

2020年度 横浜市立大学 データサイエンス学部

特別選抜入学試験

【海外帰国生／国際バカロレア／科学オリンピック／外国人留学生／社会人】

総 合 問 題

【注意事項】

1. 試験時間は90分である。
2. 試験開始の合図まで、この問題冊子を開いてはいけない。ただし、表紙はあらかじめよく読んでおくこと。
3. 問題の印刷は1ページから9ページまである。
4. 解答用紙は2枚である。
5. 試験開始後、すべての解答用紙に受験番号と氏名を所定の欄に記入すること。
6. 問題冊子に落丁、乱丁、印刷不鮮明な箇所等があった場合および解答用紙が不足している場合は、手をあげて監督者に申し出ること。
7. 解答は必ず解答用紙の指定された箇所に記入すること。解答用紙の裏面に記入してはいけない。
8. 問題番号に対応した解答用紙に解答していない場合は、採点されない場合もあるので注意すること。
9. 解答する字数に指定がある場合は、句読点も1字として数えること。英数字を記入する場合は、1字分のマス目に2文字まで記入してよい。
10. 問題冊子の中の白紙部分は下書き等に使用してよい。
11. 解答用紙を切り離したり、持ち帰ってはいけない。
12. 試験終了まで退室を認めない。試験中の気分不快やトイレ等、やむを得ない場合には、手をあげて監督者を呼び、指示に従うこと。
13. 試験終了後は問題冊子を持ち帰ること(面接時に使用するため保管しておくこと)。

[I] 以下の設間に答えなさい。

(1) 表 1a は、2010 年から 2018 年までの訪日外客数（海外から日本への来訪者数）の総数および地域別の推移である。ここで、「その他」はオセアニア、南アメリカ、アフリカ、国籍不詳の合計である。表 1b は、表 1a から計算した、2010 年の値を 100 とした総数および地域別の各年の推移である。

表 1a：訪日外客数の推移（単位：千人）

地域	2010	2011	2012	2013	2014	2015	2016	2017	2018
総数	8,611	6,219	8,358	10,364	13,413	19,737	24,040	28,691	31,192
アジア	6,528	4,723	6,388	8,115	10,819	16,645	20,429	24,716	26,758
ヨーロッパ	853	569	776	904	1,049	1,245	1,422	1,526	1,720
北アメリカ	906	685	876	981	1,112	1,311	1,570	1,757	1,940
その他	324	242	318	364	433	536	619	692	774

出典：日本政府観光局訪日外客数

表 1b：訪日外客数の 2010 年を 100 とした値

地域	2010	2011	2012	2013	2014	2015	2016	2017	2018
総数	100	72	97	120	156	229	279	333	362
アジア	100	72	98	124	166	255	313	379	410
ヨーロッパ	100	67	91	106	123	146	167	179	202
北アメリカ	100	76	97	108	123	145	173	194	214
その他	100	75	98	112	134	165	191	214	239

図 1a は表 1a の総数 x_t のグラフ ($t = 2010, \dots, 2018$), 図 1b は表 1a の総数の対前年比

$r_t = \frac{x_t}{x_{t-1}}$ のグラフ ($t = 2011, \dots, 2018$), 図 1c は表 1b の総数以外の 4 地域の値のグラフである。

