共同研究「経営実務データを用いたデータサイエンティスト育成方法の研究」報告書

2021 年 3 月 31 日

国立大学法人滋賀大学データサイエンス教育研究センター長

竹村　彰通

１．共同研究の趣旨

本共同研究の研究題目は「経営実務データを用いたデータサイエンティスト育成方法の研究」であり、本共同研究の目的は企業の経営実務データを用いて販売予測や適性な在庫管理など、これからのデータサイエンティストに求められる能力を育成するための方法を継続的に研究することである。本共同研究の期間は、令和２年 11 月２日から令和３年３月 31 日までである。

特定非営利活動法人ビュー・コミュニケーションズは実際の販売データを教育目的のために適切に処理をおこなった上で国立大学法人滋賀大学に提供し、滋賀大学データサイエンス教育研究センターはデータを 2 年生対象の「時系列解析入門」の講義の演習の形に整備して学生の教育に活用することで、データサイエンティスト育成方法に関する共同研究をおこなった。

本年度は昨年度に引き続きビュー・コミュニケーションズから、食品スーパーにおける 5 年分の酒類の月次販売データが提供された。このデータは滋賀大学のデータサイエンス学部の学生にとって、ビッグデータの一端を実感できる実務データであり、時系列解析入門の講義でのレポート課題として用いることにより、学生達は講義で学習する時系列解析の手法がどのように実際のデータ分析に役立つかについて理解を深めることができた。このような実際のデータを用いた演習は本共同研究の目的に沿うものであり、滋賀大学データサイエンス学部が掲げる実践的なデータサイエンティスト育成の観点からも非常に有用であった。その点で本共同研究の意義は大きい。

２．滋賀大学データサイエンス学部の育成人材像と教育実績

滋賀大学データサイエンス学部では、データサイエンスの基礎要素技術としてのデータエンジニアリング(情報工学)及びデータアナリシス(統計学)に加えて、データからの価値創造の能力の育成を重視している。価値創造の能力を育成するためには、実際のデータを用いた演習の中で、試行錯誤や成功体験が必要であり、さらに教科書的な理論と実際のデータとの乖離についても経験を積む必要がある。理論と実際の乖離については、ビュー・コミュニケーションズ副理事長の小松秀樹氏の著書「なぜあなたの予測は外れるのかーAI が起こすデータサイエンス革命」においても多くの例とともに示されており、データサイエンス教育において実際のデータを扱うことが非常に重要であることがわかる。

滋賀大学データサイエンス学部は日本初のデータサイエンス学部として注目をあびていたが、2017 年 4 月の開設以来この３月にはいよいよ１期生が卒業となった。卒業生の就職状況はかなり良く、データサイエンス分野に関する社会のニーズが強いことを示している。また、価値創造に重点をおいてきた学部の教育方針は、例えば卒業論文のテーマにもよく現れている。

2021 年 2 月 4 日、5 日の二日間に渡って、データサイエンス学部第１期生の学習の総括として、卒業論文の発表会を実施した。発表会では、事前に各学生が提出した卒業論文の

内容を各自が口頭発表し、質疑を行った。発表会はオンライン会議システム ZOOM で発表資料を画面共有することで実施し、発表時間は 1 人につき、発表 7 分質疑 3 分の計 10 分とした。卒業論文の内容が、グループで研究の一部のこともあったが、この発表会では各自の貢献箇所を中心に学生が個別に発表するものとした。今回の発表会では、93 名の発表者を 11 のセッションに分け、そのうちの 1 つのセッションは、指導教員から推薦された発表を集めた特別セッションとした。

特別セッションの発表者(指導教員)と発表タイトルの一覧は表の通りである。特別セッションでは、学部連携企業の方々を外部評価委員としてお招きし、質疑に加わって頂いた。評価委員からは、「実際のデータから実際の施策立案まで落とし込めたことは非常に良かったのではないかと思います」「内容、論理性、一貫性はとても高い水準だったと思います」「身近な話題を抑揚のある巧みな説明で上手に伝えられていた」などのコメントがあり、活発に質疑が行われた。

発表者 (指導教員)	発表タイトル
森本 澪二 (佐藤 智和)	実世界の仮想化に基づく高臨場 VR 型防災教育システムの開発
水口 綾乃 (市川 治)	事前学習済み分散表現を利用した学部オープンキャンパス向け質問応答システムの構築
江口 公基 (加藤 博和)	彦根市を目的地とした観光交通における鉄道利用促進のための機関選択分析
髙田 拓弥 (松井 秀俊)	関数データに基づく回帰モデルと農業・化学分野への応用
森口 翼 (河本　薫)	テナント型商業施設における会員用スマホアプリのログデータ分析による離反防止策の検討
田室 建志 (河本　薫)	自動車部品工場における機械学習を活用した異常検知モデルの構築
上田 知展 (清水 昌平)	平和堂の ID 付き POS データを活用したモバイルクーポンの改善
小西 秀明 (清水 昌平)	ID 付き POS データを利用したモバイルクーポンの仕様改善に関する施策提案と効果検証
谷口 友哉 (清水 昌平)	購買履歴データを用いたモバイルクーポン配信の最適化
仲北 昌大 (和泉 志津恵)	2018 年 7 月の西日本豪雨災害のアンケートデータから分かる発見と問題点

この特別セッションの発表題目にみられるように、多くの卒業論文の内容は、企業等の現実のデータを分析しそこから価値を引き出すものであった。外部評価委員のコメントにみられるように、これらの内容は企業の視点から見ても実践的なものであった。

このように１期生の卒業論文には滋賀大学データサイエンス学部の教育の成果が反映されているが、このような教育方針を今後も深化させていく必要がある。ビュー・コミュニケーションズから提供されたデータを用いた演習教材のような、価値創造につながる教育コンテンツの開発は滋賀大学が今後も展開していくべき活動である。

