

大型小売店における売上予測（第2報）

橋本 郁郎

Sales Forecast in a Large Size Retail Store (the 2nd Report)

Ikuro HASHIMOTO

A sales forecasting is very important for making a management plan and proceeding to research development and product planing. There are various forecasts such as technical forecast, economic forecast, sales forecast and product forecast. Succeeding to the previous report, the present study discuss further sales forecast of clothing in the department stores in Aichi Prefecture by means of a multiple regression analysis. In the previous report, such dependent variables were used as consumer expense, sales floor area, time and consumer price index. The present study added a dummy variable to the dependent variables for the improvement of autocorrelation of residual. As the result of the analysis, the autocorrelation of the residual are improved very much in which the forecast error rates are 7.13 % max. and 0.36 % min.

1. はじめに

現代社会における企業にとって、経営計画を立てる場合や研究開発・製品計画を実施するときに、予測が非常に重要なものとなっている。前報¹⁾では大型小売店のうち愛知県内の百貨店の衣料品の総売上上の予測を行なったが、本研究ではその予測手順のうちの重回帰分析にダミー変数を導入して解析を試みた。

2. 研究方法

昭和54年より61年までの月次データ²⁾により、昭和62年の愛知県内の百貨店の衣料品の総売上高を予測することである。前報で述べた予測手順³⁾のうちで本研究において異なる点は、不規則変動を除くための重回帰分析においてそのモデル式を改善するために、説明変数にダミー変数を追加導入したことである。

3. 回帰モデル式の構築

前報の重回帰分析において、そのモデル式の正当性を見るために残差を検討したところ、その時系列分布は必ずしもランダムとは言えず、回帰式に多少

問題があることが判明した。そこで本報ではそれらを改善する方法の一つとして、説明変数にダミー変数を追加導入して解析することにした。以下の統計計算はSAS⁴⁾⁵⁾ (Statistical Analysis System) によって行なった。

3・1 ダミー変数の導入

季節調整済の売上高を図1に示す。この図を見ると全体を1つの回帰式で行なうよりも、前半と後半を別々に回帰した方がより良いように考えられる。

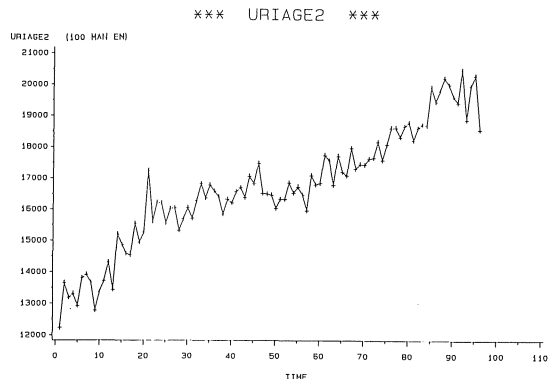


図1 季節調整済み売上データ

表1 ダミー変数の種類

種 別	ダミー変数	
	0	1
(a) DUM1	0~40	41~96
(b) DUM2	0~45	46~96
(c) DUM3	0~50	51~96
(d) DUM4	0~55	56~96
(e) DUM5	0~60	61~96

そこでダミー変数を導入することにより、1つの回帰式でそれを実施することにした。どの時点にダミー変数を導入したら最適かを探るために、図1においてTIME40, 45, 50, 55, 60の5ヶ所を選んで表1の如くダミー変数を導入し検討を行なった。

3・2 最適回帰式の選定⁶⁾

前報のダミー変数なしの重回帰分析の結果を表2に示す。これを見るとパラメータの推定値がゼロという仮説検定の有意水準確率である $PROB > |T|$ の値が、変数 SHISU では0.4172と大きく信頼性に問題があるのでこの変数は除いてダミー変数を追加し、(1)式の如き回帰式により検討を進めた。

$$Y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 + a_5x_5 + a_6x_6 + a_7x_7 \dots (1)$$

