

# 付3 標本抽出方法，結果の推定方法及び推定値の標本誤差

## 1 標本抽出の方法

この調査は，層化2段抽出法による標本調査であり，調査区\*を第1次抽出単位とし，住戸\*\*を第2次抽出単位としている。

\* 国勢調査調査区

\*\* 住宅やその他の建物の各戸で，一つの世帯が居住できるようになっている建物又は建物の一区画

### (1) 調査区の抽出（第1次抽出）

第1次抽出における調査区の抽出は，各地域\*ごとにすべての調査区を国勢調査の結果等に基づく特性により層に分けて，各地域の各層ごとに，所定の抽出率と任意の抽出起番号を用いて系統抽出法により行う。この系統抽出は，各調査区のウエイト（15世帯がほぼ1ウエイトとなるように各調査区に付されている値）に基づく確率比例抽出によっている。毎月の標本調査区数は約2,900となっている。

ただし，刑務所・拘置所等のある区域\*\*（国勢調査調査区としての後置番号が5の調査区），自衛隊区域\*\*（同6の調査区），駐留軍区域（同7の調査区）及び水面調査区（同9の調査区）については，抽出を行っていない。

\* **北海道**，**東北**（青森県，岩手県，宮城県，秋田県，山形県及び福島県），**南関東**（埼玉県，千葉県，東京都及び神奈川県），**北関東・甲信**（茨城県，栃木県，群馬県，山梨県及び長野県），**北陸**（新潟県，富山県，石川県及び福井県），**東海**（岐阜県，静岡県，愛知県及び三重県），**近畿**（滋賀県，京都府，大阪府，兵庫県，奈良県及び和歌山県），**中国**（鳥取県，島根県，岡山県，広島県及び山口県），**四国**（徳島県，香川県，愛媛県及び高知県）及び**九州**（福岡県，佐賀県，長崎県，熊本県，大分県，宮崎県，鹿児島県及び沖縄県）の10地域。

\*\* 刑務所・拘置所等のある区域及び自衛隊区域について，それぞれ法務省，防衛省からそれら施設内の居住者数の資料を得て集計に加えている。ただし，詳細結果では，調査の性格上これらについて集計対象とはしていない。

### (2) 住戸の抽出（第2次抽出）

第2次抽出における住戸の抽出は，第1次抽出で抽出された調査区（以下「標本調査区」という。）にあるすべての住戸のうちから，1調査区当たりほぼ15となるように所定の抽出率（ウエイトの逆数に等しい。）及び抽出起番号を用いて系統（等間隔）抽出により行う。抽出された住戸に居住するすべての世帯（合計約4万世帯）が調査対象となる。

ア 月や年の結果の精度と，月々及び年間の変化を見る場合の精度とを考慮し，一つの標本調査区は4か月間調査を行い，前半（2か月間）と後半（2か月間）とで調査区内の調査世帯（第2次抽出で抽出された住戸に居住する世帯。）を替えている。

イ 前年の結果との比較の精度を高めるため，標本調査区として選定された調査区は，翌年の同月に再び調査を行う\*。

すなわち，毎月の全標本調査区のうち，半数はその年新たに調査を行う調査区（したがって，翌年同月に再び調査を行う調査区。以下「1年目調査区」という。）となり，残り半数は前年同月に調査を行った調査区（以下「2年目調査区」という。）となるようにしている。

\* 各標本調査区について，翌年までに無くなった住戸に居住していた調査世帯は調査から除かれる。

一方，新設された住戸は名簿に追加され，その名簿から住戸が追加抽出されそこに居住する世帯が調査世帯に追加される。

ウ 以上の標本交替を行うため及び推定値の標本誤差の算出のため，標本調査区は，調査開始月（A，B，C又はDで表す。）及び1年目調査区か2年目調査区か（それぞれ1又は2で表す。）により区分され次のような8組の副標本で構成されている。なお，各副標本は，それぞれ同等な全国の無作為標本となるように設計されている。

