



ウォルマートの売上予測

2015年1月

慶應義塾大学濱岡豊研究会

前川千広 和田龍馬 渋谷悟

アジェンダ

1. 課題・データの概要
2. 事例研究「アメリカにおける年末商戦」
3. 分析・考察①「既存の変数、イベントダミーによる分析」
4. 分析・考察②「売上の対数変換、月ダミー追加による分析」
5. 分析・考察③「店舗ダミーの追加による分析」
6. 分析・考察④「年ダミー追加による分析」
7. まとめ・課題

要約

Kaggle.comで過去に開催されたウォルマートの売上等のデータを用いて重回帰分析により売上予測を行った。結果、値下げのデータや地域のデータが売上予測に比較的大きく影響することがわかった。また、月や年、年末商戦等のダミー変数もまた売上予測に有効であることが明らかになった。

ウォルマートとは



サム・ウォルトンがアメリカで1962年に創業した小売店 EDLP (Every Day Low Price) を掲げ急成長を遂げ現在では世界最大の小売企業となっている。

2014年3月末現在
全米で4853店舗
全世界で10984店舗
を展開している



※ ウォルマートホームページより

課題の概要・データの概要

課題: ウォルマートの売上予測 kaggle.com (<http://kaggle.com/>) より

与えられているデータ

ウォルマート45店舗の過去の売上データ 2010-02-05～2012-11-01

売場毎の売上、毎週の売上のデータを含む

値下げの実行

失業率

祝日等の情報

気温

燃料価格

消費者物価指数

※値下げについてのデータは2011年11月以降のみ

※祝日等の情報(バレンタイン、勤労感謝の日、ブラックフライデー、年末)

提言について目標

- 今回与えられた変数がウォルマートの売上にどのように影響するかを測定しその結果に合わせた売上予測を考える。
- 特に値下げ、気温、祝日やイベントが売上にどう影響するか注目して分析を行う。

※オリジナルのデータは

45店舗×98の売り場の売上×143週間で構成されているが、今回は各店舗の98の売り場の売上を合計して、45店舗の総売上×143週間のデータにして店舗ごとの売上予測を行う。

アジェンダ

1. 課題・データの概要
2. 事例研究「アメリカにおける年末商戦」
3. 分析・考察①「既存の変数、イベントダミーによる分析」
4. 分析・考察②「売上の対数変換、月ダミー追加による分析」
5. 分析・考察③「店舗ダミーの追加による分析」
6. 分析・考察④「年ダミー追加による分析」
7. まとめ・課題

アメリカにおける年末商戦

- 年末商戦・・・クリスマス頃(12月後半)商品が売れるようになる時期をあらわすもの。プレゼント、お歳暮、新年を迎えるためやボーナスなどで一時的に売り上げがあがること。

2012年年末商戦では買い物客1人当たりの支出額が423ドル(約3万5000円)、支出総額は591億ドル(約4兆8600億円)
(<http://www.cnn.co.jp/business/35024836.html>)



アメリカにおける年末商戦

青木 武 (信金中央金庫Senior Analyst)

『アメリカでは、個人消費が GDP の3分の2、小売業の取引高は GDP の 30%を占めるほど買い物好きな国民性ですが、特に 11 月下旬のサンクスギビングデーからクリスマスまでの1ヶ月間が**ホリデーシーズン**と呼ばれ、家族や友人などへのクリスマスプレゼントを購入するため、最も買い物が活発となる時期です。』

→年末商戦が売上に与える影響は大きいのでは

アジェンダ

1. 課題・データの概要
2. 事例研究「アメリカにおける年末商戦」
3. 分析・考察①「既存の変数、イベントダミーによる分析」
4. 分析・考察②「売上の対数変換、月ダミー追加による分析」
5. 分析・考察③「店舗ダミーの追加による分析」
6. 分析・考察④「年ダミー追加による分析」
7. まとめ・課題

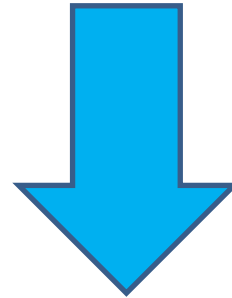
分析①

- 売上を被説明変数、その他の要素を説明変数にして回帰分析を行う。
- 売上については45店舗の売り場ごとの売上を週ごと・店舗ごとに合計する。
- 祝日、イベント等の情報が少ないので祝日・イベントに関する変数を追加する。
- 値下げについては欠損値があるので事前に補完を行う。

祝日・イベントに関する変数の追加

祝日等の情報

(バレンタイン、勤労感謝の日、ブラックフライデー、年末)



バレンタイン

イースター

母の日

父の日

勤労感謝の日

ハロウィン

ブラックフライデー

クリスマス

年末

祝日・イベントを9つの要素に分ける

欠損値の補完

欠損値補完の方法: 多重代入法 (Multiple Imputation : MI)

欠損値に異なる値を補完した複数のデータセットを作成して、そのすべてを用いて分析を行い、その結果を統合する。Rのmiパッケージを使用する。

値下げの実行

MarkDown1,2,3,4,5 についての欠損値を補完する。

※値は恐らく値下げした金額の合計値、MarkDown1,2,3,4,5と値下げについて種類が分かれているが、それぞれが何の値下げなのかは今回不明である。)

データの予測には店舗番号、売上、気温、9つの祝日・イベントの情報を用いた。

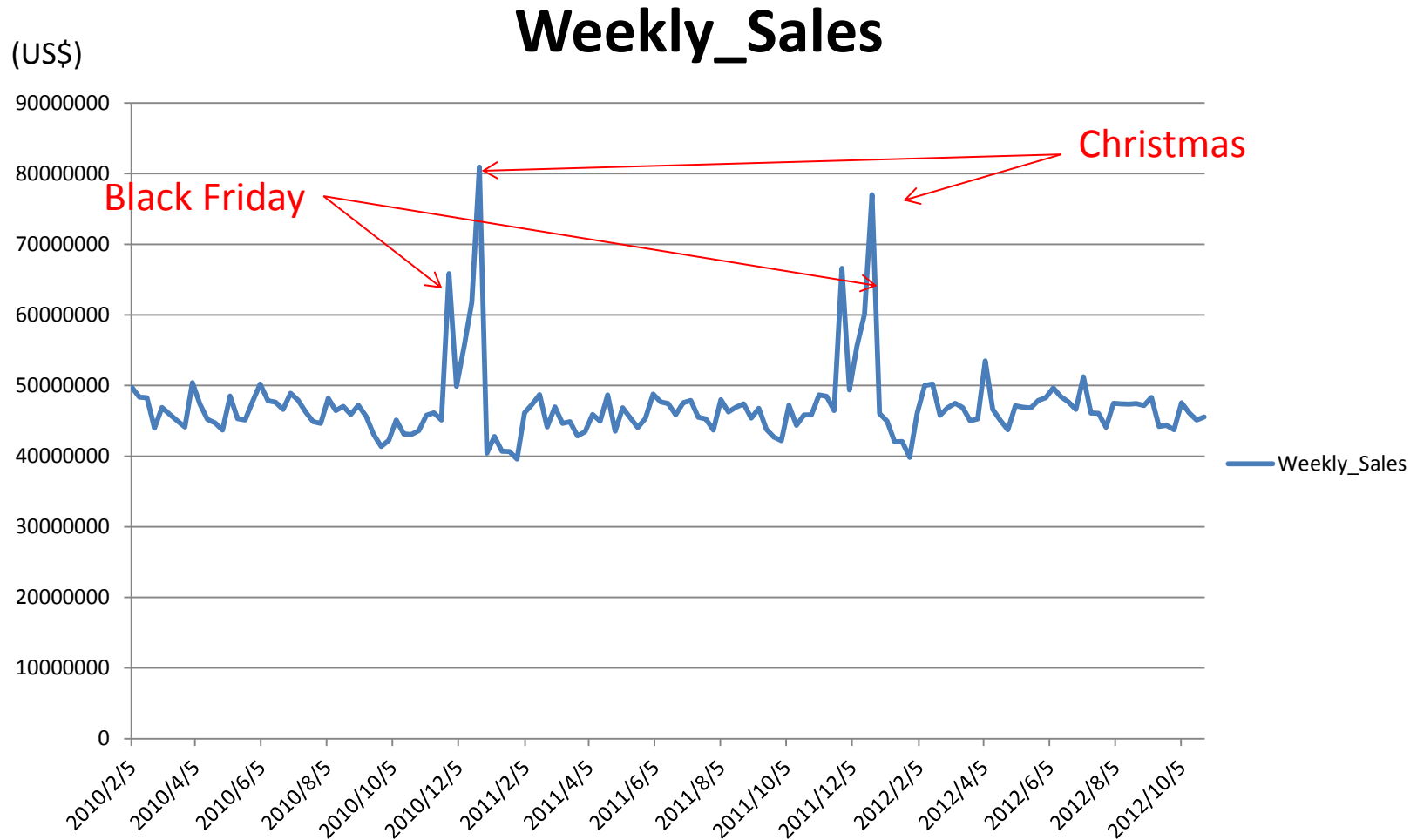
※Rにおける多重代入法については

About connecting the dots.

kaggleで予測モデルを構築してみた (5) - Rで行うMultipleImputation

[http://smrmkt.hatenablog.jp/entry/2013/01/06/164758\(2014-12-01\)](http://smrmkt.hatenablog.jp/entry/2013/01/06/164758(2014-12-01)) を参照した。

分析結果: 売上グラフ



予想通り売上の大半が年末に集中していることがわかる。

分析結果:重回帰分析

被説明変数:売上

	回帰係数	標準誤差	t値	p値
(Intercept)	1.26E+06	4.57E+04	27.486	<2.00E-16***
Temperature	2.84E+03	3.59E+02	7.908	3.05E-15***
CPI	-1.77E+03	1.62E+02	-10.919	<2.00E-16***
Unemployment	-4.76E+04	3.35E+03	-14.203	<2.00E-16***
MarkDown1	1.28E+01	6.49E-01	19.664	<2.00E-16***
MarkDown2	1.23E+01	7.35E-01	16.71	<2.00E-16***
MarkDown3	1.58E+01	1.36E+00	11.608	<2.00E-16***
MarkDown4	-1.15E+00	3.13E-01	-3.686	0.00023***
MarkDown5	3.45E+01	1.18E+00	29.168	<2.00E-16***
Black_Friday	-2.70E+05	9.54E+04	-2.831	0.00466**
Christmas	9.27E+05	5.11E+04	18.153	<2.00E-16***
Easter	6.51E+04	4.12E+04	1.578	0.11459

