

影響台灣地震之原因探討



研究動機

臺灣，坐落於板塊交界的特殊地理位置，地震是我們日常生活中的一部分。學習如何與地震和諧相處已然成為必修課程。我們收集的資料集來自臺灣地震與地球物理資料管理系統，透過這些寶貴數據，我們得以深入探討地震規模會被哪些因素影響，為我們打開地震奧秘的大門。這份資料不僅呈現了地震的複雜性，還讓我們透過科學的方法更全面地理解地震的本質。

研究目的

透過這份報告，我們期望能瞭解地震規模的形成原因及其組合效應，同時洞悉重大地震的主要成因。這使我們得以追蹤重大地震的發生地點、活動型態，以及其活動的規律性。這樣的資訊有助於我們有效的預測潛在地震風險，進而將地震可能帶來的災害損失降至最低。透過深入了解地震的多重因素。

資料介紹

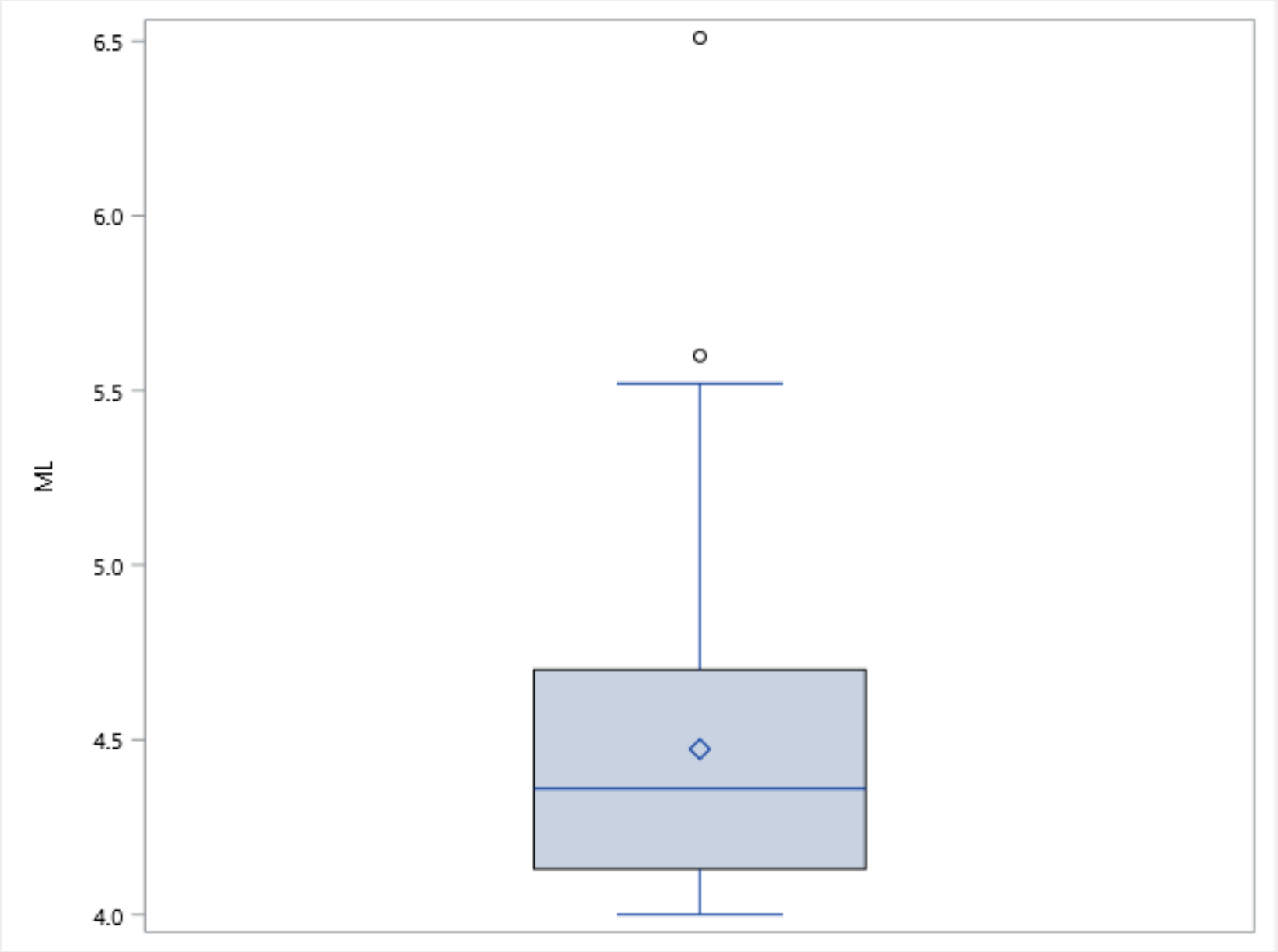
反應變數	y	地震的芮氏規模(ML)
預測變數	x_1 (屬量)	地震的緯度(lat)
	x_2 (屬量)	地震的經度(lon)
	x_3 (屬量)	震源深度(depth)
	x_4 (屬量)	使用觀測站數量(nstn)
	x_5 (屬量)	最近站震央距(Dmin)
	x_6 (屬量)	時間殘值之方均根誤差值(trms)
	x_7 (屬量)	震央之(水平)標準差(公里)(ERH)
	x_8 (屬質)	品質(Quality)

屬質變數介紹

Q	測站數量	最大間隙角度	最小震央距
A	≥ 6	≤ 90	$\leq \text{Depth or } 5 \text{ Km}$
B	≥ 6	≤ 135	$\leq 2 \text{ Depth or } 10 \text{ Km}$
C	≥ 6	≤ 180	$\leq 50 \text{ Km}$
D	others	-	-

