



はじめてのデータリテラシー実践

**THE
POWER
TO KNOW®**

CUED2J

はじめての データリテラシー実践

コースノート

本コースは米国コースノート Data Literacy in Practice (コースコード EDLESS2 : EDLESS2)を元に作成しています。

目次

- Module 1. 概要..... 2
- Module 2. データ分析の準備..... 8
- Module 3. 探索的データ分析.....25
- Module 4. 可視化でデータ関係の調査.....50
- Module 5. 文脈化とコミュニケーション.....75
- プレゼンテーション.....96

はじめての データリテラシー実践

このコースでは、データを探索し可視化するための実践的なスキルを学びます。小規模なビジネスオーナーがデータを活用して企業のパフォーマンスを向上させる過程を通じて、実際に役立つデータリテラシーのスキルを学べます。実際の状況を通じて、学んだ知識を実行可能な戦略に結びつけることができます。

データリテラシーの習得レベルに関係なく、どなたでもご参加いただけます。専門的な数学用語にとらわれず、わかりやすい概念や実践的な内容に重点を置いているため、学習者はコース全体を通して、安心して前向きに学び進めることができます。

1

Copyright © SAS Institute Inc. All rights reserved.



学習目標

- 分析のためのデータを準備する
 - 探索的なデータ分析を実施する
 - 可視化によるデータ内の関係性を調査する
 - 分析結果の文脈化と伝達する
-
- 小規模なビジネスオーナーがデータを活用して企業のパフォーマンスを向上させる過程を通じて、実際に役立つデータリテラシーのスキルを学べます。実際の状況を通じて、学んだ知識を実行可能な戦略に結びつけられます。

講師のPPTによる講義＋アクティビティ(対面／オンライン)＋ビデオ視聴

2

Copyright © SAS Institute Inc. All rights reserved.

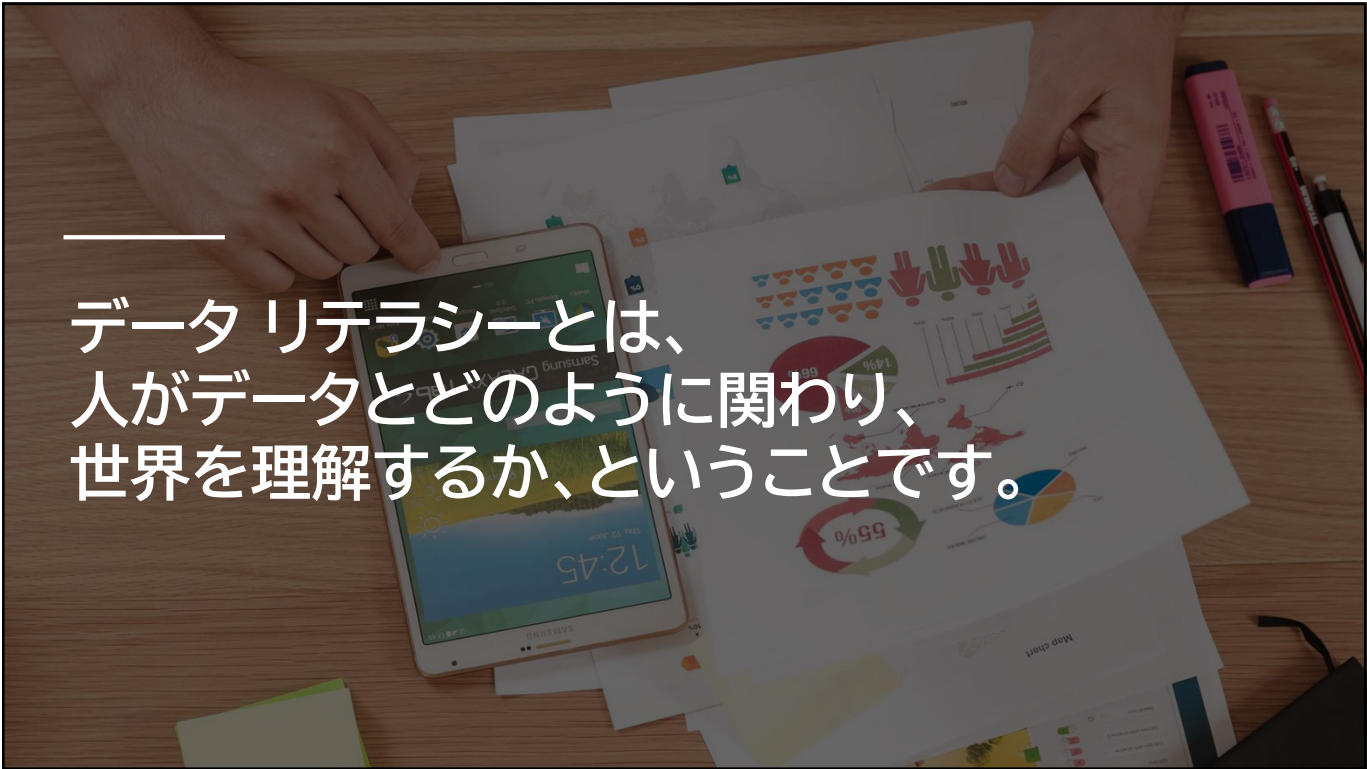


目次

- Module 1: 概要
- Module 2: データ分析の準備
- Module 3: 探索的データ分析
- Module 4: 可視化によるデータ関係性の調査
- Module 5: 文脈化と伝達
- 理解テスト

Module 1: 概要

1. データリテラシーの実践へようこそ
2. データリテラシーとのつながり
3. ビジネスシナリオ: 自動車販売



データ リテラシーとは、
人がデータとどのように関わり、
世界を理解するか、ということです。



データ



データリテラシがある、
ということは何のことでしょう？

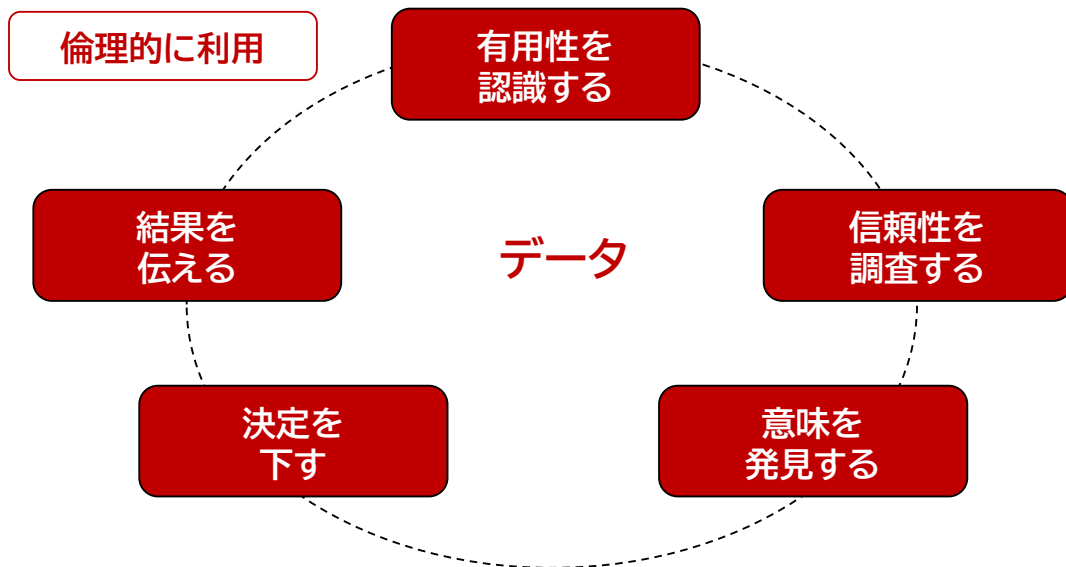


日常のデータを理解する

情報

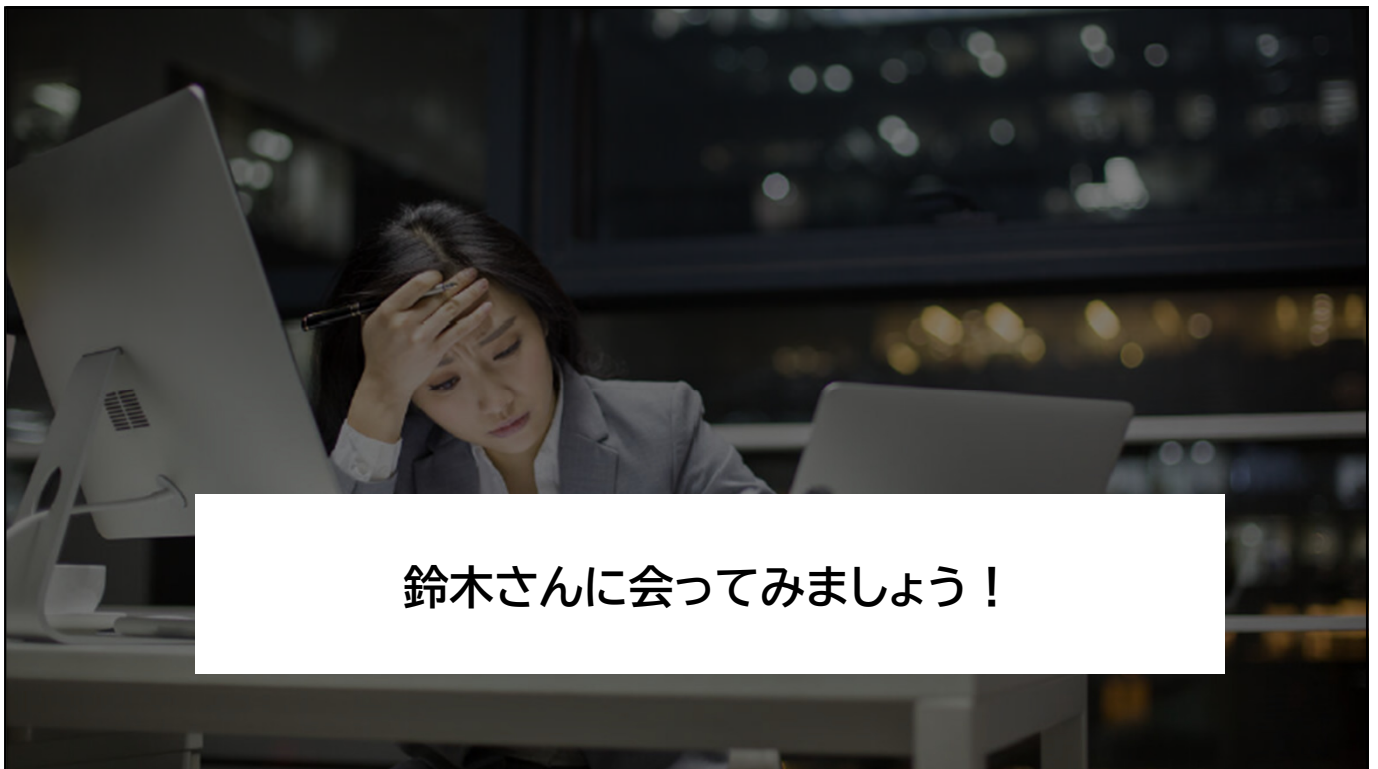


データリテラシーのスキル



9

Copyright © SAS Institute Inc. All rights reserved.



鈴木さんに会ってみましょう！

鈴木さん

中古車販売店、渚の杜自動車(Shore Globe Auto)のオーナー。エリア内の様々な地区に5つの店舗を構えており、順調に経営中。しかし、諸経費が増加、また来年にはいくつかの施設の修理を予定していることを懸念している。

彼女は、自社の「全体像」の数字、たとえば年間利益を把握している。しかし、細かいレベルで何が起きているのかを知らない。たとえば、5つの店舗のうちどの店舗が最も利益を上げているかさえない。

昨年、鈴木さんは各販売店にすべての取引に関するデータを収集するように依頼し、そのデータを調べて自分のビジネスについてさらに学ぶことに興奮しており、ビジネス上の意思決定を下すのに役立つことを期待している。

全体像: 事実

売上: 鈴木さんのビジネスモデルは、車を安く仕入れし、それを他の人に売って利益を得るというもの。彼女が昨年、自動車販売で稼いだお金の合計は、

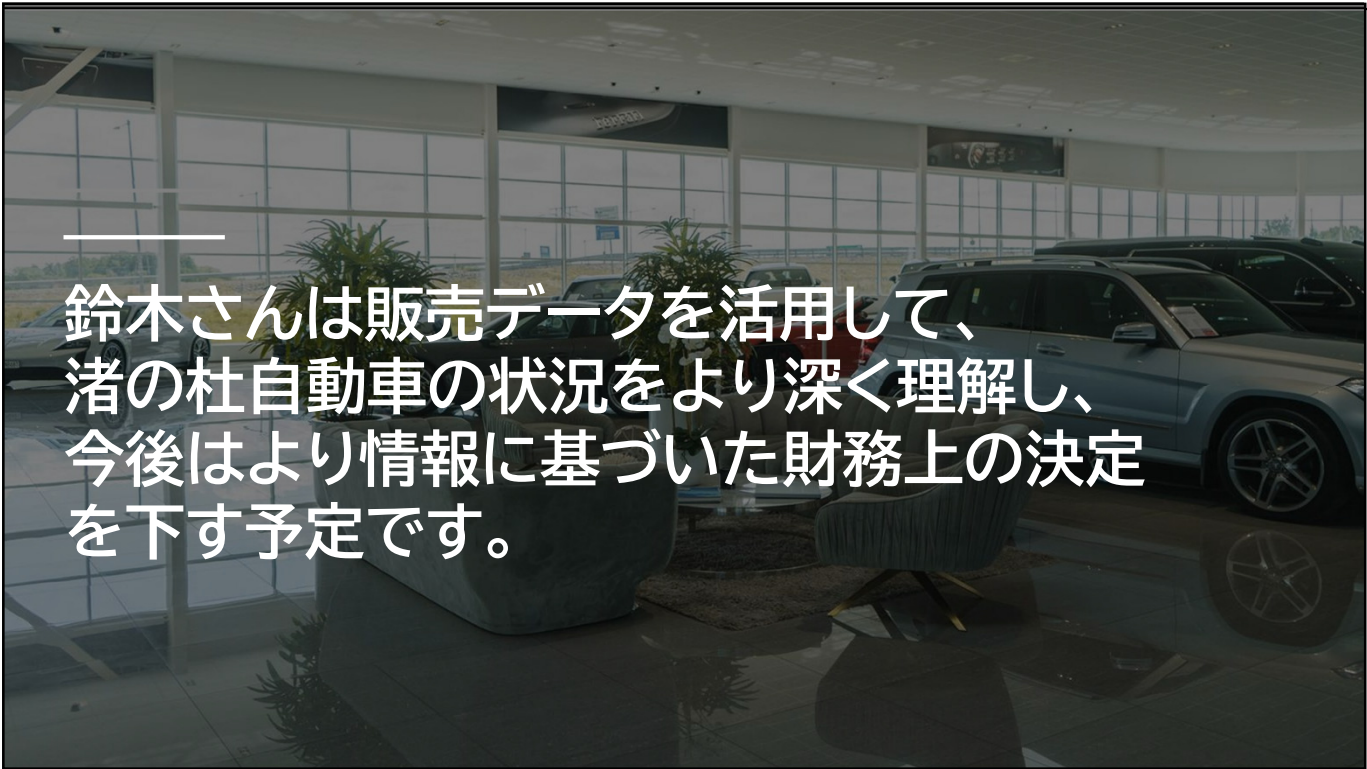
売上: 3,233百万円

経費: 仕入費用、光熱費、ライセンス、従業員の給与と手当、保険、在庫、税金など、かなりの経費がかかる。鈴木さんが昨年、事業を運営するために費やした金額の合計は、

経費: 2,599百万円

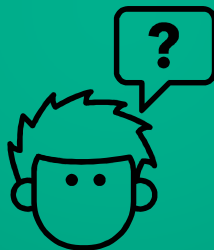
利益: 鈴木さんの経費を売上から差し引くと、残った金額が利益になる。

利益: $3,233 - 2,599 = 634$ 百万円



鈴木さんは販売データを活用して、
渚の杜自動車の状況をより深く理解し、
今後はより情報に基づいた財務上の決定
を下す予定です。

質疑応答



Module 2: データ分析の準備

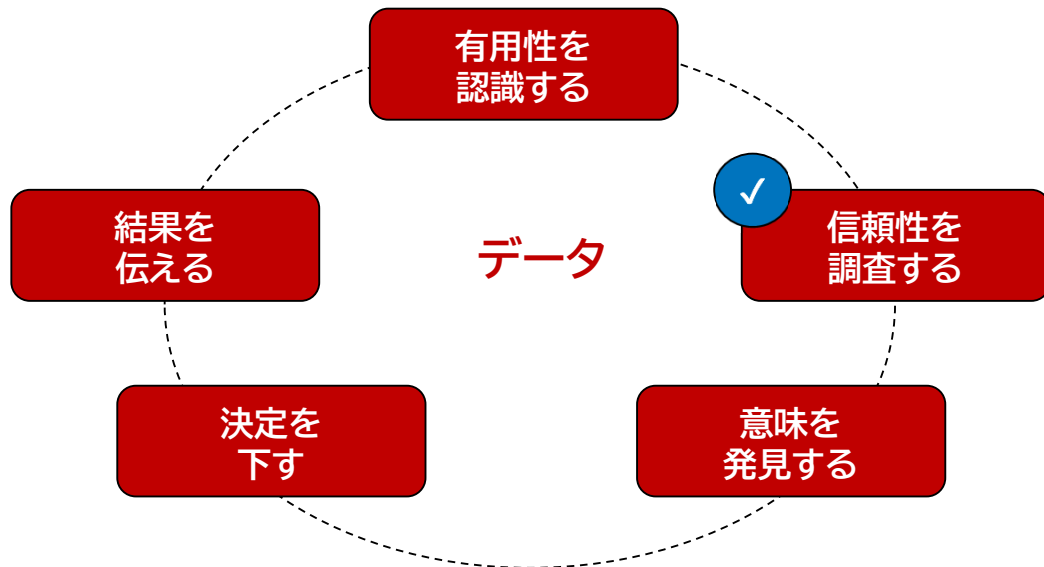
1. データ・リテラシーとのつながり
2. 自動車販売データ
3. データを知る
4. データのクリーニング
5. 質問をする
6. 分析のための整理整頓



Copyright © SAS Institute Inc. All rights reserved.

分析結果の質は、
データの質によって決まります。

データリテラシーのスキル



3

Copyright © SAS Institute Inc. All rights reserved.



鈴木さんは渚の杜自動車のデータを探求することに非常に興奮している。

鈴木さんは各販売店のマネージャーに、翌年を通しての個別の販売データを収集するよう依頼した。

渚の杜自動車販売店 – 2021年

	A	B	C	D	E	F	G
1	Sale ID	Date of Sale	Buy Price	Sale Price	Profit	Time of Sale	Dealership Name
2	10218	2021/1/1	3,223,333	4,138,788	915,455	10:48 AM	池袋（豊島区）
3	10219	2021/1/1	3,064,545	3,703,182	638,636	4:47 PM	池袋（豊島区）
4	10658	2021/1/1				2:27 PM	三鷹市
5	10794	2021/1/1	3,057,273	3,289,697	232,424	10:37 AM	中野区
6	10795	2021/1/1	2,604,545	3,128,030	523,485	11:59 AM	中野区
7	10220	2021/1/2	3,289,545	3,759,848	470,303	11:31 AM	池袋（豊島区）
8	10659	2021/1/2	3,120,909	3,788,182	667,273	10:51 AM	三鷹市
9	10796	2021/1/2		3,683,182		8:44 AM	中野区
10	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	渋谷区
11	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	
12	10221	2021/1/3	2,833,788	3,483,939	650,152	12:47 PM	池袋（豊島区）
13	10222	2021/1/3	317万円			5:49 PM	池袋（豊島区）
14	10797	2021/1/3	2,736,061	3,694,242	958,182	2:24 PM	中野区
15	10223	1月4日	2,441,818	2,870,606	428,788	3:08 PM	池袋（豊島区）
16	10660	2024/1/4	2,705,455	3,264,091	558,636	2:18 PM	三鷹市
17							
18	10418	2021/1/5	2,646,061	3,592,576	946,515	4:30 PM	新宿区
19	10002	2021/1/6	3,645,000	4,481,818	836,818	6:14 PM	渋谷区
20	10224	2021/1/6	3,135,606	4,000,303	864,697	1:10 PM	池袋（豊島区）
21	10419	2021/1/6	1,519,545	2,347,879	828,333	4:45 PM	新宿区
22	10420	2021/1/6	1,940,909	2,618,939	678,030	6:45 PM	新宿区
23	10421	2021/1/7	2,670,152	3,649,242	979,091	4:10 PM	新宿区
24	10661	2021/1/7	3,047,121	3,540,303	493,182	11:58 AM	三鷹市
25	10799	2021/1/7	2,497,121	3,042,273	545,152	6:46 PM	中野区
26	10003	2021/1/8	1,869,242	2,464,545	595,303	9:18 AM	渋谷区



データ準備の最初のステップは、データセットに慣れることである。

マネジャーがどのデータを収集するかは彼女が決めたが、何がそこにあるのかを確認するためにデータを見直す時間は必要である。

データセットの変数を理解する

	A	B	C	D	E	F	G
1	Sale ID	Date of Sale	Buy Price	Sale Price	Profit	Time of Sale	Dealership Name
2	10218	2021/1/1	3,223,333	4,138,788	915,455	10:48 AM	池袋（豊島区）
3	10219	2021/1/1	3,064,545	3,703,182	638,636	4:47 PM	池袋（豊島区）
4	10658	2021/1/1				2:27 PM	三鷹市
5	10794	2021/1/1	3,057,273	3,289,697	232,424	10:37 AM	中野区
6	10795	2021/1/1	2,604,545	3,128,030	523,485	11:59 AM	中野区
7	10220	2021/1/2	3,289,545	3,759,848	470,303	11:31 AM	池袋（豊島区）
8	10659	2021/1/2	3,120,909	3,788,182	667,273	10:51 AM	三鷹市
9	10796	2021/1/2		3,683,182		8:44 AM	中野区
10	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	渋谷区
11	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	
12	10221	2021/1/3	2,833,788	3,483,939	650,152	12:47 PM	池袋（豊島区）
13	10222	2021/1/3	317万円			5:49 PM	池袋（豊島区）

7

Copyright © SAS Institute Inc. All rights reserved.



Sale ID
販売ID

Buy Price
仕入価格

Date of Sale
販売日

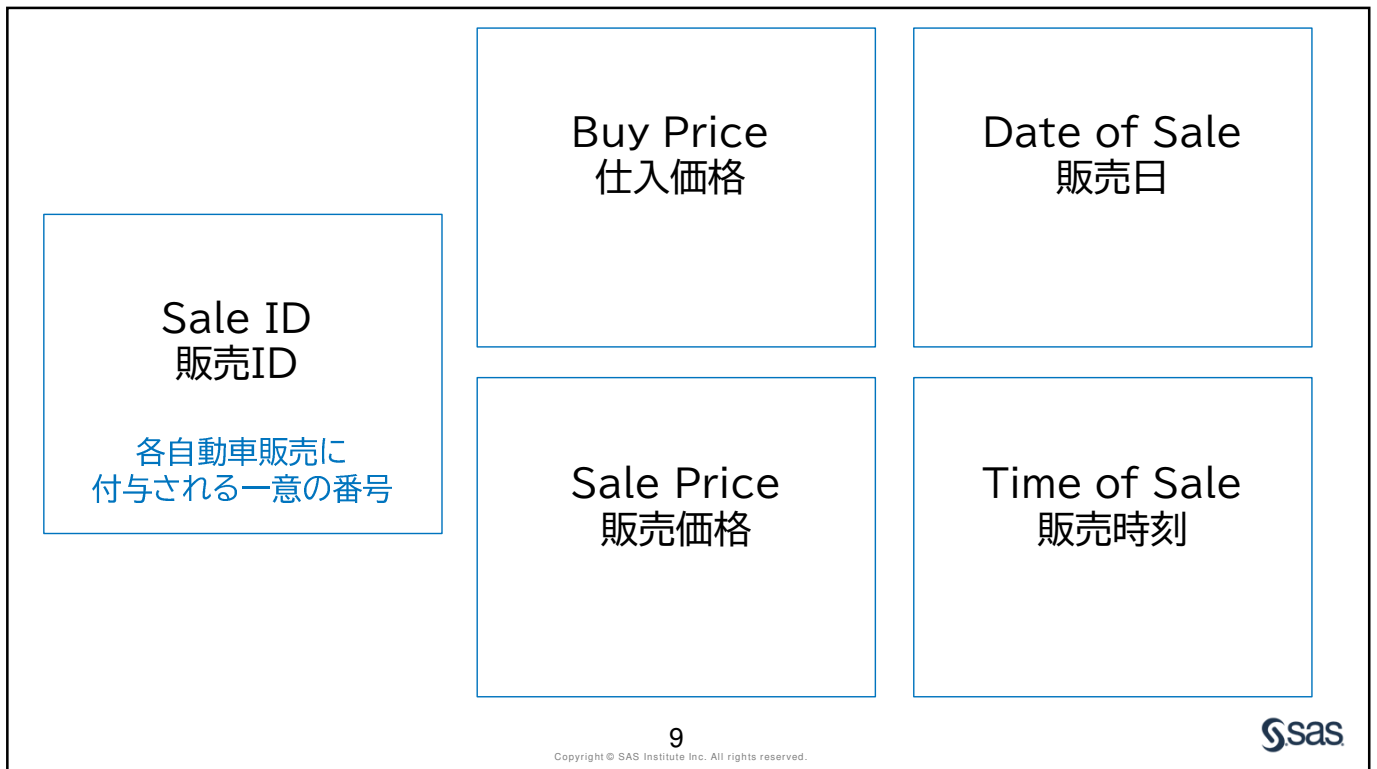
Sale Price
販売価格

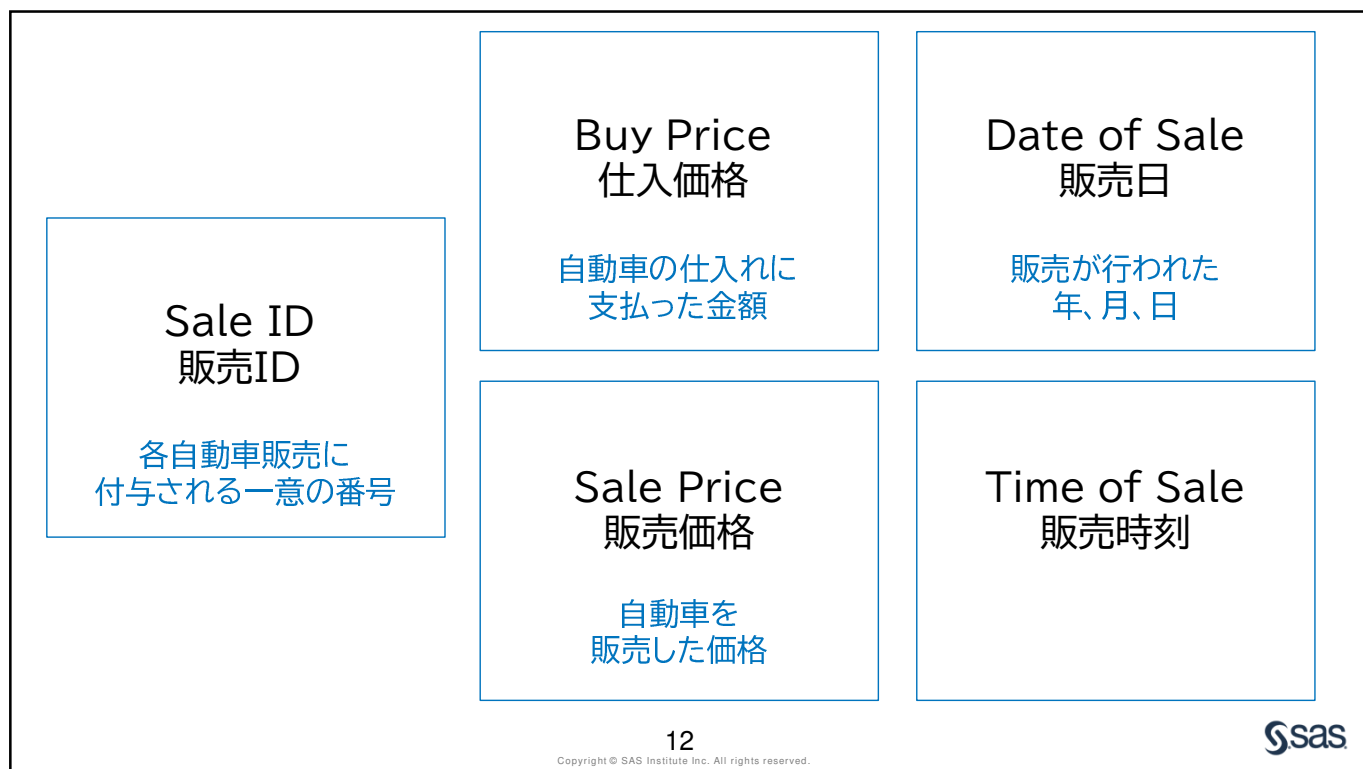
Time of Sale
販売時刻

8

Copyright © SAS Institute Inc. All rights reserved.









