太郎さんと花子さんは，社会生活基本調査の集計結果に，「睡眠」,「食事」,「通勤・通学」,「移動 (通勤・通学を除く)」などの 20 種類の行動それぞれについての総平均時間と行動者平均時間が，47 都道府県別に集計されていることを知った．

用語の説明

- 総平均時間……ある行動に費やした時間の調査対象者全員についての平均値 (分)
- 行動者平均時間……ある行動に費やした時間の調査対象者全員から、その行動に費やした時間が 0 分の人を除いた調査対象者についての平均値 (分)

例えば,「通勤・通学」に費やした時間 (分) が

75，0，90，60，0

であったとき，これらの平均値 $\dfrac{75+0+90+60+0}{5}=45$ が総平均時間であり，値が 0 である二つを除いた 75，90，60 の平均値 $\dfrac{75+90+60}{3}=75$ が行動者平均時間である．

ここでは，平日における 15 歳以上を対象とした集計結果を用いて，都道府県ごとに値を算出している．

なお，以下の図や表については、総務省の Web ページをもとに作成している．

(1) 太郎さんと花子さんは,「通勤・通学」に費やした時間について調べることにした．図 1 と図 2 はそれぞれ，令和 3 年の「通勤・通学」の総平均時間と行動者平均時間のデータをヒストグラムに表したものである．以下，ヒストグラムの各階級の区間は，左側の数値を含み，右側の数値を含まない．

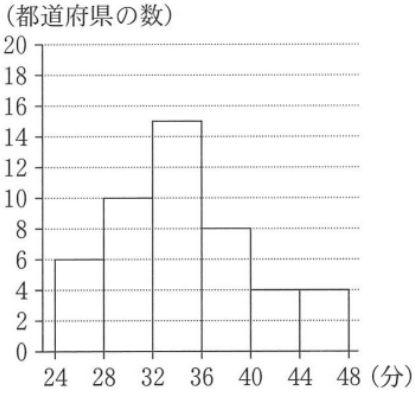

図 1　令和 3 年の「通勤・通学」の総平均時間のヒストグラム

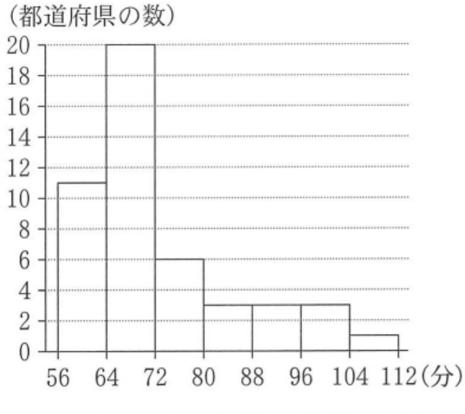

図 2　令和 3 年の「通勤・通学」の行動者平均時間のヒストグラム

(i) 図 1 から，令和 3 年の「通勤・通学」の総平均時間の最頻値は $\boxed{\text{アイ}}$ であり，同様に図 2 から，行動者平均時間の最頻値は $\boxed{\text{ウエ}}$ である．

(ii) 図 1 のヒストグラムに関して，各階級に含まれるデータの値がすべてその階級値に等しいと仮定する．このとき，令和 3 年の「通勤・通学」の総平均時間の平均値を m とすると

$$\boxed{\text{オカ}} \leqq m < \boxed{\text{オカ}} + 1$$

である．

(iii) 次に，太郎さんと花子さんは，平成 28 年と令和 3 年の「通勤・通学」に費やした時間を比較することにした．

図 3 は，平成 28 年の総平均時間，令和 3 年の総平均時間，平成 28 年の行動者平均時間，令和 3 年の行動者平均時間の箱ひげ図を並べたものである．

ここで，あるデータにおける最大値から第 3 四分位数を引いた値を H とする．そして平成 28 年の総平均時間，令和 3 年の総平均時間，平成 28 年の行動者平均時間，令和 3 年の行動者平均時間におけるを，それぞれ H_1，H_2，H_3，H_4 とする．

