

運輸應用統計---實務資料分析 SAS 講義

資料輸入、資料初步分析與處理、散佈圖、相關係數、迴歸模式與殘差分析

為了讓同學在學習運輸應用統計上能有好的效果，因此以運管系運統修課學生成績作為此次實務資料分析資料來源，用貼近學生生活的資料，讓學生能更快進入學習狀況。

妳可能會想知道：

- ◆ 期中考成績的分布情形？
- ◆ 期末考成績的分布情形？
- ◆ 期末考成績是否和期中考成績有關係？是線性關係嗎？
- ◆ 哪些因素會影響期末考成績？（期中成績、學期報告成績、班別、實習課之出席率…？）
- ◆ 那個因素影響期末考成績最多？

如何利用統計分析工具獲得上述問題答案呢？這就是運輸應用統計這門課要教會大家的。講義編寫的過程中，考慮了第一次接觸新事物的那種心中的陌生與恐懼感，SAS 程式碼及分析結果皆附上註解來說明，另外補充一些提醒學生們要注意的”高分密技”，希望能讓每個學生能詳細地瞭解每一個步驟，進而讓學生們在考試時能順利的作答。

最後提醒學生們，只要按部就班一步一步的把這份講義充分了解，多實際練習操作，就能將 SAS 軟體成為好朋友，順利通往 PASS 的道路。

SAS 最重要的三個視窗：

- ◆ SAS file
- ◆ Log (一定要看，才知道 SAS 指令執行是否有誤！)
- ◆ Output

SAS---外部讀檔&資料分析

```
*****/
/* file name: 103(1)Tr-Sta-part-one.sas */
*****/
options linesize=80 pagesize=500;
/**外部讀檔**/
data data1;
    infile'C:\Users\iris\Desktop\104(1)課程\運統\SAS 講義-成績影響因素分析\104(1)資料
    982 運統成績.txt' dlm=';';
    input ID class $ midtermexam finalreport finalexam hw1 hw2 presence end;
/*把在 word 剪下的第一列變數名稱貼上，但變數間 “,” 改為 “空格” */
run;

/**列出所讀取資料**/
/*檢查是否有異常值或類別錯誤情形*/
proc print data=data1;
    var ID class midtermexam finalreport finalexam hw1 hw2 presence end;
run;
```

```
-----
```

Obs	ID	class	midtermexam	finalreport	finalexam	hw1	hw2	presence	end
1	1	A	15	0	0	0	0	0	1
2	2	A	103	94	94	100	0	100	1
3	3	A	87	94	94	100	0	100	1
4	4	A	87	94	94	100	0	100	1
5	5	A	74	94	94	100	0	71	1
6	6	A	91	82	82	100	0	100	1
7	7	A	60	94	94	100	0	86	1
8	8	A	79	82	82	90	0	100	1
9	9	A	95	72	89	31	0	100	1
10	10	A	66	82	82	100	0	100	1
11	11	A	88	80	80	30	0	71	1
12	12	A	63	80	80	100	0	100	1
13	13	A	81	80	80	38	0	43	1

129	129	C	63	49	30	0	0	0	1
130	130	C	31	49	35	0	0	25	1
131	131	C	37	49	27	0	0	0	1
132	132	C	40	49	22	0	0	0	1
133	133	C	20	38	38	0	0	0	1
134	134	C	12	38	0	0	-1	25	1
135	135	C	0	0	0	0	0	0	1
136	136	C	33	38	52	0	0	50	1
137	137	C	62	80	80	100	0	100	1
138	138	C	31	24	58	0	0	86	1
139	139	C	24	80	16	0	0	86	1
140	140	C	17	24	26	0	0	86	1
141	141	A	0	-99	-99	50	-1	.	1
142	142	A	36	-99	-99	0	-1	.	1
143	143	B	52	-99	-99	50	0	.	1
144	144	C	22	-99	-99	0	0	.	1

```
proc freq data=data1;
  tables ID finalexam end;
run;
```

The FREQ Procedure

ID	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1	0.71	1	0.71
2	1	0.71	2	1.43
3	1	0.71	3	2.14
4	1	0.71	4	2.86
5	1	0.71	5	3.57
6	1	0.71	6	4.29
7	1	0.71	7	5.00
8	1	0.71	8	5.71

final exam	次數	百分比	累計 次數	累計 百分比
-99	4	2.78	4	2.78
0	16	11.11	20	13.89
7	1	0.69	21	14.58
10	1	0.69	22	15.28
16	1	0.69	23	15.97
17	1	0.69	24	16.67
18	2	1.39	26	18.06
20	1	0.69	27	18.75
21	1	0.69	28	19.44
22	1	0.69	29	20.14
23	1	0.69	30	20.83
25	1	0.69	31	21.53
26	2	1.39	33	22.92
27	1	0.69	34	23.61
29	1	0.69	35	24.31
30	4	2.78	39	27.08
32	1	0.69	40	27.78
35	1	0.69	41	28.47
38	1	0.69	42	29.17
39	2	1.39	44	30.56
42	1	0.69	45	31.25
43	1	0.69	46	31.94
46	1	0.69	47	32.64
47	2	1.39	49	34.03
49	1	0.69	50	34.72
52	3	2.08	53	36.81
53	1	0.69	54	37.50
55	2	1.39	56	38.89
56	1	0.69	57	39.58
58	3	2.08	60	41.67
60	3	2.08	63	43.75
64	1	0.69	64	44.44
65	1	0.69	65	45.14
66	2	1.39	67	46.53
69	1	0.69	68	47.22
70	1	0.69	69	47.92
71	1	0.69	70	48.61
73	1	0.69	71	49.31
77	1	0.69	72	50.00
80	28	19.44	100	69.44
82	9	6.25	109	75.69
83	1	0.69	110	76.39
84	7	4.86	117	81.25
88	5	3.47	122	84.72
89	5	3.47	127	88.19
90	7	4.86	134	93.06
94	10	6.94	144	100.00

end	次數	百分比	累計 次數	累計 百分比
1	144	100.00	144	100.00

/*產生新的資料檔(newdata)，並刪除異常值(-99:退選同學)*/

/*產生分群變數(每十分為一組距)以及新類別變數*/

```
data newdata;
```

```
  set data1;
```

```
  where 0<=finalexam<=100;
```

```
  finalexamgp=.
```

```
    if 0<=finalexam<=9 then finalexamgp=0;
```

```
  else if 10<=finalexam<=19 then finalexamgp=1;
```

```
  else if 20<=finalexam<=29 then finalexamgp=2;
```

```
  else if 30<=finalexam<=39 then finalexamgp=3;
```

```
  else if 40<=finalexam<=49 then finalexamgp=4;
```

```
  else if 50<=finalexam<=59 then finalexamgp=5;
```

```
  else if 60<=finalexam<=69 then finalexamgp=6;
```

```
  else if 70<=finalexam<=79 then finalexamgp=7;
```

```
  else if 80<=finalexam<=89 then finalexamgp=8;
```

```
  else if finalexam>=90 then finalexamgp=9;
```

