

2019年度春学期 統計学 第3回
クロス集計とデータの可視化

浅野 晃
関西大学総合情報学部



データの種類～尺度水準～🤔

データは数字だとは言っても

数字は、必ずしも「数量」を表しているとは限りません

例えば、選択肢の

「1番・2番・3番」は、
「a・b・c」でも
「イ・ロ・ハ」でも

同じだから、数「量」ではない

尺度水準

統計学では、数字を「数量」としての意味をどのくらい持っているかで
4つのレベルに分けている 尺度水準

比例尺度	量的データ
間隔尺度	足し算引き算ができる
順序尺度	質的データ
名義尺度	足し算引き算ができない

名義尺度

選択肢を区別するための、単なる記号。

1番・2番・3番 さあどれ？
男性：1 女性：2

2番が1番より大きいという
意味はない

順序尺度

数字の順番にのみ意味がある

この講義に満足しましたか？
1) 非常に不満・2) 不満・
3) 満足・4) 非常に満足

2番は1番より満足度が大きい

「2番と1番の満足度の差」
「3番と2番の満足度の差」は、同じではない

間隔尺度

数値の間隔にも意味がある

摂氏温度 (20°C, -10°C) 

「0°Cと10°Cの温度の差」
「-10°Cと0°Cの温度の差」
は同じ

20°Cは10°Cの2倍暖かい？
そんなことは
はない

20°Cは-10°Cの何倍暖かい？
???



2018年2月28日
フィンランド・タンペレ市



2019年2月4日
札幌市北区



※気温が+か-かは、
雪が融けるか融けないかなので、きわめて重要

2018年2月13日

比例尺度

間隔だけでなく比率にも意味がある

長さ・重さ・年齢など

40歳の方は、20歳の方の**2倍**の
年数を生きている。

マイナスの値は存在しない
(温度なら、絶対温度がこれにあたる)

簡単に平均というけれど

平均できるのは、足し算ができる
量的データ(間隔尺度・比例尺度)だけ

この講義に満足しましたか?

- 1) 非常に不満
- 2) 不満
- 3) 満足
- 4) 非常に満足

こういうのの平均は、本当は意味がない
(間隔尺度だと、近似的に考えている)

クロス集計🤔

質的データの解析について

次回以降は、平均を計算できるデータ
=量的データ を扱います

今日は、質的データを扱う
クロス集計について

クロス集計

例：商品Aが好きか嫌いか
→好きな人：60%，嫌いな人：40%

これだけでは大したことはわからない

回答者が男性か女性かも記録しておく

クロス集計

	好き	嫌い	合計
男性	25	25	50
女性	35	15	50
合計	60	40	100

差がある
のは女性

これが [クロス集計]

ひとつのデータ群を2つの項目から
見て、項目間の関係を表す

検査の感度

新しい検査法をテスト

	本当に病気である	本当は病気ではない
検査で陽性	A	B
検査で陰性	C	D
合計	A + C	B + D

$$A/(A+C) \text{ 感度}$$

感度が高ければよいというわけではない
病気であってもなくても「陽性」と答えるなら、 $C=0$ で感度100%

検査の特異度

	本当に病気である	本当は病気ではない
検査で陽性	A	B
検査で陰性	C	D
合計	A + C	B + D

$$D/(B+D) \text{ 特異度}$$

特異度：

病気でない人のうち、正しく「陰性」となる人の割合

病気であってもなくても「陰性」と答えるなら、 $B=0$ で特異度100%

感度と特異度

	本当に病気である	本当は病気ではない
検査で陽性	A	B
検査で陰性	C	D
合計	A + C	B + D

感度・特異度の両方を同時に
100%近くにするのはむずかしい

感度が90%のとき、特異度は…
という言い方で、検査の能力を表す

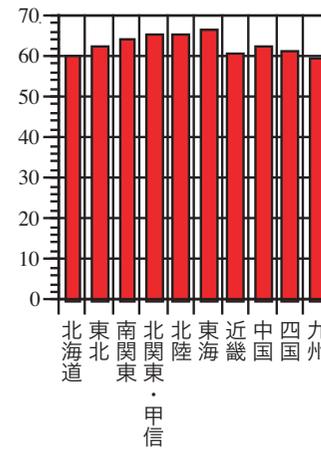
データの可視化 

データの可視化

人は、
数字の羅列をざーっと見て
即座に意味が理解できるほど
賢くはない

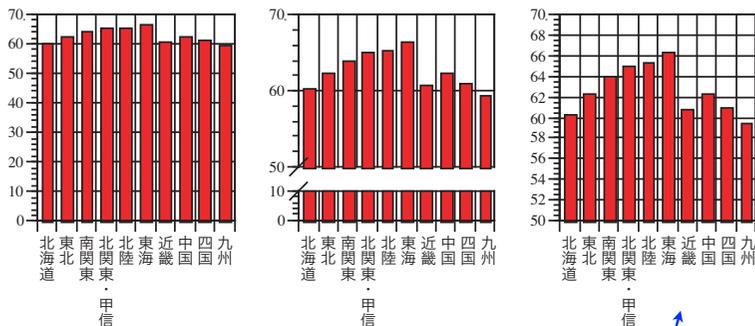
グラフなどの形に「描いて」
理解しやすくする

棒グラフ



横軸は名義尺度
でもよい

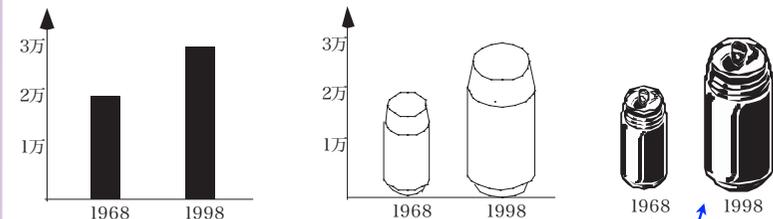
差が際立って見えるのはどれ



棒の長さが値に比例していない

言い訳すらしていない (ズル)

こんな描き方はない？



長さが2倍なら
面積は4倍
体積は8倍

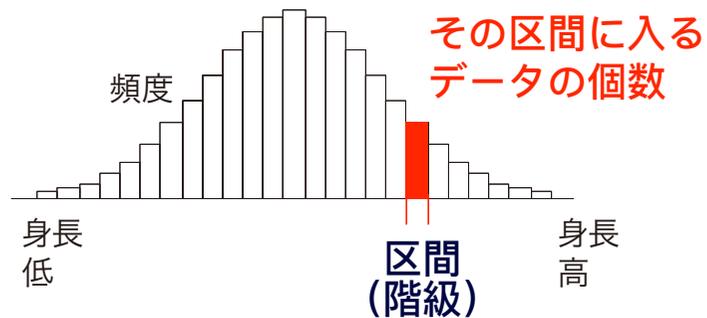
高さで量を表すはずなのに、
棒の幅や厚み感も変えて、
面積・体積で表しているかの
ように印象づけている

縦軸がないから、体積で量を表
しているように見える (ズル)

ヒストグラム

棒グラフではありません

こんなやつ 間隔尺度をもつデータに



次回重点的にやります

折れ線グラフ

(図は教室で)

右端のほう縦軸が長くなっている (ズル)

5月12日は

国際看護師の日・看護の日
フローレンス・ナイチンゲールの
誕生日

「統計学の先駆者」

ナイチンゲールのグラフ

図は教室で

出典

<http://upload.wikimedia.org/wikipedia/commons/1/17/Nightingale-mortality.jpg>