

## 回帰分析

### 1. 目的

目的変数の数量を推測するものです。

回帰分析では、目的変数、説明変数ともに数量です。

なお、予測対象の目的変数は数量ですが、説明変数が定性的なデータの場合には、数量化一類を利用します。

ここでは、店舗の売り上げデータを例題に説明します。

会社の店舗の売上高を予測します。

売り上げの要因としては、

- ・販売促進費
- ・販売スタッフの人数

で、以下のデータが提示されています。

No	営業所	売上高(万円)	広告費(万円)	販売員数(人)
1	A	8,000	500	6
2	B	9,000	500	8
3	C	13,000	700	10
4	D	11,000	400	13
5	E	14,000	800	11
6	F	17,000	1,200	13

計算対象データは、列方向（横方向）に

I D、目的変数、説明変数 1、説明変数 2.... と定義することとします。

この場合は、

- ・目的変数は 売上高 Y
- ・説明変数 1 は 広告費  $X_1$
- ・説明変数 2 は 販売員数  $X_2$

で、一番左に I D（営業所）を定義します。

説明変数の数の上限は決められてません。この場合は 2 つです。

予測式は

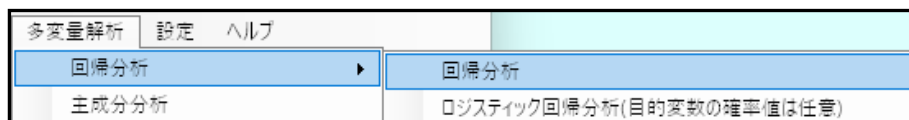
$$\cdot Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2 + \varepsilon \quad (\varepsilon \text{ は誤差})$$

であり、 $\sum \varepsilon^2$  が最小となる  $a_0$ ,  $a_1$ ,  $a_2$  を求めます。

## 2. 使用法

### (1) メニューの選択

メニューの「多変量解析→回帰分析」を選択します。



### (2) パネルが表示されます。



### (3) 計算対象データを入力

表計算ソフトに計算対象データを以下のように定義します。

No	営業所	売上高(万円)	広告費(万円)	販売員数(人)
1	A	8,000	500	6
2	B	9,000	500	8
3	C	13,000	700	10
4	D	11,000	400	13
5	E	14,000	800	11
6	F	17,000	1,200	13

上記の表の 赤い部分を コピーして、  
クリックすると、以下のように入力が完了します。

表データを貼り付け

表データを貼り付け

☒ 先頭行をラベルとして使用

クリア

	NO	営業所	売上高(万円)	広告費(万円)	販売員数(人)
▶	1	A	8,000	500	6
	2	B	9,000	500	8
	3	C	13,000	700	10
	4	D	11,000	400	13
	5	E	14,000	800	11
	6	F	17,000	1,200	13
✱					

#### (4) 計算を実行

「回帰係数の推定」ボタンを押すことで、計算結果が表示されます。

標準誤差818.1749

決定係数0.9641387

重相関係数0.9819056

自由度調整済み決定係数0.9402313

分散分析表

変動因

平方和

自由度

平均平方

F 値

P 値(%)

結果

帰無仮説の採択域

回帰

5.399177E

2

2.699588E

40.32787

0.6791071

有意

( 0 , 9.552094 )

分布関数

誤差

2008231

3

669410.1

全体

5.6E+07

5

	No.	説明変数名	回帰係数	係数下限	係数上限	t 値	p 値(%)	標準誤差	標準化回帰係数	偏相関係数
▶	0	定数項	1148.148	-3307.541	5603.837	0.8202273	47.21886	1399.793	-4.880175E-17	0
	1	広告費(万円)	7.860082	3.351945	12.36822	5.549848	1.153614	1.41627	0.6874229	0.9545914
	2	販売員数(人)	539.0947	65.63199	1012.557	3.624353	3.613658	148.7423	0.4489245	0.9022636

