

2001/2/15

## マーケティング分析における多変量解析の適用について

杉浦システムコンサルタント事務所  
システムアナリスト 杉浦 司

### 【多変量解析とは何か】

多変量解析はひとつの変数を複数の変数で予測・説明・判別するための統計解析手法である。マーケティングにおいては、広告や陳列場所の変更、値下げ、商品デザインの変更といった経営アクションが来客や売上といった目標にどのような効果を与えるのか、あるいは企画中の新製品が機能、デザイン、価格、販売チャネルなどの観点から自社あるいは同業他社製品と比べて顧客は同じあるいは異なるイメージを抱くのかといったマーケティングプログラムの策定時におけるクリティカルな意思決定を支援する情報を与えてくれる可能性を持つものであるといえる。

多変量解析においてはまず、その特性について分析される側の対象データを基準変数と呼び、基準変数の特性を決定する側の要素データを説明変数と呼ぶ。そして、基準変数と説明変数の性質、変数を測定するための尺度が量的な値か質的な値かによって、適用すべき多変量解析の種類が異なってくる。基準変数と説明変数がともに量的な値の場合には重回帰分析が、基準変数が量的な値で説明変数が質的な値の場合には数値化 類が、基準変数が質的な値で説明変数が量的な値の場合には判別分析が、そして両方とも質的な値の場合には数値化 類が適用される。

### 【多変量解析実施上の留意事項】

多変量解析を実施するには以下の事項について留意しなければならない。

#### 1. 変数間の均一化

相関分析など説明変数同士の影響度などを分析する場合、変数間のデータの質及びレベルを統一、均一化することが必要となる。量的なデータと質的なデータが混在する場合は精度の低い方である質的なデータに統一することが必要となる。また、特定の説明変数の値だけが他の変数と比べて大きい場合、基準変数がこの変数のみに影響されることになるため、重みを揃えて変数間における影響度の不公平をなくす工夫が必要である。なお、データの重みを揃える方法としては一般に、以下の式を適用する。

$$\frac{\text{あるデータの値} - \text{最小値}}{\text{最大値} - \text{最小値}} \times 10$$

#### 2. 観察数と変数の関係

Excel などコンピュータソフトを利用する場合、観察数が分析で使用する変数よりも少ない場合でも分析結果が得られるがその結果は信頼できるものではないので注意が必要である。観察数は変数より必ず多く設定する必要がある。

#### 3. 説明変数間の相関関係

説明変数を選択する場合、変数間で関連性のあるもの（正の相関関係が高い）や対立関

係にあるもの（負の相関関係が高い）を排除する必要がある。このような変数を使用した場合、誤まった高い予測精度を得てしまう危険がある。それぞれの変数は変数間において独立したものであり、基準変数にのみ関係のあるものを選び出すことが重要である。基準変数に強い関係性を見出すことができ、説明変数間に強い関係性を見出すことができない最適な説明変数セットを探し出すためには、基準変数と説明変数間、説明変数同志の単相関分析あるいは分散分析を実施して丹念に調べていくことが不可欠である。

以下、多変量解析を Excel2000 の分析ツールを使った適用方法として重回帰分析、数値化理論、判別分析について考察した。

### 【重回帰分析による多変量解析】

以下、重回帰分析による多変量解析の方法について考察するためにチェーンストアの場所別売上実績データを取り上げる。なお、表中の乗降客数は店舗の最寄駅での乗降客数を示している。

チェーンストア売上							
場所	売上高(百万円)	乗降客数(万人)	店の広さ(百平米)	駐車台数(台)	平均年収(万円)	取扱品目数(品目)	周辺人口(万人)
小田原	272	245	59	60	468	290	22.5
秦野	161	118	32	35	472	146	19.2
伊勢原	129	142	25	30	475	101	13.5
本厚木	252	249	55	45	515	224	22.6
海老名	204	174	46	40	460	172	17.3
藤沢	168	202	32	35	465	85	22.0
大和	242	254	54	45	520	172	23.8
相模大野	169	168	32	40	459	152	17.4
町田	224	224	42	50	501	158	20.3
新百合ヶ丘	202	186	45	45	783	168	16.7
成城学園前	259	212	56	50	515	285	24.5
経堂	165	145	32	30	476	103	12.3
下北沢	180	174	31	35	492	106	17.2
梅ヶ丘	131	82	38	30	452	162	9.8
代々木上原	215	177	34	40	452	192	18.6

