

## ビジネスマンの 数字活用力向上講座

- ◆ 初版 : 2011年10月
- ◆ 発行所 : 同友館
- ◆ 単行本 : 184ページ

### ビジネスマンの流行病「数字アレルギー」

私はビジネスコンサルタントとして20年やってきました。この間、何万人というビジネスマンと会ってきました。

そして多くのビジネスマンが数字を使うのが下手、と言うよりも彼らには“数字アレルギー”があることに気づきました。

私の今の仕事の中心は、経営者、マネジャー、リーダーといった企業の中核メンバーを育てることです。そのために企業内に塾のようなものを作り、指定したテキストで自己学習してもらい、そのうえでセミナーを開いて、4~6人のグループでケーススタディ（ビジネスシーンを提示して、そこで自分ならどのように考えるか、行動するかを考える）や自分の会社の課題についてディスカッションをしてもらいます。

経営戦略、マーケティング、マネジメントなどに関するケーススタディでは、セミナーがぐっと盛り上がり、議論が止まらなくなってしまうことがほとんどです。

しかしビジネスで“数字を使うシーン”をケーススタディの題材とすると、様相は一変し、次のような会話がグループでなされます。

「このテーマって難しいなあ。誰か数字の得意な人いないか？どうも昔から、数字はダメでね。あなたはたしか情報システム部ですよ。だったらよろしく願いしますよ」

「いや私はITは使えるけど、データ分析は苦手で・・・」

そしてケーススタディの結果も、他のテーマに比べ極めて低調なものです。

### 数字アレルギーの原因は学校のお勉強

なぜビジネスマンには数字アレルギーがあるのでしょうか。

それは多くの人々が「数字活用」⇒「数学」⇒「昔から苦手だった」⇒「数字活用のやり方は自分には絶対に身につかない」という条件反射を示すからです。

確かに、数字を活用する時には数学の知識が必要です。もっと言えば、数字活用のやり方を身につけるなら、数学を勉強するのが早道といえます。

だから本書にも  $n$ 乗根、母集団、積分、標準偏差、重回帰分析、正規分布、t検定・・・といったマニアックに感じるかもしれない数学用語が出てきます。

しかしビジネスマンが、数字活用とは異なり好んで学習する“マーケティング”にだって、CRM、ポートフォリオ、エリアマーケティング、フランチャイズシステム・・・など、学校ではまったく教えてくれず、基礎を学習していない人にとって

は理解するのが困難なものだってあります。

それに比べ、数学は誰でも中学校で最低3年間、高校を出ていれば小学校の算数を含めて10年以上、その基本的理論を学校で学んでいます。

ただ問題は、学校で数学を教える多くの教師（結構数学が嫌いな人もいます）です。算数が得意で計算は早いですが、数学の真髄ともいえる論理性が低い人も意外に多いようです）にプロ意識がなく（失礼！）、手を抜いて「公式」だけを覚えさせようとしてきたことです。

数学を教わった側は、その公式の意味が全くわからないまま、苦しみながら暗記し、テストでは公式を忘れてしまったり、公式の使い方がうまくできなかつたりして、点数が他の科目と比べて悪い結果となってしまいます。だから数学は勉強しても「自分にはわからない」「自分は数学に向いていない」と思い込んでいます。

そしてそれが大人になってもトラウマとなって残っているようです。

### 数字活用のコツは理解、体験、自信

数学は、凡人のために天才数学者たちが最高の「数字の使い方」を考えたものであり、人類の知恵です。

でも数学には1つ難点があります。

数学の世界では、わかりやすさよりも大切にしていることがあります。それは「まわりの人に反論させない」ということです。数学を使って出た結果に、「それはおかしい。こうするべきだ」と言わせない方法を、天才数学者たちはずっと考えてきました。これを「証明」といいます。

あらゆる反論を封じ込めるようにしたため、その「数字の使い方」は最高のやり方なのですが、キレが良すぎて、かえって凡人にはとっつきにくいものとなっているのかもしれませんが。

でもくどいようですが、数学者たちは自己満足ではなく、世の中の人々が「うまく数字を使えるように」と真剣に考えています（はずです）。だから普通の人にはわからないはずはありません。

このキレが良くてとっつきにくい“数学”をベースにした「数字の使い方」を身に付けるコツは、次の3つに要約されます。

1つは「なぜそういう“やり方”をするのか」ということに納得するまで、それを自分の頭で考えることです。絶対にやってはいけないのは、理解していないのにその“やり方”を無理矢理“体”で覚えこんでしまうことです。「習うより慣れる」は“禁じ手”です。その“やり方”を使えるだけでなく、まわりの人に説明できるまで理解すること、これがもっとも大切なポイントです。

2つ目はその理解した“やり方”を、いろいろなシーンで何度も自分で使ってみることで。

3つ目はその使った経験を基にして、「私だって使える」という自信を付けて、数字アレルギーを治してしまうことです。

つまり「数字の使い方」マスターのコツは、理解、体験、自信の3つです。見方を変えればビジネスの基本どおりに、**PLAN**（理解）、**DO**（体験）、**SEE**（自信）という**PDS**サイクルを回すことです。

### 体験すれば自信がわく

こんな背景から、私は何とかビジネスマンに数字の使い方を理解してもらい、自信を持ってもらおうと思い、これまで何冊もの本を書きました。

「数字を使える営業マンは仕事ができる」（日本経済新聞社）

「微分積分を知らずに経営を語るな」（PHP新書）

「数学を使えるビジネスマンはみな幸福である」（ベスト新書）

「会社の数字を科学する」（PHPサイエンス・ワールド新書）

「計数分析のセオリー」（同友館）

⋮

といったものです。

これらの本は「わかりやすい」と好評を得たのですが、多くの読者から次のような声が挙がりました。

「数字の使い方の教科書や、誰かがやった『数字の使い方の工夫』ではなく、もっと自分で頭を使い、手を動かして、実際に数字を使うことが実感できる本にしてほしい」

「本に書いてあるデータをエクセルに入れて、自分の力で分析したい」

先ほどの3つのコツのうち、「体験」についての声です。本を読んで理解はしても、自分自身での体験が今1つできないので、「自信がわからない」という声です。

この声に応えたのが本書です。本書は今までの本とは異なり、数字活用のノウハウよりも、その「体験」にウエイトをおいています。体験をたくさんすることで、本書を読んだ後に「私も数字が使えるそう」、もう一步踏み込んで「誰かにこれを教えたい」という“自信”を持ってもらうことが本書のねらいです。

## 本書の使い方

本書はケーススタディ方式で、「数字の使い方」を身につけるトレーニングブックです。

本書を読むうえで必要なものがあります。それはパソコンとエクセルです。本書は「エクセルを使って数字を活用すること」を前提に書かれています。

本書はLesson1～Lesson 6の6章構成です。

各Lessonの頭には、「数字活用ケース」として「ビジネスで実際に数字を使うシーン」が書かれています。読者の方はまずこのケースをよく読み、その主人公が抱えている悩みを共有し、「数字を活用してその悩みを解決する方法」を自分で考えてみてください。

そのうえで「解説」を読んでください。

解説を読む時は、ケースに記述されているデータを自分でエクセルに入れ、本文を読みながら、そこでの指示どおりに数字を使ってみてください。こうすることで先ほどのPDSサイクルを回して下さい。理解⇒体験⇒自信⇒理解⇒体験・・・です。

さらに理解、体験、自信を強固にするために、本書では各ケースの後にExerciseが付いています。これは各ケーススタディで学んだことが身に付いているかをチェックする練習問題です。今度はきっと、すべて自力で解答できるはずです。

そしてそれが自信につながるはずです。

本書にある6つのケース、13のExerciseをクリアすれば、あなたには数字の使い方とともに、次のようなキーワードが身に付いています。

Lesson1・・・母集団、推定、サンプリング、標本値、統計量、期待値、統計処理、幾何平均、標準偏差、分散

Lesson2・・・予測、散布図、最小2乗法、線形近似

Lesson3・・・相関分析、相関係数、回帰分析、重回帰分析、説明変数、被説明変数、シミュレーション、環境パラメータ、意思決定パラメータ、評価パラメータ、数量化理論

Lesson4・・・確率、離散確率分布、連続確率分布、ヒストグラム、確率密度関数、積分、微分、正規分布、安全係数、安全在庫、3シグマ、シックスシグマ

Lesson5・・・検定、検定統計量、帰無仮説、棄却率、有意水準、平均値の差の検定、有意差検定、t分布、t検定、バラツキの差の検定、F分布、F検定、適合度の検定、 $\chi^2$ 分布、 $\chi^2$ 検定

Lesson6・・・キャッシュフロー、現在価値、DCF、割引率、内部利益率

どうです？あなたもこんなおしゃれな言葉を使いながら、エクセルでカッコ良く数字を使いこなすビジネスマンに変身したいとは思いませんか？

あなたがもしそう思うなら、レッツトライです。

### 〔本書の使用上の注意〕

- ・本書ではエクセルの分析ツール機能を使用します。そのため本書でエクセルを使う前に「ツール」（または「オプション」）から「アドイン」を選び、そのメニューで「分析ツール」にチェックを入れておいてください。

- ・エクセルはそのバージョンによりユーザーインターフェース（メニューなどの表示）が少し異なります。本書執筆時点で、インターフェースが2通りある時は、本文および（ ）内にそれぞれを記してあります。（上の「ツール」「オプション」もその例です。）また本書出版後、その機能やインターフェースが変わることも予想されます。ただ現在のエクセルが持っている機能がなくなることは考えづらいので、もし本書に書いてあるインターフェースが見つからない時は、似たようなものを探してみてください。きっと同じようなものがあるはずです。

# 目次

---

## プロローグ

### Lesson1 数字の使い方の基本

#### 数字活用ケースその1

#### Exercise-1

### Lesson2 明日を読む

#### 数字活用ケースその2

#### Exercise-2

#### Exercise-3

### Lesson3 数字と数字の関係を分析する

#### 数字活用ケースその3

#### Exercise-4

#### Exercise-5

### Lesson4 未来を確率で考える

#### 数字活用ケースその4

#### Exercise-6

#### Exercise-7

#### Exercise-8

#### Exercise-9

### Lesson5 数字で相手を説得する

#### 数字活用ケースその5

#### Exercise-10

#### Exercise-11

#### Exercise-12

### Lesson6 カネの未来を考える

#### 数字活用ケースその6

#### Exercise-13

## エピローグ

# 数字の使い方の基本

## 統計をマスターしよう

統計というのは、「数字の使い方」を数学者が体系化したものです。この統計をビジネスから見ると、「たくさんある数字をうまく加工し、まわりの人にわかりやすく説明し、そして納得してもらおうツール」と考えられます。

あなたが統計というテクニックを使って誰かに数字の説明をすれば、相手はあなたに反論できません。「私はそう思わない」「私はこのように数字を加工した方がいいと思う」と相手は言うことができません。数学者はこの反論を封じ込める努力を何百年と続けてきたのです。その結果生まれたのが、統計という説明テクニックです。

さあ、あなたも統計の基本の基本、というよりもビジネスに使える「統計の面白い所」だけを勉強して、これを使ってみましょう。そうすれば、まわりの反応のちがいにびっくりするはずです。

## 数字活用ケースその1

—— 商品の強みは「伸びとコンスタント」——

山田さんは顧客へ、自社商品にパワーがあることを証明したいのですが、うまくいきません。

販売データをグラフにしてみても、「うちの商品は強い」と直感的に思うのですが、顧客にその気持ちをどう説明してよいかわかりません。

あなたなら、どうやって説明しますか？

## はっきりとした根拠が必要

A社は乳製品メーカーである。乳製品業界では、近年コラーゲン入りのヨーグルトが注目されており、A社の「Aコラーグル」とライバルX社の「Xヨーゲン」が激しいトップシェア争いをしている。

山田はA社の北陸営業所に所属するセールスマンであり、北陸地区にスーパーマーケットを10店舗展開しているBスーパーを担当している。

Bスーパーでは、毎年6月末に各店舗の商品レイアウト\*1の変更を行っている。今年も6月に入り、7月からの夏シーズンにどんな商品を品揃えしていくかの検討を始めていた。

BスーパーC店の加工食品部門を担当しているバイヤー\*2である加藤は悩んでいた。

「コラーゲンヨーグルトは商品カテゴリー\*3としては底が堅い。いや消費者の認知が進み、カテゴリー全体としては伸びている感さえある。

今のコラーゲンヨーグルトの定番\*4は、AコラーグルとXヨーゲンの2つか。この2つは完全にバッティングしているな。この2つを両方置いても、あまり意味はない。どちらか1つに絞った方が、店舗での販促\*5もしやすいし、陳列にボリューム感が出て、買物客の認知\*6度が高まるよな。

ただどちらに絞るかは難しいところだなあ」

そんな時に、山田が加藤を表敬訪問した。

もちろん山田は「それなら何とかAコラーグルに絞ってもらえませんか。商品力はうちのAコラーグルの方があはずです」とプッシュした。

加藤は渋い表情で次のように答えた。

「そうは言ってもなかなか難しいよ。私としてはどうにも決めづらい。この3ヶ月間の販売データを見る限りでは、Xヨーゲンの方が量的には売れていることは事実だ。

ただ山田さんが言うように、そして私のバイヤーとしての直感でも、Aコラーグルは何か“強み”を持っているような気もする。しかし私だって自分の“感じ”だけでは品揃えを決められない。本部\*7のスタッフに『なぜそれに絞るのか』を説明しなくてはならない。

Xヨーゲンに絞るなら、『トータル販売量が上だから』でOKだけど、販売量で負けているAコラーグルに絞るなら、はっきりとした根拠が必要だよ」

山田「それならAコラーグルとXヨーゲンのC店での販売データを見せてもらえませんか。社に持ち帰って検討してみたいと思います」

加藤「わかった。データ分析してくれるなら、販売データを貸してもいいよ。

3月から3ヶ月間の日販\*8データなら、本部からの販売実績リストが手許にある。それを使って、A社としての意見を聞かせてくれ。もちろんX社にも意見は聞くけどね」

- \* 1.商品レイアウト 店舗内に商品をどのように置くかを定めること
- \* 2.バイヤー 仕入担当責任者
- \* 3.カテゴリー 商品分類のこと
- \* 4.定番 店に常時置いてある商品
- \* 5.販促 販売促進の略。この場合は、店舗で商品が売れるように努力すること
- \* 6.認知 その商品の存在、機能を顧客が知ること
- \* 7.本部 この場合は、複数の店舗をとりまとめている部門のこと
- \* 8.日販 1日あたりの販売数または販売金額

## エクセルに入れて平均をだしてみよう

山田にAコラーグルとXヨーゲンの12週分のデータがリストで渡された。

山田「まずはこの日販データをエクセルに入れよう。データを週単位にして、個数と売上額を入れていこう。

よしできた。

さてと、週合計を取ろう。『週合計』のセルを作って、ここにカーソルをあわせて、合計だから『Σ』のマークだな。これを押して、範囲はAコラーグルの個数だから、3/7~3/13の『個数』のセルを指定して、**ENTER**を押してと。出た80だ。あとはこのセルをコピーして、第2週の『週合計』のセルに貼って...

次は平均を出そう。平均はエクセルの関数でいけるな。『週平均』のセルを作って、ここにカーソルを合わせる。『関数』は『fx』のマークだな。これを押して『関数の検索』に『平均』と入れて、『**AVERAGE**』を選ぶ。『数値1』に3/7~3/13のセルを指定して。よし出た。でも小数点以下がうっとうしいな。小数点第1位くらいにしよう。『書式』をクリックして、『セル』（または『セルの書式設定\*1』）を選んで、表示形式を『数値』にして、小数点以下の桁数を『1』にして、よし出た。『11.4』か。いい感じだな。あとは合計と同じようにコピーして、『第2週』以降の『週平均』のセルに貼る...

最後に、『12週合計』は『関数』を使おう。『fx』を押して『関数の検索』に『合計』を入れて、『SUM』を選んで、『数値1』に『第1週合計』、『数値2』に『第2週の合計』のセル…を指定し、出た。

『12週平均』は『AVERAGE』で12個の『週平均』を平均しよう。『AVERAGE』を選んで、数値を指定して、よし出た」

\*1. エクセルのメニューの後にある（ ）はプロログで述べたとおり、別のインターフェースもあることを示している。この場合は『セル』または『セルの書式設定』というメニューを選ぶことを表わしている。

		第1週									
		3/7(月)	3/8(火)	3/9(水)	3/10(木)	3/11(金)	3/12(土)	3/13(日)	週合計	週平均	
Aコラーグル	個数	4	6	10	8	12	16	24	80	11.4	
	売上額	800	1200	2000	1600	2400	3200	4800	16000	2285.7	
Xヨーゲン	個数	12	20	4	8	4	36	32	116	16.6	
	売上額	2400	4000	800	1600	800	7200	6400	23200	3314.3	
		第2週									
		3/14(月)	3/15(火)	3/16(水)	3/17(木)	3/18(金)	3/19(土)	3/20(日)	週合計	週平均	
Aコラーグル	個数	6	8	12	10	14	18	26	94	13.4	
	売上額	1200	1600	2400	2000	2800	3600	5200	18800	2685.7	
Xヨーゲン	個数	8	10	18	2	14	32	24	108	15.4	
	売上額	1600	2000	3600	400	2800	6400	4800	21600	3085.7	
		第3週									
		3/21(月)	3/22(火)	3/23(水)	3/24(木)	3/25(金)	3/26(土)	3/27(日)	週合計	週平均	
Aコラーグル	個数	10	14	12	16	12	8	28	100	14.3	
	売上額	2000	2800	2400	3200	2400	1600	5600	20000	2857.1	
Xヨーゲン	個数	48	8	36	10	40	32	36	210	30.0	
	売上額	9600	1600	7200	2000	8000	6400	7200	42000	6000.0	
		第4週									
		3/28(月)	3/29(火)	3/30(水)	3/31(木)	4/1(金)	4/2(土)	4/3(日)	週合計	週平均	
Aコラーグル	個数	24	16	12	16	28	36	32	164	23.4	
	売上額	4800	3200	2400	3200	5600	7200	6400	32800	4685.7	
Xヨーゲン	個数	16	4	2	30	12	4	14	82	11.7	
	売上額	3200	800	400	6000	2400	800	2800	16400	2342.9	
		第5週									
		4/4(月)	4/5(火)	4/6(水)	4/7(木)	4/8(金)	4/9(土)	4/10(日)	週合計	週平均	
Aコラーグル	個数	24	16	8	8	10	28	12	106	15.1	
	売上額	4800	3200	1600	1600	2000	5600	2400	21200	3028.6	
Xヨーゲン	個数	32	8	36	14	24	40	44	198	28.3	
	売上額	6400	1600	7200	2800	4800	8000	8800	39600	5657.1	
		第6週									
		4/11(月)	4/12(火)	4/13(水)	4/14(木)	4/15(金)	4/16(土)	4/17(日)	週合計	週平均	
Aコラーグル	個数	12	16	12	14	10	24	20	108	15.4	
	売上額	2400	3200	2400	2800	2000	4800	4000	21600	3085.7	
Xヨーゲン	個数	16	36	8	44	6	28	38	176	25.1	
	売上額	3200	7200	1600	8800	1200	5600	7600	35200	5028.6	

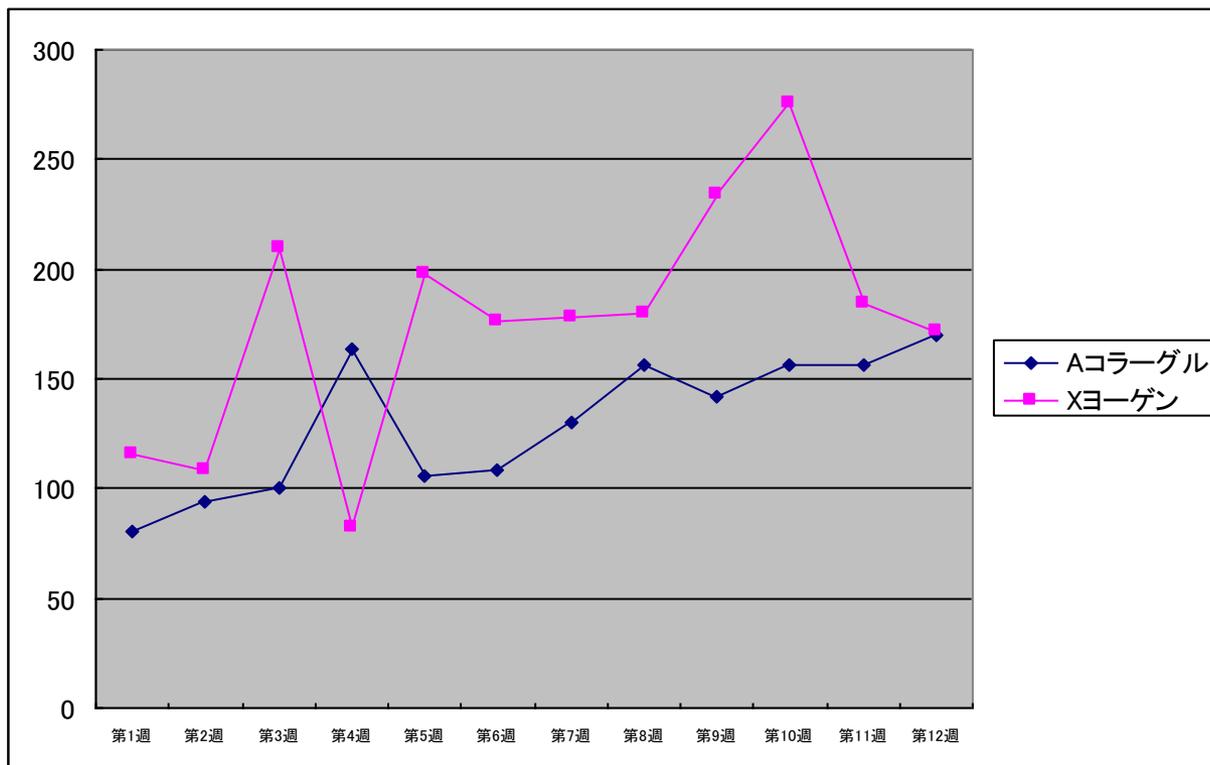
		第7週							週合計	週平均
		4/18(月)	4/19(火)	4/20(水)	4/21(木)	4/22(金)	4/23(土)	4/24(日)		
Aコーラ	個数	16	12	10	28	12	28	24	130	18.6
	売上額	3200	2400	2000	5600	2400	5600	4800	26000	3714.3
Xヨーゲン	個数	6	36	14	22	56	4	40	178	25.4
	売上額	1200	7200	2800	4400	11200	800	8000	35600	5085.7
		第8週								
		4/25(月)	4/26(火)	4/27(水)	4/28(木)	4/29(金)	4/30(土)	5/1(日)	週合計	週平均
Aコーラ	個数	16	12	12	24	24	32	36	156	22.3
	売上額	3200	2400	2400	4800	4800	6400	7200	31200	4457.1
Xヨーゲン	個数	4	40	16	42	26	40	12	180	25.7
	売上額	800	8000	3200	8400	5200	8000	2400	36000	5142.9
		第9週								
		5/2(月)	5/3(火)	5/4(水)	5/5(木)	5/6(金)	5/7(土)	5/8(日)	週合計	週平均
Aコーラ	個数	14	12	16	24	28	24	24	142	20.3
	売上額	2800	2400	3200	4800	5600	4800	4800	28400	4057.1
Xヨーゲン	個数	36	14	40	44	32	12	56	234	33.4
	売上額	7200	2800	8000	8800	6400	2400	11200	46800	6685.7
		第10週								
		5/9(月)	5/10(火)	5/11(水)	5/12(木)	5/13(金)	5/14(土)	5/15(日)	週合計	週平均
Aコーラ	個数	16	14	18	26	30	26	26	156	22.3
	売上額	3200	2800	3600	5200	6000	5200	5200	31200	4457.1
Xヨーゲン	個数	40	8	52	60	20	56	40	276	39.4
	売上額	8000	1600	10400	12000	4000	11200	8000	55200	7885.7
		第11週								
		5/16(月)	5/17(火)	5/18(水)	5/19(木)	5/20(金)	5/21(土)	5/22(日)	週合計	週平均
Aコーラ	個数	16	14	18	26	30	26	26	156	22.3
	売上額	3200	2800	3600	5200	6000	5200	5200	31200	4457.1
Xヨーゲン	個数	30	6	40	30	16	40	22	184	26.3
	売上額	6000	1200	8000	6000	3200	8000	4400	36800	5257.1



	第1週	第2週	第3週	第4週	第5週	第6週	第7週	第8週	第9週	第10週	第11週	第12週
Aコラーゲル	80	94	100	164	106	108	130	156	142	156	156	170
Xヨーゲン	116	108	210	82	198	176	178	180	234	276	184	172

(図表1-2)

さあグラフだ。まずは範囲の指定だ。この3行だな。『挿入』をクリックして、『グラフ』の中から『折れ線グラフ』にしよう。形式は『マーカー付き』にしよう。よしできた」



(図表1-3)

## 伸びを計算しよう

「これを見るとAコラーゲルは安定して伸びている。Xヨーゲンもトレンドとしては伸びているんだろうけど、大きくブレているなあ。ただこのブレをどうやって表現したらいいんだろう。わからない」

「仕方ない。ブレは置いといて、とりあえず“伸び”の方から考えてみよう。“伸び”というのは大きな商品力だよな。

伸びを表わすなら伸び率かな。週単位で前週からの伸び率を計算してみよう。さっきグラフを作った時の表にある『Aコラーゲル』の『行』の下へ、『伸び率』の『行』を追加してみよう。

第2週の下『伸び率』は、第2週を第1週で割り算だ。『=』を押して、『94』を『80』で割りたいから、『94』のセルを指定して、割り算だから『/』を押して、『80』のセルを指定する。小数点以下は第2位くらいにするか。『1.18』か。

後はこの『1.18』のセルをコピーして、第3週以降に貼り付ける。よしできた。

しかし、これだと数字がそれぞれ11個あってわかりづらいなあ。AコラーゲルとXヨーゲンの伸び率を、それぞれ平均してみよう」

	第1週	第2週	第3週	第4週	第5週	第6週	第7週	第8週	第9週	第10週	第11週	第12週
Aコーラール	80	94	100	164	106	108	130	156	142	156	156	170
Aコーラール伸び率	—	1.18	1.06	1.64	0.65	1.02	1.20	1.20	0.91	1.10	1.00	1.09
Xヨーゲン	116	108	210	82	198	176	178	180	234	276	184	172
Xヨーゲン伸び率	—	0.93	1.94	0.39	2.41	0.89	1.01	1.01	1.30	1.18	0.67	0.93

	平均 伸び率
Aコーラール	—
Aコーラール伸び率	1.10
Xヨーゲン	—
Xヨーゲン伸び率	1.15

$$=94/80$$

(図表1-4)

「Aコーラールの平均伸び率が1.10、Xヨーゲンが1.15、ということは、えっ、Aコーラールが平均10%の伸びで、Xヨーゲンが15%の伸びかあ。

Xヨーゲンの方が伸びているのか。何かおかしいなあ・・・。

Aコーラールの1週目が『80』で、12週目が『170』、Xヨーゲンが1週目が『116』で、12週目が『172』なのになあ。

だからといって1週目と12週目だけで伸び率を計算したら、せっかく見せてもらった2週～11週のデータを使ってないことになってしまうし・・・。

うーん。お手上げだ。これじゃあBスーパーの加藤さんに説明できない。」

## これは統計処理した結果です

このケースについて考える前に、統計をゼロからきちんと勉強していきましょう。

統計では、「知りたいデータ全体」のことを母集団といいます。

一般に母集団にある“すべてのデータ”はわかっていないことが多いといえます。ここに統計というテクニックが使われます。

山田さんのケースよりも、もう少し簡単な例で考えてみましょう。

納豆を生産し、これを機械で自動パッキングしているa社で、自社の納豆1パックあたりに「豆が何粒入っているか」を知りたいとします。

この時、母集団はa社で生産している「すべての納豆パック」に入っている「豆の数」です。しかし全部開けて数えるわけにはいかないので、いくつかのパックを選んで、これを開けて数えてみることにしました。

このように母集団のデータのすべてが手に入らなかつたり、手に入れようとすると膨大なコストがかかってしまったりする時には、「手に入る一部のデータ」を頼りに、母集団全体の状況を考えます。これを推定といいます。

また、この「手に入る一部のデータ」を標本（サンプル）、母集団から標本を選び出すことをサンプリングといいます。

サンプリングされた納豆パックの豆の数を数えてみると、「245個、262個、258個、235個」でした。この4つの“数字”を標本値といいます。

しかしこの4つの数字（山田さんのケースでは84個×2あります）は、このままでは多すぎて、かえって使いづらいので、何とか1つの数字で表現できるように加工します。この加工された数字を統計量といいます。

統計量の代表選手は平均値です。先ほどの4つの標本値の平均値は250であり、これが統計量です。

この250というのは標本の数字を加工したものであり、本当に知りたいのは母集団の平均値です。つまりこのメーカーで作られる「すべての納豆パックに入っている豆の数」の平均値です。この母集団の「知りたい数字」（この例では平均値）を期待値（きつとこうなるだろう）といいます。

「標本の統計量で母集団の期待値を推定する」わけです。

あなたが上の文の「 」の中に書いてあることがわかれば、統計の基本は身に付きました。

「a社が販売するすべての納豆パックに入っている豆の数の平均値は、250と考えるべきだろう」。

これが統計処理（統計というテクニックを使って数字をいじることをこう言います。ビジネスでもバンバン使いましょう！）した結果であり、数字がこの4つしかない限り、このことに誰も反論することはできません。

## 何が母集団で、何が標本だろう

さあ山田さんのケースで、勉強した「統計の考え方」を使ってみましょう。

ここで「知りたいこと」は、AコラーグルとXヨーゲンのBスーパーC店での「商品力」です。

この商品力をどう表わすかです。

まず個数と売上額という2種類のデータがありますが、山田さんが言っているように、すべて売価200円ですので、個数だけを使えばOKでしょう。

商品力としては、やはり山田さんの考えたとおり「売れ行き」が妥当でしょう。ただ「売れ行き」というのは単なる「売れた個数」ではなく、「時間を加味した日販数」（「いつの日販数か」）でしょう。

もっと厳密に言えば、本当に知りたいことは図表1-1の12週分のデータを基にした商品力ではなく、店舗レイアウト変更のための「未来の商品力」でしょう。

つまりレイアウト変更後の7月以降の「売れ行き」です。

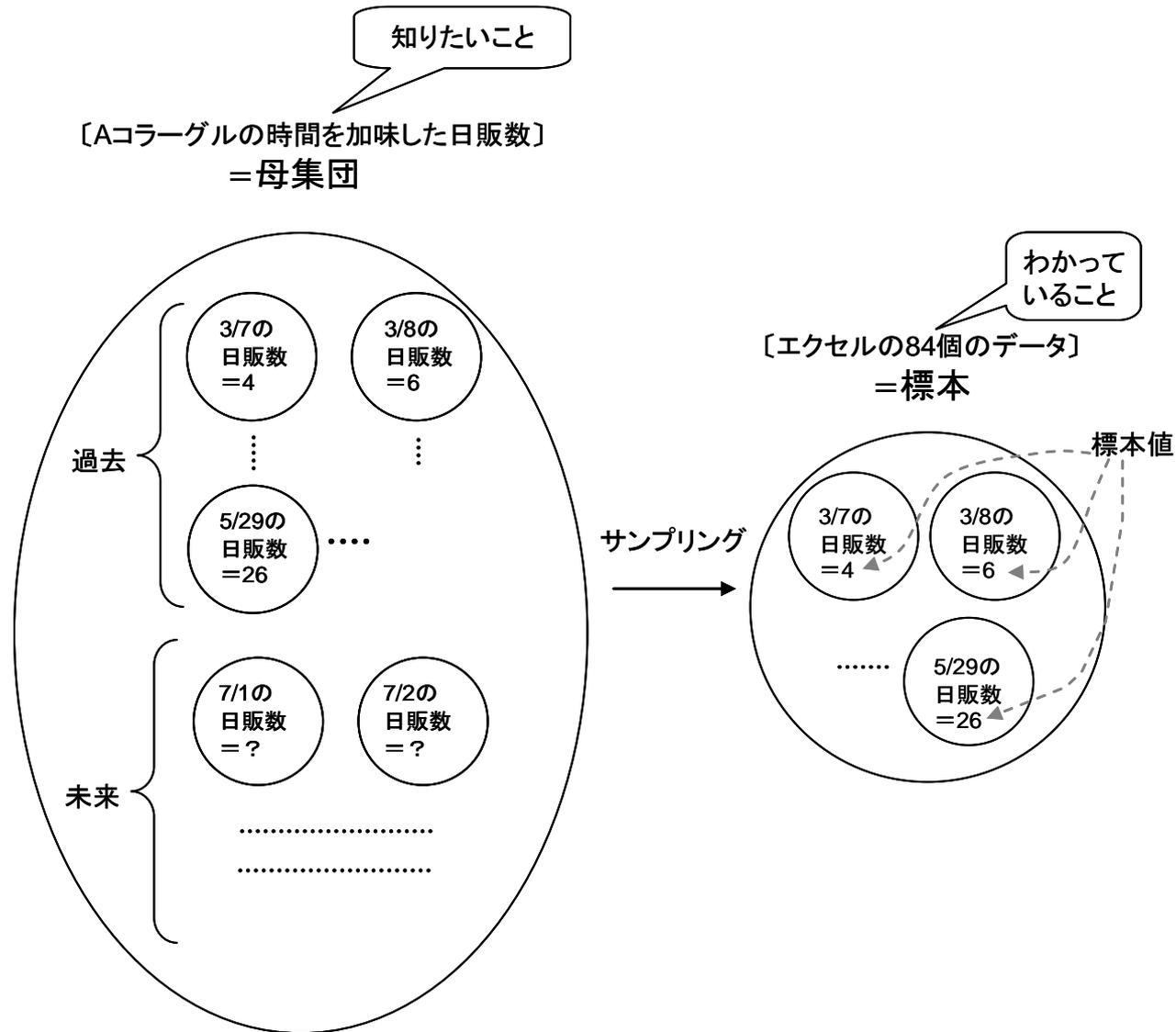
しかしこれでは「知りたいこと＝母集団」が「未来のデータ」となってしまい、その標本をサンプリングできません。「未来のデータ」なんて手に入るはずがありません。

せつかく顧客から提供された12週分のデータを使うには、何とかこれを標本とする母集団を考えなくてはなりません。

そこでBスーパーC店における過去、未来を含めた「売れ行き」、つまりすべての「時間を加味した日販数」（例えば「3月7日のAコラーグルの日販数＝4」）を母集団（先ほどの例なら納豆パック全体の豆の数）と考えます。

母集団はもちろんAコラーグル、Xヨーゲンという2つに分けます。そしてこの2つの母集団から、それぞれ図表1-1ページの12週分各84個のデータがその標本（先ほどの例なら4つの納豆パック）としてサンプリングされていると考えます。標本値はエクセルに入っている数字です。

この84個×2の標本を使って、それぞれの母集団を推定しようというものです。少しややこしいですが、統計の基本ですので頭をきちんと整理しておきましょう。



(図表1-5)

## 統計は感覚を数字に表わすテクニック

さあ次は統計量、期待値を何と考えるかです。

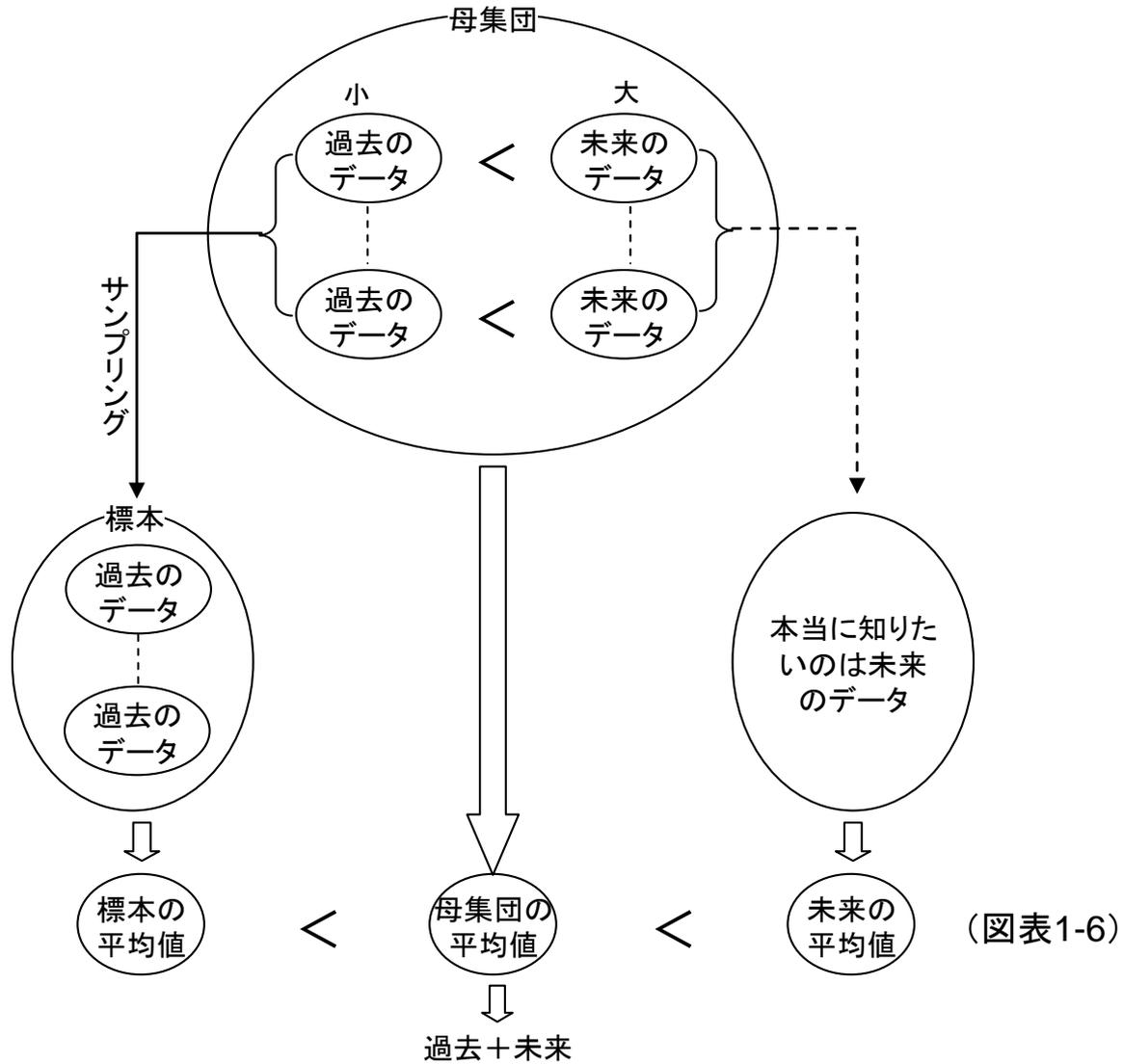
山田さんは、最初に84個の標本値（日販数）の「平均値」を考えてみましたが、どうもこの平均値が「それぞれの商品力を表わしているとはいえない」と感じています。それは決して自社のAコラーグルを“ひいき目”に見ているわけではないと思います。

7ページの図表1 - 3の折れ線グラフを見れば明らかのように、Aコラーグルは時とともに伸びています。またXヨーゲンもデコボコがありますが、トレンドとしては伸びているようにも感じます。

時とともに伸びている数字に対して、過去の標本の平均値を統計量として、「期待値＝未来を含めた母集団の平均値」を推定するのは不適切です。

時間とともに伸びているのですから、「未来の数字」の方が「過去の数字」より大きいと考えられます。だから未来を含めた母集団の平均値は、過去の84個の平均値より大きいはずです。つまり母集団の期待値（平均値）の方が統計量（標本の平均値）より大きいはずです。

もっといえば本当に知りたいのは、過去ではなく未来の商品力のはずです。



山田さんは、何とかこの「伸び」という商品力を表わしたくて「伸び率」を計算しています。

ここまではよいのですが、84個の標本値の統計量として、「伸び率の平均値」を取ると、何だか商品力という母集団を表わしている「感じ」がしません（「何かおかしいなあ」と思っています）。

数字を使う時、大切なことはこの「感じ」、つまり直感です。数字を統計処理して何かを導き出すのではなく、数字から人間が感じた直感を、統計というテクニックを使って表現するようにします。だから山田さんが「よしこれならフィーリングが合う」という数字を見つけなくてはなりません。

統計は人間の感覚を数字に表わすテクニックなのです。

### 「伸び率の平均」はかけ算で

山田さんは“伸び”という感覚を表わすテクニックを知りません。このようにいくつかある「伸び率」を平均する時は、こうやってやるではありません。このケースでいえば、12個の伸び率を足して、12で割るではありません。

