

110 學年度第二學期科學計算軟體作業九

姓名： 蕭合亭 學號： F64109527

使用 datasets package 中附加的 mtcars 資料集進行分析，透過迴歸分析探討因子對於汽車油耗量的影響。內容需包含下列項目：

*共 5 題，每題 20%；答題提醒：請使用紅框標註 **p 值(sig.)**、 **β (Beta, 估計值)** 及 **R^2** 等主要判斷依據，若未達到或錯誤皆會斟酌扣分。

*(2).~(5).需展示模型配適度(ANOVA)、模型摘要(Summary)、參數估計值(Estimate)等表。

(1). 各變數間之相關性檢定(本表結果表可不用標註相關資訊，但須說明各項因子與油耗量之間的關係正負及是否顯著)

程式碼

```
rcorr(as.matrix(HW_data), type = c("pearson"))
```

結果

```
> rcorr(as.matrix(HW_data), type = c("pearson"))
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg	1.00	-0.85	-0.85	-0.78	0.68	-0.87	0.42	0.66	0.60	0.48	-0.55
cyl	-0.85	1.00	0.90	0.83	-0.70	0.78	-0.59	-0.81	-0.52	-0.49	0.53
disp	-0.85	0.90	1.00	0.79	-0.71	0.89	-0.43	-0.71	-0.59	-0.56	0.39
hp	-0.78	0.83	0.79	1.00	-0.45	0.66	-0.71	-0.72	-0.24	-0.13	0.75
drat	0.68	-0.70	-0.71	-0.45	1.00	-0.71	0.09	0.44	0.71	0.70	-0.09
wt	-0.87	0.78	0.89	0.66	-0.71	1.00	-0.17	-0.55	-0.69	-0.58	0.43
qsec	0.42	-0.59	-0.43	-0.71	0.09	-0.17	1.00	0.74	-0.23	-0.21	-0.66
vs	0.66	-0.81	-0.71	-0.72	0.44	-0.55	0.74	1.00	0.17	0.21	-0.57
am	0.60	-0.52	-0.59	-0.24	0.71	-0.69	-0.23	0.17	1.00	0.79	0.06
gear	0.48	-0.49	-0.56	-0.13	0.70	-0.58	-0.21	0.21	0.79	1.00	0.27
carb	-0.55	0.53	0.39	0.75	-0.09	0.43	-0.66	-0.57	0.06	0.27	1.00

n= 32

P	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
mpg		0.0000	0.0000	0.0000	0.0000	0.0000	0.0171	0.0000	0.0003	0.0054	0.0011
cyl	0.0000		0.0000	0.0000	0.0000	0.0000	0.0004	0.0000	0.0022	0.0042	0.0019
disp	0.0000	0.0000		0.0000	0.0000	0.0000	0.0131	0.0000	0.0004	0.0010	0.0253
hp	0.0000	0.0000	0.0000		0.0100	0.0000	0.0000	0.0000	0.1798	0.4930	0.0000
drat	0.0000	0.0000	0.0000	0.0100		0.0000	0.6196	0.0117	0.0000	0.0000	0.6212
wt	0.0000	0.0000	0.0000	0.0000	0.0000		0.3389	0.0010	0.0000	0.0005	0.0146
qsec	0.0171	0.0004	0.0131	0.0000	0.6196	0.3389		0.0000	0.2057	0.2425	0.0000
vs	0.0000	0.0000	0.0000	0.0000	0.0117	0.0010	0.0000		0.3570	0.2579	0.0007
am	0.0003	0.0022	0.0004	0.1798	0.0000	0.0000	0.2057	0.3570		0.0000	0.7545
gear	0.0054	0.0042	0.0010	0.4930	0.0000	0.0005	0.2425	0.2579	0.0000		0.1290
carb	0.0011	0.0019	0.0253	0.0000	0.6212	0.0146	0.0000	0.0007	0.7545	0.1290	

分析

mpg-cyl：相關性=-0.85，為高度負相關； $p=0.000<0.05$ ，達顯著性。

mpg-disp：相關性=-0.85，為高度負相關； $p=0.000<0.05$ ，達顯著性。

mpg-hp：相關性=-0.78，為高度負相關； $p=0.000<0.05$ ，達顯著性。

mpg-drat: 相關性=0.68, 為高度正相關; $p=0.000<0.05$, 達顯著性。
 mpg-wt: 相關性=-0.87, 為高度負相關; $p=0.000<0.05$, 達顯著性。
 mpg-qsec: 相關性=0.42, 為中度正相關; $p=0.0171<0.05$, 達顯著性。
 mpg-vs: 相關性=0.66, 為高度負相關; $p=0.000<0.05$, 達顯著性。
 mpg-am: 相關性=0.60, 為高度負相關; $p=0.0003<0.05$, 達顯著性。
 mpg-gear: 相關性=0.48, 為高度負相關; $p=0.0054<0.05$, 達顯著性。
 mpg-carb: 相關性=-0.55, 為高度負相關; $p=0.0011<0.05$, 達顯著性。

(2). 強制輸入法(+共線性檢定(回答根據數值是否具有嚴重共線性)+列出模型完整公式)

程式碼