３．滋賀大学大学院データサイエンス研究科の教育実績と拠点形成

滋賀大学ではデータサイエンス教育を大学の戦略的な方針として位置付けており、データサイエンス教育研究拠点の整備を進めてきた。2019 年 4 月には、日本初のデータサイエンス研究科修士課程を開設し、この３月には学部の１期生と同時に修士課程の１期生も卒業となった。修士課程の開設は、データサイエンス学部の開設から 2 年後のことであり、まだ学部からの卒業生が出ていない段階での「前倒し設置」であり、学部からの進学性がいないため入学者は外部からの応募者であった。定員 20 名のところを 23 名が修士課程 1 期生として入学したが、そのうち 19 名は企業派遣の社会人であった。これらの社会人は、それぞれ派遣元の企業の実際の課題を解決するための修士論文研究をおこない、2021 年 2 月 10 日及び 12 日の二日間に渡って修士論文発表をおこなった。これらの研究の多くは、直接派遣元の企業の業績向上につながるものである。

また 2020 年 4 月には博士後期課程も開設した。博士後期課程は当初の定員は 3 名と少数だが、いずれも企業所属の３名が入学し、博士課程での研究を通じて棟梁レベルのデータサイエンティストとなることを目指している。このように滋賀大学では学部から博士課程まで一貫したデータサインス教育体制が完成した。

４．本年度ビュー・コミュニケーションズより提供されたデータ

本年度にビュー・コミュニケーションズより提供されたデータは、昨年度と同様に関東圏食品スーパー100 店舗サンプルの販売金額合計値時系列データであった。販売品目は酒類 10 品目であった。データはエクセルのファイルとして、すでに整形された形で提供されており、学生はデータを R や Python などの統計分析パッケージに読み込むことで、ARIMA モデルなどの時系列解析の手法を応用することができた。このデータは時系列解析の例題としては非常に実践的なものであった。次図は提供されたエクセルファイルの一部を示したものである。

	第3のビール	ﾋﾞｰﾙ 350ml	焼酎甲類	清酒パック	国産ワイン	輸入ワイン	国産ｳｲｽｷｰ
2014年10月	64,896,265	56,458,480	41,283,034	27,710,083	15,671,871	17,012,474	11,630,853
2014年11月	62,626,576	53,164,431	42,021,722	30,565,834	17,547,875	33,758,505	12,350,601
2014年12月	61,599,319	73,810,979	44,625,092	37,739,137	20,952,974	25,209,042	14,292,028
2015年1月	56,903,102	54,636,556	39,416,436	28,568,680	15,394,759	18,412,583	13,251,262
2015年2月	55,275,135	42,429,178	37,706,181	28,192,179	14,140,309	16,260,936	13,442,114
2015年3月	61,832,614	55,360,190	41,023,819	28,121,201	15,395,572	16,849,718	15,683,896
2015年4月	65,137,971	55,137,113	39,079,248	25,829,663	13,980,098	16,657,483	15,533,525
2015年5月	95,580,312	75,617,063	43,269,560	22,720,433	13,507,734	17,591,824	14,537,782
2015年6月	88,410,801	64,505,572	55,492,424	20,548,710	11,904,098	15,005,699	13,854,597
2015年7月	96,231,198	77,867,879	39,422,324	19,173,130	11,388,621	14,482,564	13,158,490
2015年8月	95,740,071	93,823,473	38,529,894	18,322,581	10,929,671	14,653,894	13,122,023
2015年9月	85,365,554	66,422,417	38,211,951	22,253,526	12,521,920	15,458,816	12,324,527
2015年10月	87,278,610	61,513,467	40,727,491	27,928,000	16,921,779	19,699,321	13,375,270
2015年11月	80,622,982	57,462,902	41,008,635	31,193,441	17,679,276	34,439,141	13,335,337
2015年12月	80,550,355	84,451,103	43,382,753	37,712,380	20,849,978	26,678,527	15,593,571
2016年1月	74,896,265	63,204,148	39,193,684	30,669,160	15,611,565	19,267,511	13,578,948
2016年2月	76,489,294	49,531,205	37,564,149	29,950,650	14,643,939	18,477,289	13,249,479
2016年3月	84,751,597	64,607,404	40,332,548	30,191,077	14,764,217	19,224,918	14,284,018
2016年4月	85,530,588	66,563,713	38,443,524	27,844,328	14,134,252	19,286,489	15,331,135
2016年5月	94,723,466	78,050,551	42,703,326	25,602,536	14,566,449	21,522,904	16,997,167
2016年6月	90,770,895	71,194,568	51,544,702	22,761,122	12,312,624	17,275,704	18,719,842
2016年7月	99,618,096	84,006,041	39,677,949	21,045,965	11,559,434	16,372,270	19,395,526
2016年8月	97,794,657	99,688,086	38,734,847	19,953,496	10,942,939	15,548,213	19,338,026
2016年9月	89,843,497	67,595,129	37,984,195	22,357,293	12,079,329	16,311,333	19,512,245

