

データサイエンスへのいざない 第1回

稲垣 俊介



自己紹介

■「データサイエンスへのいざない」の担当です

データサイエンスへの皆さんを誘う（いざなう）のが私の役割です。

3回の基礎編を私が担当し、その続きを品田先生にご担当いただきます。

■山梨大学教育学部の教員です

将来の学校の教員養成、現在の学校の先生方のサポートをしています。

■以前は、東京都立高校の情報科の教員でした

高校で情報を17年間教えてきました。

■私の目標は

日本の情報教育を盛り上げること、それが私の役割です（この授業もその一つです）。



稲垣俊介

検索



今日の授業で持って帰って欲しいもの

- 何を1件としたか
- 何を分母にしたか
- 何をどう分けたか
- 何を混ぜて平均したか



前回のおさらい（？）

飛行機にはなぜファーストクラスがあるの？



日本からニューヨーク便



飛行機にはファーストクラスからエコノミークラスまで様々な座席があります。

引用：JAL Webサイト

考えてみてください



なぜ飛行機にはさまざまな座席があるのでしょうか？
航空会社や乗客の私たちにメリットはあるのでしょうか？
考えてみてください。



ヒント：価格



東京ーニューヨーク便の往復価格（時期で違います）

- エコノミークラス：約27.5万円
- ファーストクラス：約313.6万円



議論を深めるためにクイズ



ファーストクラスのあるおかげでエコノミーの値段は . . .

A 高くなっている / B 安くなっている



ヒント：占有面積



エコノミークラスとファーストクラスの占有面積

- エコノミークラス：約27.5万円 占有面積1
- ファーストクラス：約313.6万円 占有面積9.6



議論を深めるためにクイズ



ファーストクラスのあるおかげでエコノミーの値段は・・・

A 高くなっている / B 安くなっている



つまり・・・

なぜ飛行機にはさまざまな座席があるのでしょうか？
航空会社や乗客の私たちにメリットはあるのでしょうか？
改めて考えてください。



エコノミークラスとファーストクラスの占有面積

- エコノミークラス：約27.5万円（1） 占有面積1
- ファーストクラス：約313.6万円（11.4） 占有面積9.6



数字の裏側

前回の結論

ファーストクラスがあるから、エコノミーが安くなる

今回は・・・

今日の目標

数字の作り方で、結論の見え方が変わる



クイズ1 就職率98%ってどういう意味？

大学生の就職率 98.0%

どういう意味でしょうか？



クイズ1 就職率98%ってどういう意味？

大学生の就職率 98.0%

では、これはどちらでしょう？



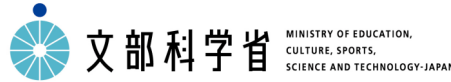
A 卒業生100人中98人が就職した

B 就職したい人100人中98人が就職した



クイズ1 就職率98%

報道発表



令和7年5月23日

大学等卒業生及び高等学校卒業生の就職状況調査結果を公表します

~大学生の就職率は98.0%~

~高校生の就職率は98.0%~

令和6年度大学・短期大学・高等専門学校及び専修学校卒業者の就職状況調査（4月1日現在）について

令和7年4月1日現在 文部科学省・厚生労働省調査

調査校は、大学62校(4,770人(国立大学1,450人、公立大学130人、私立大学3,190人))、短大20校(520人)、高専10校(400人)、専修学校20校(560人)の計112校(6,250人)である。なお、就職率とは、就職希望者に対する現時点での就職者の割合である。また、()内は前年度同期調査からの増減値(▲は減少)である。

[全体]

区 分	就職希望率	就職率	<参 考>
			前年度卒業学生の就職率 (R6.4.1現在)
大 学	75.6% (0.8)	98.0% (▲ 0.1)	98.1%
うち 国公立	53.4% (▲ 2.4)	97.6% (▲ 0.9)	98.5%
私 立	86.7% (2.5)	98.1% (0.2)	97.9%
短 期 大 学	82.5% (2.3)	97.0% (▲ 0.4)	97.4%
高 等 専 門 学 校	57.8% (0.0)	99.6% (▲ 0.4)	100.0%
計	75.0% (0.9)	98.0% (▲ 0.1)	98.1%

