

スイカ

日本人は、春夏秋冬の九十日ごとに、生活の仕方を変えていく。そして、四季のおりおりに風物がある。じっとしていても汗がにじみでる日本の夏。スイカは、涼気をさそう夏の風物である。

昨今では、冷房機など便利なものが普及して、私たちもその「恩恵」をうけている。しかし、冷房機をまつまでもなく、昔から、夏を涼しく過ごすための様々な工夫がされてきた。風鈴の音を聞きながら、浴衣姿でスイカを食べる。冷房機といえども、この自然の涼しさには及ばないのではないか。

まないたにスイカをのせ、庖丁で二つに割る。それが色あざやかにできたスイカのときなど、思わずうれしくなってくるものだ。スイカは誰にでも好まれる。アイスクリーム、コーラ、シャーベット……いろいろあるが、日本の夏からスイカを欠かすことはできない。

8月のおもな行事

- 8～10日 就業構造基本調査事務打合せ会(水戸市・土浦市・下館市)
- 15日 終戦記念日
- 20～24日 昭和55年国勢調査調査区設定事務打合せ会(神栖町・水戸市・千代田村・千代川村・水戸市)
- 23～24日 関東ブロック県民所得推計事務研究会(大洗町)
- 30～31日 小売物価統計調査ブロック会議(栃木県)

標 本 設 計 (4)

設 計 資 料 の 紹 介

一 個人、世帯、住宅に関する標本調査のための抽出用単位地域について

一般の個人、世帯、住宅について、広い地域にわたって調査しようとするとき、これらを直接の抽出単位としますと、標本が広くちらばり過ぎて、調査がしにくくなるおそれがあります。標本が広くちらばることは、結果数字の精度の点では望ましいことといえますが、実際には調査に費用がかかり過ぎるなどの理由で、標本をいくつかの地点に集約してしまう方式がとられるのがふつうです。精度で損をしても費用で得をしようという考えです。具体的には、広い地域をたくさん的小地域に区切って、その小地域を抽出単位として抽出するものです。

抽出用の単位地域としては、色々のものが考えられますが、個人、世帯、住宅関係の調査のためには、国勢調査の調査区が一番便利です。この調査区は、もとより国勢調査の調査員の受持区域として設けられた小地域ですが、設定にあたっては、後に行う各種の標本調査にも利用しやすいようにと、種々の配慮がなされています。識別は調査区番号によってなされます。各調査区は国勢調査による豊富な資料を盛り込んだ地域抽出のための基礎資料として総理府統計局に、一部は都道府県と市区町村に、整えられ一般に利用できるようになっています。そのまま利用できる資料もありますが、県、市区町村など対象地域を限定した調査においては、状況にあわせてある程度の再編成を行うのが、より効果的な利用方法です。いずれにしても、上手な標本設計を行うための鍵は、抽出単位である調査区の内容をよく知ることです。そこで、以下、昭和50年国勢調査の調査区(以下、単に調査区といえます。)について一通り説明し、具体的な標本設計はそのあとにします。

用語には特別な意味をもつものもありますから、正しくは昭和50年国勢調査の関係書類や報告書を見て下さい。

1. 1つの調査区について知りうる主要事項

(1) 調査区の所在

①調査区の所在地の名称 「調査区一覧表」という台帳があって、それに「〇〇番地～〇〇番地」などと表わされています。また、調査区が大きな建物の全部または一部であったり、調査区内に目立つ建物があるときは、そのことを付記して調査区の所在地ができるだけはっきりするようにしています。

②調査区の境界、建物の配置状況 調査区ごとに「調査区要図」という地図に示されています。調査区の所在地の確認には必須の資料です。

(2) 調査区の近辺の市街化の状況

①人口集中地区に属しているかどうか 人口集中地区というのは、厳密な定義はここでは省きますが、一言でいえば人口が多く、人口密度も高い市街地のことです。調査区番号の前に付いている前置番号でそれを表わします。前置番号は市区町村単位に、人口規模の大きい順に01, 02, 03, ……となっていますから、前置番号をみれば、その調査区が当該市区町村の中の何番目の人口集中地区の中にあるかがわかります。

