

経営データの特質と統計分析(データ解析)の可能性

東 瀧 則 之

1. はじめに

経営意思決定に役立つ情報を得るためデータ解析が用いられる。データ解析の材料であるデータに関しては、測定の直接性や実験環境の独立性など多くの点で自然科学の分野のデータと比べ、ビジネス分野のデータは質が悪いと考えられている。

たとえば, English は, 著書において, 以下に示すような高コストで低品質なデータの事例を多く紹介している。「ヨーロッパのある企業は, 注文の4%が未請求だったことを監査によって突きとめた。20億ドルの売上のあるこの企業にとって, これは8,000万ドルの注文が未収となっていることを意味する…。」

「電子データ監査によって, 一般的な顧客データベースに含まれる無効データは平均15~20%に達することが示されている。同じデータベースに対する実際のデータ監査の結果, 値が有効であったとしても, 実際のデータエラーコストは25~30%かそれ以上になることがあると示された。この低品質データのコストは, 顧客との無駄なコミュニケーションの形で, 経営上の利益を削り取る。…」

「ナッシュビル市区の一部の年金受給者は1987年から1995年に230万ドルも超過支給されていたが, 他のあるグループは260万ドルが未支給であった。これは不正確な年金計算の結果である。…」

English は, また, 低品質データのコストは, 回収不能コスト, 製品者サービスのやり直し, 仕掛かり, 損失及び取りそこなった収入などを評価すると, 収入または企業の総予算の10~25%となるかもしれないとも述べている。(English, 1999)

本稿では、このように品質が自然科学に比べて劣ると思われる経営データの特質について改めて検討し、データ解析の適用可能性について論じる。(なお、Englishは、経営者は、現状のコストと低レベルのデータ品質を、正常コストかつ許容範囲内の経営コストとみなしているか、低品質データの本当のコストに気づいていないかのいずれかである、と述べている。)

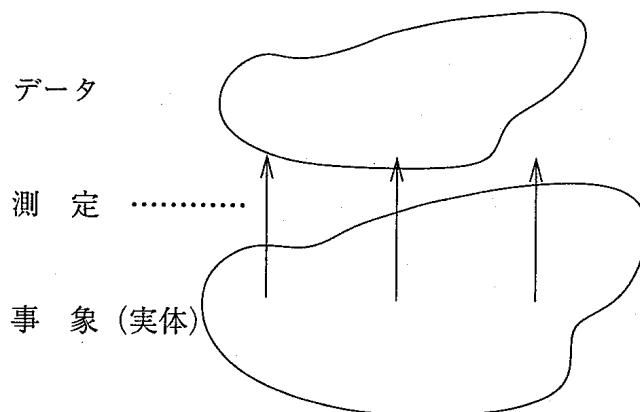
2. 経営データ品質の捉え方

ービジネス分野のデータ解析における経営データ品質の評価枠組みー

(1) データとは何か?

現実の事象をなんらかの記号によって表現することが「測定」である。表現された記号が「データ」である。また、データの品質の良否は、現実の事象を正確に再現できる程度によって判断される。

図表1 事象, 測定, データ



(2) 絶対的な経営データ品質 (誤差)

工学等の分野では、測定値の誤差の検討に当たっては、一般に、「信頼性、精度、正確さ」の各項目に分けて行われている(藤森, 1986)。これをベースとして、経営データの品質(誤差)について、次のように定義する。

a. 測定の信頼性

測定に信用が置けるかどうかという視点。自然科学分野では、測定操作に何か間違いやミスという異常がないか、測定装置の故障や狂いはないか、社会科学分野のアンケートならば、被験者のサンプリングが適切か、ワーディング等調査票は正しく作られているか、実施・回収は適切に行われたか、などの問題である。公刊データ、業務データ、会計データ等の場合、記録プロセスがどれだけ信頼できるかという問題である。

b. 測定値の精度

「精度」は、同一試料または同一とみなされる試料を数多く測定したり、あるいは同一ロットまたは同一とみなされるロットから数多くの資料を採取して1回ずつ測定した場合に得られるばらつきの尺度であって、不偏分散、標準偏差、変動係数、範囲、範囲の上方管理限界、精度 β （たとえば2シグマ）などいろいろな表現方法がある。社会科学のアンケートの場合は同一母集団からのサンプルからの回答のばらつきの問題である。

