「回帰分析」で因果関係を解き明かす

（氏名）青 野　洋 己
（所属）住友建機製造㈱
　　兼 住友重機械工業㈱
（Tel.）Tel.043-420-1792
（部門）経営工学部門

１．「回帰」とは....

「怪奇現象」の“怪奇”ではない、「回帰」とは国語辞書によれば“ひと回りして元に戻ること”とある。　ですから....

・天体が公転などによりもとの位置に戻りその運動を繰り返すこと
・動物が特定の繁殖場所から遠く離れてもそれらの位置を知って戻ってくること
　（サケや海亀が有名）

などのことを“回帰”と称することはご承知の通りです。

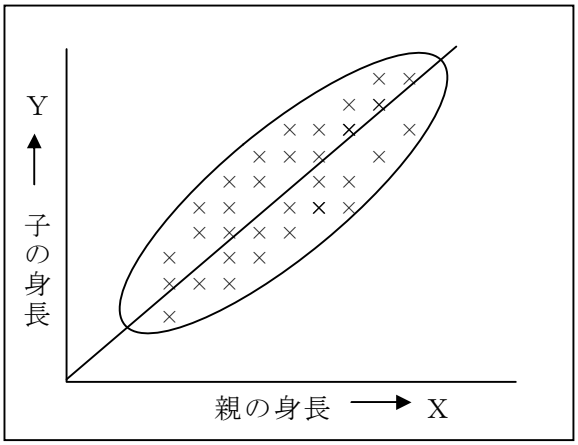
一方、統計学上の“回帰”とは次のように定義されている。

『原因となる数値と結果となる数値の関連性を統計的手法を用いて調べる方法であって、原因となる数値を説明変数、結果となる数値を目的変数と言う。説明変数が１つの場合を単回帰分析、２つ以上の場合を重回帰分析と言う。』

とあります。しかし、なぜこれが“回帰”分析なのでしょうか？

２．因と果

因果関係を考える例えとして、親の身長（因）はその子供の身長（果）に影響を与え、その関係は図―１のとおりであったとします。



図－１

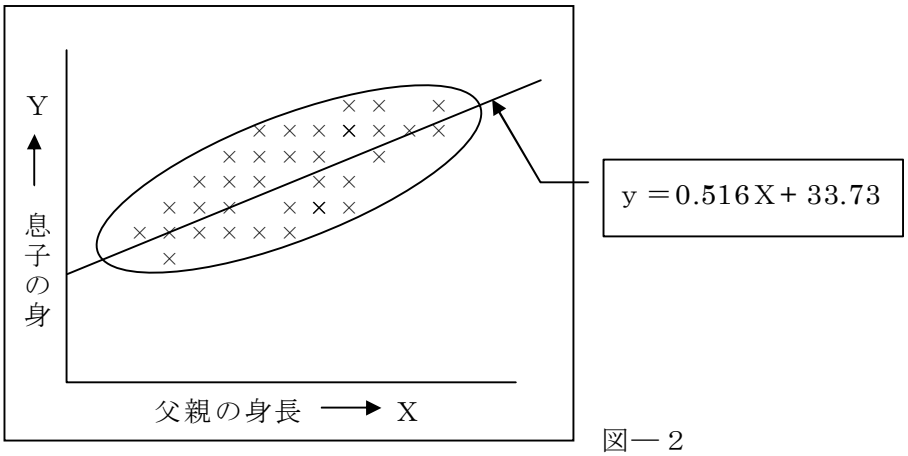
すなわち、身長の高い親の子は、親に似て身長が高いはずである。しかしまったく同じということはなく、子の身長は親の身長の上下にバラつくと仮定しましょう。この仮定はしごくもっともらしく、だれにも納得できそうである。

ところがこの仮定を認めると大変なことがおきる。身長の高い親からは、さらに身長の高い子が生まれる可能性があるし、その子が親となってさらに身長の高い子が生まれるかもしれない。その反対の身長の低いほうでも同様である。人類が誕生して数百万年における、膨大な世代交替によるバラツキの増大を想定すると、この仮定はどうもまずいということに気づくでしょう。すなわち発散してしまうからです。

３．先祖返り

この仮説を検証するためある膨大な調査が行われました。

英国の遺伝学者 Francis Galton（1822-1911）は、父親の身長が成人した息子の身長にどのよに遺伝するかを調べる研究で英国のＮ=1078 家族について調査し、散布図を作成した。（図―２）



図―２

その結果、父親の身長χとその息子の身長ｙとの間に、正の相関があり、父親の身長を１インチ毎の級に区切り、各級に入る点のｙについての平均値を求めプロットすると、ほぼ一直線になっているので、この直線の式を求めたところ次の様であった。

$$y = 0.516\chi + 33.73 \qquad \cdots\cdots \text{式（1）}$$

この式は、父親の身長χが１インチ増すと、息子の身長ｙは、０．５１６インチ増すことを示している。

この係数が０．５１６で１より小さいことと接辺 ３３．７３を有していることに

ゴールトンは注目し、背丈の高い父親の息子は、その父親ほどには高くなく、背丈の低い父親の息子はその父親ほどには低くない、すなわち父親の身長が集団全体の平均から大きく変異すると、息子の身長の変異は小さくなり集団全体の平均に近づくとし、これを“先祖がえり”の現象と考えた。

