

演習問題略解

[2章]

[1]

(1)

統計量	値
平均	56.4
標準誤差	1.5
中央値	56
最頻値	67
標準偏差	13.6
分散	184.1
尖度	-0.4
歪度	0.1
範囲	61
最小	28
最大	89
合計	4339
データの個数	77
変動係数	0.24

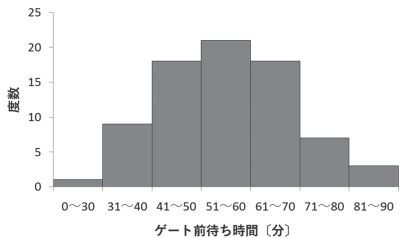
(2)

統計量	値
平均	131.67
標準誤差	2.96
中央値	134
最頻値	134
標準偏差	17.01
分散	289.29
尖度	0.07
歪度	-0.49
範囲	73
最小	92
最大	165
合計	4345
データの個数	33
変動係数	0.13

(3) 標準偏差を参照するとターミナルBのほうが大きい。しかし、両ターミナルの平均時間が大きく異なるため、変動係数で比較する必要がある。変動係数を比較すると、ターミナルAの待ち時間の変動が大きいことがわかる。

(4)

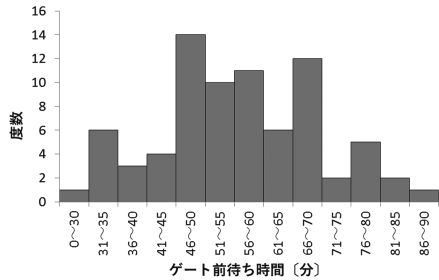
データ区間 [分]	度数	相対度数 [%]	累積相対度数 [%]
0~30	1	1.3%	1.3%
31~40	9	11.7%	13.0%
41~50	18	23.4%	36.4%
51~60	21	27.3%	63.6%
61~70	18	23.4%	87.0%
71~80	7	9.1%	96.10%
81~90	3	3.9%	100%
合計	77	100%	



2 演習問題略解

階級幅5分の場合

データ区間 【分】	度数	相対度数 【%】	累積相対度数 【%】
0~30	1	1.3%	1.30%
31~35	6	7.8%	9.09%
36~40	3	3.9%	12.99%
41~45	4	5.2%	18.18%
46~50	14	18.2%	36.36%
51~55	10	13.0%	49.35%
56~60	11	14.3%	63.64%
61~65	6	7.8%	71.43%
66~70	12	15.6%	87.01%
71~75	2	2.6%	89.61%
76~80	5	6.5%	96.10%
81~85	2	2.6%	98.70%
86~90	1	1.3%	100.00%
合計	77	100%	



(5) 歪度 (S_k) が $S_k = 0.1 > 0$ であるため、左偏りに近い分布であることがわかるが、 $S_k \approx 0$ であるため左右の歪みが小さく、正規分布に近い分布と考えられる (正規分布に従うか否かの判定は「正規性の検定」を実施する必要がある)。

【3章】

【1】

(1) 例えば、千葉県では**解表 3.1**のとおりとなる。

解表 3.1

都道府県名	万人あたり 死者数 【人/万人】	人あたり 台数 【台/人】	免許保有1人 あたり自動車 台数【台/人】	平均幅員 【m】	改良率 【%】
千葉県	0.296	0.533	0.842	6.401	60

(2) **解表 3.2**のとおり。

解表 3.2

相関係数	人あたり 台数 【台/人】	免許保有1人 あたり自動車 台数【台/人】	平均幅員 【m】	改良率 【%】
万人あたり 死者数	0.782	0.769	-0.526	-0.151

【2】

(1) F 境界値 (臨界値) $3.24 <$ 観測された分散比 (検定統計量) 43.04 となるため、交通安全教室により試験結果に差が出るといえる。

- (2) Tukey-Kramer の方法による検定統計量 (q 値) が解表 3.3 のようになり、臨界値 $q_{0.05, 4, 16}$ は付表の q 分布表より $2.86 (=4.046/\sqrt{2})$ となる。それより、「なしと実技のみ」、「なしと SS」、「座学のみと実技のみ」「座学のみと SS」の間に交通安全の試験結果の差があることが認められる。

解表 3.3

	座学のみ	実技のみ	SS
なし	-3.57	-8.66	-10.04
座学のみ		-5.09	-6.47
実技のみ			-83.1

[3]

- (1) 子ども 18 人, 成人 16 人, 高齢者 16 人。
 (2) それぞれについて p 値を求めると, 下記のとおりとなる。

年代 × 非常用持出袋の準備 $p = 0.7897$
 年代 × 家具等の転倒防止 $p = 0.0228$
 年代 × 復興支援活動 $p = 0.0027$