/*設定新變數:多類別之離散型變數*/

```
  newclassA=0;
```

```
  newclassB=0;
```

```
  newclassC=0;
```

```
  if class='A' then newclassA=1;
```

```
  if class='B' then newclassB=1;
```

```
  if class='C' then newclassC=1;
```

```
run;
```

/*檢查期末考成績(即 finalexam)異常值是否刪除，列出分群變數次數統計 */

```
proc freq data=newdata;
```

```
  tables finalexam finalexamgp;
```

```
run;
```

/*檢查新類別是否正確*/

```
proc print data=newdata;
```

```
  var class newclassA newclassB newclassC;
```

```
run;
```

```
proc freq data= newdata;
```

```
  tables class newclassA newclassB newclassC;
```

```
run;
```

finalexam	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	16	11.43	16	11.43
7	1	0.71	17	12.14
10	1	0.71	18	12.86
16	1	0.71	19	13.57
17	1	0.71	20	14.29
18	2	1.43	22	15.71
20	1	0.71	23	16.43
21	1	0.71	24	17.14
22	1	0.71	25	17.86
23	1	0.71	26	18.57
25	1	0.71	27	19.29
26	2	1.43	29	20.71
27	1	0.71	30	21.43
29	1	0.71	31	22.14
30	4	2.86	35	25.00
32	1	0.71	36	25.71
35	1	0.71	37	26.43
38	1	0.71	38	27.14
39	2	1.43	40	28.57
60	3	2.14	59	42.14
64	1	0.71	60	42.86
65	1	0.71	61	43.57
66	2	1.43	63	45.00
69	1	0.71	64	45.71
70	1	0.71	65	46.43
71	1	0.71	66	47.14
73	1	0.71	67	47.86
77	1	0.71	68	48.57
80	28	20.00	96	68.57
82	9	6.43	105	75.00
83	1	0.71	106	75.71
84	7	5.00	113	80.71
88	5	3.57	118	84.29
89	5	3.57	123	87.86
90	7	5.00	130	92.86
94	10	7.14	140	100.00

FREQ 程序

finalexamsp	次數	百分比	累計 次數	累計 百分比
0	17	12.14	17	12.14
1	5	3.57	22	15.71
2	9	6.43	31	22.14
3	9	6.43	40	28.57
4	6	4.29	46	32.86
5	10	7.14	56	40.00
6	8	5.71	64	45.71
7	4	2.86	68	48.57
8	55	39.29	123	87.86
9	17	12.14	140	100.00

Obs	class	newclass A	newclass B	newclass C
1	A	1	0	0
2	A	1	0	0
3	A	1	0	0
4	A	1	0	0
5	A	1	0	0
6	A	1	0	0
7	A	1	0	0
8	A	1	0	0
9	A	1	0	0
10	A	1	0	0
11	A	1	0	0
12	A	1	0	0
13	A	1	0	0
14	A	1	0	0
...	~	~	~	~
115	B	0	1	0
116	B	0	1	0
117	B	0	1	0
118	B	0	1	0
119	B	0	1	0
120	B	0	1	0
121	B	0	1	0
122	B	0	1	0
123	B	0	1	0
124	B	0	1	0
125	B	0	1	0
126	C	0	0	1
127	C	0	0	1
128	C	0	0	1
129	C	0	0	1
130	C	0	0	1
131	C	0	0	1
132	C	0	0	1
133	C	0	0	1
134	C	0	0	1
135	C	0	0	1
136	C	0	0	1
137	C	0	0	1
138	C	0	0	1
139	C	0	0	1
140	C	0	0	1

class	次勲	百分比	累計 次勲	累計 百分比
A	64	45.71	64	45.71
B	61	43.57	125	89.29
C	15	10.71	140	100.00

newclassA	次勲	百分比	累計 次勲	累計 百分比
0	76	54.29	76	54.29
1	64	45.71	140	100.00

newclassB	次勲	百分比	累計 次勲	累計 百分比
0	79	56.43	79	56.43
1	61	43.57	140	100.00

newclassC	次勲	百分比	累計 次勲	累計 百分比
0	125	89.29	125	89.29
1	15	10.71	140	100.00

/*期末成績的分配情形*/

成績的平均值、中位數、標準差、變異數、最大值、最小值、第1四分位和第3四分位為何?繪製期末考成績的長條圖和盒型圖

```
proc univariate plot normal data=newdata;
```

```
var finalexam;
```

```
run;
```

The UNIVARIATE Procedure
Variable: finalexam

Moments

N	140	Sum Weights	140
Mean	59.5428571	Sum Observations	8336
Std Deviation	31.3544075	Variance	983.098869
Skewness	-0.7610594	Kurtosis	-0.8576341
Uncorrected SS	633000	Corrected SS	136650.743
Coeff Variation	52.6585539	Std Error Mean	2.64993109

Basic Statistical Measures

Location		Variability	
Mean	59.54286	Std Deviation	31.35441
Median	80.00000	Variance	983.09887
Mode	80.00000	Range	94.00000
		Interquartile Range	51.50000

Tests for Location: Mu0=0

Test	-Statistic-	-----p Value-----	
Student's t	t 22.46959	Pr > t	<.0001
Sign	M 62	Pr >= M	<.0001
Signed Rank	S 3875	Pr >= S	<.0001

Tests for Normality

Test	--Statistic--	-----p Value-----	
Shapiro-Wilk	W 0.844975	Pr < W	<0.0001
Kolmogorov-Smirnov	D 0.25723	Pr > D	<0.0100
Cramer-von Mises	W-Sq 1.403793	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq 8.24421	Pr > A-Sq	<0.0050

