

太郎さんは、総務省が公表している2020年の家計調査の結果を用いて、地域による食文化の違いについて考えている。家計調査における調査地点は、都道府県庁所在市および政令指定都市（都道府県庁所在市を除く）であり、合計52市である。家計調査の結果の中でも、スーパーマーケットなどで販売されている調理食品の「二人以上の世帯の1世帯当たり年間支出金額（以下、支出金額、単位は円）」を分析することにした。以下においては、52市の調理食品の支出金額をデータとして用いる。

太郎さんは調理食品として、最初になぎのかば焼き（以下、かば焼き）に着目し、図1のように52市におけるかば焼きの支出金額のヒストグラムを作成した。ただし、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。

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。

（市の数）

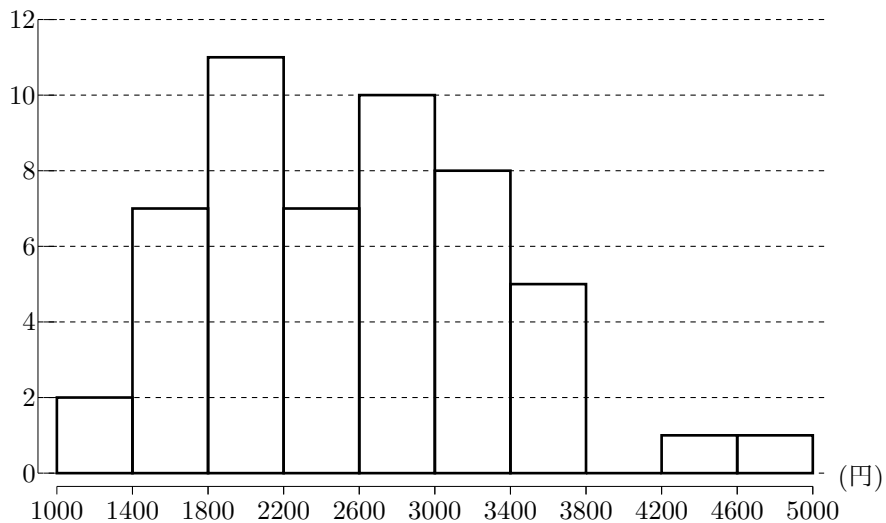


図1 かば焼きの支出金額のヒストグラム

(1) 図1から次のことが読み取れる。

- 第1四分位数が含まれる階級は である。
- 第3四分位数が含まれる階級は である。
- 四分位範囲は .

,

 の解答群（同じものを繰り返し選んでもよい。）

- | | |
|-------------------|-------------------|
| ① 1000 以上 1400 未満 | ⑤ 3000 以上 3400 未満 |
| ② 1800 以上 2200 未満 | ⑥ 3400 以上 3800 未満 |
| ③ 2200 以上 2600 未満 | ⑦ 3800 以上 4200 未満 |
| ④ 2600 以上 3000 未満 | ⑧ 4200 以上 4600 未満 |
| ⑤ 3000 以上 3400 未満 | ⑨ 4600 以上 5000 未満 |
| ⑥ 3400 以上 3800 未満 | |
| ⑦ 3800 以上 4200 未満 | |
| ⑧ 4200 以上 4600 未満 | |
| ⑨ 4600 以上 5000 未満 | |

ウの解答群

- ① 800 より小さい
 ② 800 より大きく 1600 より小さい
 ③ 1600 より大きく 2400 より小さい
 ④ 2400 より大きく 3200 より小さい
 ⑤ 3200 より大きく 4000 より小さい
 ⑥ 4000 より大きい

(2) 太郎さんは、東西での地域による食文化の違いを調べるために、52 市を東側の地域 E (19 市) と西側の地域 W (33 市) の二つに分けて考えることにした。

