

## 2群以上の関係を把握する方法-1 (重回帰分析・ロジスティック回帰分析)

2019年11月19日  
株式会社データサイエンス研究所  
伊藤嘉朗

1

Data Science Institute

### 回帰分析

例) 売上高と売上高に影響を与える要因との関係

売上高 $\Leftarrow$ 広告宣伝費、人口、セールスマン数、...

従属変数 $\Leftarrow$ 説明変数 (独立変数)

説明変数が1つ：単回帰

説明変数が2つ以上：重回帰

目的1：回帰式を求め、予測する。

回帰式 ( $y = a + b x + \dots$ ) を求める。

$a, b, \dots \rightarrow$  偏回帰係数

2

Data Science Institute

### ◇駅前コンビニの売上高と乗降客数

売上高 ← 乗降客数

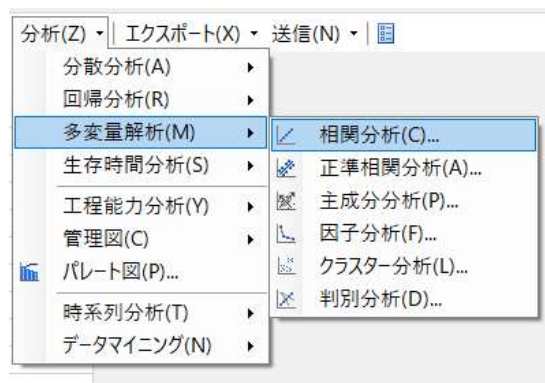
	売上高 (百万円/月)	乗降客数 (百人/日)
1	130	93
2	290	234
3	235	250
4	260	260
5	140	119
6	173	180
7	135	151
8	190	192
9	220	273
10	181	185

3

### 相関係数を求める。(SAS EG)

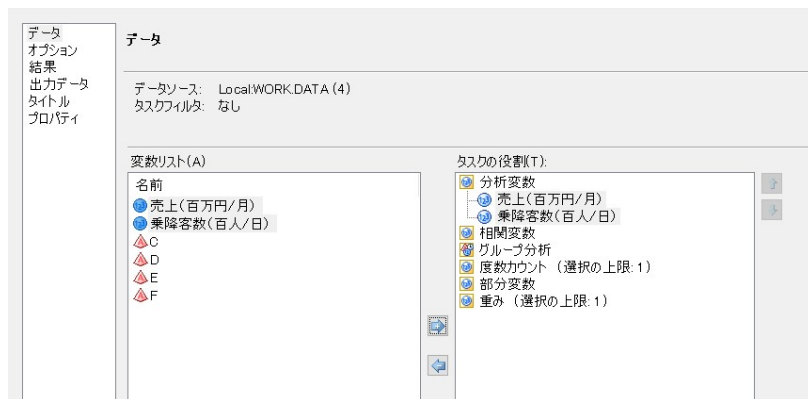
EG

1. データを入力する。
2. 「分析」-「多変量解析」-「相関分析」を選択する。



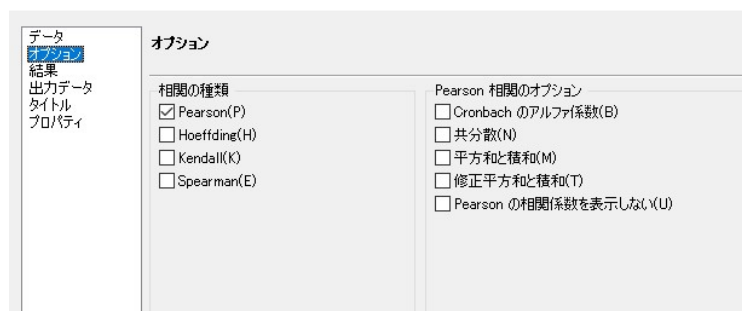
4

### 3.「データ」をクリックし、「分析変数」を設定する。



5

### 4.「オプション」をクリックし、「ピアソン (Pearson) 」を選択する。



6

### 相関分析 CORR プロシジャ

2 変数 : 売上(百万円/月) 乗降客数(百人/日)

単純統計量						
変数	N	平均	標準偏差	合計	最小値	最大値
売上(百万円/月)	10	195.40000	54.80105	1954	130.00000	290.00000
乗降客数(百人/日)	10	193.70000	60.81676	1937	93.00000	273.00000

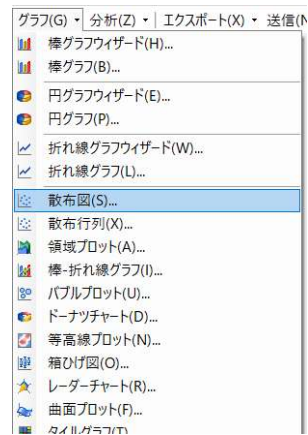
Pearson の相関係数, N = 10 H0: Rho=0 に対する Prob >  r		
	売上(百万円/月)	乗降客数(百人/日)
売上(百万円/月)	1.00000	0.86747
乗降客数(百人/日)	0.86747	1.00000
	0.0011	0.0011