c. 測定値の正確さ

同一試料あるいは同一とみなしうる試料を管理状態で多数回測定したとき、その平均値 \bar{x} の真の値 μ からの偏り、 $\bar{x} - \mu$ を「正確さ」という。ここで重要なのは、真の値とは何かという問題である。理論的に推定できない場合、真の値をどのように定義づけするかという問題がある。測定には誤差を伴うから、測定によって本当の意味での真の値を知ることは困難である。真の値に関して、次の3つの場合が考えられる。イ) われわれが求めようとする真の値が絶対的な確定数として存在し、しかも測定可能である場合、ロ) 真の値が絶対的な確定数として存在するはずであるが、測定が非常に困難な場合、ハ) 真の値が絶対的な確定数でなく、ある定義によって定まるような値である場合（真の値とは、その定義いかんによって決まる値で絶対的なものではない。したがって、標準試料の標準値などを決めるときにはまず測定方法も含めて定義することが重要である。）

(3) 相対的な経営データ品質の定義

データ品質がどの程度高ければよいかという質問自体、無条件ではほとんど意味を成さない。たとえば、アメリカにおいては住宅建築で使う木材の寸法は32分の3インチ（約2.4ミリ）までのばらつきが許容される。その一方、光通信ケーブルに使う光ファイバーの寸法は、直径±0.5ミクロン（1万分の5ミリ）以下のばらつきに収めなければならないことが仕様で決められている。市の人口数の精度は、何らかの支出金を割り当てるという観点から（人口調査のカウントに対して）正確でなければならないが、マーケティングが目的なら1,000人単位の精度でも十分であろう。精度に関して複数の要件が存在する場合、最も厳しい要件を要求精度として採用する必要がある（English, 1999）。

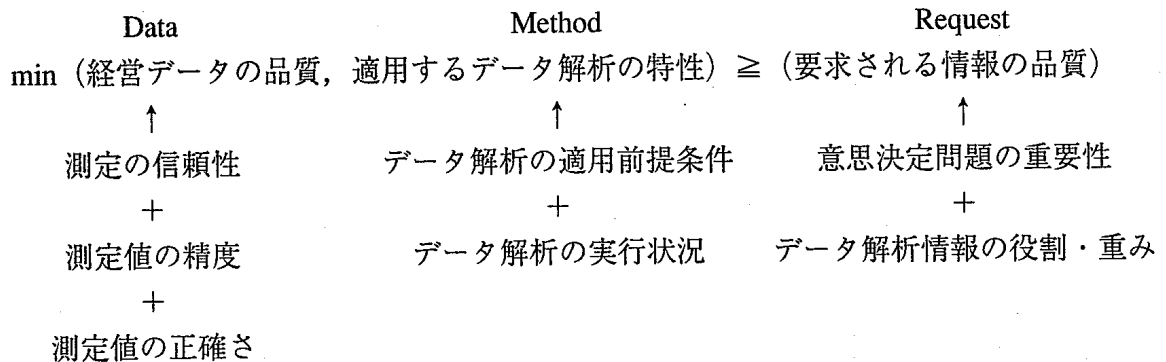
すなわち、同一のデータであっても、意思決定問題によっては要求される情報品質との比較で、品質が高いと評価されたり、低いと評価されたりする。また、多くの場合、データは何らかの解析を施されて情報抽出されるので、データ解析手法の特性との絡みも出てくる。すなわち、 $\min(\text{経営データの品質}, \text{適用するデータ解析の特性}) \geq \text{要求される情報品質}$ 、であることが必要である。要求される情報品質は、一般にビジネス分野においては自然科学に比べて低いが、当該意思決定問題の重要さやデータ解析情報のウエイトなどの違いによって高低がある（東淵, 1999）。このような必要とされる情報品質やデータ解析手法の特質と相対したときの経営データ品質を相対的な経営データ品質と呼ぶ。

(4) ビジネス分野のデータ解析における経営データ品質の評価フレーム

以上の議論をもとに、経営データの品質の評価に関するフレームワークを示すと、次図のようになる。

なお、要求される情報の品質に係る意思決定問題の重要性及びデータ解析情報の役割・重みについては東淵（1999）を参照されたい。

図表2 経営データの品質の評価に関する DMR フレームワーク



3. 経営データの品質の検討

官公庁の各種統計や業界団体などの調査データなど公刊データ、企業が独自にアンケートを実施し調査したデータ、顧客データ、日常の業務記録データ、会計データ、さらにこれらをもとに加工作成したデータなどがある。データ品質の概念に従って、これらのデータの特質について検討する。