②準人口集中地区に属しているかどうか 準人口集中地区というのは、人口が多く、人口密度も高いが、人口規模において人口集中地区に準ずる、という地区のことです。人口集中地区を大き目の市街地とすれば、準人口集中地区は小さ目の市街地ということが出来ます。やはり前置番号で表わされ、市区町村単位に人口規模の大きい順に、51, 52, 53, ……となっていますから、前置番号をみれば、その調査区の近辺にどの程度人口が集中しているかがわかります。

なお、前置番号が00となっている調査区は、人口集中地区にも、準人口集中地区にも属していないということです。

(3) 常住者の希少および地形・地物等の特殊性

①調査区が山岳・森林・原野地帯、広大な耕地、広大な埋立地・干拓地、広大な池・沼・湖、広大な公園・広場・墓地、無人島(常住者の非常に少ない島を含む)に該当していれば、調査区番号のあとに付けられた後置番号の「2」で識別されます。

②広大な工場・学校等のある地域、広大な鉄道用地・港湾施設・飛行場等のある地域に該当する調査区は、後置番号の「3」で識別されます。

(4) 居住形態および居住地域の特殊性

①社会施設、大きな病院のある地域の調査区は、後置番号の「4」で識別されます。

②刑務所・拘置所等のある地域の調査区は、後置番号の「5」で識別されます。

- ③自衛隊地域の調査区は後置番号の「6」で識別されます。
- ④駐留軍地域の調査区は後置番号の「7」で識別されます。
- ⑤おおむね50人以上の単身者が居住している寄宿舍・寮等のある地域の調査区は、後置番号の「8」で識別されます。

⑥港の区域および水上生活者のいる河川、運河の河口に近い水域の調査区は、後置番号の「9」で識別されます。

上記の(3)~(4)の⑤は特別調査区、⑥は水面調査区と呼ばれています。これらの調査区は、全体からみると例外的なものです(参考表-1)。

参考表-1 茨城県内の後置番号別調査区数

後置番号	調査区数
1	11 402
2	63
3	54
4	121
5	6
6	10
7	0
8	193
9	11
計	11 860

調査区が上記の(3)~(4)の⑥のいずれにも属さなければ、後置番号は1とし、一般調査区と呼ばれます。一般調査区の設定基準は、原則として約50世帯を目安としていますが、ある程度の幅はやむを得ないとしても、なるべく30~70世帯におさまるようにしています。しかし、これは設定の基準ですから、その後の調査の段階で世帯数がこの基準からはみ出すことはあります。

⑦大きな建物の一部からなる調査区は、その建物や地域全体の調査区番号と後置番号は同じとし、その後、51, 52, 53, ……と分割番号を、また、既設定の調査区を分割修正した場合は、調査区番号と後置番号はそのままとし、その後、1, 2, 3, ……と分割番号をつけて区別します。

(5) 調査区内の人口、世帯数

①総人口、19歳までの各歳人口、20歳から64歳までの5

歳階級人口、65歳以上人口。これらは男女計と男の別に表わされています。(女は差引きでだせる)

②15歳以上人口、労働力人口、就業者数、雇用者数、自営業主数、家族従業者数、産業大分類別就業者数および雇用者数、完全失業者数。

③世帯総数、普通世帯数およびその世帯人員、世帯人員が6人以下の普通世帯について世帯人員別世帯数、世帯人員が7人以上の普通世帯数、準世帯数およびその世帯人員、1人の準世帯数、2~29人、30人以上、50人以上の準世帯数および世帯人員。

④核家族世帯数、夫婦だけの世帯数、夫婦と子供のいる世帯数、片親と子供の世帯数。

⑤持ち家に住む普通世帯数、公営の借家に住む普通世帯数、民営の借家に住む普通世帯数、給与住宅に住む普通世帯数、間借りの普通世帯数、寄宿舍等に住む普通世帯数、1普通世帯あたり室数、同量数。

⑥自宅で働く15歳以上の就業者数、自宅以外自市区町村内の従業先で働く15歳以上の就業者数、県内他市区町村の従業先で働く15歳以上の就業者数、15歳以上の通学者数、通学地が自市区町村内にある15歳以上の通学者数、通学地が県内他市区町村にある15歳以上の通学者数、通学地が他県である15歳以上の通学者数。