#### 8組の副標本

A 1……… 1月，5月又は9月に調査開始の1年目調査区

A 2……… 1月，5月又は9月に調査開始の2年目調査区

B 1……… 2月，6月又は10月に調査開始の1年目調査区

B 2……… 2月，6月又は10月に調査開始の2年目調査区

C 1……… 3月，7月又は11月に調査開始の1年目調査区

- C 2……… 3月，7月又は11月に調査開始の2年目調査区
- D 1……… 4月，8月又は12月に調査開始の1年目調査区
- D 2……… 4月，8月又は12月に調査開始の2年目調査区

このように、副標本8組のうち、4組は1年目調査で、残り4組は2年目調査区となる。

この結果、いずれの月においても、これらの副標本のうち、2組（すなわち標本調査区の数にすると1/4）について標本調査区の交替が行われ、他の2組について同一調査区の中で調査世帯の交替が行われる。

したがって、標本調査区が交替する組と標本調査区の中の調査世帯が交替する組とを合わせると、毎月1/2の調査世帯が更新されることになる。

なお、特定調査票の調査世帯は2年目2か月目に当たる2組のもの（A 2及びC 2の組又はB 2及びD 2の組）である\*。

\* 詳細結果の調査規模は基本集計の約4分の1となっている。

## 2 結果の推定方法

### (1) 結果の推定（基本集計）

ア 毎月の全国結果は、大都市部・非大都市部\*、男女、年齢5歳階級（13区分）別に、国勢調査に基づく推計人口を基準人口（(3)参照）とする比推定によって算出している。

\* 全国を大都市部（東京都区部，横浜市，名古屋市，京都市，大阪市，神戸市及び北九州市）とそれ以外に分けたもの。

算出の基本式は、次のとおりである（就業者数の例）。

$$\text{就業者数} = \text{線型推定による就業者数} \times \frac{\text{基準人口}}{\text{線型推定による人口}}$$

（注）線型推定：調査で得られた人口に抽出率の逆数を掛け、全体の人口を推計すること

イ 四半期平均，年平均等の平均結果は，該当する期間の月次結果を単純平均して算出している。

### (2) 推定の手順（基本集計）

ア 全国結果の算出手順は、以下のとおりである。

- (i) 各標本調査区の男女，年齢階級別調査人口に線型推定用乗率を乗じて必要な合算を行い，男女，大都市部・非大都市部，年齢階級別人口の線型推定値を算出する。
- (ii) 男女，大都市部・非大都市部，年齢階級別に，基準人口をそれぞれ(i)で算出した線型推定値で除し，比推定用乗率を算出する。
- (iii) 各標本調査区の属性を有する男女，年齢階級別調査人口に，線型推定用乗率を乗じて必要な合算を行い，さらに(ii)で算出した比推定用乗率を乗じて，男女，大都市部・非大都市部，年齢階級別の比推定値  $\hat{X}$  を算出する。
- (iv) この比推定値  $\hat{X}$  を大都市部と非大都市部について合算した後，必要に応じて，男女，年齢階級について合算して各種の結果数字を得る。

（参考）上記(i)，(ii)，(iii)をまとめて計算式で表すと，次のとおりである。

$$\hat{X} = \frac{\prod_{i=1}^L \frac{1}{m_i} \prod_{j=1}^{m_i} \omega_{ij} \cdot f_{ij} \cdot x_{ij}}{\prod_{i=1}^L \frac{1}{m_i} \prod_{j=1}^{m_i} \omega_{ij} \cdot f_{ij} \cdot P_{ij}} = \frac{\prod_{i=1}^L \prod_{j=1}^{m_i} x_{ij} \cdot F_i}{\prod_{i=1}^L \prod_{j=1}^{m_i} P_{ij} \cdot F_i}$$

