全体の流れ

1つのデータを読む

- 比較する
- 分解する
- 比率で見る
- ○○あたりで見る
- データの定義
- データの前提条件

大量のデータを読む

- 平均値、中央値
- 分布
- ヒストグラム
- 標準偏差
- 傾向
- 関係性

データを正しく読む

- グラフの注意点
- サンプルの注意点
- データの偏り
- 異常値、欠損値
- 確証バイアス
- フェルミ推定

今回のポイント

データにだまされず、正しく読む

データを正しく読む

1. データにだまされない「読み方」
    - データは、
        - 誤解を招くグラフや、
        - 元々のデータに問題があるケースも多い
    - **データを誤解しないための注意点について学ぶ**
2. データにだまされない「考え方」
    - 数字を考えるときの注意点についても学ぶ
    - 確証バイアス、フェルミ推定

データを正しく読む

1. グラフ(1) 分解レベル
2. グラフ(2) 縦軸
3. グラフ(3) 累計
4. サンプル(1) サンプル数
5. サンプル(2) データの偏り
6. 異常値
7. 欠損値
8. 確証バイアス
9. 単位のちがい
10. フェルミ推定
11. まとめ

今回のポイント

グラフを読む (1)

このデータを見て、どう答える？

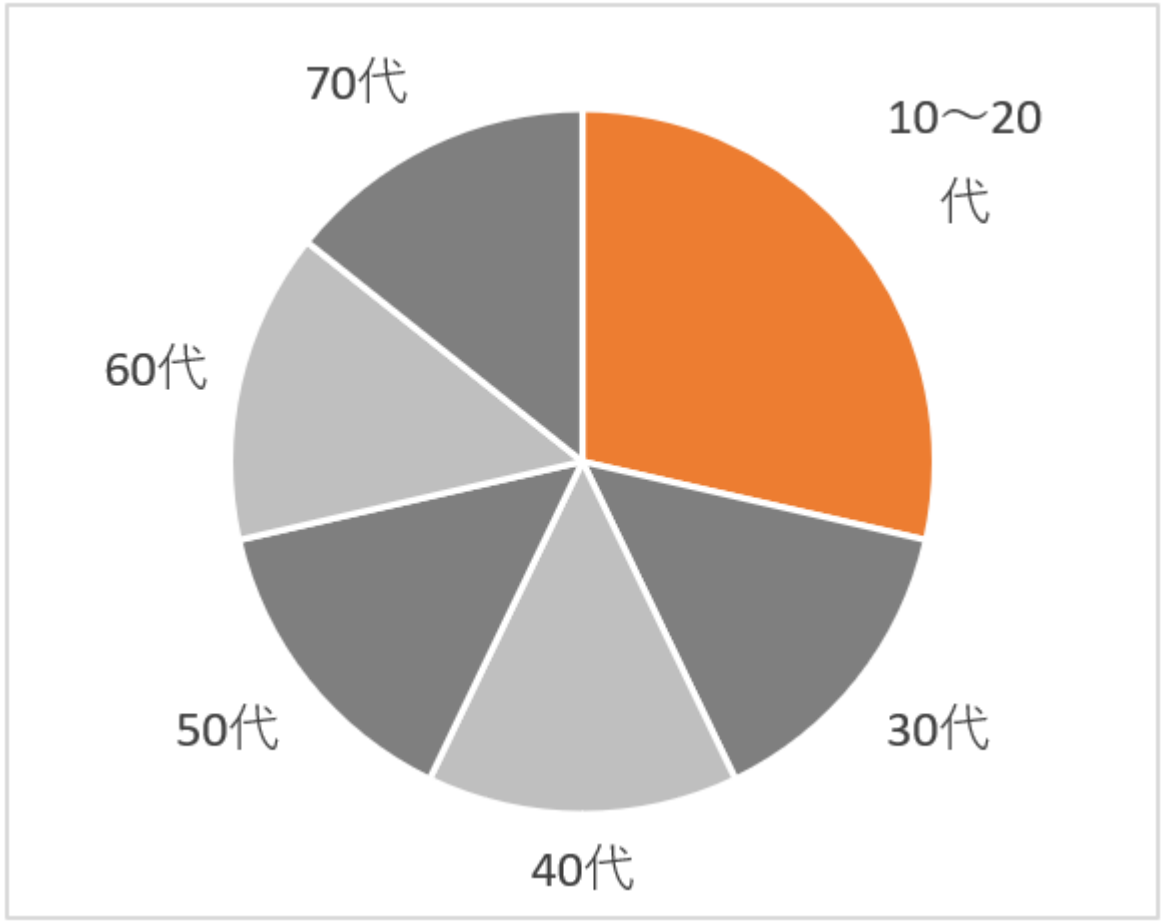
新商品の販売状況ですが、

10～20代の人たちに多く売れています！

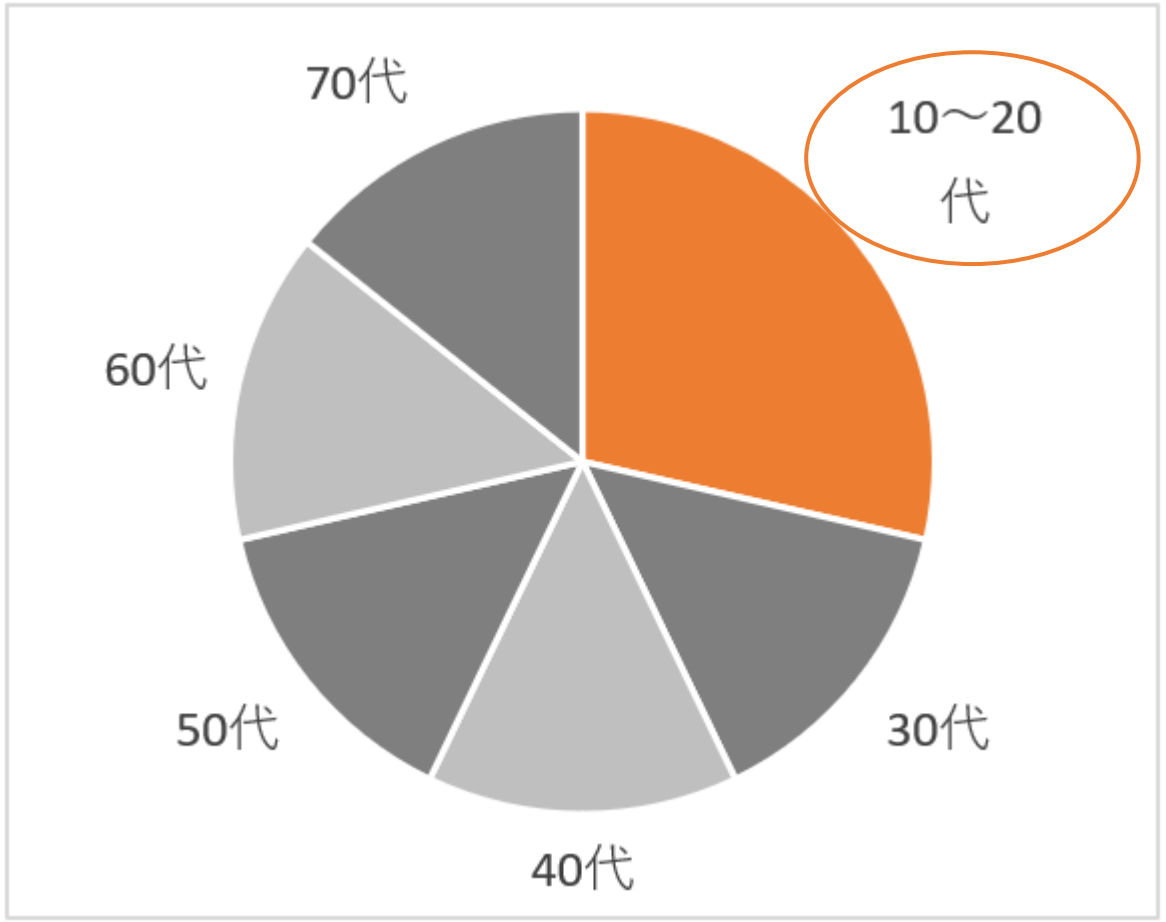
漫画：南 智恵

このデータを見て、どう答える？

新商品の販売状況ですが、10～20代の人たちに多く売れています！



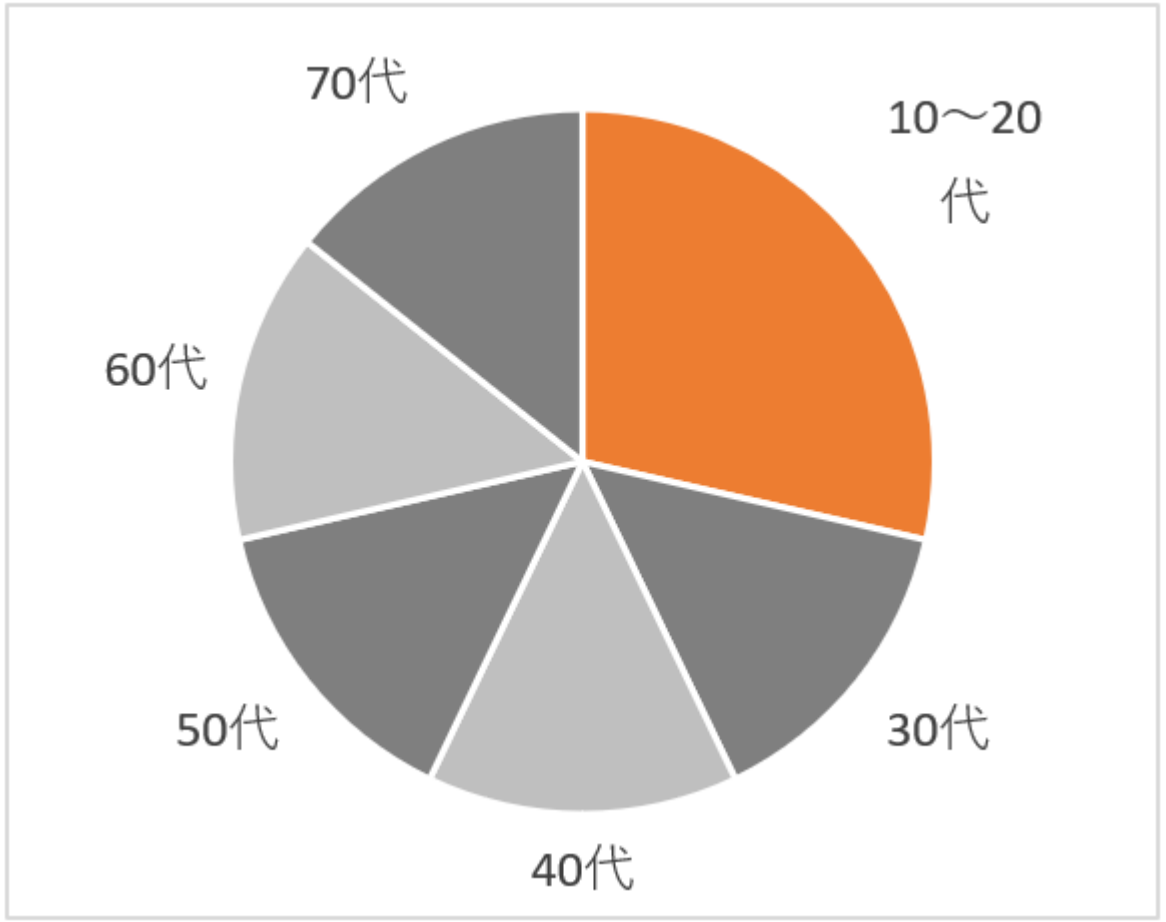
このデータを見て、どう答える？

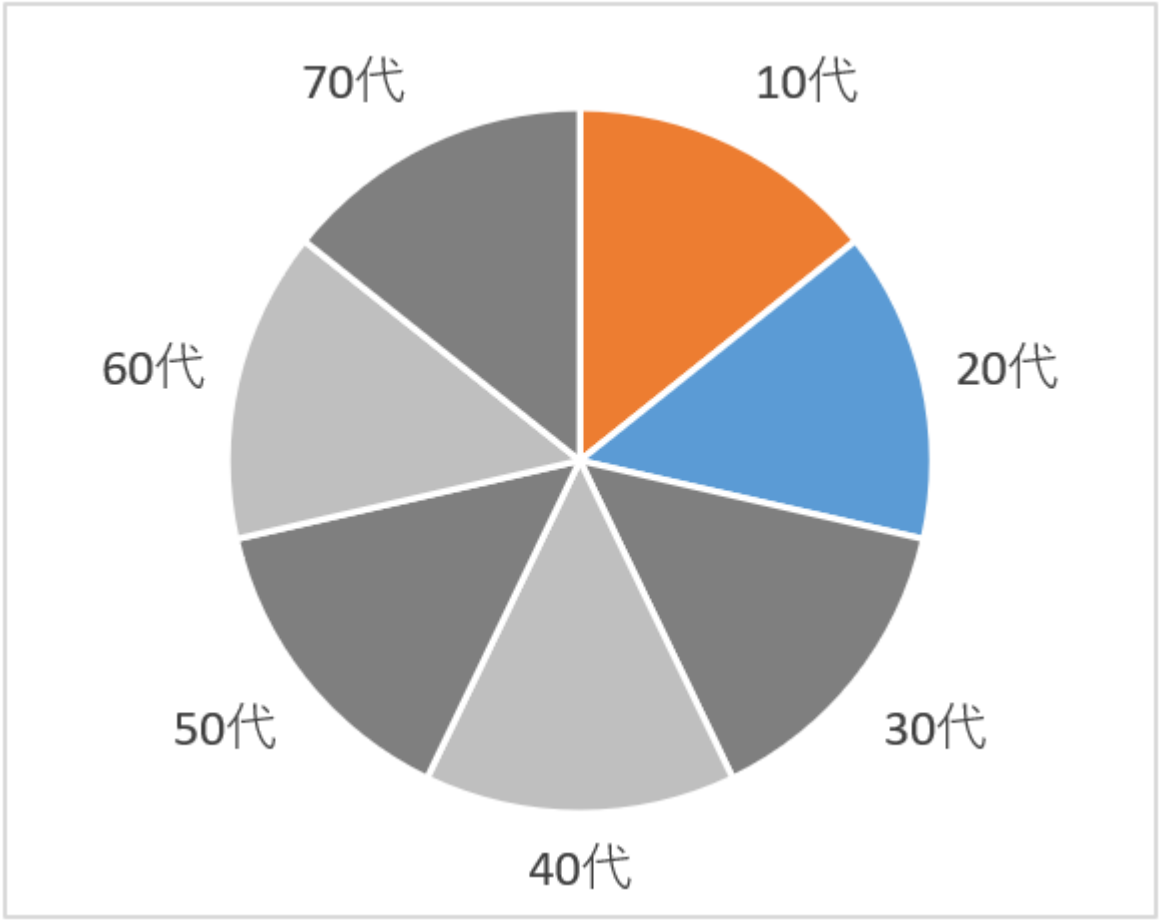


10代と20代の合計になっている

このデータを見て、どう答える？

10代と20代を別々にすると、実は他と変わらない





データを読むときのポイント

1. 比較する

- 比較するときは、**分解レベルを一緒にする**こと
- 10代、20代、30代、40代・・・

2. 注意点

- **人は、少しでも都合の良いようにデータを見せたくなる**
- そのために「10代＋20代」でまとめる、

  のは誤解を招く見せ方なので良くない

データを正しく読む

1. グラフ(1) 分解レベル
2. グラフ(2) 縦軸
3. グラフ(3) 累計
4. サンプル(1) サンプル数
5. サンプル(2) データの偏り
6. 異常値
7. 欠損値
8. 確証バイアス
9. 単位のちがい
10. フェルミ推定
11. まとめ

今回のポイント

グラフを読む (2)

