

平成28年度

卒業論文

題目	1980年以降の日本における自殺者数に影響を
	及ぼす社会・経済的要因の定量的検証

文字数

17015

学籍番号 201314057

氏名 大西勇人

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)

印

広島経済大学

要旨

本論文では自殺者数が社会・経済的要因とどのような関係にあるか重回帰分析を用いて研究した。具体的には自殺者数を目的変数、倒産件数、失業率、消費者物価指数を説明変数としどの説明変数が自殺者数に影響を及ぼすのかを調べた。使用したデータの年数はそれぞれ 1980 年～2013 年までのものであり、それぞれ日本国内の総数データを使用した。仮説の段階では最も相関があるのは倒産件数であると予想した。分析結果は予想と異なり、失業率が最も自殺者数と関係が高かった。倒産件数は任意倒産なども含むため他 2 つよりも影響が低くなったと考えられる。また倒産件数は有意性が認められず失業率、消費者物価指数は有意性が認められた。よって自殺者数は失業率、消費者物価指数と高い関係があると伺える。

目次

1 序論	1
1.1 研究の背景	1
1.2 先行研究及び本論文の位置づけ	2
2 解析のためのデータ準備	4
2.1 使用データ	4
2.2 使用データの加工	4
3 変数間の相関関係	13
4 重回帰分析	15
4.1 すべての説明変数を用いた重回帰分析	15
4.2 優位性のある説明変数のみの重回帰分析	23
5 考察	26
6 終章	28
6.1 これからの展望	28
参考文献	29

1 序論

1.1 研究の背景

表 1.1 2015 年自殺者数割合

順位	国名	全体	男性 順位	男性	女性 順位	女性
1	 ガイアナ	44.2	1	70.8	1	22.1
2	 韓国(詳細)	28.9	5	41.7	5	18.0
3	 スリランカ	28.8	3	46.4	7	12.8
4	 リトアニア	28.2	2	51.0	29	8.4
5	 スリナム	27.8	4	44.5	11	11.9
6	 モザンビーク	27.4	8	34.2	2	21.1
7	 タンザニア	24.9	13	31.6	4	18.3
	 ネパール	24.9	17	30.1	3	20.0
9	 カザフスタン	23.8	6	40.6	21	9.3
10	 ブルンジ	23.1	9	34.1	9	12.5
11	 インド	21.1	22	25.8	6	16.4
12	 南スーダン	19.8	19	27.1	7	12.8
13	 トルクメニスタン	19.6	11	32.5	32	7.5
14	 ロシア	19.5	7	35.1	46	6.2
	 ウガンダ	19.5	20	26.9	10	12.3
16	 ハンガリー	19.1	12	32.4	34	7.4
17	 日本(詳細)	18.5	20	26.9	17	10.1
18	 ベラルーシ	18.3	10	32.7	44	6.4
19	 ジンバブエ	18.1	18	27.2	18	9.7
20	 ブータン	17.8	32	23.1	13	11.2
21	 スーダン	17.2	33	23.0	12	11.5
22	 コモロ	16.9	29	24.0	15	10.3
23	 ウクライナ	16.8	16	30.3	55	5.3
24	 ポーランド	16.6	15	30.5	98	3.8
	 赤道ギニア	16.6	27	24.1	27	8.6
26	 エリトリア	16.3	22	25.8	24	8.7
27	 ラトビア	16.2	14	30.7	86	4.3
	 ケニア	16.2	26	24.4	29	8.4

世界各国の国々で自ら命を絶つ人々は多い。自殺者の割合が高い国は貧困国が多いが、先進国である日本も自殺者数の割合が高い(表 1.1)[1]。

政府は 1998 年に経済状況の悪化によって自殺死亡者数が急増したことを受けて本格的に自殺対策に取り組み始めた[2]。現在制定されている自殺対策基本法では、自殺対策の基本理念として、自殺が個人的な問題としてのみ捉えられるべきではなく、その背景に様々な社会的要因があることを踏まえ、国、地方公共団体、事業主、国民のそれぞれが責務を果たし、自殺対策を社会的な取り組みとして実施することが定めら

れている[2]。また、自殺対策基本法に基づき、政府が推進すべき自殺対策の方針として自殺総合対策大綱が 2007 年に閣議決定で定められた。ここでは、具体的な数値目標として、2016 年までに 2005 年の自殺死亡率を 20%以上減少させることが設定されている[2]。

警察庁によると[3]2015 年の全国の子殺者は前年と比べて 1456 人(5.7%)減の 2万 3971 人で、6年連続の減少となった。4 年連続で 3 万人を下回り、18 年ぶりに 2万 5千人を割り込んだ[4]減少傾向ではあるが自殺者が多いことに変わりはない。

1.2 先行研究及び本論文の位置づけ

表 1.2 CINI II における“自殺”AND“経済”AND“要因”のキーワード検索結果表

論文名	著者	出版	ページ	日付
家族の絆が支える命・都道府県別データによる自殺要因分析	藤本真代 横山和輝	オイコシカ	1-15	2016-03
自殺率に対する積極的労働市場政策の効果 OECD21カ国1980～2007年のパネルデータ分析	柴田 悠	社会学評論	116-133	2014
欧州連合における平均寿命、健康寿命と主観的健康感について	長谷川 卓志	日本老年医学会雑誌	144-150	2014
群馬県山間部の高齢化と健康寿命に関する自殺率上昇要因	瀧野 秀英子 山本 重夫子	北関東医学	325-333	2014
2000年代の高自殺リスクと男女差：既存統計資料の整理と課題抽出に向けて	小森田 龍生	専修人間科学論集	117-126	2013-03
中国農村部男女別自殺死亡率の低下と社会経済的要因に関する研究	盧綺金	秋田県公衆衛生学雑誌	21-33	2012-12
自殺増加の要因とその抑制に向けての一考察「宗教性」をめぐって	石渡照子	生命倫理	33-42	2011-09-25
健康のための社会政策：ドイツの事例から（特集）健康のための社会政策	田中 祥子	社会政策	5-18	2012-10-01
Zainの議員負担度日本語版の短縮版（ZB1.0）を用いた家族介護者の負担感分析：介護負担感要因のモデル化	徳永 初み 羽生 正宗	山口経済学雑誌	219-253	2014-09
昨今の自殺動向と生命保険	白米 知仁	日本保険医学会誌	102-119	2011-06-07
若老介護の現状分析	羽生 正宗	山口経済学雑誌	303-341	2010-01
社会経済的要因からみた自殺予防における民生委員の果たすべき役割	今田 隆一	民医連医療	54-58	2010-10
山梨県の自殺率と人口・世帯・産業・経済および医療・福祉要因に関する生態学的研究	小田切 陽一	山梨県立大学看護学部紀要	1-8	2010-03-01
山梨県の自殺率と人口・世帯・産業・経済および医療・福祉要因に関する生態学的研究	小田切 陽一	山梨県立大学看護学部紀要	1-8	2010-03-01
自殺死亡率と加齢の関連性：加齢の効果を検証する社会経済的要因	田中 剛 渡藤 京則	杏林医学雑誌	307-322	2010-01
自殺死亡率と加齢の関連性：加齢の効果を検証する社会経済的要因	田中 剛 渡藤 京則	杏林医学雑誌	307-322	2010-01
高齢者の主観的健康感とQOLとの関係：地域在住女性高齢者を対象として	津田 勝彦 村田 伸 堀江 淳 村田 潤 大田 尾浩	日本理学療法学会大会	1	2010
社会・経済的要因を変えた自殺のリスクと若者に対する精神保健的支援の可能性—心理学的アプローチに関する研究—	藤又 陽一郎 松本 俊彦 高橋 祥友	精神医学	431-440	2009-05
日本の左翼と右翼の源流	岡崎 正道	言語と文化・文学の諸相	105-119	2008-03-21
山形県における自殺死亡の地域格差と人口動態的および社会経済的要因の関連性についての研究報告	菅野 聖一	山形県衛生研究所報	62-65	2007-07
都市部における自殺の構造的説明：自殺率の異なる地区の社会・経済的要因の分析	木下 藤子 岸田 秀樹 足利 吉秀 江副 智子 飯田 英博	藍野学院紀要	47-57	2007
自殺の経済社会的要因に関する調査研究について	金子 浩之、飯島 亜希	藍野学院紀要	52-56	2006-10
自殺の心理・社会的側面：我々は自殺予防活動において何を考慮にいれるべきか（特集）自殺予防のための多角的取り組み（医学・心理・社会を軸として）	坂本 真土	ESP economy/society/policy	42-53	2006-07
自殺の心理・社会的側面：我々は自殺予防活動において何を考慮にいれるべきか（特集）自殺予防のための多角的取り組み（医学・心理・社会を軸として）	坂本 真土	ESP economy/society/policy	42-53	2006-07
貸付市場における金利上限規制の経済分析	瀬下 博之	消費者金融ジャーナル研究学会年報	43-62	2005-09-30
自殺の社会経済的要因と自殺予防の経済効果（特集）社会経済の変化と自殺予防	金子 能宏 渡部 武久 山崎 晴子	季刊社会政策研究	75-87	2004
動向 自殺の社会経済的要因と自殺予防政策の国際比較	山下 志穂 金子 能宏 辰町 吉秀	海外社会政策研究	89-97	2003
P394 日本における自殺の動向と自殺率と社会経済的要因との関連性についての検討	森村 孝英 久保田 進也 三島 徳雄 永田 靖史	産業衛生学雑誌	754	2002-03-20
東京都における自殺者の動向とその要因に関する一考察—1984年～1990年の7年間の自殺者を中心に—	島村 孝英 越水 重四郎	日本赤十字看護大学紀要	70-78	2001-03-10
資料 東京都における自殺者の動向とその要因に関する一考察—1984年～1990年の7年間の自殺者を中心に—	島村 孝英 越水 重四郎	日本赤十字看護大学紀要	70-78	2001-03-10
資料 山形県における外国死亡に関する統計的検討	菅野 聖一	山形県衛生研究所報	104-111	2011
自殺死亡率と加齢の関連性：加齢の効果を検証する社会経済的要因（博士論文）杏林大学学位論文要旨および審査要旨	田中 伸和	杏林医学雑誌	536-537	1999-12-31
自殺死亡率と加齢の関連性：加齢の効果を検証する社会経済的要因	田中 伸和	杏林医学雑誌	536-537	1999-12-31
自殺死亡率と経済変動要因に関する重回帰分析結果について	三好 久規 奥田 勝弘	日本農村医学雑誌	1046-1053	1993
自殺死亡率と経済変動要因に関する重回帰分析結果について	三好 久規 奥田 勝弘	日本農村医学雑誌	1046-1053	1993
寒冷期の日照時間と自殺の季節変動量 北海道の（1950～57）の調査から	江頭 和道 阿部 和彦	日本生気象学会雑誌	3-7	1990

