

# グラフは 見える化にも 誤解の温床にもなる

データサイエンスへのいざない 第2回  
6つのポイントと 7つの追加事件

山梨大学教育学部 稲垣俊介



# 今日の6つのポイント

## 1. 軸

どこから どこまでを 見せるか

## 2. 期間

どの時期を 切り取るか

## 3. 累計

月次か 累計か で印象が変わる

## 4. 割合

実数と シェアは 別物

## 5. 単位

名目と 実質は 違う

## 6. 二軸

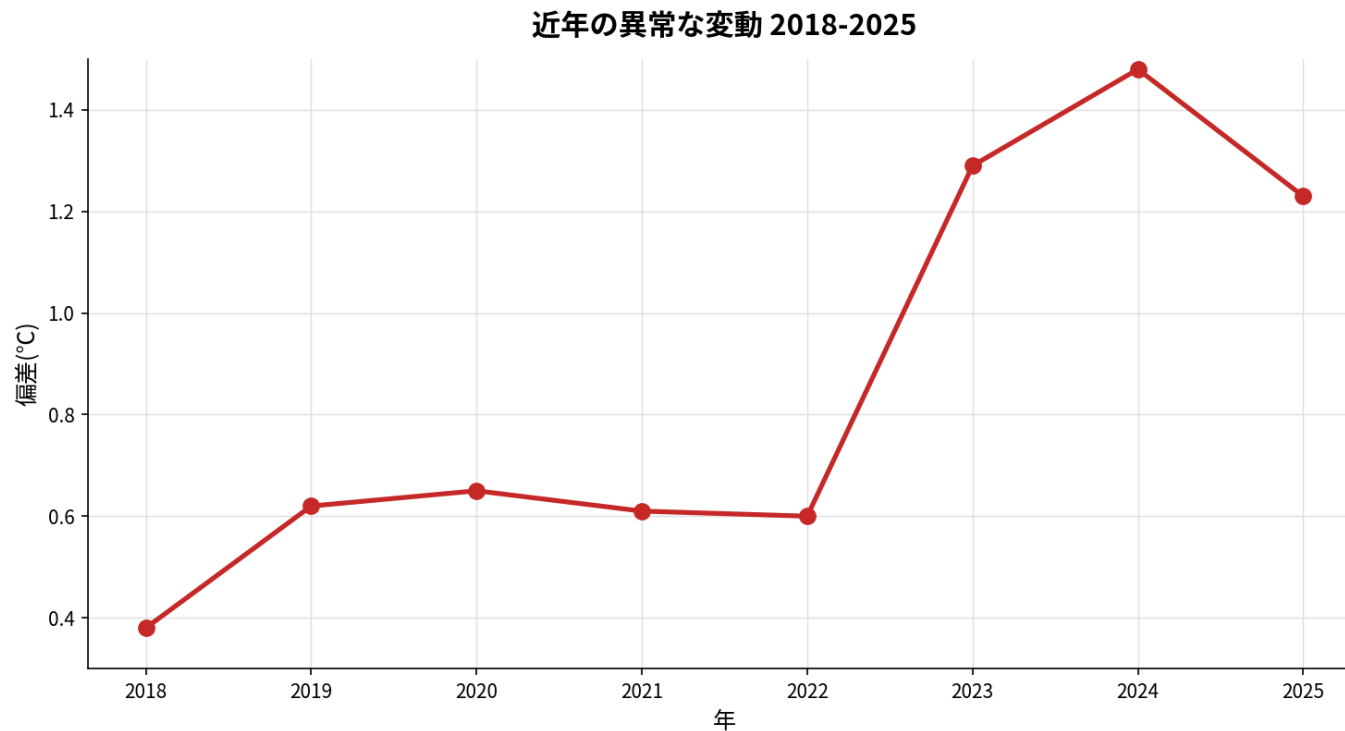
重ねると 関係ありそうに 見える

クイズ形式で進めます。ずるいグラフを先に見せますので、どこがずるいか考えてみてください。



# 事件1 気温 これは異常な温暖化？

Q. 気象庁のデータ。2018年から急に気温が上がっています。これは異常な温暖化では？



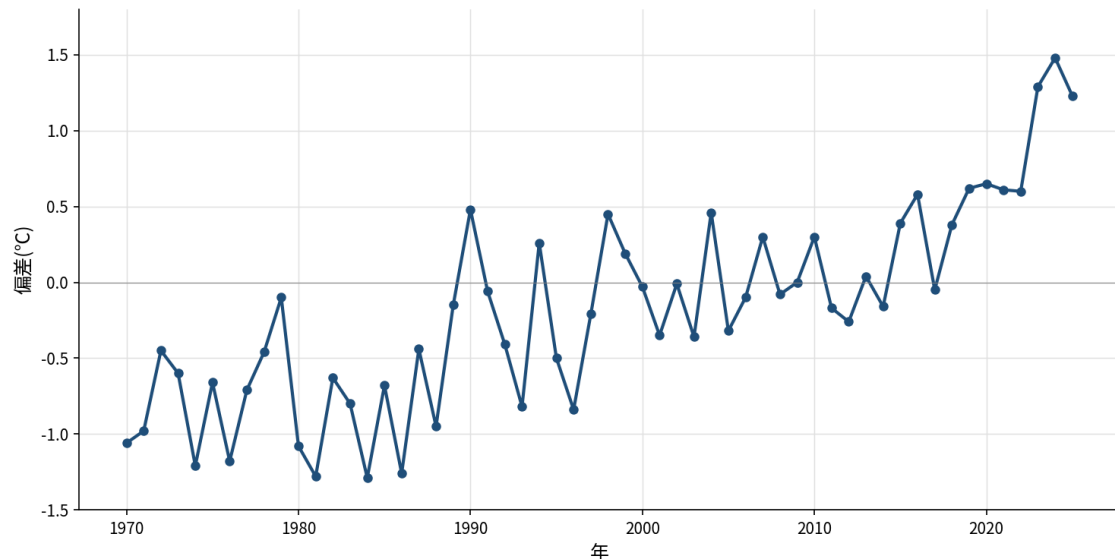
隣の人と 1分 話し合ってみよう。このグラフ どこがずるい？



# 種明かし 事件1 軸と期間を 切り取っただけ

## フェア 1970-2025の55年を見る

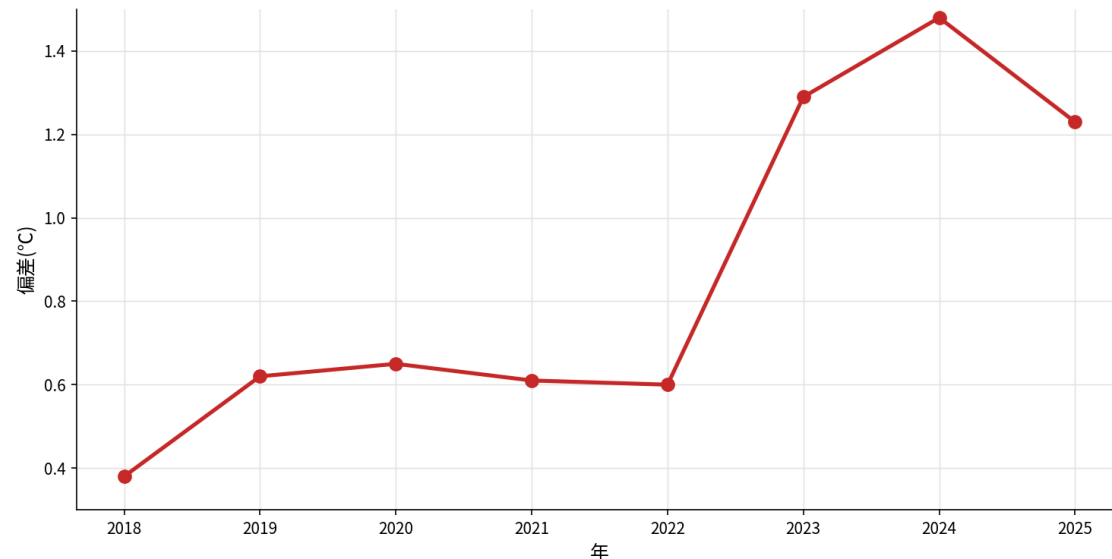
日本の年平均気温偏差 1970-2025



長期で見れば1990年代から上昇傾向。2023-2025が特に高い。異常な急変ではなく 累積した温暖化の結果。

## ずるい 2018-2025の8年に絞る

近年の異常な変動 2018-2025