このデータを見て、どう答える？

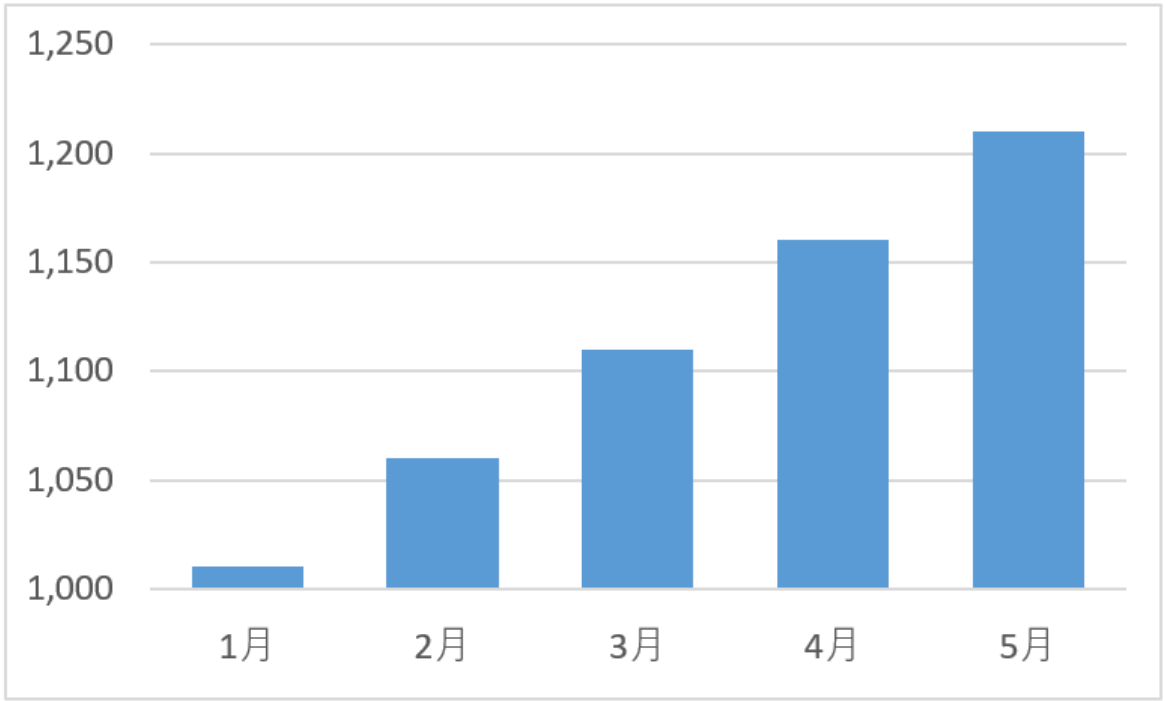
新商品の販売状況ですが、

1月～5月まですごく成長しています！

漫画：南 智恵

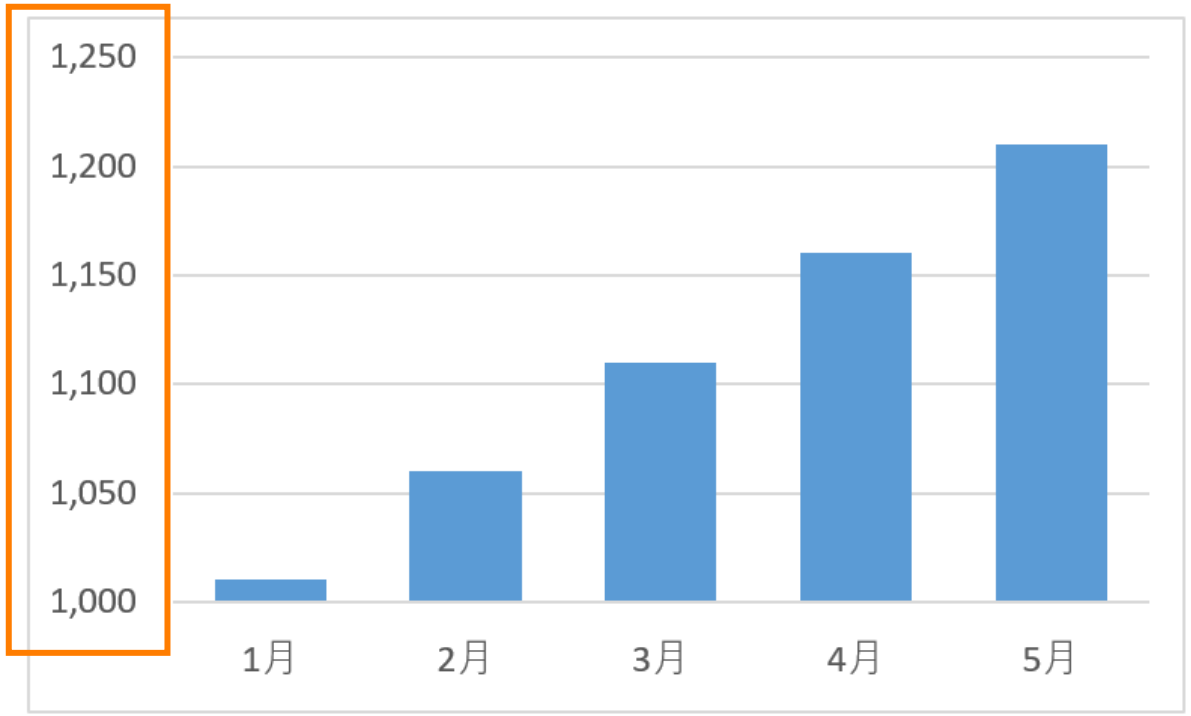
このデータを見て、どう答える？

新商品の販売状況ですが、1月～5月まですごく成長しています！



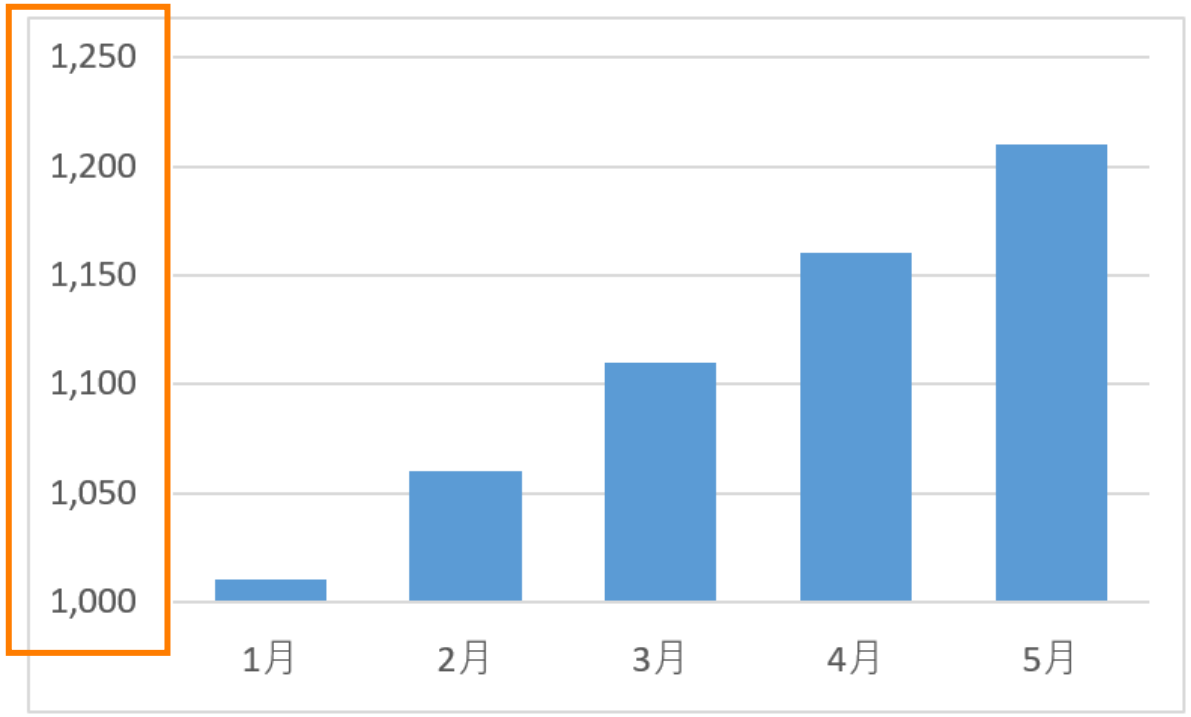
このデータを見て、どう答える？

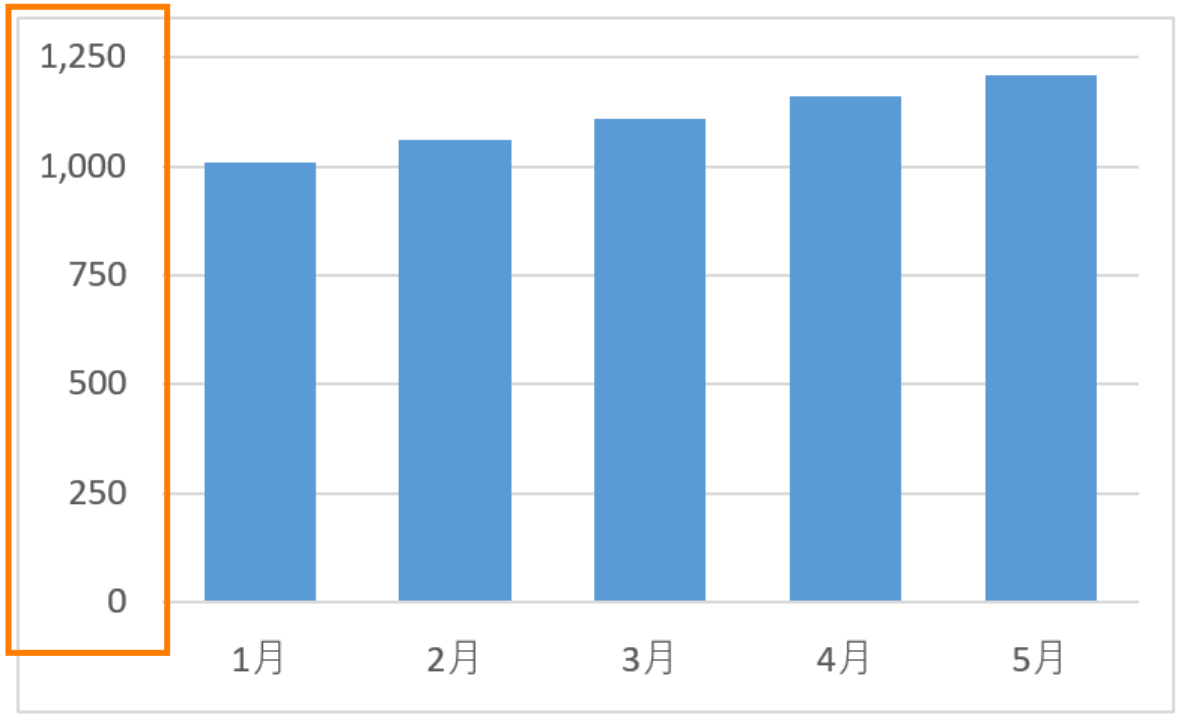
縦軸が1,000から始まっている



このデータを見て、どう答える？

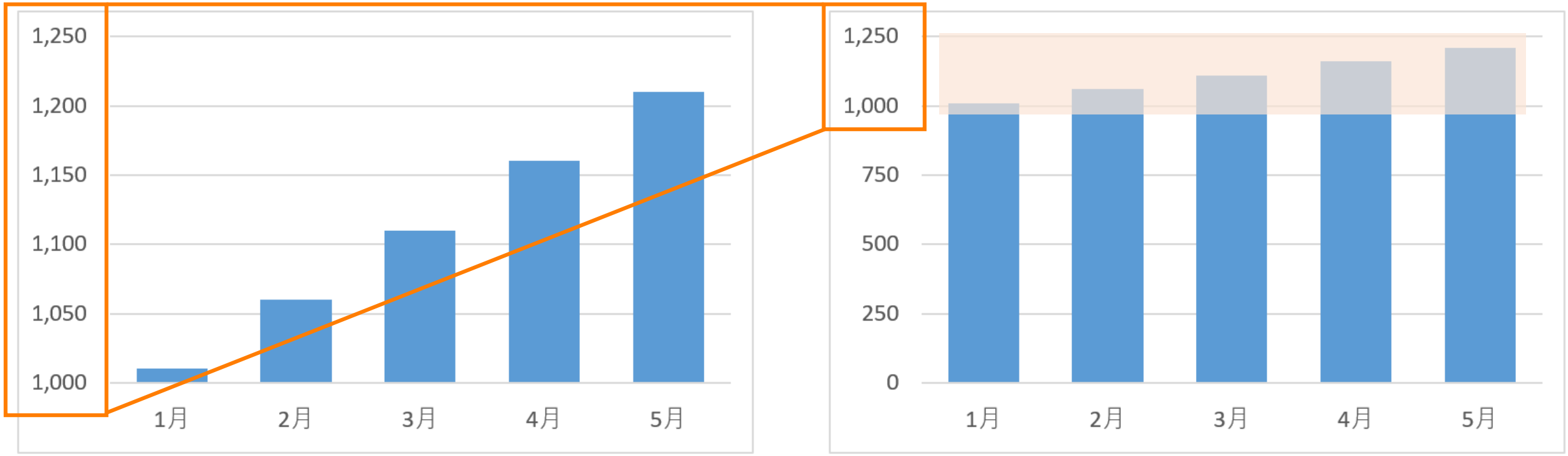
縦軸を0から始めると、実はそれほど成長していないように見える





このデータを見て、どう答える？

右図の1,000～1,250部分を、拡大したものが左図になる



データを読むときのポイント

1. グラフの読み方

- 縦軸に注意して見る
- **グラフは「少しの差を大きく見せる」ことができる**

2. 注意点

- **人は、少しでも都合の良いようにデータを見せたくなる**
- そのために縦軸を操作してデータの印象を変える、のは誤解を招く見せ方なので良くない

データを正しく読む

1. グラフ(1) 分解レベル
2. グラフ(2) 縦軸
3. グラフ(3) 累計
4. サンプル(1) サンプル数
5. サンプル(2) データの偏り
6. 異常値
7. 欠損値
8. 確証バイアス
9. 単位のちがい
10. フェルミ推定
11. まとめ

