

飲食店の売上予測

慶應義塾大学経済学部

長倉大輔研究会

佐藤 諒 濱根直志 森 顕士

要旨

ここ数年のもっとも大きな出来事のひとつに新型コロナウイルスの流行が挙げられるだろう。その被害を直接的に受けたのが外食産業で、各飲食店の売上は軒並み低下するなどして大きく変化した。そこで、新型コロナウイルス流行時を含む数年のデータを分析することで飲食店の今後の売上を予測することを試みた。分析手法として重回帰分析と時系列分析を用いた。重回帰分析では売上を被説明変数として、有意に説明する変数を探していった。時系列分析では数年分の売上の変動にだけ注目して、トレンドやパターンを見つけ出そうとした。最後にそれぞれの手法の比較、考察を行い、今後の向けての展望を行った。

目次

第 1 章	はじめに
第 2 章	先行研究の紹介
第 3 章	重回帰分析による売上に寄与する要因の分析
3-1	使用データ
3-2	モデルデータ
3-3	分析結果
第 4 章	時系列分析、特に SARIMA モデルを用いた今後の売上予測
4-1	SARIMA モデル概略
4-2	分析結果
第 5 章	2 つの分析からの考察
5-1	考察：重回帰分析
5-2	考察：時系列分析
第 6 章	今後の展望
謝辞	
参考文献	

第 1 章

はじめに

2020 年頃、新型コロナウイルスが日本列島に流入し、猛威を振るったことは記憶に新しいだろう。

新型コロナウイルスは人々の間で感染を拡大させただけでなく、新型コロナウイルスの感染の拡大を防ぐために政府から出された緊急事態宣言やまん延防止等重点措置などによって日本の経済、その中でも、観光業や飲食業に大きな被害を与えた。廃業せざるを得ない状況まで追い込まれた経営者も少なくなかっただろう。新型コロナウイルスの感染拡大が落ち着いてきた現在は、営業形態に幾分かの変更を加えながらもコロナ禍以前の営業をある程度実施出来ているだろう。しかし、実際にコロナウイルスによってどれ程の被害を受けたのかはイメージしづらい。

そこで、都内の某飲食店にコロナ禍以前から現在までの売上や来客数などを記録しているデータをいただく事が出来たため、このデータを用いて、新型コロナウイルスが与えた飲食店への実際の影響、コロナ禍を抜けた後の売上の予測、それだけでなく、飲食店の売上に寄与する要因の分析などを、重回帰分析ならびに時系列分析等の手法を用いて分析を行った。

第 2 章

先行研究の紹介

今回、論文の執筆に際して、コロナ禍の影響で売上が大きく落ち込んだとする『日本政策金融公庫 調査月報』（星田 2022.4 p4~11）の記述から着想を得た。

当記述によると、飲食業や宿泊業は約 92%の企業が 2019 年の売上に比べて 2020 年の売上が減少したとし、翌年も 2020 年の売上に比べて 2021 年の売上が減少したとする企業は約 79%もあると述べている。また、約 4 分の 3 の企業が元の売上水準に回復するまで一年以上かかるもしくはわからないと回答していることがわかる。

前章でも述べたが、コロナ禍によって様々な業界が大きな痛手を受けたのは印象だけでなく、黒田の記述にあるようにデータからも明確である。そこで、どれほどの痛手を受けたのか、また、コロナ禍から約 2 年程経った現在、どの程度まで売上の水準を回復させているのか、といったことを普段学習している重回帰分析や時系列分析といった統計的な手法を、ある 1 店舗の売上に関するデータに用いることで数値化して具体的に可視化することを試みるべく論文を執筆した。

第3章

重回帰分析を用いた売上に寄与する要因の分析

3-1 使用データ

都内の某飲食店の2020年1月～2023年4月までの売上や来客数などが記録されているデータを使用する。また、気温と降水量に関しては、気象庁の2020年1月～2023年4月までの東京都の1日の平均気温と降水量のデータも使用する。某飲食店は夜の居酒屋としての売上がメインだが、昼時にはランチメニューを提供していることもある。

