

周期時系列の統計解析 (6) ダミー変数を用いた重回帰分析

nino

2018年4月15日

これまでに、周期時系列のトレンド等を求めるため、コサイン関数モデル近似による最小二乗法（第2報）と移動平均法（第3報）の2つの統計手法により検討してきた。今回は、「ダミー変数を用いた重回帰分析」により同様に考察した。

ダミー変数を用いた重回帰について

ダミー変数とは、名義尺度（数値でないデータ）に対して、0と1を用いて数量化した変数のことである。ダミー変数に関するわかりやすい説明は参考文献1を参照されたい。重回帰の計算ソフトは、Excelの「データ分析」>「回帰分析」を用いた。エクセルによる重回帰の計算方法の解説としては、参考文献2、3などがある。以降は、ダミー変数を用いた重回帰を、単に重回帰という。

周期時系列の重回帰モデル

周期時系列の重回帰モデル Y_t は時間 t の関数として次式で表される。

$$Y_t = T_t + S_t + \varepsilon_t$$

ここで、 T_t はトレンド成分、 S_t は季節成分、 ε_t は不規則成分であり、また、 $t=1, 2, \dots, n$ である。

周期時系列の例として毎月データの気温時系列を想定し、データ数 $n=12 \times j$ (j 年間) とした場合について検討する。

重回帰モデルの条件を、

- ① トレンドは直線（一定）
- ② 月効果（季節成分）は毎年共通
- ③ 月効果の合計は0

とすると、重回帰モデル式は、

$$Y_t = a \times t + b + S_{(t \bmod 12)} + \varepsilon_t$$

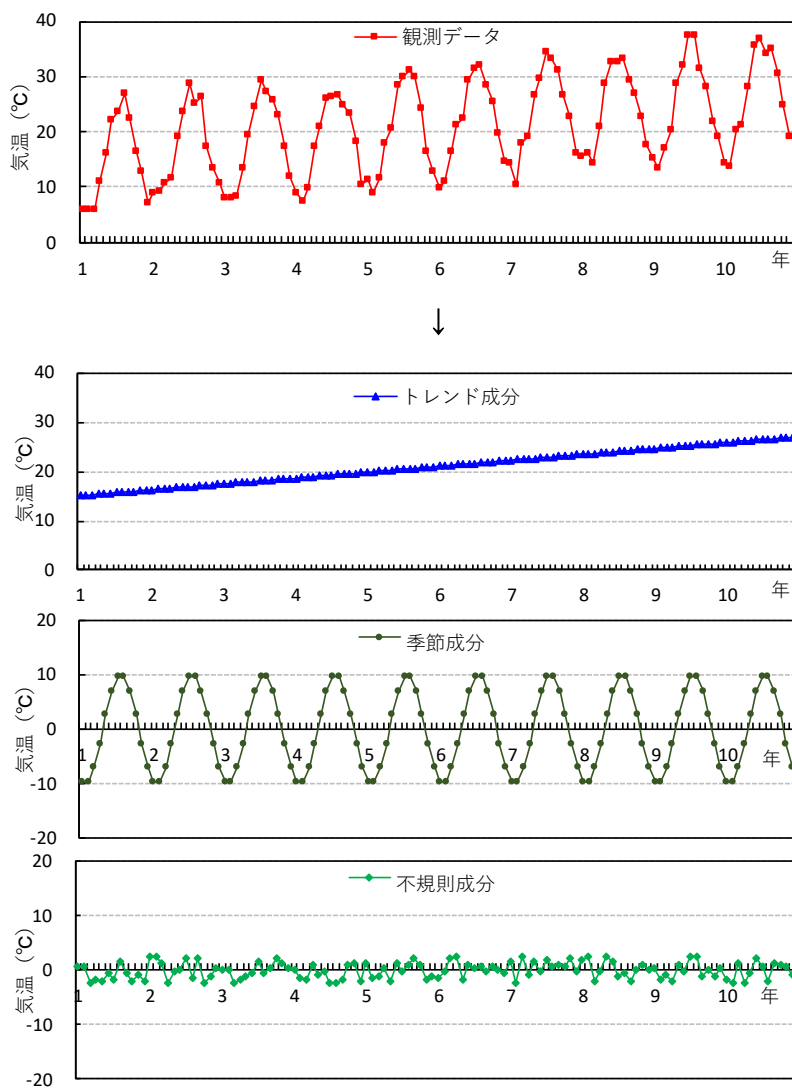
となる。ここで、 a はトレンド、 b は定数項であり、 $a \times t + b$ が回帰直線に相当する。さらに、月効果をコサイン関数で定義した場合の重回帰モデルは、

$$Y_t = a \times t + b + H \cos[2\pi/T(t + \phi)] + \varepsilon_t$$

で表される。この式は、重回帰を検証するために使用する。ここで、 H : 振幅、 T : 周期、 ϕ : 位相であり、月効果については、 $S_{(t \bmod 12)} = H \cos[(2\pi/T)(t + \phi)]$ で対応している。

