ISSN1342-0860

経 営 研 究

福山平成大学経営学部紀要
第１３号２０１７年３月

目　次

＜論文＞

周京伝　……市瀬 信子　1
－「周穆門墓志銘」を中心に－

情報リテラシー初歩教育段階のタイピング修得練習における
タッチタイピングの早期修得による効果について ……奥田 由紀恵　21

HTML5 および Javascript による幾何アニメーション表示　……尾崎 誠　33

パートタイマーの基幹労働力化が組織市民行動に及ぼす影響 ……小玉 一樹　43

広島市役所のマネジメントの改善に関する研究 …… 佐藤 幹　59
－幹部職員等のインタビュー調査結果に基づいて－

社会システム分析のための統合化プログラム２８ …福井 正康・小玉 一樹・尾崎 誠　69
－メタ分析・ロジスティック回帰分析－

社会システム分析のための統合化プログラム２９ …福井 正康・大山 知之・織田 望　99
－パラメータ設計・オンライン品質工学－

社会システム分析のための統合化プログラム３０ …福井 正康・大山 知之・織田 望 121
－異常検知－

社会システム分析のための統合化プログラム３１ …… 福井 正康・呉 暁娜 139
－生存時間分析－

フィリピン小学校英語教科書の分析 ……渡辺 清美 163
－１、２年教科書の量的分析－

2016 年度研究業績一覧 …… 176

福 山 平 成 大 学

福山平成大学経営学部紀要
第 13 号（2017），1－20 頁

周京伝
－「周穆門墓志銘」を中心に－

市瀬 信子

福山平成大学経営学部経営学科

要旨：清代の杭州詩壇は、２度の隆盛期を迎える。第１の隆盛期は、厲鶚を中心に「南宋雑事詩」が作成された康熙時代の終わりから雍正時代の始めにかけての時期である。第２の隆盛期は、それまで杭州の外に出ていた詩人たちが、杭州に再び集結し、詩社を結成して盛んに詩会を開いた、乾隆８年から 10 年余りの期間である。この時期詩会のリーダー格とされたのが、周京という詩人である。周京は、杭州の第一の隆盛期に、すでに厲鶚ら若者のリーダーとして、詩会を引っ張っていた。豪快な詩風と人柄が、多くの若者を惹きつけたのだ。その後杭州を離れた周京は、長年地方を渡り歩き、晩年になって杭州に帰り、再び詩会のリーダーとなる。その時杭州詩壇は最盛期を迎えたとされる。しかし、当時の杭州での知名度にもかかわらず、後世周京の名はほとんど知られていない。そこで、本稿では、周京に関する墓志銘などの伝記資料を中心に、周京の人生を明らかにし、またなぜ彼が杭州詩壇の領袖として愛されたのかを明らかにしたい。

キーワード：周京、墓志銘、杭州

はじめに

周京（1677-1749）は、康熙年間から乾隆年間にかけて生きた杭州の詩人であり、杭州詩会の隆盛期にその中心となって活躍し、人気を博したとされる。

清代の杭州詩壇の隆盛は大きく２期に分けることができる。第１期は、厲鶚を始めとする７人の杭州詩人によって『南宋雑事詩』が制作された雍正元年（1723）から雍正二年にかけての期間である。『南宋雑事詩』の時期を境に、厲鶚、符曽を始め、多くの杭州詩人が故郷を離れ、移動する詩人集団として、杭州以外の地で活動した。第２の隆盛期は、その 20 年後、杭州の外で活躍していた詩人達が、再び杭州に集った乾隆年間初期である。杭世駿、厲鶚を始めとし、杭州を離れて学者、芸術家、詩人として名をなした名士達が次々に杭州に戻り、地元の耆旧、方外たちと詩社を結んで盛んに詩会を開いたのだ。その様子は「乾隆初、杭州詩酒之会最盛。（乾隆の初め、杭州詩酒の会最も盛ん。）」（袁枚『随園詩話』巻三-六四）と称される。この隆盛期は、「吾郷詩社、自癸亥以後、称最盛者十年。（吾が郷の詩社、癸亥より以後、最盛と称すること十年。）」（『両浙輶軒録』巻十九「梁啓心」）と

あるように、癸亥つまり乾隆八年（1743）から約 10 年間であった。周京は、第 1 期の隆盛期には『南宋雑事詩』には、加わらなかったものの、作者たちの詩友として尊敬を集め、第 2 期の隆盛期には詩会の領袖となって活躍した。２つの隆盛期のいずれにおいても詩壇の中心にいたのである。しかし２つの隆盛期の間の約 20 年間、周京はほとんど故郷に帰らず旅に出ていた。杭州を離れていた期間が長かったにもかかわらず、なぜ杭州詩壇の領袖として迎えられたのであろうか。そこで、周京がいかなる人生を送り、なぜ詩壇の領袖たりえたかを、改めて考察する必要がある。

周京は生涯無官に終わり、また他の杭州詩人のように地方都市の塩商の庇護を受けることもなく、一人僻遠の地を旅していたため、伝記となる資料は少ない。全祖望「周穆門墓志銘」（『鮚埼亭集』内篇巻十九）（以下、全祖望「墓志銘」と表記する。）は、周京を知るための、数少ない資料の一つである。これは『無悔斎集』付録「徵士穆門周先生墓志銘」に加筆したものである。今回は情報量の多い、『鮚埼亭集』の「周穆門墓志銘」をもとに、『無悔斎集』付録「徵士穆門周先生墓志銘」と照らし合わせつつ、考察を進めることとする。[1] 他にも墓志銘としては、桑調元「周徵士穆門墓志銘」（『無悔斎集』附録）[2]（以下、桑調元「墓志銘」と表記する）がある。また全祖望「墓誌銘」に「穆門之卒也、吾友杭堇浦為之伝。（穆門の卒するや、吾が友杭堇浦 之が伝を為る。）」とあるように、杭世駿が作った伝があったはずなのだが、現存の杭世駿『道古堂集』には収録されていない。厲鶚「無悔斎詩集序」は、周京の文学的経歴を記すものだが、文学以外の情報に乏しく、やはり周京の人生全体を見渡すには、全祖望「墓志銘」が最も適切であると考えられる。そして、杭州詩壇を代表する詩人の人生を明らかにすることで、清代杭州の埋もれた文学史の一端を照らすことができると考える。

周京に関する研究を見てみると、杭州を中心とする詩派「浙派」の中心人物として、しばしば取りあげられている。[3] しかし、同時期の詩人である厲鶚、杭世駿ほどには知名度が高くなく、後世への影響も少ないため、研究は決して多くない。その中で、王小恒「論周京的人格価値取向和旅隴詩歌創作」（『甘粛社会科学』2016 年第３期）は周京の甘粛への３度の旅に的をしぼった論考で、「周京交游考」（『河西学院学報』第 32 巻第 4 期　2016）は、その交遊関係についての論考である。また同氏には、杭州詩壇で周京と並び称された顧之珽に関する論考「試顧之珽及其在浙派中的地位」（『甘粛広播電視大学学報』第 26 巻第 2 期　2016）があり、この中で周京、顧之珽両者を中心とした杭州詩人達の交遊の様子と、詩社を支えた二人の役割について論じている。拙論「周京と杭州詩壇」（『中国中世文学研究』第 67 号　2016）は、周京と杭州詩壇の詩風、また詩社の終焉について論じたものである。

１. 全祖望と「周穆門先生墓志銘」

全祖望は寧波の人である。康熙五十九年（1720）16歳の時、郷試のために杭州を訪れた。その後康熙六十一（1722）～雍正元年（1723）頃に、杭州で文人達と交流を始めたとされる。[4] その後も杭州詩人と活動を共にすることが多く、また翰林院庶吉士にもなったエリ

ートであるため、全祖望は杭州詩人の墓志銘を依頼されることが多かった。「周穆門墓志銘」もその１つである。

以下、いくつかの段落にくぎって、全祖望「墓志銘」を読んでゆく。

２．詩名と旅と不遇

〔１〕穆門以詩名天下五十余年、平生嘗徧歴秦、斉、晋、楚之墟、所至、巨公大卿皆為倒屣。顧終于蹭蹬不遇而死。

穆門詩を以て天下に名あること五十余年、平生嘗て秦、斉、晋、楚の墟を徧歴し、至る所、巨公大卿皆為に倒屣（とうし）す。顧（た）だ終（つい）于に蹭蹬として遇せられずして死す。

【通釈】周穆門（京）は世の中に詩で名を知られること50余年、終生秦（陝西）、斉（山東）、晋（山西）、楚（湖北、湖南）各地の遺跡を遍歴していたが、至るところで名士大物達が皆彼の到来を待ちかねていた。ただ最後まで運に恵まれず不遇のまま死んだ。

ここでは長く詩名があったこと、中国各地を旅し人気を得たものの、不遇のまま死んだことが記される。そこで、詩名と旅に分けて、周京の人生を詳しく見てゆくこととする。

（１）詩名

50余年に及ぶ詩名については、『無悔斎集』附録の梁文泓「哭周穆門先生」にも「歎息才名五十年、于今芸苑復誰先（歎息す才名五十年、今に于いて芸苑復た誰をか先にせん）。」と、詩壇での名声が50年に及び、死後は彼に及ぶ者がないと述べている。50年前とは、周京23歳のまだ若い時である。周京の死に際して30人の詩友が周京の為に詠んだ「哭周穆門先生」（『無悔斎集』附録）の中の桑調元「里閈聞声四十年（里閈 声を聞ゆること四十年）」も、杭州詩壇で40年にわたる詩名があったと言っている。若い頃から晩年までずっと詩人としての名声を保ち続けていたのだ。また後輩詩人との交流については、汪台「託契芝蘭四十年（契を芝蘭に託すこと四十年）」、厲鶚「如君行誼友兼師、……僂指交情三十載（君の行誼の如きは友 師を兼ぬ……指を僂る 交情三十載）」と、数十年に渡り詩友で且つ師弟の関係にあったことがわかる。周京は桑調元より18歳、厲鶚より15歳、金農より10歳年上で、若い詩人たちの良き先輩であった。周京と年下の詩人たちの交遊については、厲鶚「無悔斎集序」に、次のように見える。

乙未、丙申間、予輩数人為文字之会、暇即相与賦詩為楽。酒闌灯灺、逸韻横飛、必推周兄穆門為首唱。穆門詩主気格、以豪健為尚、淋漓排奡、往往*一座尽傾。詩成、毎擊節自歌、淵淵乎**声若出金石、予輩亦従而和之。

＊「往往」、＊＊「乎」は『無悔斎集』に無く、『樊榭山房文集』巻三によって補う。

乙未、丙申の間、予が輩数人　文字の会を為り、暇あれば即ち相与に詩を賦して楽しみと為す。酒闌（たけなわ）に灯灺（き）えんとし、逸韻横飛するに、必ず周兄穆門を推して首唱と為す。穆門 詩は気格を主とし、豪健を以て尚しと為し、淋漓として奡を排し、往往一座尽く

傾く。詩成れば、毎に節を撃ちて自ら歌ひ、淵淵乎として声 金石より出づるが若く、予が輩も亦た従ひて之に和す。

「乙未、丙申」は、康熙五十四（1715）、五十五年（1716）で、周京39～40歳、厲鶚24～25歳の時である。「文字之会」とは、いわゆる詩酒の会である。酒宴がたけなわとなり灯りも燃え尽きようとする頃合い、風雅な詩が自在に飛び交う時となると、必ず周京を首唱に推したという。周京の詩は気品があり、豪健を尚び、伸びやかで勢いがあり、しばしばその場にいる人誰もが心を奪われ、詩ができあがると、いつも拍子を取って自ら歌い、厲鶚らが唱和したという。詩会で一座の人々を魅了する詩人であったことをこの序は伝えている。この頃の詩友との交流は厲鶚『樊榭山房集』に多く見え、『南宋雑事詩』参加の詩人呉焯、沈嘉轍との交流も記録に残っている。[5)]『民国杭州府志』（巻一百四十五）「杭世駿」には、「先与同里厲鶚、周京、符曽、施安、陳撰、趙昱、趙信、呉焯、呉城、沈嘉轍、汪沆、呉穎芳、丁敬、張沅為密友近賓、言懐叙歓、各有搆属。（先に同里の厲鶚、周京、符曽、施安、陳撰、趙昱、趙信、呉焯、呉城、沈嘉轍、汪沆、呉穎芳、丁敬、張沅と密友近賓と為り、懐を言ひ歓を叙べ、各おの搆属あり）」とあり、周京が親しく交わった詩人達がいかに多かったかが見て取れる。

厲鶚『樊榭山房集』は、とくに若い頃の周京との唱酬の詩を多く収録しており、康熙五十五年、周京40歳、厲鶚25歳の時の「九里松至西山道中同金寿門、周少穆、王雪子作」（『樊榭山房集』巻一）を皮切りに、周京『無悔斎集』に含まれない唱酬の詩も多く収録しており、二人の交流の様を記録する貴重な詩集である。

周京『無悔斎集』十五巻は、康熙四十七年（1708）年、32 歳の作品から始まり、乾隆十四年（1749）73 歳で亡くなるまでの、ほぼ 40 年間の詩が収録されているが、若き日の杭州の詩会の詩会の記録は少ない。周京の記録を残す『樊榭山房集』は、雍正二年（1734）の後、雍正四年（1736）、雍正十二年（1734）と飛び飛びに周京との唱和を記すが、その後乾隆八年（1743）まで両者の唱酬がない。この空白の期間、厲鶚は杭州を離れて繁華な地方都市揚州の塩商の元で精力的に活動し、一方周京は遠く僻遠の地へと旅をしていた。

（２）遠遊

周京の旅とはどのようなものだったのか。

桑調元は、「南遊閩海、北燕趙、西秦晋、所歴名山大川、墟莽古跡、悲歌慷慨、震宕天地。書法奇逸、卓絶一時、……到処名公卿倒履虚席。（南のかた閩海に遊び、北は燕趙、西は秦晋、歴る所の名山大川、墟莽古跡に、悲歌慷慨し、天地を震宕せしむ。書法奇逸、一時に卓絶す、……到る処名公卿倒履し席を虚しくす。）」（桑調元「墓志銘」）とし、燕趙つまり北京から山西に向かった他、南方の閩つまり福建から浙江の沿海部に及んだことを記している。また、名士たちが周京を歓迎する理由として、詩人としての名声の他、「書法奇逸、卓絶一時」と、書家としての人気を挙げる。金農「哭周穆門先生」（『無悔斎集』付録）にも、「詩編伝北郭、書蹟満東呉。（詩編 北郭に伝はり、書蹟 東呉に満つ。）」と、詩と書が広く評価されていたことが見え、符曽『春鳧小稿』己巳「挽周穆門」注「穆門文章書法為

世所推、何待予言（穆門 文章書法は世の推す所と為る、何ぞ予の言を待たん。）」も、詩書が人気を記していることを伝えるものである。こうした人気によって各地で歓迎されながら旅を続けた、というのである。しかし沈徳潜は少し違う見方をしており、「少穆足跡遠到、欲結契宇内英傑。（少穆足跡遠く到り、宇内の英傑と契りを結ばんと欲す。）」（『清詩別裁集』巻二十九「周京」）と、周京自らが積極的に人脈を求めて旅に出たものとする。

厲鶚「無悔斎詩集序」は、周京が杭州を離れた後に訪れた先を次のように記す。

穆門方且臨易水、上金台。久之無所遇、遂走秦、晋之郊、極乎河、湟関塞而止。

穆門方に且に易水に臨み、金台に上らんとす。之を久しくして遇せらる所無く、遂に秦、晋の郊に走り、河、湟の関塞を極めて止む。

杭州を出て易水を訪れてから金台（北京）に赴き、そこにしばらくいたが、不遇のまま、秦、晋の城外から黄河、湟河の国境の果てまで行ったという。「久之無所遇」とあるからには、北京にまず何らかのつてを求めて行ったのではないだろうか。しかし期待した処遇を得られなかったため、そのまま西北に旅立ったのだろう。

周京は科挙に参加しておらず、生活のためには詩や書の知名度に頼り、支援者を得る必要があったと思われる。周京と同様に西北の僻地に遠遊した金農は、「中歳為汗漫游、徧走斉、魯、燕、趙、秦、晋、楚、粤、卒無所遇而帰、晩寓揚州。（中歳 汗漫の游を為し、徧く斉、魯、燕、趙、秦、晋、楚、粤に走るも、卒に遇せらる所無く帰り、晩に揚州に寓す。）」（『国朝耆献類徴初編』巻四百三十六）と、「周穆門墓志銘」と同じように、遠遊の後、不遇のまま帰ったと記している。「不遇」と記されるのは、なんらかの処遇を求めるための旅がその目的をかなえられなかったからである。周京の友人呉焯の息子で、周京の詩友でもあった呉城は、周京の旅を「衣食東西走（衣食もて東西に走る）」（「哭周穆門先生」五首其三）と、衣食を求めての旅であったと、はっきりと述べている。周京自身の言葉としては、「寄弟是諫用東坡寄子山由得告不赴商州韻」（『無悔斎集』巻一　二首其一）に、「駆飢無術苦相商、直為飢駆雁隔行。孤館乍来宵夢短、荒江翻覚旅愁長。（飢えに駆られ術無く苦だ相商し、直ちに飢えの為に駆られて雁隔行す。孤館乍ち来る宵夢の短き、荒江翻って覚ゆ旅愁の長きを。）」と、飢えのために兄弟離ればなれになって旅に出なければならないと、弟に告げている。この詩に続く「薌渓道中」詩は、江西への旅を詠じたものである。これが康熙四十七年（1708）32歳の時で、『無悔斎集』に残る最初の旅の記録である。

次に『無悔斎集』によって簡単に旅の足跡をたどってみよう。

康熙四十八年（1709）、河南の豫州に遊び、その後山東の兗州から河北の冀州、湖北の荊州も訪れたようだ。（以上『無悔斎集』巻一）。『無悔斎集』は、康熙五十三年（1714）までの作品を収録しておらず、全ての動きが見えるわけではないが、康熙五十三年～康熙六十年（1721）までは杭州にいたことがわかる。康熙六十一年（1722）には山東に赴き、泰安を経由して、河北（燕）に赴く。雍正元年（1723）には北京に赴き、9ヶ月余り滞在する。雍正三年（1725）に南下して故郷に戻る。（以上『無悔斎集』巻二、三）。

雍正四年（1726）には、再び北京方面（河北）に行く。（以上『無悔斎集』巻四）。

そして雍正五年（1727）51歳の時、更に西北に足を延ばし、陝西から甘粛に赴いた。厲鶚が「穆門方且臨易水、上金台。久之無所遇、遂走秦、晋之郊、極乎河、湟関塞而止。」（「無悔斎集序」）と、北京方面で思うような待遇を得られず、そのまま遠く旅だったといっているのは、この時のことである。年内に西安まで戻るが、雍正六年(1728)には再び甘粛方面にゆき、翌年河北から南帰する。（以上『無悔斎集』巻五）。

雍正九年（1731）からは杭州での作が続くが、雍正十二年（1734）再び山西に向かう。山西の沢州には杭州の友人である朱樟が赴任しており、彼の元で唱酬を楽しんだことがわかる。雍正十三年（1735）には再び陝西、甘粛に旅し、風物を詠じた。（以上『無悔斎集』巻六、七）。

その後北京に戻ったのは、「試後南帰」とあるように、博学鴻試が実施された乾隆元年（1736）60歳のことである。乾隆二年（1737）61歳で北京を離れ、乾隆三年（1738）には、杭州に帰る。（以上『無悔斎集』巻八）。

乾隆四年（1739）再び北上し、河北の地を経由して、朱樟が知府を務めていた山西の沢州に行く。朱樟との唱和詩は『無悔斎集』に収録されているが、他に朱樟の『沢州府志』編纂にも携わっており、朱樟「哭周穆門先生」（『無悔斎集』付録）の「分脩劇郡書（劇郡の書を分脩す）」句の自注に「謂脩郡志（郡志を脩むるを謂ふ）」とあり、桑調元「墓志銘」にも、「友人延修沢州府志、陵川邑志。」とある。乾隆六年（1941）68歳の時、南下して故郷に戻った。（以上『無悔斎集』巻八、九）

帰郷後、同じく帰郷した詩人達と杭州で盛んに唱酬を行うこれが杭州詩会の隆盛期となった。以後世を去るまで、江南各地に遊ぶことはあっても、遠遊することはなかった。（以上『無悔斎集』巻十～十五）

甘粛方面への旅は、単につてを求めるためのものではなかったようだ。厲鶚は「天時之明晦、山川之険易、人事之変遷、無不于詩発之。其豪也根于理、其健也閲乎境。（山川の険易、人事の変遷、詩に于いて之を発せざるは無し。其の豪たるや理に根ざし、其の豪たるや境を閲す。）」と、険しい地への旅が周京の詩に豪快な風格をもたらしたとし、こうした自然に向き合い、詩に詠じるために旅を続けたのではないだろうか。旅を経た周京の詩風は「詩主気格、以豪健為尚。（詩は気格を主とし、豪健を以て尚しと為す。）」（「無悔斎詩集序」）と言われ、他の杭州詩人と一線を画するものとなった。桑調元「墓志銘」も「所歴名山大川、墟莽古跡、悲歌慷慨、震宕天地。」といい、自然の荒々しさ、昔の旅人に自分を重ね、また土地の暮らしや風俗を詠ずるなど、旅が独自の詩の風格を育てたといえよう。

３．高潔

〔２〕其人淵然湛然、莫能窺其涯涘、渾淪元気、充積*眉宇、蓋古黄叔度、陳仲弓之流也。士無賢不肖、皆曰、周先生長者。乃其中則有確乎不可抜者、而不以形迹自見。大科之役、姚侍郎三辰薦之、穆門力辞不得、応徴至京、徘徊公車門下数日、称疾卒不就試以帰、莫能測也。已而始服其高。

＊「積」‥『無悔斎集』付録「徵士穆門周先生墓志銘」に「浹」に作る。以下、『無悔斎集』との異同箇所は＊で示す。

其の人淵然湛然として、能く其の涯涘（がいし）を窺ふ莫く、渾淪たる元気、眉宇に充積し、蓋し古の黄叔度、陳仲弓の流なり。士は賢不肖と無く、皆曰く、周先生は長者なり、と。乃ち其の中には則ち確乎として抜くべからざる者有りて、形迹を以て自ら見はれず。大科の役、姚侍郎三辰 之を薦め、穆門力めて辞するも得ず、徵に応じて京に至り、公車の門下を徘徊すること数日、疾と称して卒に試に就かずして以て帰るは、能く測る莫きなり。已にして始めて其の高きに服す。

【通釈】人となりはどこまでも清らかで穏やかで、渾沌たる精気がその容貌の中に充ち満ちているようであって、昔の黄叔度、陳仲弓のたぐいであろう。賢人不肖の士いずれもが、皆「周先生はすぐれた方だ」と言った。彼の中には確乎として揺るぎないものが有りながら、挙動に表わすことはなかった。博学鴻試の時に、姚三辰が周京を推薦し、穆門は懸命に断ろうとしたが断り切れず、朝廷からの召集に応じて北京に行き、試験場の入り口あたりを徘徊すること数日、ついに病気を口実に試験を受けずに帰ったのは、理解しがたいことであった。そうして始めて彼の高潔さに感服した。

周京の人となりを喩えた「黄叔度」とは、漢の黄憲（字は叔度）のことで、孝廉に挙げられ、公府に召されて上京するも、仕官せずに帰郷したという。「陳仲弓」は、後漢の陳寔（ちんしょく）（字は仲弓）のことで、宦官の専横に反対し、党錮の禁を受け、以後請われても出仕しなかった。いずれも高潔、清廉なる人格を称された名士達である。博学鴻試に推挙されながら、病と称して拒否した周京は、俗人からは理解しがたいほどの高潔な人物だという。ちなみに乾隆元年の博学鴻試には、杭州から 27 名が推挙され（『民国杭州府志』巻一百十一）、周京の他にも、杭州詩壇からは厲鶚、杭世駿、符曽、趙信、桑調元、汪台、汪沆、袁枚らが推薦を受けた。『国朝杭郡詩輯』巻十七「周京」には「乾隆丙辰薦応博学鴻試不就、試後考授州同知。（乾隆丙辰薦められて博学鴻試に応ずるも就かず、試の後州同知を考授せらる。）」とあり、博学鴻試の後、「州同知」の職を授けられたことを記す。桑調元「墓志銘」では、「職列州同要路欲挙以官之固辞、尋薦大科試保和殿放帰。（職は州同の要路に列せられ、挙て以て之に官たらんとするに固辞し、尋いで大科もて保和殿に試みらるるも放帰せらる。）」という。州同知と州同は同じ官職で、知州の副職に当たるが、桑調元はこの職を与えられることになったのは、博学鴻試より前のこととしている。桑調元によれば州同知も固辞したことになるが、いずれにせよこの職にはつかなかったと考えられる。

符曽は、「哭周穆門先生」で「曽達声名上相知、千金却聘払衣時。潞州一往成孤誼、高節他年要補遺（曽て声名 上相知り、千金却って聘す 衣を払ふの時。潞州一たび往きて孤誼を成し、高節他年補遺するを要す。）」と、名声故に推挙されても拒否したことを「高節」と称えている。符曽『春鳧小稿』所収のこの詩の自注には、（「載之史伝為高士為独行又何多譲焉。（之を史伝に載せ高士と為し独行と為すも又何ぞ多く譲らん。）」（『春鳧小稿』己巳「挽周穆門」）と言い、孤高なる高士として史伝に載せるべきとまで言う。この博学鴻試で

は、厲鶚を始め、杭州の唱酬の仲間達には応じないものが多かったのだが、その背景にはこうした行為を高節と称える考えがあったのであろう。

４．杭州詩社と周京

〔３〕杭之詩人為社集、群雅所萃、奉穆門為職志。詩成穆門以長箋写之、酔墨淋漓、姿趣頽放、或弁数語於其端、得者以為鴻宝。湖社風流、百年以来、於斯為盛、皆穆門之所鼓動也。

杭の詩人 社集を為し、群雅の萃まる所、穆門を奉じて職志と為す。詩成れば穆門　長箋を以て之を写し、酔墨淋漓、姿趣頽放たりて、或ひは数語を其端に弁ずれば、得る者以て鴻宝と為す。湖社の風流、百年以来、斯に於いて盛と為すも、皆穆門の鼓動する所なり。

【通釈】杭州の詩人が詩社の集まりをし、あまたの詩人が集まる所では、穆門を代表に奉じた。詩ができると穆門が長い用紙に書き写すのだが、酔って書く文字は墨跡淋漓として、字の趣は自在で奔放であり、紙の端に数語を評語として書き付けると、手に入れたものは宝とした。杭州西湖の詩社の風流は、百年来、ここにおいて隆盛となったのであるが、皆穆門が奮い起たせたものであった。

この段は杭州詩社の詩会について述べている。先に周京 30 代の頃、20 代の厲鶚らと開いた詩会の様子を紹介したが、ここで全祖望が記しているのは、乾隆六年に周京が 65 歳で地方から帰郷して以後のことである。詩社の集まりではリーダーと目され、詩のみならず書でも人気を博したようすが覗える。桑調元「墓志銘」には、「晩息影蓬廬、共里中詩老結吟社、策杖出西郭、扁舟草屩、縱浪湖山間。（晩に蓬廬に息影し、里中の詩老と共に吟社を結び、杖を策（つ）きて西郭を出で、扁舟草屩（きゃく）もて、浪湖山間を縱（ほしいまま）にす。）」とあり、周京が晩年故郷杭州に戻って詩社を結び､湖や山に出かけては詩を仲間達と詠じたことがわかる。よって『無悔斎集』には、郊外で同人たちと唱和した詩が多数収録されている。

杭州詩社は、各地に出ていた杭州詩人達が、次々に杭州に戻ってきたことで最盛期を迎えた。その様子は次のように記されている。

乾隆庚申、同書与先生長君仍叔同入郡庠、始獲拝先生于牀下。……越三年、金冬心先生自揚帰。……会其時、郷耆宿或致仕帰田、或倦游還里。如顧月田、沈峙公、鄭璣尺、金江声、呉東壁、周穆門、魯秋塍、厲樊榭、杭堇浦、施竹田諸先生、一時并集。而方外則有笑虚、譲山、高人則先生与金先生両人、詩社為最盛焉。　　（丁敬『硯林詩集』梁同書序）

乾隆庚申、同書　先生の長君仍叔と同に郡庠に入り、始めて先生に牀下に拝するを獲。……越ゆること三年、金冬心先生揚より帰る。……会（たまた）ま其の時、郷の耆宿或ひは致仕して帰田し、或は倦游して還里す。顧月田、沈峙公、鄭璣尺、金江声、呉東壁、周穆門、魯秋塍（しょう）、厲樊榭、杭堇浦、施竹田諸先生の如き、一時に并集す。而して方外には則ち笑（いん）虚、譲山有り、高人には則ち先生と金先生と両人有りて、詩社最盛と為す。

「乾隆庚申」とあるのは、乾隆五年（1740）のことである。その３年後の乾隆八年（1740）、金農が揚州から帰り、偶々その頃職を辞したり、旅を終えたりした多くの詩人が杭州に戻って一堂に会したことで、「詩社最盛」を迎えたという。詩社を開いたのは、まず顧之珽であり、その後に杭世駿が開いたと考えられる。顧之珽の詩社について、杭世駿は次のように記す。

憶歳壬戌、癸亥間、顧丈月田以詞場宿老、号召同里詩人為社於西湖、月必五六会。蒸然発動、転相招引、振采騰華、於時最盛。　（杭世駿『道古堂文集』巻十二「遠村吟稿序」）

憶ふ 歳壬戌、癸亥の間、顧丈月田 詞場の宿老を以て、同里の詩人を号召して社を西湖に為り、月に必ず五六たび会せしを。蒸然として発動し、転た相ひ招引し、振采騰華し、時に於いて最も盛ん。

壬戌、癸亥とは、乾隆七、八年をいう。杭世駿は乾隆八年に杭州に戻り、こうした詩社の隆盛を見、また自身も参加した。顧之珽の詩社に参加した面々は以下のようである。

顧之珽、……中蜚語罷帰、与里中周京穆門、朱樟鹿田、許大綸初観、鄭江筠谷、金志章江声、呉廷華東壁、戴廷熺鸝亭、厲鶚樊榭、汪台復園、梁啓心菉林、杭世駿堇浦、丁敬竜泓、張湄柳漁、江源敬斎、陳兆崙句山、施安竹田、汪沆西顥、顧之麟寸田、為湖南詩社、掎裳聯襼、如屏風然。西泠十子後、此為極盛矣。　（『国朝杭郡詩輯』巻八　顧之珽）

顧之珽、字は搢玉、茶園と号し、又月田と号す、仁和の人。……蜚語に中りて罷帰し、里中の周京穆門、朱樟鹿田、許大綸初観、鄭江筠谷、金志章江声、呉廷華東壁、戴廷熺鸝亭、厲鶚樊榭、汪台復園、梁啓心菉林、杭世駿堇浦、丁敬竜泓、張湄柳漁、江源敬斎、陳兆崙句山、施安竹田、汪沆西顥、顧之麟寸田と、湖南詩社を為り、掎裳聯襼、屏風の如く然り。西泠十子の後、此れ極めて盛んと為す。

顧之珽は根拠の無い蜚語によって官職を辞して帰郷し、杭州で湖南詩社を開いた。詩社には、周京ほか多くの杭州詩人が参加し、明末の西泠十社以来の「極盛」と称されたという。やがて杭世駿も、自ら詩社を開く。これが南屏詩社である。

武林自西湖八社而後、風雅衰息幾二百年。余被放帰田、於南屏開設壇坫。金江声観察、丁鈍丁隠君、周辛老、厲樊榭両徴士、牽連入社。
（杭世駿『道古堂文集』巻四十五「朝議大夫刑部貴州司主事呉君墓表」）

武林 西湖八社よりして後、風雅衰息すること幾んど二百年。余放たれて帰田し、南屏に於いて壇坫を開設す。金江声観察、丁鈍丁隠君、周辛老、厲樊榭両徴士、牽連として社に入る。

杭世駿自身が乾隆八年に帰郷して後、南屏に詩社を開いたことを記す。『道古堂集』汪沆序には、直接「南屏詩社」という名称が見える。[6] 南屏詩社には周辛老つまり周京も参加している。

周京『無悔斎集』は、詩社同人達との唱酬の作を多く収録する。また南屏詩社には、浄慈寺の僧侶が参加し、詩会も浄慈寺で開かれることが多かったため、際祥纂輯『勅建浄慈寺志』、釈篆玉『話堕集』、釈明中『㝢虚大師遺集』等、浄慈寺関係の文献は詩社の記録を多く留めている。全祖望「墓志銘」に、こうした西湖を中心とする詩社の活動を周京が鼓動したといっており、設立者ではないにもかかわらず、常に詩社の中心にいたのが周京だったようだ。梁文濂「哭周穆門先生」には「武林風雅甲天下、招邀朋輩聯詩盟。推君高座主壇坫、紀律厳于細柳営。（武林の風雅　天下に甲たり、朋輩を招邀して詩盟を聯ぬ。君を高座に推して壇坫に主たらしめ、紀律　細柳営より厳たり。）」とあり、やはり周京を杭州詩会の領袖として皆が推したと言う。また、「周京……晩歳息影湖山、里中為社集者、咸為職志。（周京……晩歳 湖山に息影し、里中の社集を為す者、咸職志と為す。）」（『国朝杭郡詩輯』巻十七　周京）の記述からも、周京が隆盛期の杭州詩壇の領袖として認識されていたことがわかる。また領袖となった時期については、厲鶚「哭周穆門先生」自注に「六年来同人結社湖上、先生為之領袖（六年来同人社を湖上に結び、先生之が領袖と為る。）」とあり、没年である乾隆十四年（1749）から溯ること６年、つまり乾隆八年の顧之珽の詩社設立時から詩社の領袖であったことがわかる。その理由の１つには、詩書の名声と並んで、周京の人間性に対する周囲の敬慕があったようだ。

５．人倫

〔４〕尤篤於人倫之誼、其娶婦也賢、而頗不得于其姑、穆門戒之曰、「黄涪翁之姉文城君困於洪氏、雖有三令子、莫能申也。汝其善事姑矣。」婦卒以是困悴而死。穆門事其母益孝、不敢有幾微見於顔色、然私憐其婦、終身不更娶以報之。有弟已析産、乗穆門之出遊而鬻其居、穆門帰更僦屋、不以一語及之。故人王、袁、許三子者死、有女皆流落、穆門贖之帰、并其二従女皆撫之如女、択婿而嫁之、以是晩景益窮、然其敦古道益摯。

尤も人倫の誼に篤く、其の婦を娶るや賢なるも、頗る其の姑に得ず、穆門之を戒めて曰く、「黄涪翁の姉文城君は洪氏に困しみ、三令子有りと雖も、能く申す莫きなり。汝其れ善く姑に事へよ」と。婦卒に是を以て困悴して死す。穆門 其の母に事ふること益ます孝、敢へて幾微も顔色に見はすこと有らず、然れども私かに其の婦を憐れみ、終身更に娶らずして以て之に報ゆ。弟有りて已に析産し、穆門の出遊に乗じて其の居を鬻ぎ、穆門帰りて更めて屋を僦り、一語を以て之に及ばず。故人王、袁、許三子なる者死し、女有りて皆流落するに、穆門之を贖ひて帰り、并びに其の二従女皆之を撫すること女の如く、婿を択びて之を嫁がしむ、是を以て晩景益ます窮するも、然れども其の古道に敦きこと益ます摯たり。

【通釈】とりわけ人倫の道理に篤く、彼が妻を娶ったところ賢婦であったのに、まったく姑に気に入られなかったので、穆門は妻を戒めて「黄庭堅の姉である文城君は姑の洪

氏に苦しめられ、息子が三人いたのだが、何も言うことができなかった。おまえは姑によくつかえておくれ」と言った。妻はついにこのために苦しみやつれて死んだ。穆門は母親につかえて益々孝行をつくし、少しも顔色に出すことはなかったものの、心ひそかに妻を憐れんで、終生再婚せずに報いた。破産してしまった弟がおり、穆門が出遊している間にその家を売り払ってしまったのだが、穆門は帰ってから改めて家を借りて、一言もこのことに触れなかった。古い友人の王、袁、許の三人が死に、娘が路頭に迷っていたが、穆門は彼女たちを買い取って帰り、また二人の姪を自分の娘のように可愛がり、婿を選んで嫁がせた。こういうわけで、晩年の暮らしは益々窮乏したのだが、古からの道義には益々あつかった。

この一段は、『無悔斎集』付録の元来の墓志銘にはなかった部分である。家族、友人に対する、周京の情の深さを伝えるエピソードではあるが、嫁姑の軋轢、弟の破産と裏切りなど、家族の恥部をさらすものでもあったため、息子の依頼による当初の墓志銘には、書かれなかったのだろう。

周京が母を優先し、妻に忍耐を要求するのは、息子として当然であったろうが、一方で妻を憐れに思って再婚しなかったというのは、実直で情の深い人間像を伝えている。更に自分を欺いて家を売ってしまう弟を許し、自分の暮らしもままならぬのに、友人の娘達を実の娘のように面倒をみて嫁がせるというから、その人間性には計り知れぬものがある。

桑調元「墓志銘」では、「姚氏性正静、事尊章以孝聞先卒、君盛年不再娶、旁無侍姫。（姚氏性正静にして、尊章に事ふるに孝を以て聞こゆるも先に卒し、君盛年なるも再びは娶らずして、旁らに侍姫無し。）」と、妻が姚氏であることを記した上で、義父母に孝行で、周京が妻の死後、終生独身を通したとのみ伝えているが、これは周京のために妻と姑の間の悲劇をあえて書くまいとしたためではなかろうか。桑調元は更に周京が家族を大切にした様子を以下のように伝える。

君事親孝、除夕考病思飴橘、苦購得之、祷于神滅己算、益考寿。翌日乃瘳、延五載、授経。奉慈幃甘旨、年垂艾依依若孺子。友愛弟、視弟子女如己出、敦族誼営葬、群従恤孤咸得所娶。

君親に事へては孝、除夕考（ちち）橘を飴めんと思ふ、苦しみ購ひて之を得、神に祷（いの）りて己が算を滅じ、考の寿を益さんとす。翌日乃ち瘳（い）え、五載を延べ、授経す。慈幃に甘旨を奉じ、年艾に垂（なんなん）とするも依依として孺子の若し。弟を友愛し、弟の子女を視ること己より出づるが如く、族誼に敦く葬を営み、群従孤を恤（あわ）れみ咸（みな）娶る所を得。

ここでは、病の父のために橘を苦労して買い求め、自分の寿命を父親に与えるよう神に祈り、母にはおいしい物を食べさせ、弟の子どもは我が子のように面倒をみるというように、家族の誰に対しても身を挺して尽くす周京の姿を描いている。いずれにせよ、

周京が身近な人々に献身的に尽くし、周囲から尊敬を集めていたことがうかがわれる。こうした人間性が、詩壇のリーダーとして慕われた大きな要因でもあろう。

ところで、この時期の杭州詩壇の隆盛を示す２度の大きな詩会があった。１度目は乾隆九年（1744）７月 12 日に、詩社同人である汪台が私園「復園」で開催した復園での詩会である。参加者 20 名の詩は「復園紅板橋詩」（『武林掌故叢編』所収）として刊行された。この詩集では周京の詩が最初に収録されている。更に２年後の乾隆十一年（1746）３月３日、杭州知府鄂敏が「西湖修禊」を開催する。集まった詩人は 61 名、浙江の詩人を中心にした、当時最大規模の詩会であった。詩会の詩は「西湖修禊詩」一巻として刊行されたが、周京はこの後序を任されている。これらも周京が詩壇の中心人物であったことを裏付けるものと言えるだろう。

６．祖先

〔５〕穆門故鄞産、前明右副都御史莓崖先生相之後、其遷杭五世、副都于先司空公為石交、副都之孫観察於先宗伯為姻家、故余于穆門*尤相愛也。近副都之後居鄞者微、不可問、穆門睠念大宗、形之寤寐。余嘗為穆門言、莓崖墓在太白山上、廿年以来、神道荒蕪、石馬眠草中、寒食麦飯、恐無挙者。穆門泫然流涕曰、吾当東帰買墓田、復置墓戸以守之。是後、歳歳相見必及此、然詘于力、竟未能也。

