

1

出題意図

統計学は記述統計と推測統計の2つに分けることができます。大学ではこの両方を習うこととなっている。この問題は統計学の理論ではなく、応用を問うものである。分析者として記述統計と推測統計の応用ができるかどうかを見る。

解答例

(1) 模擬試験を受けたところ、数学・国語・物理の得点はいずれも70点だった。各科目について、受験者全員の結果からそれぞれの平均が50, 50, 40, 標準偏差が20, 40, 20であった。基準化(標準化)変量を計算し、科目の成績の良い順を述べよ。

$$Z_{\text{数学}} = \frac{70 - 50}{20} = 1, Z_{\text{国語}} = \frac{70 - 50}{40} = 0.5, Z_{\text{物理}} = \frac{70 - 40}{20} = 1.5$$

成績のよい順: 物理, 数学, 国語

(2) あるメーカーの同じ機種デジタルビデオカメラ10台を無作為に抽出して、1回の充電で可能な最大の録画時間を調べた結果、平均=9.94時間、標本不偏分散=1.1493(時間)²であった。母分散が未知とする。母平均 μ (時間)の信頼度95%の信頼区間を求め、1回の充電で可能な平均録画時間が10時間という表示は正しくないと言えるかどうかを述べよ。なお、自由度9のt分布の $t_{0.025}=2.262$ である。

$$\left(9.94 - 2.262 * \sqrt{\frac{1.1493}{10}}, 9.94 + 2.262 * \sqrt{\frac{1.1493}{10}} \right)$$

$$(9.94 - 2.262 * 0.339, 9.94 + 2.262 * 0.339) \Rightarrow (9.94 - 0.767, 9.94 + 0.767) \Rightarrow (9.173, 10.707)$$

1回の充電で可能な平均録画時間が10時間という表示は正しくないと言えない。

(3) X大学から5人の学生を、またY大学から7人の学生を無作為に抽出して、各学生の1日の学習時間を調査した。X大学の学習時間の平均は4.24時間、不偏分散は2.278(時間)²であり、Y大学の学習時間の平均は2.171時間、不偏分散は2.762(時間)²である。母分散が等しいが未知として、共通の分散を計算したうえ、帰無仮説 $H_0: \mu_X = \mu_Y$ を対立仮説 $H_1: \mu_X \neq \mu_Y$ に対して有意水準5%で検定せよ。なお、自由度10のt分布の $t_{0.025}=2.228$ である。

$$\text{共通の分散} = \frac{4 * 2.278 + 6 * 2.762}{5 + 7 - 2} = \frac{9.112 + 16.572}{10} = 2.568$$

$$\text{検定統計量} = \frac{4.24 - 2.171}{\sqrt{2.568}} = \frac{2.069}{1.602} = 1.292$$

帰無仮説が採択され、両大学の学習時間に差があると言えない。

(4) ある大学で、全学生 5800 人に対して部活やサークルなどの課外活動をしているかどうかの調査をした。その結果は下表の通りである。

学年 課外活動	1 年生	2 年生	3 年生	4 年生	計
あり	1000	1150	880	1030	4060
なし	400	450	420	470	1740
計	1400	1600	1300	1500	5800

帰無仮説 H_0 : 課外活動と学年が独立である, 対立仮説 H_1 : 課外活動と学年は独立ではないとして有意水準 5% で検定せよ。なお、自由度 3 のカイ 2 乗分布の $\chi_{0.05}^2 = 7.8152$ である。

期待度数

学年 課外活動	1 年生	2 年生	3 年生	4 年生	計
あり	$\frac{1400 * 4060}{5800}$ = 980	$\frac{1600 * 4060}{5800}$ = 1120	$\frac{1300 * 4060}{5800}$ = 910	$\frac{1500 * 4060}{5800}$ = 1050	4060
なし	$\frac{1400 * 1740}{5800}$ = 420	$\frac{1600 * 1740}{5800}$ = 480	$\frac{1300 * 1740}{5800}$ = 390	$\frac{1500 * 1740}{5800}$ = 450	1740
計	1400	1600	1300	1500	5800

検定統計量=8.606

学年 課外活動	1 年生	2 年生	3 年生	4 年生
あり	$\frac{(1000 - 980)^2}{980}$ = 0.408	$\frac{(1150 - 1120)^2}{1120}$ = 0.804	$\frac{(880 - 910)^2}{910}$ = 0.989	$\frac{(1030 - 1050)^2}{1050}$ = 0.381

なし	$\frac{(400 - 420)^2}{420}$ = 0.952	$\frac{(450 - 480)^2}{480}$ = 1.875	$\frac{(420 - 390)^2}{390}$ = 2.308	$\frac{(470 - 450)^2}{450}$ = 0.889
----	--	--	--	--

帰無仮説が棄却され、課外活動と学年は独立ではないと言える。

2

○出題意図

- (1) 社会調査を設計するにあたり必要とされる基本的な概念の理解を問う。その際、具体例を例示させることで、その理解が妥当であるかについて確認する。
- (2) 社会調査の目的が理解できているかについて問い、社会調査を行うことの社会的意義が論理的に説明できるかについて確認する。

