

## 2015 年度秋学期 統計学—数学の補足説明 第 2 回

### 平均と分散に関連して (2)

#### Σ 記号

「算術平均」の節で、平均を求める式は次のように書かれています。

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

後半にある「Σ」(Σ (シグマ) 記号) を使った式は、数学で「合計」を表す書き方です。「Σ」の下には小さく「 $i=1$ 」、上には「 $n$ 」と書いてありますが、これで

- 「シグマ・ $i$  が 1 から  $n$  までの  $x_i$ 」
- 「 $i$  が 1 から  $n$  まで  $x_i$  を合計する」

などと読み、「添字  $i$  を 1 から順番に  $n$  までひとつずつ増やして、添字のついた数を順に足していく」ことを表します。添字はどこにあるのかといえば、後ろの  $x_i$  についています。ですから、 $x_1, x_2, \dots$  と順に添字を増やして行って、 $x_1 + x_2 + \dots$  と順に足していき、 $x_n$  まで合計することになります。つまりこれは、 $x_1 + x_2 + \cdots + x_n$  の意味になります。

わざわざ Σ 記号を使うのは、合計を計算した結果を、 $\sum_{i=1}^n x_i$  とひとかたまりに書いて、あたかも 1 つの数のように扱うためです。この使い方は、共分散と相関係数の説明 (第 6 回) に出てきます。また、ここでの  $x_i$  のように、「1 から順に数える」ための添字には  $i$  がよく用いられますが、この  $i$  は “index” から来ています。 $i$  だけでは足りない場合は、その続きの  $j, k, l, \dots$  もよく用いられます。

また、

$$E(X) = \sum_x x f(x) \quad (2)$$

というような Σ 記号の使い方もあります。ここでは、「Σ」の下に小さく  $x$  と書いてあるだけで、「 $\sum_{i=1}^n$ 」のように「 $i$  が何から何まで変化するか」とは書いてありません。この場合は、「 $x$  を可能なすべての数に変化させて、各々の場合で  $x f(x)$  を計算して、それらをすべて合計する」ことを表します。

#### 引数

数学では、 $f(x)$  や  $E(X)$  のように、文字の後ろにカッコがついた書き方がよく出てきます。このカッコの中の文字や数字は**引数 (ひきすう)** といいます。「モーメント」の節で、「階級値が  $x$  である階級の相対度数を、 $f(x)$  で表す」と書いてありますが、この場合、 $f$  だけでは「相対度数」という概念を表しているだけで、どの階級の相対度数なのかはまだ決まっていません。引数をつけてはじめて、「ある特定の階級の相対度数」となります。例えば、階級値が具体的な数値、例えば「35」であれば、その相対度数は  $f(35)$  となります。

## 変数と定数

数学では、数字のかわりに  $X$  や  $x$  という文字を使って、式を表します。文字の部分は、いろいろな数字に置き換わることになります。ただし、この「文字が数字に置き換わる」ことには2通りの意味があります。文字による数式を読む時には、各文字がどちらの意味の文字であるかを把握しておく、式の意味が理解しやすくなります。

そのひとつは、「式では文字で書いてあるが、実際のデータと問題が与えられた時は、その文字がどんな数字に置き換わるかが最初に決まっていて、問題を解いているあいだは変わらない」ものです。もうひとつは、「式では文字で書いてあって、実際の問題の中でもいろいろな数に変化する」ものです。前者は**定数**、後者は**変数**とよばれます。

たとえば、平均を求める

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

という式で、実際にデータを使って平均を求めるときには、データの個数  $n$  やデータ  $x_1, x_2, \dots$  は、データを集めた時点で決まっているはずで、一方、 $i$  は計算の途中で  $1, 2, 3, \dots$  と変化していきます。ですから、 $n$  や  $x_1, x_2, \dots$  は定数で、 $i$  は変数です。

また、分散を求める

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (4)$$

では、データの平均  $\bar{x}$  は、データを集めたあと、分散を計算する前に平均を計算して、すでにひとつの数字に決まっています。ですから、 $\bar{x}$  も定数です。

一方、標準得点や偏差値の説明（講義第5回）では、

分布そのものを代表して  $X$  であらわして、 $a, b$  を定数とするとき「分布  $X$  を、 $Z = aX + b$  という1次式で分布  $Z$  に変換する」という言い方を

と書いています。分布はいろいろな異なる数値の集まりで、データを集めて問題を解く段階になってもそれは変わりません。このような「さまざまな数字の集まり」をひとつの文字  $X$  や  $Z$  で表しているの、この  $X$  や  $Z$  も変数です。

講義の後半では、このような「分布を代表する変数」の考え方を利用して、「偶然によっていろいろな数字になる変数」というものを説明します（「確率変数」といいます）。