有意水準 5% で帰無仮説が棄却されるとすると, 年代と関連があるのは, 家具等の転倒防止, 復興支援活動の 2 点となる。

[4 章]

[1]

- (1) 「100 人あたり自動車保有台数」は多重共線性, 「高齢者割合」は F 値からモデルに適さない説明変数であることがわかる。その結果, 求められる重回帰式は, $y = 0.0223x_1 - 0.0552x_2 + 0.415x_3 - 28.8$ (y : 100 万人あたり死者数 [人], x_1 : 自動車走行台キロ [万台・km], x_2 : 取締り件数 [千件], x_3 : 事業所数 [千事業所]) となる。自由度修正済み決定係数が 0.753 であることから精度は良好であり, 分散分析表から有意なモデルであるといえる。
- (2) 標準偏回帰係数の値から, 自動車走行台キロ, 取締り件数, 事業所数の順に影響度が強いことがわかる。
- (3) (1) にて求められた重回帰モデル式に, 地域 O の 3 変数の値を代入すると, 100 万人あたり死者数は 37.0 [人] と推計される。

[5 章]

[1]

「労働力人口」と「人口密度」を説明変数とすると, p 値 = 0.0036, 判別の中率 = 11/12 となる。

「労働力人口」のみを説明変数とすると, p 値 = 0.1749, 判別の中率 = 7/12 となる。

「人口密度」のみを説明変数とすると, p 値 = 0.0075, 判別の中率 = 9/12 と

4 演習問題略解

なる。

以上より、労働力人口と人口密度を説明変数としたモデルが最も説明力が高いと考えられる。このモデルを用いると、ゾーン13に物流施設が立地する確率は34.9%と求められる。また、労働力人口の調整済みオッズ比は2.35となり、人口密度の調整済みオッズ比より高い。これより、労働力人口のほうが物流施設の立地に与える影響が大きいことがわかる。

[6章]

[1]

- (1) 判別モデル式は、 $z = 0.17x_1 + 1.82x_2 - 94.99$ (x_1 : 目的地までの距離 [m], x_2 : 駐車時間 [分]), 誤判別率は15.0%となる。
- (2) 標準化したデータから得られる判別係数を比較すると、駐車時間の影響度のほうが強いといえる。
- (3) マハラノビスの距離では誤判別率が10.0%となり、線形判別式より精度が高い。
- (4) 線形判別式、マハラノビスの距離のいずれにおいても「手数料の支払いはない」と予測される。

[7章]

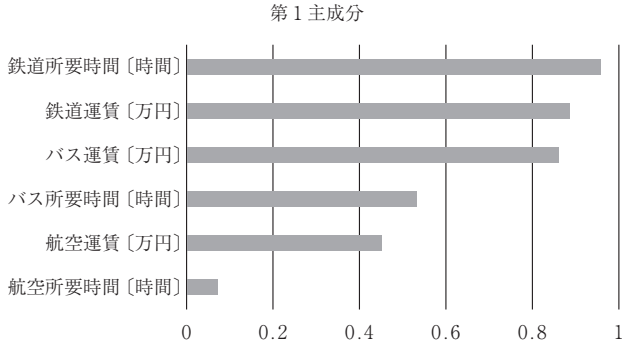
[1]

主成分分析の結果、**解表 7.1** のとおり寄与率等が得られる。第2主成分までで累積寄与率が7割を超えていること、また、第3主成分の固有値が1を下回り、標準化したデータの情報量を下回ることから第2主成分までを解釈することとする。

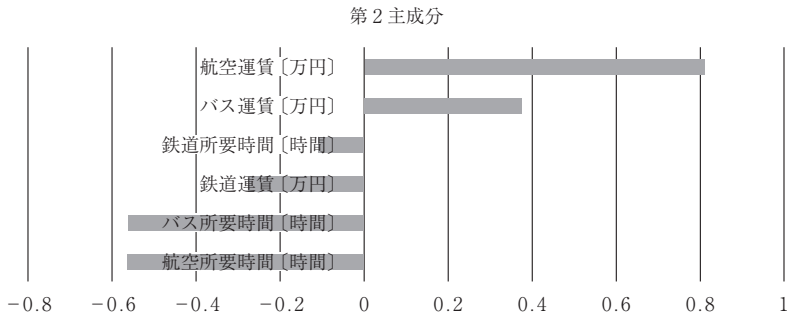
解表 7.1 固有値と寄与率

	第1主成分	第2主成分	第3主成分	第4主成分	第5主成分	第6主成分
分散 (固有値)	2.920	1.516	0.951	0.336	0.191	0.086
寄与率 [%]	48.7	25.3	15.8	5.6	3.2	1.4
累積寄与率 [%]	48.7	73.9	89.8	95.4	98.6	100.0

