

	<div>1.第2回：記述統計ーデータの全体像をつかむ</div> <div>前回は、1人で買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。</div> <div>今回は、<ul style="list-style-type: none">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、データの全体像を掴むことを目指します。</div> <div>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。</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>1 / 17</div>	<div>2.前回の測定で得た疑問</div> <div>前回は、「私」1人がポテトの重さを測りました。</div> <div>その結果、<ul style="list-style-type: none">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。</div> <div>ここで、1つ疑問が残ります。</div> <div>このばらつきは、<ul style="list-style-type: none">たままたなのか？「私」だけの測定結果なのか？</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>2 / 17</div>	<div>3.調査チームを作って確かめることにした</div> <div>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なりました。</div> <div>調査のルール<ul style="list-style-type: none">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</div> <div>こうして集まったのが、30人分のポテト重量データです。</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>3 / 17</div>
--	---	--	--

	<div>4.今日やること：集まったデータをどう見るか</div> <div>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。</div> <div>1. 数でまとめてみる（平均・最小・最大）</div> <div>2. グラフで形を見てみる</div> <div>3. 「なぜこうなったか」を考える</div> <div>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。</div> <div>→ 演習①へ</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>4 / 17</div>	<div>5.演習①：ポテト重量の基本統計量を求める</div> <div>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。</div> <div>対象とする列<ul style="list-style-type: none">重さ_g (ポテトの重量)</div> <div>作業内容<ul style="list-style-type: none">AVERAGE 関数で 平均値 を求めるMEDIAN 関数で 中央値 を求めるMIN, MAX 関数で 最小値・最大値 を求める</div> <div>確認するポイント<ul style="list-style-type: none">平均値と中央値は近い、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</div> <div>10/10/2020,</div> <div>5 / 17</div>	<div>6.【問い】平均値だけで、この店のポテトを評価できるか？</div> <div>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。</div> <div>ここで、次の問いを考えてみましょう。</div> <div>問い<div>平均値が 約133g だったとします。この数値だけを見て、<ul style="list-style-type: none">「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？</div></div> <div>ヒント：<ul style="list-style-type: none">最小値や最大値は、平均からどのくらい離れていたか？133g から大きく外れたデータは、いくつありましたか？</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>6 / 17</div>	<div>7.演習②：ヒストグラムで全体の形を確認する</div> <div>演習①で求めた数値（平均・最小・最大）をふまえて、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。</div> <div>対象とする列<ul style="list-style-type: none">重さ_g (ポテトの重量)</div> <div>作業内容<ul style="list-style-type: none">「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</div> <div>確認するポイント<ul style="list-style-type: none">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山のように、極端に離れた値は見えるか</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>7 / 17</div>
--	--	--	---	---

	<div>8.分布 (Distribution)：データの集まり方を見る</div> <div>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。</div> <div>分布とは<ul style="list-style-type: none">どの値がどのくらい多い回数（頻度）でどのように集まっているかを表したものです。</div> <div>重要なポイント<ul style="list-style-type: none">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</div> <div>ここまでは分かったこと<ul style="list-style-type: none">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</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>8 / 17</div>	<div>9.分布を見る視点①：中心（平均・中央値）</div> <div>分布を見ると、まず注目するのがデータの「中心」です。</div> <div>中心を表す代表的な数値<ul style="list-style-type: none">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</div> <div>ヒストグラムと対応づけて考える<ul style="list-style-type: none">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</div> <div>重要な点<ul style="list-style-type: none">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</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>9 / 17</div>	<div>10.分布を見る視点②：広がり（ばらつき）</div> <div>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。</div> <div>広がりを感じるための手がかり<ul style="list-style-type: none">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</div> <div>ヒストグラムで確認する<ul style="list-style-type: none">峰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</div> <div>重要な点<ul style="list-style-type: none">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</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>10 / 17</div>	<div>11.平均と分散：グラフで見るデータの「姿」</div> <div></div> <div>平均の解釈：平均値は中央値だけを見ると、2つのデータはそろっているように見えます。しかし、データの形（山の置き方）は全く違います。</div> <div>分散の姿：「分散」は山の「太り具合」を表します。分散大：山の裾（裾）に広がり、点数がバラバラな状態です。分散小：山の裾が狭く高く、平均付近に集中しています。</div> <div>結 語：データの「正確な姿」を知るためには、平均だけでなく「分散」という特徴も確認することが不可欠です。</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>11 / 17</div>
--	---	---	--	--

	<div>12.演習③：外れ値になっているデータを確認する</div> <div>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。</div> <div>作業内容<ul style="list-style-type: none">重さ_gの列を並び替える（降順・降順）特に小さい値・大きい値の行をいくつか選ぶその行の 曜日・学生ID・メモを確認する</div> <div>考えてみるポイント<ul style="list-style-type: none">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</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>12 / 17</div>	<div>13.演習④：曜日別に分けて、平均とばらつきを比べる</div> <div>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。</div> <div>作業内容<ul style="list-style-type: none">1. ポットテーブルを作成する2. 行に 曜日 を配置する3. 列に 重さ_g (平均) と 重さ_σ (分散) を配置する</div> <div>確認するポイント<ul style="list-style-type: none">曜日によって、平均値は同じか、違うか件数はどの曜日とも同じか特定の曜日だけ、重く・軽めになっていないか</div> <div>ここで考えること<ul style="list-style-type: none">もし曜日で傾向が違ったら、全体の分布は何か異なっているのか</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>13 / 17</div>	<div>14.演習⑤：学生ID別に分けて、平均を比べる</div> <div>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。</div> <div>作業内容<ul style="list-style-type: none">1. ポットテーブルを作成する2. 行に 学生ID を配置する3. 列に 重さ_σ (平均) を配置する</div> <div>確認するポイント<ul style="list-style-type: none">学生ごとに、平均値は同じか、違うか明らかに「重く」「軽め」になっている人はいないか</div> <div>ここで考えること<ul style="list-style-type: none">全体の分布は、どのような分布が混ざった結果だろうか</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>14 / 17</div>	<div>15.分布を「読む」とは何をすることか</div> <div>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。</div> <div>分布を読むときの3つの視点<ul style="list-style-type: none">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）歪り：曜日や人ごとの違いが重なっていないか</div> <div>重要なポイント<ul style="list-style-type: none">分布を見ることで、数値1つでは分からない現象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</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>15 / 17</div>
--	---	--	---	---

	<div>16.本日のまとめ：記述統計でできるようになったこと</div> <div>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。</div> <div>今日できるようになったこと<ul style="list-style-type: none">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、違いを見つける</div> <div>最も大切なポイント<ul style="list-style-type: none">数値1つだけでは、データの姿態は分からない「分布を読む」ことで、現場で起きていることが見えてくる</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>16 / 17</div>	<div>17.次回予告：分布の背後にある「起こりやすさ」</div> <div>本日は、データの「形」を読み取るところで進みました。次回は、<ul style="list-style-type: none">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。</div> <div>次回のテーマ<ul style="list-style-type: none">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</div> <div>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。</div> <div>10/10/2020, データサイエンス 2 / 101</div> <div>17 / 17</div>
--	---	--