

回帰偏差値の考案とその効用： 地域間格差を相対評価する偏差値

日下部 眞一

広島大学総合科学部

The concept and usefulness of the regression-based score and its application: A new index, “the regression-based score (R-score)” to measure the relative size of the local statistics

Shinichi KUSAKABE

*Faculty of Integrated Arts and Sciences, Hiroshima University, 1-7-1,
Kagamiyama, Higashi-Hiroshima 739-8521, Japan*

Abstract : Although the per capita statistics is usually and customarily used in the analysis on the relative evaluation of the statistics of the prefecture, these measures contain an artificial flaw due to the nonlinearity between the population number and the size of the statistics in question. For the purpose of overcoming these problems on nonlinearity, the regression-based score (R-score) was originated as a new indicator to measure the relative size of the statistics of the prefecture. This index is the application of the Studentized residuals, calculated based on the regression analysis. This index seems to be very useful in comparative evaluation on the size of the various statistics among the prefectures. The size of the various statistics of the prefecture was analyzed with this regression-based score (R-score), and the usefulness and the power of the new index was critically evaluated.

Keywords : Regression analysis, Regression-based score (R-score), Studentized residual.

1 . はじめに

都道府県の統計指標(たとえば『統計でみる県のすがた』など)を眺めてみると、かなり多くの指標(たとえば病院数、県民所得、土木費ほか)が人口当たりの数値で表されている。そして指標値と同時に、その人口当たりの数値の大小によって順位付けされている。例えば『統計でみる県のすがた』にのっている約450項目の指標値では、人口の最も少ない鳥取、島根、高知の3県のいずれかが1位か最下位の47位になっている場合がかなり多い(455指標項目の内95項目)ことに気づかされる。もちろん、この現象が真実を表しているのかもしれないが、しかし、単なる人為的な統計操作によってこのような現象がでてくる可能性が十分考えられるし、もし、そうであれば、人口当たりの数値で表して指標の比較を行って相互評価する時には、人口の大小に依存した誤った評価を行って

いることになり、これは社会的にも大きな問題である。

ここ数年の行政改革では行政が行う様々な事業の評価が取り上げられている。そして、その評価は新しい事業企画へ考慮されることになる。この時の行政単位は、ふつう都道府県が行政単位であり、都道府県単位でさまざまな地域間比較が、できるだけ客観的な基準で行われることがのぞましい。こういう観点に立って都道府県のさまざまな統計指標をながめてみると地域比較という点においてはさらに大きな改良がなされるべきであるように思われる（大友 1997）。

筆者は、特定非営利活動促進法（NPO法）施行後の各都道府県認証NPO数の地域間変異を考察し、各地域のNPOセクターの規模を客観的に相対評価できる“回帰偏差値”という評価値を考案した（日下部 2002）。これは、統計学的には回帰分析における残差分析において Studentized residual（スチューデント化残差）として理論化されている偏差であり、筆者のいう回帰偏差値は、このスチューデント化残差を10倍して50を加えた値である。この回帰偏差値は、通常の偏差値、つまり正規分布を基礎にした正規偏差値同様、きわめて有用であり、相対評価として有力な手法であると考えられるので、ここにその手法と適用例を詳細に検討して解説したい。

2. 地域指標の大きさを相対評価するにあたっての課題

従来、慣習的に一人当たりの数値で様々な指標の大小の順位付けを行ってきた相互評価は、統計的には実測数で表現して、

$$Y(\text{指標値}) = a * X(\text{人口数})$$

の式において、“さまざまな観測値aを比較する値”として相互評価しているわけで、“人口数に依存しない値を同等とすること”を暗黙の内に仮定していることを意味している。つまり100人に対する10個も1000人に対する100個も10000万人に対する1000個も、人口当たりの数値で行う相互評価では同じ0.1という評価になる。

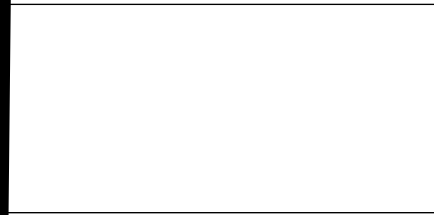
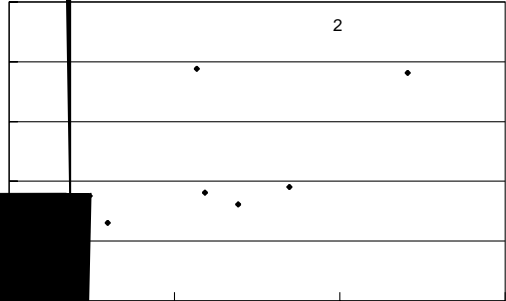
いいかえれば、“人口でわった数値を比較分析の基準とし、それらの指標値を用いて相互評価を行うこと”は、実は、“人口と指標値の間に線形性が保証されている”ことを暗黙の内に想定しているのである。したがって基本的には、この仮定、“人口要因を独立変数としたときに指標値と人口との間に線形性が成り立つのか”、つまり“従属変数を指標値、独立変数を人口としたときに1次回帰直線の傾きが一定数であることを仮定してよいかどうか”を検討しなければならない。

