

IoT導入支援セミナー ～今更聞けないIoT入門～

【データ分析基礎編】

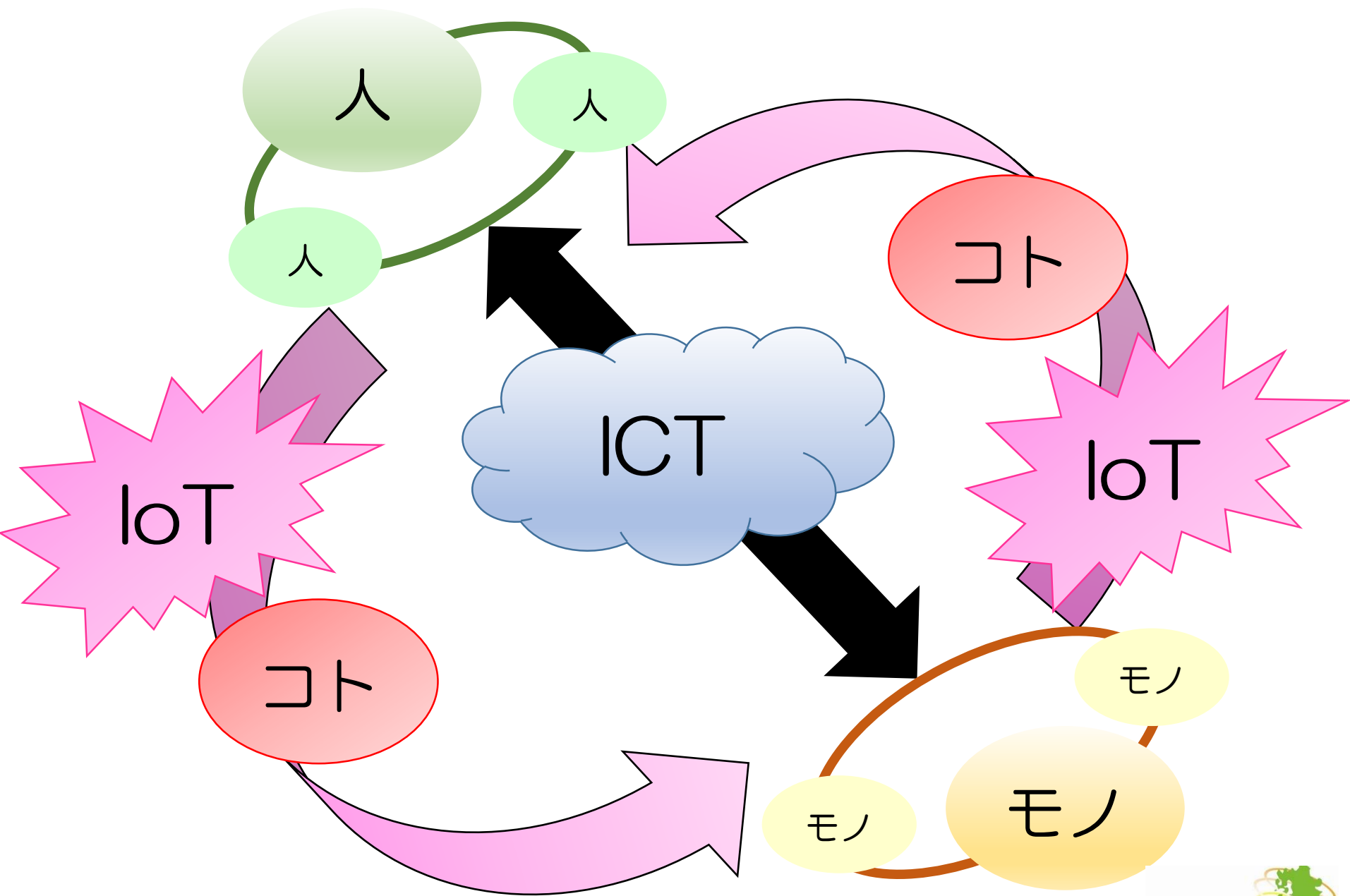
(一社)九州テレコム振興センター (K I A I)

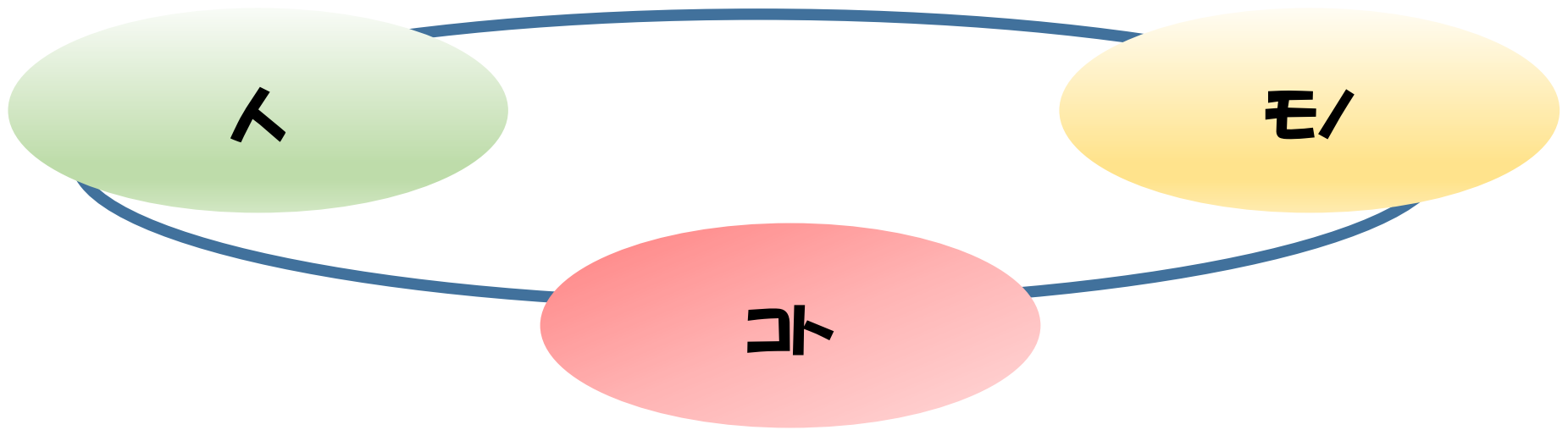
最初に突然ですが、ICT⇒IoTの
変化について少し考えてみましょう

(あくまでも私的見解ですが)

■ ICT⇒IoTとは？（私的見解その1）

人・モノ 接続の社会から
人・モノ・コト 接続の社会への変革





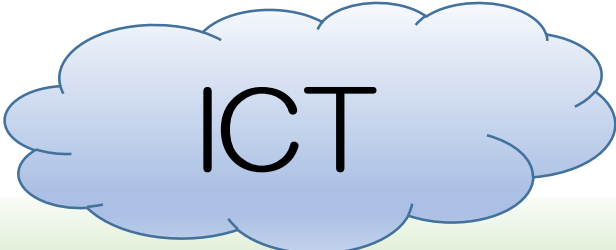
こういった3者のつながりの中
現在ネットワーク化（情報化・デジタル化）
されていない分野とは？