3-2 モデル設定

重回帰分析を行うモデルは下記のように設定した。このモデルでは、説明変数を1単位増やしたときに売上がどれくらい増えるかが分かる。

$$\text{Sale} = a + b \cdot \text{temp} + c \cdot \text{cost} + d \cdot \text{rain} + e \cdot \text{mon} + f \cdot \text{tue} + g \cdot \text{wed} + h \cdot \text{thu} + i \cdot \text{fri} + j \cdot \text{sat} + k \cdot \text{sun} + l \cdot \text{sengen} + m \cdot \text{manbo} + \varepsilon$$

Sale : 総売上からランチ・ハッピーアワーの売上を引いたもの

temp : 1日の平均気温

cost : 1日当たりのアルバイトの人件費の合計

rain : 1日の総降水量

mon : 月曜日ダミー

tue : 火曜日ダミー

wed : 水曜日ダミー

thu : 木曜日ダミー

fri : 金曜日ダミー

sat : 土曜日ダミー

sun : 日曜日ダミー

sengen : 緊急事態宣言ダミー

manbo : まん延防止等重点措置ダミー（以降、まん延防止ダミー）

非説明変数を（夜の居酒屋としての）売上とし、説明変数に気温、降水量、人件費、曜日ダミー、緊急事態宣言とまん延防止ダミーを設定し、それを立式したものである。

多重共線性の問題を避けるため、モデルから月曜日ダミー変数を減らしたも

のを最終的なモデルとした。最終的なモデルは下式である。

$$\text{Sale} = a + b \cdot \text{temp} + c \cdot \text{cost} + d \cdot \text{rain} + e \cdot \text{tue} + f \cdot \text{wed} + g \cdot \text{thu} + h \cdot \text{fri} \cdot \text{sat} + i \cdot \text{sun} + j \cdot \text{sengen} + k \cdot \text{manbo} + \varepsilon$$

また、被説明変数、説明変数ともに自然対数をとったモデルも設定した。このモデルでは、説明変数が1%増えた時に売上が何%増えるかが分かる。

$$\ln \text{Sale} = a + b \cdot \ln \text{temp} + c \cdot \ln \text{cost} + d \cdot \ln \text{rain} + e \cdot \text{mon} + f \cdot \text{tue} + g \cdot \text{wed} + h \cdot \text{thu} + i \cdot \text{fri} + j \cdot \text{sat} + k \cdot \text{sun} + l \cdot \text{sengen} + m \cdot \text{manbo} + \varepsilon$$

$\ln \text{Sale}$: Sale の自然対数をとったもの

$\ln \text{temp}$: temp の自然対数をとったもの

$\ln \text{cost}$: cost の自然対数をとったもの

$\ln \text{rain}$: rain の自然対数をとったもの

3-3 分析結果

上記の分析方法で分析した結果の中で、有意でない変数が最も少なかった結果は以下の通りである。

$$\begin{aligned} \text{Sale} = & 190440.902 - 3409.646 \times \text{temp} + 1.413 \times \text{cost} - 268.110 \times \text{rain} \\ & (7103.910) \quad (224.767) \quad (0.215) \quad (128.389) \\ & + 12101.079 \times \text{tue} + 12101.079 \times \text{wed} + 15247.770 \times \text{thu} \\ & (6429.313) \quad (6474.813) \quad (6567.245) \\ & + 75830.373 \times \text{fri} + 60358.290 \times \text{sat} + 15694.578 \times \text{sun} \\ & (6443.346) \quad (6427.920) \quad (6669.476) \\ & - 119064.585 \times \text{sengen} - 113908.443 \times \text{manbo} \\ & (4885.047) \quad (6522.927) \end{aligned}$$