今回のポイント

グラフを読む (3)

このデータを見て、どう答える？

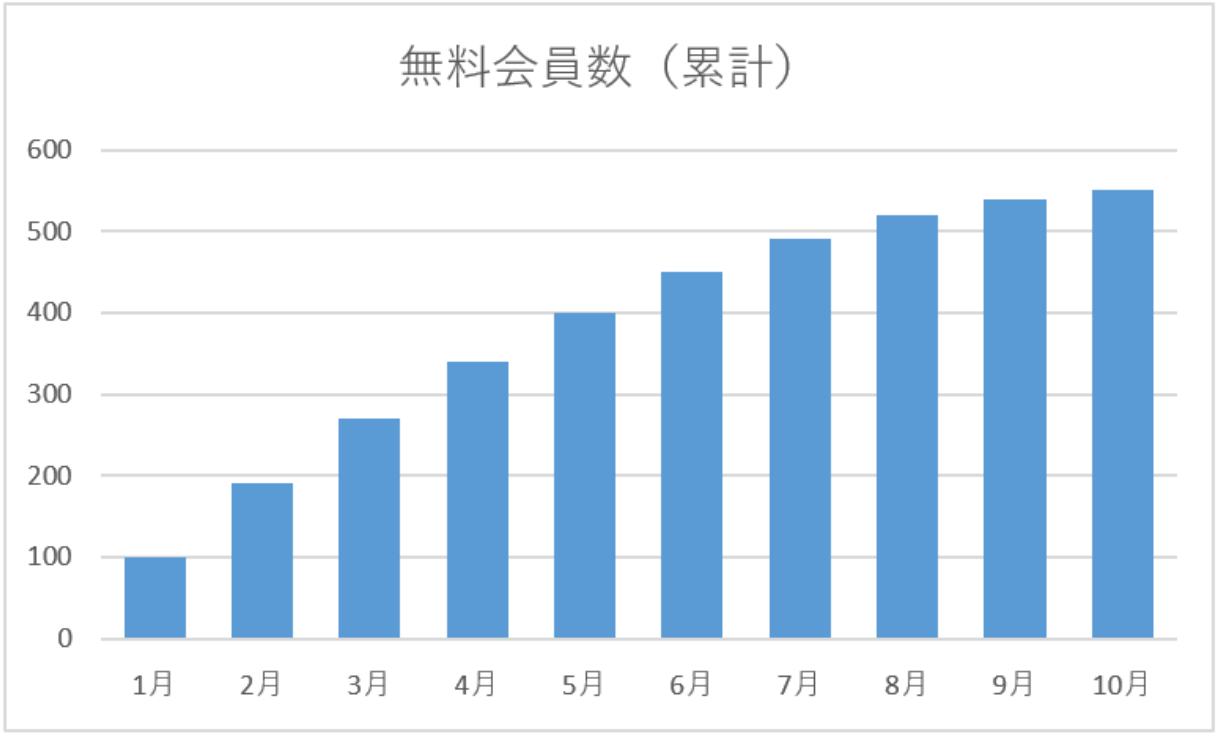
弊社のウェブサイトの無料会員数は、

ずっと過去最高が続いています！

漫画：南 智恵

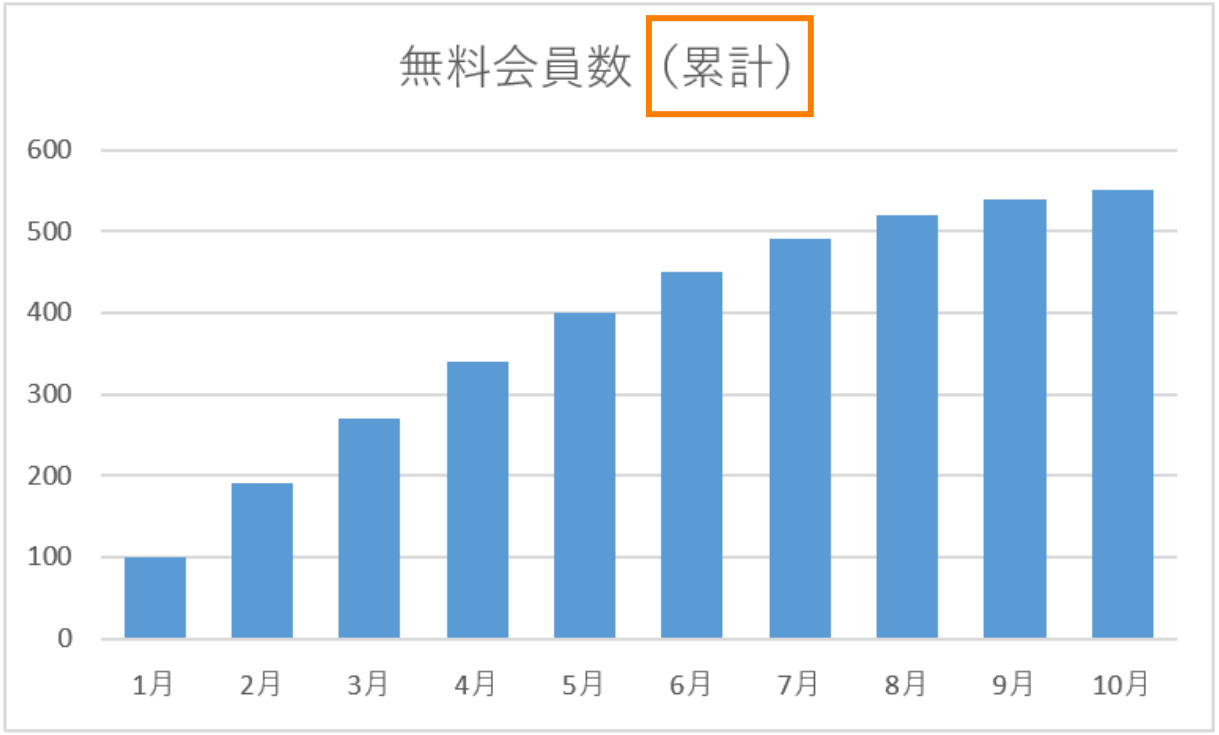
このデータを見て、どう答える？

「会員数は、過去最多を更新していて素晴らしい」



このデータを見て、どう答える？

会員数（累計）



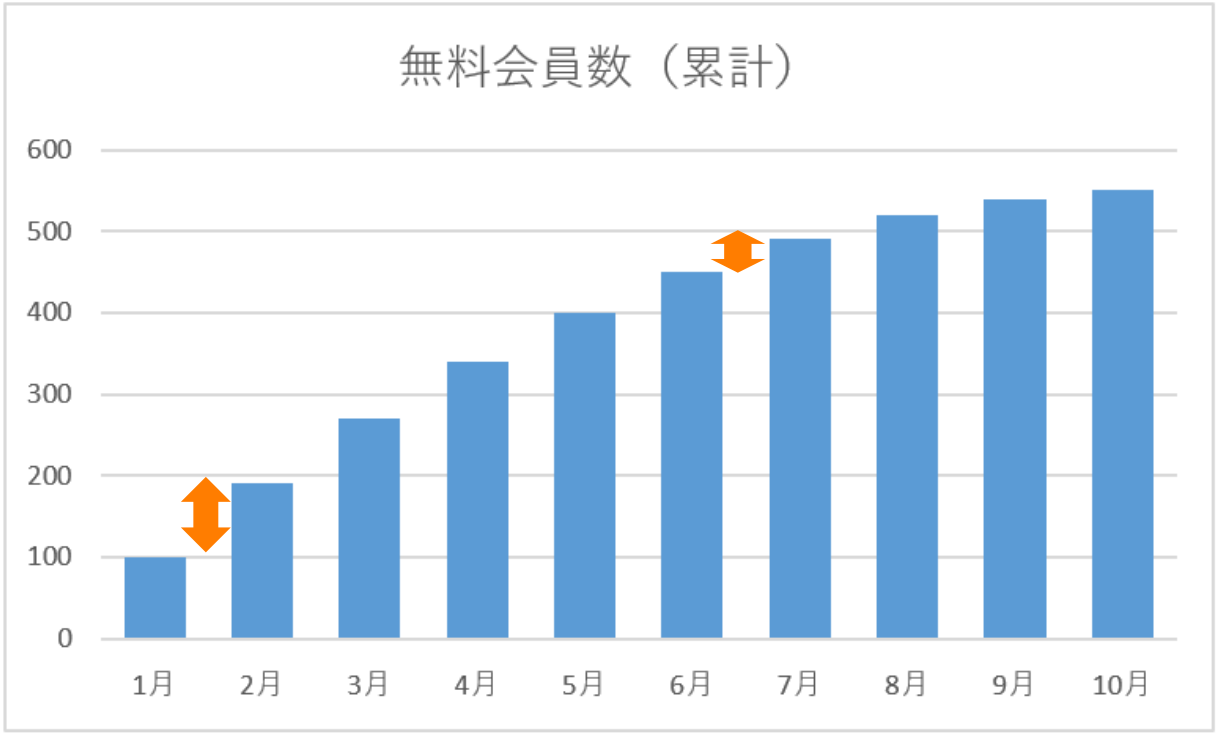
データを読むときのポイント

1. 累計

- 会員数は積みあがっていく
- 特に「無料会員登録」の場合、会員を解約する人は少ない
  （解約手続きも面倒）
- 会員数が増え続けるのは、当たり前ともいえる
- **では、「会員数は順調に増えているのか？」**
  ＝会員数の増え方のペースを見てみる

