基礎から学ぶシリーズ２

基礎からの統計学

福井正康

まえがき

経営学・経済学を専攻する学生が統計学を学ぶとき、どんなことに重点を置けばよりでしょうか。世の中には多くの統計ソフトがあり、誰でも統計学の知識があれば利用可能です。しかし、統計ソフトは分析の適用限界を超えた場合でも答えを出してくれますし、その中で何が行われているのか、ほとんどブラックボックスです。従って、統計ソフトに完全に頼ってしまっても、何をやっているのか直感的に理解することが難しいし、公式の導出に重点を置き過ぎても数学を得意としない学生からは敬遠されてしまいます。

そこで私の勤務する大学の講義用に、統計の初学者に必要な項目を選んで、平易だけれど実用にも耐えうる教科書を書いてみることにしました。基本的に公式の導出は必要最小限にとどめ、公式を利用して自分で計算ができるように、Microsoft Excel を使っての計算練習に力を入れています。数学の苦手な学生は少し数式が出ただけで最初から敬遠してしまいがちですが、この拒絶反応を少しずつ解消して行くことも本書のねらいです。

本学では、１年次に「基礎数学」という授業があり、２年次に「経営統計学」の授業があります。本書を読む前に、このシリーズの「基礎からの数学」の基本部分に目を通しておいていただければより理解が深まると思います。

統計を学ぶためにはまず確率の概念を理解しておかなければなりません。それゆえ最初は簡単な場合の数の求め方から確率の計算法を学びます。次に、集めたデータの集計方法について学習し、正規分布やそれから派生する分布の基本的性質を学びます。その後、検定と推定について詳しく学び、最後に相関や回帰分析について学習します。基本的にはデータが正規分布に従う場合について議論しますが、検定には分布形に依存しないノンパラメトリックなものも加えます。

本書の作成に当っては、一部の学生にモニターの協力をお願いしました。学生の目で分かり易さを評価してもらうことは価値のあることだと考えます。協力していただいた学生諸君には深く感謝致します。学生の学力向上に本書が少しでも役に立てば、これほどの歓びはありません。

最後になりますが、私に統計の基礎を教えて下さった元福井医科大学環境保健学教室教授の緒方昭先生に感謝致します。

福山平成大学経営情報学科
福井正康

１章 場合の数

1.1 順列

ある事象は平均して 100 回に 1 回程度起こるとか、10 回に 1 回程度だという頻度が分かっていれば、誤った判断をして大失敗をする可能性が少なくなります。そのためにはある状況で発生しうる事象が全体でどの程度の数であるかが分かれば、そのうちの 1 つの事象が起こることがどの位珍しいことなのか理解できます。この事象の発生数を場合の数といいます。ここではその例として、順列・組合せというものを学びます。

順列とは、異なったものを 1 列に並べる場合の、場合の数を表します。例として、5 個の番号の付いたボールを考えます。これを図 1-1 のように右側に 1 列に並べる場合を考えましょう。

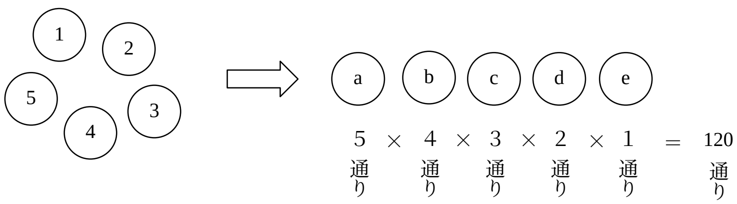

図 1-1 並べ方

最初の a の位置には 5 個のボールから 1 個を置きますので、5 通りの可能性があります。次の b の位置にはすでに 1 個を置いていますから残りで 4 通り、次の c には 3 通り、･･･となります。全体ではこれらを全部掛けて $5\times4\times3\times2\times1=120$ 通りとなります。この 5 から 1 まで掛け合わせた数を 5 の階乗といい、$5!$ と書きます。一般に n から 1 まで掛け合わせた n の階乗は以下のように表されます。

$$n! = n\times(n-1)\times(n-2)\times\cdots\times2\times1$$

但し、この式には当てはまりませんが、使いやすくするため、0 の階乗は、$0!=1$ と定義しておきます。この表式を利用すると、一般に n 個の異なるものを 1 列に並べる場合の数は、$n!$ となります。