() 内は標準誤差。

決定係数 $R^2 = 0.600$

各係数を見ると、定数項、気温、バイトの人件費、金曜日ダミー、土曜日ダミー、緊急事態宣言ダミー、まん延防止ダミーは1%有意であり、降水量、木

曜日ダミー、日曜日ダミーは 5 % 有意であり、水曜日ダミーは 10 % 有意であり、火曜日ダミーのみ有意ではなかった。

求めた式に各変数の数値を代入して、実際のデータと比較した結果は以下の通りである。

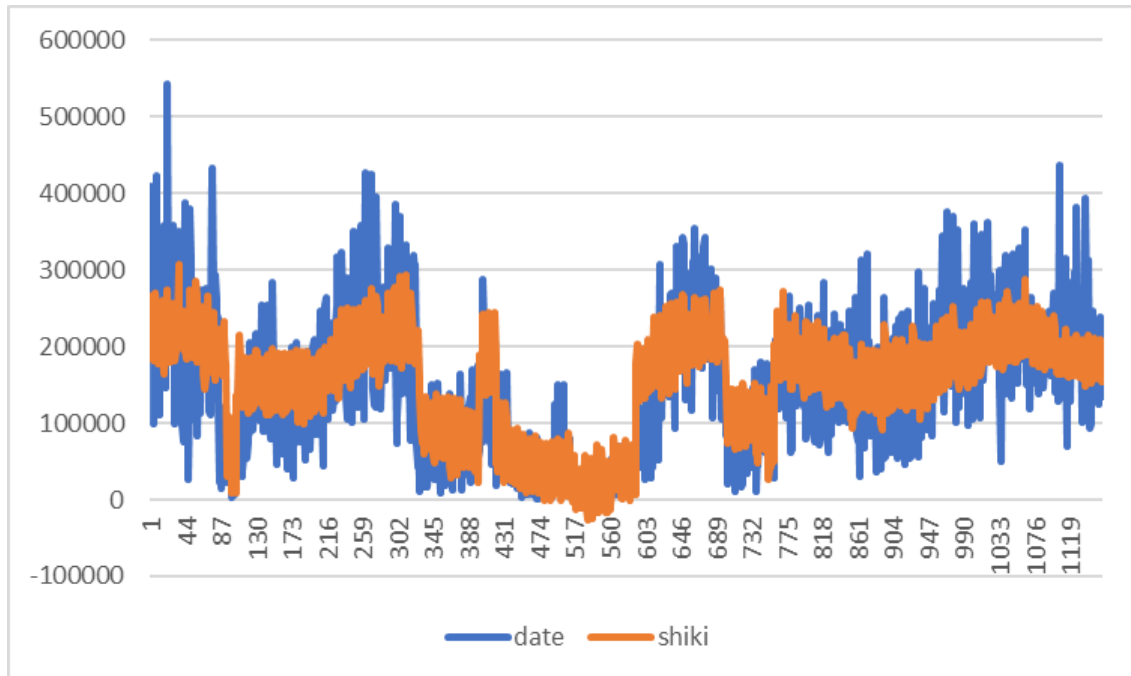


図 1：求めた式と実際のデータの比較

オレンジ色が上記で求めた式に実際のデータを当てはめた値を繋げたものであり、青色が実際のデータを繋げたものである。求めた式が実際のデータと同様の動きをしているのが分かる。

次に被説明変数と説明変数の両方の自然対数をとる。モデルの結果は以下の通りである。

$$\begin{aligned}
 \ln \text{Sale} = & 12.166 - 0.397 \times \ln \text{temp} + 0.070 \times \ln \text{cost} - 0.031 \times \ln \text{rain} - 0.013 \times \text{tue} \\
 & (0.257) \quad (0.030) \quad (0.024) \quad (0.012) \quad (0.062) \\
 & + 0.121 \times \text{wed} + 0.134 \times \text{thu} + 0.543 \times \text{fri} + 0.597 \times \text{sat} + 0.264 \times \text{sun} \\
 & (0.063) \quad (0.064) \quad (0.062) \quad (0.062) \quad (0.065) \\
 & - 1.662 \times \text{sengen} - 1.096 \times \text{manbo} \\
 & (0.047) \quad (0.063)
 \end{aligned}$$

() 内は標準誤差。

決定係数 $R^2 = 0.659$

各係数を見ると、定数項、ltemp、lcost、lrain、金曜日ダミー、土曜日ダミー、日曜日ダミー、緊急事態宣言ダミー、まん延防止ダミーは1%有意であり、木曜日ダミーは5%有意であり、水曜日ダミーは10%有意であり、火曜日ダミーのみ有意ではなかった。