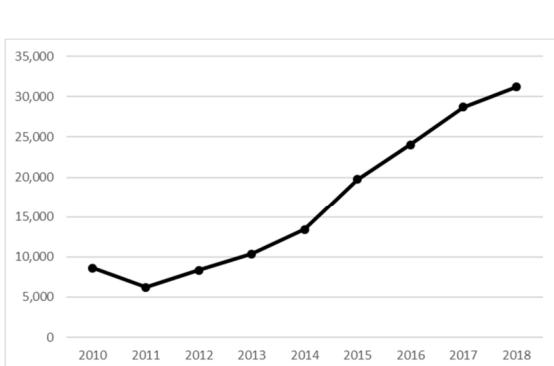


図 1a：訪日外客数の総数（千人）

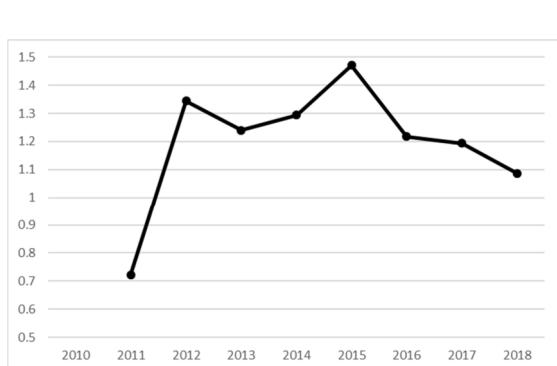


図 1b：訪日外客数の総数の対前年比

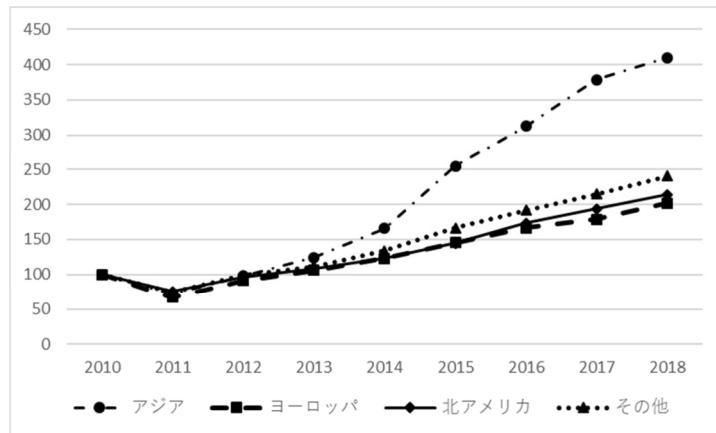


図 1c：訪日外客数の 2010 年を 100 とした値

これらの図表に関する次の文中の ~ に入る最も適当な整数を答えなさい。

訪日外客総数は、2011 年の東日本大震災の年を除いて年ごとに増加していることが分かる。2013 年には東日本大震災の前年の数字を上回り、その 5 年後の 2018 年には 2013 年の約 倍となって、初めて 千万人を超えた。

訪日外客総数の t 年における前年比の相対変化率 a_t (%) を

$$a_t = 100 \times \frac{x_t - x_{t-1}}{x_{t-1}} \quad (t = 2011, \dots, 2018)$$

とすると、変化率 a_t の最大値は $t = \boxed{c}$ における % であり、最小値は $t = \boxed{e}$ における % である。

地域別にみると、アジアからの来訪者の伸びが顕著で、アジアからの来訪者の総数に占める割合は、2010 年には全体の 76% であったものが 2018 年には約 % ポイント上昇している。また、2010 年を 100 とした値では、2018 年のアジアの値は他の 3 つの地域の 2018 年の値の約 倍となっている。

(2) 図2は、2018年7-12月期の訪日外客の各都道府県への訪問率(%)とその地での一人当たりの支出額(万円)(以下では単に、支出額という)、各都道府県への訪問率(%)とその地での平均泊数(泊)、および各都道府県での平均泊数(泊)と支出額(万円)の散布図である。表2は、各調査項目のそれぞれ上位5位の都道府県名と調査結果の値である。この調査は、全国の主要空海港の出発ロビーにて、出国直前の訪日外客に対し対面方式にて行われた。なお、ここでの「訪問率」とは、各都道府県を訪れたと回答した回答者数の全回答者数に対する比率(%)である。調査は訪れた都道府県をすべて答える複数回答のため、訪問率の総和は100ではなく255である。

