

2015 年 RSS/JSS 試験 (Higher Certificate)

HIGHER CERTIFICATE IN STATISTICS, 2015

モジュール 8 : 調査のための抽出と推定

制限時間: 90 分

4 問中 3 問を選択の上 解答のこと。

各問は合計 20 点である。小問の配点は括弧の中に記されている。

グラフ用紙と統計数値表は配布する。

解答にあたっては電卓を使用してよい。
ただし、一般財団法人統計質保証推進協会による「受験要領」に
記された範囲で使用する。

数学記号 \log は e を底とする自然対数を表す。
その他の底をもつ対数は、例えば \log_{10} のように底を明示する。

また、 $\binom{n}{r}$ は ${}_n C_r$ と同じ意味とする。

問題用紙は 8 頁からなり、それぞれの頁は片面にのみ印刷されている。
この表紙が 1 頁目である。
第 1 問は 2 頁目から始まる。

問題は全部で 4 問である。

1. 野生動物管理者たちは、アラスカ中央南部ネルチーナで群れをなすトナカイの総数を推定したい。トナカイの密度は生息場所の種類によって大きく異なっている。予備的な空中からの調査で、群れが使う地域を特定し、それを生息場所の種類にもとづいて6つの層に分けた。

主要な調査においては、地域を 4 km^2 の方形区と呼ばれる小区域に分けることになった。その調査では各層から単純無作為抽出で方形区を選び出し、その地域の方形区ごとに航空機から動物を捜し、写真を撮る。各方形区のトナカイの数、 y 、は写真の中で数えられる。

抽出された 211 の方形区による各層の測定値、 y 、の標本平均と標準偏差は以下の通りである。

層 (h)	地図内 方形区 N_h	標本 方形区 n_h	標本 平均 \bar{y}_h	標本 標準 偏差 s_h
1	400	98	24.1	74.7
2	40	10	25.6	63.7
3	100	37	267.6	589.5
4	40	6	179.0	151.0
5	70	39	293.7	351.5
6	120	21	33.2	99.0
計	770	211		

- (i) この調査の抽出の枠は土地の地図である。どのような問題がこの種の標本と関連していそうか簡潔に述べよ。
(4)

- (ii) 層化（無作為）抽出にもとづく母集団総数の推定値の分散の式は

$$V = \sum_{h=1}^L N_h (N_h - n_h) \frac{S_h^2}{n_h}$$

となる。この式の N_h , S_h , n_h はそれぞれ何を表すか定義せよ。

(2)

- (iii) 上記のデータを用いて、この群れのトナカイの総数を推定し、その総数の 95% 近似信頼区間を求めよ。

(7)

- (iv) この調査において単純無作為抽出よりも層化抽出を用いる利点を簡潔に述べよ。

(4)

- (v) この調査のために層化抽出が使われるとき、比例配分よりも最適配分を用いる利点を簡潔に述べよ。

(3)

2. 政府機関は成人における肥満の問題にますます関心を寄せている。この問題の程度を調査するため、この機関は、成人 5000 人の全国調査に資金を拠出することになっている。そこでは生活スタイルや摂取に関するものだけでなく、社会変数や人口動態変数も調査される。

参加世帯は訪問面談され、特定の期間中、世帯の個々人が摂取した食品と飲料の種類と摂取量、時間を日記に記録し続けることを依頼される。この期間の後、記録された記入項目の完全性の確認と、期間内において食事行動に影響を及ぼすかもしれない日常的でない出来事や病気についての情報を収集するため、さらに短い面談が行われる。

- (i) この調査での情報収集を、質問紙に頼るのではなく、日記を用いることの利点と欠点を概説せよ。 (6)

- (ii) 生活スタイルと摂取についての質問は微妙なものになりうる。これらの質問に回答しなかったり、食事記録を完成しなかったりする参加者が、今回の調査結果に偏りを引き起こすかもしれないのはなぜか？この偏りを調査するために、参加者の背景についての質問をどのように使うことができるか？

ここで述べたような調査において生じうる誤りの他の原因について述べよ。 (7)

- (iii) 質問項目のひとつとして身体活動への参加がある。日頃の運動について以下の質問が行われる：

あなたはまったく運動しないか、稀に運動するか、またはしばしば運動しますか？

この質問の欠点を述べよ。身体活動への参加の頻度や期間を知るための質問を提案せよ。 (7)

3. 人々の住宅状況に関する情報を集めるため、23,000 世帯からなる地域社会で世帯調査が行われた。ある研究者はこのデータを用いて、この地域社会における過密住宅の程度を研究しようとしている。単純無作為抽出された 1500 世帯から世帯規模と世帯あたりの寝室に関する以下の情報が得られた。

	世帯数
標本の全世帯数	1500
<i>1 世帯あたりの人数</i>	
1	434
2	525
3	247
4	200
5	65
6	20
7	5
8	4
<i>1 世帯あたりの寝室数</i>	
1	158
2	406
3	645
4	228
5 +	63

- (i) 以下について、点推定値と 95%近似信頼区間を求めよ。
- (a) この地域社会における世帯あたりの平均人数（過密度の指標），
- (b) この地域社会における単身世帯の割合（余剰度の指標）。
- (12)
- (ii) この研究者は(i)で与えられたものとは別に、寝室あたりの人数にもとづいた過密度の指標について聞いたことがあった。この量を上記の情報から推定する方法を説明せよ。またその推定量の性質を述べよ。
- (3)
- (iii) 以下の点において、この調査で起こり得る実際的な問題について述べよ。
- (a) “世帯”の定義，
- (b) “過密度”の定義，
- (c) 単身世帯からの情報を含めること。
- (5)

4. (a) ある調査では、人々が犯罪の被害者になったことがあるかどうかを調べ、関係した犯罪の詳細を得ることを計画している。電話による聞き取り調査と自記式の質問紙による調査とでは情報を収集するためにどちらが良いか論じよ。

(7)

- (b) あるオレンジ栽培者はトラック 1 台分のオレンジを売ることになっている。オレンジはそれぞれ 120 個ずつ 140 箱の木箱に詰められる。購入を考えている仕入れ担当者は、このオレンジの購入を決める前に、オレンジの果汁量を推定するための標本抽出検査を提案している。

便宜的抽出とは標本抽出を進めながら最も利用しやすい標本を選んで抽出することである。この標本検査において便宜的抽出よりも集落抽出が望ましい理由を述べよ。この例の文脈において、1 段階集落抽出と 2 段階集落抽出はどのように異なるか？

(7)

- (c) ある大学でフルタイムの学部生の資金に関する調査が計画されている。その目的のひとつは、学生が学期中に有給の仕事で稼ぐ平均週給を推定することである。この大学には 11,500 人のフルタイムの学部生がいる。

調査者は学生が学期中に有給の仕事で稼ぐ平均週給を ± 3 ポンド ($\pm \pounds 3$) 以内で推定したい。この調査者が単純無作為抽出を用いた場合、この目的を達成するためには約 510 人の標本が必要であることを示せ。ただし母標準偏差は $\pounds 35.38$ と仮定する。

この調査で起こり得る実際的な困難について簡潔に論じよ。

(6)

BLANK PAGE

BLANK PAGE

BLANK PAGE