

JMP を用いた ARIMA モデルのあてはめ

SAS Institute Japan 株式会社 JMP ジャパン事業部

2013 年 2 月 作成

1. はじめに

JMP の時系列分析では、一変量の時系列データに対する分析や予測を行うことができ、時系列データに対するグラフ表示、時系列モデルのあてはめ、モデルの評価、予測まで、対話的に分析を実行することができます。

時系列データにあてはめるモデルとしては、ARIMA モデルが有名であり、将来の予測のために広く用いられています。しかし、ARIMA モデルでは、自己回帰、差分、移動平均の次数を決める必要があり、季節 ARIMA モデルでは、これらの次数に加え、季節周期を決めてモデルを同定する必要があります。

JMP では、ARIMA モデル、季節 ARIMA モデルの推定の際、次数の範囲を指定して、複数の ARIMA モデルをあてはめる機能があります。さらに、あてはめたモデルについては、AIC などの適合度基準によるモデルの評価を行うことや、残差のレポートにより、モデルが適切かどうかの判断を行うことができ、これらの機能が、複数の ARIMA モデル間での適合度の比較、あてはめたモデルに対する前提の検証に役に立ちます。

本章では、JMP を用いて、時系列データの分析を行う手順と、ARIMA モデル、季節 ARIMA モデルのあてはめ、複数の ARIMA モデルのあてはめと、モデルの評価方法、予測についてご説明します。

以下、ARIMA モデルについては、 $ARIMA(p,d,q)$ と表現します。 p, d, q はそれぞれ、自己回帰の次数、差分の次数、移動平均の次数を示します。季節 ARIMA モデルについては、 $SARIMA(p,d,q)(P,D,Q)s$ と表現します。 P,D,Q はそれぞれ、季節自己回帰の次数、季節差分の次数、季節移動平均の次数を示し、 s は、1 季節の期数(季節周期)を示します。

複数の ARIMA モデルをあてはめる機能は、JMP のバージョン 9 以上で搭載されています。

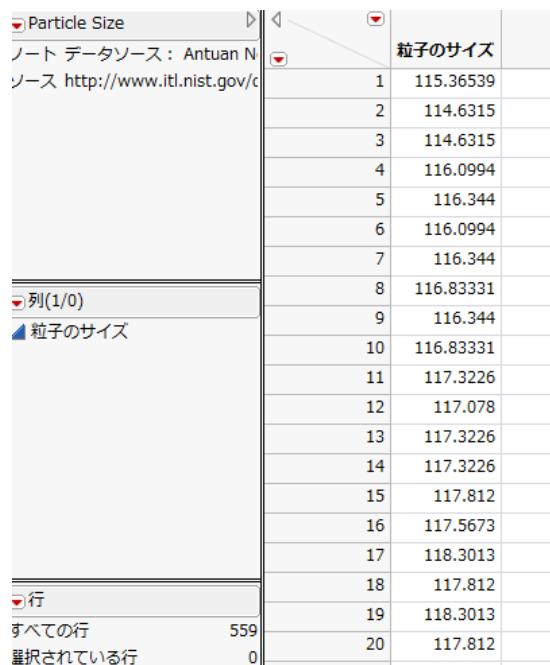
JMP の時系列分析で選択できるオプションの一覧を示します。オプションは、時系列分析レポートの左上にある赤い三角ボタンをクリックすることにより選択することができます。



2. 「時系列分析」プラットフォームの使用法

JMPの「時系列分析」プラットフォームでは、一変量の時系列に対する分析を行うことができます。この章では、JMPのサンプルデータを用いて、このプラットフォームの使用法をご説明します。

JMPのメニューbaruより、[ヘルプ] > [サンプルデータ] を選択しサンプルデータの索引のウィンドウを表示させ、[サンプルデータディレクトリ] ボタンをクリックします。ファイルの選択ウィンドウから、Time Series フォルダをクリックし、「Partial Size.jmp」(粒子のサイズ.jmp) を開きます。

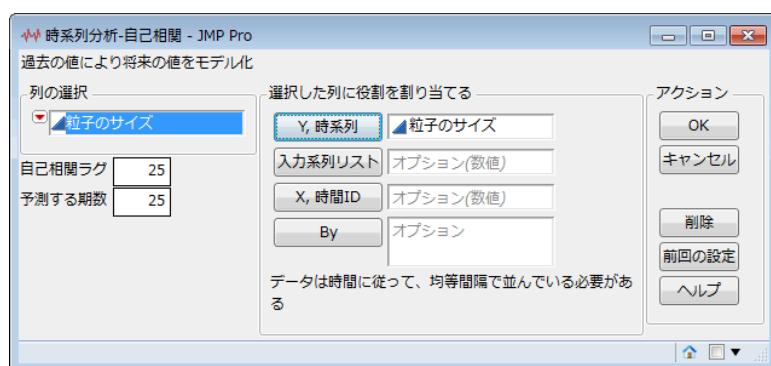


The screenshot shows the JMP Sample Data Index window. The left pane displays the dataset 'Particle Size' with a note about the data source: 'ノート データソース: Antuan N' and 'ソース http://www.itl.nist.gov/csd/datasets/nanoparticle/size/size.htm'. The right pane is a table titled '粒子のサイズ' (Particle Size) with the following data:

