

データサイエンス領域テキスト(ベータ版)

本テキストは、文部科学省の生涯学習振興事業委託費による委託事業として、学校法人河原学園河原電子ビジネス専門学校が実施した 2019 年度「専修学校による地域産業中核的人材養成事業」の成果物です。

第1章. 仕事と数値

あなたの仕事は何ですか？



美容師なら お客様の美容を整えること



イラストレーターなら イラストを描くこと



料理人なら 料理を作ること

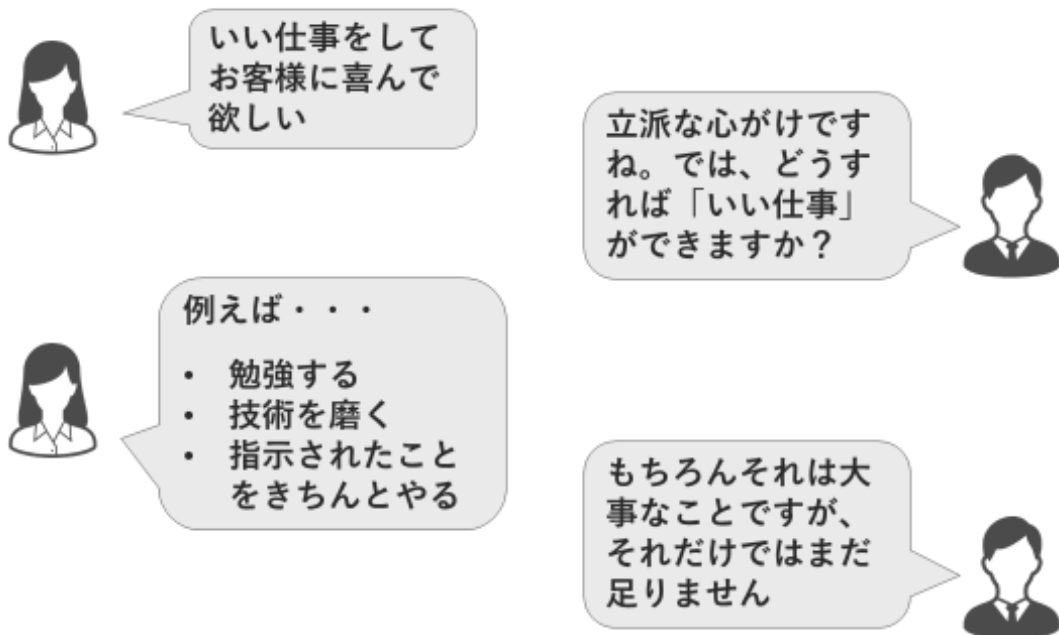
……それだけ、だと思っではいませんか？

あなたは今、美容/美術/料理/医療などさまざまな専門分野の知識・技術を学んでいて、将来その技術を活かした職業につく方が多いことでしょう。

- 美容師なら、お客様の美容を整えること。
- イラストレーターなら、イラストを描くこと
- 料理人なら、料理を作ること

それが将来自分のすべき仕事である、と考えている方が多いはずですが、もちろんそれらの仕事が一番大事な「柱」であることは間違いありません。しかし、それだけでは不十分です。「いい仕事」をしたいと思うのであれば、他にも考えなければならないことがあります。

「いい仕事」はどうすればできますか？

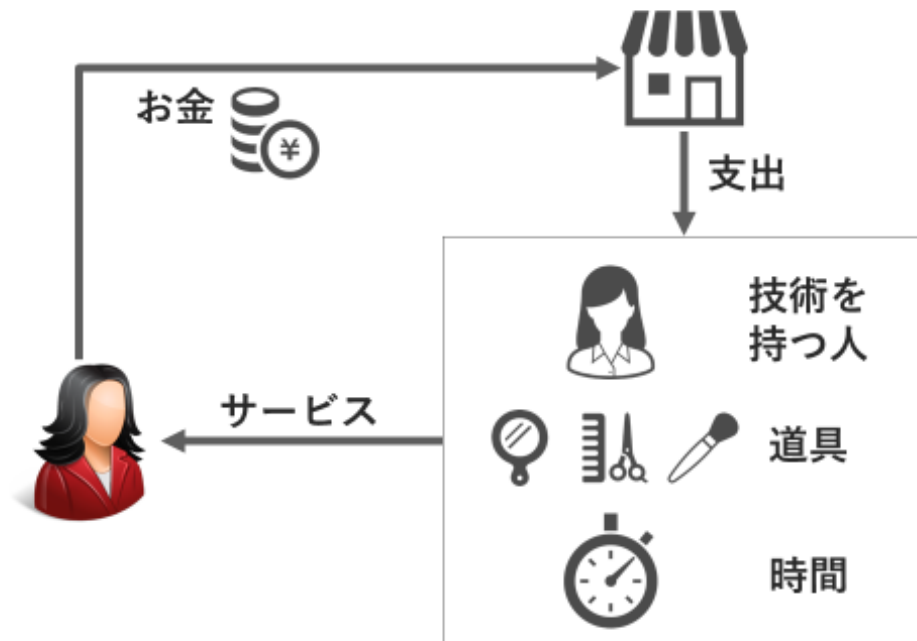


「いい仕事」をしてお客様に喜んでもらいたいというのは立派な心がけですし、そのために

- 勉強する
- 技術を磨く
- 先輩/上司に指示されたことをきちんとやる

のは大事なことです。しかしそれだけでは「いい仕事」はできません。特にあなたが5年10年と経験を積んで店長を任されたり、独立して自分の店を構えたりするようになると、「技術」だけでは足りないのです。

「いい仕事」をするには「お金」が必要

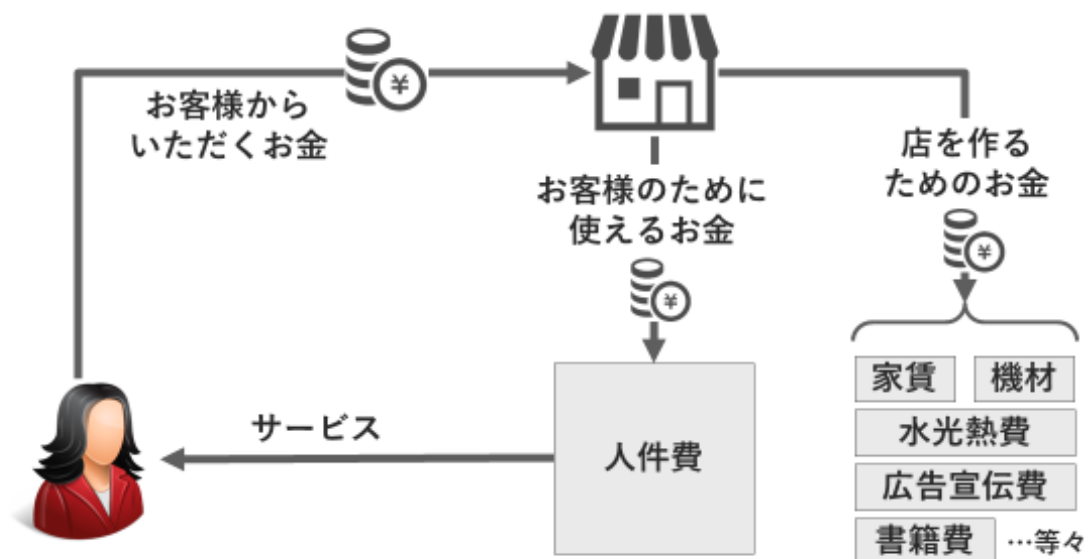


いい仕事をするためには「お金」が必要です。お金は当然、お客様からいただくものです。たとえば美容室の場合、技術を持つ人を雇い、道具を買いそろえ、時間を使ってお客様に美容のサービスをしています。人・道具・時間を確保するためには十分なお金が必要です。これは美容室に限らず、どんな仕事でも本質は同じです。

「お金がなくても精一杯やります」という心がけは必要ですが、結局のところ心がけではどうにもならない部分が多いのです。

(注: ここでいう「サービス」は仕事をするという意味であり、無料奉仕という意味ではありません)

お客様のためにお金を使うには？



いい仕事 = 「心を込めた手厚いサービス」をするためには、
「お客様のために使えるお金」を増やす必要があります。

どうすればそれができますか？

お客様からいただく「お金」は実は大まかに2種類の目的に使われます。

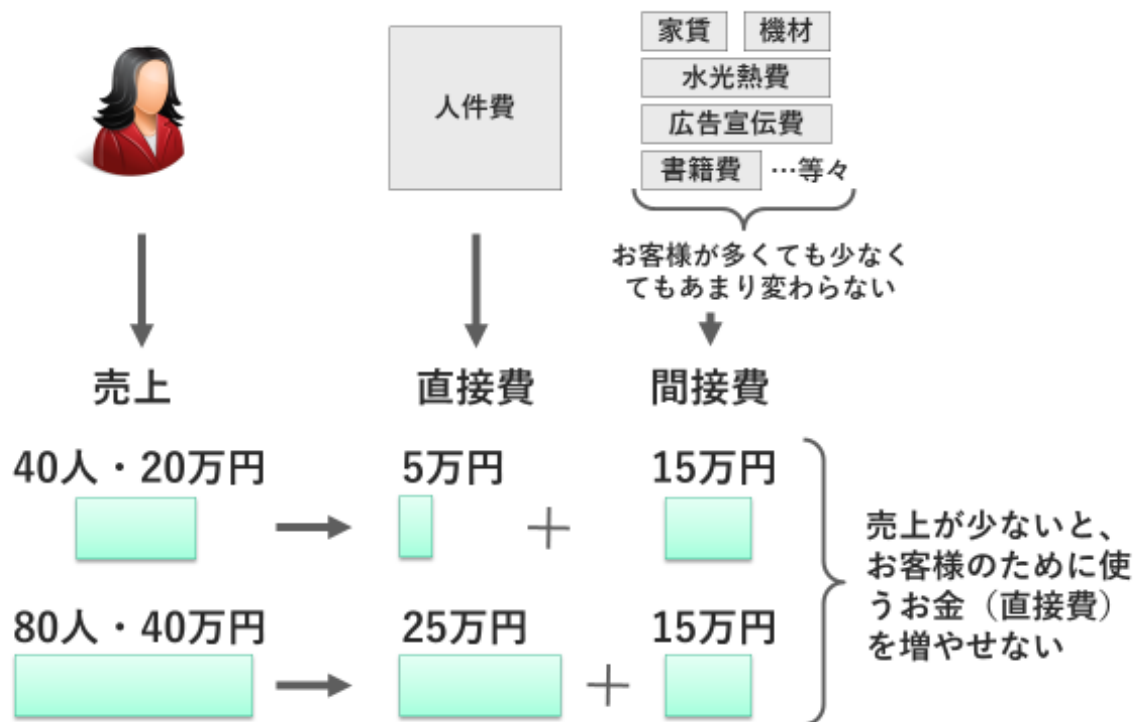
1つは家賃/機材/水光熱費/広告宣伝費等、「店」を成り立たせるために使われるお金です。

もう一つは、お客様のために使われるお金です。多くの場合、「人件費」がこれに当たります。製造業や飲食業の場合は「材料費」もここに含まれます。

「いい仕事をする」ためには「心を込めた手厚いサービスをする」必要があり、そのためには「お客様のために使えるお金」を増やす必要があります。

どうすればそれができるでしょうか？

売上を増やせば直接費を増やせる



「お客様のために使えるお金」と「店を作るためのお金」をここではそれぞれ直接費、間接費と短く呼んでおきます(注)。たとえば家賃はお客様が0人でも100人でも変わらないように、間接費はお客様が多くても少なくてもあまり変わらないため、総額で15万円固定でかかると仮定します。客単価を5,000円として、お客様が40人・売上20万円のばあいと、80人・40万円の場合でお客様一人当りに使える直接費を計算すると、

40人・20万円の場合： 直接費 5万円 ÷ 40 = 1人あたり 1,250円

80人・40万円の場合： 直接費 25万円 ÷ 80 = 1人あたり 3,125円

となり、一人当たりで2.5倍の差が出てしまいます。「お客様のために使えるお金＝直接費」が多ければその分丁寧な「いい仕事」ができます。

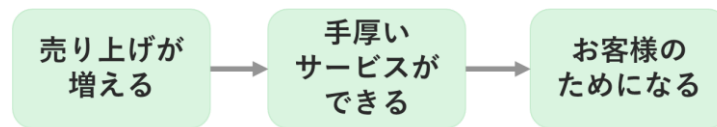
いい仕事をするには「技術を磨く」だけでなく、「売り上げを増やす」ことも非常に重要なのです

(注) 会計学上は、人件費がすべて直接費に分類されるわけではなく、水光熱費等がすべて間接費になるわけでもありませんが、大まかに言って人件費は直接費に、家賃や水光熱費等は間接費になるなる割合が大きいためこのような書き方をしています。

どうすれば売上を増やせるか？

売り上げを増やすのは「経営者の義務」です

では、どうすれば増やせるでしょうか？



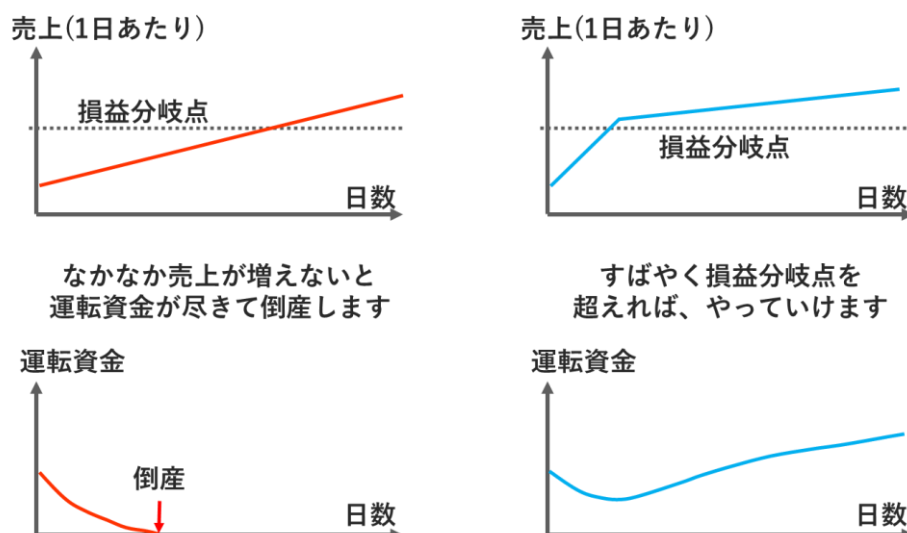
「心を込めて丁寧な仕事をする」のは大前提

しかし、それだけでは足りません

売り上げが増えれば、お客様の支払額が変わらなくても手厚いサービスができますので、お客様のためになります。したがって、売上を増やすのは「経営者の義務」と言えます。では、どうすれば増やせるでしょうか？

「心を込めた丁寧な仕事をして、お客様に満足してもらえば、また来てもらえて売上が増える」という面は確かにあります。しかしそれだけでは「お客様が増える」のに長い長い時間がかかります。集客ペースが遅く、損益分岐点を超えるのに時間がかかると運転資金が尽きて倒産してしまいます。

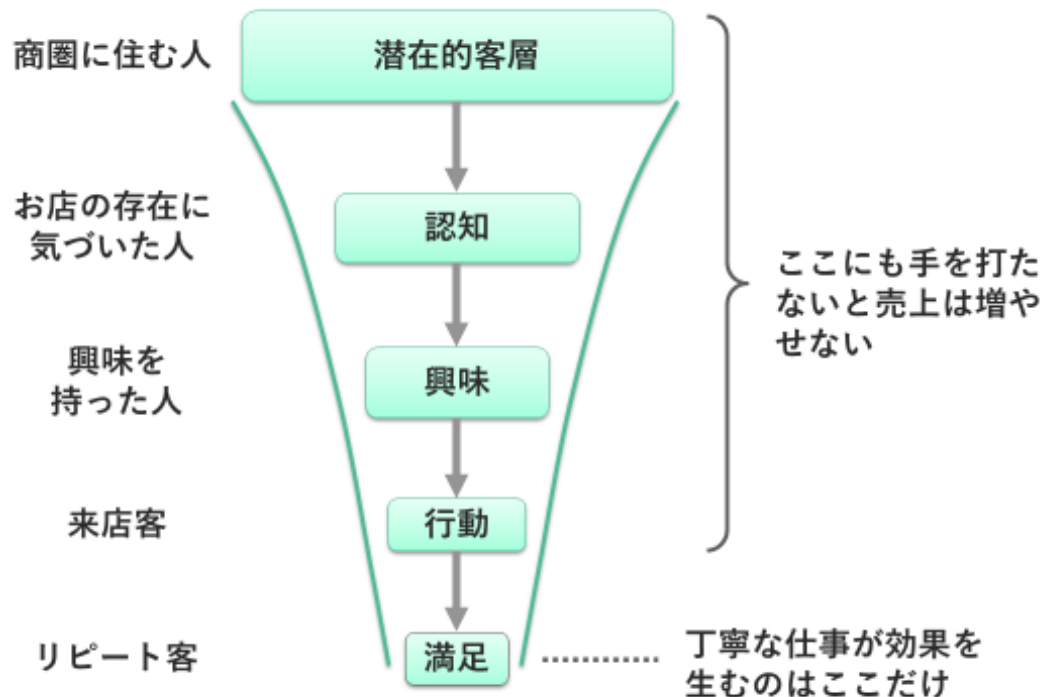
集客ペースが遅いと運転資金が尽きる



損益分岐点：売上がこの額以上になれば利益が出る、という金額のこと

運転資金：お店の定常的な支払に必要な資金。これがゼロになると倒産

丁寧な仕事が生むのは最後だけ



この図は「マーケティング・ファネル」と呼ばれるもので、「潜在顧客がリピート客になるまでのステップ」をモデル化したものです。

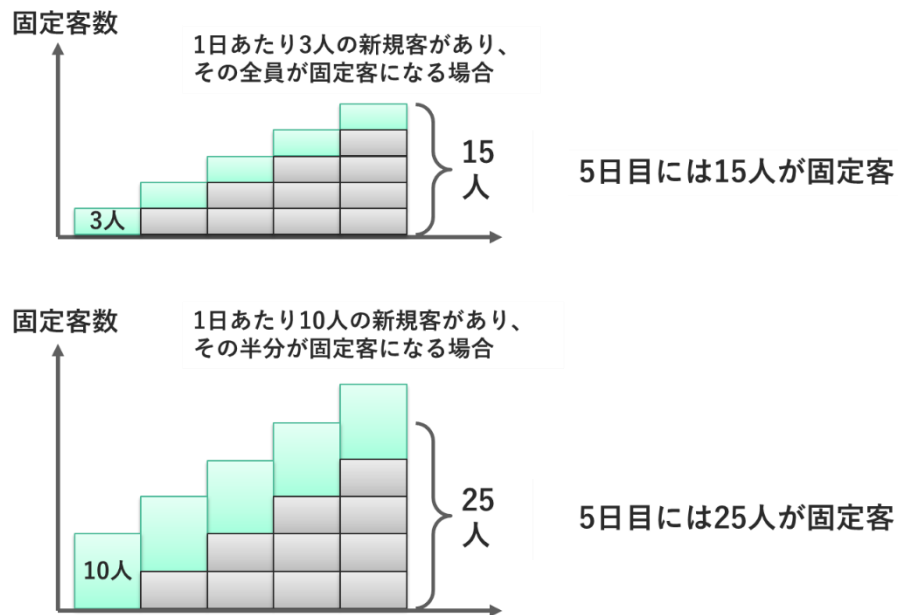
最上部の「潜在的客層」は商圏に住む人全員です。たとえば日常の買い物をするコンビニエンスストアの商圏はおおむね半径 500 メートル以内、スーパーならば 1～数 km、大規模ショッピングモールなら 10km 以上になります。美容室は日常的に来店しないためコンビニよりは広いですが、これといった特色の無い店の場合はコンビニと同程度です。ただし特色のある店ならば 10km 単位の遠方からの来店も期待できるため、店の性格によって大差があります。

しかし、いくら商圏に住んでいても、存在が知られていない店には客は来ません。通勤の路上に店がある、店の広告がある、チラシがポスティングされていた、ネットで検索したら見つかった、など、なんらかの理由で「気がついた人（認知）」のうち、「興味を持った人」が来店という「行動」を起こし、「満足」すると再来店（リピート）を期待できます。

「潜在的客層」から「認知」→「興味」→「行動」→「満足」までのそれぞれの過程でどんどん人数が減っていくため、漏斗（英語でファネル）の形に似ていることからマーケティング・ファネルと言います。新規客を獲得して固定客（継続的にリピートしてくれるお客様）へと転換していくためには、潜在的客層に認知させる、認知した者に興味を持たせる、さらに行動させ満足を得る、という各段階でどんな手を打つのかをきちんと考えていく必要があります。

このファネルの中で「心を込めた丁寧な仕事」が効果を生むのは最後の「満足」の段階だけです。1日の新規来店客が3人しかいなければ、どんなに「いい仕事」をしたとしても固定客は1日最大3人しか増えません。しかし、「認知」「興味」の段階で適切な手を打って1日に10人の新規客を獲得できれば、固定客になるのがそのうち半分しかなかったとしても1日に5人ずつ増えることになります。

新規客が増えないと固定客も増えない



「丁寧な仕事が効果を生むのは最後だけ」であり、「行動」までの段階はそれ以外の方法で集客を図らなければいけない、ということを肝に銘じてください。

事例：歯科医院の集客戦略

ビジネス街で開業したある歯科医院が
集客のために取った施策とは？

立地

ある県庁所在地、オフィスビル街

商圈特性

昼間人口は多いがほとんどがサラリーマンや公務員。ファミリーは少ない。

ニーズ

会社が終わってから歯の治療をしたい

施策

夜間診療を実施（夜10時まで）

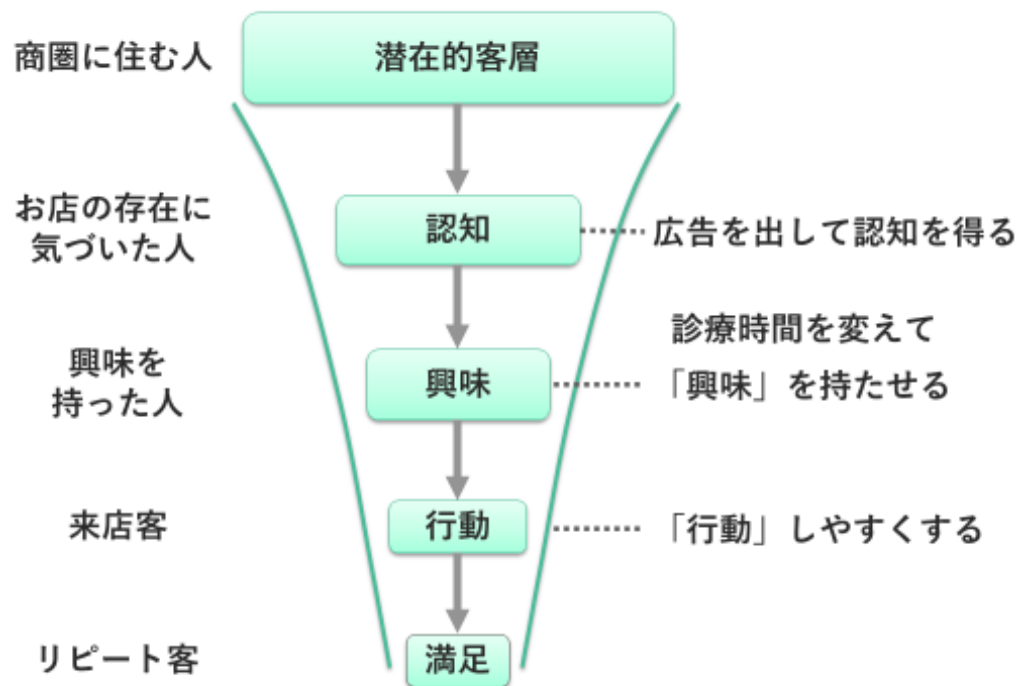


結果：夜間で昼間の2倍の患者を獲得

ここで、「認知」「興味」段階に力を入れて集客を果たした歯科医院の事例を紹介します。ある県庁所在地のオフィスビル街で開業したある歯科医院は、「昼間の人口は多いがほとんどサラリーマンや公務員であり、昼は忙しいので仕事が終わってから歯の治療をしたい」というニーズがあることに目を付けて、当時は珍しかった夜10時までの夜間診療を実施したところ、夜間で昼間の2倍の患者を獲得して一気に経営を軌道に乗せた例があります。

このように、「新規客の来店を得る」までの過程では「仕事の技術」以外の部分で集客施策を考えたほうがよいケースが少なくないのです。

何がどこで効果を生むかを考える



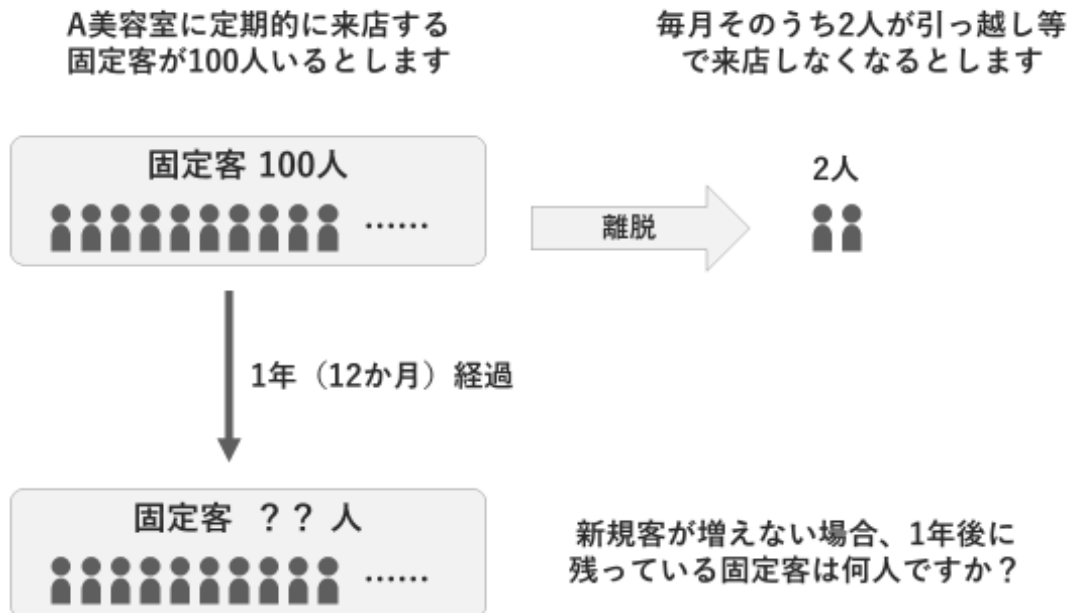
もう一度、マーケティング・ファネルに照らし合わせて考えると、夜間診療というのは「お、便利だな」という「興味」を持たせるための施策であるとともに、興味を持った人が「行動」しやすくなるという一石二鳥の効果があります。

しかし「興味」を持たせる材料があったとしてもその前に「認知」が必要です。医院そのものの前を通る人しか気づかないようでは「認知」する人数が少なすぎます。たとえば、「認知」を増やすためには「主要な通りに広告を出す」といった方法が可能です。その場合、「認知」と同時に「興味」を持たせるために、「夜間診療実施」を大々的にアピールした広告が適切でしょう。つまり下記AタイプのほうがBタイプよりも「興味」を持たせる効果が高いと思われます。広告を出す場合はこのように「強調するキーワード」の選択で効果が大きく違ってきます。



なお、歯科医院向きの方法ではありませんが、もし飲食店のような業種であれば、行動を促すために期間限定のスペシャルメニューを作る、クーポンを配布するといった方法もよく使われます。

数字で考えてみましょう



マーケティング・ファネルを踏まえて集客施策を考えるためには、「数字で考える」習慣が欠かせません。

練習問題として1つ考えてみましょう。

あるとき、A美容室に定期的に来店する固定客が100人いたとします。しかし毎月そのうち2人が遠方への引っ越し等で来店しなくなるとします。新規客が増えない場合、1年（12ヶ月）経過後に残っている固定客は何人でしょうか？

解答：月に2人ずつ×12ヶ月間減るわけですから、答えは当然

$$100 - 2 \times 12 = 100 - 24 = 76 \text{ 人}$$

となります。どんな業界でもこのように固定客は必ず減っていくものですので、常に新規客を獲得していく努力をしなければなりません。それには、

当店の固定客は現在、何人いる

通常のペースだと月に何人ぐらい減っていく

といった数字を把握する意識・習慣が必要です。

数字を把握すれば異変に気づける

A美容室では毎月末にその月の来店客を新規と既存に分けて人数を把握しています。なお、同じ人が月に2度以上来店した場合も1人と数えています。

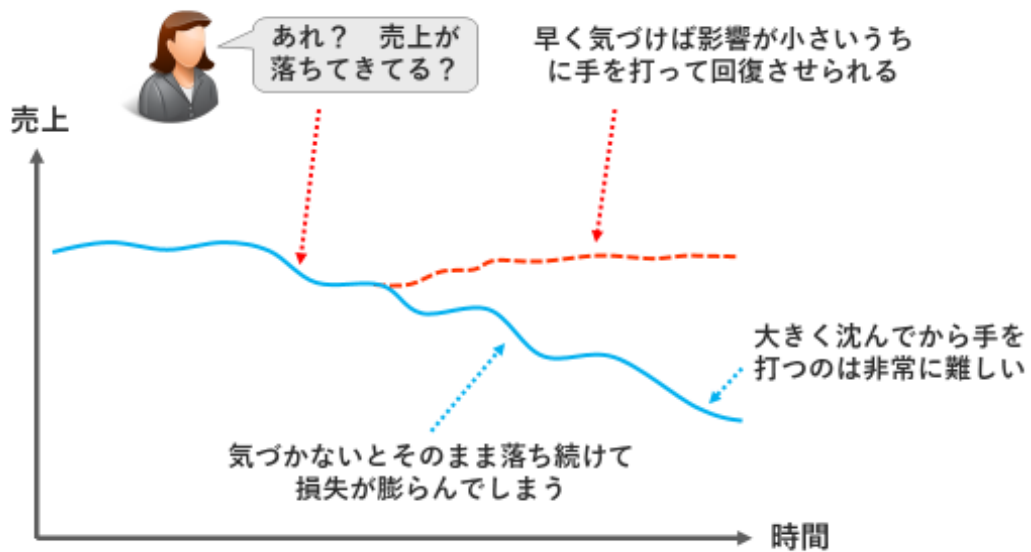
	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
既存	88	80	99	95	89	87	80	82	81	85	82	80
新規	2	3	2	10	3	20	3	2	3	4	2	2

1. この数字から気になるところを探してください。「気になる」というのは、「あれ? どうしてこうなったの?」と思うような、多すぎるか少なすぎる数字です
2. そのような異変が起きた理由としてはどんなものがありますか? ありうる理由を思いつく限り挙げてください

数字を把握するトレーニングとして今度はこの問題を考えてみましょう。

A美容室では毎月末にその月の来店客を新規と既存に分けて人数を把握しています。なお、同じ人が月に2度以上来店した場合も1人と数えています。上図中の指示に従って、1. 気になるところを探し、2. その異変が起きた理由を思いつく限り挙げてみてください。