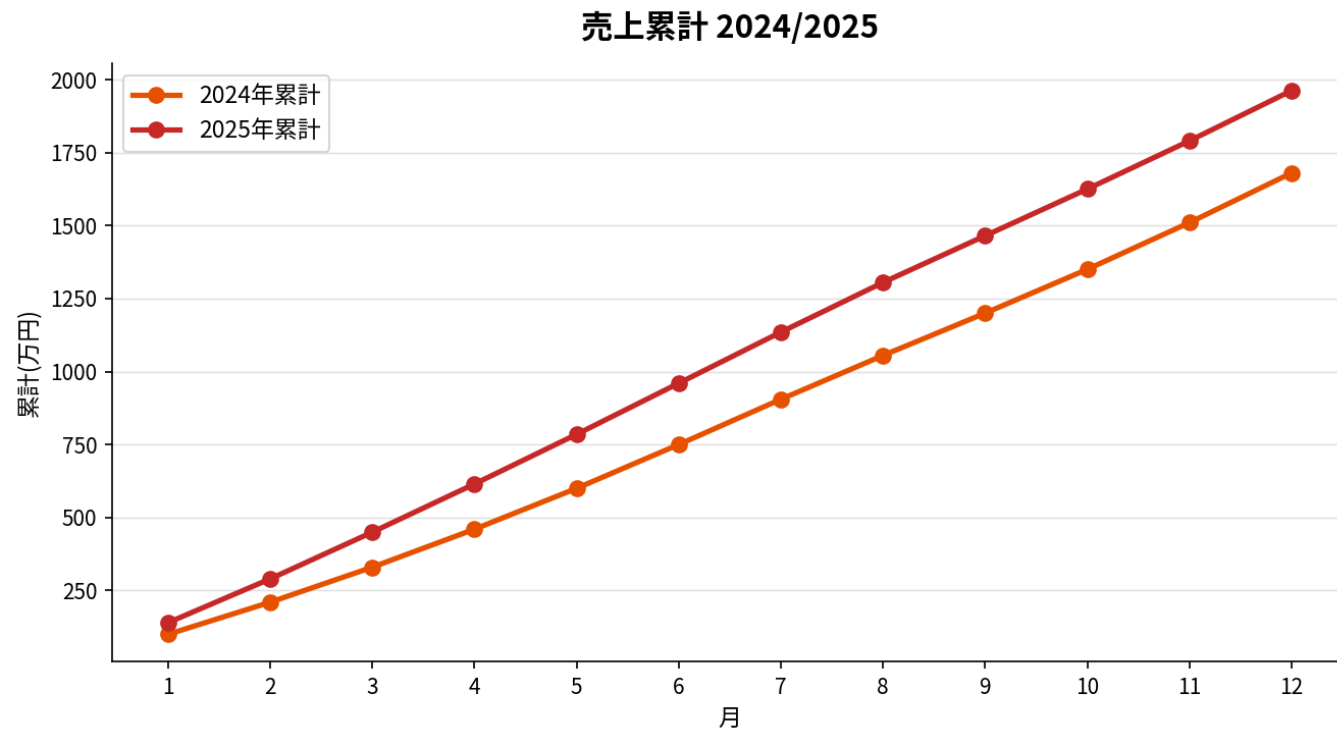
縦軸を0.3~1.5に絞り 8年だけ見せる。2023年の急上昇を強調しやすい。

鍵は 軸と期間。同じ気象庁のデータから まったく違う印象を作れる。

出典 気象庁「日本の年平均気温偏差」

## 事件2 このカフェ 順調に成長？

Q. あるカフェの売上累計。2年ともずっと右肩上がり。絶好調ですね？



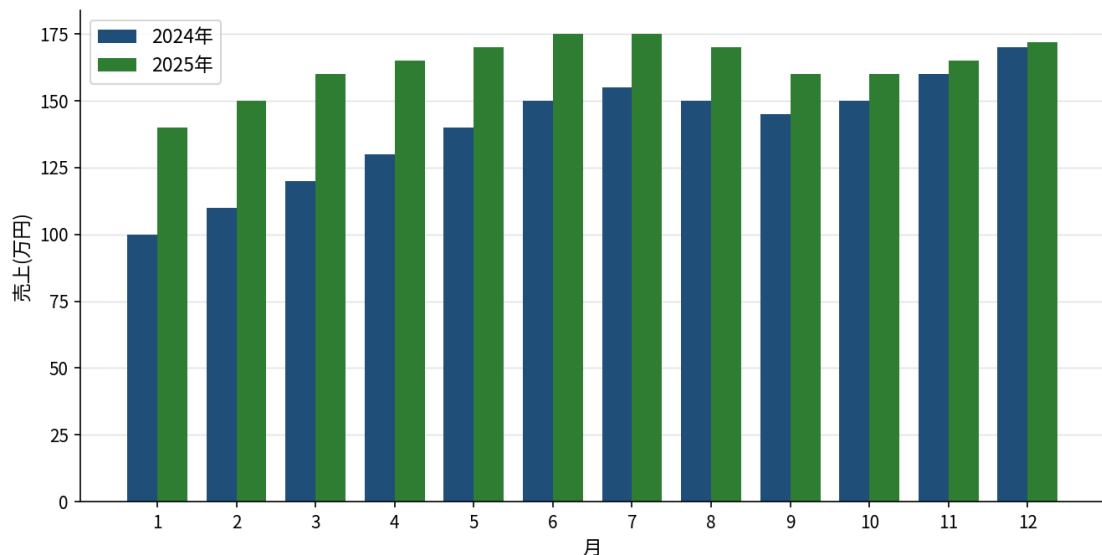
本当に 絶好調？ このグラフの どこに ひっかかりますか？



# 種明かし 事件2 累計は基本 下がらない

## フェア 月次で見ると

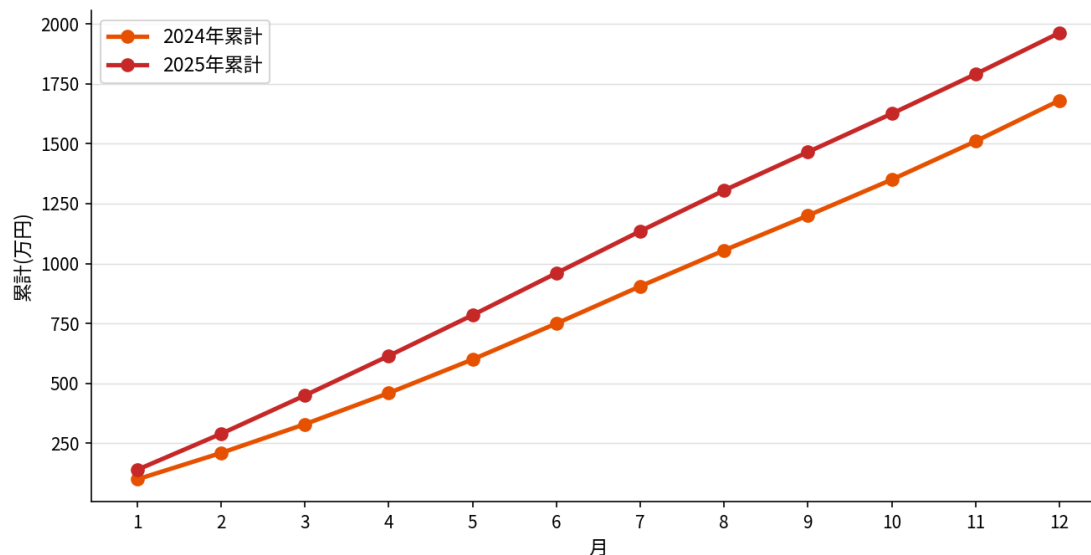
月次売上 2024/2025



1月は前年比40%増の大成長。でも12月は前年比1.2%まで鈍化。後半は成長が止まっている。

## ずるい 累計で見せる

売上累計 2024/2025



累計グラフは基本下がらない。鈍化していても ずっと右肩上がりに見える。

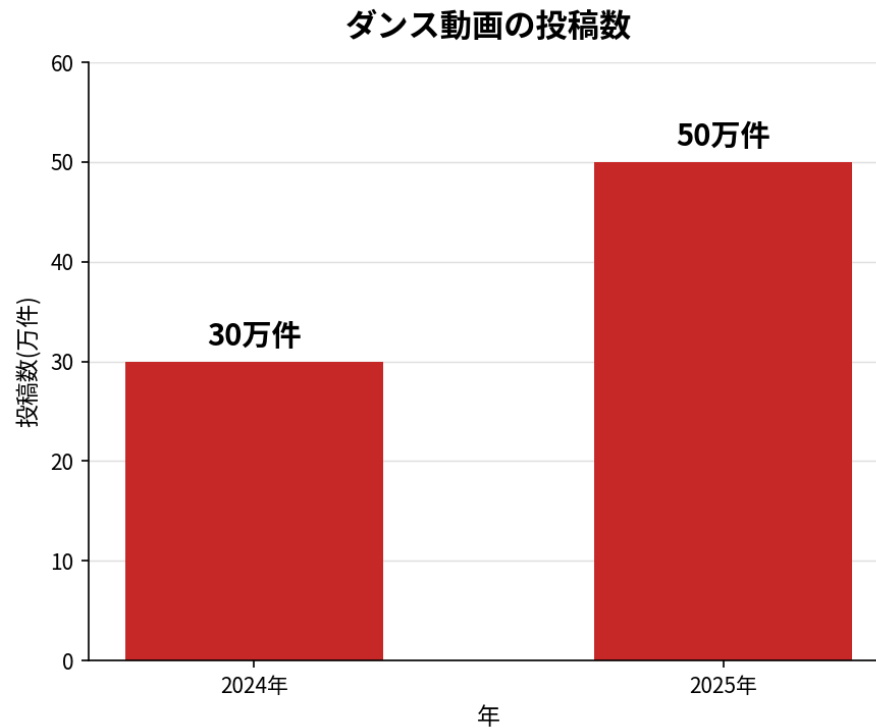
鍵は 累計。好調を演出したいときに 企業はよく使う。月次で見ると違う姿が見える。

教材用仮想データ



## 事件3 ダンス動画が大ブーム?

Q. あるSNSのダンス動画の投稿数。1年で1.7倍の大成長!ダンスブーム到来?

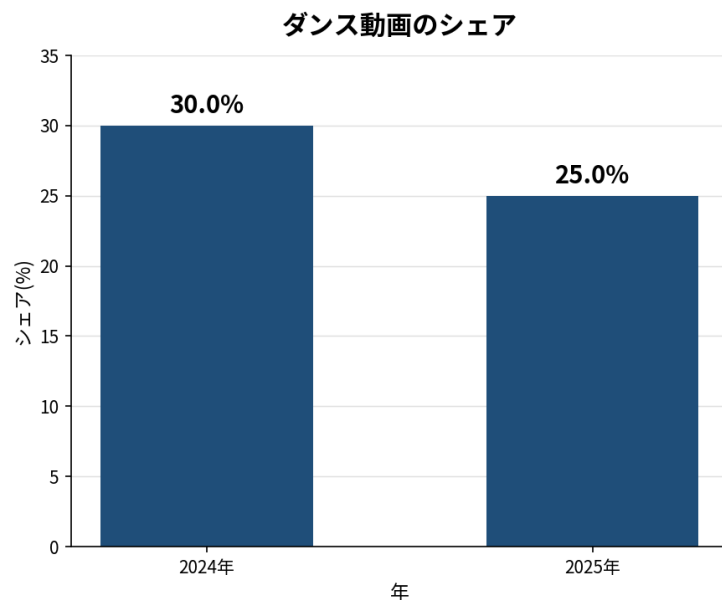


本当にブーム? 何か見えていない数字はありませんか?