第1主成分と第2主成分の主成分負荷量を**解図 7.1**、**7.2**に示す。第1主成分に関しては鉄道とバスに関する移動の所要時間と費用に係る総合指標、第2主成分に関しては運賃と所要時間の差異に関する軸と解釈することができる。所要時間と比較して運賃が高い都市間や運賃と比較して所要時間が長い都市間を明らかにすることができる。



解図 7.1



解図 7.2

[8章]

[1]

- (1) 因子分析の結果，因子負荷量，共通性および寄与率は，解図 8.1 に示す結果となった。
- (2) 解図 8.1 に示す因子負荷量の結果を見ると，因子 1 については，Q1, Q2, Q3, Q9 および Q10 で値の大きな正の因子負荷量を示しており，速達性や定時性，自動車運転の選好などに関連する変数であることから，例えば「移動の効率性重視」因子と解釈できる。つぎに，共通因子 2 については，比較的，Q4, Q5 および Q6 で値の大きな正の因子負荷量を示しており，いずれも交通安全や交通情報に関連した変数であるため，例えば「移動の確実性重視」因子と解釈できる。つぎに，各変数の共通性の値を確認してみると，Q7 を除く変数が 0.6 以上の高い共通性の値を示しており，該当する変数の 6 割以上が共通因子のみで説明されることがわかる。そして，各因子の寄与率を確認すると，因子 1 の寄与率は

因子負荷量			
変数名	因子1	因子2	共通性
Q1	0.90	0.17	0.84
Q2	0.86	0.26	0.81
Q3	0.87	0.23	0.81
Q4	-0.83	0.56	1.00
Q5	-0.78	0.46	0.81
Q6	-0.83	0.36	0.82
Q7	0.21	0.21	0.09
Q8	-0.83	-0.14	0.72
Q9	0.86	0.19	0.78
Q10	0.73	0.30	0.62

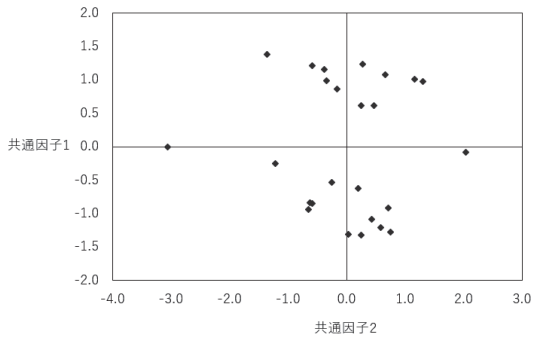
二乗和	6.33	0.98
寄与率	0.63	0.10

解図 8.1 因子負荷量, 共通性および寄与率

約 63%, 因子 2 の寄与率は約 10% で, 累積寄与率については約 73% であることから, 共通因子が観測データのおよそ七割程度の情報を説明できていることがわかる。

(3) 計算した因子得点とその散布図を解図 8.2 に示す。

因子得点		
	因子1	因子2
1	-1.08	0.43
2	-0.52	-0.24
3	-1.32	0.25
4	0.62	0.47
5	-0.84	-0.59
6	1.08	0.66
7	1.24	0.28
8	-0.91	0.72
9	1.21	-0.58
10	-0.62	0.20
11	-0.25	-1.21
12	-1.20	0.58
13	-0.93	-0.64
14	0.63	0.26
15	0.00	-3.06
16	0.99	-0.34
17	-1.27	0.75
18	1.01	1.17
19	-1.31	0.04
20	-0.84	-0.62
21	0.87	-0.16
22	0.98	1.31
23	-0.08	2.04
24	1.16	-0.38
25	1.38	-1.36

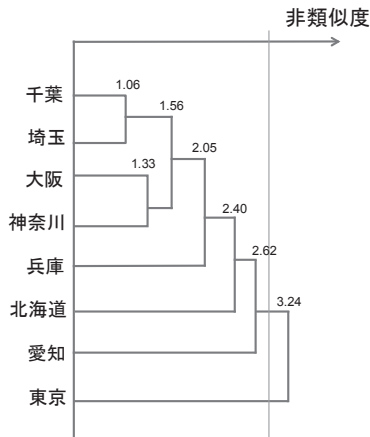


解図 8.2 因子得点の算出結果と散布図

[9章]

[1]

非類似度を横軸とすると、デンドログラムは解図 9.1 のように描くことができる。非類似度は東京とそれ以外の非類似度が最大である（ただし、千葉・埼玉間の非類似度を除く）。したがって、クラスター間の非類似度を参考にすると、{東京}{千葉, 埼玉, 大阪, 神奈川, 兵庫, 北海道, 愛知}のクラスターを形成することができる。