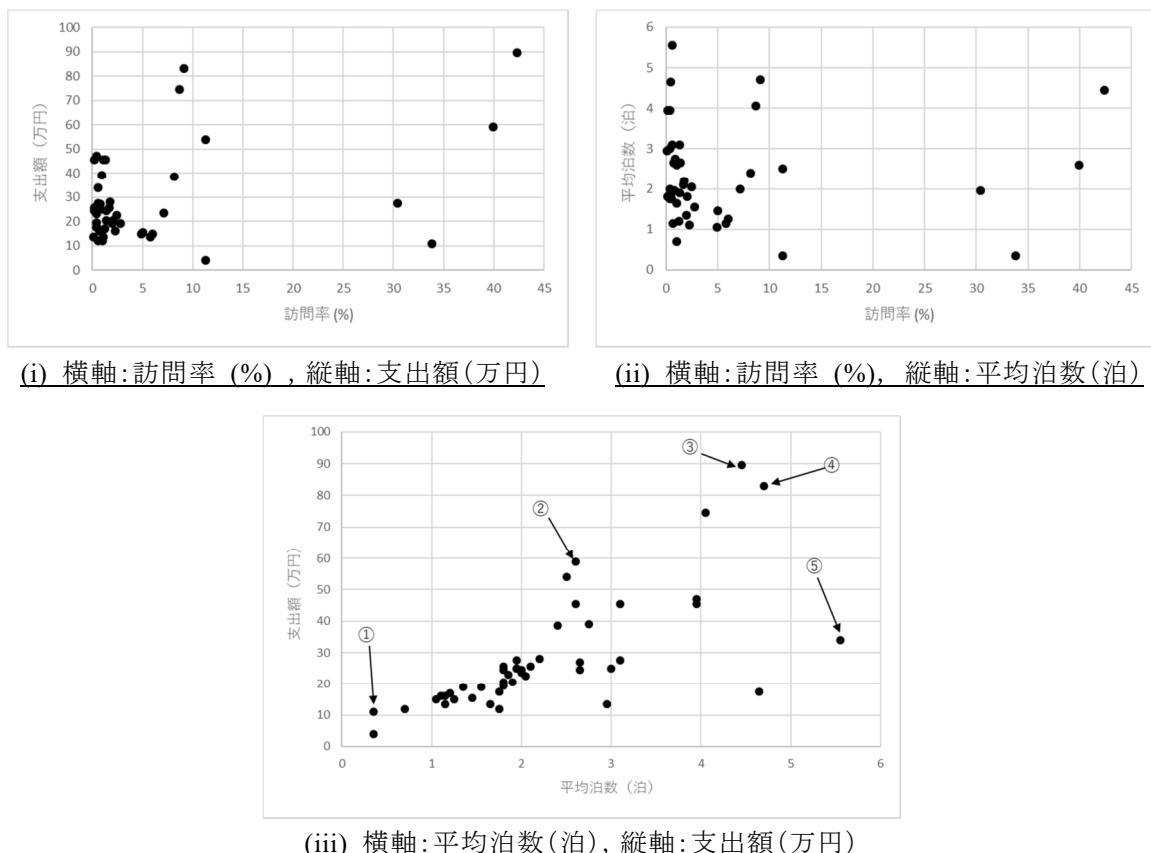


図2：訪日外客の訪問率、平均泊数、支出額の散布図
出典：観光庁【訪日外国人消費動向調査】2018年7-12月期の地域調査結果

表2：訪問率、平均泊数、支出額の各上位5都道府県

(i) 訪問率(%)

(ii) 平均泊数(泊)

(iii) 支出額(万円)

都道府県	訪問率	平均泊数	支出額
東京	42.4	4.5	90
大阪	40.0	2.6	59
千葉	33.8	0.4	11
京都	30.4	2.0	28
奈良	11.3	0.4	4

都道府県	訪問率	平均泊数	支出額
埼玉	0.6	5.6	34
北海道	9.1	4.7	83
茨城	0.5	4.7	18
東京	42.4	4.5	90
沖縄	8.7	4.1	75

都道府県	訪問率	平均泊数	支出額
東京	42.4	4.5	90
北海道	9.1	4.7	83
沖縄	8.7	4.1	75
大阪	40.0	2.6	59
福岡	11.3	2.5	54