```
ols_coll_diag(data_model)
data_model<-
lm(HW_data$mpg~HW_data$cyl+HW_data$disp+HW_data$hp
+HW_data$drat+HW_data$wt+HW_data$qsec+HW_data$vs
+HW_data$am+HW_data$gear+HW_data$carb)
ols_regress(data_model)
```

結果

```
> ols_coll_diag(data_model)
Tolerance and Variance Inflation Factor
```

	Variables	Tolerance	VIF
1	HW_data\$cyl	0.06504559	15.373833
2	HW_data\$disp	0.04625295	21.620241
3	HW_data\$hp	0.10170833	9.832037
4	HW_data\$drat	0.29632966	3.374620
5	HW_data\$wt	0.06594189	15.164887
6	HW_data\$qsec	0.13283814	7.527958
7	HW_data\$vs	0.20137444	4.965873
8	HW_data\$am	0.21512374	4.648487
9	HW_data\$gear	0.18665589	5.357452
10	HW_data\$carb	0.12644228	7.908747

VIF>5

容忍度接近 0

```
Eigenvalue and Condition Index
```

	Eigenvalue	Condition Index	intercept	HW_data\$cyl	HW_data\$disp	HW_data\$hp	HW_data\$drat	HW_data\$wt	HW_data\$qsec	HW_data\$vs
1	9.0972341796	1.000000	7.470411e-06	5.544824e-05	0.0001000700	0.0001904575	7.237917e-05	6.125887e-05	1.486910e-05	0.0005203731
2	1.1283828258	2.839399	1.303194e-06	1.660349e-04	0.0011911144	0.0011420445	1.197465e-04	1.679096e-04	9.834711e-06	0.0381139628
3	0.5639483900	4.016382	7.499117e-06	1.738962e-05	0.0002075037	0.0003771330	2.126300e-06	2.575342e-04	7.059635e-05	0.0613503974
4	0.1157787780	8.864213	1.495329e-04	1.247180e-03	0.0075728512	0.0044208629	6.946065e-04	2.175963e-04	3.407483e-04	0.1012280913
5	0.0483108351	17.22474	6.608085e-04	2.462087e-04	0.0409890279	0.0996404257	1.171260e-02	2.046533e-04	2.208025e-03	0.3357830929
6	0.0220374547	20.317673	4.208213e-04	1.277445e-02	0.0398523190	0.2293265565	9.516545e-03	9.068983e-02	5.712918e-05	0.0138389199
7	0.0096634346	30.682372	1.086739e-03	2.449967e-01	0.0840661581	0.0134842220	4.614638e-02	1.496418e-04	1.396853e-03	0.1059157259
8	0.0063040967	37.987722	5.301451e-04	7.829684e-02	0.3624911690	0.5760360686	6.739206e-02	1.809285e-01	2.595708e-02	0.2268441562
9	0.0059339704	39.154529	1.213686e-04	3.457040e-02	0.0010747718	0.0626748986	5.461145e-01	9.949836e-05	1.687102e-07	0.0108468126
10	0.0019637683	68.062761	2.548698e-02	2.460255e-01	0.4533776162	0.0087357570	2.028301e-01	6.788279e-01	2.220045e-01	0.0801584405
11	0.0004422669	143.420915	9.715273e-01	3.816038e-01	0.0090773986	0.0039715738	1.153989e-01	4.839575e-02	7.479402e-01	0.0254000274

CI>30

特徵值接近 0