データセットの変数タイプを特定する

通常、データ項目には数値と文字の2種類がある。

文字データ項目は、テキストとして扱われ、文字、数字、特殊文字(例えば、#、!)が含まれる。そう、文字データ項目には数字も含めることができる！

数値データ項目と、数字である文字データ項目の違いはどのように見分けるのか？

- データセット内の数値は、その数値がラベルとして使用されたり、値を分類するために使用され、意味のある数値的計算を生成しない場合は、文字として扱う。例えば、郵便番号は数字であっても足したり掛けたりしてはいけない。しかし、特定の地理的位置のラベルとして扱われる場合には有用である。
- 鈴木さんのデータセットでは、販売IDの値は数字である。しかし、このIDは各販売の識別子として使用されているだけで、かつ意味のある計算には使用できないことに注意しよう。

数値が、文字データか、数値データかを判断する際の確認ポイント：

- ☐ 数字は単なるラベルとして機能しているのか？
- ☐ 数字の四則演算に意味があるのか？

この数値で意味のある計算ができるか？

	A	B	C	D	E	F	G
1	Sale ID	Date of Sale	Buy Price	Sale Price	Profit	Time of Sale	Dealership Name
2	10218	2021/1/1	3,223,333	4,138,788	915,455	10:48 AM	池袋（豊島区）
3	10219	2021/1/1	3,064,545	3,703,182	638,636	4:47 PM	池袋（豊島区）
4	10658	2021/1/1				2:27 PM	三鷹市
5	10794	2021/1/1	3,057,273	3,289,697	232,424	10:37 AM	中野区
6	10795	2021/1/1	2,604,545	3,128,030	523,485	11:59 AM	中野区
7	10220	2021/1/2	3,289,545	3,759,848	470,303	11:31 AM	池袋（豊島区）
8	10659	2021/1/2	3,120,909	3,788,182	667,273	10:51 AM	三鷹市
9	10796	2021/1/2		3,683,182		8:44 AM	中野区
10	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	渋谷区
11	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	
12	10221	2021/1/3	2,833,788	3,483,939	650,152	12:47 PM	池袋（豊島区）
13	10222	2021/1/3	317万円			5:49 PM	池袋（豊島区）

15

Copyright © SAS Institute Inc. All rights reserved.



データセットのサイズを確認する

	A	B	C	D	E	F	G
1	販売ID	販売日	仕入価格	販売価格	利益	販売時間	販売店
2	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	渋谷区
3	10002	2021/1/6	3,645,000	4,481,818	836,818	6:14 PM	渋谷区
4	10003	2021/1/8	1,869,242	2,464,545	595,303	9:18 AM	渋谷区
970	10969	2021/12/18	1,977,424	2,302,424	325,000	10:12 AM	中野区
971	10970	2021/12/18	2,427,424	2,838,030	410,606	1:06 PM	中野区
972	10971	2021/12/20	2,784,697	3,612,121	827,424	4:00 PM	中野区
973	10972	2021/12/21	2,243,333	2,896,667	653,333	10:24 AM	中野区
974	10973	2021/12/21	3,094,242	3,805,909	711,667	3:45 PM	中野区
975	10974	2021/12/23	2,861,667	3,680,606	818,939	4:49 PM	中野区
976	10975	2021/12/28	3,265,909	3,965,455	699,545	5:31 PM	中野区
977	10976	2021/12/29	3,199,545	3,559,848	360,303	1:00 PM	中野区
978	10977	2021/12/31	2,923,182	3,764,242	841,061	4:42 PM	中野区
979							

16

Copyright © SAS Institute Inc. All rights reserved.



ポイント1

オブザベーション
が多いほど =  正確さ
信頼度

オブザベーション
が少ないほど =  正確さ
信頼度

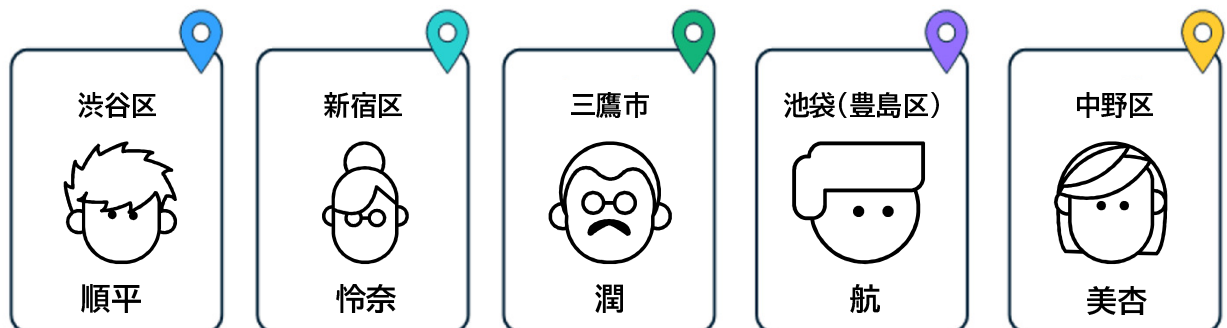
17

Copyright © SAS Institute Inc. All rights reserved.



ポイント2

渚の杜自動車
支店と店長

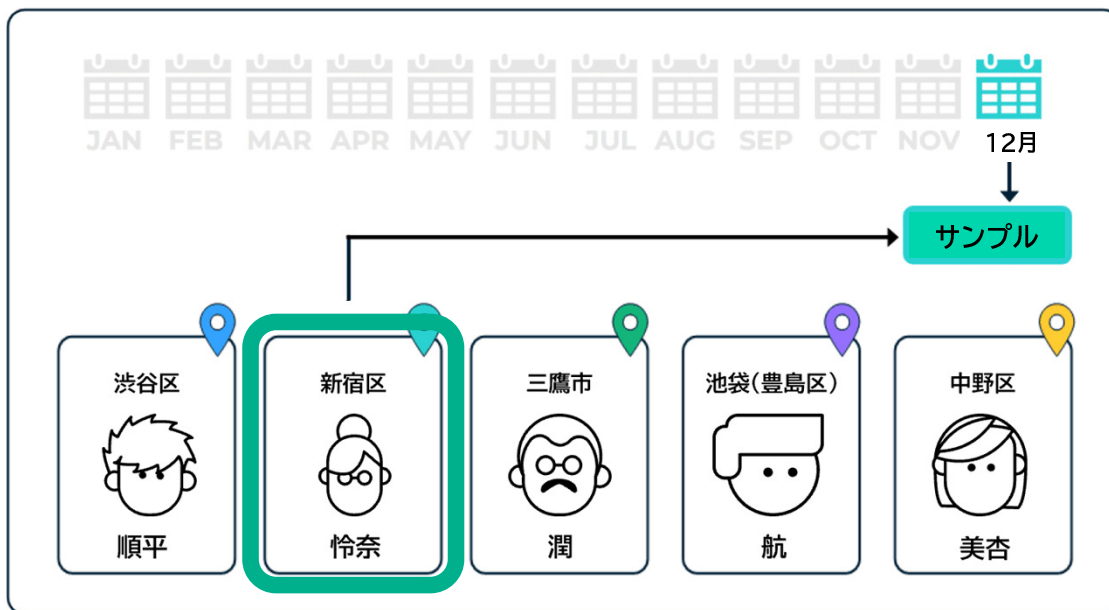


18

Copyright © SAS Institute Inc. All rights reserved.



ポイント3



19

Copyright © SAS Institute Inc. All rights reserved.

sas

ポイント4

2021年12月

Sun	Mon	Tue	Wed	Thu	Fri	Sat
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

連休中の販売

販売のスパイク = 異常

20

Copyright © SAS Institute Inc. All rights reserved.

sas



早押しクイズ

○か×か答えなさい

1. データセットは行×列で構成され、行はオブザベーションないしケース、列は変数ないしデータ項目と呼ばれる。
2. データのタイプには2種類があり、文字型と数字型に分けられる。
3. データセットのサイズは、サンプルサイズ、標本数、n数、とも呼ばれ、行の数(オブザベーション／ケースの数)を表す。

データセットにエラーがあることは珍しいことではない。

鈴木さんは2021年を通して31人の異なる従業員がデータを記録しており、意図しないミスが起こることがあります。

データ分析を始める前に、自問してください

- ① データは意味のあるものか？
- ② データは一貫したフォーマットに従っているか？
- ③ データに欠落はないか？
- ④ 重複データはないか？

23

Copyright © SAS Institute Inc. All rights reserved.



①データは意味のあるものか？

ぱっと見て、データセットに期待するものが分かるか？複雑な分析を始める前にダブルチェックするのは良い。

- ・各列のデータはタイトルから予想されるものと一致しているか？例えば、列のラベルがDate(日付)の場合、その列の値は日付か？
- ・データは有効か？例えば、データが1～10の範囲とされている場合、すべてのデータがその範囲内に収まっているか？
- ・データにタイプミスやスペースなどの小さな間違いはないか？

24

Copyright © SAS Institute Inc. All rights reserved.



②フォーマットエラー

January 1, 2021			
1-1-21	1-1-2021	01-01-21	01/01/21
01-01-2021	Jan-01	1 Jan	01JAN21

	A	B	C	D	E	F	G
1	Sale ID	Date of Sale	Buy Price	Sale Price	Profit	Time of Sale	Dealership Name
2	10218	2021/1/1	3,223,333	4,138,788	915,455	10:48 AM	池袋 (豊島区)
3	10219	2021/1/1	3,064,545	3,703,182	638,636	4:47 PM	池袋 (豊島区)
4	10658	2021/1/1				2:27 PM	三鷹市
5	10794	2021/1/1	3,057,273	3,289,697	232,424	10:37 AM	中野区
6	10795	2021/1/1	2,604,545	3,128,030	523,485	11:59 AM	中野区
7	10220	2021/1/2	3,289,545	3,759,848	470,303	11:31 AM	池袋 (豊島区)
8	10659	2021/1/2	3,120,909	3,788,182	667,273	10:51 AM	三鷹市
9	10796	2021/1/2		3,683,182		8:44 AM	中野区
10	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	渋谷区
11	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	
12	10221	2021/1/3	2,833,788	3,483,939	650,152	12:47 PM	池袋 (豊島区)
13	10222	2021/1/3	3,172,273			5:49 PM	池袋 (豊島区)
14	10797	2021/1/3	2,736,061	3,694,242	958,182	2:24 PM	中野区
15	10223	2021/1/4	2,441,818	2,870,606	428,788	3:08 PM	池袋 (豊島区)
16	10660	2024/1/4	2,705,455	3,264,091	558,636	2:18 PM	三鷹市
17							
18	10418	2021/1/5	2,646,061	3,592,576	946,515	4:30 PM	新宿区
19	10002	2021/1/6	3,645,000	4,481,818	836,818	6:14 PM	渋谷区

原データ

	A	B	C	D	E	F	G
1	Sale ID	Date of Sale	Buy Price	Sale Price	Profit	Time of Sale	Dealership Name
2	10218	2021/1/1	3,223,333	4,138,788	915,455	10:48 AM	池袋 (豊島区)
3	10219	2021/1/1	3,064,545	3,703,182	638,636	4:47 PM	池袋 (豊島区)
4	10658	2021/1/1				2:27 PM	三鷹市
5	10794	2021/1/1	3,057,273	3,289,697	232,424	10:37 AM	中野区
6	10795	2021/1/1	2,604,545	3,128,030	523,485	11:59 AM	中野区
7	10220	2021/1/2	3,289,545	3,759,848	470,303	11:31 AM	池袋 (豊島区)
8	10659	2021/1/2	3,120,909	3,788,182	667,273	10:51 AM	三鷹市
9	10796	2021/1/2		3,683,182		8:44 AM	中野区
10	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	渋谷区
11	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	
12	10221	2021/1/3	2,833,788	3,483,939	650,152	12:47 PM	池袋 (豊島区)
13	10222	2021/1/3	3,172,273			5:49 PM	池袋 (豊島区)
14	10797	2021/1/3	2,736,061	3,694,242	958,182	2:24 PM	中野区
15	10223	2021/1/4	2,441,818	2,870,606	428,788	3:08 PM	池袋 (豊島区)
16	10660	2024/1/4	2,705,455	3,264,091	558,636	2:18 PM	三鷹市
17							
18	10418	2021/1/5	2,646,061	3,592,576	946,515	4:30 PM	新宿区
19	10002	2021/1/6	3,645,000	4,481,818	836,818	6:14 PM	渋谷区

クリーンデータ

25

Copyright © SAS Institute Inc. All rights reserved.



③欠損データ

	A	B	C	D	E	F	G
1	Sale ID	Date of Sale	Buy Price	Sale Price	Profit	Time of Sale	Dealership Name
2	10218	2021/1/1	3,223,333	4,138,788	915,455	10:48 AM	池袋 (豊島区)
3	10219	2021/1/1	3,064,545	3,703,182	638,636	4:47 PM	池袋 (豊島区)
4	10658	2021/1/1				2:27 PM	三鷹市
5	10794	2021/1/1	3,057,273	3,289,697	232,424	10:37 AM	中野区
6	10795	2021/1/1	2,604,545	3,128,030	523,485	11:59 AM	中野区
7	10220	2021/1/2	3,289,545	3,759,848	470,303	11:31 AM	池袋 (豊島区)
8	10659	2021/1/2	3,120,909	3,788,182	667,273	10:51 AM	三鷹市
9	10796	2021/1/2		3,683,182		8:44 AM	中野区
10	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	渋谷区
11	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	
12	10221	2021/1/3	2,833,788	3,483,939	650,152	12:47 PM	池袋 (豊島区)
13	10222	2021/1/3	3,172,273			5:49 PM	池袋 (豊島区)
14	10797	2021/1/3	2,736,061	3,694,242	958,182	2:24 PM	中野区
15	10223	2021/1/4	2,441,818	2,870,606	428,788	3:08 PM	池袋 (豊島区)
16	10660	2024/1/4	2,705,455	3,264,091	558,636	2:18 PM	三鷹市
17							
18	10418	2021/1/5	2,646,061	3,592,576	946,515	4:30 PM	新宿区
19	10002	2021/1/6	3,645,000	4,481,818	836,818	6:14 PM	渋谷区

原データ

	A	B	C	D	E	F	G
1	Sale ID	Date of Sale	Buy Price	Sale Price	Profit	Time of Sale	Dealership Name
2	10218	2021/1/1	3,223,333	4,138,788	915,455	10:48 AM	池袋 (豊島区)
3	10219	2021/1/1	3,064,545	3,703,182	638,636	4:47 PM	池袋 (豊島区)
4	10658	2021/1/1	2,846,061	3,562,576	716,515	2:27 PM	三鷹市
5	10794	2021/1/1	3,057,273	3,289,697	232,424	10:37 AM	中野区
6	10795	2021/1/1	2,604,545	3,128,030	523,485	11:59 AM	中野区
7	10220	2021/1/2	3,289,545	3,759,848	470,303	11:31 AM	池袋 (豊島区)
8	10659	2021/1/2	3,120,909	3,788,182	667,273	10:51 AM	三鷹市
9	10796	2021/1/2	2,921,515	3,683,182	761,667	8:44 AM	中野区
10	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	渋谷区
11	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	
12	10221	2021/1/3	2,833,788	3,483,939	650,152	12:47 PM	池袋 (豊島区)
13	10222	2021/1/3	3,172,273	3,453,636	281,364	5:49 PM	池袋 (豊島区)
14	10797	2021/1/3	2,736,061	3,694,242	958,182	2:24 PM	中野区
15	10223	2021/1/4	2,441,818	2,870,606	428,788	3:08 PM	池袋 (豊島区)
16	10660	2024/1/4	2,705,455	3,264,091	558,636	2:18 PM	三鷹市
17	10418	2021/1/5	2,646,061	3,592,576	946,515	4:30 PM	新宿区
18	10002	2021/1/6	3,645,000	4,481,818	836,818	6:14 PM	渋谷区

クリーンデータ

26

Copyright © SAS Institute Inc. All rights reserved.



④重複データ

	A	B	C	D	E	F	G
1	Sale ID	Date of Sale	Buy Price	Sale Price	Profit	Time of Sale	Dealership Name
2	10218	2021/1/1	3,223,333	4,138,788	915,455	10:48 AM	池袋（豊島区）
3	10219	2021/1/1	3,064,545	3,703,182	638,636	4:47 PM	池袋（豊島区）
4	10658	2021/1/1	2,846,061	3,562,576	716,515	2:27 PM	三鷹市
5	10794	2021/1/1	3,057,273	3,289,697	232,424	10:37 AM	中野区
6	10795	2021/1/1	2,604,545	3,128,030	523,485	11:59 AM	中野区
7	10220	2021/1/2	3,289,545	3,759,848	470,303	11:31 AM	池袋（豊島区）
8	10659	2021/1/2	3,120,909	3,788,182	667,273	10:51 AM	三鷹市
9	10796	2021/1/2	2,921,515	3,683,182	761,667	8:44 AM	中野区
10	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	渋谷区
11	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	渋谷区
12	10221	2021/1/3	2,833,788	3,483,939	650,152	12:47 PM	池袋（豊島区）
13	10222	2021/1/3	3,172,273	3,453,636	281,364	5:49 PM	池袋（豊島区）
14	10797	2021/1/3	2,736,061	3,694,242	958,182	2:24 PM	中野区
15	10223	2021/1/4	2,441,818	2,870,606	428,788	3:08 PM	池袋（豊島区）
16	10660	2024/1/4	2,705,455	3,264,091	558,636	2:18 PM	三鷹市
17	10418	2021/1/5	2,646,061	3,592,576	946,515	4:30 PM	新宿区
18	10002	2021/1/6	3,645,000	4,481,818	836,818	6:14 PM	渋谷区
19	10224	2021/1/6	3,135,606	4,000,303	864,697	1:10 PM	池袋（豊島区）
20	10419	2021/1/6	1,519,545	2,347,879	828,333	4:45 PM	新宿区
21	10420	2021/1/6	1,940,909	2,618,939	678,030	6:45 PM	新宿区
22	10421	2021/1/7	2,670,152	3,649,242	979,091	4:10 PM	新宿区



	A	B	C	D	E	F	G
1	Sale ID	Date of Sale	Buy Price	Sale Price	Profit	Time of Sale	Dealership Name
2	10218	2021/1/1	3,223,333	4,138,788	915,455	10:48 AM	池袋（豊島区）
3	10219	2021/1/1	3,064,545	3,703,182	638,636	4:47 PM	池袋（豊島区）
4	10658	2021/1/1	2,846,061	3,562,576	716,515	2:27 PM	三鷹市
5	10794	2021/1/1	3,057,273	3,289,697	232,424	10:37 AM	中野区
6	10795	2021/1/1	2,604,545	3,128,030	523,485	11:59 AM	中野区
7	10220	2021/1/2	3,289,545	3,759,848	470,303	11:31 AM	池袋（豊島区）
8	10659	2021/1/2	3,120,909	3,788,182	667,273	10:51 AM	三鷹市
9	10796	2021/1/2	2,921,515	3,683,182	761,667	8:44 AM	中野区
10	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	渋谷区
11	10221	2021/1/3	2,833,788	3,483,939	650,152	12:47 PM	池袋（豊島区）
12	10222	2021/1/3	3,172,273	3,453,636	281,364	5:49 PM	池袋（豊島区）
13	10797	2021/1/3	2,736,061	3,694,242	958,182	2:24 PM	中野区
14	10223	2021/1/4	2,441,818	2,870,606	428,788	3:08 PM	池袋（豊島区）
15	10660	2024/1/4	2,705,455	3,264,091	558,636	2:18 PM	三鷹市
16	10418	2021/1/5	2,646,061	3,592,576	946,515	4:30 PM	新宿区
17	10002	2021/1/6	3,645,000	4,481,818	836,818	6:14 PM	渋谷区
18	10224	2021/1/6	3,135,606	4,000,303	864,697	1:10 PM	池袋（豊島区）
19	10419	2021/1/6	1,519,545	2,347,879	828,333	4:45 PM	新宿区
20	10420	2021/1/6	1,940,909	2,618,939	678,030	6:45 PM	新宿区
21	10421	2021/1/7	2,670,152	3,649,242	979,091	4:10 PM	新宿区

原データ

クリーンデータ

27

Copyright © SAS Institute Inc. All rights reserved.



鈴木さんはデータに精通しており、時間をかけてエラーに対処してきました。

データセットは準備できていますが、作業を開始する前に、鈴木さんは分析の計画を立てます。データを把握したので、どのような疑問／質問を調査するかをより適切に検討できます。鈴木さんは作業を開始するのが楽しみです。

全体像データ(再掲)

鈴木さんは、事業を運営するために必要な消耗品や在庫の価格が、輸入により急速に上昇していることを知っている。

鈴木さんはまた、2022年に完了する必要があるいくつかの重要な修理が遅れている。その中には、全店舗の小規模な修繕に加え、ある店舗では新しいACユニット、別の店舗では屋根が含まれる。

何らかの決定を下す前に、鈴木さんは自社の現状をより詳細に把握するため、手持ちのデータを分析する予定だ。これは、将来、より多くの情報に基づいた財務上の決断を下すために極めて重要である。

29

Copyright © SAS Institute Inc. All rights reserved.

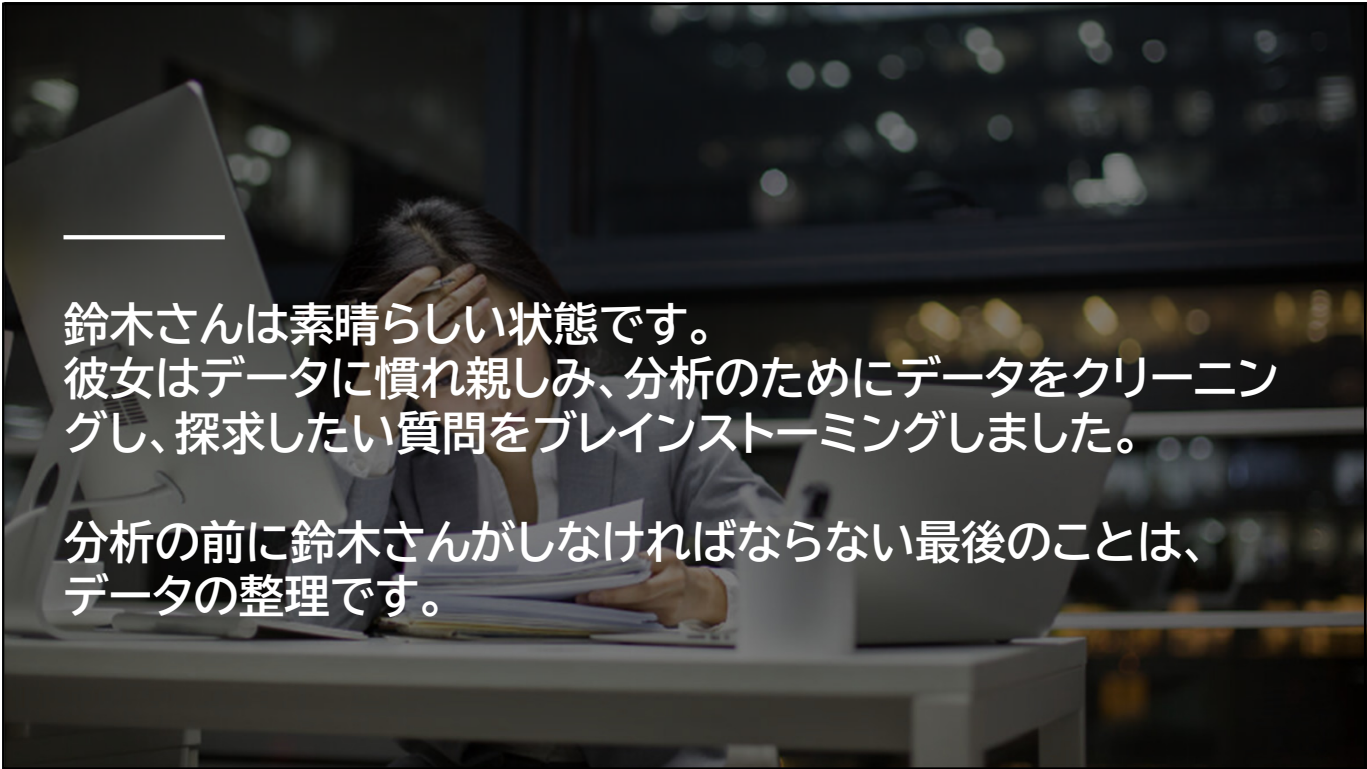


疑問	疑問(詳細)	データ(変数)	備考
個々の販売店の業績は？	他の販売店に比べてもっと利益が高い販売店は？	販売店名、仕入価格、販売価格	利益変数を作成するために、このような変数を使う必要がある。
	他の販売店に比べてもっと販売が高い販売店は？	販売店名、仕入価格、販売価格	
	特定の時間帯、日または月	販売店名、販売時間、販売日	
	仕入と販売において他と違う価格戦略を行っているか	販売店名、仕入価格、販売価格	
日中、販売に有利な時間帯があるか		販売時間	
1年のうちで、販売に有利な月があるか		販売日	
実績が高い販売員は誰か			鈴木さんはこのようなデータを収集しておらず、この質問には答えられない。

30

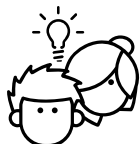
Copyright © SAS Institute Inc. All rights reserved.