出典 『Excel統計解析-回帰分析-』長谷川勝也著 共立出版

#### 1. 変数間の均一化

まず多変量解析のための準備作業として変数間の均一化を行っておく。各変数ごとの最大値と最小値を調べてデータの重みを揃える。

場所	売上高 (百万円)	乗降客数 (万人)	店の広さ (百平米)	駐車台数 (台)	平均年収 (万円)	取扱品目数 (品目)	周辺人口 (万人)
最小値	129	82	25	30	452	85	9.8
最大値	272	254	59	60	783	290	24.5

### チェーンストア売上

場所	売上高 (百万円)	乗降客数 (万人)	店の広さ (百平米)	駐車台数 (台)	平均年収 (万円)	取扱品目数 (品目)	周辺人口 (万人)
小田原	272	10	245	9	59	10	60
秦野	161	2	118	2	32	2	35
伊勢原	129	0	142	3	25	0	30
本厚木	252	9	249	10	55	9	45
海老名	204	5	174	5	46	6	40
藤沢	168	3	202	7	32	2	35
大和	242	8	254	10	54	9	45
相模大野	169	3	168	5	32	2	40
町田	224	7	224	8	42	5	50
新百合ヶ丘	202	5	186	6	45	6	45
成城学園前	259	9	212	8	56	9	50
経堂	165	3	145	4	32	2	30
下北沢	180	4	174	5	31	2	35
梅ヶ丘	131	0	82	0	38	4	30
代々木上原	215	6	177	6	34	3	40
(最小値)	129	82	25	30	452	85	9.8
(最大値)	272	254	59	60	783	290	24.5

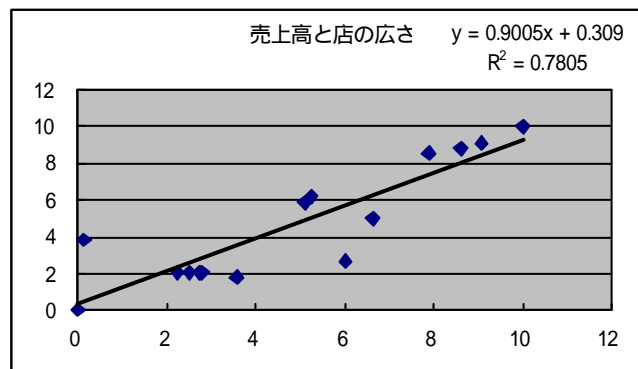
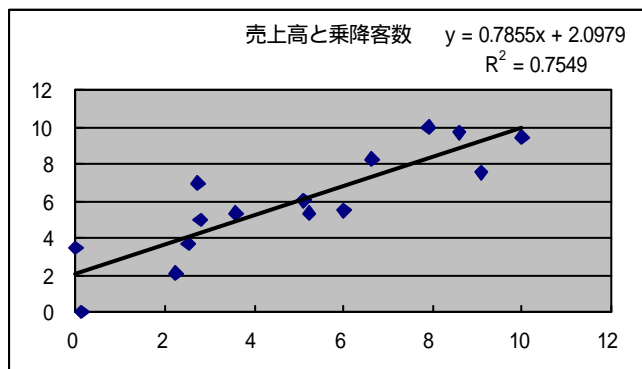
## 2. 観察数と変数の関係

