

# Excelで学ぶデータ分析

## 第4回

### 「相関・回帰分析」



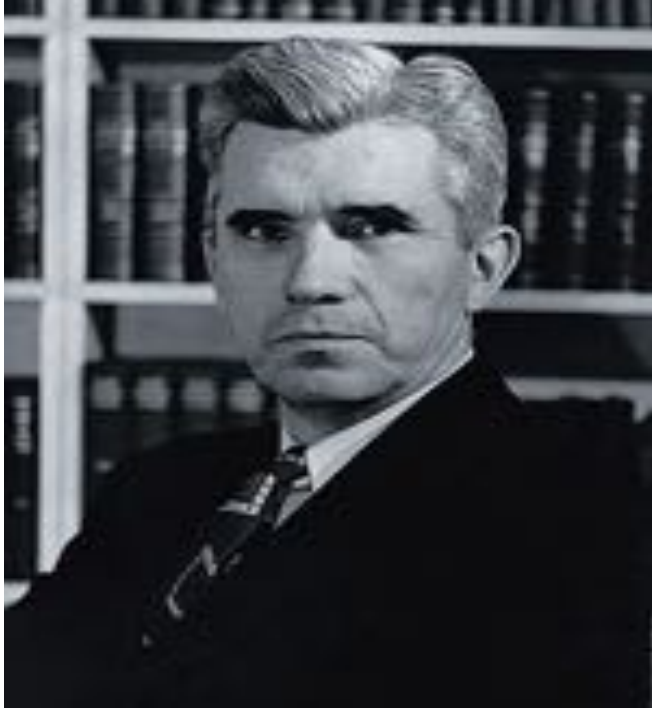
和から株式会社

# 前回のReview

---

- 「検定」の考え方を理解する

# Joseph Banks Rhine(1895-1980)



J.B.ラインはデューク大学の教授  
超心理学（テレパシー、超能力など）の分  
野を築いた

ライン教授は科学者として初めて、統計検  
定を使って、超能力の存在を証明しようと  
試みた。

# Step1 : 仮説の設定



Prof Charles

超能力があるのだ

ふむ、検証してみよう

- 帰無仮説(  $H_0$  )

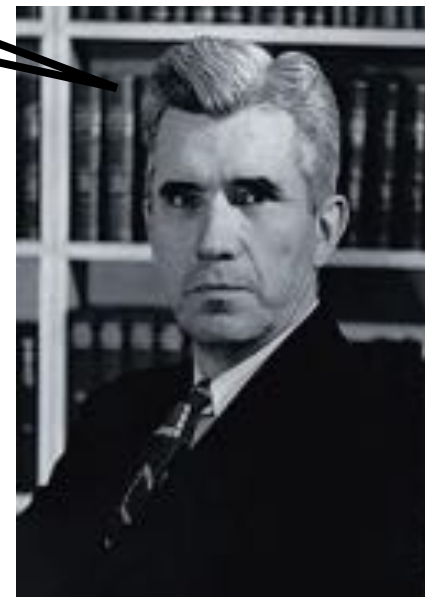
- 対立仮説(  $H_1$  )

帰無仮説

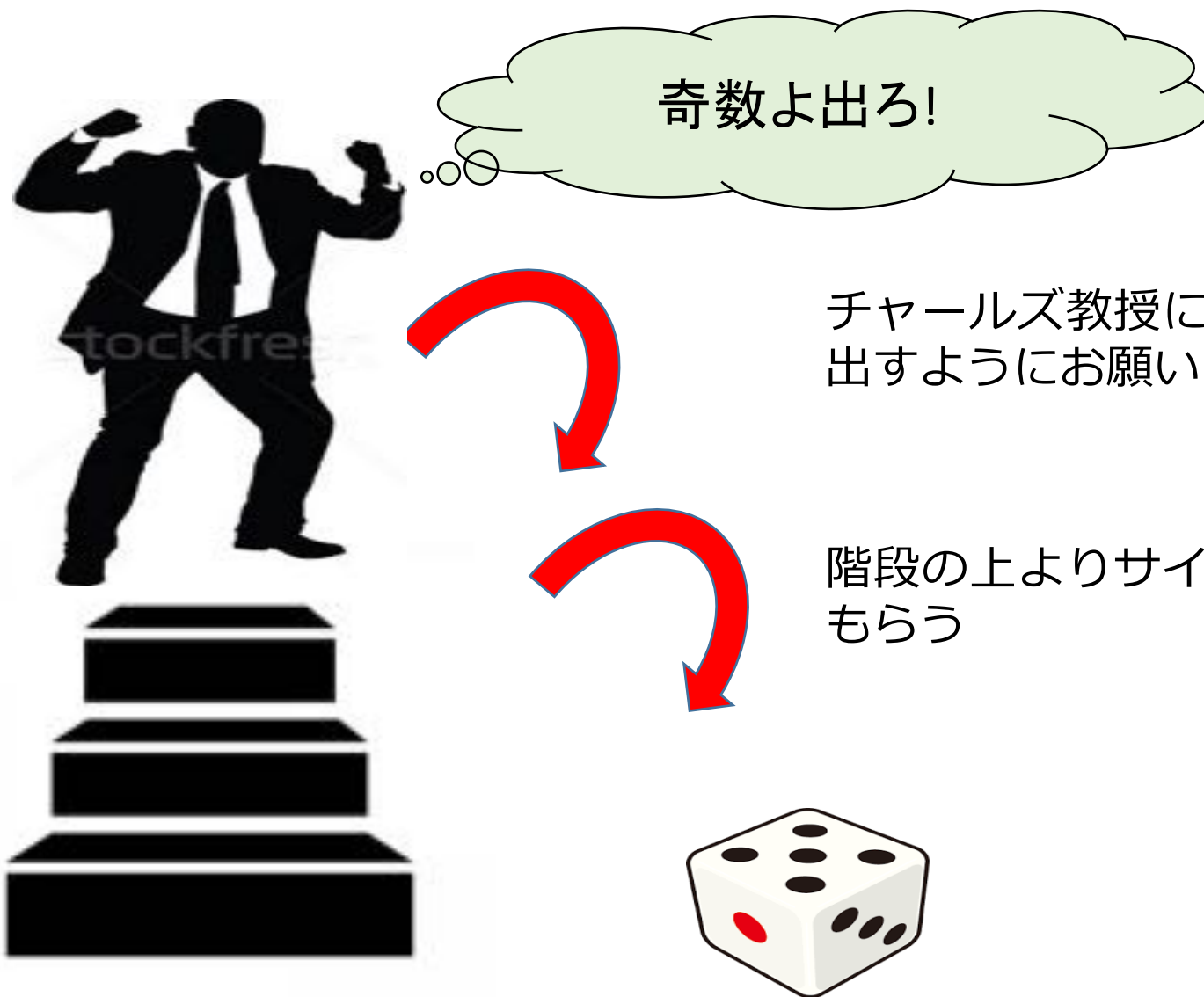
$H_0$  “チャールズ教授は超能力を持たない”

対立仮説

$H_1$  “チャールズ教授は超能力を持つ”



## Step2：実験の設計



チャールズ教授に超能力を使って、奇数を出すようにお願いします

階段の上よりサイコロを1000回投げてもらう

## Step3 : データの収集



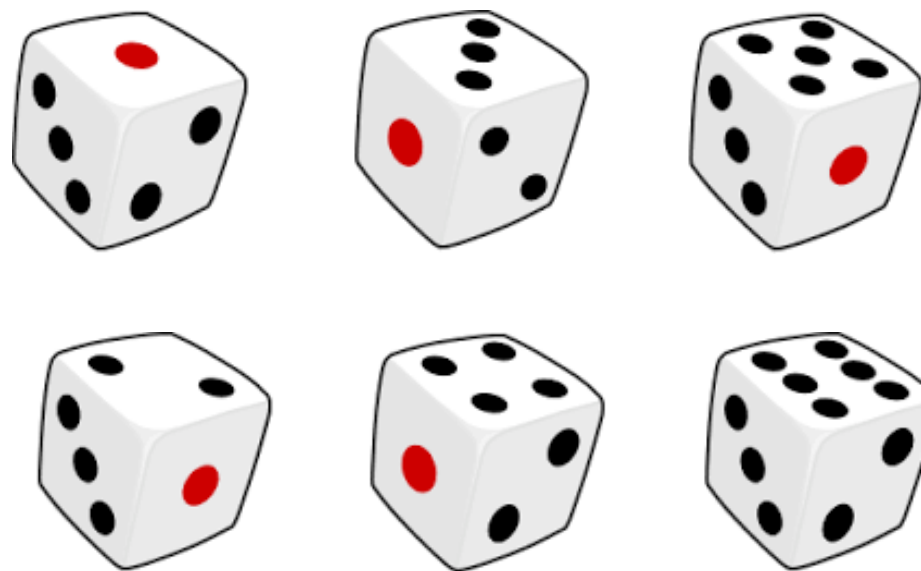
奇数  
542 回



偶数  
458 回

このデータはチャールズ教授に超能力があることを証明するに十分な根拠となるのか？

## Step4 : 結果の分析



奇数  
542 回

偶数  
458 回

帰無仮説が正しいと仮定する

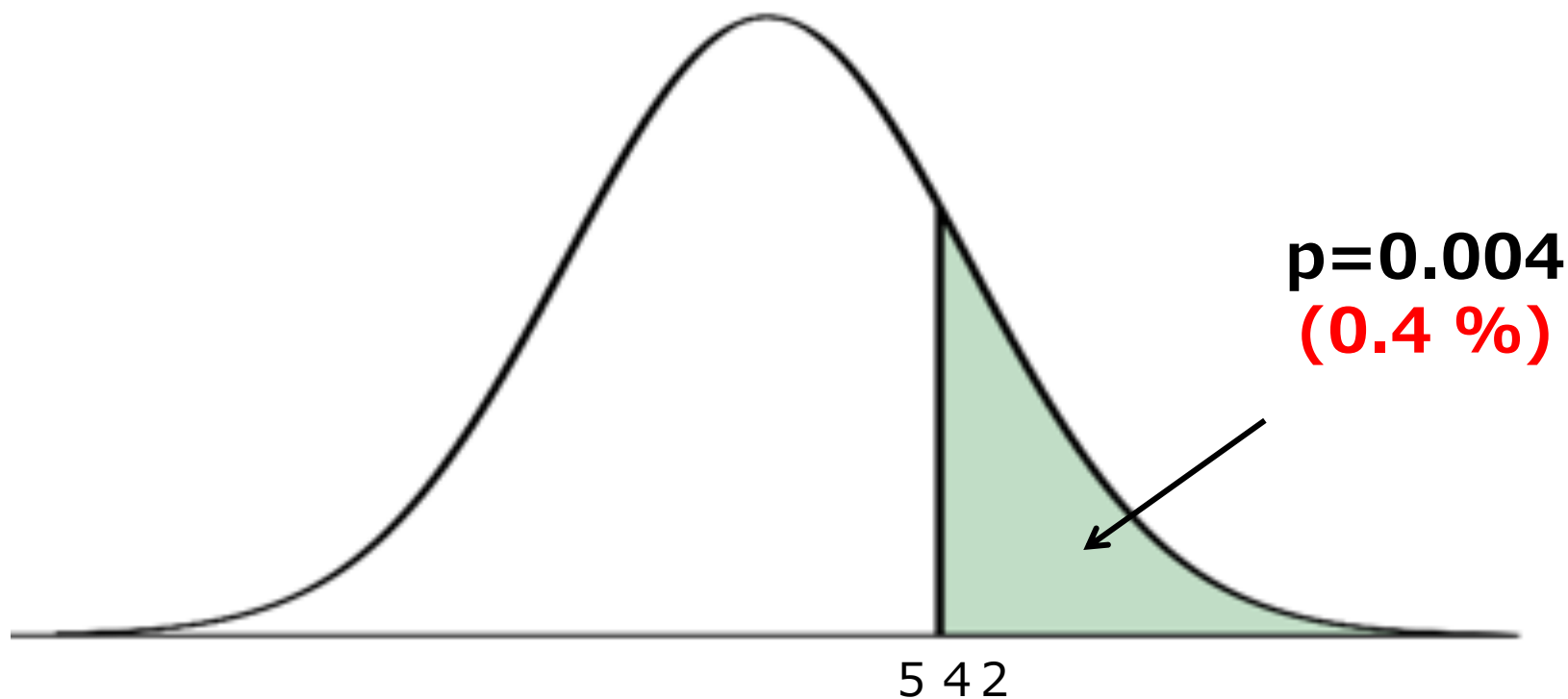
- ↔ チャールズ教授に超能力がないと仮定してるのと同じ
- ↔ 奇数と偶数の出る確率が同じと仮定してるのと同じ

# 結果の分析（二項分布・確率）

P値 = 542回以上奇数が出る確率は？

**EXCEL**

=1-binomdist(541,1000,0.5,TRUE)





## Step4 : 結論 [1]

チャールズ教授は542回奇数を出した

⇔ チャールズ教授に超能力を持たないと仮定

⇔ 1 0 0 0回サイコロを投げて542回以上奇数が出る確率は0.4%  
(小さい確率と考える)

⇔ 通常では起こりえないことが起こった

⇔ ~~帰無仮説  $H_0$   
“チャールズ教授は超能力を持たない”~~

対立仮説  $H_1$   
“チャールズ教授は超能力を持つ”

(背理法)

## Step4 : 結論 [2]

チャールズ教授は542回奇数を出した

⇔ チャールズ教授に超能力を持たないと仮定

⇔ 1 0 0 0回サイコロを投げて542回以上奇数が出る確率は0.4%  
(大きい確率と考える)

⇔ よくあることが起こった

⇔ 帰無仮説  $H_0$   
“チャールズ教授は超能力を持たない”

~~対立仮説  $H_1$   
“チャールズ教授は超能力を持つ”~~

# 有意水準(判断基準)

有意水準( $\alpha$ )は検定において帰無仮説を設定した時にその帰無仮説を棄却する基準となる確率

$\alpha=5\%$  が一般的には使われる

P value = 0.4% <  $\alpha=5\%$   
(奇数が542回以上出る確率)

~~帰無仮説  $H_0$   
“チャールズ教授は超能力を持たない”~~

対立仮説  $H_1$   
“チャールズ教授は超能力を持つ”

# 有意水準(判断基準)

有意水準( $\alpha$ )は検定において帰無仮説を設定した時にその帰無仮説を棄却する基準となる確率

$\alpha=5\%$  が一般的には使われる

P value = 11% >  $\alpha=5\%$   
(奇数が520回以上出る確率)

帰無仮説  $H_0$   
“チャールズ教授は超能力を持たない”

~~対立仮説  $H_1$   
“チャールズ教授は超能力を持つ”~~

# 統計検定の手順

- ✓Step 1 : 仮説の設定（帰無仮説と対立仮説）
- ✓Step 2 : 実験の設定とデータ収集
- ✓Step 3 : 帰無仮説が正しいと仮定し、統計量からp値を求める

✓Step 4 : 結論を下す（p値と有意水準 $\alpha$  の比較）

case 1    p-値 < 有意水準 ( $\alpha$ )

~~帰無仮説  $H_0$~~

対立仮説  $H_1$

case 2    p-値 > 有意水準 ( $\alpha$ )

帰無仮説  $H_0$

~~対立仮説  $H_1$~~

# 相関分析

# データの型と関連を探る道具

ID	満足度	他者評価	プロジェクト数	月間労働時間	Work accident	退職・在職	過去5年 昇進	所属部署	給料
1019	0.36	0.47	2	136	無	退職	無	accounting	low
6830	0.68	0.51	5	158	有	在職	無	technical	medium
9653	0.53	0.64	2	109	有	在職	無	hr	medium
12208	0.78	0.87	4	228	無	退職	無	support	low
4816	0.92	0.56	4	170	無	在職	無	marketing	medium
5637	0.98	0.92	4	175	有	在職	無	IT	medium

# データの型と関連を探る道具

ID	満足度	他者評価	プロジェクト数	月間労働時間	Work accident	退職・在職	過去5年昇進	所属部署	給料
1019	0.36	0.47	2	136	無	退職	無	accounting	low
6830	0.68	0.51	5	158	有	在職	無	technical	medium
9653	0.53	0.64	2	109	有	在職	無	hr	medium
12208	0.78	0.87	4	228	無	退職	無	support	low
4816	0.92	0.56	4	170	無	在職	無	marketing	medium
5637	0.98	0.92	4	175	有	在職	無	IT	medium

↑  
質的

↑  
質的

どうやって関係があるかを調べるか？



# データの型と関連を探る道具

ID	満足度	他者評価	プロジェクト数	月間労働時間	Work accident	退職・在職	過去5年昇進	所属部署	給料
1019	0.36	0.47	2	136	無	退職	無	accounting	low
6830	0.68	0.51	5	158	有	在職	無	technical	medium
9653	0.53	0.64	2	109	有	在職	無	hr	medium
12208	0.78	0.87	4	228	無	退職	無	support	low
4816	0.92	0.56	4	170	無	在職	無	marketing	medium
5637	0.98	0.92	4	175	有	在職	無	IT	medium

	在職	退職
Accounting	19	9
HR	20	5
IT	23	11
Management	17	2
Marketing	18	7
Product_mng	19	6
RandD	24	3
Sales	111	31
Support	50	18
Technical	81	26
総計	382	118

↑  
質的

↑  
質的

どうやって関係があるかを調べるか？

クロス集計

# データの型と関連を探る道具

ID	満足度	他者評価	プロジェクト数	月間労働時間	Work accident	退職・在職	過去5年昇進	所属部署	給料
1019	0.36	0.47	2	136	無	退職	無	accounting	low
6830	0.68	0.51	5	158	有	在職	無	technical	medium
9653	0.53	0.64	2	109	有	在職	無	hr	medium
12208	0.78	0.87	4	228	無	退職	無	support	low
4816	0.92	0.56	4	170	無	在職	無	marketing	medium
5637	0.98	0.92	4	175	有	在職	無	IT	medium

↑  
量的

↑  
量的

どうやって関係があるかを調べるか？

# データの型と関連を探る道具

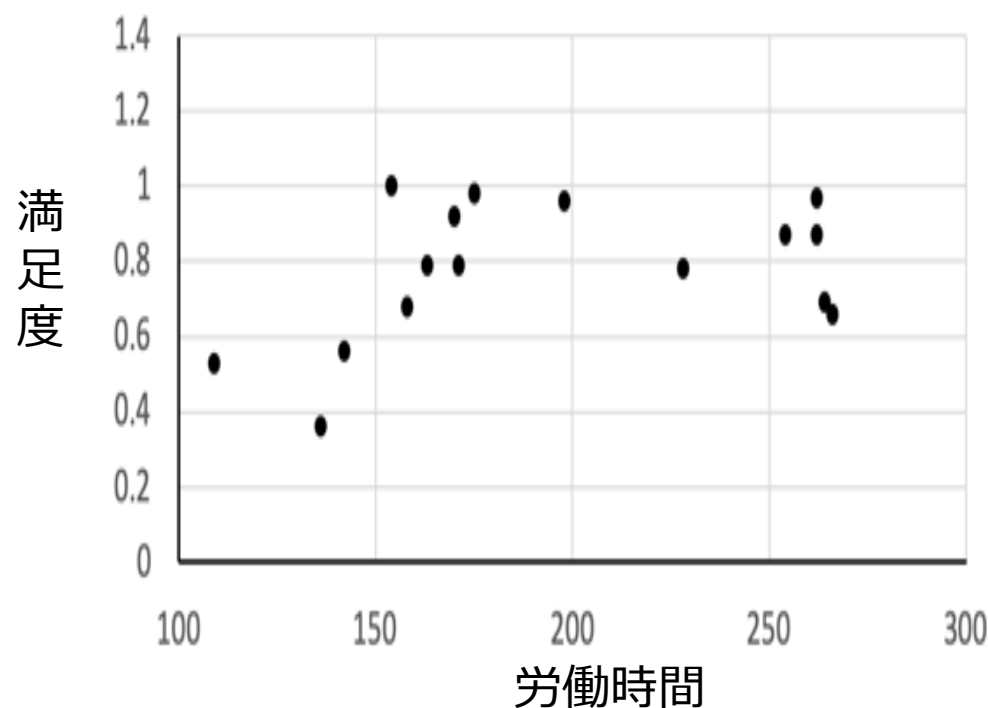
ID	満足度	他者評価	プロジェクト数	月間労働時間	Work accident	退職・在職	過去5年昇進	所属部署	給料
1019	0.36	0.47	2	136	無	退職	無	accounting	low
6830	0.68	0.51	5	158	有	在職	無	technical	medium
9653	0.53	0.64	2	109	有	在職	無	hr	medium
12208	0.78	0.87	4	228	無	退職	無	support	low
4816	0.92	0.56	4	170	無	在職	無	marketing	medium
5637	0.98	0.92	4	175	有	在職	無	IT	medium

↑  
量的

↑  
量的

どうやって関係があるかを調べるか？

**散布図**



# データ間の関係性を見たい

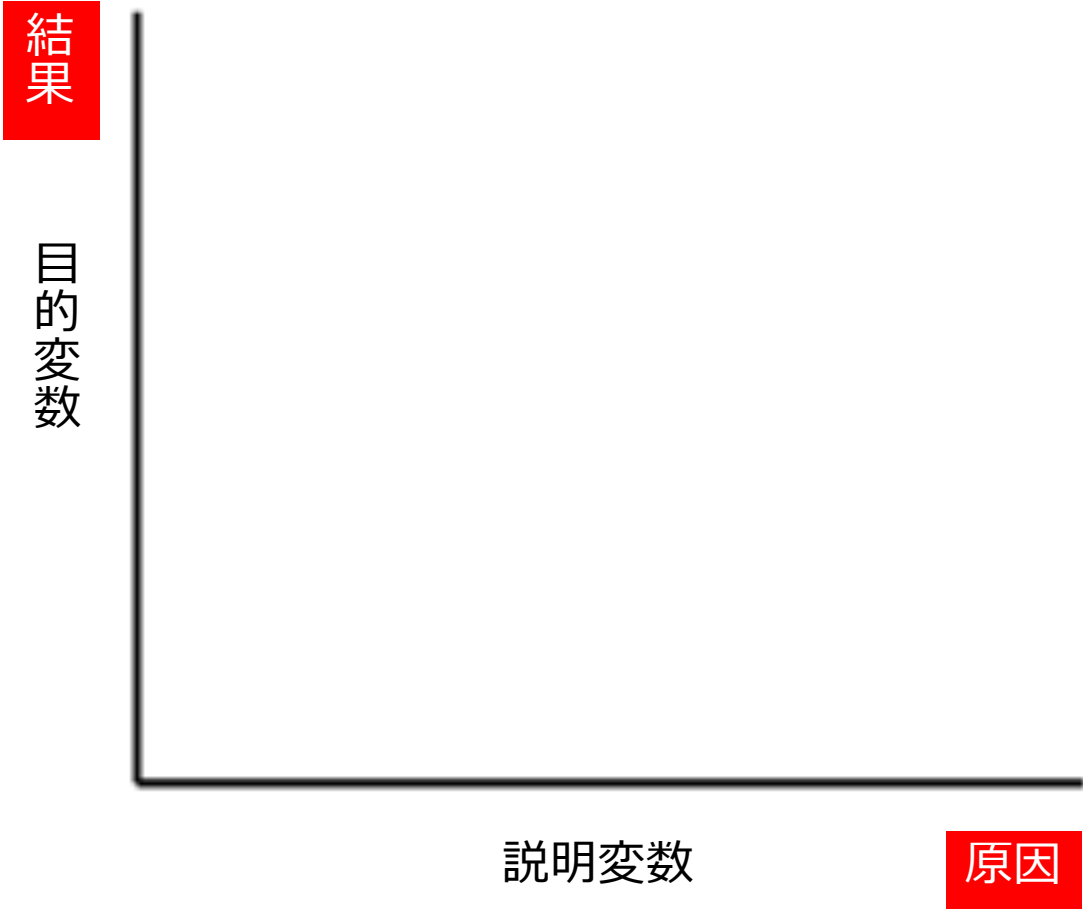
月	広告費	売上
4月	3,004	2,295
5月	3,982	5,928
6月	279,284	20,399
7月	198,374	11,245
8月	10,492	4,567
9月	78,938	8,673
10月	70,293	6,394
11月	69,283	7,124
12月	54,900	4,958
1月	112,938	8,256
2月	116,823	7,982
3月	93,847	6,504

広告費と売上は本当に関係しているのだろうか？



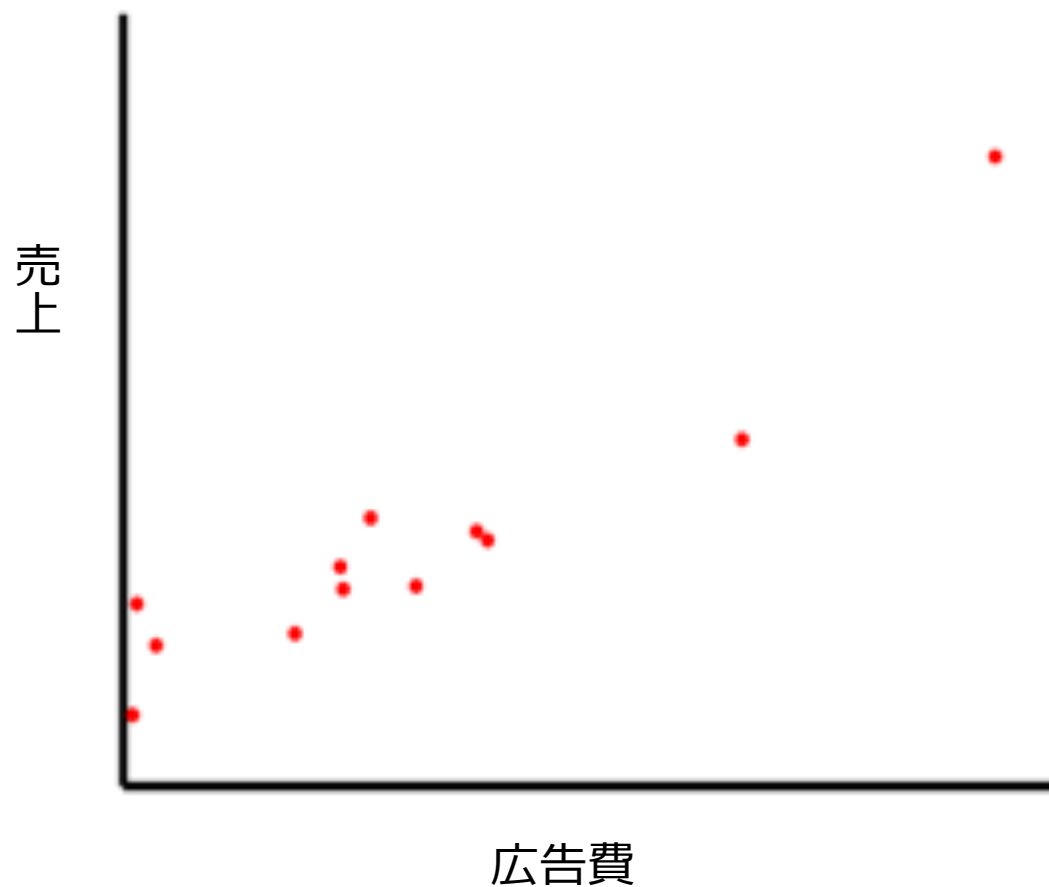
# データ間の関係性を見る

月	広告費	売上
4月	3,004	2,295
5月	3,982	5,928
6月	279,284	20,399
7月	198,374	11,245
8月	10,492	4,567
9月	78,938	8,673
10月	70,293	6,394
11月	69,283	7,124
12月	54,900	4,958
1月	112,938	8,256
2月	116,823	7,982
3月	93,847	6,504



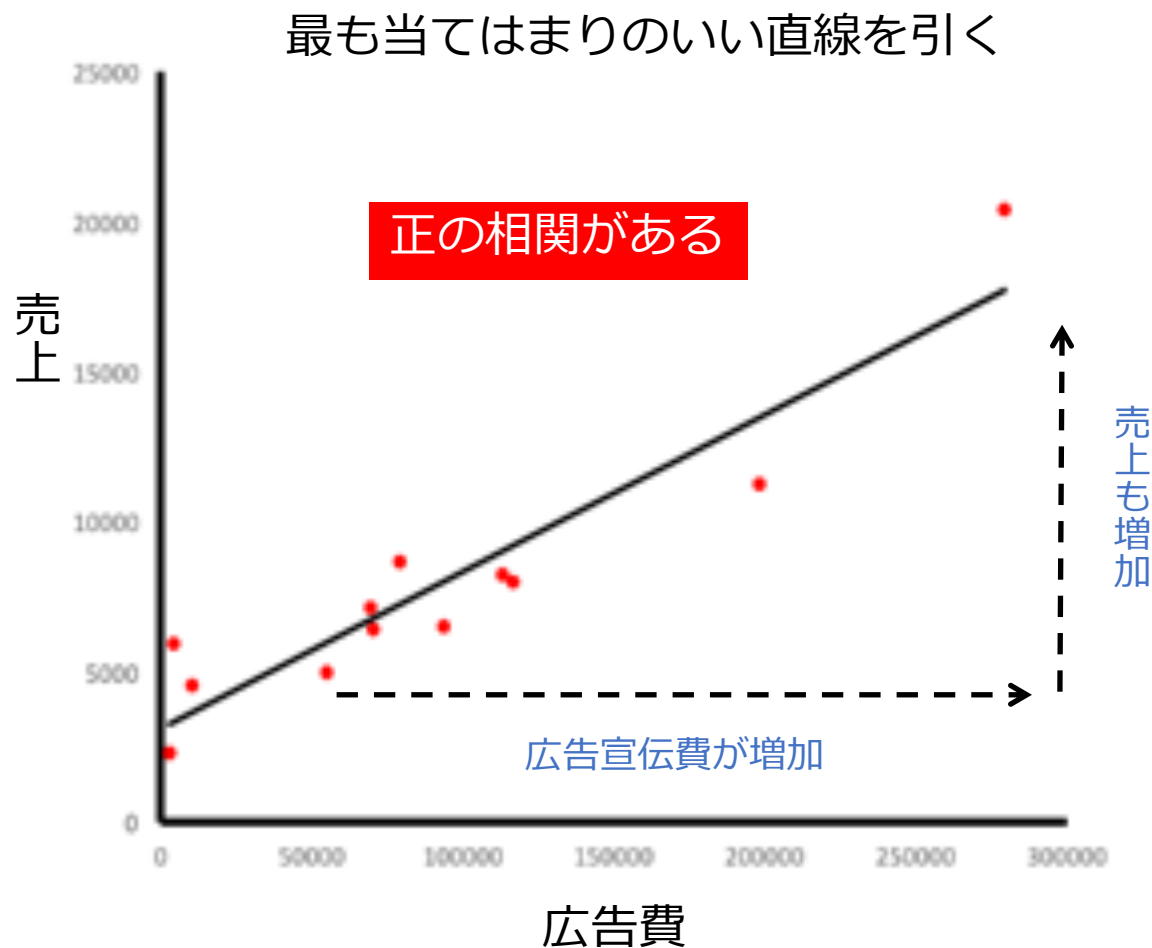
# データ間の関係性を見る

月	広告費	売上
4月	3,004	2,295
5月	3,982	5,928
6月	279,284	20,399
7月	198,374	11,245
8月	10,492	4,567
9月	78,938	8,673
10月	70,293	6,394
11月	69,283	7,124
12月	54,900	4,958
1月	112,938	8,256
2月	116,823	7,982
3月	93,847	6,504