論文、図書・雑誌、博士論文検索データベース(CINII)で”自殺”AND”経済”AND”要因”のキーワード検索を行い、2016年7月21日時点で36件の検索結果が得られた。その中でも関連のある論文、文献は3件あった。過去の研究論文において同じ様なアプローチで研究されたものとしては、「自殺死亡率と経済的要因に関する重回帰分析結果について」[5]がある。その論文によると、我が国の自殺死亡率を時系列的に観察すると、戦前戦後を通じて高い値と低い値を繰り返す、いわば循環変動が存在する。社会環境の変化によって自殺者数の割合も変動する。この論文では昭和60年の単年のデータを用い、県別訂正自殺死亡率を目的変数、社会経済的要因として県別一人当たりの所得指標、県外転出比、自殺の主な動機として取り上げられる、病苦、精神異常の指標として病死者数、精神異常者数等を説明変数として重回帰分析を行っている。本論文では単年データのみを用いるのではなく1980年以降の長期データを用いて研究を行う。

「自殺死亡率と加齢の関連性～加齢の効果を装飾する社会経済的要因～」[6]では1990から1995年の5年間において特定の条件や属性でユーザーをグループ分けし、どのような変化があるかを長期的に分析する手法であるコホート分析を用い研究している。本論文では1980年からのより長い期間でのデータと重回帰分析といった違う分析方法で研究を行うので分析方法と対象期間が異なる。

「中国農村部男女別自殺死亡率の低下と社会経済的要因に関する研究」[7]では中華人民共和国(以下中国)の1987年から2009年まで23年間の農村部男女別自殺死亡率および6つの社会経済指標の統計データ、農村部1世帯当たり人員、男女別文盲人口比率、男女別離婚人口比率、男女別職業従事者率、労働者一人当たりの扶養家族数、一人当たりの年間純所得を用い、男女別に主成分回帰分析を行っている。農村特有の自殺リスク因子の回避と低減、農村における男女別の自殺への抵抗性の向上、及び農村部における男女の社会経済的地位の改善が、農村部男女別自殺死亡率の低下の原因として考えられることが示唆された。互いに相関のある多変数の変数を互いに無相関な少数の変数に要約する、主成分分析を用いて研究を行っており、重回帰分析で検証を行う本論文とは異なる。また対象国も本論文とは異なっている。

2 解析のためのデータ準備

2.1 使用データ

目的変数は自殺者数、説明変数は以下の 3 つである。自殺者数は内閣府が発表したものであり[8]、倒産件数の推移は中小企業白書 2014 年度版の中小企業・小規模事業者の倒産件数である[9]、消費者物価指数は e - S t a t 政府統計の総合窓口の平成 22 年度基準消費者物価指数、長期時系列データである[10]、失業率は「真の失業率——2014 年 12 月までのデータによる更新」[11]の 4 データを CSV 方式でダウンロードし、使用する。またデータはすべてインターネット上に無料で公開されているもので、1980 年～2013 年までの範囲を使用する。分析には統計解析用ソフトウェアの R を用いる。

2.2 使用データの加工

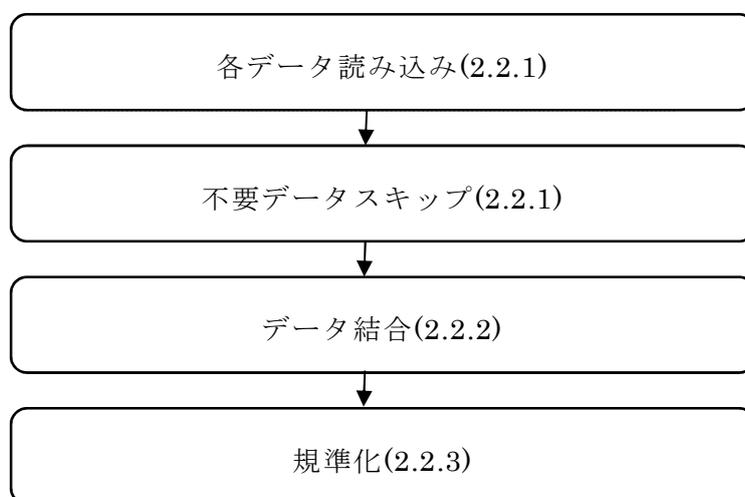


図 2.1 加工の流れ

2.2.1 データ読み込み

ダウンロードした 4 つのデータはそれぞれ数値以外の文字が入っていたり、記号が入っていたりするためそのままの状態分析を行うとエラーが起きる。そのためそれぞれのデータのいらぬ部分や行を飛ばして読み込ませる必要がある。

まずはデータフレームの作成を行う。データが格納されたオブジェクトのことをデータフレームという。

このデータフレーム名は自由に決めることができ、自殺者数はここでは *y* というデータフレームに挿入していく。下の図のコードをRに打ち込む。

```
>y<-read.csv("y.deth.csv", skip=2)
```

図 2.2 Rプログラムコード

表 2.1 自殺者数元データ

第 1-1 図 自殺者数の推移(自殺統計)			
年	自殺者数	男	女
1978	20788	12859	7929
1979	21503	13386	8117
1980
1981	20434	12942	7492
1982	21228	13654	7574

