Empowered by Innovation NEC

超入門　ビジネスデータ分析＆統計解析

NECラーニング　テクノロジー研修事業部
横垣　裕史

アジェンダ

- データ分析概要
- 代表値
- 多次元分析
- 相関分析
- 回帰分析
- 推定
- 検定
- 分散分析

データ分析概要

データ分析により、収集したデータから価値ある知見を導出する

基本的なデータ分析

分析手法	概要
代表値	平均値、最頻値、中間値、分散、標準偏差などの代表値を算出
多次元分析	様々な切り口からデータを分析し、データの特徴を把握
相関分析	データ項目間に関連があるかを調べる
回帰分析	予測値を算出するための予測式を作成する
推定	平均、分散の値を推測する
検定	仮説した平均や分散の値が正しいと言えるかを判定する
分散分析	データ集合の間に違いがあるかを判定する

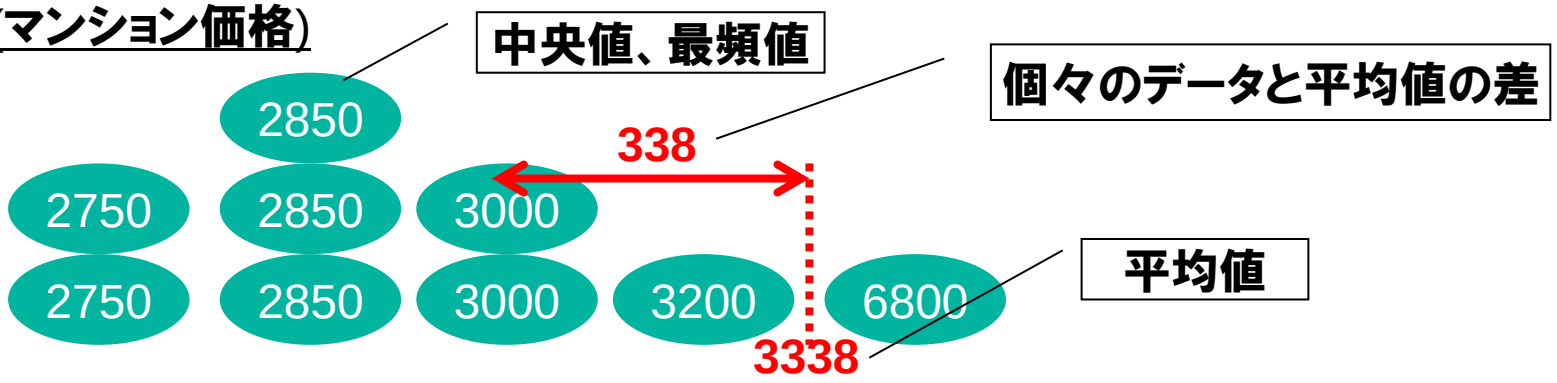
代表値①

代表値とは、データ集合の特徴を表す代表的な値

基本的な代表値

代表値	概要	サンプル値
平均値	データの合計をデータの個数で割った値	3,338
最頻値	データの中で最も頻繁に出現する値	2,850
中間値	データを順番に並べ替えたときの真ん中の値	2,850
分散	個々のデータと平均値の差を2乗した値の平均	1,704,861
標準偏差	分散を正の平方根を取った値	1,305

サンプルデータ(マンション価格)

