

# 目次

---

1. はじめに
2. 平均
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）

# 本章の目的

---

## (1) 数字の見方を学ぶ

- たくさん並んでいる数字を見て、どのように結論づけるか
- この数字は、大きいのか、小さいのか、偏っているのか・・・

## (2) 数字を見るための計算方法、グラフの作り方も解説します

- 平均値、中央値、標準偏差・・・
- ヒストグラム
- パレート図

# ケース：電話営業データ

	A	B	C	D
1				
2	電話営業にかかる時間（秒）			
3	新人A		ベテランB	
4	1回目	120	150	
5	2回目	150	140	
6	3回目	500	160	
7	4回目	100	170	
8	5回目	600	200	
9	6回目	110	160	
10	7回目	140	170	
11	8回目	800	180	
12	9回目	150	150	

## (1) 電話営業にかかる時間

A) 新人Aさんと、ベテランBさん

B) Aさん「電話時間が長くて非効率」

C) 2人の違いを定量的に評価

## (2) 評価指標

A) 平均値

B) グラフ

C) 中央値

D) 標準偏差

# 目次

---

1. はじめに
2. 平均値
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）

# ケース：電話営業データ

	A	B	C	D
1				
2	電話営業にかかる時間（秒）			
3		新人A	ベテランB	
4	1回目	120	150	
5	2回目	150	140	
6	3回目	500	160	
7	4回目	100	170	
8	5回目	600	200	
9	6回目	110	160	
10	7回目	140	170	
11	8回目	800	180	
12	9回目	150	150	

## (1) 電話営業にかかる時間

A) 新人Aさんと、ベテランBさん

B) Aさん「電話時間が長くて非効率」

C) 2人の違いを定量的に評価

## (2) 評価指標

A) 平均値

B) グラフ

C) 中央値

D) 標準偏差

# 平均値

---

## (1) もっとも代表的な指標

- とにかく **分かりやすい**
- 複雑な計算をするほど、ビジネスの現場で議論がスムーズに進まない

## (2) 計算式

- =AVERAGE (計算範囲)

## (3) 計算結果

- 新人Aさんは、ベテランBさんよりも **平均会話時間が長くなっている**

# 目次

---

1. はじめに
2. 平均値
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）

# ケース：電話営業データ

	A	B	C	D
1				
2	電話営業にかかる時間（秒）			
3		新人A	ベテランB	
4	1回目	120	150	
5	2回目	150	140	
6	3回目	500	160	
7	4回目	100	170	
8	5回目	600	200	
9	6回目	110	160	
10	7回目	140	170	
11	8回目	800	180	
12	9回目	150	150	

## (1) 電話営業にかかる時間

A) 新人Aさんと、ベテランBさん

B) Aさん「電話時間が長くて非効率」

C) 2人の違いを定量的に評価

## (2) 評価指標

A) 平均値

B) グラフ

C) 中央値

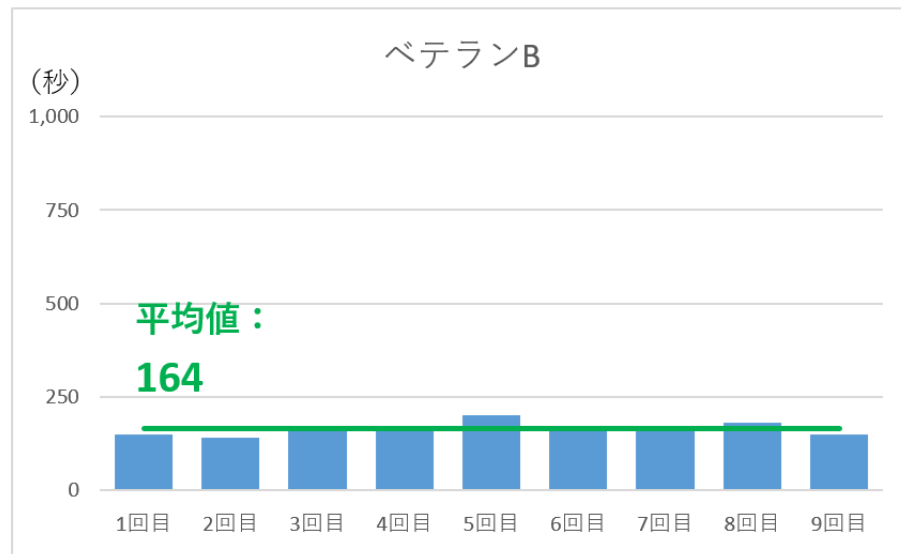
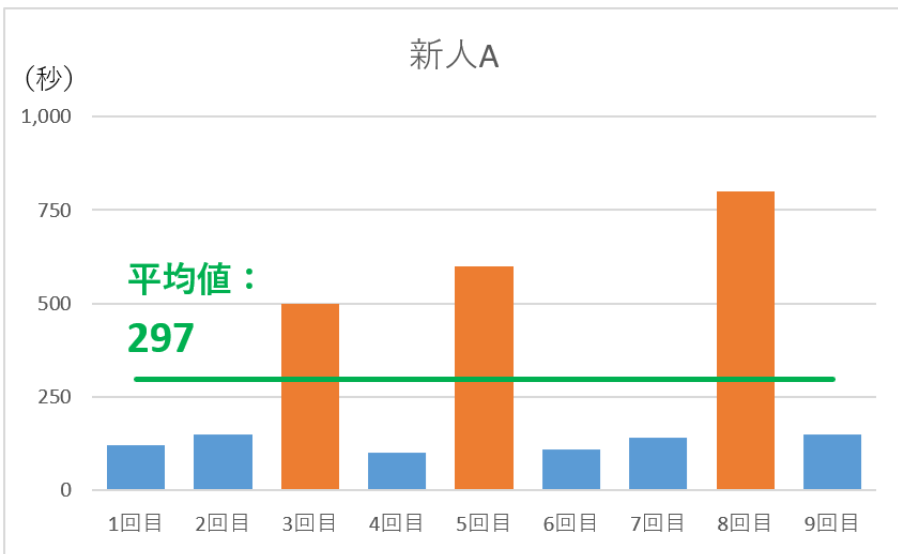
D) 標準偏差



# 平均値の課題

## (1) 平均値

- 極端に大きい数値があると、平均が引っ張られてしまう
- 新人Aさんは、一部の極端なデータを除けば、ベテランBさんと同じくら



# データは、まずグラフで見る

---

## (1) グラフのメリット

- 数字のバラつきが分かりやすい

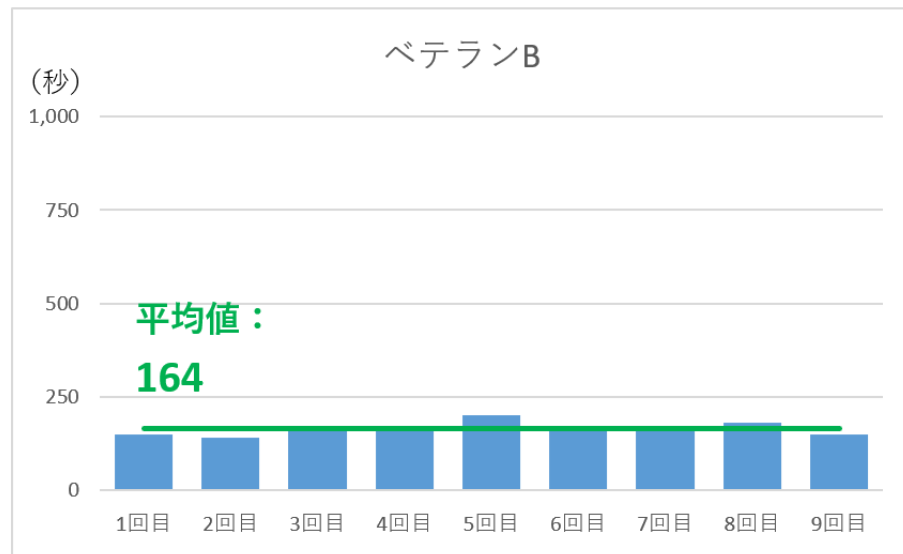
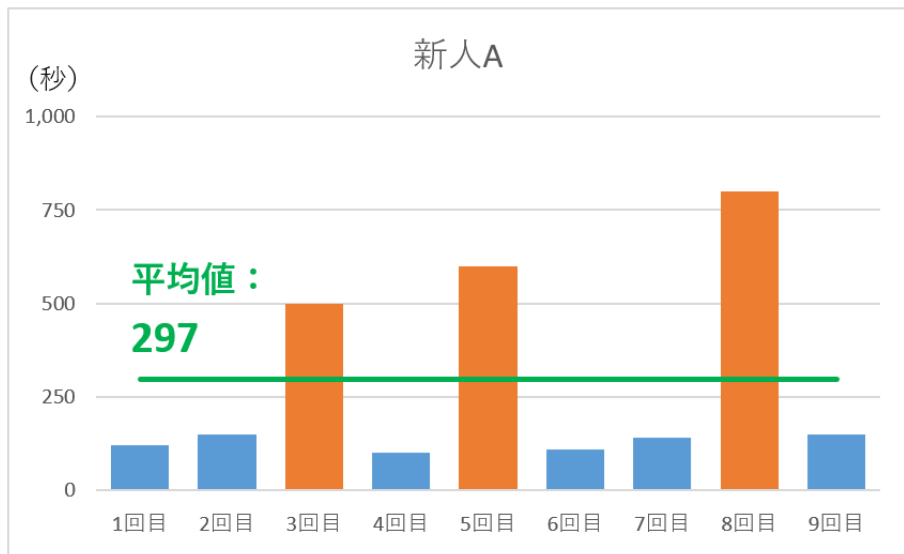
## (2) バラつきの原因を追究し、対応策を考える

- 改善できる数字かもしれない
  - 見込みがない顧客と、つい長い時間話してしまっている、など
- あるいは、ビジネスの世界では、無視していい数字もある
  - たまたまクレームがあって長電話になってしまった
  - 平均値に入れない、という判断も

# グラフ(1) 縦棒グラフ

## (1) 縦棒グラフ

- シンプルで分かりやすい
- 平均値を出すときは、組み合わせ→折れ線グラフで



# 目次

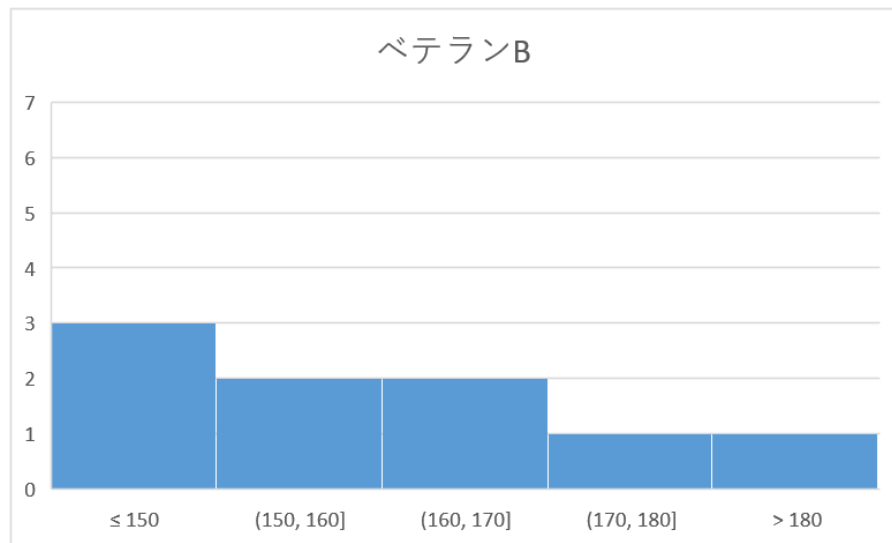
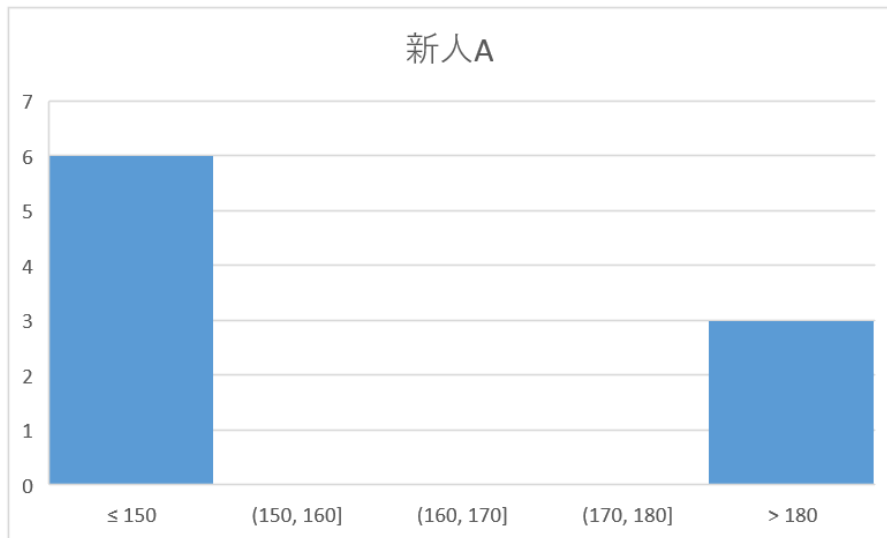
---

1. はじめに
2. 平均値
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）

# グラフ(2) ヒストグラム

## (1) ヒストグラム

- どのあたりに数字が集中しているか（分布）がわかる
- 150～180の間で、頻度を計算
- ベテランBさんに比べて、Aさんはバラつきが大きい（＝極端）



# グラフ(2) ヒストグラム

---

## (1) ヒストグラム

- どのあたりに数字が集中しているか (分布) がわかる

## (2) 作り方

- 挿入 → 統計グラフ → ヒストグラム
- 横軸を右クリック → 軸の書式設定
  - ビンの数
  - オーバーフロー (右端の最大値)
  - アンダーフロー (左端の最小値)

# 目次

---

1. はじめに
2. 平均値
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）

# ケース：電話営業データ

	A	B	C	D
1				
2	電話営業にかかる時間（秒）			
3		新人A	ベテランB	
4	1回目	120	150	
5	2回目	150	140	
6	3回目	500	160	
7	4回目	100	170	
8	5回目	600	200	
9	6回目	110	160	
10	7回目	140	170	
11	8回目	800	180	
12	9回目	150	150	

## (1) 電話営業にかかる時間

A) 新人Aさんと、ベテランBさん

B) Aさん「電話時間が長くて非効率」

C) 2人の違いを定量的に評価

## (2) 評価指標

A) 平均値

B) グラフ

C) 中央値

D) 標準偏差

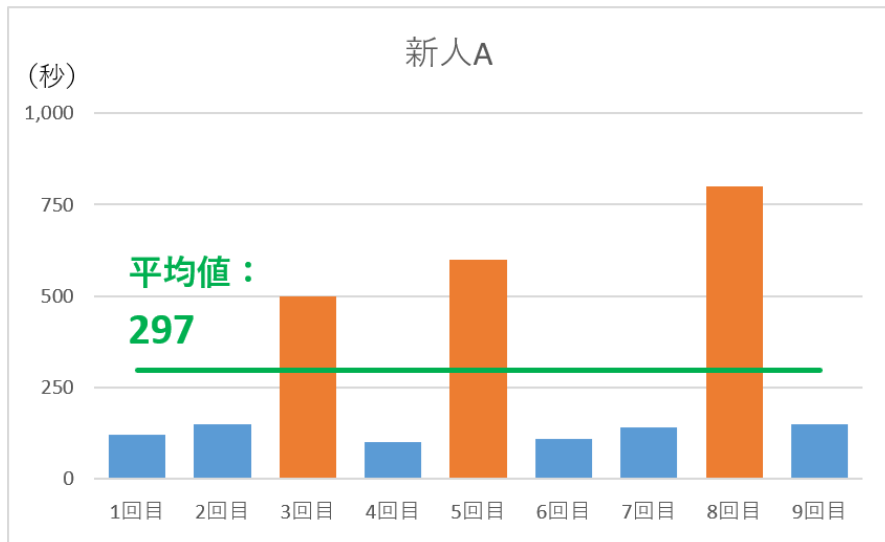


# 平均値の課題

## (1) 平均値

- 極端に大きい数値があると、平均が引っ張られてしまう  
→中央値という数字を使う

平均値：297

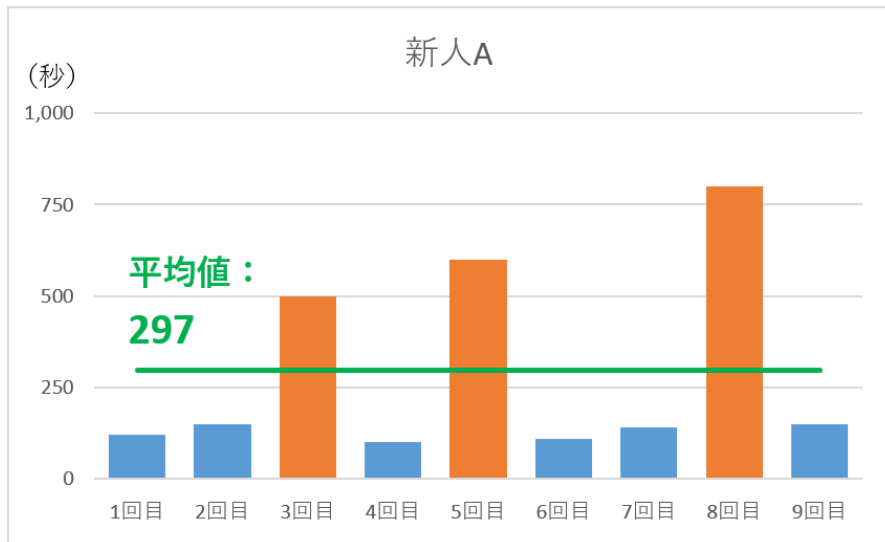


# 中央値

## (1) 中央値とは

- 大きい順に並び替えたときに、真ん中（9回中5番目に）大きい数字
- 極端な数値に引っ張られにくい

平均値：297



中央値：150



# 中央値

---

## (1) グラフの作り方

- **フィルター**で大きい順に並び替えると便利

## (2) 関数

- =MEDIAN (計算範囲)

## (3) 計算結果

- 中央値でみると、新人Aさんと、ベテランBさんでは変わらない  
→ **新人Aさんが時間かかっている理由は、一部の極端の数字のせい**

# 目次

---

1. はじめに
2. 平均値
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）

# ケース：電話営業データ

	A	B	C	D
1				
2	電話営業にかかる時間（秒）			
3		新人A	ベテランB	
4	1回目	120	150	
5	2回目	150	140	
6	3回目	500	160	
7	4回目	100	170	
8	5回目	600	200	
9	6回目	110	160	
10	7回目	140	170	
11	8回目	800	180	
12	9回目	150	150	

## (1) 電話営業にかかる時間

A) 新人Aさんと、ベテランBさん

B) Aさん「電話時間が長くて非効率」

C) 2人の違いを定量的に評価

## (2) 評価指標

A) 平均値

B) グラフ

C) 中央値

D) 標準偏差

# 分散と標準偏差

## (1) 先ほどグラフで分布をチェックしました

- グラフだけでは、数字の分散（バラつき）を定量的に表現しにくい

## (2) 分散（バラつき）を、数字で定量的に表現する

- 標準偏差
- 高いほど、バラつきが大きい（新人Aさん）

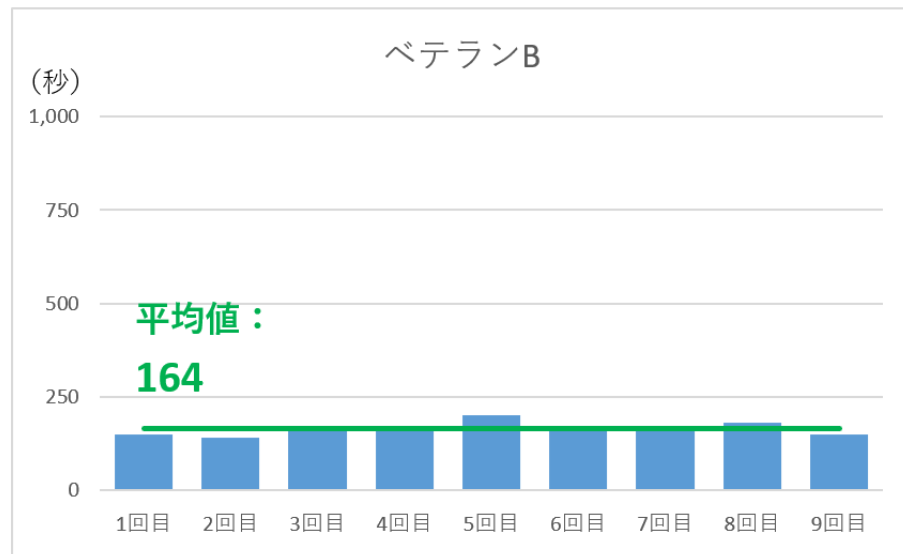
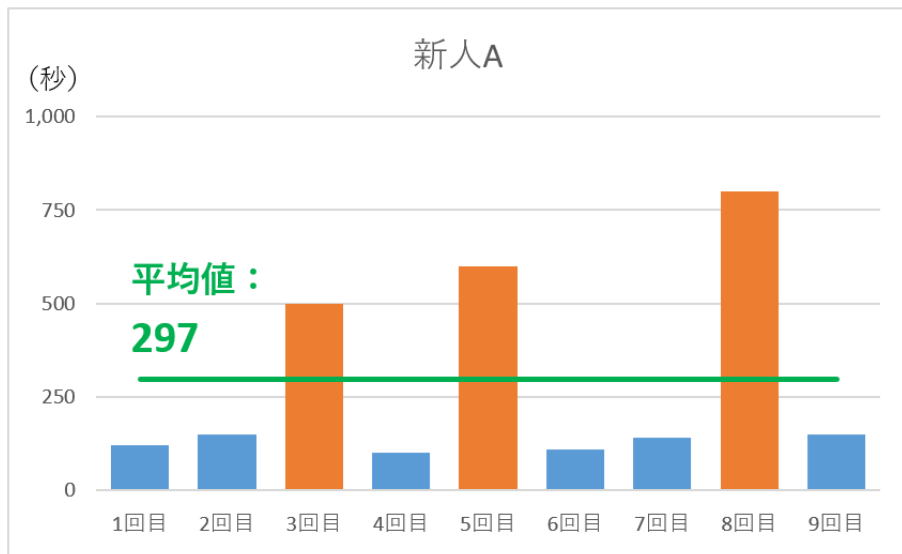
## (3) 関数（標準偏差）