#### (5) 計算結果の説明

##### 1) 予測精度について

標準誤差	818.1749	決定係数	0.9641387	重相関係数	0.9819056	自由度調整済み決定係数	0.9402313
------	----------	------	-----------	-------	-----------	-------------	-----------

予測精度を示すものとして、

- ・ 重相関係数 0.98
- ・ 決定係数 0.96
- ・ 自由度調整済み決定係数 0.94

と表示されています。

重相関係数はこの推定の当てはまり具合をしめすもので、0.0～1.0の値を取ります。1.0に近いほど、推定の精度が高いことを示します。この場合0.98ですので、高い精度で推定していることを示します。

決定係数は重相関係数の二乗で、自由度調整済み決定係数は説明変数の数を考慮して、決定係数に調整を加えたものです。

## 2) 分散分析表

分散分析表								
変動因	平方和	自由度	平均平方	F 値	P 値(%)	結果	帰無仮説の採択域	
回帰	5.399177E	2	2.699588E	40.32787	0.6791071	有意	( 0 , 9.552094 )	分布関数
誤差	2008231	3	669410.1					
全体	5.6E+07	5						

前提としている予測式  $Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2$  が意味を持つのか？を示すものです。

予測が意味を持たないとは、「回帰係数  $a_1$ 、 $a_2$  が 0 である」ことを帰無仮説として検定を行うもので、上記の「結果」欄が“有意でない”場合に、帰無仮説が採択されます。

上記では「結果」欄が“有意”となっているので、

予測が意味を持たないとは言えない  
ことを意味します。

## 3) 説明変数の評価

	No.	説明変数名	回帰係数	係数下限	係数上限	t 値	p 値(%)	標準誤差	標準化回帰係数	偏相関係数
▶	0	定数項	1148.148	-3307.541	5603.837	0.8202273	47.21886	1399.793	-4.880175E-17	0
	1	広告費(万円)	7.860082	3.351945	12.36822	5.549848	1.153614	1.41627	0.6874229	0.9545914
	2	販売員数(人)	539.0947	65.63199	1012.557	3.624353	3.613658	148.7423	0.4489245	0.9022636

モデル式  $Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2$  において、  
回帰係数は、

定数項  $a_0$  は 1148.148  
 広告費  $a_1$  は 7.86  
 販売員数  $a_2$  は 539.09

であることを示しています。

また各回帰係数の t 値は、回帰係数としての有意性を検定するもので、

広告費について 5.54  
 販売員数について 3.62

となっています。

この値は回帰係数が 0 であるかの判断の参考となり、t 値の絶対値がおおむね 2.0 以上ならば 0 でないと判断します。

少し詳しく言うと、t 値 = (回帰係数 - 0.0) / 標準誤差 は、  
 自由度(n - k - 1) (n: 個数、k: 変数) の t 分布に従うので、上記で計算される t 値が、「回帰係数 = 0.0 とする帰無仮説」の採択域から外れていれば、帰無仮説は棄却されます。

(つまり回帰係数は 0 でないと言える)

さらに各回帰係数の p 値は、t 値から計算される有意確率です。

広告費について 1.15 (%)  
 販売員数について 3.61 (%)