＊「余于穆門」‥『無悔斎集』に「穆門于余」に作る。

穆門故（も）と鄞の産、前明の右副都御史莓崖（ばいがい）先生相の後にして、其の杭に遷ること五世、副都は先の司空公と石交を為し、副都の孫観察は先宗伯と姻家と為り、故に余 穆門と尤も相愛するなり。近ごろ副都の後の鄞に居する者微にして、問ふべからず、穆門大宗を睠念し、之を寤寐に形はす。余嘗て穆門の為に言ふ、莓崖の墓は太白山上に在り、廿年以来、神道は荒蕪し、石馬は草中に眠り、寒食の麦飯は、恐らくは挙ぐる者無からん、と。穆門泫然として涕を流して曰く、吾当に東帰して墓田を買ひ、復た墓戸を置きて以て之を守るべし、と。是の後、歳歳相見れば必ず此に及ぶも、然れども力に詘（くつ）し、竟に未だ能くせざるなり。

【通釈】穆門は元々鄞（寧波）の出身で、前の明王朝の副都御史であった周相先生の子孫であり、杭州に移り住んで五世になるが、副都御史は先の司空公と固い友情で結ばれており、副都御史の孫となる観察はわがご先祖と姻戚となっていたため、私と穆門とはとりわけ仲が良かった。最近は副都御史の子孫で鄞に居住するものはごく僅かで、問題にならず、穆門は先祖を思い、いつもそれを態度に出していた。私は嘗て穆門に「莓崖殿の墓は太白山の上にあるが、二十年来、墓への道は荒れて草茫茫で、石馬は草の中に埋もれて眠っており、清明節の麦飯は恐らく供える者がいないだろう」と言った。穆門は涙をぽろぽろとこぼして「私は故郷に帰って、墓地を買い、再び墓守を置いて墓を守らせよう」と言った。この後毎年顔を合わせると必ずこの話になったが、力及ばず、ついに果たせなかった。

これは周京の祖先に関する記述である。全祖望は、寧波出身であるが、周京の祖先が寧波出身で、姻戚関係もあったことから、周京に特に親しみを抱いていた。[7] 桑調元「墓志銘」は、「其先自汝南南渡、居四明東尚里為望族。（其の先 汝南より南渡し、四明東尚里に居りて望族と為る。）」と、寧波に移る以前、河南の汝南に居り、その後寧波に来て名家となったことを記している。全祖望は祖先のルーツを記すと同時に、祖先のために墓を整えようとし、貧困のために果たせず涙を流す、周京の人となりまで描こうとしているようだ。

７．晩年と死

〔６〕暮年、別自署東双橋居士。東双橋者、副都所居鄞城北坊第也。昨年予病於杭幾死、穆門昕夕訪視。予稍進食、穆門頻齎榼来過。次年、余在越中、而穆門呉淞之訃至矣。穆門死、湖社諸人一若失其憑依*者、其為人可想見也。

＊「依」‥『無悔斎集』に「倚」に作る。

暮年、別に自ら東双橋居士と署す。東双橋とは、副都の居りし所の鄞城の北坊第なり。昨年予 杭に病み幾んど死せんとするに、穆門昕夕訪ひて視る。予稍（やや）食を進め、穆門頻りに榼（たる）を齎（もたら）して来過す。次年、余 越中に在るに、穆門呉淞の訃至る。穆門死し、湖社の諸人一へに其の憑依する者を失ふが若く、其の人と為り想見すべきなり。

【通釈】晩年には、それまでと別に自ら東双橋居士と署名した。東双橋とは、副都御史の居住していた鄞の町の北坊第である。昨年私が杭州で病気になり死にそうになったのだが、穆門は朝夕様子を見に来てくれた。私がやや食欲が出てくると、穆門は頻りに飲み物を入れた容器を持ってやって来てくれた。翌年、私が越の地にいるときに、穆門が呉淞で亡くなったという知らせが届いた。穆門が亡くなり、杭州湖社の詩人達はまるでよりどころを失ってしまったかのようであった。穆門の人となりをうかがうことができるというものだ。

ここでは、周京の晩年から死までを述べる。「東双橋居士」と称したことについては、『国朝杭郡詩輯』巻十七に「周京、字西穆、一字少穆、号穆門、亦号辛老、晩号東双橋居士。」とあり、晩年の号とある。周京が杭州を少し離れた呉淞で亡くなったことについては、厲鶚が「先生卒于雲間客舎。（先生 雲間の客舎に卒す。）」（「哭周穆門先生」自注）と記している。雲間も江蘇省松江の古称で、呉淞と同じ地方を指す。客舎とあるように短い旅に出てそのまま亡くなったようだ。

周京の死によって、杭州詩社は頼るべき領袖を失った、とあるが、詩社の同人である金志章も「湖南寂寞吟壇冷（湖南寂寞　吟壇冷ゆ）」（「哭周穆門先生」）と、その死とともに詩社が衰退したと述べている。周京は、当時詩社の最年長であり、詩才の豊かさとその人格で杭州詩社の支柱となっていた。杭州詩社の長老であった顧之珽、周京と次々に詩人を失い、杭州詩社は衰退の途を辿り始め、乾隆十七年（1752）、杭世駿が広州に行くために杭州を離れ、同年の厲鶚が世を去ったことにより、終焉を迎えるのである。

８．家系

〔７〕穆門、姓周氏、諱京、字西穆、一字少穆。曽祖某、祖某、父某。娶某氏。生于某年某月某日、卒於某年某月某日、得年七十有三、葬于湖上之某*。子宸望、諸生。

＊『無悔斎集』は「曽祖～某山」を「曽祖諱竜、祖諱楠、父諱大用。娶姚氏。生于康熙丁巳二月十四日、卒於乾隆己巳六月十有七日、得年七十有三、葬于湖上之北山青芝塢。」とする。

穆門、姓は周氏、諱は京、字は西穆、一の字は少穆。曽祖は某、祖は某、父は某。某氏を娶る。某年某月某日に生まれ、某年某月某日に卒し、年を得ること七十有三、湖上の某山に葬らる。子は宸望、諸生なり。

【通釈】穆門は、姓は周氏、諱は京、字は西穆、もう一つの字を少穆といった。曾祖父は某、祖父は某、父は某。某氏を娶った。某年某月某日に生まれ、某年某月某日に亡くなった。年は七十三、西湖の畔の某山に葬られた。息子は宸望で、諸生である。

桑調元「墓志銘」は、家族に関する更に詳しいエピソードを載せる。

有諱竜者、明太常少卿、懸車愛西湖家焉、了無長物、惟古書名画数十篋、野服長吟、詩意清遠、君曽祖也。祖考諱楠、郡痒生、考諱大用、邑痒生、両世医済人、有隠徳、妣黄太君。

諱は竜なる者有り、明太常少卿、懸車して西湖を愛してこれに家するに、了に長物無く、惟だ古書名画数十篋のみあり、野服長吟し、詩意清遠たるは、君の曽祖なり。祖考諱は楠にして郡の痒生、考諱は大用にして邑の痒生、両世医もて人を済ひ、隠徳有り、妣は黄太君なり。

これによれば、曽祖父の周竜は明の太常少卿であったが、引退すると西湖を愛して杭州に居を移したという。祖父は周楠、父親は周大用といい、父と祖父は医療に携わっていたようだ。母が黄太君であったことも、全祖望「墓志銘」にはない情報である。

周京の字については、全祖望「墓志銘」に記載される西穆、少穆以外に、呉城「哭周穆門先生」第四首第五句「字記老忘辛」自注に、「先生一字辛老」とあり、更に「辛老」という字もあったことがわかる。ただし、『国朝杭郡詩輯』巻十七では、「周京、字西穆、一字少穆、号穆門、亦号辛老、晩号東双橋居士」と、「辛老」を号としている。埋葬に関しては、桑調元「墓志銘」に「是年十二月十一日孤子将合葬考妣于北山青芝塢之陽」と、父母と合葬したことを記している。

没年月日については、桑調元「墓志銘」には、「乾隆十四年六月十七日君卒、享年七十有三。距姚孺人歿三十有五年矣。」と記している。また亡妻の死後35年であったことも記している。

息子の宸望は、乾隆十一年の「西湖修禊」に、父と共に参加しており、「周暢鶴宸望」と記されている。暢鶴は号である。また桑調元「墓志銘」は、「子宸望、邑庠生、孫方岳。」

と、方岳という孫がいたことも記している。更に『無悔斎集』巻十三に「春朝携令之従之二女湖上」と題する詩があり、令之、従之という二人の娘がいたようだ。王小恒氏は、この二人の娘には独立した詩作が見当たらないとするが、[8)]『民国杭州府志』巻九十四芸文に「天香楼外楼吟草、銭塘周京女登望従之撰」とあり、周京の娘に『天香楼外楼吟草』という詩集があったことがわかる。「登望」「従之」は娘二人の名と思われるが、『無悔斎集』の詩題にあった「令之」ではない。「登望」が「令之」と同じ娘を指すのかは不明である。

９．作品集

〔８〕穆門之卒也、吾友杭堇浦為之伝、序其事甚悉。厲樊榭、施竹田論定其詩、山陰令舒堃畝為之開雕。而宸望又以幽室之文属余、是不可以辞也、乃更為之銘曰、

重湖黯然、喪我祭酒。白雲封之、其骨不朽。

穆門の卒するや、吾が友杭堇浦之が伝を為り、其の事を序して甚だ悉たり。厲樊榭、施竹田其の詩を論定し、山陰令舒堃畝（こんほ） 之（これ）が為に開雕す。而して宸望又幽室の文を以て余に属し、是れ以て辞すべからざるなり、乃ち更に之が銘を為して曰く、

重湖黯然として、我が祭酒を喪（うしな）ふ。白雲 之を封ずるも、其の骨朽ちず。

【通釈】穆門が亡くなると、わが友杭堇浦（世駿）が彼の伝を作り、事跡を詳細に記した。厲樊榭（鶚）と施竹田（安）とがその詩を編纂し、山陰の令である舒堃畝（瞻）が彼のために刊行した。息子の宸望は更に故人のための文章を私に依頼し、これを断ることはできなかった。そこで更に彼のために墓志銘を作ったのである。

西湖の里湖と外湖は哀しみにくれる、わが祭酒を失ったために。

白雲は彼を閉じこめ、その骨は朽ちることがないだろう。

杭世駿が詳しい伝を作ったとあるが、現存する『道古堂集』には収録されていない。全祖望『鮚埼亭集』楊注は、「道古堂集中無此伝、恐刻時刪去。（道古堂集中に此の伝無し、恐らくは刻する時刪去せしならん。）」と、『道古堂集』刊刻に際して削除されたものと推測している。

周京の詩集を編纂したというのは、『無悔斎集』のことで、厲鶚、施安が編纂し、周京晩年の唱酬の仲間であり乍浦理事同知であった舒瞻が出版費用を負担して刊行した。厲鶚「無悔斎集序」に「令嗣宸望以遺稿属予点定、略為刪汰其什之二三。鏤版以行者、故人舒明府雲亭。（令嗣宸望 遺稿を以て予に点定せんことを属し、略ぼ為めに其の什の二三を刪汰す。鏤版して以て行ふ者は、故人舒明府雲亭なり。）」とあり、十分の二、三を削ったとある。現在完本として見ることができるのは、十五巻本であるが、桑調元「墓志銘」に「著有無悔斎集、古文詩詞共三十六巻、舒明府瞻為先刻詩十五巻行世、厲樊榭編定、余存于家。（著に無悔斎集有り、古文詩詞共に三十六巻、舒明府瞻為に先に詩十五巻を刻して世に行はれ、厲樊榭編定し、余家に存す。）」とあるのによれば、元々は古文詩詞合わせて三十六巻があったものを、舒瞻がまず詩集十五巻を出版した。それは厲鶚が編纂したものであり、残りの原稿は桑調元の家に保存してあったという。現在、南京図書館に『無悔斎集』三十二巻

として、巻三十一、三十二の２巻のみが所蔵されている。王小恒氏の報告によれば、『無悔斎集』の他に、周京の遺作として『絳雲集』が南京図書館に、『鉄崖詩鈔』の抄本が上海図書館に、『古侠遺詩』が、温州市図書館に所蔵されているとのことである。[9)]

なお、墓志銘最後にある「重湖」は、西湖を堤で分けた里湖、外湖を合わせて言ったもので、西湖の詩会の地を指す。「祭酒」は、詩社の年長者で徳行に優れた者を言う。個人の死を悼むというよりも、杭州詩社の人材を失ったことを悼む文で墓志銘は締めくくられる。

以上が全祖望による「周穆門先生墓志銘」からみた周京である。

10. 周京略年譜

最後に、周京と同時代の杭州詩壇の動きを略年譜とする。

・『無悔斎集』を元に作成。それ以外は必要に応じて根拠を示す。

・＊は杭州詩壇関連事項

康熙四十七（1708）〔32歳〕江西に行く。

康熙四十八（1709）〔33歳〕河南の豫州に行く。

康熙四十九（1710）〔34歳〕北遊し、山東から河北、河南、洛陽方面に行く。

康熙五十四（1715）〔39歳〕～康熙五十五年（1716）〔40歳〕杭州に居り、年下の厲鶚らと詩会で交流。常に首唱となる。（厲鶚「無悔斎詩集序」）

康熙五十八年（1719）〔43歳〕厲鶚、呉奐らと盛んに唱酬。

康熙六十一年（1722）～雍正元年（1723）〔46、47歳〕蘇州に行く。再び北上し、泰山から北京に向かう。

雍正元年（1723）〔47歳〕

＊雍正元年～二年にかけて、杭州で詩人七名（沈嘉轍、呉焯、陳芝光、符曽、趙昱、厲鶚、趙信）により『南宋雑事詩』が作られる。（『南宋雑事詩』）

雍正二年（1724）〔48歳〕北京に居る。母親の訃報を聞き、南帰する。

雍正三年（1725）〔49歳〕杭州に帰る。

雍正四年（1726）〔50歳〕再び河北に旅立ち、景物を詠じ、地元の名士たちと唱酬を行う。

雍正五年（1727）〔51歳〕陝西、甘粛に旅し、西安に戻る。

雍正六年（1728）〔52歳〕再び甘粛に旅する。

雍正七年（1729）〔53歳〕河北から南帰する。

雍正九年（1732）〔55歳〕杭州に居る。

＊杭世駿が北京から戻り、「湖山之社」を結ぶ。（『道古堂文集』巻四十七「林阮林墓碣」）

雍正十一（1733）〔57歳〕酒楼の題壁詩が人気を博す。[10)]

雍正十三（1735）〔59歳〕再び陝西、甘粛に旅する。

乾隆元年（1736）〔60歳〕博学鴻試に推挙され、北京に向かうが、受験せず。

＊博学鴻試に杭世駿、厲鶚、趙信、周京、胡天游、汪臺、汪沆、袁枚らが推挙される。

乾隆二年（1737）〔61歳〕北京を離れる。

乾隆三年（1738）〔62歳〕杭州に居る。

乾隆四年（1739）〔63歳〕山西沢州にゆき、沢州知府朱樟と唱酬を行う。

＊梁啓心、北京から杭州に帰る。

乾隆五年（1740）〔64歳〕朱樟と沢州にて唱和を繰り返す。

＊鄭江、厲鶚が杭州に帰る。

乾隆六年（1742）〔65歳〕杭州に戻り、蘇州を旅する。

＊天津水西荘にて『沽上題襟集』刊行。収録詩人８名のうち３名が杭州詩人。

乾隆七年（1742）〔66歳〕杭州に居る。

＊顧之珽が杭州に戻り、「湖南詩社」を設立、周京も参加。

乾隆八年（1743）〔67歳〕顧之珽、杭世駿らと詩社に参加。

＊２月、杭世駿が罷免され、杭州に帰り、「南屏詩社」を設立。

＊丁敬が揚州から杭州に帰る。汪沆も帰郷。

乾隆九年（1744）〔68歳〕呉淞にゆく。復園の詩会に参加。

＊金志章が北京から杭州に帰る。戴廷熺が杭州に帰る。

＊７月、汪台の復園にて詩会。参加者杭州詩人21名。『復園紅板橋詩』一巻。

乾隆十年（1745）〔69歳〕杭州に居る。詩会に参加。

＊呉震生、杭州に転居する。

＊顧之珽死去。（『道古堂集』巻四十二「文林郎行人司行人管理広東電白県知県事顧君墓志銘」）

乾隆十一年（1746）〔70歳〕西湖修禊に参加。聯句に参加。[11)]

＊閏３月３日、西湖において杭州太守鄂敏が修禊を主催。６１名が参加。鄂敏の前序、周京の後序、釈明中の絵画を附し、鄂敏が刊行。『西湖修禊詩』巻一。

＊瓶花斎で杭州詩人１２名による聯句。「十二月八日集瓶花斎主人以臘八粥供客聯句」周京、厲鶚、杭世駿ら参加。（『樊榭山房続集』巻六）

＊厲鶚『宋詩紀事』刊行。周京も参加。

＊鄭江死去。

乾隆十二年（1747）〔71歳〕杭州詩会に参加。

＊趙昱死去。

乾隆十三年（1748）〔72歳〕杭州詩会に参加。

＊鮑鈐死去。厲鶚、天津水西荘に行く。

＊揚州詩壇にて『韓江雅集』刊行。収録詩人41名中、杭州詩壇関係者15名。

乾隆十四年（1749）〔73歳〕呉淞にて死去。

＊天津水西荘の査為仁死去。

乾隆十七年（1752）

＊杭世駿が粤秀書院主講となり、杭州を離れ広州に赴く。

＊厲鶚死去。杭州詩社の主要人物が全て死去し、詩壇の衰退が顕著になる。

乾隆二十年（1755）

＊多くの杭州詩人の支援者となった揚州詩壇の塩商、馬曰琯死去。

まとめ

周京は、当時の名声にもかかわらず、死後次第に忘れられていった。同じ詩会のメンバーでも著作の多かった厲鶚、杭世駿らが後世に名を残したのと大きく異なっている。それは周京が詩会というその場限りの舞台で活躍した詩人であったためであろう。詩会の詩は、詩集に刊行されるものもあるが、ほとんどはその場限りのものである。出版されても一時的な雑誌扱いであるものが多く、保存されるのは、ごく大きな詩会のものに限られている。それとても、詩会という催しを伝える資料とはなりえても、詩人個人の名を伝えるにはその力は弱い。杭州詩社の詩会は、雍正年間はじめの２年、また乾隆８年からの約10間の隆盛期を経て、詩人の消滅とともに衰退してしまった。詩社の盛衰とともに詩社の詩人達の記憶もまた消えていかざるを得なかったのである。

周京の伝記を明らかにし、また彼が生きた杭州詩壇の詩人の動きを明らかにすることは、当時の杭州詩壇を始めとする地方の詩会や詩社の状況を知る手がかりになるだろう。文学史に名を残す著名な詩人以外に、その時だけの人気を得た詩人達が支えた、時代の文学、地域の文学に目を向けて始めて、文学史の流れは把握されるのではないだろうか。その意味で、この「周京伝」がそうした文学史の発掘にわずかながらでも貢献できることを願う。

注

１）本稿では、『鮚埼亭集』の底本として『全祖望集彙校集注』（朱鋳禹彙校集注、上海古籍出版、2000 年）を用い、『全祖望《鮚埼亭集》校注』（詹海雲校注・国立編訳館主編、鼎文書局、2003 年）を参照する。『無悔斎集』については、『清代詩文集彙編』（上海古籍出版社　2009）所収『無悔斎集』（乾隆十七年刻本影印）を底本とする。

２）桑調元の墓志銘は『清代詩文集彙編』（上海古籍出版社）本には収録されているが、『四庫全書存目叢書』（台南荘厳文化事業有限公司　1997）所収『無悔斎集』（浙江図書館蔵清乾隆刻本影印）には収録されない。

３）厳迪昌『清詩史』（浙江古籍出版社　2002）「第六章　乾嘉時期地域詩派詩群巡視　第一節　以厲鶚為代表的浙派」に「高峰期正是乾隆期、周京是個核心人物。」とあり、また、張兵等撰『文化視域中的清代文学研究』（人民出版社　2013）王小恒「第三章　浙派嬗変及厲鶚的文学思想、著述和交游」に「周京在杭州詩壇是個中心人物。他比浙派領袖厲鶚大十五歳、可称師友、也是浙派的核心成員」などとある。

４）全祖望の弟子董秉純編「全謝山年譜」には「雍正二年甲辰、先生二十歳。……自昨年再過武林、尽交樊榭、董浦、薌林、句山、谷林、意林、薏田、立甫諸先生、討論経史証明掌故、尊酒郵筒、殆無虚日。」とあり、雍正元年、全祖望 19 歳の時に武林即ち杭州に行き、杭州文人達と学問、詩酒の交流をしたとある。蔣天枢編「全謝山先生年譜」（商務印書館　1930）では、これを康熙六十一年のこととする。

５）呉焯と沈嘉轍が雍正十一年（1733）共に世を去った後、周京は「十二月二十二日同樊榭、耕民放舟湖上、時快雪初晴、風物閒静、重過酒舎、念欒城、尺鳧已下世、感今念昔、弥覚清游之足重也、分韻同作」（『無悔斎集』巻六）を作って、厲鶚、呉奐と共にその死を悼んでいる。この詩は酒楼の壁に題壁詩として書かれ、書写するものが絶えず、一時の話題となった。

６）「帰而杜門奉母、暇則偕里中耆旧襟契及方外之侶結南屏詩社、歌詠太平。」（『道古堂集』汪沆序）。南屏詩社については、鄭幸「南屏詩社考」（『厦門教育学院学報』第九巻第二期　2007）、劉正平「南屏詩社考論」（『北京大学学報（哲学社会科学版）』第五十巻第三期　2013）に論考があり。鄭江は南屏詩社は湖南詩社の別称ととらえ、劉正平は、湖南詩社が先にあり、顧之珽の死去に伴い、杭世駿の南屏詩社に合併されたとする。

７）祖先に関しては、全祖望『鮚埼亭詩集』巻六「莓厓周都御史、先司空公石友也。其従孫副使汝観。先宗伯公姻家也。今其後人已尽、偶与穆門徴君詢其世次、則莓厓之裔迁杭四世矣。為之驚喜、因勧其東帰展墓、以存故郷之遺。」という詩題でも語られている。

８）王小恒「周京交游考」（河西学院学報　2016 年第４期）「在周京詩集中、甚至有周京出游、携帯両個娘児出現、周京二女周令之、周従之、此二女雖未見独立的詩作、但想必才華出衆、這在整個浙派活動中都是絶無僅有的現象。」

９）王小恒「論周京的人格価値取向和旅隴詩歌創作」（『甘粛社会科学』2016 年第３期）に、「関于其遺集今可見者、其版本及館蔵状況如下：厲鶚為其編次、舒瞻刻印的為《無悔斎集》十五巻本、是現今可見周氏詩集之完帙、流布亦較広、存于北図、南図等処、四庫存目叢書本所収即此本；《無悔斎集》三十二巻稿本非完帙、僅残存三十一、三十二両巻、蔵于南図《絳雲集》為稿本、蔵于上図《鉄崖詩鈔》為抄本、蔵于中科院図書館、《古侠遺詩》亦為抄本、蔵于温州図書館。」とある。

10）厲鶚『樊榭山房集』巻七では、雍正十二年とする。

11）陳兆崙『紫竹山房詩集』巻四（丙寅至庚午）に「持螯聯句」があり、周京、顧之珽らが参加したとある。丙寅は乾隆十一年にあたり、顧之珽は前年世を去っている。人名の誤りか、開催時期の記録間違いかのいずれかであろう。

※本論は平成 26 年～28 年度科学研究費補助研究　基盤研究（C）26370421「乾隆時代における、移動する杭州詩人集団の変質と展開に関する研究」の研究成果の一部である。

The History of Zhou Jing
－With a Focus on “Zhou Mumen Muzhimeng ”－

Nobuko ICHINOSE

Department of Business Administration, Faculty of Business Administration, Fukuyama Heisei University

Abstract: The Hangzhou poets-circle has twice golden age. The first: the period from the end of Kangshi era to the beginning of Yongzheng, at this era *Nansong zazhishi* was made by Li E as most important person. The second: from Qianlong eight ten years continued. At this era the poets, who lived outside of Hangzhou, came together again and established poets-circle, opened actively poem-party. At this time the poet Zhou Jing was regarded as the leader of poets-circle. On the first Hangzhou era Zhou Jing led the poem-party already as the older leader of Li E and the other young poets. His amazing poem-style and personality was attractive for the young people. After that, Zhou Jing left Hangzhou, moved from place to place for many years and came back again to Hangzhou late in life and became the leader of poem-party again. They say, at that era Hangzhou poets-circle had hey-day. In spite of his fame in Hangzhou his name is not so famous until today. So we’d make clear his life history by epitaph or other biography and why he was loved as authority of Hangzhou poets-circle.

Key Words: Zhou Jing, epitaph, Hangzhou

福山平成大学経営学部紀要
第13号（2017），21－31頁

情報リテラシー初歩教育段階のタイピング修得練習におけるタッチタイピングの早期修得による効果について

奥田 由紀恵

福山平成大学経営学部経営学科

要旨：情報リテラシー教育の中で、初期段階に修得すべきであるのはタッチタイピング技術である。練習過程で、早期にタッチタイピングを習得するか否かが、本学科が合格ラインと位置付ける得点に到達するまでの過程に効果があるか否かをロジスティック回帰分析と生存分析を使用し、これまで行ってきたタイピング技術修得の方法が肯定的である可能性を探る。

キーワード：タッチタイピング、情報教育、生存分析

1. はじめに

福山平成大学では，全学科の1年次前期に情報リテラシーの授業(必修)の中で，4月の授業開始の最初にタッチタイピングの習得練習を取り入れている。タイピングソフトは、大学で自作したもの(Ptype Ver.3.0)を利用しており、初回から継続的に記録を取って、7月の中旬にタイピング試験を実施している。経営学科では、試験は、Ptype Ver3.0を使用し、ランダムに表示される基本英単語を正確に160文字タッチ／分 することを合格の目標としている。

授業は、全15回であるが、１回90分の授業時間の中でのタイピングに割く時間は、自由練習時間とテスト時間を合わせて、毎回20分程度である。練習方法は、前述の自作タイピングソフト（Ptype Ver.3.0）やオンライン上のバラエティあふれる様々なタイピングソフトで指使いの練習を十分にした後、合格ラインの160文字タッチ／分を目指して練習に入る。毎回のテスト時間は､5分から10分程度で､テスト時間内の最高得点をその日の記録とする。2から3週間の間隔で中間テストを行い、その都度、合格得点を設定しており、達成しなければペナルティとして打ち込み課題（英文と日本語の文章の混ざったもの。A4で1枚程度）を練習課題として渡している。

これまで、10年以上のタイピングテストの記録データを使用し、様々な観点から分析を行ってきた。そのうえで、疑問に感じ始めたのは、これまでの授業でタッチタイピング技術習得の練習を進める際に、キーボードを見ずに打鍵できるようになることを指導してきたが、はたしてそれが正しい方法なのか、タッチタイピングができる・できないで、どれだけタイピングテストの合否に影響するのだろうか、ということである。

そこで、今回は、2016年度前期の毎回の授業で記録したデータと、タッチタイピングが

できているかどうかを主観的に学生に判断させ、自己申告させたデータを使用して、分析を行った。

2. ロジスティック回帰分析による方法

一般化線形理論のロジスティックモデルを用いて、合格と不合格のオッズ（合格確率／不合格確率）について、タッチタイピングができる・できないのオッズ比を求める。

ロジスティックモデルではオッズは以下のように書ける。

$$\frac{p}{1-p} = \exp\left[\sum_{k=1}^{p} b_k x_k + b_0\right] \tag{1}$$

今、i 番目の変数 $x_i = \{0,1\}$ だけが異なる人があったとする。そのときの $x_i = 0$ のタッチタイピングができない人の合格確率 p_0 と、$x_i = 1$ のタッチタイピングができる人の合格確率 p_1 との違いを比較する。2人のオッズ比 r を取ると以下となる。

$$r = \frac{p_1}{1-p_1} \Big/ \frac{p_0}{1-p_0} = \exp(b_i) \tag{2}$$

これを p_1 について解くと以下となる。

$$p_1 = \frac{rp_0}{1+(r-1)p_0} \tag{3}$$

これはタッチタイピングができない人の合格確率とタッチタイピングができる人の合格確率の関係式である。

分析にあたっては、College Analysis Ver.6.3を使用した。

College Analysisでのロジスティック回帰分析のメニュー画面を図1に示す [1] 。

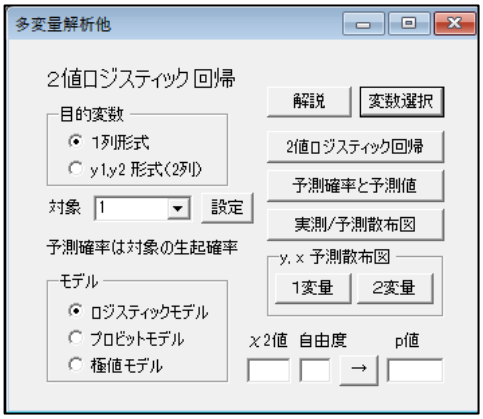


図1　ロジスティック回帰分析 実行メニュー

データは、2016年前期に測定したもので、その一部を図2に示す。

データ編集 2016経営学科前期最終データ.txt

	合否	合格回数	初期値	HP	B自己
1	1	6	146	1	1
2	1	5	84	1	1
3	1	10	85	1	0
4	1	10	77	1	1
5	1	5	91	1	1
6	1	4	55	1	1
7	0	10+	95	1	0
8	1	6	88	1	1
9	0	10+	69	1	0
10	1	9	81	1	1
11	1	10	70	1	1

4/4 (1,1)　分析:　備考:

図2　データ

データについて、

・「合否」は、合格・不合格を表す。

・「合格回数」は、授業何回目で合格したか(数値が低いほど良い)、を表す。
　最後の+記号は欠測値（不合格）を表す。

・「初期値」は、練習の第1回目に記録したタイピング得点データである。

・「HP」は、ホームポジションの指使いができている（1）・できていない（0）である。

・「B自己」は、学生に『タッチタイピングができている』と自己申告されたことを
　表し、できている（1）・できていない（0）である。

以上の項目で、分析用のデータをまとめた。

「合否」を目的変数とし、説明変数を、「初期値、HP、B自己」で、設定モデルは、「ロジスティックモデル」とした。図3は、分析結果である。

ロジスティック回帰分析結果

	偏回帰係数	標準化値	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
初期値	0.0744	2.0618	0.0361	0.0392	0.0037	0.1451	1.0772
HP	0.5780	0.3549	1.6285	0.7227	−2.6139	3.7699	1.7824
B自己	1.9154	1.5749	1.2162	0.1153	−0.4683	4.2992	6.7898
切片	−7.1209	−2.2611	3.1493	0.0238	−13.2935	−0.9482	
対数尤度値	−11.006						
逸脱度D	22.012	自由度	25	上側確率	0.6350	<-n小注意	
ピアソンχ2	21.546	自由度	25	上側確率	0.6618	<-n小注意	
C尤度比	15.350	自由度	3	上側確率	0.0015		
擬似R2	0.411						
実測予測R2	0.526						
誤判別確率	1を他と	0.105	0を他と	0.400			

図3　ロジスティック回帰分析結果

タッチタイピングができる・できないの合格のオッズ比は6.79となった。この分析結果から、$r = 6.79$ として(3)式をグラフに描くと図4のようになる。

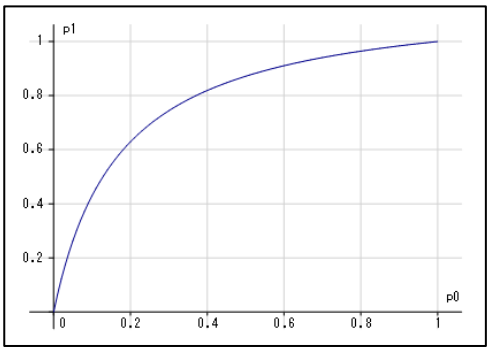


図4　合格確率を p_1 とした場合の関係式のグラフ

これを見ると、例えば、タッチタイピングができない人の合格確率が0.2であった場合でも、タッチタイピングができる人の合格確率は0.6を超えていることが分かる。ただ、これは平均を取った結果であり、データ数が少ないため、有意に可能性が高いとはいえない。

3.　生存時間分析による方法

生存時間分析の比例ハザードモデルを用いて、「B自己」（タッチタイピングの有無の自己申告）のタッチタイピングができる・できないで、1回の授業のハザード（ここでは合格確率として良い意味で使う）の比を求める。

一般に、比例ハザードモデルではハザードλを以下のように仮定する。

$$\lambda(t\,|\,\mathbf{x}) = \lambda_0(t)\exp\left[\sum_{k=1}^{p} b_k x_k + b_0\right] \tag{4}$$

ここで、分布の仮定は $\lambda_0(t)$ の中に入るが、Coxの比例ハザードモデルでは、$\lambda_0(t)$ を無視し、分析に使用する変数の影響だけを論じる。

ワイブルハザードモデルなどは $\lambda_0(t)$ の関数形を仮定し、変数の影響だけでなく、各個体の生存関数を直接推定できる。

今、i 番目の変数 $x_i = \{0,1\}$ だけが異なる人がいたとする。そのときの、$x_i = 0$ のタッチタイピングができない人のハザード $\lambda(t\,|\,\mathbf{x}_0)$ と、$x_i = 1$ のタッチタイピングができる人のハザード $\lambda(t\,|\,\mathbf{x}_1)$ の違いを比較する。

比例ハザードモデルを用いると、タッチタイピングができる・できないの合格可能性の違いを調べることができ、ワイブルハザードモデル等ではその人の不合格が続く状況を比較することもできる。

College Analysisでの生存時間分析のメニューを図5に示す [2] 。

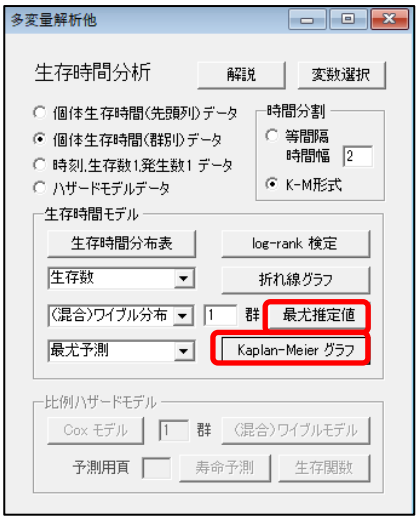


図5　生存時間分析実行メニュー画面

ロジスティック回帰分析と同じデータを使用する。「合格回数」(何回目の授業で合格したか)について、ワイブル分布予測付きのKaplan-Meierグラフを図6に示す。

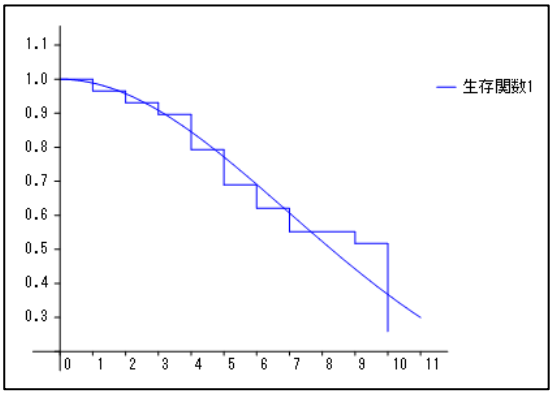


図6　「合格回数」のKaplan-Meierグラフ

これより「合格回数」は、「ほぼ、ワイブル分布で推定できる」ことが分かる。
この推定曲線の式（生存関数）$S(t)=\exp\left[-(t/b)^{a}\right]$のパラメータa,bの推定値を図7に示す。

ワイブル分布最尤推定結果

	推定値	標準偏差	2.5%下限	2.5%上限	最小2乗推
▶ 合格回数	R	0.966	R^2	0.933	
a	1.948	0.401	1.163	2.734	1.476
b=exp(-β/a)	9.995				10.853
β	-4.485	0.909	-6.268	-2.703	-4.646

図7　ワイブル分布最尤推定値結果

次に、タッチタイピングができる・できない、で分けた「合格回数」についてのKaplan-Meierグラフを図8に示す。

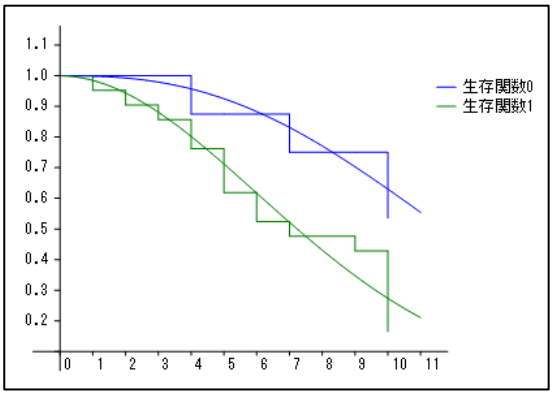


図8　タッチタイピングができる・できない、で分けたKaplan-Meierグラフ

これらから分かることは、それぞれのグラフがワイブル分布でほぼ推定され、生存関数に違いがみられるかもしれないということである。そのため、2つのグラフに違いがあるかどうか、log-rank検定を使って調べたところ、p=0.087という値であった。これは、有意な差ではないが、今後データ数を増やすことにより有意差が生じる可能性を期待させる。

次に、どの変数が「合格回数」に対して最も効いているのか、ハザードモデルを使って調べる。目的変数を「合格回数」、説明変数を「初期値、HP、B自己」として、Coxの比例ハザードモデルを使った場合の結果を図9に示す。

Cox比例ハザードモデル結果

	偏回帰係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
▶ 初期値	0.0242	0.0077	0.0016	0.0091	0.0392	1.0245
HP	0.6101	1.0944	0.5772	-1.5350	2.7551	1.8406
B自己	0.6099	0.6554	0.3520	-0.6746	1.8944	1.8403

図9　Cox比例ハザードモデル結果

これから、ハザードモデルは、初期値に大きく依存していることが分かる。Exp(b)の値から、タッチタイピングができる・できないの合格可能性は平均1.84倍違う。しかし、データ数が少ないので、この段階では「HP」と「B自己」は、はっきりと影響しているとは言えない。ワイブルモデルを使った結果を図10に示す。

ワイブル比例ハザードモデル結果

	偏回帰係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
▶ 初期値	0.0265	0.0078	0.0007	0.0112	0.0418	1.0268
HP	0.5394	1.0995	0.6237	-1.6155	2.6944	1.7151
B自己	0.5952	0.6592	0.3666	-0.6969	1.8872	1.8133
切片	-8.4697	1.5246	0.0000	-11.4578	-5.4815	2.097E-04
a	2.2640	0.4179	0.0000	1.4449	3.0831	

図10　ワイブル比例ハザードモデル結果

両側確率の値から、ハザードは初期値に大きく依存しており、Exp(b)の値から、タッチタイピングができる・できないの合格可能性は平均1.81倍違う（Exp(b)の値）。但し、データ数が少ないので、この段階では「HP」と「B自己」ははっきりと影響していると言えない。

次に、個人の生存関数を図11に表示する。

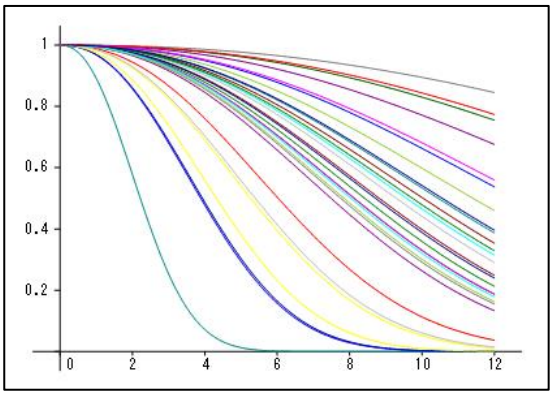


図11　個人の生存関数のグラフ

このグラフの各時点での個人の平均は、図12で見るようにKaplan-Meierグラフの近似になっている。

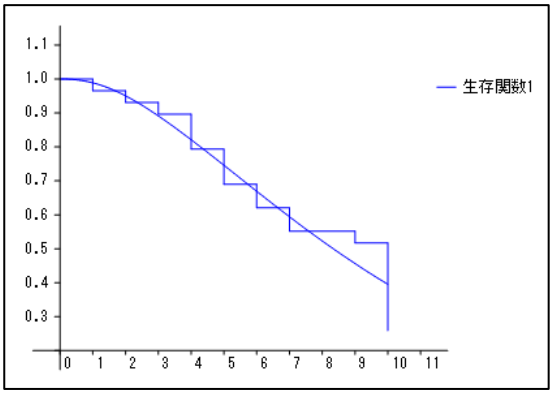


図12　K-Mグラフと平均生存関数による個人の平均

最後に、特別なデータを使って予測の生存関数を描いてみる。予測するためのデータは、図13のようなデータである。

	初期値	HP	B自己
1	50	1	1
2	50	1	0
3	100	1	1
4	100	1	0

図13　予測するためのデータ

このデータを用いて描いた生存関数を図14に示す。

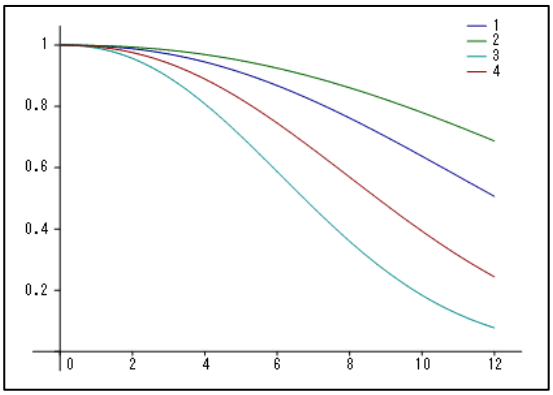


図14　予測データを使用した生存関数のグラフ

このグラフから、初期値が50の場合と、100の場合について、タッチタイピングができる・できないで分けた場合の生存関数の差が視覚化される。

ここまで、ワイブルハザードモデルを用いて、生存関数を求めてきたが、College Analysisでは、混合ワイブルハザードモデルを用いることもできる。

そこで今度は、2種混合ワイブルモデルを使って図14で与えられたデータに対して生存関数を予測してみた。結果を図15に示す。

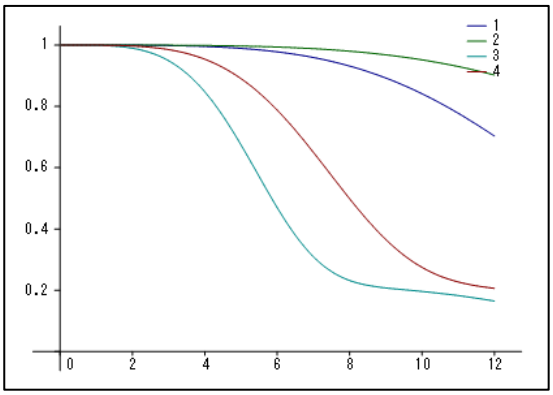


図15　特別なデータを用いた生存関数のグラフ

これによると、初期値によって生存関数は大きく異なるが、全体の適合度としてワイブルモデルから著しく改善されたとは言い難いので、ここではその結果のみを示すことにする。

４．おわりに

今まで、感覚的に、早期にタッチタイピングを習得すれば合格のタイミングも早くなるだろうと考えてやってきた。今回の、この分析で、それが肯定される可能性が表されたのは、今後タッチタイピングを指導していくうえでの重要事項の一つになると思われる。

キーの位置を早く覚えて、キーボードを見ずに打つタッチタイピングを修得すれば、コンピュータを使用する授業の多い本学科では、良い武器となり、残りの大学生活3年半で学べる量にも関わってくるはずである。今後、より一層力を入れて、タッチタイピングの修得、打鍵スピードの向上に向けて指導していきたいと思う。

本論文での分析については、本学科　福井正康 教授のご意見を伺いました。ここで謝辞を述べさせていただきます。有難うございました。

参考文献

［１］ 福井正康、小玉一樹、尾崎誠、「社会システム分析のための統合化プログラム28
－メタ分析・ロジスティック回帰分析－」、福山平成大学経営研究、第13号、(2017)

［２］ 福井正康、呉暁那、「社会システム分析のための統合化プログラム31
－生存時間分析－」、福山平成大学経営研究、第13号、(2017)

［３］ 奥田由紀恵、細川光浩、福井正康、「タイピングにおける速度変化のパターン分析」
福山平成大学経営研究、第6号, pp. 45-56、(2010)

［４］ 奥田由紀恵、細川光浩、福井正康、「タイピング速度向上の統計的検証」、日本教育情報学会第26回年会論文集、pp.264-265、(2010)

［５］ 奥田由紀恵、修士論文「4年間の記録データから検証する最良のタッチタイピング修得方法」、福山平成大学大学院、(2015)

［６］ 奥田由紀恵、福井正康、「情報処理教育初期段階におけるタッチタイピング修得法について」、日本教育情報学会第31回年会論文集、pp.214,215、(2015)

［７］ 奥田由紀恵 「情報教育初期段回におけるタッチタイピング習得法」、福山平成大学経営学部紀要、第12号、(2016)

情報リテラシー初歩教育段階のタイピング修得練習における
タッチタイピングの早期修得による効果について

Effect of Quick Touch-Typing Learning in the Information Literacy Beginning Stage

Yukie OKUDA

Department of Business Administration, Faculty of Business Administration, Fukuyama Heisei University

Abstract: In information literacy education, what should be acquired at an early stage is touch typing technology. I will use logistic regression analysis and survival analysis to study whether or not to learn touch typing in the early stage of training process is effective to reach the score that the department places as a passing line. And with the findings I will explore the possibility that the method of typing skill I have done is positive.