Y : 衣料品の総売上 (季調済み) URIAGE 2
 x_1 : 消費支出 (季調済み) SHISYU 2
 x_2 : 売場面積 MENSEKI
 x_3 : 時間 TIME
 x_4 : ダミー変数 DUMn
 x_5 : $DnSYU = DUMn * SHISYU 2$
 x_6 : $DnMEN = DUMn * MENSEKI$
 x_7 : $DnTIM = DUMn * TIME$

但し $n = 1, 2, 3, 4, 5$ とする

(1)式の重回帰分析の結果を表3に示す。表中の(a)(b)(c)(d)(e)はそれぞれ $n = 1, 2, 3, 4, 5$ に対応している。この表3において [ダミー変数] × [説明変数] の $PROB > |T|$ の大きいものは、あまり意味を持たないのでそれらを除いて再び重回帰分析を行なった。その結果を表4に示す。この中で全ての説明変数の $PROB > |T|$ の値が0.05以下のものは(e)のダミー変数がDUM5の場合のみであるので、これを最適モデルと決定する。最適モデル式を(2)式に示す。

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_6x_6 \dots (2)$$

Y, x_1, x_2, x_3 は(1)式と同じ

x_4 : DUM 5

x_6 : DUM 5 * MENSEKI

(2)式の重回帰分析の結果を表5に示す

4. 予測結果

昭和62年度の売上を予測するには、本来説明変数

表2 ダミー変数なしの重回帰分析結果¹⁾

DEP VARIABLE: URIAGE2		SAS					
		ANALYSIS OF VARIANCE					
		SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
		MODEL	4	302829563	75707390.79	207.818	0.0001
		ERROR	91	33150949.81	364296.15		
		C TOTAL	95	335980513			
		ROOT MSE		603.5695	R-SQUARE	0.9013	
		DEP MEAN		16687.27	ADJ R-SQ	0.8970	
		C.V.		3.616946			
PARAMETER ESTIMATES							
VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > T		
INTERCEP	1	-2767.51003	3851.18157	-0.719	0.4742		
SHISYU2	1	0.01050489	0.004032561	2.605	0.0107		
MENSEKI	1	38.91146459	9.83957035	3.955	0.0002		
TIME	1	32.64154622	7.71136567	4.233	0.0001		
SHISU	1	12.66940288	15.54404215	0.815	0.4172		
DURBIN-WATSON D		1.247					
(FOR NUMBER OF OBS.)		96					
1ST ORDER AUTOCORRELATION		0.317					

