不動産鑑定における Excel の活用
ーExcel を用いた統計分析ー

不動産鑑定士　堀田　勝己

※本稿は、(株)住宅新報出版より発行の『月刊 不動産鑑定』2021 年 7 月号に掲載された論文である。

１．はじめに

鑑定評価の要諦は説明責任（accountability）を果たすことにあると筆者は考える。ゆえに，収集したデータから傾向を読み取り，そこから得た結論につき，判断根拠（evidence）を明確に示すことが必要である。

無数に存在する情報からたった一つの価格という結論にたどり着く帰納的なプロセスが鑑定評価だとすると，統計分析は必須のツールであると言える。

Excel を使いこなせば，かなりの省力化ができる。本稿では，統計学の基礎を踏まえた上で，Excel を使ったいくつかの分析方法について述べる。

２．統計用語の基礎と Excel 関数

（１）基本統計量ー代表値

①平均値

〇算術平均（相加平均）・・・arithmetic mean

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

全データを合計してデータ数で割ったもの。

（例）a，b，c，d の 4 人の所持金がそれぞれ 1 万円，2 万円，2 万円，7 万円である場合の平均所持金額を求める[表 1]。

$$\mu = \frac{1+2+2+7}{4} = 3 \quad （単位万円）$$

[表 1]

Excel 関数	入力方法	計算結果（答）
AVERAGE	=AVERAGE(1,2,2,7)	3　(答)3 万円

なお，[表 2]のようにセル範囲を指定して計算結果を得ることもできる。これは，次項以降の他の関数でも基本的に同じ使い方ができる。

[表 2]

	A	B	C
1	a 所持金	1	
2	b 所持金	2	
3	c 所持金	2	
4	d 所持金	7	
5	平均	3	
6			

B5 セルに=AVERAGE(B1:B4)と入力した結果，3 が表示された。

=AVERAGE(B1:B4)と入力

○幾何平均（相乗平均）・・・geometric mean

$$\mu_G = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}} = \sqrt[n]{x_1 x_2 \cdots x_n}$$

全データを相乗して累乗根をとったもの。

（例）過去 3 年間の対前年変動率が▲10%，0%，+10%であった場合に，平均変動率を求める[表 3]。

$$\mu_G = \sqrt[3]{(1-0.1)(1+0.1)} \approx 0.9966 \quad \rightarrow ▲\ 0.0033$$

[表 3]

Excel 関数	入力方法	計算結果（答）
GEOMEAN	=GEOMEAN(0.9,1,1.1)	0.9966　(答)▲0.33%

○調和平均・・・harmonic mean

$$\mu_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

各データ値の逆数の算術平均値の逆数。

（例）片道 20km の道程を往路時速 20km，復路時速 5km で往復した場合の平均時速を求める[表 4]。

往復 40km
片道 1 として往復 2

$$\mu_H = \frac{40}{\frac{20}{20} + \frac{20}{5}} = \frac{2}{\frac{1}{20} + \frac{1}{5}} = 8$$

（単位 km/h）

速度 5km/h
速度 20km/h

[表 4]

Excel 関数	入力方法	計算結果（答）
HARMEAN	=HARMEAN(20,5)	8　(答)平均 8km/h

②中央値・・・median

全データのちょうど真ん中の値。

（例）a，b，c，d，e の 5 人の年収額がそれぞれ 300 万円，300 万円，400 万円，1000 万円，1 億円であるときの中央値（5 人のうちの 3 番目）を求める[表 5]。

[表 5]

Excel 関数	入力方法	計算結果（答）
MEDIAN	=MEDIAN(300,300,400,1000,10000)	400　(答)400 万円

③最頻値・・・mode
全データの中で最も頻繁に出現する値。

（例）a，b，c，d，e の 5 人の年収額がそれぞれ 300 万円，300 万円，400 万円，1000 万円，1 億円であるときの最頻値を求める[表 6]。

[表 6]