[ネットワーク化した際の様々な仮説を考える柔軟性も必要]

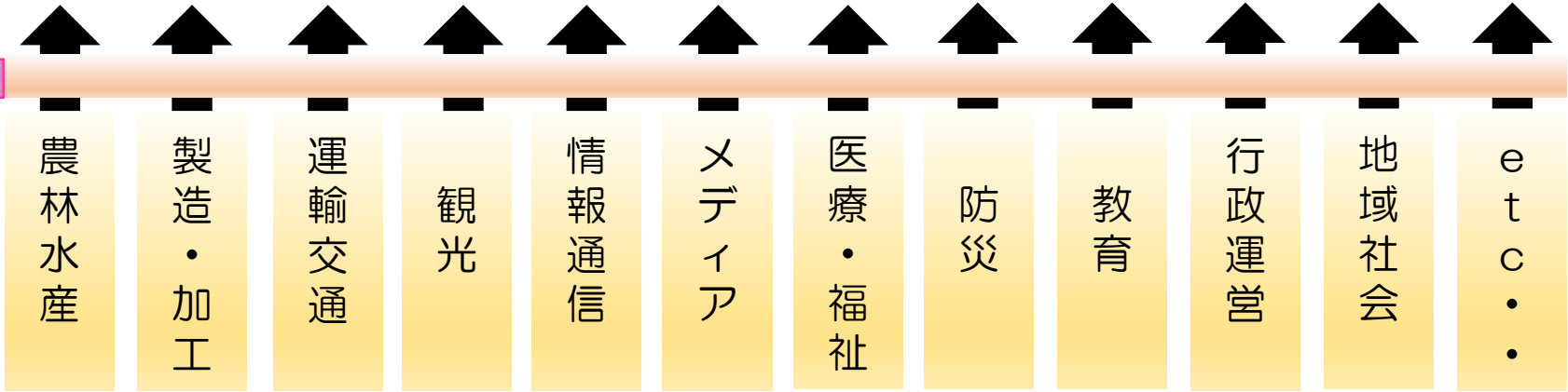
みなさんの会社（組織）の中における
人・モノ・コト
を改めて見つめ直してみると・・・

■ ICT⇒IoTとは？（私的見解その2）

**垂直型情報化社会から
水平展開的情報化社会への変革**



各分野の高度化・効率化を通じた新たな社会形成



Big Data

各分野がもたらす多様な情報orデータを総合的に活用した新たな社会形成

AI

IoT

**自社(自組織)
データ**



農林水産

製造加工

運輸交通

観光

情報通信

メディア

医療・福祉

防災

教育

行政運営

地域社会

etc...

自社の事業活動から創出（蓄積）
されているデータは果たして
自社（or同業種）に対して
のみ価値があるものなのか？

つまり、「ICT⇒IoT」における

最大のポイントとなるメインアイテムは

センサー等のIoTデバイス、それらをつなぐ通信技術も重要ですが

データ
ではないでしょうか

しかしながら・・・

以下の「3ない状態」が
未だ多く見受けられます

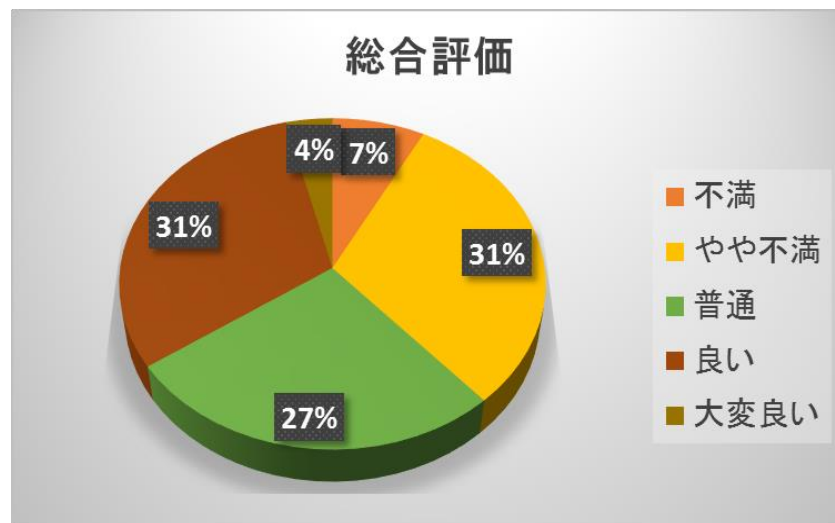
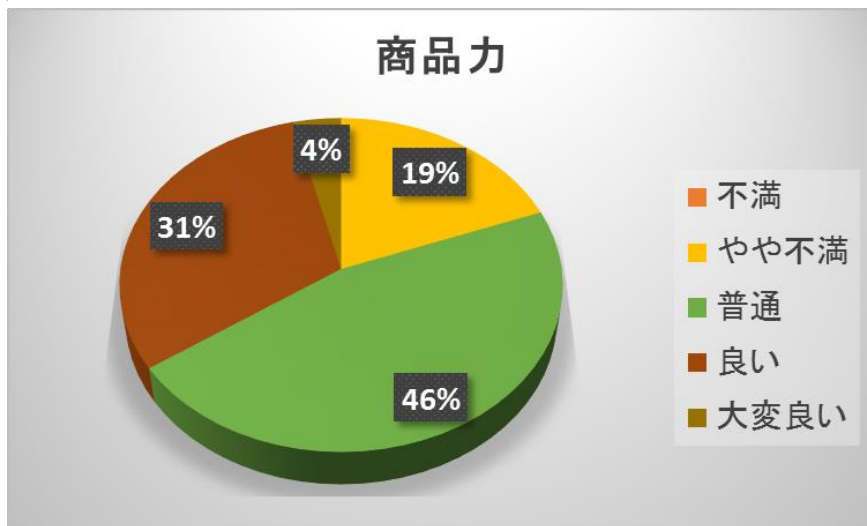
- ・データ分析に必要な基本的なプロセスが理解されていない
- ・データ分析に必要な基本的手法が理解されていない
- ・データ分析を行う必要性が理解されていない

■ こんなデータ分析で終わっていませんか？

＜アンケート調査項目例＞

項目	評価点				
	1.とても良い	2.良い	3.普通	4.やや不満	5.不満
接客	1.とても良い	2.良い	3.普通	4.やや不満	5.不満
商品力	1.とても良い	2.良い	3.普通	4.やや不満	5.不満
信頼性	1.とても良い	2.良い	3.普通	4.やや不満	5.不満
アフターサービス	1.とても良い	2.良い	3.普通	4.やや不満	5.不満
価格	1.とても良い	2.良い	3.普通	4.やや不満	5.不満
総合評価	1.とても良い	2.良い	3.普通	4.やや不満	5.不満

