

Regression Analysis

回歸分析

What is regression Analysis?

回帰分析とは、求めたい要素の値に対し、他の要素がどの程度影響を与えているかを分析する手法です。

Say what?

Examples

independent variable
(説明変数)

Advertising
広告費



dependent variable
(目的変数)

revenue
利益

independent variable
(説明変数)

dependent variable
(目的変数)

walking minutes to station
駅まで徒歩何分



rent
家賃

independent variable
(説明変数)

dependent variable
(目的変数)

hours spent partying



grades

independent variable
(説明変数)

student height



dependent variable
(目的変数)

grades

independent variable
(説明変数)

hours practiced



dependent variable
(目的変数)

points per game

Types of regressions

- 単回帰分析
- 重回帰分析
- ロジスティック回帰分析

Let's try it in Excel

Data

検索条件の設定・変更

選択中の地域 / 路線・駅

千葉県

JR常磐線
南柏

路線を変更 駅を変更

探し方の変更・選択の解除

物件種別

借りる(賃貸)

☒ マンション ☒ アパート

☒ 一戸建て ☐ もっと見る

買う

物件の条件

賃料

下限なし ~ 上限なし

☐ 共益費/管理費を含む

☐ 礼金なし

☐ 敷金なし

専有面積

下限なし ~ 上限なし

間取り

☐ ワンルーム ☐ 1K

☐ 1DK ☐ 1LDK

☐ 2K ☐ 2DK

☒ 2LDK ☐ 3K

☐ 3DK ☐ 3LDK

☐ 4K ☐ 4DK

☐ 4LDK以上

駅徒歩分

20分以内

☐ バス乗車時間含む

通勤・通学駅までの所要時間

通勤・通学時間の設定

賃貸(賃貸マンション・アパート)

この条件を保存する

この条件の新着物件を通知

LINEで受取る

メールで受取る

リスト表示

地図表示

46件 (総物件数: 158件)

並び替え

おすすめ順

1 2 次へ

PR 賃貸アパート ホワイトハイツ 0201 / 2階



賃料/管理費等

5.5万円 / -

所在地

千葉県流山市名都借232-1

交通

JR常磐線 南柏駅 徒歩13分

専有面積/間取り

49m² / 2LDK

PR 賃貸アパート ホワイトハイツ 0203 / 2階



賃料/管理費等

5.5万円 / -

所在地

千葉県流山市名都借232-1

交通

JR常磐線 南柏駅 徒歩13分

専有面積/間取り

49m² / 2LDK

チェックした物件を
(一度に20件まで)

お気に入りに追加

お問合せ(無料)

賃貸テラスハウス ウイング ヴィラ



所在地

千葉県柏市今谷上町32-91

交通

JR常磐線 南柏駅 徒歩10分
東武野田線 新柏駅 徒歩14分

築年数/階数

7年 / 2階建

Data in Excel

	A	B	C	D
1	rent	minutes	size	age
2	9.2	10	50	7
3	10.2	9	59	6
4	8.9	19	55	1
5	7	20	50	15
6	6.6	16	51	22
7	12.5	4	61	6
8	6	16	49	31
9	12.5	4	61	6
10	8.7	15	54	14
11	8.6	13	57	17
12	7.5	15	62	20
13	7	20	50	15
14	10.4	20	60	11
15	7.7	7	51	27
16	6.8	11	55	20
17	6	17	51	33
18	7.7	7	51	27
19	5.5	13	49	42
20	7	20	54	24
21	7.5	13	52	31
22	8.6	15	62	22
23	8.7	14	64	16
24	10.5	20	53	15
25	8	14	54	22
26	8.4	11	57	23
27	7.9	11	57	25
28	7.7	15	57	21
29	10.5	20	53	15
30	7.4	17	53	29
31	9	12	54	14
32	8.6	15	62	22