Excel 関数	入力方法	計算結果（答）
MODE	=MODE(300,300,400,1000,10000)	300　(答)300 万円

（2）基本統計量－散布度
①分散・・・variance

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2}{n}$$

偏差平方の平均値（各データと平均値との差を 2 乗した値の平均値）。

（例）4 人の所持金がそれぞれ 1 千円，5 千円，20 千円，50 千円であるとき，分散(*1)を求める[表 7]。

$$\sigma^2 = \frac{(1-\mu)^2+(5-\mu)^2+(20-\mu)^2+(50-\mu)^2}{4}$$

ここで,

$$\mu = \frac{1+5+20+50}{4} = 19 \quad （単位千円）$$

であるから,

$$\sigma^2 = \frac{(-18)^2+(-14)^2+1^2+31^2}{4} = 370.5$$

となる。

[表 7]

Excel 関数	入力方法	計算結果（答）
VARP	=VARP(1,5,20,50)	370.5　(答)分散 370.5×10^6

②標準偏差・・・standard deviation

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2} = \sqrt{\frac{(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2}{n}}$$

分散はデータを 2 乗するので単位がなく，その数値自体に意味がない。そこで，分散の平方根をとり元の単位に戻したものを標準偏差という。上記①と同じデータを用いて標準偏差(*2)を求める[表 8]。

$$\sigma = \sqrt{370.5} \approx 19.248$$

[表 8]

Excel 関数	入力方法	計算結果（答）
STDEVP	=STDEVP(1,5,20,50)	19.248　(答)標準偏差 19,248 円

３．Excel 分析ツール

Excel には「分析ツール」が標準添付されている。但し，通常インストールしただけでは使える状態になっていないので，オプションメニューから Excel アドインの管理画面で「分析ツール」を有効にする必要がある。

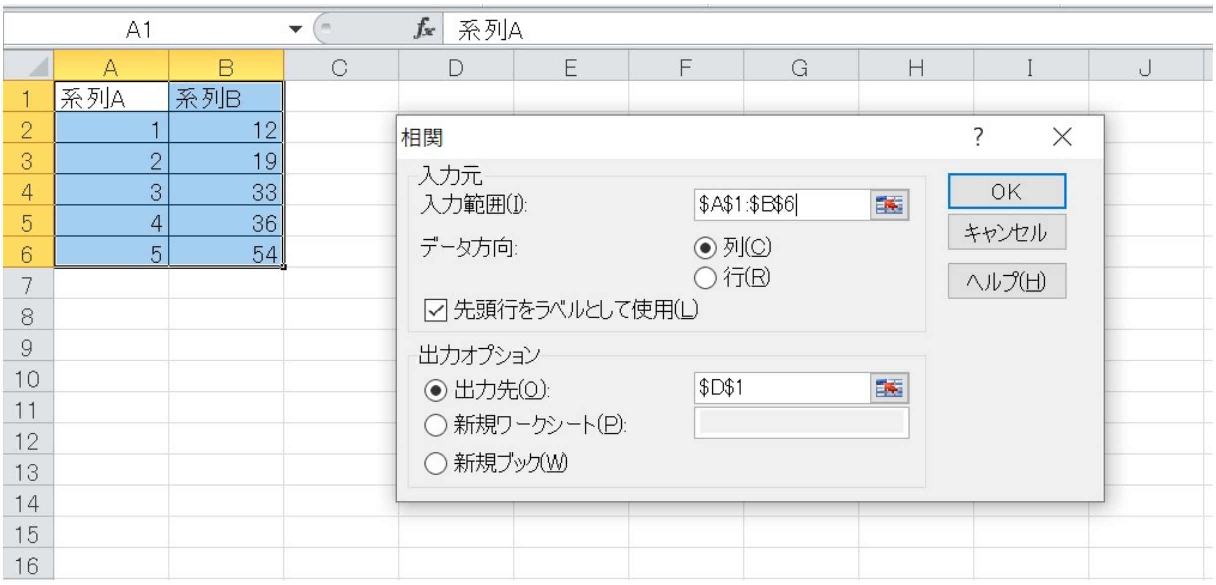
「分析ツール」には，次のようなツールが用意されており，鑑定評価における価格形成要因分析等に活用できる。

