雑感　合併 2 集団の分散

■「データの分析」が 2015 年からのセンター試験に出題される．大学入試センターが 2013 年 11 月に「試作問題」を 1 題だけ公表したが，巷に，この分野の問題は少ない．

一般社団法人統計質保証推進員会が，2011 年から「統計検定」を実施しており，その 3 級，2 級の内容に，この「データの分析」の内容が含まれる．

■ 2012 年と 2013 年に行われた統計検定の問題から「データの分析」に該当する問題をピックアップし，（著作権へ配慮して）その類題を生徒向けの練習問題用に 12 題作ってみた．

2014 年 5 月の名古屋の気温や 2013 年に発生した 31 個の台風など最新の気象データなども用いたため，生徒が関心を示してくれそうな予感があるがどうだろうか．

■ところで，2013 年 11 月 17 日に行われた統計検定第 3 級の問 12 に次のような問題がある．

ある学習塾には A 組と B 組の 2 クラスがある。先日行った数学の模擬試験の結果は次の通りであった。

A 組：受験者 19 名　平均点 65.0 点　標準偏差 9.6 点

B 組：受験者 25 名　平均点 57.0 点　標準偏差 14.4 点

〔2〕2 つのクラスをまとめた 44 名のこの試験の標準偏差の求め方として適切なものを，次の①～⑤ のうちから一つ選べ。

① $(9.6+14.4)\div 2$

② $\sqrt{(9.6\times 19+14.4\times 25)\div 44}$

③ $\sqrt{(9.6^2\times 19+14.4^2\times 25)\div 44}$

④ $\sqrt{9.6^2\times 19+14.4^2\times 25+(65.0-60.5)^2\times 19+(57.0-60.5)^2\times 25}$

⑤ $\sqrt{\{9.6^2\times 19+14.4^2\times 25+(65.0-60.5)^2\times 19+(57.0-60.5)^2\times 25\}\div 44}$

■選択肢があるので，⑤あたりだなと見当はつくものの，こんな「公式」は知らない．ネットで少し探してみたが，探し方が悪いためだろうが「公式」は見つからなかった．

となれば，導くしかない．

集団 X ：データ数 m，平均値 $\overline{x}$，標準偏差 σ_X

集団 Y ：データ数 n，平均値 $\overline{y}$，標準偏差 σ_Y

とし，2 集団 X, Y を合併した集団 Z に対して，

集団 Z ：データ数 $m+n$，平均値 M，標準偏差 σ

とするとき，

$$\sigma^2=\frac{1}{m+n}\left\{\sigma_X{}^2 m+\sigma_Y{}^2 n+(\overline{x}-M)^2 m+(\overline{y}-M)^2 n\right\}$$

を示す．

■まず，$\sigma^2=\dfrac{1}{m+n}\left(\sum_{k=1}^{m}(x_k-M)^2+\sum_{k=1}^{n}(y_k-M)^2\right)$ である．

$\sum_{k=1}^{m}x_k+\sum_{k=1}^{n}y_k=M(m+n)$，$\sigma_X{}^2=\dfrac{1}{m}\sum_{k=1}^{m}(x_k)^2-(\overline{x})^2$ から

$\sum_{k=1}^{m}(x_k)^2=\sigma_X{}^2 m+(\overline{x})^2 m$　などより

$$\begin{aligned}(m+n)\sigma^2&=\sum_{k=1}^{m}(x_k-M)^2+\sum_{k=1}^{n}(y_k-M)^2\\&=\sum_{k=1}^{m}(x_k)^2+\sum_{k=1}^{n}(y_k)^2-M^2(m+n)\\&=\sigma_X{}^2 m+\sigma_Y{}^2 n+(\overline{x})^2 m+(\overline{y})^2 n-M^2(m+n)\\&=\sigma_X{}^2 m+\sigma_Y{}^2 n+\{(\overline{x})^2-M^2\}m+\{(\overline{y})^2-M^2\}n\end{aligned}$$

示すべき式に近づいたが，ビミョーに異なる．

しかし，実用上これで十分なような気がする．

もう一息だが，示すべき式が分かっていることなので，卑怯だが，その形に強引に持ち込む．

ここで，

$$\begin{aligned}&\{(\overline{x})^2-M^2\}m+\{(\overline{y})^2-M^2\}n\\&=\{(\overline{x})^2-2\overline{x}M+M^2\}m+\{(\overline{y})^2-2\overline{y}M+M^2\}n\\&\qquad+2\overline{x}Mm+2\overline{y}Mn-2M^2(m+n)\\&=(\overline{x}-M)^2 m+(\overline{y}-M)^2 n+2M(m+n)M-2M^2(m+n)\\&=(\overline{x}-M)^2 m+(\overline{y}-M)^2 n\end{aligned}$$

となるから，

$$(m+n)\sigma^2==\sigma_X{}^2 m+\sigma_Y{}^2 n+(\overline{x}-M)^2 m+(\overline{y}-M)^2 n$$

となって，

$$\sigma^2=\frac{1}{m+n}\left\{\sigma_X{}^2 m+\sigma_Y{}^2 n+(\overline{x}-M)^2 m+(\overline{y}-M)^2 n\right\}$$

が示された．

■ニーズがないとは言えないが，こんな「公式」は覚えたくないなあ．

■書棚を探していたら『現代の綜合数学Ⅲ』（安藤洋美，現代数学社，1973）の p.192-193，例 3.17 にこの内容が扱われているのを見つけた．ただ，結果の式にミスプリントがある．

ここでの記号にしたがって記せば

$$\sigma^2=\frac{1}{m+n}\left\{\sigma_X{}^2 m+\sigma_Y{}^2 n+\frac{mn}{m+n}(\overline{x}-\overline{y})^2\right\}$$

である．

ミスプリントは，｛　｝内の $\dfrac{mn}{m+n}(\overline{x}-\overline{y})^2$ の前の符号が－になってしまっている点である．

このように見てくると，

$$\begin{aligned}&\{(\overline{x})^2-M^2\}m+\{(\overline{y})^2-M^2\}n=(\overline{x}-M)^2 m+(\overline{y}-M)^2 n\\&=\frac{mn}{m+n}(\overline{x}-\overline{y})^2\end{aligned}$$

が成り立つことになる．

さまざまな顔を持つ，妖怪みたいな式である．