表3 ダミー変数の導入時期による重回帰分析結果

	VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > T
(a)	INTERCEP	1	-6277.53748	11826.33442	-0.531	0.5969
	SHISYU2	1	0.007657137	0.008219762	0.932	0.3541
	MENSEKI	1	52.47504960	36.32719924	1.445	0.1521
	TIME	1	66.93544772	19.14391250	3.496	0.0007
	DUM1	1	-3120.76510	13180.86777	-0.237	0.8134
	D1SYU	1	-0.000632256	0.009181964	-0.069	0.9453
	D1MEN	1	16.15853096	40.83386326	0.396	0.6933
	D1TIM	1	-63.13600254	26.09972760	-2.419	0.0176
	INTERCEP	1	-10610.82450	8374.35800	-1.267	0.2085
	SHISYU2	1	0.006730787	0.007334705	0.918	0.3613
MENSEKI	1	66.09538840	24.34460476	2.715	0.0080	
TIME	1	59.89294437	14.15950099	4.230	0.0001	
DUM2	1	5812.81716	11022.19183	0.527	0.5993	
D2SYU	1	-0.000536141	0.008428770	-0.064	0.9494	
D2MEN	1	-12.90518439	33.85970887	-0.381	0.7040	
D2TIM	1	-38.27278041	27.99232835	-1.367	0.1750	
(c)	INTERCEP	1	-16045.95733	7606.13376	-2.110	0.0377
	SHISYU2	1	0.001284863	0.006165707	0.208	0.8354
	MENSEKI	1	86.02874672	21.41433397	4.017	0.0001
	TIME	1	54.91113266	12.69916064	4.324	0.0001
	DUM3	1	16605.72655	12195.99559	1.362	0.1768
	D3SYU	1	0.004468003	0.008096457	0.552	0.5825
	D3MEN	1	-51.42176829	37.63337576	-1.366	0.1753
	D3TIM	1	-11.67492106	37.24044330	-0.314	0.7546
	INTERCEP	1	-21423.84485	6649.36321	-3.222	0.0018
	SHISYU2	1	0.005009531	0.004988461	1.004	0.3180
MENSEKI	1	99.71720187	19.27417427	5.174	0.0001	
TIME	1	42.57636023	8.92016333	4.773	0.0001	
DUM4	1	22011.20201	12032.33477	1.829	0.0707	
D4SYU	1	0.001551895	0.007793712	0.199	0.8426	
D4MEN	1	-66.62407114	37.76925486	-1.764	0.0812	
D4TIM	1	4.36503465	38.95965900	0.112	0.9110	
(e)	INTERCEP	1	-23693.88110	6509.91766	-3.640	0.0005
	SHISYU2	1	0.007797769	0.004893637	1.593	0.1146
	MENSEKI	1	104.84118	19.05600226	5.502	0.0001
	TIME	1	33.82368771	7.71585036	4.384	0.0001
	DUM5	1	23697.79206	12963.03603	1.828	0.0709
	D5SYU	1	-0.001413876	0.008071425	-0.175	0.8613
	D5MEN	1	-69.28327311	40.49636181	-1.711	0.0906
	D5TIM	1	9.56567062	41.51262000	0.230	0.8183

も予測値を用いるのであるが、本論の場合は、予測式の適合度を見るため実績値を用いることにした。TC系列の予測値を図2に示す。このTC系列の予測値に季節指数¹⁾を掛けたものを売上予測値とし、(3)式により予測誤差を算出する。

$$\text{予測誤差(\%)} = \frac{\text{実績値} - \text{予測値}}{\text{予測値}} \times 100 \dots (3)$$

(予測値 = TCの補外値 × 季節調整指数)

また、予測誤差を表6に示す。

5. 残差の検討⁶⁾

重回帰分析における残差の検討は、回帰式の正当性をみるうえで重要である。

5・1 残差の時系列プロット

残差の値を時間を横軸にとってプロットすると、図3に示す如く残差が時間的にどのように変化しているかを見ることが出来る。残差は零を中心にして上下に変動するが、それにより以下に述べる事項を

