

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

ベンフォードの法則とは、自然界に現れ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、10進数以外の進数においてもベンフォードの法則が成り立つのかということに興味を持ち、研究テーマに選んだ。n進数におけるkで始まる数の出現する確率 $P(n,k)$ の式の仮説を立て、実際に市町村の人口、新聞に現れる数字を集計した。集めたデータが仮説の式に従うか確かめるために、カイ二乗検定、線形回帰分析を行った。実験の結果から、市町村の人口のデータは仮説の式に従うと考えられる。しかし、新聞のデータは、データ数の不足のためか、仮説の式に従うと判断することはできなかった。

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一様な分布ではなく、1の割合が一番多く、数字が大きくなるにつれて出現する割合が下がっていく分布になっている、という法則である(Fig1)。私達は、まだあまり研究されていない10進数以外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ついて興味を持ち、研究テーマに選んだ。

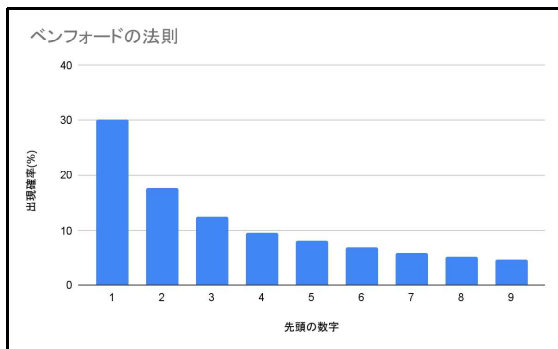


Fig1. ベンフォードの法則

2 仮説

参考文献より、10進法におけるkで始まる数が現れる確率 $P(k)$ は次の式になるとわかっている。

$$P(k)=\log_{10}(1+1/k)$$

私達は、m進法におけるkで始まる数が現れる確率 $P(n,k)$ を、上記の式の対数の底にnを代入することで求め、次の式になると考えた。

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

3 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た。数字が出現する確率が仮説の式に従っているか、以下の2つの手法を用いて確認した。

・カイ二乗検定

帰無仮説を「データがベンフォードの法則に従

う」、対立仮説を「データがベンフォードの法則に従わない」とした。次の式を用いてカイ二乗値を求め、求めた値とカイ二乗分布表を用いて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。