鈴木さんは素晴らしい状態です。
彼女はデータに慣れ親しみ、分析のためにデータをクリーニングし、探求したい質問をブレインストーミングしました。

分析の前に鈴木さんがしなければならない最後のことは、データの整理です。



各販売店のデータをスクロールしてご覧ください

このシートは「完全なデータセット」と題され、販売店の全拠点からの978行のデータ全体を示している。

- ・渋谷区のデータ
- ・新宿区のデータ
- ・三鷹市のデータ
- ・池袋(豊島区)のデータ
- ・中野区のデータ



利益のデータを生成してみましょう！

$$\text{利益} = \text{販売価格} - \text{仕入価格}$$

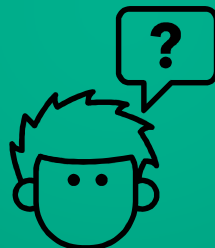
	A	B	C	D	E	F	G
1	販売ID	販売日	仕入価格	販売価格	利益	販売時間	販売店
2	10001	2021/1/3	2,894,394	3,646,970	752,576	4:36 PM	渋谷区
3	10002	2021/1/6	3,645,000	4,481,818	836,818	6:14 PM	渋谷区
4	10003	2021/1/8	1,869,242	2,464,545	595,303	9:18 AM	渋谷区
5	10004	2021/1/9	3,652,576	4,424,848	772,273	8:44 AM	渋谷区
6	10005	2021/1/11	3,040,909	3,688,333	647,424	11:22 AM	渋谷区
7	10006	2021/1/12	2,055,303	2,718,788	663,485	11:15 AM	渋谷区
8	10007	2021/1/13	3,452,424	4,131,667	679,242	2:12 PM	渋谷区
9	10008	2021/1/14	2,238,485	2,877,273	638,788	1:39 PM	渋谷区
10	10009	2021/1/15	2,921,212	3,922,121	1,000,909	6:45 PM	渋谷区
11	10010	2021/1/16	2,913,939	3,580,758	666,818	3:15 PM	渋谷区
12	10011	2021/1/16	2,575,000	3,420,606	845,606	11:22 AM	渋谷区
13	10012	2021/1/20	3,156,970	3,794,394	637,424	6:05 PM	渋谷区
14	10013	2021/1/24	3,128,939	3,858,333	729,394	4:22 PM	渋谷区
15	10014	2021/1/26	2,670,000	3,531,515	861,515	8:38 AM	渋谷区
16	10015	2021/1/27	2,419,697	2,828,788	409,091	2:42 PM	渋谷区
17	10016	2021/1/28	2,696,818	3,562,879	866,061	10:29 AM	渋谷区
18	10017	2021/1/28	2,377,121	3,017,576	640,455	10:08 AM	渋谷区

33

Copyright © SAS Institute Inc. All rights reserved.



質疑応答



Copyright © SAS Institute Inc. All rights reserved.

Module 3: 探索的データ分析

1. データリテラシーとのつながり

2. 形状を探る

3. 中心を探る

4. 広がりを探る

5. すべてをまとめる

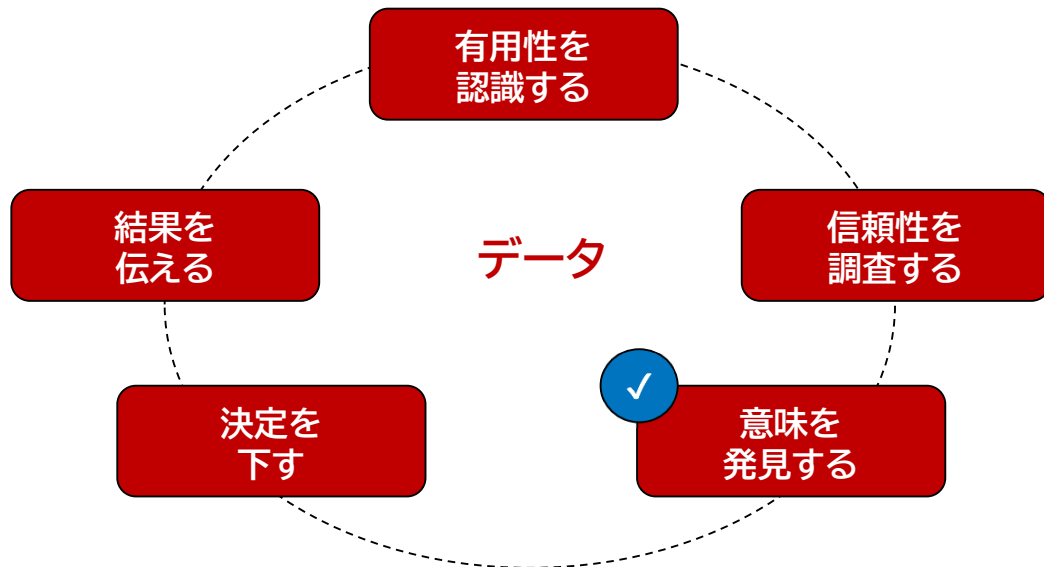


Copyright © SAS Institute Inc. All rights reserved.

A close-up photograph of a person's hands holding a tablet. The tablet screen shows a 3D bar chart with five bars of different colors (blue, red, green, yellow, pink) and numerical values on the x-axis (150, 400, 300, 200). The background is blurred, showing an indoor setting with a plant.

意味を発見するということは、
データのパターンを探索し、
そのパターンを理解することです。

データリテラシーのスキル



3

Copyright © SAS Institute Inc. All rights reserved.



人々はしばしば、特定の質問に答えるためにデータ分析にすぐに飛びつきたがる。しかし、探索的データ分析に時間をかけることは重要である。

なぜなら、探索的データ分析に時間をかけることは、データをどのように分析するかを決定するのに役立つからである。

探索的データ分析

1. 分布の形を理解する
2. 分布の中心を見つける
3. 分布におけるデータの広がりを計算する

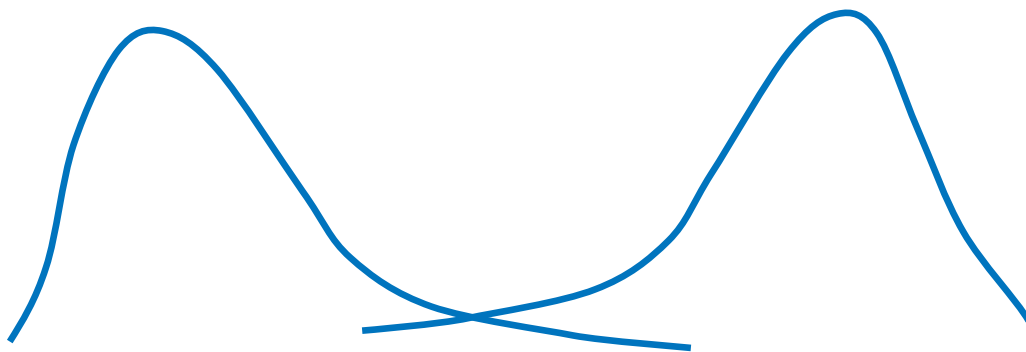
5

Copyright © SAS Institute Inc. All rights reserved.



形状を探る

Shape



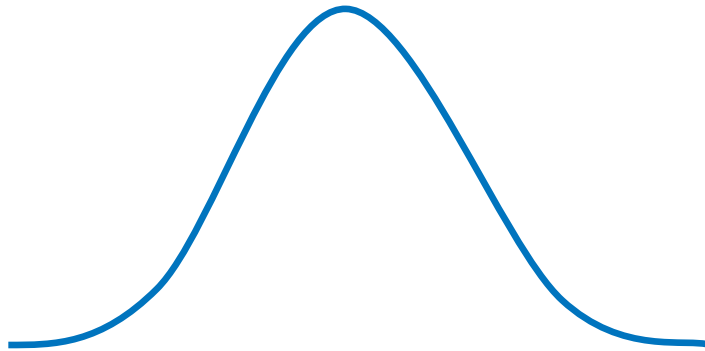
6

Copyright © SAS Institute Inc. All rights reserved.



中心を探る

Center



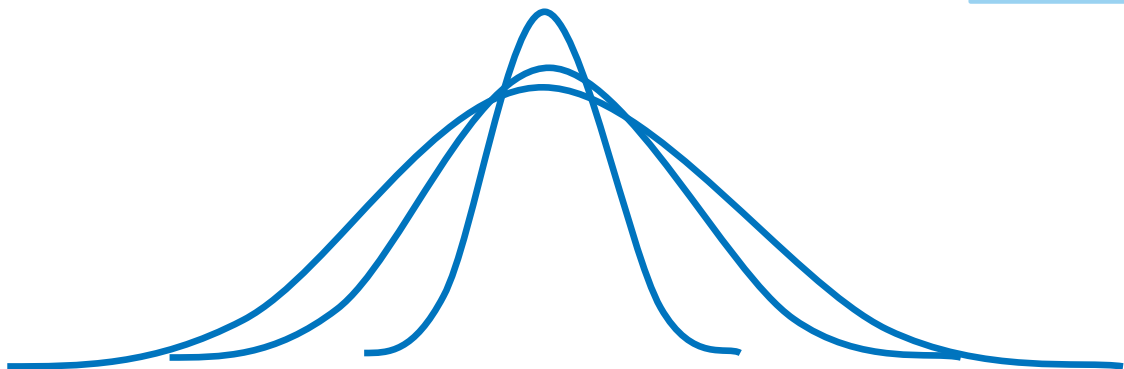
7

Copyright © SAS Institute Inc. All rights reserved.



広がりの探る

Spread



8

Copyright © SAS Institute Inc. All rights reserved.



形状を探る

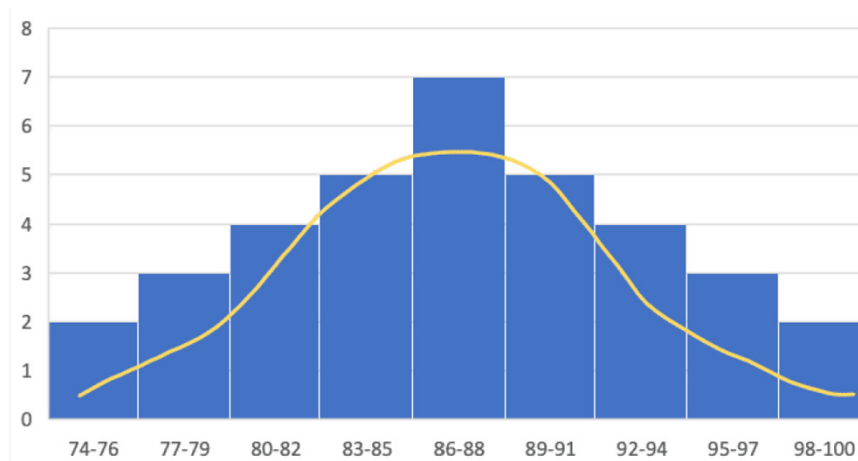
- ・度数分布表／ヒストグラム

9

Copyright © SAS Institute Inc. All rights reserved.



形状を探る



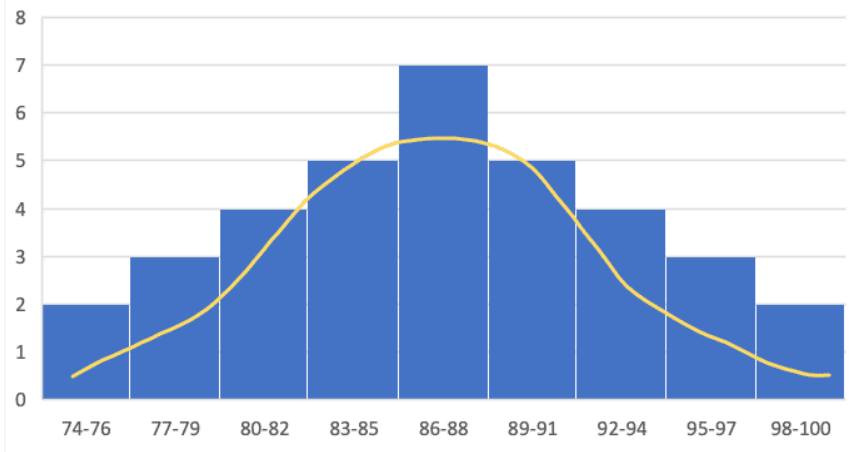
10

Copyright © SAS Institute Inc. All rights reserved.

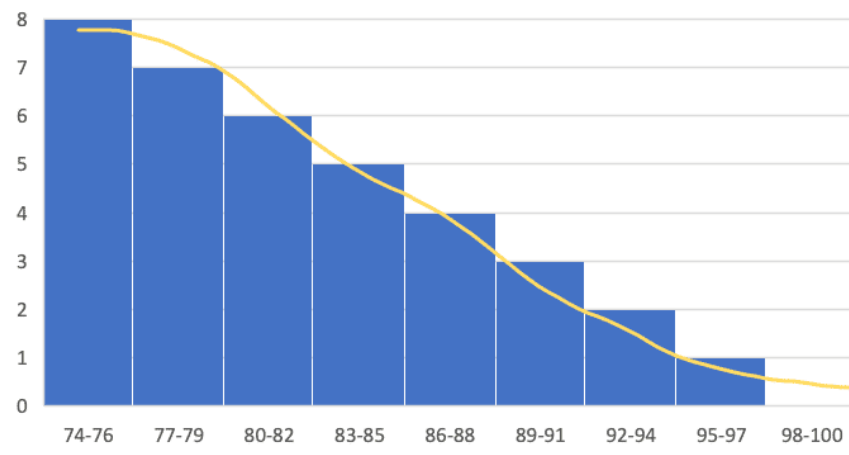


Classroom A	Classroom B	Classroom C			
74	74	78			
75	74	80			
77	74	81			
78	75	83			
79	75	84			
80	76	85			
81	76	86			
81	76	87			
82	77	87			
83	77	88			
83	77	89			
84	78	89			
84	78	90			
85	79	90			
86	79	91			
86	80	92			
87	80	92			
87	81	93			
			88	81	93
			88	82	94
			88	82	94
			89	83	95
			90	83	96
			90	84	96
			91	84	96
			91	85	96
			92	86	97
			93	87	97
			93	87	98
			94	88	98
			95	89	99
			96	90	99
			97	91	99
			98	92	99
			99	93	100
				95	100

A組



B組

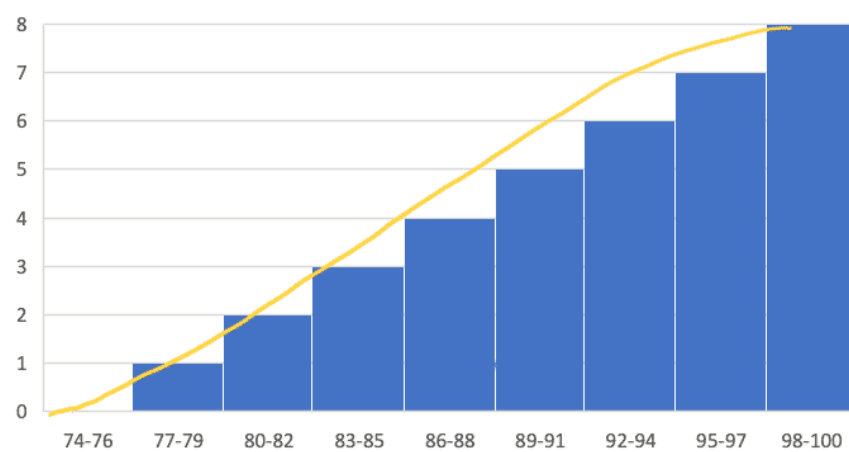


13

Copyright © SAS Institute Inc. All rights reserved.



C組



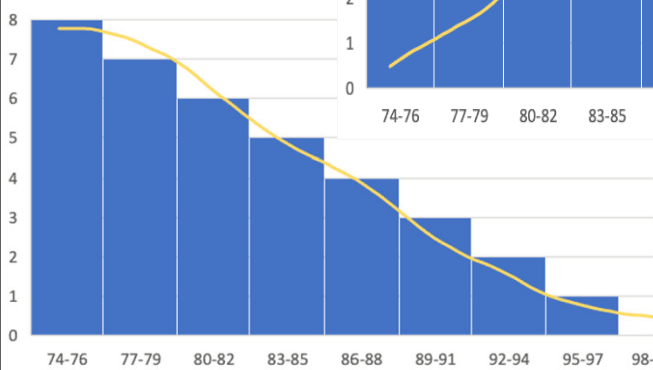
14

Copyright © SAS Institute Inc. All rights reserved.

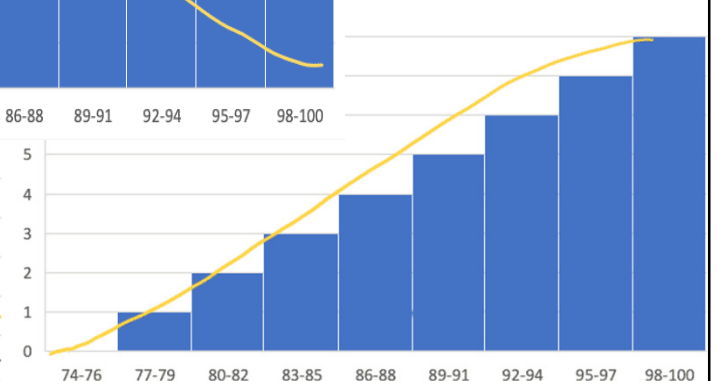


形状を探る

右に歪んだ分布

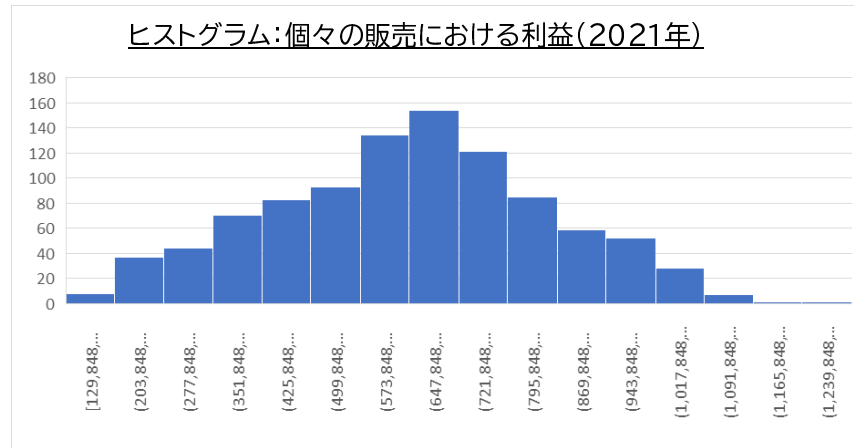


左に歪んだ分布



鈴木さんのデータの形を調べる

全店

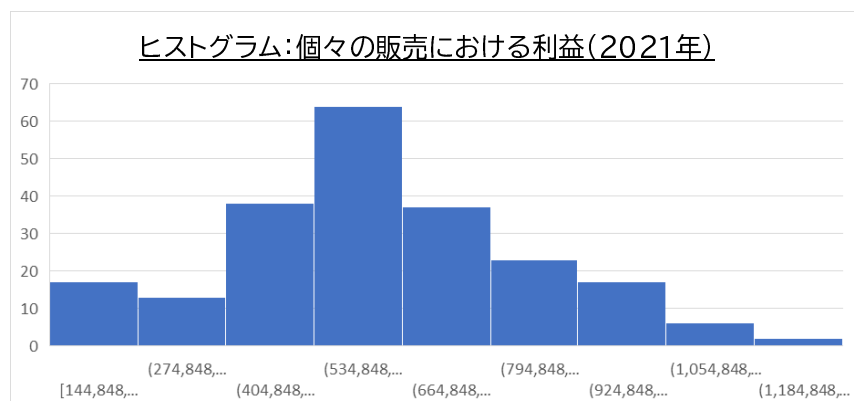


17

Copyright © SAS Institute Inc. All rights reserved.



渋谷区

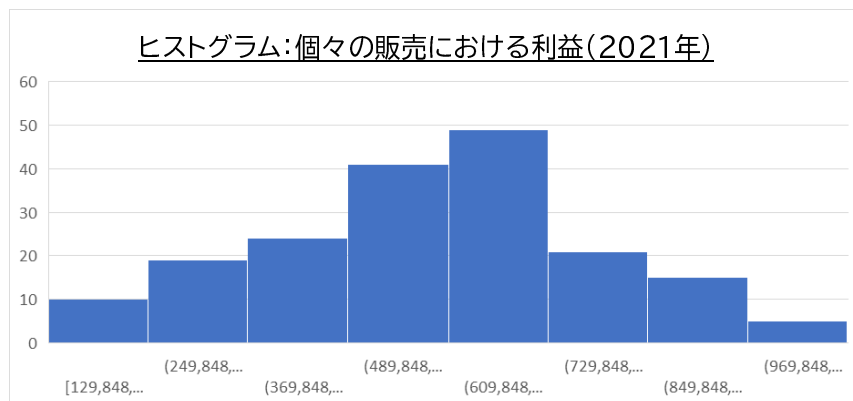


18

Copyright © SAS Institute Inc. All rights reserved.



中野区

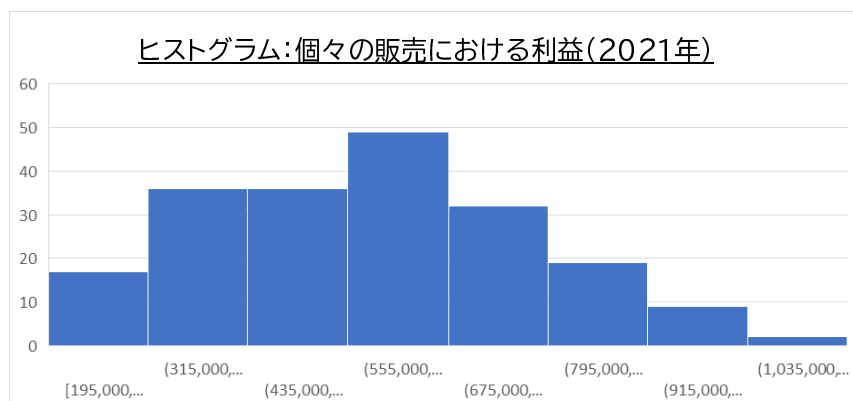


19

Copyright © SAS Institute Inc. All rights reserved.



池袋(豊島区)

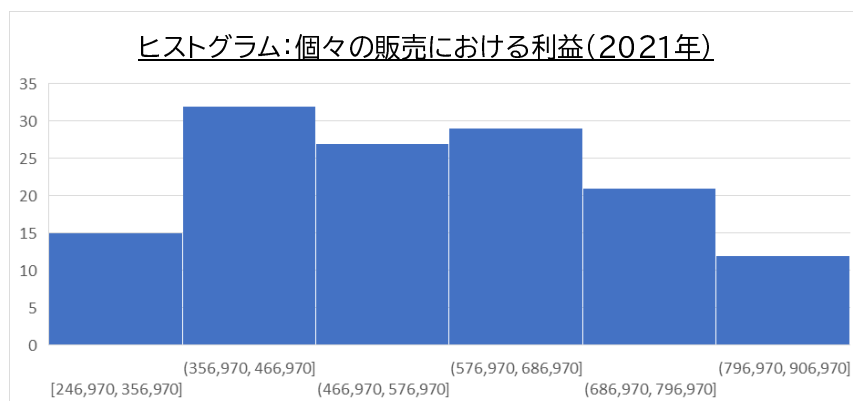


20

Copyright © SAS Institute Inc. All rights reserved.



三鷹市

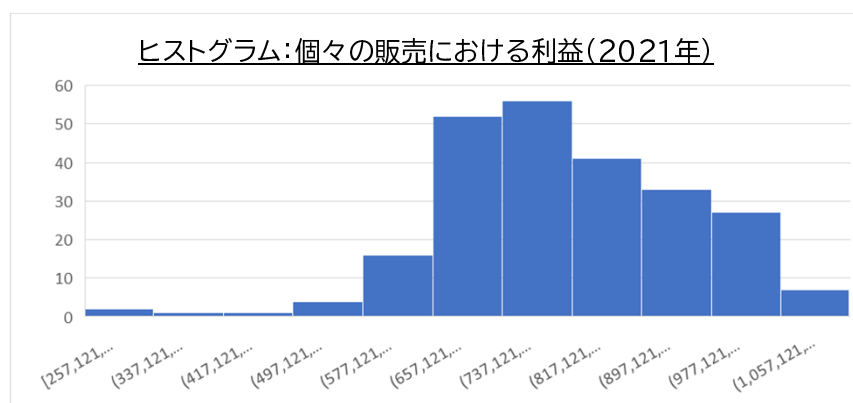


21

Copyright © SAS Institute Inc. All rights reserved.



新宿区



22

Copyright © SAS Institute Inc. All rights reserved.



中心を探る

- 平均(値)と中央値
- 平均の注意点

23

Copyright © SAS Institute Inc. All rights reserved.



平均(値)と中央値

99, 85, 85, 93, 80

$$\frac{99 + 85 + 85 + 93 + 80}{5} = 88.4$$

80, 85, 85, 93, 99

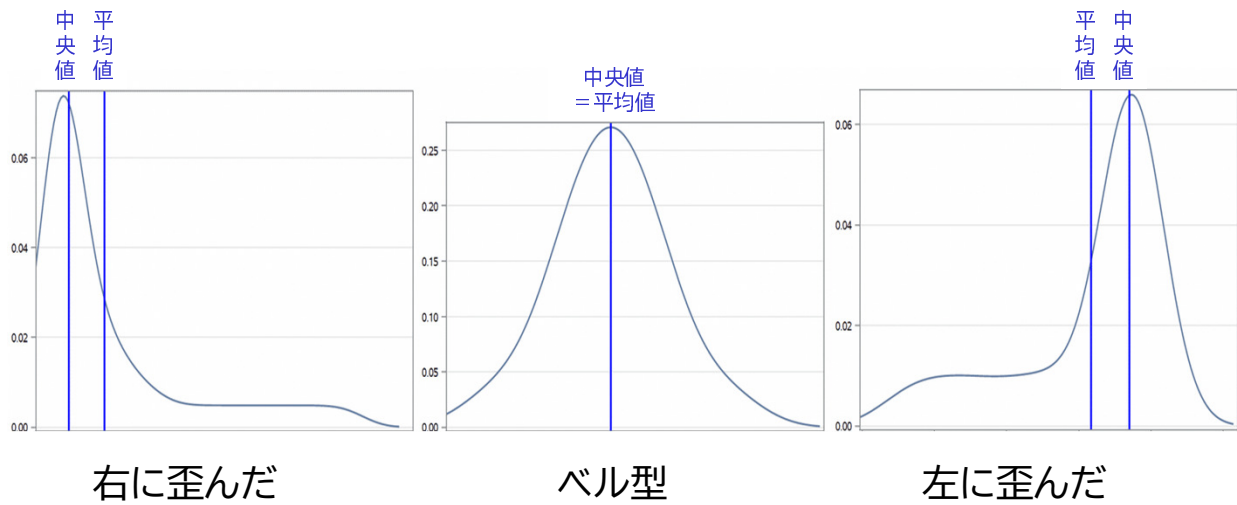
80, 85, 85, 93, 99

24

Copyright © SAS Institute Inc. All rights reserved.



