

資料集介紹



- 資料來源：UCI Machine Learning Repository
- 資料集名稱：Computer Hardware Data Set
- 資料集介紹：
Relative CPU Performance Data, described in terms of its cycle time, memory size, etc.
- 資料筆數：209筆

變數介紹

1. vendor name: 30

(adviser, amdahl,apollo, basf, bti, burroughs, c.r.d, cambex, cdc, dec, dg, formation, four-phase, gould, honeywell, hp, ibm, ipl, magnuson, microdata, nas, ncr, nixdorf, perkin-elmer, prime, siemens, sperry, sratus, wang)

2. Model Name: many unique symbols

3. MYCT: machine cycle time in nanoseconds (integer)

4. MMIN: minimum main memory in kilobytes (integer)

5. MMAX: maximum main memory in kilobytes (integer)

6. CACH: cache memory in kilobytes (integer)

7. CHMIN: minimum channels in units (integer)

8. CHMAX: maximum channels in units (integer)

9. PRP: published relative performance (integer)

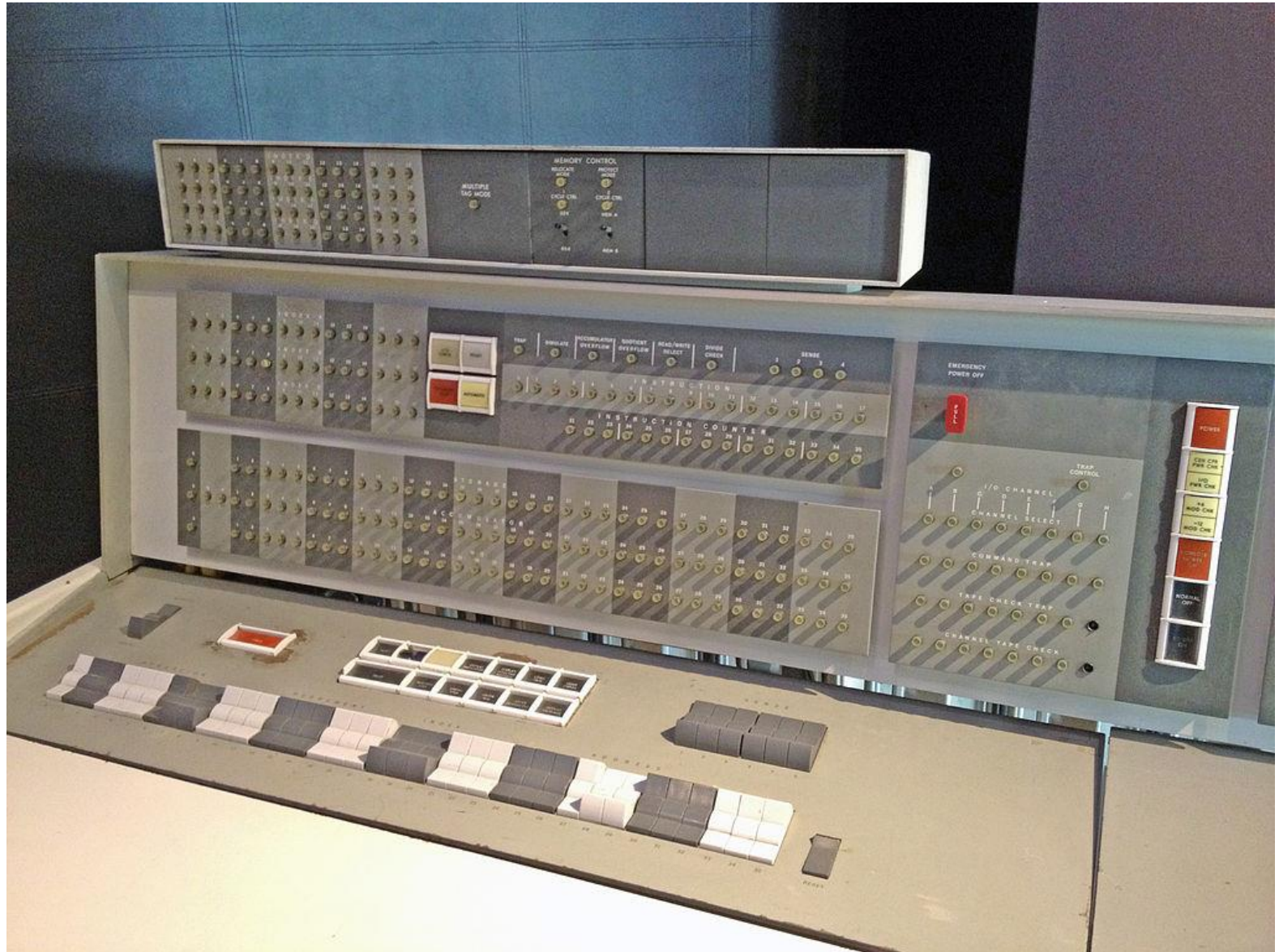
10. ERP: estimated relative performance from the original article (integer)