何が知りたい？

目的変数と説明変数を決める

independent variable
(説明変数)

dependent variable
(目的変数)

minutes

駅まで徒歩何分



size

面積



rent

家賃

age

築年数



まずは単回帰分析

independent variable
(説明変数)

dependent variable
(目的変数)

minutes
駅まで徒歩何分



rent
家賃

仮設検定

帰無仮説 【null hypothesis】

H_0 = 駅までの徒歩分数は家賃に全く影響がない

$$\beta_1 = 0$$

対立仮説 【alternative hypothesis】

H_1 = 駅までの徒歩分数は家賃に影響がある

$$\beta_1 \neq 0$$

データで見ると？

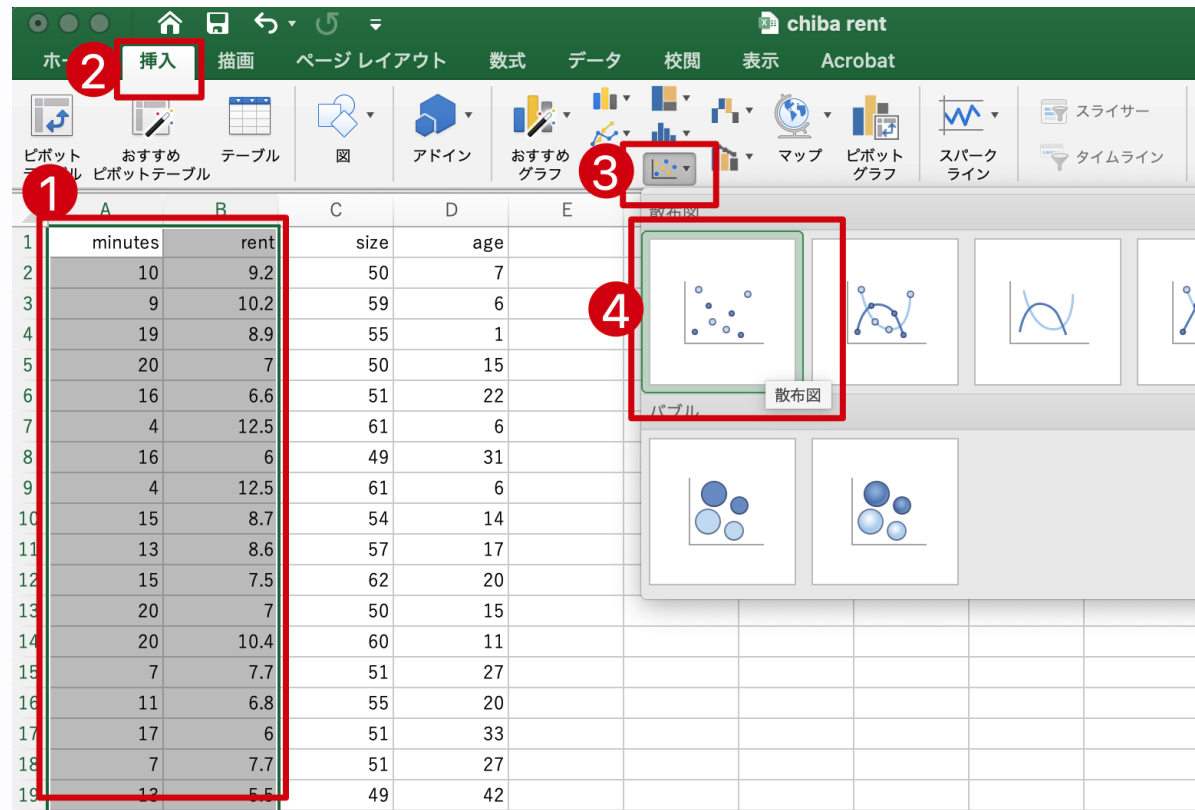
	A	B	C	D	E	F
1	minutes	rent	size	age		
2	10	9.2	50	7		
3	9	10.2				
4	15	8.9				
5	20	7				
6	16	6.6	51	22		
7	4	12.5				
8	16	8				
9	4	12.5				
10	15	8.7	54	14		
11	13	8.6	57	17		
12	15	7.5	62	20		
13	20	7	50	15		

independent variable
説明変数 **X**

dependent variable
目的変数 **Y**

First, a scatter plot

まずは散布図を作ろう



Book1

シートを検索

ホーム挿入描画ページレイアウトデータ校閲表示Acrobat

外部データの取り込み

すべて更新

クエリ 接続
プロパティ
リンクの編集

A Z
Z A
並べ替え
フィルター

クリア
再適用
詳細設定

区切り位置

フラッシュフィル

重複を削除

入力規則

統合

What-If 分析

グループ化

グループ解除

小計

分析ツール

データ分析

回復済みのブックを開きますか? 最近の変更内容が保存されました。中断したところから作業を続けますか?