深度控制,F: 未限制，即採逆推收斂至最小誤差；X:限制深度，即依據經驗給定震源深度後，該參數不為變數進行逆推收斂至最小誤差。

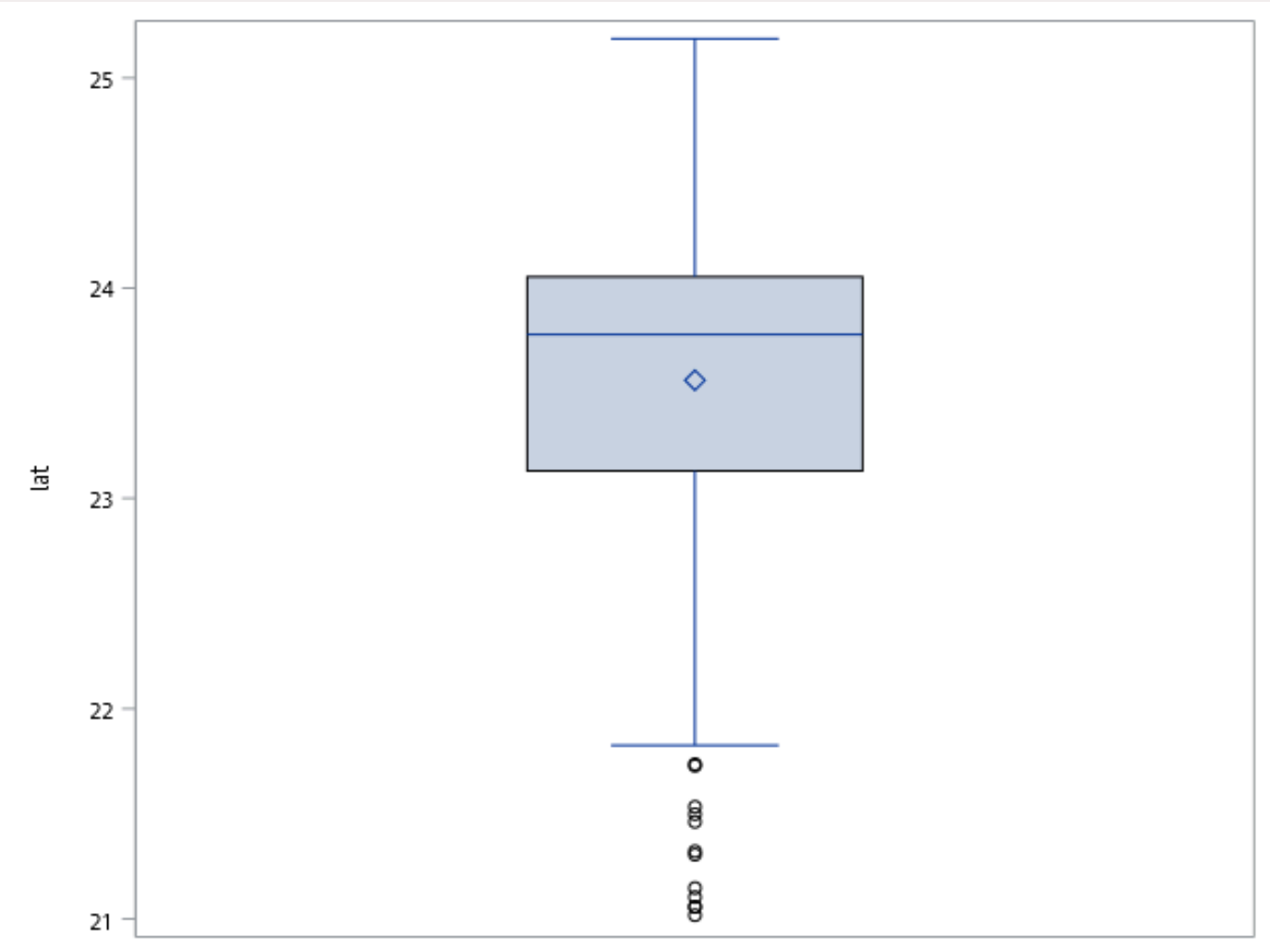
敘述統計量



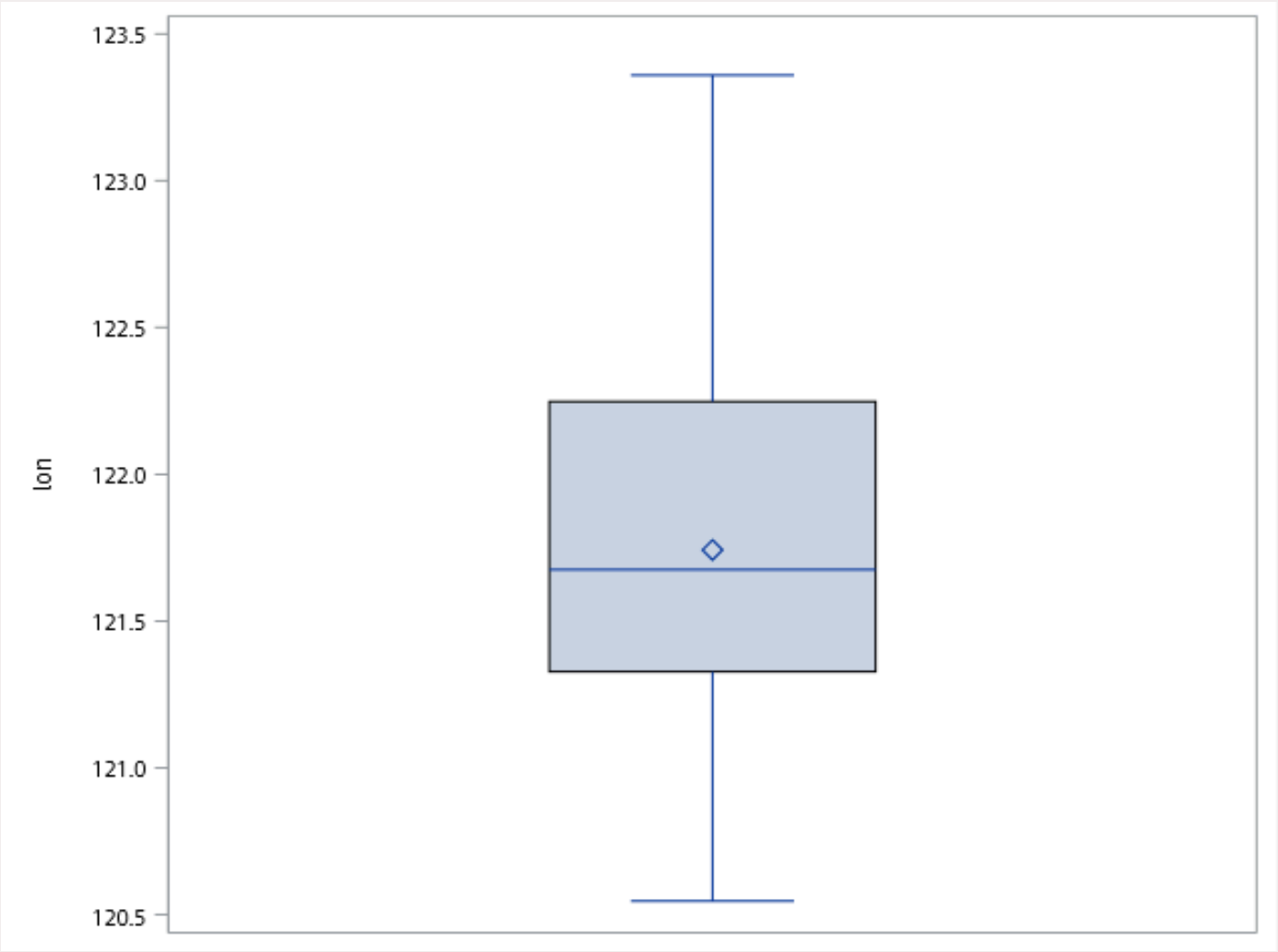
地震芮氏規模

敘述統計量

緯度



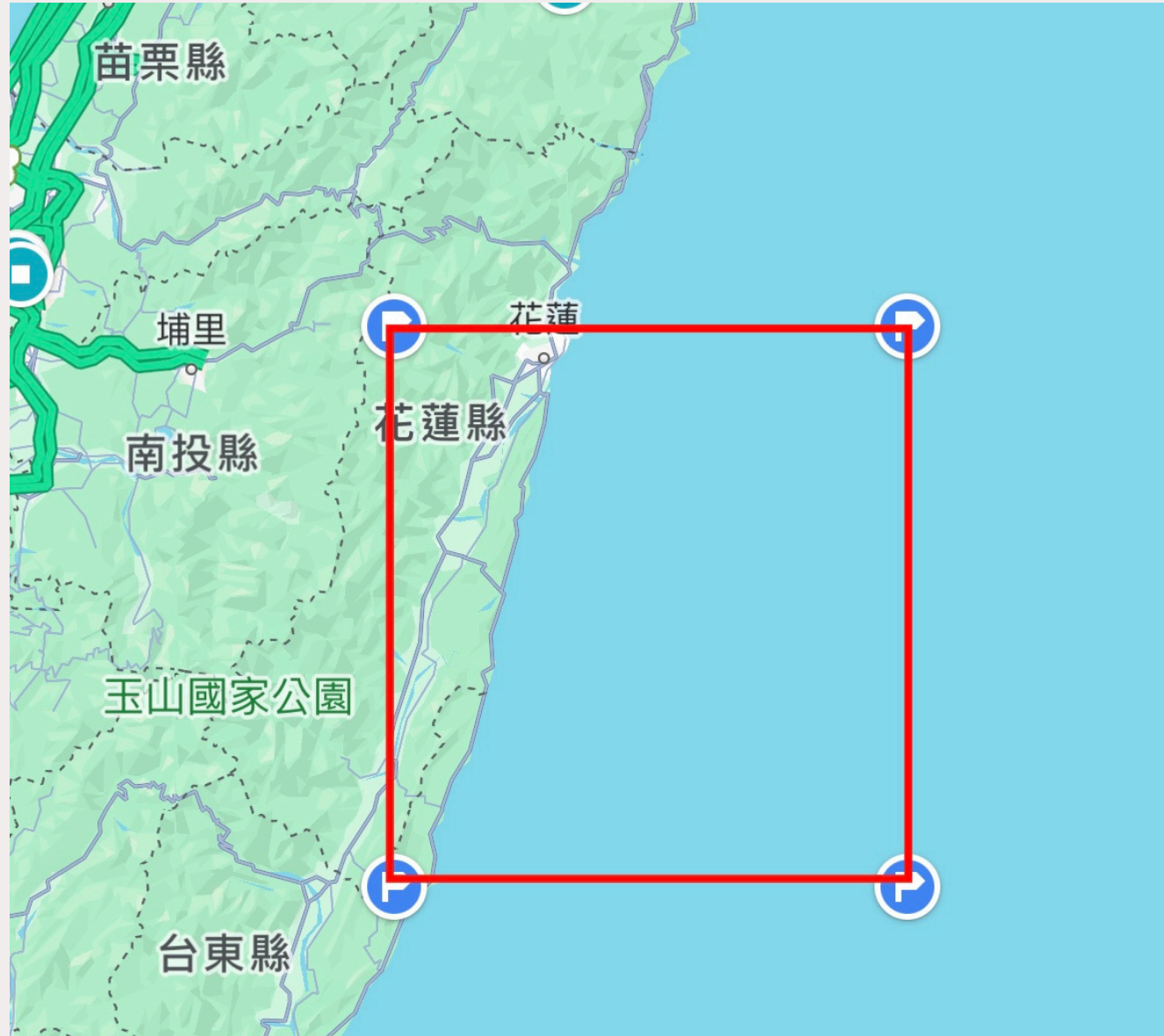
經度



平均值	中位數	眾數	標準差	最小值	最大值	四分位距
23.56	23.78	23.26	0.91	21.02	25.19	0.92

平均值	中位數	眾數	標準差	最小值	最大值	四分位距