しかし社会現象には“経験的”に、この傾きが一定でなく上に凸の指数関数的分布を示したり、下に凸の指数関数的分布を示したりする場合が多く見られる。例えば、経済学における生産要素に関する収穫逓減の法則は一般的に上に凸の指数関数で表されることが良く知られている。もし上に凸であれば、このような状況で人口当たりの数値でランキングや相互評価を行えば、人口が小さい地域が過大評価され、人口が大きい地域が過小評価されることになる。先に述べた、都道府県のさまざまな指標のランキングで人口が小さい県が上位によく並ぶのはこのような過大評価の効果ができてしまうためである。

本論は、都道府県ランキングなどの相対評価を行おうとする場合には、正であれ負であれこのような規模の効果を出来るだけ取り除いて相互評価をしたほうが指標値のより客観的な相対評価になると考え、このような影響をできるだけ克服する相対評価の指数、“回帰偏差値”を考案し、これを用いて統計指標値の相対評価を試みることを目的としている。

まずはじめにこのような課題を解決できる方法として回帰偏差値の手法を述べる。次に、正の規模の効果（効用逓減）を示す場合、規模の効果が見られない場合、負の規模の効果（効用逓増）を示す

郵便局数 蜂巣2



していないことにまず注意しておく必要がある。次に、統計的に有意となる直線回帰性(この場合は負の相関)が認められる。実際に直線回帰を行い回帰直線の傾きを求めると -0.2323 となる。この傾きの0からの統計的有意性を検定すると t 値(自由度は45)は 6.04 となり高度に有意である。相関の度合いを示す相関係数の値は、 0.4477 であり統計的に高度に有意で ($t = 3.25$, 自由度45) 統計的に有意な正の規模の効果がみられるのである。したがって、このような分布を示す場合に、人口当たりの数値の大小でもって順位付けする場合には、当然、図3における人口約300万以下の小さい地域が上位を占めるという結果になってしまう。

それでは、どのようにしてこのような分散の不均一性を除き、非線形的効果を除くことが可能であろうか。これらを克服するために、まず、両対数変換を行う。図1と同じデータを両対数変換して示したのが図3である。対数は実測値との対応が容易なように10を底とする対数を用いた。人口数は10万人あたりで表している。

対数変換により分散は安定化し、回帰直線の周辺における点の分布が正規的になり等分散性がたもたれる。この図における回帰直線の傾き 0.5901 は $t = 8.56$ (自由度45) で、統計的に高度に有意に1とは異なっている。この傾きがもし1であれば、両対数変換しないときに線形関係であることを意味しており、対数変換後の傾きが1より大きいということは両対数変換しないときに下に凸の指数関数的関係があることを意味している。この傾きが1からずれるにつれて非線形性の効果は大きい。

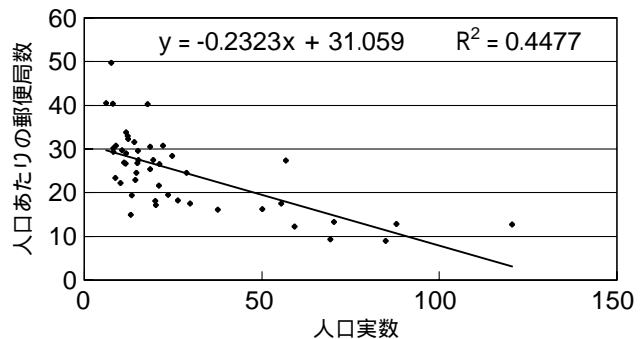


図3 各都道府県の郵便局数と人口(10万人)の両対数変換値の分布図

3.2 回帰偏差値の概念と手法

対数変換後の郵便局数と人口の分布は、観測点間の距離は実数の分布に比べると変わっているが、観測点相互の位置関係は保たれたままである。したがって、分布としては同型で線形近似化され分散が安定化されたことになる。それでは、どのようにしたらこれらの観測点を相対評価に耐え得る値にできるであろうか？それは、両対数変換値に引いた回帰直線を基準点として考えることである。つまり回帰直線を基準として、回帰直線による予測値から各観測点間の距離、残差(偏差)を相対評価の値として用いることである。しかし、統計学的には分布の中心である平均値からずれるにつれて分散が大きくなることが知られている。このような分布を、分布の同型性を保ったまま平均0、分散1の分布に変換するのが影響力係数を用いた変換で、統計学的には“残差のスチューデント化(studentized residual)”として知られている(Myers 1990, Cook and Weisberg 1994, 1999)。

直感的にわかりやすく表現すると図3の分布図で、回帰直線がx軸になるように、やや右回転させ回帰直線上の横長い楕円状の観測点の分布をx(人口)の平均値上に投影して、その分布が分散1、平均0になるように押し縮めてやる変換であると考えればよい。分布の同型性を保ったままの変換であるから厳密には正規分布とはいえないのであるが、標本数が多いと独立性の欠如を無視しても差し支えないので、たがいに独立に正規分布に従うとみなしてもよい(Chatterjee and Price 1977)といわれている。このスチューデント化残差は尺度フリーで、標準化しない残差よりも客観的な残差分析に適していると評価されていて今日では広く普及し始めている(Myers 1990)。