でいづれも 有意水準 (この場合 5 %) 以下ですので、  
 説明変数として意味を持つことを示します。

更に、標準化回帰係数は、

広告費  $a_1$  は 0.687

販売員数  $a_2$  は 0.448

となっています。

標準化しない場合の、回帰係数は

広告費  $a_1$  は 7.86

販売員数  $a_2$  は 539.09

ですが、両変数が扱うスケールや単位は異っているので、単純にその係数値で 効果の大きさを判定するのは無理があります。

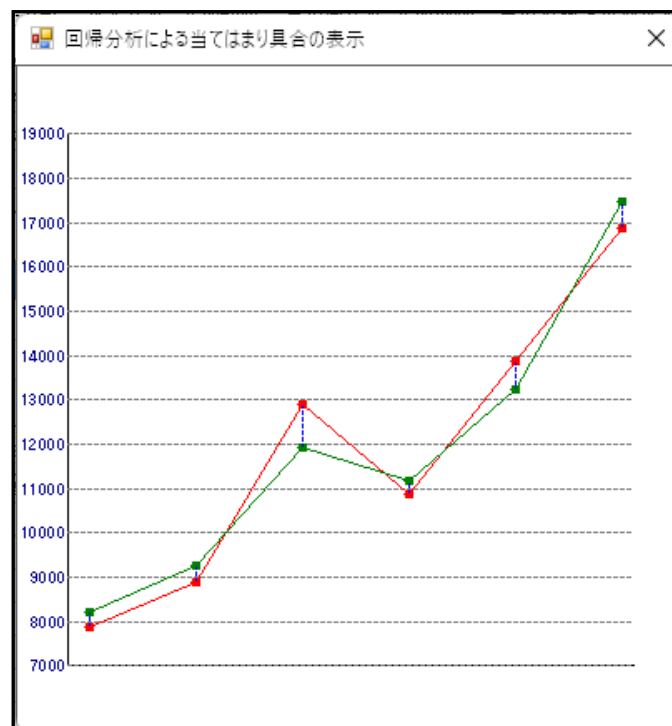
そこで数値を標準化して計算することで、係数の大小を比較しやすくするものです。

最後の偏相関係数は、目的変数と説明変数との相関係数から、説明変数間の相関関係を取り除いたものになります。

#### (6) 数式の当てはまり具合を表示

あてはまり具合表示

ボタンを押すことで表示されます。



入力値の順に左からデータがプロットされています。

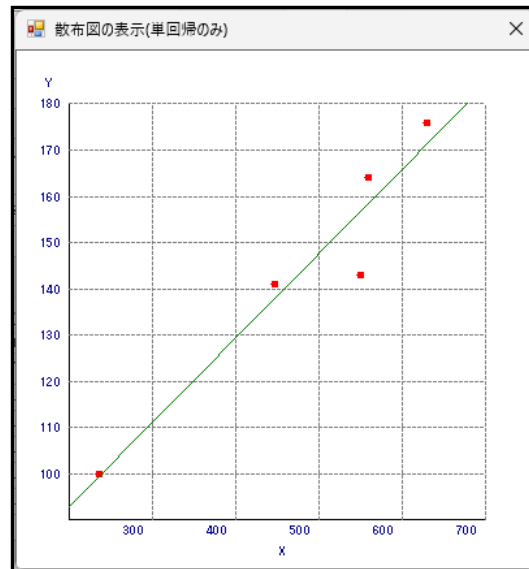
赤い点は入力値、緑の点は推計値を示すものです。

各々の点を線で結ぶことで、推計がうまくいっているのかを直観で分かりやすいように 表示しています。

### (7) 散布図（単回帰のみ）

単回帰（説明変数が1つのみ）の場合に限りませんが、計算結果と入力データとの散布図を描画できます。

**散布図(単回帰のみ)** を押すことで、散布図が描画できます。



赤い点が入力データで、緑の線が計算結果を表わします。  
多変量（説明変数が2つ以上）の場合には、変量の単位、スケールが異なる為、図での表示が意味をなさないことがあるため、単回帰のみの機能となっています。

### (8) 回帰式による予測

**回帰式による予測** を押すことで、推計を実行します。

回帰分析の結果に基づく推定

NO	説明変数名	値
1	広告費(万円)	-
2	販売員数(人)	-

計算

計算結果

期待値：

分散値：

信頼区間：(  ,  )