121.74	121.68	121.24	0.58	120.55	123.36	0.92

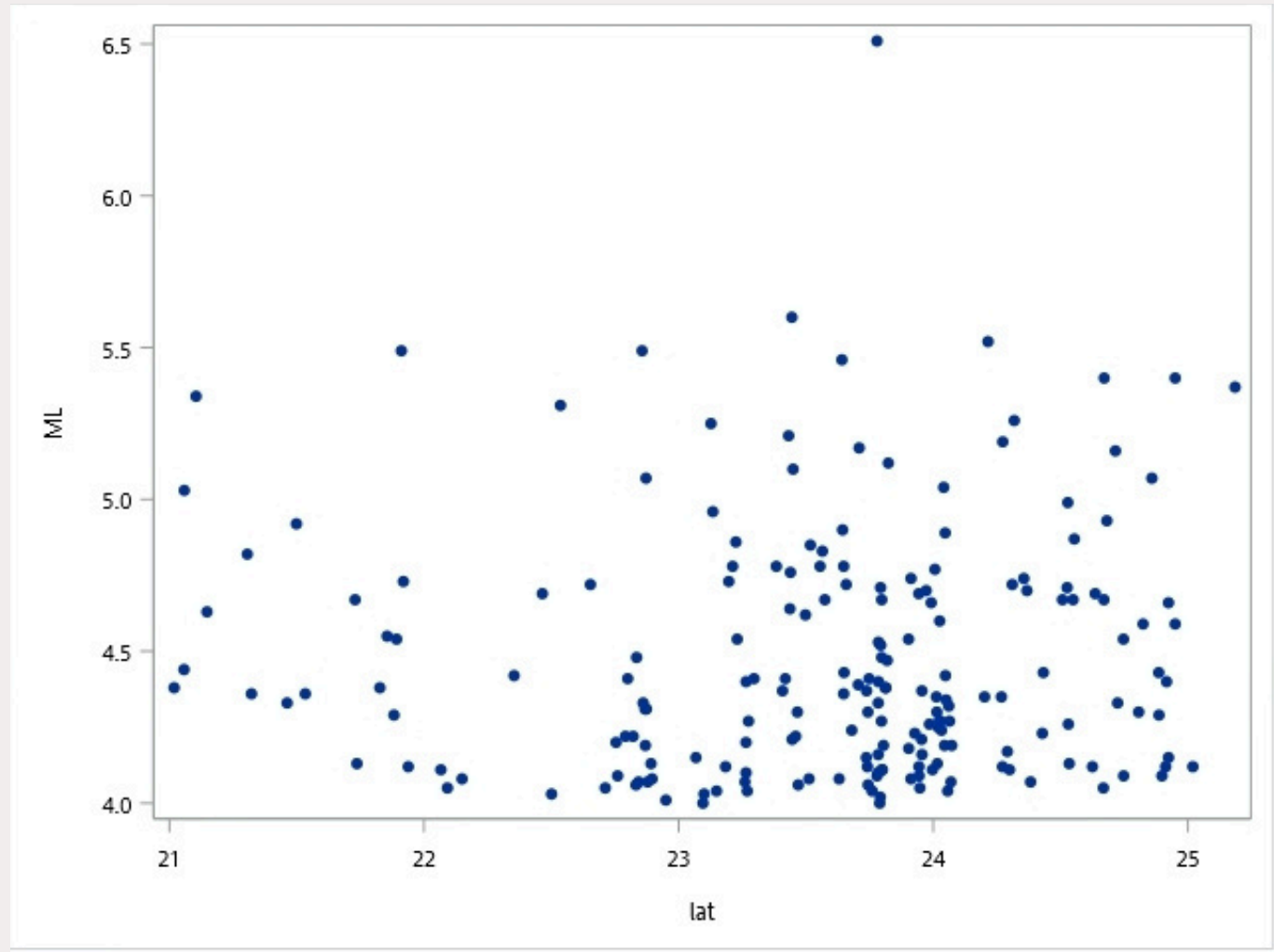
敘述統計量



根據上述資料，最容易發生地震的地區
位於臺灣東半部及外海地區

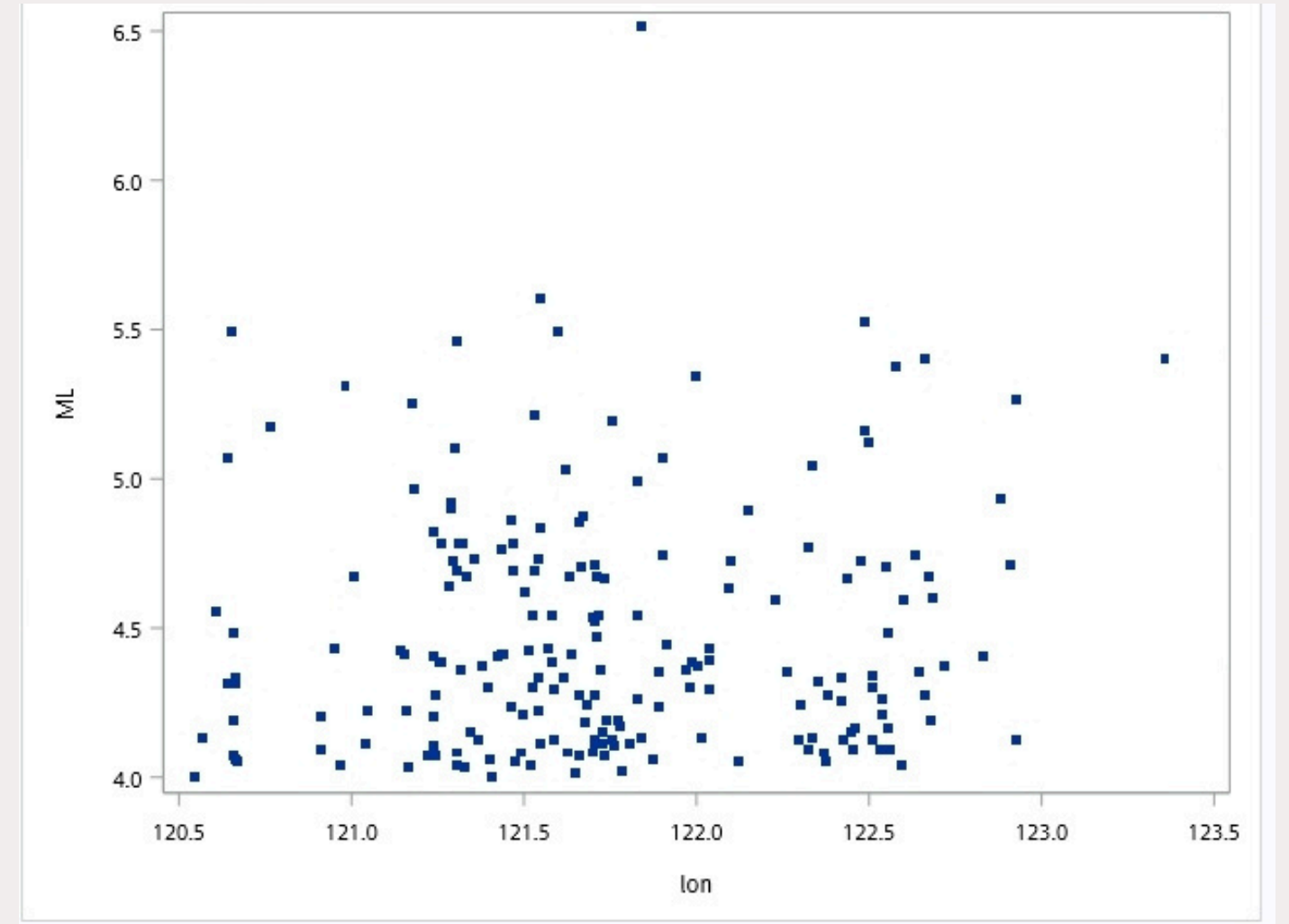
敘述統計量

緯度



$r=0.00816$

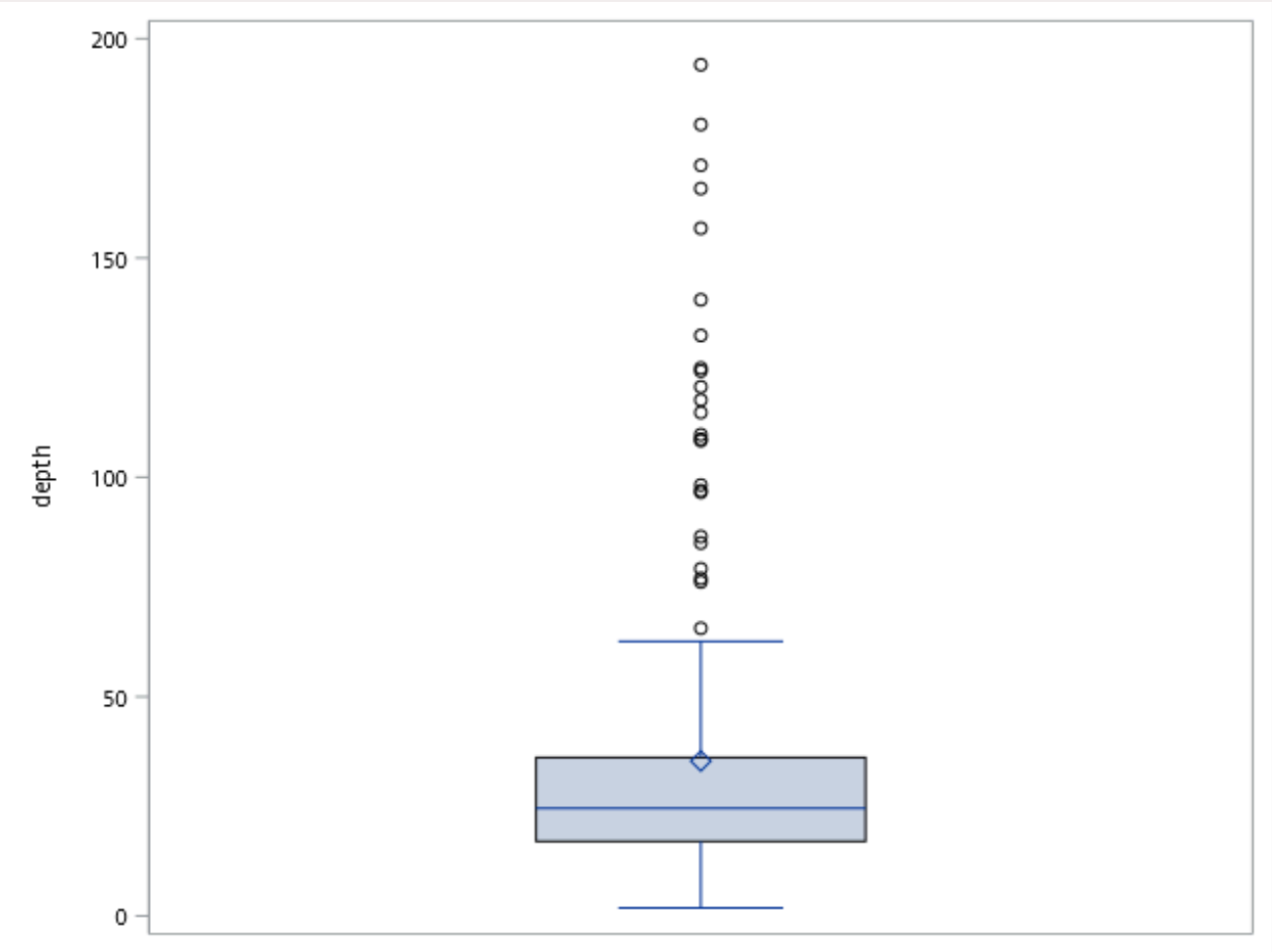
經度



$r=0.07241$

敘述統計量

震源深度



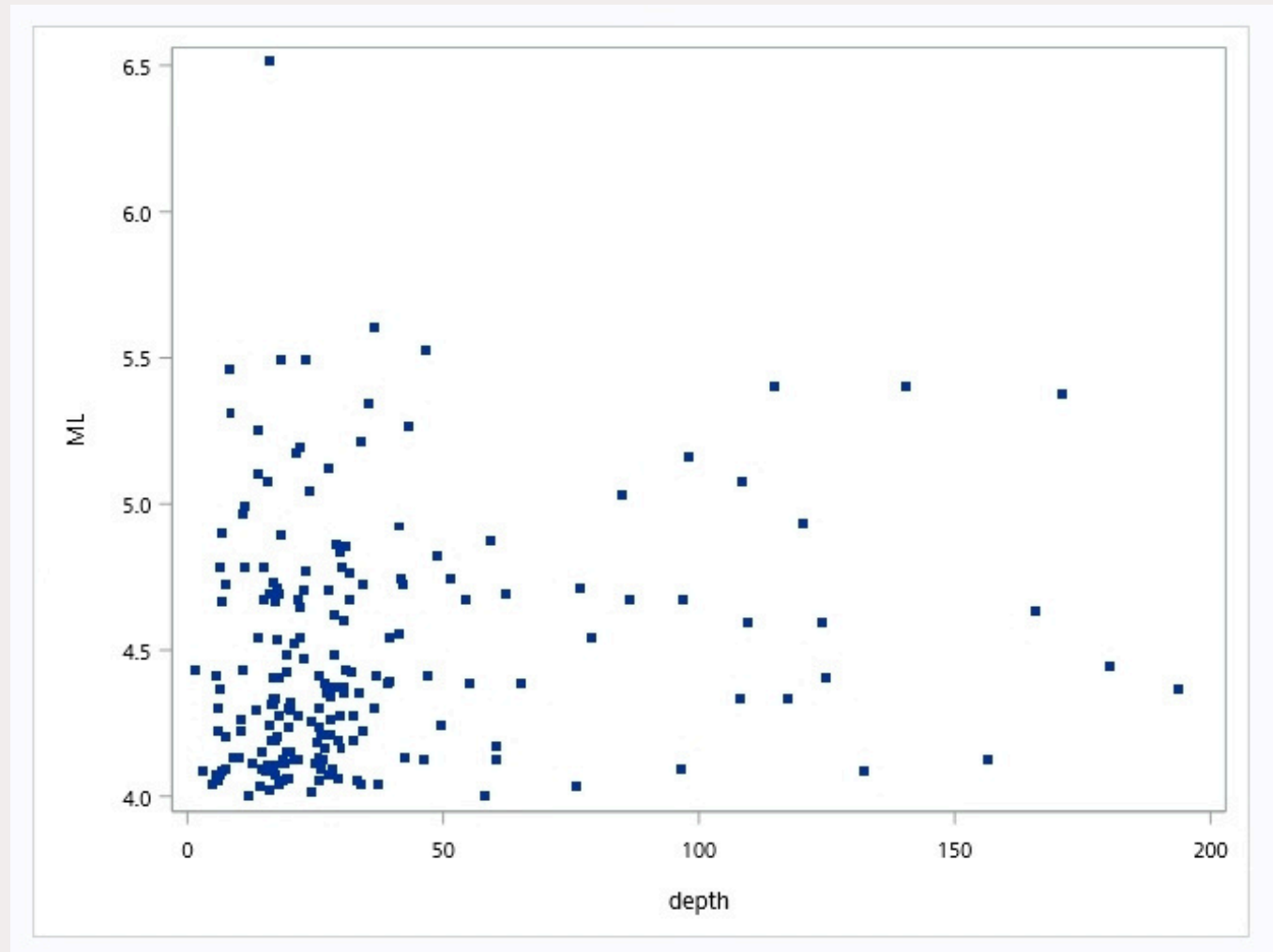
使用測站數

平均值	中位數	眾數	標準差	最小值	最大值	四分位距
