データを整える（欠けた値・外れ値・表記のゆれ）

集めたデータを分析する前に、どのような整え方が必要なのだろう

集めたままのデータ（架空のアンケート）

番号	学年	1か月の利用回数	1回の時間
1	1年	4	30分
2	1年 ③	①	20分
3	二年 ③	300 ②	0.5時間 ④
4	2年	6	40分
4 ⑤	2年	6	40分
5	3年	2	15分

① 欠けた値（空らん）
② 外れ値（300回？）
③ 表記のゆれ（1年・二年）
④ 単位のちがい（分と時間）
⑤ 重複（同じ行が2回）

整える →

整えたデータ（時間は分にそろえた）

番号	学年	1か月の利用回数	1回の時間
1	1年	4	30
2	1年	（分析で除く）	20
3	2年	（確認中）	30
4	2年	6	40
5	3年	2	15

※ 重複した行は1つにした

整え方（前処理）の例

① 欠けた値
行ごと除くか、ほかの値から補うかを決める

② 外れ値
入力のまちがいか、本当の値かを確かめる

③ 表記のゆれ
同じものは同じ書き方にそろえる

④ 単位のちがい
同じ単位に直す（時間 → 分）

⑤ 重複
同じデータは1つにする

整え方で、分析の結果が変わる

外れ値を除くかどうか、欠けた値をどう補うかで、平均などの結果が変わる。整える前のデータは消さずに残す。

どのように整えたか（理由も）を記録し、ほかの人が同じ手順をたどれるようにする。

※ 模式図。アンケートとデータは架空の例。