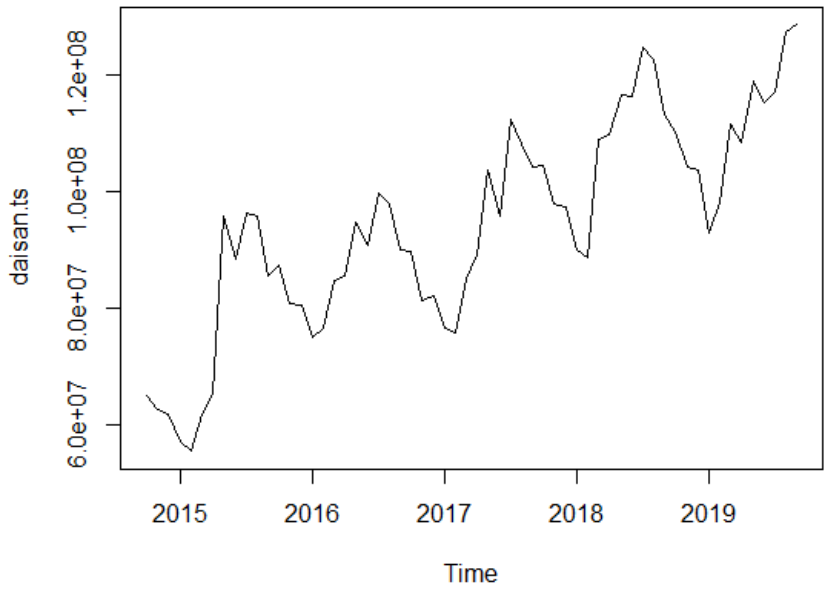
本年度提供データの一部

５．学生によるデータ分析の例

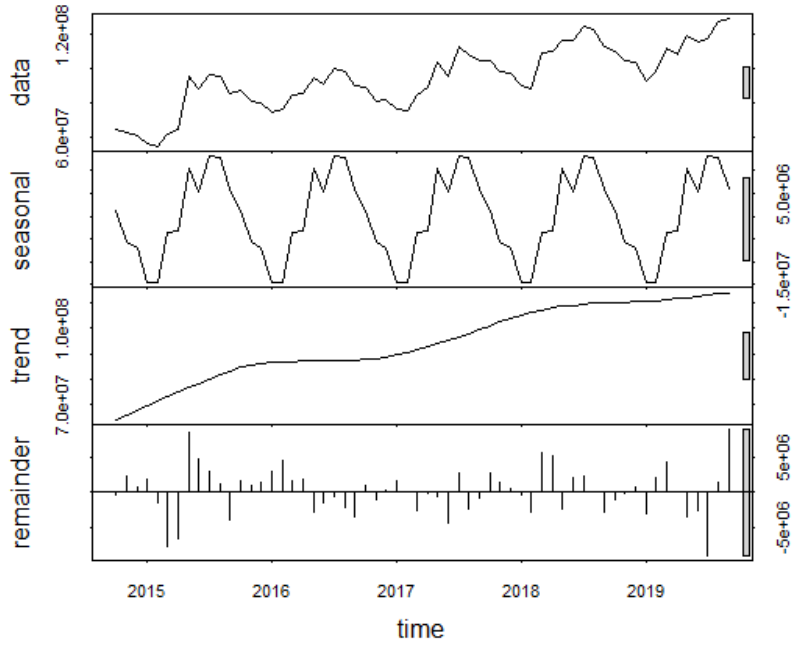
レポート課題では、ビュー・コミュニケーションズより提供された 10 品目のデータの中から３品目を選んで解析するよう学生に求めた。以下は第３のビールに関する浅野菜月さんの解析例である。

（１）第３のビールについて

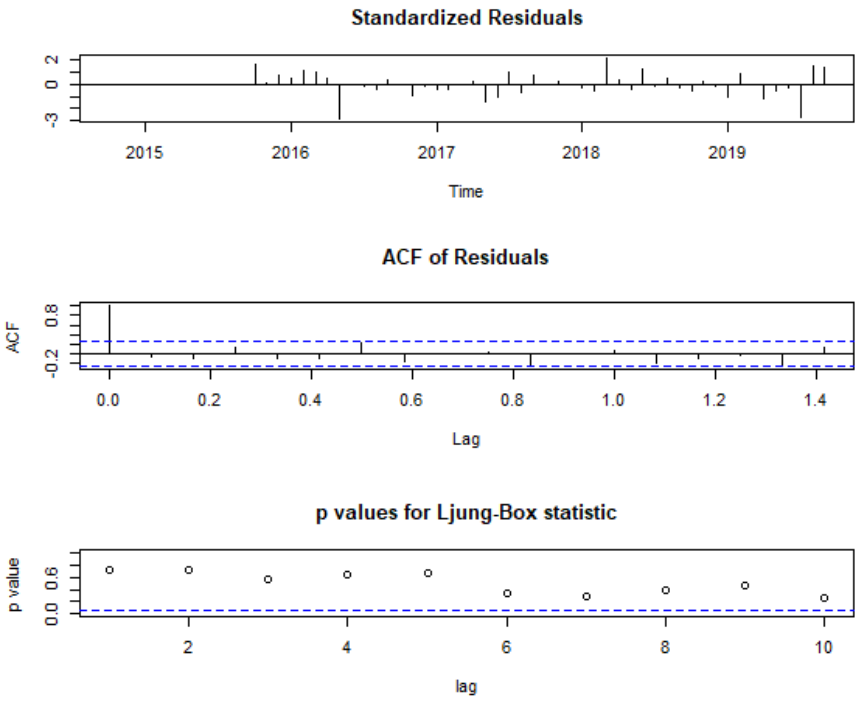
```
daisan <- scan("daisan.dat")
daisan.ts <- ts(daisan, st =c(2014,10),end = c(2019,9), fr = 12)
plot(daisan.ts)
```



daisan.stl<-stl(daisan.ts,s.window="per")
plot(daisan.stl)
推定された傾向成分より、第３のビールは近年増加傾向にあることが分かる。



daisan.arima<-auto.arima(daisan.ts,stepwise=T,trace=T)
tsdiag(daisan.arima)
以下の結果から、自己相関も小さく、誤差らしい誤差（ホワイトノイズ）であるということが分かり、よいモデルといえる。

