第７章　結果の推定方法と標本誤差等

労働力調査は標本調査であることから，その結果や誤差は統計理論に基づいて推定される。また，季節性を除去するため，季節調整値を算出している。本章では，労働力調査における結果の推定方法と標本誤差のほか，季節調整等について解説する。

１　線型推定

標本調査は，一部を調査して全体を推定しようとするものであるが，全体の推定値は，標本から得られた値に，抽出率（抽出確率－標本として選ばれる確率）の逆数を乗じることによって得ることができる。このような推定を線型推定という。

労働力調査の場合に当てはめれば，調査区の抽出の際に層別抽出を行っているので，まず各層で独立に推定を行い，次に，各層における推定値を足し合わせて全体の推定値を得ることになる。第l層における推定方法について，第l層の就業者数X_lを推定する場合を例に説明することにする。

第l層で抽出された調査区がm_l個，各調査区のウエイト注)が$w_{li}(i = 1, 2, \cdots, m_l)$であるとし，第$i$調査区において，抽出された住戸全体で就業者が$X_{li}(i = 1, 2, \cdots, m_l)$ 人居住していたとすると，線型推定値は次のようにして求めることができる。

① まず，抽出された調査区内の就業者数の合計を推定する。抽出率はウエイトの逆数としていたから，抽出率の逆数とはウエイトそのものになる。したがって，$X_{li}w_{li}$が第i調査区の就業者数になる。これは，ウエイト３の調査区の場合，三つに一つの割合で住戸を調査するから，調査した住戸に居住する就業者の数が50人であったならば，50×３＝150人をその調査区の就業者数と推定するというものである。別の見方をすれば，抽出された住戸は，３住戸分を「代表」しているのであるから，各住戸の就業者数を３倍し，それを足し合わせれば推定値が得られるともいうことができる。この「代表」の度合いがつまり抽出率の逆数なのである。

② 次に，層全体の就業者数X_lの推定値$\hat{X}_l$を求める。調査区の抽出は，確率比例抽出であったから，各調査区の抽出確率は等しくはなく，したがって推定値も同じ値を一律に乗じて求めるというわけにはいかない。そこで，同じ値を乗じて求める代わりに，調査区内の就業者数に調査区の抽出確率の逆数を乗じて推定値を求めることとする。

第i調査区の就業者数は$X_{li}w_{li}$と推定されていることから，第i調査区一つ

注）第６章の２を参照

から層内全体の就業者数$\hat{X}_l$を推定しようとするとき，層内の全調査区のウエイトの合計をw_l[注)]とすれば，第i調査区からの推定値$\hat{X}_{li}$は，

$$\hat{X}_{li} = (X_{li}w_{li}) \times \frac{w_l}{w_{li}} = X_{li}w_l$$

となり，第i調査区のウエイトw_{li}によらない値となる。このため，$\hat{X}_{li}(i = 1, 2, \cdots, m_l)$の平均値をこの層の推定値と考えることができるから，層全体の就業者数$\hat{X}_l$は，

$$\hat{X}_l = \frac{1}{m_l}\sum_{i=1}^{m_l}\hat{X}_{li} = \frac{1}{m_l}\sum_{i=1}^{m_l}X_{li}\,w_l = \frac{w_l}{m_l}\sum_{i=1}^{m_l}X_{li}$$

と推定できる。ウエイトの逆数を抽出率としたことにより，上の式の$\frac{w_l}{m_l}$は調査区によらない定数になっている。この$\frac{w_l}{m_l}$を線型推定用乗率という。

結局，第l層の就業者の推定式は，各調査区中の就業者数を単純に足し合わせ，$\frac{w_l}{m_l}$倍すればよいという簡単な式になっている。

２　比推定の考え方

１で述べたように，線型推定により推定値は得られるのであるが，補助的な情報を利用することにより，精度を高めることができる。

上の例で考えてみると，第l層の就業者数X_lの線型推定値$\hat{X}_l$を求める方法と全く同様な方法により，第l層の総人口P_lの線型推定値$\hat{P}_l$を求めることができる。すなわちi番目の標本調査区において抽出された住戸にP_{li}人の者が居住していたとすると，

$$\hat{P}_l = \frac{w_l}{m_l}\sum_{i=1}^{m_l}P_{li}$$

となる。$\hat{P}_l$も$\hat{X}_l$も標本からの推定値であるから，標本の選ばれ方によって，実際の値であるP_lやX_lより大きくなったり小さくなったりする。しかし，この二つの推定値の実際の値からのずれ方は，同じ方向であることが多いと考えられる。というのは，例えば世帯規模の大きい世帯からなる調査区がたまたま数多く抽出された場合，$\hat{P}_l$も大きい値となるが，同時に$\hat{X}_l$も大きい値となる可能性が高いからである。その結果，$\hat{P}_l$と$\hat{X}_l$の比を考えると，$\hat{P}_l$や$\hat{X}_l$そのものに比べ，かなり安定することが予想される。