2012	27858	19273	8585
2013	27283	18787	8496

関数 `read.csv("挿入したいデータ名")` でデータフレーム *y* に任意のデータを挿入することができる。図 2.2 のコードで `csv` 形式のデータ `"y.deth.csv"` (自殺者数) をデータフレーム *y* に挿入する。元のデータには数値以外にも分析に必要な文字や空白が上から 2 行分存在した。このように不必要なデータは関数 `read.csv` の引数である「`skip=`」でデータの上から何行読み飛ばして読み込むかを指定することができる。図 2.2 では `skip=2` としたため、`"y.deth.csv"` (自殺者数) の上から 2 行分を読み飛ばして、データフレーム *y* に挿入するという操作になる。

表 2.2 データフレーム *y* に挿入されたデータ

年	自殺者数	男	女
---	------	---	---

1978	20788	12859	7929
1979	21503	13386	8117
~~~~~			
2012	27858	19273	8585
2013	27283	18787	8496

これでデータフレーム  $y$  に表 2.1 の自殺者数データが挿入された。今回の分析で使用するデータは 1 列目年数、2 列目総数のみである。3 列目男、4 列目女は使用しないため、使用する 1、2 列目をデータフレーム  $y$  からこのみ抽出する。

```
> y <- y[c(1, 2)]
```

図 2.3 R プログラムコード

新しくデータフレーム  $y$  を作成する。データから必要な列を抽出するには「データフレーム[ $c$ (抽出したい列番号)]」で抽出できる。複数列を選択して抽出したい場合図 2.3 のように「,」を用いて区切ることで選択できる。

表 2.3 データフレーム  $y$  に挿入されたデータ

年	自殺者数
1978	20788
1979	21503
~~~~~	
2012	27858
2013	27283

これでデータフレーム y に表 2.2 の自殺者数データが挿入された。上記と同じ手順で倒産件数、失業率、消費者物価指数の各データをそれぞれデータフレーム x_1 、 x_2 、 x_3 に挿入し、データの取捨選択を行い分析が行える状態にする。

2.2.2 データ結合

異なる表に保存されているデータを 1 つに纏めなければ分析することができない。そのため、それぞれのデータフレーム (y (自殺者数)、 x_1 (倒産件数)、 x_2 (失業率)、 x_3 (消費者物価指数)) を 1 つに纏めていく。

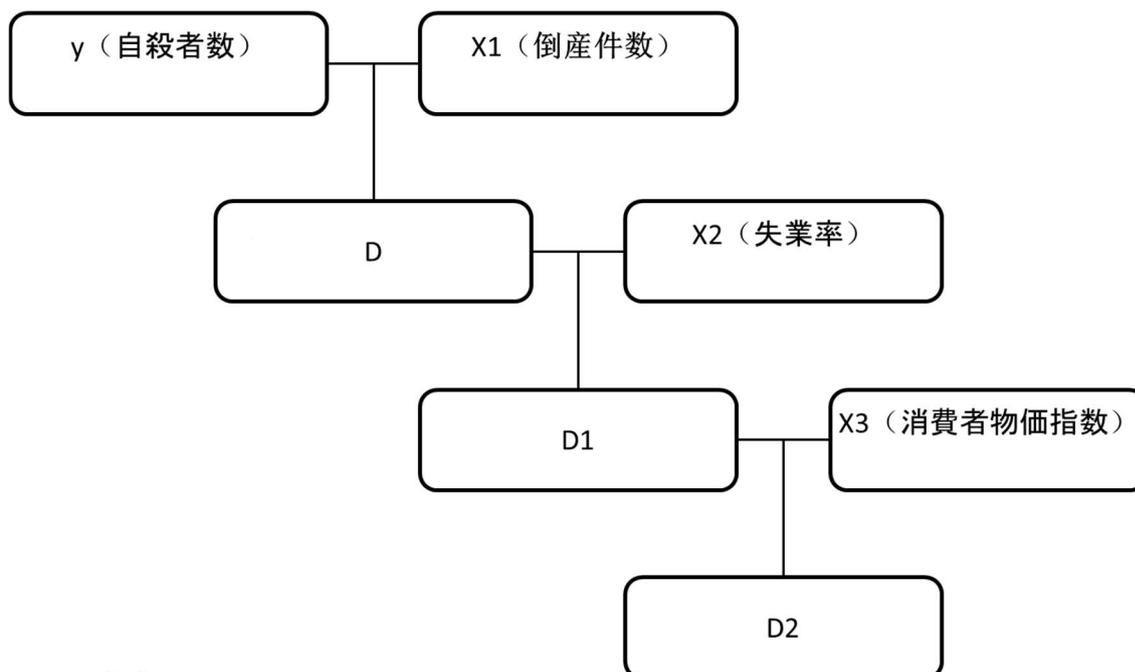


図 2.4 結合フローチャート

流れとしては図 2.4 ように二つのデータフレームを結合し、新たなデータフレームに挿入する。そのデータフレームにさらに結合する、という作業を繰り返すことになる。

```
> D <- merge(y, x1, by = "年")
```

図 2.5 結合コード

まずは y (自殺者数)、 x_1 (倒産件数)を結合する。2 つのデータフレームを結合する場合は関数 `merge()` を用いる。この関数は 2 つのデータフレームの両方にある列データ(今回は”年”) で紐付けをしてデータを結合するものである。このときに、引数 `all` に何も指定しない場合は キーが共通しているデータのみを残して結合、全てのデータを結合したい場合は `T` を引数として用いる。引数を「`by = ” 2 つのデータフレームを紐づけする列名 ”`」とすることで列名を指定することができる。

表 2.4 データフレーム D に結合されたデータ

年	自殺者数	倒産件数
1980	21048	17884
1981	20434	17610

2012	27858	12124
2013	27283	10855

図 2.5 のコードにより y (自殺者数)、 x_1 (倒産件数)のそれぞれに共通していた「年」をキーとしてデータフレーム D に結合することができた。次は先程の D と x_2 (失業率)を結合する。

年	自殺者数	倒産件数	失業率
1980	21048	17884	3.213023
1981	20434	17610	3.365118
~~~~~			
2012	27858	12124	6.465649
2013	27283	10855	5.338225

$D_1$  に失業率が追加されたのがわかる。同じように今度はデータフレーム  $D_1$  に  $x_3$ (消費者物価指数)を結合していく。

```
> D2 <- merge (D1 ,x3, by = "年")
```

図 2.7 結合コード

表 2.6 データフレーム  $D_2$  に結合されたデータ

年	自殺者数	倒産件数	失業率	消費者物価指数
1980	21048	17884	3.213023	77
1981	20434	17610	3.365118	81
~~~~~				
2012	27858	12124	6.465649	100
2013	27283	10855	5.338225	100

これですべてのデータを一つのデータフレームに挿入することができた。

2.2.3 規準化

2項で結合を行ったため、データフレーム D_2 には 1980 年から 2013 年までの自殺者数、倒産件数、失業率、消費者物価指数が入力されている。しかし、それぞれのデータは単位が異なっている。失業率は単位がパーセントだが、自殺者は人、といったようにこのままでは正しい分析結果が得られない。そのため規準化し、分析を行うことができる状態にする必要がある。

規準化とは観測地の平均値と標準偏差をある値に揃えることである。一般的には平均値を0として、標準偏差を1にするものが多く用いられる。〔12〕

自殺者数、倒産件数、消費者物価指数、実業率は x_i 、平均値を \bar{x} 、標準偏差を σ と表した場合、規準化された z は以下のようになる。

$$z = \frac{x_i - \bar{x}}{\sigma}$$

図 2.8 規準化式

まずは説明変数である倒産件数を規準化するため、下記のコードをRに打ち込む。

```
z1 <- (x1$倒産件数 - mean(x1$倒産件数)) / sd(x1$倒産件数)
```

図 2.9 規準化 R コード

`mean()` で `()` 内の変数の平均を算出する。また、データフレーム内の変数は、「データフレーム名 `(x1)` \$変数名(倒産件数)」で指定することができる。後半の「`sd()`」は `()` 内の変数について、標準偏差を算出する関数である。Rにおいて加減乗除はそれぞれ `+`、`-`、`*`、`/` を用いるため関数 `mean` の前に `-`、関数 `sd` の前に `/` を打ち込む。その結果をデータフレーム z_1 に挿入する。実際にこの平均が 0、標準偏差が 1 となっているか確認する。

```
> mean(z1)
[1] 1.328945e-16
> sd(z1)
[1] 1
```