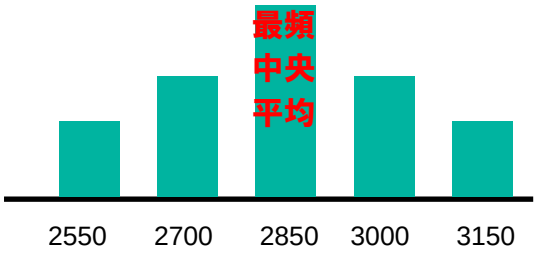


代表値②

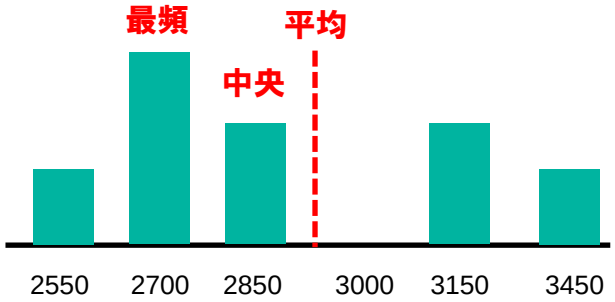
平均値、最頻値、中間値による分析

月	平均値	中央値	最頻値	標準偏差
4月	2850	2850	2850	184
5月	2900	2850	2700	290
6月	2800	2850	3150	327

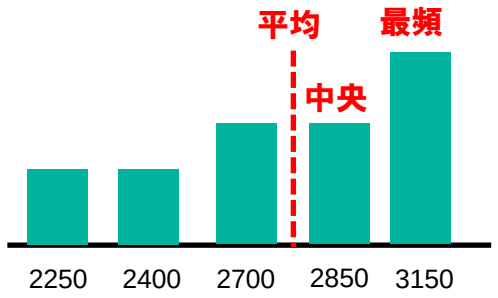
4月



5月



6月



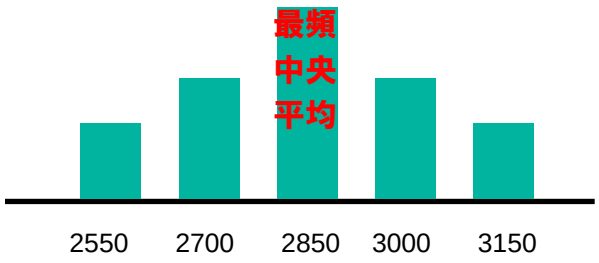
◆平均値 = 中央値 = 最頻値の場合は、平均値を中心とした左右対称の分布
◆平均値 > 中央値 > 最頻値の場合は、最頻値が左側にずれた左右非対称の分布
◆平均値 < 中央値 < 最頻値の場合は、最頻値が右側にずれた左右非対称の分布