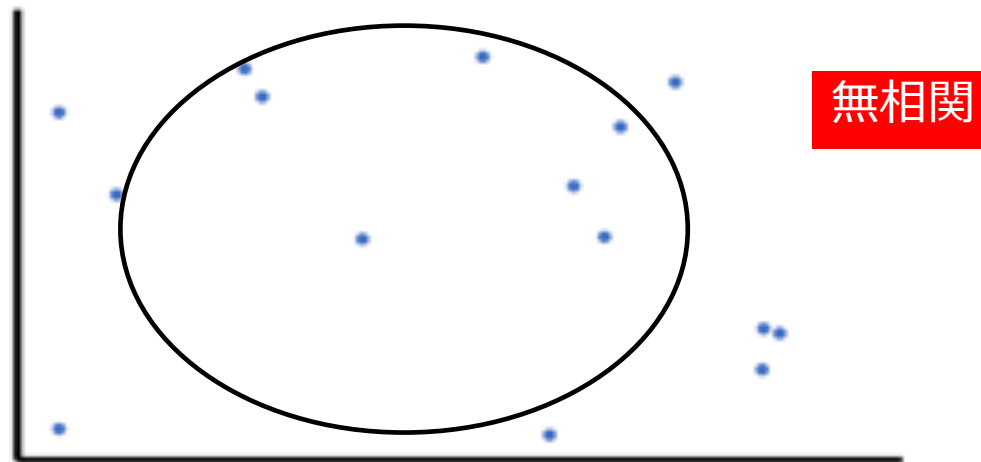
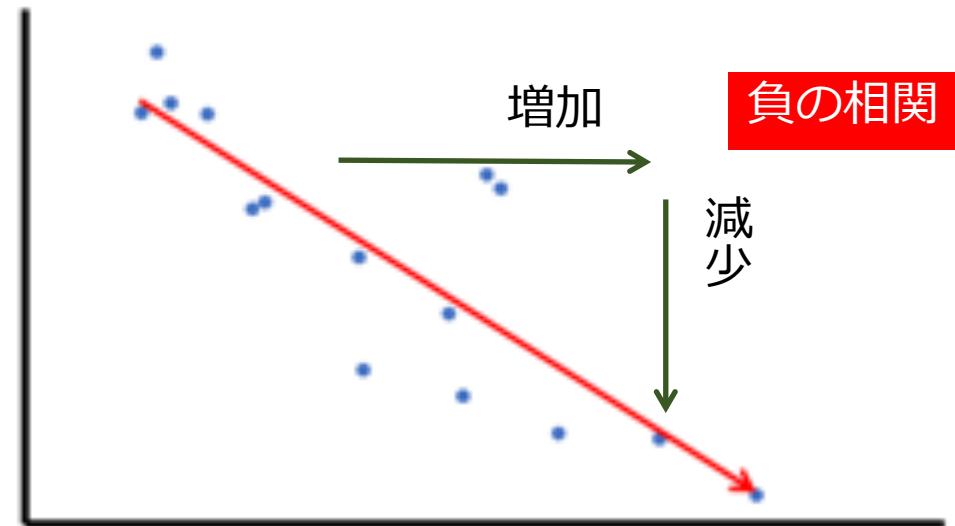
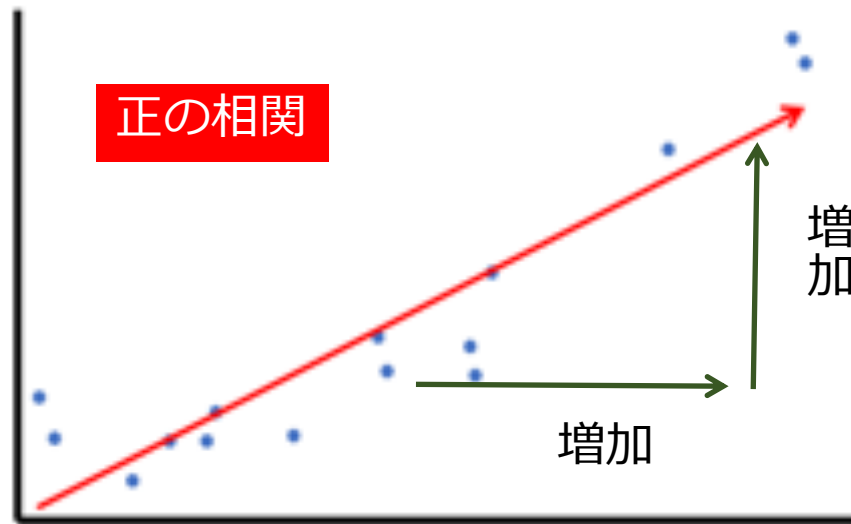


# データ間の関係性を見る

月	広告費	売上
4月	3,004	2,295
5月	3,982	5,928
6月	279,284	20,399
7月	198,374	11,245
8月	10,492	4,567
9月	78,938	8,673
10月	70,293	6,394
11月	69,283	7,124
12月	54,900	4,958
1月	112,938	8,256
2月	116,823	7,982
3月	93,847	6,504



# 関係性の方向

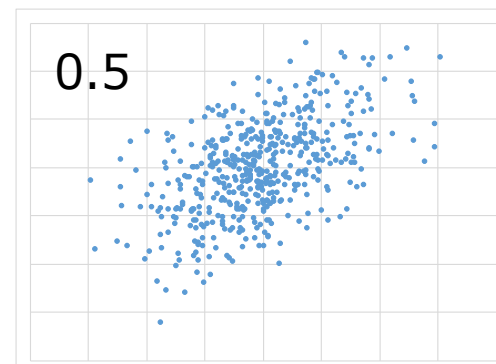
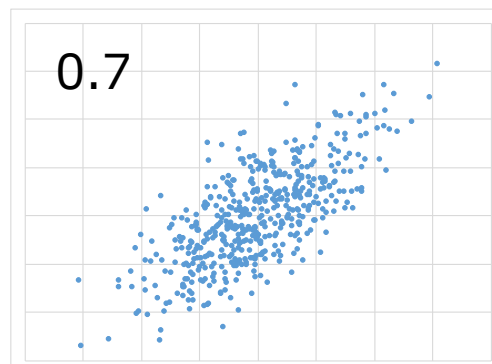
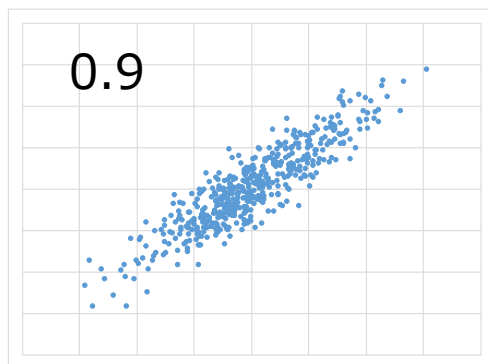




# 相関係数

- 2種類のデータの関係性を-1～1の間に収まるよう調整したもの

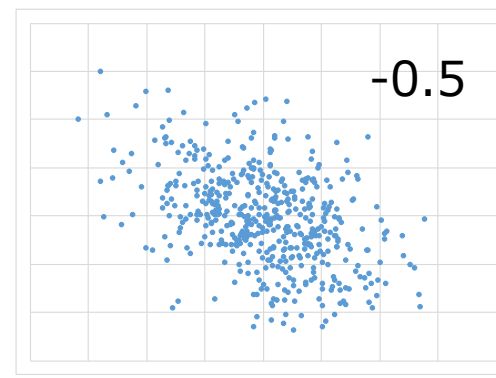
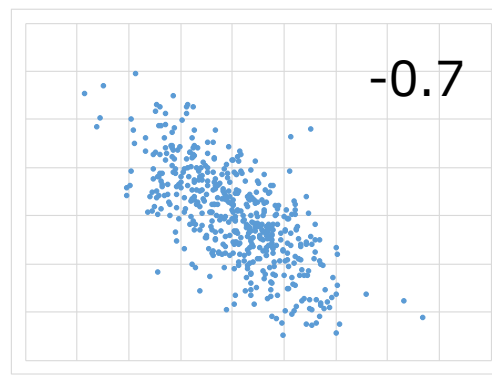
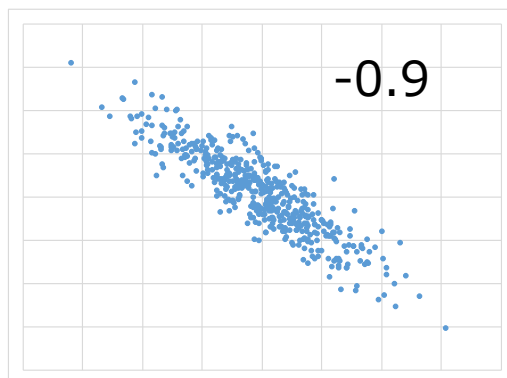
正の相関



強い

弱い

負の相関



EXCEL

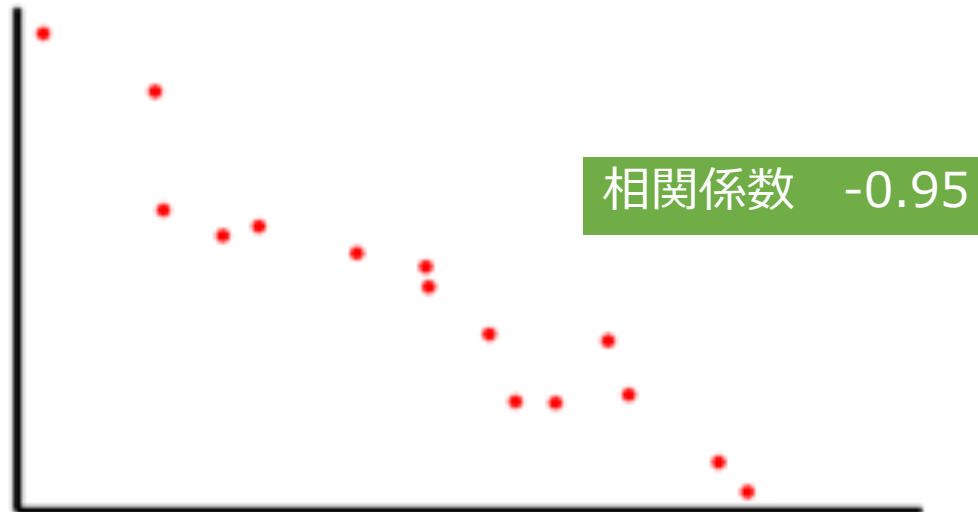
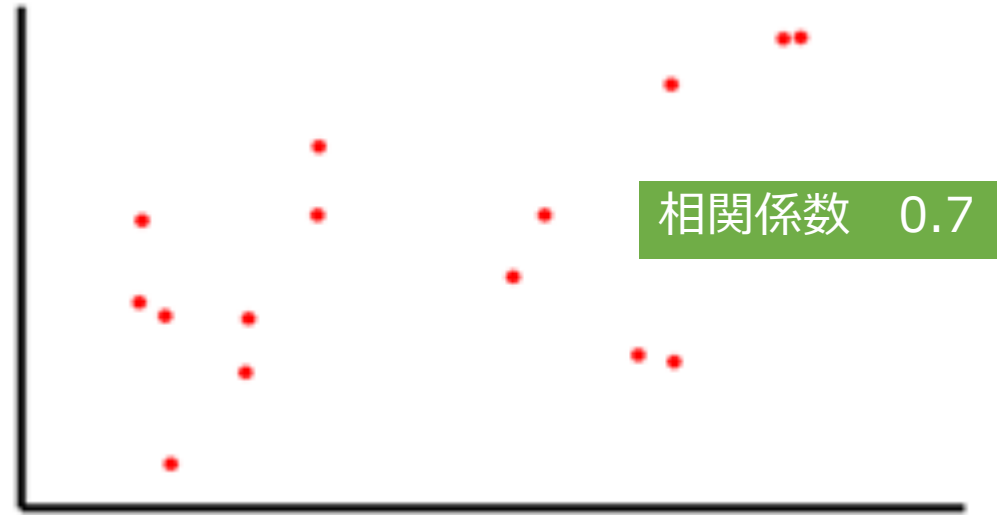
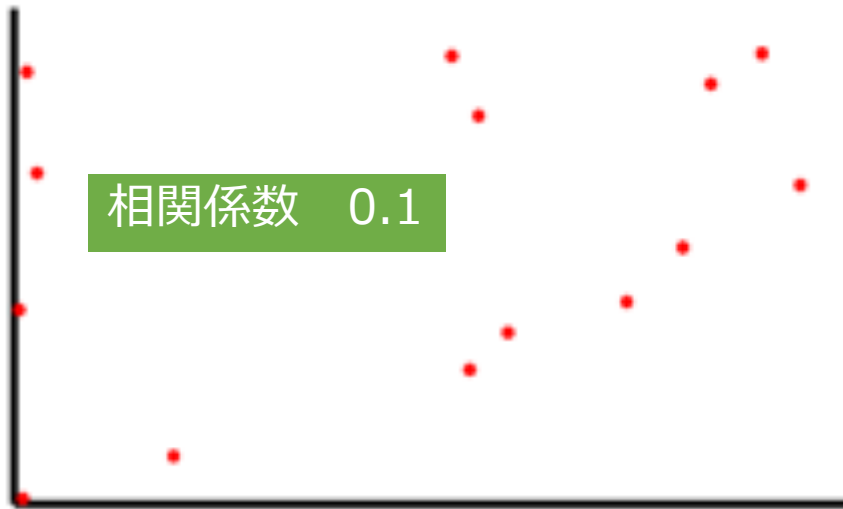
=CORREL(データ1,データ2)

散布図と相関係数の対応

# 相関係数と相関の強弱

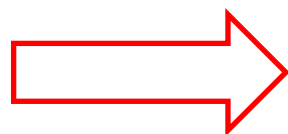
相関係数 $r$	相関の強さ
$-1 \leq r \leq -0.7$	強い負の相関
$-0.7 \leq r \leq -0.4$	負の相関
$-0.4 \leq r \leq -0.2$	弱い負の相関
$-0.2 \leq r \leq 0.2$	ほとんど相関がない
$0.2 \leq r \leq 0.4$	弱い正の相関
$0.4 \leq r \leq 0.7$	正の相関
$0.7 \leq r \leq 1$	強い正の相関

# それぞれの相関係数は？



# 相関係数の注意点

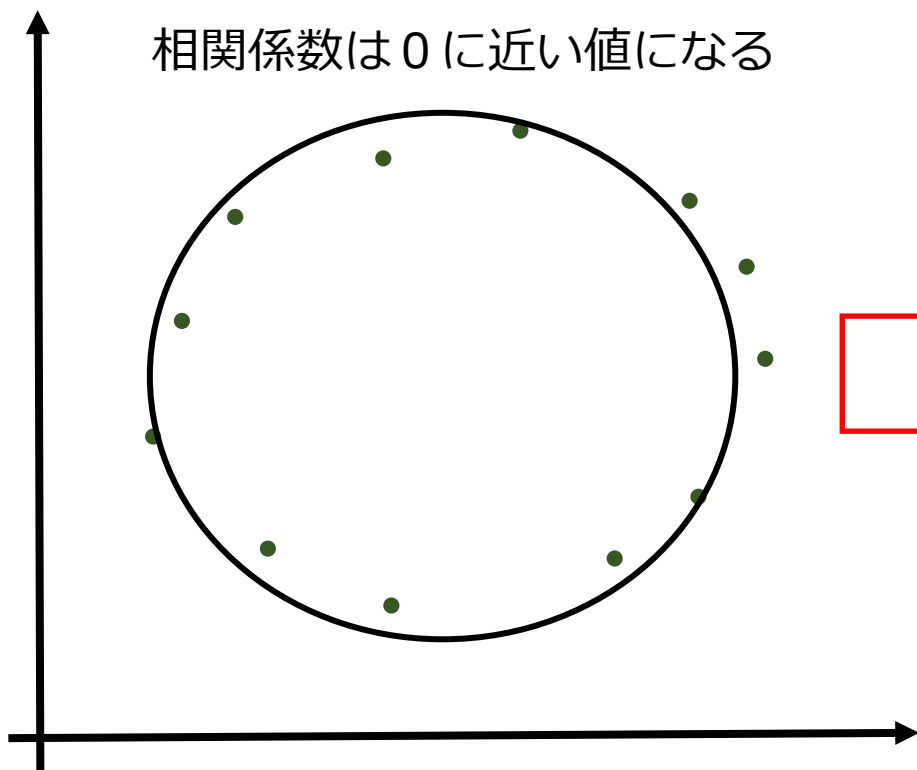
相関係数 = 0



変数間に関係がない



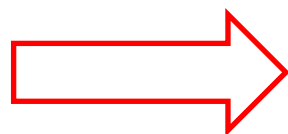
相関係数は0に近い値になる



線形関係はないが変数間に関係はある。

# 相関係数の注意点

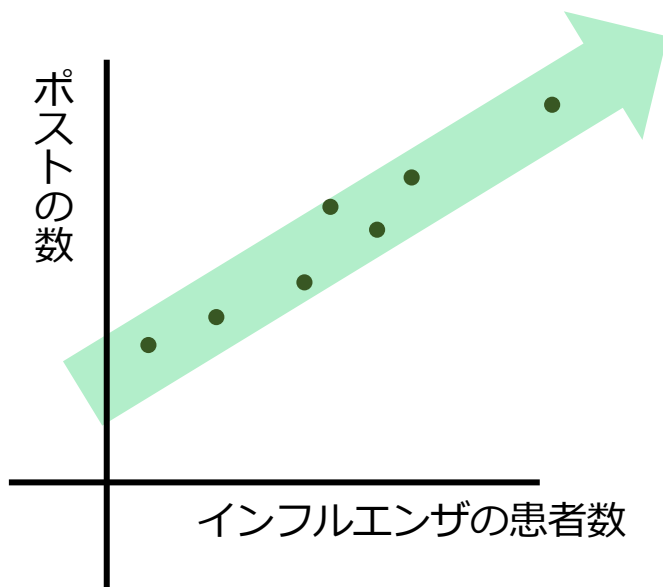
相関がある



因果関係がある



例) インフルエンザの患者数とポストの数に相関があった。ポストの数を減らせばインフルエンザの患者数は減る？



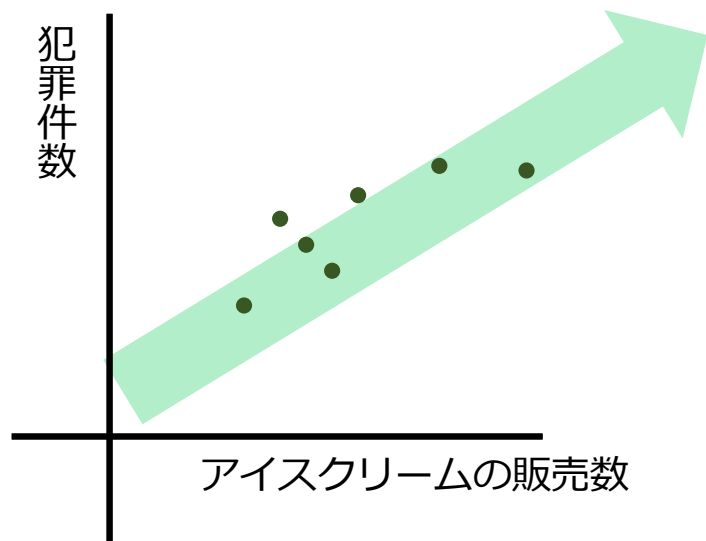
あくまで「相関＝線形関係性」であり、因果関係があるとは限らない

# 相関係数の注意点

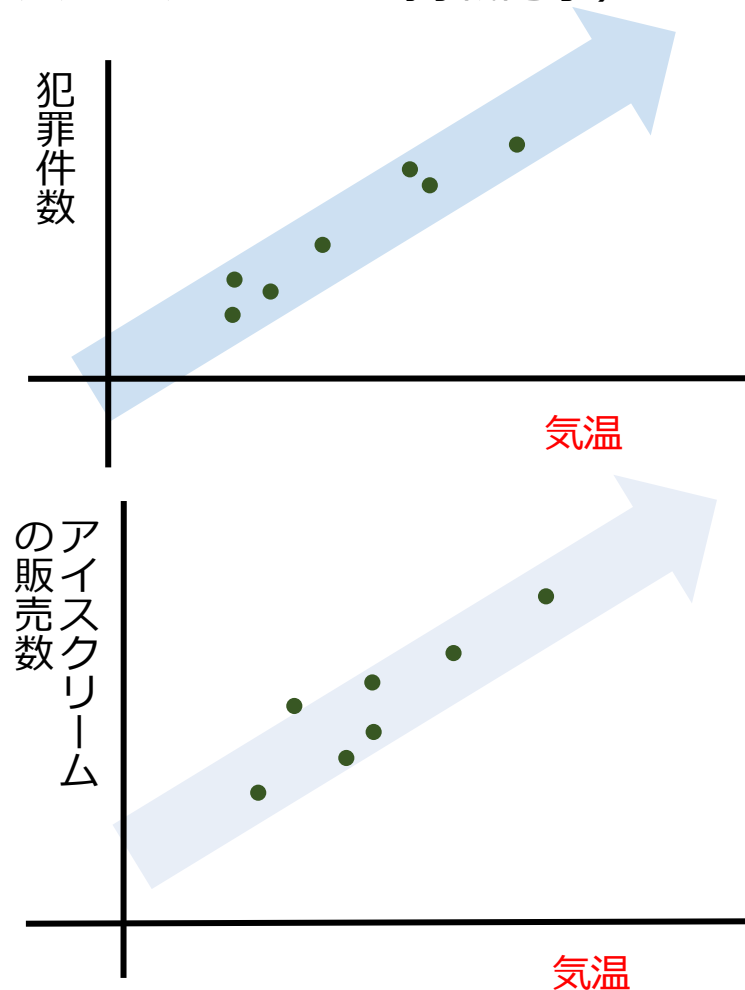
アイスクリームが殺人事件を呼ぶ？

(レイジアナ州ニューオリンズの地元紙タイムズ・ピカユーンに掲載記事)

アイスクリームの販売数と犯罪件数には強い相関がある

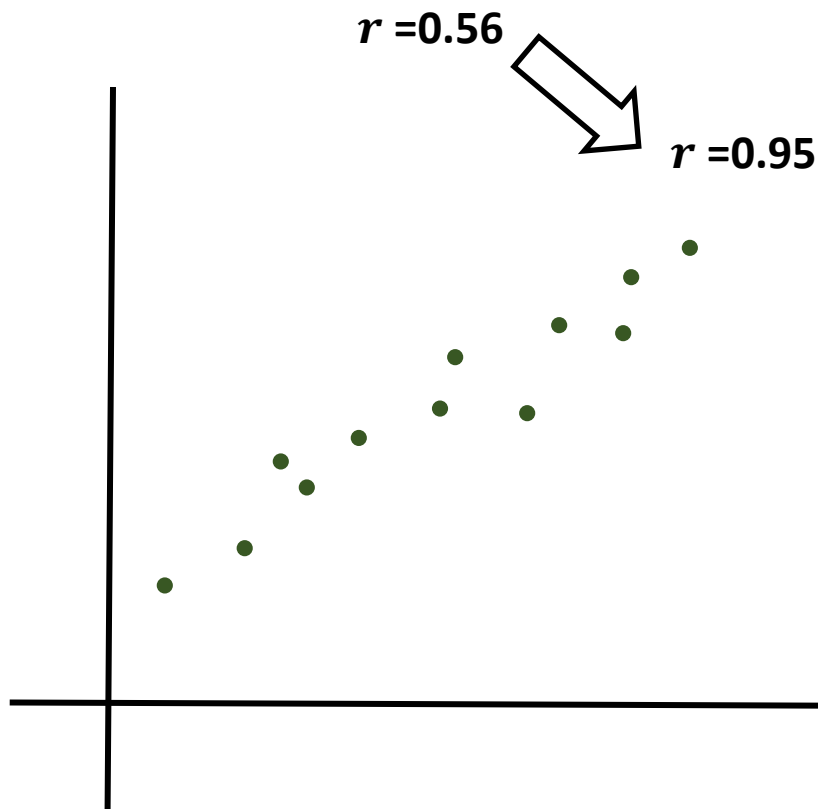


疑似相関で両方の原因となる  
第3の因子が存在するケース

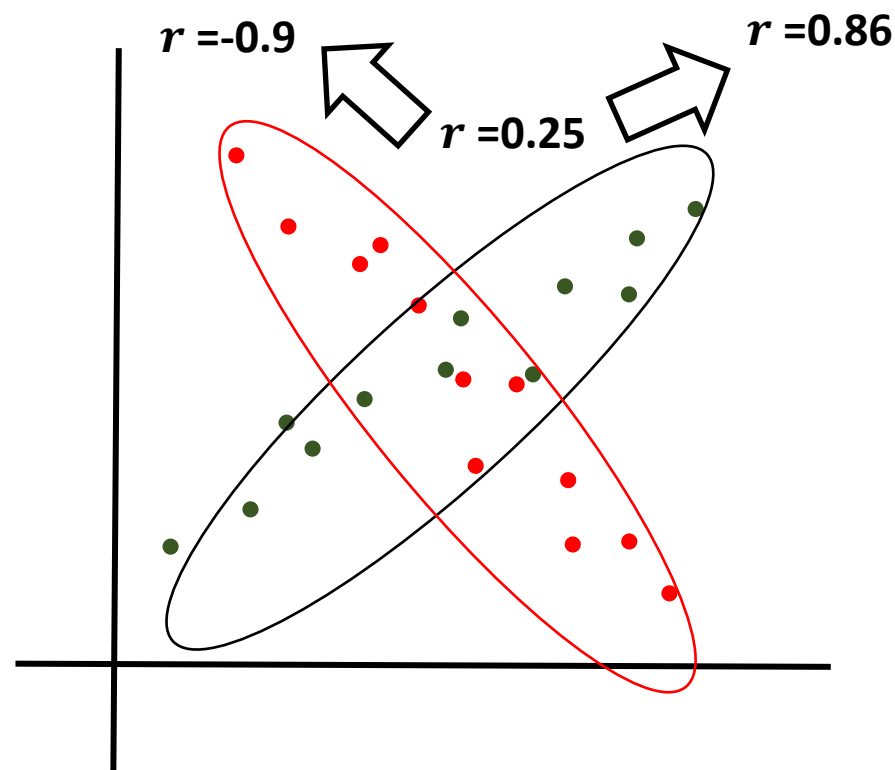


# 相関係数の注意点

相関係数は外れ値の影響を受けやすい



全体として相関は低いが、データを分けて相関を計算すると高い相関を持つ



# (演習1)2変数データ(散布図)

## 演習問題 1

下記のデータを使って次の質問に教えてください。

- (1) 散布図を作成せよ
- (2) 相関係数を求めよ

月	広告宣伝費	売上
4月	3,004	2,295
5月	3,982	5,928
6月	279,284	20,399
7月	198,374	11,245
8月	10,492	4,567
9月	78,938	8,673
10月	70,293	6,394
11月	69,283	7,124
12月	54,900	4,958
1月	112,938	8,256
2月	116,823	7,982
3月	93,847	6,504



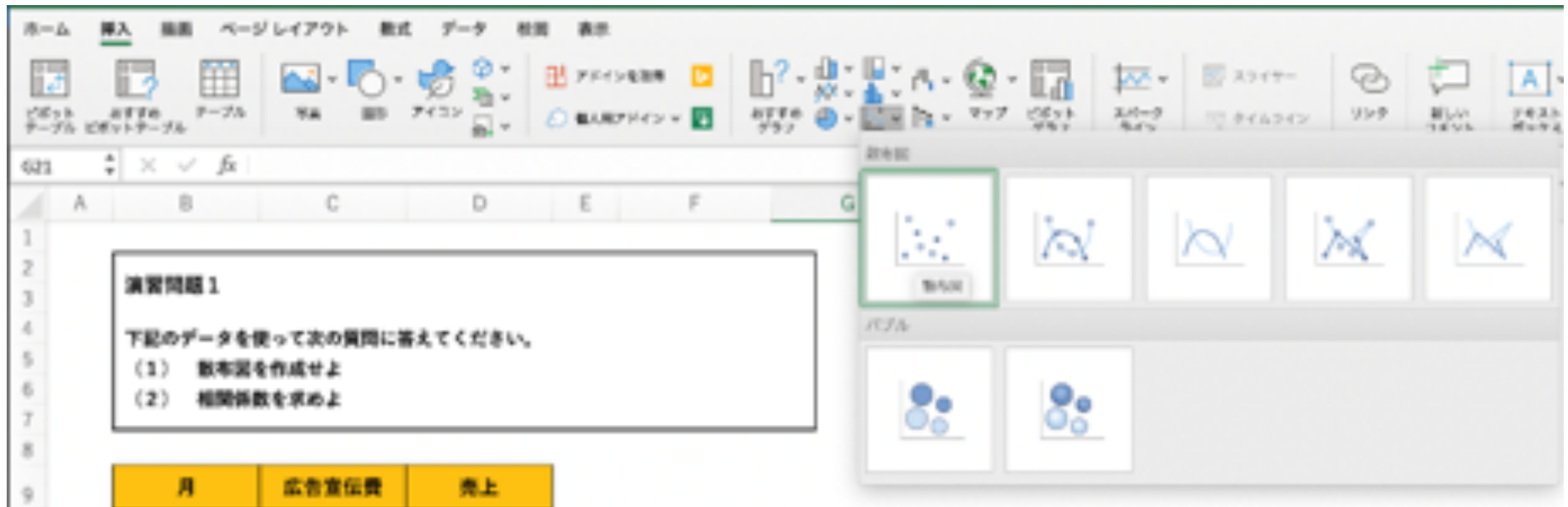
# (演習1)2変数データ(散布図)

## 1. データを選択する。

月	広告宣伝費	売上
4月	3,004	2,295
5月	3,982	5,928
6月	279,284	20,399
7月	198,374	11,245
8月	10,492	4,567
9月	78,938	8,673
10月	70,293	6,394
11月	69,283	7,124
12月	54,900	4,958
1月	112,938	8,256
2月	116,823	7,982
3月	93,847	6,504

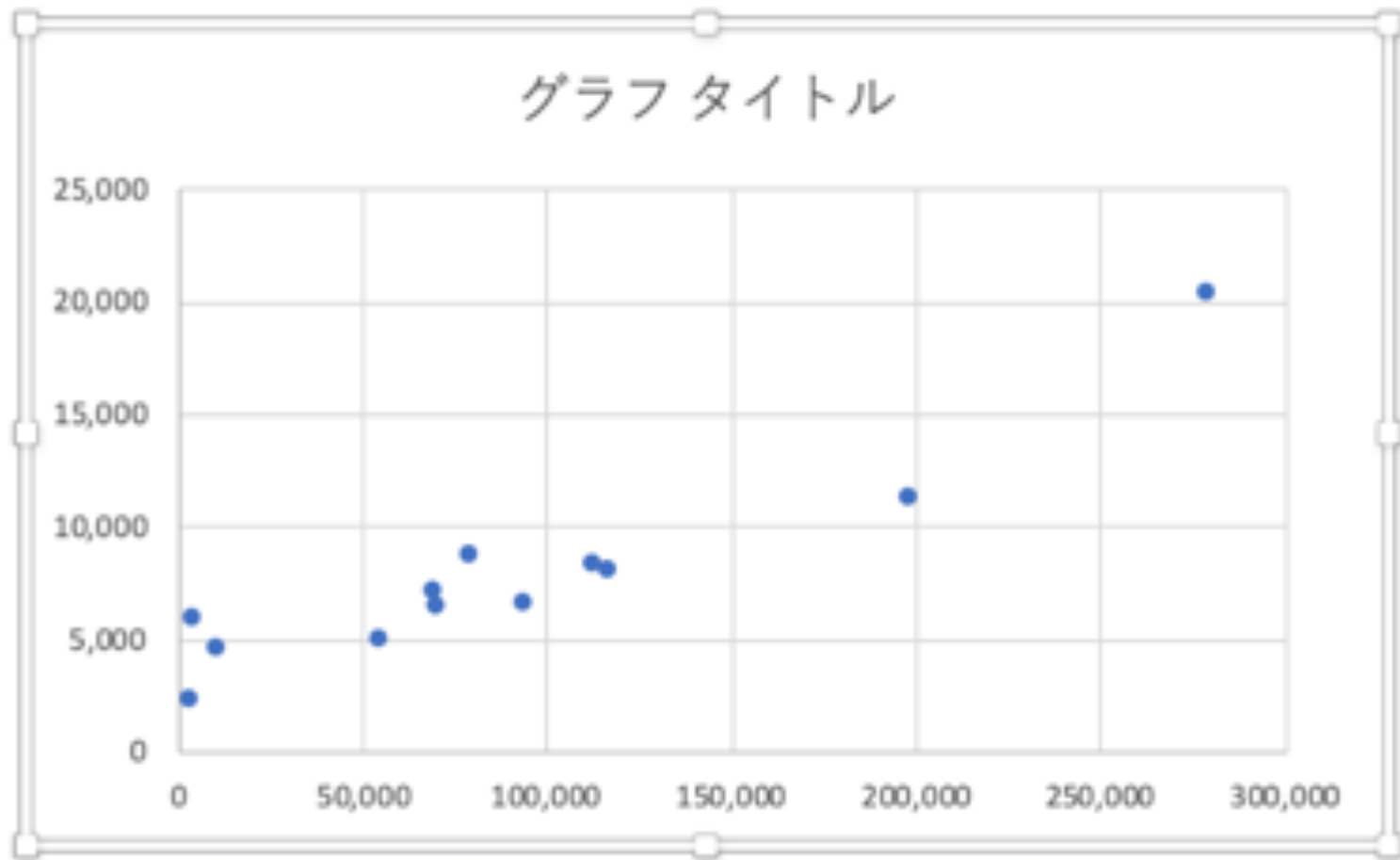
# (演習1)2変数データ(散布図)