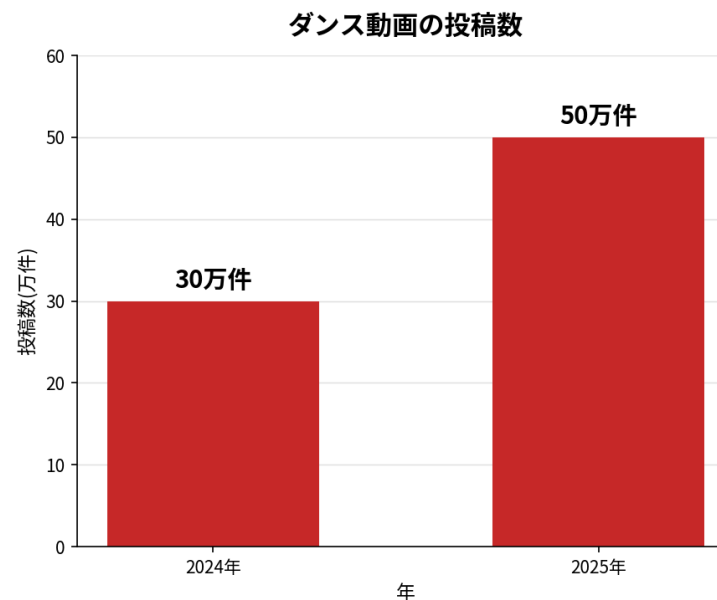
# 種明かし 事件3 シェアは下がっていた

## フェア シェアで見ると



ダンス動画のシェアは30.0%→25.0%に低下。他ジャンルがもっと伸びたから。

## ずるい 投稿数だけ見せる



投稿数は30万→50万で 1.7倍。だけど 全体の投稿数は100万→200万で 2倍に増えている。

鍵は 割合。実数が増えても シェアは下がることもある。「増えた」の裏を見る。

教材用仮想データ

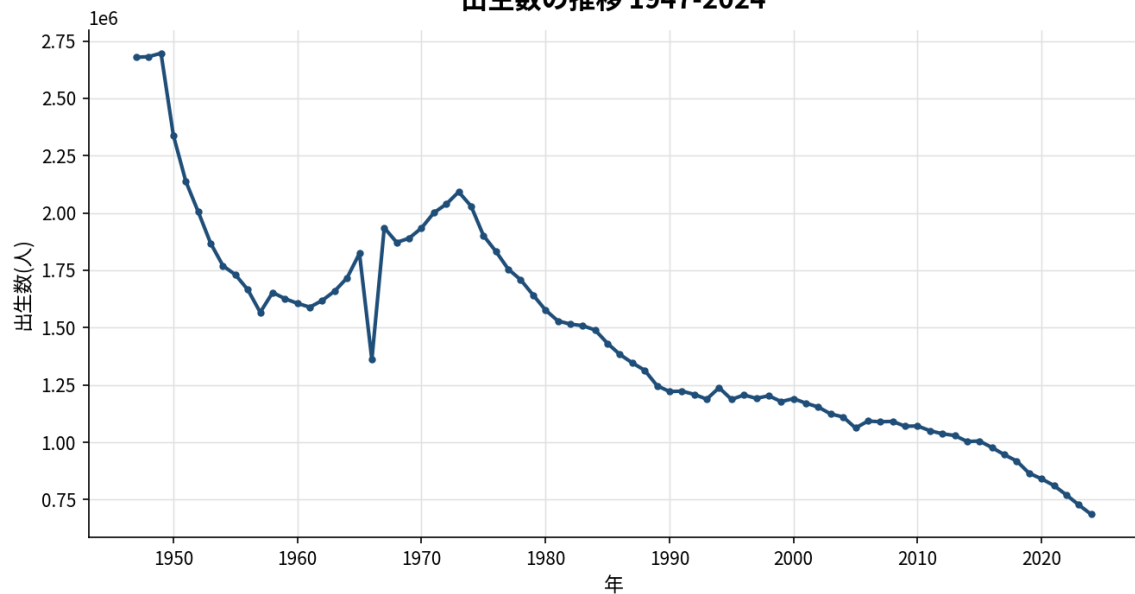


# 事件4 件数と率 どちらが本当？

どちらも厚生労働省の公式データ。同じ「少子化」を別の角度から見せる。

## 出生数（件数）

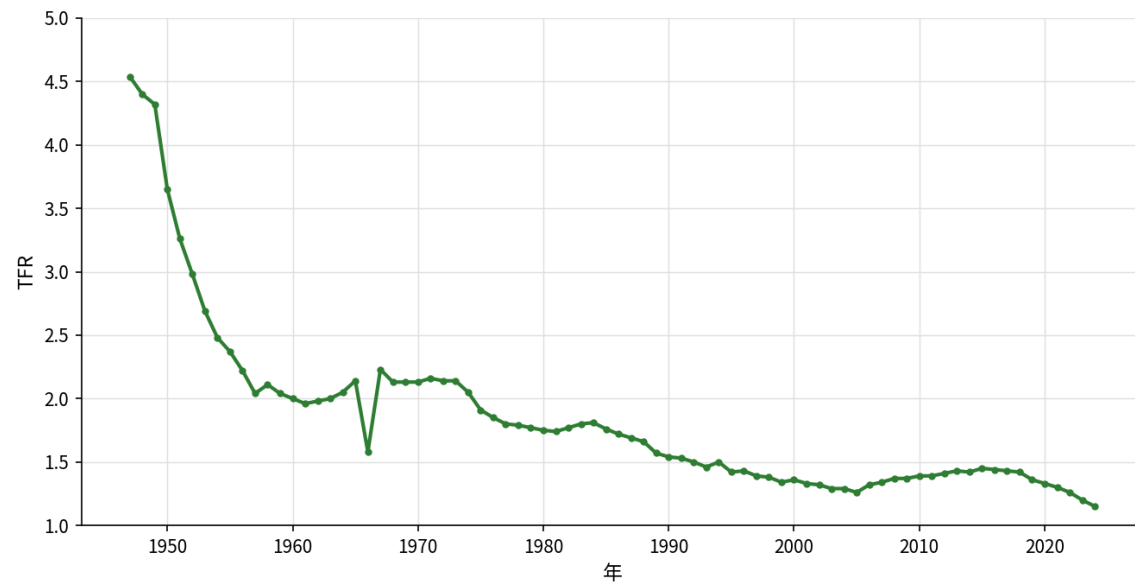
出生数の推移 1947-2024



戦後のベビーブーム以降 長期的に減少。2024年は68万人。

## 合計特殊出生率（TFR）

合計特殊出生率 1947-2024



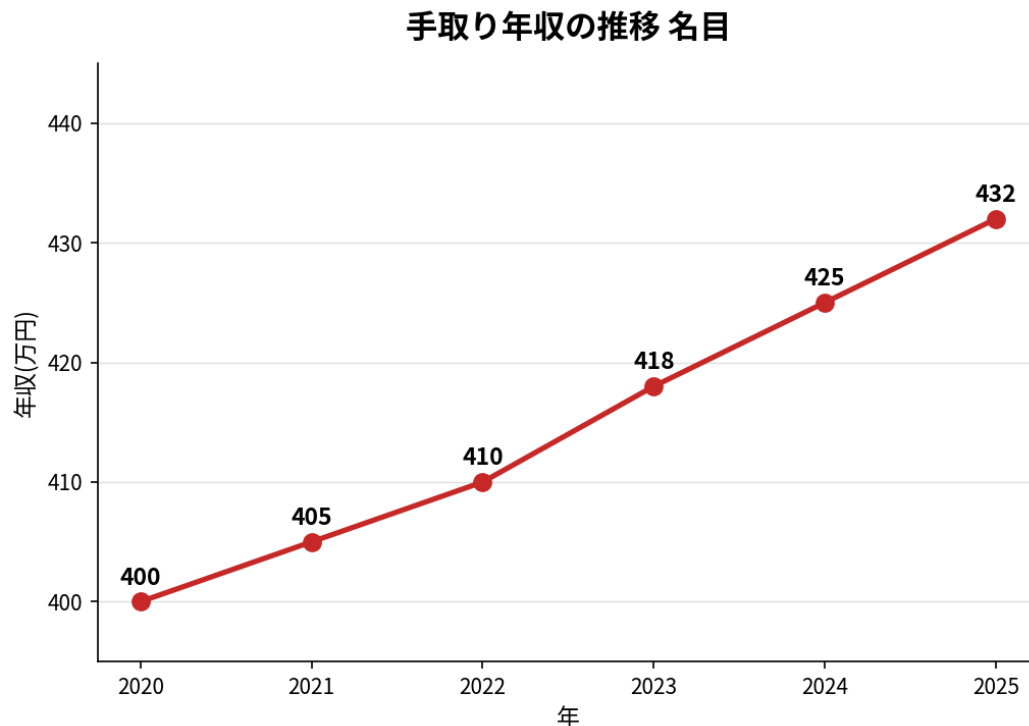
1人の女性が生涯に産む子どもの数の平均。2024年は1.15。

出典 厚生労働省「人口動態統計」 件数(出生数)と率(TFR)は、別の問いに答えている。



# 事件5 毎年 給料が上がっています!

Q. ある人の手取り年収。2020年の400万円から 2025年には432万円へ。給料UP!



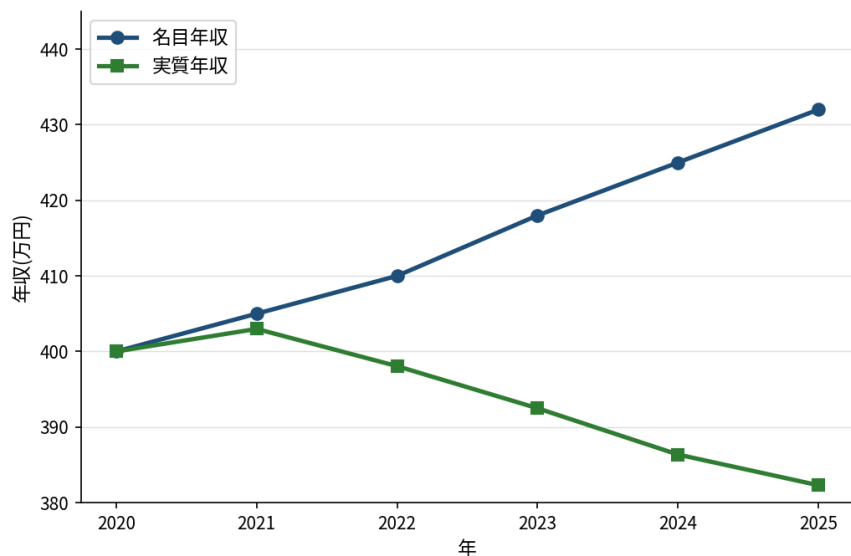
本当に豊かになっていますか？何か隠れていませんか？



