

令和6年度

宇都宮大学総合型選抜

データサイエンス経営学部データサイエンス経営学科

論 述 試 験

令和5年11月25日(土)
9時00分～10時30分

注 意 事 項

1. 試験開始の合図があるまで、この問題冊子の中を見てはいけない。
2. 「受験番号」は、各解答用紙の受験番号欄に忘れずに記入すること。
3. この冊子には二つの設問がある。乱丁、落丁、印刷不鮮明の箇所があった場合には、申し出ること。
4. 解答用紙は、2枚ある。下書き用紙は、1枚ある。解答は、必ず解答用紙の所定の解答欄に記入すること。なお、下書き用紙は回収しない。
所定の欄以外に記入したものは、無効である。

令和6年度宇都宮大学総合型選抜
データサイエンス経営学部データサイエンス経営学科
論述試験問題

第1問 以下の設問(1)～(5)に答えなさい。

(1) 図1は、厚生労働省「令和4年 国民生活基礎調査」の所得金額階級別世帯数のヒストグラム(相対度数分布)である。

この図から、相対度数が一番大きい階級は、(a) 万円以上、(b) 万円未満であることが読み取れる。また、中央値より平均値の方が大きい。
(c)

(ア) (a), (b) に入る数値を答えなさい。

(イ) 下線(c)となる理由を、ヒストグラムの形状と平均値の性質から説明しなさい。

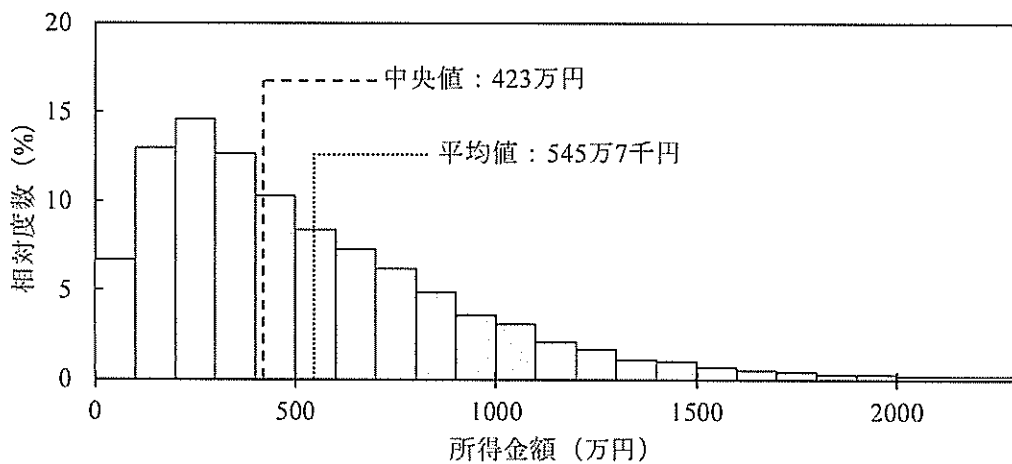


図1：所得金額階級別世帯数のヒストグラム(相対度数分布)

(注：各階級は、 n 万円以上、 $n+100$ 万円未満($n=0,100,200,\dots$)に設定されており、2000万円以上は、1つの階級としている。)

(出典：厚生労働省「令和4年 国民生活基礎調査」)

