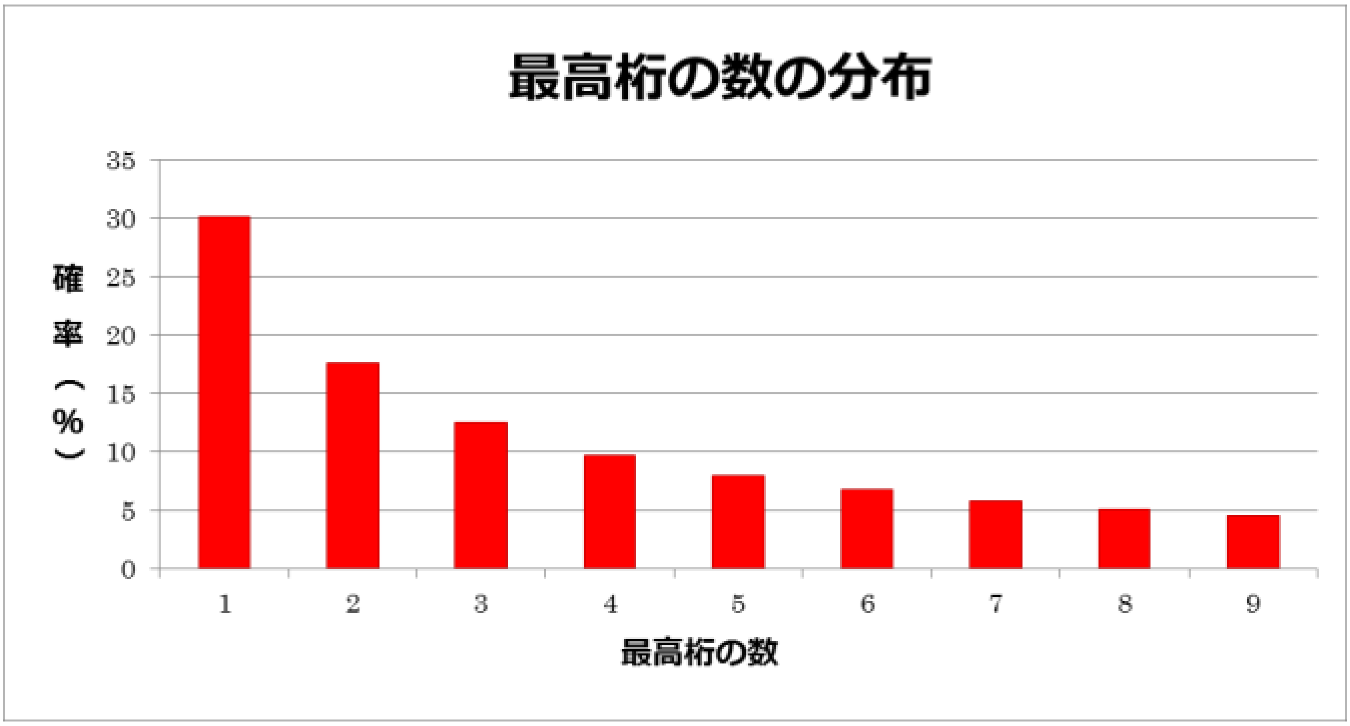
ベンフォード則の統計的分析

1 研究テーマについての説明

ベンフォード則とは、1938 年にアメリカの物理学者であるフランク・ベンフォードが発見したもので、電気料金の請求書、株価、川の長さなどの自然界に現れる多くの 10 進数の数値の最高桁の出現確率がそれぞれ 9 分の 1 ずつではなく、1 が 30.1%、2 が 17.6%、3 が 12.5%、4 が 9.7%、5 が 7.9%、6 が 6.7%、7 が 5.8%、8 が 5.1%、9 が 4.6%という分布を示すという法則である。
この法則によると、数値の分布が、
$\log_{10}(N+1) - \log_{10}(N)$
で求められる。（N には 1 から 9 までの整数が入る）



2 研究の動機と目的

自分の身近なところで現れている最高桁の数の分布は一見何も関連性がなく一様であると思っていたが、ベンフォード則によって最高桁の数の分布に規則性があることを知り驚いた。そこで、本当にこのような分布をしているのか調べてみたいと思いこの研究を始めた。
今回の研究の目的の一つ目は、まずポアンカレのルーレット定理を用いて、ベンフォード則を説明、理解し、身の回りにある数値や統計データにおいて、最高桁の数はどのように分布しているかを調査する。そして、得られたデータから、適合度検定の一つであるχ＾２検定を用いて統計学的に、ベンフォード則が成立しているかについて検証する。今回の研究では、世界の国の人口、面積、そして衆議院総選挙の小選挙区の各議員の得票についてのデータを集め、調査を行った。

二つ目は、ベンフォード則は10進数の最高桁の分布に関する法則だが、5進数などのほかの進法では規則性があるのか、また、どのような分布を示しているのかを世界の国の人口・面積において調査することである。

3 研究の結果

(1)ベンフォード則とは

ベンフォード則で求められる数値の分布になる理由は、ポアンカレのルーレット定理（下図）で説明することができる。十進数を十倍するごとに一桁上がる構造を円に表し、それを1周とする。さらに十倍すると2周すると考える。