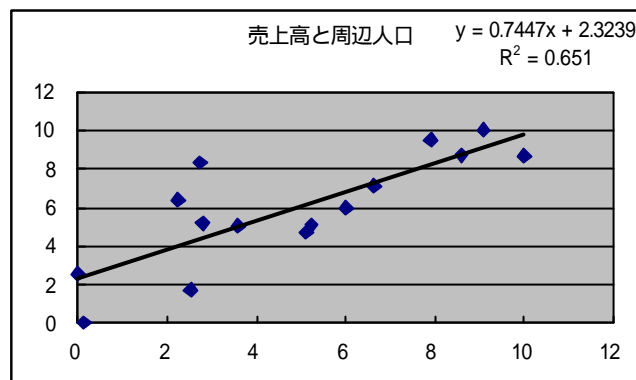
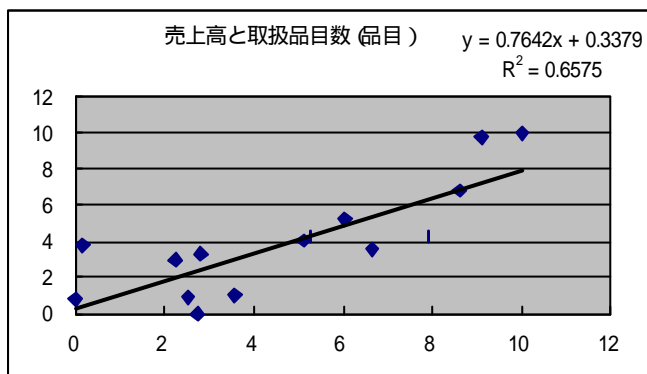
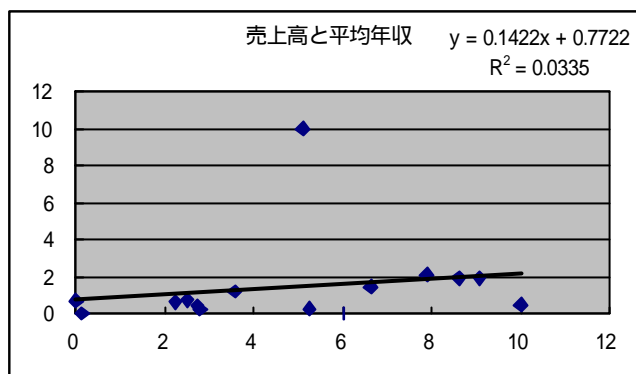
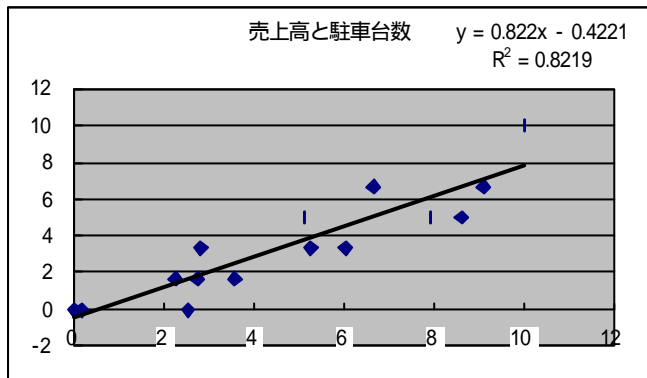
基準変数を売上高として観察数は 15、説明変数は 6 のため観察数>変数であり観察数と変数との間の関係に問題はない。

## 3. 説明変数間の相関関係

### (1) 単相関による基準変数と説明変数間の関係性分析

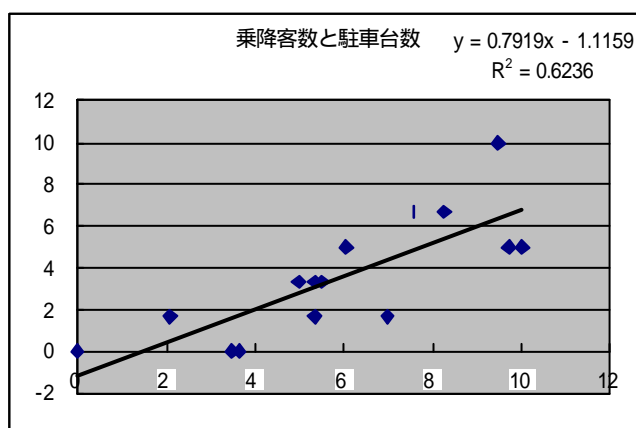
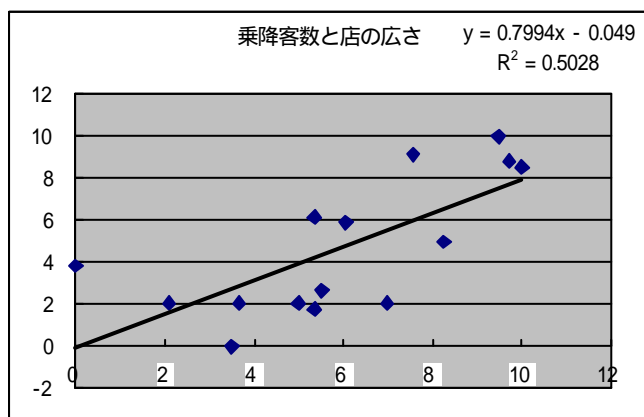
以下、売上高と各説明変数との間の単相関を調べたところ、周辺住民の平均年収以外は高い正の相関関係があることが判明した。