	粒子のサイズ
1	115.36539
2	114.6315
3	114.6315
4	116.0994
5	116.344
6	116.0994
7	116.344
8	116.83331
9	116.344
10	116.83331
11	117.3226
12	117.078
13	117.3226
14	117.3226
15	117.812
16	117.5673
17	118.3013
18	117.812
19	118.3013
20	117.812

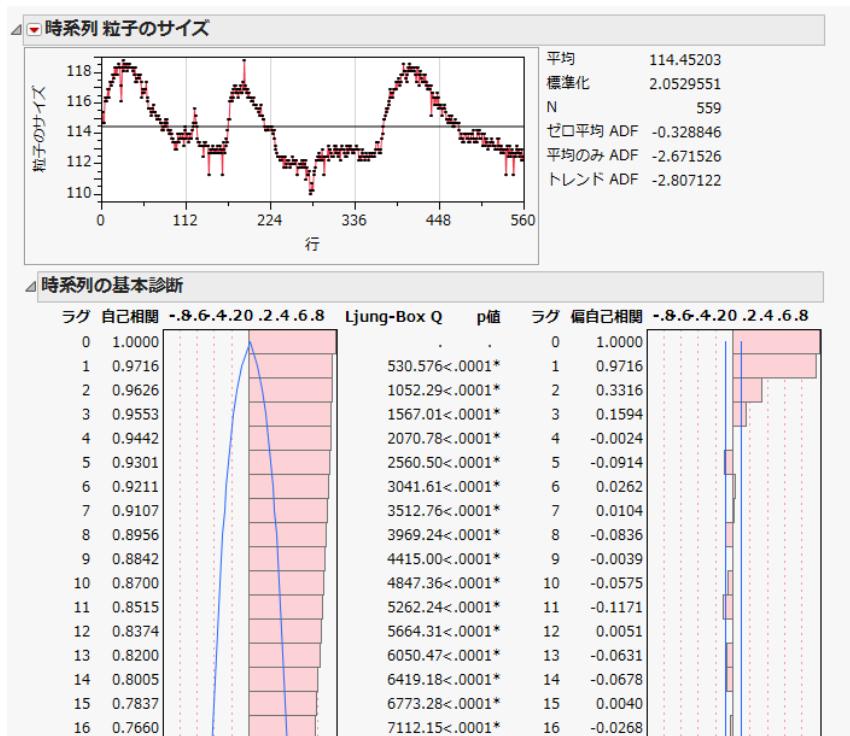
このデータは、自動噴霧器をテストするために、噴霧剤入りのミニスプレードライヤーから収集されたものです。変数「粒子のサイズ」には、等間隔の時間で測定された粒の直径が記録されています。

JMPで時系列分析を実行するには、メニューbaruより、[分析] > [モデル化] > [時系列分析] を選択します。[Y, 時系列] に列「粒子のサイズ」を指定し、[OK] をクリックします。



注意: JMPの時系列分析では、データは時間に従って、均等間隔に並んでいると仮定して分析を行います。[X, 時間 ID] には、X軸（時間軸）のラベルとなる変数があれば指定することができます。この変数の値自体が等間隔でないときでも、等間隔とみなして分析が行われます。

レポートとして、折れ線グラフや、自己相関、偏自己相関の数値、グラフが表示されます。



時系列データでは、折れ線グラフにより、データの傾向、周期などを確認します。いくつかの山や谷があり、全体的には、若干の減少傾向がありそうです。

折れ線グラフの下側に表示される「時系列の基本診断」のレポートには、左側に自己相関係数そのグラフ、右側に偏自己相関係数とそのグラフが出力されます。以下では、これらの出力を自己相関プロット、偏自己相関プロットと呼ぶことにします。

自己相関係数は、時系列をある期数だけずらし、元の時系列とずらした時系列との相関係数を求ることにより計算されます。ずらした期数はラグと呼ばれ、例えば1期ずらした場合は、ラグが1の行の自己相関係数(=0.9716)を参照します。グラフ上に表示される水色の曲線は、およそその95%信頼区間を示します。

このデータでは、自己相関係数は高く、系列間に相関があることがわかり、ラグが大きくなるについて、自己相関係数は、徐々に小さくなっています。

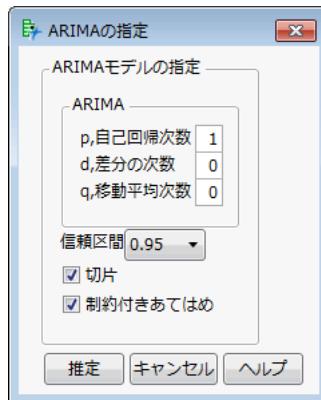
偏自己相関係数は、各ラグにおいて、そのラグよりも期数が小さいラグから受ける自己相関の影響を取り除いた上で、求めた相関係数を示します。このデータでは、ラグが1のときの偏自己相関係数が0.9716と突出しています。

自己相関プロットと偏自己相関プロットは、ARIMAモデルの自己回帰次数、移動平均次数を決める際のヒントになります。

偏自己相関プロットでは、ラグ1の値が突出していますが、ラグ2やラグ3でも信頼区間の外にあることから、AR(自己回帰)の次数として、1、2または3であると考えられます。自己相関プロットでは、突出している値はなく、徐々に減少しているため、ここでは、MA(移動平均)の次数は0とします。