2. 「挿入」→「グラフ」→「散布図」を選択する。



# (演習1)2変数データ(散布図)

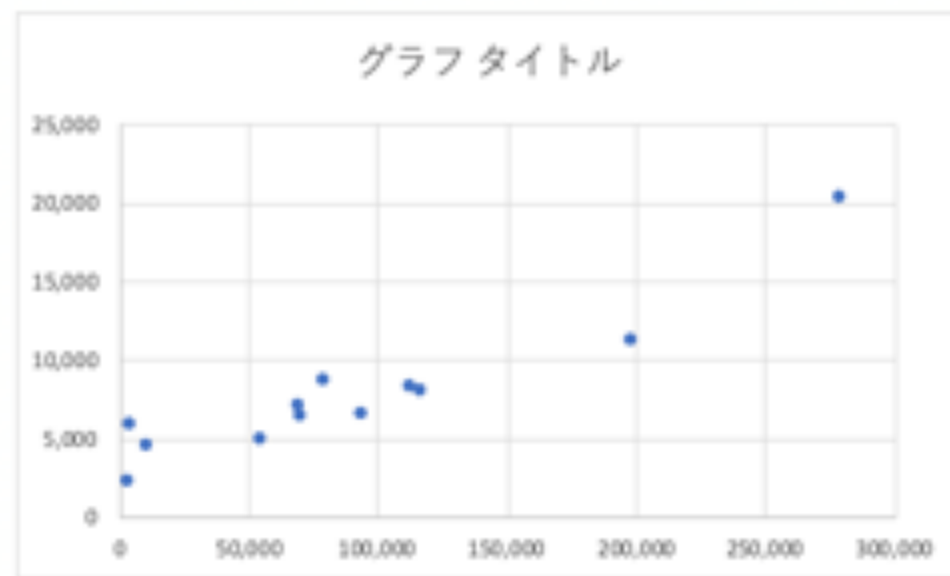
散布図が表示される。



# (演習1)2変数データ(散布図)

3. 関数CORRELを用いて、相関係数を計算する。

月	広告宣伝費	売上
4月	3,004	2,295
5月	3,982	5,928
6月	279,284	20,399
7月	198,374	11,245
8月	10,492	4,567
9月	78,938	8,673
10月	70,293	6,394
11月	69,283	7,124
12月	54,900	4,958
1月	112,938	8,256
2月	116,823	7,982
3月	93,847	6,504

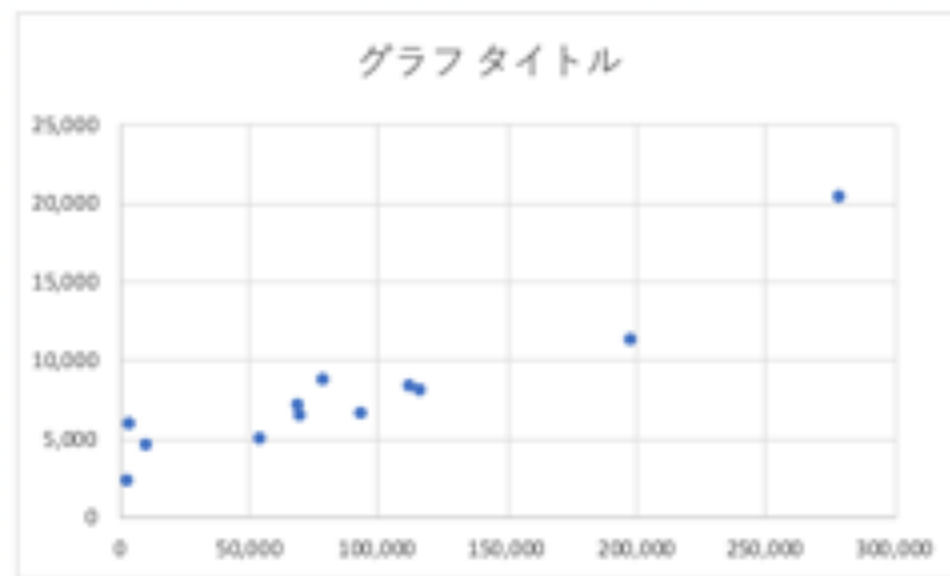


=CORREL(C10:C21, D10:D21)

# (演習1)2変数データ(散布図)

相関係数が表示される。

月	広告宣伝費	売上
4月	3,004	2,295
5月	3,982	5,928
6月	279,284	20,399
7月	198,374	11,245
8月	10,492	4,567
9月	78,938	8,673
10月	70,293	6,394
11月	69,283	7,124
12月	54,900	4,958
1月	112,938	8,256
2月	116,823	7,982
3月	93,847	6,504



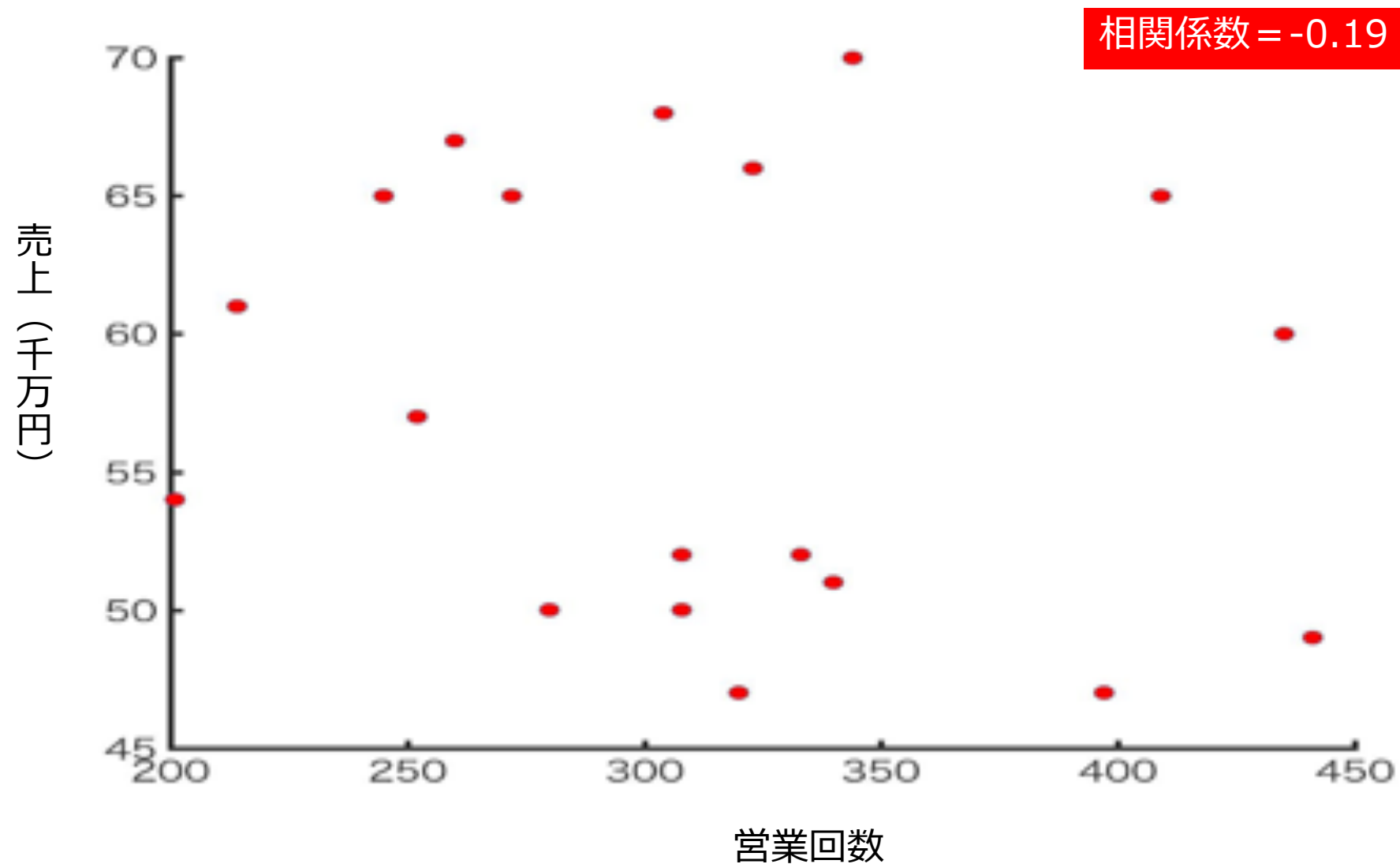
0.935174784

# 問題 1

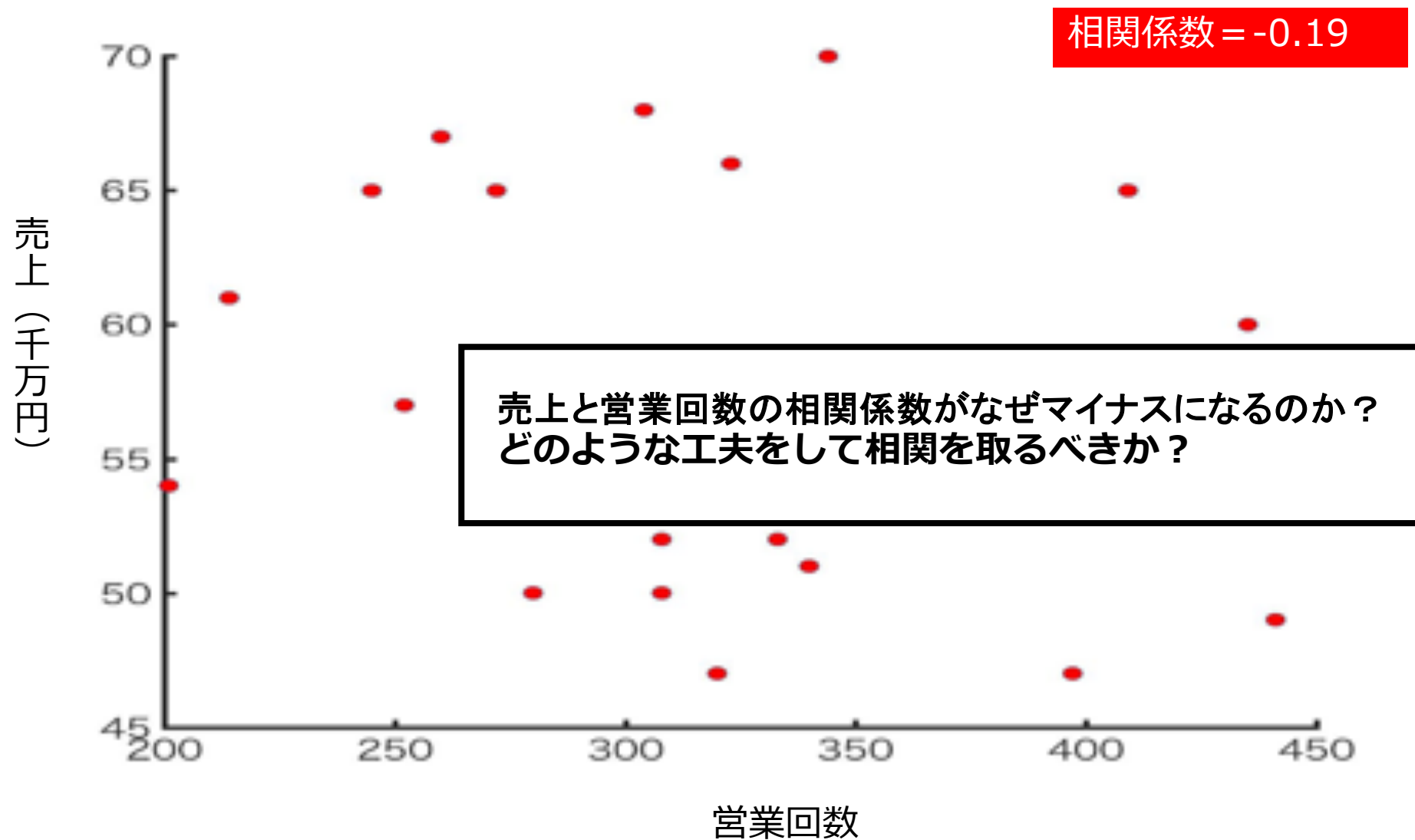
日付	売上額	営業回数
2016年1月	47	320
2016年2月	67	260
2016年3月	61	214
2016年4月	52	333
2016年5月	65	272
2016年6月	51	340
2016年7月	70	344
2016年8月	52	308
2016年9月	50	280
2016年10月	57	252
2016年11月	54	201
2016年12月	49	441
2017年1月	60	435
2017年2月	65	245
2017年3月	47	397
2017年4月	66	323
2017年5月	65	409
2017年6月	68	304
2017年7月	50	308

左はある会社の売上額と営業回数のデータです。  
2つの変数間に何らかの関係はあると言えるでしょうか？

# 問題 1 (散布図を作成)



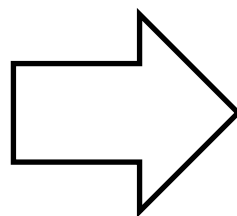
# 問題 1





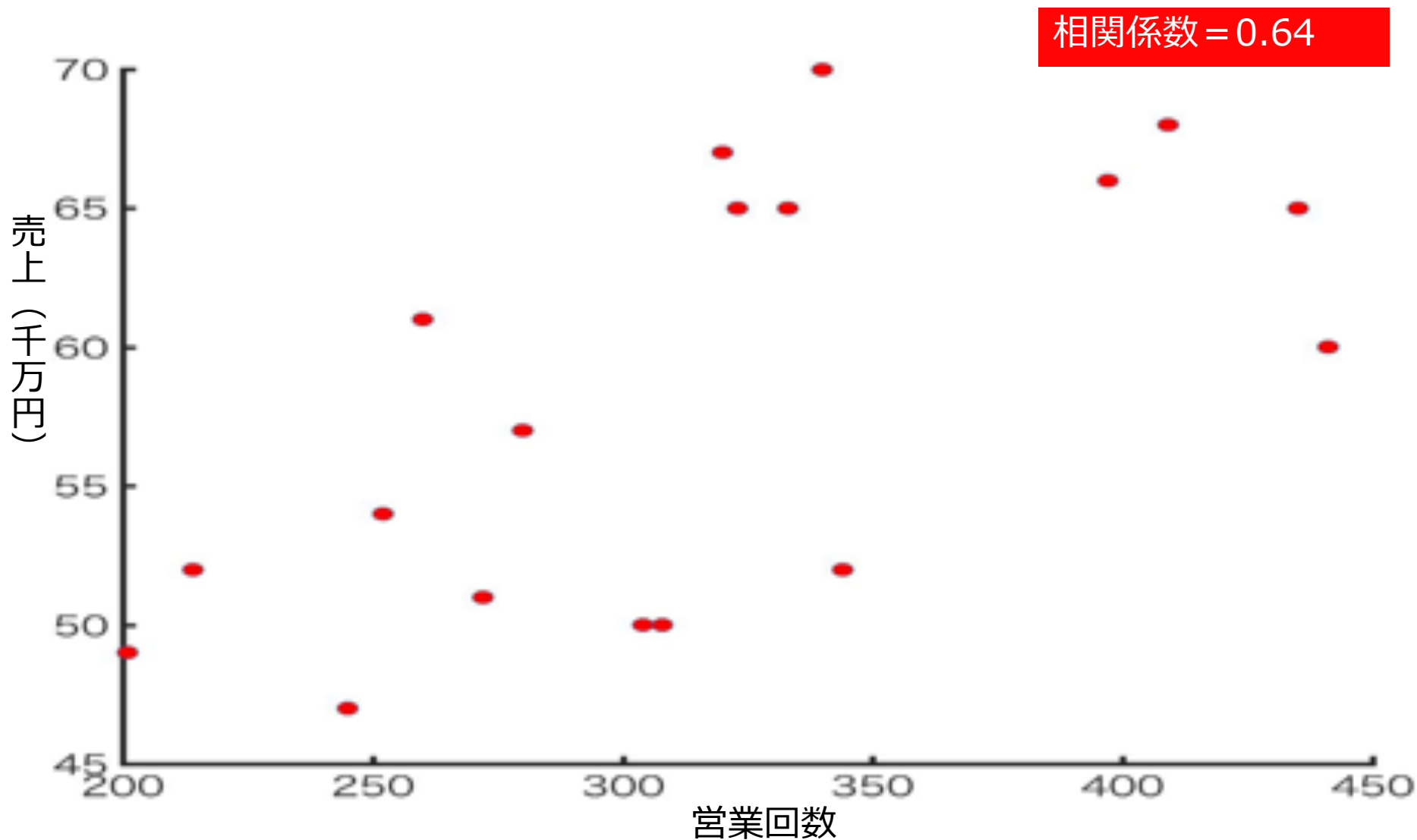
# 問題 1 (タイムラグを考慮する)

日付	売上額	営業回数
2016年1月	47	320
2016年2月	67	260
2016年3月	61	214
2016年4月	52	333
2016年5月	65	272
2016年6月	51	340
2016年7月	70	344
2016年8月	52	308
2016年9月	50	280
2016年10月	57	252
2016年11月	54	201
2016年12月	49	441
2017年1月	60	435
2017年2月	65	245
2017年3月	47	397
2017年4月	66	323
2017年5月	65	409
2017年6月	68	304
2017年7月	50	308



日付	売上額	営業回数
2016年1月	47	
2016年2月	67	320
2016年3月	61	260
2016年4月	52	214
2016年5月	65	333
2016年6月	51	272
2016年7月	70	340
2016年8月	52	344
2016年9月	50	308
2016年10月	57	280
2016年11月	54	252
2016年12月	49	201
2017年1月	60	441
2017年2月	65	435
2017年3月	47	245
2017年4月	66	397
2017年5月	65	323
2017年6月	68	409
2017年7月	50	304
		308

# 問題 1 (タイムラグを考慮した相関係数)



# 問題 2

下はスーパーの各店舗ごとの売上額とそのマーケティングデータです。データ間の関係进行分析せよ。

店舗	売上	店舗面積	アクセス数	町の規模	同業店舗数
1	231	3	294	8.2	11
2	156	2.2	232	4.1	12
3	10	0.5	149	4.3	15
4	519	5.5	260	16.1	1
5	437	4.4	567	14.1	5
6	487	4.8	237	12.7	4
7	299	3.1	451	10.1	10
8	195	2.5	433	8.4	12
9	20	1.2	212	2.1	15
10	68	0.6	502	4.7	8
11	570	5.4	378	12.3	1
12	428	4.2	577	14	7

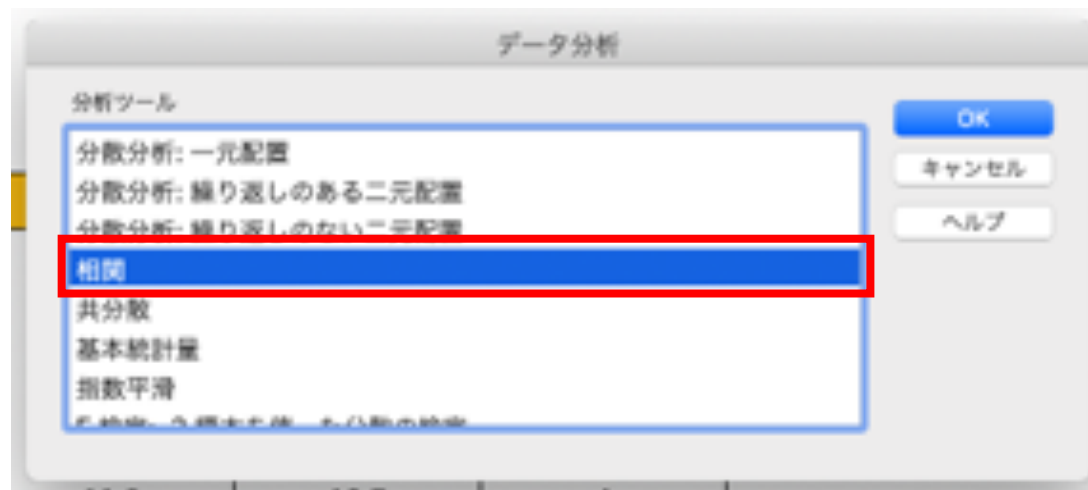
# (演習3)多変数の相関

分析ツールを開く。

1. 「データ」リボンの「データ分析」を選択する。



2. 「相関」を選択する。



## (演習3)多変数の相関

### 3. データを選択する（店舗 # は除く）。

店舗#	売上	店舗面積	アクセス数	広告	店の規模	同業者店舗数
1	231	3	294	8.2	8.2	11
2	156	2.2	232	6.9	4.1	12
3	10	6.5	149	3	4.3	15
4	519	5.5	260	12	16.1	1
5	437	4.4	567	10.6	14.1	5
6	487	4.8	237	11.8	12.7	4
7	299	3.1	451	8.1	10.1	10
8	196	2.5	433	7.7	8.4	12
9	20	1.2	212	3.3	2.1	15
10	68	6.6	502	4.9	4.7	8
11	530	5.4	378	17.4	12.3	1
12	428	4.2	577	10.5	14	7
13	464	4.7	653	11.3	15	3
14	15	6.6	126	2.5	2.5	14
15	65	1.2	168	4.7	3.3	11
16	98	1.6	651	4.6	2.7	10



## (演習3)多変数の相関

4. ラベルにチェックを入れる。



## (演習3)多変数の相関

5. 出力先を選択する。

街の規模	同業者店舗数
8.2	11
4.1	12
4.3	15
16.1	1
14.1	5
12.7	4
10.1	10
8.4	12
2.1	15
4.7	8
12.3	1
14	7
15	3
2.5	14
3.3	11
2.7	10



# (演習3)多変数の相関

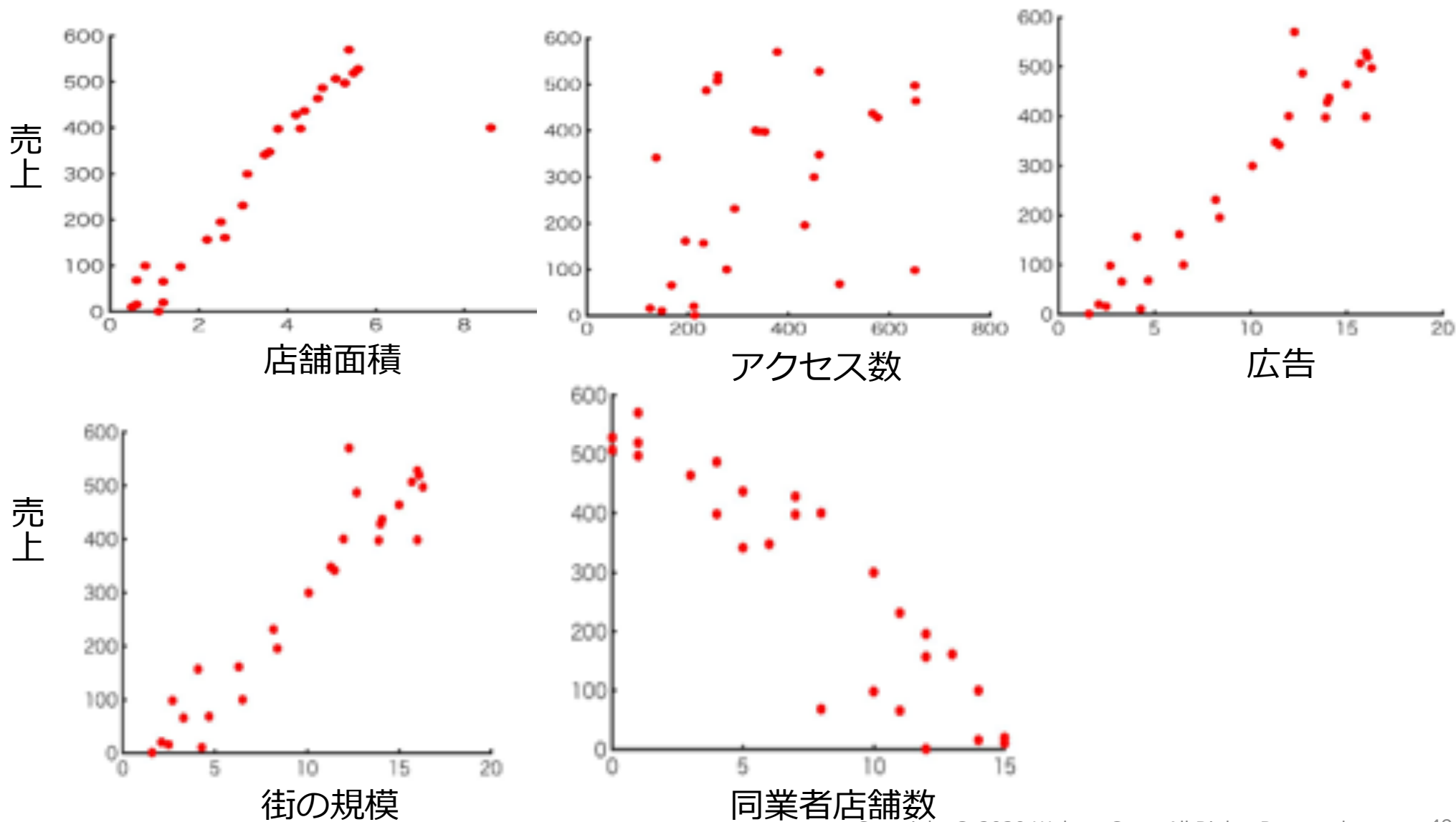
	売上	店舗面積	アクセス数	広告	街の規模	同業者店舗数
売上	1					
店舗面積	0.894092082	1				
アクセス数	0.401628632	0.312766817	1			
広告	0.914024068	0.748587237	0.359082857	1		
街の規模	0.953683059	0.838022883	0.41134901	0.795434449	1	
同業者店舗数	-0.912236392	-0.765737788	-0.416460844	-0.841279944	-0.869589611	1

※ 「ホーム」→「条件付書式」→「カラースケール」から色で識別することも可。

	売上	店舗面積	アクセス数	広告	街の規模	同業者店舗数
売上	1					
店舗面積	0.894092082	1				
アクセス数	0.401628632	0.312766817	1			
広告	0.914024068	0.748587237	0.359082857	1		
街の規模	0.953683059	0.838022883	0.41134901	0.795434449	1	
同業者店舗数	-0.912236392	-0.765737788	-0.416460844	-0.841279944	-0.869589611	1



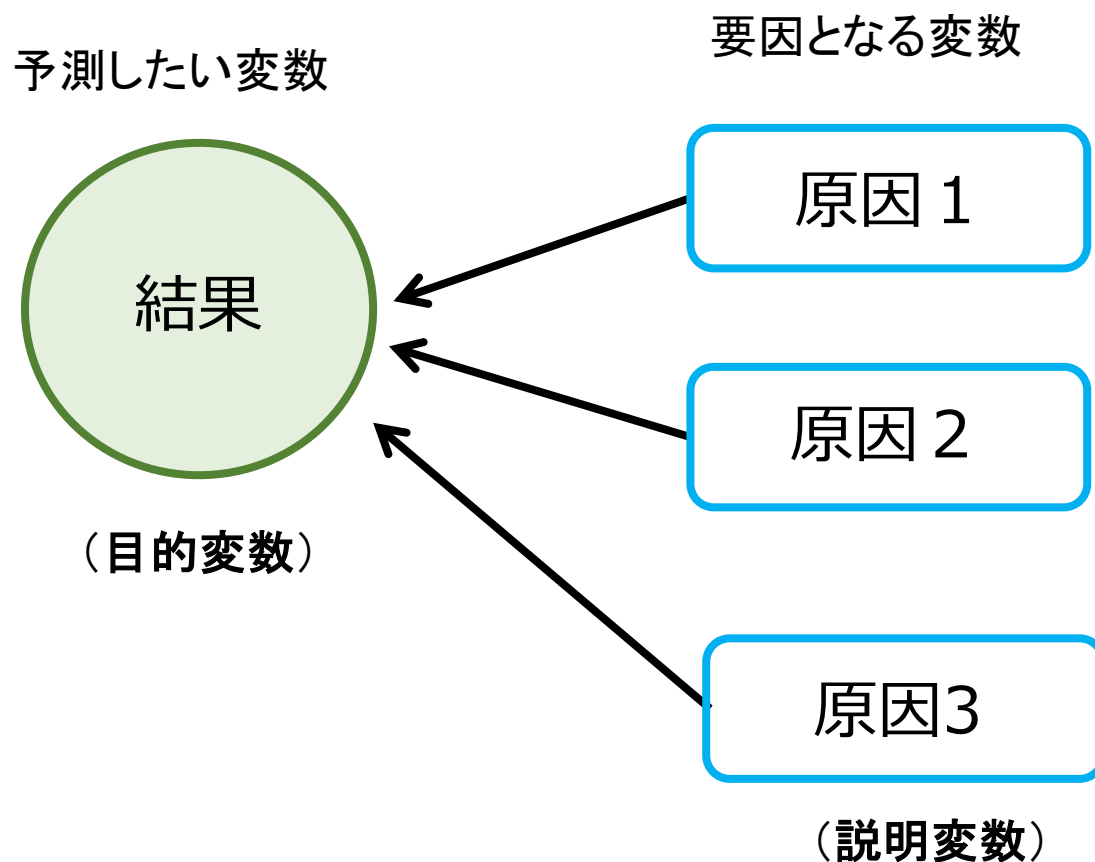
# (演習3)多変数の相関 (散布図による可視化)



# 回帰モデル

# 回帰分析モデル

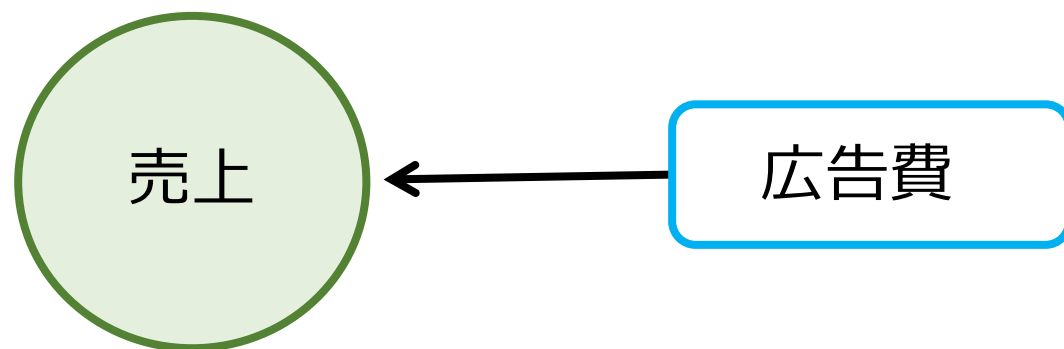
予測したい変数を要因と考えられる変数を使って予測・説明する方法



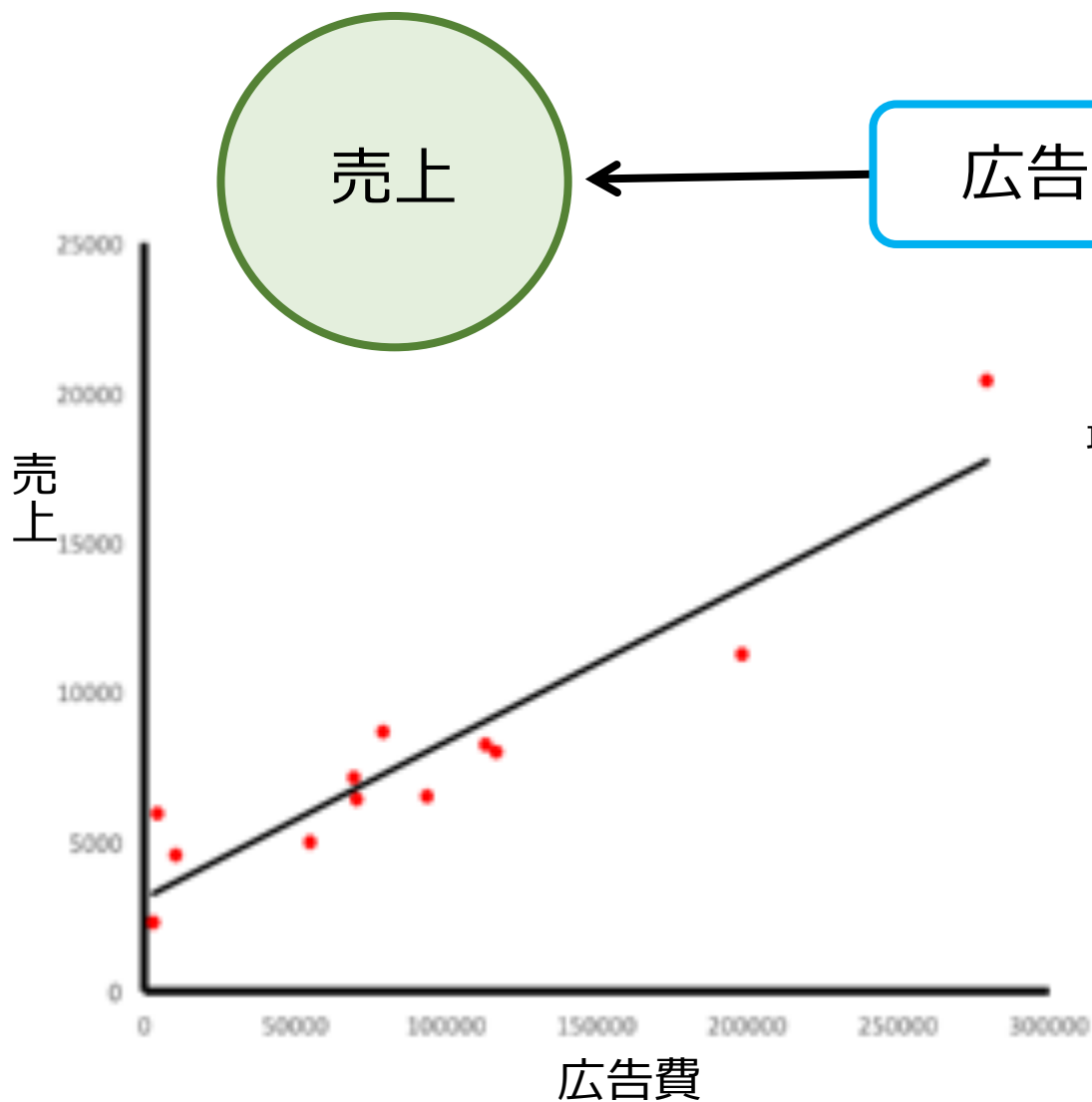
# 単回帰分析モデル

予測したい変数を1つの要因と考えられる変数を使って予測・説明する方法

月	広告費	売上
4月	3,004	2,295
5月	3,982	5,928
6月	279,284	20,399
7月	198,374	11,245
8月	10,492	4,567
9月	78,938	8,673
10月	70,293	6,394
11月	69,283	7,124
12月	54,900	4,958
1月	112,938	8,256
2月	116,823	7,982
3月	93,847	6,504