1や10、100・・・は対数関数$\log_{10} x$で円周の周回数と対応できる。

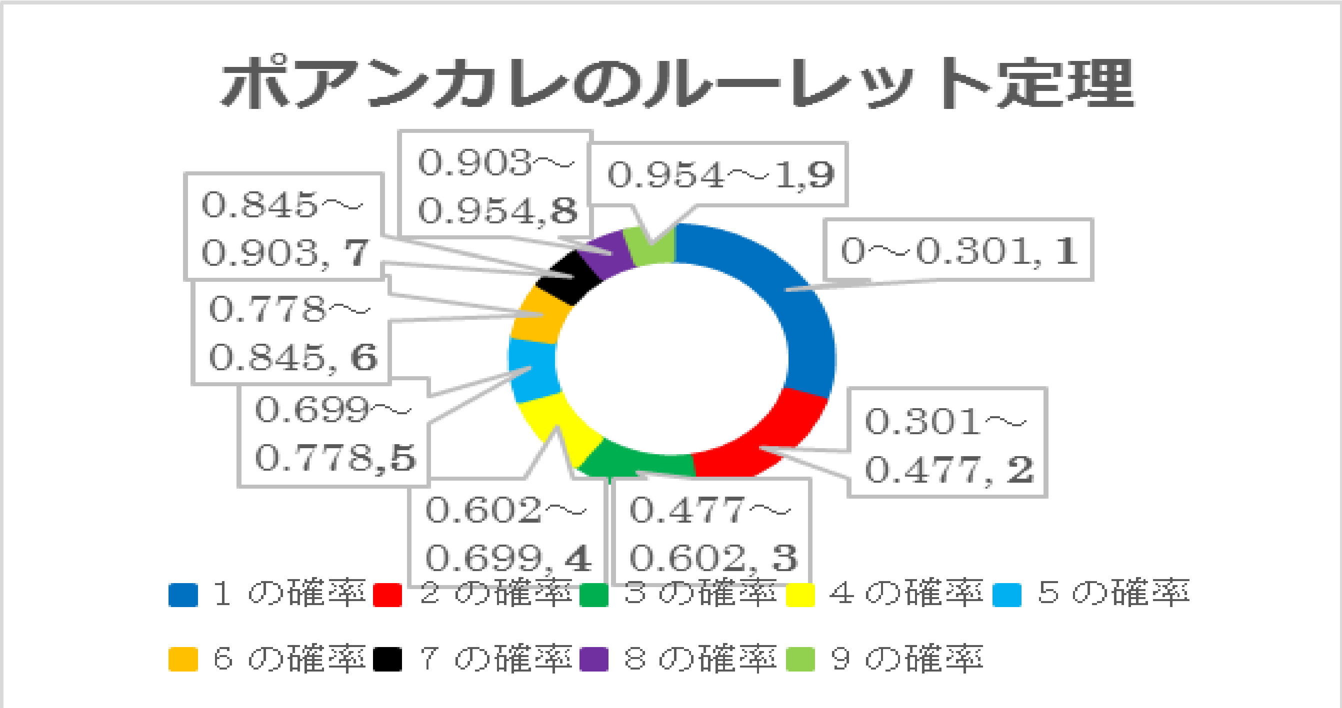
1（周）$=\log_{10} 10$、　2（周)$=\log_{10} 100$　となる。

1から9までの1桁の数は、0周から1周の間にあることになり、それらが何周に相当するかは、

0.301（周）$=\log_{10} 2$、　0.477（周）$=\log_{10} 3$、　0.602（周）$=\log_{10} 4$、

0.699（周）$=\log_{10} 5$、　0.778（周）$=\log_{10} 6$、　0.845（周）$=\log_{10} 7$、

0.903（周）$=\log_{10} 8$、　0.954（周）$=\log_{10} 9$となるので1は0から0.301周、2は0.301周から0.477周とこれ以降も同様に続いていき、9は0.954周から1周となる。また、ベンフォード則では最高桁の数だけに着目していてその数値が何桁であるかは無視することができる。例えば、無作為にとってきた数値の最高桁が3である確率は、$\log_{10} 4 - \log_{10} 3 = 0.125$という計算になる。ほかの数でも同様のことが成立するのでこのような数値の分布になる。



(2)世界の国の人口・面積がベンフォード則に従うか

ア）世界の国の人口

世界の国の人口で、最高桁の 1 から 9 の数の分布を調査した。その結果、人口では、1 が 29.0%、2 が 18.6%、3 が 11.6%、4 が 9.5%、5 が 9.9%、6 が 4.7%、7 が 6.9%、8 が 6.0%、9 が 3.4%となった。そこで、ベンフォード則が成立しているかどうかについてχ＾2 検定を用いて検証した。χ＾2 検定はそれぞれに観察度数と期待度数の差の二乗を期待度数で割ったものを求めて、これらを足し合わせたものをχ＾2とすると、χ＾2は各自由度のχ＾2分布に従うことを利用している。自由度とは、自由に動くことができる変数の個数のことである。例えば、今回の研究では、最高桁が 1 から 8 までは自由に値が変わるが、9 だけは合計の値によって制限されるので、自由度は 8 である。

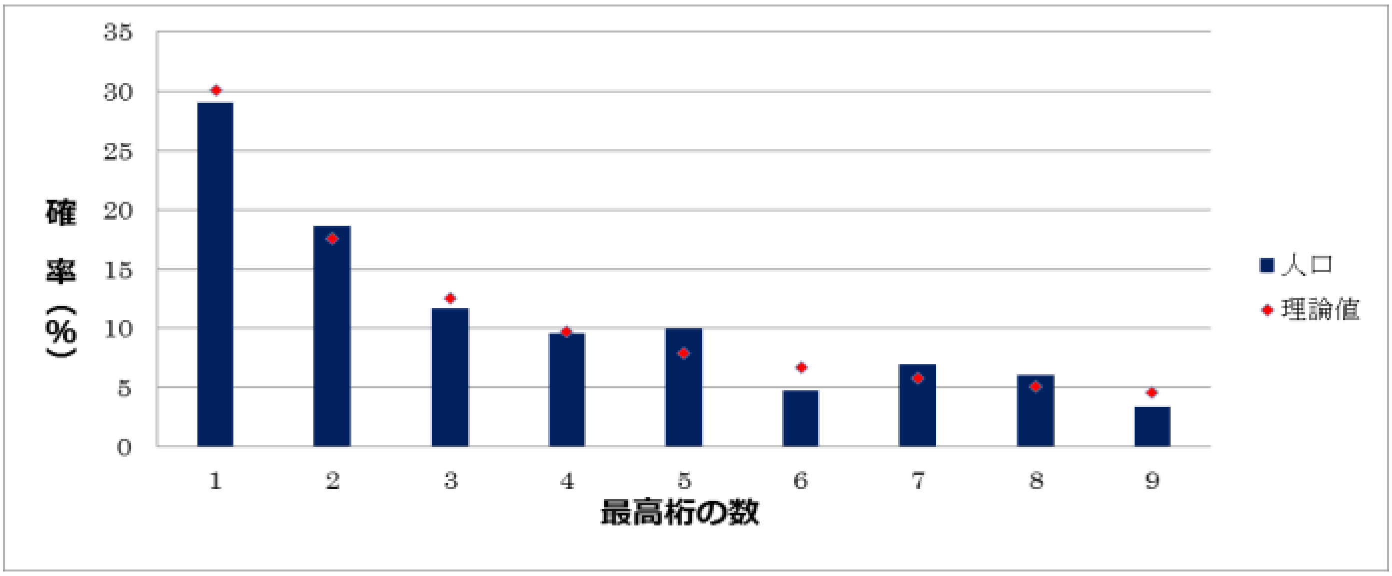
今回の研究では、最高桁の分布は、ベンフォード則に従うことを示したいので、検定の帰無仮説は、ベンフォード則に従わない、対立仮説はベンフォード則に従う、とした。