分布関数

広告費と、販売員数とに 数値を代入して、売り上げの期待値と、信頼区間を 計算できます。



The screenshot shows a software window titled "回帰分析の結果に基づく推定" (Prediction based on regression analysis results). The window is divided into two main sections. On the left, there is a table with three columns: "NO", "説明変数名" (Explanatory variable name), and "値" (Value). The table contains two rows: Row 1 with "1" in the NO column, "広告費(万円)" (Advertising cost in 10,000 yen) in the explanatory variable column, and "600" in the value column; Row 2 with "2" in the NO column, "販売員数(人)" (Number of sales staff in people) in the explanatory variable column, and "10" in the value column. Below the table is a button labeled "計算" (Calculate). On the right, there is a section titled "計算結果" (Calculation results) which displays the following values: "期待値:" (Expected value) as 11255.14, "分散値:" (Variance value) as 123362.2, and "信頼区間:" (Confidence interval) as ( 10137.14 , 12373.15 ). There is also a button labeled "分布関数" (Distribution function) below the confidence interval.

NO	説明変数名	値
1	広告費(万円)	600
2	販売員数(人)	10

計算

計算結果

期待値 : 11255.14

分散値 : 123362.2

信頼区間 : ( 10137.14 , 12373.15 )

分布関数

広告費に 600万円、販売員に 10名を指定すると、  
売り上げの期待値は 11,255万円、  
その信頼区間は (10,137～12,373)万円 と  
計算されます。

(9) 変数を計算対象から外す

入力データによっては、説明変数が回帰分析において有意でない場合があります。

その場合、説明変数を省いて 再計算するなどの処理を行います。

ここでは、高校生の身長、体重、胸囲を計測したデータを例に説明します。

20人分のデータは 以下です。

ID	身長	体重	胸囲
1	167.0	84.0	61.0
2	167.5	87.0	55.5
3	168.4	86.0	57.0
4	172.0	85.0	57.0
5	155.3	82.0	50.0
6	151.4	87.0	50.0
7	163.0	92.0	66.5
8	174.0	94.0	65.0
9	168.0	88.0	60.5
10	160.4	84.9	49.5
11	164.7	78.0	49.5
12	171.0	90.0	61.0
13	162.6	88.0	59.0
14	164.8	87.0	58.4
15	163.3	82.0	53.5
16	167.6	84.0	54.0
17	169.2	86.0	60.0
18	168.0	83.0	58.8
19	167.4	85.2	54.0
20	172.0	82.0	56.0

身長に対して 体重が大きく、特殊な部活動の生徒のデータのように見えますが、それはともかく、

モデル式  $Y = a_0 + a_1 \cdot X_1 + a_2 \cdot X_2$  として、

- ・目的変数は 身長  $Y$
- ・説明変数 1 は 体重  $X_1$
- ・説明変数 2 は 胸囲  $X_2$

で計算します。



計算結果は以下です。

標準誤差4.488991

決定係数0.4118727

重相関係数0.6417731

自由度調整済み決定係数0.3426813

分散分析表

変動因

平方和

自由度

平均平方

F 値

P 値(%)

結果

帰無仮説の採択域

回帰

239.9043

2

119.9522

5.952653

1.097757

有意

( 0 , 3.591531 )

