

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。

今回の問題は、

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

ことで、データの全体像を読みこむを目指します。

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。

導入

- ・第1回との連続性をもつて説明する。
- ・「今日は何をする日か」を元に判断する。

2020/01/20 アルゴリズム 2/103

1 / 16

1

2

3

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る
- ・でも、毎日・重い日がある

ことがわかりました。

ここで、10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？

2020/01/20 アルゴリズム 2/103

疑問提示

- ・1人データの範囲を、学生自身に気づかせる。
- ・次の行動（人数を増やす）への動機つけ。

2 / 98

4

5

6

7

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

2020/01/20 アルゴリズム 2/103

演習への接続

- ・分析手法をはじめて学び、行動の流れだけを示す。
- ・このまま前進するに自然にならぬ。

4 / 98

8

9

10

11

6. 【問い合わせ】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、次の問い合わせをしてみましょう。

問い合わせ

平均値は約132g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：

- ・最小値と最大値は、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？

2020/01/20 アルゴリズム 2/103

問い合わせ

- ・学生が自らに「何が何ではないか」と感じを発揮させる。
- ・また「自分」という言葉を使はず、他者だけを任せせる。
- ・今の状況（ヒストグラム）への感想つけを促す。

5 / 98

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回間接測定したポテト重量データ（poteto30.xlsx）を用いて、全體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か、離れているか
- ・最小値・最大値は、平均からどの程度離れているか

2020/01/20 アルゴリズム 2/103

5 / 98

12

13

14

15

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
- ・どの範囲
 - ・どのくらいの回数（頻度）
 - ・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、分布の形が違うは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

2020/01/20 アルゴリズム 2/103

分布の導入

- ・新しい統計や専門用語は出さず、直前のヒストグラム体験と結びつける。
- ・「分布の形」という言葉は使わず、「頻度」だけを任せせる。
- ・次に、分布を見ることで「どうなっているか」を予測させる。

6 / 98

9. 分布を見る視点①：中心（平均・中央値）

分布を見ると、まずは見るもののデータの「中心」です。

中心を見る代表的な数値

- ・平均値：すべての個々を均したときの中心
- ・中央値：小さい順に並べてときの真ん中

ヒストグラムと対応づける考え方

- ・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、分布の中心にあるか
- ・端のほうに、端間に離れた値は見えるか

2020/01/20 アルゴリズム 2/103

7 / 98

16

17

18

19

10. 分布を見る視点②：広がり（ばらつき）

分布を見ると、中心だけではなくデータが「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・最大値：データがどこからどこまで広がっているか
- ・中央からの距離：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

- ・横軸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

2020/01/20 アルゴリズム 2/103

広がりを見る視点

- ・他のデータ間隔を比べます。まずは平均的、他の間にばらつきを感じ取れます。
- ・安定している（していない）という評議が、なぜか広がるから生まれることをさせます。
- ・次に、広がりの大きさで特に目立つ時は幅へ着眼をつなげます。

10 / 98

11. 演習③：外れ値になっているデータを行き飛ばす確認する

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動として確認します。

作業内容

1. 重さ(g)の列を並び替える（昇順・降順）
2. 特に小さく離れた大きな値が、なぜか広がるから生まれることをさせます。
3. その日の重さ、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、他のと比べて大きくなったり小さくなったりしたのか
- ・測定の誤りがあるか
- ・特別な状況（忙しさ、落り付けのクセなど）が考えられるか

2020/01/20 アルゴリズム 2/103

11 / 98

20

21

22

23

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. 「棒」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、端間に離れた値は見えるか

2020/01/20 アルゴリズム 2/103

7 / 98

8. 演習②：組み立てる

- ・組み立てる（削除）：不要な構造を削除する。
- ・平均値を計算してではなく、分母分子の差算で計算する。
- ・次のスライドで「平均だけで十分か？」という質問を導入する準備をします。

12

13

14

15

9. 中心を見る組い

- ・演習①で求めた数値を、分布の中の「集まり方」、意味を持たせる。
- ・平均と中央値は計算方法ではなく、分母との割合で理解させる。
- ・次に「なぜそれなのか？」という質問を自然に生む。

16

17

18

19

10. 演習③：組み立てる

- ・「外れ値」：という言葉を、実際のデータと結びつける。
- ・数値だけなら、幅広か、狭いかといった分布特徴をとる意識を持たせる。
- ・次に、データを分けて見る（日立集・人別）必要性につなげる。

20

21

22

23

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
- ・特記の欄だけ、重め・軽めになっていないか

ここで覚えること

- ・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

2022/05/アルゴリズム2/103

演習④の狙い

- ・まじめである。」「自分で見る」で、見えるものが変わることを体験させる。
- ・毎日という文脈で、自分の行動を理解する慣習を意識させる。
- ・次に、人ごとの違い(測定者のかけ)へ自然につなげる。

13. 演習⑤：学生ID別に分けて、平均を比べる

これまでの演習では、同じボトでも測定する人(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ないか

ここで覚えること

- ・全体の分布は、どのような分布が混ざった結果だろうか

演習⑤の狙い

- ・人による違い、校、分野の順序を問わざることを体験させる。
- ・全集の分野ごとの傾向ではなく、複数の数の要素があると気づかせる。
- ・次に、分野を「読む」。この基礎を活用するまじめへ進む。

24

25

26

27

14. 分布を「読む」とは何をすることか

これまでの演習を通して、私たちはボト重畳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

- ・中央：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：どの程度分かれているか(最小~最大、散らばり方)
- ・選ばり：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、数値一つは分からぬ実態が見てくる
- ・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

2022/05/アルゴリズム2/103

分布を読むまとめ

- ・これまでの演習(①~③)を、3つの視点に整理して回顧する。
- ・新しい知識を追加する、「やったことを言葉にすす」実践に慣れる。
- ・次のミーティングで、授業全体の振り返り準備をす。

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週間のボト重畳データを使って、「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

- ・数値一つだけでは、データの実態は分からぬ
- ・「分布を読む」ことで、現場で起きていることが見えてくる

本日のまとめ

- ・第2回の目的と、実際に行った演習を対応づけて整理する。
- ・「平均だけでは見りない」という触に必ず立ち戻る。
- ・学生自身の体験として実践化させる。

28

29

30

31

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

本日のデータの分布が、そのまま次回の出発点になります。

2022/05/アルゴリズム2/103

次回への接続

・今日扱った分布が、そのまま次の教材になることを示す。

・「確率」別の話ではなく、過渡した学びである印象をつける。

・第3回への心構えワードを下げる。

33

16 / 36

15 / 36

32

16 / 36

15 / 36