積率相関係数 = 0.867

P値 = 0.0011

### 散布図を作成する。(SAS EG)

1. データを入力 (「ファイル」→「新規作成」→「データ」)
2. 「グラフ」→「散布図」を選択する。

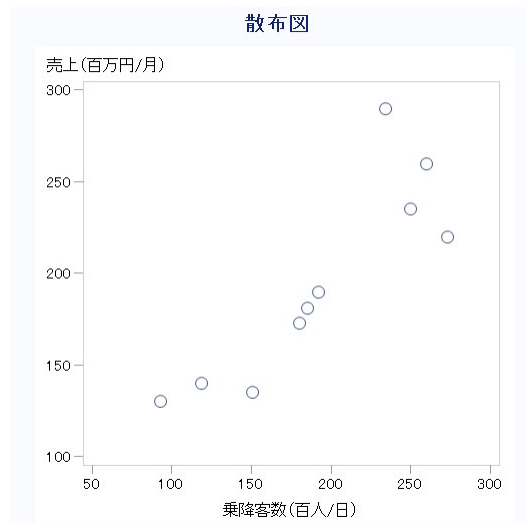


### 3. 「データ」をクリックし、乗降客数を「X軸」に、売上高を「Y軸」に設定する。



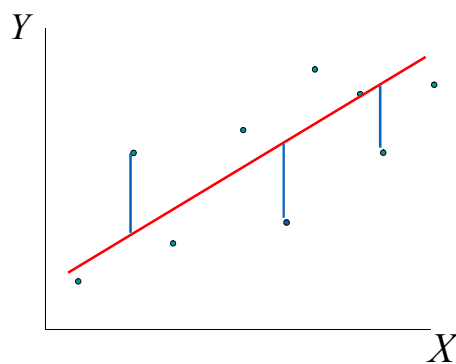
### 4. 「グラフ領域」をクリックし、「カスタムチャートサイズ」を指定する。





回帰式 ( $y = a + bx$ ) を求めて予測する。

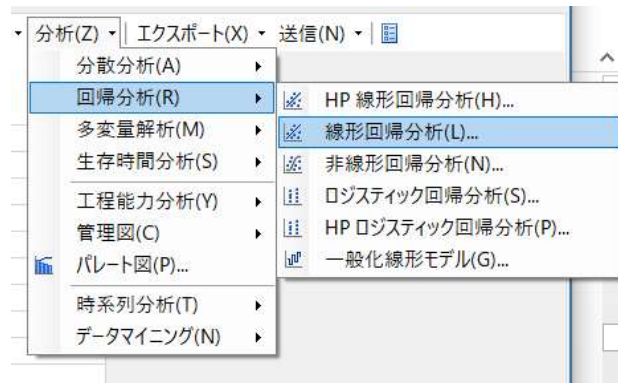
## 回帰直線の求め方 (最小二乗法)



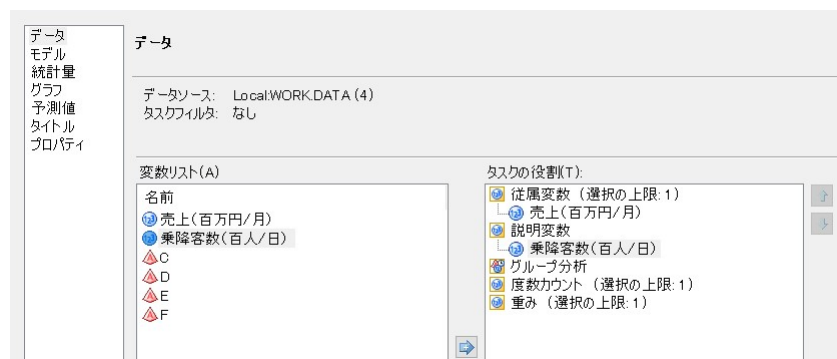
各データと回帰直線との垂直距離の2乗和を最小

## 単回帰分析 (SAS EG)

1. データを入力する。
2. 「分析」－「回帰分析」－「線形回帰分析」を選択する。



3. 「データ」をクリックし、売上高を「従属変数」、乗降客数を「説明変数」に設定する。



## 単回帰分析 結果

Root MSE	28.91642	R2 乗	0.7525
従属変数の平均	195.40000	調整済み R2 乗	0.7216
変動係数	14.79858		

パラメータの推定					
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t
Intercept	1	43.99116	32.03230	1.37	0.2069
乗降客数(百人/日)	1	0.78167	0.15849	4.93	0.0011

$$\text{売上高} = 43.99 + 0.782 \times \text{乗降客数}$$

\* 乗降客数 = 180のときの売上高の予測

$$\text{売上高} = 43.99 + 0.782 \times 180 = 184.75$$

## 回帰モデルのチェック

◇自由度調整済み決定係数

$$R^2 = 0.7216$$

約72.16%説明できる。

◇偏回帰係数のt検定

$$\text{乗降客数} : P\text{値} (P\text{値}) = 0.0011 < 0.05$$

有意水準5%において、偏回帰係数は有意。

### 説明変数に「取扱品目数」を追加

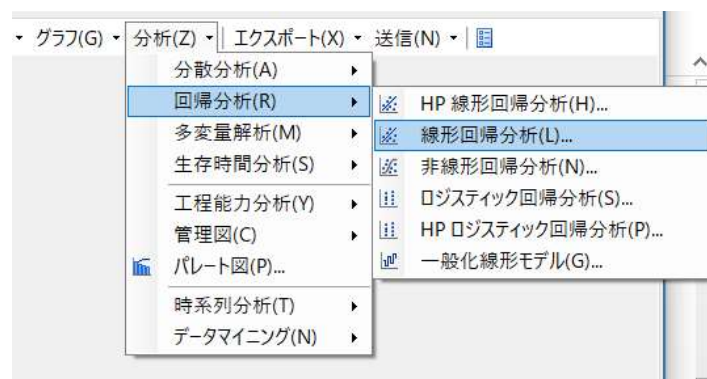
	売上高 (百万円/月)	乗降客数 (百人/日)	取扱品目数 (品)
1	130	93	150
2	290	234	311
3	235	250	182
4	260	260	245
5	140	119	149
6	173	180	160
7	135	151	98
8	190	192	180
9	220	273	113
10	181	185	105

