

# 総 問

## 総 合 問 題

令 和 7 年 度(前期)

### 注 意

1. 「解答はじめ」というまで、この問題冊子を開いてはいけない。
2. 問題は1冊（本文14ページ、白紙2枚）、解答用紙は3枚である。白紙は問題冊子の中にはさみこんであるので引き抜いて下書用紙として使ってよい。
3. 全部の解答用紙に受験番号を書くこと。受験番号は次の要領で明確に記入する。

(例) 受験番号 80001 番の場合 → 

8	0	0	0	1
---	---	---	---	---

4. 解答は解答用紙の所定の位置に書くこと。他の所に書くと無効になることがある。
5. 書き損じても、代わりの用紙は交付しない。
6. 試験終了後、問題冊子と白紙は持ち帰ること。

1 日本は、高齢化と人口減少が進む中で、地域社会に様々な課題が発生している。その課題の一つとして、空家(建築物又はこれに附属する工作物であって居住その他の使用がなされていないことが常態であるもの)の増加があげられる。表1—1は、10地域の2010年、2015年、2020年の空家率(%)を含む6つの地域に関する統計量(X1~X6)であり、実データに基づき作成した架空のデータである。表1—2は、X1からX6の統計量どうしの相関係数を示したものである。

**統計量は下記のように地域ごとに定義される：**

**X1：**空家率(%) = 空家数 ÷ 住宅数

**X2：**人口(人)

**X3：**社会増(人) = 地域に流入してきた人口 - 地域から流出していった人口であり、直近の5年間の一年当たりの平均

**X4：**児童一人当たり教育費(万円/人) = 地域が所属する自治体の予算に計上された教育費総額 ÷ その自治体の義務教育対象児童数

**X5：**高齢人口依存比率(%) = 65歳以上人口 ÷ 生産年齢(20歳以上65歳未満)人口

**X6：**地価(万円/m<sup>2</sup>)

問 1. 2010年、2015年、2020年のそれぞれにおいて、①X1と最も相関が強い統計量を、X2からX6の中から抽出しなさい。そして、②その抽出した統計量と、X1以外において最も相関が強い統計量を選びなさい。

問 2. 2010年、2015年、2020年のそれぞれにおいて、X5と①最も相関が強い統計量を、X1、X2、X3、X4、X6の中から抽出しなさい。また、②X5が上昇すると、地域社会に対してどのような影響が発生するのか。6つの統計量どうしの相関係数を踏まえて解説しなさい。

問 3. 2020年に、X1が低下した地域がある。①その地域を示しなさい。②X1が低下した理由と考えられる要因を、X1以外の5つの統計量から解説しなさい。また、③X1を低下させるために、どのような政策をとることが有効であるか考えるかを理由とともに示せ。なお、5つの統計量以外の要因も考慮すること。

表1—1：空家率(%)を含む地域別・年別の統計量

年	地域	X1：空家率(%)	X2：人口(人)	X3：社会増(人)	X4：児童一人当たり教育費(万円/人)	X5：高齢人口依存比率(%)	X6：地価(万円/m <sup>2</sup> )
2010	1	30	6050	- 30	90	35	5
2010	2	26	10400	- 70	60	30	10
2010	3	20	51750	- 45	80	33	15
2010	4	18	83060	- 35	85	30	18
2010	5	16	103275	- 25	90	25	20
2010	6	12	156465	- 15	100	28	21
2010	7	10	263500	20	85	27	23
2010	8	8	284725	50	100	23	23
2010	9	10	350770	120	130	25	25
2010	10	7	510725	160	160	23	26
2015	1	35	5650	- 40	95	38	4
2015	2	30	9750	- 50	75	33	6
2015	3	25	51100	- 30	90	35	10
2015	4	23	82110	- 40	85	35	15
2015	5	21	102050	- 45	80	28	18
2015	6	19	154790	- 35	90	32	15
2015	7	15	261300	- 40	110	30	18
2015	8	16	282875	- 20	100	28	17
2015	9	12	349920	110	120	27	18
2015	10	7	509250	155	150	23	22
2020	1	38	5150	- 50	90	40	3
2020	2	2	10250	200	150	35	5
2020	3	30	49550	- 60	80	38	8
2020	4	28	80360	- 50	85	37	10
2020	5	26	99850	- 40	90	35	15
2020	6	25	152850	- 38	100	34	10
2020	7	23	258800	- 50	85	33	16
2020	8	20	280750	- 25	100	30	15
2020	9	15	348920	100	125	28	18
2020	10	8	507500	150	130	25	25

表 1—2：地域別空家率と主要な統計量どうしの相関係数

2010 年

	X 1：空家率 (%)	X 2：人口 (人)	X 3：社会増 (人)	X 4：児童 一人当たり 教育費 (万円/人)	X 5：老齡人 口依存比率 (%)	X 6：地価 (万円/m <sup>2</sup> )
X 1	1.000	-0.865	-0.747	-0.651	0.872	-0.982
X 2	-0.865	1.000	0.965	0.883	-0.807	0.844
X 3	-0.747	0.965	1.000	0.941	-0.716	0.735
X 4	-0.651	0.883	0.941	1.000	-0.605	0.651
X 5	0.872	-0.807	-0.716	-0.605	1.000	-0.873
X 6	-0.982	0.844	0.735	0.651	-0.873	1.000

