

MDBセミナー

統計分析力養成講座

データサイエンス研究所
伊藤嘉朗

本日の内容

- 統計的意思決定法

- 統計的思考法

 - 平均、データ群の違いの大きさ
データ群の関係

- グラフの活用法

統計的意思決定法

◇あるイベント

雨が降らなければ2000万円の純益
もし雨が降れば200万円の純益

主催者は、費用500万円をかけて、雨天の場合の
損失補償（2000万円）の保険を契約すべきか？

但し、過去の記録から雨天の確率は20%

期待値を求める

①利得表を作成

万円

天気	行動	
	保険を契約	保険を契約せず
晴・曇	1500	2000
雨	1700	200

注) $1500 = 2000 - 500$
 $1700 = 2000 - 500 + 200$

②期待利得表を作成

$$\text{期待利得} = \text{利得} \times \text{確率}$$

万円

天気	確率	保険契約		保険契約せず	
		利得	期待利得	利得	期待利得
晴・曇	0.8	1500	1200	2000	1,600
雨	0.2	1700	340	200	40
計	1.0		1540		1,640

$$\begin{aligned} 1200 &= 1500 \times 0.8 & 1600 &= 2000 \times 0.8 \\ 340 &= 1700 \times 0.2 & 40 &= 200 \times 0.2 \end{aligned}$$

保険契約したときの期待利得：1540万円

保険を契約しないときの期待利得：1640万円

保険を契約しない方が100万円高い利得が期待できる。

◇あるイベント

雨が降らなければ2000万円の純益

もし雨が降れば200万円の純益

主催者は、費用500万円をかけて、雨天の場合の損失補償（2000万円）の保険を契約をすべきか。

（過去の記録から雨天の確率は20%）

気象専門会社の天気予想サービスを、費用20万円をかけても利用すべきか？

（この会社の的中率は過去の記録によれば70%）

予報に従う（予報が雨のときのみ保険を契約） 行動が増える

期待利得表を作成

万円

事象	予報	確率	保険契約		契約せず		予報に従う
			利得	期待利得	利得	期待利得	期待利得
晴・曇	晴・曇	$0.8 \times 0.7 = 0.56$	1500	840	2000	1120	1120
晴・曇	雨	$0.8 \times 0.3 = 0.24$	1500	360	2000	480	360
雨	晴・曇	$0.2 \times 0.3 = 0.06$	1700	102	200	12	12
雨	雨	$0.2 \times 0.7 = 0.14$	1700	238	200	28	238
計		1.00		1540		1640	1730

$$1500 \times 0.56 + 1500 \times 0.24 + 1700 \times 0.06 + 1700 \times 0.14 = 1540$$

$$2000 \times 0.56 + 2000 \times 0.24 + 200 \times 0.06 + 200 \times 0.14 = 1640$$

予報に従う場合の期待利得 : $1730 - 1640 = 90$ 万円増加
 気象予報サービスの費用 : 20万円 < 90万円
 ⇒ 気象予報サービスを利用

◇あるスポーツの試合

アイスクリーム、ホットドッグのいずれの場内販売権を選択すべきか。

当日の天候の確率：晴れ0.6、曇り0.3、雨0.1

利得表		万円
天気	場内販売	
	アイスクリーム	ホットドッグ
晴	300	200
曇り	100	200
雨	-50	100

期待利得表

万円

天気	確率	アイスクリーム		ホットドッグ	
		利得	期待利得	利得	期待利得
晴れ	0.6	300	180	200	120
曇り	0.3	100	30	200	60
雨	0.1	-50	-5	100	10
計	1.0		205		190

アイスクリームの期待利得 : 205万円

ホットドッグの期待利得 : 190万円

⇒ アイスクリームの販売権を選択

統計的意思決定法のまとめ

- 1) 事象・行動を整理する。
- 2) 各行動における利得表を作成する。
- 3) 事象別確率を求める。
- 4) 各行動における期待利得表を作成する。
- 5) 最も期待利得の高い行動を選択する。

データ分析とは

- 1) 代表する値を求めて検討する。
売上高の平均、参加者平均、...
- 2) 違いについて判断する。
男女の好感度、CM前後の知名度、
東京と大阪の売上高...
- 3) 関係を明らかにする。
DM発送数と売上高、広告費と売上高...

データの代表値

(1) 1,2,3,4,5 の平均は？

$(1+2+3+4+5) \div 5 = 3$ ➡ 算術平均（相加平均）

AVERAGE(算術平均)

B11		$\sum$	=AVERAGE(B3:B9)		
	A	B	C	D	
1		単位: 億円			
2	支店	売上高			
3	北海道	121			
4	東北	213			
5	関東	1,452			
6	中部	417			
7	関西	912			
8	九州	332			
9	沖縄	28			
10	合計	3,475			
11	平均	496			
12					
13					
14					

比率の平均

◇ある会社の年度別売上高。平均伸び率は？

(億円)		
年度	売上高	
2018	100	2018～2019 20% ↑ (1.2倍)
2019	120	2019～2020 30% ↑ (1.3倍)
2020	156	

$$(20\% + 30\%) / 2 = 25\% ?$$

$$100 \times 1.25 \times 1.25 = 156.25$$

比率の平均に算術平均は使用できない

幾何平均（相乗平均）

1.2と1.3の幾何平均は、

$$\sqrt{1.2 \times 1.3} = 1.2489\dots$$

$$100 \times 1.2489\dots \times 1.2489\dots = 156$$

$$\text{2つの幾何平均} \quad \sqrt{a \times b}$$

$$\text{4つの幾何平均} \quad \sqrt[4]{a \times b \times c \times d}$$

◇ある商品を価格300円で発売し、その後3回値上げをした。
値上げ率の平均は？

	価格
発売価格	300
1回目	350
2回目	380
3回目	420

	値上げ率
1回目	1.167
2回目	1.086
3回目	1.105

$$\sqrt[3]{1.167 \times 1.086 \times 1.105} = 1.1187$$

$$\sqrt[3]{\frac{350}{300} \times \frac{380}{350} \times \frac{420}{380}} = \sqrt[3]{\frac{420}{300}} = \sqrt[3]{1.4} = 1.1187$$

GEOMEAN (幾何平均)

B5		fx =GEOMEAN(B3:B4)		
	A	B	C	D
1				
2		伸び率		
3	2014/2013	1.2		
4	2015/2014	1.3		
5	幾何平均	1.2489996		
6				
7				
8				

B6		fx =GEOMEAN(B3:B5)		
	A	B	C	D
1				
2		値上げ率		
3	1回目	1.167		
4	2回目	1.086		
5	3回目	1.105		
6	幾何平均	1.119		
7				
8				

n乗根 : $^{\wedge} (1/n)$

fx		=1.56^(1/2)		
	C	D	E	
		1.2489996		

fx		=1.4^(1/3)		
	C	D	E	
		1.11868894		

◇ある事業部の売上高推移。伸び率の平均は？

年度	売上高	対前年比率
2013	29,306	
2014	31,298	1.068
2015	34,536	1.103
2016	33,971	0.984
2017	35,829	1.055
2018	36,747	1.026
2019	36,886	1.004
2020	38,925	1.055

C11					$= (B9/B2)^{(1/7)}$
	A	B	C	D	
1		売上高	対前年比率		
2	2013	29,306			
3	2014	31,298	1.068		
4	2015	34,536	1.103		
5	2016	33,971	0.984		
6	2017	35,829	1.055		
7	2018	36,747	1.026		
8	2019	36,886	1.004		
9	2020	38,925	1.055		
10	算術平均		1.042		
11	幾何平均		1.041		

算術平均 = 1.042
幾何平均 = 1.041

◇2013年度の売上高を変更

C11			$\times$	$\checkmark$	f_x	$= (B9/B2)^{(1/7)}$
	A	B	C	D		
1		売上高	対前年比率			
2	2013	12,345				
3	2014	31,298	2.535			
4	2015	34,536	1.103			
5	2016	33,971	0.984			
6	2017	35,829	1.055			
7	2018	36,747	1.026			
8	2019	36,886	1.004			
9	2020	38,925	1.055			
10	算術平均		1.252			
11	幾何平均		1.178			
12						

算術平均 = 1.252
幾何平均 = 1.178

広辞苑 ～平均～

多くの量または数の中間的な値。また、それを求める演算。中間の意味のとり方によって、相加平均（算術平均）、相乗平均（幾何平均）その他がある。ふつう、相加平均（算術平均）をさす。

幾何平均は比率の合理的な平均値であり、比率、指数等の平均においては、算術平均より優れる。

統計的思考

英語の研修前後の成績（20人）

	A君の得点	全体の平均点	得点-平均点
研修前	70	58.3	+11.7
研修後	72	58.3	+13.7

A君の成績の評価は？

平均を評価基準とする合理性？

平均の信頼度

(例) 暑い日に暑い場所で待ち合わせをした。
いつも遅れてくる人が何分後に来るのか予測。

<過去10回の遅刻データ（分）>										平均	
①	21	46	8	28	19	34	13	33	19	31	25.2分
②	25	26	23	24	26	27	26	25	24	26	25.2分

①、②それぞれにおける待ち時間の行動は？

①のデータは「バラツキ」が大きい。

「バラツキ」の大きいデータの平均は信頼できない。

全員の成績

<研修前>

70、56、89、27、69、57、69

50、33、67、37、49、98、69

68、25、65、67、33、68

<研修後>

72、31、95、36、89、88、89

76、28、47、23、28、96、48

51、20、33、91、27、98

全員の成績

<研修前>

70、56、89、27、69、57、69

50、33、67、37、49、98、69

68、25、65、67、33、68

<研修後>

72、31、95、36、89、88、89

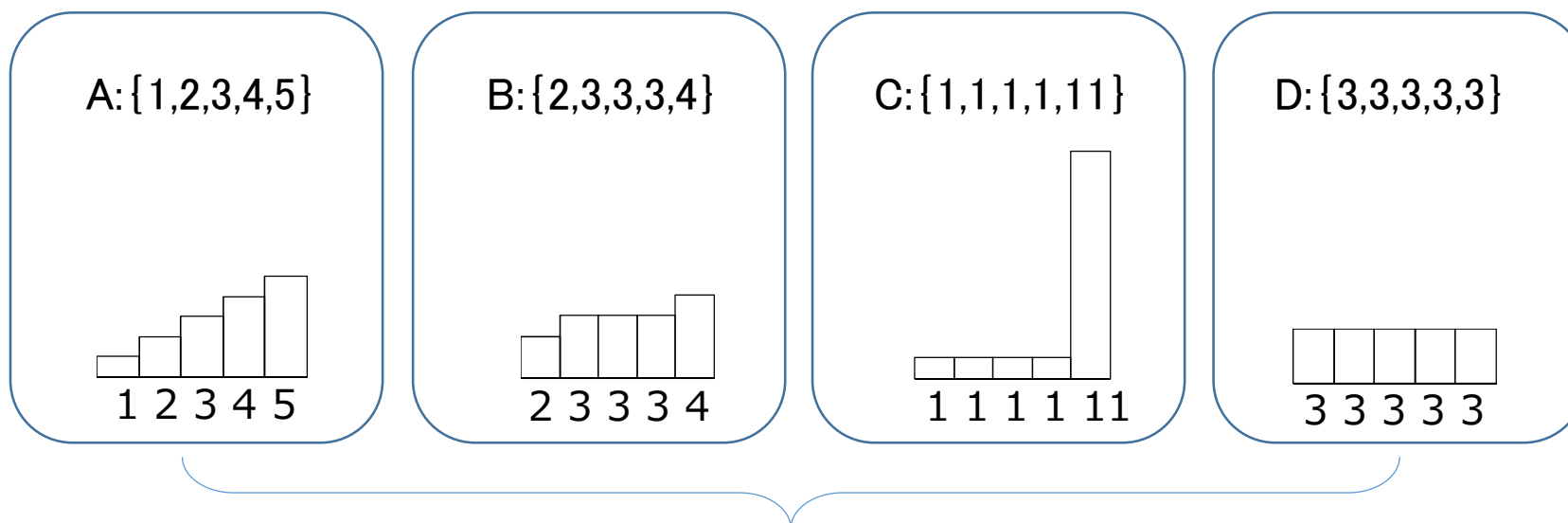
76、28、47、23、28、96、48

51、20、33、91、27、98

研修後はバラツキが大きくなった。

平均値が同じデータ群

A群～D群の平均値はすべて同じ値3



平均値は同じであるので同類集団？

各群のバラツキが異なる。 $D < B < A < C$

データのバラツキを表す代表値が必要！

バラツキの計算方法の検討

A群のバラツキの部分抽出する。

A群		平均値		値－平均値
1	－	3	=	-2
2	－	3	=	-1
3	－	3	=	0
4	－	3	=	1
5	－	3	=	2
計				= 0



合計すると0！



バラツキの部分を2乗し、合計！

A群のバラツキの部分を2乗して合計

	値－平均値	(値－平均値) ²	
	-2	4	
	-1	1	
	0	0	
	1	1	
	2	4	
計	0	10	A群のバラツキ=10

同様に、B、C、D群について計算する。

B群	平均	差	(差) ²	C群	平均	差	(差) ²	D群	平均	差	(差) ²
2	3	= -1	1	1	3	= -2	4	3	3	= 0	0
3	3	= 0	0	1	3	= -2	4	3	3	= 0	0
3	3	= 0	0	1	3	= -2	4	3	3	= 0	0
3	3	= 0	0	1	3	= -2	4	3	3	= 0	0
4	3	= 1	1	11	3	= 8	64	3	3	= 0	0
計		0	2	計		0	80	計		0	0

$$D < B < A < C \quad 0 < 2 < 10 < 80$$



「偏差平方和」

E群: $\{1,1,2,2,3,3,4,4,5,5\}$ の偏差平方和

E群	平均値	差	(差) ²
1	- 3	= -2	4
1	- 3	= -2	4
2	- 3	= -1	1
2	- 3	= -1	1
3	- 3	= 0	0
3	- 3	= 0	0
4	- 3	= 1	1
4	- 3	= 1	1
5	- 3	= 2	4
5	- 3	= 2	4
計			= 20

