

1 はじめに

前回の復習

- ・ 度数分布表、度数分布図
- ・ 基本統計量（代表値）：平均値、中央値、最頻値
- ・ Excel：複数の条件に合うセルの個数を数える方法
 - =countifs（範囲, 条件1, 範囲, 条件2）
- ・ Excel：グラフの作成方法（方法1）
 - ① データを選択
 - ※ もし、データが離れたセルであった場合は
最初の範囲を選択 → （Ctrl キーを押しながら）次の範囲を選択
 - ② グラフ種別を選択
 - ◇ 挿入 → 縦棒 → 2D縦棒（一番左のアイコン）
 - ③ グラフをお化粧（タイトル、縦軸ラベル、横軸ラベルなど）

・ Excel : グラフの作成方法 (方法 2)

① グラフ種別の選択

◇ 挿入 → 縦棒 → 2D 縦棒 (一番左のアイコン)

② データを選択

◇ グラフエリアで右クリック → データの選択

◇ 縦軸データを選択する。具体的には次のとおり。

「凡例項目」 → 追加

→ 系列名の右のアイコン → 「度数 (日)」のセルを選択 → Enter

→ 系列値の右のアイコン → データ {3, 8, . . . 1} のセルを選択 → Enter

→ OK

◇ 横軸データを選択する。具体的には次のとおり。

「横 (項目) 軸ラベル」の編集

→ 階級 (万円) のデータ {20 ~ 30, 30 ~ 40, . . . , 80 ~ 90} のセルを選択

→ Enter → OK

③ グラフをお化粧 (タイトル、縦軸ラベル、横軸ラベルなど)

2 平均値か中央値か？

演習1 以下について、平均値と中間値はどのような関係にあると予想しますか？

- ア 平均値と中央値はほぼ同じ
- イ 平均値 $>$ 中央値
- ウ 平均値 $<$ 中央値

(1) 身長

(2) 日本の世帯所得（日本の各世帯の収入額）

3 階級の数はいくつにすべきか？

- データの個数が大きい \Rightarrow 階級の数を大きく
- データの個数が少ない \Rightarrow 階級の数を小さく
- 「データの個数に対し望ましい階級の数」