こんな時には、幾何平均というテクニックを使います。

例えば、毎年10、20、80と伸びている時を考えてみましょう。伸び率は1年目「10」から2年目「20」で「2倍」、2年目「20」から3年目「80」で「4倍」という“伸び率”になります。

ここで「2倍」と「4倍」という2つの「伸び率」の平均を考えてみましょう。

「2倍」と「4倍」で単純に平均して、「3倍」とすると何か変です。「年平均3倍」の伸び率なら、1年目から3年目という2年間で $3 \times 3 = 9$ 倍となり、3年目には $10$ （1年目） $\times 9$ 倍で「90」になっているはずですが、しかし「80」、つまり「8倍」にしかありません。

算数が得意な人はもう気づいたと思います。2倍と4倍なら、 $\sqrt{2 \times 4} = \sqrt{8} \div 2.8$ 倍くらいが、「平均伸び率」となるはずですが。

「年平均2.8倍」の伸び率であれば、2年間で「 $2.8 \times 2.8 \div 8$ 」（ $\sqrt{8} \times \sqrt{8} = 8$ ）で約8倍となります。

これが幾何平均です。この「幾何」というワードは「幾何級数」（等比級数ともいい、一定の倍率で増えていく数字のこと。例えば2、4、8、16...。これは2倍ずつ増えている）から来ており、幾何平均は「伸び率の平均」（＝「一定の倍率」＝幾何）という意味です。

また幾何平均は相乗平均ともいいます。「乗」とは「かけ算」のことであり、相乗平均は「かけ算して平均する」という意味です。

これに対して先ほどの「足す」タイプの平均を、単純平均、算術平均、相加平均（「加」はたし算）などといいます。

幾何平均をしっかりと定義すれば次のようになります。

$a_1, a_2 \cdots a_n$  という  $n$  個の数字があった時、単純平均は  $\frac{a_1 + a_2 + \dots + a_n}{n}$  ですが、

幾何平均は  $\sqrt[n]{a_1 \times a_2 \times \dots \times a_n}$  となります。

$\sqrt[n]{\quad}$  は  $n$  乗根\*1を表わしています。 $\sqrt[3]{8}$  は「何を“3”回かければ“8”になるか」という“質問”のことです。その答えは2です。 $2 \times 2 \times 2 = 8$  ですので=2です。

つまりは「何を  $n$  回かけると  $a$  になるか」ということを意味しています。

（これくらいの算数でびびらないで下さい。どうしてもついてこれない人は「伸び率は幾何平均を取る」と頭に入れてしまいましょう。）

\* 1.  $n$ 乗根 本書のケースでは不要ですが、 $n$ 乗根はエクセルで計算できます。

エクセルのPOWERという関数を使います。POWERは「べき乗」（同じ数字を何回か“かける”こと）を計算するものです。例えば「 $5^4 = 『5の4乗』 = 『5 \times 5 \times 5 \times 5』$ を計算する」といったものです。この4を指数といいます。

「5の4乗」はPOWERの「数値の欄」に5、「指数の欄」に4を入れると「625」と計算してくれます。

数学では $\sqrt[n]{a} = a^{\frac{1}{n}}$ と定義されます。つまり、 $\sqrt[3]{8} = 8^{\frac{1}{3}} = 2$ 、8の $\frac{1}{3}$ 乗が2ということです。これでPOWERを使って  $n$ 乗根を計算できるようになります。

ならPOWERの関数の数値の欄に「8」、指数の欄に「1/3」と入れれば「2」と計算してくれます。

## 幾何平均の方がぴったりする

少し幾何平均の説明に手間どりました。

結論としては、この84個の標本の「伸び」を表わす統計量として、「伸び率の幾何平均」を選べばGoodということです。

それでは8ページの「週ごとの伸び率」の幾何平均を取ってみましょう。

図表1 - 4のAコーダグルの11個の伸び率（第2週～第12週）を全部かけて、11乗根を取ることです。計算式を書けば次のようになります。

$$\begin{aligned} & \text{Aコーダグルの伸び率の幾何平均} \\ & = \sqrt[11]{1.18 \times 1.06 \times 1.64 \times 0.65 \times 1.02 \times 1.20 \times 1.20 \times 0.91 \times 1.10 \times 1.00 \times 1.09} \end{aligned}$$

計算が大変そうですが、幾何平均はエクセルの「関数」で簡単に計算できます。関数としてGEOMEANを選び、対象のセル（第2週～第12週の11個の伸び率）を指定するだけです。

ここで大切なことは「エクセルの使い方」ではありません。「幾何平均は何をやっているのか」、「なぜそうしたのか」を理解することです。これを理解しないでエクセルで計算しても、まわりの人にこの幾何平均の意味を説明できません。まわりの方は幾何平均の意味がわからなければ、その数字の意味がわからず、納得しません。

さあ「GEOMEAN」を使って実際に計算してみましょう。

小数点以下は少し細かくして、小数点第3位までとしましょう\*1。次のようになりましたか？

	第2週	第3週	第4週	第5週	第6週	第7週	第8週	第9週	第10週	第11週	第12週
Aコラーゲル	1.18	1.06	1.64	0.65	1.02	1.20	1.20	0.91	1.10	1.00	1.09
Xヨーゲン	0.93	1.94	0.39	2.41	0.89	1.01	1.01	1.30	1.18	0.67	0.93
	幾何平均										
Aコラーゲル	1.071		⇒週平均7.1%の伸び								
Xヨーゲン	1.036		⇒週平均3.6%の伸び								

(図表1-7)

結果はAコラーゲルの「週平均伸び率」が7.1%、Xヨーゲンが3.6%で、Aコラーゲルの方が「2倍伸びている」となりました。

あなたが図表1-3のグラフから感じた“伸び”とフィーリングが合いましたか？

\* 1. 細かいことが気になる人がいるかもしれません。Aコラーゲルの第2週の値は1.18と小数点第2位までとなっているのに、幾何平均を小数点第3位としてよいかということです。この第2週～第12週のデータは、エクセルの表示上で小数点第2位としているだけです。エクセルで幾何平均を取っている時は、このAコラーゲルの第2週の伸び率は「1.175」、第3週の伸び率は「1.0638...」...として計算しています。

## 統計学的に言ってそう考えるしかない

ここまでを、統計っぽく整理してみましょう。

母集団は2つ、「Aコラーゲルの商品力」と「Xヨーゲンの商品力」です。ここからそれぞれ標本として、84個の日販という「売れ行き」がサンプリングされています。

「伸び」という商品力を表わす統計量として選んだものは、「平均伸び率」です。統計処理した結果、それぞれの「“伸び”に関する商品力」（平均伸び率）の期待値は7.1%、3.6%です。

つまりAコラーゲルはこれから先、7.1%の伸びが期待でき、一方Xヨーゲンは3.6%の伸びと考えるのが妥当です。

これに誰も反論することはできません。もちろんこれから先、その通りに「伸びる」とは限りません。

しかしこのデータから“統計学的に言って”そう考えることができます。この「統計学的に言って」がキーワードです。未来を当てているわけではありません。「数字を統計処理した結果がこうなった」と言っているだけであり、「そう考えるのが妥当」もつとえば「そう考えない理由がない」ということです。

## “ちがい”を感じる？

山田さんは図表1 - 3のグラフから、「伸び」とともにもう1つの商品力を感じています。それが「ブレ」です。

あなたも感じたでしょう。どう見てもAコラーゲルはブレが小さく、Xコラーゲルはブレが大きいと感じます。

このブレを統計では「バラツキ」と表現します。バラツキも「伸び」と同様に、何とか「統計量＝数字」で表わして、その期待値を推定してみましょう。

9ページのa社の納豆パックの例に戻って考えてみます。

a社の納豆パックの標本から得られた標本値は、「245、262、258、235」で、先ほどは統計量として平均値250を選択しました。

一方同じサイズの納豆パックを生産しているb社のパックも、サンプリングして豆の数を数えてみると「208、292、221、279」という標本値で、平均値はa社と同じ250でした。

このa社とb社の各4つずつの元の数字をもう一度よく見てください。

a社「245、262、258、235」とb社「208、292、221、279」です。

同じ“感じ”ですか、ちがう“感じ”ですか。

誰でも“ちがい”を感じると思います。だからa社、b社という2つの母集団も平均値は同じでも、その状態は“ちがっている”と考えられます。

## ブレに人類の知恵を使おう

そうです。その“ちがい”が、山田さんの感じた「ブレ＝バラツキ」です。

さて、このブレをどういう統計量で表わしていくかです。

こんな時は自分で考える必要はありません。山田さんと同じように悩んだ人は過去にたくさんいて、その中の賢い人が「皆が納得できるブレの計算方法」を考えているはずです。それが統計という人類の知恵です。

人類が出した結論は、ブレの統計量に「標準偏差」というものを用いることです。

標準偏差は「各数字がどれくらい平均値から離れているか」（これを偏差という）を「平均する」ものです。

納豆パックの例であれば、a社の標本の偏差は次のようになります。

平均値が250ですので、これを各標本値から引いて、 $(245 - 250)$ 、 $(262 - 250)$ 、 $(258 - 250)$ 、 $(235 - 250) = -5$ 、 $12$ 、 $8$ 、 $-15$ です。

しかしこのまま4つの偏差を平均すると、0になってしまいます。平均だから当然といえば当然です。

そこでこの偏差を2乗（2回かける）して（こうしてマイナス符号を取ってしまいます\*1）から、平均を計算します。この「マイナス符号を取るために2乗する」という“手”は、統計でよく使われます。

この「偏差の2乗を平均したもの」を「分散」といいます。そのうえで2乗したものを戻すために、分散の平方根\*2（2乗根のこと。ルートという。もう14ページで使ってしまいましたが。）を取ります。これが標準偏差です。（逆に言えば、標準偏差の2乗が分散です。）

ルートは中学校で全員が勉強しました。 $\sqrt{9}$ とは $\sqrt[2]{9}$ のことで、「何を2回かけると9になるか」という質問です。答えは「3」です。 $\sqrt{9} = 3$ です。

a社、b社の標本の標準偏差は次のように計算できます。

$$\text{a社の標本の標準偏差} = \sqrt{\frac{(245-250)^2 + (262-250)^2 + (258-250)^2 + (235-250)^2}{4}} \doteq 11$$

$$\text{b社の標本の標準偏差} = \sqrt{\frac{(208-250)^2 + (292-250)^2 + (221-250)^2 + (279-250)^2}{4}} \doteq 36$$

a社の標本の標準偏差は11、b社は36です。つまりb社の方がバラツキが3倍くらい大きいということになります。

a社= (245、262、258、235) とb社= (208、292、221、279) という2組の数字から受けるブレ=バラツキの感じが、「b社がa社の3倍のバラツキ」というのは、どうですか？あなたの“フィーリング”と合いますか？

バラツキにどんな統計量を使うかについて、最初は色々意見があったようです。そしてなるべく人間の直感に合うようにするために試行錯誤して、「標準偏差を使おう」という結論になりました。ブレを感じたら、迷わず標準偏差を使って、ブレを数字で表わしてみましよう。

\* 1. マイナスとマイナスをかけ算するとプラスになります。  $(-2) \times (-2) = 4$ です。全国民が中学までに習いました。

\* 2. 平方根、ルート 正確に言うと「平方根=ルート」ではありません。9の平方根とは「何を2回かけると9になるか」ということですので、3の他に-3もあります。  $(-3) \times (-3) = 9$ です。つまり9の平方根は3と-3の2つです。  $\sqrt{9}$  は9の平方根のうち“プラスのもの” (=3) を意味します。

またルートはエクセルを使えば、先ほどのPOWERという関数で、指数に「1/2」を入れれば計算できますが、SQRTという関数でも直接計算できます。

## 理解できたら、標準偏差もエクセルで計算しよう

標準偏差も幾何平均同様に、きちんと定義しておきましょう。

$a_1, a_2 \cdots a_n$  という  $n$  個の数字の平均が  $\bar{a}$  の時、標準偏差 ( $\sigma =$  シグマで表わす) は次のようになります。分散は  $\sqrt{\quad}$  の中を表わし、 $\sigma^2$  となります。

$$\sigma = \sqrt{\frac{(a_1 - \bar{a})^2 + (a_2 - \bar{a})^2 \cdots + (a_n - \bar{a})^2}{n}} = \sqrt{\text{分散}}$$

$\Rightarrow \text{分散} = \sigma^2$

さあAコラーグルとXヨーゲンで考えてみましょう。

標準偏差も幾何平均同様に、エクセルの関数で簡単に計算できます（でも標準偏差の意味を理解しないうちに、“簡単”に使ってははいけません）。

関数検索で「標準偏差」と入れてSTDEVP\*<sup>1</sup>を選びます。そのうえでAコラーグル、Xヨーゲンにつき、それぞれ84個ずつのデータを指定します。これで終わりです。

計算結果はAコラーグルが7.9、Xヨーゲンが15.8となりましたか？

「XヨーゲンのブレはAコラーグルの2倍」。いい感じですね。

\* 1. **STDEVP** エクセルでは標準偏差の関数として**STDEVP**だけでなく、**STDEVA**というものも用意してあります。細かいことですが、これについて簡単に説明します。

母集団から標本をサンプリングして、標本の標準偏差を計算してみると、同じ母集団からのサンプリングでも、標本数が多いほどその値が大きいくことがわかっています。そうすると母集団にあるすべての要素をサンプリングして計算した標準偏差（これが母集団の標準偏差）が一番大きくなると考えられます。だから標本の標準偏差に比べ、母集団の標準偏差は大きいと考えた方がよいはずです。

そうすると母集団の要素数が極めて大きく、そこから数少ない標本をサンプリングする時（例えば日本の全国民が母集団で、そのうちの100人を標本としてサンプリング）は、“標本の標準偏差”よりも“母集団の標準偏差”は「大きく推定した方がよい」といえます。

そこで前ページの標準偏差の定義にある分母を  $n$  ではなく、 $n-1$  として少し小さい数字にし、計算結果を少し大きくした方が、数学的には母集団の標準偏差を推定する期待値として、より適切なものとなります（これをきちんと数学者が証明しました）。

**STDEVA**という関数は、このような考えのもとに、「各標本の偏差の2乗の和」を  $n-1$  で割ってから平方根を取り、これを母集団の標準偏差の期待値としようというものです。

しかし標本数  $n$  が十分大きければ、**STDEVP**でも、**STDEVA**でも、結果はほとんど同じになります。 $n = 10000$ で割っても、 $n-1 = 9999$ で割っても、結果はあまり変わりありません。

ビジネスに使う時はこんな説明をするのがわずらわしいし、相手も理解できないので、より直感的な先ほどのSTDEVPを使った方がGoodです。

ケースの84個のデータをSTDEVA ( $n-1$  で割るタイプ) を使って標本の標準偏差を計算してみましょう。結果は小数点以下第2位を四捨五入すると、Aコラーグルが7.9、Xヨーゲンが15.9となり、STDEVP ( $n$  で割る) を使った結果である7.9、15.8とほとんど変わりません。

## さあプレゼンテーション！

山田さんはB社C店バイヤーの加藤さんに、次のようにプレゼンテーションしました。

「先日お預かりしましたAコラーグルとXヨーゲンの12週の日販データについて、弊社にて統計処理した結果を報告させていただきます。

両商品とも売り値は12週の間、ずっと200円で固定ですので、販売個数の比較を行いました。日販数は84日間の平均で、Aコラーグル18.6、Xヨーゲン25.2となります。

しかしお手許のグラフを見ていただければわかる通り、両者とも、時とともに伸びていくことが感じ取れます。つまりレイアウト変更後の7月以降の日販数は、この平均値よりも大きいことが予想されます。

そこで対前週の伸び率を出して、これを平均してみました。正確に言えば、このような伸び率の平均に、統計学で使う幾何平均というものです。結果はAコラーグルが7.1%、Bヨーゲンは3.6%となり、Aコラーグルの方が約2倍の伸びを示しています。7月以降もAコラーグルはXヨーゲンの2倍の伸びを示していくと推定されます。

またお手許のグラフで、加藤さんもお気づきになったと思いますが、両商品の間にブレのちがいが見られます。

Aコラーグルは安定した販売量を示していますが、Xヨーゲンにはブレが目立ちます。

そこでこのブレを標準偏差という数値で計算してみました。平均値からどれくらいブレているかを平均するものです。結果はAコラーグル7.9に対して、Xヨーゲンは15.8と2倍の値となっています。

標準偏差というブレが小さいということは、コンスタントに売れていることを意味します。つまりAコラーグルの方が2倍コンスタントに売れているということです。

分析の結論としましては、Aコラーグルはコンスタントで、かつ伸びが大きいということです。これはAコラーグルのファンが着実に増えていることを示しています。Aコラーグルを一度買ったお客様が次も購買し、かつ新しいお客様も増えているということです。これがAコラーグルの商品力と考えます。

したがってパワーのあるAコーダルの陳列量を増加して、C店のお客様にAコーダルをさらに強く認知してもらえば、7月以降も販売量はますます増えていくと考えられます」

どうです。山田さんはライバル社に勝てそうな気がしませんか。

c社は工場向けの機械を販売するメーカーである。機械の部品は協力会社である部品メーカー数社へ発注し、c社では主にその組立てを行っている。

c社では量産する機械だけでなく、顧客のリクエストに応じてオーダーメイドの機械も作っている。この機械の部品の中には、標準的なものだけでなく、特殊な部品を必要とするものも多い。

c社が特殊部品を発注する部品メーカーとしては現在4社あり、基本的には見積合せ\*1によって発注先を決めてきた。特殊部品といっても過去の部品のカスタマイズ\*2であり、発注してから大体1ヶ月程度でc社へ納入されている。

c社は自社を中心とするサプライチェーン\*3の構築を検討しており、SCM\*3推進室というセクションを新たに作った。木下はそのSCM推進室の部品メーカー担当に配属された。

木下は上司である室長から呼ばれ、次のような指示を受けた。

「当社のサプライチェーンでは、特殊部品の協力メーカーは1社に絞り込みたい。このパートナー会社に求められるものは、コストよりもスピードと信頼性だ。コストはサプライチェーンという仲間になるわけだから、チェーン全体での原価を考えなくてはならない。彼らのコストを下げても、うちのコストが上がってしまったら意味がない。スピードと信頼性を指標として、どの会社がよいか、候補を選定してほしい」

木下は思った。

「スピードと信頼性か。スピードは発注リードタイム\*4だよな。信頼性か、こっちはなかなか難しいなあ。

発注リードタイムはこの1年間の各社の実績を資材部に確認すればわかる」

			(単位:日)
d社	e社	f社	g社
38	28	33	16
43	26	28	23
23	32	34	45
37	40	38	17
18	18	29	22
21	35	30	40
20	51	31	17
40	38	27	21
35	29	25	23
52	16	27	37
29	29	32	17
28	28	33	15
33	38	28	40
15	19	37	22
23	39	31	
	28	28	
	34		

(図表1-8)

木下さんの調べたリードタイムの実績は上のような結果でした。  
さあ木下さんの立場でこの課題にチャレンジしてみましょう。

- \* 1.見積合せ 提示見積価格のもっとも安い会社へ発注すること
- \* 2.カスタマイズ 顧客の要求に応じて、一部変更すること
- \* 3.サプライチェーン、SCM 商品供給先の複数の会社が、1つの会社のようになって需要（最終的な商品利用者）に対応していくことをサプライチェーンという。ここではc社と部品メーカーが1つの会社のようになって、c社の顧客の需要に対応していくことをサプライチェーンと表現している。これを実現する仕組のことをサプライチェーンマネジメント、略してSCMという。
- \* 4.発注リードタイム 発注してから納入するまでの時間

まずは「数字活用ケースその1」でやったように、d社、e社、f社、g社のこれまで、そしてこれから先の発注リードタイムを母集団と考えます。だから母集団は4つです。

標本は図表1-8にある各社の発注リードタイムです。この発注リードタイムを使って、サプライチェーンを組んだ場合のスピード、信頼性を表わす統計量を考えます。

スピードの統計量としては平均値をとるのがノーマルでしょう。

信頼性は“安定感”のようなものと考えれば、その統計量としてバラツキ（標準偏差）が妥当でしょう。

4社の発注リードタイムの平均値、標準偏差を計算してみましよう。

次のようになるはずです。（やり方は先ほどの山田さんのケースでやったのでわかりますよね）

会社	d社	e社	f社	g社
平均値	30.3	31.1	30.7	25.4
標準偏差	10.2	8.6	3.5	10.0

（図表1-9）

平均値を見ると、g社が他社よりも約5日短く、残り3社はd社、f社、e社の順で短いのですが、この3社はほぼイコールと考えてもよいでしょう。

母集団としては（サプライチェーン後の発注リードタイムは）、g社のスピード（平均値）が速く、d社、e社、f社はほぼ同じと推定されます。

一方信頼性の統計量である標準偏差は、f社がもっとも小さく、安定しています。

この結果から木下さんの室長への説明は、例えば次のようになるでしょう。

「4社を1年間の発注リードタイムの実績で比較してみました。

スピードの指標として発注リードタイムの平均値を考えると、d、e、f社がほぼ30日で同じ、g社が25日で、他社より5日ほど早いと考えられます。

もう1つの信頼性という指標は、発注リードタイムのバラツキ、すなわち標準偏差で考えてみました。

結果、f社が3.5となり、d社10.2、e社8.6、g社10.0という他社に比べ  $\frac{1}{2} \sim \frac{1}{3}$

となっています。f社のブレが小さい、つまりf社の仕事は、他社と比べ2倍～3倍安定していると考えられます。

私としてはサプライチェーンの主旨から考えて、スピードよりも安定度という信頼性を重視し、パートナー相手の第一候補をf社と考えます。」

# 明日を読む

## 予測のやり方をマスターしよう

「未来を予測する」

これは現代ビジネスマンに求められる大きなテーマです。

この時、大切なことは未来を当てることではありません。未来なんて神様しかわかりません。

自分が予測した未来を、上司や顧客などまわりの人に納得してもらうことが大切です。これさえできれば、あなたはまわりから「未来を見通す力がある」と評価されます。

さああなたも予測という統計テクニックを学んで、「明日が読めるビジネスマン」に変身しましょう。

## 数字活用ケースその2

### —— 未来の数字に合意する ——

中村さんは上司から、来月の目標販売台数について、取引先と合意するよう言われました。

中村さんの会社から見れば、この目標数字を大きくしたいところです。一方、取引先から見れば、この目標数字を小さくしたいところです。

さあどうやって折り合ったらよいのでしょうか。

## 普通にやれば売れる数字を考えたい

D社は2年前、デジタルビデオカメラとパソコンを合体したパソカメという新しい商品を開発した。テレビコマーシャルを大々的に打ち、これを主に大都市にある大型家電販売店を通して販売していった。パソカメは発売以来2年間で順調に売上を伸ばしていた。

D社では販売力を高めるために、特にパソカメの売行きのよい販売店とは、新しいスタイルの販売契約を結ぶことを考えていた。それはD社と販売店の間で、月ごとに目標販売台数を設定し、その台数を達成したら、販売店に対してD社から報奨金を支払うというものである。

この報奨金のシステムは、D社営業企画部販売支援課の中村が担当することになった。

中村は営業企画部長から、次のような指示を受けた。

「今回の報奨金のポイントは、払う金額ではなく、目標販売台数の設定にある。客観的に考えて、『売れてもおかしくない数字』を出して、うちと販売店が互いに納得できる目標を作ってくれ。

まずは首都圏で一番パソカメを売ってくれているE販売店からチャレンジしてほしい」

中村は思った。

「うちとしてはもちろん目標販売台数のラインを上げたいし、販売店から見ればそのラインはできるだけ低いほうがいいに決まっている。

となると、部長が言ったように『パソカメの実力から考えて、普通にやれば売れる数字』を何とか出したい。

『普通にやれば売れる数字』の根拠となるのは、当然『これまで何台売れたか』だな。と言うよりもこの数字以外使うものがない」

### 平均値では小さすぎる

中村は営業企画部から、E店における発売以来24ヶ月のパソカメの月別販売台数入手し、次のようにエクセルに入力した。

そのうえで25ヶ月目の目標販売台数について考えてみた。

「24ヶ月間の平均は803台か。  
23ヶ月目の916台、24ヶ月目の  
1064台に比べると、ずいぶん小  
さいなあ。

あたり前か。うちのパソコン  
は伸びているからなあ。

しかしそうなると目標販売台  
数はどうすればいいんだ。平均  
が使えないなら、25ヶ月目の  
『普通にやれば売れる数字』な  
んてどう考えればいいんだろ  
う」

10ヶ月目の販売台数は  
824台ということ

月数	販売台数
1	480
2	504
3	568
4	520
5	512
6	584
7	820
8	728
9	896
10	824
11	752
12	804
13	792
14	984
15	932
16	844
17	960
18	888
19	852
20	1020
21	924
22	1112
23	916
24	1064
平均	803

25ヶ月目はいくつと考えるべきか？  
これまでの平均は803台。しかし  
23ヶ月目が916台で  
24ヶ月目が1064台で  
25ヶ月目が803台と考えるのは…

(図表2-1)

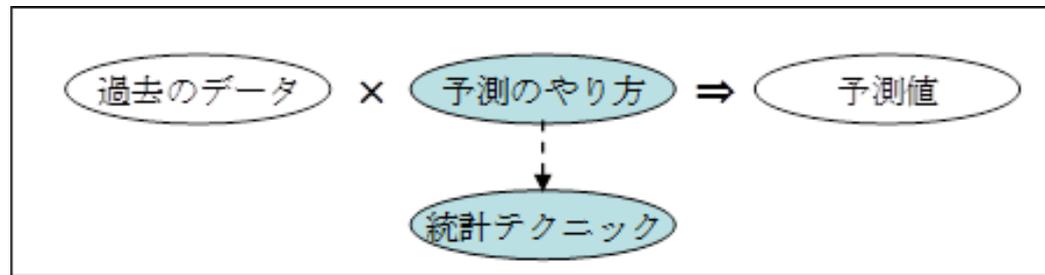
## 予測はやり方を説明すれば合意できる

さあ、この24か月間のデータを使って、25ヶ月目の「普通にやれば売れる数字」を予測しましょう。

目標はE店の店長が「そうですよね。25ヶ月目の目標数字はこうするしかないですよ」という予測値を出すことです。

まずは予測という“仕事”について考えてみましょう。

予測は次のような構造を持っています。



予測とは「過去のデータ」と「予測のやり方」を使って、予測値を出す仕事です。

中村さんのケースでは、エクセルに入っている販売台数という24個の数字が「過去のデータ」であり、これを使って何らかの「予測のやり方」で、「25ヶ月目の販売台数」という「予測値」を計算するものです。

こう考えれば、予測値を説明する相手（このケースではE店店長）には、予測値（25ヶ月目の明日の数字）そのものではなく、予測のやり方を説明し、これに納得してもらえばよいことになります。

「予測値」に納得してもらおうのは“気持ち”（「もう少し大きくしたい」「もう少し小さくしたい」）が入って、難しいかもしれません。でも「予測のやり方」に納得してもらおうのは、それほど難しいことではありません。予測のやり方に“先人たちの知恵”である統計テクニックを使えばOKです。統計テクニックは人類が出した予測に関する最終結論であり、反論も議論の余地もありません。

つまり中村さんはE店店長がわかるように、そのやり方を説明するだけでOKです。

## 予想では水かけ論

ちなみに、これ以外の方法で未来の数字を出すことを予想といいます。予想は過去のデータを使わなかったり、過去のデータを使ったとしても、やり方をはっきり決めず、人間のカンだけで数字を出すものです。

「カン」で予想すると、「どうしてその数字になったか」を説明できません。カンとはそういうものです。だから予想した時は、相手とはやり方ではなく、「予想した数字」そのものについて議論するしかありません。そして相手からは「私のカンではそうならない」と反論されてしまいます。

こうなると決め手は、“声の大きさ”や“2人の力関係”となってしまいます。中村さんのケースのように、相手が「他社の人」という“対等”な関係であれば結論が出ません。

中村さんが自分のカンで予想して、E店の店長に「24ヶ月目が1064台ですので、25ヶ月目は3%くらい伸びると考えて、区切れのいい所で1100台を目標にしましょう」と申し出れば、店長は「いや24ヶ月目の1064台は、何かの拍子にたまたま多く出ただけです。23ヶ月目の916台あたりをベースとして考えると、私のカンでは930台あたりがいい線じゃないかと思います」と言われてしまいます。こうなるとどちらのカンが当たっているか（もちろん25ヶ月目が終わってみないとわかりません）で、水かけ論となってしまいます。

### ぐうの音も出ない予測をしたい

統計テクニックを使ってE店の店長が反論できない、つまりぐうの音も出ないような予測をやってみましょう。

先ほどの24個の数字を1ヶ月目から順に見ていってください。ここからどんな感じを受けますか？

中村さんが感じている（「うちのパソコンは伸びているからなあ」）ような、「増加傾向」にあることがわかんと思います。

この「増加傾向」を表現するために、まずはグラフを書きましょう。これには、散布図（プロット図ともいう）というグラフが最適です。

散布図とは、2つの項目について「ペアとなっている数字」がある時、これをたて軸と横軸に取り、該当する所に点を打っていくものです。

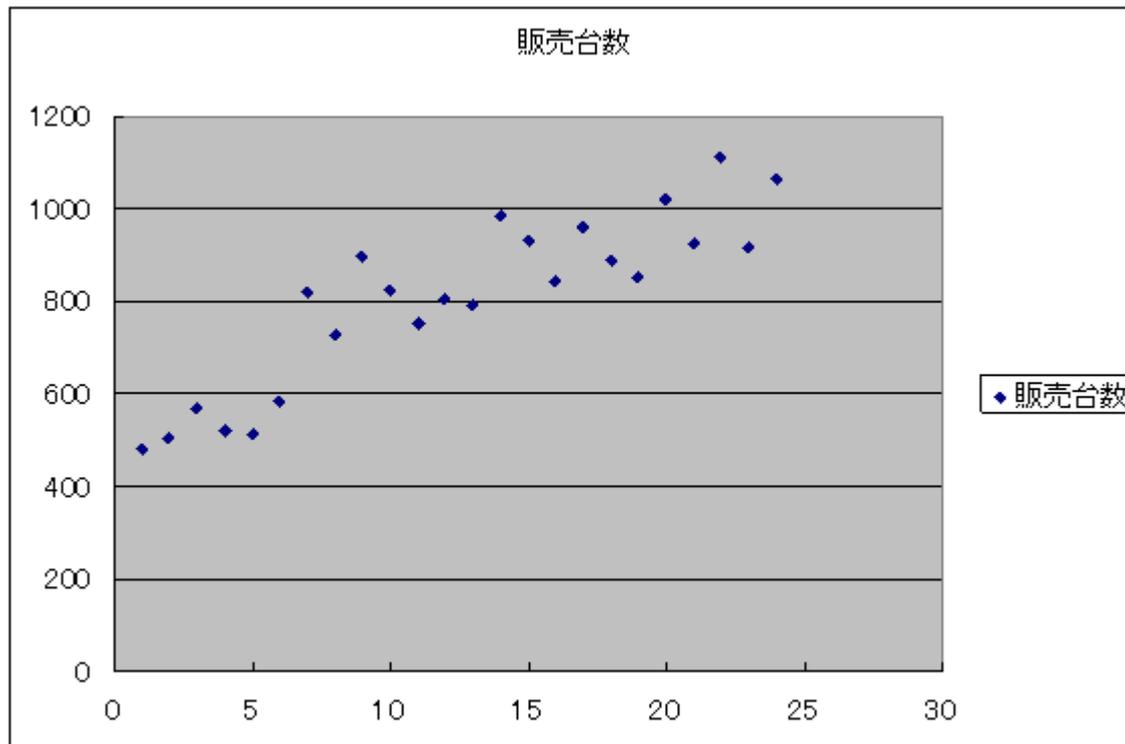
ケースで言えば、図表2-1のエクセルの表にある「月数と販売台数」というペアです。このペアの数字を、横軸に月数、たて軸に販売台数を取って、点を打っていきます。

これもエクセルに数字が入っていれば、簡単に書けます。

まずは図表2-1のエクセルの表の「月数」と「販売台数」の2列を「データ範囲」として指定します（もちろん平均は入れません）。

その上で先ほどの折れ線グラフの時と同様に、「グラフ」のメニューから「散布図」を選びます。いくつかの「形式」が出てきますので、「点を打っているだけのタイプ」を選びます。

すると次のようなグラフが出ます。



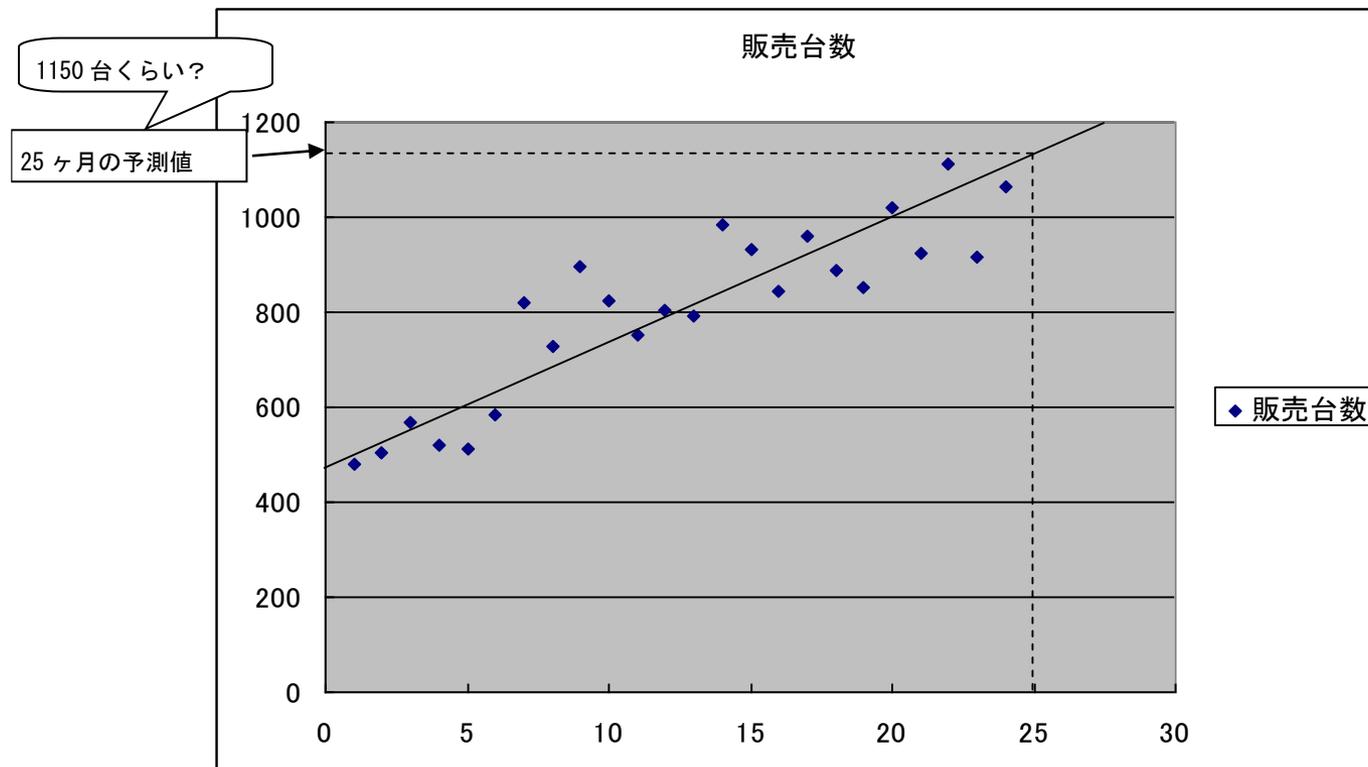
(図表2-2)

### 右肩上がりを直線で表わす

このグラフに何かが見えませんか？右肩上がりの「伸びていく線」が見えませんか？

これが先ほどの「増加傾向」の正体でしょう。

あなたが「この点を見て感じた直線」を、エクセルの表に定規を使って引いてみてください。



(図表2-3)

この直線が引ければ、25ヶ月目を予測できます。

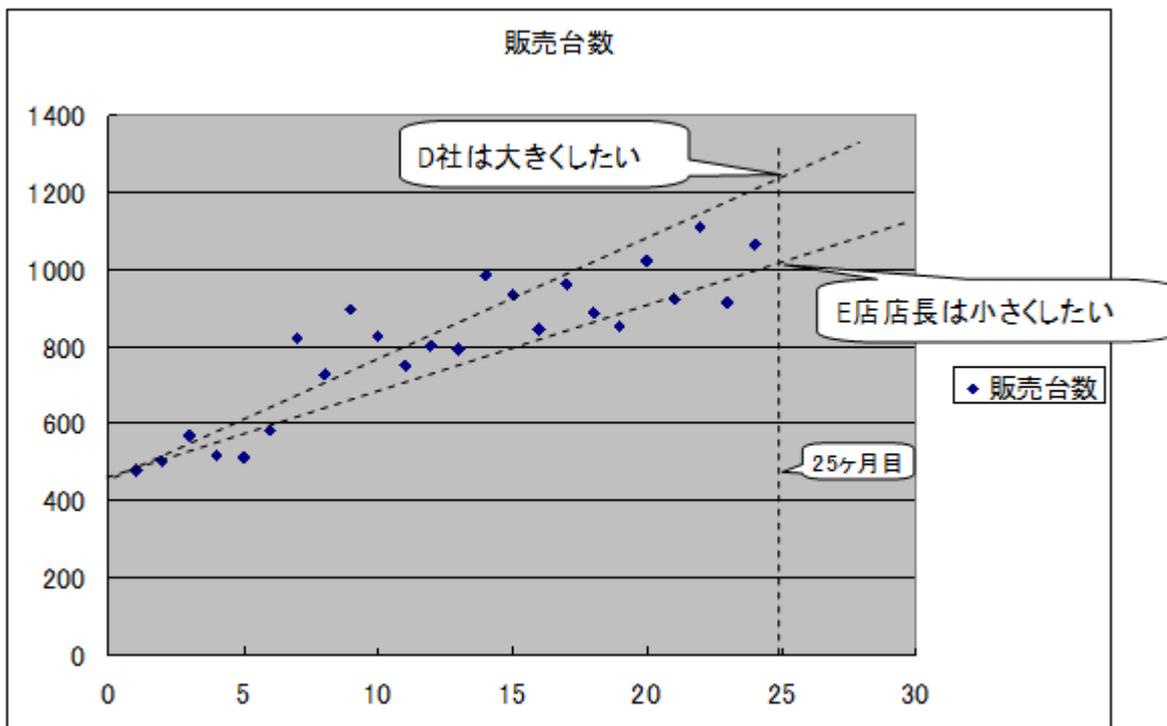
横軸の「25」（25ヶ月目）から、たてに直線を引き、引いた直線とぶつかった所で、横に線を引いていき、その台数を、たて軸の数値で読み取ればOKです。上の直線であれば1150台くらいと予測されます。

さらには26ヶ月目、27ヶ月目だって予測できます。

## 大きくしたい、小さくしたい

しかしこの直線の引き方は、人によってさまざまです。

そうすると、どうしてもD社はたくさん売ってほしいので、直線の傾きを大きくして、予測販売台数を大きくしたくなります。逆にE店店長は目標を達成しやすいように、傾きを小さくして、予測販売台数を小さくしたくなります。



上図のように直線を引くと、D社は1200台くらい、E店店長は1000台くらいと、その差が大きく出てしまいます。こうなると議論の焦点は「傾き」となってしまう、また水かけ論です。

中村さんとしては、人間の気持ち（大きくしたい、小さくしたい）が入らない、「これしかない」という直線を引きたい所です。ケースの中で営業企画部長が言っている「客観的」というやつです。