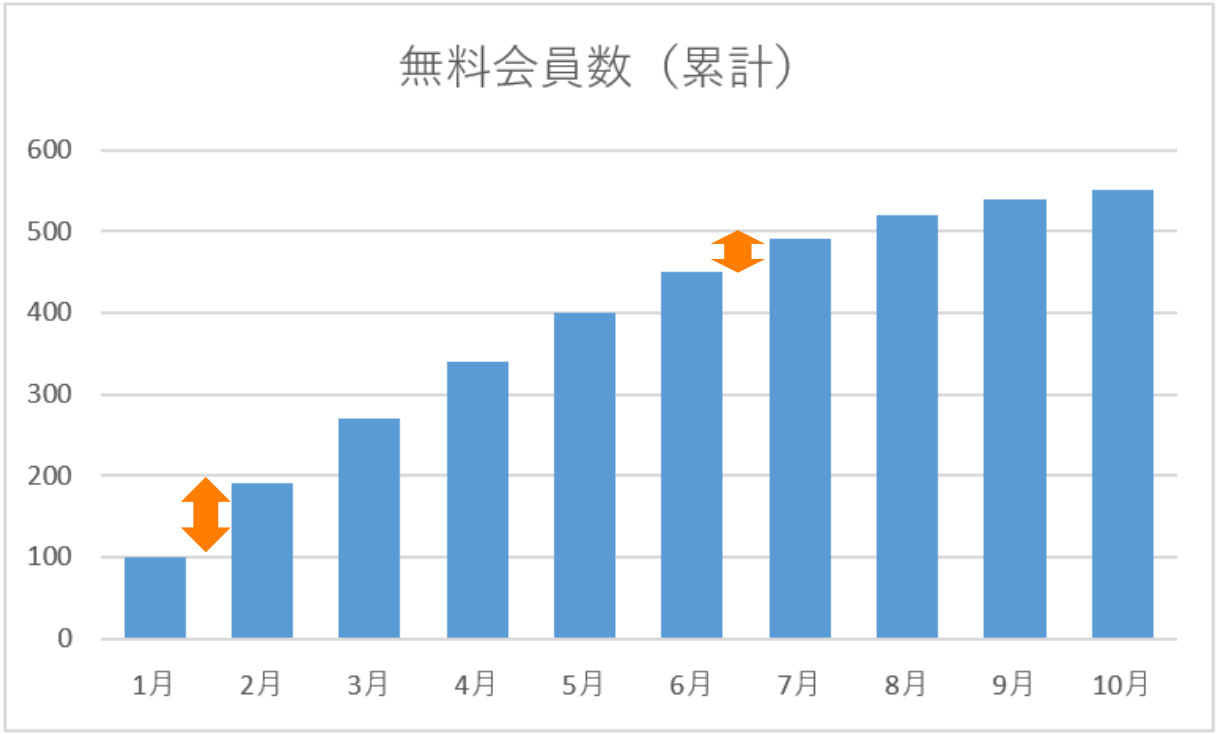
このデータを見て、どう答える？

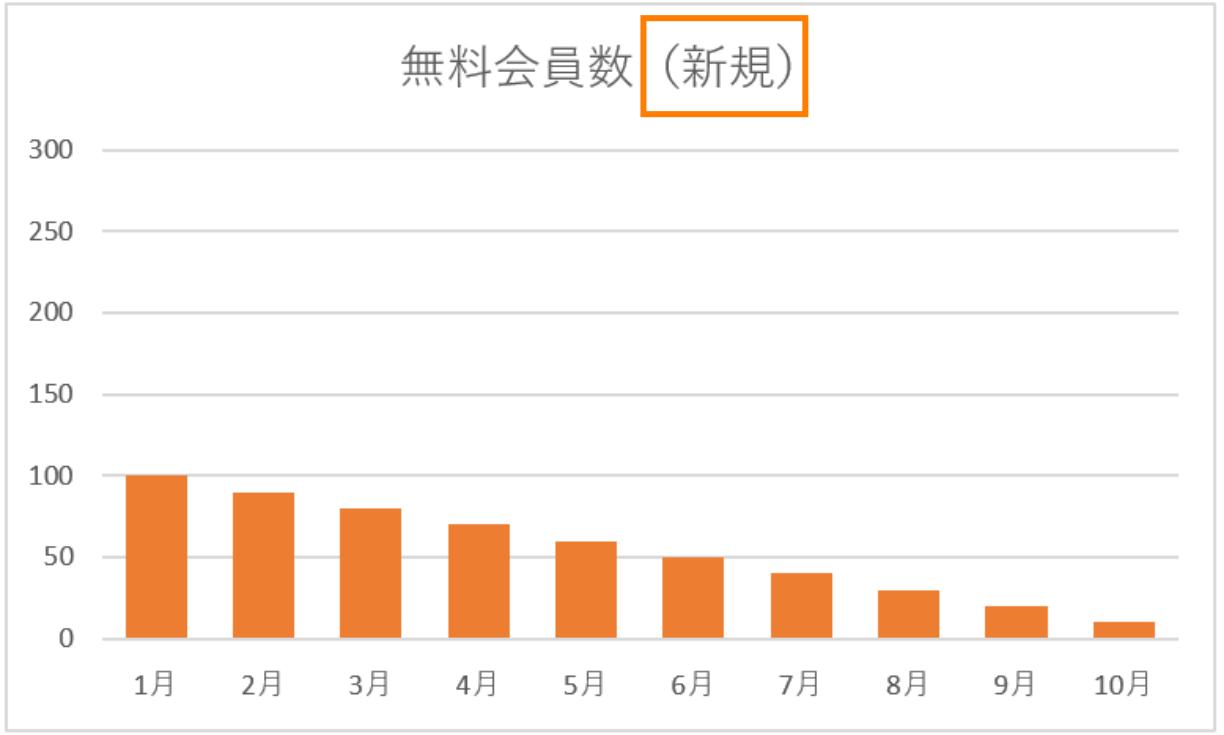
会員数の増加数（新規の会員登録）は、下がっている



このデータを見て、どう答える？

新規の会員数だけをグラフにすると分かりやすい





データを読むときのポイント

1. 累計
    - 「累計会員数が過去最多」
    - **もちろん素晴らしい数字ではあるものの、**
      その会員数の増え方のペースも見ておきたい
    - 実は成長期は過ぎてしまっているかもしれない
2. キーワード
    - サチる＝頭打ちになる
    - 数値があまり伸びずに止まってしまっていること

データを正しく読む

1. グラフ(1) 分解レベル
2. グラフ(2) 縦軸
3. グラフ(3) 累計
4. サンプル(1) サンプル数
5. サンプル(2) データの偏り
6. 異常値
7. 欠損値
8. 確証バイアス
9. 単位のちがい
10. フェルミ推定
11. まとめ

今回のポイント

サンプル数

このデータを見て、どう答える？

アンケート結果によると、

新商品を買いたいと答えた人が、

なんと60%もいました！

漫画：南 智恵

このデータを見て、どう答える？

60%の人が「新商品を買いたい」と回答

アンケート回答数	5人
新商品を買いたいと答えた人	3人
	60%

このデータを見て、どう答える？

60%の人が「新商品を買いたい」と回答

→ **回答数は、わずか5人**

→ **買いたいと回答した人が1人減るだけで、2人÷5人＝40%に低下**

→ **60%という比率は、素直に喜べる数値ではなさそう**

アンケート回答数	5人
新商品を買いたいと答えた人	3人
	60%

サンプル数

1. ポイント

- 同じ60%といっても、
  - 5人中3人が「買いたい」と回答
  - 50,000人中30,000人が「買いたい」と回答

    → 数値の信頼度が違う
  - 統計学を学ぶと、分析に必要なサンプル数を計算できる

2. キーワード

- サンプル数＝N数と呼ぶこともある

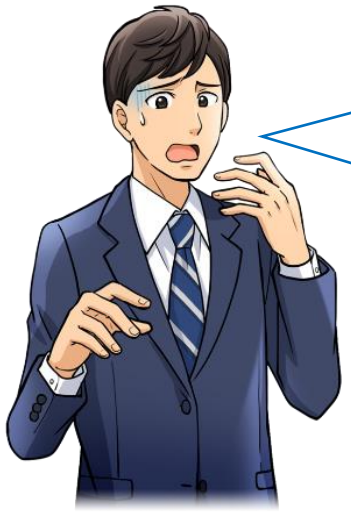
データを正しく読む

1. グラフ(1) 分解レベル
2. グラフ(2) 縦軸
3. グラフ(3) 累計
4. サンプル(1) サンプル数
5. サンプル(2) データの偏り
6. 異常値
7. 欠損値
8. 確証バイアス
9. 単位のちがい
10. フェルミ推定
11. まとめ

今回のポイント

サンプルの偏り

このデータを見て、どう答える？



アンケート結果データが大量なので、

先頭からサンプルをとって計算したら、

全員「新商品のビールを買わない」と言ってます

漫画：南 智恵

このデータを見て、どう答える？

先頭のアンケート回答を見ると、全員「買わない」と回答

回答者	年齢	回答
0001	85	買わない
0002	84	買わない
0003	84	買わない
0004	83	買わない
0005	82	買わない
0006	82	買わない
0007	81	買わない
0008	80	買わない
0125	26	買う
0126	25	買わない
0127	24	買う
0128	24	買う
0129	23	買わない
0130	22	買う
0131	21	買わない
0132	21	買わない

このデータを見て、どう答える？

よく見ると、年齢が高い人だけを計算の対象にしている

回答者	年齢	回答
0001	85	買わない
0002	84	買わない
0003	84	買わない
0004	83	買わない
0005	82	買わない
0006	82	買わない
0007	81	買わない
0008	80	買わない

0125	26	買う
0126	25	買わない
0127	24	買う
0128	24	買う
0129	23	買わない
0130	22	買う
0131	21	買わない
0132	21	買わない

このデータを見て、どう答える？

年齢が低い人たちは「買う」と言っている回答も

回答者	年齢	回答
0001	85	買わない
0002	84	買わない
0003	84	買わない
0004	83	買わない
0005	82	買わない
0006	82	買わない
0007	81	買わない
0008	80	買わない
0125	26	買う
0126	25	買わない
0127	24	買う
0128	24	買う
0129	23	買わない
0130	22	買う
0131	21	買わない
0132	21	買わない