96.51	99.00	99.00	12.13	18.00	99.00	0.00

平均值	中位數	眾數	標準差	最小值	最大值	四分位距
35.33	24.55	10.97	34.90	1.82	194.07	19.15

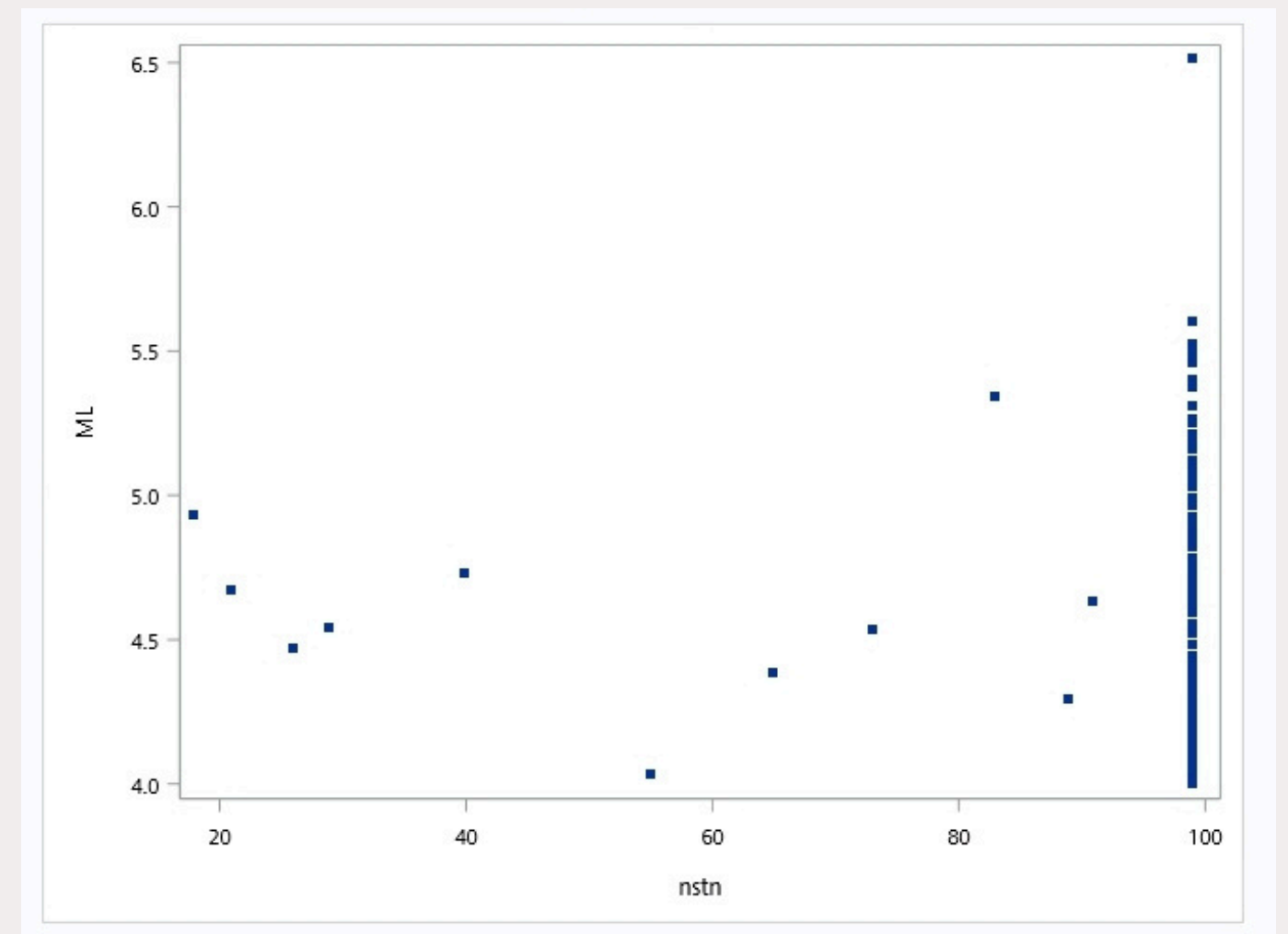
敘述統計量

震源深度



$r=0.16478$

使用測站數



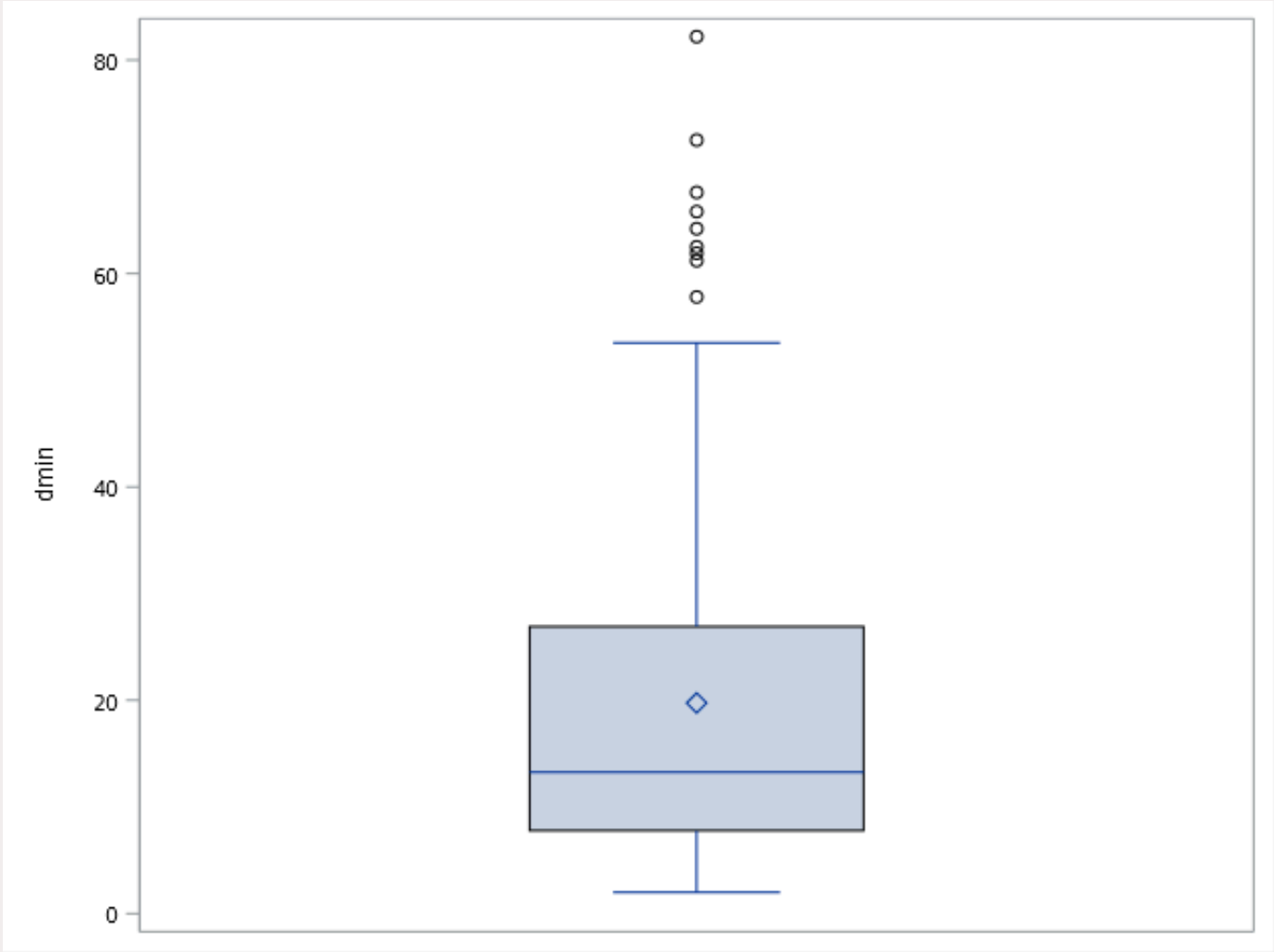
$r=-0.06485$

敘述統計量

大多數的地震都為極淺層地震及較為淺層地震

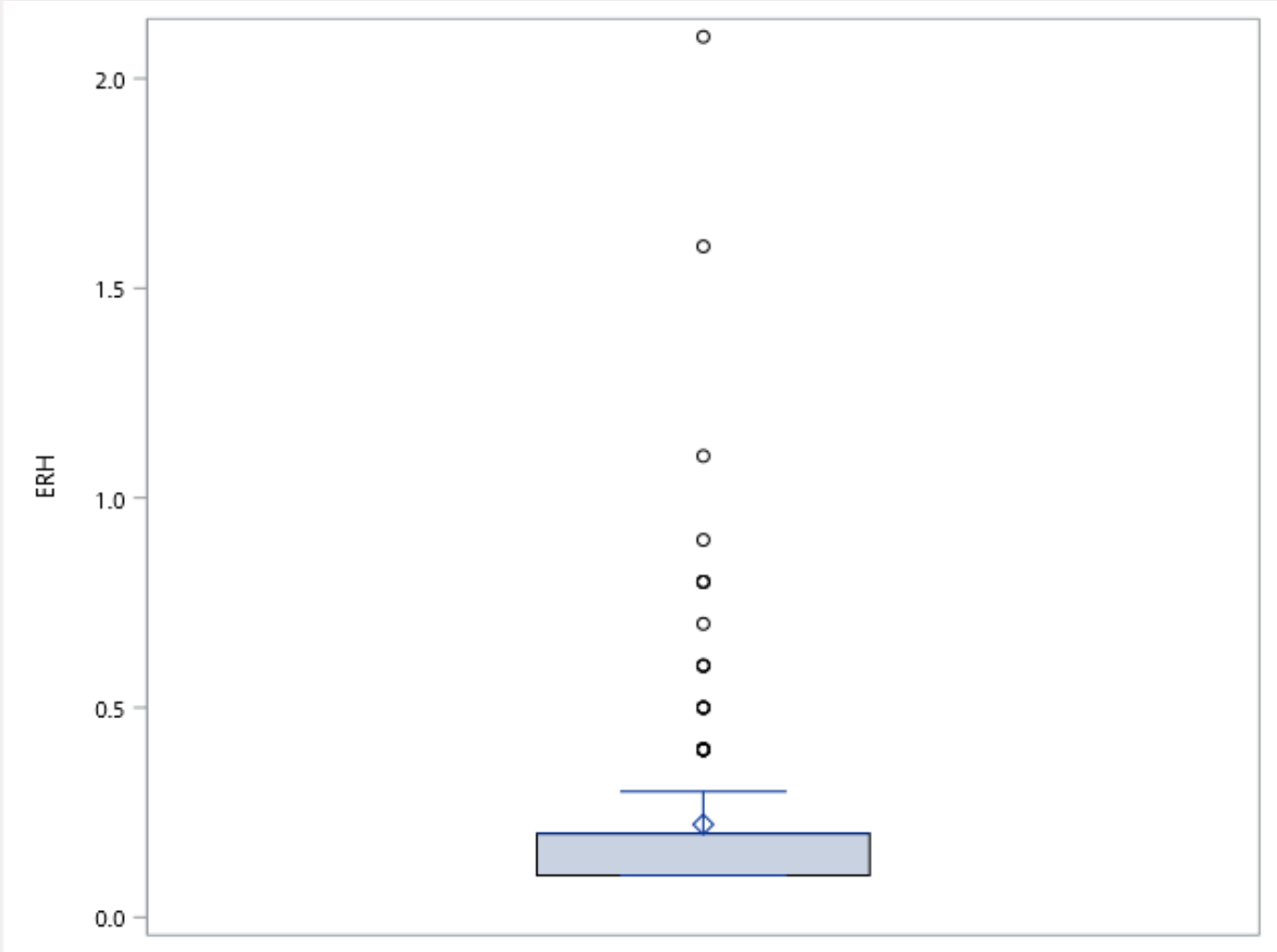
敘述統計量

最近站震央距



平均值	中位數	眾數	標準差	最小值	最大值	四分位距