χ^2 の値は、 f を観察度数、 e を期待値とすると、

$$\chi^2 = \sum_{k=1}^{n} \frac{(fn - en)^2}{en}$$

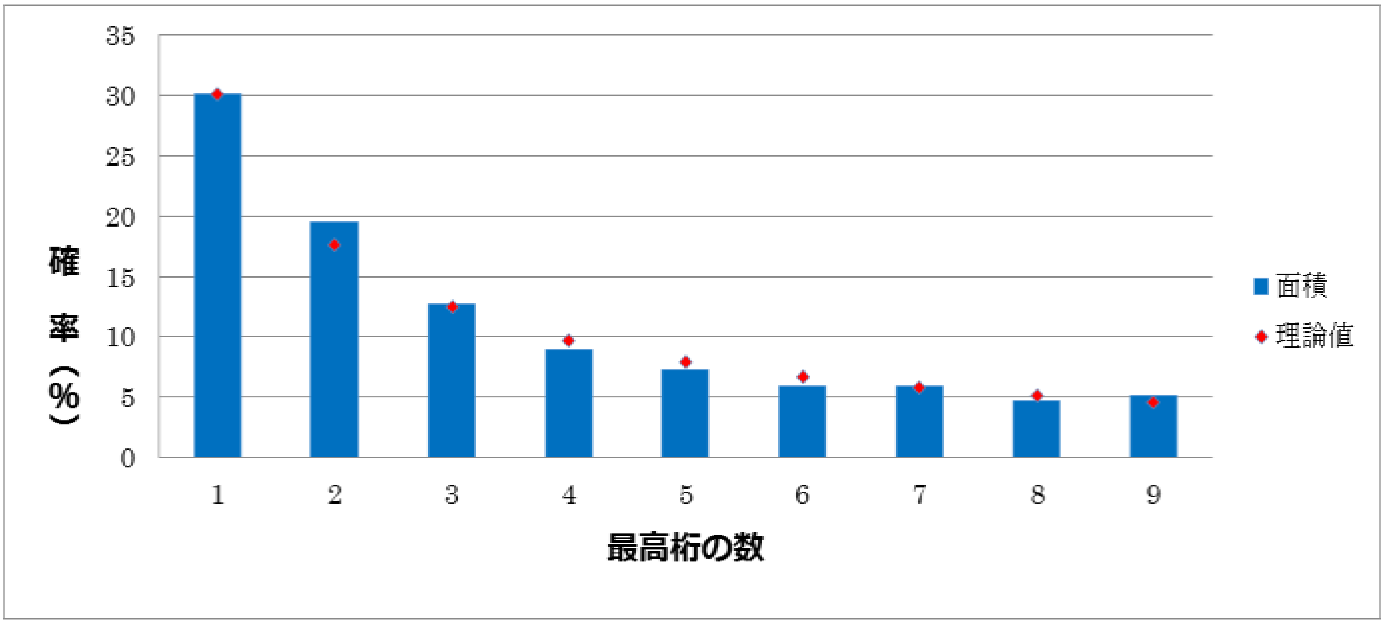
で求められる。

観察度数と期待値が一致している場合には、x＾2 の値は 0 になり、観察度数と期待値の差が大きくなると、x＾2 の値も大きくなる。棄却域とはχ＾2 の値がそこに入ったら帰無仮説を棄却する範囲のことである。自由度 8、有意水準 0.05 においてのχ＾2 分布の上側 5%の棄却域は、～15.5 となる。世界の人口においては、χ＾2 値は 4.46 となり、帰無仮説を棄却した。したがって、世界の国の人口にはベンフォード則に従っている判断した。



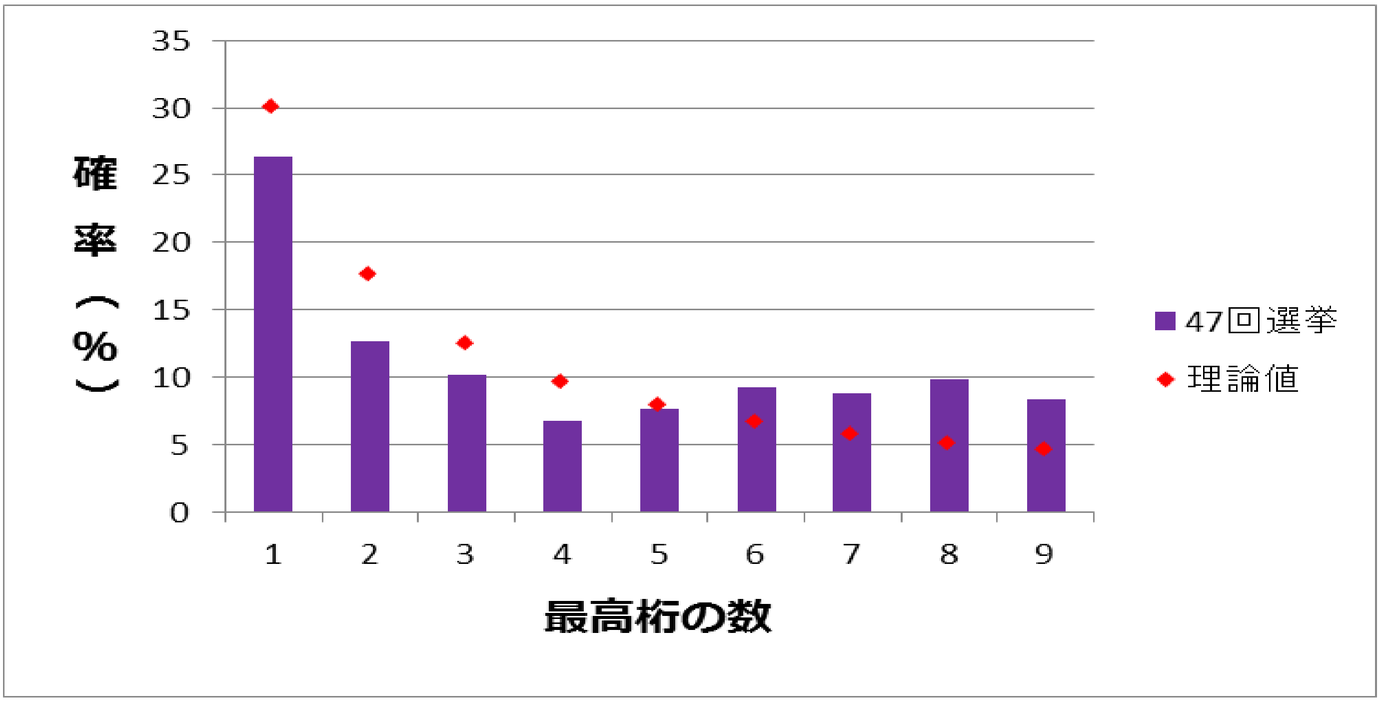
イ）世界の国の面積

世界の国の面積でも同様の実験を行った。最高桁の数の出現確率は、1 は 30.0%、2 は 19.4%、3 は 12.7%、4 は 8.8%、5 は 7.2%、6 は 5.9%、7 は 5.9%、8 は 4.6%、9 は 5.0%となった。世界の国の人口と同様にχ＾2 検定を行った。χ＾2 の値は 1.12 となり、帰無仮説を棄却し、ベンフォード則にしたがっているとした。