ここで

$i$  : 10地域，層による区分の番号 ( $i=1,2,\dots,L$ )

$j$  : 各区分中の標本調査区の番号 ( $j=1,2,\dots,m_i$ )

$x_{ij}$  : 第*i*区分，第*j*標本調査区内の属性*X*を有する調査人口（男女，年齢階級別）

$\omega_{ij}$  : 第*i*区分，第*j*標本調査区のウェイト

$f_{ij}$  : 第*i*区分，第*j*標本調査区の住戸の抽出率の逆数 ( $f_{ij} = \omega_{ij}$ )

$\omega_i$  : 第*i*区分に含まれるすべての調査区のウェイトの合計

$m_i$  : *i*区分の標本調査区数

$F_i$  : *i*区分の線型推定用乗率 ( $F_i = \omega_i / m_i$ )

$P$  : 基準人口（男女，大都市部・非大都市部，年齢階級別）

$P_{ij}$  : 第*i*区分，第*j*標本調査区内の調査人口（男女，年齢階級別）

$$\frac{P}{\prod_{i=1}^L \prod_{j=1}^{m_i} P_{ij} \cdot F_i} \text{ は比推定用乗率}$$

イ 地域別結果は、全国結果と同様の方法で各地域ごとの結果を求めた上で、10 地域の合計が全国結果と一致するように補正して算出している。

(3) 基準人口及び基準人口の切替え

ア 結果を算出するための基準人口は、「人口推計」(総務省統計局)による毎月1日(=前月末日)現在の推計人口を用いている。

この推計人口は、国勢調査による人口を基礎に、その後の各月における人口の動きを他の人口関連資料から得て、これを加減することによって算出している。

イ 基準人口となる推計人口は、最新の国勢調査の人口を基礎としており、5年ごとに新たな国勢調査の確定人口に基づく計算値に切替えが行われる。このため、労働力調査の結果の算出に用いる基準人口も5年ごとに切替えが行われる。旧基準による推計人口と新基準による推計人口との間に相違がある場合、労働力調査の基準人口の切替えが行われた年の結果には、統計上の不突合分が含まれる。

なお、労働力調査における現在の基準人口は、平成12年国勢調査の確定人口に基づくもので、旧基準の平成7年国勢調査から新基準の平成12年国勢調査への切替えは、平成14年1月に行った。

(参考1) 基準人口の切替え等による変動分(推計;概数) [単位 万人,ポイント]

	15歳以上人口	就業者数	完全失業者数	非労働力人口	完全失業率
昭和57年結果 (昭和55年国勢調査基準への切替え)	- 4	- 3	0	- 1	0.0
昭和62年結果 (昭和60年国勢調査基準への切替え)	+ 7	+ 4	0	+ 3	0.0
平成4年結果 (平成2年国勢調査基準への切替え)	- 11	- 7	0	- 4	0.0
平成9年結果 (平成7年国勢調査基準への切替え)	+ 28	+ 17	+ 1	+ 10	0.0
平成14年結果 (平成12年国勢調査基準への切替え)	- 6	- 4	0	- 2	0.0

(4) 詳細結果の推定

四半期平均及び年平均結果は、該当する期間の月次結果を単純平均して算出している。

月次結果については、毎月の基本集計結果の男女、年齢10歳階級(5区分)、就業状態(就業者、完全失業者、非労働力人口)別人口を基準人口とする比推定によって算出している。

算出の基本式は、次のとおりである。

(特定調査票A欄(就業者に係る項目)の項目の例)

$A \text{ 欄の推定値} = \text{線型推定値による} A \text{ 欄の値} \times (\text{基本集計の就業者数} / \text{詳細結果の就業者数})$