- =STDEV.P（計算範囲）

# グラフ(1) 縦棒グラフ

## (1) 縦棒グラフ

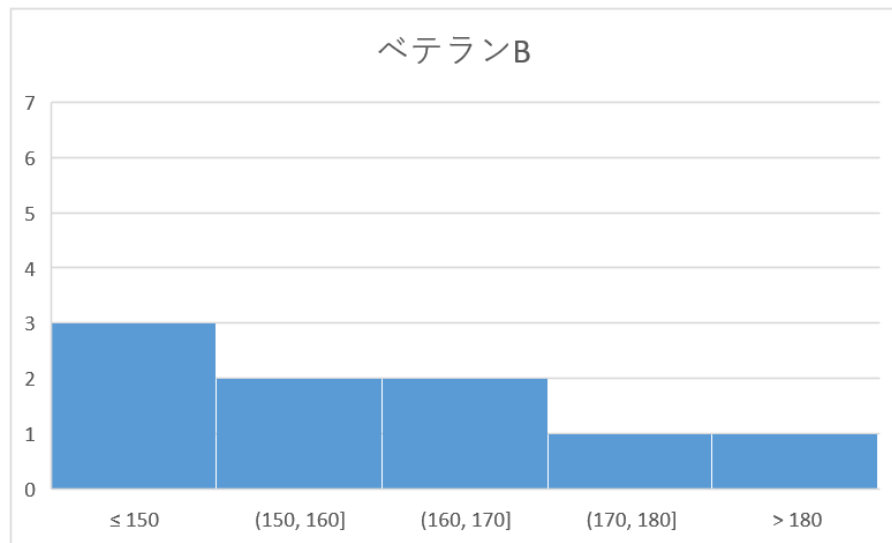
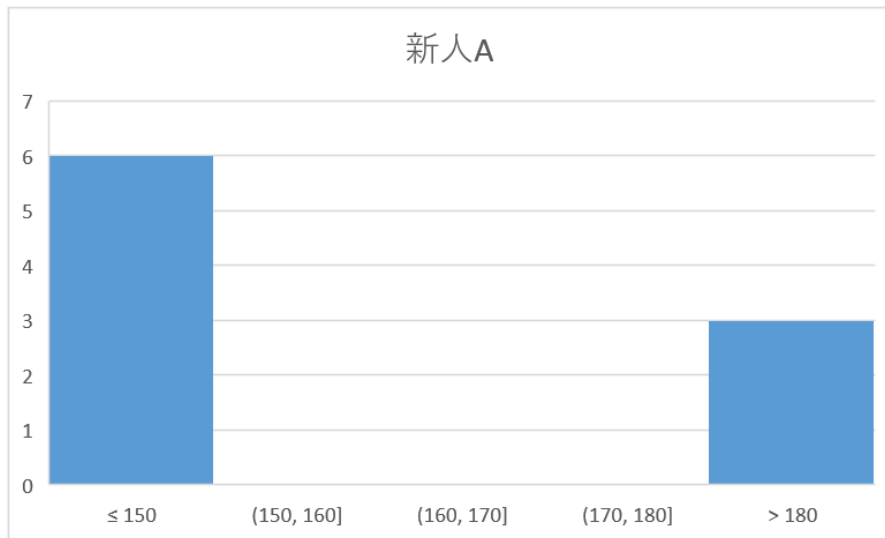
- シンプルで分かりやすい
- 平均値を出すときは、組み合わせ→折れ線グラフで



# グラフ(2) ヒストグラム

## (1) ヒストグラム

- どのあたりに数字が集中しているか（分布）がわかる
- 150～180の間で、頻度を計算
- ベテランBさんに比べて、Aさんはバラつきが大きい（＝極端）





# 標準偏差

## (1) メリット

- 分散（数字のバラつき）を定量的に表現できる
- AさんよりBさんのほうが、バラつきは少ないとはっきり説明できる

## (2) 注意点

- バラつき方が分かりにくい
  - 1つだけ極端な数字があるのかもしれない
  - 解決策のアイデアが出にくい
- あくまで、バラつきを知るきっかけとして使う
  - 標準偏差が高かったら、改めてグラフでチェックする

# 目次

---

1. はじめに
2. 平均値
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）

# ここまでのまとめ

	A	B	C	D
1				
2	電話営業にかかる時間（秒）			
3		新人A	ベテランB	
4	1回目	120	150	
5	2回目	150	140	
6	3回目	500	160	
7	4回目	100	170	
8	5回目	600	200	
9	6回目	110	160	
10	7回目	140	170	
11	8回目	800	180	
12	9回目	150	150	

## (1) 電話営業にかかる時間

A) 新人Aさんと、ベテランBさん

B) Aさん「電話時間が長くて非効率」

C) 2人の違いを定量的に評価

## (2) 評価指標

A) 平均値

B) グラフ

C) 中央値

D) 標準偏差

# ここまでのまとめ（分析例）

	A	B	C	D
1				
2		電話営業にかかる時間（秒）		
3			新人A	ベテランB
4		1回目	120	150
5		2回目	150	140
6		3回目	500	160
7		4回目	100	170
8		5回目	600	200
9		6回目	110	160
10		7回目	140	170
11		8回目	800	180
12		9回目	150	150
13				
14		平均	297	164
15		中央値	150	160
16		標準偏差	249	17

## (1) 平均値

- 新人Aさんのほうが長い

## (2) 中央値

- ふたりとも同じ水準

## (3) 標準偏差

- 新人Aさんのほうが大きい

Aさんは、一部極端に長い会話時間があるため、  
平均値は長いが、それを解決できればBさんと  
同じ水準まで改善できそう

# 目次

---

1. はじめに
2. 平均値
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）

# パレート分析

---

## (1) パレートの法則

- 全体の数値の大部分は、一部の要素が生み出している
- 売上の80%は、20%の優良顧客から生まれている
  - 80 : 20の法則

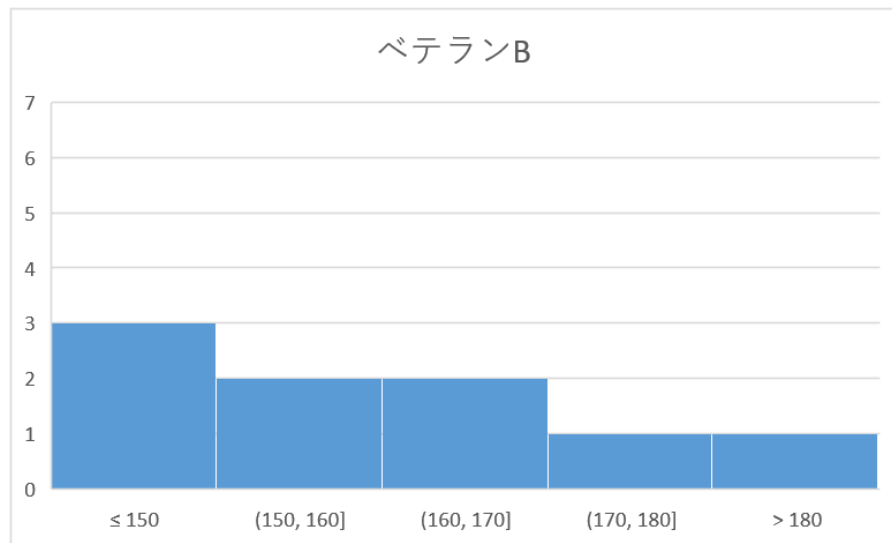
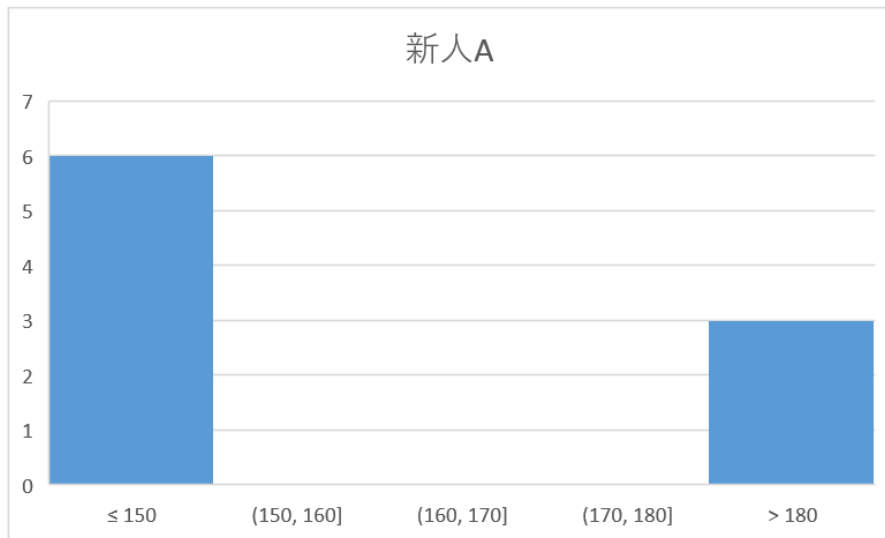
## (2) なぜ重要か？

- 優良顧客が誰か、を明確にする
- マーケティング対象の優先順位を考える
  - 50代～70代の顧客を優先的に営業して、40代以下は気にしない

# グラフ(2) ヒストグラム

## (1) ヒストグラム

- どのあたりに数字が集中しているか (分布) がわかる
- 150~180の間で、頻度を計算
- ベテランBさんに比べて、Aさんはバラつきが大きい (= 極端)



# ある百貨店の顧客数を見ると、50代以上が80%

	A	B	C	D	E
1					
2	世代別顧客数（千人）				
3			顧客数	シェア	累計シェア
4	70代以上	9,000	38%	38%	
5	60代	6,000	25%	63%	
6	50代	4,000	17%	80%	
7	40代	2,500	11%	91%	
8	30代	1,000	4%	95%	
9	20代	700	3%	98%	
10	10代以下	500	2%	100%	
11	合計	23,700	100%		



# 目次

---

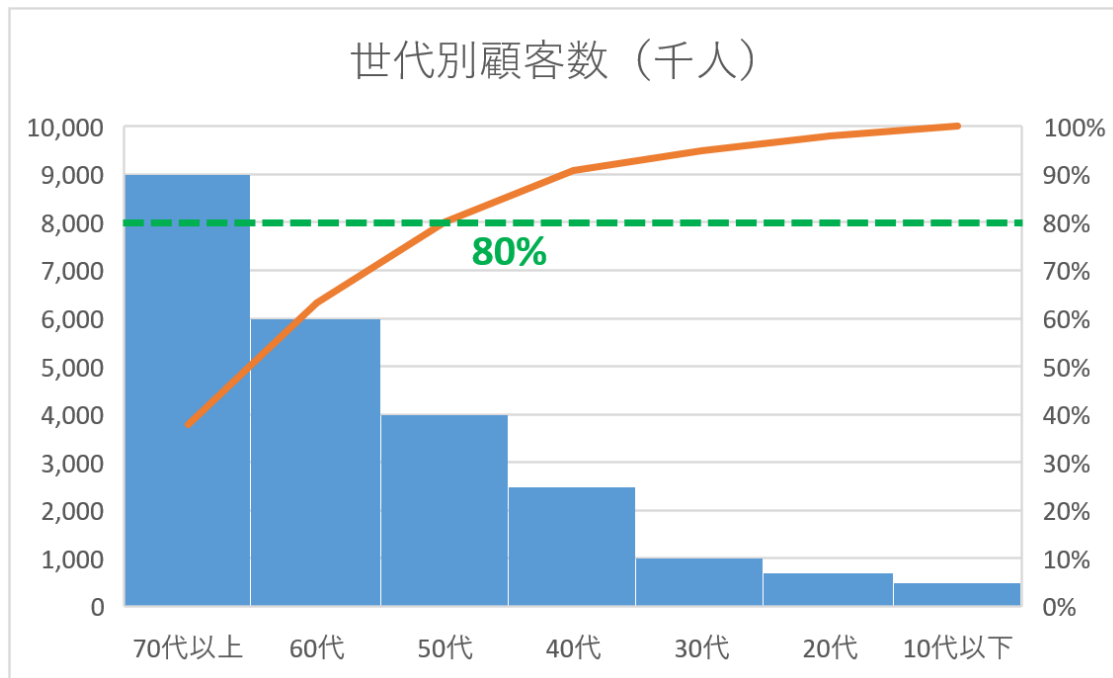
1. はじめに
2. 平均値
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）

# ある百貨店の顧客数を見ると、50代以上が80%

	A	B	C	D	E
1					
2	世代別顧客数（千人）				
3			顧客数	シェア	累計シェア
4	70代以上	9,000	38%	38%	
5	60代	6,000	25%	63%	
6	50代	4,000	17%	80%	
7	40代	2,500	11%	91%	
8	30代	1,000	4%	95%	
9	20代	700	3%	98%	
10	10代以下	500	2%	100%	
11	合計	23,700	100%		

# パレート図

- 50代～70代が80% → 最優先でマーケティングを考える
- 40代まで含めると90% → 次に優先



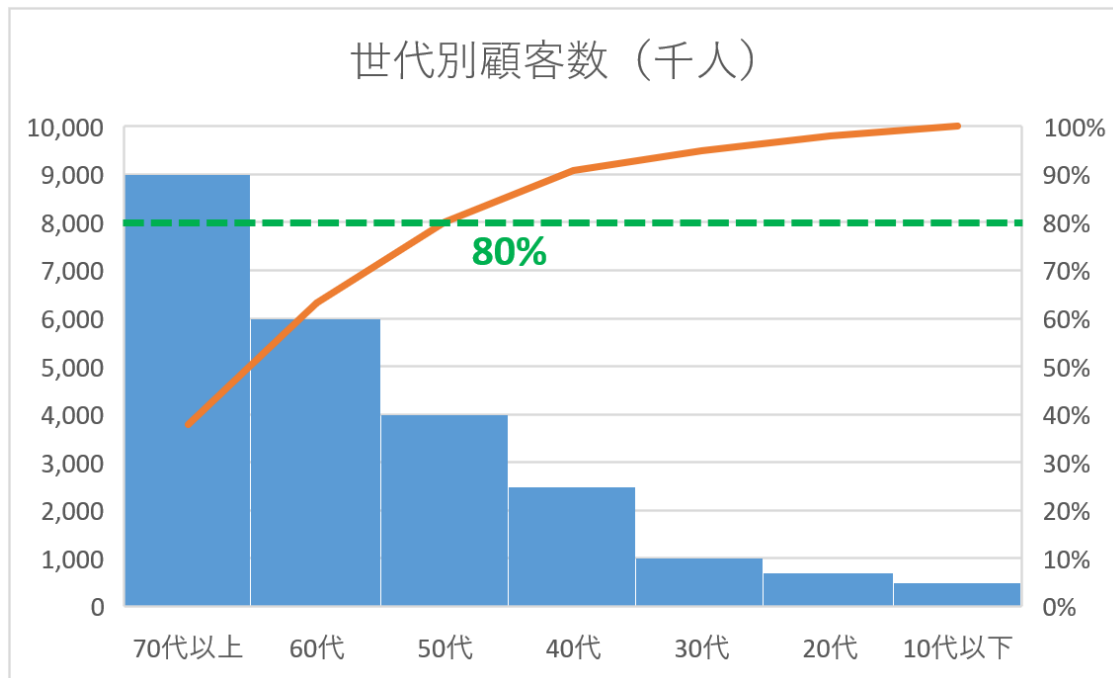
# 目次

---

1. はじめに
2. 平均値
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）

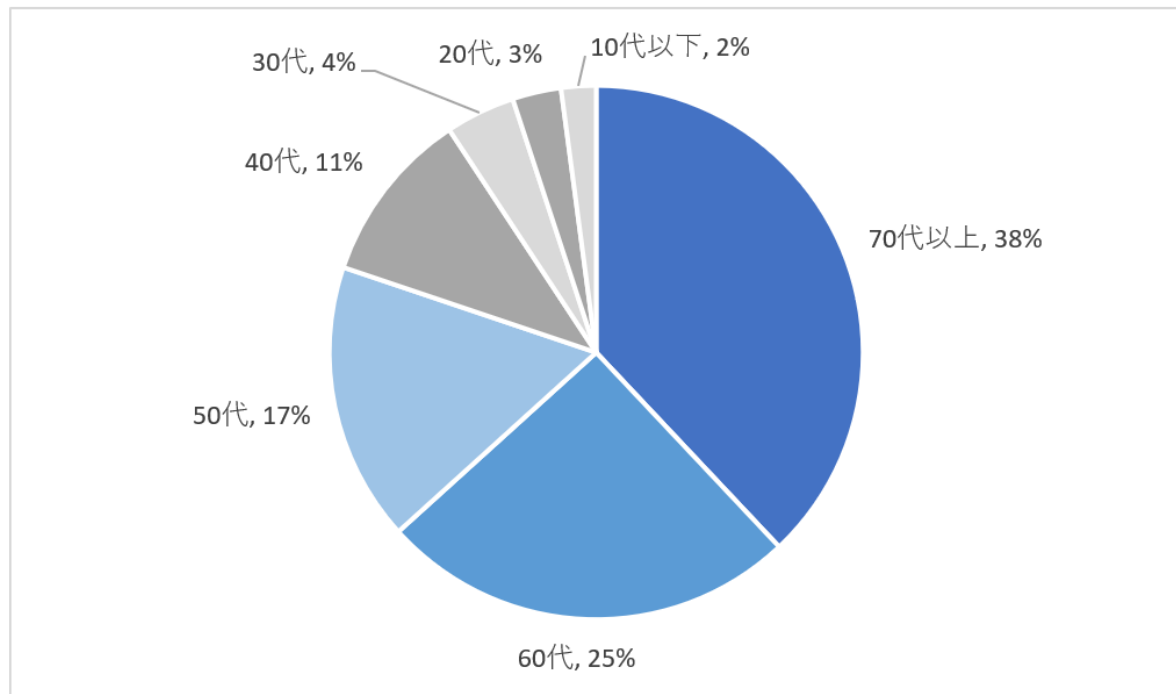
# パレート図

- 50代～70代が80% → 最優先でマーケティングを考える
- 40代まで含めると90% → 次に優先



# 円グラフ

- 50代～70代が80% → 最優先でマーケティングを考える
- 40代まで含めると90% → 次に優先



# 円グラフ作成のコツ

---

## (1) データラベル

- 右クリック → データラベルの追加

## (2) データラベルの書式設定

- 分類名
- 外部に表示

# パレート分析

---

## (1) 使用例

- クレーム内容ごとに、クレーム件数を調べる
  - 実はクレームの80%は、一部の原因によるもの
  - その原因をなくすための対策に時間をかける
- 商品ごとに、販売数を調べる
  - 一部の商品が売上の80%を占める
  - 商品開発、改善を考える



# 目次

---

1. はじめに
2. 平均値
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）

# 平均値

---

## (1) 計算の王道

- いちばん分かりやすい指標

## (2) 平均の計算の応用

- 加重平均
- 条件付き平均
- 平均成長率 (CAGR)

# 平均値

---

## (1) 計算の王道

- いちばん分かりやすい指標

## (2) 平均の計算の応用

- 加重平均
- 条件付き平均
- 平均成長率 (CAGR)

# 平均値

---

## (1) 加重平均

- 単純に平均せずに、人数や販売数などに**重みを付けて**平均を計算する
- 合計の売上を計算してから、販売数で割る

## (2) 計算式

- =SUMPRODUCT(範囲①, 範囲②)
- ※範囲①×範囲②をそれぞれ掛け合わせる

# 目次

---

1. はじめに
2. 平均値
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）

# 平均値

---

## (1) 計算の王道

- いちばん分かりやすい指標

## (2) 平均の計算の応用

- 加重平均
- 条件付き平均
- 平均成長率 (CAGR)

# 平均値

---

## (1) 条件付き平均

- 条件に合致した数字だけの平均を計算する
- 男性だけの平均売上、女性だけの平均売上
- データ分解することで、結論が変わることがある
  - 男性の売上は増加したものの、女性の売上は減少した

## (2) 計算式

- =AVERAGEIF(検索範囲, 検索条件, 平均計算範囲)

# 目次

---

1. はじめに
2. 平均値
3. グラフ化(1) 縦棒グラフ
4. グラフ化(2) ヒストグラム
5. 中央値
6. 標準偏差
7. ここまでのまとめ
8. パレート分析(1)
9. パレート分析(2)：パレート図
10. パレート分析(3)：円グラフ
11. 平均の応用(1)：加重平均
12. 平均の応用(2)：条件付き平均
13. 平均の応用(3)：平均成長率（CAGR）