イースター以外は有意な結果が得られた。値下げについても全ての変数において有意な結果となったが、多重共線性が発生した可能性がある。(スライド14で説明)

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Multiple R-squared: 0.2993, Adjusted R-squared: 0.2981 N=6435

考察①

- 売上のグラフからもわかるように売上の大半を年末商戦で売り上げる。結果各イベントダミーの内最終的に有意となったのが年末商戦関連のものに絞られてしまったと考えられる。
- 年末商戦以外の季節の売上に影響を与える要素が少ないためモデルの当てはまりがあまり高くないと考えられる。
- 値下げの欠損値を他の変数で推定したため重回帰分析では多重共線性が発生してしまった可能性が考えられる。(単回帰分析と重回帰分析で説明変数の符号が異なっていたため)

アジェンダ

1. 課題・データの概要
2. 事例研究「アメリカにおける年末商戦」
3. 分析・考察①「既存の変数、イベントダミーによる分析」
4. 分析・考察②「売上の対数変換、月ダミー追加による分析」
5. 分析・考察③「店舗ダミーの追加による分析」
6. 分析・考察④「年ダミー追加による分析」
7. まとめ・課題

分析②

- 分析1の反省を踏まえて欠損値の推定に用いるデータを変更する。(重回帰分析に用いない変数等を持用いて推定する。)
- 値下げ2, 3, 4 (MarkDown1,2,3は欠損値が多すぎるため今回は使用しない)
- 売上のグラフが極端な形であるため、logを取って分析を行う。
- 変数が不足していると分析1では考えられたため、1月を基準としたダミー変数、独立記念日のイベントダミーを追加する。

分析結果:重回帰分析

被説明変数: log(売上)

	係数	標準誤差	t値	p値
(Intercept)	1.39E+01	5.44E-02	256.196	<2.00E-16***
Temperature	-4.61E-03	6.50E-04	-7.086	1.53E-12***
CPI	-9.58E-04	1.87E-04	-5.116	3.21E-07***
Unemployment	-2.27E-02	3.85E-03	-5.882	4.26E-09***
MarkDown1	4.44E-06	3.33E-07	13.314	<2.00E-16***
MarkDown5	4.14E-05	1.37E-06	30.156	<2.00E-16***
Independence_Day	5.40E-03	5.73E-02	0.094	0.924
Black_Friday	3.89E-01	6.16E-02	6.319	2.81E-10***
Christmas	2.21E-01	6.37E-02	3.466	0.000***
February	1.86E-02	3.44E-02	0.54	0.589
March	1.35E-01	3.41E-02	3.951	7.87E-05***
April	1.64E-01	3.67E-02	4.461	8.29E-06***
May	2.26E-01	3.79E-02	5.962	2.62E-09***
June	2.48E-01	4.18E-02	5.942	2.97E-09***
July	2.43E-01	4.35E-02	5.591	2.35E-08***
August	2.61E-01	4.24E-02	6.164	7.53E-10***
September	1.63E-01	4.04E-02	4.028	5.70E-05***
October	1.76E-01	3.73E-02	4.712	2.50E-06***
November	6.76E-02	3.85E-02	1.755	0.079
December	8.73E-02	6.08E-02	1.436	0.151

独立記念日、
2月、11月、
12月以外の変
数において有
意な結果となっ
た。

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

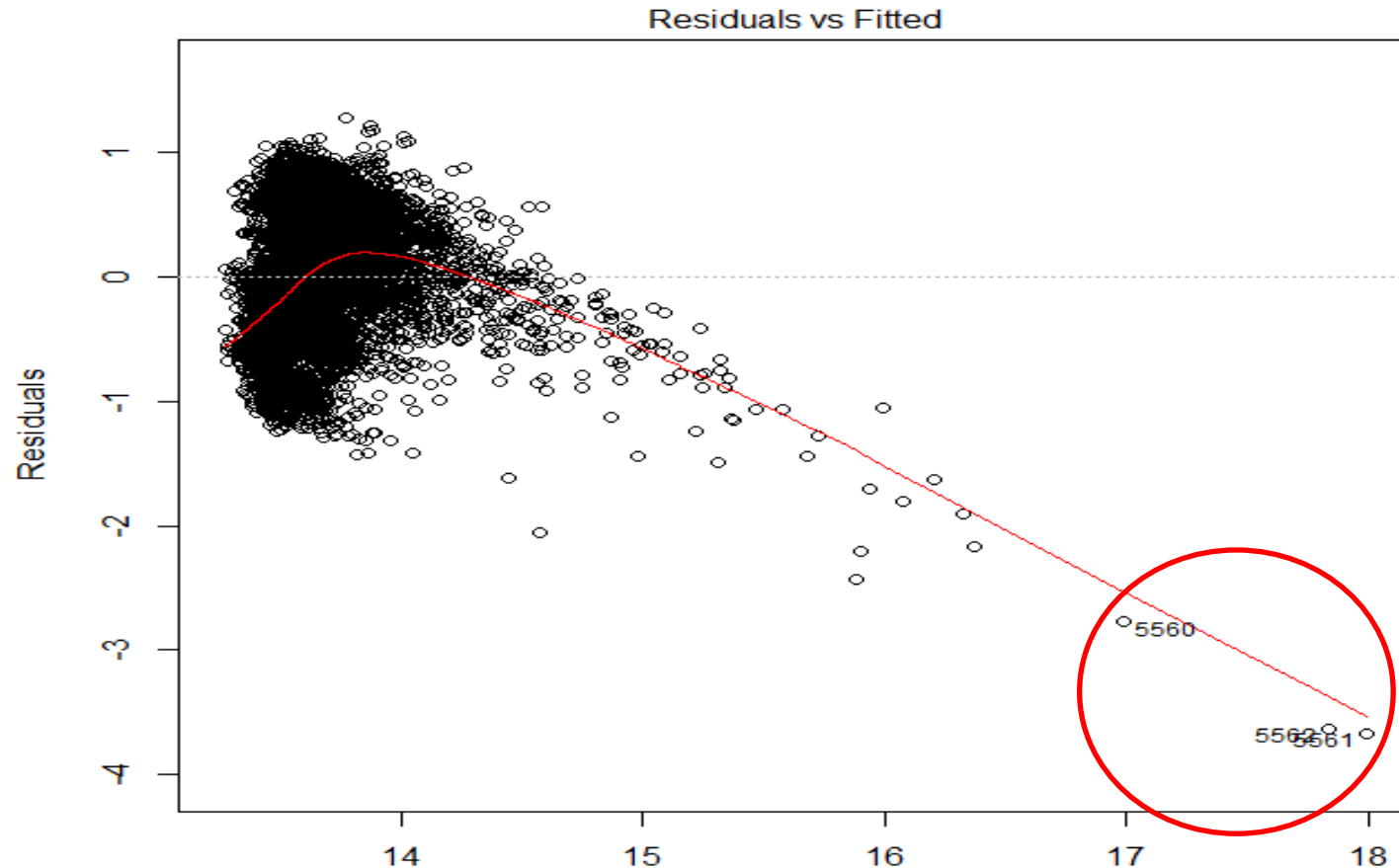
Adjusted R-squared: 0.2138

N=6435

考察②

- 売上のグラフからもわかるように売上の大半を年末商戦で売り上げる。結果各イベントダミーの内最終的に有意となったのが年末商戦関連のものに絞られてしまったと考えられる。
- 年末商戦以外の季節の売上に影響を与える要素が少ないためモデルの当てはまりがあまり高くないと考えられる。