Quantiles (Definition 5)

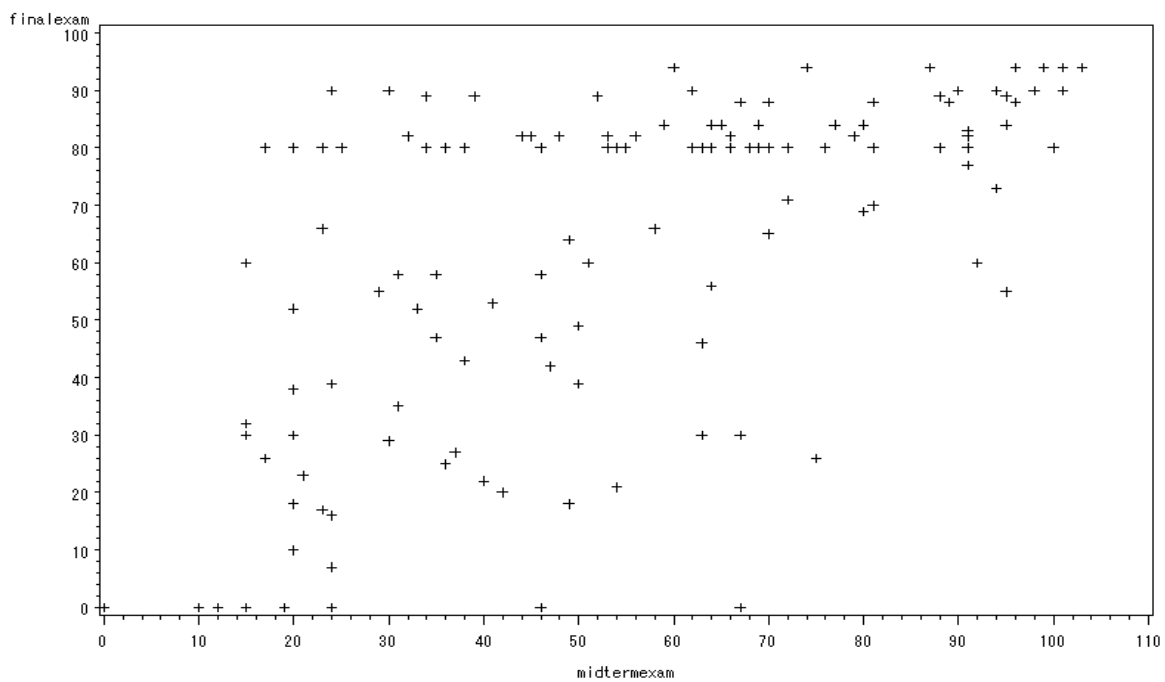
Quantile	Estimate
100% Max	94.0
99%	94.0
95%	94.0
90%	90.0
75% Q3	82.5
50% Median	80.0
25% Q1	31.0
10%	0.0
5%	0.0
1%	0.0
0% Min	0.0

Extreme Observations

----Lowest----		----Highest---	
Value	Obs	Value	Obs
0	135	94	65
0	134	94	66
0	125	94	67
0	124	94	68
0	64	94	69

/*散佈圖(scatter plot)-期中考和期末考的散佈圖為何?*/

```
proc gplot data=newdata;  
  plot finalexam*midtermexam;  
run;
```