# 平均値

---

## (1) 計算の王道

- いちばん分かりやすい指標

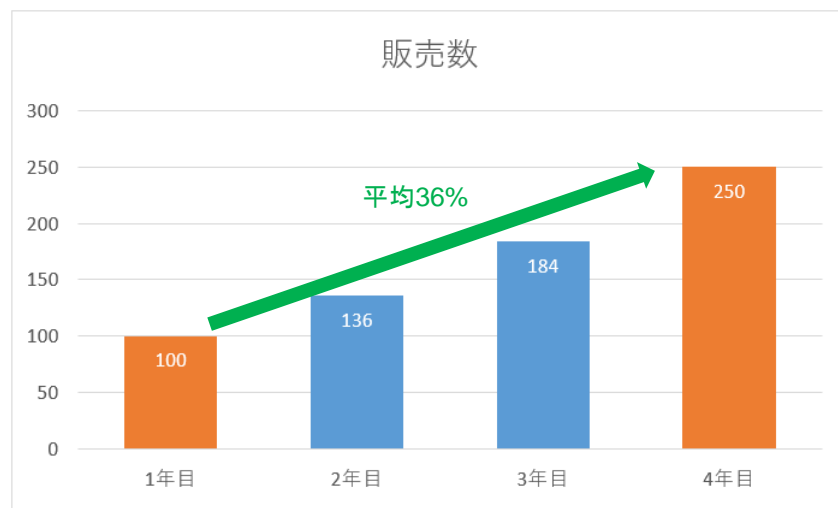
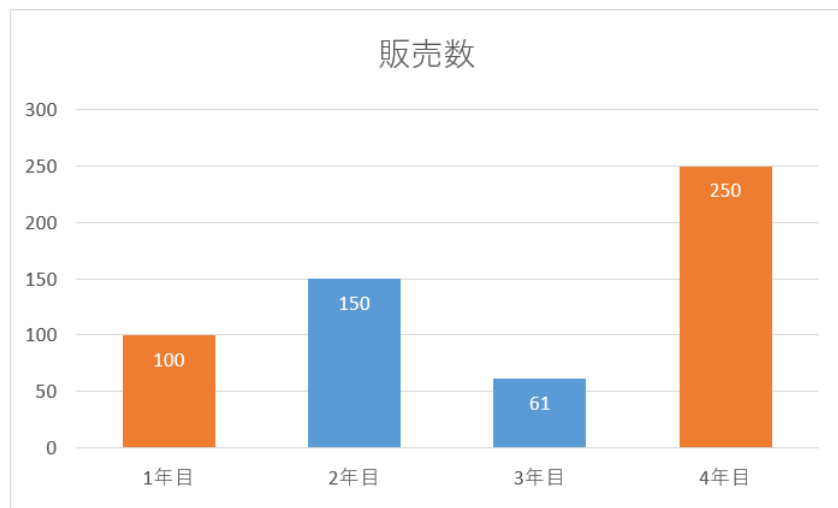
## (2) 平均の計算の応用

- 加重平均
- 条件付き平均
- 平均成長率 (CAGR)

# 平均成長率

## (1) 成長率 (CAGR = Compound Annual Growth Rate)

- 販売数の成長率の平均は36%



# 平均成長率

## (1) 成長率 (CAGR = Compound Annual Growth Rate)

- 毎年の成長率は、掛け合わせていく
- たとえば、30%の成長率が3年続く場合

×  $30\% + 30\% + 30\% = 90\%$ 成長

○  $\frac{(130\% \times 130\% \times 130\%)}{3乗} - 1 = 219\%$ 成長

## (2) 計算式

×  $(4年目 \div 1年目) \div 3 - 1$

○  $(4年目 \div 1年目) \div 3乗 - 1$

○  $(4年目 \div 1年目) ^{(1/3)} - 1$

# 目次

1. 相関分析とは
2. 近似曲線の基本(1) :  $R^2$
3. 近似曲線の基本(2) : 正の相関、負の相関
4. 近似曲線の基本(3) : 相関関係と因果関係
5. 近似曲線の基本(4) : 予測値を計算
6. データを読み解く(1) : 外れ値
7. データを読み解く(2) : グループ分け
8. データを読み解く(3) : 分解
9. データを読み解く(4) : 累計

# 目次

---

- 10. 近似曲線の応用(1)：指数近似
- 11. 近似曲線の応用(2)：対数近似
- 12. 近似曲線の応用(3)：累乗近似
- 13. 近似曲線の応用(4)：多項式近似
- 14. 最適解(1) 効率的なマーケティング
- 15. 最適解(2) マーケティング予算の分配

# 相関分析

## (1) 相関分析とは

- AとBは「本当に関係あるのか」を検証
  - 広告宣伝費をかけると売上が上がる → 本当？
  - 営業マンを増やすと売上が上がる → 本当？
- これらを、過去のデータの傾向から判断する

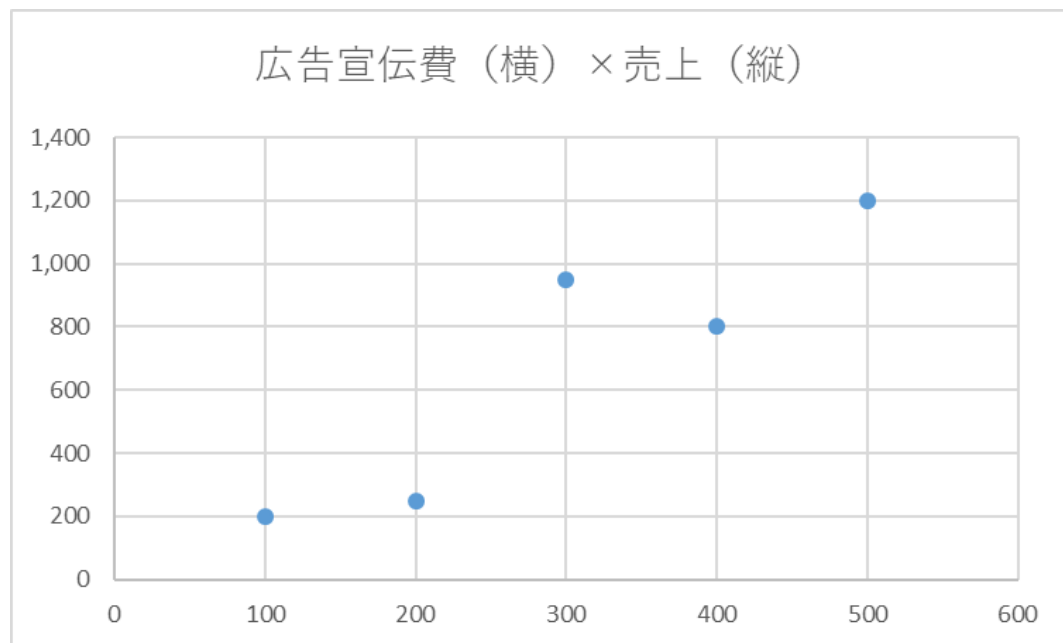
## (2) 相関分析が分かると・・・

- 自信をもってマーケティング投資の意思決定ができる
- さらに、最も効率的なマーケティング予算もシミュレーションできる

# 相関分析（例）

## (1) 広告宣伝費と売上の関係

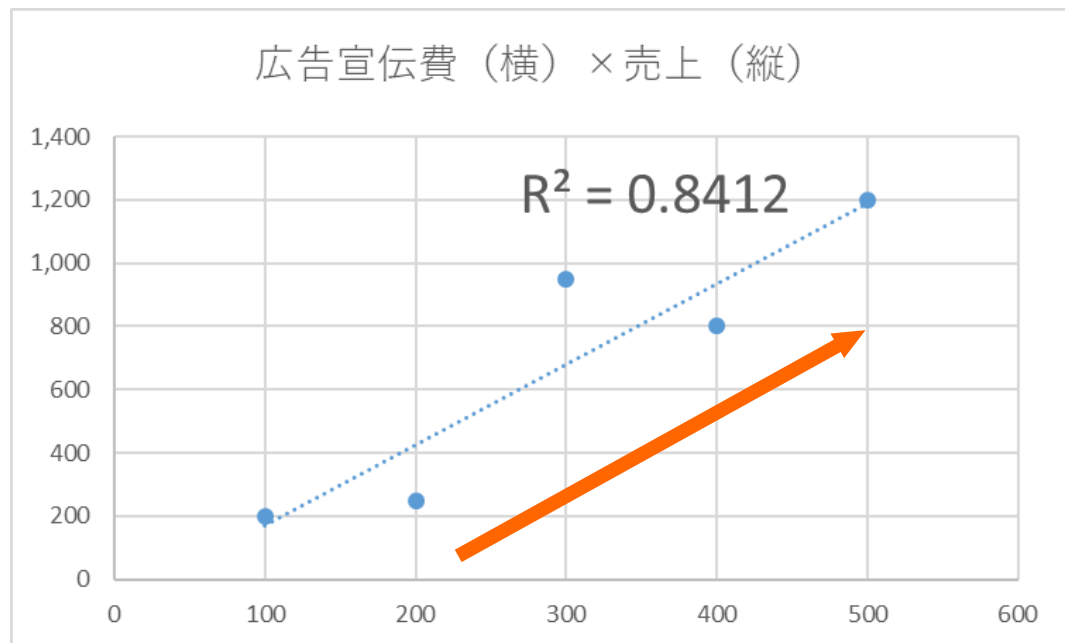
- 広告宣伝費が増えるほど売上が上がる傾向・・・？



# 相関分析（例）

## (1) 広告宣伝費と売上の関係

- 広告宣伝費が増えるほど売上が上がる傾向





# 目次

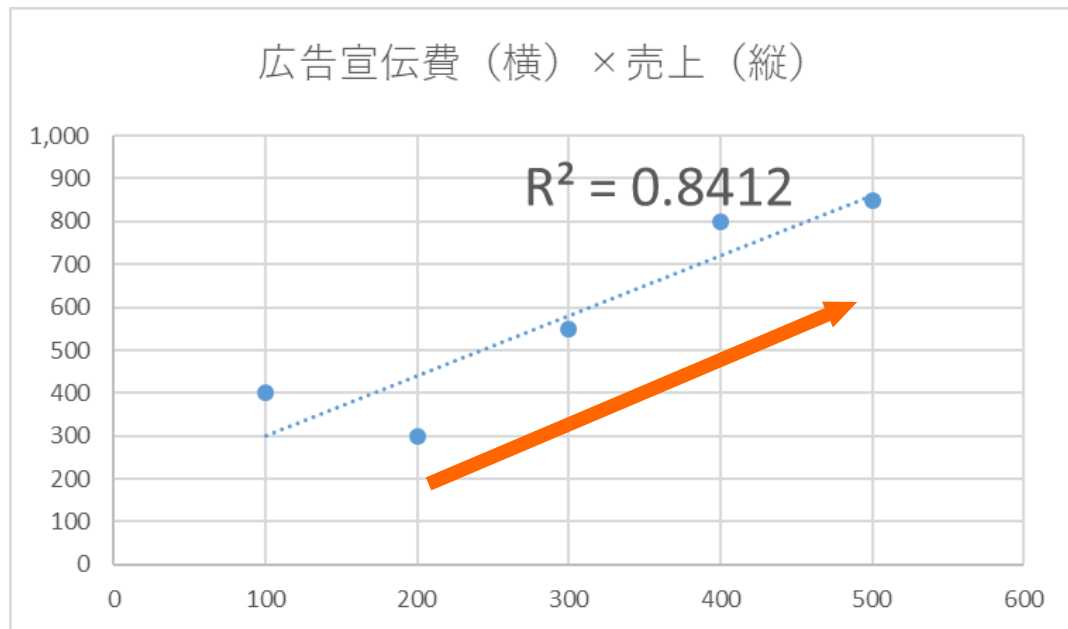
---

1. 相関分析とは
2. 近似曲線の基本(1) :  $R^2$
3. 近似曲線の基本(2) : 正の相関、負の相関
4. 近似曲線の基本(3) : 相関関係と因果関係
5. 近似曲線の基本(4) : 予測値を計算
6. データを読み解く(1) : 外れ値
7. データを読み解く(2) : グループ分け
8. データを読み解く(3) : 分解
9. データを読み解く(4) : 累計

# 相関分析（例）

## (1) 広告宣伝費と売上の関係

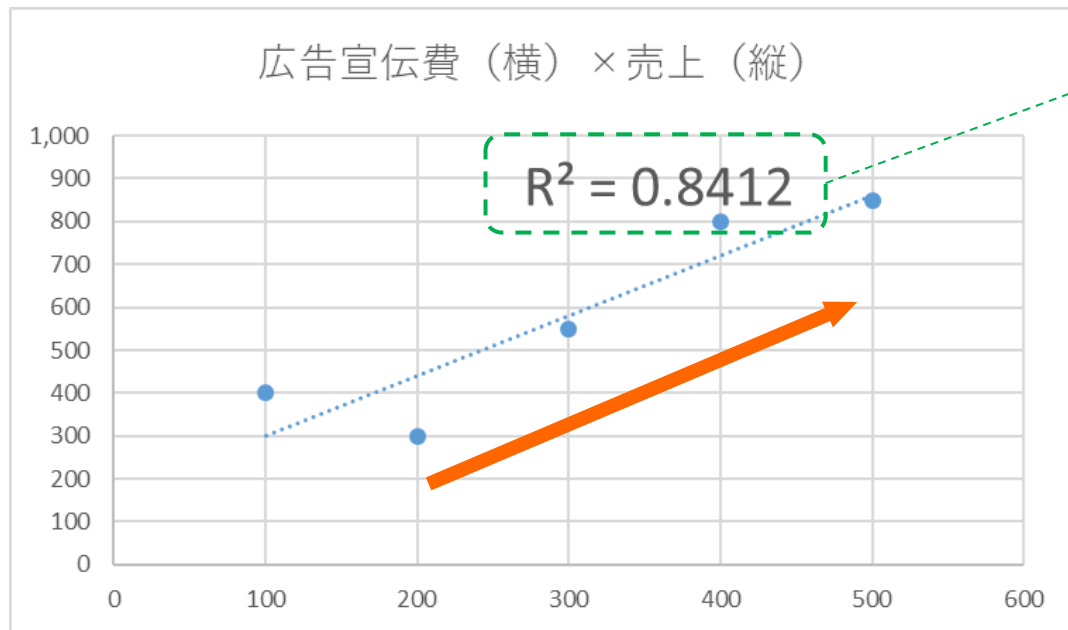
- 広告宣伝費が増えるほど売上が上がる傾向



# 相関分析（例）

## (1) 広告宣伝費と売上の関係

- 広告宣伝費が増えるほど売上が上がる傾向



2つの関係の強さを示す

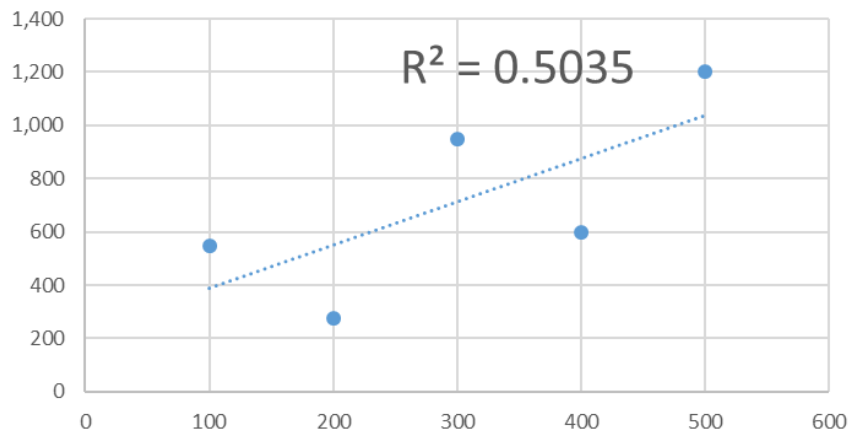
一般的には・・・

- 0.5以上 関係あり
- 0.7以上 関係が強い

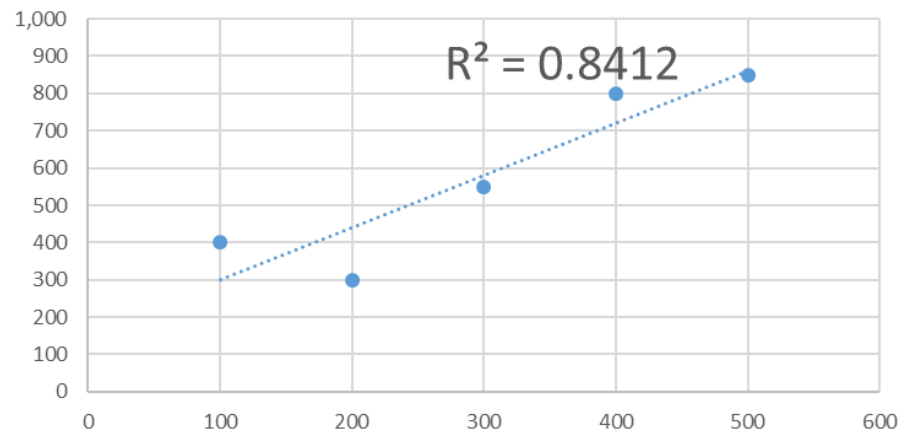
## (1) 実際にグラフで比べてみる

- 左図は0.50
- 右図は0.84 → 関係性がはっきりわかる

広告宣伝費（横） × 売上（縦）



広告宣伝費（横） × 売上（縦）



# 相関分析

## (1) 近似曲線のつくりかた

A) 散布図

B) 近似曲線

- 線形近似

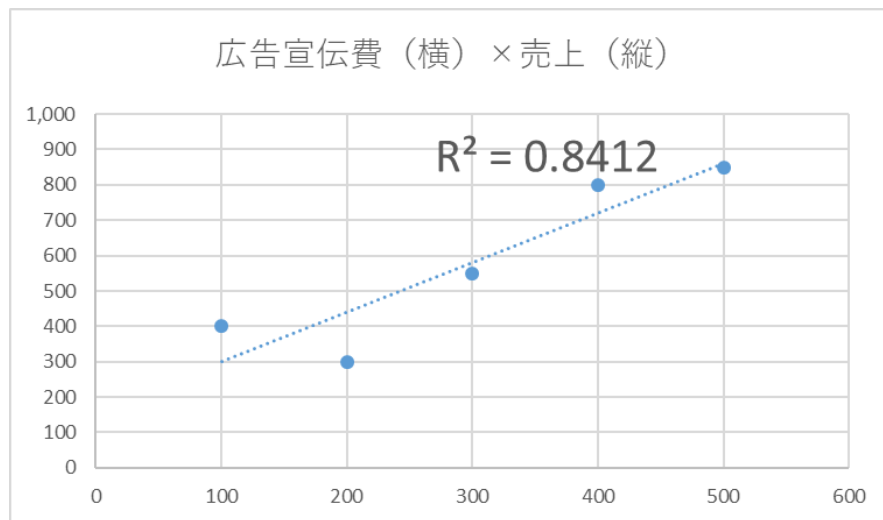
C)  $R^2$

## (2) 関数でも $R^2$ を計算できます

A)  $\text{CORREL}(\text{範囲1}, \text{範囲2}) = R$

B) このRを2乗 (^2) して $R^2$ を計算

C) 数字のバラつきは分からないので注意



# 目次

---

1. 相関分析とは
2. 近似曲線の基本(1) :  $R^2$
3. 近似曲線の基本(2) : 正の相関、負の相関
4. 近似曲線の基本(3) : 相関関係と因果関係
5. 近似曲線の基本(4) : 予測値を計算
6. データを読み解く(1) : 外れ値
7. データを読み解く(2) : グループ分け
8. データを読み解く(3) : 分解
9. データを読み解く(4) : 累計

# 正の相関、負の相関

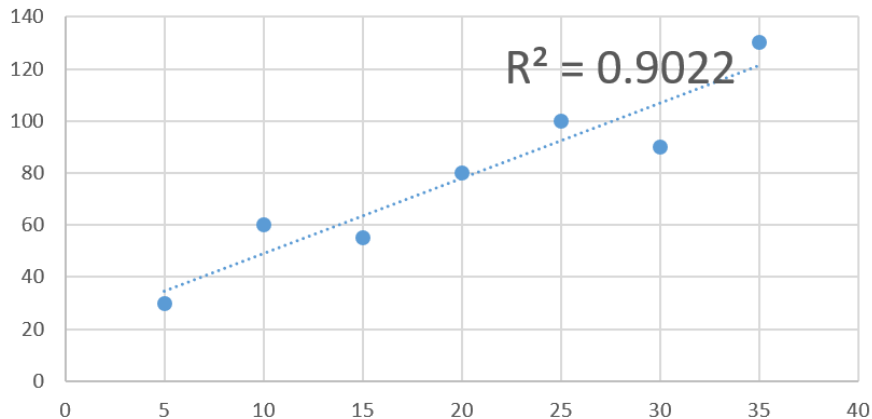
## (1) 相関の種類

A) 気温が上がれば、Tシャツが売れる → 正の相関

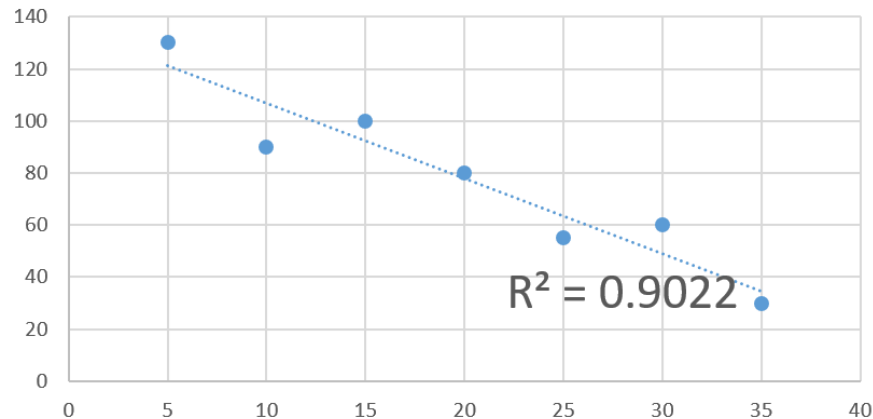
B) 気温が上がれば、コートが売れない → 負の相関

→ どちらも  $R^2$  は 0.90 なので注意！ 相関の向きもチェック！

気温（横） × Tシャツの販売数（縦）



気温（横） × コートの販売数



# 目次

---

1. 相関分析とは
2. 近似曲線の基本(1) :  $R^2$
3. 近似曲線の基本(2) : 正の相関、負の相関
4. 近似曲線の基本(3) : 相関関係と因果関係
5. 近似曲線の基本(4) : 予測値を計算
6. データを読み解く(1) : 外れ値
7. データを読み解く(2) : グループ分け
8. データを読み解く(3) : 分解
9. データを読み解く(4) : 累計