平均(値)と中央値

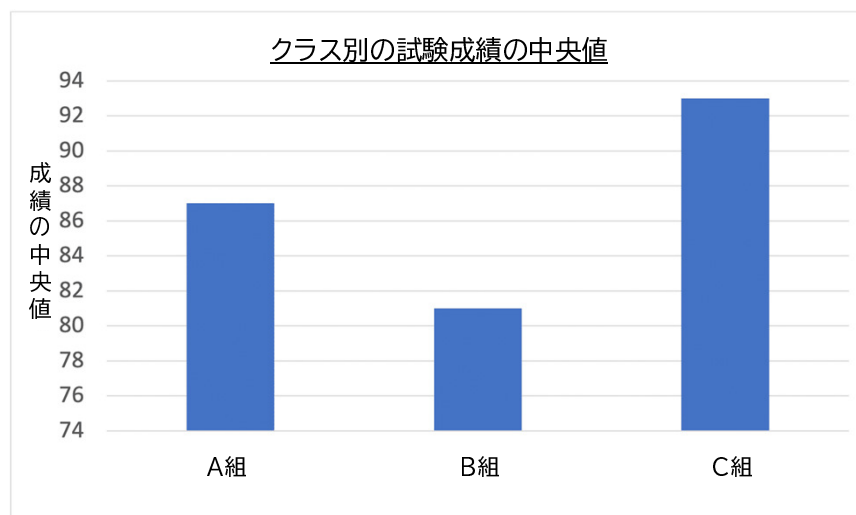


25

Copyright © SAS Institute Inc. All rights reserved.



平均(値)と中央値



26

Copyright © SAS Institute Inc. All rights reserved.



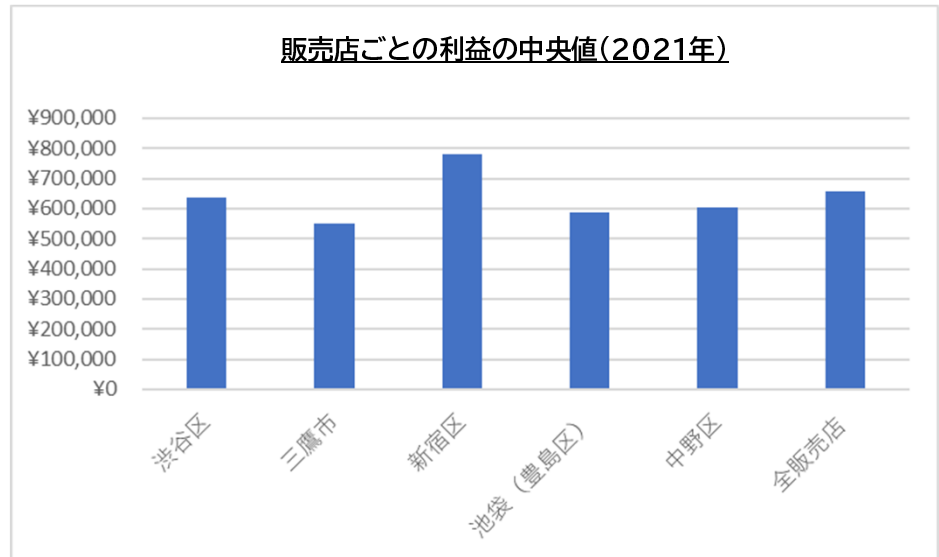
鈴木さんのデータの形を調べる

中心を探る

販売店	中央値	平均値	販売台数
渋谷区	¥635,455	¥633,242	217
三鷹市	¥551,515	¥559,610	136
新宿区	¥780,758	¥809,526	240
池袋(豊島区)	¥585,455	¥583,970	200
中野区	¥603,182	¥593,340	184
全販売店	¥656,667	¥648,695	977

中心を探る

販売店	中央値
渋谷区	¥635,455
三鷹市	¥551,515
新宿区	¥780,758
池袋(豊島区)	¥585,455
中野区	¥603,182
全販売店	¥656,667



29

Copyright © SAS Institute Inc. All rights reserved.



広がりを探る

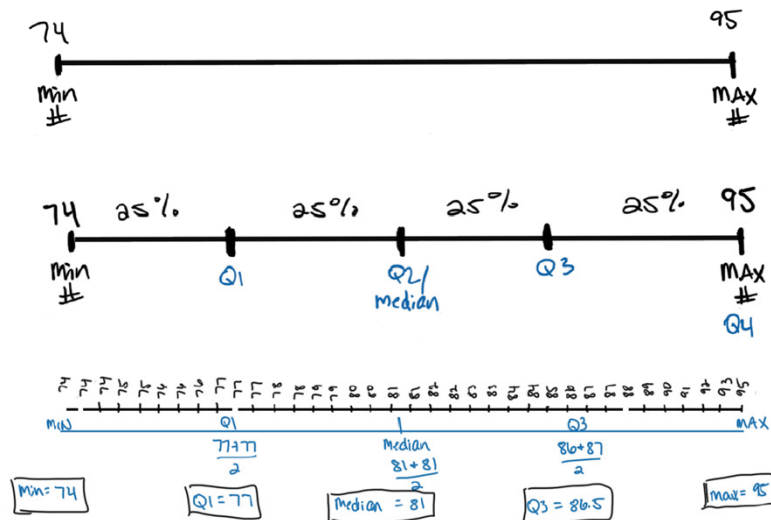
- ・範囲
- ・四分位数
- ・四分位範囲
- ・箱ひげ図

30

Copyright © SAS Institute Inc. All rights reserved.



範圍、四分位數、四分位範圍

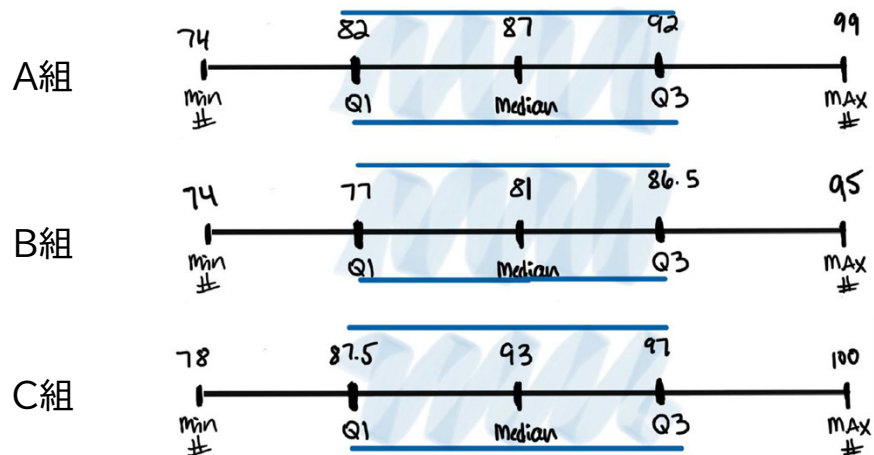
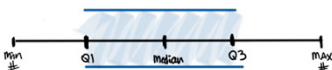


31

Copyright © SAS Institute Inc. All rights reserved.



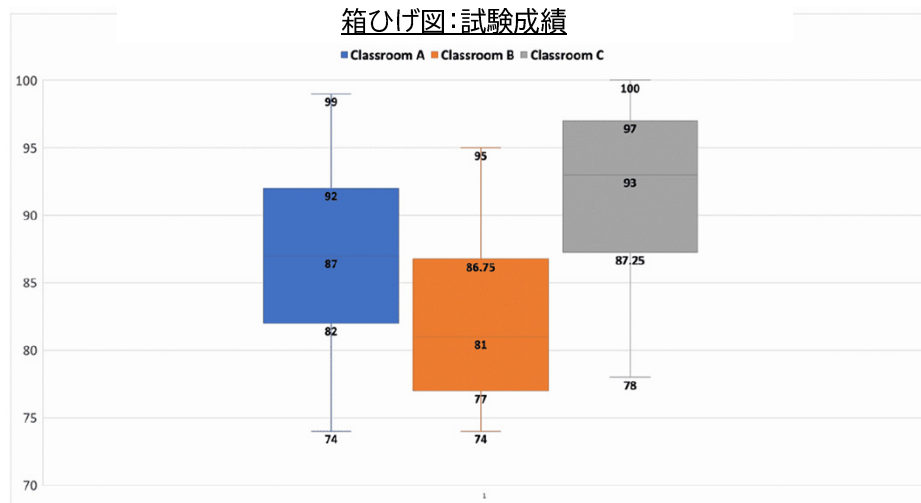
四分位範圍=第3四分位數-第1四分位數



Copyright © SAS Institute Inc. All rights reserved.



箱ひげ図

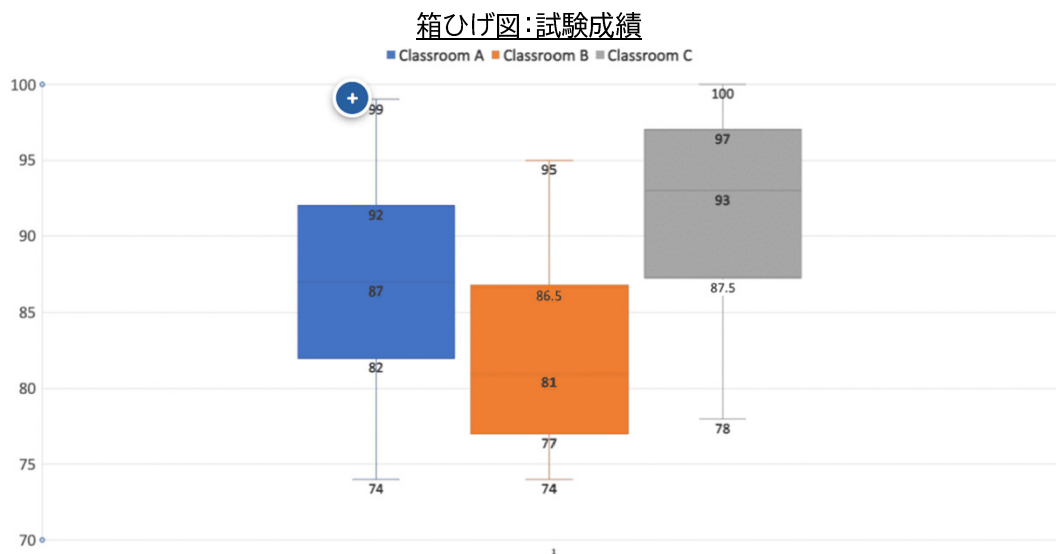


33

Copyright © SAS Institute Inc. All rights reserved.



最大値



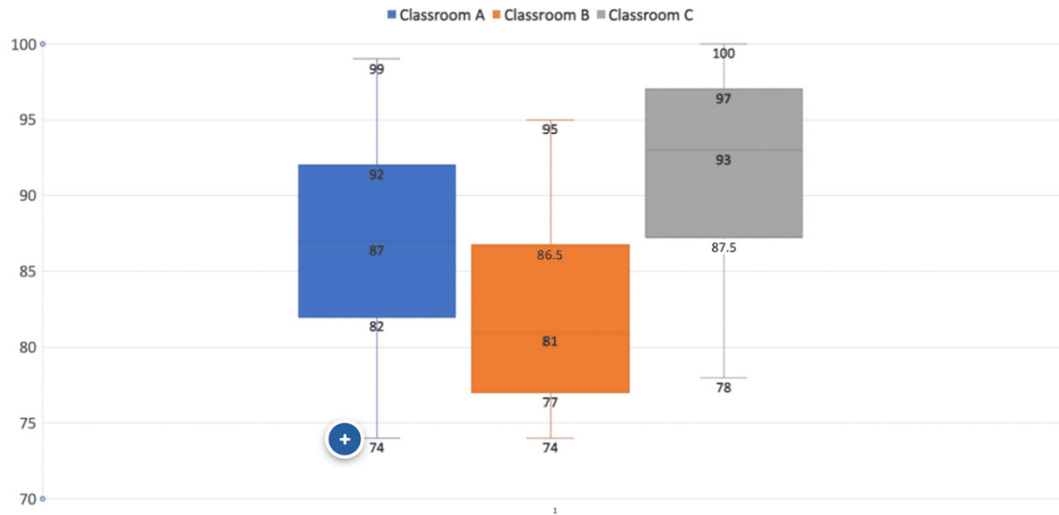
34

Copyright © SAS Institute Inc. All rights reserved.



最小値

箱ひげ図:試験成績



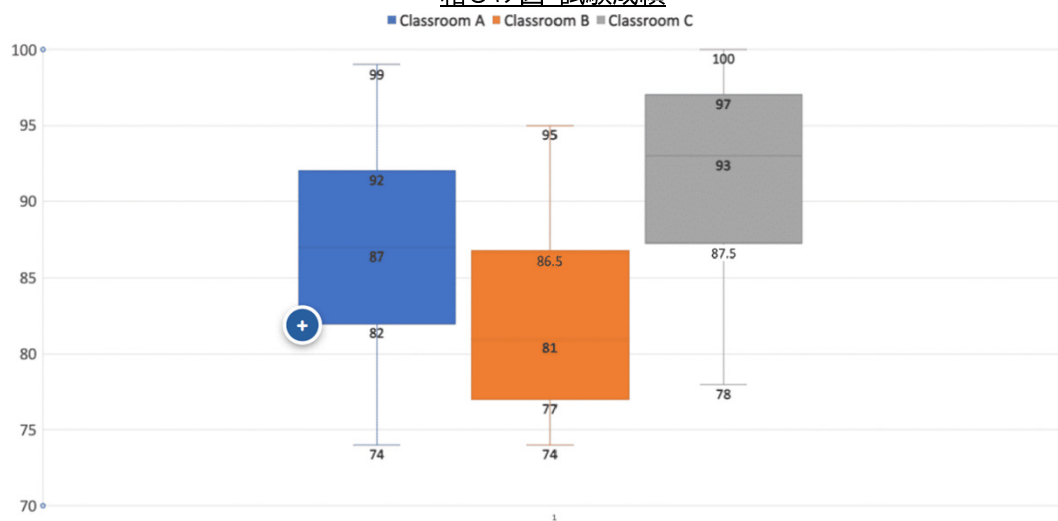
35

Copyright © SAS Institute Inc. All rights reserved.



第1四分位数

箱ひげ図:試験成績



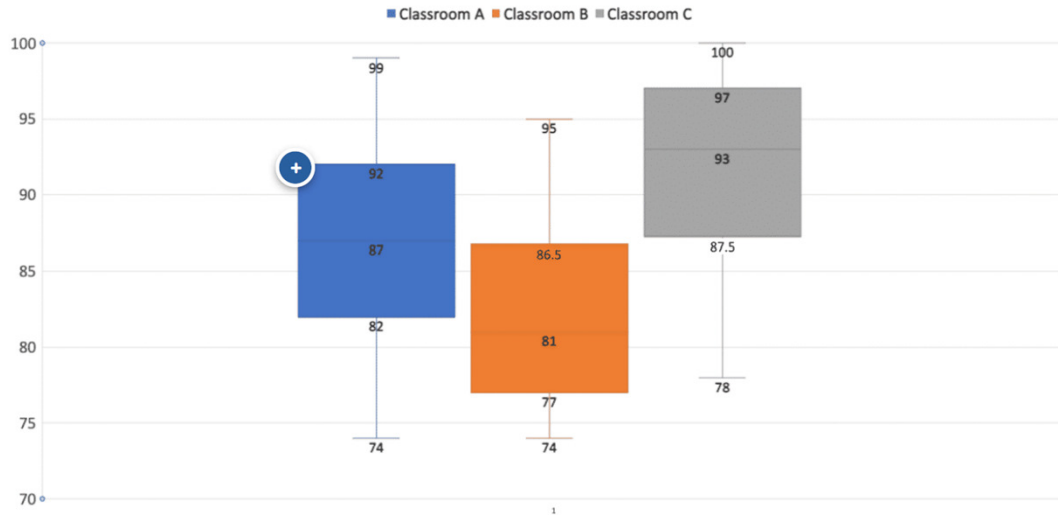
36

Copyright © SAS Institute Inc. All rights reserved.



第3四分位数

箱ひげ図: 試験成績



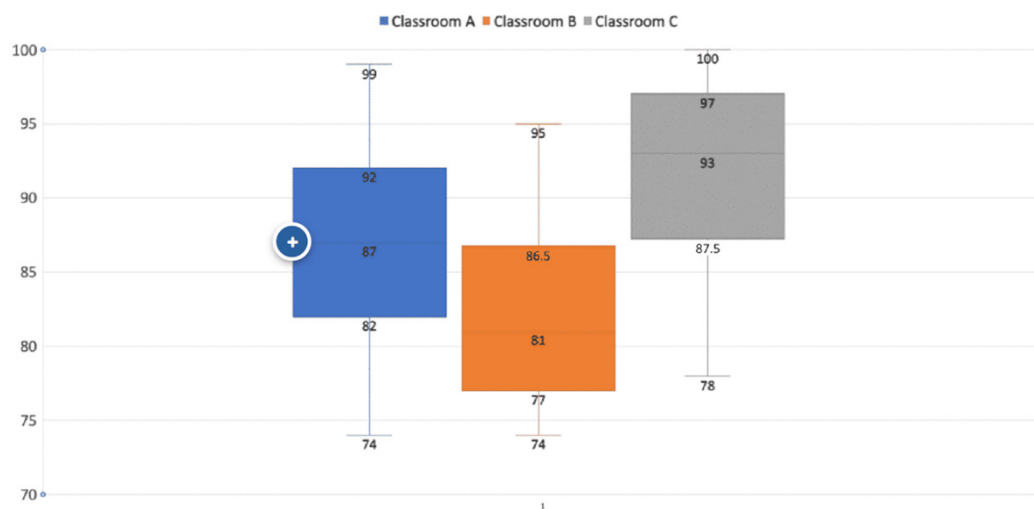
37

Copyright © SAS Institute Inc. All rights reserved.



第2四分位数 = 中央値

箱ひげ図: 試験成績



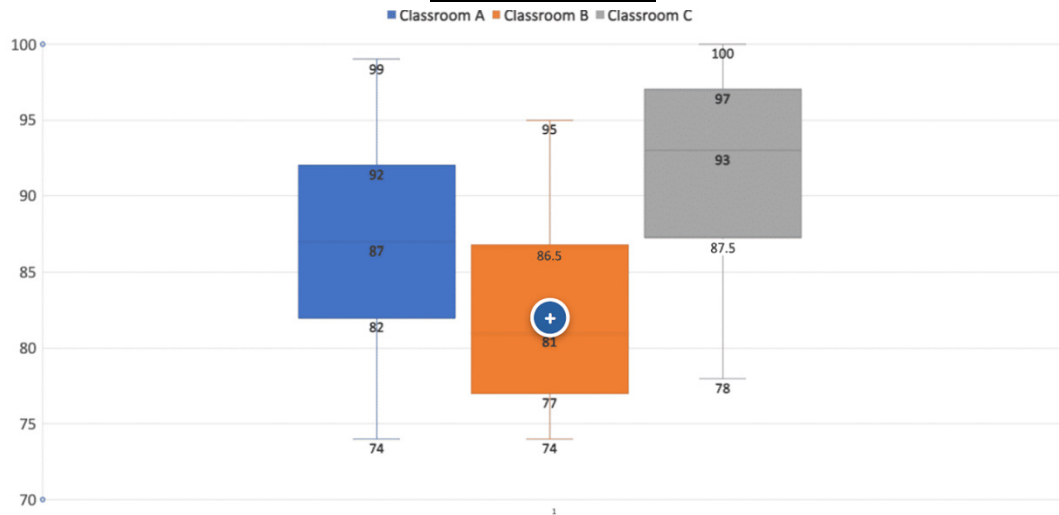
38

Copyright © SAS Institute Inc. All rights reserved.



四分位範囲

箱ひげ図:試験成績



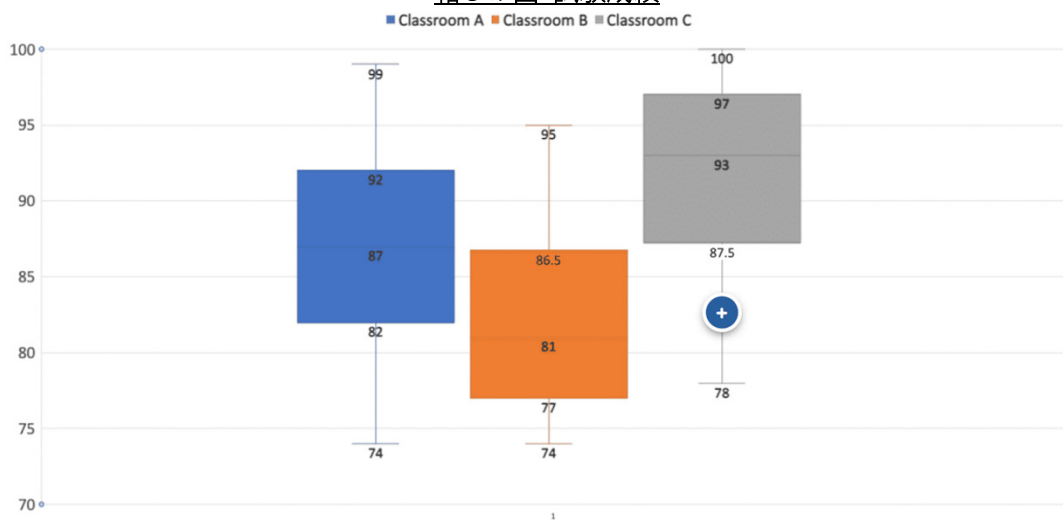
39

Copyright © SAS Institute Inc. All rights reserved.



ひげ

箱ひげ図:試験成績

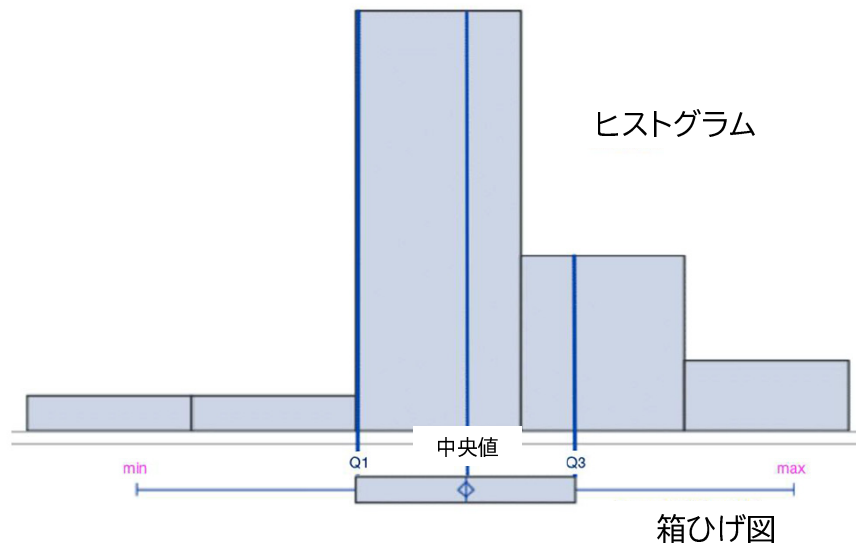


40

Copyright © SAS Institute Inc. All rights reserved.



ヒストグラムと箱ひげ図



41

Copyright © SAS Institute Inc. All rights reserved.



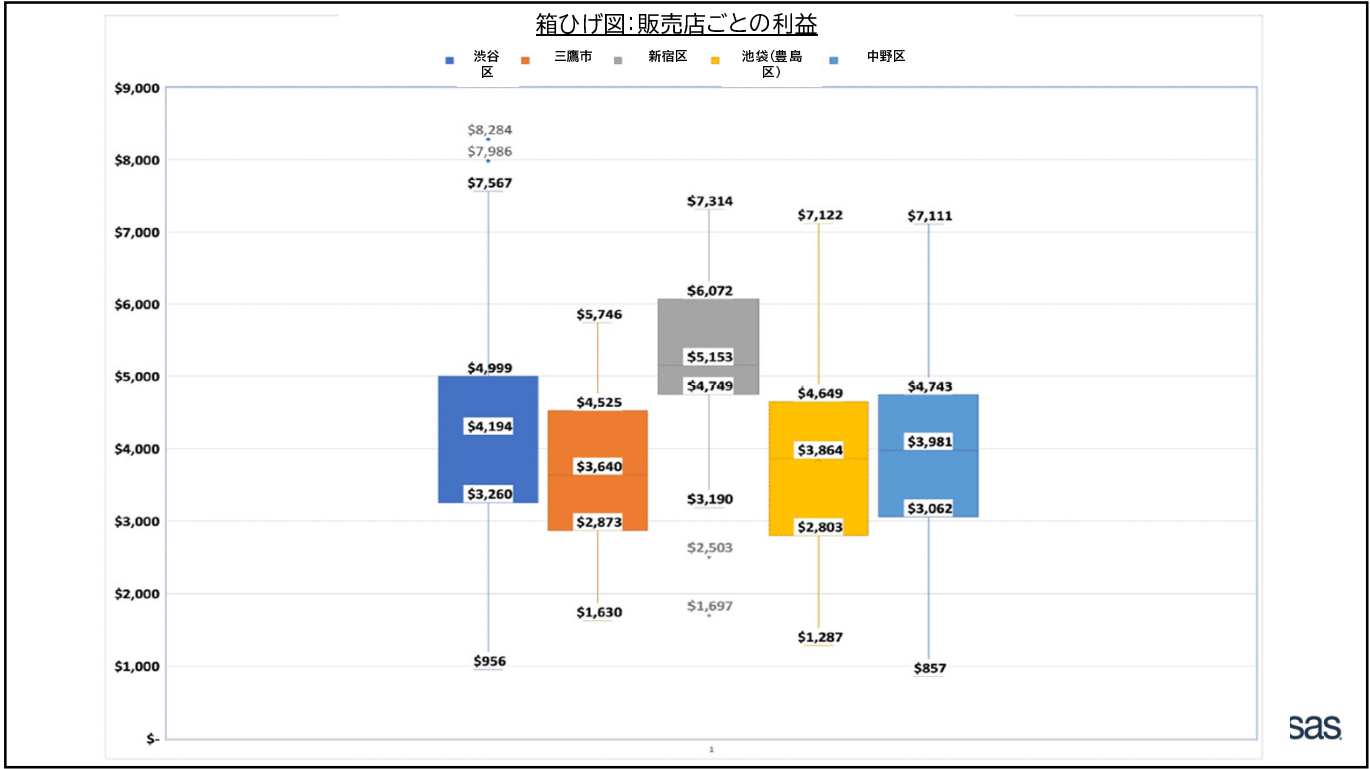
早押しクイズ ○か×か答えなさい

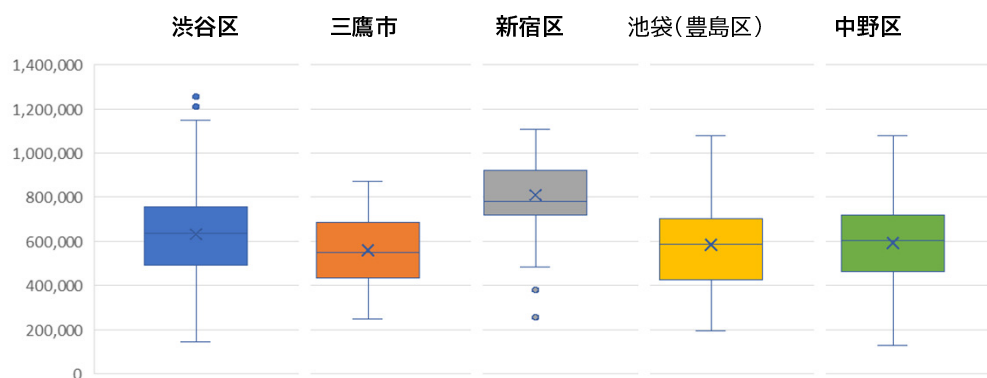
1. データセットをクリーニングする際に対処すべき主な問題は、フォーマット エラー、欠損データ、重複データ、の3つが挙げられる。
2. データの特徴は、分布の形、中心、広がりで表せる。
3. 日常でヒストグラムと箱ひげ図を見たことがあるか？

42

Copyright © SAS Institute Inc. All rights reserved.