➡ アンケートの集計結果を単にグラフ表示しただけで終わり？（深い議論ができますか？）



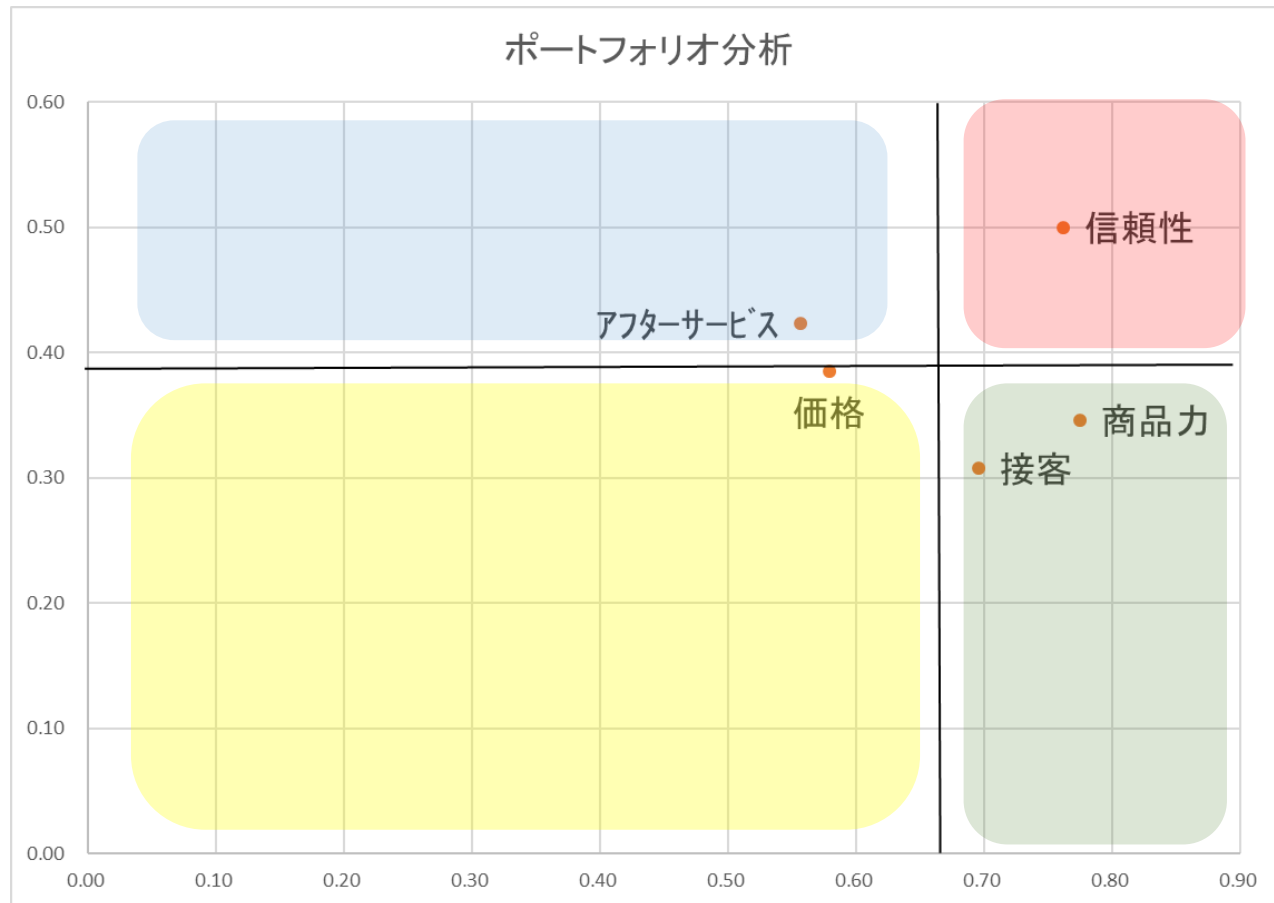
■分析スキルを少し身に着けるだけで・・・

例えば、前頁のようなアンケート調査結果に対し、データ間の関係性などを分析できる基本スキルを有しておくと、以下のような可視化分析もみなさんが普段お使いのExcel上で、簡単に行えます。（下記可視化の読み解きできますか？）

現状の組織課題
(強み・弱み)が
明確に把握できた



従来のアンケート調査結果もデータの単純なグラフ化作業で終わることなく、このようにデータとの向き合い方ひとつで、より幅広い議論ができる資料になりうるのでは？



■こういった例はわかりますか？

例1：どちらの普及促進活動が有効に機能したと推測できるか？

	普及啓発活動前	普及啓発活動後
A会場	69	70
B会場	81	82
C会場	64	66
D会場	79	81
E会場	64	68
F会場	110	128
G会場	72	73
H会場	92	91
I会場	83	83
J会場	78	79
平均来場者	79.2	82.1

	普及啓発活動前	普及啓発活動後
A会場	69	70
B会場	81	82
C会場	64	66
D会場	79	80
E会場	64	70
F会場	110	118
G会場	72	74
H会場	92	93
I会場	83	83
J会場	78	79
平均来場者	79.2	81.5

<普及啓発活動プランA実施後來場者比較>

<普及啓発活動プランB実施後來場者比較>

平均来場者数としてはプランAの方が効果があったように思えるが、本当にそうなのか・・・



例2：どの要因が各地区の購入金額に一番影響を与えているのか？
 また、各要因のデータ値から逆に購入金額の予測ができるか？

	地域世帯数	地域住民平均所得	会場面積	競合品数	購入金額計
A地区	8,850	3,100	30	3	940
B地区	3,210	1,980	31	1	479
C地区	5,890	2,700	35	2	868
D地区	1,800	1,850	24	3	359
E地区	10,020	2,980	28	4	898
F地区	4,650	2,300	24	0	795
G地区	7,820	2,050	28	2	738
H地区	2,830	2,100	20	1	404
I地区	9,320	3,000	32	3	980
J地区	3,840	1,990	32	3	325

正直、どの要因が一番影響を及ぼしているのか、データを見ただけでは判別できない・・・。



1. 課題の洗い出しについて

(1) そもそも課題とは何でしょうか？

課題が見つからない？
課題がわからない？

問題



課題

課題とは問題を解消させるために解決すべき事項

そもそも
問題が把握できてない？
問題が発掘されていない？

(2) 改めて『問題』という点について

問題を「感じる」とこと
「認識する」ことは別モノ

曖昧な問題認識の
ままでは課題自体も
曖昧なまま

改めて
「問題」って
どういうこと
なのか



あるべき姿

問題

一般的に この両者の
ギャップから生じるものが
「問題」

(あくまでもビジネスシーンに特化)