⑦世帯主が非農林漁業就業者である1人の普通世帯数、世帯主が雇用者である1人の普通世帯数、2人以上の普通世帯数、農林漁業の業主と家族従業者がいない2人以上の普通世帯数、世帯主が雇用者(産業を問わず)で世帯に農林漁業の業主と家族従業者がいない2人以上の普通世帯数、農林漁業の業主と家族従業者がいる2人以上の普通世帯数、世帯主が農林漁業の業主か家族従業者以外の就業者である2人以上の普通世帯数、世帯主が雇用者(産業を問わず)で世帯に農林漁業の業主か家族従業者がいる2人以上の普通世帯数、1人の普通世帯および準世帯に住む学生数・社会施設入居者数・入院患者数・農林漁業の業主と家族従業者数・農林漁業の業主と家族従業者を除く各産業の就業者数・その他の世帯員数。

⑧換算世帯数 世帯には普通世帯と準世帯がありますが、世帯員数が大きく異なるため、世帯がすべて2人以上の普通世帯であるとしたときの世帯数を見積るために、次の

●シリーズ「短期統計実務講座」*****

式によって指標を作り、換算世帯数と呼んでいます。

$$(2人以上の普通世帯数) + \frac{(1人の普通世帯数)}{3} \times \frac{(および準世帯人員)}{3} \quad (\text{四捨五入})$$

2人以上の普通世帯における15歳以上の世帯人員は約3人ですから、1人の普通世帯数と準世帯人員の合計を3で割り、これに2人以上の普通世帯数を加えて換算世帯数とします。就業人口などを調査する標本調査で調査区を抽出単位とするとき、調査区の大きさを示す1つの指標として使われます。なお、換算世帯数は3倍すると調査区内の15歳以上人口に近い数字になります。

⑨ウェイト 換算世帯数を、次のように15を1単位とし

換算世帯数	1—15	のとき	ウェイト	1
"	16—30	"	"	2
"	31—45	"	"	3

て表わしたもので、調査区を人口や世帯数に応じた確率で抽出する際の指標として使われます。ただし、後置番号が5, 6, 7, 9の調査区と世帯数0の調査区はすべてウェイトは1としています。したがって、ウェイトに比例した確率で調査区を抽出するような場合は、後置番号5, 6, 7, 9の調査区と世帯数0の調査区は別枠にする必要があります。なお、調査によっては、このウェイトで抽出するのでは粗っぽいということがあるかもしれません。そのときは換算世帯数を直接用いてもよいし、10か5程度の区分で新しいウェイトを作ることもできます。

(6) 調査区の種類符号

これは、調査区をその地域的特殊性、世帯の種類（普通か準か）による特性、世帯数の多少、産業構成からみた特徴を多重的に分類し、符号化したものです。したがって、調査区の種類符号をみれば、その調査区の前記の観点による特性を知ることができます。この分類の一番活きた利用は、調査区を抽出するときの層とするところにあります。特に産業構成による分類にはその意味が強調されています。（昭和45年の資料においては、産業別人口の構成比の基数は15歳以上人口でしたが、昭和50年の資料では就業者数となっています。すなわち、昭和45年の資料が産業別就業者および不就業者の精度を考慮した資料であったのに対して、昭和50年の資料では産業別就業者の精度だけを考慮するよう変更されたということです。）

符号とその意味は次に列挙するとおりです。符号はおおむね、前記の観点からみた特殊性が強いほど若くなってい

ます。もし、1つの調査区が2つ以上の符号に該当するときは、若い方に所属させることになっています。

分類は、大分類17区分、細分類11区分の計28区分からなっています。細分類の11区分は「50人以上の準世帯が1つ以上ある調査区(分類符号02)」を細分したものです。

市区町村レベルなどでの標本設計では、調査区の総数が少なくなって、細分類までを層とすると抽出に支障をきたすことがあるかもしれません。そのときは、類似の層を合併したり、調査区数は少いが特殊性が強い層に対しては全数抽出をするとか、その場に応じていろいろと工夫して利用することができます（参考表—2）。

参考表—2 茨城県内の分類符号別調査区数

分類符号	調査区数
01	27
02	208
03	77
04	200
05	34
06	158
07	8
08	26
09	413
10	1 147
11	87
12	504
13	3 881
14	1 259
15	2 274
16	1 473
17	84
計	11 860

分類符号01 調査区の後置番号が5, 6, 7, 9の調査区

分類符号02 50人以上の準世帯が1つ以上ある調査区(これは次の11の細分類の全体にあたる)