# 種明かし 事件5 実質は下がっていた

## フェア 名目と実質を並べる

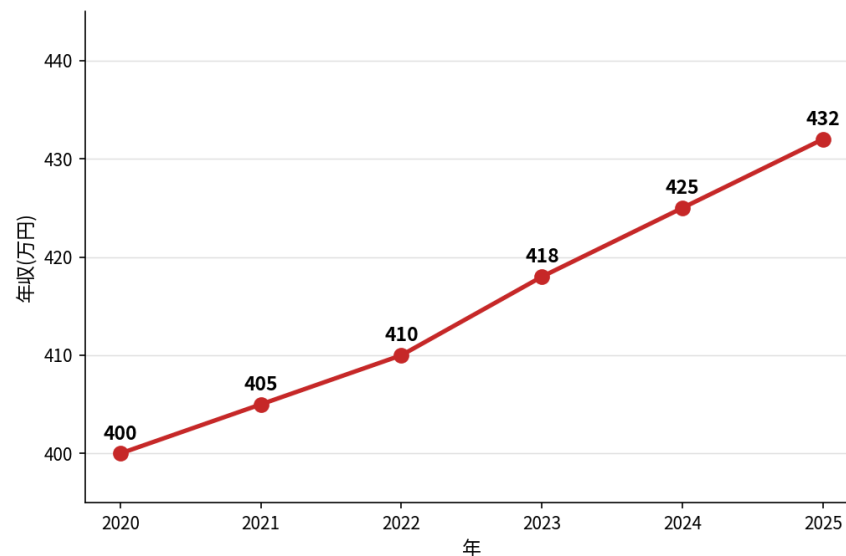
手取り年収 名目と実質



名目は上昇。でも物価上昇率のほうが大きく 実質年収は400万円→382万円に下がっている。

## ずるい 名目だけ見せる

手取り年収の推移 名目



名目年収だけ見れば 毎年きれいに右肩上がり。給料UPの印象だけ残る。

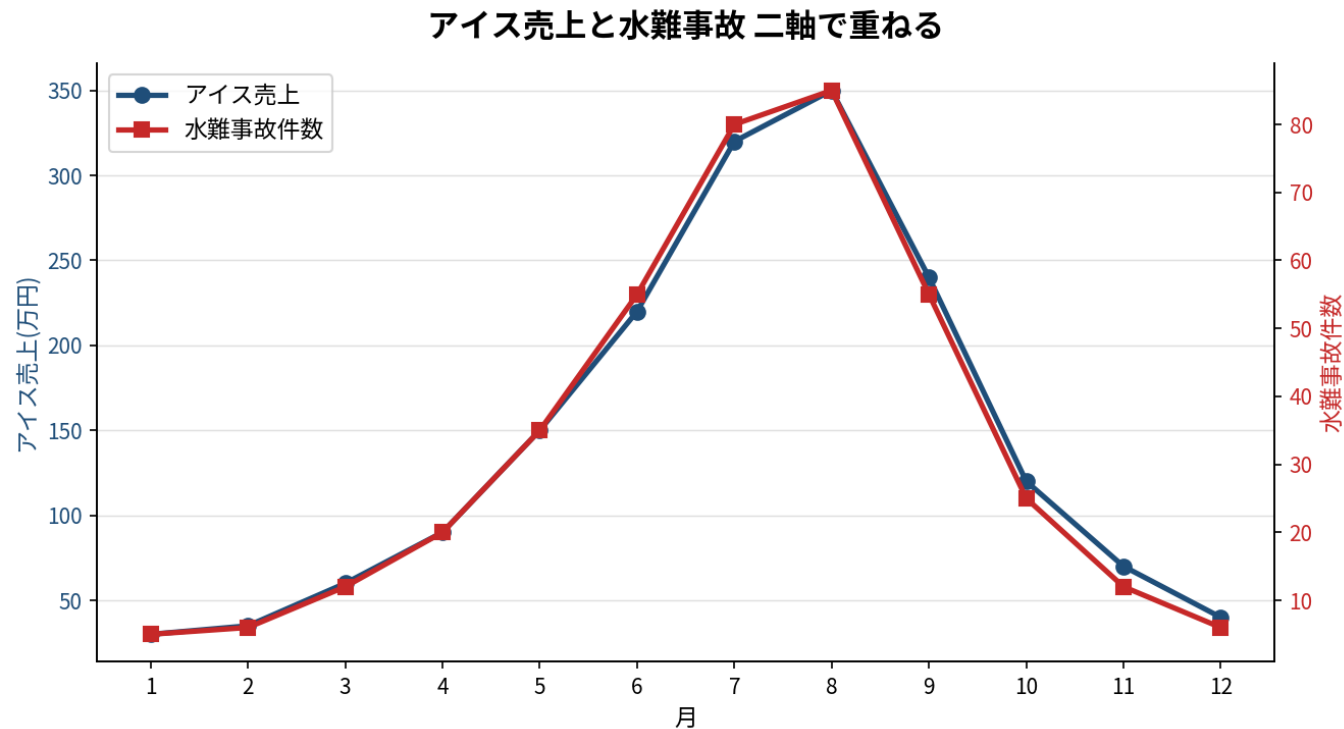
鍵は 単位。お金の数字は 物価を通して「買えるもの」に変換してはじめて意味を持つ。

教材用仮想データ



# 事件6 アイスが売れると 人が溺れる？

Q. アイス売上と水難事故件数。ほとんど同じ動き 相関係数 0.998。

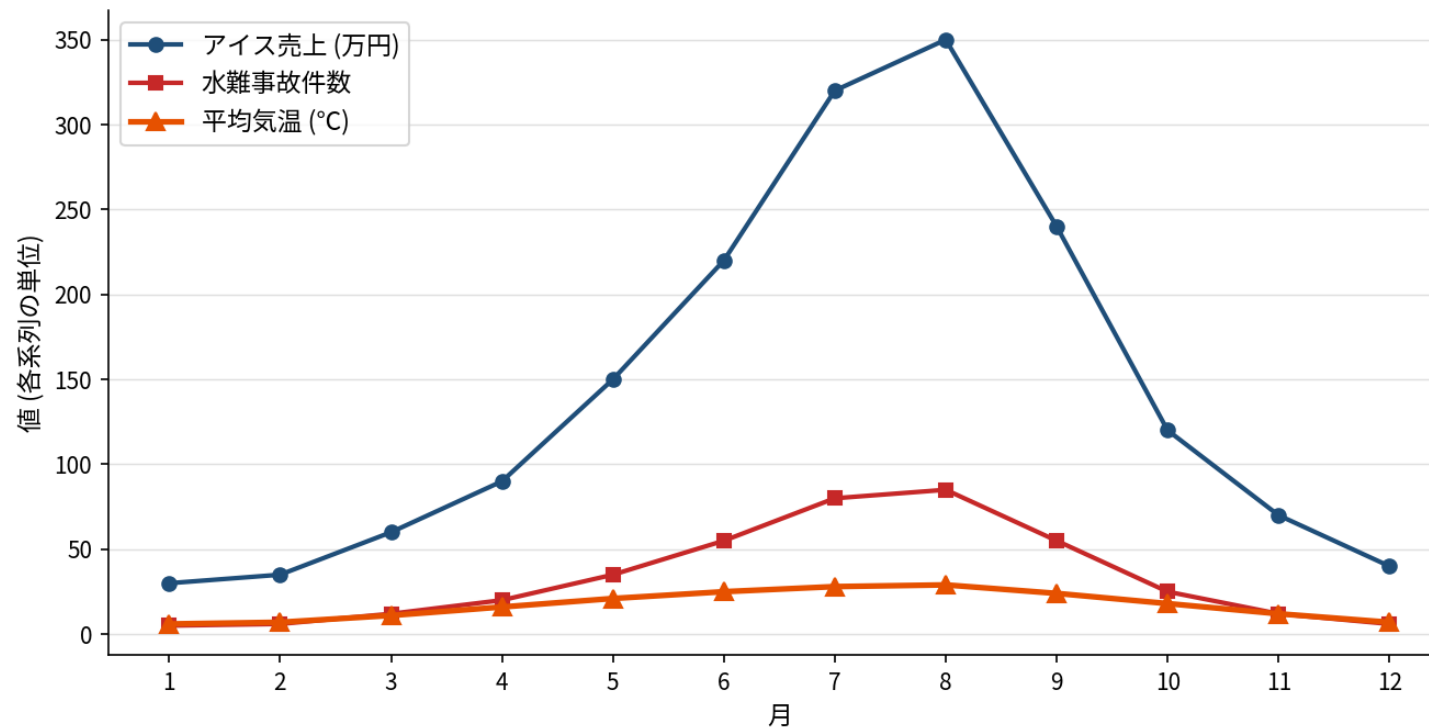


アイスが事故の原因？ それとも 別の何か？ 話し合ってみよう。



# 種明かし 事件6 第三の要因「気温」

第三の要因「気温」が両方を動かしている



夏は気温が上がる → アイスの売上も 水遊びの人も 増える → 両方が同時に増えるだけ  
アイスと事故の間に 直接の因果関係はない。第三の要因(気温)が両方を動かしている。

鍵は 二軸。相関は因果ではない。二軸で重ねるときは「第三の要因」を疑う。



# ここまでのまとめ フェアとずるいの違い

## フェアなグラフの特徴

- ・ 縦軸をゼロから適切に取る
- ・ 期間を十分に長く取る
- ・ 実数と割合の両方を見せる
- ・ 名目と実質の両方を示す
- ・ 二軸を避ける or 注釈を入れる
- ・ 出典と定義を明記する