高分密技：

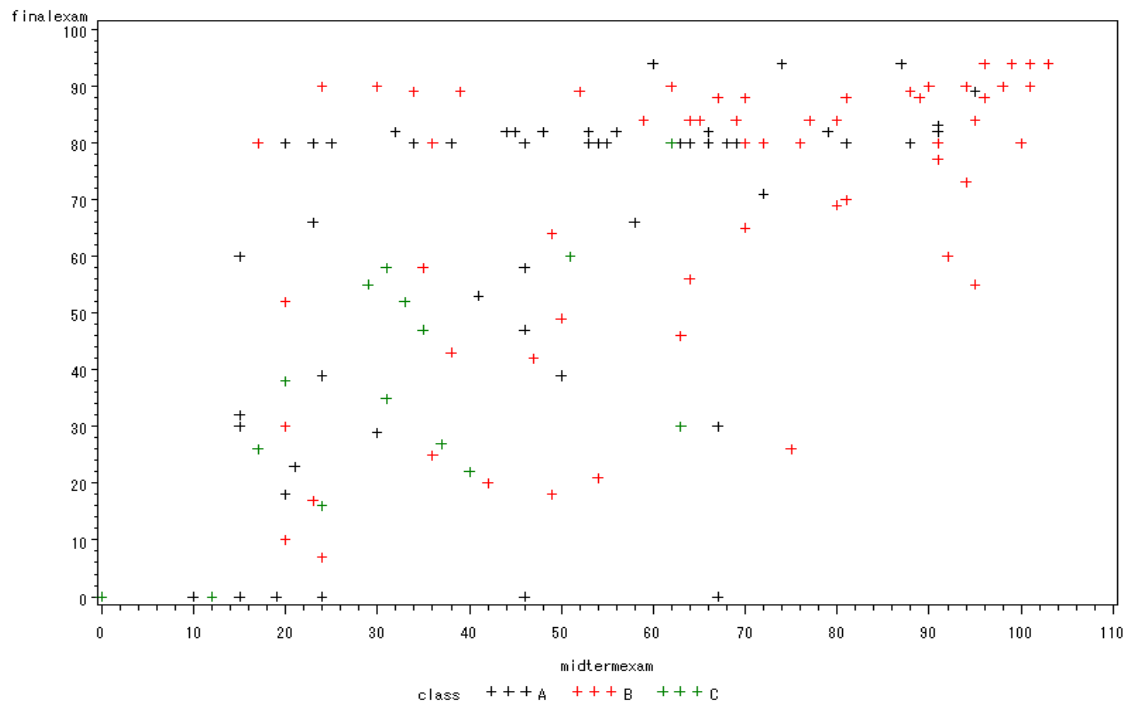
散佈圖的判斷→

1. 方向(正/負)
2. 是否為線性趨勢
3. 強度：資料散佈範圍(離直線趨勢廣/窄)
4. 是否有離群值

/*加上類別變數之散佈圖：不同班別散佈圖-期中考和期末考的散佈圖*/

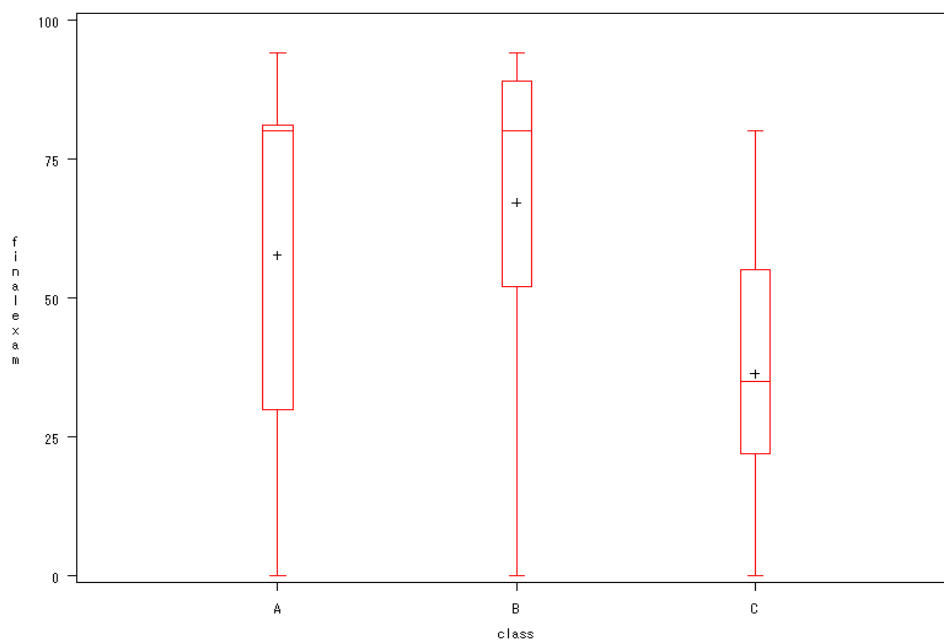
將班別以不同顏色標示方式，畫出 X 為期中考成績，Y 為期末考成績之散佈圖？
(班別是第三個變數，班別是類別變數。)

```
proc gplot data=newdata;
  plot finalexam*midtermexam=class;
run;
```



/*盒型圖(box plot)-不同班別之期末成績盒型圖為何?*/

```
proc sort data=newdata;
  by class;
proc boxplot data=newdata;
  Plot finalexam*class;
run;
```



/*相關係數-期末考成績和期中考成績的相關係數為何？這兩個變數間相關性是否顯著？記得一定還要看散佈圖！*/

```
proc corr data=newdata;
  var midtermexam finalexam ;
run;
```

The CORR Procedure

2 Variables: midtermexam finalexam

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
midtermexam	140	51.84286	29.14890	7258	0	103.00000
finalexam	140	59.54286	31.35441	8336	0	94.00000

Pearson Correlation Coefficients, N = 140
Prob > |r| under H0: Rho=0

	midtermexam	finalexam	相關係數 與 p 值
midtermexam	1.00000	0.67512 <.0001	
finalexam	0.67512 <.0001	1.00000	

高分密技：

相關係數的範圍為 1 到 -1，R=0 為無關；R>0 為正相關；R<0 為負相關。



CORR 程序

2 變數: midtermexam finalexam

簡單統計值

變數	N	平均值	標準差	總和
midtermexam	140	51.84286	29.14890	7258
finalexam	140	59.54286	31.35441	8336

簡單統計值

變數	最小值	最大值
midtermexam	0	103.00000
finalexam	0	94.00000

Pearson 相關係數, N = 140
 Prob > |r| (位於 H0 底下): Rho=0

	midtermexam	finalexam
midtermexam	1.00000	0.67512 <.0001
finalexam	0.67512 <.0001	1.00000

/*迴歸與殘差-建立期中考成績(X)與期末考成績(Y)關係的迴歸，迴歸模式殘差分析?*/

```
proc reg data=newdata;
```

```
  model finalexam=midtermexam;
```

```
run;
```

```
proc reg data=newdata;
```

```
  model finalexam=midtermexam;
```

```
  output out=database residual=res predicted=pred stdr=stdr
```

```
  STUDENT=student ;
```

```
run;
```

REG 程序
模型: MODEL1
應變數: finalexam

讀取的觀測值數目	140
使用的觀測值數目	140

變異數分析					
來源	DF	和平方	平均平方	F 值	Pr > F
模型	1	62283	62283	115.58	<.0001
誤差	138	74368	538.89569		
已校正的總計	139	136651			

根 MSE	23.21413	R 平方	0.4558
應變平均值	59.54286	調整 R 平方	0.4518
變異係數	38.98726		

參數估計值					
變數	DF	參數估計	標準誤差	t 值	Pr > t
Intercept	1	21.89463	4.01410	5.45	<.0001
midtermexam	1	0.72620	0.06755	10.75	<.0001