19.75	13.25	3.40	16.88	2.00	82.20	19.10

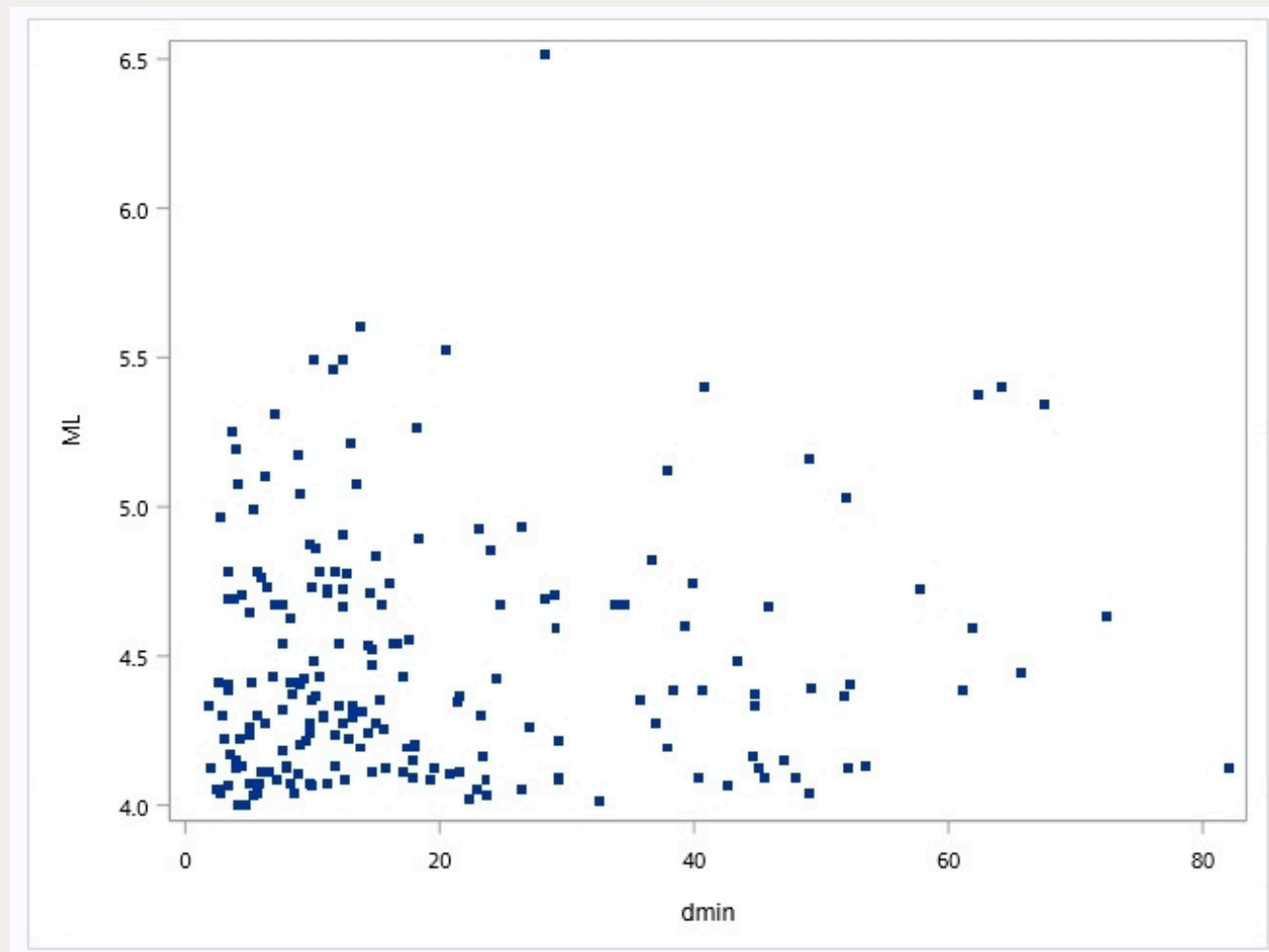
震央之(水平)標準差(公里)



平均值	中位數	眾數	標準差	最小值	最大值	四分位距
0.22	0.20	0.10	0.23	0.10	2.10	0.10

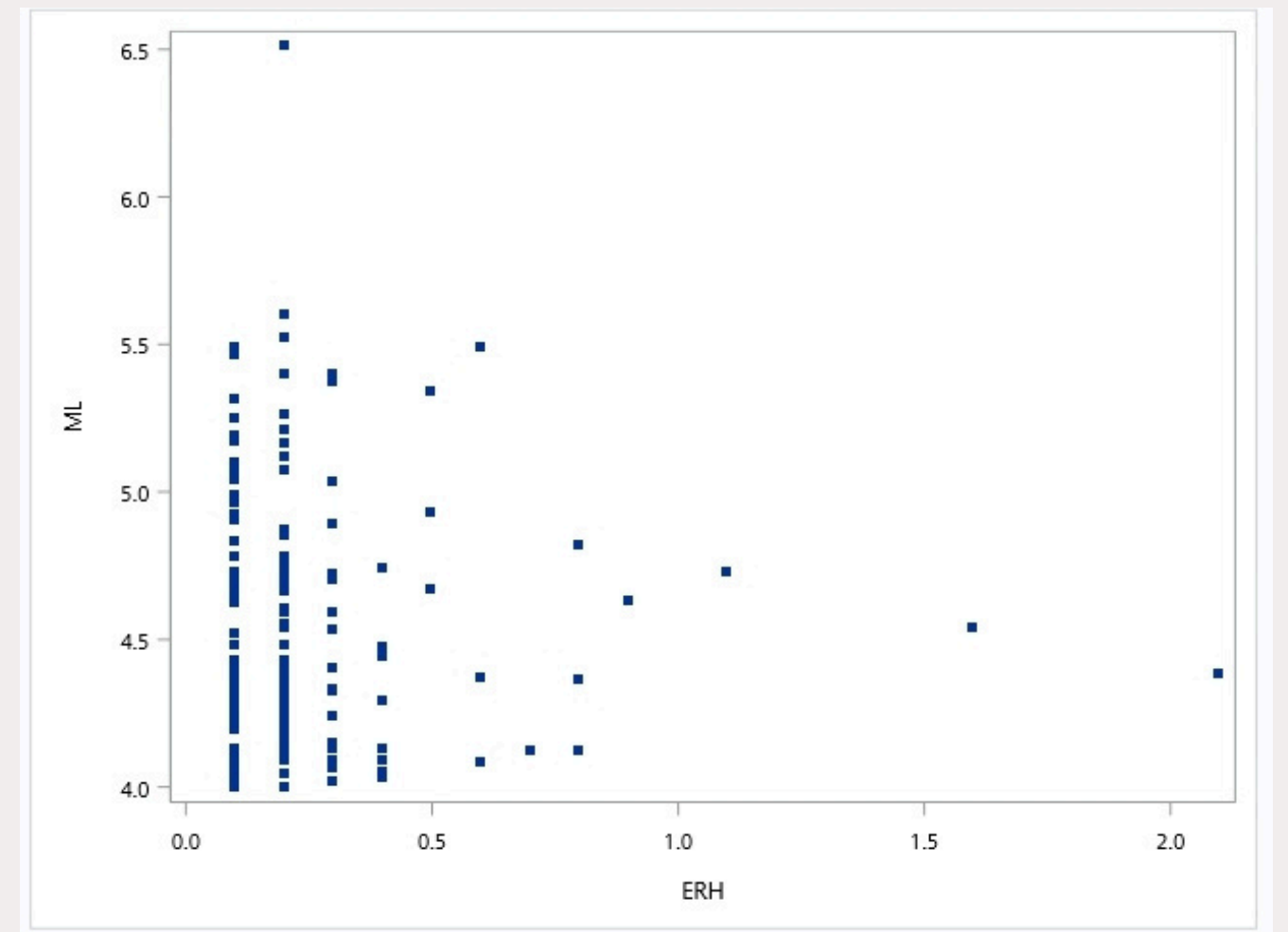
敘述統計量

最近站震央距



$r=0.08088$

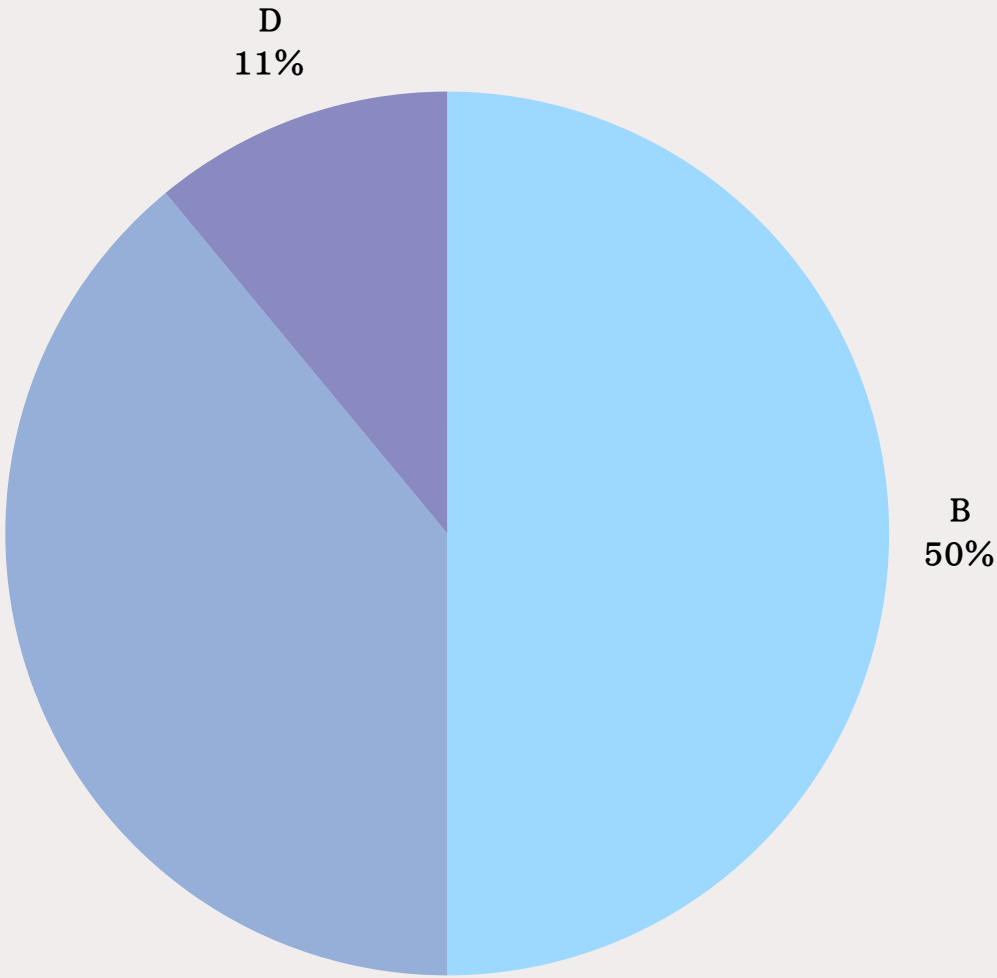
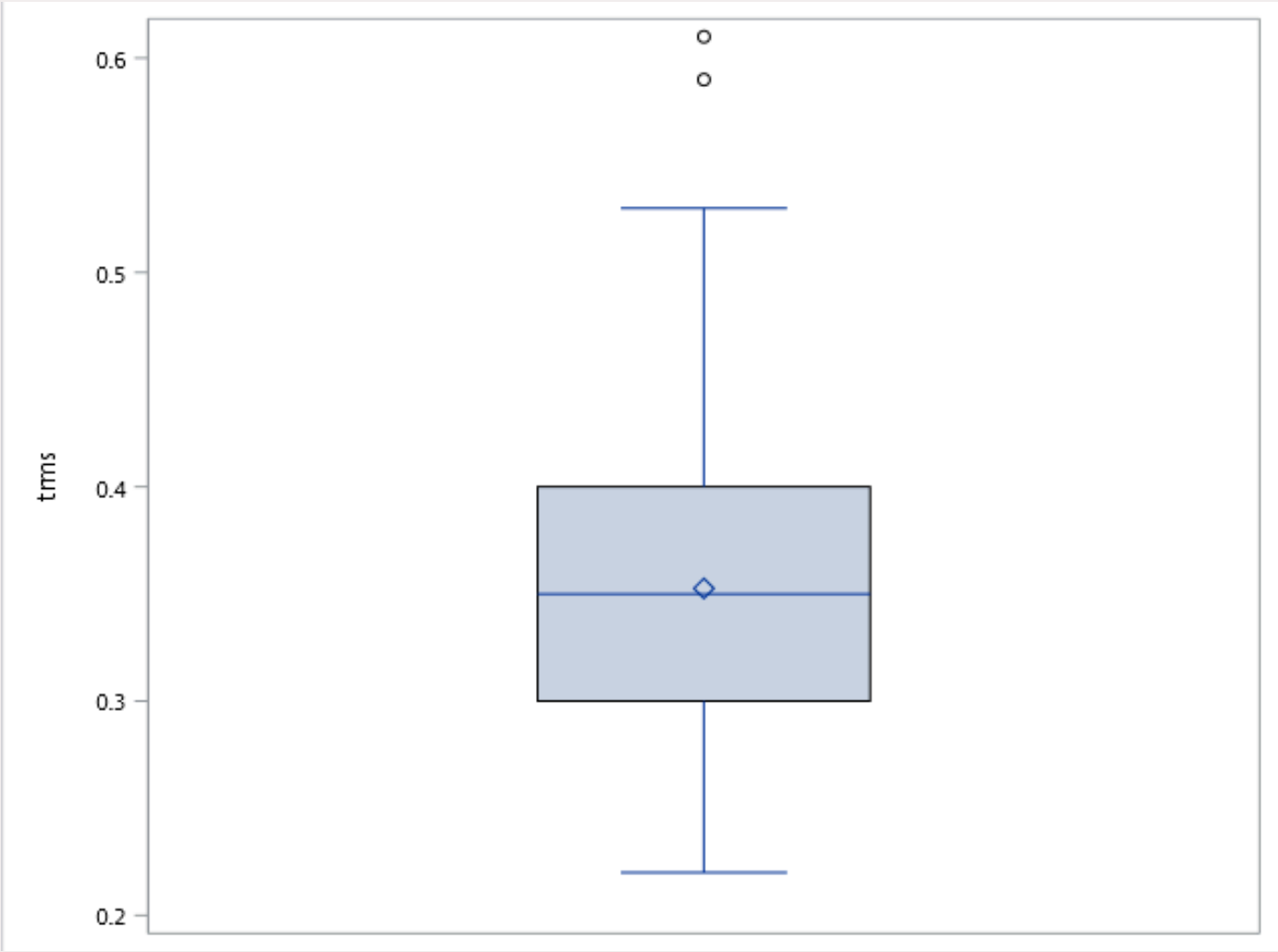
震央之(水平)標準差(公里)