サンプルの偏り

1. サンプルとは

- データが大量にある場合、

  その一部（サンプル）を抜き出して、

  平均値などを計算することがある

2. 注意点

- サンプルに偏りがあると、分析結果がおかしくなる
  - 年齢が高い人だけ
  - 年収の高い人だけ

サンプルの偏り

3. サンプルの出し方（例）

- ID番号が、(1)偶数の人と (2)奇数の人、に分ける

  ＝先頭と最後の偏りをなくす

- 女性から50人、男性から50人ずつ

  ＝性別の偏りをなくす

- それぞれの年齢層から10人ずつ（20代、30代・・・）

  ＝年齢の偏りをなくす

このデータを見て、どう答える？

回答番号が偶数のデータだけを使って計算する

回答者	年齢	回答
0001	85	買わない
0002	84	買わない
0003	84	買わない
0004	83	買わない
0005	82	買わない
0006	82	買わない
0007	81	買わない
0008	80	買わない

0125	24	買う
0126	23	買わない
0127	22	買う
0128	22	買う
0129	21	買わない
0130	20	買う
0131	19	買わない
0132	19	買う

データを正しく読む

1. グラフ(1) 分解レベル
2. グラフ(2) 縦軸
3. グラフ(3) 累計
4. サンプル(1) サンプル数
5. サンプル(2) データの偏り
6. 異常値
7. 欠損値
8. 確証バイアス
9. 単位のちがい
10. フェルミ推定
11. まとめ

今回のポイント

異常値

このデータを見て、どう答える？

顧客にアンケートを取って平均年齢を計算したのですが、50才以上になりました

20～30代が多いはずなのに・・・

漫画：南 智恵

このデータを見て、どう答える？

平均年齢は51才

アンケート番号	年齢
0001	35 才
0002	27 才
0003	30 才
0004	19 才
0005	22 才
0006	26 才
0007	221 才
0008	25 才
平均	51 才

このデータを見て、どう答える？

明らかに誤った数値がある（アンケート回答者の入力ミス？）

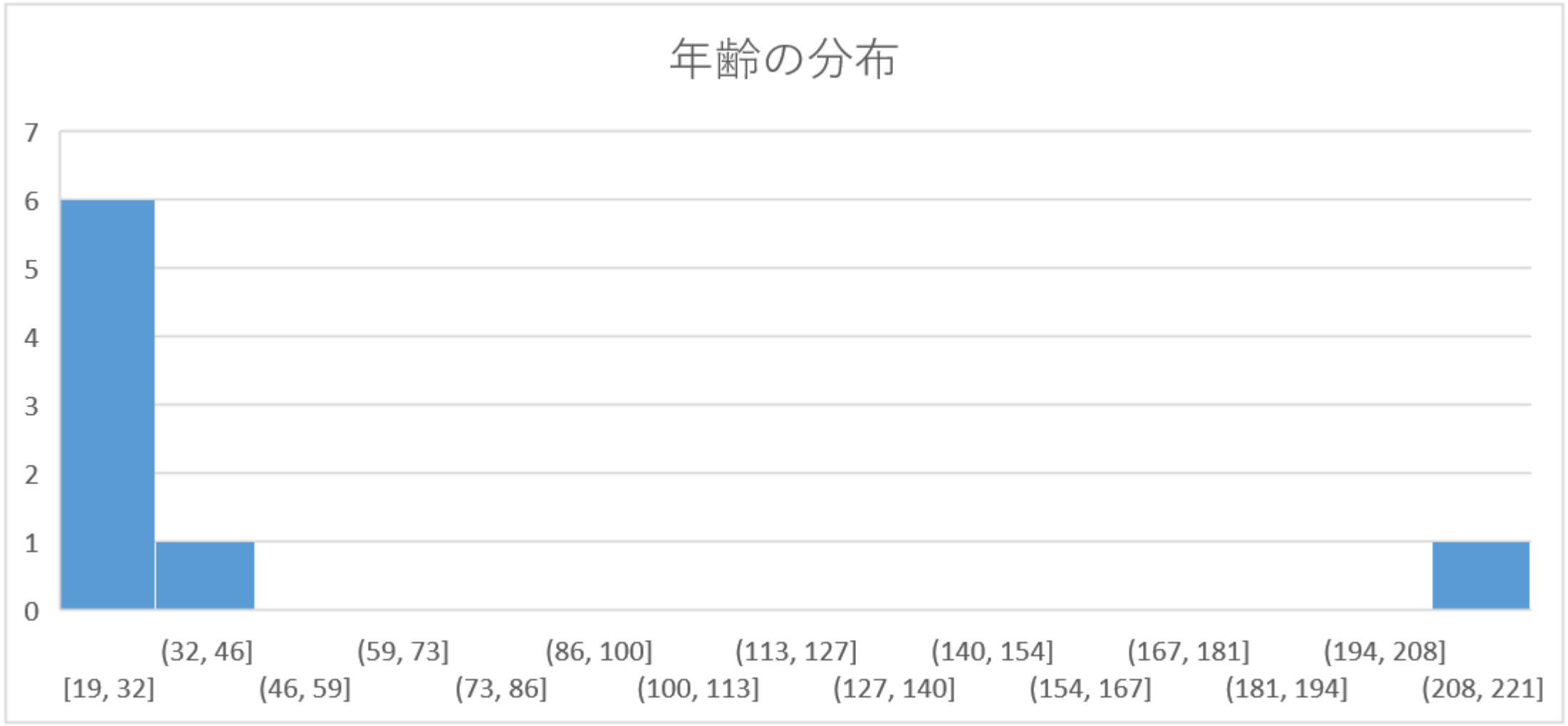
アンケート番号	年齢
0001	35 才
0002	27 才
0003	30 才
0004	19 才
0005	22 才
0006	26 才
0007	221 才
0008	25 才
平均	51 才

異常値

1. ポイント
    - **データがすべて正しいとは限らない**
        - 入力ミスなども当然ありえる
    - 平均値、中央値、標準偏差・・・計算結果に違和感を感じたら、元のデータを見る
2. 異常値チェックのコツ
    - ヒストグラムで見ると、異常値を簡単に見つけられる

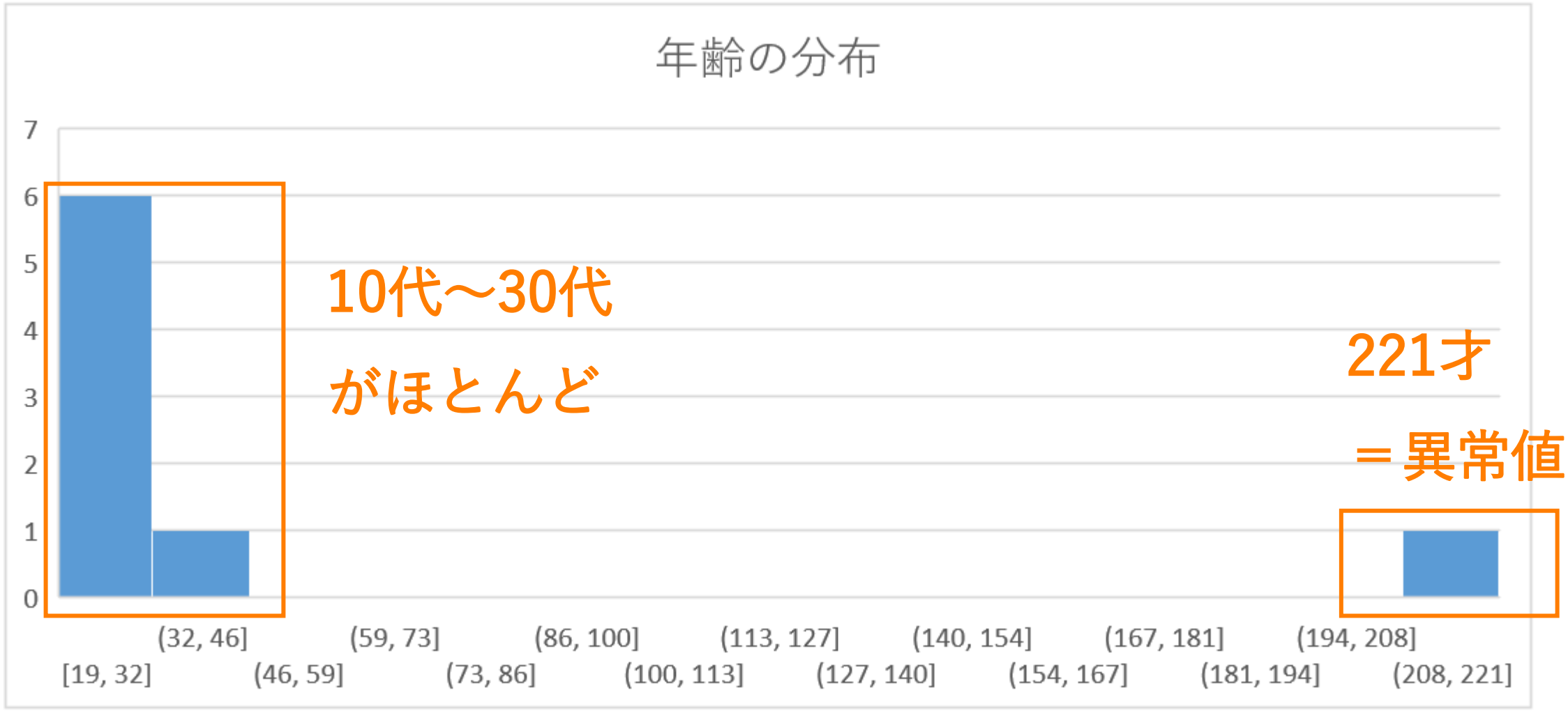
異常値

ヒストグラムで見る



異常値

ヒストグラムで見る



今回のポイント

異常値

データを正しく読む

1. グラフ(1) 分解レベル
2. グラフ(2) 縦軸
3. グラフ(3) 累計
4. サンプル(1) サンプル数
5. サンプル(2) データの偏り
6. 異常値
7. 欠損値
8. 確証バイアス
9. 単位のちがい
10. フェルミ推定
11. まとめ

