

第5章

第3章の練習問題で入手した内閣府経済社会総合研究所「県民経済計算」の「1人当たり県民所得」と「総人口」を使用します。

5-1 上記2変数の標準偏差、共分散、相関係数を求めましょう。

【解説】

最新の県民経済計算では2011～2019年度の時系列データを入手することが可能ですが、以下では現時点で最新の2019年度データを使用します。加工しやすいようにデータファイルを作成しておく便利です。Rで使用することを考えると、1行目に変数名、2行目以降にデータとし、A列に都道府県名、B列に「総人口」、C列に「1人当たり県民所得」というのが標準的です。B列とC列は逆でも構いませんが、Excelで散布図を作成する際に少々面倒となります。下が入力のイメージです。ただし、「1人当たり県民所得」は「県民所得」と略記しました。

データイメージ

都道府県	総人口2019	県民所得2019
北海道	5,259,379	2,832
青森	1,252,570	2,628
岩手	1,225,658	2,781
宮城	2,311,736	2,943
秋田	972,164	2,713
山形	1,079,751	2,909
福島	1,851,840	2,942
茨城	2,879,044	3,247
栃木	1,943,179	3,351
群馬	1,948,815	3,288
埼玉	7,342,026	3,038
千葉	6,282,575	3,058
東京	14,007,064	5,757
神奈川	9,224,455	3,199

分散と標準偏差の計算手順については3章の練習問題3-4の解説に詳しく書きましたが、ここではExcelのアドイン「データ分析」を使用する方法を記しておきましょう¹。標準偏差を得るためにはアドイン「データ分析」中の「基本統計量」を使用します。入力範囲を選ぶ際に1行目の変数名を含めて、ダイアログボックスで「先頭行をラベルとして使用」にチェックを入れると便利です。

¹ アドインを設定する方法は4章の練習問題4-4の解説を参照してください。

基本統計量

総人口2019		県民所得2019	
平均	2,692,661.2340	平均	3,016.7991
標準誤差	407,116.0741	標準誤差	71.7159
中央値（メジアン）	1,601,865.0000	中央値（メジアン）	2,972.6710
最頻値（モード）	#N/A	最頻値（モード）	#N/A
標準偏差	2,791,047	標準偏差	491.6597
分散	7,789,944,397,431	分散	241,729.2542
尖度	5.4354	尖度	21.0163
歪度	2.2559	歪度	3.8224
範囲	13,449,694.0000	範囲	3,360.4391
最小	557,370.0000	最小	2,396.3448
最大	14,007,064.0000	最大	5,756.7839
合計	126,555,078.0000	合計	141,789.5557
データの個数	47	データの個数	47

同様にアドイン「データ分析」中の「共分散」、「相関」を用いて共分散と相関係数を計算してみましょう。前と同じように、入力範囲を選ぶ際に1行目の変数名を含め、ダイアログボックスで「先頭行をラベルとして使用」にチェックを入れると便利です。計算結果は以下のようになります。