(1) 公刊データ

a. 測定の信頼性

一般的に、官公庁等の公刊する統計データの測定の信頼性は概ね高いと見られている。しかし、調査統計の場合、被験者がデータ記入する際、回答者バイアスのため期待ほど測定の信頼性が確保されていない。ある回答者がある質問に対してイエスと答えたからといって常にイエスと答えるとは限らない。例えば、林はその著書で、同じ対象者に同じ質問を2回行うと、本来ならば100%一致すべき対象者の学歴でも完全一致は75%に過ぎないと述べている（林，1993）。また、ヘックマンは、統計調査における「対象者の選択」においてすらバイアスが発生するとしている。この選択バイアスとは、ランダムに集められたと信じている情報やデータが、実はデータを収集する時点で何らかの条件が付加されたためにデータの中立性が損なわれている状態のことである。たとえば、政

府の失業統計調査の場合、サンプルを抽出する際、必然的に、短期間失業状態にある人よりも長期間失業状態にある人のほうがサンプルとして抽出されやすい傾向にあり、結果的に平均失業期間は現実よりも長くなってしまふ。

b. 測定値の精度

センサス、サンプリング調査を問わず、公刊データは、データ数が比較的多く、その精度は概ね高いと見られる

c. 測定値の正確さ

これも概ね高いと見られるが、利用の目的と整合している必要がある。例えば、先の失業率統計。失業者の中には、中小企業が倒産して退職金も出ず、本当に生活に困る人もいる。しかし、失業率を押し上げている大企業の希望退職者の場合、相当額の退職金をうる人が少なくないし、雇用保険も出る。将来はともかく、贅沢しなければ、2、3年は働かなくても何とか生活できる失業者である。突然、明日から来なくてよいという、アメリカなどでの失業とは大きく異なる。政策的に緊急に対処すべき失業状態を真の状態とするなら、それと失業統計で把握されるデータとの乖離は小さくはない（勝見，2002）。

(2) 実験データ

a. 測定信頼性

自然科学分野の実験データは、例えば実験室など外部環境から遮蔽された中で計画的に収集される場合が多いのに対して、ビジネス分野の実験データを得るには、例えば、BGMを変えて購買行動の変化を見たり、棚割を変えて売上の推移を見たりするなど、消費者を対象に店舗内実験を行うことが多い。その測定の信頼性は実験環境を操作変数以外固定しておくことができるかどうかによる。

b. 測定値の精度

測定対象が人間行動に関わることが多く、自然を測定対象にするようなレベルでの精度の高さは期待できない。変動がかなり大きくなるのも事実である。

その意味で精度は自然科学の実験よりは大きく劣ると思われる。

c. 測定値の正確さ

よく設計され管理された実験が行われると仮定すると、統計的な誤差を持つものの正確な測定値が得られる。

(3) アンケート調査データ

a. 測定の信頼性

アンケート調査においては、被調査者の抽出、アンケート票の作成、調査実施など、すべてのプロセスが適切に実施されている必要がある。しかし、実際に行われているアンケート調査を見ると、測定の信頼性が疑わしいものも少なくない。被調査者の抽出の段階では、本来ならば例えばターゲットとする消費者から無作為に抽出される必要があるが、必ずしも無作為抽出されているわけではない。アンケート票においても、質問や選択肢のワーディングが回答を誘導していることも少なくない。回答する際、例えば、調査者の製品について、本心ではよくないと思っけていても、面と向かってははっきりとそう答える人は、よほど勇気があるか正直な人で、たいていは本心を濁した回答をしがちである。生活は苦しいかと聞かれると、多くの人は明日の生活に困っているわけではないが、「はい」と答えがちである（勝見，2002）。

b. 測定値の精度

被調査者の側の問題として回答者バイアスがある。ある回答者がある質問に対してイエスと答えたからといって常にイエスと答えるとは限らない。例えば、同じ対象者に同じ質問を2回行うと、本来ならば100%一致すべき学歴でも完全一致75%、類似したものを含めてやっとならば95%になるという。1年間経過後に、同じ質問に対する回答の一致度を調べたところ、例えば、売春禁止法の改正に対する賛否では60%強、衣食住が洋風か和風か（5段階）では40%強、SD法（7段階）で取った回答などもこの程度に低いことが確認されている（林，1993）。これらのことからわかるようにアンケート調査の測定値の