➡ E群の偏差平方和=20

E群: $\{1,1,2,2,3,3,4,4,5,5\}$ の偏差平方和 = 20

A群: $\{1,2,3,4,5\}$ の偏差平方和 = 10

A群とE群の構造は同じであるが、偏差平方和の値はE群の方が大きい。

偏差平方和をデータ数で割ると同じ代表値となる。

E群 : $20 \div 10 = 2$ A群 : $10 \div 5 = 2$



バラツキの代表値①
「分散」

F群: {10,20,30,40,50} A群: {1,2,3,4,5} の10倍

F群の単位: 千円 A群の単位: 万円の場合、全く同じデータ

偏差平方和を計算すると

F群	平均	差	(差) ²
10	30	-20	400千円 ²
20	30	-10	100
30	30	0	0
40	30	10	100
50	30	20	400
計		0	1000

F群の偏差平方和: 1000千円²

F群の分散: 200千円²

A群の分散: 2万円²

分散の比較は困難

分散の平方根は、同じ値となる。

A群: $\sqrt{2}$ $\div$ 1.414万円

F群: $\sqrt{200}$ $\div$ 14.14千円



バラツキの代表値②
「標準偏差」

A群 (1,2,3,4,5) とF群 (10,20,30,40,50) の比較

	平均	標準偏差
A群 (万円)	3	1.414
F群 (千円)	30	14.14

データ群を比較するときは、単位の影響を除くために
平均と標準偏差を組み合わせる

$$\frac{1.414}{3} = \frac{14.14}{30} \div 0.471 \text{ (約47.1\%)}$$

$$\frac{\text{標準偏差}}{\text{平均}} \rightarrow \text{変動係数 (CV)}$$

STDEV.P (標準偏差)

B24						
	A	B	C	D	E	F
1		研修前	研修後			
2		70	72			
3		56	31			
4		89	95			
5		27	36			
6		69	89			
7		57	88			
8		69	89			
9		50	76			
10		33	28			
11		67	47			
12		37	23			
13		49	28			
14		98	96			
15		69	48			
16		68	51			
17		25	20			
18		65	33			
19		67	91			
20		33	27			
21		68	98			
22	平均値	58.3	58.3			
23	分散	368.4	819.0			
24	標準偏差	19.2	28.6			
25						

A君の成績の評価

	成績	平均	成績－平均	標準偏差
研修前	70	58.3	11.7	19.2
研修後	72	58.3	13.7	28.6

研修前

研修後

$$\frac{70-58.3}{19.2} = 0.609 > \frac{72-58.3}{28.6} = 0.479$$

成績と平均値の差
標準偏差

⇒ Z値

Z 値の比較によりデータの評価・比較が可能

$$Z \text{ 値} \times 10 + 50$$



偏差値

$$\text{研修前の偏差値} = 0.609 \times 10 + 50 = 56.09$$

$$\text{研修後の偏差値} = 0.479 \times 10 + 50 = 54.79$$

Z 值

————— 0 —————

(-) (+)

偏差值 50

Z 値は、平均値 = 0

平均値より大のときプラスの値、小のときマイナスの値

偏差値は、平均値 = 50

平均値より大のとき50より大きな値、小のとき50より小さな値

◇営業成績の評価

売上高実績

昨年度	3,011
今年度	3,225

7.11% ↑

評価は？

A支店の場合

	前年度	今年度	前年同月比
1月	235	250	1.064
2月	246	261	1.061
3月	251	269	1.072
4月	252	268	1.063
5月	258	279	1.081
6月	249	266	1.070
7月	267	288	1.079
8月	246	262	1.065
9月	250	271	1.084
10月	242	263	1.087
11月	249	264	1.060
12月	266	284	1.068
計	3011	3225	1.071

B支店の場合

	前年度	今年度	前年同月比
1月	235	230	0.979
2月	246	242	0.984
3月	251	241	0.960
4月	252	248	0.984
5月	258	250	0.970
6月	249	241	0.968
7月	267	259	0.970
8月	246	240	0.976
9月	250	541	2.164
10月	242	233	0.963
11月	249	240	0.964
12月	266	260	0.977
計	3,011	3,225	1.071

標準偏差（STDEV.P）を求める

B16										
	A	B	C	D	E	F	G	H	I	J
1	A支店					B支店				
2		前年度	今年度	前年同月比			前年度	今年度	前年同月比	
3	1月	235	250	1.064		1月	235	230	0.979	
4	2月	246	261	1.061		2月	246	242	0.984	
5	3月	251	269	1.072		3月	251	241	0.960	
6	4月	252	268	1.063		4月	252	248	0.984	
7	5月	258	279	1.081		5月	258	250	0.970	
8	6月	249	266	1.070		6月	249	241	0.968	
9	7月	267	288	1.079		7月	267	259	0.970	
10	8月	246	262	1.065		8月	246	240	0.976	
11	9月	250	271	1.084		9月	250	541	2.164	
12	10月	242	263	1.087		10月	242	233	0.963	
13	11月	249	264	1.060		11月	249	240	0.964	
14	12月	266	284	1.068		12月	266	260	0.977	
15	計	3,011	3,225	1.071		計	3,011	3,225	1.071	
16	標準偏差	8.788803	10.13268			標準偏差	8.788803	82.53926		
17										

売上高実績一覧表

支店	昨年度			今年度			前年比
	売上高	標準偏差	変動係数	売上高	標準偏差	変動係数	
A	3,011	8.789	0.035	3,225	10.133	0.038	1.071
B	3,011	8.789	0.035	3,225	82.539	0.307	1.071
E	30,110	87.890	0.035	32,253	825.390	0.307	1.071

・A支店とB支店の比較

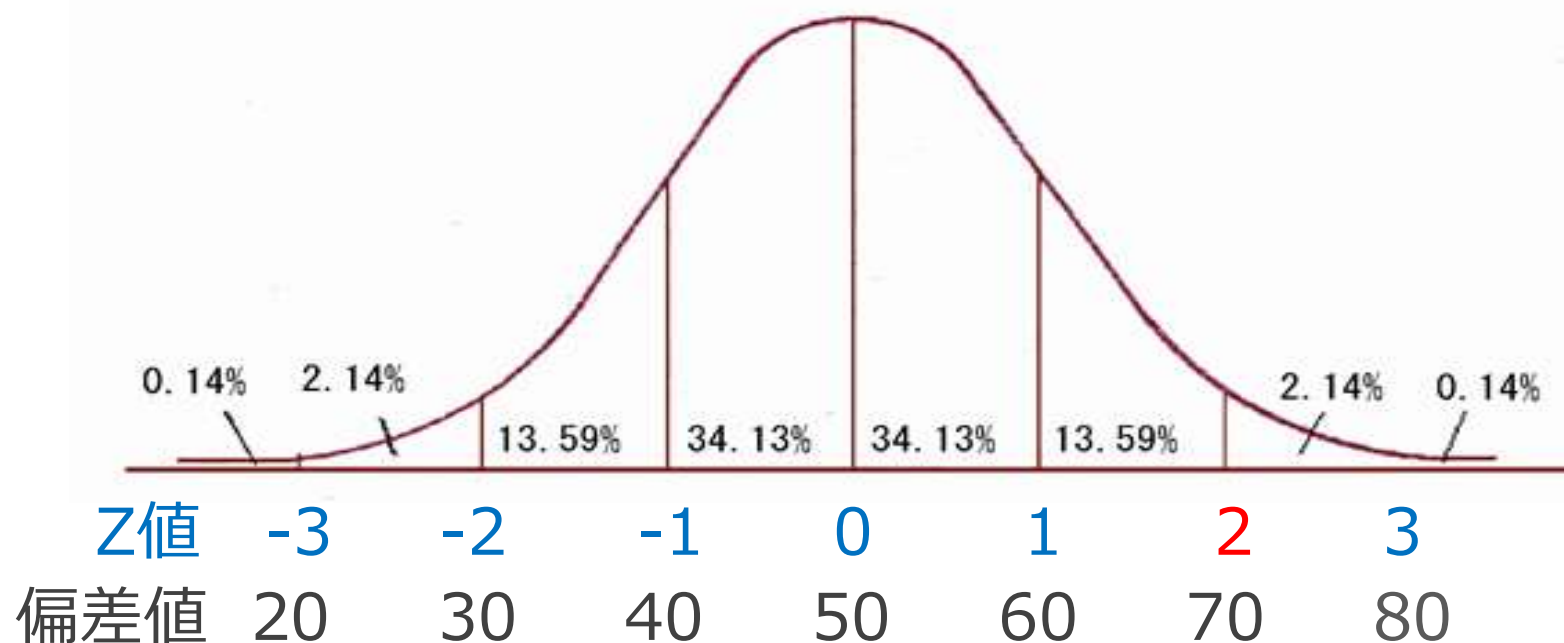
昨年度：同じ

今年度：B支店のバラツキ（標準偏差）が大きい

・B支店とE支店の比較

E支店の実績はB支店の10倍 構造は同じ（ $B=E$ ）

標準正規分布 (standard normal distribution)

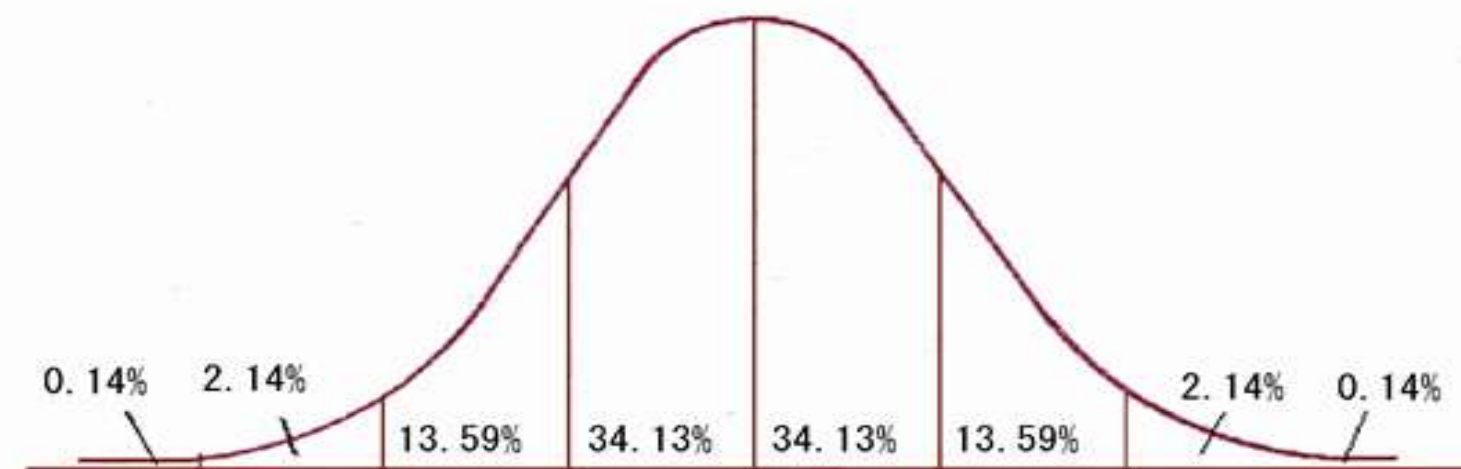


Z 値 = 1 (偏差値60) 以上の割合は約15.87%

Z 値 = 2 (偏差値70) 以上の割合は全体の約2.28%

標準正規分布の活用

- 日本の成年男子の身長が平均170cm 標準偏差6cm、
182cm以上の人は、全体の何%を占めるか？



身長(cm)

170

182

Z 値

0

2

$$Z = \frac{182 - 170 \text{ (cm)}}{6 \text{ (cm)}} = \mathbf{2}$$

182cm以上の人は約2.28%

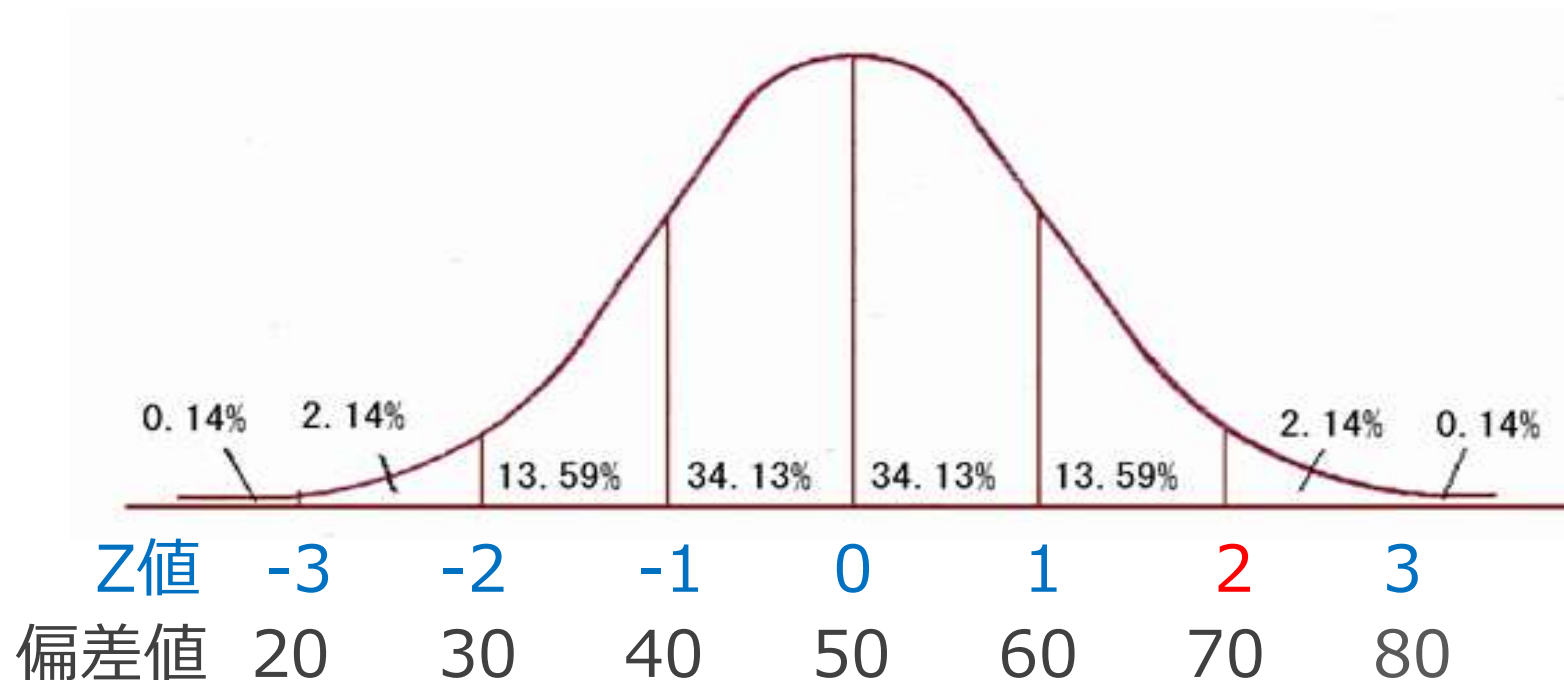
- 通信販売会社における顧客売上高の平均5,568円、標準偏差2,216円、10,000円以上購入した顧客は、全体の何%を占めるか？

$$Z = \frac{10,000 - 5,568 \text{ (円)}}{2,216 \text{ (円)}} = 2 \quad 10,000\text{円以上} : \text{約}2.28\%$$

