平成２８年度

卒業論文

題目	都道府県別総生産と在院患者延数が失業者数に与えた影響についての定量的検証

文字数
16231

学籍番号　201314056

氏　　名　小林　翔太

本卒業論文は、卒業論文提出要領の基準を満たしており、卒業論文としてふさわしい内容であることを確認しました。

担当教員
（自署）　　㊞

広島経済大学

広島経済大学

要旨

本論文の研究課題は総生産と患者数が失業者数の数にどの程度影響を与えるのかを検証することである。その影響を分析するためにインターネットで公開されている統計データを用いて重回帰分析を行う。

重回帰分析に使用したデータの出典や元のフォーマットを提示した後、それらのデータを重回帰分析が実行可能なフォーマットに変更していくプロセスを書く。

その後重回帰分析を実行し、得られた結果から傾向を分析した。そして失業者数の増減に総生産や患者数は関係していることが判明したが、データの数を増やし、分析の精度を高めようとした際に用意した都道府県別のデータの影響が強く、都道府県ごとの人口や企業数の差異が失業者数の増減の一番の要因となっていた。本来想定した内容とは異なる結果となったため失業者数ではなく失業率を用い改めて分析を行い、考察を行った。

広島経済大学

目次

広島経済大学

第1章　はじめに

1.1.　研究のテーマ

本論文のテーマは、都道府県別の総生産と患者数が失業者数に与える影響について定量的検証を行い、その影響の大小を考察し、どう影響しているのかを分析することである。失業者数の増減に景気の良さや、けが人、病人の多さが関係しているのかを検証するためこのテーマを選択した。景気の良し悪しを把握するために、都道府県別総生産のデータを使用し、けが人、病人の数を把握するために、患者数（病院を訪れた人の数）のデータを使用する。データの数を増やすことで分析の結果をより正確なものにするためにそれぞれ2001年から2012年までの12年間と47都道府県別の合計の n=564 のデータを使用する。

1.2.　研究手法

本論文ではそれぞれ都道府県別の失業者数、総生産、患者数の3つのデータの関係性を分析するために、インターネットで公開されている統計データを使用し、重回帰分析を用いて統計分析を行う。重回帰分析はR言語という統計解析向けの開発環境を用いて行う。

1.3.　本論文の構成

まず第2章で分析方法として重回帰分析を選んだ理由と、重回帰分析を行うために使用するR言語についての説明をする。次に第3章では使用するそれぞれのデータの出典や元のフォーマットなどを提示し、重回帰分析を実行するために行ったそれぞれのデータのフォーマットの統一や規準化などのプロセスを書く。続く第4章では、重回帰分析をした結果や得られた結果から散布図を表示するなどして、データから傾向を分析し考察やまとめなどを書く。第5章では第4章で得られた結果から判明した問題点を踏まえ再び重回帰分析を実行し、第4章と同じように得られた結果から散布図を表示するなどして、データの傾向を分析し考察などを書く。最後に、論文全体を通してのまとめや統計分析を実行してみて感じた問題点から今後の課題などを書く。

第2章　統計分析とは

2.1.　統計分析とは

大量のデータを分析し、その結果からデータの傾向やパターンを見出す手法である。統計分析により得た結果は調査、研究などで重大な決定を下す際に非常に参考にできる科学的根拠となる。分析する際に用いるデータの数が増えるほど、より正確な分析が可能となる（前川、2014)。統計分析には様々な手法があるが、そのなかでも今回は重回帰分析を用いる。

2.2.　重回帰分析とは

目的となるデータを他の複数のデータによって予測する関係性の式を作る手法のことである。この関係性の式のことを重回帰式といい、重回帰式は「$y = ax_1 + bx_1 + c$」のような形で表現する。

2.2.1.　重回帰式とは

重回帰式は目的変数、説明変数、係数、定数項からなる式である。

まず重回帰式の yの部分のことを目的変数と言い､他の変数により説明される変数である。被説明変数、従属変数と呼ぶこともある。

次に重回帰式のx_1とx_2の部分のことを説明変数と言い､目的変数を説明する変数である。独立変数と呼ぶこともある。

重回帰式の a と bの部分を係数と言い、説明変数が目的変数に対してどの程度の影響があるかを表す数値である。この係数が大きいほど、説明変数が目的変数に与える影響が大きくなる。

重回帰式の cの部分のことを定数項と言い、説明変数の変動に影響されない値を示している。決定項、y 切片と呼ばれることもある。

重回帰式を読み解くにあたって、影響の大小、どの程度説明できているかなどを測る要素として、決定係数・t 値・p 値がある。

説明変数が目的変数をどの程度説明できているのかを表す値のことを決定係数という。重回帰式の精度を表しており、値が 100%に近いほど精度が高く、0%に近いほど精度は低くなる。この決定係数は説明変数の数が増えるとそれに比例して値が増えるという性質を持っており、より正しい値を知るために自由度を用いて調整した自由度調整済みの決定係数を用いることが多い。寄与率、R^2と呼ばれることもある。

説明変数が目的変数に与える影響の大きさを表している値のことを t 値という。t 値の絶対値が大きいほど与える影響は強く、t 値の絶対値が 2 以下の場合、統計的には目的変数に対し、その説明変数は影響しないと判断される。t 値を構成する要素は期待値、分散、サンプルサイズによって求められる。それぞれ期待値の差は大きい方がよく、分散は小さい方がよく、そしてサンプルサイズは大きい方がよい。

説明変数の係数の有意確率を表している値のことを p 値という。p 値が 5%以下の場合、その説明変数と目的変数の間には関係性があると判断される。

この t 値と p 値はあとで説明する t-検定に使用する。

2.2.2. どういうときに用いる手法か

回帰分析は目的変数と、説明変数の間の関係を分析するものである（前川、2014）。回帰分析のうち、目的変数を 1 つの説明変数で予測・説明する際に用いる統計手法のことを単回帰分析と呼び、複数の説明変数から予測・説明する際に用いる統計手法のことを重回帰分析と呼ぶ。また、回帰分析は顕在変数という実際に測定して、データを得ている変数を用いて行うものである。

したがって重回帰分析とは複数の説明変数があり、それぞれの変数が顕在変数である場合に利用する手法である。

2.3. 最小二乗法とは

回帰分析は手法として最小二乗法を使用している。

予測値に基づく直線上の値と、実際の値との差の二乗和が最も小さくなるような直線を求める方法のことを最小二乗法といい､最小二乗法で求めた直線のことを回帰直線という。

回帰直線は 2 組のデータの、中心的な分布傾向を表す直線のことである。回帰直線は散布図に引いて将来の値の予測に利用するものである。以上は基本的な事であり、最小二乗法は直線だけに用いるものではなく、曲線をデータに当てはめることもある。

2.4. t-検定とは

2 つのグループの平均の差があるとき、その誤差が偶然、つまりよくあるデータのばらつきなのか､それとも本質的な違いがあることなのかを調べ､判断するための手法である。ここで本質的な違いがあると判断された場合、傾向の分析に成功したと言うことができ、今後その数値の増減についてある程度予測、場合によっては干渉することが可能となる。逆に偶然の範囲だった場合、傾向は分析できず今後の数値の増減を予測することは困難である。t-検定には先程説明した t 値と p 値を使用する。

t 値が大きいほどその誤差は偶然ではなく、何か差ができるだけの理由がある、つまり有意差があるということを 2.2.1.で説明したが、その t 値の大小を判断するのに用いるのが t-検定である。

t 値とサンプルサイズを計算することで p 値を求めることができる。サンプルサイズが大きいほど p 値は小さくなりやすい。そして t 値が大きいほど p 値は小さくなる。t 値には明確な大きさの基準がないが、p 値には基準がある。その基準とは p 値は 0.05 を下回ると小さいと判断する、というものである。

したがって、p 値が 0.05 を下回っているということは、逆説的に t 値は有意差があると説明するのに十分な大きさであるということができる。

2.5. R 言語とは

R 言語は統計解析に特化したプログラム言語である（金、2007)。C 言語や Java などの汎用開発言語ではなく、統計解析向けの言語である。つまり C 言語などのような汎用性はなくなっているが、統計解析に特化しているため、多量のデータを効率的に操作したり、データの配列や行列の演算をしたり、それらの結果を可視化することが比較的容易に実行可能な言語となっている。C 言語などの汎用開発言語に比べ、容易に統計解析が可能となっているが、統計解析用の言語となっているため、統計学の基礎知識がある程度必要である。