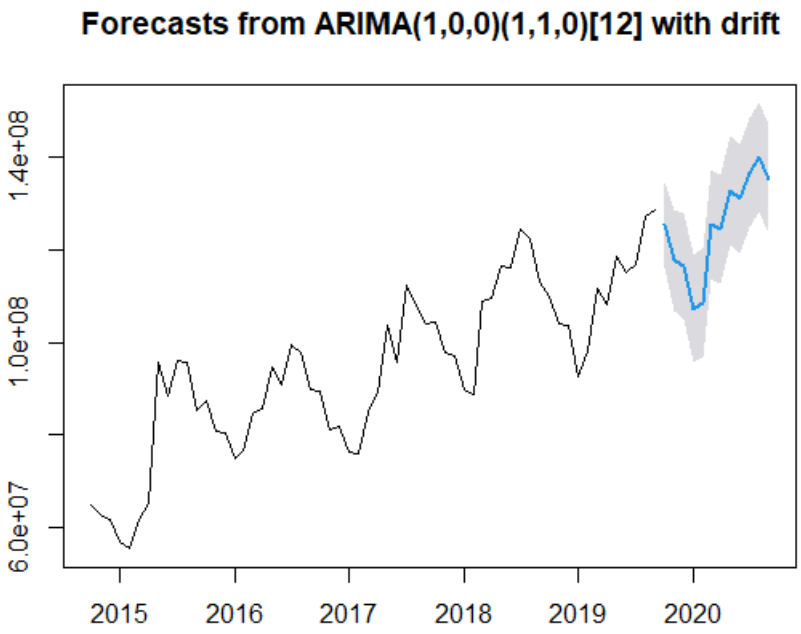


1 年先の予測をしてみると、

```
daisan.pred<-forecast(daisan.arima,level = 95,h=12)
class(daisan.pred)
daisan.pred
```

	Point Forecast	Lo 95	Hi 95
Oct 2019	125419363	116068454	134770271
Nov 2019	117881430	106918635	128844226
Dec 2019	116599663	105091219	128108106
Jan 2020	107263155	95556927	118969384
Feb 2020	108559279	96779841	120338717
Mar 2020	125676803	113870067	137483539
Apr 2020	124585109	112768167	136402051
May 2020	132980797	121160035	144801559
Jun 2020	131187141	119364950	143009333
Jul 2020	136690302	124867575	148513029
Aug 2020	140030733	128207806	151853661
Sep 2020	135191841	123368838	147014844

```
plot(daisan.pred)
```



1 年先もこれまでと同じような季節の影響等を受けながら、増加していくということが分かる。グレーの幅が信頼区間を示しているが、このグラフでは全体的に幅がほぼ同じように見えるが、下に大きく下がっている部分など所々幅が大きいところも見られる為、大きく増減しているところに関しては少し推定の精度が低いといえるのではないだろうか。

このように、浅野さんの分析は、データのプロットから始め、R の forecast パッケージで提供されている auto.arima コマンドを用いて最適な SARIMA モデルを選び、選ばれたモデルの当てはまりまでを検討しており、標準的な分析例となっている。

６．まとめ

前節でみたように、提供されたデータの分析のレポート課題に多くの学生が熱心に取り組んだ。その理由は、提供されたデータが非常に実践的で、また時系列解析の効果が実感できるものであったことである。

滋賀大学データサイエンス学部としては、今後もさらに興味深い演習課題をビュー・コミュニケーションズとの共同研究の形で開発して行きたいと考えている。修士課程の教育にも活用していきたい。さらに、滋賀大学は数理及びデータサイエンスに係る教育強化の拠点校として文部科学省より選定されており、このような教育コンテンツを他大学に展開していくことも重要な課題である。

（追加報告）

事例研究：ドリフト付き季節 ARIMA モデルの課題

2021 年 3 月 31 日

特定非営利活動法人　ビュー・コミュニケーションズ

副理事長　小松　秀樹

本稿は、首都圏Aエリアにおける第三のビール月次販売額データ（5年、スーパーマーケット）を用いて学生が行った予測分析に対して、実務的評価を行ったものである。

専門的技法の実務適用は有用であるが、技法にこだわりすぎても弊害が生じる。

本事例研究はこのような視点から書かれたものである。

学生が行った予測（R言語によるARIMA自動選択）の概要を振り返り、どのような問題があったかを検討した。また、より有用性の高いARIMAモデルの適用方法についても触れることとした。

(1)　学生が行った予測概要

(2)　予測の理論背景と複雑化の罠

(3)　予測の問題点

(4)　実務的予測とは

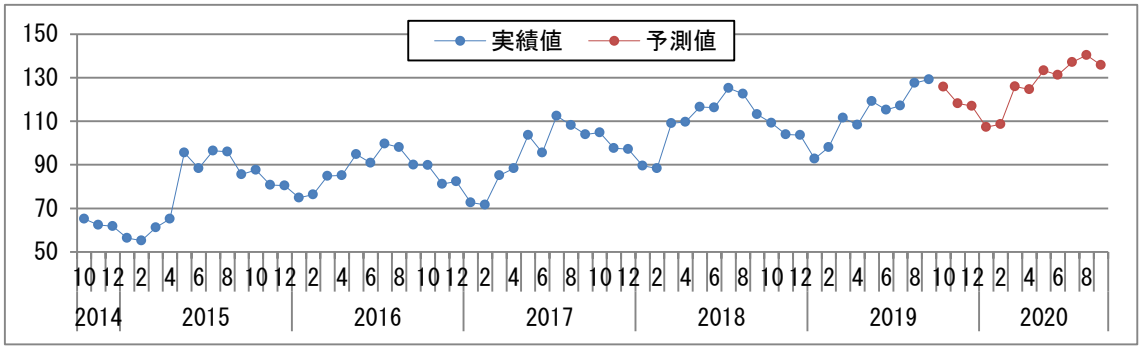
(1) 学生が行った予測概要

学生がＲ言語を用いて行ったSTL分解、及びARIMAモデルを用いた予測では、
ARIMAモデルの次数について

```
ARIMA(1,0,0)(1,1,0)[12] with drift
```