次に 5 個の番号の付いたボールから 3 個取り出して 1 列に並べる場合の数を考えましょう。

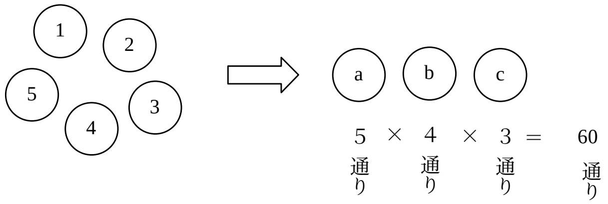

図 1-2 順列

この場合も前と同様に a の位置に 5 通り、b の位置に 4 通り、c の位置に 3 通りの可能性があり、全部で $5\times4\times3=60$ 通りの選び方があります。これを少し式を変えて表現してみましょう。

$$5\times4\times3=\frac{5\times4\times3\times2\times1}{2\times1}=\frac{5!}{2!}\quad\left(=\frac{5!}{(5-3)!}\right)$$

ここで一般に n 個の異なるものから r 個を取り出し 1 列に並べる場合の数を考えてみましょう。この場合の数を記号で ${}_n\mathrm{P}_r$ と書きます。n 個から r 個取り出しますから、n から始めて r 個分即ち、$n-(r-1)=n-r+1$ までの掛け算となります。これを階乗の記号を用いて表わして、以下の式を得ます。

$${}_n\mathrm{P}_r=n\times(n-1)\times(n-2)\times\cdots\times(n-r+1)=\frac{n!}{(n-r)!}$$

当然最初に学んだ、n 個の異なったものを 1 列に並べる場合の数は、以下のようにも計算できます。

$${}_n\mathrm{P}_n=\frac{n!}{(n-n)!}=\frac{n!}{0!}=n!$$

ここで、計算に $0!=1$ を用いています。

問題

以下の場合の数を求めよ。但し、MS-Excel では、$n!=fact(n)$ で与えられる。

1) 6 個の異なるものを 1 列に並べる場合の数

2) 7 個の異なるものから 3 個取り出して並べる場合の数

3) 10 個の異なるものから 4 個取り出して並べる場合の数

解答

1) 720 通り　　2) 210 通り　　3) 5040 通り

1.2 組合せ

次に組合せと呼ばれる場合の数について説明します。順列の場合と同様に、5 個の番号の付いたボールから 3 個取り出す場合を考えます。順列との違いは、「1 列に並べる」ことをしないところです。このことは何を意味するのでしょうか。

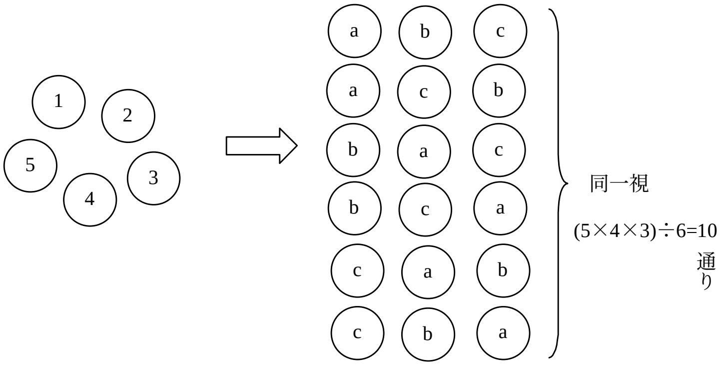

図 1-3 組合せ

順列の場合は 1 列に並べるので、その並べる順番に意味がありました。しかし、組合せの場合は並べないので順番に意味がありません。例えば図 1-3 のように 3 個選び出す場合、順列では異なっていた並べ方の $3!=6$ 通りが同じものになります。すなわち組合せの場合の数は、順列の結果の 1/6 になります。つまり 5 個の異なるものから 3 個取り出す場合の数を ${}_5C_3$ とすると、これは以下のように表わされます。

$$ {}_5C_3 = \frac{{}_5\mathrm{P}_3}{3!} = \frac{5!/2!}{3!} = \frac{5!}{2!\cdot 3!} \left(= \frac{60}{6} = 10 \right) $$