Key Words: Information Education, touch-typing, Life analysis

福山平成大学経営学部紀要
第 13 号（2017），33－42 頁

HTML5 および Javascript による 幾何アニメーション表示

尾崎 誠

福山平成大学経営学部経営学科

要旨：これまで Javascript により HTML5 上で動作する 3D Graphics Viewer[1]というプログラムを開発し、それを利用して 2 変数関数グラフや 3 次元パラメータ関数を Web ブラウザ上で 3 次元表示させることを可能にした。今回は、新たに幾何アニメーションの描画プログラムを作成した。これは図形を描き動かすことによって、数学的な関数や座標の概念を学ぶことができるプログラムである。本プログラムは、物理学シミュレーションにも応用可能であり、将来的にはシミュレーション言語に発展させる予定である。

キーワード：HTML5、Javascript、Web ブラウザ、アニメーション

1. はじめに

これまで、社会システム分析のための統合化プログラムであるCollege Analysis[2]をJavascriptによりHTML5上で動作するよう移植を進めてきた。今回は、数学的な関数や座標の概念をグラフィックの形状や動きによって学ぶことのできる2次元の幾何アニメーションについて紹介する。今回は2次元のみ対応であるが、将来的には3次元表示にも対応させる予定である。

幾何アニメーションは簡易的なマクロ言語によって描画されるが、言語仕様の作成には可能な限りの簡潔さを心掛けた。また、コマンド数もサンプルを考えながらできるだけ減らし、中学生程度の知識でも、かなりの描画が可能になるように努めた。今後もアニメーションの機能向上を図っていくが、マクロ言語の仕様はできるだけ上位互換になるように発展させる予定である。

プログラムデータは、1つのプログラムを1ページとし、パソコンであればローカルに保存されたデータの読み込みを可能としている。その際のデータフォーマットはCollege Analysisと同じとした。そのため、これまでにCollege Analysisで作成してきたデータを利用することができる。タブレット端末やスマートフォンの場合には、ローカルに保存されたデータの読み込みができないので、オンラインストレージサービスなどとの連携を視野に入れて研究を進めている。

2. 幾何アニメーション

幾何アニメーションプログラムを起動すると、図2.1のような画面が表示される。

図2.1　幾何アニメーション実行画面

デフォルトでの描画範囲は、$-4 \leq x \leq 4$，$-4 \leq y \leq 4$としており、Webブラウザ上では400×400ドットの描画範囲に正規化して描画を行う。プログラムは中央のテキストボックス内にコマンドを用いて記述する。

コマンドを用い、プログラムを記述または「ファイル選択」ボタンをクリックすることで、ローカルフォルダからプログラムデータを読み込み、「スタート」ボタンをクリックするとアニメーションが実行される。アニメーション実行後、「スタート」ボタンは「リセット」ボタンに置き換わり、「リセット」ボタンをクリックするとアニメーションが停止され、

データは初期状態に戻される。さらに、「リセット」ボタンは「スタート」ボタンに置き換わり、再度「スタート」ボタンをクリックすれば、アニメーションを初期状態から再描画できる。「停止」ボタンをクリックすると、アニメーションが一時停止し、「停止」ボタンは「再開」ボタンに置き換わる。「再開」ボタンをクリックすると、アニメーションが再実行され、「再開」ボタンは「停止」ボタンに置き換わる。

「デモ」ボタンはデモンストレーションを開始するボタンであり、クリックするとデモンストレーションが実行される。デモンストレーションは、1つのプログラムのアニメーションを12秒間実行し、プログラムが複数ある場合には順番に切り替えながらアニメーションを行う。デモンストレーション実行後、「デモ」ボタンは「解除」ボタンに置き換わり、「解除」ボタンをクリックするとデモンストレーションは終了し、「解除」ボタンは「デモ」ボタンに置き換わる。

「ページ数」の右横に配置されている「<」ボタンと「>」ボタンは、プログラムのページが複数ある場合に実行するプログラムの切り替えを行うボタンである。「現在のページ」に今選択されているプログラムのページ数が表示され、そのプログラムがテキストボックス内に表示される。「ページ数」は読み込んだファイルの最大のページ数が表示される。

「<」ボタンをクリックすると、現在のページ数が1つ減少する（最小値は1）。また、「>」ボタンをクリックすると、現在のページ数が1つ増加する（最大値は最大のページ数）。

プログラムのファイルから読み込んで使用する場合、そのファイルフォーマットは以下の通りである。

1行目	プログラムの数，最大の行数，最大の列数
2行目	1番目のプログラムの行数，1番目のプログラムの列数
3行目	1番目のプログラムの名前（空白でもよい）
4行目	コマンド
	:
m行目	2番目のプログラムの行数，2番目のプログラムの列数
m+1行目	2番目のプログラムの名前
m+2行目	コマンド
	:
n行目	i番目のプログラムの行数，i番目のプログラムの列数
n+1行目	i番目のプログラムの名前
n+2行目	コマンド
	:

次に、プログラムで利用できる各コマンドについて少し詳しく説明する。[]で囲まれたパラメータは省略可能であり、パラメータ間の空白はあってもなくてもよい。

描画範囲変更

rangex xmin，xmax（rangey ymin，ymax）

range min，max

描画範囲を一時的に変更する。メニューの描画範囲を表すテキストボックスは変化しない。2番目のように、rangeだけであれば全ての軸の描画範囲を同一に変更する。

時間変数（time=0，0.1，0.2，…）

time

時間を表す予約変数で、0，0.1，0.2，…（秒）と増加していく、アニメーションの基本となる変数である。描画は0.1秒ごとなので、簡単な動画で描画が追いつけば、実時間となる。また、描画が追いつかずコマ落ちが発生した場合には、コマ落ちに合わせて時間変数も変化するようにしてあり、その時にもなるべく実時間に合うようにしてある。

定義

define 変数名 = 定義式

変数名を定義式で置き換えることができる。プログラムの実行の最初に置換処理を実行するので、描画スピードのロスは少ない。

実行の繰り返し

playback 時間（秒）

実行メニューの「スタート」ボタンをクリックした際、またはデモンストレーションを実行している場合には、各プログラムの開始時の描画の実行時間を決める。その描画時間が終了したら再描画が繰り返される。

球体

ball(x，y)，r，color [，tr]

座標（x，y）を中心にした半径rの立体的な球を描画する。色を指定するcolorは&Hffaa80のように16進数で指定する。trは不透明度（0～1）で与える。trは省略可能であり、デフォルトで1（不透明）である。colorとtrについては、他でも同じように使用する。

円・楕円

circle(x，y)，r，color [，ratio，theta，tr]

座標（x，y）を中心にした、横半径r、色color、縦横比ratioの楕円を描画する。縦横比は、縦/横で与え、デフォルトは1である。thetaは回転角（度）、trは不透明度である。

三角形

tri(x1，y1)-(x2，y2)-(x3，y3)，color [，tr]

3点の座標(x1，y1)，(x2，y2)，(x3，y3)を繋ぐ、色colorの3角形を描画する。trは不透明度である。

正多角形

poly(x，y)，r，n，color [，theta，tr]

座標（x，y）を中心にした、半径r、色colorのn角形を描画する。thetaは回転角（度）、trは不透明度である。

連結

connect(x1，y1)-(x2，y2)，color

2点の座標(x1，y1)，(x2，y2)を繋ぐ、色colorの線分を描画する。

関数

func y=f(x)，color

色colorのy=f(x)の関数を描画する。x=f(y)の関数形も描画することができる。

パラメータ曲線 [0 <= u <= 1]

param1(x(u)，y(u))，color [，div]

予約パラメータ「u」で指定するパラメータ関数を色colorで描画する。パラメータが動く範囲は0 <= u <= 1である。divはパラメータuの0から1までの分割数である。Divは省略可能でありデフォルトでは200である。

文字列

print(x，y)，”文字列”，color，size

座標（x，y）の位置に指定された文字列（” “でくくる）を色colorで描画する。文字のフォントサイズはsizeで指定するが、デフォルトで9pointである。

コマンドを用いてプログラムを記述し、アニメーションを実行した結果を図2.2に示す。1行目は半径2の青色で半透明な（不透明度0.2）六角形を描くコマンドで、2行目は半径1 .5の緑色で半透明な（不透明度0.2）五角形を描くコマンドである。2つの多角形はtimeコマンドによって回転角を逆方向に変化させ回転している。

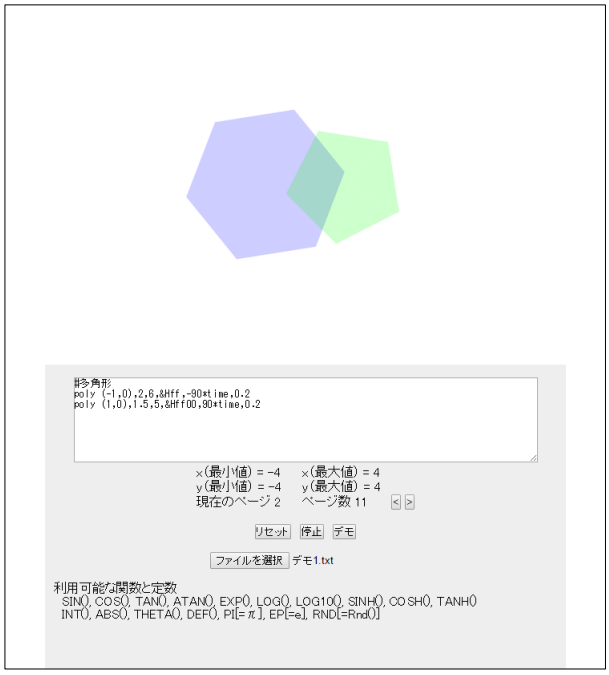


図2.2　多角形の回転

これ以外にも、プログラムの詳細は記載しないが、実行結果を図2.3、図2.4および図2.5に示す。いずれも、多角形の回転の描画のように、簡単なプログラムの記述により幾何アニメーションの描画を可能としている。

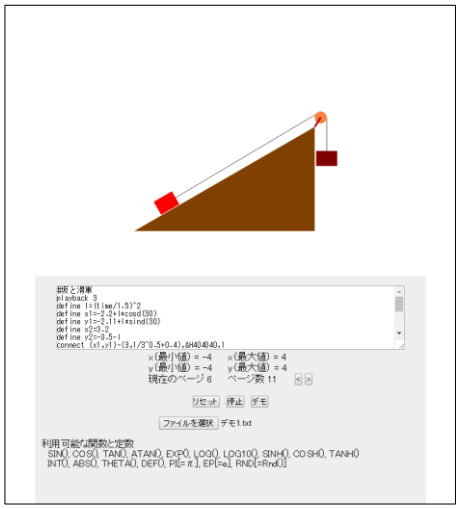


図2.3　坂と滑車

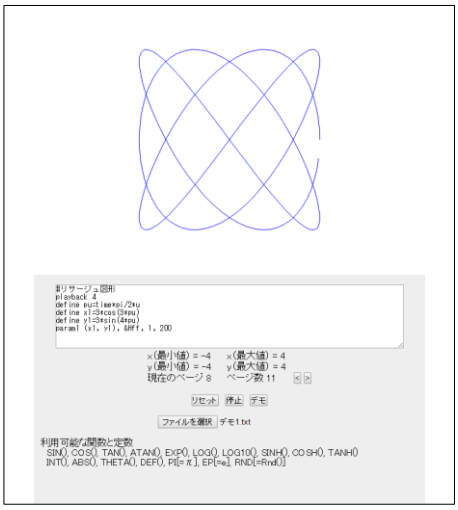


図2.4　リサージュ図形

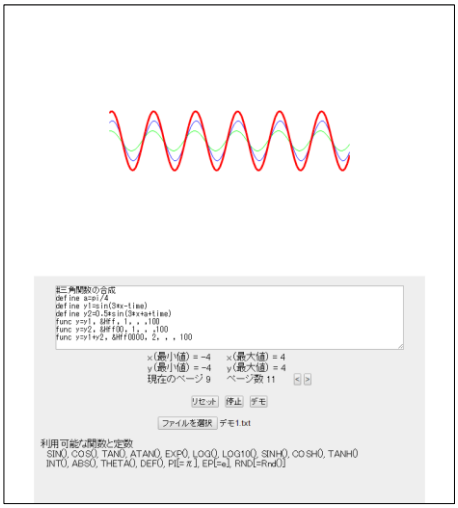


図2.5　三角関数の合成

3. おわりに

いくつかの例を用いて、幾何アニメーションのプログラムを説明してきたが、ここでの使用例はほんの一例であり、応用範囲は広い。数学の関数や座標の概念の把握、1 次変換やベクトルについての学習、物理教材作り、簡単な物理シミュレーション、さらに将来的には本格的な物理シミュレーションと可能性は大きく広がっている。

プログラムに使用できるコマンドは、基本的なものだけを作成している。本格的なアニメーション表示を行うには、さらにコマンドを追加していく必要がある。これについては、現在も追加作成中である。2 次元幾何アニメーションでは、物理教材の作成も視野に入れているため、よく利用される記号も含めておく必要がある。例えば、電流計や電圧計、電池、コンデンサ、スイッチなどである。これらについても追加作成していく必要がある。

また、現在は 2 次元表示のみ対応しているが、3 次元表示についても追加する予定である。その際、現在ではブラウザ上の描画に汎用性を重視して Canvas を使用しているが、今後幾何シミュレーションにも拡張することを考えると、より高速描画が可能な WebGL を使用した描画への切り替えを検討中である。

今回、ファイル読み込みの機能を持たせ、プログラムを外部から読み込むことを可能とし、そのファイルフォーマットを College Analysis のものと共通とした。そのため、これまで作成してきた教材用のファイルが利用できるようになっただけでなく、これまで College Analysis を用いて講義を受けた学生にとって、簡単にプログラム環境の移行ができるようになった。

その一方で、タブレット端末やスマートフォンでの利用では、データファイルの読み込みについて課題が残っており、これについて早急に対応する必要がある。例えば、Dropbox や OneDrive などのクラウドサービスが利用できるように対応するなどを検討する必要がある。本学では、Office365 の利用環境が整っているので、OneDrive への対応が望ましいように思えるが、出来るだけ利用者の負担が少なくなるよう慎重に検討していきたい。

このプログラムでは、これまで仕様を主体に考えてきたため、デバッグ機能は作成していない。しかしこのままでは、実際に初心者が利用するにはマクロ作成が難しいであろうと思われる。そのため、まずは簡単に作成できるところからデバッグ機能を追加し、実際に使用して利用者が間違えるところを見極め、順次整備していく必要がある。これについても、今後の課題である。

今回、コマンド解析および描画については、Javascript のライブラリとして、本体とは別の JS ファイルとして作成した。そのため、他のプログラムからも呼び出して利用することができる。他にどのような利用方法があるかについても検討を行っていきたい。

描画については、今回は公開されている様々なライブラリを利用することなくプログラムを作成した。しかし、プログラミング作業の負担や描画速度などを比較検討し、今後の作成方針についても十分検討する必要性が出てきた。特に、3 次元表示への拡張を考えると、それぞれのメリット、デメリットについて十分に比較検討していく必要がある。

参考文献

[1]　HTML5 および Javascript によるグラフィックスビューアとその応用、尾崎誠、福井正康、福山平成大学経営研究、第 11 号、(2015)、27-32

[2]　社会システム分析のための統合化プログラム２２　－幾何アニメーション－、福井正康、福山平成大学経営研究、第 10 号、(2014)、79-104

Geometric Animation with HTML5 and Javascript

Makoto OZAKI

Department of Business Administration、 *Faculty of Business Administration*、
Fukuyama Heisei University

Abstract: I have developed a program called 3D Graphics Viewer which runs on HTML5 by Javascript so that I can make it possible to display two-variable function graph and three-dimensional parameter function by three-dimensionally on Web browser using it. For this time, I created a geometric animation drawing program. This is a program that allows you to learn concepts of mathematical functions and coordinates by drawing figures. This program can also be applied to physics simulation and will be developed into a simulation language in the future.
Key Words: HTML5, Javascript, Web browser, animation

福山平成大学経営学部紀要
第 13 号(2017)，43－58 頁

パートタイマーの基幹労働力化が組織市民行動に及ぼす影響

小玉 一樹[*1]

[*1] 福山平成大学経営学部経営学科

要旨：本研究の目的は、パートタイマーの基幹労働力化と雇用環境が組織市民行動に及ぼす影響を明らかにすることであった。その際、パートタイマーの生計状況に着目し、サンプルを生活維持群と家計補助群に分類し探索的に分析を行った。小売業のパートタイマーから得られた 962 データを対象に階層的重回帰分析を行った結果、基幹労働力化は組織市民行動のうち“職務上の配慮”と“対人支援”に有意な正の影響を及ぼしていた。また、雇用環境が組織市民行動に及ぼす影響は、生活維持群と家計補助群とでは相違していることが明らかになった。さらに、基幹労働力化と雇用環境の交互作用が組織市民行動に及ぼす影響についても違いがみられた。

キーワード：パートタイマー、基幹労働力化、組織市民行動

1. 問題

近年、わが国では雇用の短期化や多様化がみられる。総務省が発表した平成28年4月～6月期の労働力調査によると、我が国の全雇用者のうち非正規社員が占める割合は36.8%となっている。非正規社員のうち、パート・アルバイトは68.9%を占め、年々増加傾向にあるとされる。パートタイマーの就業率が高い小売業・卸売業などでは、職務の標準化やマニュアル化などによって、これまで多くの雇用を生み出してきた。今では、正社員が担っているような判断業務や管理業務などを遂行するパートタイマーの存在も認められている。このようなパートタイマーを取り巻く労働環境の変化によって、彼らは重要な戦力として企業経営に不可欠な存在となっている。

非正規労働者が増加する一方で、正規労働者と比較した賃金格差をはじめとするさまざまな社会問題が生じている。多様な働き方が広がっていく中、このような社会問題がこれら労働者にとって著しい不均衡や処遇条件の低下が懸念されるため、1993年に短時間労働者の雇用管理の改善等に関する法律 (以下、パートタイム労働法) が制定された。このパートタイム労働法は、パートタイマーがその有する能力を一層有効に発揮することができる雇用環境を整備するため、それら労働者の納得性の向上や正社員との均等・均衡待遇の確保、正社員への転換の推進等を図ることを目的としている (厚生労働省，2015)。このパートタイム労働法が雇用主に求めているのは、1) 労働条件の文書交付・説明義務、2) 均

等・均衡待遇の確保の促進、３) 通常の労働者への転換の推進、４) 苦情処理・紛争解決援助、５) 実効性の確保である。しかし、平成26年度都道府県労働局雇用均等室でのパートタイム労働法施行状況では、改正法に対する相談が昨年より大幅に増加しており、これらの定着が難しい現状が伺える。

本研究では、パートタイマーの働き方に直接的に影響を及ぼすと考えられる、２) 均等・均衡待遇の確保の促進、３) 通常の労働者への転換の推進を採り上げる。具体的には、賃金、通常の労働者への転換を含む昇進昇格、教育訓練であり、これらの企業での取り組みがパートタイマーのパフォーマンスにどのような影響を及ぼすのかについて検討する。

既に述べたとおり、パートタイマーといえども正社員が担っていたような判断業務や管理業務などを遂行するパートタイマーの存在が認められている。また、パートタイマーの多くが主婦層であることから家計の補助を目的とした働き方なのか、世帯主などとして生活の維持を目的とした働き方なのかによっても職務行動には違いがあるであろう。こうした違いを一括りのパートタイマーとして研究を進めることは、その働き方の違いによる重大な相違を見落としてしまう可能性がある。本研究では、パートタイマーの職務内容と家計に着目し検討を進めることにした。

1.1　パートタイマーとは何か

パートタイマーという労働者のイメージは、人々がパートタイマーに抱く印象によって形成されており、その捉え方はさまざまである。たとえば、「当該事業場の一般労働者に比して労働時間が短い者」という労働時間の相対的な短さによる定義がある。この定義は、各省庁の統計調査でみられる他、「短時間労働者の雇用管理の改善等に関する法律」でも、「一週間の所定労働時間が同一の事業所に適用される通常の労働者の一週間の所定労働時間に比べ短い労働者」(パート労働法２条)と定義されている。

これに対して、総務省統計局の「労働力調査特別調査」では、「パート、アルバイト」の定義を「就業の時間や日数に関係なく、勤務先で『パートタイマー』『アルバイト』又はそれらに近い名称で呼ばれている者」としており、パートタイマーとは職場での呼称が重要な意味を持つものと考えられる。研究にあたっては、これらの人材を「パートタイマーという呼称の労働者」とすることはその実態がつかめないため、本稿ではパートタイマーを正社員と比較した所定労働時間の相違として捉え、「当該事業場の一般労働者に比較して労働時間が短い者」と定義し、議論を進めるものとする。

1.2　パートタイマーの働き方

パートタイマーが数多く働いている飲食業や小売業などでは、業務やオペレーションの標準化によって、パートタイマーの量的拡大に貢献してきた (本田，2001)。しかし、現在では、これらの業種では正規労働者が減少し、正規労働者が担ってきた管理業務などもパートタイマーに移行されるなど、これまで定型的かつ補助的な業務の担い手とされてきたパートタイマーの働き方が変化していることが指摘される (武石，2002)。このような現状からは、パートタイマーが組織運営に欠かせない専門業務や管理業務などの中心的役割を

担う時代に突入したことを意味している (村上，2004)。本田 (2001) は、パートタイマーが中心的な役割となることを、単にパートタイマーが量的に基幹労働力となることである量的基幹労働力化に対して、パートタイマーの質的基幹労働力化あるいは質的基幹化として区分している。今では、パートタイマーの基幹労働力化といえば、質的基幹労働力化を指すまでになっている (以下、「基幹労働力化」という)。

パートタイマーの基幹労働力化は、人材需給が逼迫する中で限られた人材を有効に活用することを目的としており、企業にとっては正規労働者の削減にともなうコストの軽減をもたらすことから、多くの企業ではパートタイマーの質的向上を求める向きが多い。しかし、雇用される側であるパートタイマーは、このような企業の取組みをどのように捉えているのであろうか。正規労働者と同じ職務でありながら、賃金に明らかな格差がある雇用形態であるパートタイマーには不満が生じ、その結果、職務行動も低下する可能性もある。

他方、パートタイマーの働き方を検討する上で、家計との関係を見逃すことはできない。パートタイマーとして働いているのは、主婦層を中心とした女性労働者であることが多い。これには我が国の税制が大きく関係している。我が国の税制は妻などの扶養家族がいれば夫の収入から一定額が控除され、それに課税されるという配偶者控除という仕組みがある。このため、妻が働く場合には、この控除を目当てに一定の所得範囲以内に抑え込もうとする限定的な働き方となっていることが多い。その結果、働く時間は短くなり、収入は一定額を超えないため、生計の多くは配偶者の収入に依存することとなる。一方で、独身者や配偶者控除の限度を超えパートタイマーとして働くことで得た収入で生計を立てている者もいる。このような働き方の違いは、同じパートタイマーという雇用形態であったとしても、それぞれの職務態度や職務行動に及ぼす影響には違いがあることが予想される。

1.3　組織市民行動

パートタイマーのパフォーマンスをどのように捉えることができるか。本研究では組織市民行動 (Organizational Citizenship Behavior) に着目している。組織市民行動とは、組織の成員が組織から報酬を期待せず組織のために行う貢献的行動のことである。Katz & Kahn (1966)は、「組織の機能が与えられた職務として規定されていないし要求もされていない数多くの行動に依存し、しかもそれらの行動が組織の社会的機能を円滑にする」と述べており、こうした貢献的行動の重要性を指摘している。

組織市民行動という概念の先駆的研究者である Organ (1988)は、組織市民行動を「組織市民行動は自由裁量的で、直接的ないし明確に公式的な報酬システムでは認識されておらず、全体として組織の有効的機能を促進する個人的な行動である」と定義し、愛他主義 (altruism)、誠実さ (Conscientiousness)、スポーツマンシップ (Sportsmanship)、礼儀正しさ (Courtesy)、市民の美徳 (Civic virtue)の５つの下位概念によって構成される概念であるとした。愛他主義とは、組織に関連する課題や問題を抱えている特定の他者を援助する効果のある任意の行動である。誠実さとは、出勤、規則への服従、休憩をとるといった点で、組織に関する最小限の役割要件をはるかに越えた従業員による任意の行動である。スポーツマンシップとは、従業員が理想的な環境でないことに不満を言うことなく我慢することを

厭わない行動である。礼儀正しさは、助言、情報伝達、具申といった仕事に関連した問題が他人に起こることを回避しようとして起こす任意の行動である。市民の美徳は、会社の生活に責任をもって参加あるいは関与しているか、それを気にかけている人が行う行動であるとそれぞれ定義している。我が国においては、田中 (2002, 2004)が、「日本語版組織市民行動尺度」の開発を行い、組織市民行動が、1) 対人的援助、2) 誠実さ、3) 職務上の配慮、4) 組織支援行動、5) 清潔さの５つの下位概念から構成されるとしている。

以上のように、組織市民行動の下位概念の各項目をみると、組織成員がとる組織市民行動の一つひとつは小さな行動であることが分かる。しかし、それらが組織成員間の相互の行動を促進し、役割行動を円滑にすすめることができるため、組織全体のパフォーマンスにも影響を及ぼすと考えられる。そこで、本研究ではパートタイマーのパフォーマンスの指標として、組織市民行動を採り上げることにした。

2. 方法

2.1 調査対象

A 県および周辺地域に店舗展開する B 社のスーパーマーケット 47 店舗で働くパートタイマー1,258 名を対象として調査は行われた。期間内に回収された質問紙は 1,218 通であり、すべての質問に回答したもののみを有効回答とした結果、962 データが有効回答であった(有効回答率 76.5%)。

2.2 手続き

調査は 2012 年 10 月 1 日から 2012 年 10 月 30 日に実施された。対象組織の各店舗の従業員に対して趣旨説明を掲載した質問紙を配布し、無記名で回答を求めた。対象者にはプライバシー保護の確保を文章で説明し依頼した。調査協力は対象者の自由意志とし、質問の回答を得ることにより調査への同意が得られたものとみなした。質問紙は個々が封入した上で各店の回収ボックスに投函し、調査期間最終日に店舗ごとに回収を行った。

2.3 質問紙の構成

属性に関する質問　パートタイマーの基幹労働力化とは、パートタイマーが中心的な役割となることされる。すなわち、正社員職務への接近であると捉えることができる。そのため、性別、年齢、雇用形態、勤続年数などのデモグラフィック変数に加え、職場での基幹労働力化の程度を調査するために、「私の担当している仕事の範囲は正社員と同じだ」「私の担当している仕事量は正社員と同じだ」「私の仕事上の責任は正社員と同じだ」「仕事上で私に求められる能力やスキルは正社員と同じだ」という４つの質問を、「そう思わない」＝１から「そう思う」＝４の４件法で回答を求め、４項目の平均得点を基幹労働力化として分析に用いた。

Table1 は、有効回答者の年齢、勤務時間、勤続年数の平均値、さらに属性別に集計したものである。パートタイマーという職種に就く者は圧倒的に女性が多く、その生活は配偶者の収入に依存していることが特徴であると言えるであろう。

Table 1　**調査対象者の主なプロフィール**　（n = 962）

属性等	平均 （標準偏差）
年齢（歳）	44.13歳 （11.96）
勤務時間（日）	5.31時間 （1.40）
勤続年数（年）	4.5年 （4.71）

属性等	人数 （比率：%）
性別	男性 80名 （8.32%）， 女性 882名 （91.68%）
配偶者有無	配偶者あり 652名 （67.78%）， 配偶者なし 310名 （32.22%）
子供有無	子供あり 677名 （70.37%）， 子供なし 285名 （29.63%）
学歴	中学卒 54名（5.61%）， 高校卒 608名（63.20%）， 専門学校卒 108名（11.23%） 高専・短大卒 141名（14.66%）， 大学卒 50名（5.20%）， 大学院卒 1名（0.1%）
世帯の主たる収入	主に自分の収入で暮らしている： 268名（27.86%） 主に配偶者等他者の収入で暮らしている 694名（72.14%）

Table2 は、基幹労働力化に関する４つの質問の平均を求め、平均値以上を基幹化高群、平均未満を基幹化低群に分類したものである。さらに、それらの分類ごとの平均年齢、１日の所定労働時間、勤務年数の平均と標準偏差を表している。基幹労働力化低群と高群では労働時間や勤続年数にその違いがみられる。

雇用環境に関する設問　現在の職場の雇用環境に関する 14 の質問項目について「そう思わない」＝１から「そう思う」＝５の５件法で回答を求め、因子ごとの平均得点を分析に用いた。質問項目は、昇格・正社員化が「正社員化などの活躍の場が用意されている」他の４項目、給与の妥当性が「私は仕事に見合った十分な給与を得ている」他の３項目、教育研修の充実が「仕事で必要な技術や知識については、十分な教育・研修がある」他の２項目である。尺度の信頼性を表す Cronbach の α 係数は、順に.908、.825、.811 であった。

組織市民行動　組織市民行動の測定は田中 (2002) の尺度のうち、各因子の因子負荷量の高い順に３項目を選択し、その一部について、調査を行った組織の文脈に沿った語彙に置き換えて調査に用いた。質問項目は「多くの仕事を抱えている人の手助けをする」「職場では店舗やバックヤードはいつもきれいにし、汚さないように努める」「仕事中に必要以上に休息をとらないようにする」などの 15 項目について「そう思わない」＝１から「そう思う」＝５の５件法で回答を求め、因子ごとの平均得点を分析に用いた。

Table 2 **パートタイマーの質的基幹労働力化における高群と低群の属性**

	基幹化低群 (n =729)		基幹化高群 (n =233)	
変数名	*M*	*SD*	*M*	*SD*
質的基幹化	1.90	0.58	3.33	0.39
年齢（歳）	43.49	12.20	46.12	10.97
所定労働時間（時間/日）	5.16	1.35	7.21	1.01
勤務年数（年）	4.00	4.32	6.07	5.48

3. 結果

3.1 組織市民行動の因子分析結果

パートタイマーの組織市民行動の構造を確認するため、組織市民行動に関するデータについて探索的因子分析を行った。まず、組織市民行動に関する 15 項目について、平均値から標準偏差を減じた得点はすべて下限値以上であり、床効果を示す項目はみられなかった。また、平均値に標準偏差を加えた得点は、すべて上限値となる天井効果もみられなかった。つぎに、すべての項目を最尤法・プロマックス回転法による因子分析に用いた。共通性について.35 を基準として、それを下回る項目を分析から除外し、再度、最尤法・プロマックス回転による因子分析を行った。その結果、Table 3 に示すとおり３因子解が最適であると判断した。

第１因子は「職場では自分の身の回りをきれいに清掃する」「備品・消耗品を使いやすいように整理し、配置する」などの４項目で構成されていた。これらの項目は、田中(2002)の尺度では“清潔さ”という因子に含まれていたが、対象とした業種では店内や作業スペースの清掃は日常的に行われる必要があることを考慮し、本研究ではこの因子を“職務上の配慮”と命名した。第２因子は「仕事上のトラブルを抱えている人を進んで手助けする」「休んでいる人の仕事を代わりに手伝ってあげる」などの３項目で構成されており、同僚への支援行動であるため“対人支援”と命名した。第３因子は「仕事中は必要以上に休息をとらないようにする」「仕事中は無駄な会話で時間を潰さないようにする」などの３項目から構成されているため“誠実さ”と命名した。田中(2002)の組織市民行動に関する尺度は５因子構造を想定していたが、パートタイマーを対象にした本研究では３因子のみが抽出された。抽出されなかった２つの因子は、本来の“職務上の配慮”と“組織支援行動”であった。

Table 3 組織市民行動の因子分析結果 （最尤法 プロマックス回転） *N* =962

	項目		*F1*	*F2*	*F3*	共通性
10	職場では自分の身の回りをきれいに掃除する		**.840**	-.053	-.025	.639
5	職場では店舗やバックヤードはいつもきれいにし，汚さないように努める		**.724**	.043	-.028	.540
15	備品・消耗品を使いやすいように整理し，配置する		**.661**	.163	-.086	.523
13	自分の仕事に注意を行き届かせる		**.484**	-.038	.297	.442
6	仕事上のトラブルを抱えている人を，進んで手助けする		.036	**.742**	-.032	.566
1	多くのの仕事を抱えている人の手助けをする		-.070	**.701**	.102	.493
11	休んでいる人の仕事を変わりに手伝ってあげる		.213	**.543**	-.004	.468
7	仕事中に必要以上に休息をとらないようにする		-.112	.057	**.698**	.444
12	仕事中は無駄な会話で時間をつぶさないようにする		.202	-.140	**.642**	.511
2	不必要に仕事の手を休めないようにする		-.085	.139	**.608**	.387
因子寄与			3.017	2.392	2.178	
因子間相関		*F1*	－			
		F2	.560	－		
		F3	.509	.341	－	

因子分析で抽出された３つの因子について Cronbach の信頼性係数 α を算出したところ、それぞれ α =.82、 α =.75、 α =.70 となり、一定の信頼性が確認されたため以降の分析に用いることにした。

3.2　パートタイマーの組織市民行動の階層的重回帰分析の結果

パートタイマーの基幹労働力化と雇用環境が組織市民行動に及ぼす影響を明らかにするため、組織市民行動を従属変数とし、基幹労働力化と雇用環境を独立変数とした階層的重回帰分析を生活維持群と家計補助群別に行った。その際、各サンプルの所属する店舗の規模、年齢、性別、最終学歴、職務内容、所定労働時間、勤務年数による影響を統制するため、これらをコントロール変数とし分析を試みた。

まず、測定された全体の項目について平均値、標準偏差および相関係数を Table4 に、群別の相関係数を Table5 に示す。記述統計の結果によれば、極端に分布が偏っている項目はなく、また天井効果、床効果もみられないため、すべての項目を階層的重回帰分析に用いることにした。

つぎに、基幹労働力化および雇用環境を独立変数、組織市民行動の３つの因子を従属変数とした階層的重回帰分析を行った結果を Table6 と Table7 に示す。階層的重回帰分析にあたっては、ステップ１ではコントロール変数を投入し、ステップ２ではコントロール変数と基幹労働力化、昇進・正社員化、給与の妥当性、教育研修の充実を投入した。最後のステップ３では、ステップ２の変数に加え、基幹労働力化と雇用環境の交互作用項を投入した。

3.2.1　家計補助群の階層的重回帰分析の結果

まず、家計補助群の階層的重回帰分析のステップ２では決定係数 R^2's は.061～.122 (p <.01)となり、基幹労働力化は職務上の配慮 (β =.044, p <.01) と対人支援 (β =.060, p <.01) に対する正の関係性が有意であった。つぎに、昇進・正社員化は誠実さ(β =.151, p <.05) に有意な正の影響を及ぼしていたが、給与の妥当性は対人支援 (β =-.165, p <.01) に負の影響を及ぼしていた。さらに、教育研修の充実は対人支援 (β =.090, p <.05) に正の影響を及ぼしていた (Table 6)。

