雑感

標本平均の扱い

■　数 B の統計分野の履修が増えている中，理解の難しい内容の 1 つが「標本平均の分布」ではないかと推察する．「公式」的に内容を示せば，

> 母平均 m，母標準偏差σの母集団から大きさ n の標本を無作為に復元抽出すると，標本平均$\overline{X}$の期待値は$E(\overline{X}) = m$，標準偏差は$\sigma(\overline{X}) = \sigma/\sqrt{n}$

■　証明は，$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$に対して，$X_1, X_2, \cdots, X_n$が母集団から取り出された独立な確率変数であることから,$E(\bar{X}) = \frac{1}{n}\sum_{k=1}^{n} E\ X_k\ = \frac{1}{n}\ nm\ = m$,
$V(\bar{X}) = \frac{1}{n^2}\sum_{k=1}^{n} V\ X_k\ = \frac{1}{n^2}\ n\sigma^2\ = \frac{\sigma^2}{n}$ とするだけの，何とも素っ気ない．

■　しかし，この結果や証明を見ても，期待値部分はともかく，標準偏差に至ってはその意味するところを諒解できる生徒は多くないだろう．
例の質問箱には，「母集団から 5 個の標本を選んできたとき，その 5 個の期待値(平均)が m といつも同じなのか；その 5 個の標準偏差は一々計算しなくても$\sigma/\sqrt{5}$になるのか」といったニュアンスの質問が幾つか寄せられる．

■　ここで大事だと思われるのは，分かり易く期待値で述べれば
$E(\overline{X}) = \frac{1}{n}\sum_{k=1}^{n} X_k$としているが，もともとの期待値の定義式的に言えば，大きさ n の標本の考え得る全標本 N 個の$\overline{X_j} = \frac{1}{n}(X_{j,1} + X_{j,2} + \cdots + X_{j,n})$
$(1 \leq j \leq N)$に対して, $E(\overline{X}) = \frac{1}{N}\sum_{j=1}^{N} \overline{X_j}$ ということである．
したがって，例えば母集団が 5 個の変量であり，そこから 2 個の標本を復元抽出するとしたら，標本平均は$5^2 = 25$通り考えられ，その 25 個の標本平均の期待値，標準偏差が，ここで問題となっているところの「標本平均の期待値と標準偏差」なのだということである．
上の質問は，この考え得る全標本平均のうちの 1 つについての話だから，この「公式」で対応できる話ではない．標本の選び方次第で変わった値になると言うことである．
「**大きさ n の 1 つの標本平均は，平均が m, 標準偏差が $\sigma/\sqrt{n}$ の集団から取り出された値**」という理解と認識が，以降の「母平均の推定」の学習にとって大切であるように思う．

■　100 人のテストの点数の平均を推定するとき，100 人の中から 2 人選んで 2 人の平均を考えて推定するよりも，100 人の中から 10 人を選んで 10 人の平均から推定した方が全体の平均との誤差が小さい可能性が高いと考えられるが，そのことが標準偏差の大小関係 $\sigma/\sqrt{2} > \sigma/\sqrt{10}$に結びついていると考えると，この分母の$\sqrt{n}$ の働きが分かってくる．

■　さて，教科書はこの標本平均の期待値と標準偏差について，どのように説明しているのだろうか．調べてみると，大きく次の 3 つのタイプに分類できた（出版されている全教科書を調べた）．
タイプ A：上の証明と公式，それを当てはめる例で済ませているもの
タイプ B：証明抜きで「知られている」として公式を載せているもの
タイプ C：上の公式が成り立っている例（考え得る全標本を考えている）
　　　　　を載せ，証明と公式を載せているもの[東書]（※）
であり，大半は A のタイプである．
A タイプの中には具体的な実験結果を載せて，この公式の説明も加えようという努力をしているようなものもある[啓林 708]が, 全標本ではないため，当然だが公式通りの値になっていない（分散は，理論値との差が結構大きい実験結果）．この公式は「n が大きければ…」などという公式ではなく，「≒」でもなく，きっちり「＝」の公式なのだが，誤解しそうだ．
このような教科書の記述状況では，生徒が「よく分からない」のも頷ける．

※　その教科書とは別の例を挙げる：母集団を $\{0,\ 0,\ 2,\ 2,\ 6\}$ とすると，
$m = 2, \sigma^2 = 24/5$ である. ここから大きさ 2 の標本を取り出すと, 右の表から下の分布表が出来る．これより計算すると$E(\bar{X}) = 2, V(\bar{X}) = 12/5 = \sigma^2/2$,となっていることが確認できる. きっちり「＝」なのである.

	0	0	2	2	6
0	0	0	1	1	3
0	0	0	1	1	3
2	1	1	2	2	4
2	1	1	2	2	4
6	3	3	4	4	6

$\overline{X}$	0	1	2	3	4	6	計
P	$\frac{4}{25}$	$\frac{8}{25}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{1}{25}$	1

■　＜余談＞これを書くにあたって数 B の教科書の該当部分をすべて閲覧したが，学校現場を離れた今，教科書の閲覧に苦労した．教科書採択時期にのみ開かれる教科書展示会を訪れたのだが，高校の教科書を置く展示会場は限られ，だいぶ足を延ばす必要があった．複写が出来ないのも辛い．

2024 年 10 月 1 日