

IV 回帰分析入門

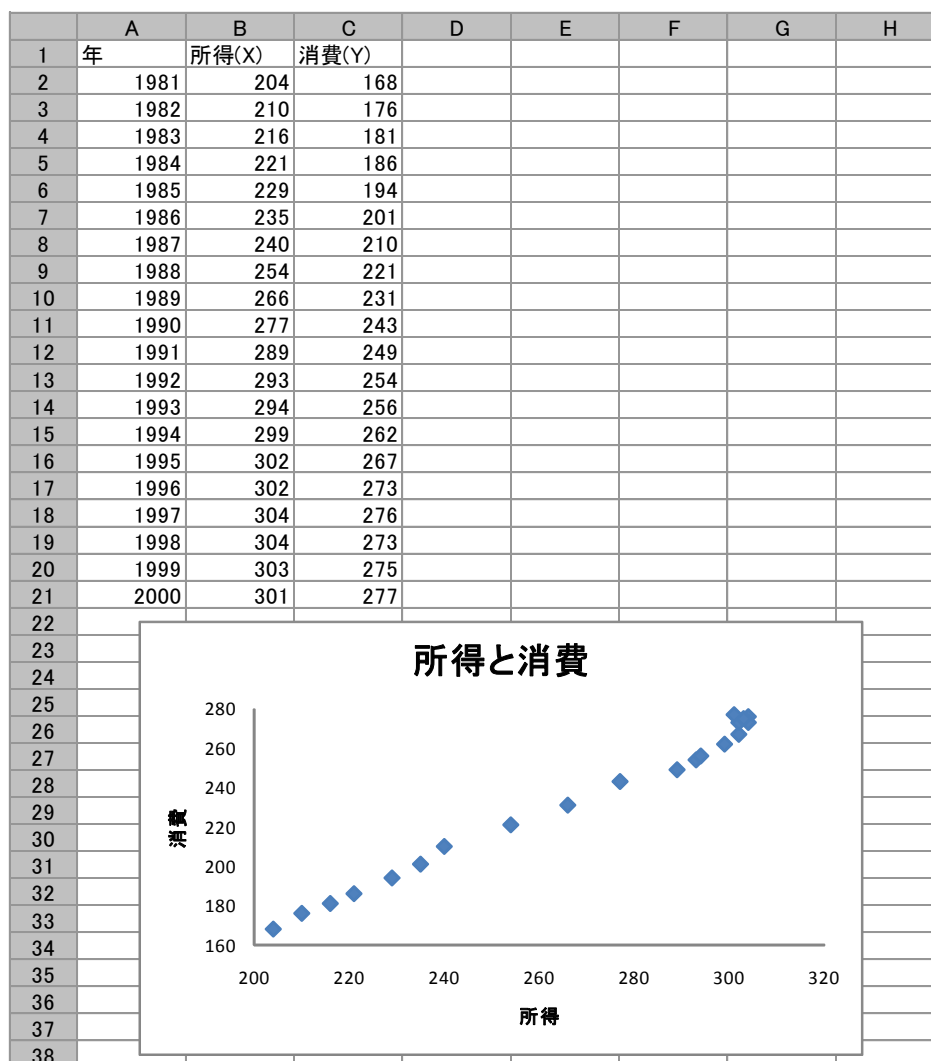
1) 2 変量データの記述

1. 散布図の描画

【課題 20】 下に示したものは、日本の実質家計可処分所得と実質家計最終消費支出のデータ（平成 12 年基準、単位：兆円）¹である。このデータを入力し、散布図を描いてみよう。

📖 散布図は次のような手順で描けばよい。

- ① B2:C21 を範囲指定し、リボン内にグラフのグループにある、散布図 のボタンをクリックする。（挿入タブをクリックすることで表示される。）
- ② 散布図グラフのフォーマット（型式）メニューにおいて、「散布図(マーカーのみ）」（左上）をクリックし、リボンの中のレイアウト 1（左端）をクリックし、タイトルや軸ラベルなどが書き込めるようにする。
- ③ 下の図のようにタイトル、軸ラベルを入力し、目盛線、凡例を非表示にする。
- ④ グラフの作成を終えた後で、それぞれの軸の書式設定をおこない、
縦軸 最小値：160 最大値：280 目盛間隔：20
横軸 最小値：200 最大値：320 目盛間隔：20
とする。