文部科学省：令和6年度大学等卒業予定者の就職状況調査（4月1日現在）について



答え1 分母が違う

文部科学省の公表では

■大学の就職希望率 75.6%

■大学の就職率 98.0%

就職率 = 就職希望者に対する就職者の割合

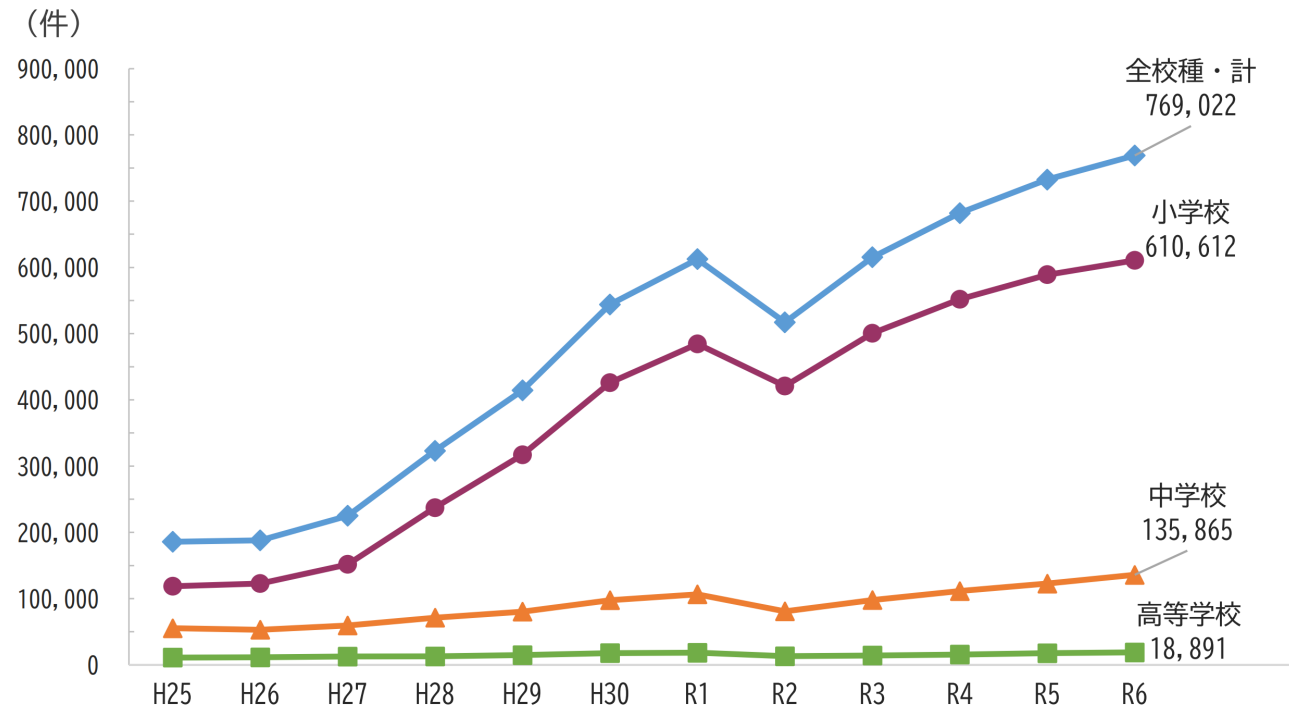
つまり、98%の分母は「卒業生全体」ではない。



クイズ2 いじめ件数が多い学校は、悪い学校？

いじめ件数が多い学校は、
悪い学校？

いじめの認知件数の推移



文部科学省：令和6年度児童生徒の問題行動・不登校等生徒指導上の諸課題に関する調査結果の概要

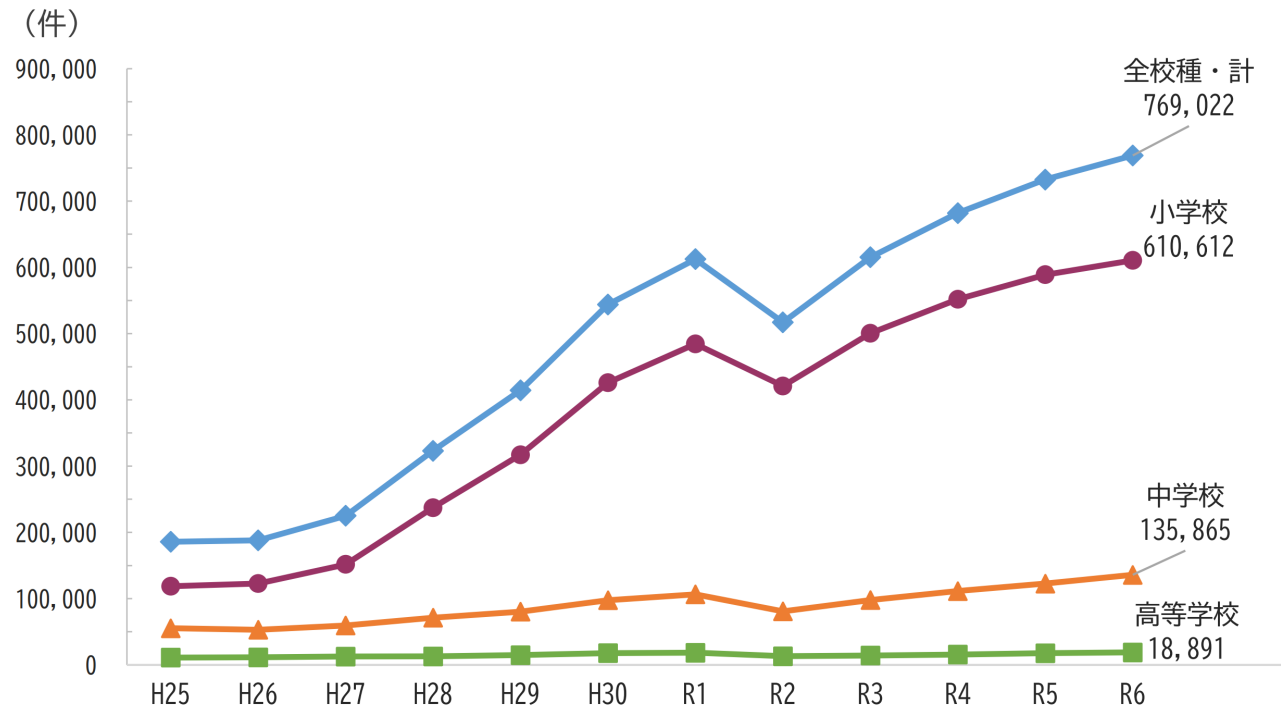
クイズ2 いじめ件数が多い学校は、悪い学校？

いじめ件数が多い学校は、
悪い学校？

A そう言える

B それだけでは言えない

いじめの認知件数の推移