ここまでは正確に言うと R 言語の説明ではなく、統計解析向けの言語の説明を行ってきた。そのためここからは R 言語の特徴を説明していく。

まず R 言語はオープンソースの統計解析向けのプログラム言語である。統計解析用の言語はいくつかあるが、R 言語はオープンソース、つまり無料で公開されている言語である。そのため統計解析を行ってみたいと思った際、他の統計解析用言語と比べ無料であるため敷居が低く、総計解析に少し興味がある程度の人でも、気軽に統計解析に触れることが可能な言語である。入手が容易なだけでなく、オープンソースであるため現在も機能の拡張が行われており、使いやすさや出来ることが増えていくのも特徴の一つである。なかでも R 言語の最大の強みは標準でサポートされている多くの一般的な統計手法とその手法が簡単なコマンド 1 つで実行可能な事である。統計学の基礎やプログラムの基礎などはあったほうがいいが、それらの専門知識に乏しい者でも比較的容易に統計分析を行うことが可能となっている。

本論文ではインターネットで公開されている統計データを使用して傾向を把握する使い方しか使用しないが、R 言語はビッグデータのような莫大な量の数値から傾向を把握するといった使い方の他にも、パッケージと呼ばれるユーザーが開発した拡張機能により、出来ることは多岐にわたる。例えば莫大な量の文字列から頻繁に出現する単語を抽出して流行を把握するなどといった使い方も存在する。

第3章　使用したデータとデータの加工

3.1. データソース

次の表 3.1.a のデータを用いた。失業者数（y）については他のデータの期間に合わせて 2001 年～2012 年までを用いた。

表 3.1.a　データの出典

変数	出典	期間
失業者数　(y)	総務省統計局　労働力調査[1]	1997 年～2014 年
総生産　(x_1)	内閣府　県民経済計算[2]	2001 年～2012 年
患者数　(x_2)	e－Stat 政府統計の総合窓口　患者調査[3]	2001 年～2012 年

注 1：http://www.stat.go.jp/index.htm
注 2：http://www.cao.go.jp/
注 3：https://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do

3.2. データの原フォーマット

3.2.1. 失業者数の原フォーマット

都道府県別の完全失業者数のデータの原フォーマットは、縦方向に 1997 年から 2014 年までの期間が並んでおり、横方向に北海道から沖縄県までの都道府県が県コードの順番に並んでいる。データの単位は千人である。

表 3.2.1.a　都道府県別失業者数の元データ

			第4表　都道府県別完全			
(千人)						
		1	2	3	46	47
		北海道	青森県	岩手県	鹿児島県	沖縄県
平成9年	1997	107	29	18	26	36
10	1998	137	36	23	34	47
11	1999	143	40	27	32	51
12	2000	159	40	28	31	50
13	2001	168	43	33	35	53
24	2012	140	36	26	37	46
25	2013	122	33	23	35	39
26	2014	109	29	19	32	37

3.2.2.　県別総生産の原フォーマット

都道府県別総生産のデータの元の書式は、縦方向に都道府県が県コードの順番に並んでおり、横方向には 2001 年から 2012 年までの期間が並んでいる。それぞれの数値には読みやすいよう 3 桁区切りのカンマが付いている。データの単位は 100 万円である。

表 3.2.2.a　都道府県別総生産の元データ

						総括表
					(単位:100万円)	
都道府県			平成13年	平成14年	平成24年	
			2001	2002	2012	
1	北海道		20,281,202	19,915,520	18,124,116	1
2	青森県		4,667,370	4,573,074	4,472,202	2
47	沖縄県		3,699,676	3,684,206	3,806,582	47
	全県計		521,397,405	516,868,391	500,158,230	

3.2.3.　患者数の原フォーマット

都道府県別患者数のデータの元の書式は、縦方向に都道府県が県コードの順番に並んでおり、横方向には患者の国内総数と公的医療機関や医療法人、個人の医者を訪れた患者などの詳細が並んでいる。患者数のデータはそれぞれの年毎に別のデータを用意しており、用意したデータは 2001 年から 2012 年までの 12 のデータである。

表 3.2.3.a　都道府県別患者数の元データ

平成15年	病院報告	年間					
平成14年	病院報告	年間					
平成13年	病院報告	年間					
閲覧　第　9表　在院患者延数，開設者(中分類)・							
	総　数	国		会社	個人	医育機関(再掲)	
		厚生労働省	その他				
総　数							
全　国	5.12E+08	23197156	16231874	3983735	29269132	28005633	
北 海 道	33223023	1007111	981335	206561	1773079	794533	
鹿 児 島	11805549	508812	275668	・	503771	233400	
沖　縄	6530132	492322	195187	・	160894	189613	

3.3. データのフォーマットの変更

使用するデータを同じフォーマットに統一し、統合することで回帰分析をすることが可能になるため､まずは使用する 3 つのデータを共通のフォーマットに変更する必要がある。

完成形となるデータのフォーマット

使用する 3 つのデータを変更する際の完成形となるフォーマットは、次で説明するように必要なところだけ抜き出し､次に 2001 年の都道府県の下に 2002 年の都道府県といったように各年の 47 都道府県を 2012 年のデータまで順番に 1 列で並べたものである。また、都道府県の並びは都道府県コードの順番で並んでいる。都道府県コードは図 3.3.a の通りである。そして完成形のフォーマットは図 3.3.b のようになる。

1 北海道	9 栃木県	17 石川県	25 滋賀県	33 岡山県	41 佐賀県
2 青森県	10 群馬県	18 福井県	26 京都府	34 広島県	42 長崎県
3 岩手県	11 埼玉県	19 山梨県	27 大阪府	35 山口県	43 熊本県
4 宮城県	12 千葉県	20 長野県	28 兵庫県	36 徳島県	44 大分県
5 秋田県	13 東京都	21 岐阜県	29 奈良県	37 香川県	45 宮崎県
6 山形県	14 神奈川県	22 静岡県	30 和歌山県	38 愛媛県	46 鹿児島県
7 福島県	15 新潟県	23 愛知県	31 鳥取県	39 高知県	47 沖縄県
8 茨城県	16 富山県	24 三重県	32 島根県	40 福岡県	

図 3.3.a 都道府県コード

			失業者数	総生産	患者数
2001	01	北海道	107	20281202	33223023
2001	02	青森県	29	4667370	6053307
2001	03	岩手県	18	4757174	6500192
2001	04	宮城県	38	8822295	7802030
2001	05	秋田県	20	3968079	5625029
2001	06	山形県	14	4044314	4550174
2001	07	福島県	29	7969590	9261999
2012	45	宮崎県	25	3531012	5815458
2012	46	鹿児島県	37	5347166	10663856
2012	47	沖縄県	46	3806582	6097865