最初に処理した $\text{mean}(z_1)$ が平均、 $\text{sd}(z_1)$ は標準偏差となる。平均の結果が $1.328945e-16$ と 0 になっていないように見えるが、 e^{-16} の e とは指数であり、この前についている数字 $\times 10^{-16}$ のことを指す。計算式で表すと下のようになる。

```
1.328945 * 0.000000000000000001 = 0.0000000000000000001328945
```

このように限りなく 0 に近くなっており、正確に処理されていることがわかる。図 2.9 のコードで倒産件数の数値が規準化された。

表 2.6 z_1 に挿入されているデータ

```
> z1
[1] 0.86937953 0.78924239 0.64651640 1.24111055 1.73421715
[31] -0.46516699 -0.63684764 -0.81525514 -1.18640121
```

ここまでは倒産件数を平均と標準偏差を用い基準化を行った。残りの自殺者数、失業率、消費者物価指数も同じように規準化を行っていく。変数の数が多くなればなるほど、毎回規準化を行っていると時間がかかってしまうことになる。そこで複数のデータフレーム内の変数をまとめて規準化することができる関数 `scale` を用いる。

<code>scale()</code>	()内のデータフレームを基準化し結果を返す関数
----------------------	--------------------------

この関数を用いて y (自殺者数)、 x_1 (倒産件数)、 x_2 (失業率)、 x_3 (消費者物価指数)をまとめて規準化を行っていく。2.2.2 でそれぞれの変数を一つのデータフレーム (D_2) にまとめた。年数は必要ないのでそれ以外を選択する。

```
D3 <- scale(D2[, 2:5])
```

図 2.10 規準化コード

表 2.7 データフレーム D_3 内の規準化されたデータ

	自殺者数	倒産件数	失業率	消費者物価指数
[1,]	-1.23078895	0.86937953	-1.0480938	-2.6433365
[2,]	-1.35747121	0.78924239	-0.9479781	-2.1002945
[3,]	-1.19365083	0.64651640	-0.8947882	-1.8287736
~~~~~				
[32,]	0.75052991	-0.63684764	1.1077880	0.4791546
[33,]	0.17427003	-0.81525514	1.0929316	0.4791546
[34,]	0.05563436	-1.18640121	0.3508106	0.4791546

### 3 変数間の相関関係

変数  $\alpha$  の数値が高くなれば高くなるほど、変数  $\beta$  の数値も高くなるといった、量的な2つ以上の変数関係を「相関関係」と言う。これに対して質的な2つ以上の変数関係を「連関」という。相関関係の評価判断表が以下になる。

表 3.1 相関関係判断表

評価	相関係数	
ほぼ相関なし	$-0.2 \leq r \leq 0.2$	
弱い相関あり	$-0.4 \leq r < 0.2$	$0.2 < r \leq 0.4$
中程度相関あり	$-0.7 \leq r < 0.4$	$0.2 < r \leq 0.7$
強い相関あり	$-1.0 \leq r < -0.7$	$0.2 < r \leq 1.0$

2.2.3 で  $D_3$  には自殺者数、倒産件数、失業率、消費者物価指数が基準化されたデータが入っている。このデータから相関係数を算出する。相関係数は以下の式で求めることができる。

$$\text{相関係数} = \frac{\text{共分散}}{\text{変数}\alpha\text{の標準偏差} \times \text{変数}\beta\text{の標準偏差}}$$

図 3.1 相関係数の算出計算式

ここで出てくる共分散とは2種類のデータ関係を示す指標である。これは2つの変数の偏差の積の平均で求めることができる。この相関係数の算出式だが R では `cor()` で求めることができる。

<code>cor()</code>	2つ以上の変数の相関係数を算出する関数
--------------------	---------------------

```
> cor(D3)
```

	自殺者数	倒産件数	失業率	消費者物価指数
自殺者数	1.0000000	0.18169922	0.87361797	0.5958303
倒産件数	0.1816992	1.0000000	0.09065834	-0.1711914
失業率	0.8736180	0.09065834	1.0000000	0.5149335
消費者物価指数	0.5958303	-0.17119144	0.51493349	1.0000000

図 3.2  $D_3$ 内データの相関係数

図 3.2 相関係数の結果を表 3.1 の相関関係判断表に当てはめる。自殺者数と倒産件数の相関係数は 0.18169922 とほとんど相関なしに判断される。一方失業率は 0.87361797 と強い相関があることがわかる。消費者物価指数は 0.5958303 で中程度相関があると判断される。説明変数同士で相関が高かったものは失業率と倒産件数の 0.09065834 であった。次に図 3.2 内を pairs 関数を用い散布図を作成する。

pairs()	( )内のデータの散布図を作成することができる関数
---------	---------------------------

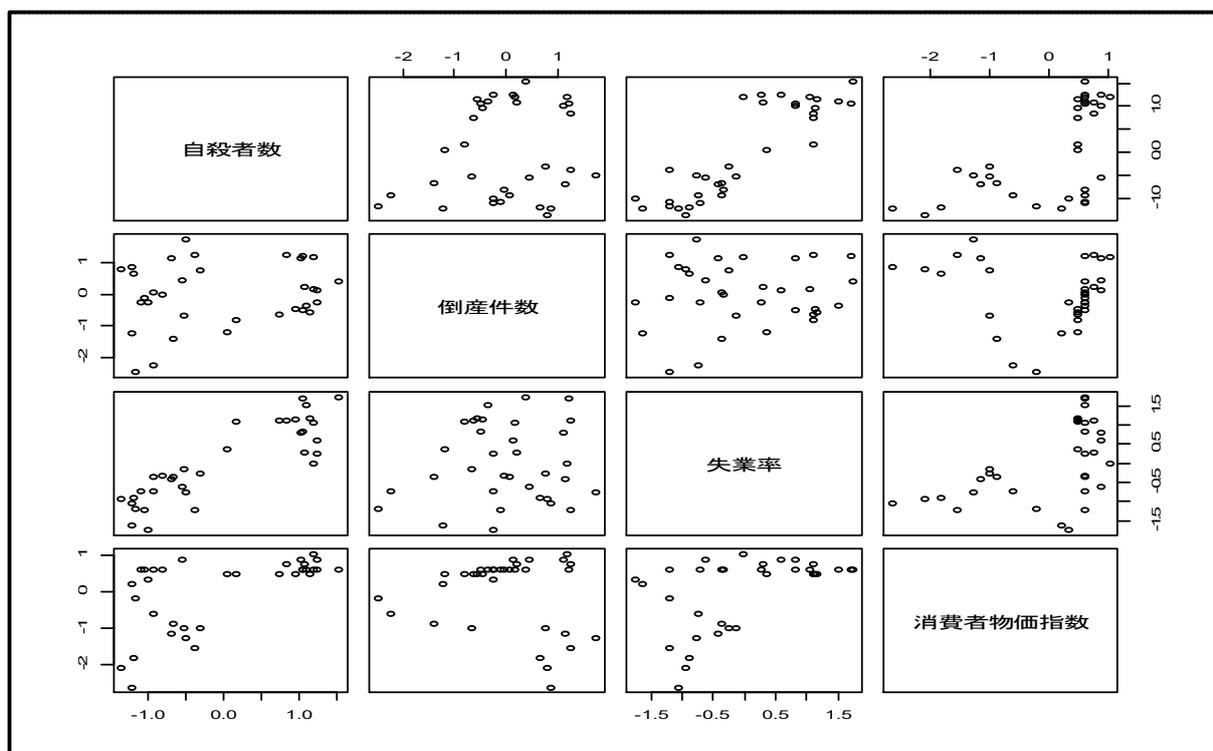


図 3.3 相関関係判断表

散布図を作成すると 2 変数間の関係が直接的なものかどうか分かる。散布図のある程度のもつた形が右肩上がりのものを「正の相関」、右肩下がりのものを「負の相関」という。グラフにまとまりがなく直線的な関係が見られないものを「無相関」という。この散布図からわかることはあくまで大まかな相関関係である。次の章では重回帰分析を用い分析を行っていく。

## 4 重回帰分析

### 4.1 すべての説明変数を用いた重回帰分析

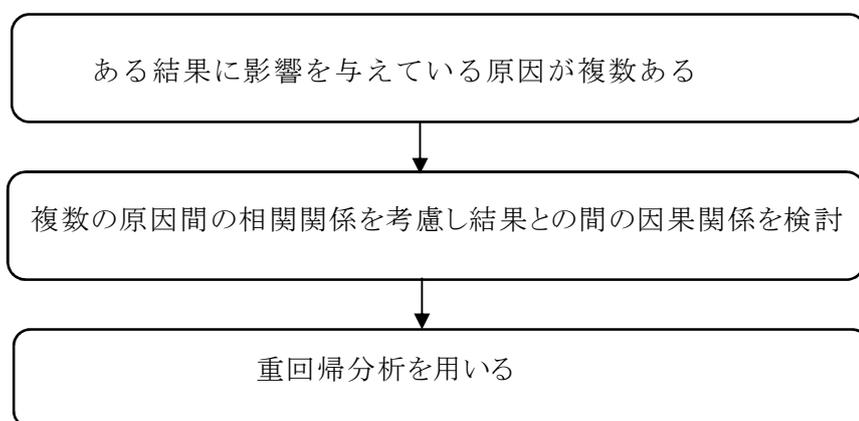


図 4.1.1 重回帰フローチャート

重回帰分析とは一つの被説明変数を複数の説明変数で説明したい際に用いる統計分析手法の一つ。本研究において被説明変数は自殺者、説明変数は失業率、倒産件数、消費者物価指数の 3 つである。

$$Y = b_0 + b_1 X_1 + b_2 X_2 \dots + b_p X_p$$

$$\text{自殺者数} = b_1 * \text{失業率} + b_2 * \text{倒産件数} + b_3 * \text{消費者物価指数}$$

図 4.1.2 重回帰式モデル

上記重回帰式モデルの各説明変数前についている  $b_1$ 、 $b_2$ ...を偏回帰係数という。 $b_0$ は切片である。偏回帰係数はどの説明変数が被説明変数に影響を及ぼしているかの直接的なものではない。例えば説明変数の測量単位がグラムからキログラムになればその偏回帰係数は 1 / 1 0 0 0 にな

る。また測量単位がグラムの説明変数とセンチメートルの説明変数では比較ができない。このような場合平均値= 0, 分散= 1 に規準化ときの偏回帰係数を求めて使用する。これが標準化偏回帰係数である[13]。