# 単回帰分析モデル



回帰モデルをつくるとは？

$$y = b_0 + b_1x$$

売上 =  $b_0$  +  $b_1$ (広告費)

回帰係数

# 単回帰分析（演習1のデータ）

演習1のデータを使って、単回帰分析を行う。

## 演習問題1

下記のデータを使って次の質問に答えてください。

- (1) 散布図を作成せよ
- (2) 相関係数を求めよ

月	広告宣伝費	売上
4月	3,004	2,295
5月	3,982	5,928
6月	279,284	20,399
7月	198,374	11,245
8月	10,492	4,567
9月	78,938	8,673
10月	70,293	6,394
11月	69,283	7,124
12月	54,900	4,958
1月	112,938	8,256
2月	116,823	7,982
3月	93,847	6,504



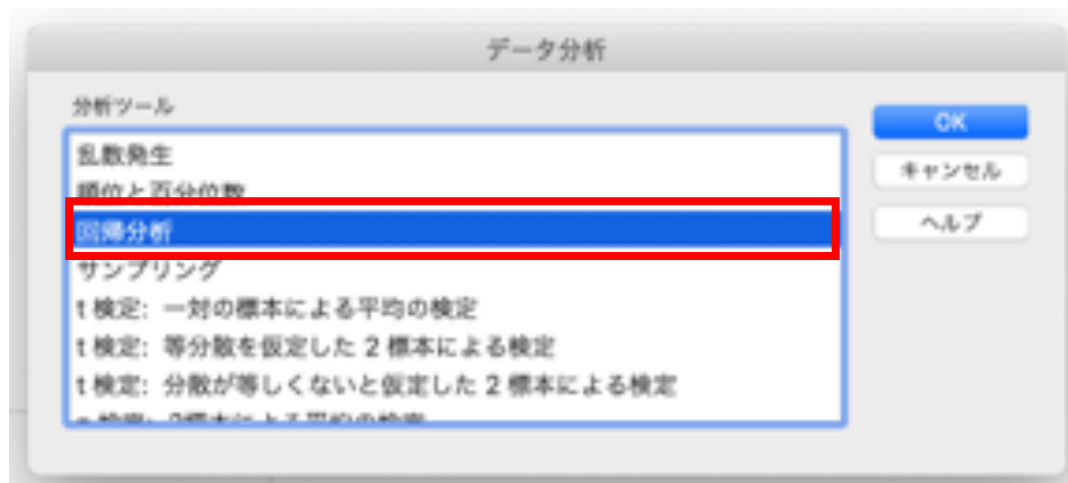
# 単回帰分析（演習1のデータ）

分析ツールを開く。

1. 「データ」リボンの「データ分析」を選択する。



2. 「回帰分析」を選択する。



# 単回帰分析（演習1のデータ）

## 3. 入力Y範囲（目的変数）を選択する。

月	広告宣伝費	売上
4月	3,004	2,295
5月	3,982	5,928
6月	279,284	20,399
7月	198,374	11,245
8月	10,492	4,567
9月	78,938	8,673
10月	70,293	6,394
11月	69,283	7,124
12月	54,900	4,958
1月	112,938	8,256
2月	116,823	7,982
3月	93,847	6,504

図解分析

入力元

入力 Y 範囲:

入力 X 範囲:

☒ ラベル ☐ 数値に 0 を使用

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 ☐ 残差値グラフの作成

正統確率

☐ 正統確率グラフの作成

OK  
キャンセル  
ヘルプ



# 単回帰分析（演習1のデータ）

## 4. 入力X範囲（説明変数）を選択する。

月	広告宣伝費	売上
4月	3,004	2,295
5月	3,982	5,928
6月	279,284	20,399
7月	198,374	11,245
8月	10,492	4,567
9月	78,938	8,673
10月	70,293	6,394
11月	69,283	7,124
12月	54,900	4,958
1月	112,938	8,256
2月	116,823	7,982
3月	93,847	6,504

図表分析

入力元

入力 Y 範囲: \$D\$1:\$D\$21

入力 X 範囲: **\$C\$1:\$C\$21**

☒ ラベル ☐ 定数に 0 を使用

☐ 配置水準 0% %

出力オプション

☒ 一覧の出力先: \$J\$7

☐ 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 ☐ 標準化された残差グラフの作成

正統確率

☐ 正統確率グラフの作成

OK  
キャンセル  
ヘルプ

# 単回帰分析（演習1のデータ）

5. ラベルにチェックを入れる。

The screenshot shows the 'Regression Analysis' dialog box. In the 'Input Range' section, the 'Y Range' is '\$D\$9:\$D\$21' and the 'X Range' is '\$C\$9:\$C\$21'. The 'Labels' checkbox is checked and highlighted with a red box. Below it, the 'Confidence Level' is set to 95%. In the 'Output Options' section, the 'Output Range' is '\$J\$7'.

6. 出力先を指定する。

The screenshot shows the 'Regression Analysis' dialog box. In the 'Input Range' section, the 'Y Range' is '\$D\$9:\$D\$21' and the 'X Range' is '\$C\$9:\$C\$21'. The 'Labels' checkbox is checked. In the 'Output Options' section, the 'Output Range' is '\$J\$7' and is highlighted with a red box. Below it, the 'Confidence Level' is set to 95%.

# 単回帰分析（結果の読み方）

## 概要

回帰統計	
重相関 R	0.9351748
重決定 R2	0.8745519
補正 R2	0.8620071
標準誤差	1691.2707
観測数	12

R2（モデルの当てはまりの良さ）  
0.5以上：当てはまりが良い  
0.2未満：モデルを再考する

## 分散分析表

	自由度	変動	分散	割かれた分散	有意 F
回帰	1	199410336.5	199410337	69.714226	8.081E-06
残差	10	28603966.39	2860396.64		
合計	11	228014302.9			

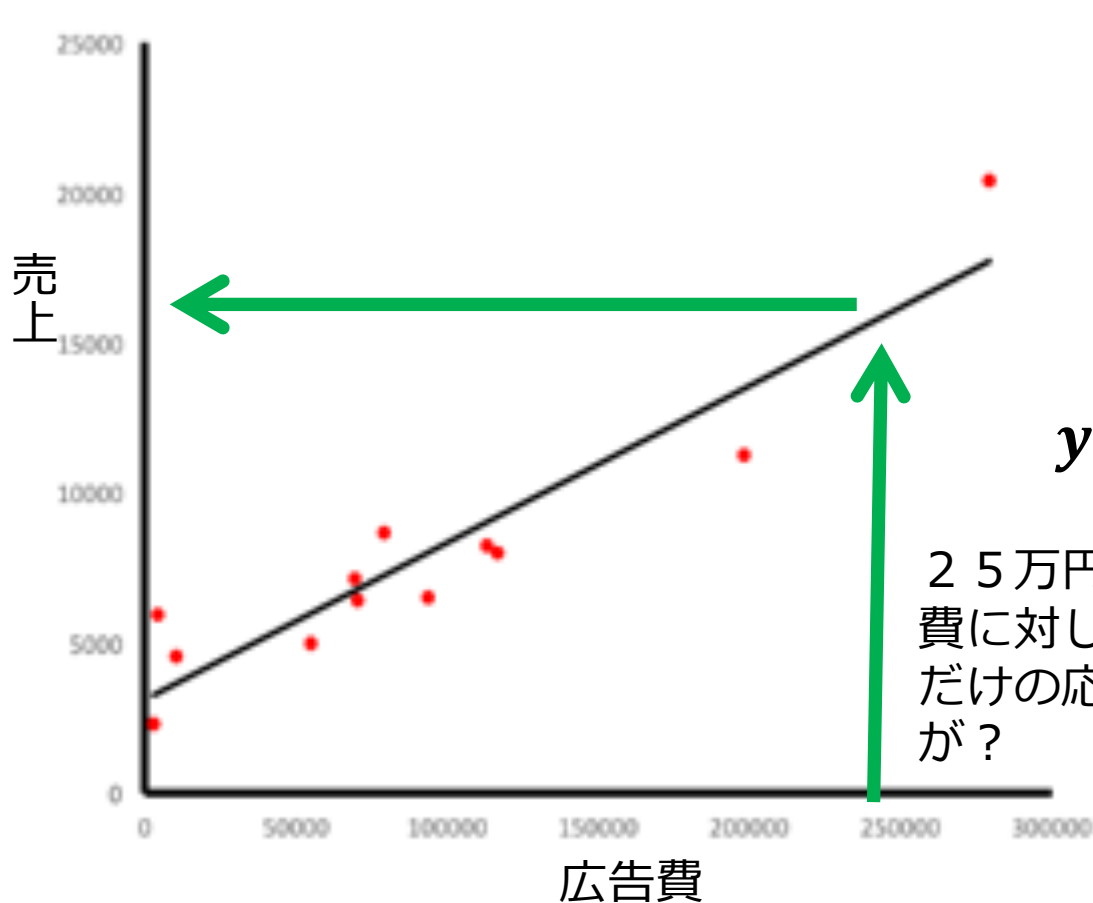
帰無仮説：回帰係数は0である  
対立仮説：回帰係数は0でない

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	3080.0412	752.4371067	4.09342016	0.0021673	1403.5069	4756.5756
広告費	0.052524	0.006290673	8.34950451	8.081E-06	0.0385075	0.0665405

回帰係数  
 $Y = 3080.0412 + 0.052524(\text{広告費})$

P値が優位水準 $\alpha$ 以下の場合、帰無仮説を棄却する

# 単回帰分析モデル



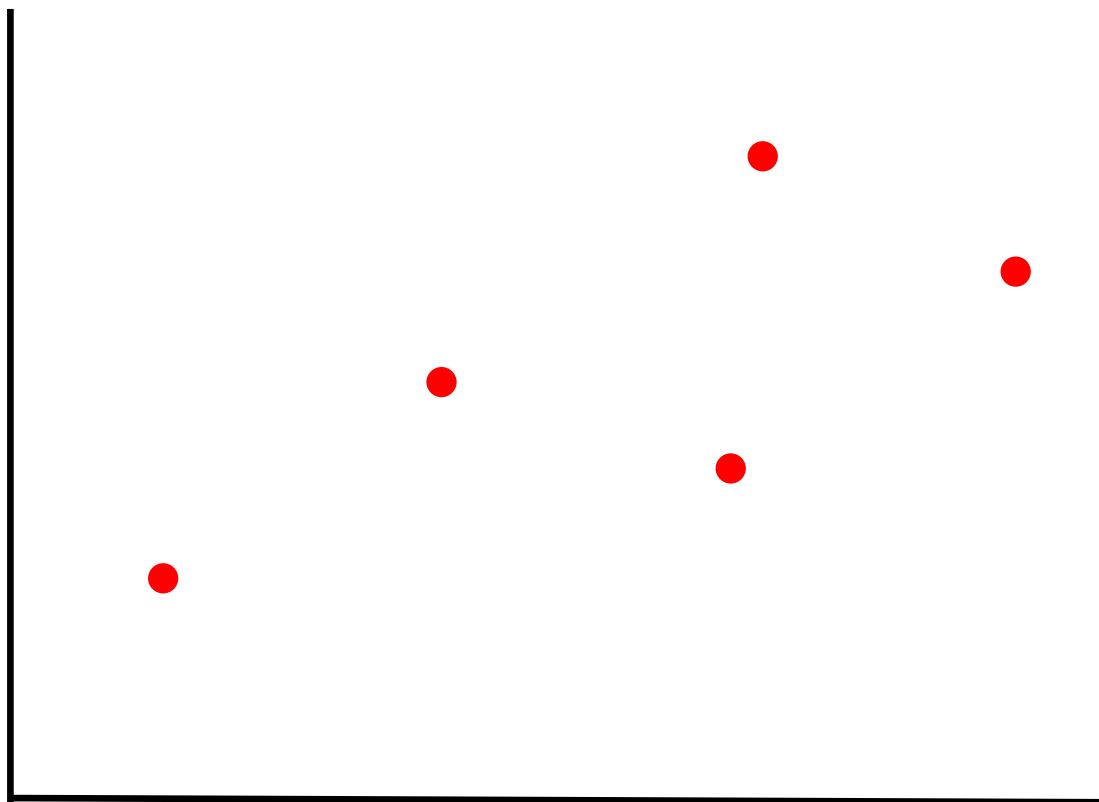
$$\text{売上} = b_0 + b_1(\text{広告費})$$

回帰変数

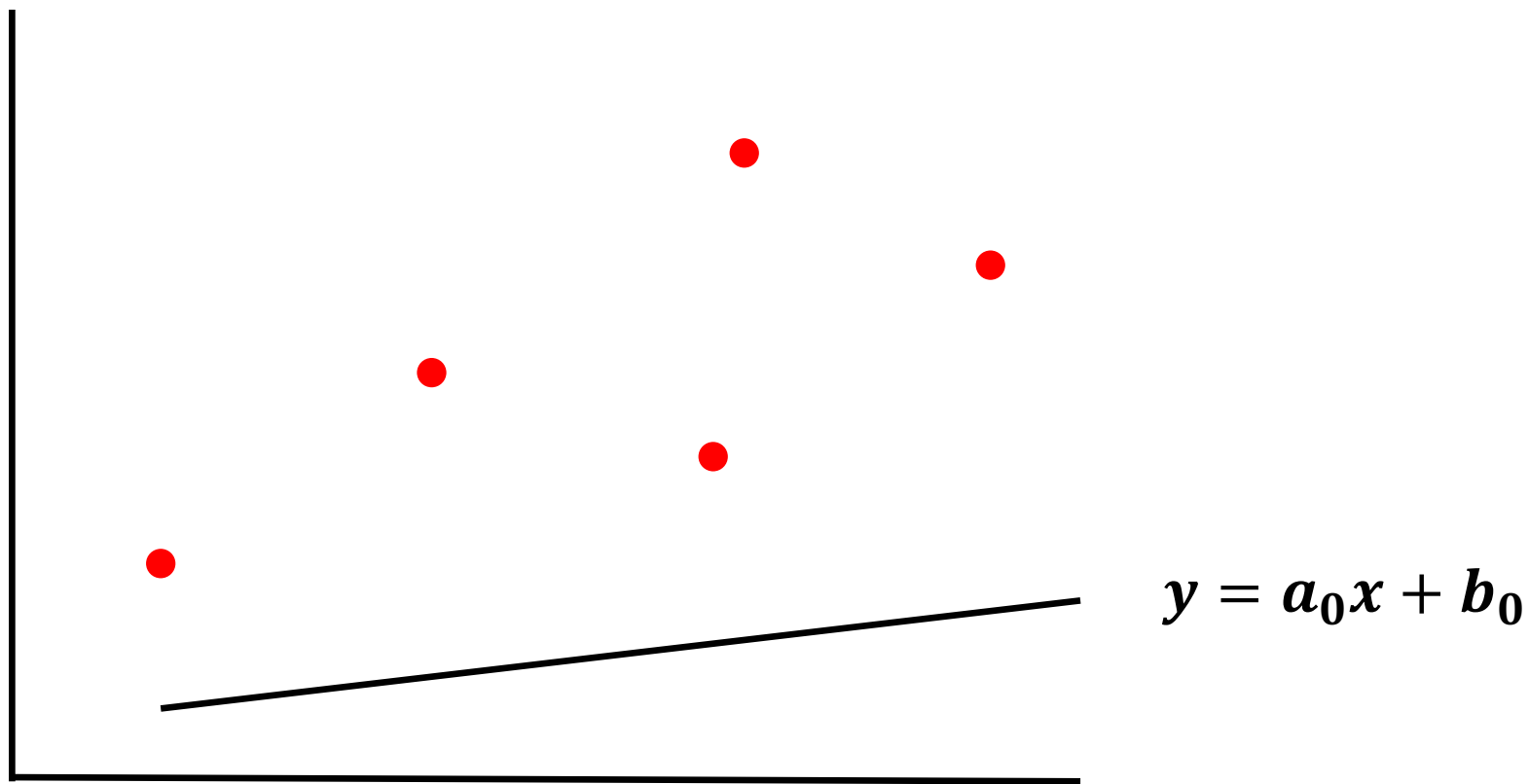
$$y = 3080.0412 + 0.052524x$$

$$y = 3080.0412 + 0.052524(250000) = 16211$$

# 最小二乗法

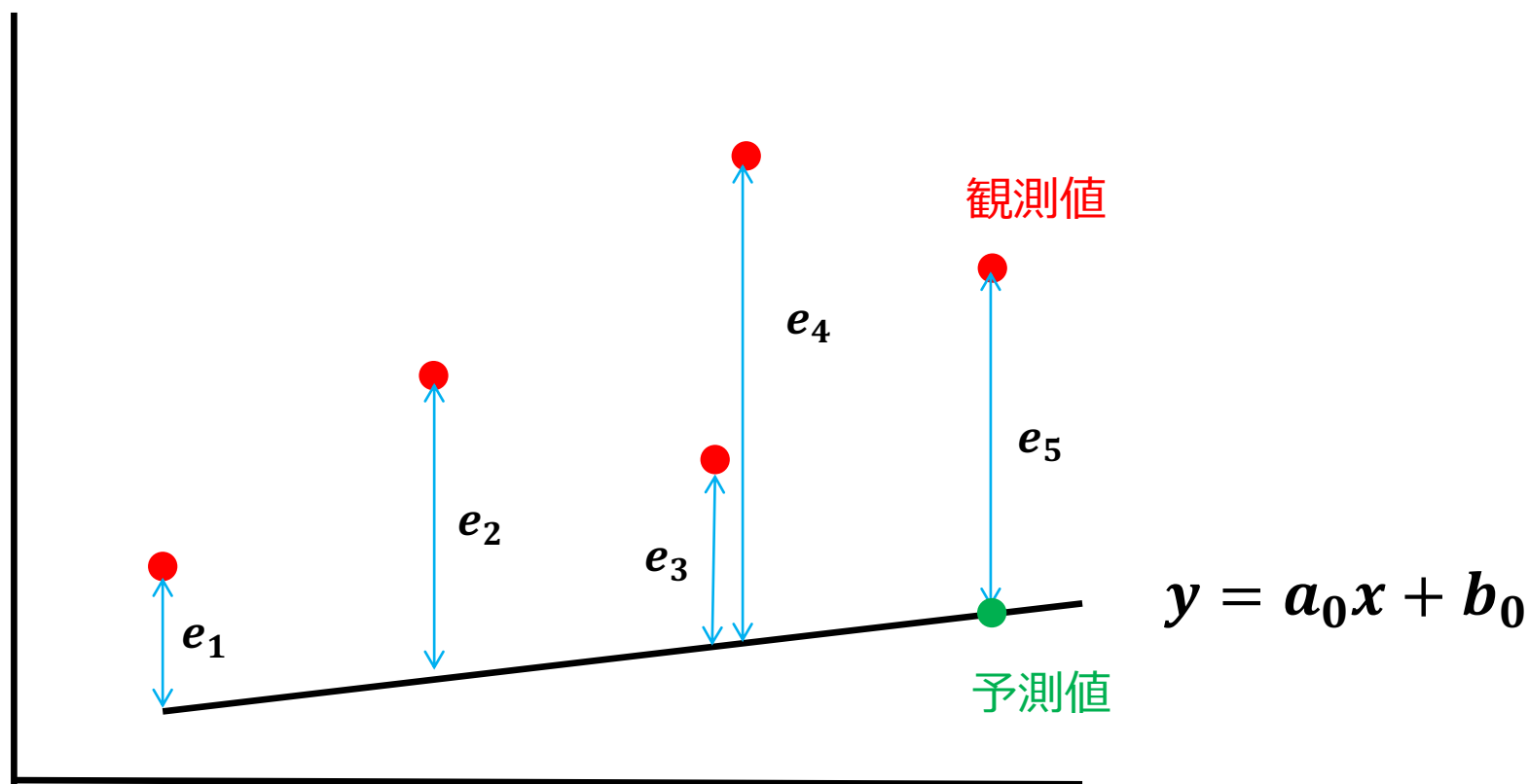


# 最小二乗法



# 最小二乗法

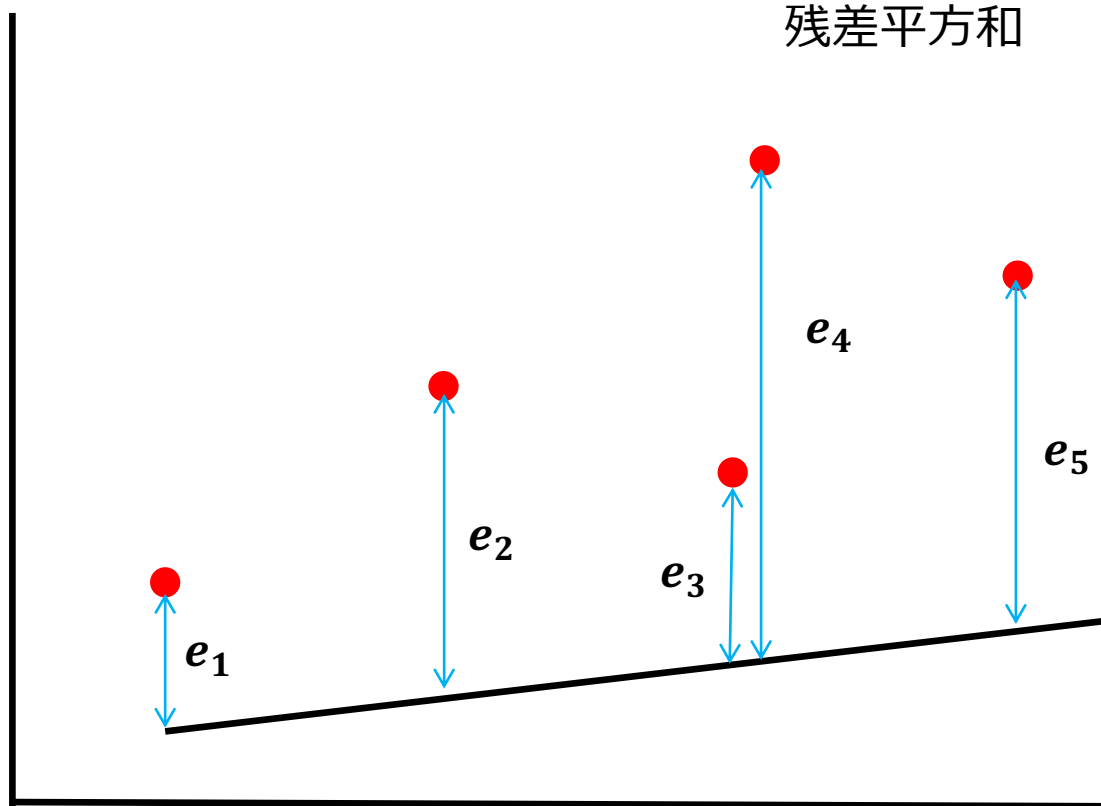
$e_i$ : 残差 = 予測値と観測値の差



# 最小二乗法

$$E = e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2$$

残差平方和



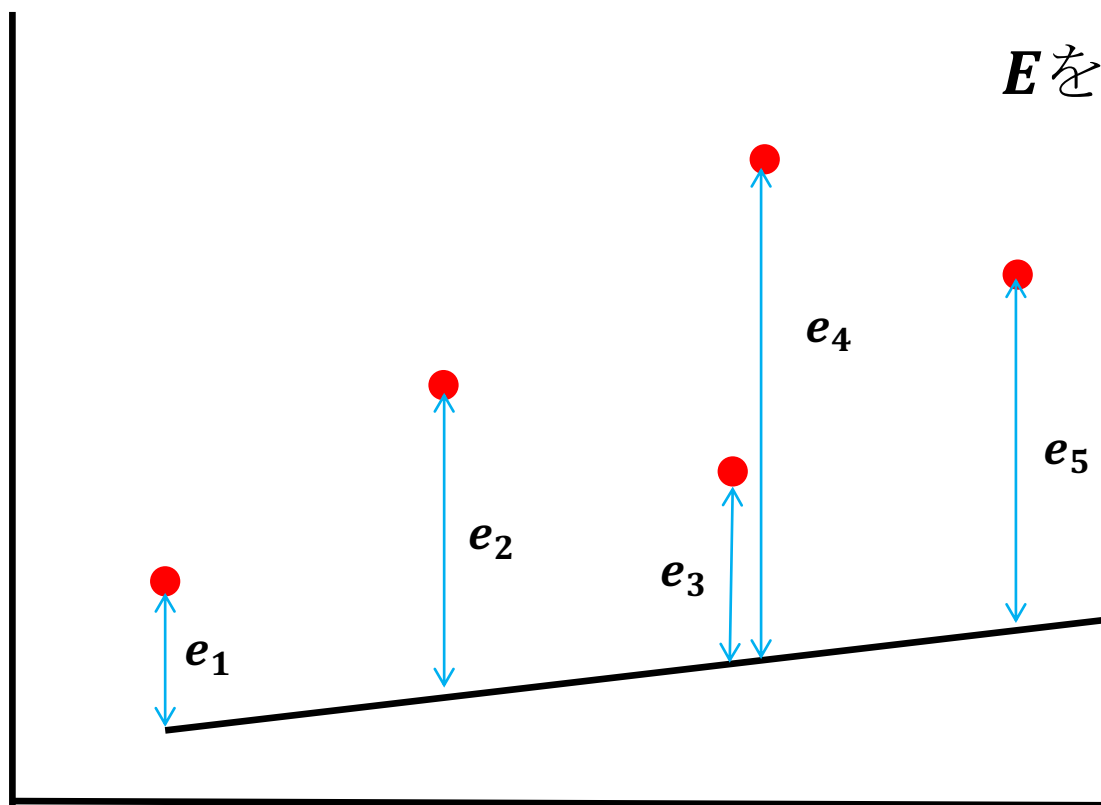
$$y = a_0x + b_0$$



# 最小二乗法

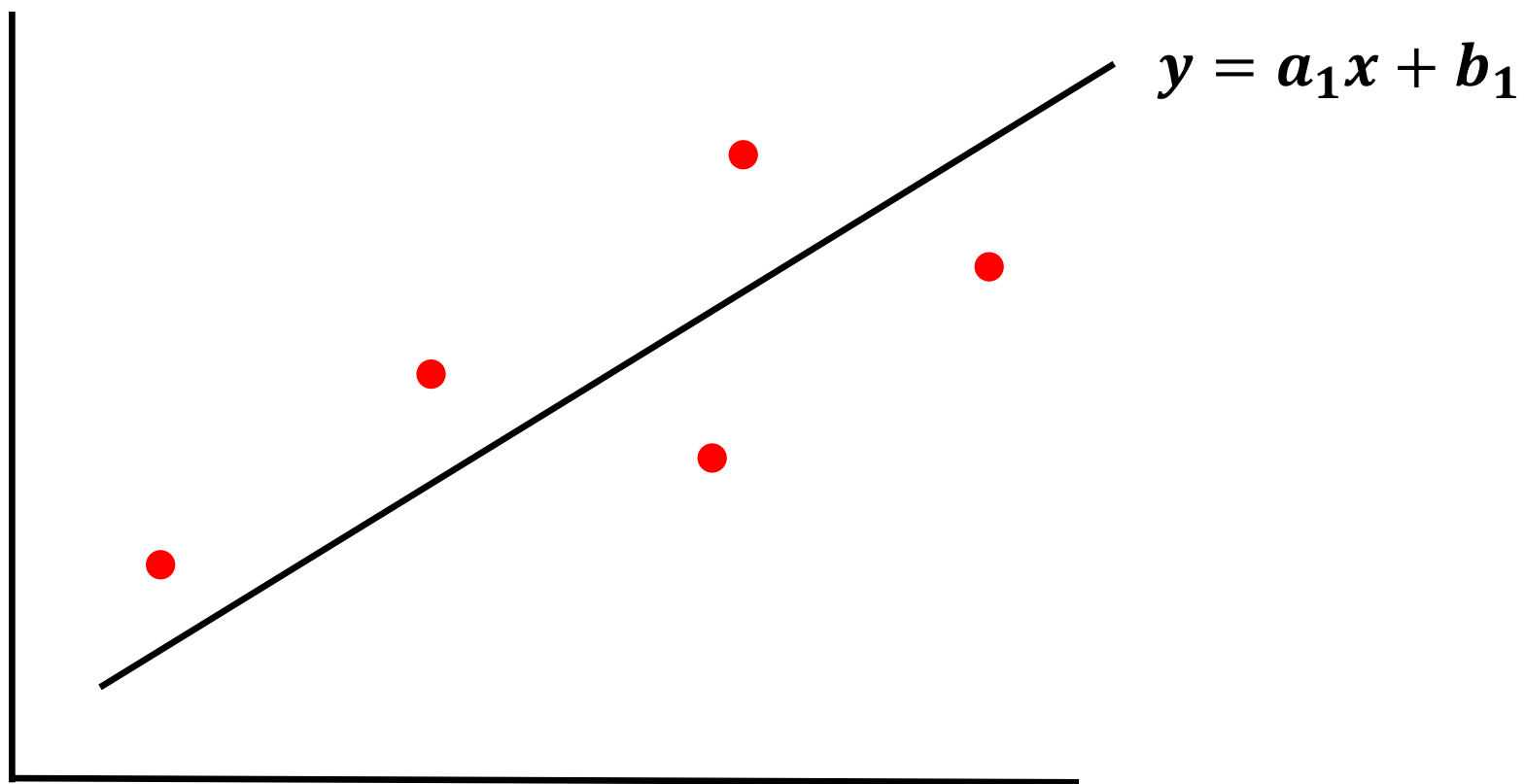
$$E = e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2$$