共分散

	総人口2019	県民所得2019
総人口2019	7,624,200,899,614	
県民所得2019	881,695,487	236,586

相関係数

	総人口2019	県民所得2019
総人口2019	1	
県民所得2019	0.656488102	1

【補足】

「総人口」の標準偏差と「1人当たり県民所得」の標準偏差を掛け合わせた値で共分散を割ると相関係数が得られるはずです。ところが、計算すると、

$$\frac{881695487}{2791047 \times 491.6597} = 0.642520 \dots$$

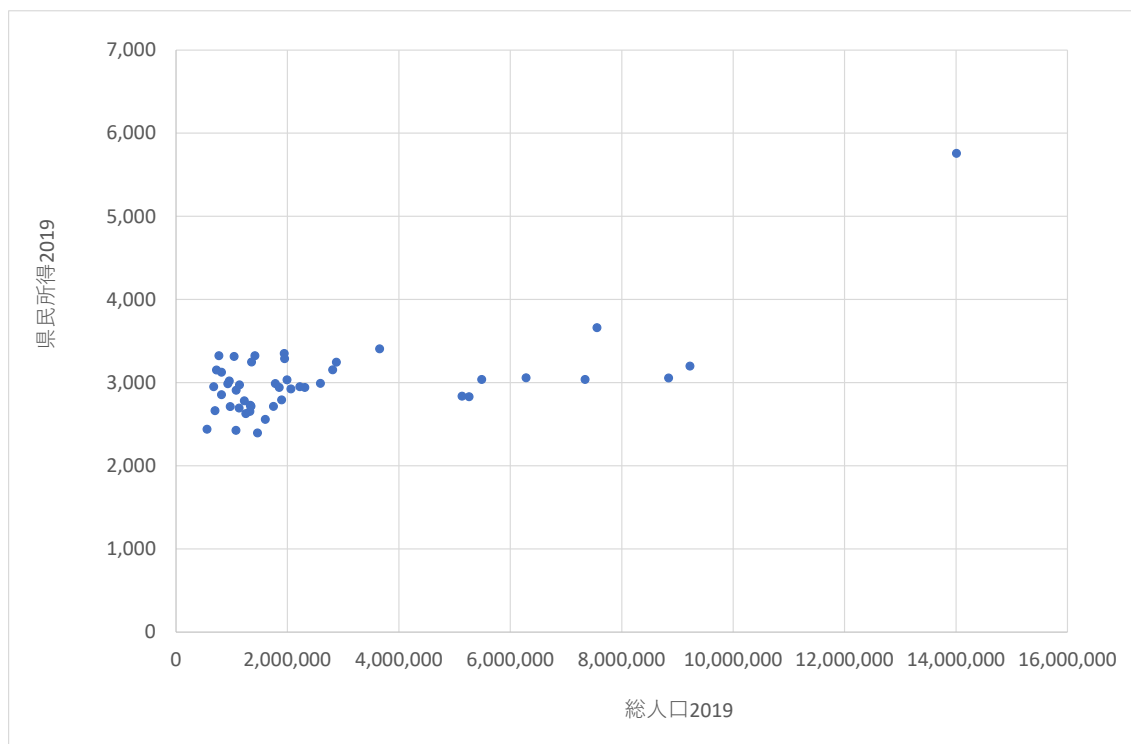
となり、上の相関係数と一致しません。この解説は別ファイル（5章補論）を参照してください。

5-2 「総人口」を横軸、「1人当たり県民所得」を縦軸にして散布図を作成しましょう。

【解説】

軸ラベルを明記した散布図は次のように描けます²。

「総人口」と「1人当たり県民所得」



5-3 上記 5-2 の図を対数目盛に変更した散布図を作成しましょう。ただし、横軸のみ対数目盛、縦軸のみ対数目盛、横軸・縦軸の両方とも対数目盛の 3 種類を作成します。