現実の姿

(3) 『問題』を正しく認識する、ということは

**あるべき姿と現実の姿がきちんと
認識されることが前提**

しかしながら・・・

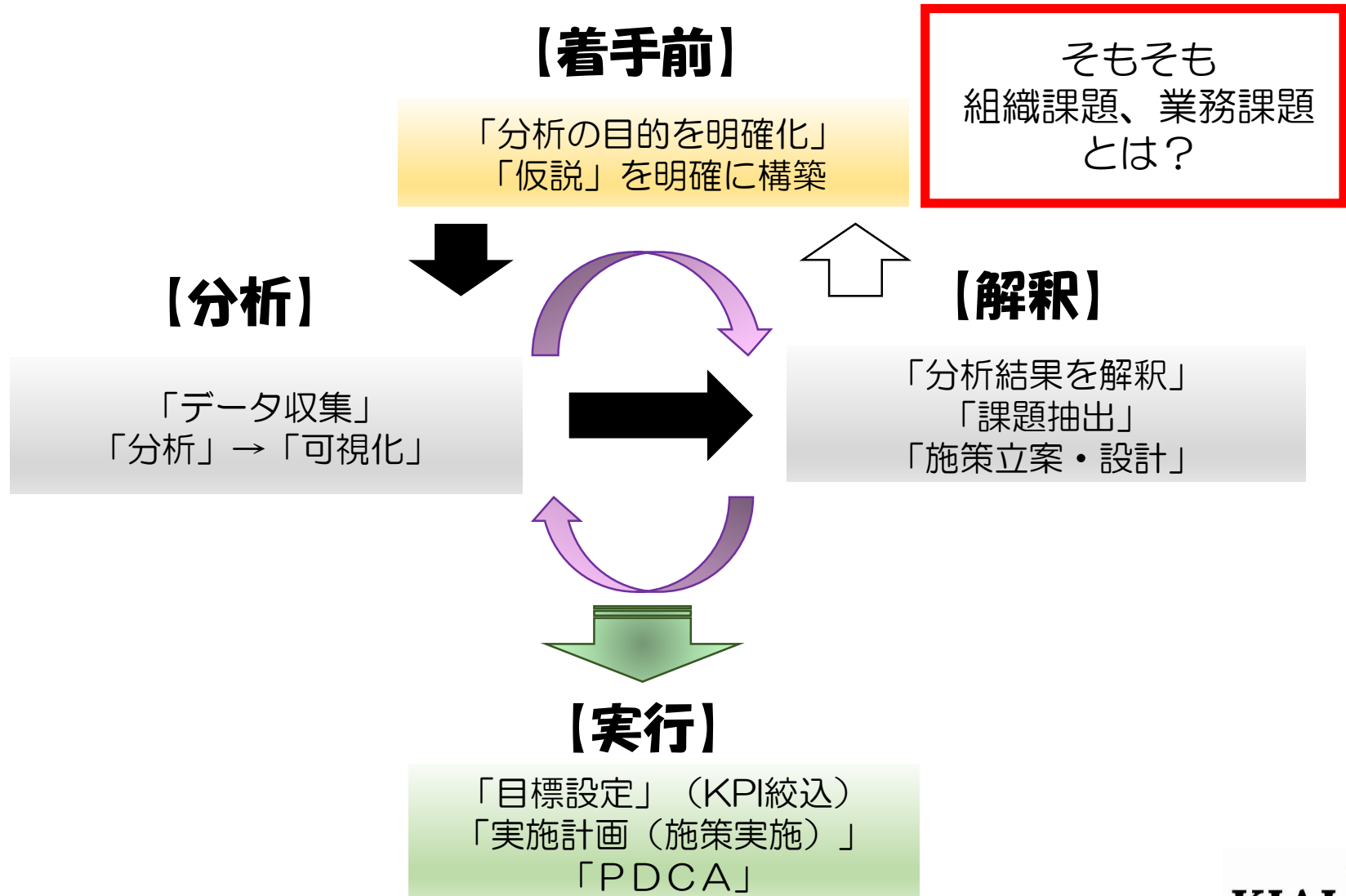
あるべき姿／現実の姿は各員それぞれ基本的に一致しないことが多い。同じ職場、同じ業務に携わっているからといってあるべき姿／現実の姿は同じ内容で認識されていない。それぞれの姿が異なるため、問題意識のズレが発生する。

課題を見出す第一歩とは、
あるべき姿／現実の姿を
関係者一同で明確化し、
共通認識する、
ということから着手すべき。



2. データ分析の基本について

(1) データ分析・活用の基本的な流れとは



(2) 習得が望ましい基礎的分析手法

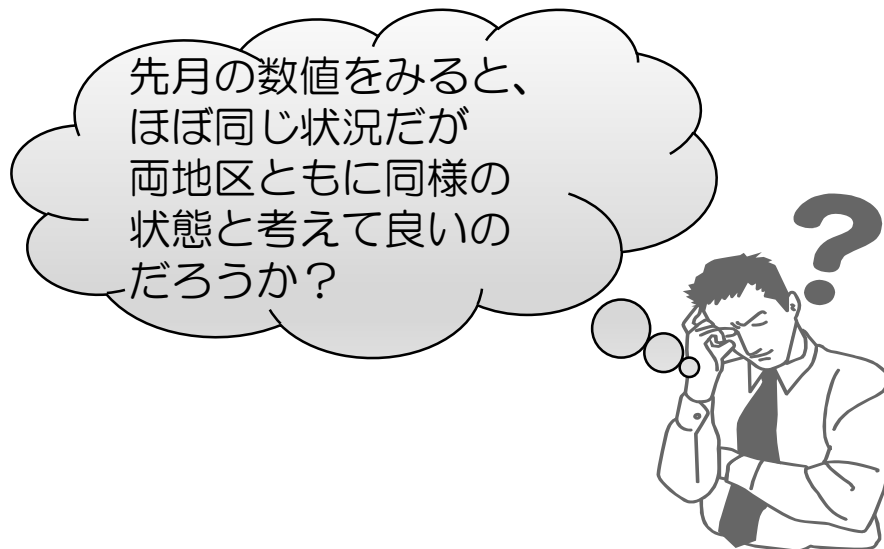
<p>■データの全体像を掌握する 平均値や合計値だけに頼らず、データ全体を幅広い視野から分析</p>	<ul style="list-style-type: none"> 標準偏差 基準化 データ全体の可視化 (ヒストグラム、散布図作成) 	<p>普段お使用の汎用ビジネスツール「Excel」で十分に分析可能です。</p>
<p>■データ間の関連性を掌握する データ間に潜む関連性を分析</p>	<ul style="list-style-type: none"> 相関分析 回帰分析 	
<p>■過去データから将来を予測する 過去データから将来の事業計画等をシミュレーション</p>	<ul style="list-style-type: none"> 回帰分析 	
<p>■平均値の違いを客観的に評価する 違いはたまたまの誤差なのか、有意的な違いなのかを分析</p>	<ul style="list-style-type: none"> 平均値検定 (t検定) 	
<p>■現状の強み、弱みを掌握する 着手すべき課題の優先順位を明確化</p>	<ul style="list-style-type: none"> ポートフォリオ分析 	

本日は限られた時間ですので、上記の中から活用頻度も比較的高い「標準偏差」「相関」の2点についてのみ、Excelを活用した具体的な分析手法を簡単にご説明いたします。

3. Excelを活用した具体的データ分析例

(1) 標準偏差

	〇〇地区	□□地区
1日	105	120
2日	103	94
3日	112	65
4日	108	122
5日	86	114
∫	∫	∫
27日	92	88
28日	98	117
29日	103	132
30日	83	92
合計	3,046	3,037
平均	102	101