文部科学省：令和6年度児童生徒の問題行動・不登校等生徒指導上の諸課題に関する調査結果の概要

答え 件数は「悪さ」だけでなく「見つけ方」も映す

文部科学省の公表では

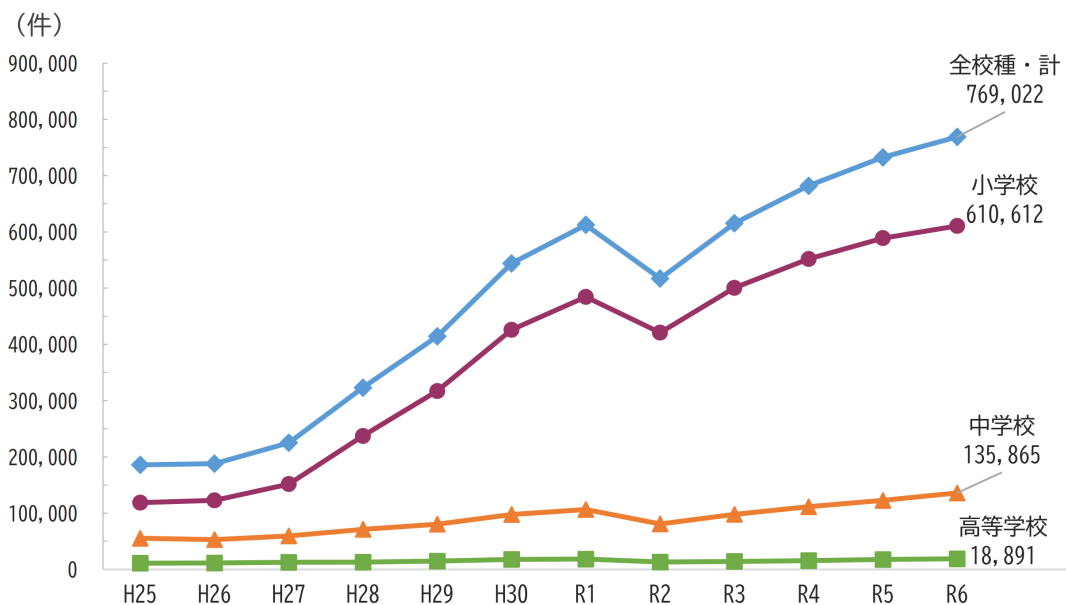
- 令和5年度のいじめ認知件数は 769,022件
- 過去最多

ただし、増加の背景として

- いじめの定義理解の広がり
- 積極的な認知
- アンケートや教育相談の充実なども挙げている。

さらに、認知件数0の学校には、潜在するいじめが多数ある懸念もある。

いじめの認知件数の推移



クイズ3 平均賃金が下がった。では、みんなの給料も下がった？

平均賃金が下がった。
では、みんなの給料も下がった？



クイズ3 平均賃金が下がった。では、みんなの給料も下がった？

平均賃金が下がった。
では、みんなの給料も下がった？

A そう言える

B そうとは限らない