図1に、パラメーター値を $a=0.1^\circ\text{C}/\text{月}$ 、 $b=15^\circ\text{C}$ 、 $H=10^\circ\text{C}$ 、 $T=12$ か月、 $\phi=4.5$ 月とした場合におけるコサイン関数の重回帰モデルの観測データと各成分の時系列の模式図を示した。月効果については、例えば、1月の月効果 S_1 は $H \cos[(2\pi/T)(t + \phi)] = 10 \cos[(2\pi/12)(1+4.5)] = -9.66$ となる。

図 1



また、重回帰モデルを行列で表すと、図 2 のようになる。

図 2

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_{11} \\ Y_{12} \\ Y_{13} \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 11 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 12 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 13 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ n & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \\ S_1 \\ S_2 \\ \cdot \\ \cdot \\ \cdot \\ S_{11} \end{pmatrix}$$

図 2 において、月効果 S_1, S_2, \dots, S_{11} に対応する行列がダミー変数に相当する。また、連立方程式では、次式で表される。

$$\begin{aligned}
Y_1 &= a \times 1 + b + S_1 \\
Y_2 &= a \times 2 + b + S_2 \\
&\vdots \\
Y_{12} &= a \times 12 + b - S_1 - S_2 \cdots - S_{11} \\
&\vdots
\end{aligned}$$

毎年 12 月の行は他の月の行とは異なり、ダミー変数がすべて -1 となっている。これは、先述したように、月効果の和は 0 ，すなわち、次式が成り立つからである。

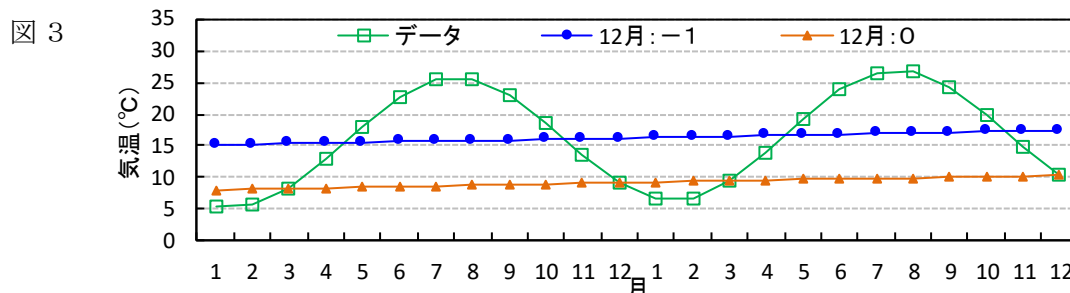
$$S_1 + S_2 \cdots + S_{12} = 0 \quad \Leftrightarrow \quad S_{12} = -S_1 - S_2 \cdots - S_{11}$$

一方、12 月行のダミー変数をすべて 0 としている報告が多く見受けられる。この場合は、例えば、 Y_{12} の式は、

$$Y_{12} = a \times 12 + b$$

となり、 Y_{12} のデータは回帰直線と一致する。そのため、毎年 12 月のデータは回帰直線を通ることになる。

12 月行のダミー変数の違いを具体例で示す。図 3 に、不規則成分を除いたコサイン関数の重回帰モデル Y_t における最初の 2 年間の観測データ（気温）の時系列と、12 月行のダミー変数を -1 （12 月： -1 ）とした場合の回帰直線および 0 （12 月： 0 ）とした場合の回帰直線を示した。



12 月行のダミー変数が -1 の場合は、回帰直線は切片 (15°C) に毎月のトレンド (0.1°C) が加算されたものとなっており、各月効果 S_t は回帰直線からの偏差を表している。一方、12 月行のダミー変数が 0 の場合の回帰直線は、12 月のデータを通る直線となっている。3 月のデータも 12 月： 0 の回帰直線上にあるのは、12 月のデータと同じ水準にあるからである。この場合の各月効果は 12 月のデータを通る回帰直線からの偏差となり、12 月： -1 の場合とは異なる値になる。

Excel による重回帰モデルの解法

重回帰モデル Y_t に Excel の回帰分析を適用した例を紹介する。

表 1 に、Excel の「回帰分析」における入力 Y 範囲 = 目的変数と入力 X 範囲 = 説明変数を示した。なお、行列では定数項 b があつたが、Excel の入力 X 範囲に該当項目が無いのは「回帰分析」に既に切片として組み込まれているためと思われる。そのダミー変数はすべて 1 なので、組み込みやすく、多くの重回帰ソフトでは、Excel と同様になっている。