$E$ を最小にする $a$ と $b$ を探す

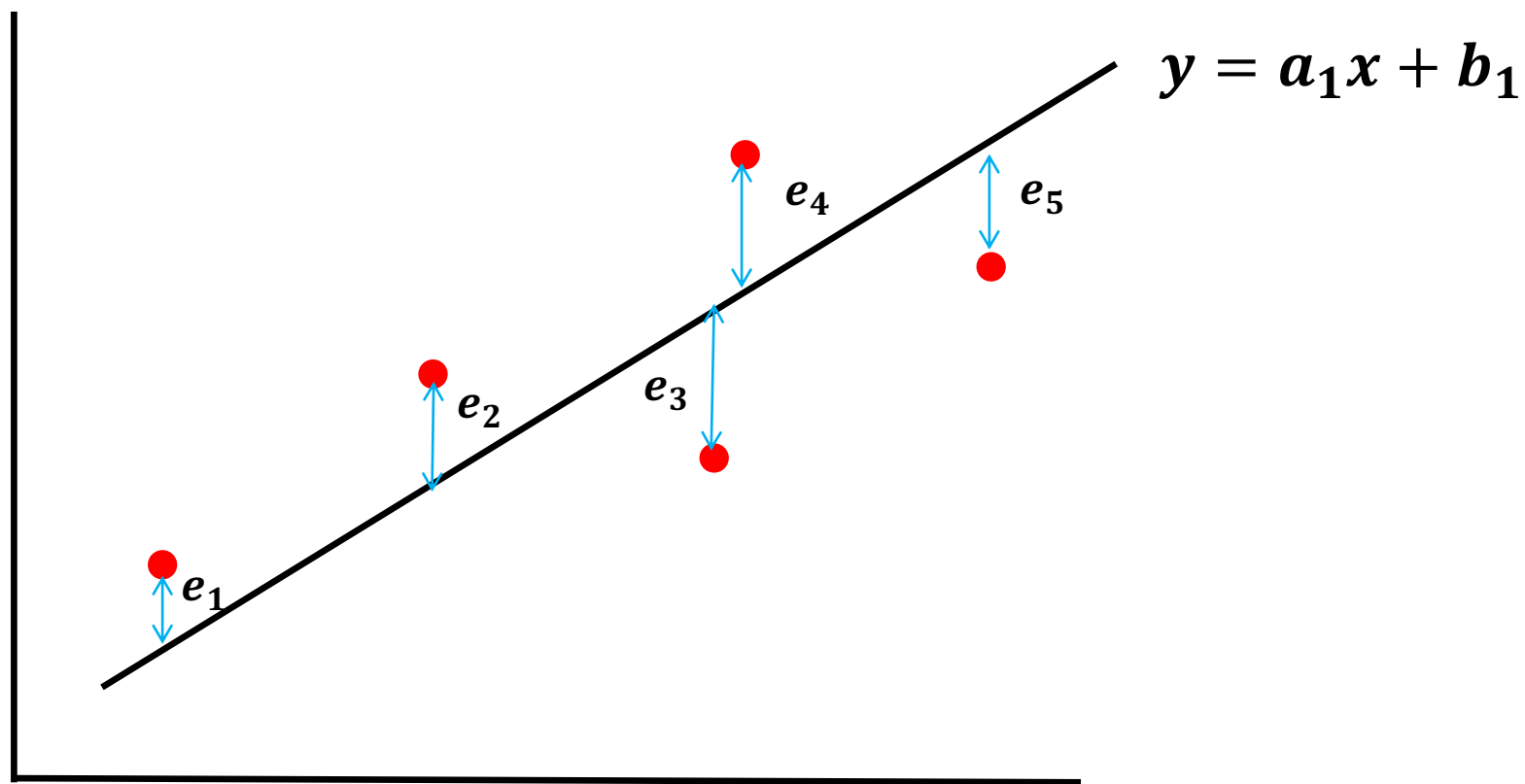


$$y = a_0x + b_0$$

# 最小二乗法

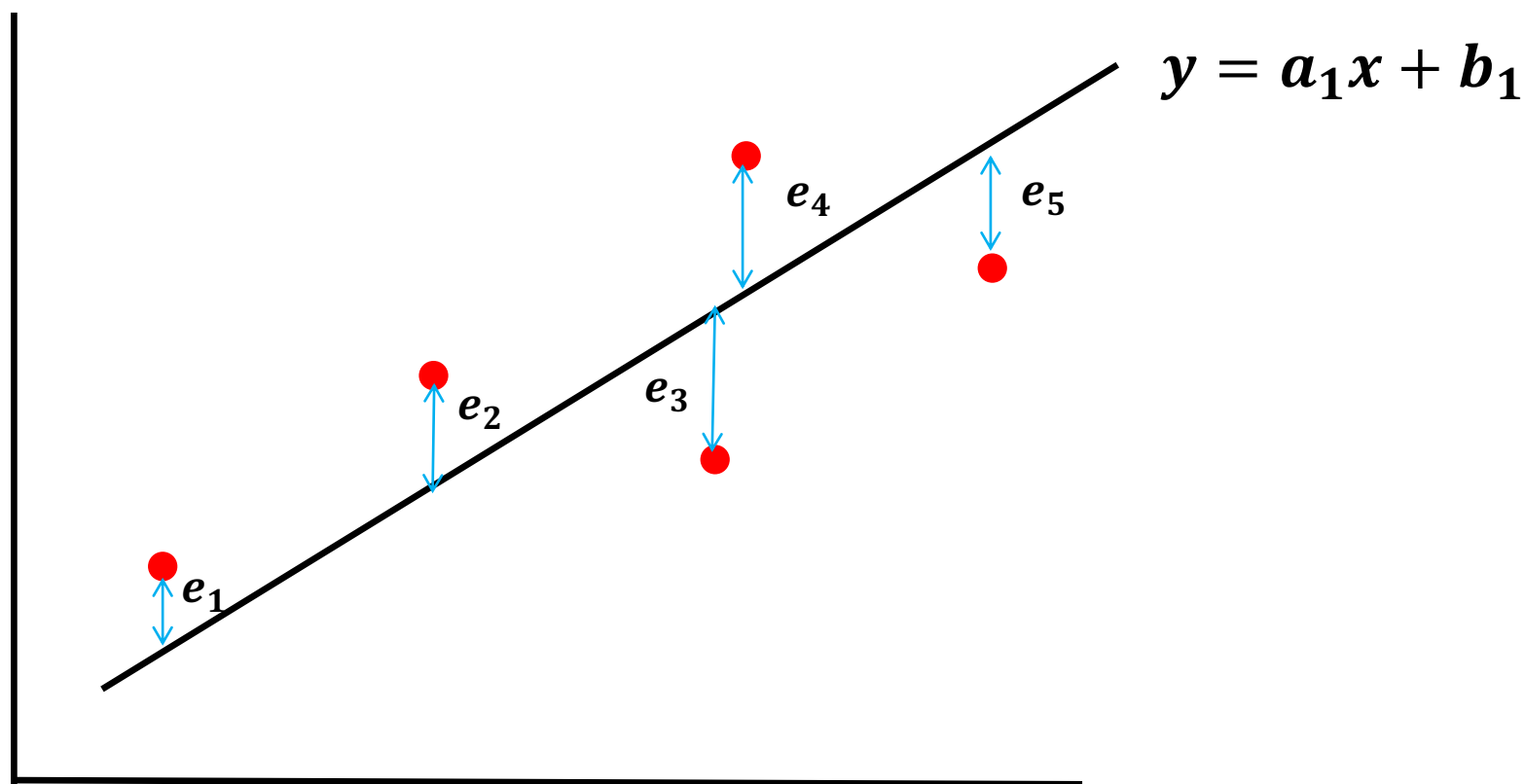


# 最小二乗法

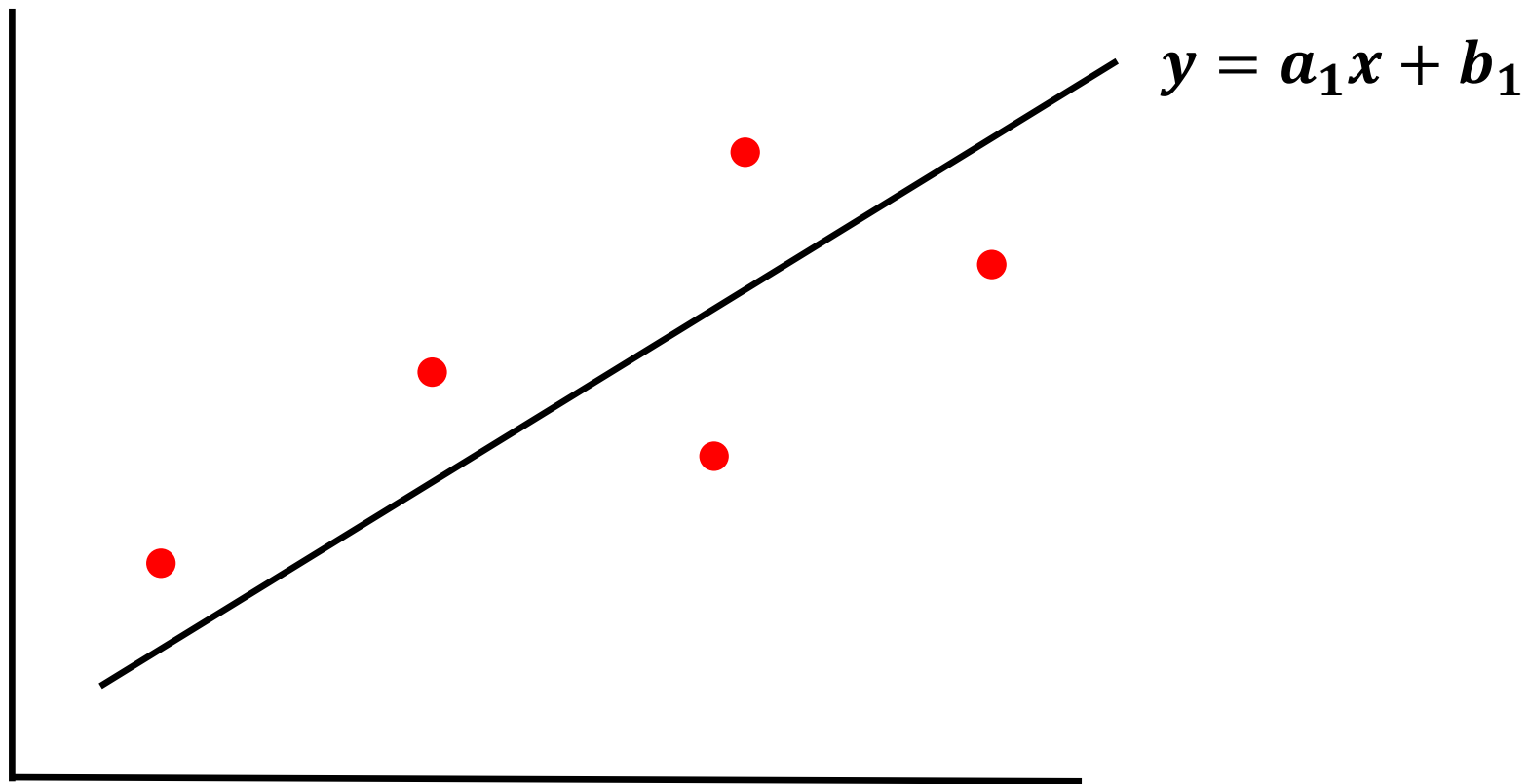


# 最小二乗法

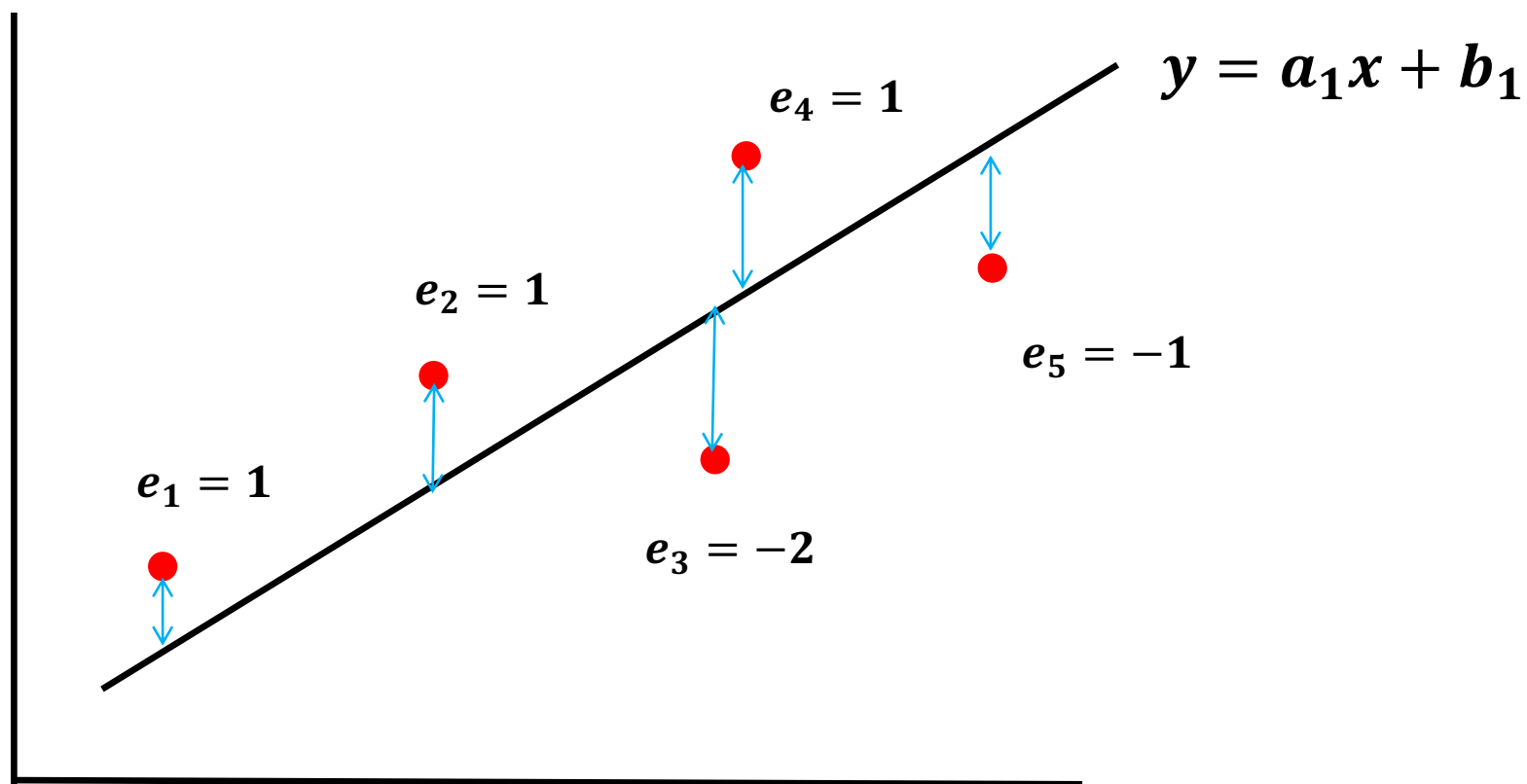
$$E = e_1^2 + e_2^2 + e_3^2 + e_4^2 + e_5^2$$



