確率分布の授業のためのプリント3点　　山田

　2学期から確率分布の授業を初めて担当することになり、準備をしていたら自分の理解不足や、教材への疑問もあったので、プリントにしてみました。

①生徒にこんな質問をされたら？

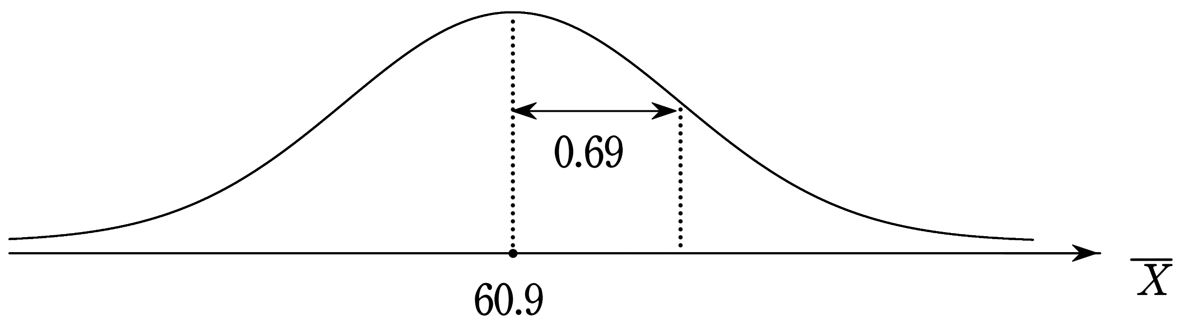
> 母平均 $\boldsymbol{m}$, 母標準偏差 $\boldsymbol{\sigma}$ の母集団から大きさ $\boldsymbol{n}$ の無作為標本を抽出するとき、その標本平均 $\overline{X}$ の期待値 $\boldsymbol{E}(\overline{X})$ と標準偏差 $\boldsymbol{\sigma}(\overline{X})$ は
>
> $\boldsymbol{E}(\overline{X})=\boldsymbol{m}$　,　$\boldsymbol{\sigma}(\overline{X})=\dfrac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}$　　数研出版　「高等学校　数学B」

［4プロセス数学Ｂ　問題146］

ある県の 16 歳の男子学生の体重は，平均値 60.9 kg，標準偏差 6.9 kg である。この母集団から無作為に 100 人からなる標本を取り出すとき，その標本平均 $\overline{X}$ の期待値と標準偏差を求めよ。

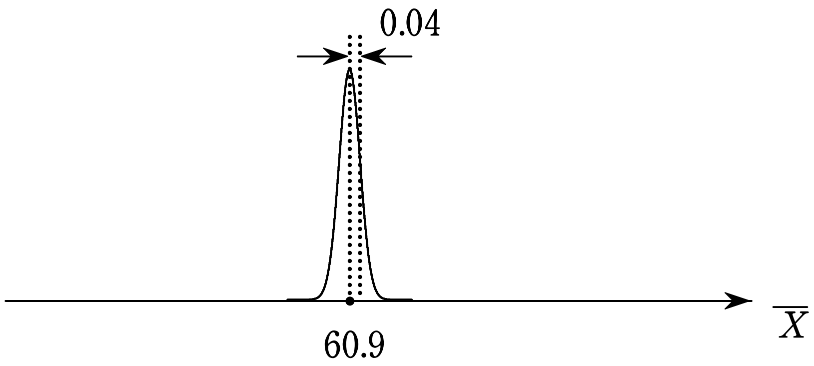
$\boxed{\text{解答}}$　母平均 60.9，母標準偏差 6.9 であるから

$$E(\overline{X})=60.9\ (\mathrm{kg}),\ \ \sigma(\overline{X})=\frac{6.9}{\sqrt{100}}=0.69\ (\mathrm{kg})$$