### これしかない直線が引けた！

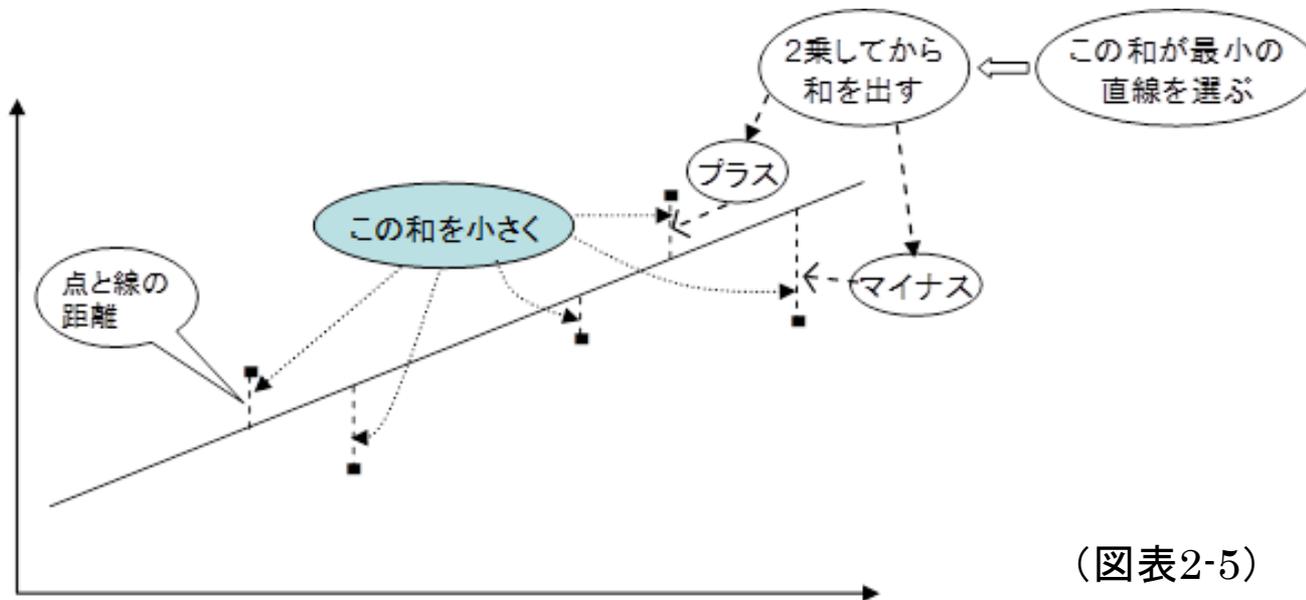
ところであなたは、上の直線を引く時どうやって引きましたか？

「点と点の間くらいをねらって、上下に点がうまく分かれるように引こう」なんて思いませんでしたか？

これを予測のやり方として、しっかりと定義し、誰が引いても同じ直線になるようにします。

それは『散布図にある24個の点から直線までの距離の和』がもっとも小さい直線を引く」ということです。

これには誰も反論できないと思います。



(図表2-5)

しかし単純に引き算で距離を計算すると、線の上にある時と下にある時でプラスとマイナスが出てきてややこしいといえます。

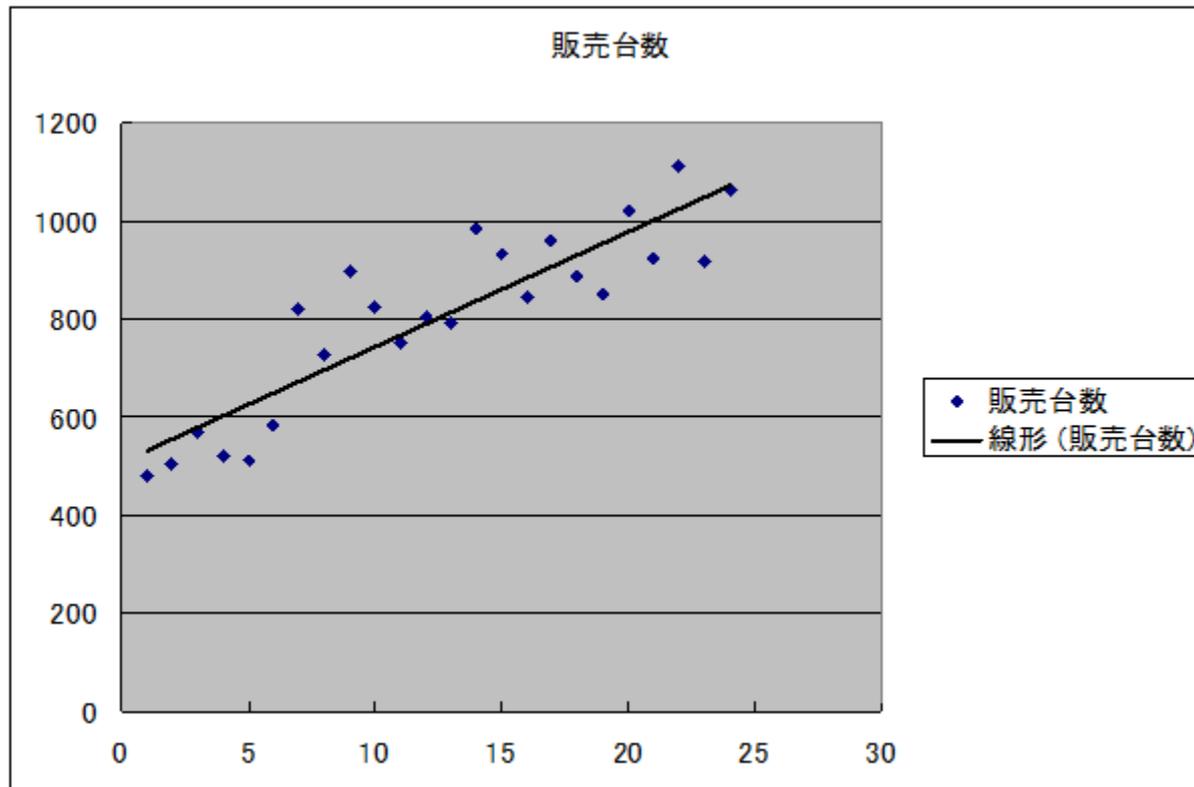
そこで19ページの標準偏差で使った“手”を用います。つまり距離をプラスとするために、その距離を2乗してから足し、その和を最小にする直線を引くようにします。このやり方を最小2乗法といいます。

さあこれをどうやって計算して、直線を引くかですが、ここまでしっかり定義すれば（そしてあなたが理解すれば）、あとはコンピュータがやってくれます。

エクセルでやってみましょう。

エクセルで出力された図表2-2の散布図の、どこかの点にカーソルをあわせて右クリックすると、メニューが出てきます。ここで「近似曲線の追加」を選びます。

そのうえで「線形近似」（「傾向を直線で表わすこと」を意味している）を選べば、下のように線を引いてくれるはずです。



(図表2-6)

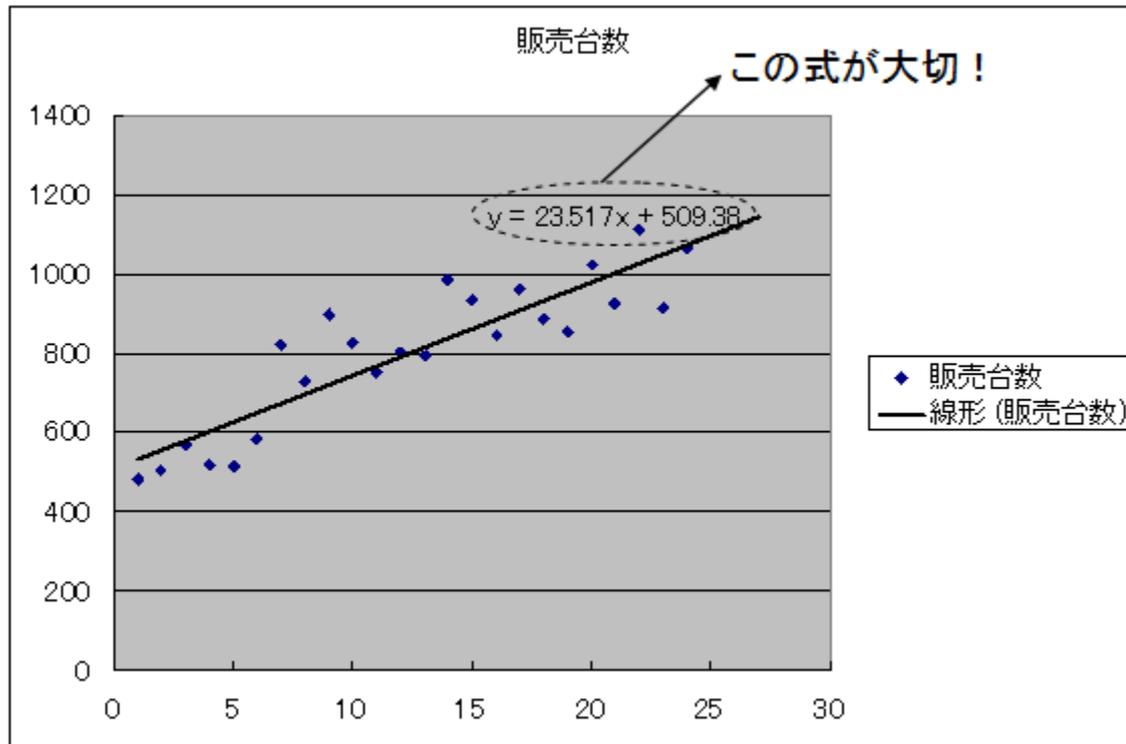
これが最小2乗法で引いた直線、つまり「各点からの距離の和がもっとも小さい直線」という「これしかない直線」です。

### 直線を式で表わせば、明日が読める

しかしこのままでは自分で直線を伸ばして、25ヶ月目の数字を読み取らなくてはなりません。こんなめんどうな仕事もエクセルがやってくれます。

この直線のどこかにカーソルをあわせて、右クリックし、出てきたメニューから「近似曲線の書式設定」を選びます。次に「オプション」のメニューで「前方補外」を“3単位”（直線を右に「何単位＝何ヶ月」まで伸ばすかということ。）くらいにセットし、「グラフに数式を表示する」をチェックすれば、次のようなグラフになります。（先ほどの「近似曲線の追加」の所で「オプション」を選んでもこの操作ができます。）

3ヶ月先の27ヶ月目までの直線を伸ばしたものです。



(図表2-7)

ここに表示されている「 $y = 23.517x + 509.38$ 」という式は、この直線の「たて軸（ $y =$ 販売台数）と横軸（ $x =$ 月数）の関係」を表わすものです。

小数点以下はあまり意味がないので「 $y = 23x + 509$ 」と考えると、次のような関係を意味しています。

$$\text{販売台数の予測値} = 23 \times \text{月数} + 509$$

したがって25ヶ月目の予測台数は「 $23 \times 25 + 509 = 1084$ 台」となります。そうです。25ヶ月目は1084台というのが、ぐうの音も出ない予測値です。

これがLesson2のテーマである「明日を読む」ということです。

### 反論なき予測値

中村さんは次のようにE店の店長へ説明しました。

「お手元の図表1は、E店における24ヶ月間のパソカメの販売台数です。

これを月ごとの販売推移のグラフで表わすと、図表2のようになります。これでおわかりのとおり、E店においてパソカメは明らかに増加傾向にあります。

そこでその傾向を直線で表現してみました。グラフの各点からの距離がもっとも小さくなるように、コンピュータで直線を引くと、図表3のようになります。

この直線を使って、25ヶ月目を予測すると、販売台数は1084台となります。

当社としてはこの1084台を目標台数として、これをクリアした時に、販売報奨金を出したいと思います。」

h社はさまざまな企業を顧客としているコールセンターであり、顧客企業の商品に関する消費者からの問い合わせに、h社の電話オペレーターが回答している。

i社は家電メーカーであり、自社商品の操作説明、クレームなどさまざまな消費者対応をh社に委託している。

新製品の家電品は、発売してからしばらく消費者からの問合せが数多く発生し、それが時間とともに減っていくのが一般的である。

そこでh社とi社の間では、新製品の週当たりの問合せが100件を切るまでは、専任オペレーターを5名確保して優先的に対応する。そして週当たりの問い合わせが100件を切った時点で、この専任オペレーターによる優先対応をやめて、すべてのオペレーターが通常の間合せとして対応することとしている。

i社の新製品Yは販売後30週が経っており、消費者からの問合せ件数は以下のとおりであった。

週当たり問合せ件数

(単位:件)

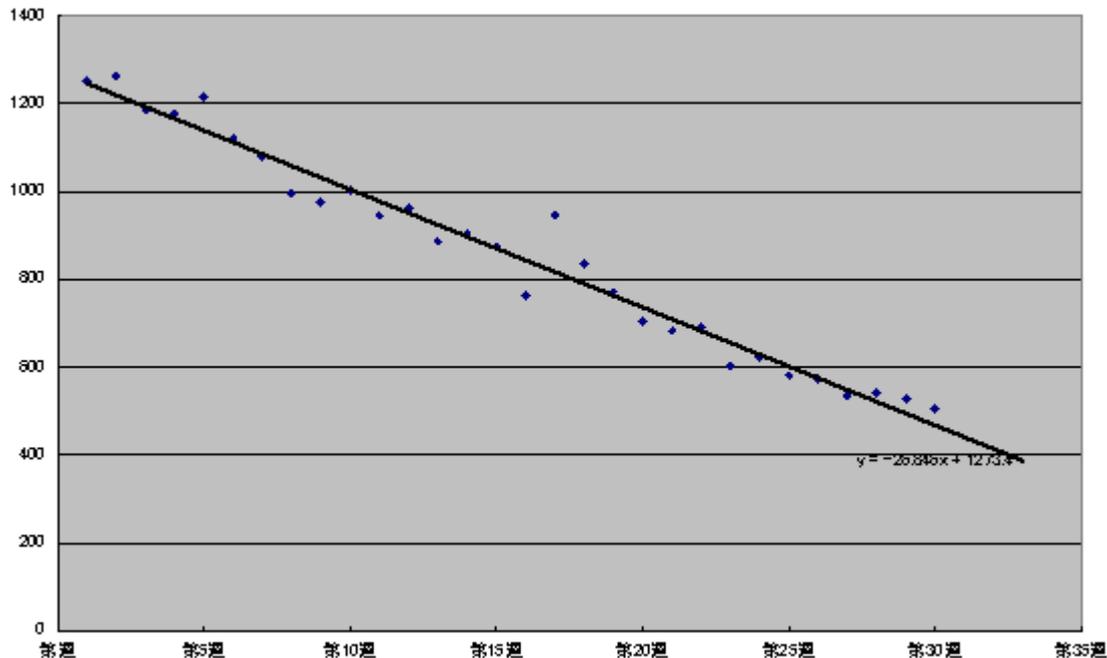
第1週	1251
第2週	1264
第3週	1185
第4週	1176
第5週	1215
第6週	1120
第7週	1080
第8週	995
第9週	976
第10週	1003
第11週	945
第12週	962
第13週	886
第14週	905
第15週	876
第16週	762
第17週	945
第18週	834
第19週	772
第20週	703
第21週	682
第22週	692
第23週	603
第24週	622
第25週	582
第26週	572
第27週	535
第28週	542
第29週	528
第30週	506

(図表2-8)

このペースで行くと、何週目になったら100件を切って、専任オペレーター制をやめることになる应考虑すべきでしょうか。

これは先ほどのケースの「増加傾向」とは逆に「減少傾向」です。この「減少傾向」を直線で表わし、未来を予測します。ただ今度は時間を決めて（ケースでは25ヶ月目）予測値を出すのではなく、予測値になる「時間」を求めるものです。

まずは上の30個のデータをエクセルに入れ、35ページでやったように散布図を書き、直線を引いて、式を出してみましよう。次のようになるはずです。



(図表2-9)

ここで大切なのは、下に出ている式です。  
小数点以下を四捨五入すると、以下のようになります。

$$y = -27x + 1273$$

ここで $x$ は「週」、 $y$ は「問合せ件数」です。そこで問合せ件数が「100」、つまり $y=100$ の時の $x$ を求めます。

$$100 = -27x + 1273$$

中学でやった方程式です。

$$27x = 1173$$

$$x = 43.4\cdots$$

$x$ は週なので整数です。

したがって44週目となります。44週目となって、問合せは100件を切ると予測されます。

44週目の問合せ件数の予測値は「 $-27 \times 44 + 1273 = 85$ 件」です。  
ちなみに43週目は「 $-27 \times 43 + 1273 = 112$ 件」となり、「100件は切らない」と予測されます。

「数字活用シーンその1」のAコラーグルとXヨーゲンの例に、もう一度戻ってみましょう。

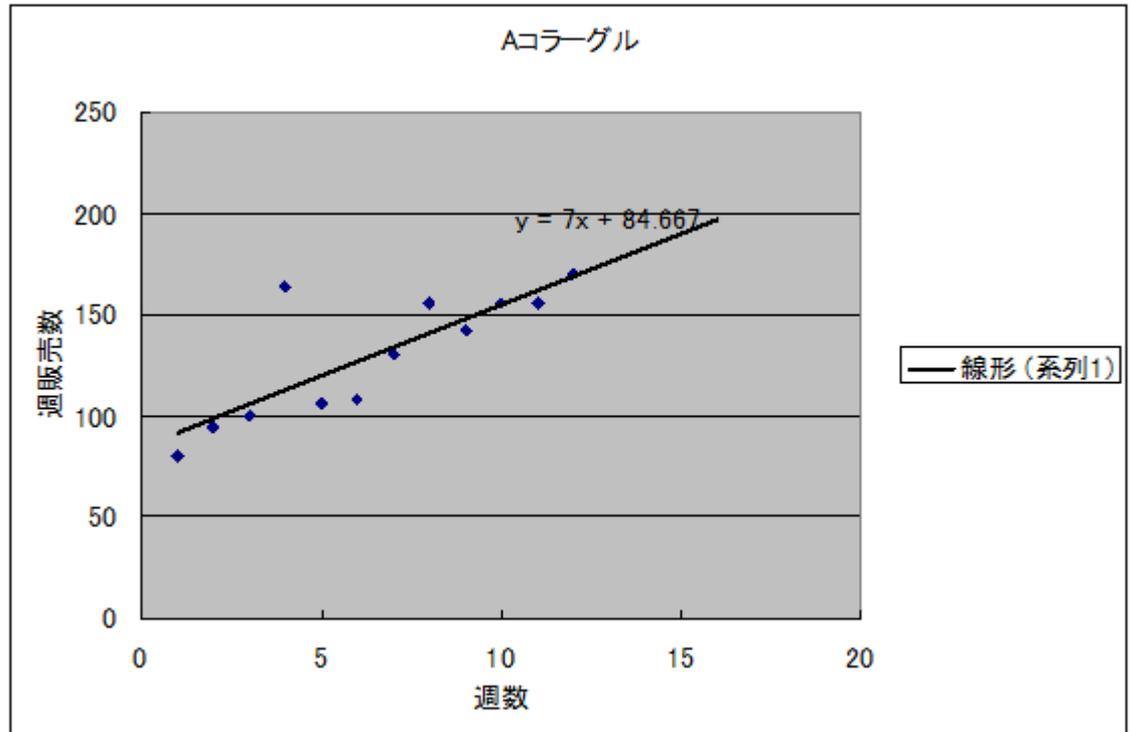
両方の商品がこのままの伸びを続けるとして、レイアウトを変えた最初の週である「7月4日の週」（「3月7日の週」から数えて18週目）には、Aコラーゲン、Xヨーゲンそれぞれの週の販売個数がいくつになるかを考えてみましょう。

まずはAカラーグルから行きましょう。  
12週のデータをエクセルに並べます。

週	第1週	第2週	第3週	第4週	第5週	第6週	第7週	第8週	第9週	第10週	第11週	第12週
Aカラーグル	80	94	100	164	106	108	130	156	142	156	156	170

(図表2-10)

先ほどやったように、  
上の2行のデータを指定してグラフの散布図を選び、「線形近似」「グラフに数式を表示する」とやっていくと次のようになります。



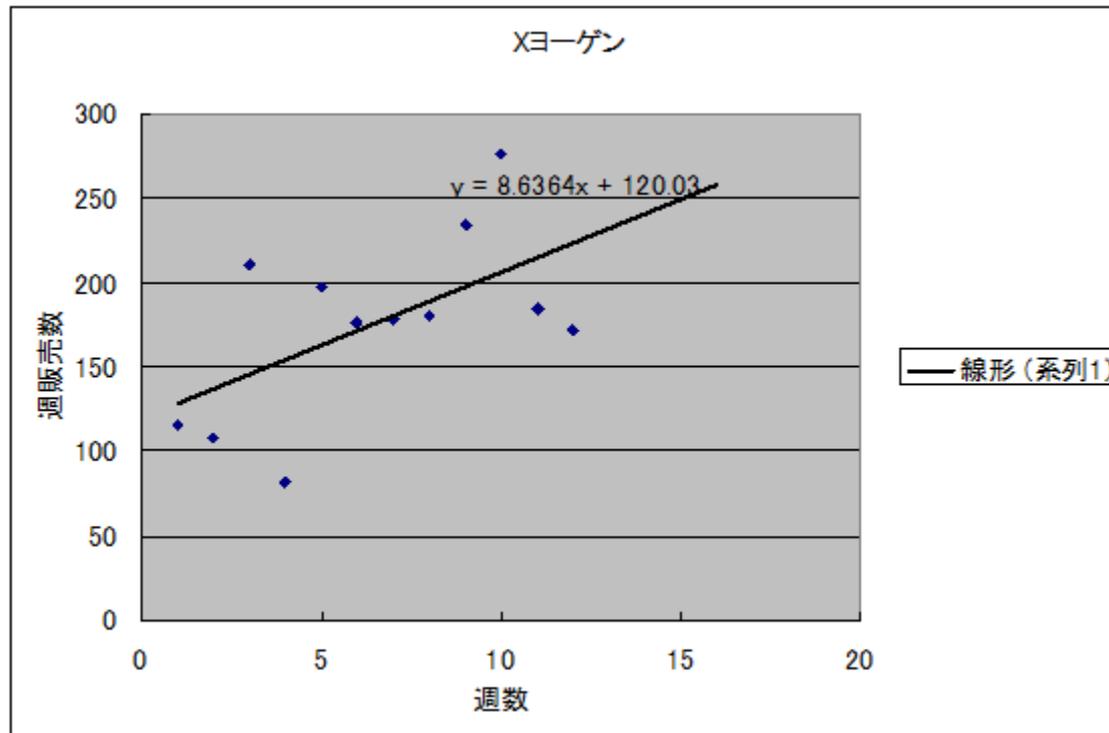
(図表2-11)

上の式から「週販売数＝ $7 \times \text{週} + 85$ 」となります。

したがって18週目の販売数は「 $7 \times 18 + 85 = 211$ 」となります。つまり211個と予測されます。

同様にXヨーゲンでもやってみましょう。

週	第1週	第2週	第3週	第4週	第5週	第6週	第7週	第8週	第9週	第10週	第11週	第12週
Xヨーゲン	116	108	210	82	198	176	178	180	234	276	184	172



(図表2-12)

「週販売数 $=8.6 \times \text{週} + 120$ 」ですので、これに「週 $=18$ 」を入れると275個となります。

しかし7月4日の週で、Aコラーグルの販売数が211個で、Xヨーゲンの販売数が275個というのは何か変です。12週目のAコラーグル170個、Xヨーゲン172個から考えると、そんなに差がつくとは思えません。と言うよりも、伸び率から考えて、AコラーグルがXヨーゲンを逆転してもいいようにも感じます。

このように自分の直感と合っていない時は、「予測のやり方」をチェックしましょう。

Aコラーグルの散布図（図表2 - 11）を見て下さい。この直線は各点の傾向を直線がよく表わしていると思います（少しはずれている点も、あるにはありますが）。

しかしXヨーゲンの散布図（図表2 - 12）を見て下さい。この直線は、はたして「各点を表わす線」として、意味があるのでしょうか。

「各点からの距離の2乗の和が最小の直線」を引いているのですが、それでも直線から各点が離れすぎています。もっとも距離が小さいように直線を引いても、こんなに離れています。

それは各点の数字が大きくブレているからです。

ブレが大きい、つまり標準偏差の大きい数字を、このような直線を使って考えるのは適切ではありません。まあ「ブレが大きすぎて予測不能」といったところ  
です。

見方を変えると、ブレの大きい商品は「明日が読みづらい」ということを意味しています。

一方Aコーラは「明日を読みやすい商品」といえます。

店舗から見れば「明日を読みやすい商品」の方が、「店に何個在庫すればよいか」がわかりやすく、「販売しやすい」といえます。Xヨーゲンでは、店に何個置けばよいかよくわかりません。

この「明日の読みやすさ」もAコーラの商品力であり、山田さんとしては、アピールしたいポイントといえます。

# 数字と数字の関係を分析する

## 相関分析と回帰分析をマスターしよう

ビジネスにおいて「複数の数字項目について、その関係を分析する」ということはよくあることです。

こんな時は、相関分析と回帰分析という2つの統計テクニックが、きれいにキレ味鋭く決まります。

相関分析、回帰分析などと言うと、何だか“難しい数学”のような気がしますが、そんなことはありません。

すでにそのコア部分はLesson2で学んでいます。Lesson2では、時間と販売台数という「複数の数字項目についてその関係を分析」しました。つまり、もうLesson3の基本部分は、Lesson2で学習済みということです。

Lesson2をクリアしたあなたなら、Lesson3は一気にクリアできます。

あなたも相関分析、回帰分析という“統計の花形テクニック”を身につけて、そのキレ味の良さを実感してください。

## 数字活用ケースその3

- ・・・まだオープンしていない店の売上を予測する

高橋さんは「これからオープンする店の売上を考えなさい」と言われています。

既に出店した店舗のデータはたくさんあるのですが、これをどうやって使ってよいのかわかりません。

「オープンしていないうちに予測なんて・・・」

高橋さんは悩んでいます。

あなたならどうやって予測しますか？

どれくらいの数字になるのか、きちんと予測してくれ

F社は老舗の百貨店である。

百貨店の売上はバブル崩壊後、低迷していた。そこでF社は10年前に、食品・衣料などの「ショッピング」、喫茶・食事などができる「レストラン」、ゲームセンターなどの「アミューズメント」という3つの施設を融合した新しいタイプの店舗として、スーパーデパートという業態を開発した。

スーパーデパートは地方の中核都市を中心に、顧客の主力ターゲットを「30代の小さな子供のいる主婦」ととらえ、この10年間順調に出店を続けてきた。

高橋はF社に入社し、いくつかの百貨店の販売員およびバイヤーなどを経験し、今はF社スーパーデパート事業部店舗開発グループのリーダーである。

スーパーデパートのこれまでの出店は30店であり、現在、ある地区への新規出店を検討している。

出店のための商圈調査を専門のリサーチ会社へ委託し、その報告書が上がってきた。高橋はこれを見て「いける」という手ごたえを感じた。

そこで執行役員スーパーデパート事業部長の中山の所へ、商圈調査の報告書を持参して、出店のGoを得ようとした。

中山事業部長は百貨店の店長をいくつか経験し、その後百貨店事業部長となり、最近横すべりでスーパーデパート事業部へ異動してきた。

中山は商圈調査の報告書を見て、高橋へこう話した。

「地区を小さく区切って、商圈人口などを細かく書いているが、商圈人口だけでその店舗が行けるかどうかわかるのか。店舗フォーマット\*1によっても売上は大きく変わるんじゃないか。

スーパーデパートだってもう30店も出しているんだから、出店したらどれくらいの数字になるのかをきちんと予測してくれないと、我々経営サイドとしては出店の判断がつかない。

もっと言わせてもらえば、店舗フォーマットによって売上が変わるんだから、『こんなフォーマットなら、このような売上になる』というのを提示してくれないと、身動きが取れないだろう」

\*1.店舗フォーマット 店舗のスタイルのこと

**データはたくさんあるけど、どうしていいかわからない**

高橋は席に戻って考えた。

「売上を予測するといっても、まだオープンしてないんだから、どうしていいかわからない。まあこれまでの既存店の売上から考えるしかないんだろうけど。

事業部長の言っている『商圈人口だけでなく、店舗フォーマットによっても売上が変わる』というのはよくわかる。

売上に影響を与えるデータにはどんなものがあるのか考えてみよう。既存店舗なら大体のデータは手に入る。

商圈人口はもちろんだなあ。ただうちの場合、広域の全商圈\*1と、自転車や徒歩中心のコア商圈\*1に分けて考えている。この2つの商圈人口は大切だ。

それとうちの主力顧客は30代の女性だから、この人口も関係あるなあ。

後は店舗フォーマットか。

小売ゾーンの売場面積は大切だな。目的買い商品\*2の食品のアイテム数、衣料のアイテム数も必要かな。レストランゾーンは面積かな。アミューズメントゾーンは面積よりゲーム機の台数の方がいいかな。

これくらいかな。よし調べてみよう」

高橋は既存店舗30店に関し、データ調査を行い、図表3-1のようにエクセルに入力した。

「しかしこのデータから、どうやって予測したらいいんだろう・・・」

\* 1. **商圈** 商圈とはその店舗に来店する顧客の地理的範囲のこと。コア商圈とは、その店の売上の70～80%くらいを占める商圈。全商圈とはそれを含めた全体。コア商圈を1次商圈、全商圈からコア商圈を除いた部分を2次商圈ということも多い。

\* 2. **目的買い商品** 顧客がその商品を買いに来店したもの

	(百万円)	(人)	(人)	(人)	(㎡)	(種)	(種)	(㎡)	(台)
店舗NO	売上	全商圈内 人口	コア商圈内 人口	30代 女性人口	売場面積	食品アイテム 数	衣料 アイテム数	レストラン 面積	ゲーム機 台数
No1	982	24564	10284	1618	1100	4490	1283	550	47
No2	660	18226	8726	1285	1890	3710	890	513	51
No3	1466	61605	9951	3408	2000	6048	1178	546	66
No4	2067	74523	15212	1844	4770	5814	3363	728	97
No5	1125	48673	9159	3502	3433	3407	508	611	37
No6	1288	17209	11076	2596	3670	2305	862	724	86
No7	1972	52656	11127	1465	2280	5762	1222	513	40
No8	1998	18033	16610	2636	5522	9097	1096	753	30
No9	872	41622	12085	512	1400	2881	1814	586	65
No10	680	33385	6996	2281	816	4157	1484	389	34
No11	1569	36657	15334	968	1490	9504	791	748	60
No12	543	27875	5287	880	666	2964	1327	513	69
No13	1664	64736	13952	2548	2520	9081	915	537	40
No14	645	33396	11976	538	546	1887	992	523	49
No15	789	6193	7445	2790	2920	7124	2275	657	45
No16	1314	51305	8384	3361	1520	7315	2692	634	43

No17	1716	31135	11618	3537	1970	7660	3103	657	64
No18	1620	20274	17035	3110	3880	6279	1445	583	42
No19	567	30223	7297	661	2280	2236	380	486	44
No20	762	28886	9220	779	537	2795	1248	425	67
No21	1276	39864	13381	859	1043	4031	2502	730	59
No22	931	21221	13037	2904	3690	5296	1838	651	56
No23	1840	51309	17437	3825	2660	11187	2269	811	71
No24	1166	52060	6986	1093	2800	5575	2525	475	37
No25	1048	17210	11329	1144	1440	4649	2525	436	38
No26	234	3517	6970	1175	523	4027	174	421	49
No27	432	20070	7869	567	1520	5124	577	396	35
No28	1694	35168	10662	1465	4220	5202	2009	687	25
No29	354	10308	8559	1711	970	4348	204	399	21
No30	298	16662	8565	1463	487	3495	434	368	60
新規出店 (予定)	?	22546	9526	1530	1485	5000	1400	750	70

(図表3-1)

## 当てる数字と、当てるために使う数字

高橋さんの悩みを解決するには、回帰分析と相関分析という2つの統計テクニックを理解する必要があります。

まずは回帰分析からいきましょう。

統計の教科書には、「回帰分析の定義」をこう書いています。

「回帰分析とは、ある変数の値にもとづいて、他の変数を説明したり、予測したりするための手法である。この時、その値が用いられる変数を説明変数、説明あるいは予測の対象となる変数を被説明変数とよぶ」

「変数」とは数字がいく通りもあるもので、図表3 - 1の売上、全商圏人口...などはすべて変数です。一方、数字が変わらないものは定数といいます。円周率のようなものが定数です。

もちろん、こんな回帰分析の定義を覚えても意味がありません。プロローグで述べたとおり理解、体験、自信のPDSサイクルを回しましょう。

高橋さんのケースの分析に入る前に、簡単な例で回帰分析を理解してしましましょう。

ある消費者向け商品を販売している企業で、「顧客の年収を知りたい」と考えています。年収を直接アンケートで聞くのが難しいので、その人の「月あたりの外

食費」を聞いて、その結果から顧客1人ひとりの年収を当てようとしています。つまり「外食費が高い人ほど年収が高い」と考えて、「外食費で年収を当てよう」ということです。

上の回帰分析の定義でいえば、年収が被説明変数（「当てる数字」のこと）、外食費が説明変数（「当てるために使う数字」のこと）となります。

あたり前の話ですが、何もデータがない状態では、外食費を聞いても年収を当てることはできません。「年収と外食費の両方がわかっている人」が必要です。

もうわかりましたね。9ページで述べたように、この「年収がわかっている人」が「標本」で、これを含め「外食費についてアンケートをする顧客1人ひとりの年収」が母集団です。

### 前にやった「散布図で直線を引く」

例えば10人（説明上少なくしています）の外食費と年収が下のようにはわかったとします。

(単位:万円)

NO.	外食費	年収
1	2.5	520
2	2.8	600
3	5.0	950
4	1.5	420
5	4.0	750
6	3.2	700
7	3.5	620
8	2.3	630
9	1.8	400
10	3.0	650

No.1の人は年収520万円で  
外食費は月2万5千円

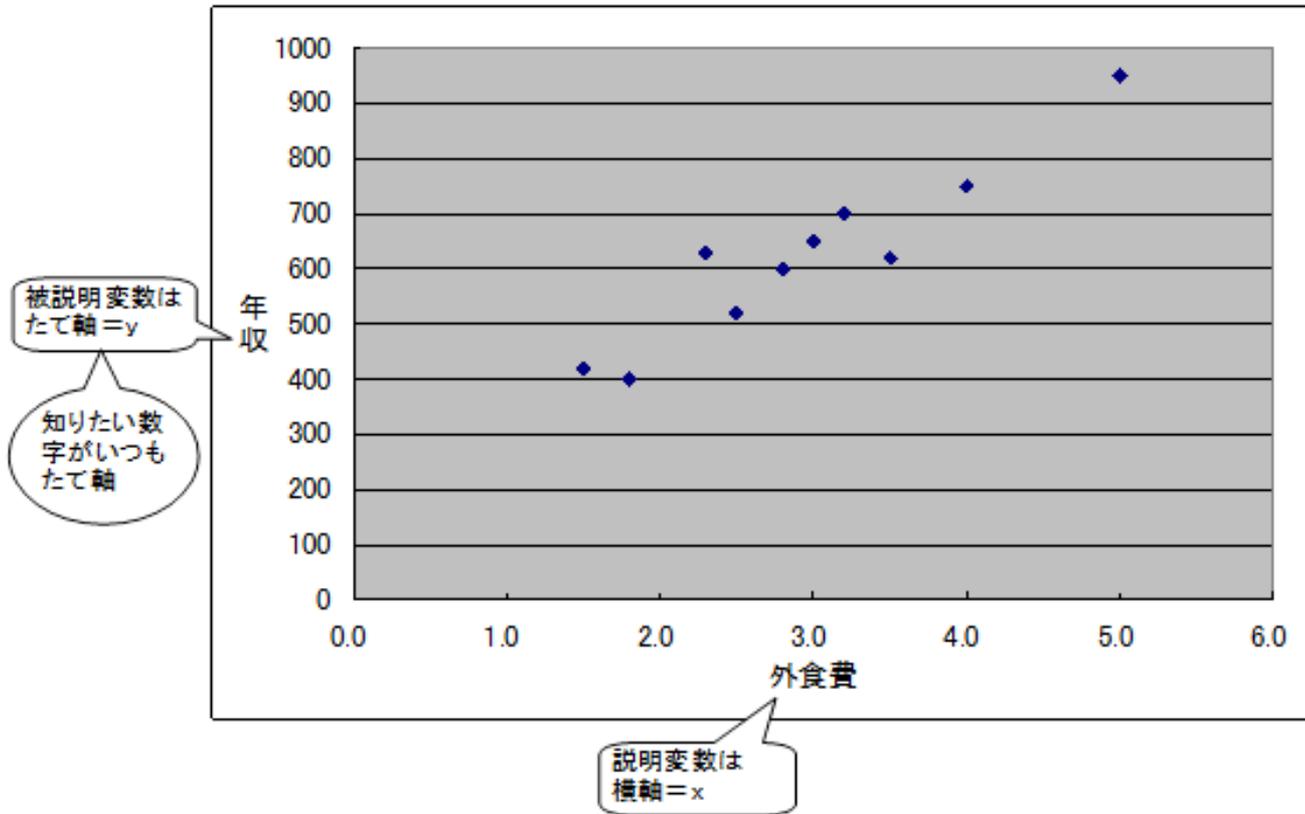
(図表3-2)

まずは上の「外食費」と「年収」を使って、35ページでやったような散布図を書いてみましょう。

必ず「説明変数」(外食費)を「横軸」(x軸)、「被説明変数」(年収)を「たて軸」(y軸)としてください。

このようなグラフを書く時は、いつも「知りたいデータ(年収)の方をたて軸にする」と覚えておきましょう。

エクセルで散布図を作ると、左側の列(上の外食費)が横軸(x軸)、右側の列(上の年収)がたて軸(y軸)となります。ですから左側の列に説明変数、右側の列に被説明変数を入れるようにしてください。

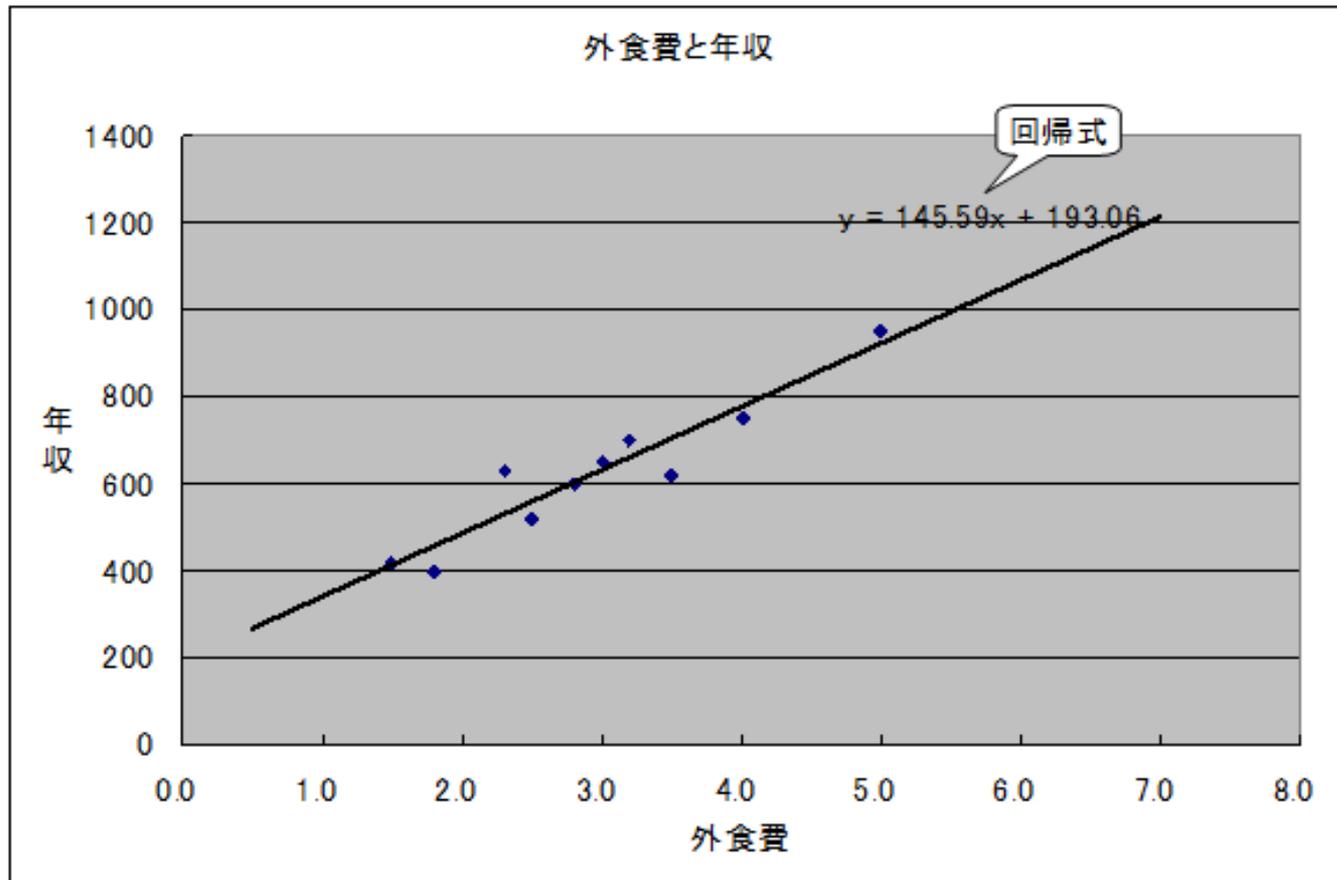


(図表3-3)

もうここまで来ると、すでに回帰分析をやっていることに気づくと思います。そうです。「数字活用シーンその2」の37ページでやったものです。あのケースは、「販売台数」が被説明変数、「月数」が説明変数の回帰分析でした。