## ずるいグラフの特徴

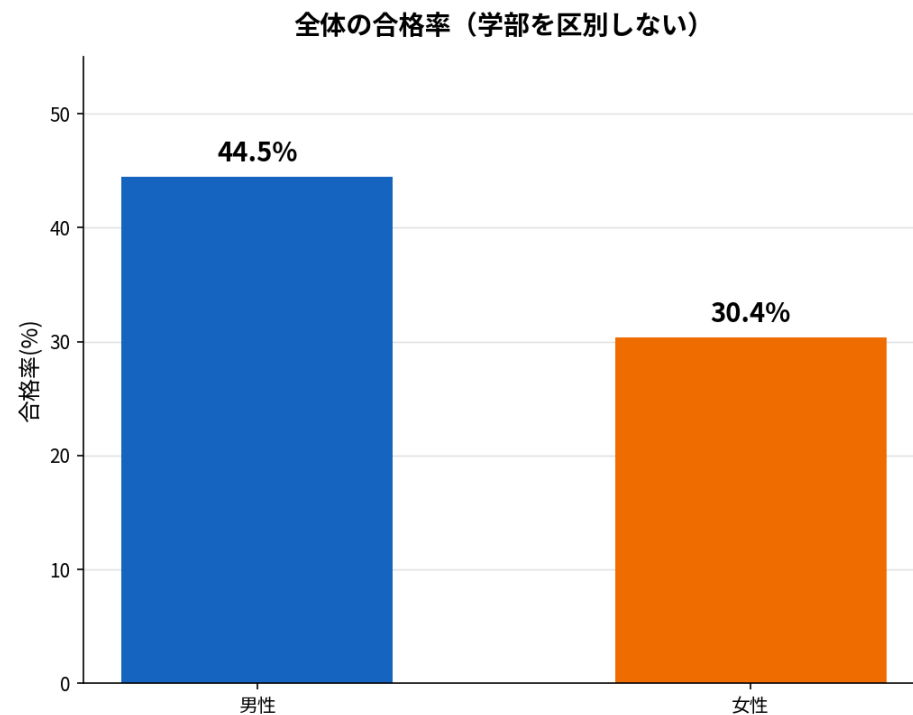
- ・ 縦軸をゼロから離す/切り取る
- ・ 都合のよい期間だけ切り出す
- ・ 実数だけ or 割合だけ見せる
- ・ 名目だけ見せ 物価を隠す
- ・ 二軸で関係ありそうに見せる
- ・ 出典と定義を曖昧にする

どちらも「嘘」ではない。でも見え方が全く違う。どの見せ方を選ぶかには「意図」がある。



# 追加事件1 男女で差別があった？ UCバークレー 1973年

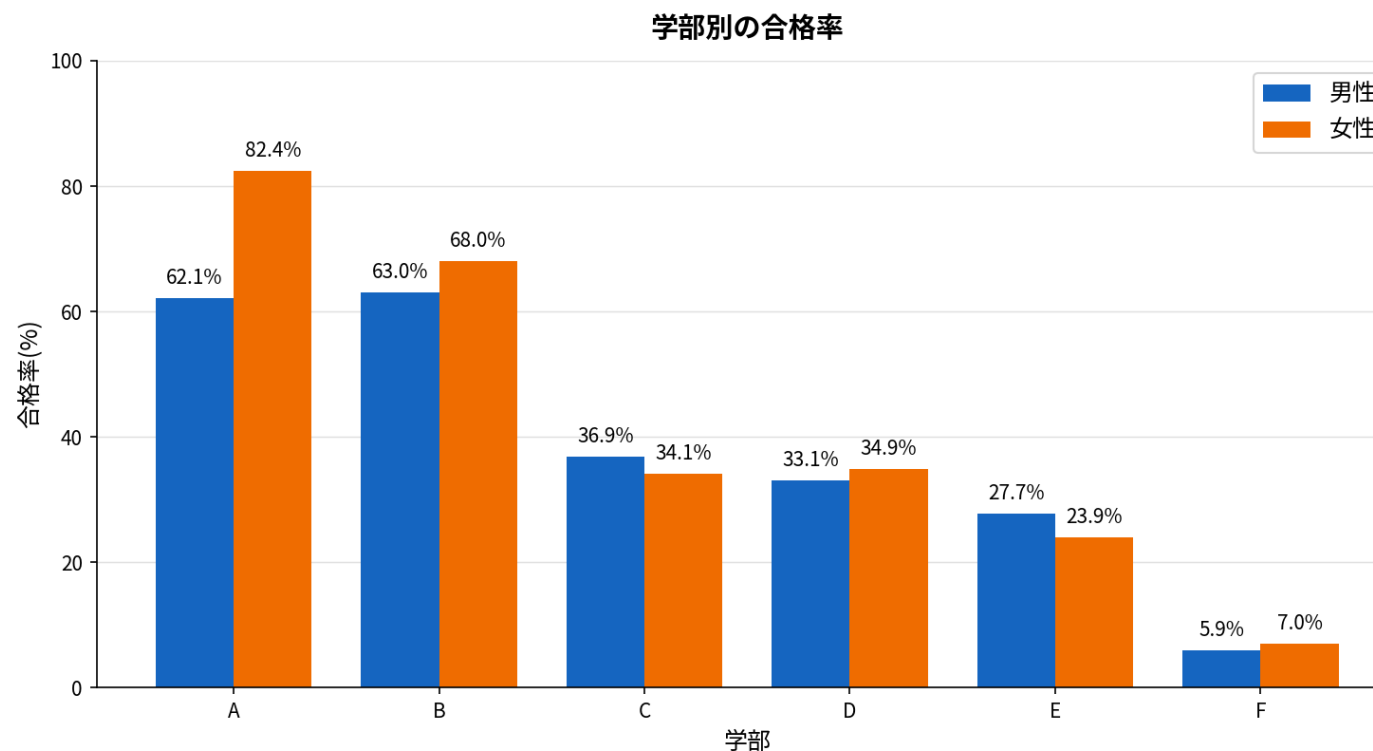
Q. 1973年 UCバークレー大学院の合格率。男女で差があった。これは入試差別？