代表値③

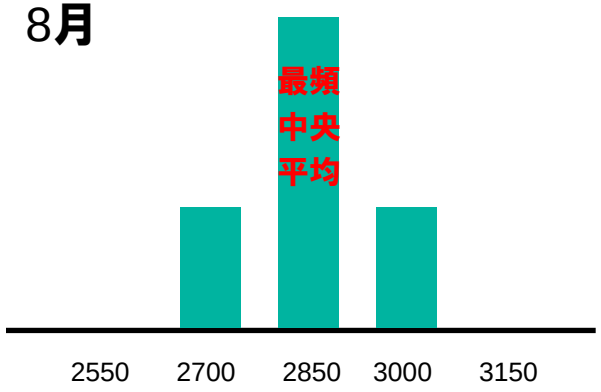
標準偏差による分析

月	平均値	中央値	最頻値	標準偏差
4月	2850	2850	2850	183
7月	2850	2850	2850	75

7月



8月



◆標準偏差が大きい場合、値のばらつきが大きく、平均値の値が出現しにくい
◆標準偏差が小さい場合　値のばらつきが小さく、平均値の値が出現しやすい

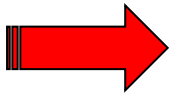
多次元分析

多次元分析とは、様々な視点からデータを観察し、特徴を調べる分析

すべて

	4月	5月	6月	7月
大阪	16	12	29	27
東京	26	28	26	27
名古屋	20	35	39	19

スライス



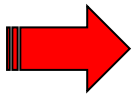
条件の変更

プール有り

	4月	5月	6月	7月
大阪	2	4	7	9
東京	7	9	2	11
名古屋	5	15	11	6

	4月	5月	6月	7月
大阪	16	12	29	27
東京	26	28	26	27
名古屋	20	35	39	19

ダイス



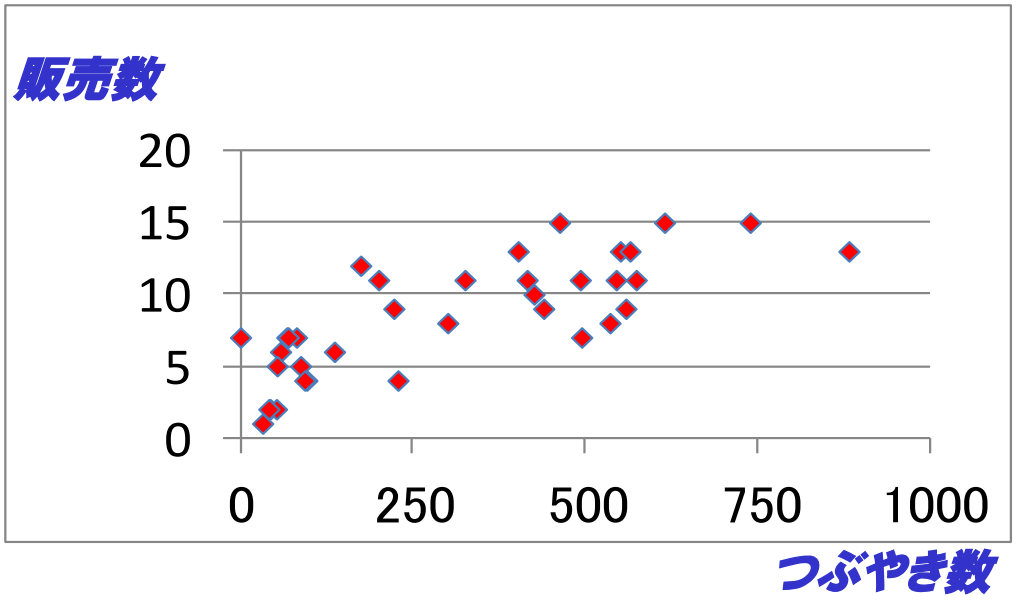
集計項目の変更

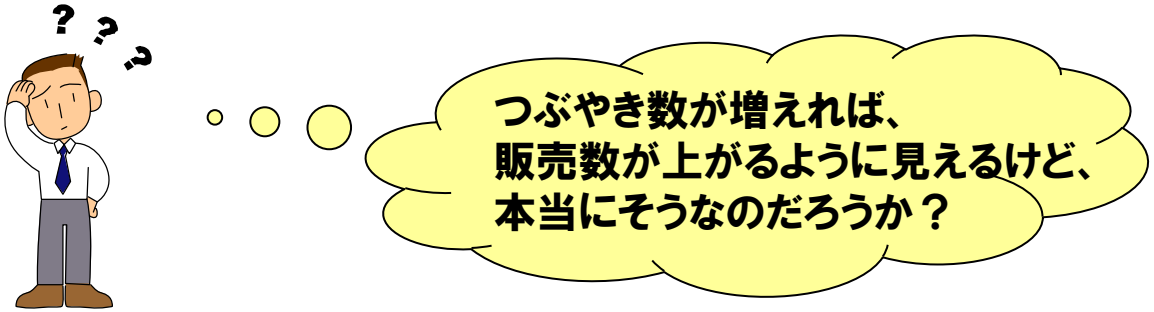
	大阪	東京	名古屋
プール有り	62	78	76
プール無し	22	29	37