上記の主効果に加え、基幹労働力化と雇用環境に関する変数との交互作用項を投入したステップ３で、決定係数の変化量 ΔR^2 が有意となったのは従属変数を誠実さとした場合のみであった。そのうち、交互作用項が有意であったのは、基幹労働力化×給与の妥当性 (β = -.067, p <.01) 、基幹労働力化×教育研修の充実 (β =.004, p <.05) であった。以上の結果から、基幹労働力化と雇用環境に対する認識の高低によって各組織市民行動の得点がどのように変化するかを検討した。

Fig.1 および Fig.2 は、家計補助群における組織市民行動のうち“誠実さ”に対する基幹労働力化と雇用環境の影響を、基幹労働力化の高群 (+1SD) と低群 (-1SD) ごとに回帰直線を描いたものである。まず、誠実さに対する基幹労働力化と給与の妥当性の影響は、高群のみに有意な関係性が認められた (Fig.1)。この結果は、基幹労働力化高群では、現在の

Table 4 分析に用いた変数の平均と標準偏差および変数間の相関係数 （n =962）

		M	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13
1	店舗規模※1	2.789	.843	1.000												
2	性別※2	1.917	.276	-.111 **	1.000											
3	年齢	44.129	11.961	-.055 +	.198 **	1.000										
4	最終学歴※3	2.509	.990	.055 +	-.031	-.011	1.000									
5	職務※4	7.168	14.598	.021	-.058 +	-.022	-.008	1.000								
6	所定労働	5.311	1.395	.074 *	-.152 **	-.117 **	-.085 **	.054 +	1.000							
7	勤務年数	4.499	4.707	-.117 **	.070 *	.324 **	-.086 **	-.025	.171 **	1.000						
8	基幹労働力化	2.249	.817	-.084 **	-.010	.062 +	-.050	.008	.279 **	.225 **	1.000					
9	昇格・正社員化	3.026	.738	.094 **	-.152 **	-.086 **	-.035	-.005	-.077 *	-.049	-.112 **	1.000				
10	給与の妥当性	3.017	.784	.066 *	-.110 **	-.106 **	-.039	.005	-.118 **	-.098 **	-.115 **	.756 **	1.000			
11	教育研修の充実	3.110	.773	.029	-.137 **	-.166 **	-.038	-.011	-.025	-.082 *	-.047	.487 **	.505 **	1.000		
12	職務上の配慮	4.091	.684	.081 *	.125 **	.268 **	.010	.005	.072 *	.145 **	.170 **	.002	-.021	.019	1.000	
13	対人支援	3.440	.774	.059 +	.020	.050	-.006	.010	.152 **	.069 *	.240 **	.018	-.036	.084 **	.521 **	1.000
14	誠実さ	4.081	.856	.074 *	.050	.167 **	.099 **	-.014	-.050	.029	.026	.063 +	.025	-.014	.409 **	.300 **

** $p < .01$, * $p < .05$, + $p < .10$

※1） 店舗規模:「1＝小型, 2＝中型, 3＝大型, 4＝特大」

※2） 男性:1, 女性2

※3） 最終学歴:1＝中学卒, 2＝高校卒, 3＝専門学校卒, 4＝専門・短大卒, 5＝大学卒, 6＝大学院卒

※4） 職務:1＝事務, 2＝レジ, 3＝水産, 4＝食肉, 5＝青果, 6＝日配, 7＝ドライグロッサリー, 8＝惣菜, 9＝その他

Table 5 分析に用いた変数間の群別の相関係数

		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	店舗規模	1	-.093 *	-.053	.071 +	.018	.092 *	-.124 **	-.085 *	.061	.054	-.001	.089 *	.052	.084 *
2	性別	-.137 *	1	.326 **	.017	-.066 +	-.153 **	.123 **	.024	-.128 **	-.128 **	-.135 **	.162 **	.074 +	.119 **
3	年齢	-.060	.057	1	-.039	-.066 +	-.131 **	.327 **	.072 +	-.105 **	-.131 **	-.167 **	.243 **	.055	.168 **
4	最終学歴	.024	-.133 *	.053	1	-.017	-.073 +	-.138 **	-.054	-.028	-.009	-.021	.023	.010	.102 **
5	職務	.024	-.041	.068	.016	1	.059	-.021	-.021	-.050	-.030	-.008	.010	.016	-.009
6	所定労働	.000	-.005	-.105 +	-.065	.022	1	.177 **	.258 **	-.068 +	-.107 **	-.033	.043	.148 **	-.033
7	勤務年数	-.104 +	.024	.317 **	.029	-.034	.166 **	1	.225 **	-.048	-.115 **	-.082 *	.112 **	.063 +	.023
8	基幹労働力化	-.100	.005	.040	-.018	.061	.251 **	.222 **	1	-.117 **	-.118 **	-.062	.168 **	.228 **	.027
9	昇格・正社員化	.176 **	-.222 **	-.047	-.058	.088	-.068	-.049	-.089	1	.746 **	.484 **	-.033	-.024	.038
10	給与の妥当性	.104 +	-.137 *	-.054	-.119 +	.081	-.094	-.060	-.087	.775 **	1	.485 **	-.053	-.091 *	-.009
11	教育研修の充実	.107 +	-.168 **	-.166 **	-.082	-.016	.003	-.083	-.007	.495 **	.554 **	1	-.001	.030	-.048
12	職務上の配慮	.055	.111 +	.326 **	-.017	-.009	.124 *	.216 **	.167 **	.086	.059	.068	1	.505 **	.394 **
13	対人支援	.069	-.018	.037	-.032	-.008	.127 *	.078	.252 **	.119 +	.101	.216 **	.559 **	1	.258 **
14	誠実さ	.055	-.062	.167 **	.085	-.021	-.046	.045	.043	.115 +	.094	.071	.452 **	.414 **	1

** $p < .01$, * $p < .05$, + $p < .10$

（左下：生活維持群 n =268，右上：家計補助群 n =694）

Table 6 パートタイマーの組織市民行動を従属変数とした階層的重回帰分析の結果 （家計補助群 : n =694）

従属変数 / 変数名	職務上の配慮			対人支援			誠実さ		
	Step1	Step2	Step3	Step1	Step2	Step3	Step1	Step2	Step3
		β			β			β	
コントロール変数									
店舗規模	.085 **	.097 **	.096 **	.044	.063 +	.064 +	.091 *	.089 *	.085 *
性別	.356 **	.378 **	.371 **	.313 *	.312 *	.296 *	.311 +	.327 *	.305 +
年齢	.013 **	.012 **	.012 **	.003	.003	.003	.011 **	.011 **	.012 **
最終学歴	.023	.033	.028	.016	.021	.025	.085 **	.088 **	.081 *
職務	.001	.002	.002	.001	.001	.001	.000	.001	.001
所定労働	.038 +	.018	.017	.093 **	.060 *	.057 *	.000	-.004	-.007
勤務年数	.005	.001	.001	.003	-.003	-.004	-.002	-.004	-.005
基幹労働力化と雇用環境									
基幹労働力化		.044 **	.043 **		.060 **	.062 **		.013	.013
昇格・正社員化		.009	.016		.098	.090		.151 *	.158 *
給与の妥当性		-.061	-.063		-.165 **	-.166 **		-.062	-.068
教育研修の充実		.036	.041		.090 *	.094 *		-.052	-.038
交互作用項									
基幹労働力化×昇格・正社員化			.007			.048 *			.035
基幹労働力化×給与の妥当性			-.029			-.041 +			-.067 **
基幹労働力化×教育・研修の充実			.018			.001			.044 *
R^2	.087 **	.122 **	.130 **	.037 **	.090 **	.097 **	.053 **	.061 **	.085 **
Adjust R^2	.078	.107	.109	.027	.074	.076	.043	.045	.063
ΔR^2	.087 **	.035 **	.008	.037 **	.053 **	.008	.053 **	.009	.024 *
F値	9.391 **	7.913 **	6.316 **	3.741 **	5.588 **	4.554 **	5.434 **	3.697 **	3.919 **

** $p < .01$, * $p < .05$, + $p < .10$

Table 7　パートタイマーの組織市民行動を従属変数とした階層的重回帰分析の結果（生活維持群 : n =268）

変数名 ＼ 従属変数	職務上の配慮			対人支援			誠実さ		
	Step1	Step2	Step3	Step1	Step2	Step3	Step1	Step2	Step3
		β			β			β	
コントロール変数									
店舗規模	.082 +	.066	.062	.072	.061	.064	.058	.039	.034
性別	.185 +	.278 **	.295 **	-.033	.091	.103	-.128	-.037	.003
年齢	.017 **	.017 **	.017 **	.003	.003	.003	.012 **	.011 *	.011 *
最終学歴	-.011	.016	.014	-.025	.011	.011	.057	.088	.092 +
職務	-.001	-.002	-.002	-.001	-.002	-.001	-.002	-.003	-.003
所定労働	.073 *	.068 *	.074 *	.073 +	.050	.056	-.015	-.012	-.002
勤務年数	.013	.011	.012	.008	.002	.004	.000	-.002	.000
基幹労働力化と雇用環境									
基幹労働力化		.038 *	.031 *		.079 **	.070 **		.027	.023
昇格・正社員化		.077	.092		.048	.064		.064	.073
給与の妥当性		-.057	-.064		-.081	-.072		-.011	-.021
教育研修の充実		.047	.041		.171 *	.153 *		-.001	.002
交互作用項									
基幹労働力化×昇格・正社員化			.006			.018			.011
基幹労働力化×給与の妥当性			-.003			-.013			-.044
基幹労働力化×教育・研修の充実			-.041 +			-.058 *			-.010
R^2	.158 **	.221 **	.236 **	.028	.153 **	.182 **	.042	.080 *	.095 +
Adjust R^2	.136	.185	.187	.002	.113	.130	.017	.036	.037
ΔR^2	.158 **	.063 *	.015	.028 **	.125 **	.029 +	.042 **	.037 +	.015
F値	6.987 **	6.038 **	4.842 **	1.058	3.828 **	3.487 **	1.649	1.837 *	1.642 +

** $p < .01$, * $p < .05$, + $p < .10$

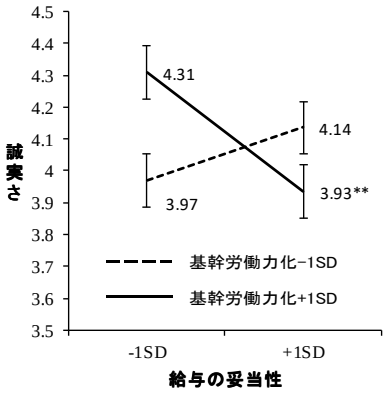


Fig. 1　家計補助群の誠実さに対する給与の妥当性と基幹労働力化の交互作用

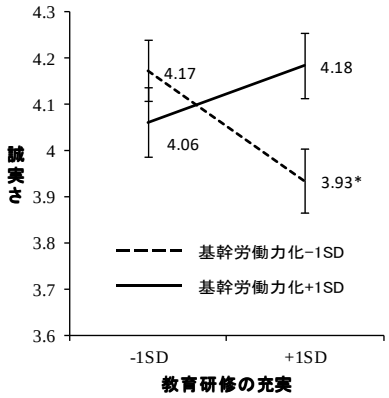


Fig. 2　家計補助群の誠実さに対する教育研修の充実と基幹労働力化の交互作用

給与水準が妥当ではないと感じているものほど誠実な行動をとる一方で、現在の給与水準に妥当性があると感じるほどに誠実な行動は抑制されるといえる。

つぎに、誠実さに対する基幹労働力化と教育研修の充実の影響は、基幹労働力化低群のみに有意な関係性が認められた (Fig.2)。この結果は、基幹労働力化低群では、組織の教育や研修が充実していないと感じているものほど誠実な行動をとろうとする一方で、教育や研修が充実していると感じるほどに誠実な行動は抑制されるといえる。

3.2.2　生活維持群の階層的重回帰分析の結果

生活維持群の階層的重回帰分析のステップ２では決定係数 R^2's は.080 (p <.1)〜.221 (p <.01)であった。基幹労働力化は職務上の配慮 (β =.038, p <.05) と対人支援 (β =.079, p <.01) に対する正の関係性が有意であった。つぎに、教育研修の充実は対人支援 (β =.171, p <.05) に正の影響を及ぼしていた。しかし、昇進・正社員化、給与の妥当性が組織市民行動に及ぼす影響は有意ではなかった (Table 7)。

上記の主効果に加え、基幹労働力化と雇用環境に関する変数との交互作用項を投入したステップ３では、従属変数を対人支援とした場合のみ決定係数の変化量が ΔR^2 =.029 (p <.1)となり有意な傾向がみられた。そのうち、交互作用項が有意であったのは、基幹労働力化×教育

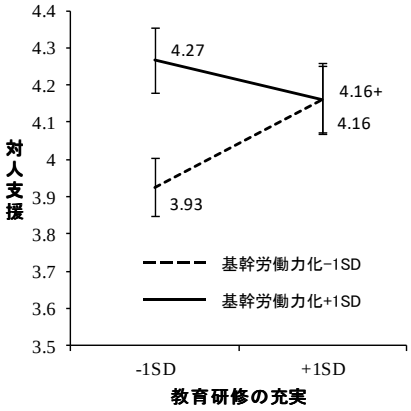


Fig. 3　生活維持群の対人支援に対する教育研修の充実と基幹労働力化の交互作用

研修の充実（β=.058, p<.05）のみであった。以上の結果から、基幹労働力化と雇用環境に対する認識の高低によって各組織市民行動の得点がどのように変化するかを検討した。

Fig.3 は、生活維持群における組織市民行動のうち“対人支援”に対する基幹労働力化と教育研修の充実の影響を、基幹労働力化の高群（+1SD）と低群（-1SD）ごとに回帰直線を描いたものである。対人支援に対する基幹労働力化と教育研修の充実の影響は、基幹労働力化低群のみに有意な関係性が認められた。この結果は、基幹労働力化低群は、組織内において十分な教育研修の機会があると感じているものほど対人支援行動をとる一方で、十分な教育研修の機会がない感じるほどに対人支援行動は抑制されるといえる。

4. 考察

本研究の目的は、パートタイマーの基幹労働力化と雇用環境が組織市民行動に及ぼす影響を明らかにすることであった。また、パートタイマーを生活維持群と家計補助群に分類し、その相違点を明らかにすることであった。以下では分析の結果について考察する。

まず、基幹労働力化はいずれの群においても、組織市民行動のうち職務上の配慮と対人支援に対して有意な正の関係性が認められた。パートタイマーの基幹労働力化は正社員業務への接近であり、判断業務や管理業務などが含まれることから、職務内容は高度化されると考えられる。こうした職務内容の変化は職務や職場に対する責任感を醸成させるため、組織貢献のためにこうした行動をとる傾向が高まると考えられる。

雇用環境では、昇格・正社員化が家計補助群において誠実さとの正の関係性が有意であった。一方、生活維持群ではいずれの組織市民行動とも有意な関係性はみられなかった。家計補助群は生活維持群と比較して、労働時間が短く収入も少ない。それでも、組織内において他者よりも高く評価されたいという自尊欲求が生じ、評価を伴う昇格や正社員化が誠実な行動に影響を与えたと考えられる。生活維持群に属するパートタイマーは既に基幹労働力となっている者も多いと考えられる。しかし、それらすべての成員が正社員となるわけではない。その意味から、彼らは正社員になることへの限界を感じ、誠実な行動に負の影響を及ぼすのかもしれない。

給与の妥当性に関しては、家計補助群において対人支援との負の関係性が有意であった。この結果は、組織内の他者への支援は自分の職務範囲ではないと考えている可能性がある。家計補助群は生活維持群に比較して、労働時間は短く賃金水準も低い。また、これらのパートタイマーの職務は主に定型的かつ補助的な業務であると考えられ、指示のあった範囲で働こうとしていることが考えられる。

教育研修の充実に関して、両群ともに対人支援に関する正の関係性が有意であった。組織内の教育は職務に関する知識を得ることができる機会であり、自分が得た知識や技能を他者に伝えようとする意識が高まり他者を支援するとも考えられる。

基幹労働力化と雇用環境の組合せによる組織市民行動への影響についても生活維持群と家計補助群とで相違が見られた。第一に、家計補助群において、組織市民行動のうち誠実さに対する基幹労働力化と給与の妥当性の影響が、基幹労働力化の高群のみに認められことである。この結果は、労働時間が短く賃金水準も低いと考えられる家計補助群であって

も、高度な職務を遂行するパートタイマーは、誠実に行動することで自分の評価を高めるためと考えることができる。なぜなら、組織市民行動そのものは公式的な報酬システムには直接的には影響を及ぼさないものの、上司の評価には間接的に影響を及ぼすと考えられるからである。逆に、給与水準に妥当性があると感じるほどに誠実な行動をとらないのは、自分自身への評価に満足しているため、積極的には誠実な行動をとらないものと推測される。

第二に、家計補助群において、誠実さに対する基幹労働力化と教育研修の充実の影響は、基幹労働力化低群のみに有意な関係性が認められた。この結果について考察すると、家計補助群における基幹労働力化低群の特徴は、短時間労働、低収入、定型的かつ補助的業務の担い手ということであろう。このカテゴリーのパートタイマーは労働時間に制約があり、社内において教育や研修が過度に充実していることは、作業時間の減少にもつながるため、誠実な行動は取り難くなると考えられる。作業時間と教育や研修に関する時間の関係は、社内教育や研修が不足していると感じ、労働時間が十分に確保できている者の方が誠実に行動するという結果からも妥当性はある。これらの層のパートタイマーには、過剰な教育や研修は組織市民行動に負の影響を及ぼすのかもしれない。

第三に、生活維持群において、対人支援に対する基幹労働力化と教育研修の充実の影響は、基幹労働力化低群のみに有意な関係性が認められたことである。このカテゴリーに属するパートタイマーは、自分の収入で生計を立てているが、職務は定型的かつ補助的な業務が中心であるため、それほど多くの収入は期待できないと考えられる。彼らが収入の水準を高めるためには、知識や技能を高め、それらを評価されることが必要となる。その意味から、組織内における教育や研修の機会は重要である。そうした機会を利用して得た知識が、積極的に他者を支援することにもつながると考えられる。一方、こうした機会に恵まれない者は、知識や技能を高めることができないため、他者への支援行動は積極的に行われないといえるであろう。

最後に本研究の課題について述べておきたい。まず、パートタイマーの基幹労働力化の問題である。パートタイマーの基幹労働力化は、パート業務の正社員業務への接近と捉えることができる。しかし、その判断は雇用する側と雇用される側で判断が分かれるところである。本研究では、その判断を雇用される側であるパートタイマーの判断に委ねている。つまり、本研究では実際に働いているパートタイマーが、正社員と同じような業務を遂行していると感じているかどうかを、基幹労働力化についての唯一の判断材料にしている。それでも研究論文である以上、一定の水準を維持する必要があるために、本研究では職務範囲、仕事量、責任、能力・技術の 4 つの要因について正社員業務との比較を行った上で、その平均点を基幹労働力化の度合いとしている。しかし、基幹労働力化の基準に違いがあれば、本研究で得られた結果も変わる可能性がある。

つぎに、生計をどのように立てているかについて、本研究では配偶者等の収入に依存して生計を立てているか、自分自身の収入で生計を立てているかによって、前者を家計補助群、後者を生活維持群に分類している。しかし、この設問では、夫婦共働き世帯では収入

の多い方が主たる収入となるため、必ずしも実態を表していないことも想定される。さらなる事態調査と設問を検討することで、これらの問題を回避できると考えられる。

引用文献

本田一成 (2001). パートタイマーの量的な基幹労働力化. 日本労働研究雑誌、 494.

本田一成 (2007). チェーンストアのパートタイマー. 白桃書房

本田一成 (2010). 主婦パート　最大の非正規雇用. 集英社新書

Katz, D., & Kahn、 R. L. (1966). The psychology of organizations. New York: HR Folks International.

厚生労働省(2014)平成 26 年度パートタイム労働法施行状況. http://www.mhlw.go.jp/topics/2007/06/tp0605-1n.html,2016.11.15.

村上忍 (2004). パートタイマー主役の店舗運営「仕事の再設計と処遇の仕組みづくり」への緊急処方せん: 次世代中心戦力の 「活用と能力開発」 の誤謬にこれだけは言わせろ! (緊急特集 疲労度を管理するリスクマネジメント 深夜そして 24 時間営業で従業員の健康と労働環境はどう変わった??) 食品商業, 33(14), 172-174.

Organ, D. W. (1988). Organizational citizenship behavior: The good soldier syndrome. Lexington Books/DC Heath and Com.

Organ, D. W., Podsakoff, P. M, & MacKenzie, S. B. (2006) Organizational citizenship behavior, （訳　上田泰）組織市民行動、白桃書房（2007）

総務省(2016)労働力調査(詳細集計)平成 28 年(2016 年)4～6 月期平均(速報) http://www.stat.go.jp/data/roudou/sokuhou/4hanki/dt/index.htm, 総務省統計局, 2016.10.25

武石恵美子 (2001). パートタイム労働者等非正規労働者の基幹労働力化の実態 (特集 パートタイム労働). Women & work, 137, 16-20.

田中堅一郎 (2002). 日本版組織市民行動尺度の研究. 産業・組織心理学研究, 15(2), 77-88.

田中堅一郎 (2004). 従業員が自発的に働く職場をめざすために: 組織市民行動と文脈的業績に関する心理学的研究. ナカニシヤ出版.

Making Part-time workers as "Core Labor Force"
- Its Effect on Organizational Citizenship Behavior -

Kazuki KODAMA[*1]

**1 Department of Business Administration, Faculty of Business Administration, Fukuyama Heisei University*

Abstract: This study aims to clarify the effect of making part-time workers as “core labor force” and the employment environment on their Organizational Citizenship Behavior (OCB). Focusing on their livelihood conditions, this study analyzed the collected samples by dividing them into a financially independent group and a financially dependent group. A total of 962 samples (independent group: n=268, dependent group: n=694) were collected from those who are part-time workers in retail stores, and a hierarchical multiple regression analysis was carried out.

The results showed that making part-time workers as “core labor force” has statistically a positive significant effect on “considerations while working” and “personal support”. In regards to the effect of work environment on OCB, it differed between the independent group and the dependent group. The results showed that the effect of interaction of the core labor force and employment environment on OCB also differed between the two groups. The theoretical and practical implications of this research are also discussed in the study.

Key words: Part-time workers, Core Labor Force, Organizational Citizenship Behavior

福山平成大学経営学部紀要
第 13 号（2017），59－68 頁

広島市役所のマネジメントの改善に関する研究
－幹部職員等のインタビュー調査結果に基づいて－

佐藤 幹

福山平成大学経営学部経営学科

要旨：2016 年中に広島市役所の幹部職員等に対して，インタビュー調査を実施した。その分析の結果，現在休止している行政評価については，インタビュィーのほぼ全員について再開する必要があるという発言がなかったことから，再開には及ばないこと，また，総合計画の実効性がある旨の発言についても，ほぼ全員からなかったことに加え，今後新たに策定するには莫大なコストと時間がかかることから，総合計画の策定は取り止めるべきではないかということを示唆するものである。

キーワード：地方自治体，予算，総合計画，行政評価，管理会計，
マネジメント・コントロール

はじめに

日本の地方自治体（以下，自治体）の運営の改善，管理会計論の観点から言い換えれば自治体のマネジメント・コントロールの改善のための定性的調査を開始してから，5 年が経過しようとしている。最初に半構造化インタビューを行ったのは，2012 年 5 月で広島市役所の幹部職員等を対象としたものであった。その後も他の市役所及び区役所の幹部職員等を対象としたインタビュー調査を断続的に実施してきた。

本稿では, 2016 年 3 月～11 月の間に行った広島市役所の幹部職員等へのインタビュー結果をもとに広島市役所のマネジメントの現状を分析し，市民のためにいかなる改善をなすべきかを示唆しようとするものである。

第 1 節　自治体におけるマネジメント・コントロール

佐藤（2013），佐藤（2014）及び佐藤（2015）において詳細に述べられているが，本節ではマネジメント・コントロール（Management Control以下，MCと略す）並びに筆者が自治体における代表的なMC手法またはシステム（仕組み）であると考える，予算，総合計画及び行政評価の意義について簡単に述べることとしたい。

第1項　マネジメント・コントロール

筆者は地方行政実務に携わるかたわら，大学院で研究をはじめ，それ以降10年間一貫して自治体運営の改善のために，MC上の問題点を追求し，改善策を探索してきた。

本稿では MC を広義に解しマネジメントと同義ととらえているが，狭義の MC とは，管理会計論における業績管理の仕組みであり，組織活動の一定単位を対象にそのパフォーマンス（Performance）を測定・評価するために，目標数値を設定した計画等を策定し，活動の結果を実績値として把握し，目標数値と実績値の差異を分析して，次期以降の活動のパフォーマンスを向上させる仕組みのことをいう。

第2項　予算，総合計画及び行政評価

本項では簡単に各用語の説明を行う。

１　予算

自治体の予算とは地方自治法の規定により，一定期間を区切って自治体の役所が作成する財政計画案，すなわち歳入及び歳出と債務負担行為の見積もり並びに執行準則を規定したものを法規に準じた形式で，議会の承認を受けたものである。したがって，自治体の予算とは資金のインプット管理のためのものであり，特定の支出を行うことの権威付けと支出限度額を示すものである。

予算は目的別分類と性質別分類の両方で住民に公表されている。目的別分類では，例えば，民生費，衛生費，労働費，農林水産業費，商工費，土木費，消防費，教育費等に区分される。

性質別分類では大きく消費的経費，投資的経費及びその他に区分され，さらに消費的経費は人件費，物件費，維持補修費，扶助費，補助費等に細分され，投資的経費は普通建設事業費と災害普及費に細分されている。

２　総合計画

旧地方自治法第 2 条第 4 項に規定されていた「その地域における総合的かつ計画的な行政の運営を図るための基本構想」に基づいて策定される基本計画と実施計画のことをいう。

すなわち基本構想，基本計画，実施計画の 3 層構造からなる自治体の管轄する「地域」の計画のことをいう。基本構想はいわばスローガンで，基本計画は概ね 10 年の計画であり，実施計画は基本計画の下部計画で概ね 4 年の計画としている自治体が多い。なお，地方自治法の改正に伴い 2011 年にこの規定は削除され，現在は各自治体で制定された条例等を根拠に同様の計画が策定されている。

なお，営利企業で総合計画という場合は，その「組織」の 3 年程度のスパンの中長期経営計画等のことをいい，自治体の総合計画と比較すると，営利企業のものでは財務情報への依存度が高い。

３　行政評価

本稿でいう行政評価とは 1996 年に三重県ではじめられた事務事業評価を先駆けとするもので，法令によらず自治体が自発的に行う，課レベルの組織単位で実施する事務事業を主な対象としたパフォーマンスの評価のことをいう。

第２節　インタビュー調査

2012 年に中規模な政令指定都市の 1 つである広島市役所[1]を対象に半構造化インタビューを実施したこともあり, 加えて, マネジメントの状況の変化を把握したいとの考えから, 2016 年においても広島市役所の幹部職員等を対象としたインタビューを行った。

インタビューは 2016 年 3 月から 11 月の間合計 9 名に対して実施した。1 回あたりのインタビュー時間は約 60 分で，2 回実施させていただいた方がいるため，延べ 10 回のインタビューを行った。なお，電話インタビューのみの方が 1 名いる。

第1項　インタビュー項目

主なインタビュー項目は以下のとおりである。

①行政評価を休止して変わったこと

②行政評価を実施して得られたもの

③行政運営の要となる MC は何か

④総合計画の実効性

⑤財政部局主導の問題点

⑥人事評価の仕組みの変化

2012 年のインタビューでの主眼の 1 つは，行政評価であった。その時点では市長が交代した直後で，行政評価を休止[2]したばかりであったため，広島市役所にとって行政評価とは何であったのかを考える時間的ゆとりを，ほとんどの方は有していなかったようであったことから，①と②の項目は改めて，現在の落ち着いた状態で振り返ってもらうための質問項目である。

③については 2012 年にも質問項目の柱の 1 つとしていたものである。

また，2014 年からは，すでに総合計画という中長期計画があるにもかかわらず，地方創生の名のもとに地方創生戦略計画を策定していることから，総合計画との重複感があるのではないかと考え, ④の質問項目を選んだ。なお, 地方創生戦略計画については後述する。

---

[1] 筆者は広島市役所に 33 年間勤務し，その間，行政評価の推進や総合計画案の策定及び補助金などの予算の見直しに取り組んだほか, 監査事務局で市政全般の点検業務を行った経験を持つ。したがって，2016 年においても予算，総合計画及び行政評価等を担当し，部局の長に昇進している複数の方と面識があり，比較的容易にインタビューに応じてもらえるとういアドバンティジを有している。

[2] 広島市役所は全 20 政令指定都市中，唯一行政評価を休止しているという特徴を持つ自治体である。

⑤の項目を選んだのは，2012 年のインタビューの経験から，このことが行政運営上，好ましからざる影響を与えているのではないかという問題意識を持ったからである。

最後に，⑥については，筆者は行政実務家時代から行政評価を進化させる方法の 1 つとして，人事評価[3]における業績評価と行政評価を融合させたいと考えていたことから，行政評価を休止した現状において，人事部局において何らかの変化があったかどうかを確認する意図の質問項目である。

第2項　インタビュィーのプロフィールの概要

インタビュィーは，計画部局，財政部局及び人事部局の経験者から選定した。ただし，2 名について事業課[4]の課長として計画部局, 財政部局及び人事部局からの指示に従ってきた経験を訊くために，計画部局，財政部局及び人事部局の経験者ではない方を選定した。

少し詳しく述べると，プロフィールのアウトラインとしては，計画部局経験者の局長 1 名，部長級職員 2 名，課長 1 名，財政部局経験者の局長経験退職者 1 名，部長 1 名，人事部局経験者の部長 1 名及び事業課の課長 2 名である。この中には行政評価担当を経験した方や監査事務局を経験した方が各 2 名いる。

したがって，合計 9 名にインタビューを実施した。なお，1 名については全 2 回のインタビューを行ったので，延 10 名に対してインタビューを行った。1 回あたりのインタビュー時間は約 60 分であった。

インタビューにあたっては，匿名性に配慮することで本音を語っていただいたので，インタビュィーが特定できないように記述している。

第 3 項　インタビュー結果のまとめ

①~⑥のインタビュー項目ごとに関連する発言，小括及び特記事項を記述する。

① 行政評価を休止して変わったこと

関連する発言

ほとんどの方が業務の負担が軽くなったと発言した。

元々，行政評価担当[5]が本気で行政改革をやるつもりがないのは見透かされていたので，本気ではやっていなかった。したがって，基本的には変化はないとの発言があった。

小括

全般的に休止して良かったという思いが伝わってきた。

これには，「評価疲れ」といわれていた，やらされ感から解放されたことを歓迎する気持ちが込められていた。

---

[3] 2011 年度以前から，広島市の人事評価は能力評価と業績評価から成り立っていた。
[4] 福祉，教育，環境及び産業などの分野で具体的な事業を推進する担当課のことである。
[5] 2011 年度に休止する以前，広島市では行政評価は計画部局が担当していた。

特記事項

行政評価を休止していなかったら，ひょっとしたら行政の継続性が保たれていたかも知れないという忸怩たる思いがあるという発言があった。これは，市長が交代したことにより，複数の仕組み等が休止・廃止されたことで，特定分野[6]の施策[7]が停滞してしまったからとのことであった。

② 行政評価を実施して得られたもの

関連する発言

インタビュィーのほとんど全てが行政評価を実施したことが，職員の意識改革につながったと思うと発言した。

小括

行政評価を実施して，アウトカム[8]志向で仕事を行っていこうという意識は，現在でも生きているようであった。

特記事項

行政評価は休止しているが，行政評価の遺産または名残として，毎年秋に「世界に誇れる『まち』の実現に向けた取組」という冊子を策定することで，施策の見直しを行っていることは有意義であるという発言があった。しかし，この冊子の意義の解釈には異論を唱える発言もあった。

この異論に関連することは第 3 節で詳細に述べる。

③ 行政運営の要となる MC は何か

関連する発言

大半の方が部門別等の中長期計画であると発言した。その中には，自分の部局の仕事は地方創生戦略計画と「世界に誇れる『まち』の実現に向けた取組」をまとめるときに行う施策の見直しにより，回っているという発言をした計画部局経験者もいた。

しかし，一方で 2 名の方が予算により広島市の行政運営が回っているという趣旨の発言をした。

小括

計画部局経験者は，財政部局は，財政難を根拠に限られた財源での予算編成を強調し，その視点により施策の取捨選択を行うため，良い施策を生み出すという発想に欠けると考えられているようであった。

---

[6] 本稿第 1 節第 1 項「1 予算」の目的別分類を参照。

[7] 行政の活動は，政策，施策，事務事業の 3 つに分けて説明され，政策とは大局的な見地から自治体が目指すべき方向や目的を示すもの，施策とは政策目的を達成するための方策，事務事業とは施策目的を達成するための具体的な手段のことであるとされ，漠然としたものである。そこで本稿では，政策，施策，事務事業の総称として施策という用語をもちいている。

[8] 目的にそくして，期待されたパフォーマンスが実際にどのくらい上がっているかを，評価しようとする概念である。詳しくは佐藤（2013）を参照されたい。

一方で，財政部局経験者は，広島市の行政運営は予算によって回っているという自負を抱いているようであった。

この他，予算により広島市の行政運営が回っているという趣旨の発言をした方の 1 名は，市政全体としては予算が MC の要であるが，現在の自分の部局では部下職員の目は予算にも計画にも向いていないとも言った。例えば，税金を徴収する収納担当の課等では，歳入予算を達成するために，営利企業のように徴収率を数値目標に掲げ，1 カ月単位で業務の進捗状況をチェックできるが，窓口業務が主体の課では，市民サービス向上のために如何にマンパワーを引き出していくかに着眼して業務を実施しており，狭義の MC という発想はないのではないかとの趣旨の発言もあった。

特記事項

計画部局主導で市政全般の運営が回っているのならば，財政課は不要である，財源の裏付けのない中長期計画が主導で，それに基づき予算を編成するのは，単なる予算の調整に過ぎず，財政課でなく予算調整課とか計数整理課と呼ぶべき課のする仕事であるとの発言もあり，財政部局が圧倒的な力を持っていることを示すものであると考えられた。

また，業務の多様性があり，狭義の MC の活用が困難な職場もあることが分かった。

④　総合計画の実効性

関連する発言

2 名の事業課の課長は総合計画を主眼において仕事はしていないと述べた。また，先に述べたように，自分の部局の仕事は地方創生戦略計画[9] と「世界に誇れる『まち』の実現に向けた取組」をまとめるときに行う施策の見直しにより回っており，総合計画を意識することはないとの発言もあった。

さらに，過去実際に総合計画案の策定に携わった方は，各部局から上がってきた作文を束ねてきれいに整理して，大分な冊子を作ってみても，果たしてそれが何になるのであろうかと思い虚しかったと述懐した。

その他のほとんどの方も，自分が所属する部局の仕事を回すうえで，総合計画を参照することはほとんどないと発言した。

小括

総合計画は機能していない。

[9] 安倍総理大臣が掲げたビジョンを具体化したもので，人口急減及び超高齢化という直面する大きな課題に対し政府一体となって取り組み，各地域がそれぞれの特徴を活かした自律的で持続的な社会を創生するために作成するように定められた計画である。地方創生戦略計画とは通称名で，正式には，「まち・ひと・しごと創生総合戦略」といい，ビジョンを踏まえ，2015 年度を初年度とする 5 か年の政策目標や施策の基本的方向を掲げ，KPI（重要業績評価指標）を設定して，各自治体がとりまとめた計画のことである。

特記事項

上で述べた，地方創生戦略計画と「世界に誇れる『まち』の実現に向けた取組」により，仕事が回っていると発言した方は，今後予定されている第 6 次総合計画は策定すべきでないとも発言した。その大きな理由は，総合計画の策定時には各部局から精鋭が選ばれ，計画部局に結集されるため，事業課で必要とされる職員が引き抜かれることとなり，事業課の仕事がスムースにできなくなるからだと言う。この方以外にも同様の発言があった。

⑤　財政部局主導の問題点

関連する発言

「目の利いた」施策が打てないという問題があるとの発言があった。また，財政部局出身者が出世して，花形部局の長になっているとの趣旨の発言があった。ビジョンのない市長の場合は計画部局が望ましい施策や中長期計画を提示しないと財政部局による財政難ありきの行政運営となるとの趣旨の発言があった。

小括

他都市の状況をみると，現在でもビジョンを持つことなく当選した市長のもとでは，財政部局の強い力で，行政運営が推し進められ，先に述べたように杓子定規な取捨選択により選定された施策の展開となってしまう弊害があるようである。

特記事項

かなり前のことであるが，明確なビジョンを持っていない市長の時代には，計画部局の職員が重用された時代もあったと言う方がいたことから，ビジョンなき自治体では，財政難の名のもと必要とされるべき施策が行われないという問題点があることが浮かび上がった。

⑥　人事評価の仕組みの変化

関連する発言

能力評価では本人評価がなくなったので，以前のように，まず本人評価が行われ，その後本人評価を修正する形ではなくなったので，客観性が増したという発言あった。

平成 26 年の地方公務員法の改正を受け，職員の業績の良し悪しを給与に反映する必要が生じたので，昨年度から課長級以上の管理職については，従来は年に 1 回であった業績評価を半年に 1 回に改正するために試行中であるとのことであった。

また，人事部局と計画部局との連携はほとんど取れていないとの趣旨の発言もあった。一方，人事部局と財政部局の予算編成最終盤における調整は，現在も当然行われていることが確認できた。

加えて，前市長時代に策定された人材育成基本方針は，改訂されていない旨の発言があった。

小括

人事部局と財政部局はある程度連携しているが，人事部局と計画部局との連携はほとんど見られない。

特記事項

業績評価の頻度は 1 年に 1 回から，1 年に 2 回へと改善されつつあるものの，かつての行政評価，現時点での用語で言い換えれば，中長期計画の達成度やその中の施策等の達成度を指標化し，人事評価上の業績評価に連携させようというような発想は現在もない。

第３節　分析結果のまとめ

地方創生戦略計画は自治体行政の全分野をカバーするものではなく，「まち・ひと・しごと」に関連する分野の施策にKPI[10]を設定して達成を図ろうとする 5 年の長期計画である。

したがって，総合計画とかなりの部分で重複が生じるという問題はあるが，総合計画の柱である基本計画は 10 年間というかなり長いスパンの計画であるため，融通がききにくいことを考えると，地方創生戦略計画の方が若干実効性があると考えられる。

産業振興担当の部局では，地方創生戦略計画でほとんどの仕事の方向性や数値目標が設定されているため，MC の要として活用されている。

しかし，その他の部局では，国の役所からの通知等に基づき作成する部門計画[11]を策定する必要もあり，また，逆に地方創生戦略計画に含まれない施策もあるため，中長期計画をMCにしようとすると各計画中の施策に重複や時には欠落が生じ混乱してしまう。

したがって，中長期計画を MC の要にするためには，どれか 1 つの計画に絞るべきである。

なお，先に述べたように，「世界に誇れる『まち』の実現に向けた取組」という冊子の位置づけについては，総合計画中の基本計画の下部計画とした 3 年程度の実行計画であり，毎年見直しを行いローリングさせているものであるとの異論もあった。このように中長期計画の解釈に混乱が見られるのは，市長が交代した時などに，行政の継続性が途切れてしまうことによるもので，MC としての中長期計画策定上の課題でもある。

一方，予算を要のMCとする場合は，すでに述べたように多くの弊害がある。それに加え，市長，特に新しく市長になった方が驚くことは，市長の裁量で編成できる予算額は，1 割にも満たない[12]（自治体によっては数パーセント）金額であり，新たなビジョンに基づ

---

[10] KPI の設定は好ましいことであるが，詳細に検討するとほとんどが 1 年サイクルでチェックするような指標を設定しており，早めの改善が困難であることが分かった。このようなことから，改めて，自治体のマネジメントサイクルをみると，1 カ月単位や 4 半期単位でも改善を行うことが前提とされている PDCA サイクルではなく，1 年単位の見直しを基本とする PDS サイクルであることが再認識できた。

[11] 広島市役所の場合，各部局単位で策定する計画としては，男女共同参画基本計画，子ども施策総合計画，地域福祉計画，高齢者施策推進プラン，環境基本計画，地球温暖化対策地域推進計画，教育振興基本計画等が挙げられる。これらの計画は必ずしも総合計画の分野別計画とはなっていない。

[12] 総務省が発表している，各自治体の経常収支比率をみれば分かるデータである。

く施策を展開することには大きな困難が伴う。

また，窓口業務などの市民サービスを行う課では，予算に縛られることは少なく，人員の確保・拡充や職員の個人的能力をアップするための研修に関心が高いようであり，MC を活用するには別途さらに工夫が必要である。

中長期計画であれ，予算であれ，今後，何を主要な MC として用いて，行政運営を行っていくにしても，行政の継続性の確保を，常に念頭に置いておく必要があることが，今回のインタビューで再確認できたことは有意義であった。

しかし，今回のインタビュー調査の最も大きな成果として示唆したいことは，できることなら第 6 次総合計画の策定を取り止め, 地方創生戦略計画の枠組みに習い KPI を重視し，できるだけ短い年数の計画としつつ，既存の部門別計画を重複なく，上手に活かしていくべきということである。

MC の根本はビジョンの実現に向けて，数値目標を掲げ，できれば 4 半期以内の短い見直し期間の中で，それを達成し続けられるように，改善のためのマネジメントサイクルを回すことこそが重要であることを，肝に銘じておかなければならない。

おわりに

2012 年 5 月からインタビュー調査を開始し，2013 年には定量調査として，人口の多い上位 300 の自治体の市役所及び区役所に対して質問票調査を実施し，佐藤（2016）でその結果をまとめた。また，2016 年中には広島市役所以外の役所の幹部職員等へのインタビュー調査を行ってきた。

今後は，さらにインタビュー調査を積み重ねつつ，これら調査の分析結果を統合して，自治体のマネジメントのあるべき姿を探索すべく，研究をとりまとめていく所存である。

引用及び参考文献

［１］　佐藤 幹（2013）『自治体・非営利組織のマネジメント・コントロール―バランスト・スコアカードの効用と限界―』創成社。

［２］　佐藤 幹（2014）「自治体予算のマネジメント・コントロール・システム化をめぐる問題」『会計プログレス』第 15 号（pp.38-48）

［３］　佐藤 幹（2015）「日本の地方自治体のマネジメントの状況―半構造化インタビューの結果から―」『保健医療経営大学紀要』第 7 号（pp. 21～28）

［４］　佐藤 幹（2016）「地方自治体のマネジメント・コントロール技法等の有効性に関する実証分析のための予備的考察」『保健医療経営大学紀要』第 8 号（pp. 57～67）

An Empirical Study of the Adoption of “Management Control Systems” in Hiroshima City Office

Tsuyoshi SATO

Department of Business Administration, Faculty of Business Administration, Fukuyama Heisei University

Abstract: In this paper, the results of a survey based on interviews from executive members of Hiroshima City Office during the year of 2016 are discussed. They indicate that the Administrative Performance Measurement, which has been put on hold, is not to be restarted since almost all the interviewee did not support it. As for the Comprehensive Municipal Plan, because not only there were hardly any supporting opinions heard from the interviewee, but also an over-proportionately enormous amount of budget and effort are required for formulating a new plan, the results incline toward calling off of the entire plan.

