

應用統計報告

Youtube 觀看次數大解析

指導老師

李信宏 教授

組員

S0522108 洪羽柔

S0522112 艾品璇

S0522131 陳晏琦

S0522143 顏均翰

2019 年 6 月 12 日

目錄

資料介紹	4
變數介紹及說明	4
範例說明	5
列舉 66 筆數據	6
模型挑選	8
逐步選擇法(Stepwise Selection)	8
向前選擇法(Forward Selection)	10
後退消除法(Backward Elimination)	12
最佳子集迴歸(Best Subsets Regression)	14
模型說明	15
迴歸分析	16
初步迴歸分析	16
刪除離群值(outlier)之迴歸分析	22
應變數之線性轉換	29
最終模型與解釋	37
結論	42
心得感想	43
資料來源	45

研究動機

身處在 e 世代，導致近幾年來 Youtube 的使用率逐漸攀升，平時我們觀看 Youtube 影片除了能夠打發時間之外，還能透過觀看 Youtube 來吸收各方面的知識，因此引起我們想要研究 Youtube 的動機，此外想要從事 Youtuber 的人也大幅增加，藉由拍攝許多不同題材的影片吸引了各年齡層的觀眾，而觀看次數的多寡決定了 Youtuber 的收入，所以我們想要了解什麼因素會影響觀看次數，藉此我們利用課堂上所學習到的迴歸分析技術，來深入探討影響觀看次數的原因。

資料介紹

資料抓取時間：以 2019 年 5 月 10 日下午 3 點左右為基準

- 變數介紹及說明

變數	變數說明	單位
Y	Youtube 影片觀看次數(抓 2019 年 3 月底 66 個 youtuber 上傳的一支影片)	萬次
X_1	Youtuber 的頻道訂閱數	萬人
X_2	Youtuber 的 facebook 追蹤人數	萬人
X_3	Youtuber 自行置入影片的廣告數(時間軸上的一個黃點就代表置入一支廣告)	個
X_4	片長	分鐘
X_5	Youtuber 副頻道數	個
X_6	影片合作人數	人
X_7	Hashtag 數	個
X_8	頻道類型：娛樂	
X_9	頻道類型：遊戲	
X_{10}	頻道類型：教育	
X_{11}	頻道類型：人物與日誌	

- 範例說明



#娃娃機 #公益 #環島 **X7=3**
 南部人的熱情！你看的見！【環島之旅-中】 | 含羞草日記X在不瘋狂就等死 **X6=1**
 觀看次數：**268,062次** **Y=26.8062(萬次)** 2540 87 分享 儲存 ...



含羞草
 發佈日期：2019年3月30日

加入

訂閱 (74萬)

【含羞草日記特別企劃系列】去你的台主
<https://goo.gl/9zFUXL>

類別 **娛樂** **X8=1**

只顯示部分資訊





含羞草 ✓

745,459 位訂閱者 **X1=74.5459(萬人)**

首頁 影片 播放清單 社群 頻道 簡介

所有頻道 ▾

合作頻道大集合



在不瘋狂就等死 x 狂人娛樂

1,034,527 位訂閱者

訂閱

X5=1



含老闆

138,886 位訂閱者

訂閱



瘋狂概念 **Crazy Concept**

95,695 位訂閱者

訂閱

- 列舉 66 筆數據

說明：

我們綜合維基百科、新知筆記 Knowledge Notes 的【台灣前 100 名 Youtuber 訂閱排行榜，你認識幾個呢? (2019 年 2 月)】影片，抓出 66 名較知名的臺灣 Youtuber。

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
2	這群人	186.5913	291.4566	197.7740	1	5.766666667	1	0	0	1	0	0	0
3	阿滴英文	76.3129	219.4632	71.5746	3	10.93333333	0	0	0	0	0	1	0
4	蔡阿嘎	49.6259	206.3970	172.0024	1	3.8	3	0	0	0	0	0	0
5	阿神	29.7891	187.7127	9.5295	5	20.9	1	0	1	0	1	0	0
6	古阿莫	57.2066	168.3972	75.8401	1	5	3	0	0	0	0	0	0
7	重量級CROWD	51.073	161.6658	132.4127	5	14.68333333	3	0	3	0	0	0	0
8	Joeman	144.6374	151.7579	29.8357	7	32.55	0	1	0	0	0	0	1
9	聖結石	47.3418	147.7294	66.8460	3	12.96666667	1	2	0	0	0	0	0
10	DE Jun	25.6967	146.8149	26.8512	2	12.16666667	0	0	0	0	1	0	0
11	小玉	129.2383	137.6543	30.3669	1	2.45	0	1	0	0	0	0	1
12	放火 Louis	67.363	134.5692	78.7551	1	3.35	1	2	0	0	0	0	0
13	安啾咪	39.9048	134.6361	53.1160	1	8.883333333	1	0	0	0	0	0	1
14	WACKBOYS 反骨男孩	39.1476	129.7082	63.8071	1	6.7	0	0	0	1	0	0	0
15	木曜4超玩	97.4912	128.1331	26.3280	4	70.36666667	1	1	3	1	0	0	0
16	菜喳	12.0357	125.8328	43.1746	0	10.55	0	1	3	0	1	0	0
17	人生肥宅X尊	52.7554	125.0779	39.4070	1	4.65	0	0	0	0	0	0	1
18	白癡公主	50.3393	124.7910	107.5830	2	11.41666667	0	0	3	1	0	0	0
19	魚乾	6.223	121.0740	35.2289	1	8.533333333	0	0	2	0	1	0	0
20	黃阿瑪的後宮生活	113.4823	117.4216	133.4785	1	7.8	0	0	0	0	0	0	0
21	HowFun	59.4892	110.3562	50.3829	1	2.95	0	0	3	1	0	0	0
22	滴妹	47.6495	108.0146	21.3899	1	7.2	0	1	0	0	0	0	1
23	理科太太 Li Ke Tai Tai	19.7645	106.2399	51.2783	0	16.58333333	0	0	1	0	0	1	0
24	三原JAPAN Sanyuan_JAPAN	60.4288	105.9070	36.6164	1	8.116666667	2	0	0	0	0	0	1
25	黃氏兄弟	18.7787	105.4942	9.5141	0	47.63333333	1	0	0	0	0	0	1
26	千千進食中	70.9482	104.7303	46.0894	2	11.5	0	0	0	0	0	0	1
27	sandy mandy	260.9366	105.2306	257.1839	1	3.333333333	0	0	2	1	0	0	0
28	在不瘋狂就等死x狂人娛樂	59.2295	103.0585	191.5539	1	13.35	13	2	1	0	0	0	0
29	舞秋風	1.9546	101.5444	22.4836	1	5.433333333	2	0	0	0	1	0	0
30	老皮	6.8438	48.4893	54.0968	4	117.6833333	1	0	0	0	1	0	0
31	啾啾鞋	15.6119	97.6745	11.8181	2	7.3	2	0	0	1	0	0	0
32	巧克力	5.8814	95.4939	15.2452	5	14.6	2	0	3	0	1	0	0
33	the劉沛	10.6631	91.9489	10.9691	3	12.2	2	0	3	0	1	0	0
34	Stopkiddinstudio	5.3362	89.6924	14.5038	2	2.316666667	1	0	0	0	0	1	0
35	古娃娃 WawaKu	31.491	88.7744	39.2893	2	8.366666667	1	0	0	0	0	0	1
36	Hello Catie	27.5375	85.2925	31.1879	4	20.85	0	0	3	0	0	0	0
37	搞神馬	42.2448	80.5950	5.1088	7	12.1	2	0	0	0	0	0	1
38	頑GAME	1.5058	74.1769	38.7879	4	10.48333333	1	4	3	1	0	0	0
39	MaoMao TV	9.0427	74.3848	12.9666	2	8.183333333	0	0	0	0	0	0	1
40	含羞草	26.4811	74.2528	13.6644	6	30	1	1	3	1	0	0	0
41	NyoNyoTV 妞妞 TV	18.7429	73.8568	4.7839	4	12.98333333	2	0	0	1	0	0	0
42	小白	17.0593	72.0594	5.8103	4	11.08333333	0	0	0	0	1	0	0
43	Taiwan Bar	8.6206	71.0374	29.6429	2	4.983333333	0	0	3	0	0	1	0
44	鬼鬼	0.4033	70.5972	12.6658	2	9.733333333	1	0	1	0	1	0	0
45	小草Yue	48.5515	70.9233	11.5197	4	12.43333333	1	0	3	0	1	0	0
46	狠愛演	78.5504	70.3629	32.5912	4	10.11666667	3	0	0	0	0	0	1
47	聖燄Dodo	11.0488	68.5488	52.0444	4	11.01666667	1	0	0	0	0	0	1
48	展榮展瑞 K.R Bros	33.9714	68.2666	69.5907	0	14.13333333	3	1	3	0	0	0	0
49	黃大謙	65.7625	68.3066	36.3344	3	10.1	0	1	0	0	0	0	0
50	星培 Jasper	10.3	65.5254	68.7464	1	4.65	0	0	0	1	0	0	0
51	英雄日常 Heroisme	11.8527	66.1975	6.4491	4	11.2	1	0	0	0	0	0	1
52	GINA HELLO!	10.3709	56.8191	44.7155	1	11.33333333	0	0	2	1	0	0	0
53	超粒方	7.0173	63.8252	12.1736	1	3.566666667	1	0	0	0	0	0	0
54	館長成吉思汗	22.9196	58.7181	97.1363	2	11.5	1	1	3	0	0	0	0
55	BuBuChaCha 傳說	17.1208	51.1362	7.1374	2	3.05	0	0	3	0	1	0	0
56	烏鴉 DoKa TV	74.043	61.0647	1.6641	1	9.733333333	0	5	0	0	0	0	1
57	愛莉莎莎 Alisasa	75.2266	61.2346	12.3827	3	10.46666667	0	1	0	0	0	0	1
58	百變沛莉 Peri	9.223	53.4387	30.8740	3	14.45	1	0	0	0	0	0	0
59	馬叔叔 UNCLE MA	1.1203	50.6677	41.2072	2	1.683333333	0	0	0	0	0	0	0
60	上班不要看 NSFW	27.1731	56.3857	39.9726	5	16.16666667	0	0	3	0	0	0	0
61	台客劇場 Tkstory	20.8992	51.5003	42.1632	0	9.1	0	0	3	1	0	0	0
62	瘋狂老爹	6.0371	50.3234	5.1204	2	8.833333333	0	5	0	0	1	0	0
63	I.C Charlie	6.8785	47.4454	4.5653	2	6.483333333	0	0	0	0	0	0	0
64	TheKellyYang	24.8384	46.9617	12.5995	5	31.65	0	0	3	0	0	0	1
65	Dinter	14.6551	46.3679	27.8490	3	10.11666667	0	0	0	0	1	0	0
66	那對夫妻	9.4587	46.6746	246.2822	1	8.4	0	3	3	1	0	0	0
67	胡子Huzi	41.4183	51.8812	3.8683	4	10.35	0	4	3	1	0	0	0

模型挑選

- 逐步選擇法(Stepwise Selection)

