

【データの取り扱い】

1. データの収集

データを活用するためには、まずはデータを集める必要がある。データを何かに利用する前提として、データの（ ）が重要となる。特に、インターネット上に存在する膨大な情報を扱う際に気を付けなければならない。

- （ ）データ…国、地方自治体や事業者が保有する官民データ。
誰もが利用できるように公開されているが、次のすべてを満たすことが条件になる。（対義語＝ **データ**）
 - ① 営利目的、非営利目的を問わず、（ ）可能
 - ②（ ）に適していること
 - ③（ ）で利用できること

[閲覧・検索ポータルサイトの例]（

など)

- 調査方法…データを集める対象の違いによって2つの方法がある。

- ①（ ）調査…対象となるすべてを調べる。
- ②（ ）調査…一部だけを選んで調べる。
（ ）ともいう。

注意 ②は標本（サンプルデータ）が（ ）の特徴をよく表すように（ ）なく抽出することが重要。

[作業] オープンデータの閲覧・検索ポータルサイトを見てみよう。

- (例 1) <https://www.e-stat.go.jp/>
- (例 2) <https://data.e-gov.go.jp/info/ja>
- (例 3) <https://kurashiki.dataeye.jp/>

2. データの整理

データには量的データと質的データがあるが、その性質によって（ 尺度水準 ）とよばれる基準で次のように分類される。

種類	尺度水準	性質	例
量的データ	() 尺度	原点の決め方が定まっていて、間隔にも比率にも意味がある。	等
	() 尺度	目盛りは等間隔だが、比率は意味をもたない。	等
質的データ	() 尺度	順序には意味があるが、間隔には意味がない数値で表す。	等
	() 尺度	数値ではない言葉、もしくは大小の意味がない数字で表す。	等

【回答課題 1】 各尺度水準のデータの例としてあてはまるものを考えてみましょう。

(Classroom で課題として出されたフォームに回答しましょう。)

- () …質的データを数値に対応させて () をすること。これによって、入力や集計などの処理効率を上げることができる。
(例) 「はい」と「いいえ」→「1」と「2」に割り当てる。
- 量的データを扱うときに注意すべきデータには次のようなものがある。
 - ① () …ほかの多数のデータから大きく離れた値
 - ② () …データが不明（未回答）もしくは収集できていない値
 これらの扱いには注意が必要。むやみに削除したり、平均値をあてはめたりすべきではない。

【データの取り扱い】

1. データの収集

データを活用するためには、まずはデータを集める必要がある。データを何かに利用する前提として、データの（ **信憑性** ）が重要となる。特に、インターネット上に存在する膨大な情報を扱う際に気を付けなければならない。

- （ **オープン** ）データ…国、地方自治体や事業者が保有する官民データ。
誰もが利用できるように公開されているが、次のすべてを満たすことが条件になる。（対義語＝ **クローズド** データ）
 - ① 営利目的、非営利目的を問わず、（ **二次利用** ）可能
 - ②（ **機械判読** ）に適していること
 - ③（ **無償** ）で利用できること

[閲覧・検索ポータルサイトの例]（ e-Stat, e-GOV データポータル, data eye, Resas… など）

- 調査方法…データを集める対象の違いによって2つの方法がある。
 - ①（ **全数** ）調査…対象となるすべてを調べる。
 - ②（ **標本** ）調査…一部だけを選んで調べる。
（ **サンプリング** ）ともいう。

注意 ②は標本（サンプルデータ）が（ **母集団** ）の特徴をよく表すように（ **偏り** ）なく抽出することが重要。

[作業] オープンデータの閲覧・検索ポータルサイトを見てみよう。

- （例 1） <https://www.e-stat.go.jp/>
- （例 2） <https://data.e-gov.go.jp/info/ja>
- （例 3） <https://kurashiki.dataeye.jp/>

2. データの整理

データには量的データと質的データがあるが、その性質によって（ **尺度水準** ）とよばれる基準で次のように分類される。

種類	尺度水準	性質	例
量的データ	（ 比率 ） 尺度	原点の決め方が定まっていて、間隔にも比率にも意味がある。	長さ，重さ，圧力 価格 等
	（ 間隔 ） 尺度	目盛りは等間隔だが、比率は意味をもたない。	暦，セルシウス温度， 偏差値 等
質的データ	（ 順序 ） 尺度	順序には意味があるが、間隔には意味がない数値で表す。	検定級，震度， 等星 等
	（ 名義 ） 尺度	数値ではない言葉，もしくは大小の意味がない数字で表す。	血液型，性別，会員番号 出席番号 等

【回答課題 1】 各尺度水準のデータの例としてあてはまるものを考えてみましょう。

（Classroom で課題として出されたフォームに回答しましょう。）

- （ **コーディング** ） …質的データを数字に対応させて（ **数値化** ）をすること。これによって、入力や集計などの処理効率を上げることができる。
（例）「はい」と「いいえ」→「1」と「2」に割り当てる。
- 量的データを扱うときに注意すべきデータには次のようなものがある。
 - ① （ **外れ値** ） …ほかの多数のデータから大きく離れた値
 - ② （ **欠損値** ） …データが不明（未回答）もしくは収集できていない値これらの扱いには注意が必要。むやみに削除したり，平均値をあてはめたりすべきではない。