

エキスパート活動 課題 A 「分散」

分散とは、データの**散らばり具合**を表す指標の一つで、**データが平均値からどれくらい離れているか**を示します。分散が大きいほどデータの散らばりが大きく、小さいほどデータの散らばりが小さいと言えます。

□分散のイメージ

例えば、以下のような2つのデータセットがあったとします。

Aさん：[10, 20, 30, 40, 50]

Bさん：[28, 29, 30, 31, 32]

どちらのデータセットも平均値は30ですが、AさんのデータはBさんのデータよりも散らばっていることがわかります。この散らばり具合を数値で表すのが分散です。

□分散の計算方法

分散は、以下の手順で計算します。

1. 各データの平均値を求める
2. 各データと平均値の差（**偏差**）を求める
3. 各偏差を2乗する
4. 2乗した偏差の平均値を求める

$$[\text{分散}] = \frac{\{(\text{各データ}) - (\text{平均値})\}^2 \text{の和}}{\text{全体の人数}}$$

□分散のメリット・デメリット

◎メリット：データの散らばり具合を**客観的に評価**できる

標準偏差（分散の平方根）と組み合わせることで、データのばらつきをより詳細に分析できる

●デメリット：単位が元のデータの2乗になるため、解釈が難しい場合がある

外れ値の影響を受けやすい

□分散の活用例

分散は、以下のような場面で活用できます。

品質管理：製品の品質のばらつきを評価する

金融：株価の変動リスクを評価する

教育：テストの成績のばらつきを評価する

☆分散は、データの散らばり具合を把握するための重要な指標です。分散を理解することで、データの特性をより深く理解し、様々な分析に役立てることができます。

発問 ある工場で製造されたある部品の直径を測定したところ、以下のようなデータが得られました。

データ A: 10.0mm, 10.1mm, 10.2mm, 9.9mm, 10.0mm

データ B: 9.8mm, 10.5mm, 9.5mm, 10.7mm, 10.0mm

- ① データ A とデータ B の平均値をそれぞれ計算しなさい。
- ② データ A とデータ B の分散をそれぞれ計算しなさい。
- ③ 分散の値を比較し、どちらのデータが平均値の周りにより散らばっていると考えられるか説明しなさい。

解答

平均値の計算:

$$\text{データ A: } (10.0 + 10.1 + 10.2 + 9.9 + 10.0) / 5 = 50.2 / 5 = 10.04\text{mm}$$

$$\text{データ B: } (9.8 + 10.5 + 9.5 + 10.7 + 10.0) / 5 = 50.5 / 5 = 10.1\text{mm}$$

分散の計算:

データ A:

$$\begin{aligned} \text{偏差の二乗: } & (0.04)^2 + (0.06)^2 + (0.16)^2 + (0.14)^2 + (0.04)^2 = 0.0016 + 0.0036 + 0.0256 + 0.0196 + \\ & 0.0016 = 0.052 \end{aligned}$$

$$\text{分散: } 0.052 / 5 = 0.0104$$

データ B:

$$\text{偏差の二乗: } (0.3)^2 + (0.4)^2 + (0.6)^2 + (0.6)^2 + (0.1)^2 = 0.09 + 0.16 + 0.36 + 0.36 + 0.01 = 0.98$$

$$\text{分散: } 0.98 / 5 = 0.196$$

分散の比較:

データ A の分散: 0.0104

データ B の分散: 0.196

データ B の分散がデータ A の分散よりも大幅に大きい。これは、データ B の値が平均値からより大きく離れていることを示している。

結論

データ Bの方がデータ Aよりも平均値の周りにより散らばっていると言える。

エキスパート活動 課題B 「標準偏差」

標準偏差とは、データの**散らばり具合**を表す指標の一つで、**平均値からどれくらいデータが散らばっているか**を示します。分散と深く関係しており、分散の平方根として計算されます。

□標準偏差のイメージ

例えば、2つのクラスのテスト結果を比較する場合を考えてみましょう。

Aクラス：平均点 70 点、標準偏差 5 点

Bクラス：平均点 70 点、標準偏差 15 点

どちらのクラスも平均点は同じですが、標準偏差が異なるため、テスト結果の散らばり具合が異なります。Aクラスは標準偏差が小さいため、ほとんどの生徒が平均点付近に集まっていると考えられます。一方、Bクラスは標準偏差が大きいため、平均点から大きく離れた点数を取っている生徒もいると考えられます。