# なぜ 2 乗するのか？

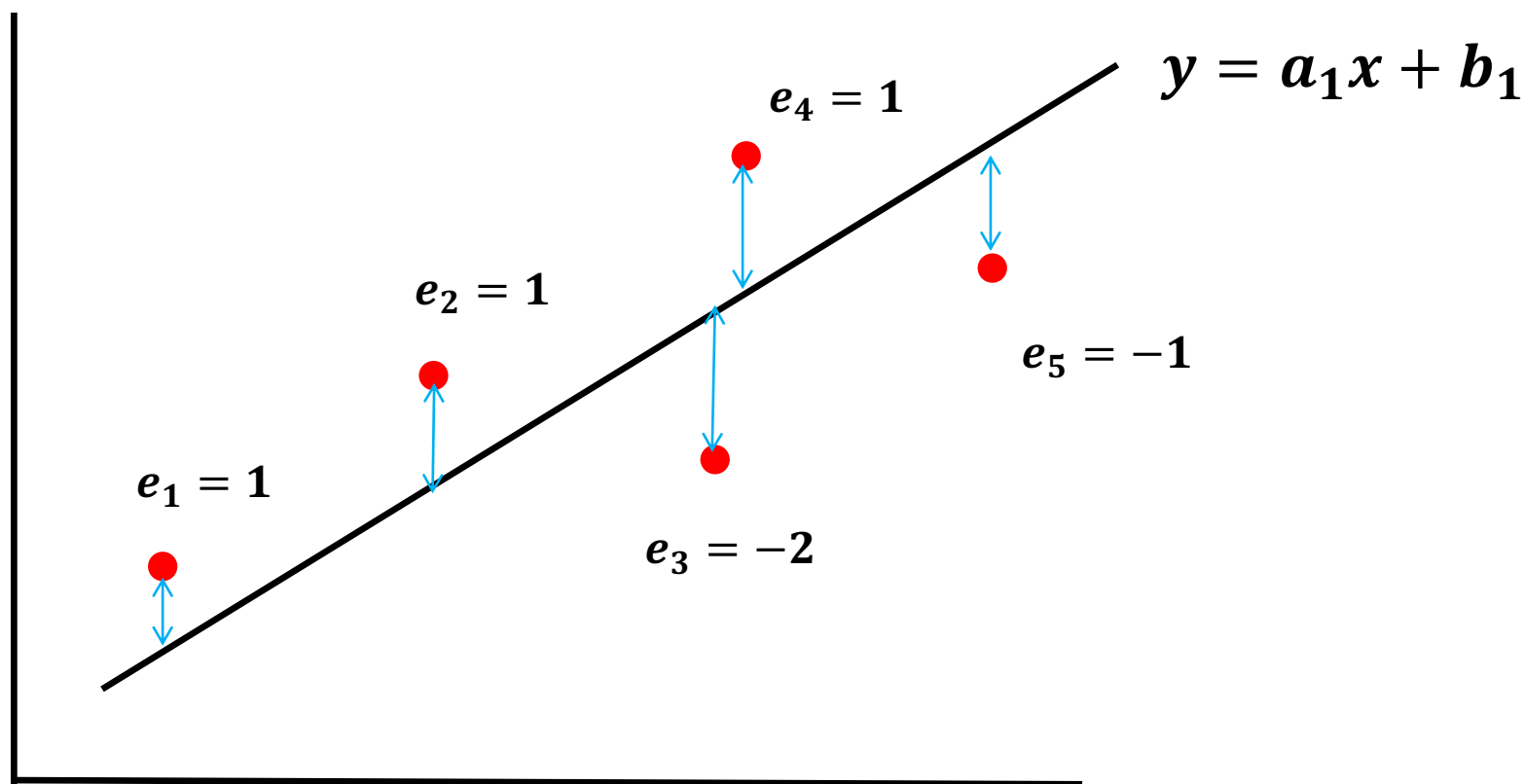


# なぜ 2 乗するのか？

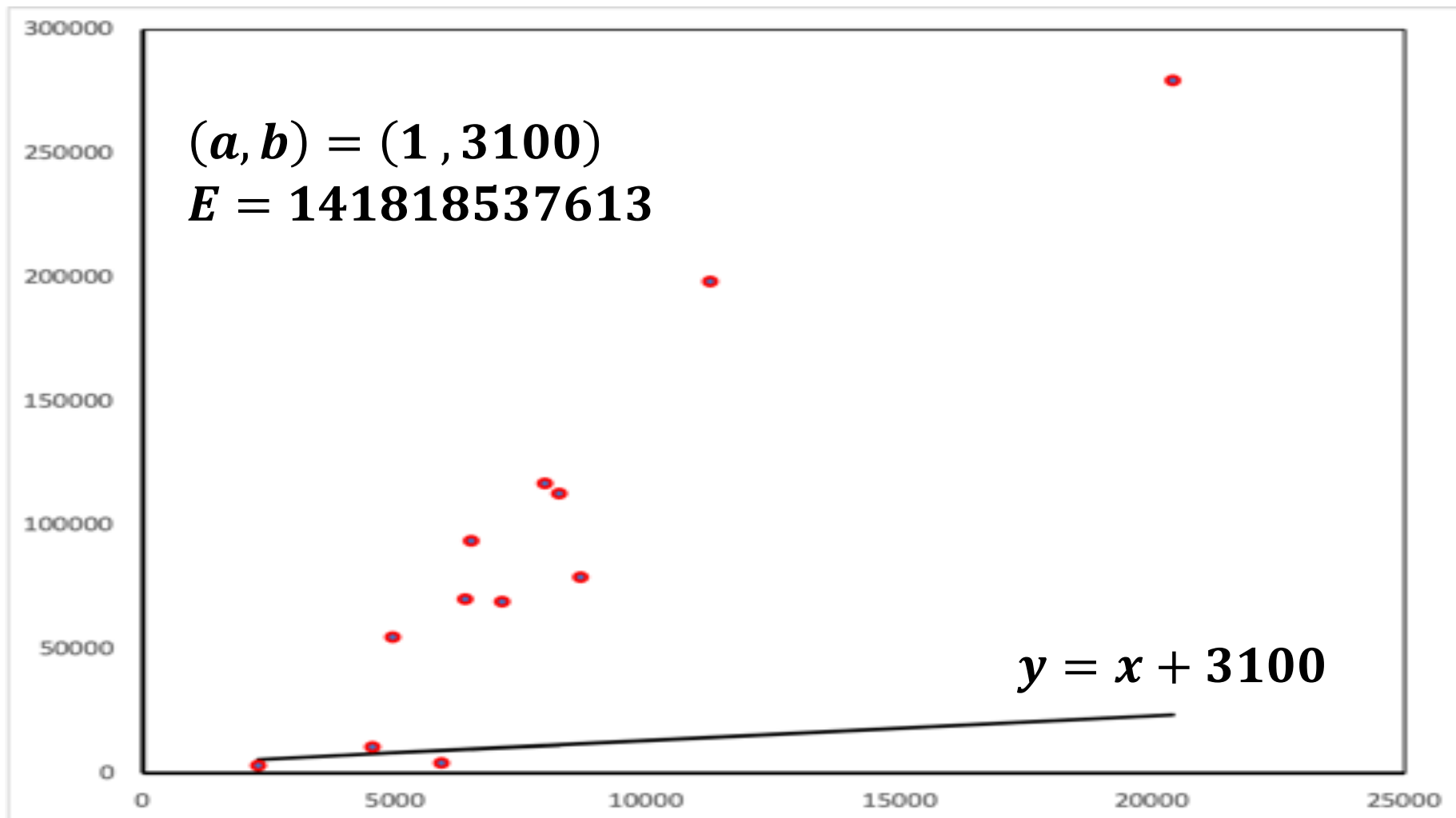


# なぜ 2 乗するのか？

$$E = e_1 + e_2 + e_3 + e_4 + e_5 = 0$$

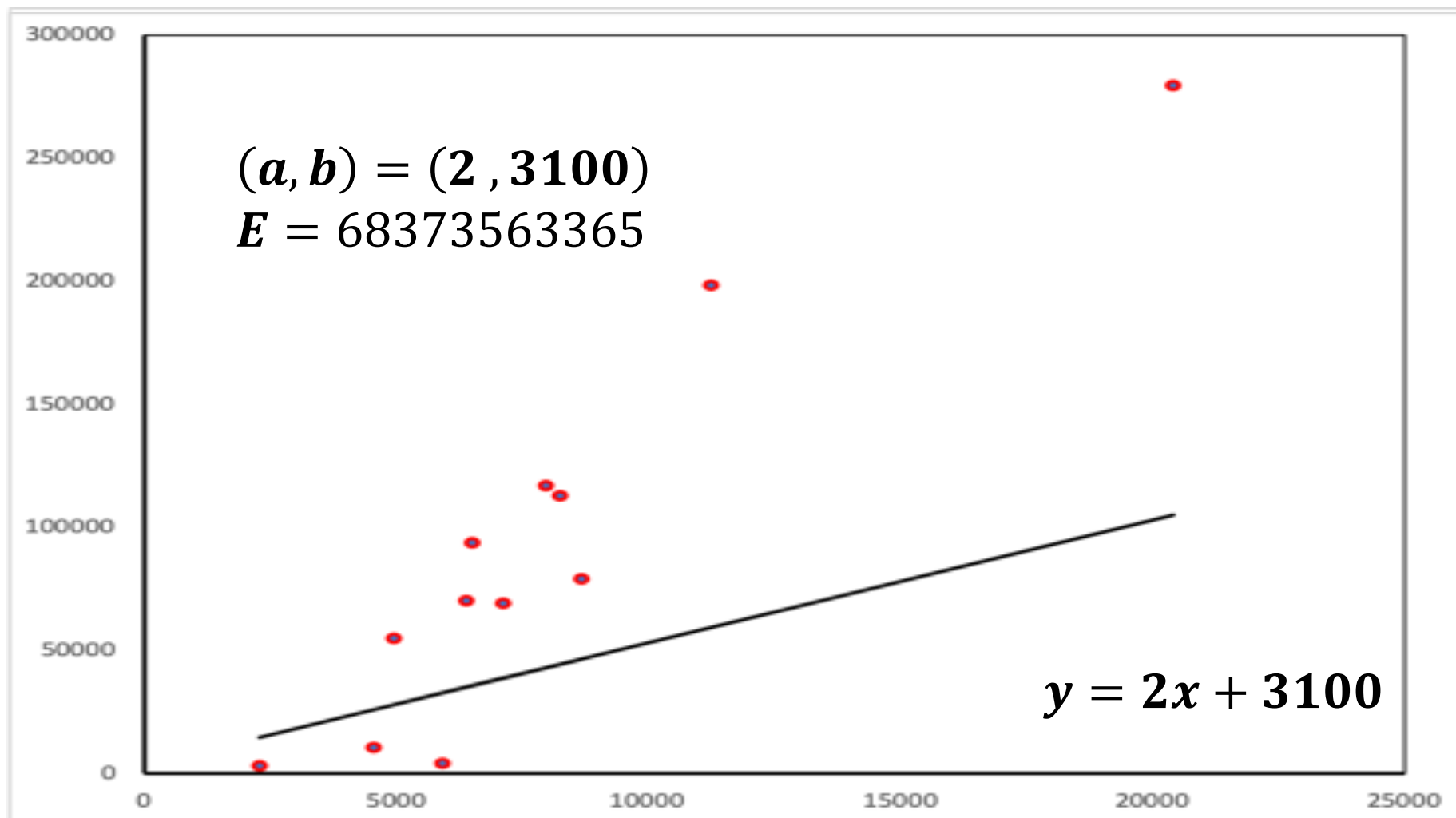


# 最小二乗法

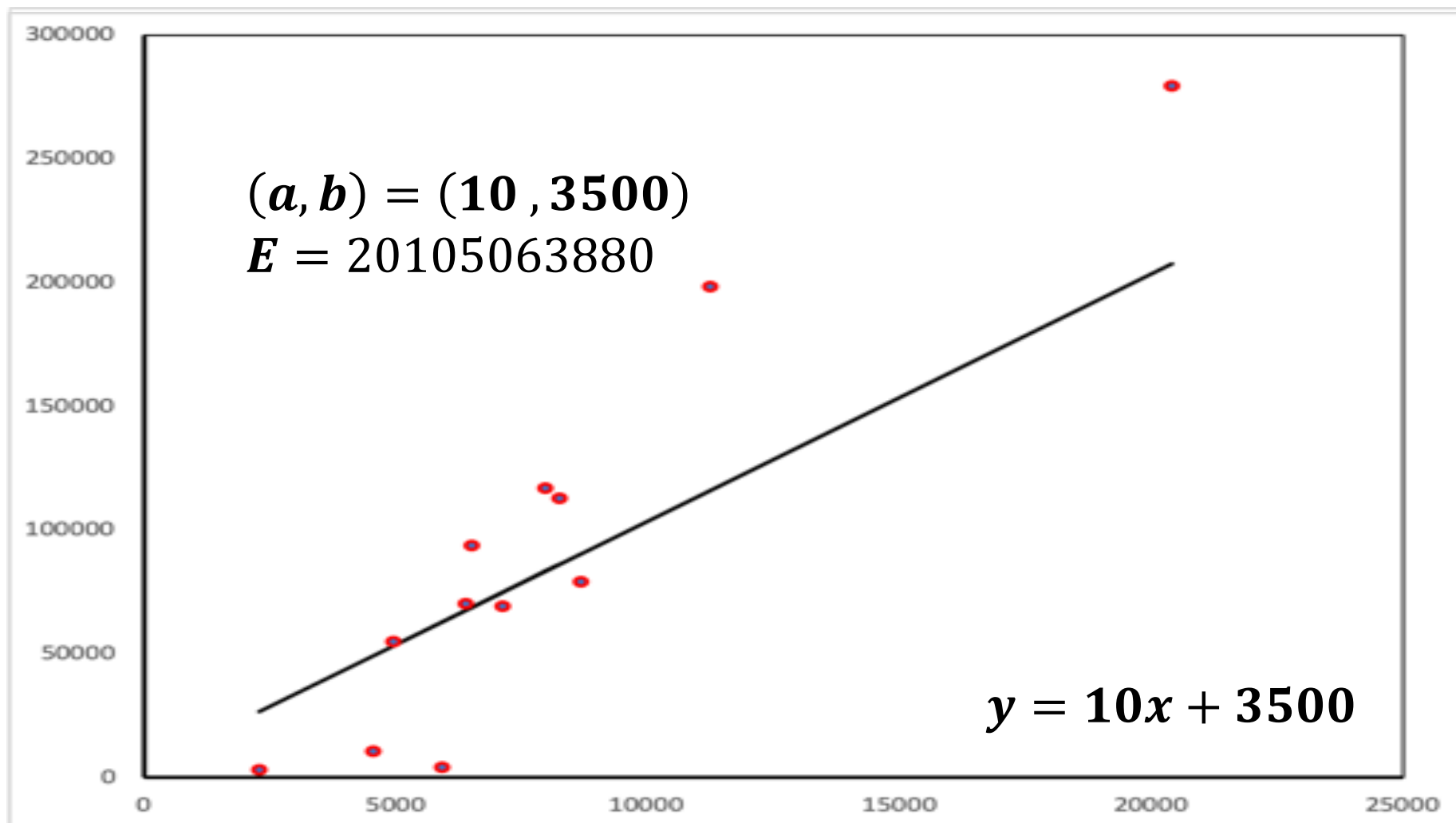




# 最小二乗法

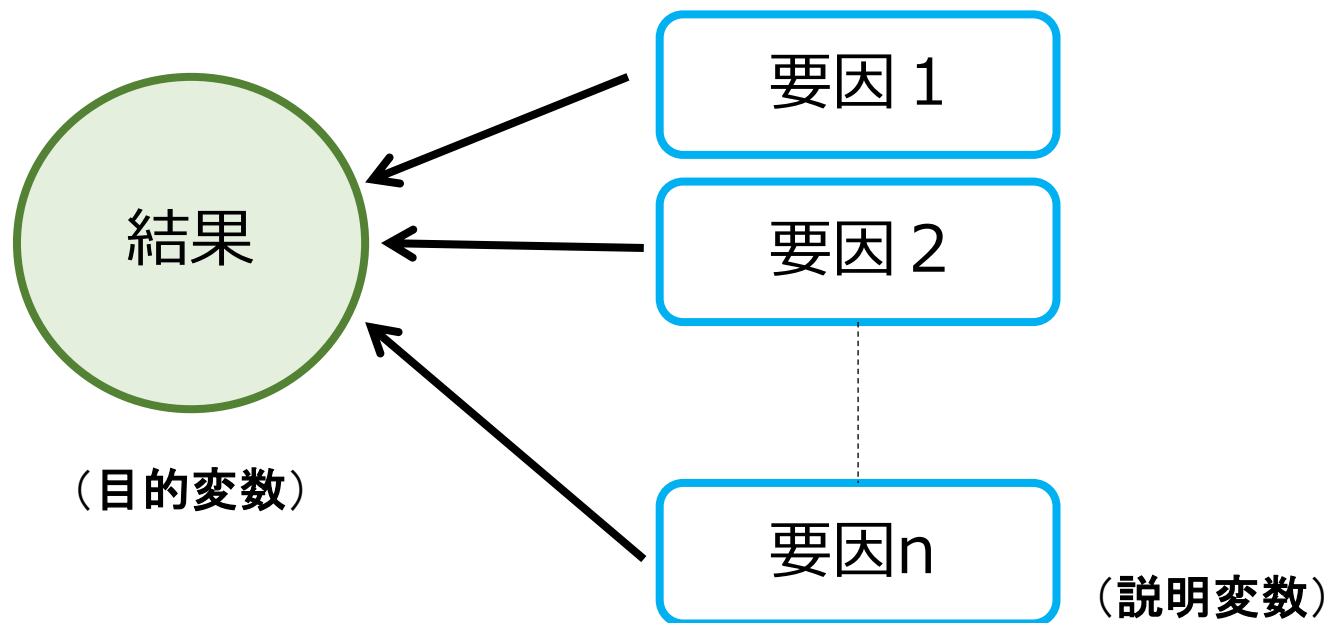


# 最小二乘法



# 重回帰分析モデル

予測したい変数をいくつかの要因となる変数の関係から予測・説明する方法

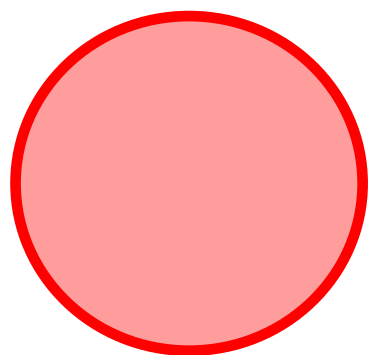


$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n$$

# 回帰分析（手順1：成果を選ぶ）

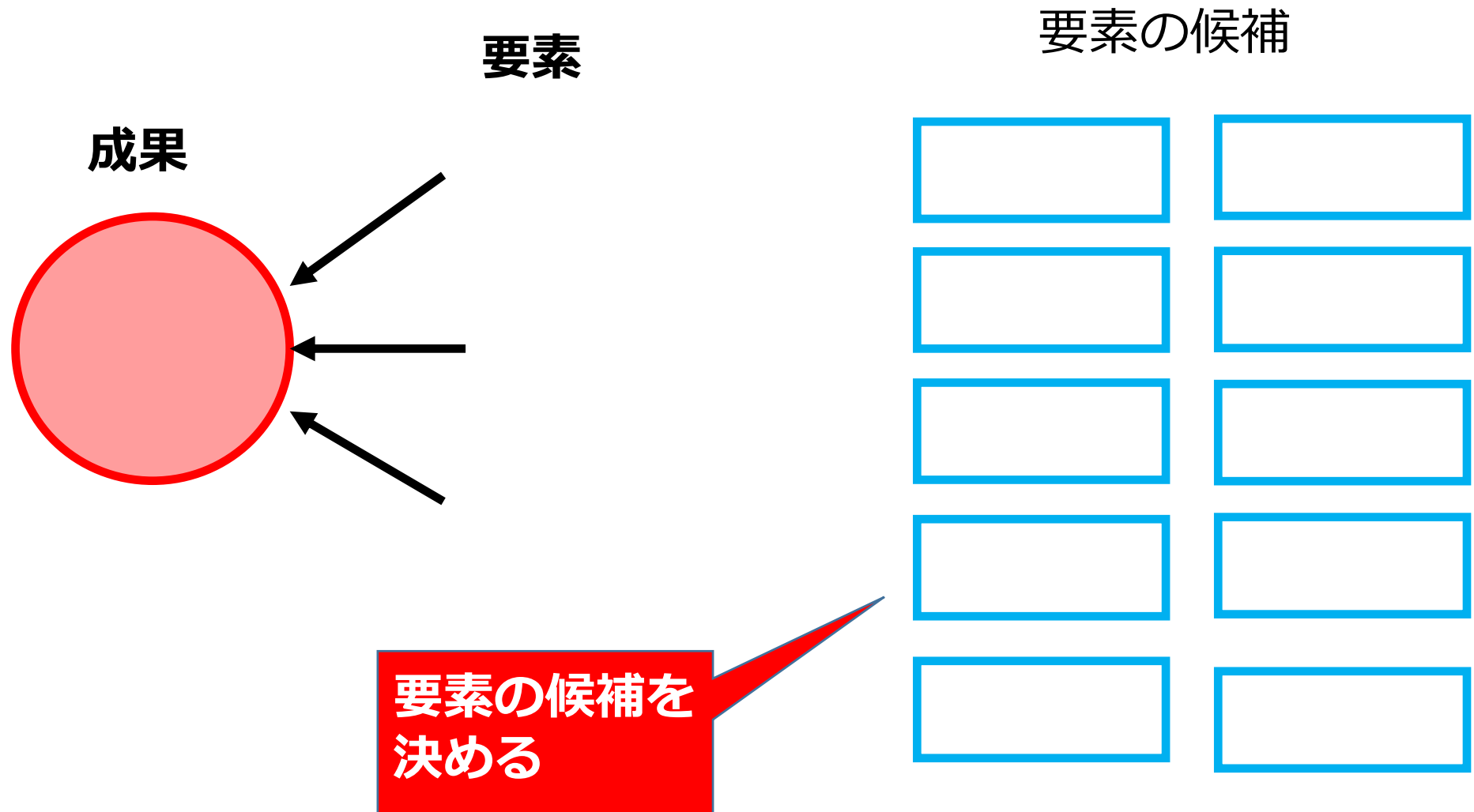
課題を明確に認識すること

**成果**

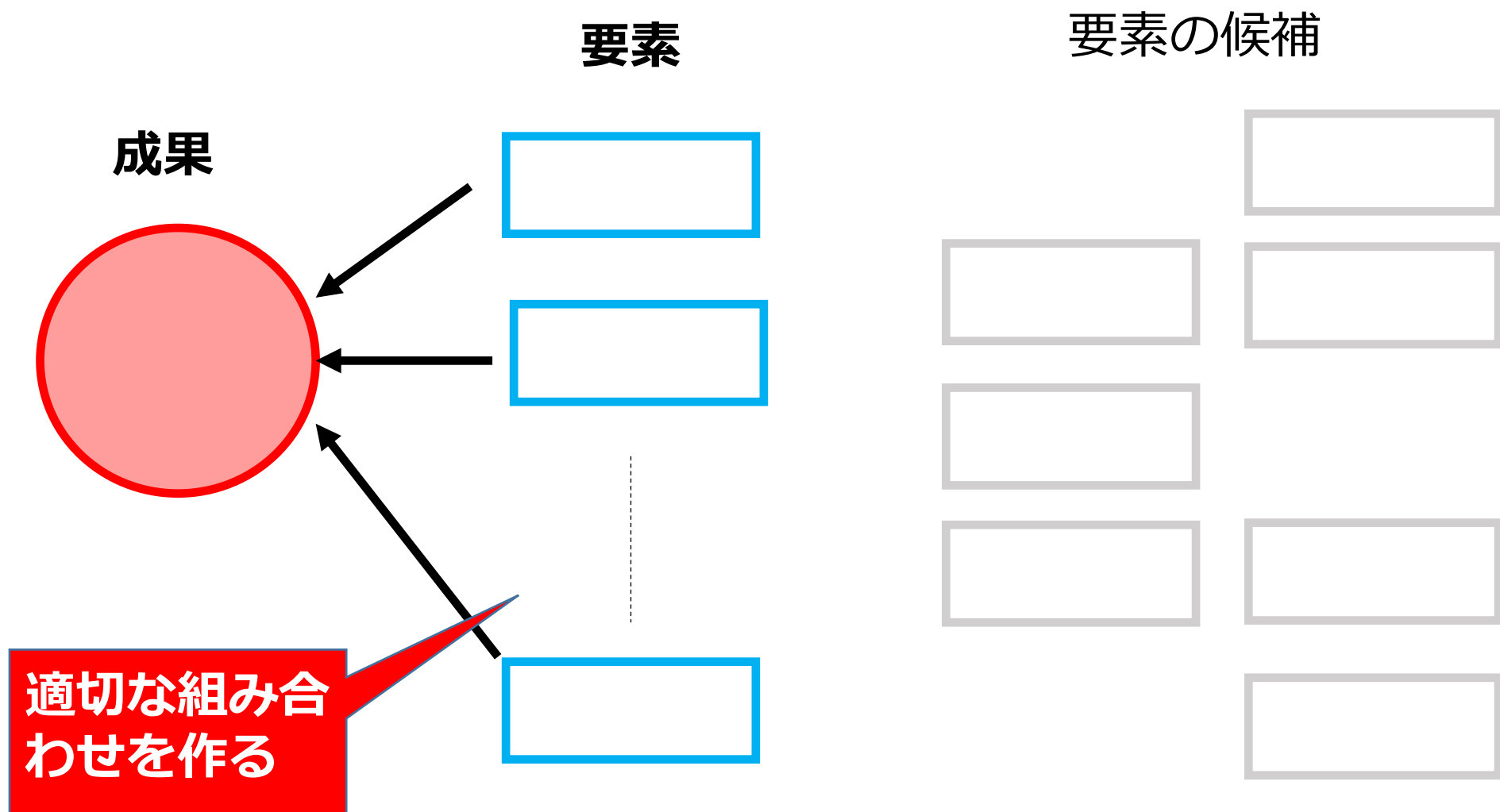


**成果を決める**

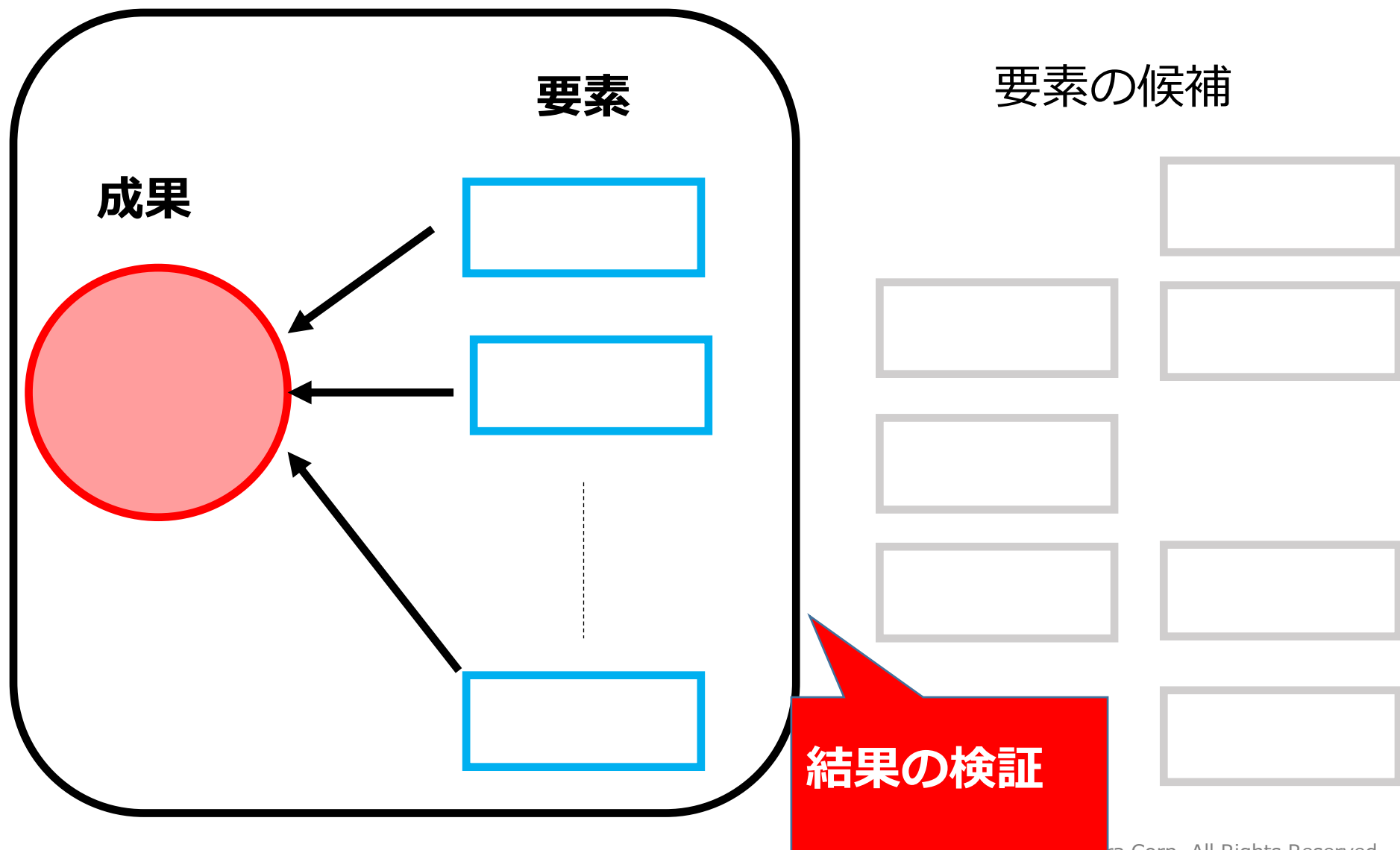
# 回帰分析（手順２：要素をリストアップ）



# 回帰分析（手順3：要素の組み合わせ）



# 回帰分析（手順4：結果の検証）



# 重回帰分析のデータ

店舗	売上	店舗面積	アクセス数	広告	街の規模	同業者店舗数
1	231	3	294	8.2	8.2	11
2	156	2.2	232	6.9	4.1	12
3	10	0.5	149	3	4.3	15
4	519	5.5	260	12	16.1	1
5	437	4.4	567	10.6	14.1	5
6	487	4.8	237	11.8	12.7	4
7	299	3.1	451	8.1	10.1	10
...						
27	400	8.6	335	7	12	8

成果

要因



# 分析結果を読む

## 概要

### 回帰統計

重相関 R	0.995085404
重決定 R2	0.990194961
補正 R2	0.987860428
標準誤差	21.16130807
観測数	27

## 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	5	949676.5317	189935.3063	424.1511825	2.4296E-20
残差	21	9403.820141	447.8009591		
合計	26	959080.3519			

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	-39.1953195	36.61749065	-1.07039884	0.296588834	-115.34556	36.95492102
店舗面積	20.43522388	3.91825741	5.215385754	3.61064E-05	12.28676152	28.58368623
アクセス数	-0.00073248	0.027729311	-0.02641534	0.979175474	-0.05839874	0.05693378
広告	16.966648	2.142086527	7.920617487	9.6594E-08	12.5119352	21.4213608
街の規模	15.68003698	1.973031864	7.947178793	9.15653E-08	11.5768926	19.78318136
同業者店舗数	-4.04959516	1.996639687	-2.02820528	0.055412839	-8.2018347	0.102644377

# 重回帰分析のデータ

店舗	売上	店舗面積	アクセス数	広告	街の規模	同業者店舗数
1	231	3	294	8.2	8.2	11
2	156	2.2	232	6.9	4.1	12
3	10	0.5	149	3	4.3	15
4	519	5.5	260	12	16.1	1
5	437	4.4	567	10.6	14.1	5
6	487	4.8	237	11.8	12.7	4
7	299	3.1	451	8.1	10.1	10
...						
27	400	8.6	335	7	12	8

成果

取り除く

要因

# 分析結果を読む

概要

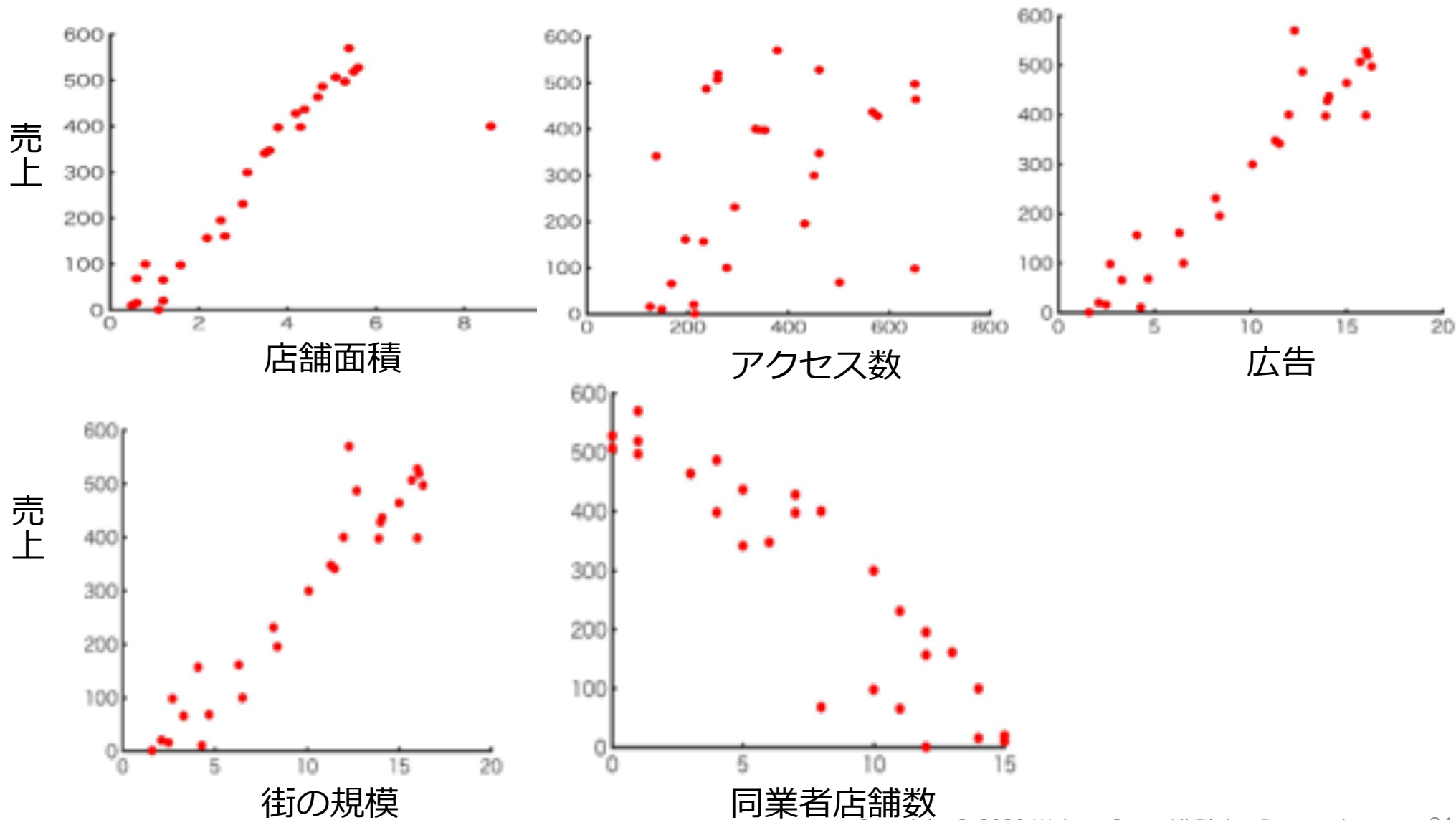
回帰統計	
重相関 R	0.99508524
重決定 R2	0.990194635
補正 R2	0.988411842
標準誤差	20.67511966
観測数	27

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	949676.22	237419.05	555.4174347	9.58E-22
残差	22	9404.1326	427.46057		
合計	26	959080.35			

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	-39.46001387	34.410562	-1.146741	0.263807961	-110.8232	31.903123
店舗面積	20.4438867	3.8148017	5.3590955	2.21824E-05	12.532472	28.355301
広告	16.96614225	2.0927877	8.106958	4.73186E-08	12.625966	21.306318
街の規模	15.67296182	1.9098557	8.206359	3.85791E-08	11.712163	19.63376
同業者店舗数	-4.043301756	1.9368286	-2.087589	0.048629059	-8.060038	-0.026565

# 散布図による可視化



# 重回帰モデルの設計

店舗面積= 6, アクセス数=400, 広告=10, 街の規模=15, 同業者店舗数=10の時の売上を予測せよ

	係数
切片	-39.46001387
店舗面積	20.4438867
広告	16.96614225
街の規模	15.67296182
同業者店舗数	-4.043301756

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$$

$$\text{売上} = -39.4 + 20.4(\text{店舗面積}) + 16.9(\text{広告}) + 15.6(\text{街の規模}) - 4.0(\text{同業者店舗数})$$

$$\text{売上} = -39.4 + 20.4(6) + 16.9(10) + 15.6(15) - 4.0(10) = 447$$

=sumproduct()

	係数	変数の値
切片	-39.46001387	1
店舗面積	20.4438867	6
広告	16.96614225	10
街の規模	15.67296182	15
同業者店舗数	-4.043301756	10

=sumproduct(配列 1、配列 2 )

447.56261384

# 偏回帰係数

$$y = b_0 + b_1x_1 + b_2x_2 + \dots$$

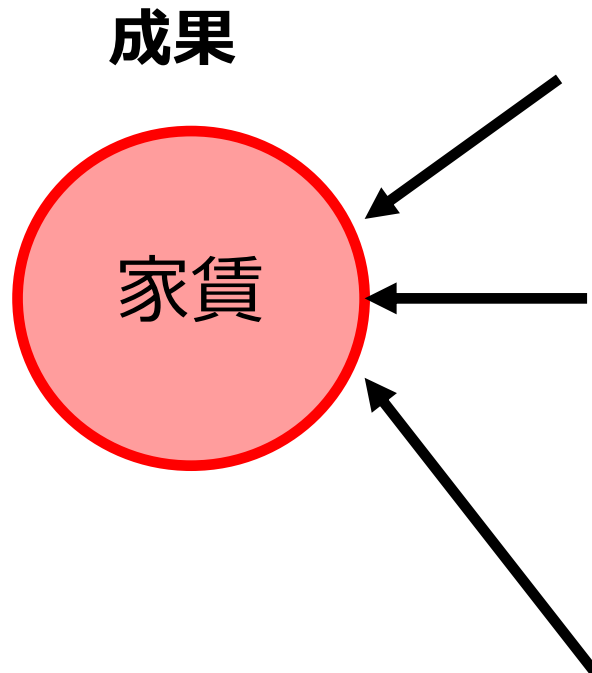
他の変数が一定という条件で、変数が1 増加した時、  
目的変数がいくら変化するかを表す値

$$\text{売上} = -39.4 + 20.4(\text{店舗面積}) + 16.9(\text{広告}) + 15.6(\text{街の規模}) - 4.0(\text{同業者店舗数})$$

他の変数が一定という条件で、(店舗面積) が1 増加した時、  
売上が+20.4変化することを表す値

# 問題：マンションの家賃を予測モデル

問題：家賃に影響を与えると考えられる要因を挙げてください



要素の候補




# 問題：マンションの家賃を予測モデル

家賃	徒歩	専有面積	築年数	階数	新築	2階以上	南向き	オート ロック	エアコン	バス トイレ別	追い焚き	フロー リング	ペット 相談可
88500	10	20.7	3	8	無	有	無	有	有	有	無	有	有
86700	10	20.7	3	7	無	有	有	有	有	有	無	有	有
87300	10	20.7	3	2	無	有	有	有	有	有	無	有	有

# ダミー変数

重回帰分析は数字などの量的データ以外にも、質的データを数字に変換して分析に取り入れることが可能。その時に使われるのがダミー変数。

ダミー変数「0」と「1」の数値で質的データを表現する

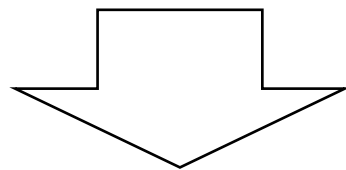
オートロック（無） 「0」  $\longleftrightarrow$  「1」 オートロック（有）

ペット相談（不可） 「0」  $\longleftrightarrow$  「1」 ペット相談（可）

女性 「0」  $\longleftrightarrow$  「1」 男性

# ダミー変数を使った重回帰分析

家賃	徒歩	専有面積	築年数	階数	新築	2階以上	南向き	オート ロック	エアコン	バス トイレ別	追い焚き	フロー リング	ペット 相談可
88500	10	20.7	3	8	無	有	無	有	有	有	無	有	有
86700	10	20.7	3	7	無	有	有	有	有	有	無	有	有
87300	10	20.7	3	2	無	有	有	有	有	有	無	有	有



家賃	徒歩	専有面積	築年数	階数	新築	2階以上	南向き	オート ロック	エアコン	バス トイレ別	追い焚き	フロー リング	ペット 相談可
88500	10	20.7	3	8	0	1	0	1	1	1	0	1	1
86700	10	20.7	3	7	0	1	1	1	1	1	0	1	1
87300	10	20.7	3	2	0	1	1	1	1	1	0	1	1

# 問題

次の条件を満たす家賃を予測してください。

- |           |     |
|-----------|-----|
| • 徒歩（分）   | 10分 |
| • 専有面積（㎡） | 2 5 |
| • 築年数（年）  | 3   |
| • 階数（階）   | 2   |
| • 新築      | 有   |
| • 2階以上    | 有   |
| • 南向き     | 有   |
| • オートロック  | 無   |
| • エアコン    | 無   |
| • バストイレ別  | 無   |
| • 追い焚き    | 無   |
| • フローリング  | 無   |
| • ペット相談可  | 無   |

# 分析結果を読む（p値による変数選択）

概要

回帰統計	
重相関 R	0.973201151
重決定 R2	0.947120479
修正 R2	0.939127064
標準誤差	7089.805967
観測数	100

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	13	77425651616	5955819355	118.4875767	4.70069E-49
残差	86	4322819984	50265348.65		
合計	99	81748471600			

20%以下を選択

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	36013.10074	5357.526756	6.721963769	1.86369E-09	25362.69004	46663.51144
徒歩（分）	-1102.804577	233.9383469	-4.714082117	9.28048E-06	-1567.858619	-637.7505356
専有面積（㎡）	2768.615702	115.4289718	23.98544888	7.39934E-40	2539.150501	2998.080903
築年数（年）	-615.0838365	325.6465636	-1.888813806	0.062287367	-1262.445792	32.27811853
階数（階）	1493.727055	250.804868	5.955733903	5.50332E-08	995.143479	1992.310631
新築	4114.591046	3095.252457	1.329323247	0.187257015	-2038.56719	10267.74928
2階以上	-1591.283103	2824.127821	-0.563460015	0.574587944	-7205.463401	4022.897194
南向き	-1037.148762	1710.210358	-0.606445141	0.545816742	-4436.934433	2362.63691
オートロック	9943.601805	2869.207823	3.4656262	0.000827665	4239.805428	15647.39818
エアコン	-1764.29321	3232.409006	-0.545813728	0.586607034	-8190.109642	4661.523222
バストイレ別	-1108.051295	2212.963955	-0.500709148	0.617854899	-5607.278038	3291.175447
追い焚き	455.6810369	2546.987933	0.178909775	0.858429365	-4607.563398	5518.925471
フローリング	-5224.498103	3101.137799	-1.684703629	0.095671703	-11389.35601	940.3598055
ペット相談可	3767.798631	1980.216482	1.902720569	0.060423229	-168.7414498	7704.338711

# 分析結果を読む（p値による変数選択）

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	36013.10074	5357.526756	6.721963769	0.000000	25362.69004	46663.51144
徒歩（分）	-1102.804577	233.9383469	-4.714082117	0.000009	-1567.858619	-637.7505356
専有面積（㎡）	2768.615702	115.4289718	23.98544888	0.000000	2539.150501	2998.080903
築年数（年）	-615.0838365	325.6455636	-1.888813806	0.062287	-1262.445792	32.27811853
階数（階）	1493.727055	250.804868	5.955733903	0.000000	995.143479	1992.310631
新築	4114.591046	3095.252457	1.329323247	0.187257	-2038.56719	10267.74928
2階以上	-1591.283103	2824.127821	-0.563460015	0.574588	-7205.463401	4022.897194
南向き	-1037.148762	1710.210358	-0.606445141	0.545817	-4436.934433	2362.63691
オートロック	9943.601805	2869.207823	3.4656262	0.000828	4239.805428	15647.39818
エアコン	-1764.29321	3232.409006	-0.545813728	0.586607	-8190.109642	4661.523222
バストイレ別	-1108.051295	2212.963955	-0.500709148	0.617855	-5507.278038	3291.175447
追い焚き	455.6810369	2546.987933	0.178909775	0.858429	-4607.563398	5518.925471
フローリング	-5224.498103	3101.137799	-1.684703629	0.095672	-11389.35601	940.3598055
ペット相談可	3767.798631	1980.216482	1.902720569	0.060423	-168.7414498	7704.338711

# 変数選択

家賃	徒歩	専有面積	築年数	階数	新築	2階以上	南向き	オート ロック	エアコン	バス トイレ別	追い焚き	フロー リング	ペット 相談可
88500	10	20.7	3	8	0	1	0	1	1	1	0	1	1
86700	10	20.7	3	7	0	1	1	1	1	1	0	1	1
87300	10	20.7	3	2	0	1	1	1	1	1	0	1	1



家賃	徒歩	専有面積	築年数	階数	新築	オートロック	フローリング	ペット相談可
88500	10	20.7	3	8	0	1	1	1
86700	10	20.7	3	7	0	1	1	1
87300	10	20.7	3	2	0	1	1	1

# 分析結果を読む

概要

回帰統計	
重相関 R	0.9726973
重決定 R2	0.9461401
補正 R2	0.9414051
標準誤差	6955.8794
観測数	100

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	8	77345504067	9668188008	199.8209394	2.843E-54
残差	91	4402967533	48384258.6		
合計	99	81748471600			

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	33880.241	4773.553426	7.09748866	2.70246E-10	24398.164	43362.319
徒歩（分）	-1125.206	221.4082399	-5.0820401	1.9813E-06	-1565.006	-685.4053
専有面積（㎡）	2753.4294	88.56517668	31.0893006	2.95882E-50	2577.5056	2929.3532
築年数（年）	-651.6833	277.6547928	-2.3470989	0.021089235	-1203.21	-100.1561
階数（階）	1384.2767	227.8510003	6.07535954	2.83457E-08	931.67873	1836.8748
新築	4091.1646	2864.307393	1.42832596	0.156621325	-1598.43	9780.7594
オートロック	9386.826	2710.941167	3.46257092	0.000817233	4001.8744	14771.778
フローリング	-5540.621	2919.876233	-1.8975534	0.060925466	-11340.6	259.3543
ペット相談可	3692.6441	1730.218482	2.13420682	0.035514633	255.77762	7129.5106



# 問題

次の条件を満たす家賃を予測してください。

- |           |     |
|-----------|-----|
| • 徒歩（分）   | 10分 |
| • 専有面積（㎡） | 2 5 |
| • 築年数（年）  | 3   |
| • 階数（階）   | 2   |
| • 新築      | 有   |
| • 2階以上    | 有   |
| • 南向き     | 有   |
| • オートロック  | 無   |
| • エアコン    | 無   |
| • バストイレ別  | 無   |
| • 追い焚き    | 無   |
| • フローリング  | 無   |
| • ペット相談可  | 無   |

# 重回帰モデルの設計

	係数
切片	33880.241
徒歩（分）	-1125.206
専有面積（㎡）	2753.4294
築年数（年）	-651.6833
階数（階）	1384.2767
新築	4091.1646
オートロック	9386.826
フローリング	-5540.621
ペット相談可	3692.6441

$$\begin{aligned}\text{家賃} = & 33880.241 \\ & -1125.26(\text{徒歩}) \\ & +2753.4294(\text{専有面積}) \\ & -651.6833(\text{築年数}) \\ & +1384.2767(\text{階数}) \\ & +4091.1646(\text{新築}) \\ & +9386.826(\text{オートロック}) \\ & -5540.621(\text{フローリング}) \\ & +3692.6441(\text{ペット相談可})\end{aligned}$$

# 解答

次の条件を満たす家賃を予測してください。

- |           |     |
|-----------|-----|
| ・ 徒歩（分）   | 10分 |
| ・ 専有面積（㎡） | 2 5 |
| ・ 築年数（年）  | 3   |
| ・ 階数（階）   | 2   |
| ・ 新築      | 有   |
| ・ 2階以上    | 有   |
| ・ 南向き     | 有   |
| ・ オートロック  | 無   |
| ・ エアコン    | 無   |
| ・ バストイレ別  | 無   |
| ・ 追い焚き    | 無   |
| ・ フローリング  | 無   |
| ・ ペット相談可  | 無   |

$$\begin{aligned}\text{家賃} &= 33880.241 \\ &\quad -1125.26(10) \\ &\quad +2753.4294(25) \\ &\quad -651.6833(3) \\ &\quad +1384.2767(2) \\ &\quad +4091.1646(1) \\ &\quad +9386.826(0) \\ &\quad -5540.621(0) \\ &\quad +3692.6441(0)\end{aligned}$$