```
summary(lm(D3[,c(2)]~D3[,c(1)]+D3[,c(3)]+D3[,c(4)]))
```

図 4.1.3 Rコード

上記のコードを打ち込み重回帰分析を行う。lm()で()内の回帰分析を行う。

```
lm(x~a,b,c...)
```

xの部分には目的変数を挿入する。2.2.3、表2.7で規準化したデータから目的変数である2行目の自殺者数を選択するため、D3[,c(2)]としている。~の後のa、b、c…には説明変数を挿入する。先ほどの自殺者数と同じくD3から1、3、4行目(それぞれ倒産件数、失業率、消費者物価指数)をD3[,c(1)]+D3[,c(3)]+D3[,c(4)]と打ち込んでいる。summary()で()内の分析結果を表示することができる。以下が出力された結果である。

表 2.8 重回帰分析結果

Call:				
lm(formula = D4[,c(2)]~D4[,c(1)]+D4[,c(3)]+D4[,c(4)])				
Residuals:				
Min	1 Q	Median	3 Q	Max
-0.83863	-0.28330	0.03545	0.19932	0.95280
Coefficients:				
	Estimate	Std. Error	Std. Error	Pr(> t )
(Intercept)	-4.900e-16	7.734e-02	0.000	1.0000
D4[,c(2)]	1.572e-01	8.153e-02	1.929	0.0633
D4[,c(3)]	7.331e-01	9.370e-02	7.824	9.91e-09***
D4[,c(4)]	2.453e-01	9.471e-02	2.590	0.0147*

```

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

Residual standard error: 0.451 on 30 degrees of freedom

Multiple R-squared: 0.8151, Adjusted R-squared: 0.7966

F-statistic: 44.09 on 3 and 30 DF, p-value: 4.101e-11
    
```

分析結果が長いので上から順に見ていく。一行目の `Call : lm(formula = D4[,c(2)] ~ D4[,c(1)] + D4[,c(3)] + D4[,c(4)])` はそのまま入力したプログラムコードの答えを呼び出すという意味になる。3行目以降の

```

Residuals:

    Min       1 Q   Median       3 Q      Max
-0.83863  -0.28330   0.03545   0.19932   0.95280
    