まず、最初に AR(1) (= ARIMA(1,0,0)) をあてはめてみます。

レポート「時系列 粒子のサイズ」の左上にある赤い三角ボタンより、[ARIMA] を選択します。「ARIMA モデルの指定」のダイアログボックスにおいて、「自己回帰次数」の値を 1 として、[推定] ボタンをクリックします。



レポート下側に、「モデルの比較」のレポートと、「モデル: AR(1)」のレポートが追加されます。

「モデル: AR(1)」のレポートを参照します。

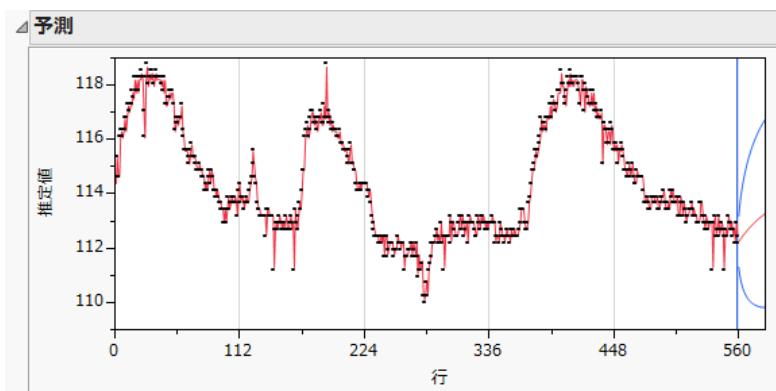


「モデルの要約」には、あてはめたモデルに対する適合度統計量が表示されます。レポートにある「赤池の情報量基準」(AIC) や SBC は、良くあてはまっているほど、値が小さくなる基準です。後述するモデルの比較の際に用います。

「パラメータ推定値」は、AR モデルにおけるパラメータの推定値、標準誤差、係数に対する有意差検定の結果が表示されます。

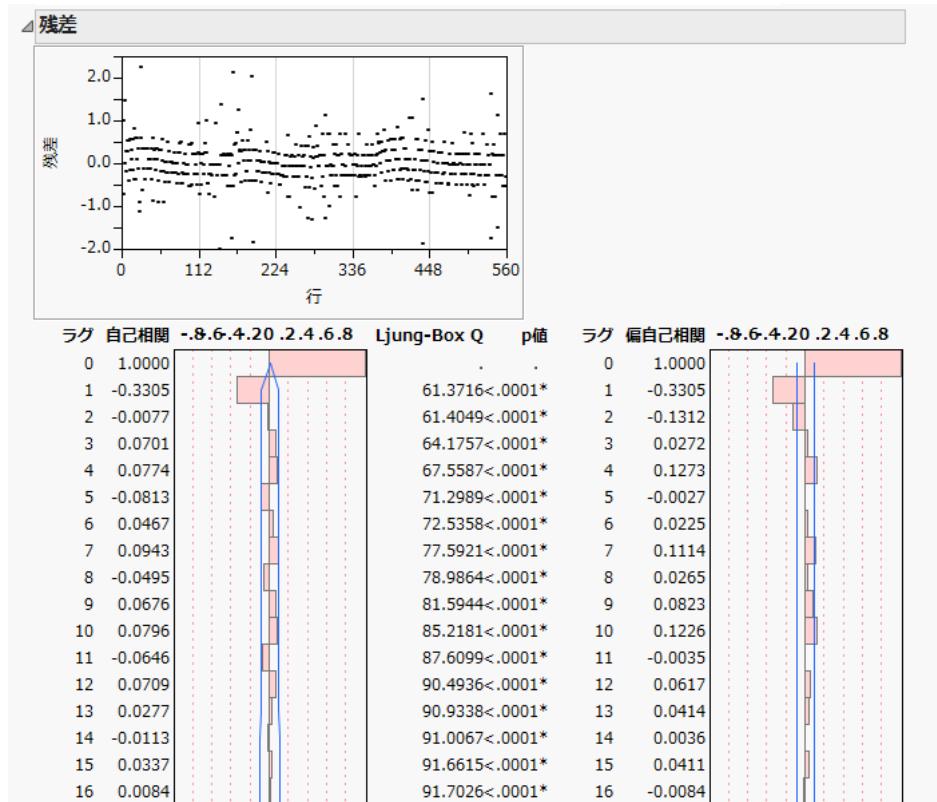
JMP では、最尤法によって ARIMA モデルのパラメータを推定しています。

あてはめたモデルに対する、「予測」のグラフが表示されます。



このグラフでは、モデルによる予測値を折れ線グラフで表示しています。横軸が 559 (= データ数) の青い垂線より左側の領域では、実測値のプロットと予測値を比較することができます。右側の領域は、将来の予測値とその 95%信頼区間が表示されます。AR(1) モデルをあてはめた場合、559 時点より先の直近の予測値は、上昇傾向にあることがわかります。

「予測」のレポートには、「残差」のレポートがあり、左側の三角ボタンをクリックすると開くことができます。上側にはモデルをあてはめたときの残差プロット、下側には、残差の自己相関プロット、偏自己相関プロットが表示されます。