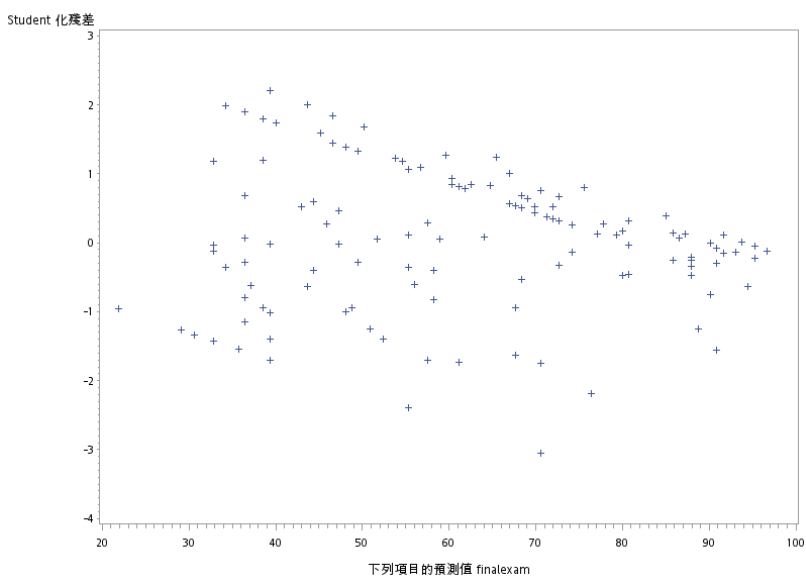
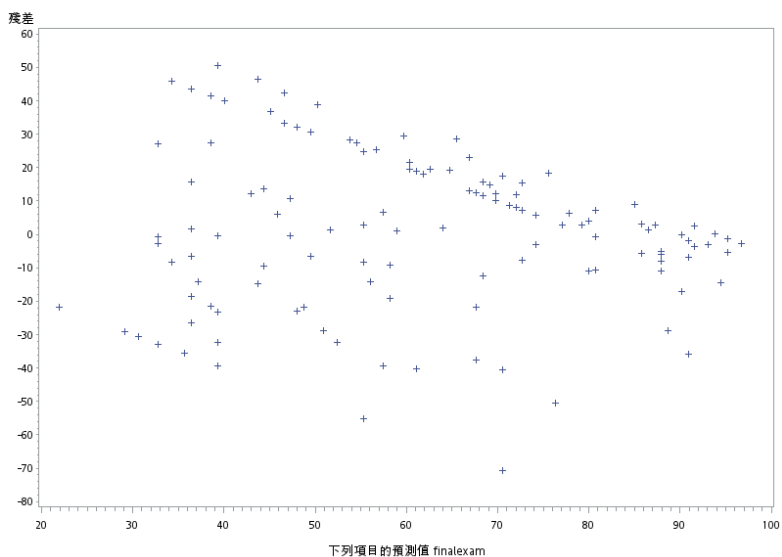
```
/*從新的資料庫(檔案名稱: database)印出殘差分析相關資料*/
proc print data=database;
  var finalexam pred res student;
run;
```

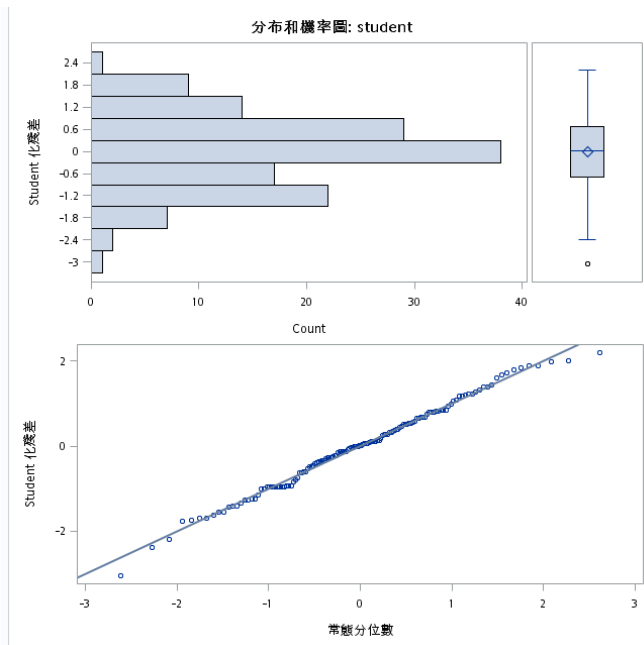
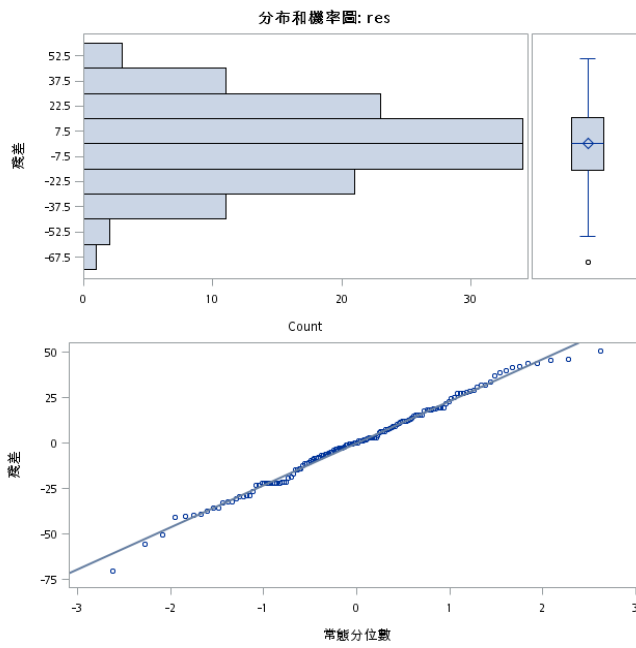
Obs	finalexam	pred	res	student
1	0	32.7876	-32.7876	-1.42575
2	94	96.6931	-2.6931	-0.11775
3	94	85.0739	8.9261	0.38794
4	94	85.0739	8.9261	0.38794
5	94	75.6334	18.3666	0.79569
6	82	87.9787	-5.9787	-0.26018
7	94	65.4666	28.5334	1.23390
8	82	79.2643	2.7357	0.11864
9	89	90.8835	-1.8835	-0.08208

```

/***** 殘差值分析 *****/
/*繪製殘插圖和Q-Q圖*/
proc gplot data=database;
  plot res*pred;
run;
proc gplot data=database;
  plot student*pred;
run;
proc univariate plot normal data=database;
  var res;
run;
proc univariate plot normal data=database;
  var student;
run;

```





```

/*類別資料初步分析*/
/*分析A,B,C三個班別的期末成績*/
proc tabulate data=newdata;
  class class;
  var finalexam;
  table (class)* finalexam,(N Min Max Median Mean std*F=6.1);
run;

```

		N	Min	Max	Median	Mean	Std
class							
A	finalexam	64	0.00	94.00	80.00	57.70	33.4
B	finalexam	61	0.00	94.00	80.00	67.16	28.1
C	finalexam	15	0.00	80.00	35.00	36.40	22.6

```

/*類別變數迴歸模式*/
proc reg data= newdata;
  model finalexam= newclassA newclassB;
run;

```

變異數分析					
來源	DF	和平方	平均值平方	F 值	Pr > F
模型	2	11793	5896.71141	6.47	0.0021
誤差	137	124857	911.36730		
已校正的總計	139	136651			