□標準偏差の計算方法

標準偏差は、以下の手順で計算します。

1. 各データの平均値を求める
2. 各データと平均値の差（**偏差**）を求める
3. 各偏差を2乗する
4. 2乗した偏差の平均値（**分散**）を求める
5. 分散の平方根を求める

$$[\text{標準偏差}] = \sqrt{\text{分散}}$$

□標準偏差のメリット

- データの散らばり具合を、平均値と同じ単位で表すことができるため、直感的に理解しやすい
- 正規分布（釣鐘型の分布）の場合、平均値±標準偏差の範囲に約 68%、平均値±2×標準偏差の範囲に約 95%のデータが含まれるという性質があり、データの分布を把握しやすい

□標準偏差の活用例

- ①品質管理：製品の品質のばらつきを評価する
- ②金融：株価の変動リスクを評価する
- ③教育：テストの成績のばらつきを評価する
- ④健康：血圧などのバイタルデータのばらつきを評価する

□分散との違い

分散もデータの散らばり具合を表す指標ですが、単位が元のデータの2乗になるため、解釈が難しい場合があります。標準偏差は、分散の平方根を取ることで、単位を元のデータと同じにしているため、より直感的に理解しやすくなっています。

☆標準偏差は、データの散らばり具合を把握するための重要な指標であり、様々な分野で活用されています。

発問 あるクラスの学生5人の身長を測定したところ、以下のようなデータが得られました。

データA: 160, 164, 156, 160, 160

データB: 150, 170, 160, 160, 160

- ① データAとデータBの平均値をそれぞれ計算しなさい。
- ② データAとデータBの標準偏差をそれぞれ計算しなさい。
- ③ 標準偏差の値を比較し、どちらのデータが平均値の周りにより散らばっていると考えられるか説明しなさい。

データセット A: 160, 164, 156, 160, 160

平均値: 160.0

標準偏差: 2.5298221281347035

データセット B: 150, 170, 160, 160, 160

平均値: 160.0

標準偏差: 6.324555320336759

エキスパート活動 課題C 「偏差値」

偏差値とは、テスト結果などのデータの**散らばり具合**を、**平均点と標準偏差を用いて数値化したもの**です。これにより、自分の成績が全体の中でどの位置にあるのかを客観的に把握できます。

□偏差値の仕組み

平均点と標準偏差：偏差値を計算するには、テストの平均点と標準偏差が必要です。

個人の得点：自分の得点から平均点を引いて、標準偏差で割ります。

偏差値の計算：上記で求めた値を10倍して、50を足すと偏差値になります。

$$[\text{偏差値}] = 50 + \frac{(\text{本人の点数}) - (\text{平均点})}{\text{標準偏差}} \times 10$$

□偏差値の解釈

50【**基準値**】：平均点と一致

50より高い：平均点より高い

50より低い：平均点より低い

□偏差値のメリット

異なるテストの比較：平均点や難易度が異なるテストでも、偏差値を用いることで成績を比較できます。

相対的な評価：自分の成績が全体の中でどの位置にあるのかを把握できます。

進路選択：高校や大学の入学試験などで、合格可能性の目安として利用されます。

□偏差値の注意点

集団によって異なる：偏差値は、テストを受けた集団によって異なります。

絶対評価ではない：偏差値は、あくまで**相対的な評価**であり、絶対的な学力を示すものではありません。

外れ値の影響：極端に高い点数や低い点数があると、偏差値が歪むことがあります。

発問 太郎は国語と数学の課題テストを受けました。その結果、数学の得点が58点（平均42点、標準偏差8点）に対して、国語の得点は65点（平均47点、標準偏差12点）でした。単純に点数を比較した場合、国語の方がいい成績のように見えますが、これは本当にそう言えるでしょうか。答えなさい。

発問 偏差値が68であることは何を意味していますか。データが正規分布に従うことを前提として説明しなさい。