

應用統計 試題

(限用答案本作答)

可使用計算機

- 一、當你設計一問卷或量表時，你如何去評估該問卷或量表之效度與信度？(請詳細說明)(10分)
- 二、探索性因素分析與驗證性因素分析有何區別？(5分)
- 三、某量販店投資者正考慮兩個可能設店地點；他隨機選擇時段，測量每一地點的路人流量，結果如下：

地點 1	35	72	79	66	69	40	90	87	75	38	92
地點 2	46	60	52	62	55	63	50	48	54	57	

- (a) 選擇最適當的方法詳細分析此資料。(10分)
 - (b) 該投資者應選擇哪一地點設店較好？(5分)
- 四、某財務顧問公司欲研究製造業、零售業、銀行業、教育業明年度計畫調薪幅度，自上述行業中各隨機抽出 5 家調查，其中銀行業有一家並未回應，調查結果經整理如下表：

製造業	零售業	銀行業	教育業
$n_1 = 5$	$n_2 = 5$	$n_3 = 4$	$n_4 = 5$
$\bar{x}_1 = 4.0\%$	$\bar{x}_2 = 3.2\%$	$\bar{x}_3 = 3.3\%$	$\bar{x}_4 = 2.3\%$
$s_1 = 1.6\%$	$s_2 = 1.3\%$	$s_3 = 1.4\%$	$s_4 = 0.8\%$

- (a) 試建立單因子變異數分析表。(10分)
 - (b) 以 5%顯著水準檢定上述四個行業平均調薪幅度是否一致。(5分)
 - (c) 試估計製造業平均調薪幅度之 95%信賴區間。(5分)
 - (d) 試以 5%聯合顯著水準，對於上述四個行業平均調薪幅度進行多重比較。(10分)
- 五、某大公司之品牌經理欲瞭解價格與促銷活動對於其產品(Brand X)市場佔有率之影響。該經理建立一資料檔(BRAND_X)，包含一組客戶在 101 週內連續被追蹤其購買行程以及在該類產品之購買活動。該資料檔所包含之變數如下表：

Variable	Definition
SHARE_X	Market share for Brand X (in percent)
WEEK	Week (from 56 to 156)
REL_PRICE	Relative price index for Brand X
LAG_PRICE	Lagged price index for Brand X
PROMO	Index of promotional activity by Brand X: 0 = no promotional activity by Brand X; 100 = all promotional activity is Brand X
LOY_HH	Number of shopping trips taken by households in the panel characterized as "loyal" to Brand X
NONLOY_HH	Number of shopping trips taken by households in the panel characterized as "not loyal" to Brand X

應用統計 試題

(限用答案本作答)

- (a) 根據以下的相關係數矩陣以及包含所有解釋變數的線性模型之分析結果，試找出一較合適的模型來解釋佔有率(SHARE_X)。(10分)

Pearson Correlation Coefficients, N = 101

	SHARE_X	WEEK	REL_PRICE	LAG_PRICE	PROMO	LOY_HH	NONLOY_H
SHARE_X	1.00000	-0.18141	-0.50164	0.18289	0.58646	0.53822	-0.38166
WEEK	-0.18141	1.00000	-0.04123	-0.02804	-0.04079	-0.06873	-0.30843
REL_PRICE	-0.50164	-0.04123	1.00000	-0.05389	-0.25405	-0.23061	0.16026
LAG_PRICE	0.18289	-0.02804	-0.05389	1.00000	0.19018	0.16875	-0.23018
PROMO	0.58646	-0.04079	-0.25405	0.19018	1.00000	0.45375	-0.18292
LOY_HH	0.53822	-0.06873	-0.23061	0.16875	0.45375	1.00000	-0.12487
NONLOY_H	-0.38166	-0.30843	0.16026	-0.23018	-0.18292	-0.12487	1.00000

Dependent Variable: SHARE_X

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	3650.70047	608.45008	30.15	<.0001
Error	94	1897.10151	20.18193		
Corrected Total	100	5547.80198			

Root MSE	4.49243	R-Square	0.6580
Dependent Mean	26.89109	Adj R-Sq	0.6362
Coeff Var	16.70602		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	101.53156	16.33624	6.22	<.0001
WEEK	1	-0.06722	0.01632	-4.12	<.0001
REL_PRICE	1	-0.60893	0.12083	-5.04	<.0001
LAG_PRICE	1	-0.02870	0.08233	-0.35	0.7282
PROMO	1	0.22242	0.04861	4.58	<.0001
LOY_HH	1	0.31480	0.08181	3.85	0.0002
NONLOY_H	1	-0.26391	0.05419	-4.87	<.0001