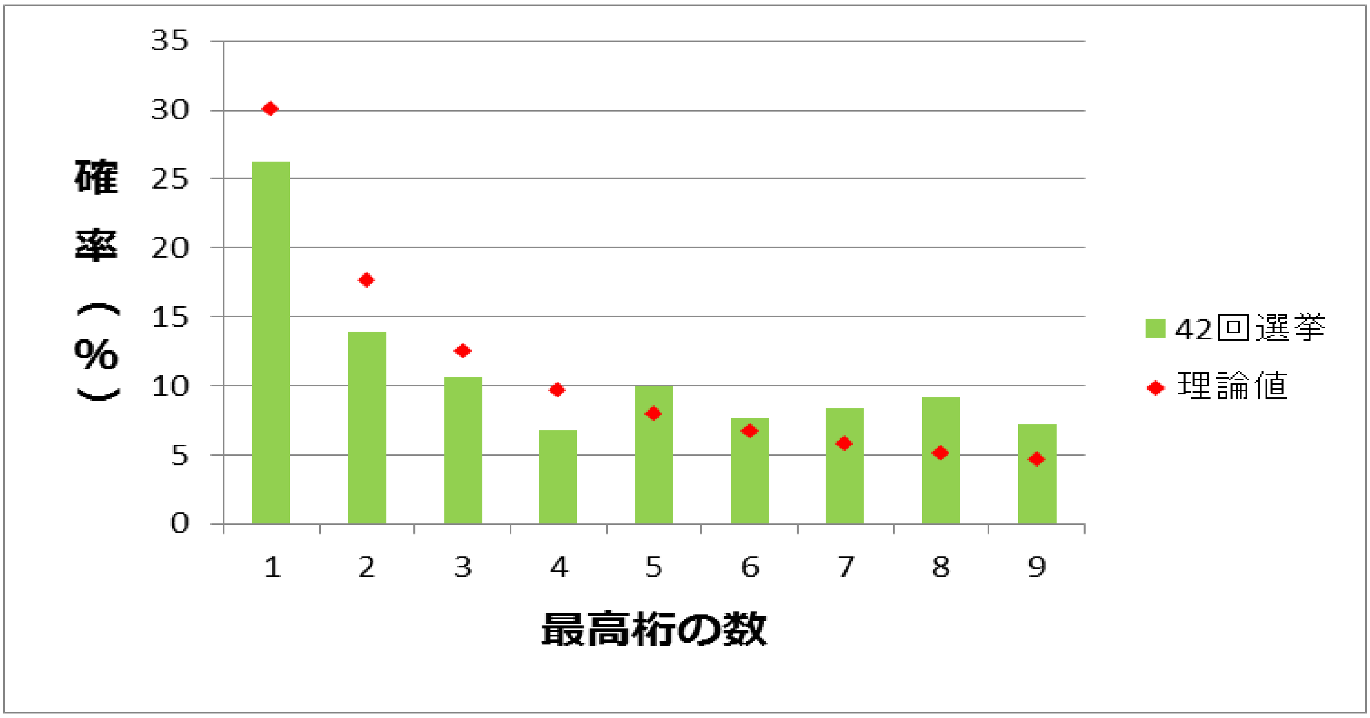


ウ）選挙結果はベンフォード則に従うか、第 47 回と第 42 回の衆議院選挙について調査

第 47 回の衆議院選挙の小選挙区制において各都道府県の全小選挙区の各候補者の得票数を調査し、ベンフォード則の理論値と比較を行った。その結果は、1 は 26.7%, 2 は 13.9%, 3 は 10.5%、4 は 6.7%、5 は 9.9%、6 は 7.7%、7 は 8.3%, 8 は 9.1%, 9 は 7.2%となった。世界の国の人口・面積よりも実測値と理論値の差は大きくなっている。そこでχ＾2 検定を行ったところ、値は 124 となり、ベンフォード則に従わないと判断した。