Key Words: Local Government, Budget, Municipal Plan, Performance Measurement, Management Accounting, Management Control

福山平成大学経営学部紀要
第 13 号（2017），69－98 頁

社会システム分析のための統合化プログラム２８
－メタ分析・ロジスティック回帰分析－

福井 正康[*1]、小玉 一樹[*1]、尾崎 誠[*1]
[*1] 福山平成大学経営学部経営学科

要旨：我々は教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた。今回は複数の研究資料をまとめてより強固な結論に導くための手法であるメタ分析や、2 項分布や多項分布の確率を説明変数の多項式の関数として求めるロジスティック回帰分析について、プログラムを作成した。この論文ではこれらの分析の理論を説明し、具体的なプログラムの利用法を紹介する。

キーワード：College Analysis，多変量解析，一般化線形モデル，メタ分析，ロジスティック回帰分析
URL：http://www.heisei-u.ac.jp/ba/fukui/

1. はじめに

我々はこれまで社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた。今回は、有意差検定を行った複数の研究資料を元に、主旨の同じ検定を集めて検定結果をより強固なものにしようと試みるメタ分析及び、一般化線形モデルに基づくロジスティック回帰分析についてプログラムを作成した。

メタ分析で問題となるのは、主旨の同じ検定でも論文によっては、検定方法が異なる場合があることである。例えば、2 つの量の関係を見る検定では、2 つの量からそのまま相関係数を求めて検定することもできるし、一方を 2 つの群に分けて 2 群の差の検定として見ることもできる。また、効率は悪いが両方を 2 つの群に分けて χ^2 検定等を行うこともできる。メタ分析では、これらの異なる検定から効果量を取り出し、それらの効果量を変換して一種類の効果量にする。その後、それらをまとめて、複数の検定結果を統合した検定結果を導出する。

このプログラムは参考文献[1] を元に作成したが、効果量として標準化平均値差、バイアス修正標準化平均値差、対数オッズ比、相関係数が取り上げられている。これらの他に、論文で扱われる指標には t 統計量や χ^2 統計量などがあるが、t 統計量は標準化平均値差に簡単に変換できる。χ^2 統計量は少なくとも分割表の非対角要素の値が分からなければ変換が困難である。

一般化線形モデルの 2 項分布モデルは、2 項分布の確率パラメータを説明変数の線形結合の関数で予測するモデルで、連結関数と呼ばれる予測の関数の形によっていろいろなモデルがあるが、ここでは、ロジスティックモデル、プロビットモデル、極値モデルを取り上げている。

分析名としては代表的なロジスティック回帰分析の名前を用いている。同様に一般化線形モデルの多項分布モデルは、多項分布の確率パラメータを説明変数の線形結合の関数で予測するモデルで、プログラムには名義尺度に対する名義ロジスティックモデルと順序尺度に対する累積ロジットモデルを取り上げている。ここでは今後のプログラム作成者の利用を考えて、一般化線形モデルの基礎から話を始めて、具体的なモデルを解説している。

この報告では、各章が独立しているため、式番号、図表番号は章ごとに付けるものとする。章を越えて参照する場合は、その旨を示す。

2. メタ分析

メタ分析は、多くの研究資料から同一の調査内容を選び出し、それらを統合して結果をより強固なものにしようとする分析手法である。１つの研究資料からは、効果量と呼ばれる統計量と、その分散またはデータ数を取り出す。代表的な効果量には標準化された平均値差、オッズ比、相関係数などがある。しかし、研究資料ごとにこれらが同じである保証はないので、必要があれば、これらを統一的な効果量に変換する。その後、各研究資料にウェイトをかけて、研究で与えられた結果が保証されるかどうか検討する。この一連の手法をメタ分析という[1]。

我々はこの過程を計算するプログラムの開発を考えた。ここでは、参考文献 [1] に従い、効果量の入力、効果量の変換、統計的分析に分けて、理論的にどのような式が使われているのかをまとめて紹介する。

2.1 効果量とその入力

我々がプログラムの中で扱う効果量は以下にまとめて示す。種々の研究資料には効果量（または検定値）とデータ数は記載されているが、効果量の分散が記載されていないことが多い。また、参考文献[1]では、後の統計的分析のために分散は記載されているが、データ数が記載されていない。これらの状況に対処するために、我々は結果表示に必要なデータは何か、またそれを得るためにはどのようなデータが必要かを検討した。結論は、比較的良い近似として、結果表示に必要なデータは、効果量と全データ数または、効果量と分散であった。ここでは、効果量と、全データ数または分散のどちらかが分かっているものとして、他方を求める近似式を与えておく。但しこの結果には関数 $y = 1/x(1-x)$ の性質を利用している。

１）標準化平均値差 d （ヘッジスの g とも呼ばれる）

効果量：$d = \dfrac{\bar{x}_1 - \bar{x}_2}{u_{pooled}}$,　$u_{pooled} = \sqrt{\dfrac{(n_1-1)u_1^2 + (n_2-1)u_1^2}{n_1+n_2-2}}$　（pooled 標準偏差）

分散：$V_d = \dfrac{n_1+n_2}{n_1 n_2} + \dfrac{d^2}{2(n_1+n_2)}$

全データ数：$N = n_1 + n_2$

変換近似式

$d, N \to V_d$ のとき、$V_d \simeq \dfrac{4\alpha + d^2/2}{N}$

$d, V_d \to N$ のとき、$N \simeq \dfrac{4\alpha + d^2/2}{V_d}$

ここで、分散の$(n_1 + n_2)/n_1 n_2 = N / n_1(N - n_1)$ の項については、例えば、$N = 100$ とすると、n_1 の関数として図１のようなグラフとなる。

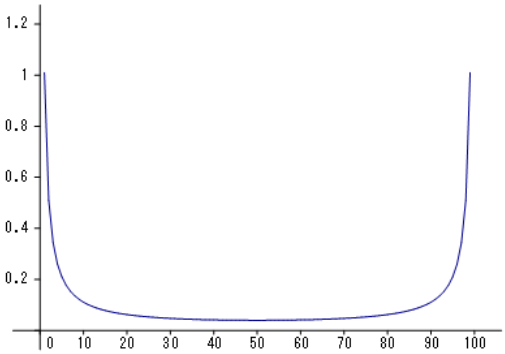


図１　$y = 100/x(100 - x)$

このグラフは、中央部で$4/N$ に近いほぼ安定な値を取っており、変動は少ないと考えられる。そこで我々は、この関数の$x = N/2$ の値を中心とした正規分布による加重平均を考え、その結果を$(n_1 + n_2)/n_1 n_2 \simeq 4\alpha/N$ とした。

α の値については、以下のように計算した。

$$\alpha = \frac{1}{4A}\int_{0.1}^{0.9} \frac{1}{x(1-x)} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-0.5)^2}{2\sigma^2}\right] dx$$

$$A = \frac{1}{\sqrt{2\pi}\sigma}\int_{0.1}^{0.9} \exp\left[-\frac{(x-0.5)^2}{2\sigma^2}\right] dx$$

この場合、例えば、$\sigma = 0.2$ （約 95%の研究で$0.1N < n_1 < 0.9N$ ）とすると、$\alpha = 1.187$ となる。我々はこの値を利用する。実際の調査では、2 群間の差の検定におけるデータ数について、群間にあまり極端な差はないものと考えられる。

標準化平均値差の代わりに、研究資料で t 統計量とデータ数が使われている場合は、以下のように簡単に標準化平均値差に変換することができる。

$$t=\sqrt{\frac{n_1n_2}{n_1+n_2}}\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{(n_1-1)u_1^2+(n_2-1)u_2^2}{n_1+n_2-2}}}=\sqrt{\frac{n_1n_2}{n_1+n_2}}d\simeq\sqrt{\frac{N}{4\alpha}}d$$

２）バイアス修正標準化平均値差 g

効果量：$g=J\times d$，　$J=1-\dfrac{3}{4(n_1+n_2-2)-1}$，　分散：$V_g=J^2\times V_d$

全データ数：$N=n_1+n_2$

変換近似式

$N\to V_g$ のとき、$J=1-\dfrac{3}{4(N-2)-1}$ として、$V_g=J^2V_d\simeq\dfrac{4\alpha J^2+g^2/2}{N}$

$V_g\to N$ のとき、$J\simeq 1$ であると考え、$N\simeq\dfrac{4\alpha+g^2/2}{V_g}$

３）対数オッズ比 LOR

以下の 2 次元分割表を考える。

	効果あり	効果なし	合計
介入群	a	b	$a+b=n_1$
統制群	c	d	$c+d=n_2$

効果量：$LOR=\ln\left(\dfrac{a/b}{c/d}\right)=\ln\left(\dfrac{ad}{bc}\right)$，　分散：$V_{LOR}=\dfrac{1}{a}+\dfrac{1}{b}+\dfrac{1}{c}+\dfrac{1}{d}$

全データ数：$N=a+b+c+d=n_1+n_2$

変換近似式

$N\to V_{LOR}$ のとき、$V_{LOR}\simeq\dfrac{16\alpha^2}{N}$　$V_{LOR}=\dfrac{1}{a}+\dfrac{1}{b}+\dfrac{1}{c}+\dfrac{1}{d}\simeq\dfrac{4\alpha}{n_1}+\dfrac{4\alpha}{n_2}\simeq\dfrac{16\alpha^2}{N}$

$V_{LOR}\to N$ のとき、$N\simeq\dfrac{16\alpha^2}{V_{LOR}}$

ここで、$1/a+1/b\simeq 4\alpha/n_1$，$1/c+1/d\simeq 4\alpha/n_2$ などとした部分については a,b や c,d の値が大きく異なっている場合は誤差が大きい。

対数オッズ比の代わりに、資料で以下の χ^2 統計量が使われていた場合は、簡単に対数オッズ比に変換することができない。分割表の度数から効果量を計算する必要がある。

$$\chi^2=\frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

４）相関係数r

効果量：$r=\dfrac{s_{xy}}{s_x s_y}$，　分散：$V_r=\dfrac{(1-r^2)^2}{n-1}$，　全データ数：$N=n$

変換近似式

$N \to V_r$ のとき、$V_r=\dfrac{(1-r^2)^2}{N-1}$

$V_g \to N$ のとき、$N=\dfrac{(1-r^2)^2}{V_r}+1$

2.2　効果量の変換

効果量は相互に変換可能である。ここではプログラムで用いられる変換について式を与える。

変換$d \leftrightarrow g$

効果量：$g=J\times d$，　分散：$V_g=J^2\times V_d$

ここに、$J=1-\dfrac{3}{4(N-2)-1}$　（入力の際にNは設定済みとする）

変換$LOR \leftrightarrow d$

効果量：$d=LOR\times\dfrac{\sqrt{3}}{\pi}$，　分散：$V_d=V_{LOR}\times\dfrac{3}{\pi^2}$

変換$r \to d$

効果量：$d=\dfrac{2r}{\sqrt{1-r^2}}$，　分散：$V_d=\dfrac{4V_r}{(1-r^2)^3}$

変換$d \to r$

効果量：$r=\dfrac{d}{\sqrt{d^2+a}}$，　分散：$V_r=\dfrac{a^2V_d}{(d^2+a)^3}$

ここに、$a=\dfrac{(n_1+n_2)^2}{n_1n_2}\simeq 4\alpha$

2.3　統計的分析

研究を統合する場合は、以下の２つの場合に分けて行われる。１つは研究間に差がなく、研究iの効果量d_iが独立に$d_i \sim N(0,V_i)$に従うと考えられる場合である。もう１つは、研究iの効果量d_iが広く拡がり、$d_i \sim N(0,V_i+\sigma^2), \sigma^2>0$に従うと考えられる場合である。

前者を固定効果モデル、後者を変量効果モデルと呼ぶ。どちらかを調べるには、帰無仮説を $\sigma^2 = 0$ として、以下の関係を利用する。

$$Q = \sum_{i=1}^{n} w_i (d_i - d)^2 = \sum_{i=1}^{n} \frac{(d_i - d)^2}{V_i} \sim \chi^2_{n-1}$$

ここで、w_i や d については、後の固定効果モデルのところで定義を示す。

固定効果モデル

固定効果モデルでは、研究間の差はなく、研究 i の効果量 d_i は独立に $d_i \sim N(0, V_i)$ に従うと仮定し、以下の集計を考える。

$$d = \sum_{i=1}^{n} d_i / V_i \Big/ \sum_{i=1}^{n} 1/V_i \sim N\left(0, 1 \Big/ \sum_{i=1}^{n} 1/V_i\right)$$

ここで、$w_i = 1/V_i$ として、これをウェイトと考え、$w = \sum_{i=1}^{n} w_i$ とすると、以下となる。

$$d_i \sim N(0, 1/w_i), \quad d = \sum_{i=1}^{n} w_i d_i \Big/ w \sim N\left(0, 1/w\right)$$

この性質より、研究を結合した検定は、検定統計量 $z = d \Big/ \sqrt{1/w} \sim N(0,1)$ を使って行う。

変量効果モデル

変量効果モデルでは、研究間に差があり、研究 i の効果量 d_i は広く拡がり、$d_i \sim N(0, V_i + \sigma^2)$ に従うと考える。$w'_i = 1/(V_i + \sigma^2)$ とおくと、$d_i \sim N(0, 1/w'_i)$ より、$\sqrt{w'_i} d_i \sim N(0,1)$ となり、以下を得る。

$$Q' = \sum_{i=1}^{n} w'_i (d_i - d)^2 = \sum_{i=1}^{n} \frac{(d_i - d)^2}{V_i + \sigma^2} \sim \chi^2_{n-1}$$

一方、

$$Q = \sum_{i=1}^{n} w_i (d_i - d)^2 = \sum_{i=1}^{n} \frac{(d_i - d)^2}{V_i}$$

は元の分散で測った量である。その差は、以下で与えられる。

$$Q - Q' = \sum_{i=1}^{n} \frac{(d_i - d)^2 \sigma^2}{V_i (V_i + \sigma^2)} = \sum_{i=1}^{n} (d_i - d)^2 w'_i w_i \sigma^2 = C\sigma^2$$

ここで、Q' と C は期待値を使って求める。即ち、

$$E(Q') = n - 1$$

$$C = E[\sum_{i=1}^{n} (d_i - d)^2 w'_i w_i] = \sum_{i=1}^{n} E[(d_i - d)^2] w'_i w_i$$

C については、

$$E[d_i^2] = V_i' = 1/w_i'$$

$$E[d_i d] = E[d_i \sum_{j=1}^{n} d_j w_j' / w'] = V_i' w_i' / w' = 1/w'$$

$$E[d^2] = E[\sum_{i=1}^{n} d_i w_i' / w' \sum_{j=1}^{n} d_j w_j' / w'] = \sum_{i=1}^{n} V_i w_i'^2 / w'^2 = \sum_{i=1}^{n} w_i' / w'^2 = 1/w'$$

より、以下となる。。

$$C = \sum_{i=1}^{n} (1/w_i' - 1/w') w_i' w_i = \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i' w_i / w' \simeq \sum_{i=1}^{n} w_i - \sum_{i=1}^{n} w_i^2 / w$$

これらより、σ^2 が以下のように求められる。

$$\sigma^2 = \frac{Q-(n-1)}{C}$$

以後、ウェイトは $w_i' = 1/(V_i + \sigma^2)$， $w' = \sum_{i=1}^{n} w_i'$ を用いて、計算を行えばよい。即ち、研究を結合した検定は、検定統計量 $z' = d/\sqrt{1/w'} \sim N(0,1)$ を使って行う。

2.4　研究群間の比較

何らかの指標の違いにより、研究が k 個のグループに分けられるとする。各グループの研究の数を n_i， 全体の研究の数を n とするとき、そのグループ間の効果量の差を検定するには、以下の性質を用いる。

$$Q_{Total} \sim \chi^2_{n-1}, \quad Q_i \sim \chi^2_{n_i-1}, \quad n = \sum_{i=1}^{k} n_i \quad \text{より、}$$

$$Q_{Total} - \sum_{i=1}^{k} Q_i \sim \chi^2_{df}, \quad df = (n-1) - \sum_{i=1}^{k} (n_i - 1) = k-1$$

この計算には、固定効果モデルではウェイト $w_i = 1/V_i$ を用い、変量効果モデルではウェイト $w_i' = 1/V_i'$ を用いる。

2.5　プログラムの利用法

メニュー［分析－多変量解析等－メタ分析］を選択すると、メタ分析の分析実行メニューが図１のように表示される。

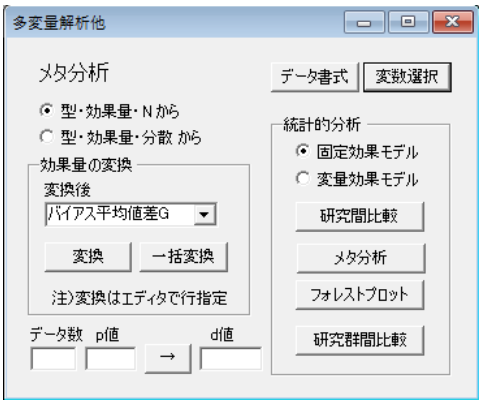
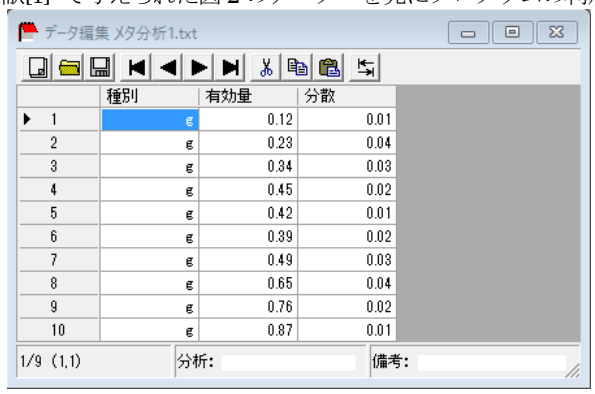


図１　分析実行メニュー

ここでは、参考文献[1] で与えられた図２のデータ[3] を元にプログラムの利用法を説明する。



	種別	有効量	分散
1	g	0.12	0.01
2	g	0.23	0.04
3	g	0.34	0.03
4	g	0.45	0.02
5	g	0.42	0.01
6	g	0.39	0.02
7	g	0.49	0.03
8	g	0.65	0.04
9	g	0.76	0.02
10	g	0.87	0.01

図２　メタ分析データ

このデータではデータ型、効果量（有効量）、分散が用いられているので、分析実行メニューのデータ型を、「型・効果量・分散」に変えて、分析を実行する。しかし、一般には分散が与えられる場合は少なく、むしろ、データ型、効果量、データ数が与えられることが多いと思われる。その場合には、分析実行メニューのデータ型を、「型・効果量・N」にして実行する。データ型には、G：バイアス修正標準化平均値差、D：標準化平均値差、LOR：対数オッズ比、R：相関係数が指定できる。なお、指定する文字は大文字でも小文字でも同じである。

また、2 群の差の検定などでは、検定統計量を省略し、検定確率だけを表示している場合もあるので、その際には、標準化平均値差 D の値を簡易的に計算できる機能をメニューの下に設けている。その他の対数オッズ比や相関係数では、殆どの場合、値を記述するので、ここでは標準化平均値差 D に限定している。また、ノンパラメトリック検定の確率から近似的に D を求めても、少し乱暴ではあるが、経験上特に大きな差は出ないように思う。

一般に各研究では効果量が同一とは限らない。異なる効果量の場合は、効果量の変換を行い、同じ効果量に合わせて分析する。そのためにプログラムには効果量の変換機能を付けている。変数選択で 3 つの変数を選択し、「変換後」コンボボックスで変換先の型を選び、「一括変換」

ボタンをクリックすると、図３のような結果が得られる。ここでは、相関係数として出力している。

効果量変換

	種別	効果量	分散
1	R	0.0551	0.0021
2	R	0.1056	0.0083
3	R	0.1549	0.0059
4	R	0.2029	0.0037
5	R	0.1896	0.0019
6	R	0.1767	0.0039
7	R	0.2204	0.0055
8	R	0.2875	0.0066
9	R	0.3302	0.0030
10	R	0.3713	0.0014

図３　相関係数への変換

グリッドの一部分のデータについて変換をしたい場合は、種別・効果量・分散の必要な行を連続的に選択して「変換」ボタンをクリックする。出力結果は省略する。

すべての研究結果を統合して検定を行いたい場合、研究間の効果量の値にばらつきがあるかどうか知らなければならない。それを調べる場合は、「研究間比較」ボタンをクリックする。結果を図４に示す。

研究間効果量比較

```
研究間効果量比較

研究数        10
研究間分散    0.054
Q値           35.839
自由度        9
有意水準      0.05
確率          0.0000
研究間の効果量に差があるといえる。
```

図４　研究間効果量の差の比較検定

この結果から、研究間の効果量に差が見られたので、分析には「変量効果モデル」を用いる。これは「固定効果モデル」に比べて差が検出しにくい検定である。

変量効果モデルを用いた最終的な分析結果を得るには、「変量効果モデル」ラジオボタンを選択し、「メタ分析」ボタンをクリックする。結果を図５に示す。

フォレストデータ

	種別	効果量	N	分散	標準誤差	p値	2.5%下限	2.5%上限
1	G	0.1200	476	0.0100	0.1000	0.2301	-0.0760	0.3160
2	G	0.2300	119	0.0400	0.2000	0.2501	-0.1620	0.6220
3	G	0.3400	160	0.0300	0.1732	0.0496	0.0005	0.6795
4	G	0.4500	242	0.0200	0.1414	0.0015	0.1728	0.7272
5	G	0.4200	484	0.0100	0.1000	0.0000	0.2240	0.6160
6	G	0.3900	241	0.0200	0.1414	0.0058	0.1128	0.6672
7	G	0.4900	162	0.0300	0.1732	0.0047	0.1505	0.8295
8	G	0.6500	124	0.0400	0.2000	0.0012	0.2580	1.0420
9	G	0.7600	252	0.0200	0.1414	0.0000	0.4828	1.0372
10	G	0.8700	513	0.0100	0.1000	0.0000	0.6740	1.0660
結合		0.4747	2773	0.0076	0.0870	0.0000	0.3041	0.6453

図５　変量効果モデルを用いた分析結果

各研究の結果がまとめて表示され、一番下の行に結合された結果が表示されている。

さらに、この結果を分かり易く表す図がフォレストプロットである。「フォレストプロット」ボタンをクリックすると、図６のような結果が表示される。

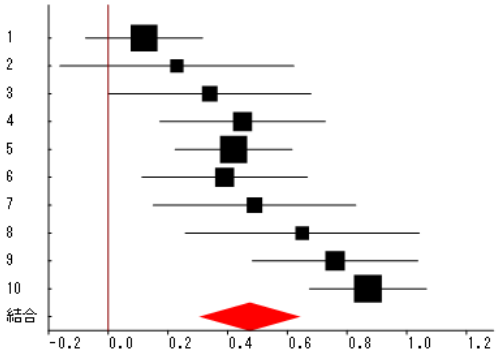
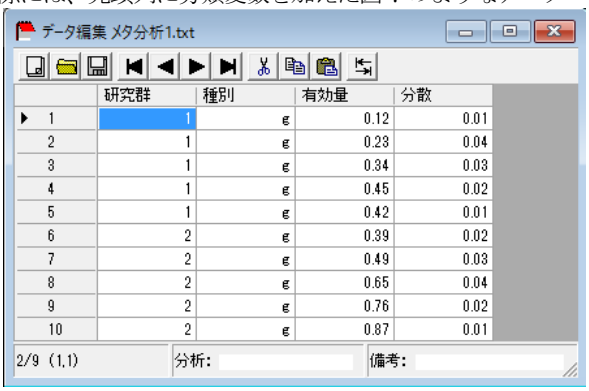


図 6　フォレストプロット

一番下のひし形が、0 をまたいでいないことから、この結果では、有意に差があるといえるということになる。

次に、研究がいくつかの特徴に分かれ、その研究群間に差があるかどうか調べてみたいと考えたとする。その際には、先頭列に分類変数を加えた図 7 のようなデータ[3] を用いる。



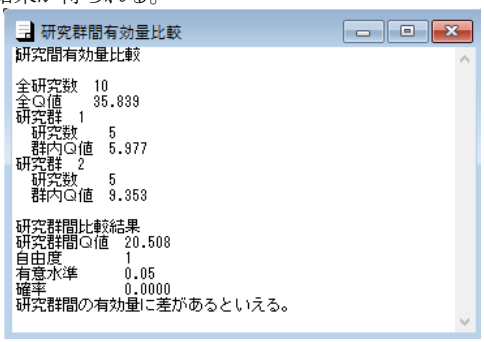
データ編集 メタ分析1.txt

	研究群	種別	有効量	分散
1	1	g	0.12	0.01
2	1	g	0.23	0.04
3	1	g	0.34	0.03
4	1	g	0.45	0.02
5	1	g	0.42	0.01
6	2	g	0.39	0.02
7	2	g	0.49	0.03
8	2	g	0.65	0.04
9	2	g	0.76	0.02
10	2	g	0.87	0.01

2/9 (1,1)　分析:　備考:

図 7　2 つの研究群による比較データ（メタ分析 1.txt）

すべてのデータを並んだ順に選択し、分析実行メニューの「研究群間比較」ボタンをクリックすると、図 8 のような結果が得られる。



研究群間有効量比較

```
研究間有効量比較

全研究数  10
全Q値     35.839
研究群  1
  研究数    5
  群内Q値  5.977
研究群  2
  研究数    5
  群内Q値  9.353

研究群間比較結果
研究群間Q値  20.508
自由度          1
有意水準       0.05
確率          0.0000
研究群間の有効量に差があるといえる。
```

図 8　研究群間の比較結果

3.　２値ロジスティック回帰

２値ロジスティック回帰分析は２項分布の確率を説明変数の１次式の関数で予測する分析手法である。この分析は、確率の大きさによって事象の出現、非出現を区別することにも使え、質的データの予測手法としても利用できる。同様の分析に判別分析があるが、これは説明変数によるマハラノビス距離を用いた判別方法で、理論的な根拠という点ではロジスティック回帰分析の方が優れている。ここでは今後のために、参考文献[2] に従って、一般化線形モデルの基礎からロジスティック回帰分析について説明しておく。

3.1　一般化線形モデル

1）　指数型分布族

ある単一のパラメータθを持つ確率変数Yが以下の密度関数に従うとき、その分布を指数型分布族という。

$$f(y;\theta)=\exp[a(y)b(\theta)+c(\theta)+d(y)]$$

指数型分布族には、ポアソン分布、正規分布、２項分布等が含まれる。特に$a(y)=y$のとき分布は正準形であると言われ、$b(\theta)$は分布の自然パラメータと呼ばれる。

確率変数$a(\mathrm{Y})$については

$$\frac{d}{d\theta}\int f(y;\theta)dy=\int[a(y)b'(\theta)+c'(\theta)]f(y;\theta)dy$$

$$=E[a(Y)]b'(\theta)+c'(\theta)=0$$

より、

$$E[a(Y)]=-c'(\theta)/b'(\theta)$$

$$\frac{d^2}{d\theta^2}\int f(y;\theta)dy$$

$$=\int[a(y)b''(\theta)+c''(\theta)]f(y;\theta)dy+\int[a(y)b'(\theta)+c'(\theta)]^2 f(y;\theta)dy$$

$$=E[a(Y)^2]b'(\theta)^2+E[a(Y)][b''(\theta)+2b'(\theta)c'(\theta)]+c''(\theta)+c'(\theta)^2$$

$$=E[a(Y)^2]b'(\theta)^2-\frac{c'(\theta)}{b'(\theta)}[b''(\theta)+2b'(\theta)c'(\theta)]+c''(\theta)+c'(\theta)^2$$

$$=E[a(Y)^2]b'(\theta)^2-E[a(Y)]^2b'(\theta)^2+\frac{1}{b'(\theta)}[-b''(\theta)c'(\theta)+c''(\theta)b'(\theta)]$$

$$=V[a(Y)]b'(\theta)^2+\frac{1}{b'(\theta)}[-b''(\theta)c'(\theta)+c''(\theta)b'(\theta)]=0$$

より、

$$V[a(Y)] = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)^3}$$

という性質がある。

対数密度関数 $l(y;\theta) = \log f(y;\theta)$ の θ に関する微分の y を確率変数とみなした

$$U(Y;\theta) = a(Y)b'(\theta) + c'(\theta)$$

は、スコア統計量とも呼ばれ、その分布の期待値と分散は上式を使うと以下となる。

$$E[U] = 0$$

$$V[U] = V[a(Y)]b'(\theta)^2 = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)}$$

さらに、

$$V[\mathrm{U}] = E[U^2] - E[U]^2 = E[U^2]$$

$$E[U'] = E[a(Y)]b''(\theta) + c''(\theta) = -\frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)} = -V[U]$$

の関係より、以下も成り立つ。

$$V[U] = E[U^2] = -E[U']$$

スコア統計量の分散 $V[U]$ は情報量とも呼ばれる。

2) 正準形の一般化線形モデル

正準形の指数型分布族の分布に従う確率変数 Y_i $(i = 1, 2, \cdots, N)$ が、パラメータ θ_i の同じ形の以下の独立な密度関数の分布に従うと考える。

$$f(y_i;\theta_i) = \exp[y_i b(\theta_i) + c(\theta_i) + d(y_i)]$$

対数密度関数 $l(y_i;\theta_i)$ は以下で与えられる。

$$l(y_i;\theta_i) = y_i b(\theta_i) + c(\theta_i) + d(y_i)$$

確率変数 Y_i の平均と分散は前節の議論より、以下のように与えられる。

$$E[Y_i] = -c'(\theta_i)/b'(\theta_i) \equiv \mu_i$$

$$V[Y_i] = \frac{b''(\theta_i)c'(\theta_i) - c''(\theta_i)b'(\theta_i)}{b'(\theta_i)^3}$$

ここで、θ_i は μ_i の関数であるとみることができる。

我々はこの μ_i に対して、ある説明変数 ${}^t\mathbf{x}_i = (1, x_{i1}, \cdots, x_{ip})$（$i = 1, \cdots, N$）とパラメータ ${}^t\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)$ を用いて以下のような仮定をする。

$$\eta_i \equiv g(\mu_i) = \sum_{j=1}^{p} \beta_j x_{ij} + \beta_0 = {}^t\boldsymbol{\beta}\mathbf{x}_i$$

この仮定により、θ_i は $\boldsymbol{\beta}$ の関数と見ることができる。またこの関係を与える関数 $\eta_i = g(\mu_i)$ を連結関数という。

確率変数 Y_i の同時密度関数（尤度関数）は以下で与えられる。

$$f(\mathbf{y};\boldsymbol{\theta}) = \exp\left[\sum_{i=1}^{N}\{y_i b(\theta_i) + c(\theta_i) + d(y_i)\}\right]$$

また対数尤度関数 $l(\mathbf{y};\boldsymbol{\theta})$ は以下のようになる。

$$l(\mathbf{y};\boldsymbol{\theta}) = \sum_{i=1}^{N} l_i = \sum_{i=1}^{N}[y_i b(\theta_i) + c(\theta_i) + d(y_i)] \tag{1}$$

この対数尤度関数の β_j による微分をスコアベクトルと呼び、U_j とすると、スコアベクトル U_j は以下のようになる。

$$U_j = \sum_{i=1}^{N}\frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{N}\frac{\partial \mu_i}{\partial \beta_j}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial l_i}{\partial \theta_i}$$

ここで、

$$\frac{\partial l_i}{\partial \theta_i} = y_i b'(\theta_i) + c'(\theta_i) = b'(\theta_i)\left[y_i + c'(\theta_i)/b'(\theta_i)\right]$$

$$= b'(\theta_i)(y_i - E[Y_i]) = b'(\theta_i)(y_i - \mu_i)$$

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{b'(\theta_i)^2}{-c''(\theta_i)b'(\theta_i) + c'(\theta_i)b''(\theta_i)} = \frac{1}{b'(\theta_i)V[Y_i]}$$

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \eta_i}{\partial \beta_j}\frac{\partial \mu_i}{\partial \eta_i} = x_{ij}\frac{\partial \mu_i}{\partial \eta_i}$$

となることから、以下の関係を得る。

$$U_j = \sum_{i=1}^{N}\frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{N}\frac{(y_i - \mu_i)x_{ij}}{V[Y_i]}\frac{\partial \mu_i}{\partial \eta_i} \tag{2}$$

また、以下も成り立つ。

$$E[U_j] = 0 \tag{3}$$

さらに U_j の β_k による微分を U_{jk} とすると、U_{jk} は以下のようになる。

$$U_{jk} \equiv \frac{\partial U_j}{\partial \beta_k} = \frac{\partial}{\partial \beta_k}\sum_{i=1}^{N}\frac{\partial l_i}{\partial \beta_j} = \sum_{i=1}^{N}\frac{\partial \mu_i}{\partial \beta_k}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial}{\partial \theta_i}\left(\frac{\partial \mu_i}{\partial \beta_j}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial l_i}{\partial \theta_i}\right)$$

$$= \sum_{i=1}^{N}\frac{\partial \mu_i}{\partial \beta_j}\frac{\partial \mu_i}{\partial \beta_k}\left(\frac{\partial \theta_i}{\partial \mu_i}\right)^2\frac{\partial}{\partial \theta_i}\left(\frac{\partial l_i}{\partial \theta_i}\right) + \sum_{i=1}^{N}\frac{\partial \mu_i}{\partial \beta_k}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial l_i}{\partial \theta_i}\frac{\partial}{\partial \theta_i}\left(\frac{\partial \mu_i}{\partial \beta_j}\frac{\partial \theta_i}{\partial \mu_i}\right)$$

$$= \sum_{i=1}^{N} \frac{\partial \mu_i}{\partial \beta_j} \frac{\partial \mu_i}{\partial \beta_k} \left(\frac{\partial \theta_i}{\partial \mu_i} \right)^2 [y_i b''(\theta_i) + c''(\theta_i)]$$

$$+ \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ik}}{V[Y_i]} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial}{\partial \theta_i} \left(\frac{\partial \mu_i}{\partial \beta_j} \frac{\partial \theta_i}{\partial \mu_i} \right)$$

また、

$$E[Y_i b''(\theta_i) + c''(\theta_i)] = E[Y_i] b''(\theta_i) + c''(\theta_i) = -c'(\theta_i) b''(\theta_i) / b'(\theta_i) + c''(\theta_i)$$

$$= \frac{-c'(\theta_i) b''(\theta_i) + b'(\theta_i) c''(\theta_i)}{b'(\theta_i)} = -b'(\theta_i)^2 V[Y_i]$$

$$E\left[\sum_{i=1}^{N} \frac{(Y_i - \mu_i) x_{ik}}{V[Y_i]} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial}{\partial \theta_i} \left(\frac{\partial \mu_i}{\partial \beta_j} \frac{\partial \theta_i}{\partial \mu_i} \right) \right] = 0$$

であることから、(2) 式を求める際の計算により、U_{jk} の変数の値を確率変数で置き換えて計算すると以下となる。

$$E[U_{jk}] = -\sum_{i=1}^{N} \frac{\partial \mu_i}{\partial \beta_j} \frac{\partial \mu_i}{\partial \beta_k} \left(\frac{\partial \theta_i}{\partial \mu_i} \right)^2 b'(\theta_i)^2 V[U_i] = -\sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{V[Y_i]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

一方、(2) の表式より、

$$E[U_j U_l] = \sum_{i=1}^{N} \sum_{k=1}^{N} \frac{E[(y_i - \mu_i)(y_k - \mu_k)] x_{ij} x_{kl}}{V[Y_i] V[Y_k]} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \mu_k}{\partial \eta_k}$$

$$= \sum_{i=1}^{N} \sum_{k=1}^{N} \frac{V[Y_i] \delta_{ik} x_{ij} x_{kl}}{V[Y_i] V[Y_k]} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \mu_k}{\partial \eta_k} = \sum_{i=1}^{N} \frac{x_{ij} x_{il}}{V[Y_i]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2$$

であることから、以下となる。

$$E[U_{jk}] = -E[U_j U_k] \tag{4}$$

ここで、$(\Im)_{jk} = -E[U_{jk}] = E[U_j U_k]$ とすると、行列 $\Im$ は情報行列と呼ばれる。

$$(\Im)_{jk} = E(U_j U_k) = \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{V[Y_i]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \tag{5}$$

今、(1) で与えられる対数尤度関数が最大となる $\boldsymbol{\beta}$ の値を求めてみよう。これには、

$$\frac{\partial l}{\partial \beta_j} = U_j = 0$$

という方程式を解くことになる。

（$f(x) = 0$ を解くには $y_m = f'(x_{m-1})(x_m - x_{m-1}) - f(x_{m-1}) = 0$ を計算することから）

解はニュートン・ラフソン法によると、

$$U_j^{(m)} = \sum_{k=1}^{p} U_{jk}^{(m-1)}(\beta_k^{(m)} - \beta_k^{(m-1)}) + U_j^{(m-1)} = 0$$