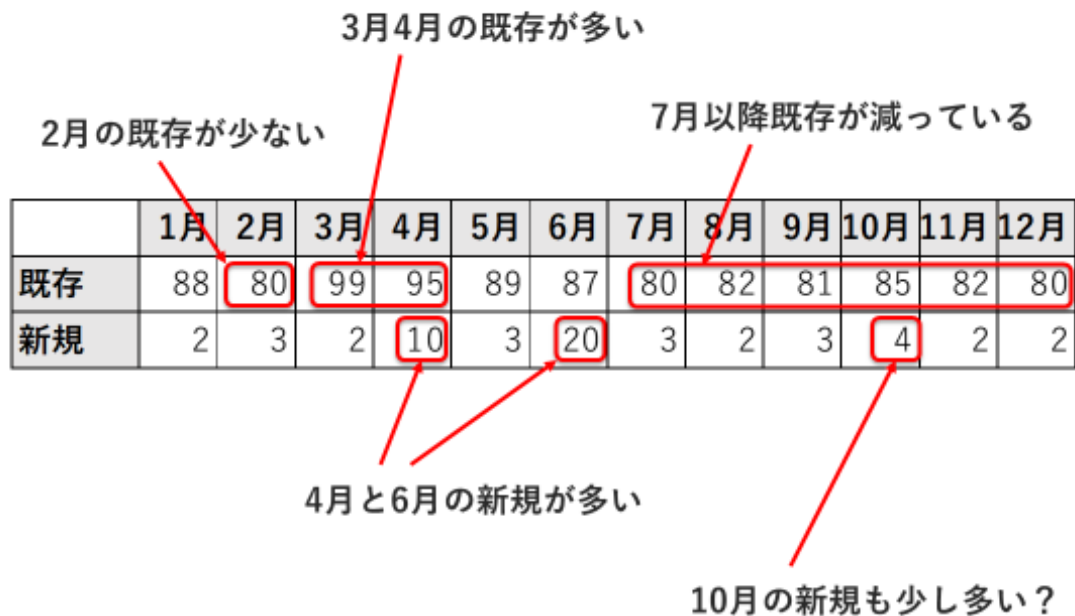
「問題」に早く気づかなければならない



では、どうすれば「問題に早く気づく」ことができるのか？

ふだんから数字に注意していると、「売上が落ちてきた」といった「問題」にいち早く気づいて手を打つことができます。早く気づけば影響が小さいうちに手を打って回復させられますが、気づかないとそのまま落ち続けて損失が膨らんでしまいます。大きく沈んでから手を打つのは非常に難しいものです。

気になる数字



答え合わせです。「気になる数字」としては、上図のような各点が挙げられます。皆様が気づいた部分が他にあれば書き足しておいてください。

では、このような数字になった理由は何なのでしょう？

「2月の既存が少ない」理由の候補

- もともと2月は1月に比べて短いので、客も少なかった
- 新年や成人式のため1月の来店客が多かった

「3月4月の既存が多い」理由の候補

- 3月4月は卒業式・入学式や人事異動といった儀式が多いから

「4月の新規が多い」理由の候補

- 進学や転勤で引っ越してきた人が多いから

などなど、このようにちょっとした数字の変化に注目して「なぜだろう？」と疑問を持ち、「理由はこれではないか？」という「仮説」を考える習慣が極めて重要です。

たとえば2月と7月の既存客は同じ数になっていますが、7月は2月よりも3日長いのに同数という

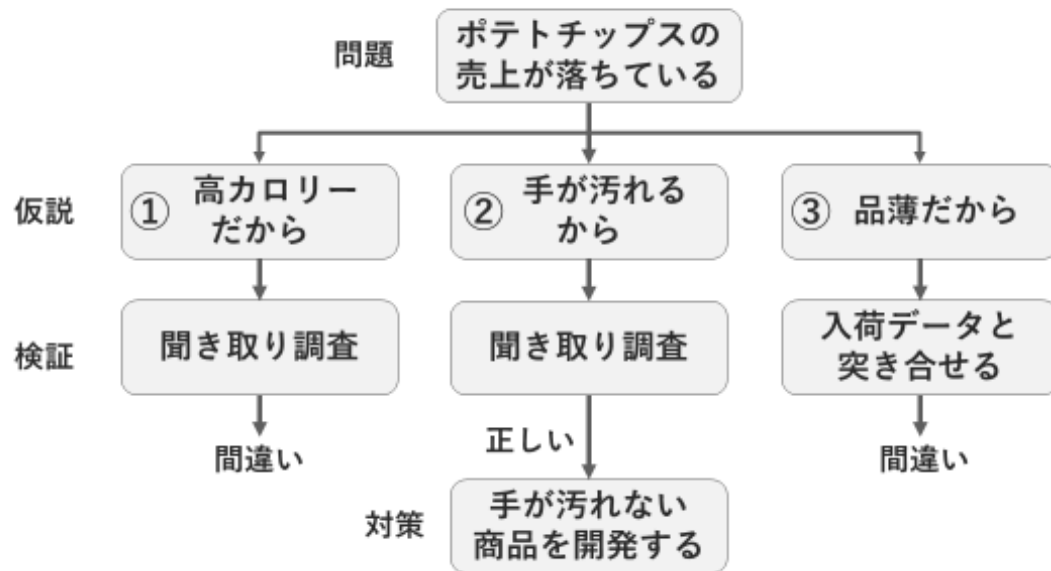
のは不自然です。1日短い6月に比べて7人減っているのも不自然です。これがたとえば「学生が夏休みに入って帰省していなくなるので減ったのだろう」といった理由であれば問題はありませんが、何か店に不満があって離れていった場合は何か手を打たなければならないでしょう。

6月の新規客が不自然に多いのも気になるポイントです。飲食店などでよくあるケースですが、TVなどに取り上げられて一時的に客が増えるとそれまでの常連客が来づらくなり、その後かえって売上が落ち込むことがあります。たとえば6月は近隣の別な美容室が一時休業していてその客が流れてきたために新規が多く、しかしその結果既存客が予約を取りにくくなってさらに別な店に流れたために7月以降は落ち込んだ、といった可能性はないでしょうか？

もちろん、数字だけでは真相はわかりません。数字はあくまでも考えるきっかけです。そのきっかけを手がかりに真相を探り、問題があれば手を打つ必要があります。

「仮説」を考える習慣を持とう

- ・ 何か困ったこと（問題）が起きたら、まず原因の「仮説」を考えます。
- ・ 仮説の大半は間違いですので、「検証」を行います。
- ・ 検証の結果、正しかった仮説を元に「対策」を考えます。



「仮説」を考える習慣は非常に大事です。何か困ったこと（問題）が起きたら、まず原因の「仮説」を考えるようにしてください。

「仮説」は間違っていておかまいません。というよりたいていの場合間違っているので、「検証」を行います。上図ではある食品スーパーでポテトチップスの売上が落ちているという問題から「仮説」を立て、「検証」して「対策」を考えたと例です。

①高カロリーだから敬遠されている、②手が汚れるから敬遠されている、③そもそも品薄で売り切れただけ、という3つの仮説を立てた場合、①と②についてはお客様への聞き取り調査をすれば、③については入荷数量のデータと突き合わせれば検証できます。検証の結果②が正しいと分かれば、手を汚さずに食べられる種類の商品を開発してそれをアピールすれば売上は回復する可能性があります。

実際、近年スマートフォンの普及にともなって手が汚れるポテトチップスが敬遠されているという事実があるため、一部のメーカーは商品のパッケージを工夫して「手を汚さず食べられるチップス」を開発しています。商品そのものではなく、パッケージ等を変えることで売上が変わることはよくあるのです。

なぜ仮説が必要なのか？

上司から部下に対して



売上が落ちてるぞ！
お前らたるんどる！
しっかり売れ！

怒って部下を叱り飛ばせば
売れる時代ではありません

お客様に対して (1)



ポテトチップスを買わ
ないのはなぜですか？

え？ 買って
ますけど？



ぼんやりした質問では原因は分かりません。
お客様が自覚しているとは限らないからです。

お客様に対して (2)



ポテトチップスを食べると
き、手が汚れるのが気にな
りますか？

そういえばスマホを見な
がら食べるときは気にな
りますね



ポイントを特定して聞くと分かることがあります。
このために「仮説」が必要です。

「仮説」を立てないとどんなことが起きるのでしょうか？

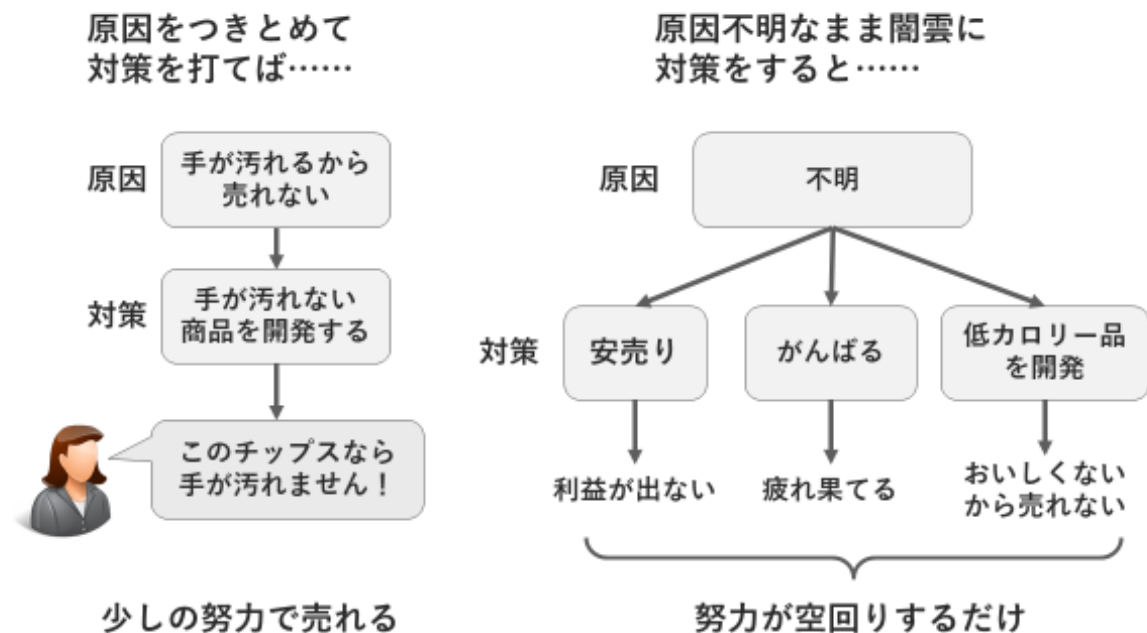
「売上が落ちている」という問題が起きたとき、1 番目の上司のように「お前らしっかり売れ！」と怒るだけというのはダメなパターンです。原因をさぐらずにただ「頑張る」だけで売れる時代ではありません。

かといって次の「お客様に対して (1)」のようにぼんやりした質問をしても原因はわかりません。人は自分の行動の理由を自覚していない場合も多いのです。

「お客様に対して (2)」のようにポイントを特定して聞くと、「そういえば……」とお客様自身もそこで気がつくことがあります。しかしこのような質問をするためにはその前に「仮説」が必要なのです。

なお、「なぜですか？」のようにキーワードを特定しない質問をオープン・クエスチョンといい、「手が汚れるのが気になりますか？」のように特定して聞く質問をクローズド・クエスチョンと言います。クローズド・クエスチョンの長所は相手が自覚していない答えがわかることで、短所はこちらが仮説を持っていない答えは得られず、誘導尋問になりかねないことです。オープン・クエスチョンはその逆の特徴があります。どちらも一長一短があるので、問題の原因を探るために聞き取り調査を行う場合は、オープンとクローズドの質問を適材適所で使い分ける必要があります。

原因不明だと努力が空回りしてしまう



何か問題が起きたときには、必ず「原因を突き止めて手を打つ」ことを心がけてください。「手が汚れるから売れない」という原因がわかれば、その一点の対策をするだけで、少しの努力で売れるようになります。これに対して原因不明なままで対策をすると

- 安売り → 売れても利益が出ない
- がんばる → 疲れ果てるだけで売れない
- 低カロリー品を開発する → おいしくないから売れない

のように努力が空回りしてしまいます。

「手に職」系の仕事をする人は、「自分は真面目に丁寧に仕事をするだけ。いい仕事をしていればお客様は評価してくれる」のように考えがちですが、いくら「真面目に丁寧に」仕事をしててもそもそもお客様の喜ぶ方向でなければ「いい仕事」にはならないので注意してください。

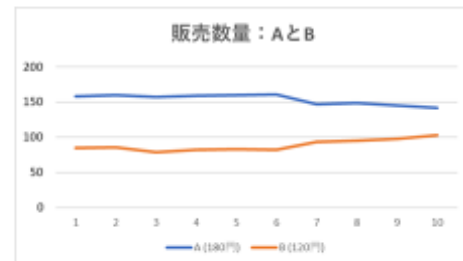
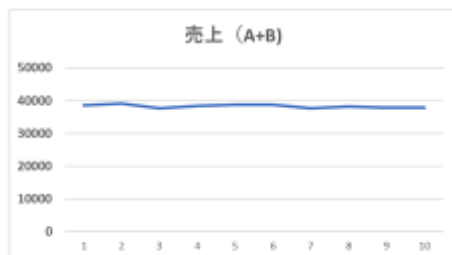
「数値」を把握していれば早く気づく



AもBも順調に売れてます。
特に変わったことはありません

というAとB（いずれもポテトチップス）の販売個数と売上総額を見てみると？

	1週	2週	3週	4週	5週	6週	7週	8週	9週	10週
A (180円)	158	160	157	159	160	161	147	149	145	142
B (120円)	85	86	79	82	83	82	93	95	98	103
売上 (A+B)	38640	39120	37740	38460	38760	38820	37620	38220	37860	37920



「問題」が起きたら原因の仮説を考えて検証し対策をするわけですが、そのためには「問題」に早く気づかなければなりません。そこで役に立つのが「数値」です。それも、いろいろな角度で数値を見る必要があります。

あるスーパーマーケットでポテトチップスをAとBの2種類だけ販売していたとします(実際には何十種類もあるのが普通ですが、話を単純化するため2種類にします)。担当者は「AもBも順調に売れています」と言っていますが、本当でしょうか。

「感覚」だけではわからない



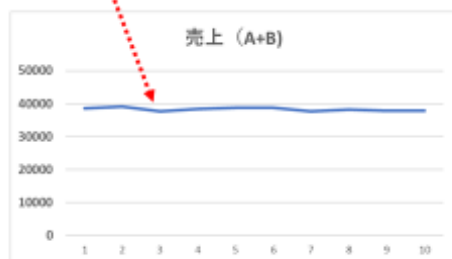
AもBも順調に売れてます。
特に変わったことはありません

人の「感覚」だけでは
わからない

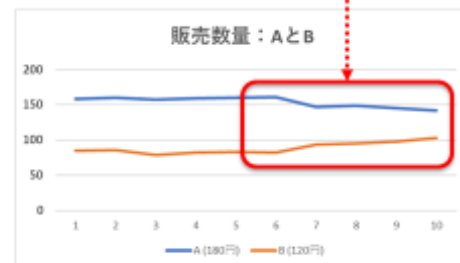
個別に数字を見ると気がつく

	1週	2週	3週	4週	5週	6週	7週	8週	9週	10週
A (180円)	158	160	157	159	160	161	147	149	145	142
B (120円)	85	86	79	82	83	82	93	95	98	103
売上 (A+B)	38640	39120	37740	38460	38760	38820	37620	38220	37860	37920

合計を見ていてもわからない



合わせて見ると意味がわかる



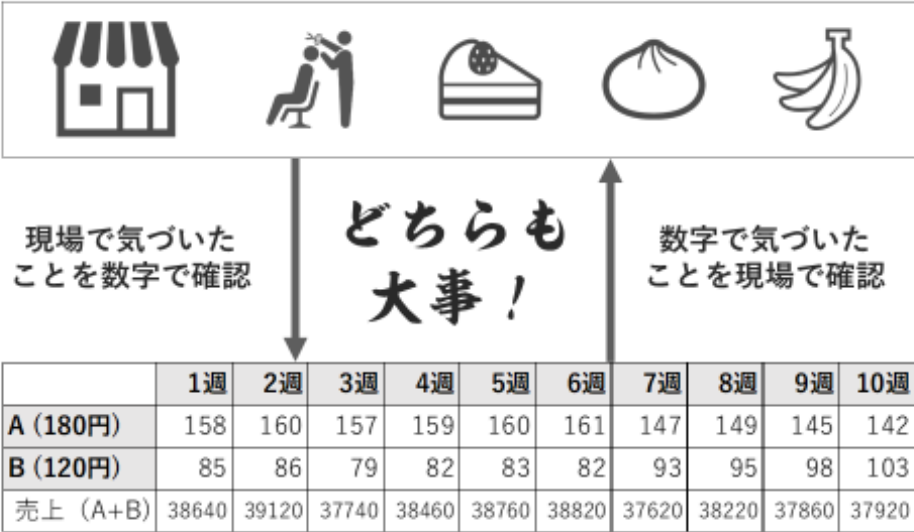
確かにAとBの合計額では特に目立った変化はありません。しかし、個別の数字では動きがあります。それも、AだけあるいはBだけだとわずかな変化ですが、2つを合わせて見ると「逆方向に動いている・・・ひょっとして低価格品への乗り替えが起きているのか？」という仮説が立てられそうです。

もし、「低価格品への乗り換えが起きつつある」のが正しければ、もう少し別の価格帯、たとえば150円や100円の商品も仕入れてみるべきかもしれません。逆に、一般的に高価格帯のほうが利幅は大きいので、売上が落ちているA商品のほうのテコ入れを図るべきかもしれません。

こうした細かい変化は数字を見ていかないと気がつきません。人間の感覚は意外に当てにならないのです。

問題を見つけるためには数字を見よ！

店頭でお客様や商品をよく見て接客・販売することも大事ですが
数字を見たほうが気がつきやすい問題もあります

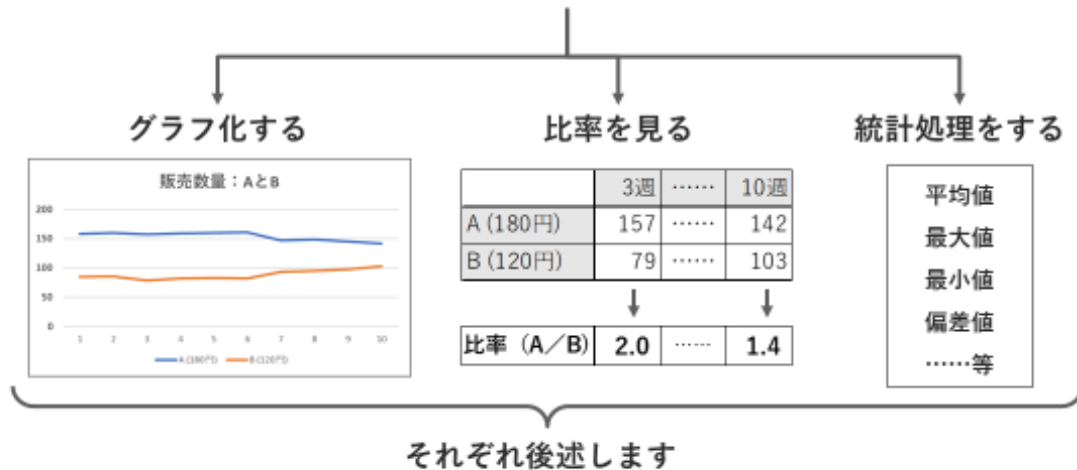


もちろん、数字だけではわからないこともあります。お客様の表情や会話は数字の上には現れませんが、それもマーケティングを考える上では非常に重要な情報です。店頭の現場で気づいたことを数字で確認すること、逆に数字で気づいたことを現場で確認することのどちらも大事です。

とはいえ、数字を見るのは面倒だ

	1週	2週	3週	4週	5週	6週	7週	8週	9週	10週
A (180円)	158	160	157	159	160	161	147	149	145	142
B (120円)	85	86	79	82	83	82	93	95	98	103
売上 (A+B)	38640	39120	37740	38460	38760	38820	37620	38220	37860	37920

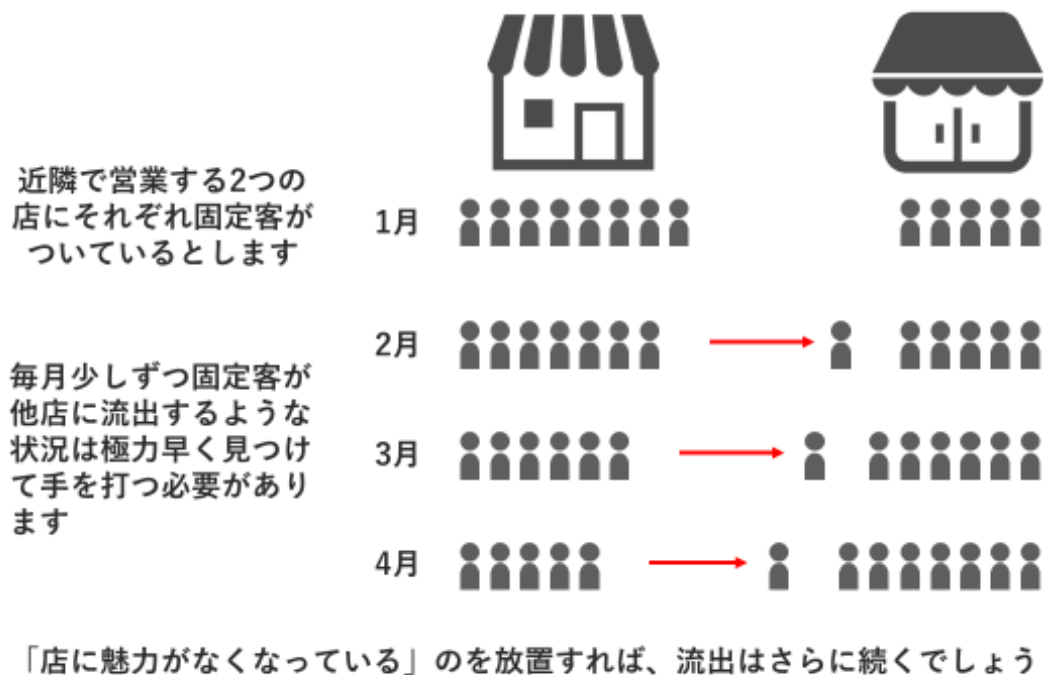
もっと分かりやすくする方法はないか？



とはいえ、数字を見るのは面倒なものです。数字だけが羅列された表をぼんやり眺めていても何もわかりませんし、集中して一つ一つ見ようとしてもすぐに目が疲れてきます。何かもっとわかりやすくする方法はないのでしょうか？

そこでよく使われるのが、グラフ化する、比率を見る、統計処理をする、などの方法です。それぞれ本講座の2章以降で扱います。

仕事と数値：固定客の減少



仕事をする上で数値に気を使わなければいけない場面をいくつか取り上げます。

まずは「固定客の減少」を察知することです。

美容室や飲食店は「いつも利用してくれる固定客」で経営が成り立つ場合が多く、固定客の維持に気を使わなければなりません。近隣で営業する2つの店があって、それぞれ固定客がついているとします。なんらかの理由で片方に魅力がなくなると、少しずつ他方に固定客が流出するため、できるだけ早く手を打たなければなりません。

たとえば「店に魅力がなくなった」の原因が「内装が古く、貧乏くさくなった」というものだとすると、「手を打つ」ためには「内装のリフォーム」が必要でかなりのお金がかかります。しかし固定客が減って売上も落ちているときは、お金のかかる投資にはなかなか踏み切れません。そうして対策を先延ばしにしたために、さらに売上が落ちてどうしようもなくなるケースも多いものです。売上があまり落ちていないうちに気づけば思い切った投資もしやすく、傷が浅いうちに回復させられます。

そこで、自店に固定客がどの程度いるのかを常に把握し、減少していないかどうかをチェックする必要があります。そのためによく使われる手法の1つに Recency（リーセンシー）分析があります。

仕事と数値：Recency分析

Recency =
最近、来ているかどうか？
を見る分析

①最終来店日を記録しておく

②経過日数でソート

会員ID	最終来店日	経過日数
1	3月30日	93
2	6月14日	17
3	5月14日	48
4	6月30日	1
5	5月25日	37
6	6月20日	11
7	3月24日	99
8	6月20日	11
9	4月29日	63
10	4月7日	85
11	5月20日	42
12	4月24日	68
13	6月2日	29
14	6月7日	24
15	6月5日	26
16	4月10日	82
17	5月14日	48
18	6月4日	27
19	5月25日	37
20	4月11日	81

会員ID	最終来店日	経過日数
4	6月30日	1
6	6月20日	11
8	6月20日	11
2	6月14日	17
14	6月7日	24
15	6月5日	26
18	6月4日	27
13	6月2日	29
5	5月25日	37
19	5月25日	37
11	5月20日	42
3	5月14日	48
17	5月14日	48
9	4月29日	63
12	4月24日	68
20	4月11日	81
16	4月10日	82
10	4月7日	85
1	3月30日	93
7	3月24日	99

③一定の日数以内の
来店者数をカウント
(期間集計)

30日以内 8
60日以内 13

Recency とは、「少し前の時間、最近」を表す英語で、Recency 分析は顧客が最近来ているか（購入しているか） どうかを見る分析です。

Recency 分析をするには、まず①顧客の最終来店日を記録しておき、来店後の経過日数がわかるようにします。さらに②経過日数が小さい順にソート（並べ替え）したり、③一定の日数以内の来店者数をカウントします。美容室であれば固定客は2ヶ月に一度程度は来ることが多いので、この数字が小さくなったときは固定客が減少していると考えられます。

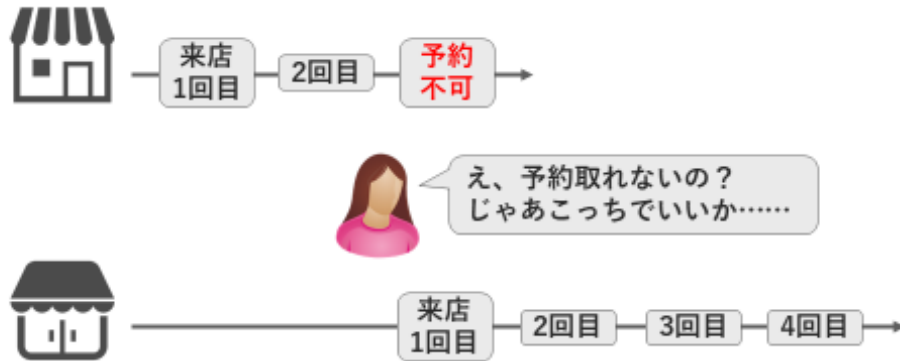
もちろん、集計期間は業種・業態によって違い、たとえば食品スーパーならばより短い期間で区切る必要があります。

小売業では Recency の他に、Frequency（フリークエンシー、来店頻度、頻繁に来ているかどうかを見る）、Monetary（マネタリー、購入金額の大きさを見る）分析を同時に行う場合もあり、総称して RFM 分析と言われています。

最終来店日を記録するのは手間がかかるため、ポイントカードを発行して自動的に記録されるシステムを作るのが一般的で、経過日数によるソートや期間集計には表計算ソフトを使ったり、専用の RFM 分析ツールを使います。

仕事と数値：多忙の反動

「予約が取れないから他店に行ってそのまま戻ってこない」パターン



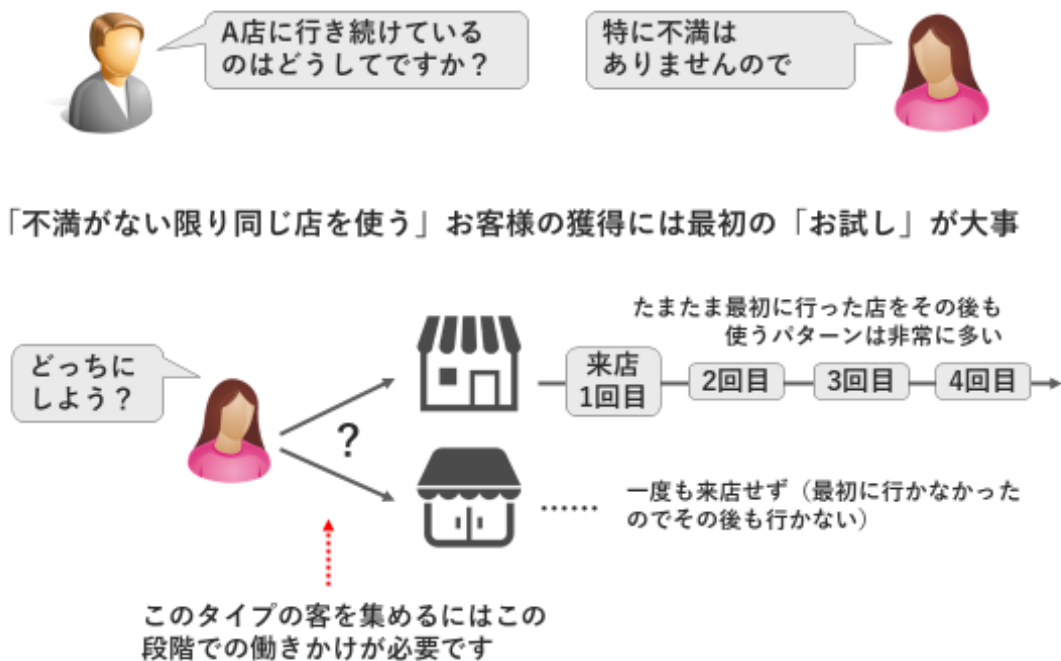
一時的に多忙な時期があった場合、既存客のその後の継続率に注意を払う必要があります

お店そのものに特にこだわりのない客の場合、何もないままそのまま固定客だったものが、何かのきっかけで他店を利用するとそのまま戻ってこない、というケースがしばしばあります。メディアに取り上げられるなど、一時的に多忙な時期（＝繁忙期）があるとこの種の「きっかけ」になりやすいため、その後の状況に注意が必要です。たとえば、

- 繁忙期に予約を受けられなかった客のリストを作っておき、繁忙期を過ぎたらそのリストに特典付お知らせメールを出す
- 常連客が事前に予約を入れやすいように、空き状況をネットで参照・予約可能なシステムを導入する

などの対策が考えられます。

仕事と数値：初回来店勧誘



「不満がない限り同じ店を使い続ける」お客様は少なくありません。このタイプの客を集めるには初回来店の獲得が大事で、それを逃がして他店に行かれてからひっくり返すのは容易ではありません。

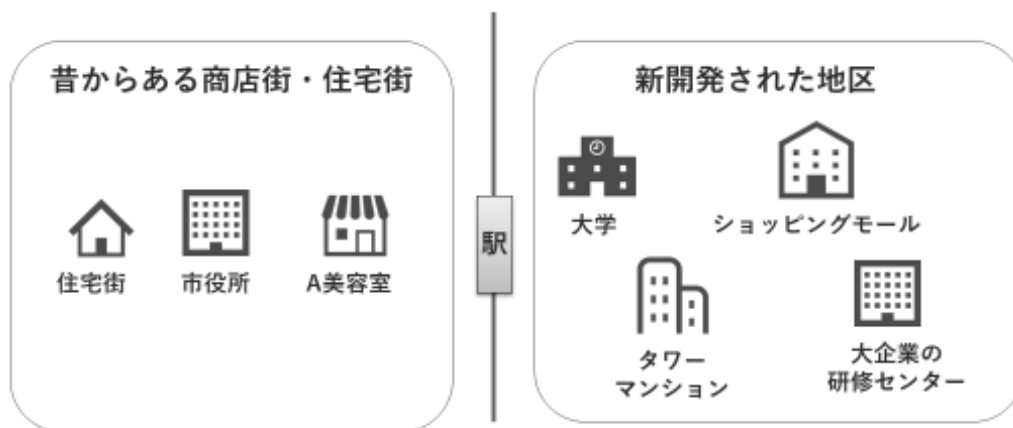
卒業・進学・就職・人事異動等で引っ越しが増える 3 月～4 月は特に新規客の獲得チャンスであり、次いで 9～10 月も転勤による引っ越しが多いため、この時期に街頭やポスティング、新聞折り込み等の手段でチラシ配布などの広告をするのは効果があります。