17

## 重回帰分析 (SAS EG)

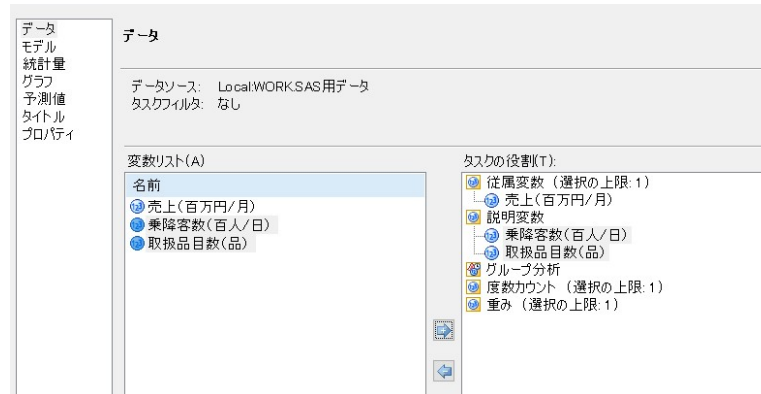
EG

1. データを入力する。
2. 「分析」－「回帰分析」－「線形回帰分析」を選択する。



18

3.「データ」をクリックし、売上高を「従属変数」、乗降客数、取扱品目数を「説明変数」に設定する。



19

Root MSE	10.04926	R2 乗	0.9738
従属変数の平均	195.40000	調整済み R2 乗	0.9664
変動係数	5.14292		

パラメータの推定					
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t
Intercept	1	7.15385	12.11739	0.59	0.5735
乗降客数(百人/日)	1	0.60153	0.05985	10.05	<.0001
取扱品目数(品)	1	0.42368	0.05505	7.70	0.0001

#### ◇偏回帰係数のt検定

乗降客数 P値 = <0.0001 <0.05

取扱品目数 P値 = 0.0001 <0.05

20

### 重回帰分析の目的1

売上高の予測

$$\text{売上高} = 7.15 + 0.602 \times \text{乗降客数} + 0.424 \times \text{品目数}$$

\* 乗降客数=200 取扱品目数が180のとき売上高の予測

$$\text{売上高} = 7.15 + 0.602 \times 200 + 0.424 \times 180 = 203.87$$

◇自由度調整済決定係数

$$R^2 = 0.9664$$

約96.64%説明できる。

21

### 重回帰分析の目的2

説明変数の影響力の比較

乗降客数と取扱品目数のどちらの方が影響力が強い？

偏回帰係数の比較

乗降客数 : 0.602 > 取扱品目数 : 0.424



乗降客数の方が売上高に与える影響は強い!?

22

## 入力単位の変更

乗降客数の単位：百人/日 ⇒ 千人/日

	売上高 (百万円/月)	乗降客数 (千人/日)	取扱品目数 (品)
1	130	9.3	150
2	290	23.4	311
3	235	25.0	182
4	260	26.0	245
5	140	11.9	149
6	173	18.0	160
7	135	15.1	98
8	190	19.2	180
9	220	27.3	113
10	181	18.5	105

23

## 重回帰分析 結果

EG

Root MSE	10.04926	R2 乗	0.9738
従属変数の平均	195.40000	調整済み R2 乗	0.9664
変動係数	5.14292		

パラメータの推定					
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t
Intercept	1	7.15385	12.11739	0.59	0.5735
乗降客数(千人/日)	1	6.01533	0.59846	10.05	<.0001
取扱品目数(品)	1	0.42368	0.05505	7.70	0.0001

24

## 偏回帰係数の比較

	変更前	変更後
乗降客数	0.60153	6.01533
取扱品目数	0.42368	0.42368

乗降客数の偏回帰係数の値が10倍

t 検定・決定係数の値は同じ。



偏回帰係数はデータの単位の影響を受ける。

偏回帰係数の単純比較は無意味

25

データの入力単位の影響を除くには



z値

データをz値変換してから重回帰分析

標準偏回帰係数



係数の相互大小比較が可能

26

## 標準偏回帰係数を求める。(SAS EG)

「統計量」をクリックし、標準偏回帰係数を選択する。

The screenshot shows the SAS EG 'Statistics' dialog box. On the left, a navigation pane lists 'データ', 'モデル', '統計量', 'グラフ', '予測値', 'タイトル', and 'プロパティ'. The '統計量' (Statistics) section is selected. It contains two main areas: '推定値の詳細' (Details of Estimates) and '診断統計量' (Diagnostic Statistics). Under '推定値の詳細', the checkbox for '標準偏回帰係数(D)' (Standardized Regression Coefficients) is checked. Other options include '平方和 Type 1(I)', '平方和 Type 2(J)', '推定値相関行列(L)', '推定値の共分散行列(V)', and 'パラメータ推定値の信頼限界(F)'. A '信頼水準(N)' (Confidence Level) dropdown is set to 95%. Under '診断統計量', several checkboxes are present but unchecked, including '共線性分析(O)', '切片を考慮し入れない共線性分析(W)', '推定値のトランス値(T)', 'VIF (分散拡大係数)(D)', '不等分散性の検定(H)', '漸近共分散行列(A)', and 'Durbin-Watson 統計量(B)'. At the bottom, there is a '相関' (Correlation) section with checkboxes for '偏相関(P)' and '半偏相関(M)'.