分布関数

誤差

342.5677

17

20.15104

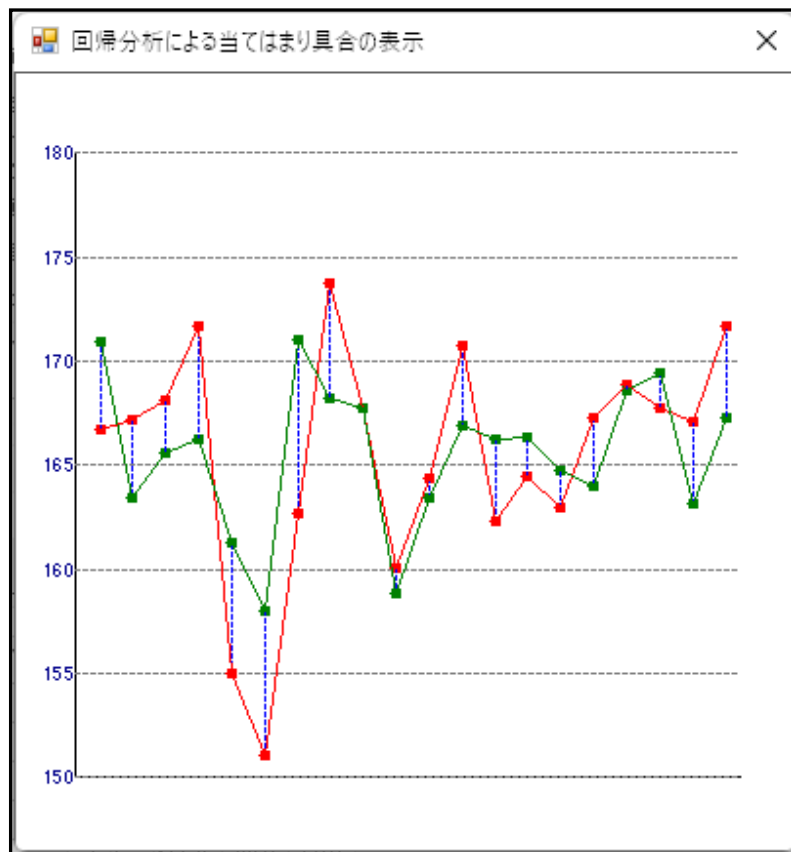
全体

582.472

19

	No.	説明変数名	回帰係数	係数下限	係数上限	t 値	p 値(%)	標準誤差	標準化回帰係数	偏相関係数
▶	0	定数項	166.5822	113.2321	219.9323	6.589587	0.000459...	25.27961	-4.91401E-15	0
	1	体重	-0.6683573	-1.523879	0.1871641	-1.648703	11.75639	0.4053836	-0.4445144	-0.371286
	2	胸囲	0.9365288	0.3562384	1.636819	3.284563	0.4373413	0.3033977	0.8855659	0.6230834

良い精度ではないようです。 当てはまり具合を表示します。



重相関係数は 0.64 で良い精度の推定とは言えなさそうです。  
分散分析の結果は有意ですので、全くダメというわけではありませんが。

回帰係数の有意性について、p 値、t 値を見えます。

- ・ 体重 については t 値 -1.64、 p 値 11.8 (%)
- ・ 胸囲 については t 値 3.28、 p 値 0.44 (%)

ということですので、少なくとも、体重は推定にあまり役に立っていないものと思われます。 (体重の行を色づけして警告します)

そこで、体重を 今回の推定から外すことにします。  
 入力のグリッドの ヘッダ部分 “体重” を 右クリックします。  
 すると 当該変数を 計算対象から 外すかどうかを聞いてきます。

説明変数の除外/復活

(体重)を説明変数から外しますか?  
 除外の場合は“はい”, 説明変数として残す場合は“いいえ”を選択してください

はい(Y) いいえ(N)

計算から除外された変数が赤く表示されます。

体重を除外して、胸囲だけで推定した結果は以下です。

標準誤差 4.698361 決定係数 0.3178338 重相関係数 0.5637676 自由度調整済み決定係数 0.2793357										
分散分析表										
変動因	平方和	自由度	平均平方	F 値	P 値(%)	結果	帰無仮説の採択域			
回帰	185.1293	1	185.1293	8.386533	0.9629218	有意	( 0	,	4.413873 )	分布関数
誤差	397.3427	18	22.07459							
全体	582.472	19								

	No.	説明変数名	回帰係数	係数下限	係数上限	t 値	p 値(%)	標準誤差	標準化回帰係数	偏相関係数
▶	0	定数項	129.8392	103.5926	156.0859	10.39588	4.897626...	12.48949	-3.939685E-15	0
	1	胸囲	0.6344086	0.1740384	1.094779	2.895951	0.9629218	0.2190674	0.5637676	0.5637676