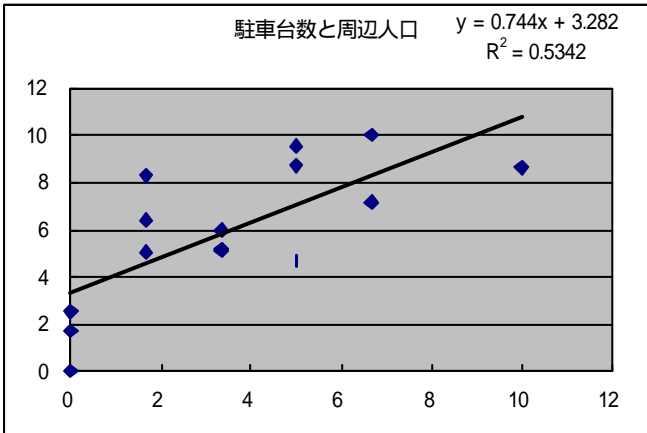
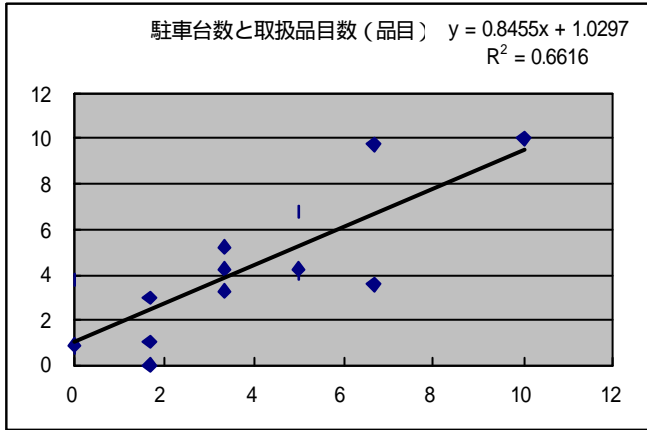
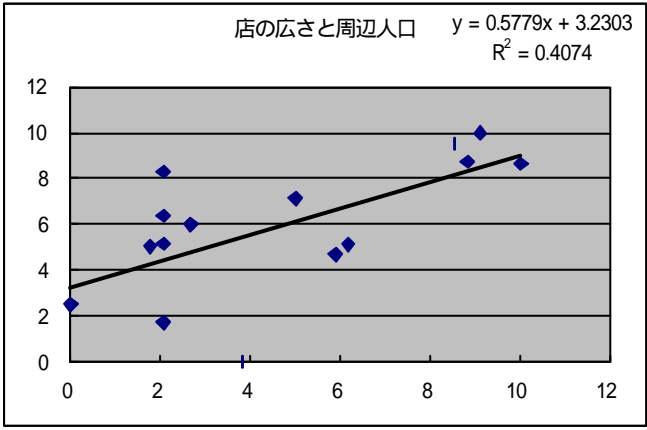
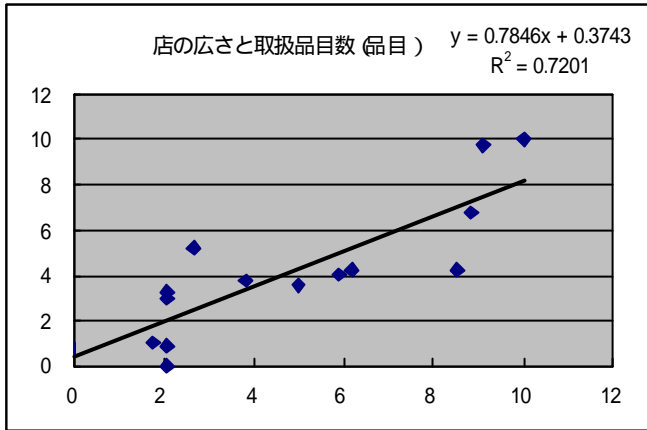
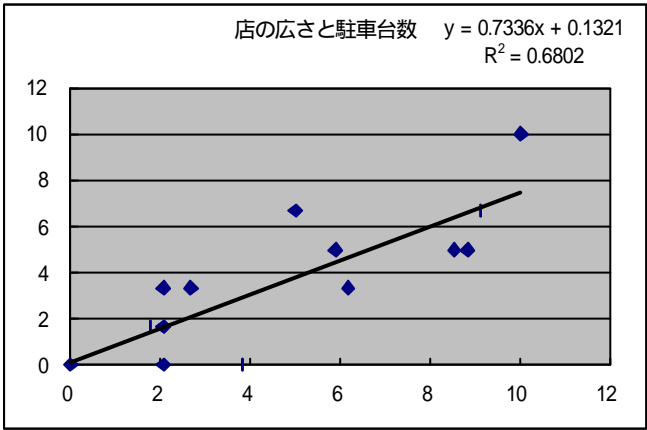
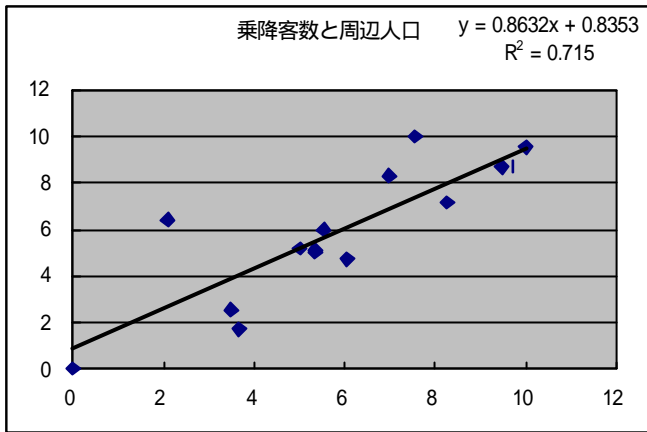
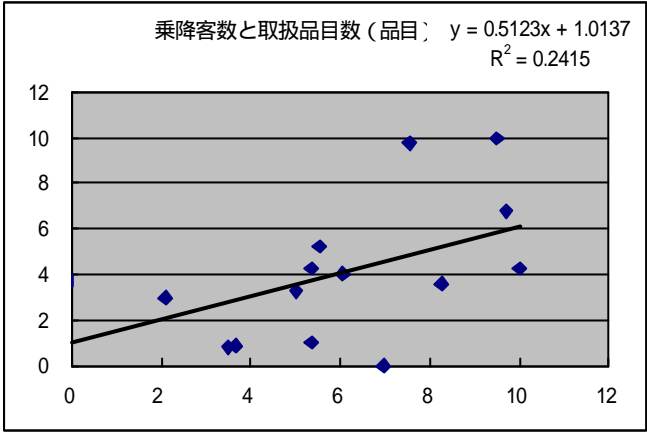