相関分析

相関分析とは、2つの項目間に関係があるかどうかを調べる分析

つぶやき数	販売数
228	4
87	5
492	11
58	6
43	2
300	8
96	4
494	7
462	15
・・・	・・・

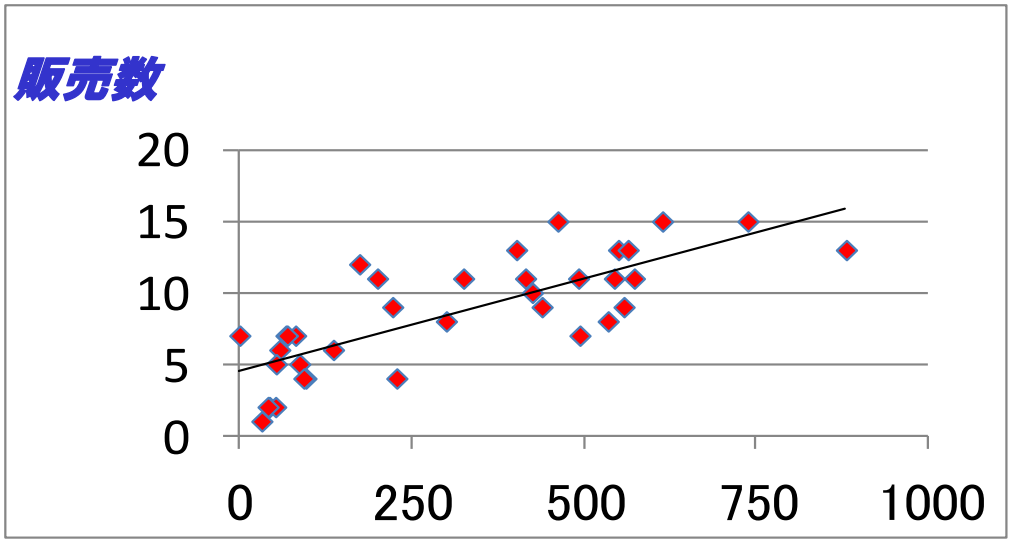




回帰分析

回帰分析とは、実績値をもとに未知の値を予測する分析

つぶやき数	販売数
228	4
87	5
492	11
58	6
43	2
300	8
96	4
494	7
462	15
・・・	・・・



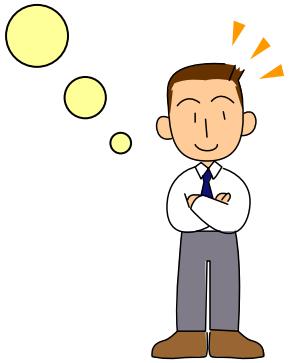
今月は、つぶやき数が
1000ぐらいになりそうだが、
販売数はどれくらいに
なるのだろうか？

推定

推定とは、抽出したデータから、全体の平均値、分散を推測すること

物件名	月	販売数
マンションA	4月	4
マンションB	4月	5
マンションC	4月	11
マンションD	4月	6
マンションE	4月	2
マンションF	4月	8
マンションG	4月	4
マンションH	4月	7
マンションI	4月	15
	・・・	・・・
	平均	8.44

9種類のマンションの月ごとの
売上から、マンション全体の
売上数の平均を推測したい！



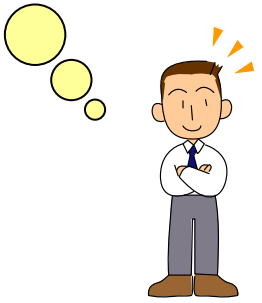
検定

検定とは、仮説した平均や分散の値が正しいと言えるかを判定すること

物件名	月	販売数
マンションA	7月	2
マンションB	7月	6
マンションC	7月	11
マンションD	7月	5
マンションE	7月	9
マンションF	7月	13
マンションG	7月	9
マンションH	7月	11
マンションI	7月	7
	平均	8.11

前月までの平均
8.56

7月の販売数の平均が、前月までの
平均より下がってしまったが、
販売数が下がってしまった
ということができるのだろうか？



分散分析

分散分析とは、データ集合の代表値の違いが、誤差によるものか、必然であるかを調べる分析

No	店有り	店無し
1	11	4
2	8	5
3	15	6
4	7	2
5	7	4
6	11	7
7	15	13
8	10	15
9	13	1
・・・	・・・	・・・
平均	*10.67*	*7.33*