27

標準偏回帰係数が表示される。

Root MSE	10.04926	R2 乗	0.9738
従属変数の平均	195.40000	調整済み R2 乗	0.9664
変動係数	5.14292		

パラメータの推定						
変数	自由度	パラメータ		t 値	Pr >  t	標準推定値
		推定値	標準誤差			
Intercept	1	7.15385	12.11739	0.59	0.5735	0
乗降客数	1	6.01533	0.59846	10.05	<.0001	0.66756
取扱品目数	1	0.42368	0.05505	7.70	0.0001	0.51117

乗降客数 (0.668) > 取扱品目数 (0.511)

28

### 説明変数に世帯数をさらに追加

	売上高 (百万円/月)	乗降客数 (千人/日)	取扱品目数 (品)	世帯数
1	130	93.0	150	143
2	290	234.0	311	284
3	235	250.0	182	320
4	260	260.0	245	302
5	140	119.0	149	182
6	173	180.0	160	225
7	135	151.0	98	190
8	190	192.0	180	242
9	220	273.0	113	320
10	181	185.0	105	235

29

### 重回帰分析 結果

EG

Root MSE	10.65938	R2 乗	0.9748
従属変数の平均	195.40000	調整済み R2 乗	0.9622
変動係数	5.45516		

パラメータの推定						
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t	標準 推定値
Intercept	1	-1.82664	23.00272	-0.08	0.9393	0
乗降客数	1	0.42311	0.38429	1.10	0.3131	0.46956
取扱品目数	1	0.42144	0.05858	7.19	0.0004	0.50847
世帯数	1	0.17978	0.38190	0.47	0.6544	0.20141

30

◇偏回帰係数のt検定

乗降客数 P値 = 0.3131 > 0.05  
取扱品目数 P値 = 0.0004 < 0.05  
世帯数 P値 = 0.6544 > 0.05



売上高を乗降客数、取扱品目数、世帯数の3つの説明変数で説明する重回帰モデルは、不成立！

説明変数の検討

- ①乗降客数 & 取扱品目数
- ②乗降客数 & 世帯数
- ③取扱品目数 & 世帯数

①説明変数：乗降客数 & 取扱品目数

Root MSE	10.04926	R2 乗	0.9738
従属変数の平均	195.40000	調整済み R2 乗	0.9664
変動係数	5.14292		

パラメータの推定						
変数	自由度	パラメータ		t 値	Pr >  t	標準推定値
		推定値	標準誤差			
Intercept	1	7.15385	12.11739	0.59	0.5735	0
乗降客数	1	6.01533	0.59846	10.05	<.0001	0.66756
取扱品目数	1	0.42368	0.05505	7.70	0.0001	0.51117

乗降客数 P値 = <0.0001 < 0.05  
取扱品目数 P値 = 0.0001 < 0.05

## ②説明変数：乗降客数&amp;世帯数

Root MSE	30.61661	R2 乗	0.7572
従属変数の平均	195.40000	調整済み R2 乗	0.6879
変動係数	15.66869		

パラメータの推定						
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t	標準 推定値
Intercept	1	23.40110	65.29769	0.36	0.7306	0
乗降客数	1	0.37915	1.10365	0.34	0.7413	0.42077
世帯数	1	0.40343	1.09329	0.37	0.7230	0.45196

乗降客数 P値 = 0.7413 > 0.05

世帯数 P値 = 0.7230 > 0.05

## ③説明変数：取扱品目数&amp;世帯数

Root MSE	10.81976	R2 乗	0.9697
従属変数の平均	195.40000	調整済み R2 乗	0.9610
変動係数	5.53723		

パラメータの推定						
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t	標準 推定値
Intercept	1	-21.00871	15.24568	-1.38	0.2106	0
取扱品目数	1	0.42041	0.05946	7.07	0.0002	0.50723
世帯数	1	0.59449	0.06403	9.28	<.0001	0.66600

取扱品目数 P値 = 0.0002 < 0.05

世帯数 P値 = <0.0001 < 0.05

説明変数	t検定
乗降客数、取扱品目数、世帯数	×
①乗降客数、取扱品目数	○
②乗降客数、世帯数	×
③取扱品目数、世帯数	○

### ピアソンの積率相関係数

	売上高	乗降客数	取扱品目数	世帯数
売上高	1.00000	0.86747	0.77224	0.86784
乗降客数	0.86747	1.00000	0.39108	0.98837
取扱品目数	0.77224	0.39108	1.00000	0.39792
世帯数	0.86784	0.98837	0.39792	1.00000

35

売上高	乗降客数	0.867
	取扱品目数	0.772
	世帯数	0.868

いずれも高い値⇒売上高を説明する説明変数として妥当

乗降客数	取扱品目数	0.391
乗降客数	世帯数	0.988
取扱品目数	世帯数	0.398

乗降客数と世帯数の値0.988は高い値

説明変数相互の積率相関係数は低い方が良い！

説明変数⇒独立変数

36

### ◇A社の昨年1年間の支店別売上高実績

支店	売上高	宣伝広告費	セールスマン数	世帯数	販促費
1	135	150	24	100	148
2	193	190	23	182	201
3	230	272	27	115	268
4	175	188	30	102	180
5	188	231	25	90	225
6	145	126	41	113	130
7	212	163	45	200	158
8	280	207	68	300	198
9	216	135	57	178	134
10	130	93	32	149	98
11	295	235	62	312	230
12	237	247	57	183	242
13	265	264	39	241	273
14	137	117	27	151	115
15	173	181	30	163	180

37

### 重回帰分析の結果

EG

Root MSE	12.22175	R2 乗	0.9626
従属変数の平均	200.73333	調整済み R2 乗	0.9477
変動係数	6.08855		

パラメータの推定					
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t
Intercept	1	1.68381	13.88187	0.12	0.9059
宣伝広告費(百万円)	1	0.52122	0.64390	0.81	0.4371
セールスマン(人)	1	1.13193	0.39405	2.87	0.0166
世帯数(万世帯)	1	0.29382	0.08471	3.47	0.0060
販促費(百万円)	1	0.03764	0.65920	0.06	0.9556

38

## ピアソンの積率相関係数

	売上(億円)	宣伝広告費(百万円)	セールスマン(人)	世帯数(万世帯)	販促費(百万円)
売上(億円)	1.00000	0.74678	0.70612	0.79438	0.73278
宣伝広告費(百万円)	0.74678	1.00000	0.16659	0.29249	0.99428
セールスマン(人)	0.70612	0.16659	1.00000	0.76203	0.13096
世帯数(万世帯)	0.79438	0.29249	0.76203	1.00000	0.29344
販促費(百万円)	0.73278	0.99428	0.13096	0.29344	1.00000