45

Copyright © SAS Institute Inc. All rights reserved.



鈴木さんは、探索的データ分析を通してデータについて多くのことを学んだ。

利益データの分布の形状、中心、広がりについて学ぶことで、何が起きているのかという「全体像」をよりよく理解することができた。

47

Copyright © SAS Institute Inc. All rights reserved

鈴木さんの探索的データ分析からの主な収穫

- 全店のデータは正規分布しているが、個々の販売店のデータはそうではない。このことから、鈴木さんは、今後の分析では平均値ではなく中央値を使用する必要があることが分かった。
- データにはいくつかの異常値があった。しかし、どれも極端ではないので、鈴木さんはデータセットから取り除く必要はないと判断した。
- 彼女は、中央値が最も高い新宿区と、最も低い三鷹市で何が起きているのかに興味を湧いた。
- 鈴木さんは、各販売店の利益についてより多くのことを学んだと認識しているが、各販売店が何台の販売を行ったか、また、販売台数が利益にどのような影響を与えるかについて興味が生じた。

47

Copyright © SAS Institute Inc. All rights reserved.



次のステップ: 関係性を探る

鈴木さんのデータセットには多くの変数があるが、彼女は利益の変数のみで探索的データ分析を行った。次の段階では、変数間の関係性に焦点を当てたい。具体的には、彼女は利益とデータセット内の他の変数(時間、データ、販売頻度など)との関係性に分析の焦点を当てる。

例えば、

- 各販売店で、販売台数と利益との関係は？
- 販売日、時間帯は利益や売上にどのような影響を与えるか？

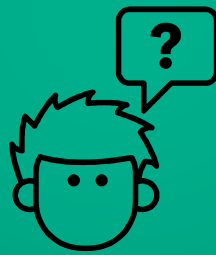
これらの変数間の関係性を見て、新宿区や三鷹市で何が起きているのかの洞察を得るようなパターンがあるのだろうか。

48

Copyright © SAS Institute Inc. All rights reserved.



質疑応答



Copyright © SAS Institute Inc. All rights reserved.

Module 4: 可視化でデータの関係の調査

1. 散布図入門
2. 散布図を深く掘り下げる
3. ヒストグラムと折れ線グラフ
4. すべてをまとめる

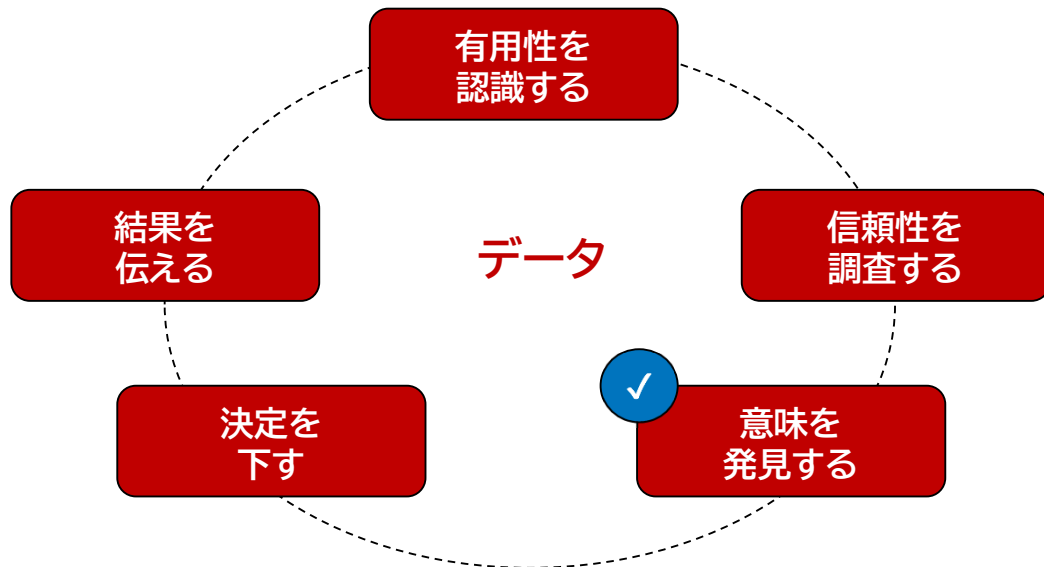


Copyright © SAS Institute Inc. All rights reserved.

A person's hands are holding a tablet computer. The screen shows a 3D bar chart with five bars of different colors (blue, red, green, yellow, and pink) representing data points. The background is a blurred office setting with a plant.

意味を発見するということは、
データ内のパターンを探索し、
そのパターンの意味を理解することです。

データリテラシーのスキル



3

Copyright © SAS Institute Inc. All rights reserved.



変数間の関係の検証

2つの変数間の関係を調べる。独立変数と従属変数の識別が必要

従属変数は、我々が興味関心を持つ結果である。

独立変数は、その結果がなぜそう振る舞うのかを説明するのに役立つ。

言い換えれば、独立変数に何が起こるかによって従属変数が変化するパターンがあるか？独立変数は従属変数に影響を与えるかもしれない。

4

Copyright © SAS Institute Inc. All rights reserved.



従属変数と独立変数の例

疑問	独立変数	従属変数
植物が太陽の光を浴びる量は植物の大きさに関係があるか？	日照時間	植物の大きさ
歯磨き時間は虫歯の本数と関係があるか？	歯磨き時間	虫歯の数
試験勉強をすることで、生徒の試験の成績に変化が見られるか？	試験勉強に費やした時間	試験の成績
カフェイン摂取量と入眠までの時間は関係あるか？	カフェインの摂取量	その人が眠りに落ちるまでにかかる時間

5

Copyright © SAS Institute Inc. All rights reserved.

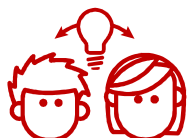


鈴木さんのデータで調べる

疑問	独立変数	従属変数
利益の変動を説明できるデータはどんなものか	販売店	利益
	仕入価格	
	販売価格	
	販売時間	
	販売日	

7

Copyright © SAS Institute Inc. All rights reserved.



グループワーク&報告 ビジネス課題への切り口

先ほどの「疑問とデータ」の整理表は、
店や人、価格、月、時間を「切り口」として
利益との関係を表している。

(一般的に)利益や販売の動きを説明するために、
他に考えられる切り口は何が挙げられるか？



8

Copyright © SAS Institute Inc. All rights reserved.



棒グラフ

鈴木さんは棒グラフを使って調査したい2つの重要な質問がある：

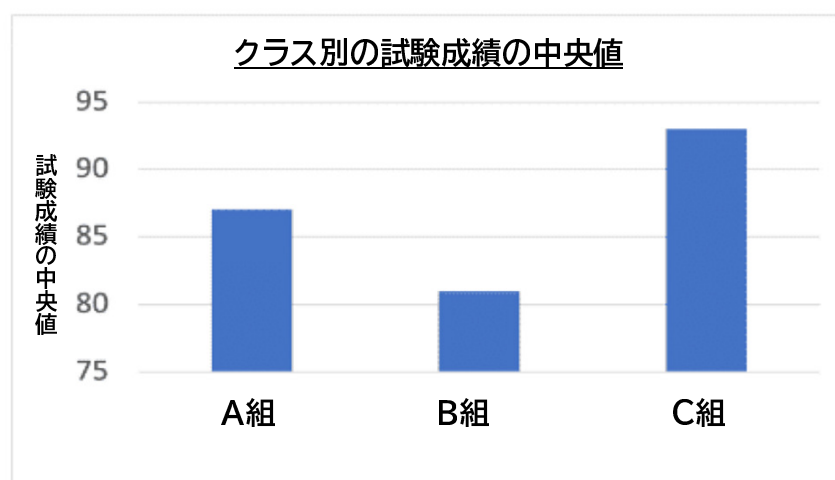
- 販売店と利益の関係は？
- 販売店と販売台数の関係は？

9

Copyright © SAS Institute Inc. All rights reserved.



棒グラフ

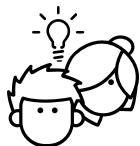


10

Copyright © SAS Institute Inc. All rights reserved.



鈴木さんは棒グラフを使用して
データ内の関係性を可視化する準備が
できています。

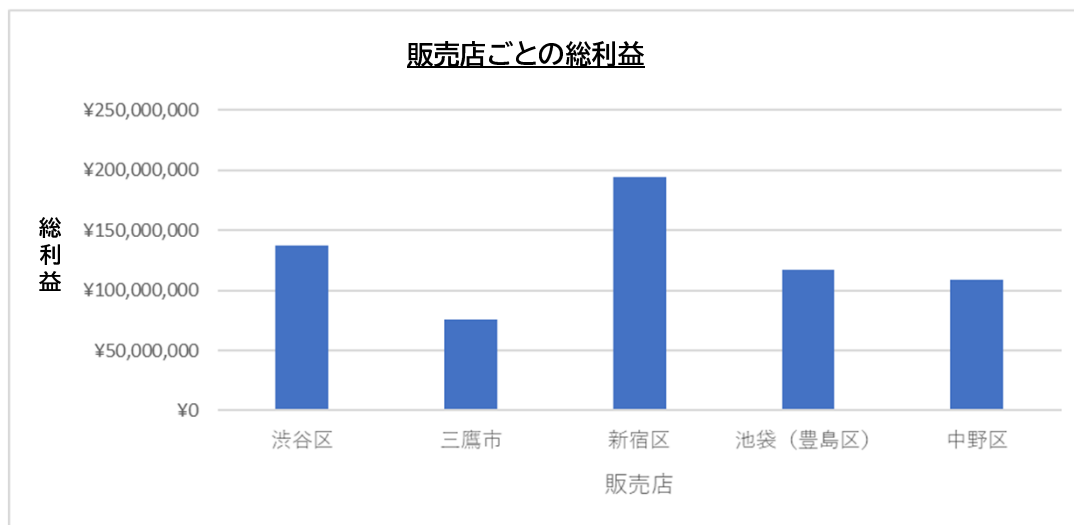


各販売店のデータで棒グラフを描く

下表は利益と販売台数に対して販売店ごとに集計したものである。

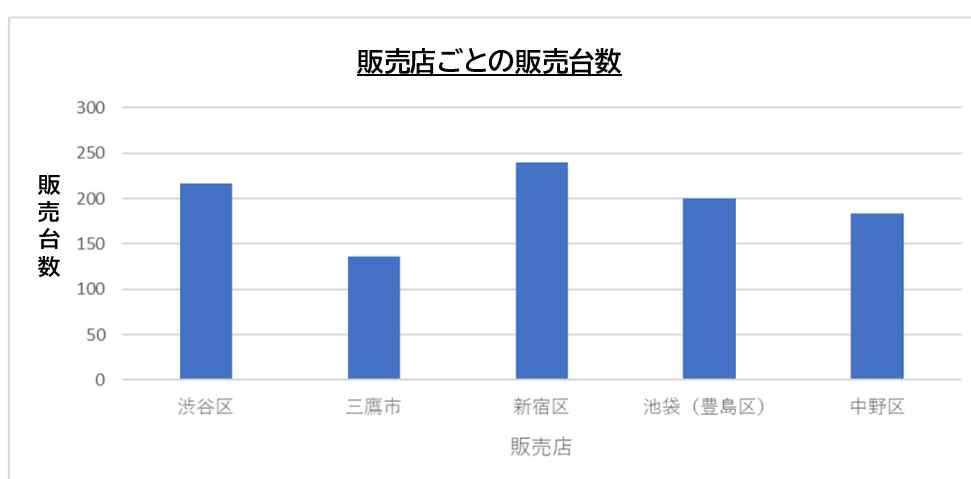
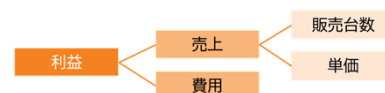
- ・統計指標は「利益」と「販売台数」
- ・販売店間、「渋谷区」「三鷹市」「新宿区」「池袋」「中野区」を比較してみましょう

販売店	利益	販売台数
渋谷区	¥137,413,485	217
三鷹市	¥76,106,970	136
新宿区	¥194,286,212	240
池袋(豊島区)	¥116,794,091	200
中野区	¥109,174,545	184
全販売店	¥633,775,303	977



13

Copyright © SAS Institute Inc. All rights reserved.

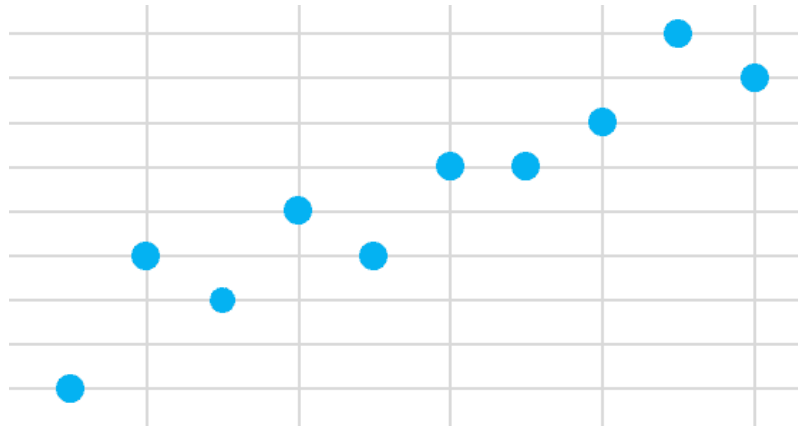


14

Copyright © SAS Institute Inc. All rights reserved.



散布図を使用して関係性を調べる



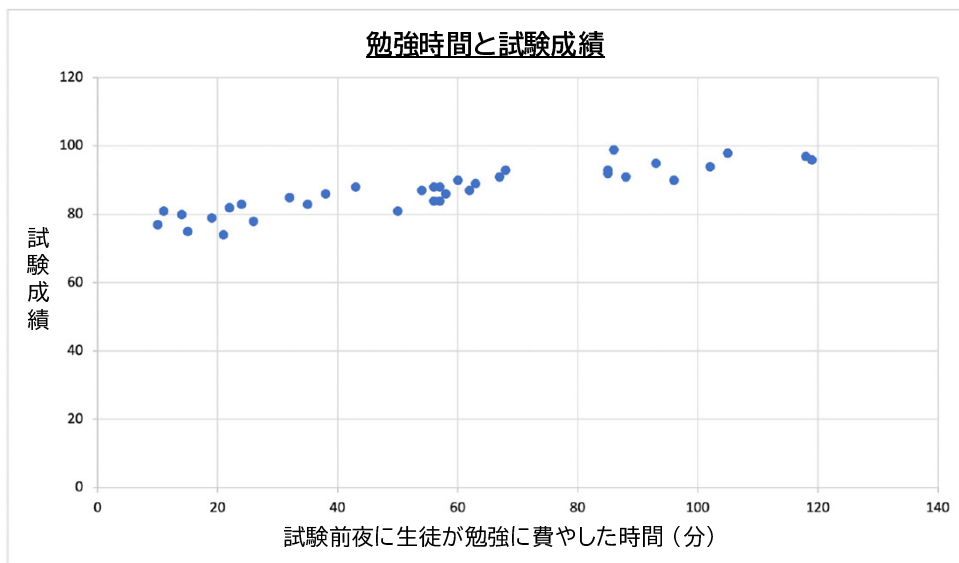
15

Copyright © SAS Institute Inc. All rights reserved.



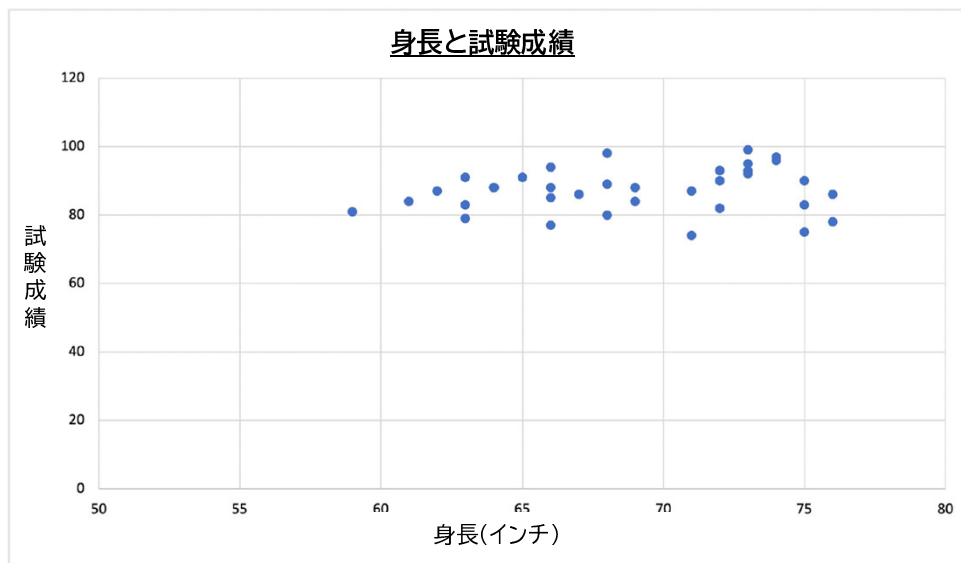
	A	B	C	D	E
1	試験成績	試験前夜に勉強に費やした時間 (分)	身長 (インチ)	学期中の欠席回数	
2	74	21	71	9	
3	75	15	75	7	
4	77	10	66	8	
5	78	26	76	6	
6	79	19	63	6	
7	80	14	68	7	
8	81	11	59	7	
9	81	50	59	5	
10	82	22	72	4	
11	83	24	63	4	
12	83	35	75	5	
13	84	57	69	4	
14	84	56	61	3	
15	85	32	66	3	
16	86	58	67	6	
17	86	38	76	3	
18	87	54	62	2	
19	87	62	71	0	
20	88	57	66	1	
21	88	43	64	1	
22	88	56	69	2	
23	89	63	68	2	





17

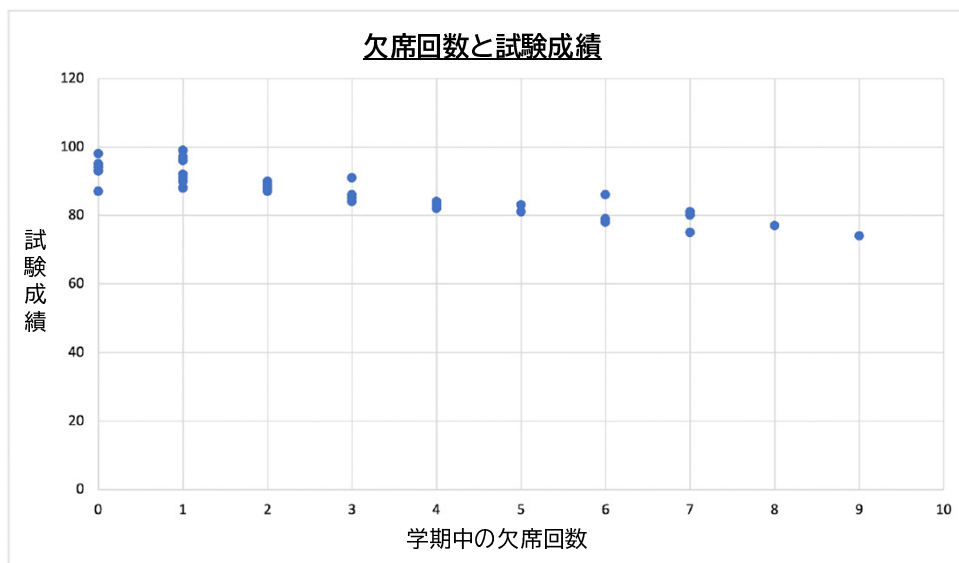
Copyright © SAS Institute Inc. All rights reserved.



18

Copyright © SAS Institute Inc. All rights reserved.





19

Copyright © SAS Institute Inc. All rights reserved.

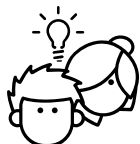


鈴木さんのデータで調べる

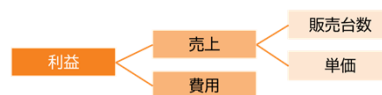
利益と販売台数の間に関係はあるか

21

Copyright © SAS Institute Inc. All rights reserved.



散布図をご覧ください



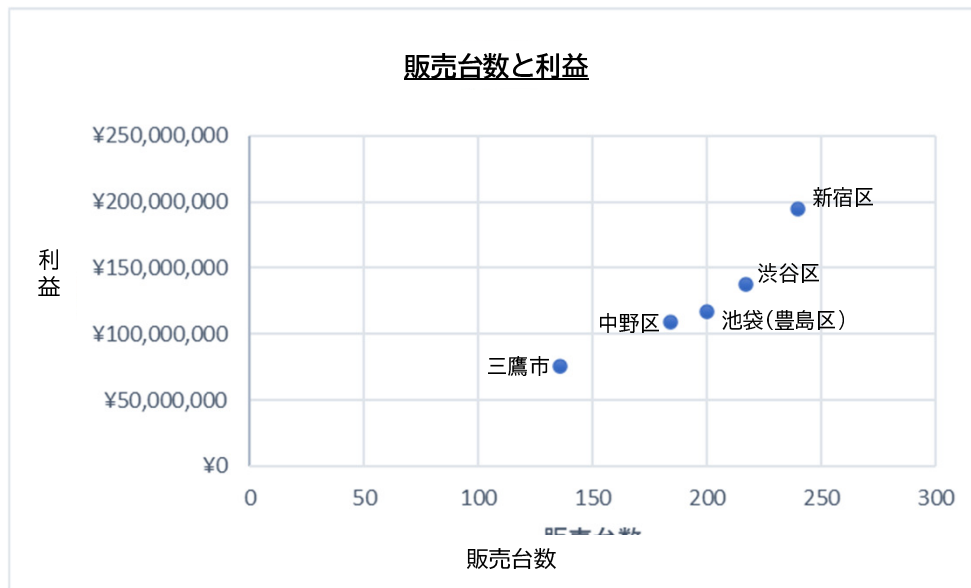
このシートは利益(従属変数)と販売台数(独立変数)に対して販売店ごとに集計したものである。両者の関係について散布図を描いて調べましょう。

販売店	利益	販売台数
渋谷区	¥137,413,485	217
三鷹市	¥76,106,970	136
新宿区	¥194,286,212	240
池袋(豊島区)	¥116,794,091	200
中野区	¥109,174,545	184
全販売店	¥633,775,303	977

22

Copyright © SAS Institute Inc. All rights reserved.





23

Copyright © SAS Institute Inc. All rights reserved.



線形性と方向性、スパース性

散布図で調べていること:

線形性: 関係は線形か、非線形か? 線形の場合は、関係がどのくらい強いのか? 線は比較的まっすぐか、それとも点の間に大きなばらつきがあるか?

方向性: 関係が線形の場合、それは正か、負か? 正の関係は右上がりが増加するのに対し、負の関係は右下がりに減少する。

スパース性: より大きなデータセットを調べると、データが満遍なく散らばり、ばらつきがあることに気付くかもしれない。例えば、グラフの中には、多くのポイントが密集しているものもあれば、ポイントが少ない領域があるものもある。

24

Copyright © SAS Institute Inc. All rights reserved.



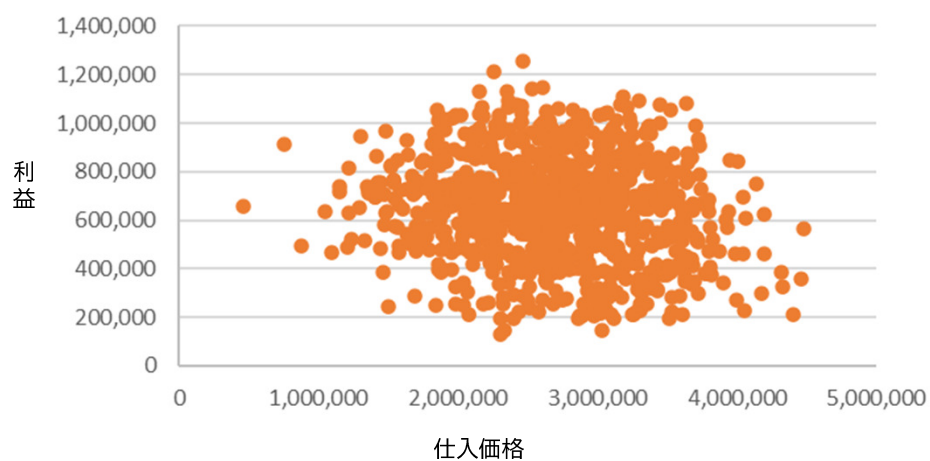
利益と仕入価格の間に関係はあるか。

25

Copyright © SAS Institute Inc. All rights reserved.