表4 信頼性の薄い変数を除いた重回帰分析結果

	VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > T
(a)	INTERCEP	1	-10397.61149	5513.18869	-1.886	0.0625
	SHISYU2	1	0.006950294	0.003589615	1.936	0.0560
	MENSEKI	1	65.25193689	16.25176323	4.015	0.0001
	TIME	1	61.45044257	11.45309714	5.365	0.0001
	DUM1	1	2038.26285	654.49968	3.114	0.0025
	D1TIM	1	-54.54767857	11.06113980	-4.931	0.0001
(b)	INTERCEP	1	-8307.03209	5753.09859	-1.444	0.1522
	SHISYU2	1	0.006422389	0.003567300	1.800	0.0752
	MENSEKI	1	59.45082892	16.73420078	3.553	0.0006
	TIME	1	62.74064974	9.91357191	6.329	0.0001
	DUM2	1	1603.20611	797.86964	2.009	0.0475
	D2TIM	1	-47.42314649	13.05448441	-3.633	0.0005
(c)	INTERCEP	1	-17234.96460	7184.99388	-2.399	0.0185
	SHISYU2	1	0.003838611	0.003874998	0.991	0.3245
	MENSEKI	1	87.90683094	20.70509651	4.246	0.0001
	TIME	1	50.76462117	10.15900674	4.997	0.0001
	DUM3	1	19907.70683	5940.95259	3.351	0.0012
	D3MEN	1	-59.05720258	16.77700692	-3.520	0.0007
(d)	INTERCEP	1	-21433.56984	6499.70081	-3.298	0.0014
	SHISYU2	1	0.005840397	0.003701867	1.578	0.1181
	MENSEKI	1	99.19227136	18.90837690	5.246	0.0001
	TIME	1	41.83963701	7.91896764	5.283	0.0001
	DUM4	1	20835.05708	5964.64998	3.493	0.0007
	D4MEN	1	-61.27958113	16.88012951	-3.630	0.0005
(e)	INTERCEP	1	-23518.78088	6398.78197	-3.676	0.0004
	SHISYU2	1	0.007442389	0.003721938	2.000	0.0486
	MENSEKI	1	104.55198	18.70484964	5.590	0.0001
	TIME	1	34.34820568	6.98719789	4.916	0.0001
	DUM5	1	21090.60952	6315.86486	3.339	0.0012
	D5MEN	1	-61.39476231	17.79639065	-3.450	0.0009

検討することが出来る。

(1) 傾向的な変化

最もマクロ的な見方で、残差プロットに右上り右下りの傾向、同期的な変化、時間と共に増大減少、曲線的な傾向などはいずれも認められない。

(2) 符号

回帰モデルが正しければ、正符号と負符号は50:50の割合で現れるはずである。本論では+符号 n_1 が48、-符号 n_2 が48になっており、この点では全く問題ない。

(3) 連の数による検定

+符号-符号が50:50で現れたとしても、その現れ方はランダムでなければならない。それを見るために連の数の検定が必要になる。+の連、-の連を合計した全体の連の数を n_R で表すと、+、-がランダムに出る場合の n_R は1つの確率変数になりある分布に従う。 $n_1 > 10$, $n_2 > 10$ の場合は連の数の分布は

近似的に(4)式の正規分布に従う。

$$\mu = \frac{2 n_1 n_2}{n_1 + n_2} + 1$$

$$\sigma^2 = \frac{2 n_1 n_2 (2 n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)} \quad \dots(4)$$

従って(5)式の如く u は標準正規分布にしたがう。

$$u = \frac{|n_R - \mu| - 0.5}{\sigma} \dots(5)$$

これに $n_1 = 48$, $n_2 = 48$, $n_R = 39$ を代入すれば、 $\mu = 49$, $\sigma^2 = 23.747$, $u = 1.9494$ となり、 u は標準正規分布の両側1%点である2.576より小さくなるので有意でない。即ち連の数から見てこの残差系列はランダムと見なせる。

(4) 連の長さによる検定

+、-の符号のランダムな系列では、一方の符号が

表5 最適モデルの重回帰分析結果

SAS

DEP VARIABLE: URIAGE2

ANALYSIS OF VARIANCE					
SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	5	308796530	61759305.90	204.471	0.0001
ERROR	90	27183983.45	302044.26		
C TOTAL	95	335980513			

ROOT MSE	549.5855	R-SQUARE	0.9191
DEP MEAN	16687.27	ADJ R-SQ	0.9146
C.V.	3.293442		

PARAMETER ESTIMATES

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > T
INTERCEP	1	-23518.78088	6398.78197	-3.676	0.0004
SHISYU2	1	0.007442389	0.003721938	2.000	0.0486
MENSEKI	1	104.55198	18.70484964	5.590	0.0001
TIME	1	34.34820568	6.98719789	4.916	0.0001
DUM5	1	21090.60952	6315.86486	3.339	0.0012
D5MEN	1	-61.39476231	17.79639065	-3.450	0.0009