根 MSE	30.18886	R 平方	0.0863
應變平均值	59.54286	調整 R 平方	0.0730
變異係數	50.70106		

參數估計值					
變數	DF	參數估計	標準誤差	t 值	Pr > t
Intercept	1	36.40000	7.79473	4.67	<.0001
newclassA	1	21.30312	8.66013	2.46	0.0151
newclassB	1	30.76393	8.70048	3.54	0.0006

```

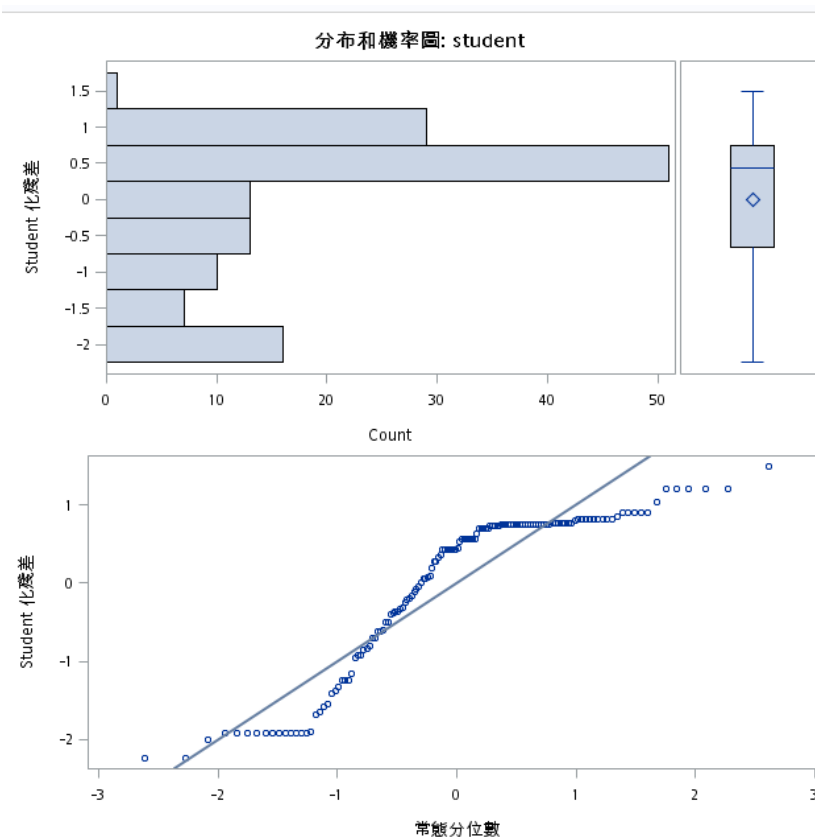
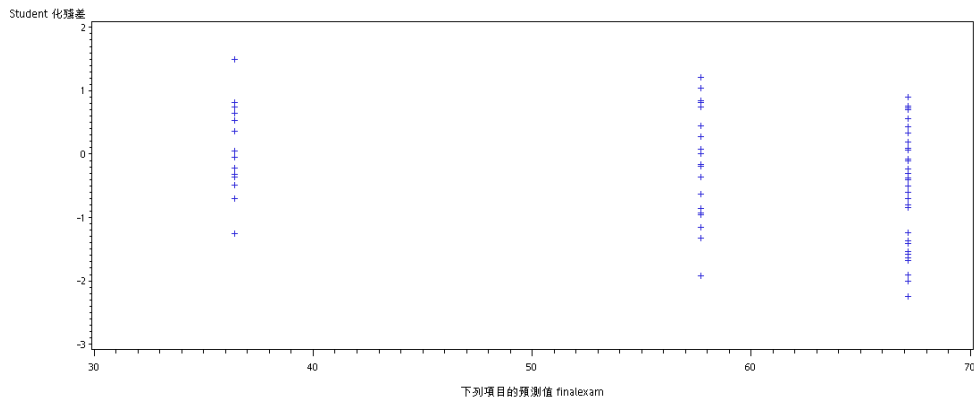
/*檢視類別變數迴歸模式是否為好模式*/
proc reg data=newdata;
  model finalexam= newclassA newclassB;
  output out=newresCatdata residual=res predicted=pred
  STUDENT=student;
run;
proc print data=newresCatdata;
  var class newclassA newclassB newclassC finalexam pred res
  student;
run;

```

		newclass	newclass	newclass				
Obs	class	A	B	C	finalexam	pred	res	student
1	A	1	0	0	0	57.7031	-57.7031	-1.92651
2	A	1	0	0	94	57.7031	36.2969	1.21183
3	A	1	0	0	94	57.7031	36.2969	1.21183
4	A	1	0	0	94	57.7031	36.2969	1.21183
5	A	1	0	0	94	57.7031	36.2969	1.21183
6	A	1	0	0	82	57.7031	24.2969	0.81119
...								
68	B	0	1	0	94	67.1639	26.8361	0.89632
69	B	0	1	0	94	67.1639	26.8361	0.89632
70	B	0	1	0	90	67.1639	22.8361	0.76272
71	B	0	1	0	90	67.1639	22.8361	0.76272
72	B	0	1	0	90	67.1639	22.8361	0.76272
73	B	0	1	0	88	67.1639	20.8361	0.69592
...								
126	C	0	0	1	60	36.4000	23.6000	0.80918
127	C	0	0	1	47	36.4000	10.6000	0.36345
128	C	0	0	1	55	36.4000	18.6000	0.63775
129	C	0	0	1	30	36.4000	-6.4000	-0.21944
130	C	0	0	1	35	36.4000	-1.4000	-0.04800
137	C	0	0	1	80	36.4000	43.6000	1.49493

```
/*繪製標準化殘差值之殘插圖和Q-Q圖*/
```

```
proc gplot data=newresCatdata;
  plot student*pred;
run;
proc univariate plot normal data=newresCatdata;
  var student;
run;
```



```
/*多元迴歸模式*/
```

```
proc reg data=newdata;
  model finalexam= midtermexam newclassA newclassB;
run;
```

變異數分析					
來源	DF	和平方	平均值平方	F 值	Pr > F
模型	3	64209	21403	40.18	<.0001
誤差	136	72442	532.65902		
已校正的總計	139	136651			