4 結果

(1) 市町村の人口

市町村の人口について、10進数、16進数のグラフと線形回帰分析・カイ二乗検定の結果は以下のようになった。

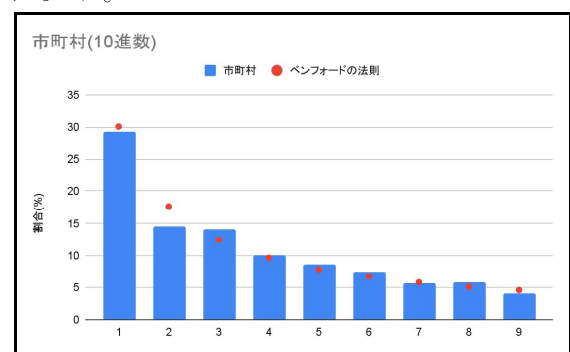


Fig2. 市町村・10進数

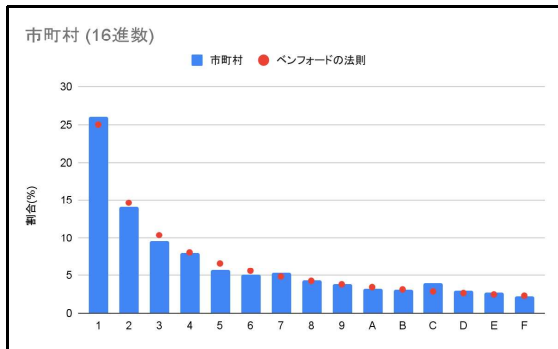


Fig3. 市町村・16進数

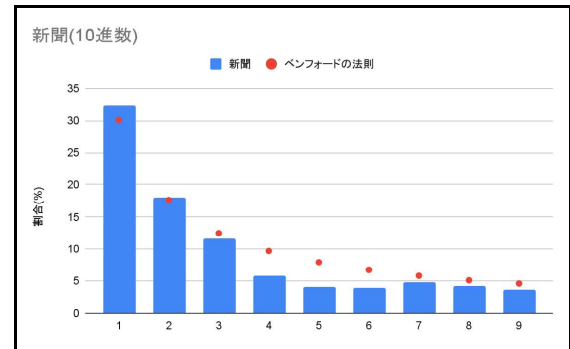


Fig6. 新聞・10進数

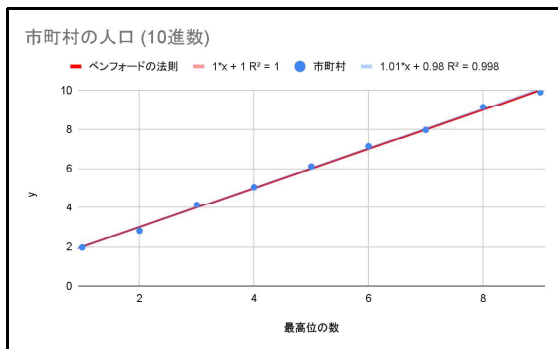


Fig4. 線形回帰分析・市町村・10進数

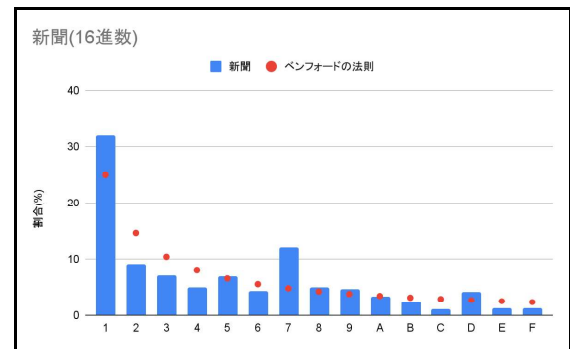


Fig7. 新聞・16進数

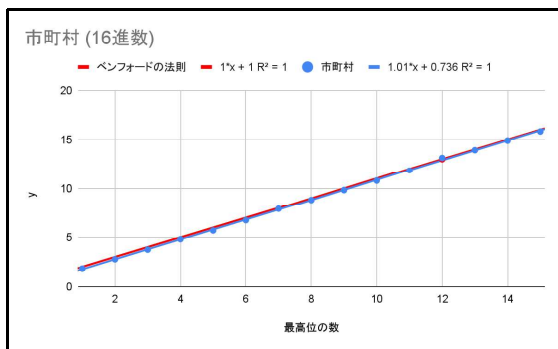


Fig5. 線形回帰分析・市町村・16進数

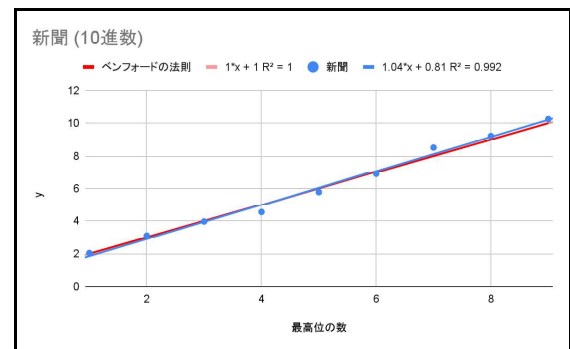


Fig8. 線形回帰分析・新聞・10進数

Table1. カイ二乗検定 (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。またp値は1にかなり近づいていた。

(2) 新聞に現れる数字

新聞に現れる数字について、10進数、16進数のグラフと線形回帰分析・カイ二乗検定の結果は以下のようになった。

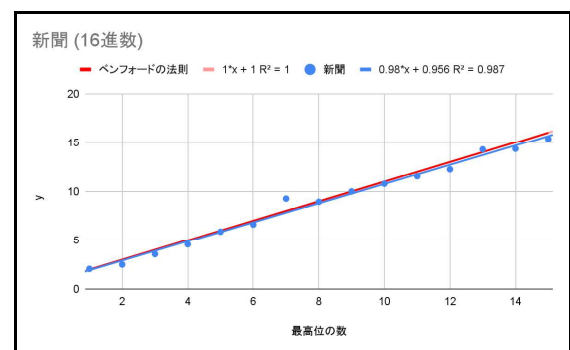


Fig9. 線形回帰・新聞・16進数

Table2. カイ二乗検定 (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。進数ごとにp値にばらつきがあった。

5 まとめ

(1) 市町村の人口のデータ

線形回帰分析では2つの線がほぼ一致し、カイ二乗検定においても、p値が極めて1に近かったことから、仮説が成り立っていると考えられる。

(2) 新聞に現れる数字のデータ

線形回帰分析では2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、仮説通りとは判断できない。ただし、新聞から集計したデータの数は市町村の人口のデータの数よりも少ないため、データ数の不足により正確な結果が得られていない可能性がある。

6 今後の展望

新聞に現れる数字での実験においては、データ数を増やしてより正確に検証していきたい。また、今回は最高位の数字の規則性について考えたが、他の位でも規則性がないか調べてみたい。

7 参考文献

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.