利益と仕入価格:全販売店

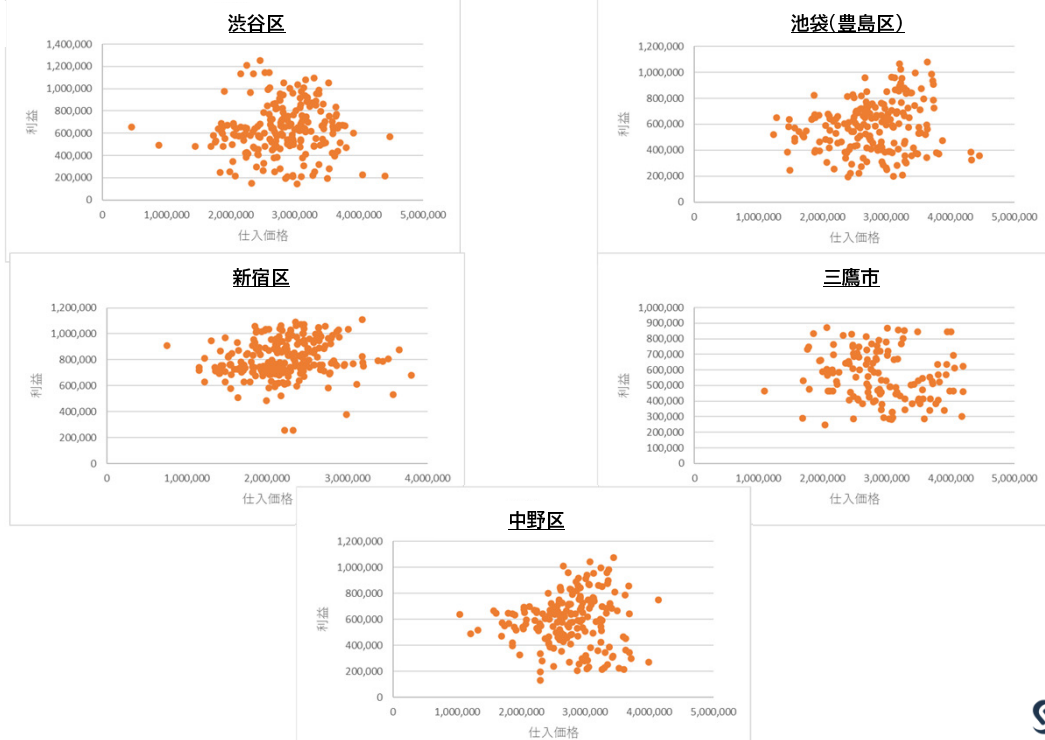


26

Copyright © SAS Institute Inc. All rights reserved.

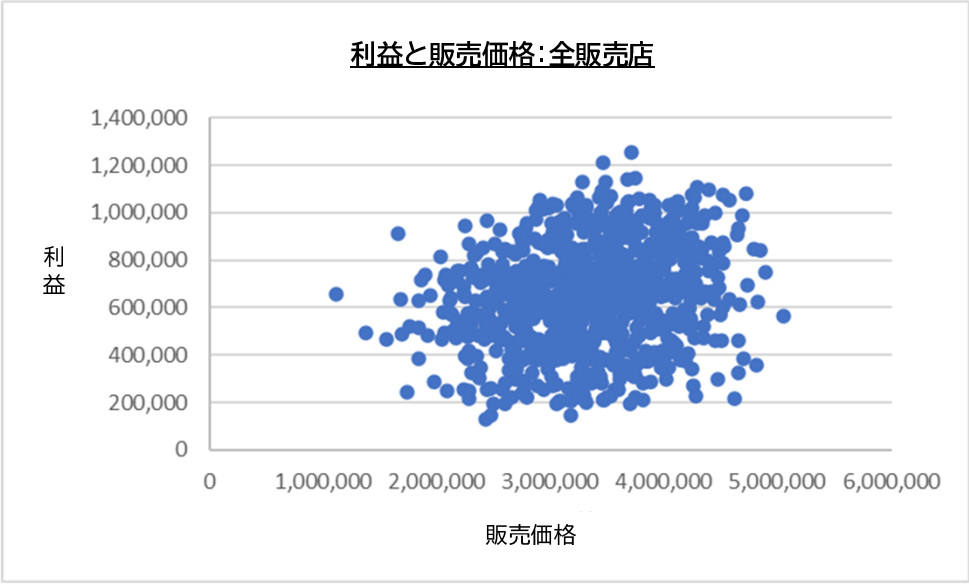


利益と仕入価格



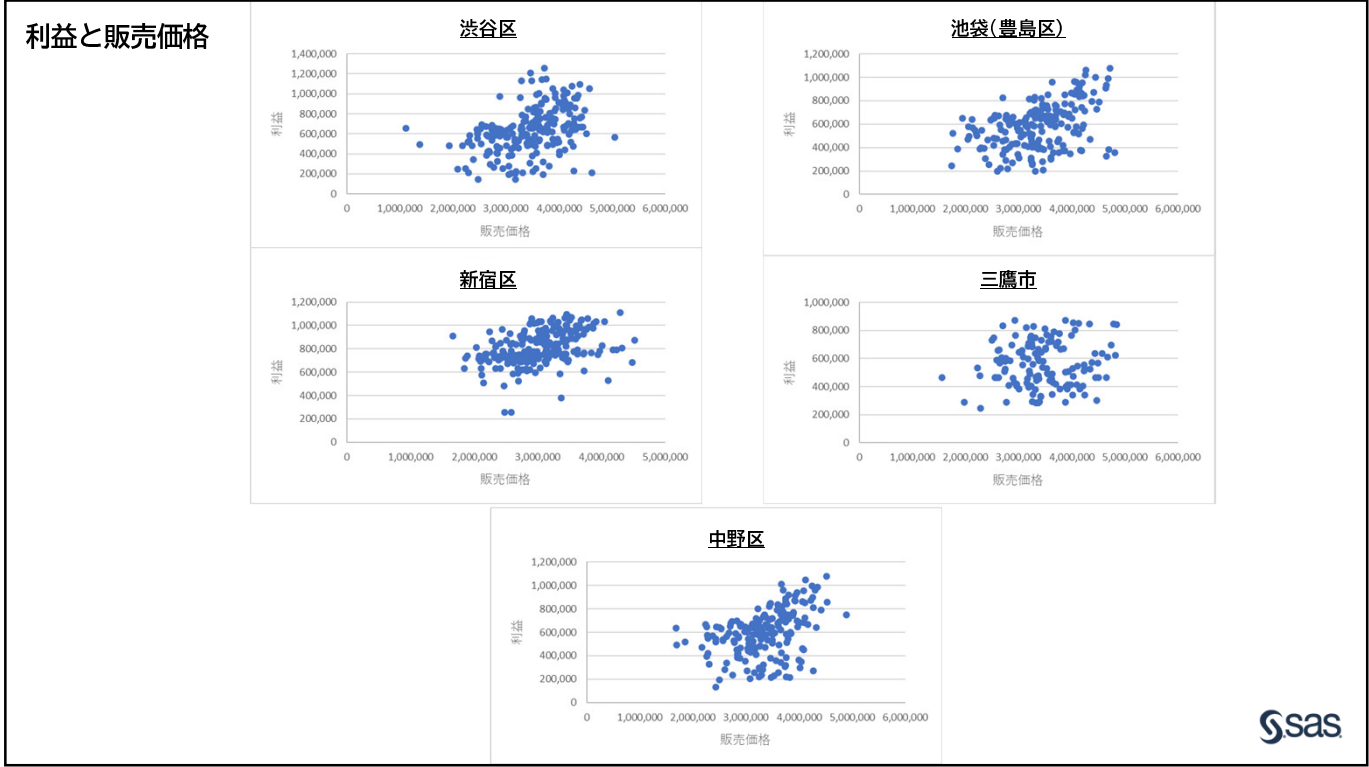
利益と販売価格の間に関係はあるか。





29

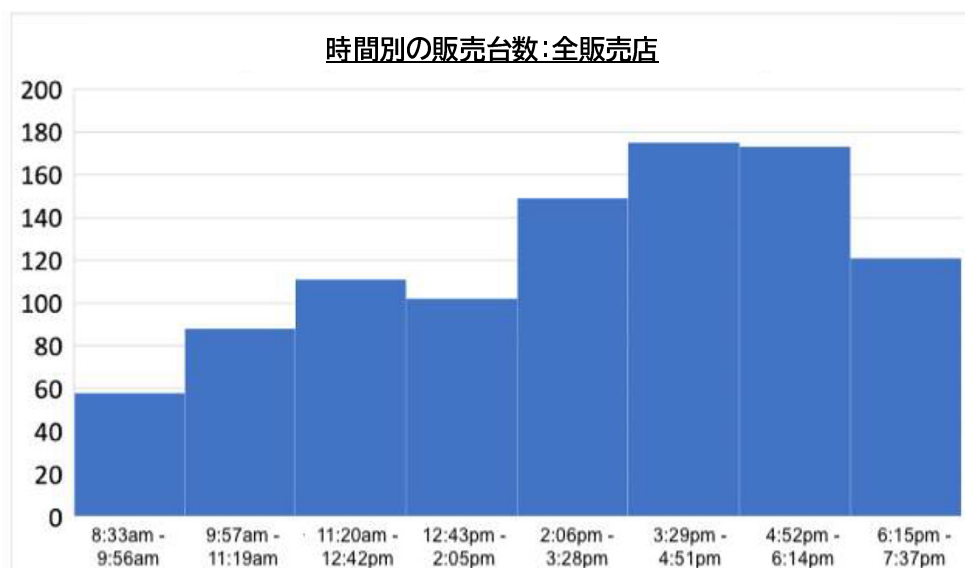
Copyright © SAS Institute Inc. All rights reserved.



鈴木さんは分析でもう1つ重要な変数を調べています。
それは時間です。

彼女は、時間帯や季節によって売上と利益が
どのように変化するかを知りたいと思っています。

販売台数に時間帯別に違いがあるか

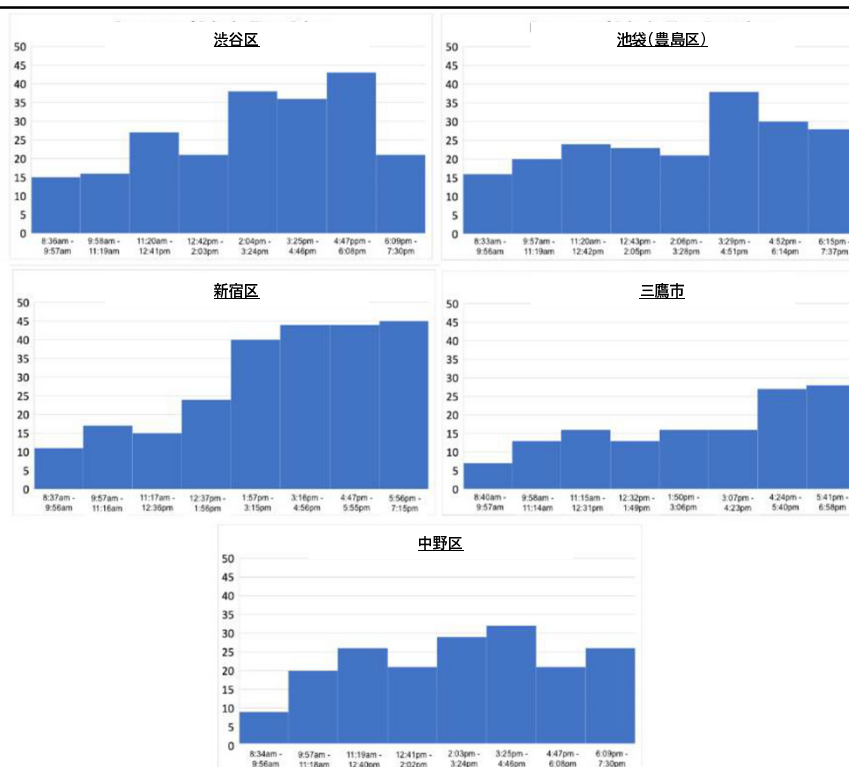


33

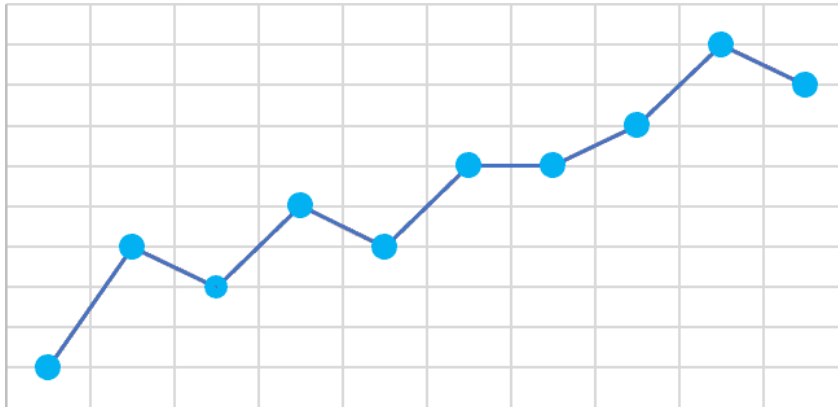
Copyright © SAS Institute Inc. All rights reserved.



時間別の販売台数

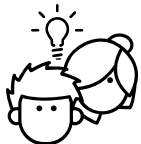


折れ線グラフによる販売と利益



35

Copyright © SAS Institute Inc. All rights reserved.



各販売店のデータで棒グラフを描きなさい

下表は月ごと(独立変数)に利益と販売台数(従属変数)を集計したものである。

- ・統計指標:「販売台数」と「利益」
- ・さらに販売店ごとに描いて比較してみましょう

月	販売台数	利益
1月	82	¥55,237,424
2月	70	¥44,843,788
3月	82	¥52,563,939
4月	80	¥51,305,909
5月	77	¥51,801,818
6月	81	¥54,890,455
7月	88	¥53,751,818
8月	85	¥53,593,939
9月	76	¥49,713,939
10月	76	¥48,414,848
11月	85	¥55,914,545
12月	95	¥61,742,879

36

Copyright © SAS Institute Inc. All rights reserved.



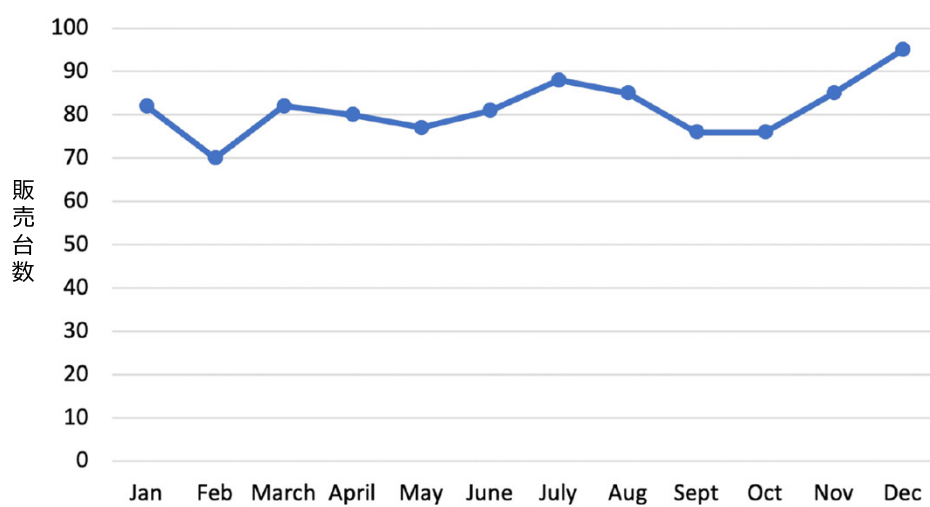
月ごとに販売台数に違いがあるか

37

Copyright © SAS Institute Inc. All rights reserved.



月ごとの販売台数:全販売店

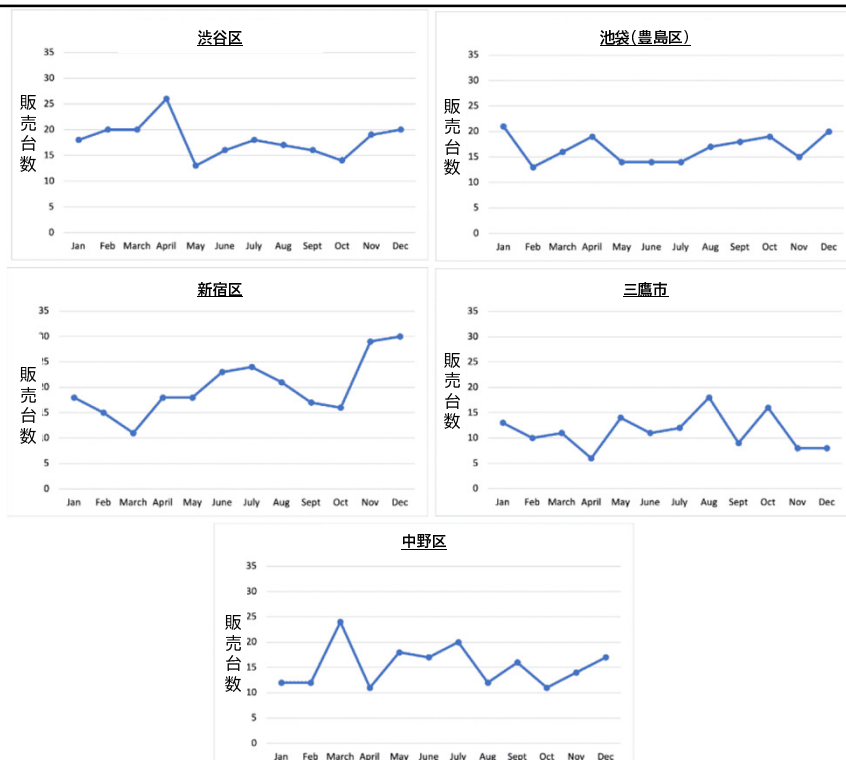


38

Copyright © SAS Institute Inc. All rights reserved.

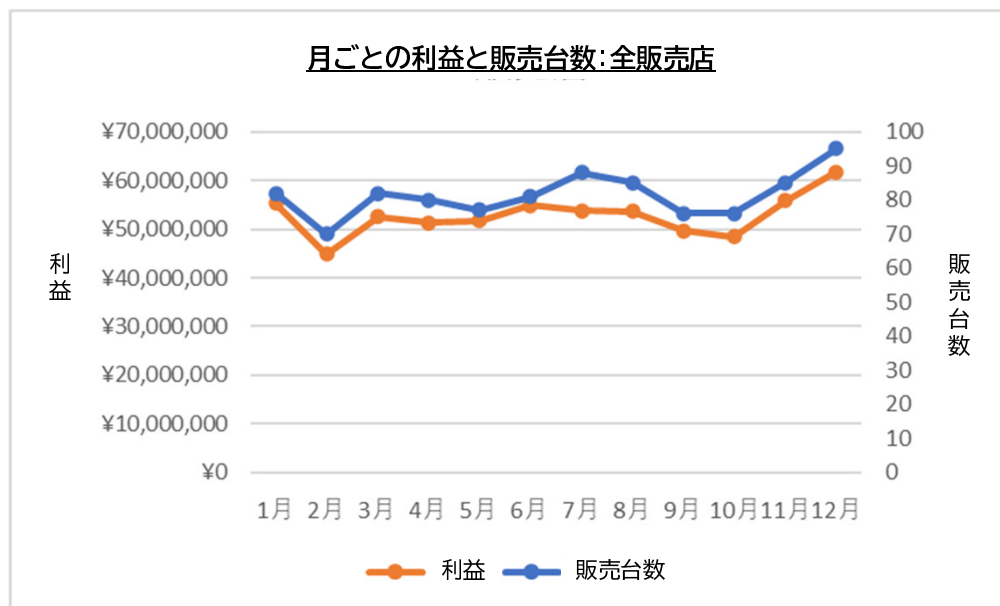


月ごとの販売台数



月ごとに販売台数や利益に違いがあるか



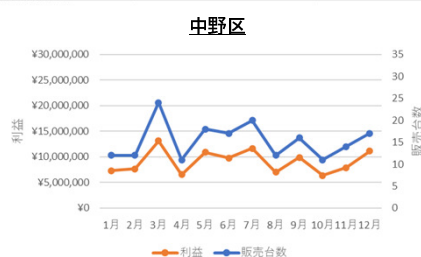
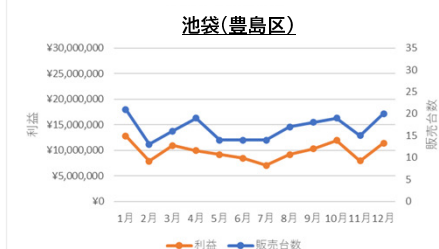
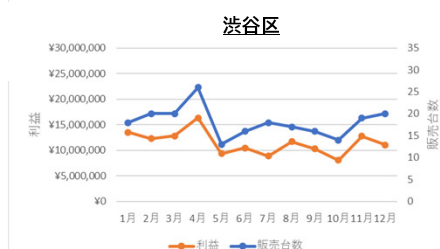


41

Copyright © SAS Institute Inc. All rights reserved.



月ごとの利益と販売台数





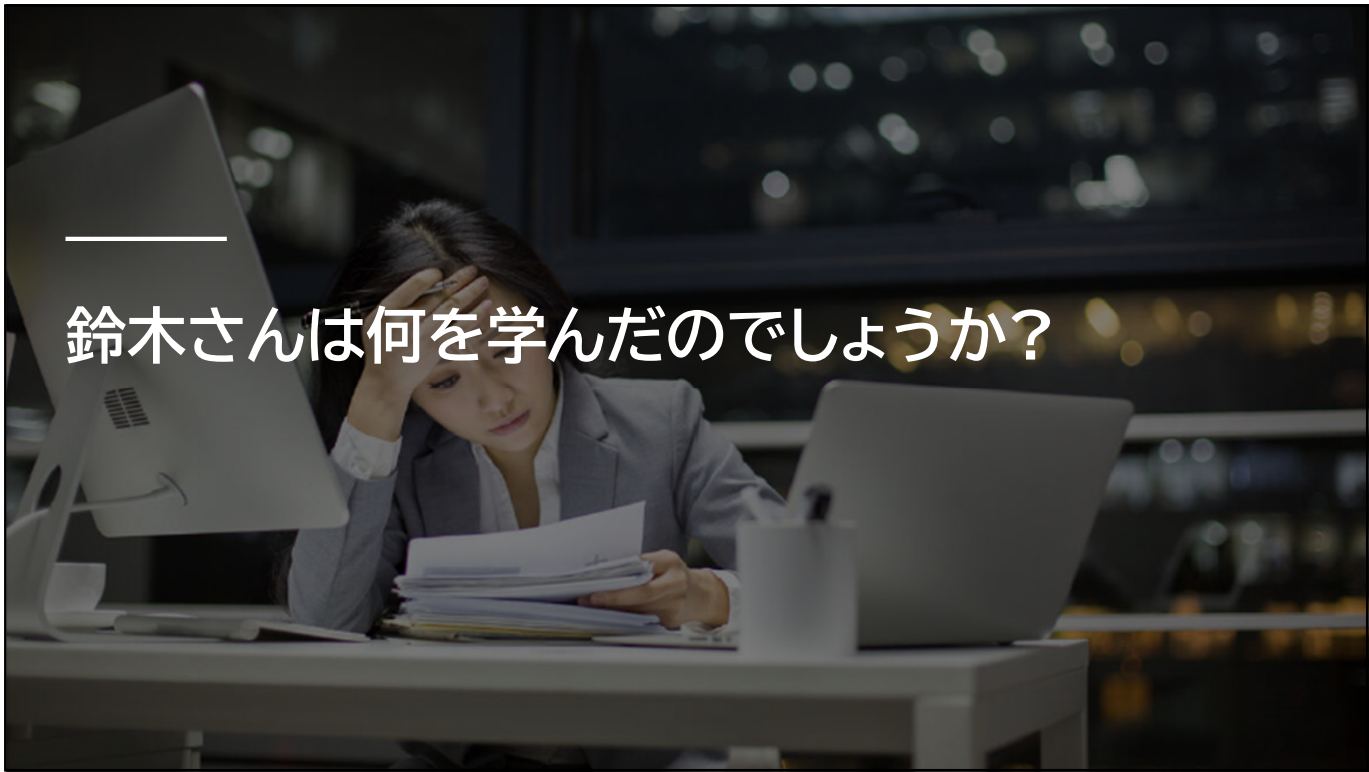
早押しクイズ

○×で答えなさい

1. 棒グラフは、一つのデータ(変数)をグループ間で比較するときに適したグラフである。一方データが複数ある場合は折れ線グラフの方が見やすい。
2. 箱ひげ図は、ヒストグラムの要約版と言えるもので、データのばらつき具合をみえる化し、複数データの比較に適したグラフである。
3. 散布図は、データ(変数)とデータ(変数)の関係を表す典型的なグラフである。

43

Copyright © SAS Institute Inc. All rights reserved.



鈴木さんは何を学んだのでしょうか？

鈴木さんの分析から分かったこと

1. 最も業績の良い販売店は新宿区で、最も業績の悪いのは三鷹市、両者の利益には、大きな差があった。

新宿区: ¥194,286,212

三鷹市: ¥76,106,970

2. 販売価格のほうがか仕入価格よりも利益と強い関係があった。

3. 販売台数が最も少ない時間帯と最も多い時間帯があった。

販売台数が最も少ない時間帯: 午前10時前

販売台数が最も多い時間帯: 午後(2時)から夕方にかけて

4. 1年のうち何ヶ月かに販売台数が大きく急増することがあるが、同じ月に急増したり急減したりしないため、明らかな傾向はなかった(ある販売店は3月、別の販売店は夏場に、他の販売店は11月と12月に販売台数が急増)。

鈴木さんは、まだ疑問を持っている。

- ・なぜ新宿区の業績は他の販売店より良いのか？
- ・ある月に売上と利益が急増したのはなぜか？
- ・時間帯によって販売店の活動はどのように違うのか？たとえば、何人のお客が店内を見て回り、何人の販売員が店頭に立っているのか。

鈴木さんは、これは他の販売店で使えるベストプラクティスを特定するのに役立つと考えている。一部の販売店で売上が急増する理由を特定することで、鈴木さんは全販売店に導入することができる。

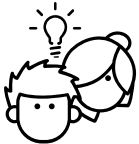
より多くの情報を収集するため、鈴木さんは各販売店を訪れ、観察し、ヒアリングすることにした。これは、彼女の発見を文脈化するのに役立つだろう！

あれから鈴木さんが学んだこと

- 鈴木さんは、新宿区の売上の20%近くは、オンライン販売によるもので、他のオンライン・プレゼンスよりもはるかに多くのリソースを投入していることが分かった。透明性のある価格開示、販売店レビューの場所、お問い合わせフォームの設置など。他の販売店もすべて在庫を掲載したウェブサイトを持っているが、いずれもナビゲートしにくく、定期的に更新されておらず、高品質の画像を提供していない。
- 鈴木さんは、渋谷区の4月の売上急増、中野区の3月の急増、そして新宿区の11月と12月の急増について各販売店を訪れ、マネジャーに尋ねた。これらの月のスパイクは、いずれも、各販売店が主催した販促によって説明できることを知った。
- 鈴木さんは販売店を訪問し、三鷹市が午後6時に、他の4つの販売店は午後7時に閉店することを知った。しかし、彼女はデータ分析から、すべての販売店が閉店後に売上を計上していることを知っている。
- 彼女はまた、観察を通じて、ある時間に働いている販売員の数が、顧客の数に基づいていないことを知る。たとえば、渋谷区では、客入りに関係なく、すべての営業時間を通して、店頭で0.5人の販売員がいる。

47

Copyright © SAS Institute Inc. All rights reserved.