答え3 平均は「顔ぶれ」で動く

例)

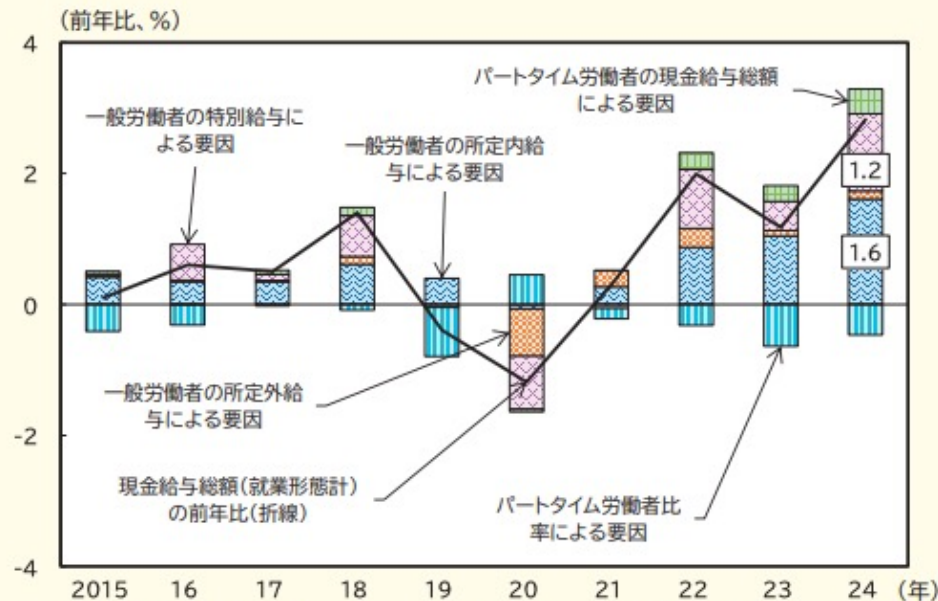
去年 一般 5人 × 40万円
パート 5人 × 10万円
→ 平均 25万円

今年 一般 5人 × 41万円
パート 10人 × 11万円
→ 平均 21万円

全員の金額は上がっても、
顔ぶれが変わると
全体平均は下がることもある

第1-(3)-9図 現金給与総額の伸び率の寄与度分解の推移

○ 2024年の就業形態計の現金給与総額は、パートタイム労働者比率の増加がマイナス寄与。



資料出所 厚生労働省「毎月勤労統計調査」をもとに厚生労働省政策統括官付政策統括室にて作成

(注) 1) 調査産業計、事業所規模5人以上の値。

2) 就業形態計、一般労働者、パートタイム労働者について、指数にそれぞれの基準数値(2020年)を乗じ、100で除して実数値を算出し、これらの数値を基にパートタイム労働者比率を集計。