# 相関関係と、因果関係

## (1) 相関

A) 気温とTシャツの販売数には関係がある

○ 気温が上がる → Tシャツが売れる

× Tシャツが売れる → 気温が上がる

B) 相関分析だけでは、どちらが原因で、どちらが結果か分かりにくい

## (2) 因果関係

A) 原因と結果

B) 納得感のある仮説を考える

C) . . . それって逆じゃないか？と疑問をもつ

# 因果関係を突き止めるためのテクニック

## (1) 反事実

A) 気温が上がれば、Tシャツが売れる

- 気温が上がらなかったら、Tシャツは売れない？ 正しい

B) Tシャツが売れば、気温が上がる

- Tシャツが売れなかったら、気温は下がる？ 誤り

## (2) テストしてみる

A) テレビCMのおかげで商品売上が増加した（本当？）

- 1ヶ月テレビCMを止めてみて、売上が減少するか検証

# 目次

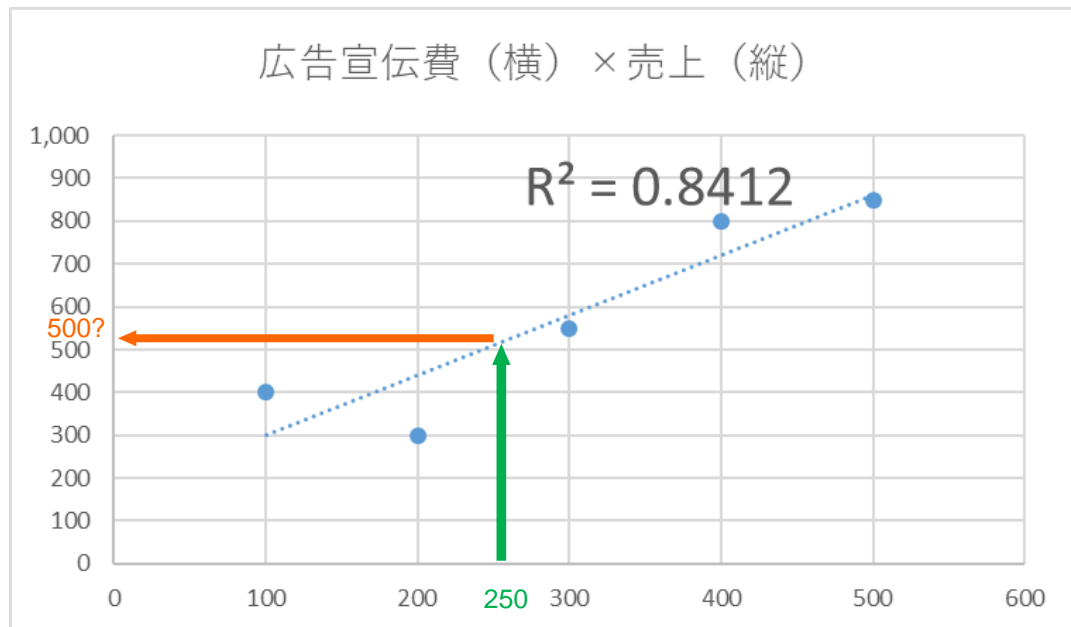
---

1. 相関分析とは
2. 近似曲線の基本(1) :  $R^2$
3. 近似曲線の基本(2) : 正の相関、負の相関
4. 近似曲線の基本(3) : 相関関係と因果関係
5. 近似曲線の基本(4) : 予測値を計算
6. データを読み解く(1) : 外れ値
7. データを読み解く(2) : グループ分け
8. データを読み解く(3) : 分解
9. データを読み解く(4) : 累計

# 予測値を計算する

## (1) 広告宣伝費と売上の関係

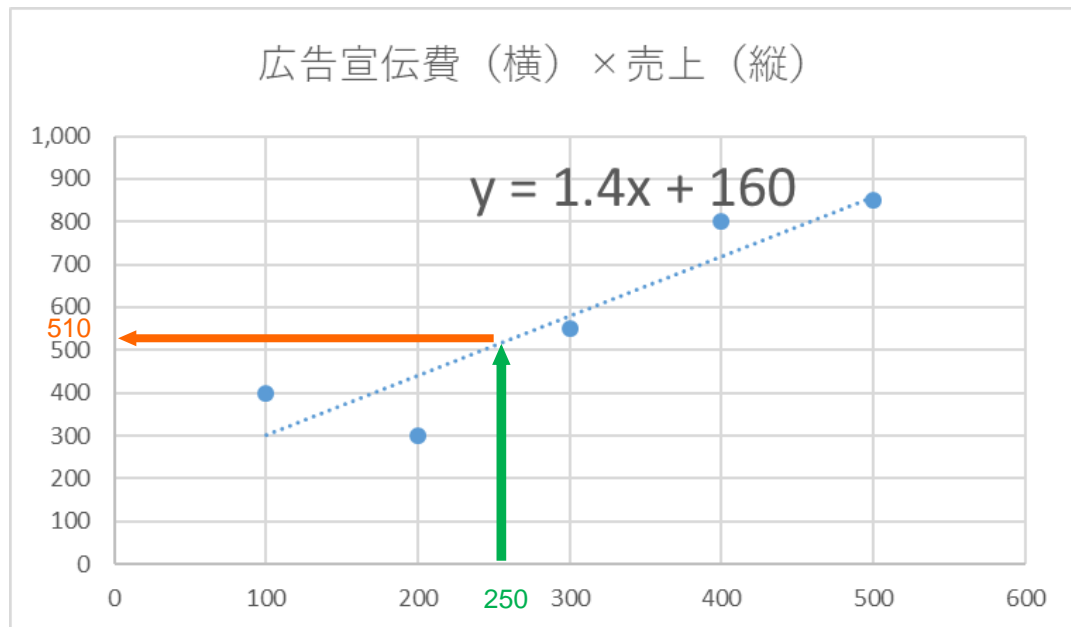
- 広告宣伝費を250にすると、売上は500くらい？



# 予測値を計算する

## (1) 近似曲線を計算式にする

- $y$  (縦軸：売上) =  $1.4 \times$  (横軸：広告宣伝費) + 160
- 510 =  $1.4 \times 250 + 160$



# 予測値を計算する

---

## (1) 近似曲線

- A) 計算式を表示する
- B) 傾き、切片を表示
- C)  $x$  を計算式に代入して、 $y$ を求める

## (2) 関数でも計算できます

- A) 傾き：SLOPE (y軸の範囲、x軸の範囲)
- B) 切片：INTERCEPT (y軸の範囲、x軸の範囲)

# 目次

---

1. 相関分析とは
2. 近似曲線の基本(1) :  $R^2$
3. 近似曲線の基本(2) : 正の相関、負の相関
4. 近似曲線の基本(3) : 相関関係と因果関係
5. 近似曲線の基本(4) : 予測値を計算
6. データを読み解く(1) : 外れ値
7. データを読み解く(2) : グループ分け
8. データを読み解く(3) : 分解
9. データを読み解く(4) : 累計

# データを読み解く

---

## (1) ビジネスにおけるデータの使い方

- A) すべてのデータが有益とは限らない
- B) データの中には例外（外れ値）も存在する
- C) あるいは、グループが異なる場合もある

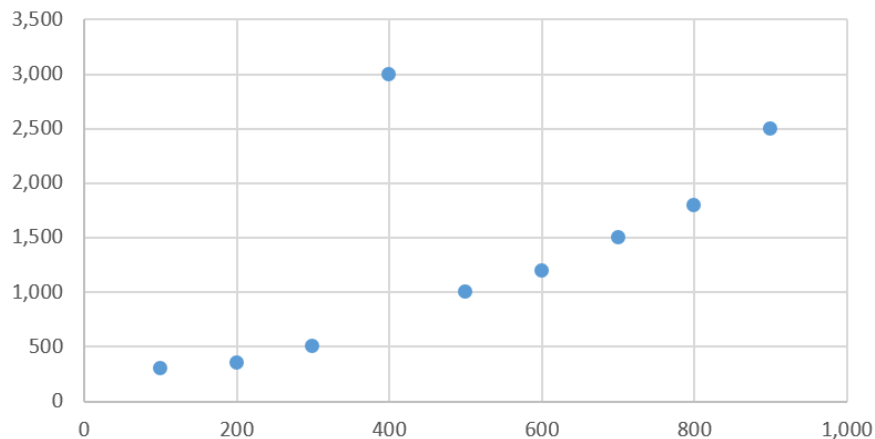


# データを読み解く

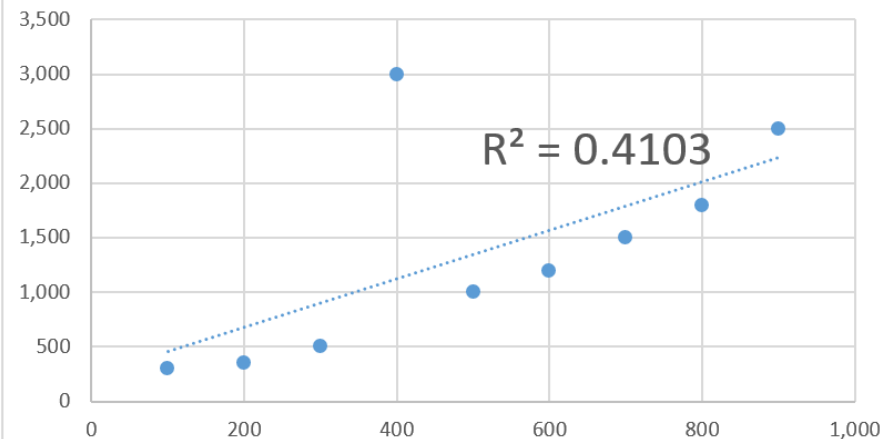
## (1) 広告宣伝費と売上

A)  $R^2$ は0.41と高くない・・・なぜか？

広告宣伝費（横）×売上（縦）



広告宣伝費（横）×売上（縦）



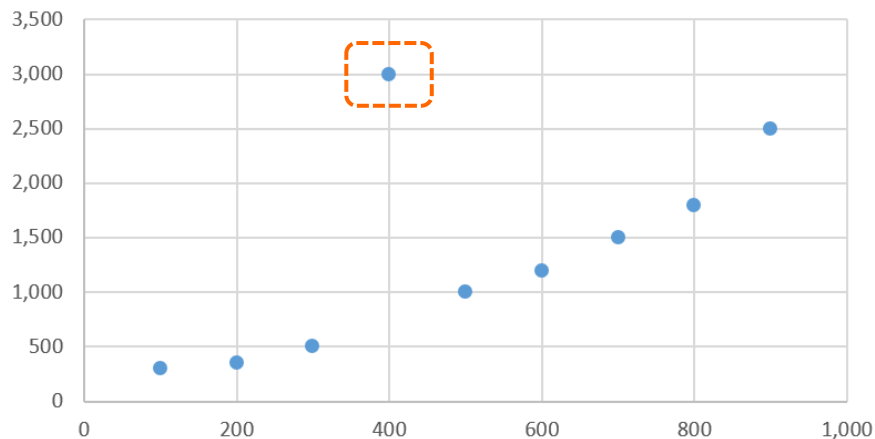
# データを読み解く

## (1) 広告宣伝費と売上

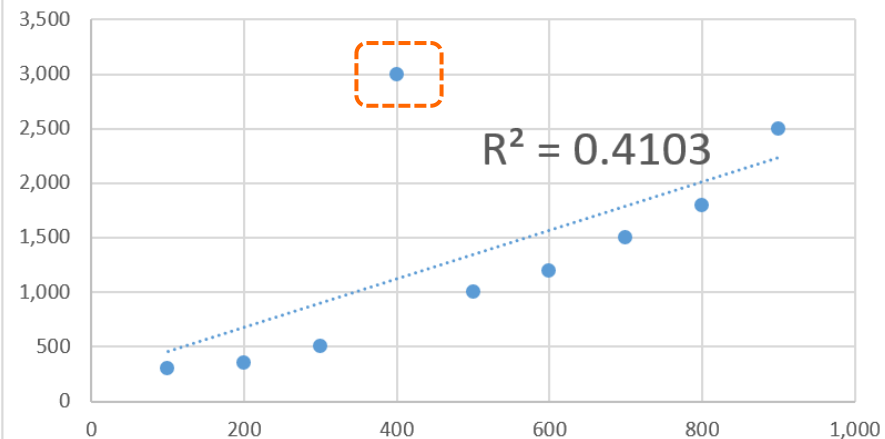
A)  $R^2$ は0.41と高くない・・・なぜか？

B) ひとつだけ外れている数字がある

広告宣伝費（横） × 売上（縦）



広告宣伝費（横） × 売上（縦）

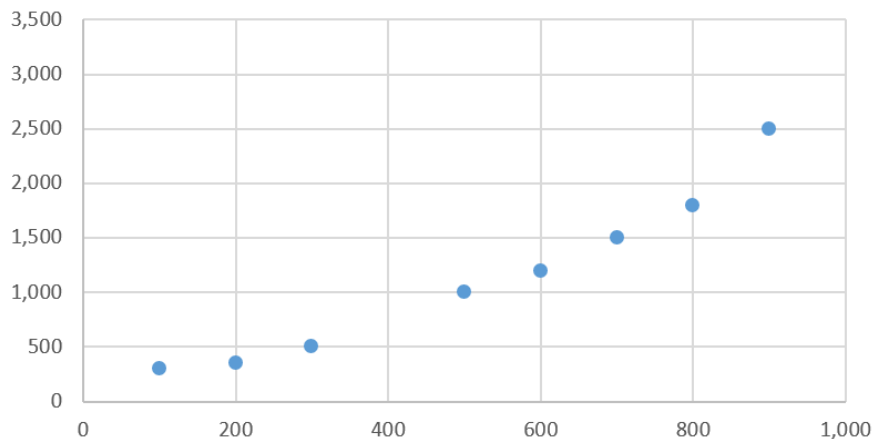


# データを読み解く

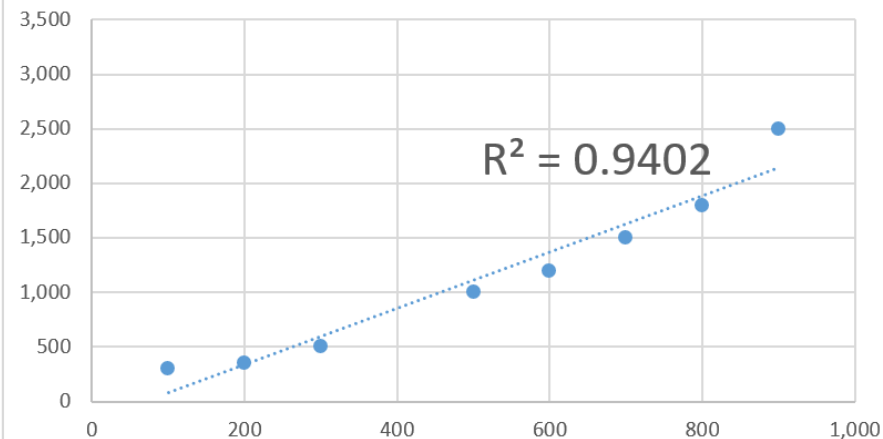
## (1) 広告宣伝費と売上

A) 外れた数字を使わないと、相関が高い

広告宣伝費（横） × 売上（縦）



広告宣伝費（横） × 売上（縦）



# データを読み解く

---

## (1) 外れ値

- A) データの中には例外（外れ値）も存在する
- B) この数字を相関の計算から外す場合があります

## (2) 季節要因

- A) クリスマスで、広告宣伝費をかけなくても商品の売上が伸びた

## (3) 一時要因

- A) 有名人が紹介してくれた

# 目次

---

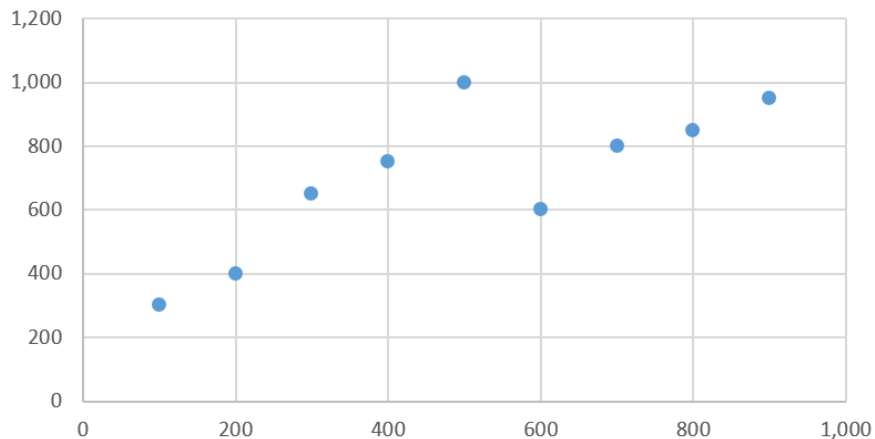
1. 相関分析とは
2. 近似曲線の基本(1) :  $R^2$
3. 近似曲線の基本(2) : 正の相関、負の相関
4. 近似曲線の基本(3) : 相関関係と因果関係
5. 近似曲線の基本(4) : 予測値を計算
6. データを読み解く(1) : 外れ値
7. データを読み解く(2) : グループ分け
8. データを読み解く(3) : 分解
9. データを読み解く(4) : 累計