$r=0.04735$

敘述統計量

時間殘值之方均根誤差值

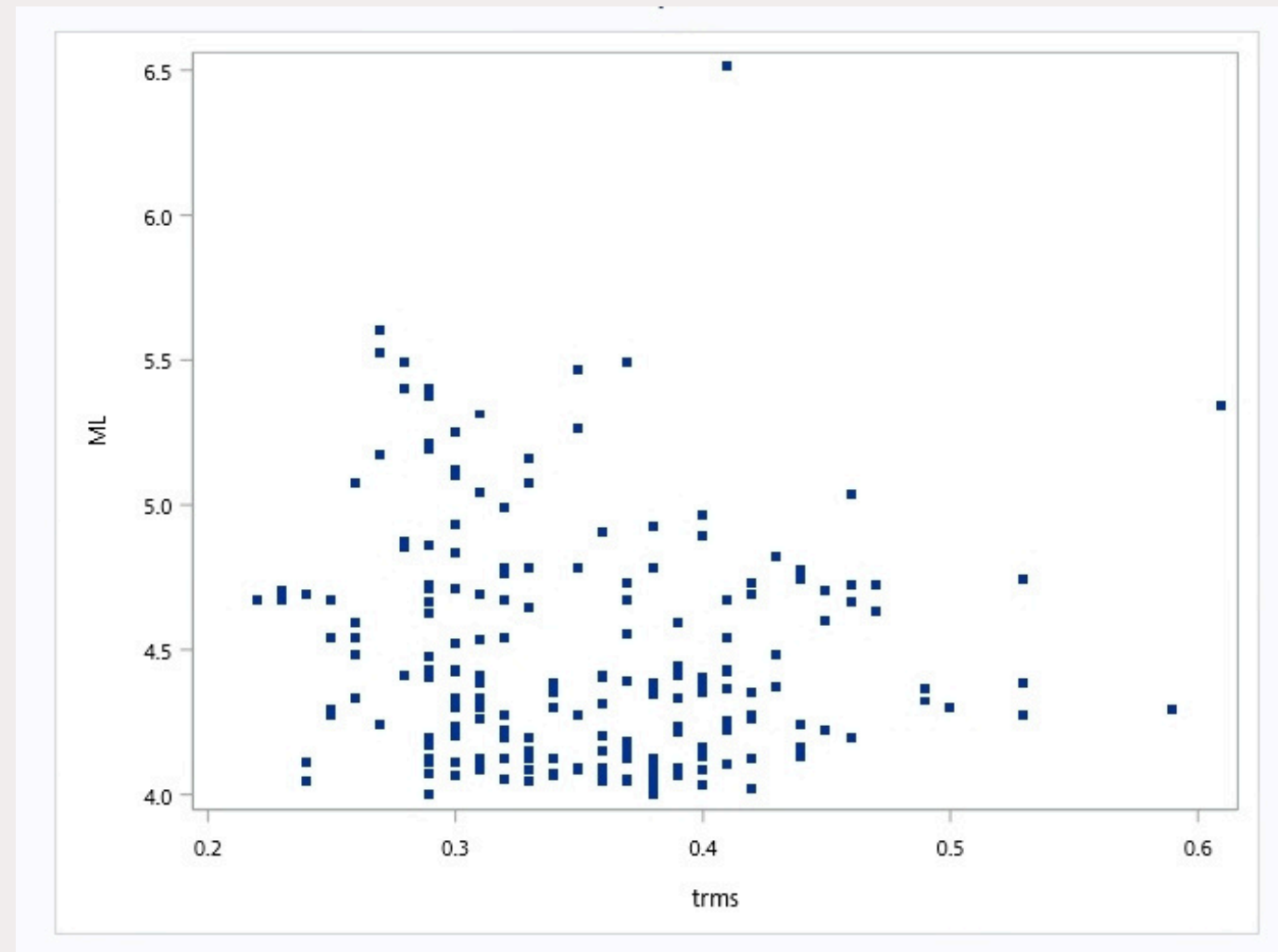


品質

平均值	中位數	眾數	標準差	最小值	最大值	四分位距
0.35	0.35	0.29	0.07	0.22	0.61	0.10

敘述統計量

時間殘值之方均根誤差值



$$r=-0.10450$$

共線性檢測

變數	DF	參數估計值	標準誤差	t 值	Pr > t	變異數膨脹
Intercept	1	-0.76733	9.83975	-0.08	0.9379	0
lat	1	-0.03527	0.05108	-0.69	0.4907	2.60952
lon	1	0.05325	0.08812	0.60	0.5464	3.17139
depth	1	0.00155	0.00105	1.48	0.1410	1.60032
nstn	1	-0.00219	0.00307	-0.71	0.4775	1.66149
dmin	1	0.00074931	0.00280	0.27	0.7895	2.67569
trms	1	-0.80250	0.53234	-1.51	0.1333	1.58366
ERH	1	-0.07920	0.19234	-0.41	0.6810	2.30381
IND	1	0.04800	0.09915	0.48	0.6289	1.15721
IND1	1	0.06569	0.06260	1.05	0.2953	1.12097

模型配適

參數估計值						
變數	標籤	DF	參數估計值	標準誤差	t 值	Pr > t
Intercept	Intercept	1	-0.76733	9.83975	-0.08	0.9379
lat	lat	1	-0.03527	0.05108	-0.69	0.4907
lon	lon	1	0.05325	0.08812	0.60	0.5464
depth	depth	1	0.00155	0.00105	1.48	0.1410
nstn	nstn	1	-0.00219	0.00307	-0.71	0.4775
dmin	dmin	1	0.00074931	0.00280	0.27	0.7895
trms	trms	1	-0.80250	0.53234	-1.51	0.1333
ERH	ERH	1	-0.07920	0.19234	-0.41	0.6810
IND	IND	1	0.04800	0.09915	0.48	0.6289
IND1	IND1	1	0.06569	0.06260	1.05	0.2953

原始模型: $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \beta_6x_6 + \beta_7x_7 + \beta_8x_8 + \beta_9x_9 + \varepsilon_i$

配適模型: $\hat{y} = -0.7673 - 0.0353x_1 + 0.0533x_2 + 0.0016x_3 - 0.0022x_4 + 0.0007x_5 - 0.8025x_6 - 0.0792x_7 - 0.048x_8 + 0.0657x_9$

假設檢定

$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0$

$H_1: \text{至少一} \beta_j \text{ 不等於 } 0, j = 1, 2, 3, 4, 5, 6, 7, 8, 9$

因為 p 值為 0.3447，所以不拒絕 H_0 ；地震的震級與地震的緯度（lat）
地震的經度（lon）、震源深度（depth）、使用觀測站數量（nstn）、
最近站震央距（Dmin）、時間殘值之方均根誤差值（trms）、震央之
（水平）標準差（公里）（ERH）、IND、IND1 之間沒有線性關係。

變異數的分析					
來源	DF	平方和	均方	F 值	Pr > F
模型	9	1.68869	0.18763	1.13	0.3447
誤差	190	31.60086	0.16632		
已校正的總計	199	33.28955			