3) 2024年の寄与度の算出においては、2023年のデータとして、「令和6年の賃金・労働時間等の前年比の算定に用いる参考値」を用いている。

4) 所定外給与=定期給与(修正実数値)-所定内給与(修正実数値)、特別給与=現金給与総額(修正実数値)-定期給与(修正実数値)として算出。このため、毎月勤労統計調査の公表値の増減とは一致しない場合あり。

厚生労働省「令和7年版 労働経済の分析」

クイズ4 死亡率が上がった

「死亡率が上がった」

では、その社会は前より不健康になったのでしょうか？



クイズ4 死亡率が上がった

「死亡率が上がった」

では、その社会は前より不健康になったと言ってよいのでしょうか。

A そう言える

B そうとは限らない



答え4 死亡率が上がった。だから「前より不健康」とは限らない

図1 粗死亡率の年次推移

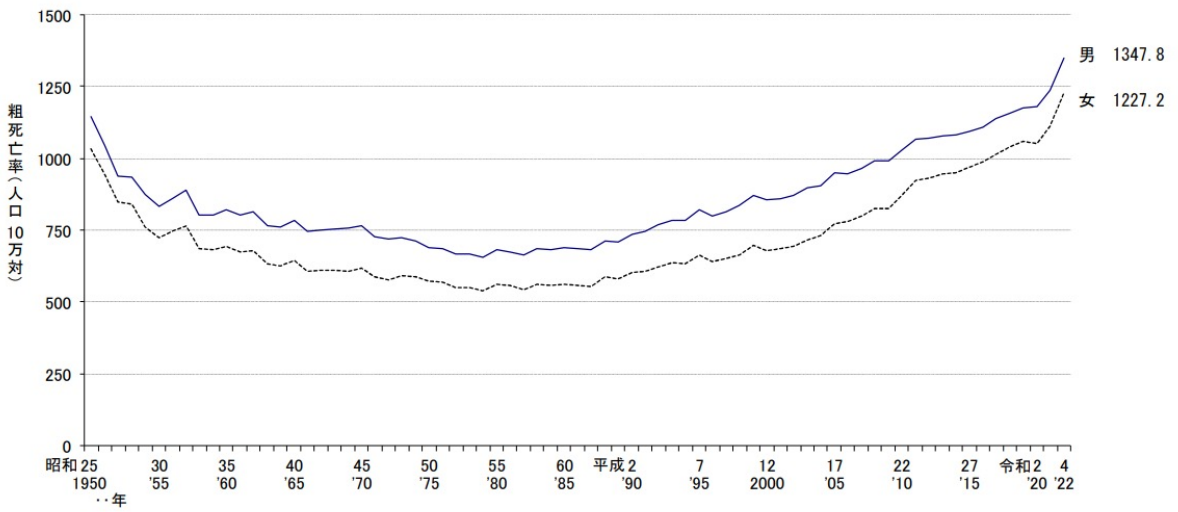
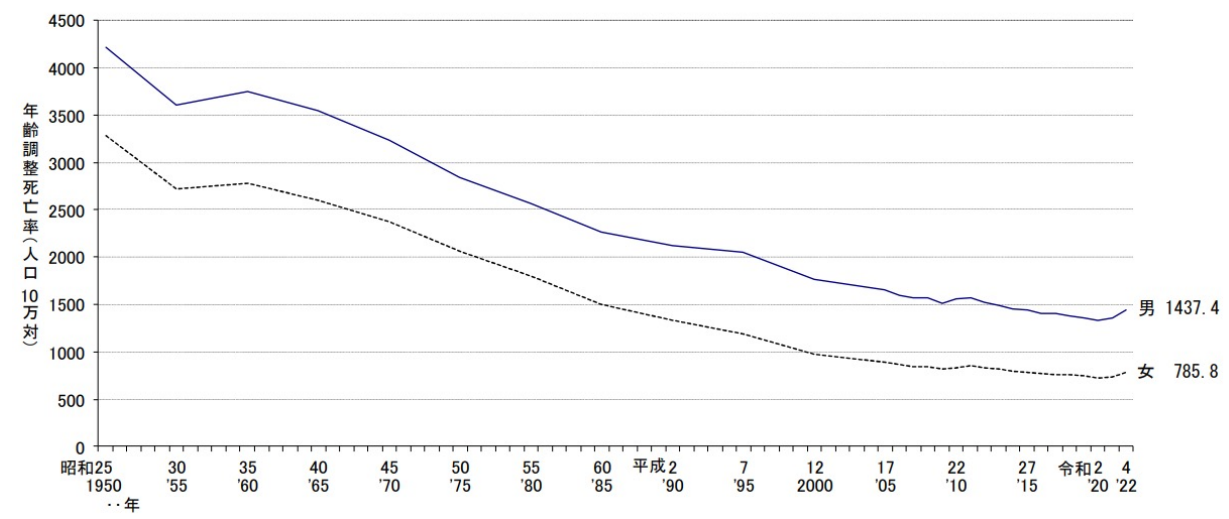


