

加工データを使用した商品分析

福永裕香¹, 荒平高章¹

概要：加工のされていないデータ（無加工データ）を収集、発見することは難しい。そこで、加工データを使用した商品分析を行うことは可能かを検討したいと考えた。本研究では、無加工データといくつかの加工データの商品分析結果をグラフや表として比較し、データの加工の程度（数値の1の位を四捨五入、10の位を四捨五入）によって分析結果にどれほど違いが生まれるかを検討した。その結果、四捨五入する位ではABC分析、デシル分析どちらの分析結果にも大きな違いが出ないことがわかった。大きな違いが生まれる場合は同じデータと言えない場合であると判明した。

キーワード：Product Analysis, Data processing, ABC Analysis, Decyl Analysis, Rounding of data

1. はじめに

自社製品の商品分析をすることになった環境と同じデータ状態で分析を行う場合、インターネットのデータはほとんどが平均や最大値、最小値、年間別などの大まかな数値とされているものである。嶋田[1]は「1次データは、利用目的に応じて必要なデータを入手できるが、2次データに比べて収集や処理に多くの時間と費用を要する」、渋谷[2]は「1次データとは…費用・時間・手間がかかりますが、目的に合った情報を入手できると、データ分析がしやすく、意思決定の精度が高まります。」と言う一文もあるように加工のされていないデータ（無加工データ）を収集、発見することは難しい。

本研究では、無加工のデータから程度の違う加工データ（数値の1の位を四捨五入、10の位を四捨五入）を作成しt検定によって比較する。次に、無加工データ、加工データの商品分析（ABC分析、デシル分析）を行う。無加工データと加工データの商品分析結果をグラフや表として作成し、データの加工の程度によって分析結果にどれほど違いが生じるかについて考察を行ったので報告する。

2. 使用する商品分析方法

2.1 ABC分析

2.1.1 概要

商品や顧客などの重要な指標を重要度によってABCの3つのランクに分類する分析方法。別名重点分析ともいわれ、売れ筋商品や死に筋商品、優良顧客の把握ができる。

2.1.2 使用方法

Excelを使用。グラフを作成時見やすくなるので軸の書式設定から累積比のみ%に変更。

商品と顧客(店舗)を指標としABCランクに分ける。ランク分けの基準値は決まっていないので自由に決めてよい。A、

B、Cを70、70~90、91~100%とする場合が多い。

A,B,Cとランク分けするには、vlookup関数を用いる。

- ①データをExcelに保存
- ②ABC分析したい項目を選択
- ③項目を降順で並び替える
- ④SUMを使用し累積を求める
- ⑤累積から累積構成比を求める
- ⑥累積構成比からVlookup(選択セル, %表示部分, 列数, TRUE)を使用してA, B, Cランクに分ける
- ⑦⑥までのデータを使用しグラフを作成する

2.2 デシル分析

2.2.1 概要

デシル分析とは、ある商品の購入金額などが高い順に顧客を並べ、10等分してグループごとに購入比率などを分析する方法。上位何%の顧客が何%の売上を占めているのか、購入金額の分布を容易に把握することができ、より効率的に集中したキャンペーンなどのマーケティング活動を行える。

2.2.2 使用方法

Excelを使用。デシルランクごとに購入金額を合計し、売上全体に占める割合(累積構成比)を算出できる。売上を基にユーザー数が等分となるようなグループを作成できる。IF関数を使用。SUMIF関数を利用して割合を算出するとしやす。

- ①データをExcelに保存
- ②顧客ごとに期間内の累積購入金額を計算する
- ③累積購入金額の降順で顧客を並べる
顧客情報の左側に順位を振る
- ④③のリストを10等分し、デシルランクを割り当てる
顧客数が10で割れない場合には、1番下のランクに余りを全て含ませる。

¹ 九州情報大学
Kyushu Institute of Information Sciences

関数 IF を使って計算する

⑤デシルランクごとの構成, 傾向を分析する

SUMIF(デシルランクの範囲, 指定するデシルランク選択, 購入金額の範囲)を使用して表を作成

2.3 t 検定

2.3.1 概要

2 つの独立した母集団があり, それぞれの母集団から抽出した標本の平均に差があるかどうかを判明させる検定
 今回は同一のデータかどうかを判断するため対応のある場合の t 検定. 帰無仮説は「対応のあるデータの差の平均は 0 である」となる.