```

は残差の四分位数を表す。残差とは重回帰分析においての実測値-予測値で定義される[14]。実測値は `lm` 関数で求めた値、予測値は元々のデータの値を指す。また、データを小さい順に並べて、最小値から  $1/4$  のところのデータを第1四分位数、 $2/4$  のところのデータを第2四分位数(中央値)、 $3/4$  のところのデータを第3四分位数という。そして、第1四分位数、第2四分位数、第3四分位数をまとめて、四分位数という[15]。`Min` は残差の最小値、`1 Q` は第1四分位数、`Median` は第2四分位数(中央値)、`3 Q` は第3四分位数、`Max` は最大値をそれぞれ表している。

Coefficients:				
	Estimate	Std.Error	t value	Pr(> t )
(Intercept)	-4.900e-16	7.734e-02	0.000	1.0000
倒産件数	1.572e-01	8.153e-02	1.929	0.0633
失業率	7.331e-01	9.370e-02	7.824	9.91e-09***
消費者物価指数	2.453e-01	9.471e-02	2.590	0.0147*

次に Coefficients を見ていく。Intercept の行には定数項が入っている。定数項とは y 切片とも呼ばれ、説明変数の変動に影響されない値のことを指す。それぞれの列については以下の通り。

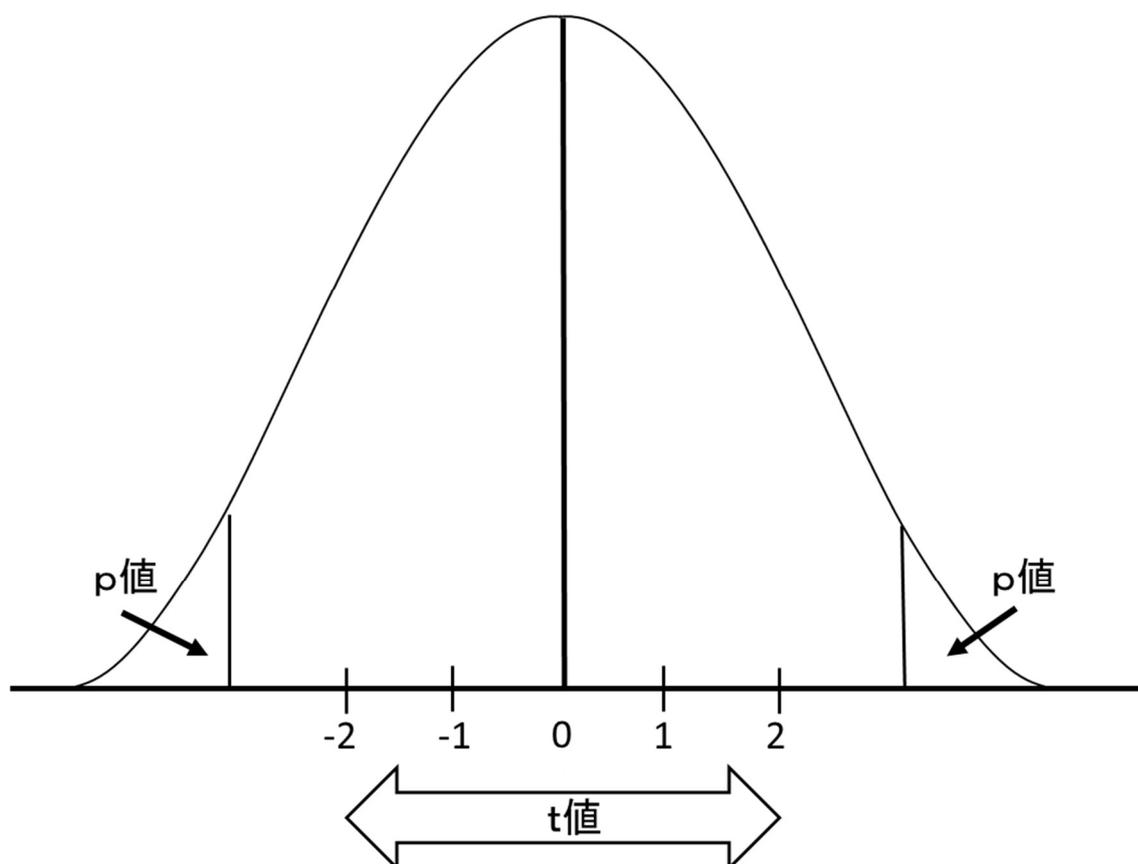


図 4.1.4 t値p値グラフ

Estimate	回帰係数のことを表す。回帰係数とは回帰分析に予測値を算出する $y = a + bx$ の式で表される回帰直線の傾き $b$ の値である。
Std. Error	標準誤差のことを表す。標準誤差とは母集団からある一定数の標本を選ぶ際、その選ぶ組み合わせによってどの程度の誤差が出るかを標準偏差で表したものを指す。
t value	$t$ 値のことを指す。この $t$ 値は説明変数の係数、定数項の確実性を判断する際に用いる数値である。 $t$ 値の絶対数が大きければ大きいほどその有意性が高いと判断することができる。一般的には $t$ 値が $-2$ 以下、 $+2$ 以上であれば有意であると判断される。
Pr ( >   t   )	$p$ 値のことを指す。この $p$ 値は説明変数の係数や定数項が偶然その値を示した確率のことを表している。ある説明変数の $p$ 値が 20 パーセント以下であった場合、その説明変数は 20 パーセント以下の確率で偶然その値になったことを示している。一般的に $p$ 値が 5 パーセント以下であれば有意性であると判断される。

回帰分析においては  $y = a + bx$  において  $b = 0$  となることはあってはならない。これは  $b = 0$  になってしまえば、 $y = a$  となり、 $x$  は  $y$  に影響を与えない変数になってしまうからである。回帰分析において  $b = 0$  となることを帰無仮説と呼ぶ。 $b = 0$  となることを棄却できるかどうか考える際に  $p$  値を用いる。 $p$  値は帰無仮説が発生する確率でもあるので、上記のように 5 パーセント未満 (0.05 未満) であれば有意性があると判断できる。このとき、統計量  $t$  の偏りを  $t$  分布の両側とも考慮するのか、片側だけ考慮するのかによって、統計量  $t$  の有意となる位置が変わる。例えば、 $A$  と  $B$  のサンプルの母集団の平均値を比較する 2 サンプル  $t$  検定では、両側検定において、以下のような仮説を立てる。

帰無仮説は  $A$  と  $B$  の母集団の平均値は等しい。対立仮説は  $A$  と  $B$  の母集団の平均値は等しくない。一方、片側検定の場合、以下のような仮説を立てる。帰無仮説は  $A$  と  $B$  の母集団の平均値は等しい。対立仮説は  $A$  よりも  $B$  の母集団の平均値のほうが大きい (または、小さい)

寄与率はデータの構成要素となる項目の変化を100とした場合にデータ全体にどの程度の影響を与えているかを示す指標である。寄与率は1に近ければ近いほど予測値と観測値が近くなることを表し、モデルの正確性を表す。

$$\text{寄与率} = \frac{\text{構成項目の増減値}}{\text{データ全体の値}} \times 100$$

図 4.1.5 寄与率式

寄与率の欠点として、説明変数の数が増加していくほど1に近づいていく。つまり、実際はあまり相関関係がない説明変数を数多く追加していけば1にどんどん近づき、誤った結果になってしまう。そこで説明変数の数量の影響を取り除いた自由度調整済み寄与率を用いる。モデルの正確性を表すこの数値の明確な基準は存在しないが、「0.5以上」を1つの目安として扱う。説明変数の数をp、標本数をn、寄与率をR、自由度調整済み寄与率を $\bar{R}$ と表すと $\bar{R}$ を求める式は以下のようになる。

$$\bar{R} = 1 - (1 - R) \frac{n-1}{n-p-1} = \frac{n-1}{n-p-1} R - \frac{n-1}{n-p-1}$$

図 4.1.6 自由調整度寄与率式

Coefficients:				
	回帰係数	標準誤差	t値	p値
(Intercept)	-4.900e-16	7.734e-02	0.000	1.0000
倒産件数	1.572e-01	8.153e-02	1.929	0.0633
失業率	7.331e-01	9.370e-02	7.824	9.91e-09 ***
消費者物価指数	2.453e-01	9.471e-02	2.590	0.0147 *
Residual standard error: 0.451 on 30 degrees of freedom				
Multiple R-squared: 0.8151, Adjusted R-squared: 0.7966				
F-statistic: 44.09 on 3 and 30 DF, p-value: 4.101e-11				

Residual standard error	残差の標準誤差のことを表す。
Multiple R-squared	寄与率(決定係数)のことを表す。
Adjusted R-squared	自由度調整済み寄与率(調整済み決定係数)
F-statistic	F 統計量のことを表す。

重回帰分析の結果、倒産件数の $p$ 値は $0.0633$ 、失業者数は $9.91e-09$ 、消費者物価指数は $0.0147$ となった。有意性は5パーセント未満で診断できることから有意性があるのは失業者数と消費者物価指数ということがわかる。有意性の強い順に並べると失業者数>消費者物価指数>倒産件数となる。

$$y = -4.900e-16 + (1.572e-01)x_1 + (7.331e-01)x_2 + (2.453e-01)x_3$$

#### 図 4.1.7 重回帰式

次に予測値と実測値の差を調べる。まずは予測値を調べるため predict 関数を用い、 $y.y$  に図 4.7 で求めた回帰式を挿入する。

```
y.y <- predict(lm(D4[,c(2)] ~ D4[,c(1)] + D4[,c(3)] + D4[,c(4)]))
```

次に $y_$ に規準化された自殺者数のデータを入れる。更にデータフレームを作成し最後に $yy$ に完成したデータを挿入し、グラフに表示させる。