図 3.3.b 完成形のフォーマット

完成系のフォーマットを作成するには図 3.3.c のように各データソースを加工し統合する。

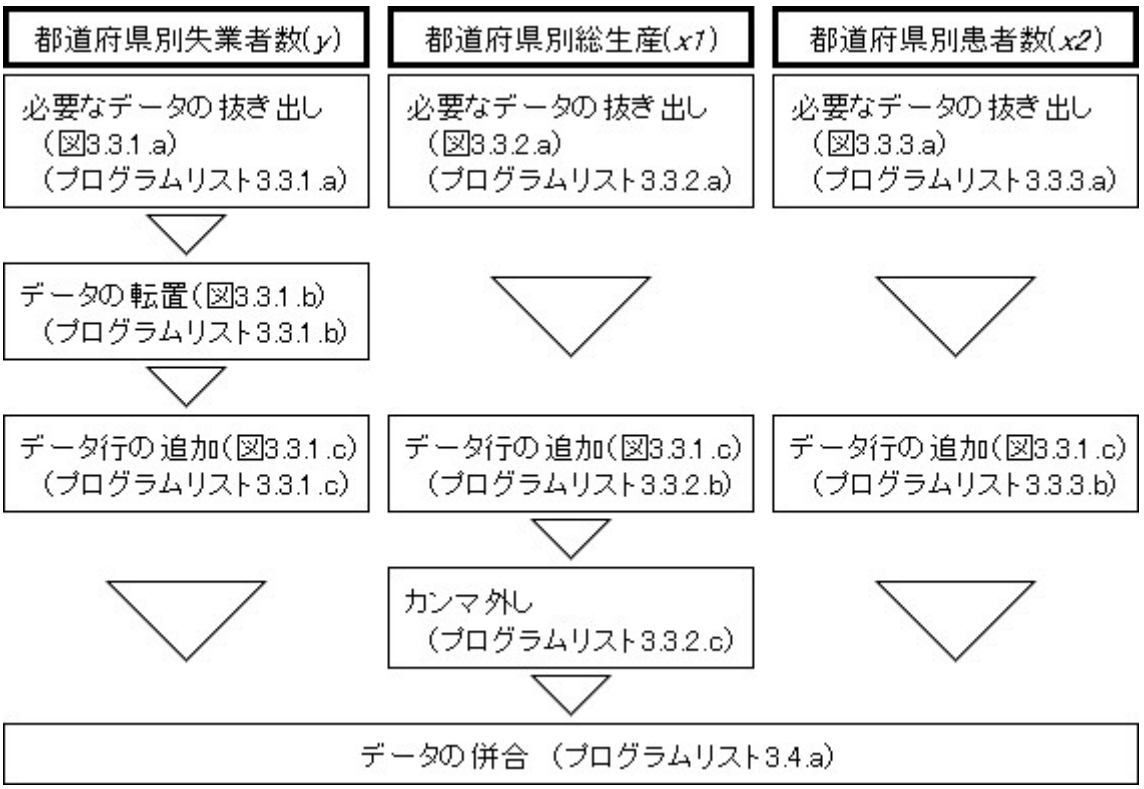

図 3.3.c　作業の流れのフローチャート

最初に目的変数となる都道府県別失業者数と総生産、患者数で共通する範囲の期間、全県合計などの余分な範囲を取り除いた 47 都道府県の範囲を抜き出す。そしてその抜き出した範囲を統一したフォーマットに変更していく。まず目的変数となる都道府県別失業者数のデータのフォーマットから変更を行う。失業者数のデータは他の 2 つの変数とは行と列の並びが異なっているため、行と列を入れ替え他の変数と同じ形式にする。この時点で全ての変数のフォーマットは統一されたが、重回帰分析を行うには目的変数 y 説明変数x_1とx_2の 3 つの変数をそれぞれ 1 列にして合計 3 列にする必要があるため、次はそれぞれの変数を 1 列にしていく。前述したように 2001 年の都道府県の下に 2002 年の都道府県といったように各年の 47 都道府県を 2012 年のデータまで順番に 1 列で並べたものである。その形にするために、都道府県ごとのまとまりで年ごとに分割してからその分割したデータを 2001 年から順番に並べなおすことでそれぞれの変数を 1 列のデータに変更する。その 1 列にした 3 つのデータを並べて、564 行 3 列の 1 つのデータにする。それが完成形のフォーマットとなる。実際に行った作業の詳細はこのあと書いていく。

3.3.1.　失業者数のデータフォーマットの変更

まず都道府県別失業者数のデータのうち余分な範囲を省き、必要な範囲だけ抜き出す。失業者数のデータは 1997 年から 2014 年までのデータが存在するが、総生産のデータは 2001 年から 2012 年までのデータしか存在しない。そのため失業者数のデータも総生産のデータに合わせて 2001 年から 2012 年の範囲を抜き出す。図 3.3.1.a の太枠を参照。

			第4表　都道府県別完全			
(千人)						
		1	2	3	46	47
		北海道	青森県	岩手県	鹿児島県	沖縄県
平成9年	1997	107	29	18	26	36
10	1998	137	36	23	34	47
11	1999	143	40	27	32	51
12	2000	159	40	28	31	50
13	2001	168	43	33	35	53
24	2012	140	36	26	37	46
25	2013	122	33	23	35	39
26	2014	109	29	19	32	37

図 3.3.1.a　都道府県別失業者数から必要な部分の抜出し

R 言語によるプログラム内容は以下の通りである。

```
> x  <-  read.csv ("lt04y.csv")      #データの読み込み
> x  <-  x [ 12：23 , ]              #2001～2012 年までのデータ行を選んで抜き出す
> x  <-  x [ , 3：49 ]               #都道府県のデータ列を選んで抜き出す
```

プログラムリスト 3.3.1.a

次に他の変数と同じく縦軸に都道府県、横軸に年数となるように行と列を入れ替える。そうすることで他の変数と同じく 47 行 12 列となるようにフォーマットを合わせる。

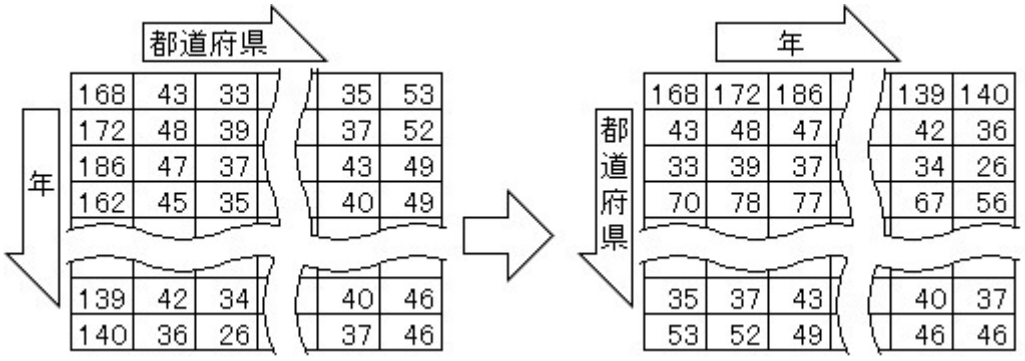

図 3.3.1.b　行と列の入れ替え

R プログラムは以下の通りである。

```
> x   <-   t (x)                          #図 3.3.1.b の左側から右側にする（転置）
> x   <-   data.frame (x)
> dim (x)
[1] 47 12
```

プログラムリスト 3.3.1.b

次に年ごとに分割し、それを図 3.3.1.c の要領で縦に並べる。詳しく説明すると失業者数のデータに限らず、使用する 3 つのデータは最初から県コードの順番に都道府県が並んでおり、先程の作業で行と列を入れ替えた状態のデータから、1 列ごとに抜き出して一度 12 個のデータに分ける。その 12 個のデータを縦 1 列になるように改めて結合する。そうすることで 47 行 12 列のデータを 564 行 1 列のデータに変更する。

図 3.3.1.c　データの並べ方

R プログラムは以下の通りである。

```
> x   <-   t (x)
> x   <-   data.frame (x)
> for (i in 1:12) {assign (paste ("x" , i , sep = "") , x [i , ] ) } #データを 1 列ずつに分割する
> #データを 1 列に結合する
> x_df   <-   data.frame (x1 , x2 , x3 , x4 , x5 , x6 , x7 , x8 , x9 , x10 , x11 , x12)
> x_df   <-   t (x_df)
> dim (x_df)
[1] 564     1
> write.csv (x_df , "situgyousya.csv" , row.names = FALSE)     #データの書き出し
```

プログラムリスト 3.3.1.c

失業者数のデータを読む単位は千人だが、後で規準化するため桁を正す必要はない。そのため今回は桁を正す作業は行わずそのままにしておく。データを 564 行 1 列にしたため失業者数のデータは他の変数のデータと結合する準備ができたので、失業者数のデータのフォーマットの変更を終了し、続いて総生産のデータのフォーマットを変更していく。

なお規準化については後で詳しく書くのでここでは説明しない。

3.3.2.　県別総生産のデータフォーマットの変更

まず都道県別総生産のデータも失業者数のデータと同じく必要となる 2001 年から 2012 年、どの都道府県のデータかの説明となっている範囲や全県合計などの不要な範囲を除いた必要な範囲だけを抜き出す。図 3.3.2.a の太枠を参照。

						総括表
					（単位:100万円）	
都道府県			平成13年	平成14年	平成24年	
			2001	2002	2012	
1	北海道		20,281,202	19,915,520	18,124,116	1
2	青森県		4,667,370	4,573,074	4,472,202	2
47	沖縄県		3,699,676	3,684,206	3,806,582	47
	全県計		521,397,405	516,868,391	500,158,230	

図 3.3.2.a　都道府県別総生産から必要な部分の抜出し

R プログラムは以下の通りである。

```
y   <-   read.csv ("soukatu1.csv") #データの読み込み
y   <-   y [5:51 , ]                #都道府県のデータ列を選んで抜き出す
y   <-   y [ , 4:15]                #2001～2012 年までのデータ行を選んで抜き出す
y   <-   data.frame (y)
dim (y)
[1] 47 12
```