DURBIN-WATSON D 1.622
 (FOR NUMBER OF OBS.) 96
 1ST ORDER AUTOCORRELATION 0.141

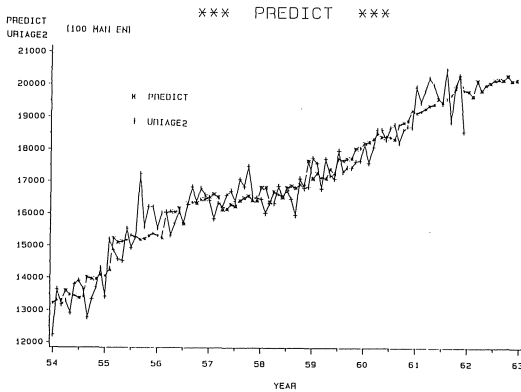


図2 予測値及び実測値 (季調済み)

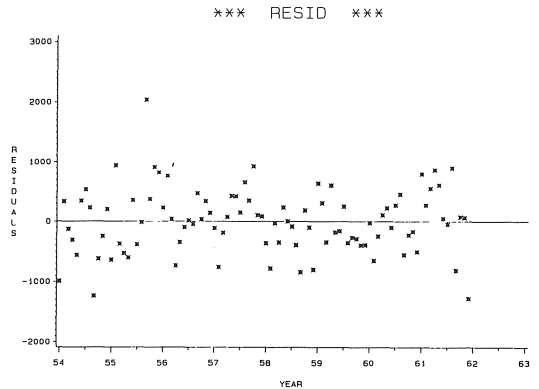


図3 残差の時系列プロット

表6 予測誤差 (%)

1 月	2 月	3 月	4 月
4.16	4.13	1.93	6.51
5 月	6 月	7 月	8 月
7.13	2.98	0.36	2.48
9 月	10 月	11 月	12 月
-1.40	3.44	6.00	-2.82

けが連続して多く現われることはまれである。本論の場合十側で9個、一側で8個が連続している。7個以上同じ符号がならぶ場合はランダムとは見なせないとする、連の長さの点からは、残差はランダムとは見なされずに、多少問題を含んでいる。

5・2 残差と従属変数の値との散布図

横軸に従属変数の予測値、縦軸にそれに対応する残差をとって散布図を画くと図4の如くなる。この図には特に問題になる様なことは含まれていないと

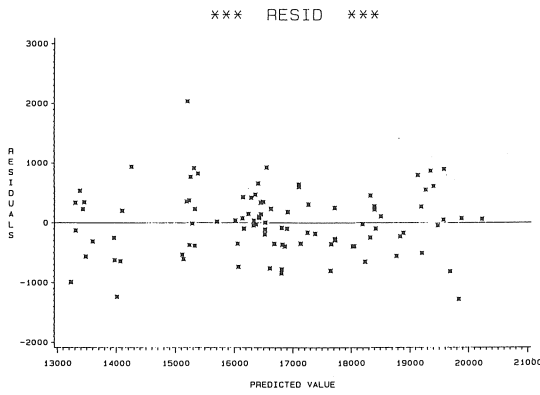


図4 残差と予測値の散布図

考えられる。

5・3 残差の度数分布

残差の度数分布を図5に示す。分散分析の表5より

$$Ve=302044.26$$

$$\sqrt{Ve}=549.5855$$

となり、残差の3シグマ限界を求めると

$$\pm 3 \sqrt{Ve}=\pm 1648.76$$

で、それを越えるものはNo.21の+2154.6が有り、異状値とみなされる。

5・4 ダービン-ワトソン比

時列データに対しては、誤差の独立性を調べるためダービン-ワトソン比(D)の検討が必要になる。誤差の間に相関がなければ、Dの値は2に近づくはずであり、正の相関があればDの値は2より小さくなり、負の相関があればDの値は2より大きくなる。データの値が100, 説明変数が5個の場合の統計限界