殘差分析

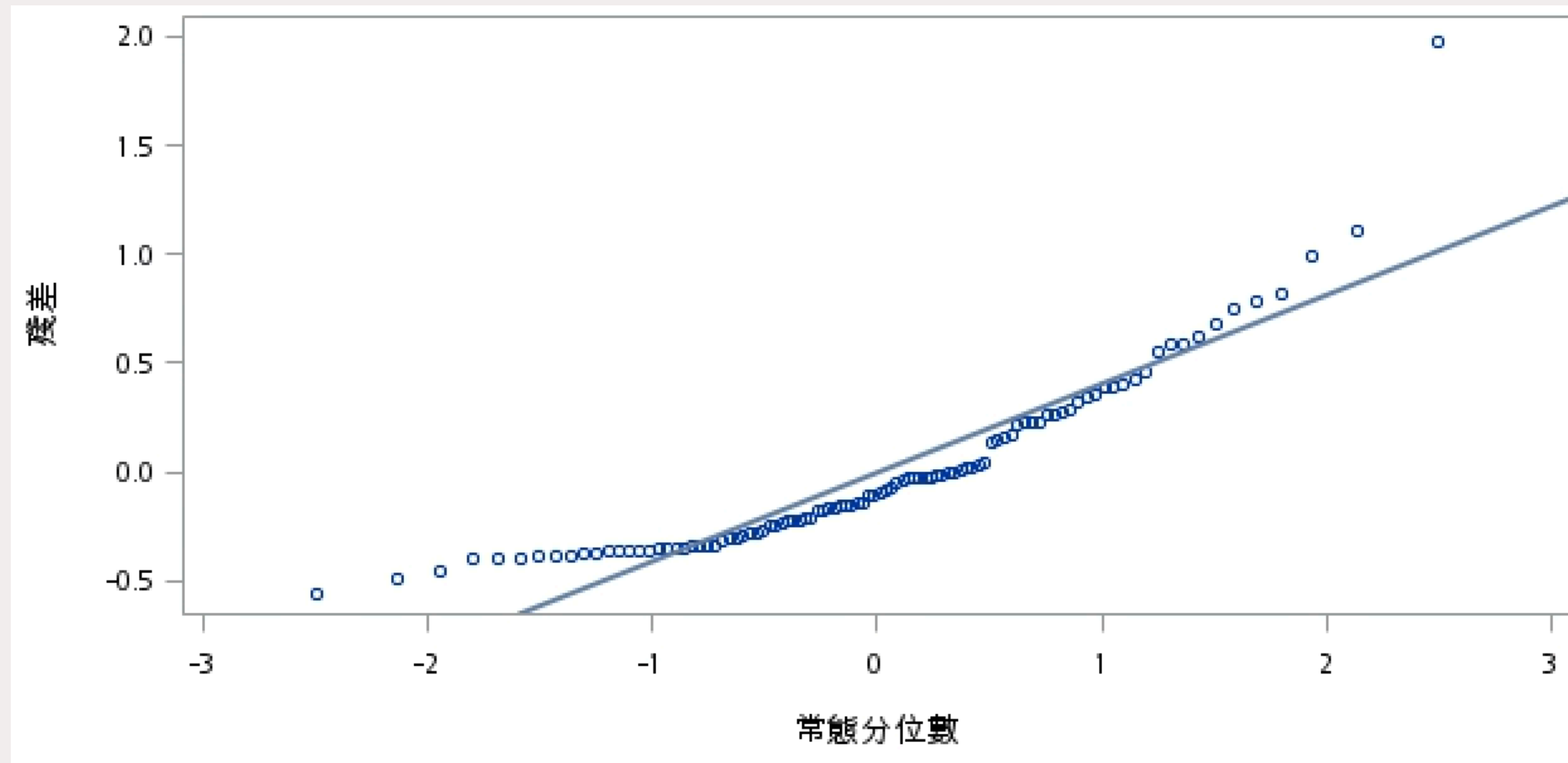
誤差項需滿足三大假設：

1.常態性(Normality)：可採用常態機率圖跟Shapiro-Wilk常態性檢定做檢查。

2.獨立性(Independency)：當樣本遠大於參數個數時，殘差之間的相依性可以忽略，此時誤差獨立性的假設成立。

3.變異數均質性(Constant Variance)：變異數若不相等會導致自變數無法有效估計應變數。

常態性(Normality)



常態機率圖近似直線，沒有違反誤差的常態假設，但還是需要進一步做檢定

常態性檢定

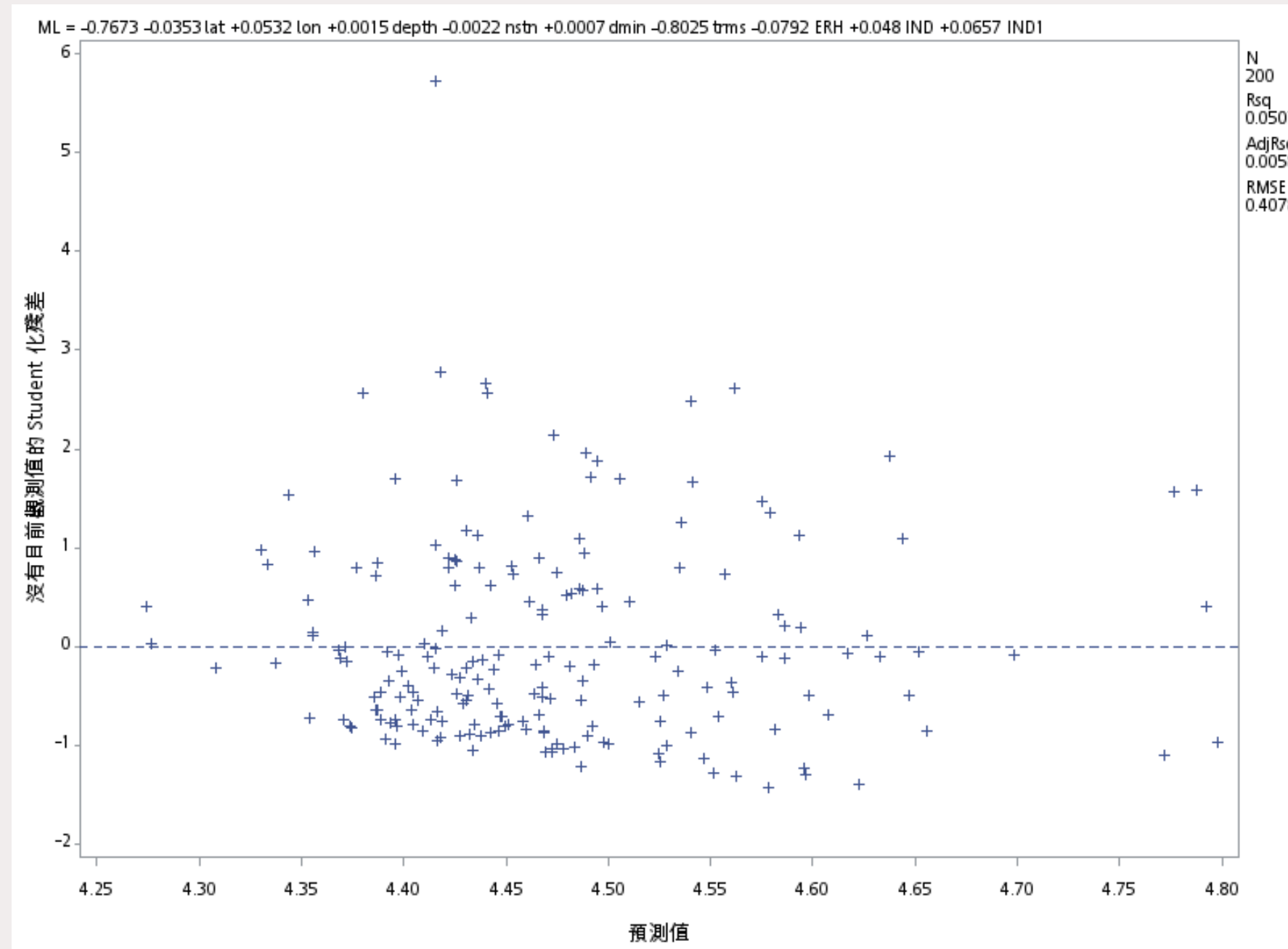
H_0 :誤差符合常態分佈

H_1 :誤差不符合常態分佈

由下表可得知因為 p-value 很小，有充分證據顯示誤差不服從常態分配。

常態性檢定				
檢定	統計值		P值	
Shapiro-Wilk	W	0.860089	Pr<W	<0.0001
Kolmogorov-Smirnov	D	0.149598	Pr>D	<0.0100
Carmer-von Mises	W-Sq	0.557593	Pr>W-Sq	<0.0050
Anderson-Darling	A-Sq	3.301743	Pr>A-Sq	<0.0050

變異數均質性(Constant Variance)



殘差隨機分佈在0的附近，表示誤差變異數可能為均質

變異數均質性檢定

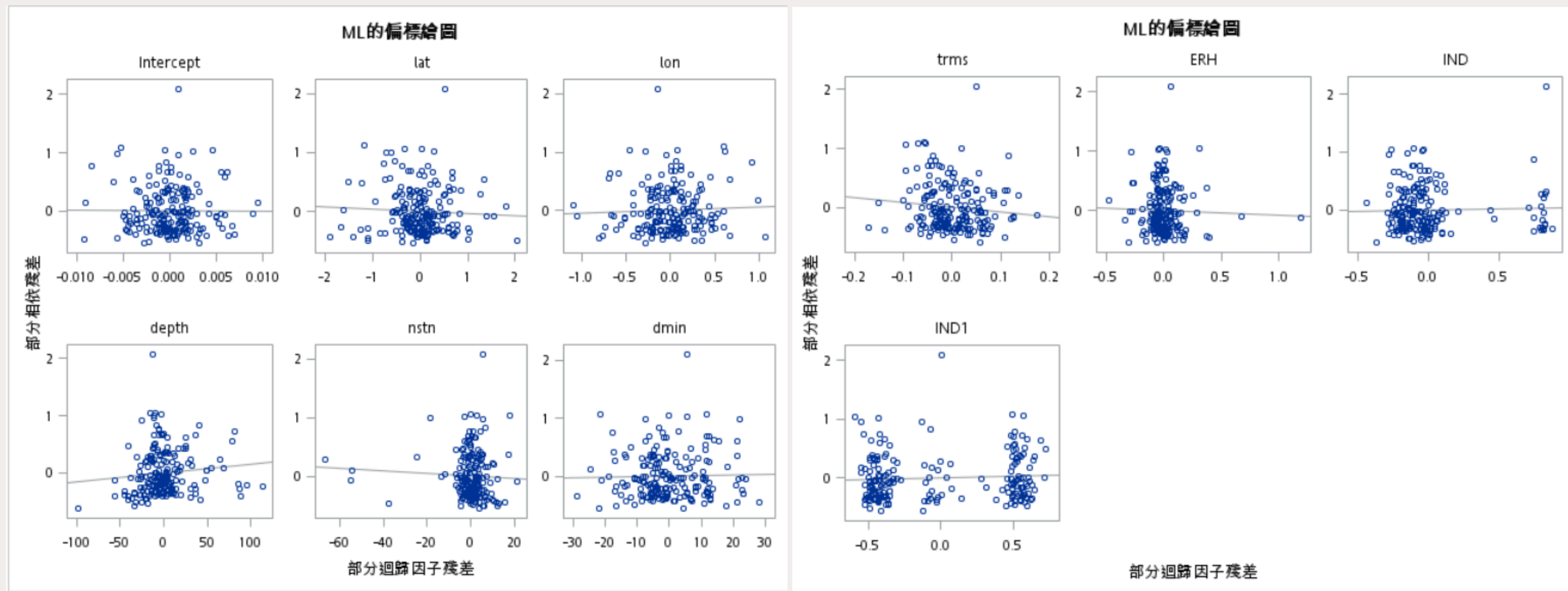
H_0 :誤差符合均質變異數的假設

H_1 :誤差符合異質變異數的假設

p-value=0.6329，不拒絕 H_0 ，

因此我們推斷誤差符合均質變異數的假設。

Residual變異數均齊性的Brown和Forsythe檢定來自群組中位數之絕對差的ANOVA					
來源	DF	平方和	均方	F值	Pr>F
Group	1	0.0177	0.0177	0.23	0.6329
誤差	198	15.2826	0.0772		