3.3 回帰偏差値をもとめる手順

新しい指数を考案した基本的な考えは、人口規模から期待される郵便局数からの現実の郵便局数の残差を客観的に標準化することにある。そして、この期待される郵便局数というのが人口への1次回帰によって説明される値、つまり1次回帰式によって推定される予測値である。標準化の方法は、回

4. 回帰偏差値による解析と結果

『統計でみる県のすがた』には約450項目の統計指標が載せられている。これらの中から13指標を選び、都道府県人口とそれら指標を両対数変換し単回帰分析を行った。この回帰直線の傾きが1より大きいときは、実数で表した両者の関係が規模の効果としては下に凸の指数関数的関係(規模の効用逓増)を意味しており、逆に、1より小さいときは、実数で表した両者の関係が規模の効果としては上に凸の指数関数的関係(規模の効用逓減)を意味している。傾きがほぼ1の場合は、実数で表した両者の関係に規模の効果が見られず、直線関係であることを意味している。これらを表す解析結果を表1にしめす。この表1の4列目には回帰直線の傾き(b)の傾き1からの有意性、5列目には回帰直線の傾き(b)の傾き0からの有意性をt値で表している。

この表の最初の5項目は、実数で表した指標と人口の関係が規模の効果としては上に凸(効用逓減)を意味している。次の4項目は規模の効果が見られない場合で、回帰直線の傾きはほぼ1である。最後の4項目は、実数で表した指標と人口の関係が規模の効果としては下に凸(効用逓増)を意味していて、回帰直線の傾きは1より有意に大きい。老人ホーム数は老人ホーム数と65歳以上人口との関係で解析したが、これ以外は全て問題指標と都道府県の総人口との関係で解析した。

表1 都道府県の統計指標値の規模の効果

	切片 : a	傾き : b	傾きbの標準誤差	t値(傾きbの1からの有意性)	t値(傾きbの0からの有意性)	AIC
(1) 回帰直線の傾きが1から有意に小さい指標						
郵便局数	1.8976	0.5901	0.0479	-8.5584	12.3218	-212.62
GSスタンド数	2.1870	0.6348	0.0458	-7.9662	13.8469	-216.73
救急車数	1.1012	0.6848	0.0393	-8.0185	17.4211	-231.19
病院数	1.1409	0.7711	0.0673	-3.3992	11.4538	-180.60
老人ホーム数	0.0583	0.7840	0.0477	-4.5250	16.4207	-226.76
(2) 回帰直線の傾きがほぼ1である指標						
薬局数	1.5437	0.9988	0.0378	-0.0328	26.4336	-234.91
電力量	1.8013	1.0182	0.0368	0.4954	27.6681	-237.38
火災死傷者数	1.7614	1.0493	0.0391	1.2587	26.8077	-231.59
交通事故数	1.7447	1.0730	0.0475	1.5384	22.6021	-213.44
(3) 回帰直線の傾きが1より有意に大きい指標						
書籍額	3.0935	1.1124	0.0385	2.9232	28.9304	-233.26
旅券数	1.3370	1.1945	0.0413	4.7091	28.9163	-226.52
郵便物数	3.7213	1.2708	0.0600	4.5101	21.1656	-191.37
ガス	0.5540	1.9773	0.1180	8.2838	16.7596	-127.87

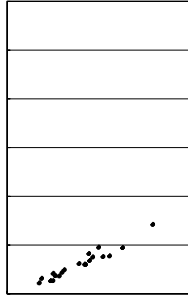
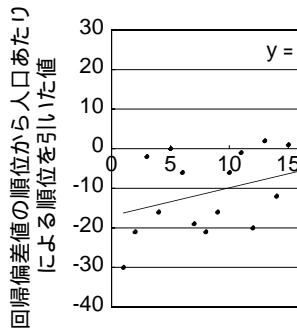
これら13項目の中から規模の効果が効用逓減の場合として郵便局数、規模の効果が見られない場合として薬局数、そして、効用逓増の場合として一般旅券発行数の場合を選んで詳細な解析を行い、回帰偏差値の効用と有効性を確認していく。