Z 値は単位とは無関係。

$$Z \text{ 値} = \frac{\text{求める値} - \text{平均値}}{\text{標準偏差}}$$

標準正規分布 (standard normal distribution)



Z 値 = 2 (偏差値70) 以上の割合は全体の約2.28%
⇒ Z値は0を基準とした違いの大きさ

違いの大きさの判断方法

◇新製品の好感度について、男女別に10点満点にて調査した。
男女間の評価に違いは見られるか？

2019年	男性	6	4	5	5	6	5	6	6	4	6	5.3
	女性	7	6	7	5	6	5	6	7	6	6	6.1



平均値の違い 0.8 ($6.1 - 5.3$) は大きい？

平均値 (5.3, 6.1) は信頼できる？

◇新製品の好感度について、男女別に10点満点にて調査した。
男女間の評価に違いは見られるか？

2020年	男性	6	4	7	3	6	3	6	6	4	8	5.3
	女性	7	3	8	5	9	2	5	7	6	9	6.1



2020年のデータは2019年に比較して
全体的にバラツキが大きい。

平均の信頼度

2019年 バラツキが小 $\Rightarrow$ 平均値は信頼できる
差 (0.8) も信頼できる。

2020年 バラツキが大 $\Rightarrow$ 平均値は信頼できない
差 (0.8) も信頼できない。



差 (0.8) が同じであっても
男女間の違いの大きさ : 2019年 > 2020年

違いの大きさの判断方法

①男性と女性の平均値の差の大きさの検討

男女の平均値の「差」が大きい程、
男女間の好感度は違うと判断できる。

②平均値の信頼性の検討

男女の各データの「バラツキ」が大きい程、
男女間の好感度は違うと判断できない。



①男女間の平均値の「差」の大きさ

②男女各データの「バラツキ」の大きさ

上記の値が大きいほど違うと判定できる。

判断に対する調査人数（標本の大きさ）の影響

男女各10人の調査、男女各100人の調査
どちらが違いに対する信頼性が高い？

100人の調査！

$$\text{違いの大きさ} = \frac{\text{①平均値の「差」の大きさ}}{\text{②データの「バラツキ」の大きさ}} \times \text{③(調査人数)}$$



- ①が大きい程、違いは大きい。
- ②が小さい程、違いは大きい。
- ③が大きい程、違いは大きい。

◇主力製品におけるCM好感度調査

- ① 昨年度100人について調査を実施。
違いがあるという判定が出来なかった。
- ② 今年度は10万人について調査を実施。
違いがあるという判定ができた！



CMの効果は？

「違い」と「効果」

「違いの大きさ」は標本の大きさの影響を受ける。

ビッグデータにおいては
「違いがある！」という分析結果が出やすい。

違いがある **≠** 効果が大さい

調査人数（サンプルサイズ）を勘案することが重要！

対応のある場合

商品の理解度について、10人に商品説明前後に、それぞれ10点満点にて調査した。
説明前後による理解度に差は見られるか。

説明前	7	3	8	3	7	2	2	6	6	9	5.3
説明後	7	4	9	3	8	3	5	8	7	9	6.3



データ数 = 20
回答者 10 人のデータ

説明前後の比較

											平均値
説明前	7	3	8	3	7	2	2	6	6	9	5.3
説明後	7	4	9	3	8	3	5	8	7	9	6.3
差	0	1	1	0	1	1	3	2	1	0	1.0

差に着目

平均値に着目する場合、差に着目する場合
分析結果が大きく異なる。

◇営業成績AとBの実績比較

											平均
A	70	55	60	50	70	65	65	70	60	50	61.5
B	65	60	50	45	65	65	50	70	45	50	56.5

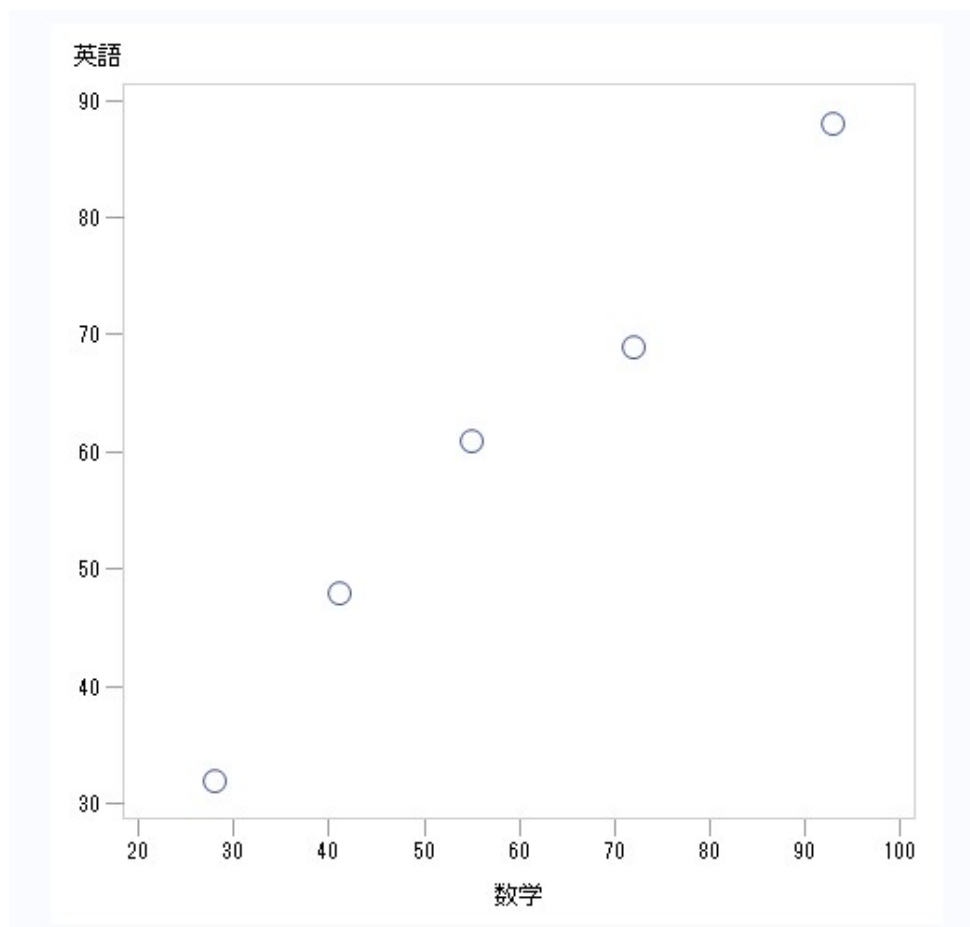
a) 過去10ヶ月のデータの場合

b) 顧客10店舗のデータの場合

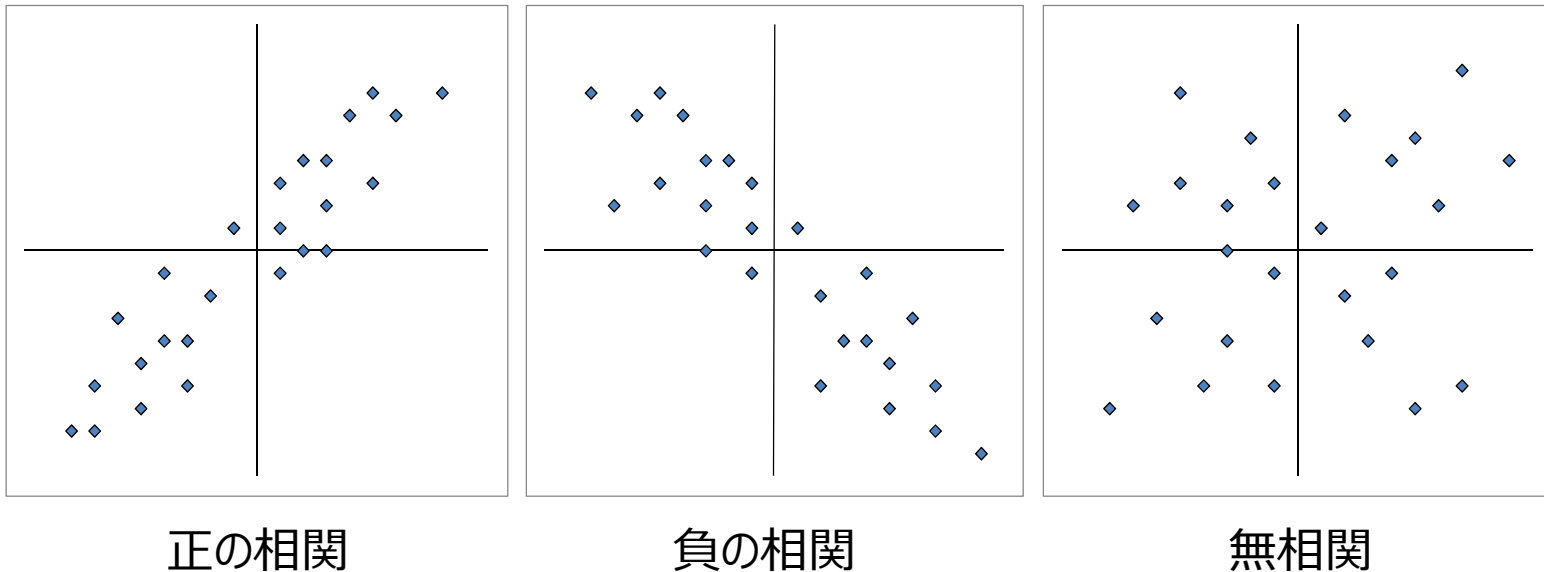
相関分析

数学	英語
28	32
55	61
93	88
72	69
41	48

散布図



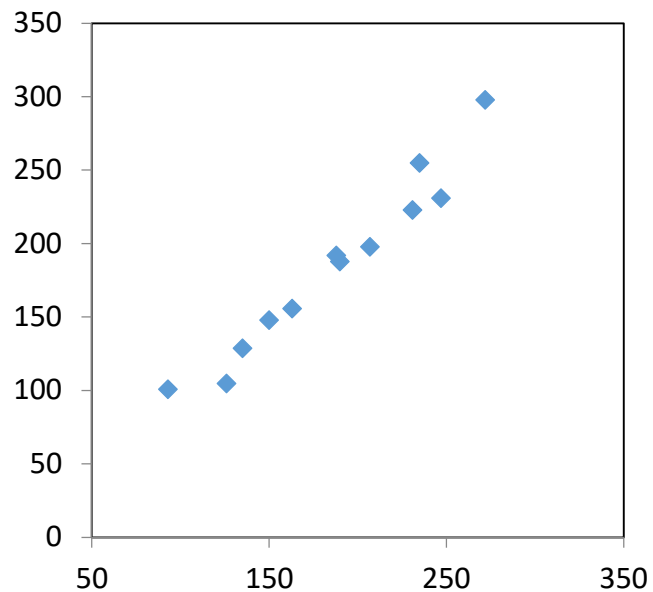
データ群の関係



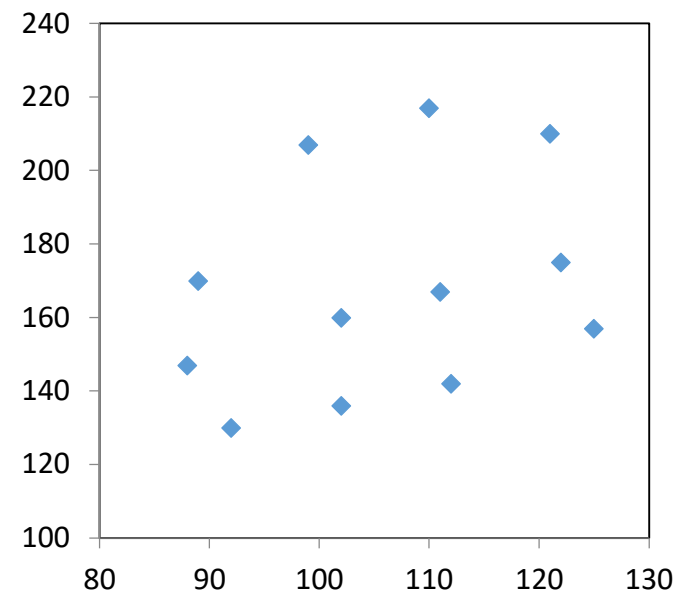
相関関係には正の相関、負の相関、無相関。
点の集中度が関係の強さを測定する手がかり。