3-7部分為在本次實驗中所採用的變數。

Dataset in SPSS

91	honeywell	dps:8/52	140	2000	32000	32	1	54	141	181
92	honeywell	dps:8/62	140	2000	32000	32	1	54	189	181
93	honeywell	dps:8/20	140	2000	4000	8	1	20	22	32
94	ibm	3033:s	57	4000	16000	1	6	12	132	82
95	ibm	3033:u	57	4000	24000	64	12	16	237	171
96	ibm	3081	26	16000	32000	64	16	24	465	361
97	ibm	3081:d	26	16000	32000	64	8	24	465	350
98	ibm	3083:b	26	8000	32000	0	8	24	277	220

IBM 3033



逐步迴歸介紹

- 逐步迴歸是多元迴歸分析中挑選自變項加入迴歸方程式的一種方法。
- 逐步迴歸是以 t 值（及其顯著水準 α 值）作為決定是否挑選某個自變項的參考指標，如果某個自變項迴歸係數的 $|t|$ 考驗值大於查表的理論 t 值（或說其值已達 α 的顯著水準）時，電腦便自動決定要挑選該自變項進入迴歸方程式中，若未達顯著水準的標準時，則電腦會自動排除該自變項於迴歸方程式之外。

SPSS 設定

式(U) 視窗(W) 說明 線性迴歸

線性迴歸：統計資料

迴歸係數

- 估計值(E)
- 信賴區間(N)
層次 (%) : 95
- 共變異數矩陣(V)
- 模型適合度(M)
- R 平方改變量(S)
- 描述性統計資料(D)
- 部分與偏相關(P)
- 共線性診斷(L)

殘差

- Durbin-Watson
- 全部觀察值診斷(C)
● 範圍外的偏離值(O) : 3 標準差
● 全部觀察值(A)

繼續 取消 說明

因變數(D) : ERP

區塊(B) 1 / 1

上一個(V) 下一個(N)

自變數(I) : MYCT, MMIN, MMAX

方法(M) : 逐步迴歸法

選擇變數(E) : 規則(U)...

觀察值標籤(C) :

加權最小平方法之權數(H) :

確定 貼上(P) 重設(R) 取消 說明

統計資料(S)...
圖形(T)...
儲存(S)...
選項(O)...
樣式(L)...
重複取樣(B)...

32000
32000
32000

變數描述性統計資料

描述性統計資料			
	平均數	標準偏差	N
ERP	99.33	154.757	209
MYCT	203.82	260.263	209
MMIN	2867.98	3878.743	209
MMAX	11796.15	11726.564	209
CACH	25.21	40.629	209
CHMIN	4.70	6.816	209
CHMAX	18.27	25.997	209
PRP	105.62	160.831	209

本表可以觀察所有變數之平均值及標準差，N表現其資料筆數為209筆。

迴歸基本假設

1. 線性關係 - 依變數和自變數之間的關係必須是線性。
2. 常態性(normality) - 若是資料呈現常態分配(normal distribution)，則誤差項也會呈現同樣的分配。
3. 誤差項的獨立性 - 自變數的誤差項，相互之間應該是獨立的。
4. 誤差項的變異數相等(Homoscedasticity) - 自變數的誤差項除了需要呈現常態性分配外，其變量數也需要相等。

ADJ-R² & Durbin-Watson

模型摘要 ^f										
模型	R	R 平方	調整後 R 平方	標準偏斜度錯誤	變更統計資料					Durbin-Watson
					R 平方變更	F 值變更	df1	df2	顯著性 F 值變更	
1	.966 ^a	.934	.934	39.833	.934	2932.573	1	207	.000	
2	.976 ^b	.952	.951	34.167	.018	75.355	1	206	.000	
3	.977 ^c	.955	.954	33.186	.003	13.355	1	205	.000	
4	.978 ^d	.957	.956	32.426	.002	10.731	1	204	.001	
5	.979 ^e	.959	.958	31.756	.002	9.687	1	203	.002	1.390

1. 調整後R平方，觀察ADJ-R²之用意為加入該變數後對模型依變數的預測性是否增加。
2. 該表中可輕易觀察到模型5得到最佳的調整後R²。
3. 模型顯著性亦小於0.05，Reject H0，模型獲得良好配適。
H0:Model fit not well
H1:Model fit well
4. Durbin-Watson統計量可以用來檢測回歸分析中的殘差項是否存在自相關，愈接近2表愈不存在自我相關性。