と示されている通り、前年同月差分のAR(1)モデルにドリフト（定数項）を加えたモデルとなっていた。

その予測結果は以下のようであった。

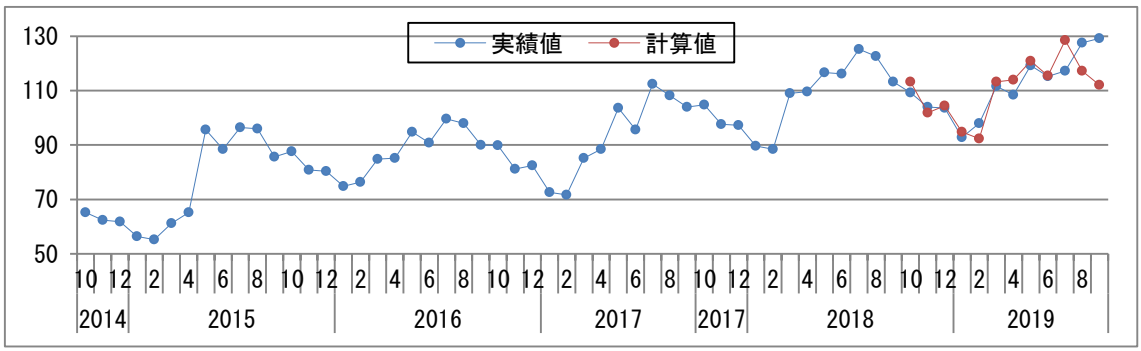


学生が予測を行ったのは2019年10月～2020年9月で実績値が不明なため事後評価ができない。そこで同モデルを用いて2018年10月～2019年9月の予測を行い実績値との比較を行った。結果は下図表にあるように誤差の状況から比較的良好な予測であった。

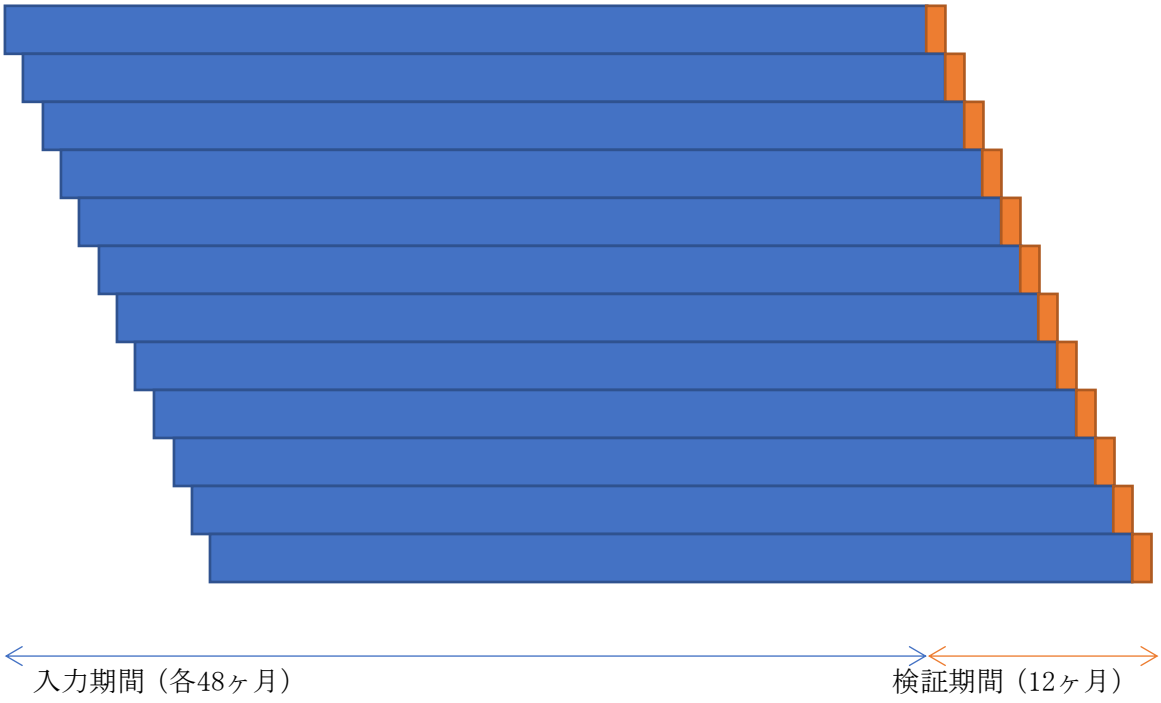
しかし、このモデルに潜む予測技法の罠が実務的には問題となる。

学生のモデル（Rのauto.arima）

	2018年			2019年									合計
	10月	11月	12月	1月	2月	3月	4月	5月	6月	7月	8月	9月	
計算値	113.2	101.9	104.4	94.7	92.3	113.2	114.0	120.9	115.5	128.5	117.2	112.1	1327.9
実績値	109.2	104.0	103.6	92.8	98.0	111.6	108.4	119.2	115.2	117.2	127.6	129.2	1336.0
誤差	4.0	-2.1	0.8	1.9	-5.7	1.6	5.6	1.7	0.3	11.3	-10.4	-17.1	-8.1
誤差率	3.7%	-2.0%	-0.8%	2.1%	-5.8%	1.4%	5.1%	1.4%	0.2%	9.6%	-8.1%	-13.3%	-0.6%



※下図のように、48ヶ月データを元に次の一ヶ月予測、を12回繰り返し
　予測精度の検証を行った結果である。



１回目：2014年10月～2018年9月の48ヶ月のデータを元に2018年10月の予測値を算出。
２回目：2014年11月～2018年10月の48ヶ月のデータを元に2018年11月の予測値を算出。
（中略）
12回目：2015年9月～2019年8月の48ヶ月のデータを元に2019年9月の予測値を算出。