(2) 図2は、茨城県、群馬県、栃木県の各市町村における納税義務者1人当たりの課税対象所得額の平均値を、県ごとに箱ひげ図にしたものである。

最小値が最も大きいのは、 県であり、第1四分位数が最も小さいのは、 県である。

(ア) (a), (b) に入る県名を答えなさい。

(イ) 茨城県、群馬県、栃木県の各県の箱ひげ図に対応するヒストグラムを、図3の (A) ~ (C) の中から1つずつ選びなさい。

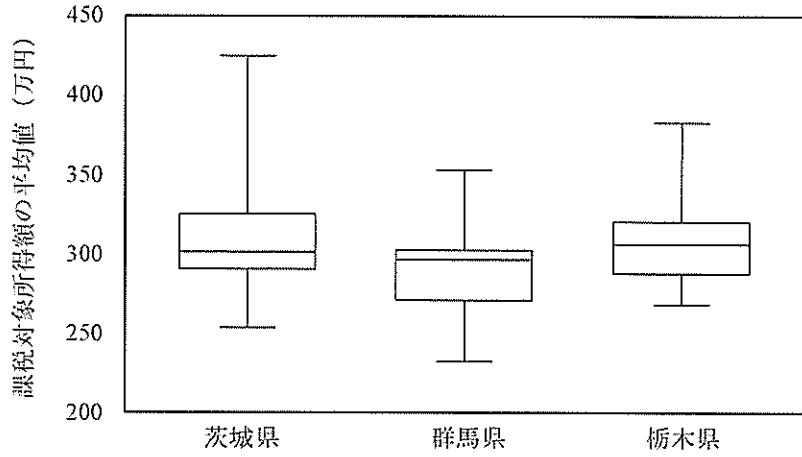


図2：課税対象所得額の平均値の箱ひげ図

(出典：総務省「令和4年度 市町村税課税状況等の調」)