集計したデータを合計値と平均値だけで分析していませんか？

データ全体のバラツキという概念を取り入れてみませんか

標準偏差の基本を理解

	売上 (万円)	偏差	偏差2乗
1	49	-1	1
2	51	1	1
3	57	7	49
4	49	-1	1
5	47	-3	9
6	52	2	4
7	48	-2	4
8	49	-1	1
9	49	-1	1
10	43	-7	49
11	53	3	9
12	54	4	16
13	48	-2	4
14	51	1	1
平均	50	0	150

『偏差』 『分散』 『標準偏差』
という順番で理解

◆偏差：データから平均値を引いた値

偏差の合計は0にしかならない。
偏差自体はデータ全体の指標としては
使いにくい

偏差の値を2乗して足し合わせていけば
0にはならないので、一定の指標として
活用できそう

◆分散：偏差の2乗の総和をデータ数
で割った値

(左図例： $150 \div 14 \div 10.7$)

分散はバラツキの指標としては
使えるものだが2乗した段階で
単位 (左図例：万円) が用いられず
実務上使いにくい

◆標準偏差：分散の平方根

(左図例： $\sqrt{10.7}$)

$\div 3.27$ 万円)

標準偏差

データ全体として平均値からどの程度のバラツキが生じているのか？

具体的算出自体はExcelの関数で簡単にできます。

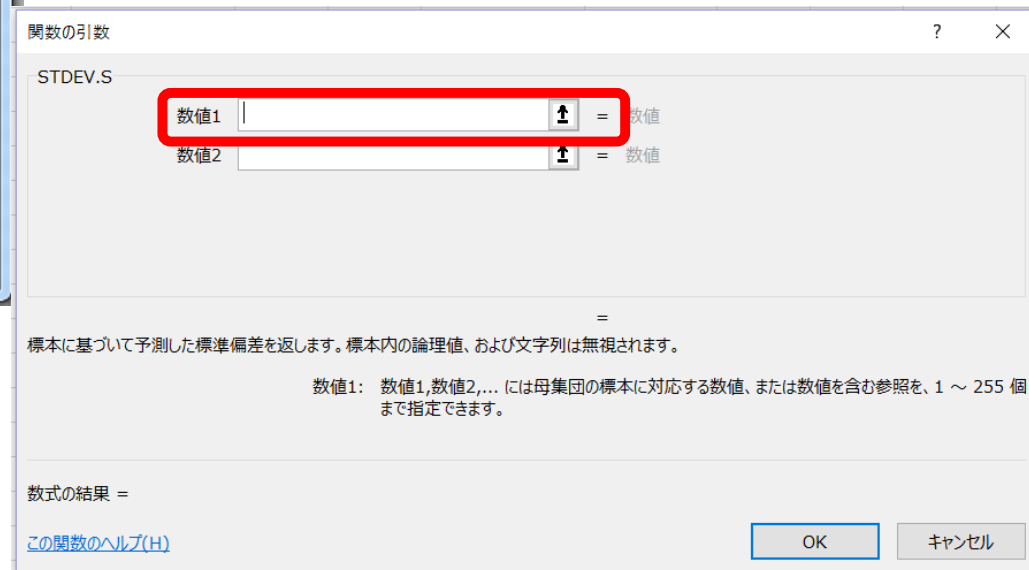


関数の分類は統計を選択し、関数自体はSTDEV.Sを使います。

※なおSTDEV.Sで算出される値は前頁での計算方法で算出される値とは若干異なるものになりますが、基本的にビジネス上で標準偏差を求める場合には、STDEV.Sを使う方が良いです。

数値1に算出したいデータ群を選択すればOKです。

(前頁の例では1~14までの売上)



さきほどの例で標準偏差をとってみると・・・

	〇〇地区	□□地区
1日	105	120
2日	103	94
3日	112	65
4日	108	122
5日	86	114
∫	∫	∫
27日	92	88
28日	98	117
29日	103	132
30日	83	92
合計	3,046	3,037
平均	102	101
標準偏差	10.3	20.8

〇〇地区と□□地区とでは、平均値は同様だが、全体のバラツキに2倍も差があることがわかった！
□□地区では何か問題が発生していないだろうか



データのバラツキ、という指標を取り入れることで課題解決に向けた新たなアプローチのきっかけが生まれてくるのではないのでしょうか

時間の都合上、本日は詳しくご説明できませんが、平均値が大きく異なるデータ間の標準偏差を比較する際は、変動係数（標準偏差／平均）という数値を活用してください。

参考：基準化（データ全体像を知る際にはこの手法も有益です）

- 異なる平均値間の個々のデータの客観比較
- 尺度や内容の異なるデータの客観的な総合評価

本日はお時間の都合上、ご説明は省略しますが・・・

支店Aと支店Bの実績をどう判断？

	支店A	支店B
第一四半期	800	380
第二四半期	580	310
第三四半期	640	370
第四四半期	680	320
平均	675	345
標準偏差	93	35

優良顧客のランキングをどう判断？

顧客名	来店回数	買上金額
A	10	220,000
B	7	146,000
C	6	166,000
D	13	249,000
E	5	230,000
F	13	50,000
G	12	267,000
H	9	72,000
I	1	77,000
J	13	100,000
平均	9	157,700
標準偏差	4	80,541

それぞれ平均値の異なる各要因を総合した評価はどう判断？

	デザイン	質感	基本スペック	耐久性	価格
1	7.46	7.58	7.93	6.78	6.54
2	7.63	5.42	7.93	6.88	6.47
3	7	5.92	8	6.75	6.42
4	7.33	4.6	7.16	6.96	6.51
5	7.48	6.08	7.48	6.96	6.48
6	7.35	6.35	7.95	6.77	6.6
7	7.38	6	7.97	7.07	6.57
8	7.17	5.61	7.91	6.13	6.3
9	7.17	6.3	8.22	6.76	6.26
10	6.75	5.72	6.44	6.36	6.25
平均	7.27	5.96	7.70	6.74	6.44
標準偏差	0.26	0.76	0.53	0.29	0.13

(2) 相関分析

- データ間の関係の強さを表す指標のひとつ
(人の思い込み等からではなく、客観的な数値で示す)
- 相関分析の手法
 - Excel等で相関係数を算出
 - 散布図を作成←これも非常に大切
- 相関係数
 - -1から1の間
 - マイナスの場合は負の相関 (一方が高くなると一方は低くなる)
 - プラスの場合は正の相関 (一方が高くなると一方も高くなる)
 - データA⇔データBの双方向の関係性