そこで，仮に人口の大きさP_lが，別の資料により正確に知ることができたとす

注）国勢調査時に得られる層内の全調査区のウエイトの合計

ると，$\hat{X}_l$そのものを推定値とするより，

$$\tilde{X} = P_l \times \frac{\hat{X}_l}{\hat{P}_l} = \hat{X}_l \times \frac{P_l}{\hat{P}_l}$$

を推定値とした方が安定した値を得ることができる。この方法は比推定と呼ばれ，P_lをベンチマーク人口という。この方法を用いると，$\hat{X}_l$の誤差は，$\hat{P}_l$の誤差が同じ方向に向かう場合，かなり縮小する。このように，別途正確な数値が得られるものと高い正の相関を持つものの推定には，比推定は非常に有効である。

線型推定値は標本調査区のウエイトの情報など国勢調査結果に基づく推定値であることから，国勢調査結果以降の人口の移動等は加味していないため，実際の推定値との間に乖離が生じている可能性がある。そのため，労働力調査では線型推定値を求めた後，比推定を用いることでその乖離を補正している。

３　推定方法

(1) ベンチマーク人口の推計方法

ベンチマーク人口には，総務省統計局が毎月公表している「推計人口」を利用している。推計人口とは，国勢調査による人口を基準とし，次の国勢調査結果が得られるまでの間，その後の出生児数・死亡者数（人口動態統計）及び出国者数・入国者数（出入国管理統計）のデータを加減することにより，男女・年齢階級別に毎月１日現在の人口を推計したものである。年齢階級別人口は，国勢調査時点を出発点として，次の式により毎月計算される。

（今月１日現在の当該年齢階級人口）＝（前月１日現在の当該年齢階級人口）－（前月１日現在の当該年齢階級人口のうち前月中に死亡した者）

＋（前月中新たに当該年齢階級に達した人口）－（前月中新たに一つ上の年齢階級に達した人口）

＋（前月中の当該年齢階級入国者数）－（前月中の当該年齢階級出国者数）

※　０～４歳階級については，「前月中新たに当該年齢階級に達した人口」の代わりに「前月中の出生児数」を用いる。

なお，月末１週間を調査期間とする労働力調査においては，翌月１日現在の概算値を用いることとしている。

また，労働力調査では，都道府県別の推計人口についても全国と同様に作成し，(2)で述べる線型推定区分と同じ11地域[注)]に合わせて，都道府県別の推計

注）標本設計での層化区分と同じ 11 地域。第６章の２を参照

人口を合算し，男女，年齢階級（16区分注)），地域（11区分）別のベンチマーク人口を算出している。

(2) 基本集計

ア　結果の推定方法

毎月の就業者数や完全失業者数等の調査結果の算出は，男女，年齢５歳階級(16区分)，地域（11区分）別人口をベンチマーク人口とする比推定によって算出している。

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。

イ　推定の手順

算出の基本的考え方は，以下のとおりである。

① 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，年齢階級，地域別人口の線型推定値を算出する。

② 男女，年齢階級，地域別に，ベンチマーク人口をそれぞれ①で算出した線型推定値で除し，比推定用乗率を算出する。

③ 各標本調査区の属性X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②で算出した比推定用乗率を乗じて，男女，年齢階級，地域別の比推定値$\tilde{X}$を算定する。

④ この比推定値$\tilde{X}$を必要に応じて，男女，年齢階級，地域について合算して就業者数や完全失業者数等の結果数値を得る。

（参考）上記①，②，③をまとめて計算式で表すと，次のとおりである。

$$\tilde{X}=\sum_{l=1}^{L}\frac{1}{m_l}\sum_{i=1}^{m_l}\frac{w_l}{w_{li}}\cdot f_{li}\cdot x_{li}\cdot r_{li}\frac{P}{\sum_{l=1}^{L}\frac{1}{m_l}\sum_{i=1}^{m_l}\frac{w_l}{w_{li}}\cdot f_{li}\cdot P_{li}\cdot r_{li}}$$

$$=\sum_{l=1}^{L}\sum_{i=1}^{m_l}x_{li}r_{li}F_l\frac{P}{\sum_{l=1}^{L}\sum_{i=1}^{m_l}P_{li}\cdot r_{li}\cdot F_l}$$