ただしこれもベースの「数字」をある程度見込んで考えなければなりません。たとえば近隣に大学の多い地域であれば 4 月に学生的大幅な異動がありますが、学生はほとんど新聞を取らないため新聞折り込みチラシでは効果が望めないのに対して、ワンルームマンションやアパートは学生の比率が高いと考えられるため、ポスティングは効果が期待できます。その場合はチラシの文面自体も学生向けに構成すべきです。一方、地方都市の旧市街地や郡部など、年齢構成が高く人口移動の少ない地域では別な方法を取らなければなりません。

居住者あるいは通勤・通学者の属性は地域によって大差があります。営業地域の特徴を、正確でなくとも構わないので、ある程度数字で把握する努力をしておきましょう。

仕事と数値：エリアマーケティング

A美容室は昔からある商店街・住宅街で営業しています。最近、駅を挟んで反対側に大学やマンションが増えて新しい人の流れができました。



チラシを撒いて集客するとして、両地区に同じチラシで良いでしょうか？

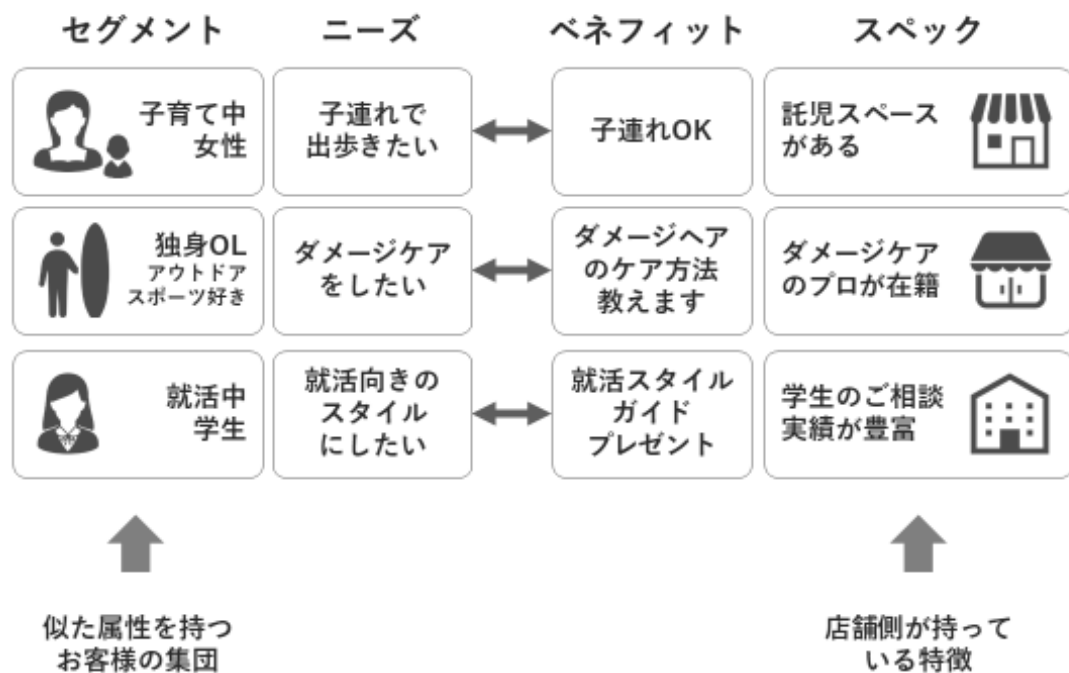
地域の特徴に合わせて営業方針を考えることを一般に「エリアマーケティング」と言います。線路や川、幹線道路などを1本またぐとまったく違う性格の地域になるケースも少なくありません。

上図のA美容室は昔からある商店街・住宅街で営業していますが、最近駅を挟んで反対側の地区が新開発されて、大学やショッピングモール、タワーマンション、研修センターなどが増えてきました。A美容室は新開発地区からも客を集められそうな立地ですが、その場合、両地区に同じチラシを使うべきでしょうか？

この場合、たとえば次のような「仮説」を立ててチラシを作り、実際に配布して反応を見て修正していくと良いでしょう。

仮説	方針
タワーマンションには30～40代の高所得者夫婦が多い	上品な装いを中心にデザインする
大学の新入生が入学後に初めて美容室に来る時期は4月下旬に集中する	4月中旬からGW期間に学生向けのチラシを配布する
ショッピングモールの客層には小さい子連れの母親が多い	子どもが遊べるスペースを用意しておく

ニーズとベネフィットのマッチング



数値を考えながら仕事をするために役立つ考え方に、「ニーズとベネフィットのマッチング」というものがあります。これは細かく言うと上図のように「セグメント、ニーズ、ベネフィット、スペック」の4つの視点で考えます。

「セグメント」というのは似た属性を持つお客様の集団で、同じセグメントの客は似たニーズを持つ、と考えます。たとえば小さな子供がいる子育て中の女性は、「子連れで出歩きたい」というニーズがある場合が多いですが、独身OLや学生には当然そのニーズがありません。一方、「スペック」というのは店舗側が持っている特徴のことで、それを「ニーズ」にマッチする形で表した言葉が「ベネフィット」です。

一般に、広告等で集客を図る場合は、1つの店が強くアピールできるベネフィットは1つだけです。したがって、自店が強みを持てるベネフィットは何なのかをよく考えて、それがどんなセグメントの客を惹きつけるのかを考えて集客手段を考えなければなりません。そうしてたとえば「当店のスペックなら、ダメージヘアを気にする女性の興味を引けるのではないかと仮説を立てると、そこで初めて「そういう客層が商圈内のどこにどれくらい住んでいるか」といった「数値」を考えたり調べたりできるようになります。つまり、セグメントからスペックまでのつながりを考えないと、「そもそも何の数値を調べたら良いのか」がわからないのです。

身近な業界で考えてみよう

飲食店（例：立ち食い蕎麦店とハンバーガーショップ、焼き肉屋とピザレストラン）、美容室（例：自分がよく行く店と行きにくい店）、英会話教室、アパレルショップなど、身近な業界で具体的なお店をイメージし、それぞれがどんなスペック（強み）を持っていてどのようなセグメントを対象に集客しているのかを考えてみましょう

セグメント	ニーズ	ベネフィット	スペック
		↔	
		↔	

練習のために、自分が利用したりする身近な業界でいくつかのお店を例にとってセグメントからスペックまでどのようになっているかを考えてみましょう。

たとえば同じ喫茶店チェーン業界でもドトールとスターバックスではハッキリと客層が違います。英会話教室には英語を使って仕事を有利にしようというチャレンジングな雰囲気を持つところもあればコンプレックス解消をテーマに集客するところもあります。ファーストフードであれば立ち食い蕎麦店と牛丼やハンバーガーではやはり客層が違います。美容室やラーメン店など、個人営業が中心の業界は特に個々の店による差が出やすいものです。

外装、内装、価格設定やメニュー構成などの具体的な店づくりのすべてをこの「セグメント～スペック」をもとに組み立てるものなので、自分が知っている店を例に考えてみると良いでしょう。

同じ業界でも行きやすい店と行きにくい店がある場合は、何がその違いを生んでいるのかを考えてみるとよいでしょう。

数値の扱いに慣れておこう

A美容室の北と南に、どちらも人通りの多い道があります。
朝の出勤時間帯に集客のためチラシを配ってみました。

北 

北の通りでは、月曜日の朝の30分間で
25人が受け取ってくれました。



A美容室

南 

南の通りでは、火曜日の朝の50分間で
33人が受け取ってくれました。

同じ時間なら、どちらのほうが多くの人に受け取ってもらえそうですか？



どちらでもかまいません！
頑張ってたっくさん
配りますよ！

……ではなく、効率よく成果を上げましょう。
それはお客様のためにも必要なことです。

それには「数値」の扱いに慣れておかなければ
なりません

「数値を手掛かりに考える」ためには、あたりまえですが数値の扱いに慣れておく必要があります。
たとえば上図のような状況ではどちらのほうが良さそうでしょうか？ 一人の人間が2か所に同時にいることはできないので、「チラシ配り」という同じ努力をするなら、より効率の良い場所を選ぶべきです。

「どちらでもかまわないので頑張ります！」という考え方はよくありません。というのは、「チラシ配り」のような集客のための作業は「客を集めるための手段」であって、お客様に直接役に立つものではないからです。「集客作業」はできるだけ短時間で済むようにして、それで浮いた時間をお客様のために使うべきです。

そこで、北と南で「1分当たり」何人にチラシを配れたかを計算してみましょう。そうすると、

北：25人 ÷ 30分 = 0.83人/分

南：33人 ÷ 50分 = 0.66人/分

となるので、同じ時間を使うなら北の方が効率が良さそうです。こういった計算は電卓でしてもかまわないので、「面倒くさがらずに数値を考える」習慣は持っておいてください。

分数の比較



北は30分で25人、南は50分で33人、
どちらが効率が良い？

これは結局のところ、2つの
分数を比較する問題です

$$\frac{25}{30}$$

$$\frac{33}{50}$$

どっちが大きい？

「どっちが大きい？」の答え
を簡単に出すには、たすきが
け掛け算をします。簡単な数
字で試してみましょう。

$$\frac{2}{3}$$

$$\frac{4}{7}$$

どっちが大きい？

$$\frac{2}{3}$$

$$\frac{4}{7}$$

たすきがけ掛け算

$$\frac{2 \times 7}{3 \times 7}$$

$$\frac{4 \times 3}{7 \times 3}$$

分母が揃う（通分）

$$\frac{14}{21}$$

$$\frac{12}{21}$$

分子だけで大小の
判断ができる

「北は30分で25人、南は50分で33人、どちらが効率が良い？」のような判断をしなければいけない場合はよくあります。これは結局のところ2つの分数のどちらが大きいかを比較する問題です。電卓をつかってもかまいませんが、ざっくりと大小比較をするだけなら簡単な計算で済むので、できれば暗算で出せるようにしておきましょう。

たとえば 「 $2/3$ と $4/7$ 、どちらが大きい？」 という問題の場合、分母を同じ数字に揃えるため、それぞれの分子に他方の分母の数字を掛けて「 2×7 、 4×3 」を計算します（小学校で習う「通分」です）。すると分母が揃うので分子だけで大小の判断ができます。 33×30 と 25×50 なら、0は無視して

「 33×3 はだいたい100、 25×5 は100より多い、だから 25×5 のほうが大きい」

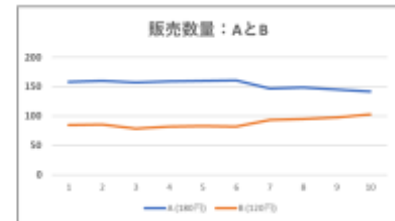
という程度の大小だけ判断できればいいので、正確な数字を出す必要はありません。暗算で大まかに大小を把握できることをめざしましょう。ですが、不安な時は遠慮なく電卓を使ってください。

数字の並びはグラフにする

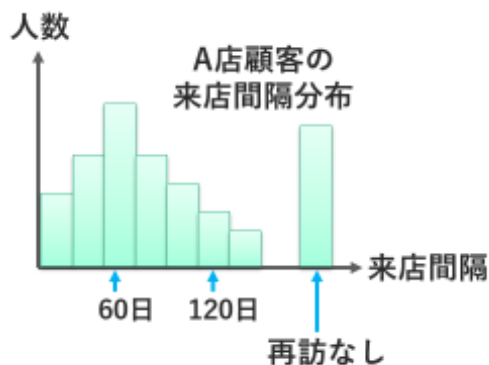
表 1

	1週	2週	3週	4週	5週	6週	7週	8週	9週	10週
A (180円)	158	160	157	159	160	161	147	149	145	142
B (120円)	85	86	79	82	83	82	93	95	98	103
売上 (A+B)	38640	39120	37740	38460	38760	38820	37620	38220	37860	37920

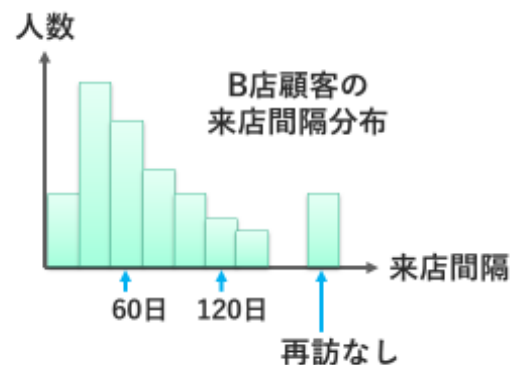
グラフ 1



グラフ 2



グラフ 3



たくさんの数値が並んでいるときに役に立つのがグラフです。表 1 は 2 つの商品の売上個数の推移、それをグラフにしたものがグラフ 1 です。数字ではわかりにくい、ちょっとした変化がグラフにするとわかるようになります。

グラフ 2 とグラフ 3 は別な店(美容室)の顧客を来店間隔別にグラフ化したものです。比べてみると、A 店は 60 日周期で来る客がもっとも多いのに対して B 店はそれより短い 40 日周期が多いこと、A 店は一度来たあと再訪しない客が B 店よりも多いことがわかります。

もし A 店の立地や客層が似ているのにこのような違いがあるのならば、A 店の接客や技術、プロモーションに何か問題があるのかもしれません。グラフを作ることによって数値を読まなくてもこうした違いが目に見えるようになります。現在は表計算ソフトウェアで簡単にグラフを作れるので、主要なグラフの作り方・使い方は知っておきましょう。本講座では後の章で扱います。

ものを「計る」尺度のいろいろ

尺度：ものを計って出てくる「数値」を分類する考え方の1つ

比例尺度：比率、間隔、順序に意味がある						
間隔尺度：間隔、順序に意味がある						
会員ID	性別	購入額	購入額順位	最終来店日	経過日数	平均来店間隔
2	男性	400	5	6月14日	17	10
4	女性	4000	1	6月28日	3	20
6	女性	3700	4	6月23日	8	30
8	女性	3800	3	6月20日	11	18
14	男性	3900	2	6月7日	24	35
順序尺度：順序に意味がある						
名義尺度：比率/順序/間隔のいずれも無意味						

数値を考えて仕事を進めようとする、と数値と言ってもさまざまな種類があることに気づきます。ここで「尺度」という考え方を知っておきましょう。

上図はある店の顧客に関するデータの一部ですが、このリストの中に「比例尺度、間隔尺度、順序尺度、名義尺度」という4種類の「尺度」が出てきます。

比例(比率)尺度

「購入額」についてはたとえば「会員ID=4番の女性の購入額4000円は、2番の女性の購入額400円の10倍である」のように、比率、倍率に意味があります。このような種類の数値が出てくる項目を「比例尺度を持つ項目」と言います。比率尺度と呼ぶ場合もあります。

間隔尺度

「最終来店日」という日付について、たとえば「6月1日は5月1日の何倍か？」のように比率を考えるのは無意味ですが、「6月1日は5月1日の何日後か？」のように間隔を考えるのは意味があります。このように「間隔に意味がある」種類の数値が出てくる項目を「間隔尺度を持つ項目」と言います。

順序尺度

「購入額順位」の項目はたとえば「順序3位と4位」の数字について「 $4/3$ 」のように比率を計算するのも「 $4-3$ 」のように間隔を計算するのも意味がなく、単に「4は3よりも後(下位)」という「順序(順位)」だけが意味を持ちます。実際、購入額を見ると1~4位はすべて3700~4000円という狭い範囲に収まっていて、5位だけが400円と極端に少ない数値です。このような場合、順位の数字で差をとっても購入額の差(間隔)とは何の関係もありません。このように順位(順位)だけが意味を持つ尺度を順序尺度と言います。

	購入額	順位	
差は300	4000	1	差は3
	3900	2	
差は3300	3800	3	差は1
	3700	4	
	400	5	

購入額の差(間隔)と順位数の差(間隔)は無関係

順序尺度はたとえば陸上競技で「上位入賞者から金・銀・銅メダルを与える」のような場面では重要です。たとえば購入額なら4000円と3900円で1位2位と順位をつける意味はほとんどありませんが、100メートル走なら9.89秒と9.90秒のようにほとんど同タイムでも順位をつけなければならず、それが金と銀という大きな名誉の差になります。

ビジネスの場面では、オンラインショッピングモール型通販サイトで複数の店舗から横断的に商品ができるような機能があると、「価格」ではなく「順位」が売れ行きに大きく影響する例があります。

オンラインショッピングモール型通販サイトでの商品検索結果例

「〇〇用レンズ」を販売している店舗が 3件見つかりました		
ショップ名	価格	最安値
A 電器	5,800	
B カメラ	5,900	
C 商事	6,500	

順位の影響は商品の種類によって違い、1円でも安いものが圧倒的に多く売れる場合もあれば、逆に最安値が敬遠される場合、中間が売れる場合もあります。

名義尺度

「性別」の項目は比率/順序/間隔のいずれも無意味です。このような項目を「名義尺度を持つ項目」と言います。性別、住所、職業などは多くの場合名義尺度になります。

尺度の判別が難しい/紛らわしい例

場合によっては尺度の判別が難しいときもあります。

【経過日数や平均来店間隔】

「6月20日」のような日付で書かれた「最終来店日」は間隔尺度ですが、それを「(最終来店後の)経過日数」として「2日、4日、10日」のような日数で表したときは、比率の計算が意味を持つてくる可能性があります。特に、たとえばそのデータを長期間取って「平均来店間隔」を産出した場合は比率に意味が出てきます。美容室や飲食店のような業界では1回来店あたりの客単価が大きく変わらないと考えられるため、平均来店間隔が2倍に伸びると売上が1/2に下がる可能性が高いためです。

【会員番号、ID】

基本的には名義尺度です。しかし、会員登録された順に1から増やしていく方法で番号をつけている場合、番号が「会員登録の時期」に連動するため、間隔や順序にある程度意味があるケースがあります。

【学校種別】

小学校・中学校・高校のような学校種別は通常は名義尺度ですが、在籍する子供の年齢が影響する情報については、順序が意味を持つ場合があります。

比例尺度についてはすべての計算が可能

4種類の尺度について、大小、差、比率のそれぞれを計算することに意味があるかどうかをまとめると下記ようになります。比例尺度についてはすべての計算が可能です。

	大小	差	比率	
名義尺度	×	×	×	性別、住所、電話番号など
順序尺度	○	×	×	金・銀・銅メダル、人気ランキングなど
間隔尺度	○	○	×	温度、日付など
比例尺度	○	○	○	金額、重量、速度、長さなど

第2章. 仕事と比較

比べてみると、違いが分かる

東京のZ町には有名な古書店街があり、専門的な古書を求める客が全国から集まります。下の図は古書店街マップで、Bは古書店、S,L,Fは大手コンビニエンスストアのチェーン店、それ以外は他の業種の店です。古書店にはそれぞれ専門があって、店によって品揃えがまったく違います。古書店とコンビニの立地を比べてみましょう。



問1) 古書店とコンビニの立地はどのように違いますか？

問2) その違いはなぜ生まれたのでしょうか？

東京のZ町には有名な古書店街があり、専門的な古書を求める客が全国から集まります。上の図はそんな古書店街の地図で、Bは古書店、S,L,Fは大手コンビニエンスストアのチェーン店、それ以外は他の業種の店です。古書店にはそれぞれ専門があって、店によって品揃えがまったく違います。古書店とコンビニの立地を比べてみましょう。

問 1) 古書店とコンビニの立地はどのように違いますか？

問 2) その違いはなぜ生まれたのでしょうか？

コンビニと古書店では客の行動が違います。専門的な古書を買うに来る客は、古書店を何軒も回って買い物をするのが普通ですが、コンビニでそのような行動を取る客はほとんどいません。それを考えると、このように立地パターンが違う理由がわかります。

コンビニと古書店街の立地の違い

	古書店街	コンビニ
商圈	全国	徒歩数分圏内
品揃え	専門的で店ごとに違う (競合しない)	似たり寄ったり (競合する)
立地の特徴	大通りの片側だけに 集中して並んでいる	分散している
その立地が 生まれた 理由	複数の店舗を回って買い 物をする客が多く、複数 の店が近くにあるほうが 集客しやすいため、大通 りの片側に集中した	1店だけで買い物を済ま せる客が多いので、他の コンビニの近くでは競合 してしまうため分散した

立地の特徴を一言で言うと、コンビニは分散しているのに対して古書店街は大通りの片側だけに集中して並んでいます。

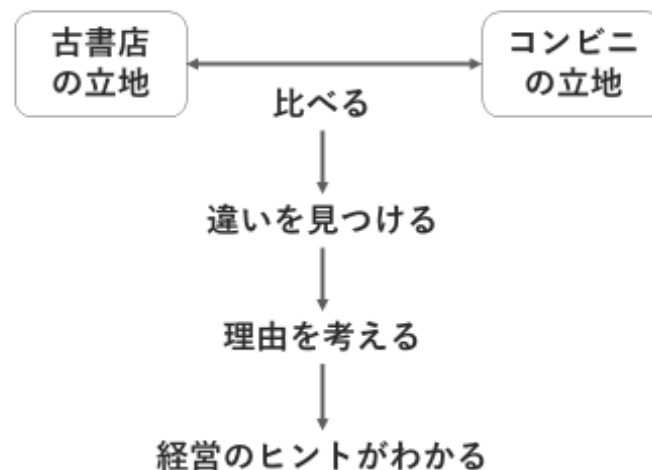
古書店街では複数の店舗を回って買い物をする客が多いため、複数の店が近くにあるほうが客にとって便利です。しかも、大通りはなかなか信号が変わらず横断するのに手間がかかるため、片側に集中するようになったと考えられます。もしこれが狭く横断しやすい通りであれば通りの両側に並んだことでしょう。一方、コンビニの客はほとんどが1店だけで買い物を済ませるので、他のコンビニの近くに出店すると競合してしまうため分散していると考えられます。

どんな業種であれ、独立開業して店を出すような場合、どこに出店するかは集客をするための最初の重要な選択になります。一度出店した店は簡単には移動できません。コンビニ的な店なのか専門古書店的な店なのかに応じて、立地の考え方は大きく変わってきます。

ちなみに、古い本を売る書店でも専門性のない「古本屋」と呼ばれる業態の場合はコンビニに近い面があります。

違うものには理由がある

「違う」ものにはたいてい何か理由があります。
理由が分かれば、それを手がかりにして、店を経営する上での「打ち手」のヒントが得られます。

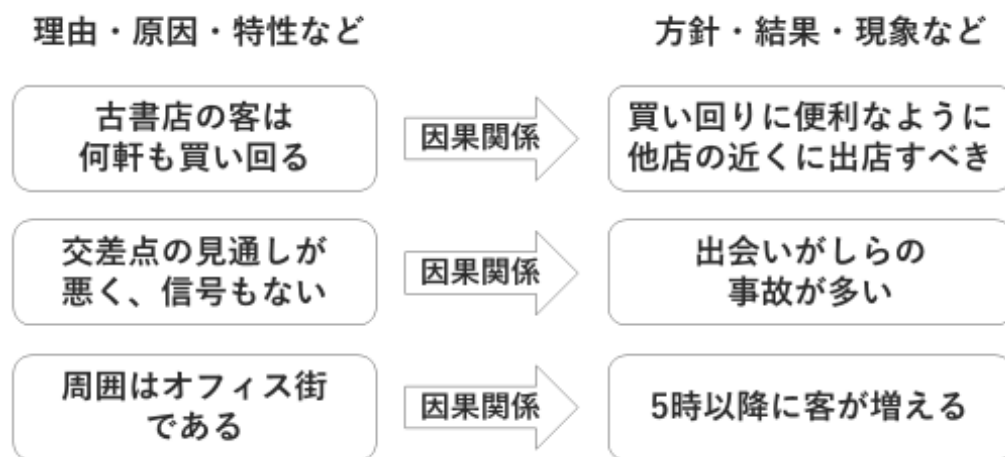


集客や仕事の進め方を改善しようとするとき、「比べてみる」ことは非常に大事です。比べてみると違いが分かり、「なぜだろう？」と考えるきっかけになります。きちんと考えたり調べてみたりすると違う理由が分かり、理由が分かれば店を経営する上での「打ち手」のヒントが得られます。

「比べる対象」は同じ業種でなくても役に立ちます。コンビニと古書店のようにまったく違う業種を比べることでヒントが得られる場合も珍しくありません。

因果関係を発見しよう

経営上の「打ち手」を考えるためには、さまざまな「因果関係」を発見しなければなりません。因果関係を正しく理解しているほど、適切な手を打てるようになります。



経営上の「打ち手」を考えるためには、さまざまな「因果関係」を発見しなければなりません。因果関係を正しく理解しているほど、適切な手を打てるようになります。

上図は因果関係の例ですが、因果関係の「原因」の側は他にも「理由」や「特性」などの名前で呼ばれることがあり、「結果」の側は「方針」や「現象」などの名前で呼ばれることがあります。呼び方は違っていても広い意味では因果関係であるというケースは非常に多いものです。上図はいずれも単純でわかりやすい例ですが、実際には何と何が因果関係になっているのかがわかりにくい例も非常に多いので、注意深く探っていかなければなりません。

数値で見ると違いがわかる(1)

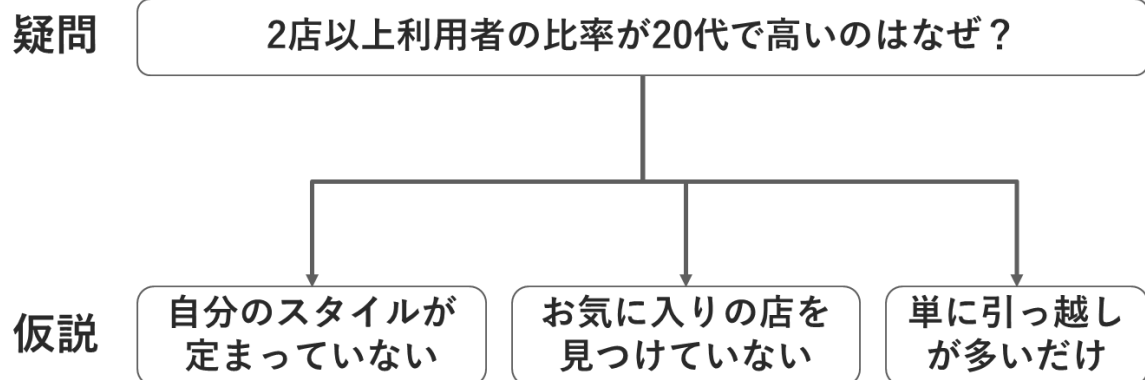
美容室利用状況に関する調査で、下記のような結果が得られたとします。
20代と50代の数値を見るとどんな違いがわかりますか？

年代	1年間の利用回数	1年間に2店以上を利用した人の割合(%)
20代	5.28	36.2
30代	4.95	28.2
40代	6.67	15.6
50代	7.69	13.4

出典：全国理美容製造者協会(2000年)

「比べて」違いを見つけようとするとき、数値を比べると違いがわかる場合が非常によくあります。
上図は美容室の利用状況に関する調査結果です。20代と50代の数値を見るとどんな違いがわかりますか？
そしてなぜその違いが生まれているのでしょうか？

違う理由について、仮説を考える



数値を見て直接わかることは大まかに下記の3つです。

- (1) 2店以上利用者の比率が20代で高い
- (2) 50代のほうが1年間の利用回数が多い

これを経営上の意味に言い換えると

- 20代の客は新規獲得しやすいが固定客化は難しい
- 50代の客は新規獲得しにくいが一度満足すれば長く使ってくれる可能性が高い

と言えそうです。もし「20代が2店以上利用している比率が高いのはなぜか？」という理由がわかったら、そこに手を打つことで「20代の新規獲得・固定客化」の双方を改善できるかもしれません。

そこでその理由の仮説を考えると、(1)自分のスタイルが定まっていないのであちこちの店を試している、(2)しかしいまだにお気に入りの店を見つけていない、(3)単に50代よりも引っ越しが多いだけといったものが考えられます。もし(1)が正しいとすると、「納得の行くスタイルを提案」することができれば、(2)お気に入りの店として固定客化できる可能性があります。一方、(3)が正しい場合は固定客化は難しいものの、引っ越しの時期に合わせた広告宣伝等が有効でしょう。

「数値」というのは小さな違いでも見つけやすく、「違いを見つけて理由を考える」ための最初のきっかけにしやすいものです。経営のさまざまな部分を数値化して比べてみることを心がけましょう。

数値の集計範囲を確認せよ

集計範囲の広い数値は、狭い地域の実情を表していないことがあります。
公的機関や業界団体等が集計した数値を参考にする場合は注意しましょう。



日本の生産年齢人口(15～64歳)は全人口の約60%

しかしN県T市K町でも同じ60%とは限りません



全国・全県の数値

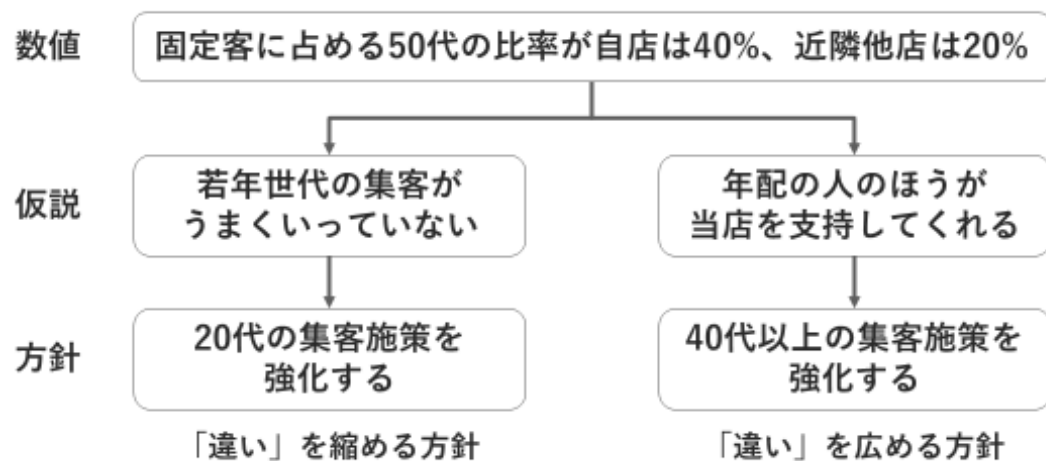


特定の店舗の営業範囲
(周囲数kmなど)

「数値を見ると違いが分かる」場合がよくあるのは事実ですが、公的機関や業界団体が「集計」した数値は集計範囲が広いため、自店の営業地域の実情と合っていない場合があることには注意してください。「周囲数 km の商圈で営業している店」にとって重要なのはその範囲の実情であって、それは全国や全県のデータとはかけ離れていることがあります。

「違い」を縮めるか広げるか？

「違い」を縮めるか広げるかに正解はありません。同じ店でも1年経てば答えが違う場合もよくあります。各自の経営環境の中で自店の強みが生きる方針をその都度自分自身で考えてください。



同じ数値から違う方針が出てくることもよくあります

同じ事実（数値）を元に、まったく逆の方針が出てくることは珍しくありません。「違い」を短所と考えると縮める方針を立てることも、長所と考えると広げる方針を立てることもありえます。店主やスタッフの個性と客の相性が大きく影響する業界・業態では「正解は人の数だけある」と言えます。どの方針であれば自分の強みを発揮できるのか？ を自分自身で考えましょう。

【事例】

アメリカの子供向け家具・おもちゃ小売店 Children's Supermart は1957年に不調であった家具から撤退し、好調であった子供向けおもちゃに絞り込んだ。同社はその後世界最大のおもちゃ小売チェーン「トイザらス」となった。→得意分野を見つけて「違いを広げた」方針の例。

ガラス製品を製造しているA社の香川工場と浜松工場にて、同じ製品の不良品率が10倍近く違っていたため、相互に技術者を派遣して改善に取り組んだ結果、不良品の大幅削減に成功した。→「違いを縮めた」方針の例。

数値で見ると違いがわかる(2)

