

## 03 記述統計（集計・可視化・分布の把握）

---

## 1. 第2回：記述統計 — データの全体像をつかむ

前回は、1人で測ったポテトの重さを使って、「同じ商品でも重さが毎回違う」ことを確認しました。

今回は、

- ・ データの数を増やし
- ・ 全体をまとめて見て
- ・ その特徴を整理する

ことで、**データの全体像を読む**ことを目指します。

このように、手元にあるデータを整理・要約する考え方を**記述統計**と呼びます。

## 2. 前回の測定で残った疑問

前回は、「私」1人がポテトの重さを測りました。

その結果、

- ・ 平均はだいたい決まっていそう
- ・ でも、軽い日・重い日がある

ことが分かりました。

ここで、1つ疑問が残ります。

このばらつきは、

- ・ たまたまなのか？
- ・ 「私」だけの測定結果なのか？

### 3. 調査チームを作って確かめることにした

そこで今回は、**クラス 30 人全員**に声をかけて、調査チームを作ることにしました。

#### 調査のルール

- ・ 同じ店のポテトを使う
- ・ 1週間、それぞれが測定する
- ・ 重さを記録し、気づいたことをメモする

こうして集まったのが、**30 人分のポテト重量データ**です。

## 4. 今日やること：集まったデータをどう見るか

今日は、この 30 人分のデータを使って、次のことを順番に行います。

1. 数でまとめてみる（平均・最小・最大）
2. グラフで形を見してみる
3. 「なぜこうなったか」を考える

まずは、**全体を 1 つの数で見たらどうなるか**を確かめてみましょう。

→ 演習①へ

## 5. 演習①：ポテト重量の基本統計量を求める

クラス 30 人が 1 週間調査したポテト重量データ（poteto30.xlsx）を用いて、全体の数値的な特徴を確認します。

### 対象とする列

- ・ 重さ\_g（ポテトの重量）

### 作業内容

1. AVERAGE 関数で **平均値** を求める
2. MEDIAN 関数で **中央値** を求める
3. MIN, MAX 関数で **最小値・最大値** を求める

### 確認するポイント

- ・ 平均値と中央値は近いのか、離れているのか
- ・ 最小値・最大値は、平均からどの程度離れているのか

## 6.【問い】 平均値だけで、この店のポテトを評価できるか？

演習①で、ポテト重量の 平均値・中央値・最小値・最大値 を求めました。

ここで、次の問いを考えてみましょう。

### 問い

平均値が 約 133g だったとします。

この数値だけを見て、

- ・ 「この店のポテトは、だいたい 133g だ」
- ・ 「品質として問題ない」

と判断してよいでしょうか？

ヒント：

- ・ 最小値や最大値は、平均からどのくらい離れていましたか？
- ・ 133g から大きく外れたデータは、いくつありましたか？

## 7. 演習②：ヒストグラムで全体の形を確かめる

演習①で求めた数値（平均・最小・最大）をふまえ、ポテト重量データ全体の「集まり方」をグラフで確認します。

### 対象とする列

- ・ 重さ\_g（ポテトの重量）

### 作業内容

1. 「重さ\_g」の列を選択する
2. 「挿入」タブから **ヒストグラム** を作成する

### 観察するポイント

- ・ データは、どのあたりに多く集まっているか
- ・ 平均値（約 133g）は、山の中央にあるか
- ・ 端のほうに、極端に離れた値は見えるか

## 8. 分布 (Distribution) : データの集まり方を見る

演習②で作成したヒストグラムは、ポテト重量データの「分布」を表しています。

### 分布とは

- ・ どの値が
- ・ どのくらいの回数（頻度）で
- ・ どのように集まっているか

を表したものです。

### 重要なポイント

- ・ 分布を見ると、「平均のまわりにどう広がっているか」が分かる
- ・ 同じ平均値でも、分布の形が違えば意味は大きく変わる

### ここまでで分かったこと

- ・ ポテトの重さは、毎回まったく同じではない
- ・ そのばらつきは、ヒストグラムの「形」として現れている

## 9. 分布を見る視点①：中心（平均・中央値）

分布を見ると、まず注目するのがデータの「中心」です。

### 中心を表す代表的な数値

- ・ 平均値：すべての値を均したときの中心
- ・ 中央値：小さい順に並べたときの真ん中

### ヒストグラムと対応づけて考える

- ・ 平均値（約 133g）は、分布のどの位置にあるか