はいいいえ

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	rent	minutes	size	age										
2	9.2	10												
3	10.2	9												
4	8.9	19												
5	7	20												
6	6.6	16												
7	12.5	4												
8	6	16												
9	12.5	4												
10	8.7	15												
11	8.6	13												
12	7.5	15												
13	7	20		50	15									
14	10.4	20		60	11									
15	7.7	7		51	27									

データ分析

分析ツール

乱数発生

順位と百分位数

回帰分析

サンプリング

t 検定: 一対の標本による平均の検定

t 検定: 等分散を仮定した 2 標本による検定

OK

キャンセル

ヘルプ

Regression options

回帰分析

入力元

入力 Y 範囲:

\$B\$1:\$B\$32

入力 X 範囲:

\$A\$1:\$A\$32

☒ ラベル

☐ 定数に 0 を使用

☐ 有意水準

95 %

OK

キャンセル

ヘルプ

出力オプション

☐ 一覧の出力先:

☒ 新規ワークシート:

☐ 新規ブック

残差

☐ 残差

☐ 残差グラフの作成

☐ 標準化された残差

☐ 観測値グラフの作成

正規確率

☐ 正規確率グラフの作成

Regression output

	A	B	C	D	E	F	G	H	I
1	概要								
2									
3	回帰統計								
4	重相関 R	0.34046331							
5	重決定 R2	0.11591527							
6	補正 R2	0.08542959							
7	標準誤差	1.63077684							
8	観測数	31							
9									
10	分散分析表								
11		自由度	変動	分散	観測された分散	有意 F			
12	回帰	1	10.1119243	10.1119243	3.80228566	0.06090971			
13	残差	29	77.1235596	2.65943309					
14	合計	30	87.2354839						
15									
16		係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
17	切片	10.0908723	0.94352867	10.6948231	1.407E-11	8.16113945	12.0206051	8.16113945	12.0206051
18	minutes	-0.1252126	0.06421338	-1.949945	0.06090971	-0.2565437	0.00611855	-0.2565437	0.00611855

なんだこりゃ？

まずは回帰式を作ろう

$$Y = a + bX$$

家賃（万円） = $a + b \times$ 駅まで徒歩分数

	A	B	C	D	E	F	G	H	I
1	概要								
2									
3	回帰統計								
4	重相関 R	0.34046331							
5	重決定 R2	0.11591527							
6	補正 R2	0.08542959							
7	標準誤差	1.63077684							
8	観測数	31							
9									
10	分散分析表								
11		自由度	変動	分散	観測された分散	有意 F			
12	回帰	1	10.1119243	10.1119243	3.80228566	0.06090971			
13	残差	29	77.1235596	2.65943309					
14	合計	30	87.2354839						
15									
16		係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
17	切片	10.0908723	0.94352867	10.6948231	1.407E-11	8.16113945	12.0206051	8.16113945	12.0206051
18	minutes	-0.1252126	0.06421338	-1.949945	0.06090971	-0.2565437	0.00611855	-0.2565437	0.00611855