年代ではなく満足度別に美容室の利用状況を調べると、下記のような結果が得られたとします。経営上気をつけるべきポイントは何でしょうか？

満足度	月1回以上利用 する人の割合(%)	1年間に2店以上を 利用した人の割合(%)
非常に満足	26.2	9.8
満足	19.0	19.4
やや満足	11.2	28.8
満足していない	11.5	47.5

出典：全国理美容製造者協会(2000年)

今度は「1年間に2店以上を利用した人の割合」という同じ項目を、年代別ではなく満足度別に見てみましょう。満足度別に見るとさらにハッキリと数値に差が出てきます。

この数値から読み取れることの1つは、当たり前ですが「不満を感じた客は他店に流れる」ということです。経営上は最低限「満足」のレベルをキープすることを目標にしなければなりません。そのためには、満足度を測る何らかの仕組みを持っておかなければなりません。

もうひとつは、「満足度を上げると来店頻度を上げられる可能性がある」ということです。今まで2ヶ月に1回だったお客様が毎月来てくれるようになれば、そのお客様からの売上は2倍か、少なくとも1.5倍にはなるでしょう。こちらの観点でもやはり「満足度」は非常に重要な指標であると言えます。

直接比較しにくいときは割合をとる

数値に大きな差があって比較しにくい場合は割合をとってみましょう

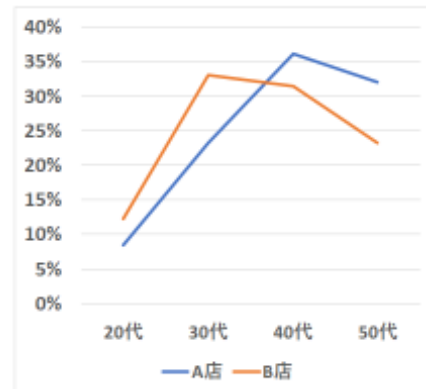
A、B両店の固定客の年齢階層分布

年齢階層	20代	30代	40代	50代	合計
A店	23	63	98	87	271
B店	82	221	210	156	669

大きな差のある数値は比較しにくい

年齢階層	20代	30代	40代	50代
A店	8%	23%	36%	32%
B店	12%	33%	31%	23%

「割合」にすると比較しやすい



グラフを作ると傾向がわかる

数値を比べてみようとしても、両者に大きな差があるときは比較しにくいことがあります。上図はA・B両店の固定客数を年齢階層別に調べたものですが、人数で表示した上の表で合計の数字を見ると271対669と2倍以上の差があるため、直接比較しにくい状態です。

そこで下の表のように「割合」にすると似た数字が揃うため、比較しやすくなります。グラフを作るとさらに傾向が分かりやすくなり、B店のほうが顧客の年齢層が若くなっていることが一目で分かります。

また、割合ではなく比率を出す方法も役に立ちます。

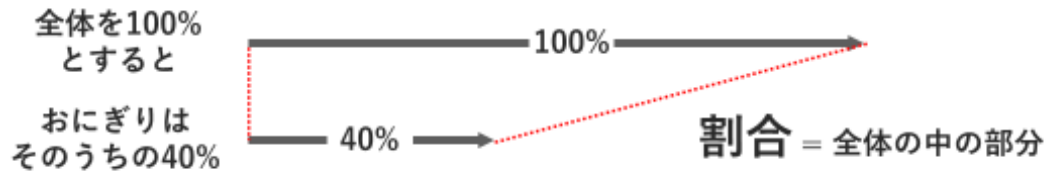
年齢階層	20代	30代	40代	50代	合計
A店	23	63	98	87	271
B店	82	221	210	156	669
B/A比率	3.6	3.5	2.1	1.8	2.5

20～30代の比率が高い

B店の固定客数がA店の何倍あるか、年齢階層別に比率を計算すると、「合計」よりも20～30代で比率の数値が高くなっているため、B店のほうが若い客が多いことがわかります。

割合と比率

ある日のある 店での売上額	おにぎり	お弁当	サンドイッチ
	8万円	6万円	6万円



	おにぎり	お弁当	サンドイッチ	
A店	4万円	2万円	3万円	
A店とB店の 売上額	↓ 2.5倍	↓ 4倍	↓ 2倍	比率 = 同種のもの の比較
B店	10万円	8万円	6万円	

「比率」と「割合」の基本的な意味は同じですが、実用的には少し違う場面で使われます。「割合」は、全体の中の部分を表す場合によく使われます。

【割合を表す表現の例】

- 当校の学生の8割は県内で就職します。
- このビールのアルコール度数は5.5%です。
- 当美容室のお客様の3割がカットのみのご利用です。

「比率」は同種のことを比較する場合によく使われます。

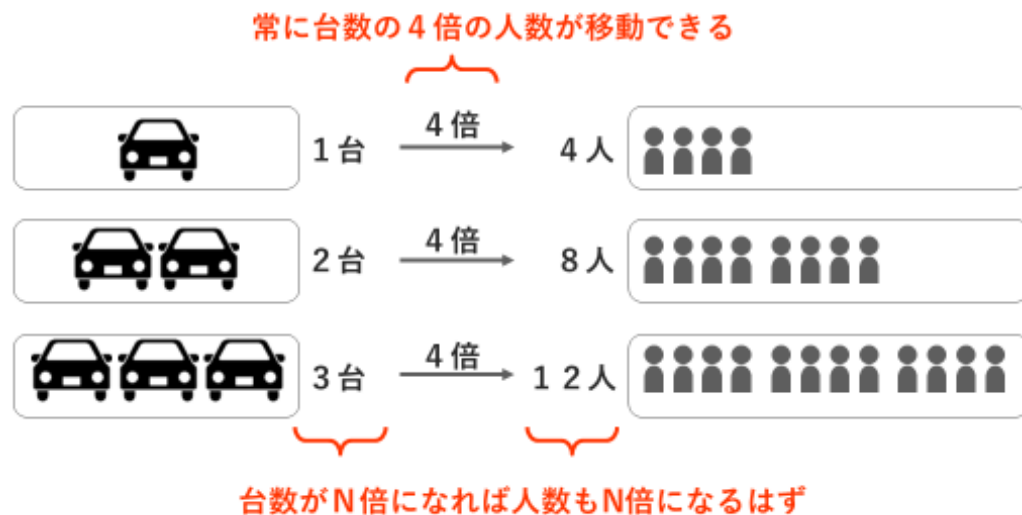
【比率を表す表現の例】

- A校の野球部員の数とB校の野球部員の数との約2倍でした。
- 減塩醤油の塩分量は一般の醤油の1/2です。
- その日の売上は前日の3倍でした。

ただしこの違いは厳密なものではなく、ある程度の目安です。いずれにしても「数値を直接比較しにくいときに、比較しやすくする手段」と言えます。

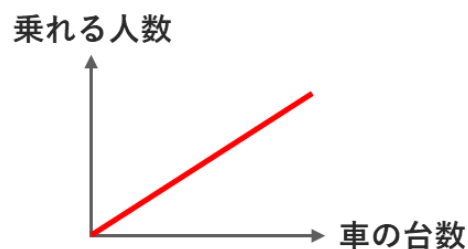
比例関係とは？

4人乗りの車が3台あります。全部使うと何人移動できますか？



……もしそうになっていなかったら、何かおかしい、とわかる

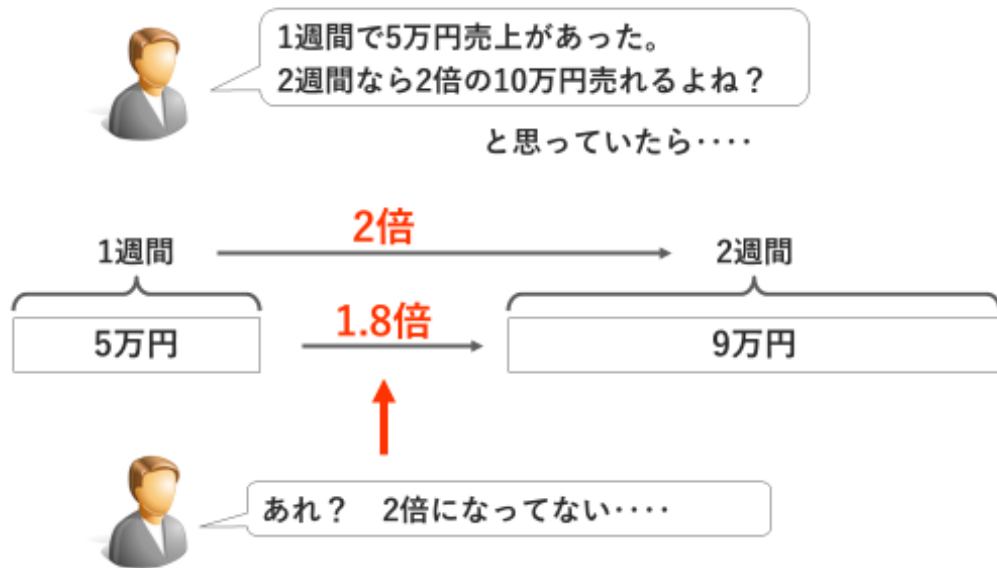
「比率」は比例関係のある数値に使うのが基本です。
比例関係というのは上図のように、一方がN倍になれば他方もN倍になるような関係です。
グラフを書くと原点を通る直線になります。



もちろん、「車の台数」というときの「車」は同じ種類のものでなければなりません。たとえば小型乗用車とバスを同じ1台と数えることはできません。

本来、同じ種類のものが複数ある場合は比例関係になるはずなので、もしそうになっていなかったら、何かおかしい、とわかります。

比率が役に立つとき



比例関係にある2つ以上の数値を比べてみると、「問題」に気づける

たとえば、1週間で5万円の売り上げがあったので、2週間では2倍の10万円売れると思っていたら9万円にしかならなかったとします。そんなときに「おかしいな？」と考えてよく調べてみると、表の看板に木の枝が伸びて見えにくくなっていることに気がついたりします。

モノやサービスの売れ行きはほんのちょっとしたさまざまな理由で落ちていくものです。落ちかけたときに素早く原因をつきとめて手を打てば、影響が大きくなる前に食い止められます。そのためには「問題」に素早く気がつかなければなりません。そこで役に立つのが「数値」です。といっても、数値が違いすぎると比べにくいのにに対して、比率は一定の数値に近くなるので比べやすいのです。

年齢階層	20代	30代	40代	50代	合計
A店	23	63	98	87	271
B店	82	221	210	156	669
B/A比率	3.6	3.5	2.1	1.8	2.5

数値が違いすぎると、比べにくい

比率は一定の数値に近くなるので比べやすい

そこで、比例関係にあるはずのものの比率を比べて異常が起きていないかを確かめることが、「問題」を見つけるための基本テクニックの1つになっています。

異常値を発見すれば原因がわかる

ある店のある月の1週目と2週目の売上に大きな差が出たので調べてみると

	1週目	2週目	2週目/1週目比率	
客数	227	218	0.96	客数があまり変わらないのに売上が落ちている
売上	287,082	259,321	0.90	
商品A	135,717	129,013	0.95	主力商品は落ちていない
商品B	126,905	120,840	0.95	
商品C	24,460	9,468	0.39	オプション商品が異常に落ちている

実際の事業の中での比例関係にはさまざまなものがあります。

【ファーストフード店】ハンバーガーを買う客はドリンクも買うことが多いので、バーガーとドリンクの売上に比例関係がある。

【携帯電話】携帯電話のついでにモバイルバッテリーを買っていくので比例関係がある

これらのさまざまな比例関係のうちの一部に異常値が出る場合があります。

上図はある店のある月の1週目と2週目の売上に大きな差が出た例です。

調べてみると、客数があまり変わらないのに売り上げが落ちています。客数が変わらないということは、集客の問題ではなく店内の問題の可能性が高いでしょう。

そこで商品別の売上を見ると、主力商品であるAとBは客数に見合った数字なのにオプション商品のCが異常に落ちています。そこでCに関してのみ調べてみると・・・

【ケース1】商品Cは、AやBを買おうとする客に対して「Cもあると便利ですよ」とお勧めして売る商品である。ところが2週目の初日にたまたま不機嫌な客がいて「余計なモノを売りつけるな！」と怒って帰ってしまわれたので、お勧めするのが怖くなってしまった・・・

【ケース2】展示入れ替え中に商品Cの宣伝パネルを一時的に撤去していて、そのまま忘れていた。主力商品ではないので気に留めていなかった。

といったことが分かったりします。このようなちょっとしたことでも売上は落ちていきます。それを早く察知して手を打つことが大事です。

間隔尺度の比には意味はない

気温は間隔尺度ですので、比率（割合）には意味はありません。
「差」には意味があります。

1		月曜	火曜	水曜	木曜	金曜
2	かき氷	52	23	2	2	2
3	ソフトクリーム	31	48	4	8	6
4	アイスコーヒー	40	37	5	28	20
5	ホットコーヒー	5	7	47	22	26

6	気温	33	30	22	25	22
---	----	----	----	----	----	----

7	気温の比率（対月曜）	100%	91%	67%	76%	67%
8	前日との温度差		-3	-8	3	-3

← 比率（割合）は無意味

← 差には意味がある

比率（割合）を計算すると数値を比べやすくなることは多いのですが、数値だからといって常にその方法が使えるわけではありません。たとえば「気温」について比率を計算しても意味はありません。

上図はある店の商品の曜日別の販売個数とその日の最高気温を表にしたものです。気温は月曜日が最高で火曜日に少し下がり、水曜日に大きく下がっています。

かき氷、ソフトクリーム、アイスコーヒーのような商品は暑いときによく売れますが、だからといって気温の比率を計算して「月曜日の 33℃に対して火曜日は 30℃だから 91%だ！」などと計算してもその比率は売れ行きとは無関係です。（参考までに：一般的には気温が 25℃を超えるとアイスクリームの売れ行きが良くなり、30℃を超えるとかき氷のほうがよく売れると言われています）

しかし、「温度差」には意味があります。水曜日は前日より 8 度下がっています。このように温度差が大きいと人間は「寒い」と感じやすくなるため、温かいものが売れます。しかし金曜日の気温は同じ 22 度ですが前日とあまり差が無いため水曜日に比べるとアイスコーヒーもよく売れています。ある商品がなぜ売れたのか、売れなかったのかを探る際、上記のように「温度差」との関係を見るとわかることがあります。

このように、「比を計算しても意味が無いが、間隔には意味がある」種類の量を間隔尺度と言います。気温は間隔尺度の代表的な数値です。

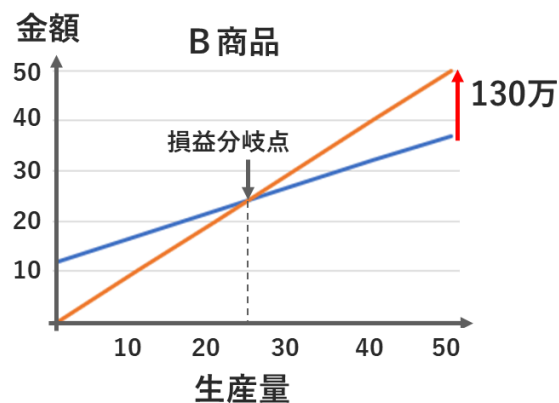
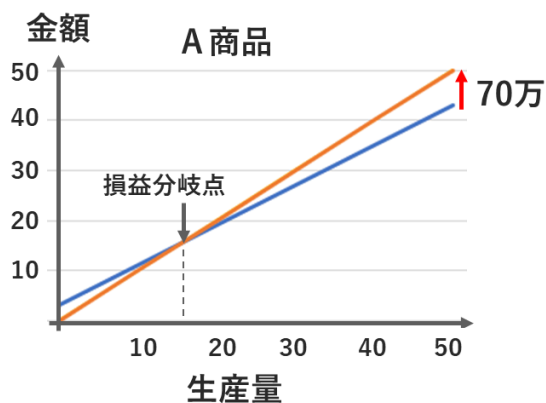
第3章. 仕事と変化

どちらが儲かりますか？

ある製麺所ではそばやうどんのような麺類を作って販売しています。現在の販売数はA・B商品とも月に3万パックで、どちらも1パックにつき10円の利益が出ます。どちらか片方を月5万パックまで増産可能で、増産分をすべて同価格で販売できるとしたら、どちらを増産する方が儲かりますか？

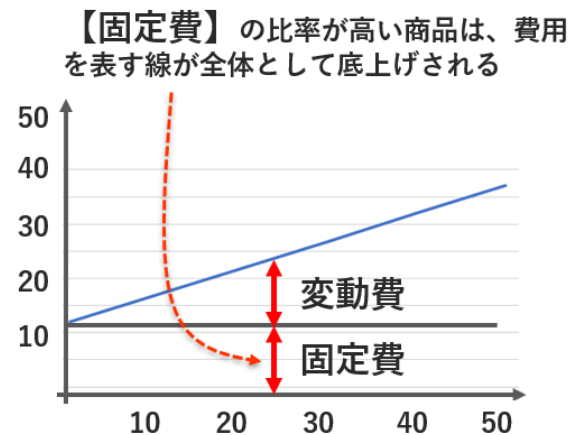
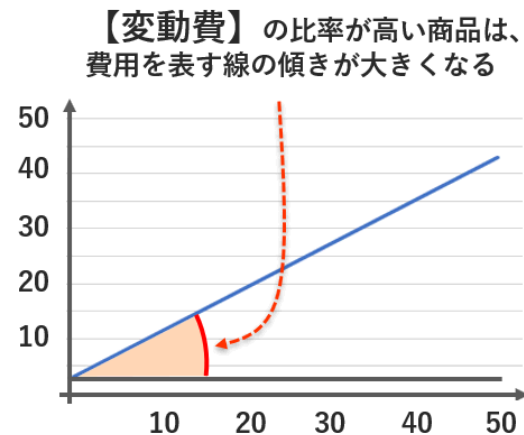
2019年1月の収支計算表			A		B	
販売価格(1パック)			100		100	
変動費	材料費		40	変動費計 80	30	変動費計 50
	労務費		30		10	
	販売費		10		10	
固定費	工場間接費		10	固定費計 10	40	固定費計 40
費用計(1パック当たり)			90		90	
営業利益(1パック当たり)			10		10	

A商品、B商品のどちらも1パックにつき同じ10円ずつ儲かるのであれば、どちらを増産してもかまわない……とはなりません。「1パックにつき10円儲かる」というのは生産量が月3万パックの場合の話であり、生産量が変わると差が出てきます。実際に生産量が0~5万までのA、B商品の費用と売上をグラフにすると下記のようになります。



費用 — 売上 生産量は1000パック、金額は1000円単位

どちらも単価100円なので、生産量に対する売上のグラフは変わりませんが、費用のグラフは大きく違います。その結果、利益（売上－費用）を計算すると、月産3万パックの場合はどちらも30万円ですが、5万パックになると、A商品の70万円に対して、B商品は130万円と大きな差がつきます。AとBでは固定費と変動費の構造が違うことに注意しなければなりません。



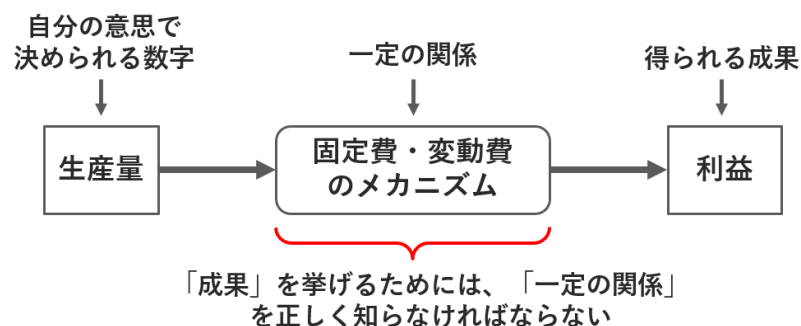
「固定費」というのは売上に関係なくかかる費用で、家賃などが代表的です。固定費の比率が高い商品について、生産量と費用のグラフを描くと、費用の線が全体として底上げされたグラフになります。

「変動費」というのは売上が増えるとそれにに応じて増える費用で、原材料費などが代表的です。変動費の比率が高い商品は、費用の線の傾きが大きくなります。その構造が異なるため、A 商品と B 商品には下記のような差が生まれます。

A 商品は固定費が低く変動費が高いので、
 損益分岐点が低い（生産量が少なくても利益が出る）
 しかし、分岐点以上に生産量が増えてもあまり儲からない

B 商品は固定費が高く変動費が低いので
 損益分岐点が高い（たくさん生産しないと利益が出ない）
 しかし、分岐点以上に生産量が増えれば急速に利益が増える

ここで、「自分の意思で決められる数字（生産量）」と「それによって得られる成果（利益）」の間に「一定の関係（固定費・変動費）」があることに注意しましょう。このような場合、高い「成果」を挙げるためには「一定の関係」がどのようなものなのかを正しく知らなければなりません。



「一定の関係」には固定費・変動費の他にも様々な種類があるため、自分の仕事のどこにどのような「関係」があるかを常に意識して考えていく必要があります。

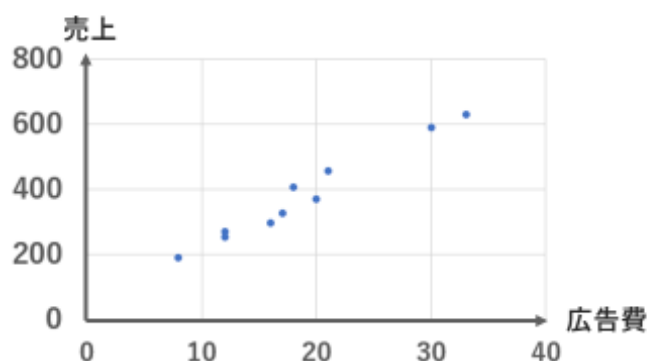
変数どうしの関係を数式で表す

ある会社の1号店から10号店までの店舗で使っている広告費と、その店舗での売上額の表とグラフです。グラフを見ると、広告費と売上額の間、ある「関係」がありそうです。どんな関係でしょうか？

	広告費	売上
1号店	10	255
2号店	12	327
3号店	11	297
4号店	5	190
5号店	24	590
6号店	18	458
7号店	17	371
8号店	33	631
9号店	8	270
10号店	16	409

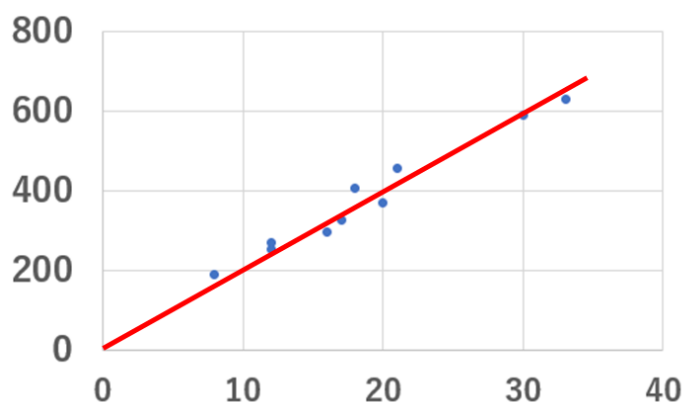
単位：万円

広告費－売上グラフ（散布図）



ある会社には1号店から10号店までの店舗があり、それぞれ個別に地域のポスティングや折り込みチラシ等の広告費を支出しています。各店の広告費と売上の関係を散布図というグラフにすると、1本の直線の上に乗るそうです。

実際、グラフの上に直線を1本引いてみるとこのように、売上＝広告費×20という式であらわされる、原点を通る一本のきれいな直線状になりました。この式は何を意味しているのでしょうか？



$$\text{売上} = \text{広告費} \times 20$$

候補 1：広告を出すと、それにかかった費用の 20 倍ぐらいの売上が上がる

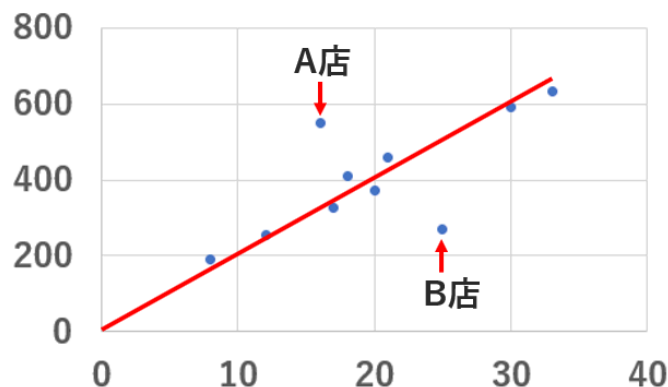
式だけを見ると上記候補 1 のようにも見えますが、実際にはほぼあり得ないと言って良いでしょう。もしこれが正しいなら、例えば今まで広告費に 10 万円かけていた店が 1000 万円かければ売上も 100 倍になる、ということですが、これは率直に言って非現実的です。

候補 2：どの店も、売上額の 20 分の 1（5%）ぐらいの費用で広告予算を出している

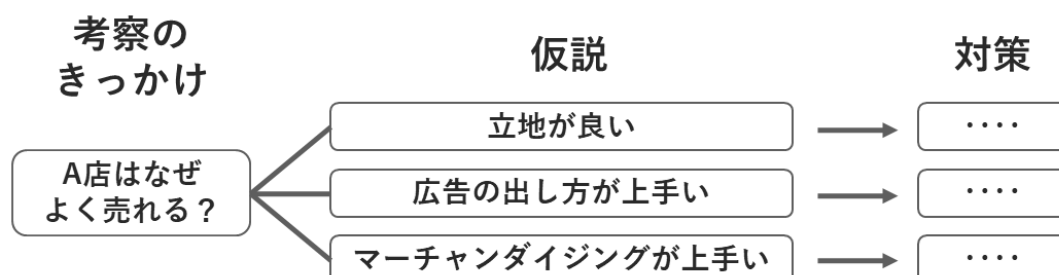
現実にはこの候補 2 が正しい可能性が高いでしょう。実際、このように「売上額の〇〇%」という基準で広告費を使っている会社は数多くあります。

さて、それではこのような数式がわかったとして、それが何の役に立つのでしょうか？

実際にリアルな会社のデータでこのような「広告費－売上」グラフを作ってみると、上記例のようにきれいな直線にはならない場合が多く、「外れ値」がしばしば出てきます。たとえば下の図では「A 店」は掛けた広告費以上によく売れている例、「B 店」はその逆です。



こうした「外れ値」からは、思いがけない改善のヒントが得られることがあります。A 店がなぜよく売れるのか？ と考えた時に、理由がおおよそ 3 つありえたとしましょう。



「立地が良い」のであれば他店では真似できませんが、「広告の出し方が上手い」のであれば、A 店の方法を学ぶことで他店の売上も改善できる可能性があります。あるいは、単に広告にとどまらず、商品の選択や店内の演出などの「マーチャンダイジング」の良さが効いているのかもしれませんが、その場合は広告の方法よりもマーチャンダイジングを学ぶべきでしょう。自立して事業を営むためには、このように何らかの「考察のきっかけ」から「仮説」を立てて「対策」をとっていくことが重要です。

その出発点となる「考察のきっかけ」は、「外れ値」からわかることが多いのですが、外れが外れであると気がつくためには「普通」がハッキリしていなければなりません。つまり、

普通の店は 売上＝広告費×20 になる

ということがわかっているからこそ、A 店や B 店はそれに当てはまらない「外れ値」である、と気がつきます。このように、「普通はこうなるよね」を数式で表せる場合は非常に多く、それができるとグラフを描いて「外れ」を発見しやすいので、「変数間の関係を数式で表す」ことはとても重要です。

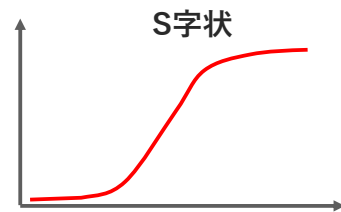
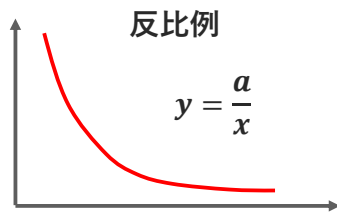
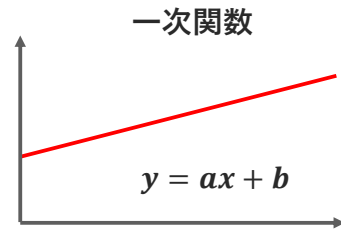
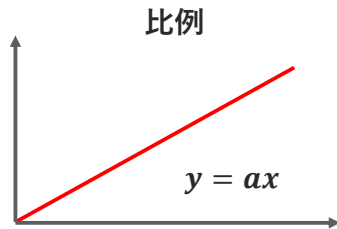
「変数」とは、「**変えられる／変わる**」数値のことを言います。「広告費」はいくら出すかを店長／社長が自分で「変えられる」数値です。「売上」はその結果「変わる」数値です。

仕事の中で出てくる様々な「数値」はすべて「変数」と言えます。それらの変数の中には、関係を数式で表せるものがあります。可能な限り、変数間の関係を数式で表していくことが重要です。

（注：なお、実際には「数値」ではない名義尺度のような値を持つ「変数」もあります）

さまざまな関係

2変数間の関係にもさまざまな種類があります。



2つの変数の関係にはさまざまな種類があります。もちろん現実にはきれいな直線や曲線を描けることは少ないですが、基本的な類型は知っておきましょう。

【比例】

原点を通る直線のグラフで表せる関係です。ある商品を1つ作るのに10分かかる→1時間あれば6個作れる、など、商品の生産に投じる時間と生産量の関係などによくあります。

【一次関数】

比例と同じく直線ですが、原点を通らないものです。実店舗では広告費をゼロにしても売り上げはゼロにはならない場合が多く、原点を通らない直線のほうが実態に近くなります。

【反比例】

商品ごとに売上額が多い順に並べると、一部の商品がよく売れて残りは急激に落ちていき、大半はゼロに近いという、反比例のグラフに似た形になりがちです。

【S字状】

広告を出しても最初はほとんど売れず、ある程度回数を重ねると売れ始め、限界に近づくと頭打ちになる場合が多いため、広告投入量と売上の関係はS字状の曲線になりがちです。

ローデータと単純集計

「変数」の値を得るためによく使うのが「集計」です。
下記はローデータを単純集計した例です。

ローデータ
(raw = 生)

売上#	ビール	ワイン	チーズ	パン	ソーセージ
1	○	×	×	×	○
2	○	○	×	○	○
3	×	○	○	×	×
4	×	○	○	○	×
5	○	×	×	×	○
6	×	○	○	○	○
7	○	×	×	×	×
8	○	×	×	○	○
9	○	○	○	○	○
10	×	○	×	×	×

集計

単純集計

集計	60%	60%	40%	50%	60%
----	-----	-----	-----	-----	-----

「変数」の中には「集計」しないとわからない数字もあります。たとえば上図はビールやワインなどの購買データですが、10 件の売上のうちで 60%がビールを、60%がワインを、40%がチーズを購入している……といった情報は「集計」しなければわかりません。

「ローデータ」のローは raw (=生) の意味で、1 件ずつの明細データのことです。たとえば売上# の欄に 1~10 までの数字がありますが、これは 10 人の個客の購入記録を表していて、1 番の客はビールとソーセージを買い、3 番の客はワインとチーズを買った、などの情報がわかります。

「単純集計」は、そのローデータをビール、ワイン、チーズ等の商品カテゴリごとに集計したものです。売れ行きが良いか悪いかを判断するには、ローデータではなくこのような集計データが必要です。

クロス集計

「クロス集計」は変数間の関係を探るためによく使われる方法のひとつです

(注：数字はいずれも架空のものです)

(1) 併買される食品

	ビール	ワイン	チーズ	パン	ソーセージ
ビール		33%	17%	50%	83%
ワイン	33%		67%	67%	50%
チーズ	25%	100%		75%	50%
パン	60%	80%	60%		80%
ソーセージ	83%	50%	33%	67%	