使用觀測站數量、震央之（水平）標準差（公里）較不符合線性關係，其餘7個變數都符合線性關係，表示他們適合放在模型中。

變數選取

向前選取

步階	已輸入變數	變數數目	偏R平方	模型R平方	C(p)	F 值	Pr > F
1	x3	1	0.0272	0.0272	-1.2813	5.53	0.0197
2	x4	2	0.0107	0.0379	-1.4271	2.20	0.1400
3	IND1	3	0.0034	0.0412	-0.0985	0.69	0.4086
4	x6	4	0.0034	0.0446	1.2242	0.69	0.4069

選取的變數有震源深度(x_3)、使用測站數量(x_4)、時間殘值之方均根誤差值(x_6)及虛擬變數(IND1)。

向前選取

考量虛擬變數只剩一個，因此我們將虛擬變數加入與剔除進行比較

模型中的數目	調整的R平方	R平方	C(p)	AIC	BIC	MSE	SSE	模型中的變數
3	0.0256	0.0403	0.0797	-358.8494	-356.5229	0.16299	31.94675	x3 x4 x6
5	0.0217	0.0463	2.8864	-356.0957	-353.5274	0.16365	31.74829	x3 x4 x6 IND IND1

只有 x_3 、 x_4 、 x_6 三個變數時表現較佳，因此選擇此模型作為向前選取的最佳模型。

向後刪除

步階	已移除變數	變數數目	偏R 平方	模型R 平方	C(p)	F 值	Pr > F
1	dmin	8	0.0004	0.0504	8.0715	0.07	0.7895
2	ERH	7	0.0006	0.0497	6.2016	0.13	0.7181
3	IND	6	0.0016	0.0482	4.5156	0.32	0.5741
4	nstn	5	0.0020	0.0461	2.9247	0.41	0.5205
5	IND1	4	0.0033	0.0428	1.5819	0.67	0.4149
6	lat	3	0.0040	0.0388	0.3794	0.81	0.3687
7	lon	2	0.0010	0.0379	-1.4271	0.20	0.6575
8	trms	1	0.0107	0.0272	-1.2813	2.20	0.1400

選取的變數只有震源深度(x_3)，其餘皆被刪除。

$$\hat{y} = 4.40498 + 0.00193x_3$$

變數	參數估計值	標準誤差	類型 II SS	F值	Pr>F
Intercept	4.40498	0.04074	1911.82805	11688.6	<.0001
x3	0.00193	0.00082144	0.90387	5.53	0.0197

逐步選取法

步階	已輸入變數	偏R 平方	模型R 平方	C(p)	F 值	Pr > F
1	x3	0.0272	0.0272	-1.2813	5.53	0.0197
2	x6	0.0107	0.0379	-1.4271	2.20	0.1400

選取的變數有震源深度(x_3)、時間殘值之方均根誤差值(x_6)。

$$\hat{y} = 4.62372 + 0.00192x_3 - 0.61967x_6$$

變數	參數估計值	標準誤差	類型 II SS	F 值	Pr > F
Intercept	4.62372	0.15313	148.23386	911.74	<.0001
x3	0.00192	0.00081898	0.89725	5.52	0.0198
x6	-0.61967	0.41825	0.35689	2.20	0.1400

模型中的數目	調整的R平方	R平方	C(p)	AIC	BIC	MSE	SSE	模型中的變數
1	0.0222	0.0272	-1.2813	-360.1202	-358.0132	0.16356	32.38568	x3
3	0.0256	0.0403	0.0797	-358.8494	-356.5229	0.16299	31.94675	x3 x4 x6
2	0.0281	0.0379	-1.4271	-360.3364	-358.1080	0.16258	32.02879	x3 x6

我們認為最適合的模型為透過逐步選取法篩選出來的：

$$\hat{y} = 4.62372 + 0.00192x_3 - 0.61967x_6$$

新模型之常態性檢定

H_0 :誤差符合常態分佈

H_1 :誤差不符合常態分佈

由下表可得知因為 p-value 很小，有充分證據顯示誤差不服從常態分配。

常態性檢定				
檢定	統計值		P值	
Shapiro-Wilk	W	0.887105	Pr<W	<0.0001
Kolmogorov-Smirnov	D	0.136903	Pr>D	<0.0100
Carmer-von Mises	W-Sq	0.964692	Pr>W-Sq	<0.0050
Anderson-Darling	A-Sq	5.68057	Pr>A-Sq	<0.0050

新模型之變異數均質性檢定

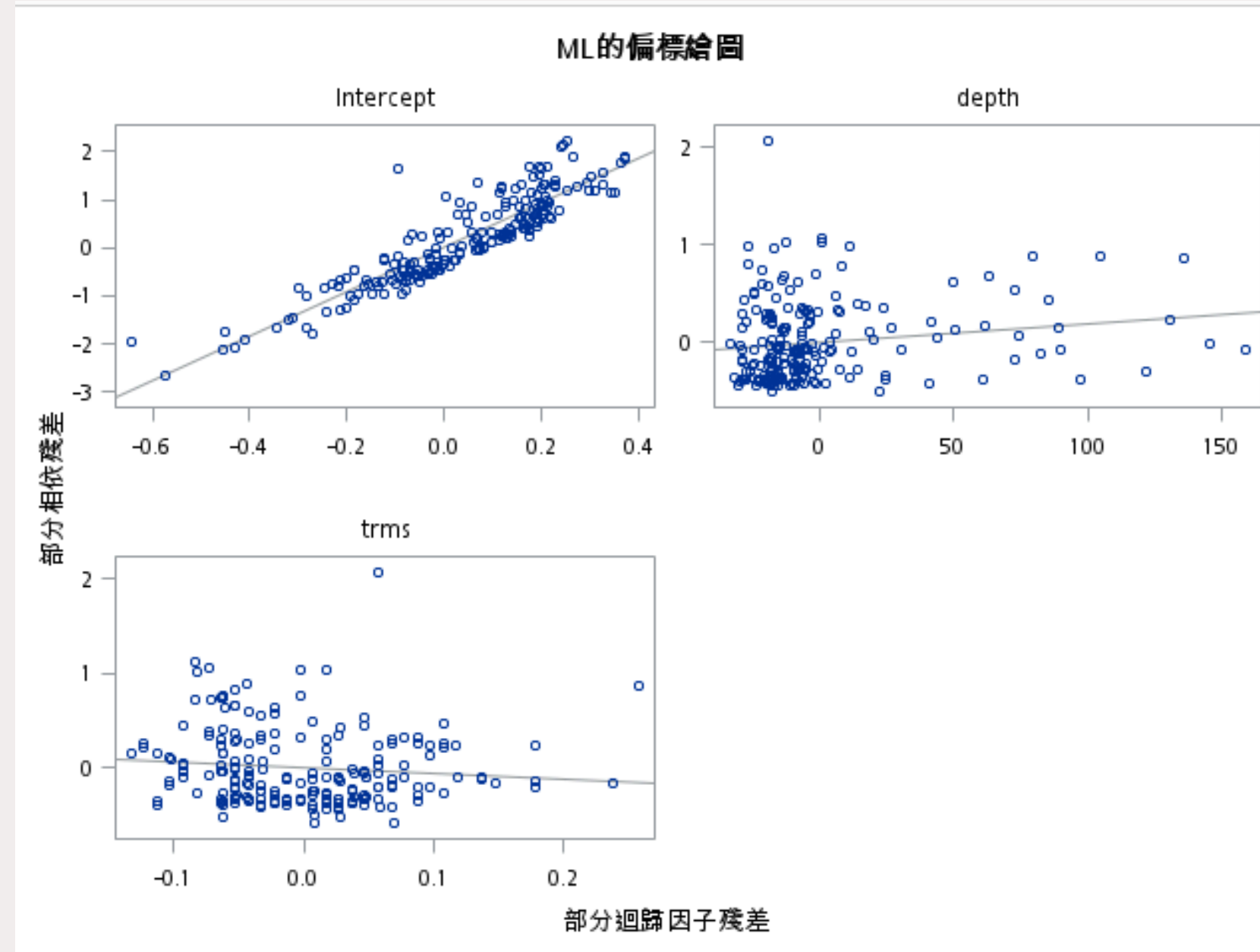
H_0 :誤差符合均質變異數的假設

H_1 :誤差符合異質變異數的假設

p-value=0.676，不拒絕 H_0

我們發現新模型的誤差一樣符合均質變異數的假設。

Residual變異數均齊性的Brown和Forsythe檢定來自群組中位數之絕對差的ANOVA					
來源	DF	平方和	均方	F值	Pr>F
Group	1	0.0141	0.0141	0.18	0.6760
誤差	198	15.9141	0.0804		



兩個變數都呈現線性關係，表示它們適合放在模型中。

離群值

觀測值	Student	RStudent
5	5.29463	5.71944
22	2.11307	2.13271
59	2.57379	2.61296
95	2.52490	2.56158
159	2.52658	2.56336
172	2.62611	2.66805
176	2.45183	2.48499

Studentized residual跟R-student residual
的絕對值大於2，代表可能為離群值。

影響點

COOK's D 的準則為： $D_i > \frac{4}{n} = \frac{4}{200} = 0.02$

當 D_i 大於 0.02 時，即為影響點；其影響點為:5、17、22、43、54、
72、74、121、130、159、172、176、197、198。

影響點

$DFFITs$ 的準則為： $|DFFITs_i| > 2\sqrt{\frac{p}{n}} = 2\sqrt{\frac{3}{200}} = 0.2449489743$

當 $|DFFITs_i|$ 大於 0.2449489743 時，即為影響點；

其影響點為：5、7、12、16、17、22、23、24、40、42、43、45、
54、59、63、64、67、69、72、74、80、91、95、102、118、121、
130、139、148、159、172、174、176、189、192、197、198、
199。

影響點

透過這兩種影響點的方法，我們發現有 14 筆資料有重複，而以 *DFFITs* 為標準則找出較多的影響點。

觀測值	COOK	DFFITs	觀測值	COOK	DFFITs
5	0.17549	1.43100	121	0.02147	0.46682
17	0.02209	-0.46991	130	0.02447	-0.49603
22	0.02752	0.52946	159	0.09686	0.99851
43	0.03850	0.62288	172	0.02781	0.53579
54	0.04014	-0.63474	176	0.02478	0.50453
72	0.02566	0.50859	197	0.02337	-0.48381
74	0.02467	-0.49700	198	0.06115	71

預測

使用此迴歸模型來預測: $\hat{y} = 4.62372 + 0.00192x_3 - 0.61967x_6$

另外選擇三筆資料來觀察模型的預測能力，我們使用的迴歸模型跑出來之地震規模預測值為 4.43、4.52、4.54，而實際上的地震規模分別為 4.23、4.65、4.15，可以看出迴歸模型的預測能力與真實地震規模非常接近。

觀察值	應變數	預測值	標準誤差 平均值預測	95%CL平均值	95%CL預測	殘差
201	.	4.4289	4.3299	4.3299 4.5280	3.6276 5.2302	.
202	.	4.5198	4.4537	4.4537 4.5859	3.7219 5.3177	.
203	.	4.5386	4.4436	4.4436 4.6337	3.7378 5.3394	.

結論

經過變數選取後，選出了 x_3 、 x_6 這兩個解釋變數較適合放入模型中，加上預測能力不錯，於是選用的模型為： $\hat{y} = 4.62372 + 0.00192x_3 - 0.61967x_6$

最後得知震源深度和時間殘值之方均根誤差值這兩個變數會影響地震的規模，相較於其他變數較為顯著。