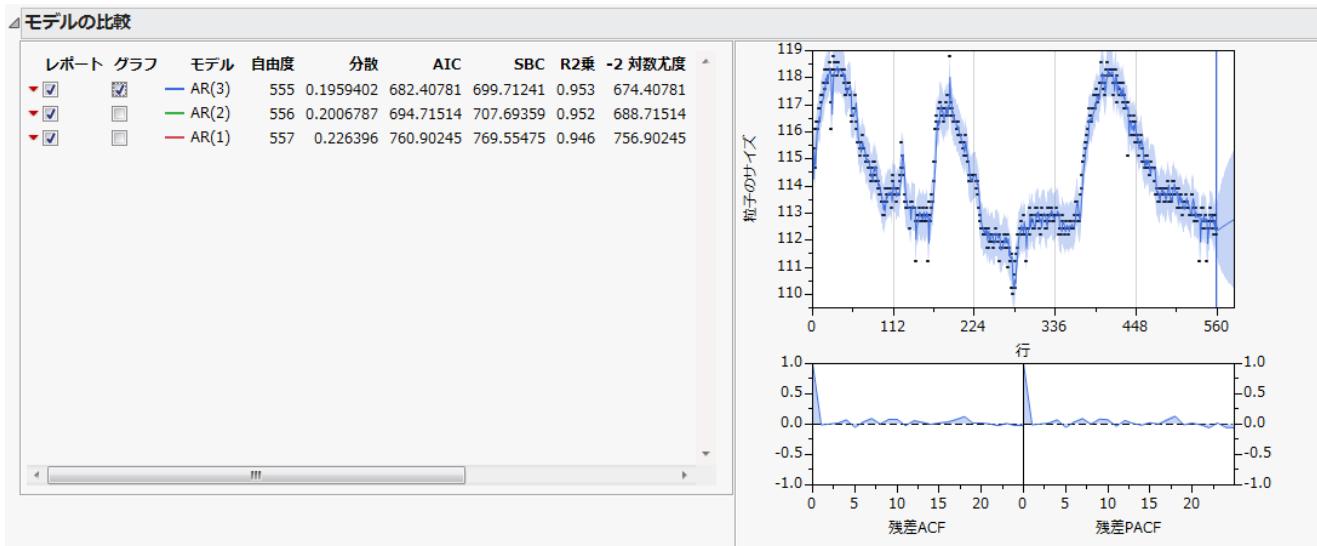
ARIMA モデルにおいて、残差は 0 を中心に正規分布に従い、かつ自己相関や偏自己相関がゼロである、いわゆるホワイトノイズであることが望ましいとされます。

残差の自己相関プロット、偏自己相関プロットをみると、ともにラグ 1 での自己相関、偏自己相関が大きくなっているため、あてはめたモデルが適切でない可能性があります。

自己相関プロットの右側には、Ljung-Box Q 統計量と統計量に対する p 値が表示されます。各ラグにおける Ljung-Box Q 統計量は、先頭(ラグ 1)から該当のラグまでの複数の自己相関がすべて 0 であるという帰無仮説を検定し、帰無仮説が棄却されると、少なく 1 つの自己相関係数が 0 と有意に異なる結論できます。この検定は、時系列がホワイトノイズであるかどうかを判定する方法として用いられます。この例では、すべてのラグで、高度に有意のため、その時系列はホワイトノイズではないと判断されます。

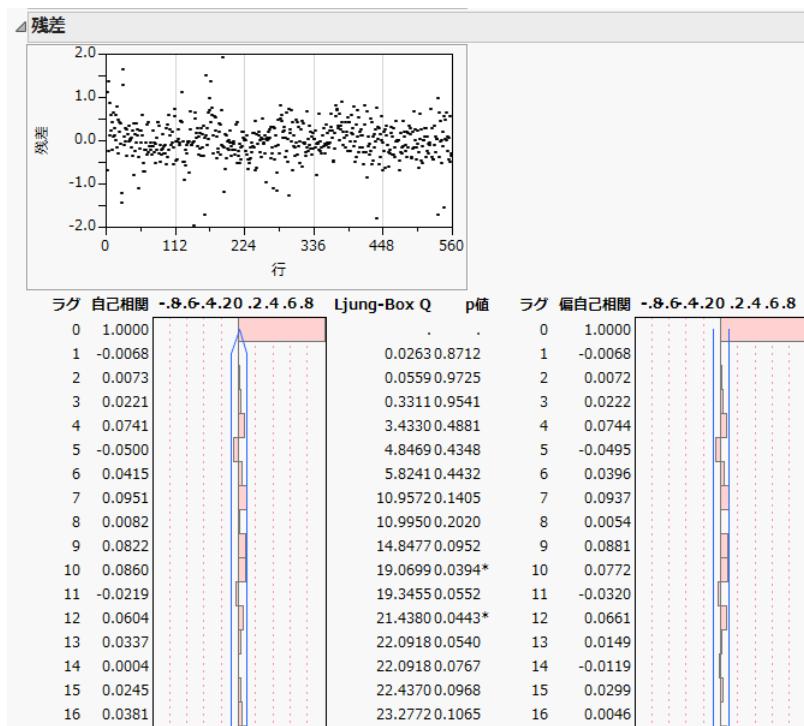
同様に [ARIMA] のオプションを選択することにより、AR(2)、AR(3) のモデルをあてはめます。