## (2) 単相関による説明変数間の関係性分析

以下、各説明変数間の単相関を調べたところ、売上高に正の相関関係のあった最寄駅での乗降客数と取扱品目数(品目)との間に強い相関関係がなく、独立な説明変数として採用すべきものと考えられることがわかった。また、売上高と極めて高い正の相関関係を示していた店の広さと駐車台数はともに正の相関関係があり、他の変数とも正の相関関係を示していることから、売上高との相関係数が高い駐車台数を説明変数として採用すべきものと考えた。





#### 4.重回帰分析の実施

基準変数に売上高、説明変数に最寄駅での乗降客数、駐車台数、取扱品目数として以下のとおり重回帰分析を実行した。回帰式の近似の度合いを示す R<sup>2</sup> 値は 0.949 と高いレベルであり、このことから、売上高は下記の回帰式によって説明あるいは予測できるといえる。

$$\text{売上高 (百万円)} = 28.54 + 0.55 \times \text{最寄駅での乗降客数 (万人)} + 0.2 \times \text{駐車台数 (台)} + 0.36 \times \text{取扱品目数 (品目)}$$

場所	売上高 (百万円)	乗降客数 (万人)	駐車台数 (台)	取扱品目数 (品目)
小田原	272	245	60	290
秦野	161	118	35	146
伊勢原	129	142	30	101
本厚木	252	249	45	224
海老名	204	174	40	172
藤沢	168	202	35	85
大和	242	254	45	172
相模大野	169	168	40	152
町田	224	224	50	158
新百合ヶ丘	202	186	45	168
成城学園前	259	212	50	285
経堂	165	145	30	103
下北沢	180	174	35	106
梅ヶ丘	131	82	30	162
代々木上原	215	177	40	192

#### 概要

回帰統計	
重相関 R	0.97434289
重決定 R <sup>2</sup>	0.94934407
補正 R <sup>2</sup>	0.93552881
標準誤差	11.5240307
観測数	15

#### 分散分析表

	自由度	変動	分散	調整された分散	有意 F
回帰	3	27377.5639	9125.85463	68.71709	2.07E-07
残差	11	1460.83611	132.803283		
合計	14	28838.4			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	28.5409165	16.1378989	1.76856459	0.104651	-6.97838	64.06021	-6.97838	64.06021
乗降客数 (万人)	0.55209428	0.11221818	4.91982931	0.000457	0.305104	0.799085	0.305104	0.799085
駐車台数 (台)	0.20194949	0.96046139	0.210263	0.837306	-1.91201	2.315912	-1.91201	2.315912
取扱品目数 (品目)	0.35863816	0.09526135	3.76478153	0.003129	0.148969	0.568307	0.148969	0.568307