$$\text{家賃} = 96368(\text{円})$$

# T検定と回帰分析の関係

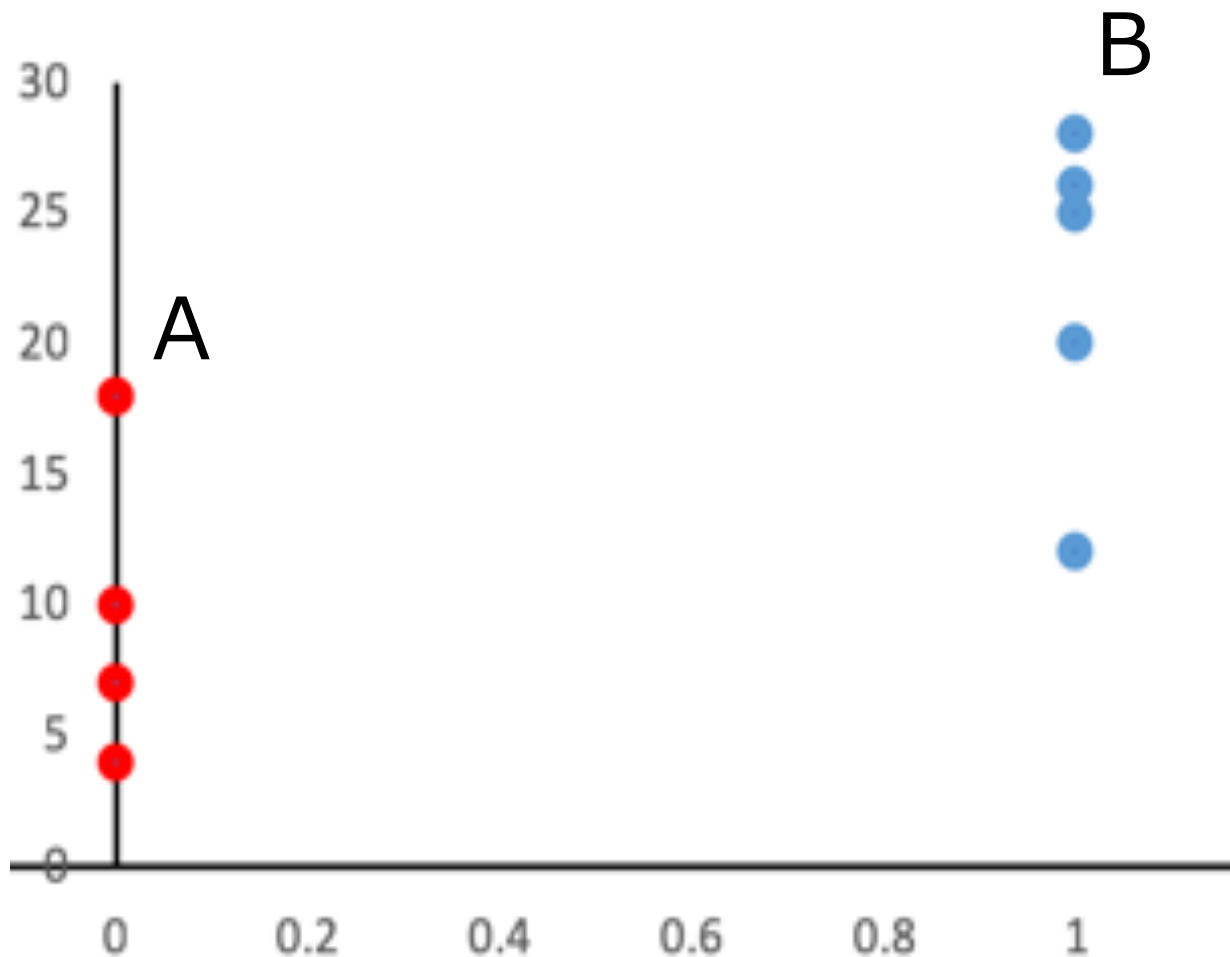
---

# T検定と回帰分析の関係？

A	B
18	12
4	26
18	28
7	25
10	20

# T検定と回帰分析の関係？

A	B
18	12
4	26
18	28
7	25
10	20

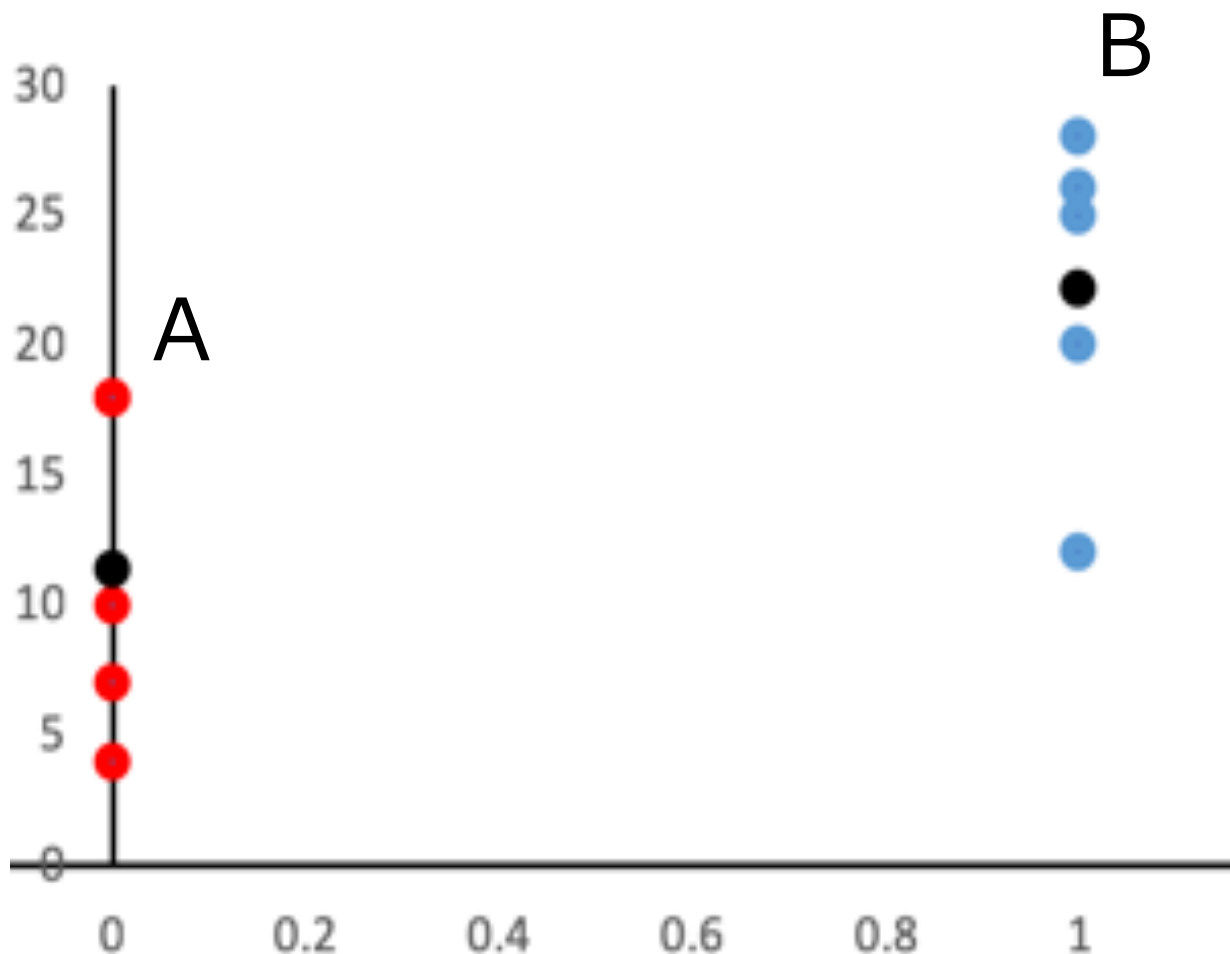


# T検定と回帰分析の関係？

A	B
18	12
4	26
18	28
7	25
10	20

平均値

11.4	22.2
------	------

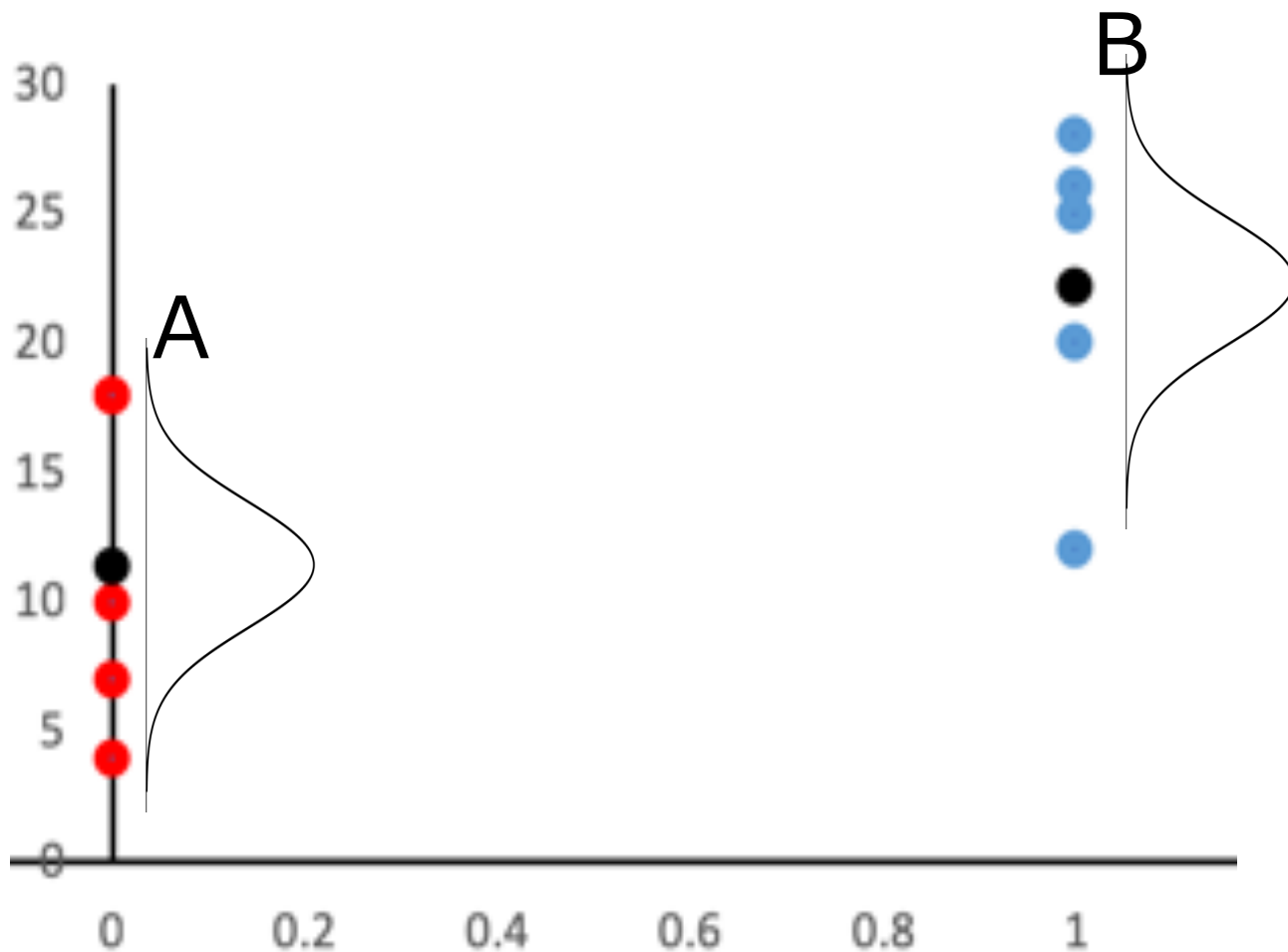


# T検定と回帰分析の関係？

A	B
18	12
4	26
18	28
7	25
10	20

平均値

11.4	22.2
------	------



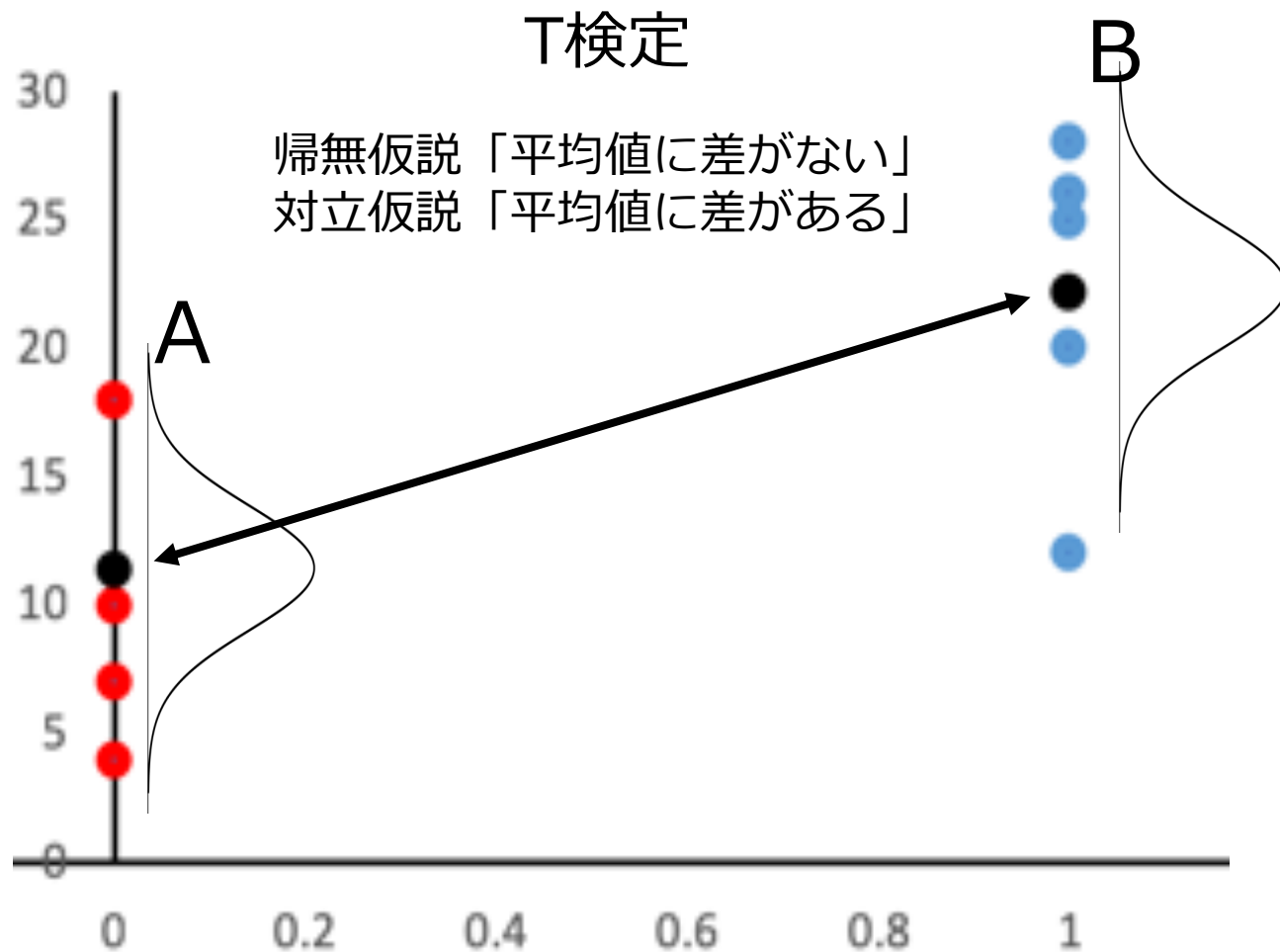


# T検定と回帰分析の関係？

A	B
18	12
4	26
18	28
7	25
10	20

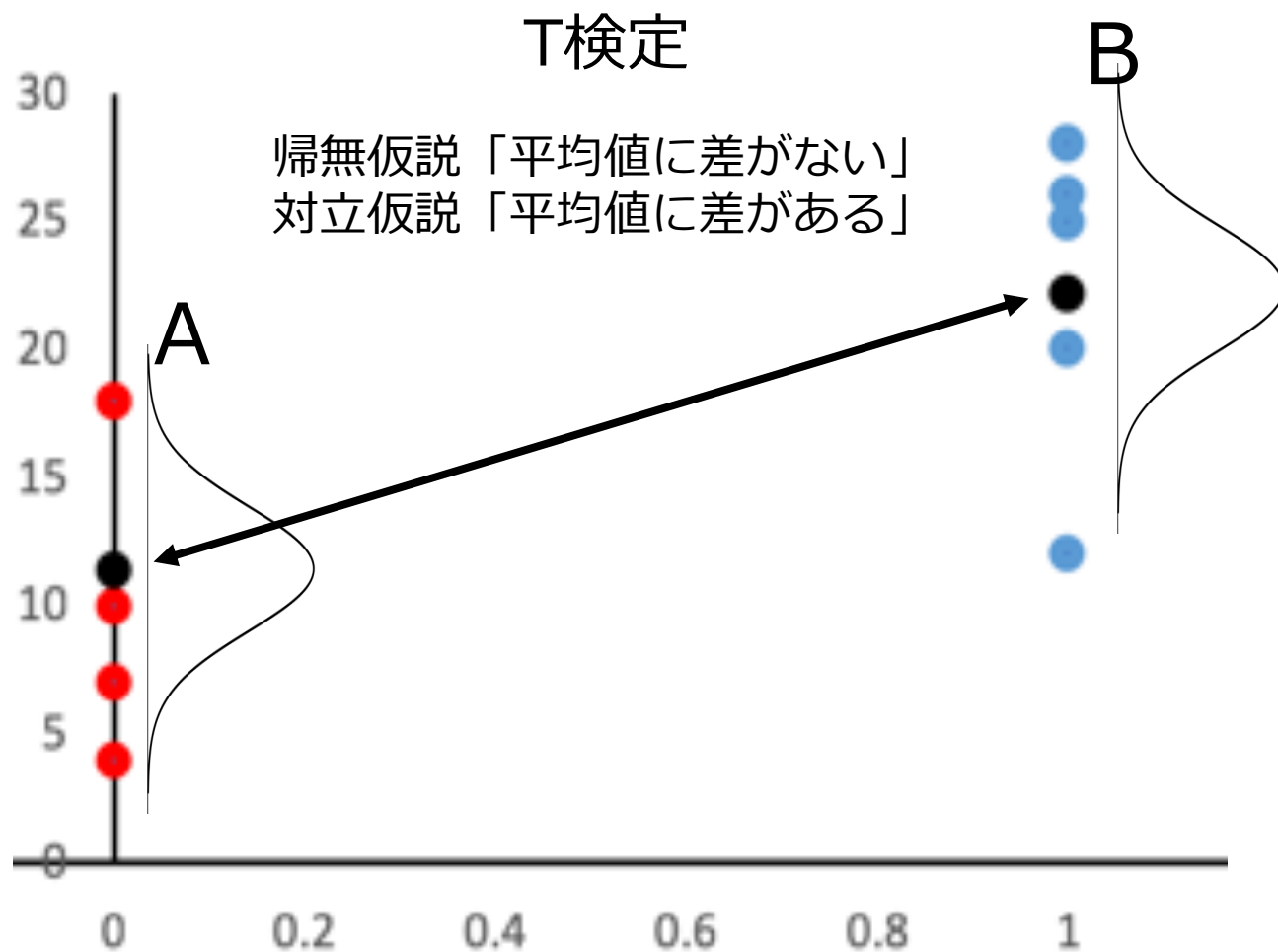
平均値

11.4	22.2
------	------



# T検定と回帰分析の関係？

t-検定: 分散が等しくないと仮定した2標本による検定		
	A	B
平均	11.400	22.200
分散	40.800	41.200
観測数	5.000	5.000
仮説平均との差異	0.000	
自由度	8.000	
t	-2.667	
P(T<=t) 片側	0.014	
t 境界値 片側	1.860	
P(T<=t) 両側	0.029	
t 境界値 両側	2.306	