図2 年齢調整死亡率の年次推移



粗死亡率
= その年の社会全体で、人口10万人あたり何人亡くなったか

年齢調整死亡率
= 年齢の違いをならして比べた死亡率

「混ぜたまま見ると、理由を見誤る」

答え4 混ぜると見えなくなる

厚生労働省は

「年齢調整死亡率が低下しているのに、粗死亡率が上昇しているのは高齢化の影響」と説明している。

高齢者の割合が増えると、全体の死亡率は上がりやすい。



今日の4つのポイント

- 1件
- 分母
- 顔ぶれ
- 分け方

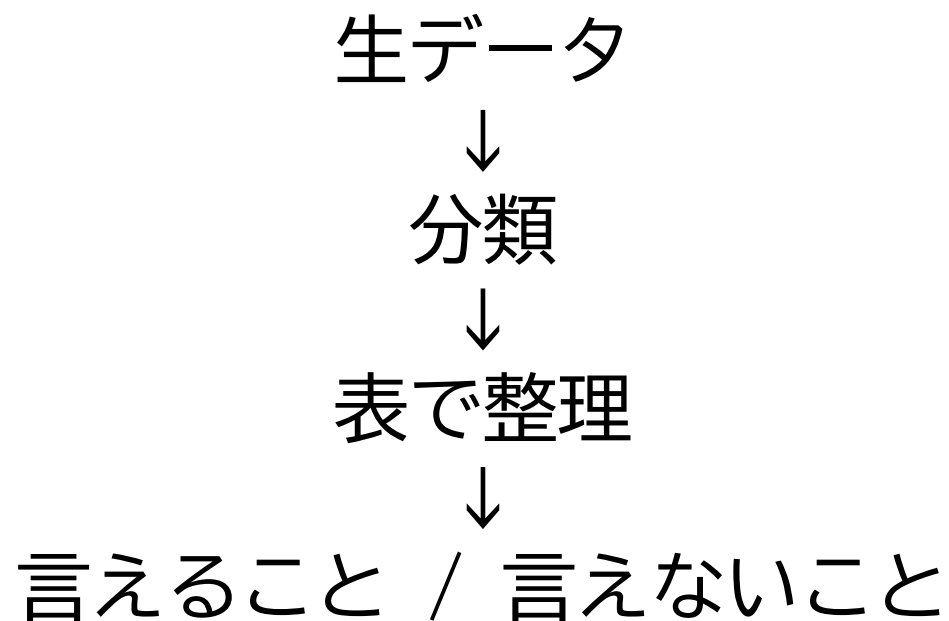
この4つで、数字の意味は大きく変わる



用意したデータで練習します

今日の主役は、操作ではありません（時間があれば練習します）

今日の主役は「どうやってデータにするか」です



今日の問い

このクラス（みなさん）は、
「朝がつらい」クラスと言えるでしょうか？

では、何を聞けばよいでしょうか？



こんな項目を文京学院大の皆さんに聞いたとします

- 通学時間（分）
- 通学手段
- 昨夜の睡眠時間（時間）
- 23時以降のスマホ利用時間（分）
- 今の眠気（1～5）
- 朝食（あり / なし）
- 自分は朝に強いと思うか（1～5）



行について

ID	通学時間	通学手段	睡眠時間	23時以降スマホ	眠気	朝食	朝の強さ
1	15	徒歩	7	20	2	あり	4
3	25	徒歩	5.5	70	4	なし	2
10	58	電車	4.5	110	5	なし	1
12	65	電車	6	50	4	あり	2
17	90	電車	7	30	3	あり	4
19	100	電車	8	20	2	あり	3

1行 = データ1件 = 学生1人

列について

ID	通学時間	通学手段	睡眠時間	23時以降スマホ	眠気	朝食	朝の強さ
1	15	徒歩	7	20	2	あり	4
3	25	徒歩	5.5	70	4	なし	2
10	58	電車	4.5	110	5	なし	1
12	65	電車	6	50	4	あり	2
17	90	電車	7	20	2	あり	4
19	100	電車	7	20	2	あり	4

1列は同じ種類の情報だけ



まだ見にくい「生データ」

ID	通学時間	通学手段	睡眠時間	23時以降スマホ	眠気	朝食	朝の強さ
1	15	徒歩	7	20	2	あり	4
3	25	徒歩	5.5	70	4	なし	2
