標準偏差と標準誤差：どちらを使うべきか？

このメモは筆者の学習用に作られたものですので，記述の正確性について保証することはできません．間違いを発見された方はご連絡ください．

卒業研究などで初めて本格的な実験を行うようになると，得られたデータを統計的にどう扱えば良いのか迷う時があると思います．今回は，標準偏差と標準誤差をどう使い分ければ良いのかという問題についてのメモしたいと思います．執筆にあたって文献 [1, 2] を大いに参考にさせていただきました．

1　標準偏差

まず標準偏差（standard deviation）とは何か考えてみましょう．ご存じのように，標準偏差はデータの分散の平方根をとったものです．

$$\text{分散：}V = \frac{\sum (X_i - \overline{X})^2}{(N-1)}$$

$$\text{標準偏差：}SD = \sqrt{V}$$

X_i は i 番目の実験によって得られたデータを，$\overline{X}$ はデータの算術平均値（標本平均と言います），N はデータのサンプルサイズを表しています．**標準偏差は得られたデータのバラつきの度合いを表す指標です．**ここまでは大丈夫だと思います．

> **♣ 注意**
> とはいえ，標準偏差はデータのバラつきを表す指標としていつも適切だとは限りません．**標準偏差がデータのバラつきを表す指標として適切なのは，そのデータの分布が正規分布に近い時に限ります．**逆にいえば，データの分布が偏っていて正規分布から離れているような場合には，標準偏差はそのデータのバラつきを表す指標として適切ではありません．またデータがこのような歪んだ分布の場合には，私たちが使い慣れている標本平均 $\overline{X}$ も，そのデータの代表値として適切ではありません．これは，$\overline{X}$ や標準偏差が外れ値に対して強く影響を受けるためです．このような場合には，データの代表値とバラつきの指標として，それぞれ中央値（median）と % 点を使うことになります．詳しくは文献 [1] などを参照してください．

さて，実験で得られたデータ系列が十分に大きく（サンプルサイズ N が十分に大きいという意味です），その分布が正規分布に近いとします．こういう時に限って $\overline{X} \pm 2SD$ の範囲に**データが入る確率**が 95% となります[*1]．これはどういうことか少し説明します．

*1 間違えてはいけないのは，$\overline{X} \pm 2SD$ **の範囲にデータの 95% が実際に入っているわけではない．**と言うことです．

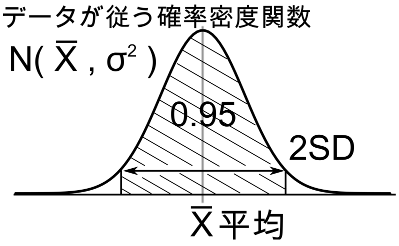

図 1　$\bar{X} \pm 2SD$ の面積が 0.95

今，データが正規分布に従って観測されるような場合を考えます（図 1）[*2][*3]．これは，データを確率変数だと考えた時に，その確率変数が従う確率密度関数が $\mathcal{N}(\bar{X}, \sigma^2)$ であることを言っています．つまり，実験を行うたびにデータの値は $\mathcal{N}(\bar{X}, \sigma^2)$ の確率密度関数に従って生成されるのです．確率変数とか確率密度関数とか，良くわからない単語が突然出てきましたので，おなじみのサイコロ投げで考えてみましょう．

サイコロのそれぞれの目が出る確率が $\frac{1}{6}$ だとします．サイコロを投げる時，偶数の目がでる確率は $\frac{1}{2}$ であると考えるのが最も合理的でしょう．これと全く同じことを言っているのです．つまり，何回も実験をした結果得られたデータの分布が，もし $\mathcal{N}(\bar{X}, \sigma^2)$ に近い形をしていたのならば，そのデータは $\mathcal{N}(\bar{X}, \sigma^2)$ の分布から生成されていたとみなすわけです．そうすると，