男性 44.5% vs 女性 30.4%。これだけで 差別と言えるでしょうか？



# 種明かし 出願先の偏りがトリック



学部別に見ると 多くの学部で 女性の方が合格率が高い。

男性は合格率の高い学部(A・B)に 女性は合格率の低い学部(C・F)に集中して出願していた。

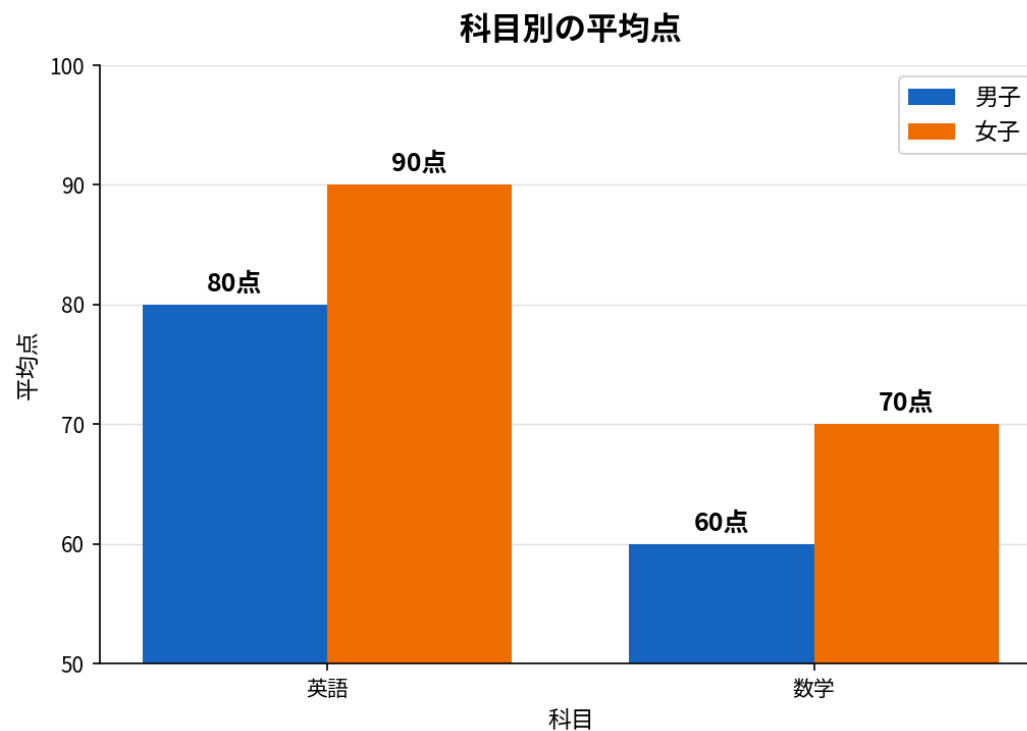
結果 各学部で公平な入試をしても 全体では女性の合格率が低く出る。

Bickel et al. (1975) Science 187, 398-404. 「差別」は入試ではなく 進学選択の偏りにあったかもしれない。



## 追加事件2 もう一つのシン普森 身近な例

Q. ある大学の英語と数学の平均点。両方の科目で 女子の方が10点高い。

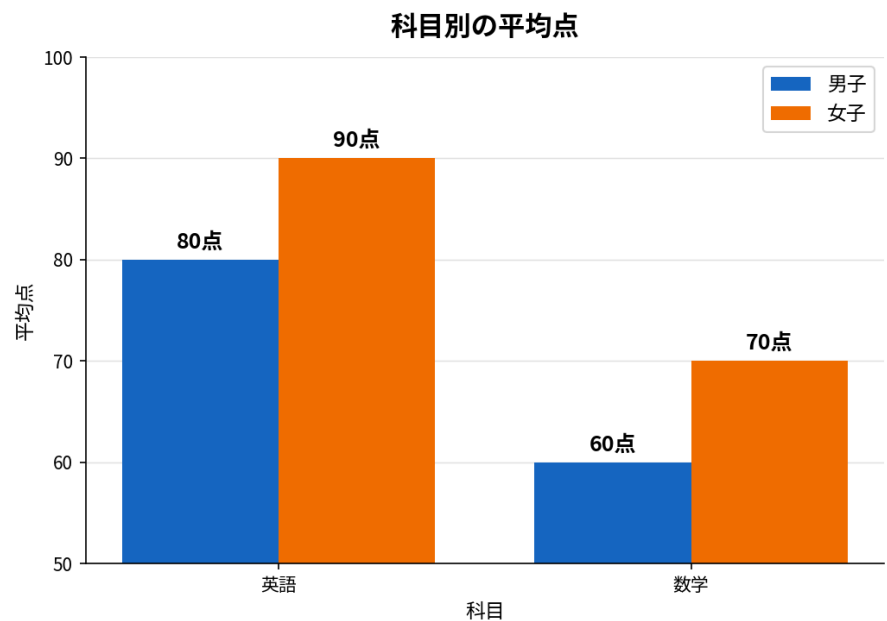


では 2科目合計の全体平均では どちらが高い？ 予想してみよう。

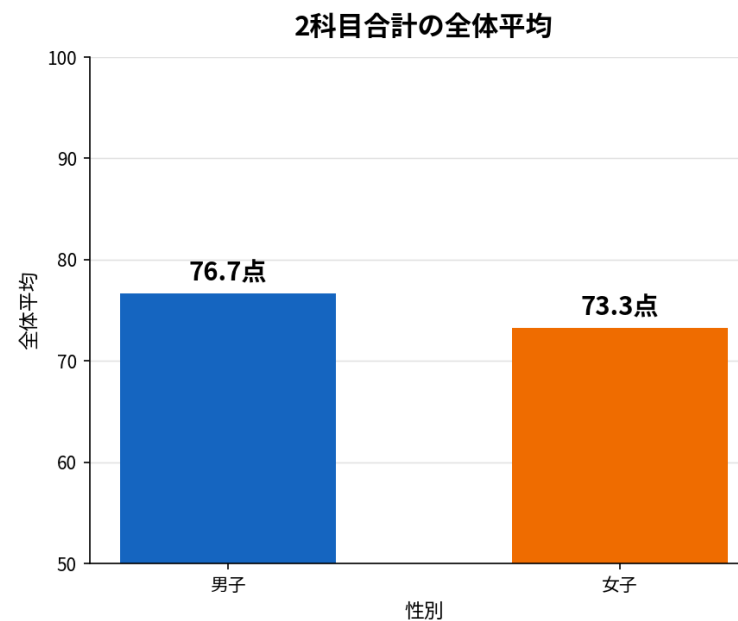


# 種明かし 全体平均は 男子の方が高い

## 科目別 女子が高い



## 全体平均 男子が高い!



英語は女子が少数派(10人)、数学は男子が少数派(10人)。人数の偏りが全体平均を歪める。  
男子 2科目合計 4600点 / 60人 = 76.7点 女子 2科目合計 4400点 / 60人 = 73.3点  
UCバークレーと同じ構造。シンプソンは 身近なところにも 潜んでいる。



## 追加事件3 アンスコムのおつ組

Q. 4つのデータセット。平均・分散・相関係数・回帰直線が すべて同じ。

4つのデータセット 統計量は すべてほぼ同じ

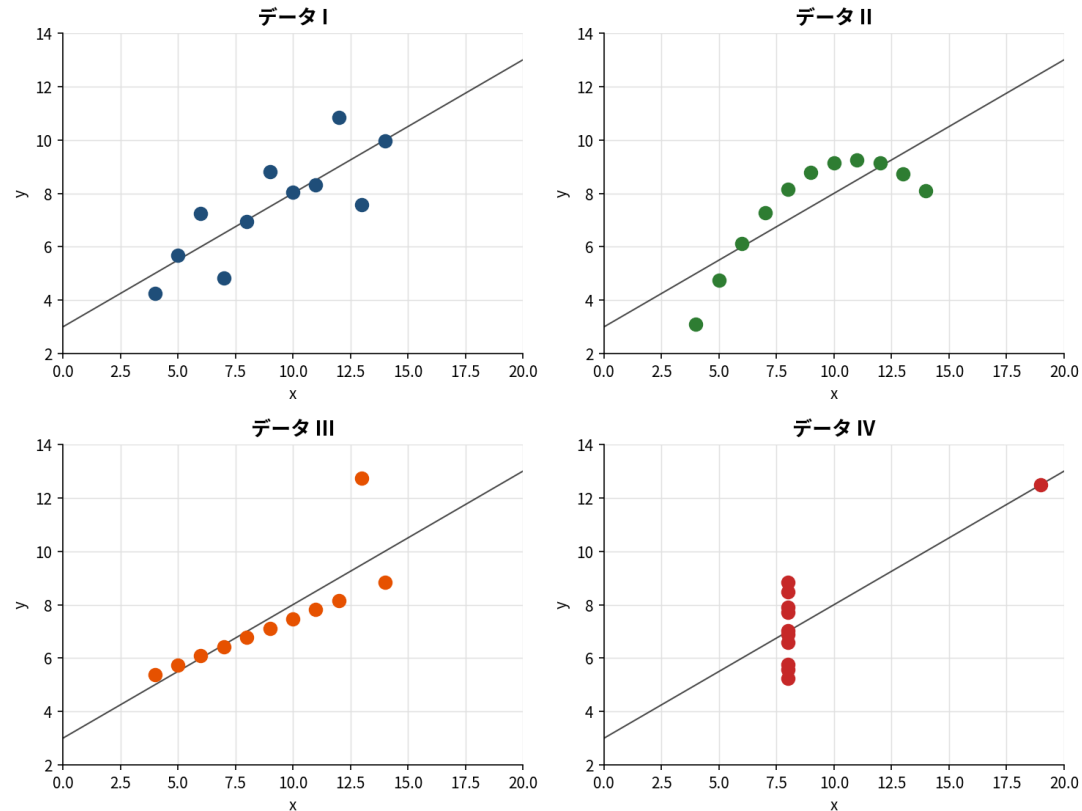
	データ I	データ II	データ III	データ IV
x の平均	9.0	9.0	9.0	9.0
y の平均	7.5	7.5	7.5	7.5
x の分散	11.0	11.0	11.0	11.0
y の分散	4.12	4.12	4.12	4.12
相関係数	0.816	0.816	0.816	0.817
回帰直線	$y=3+0.5x$	$y=3+0.5x$	$y=3+0.5x$	$y=3+0.5x$

どれも 似たような散布図になるはず？ 予想してみよう。



# 種明かし 散布図で見ると 全然違う

アンスコム の四つ組 — 数字は同じ、形は違う



同じ統計量でも 散布図の形は 全く違う。I=線形 II=曲線 III=外れ値 IV=垂直+外れ値。  
Anscombe(1973) 数字だけ見てデータを理解した気になるな、必ず目で見よ、という教え。



## 追加事件4 平均と中央値

Q. 日本の平均所得。あなたの実家は これ以上? これ以下?

日本の平均所得

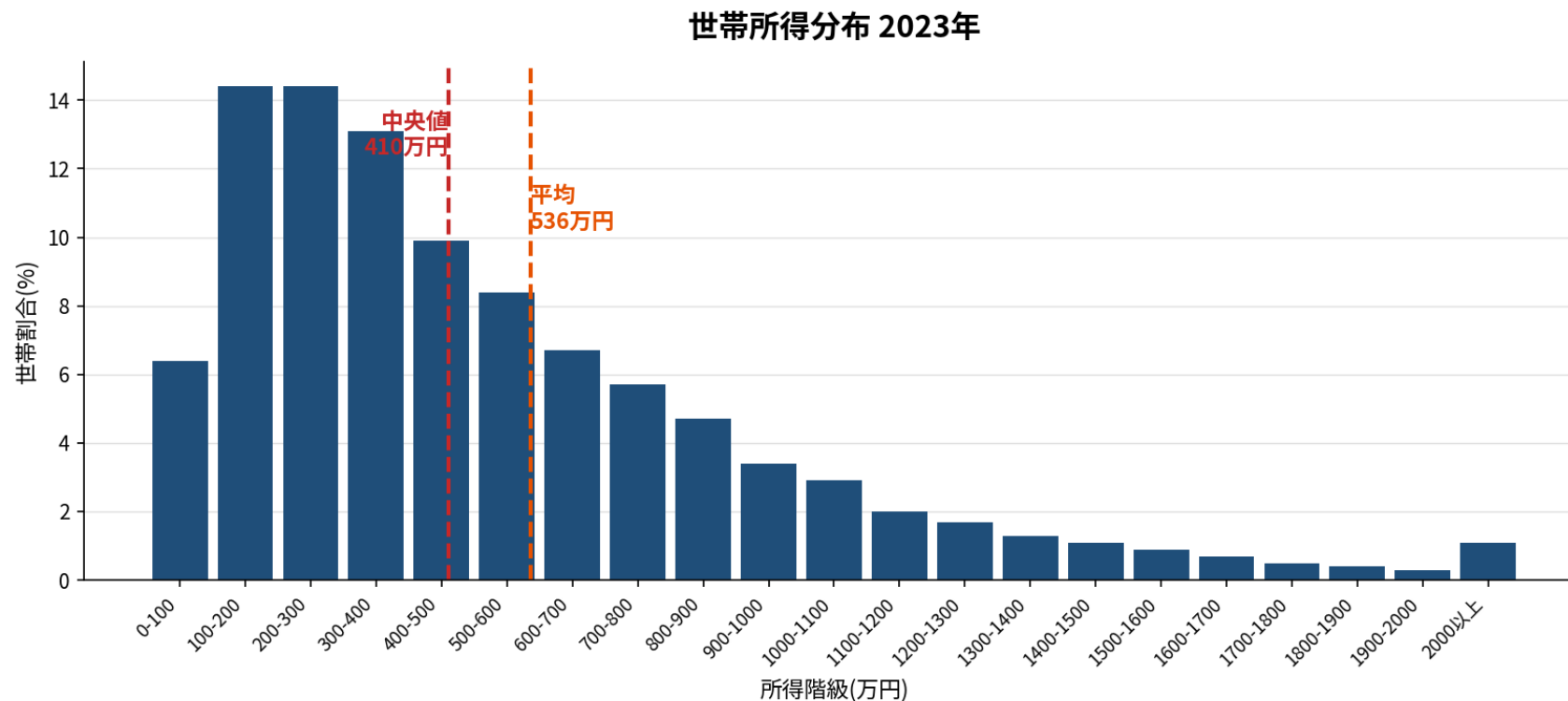
**536 万円**

(2023年 厚生労働省)

平均以下の世帯は 全体の 何%だと思えますか? 50%? 40%?