相関係数 (r)

$$-1 \leq r \leq 1$$



$r = 0.97$



$r = 0.32$

相関係数 (r) は相関関係の強さ

相関係数（ r ）の解釈

$r = 0 \quad \Rightarrow \quad \text{説明力は0\%}$

$r = 1 \quad \Rightarrow \quad \text{説明力は100\%}$

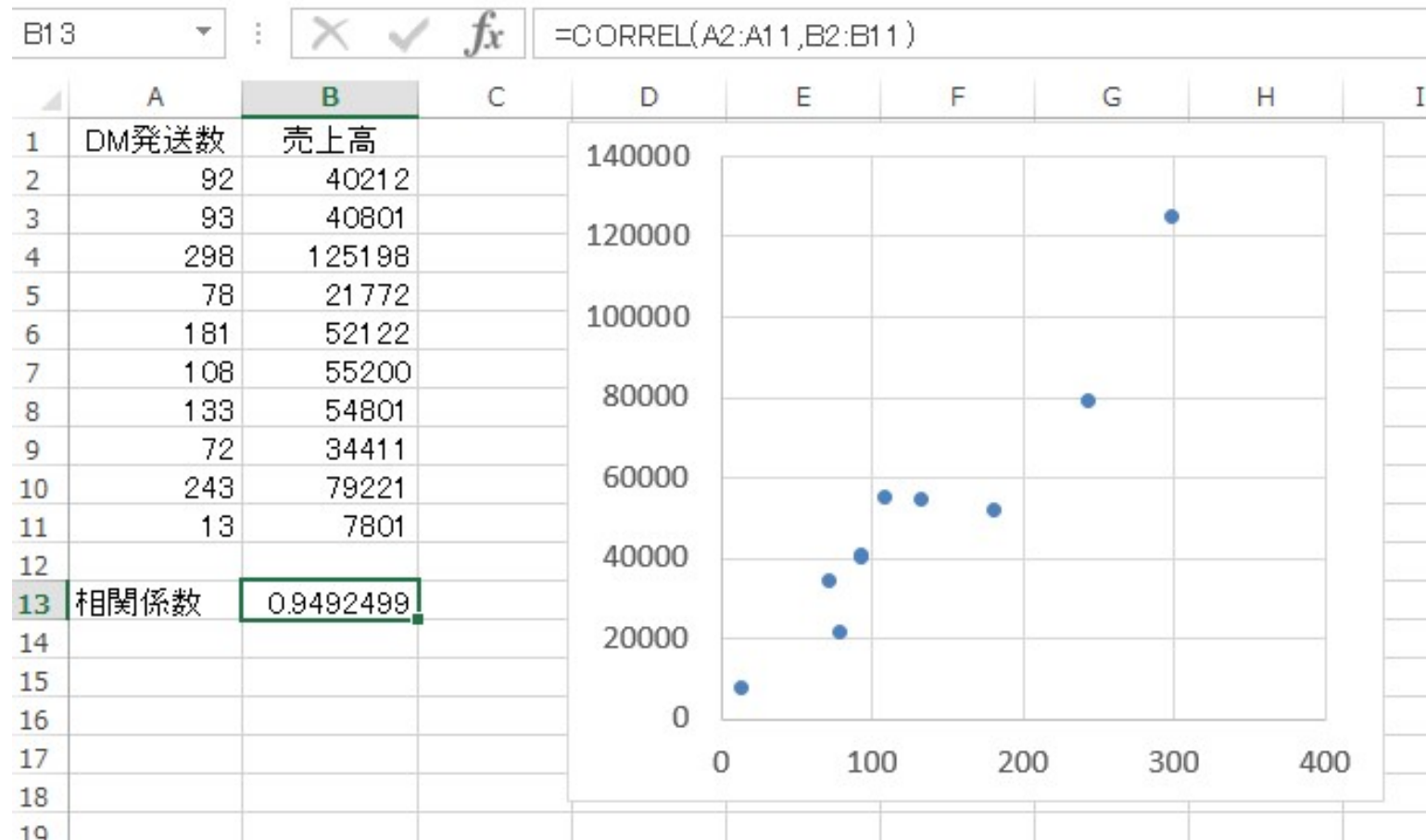
$r = 0.32 \Rightarrow \quad \text{説明力は？\%}$

$$0.32 \times 0.32 = 0.1024$$

説明力は、10.24%

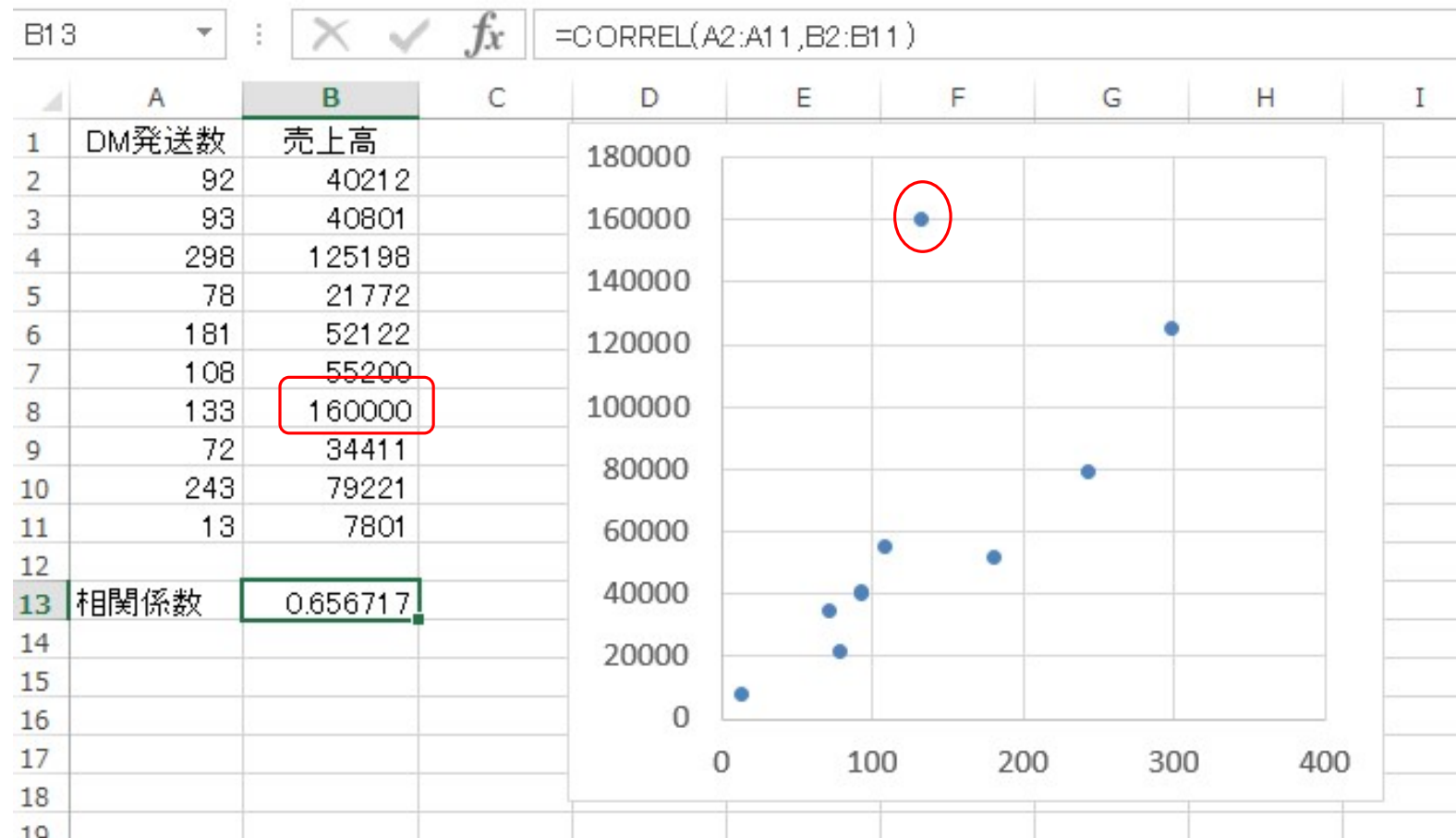
相関係数について

EXCEL : CORREL



$$r = 0.949$$

外れ値を含む場合



$r = 0.657$

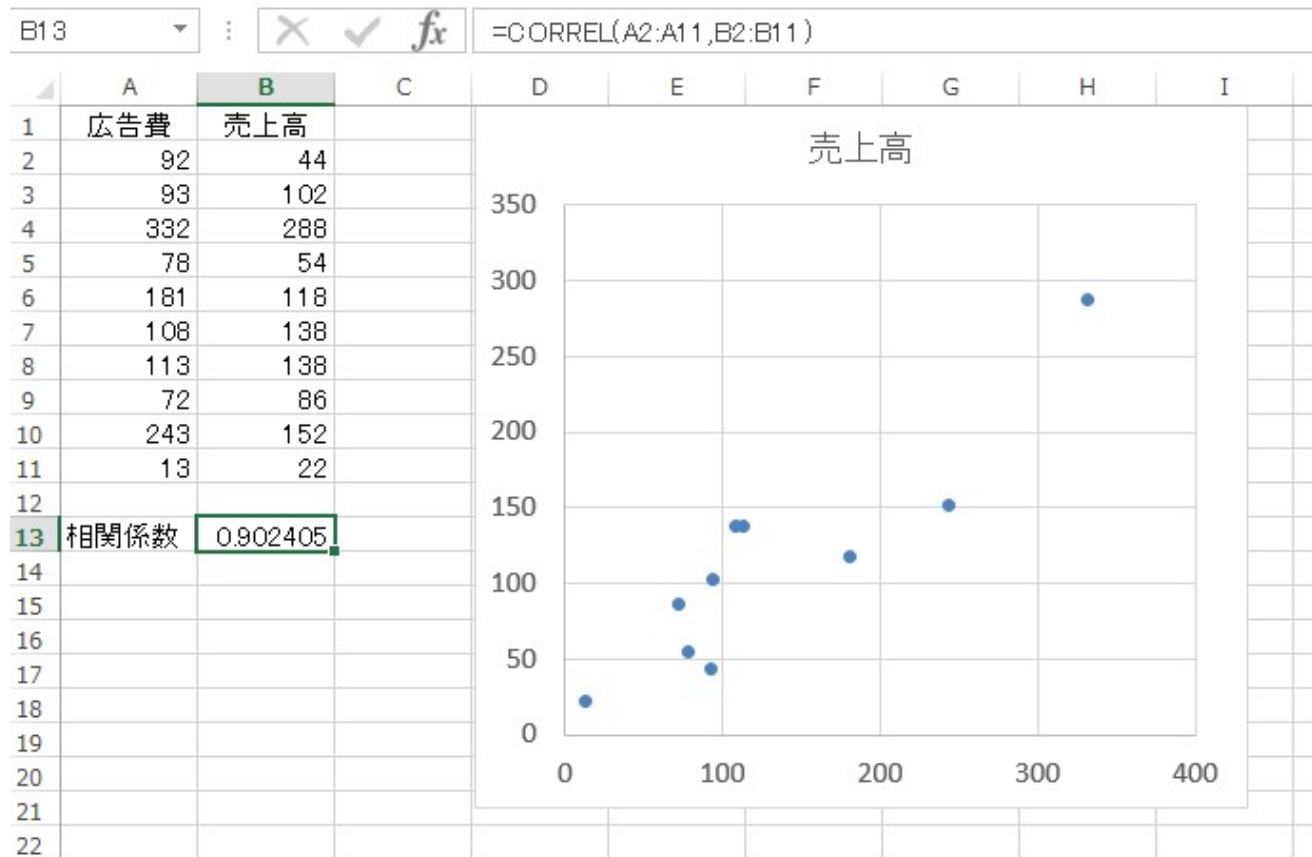
外れ値の有無の検討は必須

◇支店別広告費と売上高

支店	広告費	売上高
北海道	92	44
東北	93	102
関東	332	288
北陸	78	54
中部	181	118
近畿	108	138
中国	113	138
四国	72	86
九州	243	152
沖縄	13	22

売上高と広告費の関係？

相関係数の算出

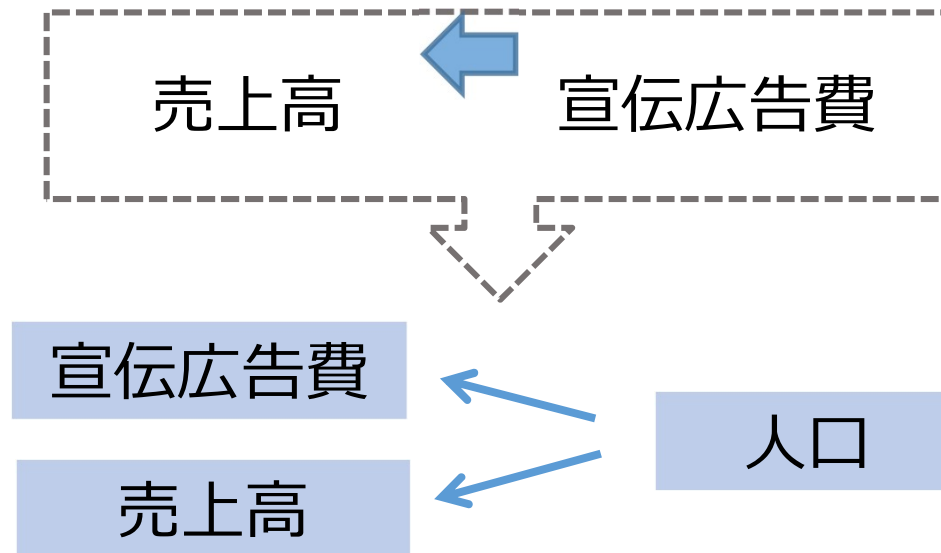


$$r = 0.902$$

広告宣伝費は売上高に貢献？

◇支店別広告費、売上高、人口

支店	広告費	売上高	人口
北海道	92	44	5506
東北	93	102	9335
関東	332	288	42604
北陸	78	54	5443
中部	181	118	18127
近畿	108	138	12912
中国	113	138	15554
四国	72	86	3976
九州	243	152	13204
沖縄	13	22	1393



