

シンポジウム6

統計モデルによる短期需要予測

荒添 悟(福岡県赤十字血液センター)

【はじめに】

日々の供給実績を記録したデータは時系列データである。

時系列データを統計処理することで、さまざまな分析を行うことが可能である。

また、時系列データに統計モデルを適用することで、未来予測へと発展させることも可能である。

過去の供給実績から将来の供給数を実用の範囲内で予測することが可能であるならば、この統計モデルによる予測は、供給計画や供給実行計画を策定する際の参考として有用である。

種々ある統計モデルの中のARIMAX (ARIMA model with exogenous variables) について、供給計画や供給実行計画を策定する際の参考としての有用性を検証する。

【対 象】

赤血球製剤を対象とする。

九州ブロック全体の供給実績(純供給数)を単位換算し、200mL献血由来と400mL献血由来とを合算して時系列データとした。

なお、供給実績(純供給数)は、血液事業情報システム情報統計の供給先別・品目別集計表より抽出・集計を行った。

使用するデータは、九州ブロックの2010年4月1日から2016年3月31日までの日次データとした。

【統計モデル】

検証する統計モデルは、ARIMAX (ARIMA model with exogenous variables) である。

外生変数(exogenous variables)を組み込んだ自己回帰和分移動平均 (ARIMA: Autoregressive Integrated Moving Average) モデルであり、外生変数として、年月日、曜日、祝日等を設定する。

また、自己回帰和分移動平均モデルとは、階差系列(Integrated) について自己回帰 (Autoregressive) モデルと移動平均 (Moving Average) モデルとを組み合わせ、さらに、残差(平均がゼロで分散が正規分布に従うランダム成分)を加えた統計モデルであり、次式で表すことができる(図1)。

統計モデルを使用した予測の計算については、フリーのソフトウェアであるR (R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.)を使用した。

【予 測】

評価期間として、2015年4月から2016年3月までの2015年度12カ月を設定した。

予測は、月々予測と一括予測を実施した。

月々予測は、予測する月の前々月までのデータを使って予測を行う。具体的には、2015年4月の

$$\Delta^d y_t = \sum_{i=1}^p \varphi_i \Delta^d y_{t-i} + \Delta^d \varepsilon_t + \sum_{j=1}^q \theta_j \Delta^d \varepsilon_{t-j}$$

ARIMAモデル

自己回帰パラメータ (p)

差分の階級 (d)

移動平均パラメータ (q)

図1 ARIMAモデルの式

予測は、2015年2月までのデータを使って予測を行う。以後、同様に2016年3月まで12回繰り返し、評価期間12カ月の予測を行う。これは、供給実行計画の策定を想定するものである。

一括予測は、2014年7月までのデータを使い2015年度の12カ月分を一括して予測し評価期間の予測とする。これは、次年度の供給計画の策定を想定するものである。

それぞれの予測データを評価期間の供給実績と比較することで評価を行う。

【評価方法】

評価方法として、2乗平均平方根誤差 (RMSE: Root mean squared error) と平均絶対パーセント誤差 (MAPE: Mean absolute percentage error) とを用いた。

それぞれ次式で表される (図2)。

予測値から観測値 (供給実績) を引いたものを予測誤差とする。予測誤差を2乗し、12カ月分の平均値の平方根が2乗平均平方根誤差 (RMSE) である。

また、予測誤差を観測値 (供給実績) で除し100を掛けたものをパーセント誤差とする。

パーセント誤差の絶対値の12カ月分の平均が平均絶対パーセント誤差 (MAPE) である。

いずれも、予測誤差が少ない場合に小さな値となる。

【結 果】

表1に評価結果を示す。

供給実行計画、ARIMAXによる月々予測、供給計画およびARIMAXによる一括予測それぞれのRMSE、MAPEそして年度計による単年度の絶対パーセント誤差 (APE) を示している。

