



ROYAL
STATISTICAL
SOCIETY
DATA | EVIDENCE | DECISIONS

2016 年 RSS/JSS 試験 (Higher Certificate)

HIGHER CERTIFICATE IN STATISTICS, 2016

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こと。

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 ${}_nC_r$ 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. 毎年，ある会社は同社の4つの事務所での従業員のストレスの程度を調査している．今年，同社は5400人の従業員から層化無作為標本を選択することを提案している．抽出された従業員それぞれは訓練を受けた調査者によって面接と身体検査を受け，0～100点の間で，ストレスの“スコア”が与えられる．

同社は面接する従業員の数を決めるにあたって，あなたのアドバイスを求めている．

これまでの年からの結果では，ストレススコアの標準偏差は事務所間で違いがあることが示唆されている．各事務所の推定標準偏差と従業員総数を以下に示す．

事務所番号	従業員総数	推定標準偏差
1	1400	25
2	500	30
3	2500	12
4	1000	20

- (i) 母平均の推定量の分散， V ，が与えられたとき，総標本サイズ n は以下の式で与えられるとする．

$$n = \frac{\sum_h W_h^2 S_h^2 / w_h}{V + \frac{1}{N} \sum_h W_h S_h^2} .$$

この式の中の N , W_h , S_h , w_h を定義せよ．

(3)

- (ii) 同社は従業員のストレススコアの全体平均を95%の確率で真値の d 単位以内（すなわち区間の幅が $2d$ ）で推定することを望んでいる．このことは母平均の推定量の分散， V ，が $V = \left(\frac{d}{1.96}\right)^2$ で与えられる必要があることを示せ．

(4)

- (iii) 最適配分を定義せよ．この調査において比例配分よりも最適配分を選択すべき理由について簡単に論ぜよ．

(4)

- (iv) これまでの年からの結果を用い，最適配分を適用して従業員のストレススコアの全体平均を95%の確率で ± 2.5 単位以内で推定するのに必要となる標本サイズ， n ，と配分のサイズ n_h を計算せよ．標本抽出のコストは各事務所で等しいとする．

(9)

2. 2014年、スコットランド政府はスコットランドの独立に関する国民投票を行った。スコットランドの人々はスコットランドが連合王国の一部として残るべきかどうか投票した。国民投票の選挙人名簿はスコットランドに居住し、適切な市民権をもつ16歳以上の全ての人々であり、その全員が投票権を有した。連合王国内において、16歳や17歳のような若者が投票することができたのは今回が初めてのことであった。

(i) 国民投票の運動期間を通して、様々な調査機関が全国紙や放送局に代わって、投票意向を把握するための世論調査を実施した。調査機関がデータ収集に用いた以下の各方法において**選択バイアス**がどのように関わるか論ぜよ。またこれら3つの方法を用いることの全体としての優劣を簡潔に論ぜよ。

(a) 割当抽出法による対面調査。

(b) 固定電話の無作為番号ダイアル（ランダムデジットダイヤリング）による電話聞き取り調査。

(c) 将来の調査で回答者になることを事前に同意していた人々からの抽出によるインターネット調査。

(14)

(ii) 国民投票に対する人々の投票意向について知ることができるよう、チェックボックス形式で回答できる質問を提案せよ。

(3)

(iii) 国民投票の結果を予測することの難しさを、議会選挙と比べながら簡単に示せ。

(3)

3. ある地域の地方自治体はそこにある公立図書館の利用に関して様々な側面の調査を行っている。

- (a) 中等教育学校の生徒の図書館サービス利用状況を調べるため、単純無作為抽出による調査が行われた。この管轄地域には全体で 12,000 人の生徒がいる。その質問項目のひとつは“過去 4 週間内に公立図書館を何回訪れましたか？”であった。結果は以下のようになった。

過去 4 週間内の 訪問回数	生徒数
0	68
1	99
2	50
3	45
4	130
>4	58
合計	450

それぞれの点推定値と 95% 近似信頼区間を求めよ。

- (i) 管轄内の生徒で過去 4 週間に公立図書館を訪れた者の割合。
(ii) 管轄内の生徒による過去 4 週間の公立図書館の訪問者総数。過去 4 週間で 5 回以上の訪問があった場合は 5 回として平均と標準偏差の計算をせよ。

(12)

- (b) この地域における成人の公立図書館訪問の割合はここ数年減少している。この地域の地方自治体は成人に対する調査を計画し、縦断的調査を検討している。この調査に参加する成人は選挙人名簿から系統抽出によって選ばれる。

- (i) 系統抽出を簡潔に記述し、それが無作為抽出または非無作為抽出のいずれとして扱われるべきか論ぜよ。
(ii) 縦断的調査が何を意味するのか、また、この状況においてそれがどのように有用になるのかを説明せよ。主な欠点は何であろうか？

(8)

4. 中東のある国において、所定の生育期間中にある地区で生産されたバナナの房の総数を推定するために小規模な調査が実施された。その地区は 289 個の一次単位に分けられ、各単位にはおよそ 500–1000 個のバナナピットがある。各ピットが 0, 1 かそれ以上のバナナの房を作る。地区全体のバナナピットの総数は 181,336 であることがわかっている。289 個の一次単位から単純無作為抽出で 20 単位が選ばれ、各単位のバナナピット数 (x) とバナナ房の総数 (y) が得られた。結果の要約は以下のとおりである。

	標本 ($n = 20$)	
	平均	SD
単位あたりのバナナピット数 (x)	644.35	115.9025
単位あたりのバナナ房数 (y)	901.70	221.8112

- (i) 単純無作為標本の平均を用いて、この地区のバナナ房の総数を推定し、その推定値の標準誤差を推定せよ。(5)
- (ii) 調査者は、この地区におけるバナナ房の総数を推定するために、単位あたりのバナナピット数の追加データをどのように使うべきか助言を求めている。彼は単位あたりのバナナピット数とバナナ房の総数（すなわち、上記の x と y ）の間の相関を推定したところ、比推定が適切であろうと考えている。
- (a) *相関*が何を意味するのか説明せよ。相関の推定値が 0.7737 だったとき、どのように答えるか？ 比推定量と(i)での標本平均に基づく推定量の性質について簡潔に論ぜよ。(5)
- (b) この地区のバナナ房の総数の比推定値が 253,760.64、その推定標準誤差が 8,550.947 で与えられている。このときこの推定値が(i)で計算されものよりも優れているかどうか、その理由とともに述べよ。またこれを用いて、この地区のバナナ房の総数の真値の 95% 近似信頼区間を構成せよ。この信頼区間が示すものを説明せよ。(4)
- (iii) このような調査において、集落抽出がどのように有用になりうるか、またそれを実施する場合に起こりうる実際的な困難について、簡単に論ぜよ。集落抽出の欠点は何であろうか？ それにも関わらず集落抽出が有用であるのはなぜであろうか？(6)

BLANK PAGE

BLANK PAGE

BLANK PAGE