- ・ 中央値は、山の中心付近にあるか

### 重要な点

- ・ 平均と中央値が近い場合、分布は比較的対称と考えられる
- ・ 大きくずれている場合、分布に偏りがある可能性が高い

## 10. 分布を見る視点②：広がり（ばらつき）

分布を見るときは、中心だけでなくデータが「どのくらい散らばっているか」にも注目します。

### 広がりを感じ取るための手がかり

- ・ 最小値と最大値：データがどこからどこまで広がっているか
- ・ 中心からの距離：平均（約 133g）から大きく離れた値があるか

### ヒストグラムで確認する

- ・ 棒が横にどれくらい広がっているか
- ・ 中心付近に集中しているか、ばらけているか

### 重要な点

- ・ 広がりが小さいほど、重さは安定していると言える
- ・ 広がりが大きい場合、同じ平均でも「ばらつきの大きい店」となる

## 11. 演習③：外れ値になっているデータ行を確認する

ヒストグラムで確認した分布の中から、中心から大きく離れている値（外れ値になりそうなデータ）を実際の行として確認します。

### 作業内容

1. **重さ\_g** の列を並べ替える（昇順・降順）
2. 特に小さい値・大きい値の行をいくつか選ぶ
3. その行の **曜日・学生 ID・メモ** を確認する

### 考えてみるポイント

- ・なぜこの値は、他と比べて大きく（または小さく）なったのか
- ・測定ミスの可能性はあるか
- ・特別な状況（忙しさ、盛り付けのクセなど）が考えられるか

## 12. 演習④：曜日別に分けて、平均とばらつきを比べる

これまでの演習では、全データをまとめて見てきました。ここでは、データを **曜日ごと** に分けて見てみます。

### 作業内容

1. ピボットテーブルを作成する
2. 行に **曜日** を配置する
3. 値に **重さ\_g (平均)** と **重さ\_g (件数)** を配置する

### 確認するポイント

- ・ 曜日によって、平均値は同じか、違うか
- ・ 件数はどの曜日も同じか
- ・ 特定の曜日だけ、重め・軽めになっていないか

### ここで考えること

- ・ もし曜日で傾向が違えば、全体の分布は何か混ざったものか

## 13. 演習⑤：学生 ID 別に分けて、平均を比べる

次に、同じポテトでも測定する人（学生）によって違いがあるかを確認します。

### 作業内容

1. ピボットテーブルを作成する
2. 行に **学生 ID** を配置する
3. 値に **重さ\_g (平均)** を配置する

### 確認するポイント

- ・ 学生ごとに、平均値は同じか、違うか
- ・ 明らかに「重め」「軽め」になっている人はいないか

### ここで考えること

- ・ 全体の分布は、どのような分布が混ざった結果だろうか

## 14. 分布を「読む」とは何をすることか

ここまでの演習を通して、私たちはポテト重量データをさまざまな角度から見てきました。

### 分布を読むときの3つの視点

- ・ **中心**：どのあたりの値が基準になっているか（平均・中央値）
- ・ **広がり**：どの程度ばらついているか（最小～最大、散らばり方）
- ・ **混ざり**：曜日や人ごとの違いが重なっていないか

### 重要なポイント

- ・ 分布を見ることで、数値1つでは分からない実態が見えてくる
- ・ 全体の分布は、複数の要因が重なった結果であることが多い

## 15. 本日のまとめ：記述統計でできるようになったこと

本日の授業では、30人×1週間のポテト重量データを使って、「記述統計」によるデータの読み取りを行いました。

### 今日できるようになったこと

- ・ データ全体を **代表値**（平均・中央値）で要約する
- ・ ヒストグラムを使って **分布の形**を確認する
- ・ 外れ値になりそうな値を見つけ、その **背景**を考える
- ・ 曜日や人ごとに **分けて見る** ことで、違いを捉える

### 最も大切なポイント

- ・ 数値1つだけでは、データの実態は分からない
- ・ 「分布を読む」ことで、現場で起きていることが見えてくる

## 16. 次回予告：分布の背後にある「起こりやすさ」

本日は、データの「形」を読み取るところまで進みました。

次回は、

- ・なぜ、このような分布の形になるのか
- ・どの重さが、どのくらいの確率で起こるのか

を考えるために、**確率**を扱います。

### 次回のテーマ

- ・分布を「感覚」ではなく「数」で表す
- ・「起こりやすさ」を確率として表現する

本日のポテトの分布が、そのまま次回の出発点になります。