のように、$\beta_k^{(m)}$ の値を逐次求めて行くことになるが、実際の計算では$U_{jk}^{(\mathrm{m}-1)}$ の代わりに、期待値を取った情報行列$-(\Im^{(m-1)})_{jk}$ を用いる。この式を書き変えると、以下となる。

$$\boldsymbol{\beta}^{(m)} = \boldsymbol{\beta}^{(m-1)} + (\Im^{(m-1)})^{-1}\mathbf{U}^{(m-1)}$$

(3)式と(5)式を元にして、大標本においては、スコアベクトルの分布は漸近的に

$$\mathbf{U} \sim \mathbf{N}(\mathbf{0}, \Im), \quad {}^t\mathbf{U}\Im^{-1}\mathbf{U} \sim \chi^2(p)$$

であることも示される。

最尤推定量$l(\boldsymbol{\beta})$ の推定値$\mathbf{b}$ の近傍でのテイラー展開近似は以下となり、

$$l(\boldsymbol{\beta}) = l(\mathbf{b}) + {}^t(\boldsymbol{\beta}-\mathbf{b})\mathbf{U}(\mathbf{b}) - \frac{1}{2}{}^t(\boldsymbol{\beta}-\mathbf{b})\Im(\mathbf{b})(\boldsymbol{\beta}-\mathbf{b})$$

スコアベクトルの推定値$U(\mathbf{b})$ の近傍でのテイラー展開近似は以下となる。

$$\mathbf{U}(\boldsymbol{\beta}) = \mathbf{U}(\mathbf{b}) - \Im(\mathbf{b})(\boldsymbol{\beta}-\mathbf{b}) = -\Im(\mathbf{b})(\boldsymbol{\beta}-\mathbf{b})$$

ここでは$E[\partial U_j / \partial \beta_k] = -\Im_{jk}$ や$\mathbf{U}(\mathbf{b}) = \mathbf{0}$ を使っている。これらより、

$${}^t(\boldsymbol{\beta}-\mathbf{b})\Im(\mathbf{b})(\boldsymbol{\beta}-\mathbf{b}) \sim \chi^2(p)$$

も示される。また、同様にして以下も示される。

$$\mathbf{b} = \boldsymbol{\beta} + \Im^{-1}\mathbf{U} \sim N(\boldsymbol{\beta}, \Im^{-1})$$

後に説明するが、モデルの最適値からのずれを表す逸脱度D を以下のように定義する。

$$D = 2[l(\mathbf{b}_{\max};\mathbf{y}) - l(\hat{\mathbf{b}};\mathbf{y})] \sim \chi^2(N-p)$$

ここに、$l(\mathbf{b}_{\max};\mathbf{y})$ はパラメータ数N の飽和モデルでの対数尤度、$l(\hat{\mathbf{b}};\mathbf{y})$ は現在考えている、パラメータ数p のモデルでの対数尤度である。同じパラメータ数では、この値が小さい連結関数のモデルほど適合が良いと判断する。但し、分布は漸近的に成り立つものである。

モデルに意味があるかどうかの検定では、以下の尤度比χ^2統計量が使われる。

$$C = 2[l(\hat{\mathbf{b}};\mathbf{y}) - l(\mathbf{b}_{\min};\mathbf{y})] \sim \chi^2(p-1)$$

ここに$l(\mathbf{b}_{\min};\mathbf{y})$ は定数パラメータ１つの最小モデルの対数尤度、$l(\hat{\mathbf{b}};\mathbf{y})$ は現在考えている、パラメータ数p のモデルでの対数尤度である。これは、帰無仮説として最小モデルが正しい（回帰式は意味がない）とする検定である。

２項分布の場合、対数尤度関数は以下で与えられる。

$$l(\boldsymbol{\beta};\mathbf{y}) = \sum_{i=1}^{N}\left[y_i \log p_i + (n_i - y_i)\log(1-p_i) + \log\binom{n_i}{y_i} \right]$$

$l(\mathbf{b}_{\max};\mathbf{y})$ と $l(\hat{\mathbf{b}};\mathbf{y})$ は、確率 p_i の中に、それぞれ実測値を用いた $p_i = y_i / n_i$ と予測値を用いた $\hat{p}_i = \hat{y}_i / n_i$ を代入して求める。$l(\mathbf{b}_{\min};\mathbf{y})$ は確率 p_i の中に同じ以下の値を代入して求める。

$$\tilde{p} = \sum_{i=1}^{N} y_i \Big/ \sum_{i=1}^{N} n_i$$

実測値と推測値の関係を与える指標として、決定係数からの類推である以下の擬似 R^2 も利用される。

$$\tilde{R}^2 = \frac{l(\mathbf{b}_{\min};\mathbf{y}) - l(\hat{\mathbf{b}};\mathbf{y})}{l(\mathbf{b}_{\min};\mathbf{y})}$$

さらにプログラムでは、実測値と予測値の相関係数も求めている。

3) 2 項分布モデル

2 項分布のパラメータを説明変数の線形結合で推測する場合、密度関数、密度関数の対数、逸脱度、目的変数の平均と分散は以下のようになる。ここで、密度関数の対数の最後の項はパラメータに依存していないので、計算上は考えないことにする（参考文献[2] の数値に従っている）。

$$f(y_i; p_i) = {}_{n_i}C_{y_i}\, p_i^{y_i} (1-p_i)^{n_i - y_i}$$

$$\begin{aligned} l(y_i; p_i) &= \log[{}_{n_i}C_{y_i}\, p_i^{y_i} (1-p_i)^{n_i - y_i}] \\ &= y_i \log p_i + (n_i - y_i)\log(1-p_i) + \log {}_{n_i}C_{y_i} \\ &= y_i[\log p_i - \log(1-p_i)] + n_i \log(1-p_i) + \log {}_{n_i}C_{y_i} \\ &\to y_i[\log p_i - \log(1-p_i)] + n_i \log(1-p_i) \end{aligned}$$

$$D = 2\sum_{i=1}^{N}\left[y_i \log\left(\frac{y_i}{n_i p_i}\right) + (n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - n_i p_i}\right)\right] \sim \chi^2(N-p)$$

$$C = 2\sum_{i=1}^{N}\left[y_i \log\left(\frac{\hat{y}_i}{n_i \tilde{p}}\right) + (n_i - y_i)\log\left(\frac{n_i - \hat{y}_i}{n_i - n_i \tilde{p}}\right)\right] \sim \chi^2(p-1)$$

$$\tilde{p} = \sum_{i=1}^{N} y_i \Big/ \sum_{i=1}^{N} n_i$$

$$E[Y_i] = n_i p_i \equiv \mu_i$$

$$V[Y_i] = n_i p_i (1-p_i)$$

ここでは n_i 回の試行に対して、y_i 回の事象が起こったとしているが、1 回の試行で起こったか起こらないかにする場合は、$n_i = 1,\ y_i = \{0, 1\}$ とすればよい。

逸脱度と同様に最適値からのずれを表す統計量に以下のピアソン χ^2 統計量がある。

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{n_i \hat{p}_i (1-\hat{p}_i)} \sim \chi^2(N-p)$$

これは逸脱度と漸近的に同じ指標であるが、逸脱度と比べてこちらの方が分布によく適合するという意見もある。

これまでは、2 項分布に基づく一般論であったが、これ以降は、説明変数との関係を与える連結関数の部分に仮定が入る。連結関数の仮定でよく利用されるモデルが、ロジスティックモデル、プロビットモデル、極値モデル等である。以下に最終的な計算で用いられる式を与えておく。

ロジスティックモデル

$$\eta_i = \log \frac{p_i}{1-p_i} = \sum_{j=1}^{p} \beta_j x_{ij} + \beta_0$$

$$p_i = \frac{e^{\eta_i}}{1+e^{\eta_i}} = \frac{1}{1+e^{-\eta_i}}, \quad 1-p_i = \frac{1}{1+e^{\eta_i}}$$

$$\mu_i = n_i p_i = \frac{n_i e^{\eta_i}}{1+e^{\eta_i}}, \qquad \frac{\partial \mu_i}{\partial \eta_i} = \frac{n_i e^{\eta_i}}{(1+e^{\eta_i})^2} = n_i p_i (1-p_i)$$

$$U_j = \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{V[Y_i]} \frac{\partial \mu_i}{\partial \eta_i} = \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{n_i p_i (1-p_i)} n_i p_i (1-p_i) = \sum_{i=1}^{N} (y_i - \mu_i) x_{ij}$$

$$(\Im)_{jk} = \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{V[Y_i]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 = \sum_{i=1}^{N} \frac{x_{ij} x_{ik} n_i^2 p_i^2 (1-p_i)^2}{n_i p_i (1-p_i)} = \sum_{i=1}^{N} x_{ij} x_{ik} n_i p_i (1-p_i)$$

プロビットモデル

$$\eta_i = \Phi^{-1}(p_i) = \sum_{j=1}^{p} \beta_j x_{ij} + \beta_0$$

$$p_i = \Phi(\eta_i)$$

$$\mu_i = n_i p_i = n_i \Phi(\eta_i), \qquad \frac{\partial \mu_i}{\partial \eta_i} = \frac{n_i}{\sqrt{2\pi}} \exp(-\eta_i^2/2)$$

$$U_j = \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{V[Y_i]} \frac{\partial \mu_i}{\partial \eta_i} = \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{p_i (1-p_i)} \frac{\exp(-\eta_i^2/2)}{\sqrt{2\pi}}$$

$$(\Im)_{jk} = \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{V[Y_i]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 = \sum_{i=1}^{N} \frac{x_{ij} x_{ik}}{p_i (1-p_i)} \frac{n_i \exp(-\eta_i^2)}{2\pi}$$

極値モデル

$$\eta_i = \log[-\log(1-p_i)] = \sum_{j=1}^{p} \beta_j x_{ij} + \beta_0$$

$$p_i = 1 - \exp[-\exp(\eta_i)] \quad (1 - p_i = \exp[-\exp(\eta_i)])$$

$$\mu_i = n_i p_i, \qquad \frac{\partial \mu_i}{\partial \eta_i} = n_i \frac{\partial p_i}{\partial \eta_i} = n_i \exp(\eta_i)\exp[-\exp(\eta_i)]$$

$$U_j = \sum_{i=1}^{N} \frac{(y_i - \mu_i)x_{ij}}{V[Y_i]} \frac{\partial \mu_i}{\partial \eta_i} = \sum_{i=1}^{N} \frac{(y_i - \mu_i)x_{ij}}{p_i(1-p_i)} \exp(\eta_i)\exp[-\exp(\eta_i)]$$

$$= \sum_{i=1}^{N} \frac{(y_i - \mu_i)x_{ij}}{p_i} \exp(\eta_i)$$

$$(\Im)_{jk} = \sum_{i=1}^{N} \frac{x_{ij}x_{ik}}{V[Y_i]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 = \sum_{i=1}^{N} \frac{x_{ij}x_{ik}}{p_i(1-p_i)} n_i \exp(2\eta_i)\exp[-2\exp(\eta_i)]$$

$$= \sum_{i=1}^{N} \frac{x_{ij}x_{ik}}{p_i} n_i (1-p_i)\exp(2\eta_i)$$

極値モデルの計算には以下の性質を利用する。

$$\lim_{p \to 0} \exp(\eta)/p = 1$$

プロビットモデルと極値モデルの場合、$p_i \to 0$ や $p_i \to 1$ のときに、計算機のまるめ誤差や分布関数の近似誤差から、除算のエラーが生じることがある。そのため、プログラムではある程度のところで、これらの極限を止めるようにしている。また最終結果でも対数尤度の計算で同様のことが起こる可能性があるので、同じように極端な値を避けるようにしている。現在のプログラムでは、$0.000001 \le p_i \le 0.999999$ の範囲に設定している。

3.2　プログラムの利用法

メニュー［分析－多変量解析他－判別手法－2 値ロジスティック回帰］を選択すると、図 1 のような、2 値ロジスティック回帰分析の分析実行メニューが表示される。

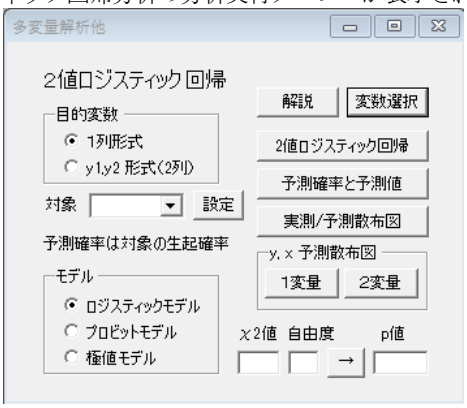
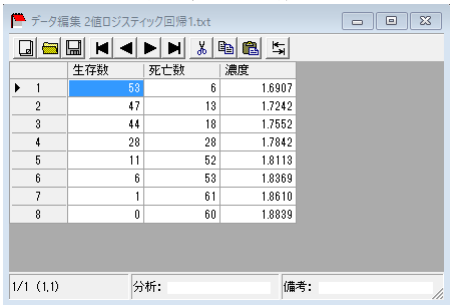


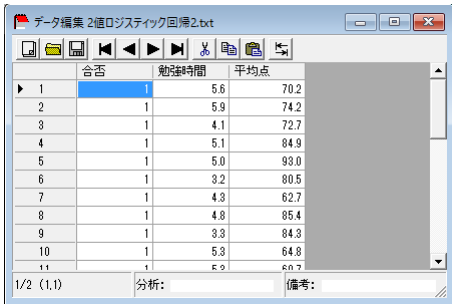
図 1　分析実行メニュー

利用するデータの形式は、「y1,y2 形式（2 列）」と「0/1 形式（1 列）」があり、それぞれ図 2a と図 2b のように、目的変数が 2 列で表されるか、1 列で表されるかの違いである。



	生存数	死亡数	濃度
1	53	6	1.6907
2	47	13	1.7242
3	44	18	1.7552
4	28	28	1.7842
5	11	52	1.8113
6	6	53	1.8369
7	1	61	1.8610
8	0	60	1.8839

図 2a　目的変数 2 列データ[3]

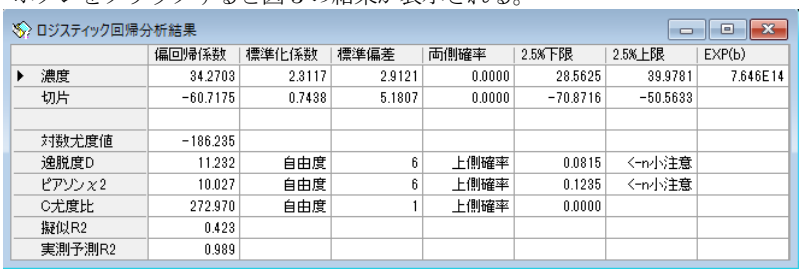


	合否	勉強時間	平均点
1	1	5.6	70.2
2	1	5.9	74.2
3	1	4.1	72.7
4	1	5.1	84.9
5	1	5.0	93.0
6	1	3.2	80.5
7	1	4.3	62.7
8	1	4.8	85.4
9	1	3.3	84.3
10	1	5.3	64.8

図 2b　目的変数 1 列データ[3]

目的変数が 2 列で表される場合は、事象 1 が何回起きて、事象 2 が何回起きたかの重複のあるデータで、1 列で表される場合は、1 回の試行で事象が起きるかどうかの重複のないデータである。2 列の場合、対象変数と非対象変数を入力し、対象変数をコンボボックスで選択しておく。対象変数の出現確率が予測確率となる。1 列のデータを事象が起きないと事象が起きたにして 2 列で表現することも可能である。目的変数が 1 列の場合は、2 列の特別な場合と考えてもよい。以後データ形式を分けて、プログラムの出力について説明する。

図 2a のデータのとき、「ロジスティックモデル」ラジオボタンを選択し、「2 値ロジスティック回帰」ボタンをクリックすると図 3 の結果が表示される。



	偏回帰係数	標準化係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
濃度	34.2703	2.3117	2.9121	0.0000	28.5625	39.9781	7.646E14
切片	-60.7175	0.7438	5.1807	0.0000	-70.8716	-50.5633	
対数尤度値	-186.235						
逸脱度D	11.232	自由度	6	上側確率	0.0815	<-n小注意	
ピアソンχ2	10.027	自由度	6	上側確率	0.1235	<-n小注意	
C尤度比	272.970	自由度	1	上側確率	0.0000		
擬似R2	0.423						
実測予測R2	0.989						

図 3　2 値ロジスティック回帰結果

ここでは回帰パラメータの値とその検定値、対数尤度値、逸脱度などが表示される。

また、「予測確率と予測値」ボタンをクリックすると、個別の実測値、予測確率、予測値が図 4 のように表示される。

	実測値	予測確率	予測値
1	6	0.059	3.457
2	13	0.164	9.842
3	18	0.362	22.451
4	28	0.605	33.898
5	52	0.795	50.096
6	53	0.903	53.291
7	61	0.955	59.222
8	60	0.979	58.743

図 4　予測確率と予測値

「実測/予測散布図」ボタンをクリックすると、この実測値と予測値が、図 5 のようにプロットされる。

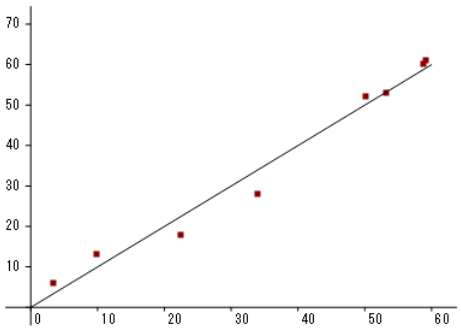


図 5　実測/予測散布図

予測の説明変数が 1 つまたは 2 つの場合、実測値と確率の予測関数（連結関数の逆関数）の関係を表示することができる。ここでは説明変数が 1 つであるので、「y, x 予測散布図」グループボックス内の「1 変量」ボタンをクリックする。結果は横軸を濃度として図 6 のようになる。

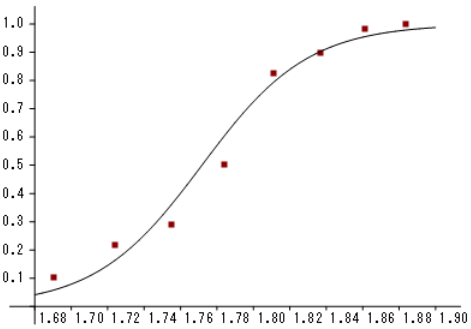


図 6　予測関数とデータ（ロジスティックモデル）

但し、ここでは軸設定を使ってグラフの軸を変更している。

この図と同様に、プロビットモデルと極値モデルの予測関数についても図 7 と図 8 で当てはまりを見てみる。

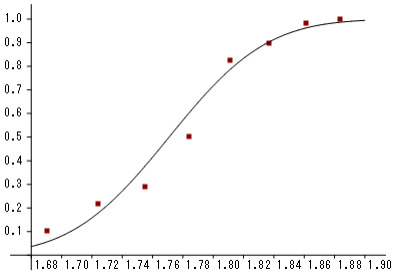


図 7　プロビットモデル

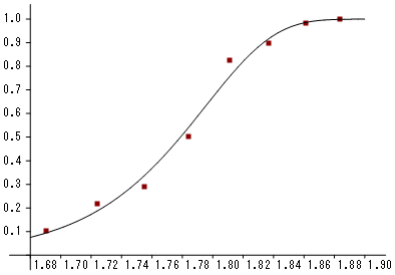


図 8　極値モデル

これらを比べると極値モデルの当てはまりが良いことが分かる。このことは、「2 値ロジスティック回帰」ボタンで表示される、対数尤度値、逸脱度 D、R^2 の値でも確認できる。

次に図 2b のデータを用いた場合のロジスティックモデルの実行結果を示す。目的変数は「0/1 形式（1 列）」を選択し、「2 値ロジスティック回帰」ボタンをクリックすると図 9 のような結果が表示される。

ロジスティック回帰分析結果

	偏回帰係数	標準化係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
▶ 勉強時間	5.0765	5.9238	2.5049	0.0427	0.1670	9.9861	1.602E02
平均点	0.4581	5.2007	0.2231	0.0400	0.0208	0.8955	1.581E00
切片	-52.7491	-20.9681	24.9005	0.0341	-101.5540	-3.9442	
対数尤度値	-5.569						
逸脱度D	11.137	自由度	27	上側確率	0.9969	<-n小注意	
ピアソンχ2	13.422	自由度	27	上側確率	0.9863	<-n小注意	
C尤度比	29.917	自由度	2	上側確率	0.0000		
擬似R2	0.729						
実測予測R2	0.864						
誤判別確率	0を1と	0.059	1を0と	0.077			

図 9　2 値ロジスティック回帰結果

このデータ形式では、確率の予測による 0/1 の判別についての誤判別確率が追加されている。

また、「予測確率と予測値」ボタンをクリックすると、個体別の実測値、予測確率、予測値が図 10 のように表示される。

予測確率と予測値

	実測値	予測確率	予測値
9	1	0.932	1.000
10	1	0.979	1.000
11	1	0.877	1.000
12	1	1.000	1.000
13	1	0.991	1.000
14	0	0.000	0.000
15	0	0.374	0.000
16	0	0.070	0.000
17	0	0.005	0.000
18	0	0.000	0.000

図 10　予測確率と予測値

ここでは、予測値として、予測確率が 0.5 未満なら 0、予測確率が 0.5 以上なら 1 が与えられている。この予測値と実測値との違いを表すのが、図 9 の誤判別確率である。

「実測/予測散布図」をクリックすると、この実測値と予測確率が、図 11 のように表示される。ここに、図 11 と図 5 の違いは、図 11 はデータが個別であるので、実測値と予測値の代わりに実測値と予測確率を用いているところである。

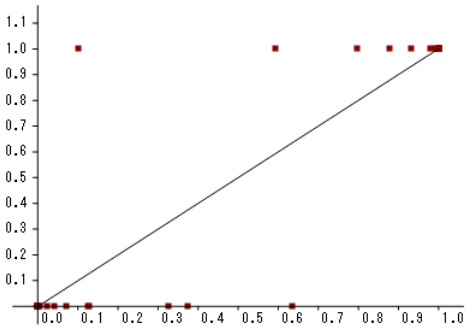


図 11　実測/予測確率散布図

このデータでは説明変数が 2 つであるので、「y, x 予測散布図」グループボックス内の「2 変量」ボタンをクリックする。結果は図 12 のようなグラフになる。

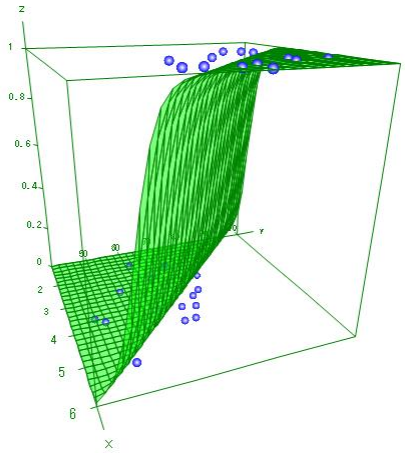


図 12　予測関数とデータ（ロジスティックモデル）

最後に、分析実行メニューの下部に、利用する可能性のある χ^2 分布の値と自由度から確率を求めるボタンを追加しておいた。必要に応じて利用してもらいたい。

4.　多値ロジスティック回帰

4.1　多項分布モデル

多項分布の密度関数、対数密度関数は以下で与えられる。ここで、対数密度関数の最後の項はパラメータに依存していないので、計算上は考えないことにする（参考文献[2] の計算に従っている）。

密度関数

$$f(y_i; p_i) = n_i! \prod_{\alpha=1}^{J} \frac{p_{i\alpha}^{y_{i\alpha}}}{y_{i\alpha}!}, \qquad \sum_{\alpha=1}^{J} y_{i\alpha} = n_i, \ \sum_{\alpha=1}^{J} p_{i\alpha} = 1$$

これより、y_i 及び p_i の中の１つは他の変数で規定される。

対数密度関数

$$l(y_i;p_i)=\log f(y_i;p_i)=\sum_{\alpha=1}^{J} y_{i\alpha}\log p_{i\alpha}+\log n_i!-\sum_{\alpha=1}^{J} y_{i\alpha}!$$

$$\rightarrow \sum_{\alpha=1}^{J} y_{i\alpha}\log p_{i\alpha}$$

以下この関係を利用して計算過程を考えてみる。

4.2　名義ロジスティックモデル

名義ロジスティックモデルでは、一般性を失わず、他の変数で規定される基準となるカテゴリを１とすると、$\alpha\neq 1$ として、

$$\frac{\partial l_i}{\partial p_{i\alpha}}=\frac{\partial}{\partial p_{i\alpha}}\left[y_{i1}\log p_{i1}+\sum_{\beta=2}^{J} y_{i\beta}\log p_{i\beta}\right]=-\frac{y_{i1}}{p_{i1}}+\frac{y_{i\alpha}}{p_{i\alpha}}$$

$$E[Y_{i\alpha}]=n_i p_{i\alpha}\equiv\mu_{i\alpha}$$

$$Cov[Y_{i\alpha}Y_{i\beta}]=n_i p_{i\alpha}(\delta_{\alpha\beta}-p_{i\beta})$$

名義ロジスティックモデルは、基準となるカテゴリに対する他のカテゴリのロジットを説明変数の線形結合で推測する。

$$\eta_{i\alpha}=\log\frac{p_{i\alpha}}{p_{i1}}=\sum_{k=1}^{p}\beta_{k\alpha}x_{ik}+\beta_{0\alpha}\quad \text{for}\ \ \alpha=2,\cdots,J$$

これより、$p_{i\alpha}=p_{i1}e^{\eta_{i\alpha}}$

また、$1=\sum_{\alpha=1}^{J} p_{i\alpha}=p_{i1}+\sum_{\alpha=2}^{J} p_{i\alpha}=p_{i1}[1+\sum_{\alpha=2}^{J} e^{\eta_{i\alpha}}]$　より、

$$p_{i1}=\frac{1}{1+\sum_{\beta=2}^{J} e^{\eta_{i\beta}}},\quad p_{i\alpha}=\frac{e^{\eta_{i\alpha}}}{1+\sum_{\beta=2}^{J} e^{\eta_{i\beta}}}$$

定義より、

$$\mu_{i1}=n_i p_{i1}=\frac{n_i}{1+\sum_{\beta=2}^{J} e^{\eta_{i\beta}}},\quad \mu_{i\alpha}=n_i p_{i\alpha}=\frac{n_i e^{\eta_{i\alpha}}}{1+\sum_{\beta=2}^{J} e^{\eta_{i\beta}}}$$

以下この関係を利用して計算過程を考えてみる。$\alpha\neq 1,\beta\neq 1$ に対して

$$\frac{\partial \mu_{i\beta}}{\partial \eta_{i\alpha}} = \frac{n_i e^{\eta_{i\beta}} \delta_{\alpha\beta}(1+\sum_{\gamma=2}^{J} e^{\eta_{i\gamma}}) - n_i e^{\eta_{i\alpha}} e^{\eta_{i\beta}}}{(1+\sum_{\gamma=2}^{J} e^{\eta_{i\gamma}})^2} = n_i p_{i\beta}(\delta_{\alpha\beta} - p_{i\alpha})$$

$$\frac{\partial \mu_{i\beta}}{\partial \beta_{j\alpha}} = \frac{\partial \eta_{i\alpha}}{\partial \beta_{j\alpha}} \frac{\partial \mu_{i\beta}}{\partial \eta_{i\alpha}} = x_{ij} \frac{\partial \mu_{i\beta}}{\partial \eta_{i\alpha}} = x_{ij} n_i p_{i\beta}(\delta_{\alpha\beta} - p_{i\alpha})$$

以上より、

$$U_j^{\alpha} = \sum_{i=1}^{N} \frac{\partial l_i}{\partial \beta_{j\alpha}} = \sum_{i=1}^{N} \sum_{\beta=2}^{J} \sum_{\gamma=2}^{J} \frac{\partial \mu_{i\beta}}{\partial \beta_{j\alpha}} \frac{\partial p_{i\gamma}}{\partial \mu_{i\beta}} \frac{\partial l_i}{\partial p_{i\gamma}}$$

$$= \sum_{i=1}^{N} \sum_{\beta=2}^{J} \sum_{\gamma=2}^{J} x_{ij} n_i p_{i\beta}(\delta_{\alpha\beta} - p_{i\alpha}) \frac{\delta_{\beta\gamma}}{n_i} \left(-\frac{y_{i1}}{p_{i1}} + \frac{y_{i\gamma}}{p_{i\gamma}} \right)$$

$$= \sum_{i=1}^{N} \sum_{\beta=2}^{J} x_{ij} p_{i\beta}(\delta_{\alpha\beta} - p_{i\alpha}) \left(-\frac{y_{i1}}{p_{i1}} + \frac{y_{i\beta}}{p_{i\beta}} \right) = \sum_{i=1}^{N} x_{ij}(y_{i\alpha} - n_i p_{i\alpha})$$

$$\Im_{jk}^{\alpha} = -\frac{\partial U_j^{\alpha}}{\partial \beta_{k\alpha}} = -\frac{\partial}{\partial \beta_{k\alpha}} \sum_{i=1}^{N} x_{ij}(y_{i\alpha} - n_i p_{i\alpha}) = \sum_{i=1}^{N} x_{ij} n_i \sum_{\beta=2}^{J} \frac{\partial \mu_{i\beta}}{\partial \beta_{k\alpha}} \frac{\partial p_{i\alpha}}{\partial \mu_{i\beta}}$$

$$= \sum_{i=1}^{N} x_{ij} n_i \sum_{\beta=2}^{J} x_{ik} n_i p_{i\beta}(\delta_{\alpha\beta} - p_{i\alpha}) \frac{\delta_{\alpha\beta}}{n_i} = \sum_{i=1}^{N} x_{ij} x_{ik} n_i p_{i\alpha}(1 - p_{i\alpha})$$

これらのスコアベクトルと情報ベクトルより、$\beta_{i\alpha}$（$\alpha \neq 1$）は推定される。

最適値からのずれを表す、逸脱度、ピアソンのχ^2統計量及び、最小モデルからのずれを表す尤度比χ^2統計量は以下のようになる。

$$D = 2\sum_{i=1}^{N} \sum_{\alpha=1}^{J} y_{i\alpha} \log \frac{y_{i\alpha}}{n_i \hat{p}_{i\alpha}} \sim \chi^2((J-1)(N-p))$$

$$\chi^2 = \sum_{i=1}^{N} \sum_{j=1}^{J} \frac{(y_{ij} - \hat{y}_{ij})^2}{n_i \hat{p}_{ij}} \sim \chi^2((J-1)(N-p))$$

$$C = 2\sum_{i=1}^{N} \sum_{\alpha=1}^{J} y_{i\alpha} \log \frac{\hat{y}_{i\alpha}}{n_i \tilde{p}_{\alpha}} \sim \chi^2((J-1)(p-1)), \quad \tilde{p}_{\alpha} = \sum_{i=1}^{N} y_{i\alpha} \Big/ \sum_{i=1}^{N} n_i$$

ピアソンのχ^2統計量は逸脱度と漸近的に同じ指標であるが、逸脱度と比べてこちらの方が分布によく適合するという意見もある。ここではn_i回の試行に対して、y_i回の事象が起こったとしているが、1回の試行で起こったか起こらないかにする場合は、$n_i = 1, y_i = \{0, 1\}$とする。但し、分布はデータ数が無限大のときの極限であるので、注意が必要である。

4.3　順序ロジスティックモデル

順序ロジスティックモデルには、累積ロジットモデル、隣接カテゴリロジットモデル、連続比ロジットモデルなどがあるが、ここではプログラムに取り入れている累積ロジットモデルについて説明する。

累積ロジットモデルでは、以下の比の対数を線形関数で予測する。

$$\frac{p_1}{p_2+\cdots+p_J}=e^{\eta_1},\quad \frac{p_1+p_2}{p_3+\cdots+p_J}=e^{\eta_2},\quad \cdots,\quad \frac{p_1+\cdots+p_{J-1}}{p_J}=e^{\eta_{J-1}}$$

これは、連続した複数のカテゴリの出現確率と残りのカテゴリの出現確率のオッズ比を説明変数の線形関数で予測することに相当する。

上の関係を以下のように書き換え、

$$\frac{p_2+\cdots+p_J}{p_1}=e^{-\eta_1},\quad \frac{p_3+\cdots+p_J}{p_1+p_2}=e^{-\eta_2},\quad \cdots,\quad \frac{p_J}{p_1+\cdots+p_{J-1}}=e^{-\eta_{J-1}}$$

$q_\alpha=p_1+p_2+\cdots+p_\alpha$ と定義すると、以下の関係が示される。

$$1-p_1=p_2+\cdots+p_J=p_1e^{-\eta_1}\ \text{より、}\ p_1=\frac{e^{\eta_1}}{1+e^{\eta_1}}=q_1$$

$$p_2=p_1e^{-\eta_1}-(p_1+p_2)e^{-\eta_2}\ \text{より、}$$

$$p_2=\frac{1}{1+e^{-\eta_2}}-\frac{1}{1+e^{-\eta_1}},\quad p_1+p_2=\frac{e^{\eta_2}}{1+e^{\eta_2}}=q_2$$

同様にして、

$$p_{J-1}=\frac{1}{1+e^{-\eta_{J-1}}}-\frac{1}{1+e^{-\eta_{J-2}}},\quad p_1+\cdots+p_{J-1}=\frac{e^{\eta_{J-1}}}{1+e^{\eta_{J-1}}}=q_{J-1}$$

また、

$$p_J=1-(p_1+\cdots+p_{J-1})=1-\frac{1}{1+e^{-\eta_{J-1}}}=\frac{1}{1+e^{\eta_{J-1}}}=1-q_{J-1}$$

これらより、q_α について考えれば、各カテゴリ α について独立に、q_α と $1-q_\alpha$ の２項分布として β_α の値を推定できることが分かる。そのためこれは２値ロジスティック回帰の拡張として捉えることができ、各カテゴリ p_α（$\alpha=1,2,\cdots,J$）については以下のように与えることができる。

$$p_1=q_1,\ p_\alpha=q_\alpha-q_{\alpha-1},\ p_J=1-q_{J-1}$$

4.4　プログラムの利用法

メニュー［分析－多変量解析等－判別手法－多値ロジスティック回帰］を選択すると図１のような多値ロジスティック回帰分析の分析実行メニューが表示される。

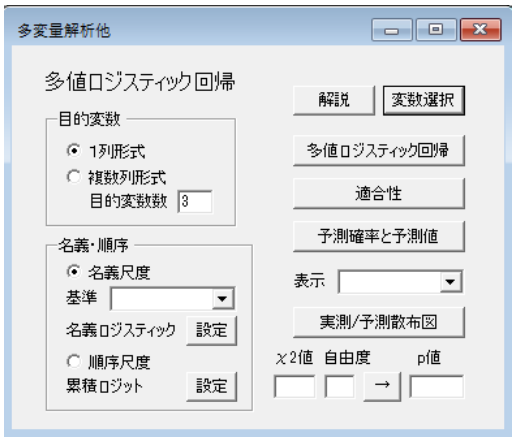


図 1　分析実行メニュー

複数列形式のデータの例を図 2 に示す。

	重要でない	重要	とても重要	性別	年齢	職業
1	26	12	7	0	0	0
2	9	21	15	0	1	0
3	5	14	41	0	0	1
4	40	17	8	1	0	0
5	17	15	12	1	1	0
6	8	15	18	1	0	1

図 2　複数列形式のデータ

「目的変数」グループボックスの「複数列形式」を選択し、「目的変数数」テキストボックスに「3」を入力する。変数選択ですべての変数を選択し、「名義ロジスティック」の設定から図 3 のように基準に「重要でない」を選択する。

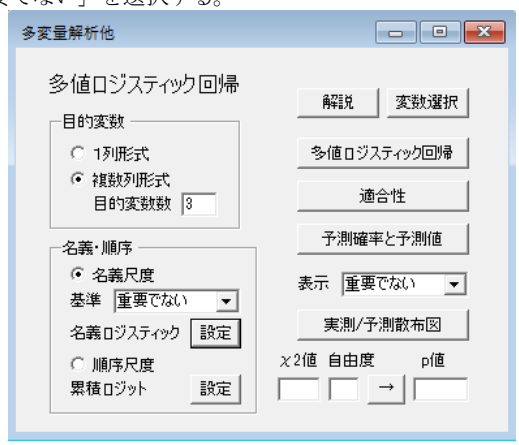


図 3　複数列目的変数の名義ロジスティック設定

ここでは、「重要でない」カテゴリの確率で、他のカテゴリの確率を割った対数オッズを説明変数の線形関数で推定することになる。

「多値ロジスティック回帰」ボタンをクリックすると図４のような分析結果が表示される。

log(確率/重要でない確率)の線形予測

	偏回帰係数	標準化係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
▶ 重要/重要でない							
性別	-0.3881	-0.2126	0.2528	0.1247	-0.8835	0.1073	0.6783
年齢	1.1283	0.5826	0.3064	0.0002	0.5277	1.7289	3.0903
職業	1.5877	0.8199	0.3108	0.0000	0.9785	2.1969	4.8925
切片	-0.5908	-0.0616	0.2595	0.0228	-1.0993	-0.0823	
とても重要/重要でない							
性別	-0.8130	-0.4453	0.2710	0.0027	-1.3442	-0.2818	0.4435
年齢	1.4781	0.7633	0.3640	0.0000	0.7647	2.1916	4.3846
職業	2.9167	1.5062	0.3463	0.0000	2.2380	3.5955	18.4805
切片	-1.0391	-0.0668	0.3072	0.0007	-1.6412	-0.4370	

図４　対数オッズの推定

ここでは、オッズ比推定の偏回帰係数、標準化偏回帰係数、偏回帰係数の標準誤差、偏回帰係数が０となる検定確率、偏回帰係数の下限と上限、説明変数の単位量変化によるオッズ比の変化量が表示される。

「適合性」ボタンをクリックすると、図５のように各種の適合性指標が表示される。

適合性

▶ 対数尤度値	-290.351				
逸脱度D	3.939	自由度	4	上側確率	0.4144
ピアソンχ2	3.927	自由度	4	上側確率	0.4160
C尤度比	77.842	自由度	6	上側確率	0.0000
擬似R2	0.118				
実測予測R2	0.981				

図５　適合性指標

「予測確率と予測値」ボタンをクリックすると、図６のような結果が表示される。

予測確率と予測値

	重要でない	予測確率	予測値	重要	予測確率	予測値	とても重要	予測確率	予測値
▶	26	0.524	23.589	12	0.290	13.066	7	0.185	8.345
	9	0.235	10.556	21	0.402	18.069	15	0.364	16.375
	5	0.098	5.855	14	0.264	15.865	41	0.638	38.280
	40	0.652	42.411	17	0.245	15.934	8	0.102	6.655
	17	0.351	15.444	15	0.408	17.931	12	0.241	10.625
	8	0.174	7.145	15	0.320	13.134	18	0.505	20.721

図６　予測確率と予測値

これには３つのカテゴリについての実測値、予測確率、予測値が表示される。「表示」カテゴリを１つ選んで「実測/予測散布図」ボタンをクリックすると、図７のような散布図が表示される。

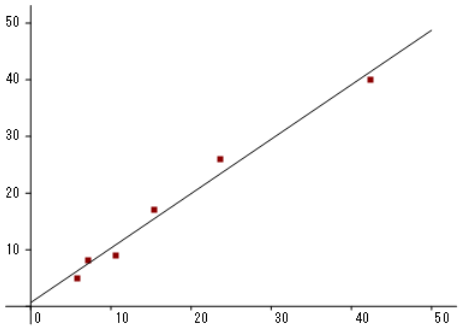


図７　実測/予測散布図

同じデータを順序尺度として、順序ロジスティックの累積ロジットモデルで分析すると図８のような結果を得る。

log(累積確率/他累積確率)の線形予測

	偏回帰係数	標準化係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
▶ 重要でない/重要～							
性別	0.5723	0.3135	0.2708	0.0346	0.0416	1.1031	1.7724
年齢	-1.2597	-0.6505	0.3064	0.0000	-1.8602	-0.6591	0.2838
職業	-2.2490	-1.1614	0.3577	0.0000	-2.9501	-1.5479	0.1055
切片	0.0746	-0.6751	0.2492	0.7647	-0.4139	0.5631	
～重要/とても重要							
性別	0.5930	0.3248	0.2710	0.0286	0.0619	1.1241	1.8095
年齢	-0.9711	-0.5015	0.3638	0.0076	-1.6841	-0.2581	0.3787
職業	-2.1137	-1.0915	0.3462	0.0000	-2.7923	-1.4351	0.1208
切片	1.5266	0.8221	0.3087	0.0000	0.9216	2.1317	

図８　累積ロジットモデルでの結果

これは最初が「重要でない」を「重要」と「とても重要」を足したカテゴリで割った対数オッズ、次が「重要でない」と「重要」を足したカテゴリを「とても重要」で割った対数オッズについて、説明変数の線形関数での推定である。

最後に目的変数が同じファイル２頁目の「1列形式」で与えられる場合、「適合性」の結果に図９のように誤判別確率の値が表示される。

適合性

▶ 対数尤度値	-8.299				
逸脱度D	16.597	自由度	54	上側確率	1.0000
ピアソンχ2	14.657	自由度	54	上側確率	1.0000
C尤度比	45.054	自由度	4	上側確率	0.0000
擬似R2	0.731				
実測予測R2	0.848				
	Aを他と	Bを他と	Cを他と		
誤判別確率	0.000	0.231	0.250		

図９　１列形式の場合の適合性結果

5.　おわりに

この報告では、メタ分析、2 値ロジスティック回帰分析、多値ロジスティック回帰分析のプログラムについて、背景となる理論も含めて紹介した。特に後者 2 つの分析については一般化線形モデルの基礎から解説したので、理論の部分が非常に長くなった。

メタ分析については、効果量とデータ数及び、効果量と分散という 2 種類のデータから分析が実行できるようにした。これらのデータ間の変換はプログラム開発中に考えたものであるが、これで非常に便利に使えるようになったものと思う。ただ、効果量として非常によく利用される t 統計量や χ^2 統計量を用いていないことが気になる。t 統計量については、全データ数が分かると、近似計算を用いて標準化平均値差に変換することができるが、χ^2 統計量については全データ数だけでは対数オッズ比に変換できず、分割表の実数が必要になる。通常分割表は文献に記載されていることが殆どなので、問題は少ないと思われるが、理論的統一性の面では気にかかる部分である。また、メタ分析は、実用としてどの程度役に立つものか、未知の部分が多い。今後何回か分析を実行することによってプログラムの改善が必要になるものと思われる。

2 値ロジスティック分析については、一般化線形モデルの 2 項分布のパラメータの推定で、3 種類の連結関数を用いたモデルを提供している。分析の名前はその中の代表的なロジスティックモデルを元にしている。データは 2 値の度数データと 2 値の個体別データを利用できる。個体別データの場合は、2 群の判別分析と同じような使い方ができる。但し、群の推測において、個体別の判別関数値ではなく、事象の出現確率が表示される。判別の分点に相当するものは出現確率で 0.5 の値である。

多値ロジスティック分析については、一般化線形モデルの多項分布のパラメータ推定で、目的変数が名義尺度の場合と順序尺度の場合に分けて結果が表示される。名義尺度の場合は名義ロジスティックモデル、順序尺度の場合は累積ロジットモデルが使われる。他のモデルについては必要が生じたら追加する。データは多値の度数データと多値の個体別データを利用できる。個体別データの場合は、多群の判別分析と同じような使い方ができるが、この分析では、事象の出現確率が最大と予測される事象が判別値となる。

参考文献

[1] 山田剛史, 井上俊哉編, メタ分析入門 心理・教育研究の系統的レビューのために, 東京大学出版会, 2012.