・相関
・共分散
・基本統計量
・回帰分析
・t 検定　など

４．相関分析

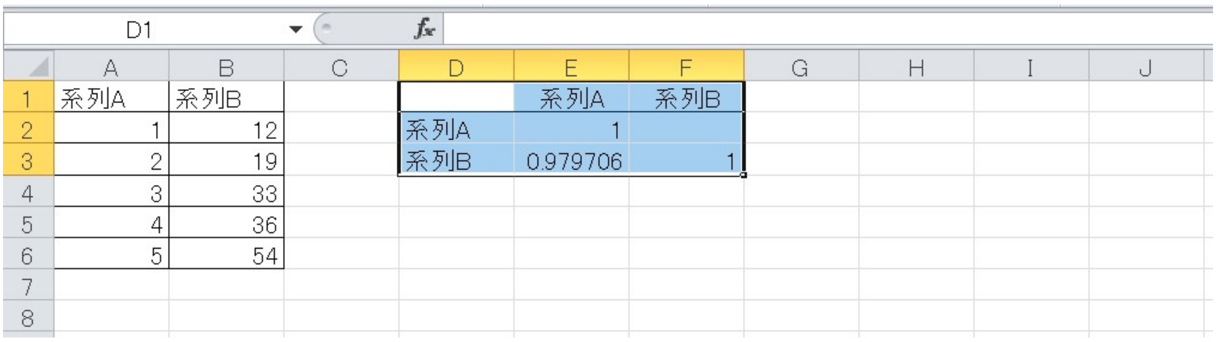
異なる 2 つのデータ系列間の関連性の強さ（線形関係の有無）(*3)を分析する方法。[図 1]は，A，B 各系列のデータ間に相関関係があるかどうかをみるために，Excel 分析ツールを利用した例である。

[図 1]



入力範囲，出力先等を指定して OK を押下するだけで，[図 2]のような結果が表示される。

[図 2]



両系列の相関係数は 0.979706 である。相関係数(r)は-1 から 1 の間の値（$-1.0 \leq r \leq 1.0$）をとり，+1 に近いほど強い正の相関が認められる（-1 に近いほど強い負の相関，0 は無相関）。なお，単に相関係数を求めるだけなら CORREL 関数(*4)が使える（=CORREL(A2:A6,B2:B6)と入力すれば 0.979706 が返される）。

本例では，説明のために単純な整数で行っているが，駅からの距離と地価，マンションの築年数と坪単価など，不動産価格に係る実際の数値を代入すれば，各価格形成要因と価格との間の相関関係や，価格形成要因相互間の相関関係の強弱を調べることができる。但し，不動産価格は多くの要因が作用して形成されているため，各条件がバラバラであると，価格との関係を見出すことは難しい。それゆえ，価格構造を解明するために，次に述べる回帰分析の手法等が用いられる。

５．回帰分析

ある変量(Y)を，他の変量(X)で説明するために，$Y = f(X)$というモデルをあてはめるもの(*5)。Yを従属変数（被説明変数），Xを独立変数（説明変数）という。最もシンプルな$Y = aX + b$という式を採用するものを単回帰分析といい，$Y = a_1X_1 + a_2X_2 + \cdots + b$と複数の独立変数を採用するものを重回帰分析という。$a_1, a_2 \cdots$は各独立変数の係数（偏回帰係数）であり，bは定数項（切片）である。この回帰分析の手法を用いて，資産価格等の成り立ちを解明し，独立変数それぞれの限界評価額（従属変数への影響の度合い）を推定する方法をヘドニック・アプローチと呼ぶ。

（１）Excel による回帰分析例～地価関数の推定

最も単純な例として，線形回帰(*6)による地価関数の推定例を紹介する。

Excel 分析ツールを利用した回帰分析では，変数は 16 個までという制約があるため，複雑な分析には適さない(*7)ものの，たとえば同一駅勢圏における住宅地価の推定等には十分に利用可能である。

[表 9]

地価(/㎡)	駅距離(m)	商店街距離(m)	幅員(m)	地積(㎡)	容積率(%)
319,810	250	200	10	250	200
310,440	100	300	6	120	200
308,000	150	90	5	88	200
300,839	300	280	5	75	200
289,920	200	380	6	320	200
288,290	550	490	8	120	100
284,900	350	330	6	290	200
283,225	800	1050	6	102	150
281,922	500	450	6	135	100
279,000	450	580	4	182	160
277,490	650	780	6	78	100
269,404	700	650	8	230	100
268,550	900	1080	9	255	200
266,493	750	880	7	198	150
260,600	950	900	4	132	160
259,900	850	750	6	103	200
259,023	600	630	4	330	160
257,520	1050	1020	5	180	200
257,448	1000	900	4	200	160
232,029	400	370	3	380	160

