



## 慶應義塾大学ビジネス・スクール

# 統計学 ノート (4)

### 1 相関と回帰

#### 1.1 線形相関

10

今まで 1 つの変数とその分布に関する統計的方法を考えてきたが、統計的な問題では複数の変数に関連するものが多い。2 つまたはそれ以上の変数に関するデータ処理の 2 つの手法を説明する。ここでは、主に 2 変数について考えるが、この方法は 3 変数以上の多変数の処理にも拡張できる。

ある問題では、いくつかの変数の相互の関係を知るために、これらの変数を同時に調べることがある。また別の問題では、関心があるのは特定の 1 変数だけであるが、それをより明らかにするために他のいくつかの変数を同時に調べることもある。これらの 2 種類の問題は一般にそれぞれ相関および回帰とよばれる。まず、相関について述べる。

相関の問題は、関心のある 1 組の変数の間に何か関連があるかどうかを知りたいときに起こる。2 つの変数  $x$  と  $y$  の間の関連を調べるには、まずデータを  $xy$  平面上の点として 20 図示し、関連のだいたいの形を発見することからはじめるべきである。このときのグラフは散布図とよばれる。このグラフを見れば、2 つの変数の間に何かきわだった関連があるかどうか、もしあればその関連は近似的に直線的関係とみなしうるかどうかを簡単に見わけができる。

点の散らばりの一般的な傾向の特徴をつかむには、散布図を見て、点全体が多少のうね 25 りをもつあるなめらかな曲線の両側に散らばる傾向があるかどうか、あるいは、これらの点が 1 本の直線の両側に散らばっているように見えるかどうか、を見極めるべきである。どちらにも見えるような場合には、ある意味で 2 つの変数が直線的に関係している度合を測定することが望ましい。このような尺度をつくり出すために、それがもつべき望ましい性質を考える。

30

関連の尺度はそれぞれの変数の原点の選び方に無関係でなければならない。この性質を実現するには、変数をそのまま用いるより、それぞれの変数の平均からの偏差を用いたほうがよい。望ましい関連の尺度をつくるには変数  $x_i, y_i$  の代わりに  $x_i - \bar{x}, y_i - \bar{y}$  を使うべきである。ここで記号  $x_i, y_i$  は第  $i$  番目の 1 組の数字を表す。

関連の尺度は、 $x, y$  に対して用いた測定の尺度にも無関係でなければならない。この 35