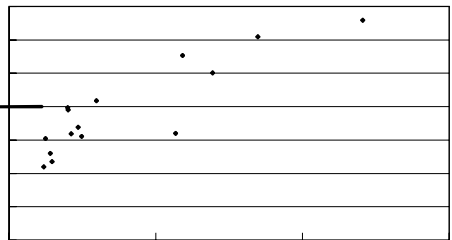
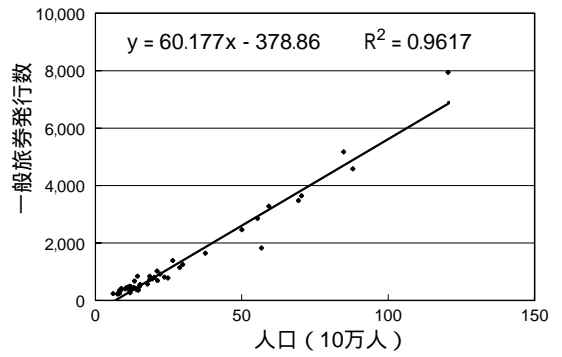
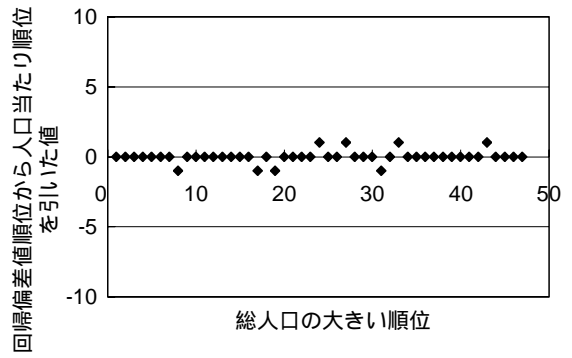
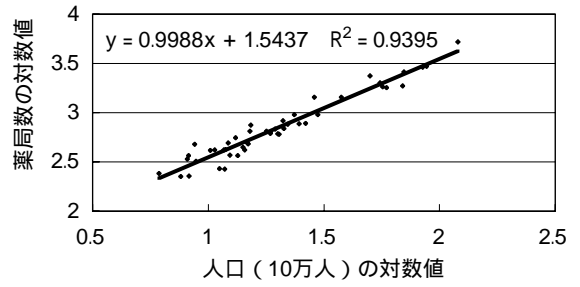
4.1 郵便局数の変異

郵便局数と人口との関係を実数値で表現すると図1のようになる。図中の点の分布で一つ大きく離れているのは北海道を示す点である。この図からは郵便局数と人口との傾向はつかみにくい。図3に示したように縦軸を人口あたりの郵便局数で表すと傾向がはっきりしてくる。人口あたりの郵便局数は人口と明らかに負の相関を示す。図3で直線回帰をすると傾きは、-0.2323となり、統計的に高度に有意に(t = 6.04)傾きは0とは異なる。これは図2に示すように両対数変換すると直線回帰の傾きが1より小さいことを意味する。

3.3で述べた手法に基づいて計算した回帰偏差値を表2に示す。偏差値が大きい方に1* 以上離れているのは、大きい方からいうと北海道、鹿児島、島根、長野、新潟、熊本の順である。逆に、偏差値が小さい方に1* 以上離れているのは、小さい方からいうと沖縄、埼玉、神奈川、滋賀、佐賀、

郵便局数	回帰偏差値	順位 (a)	人口当たりの値	人口当たり順位(b)	(a) - (b)	人口順位
北海道	75.57	1	27.29	20	-19	7
青森	47.20	29	24.53	26	3	28
岩手	57.05	9	31.50	8	1	30
宮城	45.58	33	19.45	32	1	15
秋田	56.93	10	33.73	5	5	35
山形	55.80	15	32.23	7	8	33
福島	56.81	11	26.52	24	-13	17
茨城	45.02	35	17.45	36	-1	11
栃木	39.48	40	18.00	35	5	20
群馬	37.44	41	17.09	38	3	19
埼玉	32.37	46	9.27	46	0	5
千葉	41.59	39	12.20	45	-6	6
東京	55.93	14	12.63	44	-30	1
神奈川	34.33	45	8.96	47	-2	3
新潟	62.20	5	28.31	17	-12	14
富山	46.25	31	26.85	21	10	38
石川	50.36	24	28.96	16	8	36
福井	45.89	32	30.16	12	20	43
山梨	47.93	28	30.74	9	19	41
長野	63.69	4	30.71	10	-6	16
岐阜	47.98	27	21.58	31	-4	18
静岡	45.52	34	16.06	40	-6	10
愛知	48.19	26	13.25	42	-16	4
三重	52.59	19	25.36	25	-6	23
滋賀	35.57	44	19.36	33	11	31
京都	44.59	36	18.15	34	2	13
大阪	50.76	22	12.82	43	-21	2
兵庫	55.80	16	17.42	37	-21	8
奈良	43.86	38	22.87	29	9	29
和歌山	49.61	25	29.63	13	12	39
鳥取	53.21	18	40.46	2	16	47
島根	65.80	3	49.67	1	2	46
岡山	56.65	12	27.37	19	-7	21
広島	58.77	7	24.52	27	-20	12
山口	52.51	20	27.42	18	2	25
徳島	44.47	37	29.25	15	22	44
香川	36.35	42	22.09	30	12	40
愛媛	51.03	21	26.72	22	-1	27
高知	57.98	8	40.29	3	5	45
福岡	50.69	23	16.13	39	-16	9
佐賀	36.00	43	23.38	28	15	42
長崎	55.52	17	29.53	14	3	26
熊本	60.38	6	30.50	11	-5	22
大分	56.37	13	32.92	6	7	34
宮崎	46.57	30	26.58	23	7	37
鹿児島	71.27	2	40.15	4	-2	24
沖縄	24.09	47	14.87	41	6	32





さに依存して起こっていることを示している。

これら一般旅券発行数の解析によって、人口に対して下に凸の規模の効果をもつような変数では、これを人口当たりの数値で相互評価を行えば、人口が大きい地域が過大評価され、人口が小さい地域が過小評価されることが明らかである。