- (b) 假設刪除變數 LAG_PRICE 再重新分析此模型；試根據以下分析結果詳細探討新模型配適情形以及各解釋變數如何影響佔有率。(15分)

應用統計 試題

(限用答案本作答)

Dependent Variable: SHARE_X

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	3648.24787	729.64957	36.49	<.0001
Error	95	1899.55411	19.99531		
Corrected Total	100	5547.80198			

Root MSE	4.47161	R-Square	0.6576
Dependent Mean	26.89109	Adj R-Sq	0.6396
Coeff Var	16.62860		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Standardized Estimate	Tolerance	Variance Inflation
Intercept	1	98.41482	13.60915	7.23	<.0001	0		0
WEEK	1	-0.06674	0.01618	-4.12	<.0001	-0.26253	0.88934	1.12443
REL_PRICE	1	-0.61012	0.12022	-5.07	<.0001	-0.32007	0.90607	1.10367
PROMO	1	0.22071	0.04814	4.59	<.0001	0.31660	0.75594	1.32285
LOY_HH	1	0.31244	0.08115	3.85	0.0002	0.26267	0.77427	1.29155
NONLOY_H	1	-0.25980	0.05265	-4.93	<.0001	-0.32063	0.85374	1.17131

Collinearity Diagnostics

Number	Eigenvalue	Condition Index	Proportion of Variance					
			Intercept	WEEK	REL_PRICE	PROMO	LOY_HH	NONLOY_H
1	5.71580	1.00000	0.00003242	0.00179	0.00003766	0.00375	0.00112	0.00070624
2	0.16817	5.82997	0.00019136	0.03253	0.00029063	0.64050	0.00235	0.00992
3	0.06945	9.07230	0.00020083	0.62211	0.00030034	0.02326	0.02043	0.07958
4	0.03490	12.79733	0.00017023	0.00105	0.00039465	0.23496	0.80937	0.11229
5	0.01110	22.69551	0.02289	0.32220	0.03382	0.00546	0.12073	0.79486
6	0.00059145	98.30567	0.97652	0.02031	0.96516	0.03206	0.04599	0.00264

六、在統計上有一組很有名的資料就是 Fisher 的蝴蝶花(iris)資料，這組資料包含三種蝴蝶花品種 (setosa、versicolor、virginica)，每一品種蝴蝶花都有 50 筆觀察值，每一筆觀察值包含蝴蝶花四個屬性(花萼長、花萼寬、花瓣長、花瓣寬)之測量。這組資料經主成分分析部分結果如下：

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	2.91849782	2.00446735	0.7296	0.7296
2	0.91403047	0.76727360	0.2285	0.9581
3	0.14675688	0.12604204	0.0367	0.9948
4	0.02071484		0.0052	1.0000

前兩個主成分之權重(與負荷)

Variable	Prin1	Prin2
X1 (花萼長)	0.521 (0.890)	0.377 (0.361)
X2 (花萼寬)	-0.269 (-0.460)	0.923 (0.883)
X3 (花瓣長)	0.580 (0.992)	0.024 (0.023)
X4 (花瓣寬)	0.564 (0.965)	0.067 (0.064)

應用統計 試題

(限用答案本作答)

三品種蝴蝶花之主成分平均分數

品種	Prin1	Prin2
1 (setosa)	-2.217	0.288
2 (versicolor)	0.495	-0.548
3 (virginica)	1.723	0.260

- (a) 前兩個主成分可以解釋多少資料變異比例？(5 分)
- (b) 這兩個主成分之意義分別為何？(5 分)
- (c) 將三品種蝴蝶花之主成分平均分數繪於前兩主成分平面上，描述你有何發現？(5 分)

《附表》

Quantiles of the Mann-Whitney test statistic

n_1	p	$n_2=10$	$n_2=11$
10	0.001	11	13
	0.005	17	19
	0.010	20	23
	0.025	24	27
	0.050	28	32
	0.100	33	37
11	0.001	13	16
	0.005	19	22
	0.010	23	26
	0.025	27	31
	0.050	32	35
	0.100	37	41

Critical values of t

自由度	0.05	0.025
10	1.812	2.228
11	1.796	2.201
12	1.782	2.179
13	1.771	2.160
14	1.761	2.145
15	1.753	2.131
16	1.746	2.120
17	1.740	2.110
18	1.734	2.101
19	1.729	2.093
20	1.725	2.086
21	1.721	2.080

Critical values of F for $\alpha = 0.025$

分母自由度	分子自由度	
	9	10
9	4.026	3.964
10	3.779	3.717

Critical values of F for $\alpha = 0.05$

分母自由度	分子自由度	
	3	4
15	3.287	3.056
16	3.239	3.007

本試題兩面印刷

試題完