統計学の復習

crimsonbach

2004 年 8 月 23 日

標本平均

データの位置の指標に標本平均がある。標本 $X_1,\cdots,X_n$ は母集団分布に従う独立な確率変数で、これより標本平均を

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + \cdots + X_n}{n}. \tag{1}$$

と求められる。標本分散 $\overline{X}$ は母平均 μ を正しく推計する。

標本分散

次にデータの散らばり具合を示す指標、標本分散を復習する。平均の周りの偏差 $X_i - \overline{X}$、その 2 乗和

$$S_{xx} = \sum_{i=1}^{n}(X_i - \overline{X})^2.$$

から、標本分散は

$$s_x{}^2 = \frac{S_{xx}}{n-1} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}. \tag{2}$$

と推計される。標本平均に近い観測値が多いほど標本分散の値は小さくなり、標本平均から遠い観測値が多いほど標本分散は大きくなる。因みに、標本分散では平均の周りの 2 乗和を $n-1$ で除し、 $n-1$ は自由度という。自由度は自由に動ける変数の数である。

標本標準偏差

標本標準偏差は標本分散の平方根であり、

$$s_x = \sqrt{s_x{}^2} = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}. \tag{3}$$

と推計される。標本分散と同じくばらつきを表すが、標本分散の推計で二乗したものを X と同じ次元戻す。

標本共分散

変数 X と Y との相関的な散らばりを示す標本共分散を最後に定義する。偏差の積和

$$S_{xy} = \sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

によって、標本共分散は

$$s_{xy} = \frac{S_{xy}}{n-1} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n-1}. \tag{4}$$

で表される。また 2 変数が同時にとる値の平均と換言できる。