今回のポイント

欠損値

このデータを見て、どう答える？

顧客アンケートで平均年齢を計算したのですが、普段より高い50才以上になりました

十分な数のアンケートを取ったのに・・・

漫画：南 智恵

このデータを見て、どう答える？

平均年齢は53才

アンケート番号	年齢
0001	才
0002	35 才
0003	才
0004	才
0005	才
0006	才
0007	才
0008	38 才
0009	才
0010	才
0011	才
0012	才
0013	85 才
0014	才
0015	才
0016	才
平均	53 才

このデータを見て、どう答える？

実はデータが欠損していて、回答者は3名しかいない

アンケート番号	年齢
0001	
0002	35 才
0003	
0004	
0005	
0006	
0007	
0008	38 才
0009	
0010	
0011	
0012	
0013	85 才
0014	
0015	
0016	
平均	53 才

欠損値

1. 意味

- **実はデータ数が十分ではないことがある**
- アンケートで「回答必須ではない場合」
- 年齢などプライバシー情報を入力したくない人も多い

2. ポイント

- サンプル数が少なければ、平均値などのズレも大きくなる
- 計算結果がおかしいと感じたら、元データを確認する

今回のポイント

元データを確認する

元データ

1. ポイント

- **元データ（生データ）を必ず確認しましょう**
- 平均値など分析結果だけでは、何か重要なことを見落としている可能性がある
  - 異常な数値はないか？
  - データ数は十分か？
  - サンプルの取り方に偏りはないか？

データを正しく読む

1. グラフ(1) 分解レベル
2. グラフ(2) 縦軸
3. グラフ(3) 累計
4. サンプル(1) サンプル数
5. サンプル(2) データの偏り
6. 異常値
7. 欠損値
8. 確証バイアス
9. 単位のちがい
10. フェルミ推定
11. まとめ