プログラムリスト 3.3.2.a

次に図 3.3.1.c の要領で縦に並べる。R プログラムは以下の通りである。

```
> y   <-   t (y)
> for (i in 1:12) {assign (paste ("y" , i , sep = "") , y[i , ] ) } #データを 1 列ずつに分割する
> #データを 1 列に結合する
> y_df   <-   data.frame (y1 , y2 , y3 , y4 , y5 , y6 , y7 , y8 , y9 , y10 , y11 , y12)
> y_df   <-   t (y_df)
> dim (y_df)
[1] 564     1
```

プログラムリスト 3.3.2.b

総生産のデータを読む際の単位は 100 万円だが、失業者数と同じくあとで規準化するため桁はそのままにしておく。最後に元の数値には 3 桁ごとにカンマが入っているが、そのままだと回帰分析をしようとした際エラーが起きるため、カンマを外す。

R プログラムは以下の通りである。

```
> y_df   <-   t (y_df)
> y_df   <-   gsub (",", "" , y_df)                    #カンマを取る
> y_df   <-   t (y_df)
> mode (y_df)
[1] "character"
> y_df   <-   as.integer (y_df)                        #文字列を数値に戻す
> mode (y_df)
[1] "numeric"
> write.csv (y_df , "gdp.csv" , row.names = FALSE)     #データの書き出し
```

プログラムリスト 3.3.2.c

県別総生産のデータからカンマを外そうとした場合、数列をいったん文字列として扱い、文字 A を文字 B に置き換えるというプログラムを実行する必要がある。そのプログラムを用いて、県別総生産のデータのカンマを空白（スペースではなく何も存在しない）に変更することで、県別総生産のデータからカンマを外す。

ただし、このプログラムを実行する際にデータを文字列として扱っているため、作業終了時の県別総生産のデータの型は数値ではなく文字となってしまっている。

そのままでも一応重回帰分析を実行することが可能だが、データの型が文字列のままだとエラーになる可能性があるので、念のためデータの型を数値に戻すプログラムを実行して、県別総生産のデータのフォーマットの変更を終了とする。

続いて都道府県別患者数のデータのフォーマットを変更していく。

3.3.3.　患者数のデータフォーマットの変更

都道府県別患者数のデータは、これまで加工した 2 つのデータとはちがい、1 つのデータの中に都道府県別の該当期間のデータが揃っているわけではなく、都道府県別の様々な医療機関の数値を記したデータが 1 年ごとに別のデータとして用意されている。

そのため、これまで加工したデータの該当期間である 2001 年～2012 年の 12 年分、12 個のデータを用意して、それぞれのデータから必要な範囲だけを抜き出す

もともと年ごとに別のデータを使用しており、必要な範囲は総数だけなので、それぞれのデータから総数の部分だけ抜き出す。図 3.3.3.a の太枠を参照。

平成15年	病院報告	年間						
平成14年	病院報告	年間						
平成13年	病院報告	年間						
閲覧 第 9表 在院患者延数，開設者(中分類)・								
	総 数	国			会社	個人	医育機関(再掲)	
		厚生労働省	その他					
総 数								
全 国	5.12E+08	23197156	16231874		3983735	29269132	28005633	
北 海 道	33223023	1007111	981335		206561	1773079	794533	
鹿 児 島	11805549	508812	275668		・	503771	233400	
沖 縄	6530132	492322	195187		・	160894	189613	

図 3.3.3.a　都道府県別患者数から必要な部分の抜出し

これまでと違いデータが 1 つではなく 12 個のあるため、12 個のデータを読み込み、12 個のデータから必要なところを抜き出す必要がある。少し手間はかかるが、行う作業自体は同じなのでこれまで通り作業を進めていく。

R プログラムは以下の通りである。

```
> z1  <-  read.csv ("13kanja.csv")  #2001（平成 13）年から 2012（平成 24）年までの
> z2  <-  read.csv ("14kanja.csv")  #データの読み込み
> z3  <-  read.csv ("15kanja.csv")
> z4  <-  read.csv ("16kanja.csv")
> z5  <-  read.csv ("17kanja.csv")
> z6  <-  read.csv ("18kanja.csv")
> z7  <-  read.csv ("19kanja.csv")
> z8  <-  read.csv ("20kanja.csv")
> z9  <-  read.csv ("21kanja.csv")
> z10  <-  read.csv ("22kanja.csv")
> z11  <-  read.csv ("23kanja.csv")
> z12  <-  read.csv ("24kanja.csv")
> z1  <-  z1 [6:52 , 2]#それぞれの年のデータから都道府県の範囲のみを選んで抜き出す
> z2  <-  z2 [6:52 , 2]
> z3  <-  z3 [6:52 , 2]
> z4  <-  z4 [6:52 , 2]
> z5  <-  z5 [6:52 , 2]
> z6  <-  z6 [6:52 , 2]
> z7  <-  z7 [6:52 , 2]
> z8  <-  z8 [6:52 , 2]
> z9  <-  z9 [6:52 , 2]
> z10  <-  z10 [6:52 , 2]
> z11  <-  z11 [9:55 , 2]
> z12  <-  z12 [6:52 , 2]
```

プログラムリスト 3.3.3.a

次に図 3.3.1.c の要領で縦に並べる。R プログラムは以下の通りである。

```
> z   <-   data.frame (z1 , z2 , z3 , z4 , z5 , z6 , z7 , z8 , z9 , z10 , z11 , z12)
> z   <-   t (z)
> for (i in 1:12) {assign (paste ("z" , i , sep = "") , z [i , ] ) } #データを 1 列ずつに分割する
> #データを 1 列に結合する
> z_df   <-   data.frame (z1 , z2 , z3 , z4 , z5 , z6 , z7 , z8 , z9 , z10 , z11 , z12)
> z_df   <-   t (z_df)
> dim (z_df)
[1] 564     1
> write.csv (z_df , "kanja.csv" , row.names = FALSE)       #データの書き出し
```

プログラムリスト 3.3.3.b

3.4. データの統合

上記のデータのフォーマットの変更により、3 つの変数のフォーマットは等しく 564 行 1 列の形になったため、結合が可能になった。それぞれのデータは等しく 2001 年の県コード順の都道府県の下に 2002 年の県コード順都道府県というような順番で 2012 年まで並んでいる。それを図 3.3.b の形になるように目的変数である失業者数のデータを左、説明変数である県別総生産と患者数の 2 つのデータは失業者数の右に結合することで 564 行 3 列の 1 つのデータにする。ぞれぞれのデータは先程説明した順番で並んでいるが、結合した際にズレが生じていてはいけないため、最終的に目視で正しい順番で並んで結合できているかを確認し、フォーマットの変更を終了する。

R プログラムは以下の通りである。

```
> y   <-   read.csv ("situgyousya.csv")     #失業者数のデータの読み込み
> x1   <-   read.csv ("gdp.csv")            #県別総生産のデータの読み込み
> x2   <-   read.csv ("kanja.csv")          #患者数のデータの読み込み
> a   <-   data.frame (y , x1 , x2)         #3 つのデータを結合し、1 つのデータにする
> dim (a)
[1] 564     3
```

プログラムリスト 3.4.a

3.5. データの規準化

まず規準化とは複数のデータの数値に共通の規準を与えて整える事である。また、この際に用いる共通の規準はそれぞれの変数の中の数値（測定値）からそれぞれの変数の数値の平均値を引き、出てきた数値を標準偏差で割ったものである。

なぜ規準化が必要かというと、複数のデータは単位など、それぞれ計測の規準となったものが違い、そのままでは正確な結果が得られないためである。

計算式は $\mathrm{z}_i = \frac{\mathrm{x}_i - \bar{\mathrm{x}}}{S}$ である。S は標準偏差、$\bar{x}$はその変数の平均である。

R プログラムは以下の通りである。

```
> a   <-   scale (a)        #データを規準化する
> a   <-   data.frame (a)
```

プログラムリスト 3.5.a

第4章　データ解析

4.1. 重回帰分析の結果

R プログラムは以下の通りである。

```
> colnames (a)   <-   c ("y" , "x1" , "x2")
> result   <-   lm (y ~ x1 + x2 , data = a)          #重回帰分析をする
> summary (result)                                   #重回帰分析の結果の要約を表示する
```

プログラムリスト 4.1.a

上記の方法で表示した回帰分析の結果の要約は以下の通りである。

```
Residuals：
      Min        1Q    Median        3Q       Max
 -1.16830  -0.12681  -0.02176   0.08883   1.47153

Coefficients：
             Estimate   Std. Error   t value   Pr(>|t|)
 (Intercept) 2.146e-17  1.245e-02      0.00          1
 x1          3.583e-01  2.268e-02     15.80   <2e-16 ***
 x2          6.356e-01  2.268e-02     28.03   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

    Residual standard error: 0.2956 on 561 degrees of freedom
    Multiple R-squared:  0.913,     Adjusted R-squared:  0.9126
    F-statistic:  2942 on 2 and 561 DF,  p-value: < 2.2e-16
```