```
y_ <- D4[,1]
y_ <- data.frame(y=y_)
yy <- data.frame(y_, "y-hat"=y.y)
plot(yy)
```

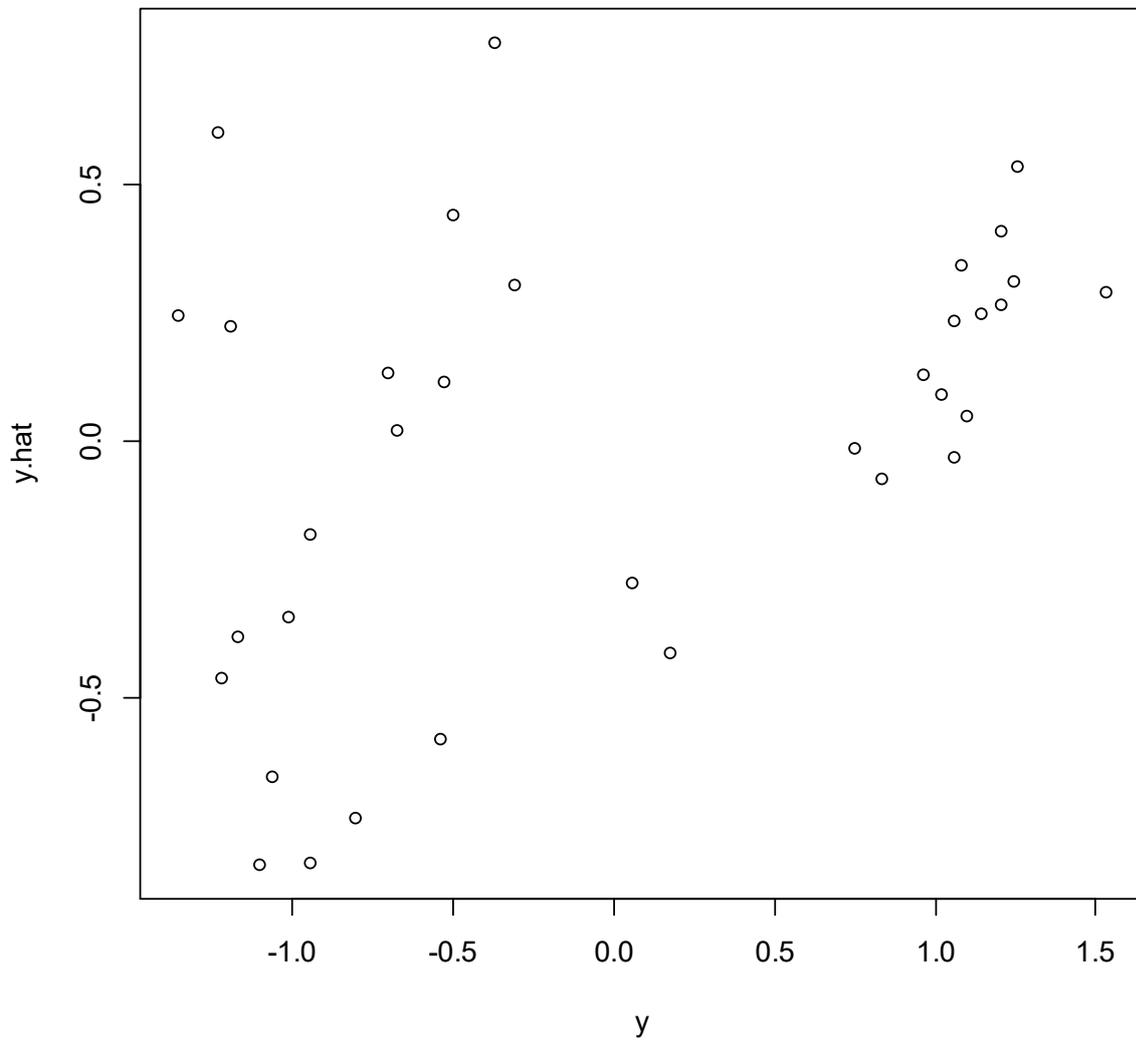


図 4.1.8 実測値、予測値グラフ

## 4.2 優位性のある説明変数のみの重回帰分析

図 4.8 で実測値と予測値のグラフではばらつきが多かったので、優位性がないと判断された倒産件数を除いて、もう一度重回帰分析を行っていく。4.1 で優位性があると判断された説明変数は失業率と消費者物価指数のためこの 2 つの説明変数を用いて行っていく。同じ流れの所は説明を省いて行っていく。まずは各データをそれぞれの変数に挿入する。

```
y <- read.csv("y.deth.csv", skip = 2)
y1 <- y[, c(1, 2)]
z <- read.csv("job.csv")
z1 <- z[, c(1, 3)]
c <- read.csv("bukka.csv", skip = 5)
c1 <- c[, c(1, 2)]
```

図 4.2.1 R 変数読み込みコード

$y_t$ (自殺者数)、 $z_t$ (失業率)、 $c_t$ (消費者物価指数)とそれぞれ挿入した。次にそれぞれを結合し規準化を行い、重回帰分析を行う。

```
D <- merge(y1, z1, by = "年")
D1 <- merge(D, c1, by = "年")
D2 <- D1[, c(2:4)]
D3 <- scale(D2)
summary(lm(D3[, c(1)] ~ D3[, c(2)] + D3[, c(3)]))
Call:
lm(formula = D3[, c(1)] ~ D3[, c(2)] + D3[, c(3)])
Residuals:
    Min       1Q   Median       3Q      Max
-0.7872  -0.2281  -0.1426   0.2533   1.0230
```

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-4.625e-16	8.066e-02	0.000	1.0000
失業率	7.713e-01	9.551e-02	8.076	4.04e-09***
消費者物価指数	1.986e-01	9.551e-02	2.080	0.0459*

---

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

Residual standard error: 0.4703 on 31 degrees of freedom

Multiple R-squared: 0.7922, Adjusted R-squared: 0.7788

F-statistic: 59.09 on 2 and 31 DF, p-value: 2.65e-11

図 4.2.2 R 重回帰分析コード

最後に実数値と観測値のグラフを作成する。

```

y.y <- predict(lm(D3[,c(1)] ~ D3[,c(2)] + D3[,c(3)]))

y_ <- D3[,1]

y_ <- data.frame(y=y_)

yy <- data.frame(y_, "y-hat"=y.y)