#### 4. 重回帰分析結果の検討

重回帰分析結果の t 値に着目すると最寄駅での乗降客数（万人）4.9、取扱品目数（品目）3.76 と非常に大きい。駐車台数（台）はこの二つと比べて 0.21 と小さい値となっている。

t 値が高いほど基準変数を説明・予測する上での貢献度が高いと考えられるため、売上高は最寄の交通機関の利用者数すなわち交通の便に大きく影響しており、次に取扱品目数の多さ、すなわち品揃えのよさに影響していると思われる。そして、交通の便、品揃えとも同条件の場合、駐車場の大きさが売上に影響すると考えられよう。

#### 【数値化理論 による多変量解析】

次に数値化理論 による多変量解析の方法について考察する。ここではビールの売上と広告効果について分析する方法について考察した。

#### ビールの売上と広告効果

日付	8/1月	8/2火	8/3水	8/4木	8/5金	8/6土	8/7日
一日当たりの売上数量	412	741	1021	666	447	678	1677
チラシ広告	×	×			×		
ケース前宣伝	×		×	×		×	×
温度	30	32	32	28	28	26	33
日付	8/8月	8/9火	8/10水	8/11木	8/12金	8/13土	8/14日
一日当たりの売上数量	319	218	263	1335	942	916	732
チラシ広告	×	×	×	×			×
ケース前宣伝		×	×	×		×	×
温度	27	26	26	34	29	30	30
日付	8/15月	8/16火	8/17水	8/18木	8/19金	8/20土	8/21日
一日当たりの売上数量	770	552	263	268	508	811	283
チラシ広告		×	×	×			
ケース前宣伝	×		×	×	×	×	
温度	31	30	26	25	25	28	30

出典 『事例でわかるパソコンデータ分析入門重回帰分析編』小幡 卓、堀合 啓一著 技術評論

#### 1. ダミー変数による数値化

曜日とチラシ広告とケース前宣伝の効果を説明変数化するために 0 と 1 のダミー変数で値を数値化した。なお、曜日についてはランク落ちによる逆行列計算で Excel がエラーを起こさないよう月曜日の列を削除している。

一日当たりの 売上数量	日付	火	水	木	金	土	日	チラシ 広告	ケース 前宣伝	温度
412	8月1日	0	0	0	0	0	0	0	0	30
741	8月2日	1	0	0	0	0	0	0	1	32
1021	8月3日	0	1	0	0	0	0	1	0	32
666	8月4日	0	0	1	0	0	0	1	0	28
447	8月5日	0	0	0	1	0	0	0	1	28
678	8月6日	0	0	0	0	1	0	1	0	26
1677	8月7日	0	0	0	0	0	1	1	0	33
319	8月8日	0	0	0	0	0	0	0	1	27
218	8月9日	1	0	0	0	0	0	0	0	26
263	8月10日	0	1	0	0	0	0	0	0	26
1335	8月11日	0	0	1	0	0	0	0	0	34
942	8月12日	0	0	0	1	0	0	1	1	29
916	8月13日	0	0	0	0	1	0	1	0	30
732	8月14日	0	0	0	0	0	1	0	0	30
770	8月15日	0	0	0	0	0	0	1	0	31
552	8月16日	1	0	0	0	0	0	0	1	30
263	8月17日	0	1	0	0	0	0	0	0	26
268	8月18日	0	0	1	0	0	0	0	0	25
508	8月19日	0	0	0	1	0	0	1	0	25
811	8月20日	0	0	0	0	1	0	1	0	28
283	8月21日	0	0	0	0	0	1	1	1	30

## 2. 回帰分析（数値化理論）の実施

基準変数に一日当たりの売上数量、説明変数に火曜日、水曜日、木曜日、金曜日、土曜日、日曜日、チラシ広告、ケース前宣伝、温度として以下のとおり回帰分析（数値化理論）を実行した。回帰式の近似の度合いを示す  $R^2$  値は 0.81 となっており有効性を示している。回帰分析の結果から、各曜日、チラシ広告、ケース前宣伝、温度から説明できる一日当たりの売上数量の予測のための回帰式は下記のとおりとなる。

$$\begin{aligned}
 \text{一日当たりの売上数量} = & -2975 + 0 \times \text{月曜日} + 118 \times \text{火曜日} + 97 \times \text{水曜日} + 217 \times \text{木曜日} \\
 & + 415 \times \text{金曜日} + 311 \times \text{土曜日} + 161 \times \text{日曜日} \\
 & + 107 \times \text{チラシ広告} + -236 \times \text{ケース前宣伝} + 120 \times \text{温度}
 \end{aligned}$$