早押しクイズ ○か×か答えなさい

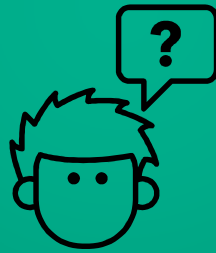
- 棒グラフは一つのデータ(変数)をグループ間で比較するときに適したグラフである。一方、データが複数ある場合は、折れ線グラフの方が見やすい。
- 箱ひげ図は、ヒストグラムの要約版と言えるもので、データのばらつき具合をみえる化し、複数データの比較に適したグラフである。
- 散布図は、データとデータの関係を表す典型的なグラフである。

48

Copyright © SAS Institute Inc. All rights reserved.



質疑応答



Copyright © SAS Institute Inc. All rights reserved.

Module 5: 文脈化とコミュニケーション

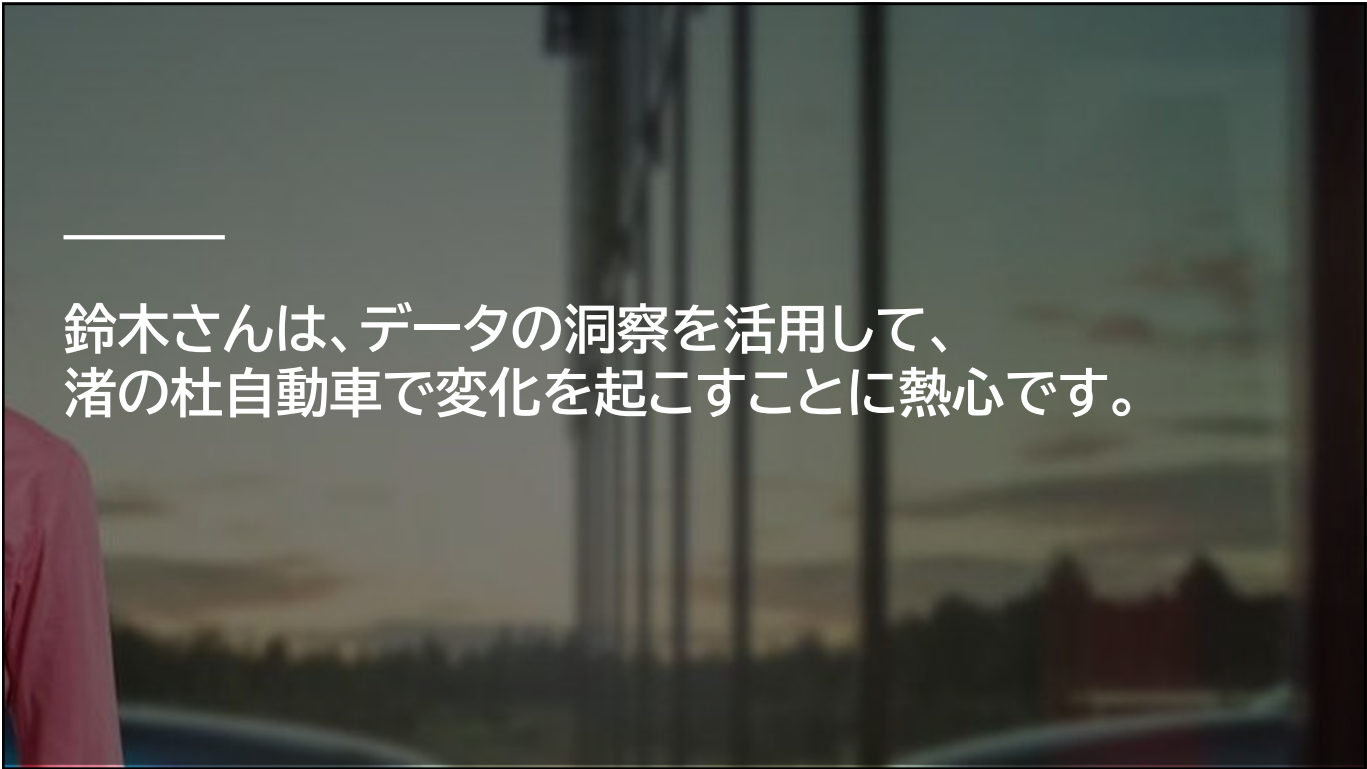
1. データ・リテラシーとのつながり
2. 結果の文脈化
3. データに基づいた意思決定
4. 洞察と決断を伝える
5. すべてのまとめる



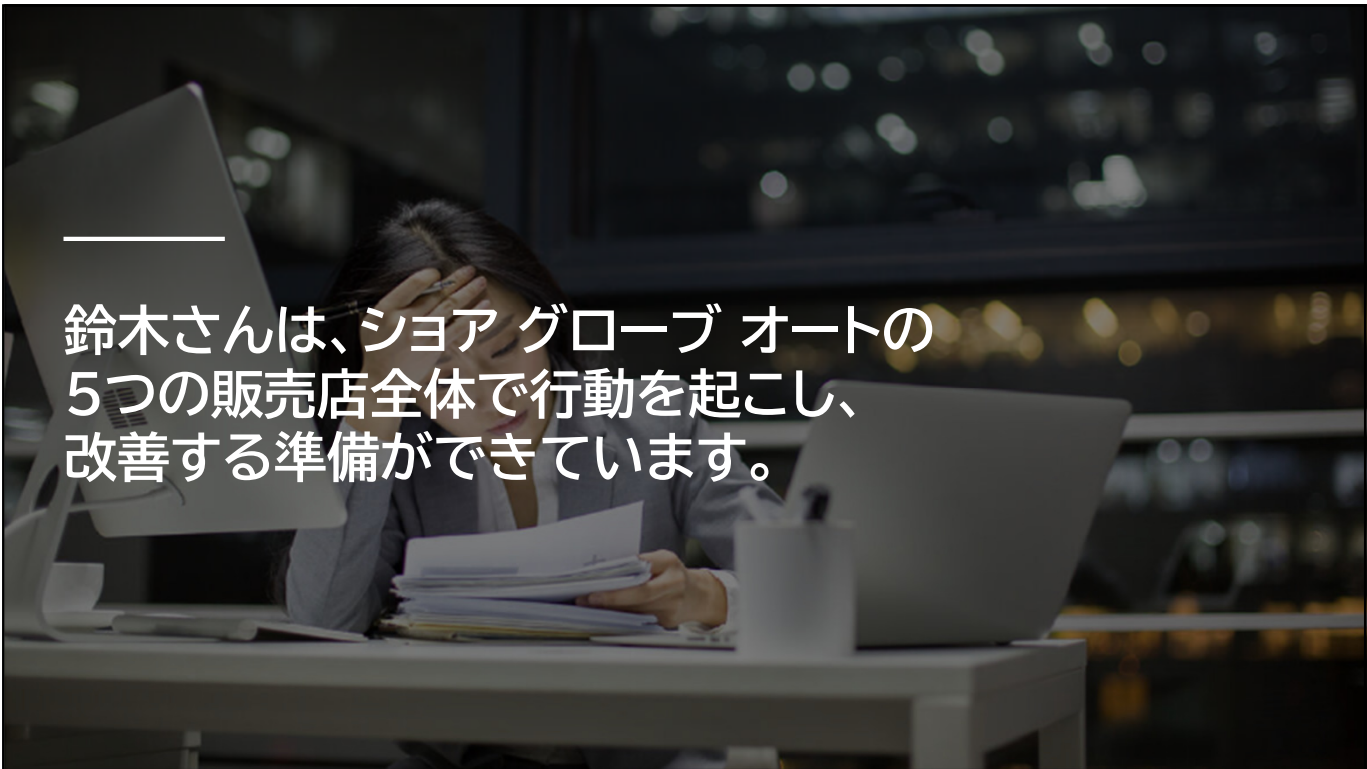
Copyright © SAS Institute Inc. All rights reserved.

A person's hands are holding a tablet computer. The screen shows a 3D bar chart with five bars of different colors (blue, red, green, yellow, pink) and numerical values on the x-axis (350, 400, 300, 200). The background is a blurred office setting with a plant.

データの洞察は、意思決定を行い、
その決定を他の人に伝えるのに
役立ちます。



鈴木さんは、データの洞察を活用して、
渚の杜自動車で変化を起こすことに熱心です。



鈴木さんは、シヨア グローブ オートの
5つの販売店全体で行動を起こし、
改善する準備ができています。

鈴木さんの発見要約

- ① 販売店の業績
- ② 月次スパイク
- ③ 時間帯
- ④ 価格戦略

5

Copyright © SAS Institute Inc. All rights reserved.



①販売店の業績

鈴木さんは、新宿区が圧倒的に業績が高く、2番目に業績の高い販売店の渋谷区よりも30%近く高い業績を上げていることを知った。

また、三鷹市が最も業績の低い販売店であることもわかった。例えば、トップの新宿区の利益は194百万円であるのに対し、三鷹市の利益はわずか76百万円である。

6

Copyright © SAS Institute Inc. All rights reserved.



②月次スパイク

鈴木さんの分析によると、販売台数と利益はいずれも、1年のうちいくつかの月に大きなスパイクを見せたが、すべての販売店で同じ月にスパイクしたり落ち込んだりするわけではなかった。

販売店みんなで統一した、明らかな傾向はなかったが、販売台数と利益が月によって変化しているのは確かである。

7

Copyright © SAS Institute Inc. All rights reserved.



③時間帯

鈴木さんは、販売店の朝は遅く、忙しくなり始める午後2時までは販売台数が少ないことを知った。

鈴木さんは、開店と閉店の時間帯を変更することで、次年度のコストを削減できると考えている。

8

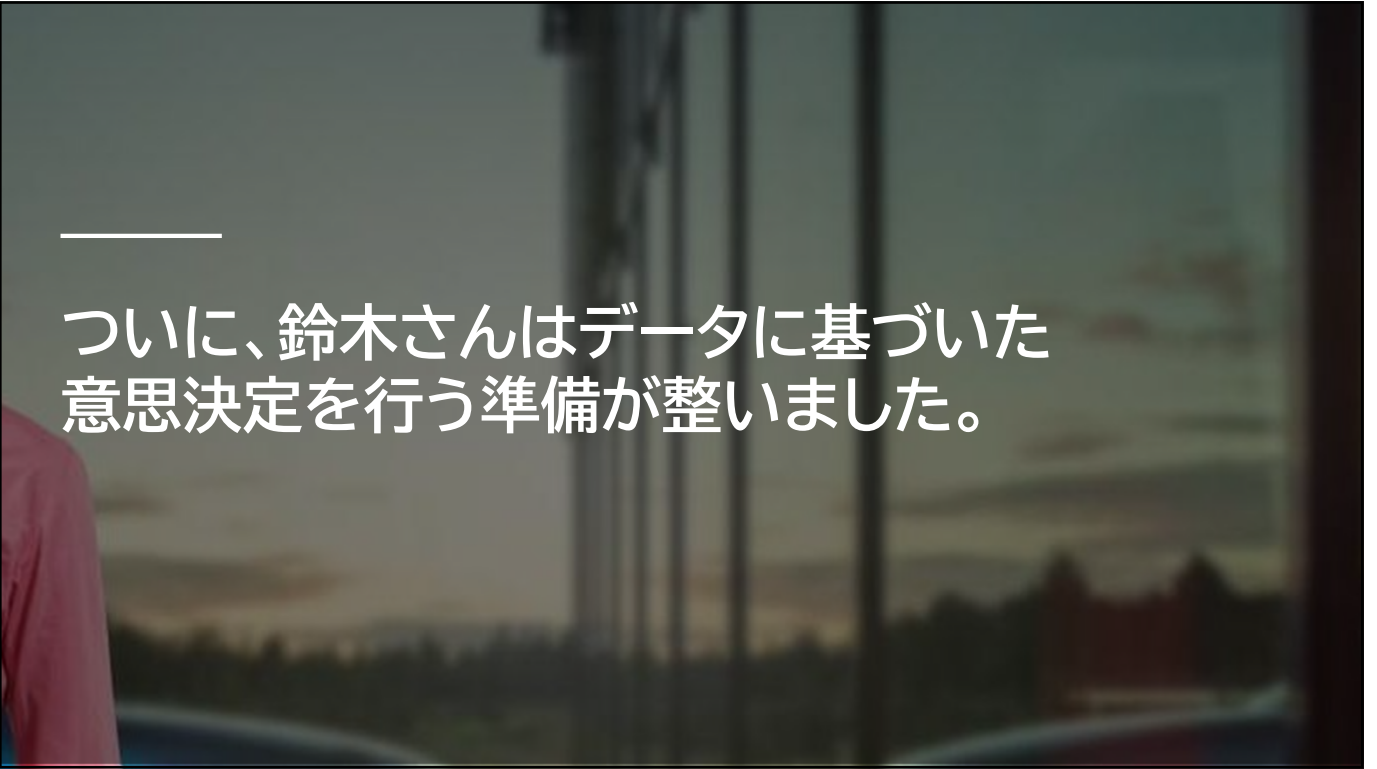
Copyright © SAS Institute Inc. All rights reserved.



④価格戦略

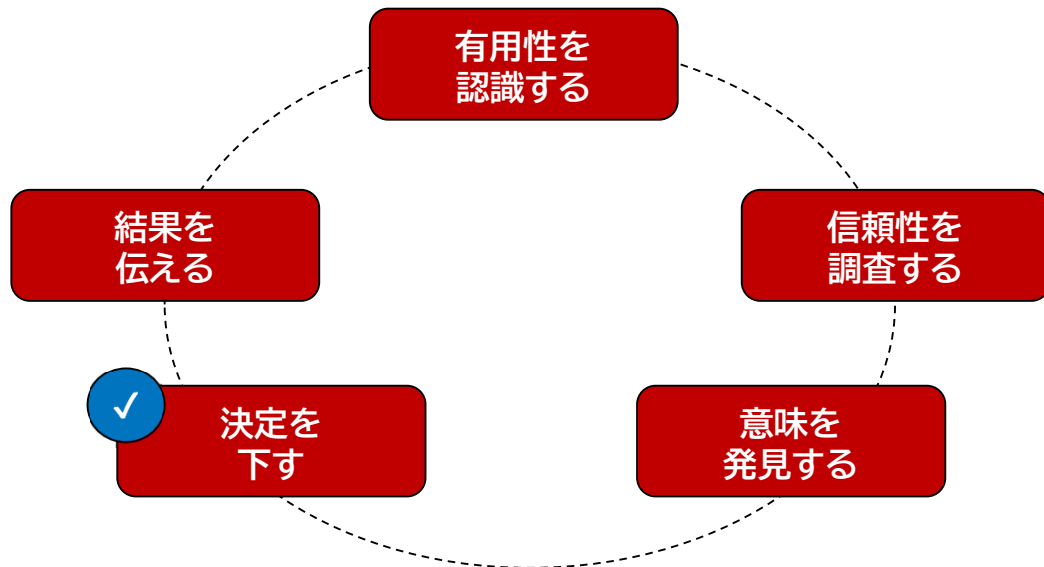
販売価格は、仕入価格よりも利益と強い関係があった。販売店が価格設定のために特定の戦略を用いている可能性を示唆している。

鈴木さんは、何かを決定する前に、何が起きているのかをよりよく理解するために、仕入価格と販売価格に関する追加のデータ分析を行う必要があることに気づいた。



ついに、鈴木さんはデータに基づいた
意思決定を行う準備が整いました。

データリテラシーのスキル



11

Copyright © SAS Institute Inc. All rights reserved.



鈴木さんのデータに基づいた決定

広範なデータ分析により、彼女はビジネスについてまったく新しい理解を深め、利益にプラスの影響を与えるベストプラクティスを特定し、効果・効率を高めることができる分野を選定することができた。

データ分析から得た洞察をもとに、鈴木さんは来年度の利益向上に役立つと思われる重要な決断をした。

1. 年に一度の販促を開催する。
2. 営業時間を延長する。
3. ダイナミック・スタッフィングを導入する。
4. 質の高いオンラインプレゼンスを確立する。
5. データ主導の文化を発展させる。

12

Copyright © SAS Institute Inc. All rights reserved.



年に一度
販促を開催

1

各販売店は、翌年中に少なくとも1回は販促を開催、
但し、競合しないよう、時期をずらして行う。

年に一度
販促を開催

1

各販売店は、翌年中に少なくとも1回は販促を開催、
但し、競合しないよう、時期をずらして行う。

営業
時間を延長

2

すべての販売店が営業時間を午後8時まで延長する。

年に一度
販促を開催

1

各販売店は、翌年中に少なくとも1回は販促を開催、
但し、競合しないよう、時期をずらして行う。

営業
時間を延長

2

すべての販売店が営業時間を午後8時まで延長する。

ダイナミック
スタッフィングを導入

3

各販売店は、暇な時間帯には販売員の数を減らし、
忙しい時間帯には販売員の数を増やす。

15

Copyright © SAS Institute Inc. All rights reserved.



年に一度
販促を開催

1

各販売店は、翌年中に少なくとも1回は販促を開催、
但し、競合しないよう、時期をずらして行う。

営業
時間を延長

2

すべての販売店が営業時間を午後8時まで延長する。

ダイナミック
スタッフィングを導入

3

各販売店は、暇な時間帯には販売員の数を減らし、
忙しい時間帯には販売員の数を増やす。

質の高いオンライン
プレゼンスを確立

4

各販売店は、新宿区のウェブサイトモデルとし、より
強力なオンラインプレゼンスを確立する。

16

Copyright © SAS Institute Inc. All rights reserved.



年に一度
販促を開催

1

各販売店は、翌年中に少なくとも1回は販促を開催、
但し、競合しないよう、時期をずらして行う。

営業
時間を延長

2

すべての販売店が営業時間を午後8時まで延長する。

ダイナミック
スタッフィングを導入

3

各販売店は、暇な時間帯には販売員の数を減らし、
忙しい時間帯には販売員の数を増やす。

質の高いオンライン
プレゼンスを確立

4

各販売店は、新宿区のウェブサイトモデルとし、より
強力なオンラインプレゼンスを確立する。

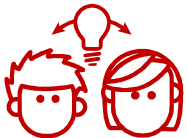
データ主導の
文化を発展

5

各販売店は、販売に関する確固としたデータを収集し、
日常のおよび中長期的な意思決定に活用する。

17

Copyright © SAS Institute Inc. All rights reserved.



グループワーク&報告

疑問/仮説(切り口)とほしいデータは？

ビジネス課題

- ・実績を踏まえて来期の販売の見通しを立てたい。
- ・売れ行きががんばしくなく売上や利益を伸ばしたい。
- ・そのためにどんな施策が打てるかリストアップしたい。

疑問/仮説(切り口)と
ほしいデータ

- ・そもそも販売の動向はどうか。
- ・個々の販売店の業績はどうか。
- ・販売が多い季節や時間帯はあるか。
- ・
- ・

分析結果と
解決策

利益

売上

費用

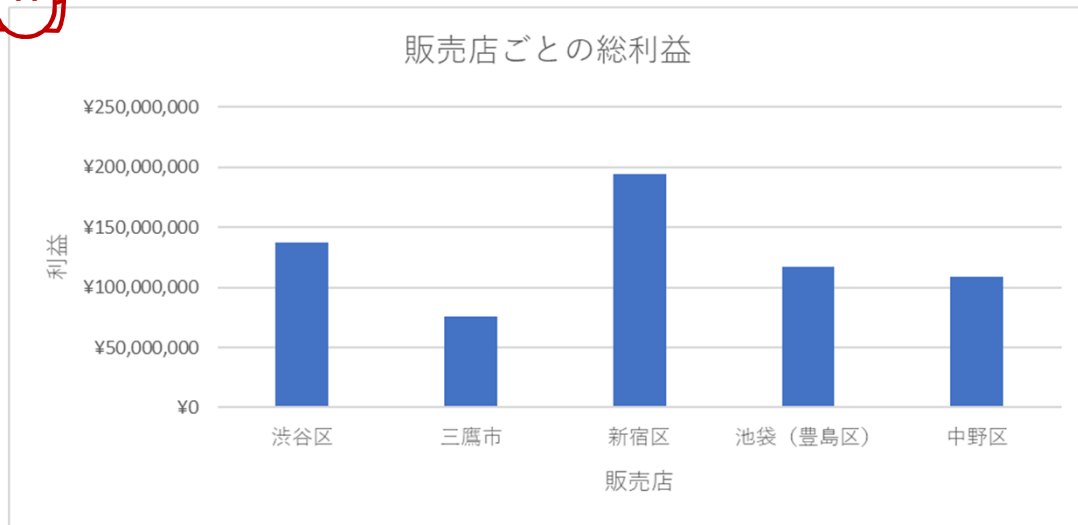
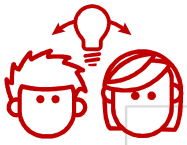
販売台数

単価

18

Copyright © SAS Institute Inc. All rights reserved.





19

Copyright © SAS Institute Inc. All rights reserved.



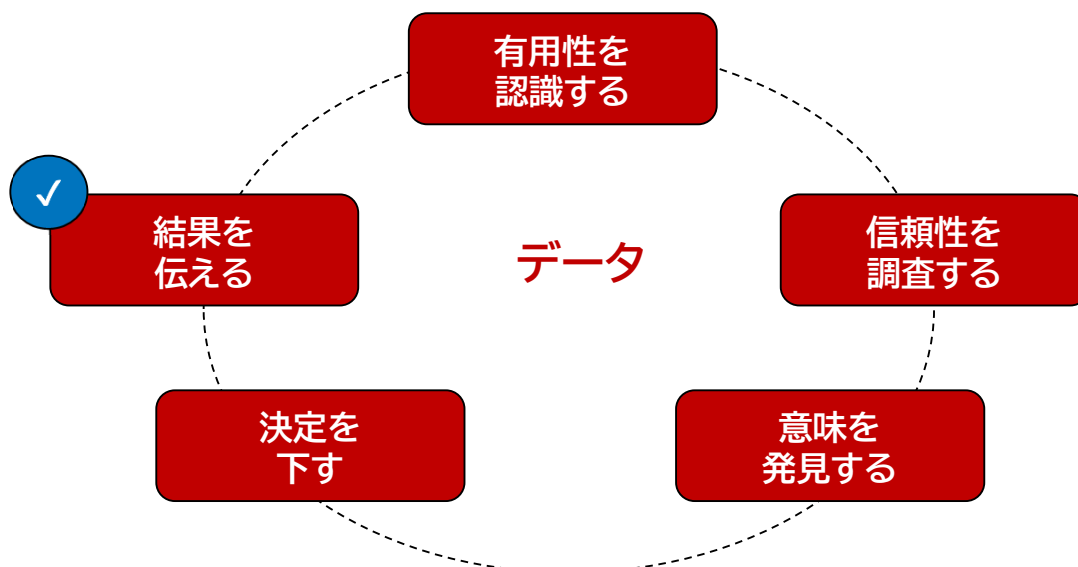
鈴木さんは、これらの変更によって売上と利益がどのように改善されるかを楽しみにしています。
また、収集している新しいデータからどのような洞察が得られるかにも興味を持っています。

彼女は、年間を通じてデータを使用して、行った変更が会社の業績にどのような影響を与えるかを確認する予定です。

鈴木さんは、来年の従業員の仕事に影響を与える重要な決定を下しました。

彼女がその決定を会社と共有する時が来ました。

データリテラシーのスキル



22

Copyright © SAS Institute Inc. All rights reserved.



洞察と決定を効果的に伝える

- 1 目的を明確にする
- 2 説得力のある証拠を提供する
- 3 説明を加える

23

Copyright © SAS Institute Inc. All rights reserved.



①目的を明確にする

意思決定を伝えるときは、まず目的を明確にすることが重要だ。

これによって、結果や決定をどのように示すか、また決定を裏付けるために分析からどのような証拠を盛り込むかが決まる。

24

Copyright © SAS Institute Inc. All rights reserved.



②説得力のある証拠を提供する

直感や勘で行う意思決定とは異なり、データに基づいた意思決定は証拠に基づいて行われる。

データの結果や決定を共有する際には、なぜそのような決定を下したのかを聴衆が理解できるような証拠を含めることが重要。

データを可視化することは、情報を伝える素晴らしい方法。多くの数字を提供する代わりに、あなたが伝えようとしていることをグラフィカルに表現してくれるからだ。

②説得力のある証拠を提供する(つづき)

意思決定や結果を伝えるためにグラフを使う際のヒント：

ビジュアルデザインを考慮し、慎重に可視化を選択する。

- グラフは読みやすく、明確にラベル付けする。グラフのタイトルと軸のラベルを付け、
- 必要であれば凡例やキーも付けよう。データに詳しくない人がグラフを理解しやすくするために含めることは重要。
- 比較するために2つ以上のグラフを入れる場合は、同じ目盛りを使っていることを確認する。
- 聴衆があなたのメッセージをより理解しやすくなるように、グラフに追加情報を含めると、あなたの目的が強調され、明確に伝わりやすくなる。

③説明を加える

別の人が同じグラフを読んだ場合、異なる視点や側面に注目し、データの解釈が異なる可能性がある。そのため、グラフの説明を加えることで、聴衆はあなたが何を伝えようとしているのかを理解しやすくなる。

解説は、グラフから聴衆に何を読み取ってもらいたいかを物語るものであるべきで、グラフに例や具体的な数字を盛り込むことで、あなたの言いたいことが十分に伝わるようにすることができる。

スライドショーで発表する場合は、口頭で説明することが多い。レポートを書く場合は、説明を文章で書くのが一般的。



鈴木さんの意思を伝える

鈴木さんの目的

鈴木さんは、全社的な2022年キックオフの際に、自分の考えを発表する。

鈴木さんの目標は、従業員の仕事に影響を与える重要な決定事項を共有し、従業員の賛同を得ることである。

社員の賛同がなければ、長期的な変化を起こすことは難しい。賛同を得るために、鈴木さんはデータ分析から得た証拠を使い、なぜこれらの行動をそれぞれ選択するのかを説明する。

29

Copyright © SAS Institute Inc. All rights reserved.



鈴木さんの施策アイデア

- 1 年に一度の販促を開催する！
- 2 営業時間を延長する！
- 3 ダイナミック・スタッフを配置する！
- 4 質の高いオンラインプレゼンスを確立する！
- 5 データ主導の文化を発展させる！

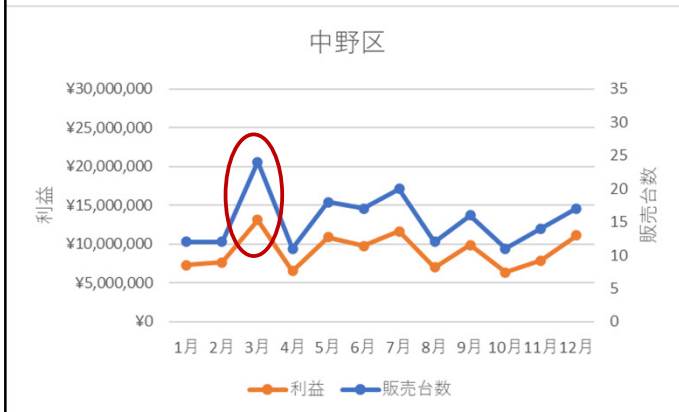
30

Copyright © SAS Institute Inc. All rights reserved.



