**平均→全データ合計÷データ数**

・外れ値に影響されやすい

**中央値→小さい順に並べたときの中央値**

・外れ値に影響されない

**最大値・最小値→データの中で最も大きい/小さい値**

・データの範囲や外れ値の確認

**不偏分散→各値と平均の差の2乗の平均（n-1で割る）**

・データのばらつきの大きさを数値で表す

**不偏標準偏差→不偏分散の平方根**

・ばらつきを元の単位で理解したいときに使う（分散より直感的）

・母集団のばらつきを正確に推定するためのもの

・不偏分散を√で戻したもの。人間にとって直感的なばらつきの指標

**四分位数→データを4等分する区切りの値**

・第1四分位数（Q1）：下位25%

・第2四分位数（Q2）：中央値

・第3四分位数（Q3）：上位25%

・データの分布の偏りや広がりを把握

**レンジ→最大値 − 最小値**

・データがどれだけ広がっているかの簡易な目安

**自由度→自由に変動可能なデータの数**

（例）

・テストの点数を３人分もっている。

・平均点が７０点と決まっているとき、自由に設定できるのは２人まで

・最後の1人の点数は、平均70になるように自動的に決まってしまうから

・この場合の自由度は２

**不偏分散と分散**

・全員のテストの点数がわかっている→普通の分散

・一部のデータで、クラス全体のばらつきを推定したい→不偏分散

**層別分析→データを性別や年齢などの「グループ（層）」ごとに分けて分析する方法**

・男女別にテストの平均点を出す、年齢層別に商品購入率を見る

**クロス集計表→2つのカテゴリ（例：性別×合格・不合格）を組み合わせて、件数をまとめる表**

**・**カテゴリ×カテゴリの関係を見たいとき

**クロス集計での比率→クロス集計の結果を「割合」に変えて比較しやすくしたもの**

・男性のうち何％が50点以上か、女性のうち何％が50点未満か

**相関分析→2つの数値の関係の強さ・方向を見る分析（どれくらい一緒に増減するか）**

・「この2つは関係ありそう？」を判断したいとき

・0.7 ～ 1.0：強い正の相関

・0.4 ～ 0.7：中程度の正の相関

・0.2 ～ 0.4：弱い正の相関

・0.0 ～ 0.2：ほとんど相関なし

|  |
| --- |
|  |