

連載 SQC手法の活用ポイント

吉富 公彦 著

第6回 相関と回帰分析

問題解決に役立つSQC手法(統計的品質管理の手法)の解説の第6回目です。

今回は、2つの特性値(長さ、重さ、など)に着目したときに、特性値間に関係性がありそうかどうかをみる相関分析と、関係性があるとするとどのような関係式で表せるかを導く回帰分析について解説します。

1. 相関分析

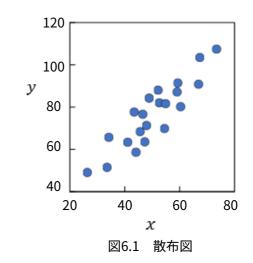
一般的に身長が高いほど体重は重く、身長が低いほど体重が軽い傾向があります。身長と体重のように2つの特性値間に関係性があるかどうかを知るためには相関分析を行います。

まずは、図6.1に示すように散布図を描きます。散布図を描く際に、どちらかの特性値が原因と思われる場合にはそれを横軸に、他方の特性値が結果と思われる場合にはそれを縦軸にとります。ここでは便宜的に横軸をx、縦軸をyとしておきます。

打点の並び方の傾向に、図6.1のような右肩上がりの傾向があればxとyの間には正の相関がある、あるいは逆に右肩下がりの傾向があれば、xとyの間には負の相関があると判断します。

散布図だけでも相関の有無を判断できますが、 相関の強さを客観的に知るために相関係数rを求めます。相関係数rは次の式で求めることができます。

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}}$$



相関係数は右肩下がりの直線上に打点がすべてある場合に r=-1、右肩上がりの直線上に打点がすべてある場合に r=+1 となります。また、打点が同心円状に広がり完全に無相関の場合は r=0 となります。因みに図6.1 の場合、相関係数は r=+0.90 となり、強い相関であることを客観的に確認できます。

なお、上式による相関係数の値+0.90は、得られたデータから計算されたものですから、真の値、つまり母相関係数であるとは限りません。第2回で紹介した推定を行っておきたいところです。正確な推定方法はないのですが、相関係数r を $z = \tanh^{-1}(r)$ という数値変換(専門的にはz 変換といいます)を行うと、近似的に正規分布とみなせることが分かっています。このことを利用してz 変換した状態で信頼限界を求め、それを逆変換($\tanh(z)$)することで母相関係数の信頼限界を見積もることができます。図6.1の場合の計算結果だけを示すと、信頼率95%での信頼下限は+0.77、信頼上限は+0.96となります。本メールマガジンをご覧の方で、推定を行わずに相関係数の点推定値(図6.1の場合は、+0.90)のみを信用し過ぎて失敗された方はおられないでしょうか。

2. 回帰分析(単回帰分析)

相関分析では、関係性の有無やその強さを知ることができました。

実務では、横軸のx が変わっていくときに、縦軸のyの値がどのようになるかを予想したくなる場面が多々あります。このようなときに適用する手法が回帰分析です。取り扱う変数がx とy の2つだけの場合は単回帰分析といいます。

単回帰分析でも相関の有無を客観的に知るために分散分析をまず行います。図6.1の データで分散分析を行うと表6.1となります。

要因	偏差平方和	自由度	分散	分散比(F比)	検定結果
回帰	3867.835	1	3867.84	81.26	1%有意
残差	904.356	19	47.598		
計	4772.191	20			

表6.1 図6.1のデータの分散分析表

検定結果が有意となっていますので、客観的に 相関ありという結論が得られます。

また、計算方法は割愛しますが、最小二乗法を 利用して次の関係式が得られます。

$$y = 1.184 + 17.086 x$$

この関係式を図示すると、図6.2(データは図6.1と同じ)の打点の中心を通る直線(破線)になります。この直線のことを回帰直線といいます。

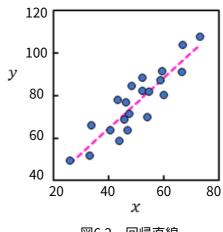


図6.2 回帰直線

ですが x と y の関係を示すこの回帰直線も得られたデータから計算して求めたものなので、真の直線 (母回帰) であるとは限りません。むしろ、この通りにならないことの方が多いことを読者の方々もご経験されているかと思います。

従って、回帰直線についても推定しておいた方 が、予測ミスを防げます。

計算過程は割愛しますが、回帰直線の信頼限界を図示すると、図6.3における回帰直線(破線)のそばにある曲線(実線)になります。

また、図6.3において一番上側の緩い曲線と一番下側の緩い曲線(共に実線)は、個々の値の最大値および最小値(信頼率95%)の見積り結果になります。

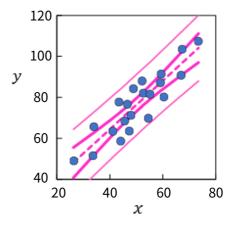


図6.3 信頼限界と予測限界 (信頼率95%)

以上に示したように、回帰分析を行うことで母回帰を推定(真の回帰直線の見積り)できますし、データの予測(個々のデータの最大値および最小値の見積り)もできます。

第3回で紹介した分散分析同様、回帰分析を積極的に業務に取り入れていただき、ビジネスリスクを少しでも軽減していただければと思います。

(以下余白)



著者紹介

吉富 公彦(よしとみ きみひこ)

1986年新日本無線㈱入社。赤外発光ダイオード、チューナーモジュールの生産技術業務に15年間従事。その後、社内SQC手法教育および品質マネジメントシステム業務に従事(ISO9001およびIATF16949の内部監査員)。元VDA6.3 Process Auditor。

現在、東京理科大学、東京情報大学非常勤講師。

(一財)日本科学技術連盟において企業向け講師派遣型研修SQCベーシックコース講師の他、品質管理セミナーベーシックコースおよび品質管理セミナー入門コースで講師をつとめる。(一財)日本規格協会 通信講座品質管理中級コース教材作成検討委員。