一般に n 個の異なるものから r 個取り出す場合の数は ${}_n\mathrm{C}_r$ と書かれますが、上の表式から、これは次のように与えられることが分かります。

$$ {}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{(n-r)!r!} $$

問題

以下の場合の数を求めよ。但し、MS-Excel では ${}_n\mathrm{C}_r = combin(n,r)$ で与えられる。

1) 10 個の異なるものから 4 個取り出す場合の数

2) 10 個の異なるものから 6 個取り出す場合の数

解答

1) 210 通り　　2) 210 通り

問題

${}_nC_r$ と ${}_nC_{n-r}$ は等しいことを示せ。

解答

$$ {}_n\mathrm{C}_r = \frac{n!}{(n-r)!\,r!} = \frac{n!}{r!\,(n-r)!} = {}_n\mathrm{C}_{n-r} $$

n 個から r 個取り出す場合の数と $n-r$ 個取り出す（r 個残す）場合の数は等しい。

1.3 重複組合せ

これまでの組合せは 1 度取り出したら取り出したままでしたが、ここではもとに戻す場合を考えます。即ち、同じものを何度でも取り出すことができることになります。それゆえこの組合せ問題は重複組合せと呼ばれています。

重複組合せの簡単な問題を考えてみましょう。3 個の数 (1,2,3) から重複を許して 2 個取り出すことにします。答えは、(1,1)(1,2)(1,3)(2,2)(2,3)(3,3)の 6 種類です。但し、右側を大きな数字として並べることにします。さて、この数字の 2 番目にすべて 1 を足してみましょう。結果は (1,2)(1,3)(1,4)(2,3)(2,4)(3,4) となります。これはよく見ると、(1,2,3,4) の中から重複を許さず 2 個取り出した場合に相当しています。

今度は 5 個の番号のついたボールから重複を許して 3 個のボールを取り出す問題を考えてみましょう。

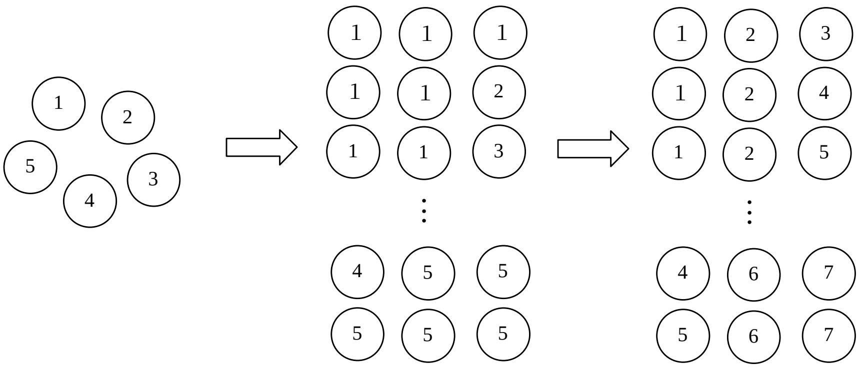

図 1-4 重複組合せ

3 個取り出す場合は 2 番目の番号に 1、3 番目の番号に 2 を足してみます。そうすると 7 個の番号の付いたボールから 3 個重複を許さずに取り出す問題に帰着します。