分類符号0251 学校の寮・寄宿舎のある地域

分類符号0252 病院・療養所のある地域

分類符号0253 社会施設のある地域

分類符号0254 鉱業の寮・寄宿舎のある地域

分類符号0255 建設業の寮・寄宿舎のある地域

分類符号0256 製造業の寮・寄宿舎のある地域

分類符号0257 卸売業、小売業、金融・保険業、および
不動産業の寮・寄宿舎のある地域

分類符号0258 運輸・通信業、および電気・ガス・水道・
熱供給業の寮・寄宿舎のある地域

分類符号0259 サービス業の寮・寄宿舎のある地域

分類符号0260 公務の寮のある地域

分類符号0261 その他の寮・寄宿舎のある地域

(上記分類符号0251～0261における寮・寄宿舎とは50人以上の準世帯をなしている寮・寄宿舎のことです)

分類符号03 世帯総数が0の調査区

分類符号04 換算世帯数が15以下の調査区

分類符号05 換算世帯数が16以上で、15歳以上就業者中に占める漁業、水産養殖業の就業者数の割合が3割以上の調査区

分類符号06 換算世帯数が16以上で、15歳以上就業者中に占める漁業、水産養殖業の割合が1割以上3割未満の調査区

分類符号07 換算世帯数が16以上で、15歳以上就業者中に占める鉱業の就業者数の割合が3割以上の調査区

分類符号08 換算世帯数が16以上で、15歳以上就業者中に占める鉱業の就業者数の割合が1割以上3割未満の調査区

分類符号09 換算世帯数が16以上で、15歳以上就業者中に占める建設業および製造業の自営業主数の割合が1割以上の調査区

分類符号10 換算世帯数が16以上で、15歳以上就業者中に占める卸売業、小売業の自営業主数の割合が1割以上の調査区

分類符号11 換算世帯数が16以上で、15歳以上就業者中に占めるサービス業の自営業主数の割合が1割以上の調査区

分類符号12 換算世帯数が16以上で、15歳以上就業者中に占める農林業就業者数の割合が7割以上の調査区

分類符号13 換算世帯数が16以上で、15歳以上就業者中に占める農林業就業者数の割合が3割以上7割未満の調査区

分類符号14 換算世帯数が16以上で、15歳以上就業者中に占める農林業就業者数の割合が1割以上3割未満の調査区

分類符号15 換算世帯数が16以上で、15歳以上就業者中に占める建設業、製造業の雇用者数の割合が3割以上の調査区

分類符号16 換算世帯数が16以上で、15歳以上就業者中に

占める第3次産業の雇用者数の割合が3割以上の調査区
分類符号17 その他の調査区

2. 抽出のための調査区リストと各種の基本数一覧表

(1) 分類符号別通し番号による調査区一覧表

これは、調査区を市区町村ごとに特に分類符号の別に並び、一連番号、ウェイト、ウェイトの累計などを記載したものです。調査区を層化して抽出するときの枠に利用できる大変便利な表です。

(2) 30人以上の準世帯のある調査区および分類符号02に該当する調査区一覧表(都道府県別)

細分類符号ごとに、都道府県内の一連番号およびウェイトの累計などが記載されています。

以下の各一覧表は、いずれも各種の抽出用基本数を表わすものです。

(3) 市区町村別分類符号別調査区数、ウェイト一覧表

(4) 市区町村別分類符号別世帯数、世帯人員一覧表

(5) 分類符号02の細分類符号別単位区数等、ウェイト総数一覧表(都道府県別)

(6) 分類符号、産業、職業別就業者数および産業、世帯の種類、住居の所有関係別世帯数一覧表(都道府県別)

これは、就業者数と世帯数の実数のみならず、1調査区あたりの平均値、調査区間変動係数、1ウェイトあたり平均値、ウェイトつき調査区間変動係数も計算されていますから、きめの細かい設計をする際、大変参考になるものです。

3. 資料の所在と利用の手続き

上記の資料はいずれも総理府統計局に整えられておりますが、このうち1の(1)の②は市区町村の統計担当係でも、2の各表は都道府県の統計主管課でも、そのほかは都道府県の統計主管課と市区町村の統計担当係でもわかります。一般の利用者でも、申請して閲覧・複写することができます。

以上、国勢調査の調査区について、地点抽出の際の抽出単位とする立場から、そのあらましを述べました。紙面の都合でふれられなかった点もありますが、それについては、具体的な標本設計の説明の際にふれることとします。

