て見分けたらいいのでしょう。パソコンを使って簡単な実験をしてみましょう。皆さんもよく使うエクセルを使い、

=randbetween（0,1）

と試しに入れて、50回乱数を出してみます。0と1が「等確率で現れるとしても」おそらく交互には出てこないはずです。0が続く場合も1が続く場合もありますね。これがでたらめの妙なのです。神様はさいころを振った直後の数字を記憶していませんから、何回も同じ数を出す場合もあります。

ギャンブルで勝ち続ける場合も、負け続ける場合も出てくるわけです。ほとんど、胴元に有り金全部取られてしまうのがオチ。ギャンブルで胴元に絶対に勝つ方法をこっそり教えましょう。それは、「絶対ギャンブルをしないこと」です。

表　エクセルによる乱数発生実験結果

第1回	第2回	第3回	第4回	第5回	第6回	第7回	第8回	第9回	第10回
0	1	0	1	0	0	0	1	1	0
第11回	第12回	第13回	第14回	第15回	第16回	第17回	第18回	第19回	第20回
0	0	1	0	1	1	0	0	0	0
第21回	第22回	第23回	第24回	第25回	第26回	第27回	第28回	第29回	第30回
1	0	1	0	0	0	0	0	1	1
第31回	第32回	第33回	第34回	第35回	第36回	第37回	第38回	第39回	第40回
0	1	0	0	0	0	0	1	0	0
第41回	第42回	第43回	第44回	第45回	第46回	第47回	第48回	第49回	第50回
1	1	1	1	1	1	0	1	0	1

連の数	1の個数	0の個数	期待値	分散	標準偏差	統計量
26	21	29	25.36	11.61	3.41	
						0.191

さて実験結果を上の表にまとめました。2つの値が等確率で出ているとすれば、50回の試行で1と0がそれぞれ25個になることも期待されますが、そうなっていません。1が少なくて21個、0が29個です。この比率は随分偏っていますから、これで「でたらめ」と言えるのでしょうか。私自身も、確かめるまで少し自信がありません。さあ「でたらめ」かそうでないか確かめましょう。では、確かめるための簡単な方法を紹介しましょう。

まず、0か1が固まって連続して現れたら、それを一まとめにして「連」と呼びます。上の表では、0、1、0、1と続いて5回目から000、となります。ここまでで回数は7ですが、「連」の数は5となります。つまり000を一つのつながりとして数えたからです。そして表にあるように、この実験の連の総数R=26です。

次に期待値Eと分散Vを、データの全個数Nと、1の現れた個数n(1)、0の現れた個数n(0)を使って下の2つの式で計算します。

期待値E＝(N+2×n(1)×n(0))/N

分散V=(2×n(1)×n(0))×(2×n(1)×n(0)-N)/((N-1)×N×N)

皆さん、N=50、n(1)=21、n(0)=29を上の2つの式に入れ計算してみてください。E=25.36とV=11.61が、ちゃんと求まりましたか。

次に連の総数R=26から期待値E=25.36を引いて、それを分散V=11.61の平方根SD（つまり標準偏差）=3.41で割ります。その結果の値（検定統計量）が0.191となります。

この数値が正規分布で5％の有意水準を与える1.96より「明らかに」小さいことから「『でたらめではない』とは言えない」という結論が出ます。「でたらめだ！」とは断定的には言えません。随分まどろっこしい言い方ですね。学問って注意深いんです。

