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon$$

$\beta_0$  : Intercept

$\beta_1$  : 廟宇密度

$\beta_2$  : 人口密度

$\beta_3$  : 直轄市(1為是, 0為否)

$\beta_4$  : 當地溫度(度c)

## ○ 套入模型

```
Call:
lm(formula = criminal ~ data)

Residuals:
    Min       1Q   Median       3Q      Max
-1041.84  -333.20   -64.66   167.06  3079.88

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.697e+01  1.668e+03  -0.010  0.991900
datatemple     1.288e+02  3.615e+01   3.562  0.000573 ***
datapopulation  6.433e-03  8.669e-03   0.742  0.459876
datacity      -2.230e+01  1.637e+02  -0.136  0.891936
datatemperature 4.446e+01  6.993e+01   0.636  0.526401
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 609.6 on 97 degrees of freedom
Multiple R-squared:  0.1907,    Adjusted R-squared:  0.1573
F-statistic: 5.715 on 4 and 97 DF,  p-value: 0.0003579
```

## ○ 評估模型

$R^2$ 值為0.1907，顯示19.07%的資料可以被此模型解釋，而Residual

standard error為609.6，小於平均  但仍偏大。

## 相關係數

	criminal	temple	population	city	temperature
criminal	1.00000000	0.4288146	0.24669614	0.04972394	0.13536726
temple	0.42881455	1.00000000	0.45761246	0.11032373	0.20918778
population	0.24669614	0.4576125	1.00000000	0.31522093	-0.07196432
city	0.04972394	0.1103237	0.31522093	1.00000000	-0.07715870
temperature	0.13536726	0.2091878	-0.07196432	-0.07715870	1.00000000

## 回歸結果

```
Call:
lm(formula = criminal ~ data)

Residuals:
    Min       1Q   Median       3Q      Max
-1041.84  -333.20   -64.66   167.06  3079.88

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.697e+01  1.668e+03  -0.010  0.991900
datatemple    1.288e+02  3.615e+01   3.562  0.000573 ***
datapopulation 6.433e-03  8.669e-03   0.742  0.459876
datacity     -2.230e+01  1.637e+02  -0.136  0.891936
datatemperature 4.446e+01  6.993e+01   0.636  0.526401
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 609.6 on 97 degrees of freedom
Multiple R-squared:  0.1907,    Adjusted R-squared:  0.1573
F-statistic: 5.715 on 4 and 97 DF,  p-value: 0.0003579
```

$$y = 16.97 + 128.8x_1 + 0.006433x_2 - 22.3x_3 + 44.46x_4$$

### Analysis of Variance Table

```
Response: criminal
      Df  Sum Sq Mean Sq F value    Pr(>F)
data     4  8495175  2123794   5.7145 0.0003579 ***
Residuals 97 36049993   371649
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA結果顯示此模型valid

然而，雖然直轄市(city)與犯罪(criminal)的相關係數為正，但是回歸結果的city係數為負，可知模型存在多元共線性。

### ○ 殘差分析

#### Run Test

$H_0$ = the randomness doesn't exists

$H_1$ = the randomness exists