相関分析注意点①

- 相関関係とは、あくまでも変数間に起こる変動を捉えた統計上の指標
- 相関関係があるからといって、必ずしも因果関係があるとはいえない

ひょっとすると以下のデータを集めたら正の相関がみられるかもしれませんが、年収が高くなることと、血圧が高くなることは因果関係のあることなのでしょうか？

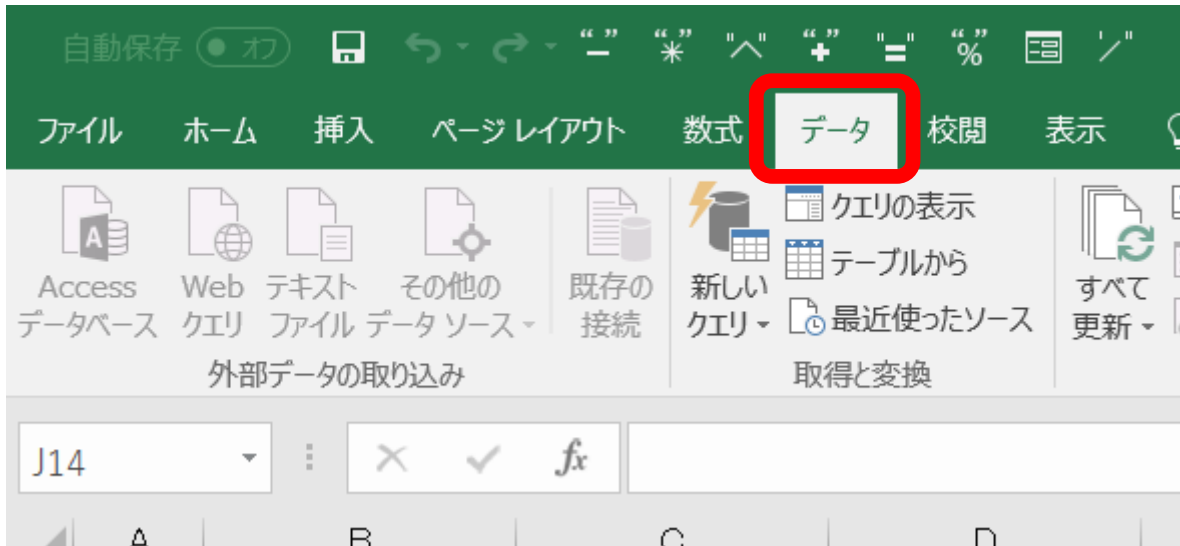
年収

血圧数値

相関分析サンプル例

	10代～ 20代	30代～ 40代	40代～ 50代	50代 以上	サービスA	サービスB	サービスC	サービスD
1	35	62	42	48	35	61	48	44
2	20	70	20	55	28	35	52	49
3	25	81	38	69	36	59	58	50
4	55	68	48	59	43	69	40	41
5	60	82	70	60	53	82	58	51
6	85	58	22	47	41	40	55	37
∫	∫	∫	∫	∫	∫	∫	∫	∫
24	83	91	77	69	63	85	61	59
25	20	121	42	71	46	62	22	58
26	66	128	39	74	58	61	58	64
27	78	80	66	75	56	78	62	49
28	120	89	34	79	61	66	60	52
29	88	112	79	80	70	90	62	59
30	90	128	83	81	75	98	59	62