# データを読み解く

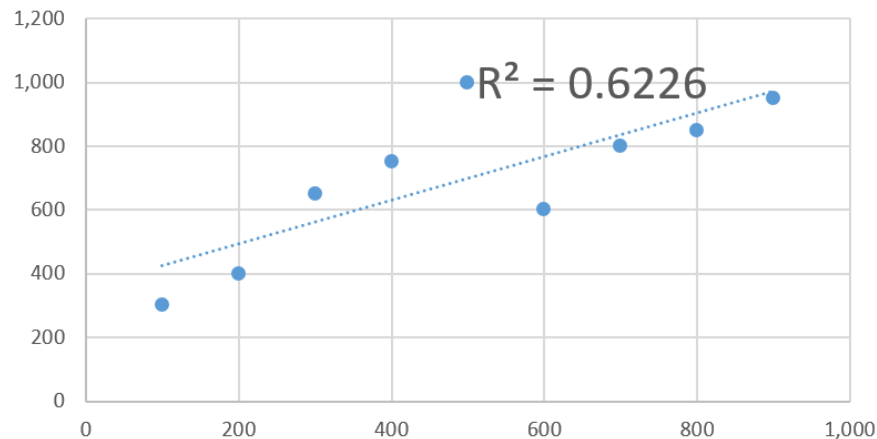
## (1) 広告宣伝費と売上

- $R^2$ は0.62と、それほど高くない・・・なぜか？

広告宣伝費（横） × 売上（縦）



広告宣伝費（横） × 売上（縦）

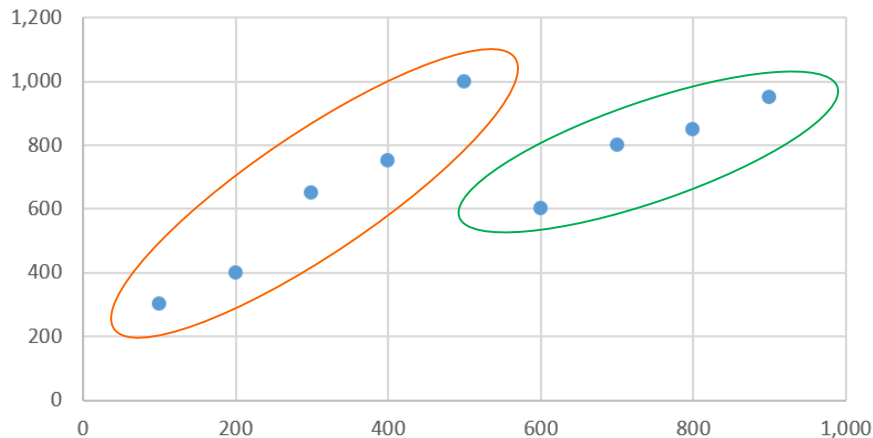


# データを読み解く

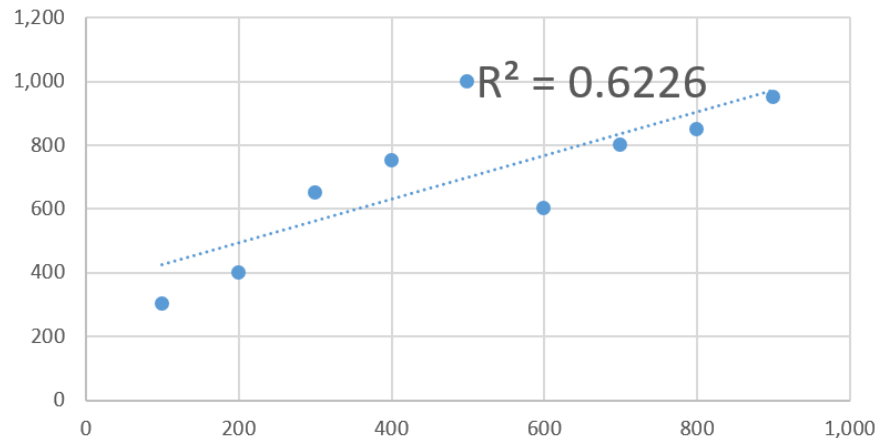
## (1) 広告宣伝費と売上

- $R^2$ は0.62と、それほど高くない・・・なぜか？
- おそらく2つのグループに分かれているのではないか（仮説）

広告宣伝費（横）×売上（縦）



広告宣伝費（横）×売上（縦）

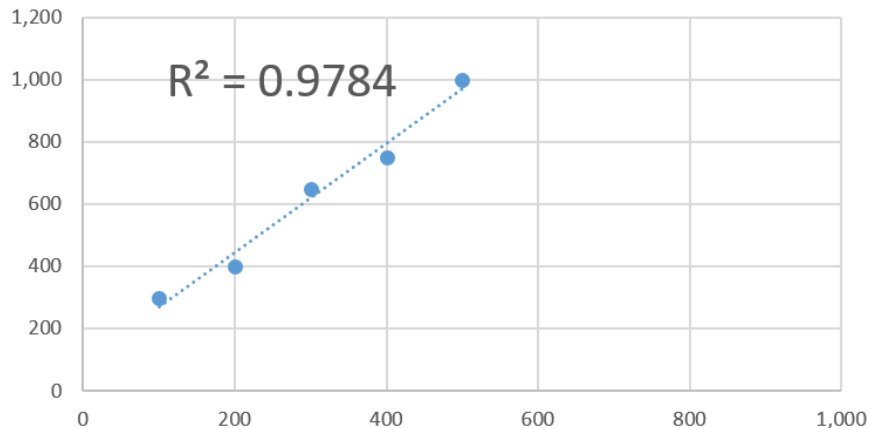


# データを読み解く

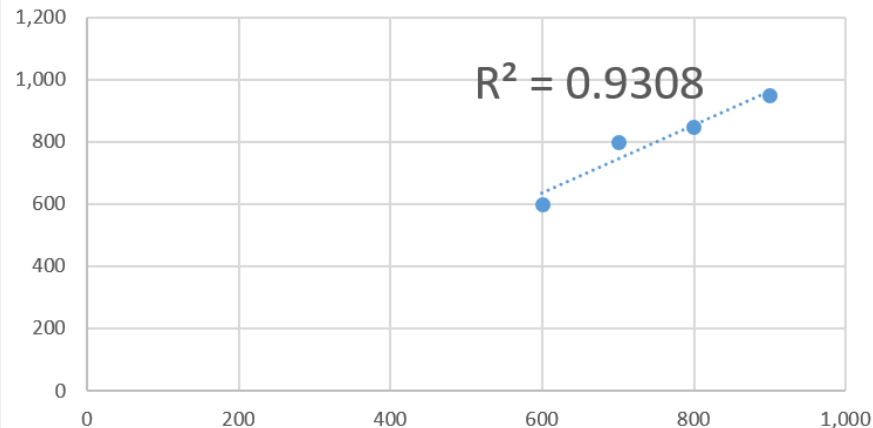
## (1) 広告宣伝費と売上

- 2つのグループを分けて相関をとると、それぞれ高い数値になる

広告宣伝費（横） × 売上（縦）



広告宣伝費（横） × 売上（縦）





# データを読み解く

---

## (1) グループ

- A) データの中には異なるグループが混在している可能性がある
- B) グループごとに分けて相関を計算する場合があります

## (2) 集団の違い

- A) 40代以下の顧客と、50代以上の顧客

## (3) 時期の違い

- A) 消費税が増税される前は、広告宣伝すれば売上は伸びた
- B) 増税してからは、なかなか売上は伸びなくなった

# 目次

---

1. 相関分析とは
2. 近似曲線の基本(1) :  $R^2$
3. 近似曲線の基本(2) : 正の相関、負の相関
4. 近似曲線の基本(3) : 相関関係と因果関係
5. 近似曲線の基本(4) : 予測値を計算
6. データを読み解く(1) : 外れ値
7. データを読み解く(2) : グループ分け
8. データを読み解く(3) : 分解
9. データを読み解く(4) : 累計

# データを読み解く

## (1) 売上と人件費

A) 「もっと人件費をかければ、売上が増えるはず」

B) 本当にそう言えるだろうか？

### 売上と人件費

		1年目	2年目	3年目	4年目	5年目	6年目	7年目	8年目	9年目	10年目
売上	百万円	300	250	280	1,200	720	1,500	1,440	1,800	650	1,500
顧客数	社	30	50	40	60	90	100	120	120	130	150
顧客単価	百万円	10	5	7	20	8	15	12	15	5	10
人件費	百万円	50	75	100	200	240	280	400	450	500	550
営業マン数	人	10	15	20	25	30	35	40	45	50	55
1人あたり人件費	百万円	5	5	5	8	8	8	10	10	10	10

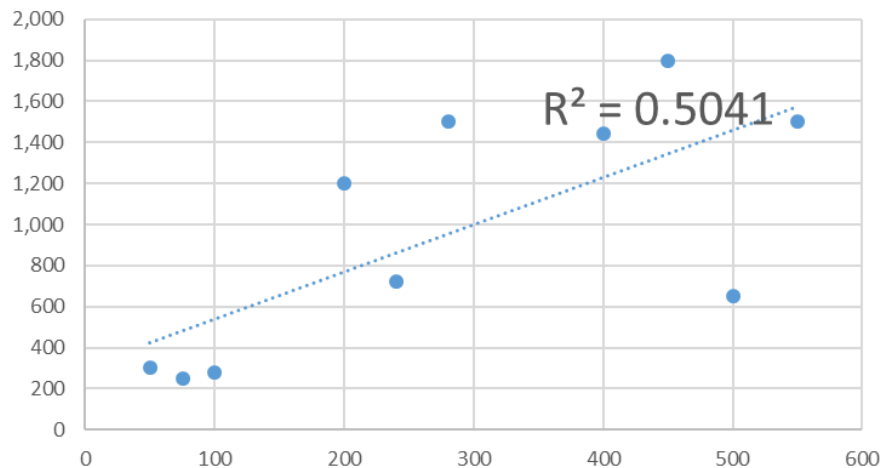
# データを読み解く

## (1) 売上と人件費

A) 「もっと人件費をかければ、売上が増えるはず」

B) あまり関係はなさそう

人件費（横） × 売上（縦）



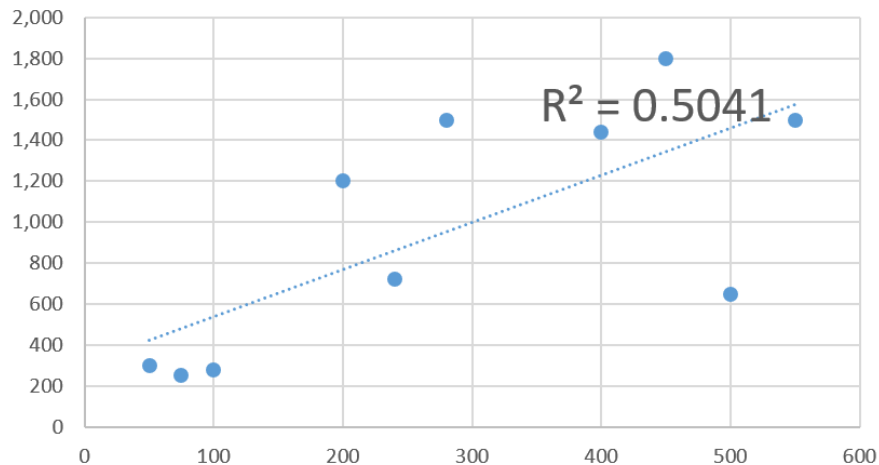
# データを読み解く

## (1) 売上と人件費

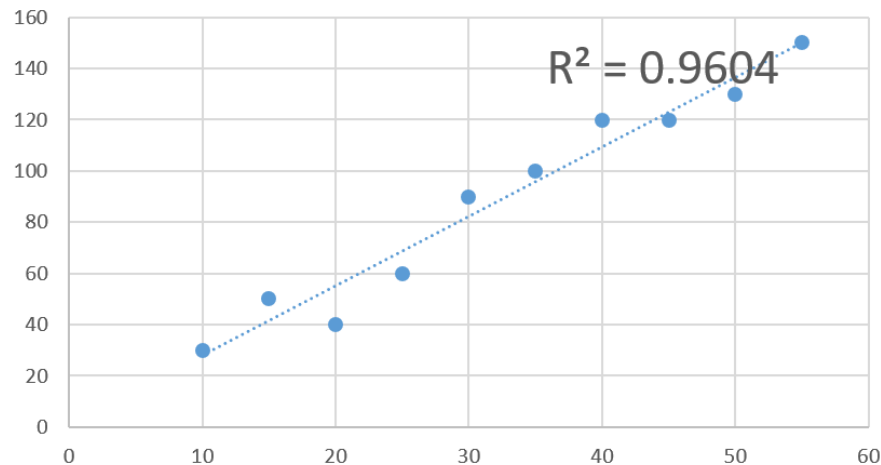
A) 「もっと人件費をかければ、売上が増えるはず」

B) でも、営業マン数を増やせば、顧客数は増えそう

人件費（横） × 売上（縦）



営業マン数（横） × 顧客数（縦）



# データを読み解く

---

## (1) データは分解して考える

- A) 人件費を増やせば、売上が増える → そうとは限らない
- B) 顧客単価が上がるとは限らないから

## (2) データを分解すれば、課題とアイデアが明確になる

A)  $\text{売上} = \text{顧客数} \times \text{顧客単価}$

- 顧客数を増やしたい → 営業マンを増やす
- 顧客単価を上げたい → 他の施策を考えたほうがよさそう

# 目次

---

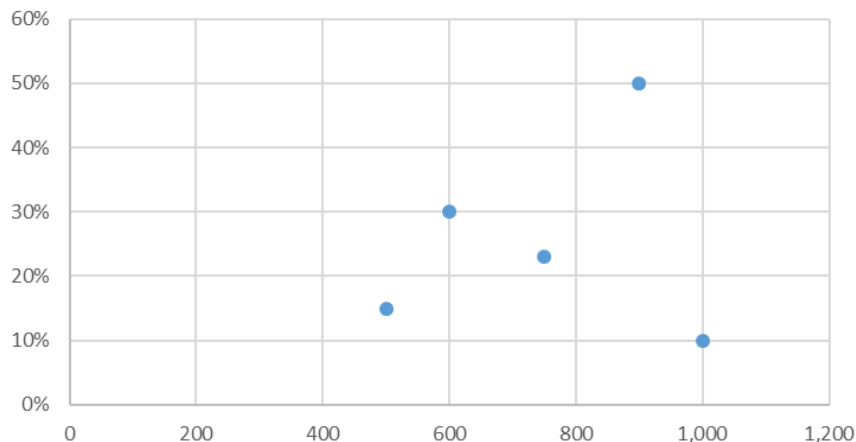
1. 相関分析とは
2. 近似曲線の基本(1) :  $R^2$
3. 近似曲線の基本(2) : 正の相関、負の相関
4. 近似曲線の基本(3) : 相関関係と因果関係
5. 近似曲線の基本(4) : 予測値を計算
6. データを読み解く(1) : 外れ値
7. データを読み解く(2) : グループ分け
8. データを読み解く(3) : 分解
9. データを読み解く(4) : 累計

# データを読み解く

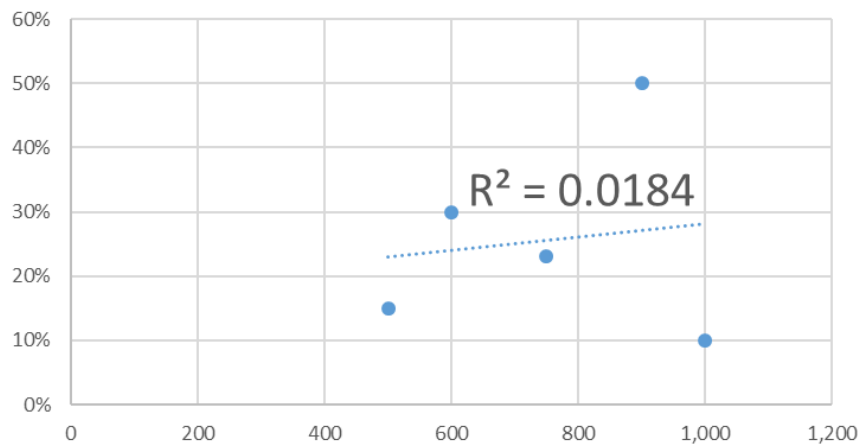
## (1) テレビCM費用（年間）と知名度の関係

A) 関係なさそう . . .

テレビCM費用（年間） × 知名度



テレビCM費用（年間） × 知名度





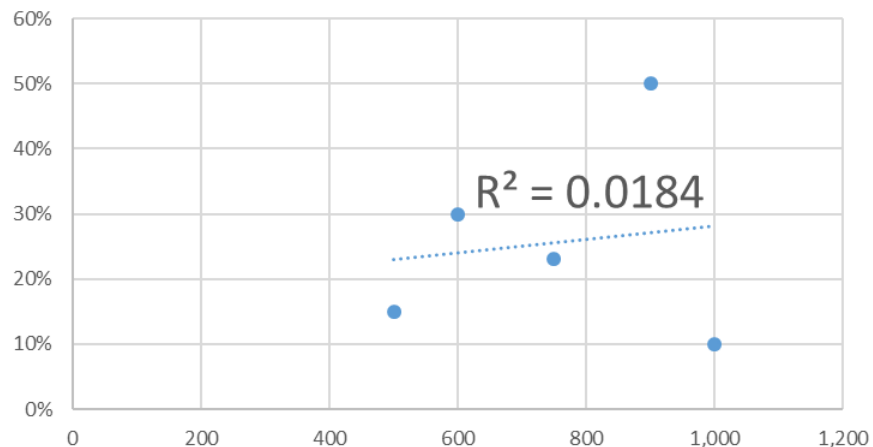
# データを読み解く

## (1) 知名度は、何で決まるのか？

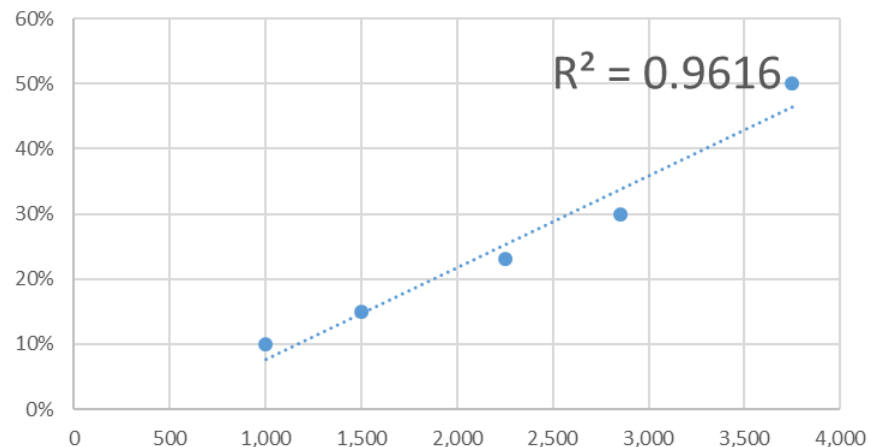
A) 知名度 = これまでのテレビCM費用の累計によって決まるのでは？

B) テレビCM費用（累計）で見ると、知名度アップにつながっている

テレビCM費用（年間） × 知名度



テレビCM費用（累計） × 知名度



# 目次

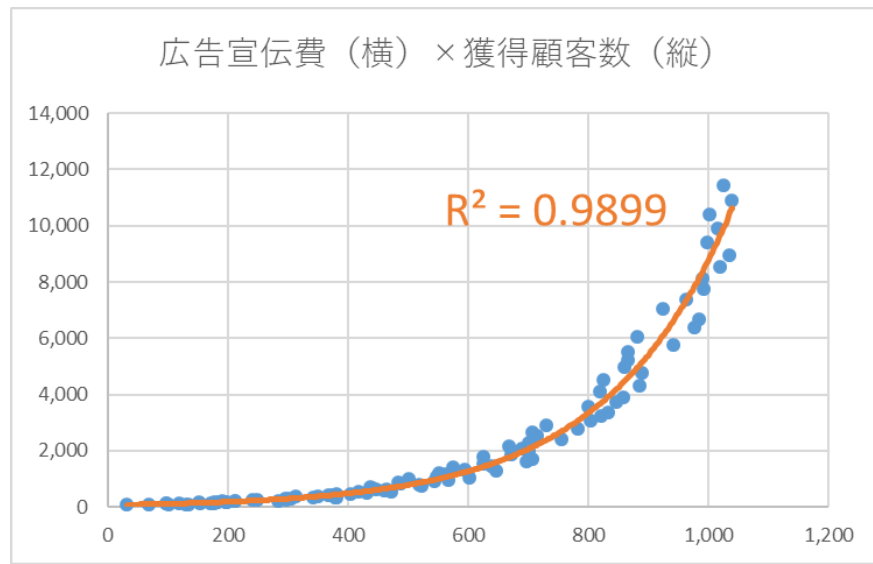
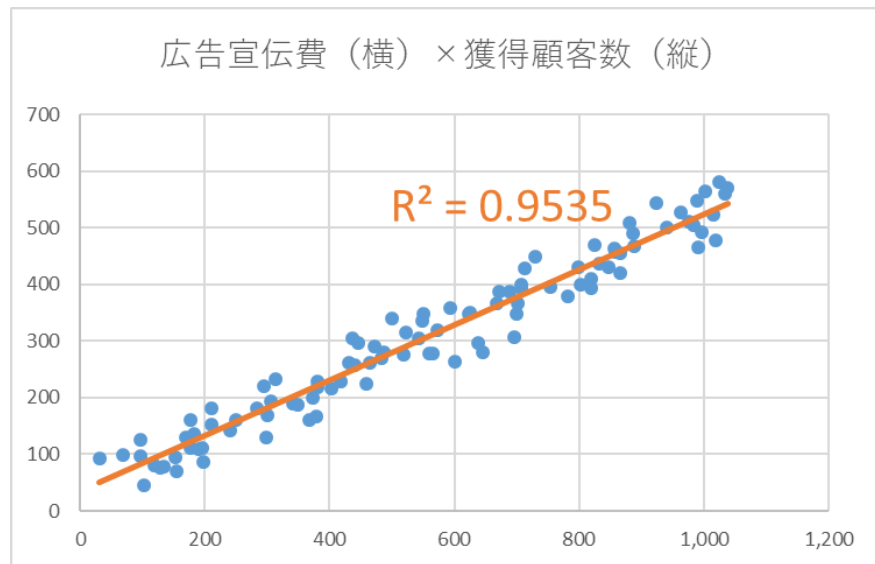
---

- 10. 近似曲線の応用(1)：指数近似
- 11. 近似曲線の応用(2)：対数近似
- 12. 近似曲線の応用(3)：累乗近似
- 13. 近似曲線の応用(4)：多項式近似
- 14. 最適解(1) 効率的なマーケティング
- 15. 最適解(2) マーケティング予算の分配

# 近似曲線の応用

## (1) 近似曲線の種類

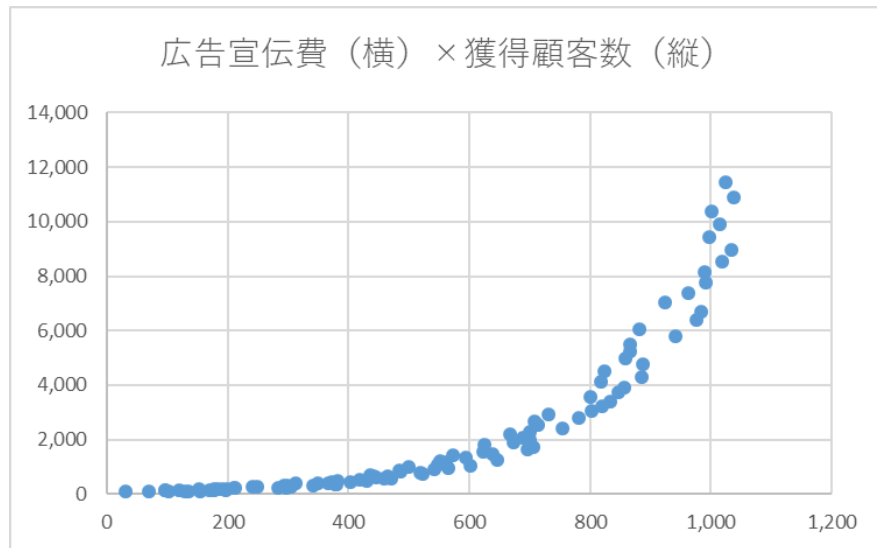
- A) 必ずしも直線（線形）とは限らない
- B) マーケティングの特徴を考えながら、適切な線を考える