今回のポイント

確証バイアス

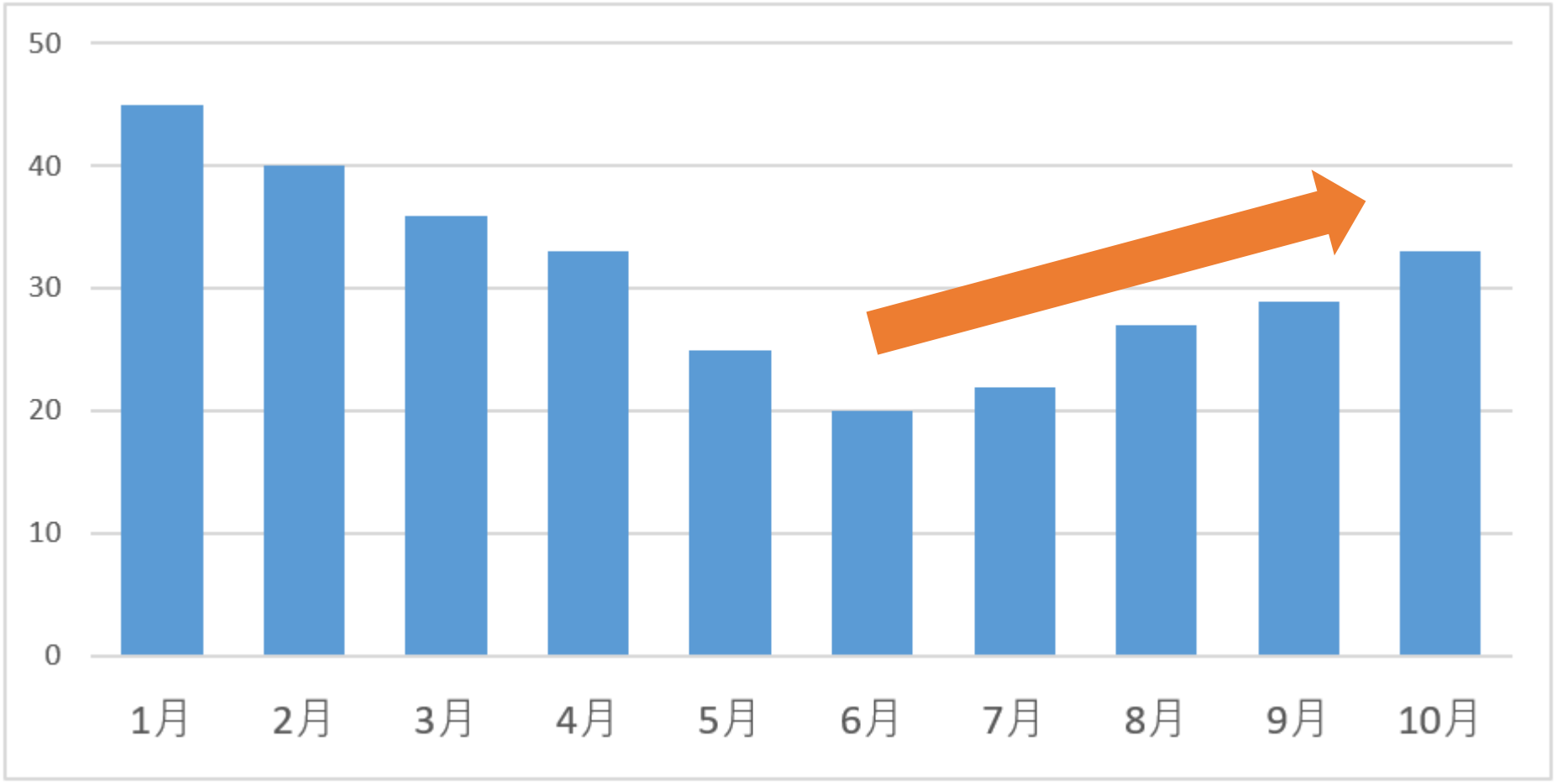
このデータを見て、どう答える？

最近すごく販売数が伸びてます！

漫画：南 智恵

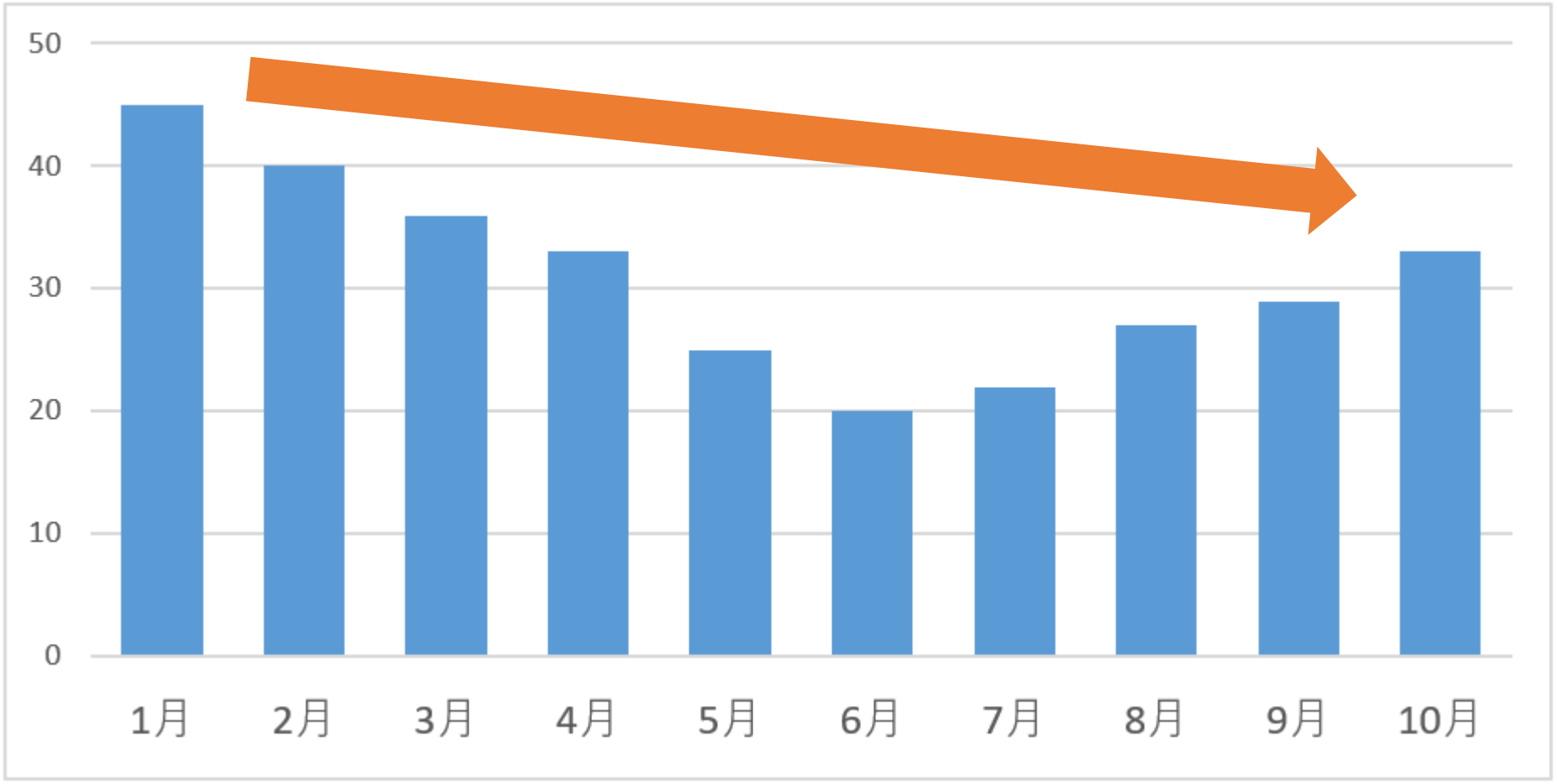
このデータを見て、どう答える？

「最近すごく販売数が伸びてます！」



このデータを見て、どう答える？

1月から比較すると、下がっている



確証バイアス

1. 意味

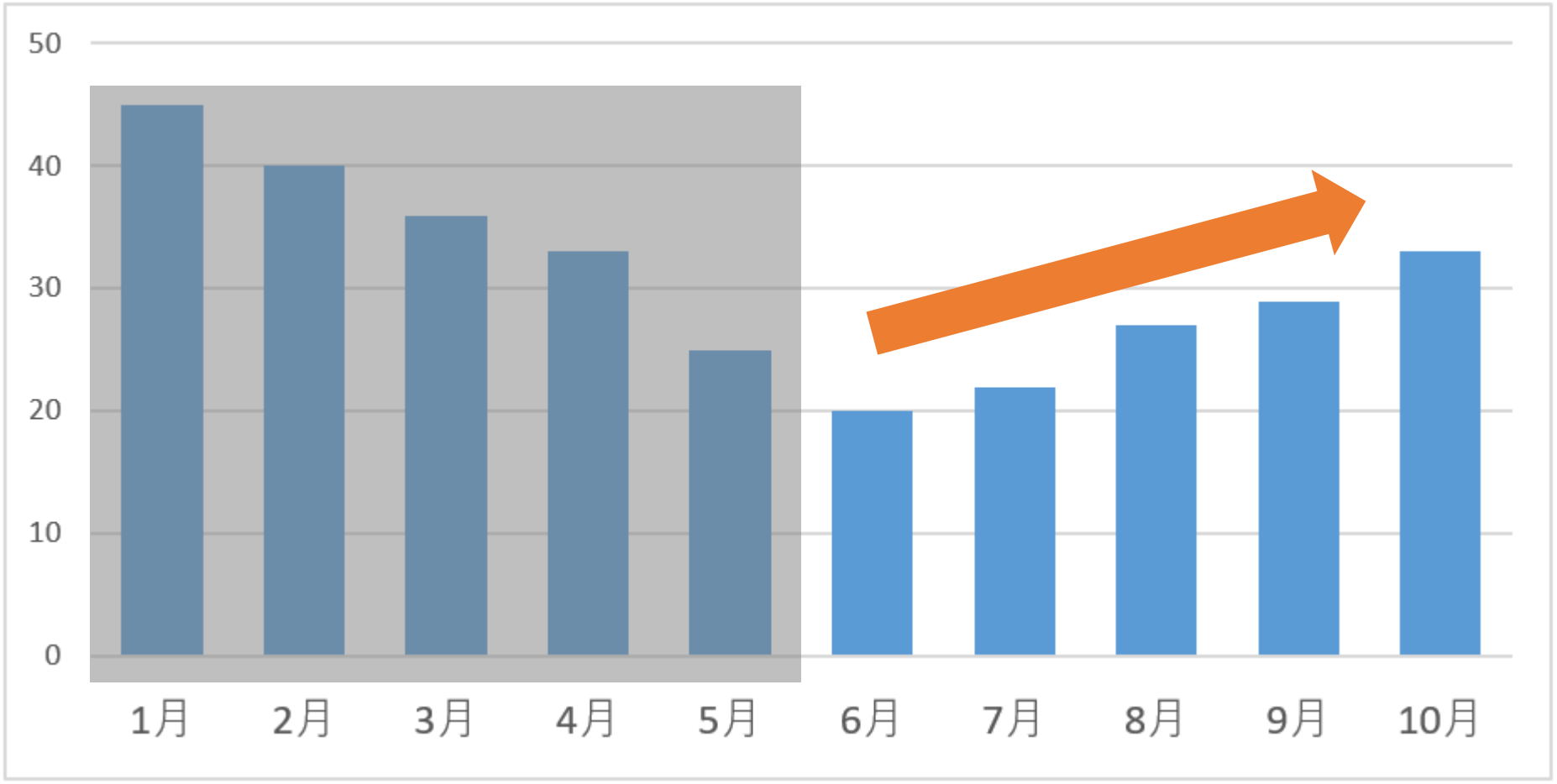
- 自分の分析が正しいと思い込み、
  その分析に都合の良いデータだけを集めてしまい、
  合理的な判断ができなくなっている

2. ポイント

- 常に批判的な考えを持つ（自分の意見は正しいのか？）
- 他者の意見を聞く

このデータを見て、どう答える？

1月～5月の販売数の減少を見て見ぬふりをしてしまっている



今回のポイント

確証バイアス

データを正しく読む

1. グラフ(1) 分解レベル
2. グラフ(2) 縦軸
3. グラフ(3) 累計
4. サンプル(1) サンプル数
5. サンプル(2) データの偏り
6. 異常値
7. 欠損値
8. 確証バイアス
9. 単位のちがい
10. フェルミ推定
11. まとめ

今回のポイント

単位のちがい

このデータを見て、どう答える？

新しい営業管理システムを導入しないか
という提案があったんですけど、
月10万円で高いんですよね・・・

漫画：南 智恵

単位のちがい

1. 例
    - 新しい営業管理システムの導入を検討
    - 毎月10万円
    - システムを導入すると、100人が毎月2時間の残業を削減

      → 会社としてシステムを導入すべきか？
2. ポイント
    - 10万円という費用を、何と比較するか？ → 残業時間
    - 費用と時間は比較できない → 残業時間を残業費用に換算

単位のちがい

1. 例
    - 新しい営業管理システムの導入を検討
    - 毎月10万円
    - システムを導入すると、100人が毎月2時間の残業を削減

      → 会社としてシステムを導入すべきか？
2. 計算例
    - 100人×2時間×時給2,500円＝50万円のコスト削減
    - 月10万円という価格は安い

単位のちがい

1. ポイント

- 「費用と時間」など、比較しにくいケースも多い
- そういうときは、すべて金額に換算すると分かりやすい

2. 例

- 営業用の車を買えば、1日1件多く受注できる（売上アップ）
- 1件×月20日×1件あたり5,000円＝月10万円
- 車を3年使うとすると、10万円×12ヶ月×3年＝360万円

  → 100万円くらいの車なら、買っても元が取れそう

データを正しく読む

1. グラフ(1) 分解レベル
2. グラフ(2) 縦軸
3. グラフ(3) 累計
4. サンプル(1) サンプル数
5. サンプル(2) データの偏り
6. 異常値
7. 欠損値
8. 確証バイアス
9. 単位のちがい
10. フェルミ推定
11. まとめ

今回のポイント

フェルミ推定

このデータを見て、どう答える？

新商品をコンビニで販売したいのですが、
そもそも東京にコンビニって何店舗あるの
だろう？調べたけど分からない・・・

漫画：南 智恵

フェルミ推定

1. 意味

- ビジネスの現場では、データが見つからないこともある
- **いま手元にある情報や経験を元に推定する手法**

フェルミ推定

1. ポイント

- 考え方の切り口はさまざま
- 柔軟に、いろいろ考えてみるのが大事

2. 注意点