「モデルの比較」のレポートには、AR(1) に加え、AR(2)、AR(3) のレポートも追加されます。レポートでは、モデルごとの適合度統計量がまとめられ、さまざまなモデルを比較することができます。モデルは、AIC の小さい順、すなわち、AIC の基準による適合度が良い順に上から下へ並べられます。



レポートでは、AR(3)、AR(2)、AR(1) の順に並んでいますので、このモデルの中では、AR(3) が最も適合しているモデルということになります。さらに、各モデルの「グラフ」のチェックをいれると、右側のグラフに、あてはめたモデルを表示、残差の自己相関プロット(残差 ACF)、残差の偏自己相関プロット(残差 PACF)が表示されます。

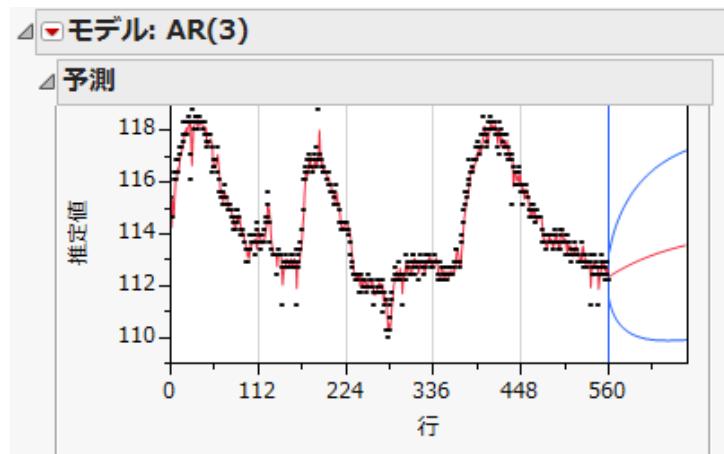
AR(3) の残差レポートを参照してみます。



ラグ 1 以降の自己相関、偏自己相関ともに小さく、Ljung-Box Q 統計量による検定の p 値も、有意でない箇所が多くなっています。

以上の分析、考察より、モデルとして AR(3) を採用することにします。

予測する期数は、レポート「時系列 粒子のサイズ」の左上の三角ボタンをクリックして、[予測する期数] オプションを選択し、予測したい期数を入力することにより変更することができます。AR(3) のあてはめの「予測」、予測する期数を 100 に指定したときのグラフを以下に示します。



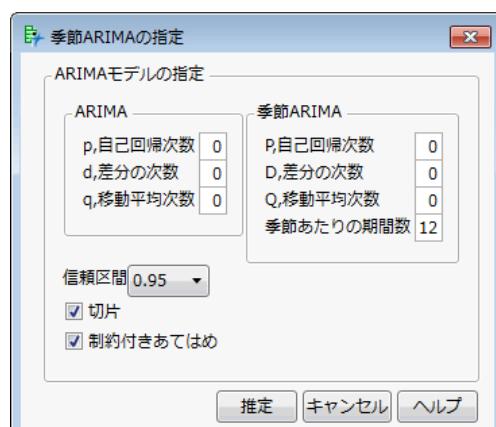
AR(3) の「予測」プロットでも、データ数である 559 時点より先の予測値は、上昇傾向にあることがわかります。

実際の予測値、その信頼区間の値を調べたい場合は、レポート「モデル: AR(3)」の左上にある赤い三角ボタンより、[列の保存]、または [予測式の保存] を選択することにより、予測値やその信頼区間をまとめたデータテーブルを出力することができます。

予測テーブル	実測値 粒子のサイズ	行	予測値 粒子のサイズ	予測値の標準誤差 粒子のサイズ	残差 粒子のサイズ	上側信頼限界 (0.95) 粒子のサイズ	下側信頼限界 (0.95) 粒子のサイズ
553	112.9189	553	112.83568168	0.4426513055	0.0832183245	113.70326229	111.96810106
554	112.4296	554	112.88239268	0.4426513055	-0.452792681	113.7499733	112.01481206
555	112.6742	555	112.61401698	0.4426513055	0.060183017	113.4815976	111.74643637
556	112.18491	556	112.68062741	0.4426513055	-0.495717408	113.54820802	111.81304679
557	112.9189	557	112.37326042	0.4426513055	0.5456395774	113.24084104	111.50567981
558	112.4296	558	112.72734079	0.4426513055	-0.297740786	113.5949214	111.85976017
559	112.18491	559	112.53605576	0.4426513055	-0.351145761	113.40363638	111.66847514
560	•	560	112.39320265	0.4426513055	•	113.26078327	111.52562203
561	•	561	112.37955574	0.5133756645	•	113.38575355	111.37335792
562	•	562	112.38195958	0.5744464379	•	113.50785391	111.25606525
563	•	563	112.41332294	0.6406472406	•	113.66896846	111.15767742
564	•	564	112.43014256	0.6948999413	•	113.79212142	111.06816371
...