Stepwise Selection of Terms

Candidate terms: x1(訂閱數)(萬), x2(FB追蹤人數)(萬人), x3(廣告數), x4(片長 分鐘), x5(副頻道數), x6(ft.人數), x7(hashtag數), x8(頻道類型:娛樂), x9(頻道類型:遊戲), x10(頻道類型:教育), x11(頻道類型:人物與日誌)

	-----Step 1-----		-----Step 2-----		-----Step 3-----	
	Coef	P	Coef	P	Coef	P
Constant	20.34		-7.5		-17.2	
x2(FB追蹤人數)(萬人)	0.4225	0.000	0.3116	0.001	0.3790	0.000
x1(訂閱數)(萬)			0.342	0.002	0.3253	0.002
x11(頻道類型:人物與日誌)					31.2	0.003
S	39.3855		36.7030		34.4038	
R-sq	26.85%		37.47%		45.93%	
R-sq(adj)	25.71%		35.48%		43.31%	
R-sq(pred)	8.79%		14.46%		21.93%	
Mallows' Cp	21.28		11.19		3.56	

α to enter = 0.05, α to remove = 0.05

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	62339	20780	17.56	0.000
x1(訂閱數)(萬)	1	12991	12991	10.98	0.002
x2(FB追蹤人數)(萬人)	1	23369	23369	19.74	0.000
x11(頻道類型:人物與日誌)	1	11484	11484	9.70	0.003
Error	62	73384	1184		
Total	65	135723			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
34.4038	45.93%	43.31%	21.93%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-17.2	10.3	-1.68	0.099	
x1(訂閱數)(萬)	0.3253	0.0982	3.31	0.002	1.18
x2(FB追蹤人數)(萬人)	0.3790	0.0853	4.44	0.000	1.26
x11(頻道類型:人物與日誌)	31.2	10.0	3.11	0.003	1.07

Regression Equation

$$Y(\text{觀看次數(萬人)}) = -17.2 + 0.3253 \times x_1(\text{訂閱數(萬)}) + 0.3790 \times x_2(\text{FB追蹤人數(萬人)}) + 31.2 \times x_3(\text{頻道類型:人物與日誌})$$

Fits and Diagnostics for Unusual Observations

Obs	Y(觀看次數(萬人))	Fit	Resid	Std Resid		
1	186.6	152.5	34.1	1.18		X
3	49.6	115.1	-65.5	-2.04	R	
7	144.6	74.7	70.0	2.13	R	
26	260.9	114.5	146.5	4.94	R	X
65	9.5	91.3	-81.8	-2.87	R	X

R Large residual

X Unusual X

- 向前選擇法(Forward Selection)

Forward Selection of Terms

Candidate terms: x1(訂閱數)(萬), x2(FB追蹤人數)(萬人), x3(廣告數), x4(片長 分鐘), x5(副頻道數), x6(ft.人數), x7(hashtag數), x8(頻道類型:娛樂), x9(頻道類型:遊戲), x10(頻道類型:教育), x11(頻道類型:人物與日誌)

	-----Step 1-----		-----Step 2-----		-----Step 3-----	
	Coef	P	Coef	P	Coef	P
Constant	20.34		-7.5		-17.2	
x2(FB追蹤人數)(萬人)	0.4225	0.000	0.3116	0.001	0.3790	0.000
x1(訂閱數)(萬)			0.342	0.002	0.3253	0.002
x11(頻道類型:人物與日誌)					31.2	0.003
S		39.3855		36.7030		34.4038
R-sq		26.85%		37.47%		45.93%
R-sq(adj)		25.71%		35.48%		43.31%
R-sq(pred)		8.79%		14.46%		21.93%
Mallows' Cp		21.28		11.19		3.56

α to enter = 0.05

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	62339	20780	17.56	0.000
x1(訂閱數)(萬)	1	12991	12991	10.98	0.002
x2(FB追蹤人數)(萬人)	1	23369	23369	19.74	0.000
x11(頻道類型:人物與日誌)	1	11484	11484	9.70	0.003
Error	62	73384	1184		
Total	65	135723			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
34.4038	45.93%	43.31%	21.93%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-17.2	10.3	-1.68	0.099	
x1(訂閱數)(萬)	0.3253	0.0982	3.31	0.002	1.18
x2(FB追蹤人數)(萬人)	0.3790	0.0853	4.44	0.000	1.26
x11(頻道類型:人物與日誌)	31.2	10.0	3.11	0.003	1.07

Regression Equation

$$Y(\text{觀看次數(萬人)}) = -17.2 + 0.3253 x_1(\text{訂閱數(萬)}) + 0.3790 x_2(\text{FB追蹤人數(萬人)}) + 31.2 x_3(\text{頻道類型:人物與日誌})$$

Fits and Diagnostics for Unusual Observations

Obs	Y(觀看次數(萬人))	Fit	Resid	Std Resid	
1	186.6	152.5	34.1	1.18	X
3	49.6	115.1	-65.5	-2.04	R
7	144.6	74.7	70.0	2.13	R
26	260.9	114.5	146.5	4.94	R X
65	9.5	91.3	-81.8	-2.87	R X

R Large residual

X Unusual X

- 後退消除法(Backward Elimination)

Backward Elimination of Terms

Candidate terms: x1(訂閱數)(萬), x2(FB追蹤人數)(萬人), x3(廣告數), x4(片長 分鐘), x5(副頻道數), x6(ft.人數), x7(hashtag數), x8(頻道類型:娛樂), x9(頻道類型:遊戲), x10(頻道類型:教育), x11(頻道類型:人物與日誌)

	----Step 1----		----Step 2----		----Step 3----	
	Coef	P	Coef	P	Coef	P
Constant	-26.0		-25.9		-27.3	
x1(訂閱數)(萬)	0.353	0.002	0.352	0.001	0.362	0.001
x2(FB追蹤人數)(萬人)	0.408	0.000	0.408	0.000	0.402	0.000
x3(廣告數)	3.80	0.189	3.87	0.157	3.65	0.168
x4(片長 分鐘)	0.023	0.935				
x5(副頻道數)	-4.36	0.108	-4.35	0.105	-4.32	0.105
x6(ft.人數)	2.39	0.523	2.38	0.522	2.33	0.526
x7(hashtag數)	-1.35	0.719	-1.35	0.716		
x8(頻道類型:娛樂)	9.8	0.469	9.9	0.459	8.8	0.495
x9(頻道類型:遊戲)	-7.0	0.618	-6.8	0.620	-7.4	0.584
x10(頻道類型:教育)	-10.9	0.596	-10.8	0.593	-11.3	0.572
x11(頻道類型:人物與日誌)	28.1	0.040	28.2	0.037	29.2	0.026
S		34.5272		34.2140		33.9483
R-sq		52.57%		52.56%		52.45%
R-sq(adj)		42.91%		43.94%		44.81%
R-sq(pred)		1.98%		5.52%		6.18%
Mallows' Cp		12.00		10.01		8.14

	-----Step 4-----		-----Step 5-----		-----Step 6-----	
	Coef	P	Coef	P	Coef	P
Constant	-31.3		-32.2		-30.0	
x1(訂閱數)(萬)	0.3570	0.001	0.3514	0.001	0.3381	0.001
x2(FB追蹤人數)(萬人)	0.4186	0.000	0.4213	0.000	0.4245	0.000
x3(廣告數)	3.64	0.166	3.74	0.149	3.71	0.151
x4(片長 分鐘)						
x5(副頻道數)	-4.14	0.114	-4.01	0.120	-3.89	0.129
x6(ft.人數)	2.37	0.518	2.46	0.497		
x7(hashtag數)						
x8(頻道類型:娛樂)	11.9	0.300	12.8	0.255	13.8	0.213
x9(頻道類型:遊戲)						
x10(頻道類型:教育)	-7.4	0.691				
x11(頻道類型:人物與日誌)	33.1	0.003	34.1	0.001	34.3	0.001
S		33.7403		33.4950		33.3433
R-sq		52.19%		52.06%		51.67%
R-sq(adj)		45.48%		46.27%		46.76%
R-sq(pred)		10.96%		11.79%		16.42%
Mallows' Cp		6.43		4.58		3.02

	-----Step 7-----		-----Step 8-----		-----Step 9-----	
	Coef	P	Coef	P	Coef	P
Constant	-26.4		-15.4		-17.2	
x1(訂閱數)(萬)	0.3269	0.001	0.3268	0.001	0.3253	0.002
x2(FB追蹤人數)(萬人)	0.4591	0.000	0.4193	0.000	0.3790	0.000
x3(廣告數)	3.89	0.134				
x4(片長 分鐘)						
x5(副頻道數)	-4.64	0.065	-4.29	0.089		
x6(ft.人數)						
x7(hashtag數)						
x8(頻道類型:娛樂)						
x9(頻道類型:遊戲)						
x10(頻道類型:教育)						
x11(頻道類型:人物與日誌)	30.89	0.002	31.26	0.002	31.2	0.003
S		33.5063		33.8642		34.4038
R-sq		50.37%		48.46%		45.93%
R-sq(adj)		46.23%		45.08%		43.31%
R-sq(pred)		15.72%		17.32%		21.93%
Mallows' Cp		2.50		2.68		3.56

α to remove = 0.05

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	62339	20780	17.56	0.000
x1(訂閱數)(萬)	1	12991	12991	10.98	0.002
x2(FB追蹤人數)(萬人)	1	23369	23369	19.74	0.000
x11(頻道類型:人物與日誌)	1	11484	11484	9.70	0.003
Error	62	73384	1184		
Total	65	135723			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
34.4038	45.93%	43.31%	21.93%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-17.2	10.3	-1.68	0.099	
x1(訂閱數)(萬)	0.3253	0.0982	3.31	0.002	1.18
x2(FB追蹤人數)(萬人)	0.3790	0.0853	4.44	0.000	1.26
x11(頻道類型:人物與日誌)	31.2	10.0	3.11	0.003	1.07

Regression Equation

$$Y(\text{觀看次數(萬人)}) = -17.2 + 0.3253 x_1(\text{訂閱數(萬)}) + 0.3790 x_2(\text{FB追蹤人數(萬人)}) + 31.2 x_{11}(\text{頻道類型:人物與日誌})$$

Fits and Diagnostics for Unusual Observations

Obs	Y(觀看次數(萬人))	Fit	Resid	Std Resid	
1	186.6	152.5	34.1	1.18	X
3	49.6	115.1	-65.5	-2.04	R
7	144.6	74.7	70.0	2.13	R
26	260.9	114.5	146.5	4.94	R X
65	9.5	91.3	-81.8	-2.87	R X

R Large residual

X Unusual X

- 最佳子集迴歸(Best Subsets Regression)