¹ 出典：『平成 21 年版 国民経済計算年報』

2. 相関係数の導出

【課題 21】 所得と消費のデータについて相関係数を求めてみよう。

相関係数の計算式は次のような式である。

$$R = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{\{n \sum X^2 - (\sum X)^2\} \{n \sum Y^2 - (\sum Y)^2\}}}$$

したがって相関係数を導出するためには、 $\sum X$ 、 $\sum Y$ 、 $\sum XY$ 、 $\sum X^2$ 、 $\sum Y^2$ をまず求める必要がある。 $\sum X$ 、 $\sum Y$ は、B 列、C 列の和を求めればよいが、 $\sum XY$ 、 $\sum X^2$ 、 $\sum Y^2$ を求めるためには、交差積(XY)と 2 乗(X^2 、 Y^2)を D 列、E 列、F 列に計算したうえで、その和を求めることになる。

手順は次のようになる。

📖 手順

- ① D 列に X と Y の交差積を求める。D2 セルに**=B2*C2** と入力し、これをコピーすればよい。
- ② E 列に X の 2 乗を、F 列に Y の 2 乗を求める。2 乗を表す演算子は ‘^’ であり、E2 セルに**=B2^2** と入力し、これをコピーする。F 列も同様である。
- ③ B23 セルから F23 セルに各列の合計を求める。これらのセルがそれぞれ $\sum X$ 、 $\sum Y$ 、 $\sum XY$ 、 $\sum X^2$ 、 $\sum Y^2$ である。
- ④ C25 セルに **=(20 * D23-B23 * C23)/SQRT((20 * E23-B23^2)*(20 * F23-C23^2))** と入力する。
この式と計算式とを見比べてみよ。

<作成見本>

	A	B	C	D	E	F
1	年	所得(X)	消費(Y)	XY	X^2	Y^2
2	1981	204	168	34272	41616	28224
3	1982	210	176	36960	44100	30976
4	1983	216	181	39096	46656	32761
5	1984	221	186	41106	48841	34596
6	1985	229	194	44426	52441	37636
7	1986	235	201	47235	55225	40401
8	1987	240	210	50400	57600	44100
9	1988	254	221	56134	64516	48841
10	1989	266	231	61446	70756	53361
11	1990	277	243	67311	76729	59049
12	1991	289	249	71961	83521	62001
13	1992	293	254	74422	85849	64516
14	1993	294	256	75264	86436	65536
15	1994	299	262	78338	89401	68644
16	1995	302	267	80634	91204	71289
17	1996	302	273	82446	91204	74529
18	1997	304	276	83904	92416	76176
19	1998	304	273	82992	92416	74529
20	1999	303	275	83325	91809	75625
21	2000	301	277	83377	90601	76729
22						
23	合計	5343	4673	1275049	1453337	1119519
24						
25		相関係数	0.994674			
26						

2) 単回帰モデル(その 1)

1. 回帰直線の導出

【課題 22】 所得と消費のデータについて $Y = \alpha + \beta X$ という 1 次式をあてはめ、回帰係数 α 、 β の推定値を求めよ。(C26 セルに β の推定値 b 、C27 セルに α の推定値 a を求めよ。)

回帰係数の推定値を求める式は次のようなものである。

$$b = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$
$$a = \frac{\sum X^2 \sum Y - \sum X \sum XY}{n \sum X^2 - (\sum X)^2}$$

この式に相関係数の導出の際に求めた $\sum X$ 、 $\sum Y$ 、 $\sum XY$ 、 $\sum X^2$ 、 $\sum Y^2$ を代入すれば回帰係数の推定値がそれぞれ計算できる。

2. 予測値と残差の計算

【課題 23】 所得と消費のデータについて各年の X のデータに対する予測値 $\hat{Y}$ と残差を求めよ。

📖 手順

- ① G 列に予測値を求める。予測値 $\hat{Y}$ は各 X_i について $a + b X_i$ を計算すればよいので、G2 セルに 1981 年の X (B2 セル)に対応する予測値を求めるなら **=C\$27+C\$26*B2** とし、これをコピーすればよい。ここでは、コピーの際に絶対参照をするので、'\$'がついている。
- ② H 列に残差を求める。残差は Y から予測値 $\hat{Y}$ を引いたものなので、H2 セルに **=C2-G2** とし、これをコピーすればよい。