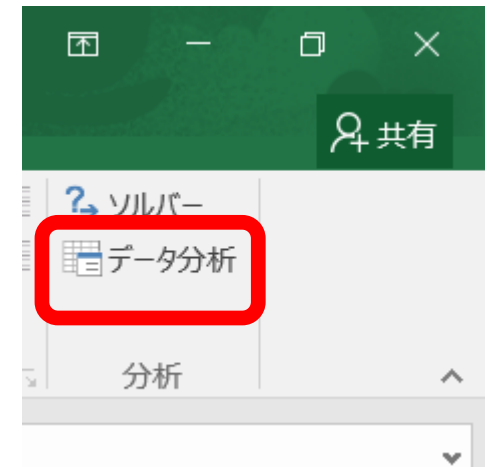
Excelのデータ分析ツールはアドインしていますか？



データタブを選択した際にメニュー右端に「データ分析」というタブが出てくれば、データ分析ツールのアドインがされています。

この「データ分析」が表示されなければ、別途アドインする必要があります。

Excelに標準搭載されている便利な機能ですので、アドインされていない方は一度アドインをし、実際にお使いになってはいかがでしょうか。





情報

新規

開く

上書き保存

名前を付けて
保存

印刷

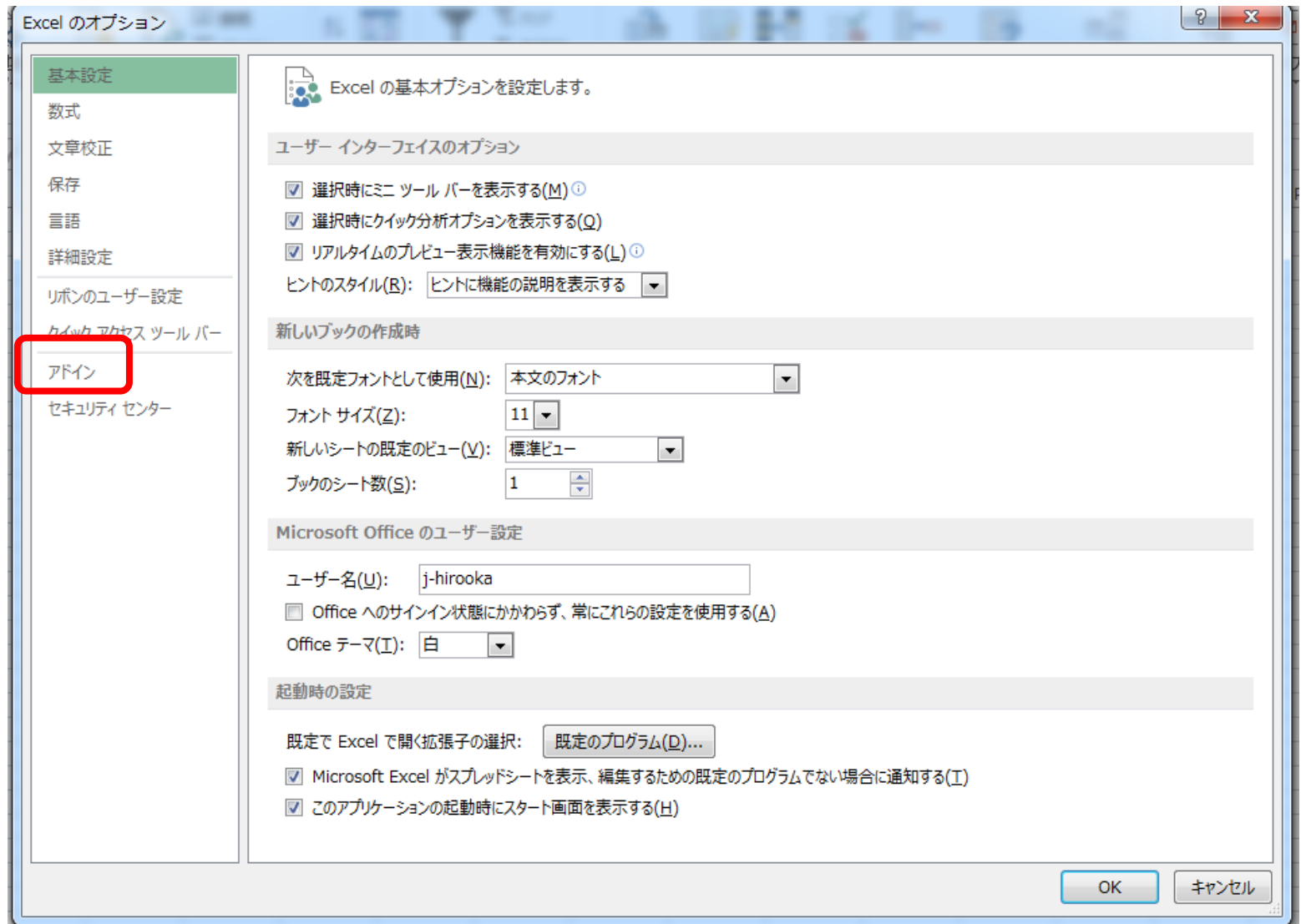
共有

エクスポート

閉じる

アカウント

オプション



ファイル画面で、オプション→アドインを選択

基本設定

数式

文章校正

保存

言語

詳細設定

リボンのユーザー設定

クイック アクセス ツール バー

アドイン

セキュリティセンター



Microsoft Office のアドインの表示と管理を行います。

アドイン

名前 ▲ 場

アクティブなアプリケーション アドイン

ソルバー アドイン

分析ツール

アクティブでないアプリケーション アドイン

Euro Currency Tools

Microsoft Actions Pane 3

Microsoft Power Map for Excel

日付 (XML)

分析ツール - VBA

ドキュメント関連アドイン

ドキュメント関連アドインはありません

無効なアプリケーション アドイン

無効なアプリケーション アドインはありません

アドイン: ソルバー アドイン

発行者:

互換性: 互換性に関する情報はありません

場所: C:\Program Files (x86)\Microsoft Office\root\O

説明: 最適化に関する数学的な手法を用いて、指定された範囲で

管理(A): Excel アドイン

設定(S)...

アドイン

? X

有効なアドイン(A):

 Euro Currency Tools ソルバー アドイン 分析ツール 分析ツール - VBA

OK

キャンセル

参照(B)...

オートメーション(U)...

分析ツール

統計学的および工学的分析を行うためのデータ分析ツールです

OK

キャンセル

Excelアドイン設定→分析ツールにチェックしてOK

Copyright (C) 2017 Kyushu Island Alliance of ICT. all rights reserved.

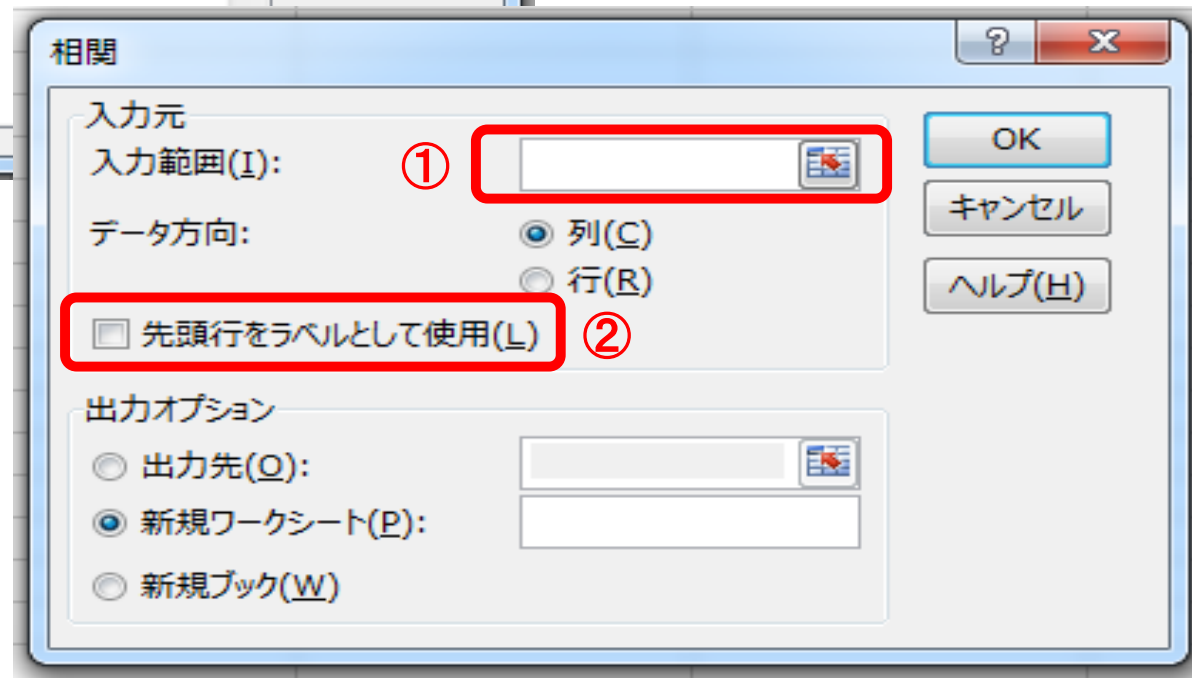
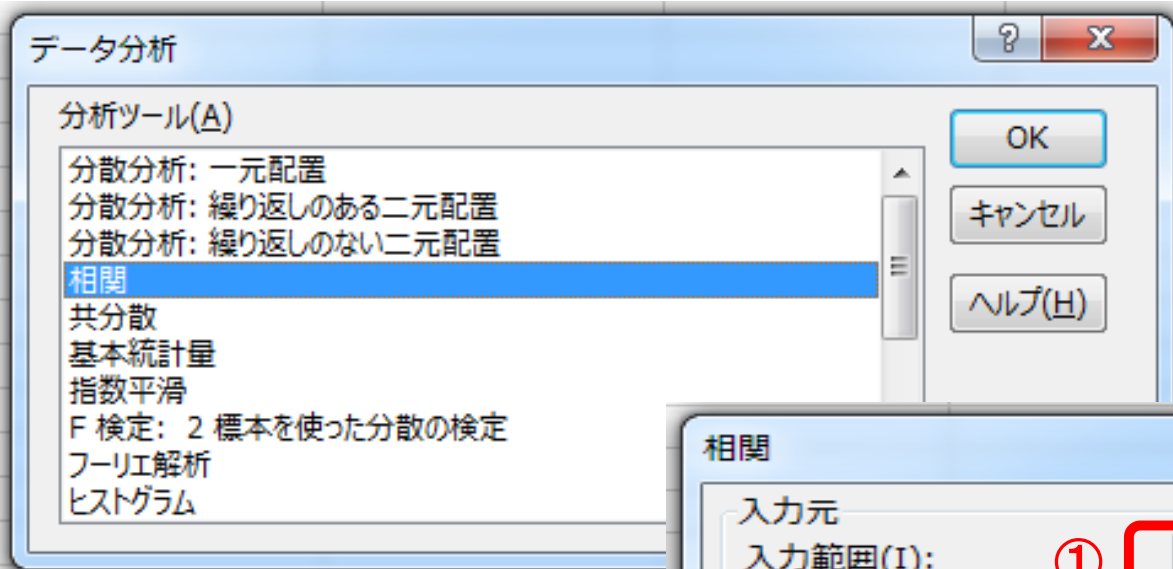
KIAI