10	58	電車	4.5	110	5	なし	1
12	65	電車	6	50	4	あり	2
17	90	電車	7	30	2	あり	4
19	100	バス	5	85	5	なし	1



「朝がつらい」を、あとで確かめられる形に変える

「朝がつらい」は、そのままだと曖昧です

だから、こう置きかえます

- ・ 眠気4~5 → 眠い
- ・ 通学60分以上 → 長い
- ・ 睡眠6時間未満 → 短い
- ・ 朝食 → あり / なし

これが「データ化 + 分類」です



まずは4つの数字で、全体の輪郭を見る

回答者 20人

- 通学60分以上 10人 (50%)
- 睡眠6時間未満 8人 (40%)
- 眠気4~5 10人 (50%)
- 朝食なし 7人 (31%)



表にすると、「関係」が見えはじめる

通学時間 × 眠気

	眠くない (1~3)	眠い (4~5)	合計
60分未満	7	3	10
60分以上	3	7	10
合計	10	10	20

眠い人の割合

60分未満 30%
60分以上 70%



でも、どこで切るかで印象が変わる

通学時間 × 眠気

	眠くない (1~3)	眠い (4~5)	合計
45分未満	4	2	6
45分以上	6	8	14
合計	10	10	20

眠い人の割合

45分未満 33%

45分以上 57%



人数だけ見ると間違える。割合を見る

朝食 × 眠気

	眠くない (1~3)	眠い (4~5)	合計
朝食あり	9	4	13
朝食なし	1	6	7
合計	10	10	20

眠い人の割合

朝食あり $4/13 = 33\%$

朝食なし $6/7 = 86\%$



このデータから言えること / まだ言えないこと

言えること

-
-
-

まだ言えないこと

-
-
-



このデータから言えること / まだ言えないこと

言えること

- この20人では、通学が長いグループの方が眠い人が多い
- この20人では、朝食なしのグループの方が眠い人の割合が高い
- 同じデータでも、分け方で印象が変わる

まだ言えないこと

- 通学時間が長いから眠くなる
- 朝食を抜くことが原因だ
- 文京学院大学の学生全体も同じだ



数字は、そのまま現実ではありません

私たちが・・・

- 何を1件とするか
- 何を分母にするか
- 何をどう分けるか
- どんな顔ぶれか

で、意味が変わります