- フェルミ推定は、高い精度で答えを出すのは難しい
- **あくまで大体の数値を推定するための手法**

「東京のコンビニは、5,000店～15,000店くらいかな？」

フェルミ推定

東京都にコンビニが何店舗あるか、推定してみる

地理状況から推定	近所の店舗数と人口の割合から考える
利用状況から推定	利用者数とコンビニ稼働から考える
他と比較して推定	他の業態と比較して考える

フェルミ推定

東京都にコンビニが何店舗あるか、推定してみる

地理状況から推定	近所の店舗数と人口の割合から考える
利用状況から推定	利用者数とコンビニ稼働から考える
他と比較して推定	他の業態と比較して考える

フェルミ推定

1. 例
    - 近所では、コンビニ1店舗の周辺に300世帯くらいある
        - 1世帯＝人口3人くらい？と仮定
    - 東京都の人口＝1,000万人と仮定

2. 計算式
    - 1,000万人÷（300世帯×3人）＝11,111店
    - だいたい1万店くらいか？

フェルミ推定

東京都にコンビニが何店舗あるか、推定してみる

地理状況から推定	近所の店舗数と人口の割合から考える
利用状況から推定	**利用者数とコンビニ稼働から考える**
他と比較して推定	他の業態と比較して考える

フェルミ推定

1. 例
    - 今日、東京都でコンビニを使ったのは、人口の1/4くらい？
        - 1,000万人 × 1/4 ＝ 250万人
    - この来店客をさばくのに、何店舗が必要か？
    - 1店舗あたり、1時間10人来店して、18時間稼働
        - 24時間営業だが、深夜の6時間はゼロ人と仮定
2. 計算式
    - 250万人÷（1時間10人×18時間）＝13,889店

フェルミ推定

東京都にコンビニが何店舗あるか、推定してみる

地理状況から推定	近所の店舗数と人口の割合から考える
利用状況から推定	利用者数とコンビニ稼働から考える
他と比較して推定	**他の業態と比較して考える**

フェルミ推定

1. 例
    - スーパーの店舗は3,000店あるというデータがあった
    - 近所を見渡すと、コンビニはスーパーの2倍くらいある

2. 計算式
    - 3,000店×2倍＝6,000店くらい？

フェルミ推定

東京都にコンビニが何店舗あるか、推定してみる

地理状況から推定	近所の店舗数と人口の割合から考える
利用状況から推定	利用者数とコンビニ稼働から考える
他と比較して推定	他の業態と比較して考える

フェルミ推定

1. ポイント

- 考え方の切り口はさまざま
- 柔軟に、いろいろ考えてみるのが大事

2. 注意点

- フェルミ推定は、高い精度で答えを出すのは難しい
- **あくまで大体の数値を推定するための手法**

「東京のコンビニは、5,000店～15,000店くらいかな？」

フェルミ推定

1. ビジネスにおける利用シーン
    - 大体の市場規模が分かるだけでも議論しやすい
2. 例えば
    - **「東京で年収1,000万円以上、アニメ好きな30代の女性を狙う」**

      → どれくらいの人数がいるのだろうか？（市場規模）

フェルミ推定

1. ビジネスにおける利用シーン
   - 大体の市場規模が分かるだけでも議論しやすい
2. 例えば
   - **「東京で年収1,000万円以上、アニメ好きな30代の女性を狙う」**
   - 年収1,000万円以上　＝　全体の5%くらい？
   - アニメ好き　＝　全体の10%くらい？
   - 30代女性　＝　全体の10%くらい？

   → 東京人口1,000万人×5% ×10% ×10%＝5,000人

   → あまり市場規模は大きくなさそう（だから狙わない）