〔注意〕

1. 上記に関するさらに詳細なことは、総理府統計局「昭和50年国勢調査 調査区関係資料利用の手引き」を見て下さい。

2. 上記資料の作成には筆者は関与しておりません。(昭和45年の調査区の分類には関与しました。)

ローレンツ曲線

度数分布表を発展させたものに累積度数表があります。ローレンツ曲線は、この累積度数表を応用したものです。今回は、累積度数表をふまえて、ローレンツ曲線について述べてみます。

1. 累積度数表

累積度数とは、普通の度数分布の級度数を变量の小さいほうから順次積み重ねていったものです。(ここで、变量には連続量と離散量があります。連続量とは、年齢、賃金、身長、体重など連続的に変化する量です。そして離散量とは、世帯の員数、事務所の従業員数、住宅の間数など常に整数値で変化する量です。)累積度数分布の例として、表-1をみてください。表-1は、「茨城の工業(昭和52年工業統計調査結果報告書)」から作成したものです。以下、この例をもとにして論をすすめます。

表-1 累積度数分布の計算(1)

従業者規模	事業所数	同百分比	同累積
人	所	%	%
総数	12 132	100.0	
1 ~ 3	4 120	34.0	34.0
4 ~ 9	4 141	34.1	68.1
10 ~ 19	1 679	13.8	81.9
20 ~ 29	783	6.5	88.4
30 ~ 99	1 044	8.6	97.0
100 ~ 299	280	2.3	99.3
300 ~	85	0.7	100.0

累積度数分布をグラフに書くときは、連続量の場合は折れ線グラフの手法をもちい、積み重ねた柱の対角線をつらねた形で表わされます。また、離散量の場合には、累積度数は段階状に変化しますから、そのグラフも段階状に表わされます。

2. ローレンツ曲線

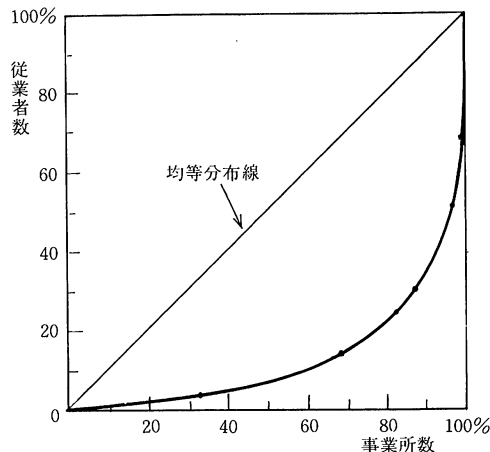
累積度数を応用したものに、ローレンツ曲線があります。本来は、所得分布をあらわす方法として考案されたものです。一般には、分布の集中度(あるいは不平等度)を示す手段として用いられています。

ローレンツ曲線を書くには、級度数と級の变量和が必要です。そしてそれぞれ累計して百分比に換算します。この百分比の組を座標として書いたものが、ローレンツ曲線です。表-1の「事業所数」を横軸に、そして表-2の「従業員数」を縦軸にすれば、従業員分布の集中度を示すローレンツ曲線は、図-1のように表わされます。

表-2 累積度数分布の計算(2)

従業者規模	従業者数	同百分比	同累積
人	人	%	%
総数	255 997	100.0	
1 ~ 3	8 545	3.3	3.3
4 ~ 9	25 305	9.9	13.2
10 ~ 19	23 510	9.2	22.4
20 ~ 29	18 861	7.4	29.8
30 ~ 90	54 593	21.3	51.1
100 ~ 299	44 341	17.3	68.4
300 ~	80 842	31.6	100.0

図-1 ローレンツ曲線



ローレンツ曲線では、分布の集中度(あるいは不平等度)は、曲線と対角線(均等分布線と言います)が作る弓形の面積で測定することができます。この面積は、対角線の下方の直角三角形の面積より大きくなることはありません。四角形の面積を1とすれば、弓形の面積は0と0.5の間で変化します。数値が0.5に近づけば近づくほど集中度