そして、第 42 回衆議院選挙において同様の実験を行った結果、値は 105 となり、この場合ベンフォード則に従わないと判断した。

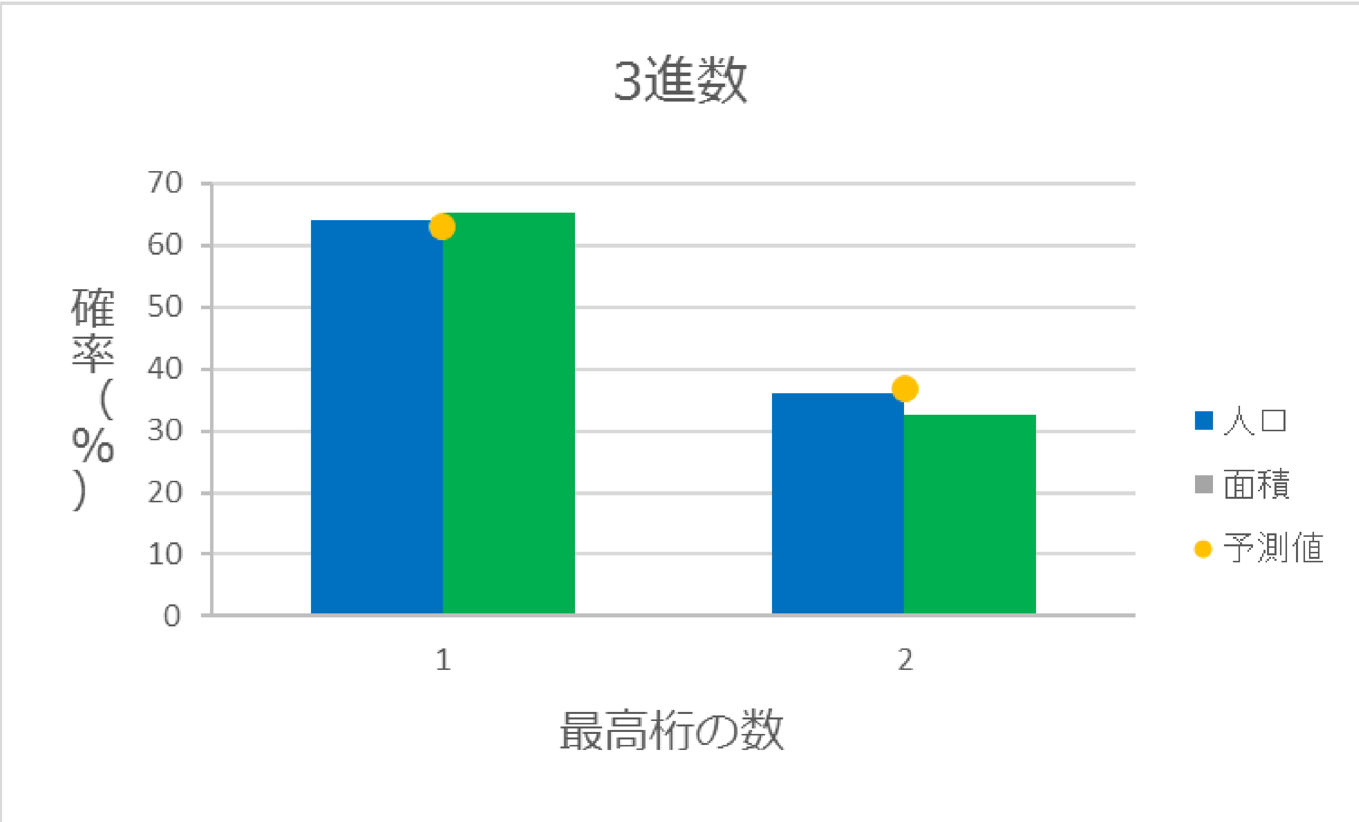


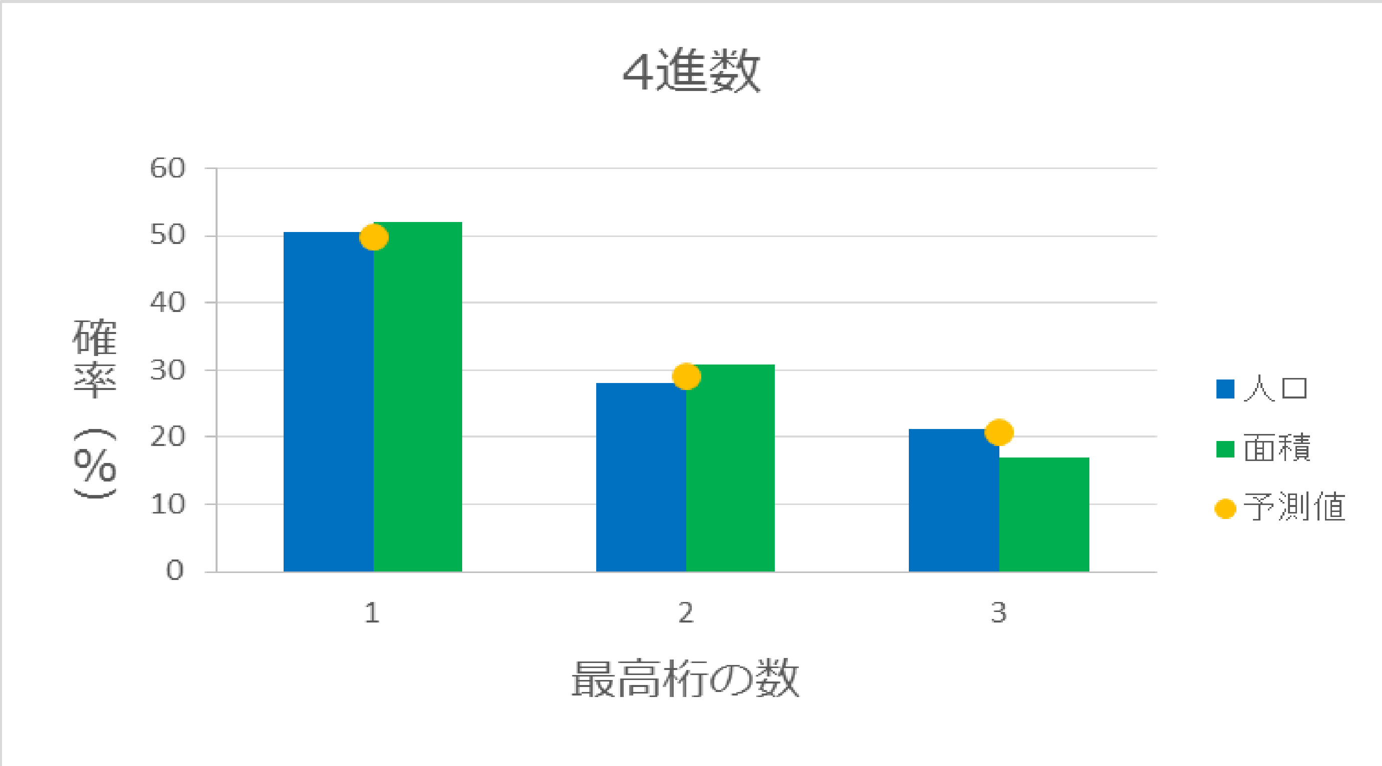
(3)世界の国の人口・面積のN進数(N＝3,4,⋯9)における調査