(i) 地域 E と地域 W について、かば焼きの支出金額の箱ひげ図を、図 2、図 3 のようにそれぞれ作成した。

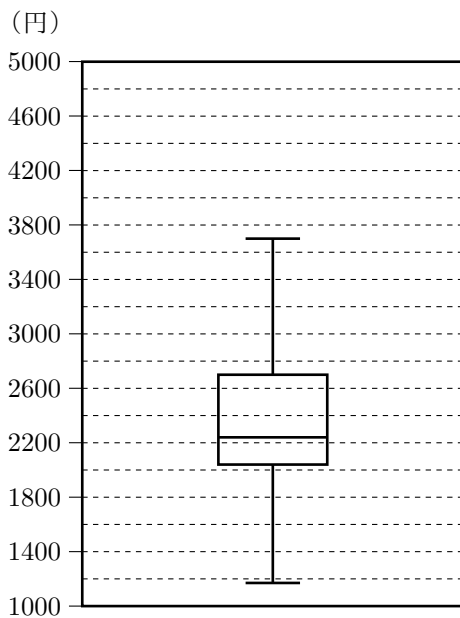


図 2 地域 E におけるかば焼きの支出金額の箱ひげ図

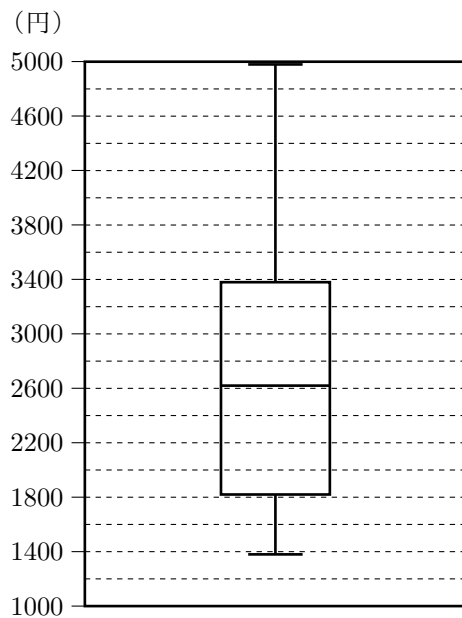


図 3 地域 W におけるかば焼きの支出金額の箱ひげ図

かば焼きの支出金額について、図 2 と図 3 から読み取れることとして、次の①～③のうち、正しいものは **エ** である。

エの解答群

- ① 地域 E において、小さい方から 5 番目は 2000 以下である。
 ② 地域 E と地域 W の範囲は等しい。
 ③ 中央値は、地域 E より地域 W の方が大きい。
 ④ 2600 未満の市の割合は、地域 E より地域 W の方が大きい。

- (ii) 太郎さんは、地域 E と地域 W のデータの散らばりの度合いを数値でとらえようと思い、それぞれの分散を考えることにした。地域 E におけるかば焼きの支出金額の分散は、地域 E のそれぞれの市におけるかば焼きの支出金額の偏差の

である。

の解答群

- ① 2乗を合計した値
 ② 絶対値を合計した値
 ③ 2乗を合計して地域 E の市の数で割った値
 ④ 絶対値を合計して地域 E の市の数で割った値
 ⑤ 2乗を合計して地域 E の市の数で割った値の平方根のうち正のもの
 ⑥ 絶対値を合計して地域 E の市の数で割った値の平方根のうち正のもの

- (3) 太郎さんは (2) で考えた地域 E における、やきとりの支出金額についても調べることにした。

ここでは地域 E において、やきとりの支出金額が増加すれば、かば焼きの支出金額も増加する傾向があるのではないかと考え、まず図 4 のように、地域 E における、やきとりとかば焼きの支出金額の散布図を作成した。そして、相関係数を計算するために、表 1 のように平均値、分散、標準偏差および共分散を算出した。ただし、共分散は地域 E のそれぞれの市における、やきとりの支出金額の偏差とかば焼きの支出金額の偏差との積の平均値である。

