統計の基本

1 導入

この講義で最重要なのは、平均だけでなく、「どれくらい散らばっているか」まで見て集団を捉えることです。

統計で不親切に感じやすいのは、平均値、分散、標準偏差の式だけが先に出てきて、「なぜこの量を見るのか」が曖昧なまま進むことです。この講義では、まず平均が何を表し、なぜそれだけでは足りないかを確認してから、散らばりの量へ進みます。

2 用語と定義

平均値 Mean は

$$\overline{x} = \frac{x_1 + \cdots + x_n}{n}$$

です。

分散 Variance は

$$\frac{(x_1 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n}$$

です。

標準偏差 Standard deviation は、分散の平方根です。

3 方針

統計では、まず代表値として平均を見て、そのあと分散や標準偏差で散らばりを見ます。

大切なのは、計算を始める前に「この問題で知りたいのは中心か、散らばりか」を決めることです。平均だけでよい場面と、散らばりまで見ないと判断できない場面は明確に違います。

4 直感的な説明

平均点が同じでも、全員がほぼ同じ点数なのか、高得点と低得点に大きく割れているのかでは意味が違います。この「平均からの離れ」を数で表したものが分散です。

たとえば、A 組も B 組も平均点が 70 点だとしても、A 組が 68,70,72 のように集まっていて、B 組が 30,70,110 のように広がっているなら、学力の分布としては別物です。統計はこの違いを言葉ではなく数で表そうとします。

5 厳密な説明

5.1 1. 平均値

$x_1, \ldots, x_n$ の平均値は

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

です。

5.2 2. 分散

各値と平均値の差をそのまま足すと 0 になるので、二乗して

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

を考えます。これが分散です。

差を二乗するのは、正と負が打ち消し合うのを防ぐためであり、平均から遠い値ほど強く効かせるためでもあります。

5.3 3. 標準偏差

分散は二乗の単位を持つので、元の値の感覚へ戻したいときは平方根を取ります。これが標準偏差です。

5.4 4. 具体例

$1, 2, 3$ の平均値は 2 です。したがって分散は

$$\frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3} = \frac{2}{3}$$

です。

標準偏差は

$$\sqrt{\frac{2}{3}}$$

です。これで「平均からどれくらい離れているか」を、元の値と同じ尺度で読めます。

6 別の見方

平均は集団の重心と見てもよいです。分散は、その重心から各点がどれくらい散らばっているかを測る量です。この見方に立つと、平均と分散が別々の数ではなく、集団の中心と広がりを表す組として見えてきます。

7 見分け方

- 集団の中心を知りたいなら平均です。
- 散らばりや安定性を見たいなら分散や標準偏差です。
- 相関や回帰が出る前でも、まず平均と散らばりを押さえると全体像が見えます。
- 平均値が同じ 2 集団を比べる問題では、分散や標準偏差を見ないと違いを見落としやすいです。

8 最終形

$$\boxed{\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

$$\boxed{\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

9 一言でいうと

- 統計では、平均で中心を、分散で散らばりを見ます。

10 関連リンク

→ 講義 **確率と期待値** lecture math probability
https://study.bem130.com/lecture/math/probability/確率と期待値-講義/

→ 講義 **相関と回帰の入口** lecture math statistics
https://study.bem130.com/lecture/math/statistics/相関と回帰の入口-講義/