解図 9.1

[10章]

[1]

解表 10.1 のように、季節アイテム、長崎港アイテムのレンジはそれぞれ 440.1, 234.7 となる。したがって、季節アイテムが目的変数（外的基準）に与える影響としては最も大きいことがわかった。

解表 10.1 分析結果のまとめ

アイテム	カテゴリー	n	カテゴリー・ウェイト	レンジ
季節	春	14	937.8	440.1
	夏	14	1235.5	
	冬	14	795.4	
長崎港への寄港の有無	有	14	234.7	234.7
	無	14	0	

* 重相関係数 $R_2 = 0.91$

【2】

解表 10.2 のように、年代アイテム、年収アイテムのレンジはそれぞれ 1.46, 1.40 となる。ほぼ値は変わらないものの、年代アイテムのレンジがやや大きく、クルーズ船利用意向に与える影響としては最も大きいことがわかる。

解表 10.2 分析結果のまとめ

アイテム	カテゴリー	n	カテゴリーウェイト	レンジ
年代	20～30代	4	1.46	1.46
	40～50代	8	0.58	
	60代以上	6	0	
年収	1000万円未満	9	1.40	1.40
	1000万円以上	9	0	

* 相関比 $\eta^2 = 0.75$

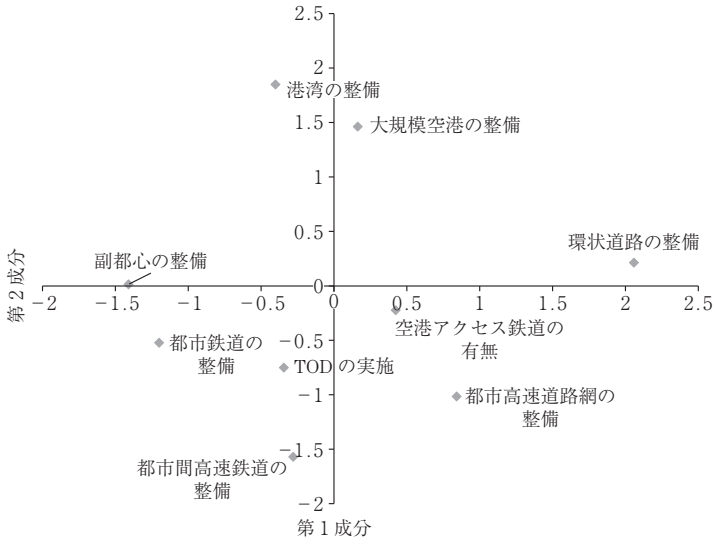
【3】

第1成分と第2成分の固有値と相関係数を解表 10.3 に取りまとめる。

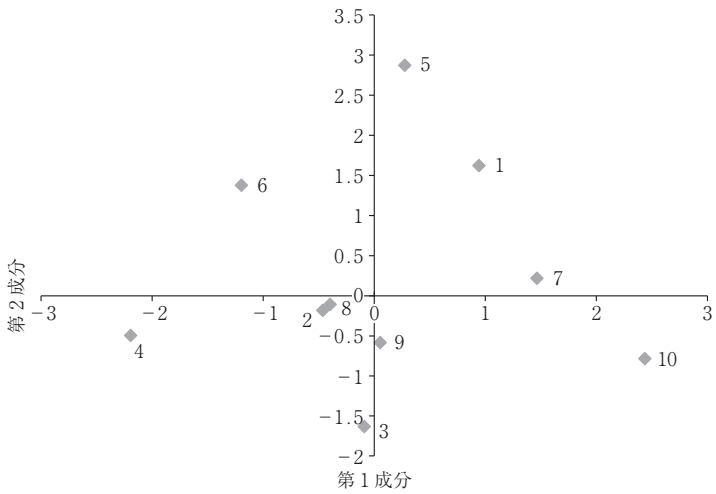
解表 10.3 固有値・相関係数の推定結果

成分(軸)	固有値	相関係数
1	0.3541	0.5951
2	0.2593	0.5092

つぎに、カテゴリスコアとサンプルスコアの散布図を解図 10.1, 解図 10.2 に示す。第1成分は環状道路の整備や都市高速道路の整備など道路系の整備が進んでいる都市がプラスに、副都心整備や都市鉄道を整備などの整備が進んでいる都市がマイナスに位置付けられるといえる。また、第2成分では港湾や空港などの拠点整備が進んでいる都市がプラスに、都市間交通の整備が進んでいる都市がマイナスに位置付けられているといえる。



解図 10.1 カテゴリスコア



解図 10.2 サンプルスコア