表 1

目的変数	説明変数											
	トレンド [*]	季節効果										
		a	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}
Y_1	1	1	0	0	0	0	0	0	0	0	0	0
Y_2	2	0	1	0	0	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Y_{11}	11	0	0	0	0	0	0	0	0	0	0	1
Y_{12}	12	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
Y_{13}	13	1	0	0	0	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Y_n	n	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

表 2 には、月効果をコサイン関数で定義した場合の重回帰モデルの式 Y_t において、不規則成分を除いた場合の重回帰結果を示す。

不規則成分を除いたモデルなので、切片 b 、トレンド a および季節効果 S_t は設定パラメーター通りとなる。例えば、1月の月効果 S_1 は、先に示したように -9.66 となっている。 S_{12} については、 $S_{12} = -S_1 - S_2 \cdots -S_{11}$ であるから、 $S_{12} = -0.707$ が得られる。

表 2

	係数	標準誤差	t	P-値
切片	15.00	4.1E-15	3.7E+15	0.00
a	0.100	5.9E-17	1.7E+15	0.00
S_1	-9.66	6.7E-15	-1.4E+15	0.00
S_2	-9.66	6.7E-15	-1.4E+15	0.00
S_3	-7.07	6.7E-15	-1.0E+15	0.00
S_4	-2.59	6.7E-15	-3.8E+14	0.00
S_5	2.59	6.7E-15	3.8E+14	0.00
S_6	7.07	6.7E-15	1.0E+15	0.00
S_7	9.66	6.7E-15	1.4E+15	0.00
S_8	9.66	6.7E-15	1.4E+15	0.00
S_9	7.07	6.7E-15	1.0E+15	0.00
S_{10}	2.59	6.7E-15	3.8E+14	0.00
S_{11}	-2.59	6.7E-15	-3.8E+14	0.00

なお、この場合のようなモデルでは、重回帰を用いなくて、先の連立方程式から係数を求めることができる。すなわち、未知数は、 a 、 b 、 S_1 、 S_2 、 \cdots 、 S_{11} の 13 個であるから、必要な方程式の数は 13 個となる。例えば、最初の 13 個の連立方程式 Y_1 、 Y_2 、 \cdots 、 Y_{13} について解くと、表 2 と同じ係数が得られる。

気温時系列への重回帰の適用

実際の観測データに重回帰を適用した。観測データは、1930年1月から2015年12月までの86年間における月平均気温（横浜：1032データ）の時系列である（図4）。

表3に、86年間の気温時系列の重回帰結果を、また、図5には、重回帰結果の予測値と残差（不規則成分）の時系列を示した。

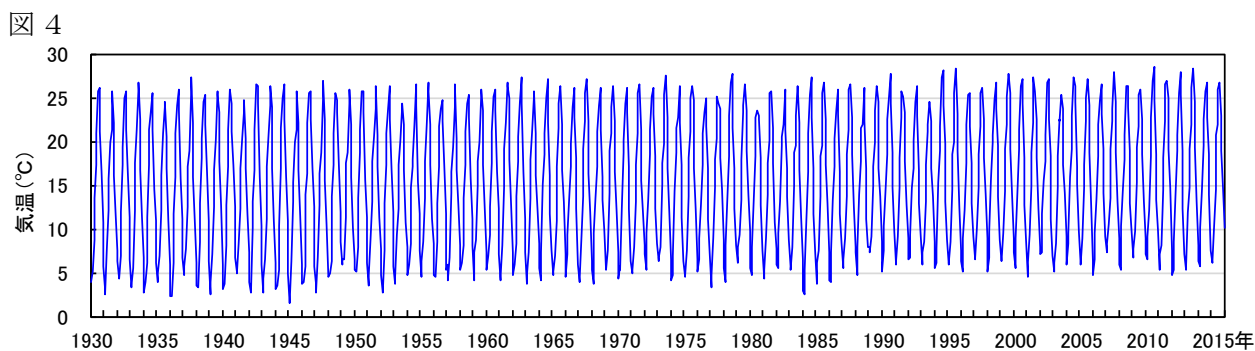


図 4 において、気温は年毎に少し変動がみられるが、わずかに上昇傾向を示している。

表 3

	係数	標準誤差	t	P-値
切片	14.03	0.06	220	0.00
a	0.0023	0.00	21	0.00
S_1	-10.19	0.11	-96	0.00
S_2	-9.84	0.11	-93	0.00
S_3	-6.86	0.11	-65	0.00
S_4	-1.64	0.11	-15	0.00
S_5	2.68	0.11	25	0.00
S_6	5.87	0.11	56	0.00
S_7	9.64	0.11	91	0.00
S_8	11.13	0.11	105	0.00
S_9	7.63	0.11	72	0.00
S_{10}	2.10	0.11	20	0.00
S_{11}	-2.88	0.11	-27	0.00