家賃（万円） = $a + b \times \text{駅まで徒歩分数}$

$$\text{家賃（万円）} = 10.09 - 0.125 \times \text{駅まで徒歩分数}$$

すなわち

南柏の駅からの徒歩分数が1分増えるごとに
家賃が1250円減る

では「a」（切片）は何？

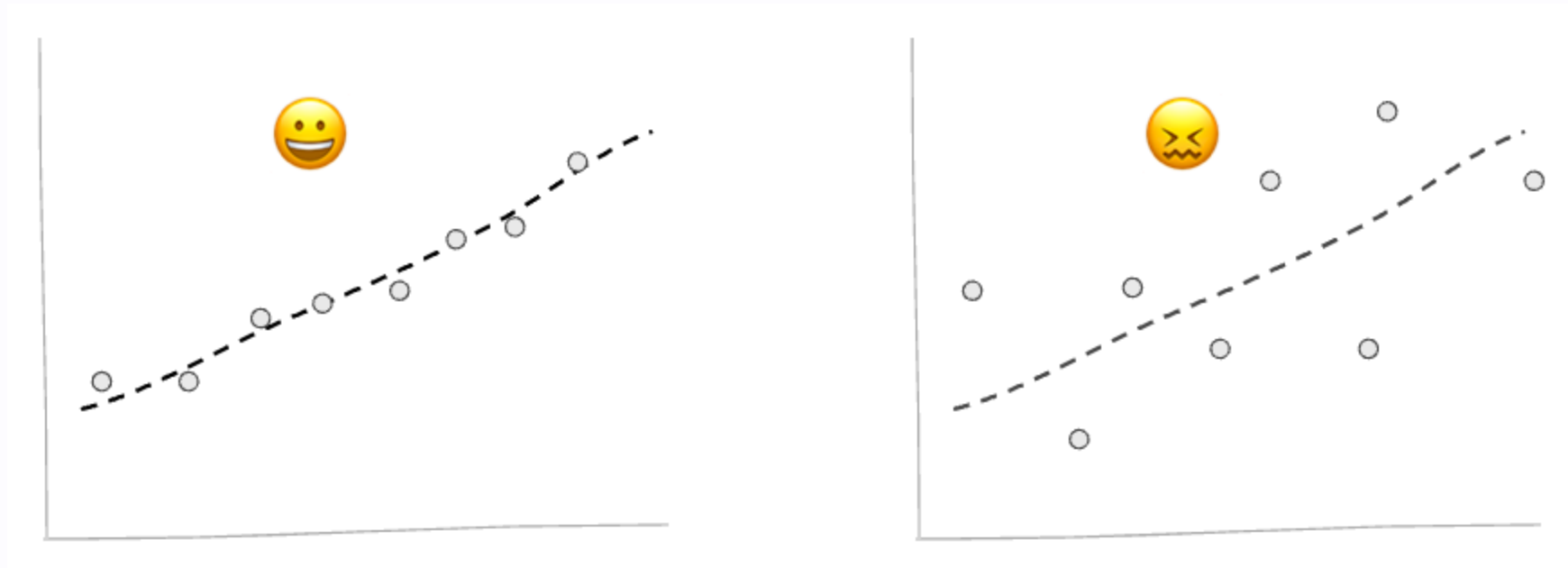
これは「x」がゼロの時の数値。

すなわち、駅から徒歩分数がゼロの賃貸（ありますか？）の
場合、家賃は10.09万円

R-Squared: How well does it fit?

$R^2 = 1$ (perfect fit 😊)

$R^2 = 0$ (bad fit 😞)



ではこの分析のR2乗は？

	A	B	C	D	E	F	G	H	I
1	概要								
2									
3	回帰統計								
4	重相関 R	0.34046331							
5	重決定 R2	0.11591527	😞						
6	補正 R2	0.08542959							
7	標準誤差	1.63077684							
8	観測数	31							
9									
10	分散分析表								
11		自由度	変動	分散	観測された分散	有意 F			
12	回帰	1	10.1119243	10.1119243	3.80228566	0.06090971			
13	残差	29	77.1235596	2.65943309					
14	合計	30	87.2354839						
15									
16		係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
17	切片	10.0908723	0.94352867	10.6948231	1.407E-11	8.16113945	12.0206051	8.16113945	12.0206051
18	minutes	-0.1252126	0.06421338	-1.949945	0.06090971	-0.2565437	0.00611855	-0.2565437	0.00611855

What about the P value?

P値で説明変数（徒歩分数）が目的変数（家賃）に対して関係があるかどうかを確認する

P値が0.05未満であれば、
「関係性がある」と判断ができる。

要するに、この関係性はランダムではないので、帰無仮説
【null hypothesis】をrejectすることができる。

でも...

	A	B	C	D	E	F	G	H	I
1	概要								
2									
3	回帰統計								
4	重相関 R	0.34046331							
5	重決定 R2	0.11591527							
6	補正 R2	0.08542959							
7	標準誤差	1.63077684							
8	観測数	31							
9									
10	分散分析表								
11		自由度	変動	分散	観測された分散	有意 F			
12	回帰	1	10.1119243	10.1119243	3.80228566	0.06090971			
13	残差	29	77.1235596	2.65943309					
14	合計	30	87.2354839						
15									
16		係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%
17	切片	10.0908723	0.94352867	10.6948231	1.407E-11	8.16113945	12.0206051	8.16113945	12.0206051
18	minutes	-0.1252126	0.06421338	-1.949945	0.06090971	-0.2565437	0.00611855	-0.2565437	0.00611855

In that case...

他の説明変数で試してみよう！