○解答例および配点

- (1) ①変数の一方の値が大きくなると、それに従ってもう一方の変数の値が大きくなる関係。
 - (例) 住民リーダーの熱意と市民協働の活発度
 - ②原因とそれによって生ずる結果との関係。
 - (例) 原因：都市化度、結果：犯罪数
 - ③2変数の関係の中には、XとYの直接の関係は薄いのに、まったく別の変数ZがXとYの両方に作用して、XとYの関係を強く見せている場合。
 - (例) 警察官の定員と刑法犯検挙件数
 - ④抽象的な概念を、観察しやすい対象や測定しやすい行為・状態に置き換えること。
 - (例) 市民協働：住民と行政が共同で事業を実行すること
- (2) 社会問題を解決するためには、まず、問題を定義し、その問題が発生した原因を探る。その原因を除去すれば、問題の解決へとつながることが想定される。また、原因が見出しにくい事例であっても、さまざま行っている対策のなかでうまくいっている対策を特定し、対策が効果を発揮するメカニズムについて、因果関係を特定することで明らかにし、問題の解決へとつながることが想定される。さらに、同種の問題とみなせる複数の事案に共通して当てはまる原因を特定し、どの事例にも適用できるような因果関係を明らかにすることで、幅広い問題の解決へとつなげることが可能となる。

・本大問の出題意図：概要

計量社会学において、質的カテゴリカルデータの分析手法に関する基礎知識は、博士前期課程における研究活動（実証的先行研究の正確な理解とそれらの方法論的問題の診断、当該手法を自ら正確に適用など）を自ら進める上で、必要不可欠である。中でも、多項ロジスティック回帰分析は、当該分野において、多用される手法の1つである。ゆえに、当該モデルを応用した先行研究の知見（分析結果の表）を用いて、当該モデルの正確な理解および適用能力を把握することを、本大問の出題意図として設定した。

・各設問の出題意図

以上の背景を踏まえて、本大問では、設問（1）～（4）ごとに、具体的な出題意図を設定した。

設問（1）：当該モデルに関する分析結果の正確な解釈およびメカニズムの的確な説明能力を確認する。

設問（2）（3）：独立変数のうち、ダミー変数の正確な解釈・作成方法を確認する。

設問（4）：交互作用効果を正確に検討するための方法論的知識および、それに付随する多重共線性の対処法の理解を確認する。

問題	解答例
Ⅲ (1)	<p>男性の分析結果では、<u>世帯収入が 800 万円以上の男性と比べて、それ以下の世帯収入の男性は、参加の理由に関わらず、親同士の友人グループに参加しにくい確率が有意に高い。</u></p> <p>女性の分析結果では、<u>世帯収入が 800 万円以上の女性と比べて、世帯収入が 399 万円以下の女性は、子供をきっかけに参加する場合よりも親同士の友人グループに参加しにくい傾向</u>がある（10%水準で有意）。</p> <p>このように、<u>男女で異なる結果が生じるメカニズムとして、以下の可能性が考えられる。</u>まず、<u>高世帯収入の男性は、平均的に高学歴であり、経験的に学歴の経済的価値を認識している。</u>そのため、<u>彼らは、子供の教育達成にも積極的な関心があり、参加の契機にかかわらず、情報交換として機能しうる親同士の友人グループに参加するインセンティブが高い。</u>他方、一般的に、女性の方が親同士のグループへの参加に関わる機会が多い。しかし、高世帯収入の女性と比べて、<u>低世帯所得の女性は、経済的・時間的な制約ゆえに、親同士の友人グループへの参加が阻害される傾向がある。</u></p>
(2)	<p>男女別の効果を比較できない変数は、<u>独立変数「就業形態」に関する①フルタイムダミー、②自営・自由ダミー、③パート・バイトダミー、④主夫/主婦・無職ダミーの4つ</u>である。これらの変数において、<u>男女別の効果を比較できない理由は、「就業形態」の比</u></p>

	<p><u>較基準となるカテゴリーが、男女間で異なるから</u>である。</p> <p><u>就業形態について、男女別の効果を比較するには、男女双方の分析において、同一の基準カテゴリー（例：パート・バイトダミー）を使えばよい。</u></p>
(3)	<p>その理由の1つとして、<u>当該ダミー変数群に関する安定的な推定値を得ることが考えられる。</u> <u>そのためには、基準カテゴリーとして、最も回答者の多いカテゴリーを選べばよい。</u> それゆえ、<u>著者らは、注釈1（※1）において、「最大数の就業形態を基準カテゴリーとした」</u>のである。</p>
(4)	<p><u>「平均値を基準とする中心化」とは、ある変数 x について、各観測値と平均値との間の距離を定量化（標準化）したものである。</u> <u>そのためには、当該変数について、個々の観測値と平均値との間の差を計算すればよい。</u></p> <p><u>著者らが当該操作を行った積極的な根拠の1つとして、多重共線性の回避</u>が考えられる。著者らは、「本人の地元ダミー」と「配偶者の地元ダミー」という主効果のみならず、<u>両変数の積（交互作用効果）の成否も検証している。</u> <u>当該交互作用効果は、各主効果から構成されているがゆえに、両者の間には、非常に高い相関関係が存在する。</u> これらの変数を加工せずにモデルに投入した場合、<u>多重共線性が生じる。</u> 他方、<u>著者らの操作（中心化）を利用すれば、両変数間の独立性を担保できる。</u></p>