上記計算式のうち，

$l(=1,\ 2,\ \cdots,\ L)$は11地域，層による区分の番号

$i(=1,\ 2,\ \cdots,\ m_l)$は各区分中の標本調査区の番号

X_{li}は第l区分，第i標本調査区内の属性Xを有する（男女，年齢階級別）調査人口

注）0～14 歳，15～19 歳から 80～84 歳までの５歳階級及び 85 歳以上

w_{li}は第l区分，第i標本調査区のウエイト

f_{li}は第l区分，第i標本調査区の住戸の抽出率の逆数（$= w_{li}$）

w_lは第l区分に含まれる全ての調査区のウエイトの合計

m_lは第l区分の標本調査区数

r_{li}は第l区分，第i標本調査区の修正倍率（調査区の分割など。０＜r_{li}≦２）

F_lは第l区分の線型推定用乗率（$= w_l/m_l$）

Pは（男女，年齢階級，地域別）ベンチマーク人口

P_{li}は第l区分，第i標本調査区内の（男女，年齢階級別）調査人口

$\frac{P}{\sum_{l=1}^{L}\sum_{i=1}^{m_l} P_{li}\cdot r_{li}\cdot F_l}$は比推定用乗率を表す。

(3) 詳細集計

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

月次結果については，毎月の男女，年齢10歳階級（６区分[注1)]），就業状態（就業者，失業者，非労働力人口），従業上の地位（５区分[注2)]），雇用形態（７区分[注3)]）別人口が基本集計結果（月別値）に合うよう比例補正して算出している。

なお，詳細集計では，刑務所・拘置所等のある区域及び自衛隊区域の施設内の居住者を除いている。

４　推定値の誤差

ある時点における就業者や完全失業者等の人数を，統計調査によって推定しようとする場合，得られる結果数値（推定値）は必ずしも真の値に一致するわけではない。このような推定値と真の値との差を誤差という。結果数値をみる際には，誤差の存在を認識して注意する必要がある。

一般的に，誤差は，標本調査であることに起因する標本誤差と，それ以外の実地調査における調査票の誤記入などに起因する非標本誤差に分けられる。

(1) 標本誤差

労働力調査では，国内に居住する全ての者を調査しているのではなく，その一部である標本を調査して全体を推定している。第６章及び本章の第３節まで

注１）　15～24 歳から 55～64 歳までの 10 歳階級及び 65 歳以上

注２）　役員を除く雇用者，役員，自営業主，家族従業者，従業上の地位不詳

注３）　正規の職員・従業員，パート，アルバイト，労働者派遣事業所の派遣社員，契約社員，嘱託，その他

に記載したように，労働力調査の標本抽出方法及び結果の推定方法はやや複雑である。推定に関して簡単な例を挙げると，抽出率が1/1000で10万人を調査したとき，そのうちの５万人が就業者であったとする。このとき，全体の就業者数は，

$$5\text{万人}\times\frac{1000}{1}=5000\text{万人}$$

と推定することができる。ただし，この5000万人という推定値は真の値に等しいとは限らない。なぜならば，調査対象として抽出された10万人は，全ての国民の完全な縮図になるとは限らないからである。標本の抽出をやり直して再び10万人を調査したとしても，就業者数が同数の５万人になるとは限らず，５万1000人あるいは４万9000人かもしれない。つまり，真の値は一つであっても，推定値はそれより大きくなることもあれば，逆に小さくなることも考えられる。このように，標本から推定することによって生じる誤差を標本誤差という。

労働力調査の結果は，このような標本誤差を持つことから，結果数値をみる際には注意が必要である。例えば，先月5000万人であった就業者数が今月は5001万人と推定された場合，仮に後述するような季節性を持たないとしても，この結果から直ちに就業者数が増加したと判断することは早計である。それは，先月と今月の結果数値の両方に標本誤差が含まれているため，真の値は逆に減少している可能性が考えられるからである。

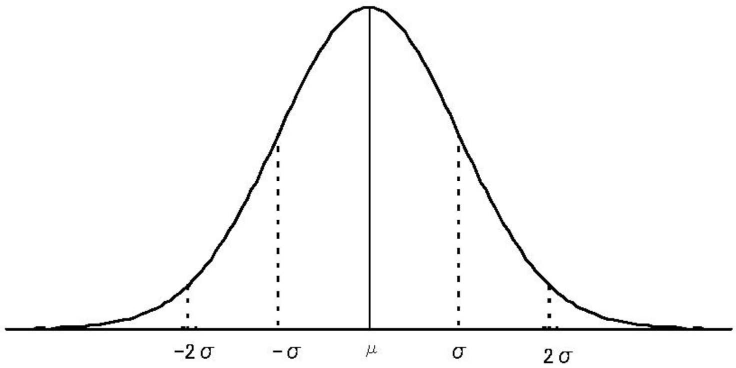
しかしながら，標本誤差の存在は，必ずしも結果数値が信頼性を持たないということにはつながらない。なぜならば，推定値が真の値に近いことは事実であり，真の値から大きく離れることが少ないことを理論的に証明できるためである。上記の例で，もし就業者数の推定値が5000万人から5100万人に増加したとすれば，真の値も増加していることが高い確率で示される。このような判断に根拠を与えるものが標本理論である。