[2] Annette J. Dobson 著, 田中豊他訳, 一般化線形モデル入門 原著第 2 版, 共立出版, 2008.

[3] ホームページ http://www.heisei-u.ac.jp/ba/fukui/analysis.html 内のサンプルデータ Samples.zip 内のファイル

Multi-purpose Program for Social System Analysis 28
- Meta-Analysis, Logistic Regression Analysis -

Masayasu FUKUI[*1], Kazuki KODAMA[*1] and Makoto OZAKI[*1]

[*1] *Department of Business Administration, Faculty of Business Administration, Fukuyama Heisei University*

Abstract: We have been constructing a unified program on the social system analysis for the purpose of education. This time, we make program of meta-analysis which is a method for leading stronger conclusion using with past research materials, and of Logistic regression analysis to calculate the probability of binomial distribution or multinomial distribution as a function of polynomial of independent variables. We explain the theory of these analyses and introduce our programs in this paper.

Keywords: College Analysis, multivariate analysis, generalized linear model, meta-analysis, logistic regression analysis

URL: http://www.heisei-u.ac.jp/ba/fukui/

福山平成大学経営学部紀要
第 13 号（2017），99－120 頁

社会システム分析のための統合化プログラム２９
－パラメータ設計・オンライン品質工学－

福井 正康[*1]、大山 知之[*2]、織田 望[*2]

[*1] 福山平成大学経営学部経営学科

[*2] 日本ニューマチック工業株式会社

概要：我々は教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた。今回は経営科学の分野で品質管理に利用されるパラメータ設計（オフライン品質工学）とオンライン品質工学についてのプログラムを紹介する。パラメータ設計は製品の品質を設計段階で高めようとする手法で、SN 比を基に最善の制御因子を求める手法である。オンライン品質工学は品質を金額ベースで表し、工場のラインでの計測とそのフィードバックを管理する手法である。

キーワード：College Analysis，経営科学，パラメータ設計，オフライン品質工学，オンライン品質工学

URL：http://www.heisei-u.ac.jp/ba/fukui/

1. はじめに

我々はこれまで社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた。今回は経営科学の分野の品質管理で利用される、オフライン品質工学とも呼ばれるパラメータ設計[1] とオンライン品質工学[2] を取り上げる。これらのプログラムは実際の製造業の分野での利用を考えて作られたものであり、今後の応用が重要である。

製品の品質は設計段階でほぼ制御できるという考えをもとに、実験計画法に基づき製品試験を行い、SN 比を基にして最善の制御因子の組み合わせを求め、製造後の故障や不良品の発生を抑えようとする手法がパラメータ設計である。これにより製品製造後の返品や品質改善のためのフィードバックをなくすことができ、効率的な製品開発を目指すことができる。

我々はこのパラメータ設計の過程を学び、実践するためのプログラムを開発した。我々のプログラムはあくまで汎用のもので、実際の問題にはその問題に特化した個別の修正が必要になるかも知れないが、実験的にパラメータ設計に取り組む際の有効性を確かめるには十分であると思われる。

工場の生産ラインでは、目標特性の製品を作るために、品質特性値を計測し、ラインを調整する。計測回数を増やしラインを調整すると、不良品の発生が低減されることは明らかであるが、これらには時間と人手を必要とし、そこには費用が発生する。オンライン品質工学では、

品質を金額ベースで表し、最適な計測間隔と調整間隔を求め、現行の値と比較し、どの程度の金額の削減につながるかを検討する。

この報告では、各章が独立しているため、式番号、図表番号は章ごとに付けるものとする。章を越えて参照する場合は、その旨を示す。

2. パラメータ設計

2.1 パラメータ設計の理論

2.1.1 動特性パラメータ設計

ここではまずゼロ点比例式の動特性パラメータ設計について考えるが、その前に SN 比と感度について、理論的な考察を加えておく。

1 つの実験では、信号水準 M_j（$j=1,\cdots,p$）と誤差水準 N_α（$\alpha=1,\cdots,n$）によって、表 1 のように pn 個のデータ $y_{j\alpha}$ が得られる。誤差水準はできるだけ広く散らばるよう配慮されるものとする。

表 1　動特性パラメータ設計におけるデータ

M_1			…	M_p		
N_1	…	N_n	…	N_1	…	N_n
y_{11}	…	y_{1n}	…	y_{p1}	…	y_{pn}

この実験について、誤差水準 α のゼロ比例式回帰直線を考える。実測値 $y_{j\alpha}$ についての推定回帰式を $Y_{j\alpha}=b_\alpha M_j$ とすると、実測値との差の 2 乗和は以下となる。

$$EV_\alpha=\sum_{j=1}^{p}(y_{j\alpha}-Y_{j\alpha})^2=\sum_{j=1}^{p}(y_{j\alpha}-b_\alpha M_j)^2$$

これを最小化するには、

$$\frac{\partial}{\partial b_\alpha}EV_\alpha=-2\sum_{j=1}^{p}M_j(y_{j\alpha}-b_\alpha M_j)=0$$

として、以下を得る。

$$b_\alpha=\sum_{j=1}^{p}M_j y_{j\alpha}\Big/\sum_{j=1}^{p}M_j^2=\frac{L_\alpha}{r},\quad \text{ここに、}L_\alpha=\sum_{j=1}^{p}M_j y_{j\alpha},\ r=\sum_{j=1}^{p}M_j^2$$

データ全体のゼロ比例式回帰直線については、推定回帰式を $Y_j=bM_j$ とすると、実測値との差の 2 乗和は以下となる。

$$EV=\sum_{j=1}^{p}\sum_{\alpha=1}^{n}(y_{j\alpha}-Y_j)^2=\sum_{j=1}^{p}\sum_{\alpha=1}^{n}(y_{j\alpha}-bM_j)^2$$

これを最小化するには、

$$\frac{\partial}{\partial b}EV = -2\sum_{j=1}^{p}\sum_{\alpha=1}^{n}M_j(y_{j\alpha} - bM_j) = 0$$

として、以下を得る。

$$b = \sum_{j=1}^{p}\sum_{\alpha=1}^{n}M_j y_{j\alpha} \Big/ \sum_{j=1}^{p}\sum_{\alpha=1}^{n}M_j^2 = \frac{1}{nr}\sum_{j=1}^{p}\sum_{\alpha=1}^{n}M_j y_{j\alpha} = \frac{1}{nr}\sum_{\alpha=1}^{n}L_\alpha = \frac{1}{n}\sum_{\alpha=1}^{n}b_\alpha$$

次にデータの変動について考察する。まず、$y=0$ からの全体の変動 S_T は以下となる。

$$S_T = \sum_{j=1}^{p}\sum_{\alpha=1}^{n}y_{j\alpha}^2 \qquad \text{自由度 } pn$$

また、$y=0$ からの全体の回帰変動 S_β は以下となる。

$$S_\beta = \sum_{j=1}^{p}\sum_{\alpha=1}^{n}(bM_j)^2 = nrb^2 = \frac{1}{nr}\left(\sum_{\alpha=1}^{n}L_\alpha\right)^2 \qquad \text{自由度 1}$$

これより、$b^2 = S_\beta \big/ nr$ となる。

誤差水準 α の回帰直線の全体の回帰直線からの変動 $S_{N\beta}$ は以下となる。

$$S_{N\beta} = \sum_{j=1}^{p}\sum_{\alpha=1}^{n}(b_\alpha M_j - bM_j)^2$$

$$= r\sum_{\alpha=1}^{n}(b_\alpha - b)^2 = r\sum_{\alpha=1}^{n}b_\alpha^2 - nrb^2 = \sum_{\alpha=1}^{n}L_\alpha^2 \Big/ r - S_\beta$$

$$\text{自由度 } n-1 \quad \text{束縛条件} \quad \sum_{\alpha=1}^{n}(b_\alpha - b) = 0$$

各点の誤差水準 α の回帰直線からの変動 S_e は以下となる。

$$S_e = \sum_{j=1}^{p}\sum_{\alpha=1}^{n}(y_{j\alpha} - b_\alpha M_j)^2 = \sum_{j=1}^{p}\sum_{\alpha=1}^{n}[y_{j\alpha} - (b_\alpha - b)M_j - bM_j]^2$$

$$= \sum_{j=1}^{p}\sum_{\alpha=1}^{n}y_{j\alpha}^2 + \sum_{j=1}^{p}\sum_{\alpha=1}^{n}(b_\beta - b)^2 M_j^2 + \sum_{j=1}^{p}\sum_{\alpha=1}^{n}b^2 M_j^2 - 2\sum_{j=1}^{p}\sum_{\alpha=1}^{n}y_{j\alpha}b_\alpha M_j$$

$$= S_T + S_{NB} + S_\beta - 2\sum_{\alpha=1}^{n}L_\alpha^2 \Big/ r^2$$

$$= S_T + S_{NB} + S_\beta - 2(S_{NB} + S_\beta) = S_T - S_{NB} - S_\beta$$

$$\text{自由度 } pn-n \quad \text{束縛条件} \quad \sum_{j=1}^{p}M_j(y_{j\alpha} - b_\alpha M_j) = 0$$

誤差水準αの回帰直線からの変動の不偏分散V_eは、変動を自由度で割って以下となる。

$$V_e = S_e/(pn-n)$$

これが各回帰直線からのずれの分散$V[\varepsilon]$の不偏推定量となることは、後に説明する。

ここまでの議論で、全変動S_Tは、全体の回帰変動S_β、全体の回帰直線からの誤差水準αの回帰変動$S_{N\beta}$、各点の誤差水準αの回帰直線からの変動S_eの和で以下のように表されることが分かった。

$$S_T = S_\beta + S_{N\beta} + S_e$$

各点の全体の回帰直線からの変動S_Nは以下となる。

$$S_N = \sum_{j=1}^{p}\sum_{\alpha=1}^{n}(y_{j\alpha} - bM_j)^2 = \sum_{j=1}^{p}\sum_{\alpha=1}^{n} y_{j\alpha}^2 + nb^2\sum_{j=1}^{p} M_j^2 - 2b\sum_{j=1}^{p}\sum_{\alpha=1}^{n} y_{j\alpha}M_j$$

$$= S_T + nrb^2 - 2nrb^2 = S_T - S_\beta = S_{N\beta} + S_e$$

自由度 $pn-1$　束縛条件　$\sum_{j=1}^{p}\sum_{\alpha=1}^{n} M_j(y_{j\alpha} - bM_j) = 0$

各点の全体の回帰直線からの変動の不偏分散V_Nは、変動を自由度で割って以下となる。

$$V_N = S_N/(pn-1)$$

これが全変動の分散$V[\sigma]$の不偏推定量になっていることは、後に説明する。

これらを使って、SN 比ηと感度Sを定義する。SN 比は、測定誤差の分散σ^2に対する有効な信号の変化の大きさβ^2の比を用いて、また感度Sはβ^2の値を用いて以下のように定義される。

SN 比：$\eta = 10\log_{10}\dfrac{\beta^2}{\sigma^2}$,　　感度：$S = 10\log_{10}\beta^2$

実際の計算ではβ^2とσ^2の値は不明であるので、これらの不偏推定量を用いて置き換える。

SN 比：$\eta = 10\log_{10}\left[\dfrac{(S_\beta - V_e)/nr}{V_N}\right]$,　感度：$S = 10\log_{10}\left[(S_\beta - V_e)/nr\right]$

一般に SN 比は大きな値ほど、有効な信号を誤差の中から拾いやすくなり、良好な結果である。また、感度は対象により、大きな値がよい場合、小さな値がよい場合、目標値がよい場合など様々であるが、感度があまり変化しない制御因子を用いて SN 比を上げることを考える場合もある。

最後に、β^2の不偏推定量を求めておく。

$$y_{j\alpha} = \beta_\alpha M_j + \varepsilon_{j\alpha}\,,\quad E[\varepsilon_{j\alpha}] = 0\,,\quad E[\varepsilon_{j\alpha}\varepsilon_{j'\alpha'}] = \delta_{jj'}\delta_{\alpha\alpha'}V[\varepsilon]$$

とすると、

$$b_\alpha = \frac{1}{r}\sum_{j=1}^{p} M_j y_{j\alpha} = \frac{1}{r}\sum_{j=1}^{p} M_j(\beta_\alpha M_j + \varepsilon_{j\alpha}) = \beta_\alpha + \frac{1}{r}\sum_{j=1}^{p} M_j \varepsilon_{j\alpha}$$

より

$$S_e = \sum_{j=1}^{p}\sum_{\alpha=1}^{n}(\beta_\alpha M_j + \varepsilon_{j\alpha} - b_\alpha M_j)^2 = \sum_{j=1}^{p}\sum_{\alpha=1}^{n}\left[(\beta_\alpha - b_\alpha)M_j + \varepsilon_{j\alpha}\right]^2$$

$$= \sum_{j=1}^{p}\sum_{\alpha=1}^{n}\left[-\frac{1}{r}M_j\sum_{j'=1}^{p}M_{j'}\varepsilon_{j'\alpha} + \varepsilon_{j\alpha}\right]^2$$

$$= \sum_{j=1}^{p}\sum_{\alpha=1}^{n}\left[\frac{1}{r^2}M_j^2\left(\sum_{j'=1}^{p}M_{j'}\varepsilon_{j'\alpha}\right)^2 - \frac{2}{r}M_j\varepsilon_{j\alpha}\sum_{j'=1}^{p}M_{j'}\varepsilon_{j'\alpha} + \varepsilon_{j\alpha}^2\right]$$

$$= -\frac{1}{r}\sum_{\alpha=1}^{n}\sum_{j=1}^{p}\sum_{j'=1}^{p}M_j M_{j'}\varepsilon_{j\alpha}\varepsilon_{j'\alpha} + \sum_{j=1}^{p}\sum_{\alpha=1}^{n}\varepsilon_{j\alpha}^2$$

となり、

$$E\left[S_e\right] = E\left[-\frac{1}{r}\sum_{\alpha=1}^{n}\sum_{j=1}^{p}\sum_{j'=1}^{p}M_j M_{j'}\varepsilon_{j\alpha}\varepsilon_{j'\alpha} + \sum_{j=1}^{p}\sum_{\alpha=1}^{n}\varepsilon_{j\alpha}^2\right]$$

$$= (pn - n)V[\varepsilon]$$

すなわち、

$$V[\varepsilon] = E[S_e/(pn-n)] = E[V_e] \tag{1}$$

また、

$$b = \sum_{j=1}^{p}\sum_{\alpha=1}^{n}M_j y_{j\alpha} \Big/ \sum_{j=1}^{p}\sum_{\alpha=1}^{n}M_j^2 = \frac{1}{nr}\sum_{j=1}^{p}\sum_{\alpha=1}^{n}M_j y_{j\alpha}$$

より、

$$b^2 = \frac{1}{n^2r^2}\left(\sum_{j=1}^{p}\sum_{\alpha=1}^{n}M_j(\beta_\alpha \mathrm{M}_j + \varepsilon_{j\alpha})\right)^2$$

$$= \frac{1}{n^2r^2}\sum_{j=1}^{p}\sum_{\alpha=1}^{n}\sum_{j'=1}^{p}\sum_{\alpha'=1}^{n}M_j M_{j'}(\beta_\alpha\beta_{\alpha'}M_j M_{j'} + 2\beta_{\alpha'}\varepsilon_{j\alpha}M_{j'} + \varepsilon_{j\alpha}\varepsilon_{j'\alpha'})$$

$$= \frac{1}{n^2r^2}\left[\beta^2n^2r^2 + 2\beta nr\sum_{j=1}^{p}\sum_{\alpha=1}^{n}\varepsilon_{j\alpha} + \sum_{j=1}^{p}\sum_{\alpha=1}^{n}\sum_{j'=1}^{p}\sum_{\alpha'=1}^{n}M_j M_{j'}\varepsilon_{j\alpha}\varepsilon_{j'\alpha'}\right]$$

であるから、

$$E\left[b^2\right]=\frac{1}{n^2r^2}E\left[\beta^2n^2r^2+2\beta nr\sum_{j=1}^{p}\sum_{\alpha=1}^{n}\varepsilon_{j\alpha}+\sum_{j=1}^{p}\sum_{\alpha=1}^{n}\sum_{j'=1}^{p}\sum_{\alpha'=1}^{n}M_jM_{j'}\varepsilon_{j\alpha}\varepsilon_{j'\alpha'}\right]$$

$$=\beta^2+\frac{1}{n^2r^2}\sum_{j=1}^{p}\sum_{\alpha=1}^{n}\sum_{j'=1}^{p}\sum_{\alpha'=1}^{n}M_jM_{j'}\delta_{jj'}\delta_{\alpha\alpha'}V[\varepsilon]$$

$$=\beta^2+\frac{1}{nr}V[\varepsilon]$$

よって、以下となる。

$$\beta^2=E[b^2]-\frac{1}{nr}V[\varepsilon] \tag{2}$$

(2)と(1)、及び $b^2=S_\beta/nr$ の関係から、

$$\beta^2=\frac{1}{nr}E\left[S_\beta\right]-\frac{1}{nr}E\left[V_e\right]=E\left[(S_\beta-V_e)/nr\right] \tag{3}$$

すなわち、β^2 の不偏推定量は $(S_\beta-V_e)/nr$ である。

同様の考え方で σ^2 の不偏推定量が $V_N=S_N/(pn-1)$ であることも示すことができる。

次に我々は SN 比を最大にする制御因子の最適設定について考える。制御因子 A, B, …について直交表を作ると、他の制御因子の影響をならして、１つの制御因子の影響を調べることができるようになる。表２に直交表を加えたデータを示す。

表２　動特性パラメータ設計における直交表とデータ

	A	B	…	M_1			…	M_p			SN 比	感度
				N_1	…	N_n	…	N_1	…	N_n		
1	1	1	…	y_{111}	…	y_{11n}	…	y_{1p1}	…	y_{1pn}	η_1	S_1
2	1	1	…	y_{211}	…	y_{21n}	…	y_{2p1}	…	y_{2pn}	η_2	S_2
:	:	:	:	:	:	:	:	:	:	:	:	:
d	2	2	…	y_{d11}	…	y_{d1n}	…	y_{dp1}	…	y_{dpn}	η_d	S_d

ここに SN 比と感度は上で述べた方法で求めて加えてあるものとする。直交表は、各制御因子の同じ番号の行を見ると、他の制御因子について、すべての番号が同じ数だけ入っているという特徴を持つ。

例えば制御因子 A が k になる行について、SN 比及び感度の平均を取ったものをそれぞれ $\bar{\eta}_{A=k}$ 、$\bar{S}_{A=k}$ と書くとすると、SN 比の補助表は表 3 のようになる。感度の補助表も同様である。

表 3　SN 比の補助表

制御因子	水準 1	…	水準 r
A	$\bar{\eta}_{A=1}$	…	$\bar{\eta}_{A=r}$
B	$\bar{\eta}_{B=1}$	:	$\bar{\eta}_{B=r}$
:	:	…	:

ここに水準の少ない制御因子の場合、その部分は空欄にしておく。

この補助表の SN 比の中で、制御因子ごとの水準値の最も大きな水準を並べたものを最適条件といい、例えば A1B2C1D3…などと表す。我々のプログラムでは制御因子名は省略して番号だけで表している。この最適な水準の SN 比を合計したものを SN 比の最適値という。感度についても SN 比の最適条件を用いて最適値を定義する。

これに対して現実の制御因子の設定を比較条件または現状条件という。この条件を用いて SN 比を合計したものを SN 比の比較値または現状値という。感度についても同様である。最適値と比較値の差は、今後の改善の可能性として検討すべき値である。

ここで述べた水準値は理論的な推測値である。この値が妥当なものかどうか、追実験をして検証しておかなければならない。また、現実的に考えて最適な制御条件が最良のものであるとは限らない。その際は、できるだけ SN 比の値を落とさず、感度で制御因子の調整を行うこともある。

2.1.2　静特性パラメータ設計

次にこれまでの動特性と異なり、静特性のパラメータ設計について述べる。これはデータをある指定された値へ近づけることを目標としたパラメータ設計である。指定された値への近づけ方として、ある値へ上下から近づける場合を一般の望目特性、0 値へ上下から近づける場合をゼロ目標の望目特性、0 値へ上から近づける場合を望小特性、できるだけ大きな値にする場合を望大特性という。

今、静特性のデータを表 4 とする。

表 4　静特性パラメータ設計における直交表とデータ

	A	B	…	N_1	…	N_n	SN比	感度
1	1	1	…	y_{11}	…	y_{1n}	η_1	S_1
2	1	1	…	y_{21}	…	y_{2n}	η_2	S_2
:	:	:	:	:	:	:	:	:
d	2	2	…	y_{d1}	…	y_{dn}	η_d	S_d

それぞれの望目特性に対する SN 比と感度を示すと以下のようにまとめられる。但し、個々の実験の番号は省略してある。

一般の望目特性

SN 比 $\eta = 10\log\dfrac{(S_m - V_e)/n}{V_e}$, $S = 10\log\,[(S_m - V_e)/n]$

ここに、 $S_T = \sum_{\alpha=1}^{n} y_\alpha^2$, $S_m = \dfrac{1}{n}\left(\sum_{\alpha=1}^{n} y_\alpha\right)^2$, $S_e = S_T - S_m$, $V_e = \dfrac{S_e}{n-1}$

ゼロ目標の望目特性

SN 比 $\eta = 10\log(1/V_e)$, 感度 $S = \bar{y}$ （平均）

ここに、 $S_T = \sum_{\alpha=1}^{n} y_\alpha^2$, $S_m = \dfrac{1}{n}\left(\sum_{\alpha=1}^{n} y_\alpha\right)^2$, $S_e = S_T - S_m$, $V_e = \dfrac{S_e}{n-1}$

望小特性

SN 比 $\eta = 10\log(1/\sigma^2) = -10\log\sigma^2$ ここに、$S_T = \sum_{\alpha=1}^{n} y_\alpha^2$, $\sigma^2 = S_T/n$

望大特性

SN 比 $\eta = 10\log(1/\sigma^2) = -10\log\sigma^2$ ここに、$S_T = \sum_{\alpha=1}^{n} 1/y_\alpha^2$, $\sigma^2 = S_T/n$

2.2 プログラムの利用法

パラメータ設計のデータは図 1 のように、左の直交表の部分と右の実験結果の部分に分けられる[3]。

	E	F	G	H	0.025	0.025	0.1	0.1	0.5	0.5
1	1	1	1	1	0.046	0.040	0.068	0.056	0.235	0.276
2	2	2	2	2	0.052	0.036	0.085	0.078	0.293	0.311
3	3	3	3	3	0.071	0.079	0.111	0.109	0.327	0.321
4	2	2	3	3	0.043	0.037	0.072	0.050	0.224	0.246
5	3	3	1	1	0.058	0.038	0.083	0.082	0.283	0.252
6	1	1	2	2	0.076	0.074	0.104	0.131	0.371	0.404
7	1	3	2	3	0.030	0.027	0.062	0.065	0.278	0.298
8	2	1	3	1	0.044	0.036	0.108	0.086	0.354	0.361

図 1 パラメータ設計のデータ

ここでは、制御因子が A～H の 8 種類、信号因子が 3 種類、誤差因子が 2 種類である。信号因子と誤差因子の部分の変数名には信号因子の数値が与えられている。最後の 2 行に制御因子が空欄の行が含まれることがあるが、これは後に再現性確認を行い実験結果を格納するためもので、あってもなくても再現性確認までの計算に問題はない。

メニュー［分析－OR－品質管理－パラメータ設計］を選択すると図２のような分析実行メニューが表示される。

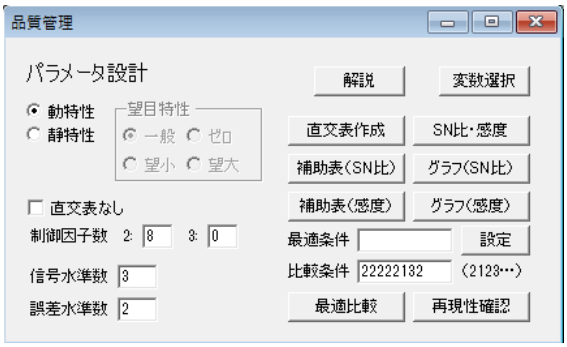


図２　分析実行メニュー

パラメータ設計には動特性と静特性の２種類あるが、最初は動特性について紹介する。まずメニュー中にある、「制御因子数」、「信号水準数」、「誤差水準数」の値を入力する。この例題の場合、デフォルトの数値がそのまま利用できる。次に「変数選択」ボタンですべての変数を選択する。現在の例では直交表が付いているが、単純に SN 比と感度のみを求める場合には、直交表を省略したデータを用いることもできる。その際には「直交表なし」チェックボックスにチェックを入れておく。

「SN 比・感度」ボタンをクリックすると、図３ の計算結果が表示される。

SN比・感度

	A	B	C	D	E	F	G	H	SN比	感度
1	1	1	1	1	1	1	1	1	26.534	-5.730
2	1	1	2	2	2	2	2	2	28.258	-4.239
3	1	1	3	3	3	3	3	3	23.149	-3.516
4	1	2	1	1	2	2	3	3	26.656	-6.431
5	1	2	2	2	3	3	1	1	25.083	-5.230

図３　各実験の SN 比・感度

ここでは各実験に対して、単純に SN 比と感度を求めて表示している。

直交表を使った SN 比の補助表は「補助表（SN 比）」ボタンをクリックすることで図４のように与えられる。感度の補助表については「補助表（感度）」ボタンをクリックして得られる。

補助表（SN比）

	水準1	水準2	水準3	MAX-MIN
A	27.494	23.169		4.325
B	23.656	22.961	29.378	6.418
C	28.332	25.959	21.704	6.628
D	22.358	26.353	27.284	4.926
E	26.649	24.186	25.160	2.463
F	25.040	27.077	23.878	3.199
G	26.675	24.008	25.313	2.667
H	23.489	26.686	25.820	3.197

図４　補助表（SN 比）

ここで制御因子 A は２水準であるから、空白が１つできている。

補助表をグラフにした図は「グラフ（SN 比）」ボタンをクリックして表示される。描画結果を図５に示す。

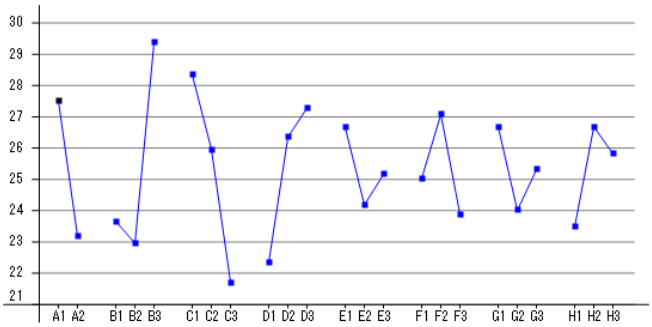


図 5　補助表のグラフ（SN 比）

図 4 と図 5 に対する感度の補助表とグラフは、それぞれ図 6 と図 7 で与えられる。

補助表（感度）

	水準1	水準2	水準3	MAX-MIN
A	-4.515	-3.975		0.540
B	-3.458	-5.041	-4.236	1.583
C	-4.337	-3.729	-4.669	0.940
D	-6.000	-4.117	-2.618	3.381
E	-3.963	-4.762	-4.011	0.799
F	-3.309	-4.239	-5.187	1.878
G	-4.045	-4.400	-4.290	0.355
H	-5.052	-3.968	-3.716	1.336

図 6　補助表（感度）

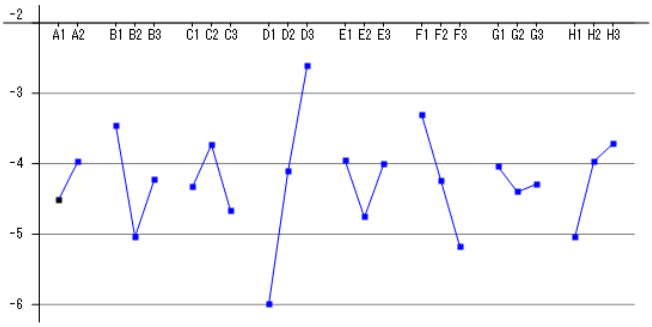


図 7　補助表のグラフ（感度）

利用者はこれらのグラフを見ながら、制御因子の調整を行う。例えば SN 比はあまり動かさず、目的の感度に近づけるような調整には、制御因子 F が適している。

SN 比の補助表やグラフを使った最適条件は「設定」ボタンをクリックすることでメニュー上の最適条件の部分に図 8 のように表示される。

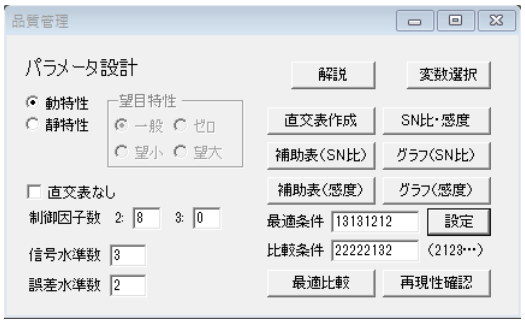


図８　SN 比の最適条件の設定

比較条件で現在の実験から得られるデータの値を求めることができるが、最適条件との比較も可能である。これらの数値は「最適比較」ボタンで得ることができる。表示結果を図 9 に示す。最適条件の制御因子の組み合わせを変えることで結果を手動で訂正することもできる。

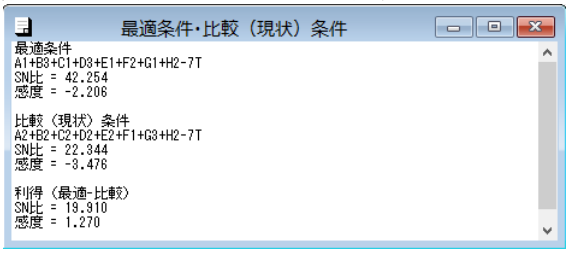


図９　最適条件と比較条件

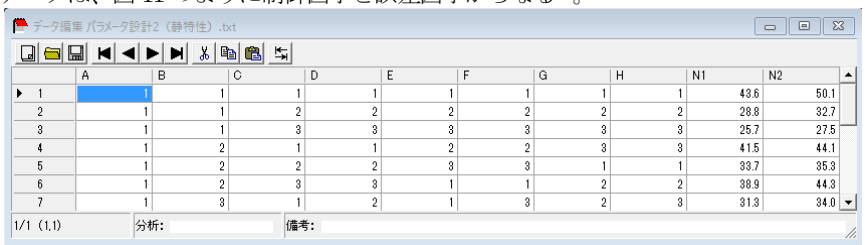
これらの最適条件と比較条件を実験で再現し、結果を得て、それをデータに追加する。当然その部分の直交表は空白になっているが、そのまま「再現性確認」ボタンをクリックすると図 10 に示す再現性確認表が得られる。

再現性確認表

	SN比推定値	感度推定値	SN比実験値	感度実験値
最適条件	42.254	-2.206	44.369	-0.962
比較条件	22.344	-3.476	26.177	-4.059
利得	19.910	1.270	18.192	3.097

図 10　再現性確認表

静特性の場合は、分析実行メニューの「静特性」ラジオボタンに切り替える。一般の望目特性のデータは、図 11 のように制御因子と誤差因子からなる[3]。



データ編集 パラメータ設計2（静特性）.txt

	A	B	C	D	E	F	G	H	N1	N2
1	1	1	1	1	1	1	1	1	43.6	50.1
2	1	1	2	2	2	2	2	2	28.8	32.7
3	1	1	3	3	3	3	3	3	25.7	27.5
4	1	2	1	1	2	2	3	3	41.5	44.1
5	1	2	2	2	3	3	1	1	33.7	35.3
6	1	2	3	3	1	1	2	2	38.9	44.3
7	1	3	1	2	1	3	2	3	31.3	34.0

1/1 (1,1)　分析:　備考:

図 11　静特性のデータ

変数をすべて選択し、実行メニューの「静特性」ラジオボタンを選択し、「望目特性」を「一般」にして、「SN 比・感度」ボタンをクリックすると、図 12 のような一般の望目特性の SN 比と感度の計算結果が得られる。

SN比・感度

	A	B	C	D	E	F	G	H	SN比	感度
1	1	1	1	1	1	1	1	1	20.145	33.393
2	1	1	2	2	2	2	2	2	20.928	29.739
3	1	1	3	3	3	3	3	3	26.398	28.493
4	1	2	1	1	2	2	3	3	27.336	32.625
5	1	2	2	2	3	3	1	1	29.682	30.754
6	1	2	3	3	1	1	2	2	20.726	32.364
7	1	3	1	2	1	3	2	3	24.653	30.270

図 12　静特性（一般の望目特性）の SN 比と感度

また「補助表（SN 比）」ボタンをクリックすると、この計算結果から、図 13 の各水準に対する SN 比の補助表が得られる。

補助表（SN比）

	水準1	水準2	水準3	MAX-MIN
A	24.257	23.949		0.308
B	23.695	24.693	23.921	0.998
C	24.672	24.319	23.318	1.354
D	24.282	24.280	23.748	0.534
E	22.210	24.018	26.082	3.872
F	23.143	23.428	25.738	2.595
G	24.156	23.260	24.894	1.634
H	23.811	23.752	24.746	0.994

図 13　補助表（SN 比）

「グラフ（SN 比）」ボタンをクリックすることで、図 14 のように補助表をグラフに表すことができる。

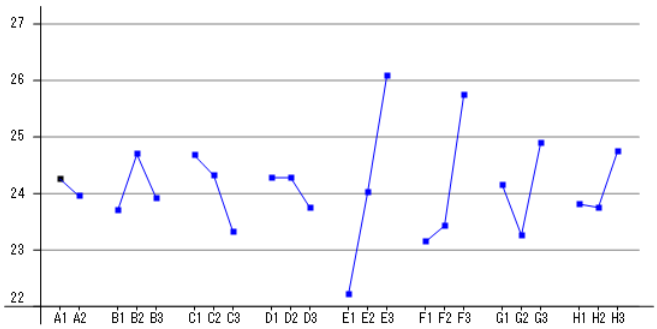


図 14　補助表（SN 比）のグラフ化

同様にして、感度の補助表とそのグラフを求めることもできる。これらを基に SN 比に対する最適条件を求めたり、最適条件と比較条件の SN 比や感度を比較したりすることは動特性のときと同様であるので省略する。

3.　オンライン品質工学

3.1　オンライン品質工学の理論

工場の生産ラインでは、目標特性の製品を作るために、品質特性値を計測し、ラインを調整する。これらには時間と人手を要し、そこには費用が発生する。オンライン品質工学では、品質を金額ベースで表し、品質特性値の最適な計測間隔と生産ラインの調整間隔を求め、現行の値と比較し、どの程度の金額の削減につながるかを検討する。

3.1.1　目的特性の規格値が$[-\Delta, \Delta]$の場合

計算には以下の量を利用するが、規格値の中央値は 0 に設定している。0 以外の値の場合は、規格値からのずれとして考えればよい。計算に当たり、我々は以下の値を利用する。

不良品損失：A，計測コスト：B，調整コスト：C

計測間隔：n 個，調整間隔：u 個，調整限界：D，計測タイムラグ：l

製品 1 個当たりの計測コストはB/n、調整コストはC/uである。また、製品 1 個当たりの品質損失（品質水準）Qを以下のように与える[2]。

$$Q = \frac{A}{\Delta^2}\left[\frac{D^2}{3} + \left(\frac{n+1}{2} + l\right)\frac{D^2}{u}\right]$$

後に述べるが、最初の項は調整限界内のばらつき、次の項は問題があった計測時の調整限界を超えた製品の予測個数と計測タイムラグによる計測が遅れた製品の個数である。この品質損失の角括弧の中は誤差分散であり、その平方根を誤差標準偏差または RMS と呼ぶ。

ここで与えた計測調整コストと品質損失を足して、製品 1 個当たりの総損失Lは以下となる。

$$L = \frac{B}{n} + \frac{C}{u} + \frac{A}{\Delta^2}\left[\frac{D^2}{3} + \left(\frac{n+1}{2} + l\right)\frac{D^2}{u}\right] \tag{1}$$

後の考察のために、品質損失（品質水準）Qを、損失関数という考え方に基づいて、少し補足的説明も加えて考えてみる。まず損失関数を以下と仮定する。

$$Q(x) = \frac{A}{\Delta^2}x^2$$

ここに係数は$Q(\Delta) = A$となるように決めている。

今、損失を調整限界内損失（限界内損失と略す）Q_{in}と調整限界外損失（限界外損失と略す）Q_{out}に分けて考える。

限界内損失のデータの確率分布を、範囲$-D \le x \le D$の一様分布と考えると、確率密度関数は$1/2D$となる。そのため損失関数の期待値は以下となる。

$$Q_{in} = \frac{1}{2D}\int_{-D}^{D}\frac{A}{\Delta^2}x^2 dx = \frac{A}{6D\Delta^2}\left[x^3\right]_{-D}^{D} = \frac{AD^2}{3\Delta^2}$$

