

# 賃貸住宅情報誌を活用した統計教育の実践Ⅱ

飯田洋市

諏訪東京理科大学 経営情報学部

## 1. はじめに

社会で役立つ数学として、社会人が真っ先に挙げるのが「統計」であろう。ところで統計には大きく分けて「記述統計」と「推測統計」がある。このうち得られたデータの平均値を計算するような記述統計は、普段の生活の中で利用されていることから馴染みやすく、理解も比較的容易である。[1]などが指摘するような誤用には配慮がほしいが、それらは知識さえあれば防げる。一方、推測統計はスキルを習得し、それらの意義を理解するためにはある程度の学習が必須である。本稿では社会科学系分野で扱われる推測統計のひとつ「(単)回帰分析」を扱う。なお、本稿は第11回年会誌(2007年)に掲載した原稿のリメイク版であり、その先を含んだ構成になっている。

## 2. 回帰分析の手順

本稿で扱う回帰分析の手順は次のようになる: 散布図の作成→回帰式  $y=ax+b$  の算出→相関係数  $r$ , 決定係数(寄与率)  $r^2$  の算出→分散分析→残差の検討(回帰診断)→具体的な  $x$  に対する  $y$  の推計

## 3. 本稿で扱う問題とデータの入力

市販の賃貸住宅にある物件の家賃を予測する。家賃を決定する要因は「専有面積」「駅からの距離」「築年数」などがあげられる。ここでは「専有面積」が大きくなれば「家賃」は高くなるという基本原則を考慮した物件の家賃予測を行う。更にある価格で賃貸したい場合の専有面積について検討する。

表1 賃貸住宅情報

物件番号	専有面積 ( $m^2$ )	家賃 (万円)	物件番号	専有面積 ( $m^2$ )	家賃 (万円)
1	16.65	6.2	9	27.8	9.8
2	17.49	6.8	10	32.9	10.0
3	18.28	7.0	11	32.99	10.5
4	20.46	7.2	12	24.25	10.5
5	20.25	7.8	13	35.1	11.0
6	18.89	8.2	14	40.96	12.6
7	40.0	9.4	15	65.01	15.0
8	32.99	9.8	16	47.63	16.5

まず表1の「専有面積」と「家賃」のデータをグラフ電卓に入力する(データは実際の賃貸住宅情報誌をもとに若干の変更を加えている)。L<sub>1</sub>に専有面積, L<sub>2</sub>に家賃を入力する(入力する位置は、目的に合わせて変更する)。まず **STAT** キーを押す。

```

EDIT CALC TESTS
1:Edit...
2:SortA(
3:SortD(
4:ClrList
5:SetUpEditor

```

Å を押す

L1	L2	L3	1
██████	-----	-----	
L1(1) =			

データを入力する

L1	L2	L3	2
32.99	10.5		
24.25	10.5		
35.1	11		
40.96	12.6		
65.01	15		
47.63	16.5		
-----	██████		
L2(17) =			

#### 4. 散布図の作成と回帰直線の描画

入力したデータをもとに、散布図を作成する。まずy [STAT PLOT]を押す。

```

STAT PLOTS
1:Plot1...On
  L1 RESID
2:Plot2...Off
  L1 L2
3:Plot3...Off
  L1 L2
4↓PlotsOff

```

Å を押す

```

Plot1 Plot2 Plot3
Off Off Off
Type: [Type icons]
Xlist:L1
Ylist:RESID
Mark: [Mark icons]

```

設定を上右図のように変える

```

Plot1 Plot2 Plot3
On Off Off
Type: [Type icons]
Xlist:L1
Ylist:L2
Mark: [Mark icons]

```

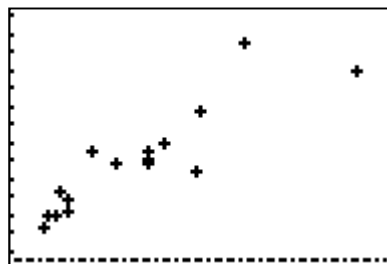
ZOOM を押す

```

ZOOM MEMORY
1:ZBox
2:Zoom In
3:Zoom Out
4:ZDecimal
5:ZSquare
6:ZStandard
7↓ZTrig