したがってあの時と同様に、“直線”を引いて“式”を求めます。この“直線”のことを「回帰直線」、この“式”のことを「回帰式」といいます。

38ページでやったことを思い出してください。点を右クリックして、「近似曲線の追加」、「線形近似」、「グラフに数式を表示する」という操作です。

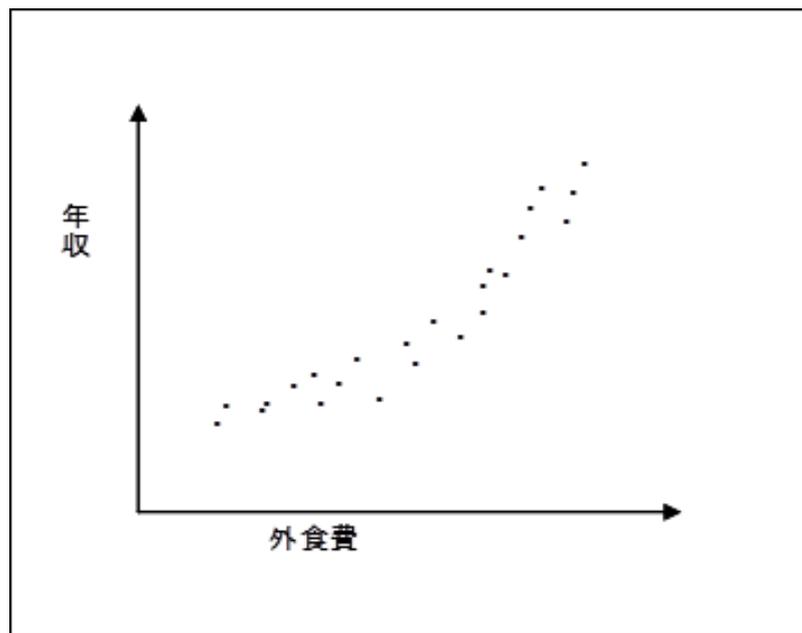


(図表3-4)

この回帰式は「 $\text{年収} = 146 \times \text{外食費} + 193$ 」ということを表わしています。  
したがって外食費が2万円の人の年収は、「 $146 \times 2 + 193 = 485$ 」となり、485万円と推定されます。

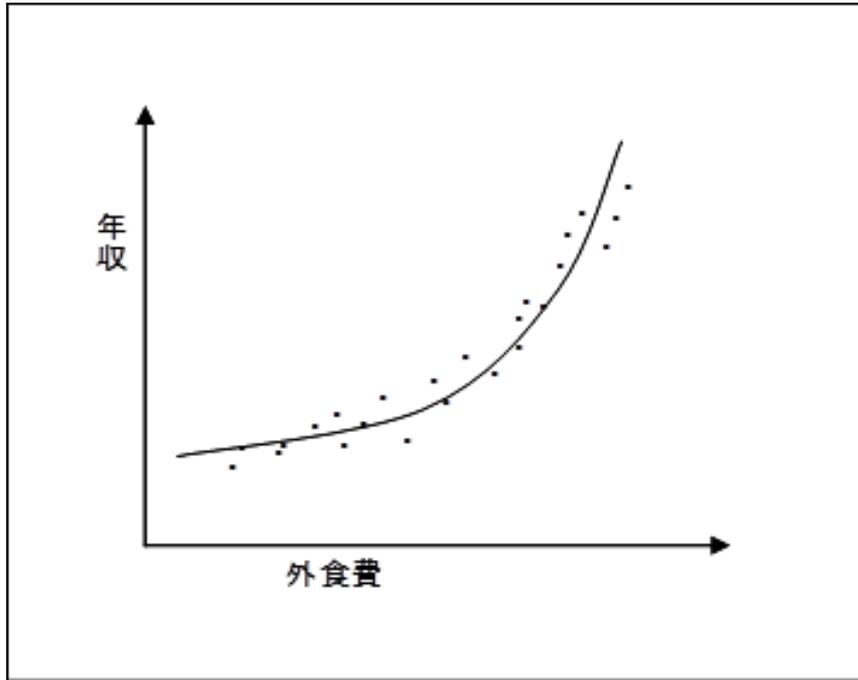
### 直線だけでなくもよいけど、ビジネスの世界では直線と決めよう

回帰分析で引く線は、直線でなくてもかまいません。例えば外食費と年収で散布図を書いてみたら、次のようになったとします。



(図表3-5)

この時は下のような曲線がフィットしているかもしれません。



(図表3-6)

これは指数曲線とよばれるものです。

指数曲線の外、エクセルでは「線のパターン」として対数、多項式、累乗といったものをサポートしています。エクセルで実際に見てみましょう。

ただビジネスで回帰分析を使う時は、「直線だけしか使わない」と心に決めていても、ほとんど問題ありません。

図表3-6でも指数曲線ではなく、直線を引いてもそれほどの違和感はありません。と言うよりもややこしい曲線を引くより、すっきりとした直線の方がかえってまわりへの説明力が高いといえます。

線を引くのが目的ではなく、まわりへ説明し、納得してもらうのが目的ということをお忘れなことです。

### 関係の強さを数字にする

高橋さんが困っているケースに戻りましょう。

このケースでは被説明変数（予測したいもの）は売上です。

では説明変数は？

この候補が全商圏内人口～ゲーム機台数まで8項目あり、このうちのどれかを選択することになります。

こういう時人間は、普通こう考えます。

「どれが売上に与える影響が大きい項目なんだろう？」。

ここに用いられるのが相関分析です。

統計の教科書には、相関分析の定義を次のように書いてあります。

「相関分析とは、2種類の母集団から標本を取り、その関係を分析することで、当該2種類の母集団の関係を推定するものである。ピアソンが考えた積率相関係数（これを単に相関係数ということが多いため。本書でもそうします）がもっとも有名である」

まあ要するに、標本を使って2種類の数字項目の関係を考えるのが相関分析、考えた結果を数字で表わしたもの（＝統計量）が相関係数です。

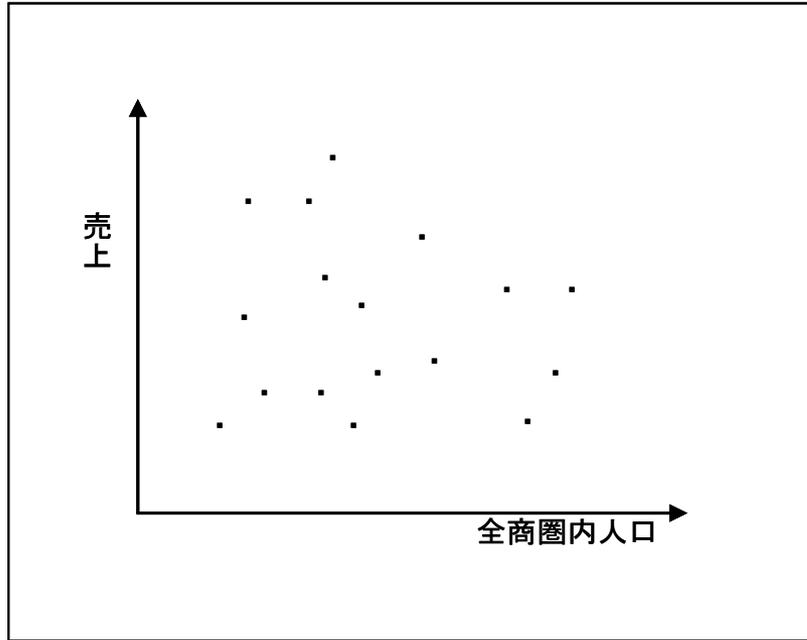
### 片方が大きくなると、片方も大きくなる関係

高橋さんのケースで相関分析を考えてみましょう。

ここでの相関分析の目的は「売上と関係の深い項目を見つけること」であり、それによって「売上の説明変数を見つけること」です。

売上と全商圏内人口という2つの数字を例にして考えてみましょう。

売上と全商圏内人口を散布図にしてみたら、次のようになったとします。

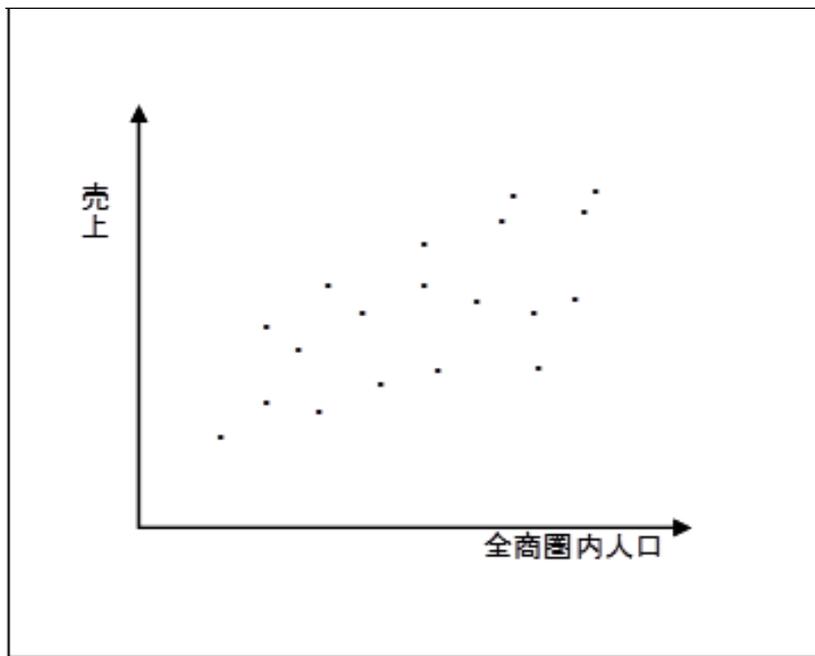


(図表3-7)

図表3-7では、どうも全商圏内人口と売上の間には「関係がなさそう」です。この状態を「全商圏人口と売上は相関がない」と表現します。「相関なし」の時は、全商圏内人口で売上を推定するのは適切ではないといえます。

このような「相関なし」の時に、相関係数が「0」になるようにピアソンさんは計算式を決めました。

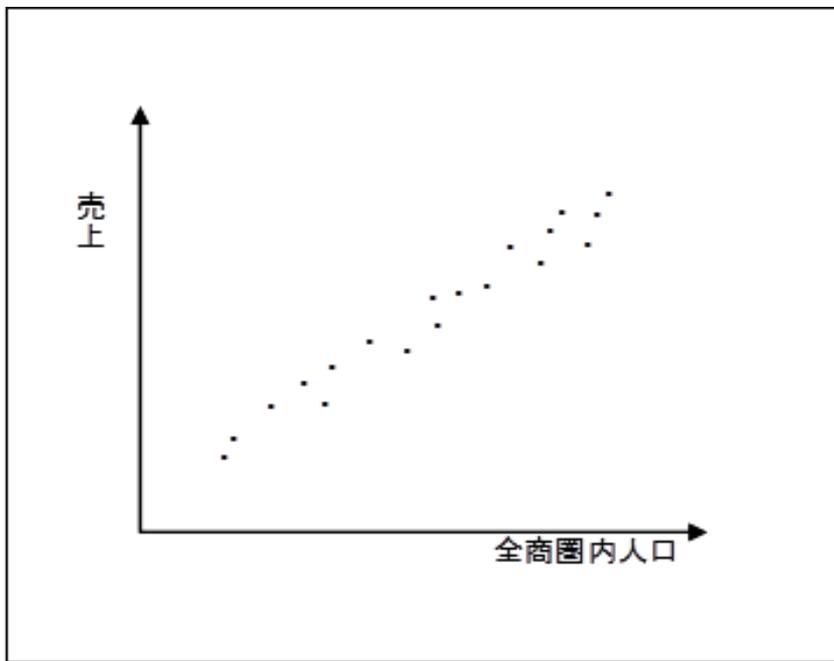
もし次のようになっていたらどうでしょうか。



(図表3-8)

これは明らかに、全商圏内人口が大きくなると、売上も大きくなっています。このように「片方が大きくなると、片方も大きくなる」という場合は「相関あり」（後で述べますが、正確に言うと「正の相関あり」）と表現します。正の相関があれば、全商圏内人口で売上を推定できそうです。

では次のような場合はどうでしょう。

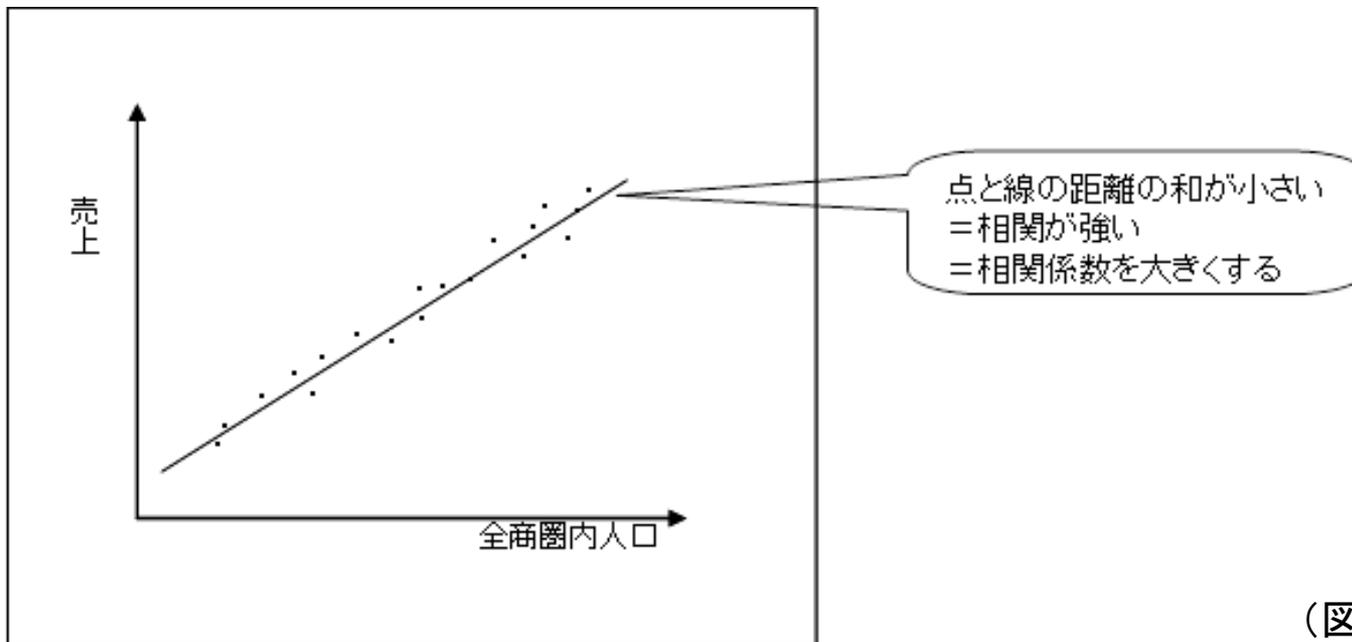


(図表3-9)

図表3-8より図表3-9は「強い相関」を感じると思います。

もしこの状態で先ほどの回帰直線を引くと、回帰直線と各点の距離が小さくなっているのがわかると思います。

つまり38ページの最小2乗法で考えた「点と線の距離の2乗の和」が小さければ、相関は強いことになります。これがピアソンさんが考えた相関係数の原点です。この距離の和が小さい時に相関係数を大きく（相関が強い）していけばよいわけです。



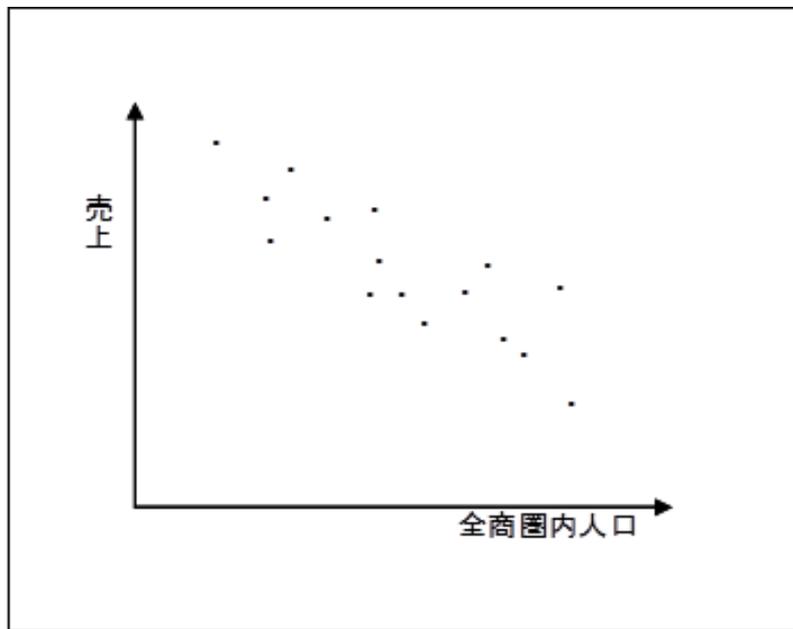
(図表3-10)

図表3-10のように、その距離がさらに小さくなると「かなり相関が強い」といえます。

そして直線と点が完全に重なっている状態の時に、相関係数が「1」となるようにしました。

### 片方が大きくなると片方が小さくなる関係

では次のようになっている場合はどうでしょう。



(図表3-11)

全商圏内人口が大きくなると、売上が小さくなっています。商圏が魅力的で、ライバル店が多くなって、こうなることがあるかもしれません。

先ほどの逆で、「片方が大きくなると、片方が小さくなる」というものです。これを「負の相関」といいます。

負の相関があっても、全商圏内人口で売上を推定できそうです。

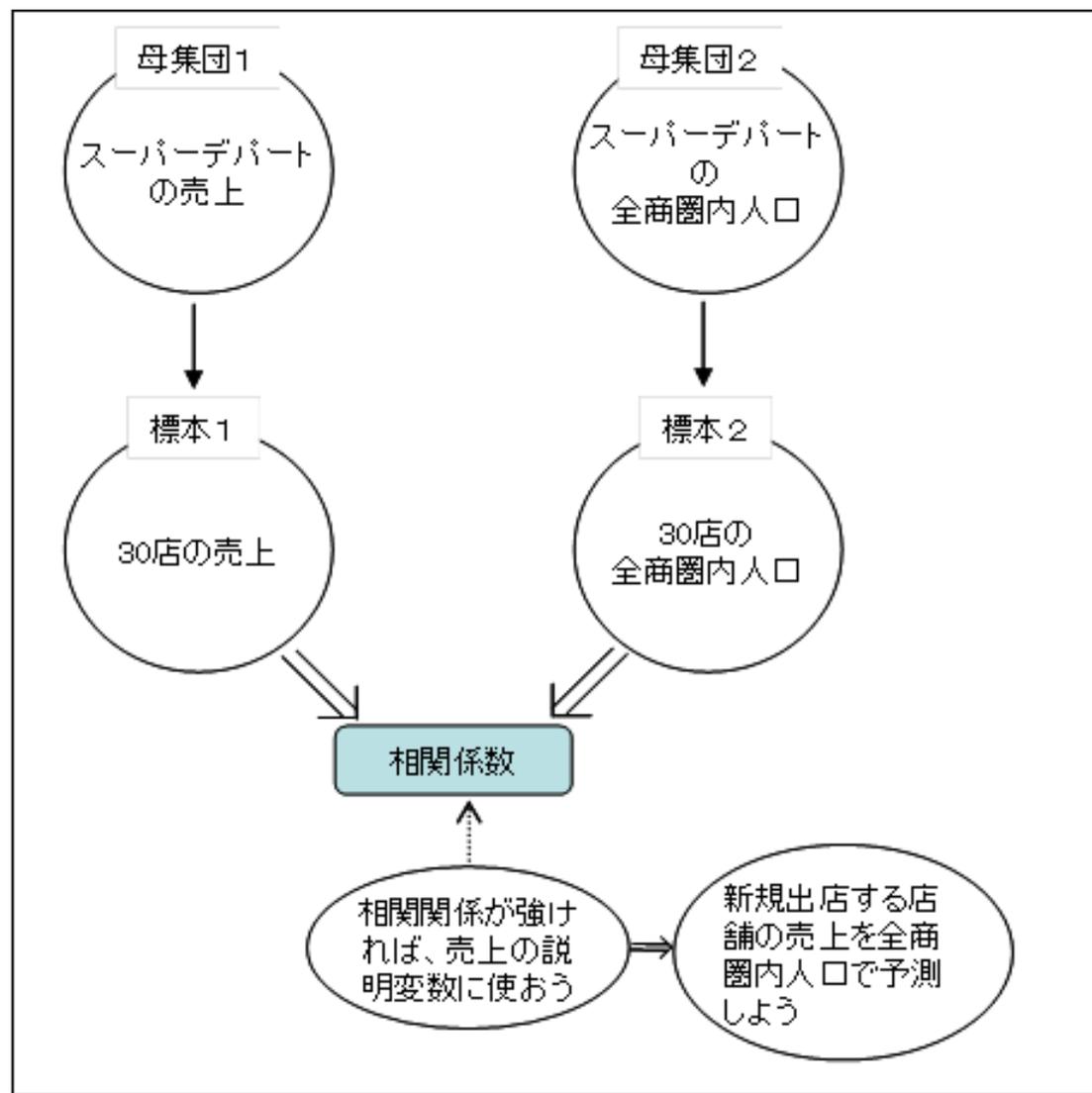
これも正の相関と同様に、点と線が完全に重なったものを $-1$ とし、離れてしまったらどんどん大きくしていった（絶対値\*1を小さくしていった）、つまり $-0.9$ 、 $-0.8\cdots$ とやっていった、「関係なし」となったら $0$ にすればいいわけです。

これがピアソンさんが考えた相関係数です。人間の「相関」というイメージを $1$ から $-1$ の数字で表わしたものです。

\* 1.絶対値 プラスの場合はそのまま、マイナスの場合はその符号をとった数値。「 $3$ 」も「 $-3$ 」も絶対値は「 $3$ 」

### 売上を何で予測しよう？

もうわかると思います。回帰分析の説明変数には、被説明変数との相関係数が「 $1$ か $-1$ に近いもの」を使えばよいことになります。



(図表3-12)

この相関係数もエクセルで簡単に計算できます。

図表3-1のエクセルの表を開いた状態で、「ツール」(または「データ」)⇒「分析ツール」(または「データ分析」)⇒「相関」とクリックします。入力範囲を聞いてきますので、「先頭行をラベルとして使用」をクリックして、「店舗NO」を除いた9列を表頭(売上、全商圏人口・・・というタイトル行)を含めて指定します。ただし新規出店の行は除きます。列が1つの項目となっていますので、「データ方向」を「列」にチェックして、OKをクリックします。これで次のような表が表示されるはずです。

	売上	全商圏内人口	コア商圏内人口	30代女性人口	売場面積	食品アイテム数	衣料アイテム数	レストランゾーン面積	ゲーム機台数
売上	1								
全商圏内人口	0.602659	1							
コア商圏内人口	0.730103	0.286337	1						
30代女性人口	0.480023	0.203207	0.315191	1					
売場面積	0.670066	0.210565	0.507894	0.470777	1				
食品アイテム数	0.660116	0.298552	0.573712	0.550444	0.382311	1			
衣料アイテム数	0.510053	0.393393	0.26913	0.255115	0.26977	0.328701	1		
レストランゾーン面積	0.719916	0.297079	0.674512	0.447467	0.620113	0.525897	0.462151	1	
ゲーム機台数	0.181524	0.277905	0.238128	0.059215	0.013142	-0.05917	0.292437	0.377714	1

ゲーム機台数とレストランゾーン面積の相関係数

(図表3-13)

ここでは売上を含めた9項目間の、それぞれの相関係数が計算されています。売上との相関係数は左端の1列です。

### 説明する数字はいくつあってもいい

一般に0.4以上の相関係数となる時は「かなり相関がある」と表現されます。売上との相関係数を見ると、ゲーム機台数を除く7項目が0.4を超えています。では何を説明変数として、売上との回帰分析を行えばよいのでしょうか。

「コア商圏内人口」がもっとも高い相関係数を示していますが、この数字だけで売上を予測したのでは、事業部長が納得してくれるとは思えません。

実は回帰分析では、説明変数を1つにしなくてもいいのです。説明変数はいくつあっても構いません。

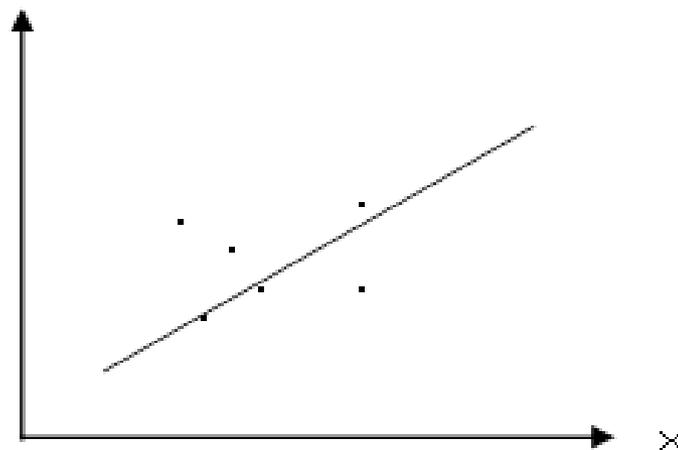
このように説明変数が複数あるものを重回帰分析といいます。一方、説明変数が1つのものを、これに対して単回帰分析といいます。

また重回帰分析に限らず、変数がたくさんあるような分析スタイルを多変量解析といいます。

### 空間に線を引く

重回帰分析の考え方を簡単に説明しましょう。

単回帰分析（説明変数が1つ）の時は、横軸（ $x$ 軸）とたて軸（ $y$ 軸）という平面で、点と直線との距離が小さくなるように、直線を引きました。



（図表3-14）

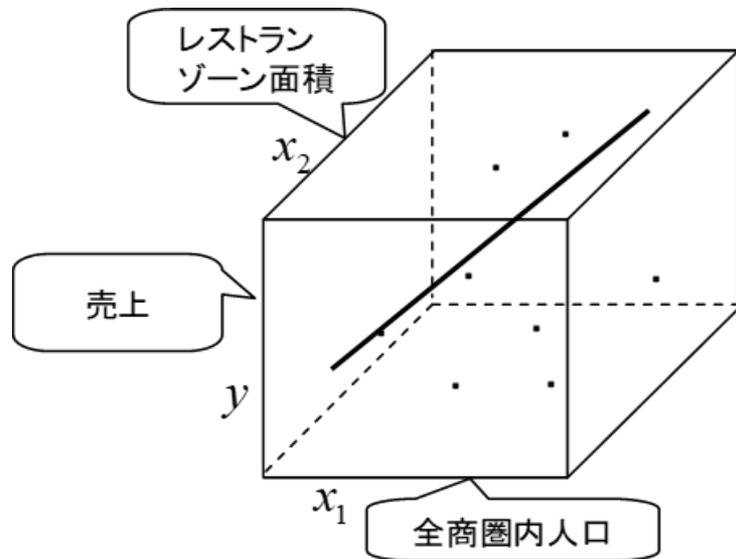
平面では、「横」（ $x$ ）と「たて」（ $y$ ）という2つの方向（これを次元という）を持っているので、2次元と表現します。

説明変数が2つ（ $x_1$ 、 $x_2$ ）となるとどうなるでしょうか。

こうなると3次元の世界になります。我々が生きている空間です。この空間には横（ $x_1$ ）、たて（ $x_2$ ）、高さ（ $y$ ）という3つの方向があります。この3次元の世界で回帰直線を引くことになります。

立方体のような箱をイメージして、そこに点が浮かんでいる感じです。

例えば全商圏内人口とレストランゾーンの面積の2つを説明変数とするなら、次のような感じです。



(図表3-15)

ここで箱の中に浮かんでいる各点からの距離の2乗の和が最小になるように直線を引けば、これが回帰直線です。

ここまで理解できれば、説明変数が3つでも4つでも・・・、いくつになっても直線を引けることはわかると思います。(頭で引いているシーンが浮かばなくても、「引ける」ことだけは理解できると思います)

## 説明変数を選ぶコツ

では説明変数は多いほどよいのでしょうか。数学的に考えれば「いくつでもOK」です。

ここではビジネスシーンで、重回帰分析を使う時のコツを3つ挙げておきます。

1つ目は仮に相関係数が大きくても、人間が「それは関係ないだろう」と思うものは、説明変数に選ばないことです。

例えば高橋さんのケースで、「売上」と「商圈内の病院の数」の相関係数が高くても、これを使わないことです。病院がたくさんある地域の店舗の売上が仮に大きくても、病院の数で新規店舗の売上を推定したのでは説得力がありません。

「売上」と「商圈内人口」の相関が高く、「商圈内人口」と「商圈内の病院の数」の相関が高ければ、「売上」と「商圈内の病院の数」の間の相関も高くなります。この場合は人間が直感的にわかりやすく、説明がしやすい「商圈内人口」だけを使うようにします。

2つ目は1つ目の後半に言ったことです。説明変数同士で高い相関を示しているものは、どちらかを使えばOKということです。

3つ目はむやみに説明変数を多くしないことです。多いほど説明力が高いと思ってやってみると、かえって説明する相手の頭が混乱してしまいます。

ビジネスにおいて、重回帰分析を使って誰か（ケースでは事業部長）に説明する時は、その相手によってもちがいますが、2~4個くらいの説明変数が適切かもしれません。細かいことを気にする相手の時は“多めに”、スパッと考えるのが得意な人の時は“少な目”にするのがセオリーでしょう。

この3つのコツからいえるのは、要するに「相手に説明しやすい変数にする」ことです。

高橋さんのケースでは「顧客」という側面から「コア商圈内人口」、事業部長の指摘した「店舗フォーマット」という側面から「売場面積」、「レストランゾーン面積」を説明変数として、重回帰分析をするのがノーマルでしょう。

もし事業部長から「他は考慮しなかったのか」と聞かれたら、図表3-13の相関係数を使って、「説明変数を選んだ理由」を説明しましょう。

### さあ重回帰分析だ

ここから先はもちろんエクセルで行います。

まずエクセルで次のような表を作ります。

店名	入力y範囲		入力x範囲	
	売上	コア商圏内人口	売場面積	レストランゾーン面積
No1	982	10284	1100	550
No2	660	8726	1890	513
No3	1466	9951	2000	546
No4	2067	15212	4770	728
No5	1125	9159	3433	611
No6	1288	11076	3670	724
No7	1972	11127	2280	513
No8	1998	16610	5522	753
No9	872	12085	1400	586
No10	680	6996	816	389
No11	1569	15334	1490	748
No12	543	5287	666	513
No13	1664	13952	2520	537
No14	645	11976	546	523
No15	789	7445	2920	657
No16	1314	8384	1520	634
No17	1716	11618	1970	657
No18	1620	17035	3880	583
No19	567	7297	2280	486

No20	762	9220	537	425
No21	1276	13381	1043	730
No22	931	13037	3690	651
No23	1840	17437	2660	811
No24	1166	6986	2800	475
No25	1048	11329	1440	436
No26	234	6970	523	421
No27	432	7869	1520	396
No28	1694	10662	4220	687
No29	354	8559	970	399
No30	298	8565	487	368

(図表3-16)

そのうえで相関分析の時と同じように、「分析ツール」（または「データ分析」）のメニューから「回帰分析」を選び、「ラベル」をクリックしておいて、「入力 y 範囲」には「表頭の売上とその下の30個の数字」を、「入力 x 範囲」には「表頭（コア商圏内人口、売場面積、レストランゾーン面積）と30×3列の数字」を選びます。

こうすると次のような表が出ます。

## 概要

回帰統計	
重相関 R	0.827031496
重決定 R2	0.683981096
補正 R2	0.647517376
標準誤差	325.2921241
観測数	30

## 分散分析表

	自由度	変動	分散	観測された分散比	有意 F
回帰	3	5954584.75	1984861.583	18.75785305	1.10265E-06
残差	26	2751189.116	105814.966		
合計	29	8705773.867			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	-504.1423104	295.2319958	-1.707614071	0.099623782	-1111.000363	102.7157424	-1111.000363	102.7157424
コア商圏内人口	0.066231746	0.025128133	2.635760676	0.013968124	0.014580128	0.117883364	0.014580128	0.117883364
売場面積	0.122610171	0.057116396	2.146672066	0.041321557	0.00520574	0.240014603	0.00520574	0.240014603
レストランゾーン面積	1.134876773	0.719379591	1.577577106	0.126753613	-0.343829139	2.613582684	-0.343829139	2.613582684

この部分が回帰式を表わしている

(図表3-17)

上の係数の列が58ページの式を表わしています。つまり次のような関係（回帰式）を表わしています。

「コア商圏内人口」の所にある係数=0.0662...

「切片」の所にある係数=-504.14...

「売上=0.066×コア商圏内人口+0.12×売り場面積+1.1×レストランゾーン面積-500」

「コア商圏内人口=0.066」「売場面積=0.12」という数字は、「コア商圏内人口の影響が小さい」ということではありません。数字の“単位”のためです。もし「コア商圏内人口」の単位を「人」ではなく、「千人」でやれば、係数は「66」となります。

ですから、係数は数字の大きさに限らず、数字の“桁”を2桁くらいに揃える(0.066、0.12、1.1、500)とよいと思います。

ここで先ほどの回帰式に、55ページ図表3-1の新規店舗のコア商圏内人口(9526)、売場面積(1485)、レストランゾーン面積(750)を入れて計算すると、次のようになります。

「新店の売上=0.066×9526+0.12×1485+1.1×750-500=1132」

つまり新店舗の売上予測値は、「1,132百万円＝11億3千2百万円」となります。もう一度図表3-1の元のデータを見てください。この新規出店の「？」の欄に「1132」と入れて、違和感を感じますか？なかなかいい感じではないですか？説明する相手が納得してくれるような数字ではないですか？

### シミュレーションだってできる

この回帰式があれば、売場面積、レストランゾーンの面積を変えることで、売上がどう変化するかも予測できます。

事業部長の発言にある「『こんなフォーマットなら、このような売上になる』というのを提示してくれないと身動きが取れないだろう」という“投げかけ”に対する“答え”です。

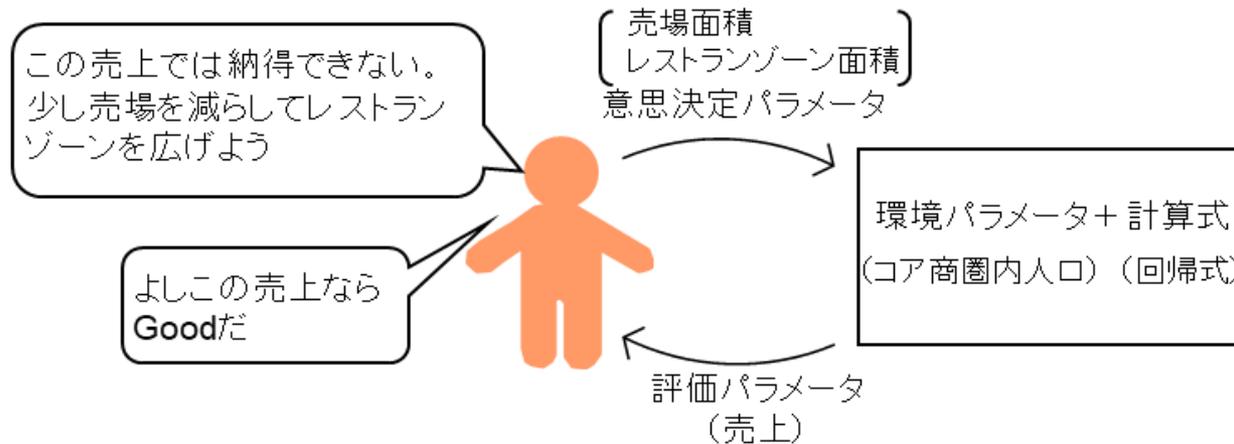
これは意思決定シミュレーションと表現されるコンピュータを使った“仕組”です。

意思決定シミュレーションとはコンピュータの中に、「環境パラメータ\*1」（動かすことのできない数字）と「計算式」をセットしておいて、「意思決定パラメータ」（自らで決められる数字）を入れて、「評価パラメータ」をアウトプットするものです。そしてこの評価パラメータが納得できないなら、意思決定パラメータを変え、納得がいったらその意思決定をするというものです。

今回のケースであれば、環境パラメータにあたるのが「コア商圏内人口」（動かすことのできない数字。出店場所は決まっている）であり、計算式にあたるのが72ページの回帰式「 $(\text{売上} = 0.066 \times \text{コア商圏内人口} + 0.12 \times \text{売り場面積} + 1.1 \times \text{レストランゾーン面積} - 500)$ 」です。

ここに「売場面積」、「レストランゾーン面積」という意思決定パラメータ（「この面積にしたらどうなるか」）を入れて、「売上」という評価パラメータ（「きっところなる」）をアウトプットするものです。

この「売上」に納得できなければ（「こんな売上ではだめ」）、可能な範囲で売場面積、レストランゾーン面積を変えていきます（「こうしたらどうなる」）。そして予測した売上に納得したら（「これならいいだろう」）、これでGo（出店）ということです。



(図表3-18)

\* 1.パラメータ この用語は色々なシーンで色々な意味に使われている。この場合は「設定する値」という意味で使われている。

## 説明変数を追加できるか

さあ事業部長への説明です。

高橋「新店の売上を重回帰分析を使って予測しました。

売上を予測するための説明変数の候補として、お手許の8つの項目を考えてみました。

まずは売上と8項目について、既存店の数字を使って相関係数を計算しました。その結果は図表のとおりです。この相関係数をもとに、顧客という切り口から

コア商圏内人口、店舗フォーマットという切り口から売場面積、レストランゾーン面積という3つを説明変数として選び、重回帰分析によって次のような予測式を得ました。

その結果、新店舗の年間売上は11億3千2百万円と予測されます」

事業部長「30代の女性人口との相関はそんなには高くないんだな。へえ一意外だなあ。でもなかなか興味深い結果だな。

ところで食品アイテム数の相関係数が結構大きいけど、これを説明変数に追加することはできるのか」

高橋「はい。できます。やってみます」

高橋さんは食品アイテム数のデータを足して、次のような表を作り、重回帰分析を行いました。

店名	売上	コア商圏内人口	売場面積	レストラン面積	食品アイテム数
No1	982	10284	1100	550	4490
No2	660	8726	1890	513	3710
No3	1466	9951	2000	546	6048
No4	2067	15212	4770	728	5814
No5	1125	9159	3433	611	3407
No6	1288	11076	3670	724	2305
No7	1972	11127	2280	513	5762
No8	1998	16610	5522	753	9097
No9	872	12085	1400	586	2881
No10	680	6996	816	389	4157
No11	1569	15334	1490	748	9504
No12	543	5287	666	513	2964
No13	1664	13952	2520	537	9081
No14	645	11976	546	523	1887
No15	789	7445	2920	657	7124
No16	1314	8384	1520	634	7315
No17	1716	11618	1970	657	7660
No18	1620	17035	3880	583	6279
No19	567	7297	2280	486	2236
No20	762	9220	537	425	2795
No21	1276	13381	1043	730	4031
No22	931	13037	3690	651	5296
No23	1840	17437	2660	811	11187
No24	1166	6986	2800	475	5575
No25	1048	11329	1440	436	4649
No26	234	6970	523	421	4027
No27	432	7869	1520	396	5124
No28	1694	10662	4220	687	5202
No29	354	8559	970	399	4348
No30	298	8565	487	368	3495

(図表3-19)

結果は次のようになりました。

概要								
回帰統計								
重相関 R	0.856386							
重決定 R2	0.733397							
補正 R2	0.69074							
標準誤差	304.6959							
観測数	30							
分散分析表								
	自由度	変動	分散	観測された分散比	有意 F			
回帰	4	6384785	1596196	17.1930592	6.77E-07			
残差	25	2320989	92839.57					
合計	29	8705774						
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	-476.472	276.8376	-1.72112	0.097583159	-1046.63	93.68607	-1046.63	93.68607
コア商圏内人口	0.047852	0.025038	1.911165	0.067513844	-0.00371	0.099418	-0.00371	0.099418
売場面積	0.118845	0.053529	2.220221	0.035699478	0.008601	0.229089	0.008601	0.229089
レストランゾーン面積	0.847327	0.686944	1.233473	0.22887115	-0.56746	2.262115	-0.56746	2.262115
食品アイテム数	0.065182	0.03028	2.152626	0.04119333	0.002819	0.127545	0.002819	0.127545

(図表3-20)

したがって

「売上=0.048×コア商圏内人口+0.12×売場面積+0.85×レストランゾーン面積+0.065×食品アイテム数-480」

となります。

ここに新店舗のデータを入れると次のようになります。

$$0.048 \times 9526 + 0.12 \times 1485 + 0.85 \times 750 + 0.065 \times 5000 - 480 = 1118 \text{ (百万円)}$$

