

02 統計で何ができるか

1. 今日のゴール：統計を「使う」ための第一歩

本日の授業では、計算の手順を覚えることよりも、以下の3点を理解することを目指します。

1. 統計学の役割：なぜ数字を使って説明する必要があるのかを知る。
2. データの形：コンピュータが処理しやすい「正しい表」の形を理解する。
3. 数える技術：Excel を使い、データの全体像を正確に把握する。

「統計は、単なる計算ではなく、事実を整理して判断を助ける道具である」という実感を持ちましょう。

導入

初回なので、数学的な難しさではなく、実務的な有用性を強調して安心感を与えます。

2. 統計学の役割：バラバラな記録を「情報」に変える

統計学とは、大量にある「事実の記録（データ）」を、意味のある「情報」に変換する技術です。

- ・要約：1000人のデータを、平均値などの「1つの数字」で表現する。
- ・判断：新しい施策に効果があったのか、偶然なのかを見極める。
- ・予測：過去の傾向から、来月の売上やリスクを予見する。

経験や勘（「たぶんこうだろう」）を、客観的な根拠（「データがこう示している」）に置き換えることが統計学の目的です。

統計の役割

「勘 vs 根拠」の対比を明確にし、ビジネスや研究での必要性を伝えます。

3. データとは「事実の記録」である

「データ」と聞くと難しい数値を想像しがちですが、実際にはもっと広い意味を持ちます。

データの具体例

- ・ **数値**：売上金額、気温、テストの点数、体重
- ・ **文字・記号**：商品名、性別（男/女）、合否（○/×）、満足度（A/B/C）

数値だけでなく、アンケートの回答や「はい/いいえ」といった記号も、集めて数えれば立派な分析対象になります。

データの定義

「数字以外もデータである」ことを認識させ、分析の幅を広げます。

4. データの種類①：質的データ（カテゴリ）

データはその性質によって大きく2つに分類されます。1つ目は「質的データ」です。

質的データ（カテゴリデータ）

- ・ 分類や区別のためのデータです。
- ・ 例：血液型、出身地、支持政党、背番号
- ・ **特徴**：足し算や平均を出すことに意味がありません。（例：血液型 A と B を足しても意味がない）
- ・ **分析方法**：「何件あるか」を数え、割合を出すのが基本です。

尺度水準

「計算して意味があるかないか」という直感的な基準で説明します。

5. データの種類②：量的データ（数値）

2つ目は、量や大きさを表す「量的データ」です。

量的データ（数値データ）

- ・ 数値として意味を持ち、計算ができるデータです。
- ・ 例：身長、価格、販売個数、試験の点数
- ・ **特徴**：合計や平均、最大・最小を求めることで、データ全体の傾向が見えてきます。

重要な区別

コンピュータは数字が入っていれば何でも計算しようとしますが、人間が「この数字は計算して良い種類か？」を判断しなければなりません。

尺度水準

来期の Python 学習でも、数値型とカテゴリ型の区別として再登場します。

6. Excel 表の基本：1行=1件の原則

コンピュータで統計分析を行うには、表が「正しい形」である必要があります。

- ・ **1行=1件（レコード）**：横一行には、ある一回分の全ての情報を並べます。
- ・ **1列=1項目（変数）**：縦一列には、同じ種類のデータだけを並べます。

見た目の美しさ（セルの結合や飾り）よりも、データが規則正しく並んでいる「構造」が重要です。

Excel の構造

「見た目のための表」と「分析のためのデータ」は別物であることを強調します。

7.【問い合わせ】なぜセルの結合や空白行は「ダメ」なのか？

Excelでデータを管理する際、セルを結合したり、見やすくするために1行あけたりすることがありますが、統計分析では強く推奨されません。

その理由は何でしょうか？ 少し考えてみてください。

(次のスライドに答えがあります)

問い合わせ

学生に一度考えさせ、実感を伴った理解を促します。

8.【答え】コンピュータは「1行」を独立した事実と捉えるから

人間は表を見て「ここは上の行と同じ意味だな」と空気を読みますが、コンピュータは1行ずつ機械的に処理します。

- ・ **結合セルがある場合**：2行目以降が「空っぽ（データなし）」と判定され、正しく集計されません。
- ・ **空白行がある場合**：そこでデータが終わっていると誤認される原因になります。

鉄則：1つの行だけを取り出したとき、それだけで内容が完結している状態（全てのセルに値が入っている状態）を目指しましょう。

答え

Excel のオートフィルタやピボットテーブルが動かなくなる原因であることを補足します。

9. 実習：使用するデータと環境

実際に Excel を使ってデータを整理してみましょう。

- ・ 使用ファイル：02_実習.xlsx
- ・ 使用シート：chap2-1
- ・ データの確認：
 - ・ A列：商品ID（質的データ：文字）
 - ・ B列：売上個数（量的データ：数値）