(2) 予測の理論背景と複雑化の罠

まずは学生の予測モデル（Rの自動選択）の理論背景を簡単に見てみる。

Rの

ARIMA(1,0,0)(1,1,0)[12] with drift

とはドリフト付き季節ARIMAを示している。

一般に、系列{X_t}が周期sのドリフト付き季節ARIMA(p,d,q)(P,D,Q)[s]過程であるとは、

$$Y_t = X_t - \delta t$$

$$Z_t = (1-L)^d(1-L^s)^D Y_t \qquad \cdots(1)$$

とするとき、系列{Z_t}が

$$\phi(L)\Phi(L^s)Z_t = \theta(L)\Phi(L^s)\varepsilon_t$$

のようなARMA過程であることである。ここで、δはドリフト率、{ε t}は平均0,分散σ²のホワイトノイズ、Lはラグ演算子（$L^i Z_t = Z_{t-i}$）、

$\phi(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$、$\Phi(z) = 1 - \sum_{i=1}^{P} \Phi_i z^i$、$\theta(z) = 1 + \sum_{i=1}^{q} \theta_i z^i$、$\Theta(z) = 1 + \sum_{i=1}^{Q} \Theta_i z^i$

である。

今回は、ドリフト付きARIMA(1,0,0)(1,1,0)[12]過程であるので、

$$Y_t = X_t - \delta t$$

$$Z_t = (1 - L^{12})Y_t = Y_t - Y_{t-12} = (X_t - \delta t) - \{X_{t-12} - \delta(t-12)\} = X_t - X_{t-12} - 12\delta$$

とするとき、系列{Zt}は

$(1 - \phi L)(1 - \Phi L^{12})Z_t = \varepsilon_t$　　但し$\{\varepsilon_t\}$は平均0、分散σ²のホワイトノイズ

のようなARMA過程に従う。

```
ARIMA(1,0,0)(1,1,0)[12] with drift
Coefficients:
         ar1     sar1   drift
      0.6375  -0.5519  0.8263
s.e.  0.1323   0.1615  0.1072
sigma^2 estimated as 23.31:  log likelihood=-144.57
AIC=297.14   AICc=298.07   BIC=304.63
```

左はR自動選択結果で、パラメータの対応は次のようになる。

φ=ar1=0.6375

Φ=sar1=-0.5519

δ=drift=0.8263

σ²=23.31

これらのパラメータをARMA過程の式に当てはめて

$(1 - 0.6375L)(1 + 0.5519L^{12})Z_t = \varepsilon_t$　　但し$\{\varepsilon_t\}$は平均0、分散23.31のホワイトノイズ

を得る。

変形すると

$$Z_t = 0.6375Z_{t-1} - 0.5519Z_{t-12} + 0.6375 \times 0.5519Z_{t-13} + \varepsilon_t$$

より、“ドリフト付き”差分系列$\{Z_t\}$のhヶ月後の予測値$\hat{Z}_{60+h}$は

$$\hat{Z}_{60+h} = \begin{cases} 0.6375Z_{60+h-1} - 0.5519Z_{60+h-12} + 0.3518Z_{60+h-13}, & h = 1 \\ 0.6375\hat{Z}_{60+h-1} - 0.5519Z_{60+h-12} + 0.3518Z_{60+h-13}, & 2 \le h \le 12 \\ 0.6375\hat{Z}_{60+h-1} - 0.5519\hat{Z}_{60+h-12} + 0.3518Z_{60+h-13}, & h = 13 \\ 0.6375\hat{Z}_{60+h-1} - 0.5519\hat{Z}_{60+h-12} + 0.3518\hat{Z}_{60+h-13}, & h \ge 14 \end{cases}$$

となる。また、$Z_t = X_t - X_{t-12} - 12\delta$の関係により、差分系列を戻すと（“ドリフト付き”

和分処理)、原系列$\{X_t\}$のhヶ月後の予測値$\hat{X}_{60+h}$は

$$\hat{X}_{60+h} = \begin{cases} \hat{Z}_{60+h} + X_{60+h-12} + 9.9156, & h \leq 12 \\ \hat{Z}_{60+h} + \hat{X}_{60+h-12} + 9.9156, & h \geq 13 \end{cases}$$

となる。

※今回はMA次数q、SMA次数Qが共に0であったが、q,Qのいずれかが1以上の場合は、ホワイトノイズ$\{\varepsilon_t\}$の推定も必要になる。

例：MA(1)過程（ドリフト無し）の場合、$X_t = \varepsilon_t + \theta\varepsilon_{t-1}$というモデル式になるので、

$$\hat{X}_{N+h} = \begin{cases} \theta\varepsilon_N, & h = 1 \\ 0, & h \geq 2 \end{cases}$$

となる。ε_Nの値を推定するには、

$$\hat{X}_i = \theta\varepsilon_{i-1}, \qquad 1 \leq i \leq N$$

$$\varepsilon_i = X_i - \hat{X}_i, \qquad 1 \leq i \leq N$$

を用いて、ε_0の値も推定しなければならない。

（簡潔には、$\varepsilon_0 = 0$（ホワイトノイズの平均値＝期待値)、$\varepsilon_1 = X_1$として、$|\theta| < 1$ならばε_0, ε_1が$\hat{X}_{N+1}$に与える影響は十分小さくなる、と考えられる）

注：前々頁の(1)の式を展開すると、

(d,D)=(1,0)のとき$Z_t = X_t - X_{t-1} - \delta$、

(d,D)=(0,1)のとき$Z_t = X_t - X_{t-s} - \delta s$となる。

d+D≧2のときはZ_tからδの項が消え、$Z_t = (1-L)^d(1-L^s)^D X_t$となる。

例：(d,D)=(2,0)のとき、

$$Z_t = (1-L)^2 Y_t = (1-L)(1-L)(X_t - \delta t) = (1-L)\{X_t - \delta t - X_{t-1} + \delta(t-1)\}$$

$$= (1-L)(X_t - X_{t-1} - \delta) = X_t - X_{t-1} - \delta - X_{t-1} + X_{t-2} + \delta$$

$$= X_t - 2X_{t-1} + X_{t-2}$$

一方、$(1-L)^2 X_t = (1-L)(X_t - X_{t-1}) = X_t - X_{t-1} - X_{t-1} + X_{t-2} = X_t - 2X_{t-1} + X_{t-2}$