# 近似曲線の応用

## (1) 広告宣伝費と、獲得顧客数の関係

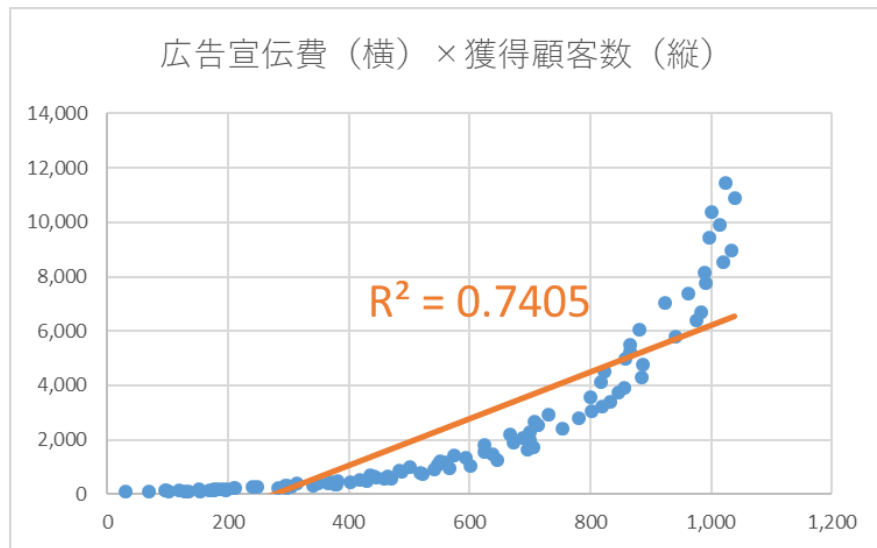
- 広告宣伝費をかけるほど・・・獲得顧客数は？



# 近似曲線の応用

## (1) 広告宣伝費と、獲得顧客数の関係

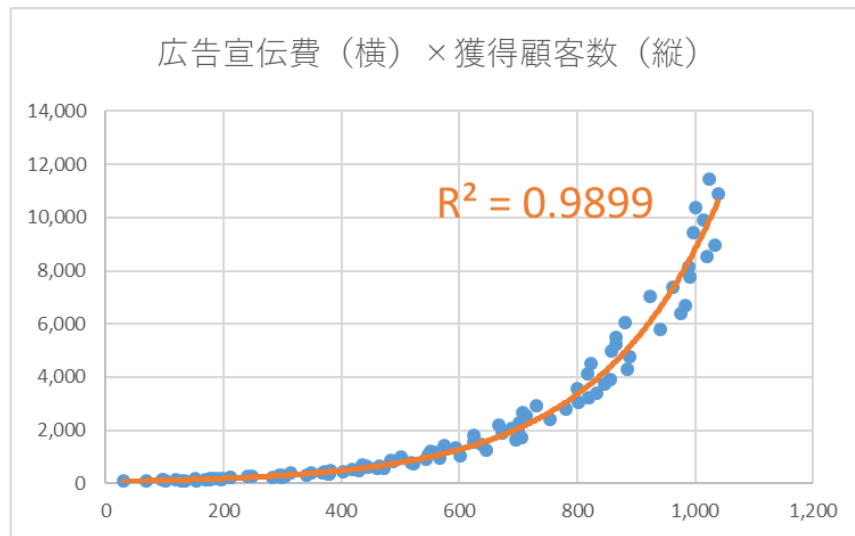
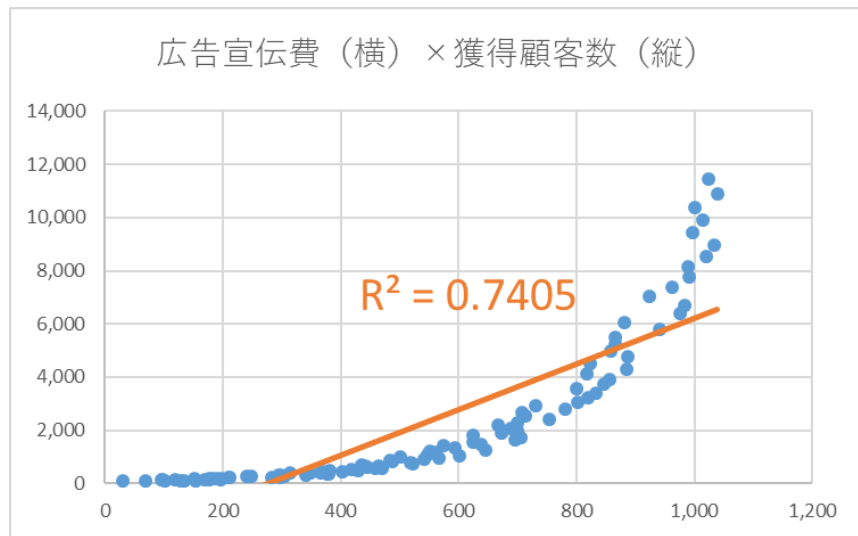
- 少しイメージと違う？



# 近似曲線の応用

## (1) 広告宣伝費と、獲得顧客数の関係

- 広告宣伝費をかけるほど、それ以上に顧客を獲得できている



# 近似曲線の応用

---

## (1) 今回のマーケティングの仮説

A) 口コミで広がるケース

B) 利用者が、どんどん友人に商品の良さを紹介

C) その友人がまた別の友人に紹介

→ 広告費を増やすと、それ以上のペースでユーザー数が増えていく

→ 指数近似

# 目次

---

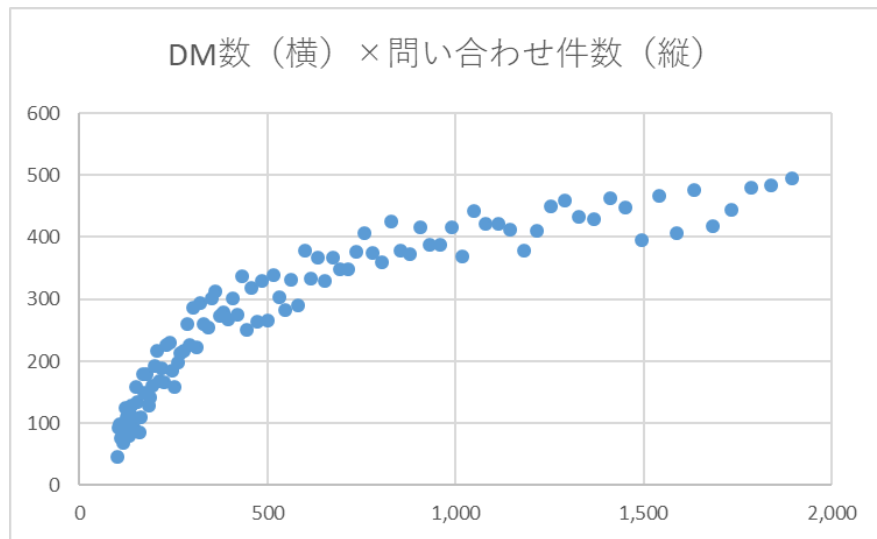
- 10. 近似曲線の応用(1)：指数近似
- 11. 近似曲線の応用(2)：対数近似
- 12. 近似曲線の応用(3)：累乗近似
- 13. 近似曲線の応用(4)：多項式近似
- 14. 最適解(1) 効率的なマーケティング
- 15. 最適解(2) マーケティング予算の分配



# 近似曲線の応用

## (1) ダイレクトメール (DM) 数と、購入希望の問い合わせ件数

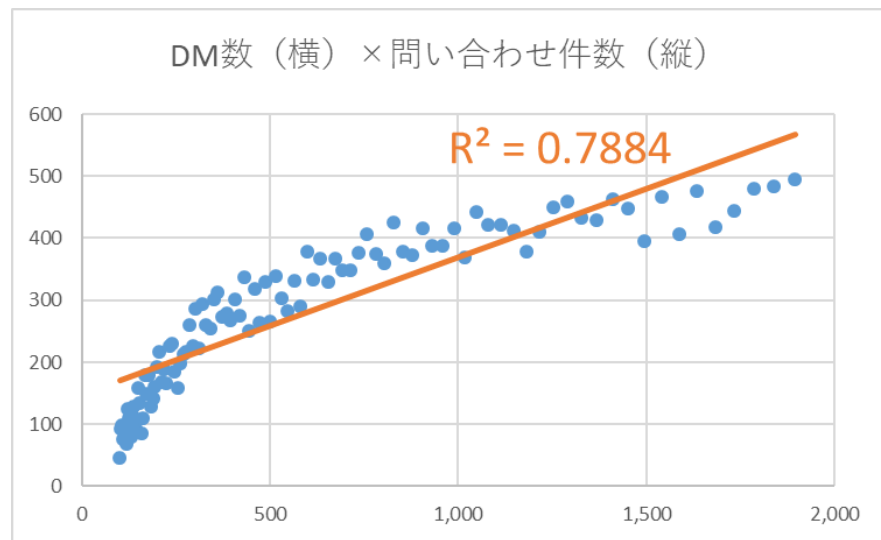
- DMを送るほど問い合わせは増えるか？



# 近似曲線の応用

## (1) ダイレクトメール（DM）数と、購入希望の問い合わせ件数

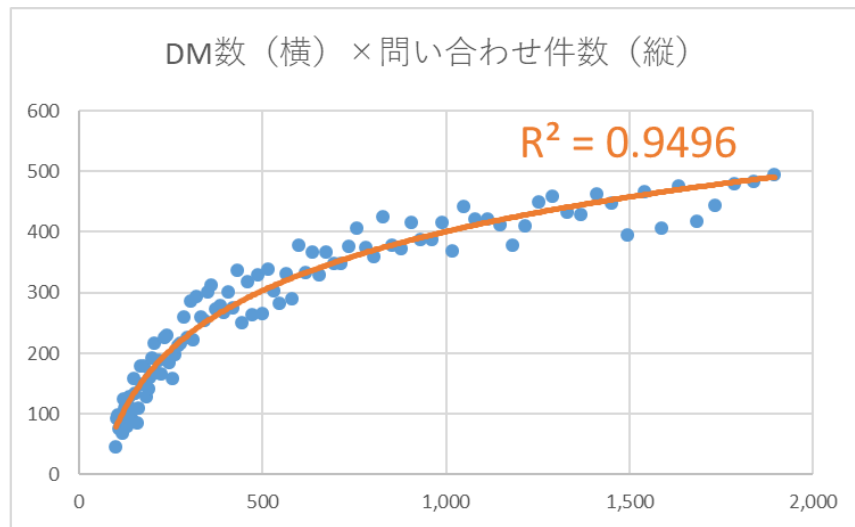
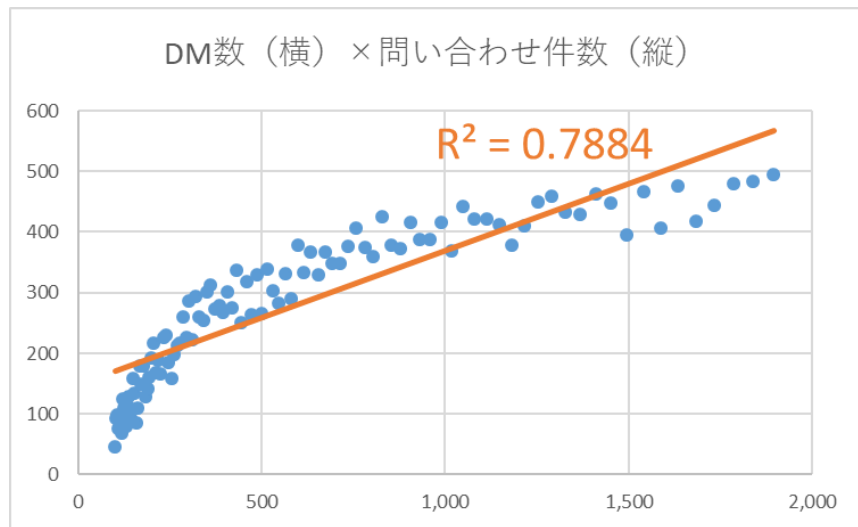
- DMが増える（右にいく）ほど、あまり問い合わせが増えていない？



# 近似曲線の応用

## (1) ダイレクトメール (DM) 数と、購入希望の問い合わせ件数

- DMが増える（右にいく）ほど、頭打ちになっている
- 500通以上のDMはムダといえるかもしれない



# 近似曲線の応用

## (1) 今回のマーケティングの仮説

- A) DMは、購入が見込める顧客から順番に送っているのでは？
- B) 最初は、DMを送るほど顧客も獲得できる
- C) DMを送りすぎると、購入が見込めない顧客にまで送ってしまう  
→対数近似

# 目次

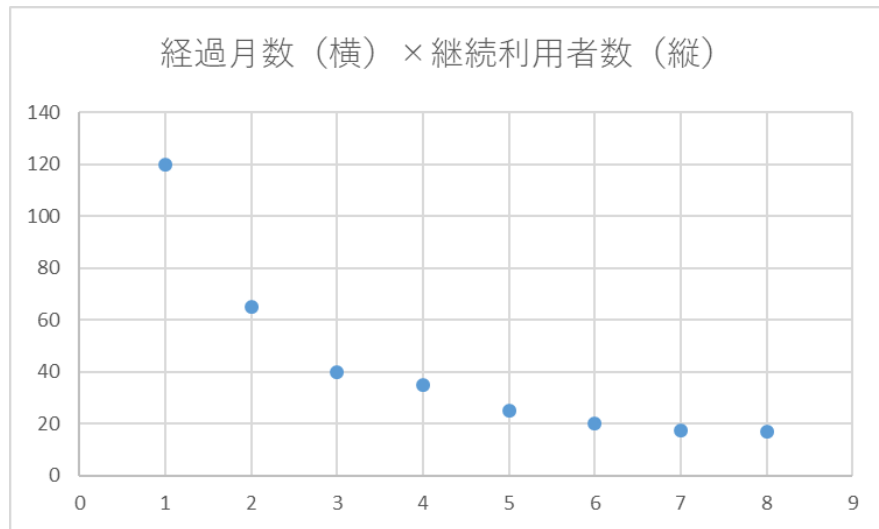
---

- 10. 近似曲線の応用(1)：指数近似
- 11. 近似曲線の応用(2)：対数近似
- 12. 近似曲線の応用(3)：累乗近似
- 13. 近似曲線の応用(4)：多項式近似
- 14. 最適解(1) 効率的なマーケティング
- 15. 最適解(2) マーケティング予算の分配

# 近似曲線の応用

## (1) ユーザー獲得してからの経過月数と、継続利用者数

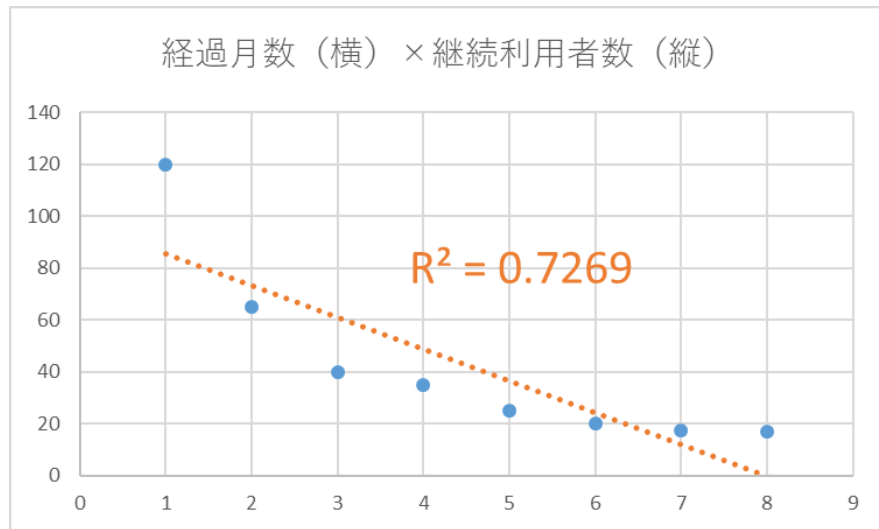
- 時間がたつほど、継続利用者数は減っている？



# 近似曲線の応用

## (1) ユーザー獲得してからの経過月数と、継続利用者数

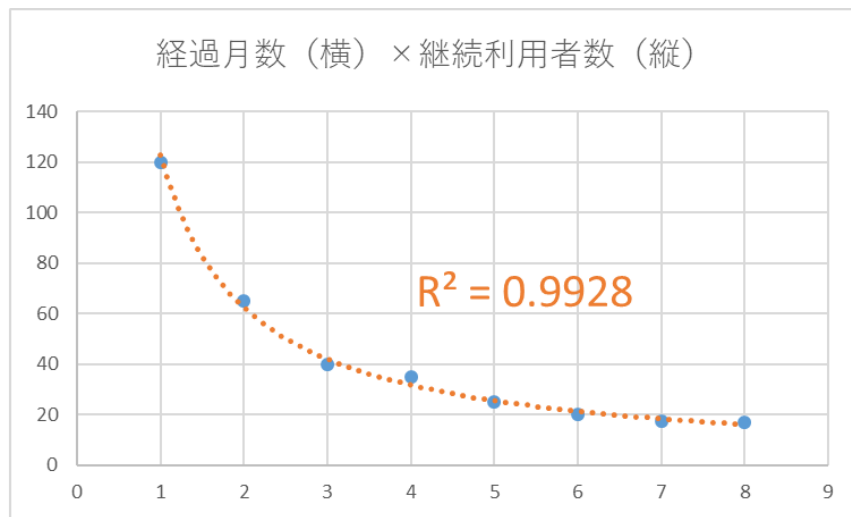
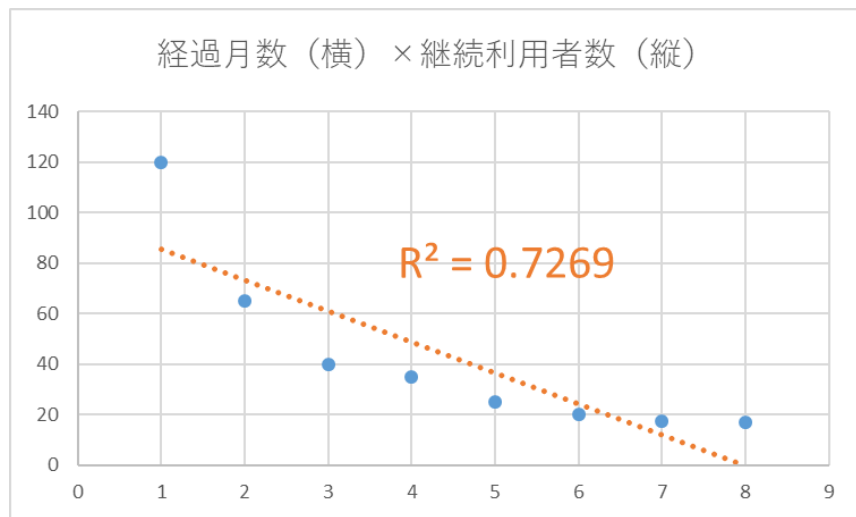
- 直線的ではない？



# 近似曲線の応用

## (1) ユーザー獲得してからの経過月数と、継続利用者数

- はじめは継続利用者数は急速に減少
- 時間がたつと、利用者数はあまり変わらない





# 近似曲線の応用

---

## (1) 今回のマーケティングの仮説

A) 継続利用ペースは、一定ではない

B) ライトユーザー（お試しで始めてみた）は、早めに離脱

C) 時間がたつと、ヘビーユーザーだけが残る

→ 累乗近似

# 目次

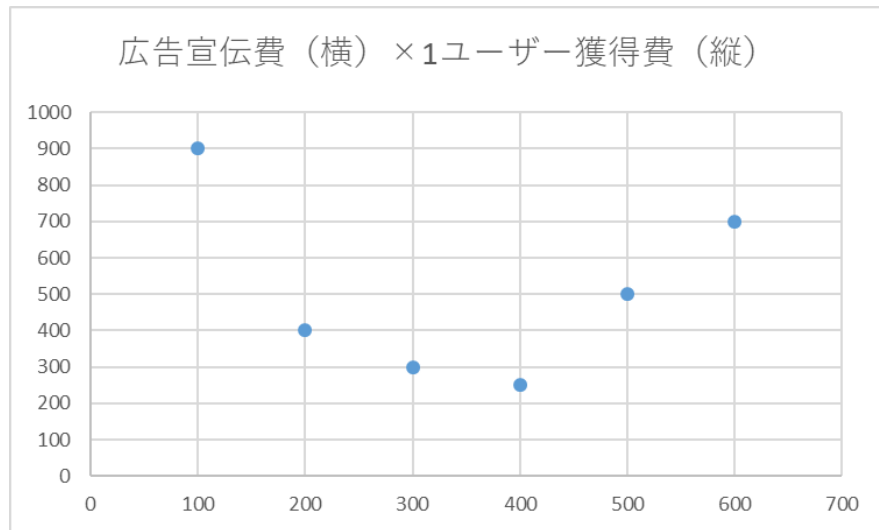
---

- 10. 近似曲線の応用(1)：指数近似
- 11. 近似曲線の応用(2)：対数近似
- 12. 近似曲線の応用(3)：累乗近似
- 13. 近似曲線の応用(4)：多項式近似
- 14. 最適解(1) 効率的なマーケティング
- 15. 最適解(2) マーケティング予算の分配

# 近似曲線の応用

## (1) 広告宣伝費と、1ユーザーあたり獲得費用の関係

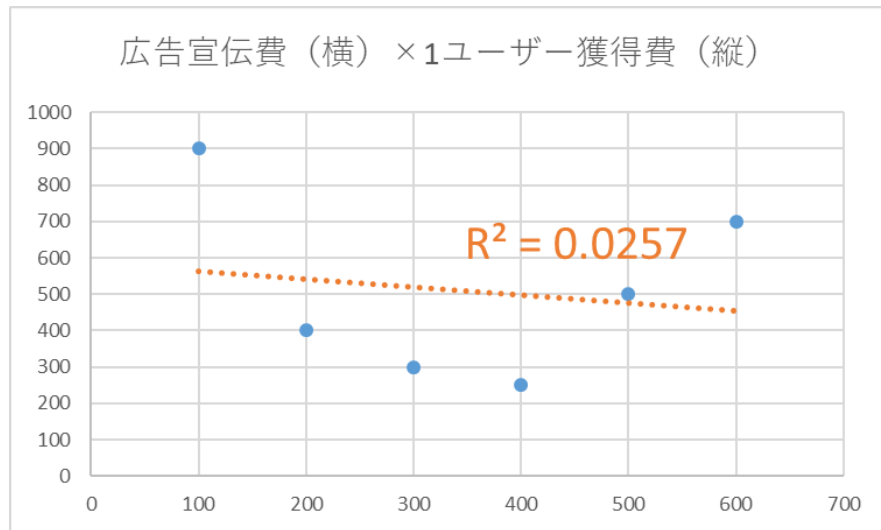
- 広告宣伝費をかけるほど・・・1ユーザーあたり獲得費用は？



# 近似曲線の応用

## (1) 広告宣伝費と、1ユーザーあたり獲得費用の関係

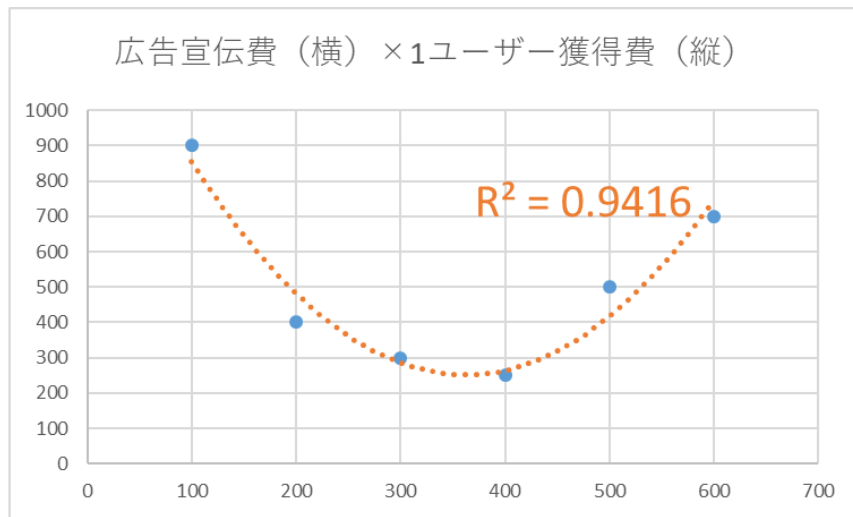
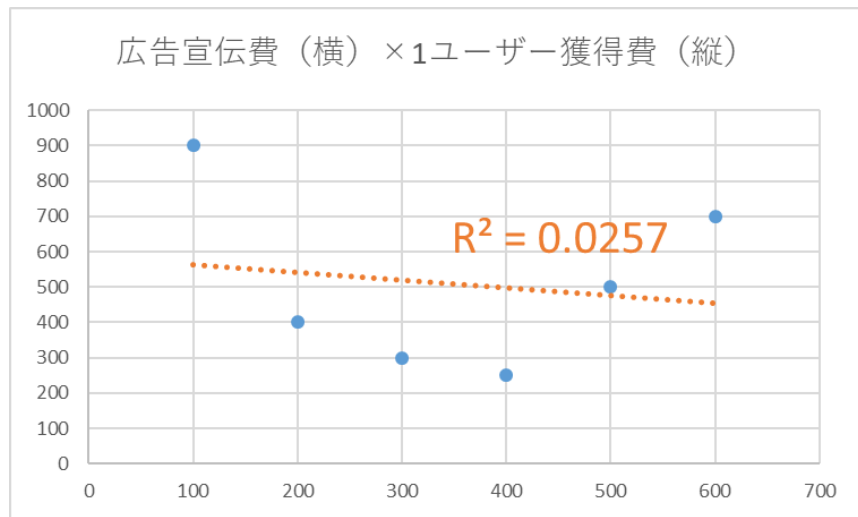
- あまり関係なさそう . . .