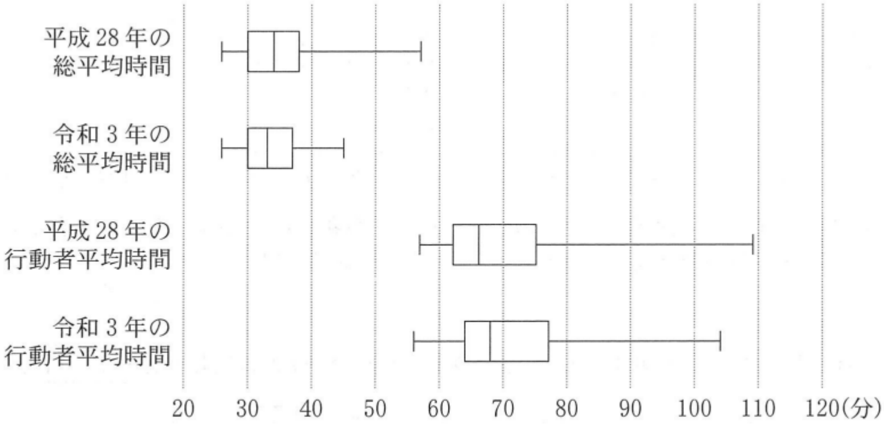

図 3　平成 28 年と令和 3 年の「通勤・通学」の総平均時間と行動者平均時間の箱ひげ図

次の (a)，(b)，(c) は，図 3 に関する記述である．

(a) 令和 3 年の総平均時間の最大値は，令和 3 年の行動者平均時間の最小値より小さい．

(b) 平成 28 年の総平均時間の四分位範囲は，平成 28 年の行動者平均時間の四分位範囲より小さい．

(c) 総平均時間と行動者平均時間それぞれの，平成 28 年と令和 3 年の H の変化を比較すると，$\dfrac{H_2}{H_1} > \dfrac{H_4}{H_3}$ となる．

(a)，(b)，(c) の正誤の組合せとして正しいものは $\boxed{キ}$ である．

$\boxed{キ}$ の解答群

	⓪	①	②	③	④	⑤	⑥	⑦
(a)	正	正	正	正	誤	誤	誤	誤
(b)	正	正	誤	誤	正	正	誤	誤
(c)	正	誤	正	誤	正	誤	正	誤

(2) 太郎さんと花子さんは，令和 3 年における「通勤・通学」と「移動（通勤・通学を除く）」(以下,「移動」) に関し，それぞれの行動に費やした総平均時間と行動者平均時間の関係について話をしている．

太郎：通勤の途中で，ふだんの経路を大きくはずれて買い物に行ったり病院に行ったりする人もいるけど，こうした行動は「通勤・通学」ではなく，「移動」になるね.「通勤・通学」に費やした時間が長いほど,「移動」に費やした時間は長いのかな.

花子：じゃあ，それぞれに費やした時間の関係を調べてみようよ.

図 4 と図 5 は「通勤・通学」と「移動」の総平均時間と行動者平均時間の散布図であり，図中の黒丸は，二つの点が完全に重なっていることを表している．なお，三つ以上の点が完全に重なっていることはない．ただし，図 4 と図 5 において，同じアルファベットを付している点は，同じ都道府県であることを表している.

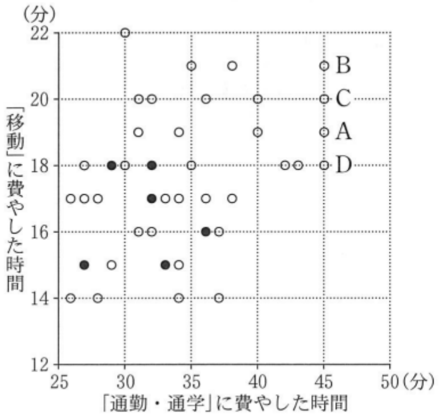

図 4　「通勤・通学」と「移動」の総平均時間の散布図

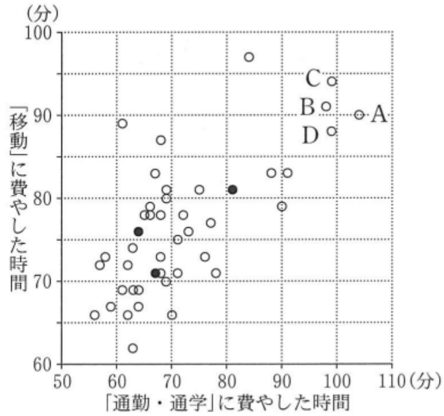

図 5　「通勤・通学」と「移動」の行動者平均時間の散布図

(i) 図 5 から,「通勤・通学」の行動者平均時間が 60 以下で，かつ「移動」の行動者平均時間が 75 以下である都道府県の数は $\boxed{\text{ク}}$ である.

(ii) 図 4 における四つの点 A，B，C，D が表す都道府県では,「通勤・通学」の総平均時間が同じ値であるが，図 5 では「通勤・通学」の行動者平均時間について，点 A が表す都道府県の値は他の三つのどの都道府県の値よりも大きくなっていることがわかる．このようになるのは，$\boxed{\boxed{\text{ケ}}}$ からである.