(2) (1)の80%以上のセルに着色

	ビール	ワイン	チーズ	パン	ソーセージ
ビール		33%	17%	50%	83%
ワイン	33%		67%	67%	50%
チーズ	25%	100%		75%	50%
パン	60%	80%	60%		80%
ソーセージ	83%	50%	33%	67%	

(3) 年代別の酒類購買傾向

	ビール	ワイン	酎ハイ	
20代	20%	31%	50%	← 表頭 (ひょうとう)
30代	35%	40%	43%	
40代	45%	45%	30%	
50代	60%	40%	30%	
← 表側 (ひょうそく)				

集計方法には単純集計の他にクロス集計という方法もあります。

(1) は併買される (同時に買われる) 商品を知るためのクロス集計の例で、縦軸横軸ともに同じ商品カテゴリーが並んでいます。このような集計により、「ビールを買った客の 33%がチーズを、83%がソーセージを同時に買っている」といったことがわかります。これがわかると「ビールの売り場でソーセージのお勧め品を展示しておくとうまく売れるかもしれない」といったアイデアが得られます。このように「併買される商品の傾向を知る」ことは小売業のマーチャンダイジングにおける重要な課題です。

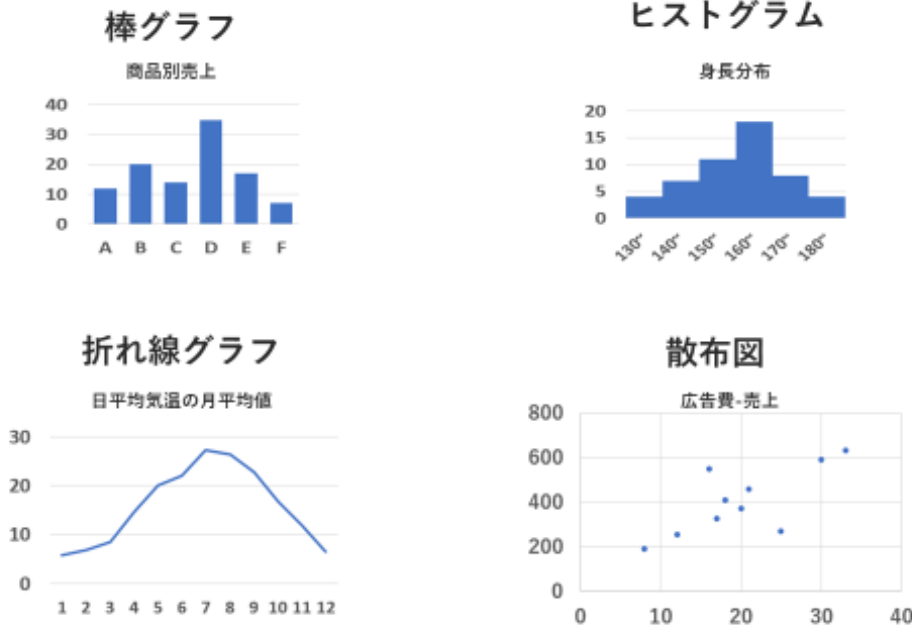
(2) は (1) と同じ集計表ですが、80%を超える数字のセルを着色して目立たせたものです。クロス集計は高い数字や低い数字に意味がある場合が多いので、単に数字を出すだけでなく、このような工夫で注目点を目立たせるとわかりやすくなります。

実用的に最もよく使うのは、(3) のように縦と横で別な分類軸を使って行うクロス集計です。年代別、性別、居住地別などで商品の好みの傾向が分かれることは非常に多いため、この方法がよく使われます。

なお、クロス集計の横軸のことを「表頭 (ひょうとう)」、縦軸のことを「表側 (ひょうそく)」と呼びます。

グラフによる表現

数値を分かりやすく見せるために、様々なグラフが使われます



数値を分かりやすく見せるためには様々なグラフが使われます。

【棒グラフ】

商品別売上など、連続性のないデータを表すときによく使われます。たとえばA、B、Cがそれぞれリンゴ、牛乳、ウィスキーのような商品だとすると、それらの売上には通常連続性がないため、棒グラフの「棒」どうしはスペースを空けて分離して書きます。順序を変えて表示するのも自由にできるのが普通です。

【ヒストグラム】

1 クラスの生徒の身長分布、日本人の年齢構成など、連続性のあるデータを人為的にいくつかの「階級」に区切って集計表示するために使われます。たとえば身長を 150cm~、160cm~のように 10cm 単位の階級に区切って集計することはよくありますが、このような区切りは集計の都合上設定しているだけで、本来何の意味もありません。そこで、本来は連続しているデータである、ということを示すために棒グラフの「棒」どうしの間にスペースを空けずに書いたものがヒストグラムです。棒グラフと違って、ヒストグラムの階級は順序を変えて表示することはできません。

【折れ線グラフ】

連続性があり、長期間継続しているデータを表すためによく使われます。たとえば「気温」は前月と当月で極端に違うことはなく、連続性があるため折れ線グラフで表現するのに向いています。

【散布図】

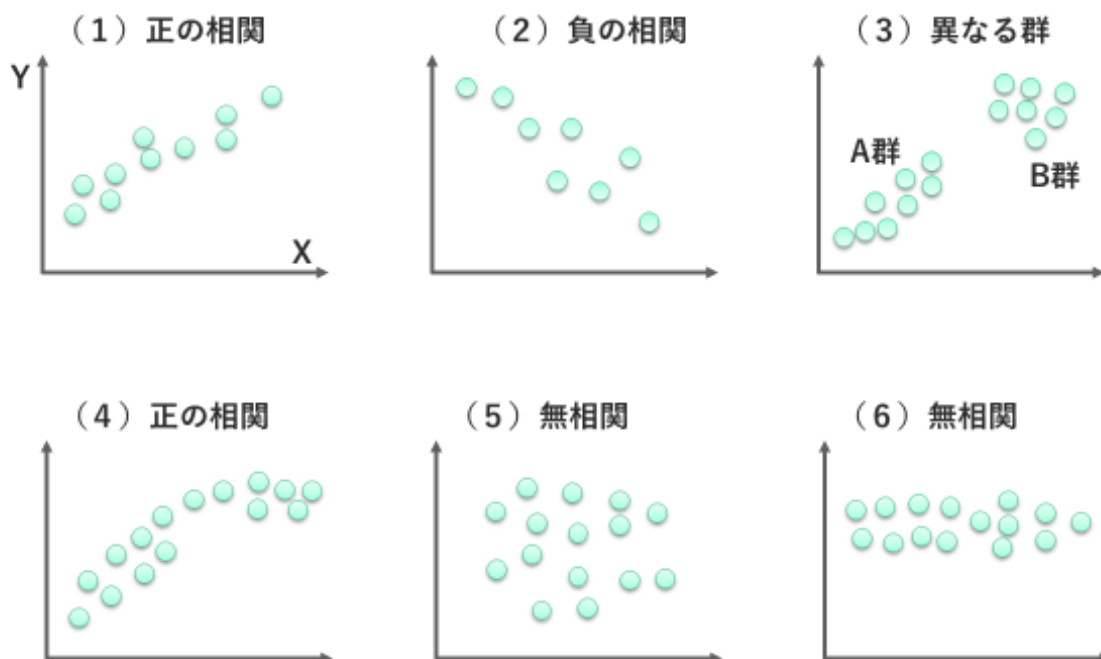
関連性があるかないかわからない複数のデータをタテ軸と横軸にとって作る、「バラバラに散らばった点」のようなグラフが散布図です。データに関連があるかないかを検証するためによく使われます。

例 1 : 「数学が得意な人は理科も強いだろう」という仮説を検証するため、数学と理科の得点を縦・横の軸にして散布図を描く

例 2 : 「高所得者はワインを好むだろう」という仮説を検証するため、「年収」と「年間のワイン消費量」を縦・横の軸にして散布図を描く

散布図を描いたときに、なんらかの直線が引けそうな分布をしていれば、関連がある可能性が高いと言えます。

散布図からわかる関係



散布図のパターンで関係の有無を判断する例です。

(1) は右上がりの直線状に分布しています。これは横軸 X が増えるほど縦軸 Y も増える関係であり、このような関係を「X と Y の間に正の相関がある」と言います。「年齢が高いほど身長が高くなる」、「年収が高いほど外食日数が増える」などがこの例です。

(2)は(1)とは逆に右下がりの直線状に分布しています。このような関係を負の相関と言います。

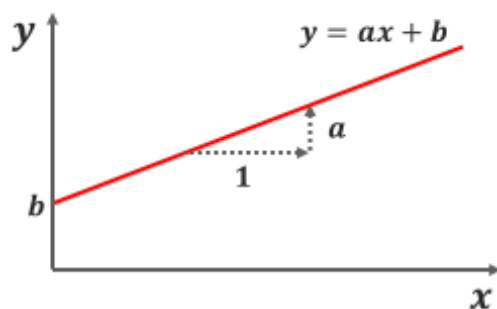
(3)ではA群とB群の間にギャップが見えます。このような場合、おそらくA群とB群は異なる性質を持った集団です。たとえば日本人と外国人に同じアンケートをとった場合など、このような形になる場合があります。(3)の例ではA・B両群の間にギャップがあるため分かりやすいですが、調査の種類によってはギャップがはっきりしないことがあります。A群だけを調べれば「正の相関」がありそうですが、AとBのギャップがはっきりせず両者を分離できないと、それを発見するのが難しくなります。

(4)の左半分はXとYに正の相関が見えますが、右半分はXが増えてもYは変わらなくなっています。たとえば携帯電話は一人1台以上持つことはあまりないため、全員が持つようになるとそれ以上は売れません。普及が頂点に達して需要が飽和したような商品でよくあるパターンです。

XとYに関係がない場合は、(5)のように線ではなく「面」として広がった形になるか、(6)のように傾きのない直線状になります。「関係がない」ことを「XとYの間には相関がない」または「無相関である」と言います。

一次関数

「一次関数」は、 $y = ax + b$ という数式で表せる関係です。
仕事の中では非常に多くの場面で出てきます。



x が 1 増えると y が a 増える。
 $x=0$ のときの y の値は b 。

代表的な例： 製造個数と売上原価の関係、
水光熱費等の従量制サービスの費用、

「さまざまな関係」の項でも出てきましたが、仕事の中で出てくる数値の関係の中には「製造個数と売上原価」「水光熱費等の従量制サービスの費用」など、「一次関数」で表せるものが非常によくあります。

一次関数は数式にすると $y=ax+b$ という形で、グラフを描くと直線になります。

【製造個数と売上原価】

うどんやそばのような商品を製造していて、「売上原価」にたとえば小麦粉やそば粉のような材料費と工場の家賃が含まれるとしましょう。通常、材料費は製造個数に比例して増えます。これが「 x が1増えると y が a 増える」の部分、 ax に該当します。一方、工場の家賃は製造個数に関係なく一定なのが普通です。これが「 $x=0$ のときの y の値は b 」の部分、 $+b$ に該当します。

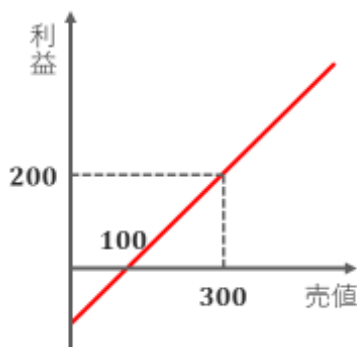
【水光熱費等の従量制サービスの費用】

水光熱費等は、まったく使わなくてもかかる基本料金+使った分に比例してかかる従量制料金という構成になっていることが多く、これも $y=ax+b$ になります。

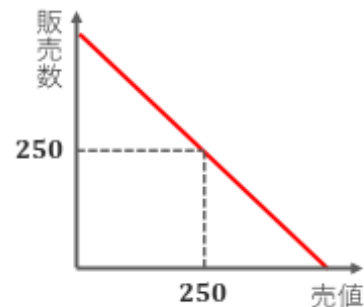
一つの変数が複数の数字に影響する

ある商品1個あたりの原価が100円だとします。
売値を高くすれば1個あたりの利益は増えますが売れる個数は減ります。
売値を安くすれば1個あたりの利益は減りますが売れる個数は増えます。
もし1個 x 円で売ると1日当たり $500-x$ 個売れるとしたら、利益を最大にするためにはいくらで売れば良いのでしょうか？

1個あたり利益 = 売値 - 100



販売数 = 500 - 売値



一次関数 $y=ax+b$ で表せる関係の場合、 x が増えたとき y は減るか増えるかどちらか片方に決まります。

$y= ax+b$ x が増えれば y は増える

$y= -ax+b$ x が増えれば y は減る

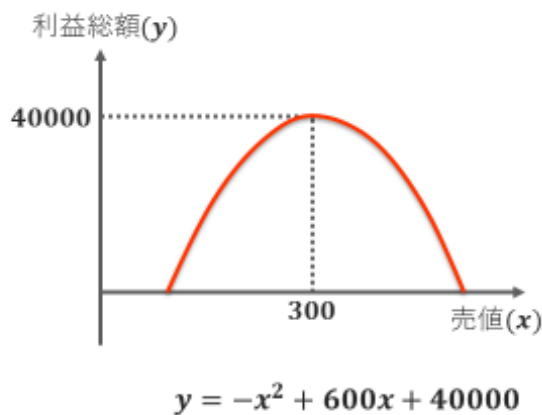
しかし、仕事で現れる数字の中にはこうならないものもあります。

たとえば「値段」と「利益」の関係はその 1 つで、値段を上げると利益は増えますが、上げすぎると売れなくなってかえって利益が減る傾向にあります。このような関係は一次関数では表せません。

二次関数

1個あたり利益×販売数 で利益の総額になります。

$$\begin{aligned}\text{利益総額} &= 1\text{個あたり利益} \times \text{販売数} \\ &= (\text{売値} - 100) \times (500 - \text{売値}) \\ &= -\text{売値}^2 + 600 \times \text{売値} + 40000 \\ &= -(\text{売値} - 300)^2 + 40000\end{aligned}$$



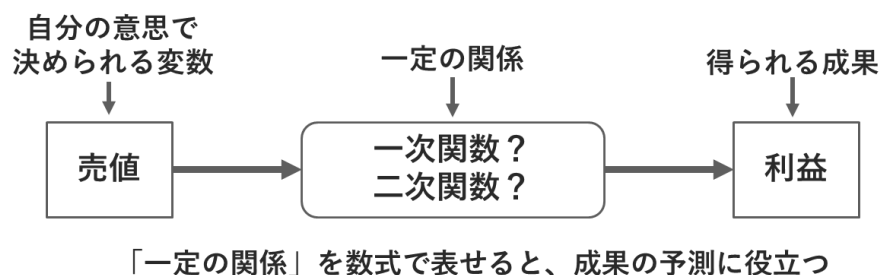
このような場合、売値と利益総額の関係は左のような曲線（放物線）を描きます。

x^2 を含む式を二次関数式といい、そのグラフは放物線になります。

1 個あたりの利益に販売数を掛けると利益総額になります。1 個あたり利益が 売値-100、販売数が 500-売値になる場合、売値と利益総額のグラフを描くと上図のような曲線になり、売値=300 円のときに利益総額が最大になります。

この場合、式は $y = -x^2 + 600x + 40000$ のようになります。 x^2 の項を含む式で表せる関数を二次関数と言い、そのグラフは放物線になります。

この例は「二次関数」を説明するための極端な設定であって、実際には売値と利益がこのような単純な放物線の関係になることはほとんどありません。しかし、経営をするためには「自分の意思で決められる変数」を変えたときに「どんな成果が得られるか」を予測する必要があります。変数と成果の関係は一次関数や二次関数である程度近似できる場合があり、この考え方を知っておくと予測・判断をするために役立ちます。



第4章. 仕事と集合 I

分類してください

あるお店のお客様リストをいくつかのグループに分類してください。
分類する基準にはどのようなものが考えられますか？

氏名	性別	住所	年齢
田中一郎	男	東区境田...	22
鈴木さとみ	女	東区山形...	25
林田亜里沙	女	西区横河...	37
佐藤浩二	男	南区磯部...	26
榊原安昌	男	東区片田...	41
伊達みどり	女	西区上町...	52
野沢公子	女	東区岩瀬...	32

上の表はあるお店のお客様リストの一部ですが、これを分類するとしたらどのような基準が考えられるでしょうか？

【性別】

女性向けファッション商品の企画をしても男性客の興味を引くことはできませんし、サッカーや野球などのスポーツ用品に興味を持つのは男性が中心です。性別で商品やサービスの好みが違うケースは非常によくあるため、性別での分類はよく使われる基準です。

【住所】

住所からもさまざまなことがわかります。店に近いか遠いかは来店頻度に影響します。幹線道路沿いの住人は騒音や大気汚染対策グッズに興味があるかもしれません。

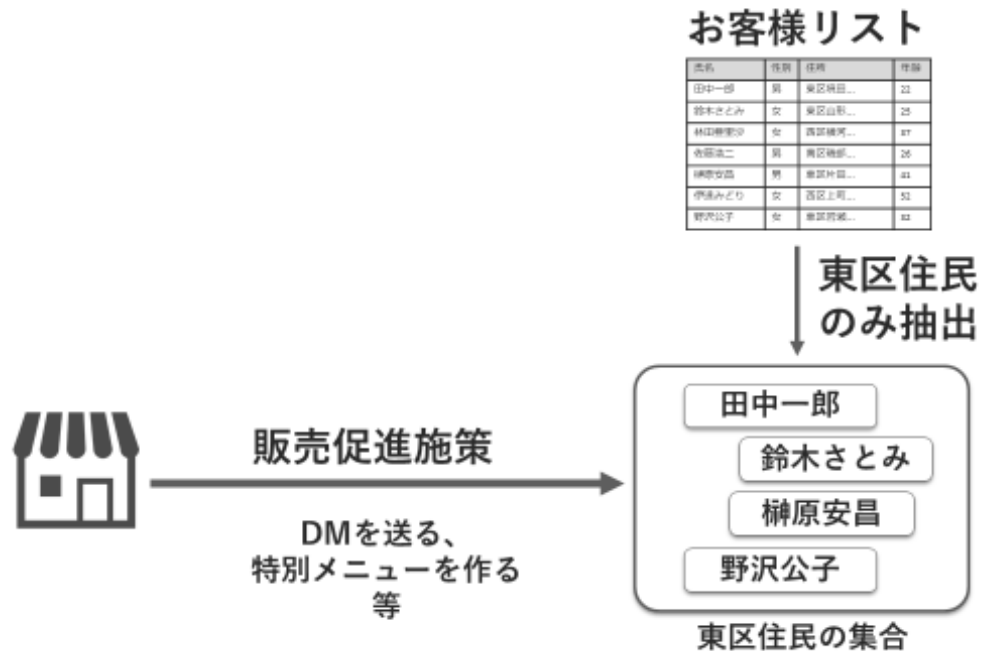
【年齢】

年齢層によっても当然さまざまなニーズが変わってきます。年齢層によって好まれる服装の傾向は違いますし、20~30代と40~50代では子供の教育に関するニーズも違うでしょう。

このように、「お客様リスト」は一つであってもそれを分類する基準は千差万別なのが普通です。

それはどのような「集合」ですか？

相手の集合に共通する性質を手がかりにして販売促進施策を考えます。



販売促進施策を考えるには、相手の「集合」に共通する性質を手がかりにします。

「集合」というのはもともとは数学の用語ですが、本書では「何らかの共通な性質を持つ要素を集めたもの」という意味で使います。たとえばお客様リストからは

- 女性客の集合
- 東区住民の集合
- 30代のお客様の集合

など、様々な「集合」を抜き出すことができます。「全員が興味を持つ商品／サービス」というのはあり得ませんので、「女性客」「音楽をやっている人」「映画好きな人」など、対象とする顧客のイメージを描いてその興味を引くように販売促進の企画を立てるのが基本です。

集合の全数を調査することは難しい

ある名簿に1000人が登録されているとします。このうち、日常的にスポーツをしている人がどのぐらいいるか調べるにはどうすれば良いでしょう？

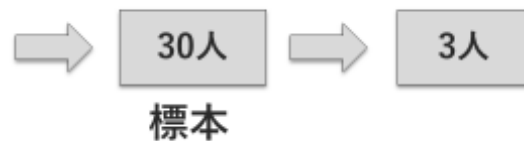
氏名	性別	住所	年齢
田中一郎	男	東区境田...	22
鈴木さとみ	女	東区山形...	25
林田亜里沙	女	西区横河...	37
佐藤浩二	男	南区磯部...	26
神原安昌	男	東区片田...	41
伊達みどり	女	西区上町...	52
野沢公子	女	東区岩瀬...	32

.....
母集団
 (全1000件)



全員に聞いて回るのは難しいから...

30人だけ聞いてみたら1割(3人)だった



1000人全員の1割なら100人かな？



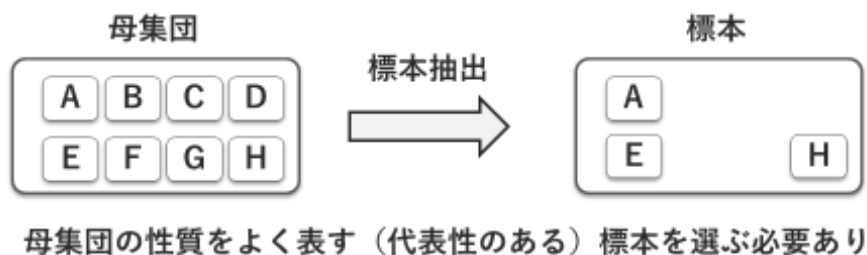
ある集合に共通する性質が分かれば経営判断の重要な手がかりになりますが、それが簡単ではない場合もあります。たとえばある店で「日常的にスポーツをしている人」向けの栄養食品を安く仕入れられる機会があって、ただし最低発注数が20個だったとしましょう。20個をすべて売ろうとすると、興味を持ちそうな人がその2倍の40人は必要だとします。

その店の顧客リストに1000人の人がいたとして、そのうち「日常的にスポーツをしている人」が40人いるかどうかを、どうすれば調べられるでしょうか？

1000人全員に聞いてみるのは不可能ですので、そのうちの30人にだけ聞いてみたらそのうちの1割、3人が該当したとしましょう。1000人全員でもこの比率が成り立つのなら、全体では100人になります。40人よりはかなり多そうなので、20個発注しても大丈夫でしょう。

このような場合、「顧客リスト全員」にあたる1000人を「母集団（ぼしゅうだん）」、実際に聞いてみた30人を「標本（ひょうほん）」と言います。母集団の数が多い場合、全数を調査するのは難しいことが多いので、一部を取り出した「標本」で全体の傾向を推定する方法をとります。

母集団と標本の代表性



目的	ある学校の生徒の平均身長を知る
標本抽出法	悪い例 バレーボール部部員の身長を計る
	良い例 出席番号下一桁が1の生徒の身長を計る

標本を選ぶときは母集団の性質をよく表す標本になるように気をつけなければなりません。

たとえば、ある学校の生徒の平均身長を知りたい場合、「バレーボール部部員」を標本として選んだのでは正しい結果は得られません。バレーボール部には通常、高身長の生徒が多く集まっているからです。それに対して、「出席番号下一桁が1」という基準で選べばそのような偏りはないと考えられるため、適切な標本と言えます。

このように、「標本が母集団の性質をよく表している」ことを「代表性がある」と言います。代表性がある標本を選ぶのは意外に難しいので、慎重に考えなければなりません。

たとえば、あるコンビニエンスストアに車で訪れる客の人数を調べたいとしましょう。その店に1日じゅう張り付いて車で来店する客を数えれば良さそうにも思えますが、おそらく平日と休日では傾向が違いますし、年末年始やゴールデンウィークのような時期もあれば、花火大会やコンサートなどの大きなイベントでも変わってきます。周辺の事故や工事で車の流れが変わる場合もあります。

代表値：最大・最小・平均・中央・最頻



最大値	32	集合に含まれる要素の最大の値
最小値	5	集合に含まれる要素の最小の値
平均値	16	集合に含まれる要素の平均の値
中央値	13	要素の中で中央に現れる値
最頻値	10	要素の中で最も頻繁に現れる値

要素が数値であるような集合の性質を表す「代表値」が何種類かあるので覚えておきましょう。(前ページの「代表性」とは別な概念です)。

「最大値 (さいだい値)」 「最小値 (さいしょう値)」 は集合に含まれる要素の中のそれぞれ最大と最小の値、「平均値」 はすべての要素の数値を足して要素数で割った値です。平均は「平均点」などの用語でも使われているので有名です。

「中央値」は、要素を大きさ順に並べたときに中央に出てくる値です。平均値と紛らわしいので区別しましょう。

「最頻値」は、最も頻繁に現れる値です。上図の例では「10」が 3 回現れているので最頻値になります。

平均値と中央値の違い

選手年俸表

選手	年俸
A	4億5000万円
B	1億2000万円
C	9000万円
D	8000万円
E	6000万円
F	5000万円
G	4000万円

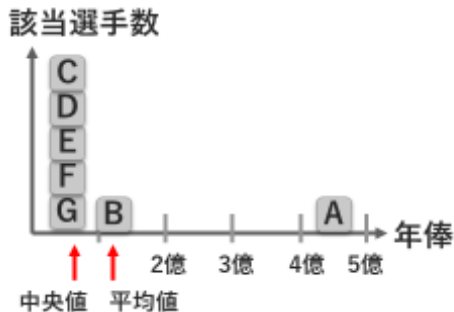
← 平均値：1億2714万円

突出した外れ値の影響を受けやすい

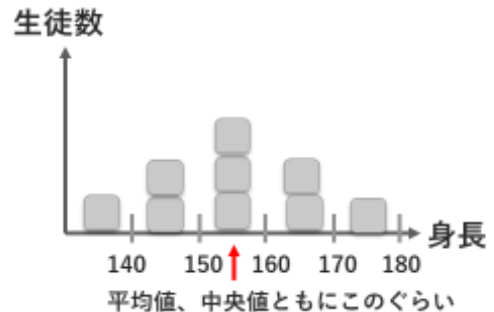
← 中央値：8000万円

外れ値が影響しにくい

年俸階級グラフ



あるクラスの男子生徒身長分布



平均値と中央値には重要な違いがあります。上図左上の「選手年俸表」はあるプロスポーツ・チームの選手の年俸表だとします。A選手が突出して大きく、他は高くても1億円を超える程度です。このようなデータの平均値を取ると1億2714万円、中央値では8000万円です。

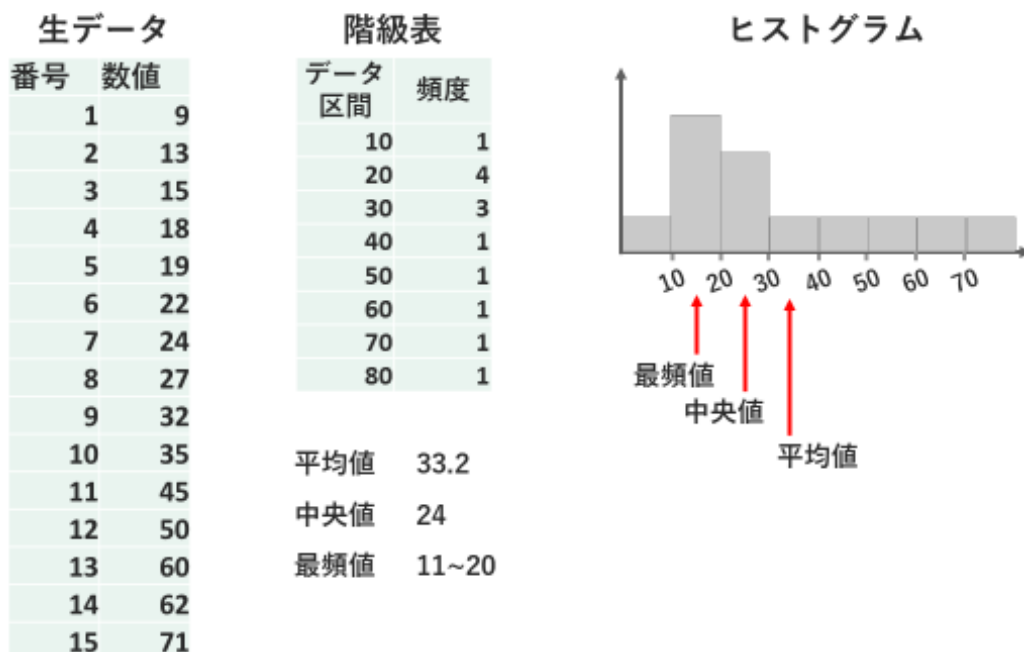
では、「このチームの普通の選手の年俸、どれぐらい？」と聞かれたとき、どちらの数字のほうが現実に近いでしょうか？明らかに中央値の「8000万円」に近い金額をもらっている選手のほうが多いので、中央値のほうがこのチームの選手年俸の実態を代表していると言えます。

上図左下の「年俸階級グラフ」は、1億、2億・・・という年俸階級別に、該当する選手数をグラフにしたものです。Aが明らかに突出していますが、「平均値」はこのように「突出した外れ値」の影響を受けやすい欠点があります。

右下の「あるクラスの男子生徒身長分布」は、身長の数字で同様のグラフを作ったものです。このように「突出した外れ値」が少なく、中央に山ができて左右になだらかに減っていくタイプのデータでは平均値と中央値に大きな差は出ません。

データの偏りが大きい場合、「平均値」は「集合の代表値」としては使えないので注意してください。

最頻値と中央値の違い



上図左側の「生データ」を見ると数値が 9~71 まで分布しています。これを 0~10、11~20、21~30 のように 10 ずつの「階級」に区切ってデータの出現個数を数えたのがその右の「階級表」です。たとえば階級表の 2 行目、データ区間 20 の頻度が 4 というのは、生データで数値 11~20 までの値をとるデータが 4 個あったという意味です。その下の平均値、中央値は生データに対して平均値と中央値を求めたものです。

一方、「最頻値」はデータの階級単位で考えます。この例では 11~20 の区間が最頻値だったことになります。

右の「ヒストグラム」は階級表をグラフにしたもので、データ区間ごとに何個のデータが出現したかが一目でわかります。前ページの「年俸階級グラフ」などのグラフもヒストグラムの一種でした。

ヒストグラム上に平均値、中央値、最頻値を描くと上図のようにやはりそれぞれ違う数値になります。「平均値は必ずしも最も普通のデータではない」、というのは前ページで記載したとおりですが、では最頻値と中央値の違いは何でしょうか？

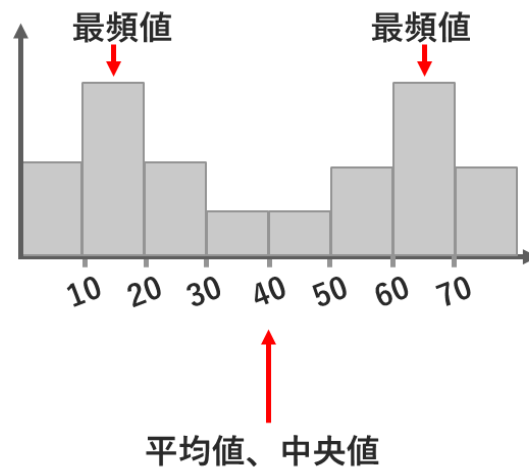
実は、販売促進の企画を立てる場合など、「中央値」よりも「最頻値」を使って考える場合がよくあります。たとえば「10 代の人向け商品」「20 代の人向け商品」のように年齢層で区切って企画を考えるとしましょう。上図の「数値」が年齢だとすると、10 代（データ区間 20 に該当）は 4 人、20