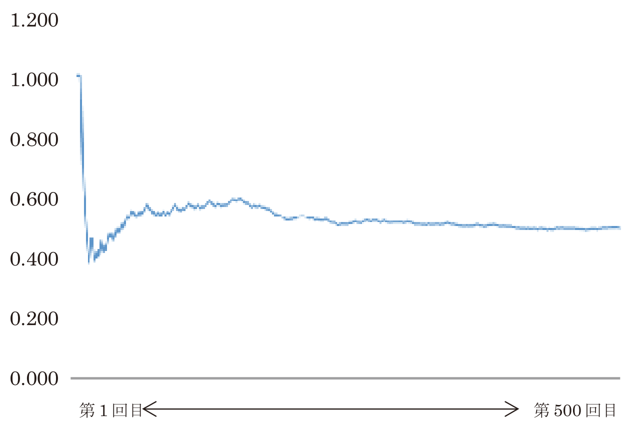
2．信頼できる標本の取り方とは？

⑴　必要標本数の求め方

さて、ある比率で存在する男性と女性、学生と社会人、条例に賛成と反対の両側から有権者を無作為に選びたい場合に、信頼に足る標本の取り方を紹介しましょう。その前に、興味深い法則を紹介します。「大数の法則」といいます。

まず、0と1の乱数を500回発生させます。理論的な平均は0.5です。出た目の累積和3（0＋0＋1＋0＋1＋1＋0＋0という具合に）をその当該回数8で割って「標本平均」（ここまでで計算すると3/8＝0.375）を順に求めていきます。すると「標本平均」がグラフのように次第に0.5にどんどん近づいてゆく様子が見られますね（図表参照）。

図表　大数の法則



この図表から、標本の数が増加するほど標本平均の精度は高まっていくことが確認できます。つまり標本数が増加するほど、信頼のおける結果が得られるということです。でも、標本数を多くすれば調査費や調査時間がバカになりません。財政当局に怒られないためには、標本数をいくつにしたら十分なのでしょうか。

ここまで我慢してついてきてくれた読者の皆さんに思いっきり簡単な式を紹介します。

必要標本数をn、政策に対して賛成m、反対(1－m)の割合とします。分散はm（1－m）です。調査で許容する誤差をεとします。また標準正規分布の信頼係数をzとします。必要標本数nは簡便式

$$n \geq \left(\frac{z}{\varepsilon}\right)^2 m(1-m)$$

で求まります。

本当に知りたいと思っている真の値を含む確率をできるだけ大きくするために信頼水準を95%に設定しますとz=1.96となります。

さて、ε＝0.02（調査の誤差2％）と誤差の大きさを思いっきり小さく、調査の質を上げるために厳しめに設定します。

ところで、意見が賛成と反対で拮抗しm=0.5の時に意見の分散が最大となりますから、必要標本数も最大になります。この厳しめの条件を入れた結果から、

$$n \geq \left(\frac{1.96 \times 0.5}{0.02}\right)^2 = 2401$$

と計算されます。必要標本数はおおよそ2,400くらいになります。つまり、世論が拮抗しているような政策に関して信頼できる調査をしようとすれば、その分標本数を多めに必要とするということです。また試しにεを緩めにして0.1で計算してみてください（なんと、たったの96！）。必要標本数はずっと少なくなりますね。これは「単純無作為抽出」の場合の標本数の目安です。同じ無作為でも「層別抽出」の場合は、分散はもっと小さくなります。「多段抽出」の場合は、分散はもっと大きくなりますから必要標本数は抽出法で違ってきます。

⑵　少数の法則

稀代の風見鶏学者で統計学の大恩人でもあるフランスの大数学者シモン・ラプラスの有名な警句によれば「目に錯覚があるように、心にも錯覚がある」ようです。大数の法則は、「偏りのない」標本数を多くとるほど信頼のおける情報が得られると教えてくれました。そして、統計学は次のようなことも教えてくれます。**標本数が少ないと標本数が多い場合よりも極端な結果が出やすくなる。**これを「少数の法則」と2002年にノーベル経済学賞をもらった心理学者ダン・カーネマン達は名付けています。人間はこの極端な結果に無意識のうちに心を惹かれてしまうかもしれません。まさしくラプラスの言う「心の錯覚」です。恋のきっかけもスポーツ選手のジンクスも同類だ（つまり誤解）と言ってカーネマンは監督たちの顰蹙（ひんしゅく）を買いました。

それはともあれ、ギリシャのように作為のでたらめ数字ではなく、偏りのない信頼できる「民意の数字」を注意深く測定し、行政に活かしたいものです。それではまたの機会に。

［1］対潜水艦戦を重視して設計・装備された軍用航空機