実際の予測値や信頼区間を数値として確認したい場合は、このデータテーブルの数値を参照します。

この章では、ARIMA のあてはめについて説明しましたが、[季節 ARIMA] オプションを選択することにより、季節 ARIMA モデルもあてはめることができます。

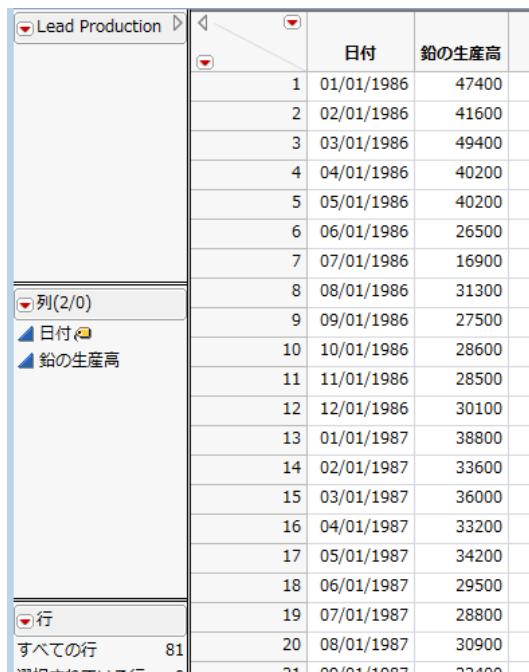


3. 複数の ARIMA モデルを一度にあてはめ

ARIMA モデルでは、ARIMA(p, d, q) の次数を、季節 ARIMA モデルでは、SARIMA(p, d, q)(P, D, Q) s の次数、期数を決める必要があります。これらの次数、期数は、自己相関プロットや偏自己相関プロットや、データの差分をとることにより推測することができますが、明確に決めるることはできないことがあります。このとき、例えば自己回帰の次数 p は 0 から 2 の範囲であると考えられるのであれば、これらの範囲を指定して、範囲内のすべてのモデルをあてはめ、適合度統計量を確認して、モデルの優先順位を決定する方法が考えられます。

JMP では、次数の範囲を指定して、複数の ARIMA モデルをあてはめる機能がありますので、本章では、この機能を用い、時系列データに対して、ARIMA モデル、季節 ARIMA モデルをあてはめてみます。

JMP のサンプルデータとして、2 章と同様のフォルダから、「Lead Production.jmp」（「鉛の生産高.jmp」）を開きます。



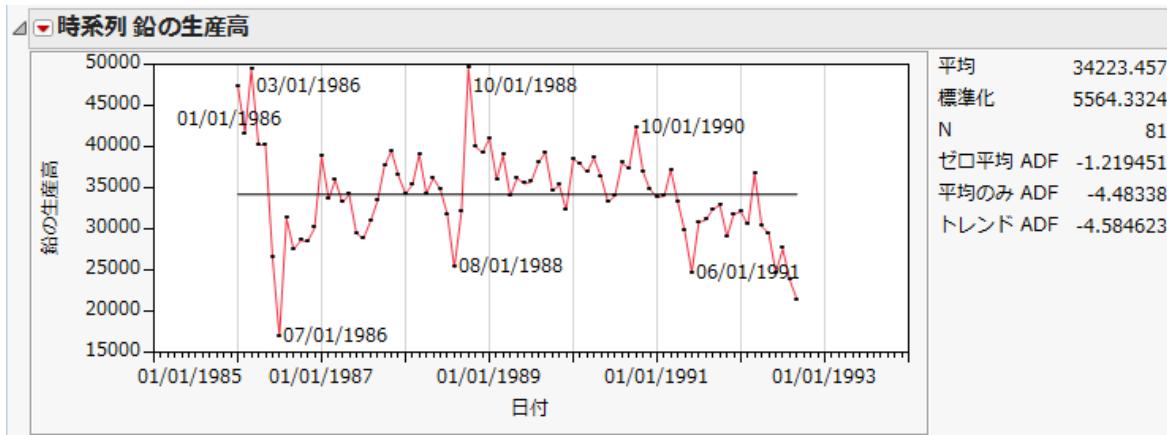
	日付	鉛の生産高
1	01/01/1986	47400
2	02/01/1986	41600
3	03/01/1986	49400
4	04/01/1986	40200
5	05/01/1986	40200
6	06/01/1986	26500
7	07/01/1986	16900
8	08/01/1986	31300
9	09/01/1986	27500
10	10/01/1986	28600
11	11/01/1986	28500
12	12/01/1986	30100
13	01/01/1987	38800
14	02/01/1987	33600
15	03/01/1987	36000
16	04/01/1987	33200
17	05/01/1987	34200
18	06/01/1987	29500
19	07/01/1987	28800
20	08/01/1987	30900
21	09/01/1987	32400

このデータは、1986 年 1 月から 1992 年 9 月まで、月ごとの鉛の生産高を記録しています。時系列分析のプラットフォームを用いて分析してみますが、分析の前に、列「日付」を選択し、[列] > [列] > [ラベルあり/ラベルなし] を選択することにより、ラベルをつけておきます。