表 3 によれば、トレンド ab は $0.0023^\circ\text{C}/\text{月}$ 、切片 b は 14.03°C と推定された。月効果は $-10.19^\circ\text{C} \sim 11.13^\circ\text{C}$ の範囲にあり、8月が最も高く1月が最低を示した。12月の月効果 S_{12} については、 $S_{12} = -S_1 - S_2 \cdots -S_{11}$ であるから、 $S_{12} = -7.64^\circ\text{C}$ となる。また、P-値は全て 0.00 であり、1% で有意となった。

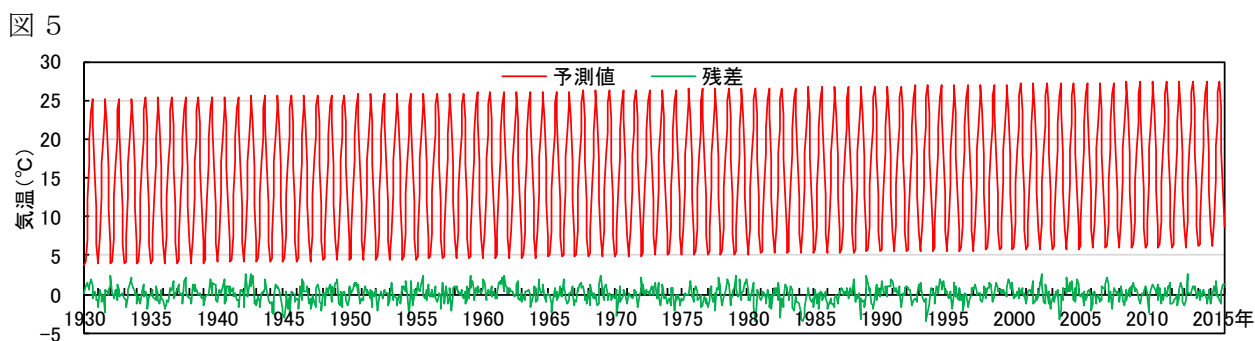


図 5 では、予測値は観測データをほぼ再現しており、また、残差に系統的な偏りはみられなかった。

異なる統計手法による気温時系列のトレンドの比較

これまでに、以下の3つの統計手法により気温時系列のトレンドを求めてきた。

- ① 気温時系列を直線とコサイン関数の合成関数として定義し、そのパラメーターとして最小二乗法によりトレンドを求める方法【最小二乗法】：第2報
- ② 気温時系列の12か月中心化移動平均(12cMA)に単回帰分析を適用してトレンドを求める方法【単回帰】：第3報
- ③ ダミー変数を用いた重回帰分析によりトレンドを求める方法【重回帰】：本報

これら3つの方法により、対象期間が10年間(1991年1月～2000年12月)と86年間(1930年1月～2015年12月)の2例について求めたトレンドを比較検討した(表4)。なお、①の方法は一般的ではなく特殊な方法であるが、参考のために、比較対象とした。また、1990年代の10年間は、気温上昇率が比較的大きい期間に相当する。

まず、10年間のトレンドについては、最小二乗法のトレンドが $0.00645^{\circ}\text{C}/\text{月}$ で最も小さかった。重回帰のトレンドは最小二乗法の1.02倍でほとんど違いがないが、単回帰のトレンドは最小二乗法の約1.3倍と比較的大きな値を示した。単回帰のトレンドが大きいのには循環変動成分による影響があるためと考えられた。また、最小二乗法と重回帰でトレンドにほとんど違いがみられないのは、両者ともに循環変動成分が不規則成分に含まれるためと推察される。

次に、86年間のトレンドは、3者ともほぼ同値($0.00226^{\circ}\text{C}/\text{月} \sim 0.00228^{\circ}\text{C}/\text{月}$)であり、気温は86年間で約 $2.3^{\circ}\text{C} \sim 2.4^{\circ}\text{C}$ 上昇したと推定された。また、86年間のトレンドの3倍程度が10年間のトレンドに相当した。1990年代の10年間の気温上昇率が大きいことが確認できた。したがって、気温のトレンドを求めるためには、短期間は避け、なるべく長期間を採用すべきである。この例では、80年程度の期間があれば、いずれの統計手法でもトレンドに大きな違いは生じないと考えられた。

表4

統計手法	トレンド($^{\circ}\text{C}/\text{月}$)	
	10年間	86年間
最小二乗法	0.00645	0.00226
単回帰	0.00857	0.00228
重回帰	0.00658	0.00227

参考文献

1. ダミー変数について
<http://aoki2.si.gunma-u.ac.jp/lecture/Regression/mreg/dummy-variable/dummy.html>
2. Excelによる重回帰分析
www2.kobe-u.ac.jp/~hamori/Jhamori/EXCEL2.pdf
3. 季節ダミーを用いた回帰分析
http://d.hatena.ne.jp/hiroshi_moriyasu/20081215/1229297706