3. 回帰直線のグラフへの書き入れ

散布図に回帰直線を書き入れる場合、Excelでは各 X に対応する予測値をグラフに書き入れ、それを直線でつなぐという手順をとる。

【課題 24】 所得と消費のデータについて散布図に回帰直線を書き入れよ。

📖 手順

- ① **グラフをアクティブ**にした状態で、リボンの中の「データの選択」ボタン(出ていない場合には上部の「グラフツール」をクリックする)をクリックし、凡例項目(系列)の「追加」ボタンをクリックする。そして、「系列Xの値」を**B2:B21**、「系列Yの値」を**G2:G21**とする。
- ② ①の操作で散布図上に赤色のマーカーが現れたはずである。これを直線で結ぶ。赤色のマーカーのひとつを右クリックして、「データ系列の書式設定」を選ぶ。そこで「マーカーのオプション」をクリックし、マーカーの種類として「なし」をチェックし、「線の色」をクリックし、線の色として「**線 (単色)**」をチェックする。グラフエリアの外をクリックすると回帰直線が引けたことがわかるはずである。

4. 決定係数の導出

決定係数は回帰における当てはまりの尺度であり、全変動のうち回帰モデルによって説明される変動の割合を示すものである。決定係数は0と1の間の値をとるが、決定係数が0.2や0.3などの小さい値であるということは、あまり関係のないXとYの間に因果関係を想定し、分析を行っているということを意味し、モデルの再検討が必要となる。

決定係数は、相関係数との間に、

$$(\text{決定係数 } R^2) = (\text{相関係数})^2$$

という関係がある。

【課題 25】(決定係数 R^2)=(相関係数)² という関係を利用して、C28セルに決定係数を求めよ。

5. 残差の表示

残差 e_i は従属変数の個々の観測データと回帰直線との間のズレの大きさをあらわすものであった。この残差の状態を調べることで、回帰直線のあてはまり具合など、さまざまな情報を入手することができる。残差を出発点としてモデルの設定やデータ間の関係を検討する分析を、残差分析(residual analysis)という。ここでは、残差を求めてそれをグラフに表示してみよう。

グラフを描くには、残差 e_i を縦軸にとり、横軸には、

① i

② X_i

③ $\hat{Y}_i$

などを用いる。

【課題 26】横軸に年をとった残差プロットを描いてみよ。

📖 手順

- ① 残差の部分、H2:H21 を範囲指定し、リボン内にグラフのグループから散布図 のボタン (出ない場合は挿入タブをクリックする。) をクリックする。そして、「散布図(マーカーのみ)」(左上) をクリックする。
- ② 年を横軸にするので、リボンの中の「データの選択」ボタンをクリックし、凡例項目(系列) の「編集」ボタンをクリックする。そして、「系列 X の値」に A2:A21 を指定する。
- ③ そして、最初に描いた散布図と同様に、レイアウトを変更し、グラフタイトル、軸ラベルをそれぞれ記入する。凡例や目盛線も、消去しておこう。

☒ 演習問題 3：消費と所得のデータについて、【課題 26】までおこなったものを、A4 用紙縦 1 枚に納まるようにレイアウトして、印刷してみよ。
次ページに見本を示す。