$\boxed{\boxed{\text{ケ}}}$ については，最も適当なものを，次の⓪～③のうちから一つ選べ.

⓪　点 A が表す都道府県の「通勤・通学」に費やした時間が 0 である人数の割合が，他の三つのどの都道府県の割合よりも大きい

①　点 A が表す都道府県の「通勤・通学」に費やした時間が 0 である人数の割合が，他の三つのどの都道府県の割合よりも小さい

②　点 A が表す都道府県の「移動」に費やした時間が 0 である人数の割合が，他の三つのどの都道府県の割合よりも大きい

③　点 A が表す都道府県の「移動」に費やした時間が 0 である人数の割合が，他の三つのどの都道府県の割合よりも小さい

(iii) 太郎さんと花子さんは，総平均時間と行動者平均時間のそれぞれの相関関係について調べることにした.

「通勤・通学」と「移動」の総平均時間の相関係数は 0.36 であった.「通勤・通学」と「移動」の行動者平均時間の相関係数を計算するために，表 1 のように平均値，標準偏差および共分散を求めた.

表 1 「通勤・通学」と「移動」の行動者平均時間の平均値，標準偏差，共分散

	平均値	標準偏差	共分散
「通勤・通学」の行動者平均時間	71.8	11.8	64.4
「移動」の行動者平均時間	76.6	7.9	

表 1 を用いると,「通勤・通学」と「移動」の行動者平均時間の相関係数は $\boxed{\text{コ}}$ である.

$\boxed{\text{コ}}$ については，最も適当なものを，次の⓪〜⑦のうちから一つ選べ.

⓪ 0.01	① 0.21	② 0.43	③ 0.58
④ 0.69	⑤ 0.78	⑥ 1.02	⑦ 1.45

(3) 総平均時間と行動者平均時間のように，0 を含むデータから平均値や分散を計算する場合には，データ全体で考える場合と 0 を除いた残りの値からなるデータで考える場合がある．ここで，データに含まれる 0 の個数によって，分散にどのような影響があるかを考察してみよう.

n, k を自然数とする．ただし，$n > k$ とする．k 個の正の値 x_1, x_2, $\cdots$, x_k と $(n-k)$ 個の 0 からなるデータ

$$\underbrace{x_1,\ x_2,\ \cdots,\ x_k}_{k\text{ 個}},\ \underbrace{0,\ 0,\ \cdots,\ 0}_{(n-k)\text{ 個}}$$

について，n 個全体で考えた場合の分散を ${s_T}^2$ とし，0 を除いた k 個のデータで考えた場合の分散を ${s_P}^2$ とする.

${s_T}^2$ と ${s_P}^2$ の関係について，次の (a)，(b) の場合について考える.

(a) 6 個のデータ 1, 2, 3, 0, 0, 0 については，${s_T}^2 = \dfrac{\boxed{\text{サ}}}{\boxed{\text{シ}}}$ であり，${s_T}^2 - {s_P}^2 = \dfrac{\boxed{\text{ス}}}{\boxed{\text{セ}}}$ となる.

(b) 12 個のデータ 1, 2, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0 については，${s_T}^2 - {s_P}^2 = \dfrac{\boxed{\text{ソ}}}{\boxed{\text{タ}}}$ となる.

${s_T}^2 - {s_P}^2$ の値について，(a)，(b) の場合で比べると，$\boxed{\text{チ}}$ の方が大きいことがわかる.

$\boxed{\text{チ}}$ の解答群

⓪ (a)	① (b)

【答】

アイ	ウエ	オカ	キ	ク	ケ	コ	サ	シ	ス	セ	ソ	タ	チ
34	68	34	1	4	0	4	4	3	2	3	1	4	0

【解答】

(1) (i) 最頻値 (モード) とは度数が最も多い階級の階級値 (階級の中央値) である.
図 1 から,「通勤・通学」の総平均時間の最頻値は

$$\frac{32+36}{2}=\mathbf{34}$$ ……(答)

であり，図 2 から，行動者平均時間の最頻値は

$$\frac{64+72}{2}=\mathbf{68}$$ ……(答)

である.

(ii) 各階級に含まれるデータの値がすべてその階級値に等しいと仮定する．令和 3 年の「通勤・通学」の総平均時間の平均値 m は

$$\begin{aligned} m &= \frac{26\times6+30\times10+34\times15+38\times8+42\times4+46\times4}{47} \\ &= \frac{156+300+510+304+168+184}{47} \\ &= \frac{1622}{47} \\ &= 34.5\cdots \end{aligned}$$

であり

$$\mathbf{34}\leqq m<34+1$$ ……(答)

を満たす.

