

**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人で測ったボ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人がボテトの重さを測りました。

その結果、

- 平均はだいたい133gで重い日がある
- 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

- 同じ店のボテトを使う
- 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

こうして集まつたのが、30人分のボテト重量データです。

**4. 今日やること：集めたデータをどう見るか**

今日は、この30人のデータを使って、次のことを順番に行います。

- 数でまとめてみる（平均・最小・最大）
- グラフで形を見てみる
- 「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：ボテト重量の基本統計量を求める

クラス30人の1週間測定したボテト重量データ（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（ボテトの重量）

作業内容

- AVERAGE 関数で 平均値 を求めろ
- MEDIAN 関数で 中央値 を求めろ
- MIN, MAX 関数で 最小値・最大値 を求めろ

確認するポイント

- 平均値と中央値は近いか、離れているか
- 最小値・最大値は、平均からの程度離れているか

6. 【問い合わせ】平均値だけで、この店のボテトを評価できるか？

演習①で、ボテト種量の 平均値・中央値・最小値・最大値 を求めました。

ここで、次の問い合わせをしてみましょう。

問い合わせ

平均値が約133gだったとします。  
この数値だけを見て、  
・「この店のボテトは、だいたい133gだ」  
・「品質として問題ない」  
と判断していいでしょうか？

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から離きれた値は、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、ボテト重量データ全体の「作り方」をグラフで確認します。

対象とする列

- 重さ\_g（ボテトの重量）

作業内容

- 「重さ\_g」の列を選択する
- 「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、どのように多く集まっているか
- 平均値（約133g）は、山の中間にあらかじめあるか
- 端っこには、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、ボテト重量データの「分布」を表しています。

分布とは

- どの値
- どのくらいの回数（頻度）で
- どのように集まっているか

を表すものです。

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、分布の形がええば意味は大きく変わる

ここまででわかったこと

- ボテトの重さは、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

ヒストグラムと確認できる

- 平均値（約133g）は、分布のどの位置にあるか
- 中央値は、山の中間に位置するか

重要な点

- 平均値と中央値が近い場合、分布は比較的対称と考えられる
- 大きくなっている場合、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）から大きさ離れた値があるか

ヒストグラムで確認する

- 棒が離れにくらい広がっているか
- 中心附近に集中しているか、ばらけているか

重要な点

- 広がりが大きいほど、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. 平均と分散：グラフで見るデータの「姿」

平均の範囲：平均値と中央値だけを見ると、2つのグラフはそっくりに見えます。しかし、グラフの形（山の高さ）は全く異なります。

分散の範囲：「分散」はより細かい、分布の幅を表します。

- 分散大：山の底くねに広がり、山の高さがバラバラな状態です。
- 分散小：山の底くねく、平均付近に集中しています。

結論：「平均」と「分散」を知るために、平均だけではなく、「分散」이라는物事を確認することが不可欠です。

12. 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

- 重さ\_gの列を並べ替える（昇順・降順）
- 特に小さい値・大きい値の行いくつか選ぶ
- その行の値：学生ID・重さ\_g・Xを確認する

考え方のポイント

- 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

13. 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います。

作業内容

- ヒストグラムを作成する
- 行に曜日を配分する
- 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
- 件数はどの曜日と同じか
- 特定の曜日だけ、重めになっていたいなか

ここで考えること

- もし曜日で傾向が違うなら、全体の分布は何が混ざったものか

14. 演習⑤：学生ID別に分けて、平均を比べる

次に、同じボ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

- ヒストグラムを作成する
- 行に学生IDを配分する
- 重さ\_g（平均）を配置する

確認するポイント

- 学年ごとに、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るのか

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

15. 分布を「読む」とは何をすることか

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

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：どの程度もついているか（最小～最大、散らばり方）
- 離れ：曜日や人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つでは分からず実態が見えてくる
- 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

16. 本日のまとめ：記述統計ができるようになったこと

本日の授業では、30人1週間のボ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てることで、違いを捉える

最も大切なポイント

- 数値だけでは、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

17. 次回予習：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、確率を学びます。

次回のテーマ

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

本日のボテトの分布が、その後次回の出発点になります。

12

9

10

11

16

13

14

15

17