なお、線型推定値は、基本集計結果の算出の際に用いた線型推定用乗率による集計値である。

### 3 推定値の標本誤差

標本誤差の大きさは、推定値の大きさのほか、調査項目の種類や調査年又は月によって異なる。その目安となる標準誤差は、1の(2)で述べた副標本を用いて計算している。

#### (1) 全国結果の推定値の大きさ別標準誤差（基本集計）

年平均結果の標準誤差			毎月の結果の標準誤差		
推定値の 大きさ (万人)	標準誤差 (万人)	誤差率 (%)	推定値の 大きさ (万人)	標準誤差 (万人)	誤差率 (%)
5000	14.2	0.3	5000	26.6	0.5
2000	8.7	0.4	2000	17.0	0.8
1000	6.0	0.6	1000	12.1	1.2
500	4.2	0.8	500	8.6	1.7
200	2.6	1.3	200	5.5	2.7
100	1.8	1.8	100	3.9	3.9
50	1.2	2.5	50	2.8	5.6
20	0.8	3.8	20	1.8	8.9
10	0.5	5.3	10	1.3	12.7

これらの表に示されている誤差率は、項目ごとの誤差率を曲線の当てはめにより平均的に評価したものである。また、標準誤差は、推定値の大きさに誤差率を乗じて算出している。

月別値及び年平均値の誤差率は下記の算式により近似的に与えられる。

#### 【年平均値用】

$$\sqrt{\frac{1}{8(8-1)} \sum_{i=1}^8 (\bar{X}_i - \bar{X})^2} / \bar{X}$$

ここで、 $\bar{X}_i$ は第*i*副標本による、 $\bar{X}$ は全標本による属性*X*を有する人口の推定値の年平均値を表す。

#### 【月別値用】

$$\sqrt{\frac{1}{8(8-1)} \sum_{i=1}^8 (\hat{X}_i - \hat{X})^2} / \hat{X}$$

ここで、 $\hat{X}_i$ は第*i*副標本による、 $\hat{X}$ は全標本による属性*X*を有する人口の月別推定値を表す。

#### (2) 全国結果の標準誤差（基本集計）

##### 主要項目の年平均結果の標準誤差

項 目	平成18年 平均結果 (万人)	標準誤差 (万人)	誤差率 (%)	項 目	平成18年 平均結果 (万人)	標準誤差 (万人)	誤差率 (%)
労働力人口	6657	18	0.3	製造業就業者	1161	8	0.7
就業者	6382	20	0.3	情報通信業就業者	181	3	1.8
自営業主	633	8	1.3	運輸業就業者	324	4	1.2
家族従業者	247	7	2.8	卸売・小売業就業者	1113	5	0.4
雇用者	5472	11	0.2	金融・保険業就業者	155	2	1.5
完全失業者	275	2	0.8	飲食店、宿泊業就業者	337	6	1.6
非労働力人口	4355	18	0.4	医療、福祉就業者	571	4	0.7
農林業就業者	250	6	2.5	教育、学習支援業就業者	287	3	1.2
建設業就業者	559	8	1.5	サービス業就業者	938	8	0.9

(3) 地域別結果の標準誤差（基本集計）

全国結果と同じ方法によって、推定値の大きさ別に平均的に評価した地域別結果（四半期平均）の誤差率は、次表のとおりである。また、年平均結果の誤差率は、各地域とも四半期結果の約 1/2 である。

地域別四半期平均結果の標準誤差率

推定値の 大きさ(万人)	誤差率 (%)									
	北海道	東北	南関東	北関東 ・甲信	北陸	東海	近畿	中国	四国	九州
2000			0.6							
1000			0.9			0.7	0.8			0.7
500		0.9	1.3	0.9		1.1	1.1	0.8		1.0
200	1.3	1.5	2.2	1.5	1.2	1.8	1.8	1.4	1.1	1.6
100	2.0	2.1	3.2	2.1	1.7	2.5	2.7	2.0	1.7	2.3
50	3.0	3.1	4.6	3.1	2.6	3.7	3.9	3.0	2.5	3.4
20	5.1	5.1	7.5	5.2	4.3	6.0	6.4	4.9	4.3	5.6
10	7.6	7.5	10.8	7.6	6.4	8.6	9.3	7.3	6.4	8.1