求めた式に各変数の数値を代入して、実際のデータと比較した結果は以下の通りである。

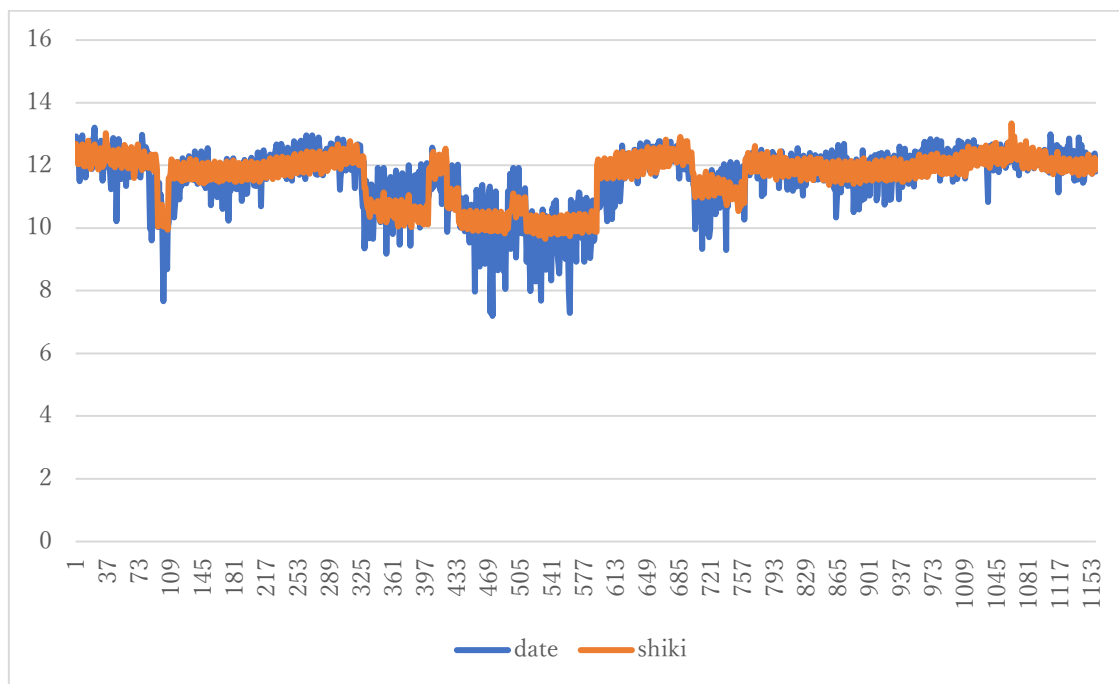


図 2：対数変換後の式と対数変換後の実際のデータの比較

オレンジ色が上記の式に実際のデータを当てはめた値を繋げたものであり、青色が実際のデータを繋げたものである。求めた式が実際のデータと同様の動きをしている。

第 4 章

時系列分析、特に SARIMA モデルを用いた今後の売上予測

4-1 SARIMA モデル概略

今回扱うデータは週次、月次の季節性を伴う経済時系列である。このようなデータに対して有用なのが SARIMA モデルである。SARIMA (Seasonal AutoRegressive Integrated Moving Average) モデルは ARIMA

(AutoRegressive Integrated Moving Average) モデルに季節性要素を加えたモデルである。

- 自己回帰 (Autoregressive) 成分 (AR) : データが直前のデータ点に依存する場合に適しており、過去のデータ点に依存する自己回帰項が含まれることを示す。
- 差分 (Integrated) 成分 (I) : データを非定常な状態から定常な状態に変換するために、差分を取ることがある。差分を取ることにより、季節性を取り除くことができる。
- 移動平均 (Moving Average) 成分 (MA) : 過去の誤差項に依存する移動平均項が含まれることを示す。
- 季節性 (Seasonal) 成分 (S) : データに季節性がある場合に適している。季節性とは、データ内の一定期間ごとのパターンやトレンドを示す。

SARIMA(p, d, q)(P, D, Q, S)モデルは通常の ARIMA(p, d, q)と S 期間ごとに過去のデータを抽出した ARIMA(P, D, Q)を重ね合わせたモデルである。一般的に、オペレータ形式を用いて次のように表される。詳しい導出過程は『現代時系列分析』(田中勝人、岩波書店)を参照されたい。

$$\varphi_p(B)\Phi_P(B)(1-B)^d(1-B^S)^D y_t = \theta_q(B)\Theta_Q(B)\varepsilon_t$$