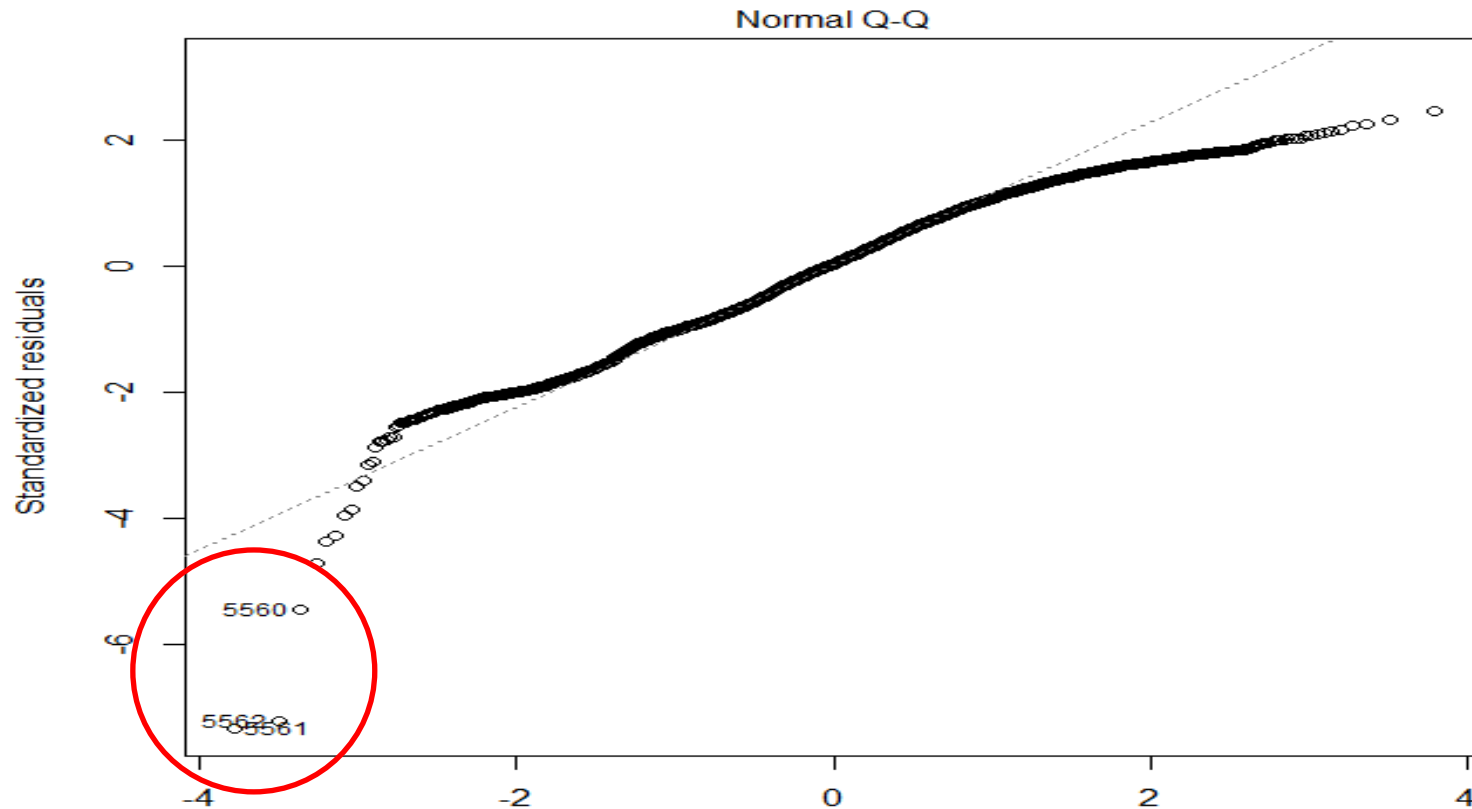
分析結果：回帰診断図



残差とフィット値のプロット

残差とフィット値の散布図で、図から残差の全体像を見る事が出来る

分析結果：回帰診断図



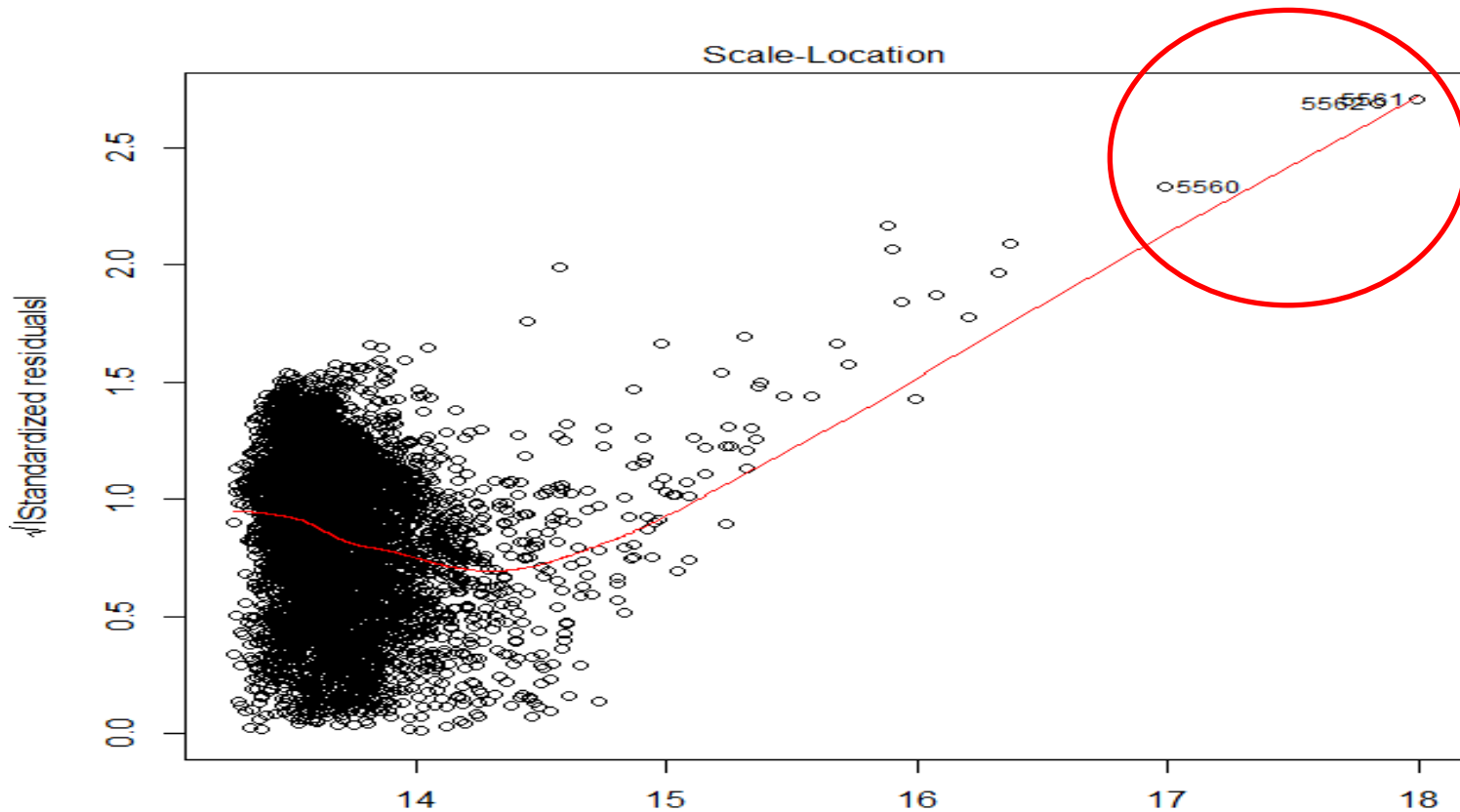
正規Q-Qプロット

データの正規性を考察するための図

データが正規分布に従う場合、点が直線上で並ぶ

回帰分析では、残差が標準正規分布に従うと仮定しているため、この図が直線に近いほど良いことになる

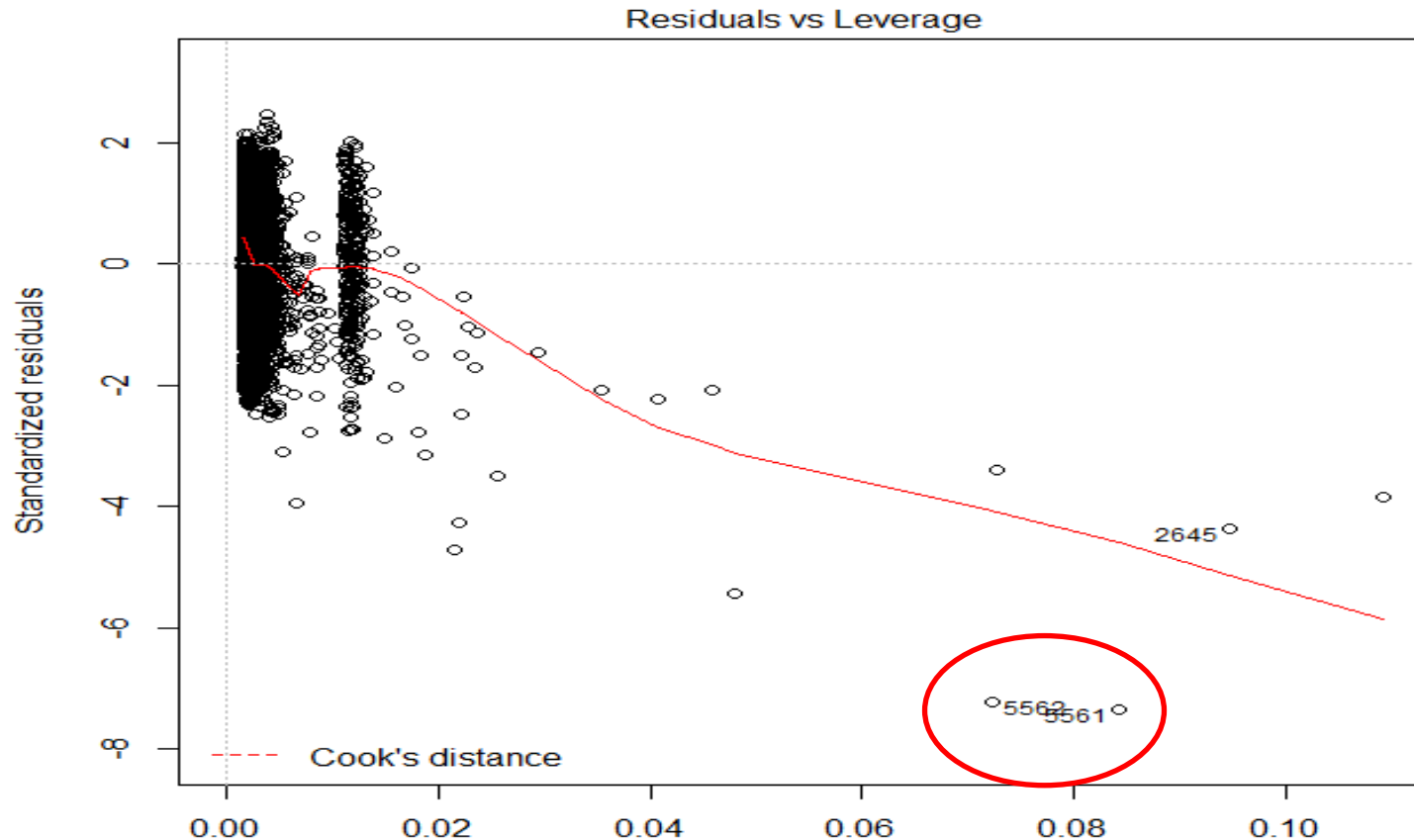
分析結果: 回帰診断図



残差の平方根プロット

標準化した残差の絶対値の平方根を縦軸にし、予測値を横軸とした散布図
この図も残差の変動状況を考察することが目的

分析結果：回帰診断図



残差と影響カプロット

横軸が、擬子値、縦軸が標準化した残差で、点線でクックの距離0.5を示している
擬子値が大きいほど当てはまりが良く、クックの距離が0.5以上で影響力が大きい、1以上だと特異に大きいと言われる

考察② 回帰診断図

- Normal Q-Q を見ると両端が直線から外れており、残差が正規分布に従うという仮定が成り立ってるとは言いにくいことがわかる。
- サンプル5560、5561、5562が外れ値となっていることがわかる。(店舗39の2012/6/9～2012/7/6 までの売上)