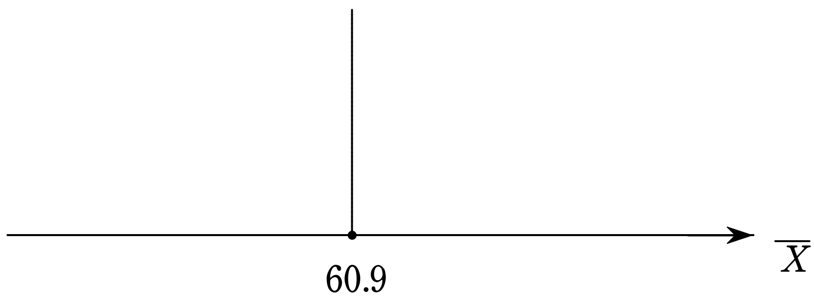


$\boxed{\text{質問}}$　この問題の県全体の16 歳の男子学生が3万人であるとする（愛知県の数とほぼ同じ）。このとき、標本数を3万とすると、

$$E(\overline{X})=60.9\ (\mathrm{kg}),\ \ \sigma(\overline{X})=\frac{6.9}{\sqrt{30000}}=\frac{6.9}{100\sqrt{3}}\fallingdotseq 0.04\ (\mathrm{kg})$$



ところが標本数は母集団の数と同じだから、
標本平均＝母平均であり、 $\sigma(\overline{X})=0$　となるのではないか？



$$\boldsymbol{\sigma(\overline{X})=\frac{\sigma}{\sqrt{n}}}$$ （nは標本の個数）はどんなときに成り立つか？

$\boxed{例}$　母集団が3人のとき、体重 $\{50kg, 60kg, 70kg\}$ につて、

母平均 $m=\dfrac{50+60+70}{3}=60$

母分散 $\sigma^2=\dfrac{(50-60)^2+(60-60)^2+(70-60)^2}{3}=\dfrac{200}{3}$

母標準偏差 $\sigma=\dfrac{10\sqrt{2}}{\sqrt{3}}$

ここから2個の標本 X_1, X_2 をとる。（$n=2$）

$\boxed{1}$ 非復元抽出のとき、確率分布は

(X_1, X_2)	$(50, 60)$ $(60, 50)$	$(50, 70)$ $(70, 50)$	$(60, 70)$ $(70, 60)$
$\overline{X}$	55	60	65
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

標本平均 $\overline{X}$ の期待値 $E(\overline{X})$ は、

$$E(\overline{X})=\frac{1}{3}\cdot 55+\frac{1}{3}\cdot 60+\frac{1}{3}\cdot 65=60$$

標本平均 $\overline{X}$ の分散 は、

$$\sigma^2(\overline{X})=\frac{1}{3}(55-60)^2+\frac{1}{3}(60-60)^2+\frac{1}{3}(65-60)^2=\frac{50}{3}$$

標準偏差は

$$\sigma(\overline{X})=\frac{5\sqrt{2}}{\sqrt{3}}\neq\frac{\sigma}{\sqrt{2}} \quad \Longrightarrow \sigma(\overline{X})=\frac{\sigma}{\sqrt{n}}$$ は成り立たない

$\boxed{2}$ 復元抽出のとき、確率分布は

(X_1, X_2)	$(50, 50)$	$(60, 60)$	$(70, 70)$	$(50, 60)$ $(60, 50)$	$(50, 70)$ $(70, 50)$	$(60, 70)$ $(70, 60)$
$\overline{X}$	50	60	70	55	60	65
P	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{2}{9}$

標本平均 $\overline{X}$ の期待値 $E(\overline{X})$ は、

$$E(\overline{X})=\frac{1}{9}\cdot 50+\frac{1}{9}\cdot 60+\frac{1}{9}\cdot 70+\frac{2}{9}\cdot 55+\frac{2}{9}\cdot 60+\frac{2}{9}\cdot 65=60$$

標本平均 $\overline{X}$ の分散 は、

$$\sigma^2(\overline{X})=\frac{1}{9}(50-60)^2+\frac{1}{9}(60-60)^2+\frac{1}{9}(70-60)^2+\frac{2}{9}(55-60)^2+\frac{2}{9}(60-60)^2$$

$$+\frac{2}{9}(65-60)^2 \quad =\frac{100}{3}$$

標準偏差は　$\sigma(\overline{X})=\dfrac{10}{\sqrt{3}}=\dfrac{\sigma}{\sqrt{2}} \Longrightarrow \sigma(\overline{X})=\dfrac{\sigma}{\sqrt{n}}$ は成り立つ