代（データ区間 30 に該当）は 3 人います。つまり「10 代向け商品」のほうが、買ってくれそうな客が一人多いわけです。

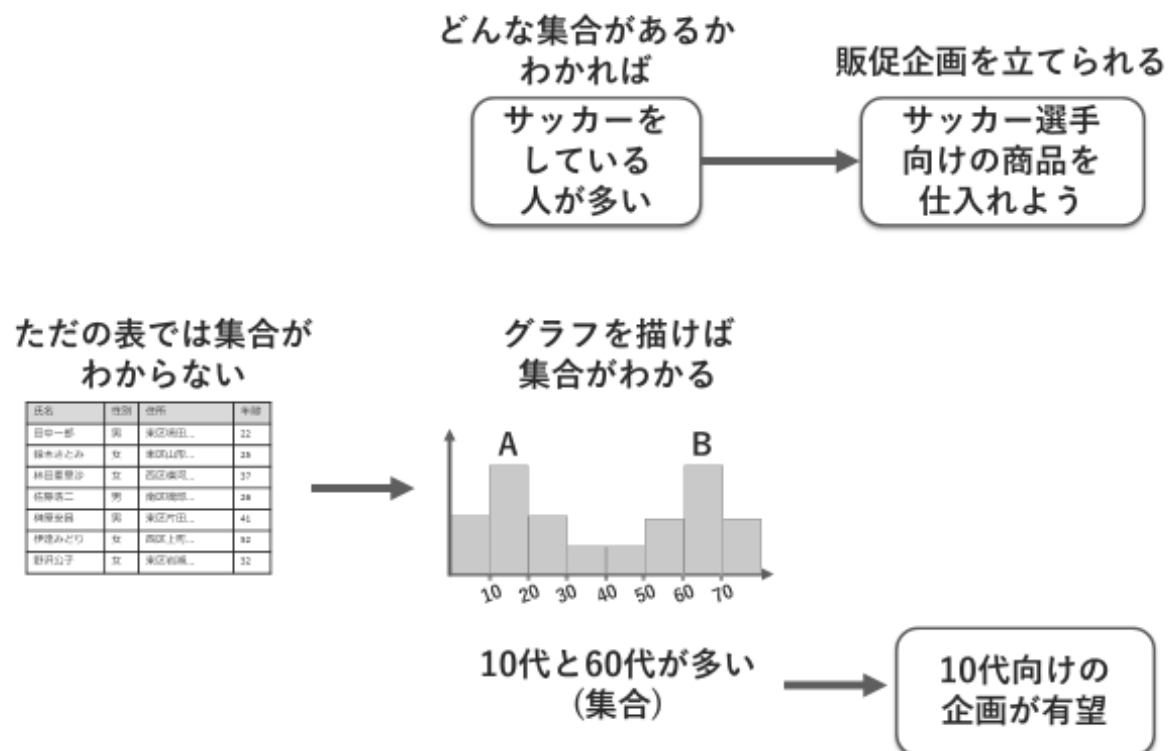
このように、データを階級で区切って考える場合は「最も数が多い階級」が重要になるので、最頻値という考え方も知っておきましょう。

最頻値は平均値や中央値とも大きく違う値になることがあります。たとえばデータが下図のように二つの山があるタイプのものの場合、平均値、中央値はいずれも真ん中付近になりますが、最頻値は右(61~70)と左(11~20)の二か所に出てきます。このような場合、平均値や中央値で企画を立ててもろくに売れません。最頻値に合わせて企画を立てなければならず、しかも最頻値は一つだけとは限らない、そんな違いがあります。

フタコブラクダ型データ



グラフを描けば集合が見つかる

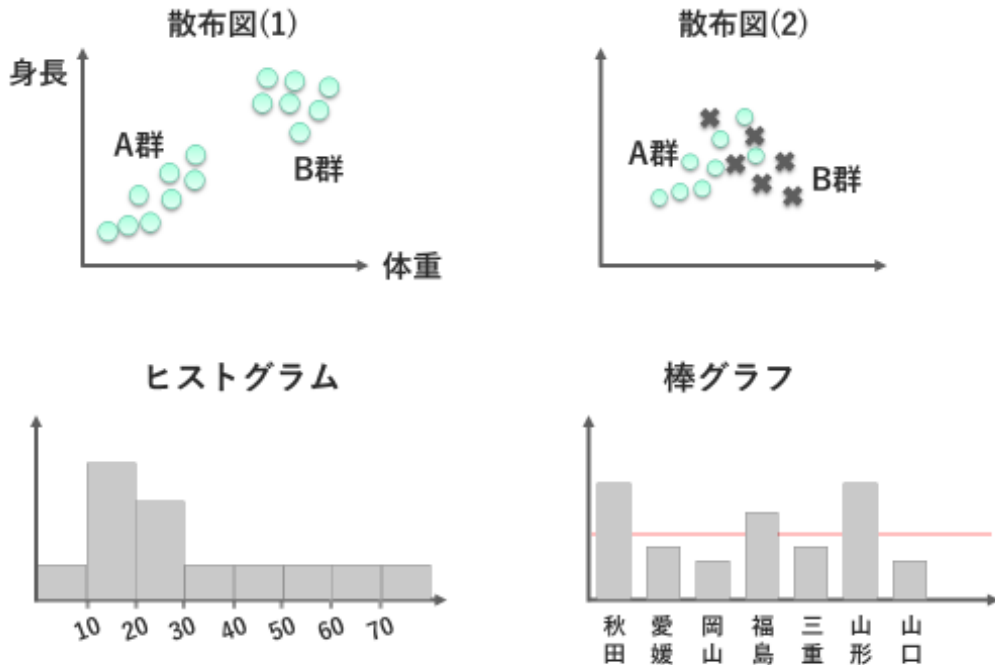


既に触れたように、販売促進の企画（販促企画）は特定の「集合」に対して立てるものです。たとえば自社の顧客の中にサッカーをしている人が多ければ、「サッカー選手向けの商品を仕入れる」など、サッカーに関わる企画が有望です。

それにはまず「集合」を見つける必要がありますが、顧客リストという「ただの表」を見ていてもなかなかわからない「集合」が、グラフを描くと分かることがあります。たとえば上図中央のヒストグラムを見ると、数字が年齢だとして、10代と60代に高い山があります。であればその集合、つまり10代や60代向けの企画が有望だろう、と分かります。

顧客リストはたとえば「Aさんは18歳、Bさんは22歳、・・・」のような細かい数字をまとめた表ですが、「集合」というのは細かい数字を捨てて「10代の顧客は〇〇人」のように大ざっぱな範囲や傾向を見る考え方です。そのためには「グラフ」のほうが扱いやすいのです。

集合を見つけるためのグラフ(1)



「グラフ」にも非常に多くの種類がありますが、「集合を見つけるためにグラフを描く」ためによく使われるグラフの種類や注意点を挙げておきます。

たとえば公園に遊びに来ている親子連れの身長と体重を量ると散布図(1)のようなグラフになるでしょう。A群は成長途中の子供で B 群は親です。子供はある程度育つと「親と一緒に公園に遊びに来る」ことは少なくなるため、A・B群の間にはギャップがあります。散布図ではこのように「ギャップ」という形で集合の区切りが見えることがあります。

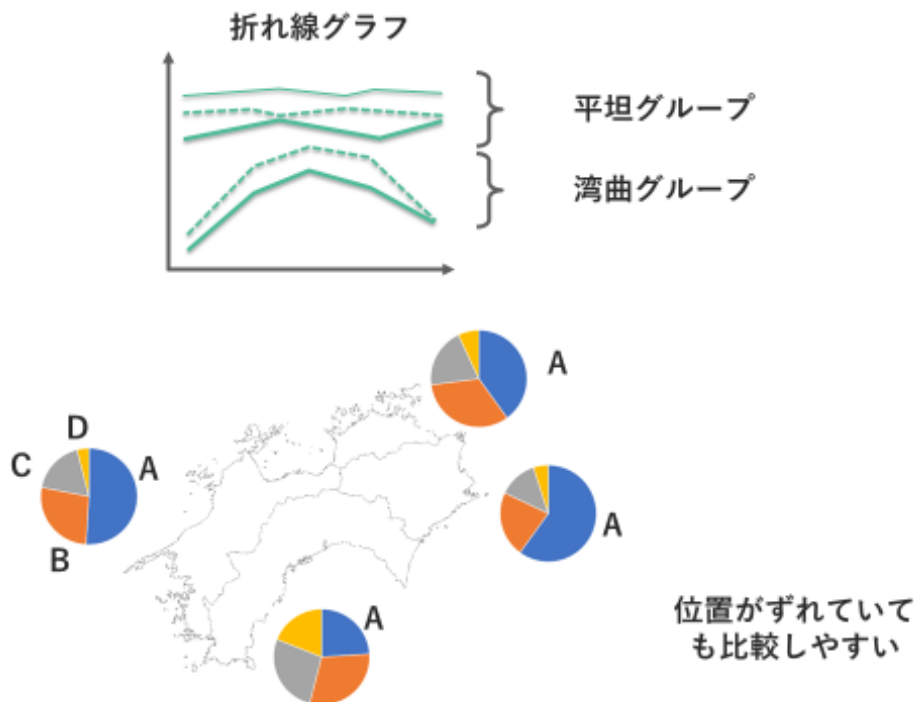
一方、散布図(2)はそのギャップが見えにくい例です。A群(○印)は右上がり、B群(×印)は右下がりになっていますが、AとBが重なっていてギャップがないので、もしすべてを○印で描いたらその区別はわかりません。このようなデータの場合、散布図以外の方法でA群・B群という集合は見つけておく必要があります。

左下のヒストグラムは、年齢のような連続的なデータを「階級」に区切ってその出現数(度数)を表すグラフです。「階級」自体が集合であり、度数が多いまたは少ない階級を見つけるために使います。

棒グラフはヒストグラムに似ていますが、データの並び順に注意が必要です。上図右下の棒グラフはいくつかの都道府県についての数値ですが、これを見てもすぐには「東北地方の県は数字が大きい」

ことには気がつかないでしょう。これは県を 50 音順に並べているからで、東北・関東・東海・・・などの地方別に並べればすぐにわかります。ヒストグラムの場合、「階級」は本来連続的なデータなので並び順は自動的に決まるのに対して、棒グラフでは並び順を自由に変えられる場合があり、その結果、集合がわかりにくくなるケースがあるので気をつけましょう。

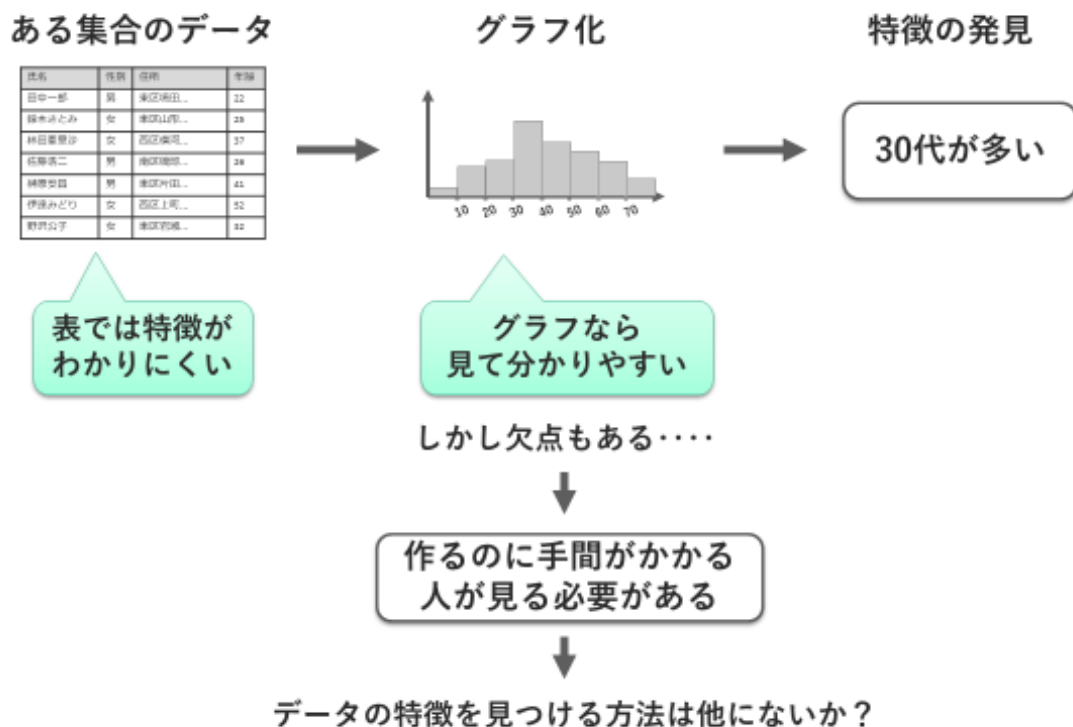
集合を見つけるためのグラフ(2)



「折れ線グラフ」で集合を探す場合もあります。上図の例では複数の折れ線グラフを載せていますが、おおまかに「平坦なグループ」と「湾曲しているグループ」という集合に分けられそうです。(ちなみに、平坦グループは熱帯地方、湾曲グループは温帯地方の年間の平均気温のグラフです)

円グラフには、位置がずれていても比較しやすいという特徴があります。たとえば上図が四国各県の産業別生産額（比率は架空のもの）だとしましょう。「A分野の比率が一番大きい県はどこ？」と聞かれた場合、円グラフならすぐにわかりますが、もしこれが棒グラフだとわかりにくくなります。棒グラフは縦か横にまっすぐ揃えて配置しないと比較しにくいのに対して、「角度」で大きさを表す円グラフは位置がずれていても比較しやすいからです。このため、円グラフは地図にひもつけてグラフを描くためによく使われます。

「グラフ」は便利だが万能ではない



ある集合のデータが表になっていたとしても、数字の並んだ表を読んでいくのは難しいものです。そこでグラフをつくとパッと見て分かりやすいので「30代が多い」といった特徴の発見が楽になります。

とはいえ、グラフは便利ですがもちろん万能ではなく、欠点もあります。それは、「作るのに手間がかかる」と、「人を見る必要がある」ということです。1枚だけなら「パッと見て分かりやすい」グラフでも、それがたとえば100枚もあったらいいかげんうんざりすることでしょう。たとえば1軒の雑貨店に約10000種類の商品があったとして、その商品別の売上推移をグラフにすると10000枚のグラフになります。特徴を見つけようと思っても、とても全部見ている時間はありません。

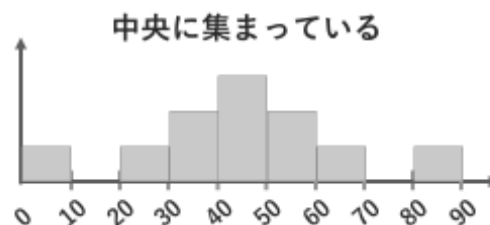
そこで、「人が見なくてもデータの特徴を見つけられる方法」があると便利ですが、そんな方法はないのでしょうか？

データの「バラツキ」を考えよう

データ1

10	30	40	40	50	50	50	60	60	70	90
----	----	----	----	----	----	----	----	----	----	----

最小値 10
 最大値 90
 平均値 50
 中央値 50
 最頻値 50

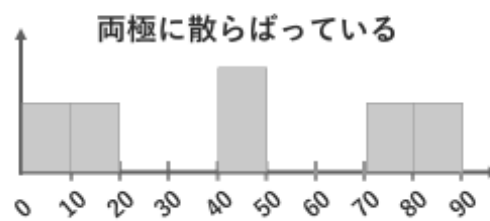


バラツキ度合が
違う

データ2

10	10	20	20	50	50	50	80	80	90	90
----	----	----	----	----	----	----	----	----	----	----

最小値 10
 最大値 90
 平均値 50
 中央値 50
 最頻値 50



ここで、上図中のデータ 1 とデータ 2 を比べてみましょう。どちらもデータ個数は 11 個で値は 10 ～90 の範囲です。ヒストグラムを作ってみるとデータ 1 は中央に集まっているのに対してデータ 2 は両極に散らばっている（バラついている）という大きな違いがあります。しかし最小値、最大値、平均値、中央値、最頻値を計算してみると完全に一致してしまうため、これらの代表値ではその違いはわかりません。

仮にこの数字が年齢であれば、データ 1 の客層を持つ店なら 50 歳前後を主な客と考えれば済みそうですが、データ 2 の客層を持つ店なら低年齢層と高齢者層も別に考えなければうまく行かないでしょう。

データ 1 と 2 の違いは「集まっているか、散らばっているか」です。実はこのようなバラツキ度合を計算するための特別な方法があります。

平均値との差を考える

データ1	10	30	40	40	50	50	50	60	60	70	90
平均値との差	-40	-20	-10	-10	0	0	0	10	10	20	40
その2乗	1600	400	100	100	0	0	0	100	100	400	1600
合計 ↓ 4400 データ数で割る → 400 分散 平方根 → 20 標準偏差											
データ2	10	10	20	20	50	50	50	80	80	90	90
平均値との差	-40	-40	-30	-30	0	0	0	30	30	40	40
その2乗	1600	1600	900	900	0	0	0	900	900	1600	1600
合計 ↓ 10000 データ数で割る → 909 分散 平方根 → 30.2 標準偏差											

しばらく無味乾燥な計算が続きますが、いったんは下記で説明する計算過程を地道にたどってみてください。

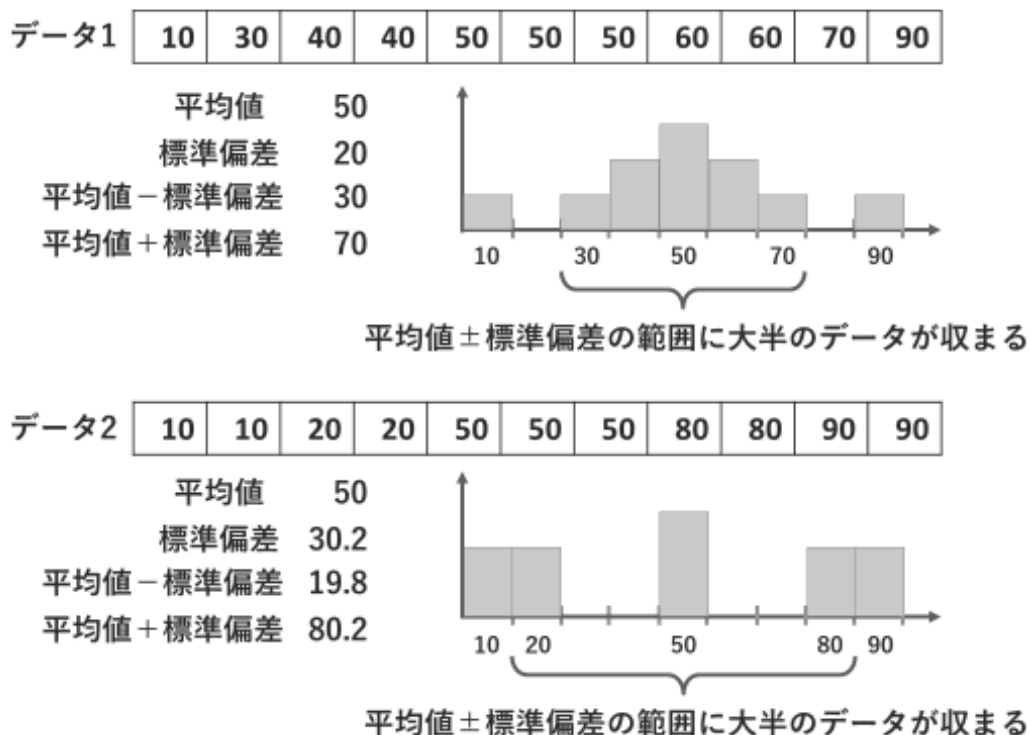
「データ1」が元になるデータで、個数は11、平均値は50です。次に、11個の各データと平均値との差を求めます。10, 30, 40・・・と平均値の差は-40、-20、-10・・・になるわけです。

次に、その結果（平均値との差）を2乗します。-40の2乗は1600、-20の2乗は400、以下同様に11個のデータすべてについて平均値との差の2乗を計算して、それをすべて合計すると4400になります。この「4400」をデータ個数11で割ると400になり、これを「分散」と言います。さらにその平方根（2乗すると400になる数）を計算すると20になり、これを「標準偏差」と言います。

「データ2」について同様の計算をすると、分散は約909、標準偏差は30.2になります。

「標準偏差」と「分散」という2つの専門用語が出てきましたが、ここでは「標準偏差」に注目してください。データ1と2の計算結果を比べてみると、中央に集まっているデータ1よりも、両極に散らばっているデータ2の標準偏差のほうが大きいですね。これが重要なポイントで、「標準偏差はデータのバラツキの大きさを表している」のです。

標準偏差が意味するもの



あらためて、ヒストグラムを使って標準偏差の意味を図示したのが上図です。

データ1と2の平均値は同じですが、標準偏差は違います。しかしどちらにしても、

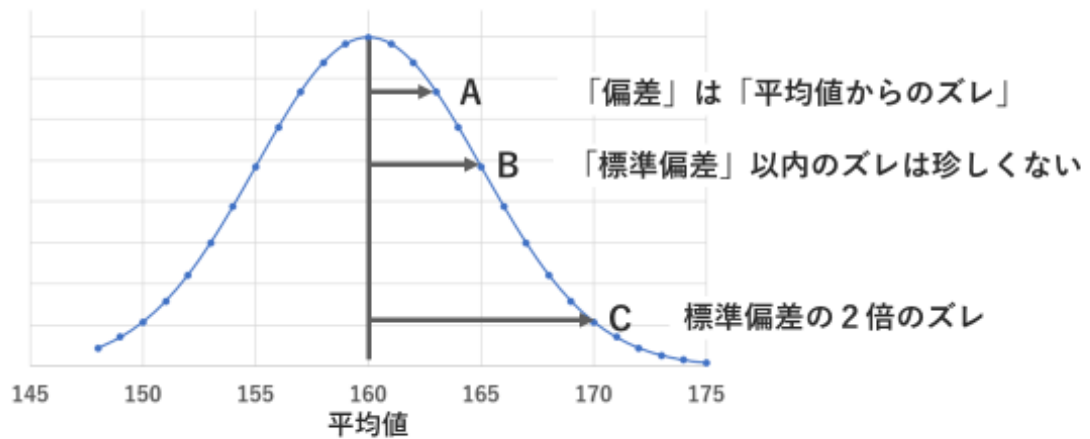
平均値±標準偏差の範囲に大半のデータが収まっている

ことに注意してください。前ページの手順で計算した「標準偏差」にはそんな性質があります。そのため、標準偏差が大きいほど「データのバラツキが大きい」ことを意味します。

たとえばサッカー選手しかいない集合に対してならばサッカー選手向けの商品企画だけで済みますが、野球、ラグビー、バレーボールなど他のスポーツの選手も混じっているなら、それぞれに向けた商品企画をしなければなりません。「バラツキが大きい」というのはそういうことです。「標準偏差」は、グラフを見なくても「計算」だけでバラツキの大きさを判断できる指標です。

「標準」と「偏差」の意味

平均値 160 標準偏差 5 のグラフ
(日本人成人女性の身長分布のイメージ)



数式では標準偏差は σ (シグマ) という記号で表記します

「標準偏差」という用語は難しそうに見えますが、

「偏差」は「平均値からのズレ」

「標準」は「普通に良くある範囲 (珍しくない)」

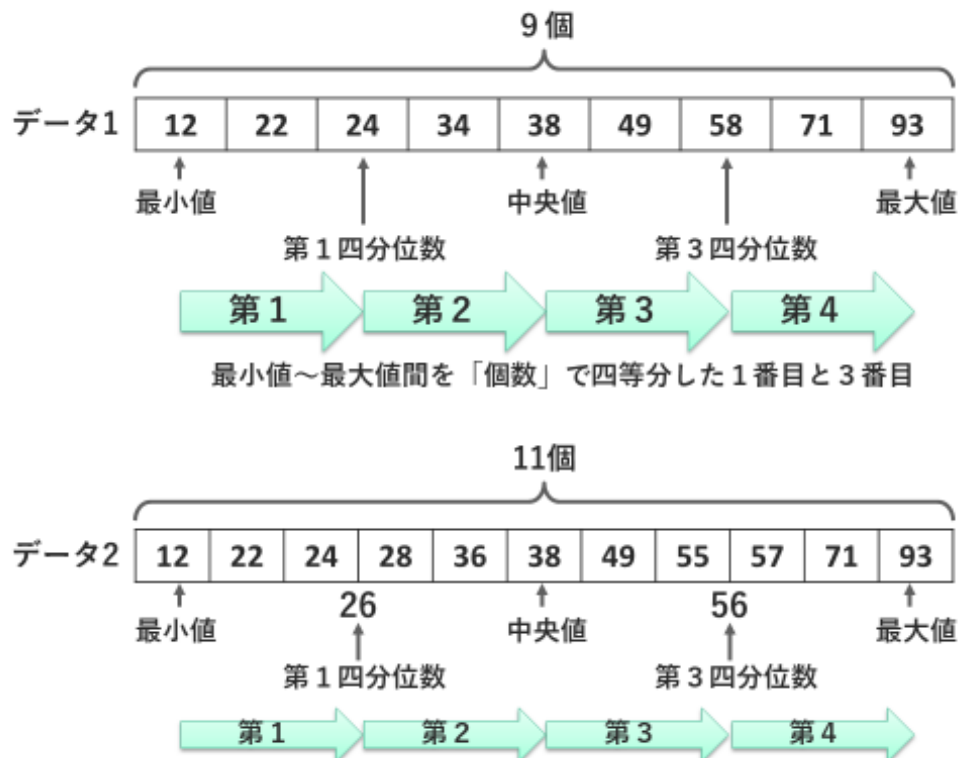
のような意味と考えてください。上図は、「平均値 160、標準偏差 5 の正規分布」と呼ばれるデータをグラフにしたものです。日本人の成人女性の身長分布をイメージしてもらえると良いでしょう。

A 点は 163cm ぐらいの人で、この場合「平均値からのズレ=偏差」は 3cm です。

B 点は平均値(160)+標準偏差(5)=165cm です。このぐらいの身長ならそれほど珍しくありません。これを越えるとだんだん「背が高いですね」というイメージが出てきて、C 点(170cm、つまり標準偏差の 2 倍のズレ)になるとかなり少なくなるのがわかります。

「標準偏差」は今後もよく使う非常に重要な概念なので覚えておいてください。数式の中で標準偏差を表す場合は σ という記号を使い、「シグマ」と呼びます。

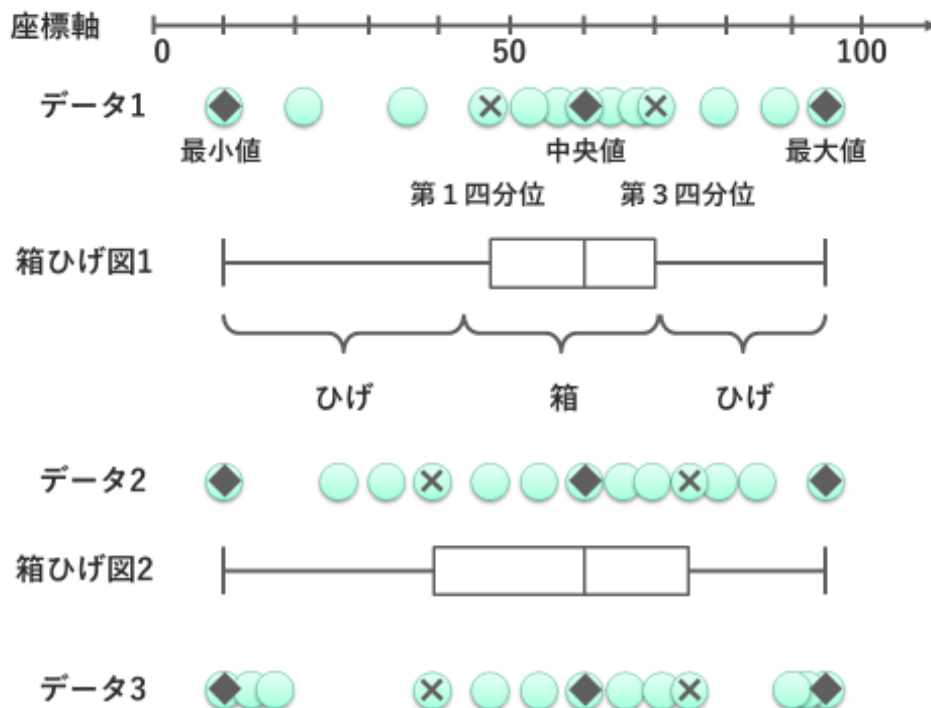
四分位数の考え方



データのバラツキを計算で示せる指標として、標準偏差の他に四分位数（しぶんいすう）というものも使われます。これは最小値～最大値間をデータの個数で4分割した各点の値のことです。上図のデータ1の場合、4分割した小さい方から1番目の「24」が第1四分位数、3番目の「58」が第3四分位数です。一般に「四分位数」というとこの第1と第3の2つを指します。「第2四分位数」は中央値、「第4四分位数」は最大値のことですので、これらについてはそのまま中央値、最大値と呼ばれます。また、最小値は「第0四分位数」に該当します。

11個のデータで構成されているデータ2のほうは、第1四分位の位置が24と28の間なので両者の平均をとって26、第3四分位は55と57の平均をとって56になります。

箱ひげ図



四分位数を用いて「箱ひげ図」というグラフを書く方法があります。

上図の最上段にあるのは、0 から 100 までの目盛りを刻んだ単なる座標軸です。

その下の「データ 1」は、その座標軸を使ってあるデータをプロットしたものです。最小値・中央値・最大値に◆印を、第 1・第 4 四分位数に×印を付してあります。

このデータを「箱ひげ図」法でグラフ化したのが「箱ひげ図 1」です。この図で次のようなことがわかります。

- おそらく中央値の周囲にデータが集中していて、最小値・最大値付近は少ない
- 最小値よりも最大値側に寄ったデータが多い

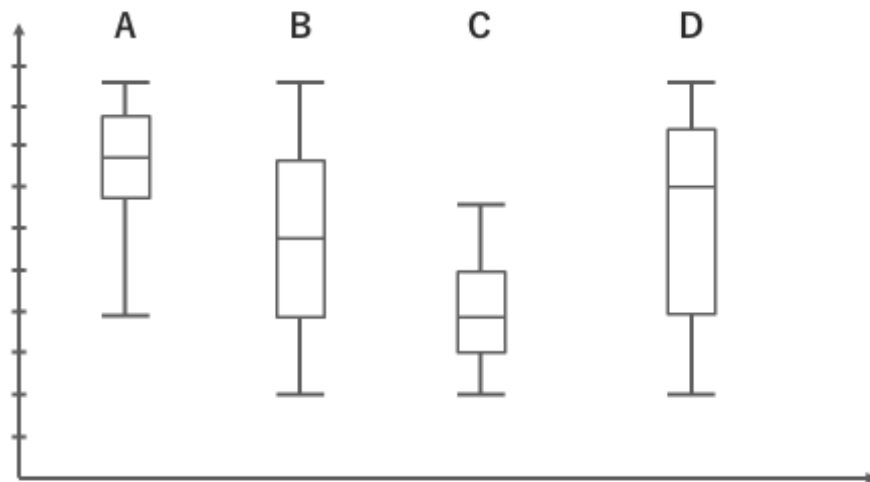
一方、その下の「データ 2」を基にした箱ひげ図 2 を見ると「箱」の範囲が箱ひげ図 1 よりも大きいので、最小・最大・中央値は同じでもデータ 1 よりもバラつきが大きいだろうということがわかります。

ただし箱ひげ図には欠点もあります。「データ 3」を使って箱ひげ図を描くと実は箱ひげ図 2 とまったく同じになりますが、データ 2 とデータ 3 を見比べるとずいぶん分布が違います。データ 3 では最小値・最大値付近にもデータがまとまっていることに注意してください。おそらくデータ 3 は最小値付近、中央値付近、最大値付近の 3 つのグループに分けて考えるべきです。箱ひげ図（およびその下になる四分位法）ではこのような違いが見えなくなってしまいます。

実は、箱ひげ図は中央値付近にだけ集中するタイプのデータを表現するのに向いた手法であり、データ 3 のように「複数の集中点があるデータ」のバラツキを表すのには不向きです。