精度はあまり高くはない。

c. 測定値の正確さ

一般にアンケート調査の回収率は30%前後であることが多く決して高くない。アンケート調査の測定値の正確さは一般に高くはない。回収率60%未満のアンケート調査結果の鵜呑みは極めて危険であるが、低回収率を理由にしてこれをまったく無視してしまうのは問題であろう。小標本データの利用に当てはまることだが、可能な限り他の関連情報を活用して回収不良からくる調査結果の偏りを検討評価した上で、これを利用していくというのが、より妥当な態度というべきであろう(古寺, 1981)。

(4) 顧客データ

a. 測定の信頼性

日常の業務やアンケート調査によって収集した顧客情報が山のようにあるからと言ってもまったく安心できない。例えば、顧客がアンケートに適当に記入したり、書き間違いをしてしまうことは容易に想像できる。アンケートではなく正式な会員制プログラムの加入申込書であっても、年齢、性別、連絡先を記入することに抵抗を感じる人は意外と多い。企業側も強制できずに記入漏れがあっても受け付けざるを得ないことが多い。また、住所変更がメンテナンスされていないことも多い。結局データそのものが信頼できなくなってしまう(アンダーセン, 1998)。顧客データの測定の信頼性は、企業の情報資源に対する意識と行動が大いに反映する。

b. 測定値の精度

企業ごとの取り組みの差によって、顧客データの精度は様々である。顧客データベースを構築しているメーカーは多くはない。消費財メーカー担当者の話を聞くと「データがない」「簡単にデータを取れる方法はないか」「データを購入できないか」など悩みは多い。データは企業にとってお金と同様の貴重な資産であると認識し、何が何でも一心不乱にデータを集める強い意思と全社的な

協力が何より必要である。営業マンの日報や販売記録、クライアントとの取引伝票、契約書や注文書、配達伝票、お客様からのクレーム情報、外交員の手帳メモ、会社のあちこちの部署に存在するデータを集めるだけでも結構な顧客データになると思われる（荒川，2002）。

c. 測定値の正確さ

精度と同様である。

(5) 業務記録データ

日常業務の遂行と共に自然発生的に蓄積される業務データは大きなデータ源である。

a. 測定の信頼性

業務の遂行を記述するという意味で「測定」と考えられる。しかし、業務の遂行が主たる行為であるため、それを記述する記録は、ややもすると関心外になることも少なくない。業務データが業務の測定という点で不完全であるケースの例を挙げる。これらは珍しいことではなく、むしろ普通である。

品切れの存在…品切れを起こしている売り場でいくら POS データ収集しても、真の売れ筋は把握できない。

売上げ操作…月末になるとノルマ達成のため商品を小売に無理やり押し込んだり、逆に、今月の売上げを来月に繰り延べることもある。これでは真の需要量は得られない。

販売状況…通常価格での売上げと特価での売上げを区別して記録できていなければ、価格と売上げとの関係は分析できないが、そうならない場合がある。

定義の不一致…「売上げ」は「注文を受けた時点か」「伝票を切った時点か」「実際納品した時点か」が統一されていない場合がある。

人為的ミス…例えば、注文処理を発注システムで実行せずに電話やファクシミリで済ましてしまいコンピュータ上に未登録となってい

ることも珍しくない。

販売記録ミス…返品処理等を正しく記録していない。

人間特有のケアレスミス…聞き間違い、記入間違いなど（一般に、伝票の2～3%はミスがあると言われている。筆者の実態ヒアリングでも同様。）。

b. 測定値の精度

企業における管理の厳格さや情報資源への認識の違いによって、企業間でかなり差がある。

例えば、セブンイレブンはPOSデータの徹底的分析・活用で有名であるが、本当のすごさはデータの正確さを維持するケイパビリティにある。商品の読み取りに加え、必ず顧客の性別・年齢層を入力しなければレジが開かない仕組みになっている。ところがデータの正確さという面では大きな抜け穴がある。正しいキー入力が行われる保障はどこにもない。他のコンビニでも同種のオペレーションを行っているが、有効な情報を集められているところは極めて少ない。その点、セブンイレブンは違う。その店のロケーションや販売の日付・時間帯といった情報をこれまでに蓄積した全国の膨大な購買データのパターンと組み合わせることで、入力されたデータの異常を識別するケイパビリティを持っている。この他にも店内ビデオの活用や地区のスーパーバイザーといった人間系での対応も駆使して、データの正確性を確保している（アンダーセン、1998）。