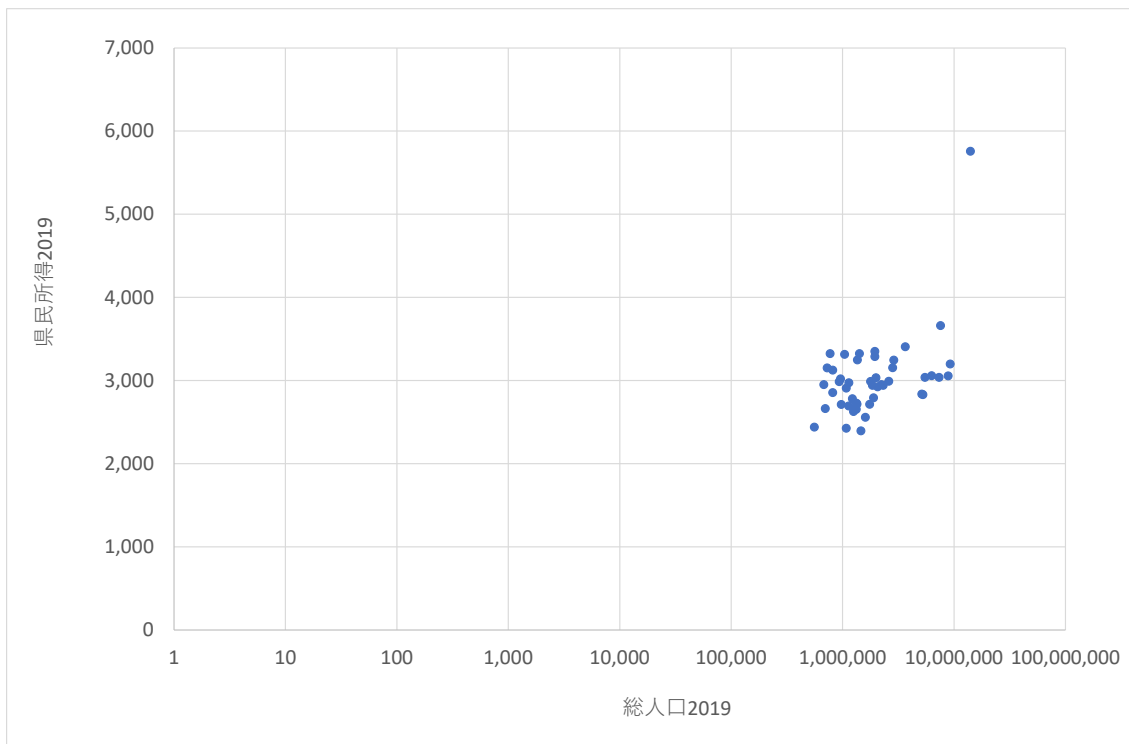
【解説】

まず 5-2 で作成したグラフのコピーを 3 つ作成しておきます。コピーしたいグラフのタブ上で右クリック、「移動またはコピー」を選び、「コピーを作成する」にチェックを入れます。次に、横軸または縦軸の目盛り数字上をクリックします。「軸のオプション」の最右の棒グラフ型の絵をクリックすると、中に「軸のオプション」があり、下の方に「対数目盛」のチェックボックスがあります。チェックボックスにチェックを入れて、「基数」に対数の底となる数字を入力します³。

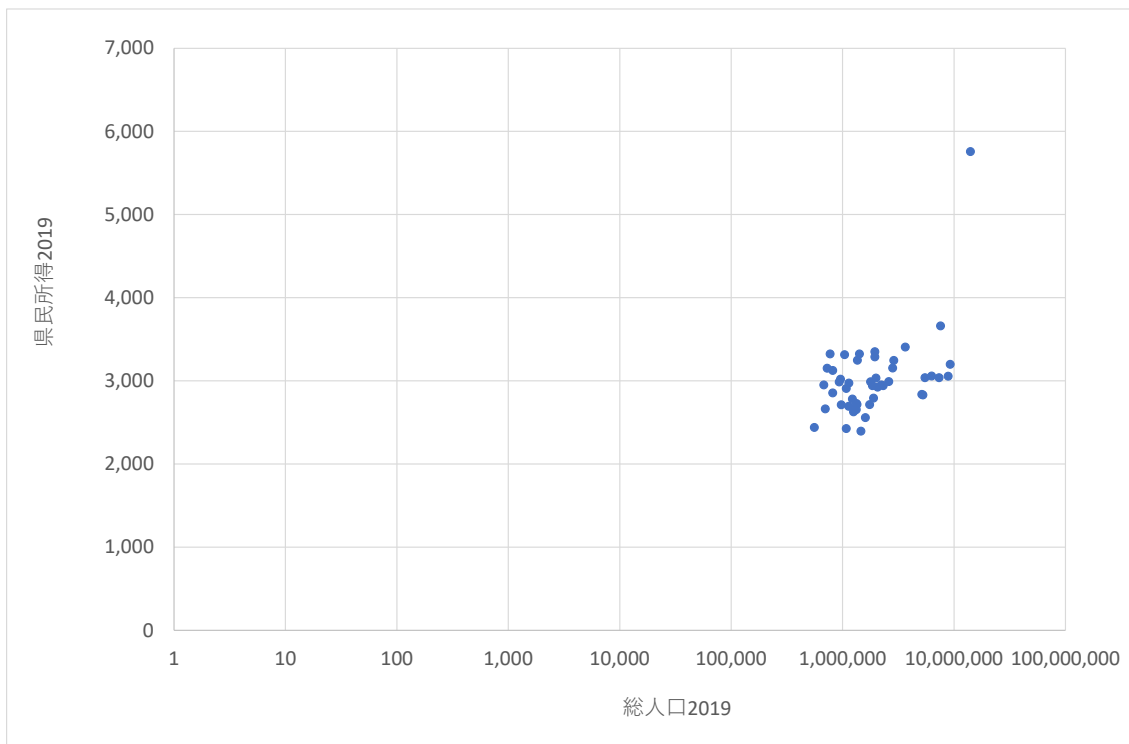
² 散布図の作成方法について 1 章の練習問題 1-3～1-4 で解説しました。