5. 回帰偏差値の適用例

5.1 動物の脳化指数：脳重量と生体重との相関

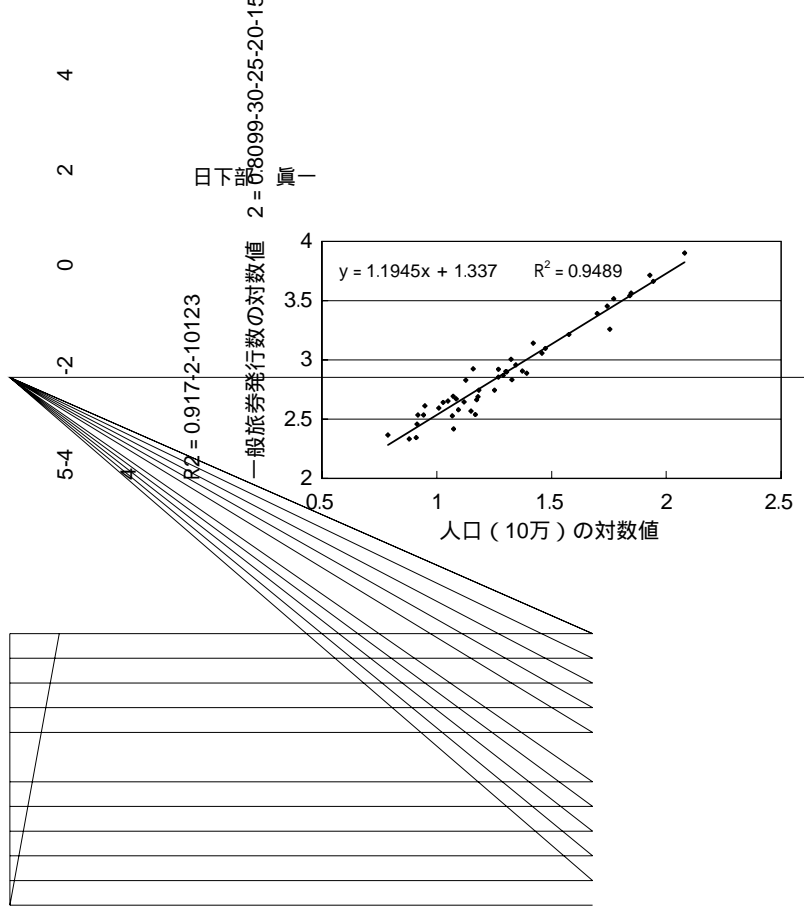
Weisberg (1985)が、かれの著書の回帰分析の残差分析で引用している動物の脳の重量と生体重の一覧表は動物における脳の進化を考える上で大変興味深い課題を提起している。このデータをもとに目的変数を脳重量、独立変数を生体重として両対数変換した値の回帰分析を行って図13に示す。生体重に対する脳重量の回帰偏差値を計算して表5に示す。これは、脳重量の実測値が、生体重で期待される脳の重量よりも大きいか小さいかの回帰偏差値を示す指数である。したがって、進化的観点から見れば、動物一般で期待されている以上に脳の発達が見られるかどうかの目安

表3 薬局数の回帰偏差値による順位と人口当たりの数値による順位との差

薬局数	回帰偏差値	順位 (a)	人口当たりの値	順位 (b)	(a) - (b)	人口順位
北海道	45.71	32	32.17	32	0	7
青森	47.29	29	33.13	29	0	28
岩手	44.22	34	31.29	34	0	30
宮城	57.84	11	40.34	11	0	15
秋田	50.58	22	35.24	22	0	35
山形	41.34	39	29.66	38	1	33
福島	45.80	30	32.20	31	-1	17
茨城	45.01	33	31.73	33	0	11
栃木	42.67	36	30.37	36	0	20
群馬	41.35	38	29.63	39	-1	19
埼玉	35.65	45	26.82	45	0	5
千葉	42.26	37	30.21	37	0	6
東京	62.50	6	43.24	6	0	1
神奈川	48.74	25	34.00	25	0	3
新潟	43.59	35	30.90	35	0	14
富山	30.11	46	24.09	46	0	38
石川	26.74	47	22.61	47	0	36
福井	36.39	44	27.14	43	1	43
山梨	51.43	20	35.81	20	0	41
長野	48.93	24	34.15	24	0	16
岐阜	56.41	13	39.28	13	0	18
静岡	54.40	16	37.78	16	0	10
愛知	52.77	17	36.58	17	0	4
三重	47.63	26	33.33	26	0	23
滋賀	36.52	43	27.10	44	-1	31
京都	41.05	40	29.46	40	0	13
大阪	47.53	27	33.28	27	0	2
兵庫	51.98	18	36.08	19	-1	8
奈良	39.97	42	28.90	42	0	29
和歌山	55.73	15	38.79	15	0	39
鳥取	56.10	14	38.99	14	0	47
島根	40.76	41	29.43	41	0	46
岡山	49.94	23	34.80	23	0	21
広島	69.13	2	49.77	2	0	12
山口	68.10	3	48.89	3	0	25
徳島	62.67	5	44.05	5	0	44
香川	58.03	10	40.47	10	0	40
愛媛	45.79	31	32.22	30	1	27
高知	59.31	9	41.40	9	0	45
福岡	66.44	4	47.09	4	0	9
佐賀	73.72	1	54.05	1	0	42
長崎	60.88	7	42.72	7	0	26
熊本	47.32	28	33.14	28	0	22
大分	57.55	12	40.13	12	0	34
宮崎	50.93	21	35.47	21	0	37
鹿児島	51.91	19	36.11	18	1	24
沖縄	60.03	8	42.03	8	0	32