- (ア) 調査対象者が訪れた都道府県の数は平均すると である。 に当てはまる数値を答えなさい。
- (イ) 図 2 の散布図の相関係数の値は (i) , (ii) , (iii) である。
 \sim に当てはまる数値として最も適当なものを、下の選択肢から一つずつ選び記号で答えなさい。ただし同じ記号を 2 度以上用いないこと。
- ① -0.33 ② 0.04 ③ 0.42 ④ 0.71 ⑤ 0.98
- (ウ) 平均泊数 1 泊あたりの支出額が最も多い都道府県は である。 に当てはまる記号を、図 2 (iii) の散布図の中の ① ~ ⑤ から一つ選びなさい。逆に、平均泊数 1 泊あたりの支出額が最も少ない都道府県は である。 に当てはまる都道府県名を書きなさい。
- (エ) 図 2 (i) の訪問率と支出額の関係および図 2 (iii) の平均泊数と支出額の関係に見られる相関の大小関係の理由を述べるとともに、訪日外客に都道府県内でお金をたくさん使ってもらうための方策として考えられることを 100 字以内で記述しなさい。

(3) データはその収集法によっては異なる結果を示すことが多い。図 2 および表 2 の平均泊数は帰国直前の訪日外客から得たデータであるが、これを訪日外客の宿泊施設で調査したらどのようになるか簡単な数値例で考える。

訪日外客の宿泊施設で調査した泊数とその比率を表 3 に示す。以下の文の ~ に入る数値を答えなさい。

表 3 : 泊数とその比率

泊数	比率
1	0.4
2	0.3
3	0.2
4	0.1

一般に、自然数 $1, \dots, K$ の値で回答される調査項目 X において、 $1, \dots, K$ の比率がそれぞれ p_1, \dots, p_K であるとき、 X の平均 M は

$$M = 1 \times p_1 + 2 \times p_2 + \dots + K \times p_K \quad (*)$$

で求められる(表 3 では $K=4$ である)。表 3 の比率から式 (*) によって求めた泊数の平均 M は である。

ある日に、宿泊施設からランダムに訪日外客を選んで泊数を調査するとき、2 泊の人が選ばれる確率は、1 泊の人の選ばれる確率の 2 倍である。同様に確率は、3 泊の人は 1 泊の人の 3 倍、4 泊の人は 4 倍となる。よって、 k 泊の人が選ばれる相対比率は、 k 泊の人の比率は p_k であるので、 $r_k = kp_k$ となる。表 3 の数値で計算すると合計は $r_1 + r_2 + r_3 + r_4 =$ となる。よって、 k 泊の人が選ばれる確率 q_k は各相対比率 kp_k をそれらの合計 で割って、

$$q_1 =$$

となる。これより、この場合の泊数の平均は、式 (*) の計算により となる。

このように、泊数を滞在地で直接聞くと、長期滞在者のほうが調査で選ばれる確率が高くなるので、泊数の平均値は大きくなる。

[II] 以下の設問に答えなさい。

(1) 1995年1月17日に明石海峡を震源とするマグニチュード7.3の兵庫県南部地震が発生した。

この地震では、近畿圏の広域、兵庫県を中心に、大阪府、京都府も大きな被害を受けた。特に震源に近い神戸市市街地の被害は甚大で、日本国内のみならず世界中に衝撃を与えた。ここでは、国土交通省気象庁の特設サイト「阪神・淡路大震災からの20年」に掲載されている震度のデータを使用する。図1は宝塚市を含む各地の震度のデータから作成したものである。このような図示は、データの要約を行うためにしばしば用いられる。

(注：当時の震度階級は、震度0から震度7までの8階級であった。)