..... 水戸県税事務所 上 沢 光 男

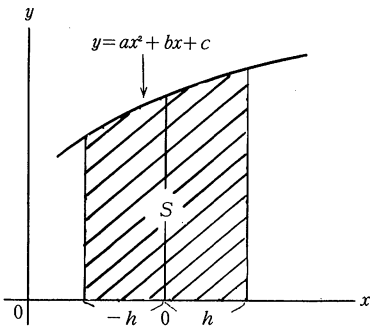
は大きく、逆に 0 に近ければ近いほど集中度は小さいとい
うことになります。

この面積を λ (ラムダ) で示すと、ローレンツ曲線の集
中度 λ は、シンプソンの近似式を応用した次の式で計算す
ることが出来ます。

$$\lambda = \frac{1}{6} \{ 2.5 - (2y_{25} + y_{50} + 2y_{75}) \}$$

(ただし、 y_{25} 、 y_{50} 、 y_{75} は、それぞれ横軸の 25%、50%、75%
点における曲線の高さ(百分比)です。)

ここで、シンプソンの近似式をもちいて、この式を導い
てみましょう。ローレンツ曲線は二次曲線ではありません
から、その極小部分を取りだして、それを二次曲線と近似
させます。



斜線の部分の面積 S は、次のように表わされます。

$$\begin{aligned} S &= \int_{-h}^h (ax^2 + bx + c) dx \\ &= \left[\frac{a}{3} x^3 + \frac{b}{2} x^2 + cx \right]_{-h}^h \\ &= \left(\frac{a}{3} h^3 + \frac{b}{2} h^2 + ch \right) - \left(-\frac{a}{3} h^3 + \frac{b}{2} h^2 - ch \right) \\ &= \frac{2}{3} ah^3 + 2ch \\ &= \frac{h}{3} (2ah^2 + 6c) \dots\dots\dots(1) \end{aligned}$$

$f(x) = ax^2 + bx + c$ とすると、

$$f(-h) = ah^2 - bh + c$$

$$4f(0) = 4c$$

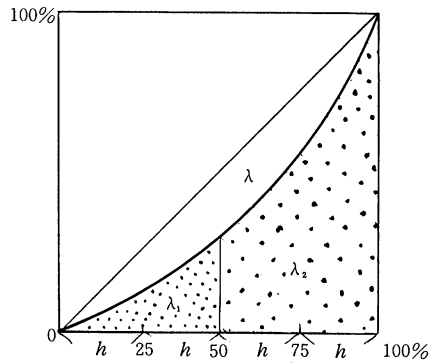
$$f(h) = ah^2 + bh + c$$

$$\therefore f(-h) + 4f(0) + f(h) = 2ah^2 + 6c \dots\dots\dots(2)$$

(1)(2)から、

$$S = \frac{h}{3} \{ f(-h) + 4f(0) + f(h) \} \dots\dots\dots(3)$$

(3)式がシンプソンの近似式と言われているものです。この
式を用いて、 λ (集中度) の面積を求めます。



λ_1 と λ_2 の面積は(3)式により、

$$\lambda_1 = \frac{0.25}{3} (0 + 4y_{25} + y_{50})$$

$$\lambda_2 = \frac{0.25}{3} (y_{50} + 4y_{75} + 1)$$

λ の面積は、 $\lambda = \frac{1}{2} - (\lambda_1 + \lambda_2)$ ですから、

$$\begin{aligned} \lambda &= \frac{1}{2} - \left\{ \frac{0.25}{3} (4y_{25} + y_{50} + y_{50} + 4y_{75} + 1) \right\} \\ &= \frac{1}{2} - \left\{ \frac{0.25}{3} (4y_{25} + 2y_{50} + 4y_{75} + 1) \right\} \\ &= \frac{1}{2} - \frac{1}{3} y_{25} - \frac{0.5}{3} y_{50} - \frac{1}{3} y_{75} - \frac{0.25}{3} \\ &= \frac{1}{6} \{ 2.5 - (2y_{25} + y_{50} + 2y_{75}) \} \end{aligned}$$

以上が、シンプソンの近似式を用いた λ (集中度) の導き
方です。

この算式から、図-1 の λ (ここでは従業員分布の集中
度) を計算してみましょう。 $y_{25} \approx 0.02$ 、 $y_{50} \approx 0.10$ 、 $y_{75} \approx 0.20$
ですから、この数値を算式に代入して、

$$\begin{aligned} \lambda &= \frac{1}{6} \{ 2.5 - (0.02 \times 2 + 0.10 + 0.20 \times 2) \} \\ &= \frac{1}{6} (2.5 - 0.54) \\ &\approx 0.327 \end{aligned}$$

となります。