c. 測定値の正確さ

測定値の精度と同様である。

(6) 会計データ

a. 測定信頼性

会計は活動を貨幣尺度で記録したものである。その記録作業が、一定の会計処理方法に則って行われている限りは、一定の信頼性を持つと言える。

b. 測定値の精度

会計記録が厳格に行われている企業では測定値としての精度は高いと言える。管理の行き届いた企業であれば、勘定科目の内容説明や取り扱い方法について説明した勘定科目台帳が作成されている。反対に、勘定科目はあってもきちんと体系的に管理されていない企業も少なくない。

c. 測定値の正確さ

1つの事柄に対して複数の会計処理方法が認められており、その選択が経営者の判断にゆだねられている点は注意を要する。基本的には取得原価主義であり、貸借対照表の評価額が現実と乖離する場合があること、また、取引自体が計上されないオフ・バランス取引の存在もあり、財務諸表データの活用には注意を要する（鳥邊ほか、1996）。

(7) その他、収集されたデータ・加工して得られたデータ

データ収集には困難がつきものである。例えば、運動靴を仕入れるのに際して、近隣の小学校の下駄箱にある靴のサイズを、先生に承諾を得た上で数えさせてもらうこともあると聞く。また、他社の缶ジュースの銘柄別の売れ行きを調べるため自販機横の空缶入れを漁って、捨てられた空き缶を銘柄別に数えることもありうるという。データを集められない、できないという理由には事欠かない。しかし、それでもなおギリギリのところまでデータを収集しようとする執念が一步上をいけるかどうかの分岐点である。また、直接必要なデータがなくとも、わずかの加工で演繹的に見られるデータは数多くある。売上げデータと客数のデータがあれば、結果として客単価のデータをつくることができる。過去の新製品の売上げ推移をいくつか集めることにより、自社商品のライフサイクルに関する基礎データを作ることもできる。社内に直接的なデータが存在しなくても、営業セクションなどの現場を通して、取引先の持つ競合品に関する基礎データを収集することもできる。必要なデータは、必ずしも必要な形で存在するとは限らない。大切なことは、常に目的を明確にし、執念を持ってデ

ータを収集することである。

a. 測定の信頼性

ケースバイケースで測定の信頼性は異なる。例えば、先の運動靴調査は高い信頼性を持つと思われるが、自販機調査ではある程度割り引いて評価されるべきである。加工されてえられるデータに関する測定の信頼性は、そのもとになるデータの測定の信頼性に依存する。

b. 測定値の精度

測定の信頼性と同様である。

c. 測定値の正確さ

測定の信頼性と同様である。

4. 経営データの解析可能性

前節において経営データの品質について検討した結果、工学や理学、生物学等の調査・実験データの「測定の信頼性、測定値の精度、正確さ」に比べて劣ることが示された。しかし、このことは経営データを解析することが無意味であることを意味するものではない。2節において示されたように、経営データの解析情報が意思決定において果たす役割との対比で経営データの実用上の品質あるいは満足度は判断されるべきである。

(1) 意思決定におけるデータ解析情報の役割の観点からの評価

経営意思決定に用いられる情報として、データ及びデータ解析情報のほかに、定性情報・事実情報、経験、直感を挙げた。自然科学の場合には、データ解析情報は、仮説の発見や検証に際して第一に位置づけられ、そこを基盤に、定性情報や事実情報、経験、直感が付加されることが多い。しかし、経営管理の場面では、必ずしも第一位に位置づけられるとは限らない。確かに統計的品質管理の場合には第一位に位置づけられるだろうが、そうでない場合も多い。経営意思決定に関する情報は数値データ化できるものばかりではない。しか

も、その信頼性も必ずしも高くはない。他の情報源との比較において、自然科学の場合のように扱われない。

逆に言うと、データ解析情報は概略が正しければ十分その職責を果たすとも考えられる。この点について、ホワイトヘッドらは、「しばしば、正確な数値は重要とは考えられない。むしろトレンドが重要である。経営者や政府は誤った方向のトレンドを正しい方向に向けるよう行動をとる。速やかに行わなければならない。時間は大切な要素である。いたずらに精度を求めるのはよくない。」と述べている。また、内田は「私がよく話すのは「データは有効数字1桁でいい」ということです。例えば、「支持する人53.2%、指示しない人22.6%」というデータは「50%以上の人が支持、20%の人が不支持」というシンプルな数字にしても、判断材料としては十分である。データの細かい部分にエネルギーを費やしても、かえってデータを複雑でわかりにくいものにする。」と述べている（内田，2001）。また、われわれの行った調査でも、50名のビジネスパーソンの回答（A28社は統計分析利用企業、B22社は未利用企業）は、その傾向を支持していた。すなわち、