2.3.2 使用方法

Excel を使用. 無加工データと 1 の位を四捨五入したデータ, 無加工データと 10 の位を四捨五入したデータを比較.
 Excel のデータ分析機能の一对の標本による平均の t 検定ツールを使用.

- ① データを Excel に保存
- ②累積購入金額の降順で顧客を並べる
- ③無加工データと加工したデータをデータ分析ツールで分析する

t 検定: 一对の標本による平均の t 検定ツールを使用

3. 使用データ

データ名 Wholesale customers data

Csv データ. 列 8 行 440.

1) FRESH: 新鮮な製品に対する年間支出 (mu) (連続);

2) MILK: 乳製品の年間支出 (mu) (連続);

記述統計量:(最小、最大、平均、標準偏差)

新鮮な製品 (3, 112151, 12000.30, 12647.329)

乳製品 (55, 73498, 5796.27, 7380.377)

4. 分析

Table1 と Table2 より新鮮な食品の支出データの無加工データと 1 の位を四捨五入したデータ, 無加工データと 10 の位を四捨五入したデータをデシル分析した結果を比較するとデシルランク別の差も合計差も 1 の位を四捨五入したデータの方が小さくなった. しかし, Table3 と Table4 から乳製品の支出データの無加工データと 1 の位を四捨五入したデータ, 無加工データと 10 の位を四捨五入したデータをデシル分析した結果を比較すると各デシルランク差が 10 の位を四捨五入したデータの方が小さい場合もあった. 合計の差で比べても 10 の位を四捨五入したデータの方が小さいことが分かった. Table5 と table6 から新鮮な食品の支出データの無加工データと 1 の位を四捨五入したデータ, 無加工データと 10 の位を四捨五入したデータを ABC 分析し

た結果では, 合計差は 1 の位を四捨五入したデータの方が小さくなったもののランクごとの差では 1 の位を四捨五入したデータの方が大きな値が出た. 以上から ABC 分析でのデシル分析でも 10 の位を四捨五入したデータより 1 の位を四捨五入したデータの方がクラスごとの差や合計差も小さくなるとは言えないことが明らかとなった.

Fig1 から新鮮な製品の支出データの無加工データと 1 の位を四捨五入したデータを t 検定すると 0.05 よりも小さくなり同じデータと判断されないことがわかった. これが原因で ABC 分析の結果が無加工データや 10 の位を四捨五入したデータものと数値が変わった可能性があると考えられる. 同様に, Fig2 から乳製品の支出データの無加工データと 1 の位を四捨五入したデータを t 検定すると 0.05 よりも小さくなり同じデータと判断されないため, デシル分析の結果が無加工データや 10 の位を四捨五入したデータものと数値が変わった可能性があると考えられる.

データの数値を 1 の位や 10 の位で四捨五入して使う場合には, t 検定などを先に行い四捨五入したデータが無加工データと少なくとも同じデータかもしれないと判明してから分析に使う必要があることが判明した.