原則：・従属変数と説明変数の積率相関係数の値は高い

・説明変数相互の積率相関係数の値は低い



宣伝広告費と販促費の値0.994は高い値

説明変数： 宣伝広告費、セールスマン数、世帯数

Root MSE	11.65488	R2 乗	0.9626
従属変数の平均	200.73333	調整済み R2 乗	0.9524
変動係数	5.80615		

## パラメータの推定

変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t
Intercept	1	1.93002	12.58303	0.15	0.8809
宣伝広告費(百万円)	1	0.55782	0.05865	9.51	<.0001
セールスマン(人)	1	1.11993	0.31783	3.52	0.0048
世帯数(万世帯)	1	0.29597	0.07233	4.09	0.0018

説明変数：セールスマン数、世帯数、販促費

Root MSE	12.02870	R2 乗	0.9602
従属変数の平均	200.73333	調整済み R2 乗	0.9493
変動係数	5.99238		

#### パラメータの推定

変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t
Intercept	1	-1.07348	13.24491	-0.08	0.9369
セールスマン(人)	1	1.29898	0.33038	3.93	0.0023
世帯数(万世帯)	1	0.26496	0.07562	3.50	0.0049
販促費(百万円)	1	0.56880	0.06197	9.18	<.0001

#### t検定におけるサンプルサイズの影響

売上高	乗降客数	間口の広さ
130	93	150
290	234	148
235	250	182
260	260	245
140	119	149
173	180	160
135	151	135
190	192	180
220	273	113
181	185	105

## ◇データ10組のとき

Root MSE	29.46589	R2 乗	0.7751
従属変数の平均	195.40000	調整済み R2 乗	0.7109
変動係数	15.07978		

パラメータの推定						
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t	標準 推定値
Intercept	1	17.72278	45.22169	0.39	0.7068	0
乗降客数	1	0.74276	0.16802	4.42	0.0031	0.82429
間口の広さ	1	0.21573	0.25704	0.84	0.4290	0.15650

間口の広さ : P値 = 0.4290

## ◇データ50組のとき

Root MSE	25.42756	R2 乗	0.7751
従属変数の平均	195.40000	調整済み R2 乗	0.7656
変動係数	13.01308		

パラメータの推定						
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t	標準 推定値
Intercept	1	17.72278	17.45207	1.02	0.3151	0
乗降客数	1	0.74276	0.06484	11.45	<.0001	0.82429
間口の広さ	1	0.21573	0.09920	2.17	0.0347	0.15650

間口の広さ : P値 = 0.0347

t検定のP値はサンプルサイズの影響を受ける。

## 訪問回数と成約件数 (1)

成約件数	訪問回数
2	1
5	2
5	3
0	3
3	4
3	5

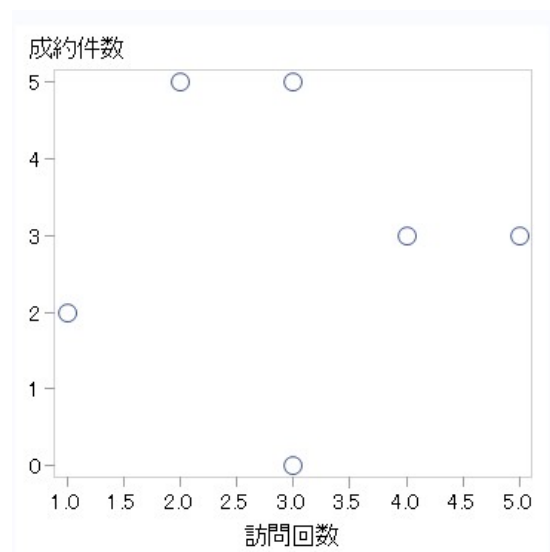
訪問回数と成約件数の関係は？

45

Data Science Institute

## 散布図

EG



46

Data Science Institute

## 積率相関係数

2 変数：成約件数 訪問回数

Pearson の相関係数, N = 6

	成約件数	訪問回数
成約件数	1.00000	0.00000
訪問回数	0.00000	1.00000

積率相関係数 = 0

## 訪問回数と成約件数 (2)

成約件数	訪問回数	性別
2	1	女性
5	2	女性
5	3	女性
0	3	男性
3	4	男性
3	5	男性

成約件数と、訪問回数、性別の関係は？

性別データの数値化



男性=1 女性=0

成約件数	訪問回数	男性ダミー
2	1	0
5	2	0
5	3	0
0	3	1
3	4	1
3	5	1

49

EG

Root MSE	1.00000	R2 乗	0.8333
従属変数の平均	3.00000	調整済み R2 乗	0.7222
変動係数	33.33333		

パラメータの推定					
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t
Intercept	1	1.00000	1.15470	0.87	0.4502
訪問回数	1	1.50000	0.50000	3.00	0.0577
性別	1	-5.00000	1.29099	-3.87	0.0305

成約件数 =  $1 + 1.5 \times \text{訪問回数} - 5 \times \text{男性ダミー}$

50

### 偏回帰係数の解釈

○訪問回数の偏回帰係数 = 1.5



性別が同じであれば、訪問回数が1回増えるごとに平均的に成約件数が1.5件ずつ多い。

○男性ダミーの偏回帰係数 = -5



訪問回数が同じであれば、男性は女性に比べて平均的に成約件数が5件少ない。

51

Data Science Institute

### ダミー変数の作成方法

(2区分)	ダミー1
有	1
無	0

区分数 - 1 ⇒ ダミー変数

(3区分)	ダミー1	ダミー2
大	1	0
中	0	1
小	0	0

(4区分)	ダミー1	ダミー2	ダミー3
20歳代	1	0	0
30歳代	0	1	0
40歳代	0	0	1
50歳代	0	0	0

52

Data Science Institute

売上個数 ← 曜日、温度

売上個数	曜日	温度
356	月	28
245	火	21
128	水	15
189	木	15
215	金	28
412	土	24
388	日	22
312	月	19
301	火	22
355	水	25