根 MSE	23.07941	R 平方	0.4699
應變平均值	59.54286	調整 R 平方	0.4582
變異係數	38.76100		

參數估計值					
變數	DF	參數估計	標準誤差	t 值	Pr > t
Intercept	1	13.05325	6.40701	2.04	0.0436
midtermexam	1	0.72206	0.07279	9.92	<.0001
newclassA	1	12.19082	6.68410	1.82	0.0704
newclassB	1	7.99325	7.03647	1.14	0.2580

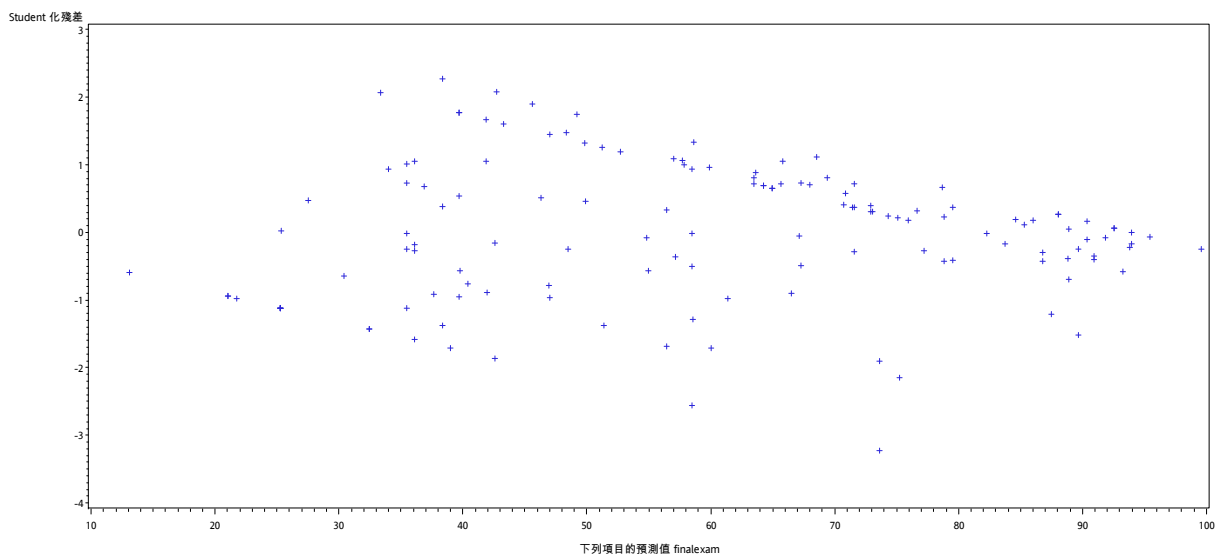
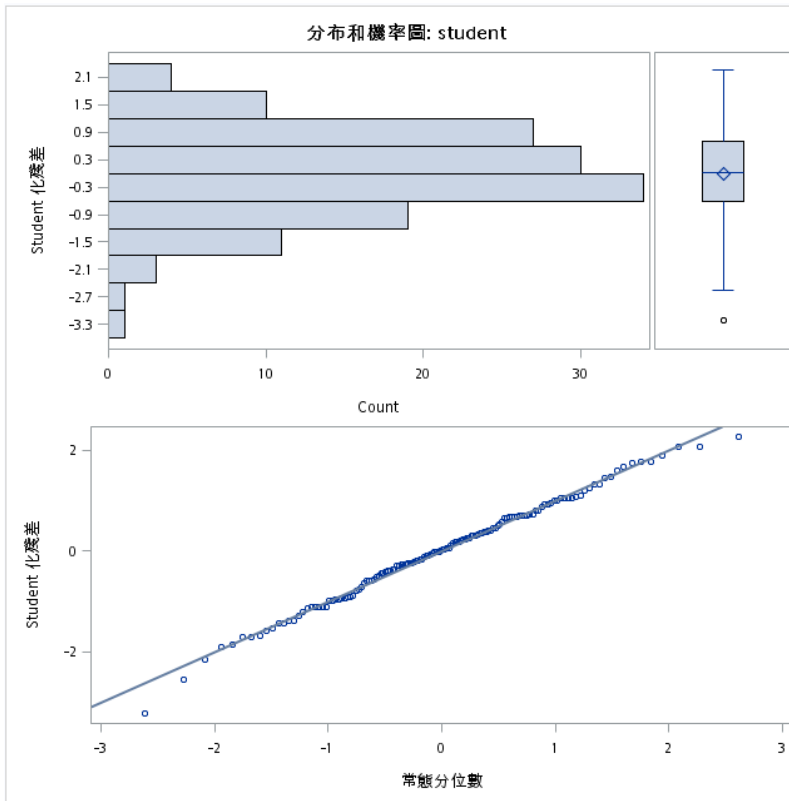
/*檢視多元迴歸模式是否為好模式*/

```
proc reg data=newdata;
  model finalexam=midtermexam newclassA newclassB;
  output out=newresdata residual=res predicted=pred
  STUDENT=student;
run;
proc print data=newresdata;
  var finalexam pred res student;
run;
```

Obs	finalexam	pred	res	student
1	0	36.0750	-36.0750	-1.58263
2	94	99.6167	-5.6167	-0.24957
3	94	88.0637	5.9363	0.26159
4	94	88.0637	5.9363	0.26159
5	94	78.6768	15.3232	0.67205
6	82	90.9519	-8.9519	-0.39520
7	94	68.5679	25.4321	1.11192
8	82	82.2872	-0.2872	-0.01262
9	89	93.8402	-4.8402	-0.21410

```
/*繪製標準化殘差值之殘插圖和Q-Q圖*/
```

```
proc gplot data=newresdata;  
  plot student*pred;  
run;  
proc univariate plot normal data=newresdata;  
  var student;  
run;
```



```

/*****/
/* file name: 104(1)Tr-Sta.sas
*/
/*****/
options linesize=80 pagesize=500;
/**外部讀檔**/
data data1;
    infile 'C:\Users\iris\Desktop\104(1)課程\運統\SAS講義-成績影響因素
    分析\104(1)資料982運統成績.txt' dlm=' ';
    input ID class $ midtermexam finalreport finalexam hw1 hw2
    presence end;
run;

/**列出所讀取資料**/
/*檢查是否有異常值或類別錯誤情形*/
proc print data=data1;
    var ID class midtermexam finalreport finalexam hw1 hw2 presence
    end;
run;

proc freq data=data1;
    tables ID finalexam end;
run;

/*產生新的資料檔(newdata)，並刪除異常值(-99:退選同學)*/
/*產生分群變數(每十分為一組距)以及新類別變數*/
data newdata;
    set data1;
    where 0<=finalexam<=100;
    finalexamgp=.;
        if 0<=finalexam<=9 then finalexamgp=0;
    else if 10<=finalexam<=19 then finalexamgp=1;
    else if 20<=finalexam<=29 then finalexamgp=2;
    else if 30<=finalexam<=39 then finalexamgp=3;
    else if 40<=finalexam<=49 then finalexamgp=4;
    else if 50<=finalexam<=59 then finalexamgp=5;
    else if 60<=finalexam<=69 then finalexamgp=6;
    else if 70<=finalexam<=79 then finalexamgp=7;
    else if 80<=finalexam<=89 then finalexamgp=8;
    else if finalexam>=90 then finalexamgp=9;

/*設定新變數:多類別之離散型變數*/
newclassA=0;