ここでは[表 9]のデータを題材(*8)に，分析をおこなう。時点は同一のクロスセクションデータ(*9)である。実際には，変数選択も回帰式の設定も一意に定まるようなものではなく，試行錯誤によってあてはまりを見ながら探索し，最良のものを選択する。

まず，独立変数間の相関関係を確かめるため，相関分析を行った結果が，[表 10]（相関行列）である。

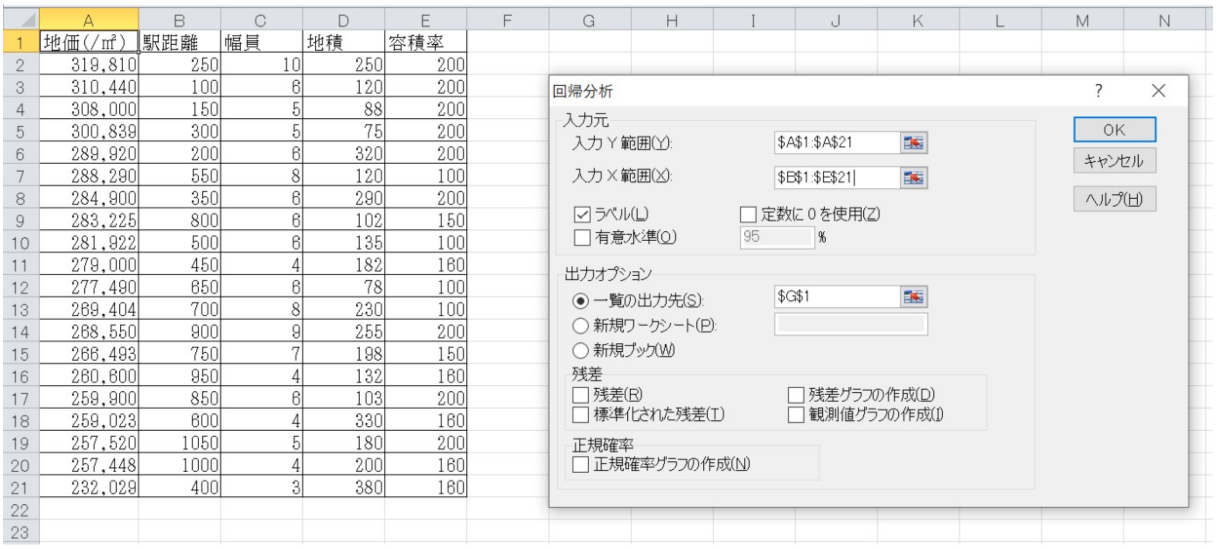
[表 10]

	地価(/㎡)	駅距離	商店街距離	幅員	地積	容積率
地価(/㎡)	1					
駅距離	-0.67907	1				
商店街距離	-0.59585	0.929869	1			
幅員	0.471397	-0.05919	-0.01049	1		
地積	-0.38435	-0.10132	-0.09262	-0.04173	1	
容積率	0.233464	-0.2343	-0.20099	-0.06028	0.17209	1

駅距離と商店街距離の間に強い正の相関（0.929869）があるので，多重共線性(*10)を回避するため，いずれかの変数を取り除いた上で回帰分析を実行する必要がある。後者を取り除き，駅距離を採用して分析をおこなう。

[図 3]の「入力 Y 範囲」には従属変数が入力されているセル範囲（$A$1:$A$21），「入力 X 範囲」には独立変数が入力されているセル範囲（$B$1:$E$21）を指定する。最上行に数値ではなくデータラベル（項目名）が入力されているので，「ラベル」のチェックボックスにチェックを入れる。

[図 3]



	A	B	C	D	E
1	地価(/㎡)	駅距離	幅員	地積	容積率
2	319,810	250	10	250	200
3	310,440	100	6	120	200
4	308,000	150	5	88	200
5	300,839	300	5	75	200
6	289,920	200	6	320	200
7	288,290	550	8	120	100
8	284,900	350	6	290	200
9	283,225	800	6	102	150
10	281,922	500	6	135	100
11	279,000	450	4	182	160
12	277,490	650	6	78	100
13	269,404	700	8	230	100
14	268,550	900	9	255	200
15	266,493	750	7	198	150
16	260,600	950	4	132	160
17	259,900	850	6	103	200
18	259,023	600	4	330	160
19	257,520	1050	5	180	200
20	257,448	1000	4	200	160
21	232,029	400	3	380	160
22					
23					