まずは、この表が「1行 = 1件」の形になっているか、目で見て確認してください。

実習準備

ファイルを開く時間を確保し、列の性質（質的/量的）を再確認させます。

10. 実習①：データが全部で何件あるか数える

分析の第一歩は「全体（分母）」を知ることです。商品が何種類あるか数えます。

使用関数：**COUNTA**（カウント・エー）

- ・ 役割：指定した範囲の中で、空白ではないセルの個数を数える。
- ・ 入力：`=COUNTA(A2:A9)`

ポイント：A列は「商品名（文字）」なので、数値専用の COUNT 関数ではなく、文字も数えられる COUNTA を使います。

COUNTA

「A は All (全て) の A」と覚えると忘れにくいことを伝えます。

11. 実習②：数値が入っている件数を数える

次に、売上個数が入力されている件数を確認します。

使用関数：**COUNT**（カウント）

- ・ 役割：指定した範囲の中で、「数値」が入っているセルの個数だけを数える。
- ・ 入力：`=COUNT(B2:B9)`

もし B 列に文字（「不明」など）が入っていた場合、この関数ではカウントされません。データの欠損（入力漏れ）を見つけるのにも役立ちます。

COUNT

COUNTA との使い分けを意識させます。

12. 実習③：売上の合計を求める

量的データ（数値）の大きな特徴は、足し合わせることで「規模」が見えることです。

使用関数：SUM（サム）

- ・ 役割：数値をすべて合計する。
- ・ 入力：`=SUM(B2:B9)`

これがデータの「総量」になります。この合計値を各商品の個数で割ることで、後に「シェア（構成比）」を計算できるようになります。

SUM

基本関数ですが、統計における「総量把握」の意味を添えます。

13. 実習④：極端な値（最大・最小）を確認する

平均を見る前に、データの「端」を確認しましょう。

- ・ 最大値 (MAX) : =MAX(B2:B9)
- ・ 最小値 (MIN) : =MIN(B2:B9)

一番売れた商品と売れなかった商品の差はどれくらいあるでしょうか？ この「差」が大きいほど、ばらつきの激しいデータであると言えます。

MAX/MIN

「範囲（レンジ）」という概念への伏線です。

14. 実習⑤：代表値（平均と中央値）を出す

データを一言で表す「代表値」を計算します。

- ・ 平均値 (AVERAGE) : 全て足して件数で割った値。
- ・ 中央値 (MEDIAN) : データを並べ替えてちょうど真ん中にくる値。

自習課題：自分の Excel の結果を見てください。平均値と中央値は同じ数字になりましたか？ 大きくズレていますか？ なぜズレるのかは次回詳しく扱います。

代表値

同じ「真ん中」でも計算方法が違うことを体験させます。

15. 分析のポイント：数値を「割合」に変換する

実習で「全体件数」や「合計」を出したのは、数値を比較可能な形にするためです。

比較の例：どちらの商品がすごい？

- ・ A 商品：10 個売れた
- ・ B 商品：50 個売れた

これだけでは B 商品がすごいように見えますが、もし A 商品が「限定 20 個中」で、B 商品が「1000 個中」の売上だったらどうでしょうか？ このように、「**全体に対してどれくらいか（割合）**」を出すことで、初めて公平な比較が可能になります。

割合の重要性

実習で行った COUNTA や SUM が「分母」を作る作業だったことを言語化します。

16. 言葉の進化：過去の「割合」は未来の「確率」

統計学の後半で学ぶ「確率」は、実は今日計算した「割合」と同じものです。

- ・ **割合**：過去に起こった事実の集計結果。
 - ・ 例：昨日まで 100 回中 30 回雨が降った（割合 = 0.3）
- ・ **確率**：過去の割合を、未来の予想に当てはめたもの。
 - ・ 例：明日雨が降る「もっともらしさ」は 30% だ（確率 = 0.3）

計算式はどちらも「部分 ÷ 全体」です。視点が「過去」か「未来」かだけで、呼び方が変わります。

確率への橋渡し

確率を身近なものとして定義し直します。

17. なぜ確率を学ぶのか：不確定な未来を判断するため

私たちは未来を 100% 予言することはできませんが、データを使えば「高い確率でこうなる」と予想できます。

- ・ 過去の故障率（割合）から、次の製品が壊れる「確率」を出す。
- ・ 昨日の来客数（割合）から、今日の仕入れ量を「確率的」に決める。

「たぶん」という曖昧な言葉を、「〇〇%の確率」という数値に変える。これが、統計学がビジネスや科学で重宝される理由です。

確率の意義

「不確実性への対処」という統計の核心に触れます。

18. 本日のまとめ

- ・ 統計の目的：データを整理し、客観的な判断の根拠を作ること。
- ・ 正しい表の形：1行1件。セルの結合はNG。
- ・ 質的データ：カテゴリ。COUNTAで件数を数える。
- ・ 量的データ：数値。SUM や AVERAGE で傾向を見る。
- ・ 割合と確率：過去の集計（割合）を未来の予測（確率）に使う。

次回からは、これらのデータがどのように「ばらつく」のか、その形（分布）について学んでいきます。

まとめ

全体を振り返り、次回の予告をすることで学習の継続性を高めます。

19. 学習カルテ：本日の振り返り

本日の学習内容について、LMS から「学習カルテ」に回答してください。

確認項目

1. 量的データと質的データの違いを説明できますか？
2. Excel で「1行1件」にする理由を覚えていますか？
3. 割合と確率の関係を一言で言うと何ですか？

お疲れ様でした。質問がある場合は、終了後に受け付けます。

終了

授業の締めくくり。LMS 等への誘導を行います。