# 種明かし 平均以下が 61.9%

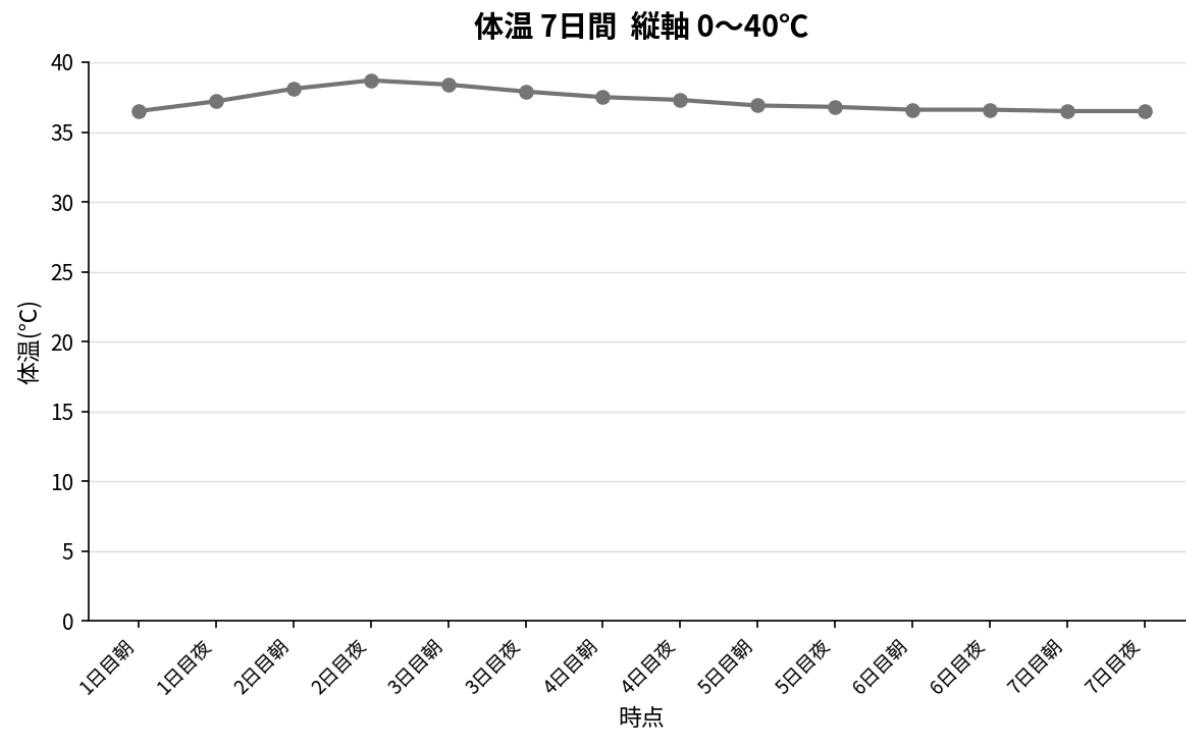


平均 536万円 中央値 410万円 差が126万円もある。  
平均以下の世帯は 61.9%。一部の高所得者が 平均を押し上げている。  
「平均年収」と聞いたら 中央値はいくらか も確認する。