英語を日本語に訳すとそれぞれ以下のようになっている。

まず Residuals は残差の四分位数を表しており、詳細は左から最小値、第一四分位数、中央値、第三四分位数、最大値を表している。次に Coefficients は推定された回帰係数を表しており、詳細は左から係数、標準誤差、t 値、p 値を表している。そして Residual standard error は残差の標準誤差と自由度を表している。Multiple R-squared は寄与率、決定係数を表しており、Adjusted R-squared は調整済みの寄与率、決定係数を表している。F-statistic は F 統計量を表している。p-value は F 統計量に対する p 値を表している。

以上を踏まえ日本語に手直しした要約が以下である。

残差の四分位数:

最小値	第一四分位数	中央値	第三四分位数	最大値
-1.16830	-0.12681	-0.02176	0.08883	1.47153

推定された回帰係数:

	切片の推定値	標準誤差	t 値	p 値
(Intercept)	2.146e-17	1.245e-02	0.00	1
x1(総生産)	3.583e-01	2.268e-02	15.80	<2e-16 ***
x2(患者数)	6.356e-01	2.268e-02	28.03	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

残差の標準誤差は 0.2956 で自由度は 561 である
寄与率、決定係数: 0.913, 調整済みの寄与率、決定係数: 0.9126
F 統計量: 2942 on 2 and 561 DF, F 統計量に対する p 値: < 2.2e-16

4.2. 失業者数への影響の分析

4.2.1 重回帰分析の結果からの考察

重回帰分析により都道府県別の総生産が多いほど失業者数が多く、在院患者延数が多いほど失業者数が多いという結果が得られた。

総生産が多いほど失業者数が多かったのは想定と違ったが、これは企業が都市部に集中しているためだと考えられる。

反対に患者数が多いほど失業者数が多いのは想定通りで、これはケガや病気などで仕事をつづける事ができなくなった人がいるためだと考えられる。

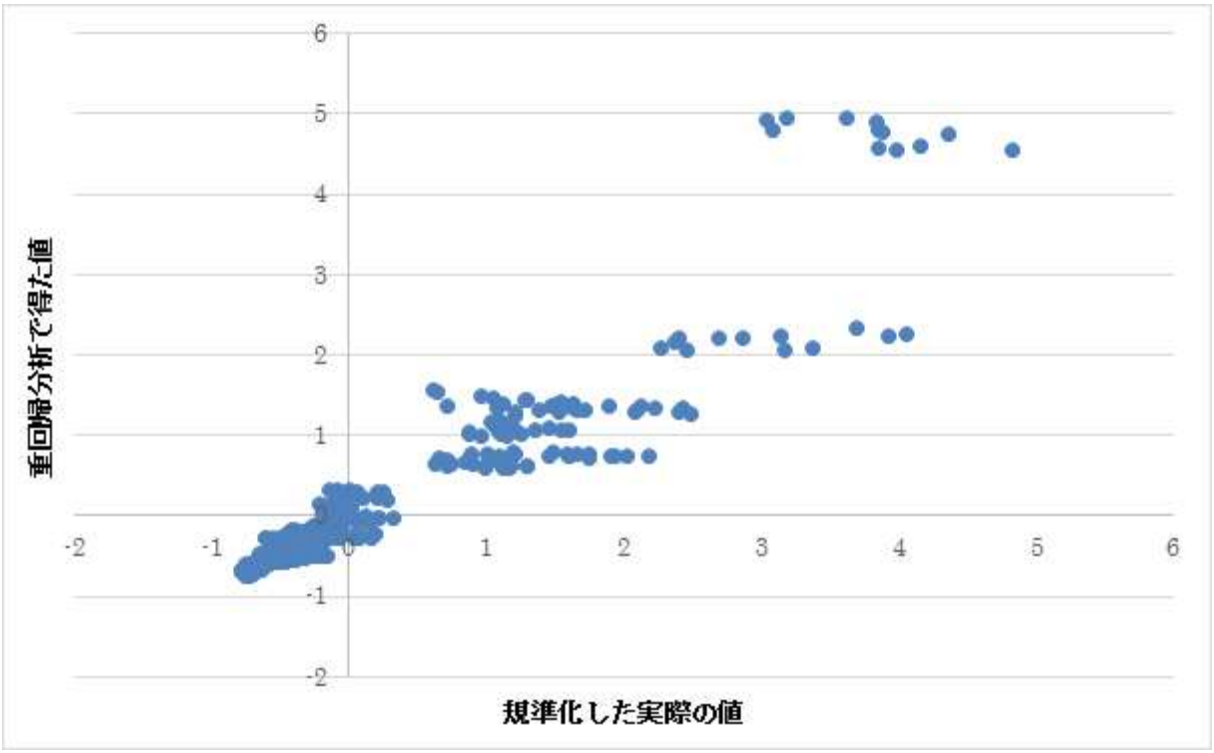

図 4.2.a　失業者数の散布図

図 4.2.a は規準化した失業者数の値と重回帰分析で得た値をグラフ化したものである。この散布図の中の点を 4 つのグループに分割する。A グループは 0.5 以下の点、B グループは 0.5～2 の間の点、C グループは 2～4 の間の点 D グループは 4 以上の点となっている。