これを平均への「回帰（regression)」現象と呼びました。

４．回帰モデル

目的変数をｙ、説明変数をＸとし、相互の線形関係を仮定すると次のようなモデル式が予測される。これを「単回帰モデル」という。

$$y = \beta_0 + \beta_1 X + \varepsilon \qquad \cdots\cdots \text{式（２）}$$

ここで、β_0は切片（回帰定数)、β_1は回帰係数と言い、εはこの回帰式では説明がつかない誤差である。

回帰分析の中身は、この誤差εを最小にするようなβ_0・β_1を探索する作業（最小二乗法）そのものなのである。

そして、このεのバラツキがｙの変動に対して十分小さければ回帰式は有意であると解釈する。

また、説明変数Ｘの数が複数になった拡張モデルを「重回帰モデル」という。

５．回帰式の解釈

β_0はＸの値が“０”の時も存在するｙの値で、そのモデルの基礎的エネルギーとでもいうもので、全体に平均をシフトするにはこの値を嵩上げできれば良いわけです。人間で言うと基礎体力・基礎代謝量のような感じです。

ところで、式（１）で示した、ゴールトンモデルにおけるβ_0すなわち“３３．７３インチ”は何を意味するのでしょうか？

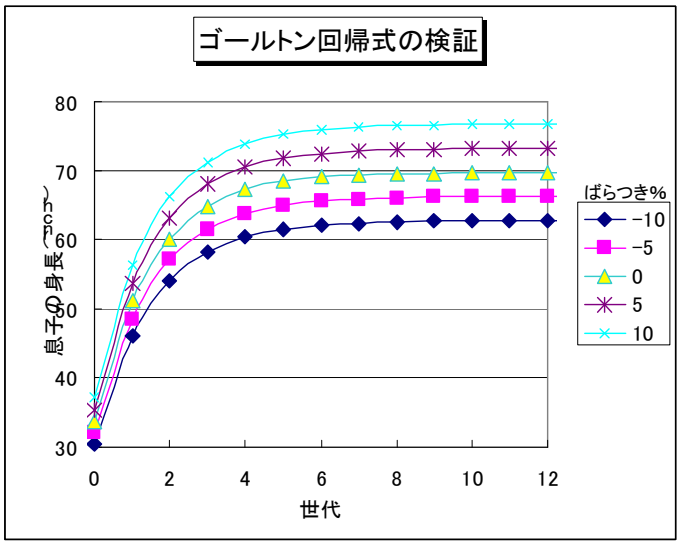
もしかすると人類の進化の過程の原点....すなわち猿人から人類に突然変異した時の初期の背丈なのかもしれません。(33.73×2.54cm/inc＝85.7cm...小さいなぁ)

またβ_1の回帰線の傾きの値“０．５１６”は何を意味するのでしょうか？

これはいわば進化の過程の成長速度を表しているかもしれません。そうだとするとこの値は大きいと言えるのかはたまた小さいのか、またこのモデルは収束するのだとすると人類はどの程度の進化（背の高さ）を遂げるのであろうか。

そこで、人類の原点をβ_0の切片３３．７３inchに置き、次世代での身長のばらつきを±１０％と仮定（トンビがカラスを産むということもある）し、２０世代の間にどの程度の身長の変化が起こるのかシミュレーションを行ってみた。

その結果のグラフを図―３に示す。



図－３

これより、世代は１２世代（ざっと６００年）を越えると概ね収束し、最終的には、高い人で７７inch（１９６cm）低い人でも６３inch（１６０cm）になることが分かる。

この仮定には、増分減分のランダムな“振れ“を考えておらず、一方向に進化の過程が進行した場合を想定しているのであって、現実にはありえない。すなわちこの進化の速度はもっとゆっくり進行するはずである。

増減方向の振れは、増分側と減分側にそれぞれ５０％の確率で現れ、また振れ幅の［Ｍａｘ／Ｍｉｎ］の出現確率を３σの片側確率すなわち１０００回に１．３回で考えると収束年数は次のようになり、概ね９２万年かかると試算される。

収束年数＝６００年×（１／０．５）×（１０００／１．３）＝９２３，０７０年

また、このモデルは地球環境変化や科学の進歩に伴う生活習慣の変化などが影響し、回帰モデルは時々刻々と変化していると考える方が妥当であろう。

いずれにしても、この単純なモデル式の中に潜むひとつひとつの数値・符号の本質的な意味を考えると、ただの一次関数としての数学的意味合いを越え、実に興味ある示唆を与えてくれる。

回帰分析は、事物の因果関係を解き明かす恐るべきツールなのである。

以上