2015 年

	X 1：空家率 (%)	X 2：人口 (人)	X 3：社会増 (人)	X 4：児童 一人当たり 教育費 (万円/人)	X 5：老齡人 口依存比率 (%)	X 6：地価 (万円/m <sup>2</sup> )
X 1	1.000	-0.941	-0.723	-0.765	0.895	-0.943
X 2	-0.941	1.000	0.848	0.914	-0.883	0.826
X 3	-0.723	0.848	1.000	0.898	-0.721	0.562
X 4	-0.765	0.914	0.898	1.000	-0.688	0.608
X 5	0.895	-0.883	-0.721	-0.688	1.000	-0.853
X 6	-0.943	0.826	0.562	0.608	-0.853	1.000

2020 年

	X 1：空家率 (%)	X 2：人口 (人)	X 3：社会増 (人)	X 4：児童 一人当たり 教育費 (万円/人)	X 5：老齡人 口依存比率 (%)	X 6：地価 (万円/m <sup>2</sup> )
X 1	1.000	-0.462	-0.916	-0.905	0.676	-0.419
X 2	-0.462	1.000	0.356	0.317	-0.950	0.938
X 3	-0.916	0.356	1.000	0.978	-0.555	0.269
X 4	-0.905	0.317	0.978	1.000	-0.543	0.203
X 5	0.676	-0.950	-0.555	-0.543	1.000	-0.894
X 6	-0.419	0.938	0.269	0.203	-0.894	1.000

2 次の文章を読んで以下の問いに答えなさい。

Byte pair encoding (BPE : バイト対符号化) と呼ばれるアルゴリズムがある。BPE はデータ圧縮法の一つで、可逆圧縮のアルゴリズムであり、OpenAI の ChatGPT (大規模言語モデルの一つ) のテキスト前処理にも使われている。BPE の疑似コードを表 2-1 に示す。疑似コードにおける各行の“//”以降はコメントである。

表 2-1 : Byte pair encoding

```
入力：文字列集合  $D$ ，語彙サイズ  $k$ 
出力：語彙集合  $V$ 

1 : procedure BPE( $D$ ,  $k$ )
2 :    $D$  に含まれるすべての異なり文字を辞書順で  $V$  に追加
3 :   while  $V$  のサイズ  $< k$  do
4 :      $D$  の最頻出ペア  $t_L, t_R$  を抽出 // 同頻度のペアは  $V$  への登録が若い順
5 :      $t_{NEW} \leftarrow t_L + t_R$  // ペアを連結して新しい語彙を作る
6 :      $t_{NEW}$  を  $V$  に加える
7 :      $D$  に出現するすべての  $t_L, t_R$  を  $t_{NEW}$  で置き換える
8 :   end while
9 :   return  $V$ 
```

以下の表 2-2 は  $D = \{\text{"cooler", "coolest", "eldest", "highest"}\}$ ,  $k = 16$  とした場合の各マージ(連結)ステップを示したもので、BPE による符号化後の語彙集合は  $V = \{\text{"c", "d", "e", "g", "h", "i", "l", "o", "r", "s", "t", "es", "est", "co", "ol", "cool"}\}$  となり、それを用いた文字列集合の分割結果は  $\{\text{"cool-e-r", "cool-est", "e-l-d-est", "h-i-g-h-est"}\}$  となる(語彙となる文字列の切れ目に“-”を入れている)。

表 2—2 : BPE の適用例 (下線は各ステップでマージされる語彙のペアを示している。)

初期状態	c-o-o-l-e-r	c-o-o-l-e-s-t	e-l-d-e-s-t	h-i-g-h-e-s-t
マージ 1	c-o-o-l-e-r	c-o-o-l- <u>e-s-t</u>	e-l-d- <u>e-s-t</u>	h-i-g-h- <u>e-s-t</u>
マージ 2	c-o-o-l-e-r	c-o-o-l- <u>es-t</u>	e-l-d- <u>es-t</u>	h-i-g-h- <u>es-t</u>
マージ 3	<u>c-o-o-l</u> -e-r	<u>c-o-o-l</u> -est	e-l-d-est	h-i-g-h-est
マージ 4	co- <u>o-l</u> -e-r	co- <u>o-l</u> -est	e-l-d-est	h-i-g-h-est
マージ 5	<u>co-ol</u> -e-r	<u>co-ol</u> -est	e-l-d-est	h-i-g-h-est

また、表 2—3 に Common Crawl(ウェブ上のデータを大規模にダウンロードしたデータベース)におけるテキストの言語別の割合を示す。ChatGPT のベースとなった GPT-3 で用いられていた学習データのうち、60 % が(フィルタリング済みの)Common Crawl 由来であると言われている。

表 2—3 : Common Crawl(2024) データにおける上位 10 言語の割合

言語	割合
英語	43.8 %
ロシア語	6.1 %
ドイツ語	5.3 %
中国語	5.1 %
日本語	4.8 %
スペイン語	4.6 %
フランス語	4.2 %
(不明)	3.2 %
イタリア語	2.6 %
ポルトガル語	2.1 %
オランダ語	1.8 %

- 問 1. 表 2—1 のアルゴリズムに従い BPE を  $D = \{\text{“picked”}, \text{“pickled”}, \text{“pickles”}, \text{“pushed”}\}$ ,  $k = 16$  として適用した場合の各ステップを表 2—2 のように示し, 構築される語彙集合  $V$  およびそれを用いた  $D$  の分割を求めよ。
- 問 2. 表 2—3 を参考に, ウェブ上のテキストデータを用いて BPE によってデータ圧縮を行う際の問題点について, 政治・経済的観点及び倫理的観点から考察せよ。

3

以下の文章(日本の行政機関である個人情報保護委員会が2024年3月27日に  
発表した文章)を読み, 問いに答えなさい。

著作権保護の観点から、  
公開していません。

# 著作権保護の観点から、 公開していません。

問 以上の文章の要点を踏まえたうえで、グローバル化時代の日本における情報流通の課題について 250 字以上 350 字以内で書きなさい。