回帰統計	
重相関 R	0.902014
重決定 R2	0.81363
補正 R2	0.661145
標準誤差	220.7629
観測数	21

#### 分散分析表

	自由度	変動	分散	F値	有意 F
回帰	9	2340430	260047.8	5.335819	0.005791
残差	11	536098.8	48736.26		
合計	20	2876529			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	-2974.97	627.8087	-4.73866	0.000611	-4356.77	-1593.17	-4356.77	-1593.17
火	117.7526	189.1972	0.62238	0.546377	-298.668	534.173	-298.668	534.173
水	96.52729	187.0462	0.516061	0.616026	-315.159	508.2134	-315.159	508.2134
木	217.2504	185.8696	1.168832	0.267178	-191.846	626.3469	-191.846	626.3469
金	414.9295	201.1858	2.062419	0.063599	-27.8777	857.7368	-27.8777	857.7368
土	311.1504	205.5251	1.513929	0.158235	-141.208	763.5083	-141.208	763.5083
日	161.4056	186.3516	0.866135	0.404912	-248.752	571.5629	-248.752	571.5629
チラシ広告	107.0654	126.4925	0.846417	0.415355	-171.343	385.4737	-171.343	385.4737
ケース前宣伝	-236.192	137.9945	-1.71161	0.114982	-539.916	67.53149	-539.916	67.53149
温度	119.9436	21.49808	5.579268	0.000165	72.62659	167.2606	72.62659	167.2606

### 3. 回帰分析（数値化理論）結果の検討

回帰分析結果の t 値に着目すると温度 21.5、金曜日 2、土曜日 1.5、木曜日 1.2、日曜日 0.86、チラシ広告 0.8 あたりの貢献度が大きくなっている。ケース前宣伝は -1.7 であり、あまり効果がでていないと思えない。

回帰分析の結果からは、ビールの日当たりの売上数量は気温が高い日、また木、金、土、日曜日に、そしてチラシ広告を打った日によく売れていると考えられる。

#### 【判別分析による多変量解析】

判別分析はいくつかのグループにあらかじめデータがある場合に、それらのグループの特徴をあらわす説明変数を使って、新しいデータがどのグループに属するかを判別するための手法である。判別分析は基準変数が質的な値で説明変数が量的な値の場合の多変量解析の方法である。以下、新商品のブランドイメージについて判別分析する事例について考察した。

	人気度	認知度	所有率	高級感	誇らしさ	品質の信頼性	センスのよさ	親しみやすさ	広告が魅力的
シャネル	159	377	209	318	136	150	123	36	86
エルメス	145	327	136	245	104	154	127	27	41
ティファニー	145	327	136	182	86	136	136	77	59
ルイ・ヴィトン	136	359	186	177	77	186	82	109	18
グッチ	123	350	154	163	73	141	114	68	32
ラルフローレン	114	295	200	54	27	114	91	154	36
カルティエ	109	291	109	232	95	150	95	14	23
フェラガモ	109	286	68	159	64	109	77	32	18
ブラダ	104	245	45	104	50	77	82	59	18
カルバン・クライン	100	263	123	32	23	64	118	132	54
ベネトン	86	327	241	18	5	54	59	227	95

454人（複数回答）

出典 96年8月31日の日経流通新聞 首都圏在住の15歳以上の女性モニター454人に聞いたアンケート調査データ

### 1. 基準変数の質的な値への変換

人気度の高いブランドを判別するために人気度が120点以上のブランドを1、100点以下を0とし、101点以上119点以下の人気度があいまいな製品を削除した。また、ティファニーは判別式の検証用に抜いておく。

	人気度	認知度	所有率	高級感	誇らしさ	品質の信頼性	センスのよさ	親しみやすさ	広告が魅力的
シャネル	1	377	209	318	136	150	123	36	86
エルメス	1	327	136	245	104	154	127	27	41
ルイ・ヴィトン	1	359	186	177	77	186	82	109	18
グッチ	1	350	154	163	73	141	114	68	32
カルバン・クライン	0	263	123	32	23	64	118	132	54
ベネトン	0	327	241	18	5	54	59	227	95

ティファニー	145	327	136	182	86	136	136	77	59
--------	-----	-----	-----	-----	----	-----	-----	----	----