これで11億1千8百万円が予測値となります。

予測で大切なのは結果ではありません。その根拠です。この場合は「『コア商圏内人口、売り場面積、レストランゾーン面積、食品アイテム数』を説明変数として、重回帰分析で売上を予測した」ということです。33ページで述べた「予測のやり方」であり、これを事業部長とは議論します。つまり説明変数を変えたり、74ページのように意思決定パラメータを変えたりして、互いが合意できるように調整します。

## 数字でなくても大丈夫

では事業部長がなかなか細かい人で、「店舗の近くに競合店があるかどうかで、売上は変わるだろう。競合店も百貨店とファミリーレストランではちがうだろう」と言ったらどうでしょうか。

もちろん回帰分析では、数字以外を使うことはできません。ですから「百貨店があるかないか」を数字にします。数字になれば説明変数として追加できます。

「近くに百貨店がある」なら「1」、「ない」なら「0」です。これでもう数字です。数字になれば説明変数として追加できます。

ファミリーレストランも同様です。

百貨店変数（1、0）、ファミリーレストラン変数（1、0）をエクセルの表に追加して、重回帰分析をやればOKです。

このように定性的な項目（数字になっていない項目）を数字に表現していくことを、統計の世界では数量化理論といっています。

j社は情報システムをトータルで提供し、それをサポートしていくITベンダー\*1である。

j社では現在、生産財\*2メーカーのk社に対して、新しいスタイルの情報システムを提案している。

j社からの提案内容はk社にほぼ合意された。そこでj社の営業担当の横山は、k社情報システム部長とシステム投資の金額の詰めに入っていた。

部長の意見は「情報システムへの投資は、カネを使えばキリがない。社長からは『世間並みの平均的な情報システム投資額をまず考えろ』と言われている。だから世の中の一般的な投資金額を考え、これをベースに予算化を検討していきたい。

御社の方で同業他社を調べて、平均的な情報システム投資額を考えてみてくれないか」であった。

帰社後、横山は思った。「しかし平均といっても、k社と同業種の会社は、売上や従業員数が随分ちがう。規模がちがう会社の情報システム投資額の平均を取っても意味がないと思う。

でも、まあとりあえず同業種の情報システム投資額とその企業サイズに関するデータを集めてみよう」

その結果は以下の通りであった。

会社No.	情報システム投資額(百万円)	年間売上高(億円)	従業員数	事業所の数	営業資産 <sup>*3</sup> (億円)	情報システム部員人数
1	120	180	980	10	50	20
2	180	340	1450	10	45	15
3	720	1420	3280	15	70	25
4	85	120	285	7	28	20
5	145	720	1620	18	45	18
6	480	1820	4100	28	60	28
7	200	420	780	18	40	20
8	1100	2880	4820	40	120	25
9	280	520	1220	32	48	20
10	260	690	1680	25	72	50
k社	?	920	1420	22	50	40

(図表3-21)

横山「何とかこのうちの2つくらいの項目を考慮して、うまく情報システム投資額の平均がとればいいのになあ」

さあ横山さんの立場で、情報システムの平均投資額を計算してみましよう。ヒントはLesson4のExerciseですので、重回帰分析を使うことです。

- \* 1. ITベンダー IT商品を販売している会社のこと
- \* 2. 生産財 工場向け製品のこと
- \* 3. 営業資産 その会社が持っている「事業に使用される設備などの財産」

回帰分析は、「状況の異なる数字の“平均”を取っている」と考えることもできます。

「数字活用シーンその3」の新店舗の売上の例で考えてみましょう。既存店舗の売上の平均値を計算して、31店目の売上予測をしたいのですが、状況のちがう店舗の売上をそのまま平均しても意味がありません。そこで30店舗のコア商圈内人口、売場面積、レストランゾーン面積という特徴を加味して、30店舗の平均を取り、この平均を31店舗目に使うようにする、と回帰分析を考えることができます。

この時、平均は「1つの数字」ではなく、回帰式で表わされます。

これが回帰分析のもう1つの見方です。

このExerciseのように、回帰分析は状況が異なっている数字を、その状況を考慮して平均を取りたい時にもよく利用されています。

今回の情報システム投資額を被説明変数として、2つくらいの説明変数を使って重回帰分析をすることで、「平均を取る」と考えてみましょう。

まずは何を説明変数とするかです。

もちろん相関分析です。次のような表が出せましたか？出し方を忘れた人は67

ページで復習してください。

	情報システム投資額(百万円)	年間売上高(億円)	従業員数(人)	事業所の数	営業資産(億円)	情報システム部員人数
情報システム投資額(百万円)	1					
年間売上高(億円)	0.944902093	1				
従業員数	0.902940084	0.968498479	1			
事業所の数	0.67273185	0.744057462	0.664007435	1		
営業資産(億円)	0.909715635	0.88506919	0.834473268	0.731366872	1	
情報システム部員人数	0.17083189	0.208971677	0.225780432	0.347675204	0.397959947	1

(図表3-22)

これを見ると年間売上高、従業員数、営業資産が情報システム投資額との間で0.9以上という「強い相関」を示しています。

また年間売上高と従業員数の間も0.97という「極めて強い相関」を示しています。つまり説明変数として考えると年間売上高、従業員数はどちらかを使えばOKです。

そこで年間売上高と営業資産の2つを説明変数として重回帰分析を行うことにします。

まずはこの3項目をエクセルに次のようにセットします。

情報システム 投資額	年間売上高	営業資産
120	180	50
180	340	45
720	1420	70
85	120	28
145	720	45
480	1820	60
200	420	40
1100	2880	120
280	520	48
260	690	72

(図表3-23)

72ページでやったように「回帰分析」を選び、「ラベル」にチェックを入れ、入力範囲  $y$  には見出しを含めて「情報システム投資額」を、入力範囲  $x$  は「年間売上高」、「営業資産」の列を選びます。

すると次のような結果となります。

概要								
回帰統計								
重相関 R	0.957975							
重決定 R2	0.917715							
補正 R2	0.894205							
標準誤差	105.4166							
観測数	10							
分散分析表								
	自由度	変動	分散	観測された分散比	有意 F			
回帰	2	867571.4	433785.7	39.0352551	0.00016			
残差	7	77788.65	11112.66					
合計	9	945360						
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	-107.691	112.3071	-0.9589	0.369549934	-373.255	157.8729	-373.255	157.8729
年間売上高	0.238267	0.086046	2.769051	0.027731995	0.034799	0.441735	0.034799	0.441735
営業資産	4.284256	2.945108	1.454702	0.189078275	-2.67982	11.24833	-2.67982	11.24833

(図表3-25)

これから次のような回帰式となります。

$$\text{「情報システム投資額 百万円）} = \text{年間売上高(億円)} \times 0.24 + \text{営業資産（億円)} \times 4.3 - 110\text{」}$$

したがってk社については次のようになります。

$$\begin{aligned} \text{k社に関する情報システム投資額の平均} &= 920 \times 0.24 + 50 \times 4.3 - 110 \\ &= 326 \text{ (百万円)} \end{aligned}$$

つまり3億2千6百万円が、k社としての「平均的な情報システム投資額」となります。

図表3-22の表の「？」に「326」（百万円）を入れるとどんな感じですか？中規模タイプのk社の数字としてはGoodな感じがしませんか？

私はk社の社長の「世間並み」「平均」というキーワードにぴったり合った“感じ”だと思います。

1社は菓子メーカーであり、昨年子供向けのコーヒーキャンディを新発売した。「コーヒーの成分が頭脳の活性化に効果がある」とのテレビ報道もあり、販売は順調に推移した。

コーヒーキャンディは発売後の1年間で、大都市圏内の一部ではマーケットは飽和状態\*1となっていた。一方、その他の地域では、まだそれほど浸透しておらず、マーケットはゆっくりとした伸びを示していた。

そのような中、テレビ報道を見たライバルメーカーが同タイプのコーヒーキャンディを開発して、ここに来て次々と追随してきた。

1社では、本年度よりエリアマーケティングというマーケティングシステムを導入することとなった。これは全国を市町村単位くらいの「エリア」という地域に割り、このエリア単位にマーケティングを考えるというものである。

エリアマーケティングにおけるポイントは、ポテンシャルパイというものである。それはそのエリアが本来持っている「潜在的な需要」のことであり、そのエリアの当該商品を「買う力」と考えてもよい。

このポテンシャルパイをベースとして、各エリア単位に販売目標、プロモーション費用の配分といったすべてのマーケティング活動を行うのがエリアマーケティングである。

I社ではまず販売好調のコーヒーキャンディで、テスト的にエリアマーケティングを実施してみることにした。

山中は新設されたエリアマーケティング推進部に配属された。

山中はエリアマーケティングの本を読み、大体の考え方は理解した。

「コーヒーキャンディのポテンシャルパイはそのエリアで売れておかしくない総量のことだな。他社の販売量を含めるだけでなく、今は売れていないが、何かマーケティング行為を行えば売れる量も考えるんだな。コーヒーキャンディがあることを知らない子供だっているだろう。これが頭脳の活性化に役立つことや、苦味は全くないことを知れば買うかもしれない・・・。

しかし『買ってでもいいはずの客が買う量』なんてどうやって計算するんだろう。もう1度エリアマーケティングの本を読んでみよう。

『ポテンシャルパイはすでにその商品が欲しい人に行き渡っているエリア、つまりマーケットが飽和されたエリアの販売量をベースにする。他のエリアはこれを使って推測すればよい。』

そうか、飽和しているマーケットの販売量がポテンシャルパイか。

よし、コーヒーキャンディの売れ行きが頭を打っているエリアを、営業企画部に教えてもらおう。このエリアはまだうちの独占状態のはずだ。だから月間の販売個数をこのエリアのポテンシャルパイと考えていいはずだ。ただ推測するには

販売量だけでなく、販売に影響を与えるデータが必要だと書いてあるから、それも頼もう。」

営業企画部「販売に影響を与えるデータね。それならエリアマーケティング導入時に、リサーチ会社に頼んでエリア属性を調べてもらっているよ。えーと、あるのは小中学生人口、小中学校の数、学習塾の数、小中学生の携帯電話所有率の4つだ」

「小中学生の数、学校の数わかるけど“塾”と“携帯”…。塾はそうか、頭を使うからか。携帯は小中学生のお金持度みたいなものだな。

よしこれを一覧表にしよう」

山中は30エリアのデータを、次のようにエクセルに入れた。

No	ポテンシャル パイ	小中学生数	小中学校数	学習塾数	携帯 保有率(%)
1	4100	7000	20	30	35
2	5700	16000	53	25	25
3	3550	5700	19	19	21
4	6297	20613	76	8	32
5	4159	11556	38	25	9
6	8643	19978	62	7	65
7	4652	9528	30	16	23
8	6838	19055	62	23	53
9	5076	11841	33	67	34
10	5317	16553	50	23	22
11	6357	16237	52	32	27
12	6210	20381	61	43	21
13	5997	14920	43	5	37
14	4251	7214	21	34	45
15	5218	8503	24	22	48
16	6066	10117	32	17	51
17	4703	6368	17	13	53
18	5752	17450	52	34	21
19	3449	3035	9	13	24
20	5157	5138	17	15	63
21	4582	3128	10	21	41
22	9066	20172	61	55	88
23	4823	4071	11	21	57
24	6276	14732	43	12	55
25	5731	10773	33	23	63
26	9964	17710	53	72	92
27	6481	13951	40	13	73
28	7935	20277	58	18	60
29	4537	7286	20	24	31
30	6965	18656	55	16	44

「でもどうやって他のエリアのポテンシャルパイを推測しようか」

さあ山中さんの立場でポテンシャルパイの算出方法を考えてみましょう。最終アウトプットは各エリアのポテンシャルパイを計算する式（ポテンシャルパイ＝・・・）です。

\* 1.飽和状態 買いたい人に商品が行き渡った状態。したがって商品売上は横バイとなる。

(図表3-26)

もう慣れてきたと思います。ポテンシャルパイを被説明変数として、その説明変数を相関分析で選んで、回帰分析します。

まずは相関係数を計算しましょう。

	ポテンシャルパイ	小中学生数	小中学校数	学習塾数	携帯保有率
ポテンシャルパイ	1				
小中学生数	0.77997596	1			
小中学校数	0.741044673	0.986124757	1		
学習塾数	0.312514843	0.192174972	0.146293231	1	
携帯保有率	0.686262248	0.166495253	0.11775341	0.197648	1

(図表3-27)

ポテンシャルパイとの相関は小中学生数0.78、小中学校数0.74、学習塾数0.31、携帯保有率0.69です。しかし小中学生数と小中学校数には0.99という「極端に高い相関」を示しています。

そこで被説明変数には、説明しやすさから小中学生数と携帯保有率とします。これをエクセルの表にします。

No	ポテンシャルバイ	小中学生数	携帯保有率(%)
1	4100	7000	35
2	5700	16000	25
3	3550	5700	21
4	6297	20613	32
5	4159	11556	9
6	8643	19978	65
7	4652	9528	23
8	6838	19055	53
9	5076	11841	34
10	5317	16553	22
11	6357	16237	27
12	6210	20381	21
13	5997	14920	37
14	4251	7214	45
15	5218	8503	48
16	6066	10117	51
17	4703	6368	53
18	5752	17450	21
19	3449	3035	24
20	5157	5138	63
21	4582	3128	41
22	9066	20172	88
23	4823	4071	57
24	6276	14732	55
25	5731	10773	63
26	9964	17710	92
27	6481	13951	73
28	7935	20277	60
29	4537	7286	31
30	6965	18656	44

(図表3-28)

さあ重回帰分析です。

概要									
回帰統計									
重相関 R	0.962689								
重決定 R2	0.92677								
補正 R2	0.921345								
標準誤差	439.6253								
観測数	30								
分散分析表									
	自由度	変動	分散	観測された分散比	有意 F				
回帰	2	66040644	33020322	170.8503535	4.71 E-16				
残差	27	5218302	193270.4						
合計	29	71258946							
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%	
切片	1587.679	241.8419	6.564946	4.84452E-07	1091.46	2083.897	1091.46	2083.897	
小中学生数	0.182988	0.014115	12.96375	4.15323E-13	0.154025	0.21195	0.154025	0.21195	
携帯保有率	43.45982	4.011062	10.83499	2.47275E-11	35.2298	51.68984	35.2298	51.68984	

(図表3-29)

したがって次のような式が得られます。

「エリアのポテンシャルパイ = 小中学生数 × 0.18 + 携帯保有率 × 43 + 1600」

これですべてのエリアのポテンシャルパイを推定します。

# 未来を確率で考える

## 正規分布と積分をマスターしよう

そろそろ統計テクニックもプロの領域に達しようとしています。

Lesson4では確率、確率分布、正規分布、それに積分という本格的な数学の分野も登場します。

でも心配することはありません。何度も言うように、数学というのは、世界でもっとも頭が賢くて、まわりに説明能力の高い人が学者となり、数字の使い方をきちんと“見える化”したものです。だからわからないはずはありません。

重回帰分析、多変量解析までマスターしたあなたなら、ここでもう1歩踏み込んでみましょう。これを達成すればそろそろゴールが見えてきます。

## 数字活用ケースその4

### —— 需要にフィットした生産量 ——

村田さんは「おかあさん弁当」の欠品や売れ残りをなるべく減らすように、最適な生産量というものを考えています。そこで「おかあさん弁当」の2ヶ月分の売れ行きデータを集めてみたのですが、どうしていいかわかりません。1日あたりの販売量の平均値を生産量とすると欠品が多すぎます。でも平均以外に数字の加工のやり方が思いつきません。あなたなら生産量をどう決めますか？

## 「これなら妥当」と皆が認める生産量

G社は弁当、惣菜などを製造しているメーカーであり、直営店舗も展開している。

現在G社では、主力商品である昼食向け「おかあさん弁当」の生産量に悩んでいた。おかあさん弁当の売れ行きは順調だが、店舗での欠品（品切れ）を恐れるあまり、売れ残りによる処分ロスが問題となっていた。

G社生産調整室の村田は、室長からの指示で「おかあさん弁当」の生産量の調整を行うことになった。

室長「村田君、今までのうちの生産量の決め方はあまりにもいいかげんすぎる。生産調整室がここ数日の売れ行きを見て、適当に生産量を決めている。

しかも生産調整室は、欠品を恐れる現場の販売サイドの声に押されて、結局作りすぎてしまい、大量の廃棄処分をしている。

欠品や売れ残りを一切出さないようにするには、売れる量をピタッと当てなくてはならない。だからそんなことが無理なことはわかっている。

これまでの販売量、つまり需要をきちんと分析して、『最適生産量』というものを考えるべきだろう。それが君の仕事だ。

販売サイドも文句を言うだろうから、誰が考えても『これなら妥当だ』と思える量を見つけてくれ。

とりあえず主力商品の『おかあさん弁当』を対象として、1日あたりの標準的な生産量というのを考えてみてくれ」

## 平均では欠品が多すぎる

村田は席に戻って考えた。

「生産量だから1日どれくらい売れているか、つまり1日あたりのおかあさん弁当の需要を考えるべきだな。おかあさん弁当は平日と休日では売れ方が大きくちがうので、まずは平日の需要から考えてみよう。

おかあさん弁当は、ここ2ヶ月は比較的需要が安定している。ここ2か月分のデータを使おう。ただ欠品している日は、その分の需要がわからないことになってしまう。売れ残りがゼロの日を調べてみよう。」

「何だ、ここ2ヶ月は毎日売れ残りを処分しているじゃないか。これじゃ室長が怒るのも無理はない。

ということは、需要は毎日の販売数を見ればいいんだな。この販売数分だけ、顧客がおかあさん弁当を求めていることになる。

まずは表計算ソフトにデータを入れて平均を取ってみよう。」

9/5	月	325
9/6	火	388
9/7	水	335
9/8	木	304
9/9	金	315
9/12	月	345
9/13	火	415
9/14	水	386
9/15	木	354
9/16	金	378
9/19	月	307
9/20	火	388
9/21	水	324
9/22	木	336
9/26	月	358
9/27	火	334
9/28	水	326
9/29	木	318
9/30	金	354
10/3	月	288
10/4	火	276
10/5	水	334
10/6	木	355
10/7	金	334

388個が需要。  
だから338個の生産量  
では欠品！

最大値！  
でも416個になる日  
だってあるかもしれ  
ない

10/11	火	295
10/12	水	326
10/13	木	378
10/14	金	364
10/17	月	339
10/18	火	308
10/19	水	368
10/20	木	326
10/21	金	288
10/24	月	345
10/25	火	309
10/26	水	310
10/27	木	372
10/28	金	328
<b>平均値</b>		<b>338</b>

(図表4-1)

「平均は338個か。しかし338個を生産量とすると欠品の日が多くなるなあ。頭から欠品の日、つまり338個より需要の多い日を見ていくと、9月6日、12日、13日、14日、15日、16日、何だこの週は1週間ずっと欠品じゃないか。

よく考えたら平均なんだからあたりまえか。毎日の販売数が4個、6個、7個、3個なら平均は5個だ。5個を生産量とすると、需要が6個、7個の時には欠品か。2日に1回は欠品だ。

それじゃあ生産量を需要の最大値にするのか。この2ヶ月間の最大値は415個か。しかし415個の日があるということは、これから先、416個の日だってありそう。室長の言っているとおり、絶対に欠品しない生産量なんてないし、最大値415個が最適生産量とはとても思えないよなあ。

何かよい方法がないかなあ。」

## 未来を知るコツでエクセレントカンパニーへ

生産量や仕入量と、実際の販売量の差を「在庫」といいます。この在庫を考える仕事を在庫管理といいます。

在庫管理は、その商品が「未来に売れる量」＝「需要」を考えて、その分を作ったり、仕入れたりするという仕事です。だから在庫管理は、需要予測そのものです。まさに村田さんが悩んでいるテーマです。

このケースのような需要予測という仕事は、ビジネスにおいて「統計がもっともうまく使われている分野」です。統計をうまく使うことでエクセレントカンパニーとなったのが、トヨタ、花王、セブン-イレブンジャパンといった会社です。

在庫管理、需要予測は、統計、というよりも数学に関する基本的な知識がないとよく理解できません。だから「数学が苦手な人」（声の大きい人が多い）がリーダーシップを取っている多くの日本の会社では、この在庫管理をうまくやっていくことができません。ということは、逆にこれをうまくやれば、上のエクセレントカンパニーのようにライバル会社を大きく引き離すことができます。

数学が苦手な人も、このLesson4で必ず理解できるようになりますから、がんばって、需要予測という「未来を知るコツ」を手に入れましょう。そして自分の会社をリードして、エクセレントカンパニーへと導きましょう。少し言いすぎのように聞こえるかも知れませんが、そんなことはありません。あなたが数字を使えるようになるだけでなく、その数字の使い方をまわりへ説明していけば、きっとあなたの会社は「数字に強い会社」に変わります。

## 確率といえばサイコロ

需要予測を理解するための数学の基礎知識として、まずは確率からスタートです。

確率は極めて直感的です。

確率はよくサイコロの例で説明されます。サイコロの目は1、2、3、4、5、6という6つの数字が出ます。だからサイコロの目は56ページで述べた“変数”です。

小学校か中学校でやったように、サイコロを振って「目が1」と出る確率は $1/6$ です。

サイコロの目のように、取りうる数字が限られていて、これらの値（1～6）を出す確率（この場合はすべて $1/6$ ）が決まっている時は、この変数を離散確率変数といいます。またこの「1と $1/6$ 」「2と $1/6$ 」・・・という「変数（サイコロの目）とその確率の組み合わせ」を離散確率分布といいます。

離散確率分布と用語は大げさですが、大して難しいことはありません。離散とは「つながっていない」という意味で、後で出てくる「連続」の反対です。分布とは、2種類の数（「サイコロの目」と「確率」）が“ペア”で変わっていくという状態を表わしています。分布の例は、これまで何度も出てきたことはわかりますよね。

一方、変数が連続数で、その変数の値によって確率が決まっているものを連続確率変数、この連続数と確率が変わっていく状態を連続確率分布といいます。

連続数とは長さ、重さのように数字がつながっているものです。さいころの目には2.5というのはありませんが、長さには2.5cmというものがあります。

## 正規分布＝すっきり分布

連続確率分布の代表は正規分布です。まあ、これも大して難しいことはありません。正規分布を「正規」と「分布」に分けてみましょう。

先ほど説明したように「分布」というくらいなので、2つの数字がペアで変わっていく状態を表わしたものです。正規分布は確率分布ですので、このうちの1つの数字は確率です。

「正規」とはnormalの和訳です。ここでいう「ノーマル」とは「標準」といった意味です。だから正規分布とは「標準的な連続確率分布」という意味です。

ついでに言うと、数学では「正規化」という言葉を使います。これは“化”（「変える」ということ）というくらいなので、「正規ではないもの」を「正規にする」（＝「標準にする」）という意味です。数学では正規化を、主に「冗長性（余分なもの）を取って標準スタイルにする」といった意味で使います。まあ要するに“ぜい肉”を取って、すっきりさせることです。

ビジネスの世界では“正規”のイメージを、「すっきり」「単純化」と考えて問題ありません。

ですから正規分布は“すっきり分布”、“単純分布”といった意味です。

## 度数はヒストグラムで

ケースに入る前に、いつものように簡単な例で考えてみましょう（言ってみれば先ほどの「正規化」＝単純化です）。

「日本人の20代の男性の体重がどうなっているのか」を知りたいとします。もうわかりますね、これが母集団です。

しかし日本人20代男性全員の体重を調べるわけにはいかないので、100人をサンプリングして、体重を測りました。この100人が標本です。

標本の100人の体重で、日本人の20代男性の体重を推定しようというものです。

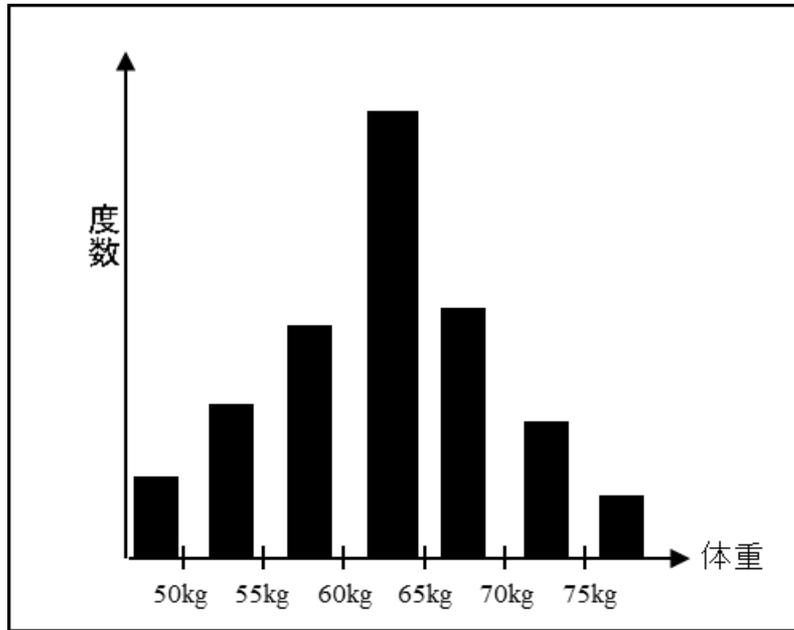
体重は連続数です。これを5kg単位にランク分けして、カウントしてみたら、次のようになりました。カウントした結果を「度数」といいます。

何人いたか = 何度カウントしたか

ランク	度数
50kg以下	6
50~55kg	11
55~60kg	17
60~65kg	32
65~70kg	19
70~75kg	10
75kg以上	5
合計	100

(図表4-2)

これを下のようなヒストグラムという「度数」を表わす帯グラフにします。



(図表4-3)

このヒストグラムも、もちろんエクセルを使って書くことができます。書き方は2つあります。

1つは図表4-2の度数表を用いるものです。この時はヒストグラムを「たて形の棒グラフ」で考えて書きます。

図表4-2の表を範囲として指定して、35ページの散布図同様に、グラフのメニューから「縦棒」を選べばOKです。

もう1つは元データを使う方法です。

これは「標本100人の体重データ」をエクセルの1列に入れ、隣の1列にランク分けのためのデータを作ります（上の例では50から5刻みで50、55、60、65、70、75、80と入れます）。

そのうえで「分析ツール」（または「データ分析」）から「ヒストグラム」をクリックして選びます。

次に「ヒストグラム」の「メニュー」で、「入力範囲」に「100人の体重データ」の「列」を、「データ区分」に「ランク」（5刻みの列）を入れた「列」を選択し、「グラフ作成」をチェックして、OKを押せばできます（度数表もあわせて出てきます）。

後者の方が元データの度数をカウントしなくてすむので、“楽”ですね。

### 棒の高さと確率の関係

図表4-2の度数表と、図表4-3のヒストグラムを見てください。

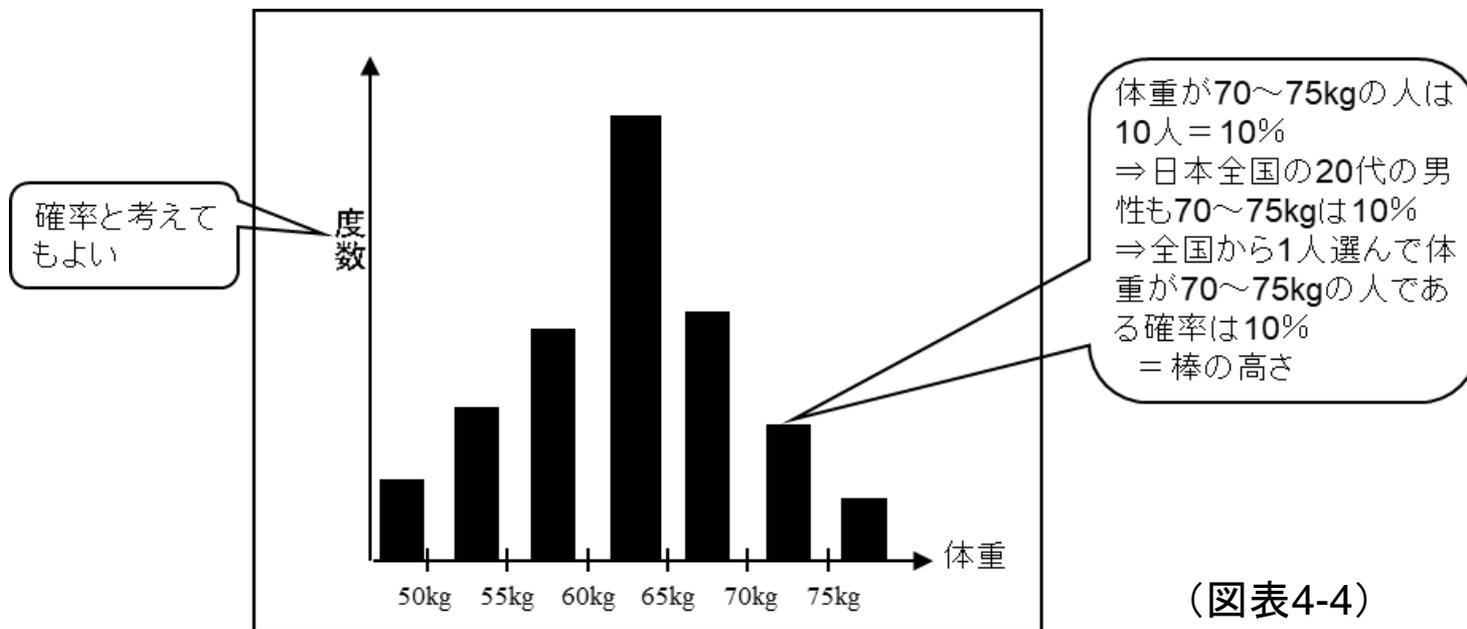
ここでは70kg～75kgの人が、100人のうち10人います。全体の10%です。

「全国の20代男性で、体重が70～75kgの間にある人は何%いる」と考えるべきでしょうか。この標本（100人）で母集団（全国）を推定するのですから、10%と考えるべきでしょう。と言うよりも、それしか考えようがありません。

それでは全国の20代男性の中から1人選んで、その人の体重が70～75kgに入っている“確率”はいくつでしょうか。

これも「10%」と考えるしかありません。

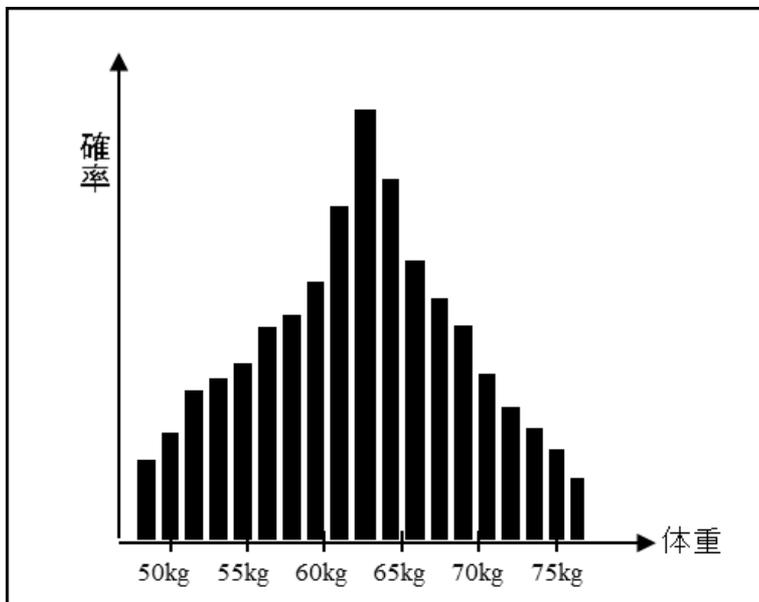
ということは、このヒストグラムの「高さ」（＝度数）は確率を表わしているともいえます。



(図表4-4)

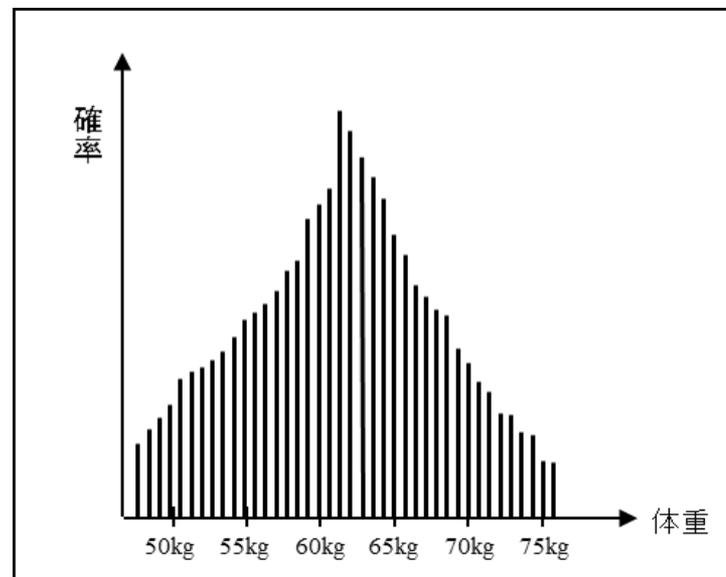
### 棒をつなげば線が見える

ここで標本の数をもっと多くして、ランク分けを細かくしていくと、図表4-5のようになります。



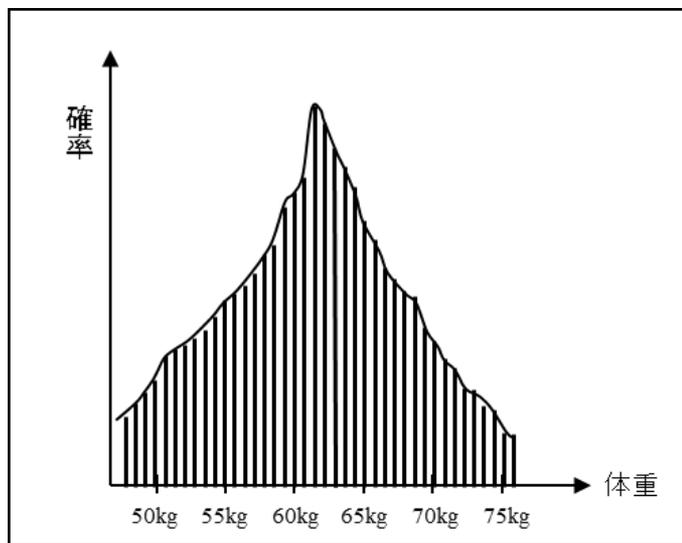
(図表4-5)

さらに標本を増やして、ランクを細かくしていきくと、図表4-6のようにそれぞれの棒がだんだんつながってきます。



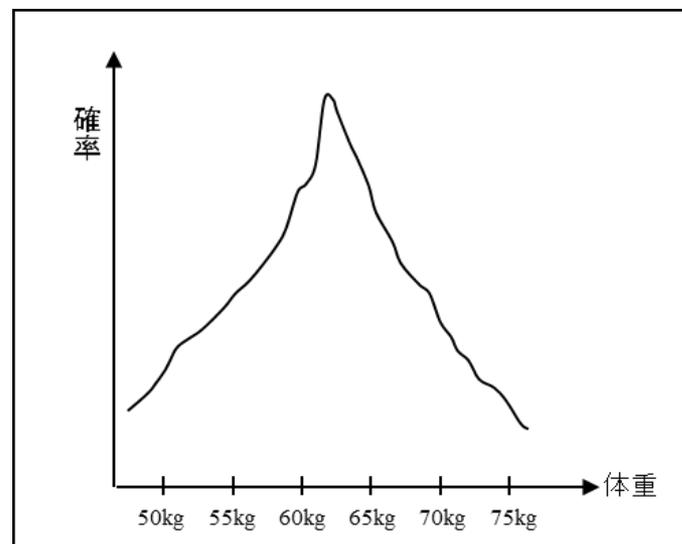
(図表4-6)

この棒の頭をつないでみましょう。



(図表4-7)

こうなると棒の意味がないので、棒を取ってしましましょう。



(図表4-8)

これが全国20代男性の体重を表わす連続確率分布、つまり「体重と確率という2つの数字が変わっていく姿」です（体重は連続数です）。

そしてこの曲線のことを確率密度関数といいます。

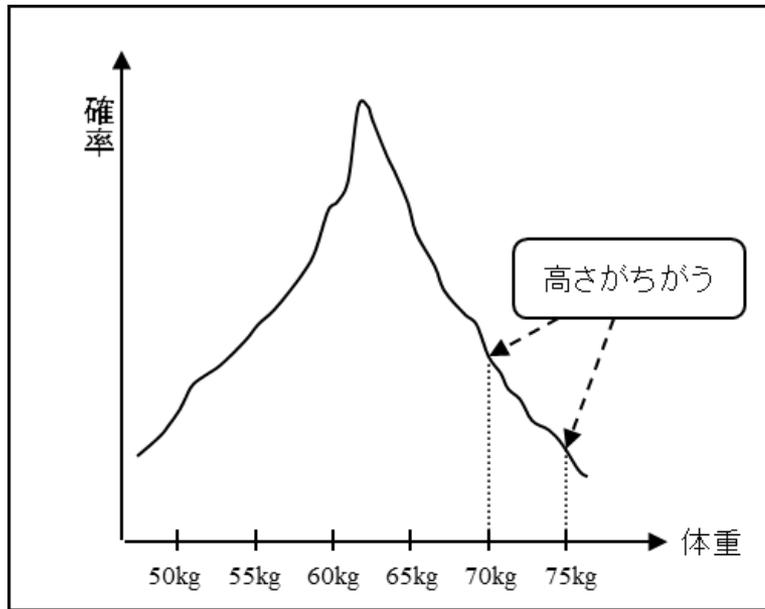
確率密度関数も大げさな表現ですが、簡単に説明しましょう。

「関数」とは2つの数字の「関係」を表わしたものです（もうエクセルのメニューで、この用語を使っていますが・・・）。この「関係」は「式」（回帰分析では回帰式）で表わしたり、上のように線（回帰分析では回帰直線）で表わしたりします。「関数」と「分布」は同じような意味ですが、分布は2つのペアの数字が変わっていく「状態」、関数はそれを“式”や“線”という「関係」で表わしたものと理解しましょう。

確率密度とは、「棒」という「確率をあらわすもの」が「ぐっつつまっが入っている」（図表4-7のような感じ）という意味です。

### 積分＝面積＝確率

この曲線（確率密度関数）を使って、「全国20代の男性の中から1人選んで体重が70～75kgに入っている確率」を考えましょう。



(図表4-9)

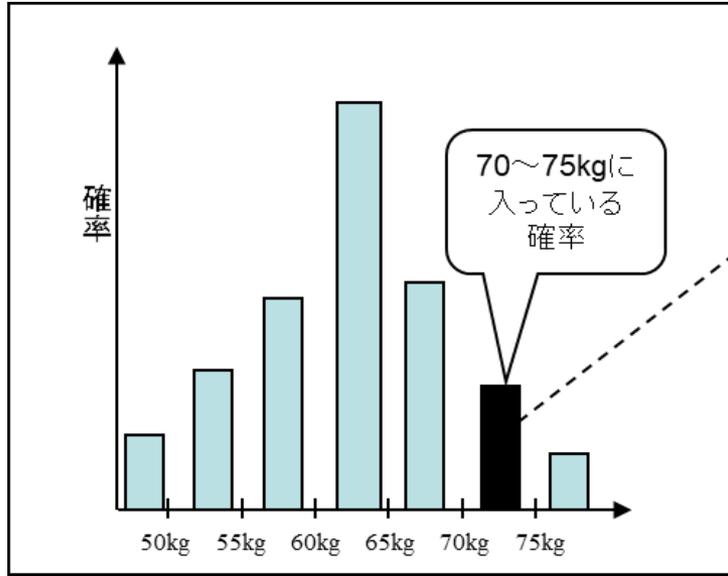
ここでたて軸は確率でした。つまり「曲線の高さ」が確率です。

70～75kgに入っている確率ですので、この部分の「高さ」を測りたいところです。しかし先ほどの棒とちがって、70kgと75kgの所では高さがちがいます。

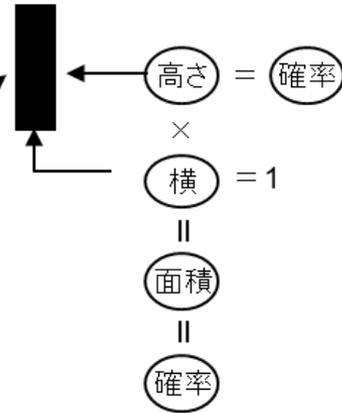
もう一度図表4-3のグラフを見てみましょう。ここでは棒の高さが確率でした。

しかしこの確率（棒の高さ）を「横の長さが“1”の長方形（棒）の面積」と考えることもできます。

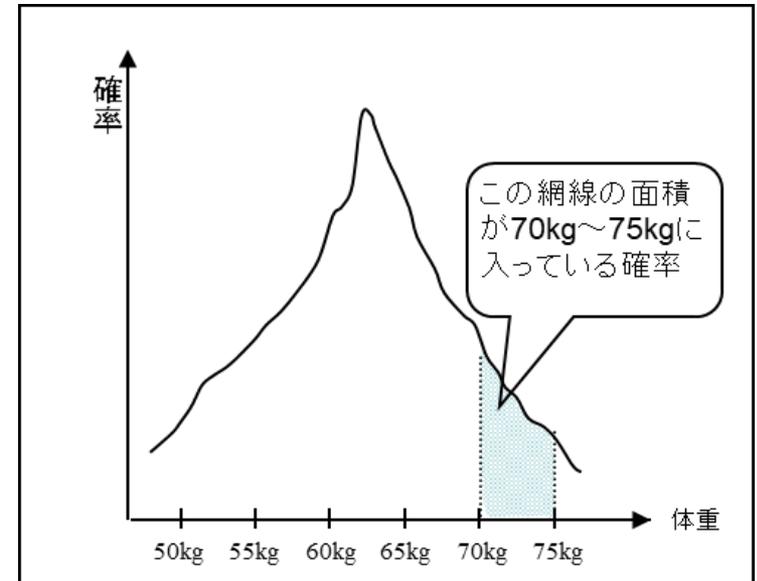
長方形の面積は小学校で勉強したように「たて×横」です。「横=1」なら「棒の面積=棒の高さ（たて）=確率」です。