Best Subsets Regression: Y versus X1, X2, X3, X4, X5, X6, ... X9, X10, X11

Response is Y

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	X 1	X 2	X 3	X 4	X 5	X 6	X 7	X 8	X 9	X 10	X 11
1	26.9	25.7	8.8	21.3	39.385		X									
1	25.0	23.9	19.7	23.3	39.872	X										
2	37.5	35.5	14.5	11.2	36.703	X	X									
2	36.4	34.3	16.8	12.5	37.028	X										
3	45.9	43.3	21.9	3.6	34.404	X	X									
3	41.0	38.1	16.6	9.2	35.947	X	X						X			
4	48.5	45.1	17.3	2.7	33.864	X	X			X						
4	48.3	44.9	23.2	2.8	33.907	X	X						X			
5	50.4	46.2	15.7	2.5	33.506	X	X	X		X						
5	49.9	45.8	18.0	3.0	33.654	X	X			X			X			
6	51.7	46.8	16.4	3.0	33.343	X	X	X		X			X			
6	51.0	46.0	11.5	3.8	33.587	X	X	X		X	X					
7	52.1	46.3	11.8	4.6	33.495	X	X	X		X	X		X			
7	51.8	46.0	15.7	4.8	33.572	X	X	X		X			X		X	
8	52.2	45.5	11.2	6.4	33.729	X	X	X		X	X	X	X			
8	52.2	45.5	11.0	6.4	33.740	X	X	X		X	X		X		X	
9	52.4	44.8	6.2	8.1	33.948	X	X	X		X	X		X	X	X	
9	52.3	44.7	10.4	8.3	33.984	X	X	X		X	X	X	X		X	
10	52.6	43.9	5.5	10.0	34.214	X	X	X		X	X	X	X	X	X	
10	52.5	43.8	3.6	10.1	34.253	X	X	X	X	X	X		X	X	X	
11	52.6	42.9	2.0	12.0	34.527	X	X	X	X	X	X	X	X	X	X	

- 模型說明

根據 逐步選擇法(Stepwise Selection)、向前選擇法(Forward Selection) 和 後退消除法(Backward Elimination) 三種挑選模型的方式中，分析出的模型應變數 Y (影片觀看次數)都是與 X_1 (頻道訂閱數)、 X_2 (facebook 追蹤人數)和 X_{11} (頻道類型：人物與日誌) 這三個自變數有關。而透過最佳子集迴歸(Best Subsets Regression)的數據顯示中，我們可以觀察到此模型的 C_p 值為 3.6，相較於其他模型的 C_p 值，相對接近其 p 值($p=4$)。加上，其他的模型因為 C_p 值與其 p 值相差較大，或是變數增加但 $R\text{-sq}(\text{adj})$ 值卻沒有明顯的提升等因素考量，例如：變數個數為 6 的第一筆資料，雖然它的 $R\text{-sq}(\text{adj})$ 上升了 3.5，但多了三個變數，且 C_p 值與 p 值差了 4，因此我們最終選的模型為：

$$\hat{Y} = -17.2 + 0.3253X_1 + 0.3790X_2 + 31.2X_{11}$$

迴歸分析

- 初步迴歸分析

Regression Analysis: Y(觀看次數(萬人) versus ... 1(頻道類型:人物與日誌)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	62339	20780	17.56	0.000
x1(訂閱數)(萬)	1	12991	12991	10.98	0.002
x2(FB追蹤人數)(萬人)	1	23369	23369	19.74	0.000
x11(頻道類型:人物與日誌)	1	11484	11484	9.70	0.003
Error	62	73384	1184		
Total	65	135723			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
34.4038	45.93%	43.31%	21.93%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-17.2	10.3	-1.68	0.099	
x1(訂閱數)(萬)	0.3253	0.0982	3.31	0.002	1.18
x2(FB追蹤人數)(萬人)	0.3790	0.0853	4.44	0.000	1.26
x11(頻道類型:人物與日誌)	31.2	10.0	3.11	0.003	1.07

Regression Equation

Y(觀看次數(萬人)) = -17.2 + 0.3253 x1(訂閱數)(萬) + 0.3790 x2(FB追蹤人數)(萬人) + 31.2 x11(頻道類型:人物與日誌)

Fits and Diagnostics for Unusual Observations

Obs	Y(觀看次數(萬人))	Fit	Resid	Std Resid	
1	186.6	152.5	34.1	1.18	X
3	49.6	115.1	-65.5	-2.04	R
7	144.6	74.7	70.0	2.13	R
26	260.9	114.5	146.5	4.94	R X
65	9.5	91.3	-81.8	-2.87	R X

R Large residual
X Unusual X

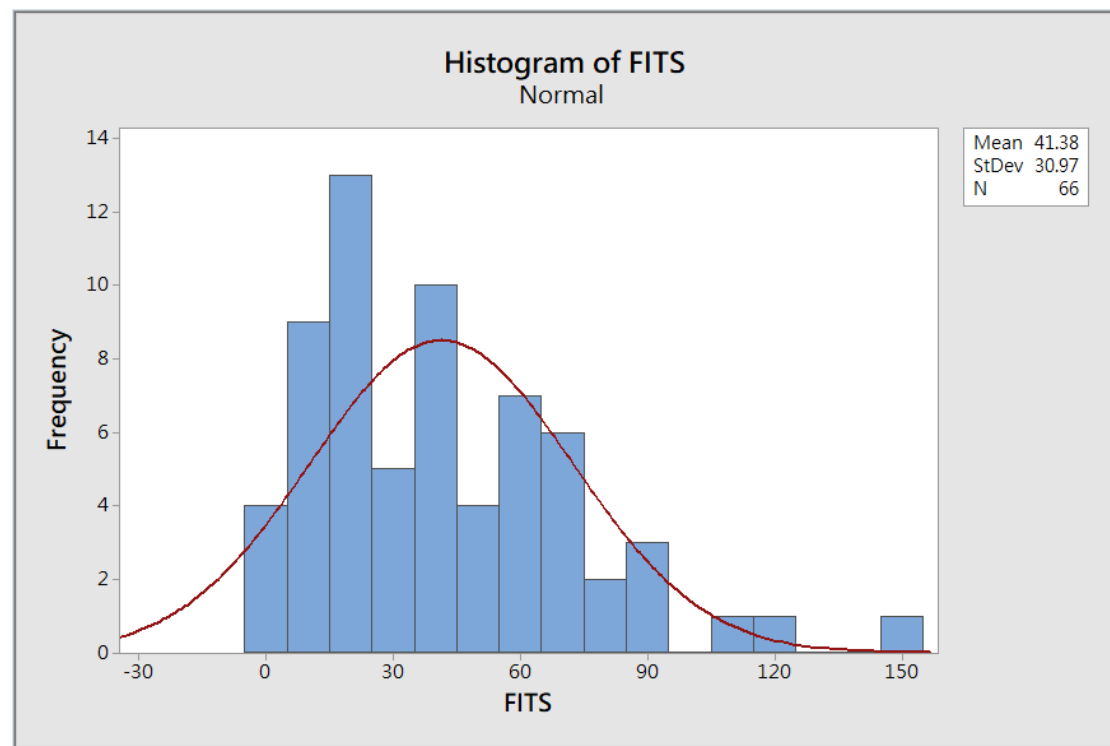
分析：

$$17.56 > F_{0.05}(3,62) = 2.75297$$

$$3.31、4.44、311 \text{ 皆} > T_{0.975}(62) = 1.99897$$

藉由 F 檢定可以發現有很大的機會模型中自變數的係數至少有一個非零，且根據 t 檢定可以知道自變數與應變數之間是有關係的。再者，我們也觀察到各別的 P-Value 都相當小，因此我們更加確認我們的模型是顯著的。從各別的 VIF 值得知我們的自變數之間沒有存在共線性關係。而我們也發現第 3、7、26、65 筆資料的標準化殘差較大，因此我們將在後頭藉由圖來觀察這四筆資料是否為離群值。

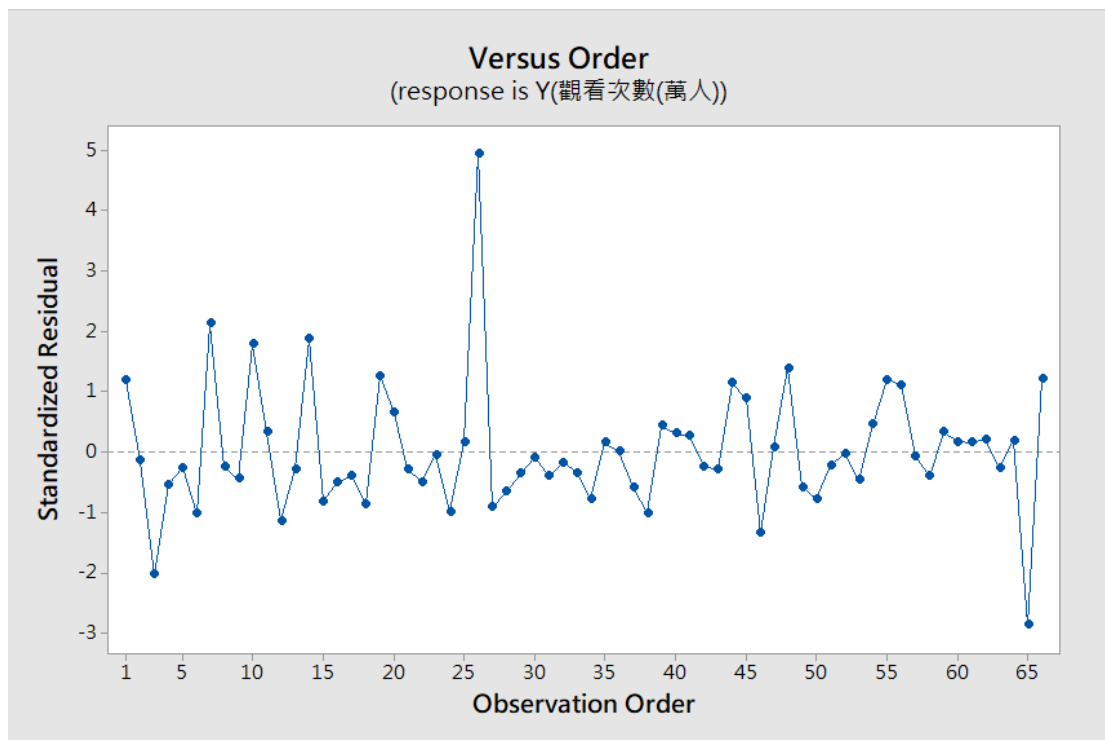
➤ 擬合值之長條圖(Histogram of FITS)



分析：

透過擬合值(fits)的長條圖中，我們可以觀察到擬合值(fits)並非完整呈現標準的常態的樣貌，也就是未明顯表現出鐘形的模樣，但由於資料整體上並沒有某一個擬合值(fits)特別偏離其他資料，因此藉由這張擬合值(fits)的長條圖所給的資訊，我們目前沒有打算將 Y 做取 \ln 等轉換的步驟。