データの個数	8	16	32	64	128	256	512	1024	2048
階級数	4	5	6	7	8	9	10	11	12

※「スタージェスの公式」という公式で求めたもの

- あくまでも目安。（この表に従う必要はない）

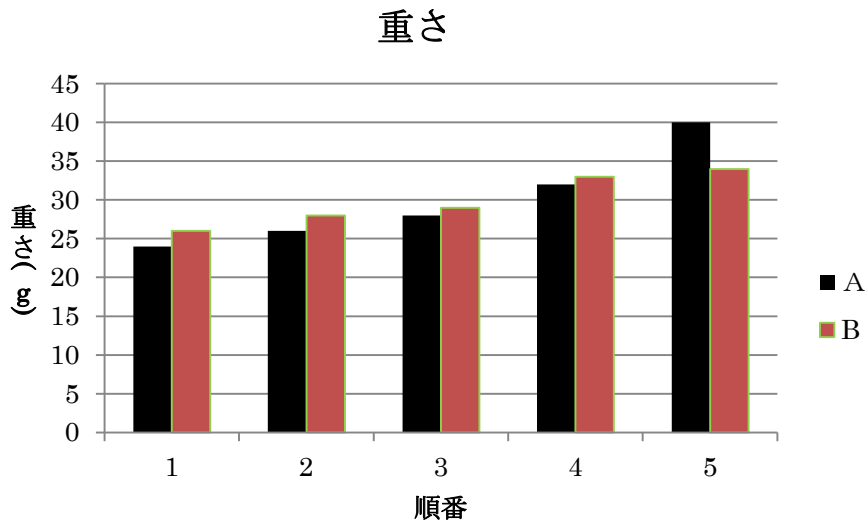
4 分散、標準偏差

- データの「バラツキ度合い」を表す指標です。

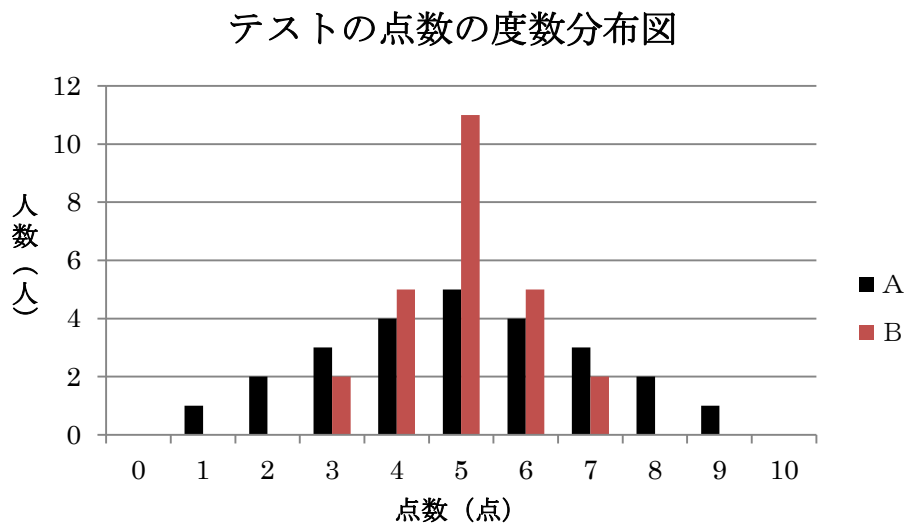
演習 2 グループAとグループBでは、どちらが重さのバラツキが大きいですか？

A : 24g, 26g, 28g, 32g, 40g

B : 26g, 28g, 29g, 33g, 34g



演習 3 クラスA (25人)とクラスB (25人) では、どちらが点数のバラつきが大きいですか？



※AとBの比較がしやすいように、棒の間隔を空けている

演習4 例にならって言い換えて下さい。

(例) ネジの長さのバラツキ (分散) が大きい

➤ 長いネジもあれば短いネジもある

(1) メロンパンは重さのバラツキ (分散) が大きい。

(2) テストの点数のバラツキ (分散) が大きい。

(3) 毎日の売上個数のバラツキ (分散) が大きい。

演習5 次の文の正誤を答えて下さい。

(1) メロンパンの重さの分散は大きい方が、販売するのに都合がよい。

(2) テストの点数の分散は大きい方が、教える側は楽である。

(3) 毎日のメロンパンの売上個数の分散は大きい方が、生産量を決めるのに都合がよい。

偏差、平均偏差、分散、標準偏差の計算方法

- ・ 偏差 = (各データ値) - (平均値)
- ・ 平均偏差 = $\frac{|偏差| + |偏差| + \dots + |偏差|}{データの個数}$

- ◇ |偏差|とは、偏差の絶対値のこと。
- ◇ Excel 関数で、絶対値は、=abs (数値)
- ◇ Excel 関数で、平均偏差は、=avedev(範囲)

- ・ 分散 = $\frac{(偏差)^2 + (偏差)^2 + \dots + (偏差)^2}{データの個数}$

- ◇ Excel 関数で、2乗は、^2 と入力する。 (例) 5 の 2乗は、5^2
- ◇ Excel 関数で、分散は、=varp(範囲)
(注) 最新の excel では=var.p(範囲)

- ・ 標準偏差 = $\sqrt{\text{分散}}$

- ◇ Excel 関数では、ルート (平方根) は、=sqrt (数値)
- ◇ Excel 関数で標準偏差は、=stdevp(範囲)
(注) 最新の excel では=stdev.p (範囲)

バラツキの指標として、どれを用いたらよい?

- 平均偏差 : ×
- 分散 : ○
- 標準偏差 : ◎ (単位がデータと同じ)

演習 6 あなたは学園祭で「かまぼこ切り競争」を企画しました。参加者に 150 g のかまぼこを 5 つに切ってもらい、その重さのバラツキが小さいほど勝ちという競争です。下の表の空欄を埋め、各人の「かまぼこの重さの標準偏差」を求めてください

表 Aさんのかまぼこの重さ(g)

	1	2	3	4	5	合計	平均値	
データ	24	26	28	32	40			←平均値
偏差								
偏差								←平均偏差
(偏差) ²								←分散

$$\text{標準偏差} = \sqrt{\text{分散}}$$

表 Bさんのかまぼこの重さ(g)

	1	2	3	4	5	合計	平均値	
データ	26	28	29	33	34			←平均値
偏差								
偏差								←平均偏差
(偏差) ²								←分散