³ 常用対数であれば基数を 10、自然対数であれば 2.71828…（ネイピア数）とします。

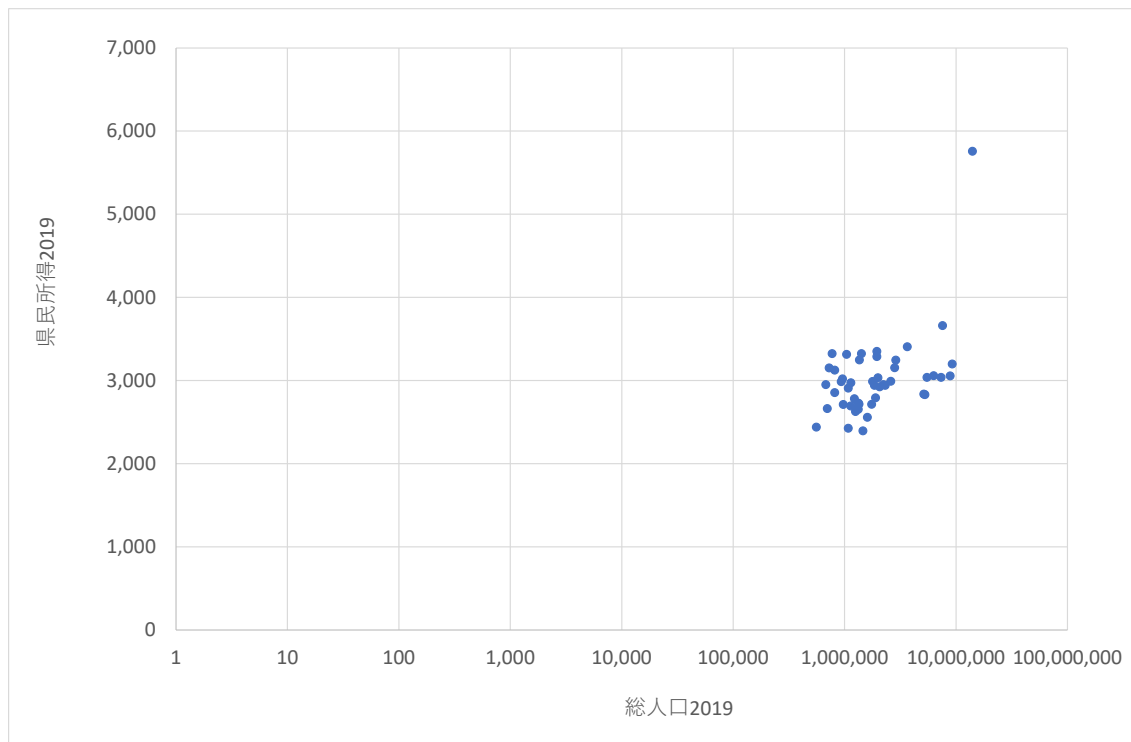
横軸のみ対数目盛



縦軸のみ対数目盛



横軸・縦軸の両方とも対数目盛



5-4 本書のウェブサポートページにある都道府県別スターバックス店舗数のデータを被説明変数とし、「総人口」と「1人当たり県民所得」を説明変数として回帰分析を実行しましょう。また、本文中の「総人口」のみを説明変数とした回帰分析と比較しましょう。ただし、すべての変数は自然対数に変換してください。

【解説】

まず、都道府県別スターバックス店舗数のファイルをダウンロードしてください。2011年5月13日から2022年6月30日までの不定期なデータとなっています。本書では、2021年6月のスターバックス店舗数、2021年1月の住民基本台帳人口を使用しました。ここで手元にある県民経済計算のデータは2019年度のものなので、店舗数として最も時点の近い2020年7月16日のデータを使用することにします。読者の皆さんは、他の年についてもチャレンジしてみてください。

Excelで自然対数変換を行うためには、LN関数を利用します。例えば、B2セルの値を自然対数に変換する場合には、任意のセルに

$$=LN(B2)$$

と入力することになります。

まず、総人口のみを説明変数として回帰分析を行った結果です。係数を見ると、本書150ページ図5-6(2)自然対数の結果とほぼ同じであることが分かります。

都道府県別スターバックス店舗数の回帰分析（総人口）

回帰統計									
重相関 R	0.9578								
重決定 R2	0.9174								
補正 R2	0.9156								
標準誤差	0.2939								
観測数	47								
分散分析表									
	自由度	変動	分散	観測された分散比	有意 F				
回帰	1	43.1698	43.1698	499.9522	0.0000				
残差	45	3.8857	0.0863						
合計	46	47.0555							
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%	
切片	-14.8087	0.7921	-18.6955	0.0000	-16.4041	-13.2133	-16.4041	-13.2133	
ln総人口	1.2239	0.0547	22.3596	0.0000	1.1137	1.3342	1.1137	1.3342	