さて、ドリフト付き季節ARIMAは変動量をトレンド成分、季節成分、規則的・不規則的成分からなると考え、モデルのパラメータを推定しようとするものである。

ここで、モデル評価の基本的視点AICと現実的変動特性からこのモデルについて考えてみる。

赤池情報量基準AICは

$$\mathrm{AIC} = -2\ln L + 2K$$

Lは最大尤度、Kは推定パラメータ数（ホワイトノイズ分散も推定パラメータに含む）

と定義されモデルの基礎的な判定基準である。

厳密性と簡便性を併せ持った基準であるが、モデル精度を高めようとすればする程複雑なモデルになりやすくパラメータも増える。なるべくシンプルなモデルで精度が高いことが理想的モデルとなる。

規則・不規則成分から構成されるARIMAモデルと今回のドリフト付き季節ARIMAモデルを比べるとドリフト（トレンド）と季節性がモデルを複雑化している。

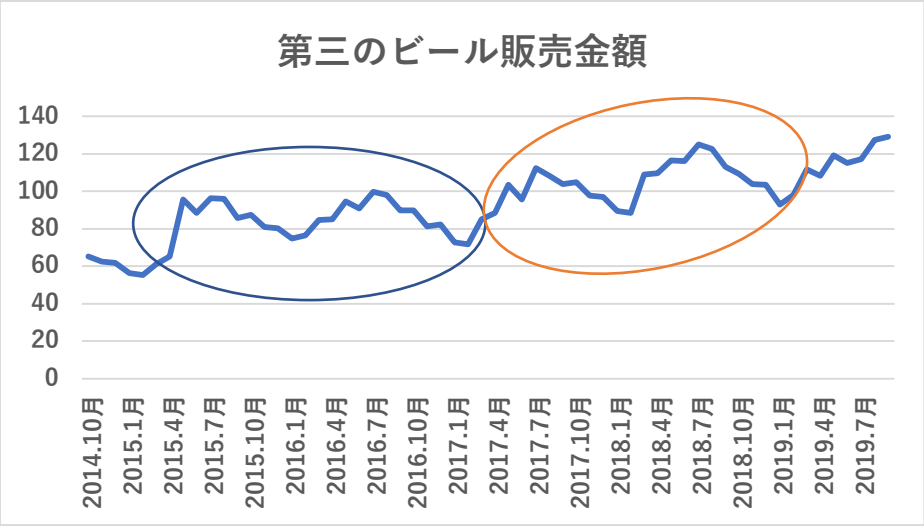
ARIMAでは、差分系列を工夫することでトレンド、季節性を考慮できる。ただし、トレンド、季節性が明瞭かつ定量的に存在する場合はモデルの厳密性（尤度が低い）に欠ける。

トレンド、季節性の現実的特性は、現代市場の多くで複雑化・不確実化が進み変化のスピードも高まっているので、明確なトレンド、季節性が観測しにくくなっている。地球温暖化の影響かもしれないが、極端な天候不順も多くなり季節性が崩れてきているようにも見える。

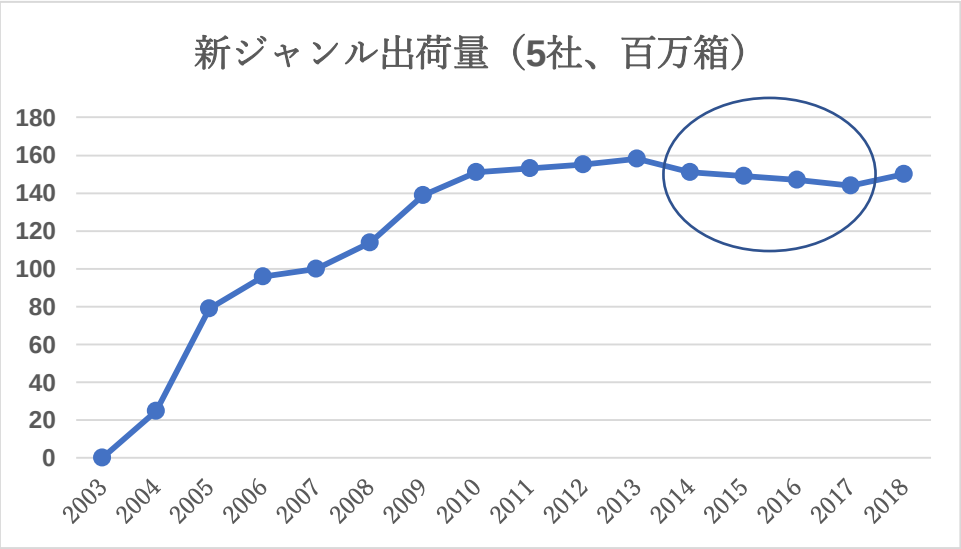
学生がより高度・複雑な予測モデルを学習し使いたくなるのもうなずけるところであるが、情報量基準、現実的適合性から思わぬ落し穴（罠）が見え隠れする。