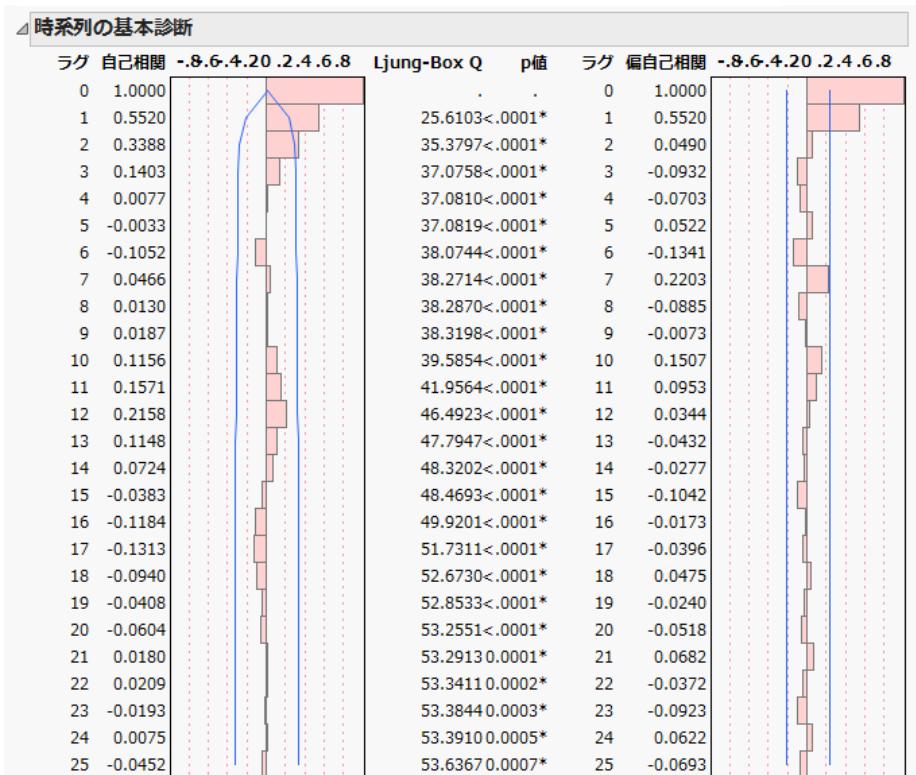
時系列分析のプラットフォームを起動し、次のように列を指定して [OK] をクリックします。



時系列グラフが表示されます。



自己相関プロット、偏自己相関プロットが出力されます。



偏自己相関プロットでは、ラグが 1 のときの偏自己相関が突出していて、ラグ 2 以降の偏自己相関は、ほぼ信頼区間にあります。そのため自己回帰の次数 p の候補は、1 になります。

自己相関プロットをみると、ラグ 1、ラグ 2、ラグ 3 と徐々に自己相関は小さくなり、ラグが 4 のときに自己相関は 0 に近くなっています。また、信頼区間より外にある自己相関は、ラグ 1、ラグ 2 のときのため、移動平均の次数 q の候補は 1 または 2 とします。差分の次数 d については、0 から 2 の範囲で検討してみます。

さらに自己相関プロットでは、ラグ 6 やラグ 12 で値が大きくなっています。このことから、データに半年または年周期の要素があることが考えられます。このグラフから季節性の次数については、判断しにくいため、季節自己回帰、季節差分、季節移動平均のそれぞれの次数(P,Q,R)の範囲を 0 から 2 の間に設定し、季節周期(s)を 6 または 12 に設定して、複数の ARIMA モデルをあてはめてみます。

レポート「時系列 鉛の生産高」の左上にある赤い三角ボタンをクリックし、[複数の ARIMA モデル] を選択します。ARIMA モデルの次数や期数を指定するダイアログボックスが表示されますので、ここに次数や期数の範囲を指定します。

まずは、半年周期であると仮定し、「季節あたりの期間数」には 6 を指定します。

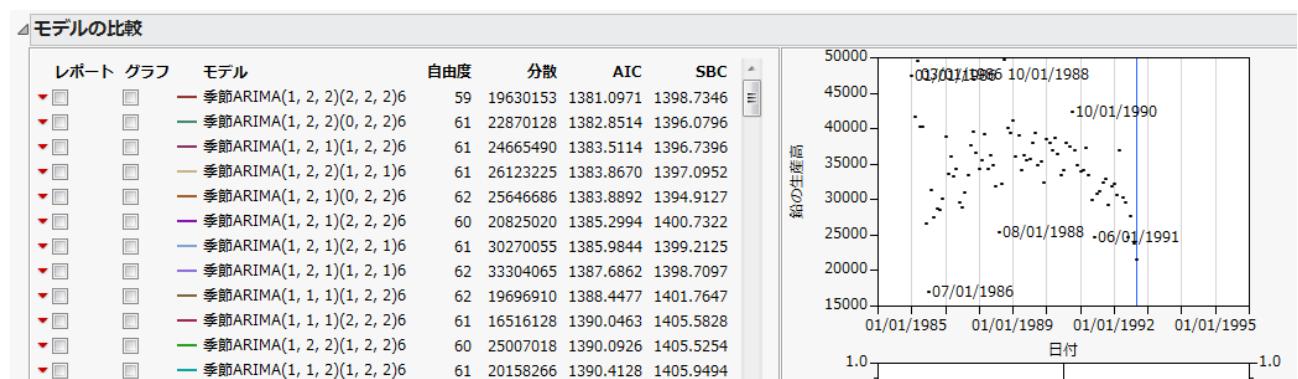
ダイアログ下には、モデルの総数 162 と表示され、162 個の ARIMA モデルをあてはめることを示しています。



[推定] ボタンをクリックすると、162 個の ARIMA モデルの推定が行われます。

注意:複数の ARIMA モデルのあてはめは、たくさんのモデルをあてはめると、より多くのメモリを消費します。

「モデルの比較」のレポートには、162 個のあてはめに対する適合度統計量が表示されます。2 章で記載したとおり、AIC の値が小さい順に並びます。レポートより、162 個のあてはめの中で、最も良いモデルは、SARIMA(1,2,2)(2,2,2) 6 であることがわかります。