(参考2) 主な項目別の不詳などの数

〔単位 万人〕

	就業状態 不詳	就業時間 不詳	従業上の地位 不詳	従業者規模 不詳	分類不能 の産業	分類不能 の職業
平成 18 年平均	8	37	30	39	71	66

4 季節調整値の算出方法

(1) 季節調整の方法は、センサス局法 (X12-ARIMA) の X11 デフォルトを用いている。なお、特異項の管理限界は、下限 9.8 ，上限 9.9 としており、これ以外は標準オプションとしている。

(2) 各系列の季節調整はそれぞれ系列ごとに独立して行っている。

(3) 毎月公表する季節調整値は、前年 12 月までのデータから推計した当該年の予測季節指数により算出している。毎年年初には、前年 12 月までのデータに基づき当該年の予測季節指数とともに、過去にさかのぼって各月の季節指数及び季節調整値の再計算\*を行っている。

なお、長期時系列に示す季節指数及び季節調整値は、平成 18 年 12 月までのデータに基づき算出したものである。

\* 原則として 29 年前までの原数値を用いて再計算を行い、直近 10 年分について再計算結果に改定している。

## 5 都道府県別結果(モデル推計値)の推定方法

### (1) 経緯

都道府県別結果については、平成 14 年から参考として比推定による年平均結果(試算値)の公表を開始した。さらに、時系列回帰モデルによる推計手法を採用して精度の向上を図った上で、平成 18 年 5 月から新たに参考として四半期平均結果(モデル推計値)の公表を開始した。

### (2) 公表系列

モデル推計値は、平成 9 年以降の以下の項目について、都道府県別四半期平均及び年平均結果を公表している。

労働力人口、就業者、完全失業者、非労働力人口、完全失業率

### (3) 推定方法

労働力調査の都道府県別結果を推計する方法については、以下のような 5 つの要素からなる時系列回帰モデルを採用している。

$$\underbrace{Y(t)}_{\text{観測値}} = \underbrace{X(t)}_{\text{回帰}} \underbrace{(t)}_{\text{トレンド}} + \underbrace{T(t)}_{\text{季節変動}} + \underbrace{S(t)}_{\text{不規則変動}} + \underbrace{I(t)}_{\text{標本誤差}} + \underbrace{e(t)}_{\text{標本誤差}}$$

注：観測値とは全国等の結果を求める方法(比推定)による調査結果数値である。

それぞれの要素は次のような変動を表している。

回帰項： 各都道府県の動きと都道府県が属する地域のトレンドとの関係を表す。

トレンド項： 経済の成長などに伴い長期的に変動を示す趨勢変動と、景気の循環に伴う変動などほぼ一定の周期を持つ変動で、周期が 12 か月を超える循環変動とを併せた変動。  
景気の後退と回復によって、完全失業者が傾向的に増加したり、減少したりするような動きのことである。

季節変動項： 12 か月を周期とする季節変動。

例えば、就業者数は新年度の始まる 3 月～4 月から増加し、6 月～7 月にピークとなり、その後の年後半に減少するような動きのことである。

不規則変動項： 趨勢変動、循環変動、季節変動以外の変動で、突発的な出来事による変動や景気の短期的変動。

過去の石油ショックや地震などの自然災害の影響によって起こる一時的な動きのことである。

標本誤差項： 推計モデルでは最も重要な部分で、標本誤差の変動パターンと変動の振れ幅を表している。

回帰項は、トレンドに近い変動を捉えており、回帰項とトレンド項とで趨勢変動及び循環変動を併せた変動と考えることも可能である。回帰項により、時系列的な変動要素に空間(地域)情報も取り入れることになり、より多面的な情報を推計に利用できるものになっている。