曜日 : 7区分



ダミー変数 : 6個

53

入力データ

売上個数	ダミー-1	ダミー-2	ダミー-3	ダミー-4	ダミー-5	ダミー-6	温度
356	1	0	0	0	0	0	28
245	0	1	0	0	0	0	21
128	0	0	1	0	0	0	15
189	0	0	0	1	0	0	15
215	0	0	0	0	1	0	28
412	0	0	0	0	0	1	24
388	0	0	0	0	0	0	22
312	1	0	0	0	0	0	19
301	0	1	0	0	0	0	22
355	0	0	1	0	0	0	25

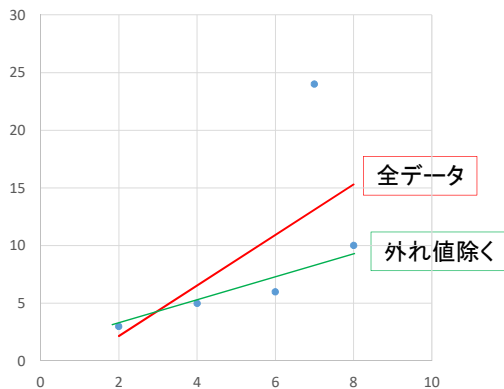
54

## 外れ値を含む場合

売上高	広告費
3	2
5	4
6	6
10	8
24	7

$$\text{売上高} = -2.22 + 2.19 \times \text{広告費}$$

$$\text{売上高} = 0.5 + 1.1 \times \text{広告費}$$



回帰直線は外れ値の影響を受ける。

55

## ダミー変数の設定

売上高	広告費	ダミー
3	2	0
5	4	0
6	6	0
10	8	0
24	7	1

Root MSE	0.94868	R2 乗	0.9937
従属変数の平均	9.60000	調整済み R2 乗	0.9874
変動係数	9.88212		

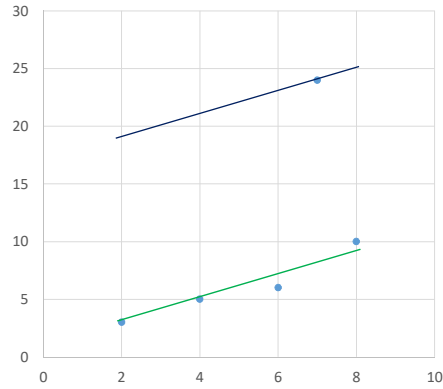
パラメータの推定					
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t
Intercept	1	0.50000	1.16190	0.43	0.7089
広告費	1	1.10000	0.21213	5.19	0.0352
ダミー	1	15.80000	1.14237	13.83	0.0052

$$\text{売上高} = 0.5 + 1.1 \times \text{広告費} + 15.8 \times \text{ダミー}$$

自由度調整済み決定係数 0.9874

56

ダミー : 0 売上高 = 0.5 + 1.1 × 広告費  
 ダミー : 1 売上高 = 16.3 + 1.1 × 広告費



57

◇満足度調査

男性19名  
女性21名

<従属変数>  
満足度

<説明変数>

- ・機能
- ・デザイン
- ・性別  
(男性1 女性0)

満足度	機能	デザイン	男性 ダミー	満足度	機能	デザイン	男性 ダミー
5	5	4	1	5	2	5	0
4	4	2	1	5	2	4	0
4	4	3	1	5	4	4	0
4	3	5	1	5	3	3	0
4	3	2	1	4	5	3	0
3	4	3	1	4	2	5	0
3	3	5	1	4	3	4	0
3	3	2	1	4	5	3	0
3	3	4	1	4	3	5	0
3	3	3	1	3	2	4	0
3	3	3	1	3	1	4	0
3	3	5	1	3	5	2	0
3	3	3	1	3	5	3	0
2	2	4	1	3	3	2	0
2	2	3	1	2	3	3	0
2	2	3	1	2	2	3	0
2	2	3	1	2	3	2	0
1	1	3	1	1	1	3	0
1	1	4	1	1	3	2	0
5	5	5	0	1	4	1	0

58

## 重回帰分析の結果

Root MSE	0.91407	R2 乗	0.4813
従属変数の平均	3.10000	調整済み R2 乗	0.4533
変動係数	29.48628		

パラメータの推定						
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t	標準 推定値
Intercept	1	-0.74452	0.67665	-1.10	0.2783	0
機能	1	0.58538	0.12661	4.62	<.0001	0.55720
デザイン	1	0.62340	0.14172	4.40	<.0001	0.53013

有意水準5%において、「機能」、「デザイン」の偏回帰係数のP値は共に小さい。

説明変数に男性ダミー（男性：1、女性：0）を追加

Root MSE	0.91812	R2 乗	0.4908
従属変数の平均	3.10000	調整済み R2 乗	0.4484
変動係数	29.61670		

パラメータの推定						
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t	標準 推定値
Intercept	1	-0.58665	0.70629	-0.83	0.4117	0
機能	1	0.57186	0.12823	4.46	<.0001	0.54433
デザイン	1	0.62253	0.14235	4.37	0.0001	0.52939
男性ダミー	1	-0.24083	0.29317	-0.82	0.4168	-0.09853

## データ（男性のみ）の場合

Root MSE	0.42484	R2 乗	0.8541
従属変数の平均	2.89474	調整済み R2 乗	0.8358
変動係数	14.67616		

パラメータの推定						
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t	標準 推定値
Intercept	1	0.05533	0.47289	0.12	0.9083	0
機能	1	0.95629	0.09883	9.68	<.0001	0.92526
デザイン	1	0.03608	0.10498	0.34	0.7356	0.03287