標本理論は，標本調査から得られる推定値が真の値からどの程度離れる可能性があるかを理論的に示すものである。このとき，真の値からの距離を測定する際の物差しになるものが標準誤差である。標本抽出を何度も繰り返して推定を行えば，得られる推定値は真の値を中心とした位置に分布する。この分布の広がりが小さければ精度の良い推定ということができる。分布の広がり具合は，通常，分布の標準偏差で示される。この標準偏差を標本理論では標準誤差という。標準誤差は精度を示す指標であると同時に，誤差を測る尺度となる。また，標準誤差を真の値に対する比率で表したものを標準誤差率という。

標本理論によれば，推定値$\bar{x}$は真の値μの周りにほぼ正規分布をしていると考えられる。また，標準誤差σが分かれば，$\bar{x}$とμの差がσ未満となる確率は，

$|\bar{x}-\mu|<\sigma$が約68%，$|\bar{x}-\mu|<2\sigma$が約95%と示される。したがって，推定値の誤差は，３回中２回はσの範囲に収まっており，2σの範囲を超えることは20回に１回程度しかないことがいえる。

推定値の分布



労働力調査では，月次結果の推定値が5000万人のとき，標準誤差は約30万人と推定されている。そのため，調査結果をより正確に記述するならば，単に推定値を5000万人と表すのではなく，例えば，5000万人±30万人というように表す必要がある。

標本の抽出方法あるいは結果の推定方法が複雑な場合，標準誤差は簡単には算出できない。労働力調査のような人数を推定する場合，15歳以上人口をN，ある属性を持つ人口をX，標本の大きさnの標本によるXの推定値を$\bar{x}$としたとき，$\bar{x}$の標準誤差$\sigma(\bar{x})$は，

$$\sigma(\bar{x}) \fallingdotseq N \times \sqrt{\frac{p(1-p)}{n}} \qquad p = \frac{X}{N}\ (X\text{の15歳以上人口に占める割合})$$

となる。また，標準誤差$\sigma(\bar{x})$を真の値Xで割った標準誤差率は，

$$\frac{\sigma(\bar{x})}{X} \fallingdotseq \frac{N\sqrt{\frac{p(1-p)}{n}}}{X} = \frac{N\sqrt{\frac{p(1-p)}{n}}}{Np} = \sqrt{\frac{1-p}{pn}}$$

と表される。この式から，標準誤差は標本の大きさの平方根に反比例して小さくなり，例えば，標本の大きさが４倍になれば標準誤差は半分になることが分かる。また，全体に占める割合pが小さい場合，標準誤差は小さくなるが，逆に標準誤差率は大きくなることが分かる。

なお，労働力調査では２段抽出法で標本を抽出しているため，実際の標準誤差は上の式で示される値よりも少し大きくなる。

ア　全国結果の推定値の大きさ別標準誤差

年平均値及び月別値の標準誤差率は，標本の交代を行うために設けられた８組の副標本を利用して，下記の算式により推定されたものである（ただし，組別の推定値が独立に正規分布していると仮定している。）。

年平均値用

$$\sqrt{\frac{1}{8(8-1)}\sum_{k=1}^{8}\left(\bar{\bar{X}}_k-\bar{\bar{X}}\right)^2}\Big/\bar{\bar{X}}$$

ここで，$\bar{\bar{X}}_k$及び$\bar{\bar{X}}$は，それぞれ第k副標本及び全標本による属性Xを有する人口の推定値（年平均値）を表す。

月別値用

$$\sqrt{\frac{1}{8(8-1)}\sum_{k=1}^{8}\left(\tilde{X}_k-\tilde{X}\right)^2}\Big/\tilde{X}$$

ここで，$\tilde{X}_k$及び$\tilde{X}$は，それぞれ第k副標本及び全標本による属性Xを有する人口の推定値（月別値）を表す。

属性ごとの標準誤差率を曲線の当てはめにより平均的に評価し，推定値の大きさ別に標準誤差及び標準誤差率（2018年）を算出すると，次表のとおりとなる。

(ｱ) 基本集計

年平均結果の標準誤差

推定値の大きさ（万人）	標準誤差（万人）	標準誤差率（%）
5000	16.2	0.3
2000	9.7	0.5
1000	6.6	0.7
500	4.5	0.9
200	2.7	1.3
100	1.8	1.8
50	1.2	2.5
20	0.7	3.7
10	0.5	5.1