①年に一度の販促を開催する！

月ごとの利益と販売台数



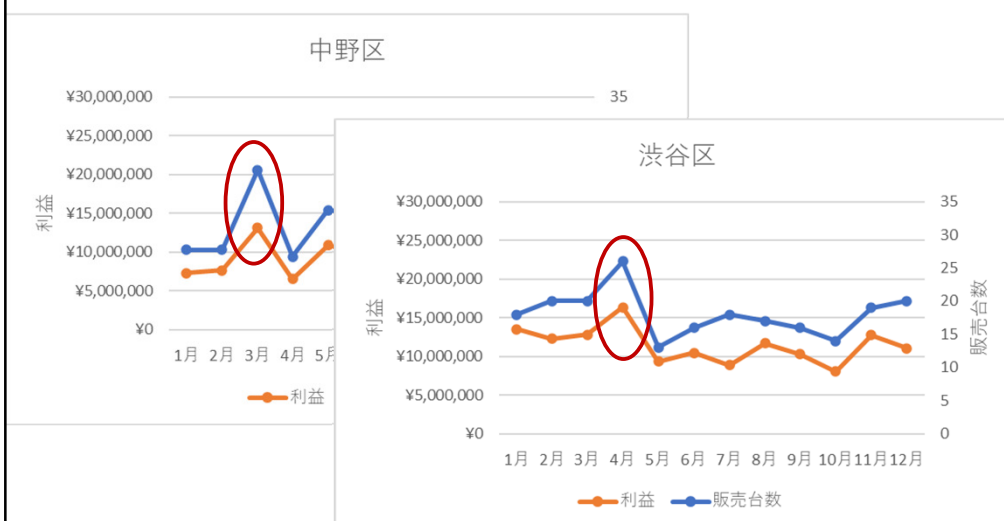
31

Copyright © SAS Institute Inc. All rights reserved.



①年に一度の販促を開催する！

月ごとの利益と販売台数



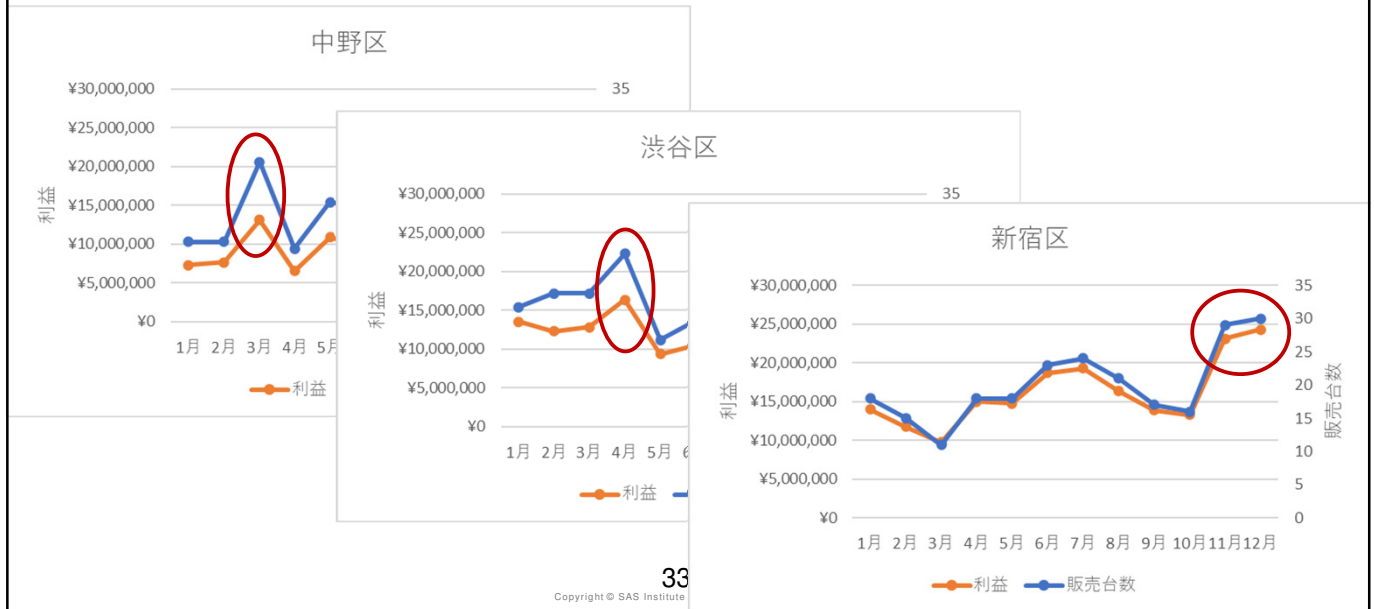
32

Copyright © SAS Institute Inc. All rights reserved.



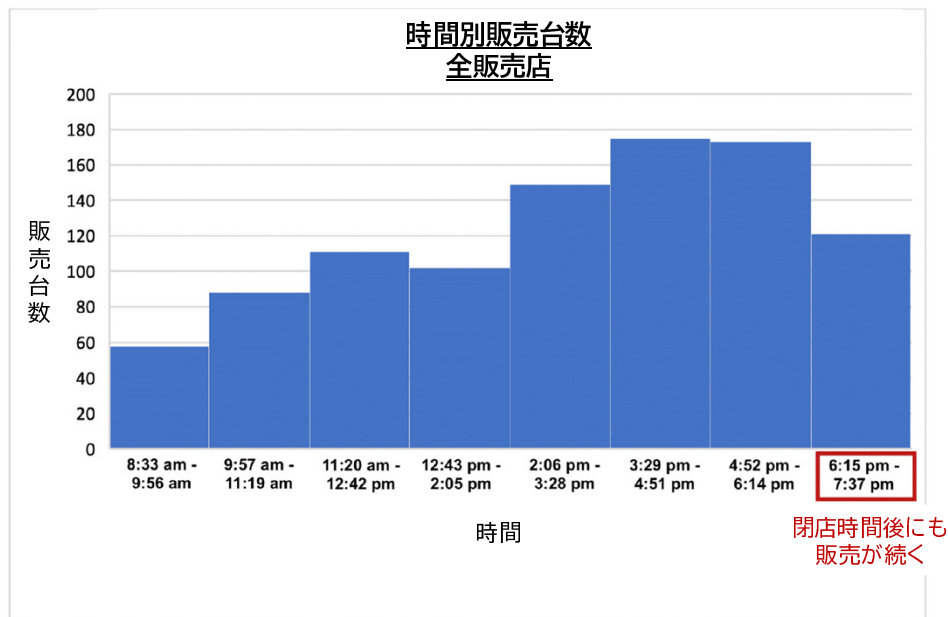
①年に一度の販促を開催する！

月ごとの利益と販売台数



②営業時間を延長する！

時間別販売台数
全販売店



34

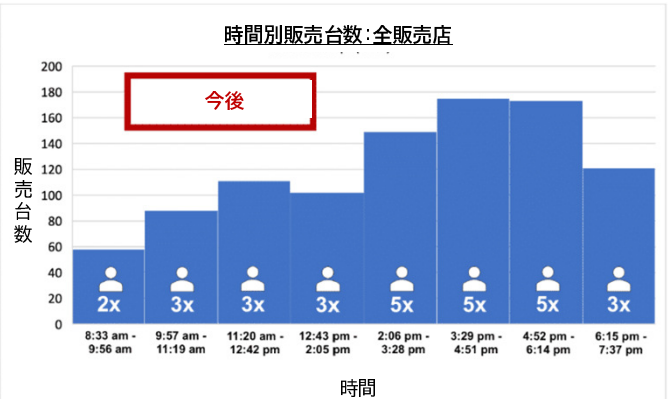
Copyright © SAS Institute Inc. All rights reserved.



③ダイナミック・スタッフを配置する！



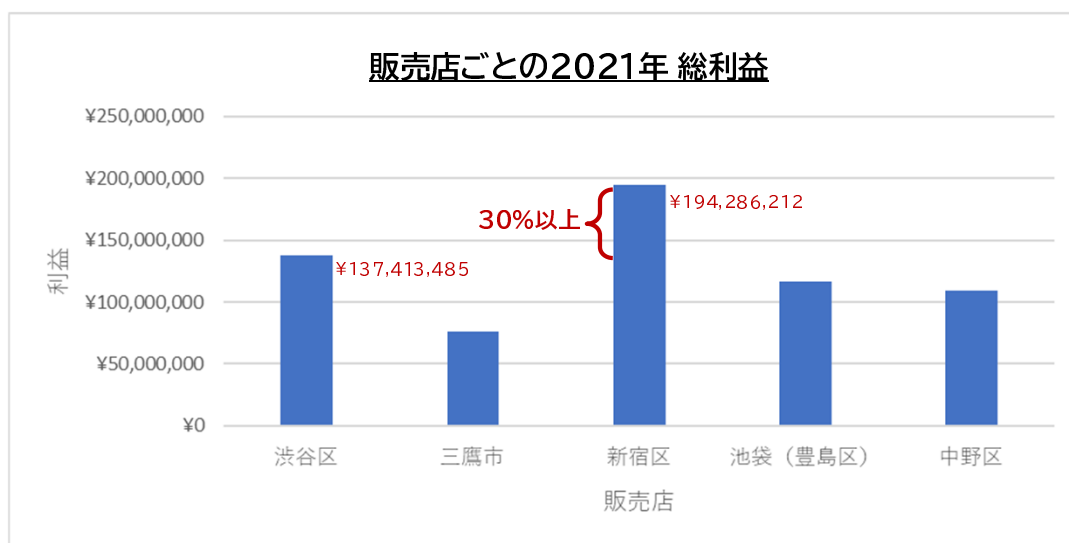
ダイナミック・スタッフィング



35

Copyright © SAS Institute

④質の高いオンラインプレゼンスを確立する！



36

Copyright © SAS Institute Inc. All rights reserved.



⑤データ主導の文化を発展させる！

- 最後に、鈴木さんは、渚の杜自動車のために十分な情報に基づいた意思決定を行うために、自社のビジネスを理解する上でデータ分析がいかに強力であるかを理解した。彼女は、全従業員が日々の意思決定を行うためにデータを収集し、使用することにもっと慣れてほしいと考えている。
- 各販売店は、販売に関する確実なデータを収集し、それを日々の意思決定や長期的な意思決定に活用する。データには、販売を行ったスタッフ、販売店で1日のアクセス数、オンラインのアクセス数などが含まれる。各マネージャーは、日常的・長期的な意思決定により強い洞察をもたらすと思われるデータを収集することを決定する。
- さらに、鈴木さんは1年を通してデータを分析し、行った変更の影響を評価する。また、全従業員が安心してデータを使って洞察し、意思決定できるよう、専門的能力開発も行う。

すべてをまとめる

鈴木さんがこのプロセスを始めたとき、会社全体の利益と経費についてはある程度理解していたが、複数の販売店の業績についてはほとんど知らなかった。彼女は、販売に関連するデータセット(思い起こせば、データが欠落しており、重複していた)を持っており、どうすれば利益を向上させられるかをよりよく理解するのに役立つことを期待していた。

鈴木さんはかなり長い道のりを歩んできた。彼女はデータをきれいにし、分析し、分析から得た洞察を実行可能な、情報に基づいた意思決定に変えてきた。彼女は自分が行った仕事に誇りを持ち、それが会社の成功に強い影響を与えると信じている。彼女は、今度の会社キックオフで、自分が下した決断を社員と共有する準備ができている！

覚えておいてほしいのは、彼女の目的は、決断を下した理由(データを証拠に！)と共に共有し、社員の賛同を得ることだ。結局のところ、社員からの賛同がなければ、インパクトのある変化を起こすことは不可能だということを、彼女は知っているのだ。

最終プレゼンテーション



会社のキックオフ： 2022年の重点分野

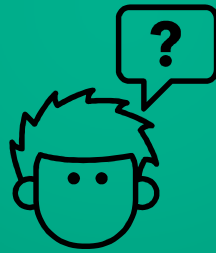
鈴木 佳乃
創業者、渚の杜自動車

40

Copyright © SAS Institute Inc. All rights reserved.



質疑応答



Copyright © SAS Institute Inc. All rights reserved.

会社のキックオフ： 今期の重点分野

鈴木 佳乃

創業者、渚の杜自動車

今期の重点分野



年間販売目標
達成



営業時間の
延長



柔軟な
人員配置



オンラインでの
存在感の強化



データ文化の
醸成

今期の重点分野



年間販売目標
達成



営業時間の
延長



柔軟な
人員配置



オンラインでの
存在感の強化



データ文化の
醸成



今期の重点分野



年間販売目標
達成



営業時間の
延長



柔軟な
人員配置

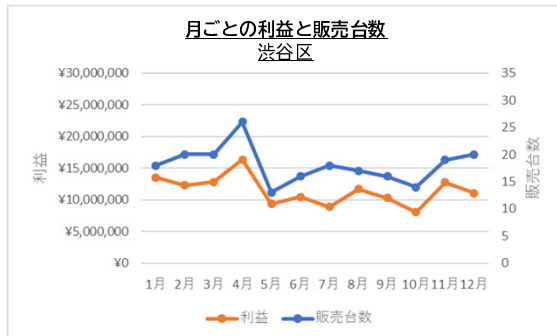


オンラインでの
存在感の強化

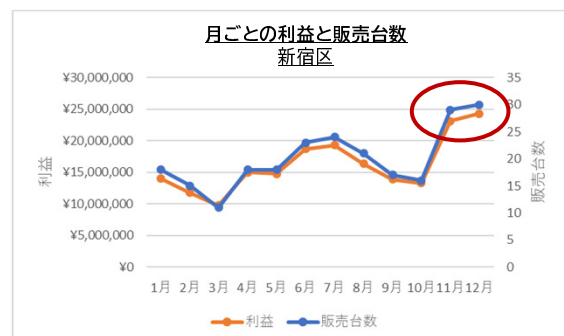
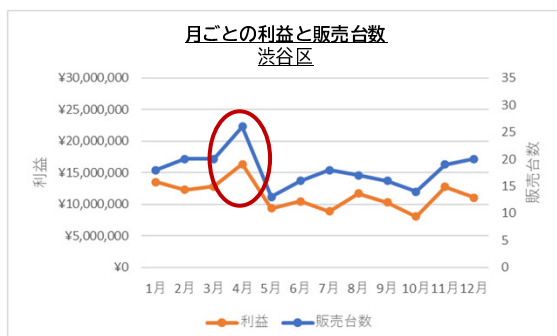


データ文化の
醸成

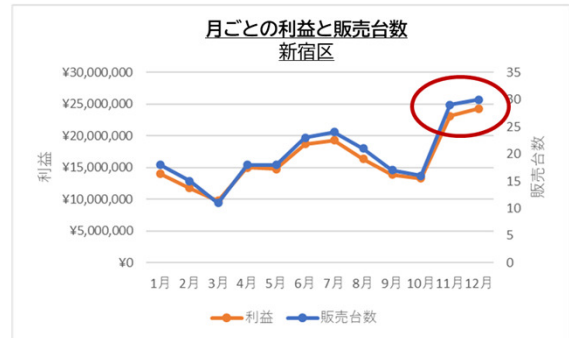
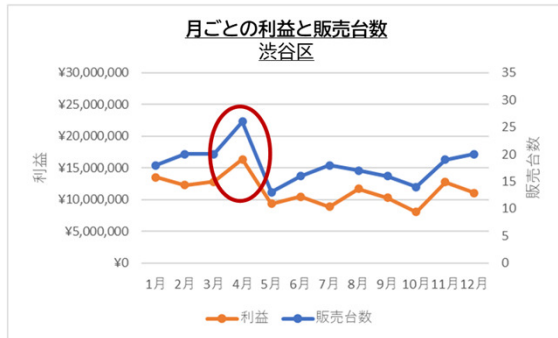
各販売店で年に一度のセールを開催する



各販売店で年に一度のセールを開催する

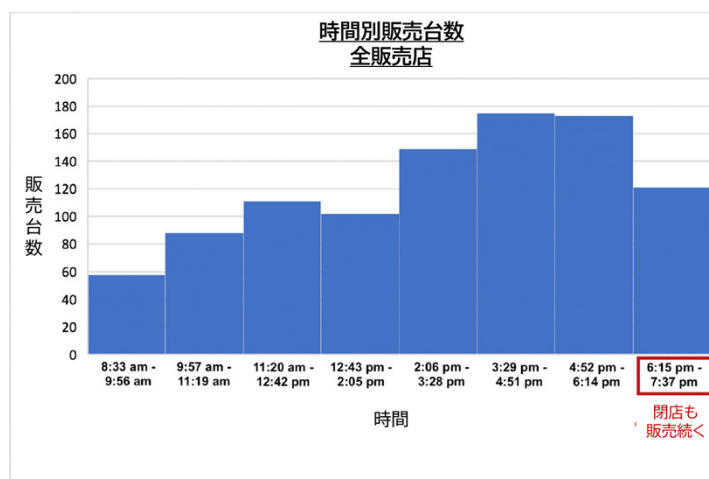


各販売店で年に一度のセールを開催する

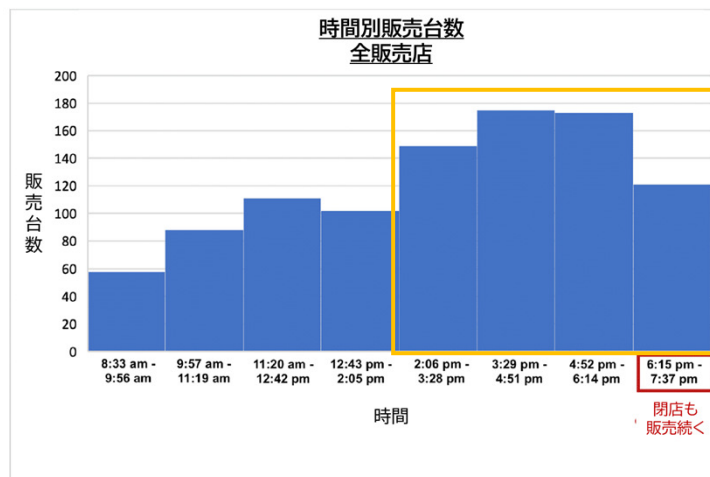


販売店ごとの販売促進 = 利益の増加

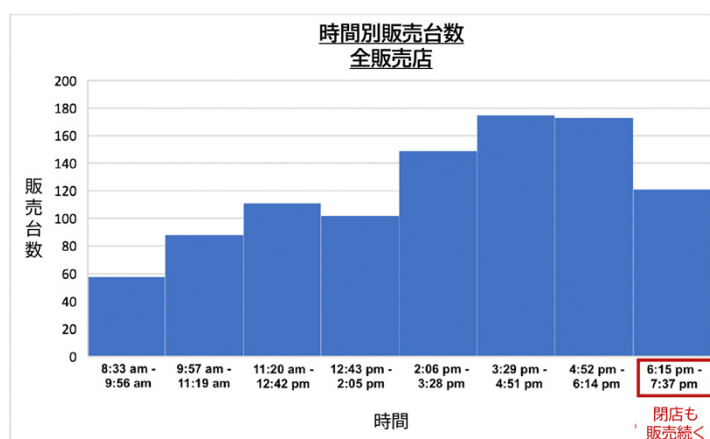
営業時間を延長する！



営業時間を延長する！



営業時間を延長する！

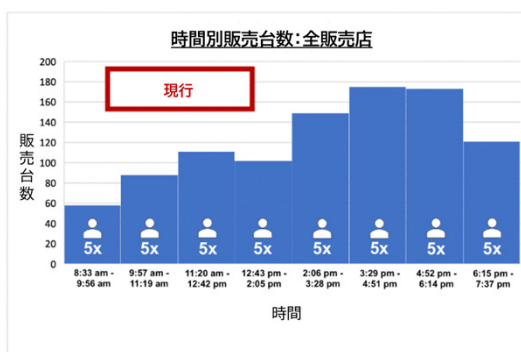


必要に応じて調整するために、毎月データを確認する。

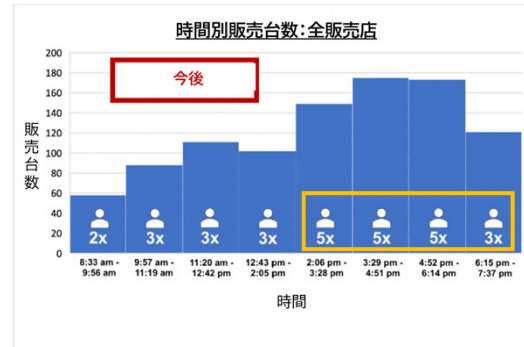
柔軟な人員配置を導入する



柔軟な人員配置を導入する



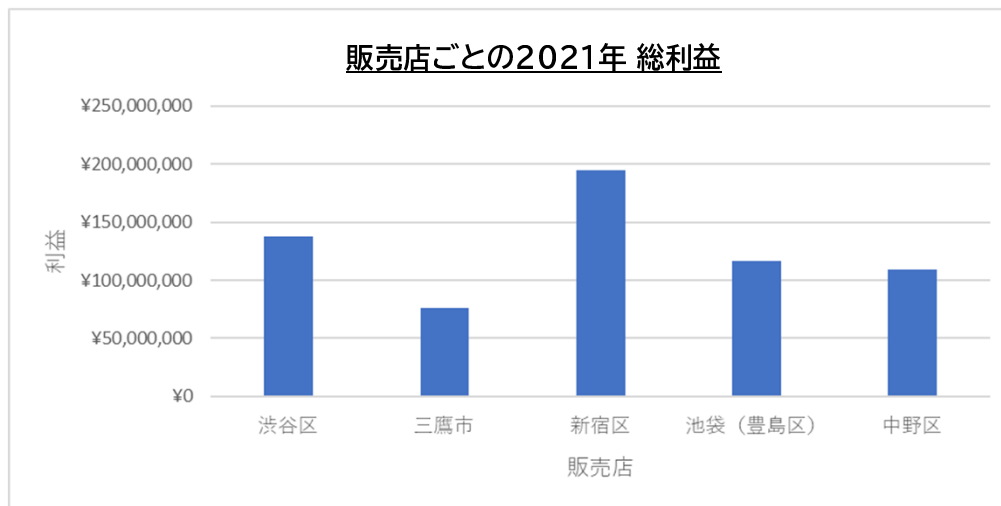
柔軟な人員配置を導入する



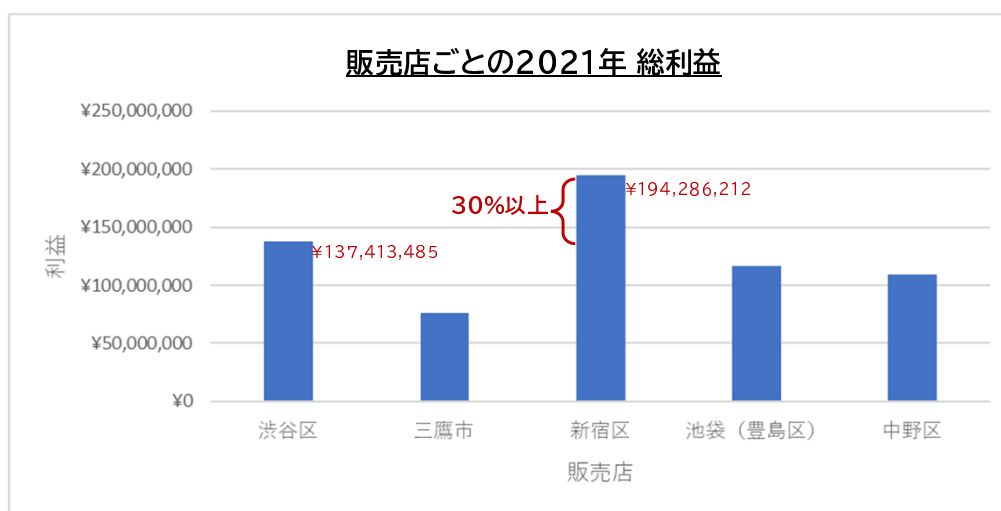
柔軟な人員配置を導入する



質の高いオンラインプレゼンスを確立する



質の高いオンラインプレゼンスを確立する



質の高いオンラインプレゼンスを確立する

重要な要素:

質の高いオンラインプレゼンスを確立する

重要な要素:

- ➡ オンライン在庫
- ➡ 高解像度の画像
- ➡ 透明性のある価格表示
- ➡ お問い合わせフォーム

データ文化を醸成する

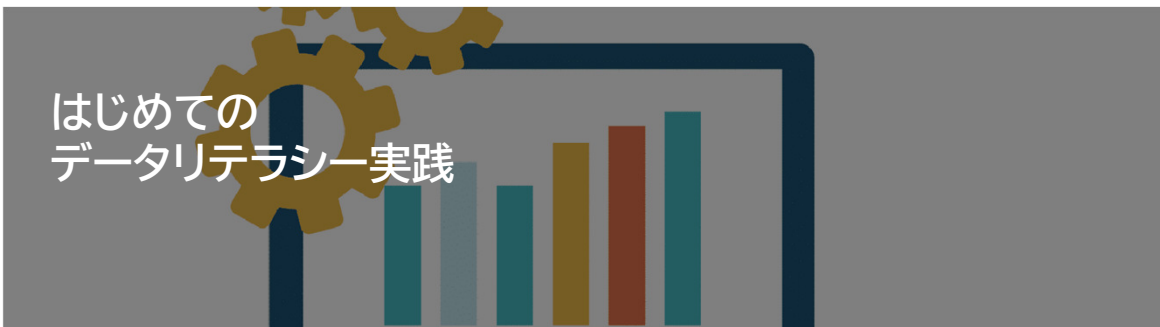


ご清聴
ありがとうございました！

最後に

- このコースを通して、皆さんはデータから重要な洞察を引き出す方法を学んできました。このモジュールでは、これらの洞察を意思決定や他者への伝達によって行動に移す方法を学びます！
- ビジネスでは、データ分析のゴールは一般的にデータの洞察を行動に移すことです。人はしばしば「直感」に基づいて意思決定を行います。これが効果的な場合もあるが、そうでない場合も多い。何が起きているのかを理解するためにデータに目を向けるとき、人々はより良い情報に基づいた意思決定をする力を与えられる。意思決定にデータを活用することは第一歩だが、それで終わりではない！明確で説得力のある根拠をもって意思決定を伝えることは、組織全体の賛同を得、変化を実現する鍵となります。



A decorative graphic featuring a grey background with a dark blue L-shaped frame. Inside the frame, there are several vertical bars of different heights and colors (teal, grey, brown, red, teal). To the left of the bars, there are two interlocking gears, one gold and one brown.

はじめての データリテラシー実践

以上で、コースレッスンは終了です。

理解テストに進んでください！

はじめてのデータリテラシー実践

2026年6月16日 初版第1刷

発行元	SAS Institute Japan株式会社
〒106-6111	東京都港区六本木6-10-1 六本木ヒルズ森タワー11F
	TEL 03(6434)3690

本書内容の一部、全体を問わず、SAS Institute Japan株式会社の文書による承諾なく引用複製することを禁じます。