<作成見本>

日本の実質家計可処分所得と実質家計消費支出

E40-000 徳山 太郎

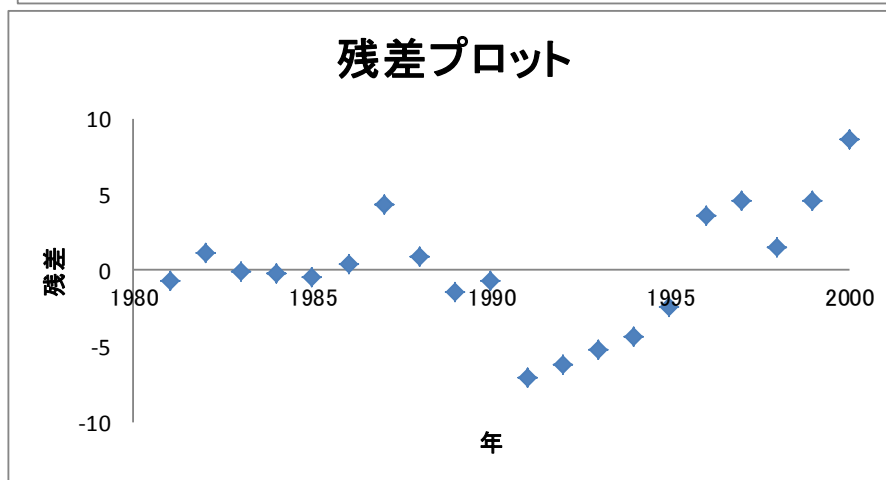
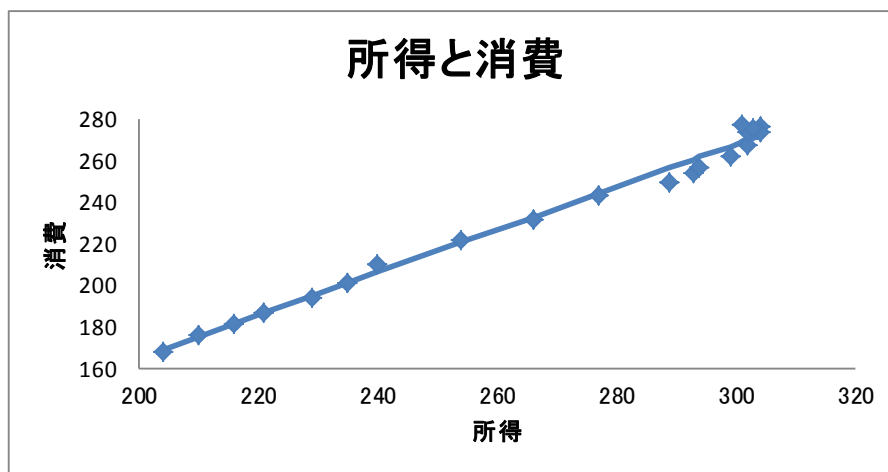
年	所得(X)	消費(Y)	XY	X ²	Y ²	予測値	残差
1981	204	168	34272	41616	28224	168.7907	-0.79075
1982	210	176	36960	44100	30976	174.9531	1.046853
1983	216	181	39096	46656	32761	181.1155	-0.11555
1984	221	186	41106	48841	34596	186.2509	-0.25088
1985	229	194	44426	52441	37636	194.4674	-0.46741
1986	235	201	47235	55225	40401	200.6298	0.370189
1987	240	210	50400	57600	44100	205.7651	4.234857
1988	254	221	56134	64516	48841	220.1441	0.855925
1989	266	231	61446	70756	53361	232.4689	-1.46887
1990	277	243	67311	76729	59049	243.7666	-0.76661
1991	289	249	71961	83521	62001	256.0914	-7.0914
1992	293	254	74422	85849	64516	260.1997	-6.19967
1993	294	256	75264	86436	65536	261.2267	-5.22674
1994	299	262	78338	89401	68644	266.3621	-4.36207
1995	302	267	80634	91204	71289	269.4433	-2.44327
1996	302	273	82446	91204	74529	269.4433	3.556731
1997	304	276	83904	92416	76176	271.4974	4.502598
1998	304	273	82992	92416	74529	271.4974	1.502598
1999	303	275	83325	91809	75625	270.4703	4.529664
2000	301	277	83377	90601	76729	268.4162	8.583797
合計	5343	4673	1275049	1453337	1119519		

相関係数 0.994674

b 1.027067

a -40.7308

決定係数 0.989376



3) 単回帰モデル(その 2)

前章では、相関係数と回帰係数の推定値を交差積和($\sum XY$)、2乗和($\sum X$, $\sum Y$)を求め、それを計算式に代入することによって求めた。

しかしExcelによって相関係数と回帰直線を求めるには、以下に説明するような Excelが備えている関数を用いることもできる。ここでは**reg**の例について、統計関数を用いた方法についても行ってみたいことにする。その際には、データを入力したセルの範囲に名前をつけておくと便利である。まず、実習の準備として、**reg**のSheet1から、年次、所得X、消費Yの部分(A3:C23)をSheet2のA3:C23にコピーし、所得のデータに **_X**、消費のデータに **_Y** という名前を定義しておく。