アジェンダ

1. 課題・データの概要
2. 事例研究「アメリカにおける年末商戦」
3. 分析・考察①「既存の変数、イベントダミーによる分析」
4. 分析・考察②「売上の対数変換、月ダミー追加による分析」
5. 分析・考察③「店舗ダミーの追加による分析」
6. 分析・考察④「年ダミー追加による分析」
7. まとめ・課題

分析③

- 分析2の回帰診断図の結果を踏まえてサンプル5560, 5561, 5562(店舗39の2012/6/9~2012/7/6までの売上)の特徴を考えたが不明であったためこれらをサンプルから取り除く。
- 新たに店舗をグループ分け(地域ごとに分ける)して、ダミー変数として追加する。

変数の追加

店舗をグループ分けする(地域分けを行う)

ガソリン価格を基準に9つの地域に分けられる。

ダミー変数の基準となる店舗 店舗36

Area_A 1,2,3,5,6,8,9,11,21,30,31,37,39,43

Area_B 7,16,32,41

Area_C 4,34

Area_D 13,17,44

Area_E 14,20,25,35,45

Area_F 18,22,23,26,29,40

Area_G 15,19,24,27

Area_H 10,12,28,33,38,42

分析結果: 重回帰分析

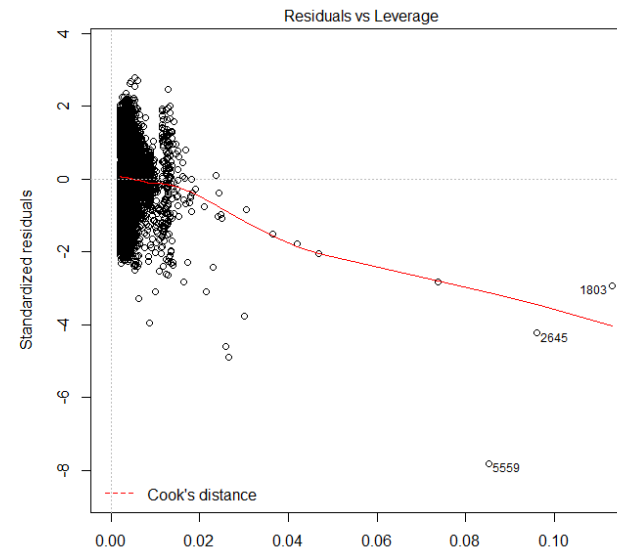
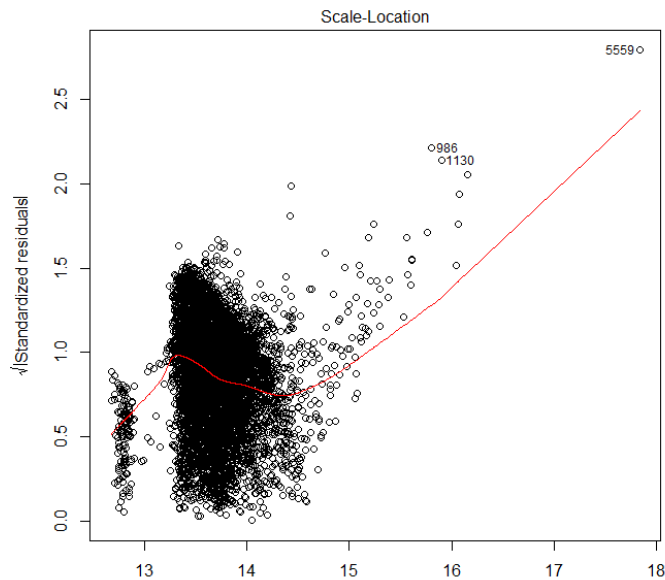
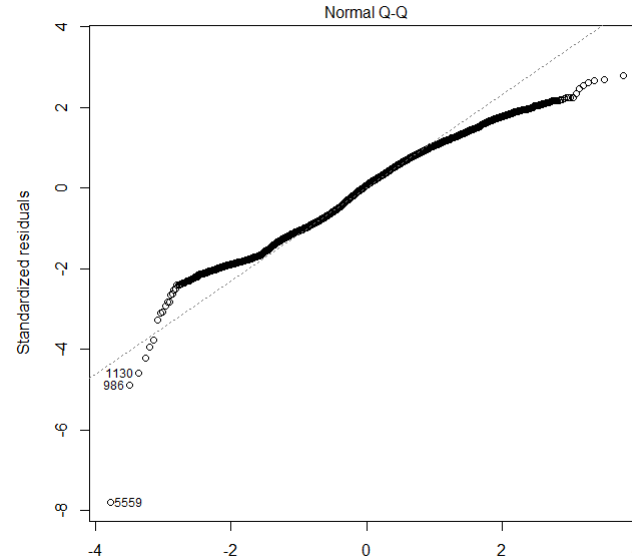
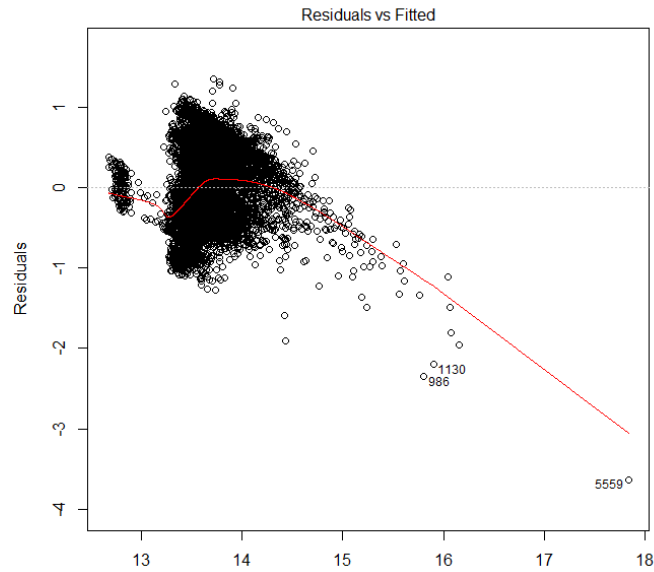
被説明変数: log(売上) N=6432

	係数	標準誤差	t値	p値
(Intercept)	1.18E+01	1.71E-01	68.983	<2.00E-16***
Temperature	2.16E-03	1.01E-03	2.138	0.032*
CPI	3.74E-03	6.75E-04	5.534	3.26E-08***
Unemployment	-7.76E-03	4.24E-03	-1.829	0.06749.
MarkDown1	3.65E-06	3.14E-07	11.611	<2.00E-16***
MarkDown5	4.19E-05	1.39E-06	30.14	<2.00E-16***
Black_Friday	3.92E-01	5.75E-02	6.82	9.93E-12***
Christmas	2.05E-01	5.95E-02	3.441	0.000***
February	1.96E-02	3.21E-02	0.611	0.541
March	5.29E-02	3.33E-02	1.59	0.111
April	2.53E-02	3.80E-02	0.665	0.506
May	3.49E-02	4.20E-02	0.83	0.406
June	1.87E-03	4.93E-02	0.038	0.969
July	-3.34E-02	5.25E-02	-0.636	0.524
August	-1.42E-02	5.14E-02	-0.277	0.781
September	-6.35E-02	4.61E-02	-1.375	0.169
October	2.14E-02	3.92E-02	0.544	0.586
November	-4.03E-03	3.69E-02	-0.109	0.913
December	9.17E-02	5.68E-02	1.615	0.106
Area_A	6.79E-01	4.24E-02	16.007	<2.00E-16***
Area_B	8.01E-01	5.40E-02	14.831	<2.00E-16***
Area_C	1.47E+00	7.76E-02	18.936	<2.00E-16***
Area_D	1.02E+00	7.87E-02	12.904	<2.00E-16***
Area_E	1.11E+00	5.12E-02	21.675	<2.00E-16***
Area_F	1.16E+00	7.29E-02	15.945	<2.00E-16***
Area_G	1.32E+00	7.23E-02	18.213	<2.00E-16***
Area_H	8.94E-01	7.15E-02	12.504	<2.00E-16***