RMSE、MAPE、APEいずれにおいても、月々予測は供給実行計画より小さな値となり、一括予測は供給計画より小さな値となっている。

このことから、月々予測は供給実行計画の策定に於いて、そして、一括予測は次年度の供給計画の策定においてそれぞれ実用的な予測精度にあるといえる。

同様に、福岡県全体と北九州事業所管轄の供給実績について、2015年度の月々予測と一括予測を行い、九州ブロック全体の結果と比較した。

表2にその結果を示している。

九州ブロック全体、福岡県全体、北九州事業所の順で、供給規模は小さくなるが、MAPEはこの順で大きくなる。

これは、供給規模が大きくなるにつれて、統計モデルによる予測精度はよくなることを示している。

なお、ここにはデータを示していないが、ARIMAXモデルの外生変数を変更することでも、予測精度は変化することがわかっている。

【考 察】

統計モデルによる予測は過去の実績 (時系列データ) に統計学的処理を行うことにより将来の予測を試みるものである。

予測の材料となるデータは過去のものであり、過去の状態を反映したものである。

現在・未来の状態は過去の状態と同じではなく、過去の状態から変化していると考えerことは自然なことである。

したがって、過去の状態を現在・未来へそのまま当てはめることはできない。

また、時系列データを統計モデルによって数式化しても、そこには必ず残差 (平均がゼロで分散が正規分布に従うランダム成分) が含まれており、予

$$RMSE = \sqrt{\text{mean}(e_i^2)}$$

$$MAPE = \text{mean}(|p_i|)$$

予測誤差 $e_i = \hat{y}_i - y_i$ [\hat{y}_i : 予測値、 y_i : 観測値 (実績)]

パーセント誤差 $p_i = 100 e_i / y_i$

図2 予測の評価式

表1 2015年度における予測の評価結果

	供給実行計画	月々予測 (ARIMAX)	供給計画	一括予測 (ARIMAX)
RMSE	1,603	625	2,441	704
MAPE	1.87	0.81	2.87	0.87
APE (年度計)	0.92	0.17	2.34	0.19

表2 供給実績の規模の違いによる予測の評価結果

	九州ブロック	福岡県	北九州
月々予測 (MAPE)	0.81	1.92	3.49
一括予測 (MAPE)	0.87	2.01	2.86

測できない誤差が存在する。つまり、統計モデルによる予測が成功するためには、少なくとも二つの条件をクリアすることが必要となる。

(1) 将来における状態の変化が少ないこと

(2) 残差が小さくなる統計モデルであること

遠い将来においては、状態の変化の程度を推定することは困難であるが、近い将来においては、その変化は小さいことが期待できる。

したがって、統計モデルによる予測は、中・長期の予測よりは、短期の予測において高い精度を示すと思われる。

また、供給実績の規模によって予測精度に差が出ることもわかるように、同じ統計モデルを使うにしても、その対象、条件設定等を工夫することで、残差が小さくなることが期待できる。

精度の高い予測を行うためには、使用する統計モデル自体も含め、さまざまな条件設定の見直しを継続していくことが必要である。

【まとめ】

今回の検討で、統計モデルによる短期需要予測

は、供給実行計画や供給計画の参考として、十分な予測精度にあることが分かった。

しかし、最新のデータを追加しながら予測を更新していくことは大事である。

また、より高い精度の予測のためには、供給実績の規模が大きければ予測精度が高い傾向にあることから、地域センター単位の予測に加え、ブロック全体としての予測を考慮することも必要である。

さらに、血液製剤の供給にかかる状態は常に変化していると考えるべきで、状態の変化に応じた統計モデルの選択を含め、予測にかかる条件設定のアップデートが重要と考える。

謝 辞

日本赤十字社九州ブロック血液センター田村和士氏には、血液事業情報システム情報統計の供給先別・品目別集計表よりデータを抽出・集計するにあたって協力をいただきました。

ここに深謝いたします。