B を後退オペレータといい、 y_t に対して $B^k y_t = y_{t-k}$ で定める。 ε_t は確率的なバラつきを表すノイズである。ここで、自己回帰オペレータ $\varphi_p(B)$ と移動平均オペレータ $\theta_q(B)$ は以下のように定義される。

$$\begin{aligned}\phi_p(B) &= 1 - \phi_1 B^s - \phi_2 B^{2s} - \dots - \phi_p B^{ps} = \sum_{i=0}^p (-\phi_i) B^{si} & \phi_0 &= -1 \\ \theta_Q(B) &= 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_Q B^{Qs} = \sum_{i=0}^Q (-\theta_i) B^{si} & \theta_0 &= -1\end{aligned}$$

それぞれのパラメータは以下のである。

- p : 自己回帰 (AR) 過程の階数
- d : 定常化に必要な階差 (I) の階数
- q : 移動平均 (MA) 過程の階数
- P : 季節性自己回帰 (SAR) 過程の階数
- D : 季節性差分 (SI) 過程の階数
- Q : 季節性移動平均 (SMA) 過程の階数
- S : 季節性の周期単位

4-2 分析結果

今回のモデルでは、コロナウイルスの影響をできる限り減らすために 2022 年 5 月～2023 年 4 月までの 1 年間の日次データのみを使用した。多重共線性の問題を避けるために月曜日ダミーを外してモデルを作成した。

求められたモデルは SARIMA(0,1,1)(1,0,0,7)モデルである。

SARIMA(0,1,1)(1,0,0,7)モデルの各係数は以下の通りである。

=====

SARIMA(0,1,1)(1,0,0,7)モデル

ma1	- 0.929 *** (0.021)
sar1	0.060 (0.054)
tue	8394.748 (10665.542)
Wed	9328.226 (10677.201)
thu	- 9274.079

	(10680.839)
fri	94503.798 ***
	(10685.618)
sat	23531.463 **
	(10627.584)
sun	-36821.019 ***
	(10620.087)

AIC	8997.186
AICc	8997.693
BIC	9032.285
Log Likelihood	- 4489.593
Num. obs.	365

*** p < 0.01; ** p < 0.05; * p < 0.1

() 内は標準誤差。

ma1、金曜日ダミー、日曜日ダミーが1%有意で、土曜日ダミーが5%有意である。その他の説明変数は有意ではない。

ma1の係数は-0.929である。これは1期前の誤差項に対する移動平均成分の係数が-0.929であることを示している。係数の符号がマイナスなので、1期前の誤差が増加すると現在の売上が下がる。

sar1の係数は0.060である。これは季節性の自己回帰成分の影響を示している。今回は係数が有意ではないので7期前、つまり1週間前の誤差が現在の売上に影響を与えない、もしくは与える影響が小さいといえる。

第 5 章

2 つの分析からの考察

5-1 考察：時系列分析

重回帰分析の各変数の係数について妥当かどうかを検討する。

平均気温に関して、1 日の平均気温が 1 単位下がると売上が 3409.646 円上がる。これは妥当だと考える。某飲食店は鍋や焼き鳥を提供しているため、気温が低い時に来客数が増えて売上が上がると考えられるからである。

アルバイトの人件費に関して、アルバイトの人件費が 1 単位増えると売上が 1.413 円上がる。これは、妥当とは言い切れない。売上が高くなると予想される日にはバイトのシフトを増やすという逆因果の可能性があるからである。

降水量に関して、降水量が 1 単位増えると売上が 268.110 円下がる。これは妥当だと考えられる。降水量が増えることで外食する意欲が減少して売上が下がると考えられるからである。

各曜日のダミー変数に関して、金曜日は 75830.373 円、土曜日は 60358.290 円と売上が大きく上がっていることが分かる。他の曜日に関しては、月曜日より売上に対してプラスの効果を与えていることが分かる。これは妥当だと考えられる。社会人にとって金曜は休みの前日、土曜日もある日があるため来客数が多くなると考えられる。また、月曜日から働き始めるので、月曜日が最も来客数が少なく売上も低くなると考えられる。金曜日に近づくにつれて売上に与える影響が大きくなっているのは、休日が近くなることで外食への意欲が高まるからだと考えられる。