OK を押下すると分析が実行され，[表 11]のような結果が表示される。

回帰式の良否は，一般に決定係数（[表 11]では「重決定 R2」と表示）の大きさで判断され，1 に近いほどあてはまりが良いとされる。また，この決定係数は独立変数を増やすだけで大となるため，それに左右されない指標として自由度調整済決定係数（[表 11]では「補正 R2」と表示）が用いられる。

採用した独立変数が，採用に値するものであるかどうかをみる簡便な指標として，t 値（[表 11]の「t」）に着目し，これがおおむね絶対値 2 以上であれば統計的に有意と判断される。

[表 11]

概要

回帰統計	
重相関 R	0.931977
重決定 R2	0.868581
補正 R2	0.833535
標準誤差	8781.359
観測数	20

分散分析表

	自由度	変動	分散	観測された分散比	有意 F
回帰	4	7.64E+09	1.91E+09	24.7846	1.84E-06
残差	15	1.16E+09	77112271		
合計	19	8.8E+09			

	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
切片	278983.7	13560.94	20.5726	2.1E-12	250079.3	307888.2	250079.3	307888.2
駅距離	-47.8581	7.041114	-6.79695	6.02E-06	-62.8659	-32.8503	-62.8659	-32.8503
幅員	5062.428	1123.044	4.507775	0.000417	2668.718	7456.139	2668.718	7456.139
地積	-109.098	22.34804	-4.88177	0.000199	-156.732	-61.4643	-156.732	-61.4643
容積率	102.7914	54.27931	1.893749	0.077709	-12.9022	218.485	-12.9022	218.485

当分析によって，次の回帰式が求められた。

地価(/㎡)=278,983.7+駅距離×(-47.8581)+幅員×5,062.428+地積×(-109.098)+容積率×102.7914

当分析では，駅距離と地積は地価に対して負の影響を，幅員と容積率は地価に対して正の影響を与えていることがわかる。

誤解して頂きたくないのは，このようなモデルを組み立てれば，自動的に評価額が算出されると主張したいわけではないということである。価格形成要因は不動産ごとに異なり，時期によっても異なり，また，その影響の程度も異なる。計量化できない要因も多数ある。そもそも採用できるデータ数にも限りがある。隠れた要因を探り出し，各要因の影響度の強弱をデータから測定すれば，鑑定評価プロセスの可視化や精度向上が期待できるということに尽きる。

（2）Excel による回帰分析～応用編

上記（1）の分析例では，従属変数である地価を，複数の独立変数（価格形成要因）の和として表現する式となっているため，理論上，従属変数は負の値も取り得る。これを回避するために，次のような工夫が考えられる。

収益物件の利回り（キャップレート$R = y$）を推定するため，例えば築年数(X_1)，階層(X_2)，駅距離(X_3)を独立変数として採用する際に，次のような式（lnは自然対数）を想定する。

[1 式]　$ln(y) = ln(a_1)X_1 + ln(a_2)X_2 + ln(a_3)X_3 + ln(b)$

ここで，$ln(y) = Y$，$ln(a_1) = A_1, ln(a_2) = A_2, ln(a_3) = A_3, ln(b) = B$と書き直せば，上記[1 式]は，
[2 式]　$Y = A_1X_1 + A_2X_2 + A_3X_3 + B$

という通常の回帰式であることがわかる。
Excel にて通常の回帰分析を実行後，真数に戻せば，[1 式]は，

[3 式]　$y = {a_1}^{X_1} \cdot {a_2}^{X_2} \cdot {a_3}^{X_3} \cdot b$　（ただし，$a_i = exp(A_i), b = exp(B)$）