(図表4-10)



つまり確率＝面積です。  
 そう考えれば1人の20代男性を選んで、その人の体重が70～75kgである確率は、下の図の網線の部分の面積となります。

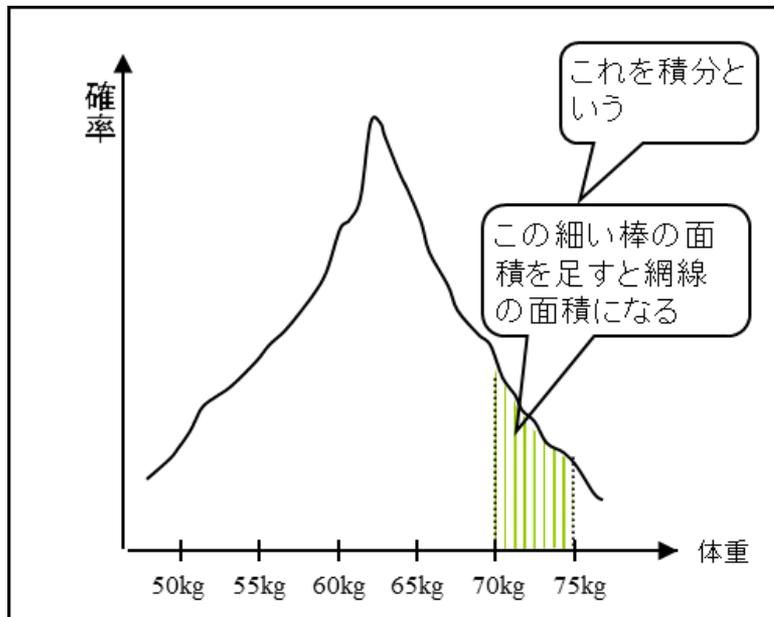


(図表4-11)

見方を変えれば、70～75kgの網線部の面積（確率）は、図表4-7のグラフにある「細い棒の面積」を「細かく足していったもの」といえます。これが積分です。つまり「分けているもの」（分）を「つなぐ」（積）という意味です。

積分\*1とは曲線を細かく区切られた「たて棒」を足していくことであり、それによって面積を出すことです。

ちなみに細かく切っていくことを微分\*1といいます。“微”に“分ける”という意味です。



(図表4-12)

積分は上の曲線の式が決まれば、“数学的”に計算できます。やり方はもう少し後で説明します。

\*1. 微分、積分について詳しく知りたい人は「『微分・積分』を知らずに経営を語るな」（内山力著 PHP新書）を読んでみて下さい。

## ヒストグラムをエクセルで書こう

さあケースの「おかあさん弁当の生産数」を考えてみましょう。

ここに先ほどの「積分の考え方」を入れてみたいと思います。

ただ体重は連続数ですが、「おかあさん弁当の生産量＝需要」は離散数（1.5個がない）です。したがってこのままでは体重と同様に「連続確率分布→積分」というものが使えません。

しかしよく考えてみると、図表4-2にある「度数」という離散数が、図表4-8では連続確率分布という“曲線”に変わっています。これと同じことをやれば、需要という離散数も“曲線”に変わり連続確率分布として考えることができそうです。

だからおかあさん弁当の需要も20代の男性の体重のように度数を取って、ヒストグラムを作り、これを細かくしていきましょう。

それでは図表4-1の毎日の需要のヒストグラムを、100ページの「書き方」にある「後者のやり方」で書いてみましょう。

「ランク」は「260」から「440」まで「20個区切り」とします。  
エクセルの2列に次のように並べます。

入力範囲	データ区分
需要	ランク
325	260
388	280
335	300
304	320
315	340
345	360
415	380
386	400
354	420
378	440
307	
388	
324	
336	
358	
334	
326	
318	
354	
288	
276	
334	
355	
334	
295	
326	
378	
364	
339	
308	
368	
326	
288	
345	
309	
310	
372	
328	

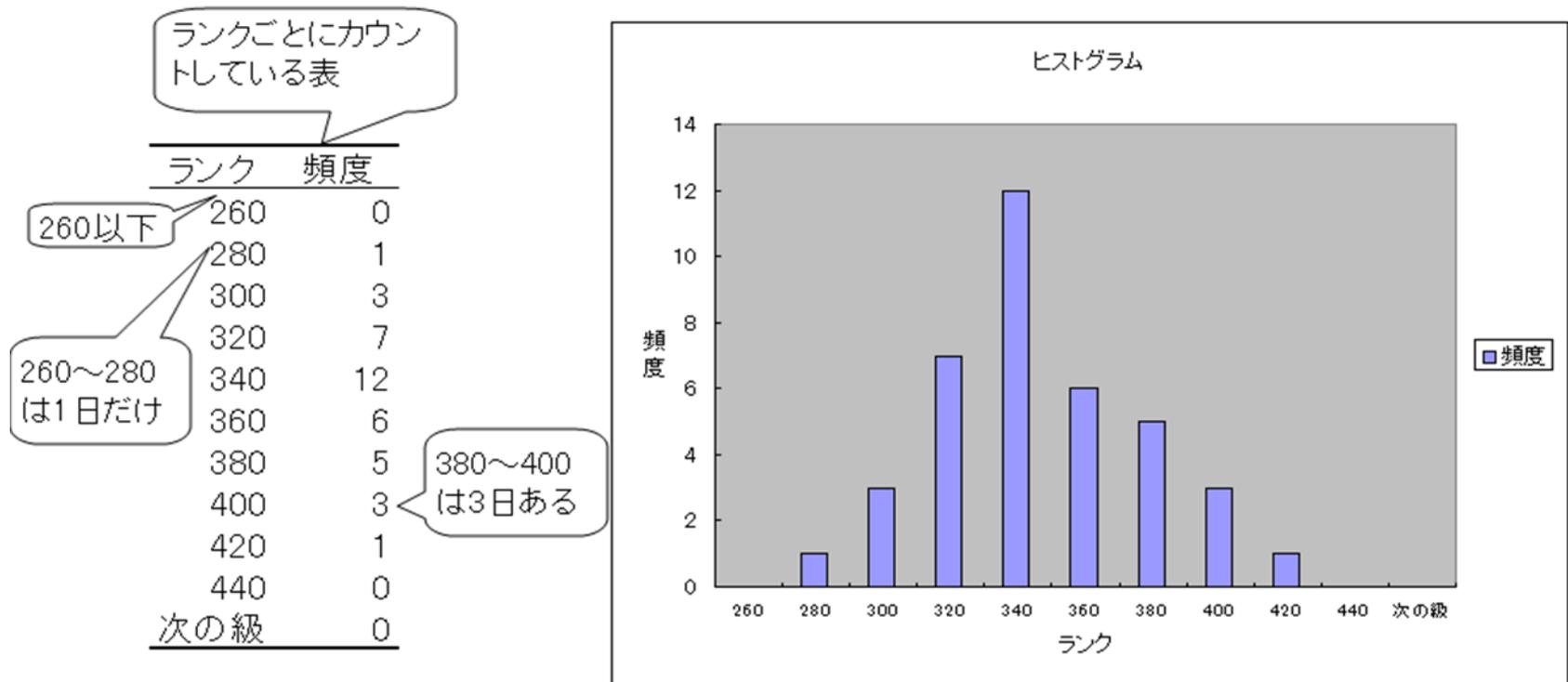
  

入力範囲	データ区分
需要	ランク
325	260
388	280
335	300
304	320
315	340
345	360
415	380
386	400
354	420
378	440
307	
388	
324	
336	
358	
334	
326	
318	
354	
288	
276	
334	
355	
334	
295	
326	
378	
364	
339	
308	
368	
326	
288	
345	
309	
310	
372	
328	

(図表4-13)

そのうえで100ページで述べたように、「分析ツール」（または「データ分析」）から「ヒストグラム」を選んで、入力範囲に左の列の「需要という表頭と数字」を、「データ区分」に右の列の「ランクという表頭と数字」を入れ、「ラベル」（需要、ランクという表頭を含めて指定していることを意味する）、「グラフ作成」をチェックして、OKを押します。

すると次のようなグラフになります。



(図表4-14)

## 山形で考える

このヒストグラムに対して、先ほどやったように「棒を細かくして行って、つなげて曲線にして、積分して確率を求める」とやっていきたいところです。

しかしそのためには、この曲線を確率密度関数という「式」や「線」に表わさないと、積分ができません。基本的には「どんな式にするか」（回帰式のような $x$ と $y$ の関係）は回帰分析と同様に、「図表4-14の需要の数字」からコンピュータ（エクセル）が決めてくれます。

しかし人間が大体どんな“線のパターン”にするかを決めないとできません。回帰分析の時は“直線”という“線のパターン”を選びましたが、60ページで述べたように、そのメニューでは直線の他に、指数曲線、放物線といった“線のパターン”を選ぶこともできました。

このケースのように確率を考える時は、97ページでさらっと述べた「正規分布」という「山形の曲線」（正確に言うと確率密度関数）を選択するのがノーマル（＝正規＝標準＝普通）です。

一般的に、ヒストグラムを作ってみて山のような形（中心が盛り上がって左右両側に下がっていく）の時は、正規分布を考えればOKです。

先ほどの20代男性の体重も、おかあさん弁当の需要も、そのヒストグラムを書くと“山形”（つり鐘型という人もいますが、山形のほうが直感的です）になりま

ます。ヒストグラムを作る目的は、この“山形”であることを確認することにあると言えます。

もしヒストグラムが山形にならない時は、標本のボリュームが少ないと思われるので、標本をもっとたくさん集めましょう。体重の例なら100人ではなく1000人、ケースの例なら2ヶ月ではなく1年にしてみるといったことです。それでもだめな時は、曜日単位（月曜日だけ）に標本を取るなどして、同じ特徴（数字の値が近いのではなく、「曜日が同じ」、「月の初旬」といったこと）を持ったデータを取れば、きっと山形となると思います。

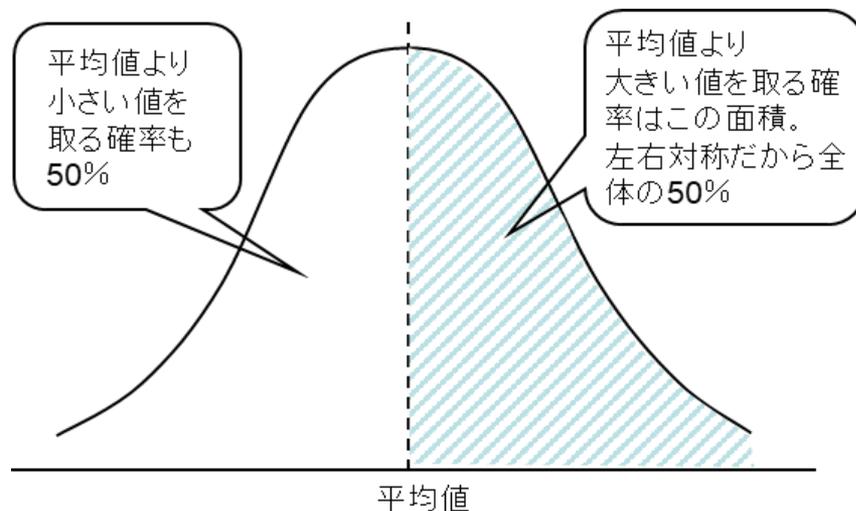
## 山形は左右対称

この正規分布には2つの特徴があります。

1つはその山形曲線が、「平均値を中心にして左右対称となること」です。

これはその値が平均値より大きい確率と小さい確率が、ともに50%（ $1/2$ ）ということの意味しています。（94ページの村田さんのせりふ「2日に1回は欠品だ」を思い出してみてください。）

確率は面積で表わされることは前にやりましたね。したがって次のようなイメージです。



(図表4-15)

## 平均値と標準偏差で形が決まる

2つ目は、正規分布は「平均値と標準偏差が決まると山の形（曲線）が1つに決まること」です。

曲線が1つになれば、あらかじめ「面積＝確率」を計算しておくこと、つまり積分の値を計算しておくことができます。

これを表にまとめたのが、次の標準正規分布表です。

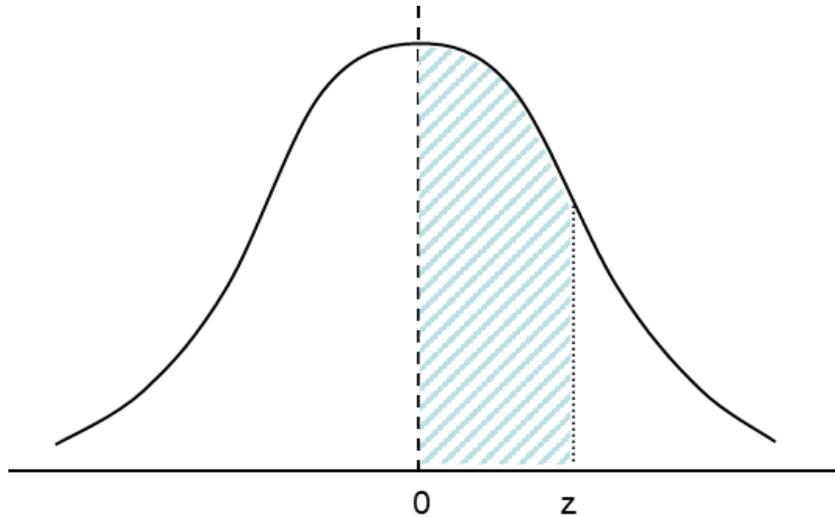
標準正規分布表

<i>z</i>	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993

② 0から0.5までを取る  
確率は0.1915≒19%

① 0から1.51までを取る  
確率が0.4345

これは「平均値が0、標準偏差が1の正規分布」について、次のような部分の確率（=面積）を計算しておいた表です。

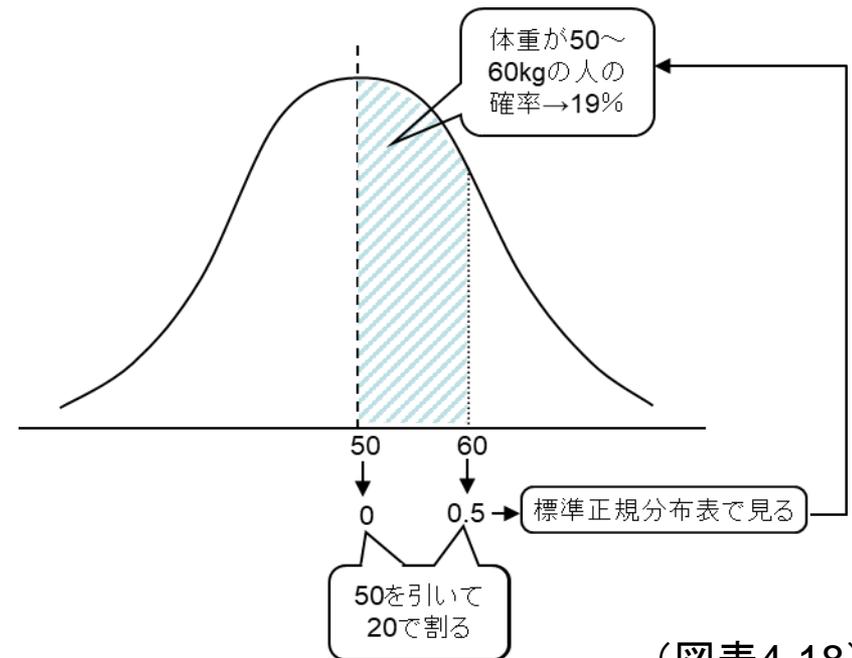


(図表4-17)

例えばその変数（図表4-17、図表4-18の  $z$ ）が、「0から1.51までを取る確率」（図表4-17の **a**）は0.4345、つまり約43%（網線部の面積）ということを意味しています。

では先ほどの20代男性の体重の例で、仮に平均値が50、標準偏差が20であった場合は、どうやってこの表を使ったらよいのでしょうか。

この時はすべての体重データから50を引けば、「平均値は0」となり、これを20で割れば「標準偏差は1」となります。



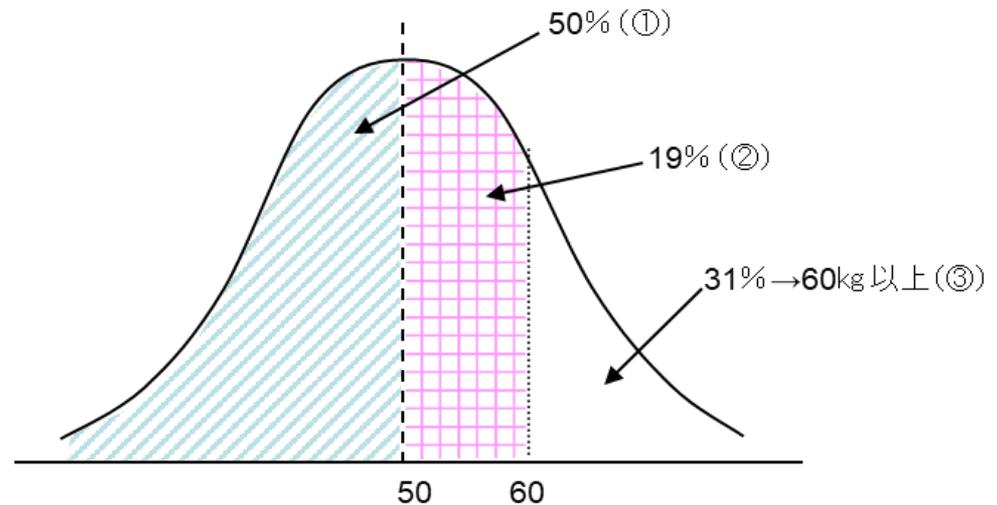
(図表4-18)

体重が「50kg～60kg」の人の確率を知りたいとします。まずこの「50～60」から50を引いて、「0～10」とし、これを20で割って「0～0.5」とします。

つまり標準正規分布表の0.5の部分を見ればよいこと（図表4-17の**①**）になります。この確率は約19%とわかります。

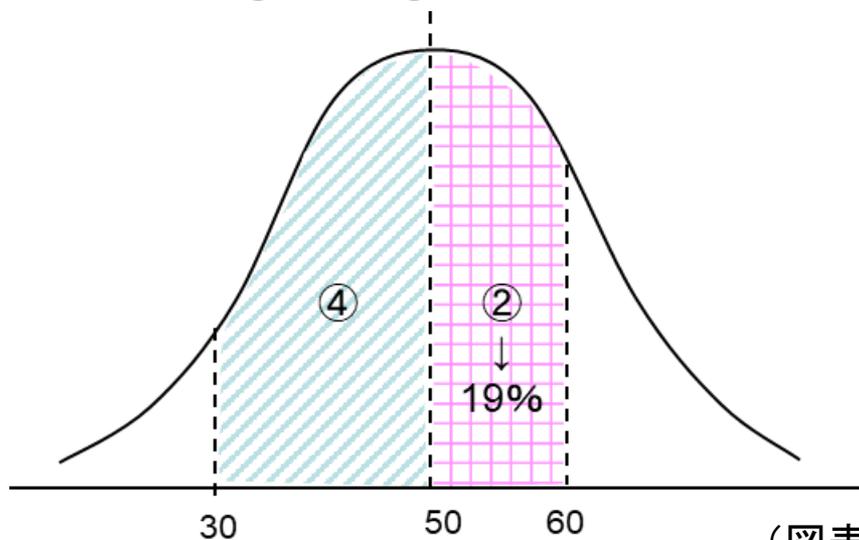
左右対称だから、いろいろな確率が出せる  
では体重が60kg以上の確率は？

平均が50kgですので、50kg以下は50%（図表4-20の**①**）、50～60kgが19%（**②**）なので、「60kg以下の人」は69%となります。したがって「60kg以上の人」は「 $100 - (50 + 19) = 31\%$ 」（**③**）となります。



(図表4-19)

では30 kg～60kgの確率は?

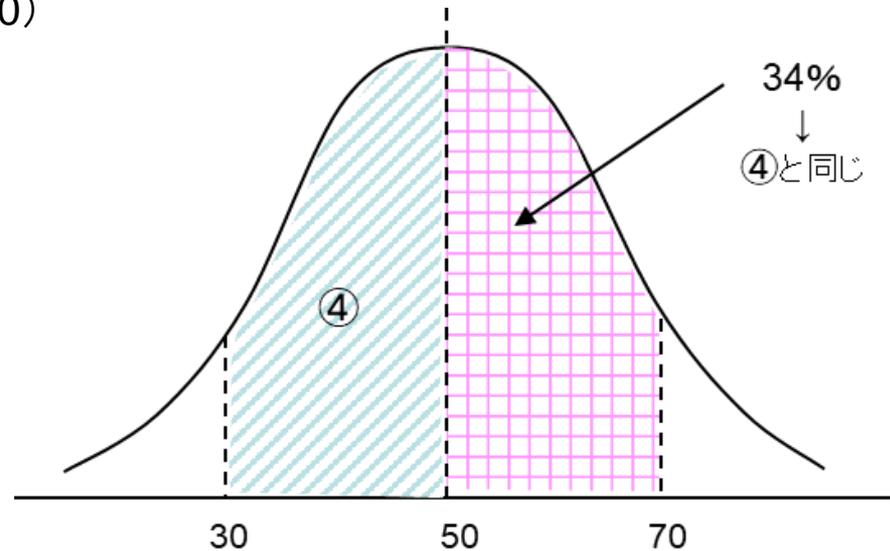


(図表4-20)

70から50を引いて20、それを標準偏差20で割って1.0となります。標準正規分布表で1.0の所を見ると0.34...なので、④の確率は約34%となります。

したがって30～60kgは34% (④) + 19% (②) = 53%となります。(図表4-21)

これは30～50kg (④) と50～60kg (②=19%) に分けて考えればOKです。  
30～50kg (④) は50kgから20kg離れています。左右対称ですので50～70kgの確率と同じです。



(図表4-21)

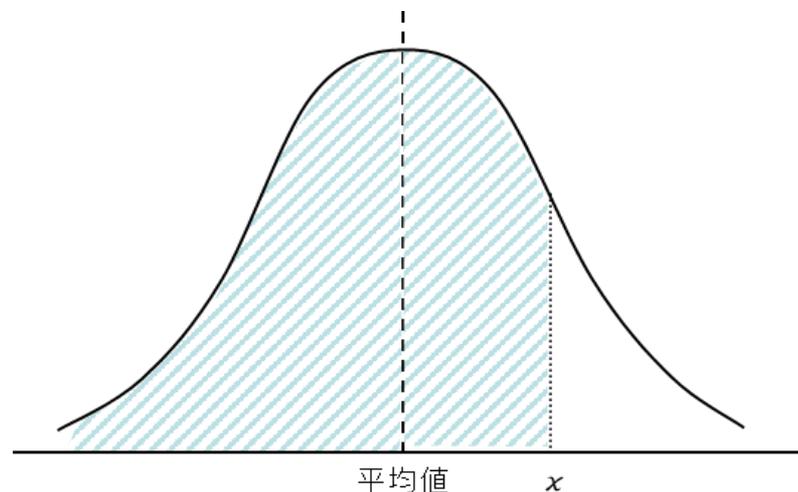
## 理解したら、エクセルを使おう

しかし引いたり、割ったりして、標準正規分布表を使うのは大変なので、この確率もエクセルを使って計算してしまいましょう。

でも前にも言ったとおり、大切なことは「エクセルで何を計算しているか」を、まわりの人に説明できることです。「需要を正規分布と仮定して、その面積から確率を出していること」を説明して、「なぜ面積が確率なのか」と聞かれたら、これまでのことを説明してください。

ここまでのことをしっかり理解できたら、やっとエクセルの「**NORMDIST**」という関数を使いましょう。

この**NORMDIST**は次のような網線部分の確率を計算しています。つまり「 $x$ 以下の値を取る確率」です。



(図表4-22)

20代男性の体重の例（平均値50、標準偏差20）で、NORMDIST を使って60kg以下の人の確率を求めてみましょう。

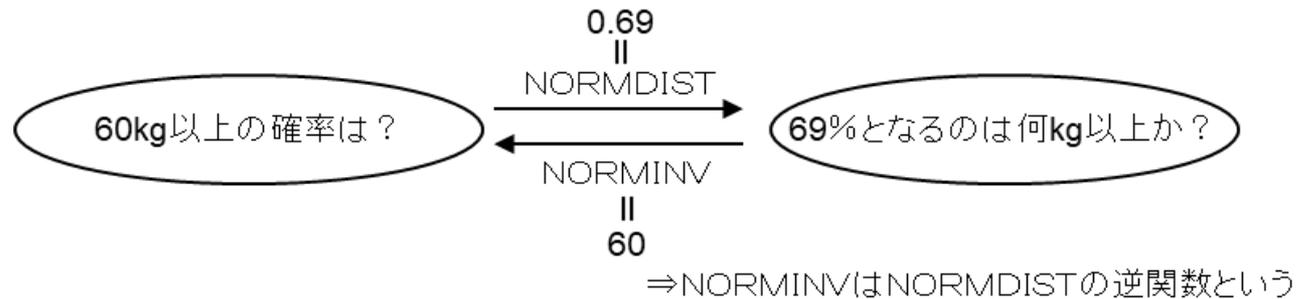
関数ですから、fxをクリックして「関数の検索」の欄に「正規分布」と入れ、NORMDISTを選びます。その「xの欄」に「60」、「平均」に「50」、「標準偏差」に「20」を入れます。関数形式の欄には「true」と入れます。

そうすると指定したセルに「0.69…」と出るとおもいます。つまり69%です。先ほど正規分布表から計算した結果も60kg以下の人の確率は69%でした。

エクセルでは逆に「ある確率を示す値」も計算できます。つまり先ほどとは逆に、「確率からxを求めるもの」です（xと確率の位置関係は図表4-22と同じです）。だからNORMINVをNORMDISTの逆関数といいます（NORMDISTはNORMINVの逆関数です）。

「体重を何kg以下とすれば、その確率が69%となるか」を考えてみましょう。

先ほどの正規分布の関数の欄でNORMINVを選び、確率に0.69、平均に50、標準偏差に20を入れると、「数式の結果=59.91…」と出ます。つまり約60kgとなります。



(図表4-23)

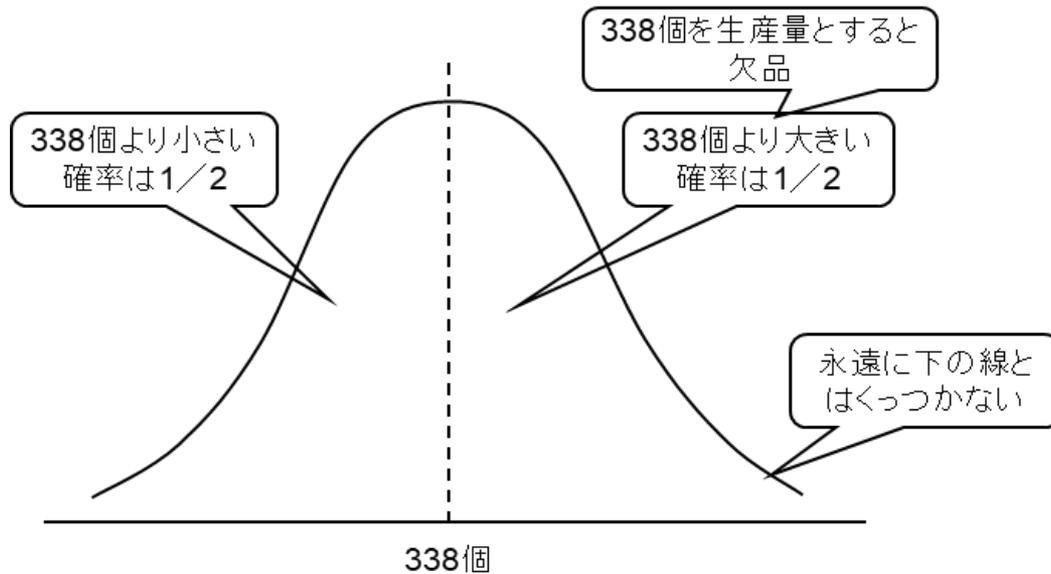
## 欠品はなくせない

さあ村田さんのケースに戻しましょう。

「おかあさん弁当」の需要を正規分布と考えます。

したがっておかあさん弁当の平均値だけでなく、標準偏差も計算しておく必要があります。21ページでやったように、エクセルの関数STDEVPで94ページの60個のデータを指定して計算してみると、標準偏差は31となります。

平均値は338でした。つまり「おかあさん弁当」を338個作ると、それより右の確率（需要が338より大きい）が50%あるので、2日に1回欠品します。



(図表4-24)

絶対欠品しないようにするには、この「正規分布の曲線と下の線のくっついた所」を生産量とする必要があります。

しかし正規分布の曲線には「右はし」というものがなく、永遠に下の線とはくっつきません。つまり「欠品を出さない生産量」というものはありません。

G社の生産調整室長もこれはわかっているようです。

1日平均338個しか売れていないのに、何かの拍子に「1000個欲しい」という需要が生まれることもあります。東日本大震災の時、「すべての店でトイレットペーパーやミネラルウォーターが欠品してしまった」といった現象です。

しかし、だからと言って「1000個売れる」という“めったにない現象”のために、1000個作っておくのは賢くありません。

### どこまで欠品を許すか

欠品をなくすことはできないので、「どれくらいの欠品まで許すか」ということを考えなくてはなりません。これを確率で表わしたものを、許容欠品率とよびます。

許容欠品率はケースバイケースで決めていくのですが、10%、5%、1%といった「切れのよい数字」を使います。

「許容欠品率10%」ということは「10日に1回欠品することは仕方がない」と考えることです。

しかし“需要”のような「世の中で起きている一般的な現象」は、数学が求める「正確な正規分布」を示すわけではありません。正規分布と仮定して計算すると、ほとんどの場合に現実より“少し大きめに”欠品率が計算されます。

欠品率を10%に設定すると、本来は10日に1回欠品するということですが、現実の世界では、もう少し欠品の回数は少なくなると思います。感覚的には「あまり欠品しない」という感じですが。

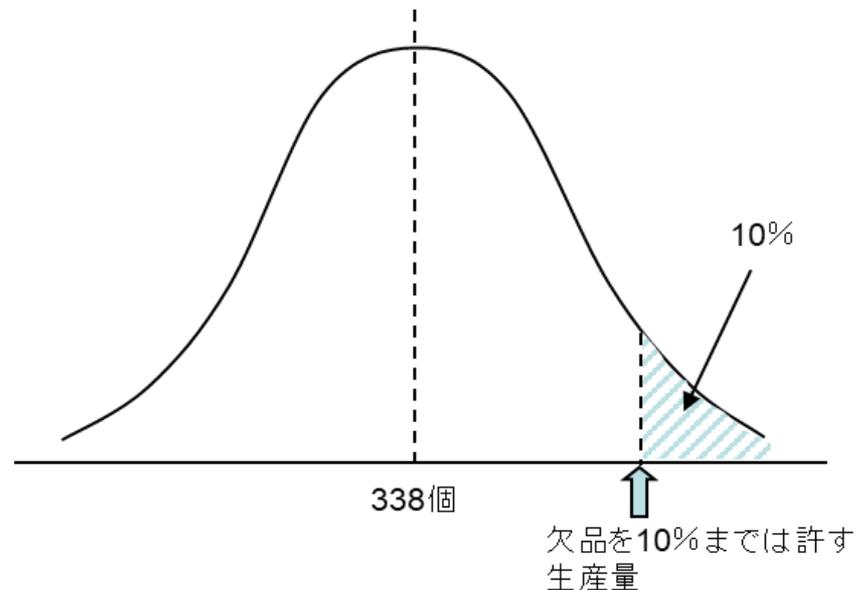
5%（20日に1回）にすると「めったに欠品しない」という感じであり、1%（100日に1回）に設定するのは「欠品は許されない」という状態の時です。

許容欠品率を小さくしていくと、生産量は増えていきます。需要(販売量)は生産量を増やしても変わらないので、その増えた分だけ「売れ残る量も増えていく」こととなります。

### さあ生産量を計算しよう

今回の「おかあさん弁当」では、「あまり欠品しない」という状況を作るために許容欠品率を10%に設定します。

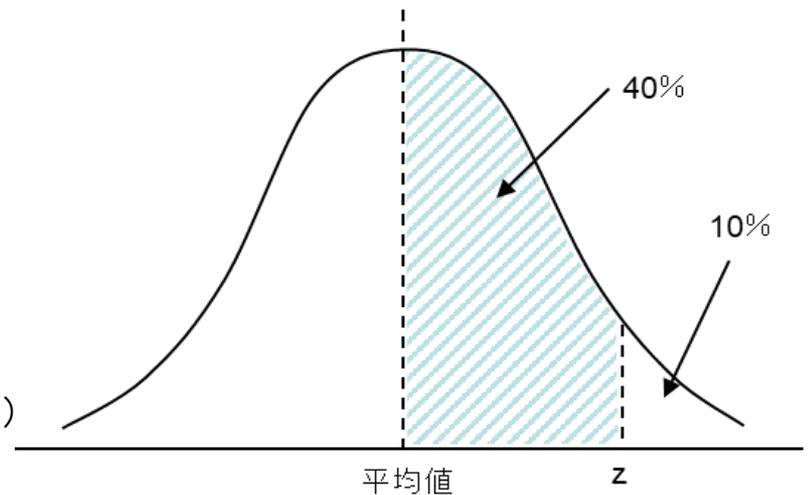
これで生産量を計算していきます。



(図表4-25)

まずは計算プロセスを理解するために、110ページの標準正規分布表を使いましょう。右はしを10%にしますので、図表4-17の標準正規分布表の斜線部分は、下のように40%となります。

(図表4-26)



そこで標準正規分布表（図表4-16）の数字の中で、0.40にもっとも近いものを探します。「1.28」（この時0.3997）が見つかりましたか？

「平均値0、標準偏差1」で1.28ですので、先ほどの逆の計算（「平均値を引く」→「平均値を足す」、「標準偏差で割る」→「標準偏差をかける」）をやりまます。つまり平均値338に「1.28におかあさん弁当の標準偏差31をかけた数」を足します。

$$338 + 1.28 \times 31 = 378$$

378個が「欠品を10%許す生産量」です。

この「1.28」を「許容欠品率10%の安全係数」といい、「 $1.28 \times 31 = 40$ 個」を「安全在庫」といいます。1日平均338個しか売れないのに、378個を在庫量（生産

量) とすれば、1日平均して40個の売れ残りが出ます。つまり安全在庫は「売れ残り量の平均値」といえます。

一般に「10%の欠品を許す生産量(需要)」は  
「**需要の平均値+1.28 (=安全係数) × 標準偏差**」  
で計算されます。

ここで許容欠品率を5% (0.05) としてみましよう。

標準正規分布表で0.45 (0.5-0.05) にもっとも近い所を探してみると1.64となり、これを安全係数として計算します。

$$338 + 1.64 \times 31 = 338 + 51 = 389$$

389個となります。この時安全在庫は51個となり、10%の時よりも11個増えます。

これで欠品の可能性が減りますが、欠品率10%の時より、1日平均11個分の売れ残りが増えることとなります。

この欠品率と平均売れ残り量 (=安全在庫) を考えあわせて、村田さんは生産量を決定することになります。

## 理解できれば、あとはエクセルで計算

このイメージを理解できたら、表を見るのはめんどうなので、今後はエクセルで計算するようにしましょう。

先ほど標準正規分布表でやったことを、エクセルでやってみましょう。

まずは許容欠品率10% (0.1) です。115ページで述べたとおり、「特定の確率を取る数値」を探す時はNORMINVです。

fxを押して、NORMINVを選び、確率に0.9 (115ページの図表4-21を見てください。「1-0.1」です)、平均値に338 (エクセルで計算した平均値のセルを指定してもOK)、標準偏差に31 (標準偏差のセルを指定してもOK) を入れると、「378」、つまり先ほど標準正規分布表で計算したものと同じ結果となります。

次は5%にします。NORMINV (0.95, 338, 31) で計算すると、389個と同じになります。(STDEVPで計算した平均値、標準偏差のセルを指定すると390個になります。)

ついでに1%でやってみると410個 (STDEVPで計算した平均値、標準偏差のセルを指定すると411個) になり、安全在庫は72個 (73個) となります。

m社は大都市圏にホテルを10ヶ所展開している。

前田はm社に入社し、主にフロント業務を担当してきたが、今回の定期人事異動で本社客室運営部へ配属となった。

客室運営部のスローガンは「ムダ、ムリ、ムラをなくしてホスピタリティ\*1を上げよう！」である。

前田は客室のアメニティグッズの在庫量の適正化を担当することになった。

上司から言われたのは次のようなことである。

「これまで当社では、アメニティグッズの在庫量に関して、何も考えてこなかった。ここにはムダが多くあり、1つ1つの単位は小さいが、『ちりも積もれば山となる』だ。君はこのアメニティグッズの在庫削減に当たってくれ。とは言っても、ホテル全体でいきなりやるのは大変なことだ。まずはn店の歯ブラシからやってくれ。そのうえですべてのアメニティグッズ、すべての店舗にこれを展開してくれ」

前田は上司に質問した。

「アメニティグッズの在庫管理はどうやっているんですか？」

「何も考えずに適当にやっているんじゃないか。倉庫へ行くと、いつも歯ブラシが山のようにあるからな」

前田は在庫管理について、本を買って勉強した。その本には次のようなことが書いてあった。

「コンスタントに使われる資材などには、一般に定量発注方式が取られる。ある一定量の在庫量を切ると、一定量の資材を発注するものである。

この発注するタイミングの在庫量を発注点といい、発注リードタイム分の需要をこれに充てる。発注量は仕入先との交渉で決まる。

発注リードタイムを短かくし、発注量を小さくすると、資材の在庫は削減できる」

前田は思った。

「よし、歯ブラシは定量発注方式としよう。発注リードタイムと発注量を確認してみよう。」

「発注リードタイムは3日か、これは納入先と交渉すれば1日にできるはずだ。発注量は2000本か、ちょうどダンボール1箱だ。これを小さくするのはちょっと難しいなあ。

発注点を考えてみよう。こちらはうちの意思で決定できる。発注点を決めておいて、これを切ったら、うちが発注するのだから。

そのうえで発注リードタイムを3日から1日にしたら、どれくらい在庫が削減されるかを考えてみよう。

まずはとりあえず先月の歯ブラシの使用実績を調べてみよう。」

結果は次のとおりであった。

日	1日	2日	3日	4日	5日	6日	7日	8日	9日	10日	11日	12日	13日	14日	15日
使用量	138	134	85	98	133	122	140	135	145	125	110	98	85	116	108
日	16日	17日	18日	19日	20日	21日	22日	23日	24日	25日	26日	27日	28日	29日	30日
使用量	120	63	124	122	118	116	55	70	100	106	108	114	152	85	78

(図表4-27)

ここから先はあなたが前田さんの立場になってやってみましょう。

発注点を求め、次に発注リードタイムを3日から1日にしたら在庫量がどう変わるかを計算してみてください。

(ヒント) 3日間の需要の標準偏差は、1日の需要の標準偏差の $\sqrt{3}$ 倍となります。

\* 1.ホスピタリティ 心をこめてサービスをする精神のこと

まずは1日あたりの歯ブラシの需要の平均値と標準偏差を、エクセルで求めてみましょう。

110と24になりましたか？

さあ発注点です。問題文に書いてあったとおり、発注点は発注リードタイムである「3日間の需要」がしのげればOKです。もちろん平均需要だけでなく、安全在庫も加味しなくてはなりません。