```
HW_data$am HW_data$gear HW_data$carb
1 0.000644664 8.106547e-05 0.0003091537
2 0.028249206 1.340825e-04 0.0010097688
3 0.116394472 7.683592e-05 0.0045589756
4 0.060332928 3.221708e-05 0.1639262366
5 0.184543665 8.019726e-03 0.0219539142
6 0.261153865 1.663697e-02 0.0725189384
7 0.258583700 2.077448e-01 0.0279297018
8 0.023785086 6.541188e-03 0.1746454163
9 0.001427952 4.823324e-01 0.0096870513
10 0.004501660 1.608152e-01 0.5232687171
11 0.060382802 1.175855e-01 0.0001921261
> |
```

```
> data_model<-lm(HW_data$mpg~HW_data$cyl+HW_data$disp+HW_data$hp
+ +HW_data$drat+HW_data$wt+HW_data$qsec+HW_data$vs+HW_data$am
+ +HW_data$gear+HW_data$carb, data=HW_data)
> ols_regress(data_model)
```

Model Summary

R	0.932	RMSE	2.650
R-Squared	0.869	Coef. Var	13.191
Adj. R-Squared	0.807	MSE	7.024
Pred R-Squared	0.654	MAE	1.723

RMSE: Root Mean Square Error
MSE: Mean Square Error
MAE: Mean Absolute Error

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	978.553	10	97.855	13.932	0.0000
Residual	147.494	21	7.024		
Total	1126.047	31			

Parameter Estimates

model	Beta	Std. Error	Std. Beta	t	Sig.
(Intercept)	12.303	18.718		0.657	0.518
HW_data\$cyl	-0.111	1.045	-0.033	-0.107	0.916
HW_data\$disp	0.013	0.018	0.274	0.747	0.463
HW_data\$hp	-0.021	0.022	-0.244	-0.987	0.335
HW_data\$drat	0.787	1.635	0.070	0.481	0.635
HW_data\$wt	-3.715	1.894	-0.603	-1.961	0.063
HW_data\$qsec	0.821	0.731	0.243	1.123	0.274
HW_data\$vs	0.318	2.105	0.027	0.151	0.881
HW_data\$am	2.520	2.057	0.209	1.225	0.234
HW_data\$gear	0.655	1.493	0.080	0.439	0.665
HW_data\$carb	-0.199	0.829	-0.053	-0.241	0.812

```
> |
```

分析

sig=0.000，mpg 對於其他資料達到統計上的顯著性

R-squared=0.869，表示可以解釋約 87%。

Adj R-squared=0.807，表示可以解釋約 81%。

ANOVA 的 sum of squares

(3). 順向進入法(+變數選入摘要表(Selection summary))

程式碼

```
ols_step_forward_p(data_model,penter=0.05,details=TRUE)
```

結果

Parameter Estimates							
model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	39.686	1.715		23.141	0.000	36.179	43.194
HW_data\$wt	-3.191	0.757	-0.518	-4.216	0.000	-4.739	-1.643
HW_data\$cyl	-1.508	0.415	-0.447	-3.636	0.001	-2.356	-0.660

Selection Summary						
Step	Variable Entered	R-Square	Adj. R-Square	C (p)	AIC	RMSE
1	HW_data\$wt	0.7528	0.7446	11.6270	166.0294	3.0459
2	HW_data\$cyl	0.8302	0.8185	1.2187	156.0101	2.5675

分析

使用順向進入法，當 p 值 < 0.05 時才會進入，選取了 wt、cyl 欄。

wt : sig = 0.000，達統計上顯著標準。

cyl : sig = 0.001，達統計上顯著標準。

選取摘要表：

R-square：

wt 的 R-square = 0.7528，表示可以解釋約 75%。

cyl 的 R-square = 0.8302，表示可以解釋約 83%。

Adj R-square：

wt 修正過後的 R-square = 0.7446，表示可以解釋約 74%。

cyl 修正過後的 R-square = 0.8185，表示可以解釋約 82%。

(4). 反向淘汰法(+變數淘汰摘要表(Elimination summary))

程式碼