次に練習問題の課題に答えましょう。「総人口」と「1人当たり県民所得」（表では「県民所得」と表示）の2つを説明変数として回帰分析を行った結果は下のようになります。決定係数が若干上昇し、「1人当たり県民所得」の係数は1%水準で統計的に有意です。説明変数として「1人当たり県民所得」には意味がありそうです。

都道府県別スターバックス店舗数の回帰分析（総人口+1人当たり県民所得）

回帰統計									
重相関 R	0.9643								
重決定 R2	0.9298								
補正 R2	0.9266								
標準誤差	0.2740								
観測数	47								
分散分析表									
	自由度	変動	分散	観測された分散比	有意 F				
回帰	2	43.7516	21.8758	291.3324	0.0000				
残差	44	3.3039	0.0751						
合計	46	47.0555							
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%	
切片	-21.3208	2.4534	-8.6903	0.0000	-26.2653	-16.3762	-26.2653	-16.3762	
ln総人口	1.1412	0.0591	19.3222	0.0000	1.0222	1.2603	1.0222	1.2603	
ln県民所得(1人当たり)	0.9631	0.3460	2.7834	0.0079	0.2658	1.6605	0.2658	1.6605	

念のため、説明変数を「1人当たり県民所得」のみにして回帰分析を行ってみました。下の結果を見ると、決定係数は小さくなり、説明力は大幅に低下しました。「1人当たり県民所得」の係数は統計的に有意ですが、スターバックス店舗数の説明モデルとしては適切でないかもしれません。店舗数の決定要因として、やはり人口要因は無視できないようです。

都道府県別スターバックス店舗数の回帰分析（1人当たり県民所得）

回帰統計									
重相関 R	0.5779								
重決定 R2	0.3340								
補正 R2	0.3192								
標準誤差	0.8345								
観測数	47								
分散分析表									
	自由度	変動	分散	観測された分散比	有意 F				
回帰	1	15.7175	15.7175	22.5697	0.0000				
残差	45	31.3380	0.6964						
合計	46	47.0555							
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%	
切片	-31.7455	7.2887	-4.3554	0.0001	-46.4256	-17.0653	-46.4256	-17.0653	
ln県民所得(1人当たり)	4.3267	0.9107	4.7508	0.0000	2.4924	6.1610	2.4924	6.1610	

ところで、「総人口」と「1人当たり県民所得」の2変数を説明変数としたとき、2つの係数がともに1に近いことに気付いたでしょうか。もし、この2変数の係数が同じだと仮定すると、以下のように式を変形することが可能です。いま、自然対数の記号をln、共通の係数を β と書くことにしましょう。

$$\begin{aligned} & \beta \ln \text{総人口 2019} + \beta \ln \text{県民所得 2019} \\ & = \beta (\ln \text{総人口 2019} + \ln \text{県民所得 2019}) \end{aligned}$$

となります。対数の足し算は、

$$\begin{aligned} & \ln \text{総人口 2019} + \ln \text{県民所得 2019} \\ & = \ln(\text{総人口 2019} \times \text{県民所得 2019}) \end{aligned}$$

のように掛け算の形に変形できますから、カッコの中は県民所得の（1人当たりではなく）総人口分、つまり県民所得の合計（県民所得計）ということになります。

下の回帰分析結果は、説明変数に「県民所得計」（の対数）のみを採用したものです。「県民所得計」の係数は統計的に有意です。また、「総人口」と「1人当たり県民所得」の2つを説明変数として行った回帰分析と比べて、決定係数の値も同じ程度であることが分かります。

都道府県別スターバックス店舗数の回帰分析（県民所得計）

回帰統計									
重相関 R	0.9641								
重決定 R2	0.9294								
補正 R2	0.9279								
標準誤差	0.2716								
観測数	47								
分散分析表									
	自由度	変動	分散	観測された分散比	有意 F				
回帰	1	43.7350	43.7350	592.7110	0.0000				
残差	45	3.3205	0.0738						
合計	46	47.0555							
	係数	標準誤差	t	P-値	下限 95%	上限 95%	下限 95.0%	上限 95.0%	
切片	-22.3628	1.0375	-21.5553	0.0000	-24.4524	-20.2733	-24.4524	-20.2733	
ln県民所得(計)	1.1242	0.0462	24.3457	0.0000	1.0312	1.2172	1.0312	1.2172	