歯ブラシ1日の需要の平均値は110ですので、3日間の需要の平均値は330です。問題は標準偏差です。1日の需要の標準偏差が24ですので、3日間の需要の標準偏差は $24 \times 3 = 72$ でしょうか？

残念ながらちがいます。3日間の需要の「分散」が3倍になるので、ヒントに書いてあったように「標準偏差」は $\sqrt{3}$  倍となります。したがって3日間の需要の標準偏差は42 ( $24 \times \sqrt{3}$ ) となります\*1。

\*1. どうして $\sqrt{3}$ 倍になるか気になる人向けに、もう少しだけ説明します。

正規分布には加法性があります。加法性とは「足し算ができる」ということです。この加法性は「平均値」と「分散」で見られます。つまり2つの独立した正規分布がある時（独立とは互いに影響を与えないという意味）、「その2つの和の正規分布」の平均値、分散は、それぞれ2つの分布の和となります。

もう少し数学っぽく書きましょう。

$N(\mu, \sigma^2)$  とは平均値が  $\mu$ 、分散が  $\sigma^2$  (標準偏差が  $\sigma$ ) の正規分布を表現しています。これを使って上の加法性を表すと次のようになります。

「独立した確率変数  $x$ 、 $y$  がそれぞれ  $N(\mu_1, \sigma_1^2)$ 、 $N(\mu_2, \sigma_2^2)$  に従う時、 $z = x + y$  という確率変数は、 $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$  に従う。」

ここで歯ブラシの毎日の需要は独立しています。つまり昨日は歯ブラシをたくさん使ったからといって、今日、歯ブラシの需要が変わるわけではありません。したがって3日間の需要の平均値、分散は、1日の需要のそれぞれ3倍となります（3日分を足したものになる）。

1日の需要の分散は「 $24 \times 24$ 」（標準偏差は24。これの2乗が分散）ですので、3日分の需要の分散は「 $24 \times 24 \times 3$ 」となります。したがって標準偏差は「 $\sqrt{24 \times 24 \times 3} = 24 \times \sqrt{3} \doteq 42$ 」となります。

まあビジネスで使うことを考えると、このあたりは読みとばしてもOKです。要するに3日間の標準偏差は1日の $\sqrt{3}$ 倍、4日間の標準偏差は1日の2倍です。

次は許容欠品率です。歯ブラシがなくてはホテルになりませんから、「品切れは許されない状況」ということで「1%」としましょう。

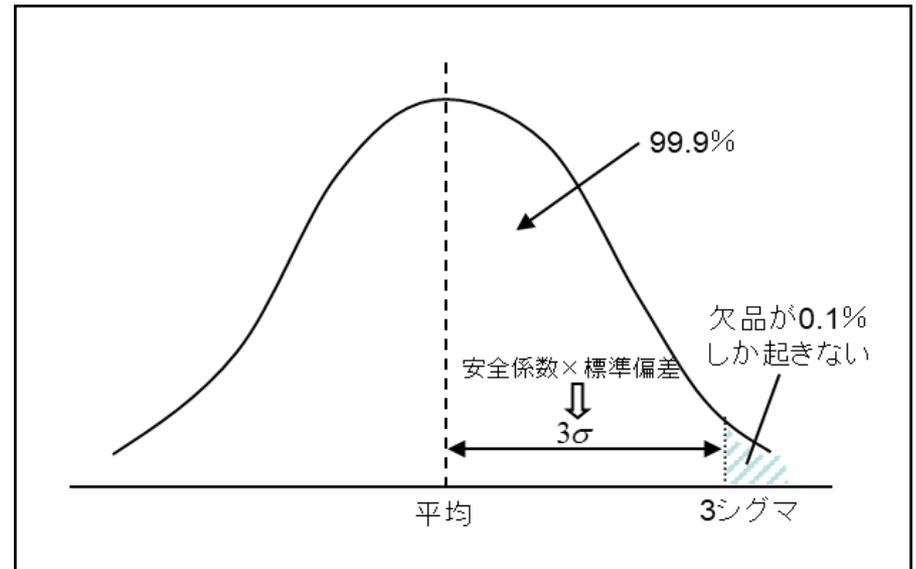
さあエクセルのNORMINVで計算です。



$$\text{NORMINV}(0.99, 330, 42) = 428$$

つまり428個を発注点にすればOKです。

これでも欠品が心配なら、許容欠品率を1%から0.1%にまで下げると、460個になります。99.9%大丈夫ということで、この時安全係数は「3」となります。



(図表4-28)

「3シグマ」という言葉を聞いたことはありませんか？

このシグマは標準偏差のことです。21ページで述べたように標準偏差には $\sigma$ (シグマ)という文字が使われています。しっかり管理すること(欠品を0.1%に抑える)や、想定外のことが起こること(「歯ブラシが倉庫からなくなってしまう」)をよく「3シグマ外」といいます。

「君は3シグマ外だな」といわれたら、それは「3シグマの外」にあるということで「規格外」というような意味です。まあ「普通ではない」といったところです。(けなしているのか、ほめているのかはよくわかりません。)

シックスシグマという言葉もあります。標準偏差の6倍まで持っていくことです。これは製品の不良発生率を6シグマの右側(図表4-28の斜線部分)まで持っていくことです。「不良を出さない驚異の管理法」というような意味で使っています。

話はそれてしまいましたが、まあ歯ブラシは許容欠品率1%で、428個を発注点としましょう。つまり428個の在庫を切ったら、2000個の発注です。

さあ在庫量です。

定量発注方式では、在庫量は毎日、図表4-29のように変化します。

在庫量の最大が「発注量+発注点」、最小は0なので、平均在庫量は「(発注量+発注点)÷2」と考えられます。

したがって平均在庫量 =  $(2000 + 428) \div 2 = 1214$ 個です。

また歯ブラシの在庫スペース (置いておく場所の広さ) は、「最大在庫分」だけ必要となります。つまり2428個分必要です。

ここで発注リードタイムを1日 (需要の平均値110、標準偏差24) にした場合を考えてみます。

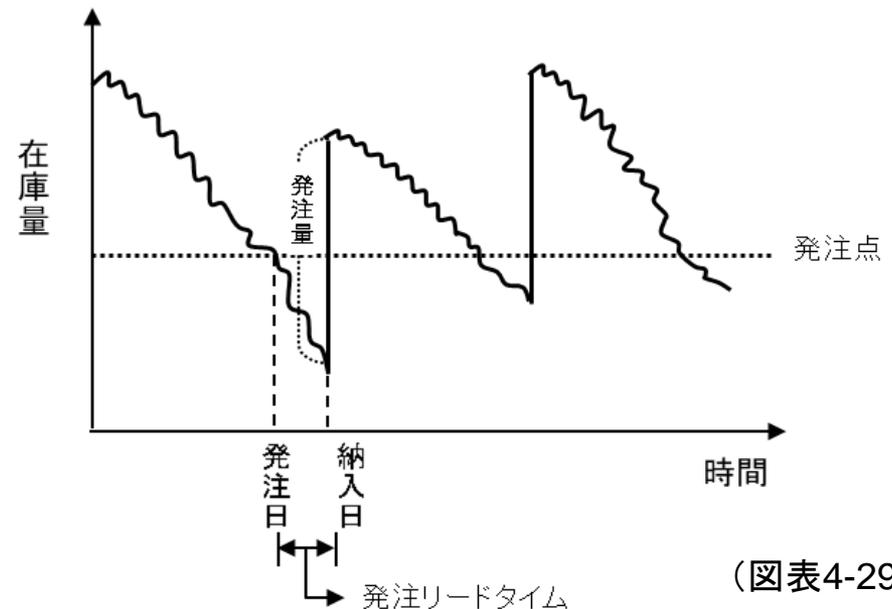
発注点 =  $\text{NORMINV}(0.99, 110, 24) = 166$ 個

となります

平均在庫量は

「 $(2000 + 166) \div 2 = 1083$ 個」

となります。



(図表4-29)

したがって発注リードタイムを3日から1日にすると、「 $1214 - 1083 = 131$ 個」だけ平均在庫が削減されます。

また最大在庫量は2166個分となり、262個 ( $2428 - 2166$ ) 分の在庫スペースが削減できます。

このように発注リードタイムの短縮でどれだけ在庫量が削減できるかが、数字としてつかめることができます。

前田さんがこのように確率を使って説明すれば、上司は「やるなあ」と評価するでしょう。

そして前田さんがこのノウハウをまわりに伝えていけば、m社は数字に強い、需要予測に強い、明日が読める会社へと変身していくはずです。

再び「数字活用ケースその1」のAコラーグルとXヨーゲンで考えてみましょう。

Aコラーグルの標準偏差は7.9、Xヨーゲンの標準偏差は15.8でした。

ここでAコラーグルの伸びが高いので、仮に両者の平均日販が同じになって、伸びが止まった状態を想定します。また標準偏差は相変わらず、それぞれ7.9、15.8だったとします。

この時AコラーグルとXヨーゲンの適正在庫量はどちらがうでしょう。

つまり平均値が同じで標準偏差が異なる時に、どれくらい在庫量がちがうかを考えてみて下さい。

許容欠品率は10%、5%の2通りで考えてみましょう。

この時には、AコラーゲンとXヨーゲンは平均需要が同じなので、在庫量は安全在庫の分だけ異なることになります（在庫量＝需要の平均値＋安全在庫）。

許容欠品率を10%と考えると、安全係数は1.28ですので、安全在庫は次のようになります。

$$\text{Aコラーゲルの安全在庫} = 1.28 \times 7.9 = 10 \text{個}$$

$$\text{Xヨーゲンの安全在庫} = 1.28 \times 15.8 = 20 \text{個}$$

つまり平均して同じ量だけ売れる商品でも、店舗に置いておく量は10個ちがうということです。つまり同じだけ売れても、ブレの大きいXヨーゲンの方が「10個分の広さ」の陳列スペースが必要となります。

これを許容欠品率5%にすると

$$\text{Aコラーゲンの安全在庫} = 1.64 \times 7.9 = 13$$

$$\text{Xヨーゲンの安全在庫} = 1.64 \times 15.8 = 26$$

となって13個もちがうことになります。

このようにAコラーゲルの「コンスタントさ」という商品力は、「在庫を圧縮できる力」ともいえます。この圧縮した商品の陳列スペースに、他の商品を置けば、商品の品揃えが増えます。

CONSTANTに売れる商品に絞り、商品種類数を増やし、売上を伸ばしていったのがコンビニエンスストアというお店です。コンビニエンスストアはこの方法で小売業界のチャンピオンとなりました。

24ページの「Exercise - 1」のd社、e社、f社、g社の4社についてもう一度考えてみましょう。

ここで考えたのはスピードと信頼性という指標でした。

このスピードと信頼性をセットにして、各社が「十中八九終えることのできる日数」を考えてみましょう。

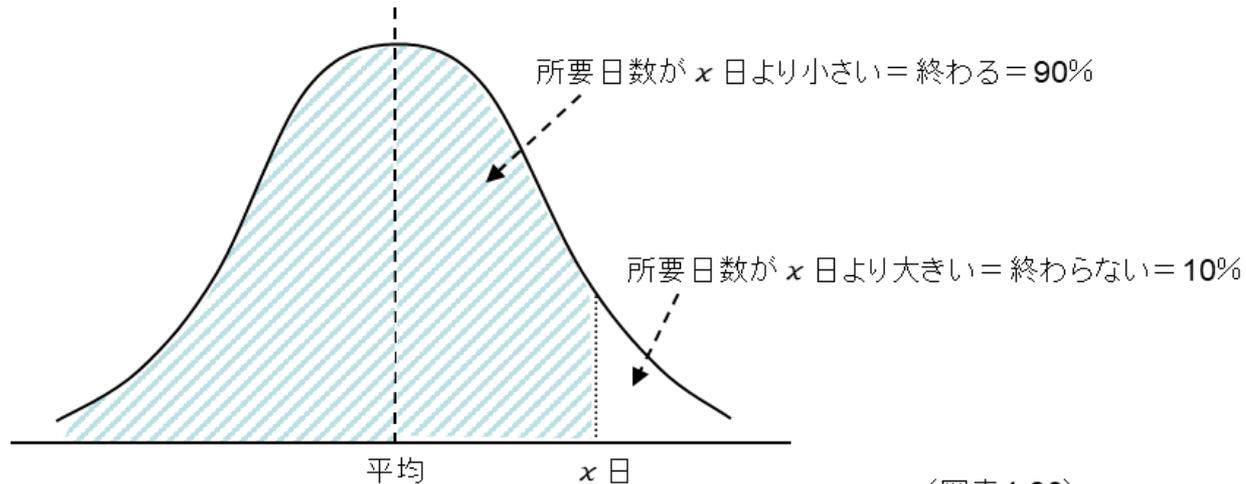
「十中八九」とは「10のうち8、9」は終わるというものですので、「90%の確率で終わる日数」を考えてみましょう。つまり各社は納期を何日以内とすれば、90%の確率で終わるかです。

さあ計算してみましょう。

25ページの発注リードタイムは、その仕事の所要日数と考えられます。

したがって90%の確率で終わる日数は、各社の所要日数から下の図の $x$ 日を求めることとなります。

もうわかりますね。NORMINVです。



(図表4-30)

	90%	平均値	標準偏差	
d社	↓	↓	↓	
...				
d社	NORMINV	(0.9, 30.3,	10.2)	= 43日
e社	NORMINV	(0.9, 31.1,	8.6)	= 42日
f社	NORMINV	(0.9, 30.7,	3.5)	= 35日
g社	NORMINV	(0.9, 25.4,	10.0)	= 38日

つまりf社の場合、「35日あれば、90%の確率で（十中八九）終わる」といえます。

スピードがもっとも速い(平均25.4日)と思われるg社でも、「38日」となっています。

さらには「ほとんど終わる」という感覚で95%にすると、各社は47日、45日、36日（平均値、標準偏差をセルで指定すると37日）、42日となります。

「絶対に終わる」という感じの99%にすると54日、51日、39日、49日となります。

99%に設定するとf社（39日）と他社（54日、51日、49日）では10日～15日も差が出ます。

「納期を何日以内とすれば終わるか」というのは、スピードと信頼感（安定感）をセットにした指標といえます。この指標で選んでもやはりf社となります。

o社は冷凍食品メーカーであり、現在、「冷凍とんこつラーメン」という新商品を開発中である。

開発担当の宮本はスープの味に悩んでいた。

とんこつスープはその味を「好きか嫌いか」ということもあるが、好きな人でも「こってりしたものを求める人」、「あっさりさを求める人」がいて、その調整は極めて難しい。万人に受ける味というのはないのであろうが、できるだけ多くの人の心をとらえたいと思っている。

評価点	評価内容
0	あっさりしすぎて食べたくない
1	かなりあっさりしている
2	あっさりしている
3	少しあっさりしている
4	気持ちだけこってりさせた方がいい
5	ちょうどいい
6	気持ちだけあっさりさせた方がいい
7	少しこってりしている
8	こってりしている
9	かなりこってりしている
10	こってりしすぎて食べられない

宮本は上司の了解を得て、500人に試作品を食べてもらうことにした。

いくつかのアンケート項目があるが、最大の関心事はスープであった。

このスープについて次のような評価点を使うこととした。

宮本は他のアンケート項目の評価を含め、500人分の評価点をエクセルに入れていった。

消費者No.	スープ 評価点	麺 評価点	.....
No.1	3	4	
No.2	5	3	
No.3	⋮	⋮	⋮
	⋮	⋮	⋮
	⋮	⋮	⋮

スープについて、500人の評価点の平均値と標準偏差を取ってみると、それぞれ4.8、1.3となった。

また評価点ごとにカウントしてみると次のようになった。

1回目調査の度数表

評価点	度数
0	0
1	2
2	9
3	75
4	115
5	179
6	68
7	43
8	8
9	1
10	0
合計	500

評価1の  
人が2人

「やはりスープの味は人によってさまざまだな。平均点は4.8か、まあまあだな。5より小さいんだから、もう少しだけこってり感を付けてみよう」

宮本はこってり感を少し強めて、再度500人に試食してもらうことにした。

2回目アンケート調査の結果は次のようになり、平均値は5.2、標準偏差は1.6と計算された。

2回目調査の度数表

評価点	度数
0	0
1	1
2	5
3	53
4	108
5	155
6	72
7	57
8	38
9	8
10	3
合計	500

(図表4-32)

「平均は5.2か。1回目が4.8で2回目が5.2か。5.0となるのが最高だけど、それには500人の全員が5をつけなくてはならないから、もちろん無理に決まっている。

1回目、2回目とも、5.0という満点から、0.2だけ離れているから同じなのかな」

そこでマーケットリサーチをよくやっているマーケティング部に聞いてみたら、次のような答えだった。

「うちはこんな時、アンケート評価点が4.5～5.5のゾーンに、回答者の何%が入るかを考えている。これをジャストフィットゾーンと呼んでいるんだ」

宮本は思った。

「ジャストフィットゾーンってどうやって計算するんだろう？」

さあアンケートの1回目調査と2回目調査について、ジャストフィットゾーンに何%いるかを計算してみましょう。

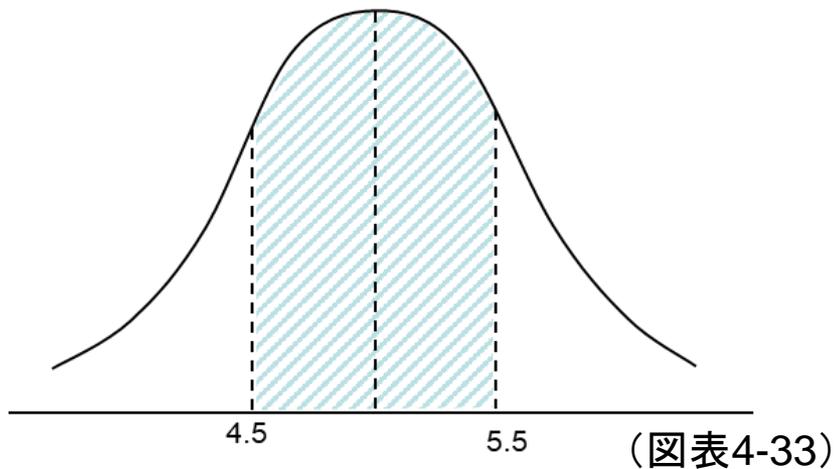
ここまで本書で勉強してきたあなたなら、もうジャストフィットゾーンの確率計算はできますよね。

この計算を自分で最初から最後まできちんとやるには、ばかばかしいかもしれませんが、1回目と2回目の度数表から元のデータを作りましょう。つまり1回目調査なら、133ページの1回目調査度数表を元に、数値が「1」のセルを2個、「2」のセルを9個、「3」のセルを75個・・・と作っていきます。

そのうえで1回目と2回目の500個のデータの平均値と標準偏差を出します。

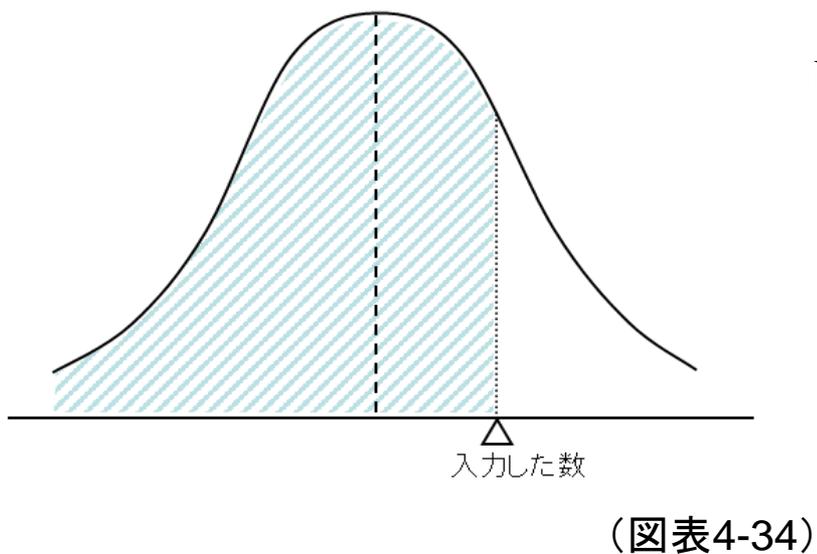
1回目調査の平均値が4.76、標準偏差が1.29、2回目調査の平均値が5.23、標準偏差が1.55となったのでしょうか。（めんどくさい人は問題文にある平均と標準偏差を使ってもOKです。）

1回目調査から、4.5～5.5にある確率を考えましょう。



115ページで述べたようにNORMDISTでは次のような部分の確率が計算されます。

しがたって次のように計算すればOKです。



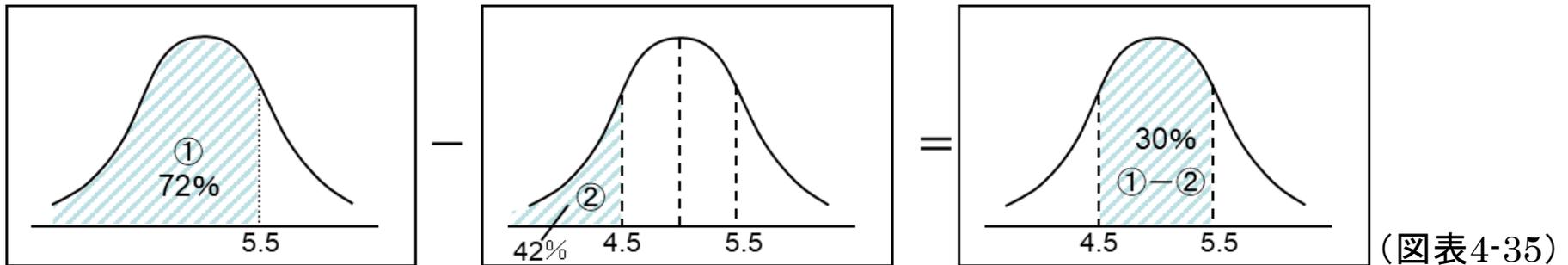
1回目調査の平均値が入っているセル

1回目調査の標準偏差が入っているセル

NORMDIST(5.5, , , TRUE) = 0.716 ← 5.5以下の確率 = 下図①

— NORMDIST(4.5, , , TRUE) = 0.419 ← 4.5以下の確率 = 下図②

0.297 ← 5.5～4.5の確率 = 下図① - ②



つまり4.5～5.5のジャストフィットゾーンは30%です。

同様に2回目調査でもやってみましょう。結果は25%となりましたか？

つまり1回目の30%から、2回目は25%へと落ちていきます。

このジャストフィットゾーンという指標を使うのであれば、1回目の味を採用すべきといえます。

# 数字で相手を説得する

## 検定をマスターして説得力を高めよう

ここまで学んできた統計テクニックは、主に「推定」とよばれる分野です。標本をサンプリングして、それを使って母集団という本当に知りたい数字を考えるものです。

「推定の裏返し」ともいえるものが、Lesson5で学ぶ検定です。検定とは、推定したことが正しいかをチェックするものであり、自分が考えたことについて「統計学的に言って正しい」ということを証明するものです。

検定は推定に比べ、まわりくどくて“ややこしい”のですが、ゆっくり読んでもらえば必ずわかります。これによって数字活用力だけでなく、論理性という能力の向上も期待できます。

あなたがLesson5で検定というテクニックを身に付ければ、人を説得するには「最高の道具」だということを実感してもらえらると思います。

## 数字活用ケースその5

——「効果があった」と言いたい——

内田さんは自社の商品パッケージを変更したいと考えています。そこでパッケージを変更してテスト的に販売してみると、売上が少し伸びました。

その結果を基に、まわりの人へ「パッケージを変えると効果がある」と言ったのですが、「そんなのたまたまじゃないの」と言われて、返す言葉がありません。

何とかこのデータだけで説得したいと思うのですが……。

あなたが内田さんならどうやって説得しますか？

## パッケージを変えれば売上は伸びる？

H社は洗剤、化粧品などのトイレタリー用品を生産しているメーカーである。H社の基幹商品である「しっとりシャンプー」は、10年前に発売して以来コンスタントに販売が推移していた。

H社マーケティング室の内田は、他社のシャンプーを見ながら考えていた。

「うちの『しっとりシャンプー』はパッケージが白地で地味だなあ。今1つインパクトがない。『シャンプー＝白＝清潔』というイメージにこだわりすぎている。

ドラッグストア、コンビニ、スーパーなどでは、消費者は数あるシャンプーの中からたった1つを選ぶのだから、もっと目立つパッケージにすべきだと思う。思い切ってベースをオレンジ色にしたらどうだろう。

そうすれば商品認知力が高まって売上が伸びると思う。」

内田はこれを商品会議に提案した。

商品会議での意見は2つに分かれた。「オレンジにすると売上が落ちる」という意見はなかったが、「パッケージを変えても売上は変わらないと思う。変えるにはコストがかかりすぎるから、やめた方がいい」と「目立てば、きっと売上が伸びるから変更しよう」という意見に2分された。

結局、商品会議では合意が得られず、とりあえずテストマーケティング\*<sup>1</sup>を実施してみることにした。

内田はオレンジ色のパッケージを作り、従来から良好な取引関係にあるドラッグストアI社に、10店舗でのテスト販売を依頼し、了承を得た。

I社の10店舗で、従来の白色のパッケージを標準価格で4週間販売した後、オレンジ色のパッケージを同価格で4週間販売し、それを比較してみることにした。

**\*1.テストマーケティング** 実際に販売する前に商品をテスト的に販売してみること。

**よし、いい感じだ！**

結果が出たので、内田はこれをエクセルに入れて次のように整理した。

	1週目							第1週 合計	第1週 平均
	1日目	2日目	3日目	4日目	5日目	6日目	7日目		
従来パッケージ	191	192	194	195	197	198	200	1367	195
新パッケージ	206	203	200	197	194	191	188	1379	197

	2週目							第2週 合計	第2週 平均
	8日目	9日目	10日目	11日目	12日目	13日目	14日目		
従来パッケージ	197	198	200	203	206	207	209	1420	203
新パッケージ	189	191	192	194	195	197	198	1356	194

	3週目							第3週 合計	第3週 平均
	15日目	16日目	17日目	18日目	19日目	20日目	21日目		
従来パッケージ	174	176	177	179	180	182	183	1251	179
新パッケージ	189	191	192	194	195	197	198	1356	194

	4週目							第4週 合計	第4週 平均
	22日目	23日目	24日目	25日目	26日目	27日目	28日目		
従来パッケージ	197	194	191	188	185	182	179	1316	188
新パッケージ	200	203	206	209	212	215	218	1463	209

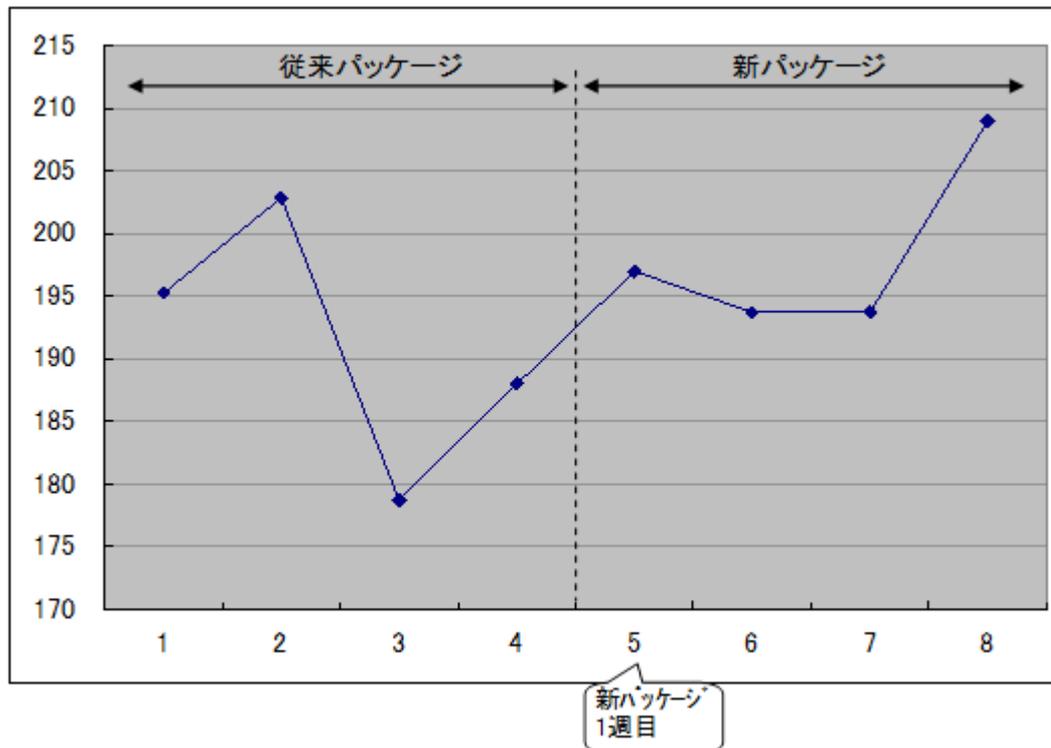
	28日 合計	28日 平均
従来パッケージ	5354	191
新パッケージ	5554	198

(図表5-1)

「従来パッケージが28日平均の日販数で191個、新パッケージが198個か。よし4%伸びているぞ。いい感じだな。

グラフにしてみよう。毎日じゃ細かすぎてかえってわかりづらいから、週ごとの平均をグラフにしてみよう。まずはグラフ用のデータを作ると。さあグラフだ。」

従来パッケージ				新パッケージ			
1週目平均	2週目平均	3週目平均	4週目平均	1週目平均	2週目平均	3週目平均	4週目平均
195	203	179	188	197	194	194	209



(図表5-2)

(図表5-3)

「グラフでは、従来パッケージの3週目の落ち込みと、新パッケージの最後の週が目立つなあ。それでかえってわからなくなっている。まあ全体として見れば、新パッケージの方が勝っていることは事実だけだ」

### たまたまじゃないの

内田はこのデータを使って、商品会議で説明した。

「I社の10店舗の定番棚\*<sup>1</sup>で、しっとりシャンプーを従来パッケージおよび新パッケージで4週間ずつ、標準価格で販売したところ、28日の平均で日販数が4%くらい伸びました。したがって新パッケージへ変更すれば、売上は伸びると考えられます」

しかし商品会議でのメンバーの反応はあまりよくなかった。

「両方の日販数を見ると、従来パッケージでも209の時もあるし、新パッケージで188の時もある。

必ずしも『新パッケージで売上が伸びる』とはいえないんじゃないの。この結果はたまたまじゃないの。4%の伸びなんてびびるもので、誤差範囲だろう」

内田は返す言葉がなかった。

そして頭の中で考えていた。

「こんなにデータがあるのに『パッケージ変更すれば、売上を伸ばす効果がある』と言い切れないのは悔しいなあ。平均が増えているんだから『効果がある』と言ったけど、『たまたま』と言われれば、その通りかもしれないし…。

だからと言って、テストをこれ以上続けても、結局『たまたま』で終わってしまうような気がする。

パッケージの変更で、『2倍に売上が伸びる』といった歴然とした差が出るわけではないなあ。

しかも、これ以上I社で実験を続けさせてもらうわけにもいかない。何とかこのデータで、『新パッケージは売上を伸ばす効果がある』と言ってまわりを説得したいなあ」

\* 1.定番棚 特売などではなく、店で常に売っている商品を置いてある棚のこと。

## 逆から攻める

こういう時には「検定」というテクニックを使います。

まずは検定の対象となる母集団を考えます。ここでの母集団は何でしょうか。もう大丈夫でしょう。このケースでの母集団は2つです。1つは「従来パッケージでの日販数」、もう1つは「新パッケージでの日販数」です。

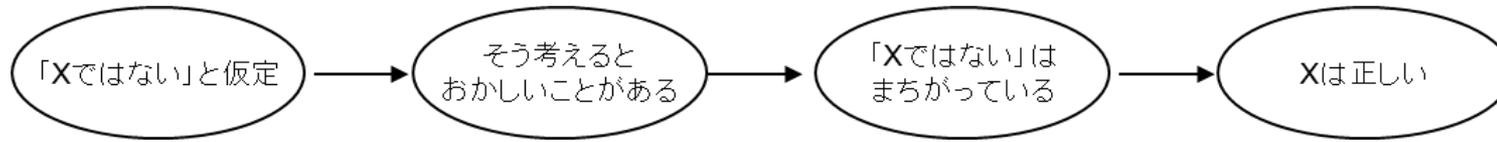
ここで検定の“方針”を決めます。

ここから先は少しまわりくどいので、自分の頭を混乱させないようにゆっくりと読んでいって下さい。

まずは「パッケージ変更をしても効果はない」と考えます。つまり「従来パッケージでも、新パッケージでも日販数は変わらない」ということです。

これは数学でよく使う手法で背理法というものです。X（この場合は「効果がある」）という仮説（正しいかどうか証明されていないのでこう表現します）を証明するために、「Xではない」（「効果がない」）という仮説を立て、どこかおかしい所がないかを探します。

仮にXが正しいなら、どこかにおかしいことがあるはずですが、もしおかしいことが見つかれば「Xではない」という仮説が「まちがっているから」と考えます。したがって「Xは正しい」というものです。



(図表5-4)

## 「効果がない」と考える

今回のケースで考えてみます。

「新パッケージにしても効果がない」という仮説を立てて、「おかしいこと」を探してみます。そのために標本を使って、何らかの統計量（「検定のための統計量」なので検定統計量といいます）を計算します。

次にその検定統計量が、その値を取る“確率”を出します。その“確率”が一定値より小さければ、「おかしい」と考えます。確率が一定値よりも小さいということは、「そんなことはめったに起きない」ということです。「めったに起きないこと」が「たまたま起きた」（先ほどの商品会議での発言）というのは「おかしい」、そうなる「仮説はまちがいと考えた方が自然だ」と結論づけます。

「効果がない」という仮説がまちがいということは、「パッケージ変更は効果がある」＝「パッケージを変更すれば売上が伸びる」を結論としようというものです。

この時「効果はない」という最初に立てる反対の仮説は、「まちがい」として棄てることが多いので、「帰無仮説」（「無に帰す」仮説という意味らしい。なかなか文学的な表現です）といい、仮説を棄てることを「棄却」といいます。

## ややこしいので再度説明

ストーリーがややこしいので、もう一度説明しましょう。

「新パッケージにしても効果がない」ということが「正しい」として、今回のような結果、つまり141ページの図表5-1のエクセルの表にある「28個ずつの日販数」となる確率を出します。

もしそれが1%なら、100回に1回しかないということです。「100回に1回しか起きないことが、たまたま今回のI社の10店舗で起こった」と考えるよりも、そもそも「効果がない」と仮定したことがまちがいと考えた方が自然でしょう。だから「効果がある」ということです。

そう考えれば効果がなくても、このようなデータになることは100回に1回あります。これを第1種の誤りといいます。つまり「効果がある」と結論づけても、「実は効果がない」ことも1%の確率でありうるということです。

ちなみに「第2種の誤り」とは、これとは逆に、本当は「効果がない」がまちがっているのに、棄却されず、「効果がない」と結論づけることです。

まあそれほど役に立つわけではありませんが、ちょっと数学を知っている人から「第1種の誤りは・・・」なんて言われた時のために、一応知っておきましょう。

【「効果がない」という仮説の検定エラー】

		本当は正しいこと	検定の結論
仮説を棄却⇒	第1種の誤り	効果がない	効果がある
仮説を棄却しない⇒	第2種の誤り	効果がある	効果がない

 (図表5-5)

### 「たまたま仮説」の正体

検定の方針が決まったので、実際にやってみましょう。

まず標本値（図表5-1の各4週分の日販数）を母集団に合わせて2つに分け、28個ずつのデータをエクセルに次のように2列に入れておきます。

A ↓	B ↓
従来パッケージ	新パッケージ
191	206
192	203
194	200
195	197
197	194
198	191
200	188
197	189
198	191
200	192
203	194
206	195
207	197
209	198
174	189
176	191
177	192
179	194
180	195
182	197
183	198
197	200
194	203
191	206
188	209
185	212
182	215
179	218

ここで従来パッケージの標本をA、新パッケージの標本をBとします。

そして先ほど考えたように、「新パッケージにしても効果がない」と仮定します。

「効果がない」ということは、Aの28個の標本とBの28個の標本は、「同じ母集団からサンプリングされた標本」と考えてもよいこととなります。（「効果がない」のだからAとBは同じ性質を持っているはずです）

ここでAの標本の平均値は191、Bの標本の平均値は198です。

今回の帰無仮説は「効果がない」＝「AとBは同じ母集団」＝「平均値の差はたまたま」ということですから、これは「AとBの平均値には差がない」という仮説と同じことです。そこで今回のような検定パターンを、「平均値の差の検定」といいます。

(図表5-6)

つまりこの191と198という平均値の結果は似たようなもので、サンプリングの仕方によって「たまたまちょっとちがっただけ」という仮説です。

確かに同じ母集団から2回に分けて標本を取り出して、その2つの標本の平均値が少しくらいちがっていても不思議でも何でもありません。

またこのケースでは棄却がその目的で「母集団に差があること」を証明するので、この検定のことを有意差検定ともいいます。

この「有意」という言葉も統計ではよく使う用語です。「有意」とは「偶然」(＝「たまたま」)の反対語です。だから有意差検定とは「たまたまを否定する検定」のことです。

### 「たまたま仮説」を否定

さあこの「たまたま仮説」を、検定という統計テクニックを使って否定します。

否定は先ほどの方針どおり、同じ母集団から28個ずつ標本を2回取って、図表5-6のような「56個の数字になる確率」が一定値以下なら、その仮説をまちがいとします。

この「正しい」と「まちがい」の境目となる確率を、棄却率または有意水準といいます。

棄却率をいくつにするかですが、5%（一般的な場合）、1%（少し厳しく考える場合）あたりが普通です。このケースでは5%としましょう。

残された課題はどうやって確率を出すかです。ここでもエクセルを使います。

エクセルを使うには正規分布の時のように、確率分布（線の形）を決める必要があります。

「平均値の差の検定」では、正規分布とよく似た形のt分析というものを使います。（t分布は標本数がかなり大きくなると、ほとんど正規分布と同じ形になります）

t分布はt値という検定統計量の分布です。t値は「A-B」の分布と考えてOKです。つまりA、Bという28個の標本の差を意味しています。

t分布と決めれば、114ページの正規分布のように確率をエクセルで計算することができます。

そこでこのようにt分布を使う検定を、t検定といいます。

これで準備OKです。さあエクセルの出番です。

図表5-6の2列のエクセルシートで、「分析ツール」（または「データ分析」）のメニューから「t検定 等分散を仮定した2標本による検定」を選びます。

ここで出て来た「変数1の入力範囲」として図表5-6の「従来パッケージの28個のデータ」(A)、「変数2の入力範囲」として「新パッケージの28個のデータ」(B)を指定し、OKをクリックします。

そうすると次のような表が出力されます。

t-検定：等分散を仮定した2標本による検定

	変数 1	変数 2
平均	191.2143	198.3571
分散	101.5079	64.38624
観測数	28	28
プールされた分散	82.94709	
仮説平均との差異	0	
自由度	54	
t	-2.93451	
P(T<=t) 片側	0.002448	
t 境界値 片側	1.673565	
P(T<=t) 両側	0.004897	
t 境界値 両側	2.004879	

←----- これが求める確率！

(図表5-7)

求めたかった確率は「P(T<= t)両側」というところにある数字、「0.004897」です。つまり約0.5%です。エクセルではこのような56個の数字になる確率を、0.5%と計算しています。