月次結果の標準誤差※

推定値の大きさ（万人）	標準誤差（万人）	標準誤差率（%）
5000	27.9	0.6
2000	17.7	0.9
1000	12.6	1.3
500	8.9	1.8
200	5.7	2.8
100	4.0	4.0
50	2.9	5.7
20	1.8	9.1
10	1.3	13.0

※2018年１月～12月分を単純平均したもの

(ｲ) 詳細集計

年平均結果の標準誤差

推定値の大きさ（万人）	標準誤差（万人）	標準誤差率（%）
5000	18.7	0.4
2000	11.4	0.6
1000	7.8	0.8
500	5.4	1.1
200	3.3	1.6
100	2.2	2.2
50	1.5	3.1
20	0.9	4.7
10	0.6	6.4

四半期平均結果の標準誤差※

推定値の大きさ（万人）	標準誤差（万人）	標準誤差率（%）
5000	37.6	0.8
2000	22.9	1.1
1000	15.7	1.6
500	10.7	2.1
200	6.5	3.3
100	4.5	4.5
50	3.1	6.1
20	1.9	9.3
10	1.3	12.8

※2018年第1四半期から第4四半期までのそれぞれの標準誤差率を単純平均したもの

イ　地域別結果の推定値の大きさ別標準誤差（基本集計）

全国結果と同様の方法で，地域別結果の推定値の大きさ別に標準誤差率（2018年）を算出すると，次表のとおりとなる。

年平均結果の標準誤差率

推定値の大きさ（万人）	標準誤差率（%）										
	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	沖縄
2000			0.4								
1000			0.6			0.5	0.5			0.5	
500	0.6	0.5	0.8	0.5		0.7	0.7	0.6		0.7	
200	1.0	0.8	1.3	0.9	0.7	1.1	1.2	0.9	0.7	1.1	
100	1.4	1.2	1.9	1.2	1.0	1.6	1.6	1.3	1.0	1.5	0.5
50	1.9	1.7	2.6	1.8	1.5	2.3	2.4	1.9	1.5	2.1	0.7
20	3.1	2.8	4.1	2.9	2.5	3.6	3.8	3.0	2.4	3.3	1.3
10	4.4	4.0	5.7	4.2	3.6	5.1	5.4	4.3	3.5	4.7	1.9

四半期平均の標準誤差率※

推定値の大きさ（万人）	標準誤差率（%）										
	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	沖縄
2000			0.6								
1000			0.9			0.7	0.8			0.7	
500	0.8	0.8	1.3	0.8		1.1	1.2	0.9		1.1	
200	1.4	1.4	2.1	1.4	1.1	1.8	2.0	1.4	1.1	1.7	
100	2.1	2.0	3.0	2.0	1.6	2.5	2.8	2.1	1.6	2.5	0.8
50	3.0	3.0	4.3	3.0	2.4	3.7	4.1	3.0	2.4	3.6	1.2
20	5.0	4.9	7.0	5.0	4.1	6.1	6.6	5.0	4.1	5.8	2.2
10	7.4	7.2	10.1	7.3	6.2	8.8	9.5	7.3	6.0	8.3	3.4

※ 2018年第１四半期から第４四半期までのそれぞれの標準誤差率を単純平均したもの

(2) 非標本誤差

非標本誤差とは，誤差の要因のうち標本抽出（偶然性）に起因するものを除いた全ての要因により生じる誤差をいう。非標本誤差は，その要因により幾つかに分類することができる。回答者が質問を誤解したり懸念したりして事実と異なる記入をした場合の誤りや，無回答，調査員の面接の拙さによる誤り，不慣れによる標本の脱落・把握誤り，連絡・指導の不徹底による誤り，調査票の処理及び集計上の誤りなどに分類することができる。このように，非標本誤差は調査のあらゆる段階で発生する可能性がある。

非標本誤差の特徴は，標本誤差とは対照的である。標本誤差の特徴は，①標本の大きさと密接な関係があり，避けられないものであること，②量的な測定ができ，そのコントロールができることなどが挙げられる。一方，非標本誤差は，①標本の大きさと直接関係がなく，原因を究明すれば避けられるものがあること，②量的な測定が難しくそのコントロールができないことなどが特徴として挙げられる。

調査が大規模になって調査関係者の人数が増えるほど，非標本誤差の発生源も増加することになる。調査の各段階での誤りを少なくして非標本誤差を小さく抑えるには，調査関係者の努力と回答者の統計に対する理解に懸かっている。

５　季節調整値

労働力調査の結果をみる場合，本章の４で述べた誤差の存在とともに注意する必要があるのは季節性の存在である。例えば，農業就業者は夏多く冬少ないが，これは毎年ほぼ決まって繰り返されるパターンであって，農林業就業者数のすう勢とは関係ない。したがって各月のデータを時系列的にみて農林業就業者が増加する傾向にあるのか否かを判断しようとする場合，季節的な変動による部分は除去して考えることが必要となる。