```
cat(Run_Test(SE))
```

```
1
```

Don't reject the null hypothesis since  $1 > 0.05$ .

我們有95%的信心資料為隨機。

## Chi-square test for normality

$H_0$  = the standardized errors are normally distributed

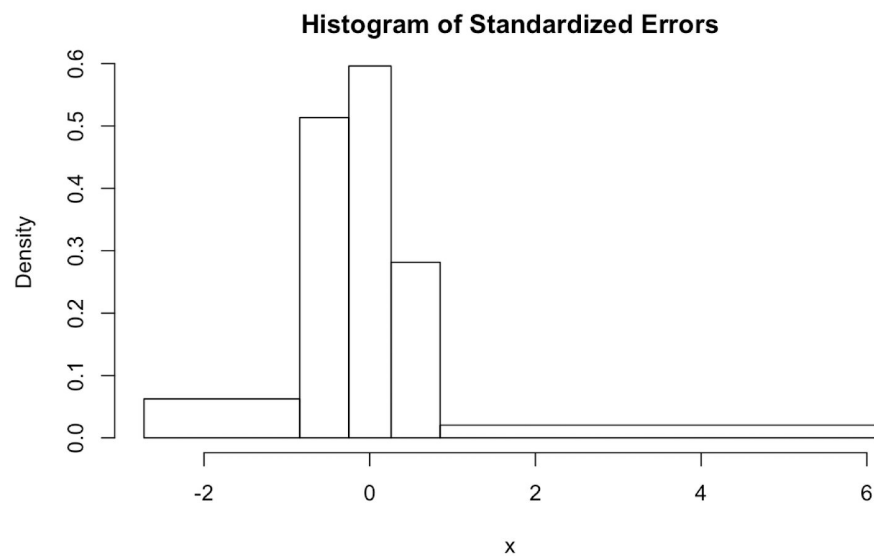
$H_1$  = the standardized errors are not normally distributed

#### Chi Square Test ####

Hide

Chi\_Square\_Normal\_Test(SE,5)

The Expected Counts: 20.4 20.4 20.4 20.4 20.4  
X-squared  
6.213044e-05



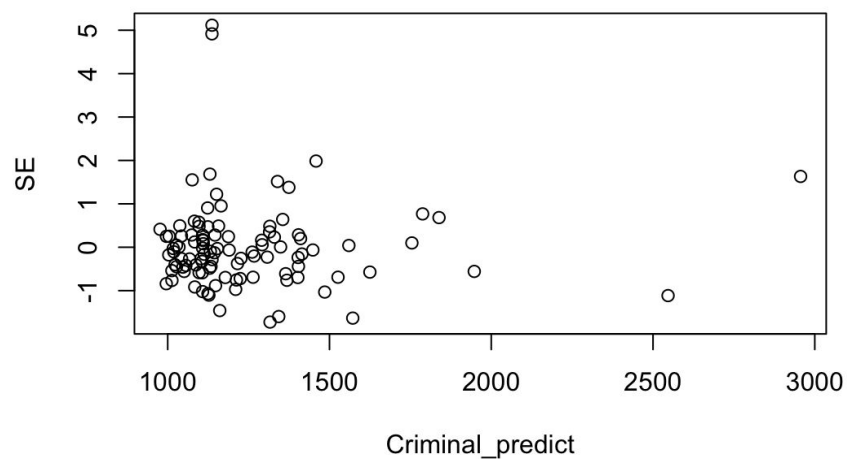
Don't reject the null hypothesis since  $6.213044e-05 < 0.05$ .

我們無法做出常態分佈的結論。

## Test for Homoscedasticity

$H_0$  = Homoscedasticity

$H_1$  = Heteroscedasticity



沒有reject  $H_0$

## Outliers

```
Outliers <- abs(SE) > 2  
cat("Outliers: ", which(Outliers), "\n")
```

```
Outliers: 60 61
```

Outlier為第60、61筆資料

綜合以上，此模型存在多元共線性，因此，我們必須套入新模型(刪除直轄市city)。

除此之外，我們也刪除了60, 61筆這兩筆outlier(分別為佳里分局，學甲分局)。

```
data2 = data[-60,]  
data2 = data2[-60,]  
data2 = data2[, -3]  
criminal2 = criminal[-60]  
criminal2 = criminal2[-60]
```

## Influential data

新北市永和分局、台南市第二分局、高雄三民二分局、台南鹽埕分局，在還沒去除outlier時是influential data。

新北市永和分局、台南市第二分局、台南市苓雅分局、高雄市前鎮分局，在去除outlier後是influential data。

### ○ 套入新模型

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

$\beta_0$ ：當其他獨立變數為0時y的值，但是此模型不存在所有獨立變數為0的情況，所以不能interpret intercept.

$\beta_1$ ：廟宇密度

$\beta_2$ ：人口密度

$\beta_3$ ：當地溫度(度c)



## ○ 評估模型

```
Call:
lm(formula = criminal2 ~ data2)

Residuals:
    Min       1Q   Median       3Q      Max
-958.78 -282.68  -18.16   210.01 1159.90

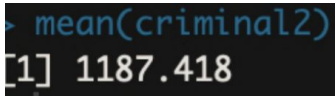
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  8.908e+02  1.153e+03   0.773   0.4415
data2temple   1.360e+02  2.508e+01   5.422 4.39e-07 ***
data2population 1.041e-02  5.779e-03   1.802  0.0747 .
data2temperature 1.311e+00  4.865e+01   0.027  0.9786
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 423 on 96 degrees of freedom
Multiple R-squared:  0.3696,    Adjusted R-squared:  0.3499
F-statistic: 18.76 on 3 and 96 DF,  p-value: 1.171e-09
```

$$y = 890.8 + 136x_1 + 0.01041x_2 + 1.311x_3$$

$|r| = 0.60795$ , model is moderately strong.

R-squared = 0.3696, 顯示36.96%的資料可以被此模型解釋。

標準差為423, 小於犯罪率之平均  , 與先前的模型相比也有降低。

