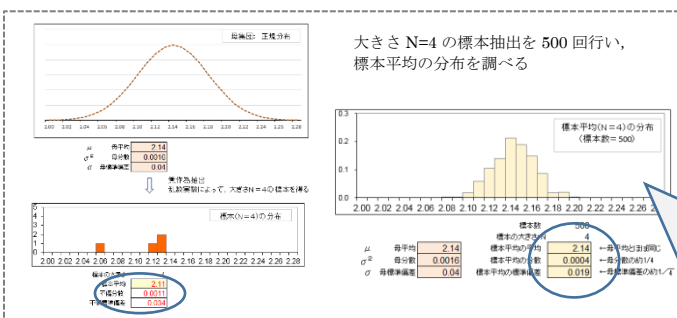


10月8日の推測統計の話は難しかったですね。普通の数学では、答えが1つに決まりますが、推測統計(「推定」)では「答えはa~bまでの範囲。なおかつ、数%の確率で答えは間違っていると思うけど、我慢してね」といった具合ですから、初学者にとって、その独特の考え方が理解しがたいのだらうと思います。

こう言うといい加減な学問に聞こえるかもしれませんが、しかし、推定に使うための“確率分布”(正規分布、t分布など)は、厳密な論理によって考えられたもので、難しい数学によって裏付けられています。統計ユーザーにとって、その数学的裏付けを理解することは必要ないですが、「推定」や「検定」において確率的に答えを導く手順を学び、その手順の中でどのように“確率分布”が使われているのかを理解することが必要です。

【演習6】2)と【演習7】を使って、“推測統計”基礎の基礎をマスターして欲しいと思います。下の解説と授業で使ったワークシートp8の解答例を参考に、自分の答えをもう一度見直してみてください。

【演習6】正規母集団の平均 μ を標本平均から推定する(定理1) …エクセル「正規母集団からの抽出実験」を使用



「定理1」を数学的に証明することは難しいので、乱数実験によるシミュレーションで確かめています。今回は標本平均を500個集めた分布で調べていますが、母集団の抽出を無限に繰り返せば、「標本平均の分布」は正規分布になります。また、この「標本平均の分布」の平均と分散は、「母集団の分布」の平均や分散から理論的に導くことができます。⇒これが「定理1」
 ここまでは、理論の話。このように「標本平均の分布」(正規分布)を理論的に知ることができるので、「手にした標本平均」を確率的に扱うことができるようになる訳です。実験者(統計ユーザー)は、1つの標本(今回は4つのデータ)と理論的な分布(今回は正規分布)から推定を行うという手順を理解しよう。

2) 母分散 $=0.0016$ (母標準偏差 $\sigma=0.04$)であることが分かっているとして、得られた標本平均から母正規分布の平均 μ を、信頼度95%で区間推定せよ。ただし、正規分布において区間 $(\mu - 2\sigma, \mu + 2\sigma)$ の区間確率は0.95であるとする。

標本平均 \bar{x}	乱数実験の結果を記入
標本平均の分布の分散 (定理1を使って計算)	0.0004

<解説>

【定理1】が成立つならば、標本平均($n=4$)の分布の分散は $\sigma^2=0.0004$ 。この値の平方根をとって、この分布の標準偏差 $\sigma=0.02$ 。

正規分布の性質より、ある標本平均 \bar{x} (毎回違います)は、95%の確率で、区間 $(\mu - 2 \times 0.02, \mu + 2 \times 0.02)$ に入るはず。

$$\therefore \mu - 2 \times 0.02 < \bar{x} < \mu + 2 \times 0.02 \quad \bar{x} - 2 \times 0.02 < \mu < \bar{x} + 2 \times 0.02$$

※ 答えは、上の式に \bar{x} の値を代入して求めて下さい

【定理1】「正規分布する母集団(平均 μ 分散 σ^2)から、標本サイズ n の標本抽出を無限回行い、各標本の平均を無限個集めた分布(理論的につくる確率分布)をつくと、平均 μ 分散 σ^2/n の正規分布となる」

⇒ $\sigma^2=0.0016, n=4$ を上式に代入。 $0.0016/4=0.0004$

【演習7】正規母集団の平均 μ を、t分布を用いて推定する(母標準偏差が未知の場合)

【演習6】2)のように母分散(母標準偏差)が既知でなければ、“標本平均の分布の分散”を求めることは出来ない。この場合は、t分布を用いて推定する。ただし、t分布は標本の大きさ n ごとに定まり、 $n=4$ のt分布において区間 $(-3.2 < t < 3.2)$ の区間確率は0.95であるとする。

t分布については、「事前プリント その2」に簡単な説明があります

t値は $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ と定義されるので、推定する母平均は $\mu = \bar{x} - t \frac{s}{\sqrt{n}}$ となる。

標本の大きさ n	4
標本平均 \bar{x}	乱数実験の結果を記入
不偏標準偏差 s	乱数実験の結果を記入

<解説>

【演習6】2)では、「標本平均の分布」として、【定理1】に基づいた「正規分布」を考えましたが、今回は母集団の σ が未知なのでこの「正規分布」を使って推定することができません。そこで、新たな理論的な確率分布としてt分布が登場しました(t分布について、詳しく分からなくても、大丈夫です)。t分布も、正規分布と同様に、ある区間の確率が求められており、

本などで調べることができます(統計学者の先生、ありがとう)。ここで用いるt値は、 $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ と定義されており、 n, \bar{x}, s と't値の範囲'はすでに手に入れている(上の表と問題文)のですから、求めるべき μ の範囲も決まりますね。答えは、ワークシートp8の解答例で確認を…。