になるのを“脳化指数”とも表現できる指数である。図13の回帰直線から上の方にはなれて見える点がヒトの脳重量・生体重の点であることは容易に理解できる。ヒトの偏差値は78であり、平均から2* (20)以上はなれている。偏差値で、60以上である11種の動物で8種類がサルの仲間である。これはサルの仲間では平均以上に加速していたこと示しており、進化学的に興味深い課題を提出する。

回帰偏差値 順位 (a)





と下に凸の变量との間で相関をとるとどのような变量間でも必然的に負の相関になってしまうのである。

公共事業費は図 15 で示したように人口に対して上に凸の規模の効果を示していた。筆者の指摘が正しいならば県民所得は人口に対して下に凸の規模の効果を示しているはずである。これを確認するために、県民所得と人口との関係を検討する。図 17 は実数での県民所得と人口との相関で、やや下に凸の傾向を示している。さらに、両対数変換した県民総所得と人口との相関をみると図 18 のようになり、回帰直線の傾きは 1.1251 で、統計的に有意に 1 より大きく、確かに下に凸の規模の効果を示していることがわかる。

筆者が指摘するように“規模の効果が上に凸の变量と下に凸の变量との間で相関をとると必然的に負の相関を示すことになる”

ことが示された。さらに、これが普遍的事実であることを示すために表 7 に、このような見かけ上の負の相関がおこる例を、今まで解析してきた例の中から示した。一人当たり県民所得と一人当たり老人ホーム数、一人当たり県民所得と一人当たり病院数数、一人当たり県民所得と一人当たりガソリンスタンド数、一人当たり県民所得と一人当たり郵便局数の 4 つの例で、すべて上に凸の規模の効果を示す量と下に凸の規模の効果を示す量の間には有意な負の相関が検出される。すべての場合で、相関に意味があるようには思えない組み合わせであるが、単純に統計的操作から見かけ上の負の相関が導かれる。この中で、一例として図 19 に一人当たり県民所得と一人当たり老人ホーム数の負の相関を図示する。

さて、それでは公共事業費と県民所得の関係を正しくはどのように理解すれば良いのであろうか？このためには、公共事業費の人口に対する回帰偏差値を考えたように、県民所得の人口に対する回帰偏差値を考えると良い。河野が指摘するように、もし、公共事業費の効果がなかったとすれば公共事業費の回帰偏差値と県民所得の回帰偏差値との間に相関がないか、負の相関が見られてもよいはずであろう。これを検討するために表 8 に県民所得の回帰偏差値を表し、図 20 に公共事業費の回帰偏差値と県民所得の回帰偏差値との相関を図示する。相関は弱く、回帰直線の傾きはやや正である。統計的検定では、この傾き(0.1364)は有意に 0 とは異なっている。したがって、公共事業費と県民所得の間には弱いながらも正の相関が検出されるのである。

公共事業費についての問題としては、朝日新聞や河野が主張するような“都市に厚く地方に薄い公共事業費が地方の資本形成にマイナスの効果をもたらしている”というような結論を導くことはできない。しかも、回帰偏差値の相関からは、有意な弱い正の相関が見られたことは、“公共事業費は、都市、地方を問わず全国に一樣に、資本形成にわずかなながらも県民所得を高める効果をもたらしている”という結論になる。

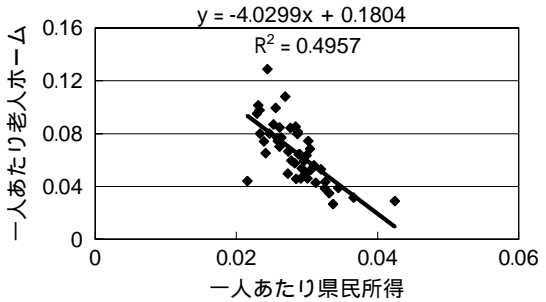


図 19 一人当たり県民所得と一人当たり老人ホーム数

表 7 みかけ上の負の相関がおこる場合

統計指標の組み合わせ	相関係数
一人あたり県民所得と一人あたり老人ホーム数	-0.7041
一人あたり県民所得と一人あたり病院数数	-0.5318
一人あたり県民所得と一人あたりガソリンスタンド数	-0.6125
一人あたり県民所得と一人あたり郵便局数	-0.6585