男性は、機能重視

## データ（女性のみ）の場合

Root MSE	0.89023	R2 乗	0.6274
従属変数の平均	3.28571	調整済み R2 乗	0.5860
変動係数	27.09389		

パラメータの推定						
変数	自由度	パラメータ 推定値	標準誤差	t 値	Pr >  t	標準 推定値
Intercept	1	-1.23401	0.88924	-1.39	0.1822	0
機能	1	0.42718	0.15695	2.72	0.0140	0.40593
デザイン	1	0.95315	0.17871	5.33	<.0001	0.79548

女性は、機能、デザイン共に重視  
機能 < デザイン

男性は、機能重視

女性は、機能、デザイン共に重視

性別によって異なる。



交互作用

63

Data Science Institute

## ロジスティック回帰分析

従属変数 : 2値データ(0, 1)

従属変数 ← 説明変数

説明変数を用いて、発生確率を求める。

$$\text{回帰式} : \log_e \left( \frac{Y}{1-Y} \right) = a + b_1 X_1 + \dots$$

Y 成約    1-Y 不成約

log<sub>e</sub> : 自然対数(底がe)

64

Data Science Institute

## ネイピア数(e)

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e$$

利息の計算（1年後の元利）

- ・年100%       $(1+1)^1 = 2$       2倍になる
- ・半年50%     $(1+0.5)^2 = 2.25$     2.25倍
- ・3カ月25%    $(1+0.25)^4 = 2.44$     2.44倍
- ・1カ月8.33%  $(1+0.083)^{12} = 2.61$  2.61倍



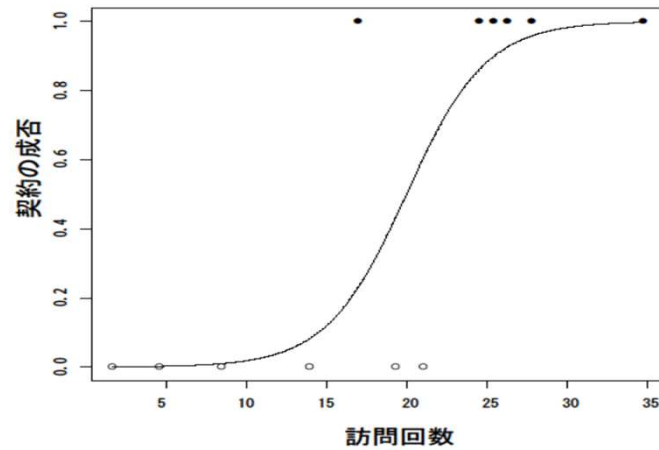
e (2.713...)

65

	No.	契約	訪問回数	No.	契約	訪問回数
	1	0	24	14	1	12
	2	0	18	15	1	6
(従属変数)	3	0	15	16	1	6
成約 : 0	4	0	16	17	1	12
不成約 : 1	5	0	10	18	1	12
	6	0	26	19	1	18
(説明変数)	7	0	2	20	1	3
訪問回数	8	0	24	21	1	8
	9	0	18	22	1	9
	10	0	22	23	1	12
	11	0	3	24	1	6
	12	1	6	25	1	8
	13	1	15	26	1	12

66

## 訪問回数と契約成否



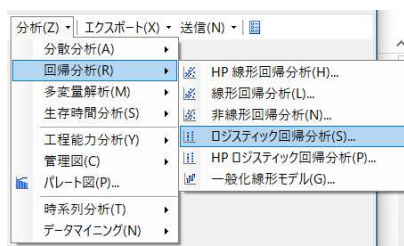
67

## ロジスティック回帰分析 (SAS EG)

EG

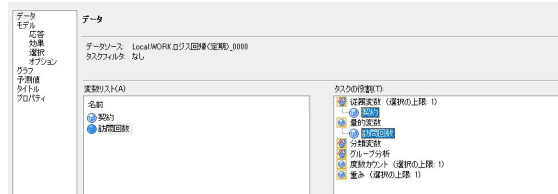
1. データを入力。
2. 「分析」-「回帰分析」-「ロジスティック回帰分析」を選択。

	契約	訪問回数
1	0	24
2	0	18
3	0	15
4	0	16
5	0	10
6	0	26
7	0	2
8	0	24
9	0	18
10	0	22
11	0	3
12	1	6
13	1	15
14	1	12
15	1	6
16	1	6
17	1	12
18	1	12
19	1	10



68

3.「データ」をクリックし、契約を「従属変数」、訪問回数を「量的変数」に設定する。



4.「効果」をクリックし、訪問回数を「主効果」に設定する。



## ロジスティック回帰分析の結果

最尤推定値の分析					
パラメータ	自由度	推定値	標準誤差	Wald 乗	Pr > ChiSq
Intercept	1	-2.5009	1.0993	5.1754	0.0229
訪問回数	1	0.1719	0.0780	4.8572	0.0275

オッズ比の推定			
効果	点推定	95% Wald 信頼限界	
訪問回数	1.188	1.019	1.384

◇ 偏回帰係数の検定

訪問回数 P値=0.0275 < 0.05

◇ オッズ比 : 1.188 > 1

信頼区間内に1を含まない

予測値 (Y) を求める。

$$\text{loge} \left( \frac{Y}{1-Y} \right) = -2.5 + 0.1719 \times X$$

$$Y = \frac{1}{1 + e^{-(-2.5 + 0.1719 \times X)}}$$



71

### 予測値 (確率)

No.	契約	回数	確率	No.	契約	回数	確率
1	0	24	0.835	14	1	12	0.392
2	0	18	0.644	15	1	6	0.187
3	0	15	0.519	16	1	6	0.187
4	0	16	0.562	17	1	12	0.392
5	0	10	0.314	18	1	12	0.392
6	0	26	0.877	19	1	18	0.644
7	0	2	0.104	20	1	3	0.121
8	0	24	0.835	21	1	8	0.245
9	0	18	0.644	22	1	9	0.278
10	0	22	0.783	23	1	12	0.392
11	0	3	0.121	24	1	6	0.187
12	1	6	0.187	25	1	8	0.245
13	1	15	0.519	26	1	12	0.392