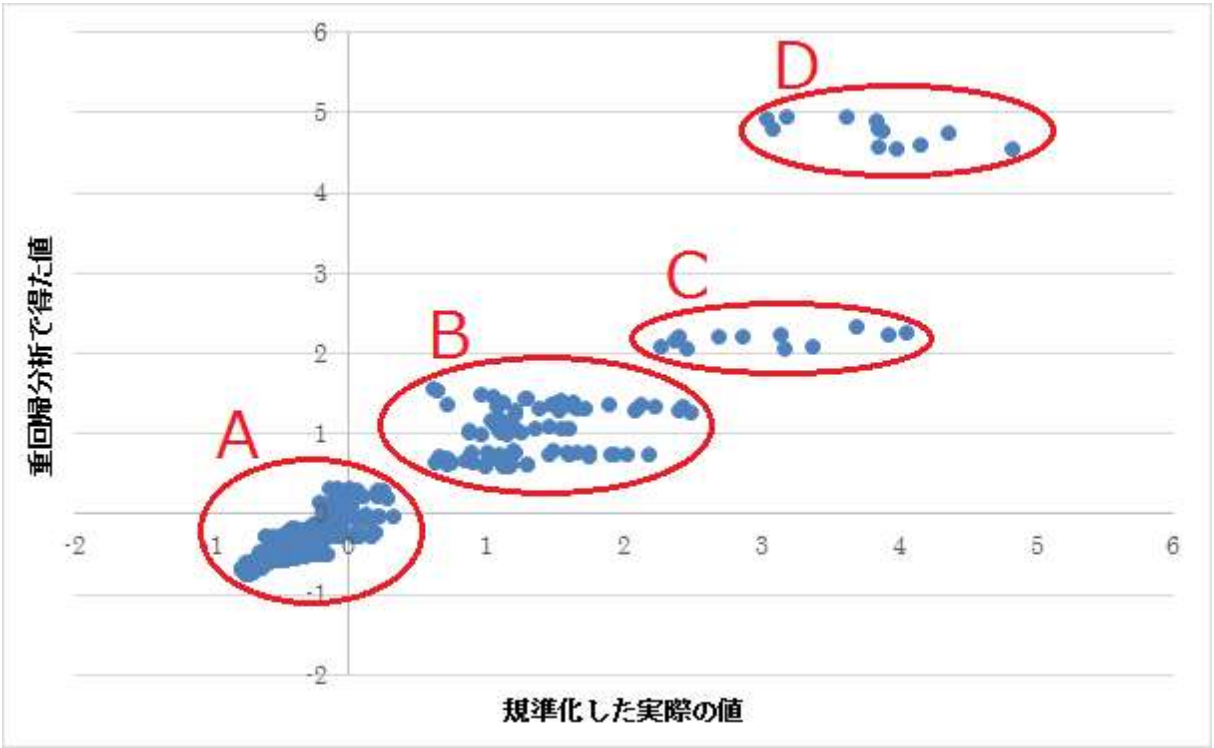

図 4.2.b　グループ分けをした失業者数の散布図

それぞれのグループの内容を表にしたものが以下の表 4.2.a である。

表 4.2.a　失業者数の散布図内のグループの詳細な内容

A	青 森 県01～青 森 県12、岩 手 県01～岩 手 県12、宮 城 県01～宮 城 県12、 秋 田 県01～秋 田 県12、山 形 県01～山 形 県12、福 島 県01～福 島 県12、 茨 城 県01～茨 城 県12、栃 木 県01～栃 木 県12、群 馬 県01～群 馬 県12、 新 潟 県01～新 潟 県12、富 山 県01～富 山 県12、石 川 県01～石 川 県12、 福 井 県01～福 井 県12、山 梨 県01～山 梨 県12、長 野 県01～長 野 県12、 岐 阜 県01～岐 阜 県12、静 岡 県01～静 岡 県12、三 重 県01～三 重 県12、 滋 賀 県01～滋 賀 県12、京 都 府01～京 都 府12、奈 良 県01～奈 良 県12、 和 歌 山 県01～和 歌 山 県12、鳥 取 県01～鳥 取 県12、島 根 県01～島 根 県12、 岡 山 県01～岡 山 県12、広 島 県01～広 島 県12、山 口 県01～山 口 県12、 徳 島 県01～徳 島 県12、香 川 県01～香 川 県12、愛 媛 県01～愛 媛 県12、 高 知 県01～高 知 県12、佐 賀 県01～佐 賀 県12、長 崎 県01～長 崎 県12、 熊 本 県01～熊 本 県12、大 分 県01～大 分 県12、宮 崎 県01～宮 崎 県12、 鹿 児 島 県01～鹿 児 島 県12、沖 縄 県01～沖 縄 県12
B	北 海 道01～北 海 道12、埼 玉 県01～埼 玉 県12、千 葉 県01～千 葉 県12、 神 奈 川 県01～神 奈 川 県12、愛 知 県01～愛 知 県12、兵 庫 県01～兵 庫 県12、 福 岡 県01～福 岡 県12
C	大 阪 府01～大 阪 府12
D	東 京 都01～東 京 都12

4.2.2.　散布図を用いたデータの説明

グループ D は東京都、グループ C は大阪府、B グループは埼玉県、千葉県、神奈川県などの都市部であることが分かった。そのため、失業者数と県別総生産や患者数の関係は、総生産や患者数自体ではなく、都道府県ごとの人口や企業数の差が失業者数の大小の原因の大半を占めていると考えられる。

これにより、当初想定していた内容の統計分析で必要だったのは失業者数ではなく、失業率であった可能性がでてきた。失業率のデータを使用することで本来想定していた分析を行う事ができるかは不明だが、2 つの結果を得る事で把握できる事実もあると考えられるため、新たに失業率のデータを用意し、改めて重回帰分析を実行してみる。

失業率のデータを目的変数とした重回帰分析行うことで、失業者数のデータの重回帰分析の結果と比較して、新たな問題点や改善点などを把握することが出来ると考えられる。

第5章　失業率を用いた解析

この章では、都道府県別失業者数のデータでは想定した結果が得られなかったため、新たに失業率のデータを用意し、今度は目的変数を失業率、説明変数を県別総生産と患者数で改めて重回帰分析を行っていく。基本的には今までの章で行ってきた事と同じ作業を行う。最初に都道府県別失業率のデータソースを提示し、次に失業率のデータの元のフォーマットを提示する。続いて失業率のデータのフォーマットを第 3 章で加工した県別総生産、患者数のフォーマットに合わせて加工し、データのフォーマットを統一する。そしてフォーマットを統一したあと、データを結合、規準化して重回帰分析を実行する。重回帰分析で得た結果から一度考察を行い、その後散布図を表示してその内容を把握、傾向の分析を行ってみる。最後に実行した作業から得られた結果をもとに考察を行い、失業者数のときに得た結果と比較して、さらに考察を行う。

5.1.　失業率のデータソース

新たに次の表のデータを用意し、改めて重回帰分析を行う。

変数	出典	期間
失業率　（y）	総務省統計局　労働力調査[1]	1997 年～2015 年

注 1：http://www.stat.go.jp/index.htm

5.2.　失業率の原フォーマット

都道府県別の失業率のデータの原フォーマットは、縦方向に 1997 年から 2015 年までの期間が並んでおり、横方向に北海道から沖縄県までの都道府県が県コードの順番に並んでいる。データの単位は%である。

失業率のデータの期間は第 3 章で加工した県別総生産、患者数の期間である 2001 年～2012 年を含んでいるため、それにあわせてデータを加工していけばいい。したがって、すでに加工している県別総生産、患者数のデータには手を付けず、失業率のデータのフォーマットをその 2 つと同じフォーマットに加工していく。

表 5.2.a　都道府県別失業率の元データ

		労働力調査参考資料					
		[年平均]					
			第6表　都道府県別完全				
	(%)						
		1	2	3	46	47	
		北海道	青森県	岩手県	鹿児島県	沖縄県	全国
平成9年	-1997	3.7	3.9	2.4	2.9	6	3.4
10	-1998	4.8	4.8	3	3.8	7.7	4.1
11	-1999	5	5.3	3.5	3.6	8.3	4.7
12	-2000	5.5	5.3	3.6	3.6	7.9	4.7
13	-2001	5.8	5.7	4.3	4.2	8.4	5
24	-2012	5.2	5.3	3.9	4.5	6.8	4.3
25	-2013	4.6	4.9	3.3	4.4	5.7	4
26	-2014	4.1	4.4	2.9	4	5.4	3.6
27	-2015	3.4	4.5	2.9	3.5	5.1	3.4

5.3.　失業率のデータフォーマットの変更

第 3 章で加工した県別総生産、患者数にあわせて、データは 2001 年から 2012 年の範囲を抜き出す。図 5.3.a の太枠を参照。

		労働力調査参考資料					
		[年平均]					
			第6表　都道府県別完全				
	(%)						
		1	2	3	46	47	
		北海道	青森県	岩手県	鹿児島県	沖縄県	全国
平成9年	-1997	3.7	3.9	2.4	2.9	6	3.4
10	-1998	4.8	4.8	3	3.8	7.7	4.1
11	-1999	5	5.3	3.5	3.6	8.3	4.7
12	-2000	5.5	5.3	3.6	3.6	7.9	4.7
13	-2001	5.8	5.7	4.3	4.2	8.4	5
24	-2012	5.2	5.3	3.9	4.5	6.8	4.3
25	-2013	4.6	4.9	3.3	4.4	5.7	4
26	-2014	4.1	4.4	2.9	4	5.4	3.6
27	-2015	3.4	4.5	2.9	3.5	5.1	3.4

図 5.3.a　都道府県別失業率から必要な部分の抜出し

R 言語によるプログラム内容は以下の通りである。

```
y   <-   read.csv ("lt06y.csv")   #データの読み込み
y   <-   y [12:23 , ]
y   <-   y [ , 3:49]
```

プログラムリスト 5.3.a

次に第 3 章で失業者数のデータに行ったように行と列を入れ替える。

R プログラムは以下の通りである。

```
y   <-   t (y)
y   <-   data.frame (y)
dim (y)
[1] 47 12
```

プログラムリスト 5.3.b

次に年ごとに分割し、それを図 3.3.1.c の要領で縦に並べる。

R プログラムは以下の通りである。

```
> y   <-   t (y)
> for (i in 1:12) {assign (paste ("y" , i , sep = "") , y[i , ] ) } #データを 1 列ずつに分割する
> #データを 1 列に結合する
> y_df   <-   data.frame (y1 , y2 , y3 , y4 , y5 , y6 , y7 , y8 , y9 , y10 , y11 , y12)
> y_df   <-   t (y_df)
> dim (y_df)
[1] 564    1
> write.csv (x_df , "situgyouritu.csv" , row.names = FALSE)    #データの書き出し
```

プログラムリスト 5.3.c

5.4. データの統合

上記のデータのフォーマットの変更により、失業率のデータも県別総生産や患者数と同じフォーマットになったため、結合が可能となった。第 3 章の時と同じようにデータを統合し、重回帰分析が可能な状態にする。R プログラムは以下の通りである。

```
y   <-   read.csv ("situgyouritu.csv")
x1   <-   read.csv ("gdp2.csv")
x2   <-   read.csv ("kanja1.csv")
a   <-   data.frame (y , x1 , x2)
dim (a)
[1] 564    3
```