```

} } Í を押す



「専有面積」と「家賃」が直線関係にあることを確認できる。統計量が等しくても全く異質なデータ群は多々存在する。単回帰分析では、このような関係を散布図で確かめることが重要である。ここでグラフ電卓を利用して、最小2乗法で（近似）直線の式を求める。ところで「専有面積」と「家賃」の関係の強さを表す指標として相関係数  $r$  がある。相関係数は-1 から 1 までの値をとり、絶対値が大きいほど関係が強いことを意味する。なお相関係数  $r$  が正の値であれば正の相関があるといい、負の場合は負の相関があるという。相関係数を2乗した  $r^2$  を決定係数（寄与率）と呼ぶ。相関係数と決定係数を出力するために次の操作を行う。y [CATALOG]を押す。

```

CATALOG
abs(
and
angle(
ANOVA(
Ans
Archive
Asm(

```

- ,† ×9 を押す（右図参照）

```

CATALOG
Degree
DelVar
DependAsk
DependAuto
det(
DiagnosticOff
DiagnosticOn

```

Í を押す

```

DiagnosticOn

```

Í を押す

```
DiagnosticOn
Done
```

```
EDIT CALC TESTS
1:Edit...
2:SortA(
3:SortD(
4:ClrList
5:SetUpEditor
```

```
EDIT CALC TESTS
1:1-Var Stats
2:2-Var Stats
3:Med-Med
4:LinReg(ax+b)
5:QuadReg
6:CubicReg
7↓QuartReg
```

[STAT]を押す

~を押す

⇩を押す

```
DiagnosticOn
Done
LinReg(ax+b)
```

```
DiagnosticOn
Done
LinReg(ax+b) L1,
L2,
```

```
VARs Y-VARS
1:Window...
2:Zoom...
3:GDB...
4:Picture...
5:Statistics...
6:Table...
7:String...
```

y À c y Á c と押す

[VARs]を押す

~を押す

```
VARs Y-VARS
1:Function...
2:Parametric...
3:Polar...
4:On/Off...
```

```
FUNCTION
1:Y1
2:Y2
3:Y3
4:Y4
5:Y5
6:Y6
7↓Y7
```

```
DiagnosticOn
Done
LinReg(ax+b) L1,
L2, Y1
```

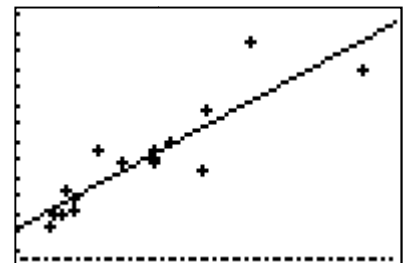
Àを押す

Àを押す

Í を押す

```
LinReg
y=ax+b
a=.1949465557
b=3.905235493
r²=.7954905234
r=.8919027544
```

```
Plot2 Plot3
Y1 19494655565
X+3.9052354933
47
Y2=
Y3=
Y4=
Y5=
```



[Y=]を押す

[GRAPH]を押す

y z を押す

決定係数  $r^2$  を値を確認し、直線のあてはまりの良さを検討する。ところで回帰式は直線であるから、 $x=0$  や  $y=0$  などに対応する値が計算上は得られる。しかしこれらは意味を持たない。この意味で、説明変数  $x$  は元のデータの最小値と最大値の間の領域が意味を持つ。この領域内で予測することを内挿とよび、それ以外での予測を外挿とよぶ。外挿では十分な注意が必要である。

### 5. 単回帰分析の分散分析

ここでは分散分析表の一般形と分散分析表を作成するための数式を与えた上で、それらを用いた回帰直線の妥当性の検定を行う（例えば[2]を参照のこと）。ここで  $y_i$  の真値を  $\hat{y}_i$  と書く。

$$S_{xx} = \sum (x_i - \bar{x})^2, \quad S_{yy} = \sum (y_i - \bar{y})^2, \quad S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

$$S_T = S_{yy} = \sum (y_i - \bar{y})^2 = \left( \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2 = S_e + S_R \right), \quad S_R = \frac{S_{xy}^2}{S_{xx}}, \quad S_e = S_T - S_R$$

$S_T$ は総平方和（自由度  $n-1$ ）、 $S_R$ は回帰平方和（自由度 1）、 $S_e$ は残差平方和（自由度  $n-2$ ）と呼ばれ、 $S_R$ は直線で近似した場合のデータのバラつき具合、 $S_e$ は誤差によるバラつき具合を表わす。残差平方和は  $S_e=S_T-S_R$  より求める。回帰による分散  $V_R=S_R/1=S_R$  と残差に関する分散  $V_e=S_e/(n-2)$  から、 $F$  検定のための分散比は  $F_0=V_R/V_e$  を得る。これらをまとめ分散分析表を得る（表 2）。