箱ひげ図によるバラツキの表現

さまざまなスポーツ選手の身長分布のイメージ



A～Dまで、それぞれどんなスポーツをイメージしますか？

上図は、さまざまなスポーツの選手の身長分布を箱ひげ図で表したイメージです。

A は高身長に集中していますが低身長の選手も少しいる、これはバレーボールやバスケットボールなどに多いパターンです。

B は低身長から高身長までまんべんなくいるスポーツで、卓球や柔道が当てはまりそうです。

低身長側に寄った C は競馬や競艇など、体重制限のある分野でこの傾向があります。

D は B に似て低身長から高身長まで範囲が広いですが、中央値が高いので B に比べると高身長側が多くなっています。サッカーやアメリカンフットボールはこの傾向があります。ただし、アメリカンフットボールの場合はポジションによって身長の傾向がまったく異なり、「複数の集中点があるデータ」になるため、本来は箱ひげ図を使うのには向いていません。

確率とは？

「確率」とは、ある出来事が「起きる」と期待できる度合のこと



ある出来事が「起きる」と期待できる度合いのことを「確率」と言います。たとえば天気予報では「明日の降水確率は 80%」のように使われます。「降水確率 80%」と言えば「おそらく雨が降る」という意味ですし、「降水確率 0%」なら「絶対に降らない」という意味です。このように日常生活でも「確率」という言葉は使われます。

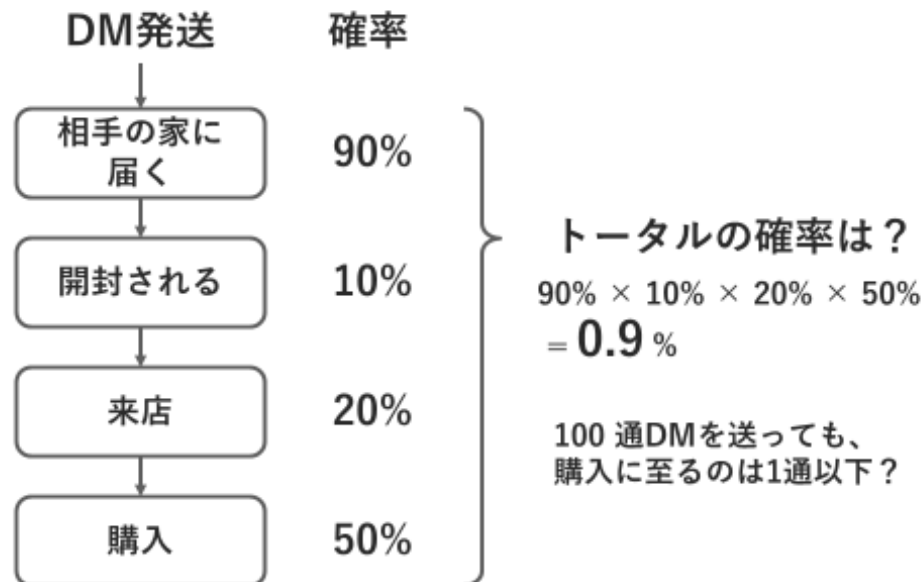
確率の中には分かりやすいものもあります。たとえば正しく作られたサイコロを 1 回振ったときには、1～6 までどの目が出る確率も $1/6$ です。

仕事をする上でもこの「確率」が重要な場面があります。たとえば見込み客に DM（ダイレクトメール）を送ったとして、その人が来店してくれる確率はどのくらいでしょうか？ もし DM を 1 通送るのに 100 円の費用がかかり、受け取った客の 10 人に 1 人が来店する（来店確率 = $1/10$ ）なら、来店客 1 人を集めるために 1000 円の費用がかかる計算なので、1000 円以上の利益が出る買い物をしてもらわないと採算が合いません。

DM に限らず、販売促進施策による来店確率は非常に変わりやすいので、十分考えて販促施策の手を打つ必要があります。

仕事の成否を左右する確率を考えよう

DM発送が購入に至るプロセスの確率を考える



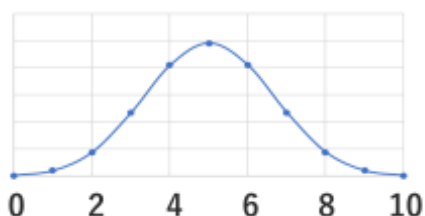
1つの販促施策が成果を上げるまでには複数の段階があり、それぞれに確率があります。たとえばDMを発送してそれが相手の家に届く確率、開封されて読まれる確率、興味を持たれて来店する確率、店頭で購入される確率がそれぞれ90%、10%、20%、50%だとすると、それを全部掛けたトータルの確率は0.9%となり、100通DMを送っても購入に至るのは1通以下ということになります。

この数字は具体的な内容によって大きく変わります。たとえば「DMが相手の家に届く確率」の90%は、最近会員登録された自社の顧客リストなら高めに出来ますし、古いリストや他社から購入したものなら下がります。「開封される」から後の確率には、たとえばDMのデザイン、顧客ロイヤルティ、企画内容が大きく影響します。「DMを打っても売上につながらない」という場合、どのポイントで確率が低くなっているのかを突き止めて、ピンポイントでそこに手を打つことが求められます。

確率の基本：二項分布

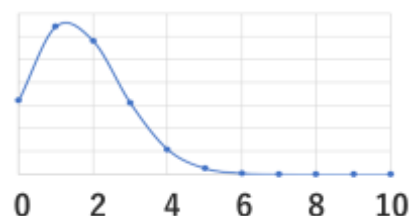
コインを10回投げたとき、
表がN回出る確率は？

表の回数	確率
0	0.0010
1	0.0098
2	0.0439
3	0.1172
4	0.2051
5	0.2461
6	0.2051
7	0.1172
8	0.0439
9	0.0098
10	0.0010



サイコロを10回振ったとき、
6の目がN回出る確率は？

6が出る回数	確率
0	0.1615
1	0.3230
2	0.2907
3	0.1550
4	0.0543
5	0.0130
6	0.0022
7	0.0002
8	0.0000
9	0.0000
10	0.0000



仕事上のプロセス、特に集客のようなものが「うまくいく確率」を見積もるのは非常に難しいものですが、単純なプロセスであれば確率は計算で正確に求められるものがあります。

たとえば10円玉のようなコインを10回投げたとき、表が出る回数は0回から10回まで考えられますがその確率は上図左側、同様にサイコロを10回振ったときに6の目がN回出る確率は上図右側です。

コインを1回振って表が出る確率は1/2ですので、10回振ったら5回が一番多くなりそうです。実際その通りなのですが、4回・6回も意外に多く出ることがわかります。

一方、サイコロを1回振って6の目が出る確率は1/6ですから、10回振るとその10倍で $10/6 = 1.66$ 回になりそうです。この数字は1よりも2に近いので、直感的には「2回出る確率が高い」ような気がしますが、実際には1回のほうが少し高確率で出ます。人間の直観が実際の確率と大きくズレるケースは珍しくありません。直観だけで確率を考えていると、経営上の判断を間違えがちです。少なくともこのような単純なプロセスの確率については、計算方法を知っておきましょう。今回例示した「表 or 裏」「6 or その他」のように「結果が2種類しかない操作を複数回繰り返すと一方の結果が何回出るか」を示す分布は二項分布と呼ばれていて、その確率は簡単な式で計算できます。

二項分布の計算式

例：コインを1度投げる

例：表が出る

1/2

ある操作を1回行って結果が真になる確率が p であるとする。
同じ操作を n 回行って、そのうち k 回で結果が真になる確率(X)はいくつか？

この答えは次の式で求められます。

$$\begin{aligned} X &= \binom{n}{k} P^k (1 - P)^{n-k} \\ &= \frac{n!}{k! (n - k)!} P^k (1 - P)^{n-k} \end{aligned}$$

(Excel では BINOM.DIST 関数で計算できます)

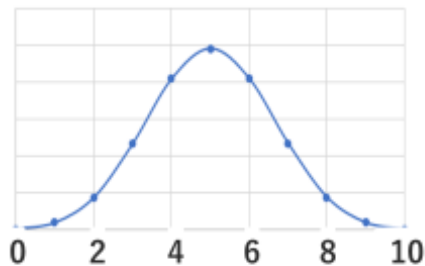
二項分布は「結果が二種類しかない操作を独立して複数回行う」場合を想定しています。「結果が二種類」のところという「結果」は実際には Yes/No、表/裏、真/偽 などの名前と呼ばれることが多いですが、「6の目が出るか否か」のような想定もその結果は「真/偽」と見なせるので適用できます。

たとえば「サイコロを1000回振って5か6が100回出る確率」ならば、 $n=1000$, $p=1/3$, $k=100$ とすれば計算できます。この式を手で計算するのは大変ですが、現在は表計算ソフト等でやれるので活用しましょう。例えば Excel なら BINOM.DIST という名前の関数でこの2項分布を計算する機能があります。

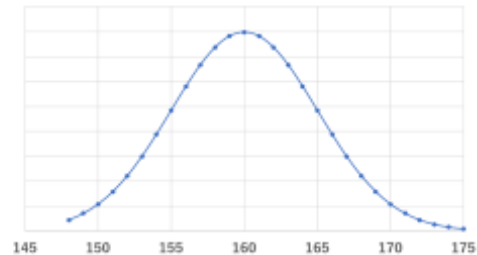
二項分布と正規分布は似ている

二項分布のグラフは、以前出てきた正規分布のグラフによく似ています。

二項分布のグラフ



正規分布のグラフ



実は二項分布を「正規分布」に近似する方法が実務的によく使われます。

正規分布のグラフを表す式

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

μ は平均値
 σ は標準偏差

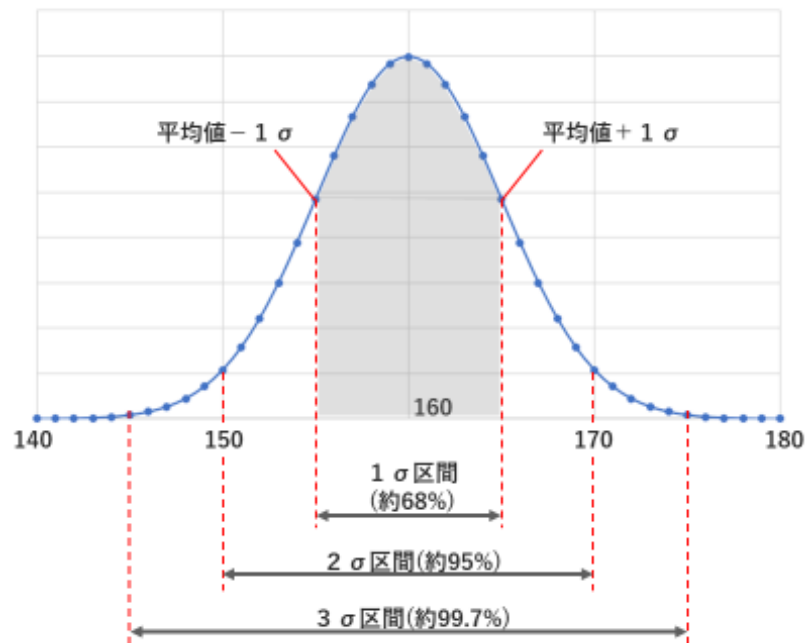
標準偏差の説明(p.16, 「標準」と「偏差」の意味)の中で「正規分布」という用語が出てきたのを覚えていませんか？ 正規分布というのは平均値と標準偏差で決まる関数で、数式では上図最下部の式で表され、山形のグラフになります。二項分布とは違うものなのですが、並べてみるとよく似ているのがわかります。

実は、二項分布は n (操作を試行する回数) が多くなるほど正規分布に近づいていく性質があります。さらに、正規分布のほうが二項分布よりも実務的に扱いやすい場合が多いので、大まかな傾向が分かればいい場合は二項分布ではなく正規分布を使って確率の計算を初めとする判断を行うことが多いのです。

正規分布とシグマ区間

シグマ区間：平均値の左右に標準偏差の整数倍の幅を持った区間

正規分布のグラフ



「コインを複数回投げて表が出る数」は、投げる回数が多くなるほど正規分布に近い結果が得られます。他にも、(人間を含む) 動物の身長や体重、雨粒の大きさなど、自然界のランダムな現象は正規分布を示すことが多いので、ある分野について「どのぐらいの範囲のデータがありうるか」という見当をつけるには正規分布を知っておくことが役に立ちます。

たとえば、「日本人の成人女性の平均身長は 160cm、標準偏差は 5cm」だったとしましょう（この数字は仮定です）。その場合、身長 170cm の女性は何人に一人ぐらいいると考えればよいのでしょうか？ 少ないのは間違いありませんが、具体的には 20 人に一人なのか、それとも 100 人に一人あるいはもっと少ないのでしょうか？ その答えによって、経営上の判断が変わってることがあります。たとえば「170cm 以上の人が 50 人に 1 人よりも少ないなら、そのサイズの商品を作っても売れる見込みがないから作らない。それより多いなら作る」のような基準で商品企画をしている会社にはこの答えが必要です。そこで役に立つのが、

平均値を中心とする 1 σ 区間に約 68%、2 σ 区間に約 95%、3 σ 区間に約 99.7%

のデータが集まるという、正規分布の特徴です。170cm はちょうど 2 σ 区間の境界ですので、150~170cm の範囲に 95% の人が収まっているはずで、ということは残り 5% つまり 20 人に一人はそこからはみ出します。これは低いほう（150cm 以下）と高いほう（170cm 以上）の両方を含むの

で、高いほうだけであれば半分の「40 人に一人」程度はいるでしょう。したがって「日本人の成人女性で 170cm 以上の方は 50 人に一人以上いそうだ」ということが推定できるため、その商品を作ろう、という方向で経営判断ができます。

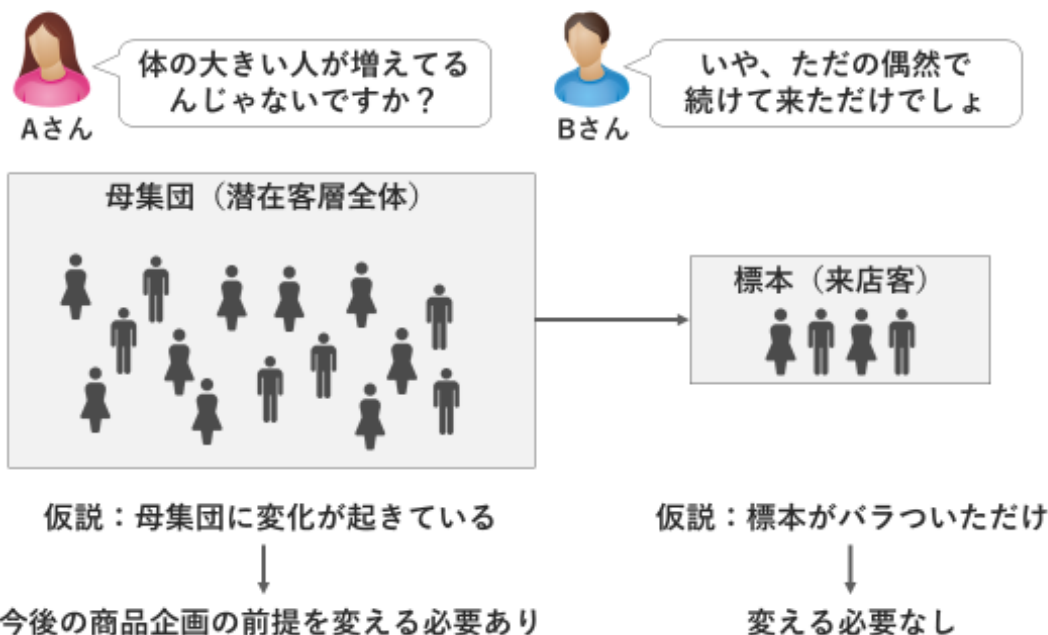
このように「平均値と標準偏差がわかれば、どの範囲にどれぐらいのデータがありそうかを簡単に推定できる」のが正規分布のメリットの一つです。1 σ 、2 σ 、3 σ と「標準偏差の整数倍」単位で境界を考えることが多いので、その範囲の概数とともにこれらの用語を覚えておきましょう。

名前	定義	含まれるデータ数
1 σ 区間	平均値 $\pm$ 1 σ の範囲	約 68%
2 σ 区間	平均値 $\pm$ 2 σ の範囲	約 95%
3 σ 区間	平均値 $\pm$ 3 σ の範囲	約 99.7%

第5章. 仕事と集合Ⅱ

「珍しい出来事」をどう解釈する？

ある洋服店である週に、ふだんはなかなか売れない商品(例：XXLサイズのシャツ)が連続して売れました。これは何を意味するのでしょうか？



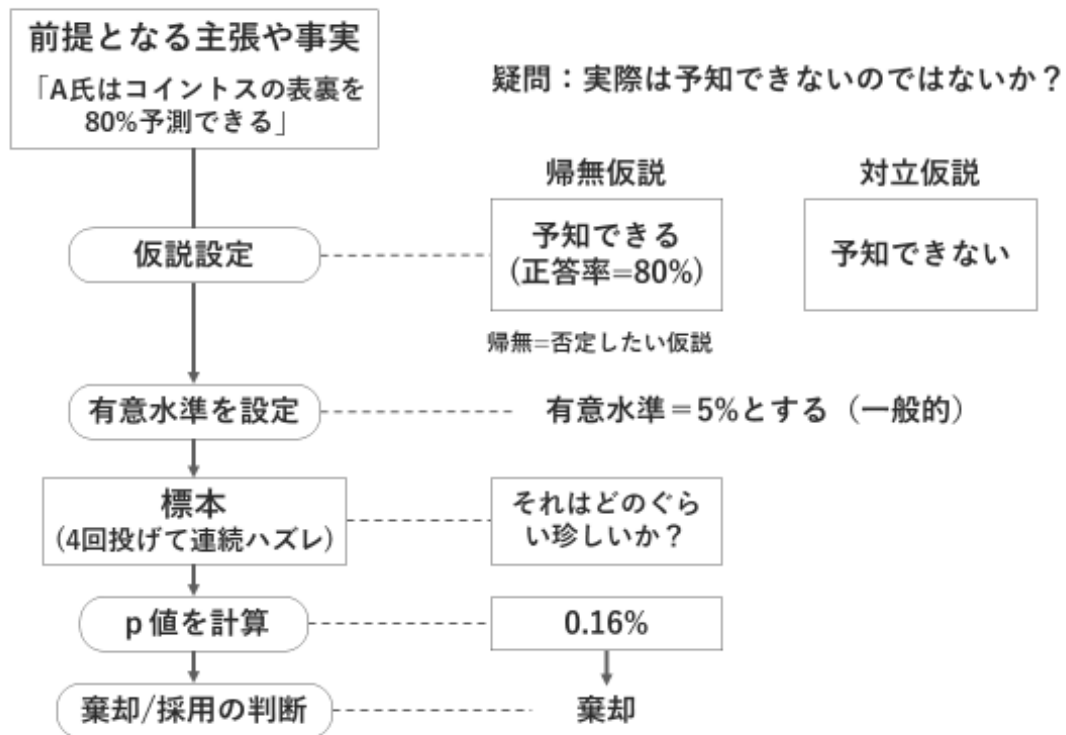
あるときある洋服店で、ふだんはなかなか売れない XXL サイズのシャツが連続して売れたとします。これに対して A さんは「体の大きい人が増えているんじゃないか」と考えたのに対して B さんは「ただの偶然でしょ」と意見が分かれました。

お店への来店客というのは、「母集団」に対する「標本」と考えることができます。母集団は「そのお店に来る可能性がある客層全体」です。通常はその店の近くに住んでいる、通勤している、近くに遊びに来ることがある人が該当します。一方、標本はその週に実際にその店に来た客です。

A さんの意見が正しいなら、母集団に変化が起きているということですから、今後の商品企画の前提を変えて「大きなサイズを増やす」などの手を打つ必要があります。B さんの意見が正しいなら、特に何も変える必要はないでしょう。

こんなときに、ただの「カン」ではなくデータを分析してどちらの意見がより「確からしい」かを判断することはできるでしょうか？ このような目的に対して使えるのが「仮説検定」という統計的なデータ処理の方法です。「検定」とは「正しいかどうかを確かめる」ことをいうので、「仮説検定」とは「仮説が正しいかどうかを確かめる」一般的な方法です。

仮説検定の基本



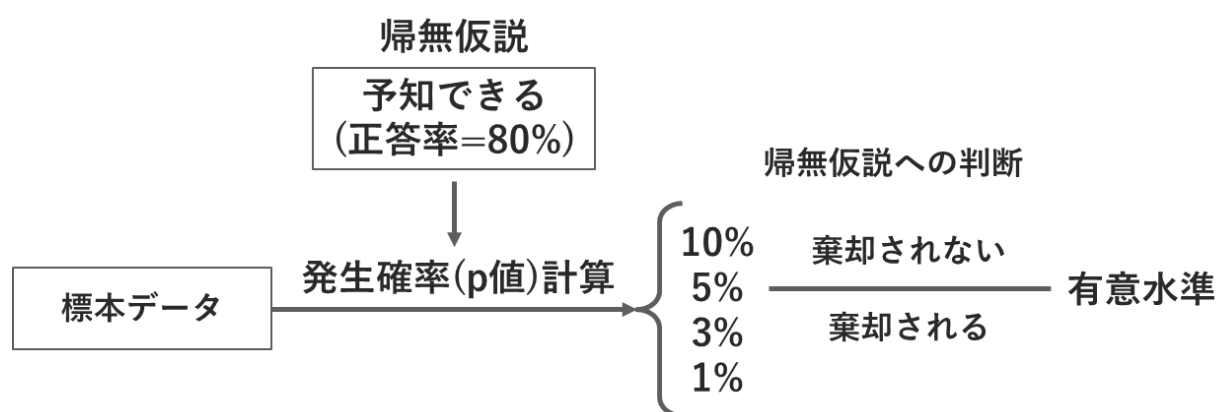
仮説検定の基本を知るため、わかりやすい極端な例から考えましょう。ある人（A氏）が「私はコイントスの表裏を 80%予測できる」と予知能力があることを主張していたとします。当然、誰もが「出来るわけじゃないか」と疑問を持つはずですので、ここで仮説検定を行います。

仮説検定を行う対象は何らかの「前提となる主張や事実」です。たとえば「私はコイントスの表裏を 80%予測できる」というのは A 氏の主張です。他には「文部科学省によると、男子の小学校卒業時の平均身長は XXcm、標準偏差は YYcm である」というようなものなら「過去に計測された事実」です。過去の事実は現在までの間に変わることがあるので、いったんは「事実」と確かめられたものでも、時間が経つなどして状況が変われば仮説検定の対象になります。

まずはその「主張や事実」に対して仮説を設定します。このとき設定する仮説には「帰無（きむ）仮説」と「対立（たいりつ）仮説」の 2 種類があります。「帰無」というのは「おそらく間違っている」という意味あいであらえましょう。つまり「無に帰る」、ざっくりばらんに言えば「無い無い、それ絶対ありえない」というような種類の仮説が帰無仮説です。したがって「コイントスを 80%予知できる」というような荒唐無稽な仮説が該当します。それに対する「おそらく正しい、普通の意見」を「対立仮説」と言います。今回は「予知なんかできるわけがない」という、いかにも普通の常識的な意見が対立仮説です。この段階ではどちらもただの「仮説」であって、正しいとも間違っているとも言えません。

今回は極端な例を出しているのので「帰無と対立」の違いがわかりやすいですが、現実には仮説検定を行う場面ではこの差ははっきりしないときもあります。たとえば「前提となる主張」が「地球は温暖化している」というものだったらどうでしょうか。これに対しては人によって「している」「していない」と意見が分かります。基本的に、帰無仮説には「自分が否定したい仮説」を選びます。自分が「地球は温暖化している」という意見を持っているなら「温暖化していない」を帰無仮説に選び、逆の意見なら逆にします。したがって意見が分かれる問題の場合は人によって帰無と対立の選択が逆になります。

次に「有意水準」を設定します。有意水準の位置づけは下記の図で説明します。



帰無仮説を立てた後、「標本データ」の「発生確率（p 値）」を計算します。p 値は「帰無仮説が正しいと仮定したとき、標本データの事象が発生する確率」のことです。たとえば「正答率=80%」という帰無仮説のもとで 1 回コイントスを行い、予測がハズレた場合の p 値は 20%（80%正答できるはずなのにハズレたので、残り 20%に該当する事態が起きた、と考える）。2 回続けてハズレた場合は 20%の 2 乗ですから p 値=4%です。これに対して、同じ 1 回ハズレ、2 回ハズレでも、帰無仮説が「正答率=60%」というものだった場合の p 値はそれぞれ 40%、16%です。つまり標本データが同じでも、帰無仮説が異なれば発生確率（p 値）の計算結果は異なります。

標本が同じでも帰無仮説が違えばp値は異なる

標本データ	帰無仮説 正答率=80%	帰無仮説 正答率=60%
1回ハズレ	p値=20%	p値=40%
2回連続ハズレ	p値=4%	p値=16%

ある帰無仮説をもとに具体的な標本データの p 値を計算したあと、それが「有意水準」より高いか

低いかを判断します。もし有意水準より低ければ、

**帰無仮説が正しいという仮定の下では、非常に低い確率でしか発生しない
標本が得られてしまったので、おそらく帰無仮説は正しくない**

というロジックで帰無仮説を「棄却」します。「棄却」とは捨て去ることですので、帰無仮説は「無かったこと」になります。

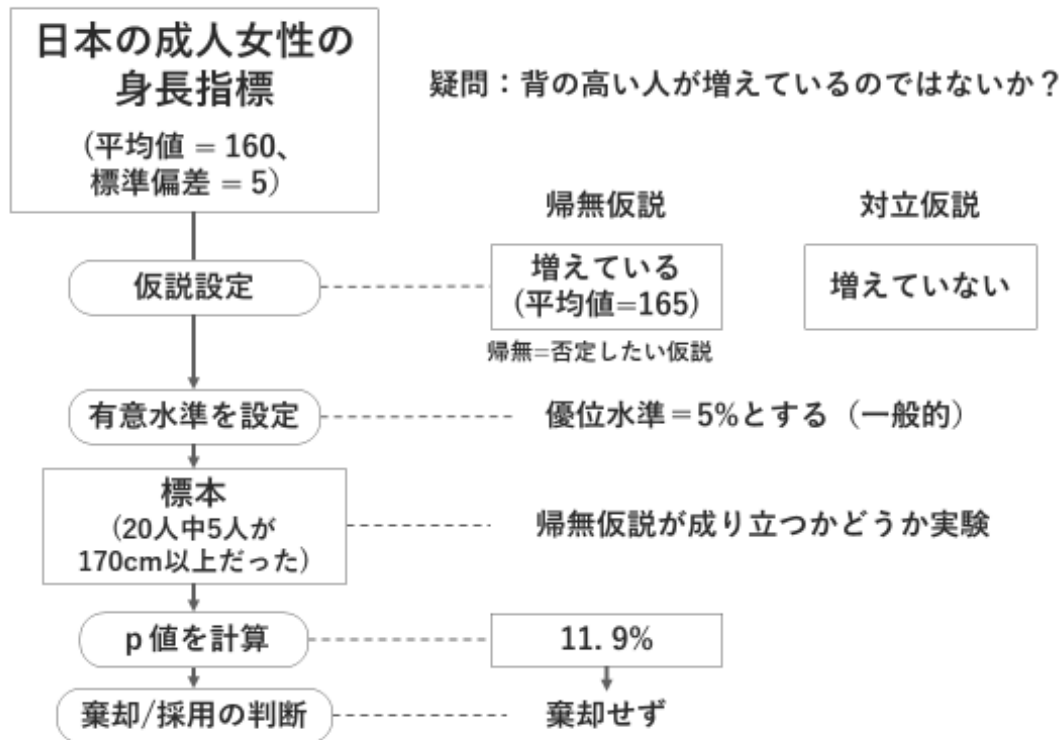
このように「帰無仮説を捨てるかどうか」を左右する重要な数字が「有意水準」です。一般的には 5% に設定されますが、人命が関わるような重大な問題については 1% がよく使われます。有意水準は勝手に決めてよいものではなく、問題の性質に応じて適切に選ばなければなりません。

【注意！】 p 値が有意水準より高ければ帰無仮説は棄却されませんが、これは棄却するには証拠が不十分というだけで、「帰無仮説が正しい」という意味ではありません。たとえば p 値=10%つまり 10 回に 1 回というのは偶然でも十分起こりうる数字です。「俺はコイントスで必ず表を出せる」と宣言した人がいたとして、その人が実際に 3 回連続して表を出したとしても「必ず表を出せる」と信じる人はあまりいないでしょう。3 回連続偶然に表が出る確率は $1/8$ あるので、この程度では「必ず出せる」能力の証明にはなりません。4 回連続でもまだ不十分で、多くの人は 5 回連続つまり $1/32$ ぐらいになってようやく「ひょっとして本当かも？」と信じ始めます。

さて、有意水準を設定した上で標本を採るのは、「帰無仮説が成り立つかどうか」を実験することです。4 回連続でハズレた場合の p 値は 0.16% となり、有意水準 5% より小さいので帰無仮説は棄却され、対立仮説が採用されます。

以上が「仮説検定」の基本ですが、次はもう少し複雑で現実的なケースを考えましょう。

ある洋服店での仮説検定



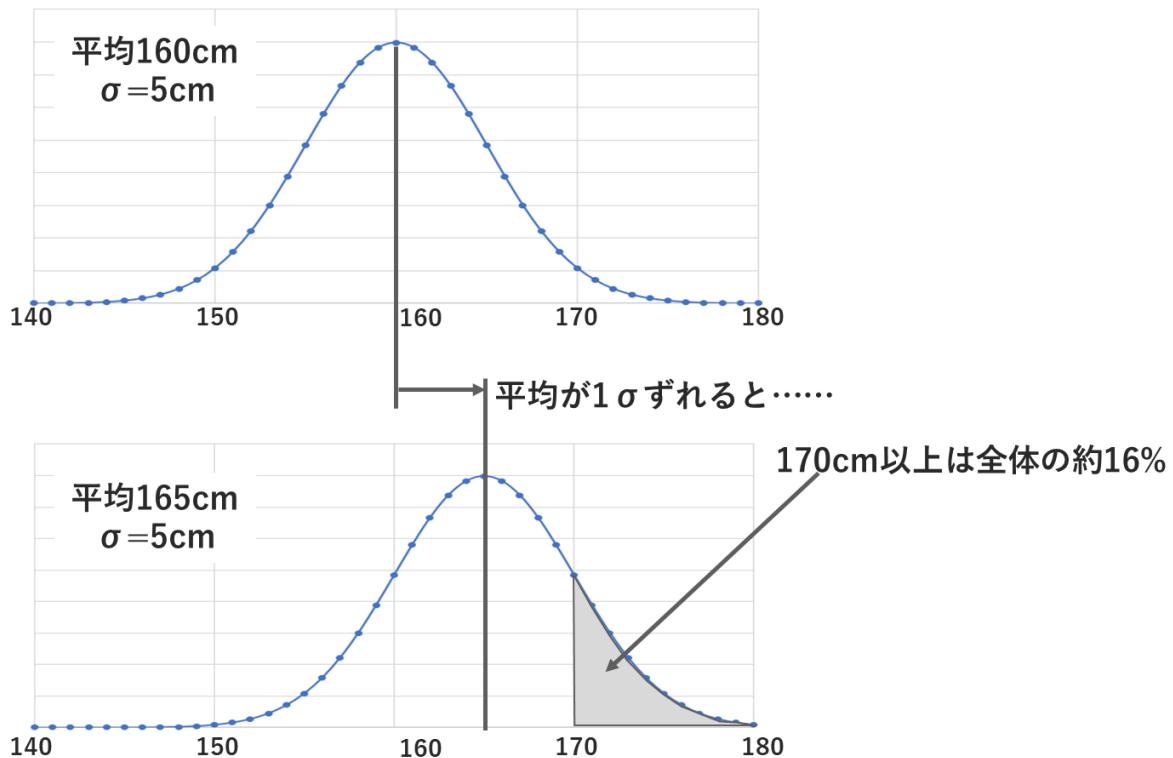
ある洋服店で短期間に大きなサイズが売れたために、「背の高い人が増えているのではないかな？」という仮説を立てた場合の仮説検定の流れです。