```

```

newclassB=0;
newclassC=0;
if class='A' then newclassA=1;
if class='B' then newclassB=1;
if class='C' then newclassC=1;

/*設定新變數：多類別之離散型變數*/
newhw2_good=0;
newhw2_fair=0;
newhw2_bad=0;
if hw2=1 then newhw2_good=1;
if hw2=0 then newhw2_fair=1;
if hw2=-1 then newhw2_bad=1;
run;

/*檢查期末考成績(即 finalexam)異常值是否刪除，列出分群變數次數統計 */
proc freq data=newdata;
  tables finalexam finalexamgp;
run;

/*檢查新類別是否正確*/
proc print data=newdata;
  var class newclassA newclassB newclassC;
run;

proc freq data= newdata;
  tables class newclassA newclassB newclassC;
run;

proc freq data= newdata;
  tables hw2 newhw2_good newhw2_fair newhw2_bad;
run;

/*期末成績的分配情形-期末成績的平均值、中位數、標準差、變異數、最大值、最小值、第1四分位和第3四分位為何?繪製期末考成績的長條圖和盒型圖*/
proc univariate plot normal data=newdata;
  var finalexam;
run;

/*長條圖-期末考成績的長條圖為何散佈?*/
/*Q：SAS輸出的長條圖有何問題？如何改善？*/
proc gchart data=newdata;
  vbar finalexam;

```

```

run;
/*期末考成績的長條圖資料：每十分為一組距*/
proc freq data=newdata;
  tables finalexamgp;
run;

/*散佈圖(scatter plot)-期中考和期末考的散佈圖為何?*/
proc gplot data=newdata;
  plot finalexam*midtermexam;
run;

/*加上類別變數之散佈圖：不同班別散佈圖-期中考和期末考的散佈圖*/
proc gplot data=newdata;
  plot finalexam*midtermexam=class;
run;

/*盒型圖(box plot)-不同班別之期末成績盒型圖為何?*/
proc sort data=newdata;
  by class;
proc boxplot data=newdata;
  plot finalexam*class;
run;

/*相關係數-期末考成績和期中考成績的相關係數為何？這兩個變數間相關性是否顯著？記得一定還要看散佈圖！*/
proc corr data=newdata;
  var midtermexam finalexam;
run;

/*迴歸與殘差-建立期中考成績(X)與期末考成績(Y)關係的迴歸，迴歸模式殘差分析?*/
proc reg data=newdata;
  model finalexam=midtermexam;
run;
proc reg data=newdata;
  model finalexam=midtermexam;
  output out=database residual=res predicted=pred stdr=stdr
  STUDENT=student ;
run;
/*從新的資料庫(檔案名稱：database)印出殘差分析相關資料*/
proc print data=database;
  var finalexam pred res student;
run;

```

```

/***** 殘差值分析 *****/
/*繪製殘插圖和Q-Q圖*/
proc gplot data=database;
    plot res*pred;
run;
proc gplot data=database;
    plot student*pred;
run;

proc univariate plot normal data=database;
    var res;
run;
proc univariate plot normal data=database;
    var student;
run;

/*****類別變數之迴歸模式 *****/
/*分析A,B,C三個班別的期末成績*/
proc tabulate data= newdata;
    class class;
    var finalexam;
    table (class)* finalexam, (N Min Max Median Mean Q1*F=6.1 Q3*F=6.1
std*F=6.1);
run;

/*類別變數迴歸模式*/
proc reg data=newdata;
    model finalexam= newclassA newclassB;
run;

proc reg data=newdata;
    model finalexam= newclassA newclassC;
run;

proc reg data=newdata;
    model finalexam= newclassB newclassC;
run;

/***** hw2 分析 *****/
proc tabulate data= newdata;
    class hw2;
    var finalexam;

```

```

    table (hw2)* finalexam, (N Min Max Median Mean Q1*F=6.1 Q3*F=6.1
std*F=6.1);
run;
proc reg data=newdata;
    model finalexam=newhw2_good newhw2_fair;
run;

proc reg data=newdata;
    model finalexam=newhw2_fair newhw2_bad;
run;

/*檢視類別變數迴歸模式是否為好模式*/
proc reg data=newdata;
    model finalexam= newclassA newclassB;
    output out=newresCatdata residual=res predicted=pred
STUDENT=student;
run;
proc print data=newresCatdata;
    var finalexam pred res student;
run;
/*繪製標準化殘差值之殘插圖和Q-Q圖*/
proc gplot data=newresCatdata;
    plot student*pred;
run;
proc univariate plot normal data=newresCatdata;
    var student;
run;

/*多元迴歸模式*/
proc reg data=newdata;
    model finalexam= midtermexam newclassA newclassB;
run;
/*檢視多元迴歸模式是否為好模式*/
proc reg data=newdata;
    model finalexam=midtermexam newclassA newclassB;
    output out=newresdata residual=res predicted=pred
STUDENT=student;
run;
proc print data=newresdata;
    var newclassA newclassB midtermexam finalexam pred res student;
run;
/*繪製標準化殘差值之殘插圖和Q-Q圖*/
proc gplot data=newresdata;

```

```
plot student*pred;  
run;  
proc univariate plot normal data=newresdata;  
var student;  
run;
```