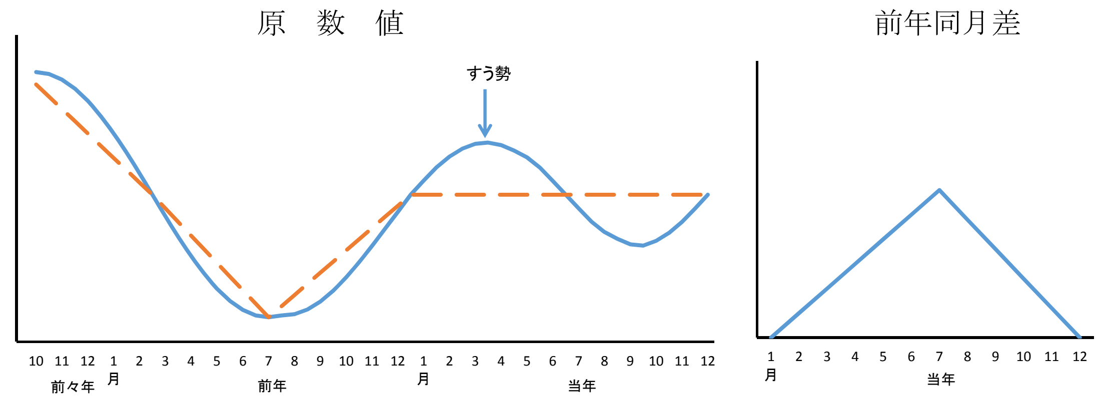
(1) 前年同月差の利用

前年同月差又は前年同月比（以下「前年同月差」という。）を見るというのは，季節性を除いてデータを見る最も手軽な方法である。同じ月同士を比較すれば季節的な変動は自動的に除かれる。このため労働力調査の結果は，前年同月差の形にして動向をみることが多い。労働力調査の標本設計においても，標本のうち半分を前年同月との継続標本とするなど，前年同月との比較の精度が向上するような工夫がなされている。

前年同月比較を行う場合，いくつか注意すべき点がある。例えば，１年前との比較であるため，前年同月差の変化方向は，基本的なすう勢の変化方向と必ずしも一致していない。前年同月差が拡大したといっても，それは１年前の動

きによって引き起こされているかもしれない。例えば，下図のような例の場合，当年の前年同月差の動きは前年におけるすう勢的変化によって引き起こされたものであって，当年における原数値の動きは季節性を除けば安定したものとなっている。また，季節パターンが年とともに変化していく場合は，前年同月との比較では十分でないことに注意する必要がある。

原数値のすう勢と前年同月差



(2) 季節調整の考え方

もう一つの一般的な方法として，季節パターンを除去する手法を原数値に適用して季節調整値を得る方法がある。季節調整値が得られれば，前月との直接比較が可能となる。

まず，原数値の動きが，次の四つの要素から構成されていると仮定する。

○すう勢変動（T：Trend）：経済の成長などに伴い，長期的に上昇・下降を示す変動

○循環変動（C：Cycle）：景気の循環に伴う変動など，ほぼ一定の周期を持つ周期変動で，周期が12か月を超えるもの

○季節変動（S：Seasonal）：12か月を周期とする変動

○不規則変動（I：Irregular）：上の三つ以外の変動で，突発的な出来事や，標本誤差などの変動

そして，原数値の系列O（Original）は

$$O = T \times C \times S \times I$$

という形で，各要素が乗法的に結び付いたものと仮定する。原数値の系列が，絶対水準が高くなるにしたがって季節的な変動の振幅も大きくなっているような場合は，このように仮定するのが一般的で，労働力調査の結果の季節調整もこの「乗法モデル」を用いている。