$\boxed{1}\boxed{2}$より、

$$\boxed{\sigma(\overline{X})=\frac{\sigma}{\sqrt{n}} \quad \text{は復元抽出のときに成り立つ}}$$

（理由）　復元抽出の場合は標本 $X_1, X_2, \cdots X_n$ は独立。

よって　$\sigma^2\left(\dfrac{X_1+X_2+\cdots X_n}{n}\right)=\dfrac{1}{n^2}\{\sigma^2(X_1)+\sigma^2(X_2)+\cdots\sigma^2(X_n)\}$

$=\dfrac{n\sigma}{n^2}=\dfrac{\sigma}{n}$　が成り立つ

質問の答

問題146のように 無作為に 100 人からなる標本を取り出すとき，生徒は全て異なるから非復元抽出であるが、母集団の数が大きいから例えば100人とも60.0 kgとなることもありうる。よって復元抽出とみなしても計算に大きな差はない。しかし標本が3万人のとき、全て60.0kgとなることは有り得ないから復元抽出とみなせず、$\sigma(\overline{X})=\dfrac{\sigma}{\sqrt{n}}$ は成り立たない 。

教科書では、この理屈が、本文を注意深く読めば分かるように書いてあるが、見落としやすい。

実際、最初の質問　に対して、「n が大きくなれば $\dfrac{\sigma}{\sqrt{n}} \to 0$ だから問題ない」と誤解する人が多いのではないか。以上のことから、冒頭の教科書の説明は、次のように書いた方が混乱が少ないと思う。

> 母平均 m, 母標準偏差 σ の母集団から大きさ n の無作為標本を復元抽出するとき、その標本平均 $\overline{X}$ の期待値 $E(\overline{X})$ と標準偏差 $\sigma(\overline{X})$ は
>
> $E(\overline{X})=m \quad , \quad \sigma(\overline{X})=\dfrac{\sigma}{\sqrt{n}}$
>
> ただし、非復元抽出のときも、標本の大きさ n に比べて
> 母集団の大きさが十分大きいときは、復元抽出とみなしてよい。

②元に戻さないくじ引き

［4プロセス数学Ｂ　問題121］

50 本のくじの中に 20 本の当たりくじがある。このくじから 10 本のくじを続けて取り出すとき，その中の当たりくじの本数を Y とする。確率変数 Y の期待値を求めよ。ただし，取り出したくじはもとにもどさないとする。

解答

$i=1,\ 2,\ \cdots\cdots,\ 10$ に対して，i 番目に取り出したくじが，

当たりくじのとき　$X_i=1$　　当たりくじでないとき　$X_i=0$

とすると，$Y=X_1+X_2+\cdots\cdots+X_{10}$ である。

1 本ずつ引くくじ引きにおいて，当たりくじを引く確率，およびはずれくじを引く確率はくじを引く順番に関係なく，それぞれ一定であるから，　…①

$i=1,\ 2,\ \cdots\cdots,\ 10$ の各場合に　$P(X_i=1)=\dfrac{20}{50}=\dfrac{2}{5}$，$P(X_i=0)=\dfrac{3}{5}$

よって　$E(X_i)=1\cdot\dfrac{2}{5}+0\cdot\dfrac{3}{5}=\dfrac{2}{5}$

したがって　$E(Y)=E(X_1)+E(X_2)+\cdots\cdots+E(X_{10})=10\cdot\dfrac{2}{5}=4$

この解答への疑問　　上記①を、説明なしで利用してよいのか？

例えば以下のように生徒には確認させた方がよい。

（例）当たりくじ2本、はずれくじ1本の計3本から、A,B,Cの順に1本ずつ元に戻さずに引くとき、A,B,Cのあたる確率を求める。

（解）当たりが①②、はずれが●として、A,B,Cの引き方は３！＝６通りで、
全て書き出してみる。

①②●
①●②
②①●
②●①
●①②
●②①

このうち、①が１，２，３番目のどこにくるかは、平等で、2通りずつ。
②も同じ。

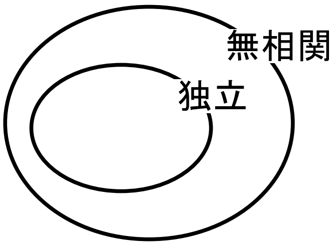
よってA,B,Cの当たる確率はすべて $\dfrac{4}{6}=\dfrac{2}{3}$