1. 統計関数による相関係数と回帰直線の導出

Excelが備えている関数を用いた相関係数と回帰直線の導出を行ってみることにする。**reg**のSheet2のA3:C23に、年次、所得X、消費Yのデータが入力されているものとする。

(1) 関数 PEARSON (CORREL), RSQ

相関係数を求めるには、関数 **PEARSON**(引数1, 引数2) を用いる。**PEARSON** は相関係数を最初に導出した Karl Pearson (イギリス;1851-1936) にちなんでつけられた名前である。または、**CORREL** という名前の関数もあるが、どちらも全く同じものである。引数は2個あり、それぞれがデータの範囲(名前でもよい)である。戻り値は r_{xy} である。

```
= PEARSON( B4:B23, C4:C23 )  
= PEARSON( _X, _Y )
```

範囲B4:B23に名前_X、範囲C4:C23に名前_Yを付けてあれば、どちらの式でも結果は同じである。以下の説明では下式の書き方で示す。

関数 **RSQ**は相関係数の2乗(=**決定係数**)を求める関数であるが、引数は **PEARSON** と同じである。したがって、べき乗を求める演算子 $\wedge$ を用いれば **RSQ** は不要となる。

```
= RSQ(_X, _Y)  
= ( PEARSON(_X, _Y) ) ^ 2
```

どちらも全く同じ結果を与える。

(2) 関数 SLOPE とINTERCEPT²

SLOPE は回帰直線の傾き(回帰係数) **b** を、**INTERCEPT** は切片(回帰定数) **a** を求める関数で、どちらも引数は2個あるが、最初の引数が従属変数の範囲で、2個目の引数が独立変数の範囲をとる。引数の順序に注意しなければならない。

```
= SLOPE( _Y, _X )  
= INTERCEPT( _Y, _X )
```

² 回帰直線の傾きと切片を求める関数には、**LINEST** という関数がある。この関数は傾きと切片以外に分析結果に関する多くの情報量を与えてくれる、非常に便利な関数である反面、使用法および結果の解釈の仕方が難しい。**LINEST** 関数の説明はここでは省略する。

(3) 関数 FORECAST と TREND

予測値 $\hat{Y}$ を求める関数には2種類のものが用意されている。関数**FORECAST** は引数を3個とり、FORECAST(X_i , Y範囲, X範囲)として用いる。戻り値は $a + bx_i$ として求められた数値1個である。

=FORECAST(B4, _Y, _X) セルB4の値を x としたときの $a + bx$ が戻り値
残りの X の値に対する予測値は、これをコピーして求めればよい。

あるいは、 n 個の予測値を書き込む範囲を指定しておき、配列数式とすることもできる。たとえば、D4:D23の範囲を指定して、

=FORECAST(_X, _Y, _X)

を入力して、**Ctrl** + **Shift** + **Enter** とする。

関数 **TREND** も同じ予測値を求めるものであるが、引数の数が **FORECAST** より1個多く、計4個となる。一般的な型式は TREND(Y範囲, X範囲, X_i , 1)となる。最後の引数は0か1で、0のときは、原点を通る直線 $Y = bX$ による予測値、1のときはこれまで通りの $Y = a + bX$ による予測値を戻り値として求める。第4引数を省略した場合は、1を指定したものとみなす。

TRENDの第3引数として、 a , b の計算に用いない任意の数値を指定することもできる。たとえば、

=TREND(_Y, _X, 190, 1)

とすれば、 $\hat{Y} = a + b \times 190$ を求めることになる。**reg**の X のデータの中には190という数値はなく、これによって求まる $\hat{Y}$ は未知の X の値に対する**予測値**（外挿値）である。同様のことを、FORECASTを用いてもおこなうことができる。

=FORECAST(190, _Y, _X)

FORECASTでは190が第1引数となる点に注意されたい。また、190という数値を直接指定するのではなく、セル番地で指定することもできる。セルN4に190が書き込んであれば、

=TREND(_Y, _X, N4, 1)

=FORECAST(N4, _Y, _X)

とすればよい。また、N4からN13に $\hat{Y} = a + bx$ として求めたい x の値が連続して書き込まれていれば、

=TREND(_Y, _X, N4:N13, 1)

=FORECAST(N4:N13, _Y, _X)