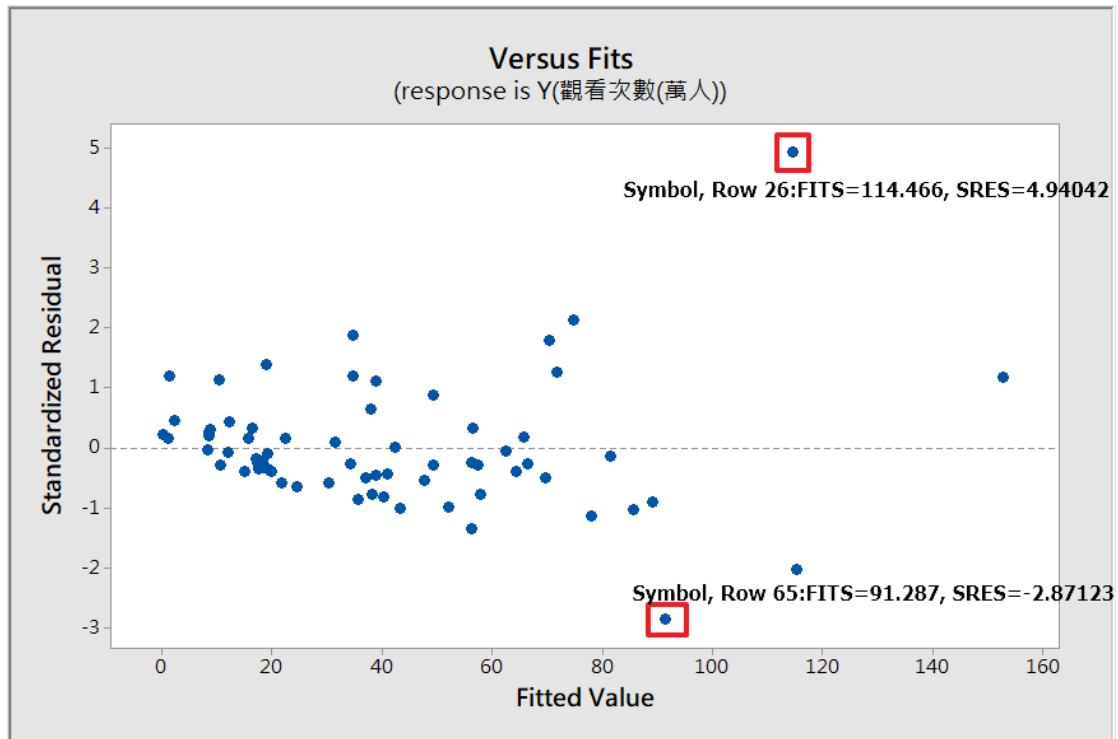
➤ 次序比較圖 (Versus Order)



分析：

透過此次序比較圖的呈現，我們可以發現數據是隨機(random)的，不會因為觀察的次序(order)而有趨勢性的逐漸上升或是下降。

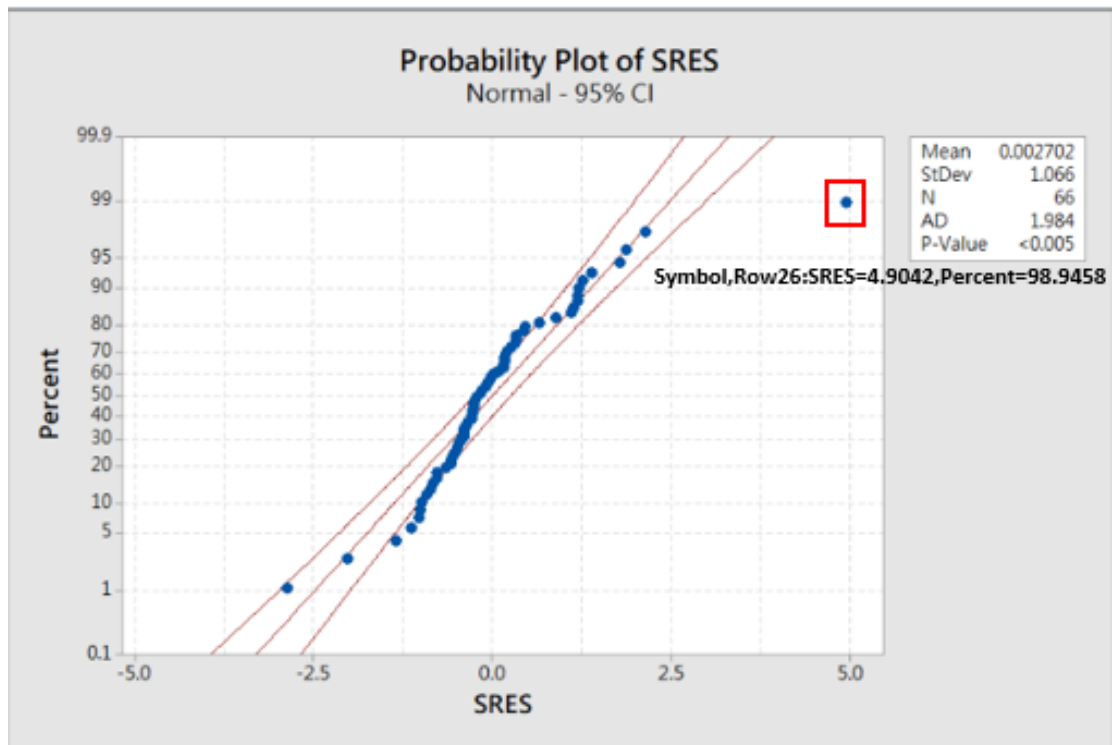
➤ 擬合值比較圖 (Versus Fits)



分析：

藉由此圖，我們發現有其中兩點，分別是第 26 筆資料和第 65 筆資料，與其他點相比之下較為偏離，且這兩筆資料的標準化殘差也落在 $(-2, 2)$ 的區域外，相對而言這兩筆數據的殘差較為極端，因此可做為往後判斷是否有離群值的依據。而這些點的分佈有向右開的趨勢 (right-opening magaphone)，可以得知標準化殘差的變異數並非常數 (nonconstant variance)，這表示其變異數可能會和應變數 (Y) 、單個或是多個自變數 (X_i) ，甚至是其他因素有關聯。

➤ 標準化殘差之機率圖(Probability Plot of SRES)



分析：

透過機率圖我們可以觀察到第 26 筆資料嚴重偏離其他的資料，且藉由 95% 的信賴區間帶可以看出資料並非被覆蓋在範圍之內，因此可推測第 26 筆資料有可能是離群值。此外 $P\text{-Value} < 0.005$ 可以知道我們所建立出來的模型其標準化殘差有很大的機率不屬於常態分配。

- 刪除離群值(outlier)之迴歸分析

➤ 刪除原因

第 26 筆資料為兩位年輕女生跳舞的頻道，我們所選的其餘 65 筆資料都沒有以跳舞為主軸的頻道，再加上該影片的拍攝主題為 BLACKPINK - 'Kill This Love' Dance cover，而 BLACKPINK 為當紅韓團，當時 Kill This Love 膾炙人口，可以發現這首歌的官方影片觀看次數高達 3 億，並且跳這首歌的影片觀看次數也都相當可觀，因此我們認為資料 26 是會影響我們迴歸模型的離群值，決定將它刪除以降低誤差。

另外我們還刪掉了第 65 筆資料，因為剛好選到那對夫妻的回顧影片，這支影片就是把之前他們上傳過的多支影片，加以剪接成一支新的影片，主要是在做重點回顧，在加上其他 65 支影片都不屬於回顧類型，因此我們也將它視為離群值刪除。

(已刪除資料的後方資料皆自動往前排序。)

➤ 迴歸分析

Regression Analysis: Y(觀看次數(萬人) versus ... 1(頻道類型:人物與日誌)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	42792	14264.1	19.83	0.000
x1(訂閱數)(萬)	1	12818	12818.4	17.82	0.000
x2(FB追蹤人數)(萬人)	1	3787	3786.9	5.26	0.025
x11(頻道類型:人物與日誌)	1	9886	9885.8	13.74	0.000
Error	60	43158	719.3		
Total	63	85951			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
26.8199	49.79%	47.28%	38.80%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-15.24	8.12	-1.88	0.065	
x1(訂閱數)(萬)	0.3687	0.0873	4.22	0.000	1.51
x2(FB追蹤人數)(萬人)	0.2243	0.0978	2.29	0.025	1.60
x11(頻道類型:人物與日誌)	29.17	7.87	3.71	0.000	1.07

Regression Equation

$$Y(\text{觀看次數(萬人)}) = -15.24 + 0.3687 x1(\text{訂閱數(萬)}) + 0.2243 x2(\text{FB追蹤人數(萬人)}) + 29.17 x11(\text{頻道類型:人物與日誌})$$

Fits and Diagnostics for Unusual Observations

Obs	Y(觀看次數(萬人))	Fit	Resid	Std Resid	R	X
1	186.59	136.60	49.99	2.25	R	X
3	49.63	99.45	-49.83	-2.03	R	
7	144.64	76.58	68.05	2.66	R	
10	129.24	71.50	57.74	2.24	R	
14	97.49	37.92	59.58	2.27	R	
19	113.48	58.00	55.48	2.18	R	
26	59.23	65.74	-6.51	-0.29		X

R Large residual

X Unusual X

分析：

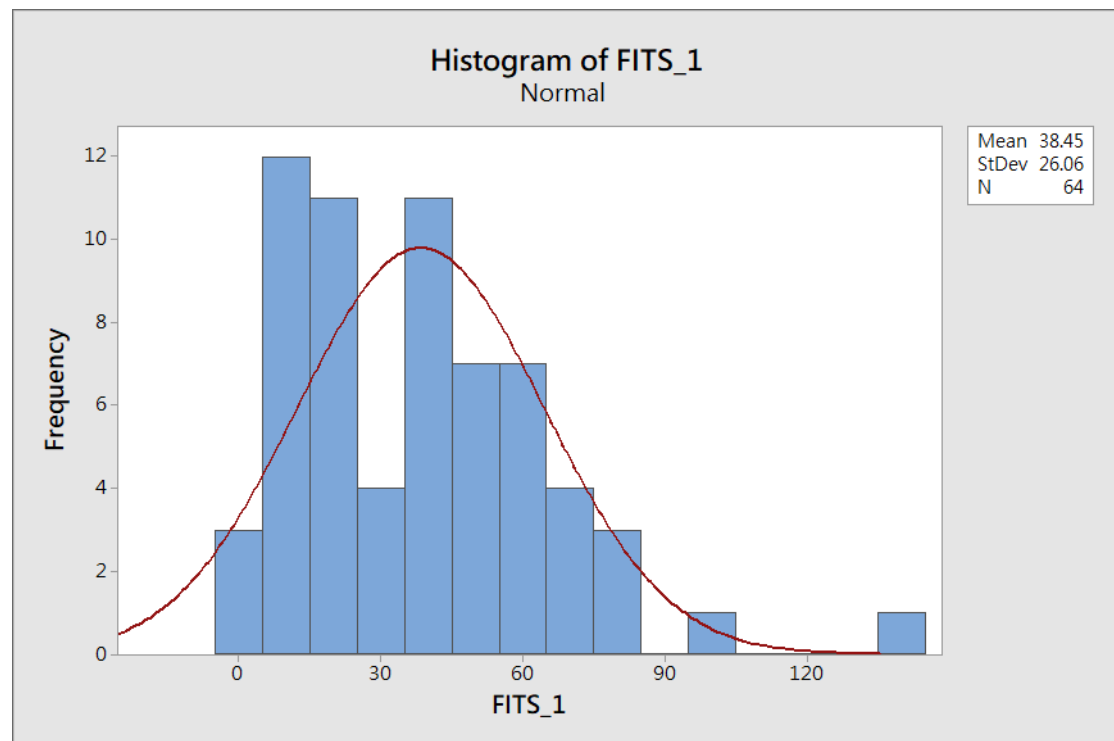
$$19.83 > F_{0.05}(3,60) = 2.75808$$

$$4.22、2.29、3.71 \text{ 皆} > T_{0.975}(60) = 2.00030$$

藉由 F、T 檢定及 P-Value 得知刪掉離群值後的模型依然是顯著的。

再來我們可以發現 R-sq(adj)從 43.31%提升至 47.28%。而我們的自變數依然保持沒有共線性的關係。雖然第 1、3、7、10、14、19 筆資料的標準化殘差較其他資料來的大，但我們認為是在合理範圍內，因次我們不打算將這六筆資料視為潛在的離群值。