$$\text{標準偏差} = \sqrt{\text{分散}}$$

5 Excel の豆知識

- ・ シートのコピーを作成する
 - (下部のシート名を右クリック) → 「移動又はコピー」 → 「コピーを作成する」にチェック

- ・ 3 を 3.0 と表示する

- ホーム → セルの選択 → 

【演習 1】（解答例）

(1) ア

(2) イ

- ・ 所得金額階級別にみた世帯数の相対度数分布図（2013 年度）
（出典：国民生活基礎調査（厚生労働省））は以下のとおりです。

➤ 平均値=538 万円

◇ 平均値を下回る世帯の割合=67.9%

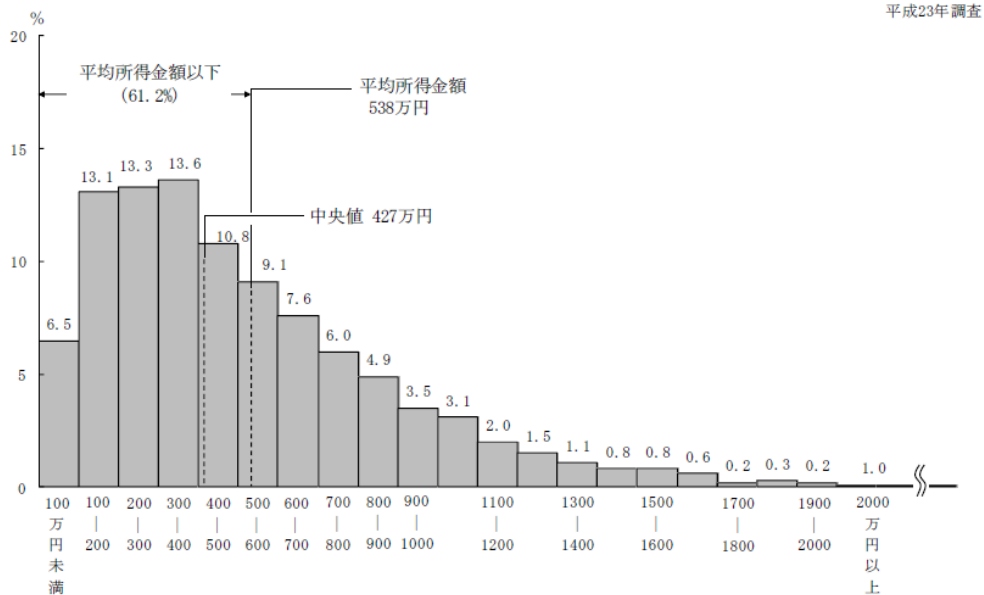
➤ 中央値=427 万円（平均値よりも 100 万円も低い！）

➤ 最も度数の大きい階級=200 万円以上 300 万円未満

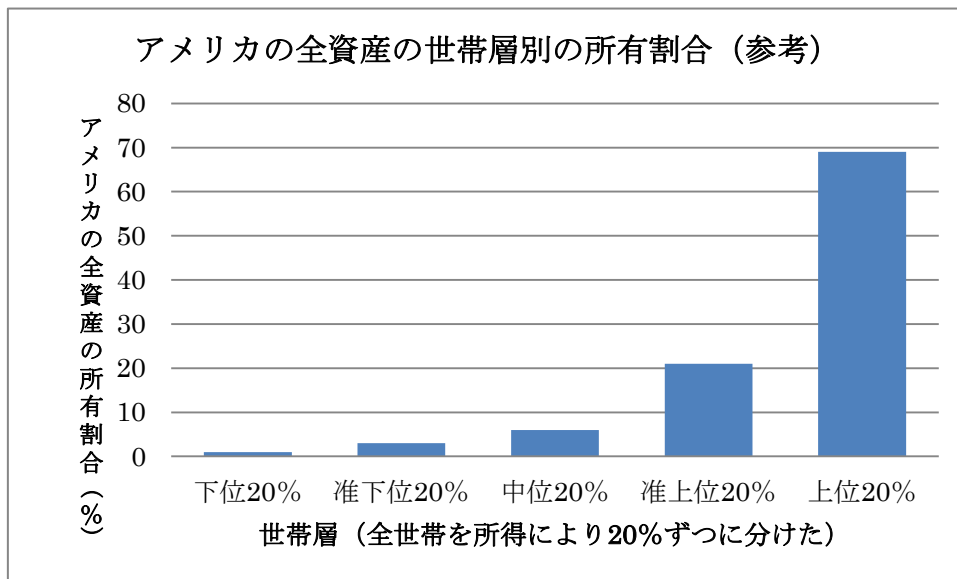
- ・ 平均値は、超高額所得者の所得の影響を受けやすいです。
すなわち、中央値を用いる方がよいです。

➤ 平均値は、外れ値（他の値より著しく異なる値）に大きく影響され、無意味な値となることがあります。一方、中央値は外れ値にほとんど影響されません。

図11 所得金額階級別にみた世帯数の相対度数分布



(参考)



【演習2】(解答例)

A

【演習3】(解答例)

B

【演習4】(解答例)

(1) 重いメロンパンもあれば軽いメロンパンもある

⇒同じ重さになるように努力すべき

(2) よい点数の人もいれば悪い点数の人もいる

⇒教え方に工夫が必要

(3) たくさん売れる日もあればあまり売れない日もある

⇒売れ残りや品切れが出やすいので、売れる量の予測をしっかりと行う必要がある。

【演習5】(解答例)

(1) ×

(2) ×

(3) ×