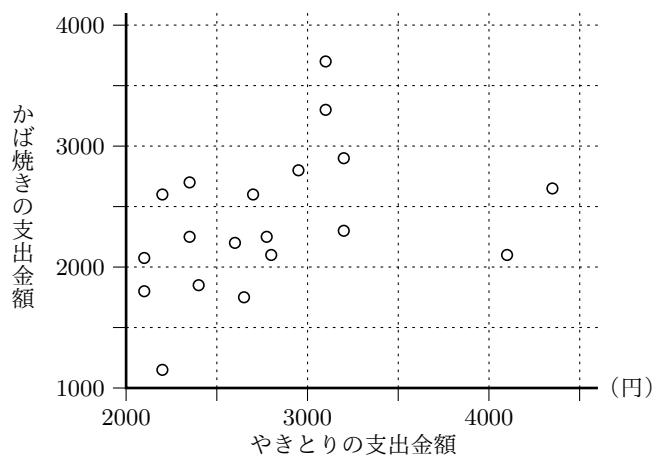


図 4 地域 E における、やきとりとかば焼きの支出金額の散布図

表 1 地域 E における、やきとりとかば焼きの支出金額の平均値、分散、標準偏差および共分散

	平均値	分散	標準偏差	共分散
やきとりの支出金額	2810	348100	590	124000
かば焼きの支出金額	2350	324900	570	

表1を用いると、地域Eにおける、やきとりの支出金額とかば焼きの支出金額の相関係数は **カ** である。

カ については、最も適当なものを、次の①～⑩のうちから一つ選べ。

①	-0.62	④	-0.50	⑦	-0.37	⑩	-0.19
②	-0.02	⑤	0.02	⑧	0.19	⑨	0.37
③	0.50	⑥	0.62				

(23 共通テスト 本試験 IA2[1])

【答】

ア	イ	ウ	エ	オ	カ
2	5	1	2	2	7

【解答】

- (1) ヒストグラムから度数分布表をつくると右表となる。データは52個あるから、すべてのデータの値を小さい方から並べたとき、下位のデータは26個、上位のデータも26個ある。

第1四分位数は、13番目と14番目の値の平均である。13番目と14番目は同じ階級にあるから、第1四分位数が含まれる階級は累積度数13、14が含まれる階級と一致し

1800 以上 2200 未満の階級 ② ……(答)

である。第3四分位数は、39番目と40番目の値の平均である。39番目と40番目は同じ階級にあるから、第3四分位数が含まれる階級は、累積度数39、40が含まれる階級と一致し

3000 以上 3400 未満の階級 ⑤ ……(答)

である。

四分位範囲とは、第3四分位数 Q_3 から第1四分位数 Q_1 を引いた値である。

$$\begin{cases} 1800 \leq Q_1 < 2200 \\ 3000 \leq Q_3 < 3400 \end{cases}$$

であるから

$$3000 - 2200 < Q_3 - Q_1 < 3400 - 1800$$

すなわち、四分位範囲 $Q_3 - Q_1$ は

800 より大きく 1600 より小さい。 ① ……(答)

- (2) (i) 各項目の真偽を確かめる。

① 地域Eには19市があるから、第1四分位数は支出金額を小さい順に並べたときの5番目の値であり、箱ひげ図をみると2000より大きい。正しくない。

② 範囲とはデータの最大値と最小値の差であるから、地域Eの範囲は $3700 - 1200 = 2500$ よりやや大きく、地域Wの範囲は $5000 - 1400 = 3600$ である。正しくない。

③ 中央値(メジアン、第2四分位数)とは箱を上下に分ける線分が表す支出金額である。地域Eの中央値は2200よりやや大きく、地域Wの中央値はほぼ2600である。正しい。

階級	度数	累積度数
1000～1400	2	2
1400～1800	7	9
1800～2200	11	20
2200～2600	7	27
2600～3000	10	37
3000～3400	8	45
3400～3800	5	50
3800～4200	0	50
4200～4600	1	51
4600～5000	1	52

- ③ 2600 は、地域 E では中央値より大きく第 3 四分位数より小さい。地域 W では中央値よりやや小さい。したがって、2600 未満の市の割合は、地域 E では $\frac{1}{2}$ より大きく、 $\frac{3}{4}$ よりやや小さい。地域 W では $\frac{1}{2}$ よりやや小さい。正しくない。

以上より、正しいものは ② である。…………(答)

- (ii) X の平均 $E(X)$ を m とおくと、分散 $V(X)$ の定義は

$$V(X) = E((X - m)^2)$$

である。すなわち、地域 E の支出金額の分散は、それぞれの市の支出金額の偏差 $X - m$ の 2 乗を合計して地域 E の市の数で割った値である。②…………(答)

- (3) やきとりの支出金額を x 、かば焼きの支出金額を y とおくと、 x と y の共分散 s_{xy} は

$$s_{xy} = \frac{1}{n} \{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \cdots + (x_n - \bar{x})(y_n - \bar{y})\}$$

であり、 x の標準偏差 s_x と y の標準偏差 s_y の相関係数 r は

$$r = \frac{s_{xy}}{s_x s_y}$$

である。表 1 より

$$r = \frac{124000}{590 \cdot 570} = 0.368 \cdots$$

であり、解答欄の中で最も適当なものは

0.37 ⑦…………(答)

である。