表 2 単回帰分析の分散分析表

要因	平方和	自由度	分散	分散比
回帰	$S_R$	1	$V_R$	$F_0$
残差	$S_e$	$n-2$	$V_e$	
総和	$S_T$	$n-1$		

ここで分散分析，すなわち  $F$  表から  $F(1, n-2; 0.05)$  を求め  $F$  検定を行う。帰無仮説「直線関係がない（=等分散）」に対し、 $F_0 > F(1, n-2; 0.05)$  であれば帰無仮説を棄却（有意，直線関係がある）とし、 $F_0 \leq F(1, n-2; 0.05)$  であれば帰無仮説を採択する（有意でない，直線関係がない）。ここでは有意水準を 95% としている。ところで以下の式より，総平方和における回帰平方和の割合が寄与率に一致することが分かる。回帰平方和は直線関係によるバラつきだから，寄与率  $r^2$  の意味が実感できる。

$$\frac{S_R}{S_T} = \frac{S_{xy}^2}{S_{xx} S_{yy}} \times \frac{1}{S_{yy}} = \left( \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \right)^2 = r^2$$

今の問題では表 3 となる。 $F_0=54.31 > F(1, 14; 0.05)=4.60$ ※注より「専有面積」と「家賃」は直線関係があるといえる。

表 3 表 1 から得られる分散分析表

要因	平方和	自由度	分散	分散比
回帰	99.68	1	99.68	54.31
残差	25.69	14	1.84	
総和	125.37	15		

これらの結果をグラフ電卓で確かめる。まず **STAT** キーを押す。

```

EDIT CALC TESTS
1:Edit...
2:SortA(
3:SortD(
4:ClrList
5:SetUpEditor
    
```

～ を押す

```

EDIT CALC TESTS
1:Z-Test...
2:T-Test...
3:2-SampZTest...
4:2-SampTTest...
5:1-PropZTest...
6:2-PropZTest...
7:ZInterval...
    
```

} } を押す

```

EDIT CALC TESTS
1:2-SampInt...
A:1-PropZInt...
B:2-PropZInt...
C:X^2-Test...
D:2-SampFTest...
2:LinRegTTest...
F:ANOVA(
    
```

í を押す

```

LinRegTTest
Xlist:L1
Ylist:L2
Freq:1
B & P:EQ <0
RegEQ:Y1
Calculate
    
```

```

LinRegTTest
y=a+bx
B≠0 and P≠0
t=7.379463761
P=3.4598299E-6
df=14
↓a=3.905235493
    
```

```

LinRegTTest
y=a+bx
B≠0 and P≠0
↑b=.1949465557
s=1.353282392
r^2=.7954905234
r=.8919027544
    
```

上図のように入力し，最下段にカーソルを合わせてí

† で検定結果を確認する

なお、 $F_0 = t^2 = 7.37^2 = 54.31$  (小数点第3位以下を切り捨て) であることに注意する。

## 6. 残差の検討

残差の分布を検討することで回帰式が説明しきれない程度がわかる。「専有面積」と「残差」、実際の家賃から回帰式より求まる家賃の差、の相関係数を検討する。まず **STAT** を押す。

L1	L2	L3	1
16.5	6.2	-----	
17.49	6.8		
18.28	7		
20.46	7.2		
20.25	7.8		
18.89	8.2		
40	9.4		
L1()=16.5			

L1	L2	L3	3
16.5	6.2	-----	
17.49	6.8		
18.28	7		
20.46	7.2		
20.25	7.8		
18.89	8.2		
40	9.4		
L3 =			

NAME	OPS	MATH
1: L1		
2: L2		
3: L3		
4: L4		
5: L5		
6: L6		
7: RESID		

~ ~ } **I** を押す

y **[LIST]** を押す

- を押す

L1	L2	L3	3
16.5	6.2	-----	
17.49	6.8		
18.28	7		
20.46	7.2		
20.25	7.8		
18.89	8.2		
40	9.4		
L3 = LRESID			

L1	L2	L3	3
16.5	6.2	-----	
17.49	6.8	-.5149	
18.28	7	-.4689	
20.46	7.2	-.6938	
20.25	7.8	-.0529	
18.89	8.2	.61222	
40	9.4	-2.303	
L3() = -.921853661...			