# 近似曲線の応用

## (1) 広告宣伝費と、1ユーザーあたり獲得費用の関係

- 広告費をかけると最初は獲得費は下がるが、
- 広告費をかけすぎると逆に上がってしまう



# 近似曲線の応用

## (1) 今回のマーケティングの仮説

A) まったく費用をかけないと、効率的なマーケティングができない

B) 費用をかけると、効率的に獲得できる

- 40代女性をターゲットにすると効率的
- 関西地方をターゲットにすると効率的

C) 一方、費用をかけすぎると、むしろ非効率になる

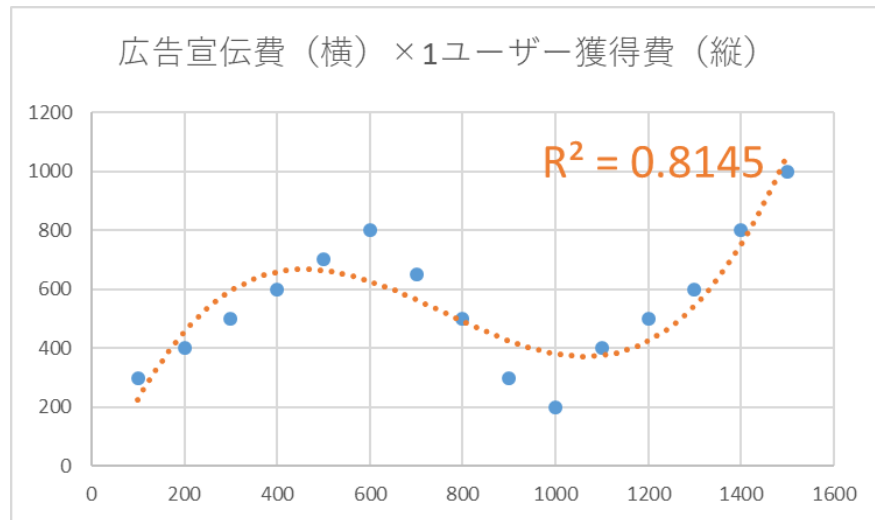
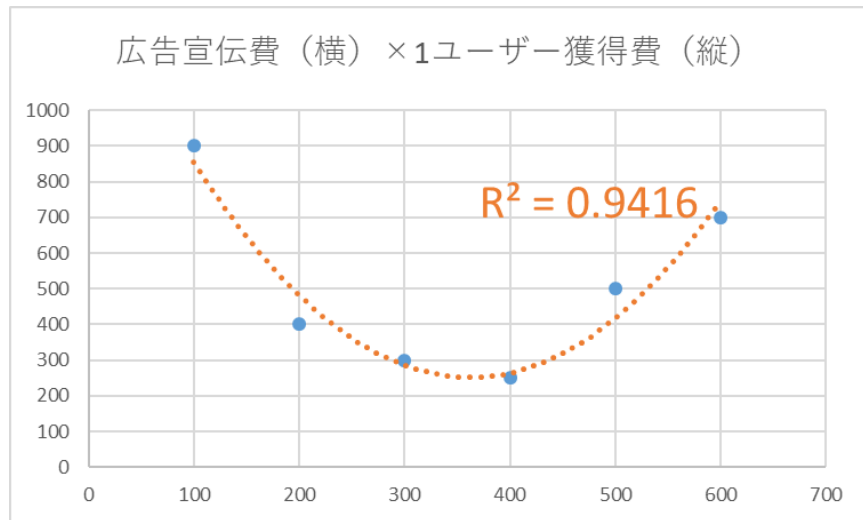
- ターゲットではない20代男性にコストをかけてしまう
- 東北地方までターゲットを広げてしまった

# 近似曲線の応用

## (1) 多項式近似

A) 〇次関数

B) 左図は2次関数、右図は3次関数



# 目次

---

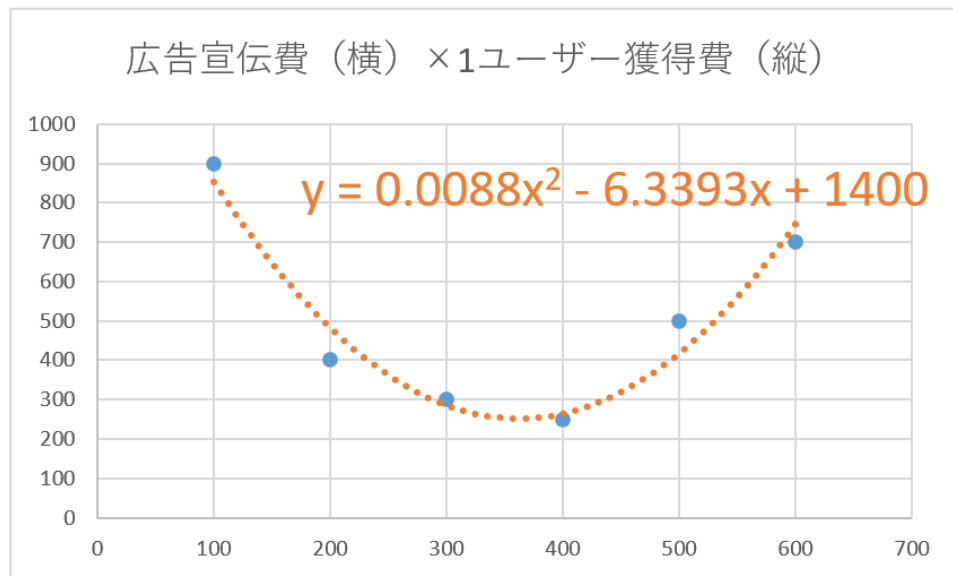
- 10. 近似曲線の応用(1)：指数近似
- 11. 近似曲線の応用(2)：対数近似
- 12. 近似曲線の応用(3)：累乗近似
- 13. 近似曲線の応用(4)：多項式近似
- 14. 最適解(1) 効率的なマーケティング
- 15. 最適解(2) マーケティング予算の分配



# マーケティングの最適化

## (1) 広告宣伝費と、1ユーザーあたり獲得費用の関係

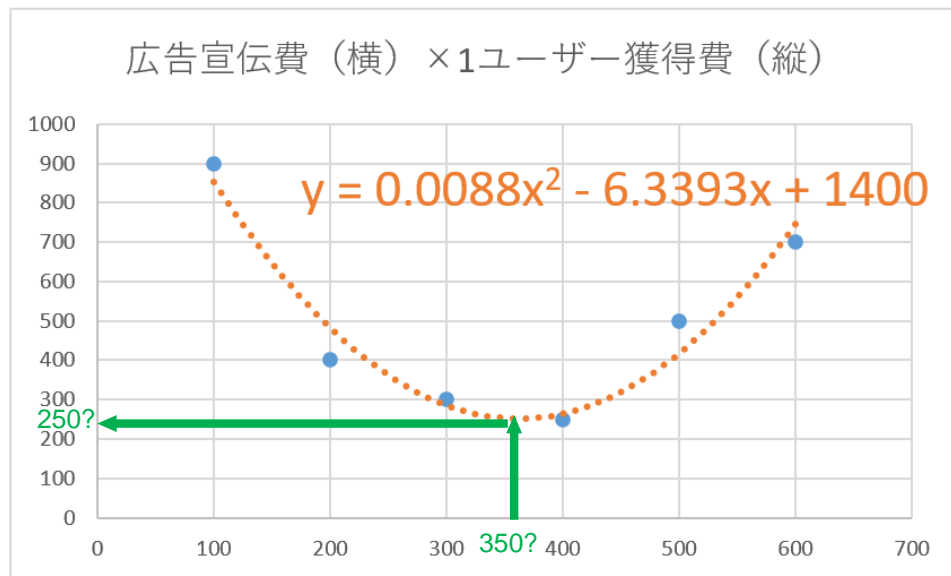
- 広告費をかけると最初は獲得費は下がるが、
- 広告費をかけすぎると逆に上がってしまう



# マーケティングの最適化

## (1) 広告宣伝費と、1ユーザーあたり獲得費用の関係

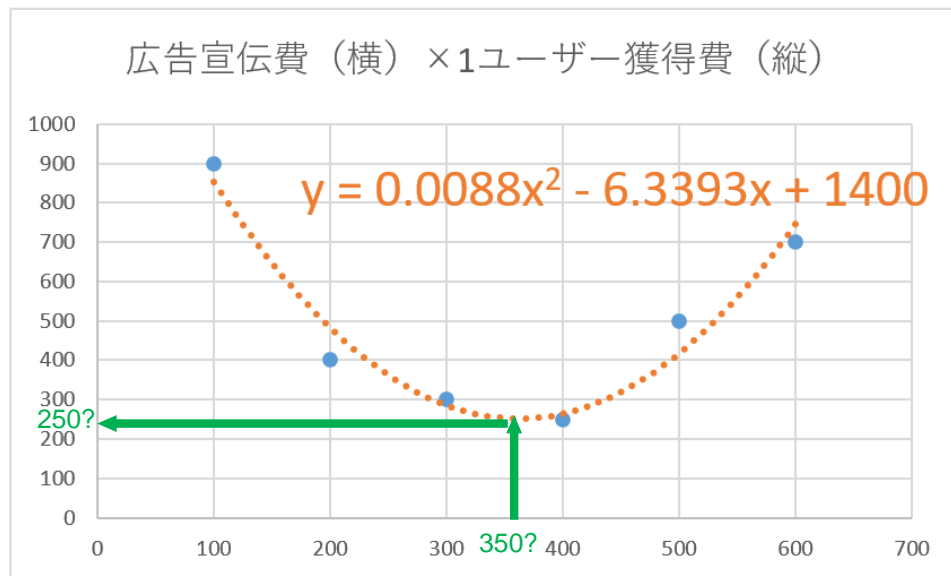
- いちばん獲得費用が安くなるのは、広告宣伝費350くらい？



# マーケティングの最適化

## (1) もっとも獲得費が安いポイントを調べる

- $y = 0.0088x^2 - 6.3393x + 1400$  から、 $y$  が一番小さくなる  $x$  を計算 (数学)
- ソルバーを使うと、簡単に計算できます



# マーケティングの最適化

---

## (1) ソルバー

- 計算式を基に、もっとも最適な数値を求める
- $y$  を最小にするためには  $x$  の数字は？

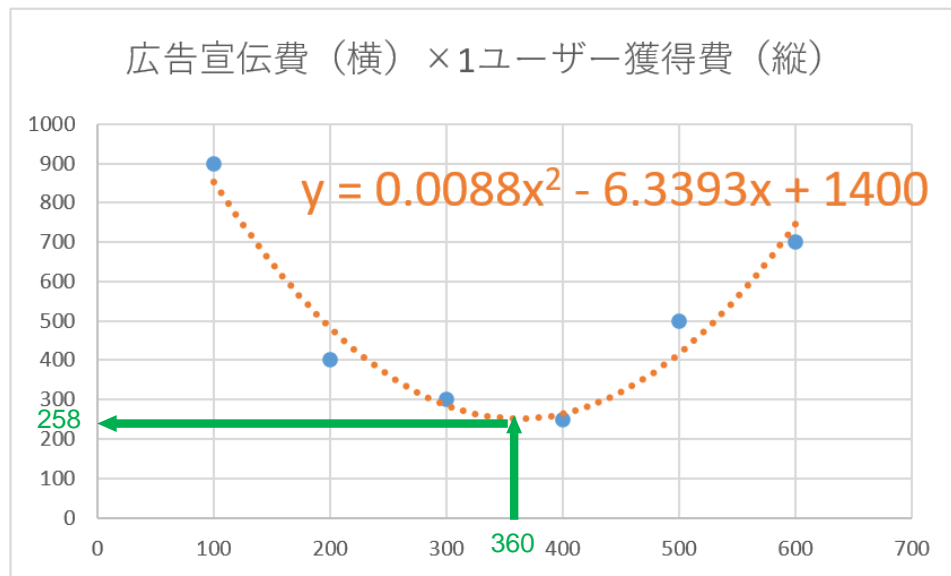
## (2) Excelの使用方法

- ソルバーの設定 [オプション] → [アドイン] → [Excelアドイン]
- [データ] → [ソルバー]

# マーケティングの最適化

## (1) もっとも獲得費が安いポイントを調べる

- $y = 0.0088x^2 - 6.3393x + 1400$  から、 $y$  が一番小さくなる  $x$  を計算 (数学)
- 広告宣伝費360のときに、1ユーザー獲得費は258と最小になる



# 目次

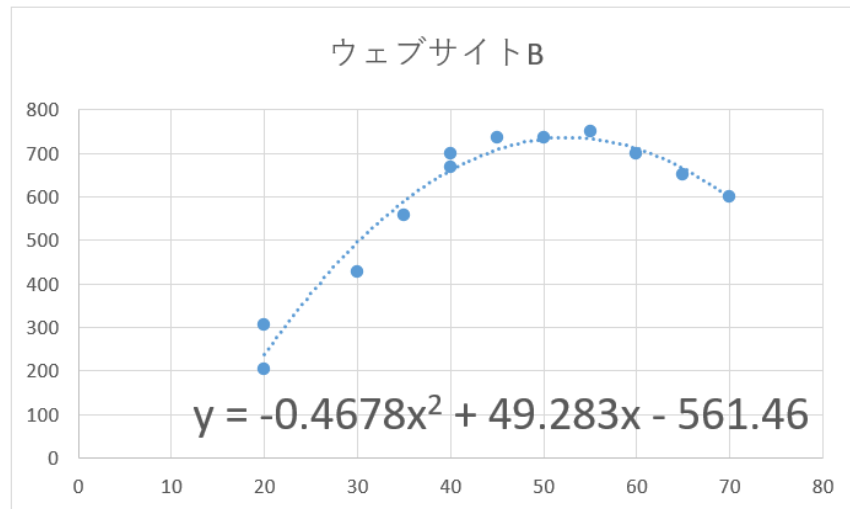
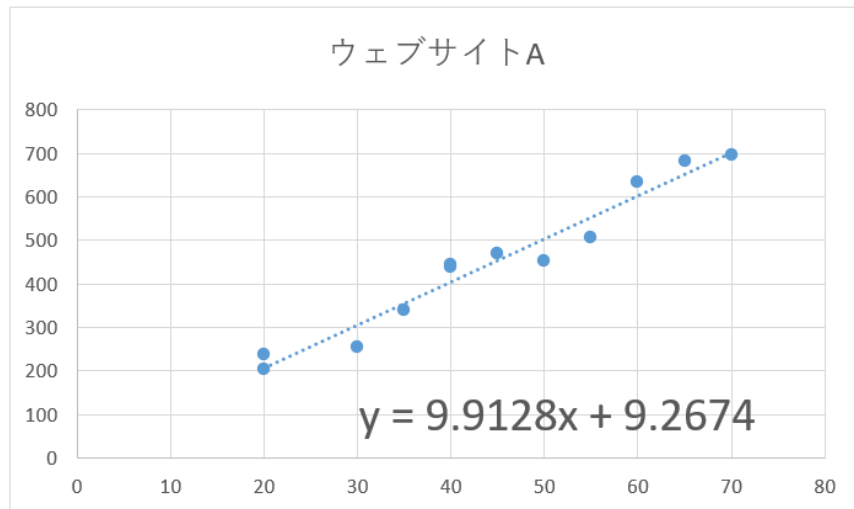
---

- 10. 近似曲線の応用(1)：指数近似
- 11. 近似曲線の応用(2)：対数近似
- 12. 近似曲線の応用(3)：累乗近似
- 13. 近似曲線の応用(4)：多項式近似
- 14. 最適解(1) 効率的なマーケティング
- 15. 最適解(2) マーケティング予算の分配

# マーケティングの最適化

(1) 販売数を増やすために、2つのウェブサイトA, Bに広告を出そうと考えている

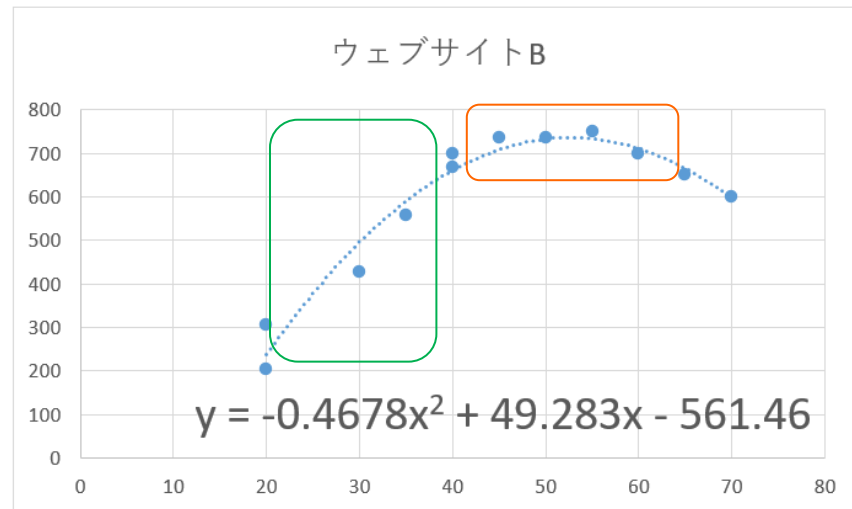
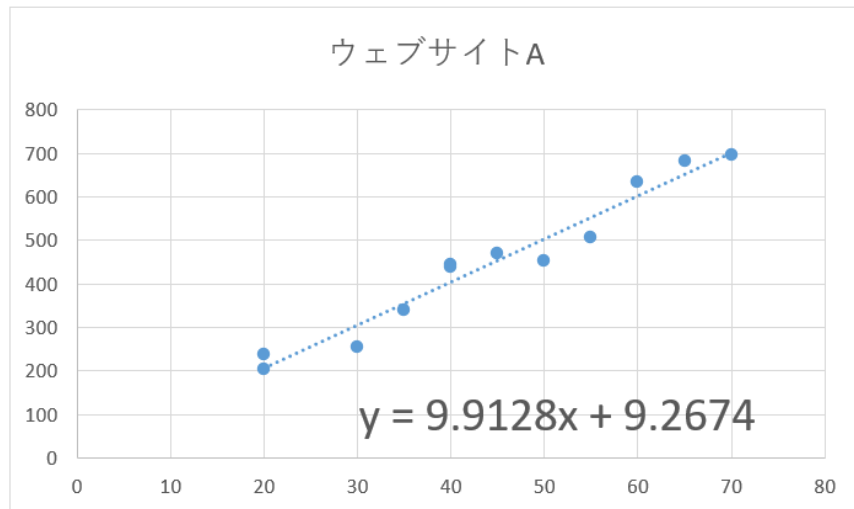
- 広告宣伝費の予算100を、AとBにいくらずつ振り分けるべきか
- 過去の広告宣伝費と、販売数の関係は以下の通り



