

RSS Higher Certificate in Statistics, 2010

Module 8 : Survey sampling and estimation

1. キングスタウンの中小企業の広告費 (advertising expenditure) の研究では, 層化抽出によって以下の結果が得られた.

<i>Employment size-band</i> h	<i>Number of businesses</i> N_h	<i>Achieved sample size</i> n_h	<i>Mean advertising expenditure in 2008 (£000's)</i> \bar{y}_h	<i>Standard deviation of advertising expenditure</i> s_h
Fewer than 5	10 000	100	3.5	3
Between 5 and 9	6 000	60	10.0	5
Between 10 and 20	2 000	15	35.0	15
TOTAL	18 000	175		

- (i) 従業員が 5 人未満の全ての企業の広告費の平均に対する 95%信頼区間を求めよ.
- (ii) この信頼区間とは何を意味しているのか, わかりやすく説明せよ.
- (iii) この信頼区間の幅が £ 500 より大きくならないためには, 標本サイズについてどんな条件が必要か.
- (iv) キングスタウンの全ての企業の広告費の平均に対する 95%信頼区間を求めよ.
- (v) この研究の標本は, 地元企業の詳細な抽出枠から選ばれた. この枠によって, 比推定値が今後は使えるようになるかもしれないと示唆されている. 詳細な枠に含まれている可能性が高く, 比推定値を用いて広告費の平均 \bar{y} の予測を調整するのに役立つ, 補助変数 x の例を挙げよ. 補助変数に必要な性質を説明しながら, 上記の解答の理由を述べよ.

2. チップタウンという地域を囲む新たな道路の提案についていくつかの議論が起きている。町の交通量を減らすはずだという意見もあれば、景観の美しい地域を横切るだろうという意見もある。

地元紙では提案されている道路について読者の意見を決定するために彼らへの調査を行った。「チップタウンの危険な交通状態を軽減するために提案されている道路について賛成しますか?」という質問に対して、単純な「はい」か「いいえ」という投票をテキストメッセージで読者に送ってもらうことによって標本は抽出された。

調査の結果、150 の標本のうち 110 が「はい」と答えた。町の母集団は 25,000 である。

- (a) 地元の学者は調査過程の正当性について深刻な疑問を抱いているということを書き伝えた。彼は抽出の枠とデータの収集法と行われている質問を批判した。その調査結果は非常に偏りがありそうだと彼は結論付けた。
 - (i) 偏りとは何を意味するか説明せよ。
 - (ii) 調査過程に対する彼の特定の批判がどのようなものであったかについて、あなたが思うことを簡単に説明せよ。彼の挙げた要因がどのようにバイアスに影響するかについて述べ、またバイアス以外の点についても触れよ。
- (b) チップタウンは過去 100 年間のうち、さまざまな時代に建設されたエリアの混じった町であると他の居住者が投書で指摘をした。彼は集落抽出を行うべきであると述べている。このアプローチで有利になりそうな点を議論し、クラスタをどのように構成して抽出するか示せ。

3. The Learning and Skills Office (LSO) はある国が将来必要な技術を得ることを確実にするための役割を担っている. 2009 年に中等教育を受けている学生から無作為抽出を行った. 彼らには「あなたは大学の学位を目指して勉強をするつもりですか?」という質問を行った. ある地域では全体で 12,000 人の学生がこれらのコースに所属しており, 8,000 人が都会とされているエリアの学校に通っており, 4,000 人が地方の学校に通っている. この地域での結果は次のようになった.

<i>School location</i> h	<i>Population of secondary students</i> N_h	<i>Achieved sample size</i> n_h	<i>Proportion intending to study for a university degree</i> p_h
Urban	8 000	300	0.3
Rural	4 000	150	0.6
TOTAL	12 000		

- (i) 層化の 2 つの目的を確認せよ. この調査ではこれらはどのように関係するだろうか?
- (ii) 各エリアの学校に通う学生のうち大学の学位を目指しているものの割合についての 95% 信頼区間を求めよ.
- (iii) 大学で勉強するつもりのある学生の割合が都会と地方の間で違いがあるかどうか検定を行え.
- (iv) 大学で勉強するつもりのあるこの地域の学生の総数について推定を行え.
- (v) LSO はこの地域の標本サイズを 500 として来年も同類の調査を行うことを企画している. 統計官は調査を企画する際に *比例配分*か*最適配分*のどちらかを用いるように忠告された. これらの用語が何を意味するか説明し, 2009 年の標本から得られる情報を用いて各方法によって与えられる階層の標本サイズを求めよ.

4. 南イングランド商工会議所は、不景気を考慮してその地域の企業信頼度について評価を企画した。特にその地域の雇用状況の見通しについて興味を持っている。

調査は商工会議所に属している企業からの単純無作為抽出により行われた。これらの企業のうち 100 社から返答があり、そのうち 40%の企業が次の 12 ヶ月間に従業員を増やすと予想した。

このため、地元新聞紙では次のような見出しをつけた：「南イングランド商工会議所が発表した数字によると、ほとんどの企業では次の 12 ヶ月後に雇用を減らすであろうとのこと。」

- (i) 新聞の見出しについて、言い回し、ベースとなっている調査について知っていること、の両方の観点からあなたが疑念を抱く点は何か？
- (ii) 商工会議所が調査の目的についての答えをより強固なものにするために用いることのできた可能な調査方法を簡単に説明せよ。母集団を特定したり、適切な抽出の枠について工夫したり、調査を指揮したり、結果を分析したりするといった問題を含めて議論すること。