```
print(anova(model2))
```

Analysis of Variance Table

Response: criminal2

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
data2	3	10072555	3357518	18.763	1.171e-09 ***
Residuals	96	17178466	178942		

---

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

p-value < 0.05, This model is valid.

	criminal2	temple	population	temperature
criminal2	1.0000000	0.5896293	0.40041274	0.10200377
temple	0.5896293	1.0000000	0.45536823	0.21490932
population	0.4004127	0.4553682	1.0000000	-0.06322616
temperature	0.1020038	0.2149093	-0.06322616	1.0000000

多元共線性的問題已被解決。

○ 殘差分析

**Chi-square test for normality**

$H_0$  = the standardized errors are normally distributed

$H_1$  = the standardized errors are not normally distributed

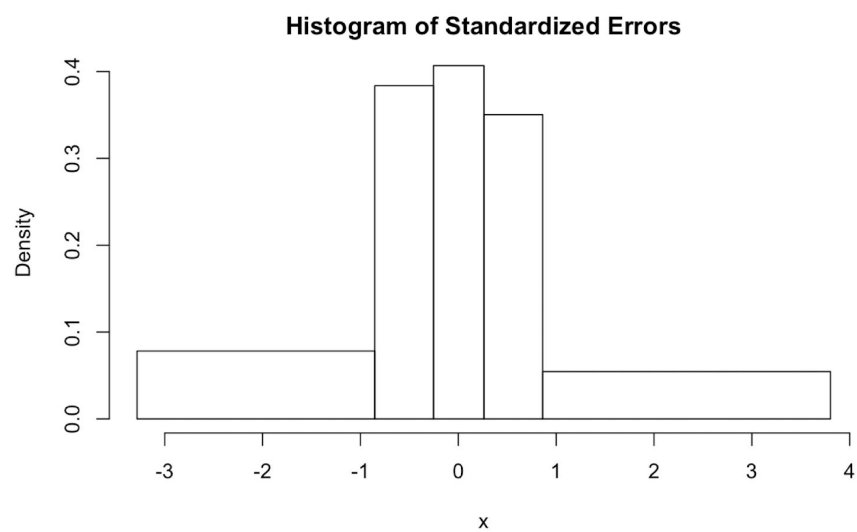
```
cat("\n", "#### Chi Square Test ####", "\n")
```

```
#### Chi Square Test ####
```

Hide

```
Chi_Square_Normal_Test(SE2,5)
```

```
The Expected Counts: 20 20 20 20 20
X-squared
0.4965853
```

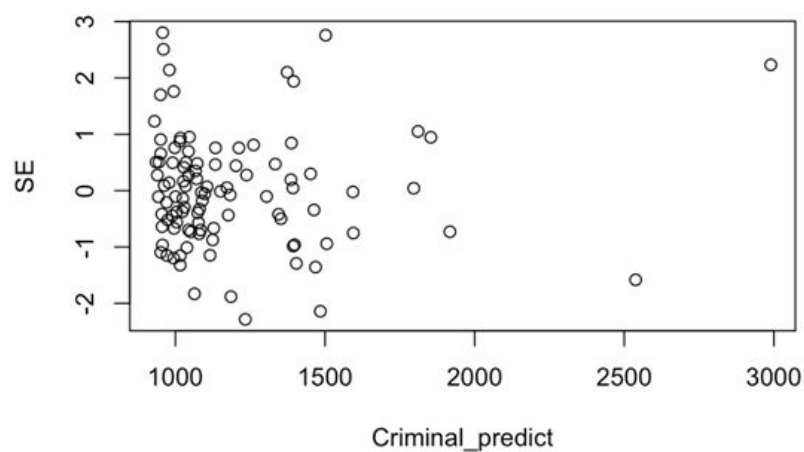


p-value = 0.4965853 > 0.05, 沒有reject  $H_0$ , 數據是常態分佈。

**Test for Homoscedasticity**

$H_0$  = Homoscedasticity

$H_1$  = Heteroscedasticity



沒有reject  $H_0$ 。

## Run Test

$H_0$  = the randomness doesn't exist

$H_1$  = the randomness exists

```
cat("##### Run Test #####", "\n")
```

```
##### Run Test #####
```

Hide

```
cat(Run_Test(SE))
```

```
0.6876578
```

p-value = 0.68766 > 0.05, 沒有reject  $H_0$ 。

### ○ 解釋模型

- $y = 890.8 + 136x_1 + 0.01041x_2 + 1.311x_3$
- $y$  = 犯罪率（刑事案件數/十萬人）， $x_1$  = 該區域宮廟密度（宮廟數/平方公里）， $x_2$  = 該區域人口密度（人口/平方公里）， $x_3$  = 該區域平均氣溫（攝氏溫度）
- 其中，只有宮廟這項獨立變數的p-value( $4.39e-07$ ) < 0.05，因此在蒐集的所有因素中，我們有95%的信心認為宮廟密度會影響犯罪率。
- 我們發現，其他情況相同下，宮廟密度每上升一個單位，每十萬人的刑事案件數就會提高136件。
- 而我們的解釋力其實並不高，這可能的原因是我們並沒有找到真正和犯罪率相關的變數。雖然我們有想到像是警察佈署的密度、教育程度、平均收入等的變數可能會對犯罪率有顯著的影響，但是這些資料沒有辦法針對每個不同的轄區分別記錄，因此無法和犯罪率做比較，這是我們在這次報告感到最遺憾的地方。