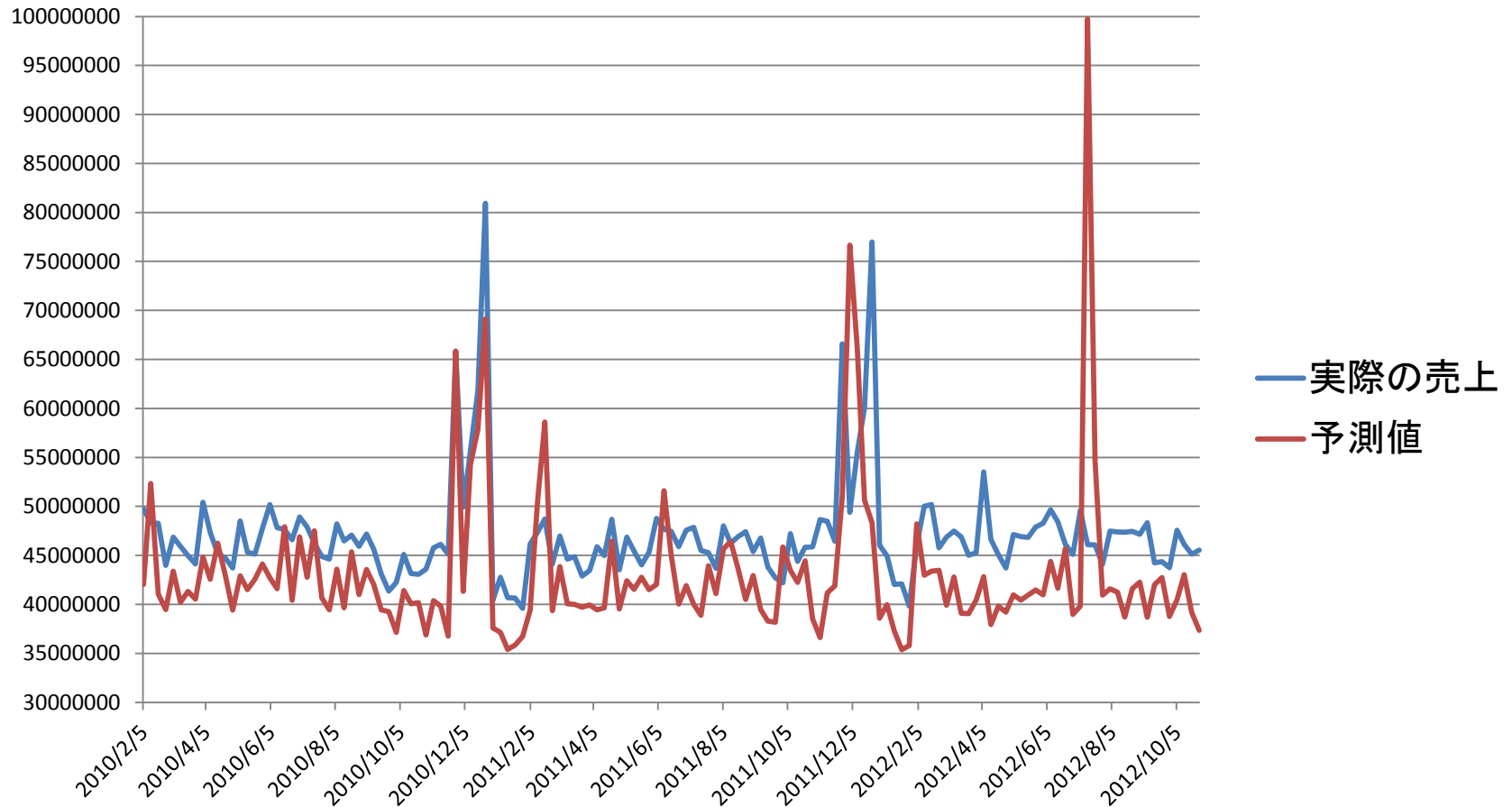
1月を基準にした月ダミーは有意な結果とならなかった。

一方地域ダミーやその他の変数は有意な結果となった。

分析結果：回帰診断図



実際の売上と予測値の比較



※予測には分析3の結果を用いた。

※店舗1-45の売上を週ごとにそれぞれ合計した上で比較した。

2012年8月の売上の予測値が異常な値となってしまうことがわかった。

考察③

- 店舗ダミーが有意な結果となった一方で、月ダミーが有意な結果とならなかった。
- 店舗ダミーは当てはまりを高くすることがわかった。
- 回帰診断図を見ると先ほど取り除いた店舗39の他の週の値が外れ値となっていた。
- 実際の売上と予測値の比較から、2012年8月の予測値に異常が見られた。
- 2011年の年末商戦の実際の売上とその予測値が1週間分ほどずれていることが分かった。
- 2010年に比べて2011年、2012年の当てはまりが悪いことが予測値との比較で明らかとなった。

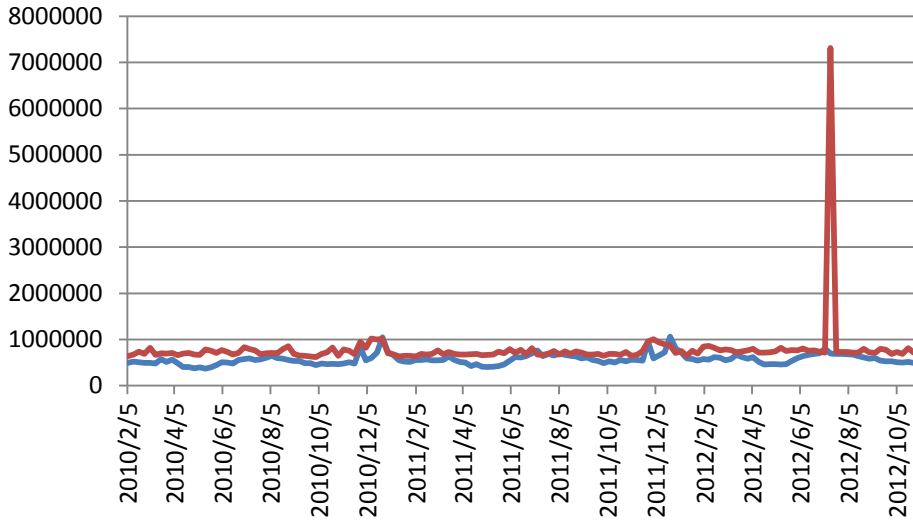
アジェンダ

1. 課題・データの概要
2. 事例研究「アメリカにおける年末商戦」
3. 分析・考察①「既存の変数、イベントダミーによる分析」
4. 分析・考察②「売上の対数変換、月ダミー追加による分析」
5. 分析・考察③「店舗ダミーの追加による分析」
6. 分析・考察④「年ダミー追加による分析」
7. まとめ・課題

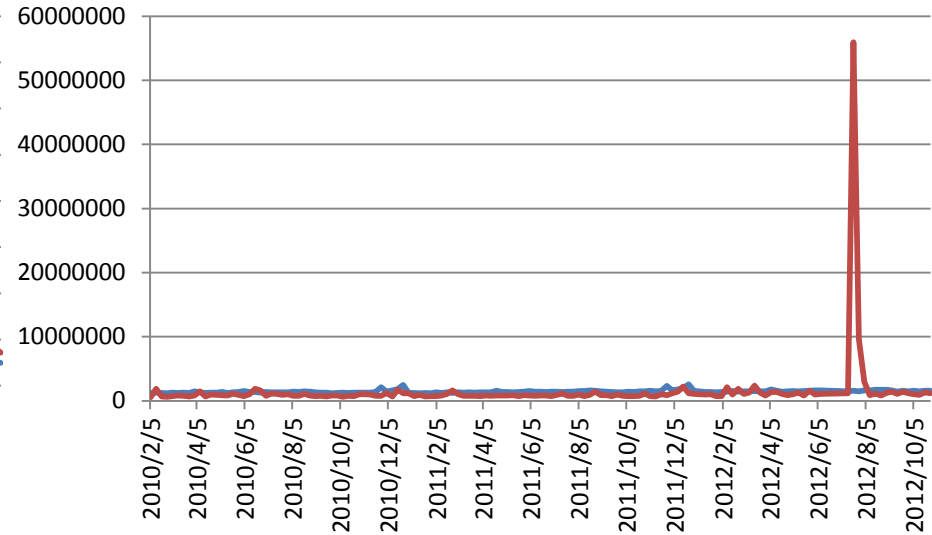
分析④

- 2012年8月の予測値に異常が見られたため各店舗ごとの売上の予測値をグラフにして異常を見つける。
- 2011年の年末商戦は予測値の時期がずれていたなのでダミー変数を調整する。
- 2010年に比べて2011年と2012年の予測値の実際の売上との差が大きいのので2010年を基準とした年ダミーを変数に追加する。