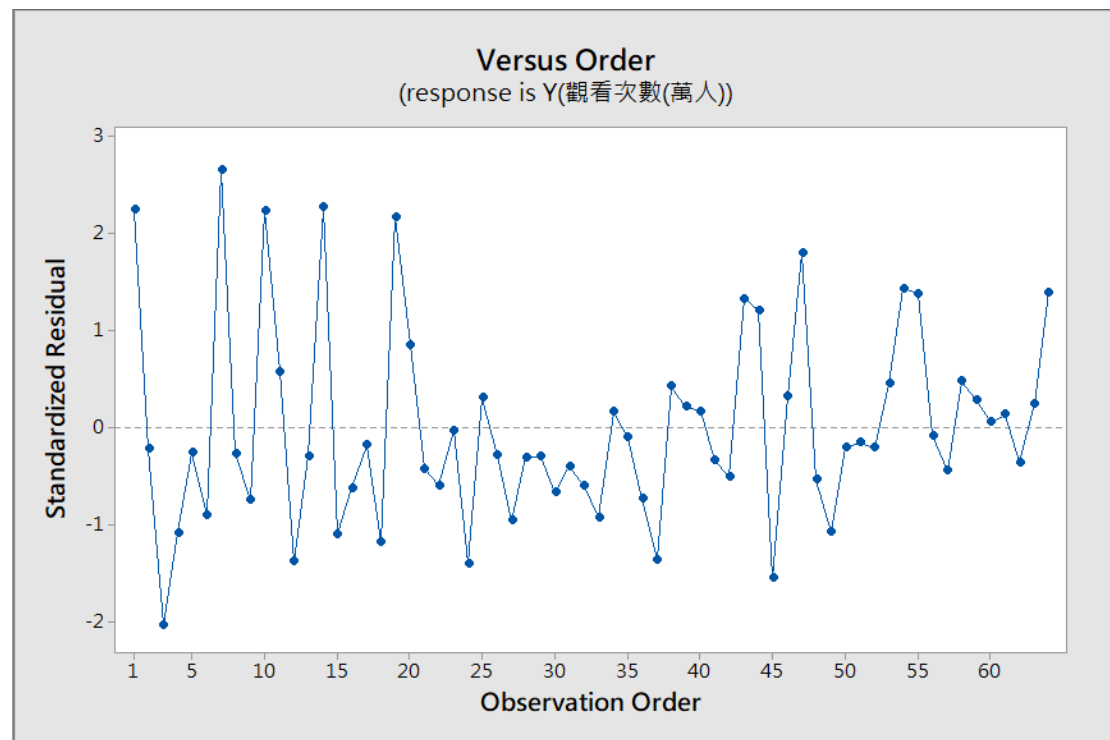
➤ 擬合值之長條圖(Histogram of FITS)



分析：

刪除離群值後，可以發現標準差從 30.97 降至 26.06，大約減少了 4 個單位，因此可得知資料的分散程度降低了。

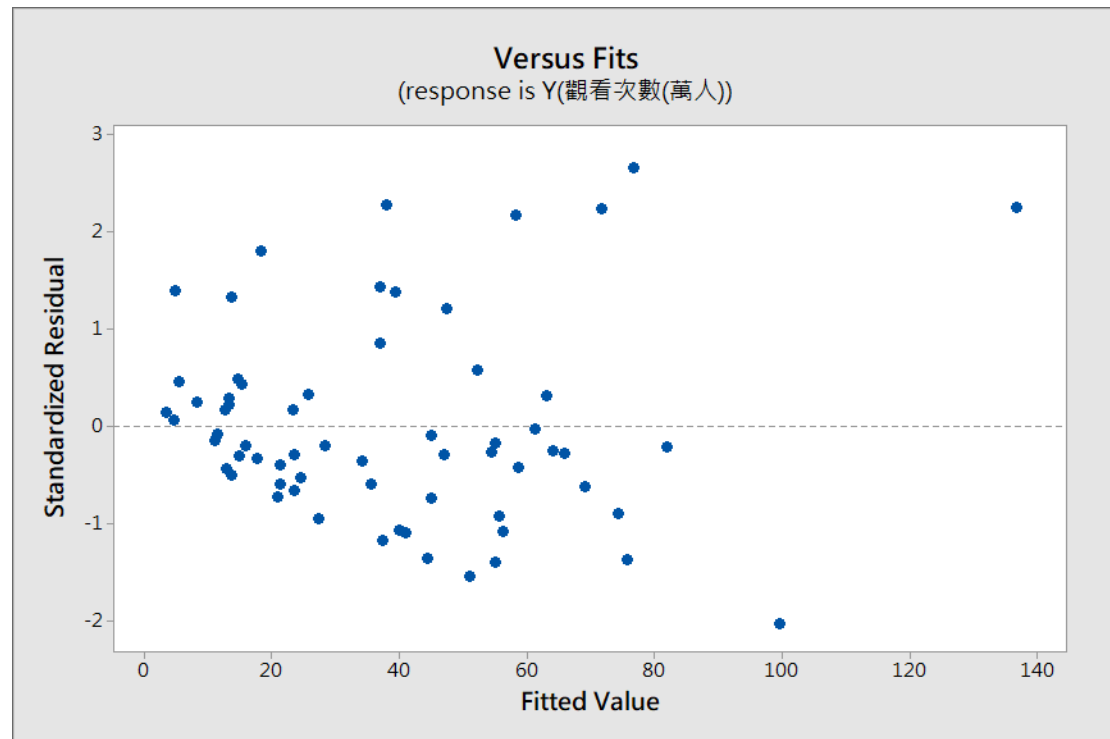
➤ 次序比較圖 (Versus Order)



分析：

刪掉離群值後，標準化殘差範圍縮小至 $(-2, 3)$ 左右，並且可觀察出此圖並沒有任何趨勢，因此我們得知我們的數據是隨機的，不會因為次序的增加而標準化殘差就隨之上升或下降。

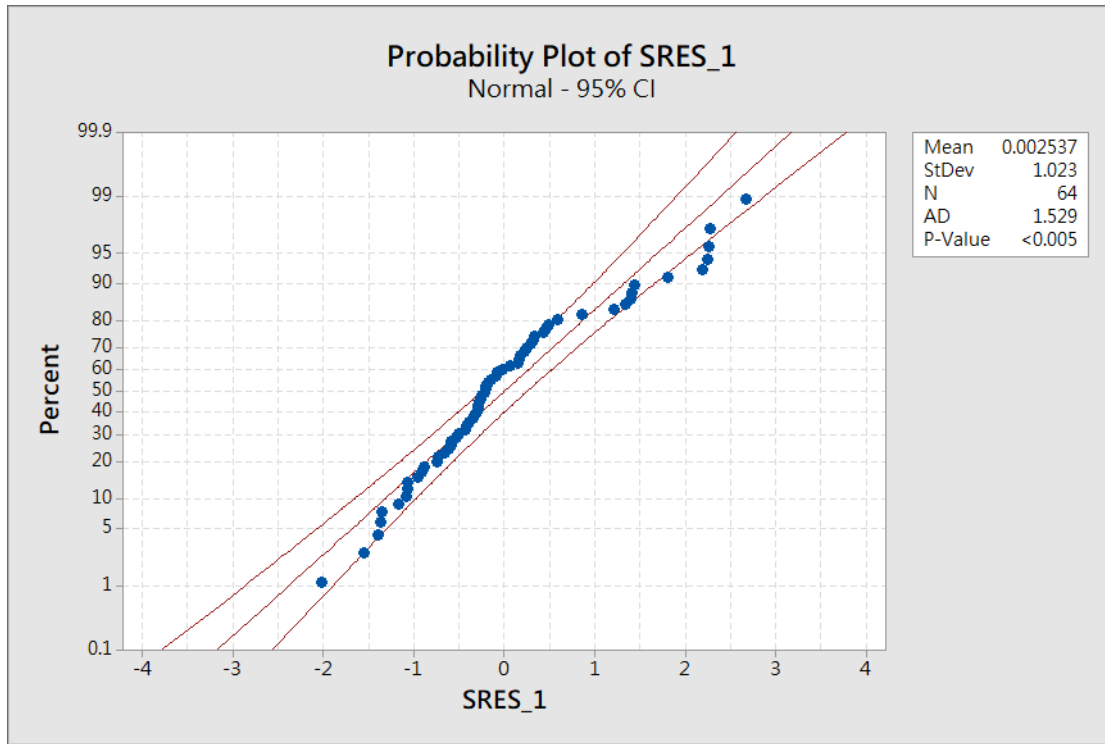
➤ 擬合值比較圖 (Versus Fits)



分析：

刪掉離群值後，標準化殘差範圍縮小至 $(-2, 3)$ 左右，但我們發現點的分佈依舊有右開的趨勢，可以得知標準化殘差的變異數並非常數，因此我們考慮做變數轉換以符合標準化殘差的變異數為常數的迴歸假設。

➤ 標準化殘差之機率圖(Probability Plot of SRES)



分析：

根據此圖的數據，刪掉離群值後的 P-Value 依舊小於 0.005，因此我們將 Y 做變數變換以提升 P-Value，藉此滿足標準化殘差屬於常態分配的迴歸假設。