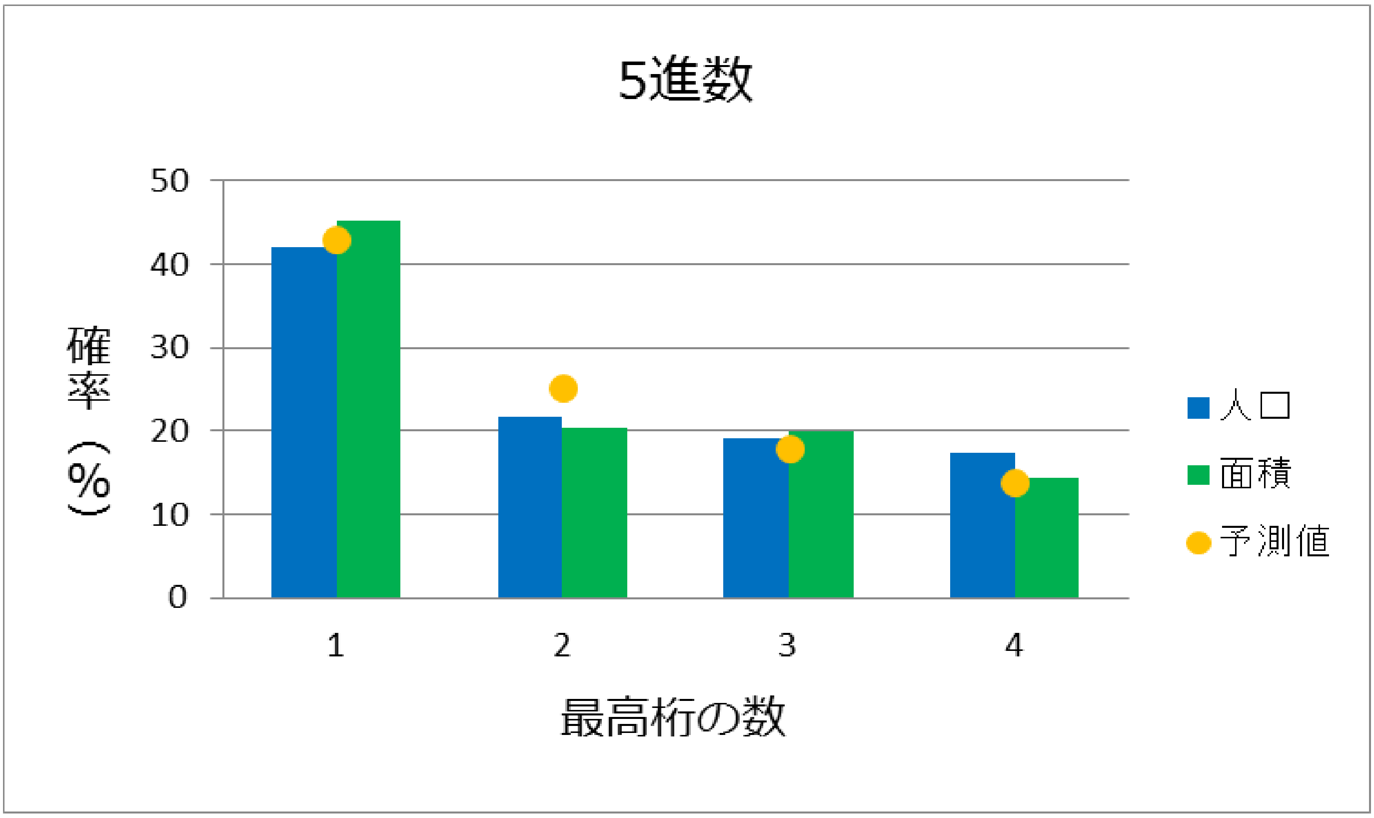
次に、世界の国の人口、面積において 3 から 9 進数の最高桁の数の分布について調査した。それぞれの進法での最高桁の分布は

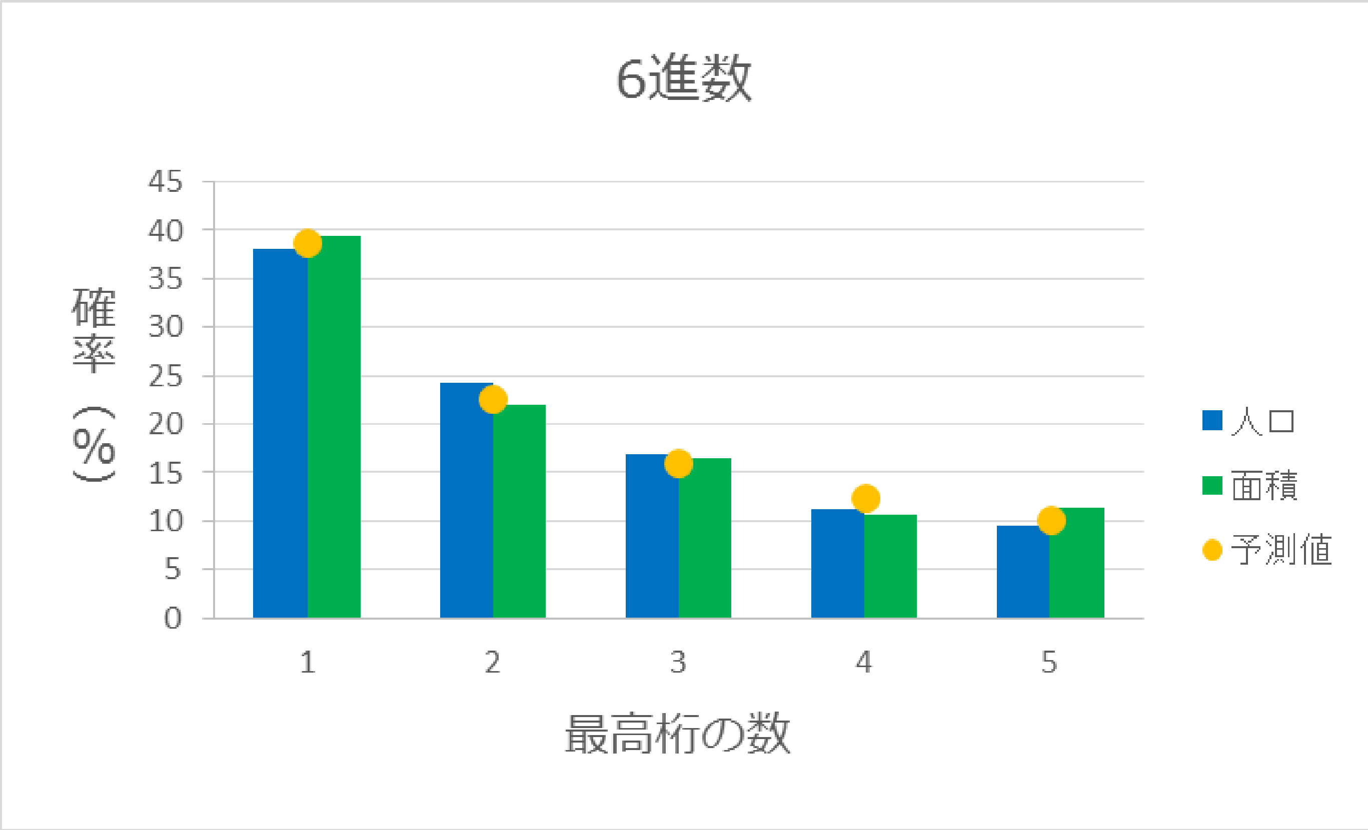
$\log_m(N+1)-\log_m(N)$　（mは進数の数、N は最高桁の数に対応している）