(iii) 図 3 より

(a) 令和 3 年の総平均時間の最大値は 46，令和 3 年の行動者平均時間の最小値は 56 であるから,「令和 3 年の総平均時間の最大値は，令和 3 年の行動者平均時間の最小値より小さい」という記述は**正しい**.

(b) 平成 28 年の総平均時間の四分位範囲 (箱の長さ) は $37-30=7$，平成 28 年の行動者平均時間の四分位範囲は $75-62=13$ であるから,「平成 28 年の総平均時間の四分位範囲は，平成 28 年の行動者平均時間の四分位範囲より小さい」という記述は**正しい**.

(c) $H=$ (最大値) $-$ (第 3 四分位数) $=$ (右ひげの長さ) であり，図 3 より

$$H_1=57-38=19,\ H_2=45-37=8 \qquad \therefore\ \frac{H_2}{H_1}=\frac{8}{19}=0.42\cdots$$

$$H_3=109-75=34,\ H_4=104-77=27 \qquad \therefore\ \frac{H_4}{H_3}=\frac{27}{34}=0.79\cdots$$

であり，$\dfrac{H_2}{H_1}<\dfrac{H_4}{H_3}$ であるから,「$\dfrac{H_2}{H_1}>\dfrac{H_4}{H_3}$ となる」という記述は**誤り**である.

よって，(a)，(b)，(c) の正誤の組合せとして正しいものは 1 である. ……(答)

(2) (i) 図 5 から,「通勤・通学」の行動者平均時間が 60 以下で，かつ「移動」の行動者平均時間が 75 以下である都道府県の数は **4** である. ……(答)

(ii) A，B，C，D について総平均時間が同じで A の行動者平均時間が大きいということは A での 0 の個数が多いということであるから，最も適当なものは 0 である. ……(答)

(iii) 「通勤・通学」と「移動」の行動者平均時間をそれぞれ x, y，標準偏差をそれぞれ s_x, s_y および共分散を s_{xy} とおくと，相関係数 r は

$$r=\frac{s_{xy}}{s_xs_y}=\frac{64.4}{11.8\times7.9}=\frac{6440}{9322}=0.690\cdots\fallingdotseq\mathbf{0.69}\quad \text{コは}4$$ ……(答)

である.

(3) (a) 6 個のデータ 1, 2, 3, 0, 0, 0 の分散 ${s_T}^2$ は

$$\begin{aligned}{s_T}^2 &= \overline{x^2} - \overline{x}^2 \\ &= \frac{1^2+2^2+3^2+0^2+0^2+0^2}{6} - \left(\frac{1+2+3+0+0+0}{6}\right)^2 \\ &= \frac{14}{6} - 1 \\ &= \boldsymbol{\frac{4}{3}}\end{aligned}$$

……(答)

である．0 を除いた 3 個のデータで考えた場合の分散 ${s_P}^2$ は

$$\begin{aligned}{s_P}^2 &= \frac{1^2+2^2+3^2}{3} - \left(\frac{1+2+3}{3}\right)^2 \\ &= \frac{14}{3} - 4 \\ &= \frac{2}{3}\end{aligned}$$

であり

$${s_T}^2 - {s_P}^2 = \frac{4}{3} - \frac{2}{3} = \boldsymbol{\frac{2}{3}}$$

……(答)

となる．

(b) 12 個のデータ 1, 2, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0 については

$$\begin{aligned}{s_T}^2 &= \overline{x^2} - \overline{x}^2 \\ &= \frac{1^2+2^2+3^2}{12} - \left(\frac{1+2+3}{12}\right)^2 \\ &= \frac{14}{12} - \frac{1}{4} \\ &= \frac{11}{12}\end{aligned}$$

$$\begin{aligned}{s_P}^2 &= \frac{1^2+2^2+3^2}{3} - \left(\frac{1+2+3}{3}\right)^2 \\ &= \frac{2}{3} \quad (\because \text{ (i) と同じ計算})\end{aligned}$$

であり

$${s_T}^2 - {s_P}^2 = \frac{11}{12} - \frac{2}{3} = \frac{3}{12} = \boldsymbol{\frac{1}{4}}$$

……(答)

となる．

${s_T}^2 - {s_P}^2$ の値について，(a)，(b) の場合で比べると，

(a)　チは0　……(答)

の方が大きいことがわかる．