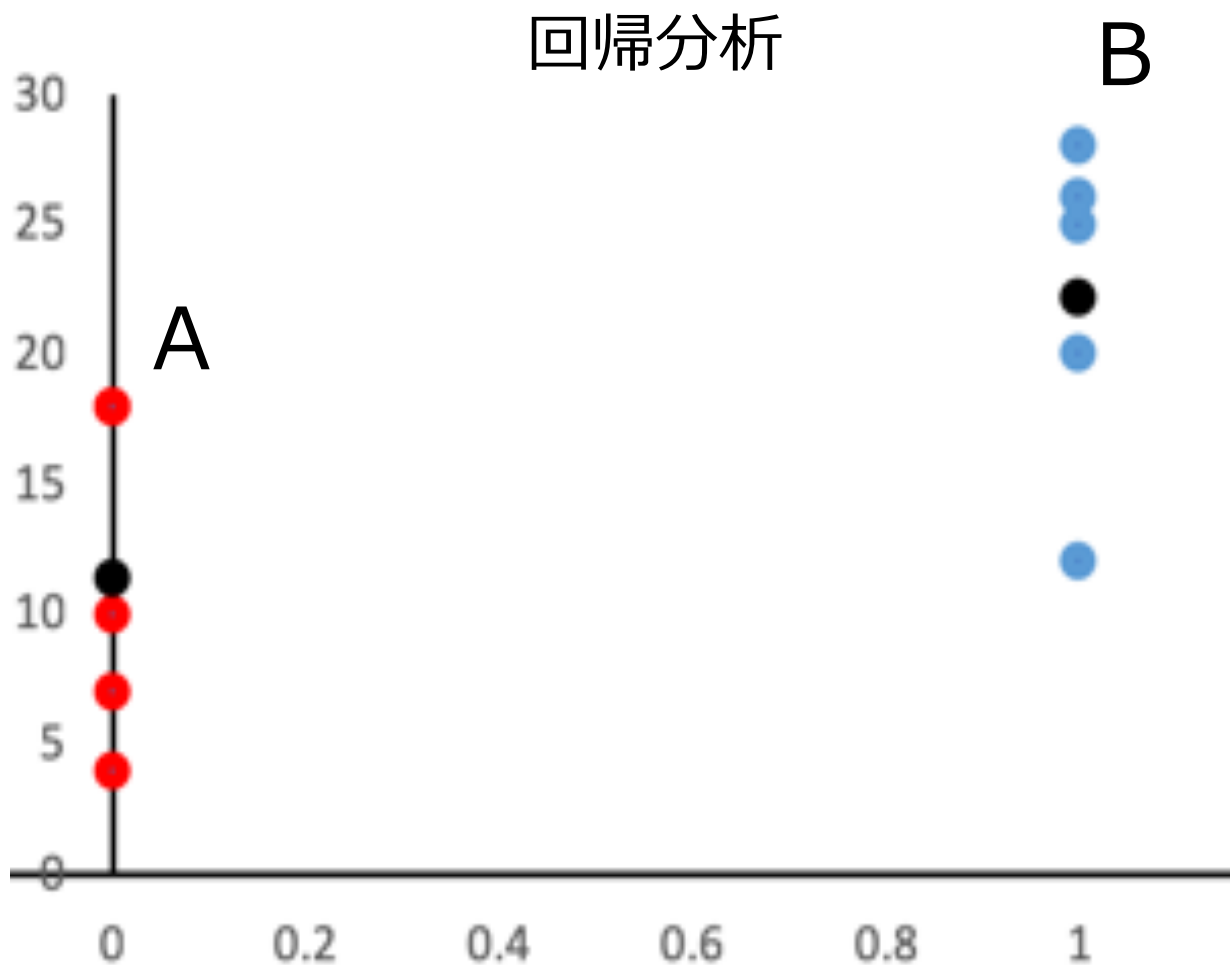


# T検定と回帰分析の関係？

A	B
18	12
4	26
18	28
7	25
10	20

平均値

11.4	22.2
------	------

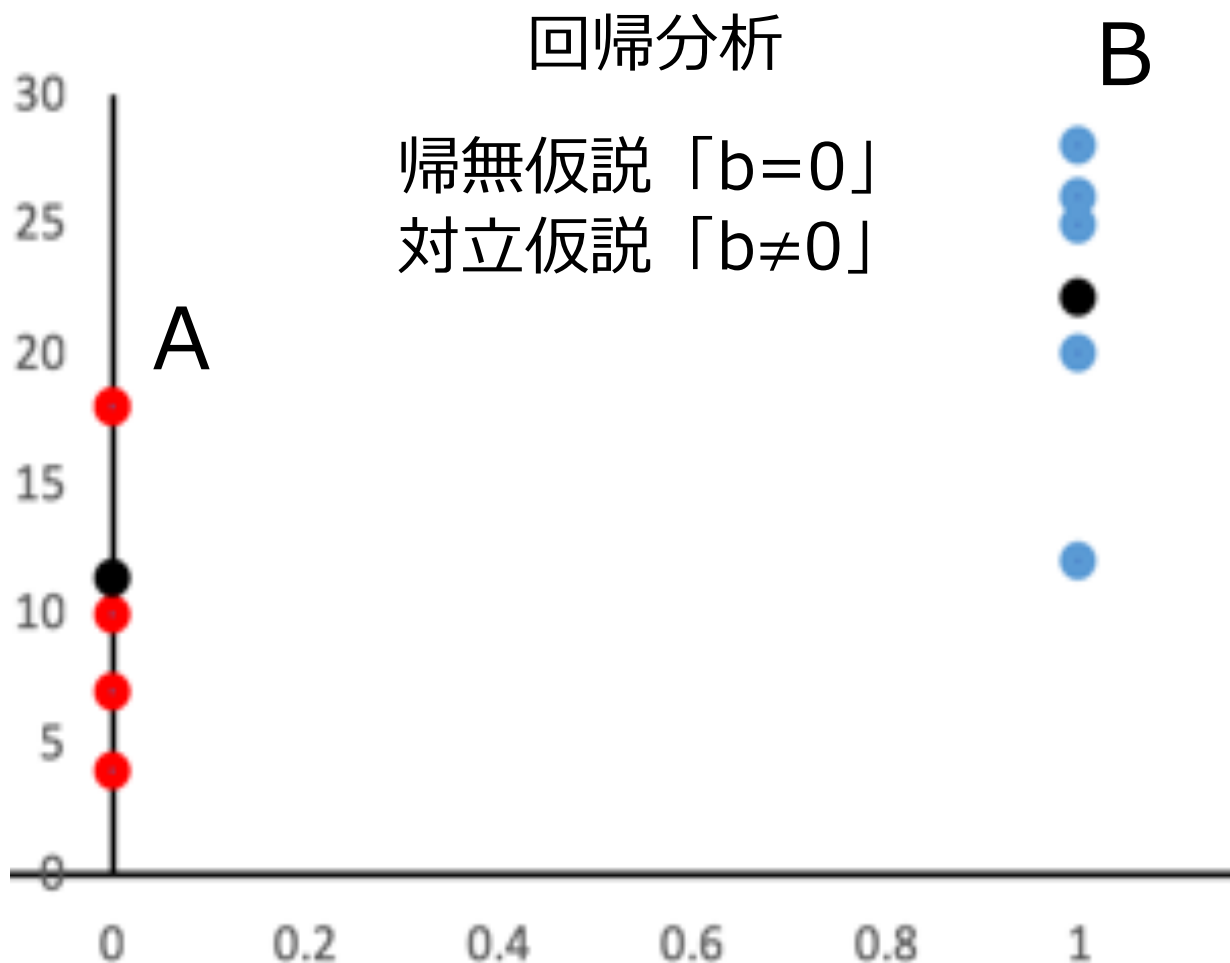


# T検定と回帰分析の関係？

A	B
18	12
4	26
18	28
7	25
10	20

平均値

11.4	22.2
------	------

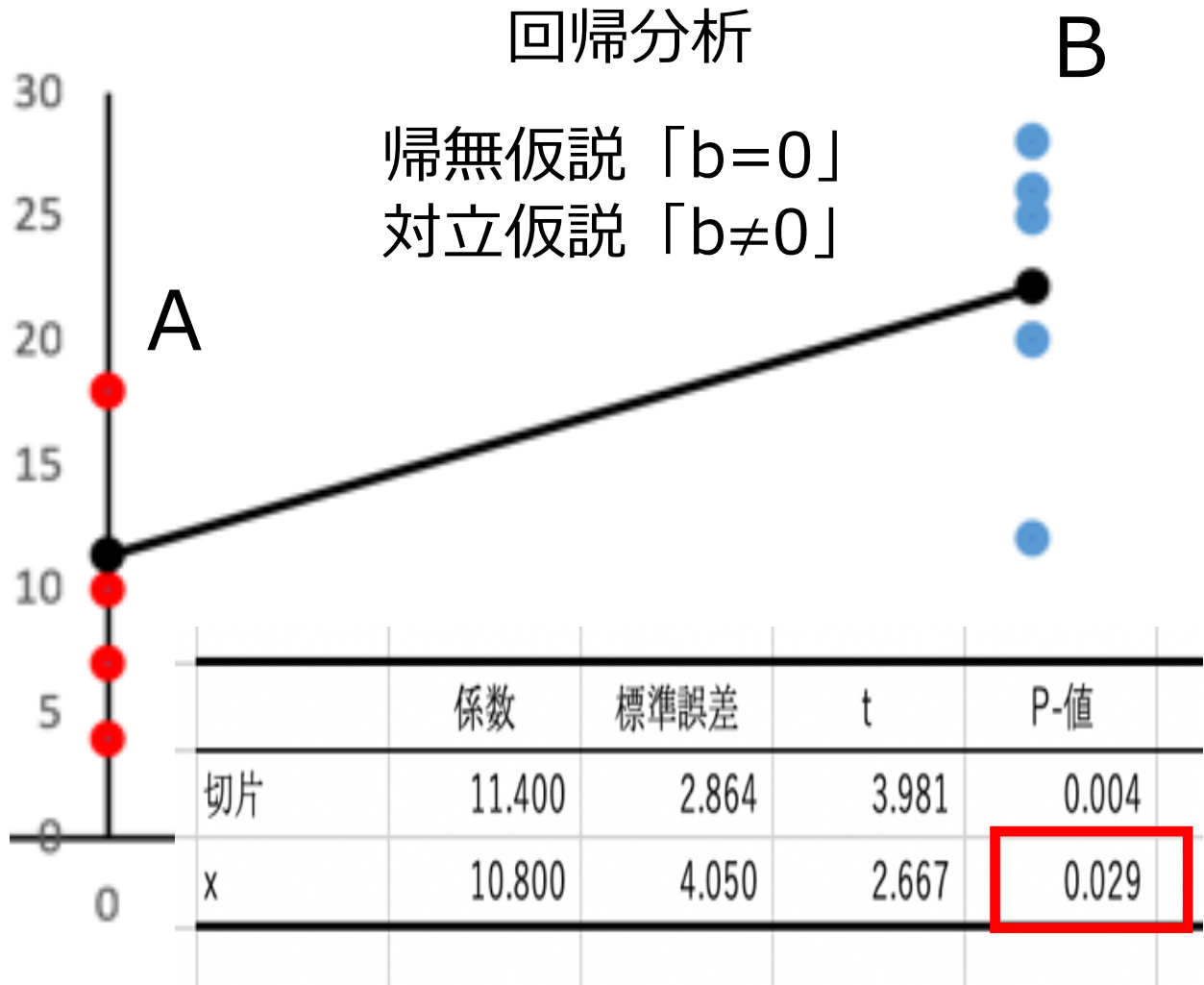


# T検定と回帰分析の関係？

A	B
18	12
4	26
18	28
7	25
10	20

平均値

11.4	22.2
------	------

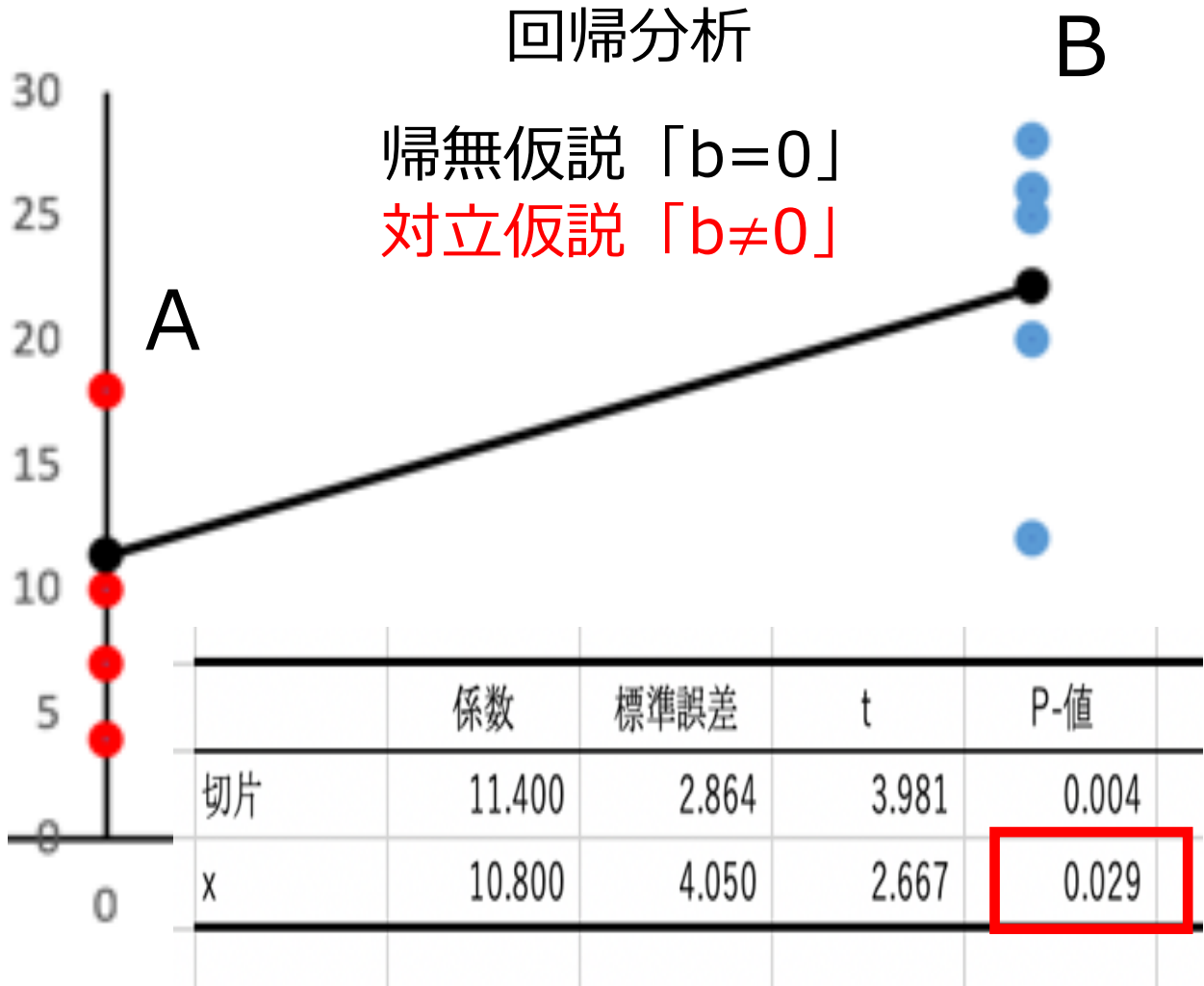


# T検定と回帰分析の関係？

A	B
18	12
4	26
18	28
7	25
10	20

平均値

11.4	22.2
------	------

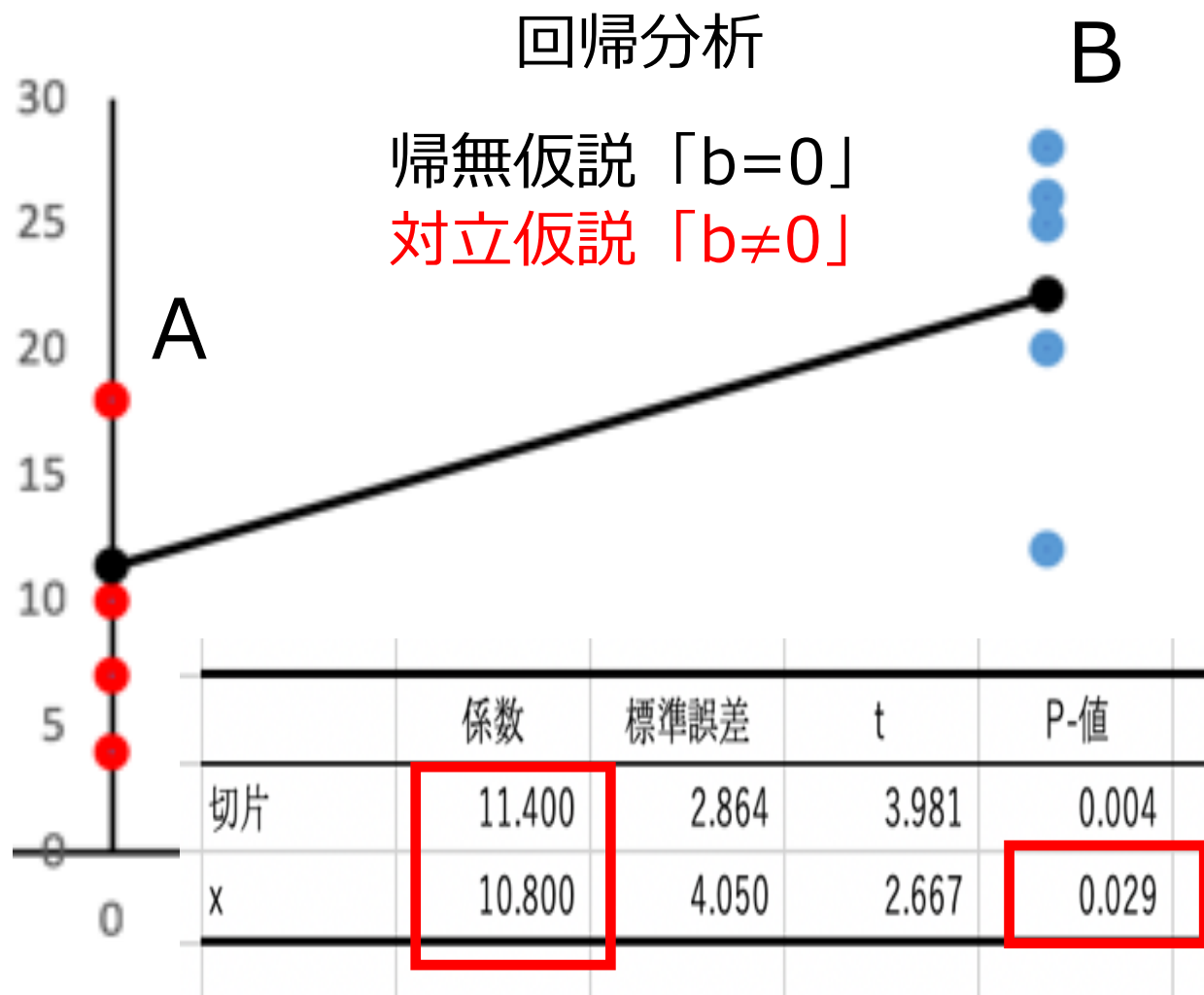


# T検定と回帰分析の関係？

A	B
18	12
4	26
18	28
7	25
10	20

平均値

11.4	22.2
------	------

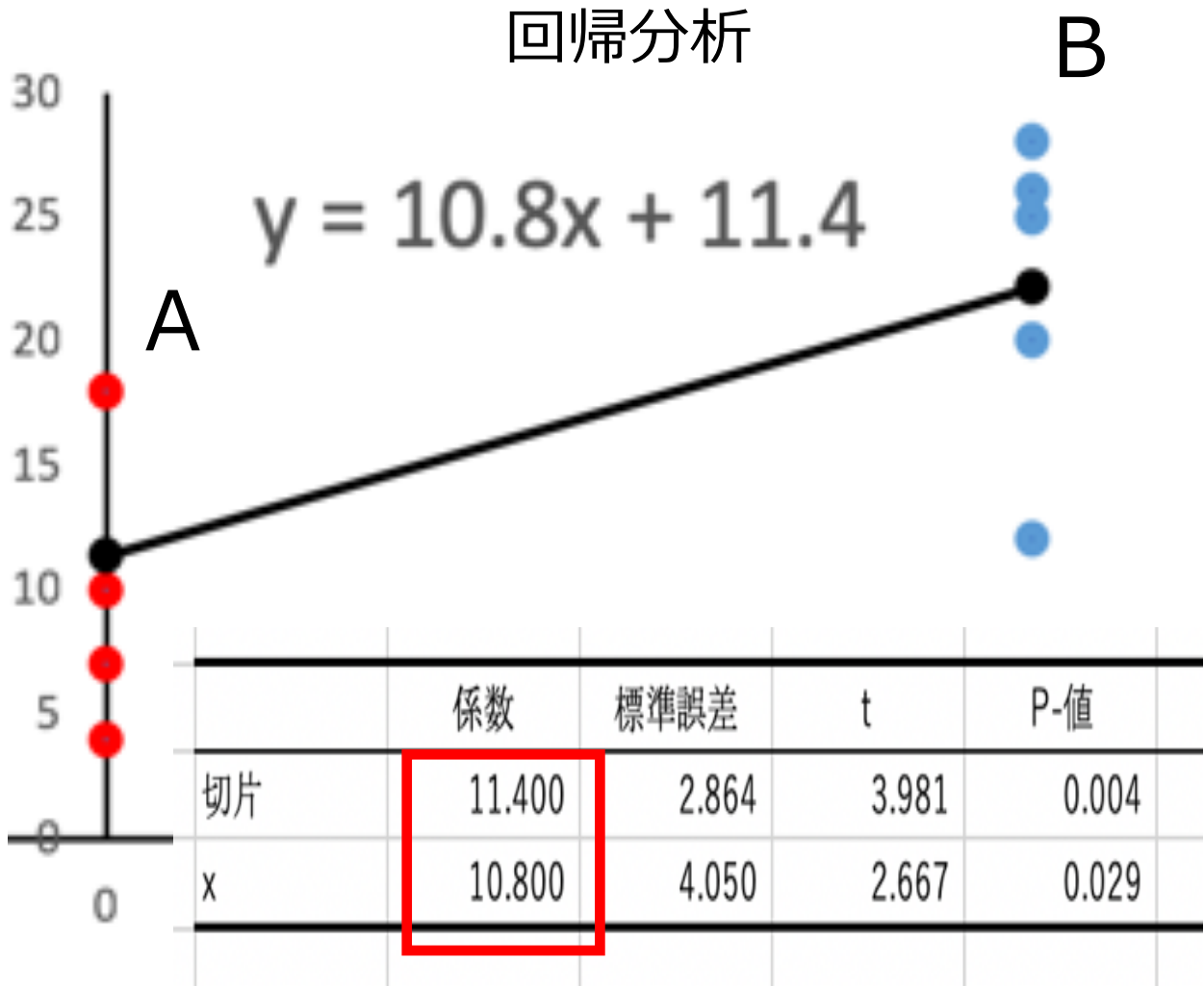


# T検定と回帰分析の関係？

A	B
18	12
4	26
18	28
7	25
10	20

平均値

11.4	22.2
------	------





# T検定と回帰分析の関係？

## T検定

帰無仮説「平均値に差がない」  
対立仮説「平均値に差がある」

t-検定: 分散が等しくないと仮定した2標本による検定		
	A	B
平均	11.400	22.200
分散	40.800	41.200
観測数	5.000	5.000
仮説平均との差異	0.000	
自由度	8.000	
t	-2.667	
P(T<=t) 片側	0.014	
t 境界値 片側	1.860	
P(T<=t) 両側	0.029	
t 境界値 両側	2.306	

## 回帰分析

帰無仮説「b=0」  
対立仮説「b≠0」

	係数	標準誤差	t	P-値
切片	11.400	2.864	3.981	0.004
x	10.800	4.050	2.667	0.029

# ダミー変数を使った分析事例

ある調査によると、男性と女性看護師の間で賃金格差が存在することが報告された。この調査報告を検証するため、看護師男性120人、女性120人を無作為に抽出し、経験年数、給与のデータが集計された。男女間に賃金格差があるかを検証せよ？

男性看護師	経験年数 (ヶ月)
\$70,124.00	46
\$66,992.00	13
\$71,922.00	60
\$68,601.00	21

⋮

\$68,601.00	21
-------------	----

女性看護師	経験年数 (ヶ月)
\$67,719.00	86
\$67,523.00	63
\$68,502.00	102
\$67,101.00	82

⋮

\$68,601.00	21
-------------	----

# ダミー変数を使った分析事例

男性看護師	経験年数 (ヶ月)
\$70,124.00	46
\$66,992.00	13
\$71,922.00	60
\$68,601.00	21

女性看護師	経験年数 (ヶ月)
\$67,719.00	86
\$67,523.00	63
\$68,502.00	102
\$67,101.00	82

給料	経験	性別
\$70,124.00	46	1
\$66,992.00	13	1
\$71,922.00	60	1
\$68,601.00	21	1
\$67,719.00	86	0
\$67,523.00	63	0
\$68,502.00	102	0
\$67,101.00	82	0

# 性別の効果を考慮しない場合

給料	経験
\$70,124.00	46
\$66,992.00	13
\$71,922.00	60
\$68,601.00	21
\$67,719.00	86
\$67,523.00	63
\$68,502.00	102
\$67,101.00	82

概要

回帰統計	
重相関 R	0.6881586
重決定 R2	0.4735621
補正 R2	0.4713501
標準誤差	2564.2435
観測数	240

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	1	1407750033	1407750033	214.0952503	5.17383E-35
残差	238	1564931999	6575344.53		
合計	239	2972682032			

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	63991.471	360.825063	177.347634	8.2686E-255	63280.65263	64702.29
経験日数 (月)	73.002391	4.98923051	14.6319941	5.17383E-35	63.17369921	82.831083

# 性別の効果を考慮した場合

給料	経験	性別
\$70,124.00	46	1
\$66,992.00	13	1
\$71,922.00	60	1
\$68,601.00	21	1
\$67,719.00	86	0
\$67,523.00	63	0
\$68,502.00	102	0
\$67,101.00	82	0

概要

回帰統計	
重相関 R	0.9695049
重決定 R2	0.9399398
補正 R2	0.9394329
標準誤差	867.9477
観測数	240

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	2	2794142062	1397071031	1854.519376	1.8305E-145
残差	237	178539970	753333.208		
合計	239	2972682032			

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	60978.807	140.88328	432.832104	0	60701.26315	61256.35
経験日数 (月)	82.183172	1.70226582	48.2786947	1.2852E-124	78.82966743	85.536677
性別	4845.3696	112.947713	42.8992273	1.0547E-113	4622.859919	5067.8793

# 回帰係数の解釈

	係数
切片	60978.807
経験日数（月）	82.183172
性別	4845.3696

$$y = 60978.807 \\ + 82.183172(\text{経験月数}) \\ + 4845.3696(\text{性別})$$



男性と女性では同じ経験月数でも給料に4845.3696ドルの差が存在する

# ダミー変数を使って多数カテゴリーを表現する

- 初任給と成績、卒業学部の間に関係があるかを調べるために次のようなデータを収集しました。

工学部の卒業生

初任給	成績
\$52,757.00	3.91
\$49,358.00	3.3
\$49,262.00	3.8

経済学部の卒業生

初任給	成績
\$50,479.00	3.04
\$48,049.00	3.1
\$48,048.00	3.71

芸術学部の卒業生

初任給	成績
\$45,874.00	2.69
\$45,846.00	2.62
\$45,895.00	3.51

農学部の卒業生

初任給	成績
\$46,437.00	3.04
\$46,275.00	3.1
\$46,770.00	3.71

# ダミー変数に関するルール

- ・ (ダミー変数の列数) = 条件数 - 1

男性と女性  $\Longrightarrow$  条件の数 = 2  $\Longrightarrow$  ダミー変数の列数 = 1

晴れ、曇り、雨  $\Longrightarrow$  条件の数 = 3  $\Longrightarrow$  ダミー変数の列数 = 2

工学・経済  
芸術・農学部  $\Longrightarrow$  条件の数 = 4  $\Longrightarrow$  ダミー変数の列数 = 3



# ダミー変数

初任給	成績	工学部	経済学部	芸術学部
\$ 52,757.00	3.91	1	0	0
\$ 49,358.00	3.3	1	0	0
\$ 49,262.00	3.8	1	0	0
\$ 50,479.00	3.04	0	1	0
\$ 48,049.00	3.1	0	1	0
\$ 48,048.00	3.71	0	1	0
\$ 45,874.00	2.69	0	0	1
\$ 45,846.00	2.62	0	0	1
\$ 45,895.00	3.51	0	0	1
\$ 46,437.00	2.44	0	0	0
\$ 46,275.00	3.59	0	0	0
\$ 46,770.00	2.72	0	0	0

工学部 (1、0、0)

経済学部 (0、1、0)

芸術学部 (0、0、1)

農学部 (0、0、0)

# 重回帰モデルの結果

概要

回帰統計	
重相関 R	0.911821
重決定 R2	0.831417
補正 R2	0.824319
標準誤差	1145.058
観測数	100

分散分析表

	自由度	変動	分散	割られた分散	有意 F
回帰	4	6.14E+08	1.5E+08	117.13	7.6E-36
残差	95	1.25E+08	1311158		
合計	99	7.39E+08			

	係数	標準誤差	t	P-値	下限 95%	上限 95%
切片	43244.91	733.5203	58.9553	1.3E-76	41788.69	44701.13
成績	1142.315	239.8636	4.76235	6.9E-06	666.1254	1618.5045
工学部	2830.257	398.3591	7.10479	2.2E-10	2039.414	3621.0997
経済学部	1304.71	411.1126	3.17361	0.00203	488.5482	2120.8718
芸術学部	-2321.8	344.4478	-6.7406	1.2E-09	-3005.62	-1637.986

# 重回帰モデルの設計

	係数
切片	43244.90794
成績	1142.314932
工学部	2830.256883
経済学部	1304.709995
芸術学部	-2321.800796

$$\begin{aligned}\text{初任給} &= 43244 \\ &+ 1142(\text{成績}) \\ &+ 2830(\text{工学部}) \\ &+ 1304(\text{経済学部}) \\ &- 2321(\text{芸術学部})\end{aligned}$$



同じ成績でも工学部を卒業すると農学部を卒業する学生よりも給料は2830ドルいい。

# 参考資料

---

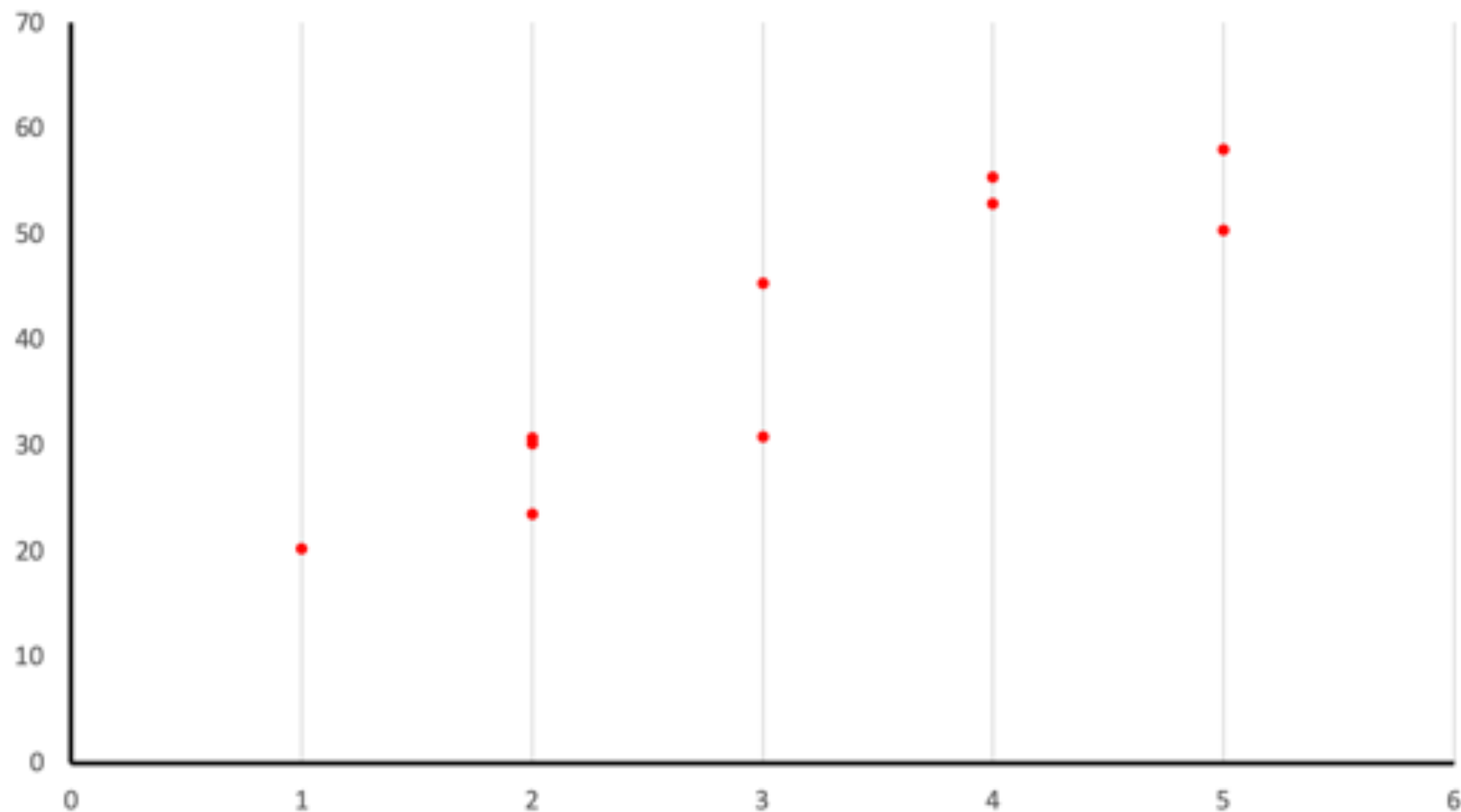
# 回帰分析の仮定

## 正規分布の仮定

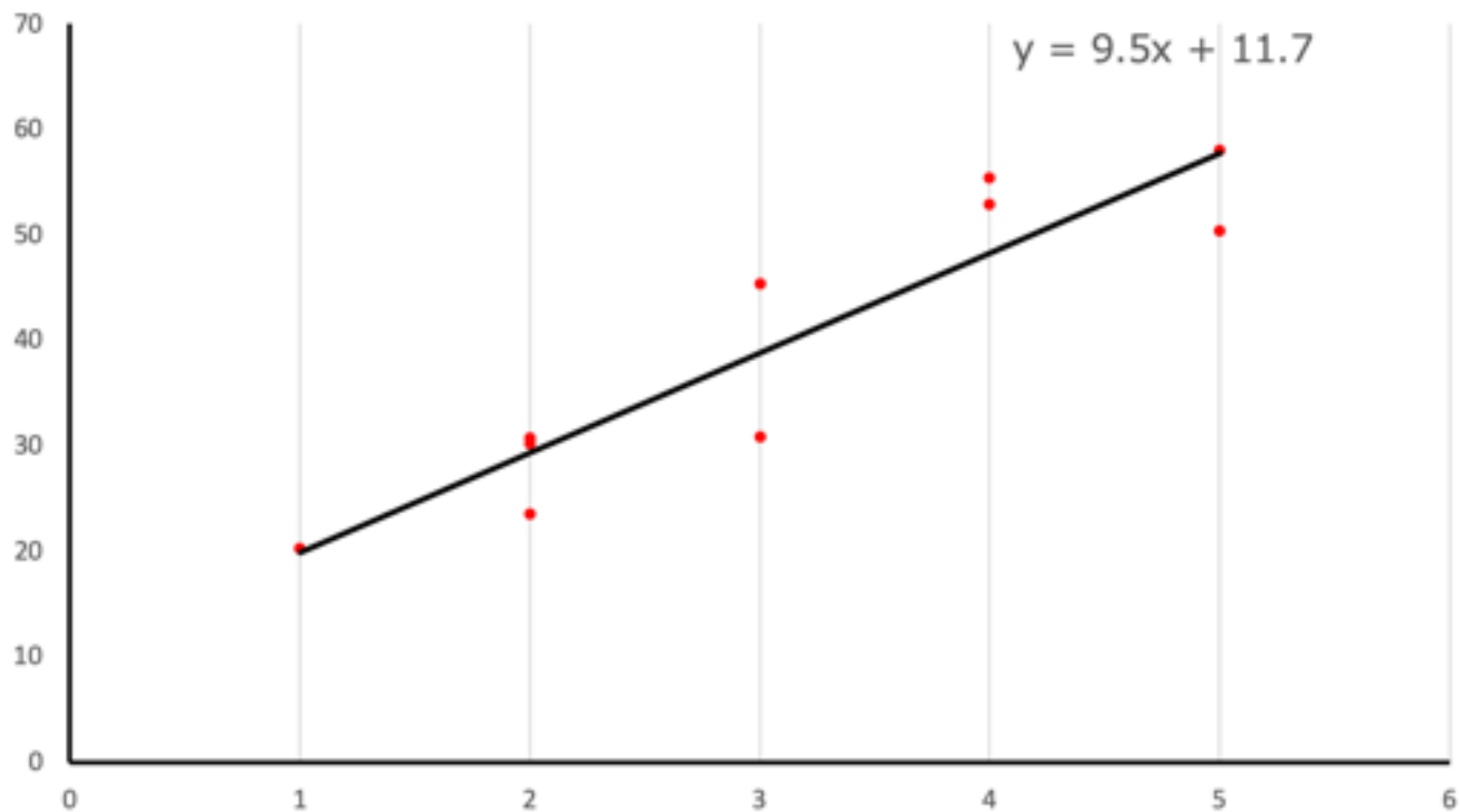
誤差項の確率分布は正規分布に従い、誤差項は互いに独立で同一の正規分布 $N(0, \sigma)$ に従う

$$e_i \sim N(0, \sigma)$$

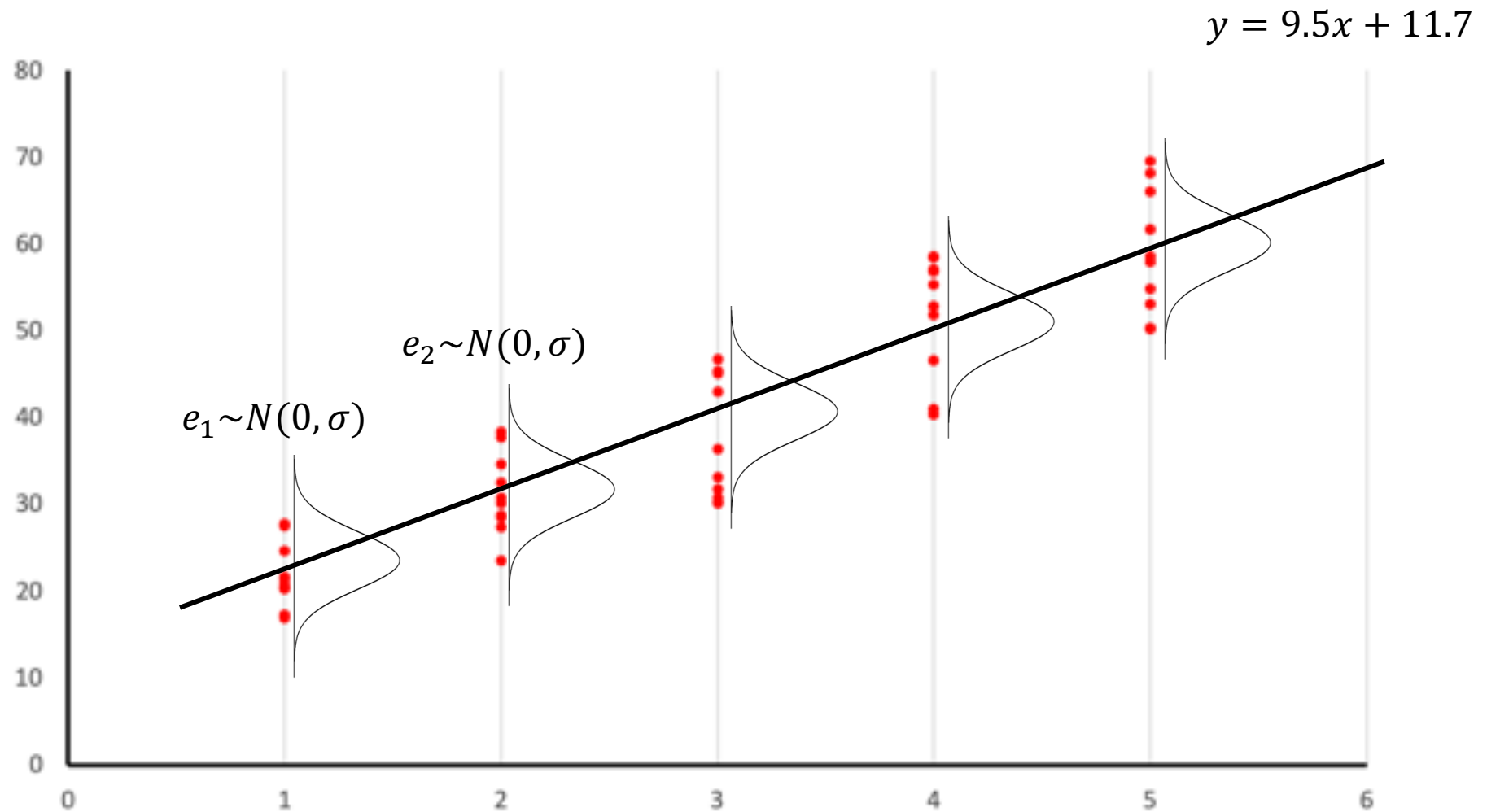
# 残差の満たすべき条件 $e_i \sim N(0, \sigma)$



# 残差の満たすべき条件



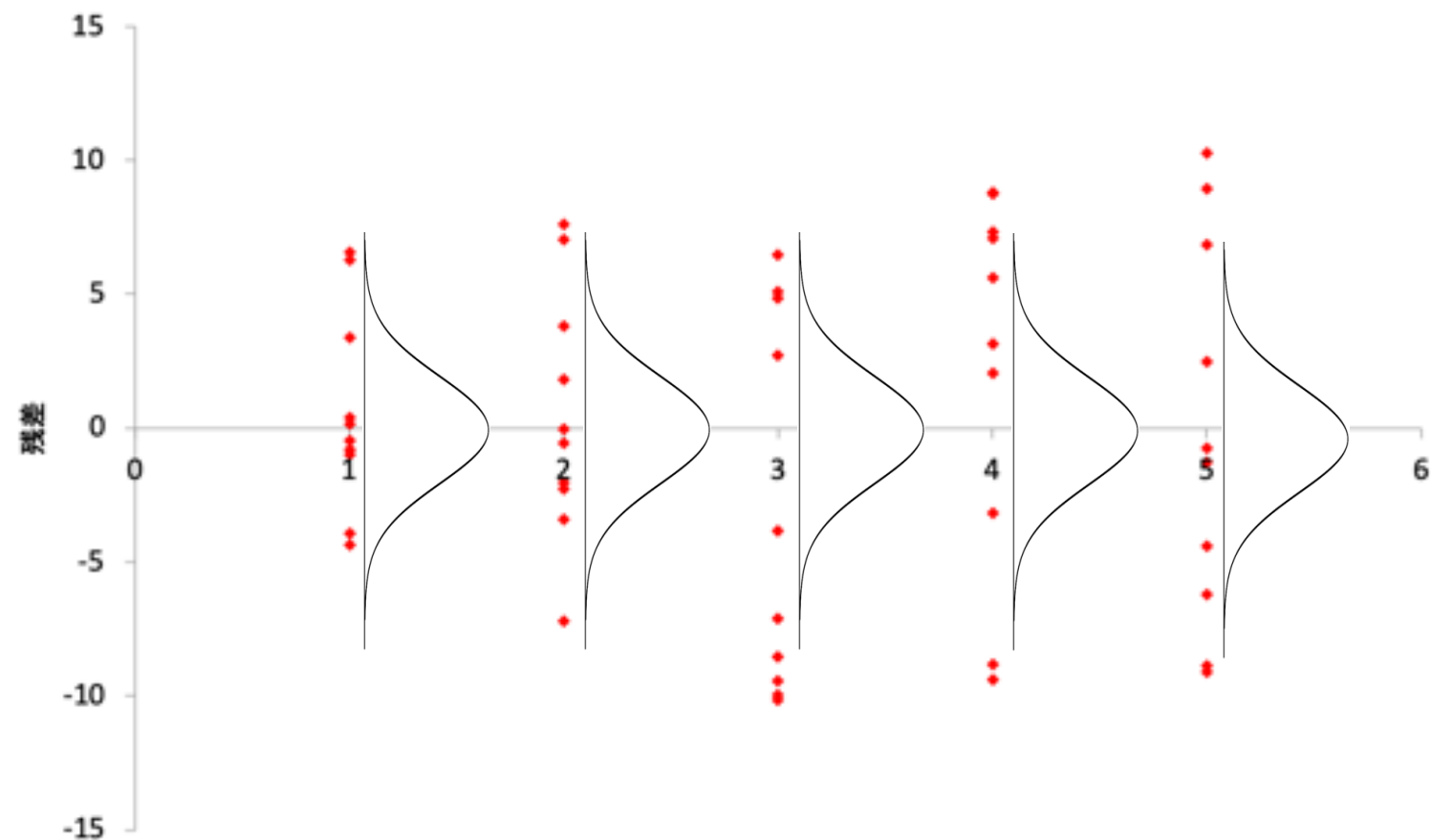
# 残差の満たすべき条件





# 残差プロット

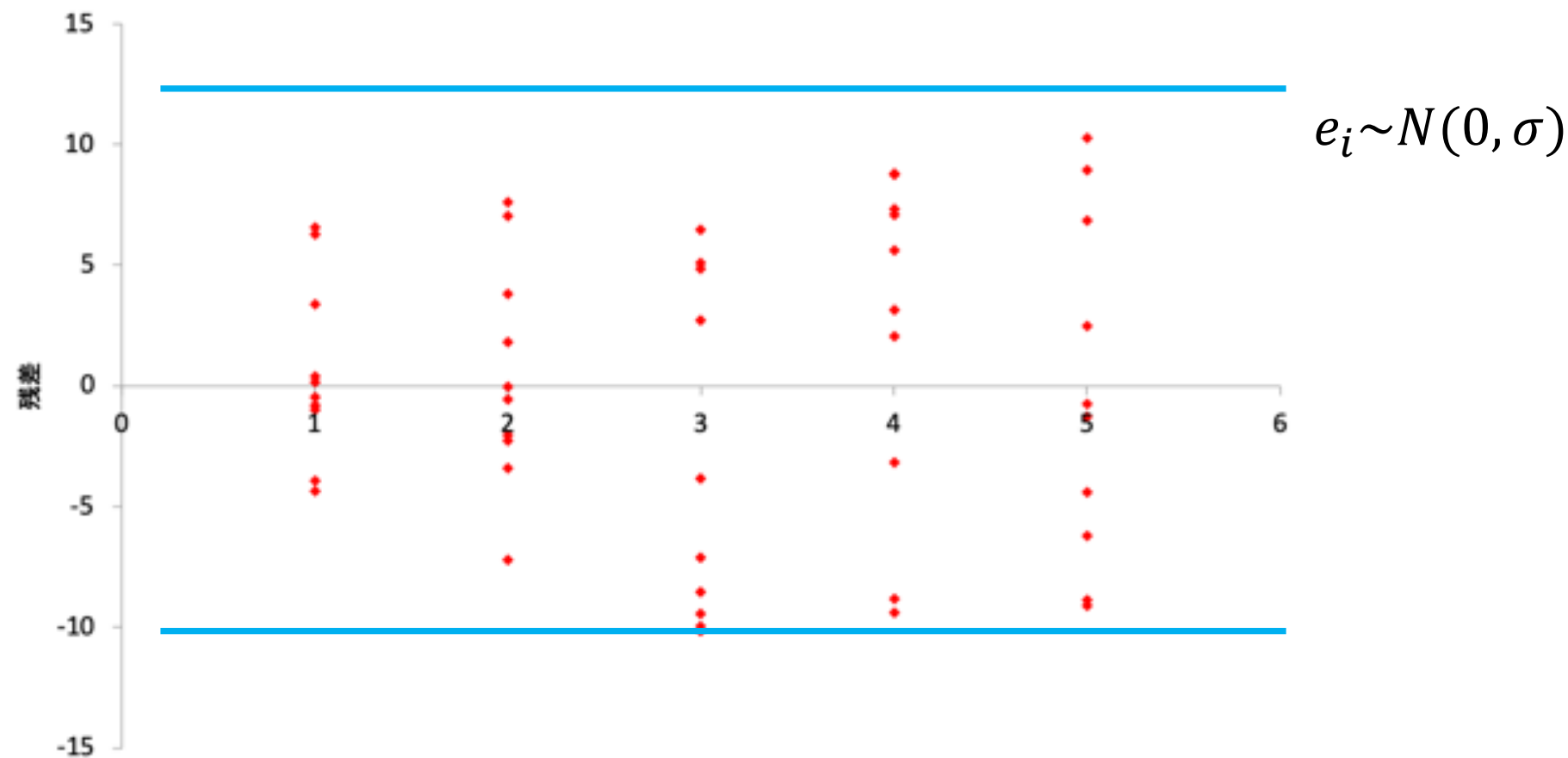
残差が縦軸「0」に対して均一に分散している場合、問題はない



$$e_i \sim N(0, \sigma)$$

# 残差プロット

残差が縦軸「0」に対して均一に分散している場合、問題はない



# 残差プロット

残差が徐々に大きくあるいは小さくなっていたり、何かしらの傾向が見られる場合は要注意。この場合、推定された回帰式が妥当ではない可能性が<sup>高い</sup>。