### 2. 観察数と変数の関係

観察数が6、説明変数が9であるため、観察数<変数となり回帰分析ができない。人気度との相関係数の大きい変数を選択する必要がある。ここでは認知度、高級感、誇らしさ、

品質の信頼性の四つの変数を選択する。

	人気度	認知度	所有率	高級感	誇らしさ	品質の信頼性	センスのよさ	親しみやすさ	広告が魅力的
人気度	1								
認知度	0.758246	1							
所有率	-0.12232	0.544737	1						
高級感	0.882299	0.718827	-0.02011	1					
誇らしさ	0.880045	0.680616	-0.08224	0.996703	1				
品質の信頼性	0.956728	0.707582	-0.12072	0.810804	0.811216	1			
センスのよさ	0.436918	-0.07663	-0.71347	0.576694	0.627465	0.317592	1		
親しみやすさ	-0.82934	-0.38547	0.483261	-0.87825	-0.90507	-0.73874	-0.85827	1	
広告が魅力的	-0.51205	-0.01184	0.630871	-0.1421	-0.17559	-0.61938	-0.23241	0.395456	1

### 3. 回帰分析（判別分析）の実施

基準変数に人気度、説明変数に認知度、高級感、誇らしさ、品質の信頼性として以下のとおり回帰分析（判別分析）を実行した。回帰分析の結果から、認知度、高級感、誇らしさ、品質の信頼性から説明できる人気度の判別のための回帰式は下記のとおりとなる。

$$\text{人気度} = -0.85 + 0.001 \times \text{認知度} + 0.0008 \times \text{高級感} + 0.047 \times \text{誇らしさ} + 0.006 \times \text{品質の信頼性}$$

回帰統計	
重相関 R	0.975782
重決定 R2	0.952151
補正 R2	0.760754
標準誤差	0.252585
観測数	6

### 分散分析表

	自由度	変動	分散	割られた分散	有意 F
回帰	4	1.269534	0.317384	4.974746	0.322884
残差	1	0.063799	0.063799		
合計	5	1.333333			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	-0.85089	1.760939	-0.4832	0.713446	-23.2256	21.52387	-23.2256	21.52387
認知度	0.001564	0.005944	0.263168	0.836177	-0.07397	0.077095	-0.07397	0.077095
高級感	-0.0008	0.016906	-0.04706	0.970063	-0.21561	0.214014	-0.21561	0.214014
誇らしさ	0.0047	0.039333	0.119492	0.924288	-0.49507	0.50447	-0.49507	0.50447
品質の信頼性	0.006362	0.004066	1.564662	0.362037	-0.04531	0.05803	-0.04531	0.05803

### 4. 回帰分析（判別分析）の検討

回帰分析によって得られた判別式を利用して検証用に抜いておいたティファニーのデー

タを判別してみる。人気度は 0 か 1 のため、判別式の値が 0.5 より大きければ人気ブランドであると判別することができる。判別式の結果は 0.5516 となるためティファニーは人気ブランドのグループに属すると判別できる。

	人気度	認知度	高級感	誇らしさ	品質の信頼性
ティファニー	145	327	182	86	136

ティファニーの人気度

$$0.5516 = -0.85 + 0.001 \times 327 + 0.0008 \times 182 + 0.0047 \times 86 + 0.006 \times 136$$

### 【多変量解析実施上の留意点】

以上、考察してきた多変量解析の手法についてはその適用において説明変数の候補から最も適した変数を選択することが分析結果の信頼性を大きく左右することがわかる。しかし、説明変数の候補が多くなればなるほど単相関などによる総当たり方式によるチェックは困難となってしまう。多変量解析における説明変数の効率的な選択方法については逐次変数選択法や対話型変数選択法、説明変数選択基準（上田太一郎氏など）などが提唱されているが、絶対的な手法は存在しない。逐次変数選択法には変数増加法、変数減少法、変数増減法、変数減増法の四つがあるが何れも説明変数の足し引きにおける寄与率  $R^2$  値の変化をみて最適な変数組み合わせを探し出そうというものである。説明変数選択基準も寄与率評価の変形であり、逐次変数選択法と基本的な考え方は同じである。対話型変数選択法はコンピュータと対話しながら変数減増法あるいは変数減増法を行う方法である。現在のところ、コンピュータと対話しながら試行錯誤的に変数減増法あるいは変数減増法を行う対話型変数選択法が妥当な方法といえるだろう。将来的にはニューラルネットワークあるいは複雑系の理論を応用した人工知能による説明変数の探索が実現されることを期待したい。

以上