とすればよい。範囲 N4:N13 に名前を付けて、それを使用してもよい。

2. 分析ツールの利用

Excel には統計分析を行うためのいくつかの分析ツールが付属している。これらのツールを使えば一度に詳細な分析結果を得ることができる。

分析ツールを最初に使用する場合には、アドイン(有効にすること)しなくてはならない。分析ツールのアドインは次のようにおこなう。

- ① 「ファイル」のタブをクリックし、下にある「オプション」のボタンをクリックする。
- ② 「Excel のオプション」のウィンドウが開くので、左側の「アドイン」をクリックする。
- ③ 一番下に表示される「Excel アドイン」の右の**設定**ボタンを押す。
- ④ 「分析ツール」にチェックをつけ、OK ボタンをクリックする。

すると、データタブの中に「データ分析」のボタンが出てくるので、下のほうにある、回帰分析を選べばよい。

アドインを行った後で、再びメニューバーから「ツール」を選ぶと、下のほうに「分析ツール」と表示される。ここで分析ツールを選び、回帰分析を選べばよい。

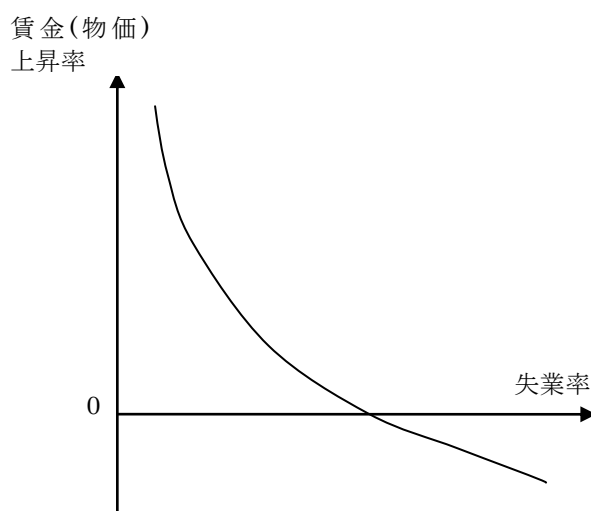
4) 単回帰モデルの他の実例

1. フィリップス曲線

フィリップス(A. W. Phillips) は 1958 年、名目賃金上昇率と失業率の間に右図のような関係があることを、1861 年から 1957 年のイギリスのデータを用いて示した。この曲線は、発見者の名前をとって**フィリップス曲線**と呼ばれる。

その後の研究により、賃金上昇率は物価上昇率との関係が強いことなどから、最近では縦軸に物価上昇率を用いることが多い。

この関係は、インフレ率が高い状況では失業率が低下し、失業率が高い状況ではインフレ率が低下するという、インフレーションと失業とのトレードオフの関係を示している。



2. フィリップス曲線の定式化

フィリップス曲線は、物価上昇率を Y 、失業率を X とするとき、

$$Y = \alpha + \beta \frac{1}{X}$$

という式をあてはめることが考えられる。この式において、 Y と X の関係は直線では表されない(Y と X は**線形関係ではない**という)が、 $X' = \frac{1}{X}$ という変数変換をおこなえば、この式は

$$Y = \alpha + \beta X'$$

という直線(線形式)で表すことができる。

3. フィリップス曲線の実証分析

近年の日本のデータにおいて、フィリップス曲線のような関係が成り立っているかどうかを分析してみよう。

【課題 27】講義用 HP に **Phillips.xlsx** という名前で 1985 年から 2010 年までの物価上昇率と失業率のデータ³が保存されている。このデータを各自のフォルダにダウンロードし、分析をおこなってみよう。

手順は次のようになる。

📖 手順

- ① 失業率を横軸に、物価上昇率を縦軸にとった散布図を描く
- ② D 列に $X' = \frac{1}{X}$ という変換をおこなったデータを作成する。D2 セルに ' $=1/C2$ ' と入力し、それを D 列全体にコピーすればよい。
- ③ 分析ツールを用いて Y を被説明変数、 X' を説明変数とする回帰分析をおこなう。
- ④ 分析ツールの出力結果にある予測値を、散布図に描き入れ、線でつなぐ。

³ 物価上昇率 — 総務省統計局『消費者物価指数』(全国・総合)対前年比
失業率 — 総務省統計局『労働力調査』完全失業率(年平均・男女計)