72

### オッズ比

$$\text{回帰式} : \log_e\left(\frac{Y}{1-Y}\right) = -2.5 + 0.1719 \times \text{訪問回数}$$

$$\frac{Y}{1-Y} = \text{EXP}(-2.5 + 0.1719 \times \text{訪問回数})$$

◇訪問回数1回と2回の比較

$$\frac{\text{EXP}(-2.5 + 0.1719 \times 2\text{回})}{\text{EXP}(-2.5 + 0.1719 \times 1\text{回})}$$
$$= \frac{\text{EXP}(-2.1562)}{\text{EXP}(-2.3281)} = \frac{0.115764}{0.097481} = 1.188$$

契約の確率が1.188倍になる。

73

Data Science Institute

### ◇疾患の有無と2種類の検査結果

	有無	検査A	検査B	有無	検査A	検査B
(従属変数)	0	27	25	1	40	25
無 : 0	0	14	18	1	20	35
有 : 1	0	22	10	1	12	36
	0	31	10	1	30	29
	0	25	20	1	22	28
(説明変数)	0	32	17	1	18	31
検査A	0	24	15	1	25	26
検査B	0	12	12	1	28	18
	0	16	23	1	35	27
	0	10	12	1	32	24
データ	0	15	9	1	24	25
26組	0	18	8	1	20	30

74

Data Science Institute

## ロジスティック回帰分析の結果

EG

### データ26組のとき

パラメータ	自由度	推定値	標準誤差	Wald カイ 2 乗	Pr > ChiSq
Intercept	1	15.9314	7.9050	4.0617	0.0439
検査A	1	-0.2099	0.1775	1.3982	0.2370
検査B	1	-0.4798	0.2011	5.6915	0.0170

効果	点推定	95% Wald 信頼限界	
検査A	0.811	0.572	1.148
検査B	0.619	0.417	0.918

#### ◇オッズ比

検査A 0.811 信頼区間内に1を含む  
 検査B 0.619 信頼区間内に1を含まない

75

## 予測値

有無	検査A	検査B	確率	有無	検査A	検査B	確率
0	27	25	0.151	1	40	25	0.011
0	14	18	0.987	1	20	35	0.006
0	22	10	0.999	1	12	36	0.021
0	31	10	0.990	1	30	29	0.014
0	25	20	0.748	1	22	28	0.107
0	32	17	0.742	1	18	31	0.062
0	24	15	0.976	1	25	26	0.143
0	12	12	1.000	1	28	18	0.805
0	16	23	0.823	1	35	27	0.013
0	10	12	1.000	1	32	24	0.091
0	15	9	1.000	1	24	25	0.250
0	18	8	1.000	1	20	30	0.065

76

## ロジスティック回帰分析の結果

EG

データ74組のとき

パラメータ	自由度	推定値	標準誤差	Wald カイ2乗	Pr > ChiSq
Intercept	1	15.9314	4.5639	12.1851	0.0005
A	1	-0.2099	0.1025	4.1945	0.0406
B	1	-0.4798	0.1161	17.0745	<.0001

効果	点推定	95% Wald 信頼限界	
A	0.811	0.663	0.991
B	0.619	0.493	0.777

◇オッズ比

検査A 0.811 信頼区間内に1を含まない  
 検査B 0.619 信頼区間内に1を含まない

77

Data Science Institute

## ロジスティック回帰分析の解釈

説明変数	回帰係数	オッズ比	オッズ比信頼限界	
			下限	上限
男性ダミー	2.00	7.39	5.24	9.54
年齢	-0.05	0.95	0.93	0.97
利用月数	0.20	1.22	1.15	1.29
DM送付数	0.16	1.17	0.55	1.79

- 男性は女性に比べて約7.39倍来店しやすい
- 年齢が1歳上がるごとに来店確率は0.95倍に小さくなる。
- 利用月数は1カ月増えるごとに来店確率は1.22倍になる。
- DM送付数は、信頼区間に1を含むので判断ができない。

78

Data Science Institute

### ◇不整脈症状有無

不整脈症状有無	喫煙有無	飲酒有無	ギャンブル嗜好
1	1	1	1
1	1	1	1
1	1	1	1
1	1	0	1
1	1	0	1
1	1	0	0
1	0	1	0
1	0	0	0
0	1	0	1
0	1	0	0
...	...	...	...

79

### ロジスティック回帰分析の結果

最尤推定値の分析					
パラメータ	自由度	推定値	標準誤差	Wald カイ2乗	Pr > ChiSq
Intercept	1	2.3743	0.7650	9.6322	0.0019
喫煙有無	1	-2.6428	1.1039	5.7312	0.0167
飲酒有無	1	-1.9484	0.9857	3.9071	0.0481
ギャンブル嗜好	1	-0.5035	1.0992	0.2098	0.6469

オッズ比の推定			
効果	点推定	95% Wald 信頼限界	
喫煙有無	0.071	0.008	0.619
飲酒有無	0.143	0.021	0.984
ギャンブル嗜好	0.604	0.070	5.212

80

## 2群以上の関係を把握する方法

### ◇重回帰分析

- 従属変数、説明変数（独立変数）
- 回帰モデルのチェック
  - t 検定 p値（有意確率）
- 自由度調整済み決定係数
- 標準偏回帰係数
- 説明変数相互の積率相関係数
- ダミーデータを用いた回帰分析
- 交互作用

### ◇ロジスティック回帰分析

- 従属変数（0 or 1）
- オッズ比