$$2\int_{\bar{X}}^{2SD} \mathcal{N}(\bar{X}, \sigma^2)dx = 0.95^{*4}$$

となりますから，$\bar{X} \pm 2SD$ の範囲にデータが入る確率が 95% と考えられるのです．

このように，もしデータの分布が正規分布に近い時には，標準偏差はデータのバラつき（範囲）を示す指標として非常に有効です．

2　標準誤差

次に標準誤差について考えてみましょう．標準誤差とは，標準偏差を $\sqrt{N}$ で割ったものです．

$$\text{標準誤差}：SE = \frac{SD}{\sqrt{N}}$$

式から明らかなように，標準誤差は標準偏差よりも**必ず小さくなります**．この性質のため，標準誤差を用いてグラフを作成したほうがデータのバラつきが少ないように見え，カッコよくなります．実際，多くの論文で標準誤差が使われているようです．ではなぜ標準偏差ではなく，標準誤差を使うのでしょうか？グラフをカッコ良く見せるためではありません．

ズバリその理由は，**標準誤差によって母集団の平均値の区間推定を行えるから**です．母集団の平均値の区間推定とはどういうことでしょうか？標準誤差の説明に入る前に，実験によってデータを得るとはどういうことか少し考えてみましょう．

2.1　データを得る．ということ

実験とは，それによっていくつかのデータを取得して，（我々が決して知りえない）**真の値を推定する行為**と言えます．実験によってデータを得る．と言うことを簡単に表した絵が図 2 です．

[*2] ” データが正規分布に従って観測される ” という意味は良く考えなければなりません．
[*3] 平均 $\bar{X}$ と分散 $V = \sigma^2$（よって $SD = \sigma$）の正規分布を $\mathcal{N}(\bar{X}, \sigma^2)$ と書きます．
[*4] これはご自分で確かめてみてください．正確には $\bar{X} \pm 1.96SD$ となるはずです．

例えば，東京都の住人の平均身長を知りたい場合を考えます．この場合，母集団は東京都のすべての住人となります．東京都の住人は有限のはずですから，その平均身長は厳密に存在するはずです．この**真の平均身長が，母平均** μ です．理論的には，東京都に住む人間のすべての身長がわかれば計算できるでしょう．

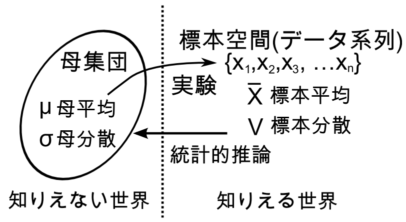

図2　データを得る．ということ

しかし，実際にすべての住人の身長を測定することができないときどうすればよいでしょうか？こういう時は，何人かの住人を選び出して身長を測定し，その平均（算術平均）を母平均 μ の推定量とするのが最も妥当でしょう．選び出した何人かの住人の身長の算術平均を，標本平均 $\bar{X}$ と呼びます．実験によってデータを取得してその平均を計算するという行為は，未知の母平均 μ を推定する行為なのです[*5]．

さて，すぐに気がつくように**標本平均 $\bar{X}$ は，母平均 μ とは違います**．標本平均 $\bar{X}$ は，母平均 μ の推定値なのですから，当然と言えば当然です．しかし違うということだけわかっていてもどうしようもないのですから，**母平均 μ は，どれくらいの幅の中にどれくらいの確からしさで存在していると推定できるのか**[*6]．ということが知りたいでしょう．それを叶えてくれるのが，標準誤差なのです．

2.2　標準誤差による母平均 μ の区間推定

今，母平均 μ，母分散 σ_0^2 の母集団を考えましょう．このような母集団から，i 回目の実験で観測されたデータを X_i とします．n 回の実験で得られたデータの標本平均 $\bar{X}$ は