6 . 総合考察

本論の主眼は、考案した回帰偏差値の効用を確認し、その有効性を検討することにある。このため都道府県の統計指標から人口増加に対して正の規模の効果（効用逓減）を示す場合（郵便局数）規模の効果を示さない場合（薬局数）負の規模の効果（効用逓増）を示す場合（一般旅券発行数）の三つの例を具体例として用いて詳細な解析を行った。そして従来の人口当たりの値でもって相互評価を行う場合には、人口の大きい地域や小さい地域は、その規模の効果の大小・正負に応じて過大評価や過小評価されること、したがって従来の人口当たりの値でもって順位付けする場合は相対評価が誤ることを指摘した。

このような、いわゆる“規模の効果”と言われてきた二変量間の非線形性の度合いは、両対数変換を行って得た回帰直線の傾きの1からのずれで表される。この回帰直線の1からのずれの統計的有意性は通常のt検定で判断できる。表1に示したように都道府県の様々な指標値の場合には、この回帰直線の傾きの標準誤差は おおまかに0.04 0.06 である。都道府県の相対評価を行う場合は標本数が47であるから検定の自由度は45である。自由度45の5%有意水準を示すtの値は2.02であるから傾きが1.1か0.9ほどであれば統計的に有意な非線形性が認められることになる。このような非線形性は統計指標ではよく見られる現象で、本論で詳細に検討したように、従来、そして今でも慣用的・慣習的に使用されている人口当たりの値でもって順位付けすることは全く間違った順位付けを与える可能性が強いので地域相対評価

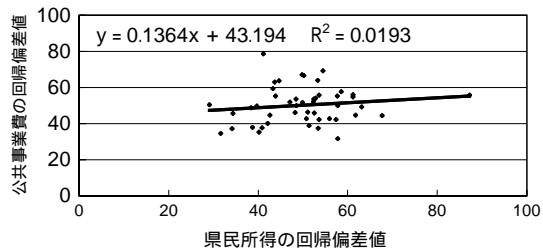


図20 県民所得と公共事業の回帰偏差値間の相関

表8 県民総所得の回帰偏差値と一人当たりの県民所得の順位差

	回帰偏差値	順位 (a)	あたり県民所得	順位 (b)	(a) - (b)	人口順位
北海道	32.92	45	27.38	29	16	7
青森	37.61	41	24.18	41	0	28
岩手	45.37	35	25.91	35	0	30
宮城	48.75	29	28.47	24	5	15
秋田	46.06	33	25.56	37	-4	35
山形	47.90	31	26.12	34	-3	33
福島	45.67	34	27.33	30	4	17
茨城	52.97	18	30.60	11	7	11
栃木	62.81	7	32.03	8	-1	20
群馬	57.60	10	30.47	12	-2	19
埼玉	50.74	26	33.18	5	21	5
千葉	50.77	25	32.56	7	18	6
東京	70.54	1	42.49	1	0	1
神奈川	49.94	27	33.75	4	23	3
新潟	51.94	23	29.61	17	6	14
富山	62.88	6	29.82	16	-10	38
石川	63.34	3	30.11	14	-11	36
福井	62.98	5	28.68	23	-18	43
山梨	63.27	4	29.00	20	-16	41
長野	54.84	14	29.99	15	-1	16
岐阜	52.55	19	29.17	19	0	18
静岡	52.30	21	31.27	9	12	10
愛知	61.19	8	36.65	2	6	4
三重	53.10	17	28.87	21	-4	23
滋賀	70.22	2	32.63	6	-4	31
京都	56.29	12	31.01	10	2	13
大阪	51.57	24	34.44	3	21	2
兵庫	40.62	37	29.32	18	19	8
奈良	54.33	15	28.32	26	-11	29
和歌山	40.54	38	23.92	42	-4	39
鳥取	56.08	13	25.87	36	-23	47
島根	47.31	32	24.45	40	-8	46
岡山	48.03	30	27.66	28	2	21
広島	52.21	22	30.23	13	9	12
山口	53.93	16	28.44	25	-9	25
徳島	56.42	11	26.97	31	-20	44
香川	60.13	9	28.72	22	-13	40
愛媛	39.76	39	24.72	39	0	27
高知	41.74	36	23.37	43	-7	45
福岡	36.48	42	27.81	27	15	9
佐賀	52.45	20	26.14	33	-13	42
長崎	33.44	44	23.29	44	0	26
熊本	39.45	40	25.29	38	2	22
大分	48.97	28	26.37	32	-4	34
宮崎	35.14	43	22.93	46	-3	37
鹿児島	30.75	46	23.11	45	1	24
沖縄	27.69	47	21.57	47	0	32

として順位付けを行う場合は十分注意しなければならない。

このような規模の効果に依存しないで相対評価を行うことが出来る指数として“回帰偏差値”を考案した。これは、二変量を両対数変換して実数で見られる規模の効果による非線形性をのぞき分散を安定化させ、対数変換値にひいた回帰直線からの標準化偏差 (Studentized residuals) でもって相対評価を行い、分布の中の順位付けを行う方法である。統計的に有意な、負や(一般旅券発行数)正の(郵便局数)規模の効果の場合に、従来の人口当たりの数値を用いてつけた順位と回帰偏差値による順位との間に、人口の大小に依存した変動が起こることが明白に示され、回帰偏差値の有効性がはっきりと示された。