プログラムリスト 5.4.a

5.5. データの規準化

同じように規準化も行う。R プログラムは以下の通りである。

```
a  <-  scale (a)
a  <-  data.frame (a)
```

プログラムリスト 5.5.a

5.6. 重回帰分析

5.6.1. 重回帰分析の結果

```
> colnames (a)  <-  c ("y" , "x1" , "x2")
> result  <-  lm (y ~ x1 + x2, data = a)          #重回帰分析をする
> summary (result)                                #重回帰分析の結果の要約を表示する
```

プログラムリスト 5.6.a

上記の方法で表示した回帰分析の結果の要約は以下の通りである。

```
Residuals：
      Min        1Q      Median      3Q     Max
  -1.8635   -0.5625   -0.1046   0.4164  4.0528

Coefficients：
              Estimate    Std. Error   t value   Pr(>|t|)
(Intercept)  -1.020e-17   3.818e-02     0.000          1
x1           -4.635e-01   6.957e-02    -6.662   6.45e-11 ***
x2            7.279e-01   6.957e-02    10.464    <2e-16  ***
---
Signif. codes：  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9067 on 561 degrees of freedom
Multiple R-squared:  0.1809,     Adjusted R-squared:  0.1779
F-statistic:  61.93 on 2 and 561 DF,  p-value: < 2.2e-16
```

そして英語をそれぞれ日本語に手直しした要約が以下である。

```
残差の四分位数:
    最小値      第一四分位数      中央値      第三四分位数      最大値
  -1.8635        -0.5625       -0.1046        0.4164        4.0528

推定された回帰係数:
                切片の推定値    標準誤差      t 値      p 値
(Intercept)     -1.020e-17   3.818e-02     0.00          1
x1(総生産)      -4.635e-01   6.957e-02    -6.662   6.45e-11 ***
x2(患者数)       7.279e-01   6.957e-02    10.464    <2e-16 ***
---
Signif. codes:  0  ‘***’  0.001  ‘**’  0.01  ‘*’  0.05  ‘.’  0.1  ‘ ’  1

残差の標準誤差は 0.9067 で自由度は 561 である
寄与率、決定係数:  0.1809,  調整済みの寄与率、決定係数:  0.1779
F 統計量:  61.93 on 2 and 561 DF,  F 統計量に対する p 値: < 2.2e-16
```

5.6.2. 失業率への影響の分析

都道府県別総生産が少ないほど失業率が高く、患者数が多いほど失業率が高いという結果を得ることができた。今回は失業者数の時とは違い、県別総生産と患者数と失業率の関係は想定した通り、景気が悪いほど失業率が高く、けが人や病人が多いほど失業率が高いというものだった。

しかし、決定係数の項目が 0 に近い値となっているため、これらの変数はほとんど関係がないといえるほどに微弱な関係性しかないという結果であった。これにより想定通りの結果が得られたとはいえ、それは偶然の範囲である可能性も高い。

そのため、微弱な関係性しかないという結果が出たが、失業者数と同じく一応散布図を用意して、都道府県がそれぞれどういうグループに属しているかを調べ、微弱な関係性の中でなにか傾向のようなものが確認できないかを調べてみる。

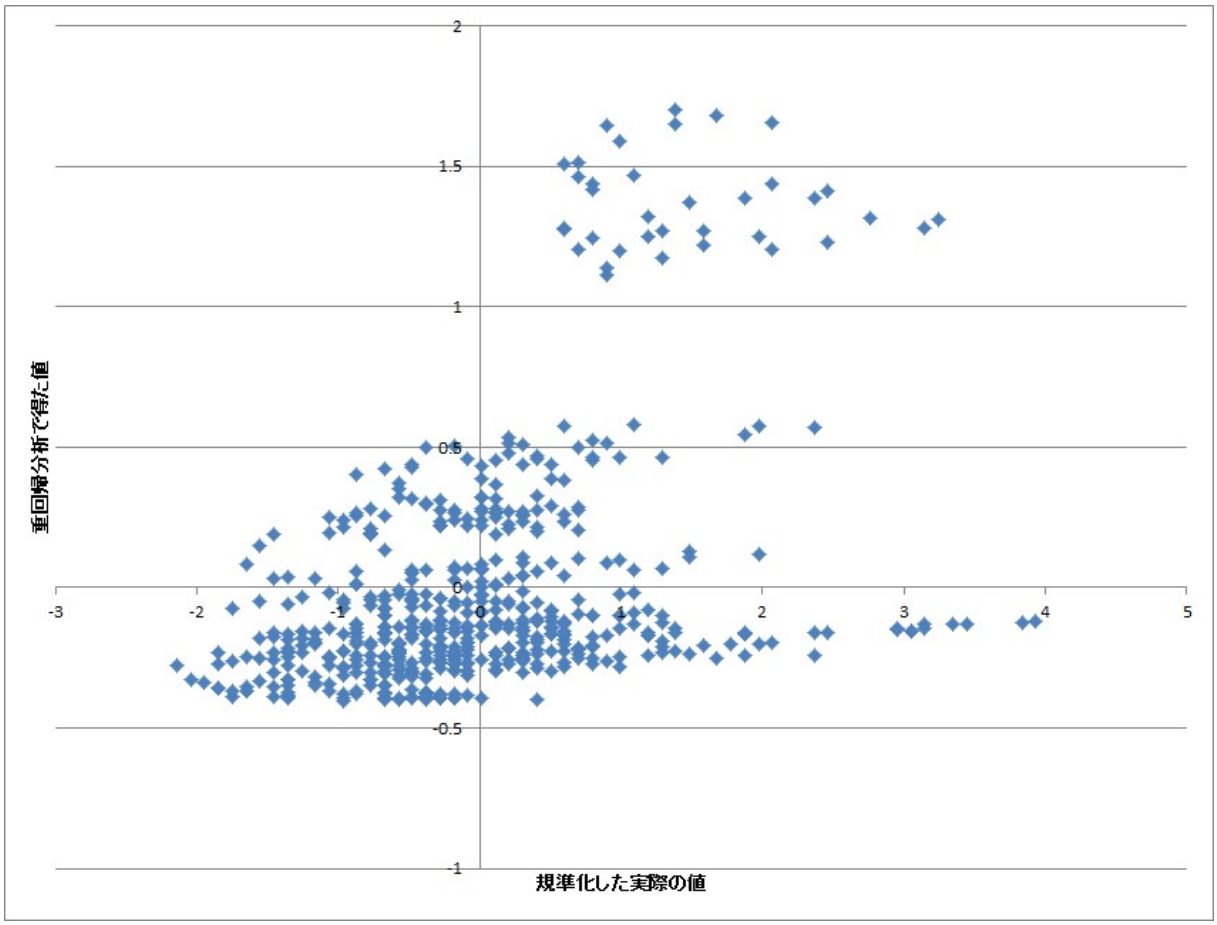

図 5.6.a　失業率の散布図

散布図を表示してみたが、点は全体的に散っており、これといった規則性がなく、やはり、ほとんど関係がないといえるほど微弱な関係性しか存在していないという結果に間違いはないと考えられる。

上の方に集まっている点が周囲から浮いていると言えるため、失業者数の時と同じくグループ分けを行い、内容を調べる。グループ分けが 2 つだけでは、傾向の分析をしても大した情報は得られず、恐らく今以上の傾向の分析ができないと考えられるため、下のまとまりの中でも一応の切れ目が確認できる重回帰分析で得た値の 0.2 付近でグループを分け、合計 3 グループに分けて内容を調べてみる。

重回帰分析で得た値が 0.2 以下の点をグループ A、0.2 以上 1 以下の点をグループ B、1 以上の点をグループ C に分けたものが図 5.6.b である。このグループにそれぞれどの都道府県が属しているかを調べることで出来る限りの傾向の分析を行ってみる。