「統計分析情報に要求される精度は概略的でよい」という意見に、まったく反対である(1)⇔まったくその通りである(7)の尺度で賛否を問うたところ、5.13（標準偏差1.13）となった。（東淵，1992）

このことから、入手した経営データの品質が仮によくなくても、その品質の程度を踏まえたうえで解析を行い、情報をできる限り抽出できれば、データ解析は経営意思決定に役立ったと言えるだろう。データ収集と並んでここでも執拗さが不可欠であるということになる。

(2) データ解析手法の特性との観点からの評価

データ解析手法の多くは統計学に基づいている。推測統計学の分野では、適用前提条件が満たされないままデータ解析を実施しても正しい結果は得られないし、場合によっては大いなる損害をもたらすことさえある。また、統計手法

には、データの誤差や品質の悪さに鋭敏に反応するものもあれば、ノンパラメトリック手法やテューキーの抵抗直線のように比較的鈍感なものもある。比較的品質の悪い経営データの場合、鋭敏に反応する手法を適用することなく、大雑把ではあるが、外れの少ない手法を適用することによって、内包する情報をそれなりに抽出しうると言える。

①データの検索・集計による解析

データベースに対して検索・集計することによって、例えば、「購買金額の多い顧客は？」[最近購入のない顧客は？]「過去3回以上のクレームが発生した顧客は？」などの情報が一瞬にして抽出できる。前者については、顧客ごとに購入金額を集計し、金額の多い順にソートすればよい。後の2者については検索をかければはじき出せる。また、自社の顧客の男女比や年齢構成、地域ごとの顧客分布なども、導き出すことができる。データベースを使ったこのような解析においては、経営データの品質の悪さが解析によって拡大されるということはない。もちろんデータのメンテナンスが不十分で、現実との乖離が大きいと使いものにならない。

②併売分析（マーケットバスケット分析）

マーケットバスケット分析は買い物カゴにどのような商品が入っているか、つまり同時購買の商品の組合せを解析するものである。もとデータは、顧客のレシートレベルでの購買データである。POSやデータウェアハウス等がかなり普及している現在、データの品質はかなり高いと考えられる。仮にやや低くても、同時購買数をカウントする処理なので、データ品質の悪影響はあまりないと考えられ、概略的な同時購買傾向は抽出できるであろう。

③グラフによる分析

グラフを描いて情報の抽出を図ることは極めて重要である。グラフ化は基本的にはデータを何らかの形で再表現することである。したがって、ここでもデータの品質の悪さが、グラフを描くことで拡大されるとは考えにくい。清水は散布図の具体的な使い方として売れ筋の商品を見つける方法を紹介している。

ある期間で平均した店頭在庫を x 軸にし、同期間の売上数を y 軸にして、散布図を作成する。売れ筋商品は売り切れで平均店頭在庫が少なくなるが、少ない店頭在庫の割には、高い売上げであることを示すことになるというものである。売上高と在庫量の誤差や品質の悪さは拡大されることはないので、ある程度の目安になる情報が得られるであろう。

④データの加減乗除加工による解析

データの加減乗除加工による解析とは、例えば、比率（例：人口密度）、構成比（例：産業構成比）や相対比（例：性比）、指数（例：価格指数）や変化率などのようにデータに加減乗除を施し、データを加工し見やすくして情報を抽出することである。乗除の結果、誤差が拡大する可能性があるものの、加工によって情報がわかり易く抽出されるメリットに比べれば、その悪影響は無視できる程度である。

⑤記述統計手法による解析

平均値や標準偏差、相関係数など記述統計手法は比較的データの誤差に鋭敏に反応する。例えば、「アンケートで 10 人に 1 日あたりの労働時間を聞いたところ、9 人は 9 時間と記入したが、1 人が 90（時間）と書き間違えて答えた」とすると、1 日の平均労働時間が、 $171 \div 10 = 17.1$ 時間となってしまう」。このように算術平均値はデータの品質に比較的敏感である。通常の場合も、1 個の外れ値があれば、ない場合に比べてかなりの違いが生じる。また、変数に誤差があれば、相関係数は希薄化される傾向がある。データの品質がよくない場合には、平均値の代わりに中央値を使ったり、順位相関係数を使ったりすることで、概略的に正しい情報抽出が可能になる。