人口は交絡要因 $\Rightarrow$ 疑似相関

相関係数の算出

広告費と売上高 $r = 0.902$

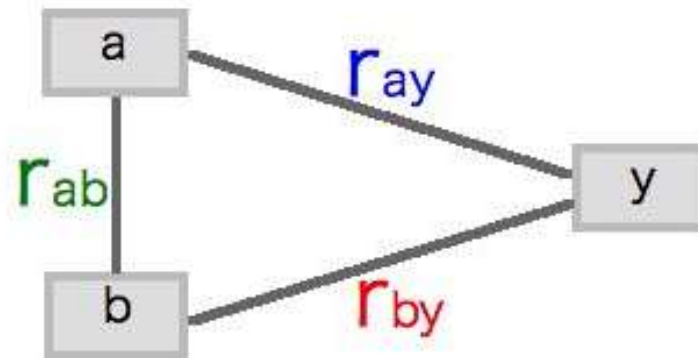
人口と広告費 $r = 0.895$

人口と売上高 $r = 0.951$

人口は交絡要因

人口の影響を除いたときの広告費と売上高の
相関係数 ➡ 偏相関係数

偏相関係数の求め方 (EXCEL)



aの影響を除いた、bとyの偏相関係数

$$\frac{r_{by} - (r_{ay} \times r_{ab})}{\sqrt{1 - r_{ay}^2} \times \sqrt{1 - r_{ab}^2}}$$

偏相関係数の計算

$$\frac{0.902 - (0.951 \times 0.895)}{\sqrt{1 - 0.951^2} \times \sqrt{1 - 0.895^2}} \\ = 0.368$$

人口の影響を除いたときの広告費と売上高の
相関係数 : 0.368

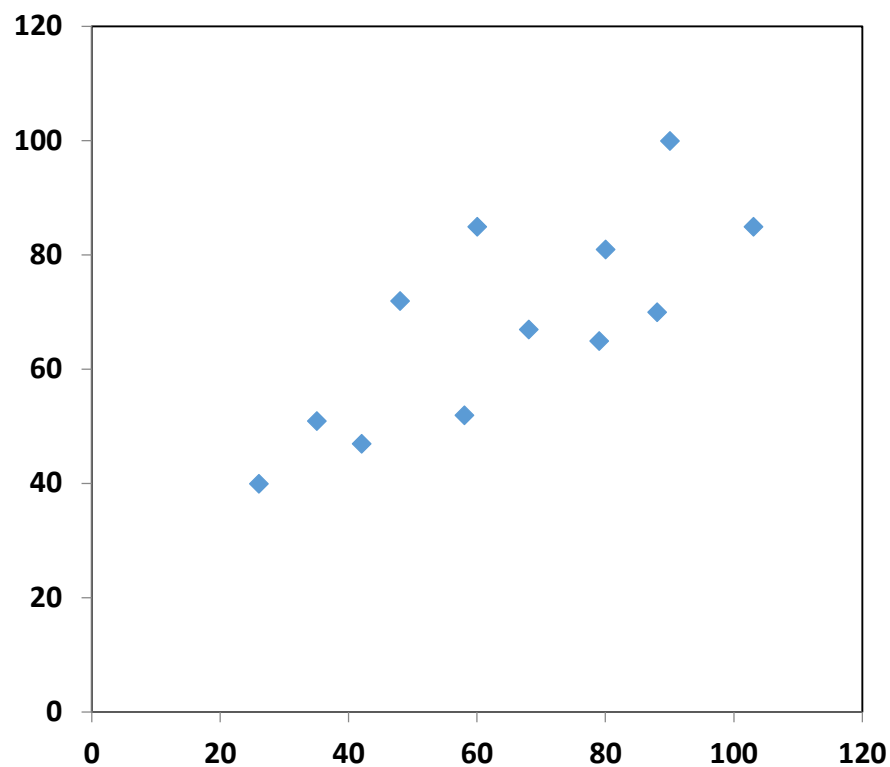
A支店の広告費と売上高

月	4	5	6	7	8	9	10	11	12	1	2	3
広告費	42	26	35	68	103	90	88	58	79	80	48	60
売上高	47	40	51	67	85	100	70	52	65	81	72	85

$$r = 0.775$$



強い正の相関！



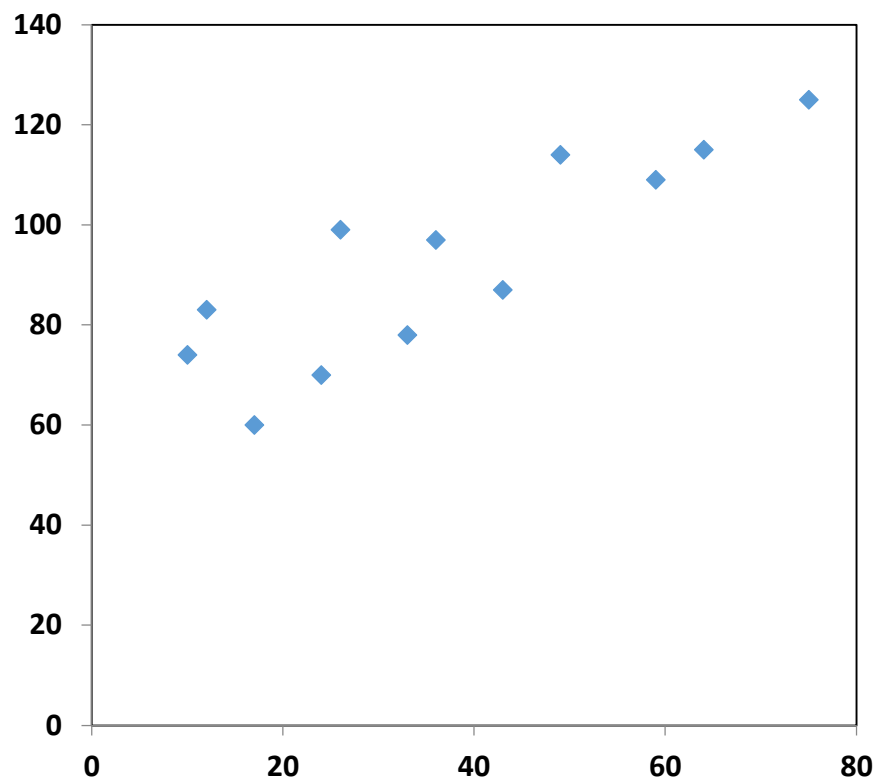
B支店の広告費と売上高

月	4	5	6	7	8	9	10	11	12	1	2	3
広告費	12	10	17	59	75	64	33	24	49	43	36	26
売上高	83	74	60	109	125	115	78	70	114	87	97	99

$$r = 0.863$$



強い正の相関！



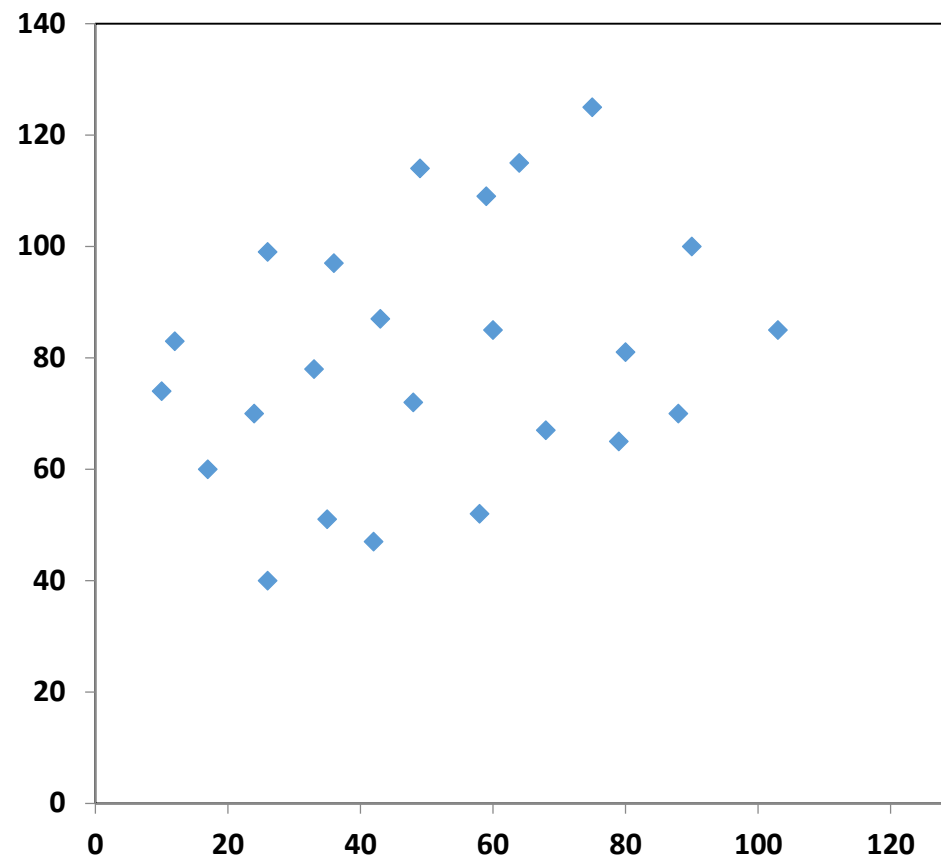
70

(A支店 + B支店) の広告費と売上高

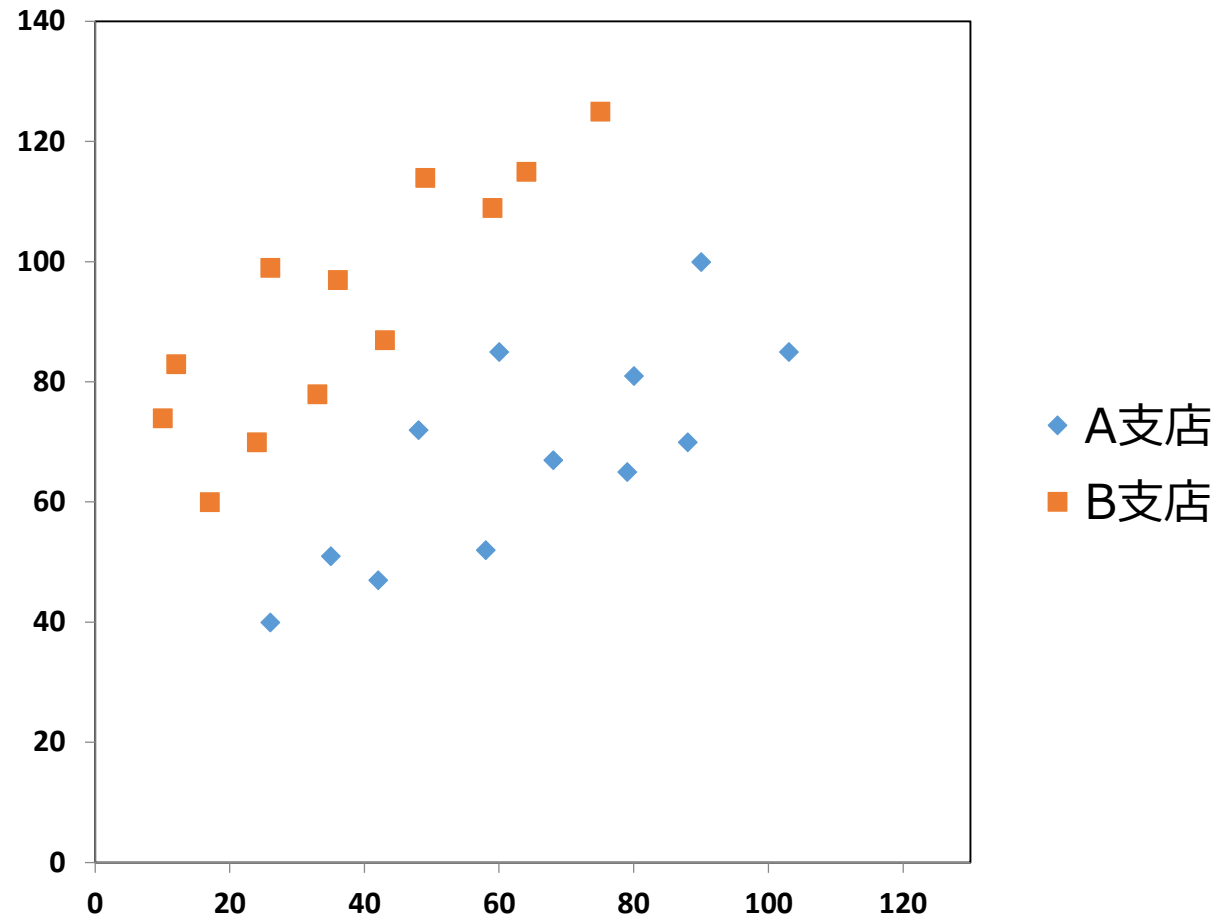
$$r = 0.272$$



???