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n}$$

となります．n が十分大きい時には，中心極限定理により

$$Z_n = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

は平均０，分散１の正規分布（標準正規分布 $\mathcal{N}(0,1)$）に従います[*7]．これはどういう意味かと言うと，実験を n 回行うとある一つの $\bar{X}$ を得るのですが，そうして得た $\bar{X}$ から Z_n^1 というものが計算できます．同じようにまた別に n 回の実験を行うと，ある別の Z_n^2 が得られます．このようにして何度も n 回の実験を m 回繰り返して $Z_n^1, Z_n^2, ..., Z_n^m$ を得ると，Z_n の分布は標準正規分布に従っているということです．つまり話を Z_n から $\bar{X}$ に戻すと，何度も $\bar{X}$ を得ると $\bar{X}$ の出かたはある正規分布に近づいていくということです.（実際には $\bar{X}$ を得る実験は一度しかしないのですから，そのようにして得られた $\bar{X}$ は，ある正規分布から生成された $\bar{X}$ の実現値だと考えられます.）

そうすると，標準正規分布の $x = -2$ から $x = 2$ までの積分は

$$\int_{-2}^{2} \mathcal{N}(0,1) dx > 0.95$$

[*5] 平均 $\bar{X}$ として，標本平均（算術平均）以外にもいくつか平均の取り方が存在します．例えば，トリム平均とか中央値とかいったモノです．母平均 μ の最良の推定量が何なのかという問題は難しいのですが．データの分布がわからないような場合，標本平均が最も良い推定量となることが知られています．詳しくは文献 [2] を参照してください.

[*6] これが**区間推定**です.

[*7] 中心極限定理まで説明すると非常に長くなってしまうのでやめますが，この議論が気になる方は確率論の本をなんでもいいから眺めてみてください.

ですから，Z_n が-2 から 2 の間

$$-2 < \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma_0} < 2$$

にある確率は，0.95 以上であることになります．少し式展開を行うと

$$\bar{X} - \frac{2\sigma_0}{\sqrt{n}} < \mu < \bar{X} + \frac{2\sigma_0}{\sqrt{n}}$$

となります．つまり，**母平均 μ が $\bar{X} - \frac{2\sigma_0}{\sqrt{n}}$ と $\bar{X} + \frac{2\sigma_0}{\sqrt{n}}$ の間にある確率が** 95% ということなのです．

さてこの式を良く見ると，標準誤差（$SE = \frac{\sigma}{\sqrt{n}}$）と同じ形が表れていますね．

母平均 μ は，95% **の確率で標本平均 $\bar{X}$ から $\pm 2SE$ の範囲に存在する．**と言えるのです*8.

このように，標準誤差によって母平均 μ の区間推定が行えるのです．

3　結局どちらを使えばいいの？

さて，ようやく標準偏差と標準誤差のどちらを使うべきなのか？という問いに対する答えが得られました．

> **問．標準誤差と標準偏差，どちらを使うべきか？**
> **答．何を推定したいのかによる．**

標準偏差とはデータのバラつきの指標でした．$\bar{X} \pm 2SD$ の範囲にデータが入る確率が 95% となるのです．ですから，データのバラつきを示したい，あるいは比べたい時は標準偏差を使うべきです．例えばある因子によって，データがバラつかなくなるというようなことを示したい場合などです．

一方で，**標準誤差とは，母平均 μ の区間推定量**でした．$\bar{X} \pm 2SE$ の範囲に母平均 μ がある確率が 95% となるのです．よって，母平均 μ の推定をしたい，あるいは比べたいという時は標準誤差を示せばいいのです．多くの場合，実験によって知りたいのは母平均 μ であるので，多くの論文で標準誤差が示されているのですね．

参考文献

[1] 浜田知久馬『学会・論文発表のための統計学』（真興交易，1999）

[2] 渡辺澄夫，村田昇『確率と統計ー情報学への架橋ー』（コロナ社,2005）

*8 ちょっと待ってください．ここでの σ_0 は母分散 σ_0^2 から得られたものだということを思い出してください．一方で SE の計算は，標本分散 σ^2 によって行われますよね．母分散 σ_0^2 は未知で，標本分散 σ^2 とは違うはずなのです．ここからの議論は，自分の中でまだ納得できていないのですが，おそらく標本分散 σ^2 を母分散 σ_0^2 の推定値としているのだと思います．つまり母平均 μ の区間推定のために，母分散 σ_0^2 の推定値である標本分散 σ^2 を使っているのですね．