棄却率は5%です。0.5%はこれより小さいので、この帰無仮説は棄却されず。「めったに起きない現象」なので「仮説はまちがい」と考えます。

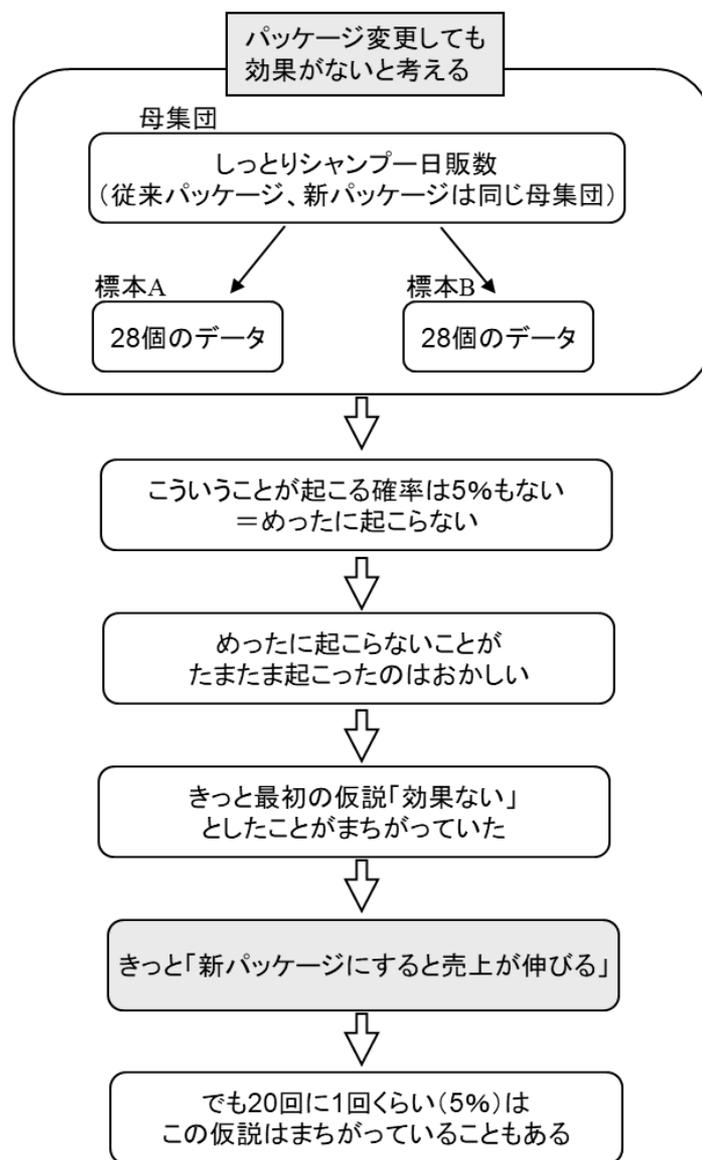
「パッケージ変更の効果はない」という仮説はまちがい、よって「パッケージ変更によって売上は伸びる」ということが結論です。

さあこれで高橋さんは商品会議で次のように発言できます。

「I社10店舗で行ったテストマーケティングの結果を分析しました。従来パッケージの28日分の日販数と、新パッケージの28日分の日販数を、t検定によって統計処理しました。その結果、この2つには統計上の有意差が見られます。すなわち新パッケージの日販数の方が大きいと結論づけられます。したがってこのテストの結果から、統計学的に言って『新パッケージの変更によって売上は伸びる』と判断するのが妥当といえます。」

**これしか人類は思いついていない**

くどうようですが、もう1回検定のフローを整理しましょう。



(図表5-8)

よくこういう説明をすると、説明した相手から「こんなややこしい方法しかないのか」という質問を受けます。

そういう時はこう答えましょう。

「あるかもしれませんが。しかしまだこれ以上良いやり方を思いついた人は人類史上いません」

つまり天才でもない限り、内田さんのようなケースではこのt検定というやり方を使う“しか”ないのです。

正確に言えばもう1つあります。こんな面倒くさいことをやらずに、ただ声を大きくして「効果がある」と言い張ることです。

あなたはこちらを選びますか？それともすっきりと「検定」を選びますか？

### ささいなことだが補足説明

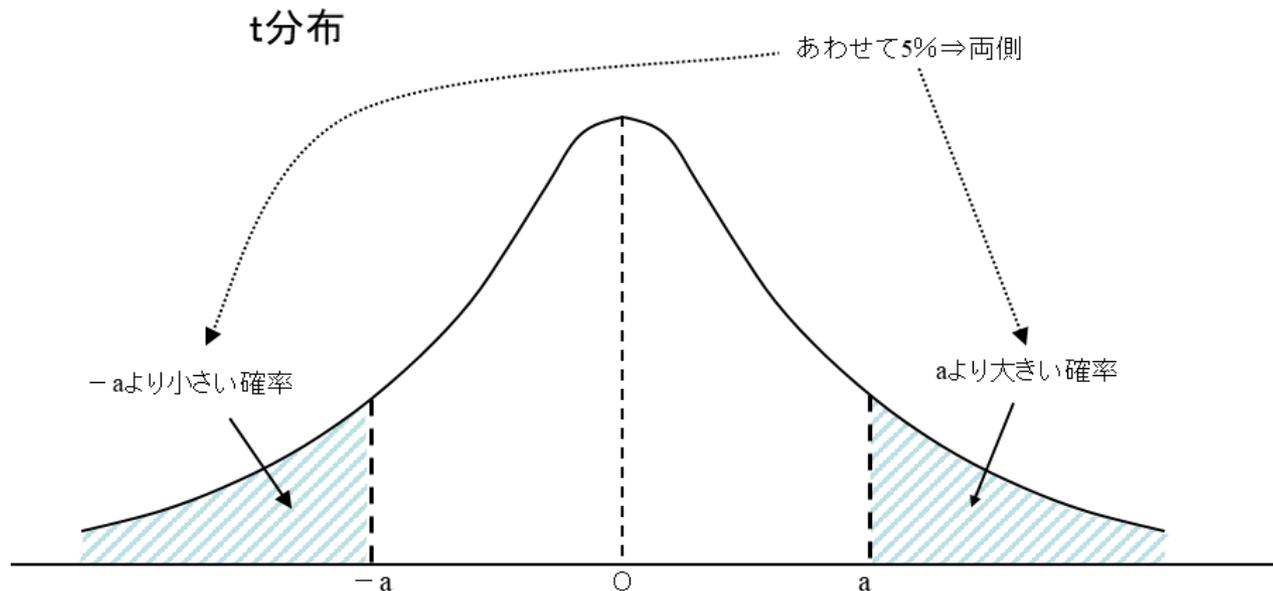
細かいことですが、ここで2つのことを補足しておきます。

1つ目は先ほどの検定結果の表にあった「 $P(T \leq t)$ 両側」の「両側」の意味です。

今回の帰無仮説は「AとBの平均値に差がない」ということであり、t分布は「A - B」の分布でした。

$A=B$ を否定するには、2つのパターンがあります。 $A>B$ （「 $A-B$ 」がプラス）か、 $A<B$ （「 $A-B$ 」がマイナス）です。

両側とは下の図の左側と右側の面積を足して5%になる $a$ と $-a$ （ $t$ 分布も左右対称）を境界線としています。



(図表5-9)

従来パッケージと新パッケージであれば「平均値に差がない」という帰無仮説は、「新パッケージ>従来パッケージ」（新パッケージにすると売上が伸びる）だけでなく、「従来パッケージ>新パッケージ」（新パッケージにすると売上が減る）の場合でも棄却されるということです。

ただこの場合、「平均値に差がある」が結論なら、そのデータから見て明らかに、「新パッケージ>従来パッケージ」です。だからこれを「正しい」と考えます。

もう1つはやはり検定結果にあった「自由度」です。自由度は「母集団から取り出した標本数から1を引いたもの」と考えればOKです。どうしても気になる人には、その意味を簡潔に説明しましょう。

「標本数から1を引く」という発想は、21ページで述べた分散（標準偏差の2乗）の分母を「 $n$ から $n-1$ に変える時」に考えられたものです。つまり $n$ 個の標本の自由度を $n-1$ と考えて、標準偏差を計算することです。標準偏差を計算するには平均値が必要です。ここでは標本の平均値=母集団の平均値と仮定しています。つまり標本の平均値はわかっている状態で、母集団の標準偏差を推定しています。そう考えると、 $n$ 個の標本のうち、「 $n-1$ 個の値」と「わかっている平均値」が決まれば、 $n$ 個目の標本の値は自動的に計算されてしまいます。だから標準偏差を計算する時の $n$ 個の標本が取る値の自由度は、 $n-1$ 通りというものです。

まあこんな理屈を知ってもそれほどの幸せはないので、一般人なら自由度は「標本数-1」と考えてしまいましょう。

先ほどt分布の例なら、Aの自由度が「28-1」で27、Bの自由度が「28-1」で27、合わせて自由度は54となります。これが149ページの表の自由度の所に出ています。

## バラツキだって検定する

Lesson1~4で述べてきた「推定」に何パターンもあったように、その“逆”とも言える「検定」にもいくつものパターンがあります。

平均値の差の検定（t検定）をやったので、次に「バラツキの差の検定」を考えてみましょう。

「バラツキの差の検定」では統計量として分散を使います。そのため分散検定ともいいます。分散は前に勉強したように、バラツキを表わす統計量である標準偏差を2乗したものです。

先ほどと同様に「AとBの標本の間バラツキの差はない」という帰無仮説を立て、棄却率を決め、その発生確率が棄却率以下なら棄却する、と進めていくのが分散検定です（何度も説明したので、この文章も一気に読めたと思います）。

平均値の差の検定では「 $A-B$ 」という差を表わすt値を使いましたが、ここでは $\frac{A}{B}$ という「分散の比」を表わすF値というものを使います。このF値の確率分布はF分布とよばれます。そのためこの検定のことをF検定ともいいます。

### Aコラーグルのコンスタントさを検定する

5ページ図表1-1のAコラーグルとXヨーゲンの1日あたりの個数という「84個×2のデータ」でF検定をやってみましょう。

「AコラーグルはXヨーゲンよりコンスタント」と言った時に、「今回たまたまそうなったんじゃないの」という反論を封じ込めるやり方です。

帰無仮説は「AコラーグルとXヨーゲンのバラツキに差はない」、棄却率は5%です。

エクセルを使います。あなたも自分でやってみましょう。

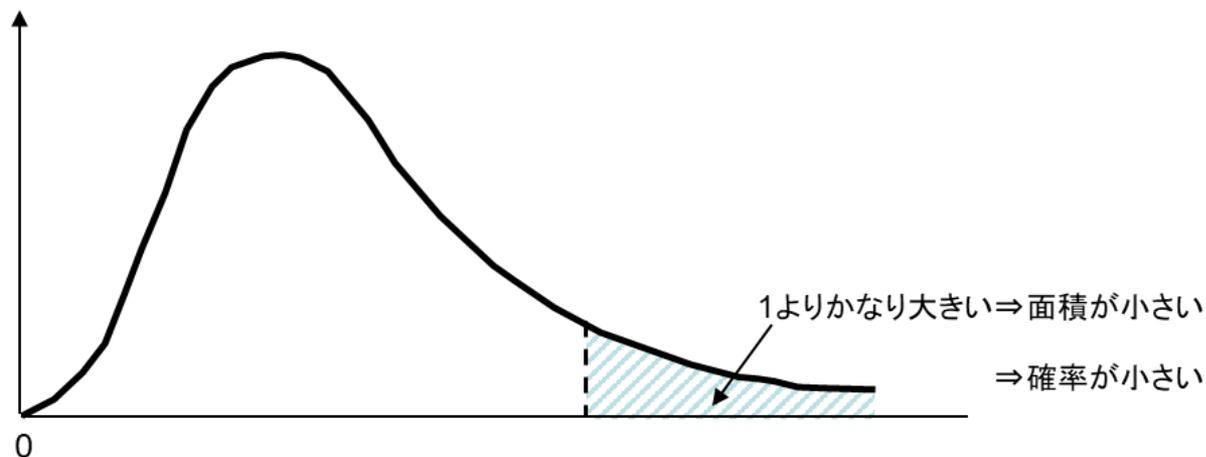
図表1-1のAコラーグル、Xヨーゲンの個数データ各84個をそれぞれ2列に並べます。

次に分析ツールのメニューから「F検定 2標本を使った分散の検定」を選びます。ここでデータを指定します。

ところでF分布は図5-10のような曲線で、左右対称ではありません。

そこで右端の確率を求めていくようにします。t検定が両側検定なら、F検定は片側検定です。つまりF値が一定以上大きいものを棄却します。

だから「分散の比」=F値を計算する時には、「分子>分母」として1より大きくし、1に近い時は「棄却しない」、1より一定以上大きい時は「棄却する」とします。



(図表5-10)

そのためエクセルのF検定でデータを指定する時には、F値を1より大きくするために、変数1（これが分子です）には分散（標準偏差）の大きい方を選びます。この例ですとXヨーゲンの84個のデータです。

変数2には分散の小さいAコラーグルを選びます。これでF値は1よりも大きくなります。

そのうえでOKをクリックします。すると次のような結果が出力されます。

F-検定 : 2 標本を使った分散の検定		
	変数 1	変数 2
平均	25.16667	18.59524
分散	253.7068	62.53299
観測数	84	84
自由度	83	83
観測された分散比	4.057168	←-----これがF値⇒1より大きいようにセットする
P(F<=f) 片側	4.49E-10	←-----これが求める確率
F 境界値 片側	1.437879	

(図表5-11)

「観測された分散比」というのが、先ほどのF値です。約4であり、1よりかなり大きく出ています。

ここで「P(F<=f)片側」を見ると「4.49E-10」となっています。これは4.49に $10^{-10}$ を「かけること」を表わしています。 $10^{-1}$ は0.1、 $10^{-2}$ は0.01ですので、4.49に $10^{-10}$ をかけると「0.000000000449」を表わしています。要するに、確率は極めて小さいので、「バラツキに差はない」という帰無仮説は棄却されます。

つまり「AコラーゲルとXヨーゲンのバラツキには差があり、 $A < X$ 」ということです。

「数字活用ケースその1」で山田さんが、もしバイヤーの加藤さんから「コンスタントなんてたまたまじゃないの」と言われたら、「F検定をした結果、統計学的に言って、AコラーゲルとXヨーゲンのバラツキには有意差が見られます」なんておしゃれに言い返すことができます。そして加藤さんは統計学的には「反論できません」。

### エクセルの「等分散の仮定」の意味がわかった

149ページでt検定をやった時に、エクセルの分析メニューの表示に「等分散を仮定した2標本による検定」という表現があったと思います。

これはA、Bという母集団の平均値の差の検定をやる時には、AとBの分散を「等しい」と考えるか、「等しくない」と考えるかでやり方がちがうということを意味しています。

そのため、本来は平均値の差の検定をやる前に、バラツキの差の検定をしなくてはなりません。

今回のケースでF検定をやってみましょう。

「従来パッケージと新パッケージのバラツキに差はない」が帰無仮説、棄却率は5%です。

エクセルでF検定を選び、147ページの2列を指定します。

ここで「観測された分散比」が1より大きくなっていることを確認します（変数1の分散 > 変数2の分散）。1より小さいなら、変数1と変数2を取りかえます。

F-検定 : 2 標本を使った分散の検定		
	変数 1	変数 2
平均	191.2143	198.3571
分散	101.5079	64.38624
観測数	28	28
自由度	27	27
観測された分散比	1.576547	→ 1より大きい
P(F<=f) 片側	0.121628	→ 12%
F 境界値 片側	1.904823	

(図表5-12)

やってみると「P (F<=f) 片側」は0.12 (12% > 5%) となり、棄却することはできません。

ということは「従来パッケージと新パッケージのバラツキに差はない」という仮説がまちがっているとは言えません。「証拠不十分で否定できない」といった感じですか。

まあ要するに「バラツキの差はない」が結論です。

したがってAコラーグルとXヨーゲンのt検定は、先ほど使った「t検定 等分散を仮定した（＝「バラツキに差がない」）2標本による検定」でOKということになります。

もし「バラツキに差がある」と結論になったのなら、エクセルのメニューで「t検定 分散が等しくないと仮定した2標本による検定」を選んで同じようにやればOKです。

### 認知力に差はあるのか？

もう1つだけ検定のパターンについて解説しましょう。ケースのパッケージ変更の例を使います。

内田さんは新パッケージの効果について、1日しか店舗で実験させてもらえないことになりました。

そこで店舗に新旧パッケージの両方を並べ、消費者がどちらを手にとって、見るかを調べました。結果は次のとおりでした。

従来パッケージを取った人	65	人
新パッケージを取った人	109	人
合計	174	人

(図表5-13)

これで「新パッケージの方が認知力が高く、手に取る確率が高い」と言いたいのですが、例によって「たまたまじゃないの。だって従来パッケージを取った人だって65人もいるんだから」と言われそうです。

こんな時は「適合度の検定」というものを使います。

この検定は計数値の検定に向いているものです。

計数値とは99ページでやった“度数”のようなものです。これに対して計量値とはその数値を測るものをいいます。

141ページの日販数は計数値であり、上の数字（65、109）は計量値（何人いるかという度数）です。

ここでは  $\chi^2$ （カイ2乗と読む）という検定統計量、 $\chi^2$  分布という確率分布を使います。そこでこのような検定を  $\chi^2$  検定ともいいます。

$$\chi^2 = \sum \frac{(\text{実測値} - \text{期待値})^2}{\text{期待値}} \quad (\Sigma \text{はすべて「足す」ということ})$$

で定義されます。実測値（観測値ともいう）は計数値そのもの、期待値（理論値ともいう）は仮説が正しければ「どういう値になるか」ということです。

## 統計学的に言って「認知力あり」

この例で $\chi^2$ 検定をやってみましょう。

まず帰無仮説は「従来パッケージ、新パッケージに認知力の差はない」です。

そうすると、消費者はこれを区別なく取り上げるので、期待値は全体の半分ずつとなります。つまり期待値は従来パッケージ、新パッケージとも $174 \div 2 = 87$ 人となります。

$\chi^2$ の確率計算もエクセルでできます。

エクセルに次のようなデータを入れます。

実測値	期待値
65	87
109	87

(図表5-14)

$\chi^2$ 検定では、関数にCHITEST（「関数の検索」で「カイ2乗検定」と入れると出てきます）を選びます。これをクリックし、「実測値範囲」に（65,109）の列を、「期待値範囲」に（87,87）の列を選ぶと、確率が計算されます。

0.000851となったでしょうか？つまり確率は0.09%となり、5%以下なので「認知力に差はない」という帰無仮説は棄却です。

だから「統計学的に言って、新パッケージの方が認知力が高い」といえます。

p社はレストランチェーンであり、p社のq店はビジネス街にあるイタリアンレストランである。

q店の主力商品はパスタランチであり、ベースとなるパスタにさまざまなトッピングを付ける形をとっている。

q店の顧客の男女構成比はほぼ半数ずつである。

q店店長の中野は思っていた。

「直感的には、女性客の客単価が高いような気がする。だったら女性向けの、ちょっと値段が高めのスペシャルセットメニューを考えたい。ただメニュー変更は本部決裁だ。相談してみよう」

本部の回答は次のとおりであった。

「男女別の客単価の状況を見て、考えたい」

中野は思った。

「男性客と女性客の単価なんて数字としてはない。ランチの時にレジで調べるしかないか。

しかし手間がかかるなあ。でもやるしかないか。

明日ランチに来た男性、女性の30人分の客単価を調べてみよう」

結果は次のとおりであった。

男性	950	970	910	940	950	930	1000	940	940	930	940	940	960	970	970
女性	1000	970	970	960	940	930	920	910	900	900	910	910	1000	930	930
男性	980	840	850	850	860	870	870	880	930	920	910	900	880	870	860
女性	940	900	910	1000	920	930	930	940	940	960	970	980	990	980	1020
	平均単価														
男性	917														
女性	946														

(図表5-15)

「客単価は平均で女性が946円、男性が917円か。やっぱり女性が高いな」

これを本部へ持っていくと、返ってきた答えは次のようなものだった。

「調べたサンプル数が少なすぎるんじゃないの。高々30人くらいのデータじゃ・・・」

さあ、あなたが中野さんならどうしますか。

もう大丈夫ですね。男性客単価と女性客単価の30ずつのデータについて「バラツキの差の検定」→「平均値の差の検定」と進めていきます。

まずはバラツキの差の検定、F検定からです。帰無仮説は「男女の客単価にバラツキの差はない」です。棄却率を5%としましょう。

次のような結果になりましたか？（まちがえた人はもう一度153ページを読みましよう）

F-検定：2 標本を使った分散の検定

	変数 1	変数 2	
平均	917	946.3333	
分散	1987.241	1203.333	
観測数	30	30	
自由度	29	29	チェック
観測された分散比	1.651447		1より大きい
P(F<=f) 片側	0.091383		9%
F 境界値 片側	1.860811		

(図表5-16)

9%（棄却率5%より大きい）となって、「バラツキに差はない」は棄却されません。したがって「バラツキに差はない」が結論です。

次は「平均値の差の検定」です。

「男性と女性の客単価に差はない」を帰無仮説として、「t検定 等分散を仮定した2標本による検定」を使います。

次のような結果となりましたか？

t-検定：等分散を仮定した2標本による検定

	変数 1	変数 2
平均	917	946.3333
分散	1987.241	1203.333
観測数	30	30
プールされた分散	1595.287	
仮説平均との差異	0	
自由度	58	
t	-2.84438	
P(T<=t) 片側	0.003069	
t 境界値 片側	1.671553	
P(T<=t) 両側	0.006138	
t 境界値 両側	2.001717	

(図表5-17)

0.6% (5%より小さい) となって棄却されます。つまり「男性と女性の客単価の

平均値は等しくない」となります。つまり「女性の客単価の方が高い」、これが結論です。

ちなみに先ほどのF検定が9%と少し小さいので、「分散が等しくないと仮定したt検定」でもやってみましょう。

t-検定：分散が等しくないと仮定した2標本による検定

	変数 1	変数 2
平均	917	946.3333
分散	1987.241	1203.333
観測数	30	30
仮説平均との差異	0	
自由度	55	
t	-2.84438	
P(T<=t) 片側	0.003119	
t 境界値 片側	1.673034	
P(T<=t) 両側	0.006237	
t 境界値 両側	2.004045	

(図表5-18)

これでも0.6%でやはり棄却です。

中野さんは本部にこう言いました。

「当店の男性と女性の客単価は、統計学的に言って有意差があり、女性のほうが客単価が高いと言えます。」

r社は海外から生鮮品を輸入する商社である。r社では、今回××ガニを継続的にs国から輸入する契約を結ぼうと考えていた。

××ガニは生態上はオス、メス同数であるが、消費者にはメスの方が圧倒的に人気がある。

r社としてはメスだけを輸入したいところだが、輸出元のs国t社ではオスとメスを一緒にないと販売できないという。

r社の××ガニ担当の渡辺は悩んでいた。

「オスカメスカは買ってみたいとわからない。t社は性を選んでないから、ほぼ同数のはずだと言っている。しかしオスばかりでは困るしなあ。

そうだ、私がs国へ行ってオスとメスをカウントしてみよう。

いくらなんでも、すべてカウントするわけにはいかないから、適当にサンプルガニを抜き出して、オスカメスカを調べてみよう。」

結果はオス70匹、メス56匹だった。

「やはりオスのほうが多い」とt社に言ったところ、「調査の仕方で、たまたまそうなったのでしょう」と言われた。

さああなたが渡辺さんならどうしますか？

大丈夫ですね？

$\chi^2$ 検定です。

期待値は「 $(70+56) \div 2 = 63$ 匹」です。棄却率は5%とします

実測値	期待値
70	63
56	63

(図表5-19)

関数はCHITESTです。

結果は21%となりましたか？

だからこういう結果になる確率は5%以上あることになります。したがって「オスとメスが同じ数の母集団から取り出しても、オス70匹、メス56匹になることもある」ということです。

だから今回輸入するカニの「オスとメスの数が等しい」というt社の言っていることは否定できないということです。

Exercise1のd社～g社の4社の例をまたまた使って、検定をやってみましょう。各社の発注リードタイムの平均値とバラツキに差があると考えていいかです。棄却率5%でやってみましょう。

まずはd社、e社のペアで検定してみましょう。

バラツキの差の検定、F検定からです。

25ページのエクセルのd社、e社のデータを、変数1、変数2に指定してやってみましょう。（標準偏差はd社>e社です。）

$P(F \leq f)$ 片側=26%となると思います。つまり「d社とe社のバラツキが同じ」という仮説は棄却できません。すなわち「バラツキは同じ」と考えます。

そのうえで「t検定 等分散を仮定した2標本による検定」をやってみると「 $P(T \leq t)$ 両側=83%」となって、やはり「平均値には差がない」という仮説が棄却できません。

したがってd社、e社の母集団のバラツキ、平均値に差はないと考えられます。

同様にe社とf社でやってみると、F検定では、「 $P(F \leq f)$ 片側=0.07%」となって棄却されます。（標準偏差はe社>f社ですので、それぞれ変数1、変数2に指定します。）つまり「バラツキに差はある」です。

そこで「t検定 分散が等しくないと仮定した2標本による検定」をやってみると、 $P(T \leq t)$ 両側=88%となって、「平均値に差はない」は棄却できません。つまり「平均値に差はない」が結論です。

ここまでの結果で「d社、e社、f社の平均値には差がなく、バラツキはd社、e社は差はなし、f社は2社よりも小さい」という検定結果になります。

次に安定のf社とスピードのg社で検定をしてみましょう。

F検定は確率が0.01%で棄却、つまり「バラツキに差があり」です。（標準偏差はg社>f社ですので、それぞれ変数1、変数2に指定します。）

次に「t検定 分散が等しくないと仮定した2標本による検定」をやってみると9%となり、棄却率5%では棄却できません。「f社、g社の平均値に差はない」（正確には「差がある」とは言い切れない）が結論です。

これらの検定の結論は、4社とも平均値に差はなく、バラツキはf社が小さいということになります。

まあ27ページの木下さんの判断は、検定の考え方を使っても誤っていないようです。

# カネの未来を考える

## 現在価値、DCFをマスターしよう

数字活用トレーニングの最後に、少し毛色の違うものをやりましょう。

今までの数字活用テクニックが「統計の世界」なら、今回は投資、ファイナンスという「カネの世界」で使われてきたテクニックです。

現在価値、**DCF**といったものですが、一般ビジネスの世界で近年脚光を浴びている数字活用テクニックです。

ビジネスにおいては、このテクニックもこれまでの統計分野と特に境のあるものではありません。これまで同様「エクセルを使った数字活用」であり、その主力テーマは「数字で未来を考える」です。

Lesson6は「数字活用」の「カネ」バージョンといったところです。

## 数字活用ケースその6

### —— どの案に投資するか ——

上田さんは3つの居酒屋タイプのうち、どれを出店するかを悩んでいます。

店舗を開いたら、それぞれどのような業績になるかは予測できているのですが、その数字を見ても、どのタイプにしたらよいかかわかりません。

あなたならどのタイプを選ぶかを考えてみましょう。そしてそのタイプを選んだ理由をまわりに説明して、納得してもらえそうかも考えてみましょう。

## どのタイプがカネを増やしてくれるか

J社は20年前に創業した居酒屋チェーンであり、過去さまざまなタイプの居酒屋を展開してきた。

J社では3年前の上場を機に、今後新規出店する店舗を3つの業態に分けて展開することとした。海鮮タイプ、焼肉タイプ、中華タイプの3つである。

J社ではこの各業態ごとにリニューアルまでの標準期間を決めており、それぞれ5年、7年、10年となっている。

現在J社ではK市の繁華街への出店を計画している。

J社業態開発部で今回の出店を担当する上田は、K市の出店を3業態のうちのだれにするか悩んでいた。

上田はマーケットリサーチ部へ、3業態ごとに「K市へ出店した場合の投資額とそのリターンの予測」を計算してもらうように依頼した。

しばらくして上田のもとに、エクセルで次のような資料が届いた。

業態	海鮮タイプ	焼肉タイプ	中華タイプ
期間	5年	7年	10年
初期投資額 (単位百万円)	120	160	200
年間に期待される 営業キャッシュ フロー (単位百万円)	1年目	25	25
	2年目	30	30
	3年目	30	30
	4年目	35	35
	5年目	35	35
	6年目		35
	7年目		35
	8年目		
	9年目		
	10年目		

(図表6-1)

## 足し算はできない??

上田はこの表の“営業キャッシュフロー”の意味がわからないので、マーケットリサーチ部に問合せした。

「キャッシュフローとは1年あたりの現金増加量です。このうち営業キャッシュフローとは事業で増やす現金の量です。その居酒屋店の1年間の商売によって、当社にもたらされるカネの量と考えてください。」とのことだった。

上田「しかしどうやって出店する業態を決めたらいいんだ。海鮮タイプは120百万円投資して、5年間で25百万円、30百万円、30百万円、35百万円、35百万円だけ現金が増えるから、トータル155百万円か。とすると、差し引き35百万円増えるわけだな。

焼肉タイプは  $(25+30+30+35+35+35+35) - 160 = 65$ 百万円、中華タイプは  $(25+25+30+30+30+30+30+30+30+30) - 200 = 90$ 百万となる。

これなら90百万円の中華タイプかな？」

上田は、最近金融機関から転職してきた業態開発部長へ、自らの意見を伝えた。部長は次のように答えた。

「何をやってんだ。  
海鮮タイプの営業キャッシュフローでいえば1年目の25百万と2年目の30百万を足し算なんかできないだろう。  
君は何年業態開発に携わってきたんだ。DCFも知らんのか。」

## 未来のカネは割り引く

あなたが「『今日の100万円』と『1年後の100万円』のどちらを取ってもいいですよ」と言われたら、もちろん「今日の100万円」を取るでしょう。

では「今日の100万円」と「1年後の200万円」では？

多く人は「1年後の200万円」を取るでしょう。

それでは1年後の150万円なら、140万円なら・・・とやっていくと、どんな人にも（もちろん人によって額はちがいますが）「何ともいえない」というラインがあると思います。

これが110万円の時、その人にとって「今日の100万円＝1年後の110万円」ということを意味しています。つまりその人から見ると「1年後の110万円」は、今日の価値に換算すると「100万円」ということです。

この「100万円」を「1年後の110万円の現在価値」といい、 $110 / 100 = 1.1$ の小数点以下「0.1」（＝10%）を、この時の割引率（ディスカウントレート）といいます。

1年後の110万円という未来のカネを、現在価値にするために10%割引くという意味です。

割引くカネがケースのような「将来のキャッシュフロー」の時、これをディスカウントキャッシュフロー、略してDCFといいます。

DCFでは2つの前提をとります。

1つはキャッシュフローの額によって、一度決めた割引率を変えないことです。先ほどの例で、割引率を10%と決めたら、1年後の220万円は「 $220 / 1.1 = 200$ 万円」で、現在価値は200万円と考えます。

もう1つは年によって割引率を変えないことです。つまり2年後の110万円は、10%で2回割引いて、 $110 / (1.1)^2 \doteq 91$ 万円となります。

### 今日の価値にしてから決めよう

部長はこれを指摘したのです。つまり海鮮タイプの1年目（出店してから1年後）の営業キャッシュフロー25百万円と2年目（出店してから2年後）の30百万円は、「そのままでは足せない」ということです。「将来のカネは割り引いて考える必要がある。だからすべてのカネを現在価値にしてから考えなさい」と金融機関でやってきた経験から言っているのです。

DCFの考え方を使得、先ほどのケースを考えてみましょう。

まずは割引率を決めなくてはなりません。

どうやって決めるかはケースバイケースです。J社で「店舗投資については割引率を5%とする」と統一しておいもよいですし、この出店のためのカネを銀行から借り入れるなら、その利率を使ってもOKです。

ここではもっとも一般的な「5%」と設定しましょう。

海鮮タイプの営業キャッシュフローの現在価値は

$$\frac{25}{1.05} + \frac{30}{(1.05)^2} + \frac{30}{(1.05)^3} + \frac{35}{(1.05)^4} + \frac{35}{(1.05)^5} = 133 \text{ (百万円)}$$

となります。

海鮮タイプでは現在価値で（投資は0年目です）120百万円の投資をしますので、 $133 - 120 = 13$ 百万のカネが増加することになります。

つまり海鮮タイプは、この投資によって、現在のカネの価値に換算して、13百万円のカネを増やすことができます。この増加したカネをネットキャッシュフローといいます。

この計算もエクセルでできます。焼肉タイプでやってみましょう。171ページの図表6-1を開いて、関数検索で「現在価値」と入れ、NPVという関数を選びます。

ここで「割引率」に「0.05」、「値1」に焼肉タイプの「25、30、30、35、35、35、35」という列」を指定します。OKを押すと「184.1・・・」 $\div$ 「184」と計算されます。

焼肉タイプの営業キャッシュフローの現在価値の和は184百万円となります。現在価値で160万円を投資するので、24百万円の増加です。

一方中華タイプは222百万円となり、22百万円の増加となります。

したがって次のような結果となります。

業態	海鮮タイプ	焼肉タイプ	中華タイプ
期間	5年	7年	10年
初期投資額	120百万円	160百万円	200百万円
ネットキャッシュフロー(DCFベース)	13百万円	24百万円	22百万円

(図表6-2)

上田さんの判断は難しい所です。

中華タイプは焼肉タイプに明らかに負けています。焼肉タイプは7年で24百万円カネを増やし、中華タイプは10年で22百万円しか増やせません。

しかし海鮮タイプか焼肉タイプは微妙です。海鮮タイプの「120百万円投資して5年間で13百万円のカネを増加させる」を取るか、焼肉タイプの「160百万円投資して7年間で24百万円増加させる」を取るかです。

上田さんがケースで述べていた中華タイプは“問題外”といえます。

### 内部利益率ですっきり決めよう

先ほどの「決め方」を正味現在価値法といいますますが、もう1つ内部利益率法というものがあります。こちらの方がすっきりと決定できます。

海鮮タイプでは、割引率5%でネットキャッシュフローは13百万円となりました。この割引率を6%にすると、営業キャッシュフローの増加額が129百万円となり、ネットキャッシュフローは9百万円と下がります。

こうやって割引率を上げていくと、このネットキャッシュフローがちょうど0になる割引率があるはずです。これを内部利益率といいます。

投資額が少なく、営業キャッシュフローが多いと、この割引率（＝内部利益率）は高くなります。つまり内部利益率は「その投資の効率性」を表わしたものといえます。

内部利益率もエクセルで計算できます。  
まず次のようにエクセルをセットします。

海鮮タイプ	焼肉タイプ	中華タイプ
-120	-160	-200
25	25	25
30	30	25
30	30	30
35	35	30
35	35	30
	35	30
	35	30
		30
		30
		30

← 0年目の投資額は  
カネが減るのでマイナス

(図表6-3)

「関数の検索」で「内部利益率」と入れて、**IRR**という関数を選びます。  
 海鮮タイプであれば、「範囲」に「-120から35までの列」を指定します。  
 そうすると8.7%と計算されます。同様に焼肉タイプは8.8%、中華タイプは7.2%  
 と計算されます。

内部利益率は「投資額を加味した1年あたりのカネの増加率」という投資効率の  
 指標と考えられます。

この内部利益率法を使うと、微妙ですが海鮮タイプの8.7%をおさえ、ギリギリ  
 8.8%の焼肉タイプが選定されることになります。

内部利益率法は投資案が1つしかなくても使うことができます。

ケースのJ社で、「海鮮タイプの居酒屋を出すかどうか」を考えているとします。

この時一定の率（例えば6%）をあらかじめ決めおいて、「これを超える内部利益率であれば出店する」と決めます。この6%をハードルレートといいます。

この時海鮮タイプは8.7%の内部利益率ですので「GO」という判定になります。

ソフトハウスt社は同業のu社を買収すべく検討を進めていた。つまりu社の株をt社が買って、u社の親会社になることである。

t社は創業者が未だ大株主となっているが、ベンチャー企業向けの証券市場に上場している。

t社はu社の創業者および経営陣の買収合意を得て、現在は「u社をいくらかで買収するか」、つまり「t社の株を1株いくらかで買うか」を検討している。

t社の証券市場での株価が1つの目安とも考えられるが、t社の株価はここ数年大きく乱高下している。そのうえ昨年度は赤字となった大型プロジェクトがあり、最終利益も赤字となってしまう、株価を大きく下げてしまっている。

大株主のt社創業者は、「この最悪の株価で売却するつもりはない。」と言っている。

そこでt社とu社ではDCF法をベースとして考えることにした。具体的には次のような方法である。

- ・u社の将来のディスカウントキャッシュフローをベースとして、買収企業価値を算定する。
- ・割引率は「加重平均した資本コスト\*1」をベースにする
- ・対象期間は5年として、5年間の事業によるキャッシュフローを対象とする。さらに5年後に資産（財産）をすべて売却し、すべての負債（借金）を返済すると考える。すなわち純資産\*2（=資産－負債）を5年目のキャッシュフローに加算する

- 。資産はのれん\*3を含めてすべて時価評価\*4する
- ・上記の企業価値を発行済株数（50万株）で割り、買収予定株価とする

t社M&A準備プロジェクトでは、急ピッチでu社の評価が進められていた。

このプロジェクトで出てきた評価結果は以下のとおりであった。

- ・加重平均資本コスト（割引率）は8.5%。
- ・u社のこの先5年間の事業によるキャッシュフローは1年目1,000百万円、2年目1,000百万円、3年目1,200百万円、4年目1,200百万円、5年目1,500百万円と予測。
- ・5年後の純資産総額は1,400百万円。

さあ1株あたりいくらで買収することで、t社とu社は合意したらよいかを計算してみましよう。

\* 1.加重平均した資本コスト 資本コストとは会社のカネを集めるのにかかったコスト。借入金なら金利。加重平均とはその割合を加味して平均を取ること。

\* 2.純資産 資産（財産）から負債（借金）を引いた金額。u社を買収後、これを解散すれば、この純資産分がu社の株主であるt社のものとなる

\* 3.のれん 目に見えない財産のこと。ブランド、ノウハウ、情報・・・

\* 4.時価評価 その財産を売った場合の価格を考えること

まずは次のようなエクセルを作りましょう。

	1年目	2年目	3年目	4年目	5年目	
事業によるキャッシュフロー	1000	1000	1200	1200	1500	(A)
純資産売却によるキャッシュフロー					1400	(B)
ネットキャッシュフロー	1000	1000	1200	1200	2900	(A+B)

企業価値  (C)

株数  (D)

株価  (C/D)

(図表6-4)

そのうえでネットキャッシュフロー (A+B) を割引率8.5%で割引きます。

カーソルを上企業の企業価値のセル (C) に合わせて、関数にNPVを選びます。

割引率は0.085 (8.5%)、データの範囲は「ネットキャッシュフローの1年目～5年目の行」です。

これで上のように (Cのセルに表示されている) 5,505 (“¥” はデータが円、つまりカネであることを表わしている) と出ます。この5,505百万円 = 55億5百万円がu社の企業価値、つまり買収予定価格です。

これを0.5百万株（50万株、単位は百万なのでこれに揃える）で割ります。その結果の11,010円（C/D）が買収予定の「1株あたりの価格」となります。

t社がu社と交渉すべき点は、企業価値や株価といった数字の結果ではなく、事業によるキャッシュフロー、5年後の純資産額、割引率といった「その数字を出すプロセス」についてです。このプロセスに合意すれば、自ずと買収価格（企業価値、株価）は決まります。

このパラダイムは30ページの目標販売台数と同じです。「片方が高くしたい、片方は低くしたい」という交渉シーンでは、力づくではうまく行きません。互いに合意するには、人類が納得した“やり方”を使うことが求められます。

（このExerciseに書いてある「カネに関する言葉」の意味がよくわからなかった人は、拙著「ビジネスマンのナレッジ（基本編）」（内山力著 同友館）を読んでください。）

あなたも自分の会社やよく知っている上場会社の株価を計算してみましょう。計算するのに必要な情報はその会社のWebサイトを見ればほとんどが手に入るはずです。

Webサイトを見ればほとんどが手に入るはずですよ。

キャッシュフローや純資産額は有価証券報告書（上場企業であればインターネットでダウンロードできる）にあるキャッシュフロー計算書や経営計画書などから予測します。発行済株数は有価証券報告書に載っています。

割引率は期待収益率（あなたが1年間で何%くらいカネを増やしてほしいかという利率）などを使ってもOKですし、よくわからなかったら5%くらいが妥当でしょう。

こうやって自分で株価を出して、現在の証券市場での株価と比較してみましょう。

あなたの計算した株価 > 市場の株価なら「買い」です。逆なら「売り」です。この結果（「この会社、“買い”だね」）をさりげなくまわりの人に言えば、それだけであなたは「カネに強い」「数字に強い」ビジネスマンと見られます。

「やるね。いつの間にそんなに数字に強くなったの？ どうやって勉強したの？」と聞かれたら、そっと本書を出しましょう。

これで本書の数字活用トレーニングは終わりです。どうですか？数字活用の自信がつけましたか？数字アレルギーだった方は完治しましたか？

最後に本書で伝えてきた“数字活用のコツ”をまとめておきましょう。

- ・数字を見た時の**直感**を大切にす。どうやって数字を使うかを考える前に、その数字を肌で感じる。
- ・その直感がどうやったらまわりの人へ伝わり、**納得**してもらえるのかを考える。
- ・数字活用のやり方は自分で考えない。**先人の知恵**を使う。
- ・やり方は使う前に**理解**する。理解できないやり方は絶対に使わない。
- ・計算結果をプッシュしない。**計算プロセス**を説明し、合意を得る。
- ・数字活用の最大の目的は**未来**を考えること。過去を分析するものではない。
- ・未来は誰にもわからない。未来を当てるのではなく、「過去の数字を使えば**未来はこうなると考えるしかない**」という数字を作る。

数字活用はスポーツのようなものです。やり方、コツを学び、それを使って自分でやってみることで、どんどん上達していきます。

あなたも本書を読破したら、仕事で数字を使うことに自信を持って、積極的にチャレンジしましょう。そうすれば苦手だった数字活用がいつの間にか快感に変わり、まわりを説得できないというストレスがきれいに解消できると思います。