(A支店 + B支店) の広告費と売上高



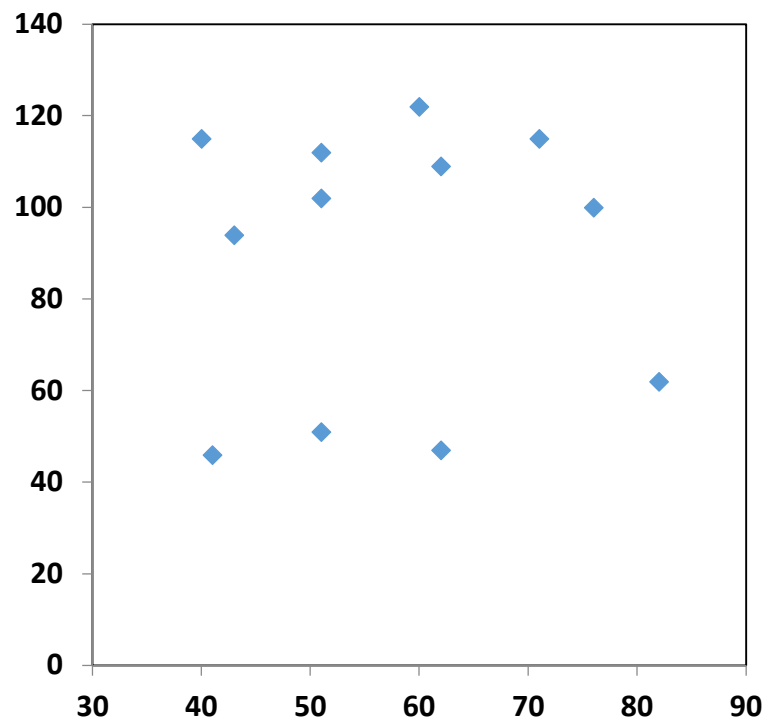
A支店の広告費と売上高

月	4	5	6	7	8	9	10	11	12	1	2	3
広告費	43	51	62	76	41	51	62	82	40	51	60	71
売上高	94	102	109	100	46	51	47	62	115	112	122	115

$$r = 0.016$$



無相関！



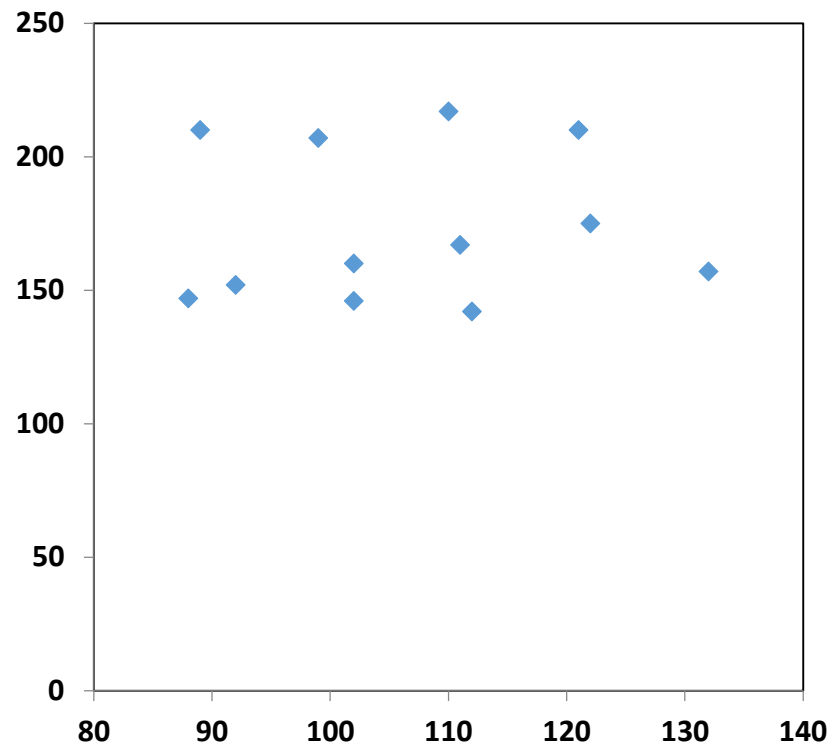
B支店の広告費と売上高

月	4	5	6	7	8	9	10	11	12	1	2	3
広告費	92	102	111	122	88	102	112	132	89	99	110	121
売上高	152	160	167	175	147	146	142	157	210	207	217	210

$$r = 0.040$$



無相関！

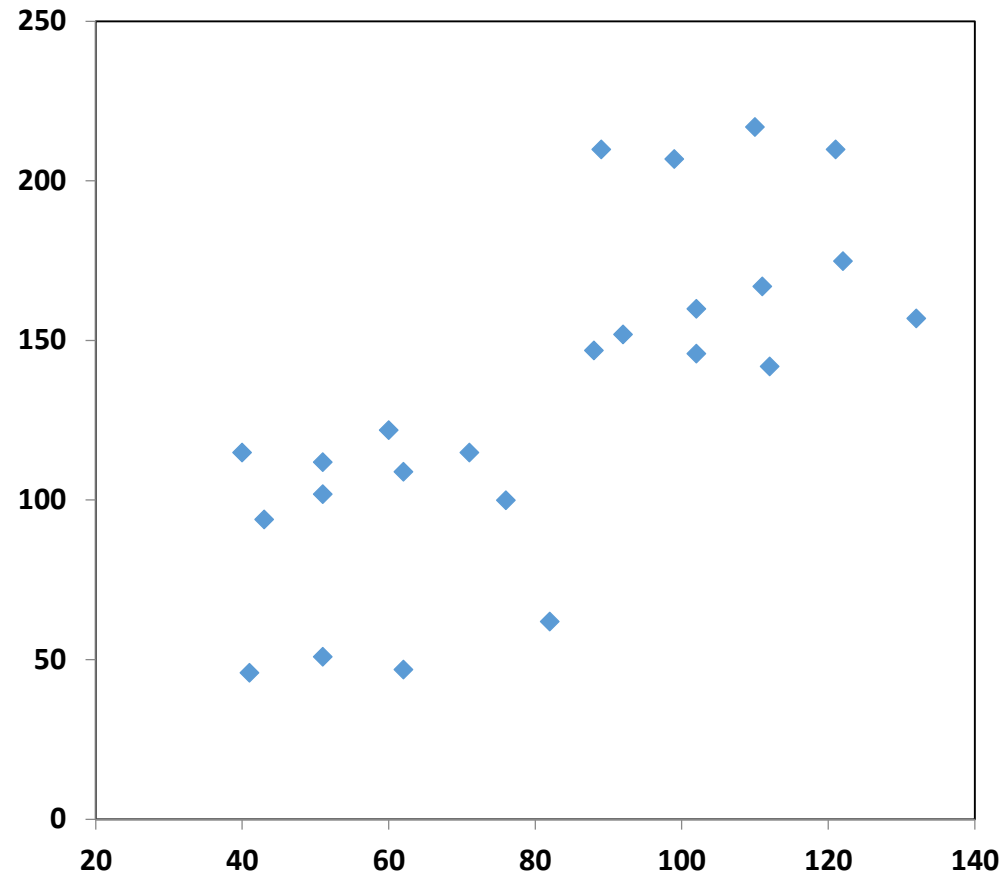


(A支店 + B支店) の広告費と売上高

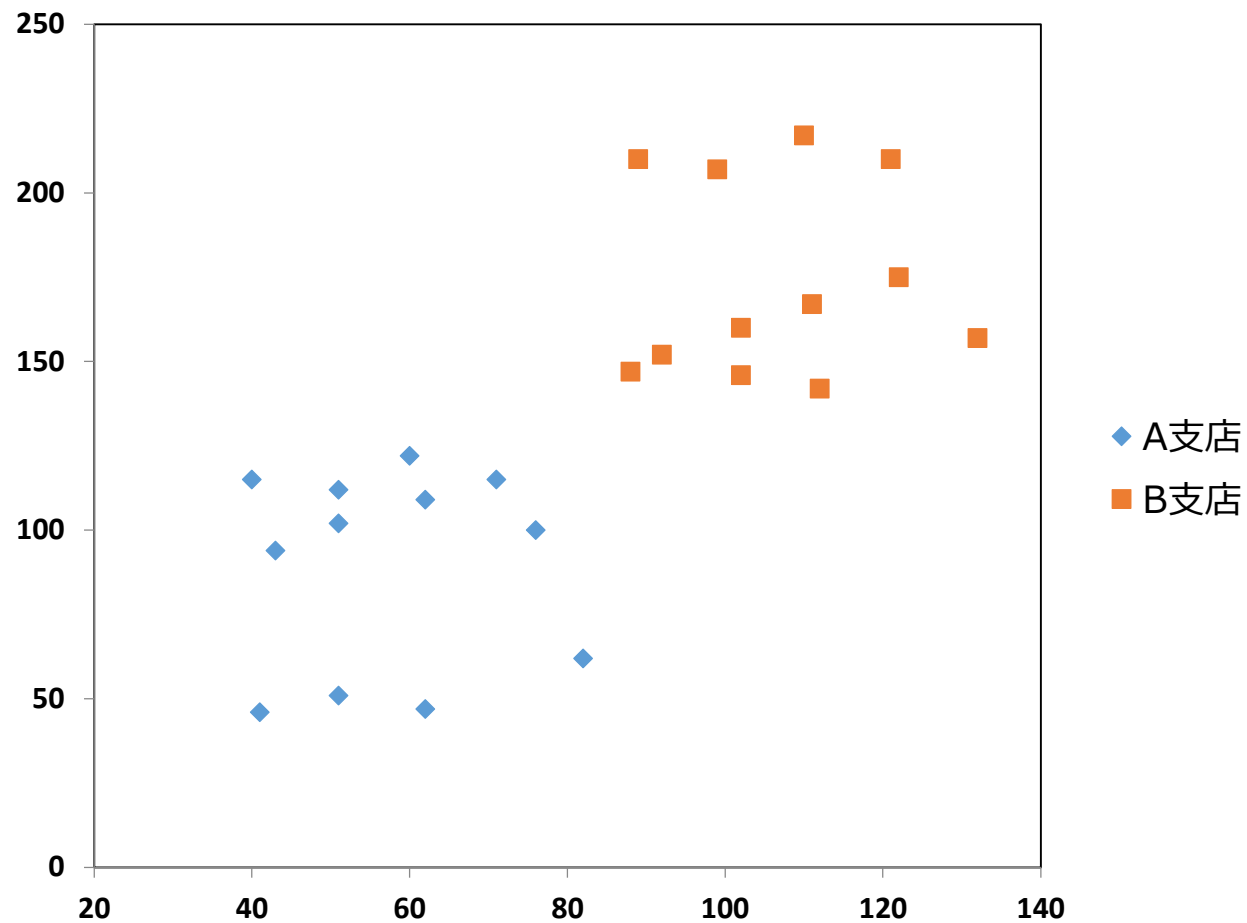
$$r = 0.823$$



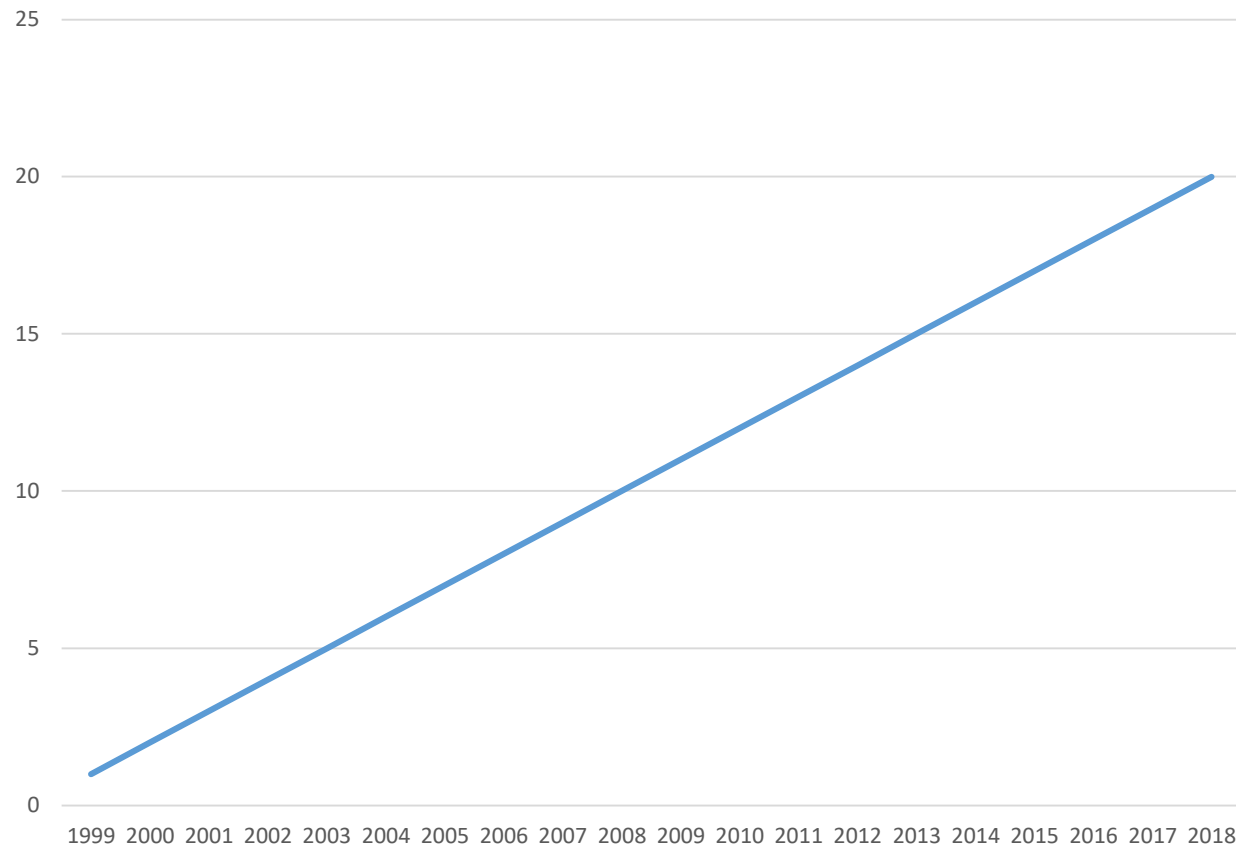
???



(A支店 + B支店) の広告費と売上高



売上高推移（設立～20年）



売上高は順調に伸びている!?