```
ols_step_backward_p(data_model, prem=0.1, details=TRUE)
```

結果

Parameter Estimates							
model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	9.618	6.960		1.382	0.178	-4.638	23.874
HW_data\$wt	-3.917	0.711	-0.636	-5.507	0.000	-5.373	-2.460
HW_data\$qsec	1.226	0.289	0.363	4.247	0.000	0.635	1.817
HW_data\$am	2.936	1.411	0.243	2.081	0.047	0.046	5.826

Elimination Summary						
Step	Variable Removed	R-Square	Adj. R-Square	C (p)	AIC	RMSE
1	HW_data\$cyl	0.8689	0.8153	9.0114	161.7271	2.5900
2	HW_data\$vs	0.8687	0.823	7.0496	159.7853	2.5353
3	HW_data\$carb	0.8681	0.8296	5.1472	157.9333	2.4877
4	HW_data\$gear	0.8667	0.8347	3.3700	156.2687	2.4503
5	HW_data\$drat	0.8637	0.8375	1.8462	154.9740	2.4293
6	HW_data\$disp	0.8579	0.8368	0.7900	154.3274	2.4348
7	HW_data\$hp	0.8497	0.8336	0.1026	154.1194	2.4588

分析

使用反向淘汰法，逐步將 p 值>0.1 的項目刪除，最後淘汰了 cyl、vs、carb、gear、drat、disp、hp 項。

wt : sig= 0.000，達統計上顯著標準。

qsec : sig= 0.000，達統計上顯著標準。

am : sig=0.047，達統計上顯著標準。

淘汰摘要表：

R-square：

cyl 的 R-square = 0.8689，表示可以解釋約 87%。

vs 的 R-square = 0.8687，表示可以解釋約 87%。

carb 的 R-square = 0.8681，表示可以解釋約 87%。

gear 的 R-square = 0.8667，表示可以解釋約 87%。

drat 的 R-square = 0.8637，表示可以解釋約 86%。

disp 的 R-square = 0.8579，表示可以解釋約 86%。

hp 的 R-square = 0.8497，表示可以解釋約 85%。

Adj R-square：

cyl 修正過後的 R-square = 0.8153，表示可以解釋約 82%。

vs 修正過後的 R-square = 0.823，表示可以解釋約 82%。

carb 修正過後的 R-square = 0.8296，表示可以解釋約 83%。

gear 修正過後的 R-square = 0.8347，表示可以解釋約 83%。

drat 修正過後的 R-square = 0.8375，表示可以解釋約 84%。

disp 修正過後的 R-square = 0.8368，表示可以解釋約 84%。

hp 修正過後的 R-square = 0.8336，表示可以解釋約 83%。

(5). 逐步分析法(+變數選入/淘汰摘要表(Stepwise selection summary))

*選入(Enter)、移除(Remove)條件均 0.05 及 0.1

程式碼

```
ols_step_both_p(data_model, penter=0.05, prem=0.1, details=TRUE)
```

結果

Parameter Estimates							
model	Beta	Std. Error	Std. Beta	t	Sig	lower	upper
(Intercept)	39.686	1.715		23.141	0.000	36.179	43.194
HW_data\$wt	-3.191	0.757	-0.518	-4.216	0.000	-4.739	-1.643
HW_data\$cyl	-1.508	0.415	-0.447	-3.636	0.001	-2.356	-0.660

Stepwise Selection Summary							
Step	Variable	Added/ Removed	R-Square	Adj. R-Square	C(p)	AIC	RMSE
1	HW_data\$wt	addition	0.753	0.745	11.6270	166.0294	3.0459
2	HW_data\$cyl	addition	0.830	0.819	1.2190	156.0101	2.5675

分析

使用逐步分析法(同時衡量選入與淘汰)

將 wt、cyl 加入參數：

wt : sig= 0.000，達統計上顯著標準。

cyl : sig= 0.001，達統計上顯著標準。

摘要表：

加入項

R-square：

wt 的 R-square = 0.753，表示可以解釋約 75%。

cyl 的 R-square = 0.830，表示可以解釋約 83%。

Adj R-square：

wt 修正後的 R-square = 0.745，表示可以解釋約 75%。

cyl 修正後的 R-square = 0.819，表示可以解釋約 82%。

```
>> data(mtcars, package = "datasets")
```

```
>> HW_data <- mtcars
```

```
>> View(HW_data)
```

row.names	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2

#Variables:

mpg	油耗，每加侖里程數(英里/每加侖)	qsec	1/4英里行駛時間
cyl	汽缸數	vs	引擎類型(0= V型; 1= 直型) *類別變數
disp	排氣量(立方英寸)	am	變速箱類型(0= 自動; 1= 手動) *類別變數
hp	馬力	gear	前進檔數
drat	後軸比	carb	化油器數量
wt	重量		