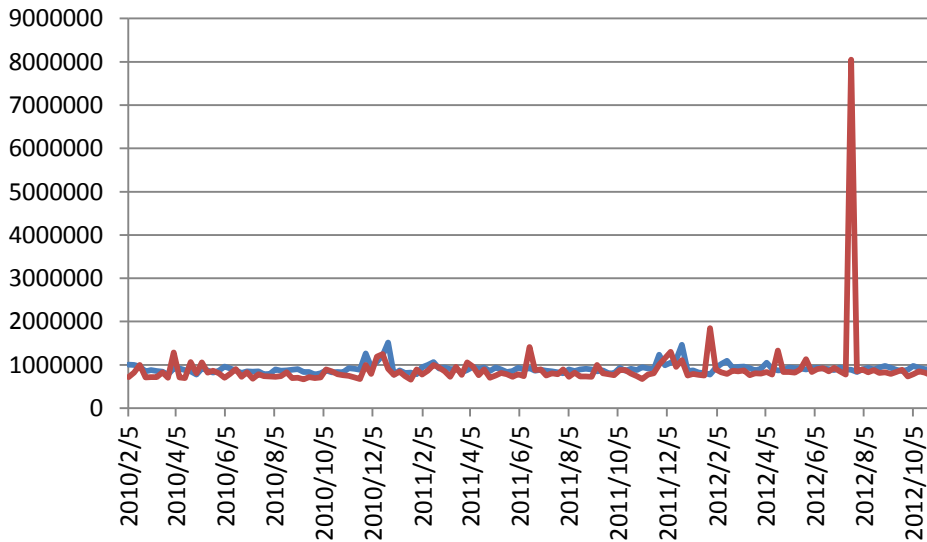
店舗7の売上予測と実際の売上



店舗39の売上予測と実際の売上



店舗8の売上予測と実際の売上



— 実際の売上
— 売上予測

分析3を用いて各店舗の実際の売上と売上予測を比較したところ店舗7, 8, 39の2012年7、8月に異常が見られた。分析4ではこれらの店舗を削除する。

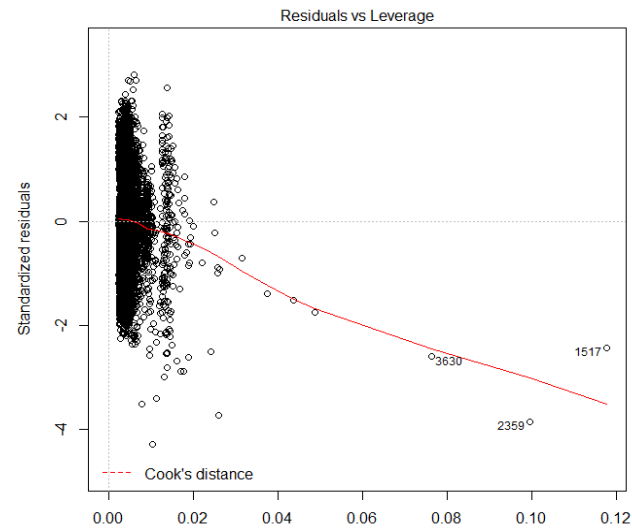
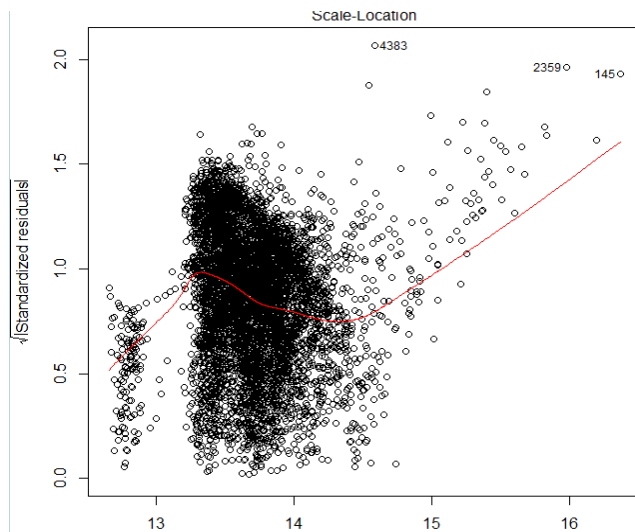
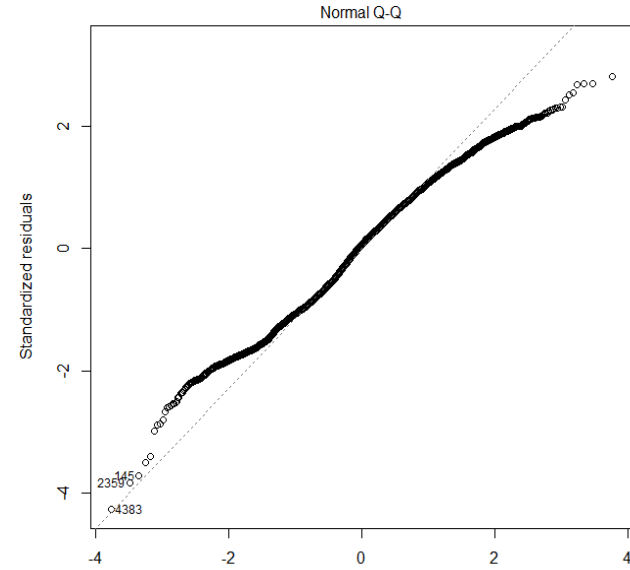
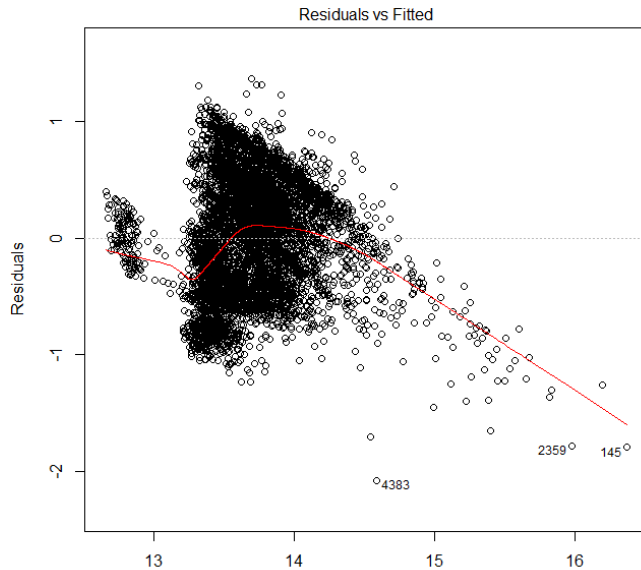
分析結果:重回帰分析

被説明変数: log(売上) N=6004

	係数	標準誤差	t値	p値
(Intercept)	1.19E+01	1.76E-01	67.722	<2.00E-16***
Temperature	-2.15E-04	1.10E-03	-0.196	0.8449
MarkDown1	3.16E-06	3.24E-07	9.769	<2.00E-16***
MarkDown5	5.00E-05	1.60E-06	31.207	<2.00E-16***
CPI	3.65E-03	7.05E-04	5.18	2.29E-07***
Unemployment	-2.71E-03	4.48E-03	-0.606	0.5448
Black_Friday	4.01E-01	5.97E-02	6.72	1.99E-11***
Christmas	1.93E-01	6.18E-02	3.123	0.0018**
Area_A	6.27E-01	4.27E-02	14.663	<2.00E-16***
Area_B	8.24E-01	5.59E-02	14.738	<2.00E-16***
Area_C	1.40E+00	8.04E-02	17.467	<2.00E-16***
Area_D	9.50E-01	8.16E-02	11.643	<2.00E-16***
Area_E	1.05E+00	5.25E-02	19.939	<2.00E-16***
Area_F	1.09E+00	7.59E-02	14.392	<2.00E-16***
Area_G	1.24E+00	7.53E-02	16.523	<2.00E-16***
Area_H	8.61E-01	7.46E-02	11.549	<2.00E-16***
February	1.86E-02	3.37E-02	0.55	0.5821
March	8.80E-02	3.55E-02	2.479	0.0132*
April	7.97E-02	4.08E-02	1.954	0.0507.
May	1.15E-01	4.55E-02	2.526	0.0115*
June	9.67E-02	5.37E-02	1.802	0.0716.
July	8.79E-02	5.73E-02	1.534	0.125
August	7.98E-02	5.63E-02	1.419	0.1559
September	1.37E-02	5.02E-02	0.273	0.7852
October	8.39E-02	4.24E-02	1.979	0.0478*
November	1.77E-02	4.01E-02	0.441	0.6592
December	1.09E-01	5.97E-02	1.832	0.067.
Year2011	-3.08E-03	1.56E-02	-0.197	0.844
Year2012	4.09E-02	1.83E-02	2.233	0.0256*

分析①-③と比較して気温や失業率が有意な結果とならなかった。地域や値下げ年末商戦関連の変数は有意となった。月に関しては有意となった月とそうでない月に分かれた。

分析結果：回帰診断図



考察④

- 年ダミーが2012年に関しては有意な結果となり、月ダミーも全てではないが有意となる月があった。異常な値を示した店舗を削除したことが影響しているのであろう。
- 決定係数は一番高い値を示したが、回帰診断図の結果より依然としてモデルの当てはまりは良くないことがわかる。(正規QQプロットが直線になっていないことなどから)

アジェンダ

1. 課題・データの概要
2. 事例研究「アメリカにおける年末商戦」
3. 分析・考察①「既存の変数、イベントダミーによる分析」
4. 分析・考察②「売上の対数変換、月ダミー追加による分析」
5. 分析・考察③「店舗ダミーの追加による分析」
6. 分析・考察④「年ダミー追加による分析」
7. まとめ・課題

まとめ・課題

• 売上の予測に有効だった変数

値下げの実行

失業率

消費者物価指数

気温

ブラック・フライデー

クリスマス

地域

年・月

特に値下げや地域が
どのモデルにおいても
大きく影響した。

• 予測に有効ではなかった変数

燃料価格

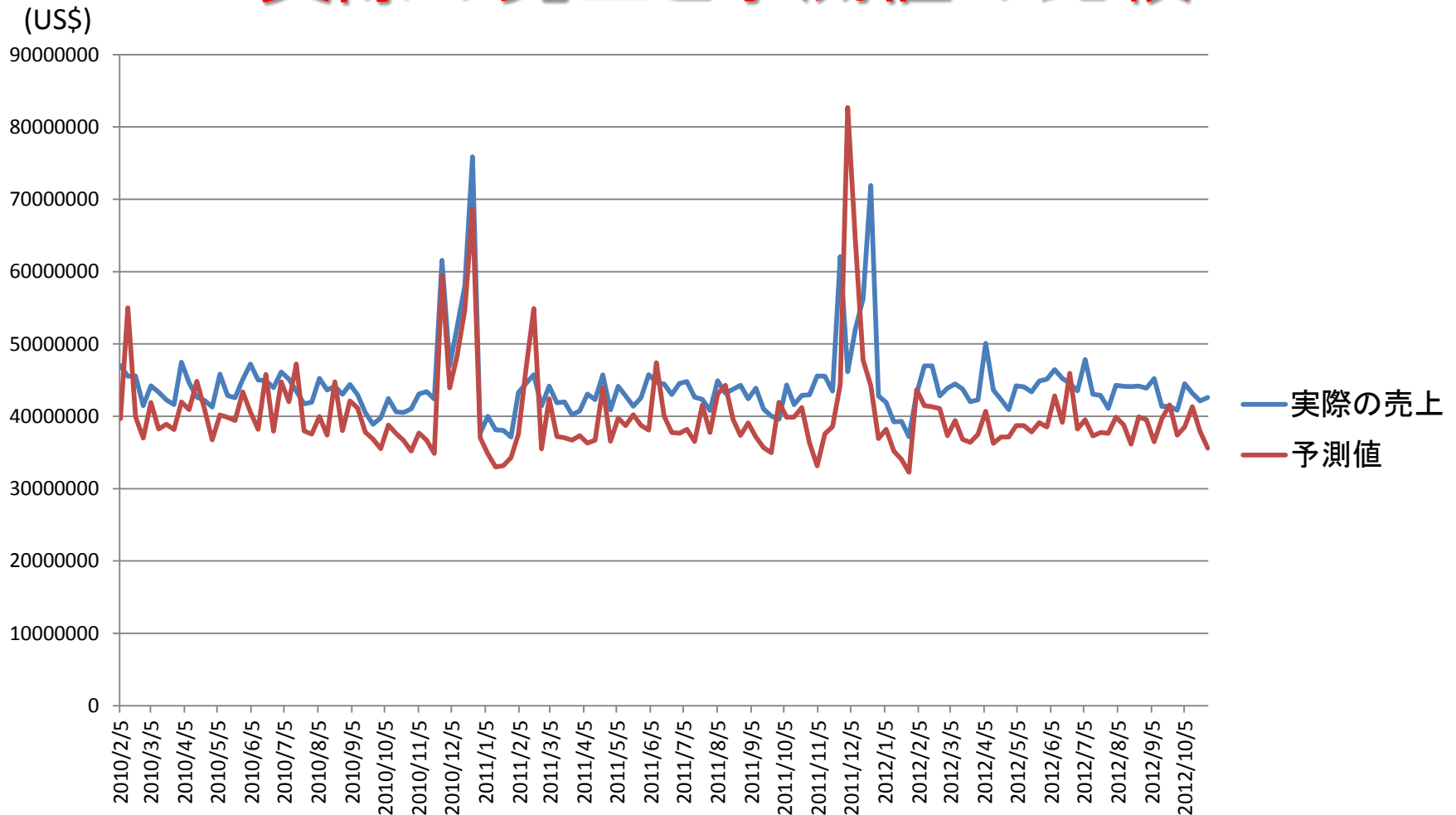
上記以外のイベント

まとめ・課題 分析①～④の比較(売上は各店舗の週ごとの売上)