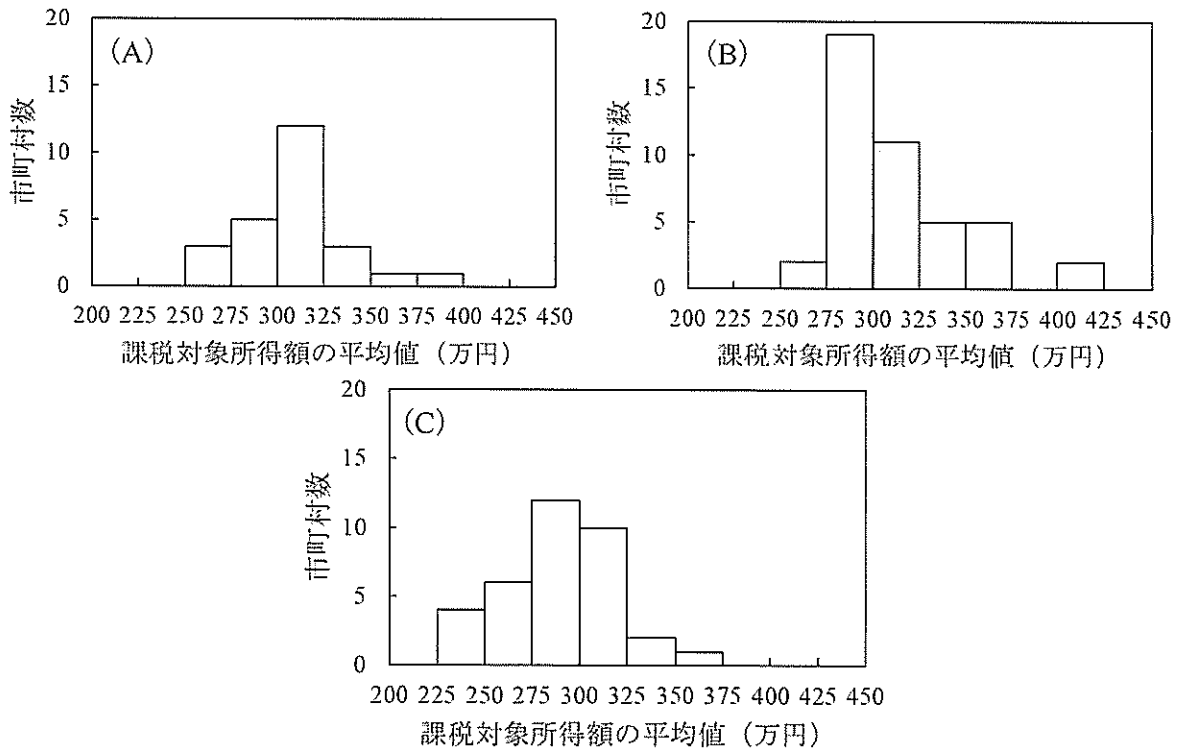


図3：「課税対象所得額の平均値」階級別市町村数のヒストグラム

(3) 次のような2つの変数 x, y の値の n 個の組を1セットのデータとする。

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

今、(A)、(B)、(C) の3セット分のデータがあり、それらのそれぞれの散布図を描いたものが図4である。

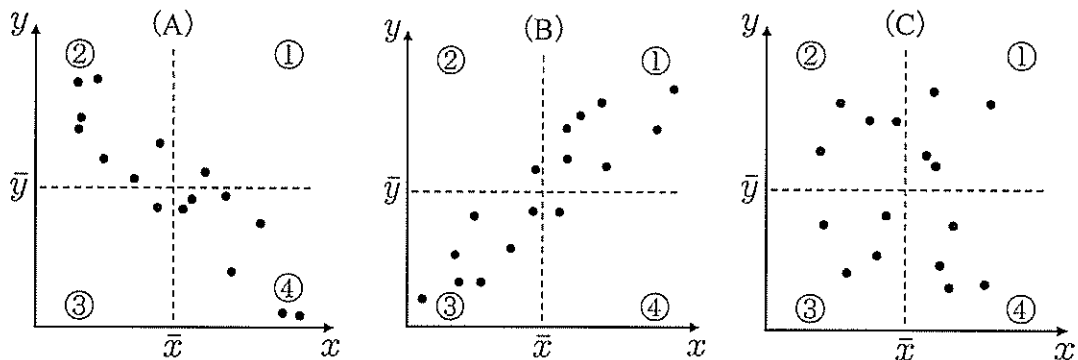


図4: 3つの散布図 ($n = 16$ の場合)

ここで、(A)、(B)、(C) のデータの共分散の値を比較したい。共分散は次の式の S_{xy} で定義される。

$$S_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

ただし、 \bar{x} は x_1, x_2, \dots, x_n の平均値、 \bar{y} は y_1, y_2, \dots, y_n の平均値である。

図4において、各データの \bar{x} は等しく、同様に \bar{y} も等しいとする。また、各散布図で描かれている x 軸の範囲は全て等しく、同様に y 軸の範囲も全て等しいものとする。散布図に描かれた十字の破線はそれぞれ平均値 \bar{x} と \bar{y} の位置を示しており、この破線によって平面を ①～④ の区域に分ける。

(ア) データ (A)、(B)、(C) の共分散をそれぞれ S_A, S_B, S_C とする。これらの大小関係として最も適当なものを次の6つの選択肢の中から1つ答えなさい。

- 1) $S_A < S_B < S_C$
- 2) $S_A < S_C < S_B$
- 3) $S_B < S_A < S_C$
- 4) $S_B < S_C < S_A$
- 5) $S_C < S_A < S_B$
- 6) $S_C < S_B < S_A$

(イ) 問(ア)でそのような選択肢を選んだ理由を、共分散 S_{xy} の定義式と散布図の区域 ①～④ のそれぞれにある点の数から説明しなさい。

- (4) 10人分(番号1~10)の身長(cm)と体重(kg)のデータについて考える。表1はそのデータの一部を示したものである。ここで、このデータの身長と体重の単位をそれぞれ、m(メートル)とg(グラム)に変換したデータを新たに考える。

表1: 変換前のデータ

番号	身長 (cm)	体重 (kg)
1	180.1	72.3
2	156.2	46.5
⋮	⋮	⋮
9	168.7	73.6
10	148.4	42.4

変換後のデータにおいて、番号9の人の身長は (m)、体重は (g) となる。

変換前のデータの身長の標準偏差が P (cm) であったとすると、変換後のデータでは、身長の標準偏差は、 $\times P$ (m) となる。

変換前のデータの体重の標準偏差 Q (kg) であったとすると、変換後のデータでは、体重の標準偏差は、 $\times Q$ (g) となる。

ただし、変量 x の n 個の値 x_1, x_2, \dots, x_n の標準偏差 s は、平均値を \bar{x} とすると、次の式で定義される。

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

変換前の身長 (cm) と体重 (kg) の共分散が S (cm·kg) であったとすると、変換後のデータでは、身長 (m) と体重 (g) の共分散は、 $\times S$ (m·g) となる。