ANOVA 分析

變異數分析 ^a						
模型		平方和	df	平均值平方	F	顯著性
1	迴歸	4653103.994	1	4653103.994	2932.573	.000 ^b
	殘差	328446.226	207	1586.697		
	總計	4981550.220	208			
2	迴歸	4741070.993	2	2370535.497	2030.655	.000 ^c
	殘差	240479.227	206	1167.375		
	總計	4981550.220	208			
3	迴歸	4755779.158	3	1585259.719	1439.415	.000 ^d
	殘差	225771.062	205	1101.322		
	總計	4981550.220	208			
4	迴歸	4767061.478	4	1191765.369	1133.487	.000 ^e
	殘差	214488.742	204	1051.415		
	總計	4981550.220	208			
5	迴歸	4776830.116	5	955366.023	947.339	.000 ^f
	殘差	204720.104	203	1008.473		
	總計	4981550.220	208			

本表顯示在使用逐步回歸法過程中所產生的每個模型，皆呈現顯著性小於0.05，指模型皆獲得良好的配適。

係數

係數^a

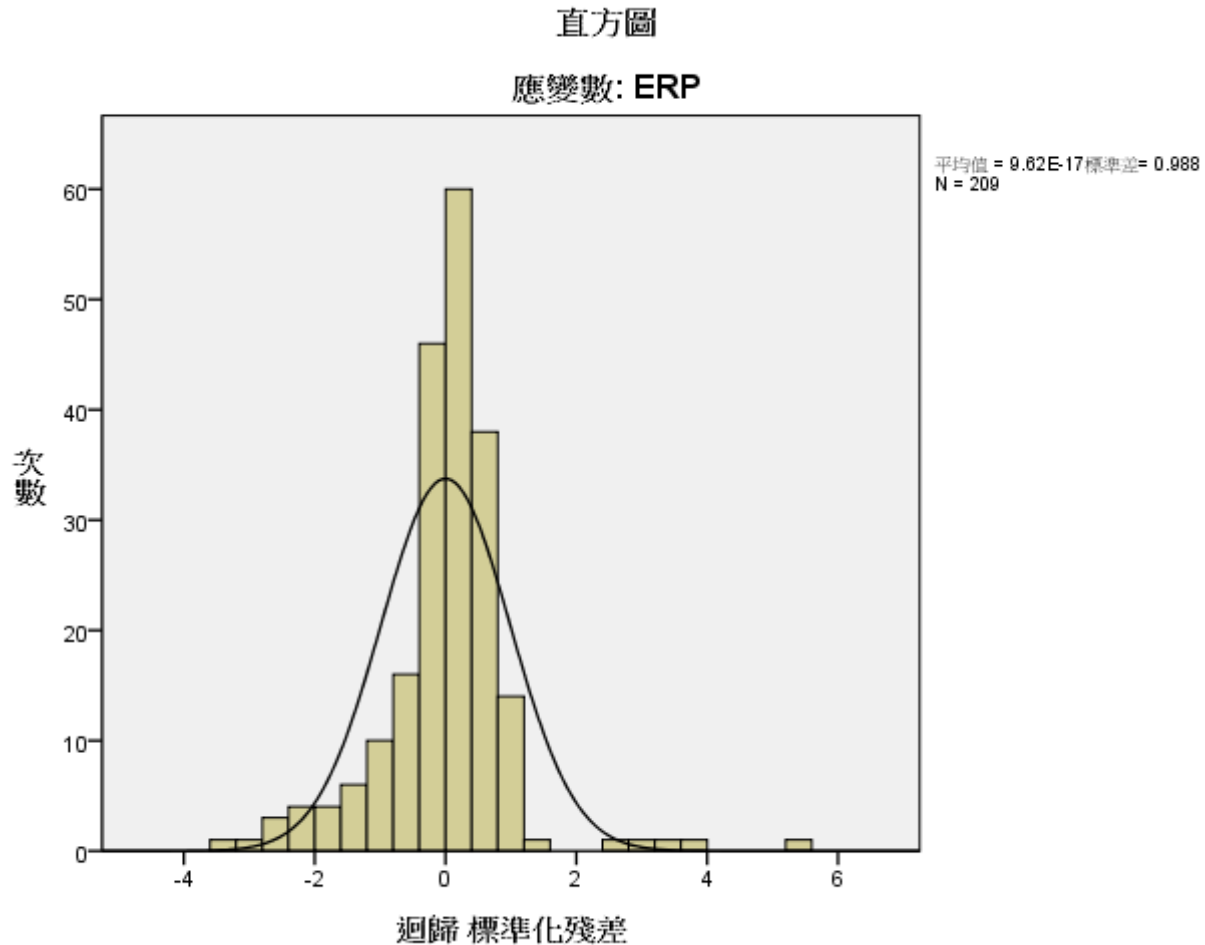
模型		非標準化係數		標準化係數	T	顯著性	共線性統計資料	
		B	標準錯誤	Beta			允差	VIF
1	(常數)	1.104	3.299		.335	.738		
	PRP	.930	.017	.966	54.153	.000	1.000	1.000
2	(常數)	-16.773	3.500		-4.793	.000		
	PRP	.712	.029	.739	24.404	.000	.255	3.918
	MMAx	.003	.000	.263	8.681	.000	.255	3.918
3	(常數)	-18.103	3.419		-5.295	.000		
	PRP	.663	.031	.689	21.160	.000	.209	4.791
	MMAx	.003	.000	.237	7.826	.000	.241	4.148
	MMIN	.004	.001	.092	3.654	.000	.348	2.876
4	(常數)	-27.686	4.440		-6.235	.000		
	PRP	.654	.031	.680	21.301	.000	.207	4.826
	MMAx	.003	.000	.258	8.515	.000	.231	4.334
	MMIN	.004	.001	.101	4.075	.000	.344	2.911
	MYCT	.031	.009	.052	3.276	.001	.845	1.183
5	(常數)	-32.541	4.620		-7.044	.000		
	PRP	.596	.035	.620	16.855	.000	.150	6.675
	MMAx	.003	.000	.248	8.334	.000	.228	4.379
	MMIN	.006	.001	.142	5.138	.000	.267	3.750
	MYCT	.035	.009	.059	3.781	.000	.825	1.212
	CHMAX	.377	.121	.063	3.112	.002	.488	2.048