STAT PLOTS	
1: Plot1...On	↖ L1 L2 +
2: Plot2...Off	↖ L1 L2 □
3: Plot3...Off	↖ L1 L2 □
4: PlotsOff	

**I** を押す

y **[STAT PLOT]** を押す

**Δ** を押す

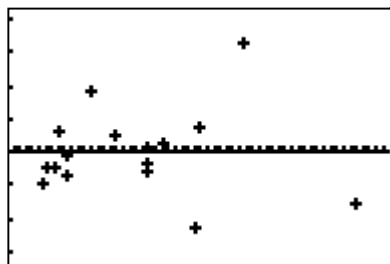
Plot1	Plot2	Plot3
Off	Off	Off
Type:		
Xlist: L1		
Ylist: L3		
Mark: □ ■ .		

Plot1	Plot2	Plot3
Off	Off	Off
Type:		
Xlist: L1		
Ylist: RESID		
Mark: □ ■ .		

ZOOM MEMORY
1: ZBox
2: Zoom In
3: Zoom Out
4: ZDecimal
5: ZSquare
6: ZStandard
7: ZTrig

上図のように設定を変更する (右図でも OK) **ZOOM** を押す

**Ⓢ** を押す



y=ax+b
a=.1949465557
b=3.905235493
r <sup>2</sup> =.7954905234
r=.8919027544
LinReg(ax+b) L1,
L3, Y1

LinReg
y=ax+b
a=0
b=1.25E-14
r <sup>2</sup> =0
r=0

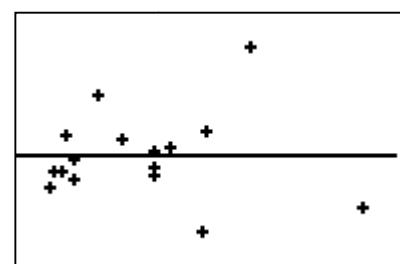
y **[QUIT]** を押す

上記のように入力し

y **[FORMAT]** を押す

PolarGC
CoordOn CoordOff
GridOff GridOn
AxesOn AxesOff
LabelOff LabelOn
ExprOn ExprOff

PolarGC
CoordOn CoordOff
GridOff GridOn
AxesOn AxesOff
LabelOff LabelOn
ExprOn ExprOff



上右図のように設定を変更

**GRAPH** を押す

無相関であることがわかる

## 7. 回帰式の検討と活用

求めた回帰式を活用して「家賃」を予測してみる。これは検証作業ともいえる。表4のデータを利用するが、この専有面積は、表1の16と15のデータ間にあるので、この予測は内挿となる。

表4

物件番号	専有面積 (m <sup>2</sup> )	家賃 (万円)
17	54.5	16.0

改めてL<sub>1</sub>とL<sub>2</sub>の回帰式を求める(6章直後に実行するとうまくいかない)。

I を押す
[VARS] ~ を押す
Δ を押す

Δ を押す
£ · ¶ ÷ · × I と押す

以上より、この物件の相場の予想は14.5万円であることがわかる。実際には16.0万円であった(表4)。この結果を受けて、例えばこの物件に関して有面積以外の要因を検討できる。この物件は南向きであるなど。なお、本来「専有面積」についての予測は母回帰式を利用する(区間推定)<sup>[2]</sup>。

## 8. おわりに

回帰分析についていえば、本稿では触れた内容が全てではない。回帰母数の検定と推定、母回帰の区間推定、さらにはデータの予測値の区間推定などについての分析が必要である。また本稿では「家賃」を予測するため「専有面積」と「家賃」の関係に注目した。しかし他にも「駅からの距離」や「築年数」と「家賃」の関係なども考えられる。実際「専有面積」「駅からの距離」「築年数」から「家賃」を予想するほうが予想精度(寄与率)を高められそうなことは容易に想像できる。このような分析は重回帰分析と呼ばれる。ただし統計処理は単純であることも重要である。このあたりも加味して、本稿の最後に、統計の活かし方を解説した書籍[3]を紹介しておく。

※注 Excelでは、 $F(1, n-2; 0.05)$ は「=FINV(0.05,1,n-2)」により求められる。

参考文献 [1] 谷岡一郎(2007)「データはウソをつく」筑摩書房。

[2] 奥原正夫(2007)「JUSE-StatWorksによる回帰分析入門」日科技連出版社。

[3] 林知己夫(2011)「調査の科学」筑摩書房。