という相乗式となり，従属変数たる$y = R$（キャップレート）の非負が保証される。

6．おわりに

統計で鑑定評価などできないという意見（批判？）を業界内から頂くことがある。鑑定は算定ではないということなのであろうが，そこには大きな誤解がある。一方，ビッグデータの整備が進めば，不動産鑑定士など不要であるとの声も聞こえてくるが，そこにもまた誤解がある。
まず，統計で鑑定評価ができないのは当然である。それに対して異論はない。
筆者が考えるに，鑑定評価は，①必要なデータを収集し，できうる限り客観的に分析して法則性をさぐり，②専門知識に基づき鑑定評価方式を適用して，最終判断に到達する。その 2 段階のプロセスに大きく分けることができるが，鑑定評価の本質は，②のプロセス，すなわち知識と経験に裏打ちされた「判断」にある。
統計学が重要なツール（道具）であることは間違いないが，道具は単なる道具であって，それ以上でも以下でもない。よって，鑑定評価の本質を知らず，数的処理だけに長けた人たちに算定はできても，鑑定はできない(*11)。鑑定評価のプロセスの中に統計分析は当然にあるが，それは一部に過ぎないということを，われわれ不動産鑑定士はもっと発信していかなくてはならないだろう。

以上

（注）
*1：この 4 つのデータを，母集団から抽出した標本（sample）だとすると，ここで求めている分散は標本分散と呼ばれるものである。これらから母集団の分散（母分散）を推定する場合の最良の推定値を不偏分散(v^2)といい，次の式で表される。また，この場合の Excel 関数は “VAR”または“VAR.S”である。

$$v^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2 = \frac{(x_1-\mu)^2 + (x_2-\mu)^2 + \cdots + (x_n-\mu)^2}{n-1}$$

*2：分散と同様，ここで求めているのは標本標準偏差であり，母集団の標準偏差を推定する場合の Excel 関数は“STDEV”または“STDEV.S”である。
*3：相関は，データ系列間に線形（1 次関数）関係があるかどうかをみるものであり，非線形の関係がある場合には，相関係数の絶対値は小さくなる。一般的に，相関係数を算出する前に散布図（scatter plot）を作成して相関関係の有無を視覚的にチェックすることがおこなわれるが，本稿では紙幅の関係で省略した。なお，相関関係が認められるからといって，因果関係があるとは限らないので注意する。
*4：相関係数rは，系列Aの標準偏差をσ_A，系列Bの標準偏差をσ_B，両系列の共分散をσ_{AB}とすると，$r = \sigma_{AB}/\sigma_A\sigma_B$である。それゆえ，CORREL 関数の代わりに COVAR 関数（または COVARIANCE.P 関数）と STDEVP 関数を用いて，=COVAR(A2:A6,B2:B6)/STDEVP(A2:A6)/STDEVP(B2:B6)と入力しても同じ結果が得られる。
*5：回帰分析によって唯一最良の式が導出されるのは，$Y = f(X) + \varepsilon$（ただし，εは実測値とモデル関数との誤差）としたとき，εの二乗和が最小となるような$f(X)$が選択されるからであり，これを最小二乗法という。
*6：重回帰式$Y = a_1X_1 + a_2X_2 + \cdots + b$からわかるとおり，被説明変数$Y$は各独立変数の係数$a_1, a_2 \cdots$に対して線形関係であるため線形回帰という。
*7：ソルバーやピボットテーブルの機能を使えば，Excel でもさらに制約の少ない分析ができるが，本稿の範囲を超えるため，別稿に譲る。

*8：このような地価推定が，限られたデータ（標本）から母集団を推定する作業であるとすると，標本の選択は推測統計の大原則である無作為抽出によらなければならない。好ましい結果を得るために，意図的に標本を選択（統計的な外れ値は除く）するような行為は，捏造という。なお，本例はあくまでも分析例であって，実際の標本数は少なくとも 30 以上はほしい。
*9：同一時点でデータを切り取っておこなうものをクロスセクション分析といい，異なる時点のものを混ぜておこないたい場合には，独立変数に時間ダミーを投入しておこなう。
*10：独立変数間に強い相関（線形関係／一次従属の関係）があると，偏回帰係数が求められない，回帰式自体が求められない等の現象が起こることがある。これを，多重共線性（multicollinearity : マルチコ）という。
*11：たとえば駅距離が地価に与える負の影響は，一般に逓減するため対数関数のあてはまりが良いが，住宅地よりも商業地のほうがグラフの傾斜は急であることが多い。前面道路幅員が地価に与える影響は 4m, 6m, 12m で断裂が生じる可能性があるため，そこで式を分けたほうがあてはまりが良くなる場合がある。容積率を独立変数に採用すると，高級住宅地では負の影響が観測されることがあるなど，不動産に関する知識（法的知識を含む）や経験がなければ理解できない事柄は多くある。