このとき，次の(3)で述べる方法などにより，年の各月のＳが得られたとする。Ｓはその年の季節パターンを示し，季節指数と呼ばれる。季節調整値は，

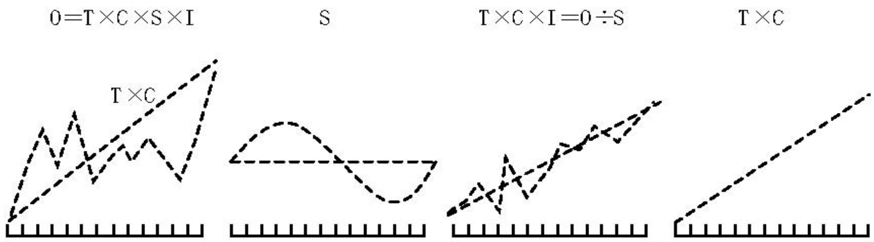
各月において原数値を季節指数で除することによって得られる。式で示せば，

$$O \div S = (T \times C \times S \times I) \div S = T \times C \times I$$

となる。このようにして得られた系列は，ＴＣＩ系列ともいう。

これを図で示せば，次のようになる。

季節調整の概要



さらに，Ｉを除去すれば，すう勢変動と循環変動のみが残り，傾向的な動きを知ることができる。この系列をＴＣ系列ともいう。

(3) 季節調整の方法

季節調整を行う方法は幾つかあるが，一般的な方法としては，アメリカ商務省センサス局が開発したセンサス局法[注)]がある。センサス局法は，移動平均比率法に基づくもので，移動平均を繰り返すことによって季節変動成分を取り出そうというものである。センサス局法による基本的な季節調整の考え方は以下のとおりである。

① 原数値の系列 $O = T \times C \times S \times I \xrightarrow{\text{12か月移動平均}} (T \times C)^*$

② $O \div (T \times C)^* \longrightarrow S \times I^*$

③ $S \times I^* \xrightarrow{\text{移動平均}} S$

④ $O \div S \longrightarrow T \times C \times I$　（ＴＣＩ系列）

⑤ $T \times C \times I \xrightarrow{\text{移動平均}} T \times C$　（ＴＣ系列）

※「*」は暫定値であることを示す。

実際には，このような手続を１回行っただけでは純粋な季節変動成分を安定的に取り出すことはできないので，⑤からまた②に戻るといったように，手続を何度か繰り返すようになっている。また，移動平均も様々な種類のものが使われている。

注）センサス局は，1950 年代から研究を始め，1954 年にセンサス局法Ｉを完成し，複数次にわたる改正を行い，1965 年にセンサス局法Ⅱ（X－11）を公表している。その後 1996 年に X－12－ARIMA が公表された。

さらに，Ｏには，大規模なストライキや天候不順等による例外的な変動が含まれていることがあり，その影響で季節変動成分，すう勢・循環変動成分の推定値がゆがんでしまう場合がある。センサス局法では，各成分を正しく推定するため，例外的な変動の部分を特異項として取り出し，それらを修正，あるいは除去して季節調整を行うようにしている。その場合，特異項と判定するためには，ベースとなるＴＣ系列が安定したものである必要があり，逆に安定したＴＣ系列を得るためには，特異項を修正しておく必要がある。このため，前述の①～⑤のステップを，特異項を検出修正しつつ何度か繰り返すといった方法を採っている。

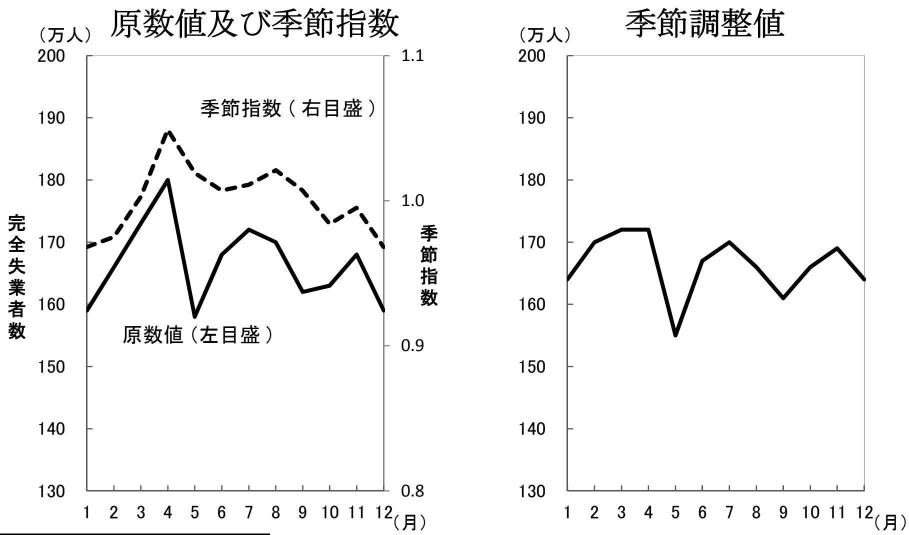
なお，センサス局法による移動平均の仕方，特異項の修正法については，「付録６　センサス局法の概要」で解説している。

(4) 労働力調査結果の季節調整値

労働力調査では，センサス局法（Ｘ－12－ARIMA）[注1]のＸ－11デフォルトを用いて季節調整を行っているが，主要系列については，季節調整の安定性などの向上を図るために，reg-ARIMAモデルを導入している。現在約130系列[注2]について，季節調整済みの月次又は四半期データがあり，うち18系列[注3]がreg-ARIMAモデルの導入系列である。