年度	売上高	前年比
1999	1	-
2000	2	2.00
2001	3	1.50
2002	4	1.33
2003	5	1.25
2004	6	1.20
2005	7	1.17
2006	8	1.14
2007	9	1.13
2008	10	1.11
2009	11	1.10
2010	12	1.09
2011	13	1.08
2012	14	1.08
2013	15	1.07
2014	16	1.07
2015	17	1.06
2016	18	1.06
2017	19	1.06
2018	20	1.05

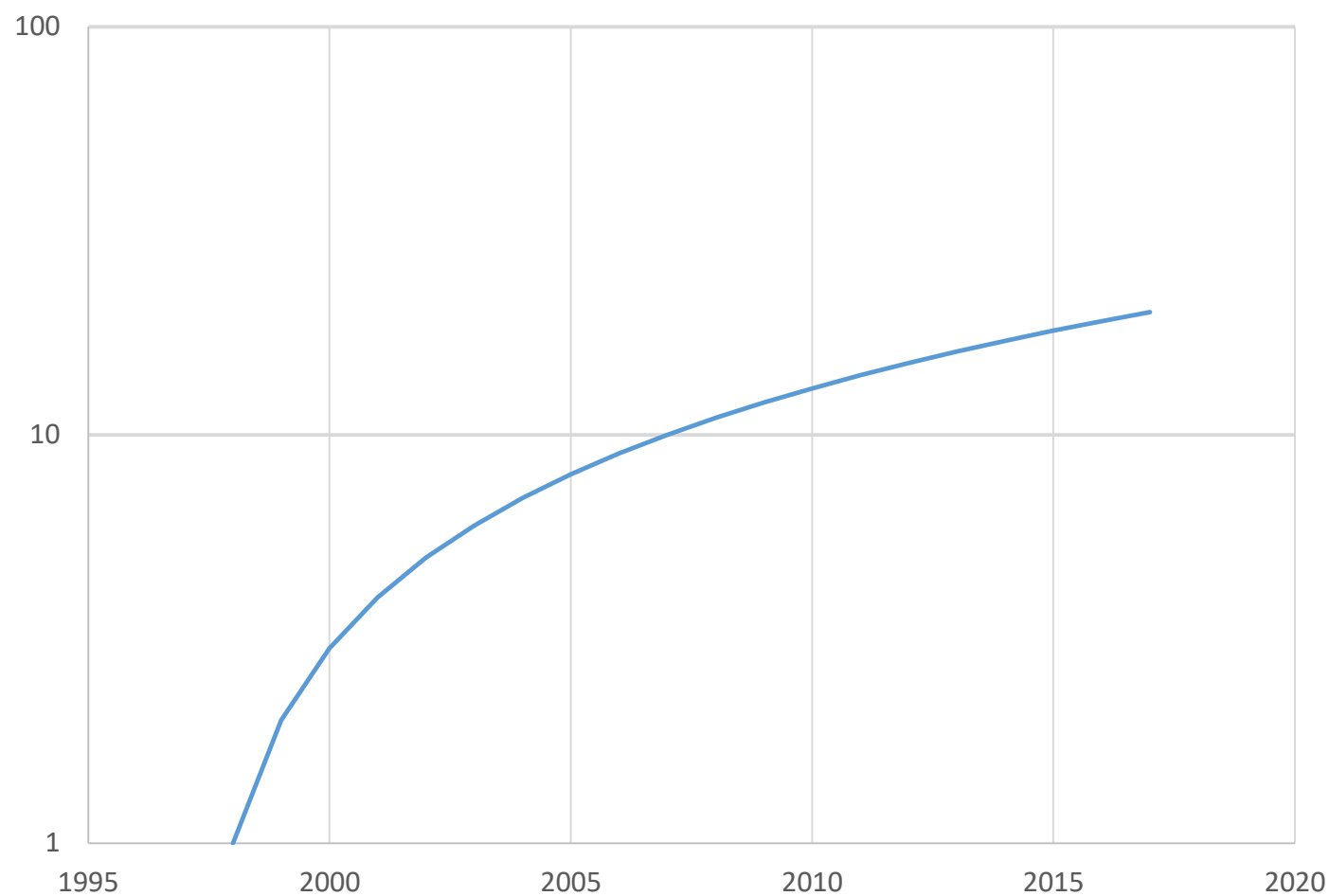
・年度間の差に着目

+1

・年度間の比に着目

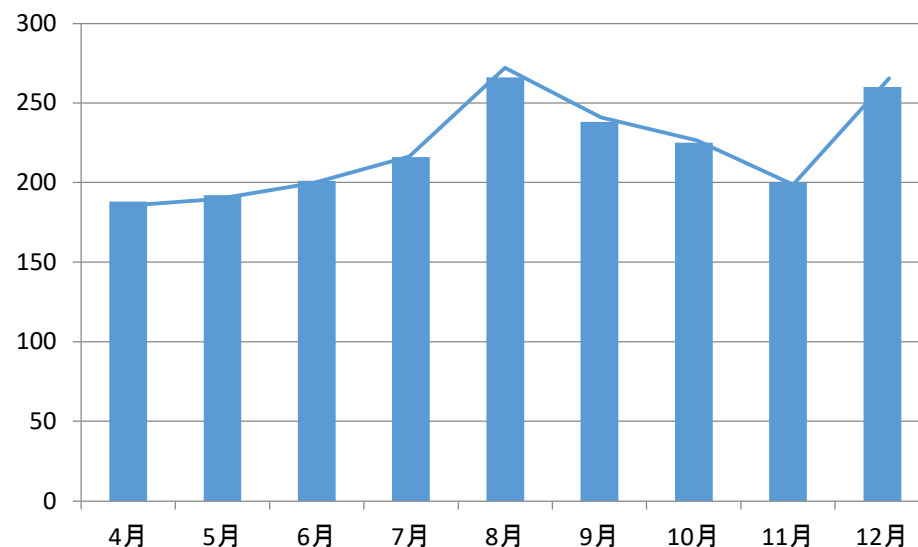


比率は年々減少！



増加率は年々減少傾向！

折れ線グラフと棒グラフの違いは何か？



目の動きは棒をイメージして上下に動かすべき



目の動きを斜めに動かすグラフは？

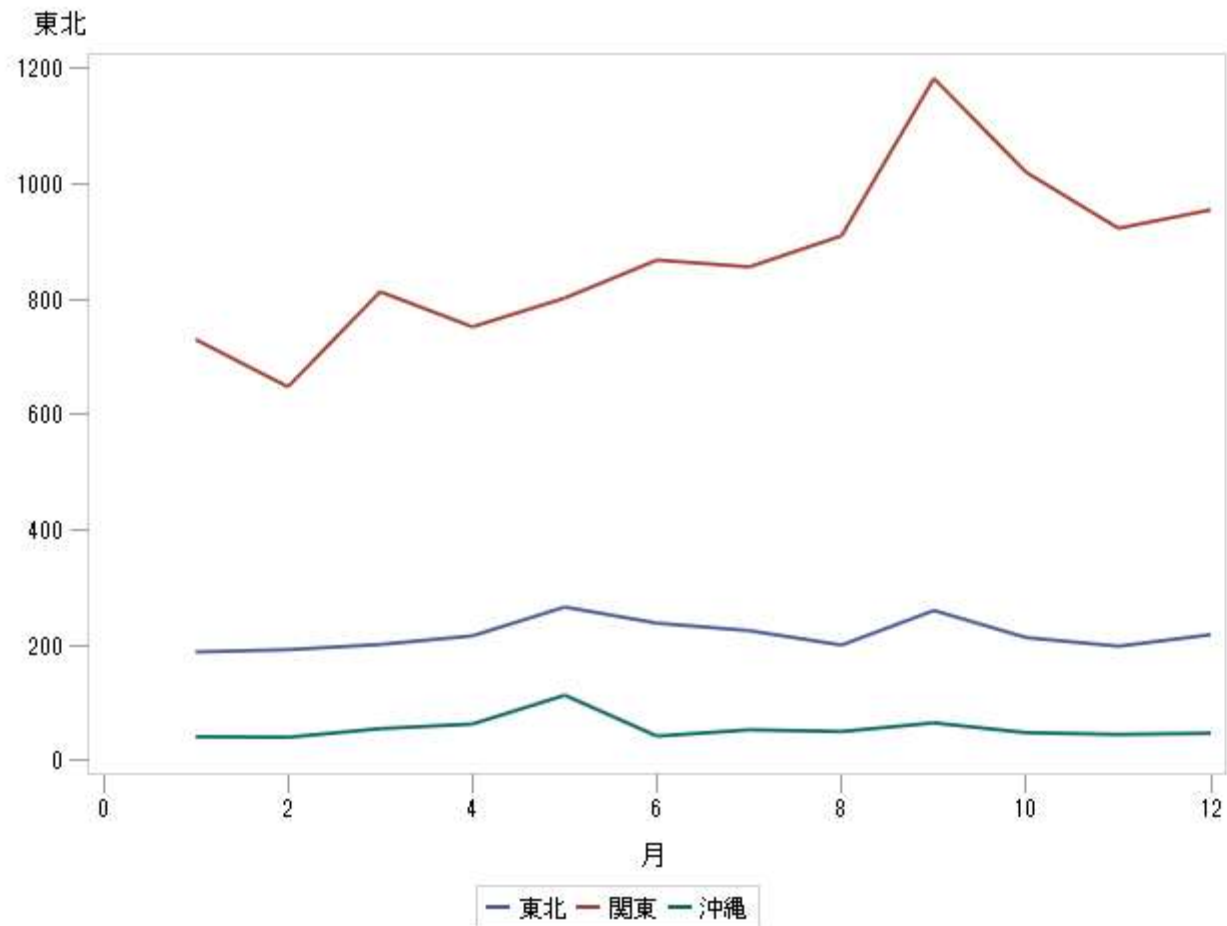
支店（東北、関東、沖縄）別売上高

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
東北	188	192	201	216	266	238	225	200	260	213	198	218
関東	730	648	813	752	802	868	856	910	1183	1020	923	955
沖縄	41	40	55	63	113	42	53	50	65	48	45	47

	4月	5月	差	前月比
東北	216	266	50	1.231
関東	752	802	50	1.066
沖縄	63	113	50	1.794

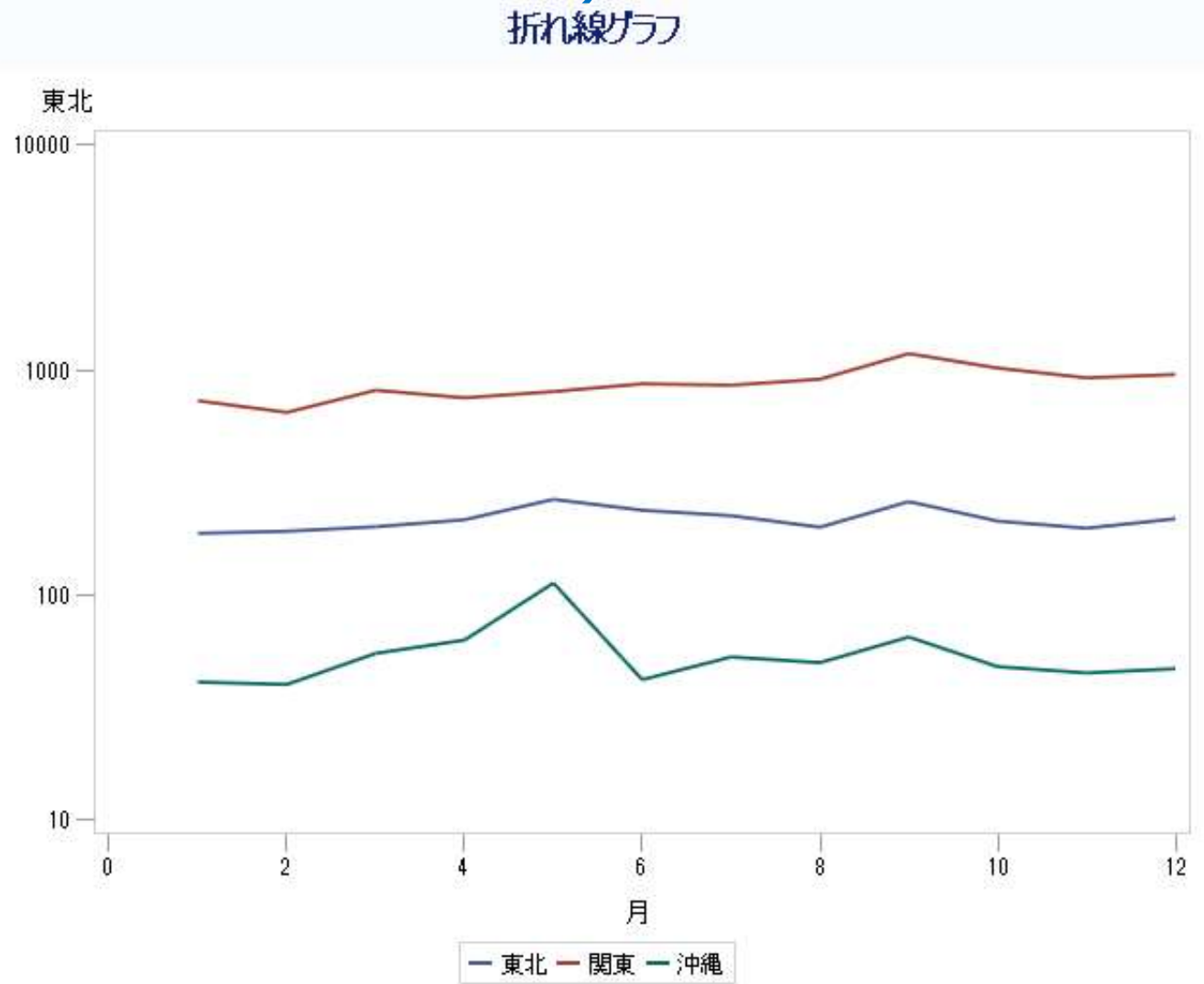
	8月	9月	差	前月比
東北	200	260	60	1.3
関東	910	1183	273	1.3
沖縄	50	65	15	1.3