		分析①	分析②	分析③	分析④
被説明変数		売上	log(売上)	log(売上)	log(売上)
説明変数	(Intercept)	1.26E+06***	1.39E+01***	1.18E+01***	1.19E+01***
	Temperature	2.84E+03***	-4.61E-03***	2.16E-03*	-2.15E-04
	CPI	-1.77E+03***	-9.58E-04***	3.74E-03***	3.65E-03***
	Unemployment	-4.76E+04***	-2.27E-02***	-7.76E-03.	-2.71E-03
	MarkDown1	1.28E+01***	4.44E-06***	3.65E-06***	3.16E-06***
	MarkDown2	1.23E+01***			
	MarkDown3	1.58E+01***			
	MarkDown4	-1.15E+00***			
	MarkDown5	3.45E+01***	4.14E-05***	4.19E-05***	5.00E-05***
	Black_Friday	-2.70E+05***	3.89E-01***	3.92E-01***	4.01E-01***
	Christmas	9.27E+05***	2.21E-01	2.05E-01***	1.93E-01**
	Easter	6.51E+04			
	Independence_Day		5.40E-03		
	February		1.86E-02	1.96E-02	1.86E-02
	March		1.35E-01***	5.29E-02	8.80E-02*
	April		1.64E-01***	2.53E-02	7.97E-02.
	May		2.26E-01***	3.49E-02	1.15E-01*
	June		2.48E-01***	1.87E-03	9.67E-02.
	July		2.43E-01***	-3.34E-02	8.79E-02
	August		2.61E-01***	-1.42E-02	7.98E-02
	September		1.63E-01***	-6.35E-02	1.37E-02
	October		1.76E-01***	2.14E-02	8.39E-02*
	November		6.76E-02.	-4.03E-03	1.77E-02
	December		8.73E-02	9.17E-02	1.09E-01.
	Year2011				-3.08E-03
	Year2012				4.09E-02*
	Area_A			6.79E-01***	6.27E-01***
	Area_B			8.01E-01***	8.24E-01***
	Area_C			1.47E+00***	1.40E+00***
	Area_D			1.02E+00***	9.50E-01***
	Area_E			1.11E+00***	1.05E+00***
	Area_F			1.16E+00***	1.09E+00***
Area_G			1.32E+00***	1.24E+00***	
Area_H			8.94E-01***	8.61E-01***	
N		6435	6435	6432	6004
R2		0.2981	0.2138	0.3159	0.3353

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

実際の売上と予測値の比較



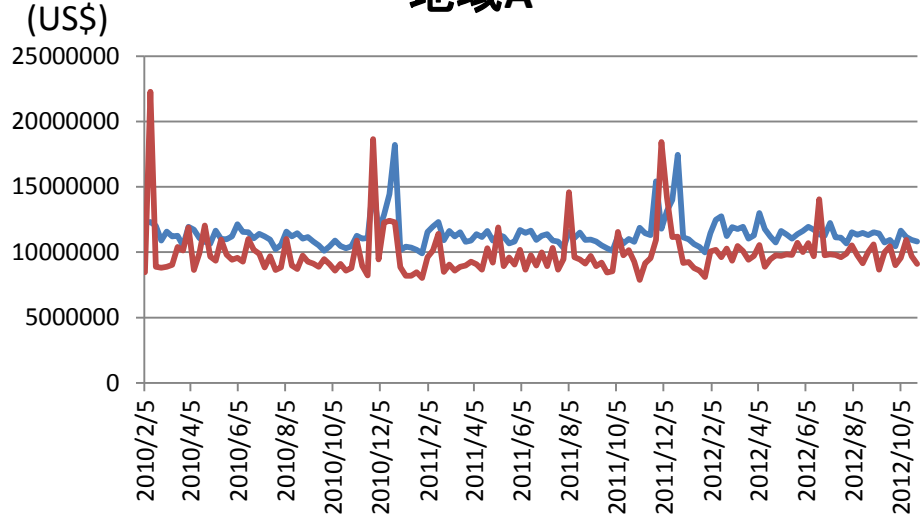
※比較には決定係数が最も高かった分析④の結果を用いた。

※店舗1-45の売上を週ごとにそれぞれ合計した上で比較した。

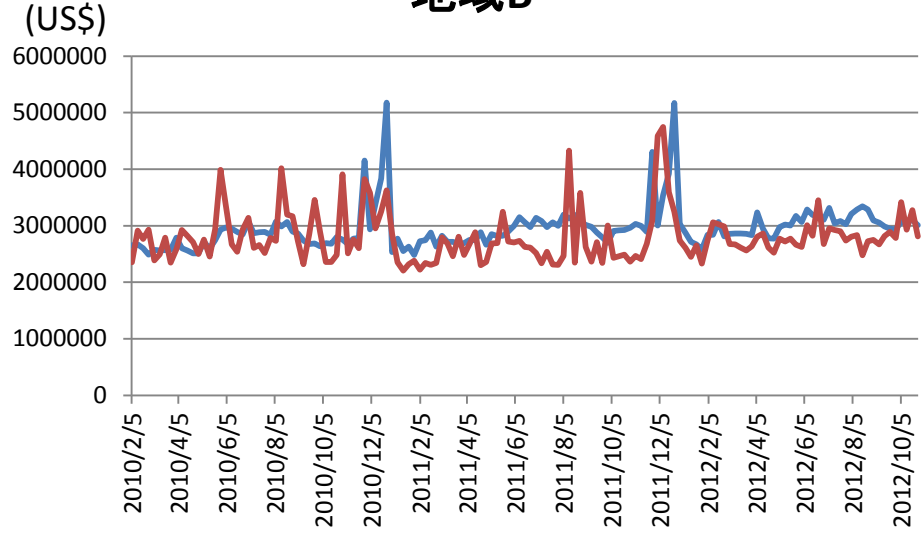
分析3で確認された2012年8月の売上の予測値が異常な値は解決した。

実際の売上と予測値の比較(地域ごと:地域A~D)