変換前の身長 (cm) と体重 (kg) の相関係数が R であったとすると、変換後のデータでは、身長 (m) と体重 (g) の相関係数は、 $\times R$ となる。

ただし、2つの変量 x, y の相関係数 r は、次の式で定義される。

$$r = \frac{x \text{ と } y \text{ の共分散}}{(x \text{ の標準偏差}) \times (y \text{ の標準偏差})}$$

- (ア) (a)~(f) に入る数値を答えなさい。
 (イ) 2つの変量の相関関係を表すのに適しているのは、共分散と相関係数のどちらか、その理由とともに答えなさい。

- (5) 3組の2つの変量 x, y の値 $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ の相関係数について考える。 \bar{x} を x_1, x_2, x_3 の平均値, \bar{y} を y_1, y_2, y_3 の平均値とすると, 変量 x, y の相関係数 r_3 は次の式で表される。

$$r_3 = \frac{\frac{1}{3} \sum_{i=1}^3 (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{3} \sum_{i=1}^3 (x_i - \bar{x})^2} \sqrt{\frac{1}{3} \sum_{i=1}^3 (y_i - \bar{y})^2}}$$

ここで新たに, $x_i - \bar{x}$ と $y_i - \bar{y}$ ($i = 1, 2, 3$) をそれぞれ空間ベクトルの成分として, 空間ベクトル \vec{X} と \vec{Y} を次のように定義する。

$$\vec{X} = (x_1 - \bar{x}, x_2 - \bar{x}, x_3 - \bar{x})$$

$$\vec{Y} = (y_1 - \bar{y}, y_2 - \bar{y}, y_3 - \bar{y})$$

- (ア) 空間ベクトル \vec{X} と \vec{Y} の内積をそれぞれの成分を用いて表しなさい。
(イ) 空間ベクトル \vec{X} の長さをその成分を用いて表しなさい。
(ウ) 空間ベクトル \vec{Y} の長さをその成分を用いて表しなさい。
(エ) 空間ベクトル \vec{X} と \vec{Y} のなす角を θ とするとき, $\cos \theta$ と 3組の2つの変量 x, y の値 $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ の相関係数 r_3 が等しくなることを, 問 (ア) ~ (ウ) の解答を用いて示しなさい。

第 2 問

今日、データはその種類を問わず、私たちの生活に欠かせないものとなっている。例えば、天気予報は気象データに基づいて予測し、渋滞予測は過去のデータによる統計的予測が中心である。なかでも、市場の大きな主体である企業では、これまで以上にデータの利活用いわゆるデータサイエンスに基づいた経営の必要性が叫ばれている。

経営は企業経営に限らず、目的達成のための枠組み、考え方として使われることがある。例えば、NPO 法人（非営利法人）や部活動、文化祭の運営なども経営である。NPO 法人は、市民の社会貢献活動により社会の健全な発展を促進させることを目的とする。その際、経営資源といわれるヒト・モノ・カネ・情報を効果的に組み合わせ、運営を行う。また、部活動、文化祭の運営も、目的に対して計画を立て、資源を組み合わせ実行していく。これらは、目的達成のための枠組みとして経営そのものといえる。

そこで、以下の問いに答えよ。

経営においてデータの利活用を行う場合、どのようなメリットがあるか。また、経営のなかには、データ、数字として表すことが困難な要素がある。その要素について、どのような事例があると考えられるか。以上の二点を具体的に句読点、空欄を含め、600 文字以上 800 文字以内で記述しなさい。