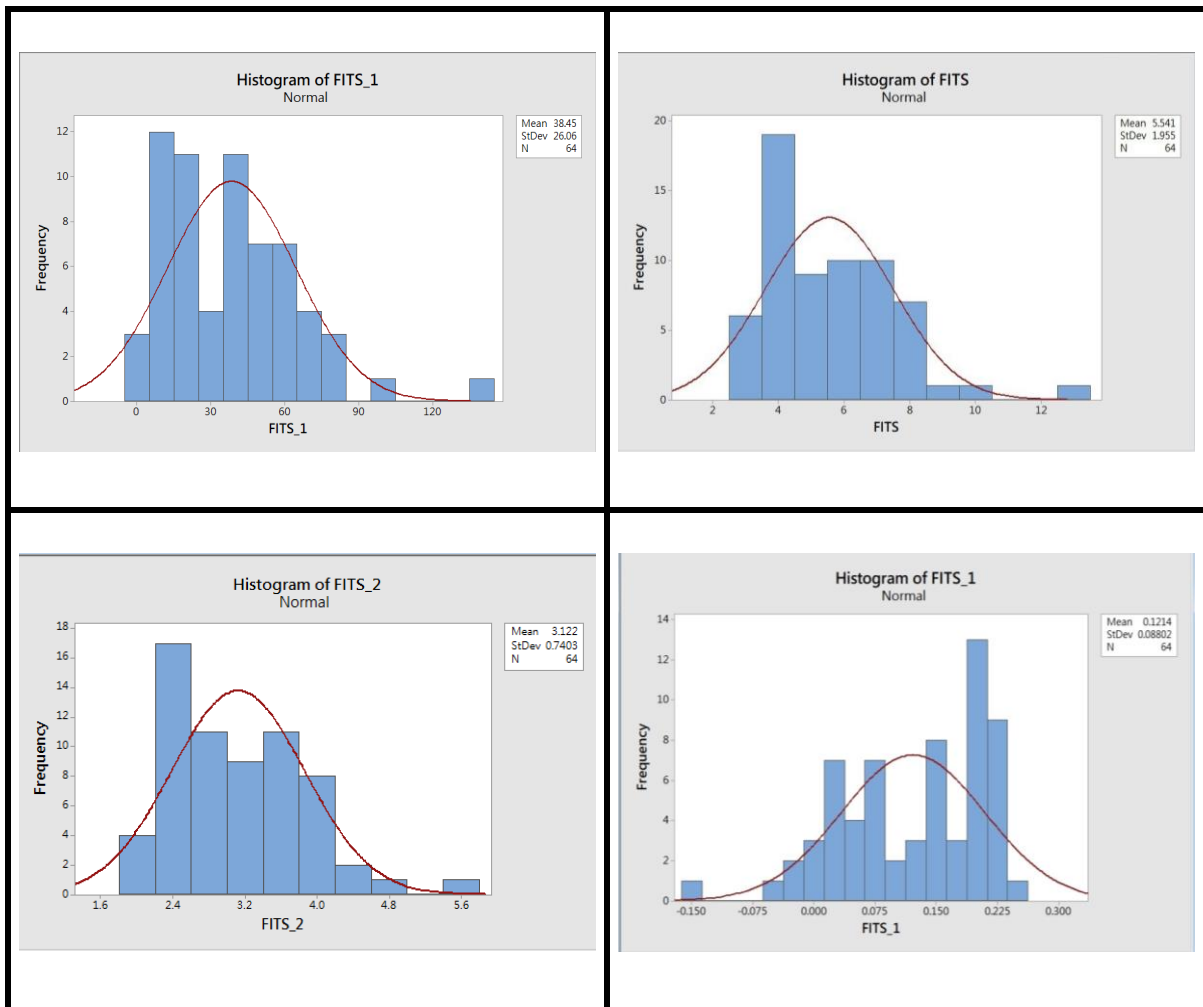
- 應變數之線性轉換

我們考慮三種常見的轉換方式：

分別是對 Y 開根號、對 Y 取 \ln 、取 Y 的倒數。

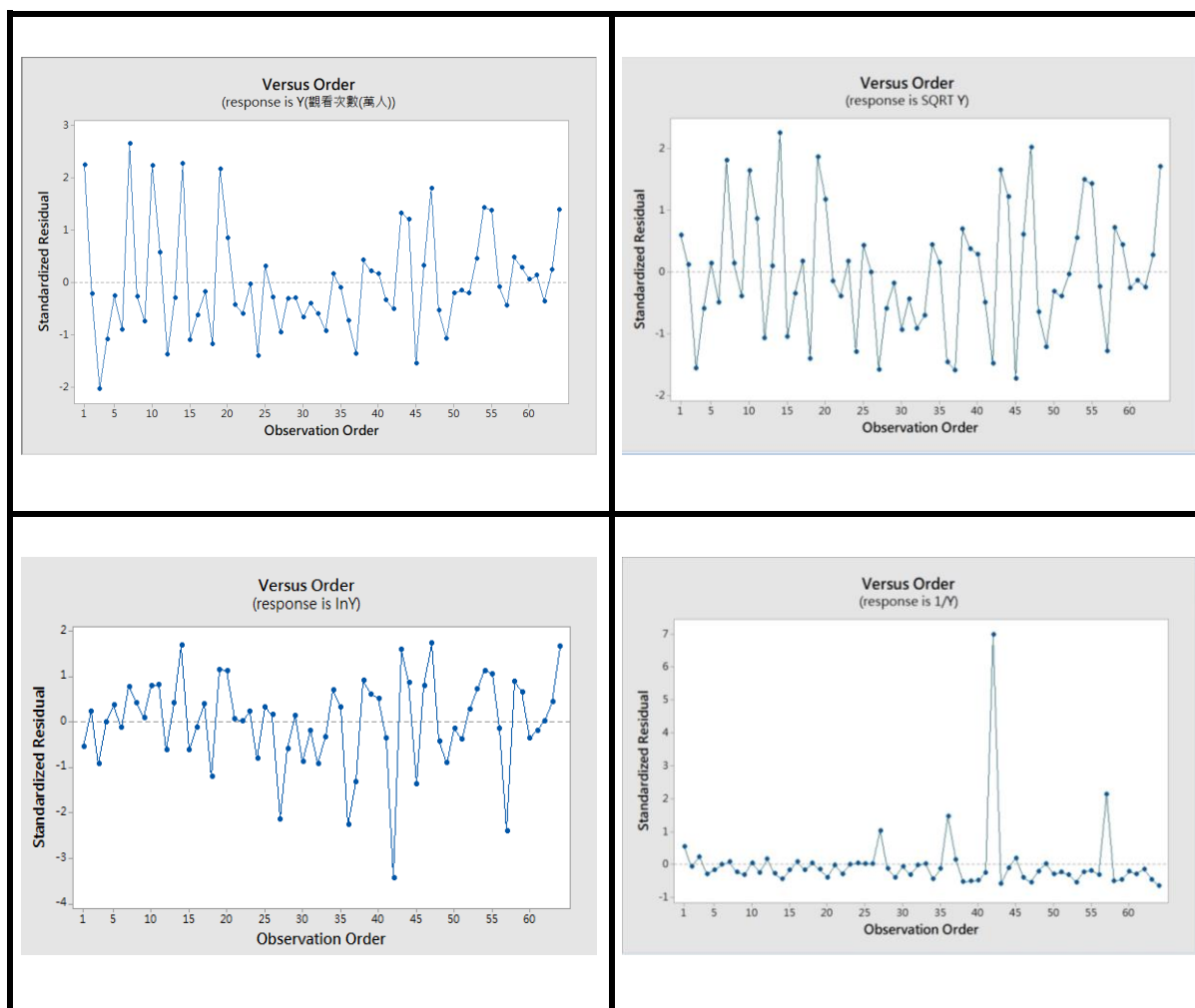
下列表格為 Y 與經過三種不同轉換後的 Y 的比較示意圖。

➤ 擬合值之長條圖(Histogram of FITS)



(左上： Y 、右上： \sqrt{Y} 、左下： $\ln Y$ 、右下： $\frac{1}{Y}$)

➤ 次序比較圖 (Versus Order)



(左上： Y 、右上： \sqrt{Y} 、左下： $\ln Y$ 、右下： $\frac{1}{Y}$)

分析：

從四張圖中皆顯示數據隨機的，不會因為 order 的增加而標準化後的殘差就隨之上升或下降。

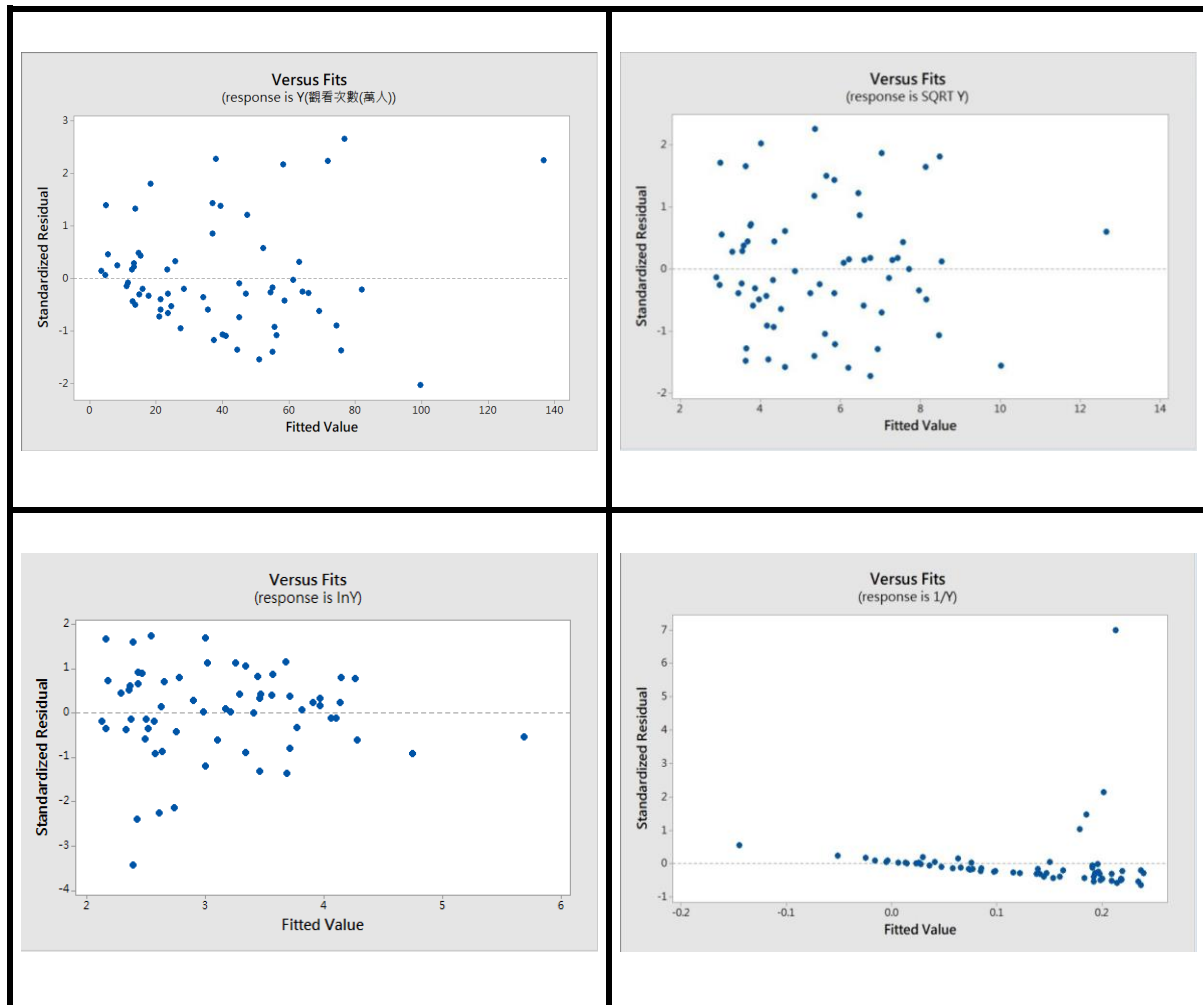
Y 的標準化殘差範圍大約為 $(-2, 3)$ ；

\sqrt{Y} 的標準化殘差範圍大約為 $(-2, 2)$ ；

$\ln Y$ 的標準化殘差範圍大約為 $(-4, 2)$ ；

$\frac{1}{Y}$ 的標準化殘差範圍大約為 $(-1, 7)$ 。

➤ 擬合值比較圖 (Versus Fits)