完全失業者数の季節調整値の例を示すと，下図のとおりである。

季節調整値の例（完全失業者数－2018 年（2019 年１月改定））



注１）特異項の管理限界は，下限 9.8σ，上限 9.9σとし，X-11 デフォルトではこれ以外は標準オプションとしている。reg-ARIMA モデルについては「付録６　センサス局法の概要」を参照

注２）開始年は系列により異なる。なお，最も長い系列は 1953 年1月からの月次データがある。

注３）reg-ARIMA モデルを導入している系列は，（労働力人口，就業者，雇用者，完全失業者，非労働力人口，完全失業率）×（男女計，男，女）の計 18 系列

季節調整値をみる場合，二つの点に注意する必要がある。第一は，過去に公表された季節指数及び季節調整値は，年１回改定されるという点である。センサス局法では，過去の傾向から，将来の季節指数の推定値(推計季節指数)も計算できる。労働力調査では，例えば当年12月までの月次データがそろうと，それを使って翌年１月から12月までの推計季節指数を計算し，その推計季節指数により翌年の各月の季節調整値を公表している。そして，当年12月までのデータがそろうと，当年12月までのデータに基づき翌年の推計季節指数を計算するとともに，過去に遡って各月の季節指数及び季節調整値の再計算[注１)]を行っている。このように，新たなデータが加わることにより，過去に公表した数値が改定されることに注意する必要がある。

第二は，各系列に対して独立に季節調整しているので，例えば男性の系列の季節調整値と女性の系列の季節調整値が，男女計の季節調整値に一致しないこともあり得る点（不加法性）である。また，完全失業率も，独立した一つの系列として季節調整している。このため，完全失業者数の季節調整値を労働力人口の季節調整値で除したものとは一致しない場合があることにも注意する必要がある。

６　時系列回帰モデルによる都道府県別結果の推定

(1) 経緯

労働力調査では，都道府県別に標本設計をしておらず（北海道及び沖縄県を除く。）[注２)]，都道府県ごとの標本規模も小さいことから，当初は都道府県別結果を集計していなかった。しかし，2001年中頃に完全失業率(季節調整値)が調査開始以来初めて５％に達し，その後も高水準で推移し，厳しい雇用情勢が続いた。このような状況の中，雇用・失業情勢の詳細な把握に必要であることから，各方面から都道府県別の結果の公表が要請されるようになった。

このような背景から，都道府県別結果について，2006年５月から時系列回帰モデルによる推定手法を採用し，より安定的な結果が得られるようにした上で，四半期平均結果（モデル推計値）の公表を開始[注３)]した。

(2) 公表系列

モデル推計値は，1997年以降の労働力人口，就業者，完全失業者，非労働

注１）　原則として，当年から29年前までの原数値を用いて再計算を行い，直近10年分の結果について改定している。

注２）　第６章の２を参照

注３）　2002年から，参考として比推定による都道府県別の年平均結果（試算値）を公表していたが，モデル推計値の時系列データが十分に整備されたことに伴い，2007年平均結果をもって廃止した。

力人口，完全失業率の５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。

(3) 推定方法

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

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

※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すう勢変動と，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合わせた変動。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３月から増加し，５月～６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：すう勢変動，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。地震などの自然災害や石油ショックなど一時的な現象の影響によって起こる生産の減少といった動きのことである。

○標本誤差項：労働力調査は，当月調査世帯の半分が前月・前年同月にも調査世帯となるような標本設計となっている。したがって，標本誤差は自己相関を持つ（前月・前年同月の標本誤差が大きければ，当月の標本誤差も大きい）とみなすことが可能である。そこで，これを仮定した時系列モデルにより，標本誤差と考えられる変動パターンと変動幅を前後の時系列データから推計したものである。

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すう勢変動及び循環変動を合わ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）から標本誤差の推計値（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，大阪府及び沖縄県については，比推定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，(3)に示したように，時系列モデルに基づいて推計された標本誤差項を取り除くことで，比推定結果よりも安定的な結果が得られるようにしている。しかし，労働力調査は，都道府県別に結果表章するための標本設計を行っておらず（北海道及び沖縄県を除く。），標本規模も小さいことなどにより，都道府県別結果（モデル推計値）については，全国結果に比べ結果精度が十分に確保できないとみられることから，結果の利用に当たっては注意を要する。

また，時系列回帰モデルは，推定時点以前のデータに加え，推定時点以降のデータをモデル計算に算入することで，より安定的な結果を得ることができる。このため，毎年１～３月期平均結果の公表時に，新たな結果を追加して再計算を行い，前年までの過去５年間の四半期平均及び年平均結果を遡って改定している。