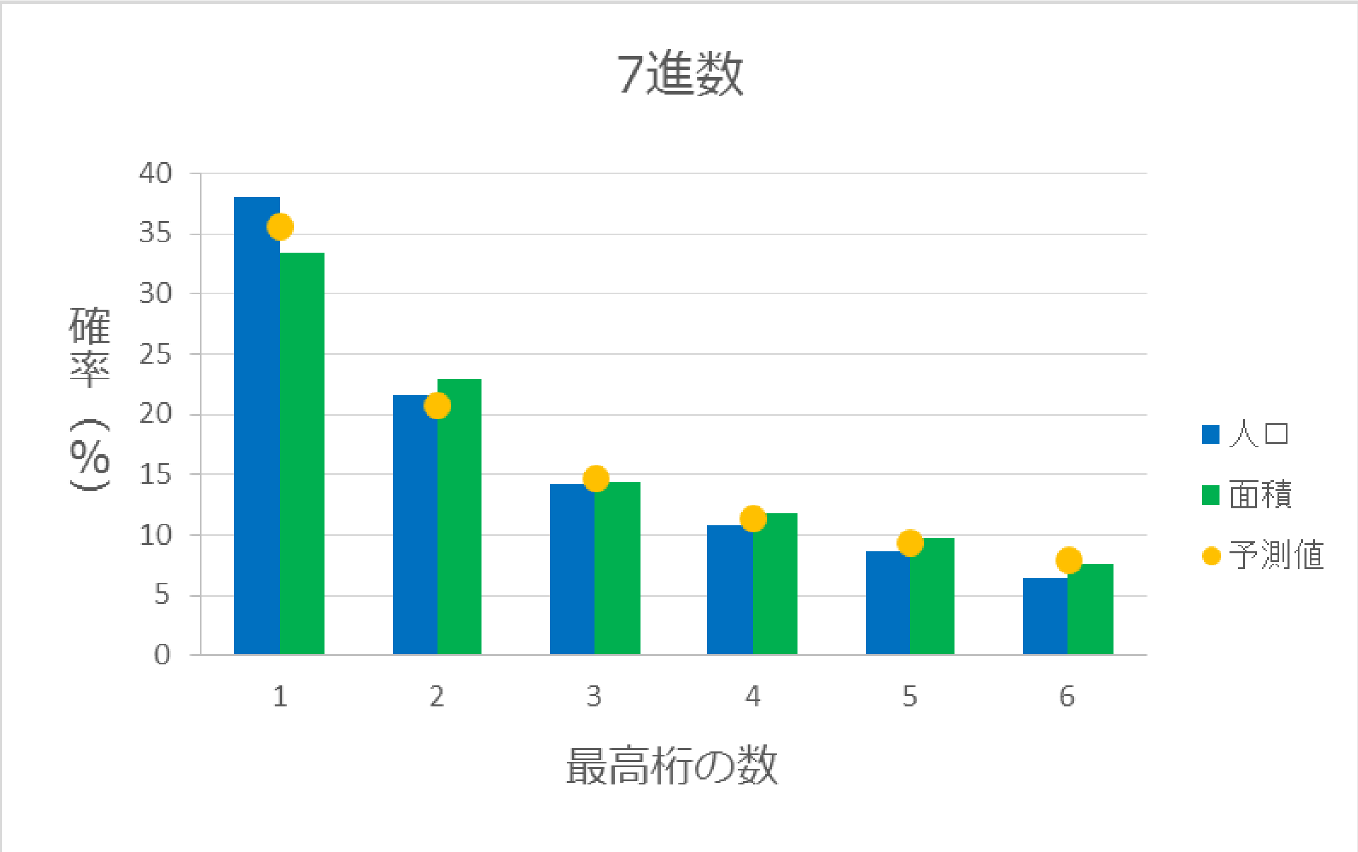
と予測した。

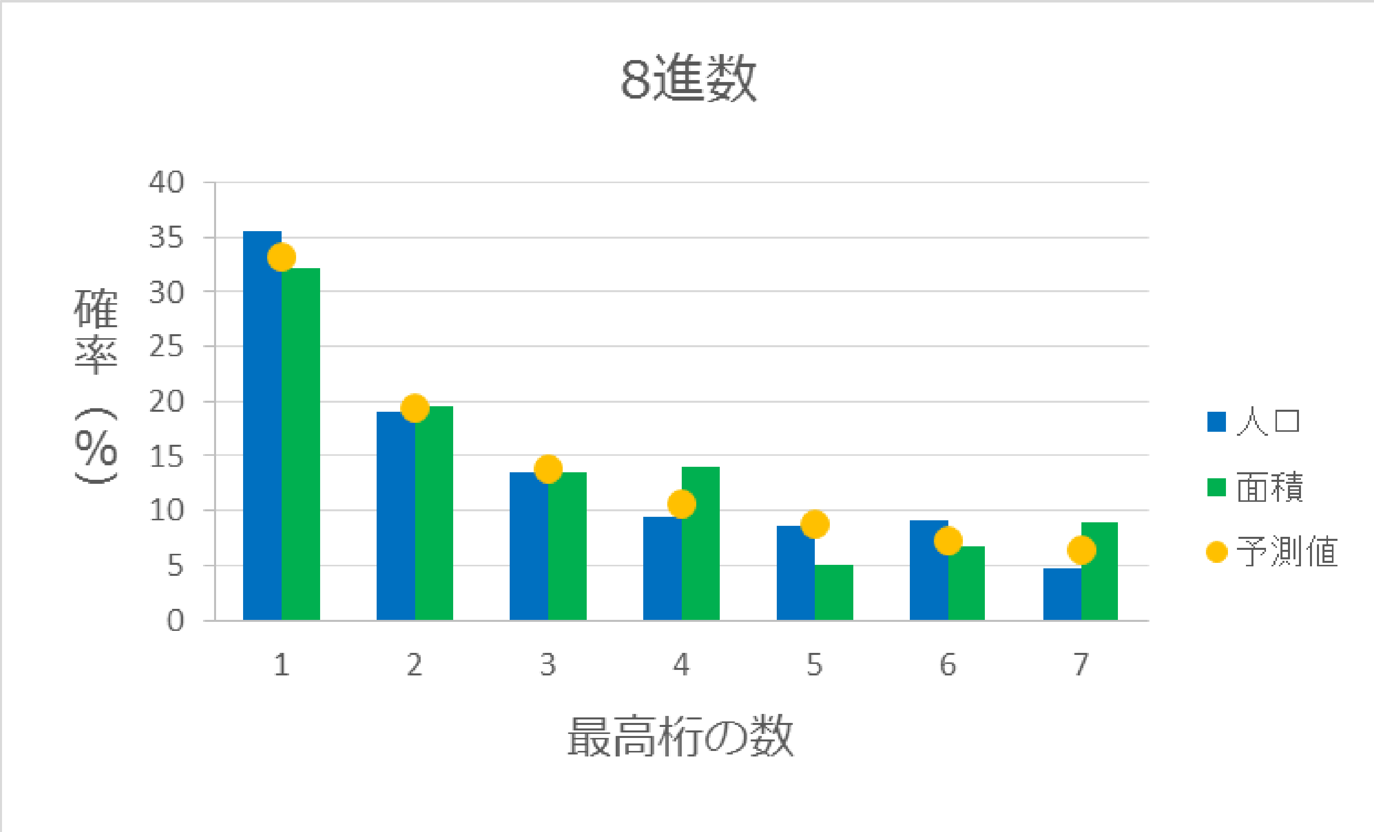


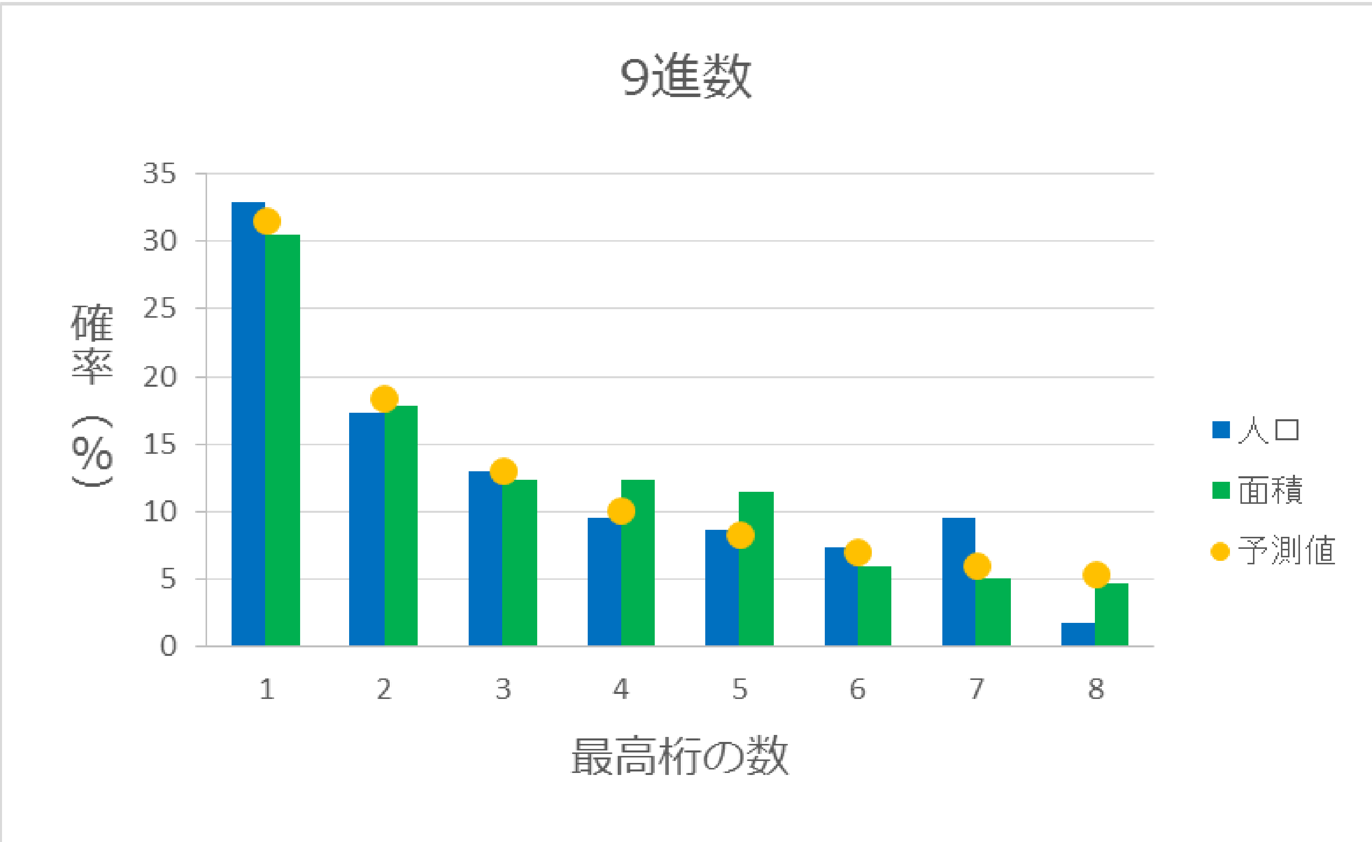

4進数
60
50
40
30
20
10
0
確率（%）
1
2
3
最高桁の数
人口
面積
予測値


5進数
50
40
30
20
10
0
確率（%）
1
2
3
4
最高桁の数
人口
面積
予測値


6進数
45
40
35
30
25
20
15
10
5
0
確率（%）
1
2
3
4
5
最高桁の数
人口
面積
予測値


7進数
40
35
30
25
20
15
10
5
0
確率（%）
1
2
3
4
5
6
最高桁の数
人口
面積
予測値





予測値が正しいことを統計的に示すためにχ＾2検定を行った。有意水準は 0.05 とし、自由度が、3 進数は 1、4 進数は 2、5 進数は 3、6 進数は 4、7 進数は 5、8 進数は 6、9 進数は 7 となる。そして棄却域の値は 3 進数から 3.84　5.99　7.81　9.48　11.0　12.5　14.0 となる。そして、