出典 厚生労働省「2024年 国民生活基礎調査の概況」

# 追加事件5 体温の変化 見えますか？

Q. 風邪で7日間 体温を記録した。この人 発熱した？

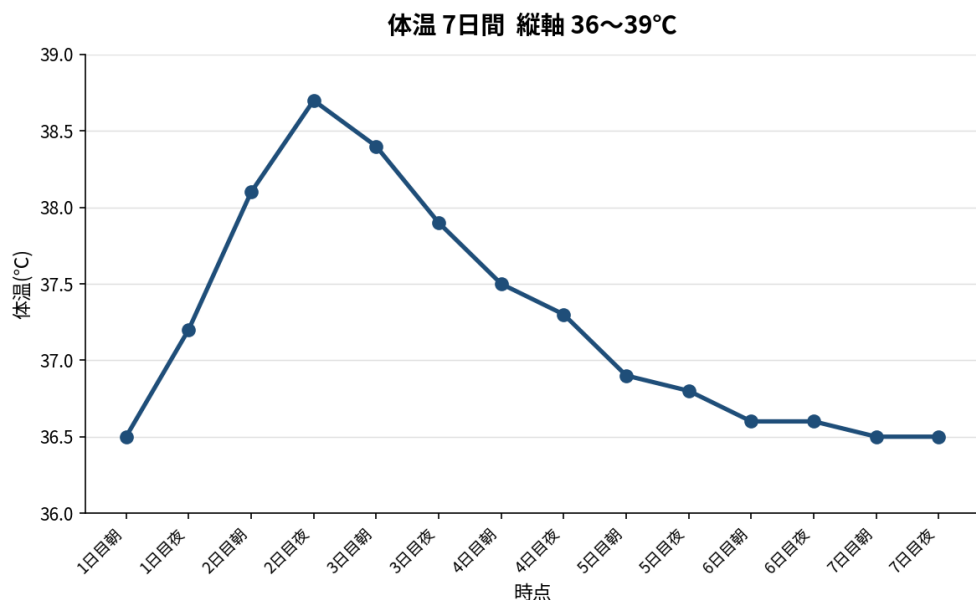


このグラフ 平らに見えるけど 本当に 変化なし？



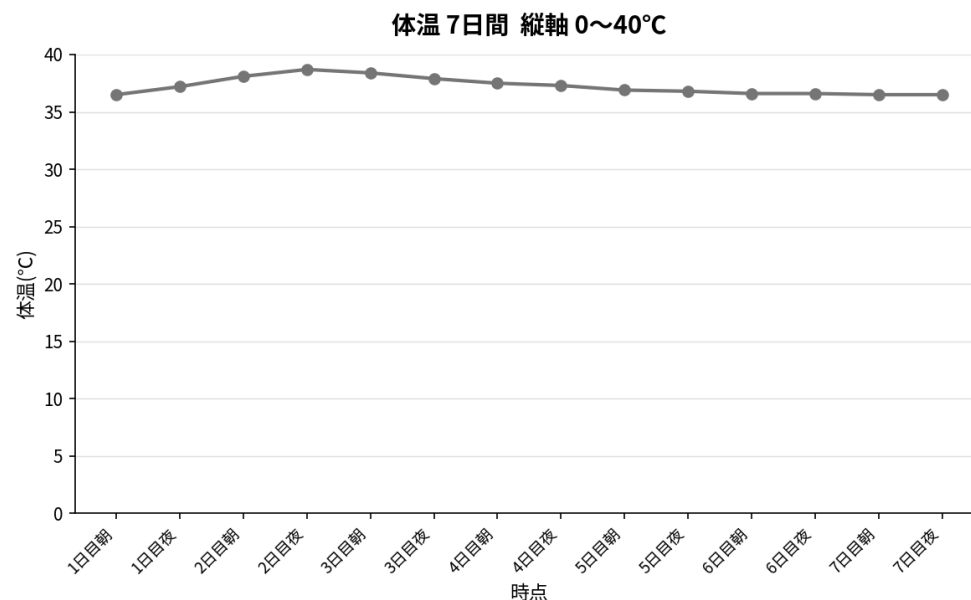
# 種明かし 縦軸を絞ると 発熱が見える

## 縦軸 36-39°C に絞る



発熱ピークがはっきり見える。2日目夜に38.7°C。解熱剤で下がり 5日目に平熱へ。

## 縦軸 0-40°C は広すぎ



事件1では「縦軸はゼロから」が原則だった。でも体温は 意味のある範囲で絞るのが正しい。

鍵は 縦軸。「ゼロから」が原則だが 目的によって変える。体温や pH は絞るのが正解。

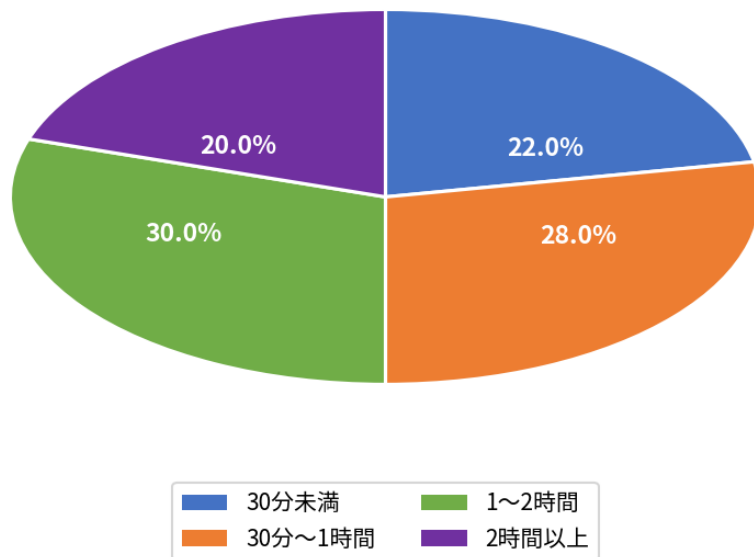
教材用仮想データ



## 追加事件6 3D円グラフの印象

Q. 学生アンケート「夜寝る前のスマホ利用時間」。3D円グラフで表示。

3D円グラフ ずるい版



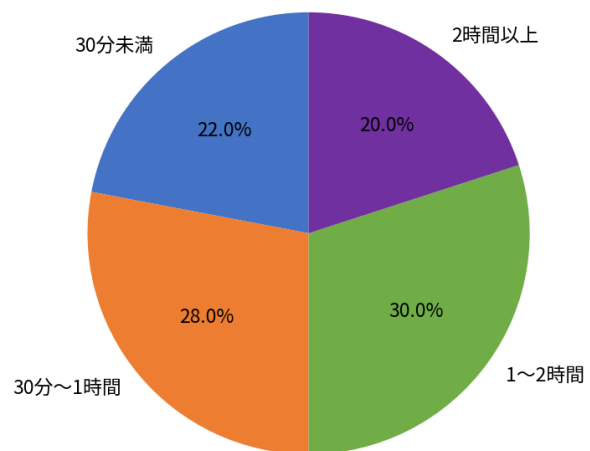
手前の部分と 奥の部分 どちらが 大きく見えますか？



# 種明かし 2Dで見ると 実は 4等分

## フェア 2D円グラフ

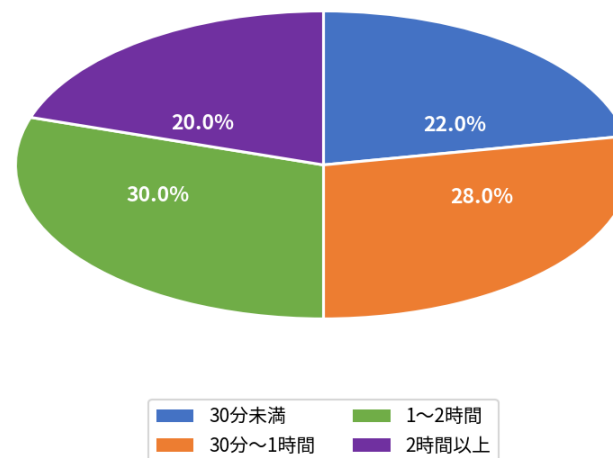
2D円グラフ フェア版



実際は 22% 28% 30% 20% で ほぼ4等分。  
2Dなら公平に見える。

## ずるい 3D円グラフ

3D円グラフ ずるい版



奥行きのおかげで 手前が大きく 奥が小さく見える錯覚が生じる。

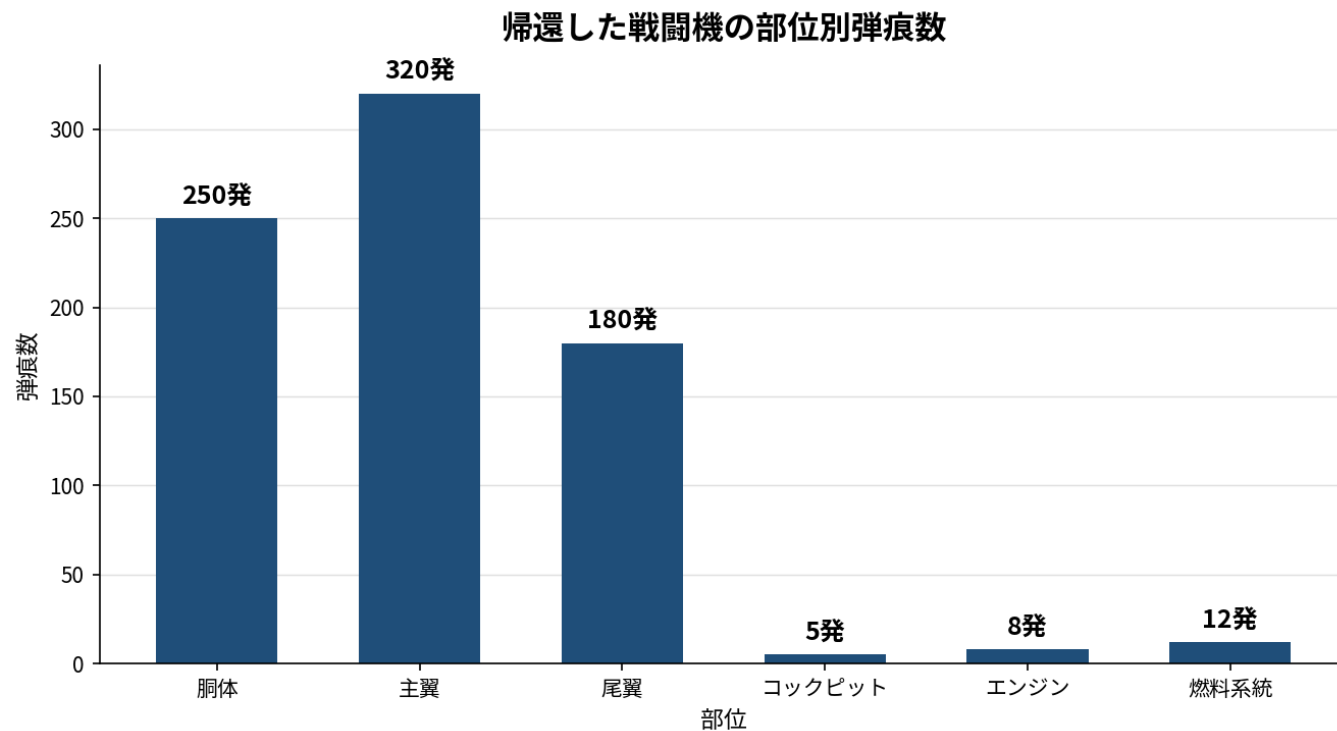
鍵は 見せ方。3Dは映えるが 正確さを犠牲にする。伝えたいのが正確さなら 2Dを選ぶ。

可視化の専門家は 円グラフ自体を避け 棒グラフを推奨する。



# 追加事件7 戦闘機を どこで補強する？

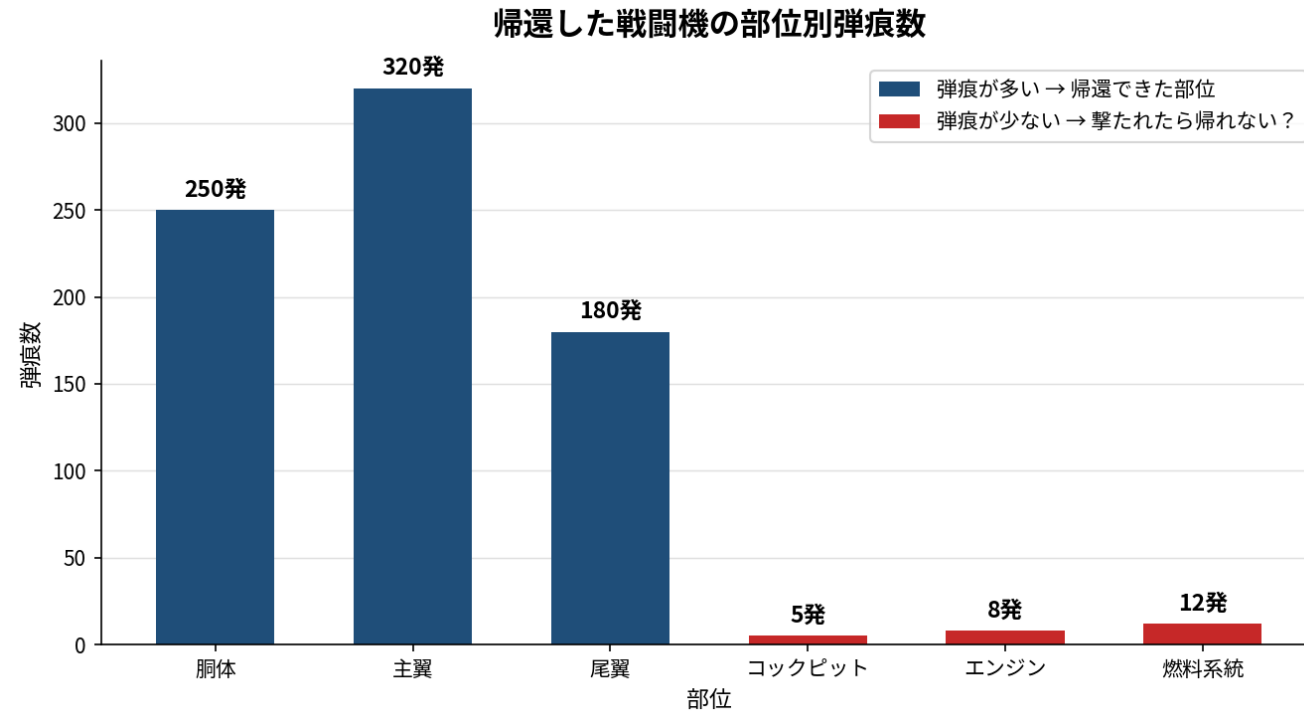
Q. 第二次大戦中 米軍は帰還した戦闘機の弾痕数を部位別に調べた。どこを補強すべき？



素直に考えれば 弾痕が多い場所？ 隣の人と 話し合ってみよう。



# 種明かし ウォルドの逆説



**統計学者ウォルドの答え「弾痕が少ない場所こそ補強すべき」**  
理由 撃たれて帰ってこられなかった機体はこのデータに含まれていない。  
コックピット エンジン 燃料系統 を撃たれた機体は 撃墜されて データに残らない。  
鍵は データに映っていないもの を考えること。

Abraham Wald (1943) に着想を得た話



# データを見る目 今日の総まとめ

- 1 軸と期間 を変えると 同じデータでも違う印象
- 2 実数と割合 件数と率 名目と実質 は 別物
- 3 二軸で重ねると 関係ありそうに見える 疑似相関に注意
- 4 全体と部分 で結論が逆になる シンプソンのパラドックス
- 5 数字だけ見ず 目で見る アンスコム の四つ組の教え
- 6 平均 と 中央値 は違う。外れ値に注意
- 7 データに映っていないもの を考える 生存者バイアス

これらの視点を持って 普段ニュースや広告で見るグラフを 読み解いてみてください。



# 実習 フェアなグラフとずるいグラフを作ってみよう

配布したExcelファイルから 好きな事件を1つ選んでください

## 手順

1. Excelのシートを見て データを確認する
2. 同じデータから フェアなグラフ を作ってみる
3. 同じデータから ずるいグラフ も作ってみる(軸 期間 累計 など进行操作)
4. どちらを企業が宣伝に使いそうか 想像してみる

提出は不要です。授業の最後に 気づいたことをGoogleフォームで答えてもらいます。

Excelの基本操作に戸惑ったら、近くの人に話しかけて互いに協力しよう。



# 振り返り Googleフォームに答えてください

振り返り用のGoogleフォームから回答してください。

## 問いかけ

- ・今日のクイズで 一番印象に残った事件は？
- ・驚いた話 新しく知った話は？
- ・今後 グラフを見るとき どこに気をつけますか？
- ・自分でグラフを作るとき どう気をつけますか？

率直な感想 気づきを書いてください。

次回 第3回は「相関と因果」。今日の疑似相関の続きをやります。



## 同じデータでも 見せ方で 印象は変わる

### グラフを見るとき3つの問い

1. 何が省略されているか 期間 軸 分母...
2. 代わりに何を見せたら 結論は変わるか
3. このグラフを作った人の 意図 は何か

グラフは武器にもなる 時に嘘にも使える道具。正しく使うリテラシー を身につけよう。

では、お疲れ様！ 次回は 相関と因果です。

