

n進数におけるベンフォードの法則

宮城県仙台第三高等学校 理数科1班

1. 背景と目的・仮説

①背景

ベンフォードの法則とは、自然界に出てくる数値の最初の桁の分布がある特定の分布になっている、という法則である。私達は、まだあまり研究されていない他の進数においてもベンフォードの法則が成り立つのかに興味を持ち研究テーマに選んだ。

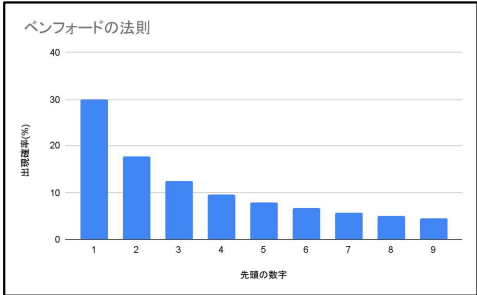


Fig1.ベンフォードの法則

②仮説

私達は、n進法におけるkで始まる数の確率P(n,k)が次の式になると考えた。

$$P(n,k)=\log_n(1+1/k)$$

2. 実験

市町村の人口 (1741データ)と新聞に現れる数字 (764データ)について、進数を変えたデータの最高位の数字を数え、カイニ乗検定と線形回帰分析を行い、進数を変えてもベンフォードの法則に従うか確かめた。

< 実験に用いた手法 >

・カイニ乗検定

次の式を用いてカイニ乗値を求め、求めた値とχニ乗分布表を用いてp値を求めた。

$$\sum_{i=1}^r \frac{(n_i - E_i)^2}{E_i}$$

・線形回帰分析

仮説の式を変形し $P(n,k)=\log_n(1+1/k) \Leftrightarrow n^{P(n,k)}=1+1/k$
 $\Leftrightarrow k \times n^{P(n,k)}=k+1$ 。

ここで $(k, k \times n^{P(n,k)})$ をx-y平面にプロットすると、もし観測値がベンフォードの法則に従うなら、この線は理論値 $(k, k+1)$ に近づく。

3. 結果・考察

(1)市町村の人口

16進数のグラフと線形回帰分析の結果は下のようになった。

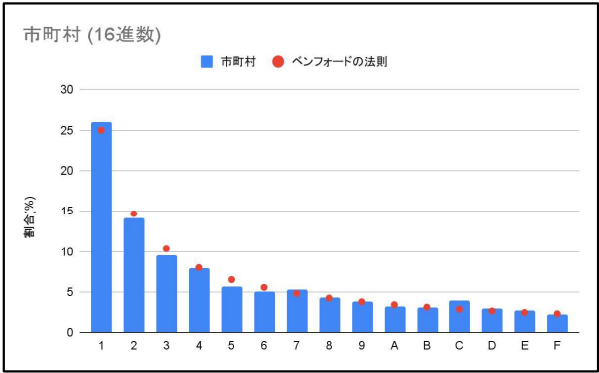


Fig2. ベンフォードの法則 (市町村・16進数)

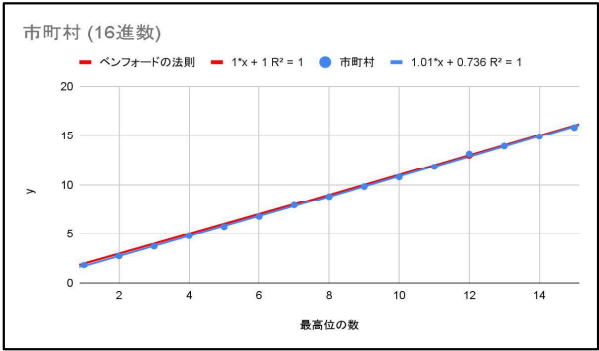


Fig3. 線形回帰分析 (市町村・16進数)

Fig2, Fig3はどちらもベンフォードの法則に従っているように見えるが、実際に従っているか確認するためにカイニ乗検定を行った。帰無仮説は「仮説に従っている」対立仮説は「仮説に従っていない」である。

table1.p値(市町村)

	5進数	10進数	16進数	36進数
p値	0.997	0.998	0.999	1

全ての進数でp値が有意水準(0.05)を超え、帰無仮説が棄却されなかった。ほとんどの進数でp値が0.99を超えていた。

(2)新聞に現れる数字

16進数のグラフと線形回帰分析の結果は下のようになった。

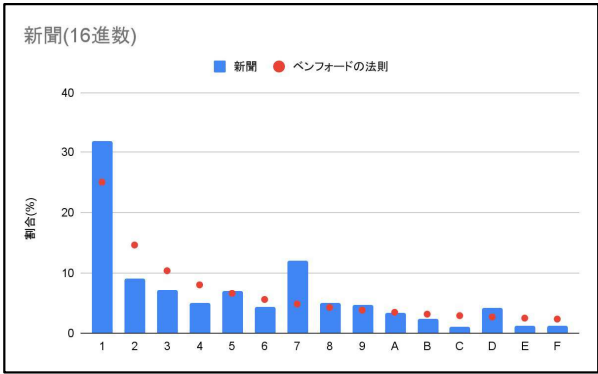


Fig4. ベンフォードの法則(新聞・16進数)

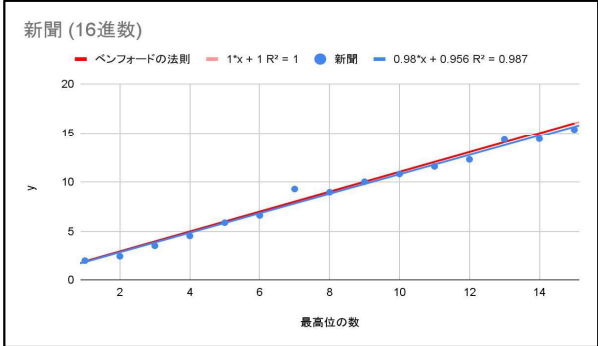


Fig5. 線形回帰分析(新聞・16進数)

(1)と同様にカイニ乗検定を行った。帰無仮説は「仮説に従っている」対立仮説は「仮説に従っていない」である。

table2.p値(新聞)

	5進数	10進数	16進数	36進数
p値	0.554	0.846	0.108	0.963

全ての進数でp値が有意水準である0.05を超え、帰無仮説は棄却されない。進数ごとでp値にばらつきがあった。

4. まとめ・今後の展望

今回は、市町村の人口と新聞に現れる数字をデータとして、実験を行った。

市町村の人口のデータでは、線形回帰分析ではつの線がほぼ一致しp値も極めて1に近かったことから、仮説通りと考えられる。

新聞に現れる数字のデータでは、線形回帰分析ではつの線がほぼ一致したものの、グラフ上での誤差が目立ち、p値にもばらつきがあるため、仮説通りとはいえない。ただし、新聞のデータの数は人口のデータの数よりも少ないため、データ不足により誤差が生じている可能性がある。

今後はデータ数を増やして検証するとともに、最高位以外でも規則性がないかを検証していきたい。

参考文献

1)早苗 雅史・林 一雄・堀部 一経．数学の課題研究:テーマ選びのヒント 第1集．デザインエッグ．2019.