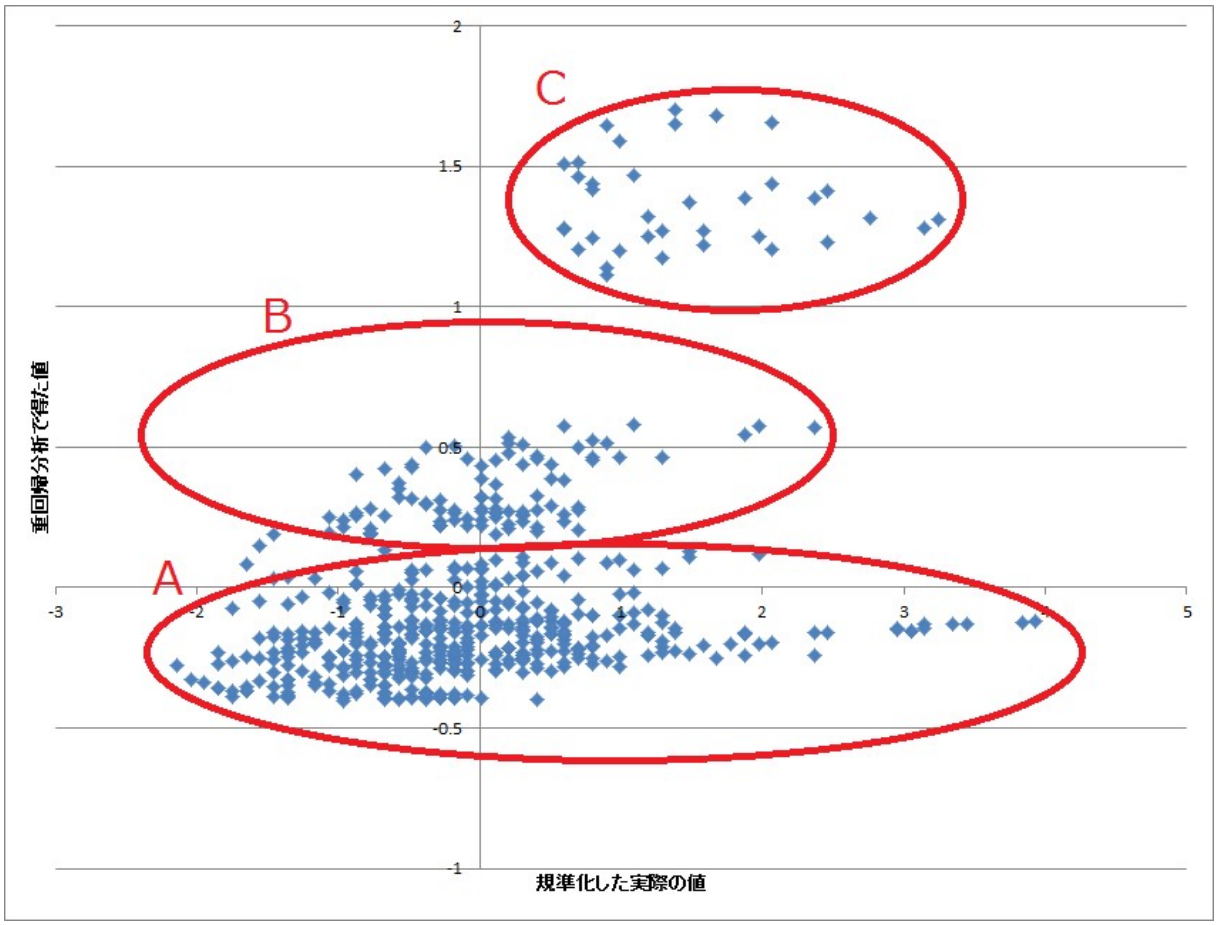

図 5.6.b　グループ分けをした失業率の散布図

それぞれのグループの内容を表にしたものが以下である。

表 5.6.a　失業率の散布図内のグループの詳細な内容

A	青森県01～青森県12、岩手県01～岩手県12、宮城県01～宮城県12、 秋田県01～秋田県12、山形県01～山形県12、福島県01～福島県12、 茨城県01～茨城県12、栃木県01～栃木県12、群馬県01～群馬県12、 東京都01～東京都12、新潟県01～新潟県12、富山県01～富山県12、 石川県01～石川県12、福井県01～福井県12、山梨県01～山梨県12、 長野県01～長野県12、岐阜県01～岐阜県12、静岡県01～静岡県12、 愛知県06～愛知県08、愛知県11～愛知県12、三重県01～三重県12、 滋賀県01～滋賀県12、京都府01～京都府12、奈良県01～奈良県12、 和歌山県01～和歌山県12、鳥取県01～鳥取県12、島根県01～島根県12、 岡山県01～岡山県12、広島県07、広島県11～広島県12、 山口県01～山口県12、徳島県01～徳島県12、香川県01～香川県12、 愛媛県01～愛媛県12、高知県01～高知県12、佐賀県01～佐賀県12、 長崎県01～長崎県12、大分県01～大分県12、宮崎県01～宮崎県12、 鹿児島県11～鹿児島県12、沖縄県01～沖縄県12
B	埼玉県01～埼玉県12、千葉県01～千葉県12、神奈川県01～神奈川県12、 愛知県01～愛知県05、愛知県09～愛知県10、兵庫県01～兵庫県12、 広島県01～広島県06、広島県08～広島県10、熊本県01～熊本県12、 鹿児島県01～鹿児島県10
C	北海道01～北海道12、大阪府01～大阪府12、福岡県01～福岡県12

グループ C の内容は北海道、大阪、福岡となっており、他の都府県と比べて実際の失業率と回帰結果の失業率がともに高い地域であり、対象とした期間は経済の停滞が影響した可能性が高い。

グループ B は埼玉、千葉、神奈川と比較的都市部に近い県が属していたが、東京は属しておらず、拡大都市部の特徴が出ていると思われる。

グループ A は点が万遍なく展開している。

ここまで関係が薄いというのは想定外であったが、散布図を用意したことで、失業者数の時の内容と比べ、地域間の特徴をはっきりと捉えることができた。当初想定していた結果とは異なったが、2 通りの分析を行ったことによって、統計分析により傾向の分析が可能であるという事実をしっかりと体験する事ができた。反省点も見つかったため、今後統計分析をする際は、今回の経験を踏まえ、より正確な分析ができるように努力しようと思う。

第6章　おわりに

6.1.　まとめ・結論

失業者数のデータを用いた重回帰分析はしっかりと相関関係にあり、傾向の分析もしっかりと行う事ができた。大体は想定通りの結果を得る事ができたが、総生産が高い、つまり景気がいいと失業者も多いという事だけは想定外だった。

そのためなぜこの結果になったかを調べるため、散布図を表示してさらに細かく分析を行った。その結果、都市部と地方で差がついていることが判明し、人口や企業の数がそのまま失業者の数に影響している可能性が高い事が分かった。

失業者数で行った重回帰分析から、失業者の数では人や企業の数に比例して値が大きくなっていると分かったため、失業者の数ではなく、失業率のデータを使用することで人や企業の数に比例した値の増加の問題は解決できると考え、改めて失業率で重回帰分析を行っていく。

失業率のデータを用いて改めて行った重回帰分析では、失業者数の時に生じた問題点である総生産が多いほど失業者数が多いという問題を改善する事ができた。

6.2.　今後の課題

失業者数のデータを用いた重回帰分析で生じた問題は、本来想定した組み合わせである失業者数と総生産と患者数の関係性を調べようとしたが、結果的には人口や企業の数などが理由で説明出来るものであり、総生産や患者数が失業者数を説明したとは言い難かったことである。

だがこれは改めて考えると事前に予想が可能な範囲の問題であった。失業率のデータも同じで、失業者数のデータで行った重回帰分析での失敗から急いで用意したデータだったため確認を怠ったが、相関関係があるかどうかは事前に調べ重回帰分析を行う意義があるかしっかりと検討する必要があると痛感した。

これらの問題点から、次回以降、統計分析する際はこれと同じ失敗をしないよう、目的変数と説明変数の選択には気を配っていこうと思う。

今回は事前準備の段階での失敗が目立ち、それ以外は順調に重回帰分析を行う事ができたため、事前準備の重要性を感じたが、これ以外にも単純な失敗は起こりうる可能性が高いので、事前準備だけでなく、作業のさまざまな工程で必要に応じた確認を行い、正しい分析を行う事ができるように心がけていこうと思う。

そして、図 5.6.b をさらにクラスタ分けする分析方法も今後検討する必要がある。

謝辞

最後に、2 年次にお世話になった伊藤則之教授、3 年次、4 年次で統計分析と卒業論文の作成について指導をしてくださった田中章司郎教授に深く御礼申し上げます。

なお、本論文、本研究で作成したプログラム及びデータ、資料などの全ての知的財産権を本ゼミナールの指導教員である田中章司郎教授に譲渡致します。本論文をインターネット等で公開しても差し支えありません。

参考文献

(1) 前川功一（2014）『経済・経営のためのよくわかる統計学』

(2) 金明哲（2007）『R によるデータサイエンス-データ解析の基礎から最新手法まで』

参考ホームページ

(1) e-Stat 政府統計の総合窓口 https://www.e-stat.go.jp/SG1/estat/eStatTopPortal.do

(2) 総務省統計局ホームページ http://www.stat.go.jp/index.htm

(3) 内閣府ホームページ http://www.cao.go.jp/