(左上： Y 、右上： \sqrt{Y} 、左下： $\ln Y$ 、右下： $\frac{1}{Y}$)

由上圖可觀察出，

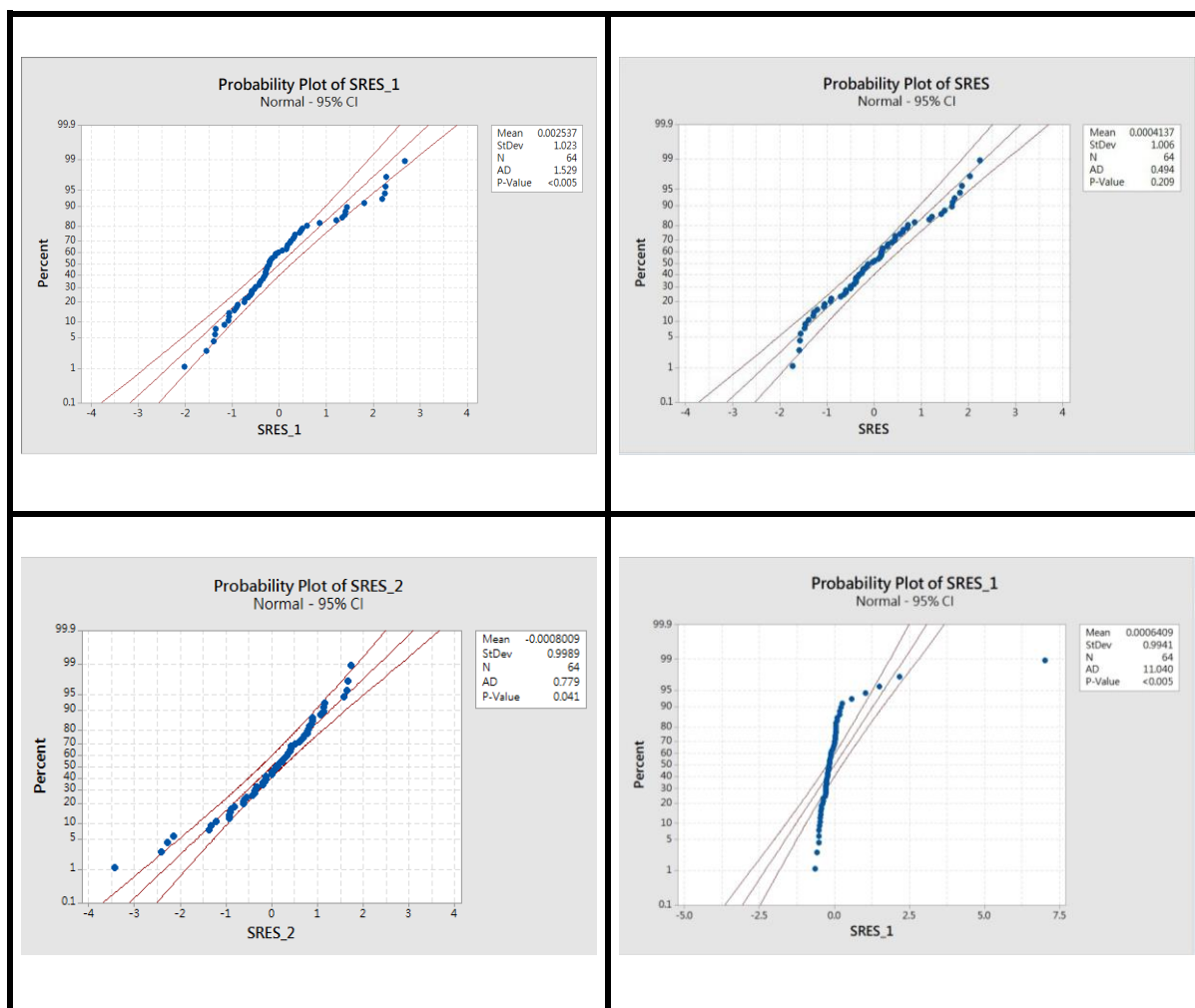
Y ：此圖形有右開的趨勢，可推測標準化殘差的變異數非常數

\sqrt{Y} ：圖形中的點像一盤散沙，可推測標準化殘差的變異數為常數

$\ln Y$ ：此圖形有左開的趨勢，可推測標準化殘差的變異數非常數

$\frac{1}{Y}$ ：經轉換後並未改善原本點的分佈

➤ 標準化殘差之機率圖(Probability Plot of SRES)



(左上： Y 、右上： \sqrt{Y} 、左下： $\ln Y$ 、右下： $\frac{1}{Y}$)

我們可發現 \sqrt{Y} 機率圖中的 P-Value=0.209 相較於其他三個的 p-value 都來的大，也就是說 \sqrt{Y} 的模型改善了原本未符合迴歸假設(標準化殘差屬於常態分配)的情況。

變數	Y	\sqrt{Y}	$\ln Y$	$\frac{1}{Y}$
P-Value	<0.005	0.209	0.041	<0.005

➤ 轉化後的迴歸分析

Y 的迴歸分析

Regression Analysis: Y(觀看次數(萬人) versus ... 1(頻道類型:人物與日誌)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	42792	14264.1	19.83	0.000
x1(訂閱數)(萬)	1	12818	12818.4	17.82	0.000
x2(FB追蹤人數)(萬人)	1	3787	3786.9	5.26	0.025
x11(頻道類型:人物與日誌)	1	9886	9885.8	13.74	0.000
Error	60	43158	719.3		
Total	63	85951			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
26.8199	49.79%	47.28%	38.80%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-15.24	8.12	-1.88	0.065	
x1(訂閱數)(萬)	0.3687	0.0873	4.22	0.000	1.51
x2(FB追蹤人數)(萬人)	0.2243	0.0978	2.29	0.025	1.60
x11(頻道類型:人物與日誌)	29.17	7.87	3.71	0.000	1.07

Regression Equation

$$Y(\text{觀看次數(萬人)}) = -15.24 + 0.3687 x1(\text{訂閱數(萬)}) + 0.2243 x2(\text{FB追蹤人數(萬人)}) + 29.17 x11(\text{頻道類型:人物與日誌})$$

Fits and Diagnostics for Unusual Observations

Obs	Y(觀看次數(萬人))	Fit	Resid	Std Resid	
1	186.59	136.60	49.99	2.25	R X
3	49.63	99.45	-49.83	-2.03	R
7	144.64	76.58	68.05	2.66	R
10	129.24	71.50	57.74	2.24	R
14	97.49	37.92	59.58	2.27	R
19	113.48	58.00	55.48	2.18	R
26	59.23	65.74	-6.51	-0.29	X

R Large residual

X Unusual X

\sqrt{Y} 的迴歸分析

Regression Analysis: SQRT Y versus x1(訂閱數)(萬), ... 類型:人物與日誌 Menu ▾

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	240.86	80.285	18.89	0.000
x1(訂閱數)(萬)	1	62.02	62.017	14.59	0.000
x2(FB追蹤人數)(萬人)	1	24.59	24.593	5.79	0.019
x11(頻道類型:人物與日誌)	1	69.70	69.696	16.40	0.000
Error	60	255.04	4.251		
Total	63	495.89			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2.06171	48.57%	46.00%	41.74%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.587	0.624	2.54	0.014	
x1(訂閱數)(萬)	0.02565	0.00671	3.82	0.000	1.51
x2(FB追蹤人數)(萬人)	0.01808	0.00752	2.41	0.019	1.60
x11(頻道類型:人物與日誌)	2.449	0.605	4.05	0.000	1.07

Regression Equation

$$\text{SQRT Y} = 1.587 + 0.02565 \times \text{x1(訂閱數)(萬)} + 0.01808 \times \text{x2(FB追蹤人數)(萬人)} + 2.449 \times \text{x11(頻道類型:人物與日誌)}$$

Fits and Diagnostics for Unusual Observations

Obs	SQRT Y	Fit	Resid	Std Resid	
1	13.660	12.638	1.022	0.60	X
14	9.874	5.350	4.524	2.25	R
26	7.696	7.694	0.003	0.00	X
47	8.109	3.996	4.113	2.02	R

R Large residual
X Unusual X

Residuals vs Fits for SQRT Y

Residuals vs Order for SQRT Y

透過 P-Value，我們得知 \sqrt{Y} 的模型是顯著的，而且可以從 VIF 觀察到變數間依舊沒有共線性關係。

lnY 的迴歸分析

Regression Analysis: lnY versus x1(訂閱數)(萬), ... 頻道類型:人物與日誌)

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	34.526	11.5085	12.13	0.000
x1(訂閱數)(萬)	1	7.457	7.4565	7.86	0.007
x2(FB追蹤人數)(萬人)	1	3.919	3.9194	4.13	0.046
x11(頻道類型:人物與日誌)	1	12.366	12.3655	13.04	0.001
Error	60	56.906	0.9484		
Total	63	91.432			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.973875	37.76%	34.65%	31.60%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	1.661	0.295	5.64	0.000	
x1(訂閱數)(萬)	0.00889	0.00317	2.80	0.007	1.51
x2(FB追蹤人數)(萬人)	0.00722	0.00355	2.03	0.046	1.60
x11(頻道類型:人物與日誌)	1.032	0.286	3.61	0.001	1.07

Regression Equation

$$\ln Y = 1.661 + 0.00889 x1(\text{訂閱數})(\text{萬}) + 0.00722 x2(\text{FB追蹤人數})(\text{萬人}) + 1.032 x11(\text{頻道類型:人物與日誌})$$

Fits and Diagnostics for Unusual Observations

Obs	lnY	Fit	Resid	Std Resid	
1	5.229	5.681	-0.452	-0.56	X
26	4.081	3.960	0.121	0.15	X
27	0.670	2.727	-2.056	-2.15	R
36	0.409	2.601	-2.192	-2.28	R
42	-0.908	2.381	-3.289	-3.44	R
57	0.114	2.409	-2.296	-2.41	R

R Large residual

X Unusual X

$\frac{1}{Y}$ 的迴歸分析

Regression Analysis: 1/Y versus x1(訂閱數)(萬), ... (頻道類型:人物與日誌) Menu ▾

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	0.48805	0.16268	1.49	0.225
x1(訂閱數)(萬)	1	0.06129	0.06129	0.56	0.456
x2(FB追蹤人數)(萬人)	1	0.07085	0.07085	0.65	0.423
x11(頻道類型:人物與日誌)	1	0.25002	0.25002	2.30	0.135
Error	60	6.52944	0.10882		
Total	63	7.01749			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.329885	6.95%	2.30%	0.00%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	0.2817	0.0999	2.82	0.006	
x1(訂閱數)(萬)	-0.00081	0.00107	-0.75	0.456	1.51
x2(FB追蹤人數)(萬人)	-0.00097	0.00120	-0.81	0.423	1.60
x11(頻道類型:人物與日誌)	-0.1467	0.0968	-1.52	0.135	1.07

Regression Equation

$$1/Y = 0.2817 - 0.00081 \text{ x1(訂閱數)(萬)} - 0.00097 \text{ x2(FB追蹤人數)(萬人)} - 0.1467 \text{ x11(頻道類型:人物與日誌)}$$