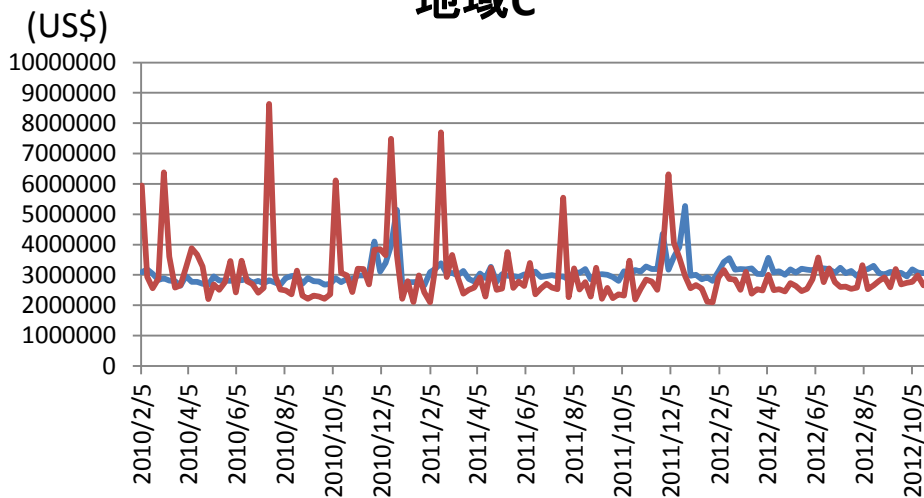
地域A



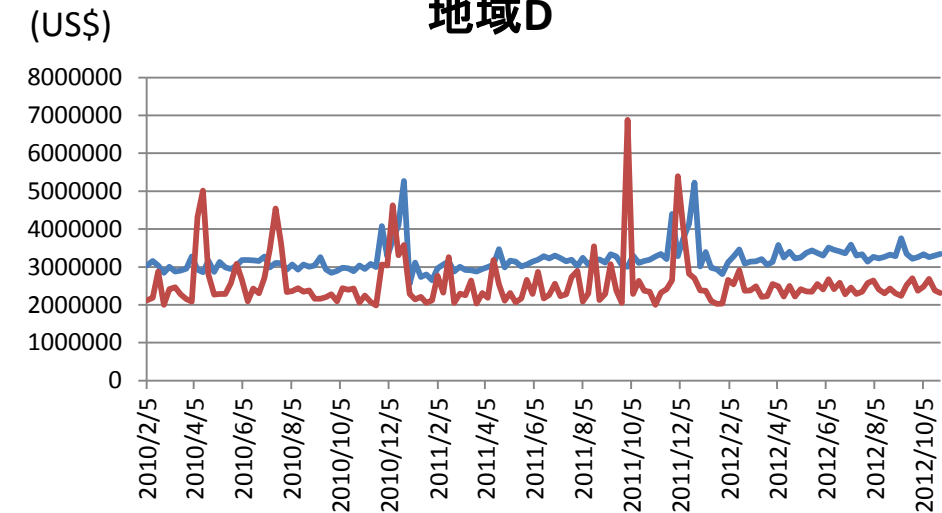
地域B



地域C



地域D



 実際の売上

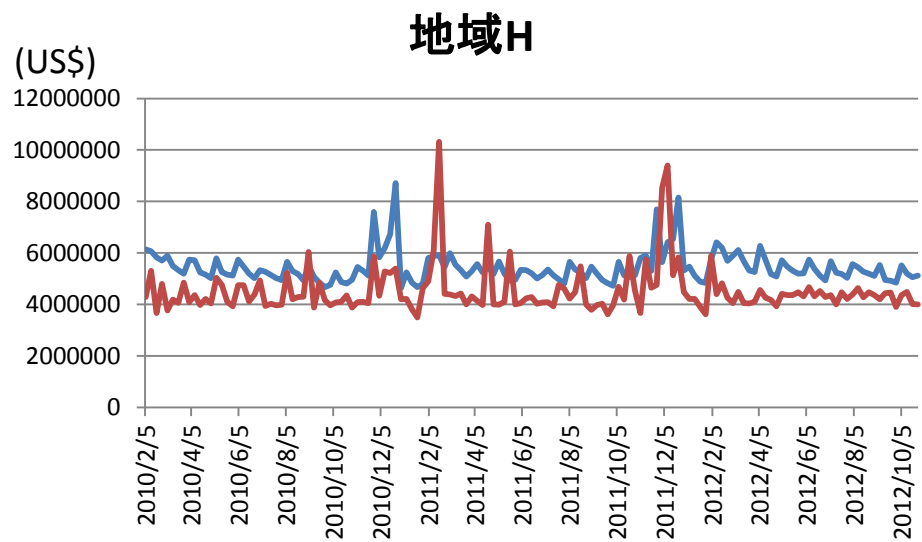
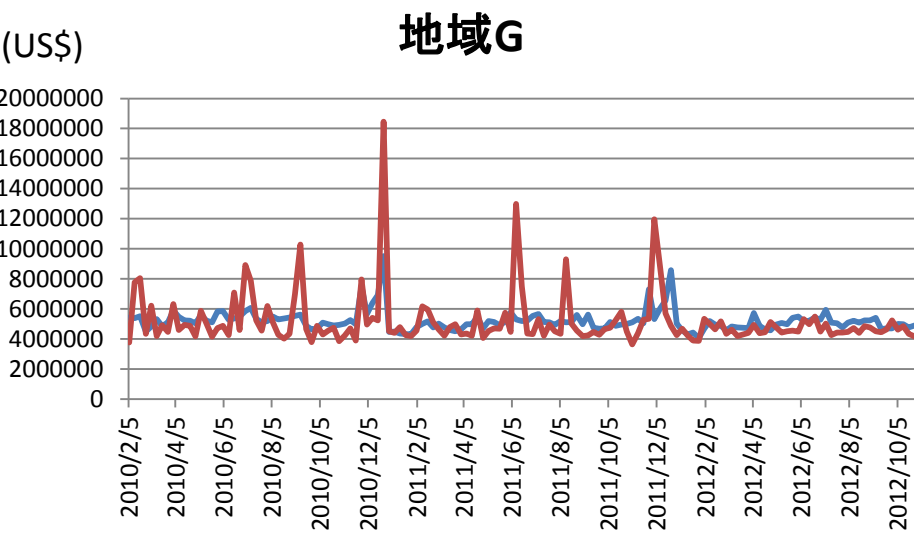
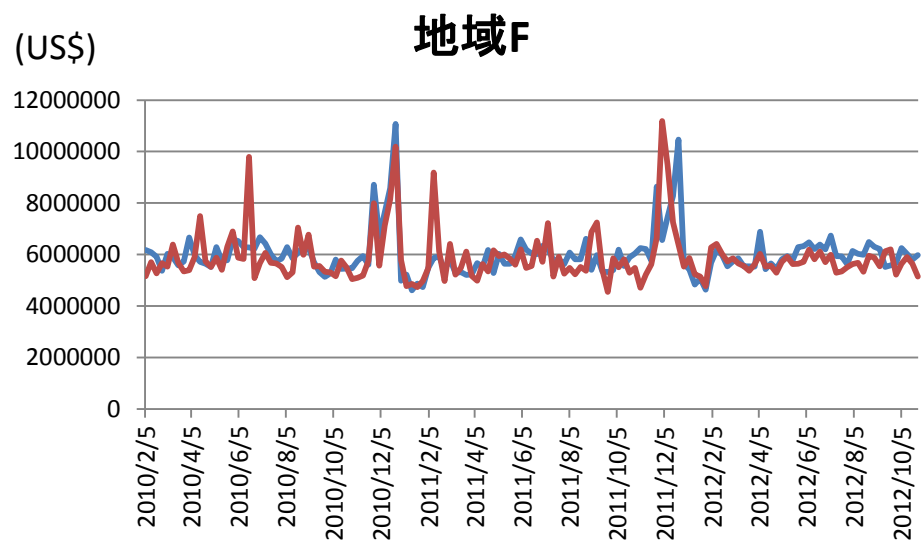
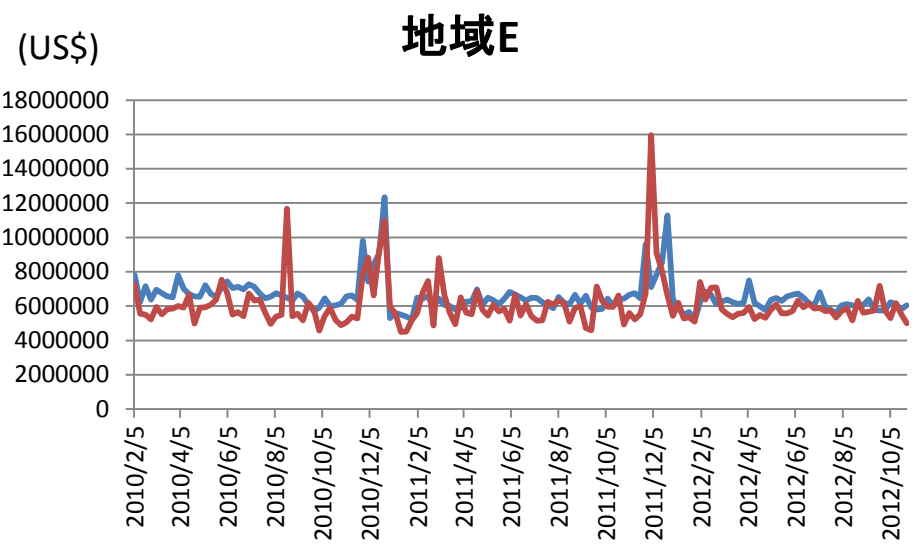
実際の売上

 売上予測

売上予測

※分析④の結果で
予測値を求めた。

実際の売上と予測値の比較(地域ごと:地域E~H)



実際の売上

売上予測

※分析④の結果で
予測値を求めた。

まとめ・課題

- 予測値と実際の売上を比較するとまだまだ当てはまりがよくない店舗が存在していることがわかる。
- 今回のデータは値下げの欠損値が多いなどの問題があったため値下げのデータを完全に使用することができなかった。より詳細な値下げデータが存在することでさらに当てはまりのよいモデルが構築できるであろう。
- 地域についても詳細なグループ分けができれば気温や降水量等のデータと組み合わせることが可能であった。
- 予測値の一部に異常が見られたためその要因を取り除く必要がある。

参考文献

- About connecting the dots.
kaggleで予測モデルを構築してみた (5) - Rで行うMultipleImputation
<http://smrmkt.hatenablog.jp/entry/2013/01/06/164758>(2014-12-01)
- CNN.co.jp「米年末商戦スタート 買い物客数、支出額とも過去最高に」<http://www.cnn.co.jp/business/35024836.html> (2014-11-24)
- Kaggle.com <http://www.kaggle.com/> (2014-11-13)
- USAJPN.COM「アメリカ生活情報」アメリカの行事・イベント一覧
<http://usajpn.com/event.php> (2014-11-30)
- Walmart本社 <http://www.walmart.com/> (2014-11-23)
- yokkunsの日記 回帰分析 - 線形単回帰分析
<http://d.hatena.ne.jp/yokkuns/20110301/1298938050> (2014-12-15)
- 信金中央金庫NewYorkコラム16-18号「アメリカの年末商戦」
<http://www.scbri.jp/HTMLcolumnNY/16-18.pdf> (2014-11-24)

付属資料: 分析①で行った単回帰分析の結果

被説明変数	説明変数	回帰係数	標準誤差	t値	p値	R2
売上	(Intercept)	1165260	24285.4	47.982	< 2e-16***	0.004
	気温	-2052.8	382.9	-5.362	8.53E-08***	
	(Intercept)	1006186	52319	19.232	<2e-16***	-8.62E-05
	燃料価格	10262	15434	0.665	0.506	
	(Intercept)	1244309	31451	39.563	< 2.00E-16***	0.006
	消費者物価指数	-1173.3	178.4	-6.575	5.24E-11***	
	(Intercept)	1280302	30738	41.652	< 2.00E-16***	0.009
	失業率	-29724	3744	-7.938	2.40E-15***	
	(Intercept)	9.34E+05	8.19E+03	114.01	<2e-16***	0.083
	値下げ1	1.56E+01	6.43E-01	24.26	<2e-16***	
	(Intercept)	1.02E+06	7.15E+03	142	<2e-16***	0.040
	値下げ2	1.31E+01	7.92E-01	16.5	<2e-16***	
	(Intercept)	1.04E+06	6.99E+03	148.2	<2e-16***	0.025
	値下げ3	1.10E+01	8.47E-01	13	<2e-16***	
	(Intercept)	1.03E+06	7.20E+03	143.475	<2e-16***	0.011
	値下げ4	2.82E+00	3.29E-01	8.562	<2e-16***	
	(Intercept)	8.58E+05	8.84E+03	97.03	<2e-16***	0.135
	値下げ5	4.04E+01	1.27E+00	31.82	<2e-16***	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 N=6435

付属資料: 分析①で行った単回帰分析の結果

被説明変数	説明変数	回帰係数	標準誤差	t値	p値	R2
売上	(Intercept)	1040945	7057	147.502	< 2e-16***	0.007
	Black_Friday	430285	59674	7.211	6.21E-13***	
	(Intercept)	1036923	7006	148	<2e-16***	0.022
	Christmas	717832	59242	12.12	<2e-16***	
	(Intercept)	1048185	7084	147.956	<2e-16***	0.000
	Year_End	-87358	59904	-1.458	0.145	
	(Intercept)	1045509	7162	145.989	<2e-16***	2.779
	Valentine_Day	41583	38299	1.086	0.278	
	(Intercept)	1045175	7109	147.017	<2e-16***	0.000
	Easter	85215	49083	1.736	0.0826.	
	(Intercept)	1046725	7111	147.19	<2e-16***	-0.000
	Mother_Day	11267	48917	0.23	0.818	
	(Intercept)	1046623	7111	147.19	<2e-16***	-0.000
	Father_Day	16202	49094	0.33	0.741	
	(Intercept)	1047033	7111	147.244	<2e-16***	-0.000
	Labor_Day	-3322	49094	-0.068	0.946	
	(Intercept)	1047730	7136	146.82	<2e-16***	-9.13E-05
	Halloween	-27406	42668	-0.642	0.521	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1 N=6435