（3） 予測の問題点

今回予測の最大の問題はモデルにドリフト（トレンド）成分を線形的に組み込んだ点にある。より詳しく実データを見てみよう。



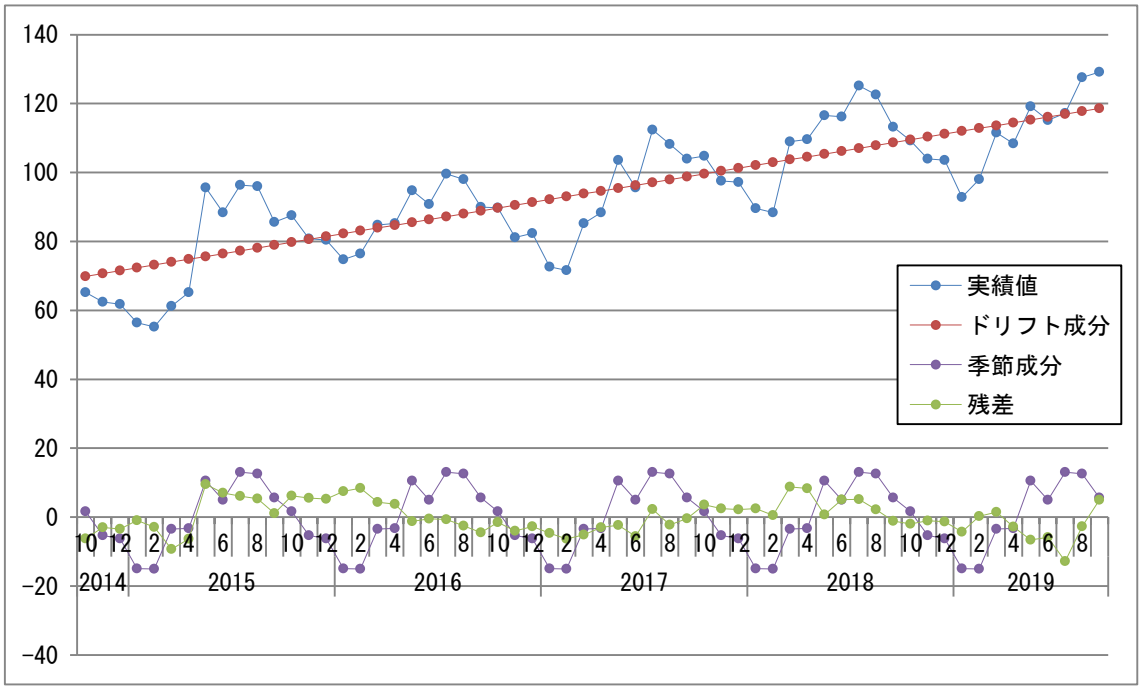
販売トレンドは、2015年春～2017年冬はほぼ横ばい、2017年春～2018年夏は単調増加傾向と伺える。全期間にわたり線形回帰線（トレンド線）を引くのは無理がある。
更に、限定されたスーパーマーケット市場ではなくビール大手5社の出荷量を見るとより分かりやすいかもしれない。



このように2014年～2017年はやや下降傾向で2018年に反転増加したように見える。
現実世界を線形的に捉えると、しばしば現実との乖離が大きくなると考えるのべきである。
次に、ドリフト（トレンド）予測量が全体予測量に占める大きさを考えてみよう。

　予測＝トレンド+季節+ARIMA

という単純な加法モデルを用いて見てみた。



それぞれの成分の大きさ上図の様になり、トレンド成分が予測値の大部分を占めてしまい、季節成分や規則・不規則成分の予測に占める割合は小さくなる。もし、季節成分、規則・不規則成分のモデル精度が高くてもトレンド成分のモデル精度が低いと、全体の精度は高くなりにくい。

もともとトレンドはかなり曖昧な概念であり、トレンドの線形回帰は変化の激しい現代市場では適合性が悪くなる。

この場合は、トレンド成分の誤差を季節成分、規則・不規則成分の誤差で相殺し、結果的に一定の予測精度に保たれたと考えられる。

(4) 実務的予測とは

実務的には予測を行う際はなるべく理論的仮定を少なくしてシンプルなモデルを構築するように努めるのが肝要である。時系列予測については差分系列に対するARMAモデルを用いることが望ましい。トレンド、季節性が不明瞭な場合は、様々な差分系列を作成しそれぞれについてパラメータ推定をし各種判定基準を用いてベストモデルを選択する。もし明確な季節性がある場合は、7日差分、12ヶ月差分のような限定された差分系列に対しARMAモデルを構築すれば良い。

この事例では、「ここ最近第三のビールの売れ行きが好調」、「夏場にはビール出荷量が伸びる」といった認識が一般的で、明確なトレンド、季節性を定義するには心もとない

ケースが多い。

実務的には考えられる全ての差分系列を検討することとなる。

繰り返しになるがトレンドや季節性に強い仮定は設けていない点が留意点である。

全ての差分系列のなかで選択されたARIMAモデルは

　前年同月差分+前月差分に対するARMAモデル（但し原系列に対数変換）

であった。

このモデルの予測結果は下図表のようになった。

	2018年			2019年									合計
	10月	11月	12月	1月	2月	3月	4月	5月	6月	7月	8月	9月	
計算値	114.5	102.1	103.4	95.5	91.8	119.9	117.1	116.7	117.4	125.5	117.0	113.0	1333.9
実績値	109.2	104.0	103.6	92.8	98.0	111.6	108.4	119.2	115.2	117.2	127.6	129.2	1336.0
誤差	5.3	-1.9	-0.2	2.7	-6.2	8.3	8.7	-2.5	2.2	8.3	-10.6	-16.2	-2.1
誤差率	4.9%	-1.9%	-0.2%	2.9%	-6.3%	7.5%	8.1%	-2.1%	1.9%	7.1%	-8.3%	-12.5%	-0.2%

全体としては良好な予測結果となっている。少しコメントすれば、2019 年 9 月の誤差が大きく出ているのは例年に比べやや異常に販売が増加しており、当年当月の特殊事情があったものと推測される。過去データの規則性からは妥当な予測値と評価されよう。

この結果は一見すると学生の予測結果と大差ない。しかし学生のモデルは線形トレンドから現実世界が乖離すると大きな誤差が生じかねない欠点を持っている。

トレンド成分を除去して予測する方法（差分系列）が実務では採用されることになる。

末尾になったが、理論学習、専門的ソフト実習を行う学生にとり、このような視点を考えておくことも将来的に役立つことと思われる。

期待を込めて。