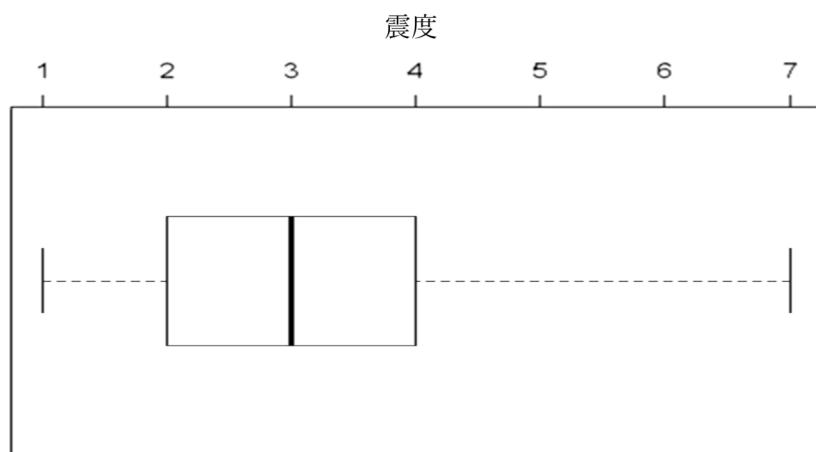


図1：震度のデータの要約

次の文章は図1を説明したものである。

このような図のことを と呼ぶ。箱の左端が を表し、箱の右端が を表す。中央の線は を表している。左右のひげはそれぞれ と を表している。

は四分位範囲から よりも離れているように見える。
(A)

(ア) 文章中の空欄 ~ に当てはまる語句を答えなさい。

(イ) 下線部(A)の四分位範囲はどのように定義されるか。 ~ の語句のうち必要なものを用いて説明しなさい。

(2) 震源地からの距離と震度の関係を調べることにした。表1は、震源地から100km単位で ± 5km 以内の距離に収まる地点の震度を抽出したものである。

表1：震源地からの距離と震度

地点	震源地からの距離(100km)	震度
A	9	1
B	9	1
C	8	2
D	7	2
E	7	3
F	6	3
G	3	3
H	1	3
I	0	5
J	0	7

(データを一部改変)

- (ア) 縦軸に震度、横軸に震源地からの距離をとって散布図を描きなさい。
- (イ) 震源地からの距離を x とする。 x の分散を求めなさい。
- (ウ) 震度を y とする。 y の分散を求めなさい。
- (エ) x と y の相関係数 r_{xy} を求めることにした。以下の問いに答えなさい。
- (a) x と y の共分散 (x と y の偏差の積の平均値) を求めなさい。(小数第1位まで)
- (b) 相関係数 r_{xy} を求めなさい。(小数第2位まで)
- (c) 相関係数 r_{xy} と散布図をもとに、震源地からの距離と震度の関係について50字程度で述べなさい。(ただし震度0については考慮に入れない)

(3) 前問(2)では、相関係数や散布図から2つの変量間の関係を読み取った。相関係数の性質などについて、以下の問いに答えなさい。

(ア) 表1で、地点Hの震度は3ではなく、実は5が正しい値であった。このとき、計算し直した相関係数 r'_{xy} と前問(2)で求めた r_{xy} との関係として適当なものを、次の①～③のうちから一つ選びなさい。

① $r'_{xy} > r_{xy}$

② $r'_{xy} < r_{xy}$

③ $r'_{xy} = r_{xy}$

(イ) 今まで相関係数を用いて2つの変数間の関係性をとらえてきた。以下の①～③のそれぞれについて、正しいのものには○、誤りであるものには×を記入しなさい。

① 相関係数からは2変数の直線的な関連性をとらえることができるが、曲線的な関係は必ずしもとらえることはできない

② 相関係数は外れ値（とても大きな値や、とても小さな値）の影響を受けにくいので、外れ値について考慮する必要はない

③ 変数がとることのできる値の範囲に制約があると、相関係数が影響を受けることがある

(ウ) 2つの変数xとyの相関係数の性質として、以下の①～③それぞれについて、正しいものには○、誤りであるものには×を記入しなさい。

① xとyの相関係数が0.8であれば、-0.8とは方向が逆であるが同じ強さの関係がある

② 変数xとyを入れ替えて相関係数を計算しても相関係数の値は変わらない

③ xの測定単位を変更する、たとえばkmからmに変更してyとの相関係数を求めた場合、変更後の値は変更前の値から変化しない