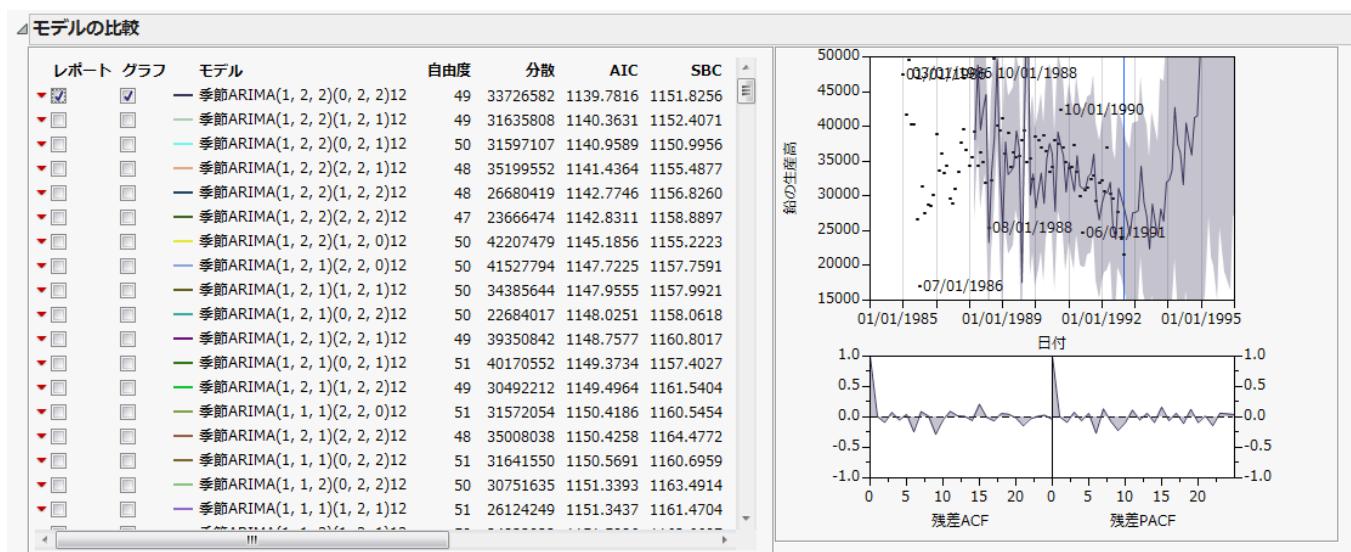


さらに、[複数の ARIMA] のオプションを用いて、期数を 12(1 年周期)に設定し、他の次数の範囲は同様の設定にして、複数の ARIMA モデルを推定してみます。



「モデルの比較」のレポートには、期数が 6 のときの結果と、期数が 12 のときの結果がまとめて表示されます。そのため、期数が 6 のときの 162 個のモデル、期数が 12 のときの 162 個のモデルの計 324 個のモデルに対する、適合度統計量が出力されています。

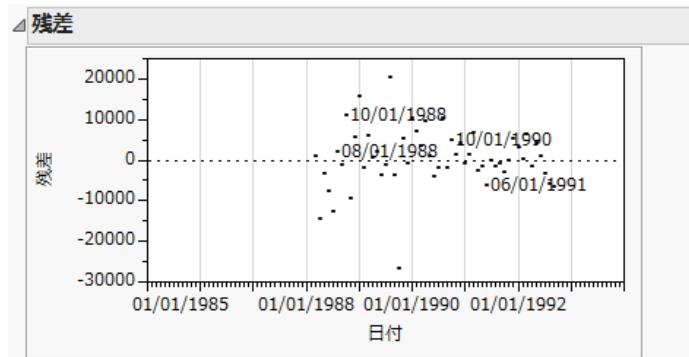
注意:あてはめるモデルの総数が多いと、それだけ多くのメモリを消費します。そのため、多くのモデルをあてはめるとメモリが不足する可能性もあります。



モデルの上位には、期数が 12 のモデルが多く含まれています。最良のモデルは、SARIMA(1,2,2)(0,2,2) 12 です。このモデルのグラフとレポートを表示させるために、左側のチェックボックスにチェックをいれています。右側のグラフは、このモデルをあてはめたときの予測値と信頼区間、残差 ACF(残差自己相関係数)、残差 PACF(残差偏自己相関係数) を表示しています。将来の予測値は、上下に変動しますが、長期的には増加傾向になるようです。

SARIMA(1,2,2)(0,2,2) 12 の「残差」のレポートを確認してみます。

下図の残差プロットでは、Y 軸をダブルクリックし、0 に参照線を引いています。



残差(の絶対値)が大きいプロットがいくつか見られますが、多くのデータが 0 を中心にランダムにばらついているようです。

残差の自己相関プロット、偏自己相関プロットを参照してみます。



いくつかのラグ値で自己相関や偏自己相関が高いところが見られますが、Ljung-Box Q 検定による p 値は、すべてのラグで有意水準である 0.05 より大きくなっています。そのため、残差にはホワイトノイズの特徴が見てとれます。

このように、JMP では、複数の ARIMA モデルを一度にあてはめを行い、モデルの評価、検証を行うことができます。