緊急事態宣言とまん延防止のダミー変数に関しては、緊急事態宣言の時は 119064.585 円、まん延防止の時は 113908.443 円売上が下がることが分かる。新型コロナウイルスの流行によって外食の機会が減ったことで、売上にマイナスの影響を与えていたと考えられる。

1 つ目のモデルは決定係数が 0.600 なので、実際のデータの動きの 60%を今回求めた式で説明できていることになる。対数変換後のモデルの決定係数は 0.659 なので、対数変換した実際のデータの 65.9%を説明できていることになる。

対数変換後の式について考える。説明変数の係数を比較して、売上に与える影響の大きさ確かめる。係数が最も大きいのは緊急事態宣言ダミーであり、次に大きいのはまん延防止ダミーである。この結果からもコロナ禍が飲食店

の売上に与えた影響の大きさが分かる。また、曜日ダミーで係数が大きかったのは金曜日と土曜日である。

5-2 考察：SARIMA モデル

ma1 項の p 値が有意であるので 1 期前の移動平均は売上に関係している。一方係数が負であることから過去の誤差が未来の値に関して影響を与えている。

sar1 の係数から季節性はあるが、有意ではないといえる。売り上げの要因としてあまり重要ではない。

各曜日ダミーを比べると、金曜日土曜日は係数が有意でかつ値が大きい。先ほどの重回帰分析と同じような解釈ができるので妥当な結果である。

第 6 章

今後の展望

今回、重回帰分析と時系列分析を用いて分析を行ったが、結果として予想通りにはいかない部分も多々あった。

重回帰分析に関しては、説明変数として抽出できる要素が少なかったために、モデルの多くがダミー変数を用いた少し不自然な分析となってしまう。その結果、売上の要因として曜日効果や降水量などといった、お店側が調整することの出来ない要素が大きく影響するといった、お店側へのフィードバックが少ない分析であった。

時系列分析に関しても、1 週間の中で売上に周期性が存在すると予測して、季節性を含む時系列の分析が可能な SARIMA モデルを用いた分析を試みたが、今回用いたデータでは、統計的に有意な周期性は存在せず、また 1 期前の売上が直接現在の売上には影響しないという結果となり、過去の売上から今後の売上の予測が正確には出来ず、終わってしまった。

今回、用いたデータがコロナの影響を受けている時期のデータが多かったこと、1 店舗だけのデータであったこと、約 3 年間のみのデータであったこと、など予想通りの分析が出来なかった原因は多く挙げられる。今回の失敗を踏まえた改善点として、店舗の周りにどんな人がいるか、どんなものがあるかなど現地に足を運んだ調査を行うことでより正確な分析が行えるかもしれないという点や、今回用いた統計的な分析手法を正確に適用できるようにするために、より長期の、より多くのデータの収集が必要だったという点が考えられる。

<謝辞>

今回の論文の執筆に際して、長倉大輔教授および長倉大輔研究会の皆様から多大なるご指導いただいたこと、心からお礼申し上げます。また、貴重なデータを提供してくださった某飲食店の関係者様に深く感謝致します。

<参考文献>

星田佳祐「日本政策金融公庫 調査月報」p4~11、2022 年

山口類・土屋映子・樋口知之「状態空間モデルを用いた飲食店売上の要因分解」オペレーション・リサーチ

W.ヴァンデール「ボックスジェンキンスモデルの応用」箕谷千鳳彦・廣松毅訳『時系列入門』多賀出版、1988 年

田中勝人『現代時系列分析』岩波書店、2006 年

https://qiita.com/suzuki_/items/14998f8120b33a2e5668

kanbe(2020.4.14)「ARIMA モデル（自己回帰和文移動平均モデル）について分かりやすく解説」https://ai-trend.jp/basic-study/time-series-analysis/arima_model/

suzuki(2021.5.2)「SARIMA モデルを導出したい」[SARIMA モデルを導出したい #時系列解析 - Qiita](#)