この推計方法による都道府県別の推計値は、比推定値(全国と同様の推計方法)から標本誤差項を除くことにより得られる。

なお、相対的に標本規模の大きい北海道、東京都、神奈川県、愛知県、大阪府、沖縄県については、比推定による推計を用いている。これら以外の 41 府県は、標本規模が小さく四半期ごとの結果としては標本誤差が大きいことから、時系列回帰モデルによる推計を用いている。

### (4) 利用上の注意

時系列回帰モデルによる推計では、推計された標本誤差を取り除くことで比推定より結果精度の向上を図っている。しかし、労働力調査は、都道府県別に表章するための標本設計を行っておらず(北海道、沖縄県を除く)、標本規模も小さいことから、都道府県別結果(モデル推計値)については、全国結果に比べ標本誤差が大きいので、結果の利用に当たっては注意を要する。

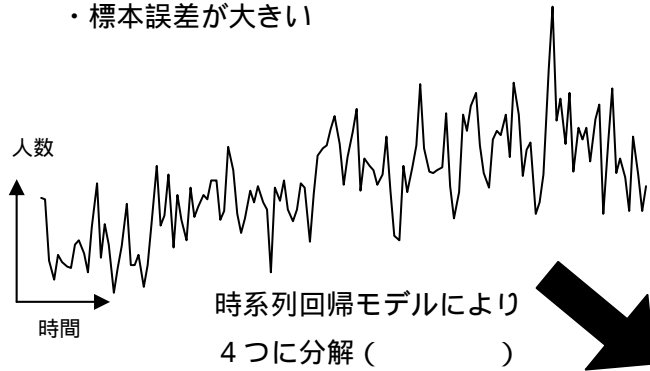
また、毎年 1～3 月期平均公表時に、新たな結果を追加して再計算を行い、前年までの四半期平均及び年平均結果を遡及改定している。

# 都道府県別結果の時系列回帰モデルによる推計方法

## 従来の手法による推計

試算値として公表してきた値の推計方法（都道府県、男女、年齢5歳階級による比推定）では、標本数が少ないため下図のようにぶれが大きい。

従来の手法による推計値（基となるデータ）  
・ 標本誤差が大きい



$$Y_{(t)} = T'_{(t)} + S_{(t)} + I_{(t)} + e_{(t)}$$

- $Y_{(t)}$  比推定値
- $T'_{(t)}$  トレンド (+ 回帰)
- $S_{(t)}$  季節変動
- $I_{(t)}$  不規則変動
- $e_{(t)}$  標本誤差

## 新たな統計的手法による推計

近隣地域の情報を取り入れた回帰式を用いた時系列モデル（時系列回帰モデル）による推計

アメリカ労働統計局も同様の手法を用いている。

トレンド (+ 回帰)  
中長期的な変動を表す。回帰は当該都道府県を含む地域のトレンドにより説明される部分。



季節変動  
1年周期の変動を表す。



不規則変動  
文字通り不規則な変動を表す。短期的な変動などがここに入る。

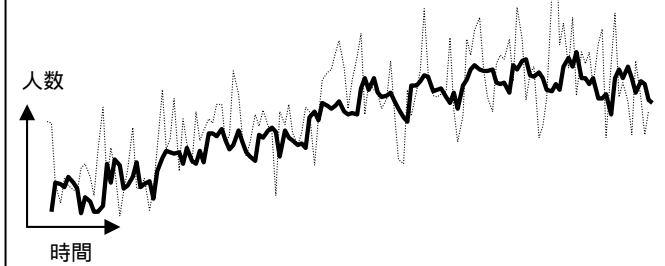


標本誤差【推計により除去する部分】



推計された標本誤差( )を除く  
( )

時系列回帰モデルによる推計  
・ 標本誤差が小さくなり、平滑化される。



- ・ 従来の手法による推計（点線）
- ・ 時系列回帰モデルによる推計（実線）