Table 1. Decyl Analysis, Fresh, raw data and data rounded to the nearest 1

| | 新鮮な食品の支出 (無加工) | 新鮮な食品の支出 (1 の位を四捨五入) | 差 |
|-----|-------------------|-------------------------|------|
| D1 | 1790532 | 1790550 | -18 |
| D2 | 990688 | 990740 | -52 |
| D3 | 739252 | 739280 | -28 |
| D4 | 546980 | 547010 | -30 |
| D5 | 425320 | 425340 | -20 |
| D6 | 319807 | 319810 | -3 |
| D7 | 229647 | 229650 | -3 |
| D8 | 142843 | 142900 | -57 |
| D9 | 77807 | 77820 | -13 |
| D10 | 17255 | 17250 | 5 |
| | | 合計差 | -219 |

Table 2. Decyl Analysis, Fresh, raw data and data rounded to the nearest 10

| | 新鮮な食品の支出 (無加工) | 新鮮な製品の支出 (10の位を四捨五入) | 差 |
|-----|-------------------|-------------------------|------|
| D1 | 1790532 | 1790700 | -168 |
| D2 | 990688 | 990500 | 188 |
| D3 | 739252 | 739300 | -48 |
| D4 | 546980 | 546900 | 80 |
| D5 | 425320 | 425900 | -580 |
| D6 | 319807 | 320100 | -293 |
| D7 | 229647 | 229600 | 47 |
| D8 | 142843 | 142700 | 143 |
| D9 | 77807 | 77600 | 207 |
| D10 | 17255 | 17200 | 55 |
| | | 合計差 | -369 |

Table 3. Decyl Analysis, Milk, raw data and data rounded to the nearest 1

| | 乳製品 (無加工) | 乳製品 (1の位を四捨五入) | 差 |
|-----|--------------|-------------------|-------|
| D1 | 972960 | 911730 | 61230 |
| D2 | 448353 | 444380 | 3973 |
| D3 | 318813 | 316830 | 1983 |
| D4 | 246914 | 245500 | 1414 |
| D5 | 182821 | 181600 | 1221 |
| D6 | 141103 | 140190 | 913 |
| D7 | 97192 | 96420 | 772 |
| D8 | 69380 | 68790 | 590 |
| D9 | 47036 | 46680 | 356 |
| D10 | 25785 | 24990 | 795 |
| | | 合計差 | 73247 |

Table 4. Decyl Analysis, Milk, raw data and data rounded to the nearest 10

| | 乳製品 (無加工) | 乳製品 (10の位を四捨五入) | 差 |
|-----|--------------|--------------------|--------|
| D1 | 972960 | 985400 | -12440 |
| D2 | 448353 | 444400 | 3953 |
| D3 | 318813 | 317000 | 1813 |
| D4 | 246914 | 245800 | 1114 |
| D5 | 182821 | 181500 | 1321 |
| D6 | 141103 | 140300 | 803 |
| D7 | 97192 | 96200 | 992 |
| D8 | 69380 | 68900 | 480 |
| D9 | 47036 | 46300 | 736 |
| D10 | 25785 | 25000 | 785 |
| | | 合計差 | -443 |

Table 5. ABC Analysis, Fresh, raw data and data rounded to the nearest 1

| | 新鮮な製品 (無加工) | 新鮮な製品 (1の位四捨五入) | 差 |
|---|----------------|--------------------|-------|
| A | 4274713 | 4265170 | 9543 |
| B | 525527 | 535220 | -9693 |
| C | 479891 | 479460 | 431 |
| | | 合計差 | 281 |

Table 6. ABC Analysis, Fresh, raw data and data rounded to the nearest 10

| | 新鮮な製品 (無加工) | 新鮮な製品 (10の位四捨五入) | 差 |
|---|----------------|---------------------|------|
| A | 4274713 | 4275000 | -287 |
| B | 525527 | 526100 | -573 |
| C | 479891 | 479400 | 491 |
| | | 合計差 | -369 |

| t-検定: 一対の標本による平均の検定ツール | | |
|------------------------|-------------|---------|
| | 無加工 | 1の位 |
| 平均 | 12000.2977 | 12000.8 |
| 分散 | 159954927 | 1.6E+08 |
| 観測数 | 440 | 440 |
| ピアソン相関 | 0.99999997 | |
| 仮説平均との差異 | 0 | |
| 自由度 | 439 | |
| t | -3.60209128 | |
| P(T<=t) 片側 | 0.00017588 | |
| t境界値 片側 | 1.64833201 | |
| P(T<=t) 両側 | 0.00035175 | |
| t境界値 両側 | 1.96538247 | |