即ち、重複を許した組合せの数を ${}_5\mathrm{H}_3$ と表わすと、結果は以下のように表わされます。

$$ {}_5\mathrm{H}_3 = {}_7\mathrm{C}_3 = \frac{7!}{4!\times 3!} = 35 $$

上の結果を参考にすると、一般に n 個の異なるものから重複を許して r 個取り出す場合の数は、以下のようになります。

$$ {}_nH_r = {}_{n+r-1}C_r = \frac{(n+r-1)!}{(n-1)!r!} $$

問題

(1, 2, 3)の中から重複を許して 3 個取り出す場合の数を上の方法に基づいて考えよ。

解答

(1,2,3) の中から重複を許して 3 個取り出す組合せは以下となる。

(1,1,1)(1,1,2)(1,1,3)(1,2,2)(1,2,3)(1,3,3)(2,2,2)(2,2,3)(2,3,3)(3,3,3)　10 種類

2 番目の数に 1、3 番目の数に 2 を足すと、上の結果は以下となる。

(1,2,3)(1,2,4)(1,2,5)(1,3,4)(1,3,5)(1,4,5)(2,3,4)(2,3,5)(2,4,5)(3,4,5)　10 種類

これは (1,2,3,4,5) の中から 3 個取り出す組合せになっている。

問題

1) 5 個の異なるものから重複を許して 4 個取り出す場合の数を求めよ。

2) 3 個の異なるものから重複を許して 5 個取り出す場合の数を求めよ。

解答

1) ${}_5\mathrm{H}_4 = {}_8\mathrm{C}_4 = 70$　　2) ${}_3\mathrm{H}_5 = {}_7\mathrm{C}_5 = 21$

1.3 ２項定理

ここでは組合せの考え方を応用した､2 項定理と呼ばれる展開公式について学びます。以下の式を見てみましょう。

$$ (a+b)^2 = (a+b)(a+b) = a^2 + 2ab + b^2 $$

この右辺の式の係数について、a^2 は 2 つの $(a+b)$ の中から 2 つの a を取り出して掛け合わせる訳ですから、2 個のものから 2 個取り出す組合せの数 ${}_2\mathrm{C}_2 = 1$ 個だけ出てきます。ab については a をどちらの $(a+b)$ から取ったかで（残りは必然的に b になります）、2 個のものから 1 個取り出す組合せの数 ${}_2\mathrm{C}_1 = 2$ になります。b^2 については a はどちらからも取りませんので ${}_2\mathrm{C}_0 = 1$ 個だけです。

次に少し複雑になりますが、以下の式を見て下さい。

$$ (a+b)^3 = \underset{1}{(a+b)}\,\underset{2}{(a+b)}\,\underset{3}{(a+b)} $$

$$ = a^3 + \underset{(1,2)(1,3)(2,3)}{3a^2b} + \underset{(1)(2)(3)}{3ab^2} + b^3 $$

ここで、 a^2b の項について 2 つの a をどの $(a+b)$ から取り出したかについて、

${}_3\mathrm{C}_2=3$通りの方法があり、係数が 3 になっています。数式の下に数字の組合せが書いてありますが、それぞれaがどの$(a+b)$から取り出されたか示しています。他の係数についてはa^3が${}_3\mathrm{C}_3=1$、ab^2が${}_3\mathrm{C}_1=3$、b^3は${}_3\mathrm{C}_0=1$となることが分かると思います。

これらのことを元に、$(a+b)^n$について考えてみると以下の公式を得ることができます。

$$
\begin{aligned}
(a+b)^n &= \underset{1}{(a+b)}\underset{2}{(a+b)}\cdots\underset{n}{(a+b)} \\
&= {}_nC_n a^n b^0 + {}_nC_{n-1} a^{n-1} b^1 + {}_nC_{n-2} a^{n-2} b^2 + \cdots + {}_nC_0 a^0 b^n \\
&= \sum_{i=0}^{n} {}_nC_{n-i} a^{n-i} b^i
\end{aligned}
$$

最後のΣ記号はこの他にもいろいろな表わし方があります。例えば、

$$
(a+b)^n = \sum_{i=0}^{n} {}_nC_{n-i} a^{n-i} b^i = \sum_{i=0}^{n} {}_nC_i a^{n-i} b^i = \sum_{i=0}^{n} {}_nC_i a^i b^{n-i}
$$

などです。2 番目の等号は${}_n\mathrm{C}_r={}_n\mathrm{C}_{n-r}$の性質を用いました。3 番目の等号は式を反対側から書き下したものに対応します。

問題

以下の質問に答えよ。

1) $(x+y)^9$のx^3y^6の係数を求めよ。
2) $(2x+y)^9$のx^3y^6の係数を求めよ。
3) $(2x+3)^5$のx^3の係数を求めよ。

解答

1) ${}_9\mathrm{C}_3=84$　　2) ${}_9\mathrm{C}_3\cdot 2^3=672$　　3) ${}_5\mathrm{C}_3\cdot 2^3\cdot 3^2=720$