a. 應變數: ERP

本表針對回歸之係數描述，以Model5為例， $ERP = -32.541 + 0.596 \times PRP + 0.003 \times MMAx + 0.006 \times MMIN + 0.035 \times MYCT + 0.377 \times CHMAX$

所有變數之VIF皆小於10，表變數之間並無共線性。

殘差

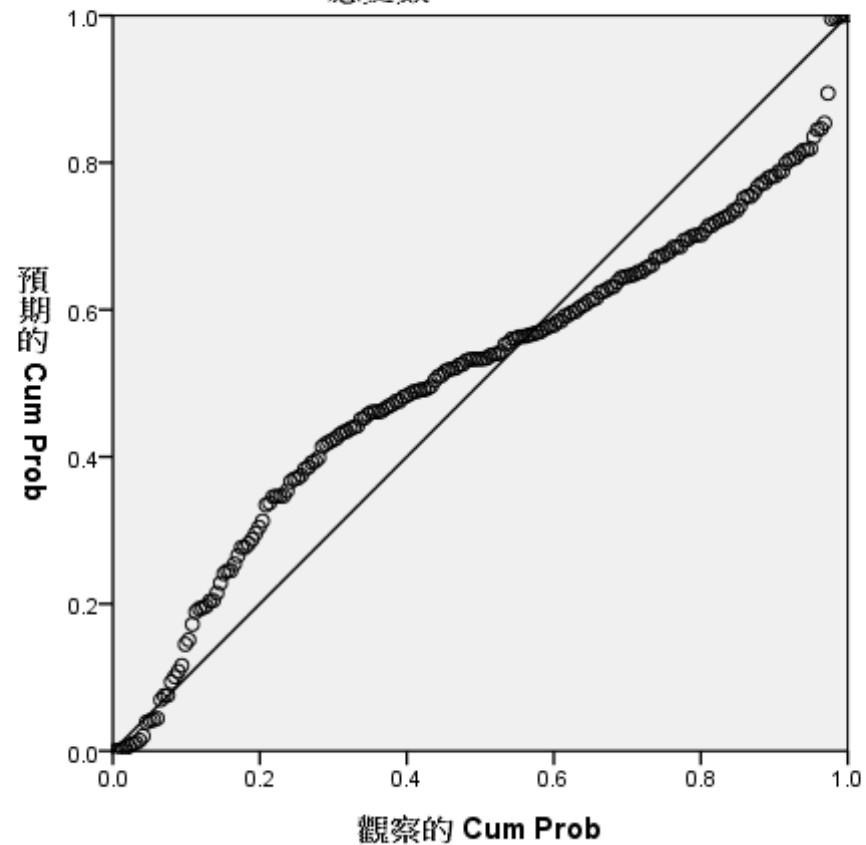


本圖為預測數值與實際值之殘差，圖形顯示接近常態分配。

P - P 圖

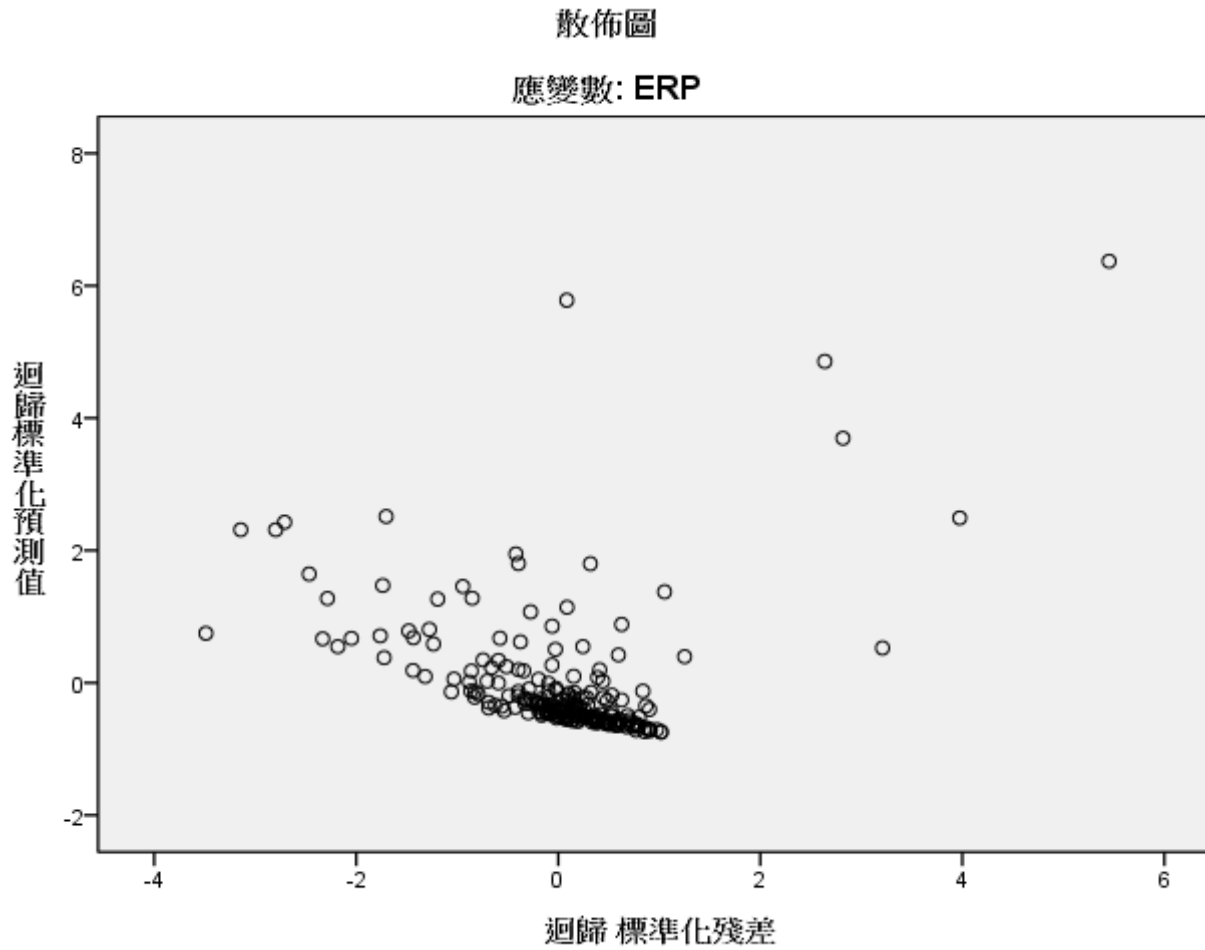
迴歸標準化殘差的常態 P-P 圖

應變數: ERP



本圖為預期與實際之變化，愈接近45度線表愈優良。

散佈圖



實際與預測之標準差殘差愈接近零，表模型之預測能力優良。

參考

1. 余民寧 逐步迴歸 <http://terms.naer.edu.tw/detail/1310873/>
2. <http://www.gotop.com.tw/epaper/e0719/AEM000900n.pdf>
3. 林淑玲 台北科技大學 上課講義
4. Durbin Watson Statistic <http://www.investopedia.com/terms/d/durbin-watson-statistic.asp>
5. IBM 303X wiki https://en.wikipedia.org/wiki/IBM_303X

Thank You