# マーケティングの最適化

## (1) ウェブサイトBの特徴

- 広告宣伝費をかけ始めたときは、販売数はAよりも大きく伸びる
- ところが、広告宣伝費をかけすぎると、販売数が減少
- おそらく、40くらいまでBに投資して、残り60をAに投資するのが正解？





# マーケティングの最適化

---

## (1) 目的セル

- 販売数の合計

## (2) 目標値

- 最大化させたい

## (3) 変数セル

- サイトAの予算、サイトBの予算（Ctrlキーを押しながらクリック）

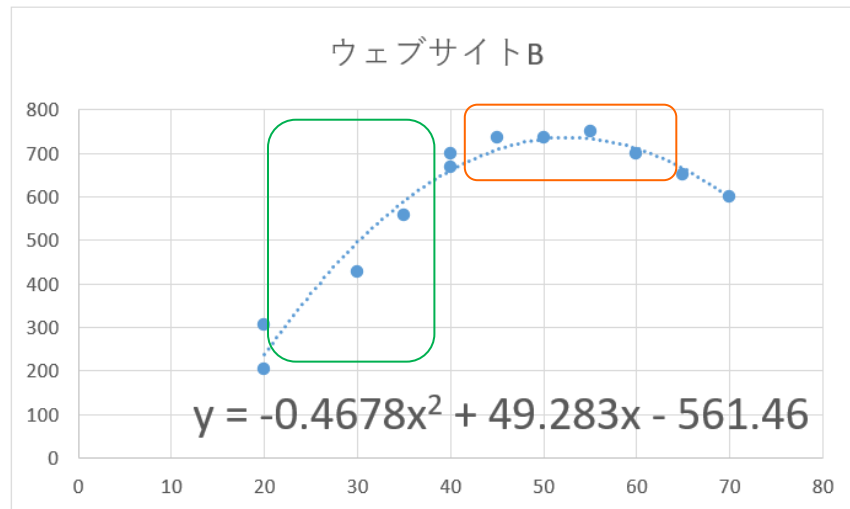
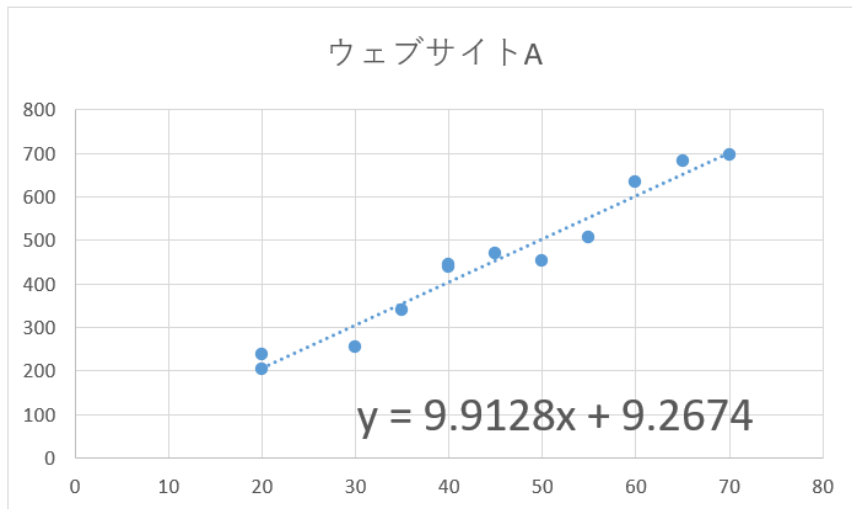
## (4) 制約条件

- 予算の合計 = 100

# マーケティングの最適化

## (1) ウェブサイトBの特徴

- 広告宣伝費をかけ始めたときは、販売数はAよりも大きく伸びる
- ところが、広告宣伝費をかけすぎると、販売数が減少
- おそらく、42くらいまでBに投資して、残り58をAに投資するのが正解？



# 目次

---

1. 数字の信頼性とは
2. 信頼区間(1)：平均値の「幅」
3. 信頼区間(2)：エクセルで計算
4. 信頼区間(3)：シミュレーション
5. P値(1)：テストの検証
6. P値(2)：エクセルで計算
7. P値(3)：シミュレーション
8. P値(4)：有意差の注意点

# 数字の信頼性

## (1) 目的

数字をより正確に理解し、思い込みをなくし、正しく伝える

## (2) 信頼区間

1日あたり平均の販売数は、100個

→ 95%の確率で、平均の販売数は96個～104個の間に収まります

## (3) P値

100人にアンケートとったら、新商品が欲しいという声が多数

→10,000人に聞かないと、統計的に意味があるとは言えない

# 目次

---

1. 数字の信頼性とは
2. 信頼区間(1)：平均値の「幅」
3. 信頼区間(2)：エクセルで計算
4. 信頼区間(3)：シミュレーション
5. P値(1)：テストの検証
6. P値(2)：エクセルで計算
7. P値(3)：シミュレーション
8. P値(4)：有意差の注意点

# 平均値の「幅」

---

## (1) 平均値の幅を考える

A) 同じ平均値でも幅が変わる

- データのバラつき
- データ数

## (2) 幅が分かることのメリット

A) 将来の売上の予測

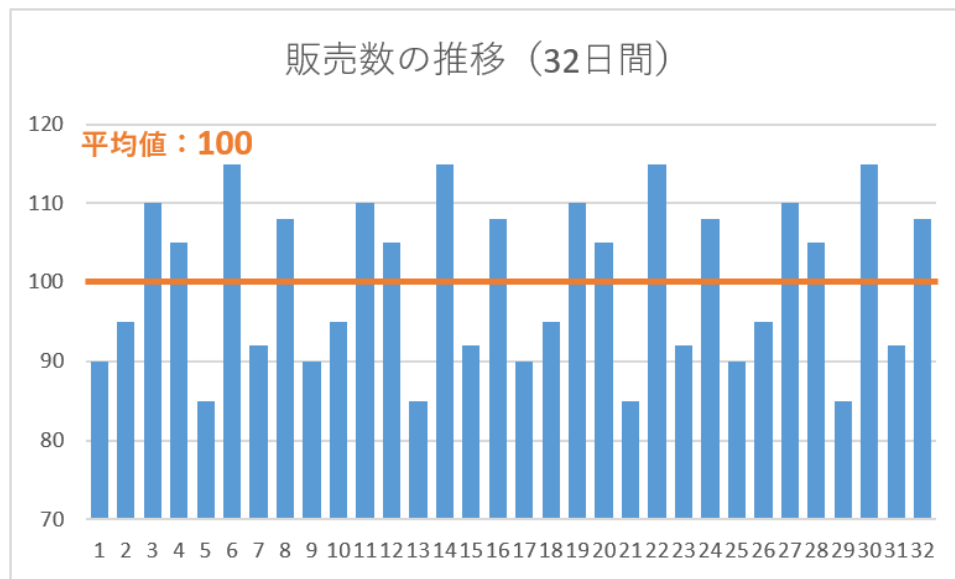
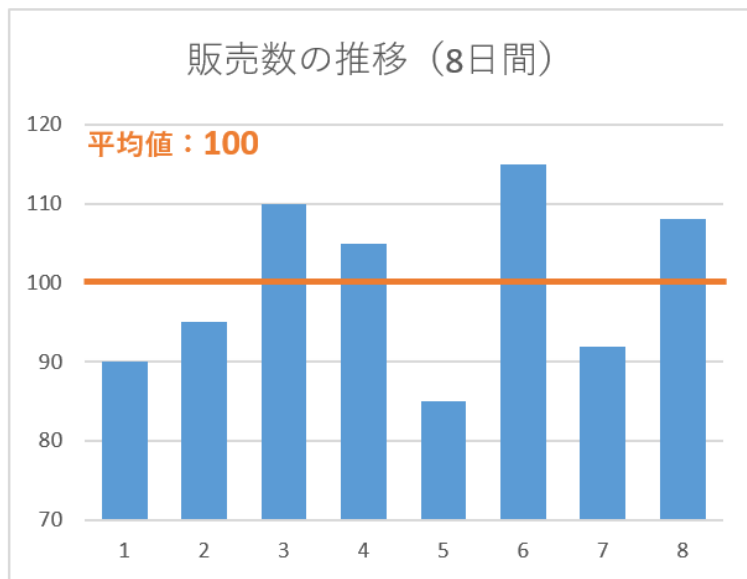
B) 必要な材料の調達、社員の採用

# 平均値の「幅」

## (1) 平均値の幅を考える

A) 過去8日間のデータと、過去32日間のデータ（4倍）

B) どちらも平均値は100

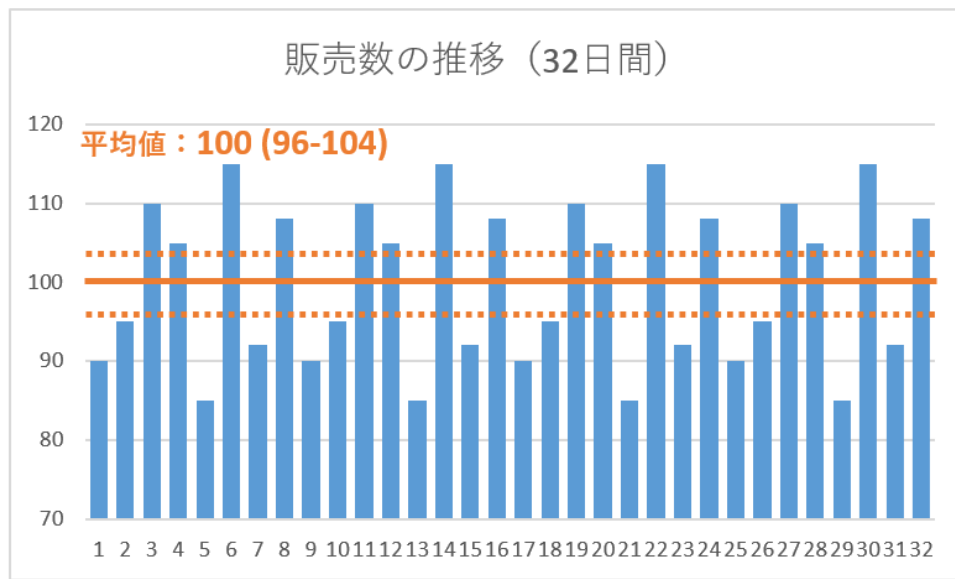
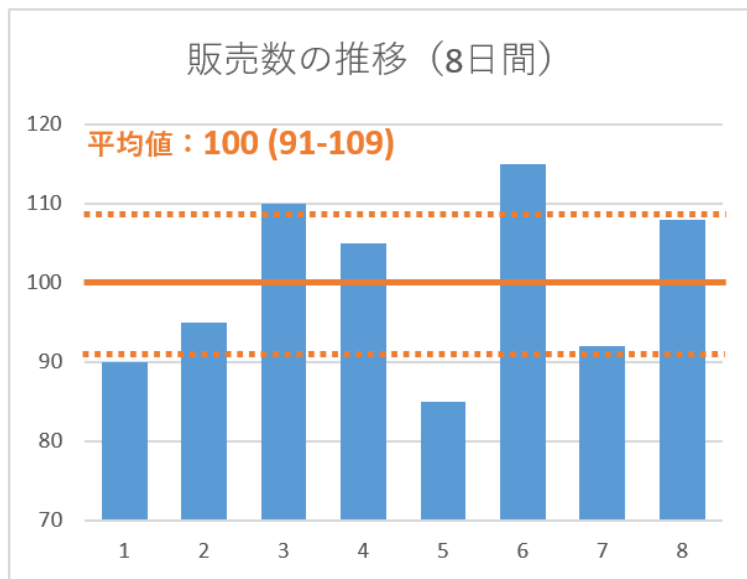


# 平均値の「幅」

## (1) 平均値の幅を考える

A) 過去8日間のデータと、過去32日間のデータ（4倍）

B) データが多いほど、平均値の幅はせまくなる（精度が高くなる）





# 目次

---

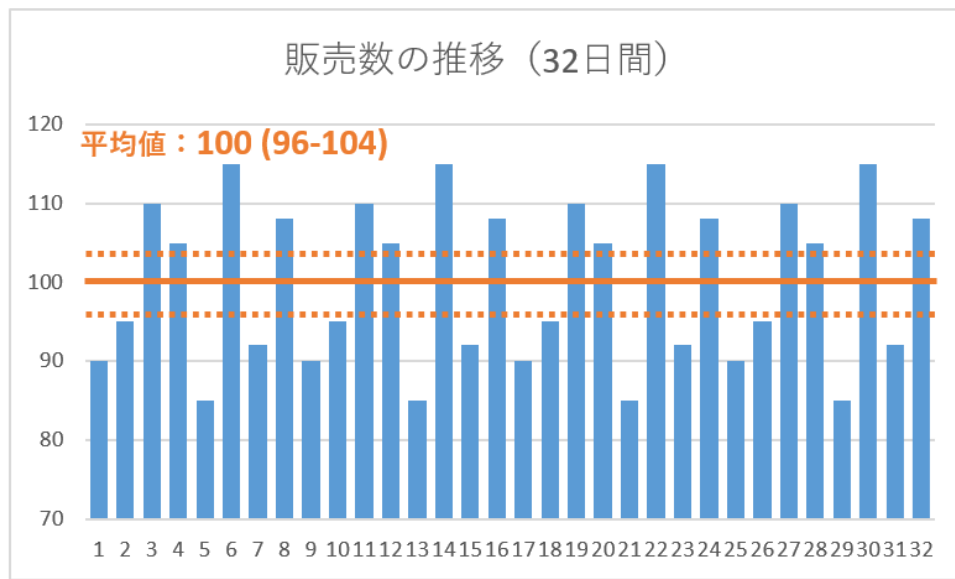
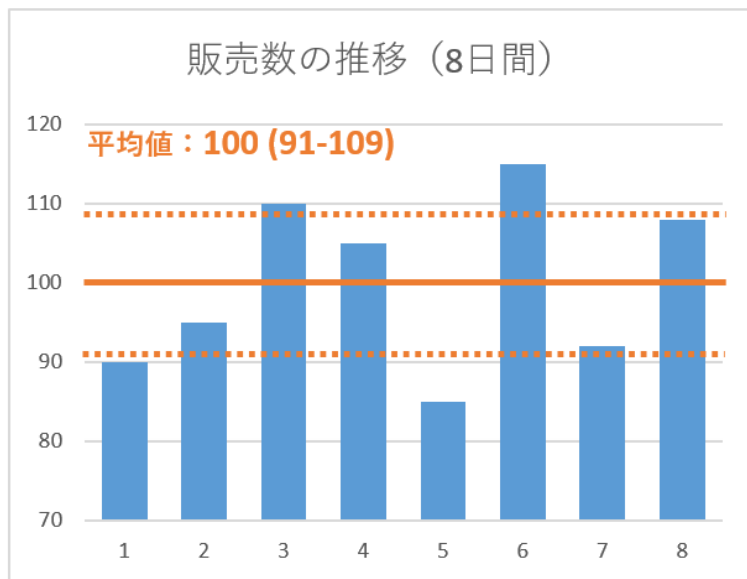
1. 数字の信頼性とは
2. 信頼区間(1)：平均値の「幅」
3. 信頼区間(2)：エクセルで計算
4. 信頼区間(3)：シミュレーション
5. P値(1)：テストの検証
6. P値(2)：エクセルで計算
7. P値(3)：シミュレーション
8. P値(4)：有意差の注意点

# 平均値の「幅」

## (1) 平均値の幅を考える

A) 過去8日間のデータと、過去32日間のデータ（4倍）

B) データが多いほど、平均値の幅はせまくなる（精度が高くなる）



# 信頼区間

## (1) 平均値の幅をどのように計算するか

- A) 〇〇%の確率で、平均値はこの範囲（幅）に収まります、という考え
- B) 信頼区間

## (2) 信頼区間

- A) 95%を使うことが多い
- B) 「ほとんどの場合、平均値は、この区間の範囲に収まる」
- C) 95%信頼区間の場合、平均値は91～109の間です
  - 平均値100、信頼区間 $\pm 9$

# エクセルで信頼区間＝「平均値の幅」を計算

## (1) 信頼区間の計算は、以下の要素で決まる

- A) データの数 (COUNT関数) ※標本数
- B) 平均値 (AVERAGE関数)
- C) 標準偏差 (STDEV.S関数)
- D) 有意水準 (1 - 信頼区間%)
  - 95%信頼区間の場合は、有意水準5%

## (2) 信頼区間の計算

- A) CONFIDENCE.T (有意水準、標準偏差、データの数)

# 目次

---

1. 数字の信頼性とは
2. 信頼区間(1)：平均値の「幅」
3. 信頼区間(2)：エクセルで計算
4. 信頼区間(3)：シミュレーション
5. P値(1)：テストの検証
6. P値(2)：エクセルで計算
7. P値(3)：シミュレーション
8. P値(4)：有意差の注意点

# エクセルで信頼区間シミュレーション

## (1) 信頼区間%を変えてみる

- A) 95%信頼区間 → 平均値は、91～109の間（ほぼ確実）
- B) 99%信頼区間 → 平均値は、87～113の間（絶対に確実！）
- C) 信頼区間を上げるほど、平均値の幅は広がる

## (2) データの数を変えてみる

- A) 8個 → 平均値は、91～109の間
- B) 32個 → 平均値は、96～104の間
- C) データの数が増えるほど、平均値の幅はせまくなる（精度が上がる）

# 目次

---

1. 数字の信頼性とは
2. 信頼区間(1)：平均値の「幅」
3. 信頼区間(2)：エクセルで計算
4. 信頼区間(3)：シミュレーション
5. P値(1)：テストの検証
6. P値(2)：エクセルで計算
7. P値(3)：シミュレーション
8. P値(4)：有意差の注意点

# テストは重要

---

## (1) 現在のマーケティングはテストを重視しています

- アンケート
- 試作品
- A/Bテスト
  - ウェブサイト上で、複数の広告をランダムに表示して反応を計測
  - 広告のクリック率、購入率に違いが見られるか

## (2) テストマーケティングにおける統計×エクセル

- 「そのテスト結果、正しいといえるのか？」を解説します



# テストマーケティング

---

## (1) 今回のマーケティングの仮説

- 新商品を開発中
- テストをして、新商品が現在の商品よりも売れそうかチェックしたい

## (2) テスト結果

- 100人に試してもらって「新商品と現商品どちらを買いたい？」
- 新商品：51人      現商品：49人

→ 果たして、新商品は現商品より売れそうといえるか？

# P値

---

## (1) P値とは

- 統計的に有意か、を示す指標
- 有意 = たまたま起こった可能性は低い（明確に差が発生している）
- $P\text{値} = \text{たまたま起こる確率}$
- このP値が低いほど、たまたま起こったとは言えない

## (2) P値の基準（例）

- 5%を下回ると、たまたま起こった可能性は低い（差は有意である）

# P値

---

## (1) たとえば

- サイコロを振ったら、なぜか10回連続で1が出た
- たまたま起きるだろうか？
  - おそらく起きない
  - 細工がしてあるサイコロの可能性が高い
- P値はすごく低い
  - 差は有意

# 目次

---

1. 数字の信頼性とは
2. 信頼区間(1)：平均値の「幅」
3. 信頼区間(2)：エクセルで計算
4. 信頼区間(3)：シミュレーション
5. P値(1)：テストの検証
6. P値(2)：エクセルで計算
7. P値(3)：シミュレーション
8. P値(4)：有意差の注意点

# P値

---

## (1) P値の計算方法

- カイ2乗検定という計算方法を使います
- 関数 = CHISQ.TEST (実測値、期待値)
- CHI : カイ
- SQ : 2乗 (スクエア)
- TEST : 検定

# P値

## (1) 実測値と期待値

- 実測値とは、実際に測定された結果
- 期待値とは、差がない場合の結果（理論的な）

## (2) 今回のケース

- 新商品と現商品のどちらを買いたいのか？

- 実測値

新商品：51人

現商品：49人

- 期待値

新商品：50人

現商品：50人

この違いがP値

# P値

---

## (1) P値の判断基準

- 一般的には、5%以下であれば差は有意（と言う場合が多い）

# 目次

---

1. 数字の信頼性とは
2. 信頼区間(1)：平均値の「幅」
3. 信頼区間(2)：エクセルで計算
4. 信頼区間(3)：シミュレーション
5. P値(1)：テストの検証
6. P値(2)：エクセルで計算
7. P値(3)：シミュレーション
8. P値(4)：有意差の注意点



# P値と、必要なテスト数

## (1) 51%が欲しいと答えたなら・・・

- 100 人中 51 人が欲しい → 有意差はない
- 10,000 人中 5,100 人が欲しい → 有意差はある

## (2) 54%が欲しいと答えたなら・・・

- 1,000人のテストで十分
- 10,000人は必要なし

## (3) ざっくりいうと

- 少しの差なら、多くのテスト数が必要
- 大きな差なら、少しのテスト数で十分

# 目次

---

1. 数字の信頼性とは
2. 信頼区間(1)：平均値の「幅」
3. 信頼区間(2)：エクセルで計算
4. 信頼区間(3)：シミュレーション
5. P値(1)：テストの検証
6. P値(2)：エクセルで計算
7. P値(3)：シミュレーション
8. P値(4)：有意差の注意点

# 有意差の注意点

---

## (1) テスト結果

- A) 新商品を買いたい人は、100人中51人
- B) 有意差はない

## (2) 誤解しないように注意

- A) 有意差がない = 新商品を買いたくない、と言ってるわけではない
  - 新商品の販売をあきらめるのは早い
  - まだ分からないだけなので、テストを続ける必要あり
- B) 新薬が効くか分からない  $\neq$  新薬が効かない