Fig 1. T-test, Fresh, raw data and data rounded to the nearest 1

| | 無加工 | 10の位 |
|------------|-------------|----------|
| 平均 | 12000.2977 | 12001.14 |
| 分散 | 159954927 | 1.6E+08 |
| 観測数 | 440 | 440 |
| ピアソン相関 | 0.99999747 | |
| 仮説平均との差異 | 0 | |
| 自由度 | 439 | |
| t | -0.61699091 | |
| P(T<=t) 片側 | 0.26878029 | |
| t 境界値 片側 | 1.64833201 | |
| P(T<=t) 両側 | 0.53756058 | |
| t 境界値 両側 | 1.96538247 | |

Fig 2. T-test, Fresh, raw data and data rounded to the nearest 10

5. まとめ

本研究では、無加工データといくつかの加工データの商品分析結果をグラフや表として比較し、データの加工の程度（数値の1の位を四捨五入、10の位を四捨五入）によって分析結果にどれほど違いが生まれるかを検討した。その結果、四捨五入する位ではABC分析、デシル分析どちらの分析結果にも大きな違いが出ないことがわかった。大きな違いが生まれる場合は同じデータと言えない場合であると判明した。t検定の結果が帰無仮説と証明されず同じデータといえなかったデータと無加工データの分析結果がABC分析、デシル分析とも帰無仮説が証明された結果と無加工データを比べたものより差が大きくなったことが裏付けである。本研究では、2つのデータを使用しているがどちらもデータ数が440で同じものであり、データ数が異なるもので検証することが課題である。

参考文献

- (1) 嶋田毅, "MBA 定量分析と意思決定", グローバル・マネジメント・インスティテュート(2014), 定量分析のステップ, Step3 情報を収集する.
- (2) 渋谷智之『データ利活用の教科書』(2022), データと20年向き合ってきたマクロミルならではの成功法則, pp243.
- (3) ID レシート BI ツール, Felice Networks, 商品分析方法7選! マーケティングに活かせる知識を学ぼう.
<https://receiptreward.jp/solution/column/product-analysis.html>
- (4) Udemy メディア, 株式会社ベネッセコーポレーション, "ABC分析とは? Excelのやり方を覚えれば在庫管理が楽になる".
<https://udemy.benesse.co.jp/marketing/basic/abc-analysis.html>.

(5) ビジネス+IT, SB Creative, "ABC分析とは? 目的とやり方と手順、パレード図作成方法を解説".

<https://www.sbit.jp/article/cont1/53364>

(6) 365日 Office ライフ!, "ROUND関数で簡単に四捨五入! わかりやすく図解します. 四捨五入(=ROUND)".

<https://www.onamae-office.com/office365column/excel/round/>

(7) FUSION デシル分析・RFM分析, "Excelでできる顧客分析入門".

<https://www.fusion.co.jp/column/2015/10/excel-decile-rfm>

(8) CUI ホールセール顧客データセット, Margarida GMS Cardoso, margarida.cardoso '@' iscte.pt, ISCTE-IUL, リスボン, ポルトガル.

<https://archive.ics.uci.edu/ml/datasets/Wholesale+customers>

(9) データに築くビジネス戦略, アイトクコンサルティング, "2つのグループを比較, 差は本当にあるのか? t検定: 1対の標本による平均の検定ツール".

<https://aitokuconsult.com/dataanalysis/statistic0908/>

(10) AVILEN AI Trend, ライター古澤嘉啓, "t分布表".

<https://ai-trend.jp/basic-study/t-distribution/t-table/>