Kyushu Island Alliance of ICT





データ分析を選択すると標準搭載されている様々な分析ツールが表示されるのでその中から相関を選択



①相関係数を算出したい対象データを選択
(前頁の例では10代~20代からサービスDまでの全データ)

②関数と異なり分析ツールでは先頭行の日本語タイトルを認識できるので、ラベルとして使用をチェック
(日本語行がなければ不要)

さきほどの例で相関分析を行ってみると・・・

	10代～20代	30代～40代	50代以上	サービスA	サービスB	サービスC	サービスD	サービスE
10代～20代	1							
30代～40代	0.00179627	1						
50代以上	-0.11281081	-0.00265079	1					
サービスA	0.074881149	0.832035848	0.142314791	1				
サービスB	0.518038677	0.510223326	0.606100821	0.56863546	1			
サービスC	0.234293137	0.318834025	0.787528451	0.433708478	0.851898538	1		
サービスD	0.637685977	0.027879519	0.102324658	0.076384839	0.441367848	0.279063068	1	
サービスE	0.058503753	0.777573194	0.07747135	0.693060307	0.469281662	0.330585437	0.222550573	1

分析ツールだけでなく、単純に関数を用いても相関係数の算出は可能です。関数：CORREL

10代～20代と関連性あるサービスは ()
 50代以上と関連性あるサービスは ()
 サービスAと関連性高いのはサービス ()



単純なグラフ表示等だけではつかみにくいデータ間の関係性を掌握することで様々な課題抽出、事業企画等に役立つのでは

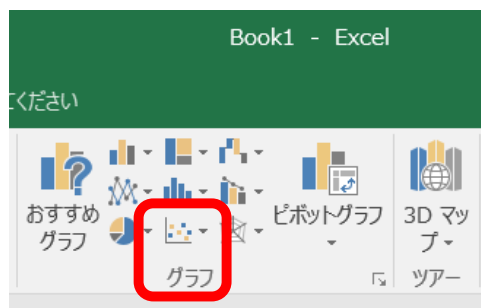
相関分析注意点②

「見せかけの相関」に注意

- 相関係数だけを求めるのは危険
- 必ず散布図を描くクセをつける

※ 散布図はデータの可視化においても大変有用なツール

※ これまで散布図をあまり活用したことがない方は、今後、是非活用してみてください。Excelのグラフ作成機能の中に入っています。

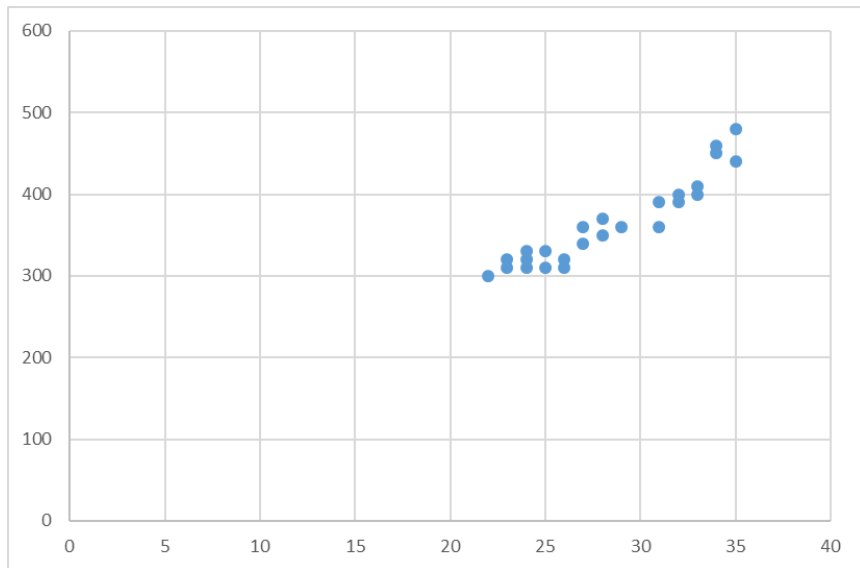


散布図の種類をきいてくるので標準的な散布図を指定（作成自体はその他棒グラフや円グラフ等と同様）



グラフ作成のこのスタイルを選択すると

- ◆ 外れ値の存在がいたずらに相関係数を高く（低く）している場合がある



左図：相関係数0.94の散布図

正の相関：右肩上がり

負の相関：右肩下がり

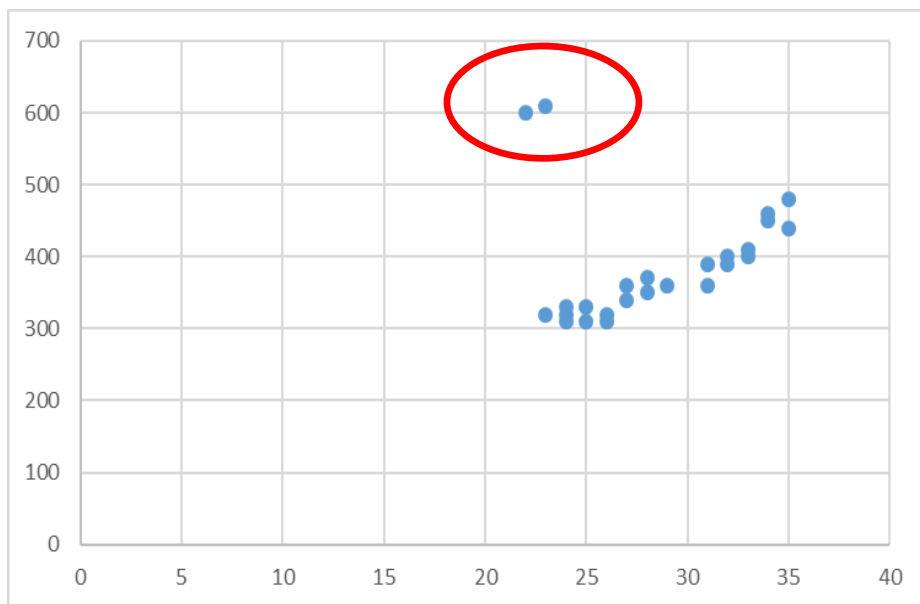
相関強い：データの密着度が高い

相関弱い：データの密着度が低い

右図：相関係数0.17の散布図

データ全体として相関がないといえるのか？

外れ値の存在が全体の相関係数を低くしているが、データ全体の傾向としては非常に高い相関があるデータ群と推測できる



大切なことは・・・

分析業務自体に重きを置くのではなく
分析結果に基づく



さいごに

本日はデータ分析のほんの一部だけを
簡単にご紹介しました。

先に述べたようにまずは最低限、習得して
おくと望ましい手法がまだまだあります。

これを機会に一人でも（一機関でも）多く
の方々がデータ分析に対しより具体的に
関心をお持ちいただき、従来以上にデータとの
向き合い方を見つめ直していただければ
と思っております。

(一社)九州テレコム振興センター(KIAI)
TEL:096-322-0120 Email:jim@kiai.gr.jp
<https://www.kiai.gr.jp/>