この回帰偏差値は異質な属性(例えば、人口がちがうとか、GDPがちがうとか、体重がちがうなど)を持つ対象標本からの統計指標を相対評価することが出来るので大きな有効性・有用性を持っている。

まず、回帰偏差値は、回帰分析における残差分析として統計学的理論付けが十分行われている。通常の回帰分析(Draper and Smith 1966, Sokal and Rohlf 1973, Chatterjee and Price 1977, Dixon and Massey 1983, 広津 1992, 大友 1997)では、残差の分析・評価は回帰からはずれた外れ値を検出するのに用いられてきただけで、筆者が考案したような偏差値化して相対評価に用いる手法はいまだ例を見ない。標準化した偏差は標本数が多いと正規分布(0,1)にほぼ従う統計量とみなして良いので(Chatterjee and Price 1977)、原理としては、さまざまな試験などで従来用いられてきた正規分布に基づくZスコア(Z score)(eg Dixon and Massey 1983)を平均値50とした通常の偏差値と異なることはない。特に本論で考察したような、地域間の相対評価を考える際には有効かつ有力である。したがって、筆者が本論で考案した回帰偏差値の手法は、すべての回帰分析において適用することが可能である。本論では、一次回帰だけを考えたが、二重回帰でも回帰偏差値の考え方は有効である。

回帰直線を利用した方法は、回帰直線からはずれた2-3の異常点に大きく左右される傾向にあるから危険性が大きいという批判もあり得る。しかし、そのための残差を用いた残差分析であり、この残差を影響力係数を用いて有効利用した回帰偏差値なのである。従来の“人口当たり”という慣用的手法にしたがって“暗黙の内に”非線形効果をふくんだ相対評価を行って過大評価をしたり過小評価をして意味のない順位付けを行い都市対地域という歪んだ印象を与えていることの方が、回帰偏差値の危険性よりももっと危険である。

また、回帰偏差値は偏差値であるから、尺度フリーであり、したがって、回帰偏差値の平均値や総和として異質な指標を混みにした総合的な相対評価を行うことも可能である。これは、一般の学力についての正規分布に基づく偏差値と同様のことで、ある学生の個別科目の偏差値の総和をとって全体の学力とみなす方法と同じことが可能である。しかし、従来の“人口当たりの数値”では、異質な指標間では規模の効果がそれぞれに異なるであろうから、回帰偏差値で可能であるような総和をもって総合評価とするようなことは出来ないのである。

本論では、従来の慣用的に用いられている“人口当たりの数値”は、人口の小さい地域や大きい地域を過大評価したり過小評価したりする危険性があることを指摘した。そしてこの危険性をのぞき、相対評価を可能にする回帰偏差値を提案したのである。今まで、慣用的に用いられてきた“一人当たり”の指標値に替え、ここで詳述してきた“回帰偏差値”を用いることによって、より客観的な相対評価が可能となることが望ましい。そして、“都市対地域”というわかりやすい図式に十分気をつけて正当な相対評価を行わねばならない。

参考文献

- Chatterjee,S., Price,B. (1977) *Regression Analysis by Example*. John Wiley & Sons, Inc. (佐和隆光・加納悟 訳 (1981) 『回帰分析の実際』新曜社)
- Cook,R.D. and Weisberg,S. (1994) *An introduction to regression graphics*. John Wiley & Sons, Inc. pp. 203-216.
- Cook,R.D. and Weisberg,S. (1999) *Applied regression including computing and graphics*. John Wiley & Sons, Inc. pp. 354-372.
- Dixon, W.J., Massey,F.J. (1983) *Introduction to Statistical Analysis*, 4th ed. McGraw-Hill.
- Draper,N.R., Smith,H. (1998) *Applied Regression Analysis*, 3rd ed. John Wiley & Sons, Inc. (中村慶一訳 (1968) 『応用回帰分析』森北出版)
- 広津千尋 (1992) 『実験データの解析 - 分散分析を超えて - 』共立出版 .
- 河野龍太郎 (2001) 90年代の公共投資は景気対策ではなかった!、『金融ビジネス』JUN. 2001、76-80 .
- 日下部眞一 (2002) NPOの規模をはかる回帰偏差値、“NPO指数”の考案：NPO指数を通して見えてきた地域格差、日本NPO学会機関誌『ノン・プロフィットレビュー』(投稿中)
- Myers, R.H. (1990) *Classical and modern regression with applications*, 2nd ed. PWS-KENT Publishing Company, Boston.
- 大友篤 (1997) 『地域分析入門(改訂版)』東洋経済新報社 .
- 総務省統計局 (2002) 『統計でみる県のすがた』(財)日本統計協会 .
- Sokal,R.R. , Rohlf,F.J. (1973) *Introduction to Biostatistics*. Freeman, New York. (藤井宏一訳(1983)『生物統計学』共立出版)
- Snedecor,G.W., Cochran,W.G. (1980) *Statistical Methods*, 7th ed. The Iowa State University Press, Iowa. (奥野ら共訳(1972)『統計的方法原書第6版』岩波書店)
- Weisberg,S. (1985) *Applied Linear Regression*, 2nd ed. John Wiley & Sons, Inc.