⑥探索的データ解析（EDA）

EDA を用いることによって、データの品質の悪さを鋭敏に反応させずに、概略正しい情報を認識することが可能である。テューキーの抵抗直線は、伝統的な最小 2 乗法では、回帰直線が少数の外れ値によって大きな影響を受けるため、その弊害を極力抑え、概略正しい回帰直線、さらには因果関係についての

情報を求めることができる。

⑦推測統計手法による解析

統計手法の適用前提条件が満たされない場合に結果が大なり小なり歪んでしまうが、その影響が比較的少ない性質を通常「頑健性」と呼んでいる。推測統計学は他の手法に比べて頑健性に劣る場合が多い。パラメトリック推定や検定等を用いる場合には、データの品質に注意を払うことが必要である。解析結果の解釈においては、品質に劣るデータがある場合には、慎重な態度が必要となる。このような態度をもって臨めば、品質に劣る経営データの場合であっても、概略正しい情報を獲得することができる。

⑧多変量解析及び数量化理論による解析

多変量のデータを縮約する過程においてデータが持つ誤差や不正確さ等が影響し、解析結果に歪みをもたらす可能性は否めない。品質に劣るデータが一部でもある場合には、解析結果を評価する際に慎重な態度が必要である。このような態度をもって臨めば、品質に劣る経営データであっても、概略正しい情報を獲得することができる。

5. ま と め

本稿において、経営データの品質は必ずしも高くないが、ビジネス実践においてデータ解析情報に求められるものが「概略的に正しい情報」であり、決して高度な精度を求められているのではないこと、また、データ解析手法にはそもそもデータの品質に敏感でない手法も多く、仮に敏感であっても結果の解釈において控えめな態度で臨むことによって、多少品質に劣る経営データであっても有益な情報抽出が可能になることを述べた。

ビジネス分野における経営データの解析に当たって、分野は異なるが、ケインズの言葉はまことに示唆に富むように思われる。

I'd rather be vaguely right than precisely wrong.

John Maynard Keynes

参 考 文 献

- (1) English, L. P., Improving Data Warehouse and Business Information Quality, John Wiley & Sons, 1999 (NTT データ『高品質データウェアハウス戦略』2000年)
- (2) Whitehead, P. & G. Whitehead, Statistics for Business, Pitman, 1984
- (3) 荒川圭基『データベース・マーケティング実践ガイド』PHP 研究所 2002年
- (4) アンダーセン・コンサルティング『CRM 顧客はそこにいる』東洋経済新報社 1998年
- (5) 内田和成「シンプルに視覚化して多面的な考察を加えよ」The 21 2002年2月
- (6) 勝見明「半歩先を読むー鈴木敏文流統計学ー」PRESIDENT 2002.2.18 pp.54~59
- (7) 古寺雅美『統計学以前の統計入門』東京法令出版 1981年
- (8) 清水理『Excel でわかる統計入門』ナツメ社 2000年
- (9) 東瀨則之「ビジネスにおける統計手法の利用に関する調査ープリテストとしてー」松山大学論集第4巻第5号 1992年12月 pp.227~262
- (10) 東瀨則之「需要予測における需要量と販売量の乖離とその影響」松山大学論集第7巻第6号 1996年2月 pp.43~53
- (11) 東瀨則之「ビジネスにおける初期的データ解析の重要性」松山大学論集第10巻第6号 1999年2月 pp.69~83
- (12) 東瀨則之「ビジネスにおける統計分析の分類」松山大学論集第11巻第1号 1999年4月 pp.151~157
- (13) 鳥邊晋司, 東原英子『会計情報と経営分析』中央経済社 1996年
- (14) 林知己夫『行動計量学序説』朝倉書店 1993年
- (15) ジェームズ・ヘックマン「データ読解力養成講座 Don't Miss the Missing Data」ダイヤモンド・ハーバード・ビジネス・レビュー 21 2002年6月 pp.94~104
- (16) 藤森利美『分析技術者のための統計的方法』丸善 1988年

(本稿は松山大学特別研究助成(1999年度)の成果の一部である。)