くじの本数や引く人数が変わっても同様だから、上記①は正しい。

③　独立と無相関は同じではない

（1）　XとYが独立 $\Longleftrightarrow$ $P(X=a\,,\,Y=b)=P(X=a)\,P(Y=b)$
　が $a\,,\,b$ のとり方に関係なく常に成り立つ

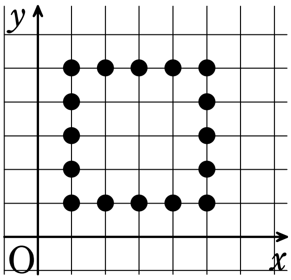
（2）　XとYが独立 $\Rightarrow$ $E(XY)=E(X)\,E(Y)$



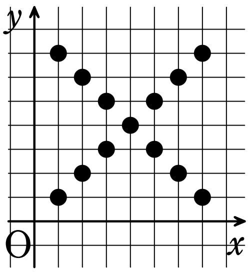
この定理(1)(2)の関係が、わかりにくい。ところで
数学Ⅰ「データの分析」の「相関関係」で学ぶ
共分散 $\sigma_{xy}=\overline{xy}-\overline{x}\ \overline{y}$ を使うと $E(XY)=E(X)\,E(Y)$ は $\overline{xy}=\overline{x}\ \overline{y}$ 　だから
xとyが独立$\Rightarrow\sigma_{xy}=\overline{xy}-\overline{x}\ \overline{y}=0$ 　（x と y は相関関係が無い：無相関）
と考えればイメージしやすい。ここで注意すべきは、「独立」と「無相関」の関係（上図）である。

x と y は無相関だが、独立ではない例（$\boxed{1}$～$\boxed{5}$ はいずれも $\sigma_{xy}=0$ ）

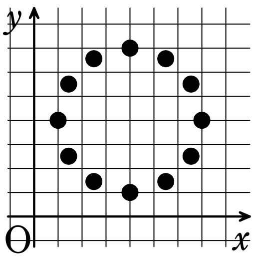
$\boxed{1}$



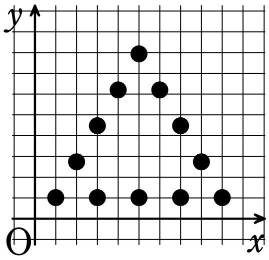
$\boxed{2}$



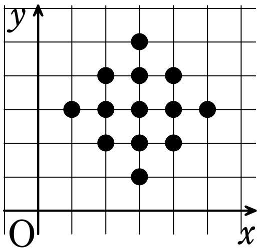
$\boxed{3}$



$\boxed{4}$



$\boxed{5}$



$\boxed{1}$ では、$\sigma_{xy}=\overline{xy}-\overline{x}\ \overline{y}=0$ で無相関であるが、

$P(x=1)=\dfrac{5}{16}\,,P(y=1)=\dfrac{5}{16}\,,\ P(x=1\,,\,y=1)=\dfrac{1}{16}$

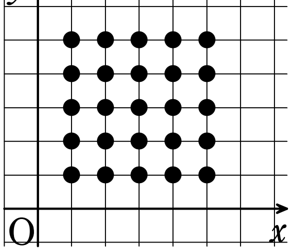
$P(x=1\,,\,y=1)\neq P(x=1)\,P(y=1)$ より独立でない。$\boxed{2}$～$\boxed{5}$ も同様。

「無相関」=「x と y に直線的関係がない」であり、「独立」=「x と y に関係がない」は意味が異なる。

$\boxed{1}$～$\boxed{5}$は　「直線的関係」はないが、「関係」はあるから、独立ではない。

x と y は無相関で、独立でもある例（$\boxed{6}$）。

$\boxed{6}$



$\sigma_{xy}=\overline{xy}-\overline{x}\ \overline{y}=0$ より無相関。

また、$P(x=1)=\dfrac{1}{5}\,,P(y=1)=\dfrac{1}{5}\,,$

$P(x=1\,,\,y=1)=\dfrac{1}{25}$ 　で、

$P(x=1\,,\,y=1)=P(x=1)\,P(y=1)$ より独立である。