Fits and Diagnostics for Unusual Observations

Obs	1/Y	Fit	Resid	Std Resid	
1	0.005	-0.145	0.151	0.55	X
26	0.017	0.013	0.004	0.01	X
42	2.480	0.213	2.267	6.99	R
57	0.893	0.201	0.692	2.14	R

R Large residual
X Unusual X

Residuals vs Fits for 1/Y

Residuals vs Order for 1/Y

最終模型與解釋

藉由上述的比較，我們發現 \sqrt{Y} 的表現不管是在四種圖形的比較上還是迴歸分析，皆優於其他兩種轉換，且做 \sqrt{Y} 此轉換後，修正了原本未符合假設的情況，因此我們最終選擇的模型為

$$\sqrt{Y} = 1.587 + 0.02565X_1 + 0.01808X_2 + 2.449X_{11}$$

【模型解釋】

$$b_0 = 1.587 \quad b_1 = 0.02565 \quad b_2 = 0.01808 \quad b_{11} = 2.449$$

① 觀察訂閱人數(X_1)增加 1 單位，影片觀看次數(Y)會增加多少：

固定 facebook 追蹤人數(X_2) 和 頻道類型：人物與日誌(X_{11}) 的值

$$\sqrt{Y} = b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11}$$

$$\Rightarrow Y = (b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11})^2$$

設 Y' 為 X_1 代 $(X_1 + 1)$ 所得結果

$$Y' = [b_0 + b_1(X_1 + 1) + b_2X_2 + b_{11}X_{11}]^2$$

$$\Rightarrow Y' = [b_1 + (b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11})]^2 = (b_1 + \sqrt{Y})^2$$

∴當 X_1 增加 1 單位時 Y 的變化量為

$$\begin{aligned} Y' - Y &= (b_1 + \sqrt{Y})^2 - Y = b_1^2 + 2b_1\sqrt{Y} + Y - Y = b_1^2 + 2b_1\sqrt{Y} \\ &= b_1^2 + 2b_1(b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11}) \end{aligned}$$

範例：

當 facebook 追蹤人數(X_2)為 50 萬人(取 X_2 的平均數)

頻道類型不為人物與日誌($X_{11} = 0$)

若 訂閱人數(X_1)為 290 萬人(取 X_1 的最大值)，則訂閱人數(X_1)增加 1 萬人時，影片觀看次數(Y)會增加約 5100 次。

<Sol.>

$$Y' - Y$$

$$= b_1^2 + 2b_1(b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11})$$

$$= 0.02565^2 + 2 \times 0.02565 \times (1.587 + 0.02565 \times 290 + 0.01808 \times 50 + 2.449 \times 0)$$

$$\approx 0.5100(\text{萬次})$$

若 訂閱人數(X_1)為 47 萬人(取 X_1 的最小值)，則訂閱人數(X_1)增加 1 萬人時，影片觀看次數(Y)會增加約 1903 次。

<Sol.> $Y' - Y$

$$= b_1^2 + 2b_1(b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11})$$

$$= 0.02565^2 + 2 \times 0.02565 \times (1.587 + 0.02565 \times 47 + 0.01808 \times 50 + 2.449 \times 0)$$

$$\approx 0.1903(\text{萬次})$$

② 觀察 facebook 追蹤人數(X_2)增加 1 單位，影片觀看次數(Y)會增加多少：

固定 訂閱人數(X_1) 和 頻道類型：人物與日誌(X_{11}) 的值

$$\sqrt{Y} = b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11}$$

$$\Rightarrow Y = (b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11})^2$$

設 Y' 為 X_2 代 $(X_2 + 1)$ 所得結果

$$Y' = [b_0 + b_1X_1 + b_2(X_2 + 1) + b_{11}X_{11}]^2$$

$$\Rightarrow Y' = [b_2 + (b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11})]^2 = (b_2 + \sqrt{Y})^2$$

\therefore 當 X_2 增加 1 單位時 Y 的變化量為

$$\begin{aligned} Y' - Y &= (b_2 + \sqrt{Y})^2 - Y = b_2^2 + 2b_2\sqrt{Y} + Y - Y = b_2^2 + 2b_2\sqrt{Y} \\ &= b_2^2 + 2b_2(b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11}) \end{aligned}$$

範例：

當 訂閱人數(X_1)為 97 萬人(取 X_1 的平均數)

頻道類型不為人物與日誌($X_{11} = 0$)

若 facebook 追蹤人數(X_2)為 257 萬人(取 X_2 的最大值)，則追蹤人數(X_2)增加 1 萬人時，影片觀看次數(Y)會增加約 3157 次。

<Sol.>

$$Y' - Y$$

$$= b_2^2 + 2b_2(b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11})$$

$$= 0.01808^2 + 2 \times 0.01808 \times (1.587 + 0.02565 \times 97 + 0.01808 \times 257 + 2.449 \times 0)$$

$$\approx 0.3157(\text{萬次})$$

若 facebook 追蹤人數(X_2)為 2 萬人(取 X_2 的最小值)，則追蹤人數(X_2)

增加 1 萬人時，影片觀看次數(Y)會增加約 1490 次。

<Sol.> $Y' - Y$

$$= b_2^2 + 2b_2(b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11})$$

$$= 0.01808^2 + 2 \times 0.01808 \times (1.587 + 0.02565 \times 97 + 0.01808 \times 2 + 2.449 \times 0)$$

$$\approx 0.1490(\text{萬次})$$

③ 頻道類型是否為人物與日誌(X_{11})對影片觀看次數(Y)的影響：

$$\sqrt{Y} = b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11}$$

$$\Rightarrow Y = (b_0 + b_1X_1 + b_2X_2 + b_{11}X_{11})^2$$

設 Y_1 為 $X_{11} = 0$ 所得結果， $Y_1 = (b_0 + b_1X_1 + b_2X_2)^2$

設 Y_2 為 $X_{11} = 1$ 所得結果，則

$$Y_2 = (b_0 + b_1X_1 + b_2X_2 + b_{11})^2 = [b_{11} + (b_0 + b_1X_1 + b_2X_2)]^2$$

$$= (b_{11} + \sqrt{Y_1})^2 = b_{11}^2 + 2b_{11}\sqrt{Y_1} + Y_1$$

∴頻道類型是「人物與日誌($X_{11} = 1$)」較「非人物與日誌($X_{11} = 0$)」

的影片觀看次數(Y)增加量為

$$\begin{aligned} Y_2 - Y_1 &= (b_{11}^2 + 2b_{11}\sqrt{Y_1} + Y_1) - Y_1 = b_{11}^2 + 2b_{11}\sqrt{Y_1} \\ &= b_{11}^2 + 2b_{11}(b_0 + b_1X_1 + b_2X_2) \end{aligned}$$

範例：

當 訂閱人數(X_1)為 97 萬人(取 X_1 的平均數)

facebook 追蹤人數(X_2)為 50 萬人(取 X_2 的平均數)

則 頻道類型為人物與日誌($X_{11} = 1$) 相較 頻道類型非人物與日誌

($X_{11} = 0$)影片觀看次數(Y)增加約 303850 次。

$$\begin{aligned} \langle \text{Sol.} \rangle Y_2 - Y_1 &= b_{11}^2 + 2b_{11}(b_0 + b_1X_1 + b_2X_2) \\ &= 2.449^2 + 2 \times 2.449 \times (1.587 + 0.02565 \times 97 + 0.01808 \times 50) \\ &\approx 30.3850(\text{萬次}) \end{aligned}$$

- 結論

從迴歸分析結果，我們可以得知 Youtuber 的頻道訂閱數 (X_1)、Youtuber 的 Facebook 追蹤人數 (X_2) 以及 頻道類型：人物與日誌 (X_{11})，可以解釋 46% Youtube 影片觀看次數 (Y) 的變化量。因此，想要有較高的觀看次數，就需要較高的頻道訂閱數和 Facebook 追蹤人數，且拍攝人物與日誌類型的影片，相較於其他類型的影片會擁有較高的觀看次數。可惜的是我們的迴歸模型只能解釋不到一半的變化量，代表還有其他我們尚未探討到的因素會影響影片的觀看次數。

心得感想

【洪羽柔】

透過這次的報告，增加了我對統計的了解與熱情，並且接觸了從未觸碰過的 MINITAB，更加明白迴歸分析要如何運用在生活實際應用裡，雖然一開始光是確定主題就遇到了很大的問題，而中間過程中也發生許多狀況，但在與組員們和老師的多次討論後，終於完成這次的報告，也許成果看似簡陋或是操作上輕而易舉，但其實都花了大家許多精力與時間，不過最重要的是我們都在這個過程中一起學習到團隊合作、溝通協調、問題處理能力、當然還有統計相關的背景知識，謝謝老師以及組員們的指導與合作。

【艾品璇】

在這次的分組報告有很不一樣的經驗，學習到許多事，例如透過這次報告，知道如何操作 MINITAB，也藉由 MINITAB 加深了課堂上所學習到的知識，雖然過程並不是那麼的順利，從一開始找主題就四處碰壁，中間在操作上也發生一些錯誤，導致分析結果是錯的而重新來過，但經過老師耐心的講解，這些問題最後都解決了。也因為這次報告發現每個人都有各自擅長的地方，有的擅於美化 ppt，有的善於修飾文字，很開心能和他們一起完成這次的報告。

【陳晏琦】

透過這次的專題報告，讓我學習到統計相關的資料分析，從一開始找主題、蒐集相關的資料，到使用 MINITAB 進一步的數據分析，從中的每一個過程，我一邊學習新的內容，一邊實際套用在我們的主題上。而 Youtube 觀看次數這個主題，對於我而言是相當生活化的，因為平時我也是看 Youtube 居多，因此在做相關分析時會更有興趣，也因此領悟到什麼因素會去影響觀看次數。雖然最後做出來的模型在 R^2 的表現上不是非常突出，但在做報告的過程中，對於一些統計的圖表、專有名詞都更加了解，因為未來我也是要朝統計這方面繼續鑽研，因此這樣的學習經驗是我覺得收穫最多的地方。

【顏均翰】

第一次作統計相關的實驗報告，而在這報告過程中，發現原來要做出一個漂亮的迴歸，是一件不容易的事，但利用 MINITAB 軟體讓我們能夠去找出一個比較合適的迴歸，而我們也是初次學習 MINITAB，很慶幸我的組員們！真得很厲害透過合作，一起思考，完成了這一份報告，雖然過程中有些不順遂，但透過詢問老師也一步一步帶我們慢慢解危，因此在這份期末報告學得了一些基本 MINITAB 的操作並能簡單分析相關內容，並也深深感謝我的組員們，辛苦了！

資料來源

<https://www.youtube.com/>

<https://www.facebook.com/>

<https://zh.wikipedia.org/wiki/%E5%8F%B0%E7%81%A3YouTube%E9%A0%BB%E9%81%93%E8%A8%82%E9%96%B1%E4%BA%BA%E6%95%B8%E6%8E%92%E8%A1%8C%E6%A6%9C>

<https://www.youtube.com/watch?v=brTYxkUYIT0>

<http://netstat.stat.tku.edu.tw/prob.php>