plot(yy)

```

図 4.2.3 R 重回帰分析コード

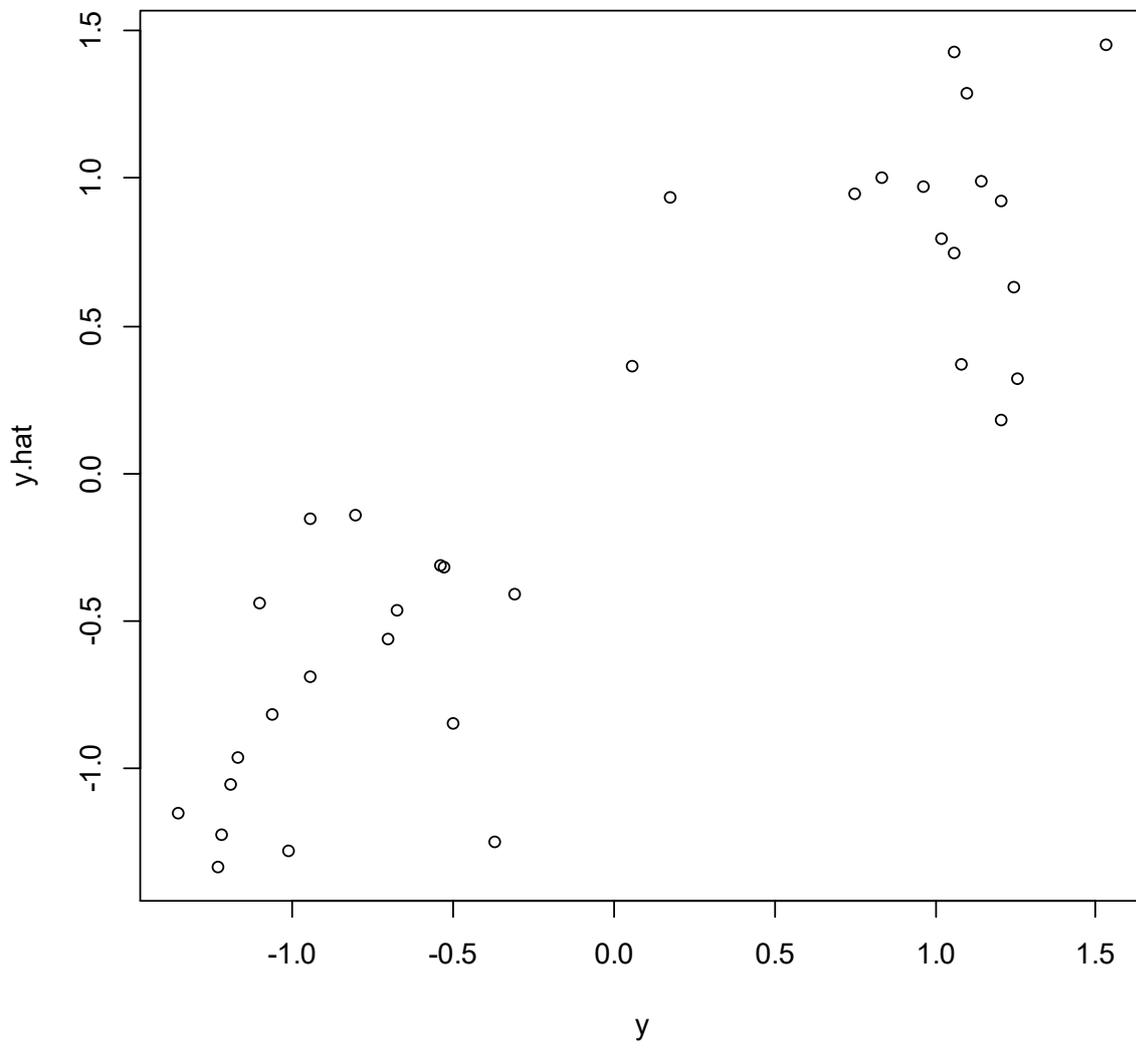


図 4.2.3 実測値、観測値グラフ

5 考察

図 4.2.3 実測値観測地グラフではほかの数値よりも少し離れた位置にあるデータを見ていく。

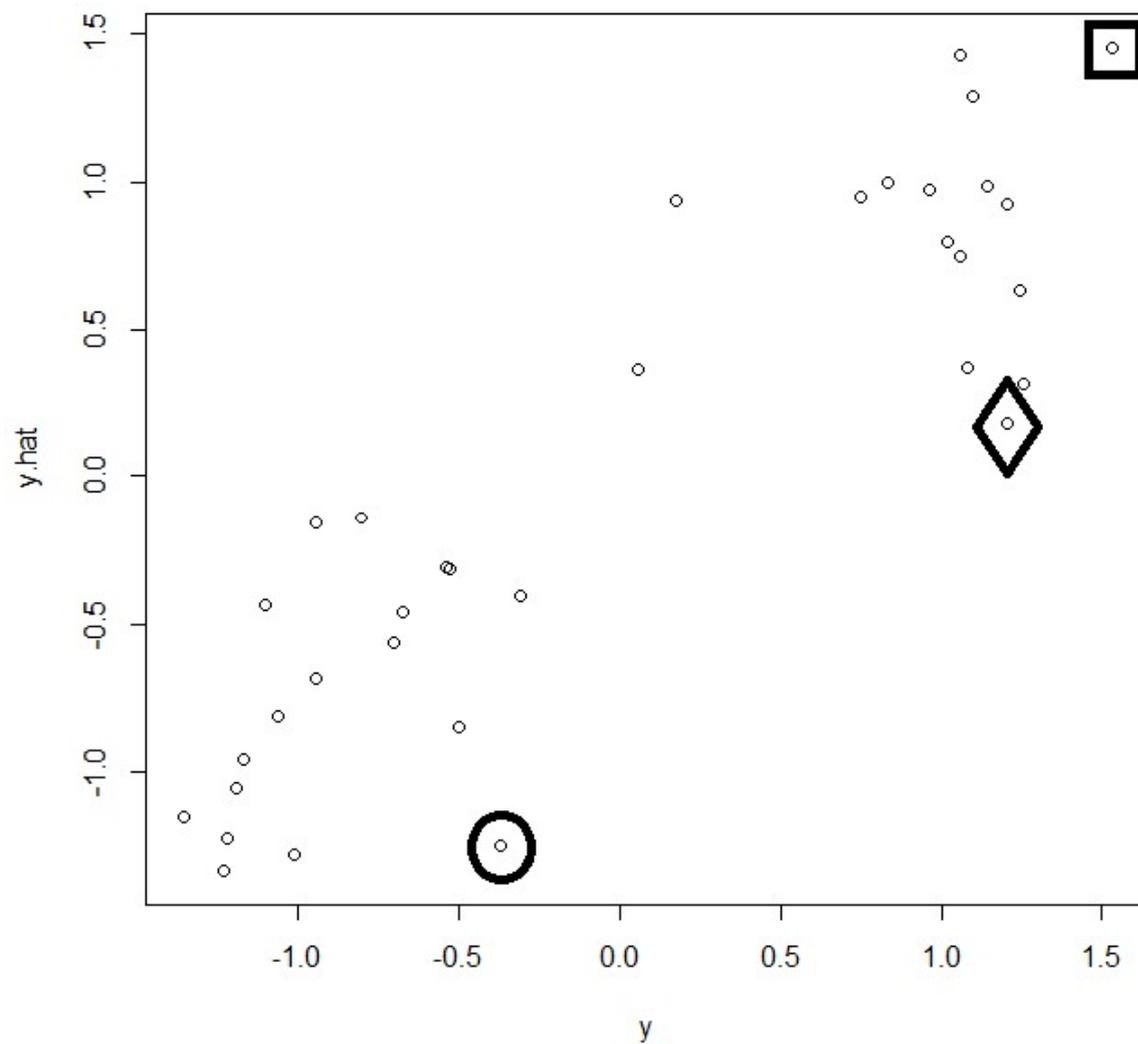


図 5.1 実測値、観測値グラフ

丸、ひし形、四角とデータを囲った。それぞれ 1983 年、1998 年、2003 年の自殺者数データを表している。それぞれに共通する点があった。



自殺者数に相関がある説明変数は失業率＞消費者物価指数＞倒産件数という結果であった。本研究で自殺者数は失業率が高くなれば多くなる傾向があることが分かった。職を失うことは多くの人々にとってとても重い事態であり、大きなストレス、将来の不安などから自殺者が増えるのではないかと考える。

## 6 終章

### 6.1 これからの展望

今回の研究では重回帰分析を用いて自殺者数と説明変数の相関関係について分析を行ってきた。自殺者数といくつかの要因との相関関係を定量化し、約 8 割の信頼がある結論を出せたことが成果である。

今回、説明変数は倒産件、消費者物価指数、失業率の 3 つのみであった。これらの説明変数以外にも様々な要因があると考えられる。これからの課題としてそれらを加え、説明変数の数を吟味し分析の精度をあげたい。また図 4.2.3 で右上と左下で 2 つまとまりができた。これらがどのようなまとまりであるかを解析していくことが今後の展望となる。

## 謝辞

本研究を進めるに際してお世話になりました田中章司郎教授に感謝します。

また本論文、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] 国の自殺率順リスト <https://ja.wikipedia.org/wiki/>
- [2] 自殺対策基本法及び自殺総合対策大綱  
[http://www8.cao.go.jp/jisatsutaisaku/whitepaper/w012/html/honpen/part2/s2_1_2.html](http://www8.cao.go.jp/jisatsutaisaku/whitepaper/w012/html/honpen/part2/s2_1_2.html)
- [3] 警察庁 <https://www.npa.go.jp/>
- [4] 内閣府による統計  
[http://www8.cao.go.jp/jisatsutaisaku/whitepaper/w2014/html/chapter1/chapter1_01_01.html](http://www8.cao.go.jp/jisatsutaisaku/whitepaper/w2014/html/chapter1/chapter1_01_01.html)
- [5] 三好、真田,1999
- [6] 田中伸和,1998
- [7] 黄 誠金,2012
- [8] 自殺者数  
[http://www8.cao.go.jp/jisatsutaisaku/whitepaper/w2014/html/chapter1/chapter1_01_01.html](http://www8.cao.go.jp/jisatsutaisaku/whitepaper/w2014/html/chapter1/chapter1_01_01.html)
- [9] 倒産件数  
[http://www.chusho.meti.go.jp/pamflet/hakusyo/H26/h26/html/b1_1_2_5.html](http://www.chusho.meti.go.jp/pamflet/hakusyo/H26/h26/html/b1_1_2_5.html)
- [10] 消費者物価指数  
<http://www.e-stat.go.jp/SG1/estat/List.do?bid=000001033702&cycode=0>
- [11] 失業率 [http://d.hatena.ne.jp/kuma_asset/20150130/1422641351](http://d.hatena.ne.jp/kuma_asset/20150130/1422641351)  
<http://lang.sist.chukyo-u.ac.jp/classes/R/Rstatistics-01.html>
- [12] ファイルからデータを読み込む <http://cse.naro.affrc.go.jp/takezawa/r-tips/r/40.html>
- [13] 基準化 <http://www.otoiwase-portal.jp/words/2013/06/post-264.html>
- [14] データの結合 (マージ) と整列 (ソート)  
<http://cse.naro.affrc.go.jp/takezawa/r-tips/r/43.html>

[15] <https://ja.wikipedia.org/wiki/%E6%8C%87%E6%95%B0%E8%A1%A8%E8%A8%98>

[16] 重回帰分析

<https://ja.wikipedia.org/wiki/%E9%87%8D%E5%9B%9E%E5%B8%B0%E5%88%86%E6%9E%90>

(17) 総務省統計局 <http://www.stat.go.jp/teacher/c2dic.htm>

(18) 残差

<http://www.weblio.jp/content/%E6%A8%99%E6%BA%96%E5%8C%96%E6%AE%8B%E5%B7%AE>