χ＾2 の値は各進法で(人口,面積)の順に 3 進数(0.04,0.57)　4 進数(0.06,0.88)　5 進数(1.46,1.31)　6 進数(0.32,0.48)　7 進数(0.57,0.38)　8 進数(1.11,3.59)　9 進数(4.61,2.17)となりすべて棄却域に入ったので、仮説が正しいと考察した。つまり、いずれもベンフォードに従うと結論づけた。

5 まとめ・今後の展望

人口、面積の最高桁の分布は、χ^2 検定から統計学的にはベンフォード則に従っていることが言えるが、選挙では、データ数を増やしても、棄却域に入らなく、二回とも同じような分布を示したので、衆議院総選挙の小選挙区の人の分け方によって特有の分布が存在していると考察した。

10 進数の 10 倍すると 1 桁上がる特性を、円周を用いた対数関数による説明が可能なことから、ベンフォード則は最高桁の分布の特徴を示しているものだと考えられる。他の進法について最高桁の数の分布を調査したところ、他の進法でも最高桁の分布が一様なものでなく特定のものになっているので、ベンフォード則を応用できると考察した。

今回の研究では、最高桁の数の分布について調査したが、今後は最高桁以降の数の分布はベンフォード則のような法則性があるかどうかについて調べていきたい。

6 参考文献

1）数学で身に着ける柔らかい思考力　ダイヤモンド社
ロブ・イースタウェイ　ジェレミー・ウィンダム　著　水谷淳　訳
2）第 6 回マス・フェスタ資料
3）例題で学ぶ統計的方法　創成社　井上洋　野澤昌弘　著