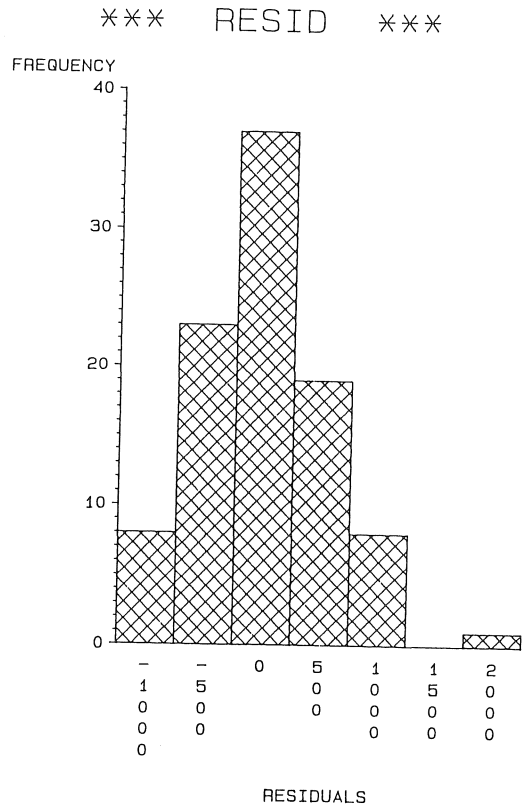


図5 残差の度数分布

は有意水準1%で $D_L=1.44$, $D_U=1.65$ となっている。正の相関を調べるには

$$D \leq D_L \quad \text{正の相関あり}$$

$$D \geq D_U \quad \text{正の相関なし}$$

$$D_L < D < D_U \quad \text{判定出来ない}$$

表7 ダミー変数による比較

		ダミー変数なし	ダミー変数有り
寄与率 (R ²)		0.9013	0.9191
残	DURBIN-WATSON (D)	1.247	1.622
	傾向的变化	無し	無し
	符号	+48, -48	+48, -48
差	連の数による検定	ランダムとは見なせず	ランダムと認められる
	連の長による検定	ランダムとは見なせず	ランダムとは見なせず

と検定すればよい。

本論の場合にこれを近似的に適用すると、 $D = 1.622$ なので

$$D_L (1.44) < D (1.622) < D_U (1.65)$$

となり、正確には判定出来ない。

6. 考察

重回帰分析において、前報におけるダミー変数なしと、本論におけるダミー変数を導入した場合の結果を比較すると表7の如くなる。

これによると、回帰分析全体の信頼性を表す寄与率は90.13%から91.91%へと向上している。また残差についてもその独立性は大幅に向上し、回帰式はより信頼性が増したものと考えられる。

7. おわりに

企業が経営計画を立てる場合や、研究開発・製品計画等を実施するときには、予測が重要なものとなっている。百貨店の衣料品の売上予測をする場合に不規則変動を除去するための重回帰分析において、説明変数として消費支出、売場面積、時間を用い、さらに残差の自己相関を改善する為に、ダミー変数を追加して解析を行なった。その結果残差の自己相関は大幅に改善され、回帰式の信頼性は向上することが出来たが、予測誤差は前報の-0.05%~6.70%が本論では0.36%~7.13%となり改善が見られなかった。今後の課題として、さらにモデル式の改善の必要が認められる。

参考文献

- 1) 橋本郁郎：大型小売店における売上予測，愛知工業大学研究報告 Vol. 25, Part B, p. 71~77, 1990
- 2) 愛知県：愛知県統計年鑑，愛知県，1979~1987
- 3) 本多正久：BASICによる予測入門，共立出版，東京，1986
- 4) SAS USER'S GUIDE STATISTICS ver. 5, SAS Institute Inc.
- 5) SAS USER'S GUIDE GRAPH ver. 5, SAS Institute Inc.
- 6) 奥野忠一，久米 均，芳賀敏郎，吉澤 正：多変量解析法，日科技連，東京，1986

(受理 平成3年2月20日)