まず、「前提となる主張や事実」としては、「日本の成人女性の身長指標」として一般的な数字を使います。ここでは、話を単純化するために平均値=160cm、標準偏差=5cm としておきます。一方、疑問は「背の高い人が増えているのではないかな？」というものでしたが、成人の身長はそう簡単に伸び縮みするものではないので、ここでは「いやそれはただの偶然で、実際は増えていないだろう」と反論する立場で考えてみましょう。そこで、「増えていない」を対立仮説、「増えている」を帰無仮説とします。

「増えている」という帰無仮説の枠内で「平均値=165」と書かれているのに注目しましょう。増えているとしてそれは何 cm なのかというこの数字を決めないことには p 値の計算ができないので、決めておく必要があります。ここでは+5cm つまり 1σ 分増やしていますが、σ 単位でなければならぬわけではなく、p 値の計算ができさえすれば何でもかまいません。

有意水準は一般的な 5% に設定しましょう。標本として過去 1 週間の採寸データを調べたところ、20 人中 5 人が 170cm 以上だったとしましょう。帰無仮説 (平均値=165cm、標準偏差は変わらず 5cm) が正しいとしたら、標本がこの分布を示す確率 (p 値) はいくらになるのでしょうか？

平均値が 160cm から 165cm へと 1σ ずれると、身長 170cm 以上は全体の約 16%になります。これは、 1σ 区間=160~170 が 68%で、170cm 以上はその残り 32%のうち右側（高身長側）半分だからです。



それがわかれば、20 人のうち 5 人が 170cm 以上である確率は二項分布の式で計算できます。「1 人の客」を 1 回の試行と考え、1 回あたり 16%の確率で真になる試行を 20 回繰り返して 5 回真になると考えればよいわけです。これを計算すると p 値は 11.9%で、優位水準 5%よりも大きいため、帰無仮説「平均身長が 165cm に増えている」は棄却できません。

ちなみに帰無仮説と対立仮説を逆にして、「平均身長は 160cm のまま、増えていない」を帰無仮説にして p 値を計算すると 0.01%となり、優位水準より極めて小さいため棄却されます。したがって、この検定によれば、「20 人中 5 人が 170cm 以上」であった場合、その店の客層の平均身長は具体的な数値は不明ながら 160 よりは上である可能性が高いと考えられます。

なお、実際には「標本」に偏りがある状態で仮説検定を行うと非現実的な答えが出がちなので注意が必要です。たとえば標本が「ある店のある 1 週間の来店客」のようなデータだと、たまたまその週に近所のホテルに泊まっていたバレーボールの選手団が買いに来ていた、といったバイアスを排除できません。

片側検定と両側検定

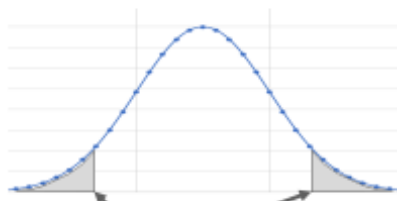


今週からA商品をレジ前にも置いたんですけど、売上増えてますか？



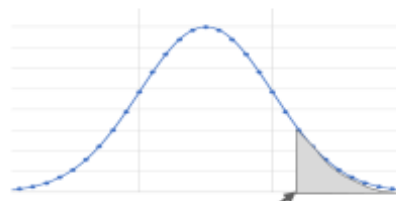
前週までの売上と比較してみよう

A商品を一般売り場から撤去しレジ前だけに置いた場合、売り上げが減る可能性もあるので両側検定を行う



両側合わせて5%の範囲に有意水準を設定（片側2.5%ずつ）

A商品を一般売り場に加えてレジ前にも置いた場合売り上げが減る可能性はほぼゼロなので片側検定を行う



片側だけで5%の範囲に有意水準を設定（減少側は考慮しない）

スーパーマーケットなどの業態では、「レジ前商品」というジャンルがあります。店内を回って目的の商品を探し終えた後、会計のためレジ前に並んでいるときに目に留まると衝動買いされやすい商品をレジ前に置きます。通常、レジ前商品はレジ前「にも」置くので、それで売上が下がることはほとんど考えられません。このような場合は減少側を考慮に入れる必要がないので、「売上が上がったかどうか？」と、片側だけを判定します。

一方、商品を一般売り場から撤去してレジ前に置く場合はかえって売上が下がることもあり得ます。そのような場合は両側を判定しなければなりません。

【両側検定】

正規分布の下位に外れた場合、上位に外れた場合の両方に有意水準を設定する。

【片側検定】

正規分布の下位・上位のどちらか片方にだけ有意水準を設定する。

有意水準は通常 5%や 1%で設定します。よく使われる「5%」というのは両側検定の場合で平均値± 2σ を外れた範囲です。片側検定の場合、同じ 5%でも片側検定だと片側だけで 5%の範囲を有意と判定するため、両側検定よりも甘い有意条件を設定していることになるため注意してください。

第6章. 仕事と集合Ⅲ

A店でよく売れているのはなぜだろう？

コンビニのように支店をいくつか持つチェーン店のうち1店で、Xという商品がよく売れているそうです。なぜA店でだけ売れているのでしょうか？

各店舗の基礎データとX商品の売上

	A店	B店	C店	D店		相関係数
X商品売上	100	80	82	75		
来店客数	4200	4300	4200	4100		0.19
客単価	670	660	680	620		0.56
年齢：平均値	40	42	41	39		-0.04
年齢：標準偏差	12	14	14	11		-0.08
年齢：中央値	39	40	42	37		0.11
駐車場の広さ	3	1	2	0		0.91
商圏人口	3300	3200	2900	3400		0.06

基礎データを見てもA店でXが
売れる理由はよくわからない

しかし、相関係数を
取ってみると・・・

コンビニのように支店をいくつか持つチェーン店のうち1店で、Xという商品がよく売れているそうです。なぜXでだけ売れているのでしょうか？

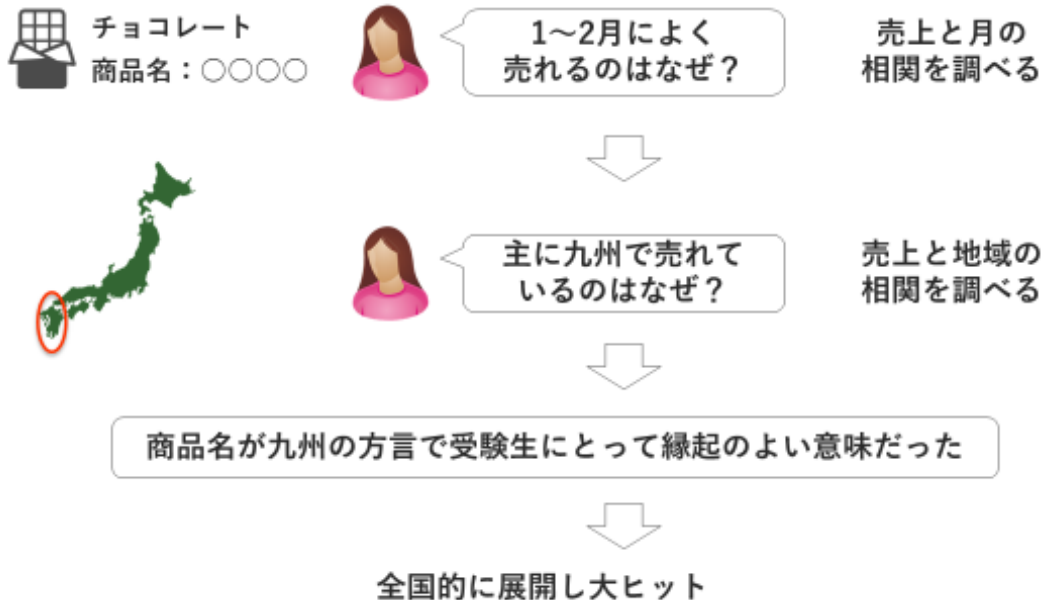
来店客数や商圏人口といった基礎データを見ても、A店でだけ売れる理由はよくわかりません。

「特定の商品が一部の店でだけよく売れる」というケースはよくありますが、食品や雑貨、衣料品などを売っている普通の小売店には何千種類もの商品があり、よほど目立つものでなければ、特定の商品が売れた理由を調べることはないでしょう。そもそも、上図の表のように売上がA店の100に対して他店は80前後という程度では、「A店でよく売れている」ことに気づいてさえいないことも多いものです。

しかし、もしその理由が分かれば適切な手を打ってB～D店でも売れるようになるかもしれません。実はそんな「なぜ売れている？」という疑問への答えを探るためによく使われるのが相関分析という手法です。上図右側に「相関係数」という列がありますが、これは「X商品売上」と「来店客数、客単価、年齢・・・」などの基礎データにどのくらい「関連性があるか」を計算した数値です。相関係数について詳しくは後述しますが、見ると「駐車場」の項目で0.91という最も大きな数字が出ています。となるとX商品は駐車場が多い店で売れる＝車で来店する客が買っているのかもしれません。だとしたら、他にも駐車場の多い店を中心に販促に力を入れることで売上を増やせる可能性があります。このような手がかりを見つけるのに役立つのが相関分析です。

異変に気づけばヒットにつながる

あるお菓子メーカーのチョコレートがなぜか1~2月によく売れている
ことに気づいた担当者が調べてみると・・・

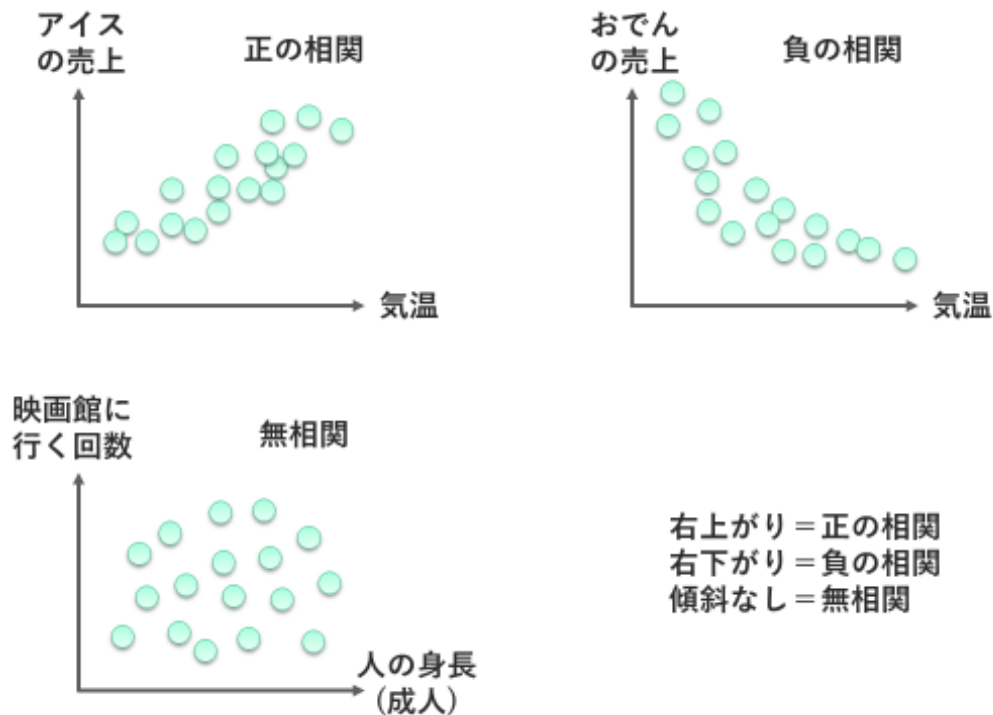


あるお菓子メーカーのチョコレートが特別な施策もしていないのになぜか1~2月によく売れているという現象がありました。担当者が不思議に思ってさらに調べてみるとその売上のほとんどが九州であることがわかりました。さらに調べてみると、商品名が九州の方言で受験生にとって縁起の良い意味になっていたために、受験シーズンに縁起物として大量に買われていたことがわかりました。

その後、そのメーカーはそれをヒントに商品パッケージを変え広告キャンペーンも変えて全国的に展開した結果大ヒット商品となりました。

しかしその大ヒットのきっかけは九州でのちょっとした異変でした。商品別の売上と月の相関を調べる、地域との相関を調べる、などの方法で、そんな「ちょっとした異変」に気づくことができます。「相関係数」は簡単な計算で出せますから、生の数字だけでは見逃してしまいそうなちょっとした異変を見つけるために活用できます。

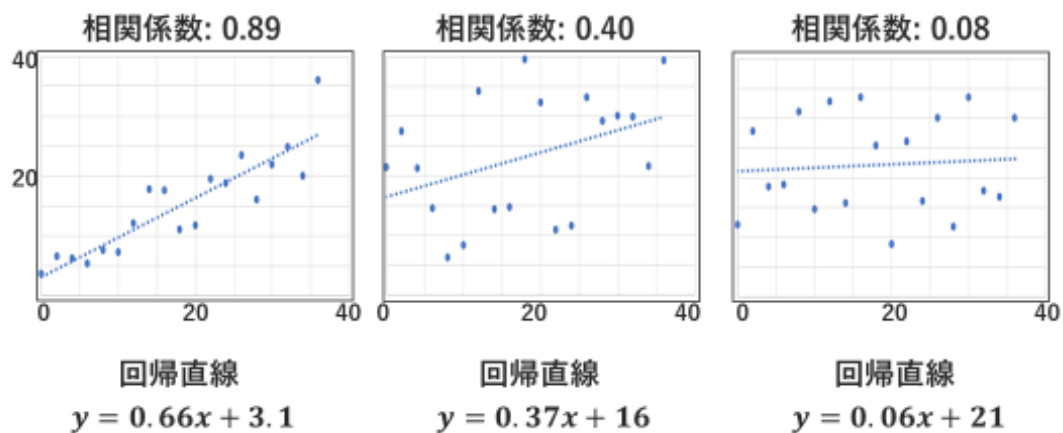
相関分析



「相関」とは「相互に関連がある」という文を略した用語です。たとえば「夏の暑い日にはアイスクリームがよく売れる」ことから、「アイスの売上と気温の間には関連（相関）がある」と言えます。実際、ある店の一か月間の記録をもとに、一日の平均気温を横軸に、その日のアイスの売上を縦軸にとって散布図を描くと上図左上のようになりました。右上がりの傾向がはっきりとみて取れます。このように、ある変数（気温）が増えると他方の変数（アイスの売上）も増える、というような関係を「正の相関」と言います。逆に寒い時期のほうがよく売れる「おでん」について同じことをすると上図右上のように右下がりの傾向が見えます。この場合は「負の相関」があると言えます。

一方、人の身長（成人）とその人が映画館に行く回数にはほとんど関係がない（無相関である）と考えられます。このような場合は上図左下のように散布図では面上に広がるなど、傾斜のない形になります。

相関係数・回帰式



相関係数は-1～1の値を取る

-1または1に近いほど相関が強い。0は無相関

ばらばらの点の動きに最も近い一次関数が回帰直線

「2つの変数の間に関連がある」といっても、ではどの程度の強さの関連なのか、という「程度」の差はあります。その相関の強さを数値化したものが「相関係数」です。

上図では3つの散布図の例を示していますが、左側のはっきりとした右上がりの傾向が見えるのに対して中央は「右上がりと言われればそう見える」程度、右側は「上がっても下がってもいないように見える」グラフです。これらの散布図について「どの程度相関が強いを示す相関係数」をそれぞれ計算すると 0.89, 0.40, 0.08 になります。相関係数は±1の範囲の値をとり、1に近いほど強い正の相関、-1に近いほど強い負の相関、0に近いほど無相関であることを示します。0.89は+1に近いのではっきりとした右上がりが見えるのに対して、弱い相関だと散布図でも右上がりの形が見えにくいものです。実際、中央(0.40)と右(0.08)の散布図を比べて見ても、中央のほうが相関が強いとはわかりにくいですね。しかし数字であれば0.40と0.08はハッキリとした違いなので、散布図ではわかりにくい「弱い相関」も発見することができます。

一方、点線で引いた「回帰直線」は、散布図の点の分布にできるだけ近くなるように1次関数の線を引いたものです。左側の散布図の中の回帰直線は「点」の動きと大筋合っているとと言えるでしょう。相関の強いデータに関してはこのように回帰直線を引くことで「明日は気温25度の予想だからこのくらい売れるだろう」といった予想の道具として使えます。相関が弱くても回帰直線を引くこと自体は可能ですが、中央や右のグラフで分かるとおり、その場合は回帰直線が現実と合わないケースが多発します。

説明変数と目的変数

問：AステーキよりもBステーキの値段が高いのはなぜ？
答：Bステーキのほうが大きいいからです



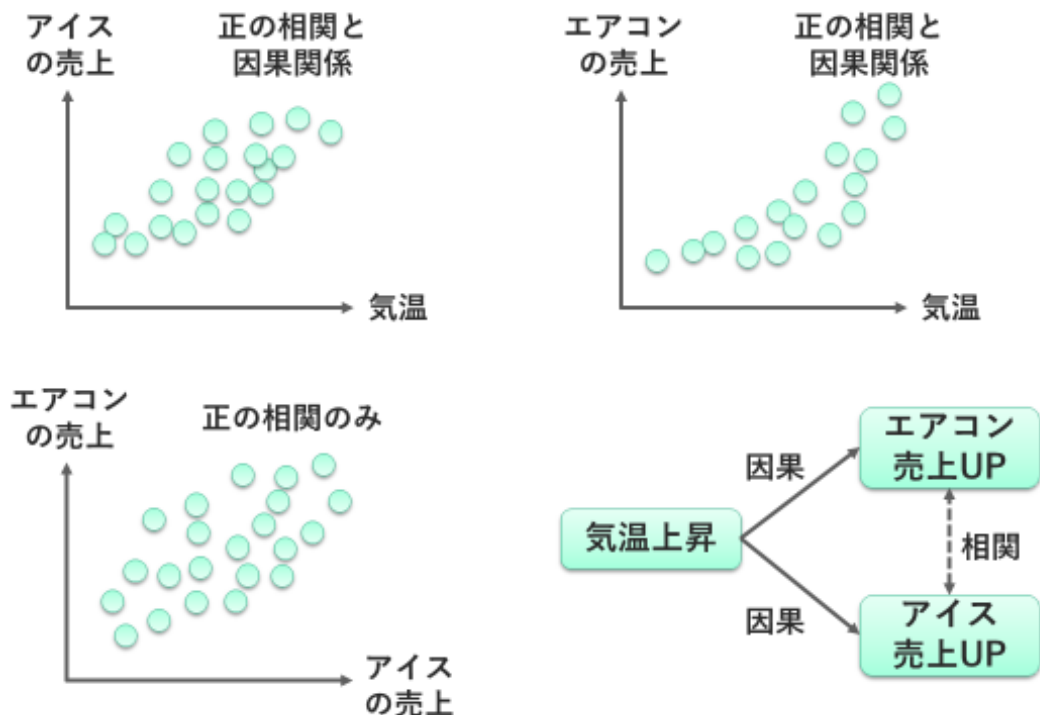
因果関係が存在する場合が多いが、必ずそうなるわけではない

相関分析の話をする前に、「説明変数と目的変数」という考え方を知っておきましょう。

同じ店を出している A ステーキよりも B ステーキのほう値段が高いのはなぜ？ と問われたら、単純に「B ステーキのほうが大きいいから」というのが自然な答えでしょう。これは値段の違いを内容量によって説明しているもので、ここでいう「値段」を目的変数、「内容量」を説明変数と言います。

本章の最初に出てきた「なぜ A 店で X 商品が売れるの？」という疑問の例で言えば、「X 商品の売上」が目的変数、「駐車場の広さ」が説明変数です。目的変数と説明変数の間には因果関係が存在する場合がありますが、必ずそうなるわけではありません。

相関関係と因果関係

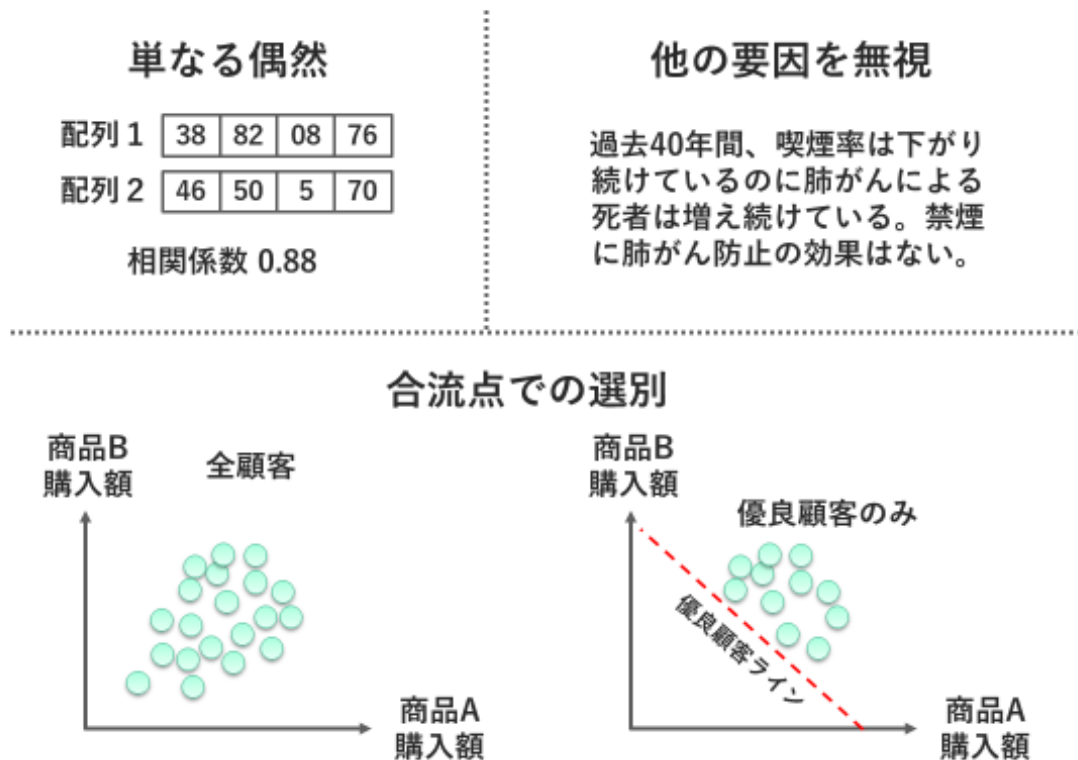


相関関係と因果関係の違いはよく誤解されています。あるお店のアイスクリームの売上と、その日の平均気温データを数十日分記録してその組み合わせを散布図にすると図中左上のようになったとします。全体として右上がりの「正の相関」を示しています。「暑くなると冷たいデザートが欲しくなる」のは自然な感覚ですからおそらく因果関係もあるでしょう。

「気温とアイス」が典型的な「因果関係のある正の相関」であるのに対して、電機店では「気温が高いとエアコンが売れる」ことも知られています。これも「因果関係のある正の相関」なのは間違いありません。

では、「エアコンの売上」と「アイスの売上」で散布図を描いたらどうなるのでしょうか？ 実はこれもおおむね右肩上がりの正の相関を描きます。しかしこの 3 種類のうち、「気温とアイス」「気温とエアコン」には因果関係がありますが「アイスとエアコン」には因果関係がないことに注意してください。共通の原因（気温上昇）を持つ 2 つの変数（エアコンとアイスの売上）の間には相関がありますが、だからといってがんばってアイスクリームを多く売ってもエアコンの売上は増えません。当たり前すぎる話ですが問題が複雑になるとこのレベルの間違ったロジックで「がんばれ！」と無駄な努力をしているケースも少なくないので、因果と相関の違いは知っておく必要があります。

相関・因果関係を誤認するパターン



相関関係や因果関係を間違って認識するパターンでもっとも単純なのが「単なる偶然」です。上図左上の配列 1 と配列 2 について相関係数を計算すると 0.88 と非常に高い相関を示しますが、実はこれはある日のある道路上で 1 分間上り方面と下り方面へ向かう車のナンバー下 2 桁をとっただけで、何の関係也没有。単なる偶然でも相関係数が高くなりうる場合があるため、標本調査は何度も行って検証する必要があります。

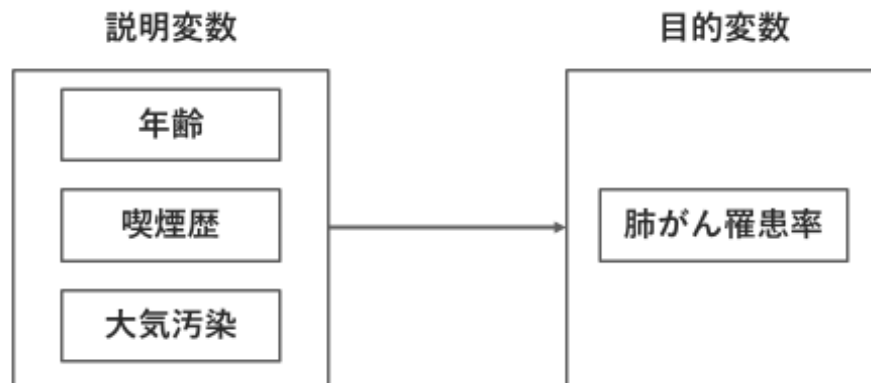
「喫煙率は下がり続けているのに肺がんの死者は増え続けている」というのは医療不信の文脈で「喫煙→肺がん」の因果関係を否定するためによく言われているものです。実際には、「喫煙習慣→肺がん発病」の間には何十年単位の長い時間がかかるため、現在増えている肺がんの多くは何十年も昔に喫煙していた人に発症しています。また、医療が進歩して「がん以外の病気ではなかなか死ななくなった」ため、結果として肺がんによる死者が増えているように見えています。これらの要因を無視して単純な相関関係だけを見ても適切な判断はできません。

上図左下はある店の全顧客に関して商品 A の購入額と商品 B の購入額で散布図を描いたものです。これを見ると A と B の購入額には正の相関があります。あるときその店では A と B の購入額の合計が一定ラインを越えたら優良顧客として特別なサービスをすることにしました。そこで優良顧客のみを見ると A と B の購入額には負の相関があるように見えてしまい、まったく逆の判断をしてしまうかもしれません。

そうすると何が困るのでしょうか？ もし「A と B の購入額に正の相関がある」ならそれは「A を買う客は B も買いたがる可能性が高い」ことを意味するので、「A と B、セットでお得キャンペーン」のような企画が有望です。一方、「A と B の購入額に負の相関がある」ならそれは「A と B の一方を買った人は他方を必要としない傾向がある」ことを意味するので、「今月の来店者には A か B のどちらか 1 つをお値引きします」というような販促施策が有望です。相関が正か負かによって施策の打ち方が変わってくるので、その読みを間違えると適切な手を打つことができません。

このようなミスは 2 つ（以上）のグループを合わせてある条件で選別したときに起こりやすく、「合流点での選別」のと言われています。「商品 A の購入」と「商品 B の購入」がそれぞれグループに当たり、「合計購入額が高い客を優良とする」が「合流点での選別」にあたります。

説明変数は複数ありうる

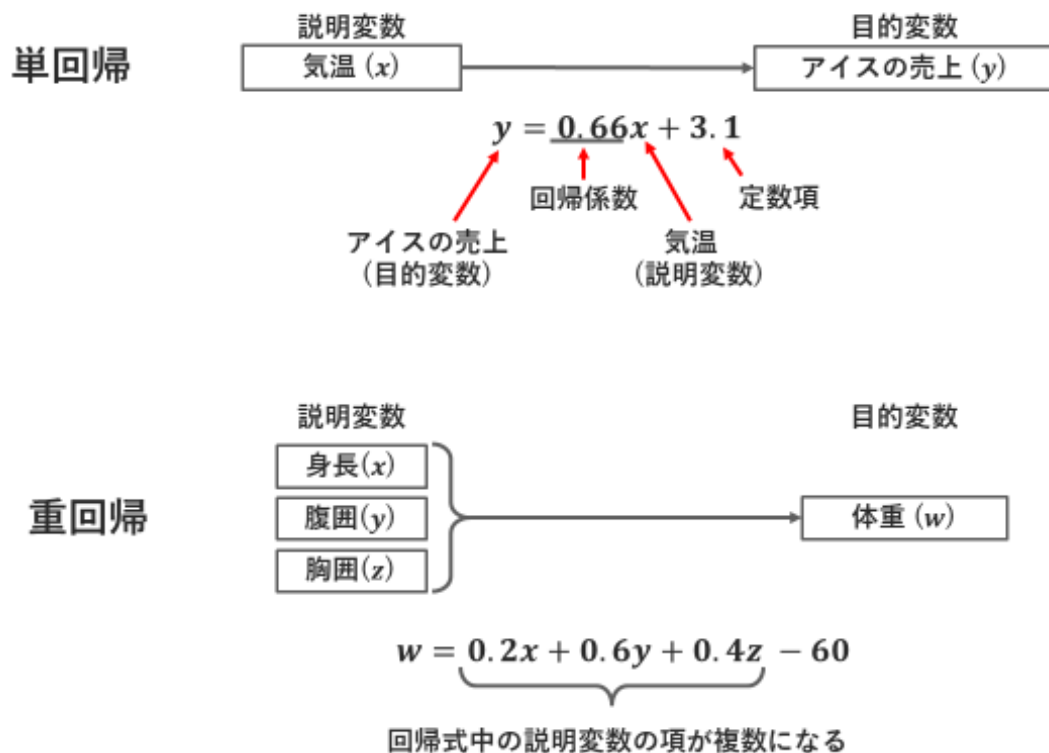


目的変数 1 つに対して説明変数が複数存在することはよくあります。たとえば「肺がん」という病気には、年齢、喫煙歴、大気汚染などの要素が関係していると考えられています。

なお、A（目的変数）の変化を B（説明変数）で説明できるとしても、B と A の間に因果関係があるとは限りません。前ページでも書いたように、相関関係と因果関係は混同しやすく、単なる相関関係である場合もよくあります。

それでも、目的変数と説明変数の関係を明らかにすることには意味があります。たとえばある人が「肺がんにかかっているかどうか」は手間とお金のかかる検査をしなければわかりませんが、「若い時期に長い喫煙歴がある」かどうかは簡単なアンケートをとるだけでわかります。該当する集団は「肺がん罹患率の高い集合」になるので、それに対して本格的な検査をすれば効率良く肺がんを早期発見し治療することが可能で、結果として多くの人を救うことができます。

単回帰分析と重回帰分析



相関のある変数どうしの関係から回帰式を導くと、さまざまな予測や意思決定に使えます。回帰式のうち、説明変数が 1 個しかないものを単回帰式、複数あるものを重回帰式といい、それらを導く分析のことを単回帰分析、重回帰分析などと言います。

上図では体重を予測する重回帰式の例を示していますが、身長と体重に相関があるのは当然として、腹囲や胸囲も体重との相関があるため、回帰式中の説明変数の項が 3 つになっています。なお、説明変数の項に掛ける定数（アイスの売り上げの式で言えば 0.66 の部分）のことを回帰係数と言います。

世の中のさまざまなものごとのうち、「1 つの説明変数だけで結果を見通せる」というものは多くはありません。肺がんには年齢・喫煙歴・大気汚染が影響し、店の売上には商圏人口・駐車場・店舗面積が影響し、などのように複数の説明変数が目的変数に影響するのが普通です。重回帰分析はその複雑な関係を数式にして扱いやすくする分析方法です。