折れ線グラフ



⇒ 普通目盛の折れ線グラフでは比較困難

折れ線グラフ（対数目盛）

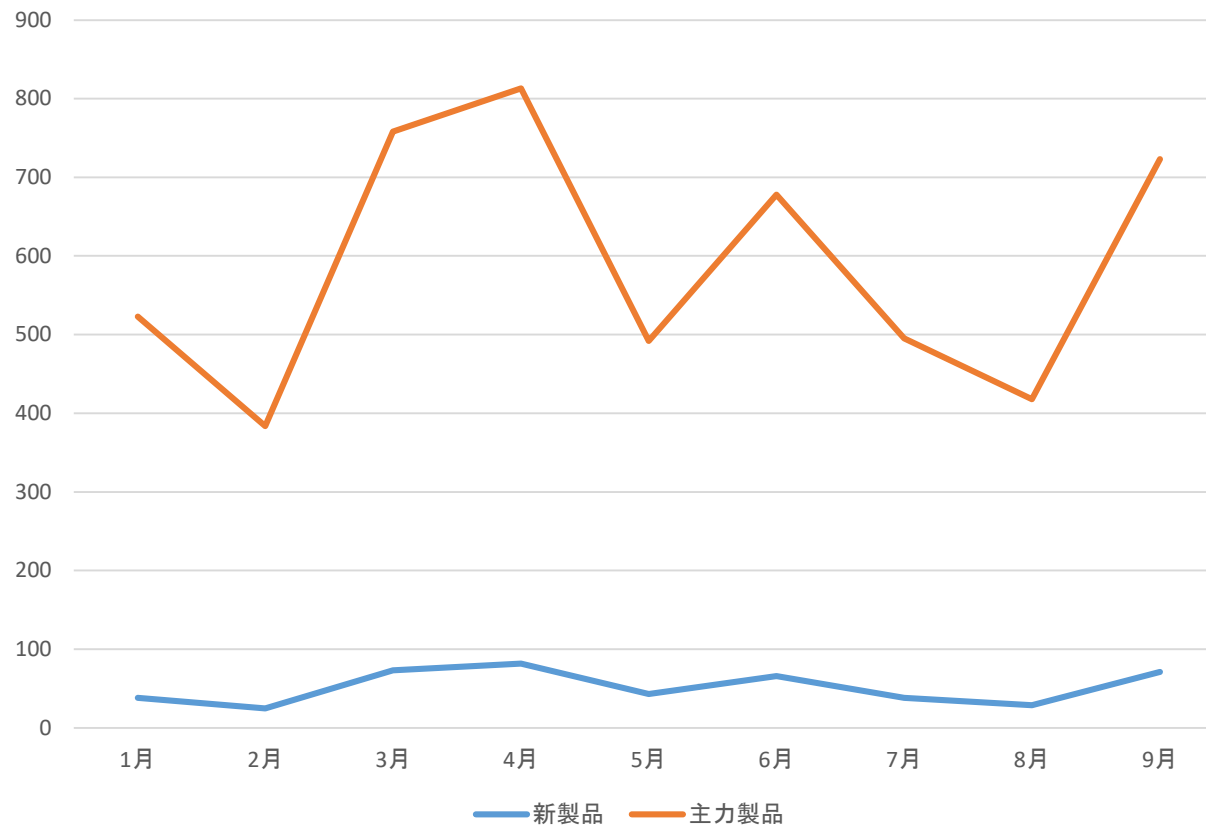


◇新製品と主力製品の売上高

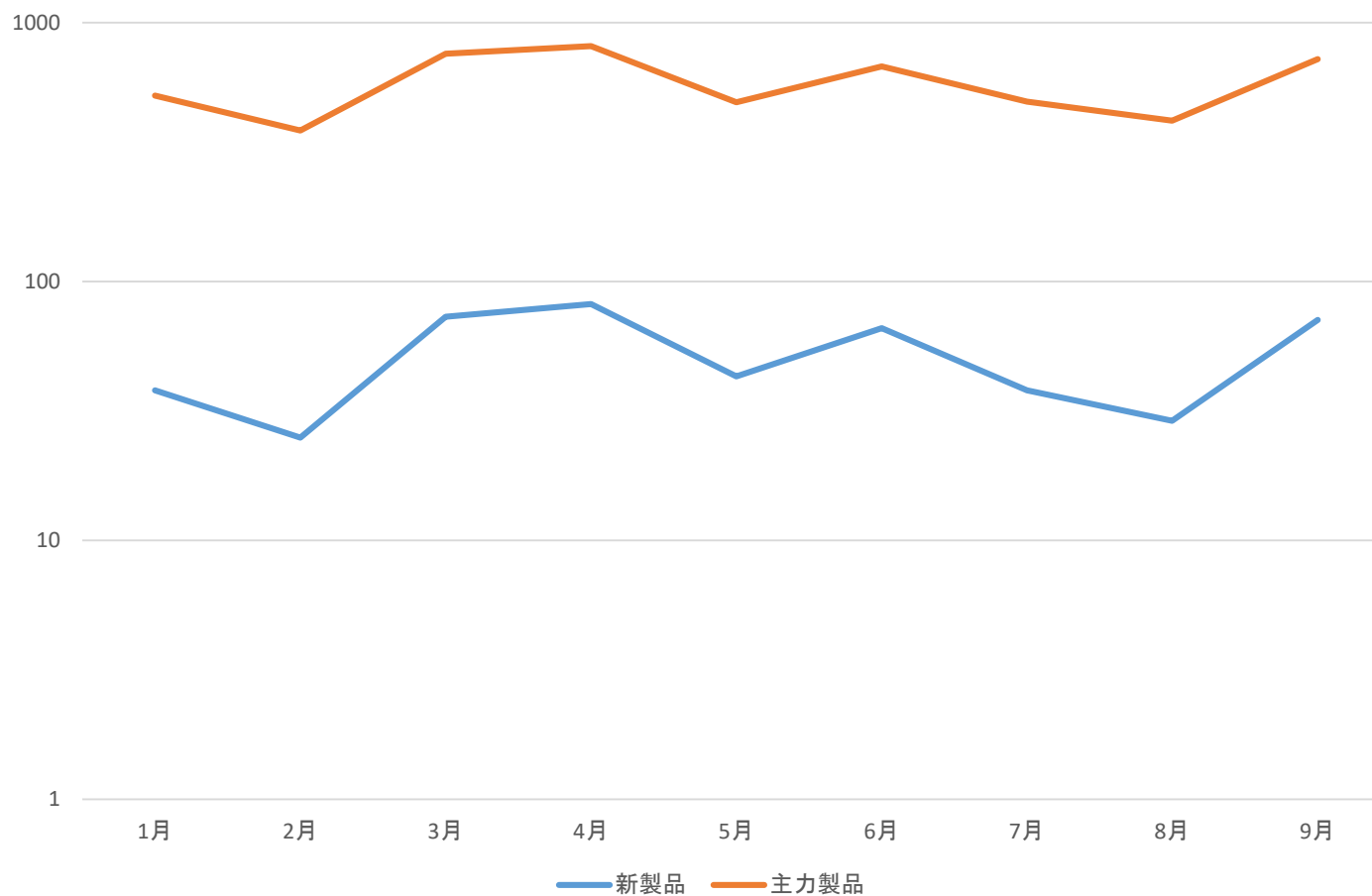
(百万円)

	新製品	主力製品
1月	38	523
2月	25	384
3月	73	758
4月	82	813
5月	43	492
6月	66	678
7月	38	495
8月	29	418
9月	71	723

折れ線グラフ（普通目盛）



折れ線グラフ（対数目盛）



人間の五感は対数に変換されている

◇ウェーバー・フェヒナーの法則

弁別閾（気づくことができる最小の刺激差）は刺激の値に比例

手に重りを100gのせ、少しずつ重りを加え、
重さを感じたのが110gのとき、
手に重りを200gのせ、1gずつ重りを加えると、
重さを感じるのは220gのときである。

デシベル、P H、マグニチュード、等星・・・

分析結果の検証

{レトルトカレー}⇒{洗濯用洗剤}

- ✕ カレーを食べると服が汚れるから
- たまたま、カレーと洗剤を同じ日に特売した

分析から得た知識は、無条件に受け入れずに必ず検証し、より妥当な知識を見つけるよう試行錯誤

見つけた知識の合理性は必ず検証

まとめ

有用なデータ = 数値データ + 背景

- ・人材

統計学（機械学習）、コンサル力（翻訳力）
システム系

データの活用 ⇒ 統計学基本知識（機械学習）
× 現場実践力

社内インストラクターの養成も重要！

■ ■ 統計分析力養成講座 ■ ■

=====

2021年7月10日 初版 第1刷

発行元： 株式会社 データサイエンス研究所

本社 〒152-0021 東京都千代田区平河町2-5-5 全国旅館会館
tel : 03-3265-3908 mail : info@datascience.co.jp

=====

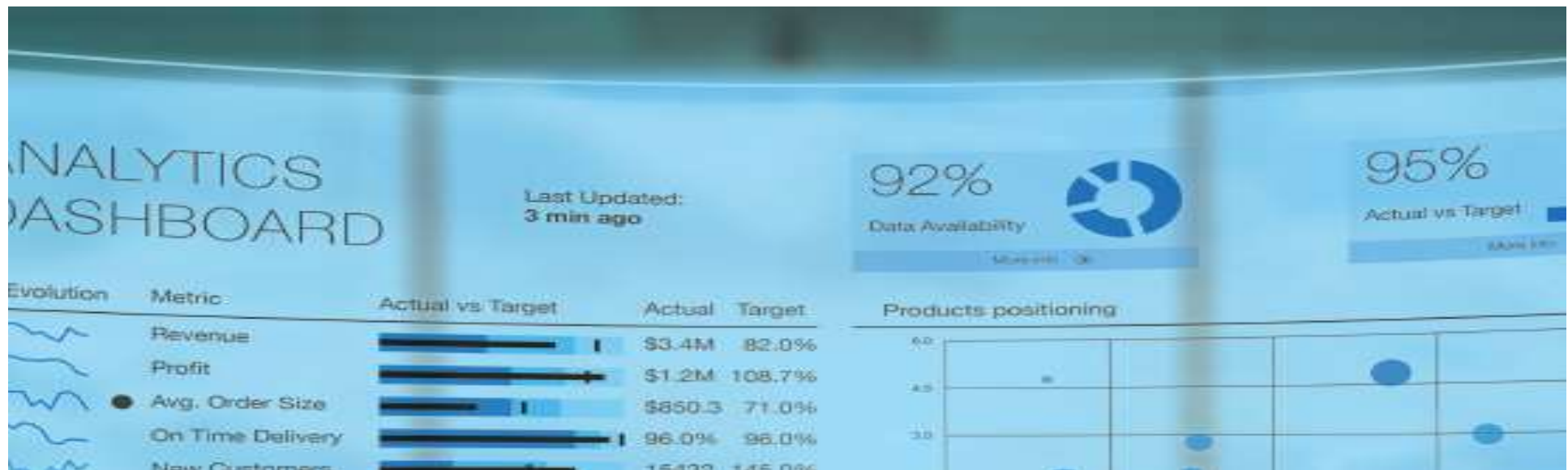
本書内容の一部、全体を問わず、株式会社 データサイエンス研究所の文書
による承諾なく引用複製する事を禁じます。

大学における統計系学部を設置数（2016年）

米国	130
中国	300
韓国	50
日本	0

＊統計学博士取得者数

米国 年間600名 日本5名程度



第四次産業革命スキル習得講座

分野別データサイエンティスト養成講座

(マーケティング、マニュファクチャリング、メディカル)

株式会社データサイエンス研究所

■ 第四次産業革命スキル習得講座（Reスキル講座）

IT・データを中心とした将来の成長が強く見込まれる分野において、社会人が高度な専門性を身に付けてキャリアアップを図る、専門的・実践的な教育訓練講座を経済産業大臣が認定する制度。（経済産業省HP）

■ 専門実践教育訓練指定講座

■ 分野別データサイエンティスト養成講座の受講対象

- ・データサイエンティストを目指す方
- ・分析実務担当者
- ・実践力・問題解決力を習得したい方

- ・分析初心者の方
- ・プログラミング未経験の方

研修室受講、WEB受講、アーカイブ受講
昼夜受講可能

■ 分野別データサイエンティスト養成講座の特徴

◎ 実践力（課題解決力）の養成



- ① データの取捨選択、特性を把握した分析手法の選択
- ② 各分析手法のしくみの理解
- ③ 分析結果を正しく解釈、分析目的との整合

分野別データの特徴に合わせたトレーニング

マーケティングデータサイエンティスト養成講座

マーケター必須のデータサイエンスを実践的・効率的に習得

マーケティング分野において用いられる解析手法のしくみ・活用方法の理解と共に、マーケティング分野の特性に合わせたデータを用いた具体的演習により、実践的解析力、問題解決力の習得を目指す。

■ 料金 50万円（税込）

■ 研修時間 54時間（4カ月）

昼間コース（10：00～17：00） 6時間× 9日 = 54時間

夜間コース（18：30～21：30） 3時間×18日 = 54時間

* オンラインによる受講が可能です。

■ 演習事例

- ・売上高を支店別に前年度実績と比較し問題点を明らかにする。
 - ・新製品の好感度について男女による違いについて比較する。
 - ・DMの送付前後による売上高の違いについて検討する。
 - ・売上実績、営業員数、広告費から、売上高を予測するモデルを構築する。
 - ・物件情報から中古マンションの価格を予測する。
 - ・製品評価に与える因子について分析する。
 - ・顧客情報から顧客離反要因について分析する。
 - ・小売店における同時購入アイテムについて分析する。
 - ・ソーシャルメディアの口コミについて分析する。
 - ・売上高の構造について年度別に比較し問題点を明らかにする。
 - ・コンサート会場アンケートから、満足度と様々な要因との関係进行分析する。
 - ・
-

■ 修了認定要件

修了認定要件

要件 1

全科目受講
(各科目 90%以上出席)



要件 2

総合テスト 70 点以上



要件 3

課題解決
プレゼンテーションに
合格



合格の場合

修了認定証を発行します。

*統計検定 2 級取得を無料サポートします。



不合格の場合

1. 受講科目不足の場合は、オンラインによるアーカイブ受講をして頂きます。
2. 総合テスト不合格の場合は、不合格分野の再受講をお勧めし、再試験を受けて頂きます。
3. 課題解決プレゼンテーションに不合格の場合は、課題レポートを提出して頂きます。
4. 総合的に審査し、合格と認められた場合は修了認定証を発行します。

【受講お申込み】

<https://datascience.co.jp/reskill/>