次に限界外損失について考えるが、これは一度管理外データが出るとその後も出続けると考える。n 個に 1 回の計測で管理外データが発見された場合、前の $n-1$ 個については計測されていないのでどの時点で最初のエラーが発生したか分からない。そのため確率を $1/2$ として、エラーが発生してからの個数の期待値をとると、$(n-1)/2$ となる。これと管理外データが発見された 1 個を足して、期待値は $(n+1)/2$ となる。また、これにタイムラグ l による調整の遅れで管理外データが発生すると考えると、これらを合わせて、1 回の調整での管理外データの個数の期待値 n_{out} は以下となる。

$$n_{out} = \frac{n+1}{2} + l$$

これが調整間隔 u 個に 1 回起こるので、商品 1 個当たりの管理外データの発生確率 p_{out} は以下となる。

$$p_{out} = \frac{n_{out}}{u} = \left(\frac{n+1}{2} + l\right)\frac{1}{u}$$

管理外データの損失関数 $Q(x)$ の値は $x = D$ として、1 回につき AD^2/Δ^2 で与えられると考えると、商品 1 個当たりの限界外損失 Q_{out} は以下で与えられる。

$$Q_{out} = \frac{A}{\Delta^2}\left(\frac{n+1}{2} + l\right)\frac{D^2}{u}$$

限界内損失と限界外損失を合わせて品質損失は以下となる。

$$Q = Q_{in} + Q_{out} = \frac{A}{\Delta^2}\left[\frac{D^2}{3} + \left(\frac{n+1}{2} + l\right)\frac{D^2}{u}\right]$$

これは、分散という考えから求めた結果と同じである。これに計測費用と調整費用を加えて総費用は以下となる。

$$L = \frac{B}{n} + \frac{C}{u} + \frac{A}{\Delta^2}\left[\frac{D^2}{3} + \left(\frac{n+1}{2} + l\right)\frac{D^2}{u}\right]$$

現行計測間隔を n_0，現行調整限界を D_0，現行調整間隔を u_0 とすると、現行総損失 L_0 は以下のようになる。

$$L_0 = \frac{B}{n_0} + \frac{C}{u_0} + \frac{A}{\Delta^2}\left[\frac{D_0^2}{3} + \left(\frac{n_0+1}{2} + l\right)\frac{D_0^2}{u_0}\right]$$

次に、(1)式に基づく最適な総損失を求めてみよう。そのために、我々は $u = \lambda D^2$ という仮定を考える。パラメータ λ は、現行の値を使って $\lambda = u_0/D_0^2$ で与える。この関係は調整限界

間隔を小さく設定すると、平均調整間隔も小さくする必要があるということに基づく。しかし、これには計算上の理由もある。(1)式にこの関係を代入すると以下となり、

$$L = \frac{B}{n} + \frac{C}{\lambda D^2} + \frac{A}{\Delta^2}\left[\frac{D^2}{3} + \left(\frac{n+1}{2} + l\right)\frac{1}{\lambda}\right] \qquad (2)$$

n と D とは和の項として分離され、極小化条件を求めるために微分した後、単独に解を求めることが可能となる。他の関係では、解を求めるために、数値計算の必要が生じ、解析的には困難となる。

我々は(2)式を n と D とで微分して 0 とおき、最適計測間隔 $\hat{n}$ と最適調整限界 $\hat{D}$ を求める。これらは以下のように与えられる。

$$\hat{n} = \sqrt{\frac{2B\lambda\Delta^2}{A}} \quad , \quad \hat{D} = \left(\frac{3C\Delta^2}{A\lambda}\right)^{1/4}$$

これより、最適総損失 $\hat{L}$ は以下のようになる。

$$\hat{L} = \frac{B}{\hat{n}} + \frac{C}{\lambda \hat{D}^2} + \frac{A}{\Delta^2}\left[\frac{\hat{D}^2}{3} + \left(\frac{\hat{n}+1}{2} + l\right)\frac{1}{\lambda}\right]$$

ここで、ハットの付いた量を含む式は $u = \lambda D^2$ の仮定を入れていると考えてもらいたい。

次に $u = \lambda D^2$ の条件を外してその他の指標をみてみよう。誤差分散 σ^2 は(1)式の[ ]の中で、以下のように与えられる。

$$\sigma^2 = \frac{D^2}{3} + \left(\frac{n+1}{2} + l\right)\frac{D^2}{u}$$

前にも述べたが、誤差の標準偏差は RMS と呼ばれ、以下で与えられる。

$$RMS = \sigma = \sqrt{\frac{D^2}{3} + \left(\frac{n+1}{2} + l\right)\frac{D^2}{u}}$$

また、規格値の範囲 $\pm\Delta$ とこの標準偏差の $\pm 3\sigma$ の範囲との比を工程能力指数 C_p と呼び、精度評価の１つの基準としている。

$$C_p = \frac{\Delta}{3\sqrt{\frac{D^2}{3} + \left(\frac{n+1}{2} + l\right)\frac{D^2}{u}}}$$

バッチ処理の場合には、A, n, u 等はバッチ単位の値として総損失等は計算される。計測方法としてバッチ内のいくつかの製品についてサンプリング検査されるので、バッチ処理の場合

には計測のバッチ内分散 s_m^2 も考慮しなくてはならない。そのため、バッチ内分散の値を数値として与え、品質損失（品質水準） Q を以下のように考える。

$$Q = \frac{A}{\Delta^2}\left[\frac{D^2}{3} + \left(\frac{n+1}{2} + l\right)\frac{D^2}{u} + s_m^2\right]$$

これに関連して、総損失 L や RMS、工程能力指数 C_p も変更を受ける。プログラムではバッチ内分散の値が全体の分散の値に占める比率をバッチ内分散比率 r_B として以下のように与えている。

$$r_B = \frac{s_m^2}{\frac{D^2}{3} + \left(\frac{n+1}{2} + l\right)\frac{D^2}{u} + s_m^2}$$

3.1.2　目的特性の規格値が $[0, \Delta]$ の場合

目的特性の規格の下限値は 0 に設定しているが、特にどんな値でもずらして考えればよい。ここでは、区間が正に限られた場合（製品によって負になる場合も含む）の損失関数を以下と考える。

$$Q(x) = \frac{A}{\Delta^2}x^2$$

限界内損失のデータの確率分布を、範囲 $m - D \le x \le m + D$ の一様分布と考えると、確率密度関数は $1/2D$ となる。そのため限界内損失関数の期待値は以下となる。

$$Q_{in} = \frac{1}{2D}\int_{m-D}^{m+D}\frac{A}{\Delta^2}x^2 dx = \frac{A}{6D\Delta^2}\left[x^3\right]_{m-D}^{m+D} = \frac{A}{3\Delta^2}(D^2 + 3m^2)$$

定数 Δ の考え方が異なるので注意を要するが、形式的には以前の結果に比べて $3m^2$ だけ追加されている。

限界外損失の場合の損失関数の値は以下となる。

内側（0 に近い側）　$Q(m-D) = A(m-D)^2/\Delta^2$

外側（0 から遠い側）　$Q(m+D) = A(m+D)^2/\Delta^2$

平均的に内側と外側で、確率 $1/2$ で管理外データが発生すると考えると、前節の p_{out} を用いて限界外損失 Q_{out} は以下のようになる。

$$Q_{out} = \frac{1}{2}\left[Q(m-D) + Q(m+D)\right]p_{out} = \frac{A}{2\Delta^2}\left[(m+D)^2 + (m-D)^2\right]p_{out}$$

$$= \frac{A}{\Delta^2}\left(\frac{n+1}{2} + l\right)\frac{D^2 + m^2}{u}$$

ここで値が 0 未満になることも考えられるが、これは廃棄になることはないと考えて、損失関数がそのまま適用できると考える。廃棄になる場合は、それを見越して精度や中心の位置を調整する必要がある。

これより、品質損失は以下となる。

$$Q = Q_{in} + Q_{out} = \frac{A}{\Delta^2}\left[\frac{D^2+3m^2}{3} + \left(\frac{n+1}{2}+l\right)\frac{D^2+m^2}{u}\right]$$

これに計測費用と調整費用を加えて総費用は以下となる。

$$L = \frac{B}{n} + \frac{C}{u} + \frac{A}{\Delta^2}\left[\frac{D^2+3m^2}{3} + \left(\frac{n+1}{2}+l\right)\frac{D^2+m^2}{u}\right] \tag{3}$$

m が定数の場合（例えば Δ に比例するとしても）、前節で述べた、n と D との和の項としての独立性は得られず、(3)式の最小化のためには連立方程式の数値計算が必要になる。しかし、$m = \alpha D$ のように決めることができると、

$$L = \frac{B}{n} + \frac{C}{u} + \frac{A}{\Delta^2}\left[\frac{(1+3\alpha^2)D^2}{3} + \left(\frac{n+1}{2}+l\right)\frac{(1+\alpha^2)D^2}{u}\right]$$

のようになり、前と同じ方法で計算が容易になる。

今、$u = \lambda D^2$ を仮定すると

$$L = \frac{B}{n} + \frac{C}{\lambda D^2} + \frac{A}{\Delta^2}\left[\frac{(1+3\alpha^2)D^2}{3} + \left(\frac{n+1}{2}+l\right)\frac{1+\alpha^2}{\lambda}\right] \tag{4}$$

我々は(4)式を n と D とで微分して 0 とおき、最適計測間隔 $\hat{n}$ と最適調整限界 $\hat{D}$ を求めると、以下のように与えられる。

$$\hat{n} = \sqrt{\frac{2B\lambda\Delta^2}{A(1+\alpha^2)}} \quad , \quad \hat{D} = \left(\frac{3C\Delta^2}{A\lambda(1+3\alpha^2)}\right)^{1/4}$$

誤差分散 σ^2 については、損失関数に関係なく、(1)式の[ ]の中で、以下のように与えられる。

$$\sigma^2 = \frac{D^2}{3} + \left(\frac{n+1}{2}+l\right)\frac{D^2}{u}$$

誤差の標準偏差 RMS も同様である。

$$RMS = \sigma = \sqrt{\frac{D^2}{3} + \left(\frac{n+1}{2}+l\right)\frac{D^2}{u}}$$

工程能力指数 C_p については目的特性の範囲が異なるため、以下とする。

$$C_p = \frac{\Delta}{6\sqrt{\frac{D^2}{3} + \left(\frac{n+1}{2} + l\right)\frac{D^2}{u}}}$$

バッチ処理の場合には、前節と同様に計測のバッチ内分散 s_m^2 も考慮して、品質損失（品質水準）Q を以下のように考える。

$$Q = \frac{A}{\Delta^2}\left[\frac{D^2 + 3m^2}{3} + \left(\frac{n+1}{2} + l\right)\frac{D^2 + m^2}{u} + s_m^2\right]$$

バッチ内分散の値は定数として与えるので、$u = \lambda D^2$ と $m = \alpha D$ の仮定の元に与えられた最適計測間隔 $\hat{n}$ と最適調整限界 $\hat{D}$ には変更がない。

3.2　プログラムの利用法

メニュー［分析－ＯＲ－品質管理－オンライン品質工学］を選択すると図１に示す分析実行メニューが表示される。また、このまま「実行」ボタンをクリックすると、画面上で与えられた数値を使って、図２に示す結果が表示される。

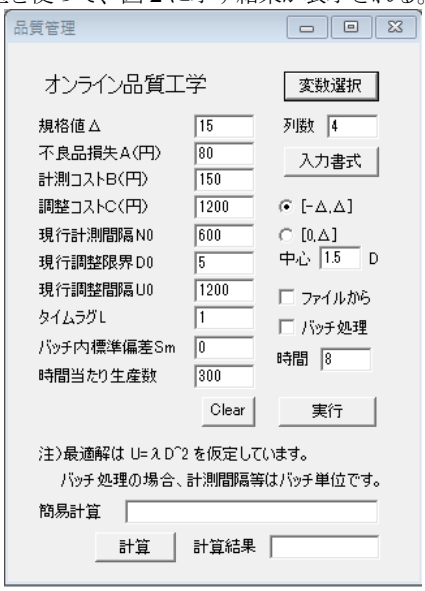


図１　分析実行メニュー

計算結果

	現行	最適
計測間隔N	600.000	201.246
調整限界D	5.000	3.811
調整間隔U	1200.000	697.137
パラメータλ	48.000	48.000
計測費/個	0.250	0.745
調整費/個	1.000	1.721
品質損失/個	5.196	2.478
総損失/個	6.446	4.944
工程能力指数Cp	1.308	1.894
RMS	3.823	2.640
バッチ内分散比率	0.000	0.000
累積計測回数	4.000	11.926
累積調整回数	2.000	3.443
累積計測費	600	1789
累積調整費	2400	4131
累積品質損失	12471	5947
累積総損失	15471	11867

図２　画面からの実行結果

累積となっている部分は、メニューの「時間当たり生産数」と「時間」の値を使って全時間で発生する回数と金額を計算している。例えば１日８時間稼働とすると「時間」を８にして、「時間当たり生産数」を１時間当たりの生産数にする。単位時間を１日にして、年間の稼働日数を「時間」として設定してもよい。

現行のデータを複数与えて、最適解を求めるときには、「ファイルから」チェックボックスをチェックして、変数選択して「実行」ボタンを押す。しかし、このプログラムは最適な結果を出すだけでなく、徐々に最適な結果に近づけて行くときにも利用できる。この方法を参考文献[2]に従って考えて行く。

データファイル[3]を開くと、図３のように現行のデータが入力されている。

データ編集 オンライン品質工学1.txt

	製品1
規格値Δ	15
不良品損失A(円)	80
計測コストB(円)	150
調整コストC(円)	1200
現行計測間隔N0(個)	600
現行調整限界D0(個)	5
現行調整間隔U0(個)	1200
タイムラグL(個)	1
バッチ内標準偏差Sm	0
時間当たり生産数	300

1/3 (1,1)　分析:　備考:

図３　変更用ファイルデータ

データは１品目についてであるが、このデータを変更しながら最適な結果に近づけて行く。

まず、最適計測間隔と最適調整限界を求めて、現行よりいくらの改善になるかを推定する。この推定の最後には簡単な金額計算をする必要があるので、メニュー画面の下に簡単な電卓機能を付けておいた。

図３のデータで変数選択を「All」にして、「ファイルから」チェックボックスにチェックを入れ、「実行」ボタンをクリックすると、図４のような結果が出る。

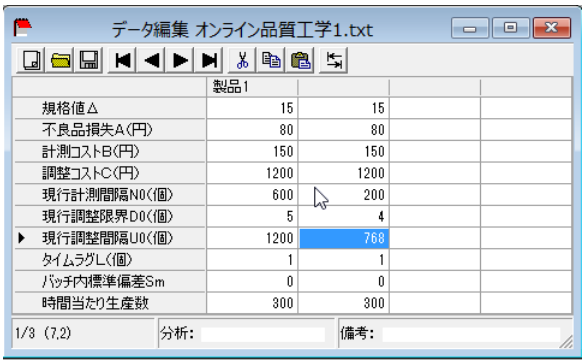
計算結果

	現行製品1	最適製品1	現行var2	最適var2	現行var3	最適var3	現行var4	最適var4
計測間隔N	600.000	201.246						
調整限界D	5.000	3.811						
調整間隔U	1200.000	697.137						
パラメータλ	48.000	48.000						
計測費/個	0.250	0.745						
調整費/個	1.000	1.721						
品質損失/個	5.196	2.478						
総損失/個	6.446	4.944						
工程能力指数Cp	1.308	1.894						
RMS	3.823	2.640						
バッチ内分散比率	0.000	0.000						
累積計測回数	4.000	11.926						
累積調整回数	2.000	3.443						
累積計測費	600	1789						
累積調整費	2400	4131						
累積品質損失	12471	5947						
累積総損失	15471	11867						

図４ファイルからの実行結果

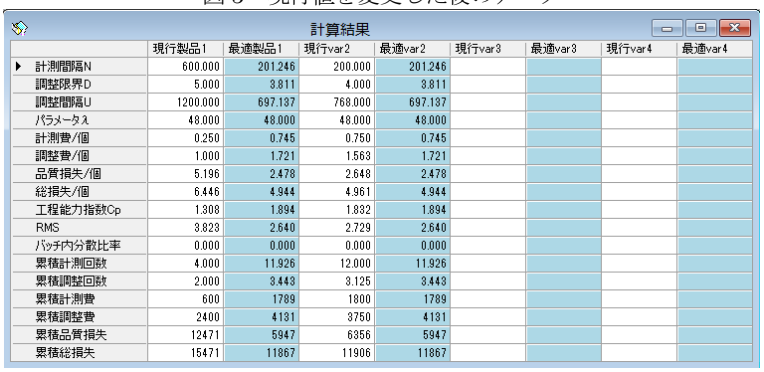
変数が現行と最適合わせて４つ表示されるのは図３ですべての変数(列)を選んでいるためで、変数を１つずつ追加で選択して行く必要がないように、データの「規格Δ」の値が空欄のときは、関連する２列すべてを空欄になるように設定しているためである。

最適な結果が出ているので、その値を参考にして、現行の値を変更することができる。グリッドエディタ上で、現行の値を次の列にコピーして必要な部分を変更した結果を図 5 に示す。また、図 5 のデータを元に計算した結果が図 6 である



	製品1			
規格値Δ	15	15		
不良品損失A(円)	80	80		
計測コストB(円)	150	150		
調整コストC(円)	1200	1200		
現行計測間隔N0(個)	600	200		
現行調整限界D0(個)	5	4		
現行調整間隔U0(個)	1200	768		
タイムラグL(個)	1	1		
バッチ内標準偏差Sm	0	0		
時間当たり生産数	300	300		

図 5　現行値を変更した後のデータ



	現行製品1	最適製品1	現行var2	最適var2	現行var3	最適var3	現行var4	最適var4
計測間隔N	600.000	201.246	200.000	201.246				
調整限界D	5.000	3.811	4.000	3.811				
調整間隔U	1200.000	697.137	768.000	697.137				
パラメータえ	48.000	48.000	48.000	48.000				
計測費/個	0.250	0.745	0.750	0.745				
調整費/個	1.000	1.721	1.563	1.721				
品質損失/個	5.196	2.478	2.648	2.478				
総損失/個	6.446	4.944	4.961	4.944				
工程能力指数Cp	1.308	1.894	1.832	1.894				
RMS	3.823	2.640	2.729	2.640				
バッチ内分散比率	0.000	0.000	0.000	0.000				
累積計測回数	4.000	11.926	12.000	11.926				
累積調整回数	2.000	3.443	3.125	3.443				
累積計測費	600	1789	1800	1789				
累積調整費	2400	4131	3750	4131				
累積品質損失	12471	5947	6356	5947				
累積総損失	15471	11867	11906	11867				

図 6　現行値を変更した後の結果

ここでは現行の総損失の差を見てみる。1 個当たりの改善は 6.446 −4.961=1.485 円、1 時間 300 個、1 年 2000 時間として 89.1 万円となる。これはメニューの「時間」のところを「2000」に変えて出力すると、そのまま累積総損失に全時間の合計が計算されるので、差を見つけ易い。

工程能力指標 C_p を見てみると現行 1.308、改善 1.832 となる。

最後に、計測に必要なマンパワーを求めてみよう。例えば計測に 3 分、調整に平均 15 分かかるとする。計測回数は 1 日 8 時間として 1 日当たり 12 回、時間は 12×3=36 分である。調整回数は 1 日当たり 3.443 回、時間は 3.443×15=約 47.9 分である。両方合わせて、約 82.9 分となる。そのため、1 日当たり 82.9 分/480 分=0.173 人必要である。

規格区間が $[0,\Delta]$ となる場合については、図 1 の「$[0,\Delta]$」ラジオボタンを選択し、その下の「中心」の位置を調整限界値の何倍に設定するかを決めて同じ処理を行うが、操作は全く同じなので、説明は省略する。

4.　おわりに

我々は今回、経営科学の中の品質管理の手法に、品質改善のためのフィードバックをなくすことを目的とするパラメータ設計のプログラムと、製品の品質を金額ベースでとらえ、最適の計測と調整を補助するオンライン品質工学のプログラムを加えた。これらは企業からの要請によるものであるが、仕様がまだ不十分な気がしている。

例えば、パラメータ設計では、データの測定結果の変数名に信号因子の値そのものを使っている。これはこれまでの分析ではなかったことで、整合性に反する部分である。また、メニューの中で制御因子の最適条件と比較条件のところは、制御因子の番号を連続して記述するようにしているが、どの制御因子か分かりにくいし、欄の長さが適切かどうか不明である。これらは、実際に使ってみて不便さが見えてくる部分である。

また、オンライン品質工学では、損失の計算が一意的であり、柔軟性に欠ける。この方式に合わない製造過程に対してどのように対処すべきか、対応に苦慮する。もう少し汎用的に損失関数を変更できないものかと思う。

このように、これらのプログラムは実際に使ってみて、その後改良点が見つかるものである。また、業種により個別に対応しなければならない場合もあると思われ、今後の改良は必須であろう。ただ、教育的に考えた場合、品質管理の基本的考え方を学ぶツールとしては、役に立つものと考える。

参考文献

[1] 井上清和・中野惠司他, 入門パラメータ設計, 日科技連出版社, 2008.

[2] 矢野宏, 品質工学概論, 日本規格協会, 2009.

[3] ホームページ http://www.heisei-u.ac.jp/ba/fukui/analysis.html 内のサンプルデータ Samples.zip 内のファイル

Multi-purpose Program for Social System Analysis 29
- Parameter Design, On Line Quality Engineering -

Masayasu FUKUI[*1], Tomoyuki OYAMA[*2] and Nozomu ODA[*2]

[*1] *Department of Business Administration, Faculty of Business Administration, Fukuyama Heisei University*

[*2] *Nippon Pneumatic Manufacturing Co., Ltd.*

Abstract: We have been constructing a unified program on the social system analysis for the purpose of education. This time, we make programs of parameter design (offline quality engineering) and online quality engineering. The parameter design is a technique to increase product quality at the design stage, and is a method which finds the best control factor based on the SN ratio. The online quality engineering is a method which expresses quality on a monetary basis and manages anomaly detection and equipment adjustment at the factory line.

Keywords: College Analysis, management science, parameter design, offline quality engineering, online quality engineering

URL: http://www.heisei-u.ac.jp/ba/fukui/

福山平成大学経営学部紀要
第 13 号（2017），121－138 頁

社会システム分析のための統合化プログラム３０
－異常検知－

福井 正康[*1]、大山 知之[*2]、織田 望[*2]
[*1] 福山平成大学経営学部経営学科
[*2] 日本ニューマチック工業株式会社

要旨：我々は教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた。今回は品質管理で実践される異常検知についてのプログラムを紹介する。ここでは複数変数の正規性を持つデータと正規性を持たないデータ、時系列データ、入出力がある場合のデータの異常検知についてその手法を示し、プログラムの利用法について紹介する。

キーワード：College Analysis、経営科学、品質管理、異常検知、機械学習、EM アルゴリズム
URL：http://www.heisei-u.ac.jp/ba/fukui/

1. はじめに

製造現場における検査過程では多くのデータが測定されるが、正常なデータと異常なデータの迅速な選別は品質管理の上で非常に重要である。ここではその主要な、複数変数の異常検知、時系列データの異常検知、入力と出力がある場合の異常検知について、その方法を説明し、分析を実行するプログラムを紹介する。このプログラムは参考文献 [1] の手法に基づいている。

複数変数の異常検知では、データが多変量正規分布に従うと仮定される場合とそうでない場合を扱う。データが多変量正規分布に従う場合、マハラノビス距離の 2 乗を元にしたホテリングの t^2 統計量に基づく判定法を用いる。また、多変量正規分布に従わない場合は、混合正規分布モデルを仮定する方法を用いている。ここでは EM アルゴリズムを用いて混合正規分布のパラメータを推定している。また、１次元データについては、ガンマ分布による異常検知の方法も加えている。

時系列データの異常検知では、周波数変化や波形変化に対応した検知法に特異スペクトル変換法を用いている。

入力と出力がある場合のデータでは、重回帰分析を用いている。但し、問題は説明変数の多重共線性である。これに対してプログラムではリッジ回帰分析と PLS 回帰分析を加えてある。これらを利用することにより、多重共線性の問題は回避される。

2. 異常検知の理論

2.1 複数変数の異常検知

1) 多変量正規分布に基づく異常検知

一般に p 変量正規分布の密度関数は以下で与えられる。

$$f(\mathbf{x}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^p\,|\,\boldsymbol{\Sigma}\,|}}\exp\left[-\frac{1}{2}\,{}^t(\mathbf{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]$$

データ $D=\left\{\mathbf{x}_1,\cdots,\mathbf{x}_N\right\}$ が与えられた場合の対数尤度関数は以下で与えられる。

$$L(\mu,\Sigma\,|\,D) = -\frac{pN}{2}\log(2\pi) - \frac{N}{2}\log|\,\boldsymbol{\Sigma}\,| - \frac{1}{2}\sum_{\lambda=1}^{N}{}^t(\mathbf{x}_\lambda-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x}_\lambda-\boldsymbol{\mu}) \quad (1)$$

我々は最尤法を用いて(1) 式を最大化するが、その解は以下となる。

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{\lambda=1}^{N}\mathbf{x}_\lambda\,,\quad \hat{\boldsymbol{\Sigma}} = \frac{1}{N}\sum_{\lambda=1}^{N}(\mathbf{x}_\lambda-\hat{\boldsymbol{\mu}})^t(\mathbf{x}_\lambda-\hat{\boldsymbol{\mu}})$$

ここで、同じ正規分布の確率変数 $\mathbf{x}'$ に対する異常度 $a(\mathbf{x}')$ を $-2\log f(\mathbf{x}'\,|\,\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}})$ を元に以下のように定義する。

$$a(\mathbf{x}') = {}^t(\mathbf{x}'-\hat{\boldsymbol{\mu}})\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}'-\hat{\boldsymbol{\mu}})$$

ここで上式と $-2\log f(\mathbf{x}'\,|\,\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}})$ の差は定数であるので、評価関数として本質的な差はない。また上式は１次元変数の場合の変数の標準化の一般形である。

異常度の式については、以下のように定数を掛けると、分布が自由度 $p, N-p$ の F 分布に従うことが知られている。

$$T^2 \equiv \frac{N-p}{(N+1)p}\,{}^t(\mathbf{x}'-\hat{\boldsymbol{\mu}})\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{x}'-\hat{\boldsymbol{\mu}}) \sim F_{p,N-p}$$

この T^2 をホテリング統計量という。

異常検知には、この統計量を使って確率の値を指定するか、直接 T^2 値を指定して閾値とする。

2) 混合多変量正規分布に基づく異常検知

p 変数、n 種混合多変量正規分布の密度関数は、種類 α の確率密度関数

$$f_\alpha(\mathbf{x}\,|\,\boldsymbol{\mu}_\alpha,\boldsymbol{\Sigma}_\alpha) = \frac{1}{\sqrt{(2\pi)^p\,|\,\boldsymbol{\Sigma}_\alpha\,|}}\exp\left[-\frac{1}{2}\,{}^t(\mathbf{x}-\boldsymbol{\mu}_\alpha)\boldsymbol{\Sigma}_\alpha^{-1}(\mathbf{x}-\boldsymbol{\mu}_\alpha)\right]$$

（$\alpha=1,\cdots,n$）

を利用して以下で与えられる。

$$f(\mathbf{x}\,|\,\boldsymbol{\mu}_1,\cdots,\boldsymbol{\mu}_\alpha,\boldsymbol{\Sigma}_1,\cdots,\boldsymbol{\Sigma}_\alpha)=\sum_{\alpha=1}^{n}\pi_\alpha f_\alpha(\mathbf{x}\,|\,\boldsymbol{\mu}_\alpha,\boldsymbol{\Sigma}_\alpha)$$
$$=\sum_{\alpha=1}^{n}\frac{\pi_\alpha}{\sqrt{(2\pi)^p\,|\,\boldsymbol{\Sigma}_\alpha\,|}}\exp\left[-\frac{1}{2}\,{}^t(\mathbf{x}-\boldsymbol{\mu}_\alpha)\boldsymbol{\Sigma}_\alpha^{-1}(\mathbf{x}-\boldsymbol{\mu}_\alpha)\right]$$

ここに、π_α は群α の生起確率である。

この密度関数に従うデータ $D=\{\mathbf{x}_1,\cdots,\mathbf{x}_N\}$ による対数尤度は以下である。

$$L(\pi_1,\cdots,\pi_n,\boldsymbol{\mu}_1,\cdots,\boldsymbol{\mu}_n,\boldsymbol{\Sigma}_1,\cdots,\boldsymbol{\Sigma}_n\,|\,D)=\sum_{\lambda=1}^{N}\log\left[\sum_{\alpha=1}^{n}\pi_\alpha f_\alpha(\boldsymbol{\mu}_\alpha,\boldsymbol{\Sigma}_\alpha\,|\,\mathbf{x}_\lambda)\right]$$
$$=\sum_{\lambda=1}^{N}\log\left[\sum_{\alpha=1}^{n}\frac{\pi_\alpha}{\sqrt{(2\pi)^p\,|\,\boldsymbol{\Sigma}_\alpha\,|}}\exp\left\{-\frac{1}{2}\,{}^t(\mathbf{x}_\lambda-\boldsymbol{\mu}_\alpha)\boldsymbol{\Sigma}_\alpha^{-1}(\mathbf{x}_\lambda-\boldsymbol{\mu}_\alpha)\right\}\right] \tag{2}$$

最尤法を用いてこの対数尤度の最大値を求めるが、その際以下のアルゴリズムを利用する。

①パラメータ $\pi_\alpha,\boldsymbol{\mu}_\alpha,\boldsymbol{\Sigma}_\alpha$ に初期値 $\hat{\pi}_\alpha,\hat{\boldsymbol{\mu}}_\alpha,\hat{\boldsymbol{\Sigma}}_\alpha$ を与える。

② $\hat{\pi}_\alpha,\hat{\boldsymbol{\mu}}_\alpha,\hat{\boldsymbol{\Sigma}}_\alpha$ の値を用いて、各データの群α への帰属度 $q_\alpha(\mathbf{x}_\lambda)$ を以下で求める。

$$q_\alpha(\mathbf{x}_\lambda)=\frac{\hat{\pi}_\alpha f_\alpha(\mathbf{x}_\lambda\,|\,\hat{\boldsymbol{\mu}}_\alpha,\hat{\boldsymbol{\Sigma}}_\alpha)}{\sum_{\beta=1}^{n}\hat{\pi}_\beta f_\beta(\mathbf{x}_\lambda\,|\,\hat{\boldsymbol{\mu}}_\beta,\hat{\boldsymbol{\Sigma}}_\beta)}$$

③この帰属度を使い、新しいパラメータを以下のように決定する。

$$\hat{\pi}_\alpha=\frac{1}{N}\sum_{\lambda=1}^{N}q_\alpha(\mathbf{x}_\lambda),\qquad \hat{\boldsymbol{\mu}}_\alpha=\sum_{\lambda=1}^{N}q_\alpha(\mathbf{x}_\lambda)\mathbf{x}_\lambda\Big/\sum_{\lambda=1}^{N}q_\alpha(\mathbf{x}_\lambda),$$
$$\hat{\boldsymbol{\Sigma}}_\alpha=\sum_{\lambda=1}^{N}q_\alpha(\mathbf{x}_\lambda)(\mathbf{x}_\lambda-\hat{\boldsymbol{\mu}}_\alpha)^t(\mathbf{x}_\lambda-\hat{\boldsymbol{\mu}}_\alpha)\Big/\sum_{\lambda=1}^{N}q_\alpha(\mathbf{x}_\lambda)$$

④新しいパラメータと元のパラメータを比較し、十分近ければ（プログラムではすべての成分が 0.001 未満）終了し、そうでなければ②へ戻る。

この方法によって求めたパラメータを使って、異常判定には以下の指標を用いる。

$$a(\mathbf{x}')=-\log\sum_{\alpha=1}^{n}\hat{\pi}_\alpha f_\alpha(\mathbf{x}'\,|\,\hat{\boldsymbol{\mu}}_\alpha,\hat{\boldsymbol{\Sigma}}_\alpha)$$

判定基準は各個体についてこの指標を小さい方から順番に並べ、分位点を閾値として決めるか、直接指標の閾値を指定する。

このモデルの適合度は赤池情報量基準 AIC、ベイズ情報量基準 BIC などを使って求める。今このモデルのパラメータ数を M_n とすると、AIC と BIC はそれぞれ以下のように表現される。

$$AIC = -2L(\Theta \mid D) + 2M_n$$

$$BIC = -2L(\Theta \mid D) + M_n \log N, \quad M_n = \frac{p}{2}(n+1)(n+2)$$

ここに、$L(\Theta \mid D)$ はパラメータを Θ で代表させて書いた対数尤度である。具体的には(2)式に、求めたパラメータの値を代入したものである。適合度を求めるために、交差検証を使う方法も考えられるが、プログラムでは使用していない。

2.2 時系列データの異常検知

時系列データの異常検知で、周波数変化や波形変化に対応した検知法に特異スペクトル変換法がある。これは、現時刻 t に対して、時系列データ x_{t-w} から x_{t-1} までの w 個のデータを抽出し、それを１つのベクトル ${}^t\mathbf{x}_{t-w} = (x_{t-w}\ x_{t-w+1} \cdots x_{t-1})$ とする。そのベクトルの開始時点を１つずつずらして k （$k < w$）本並べ、以下のような行列を作る。

$$\mathbf{X}_1^{(t)} = (\mathbf{x}_{t-k-w+1}\ \mathbf{x}_{t-k-w+2} \cdots \mathbf{x}_{t-w})$$

これに対して時間 L だけ経過した時点から作ったデータを以下のように $\mathbf{X}_2^{(t)}$ とする。

$$\mathbf{X}_2^{(t)} = (\mathbf{x}_{t-k-w+1+L}\ \mathbf{x}_{t-k-w+2+L} \cdots \mathbf{x}_{t-w+L})$$

この二つのデータの間で違いを見ることになる。これらのデータ数やずれ等は、状況によって適当な値に設定する。

今、${}^t\mathbf{X}_1^{(t)}$，${}^t\mathbf{X}_2^{(t)}$ の各列ベクトルに係数をかけて足し合わせ、特徴的な量を求めるために、その大きさを以下のように最大化する。

$$\left\| {}^t\mathbf{X}_1^{(t)}\mathbf{u}^{(t)} \right\|^2 \rightarrow \text{最大化}\quad \text{但し、} {}^t\mathbf{u}^{(t)}\mathbf{u}^{(t)} = 1\ ,\ \mathbf{u}^{(t)} \text{ は } (w \times 1) \text{ ベクトル}$$

$$\left\| {}^t\mathbf{X}_2^{(t)}\mathbf{v}^{(t)} \right\|^2 \rightarrow \text{最大化}\quad \text{但し、} {}^t\mathbf{v}^{(t)}\mathbf{v}^{(t)} = 1\ ,\ \mathbf{v}^{(t)} \text{ は } (w \times 1) \text{ ベクトル}$$

このベクトルは、以下の固有方程式から求められ、

$$\mathbf{X}_1^{(t)}\,{}^t\mathbf{X}_1^{(t)}\mathbf{u}^{(t)} = \lambda\mathbf{u}^{(t)}$$

$$\mathbf{X}_2^{(t)}\,{}^t\mathbf{X}_2^{(t)}\mathbf{v}^{(t)} = \mu\mathbf{v}^{(t)}$$

固有値の大きい順に固有ベクトルを m 個並べて表した $(w \times m)$ 行列を以下のように定義する。

$$\mathbf{U}_m^{(t)}(w \times m) = \left(\mathbf{u}_1^{(t)}\ \mathbf{u}_2^{(t)} \cdots \mathbf{u}_m^{(t)}\right)$$

$$\mathbf{V}_m^{(t)}(w \times m) = \left(\mathbf{v}_1^{(t)}\ \mathbf{v}_2^{(t)} \cdots \mathbf{v}_m^{(t)}\right)$$

これら２つの行列を用いて、変化度 $a(t)$ は行列 ${}^t\mathbf{U}_m^{(t)}\mathbf{V}_m^{(t)}\ (m \times m)$ の２ノルムを用いて以下のように定義される。

$$a(t)=1-\left\|{}^t\mathbf{U}_m^{(t)}\mathbf{V}_m^{(t)}\right\|_2^{\ 2}=1-\left({}^t\mathbf{U}_m^{(t)}\mathbf{V}_m^{(t)}\text{ の最大特異値}\right)^2$$

ここに、実行列 $\mathbf{A}$ の 2 ノルム $\|\mathbf{A}\|_2$ は以下のように定義される。

$$\|\mathbf{A}\|_2=\max_{\mathbf{u}}\sqrt{\frac{{}^t(\mathbf{Au})\mathbf{Au}}{{}^t\mathbf{uu}}}=\sqrt{\lambda_{\max}}$$

注）行列 $\mathbf{A}$ の特異値とは、行列 $\mathbf{A}$ と $\mathbf{A}$ の随伴行列 $\mathbf{A}^*$ の積 $\mathbf{AA}^*$ （$\mathbf{A}^*\mathbf{A}$）の非負の固有値の平方根である。

2.3　入力と出力がある異常検知

入力と出力の関係で生じる異常の検知方法については、重回帰分析を用いる手法が考えられる。しかし、重回帰分析は、入力変数が多くその値が似通っている場合に、多重共線性の問題が発生する可能性があり、予測が不安定となる。これに対して改善方法と考えられている代表的な手法がリッジ回帰分析と PLS 回帰分析である。リッジ回帰分析は、多重共線性の元となる分散共分散行列に手を加える手法であり、PLS 回帰分析は多重共線性を与える変数間の自由度を制約する手法である。我々のプログラムは 3 者を比較するように作成しており、その違いを理解し易くなっている。

1）重回帰分析

重回帰分析の目的変数を y_λ（$\lambda=1,2,\cdots,N$）、説明変数を $x_{i\lambda}$（$i=1,2,\cdots,p$）とし、それらの関係を ε_λ を誤差項として以下とする。

$$y_\lambda=\sum_{i=1}^{p}b_ix_{i\lambda}+b_0+\varepsilon_\lambda$$

最小 2 乗法としての重回帰分析では、以下の値 D が最小になるように、パラメータ b_i, b_0 を決定する。

$$D=\sum_{\lambda=1}^{N}\left(y_\lambda-\sum_{i=1}^{p}b_ix_{i\lambda}-b_0\right)^2={}^t(\mathbf{y}-\mathbf{Xb})(\mathbf{y}-\mathbf{Xb})$$

ここに、

$$(\mathbf{X})_{\lambda i}=\tilde{x}_{i\lambda}=x_{i\lambda}-\overline{x}_i\ ,\ (\mathbf{y})_\lambda=\tilde{y}_\lambda=y_\lambda-\overline{y}\ ,\ \mathbf{b}={}^t(b_1,b_2,\cdots,b_p)$$

である。パラメータは以下で与えられる。

$$\mathbf{b}=({}^t\mathbf{XX})^{-1}\,{}^t\mathbf{Xy}\ ,\ b_0=\overline{y}-\sum_{i=1}^{p}b_i\overline{x}_i$$

問題となる多重共線性は、行列 ${}^t\mathbf{XX}$ の非正則性から生じる。

多重共線性の判定については、i 番目の説明変数を、他の説明変数で予測して重相関係数 r_i を求め、以下の式で定義される VIF 指標を利用している。

$$VIF_i = 1/(1-r_i^2)$$

一般に VIF 指標が 10 以上であれば多重共線性の疑いがあるとみなされる。

異常度について、通常の重回帰分析では以下で定義する。

$$a(y', \mathbf{x}') = \frac{(y' - b_0 - \mathbf{b}\mathbf{x}')^2}{\sigma^2}$$

ここで分散 σ^2 については以下で推測する。

$$\sigma^2 = \frac{D}{N} = \frac{1}{N}\,{}^t(\mathbf{y} - \mathbf{X}\mathbf{b})(\mathbf{y} - \mathbf{X}\mathbf{b})$$

2) リッジ回帰分析

リッジ回帰分析は重回帰分析の多重共線性の問題に対して、以下のように置くことによって正則性を確保しようとする手法である。

$$\mathbf{b}' = ({}^t\mathbf{X}\mathbf{X} + \eta\mathbf{I})^{-1}\,{}^t\mathbf{X}\mathbf{y}$$

これは、以下を最小化する解でもある。

$$D' = {}^t(\mathbf{y} - \mathbf{X}\mathbf{b}')(\mathbf{y} - \mathbf{X}\mathbf{b}') + \eta\,{}^t\mathbf{b}'\mathbf{b}'$$

ここでパラメータ η の値は以下のようにして求められる。λ 番目の個体を抜いた 1 個抜き交差検証のリッジ回帰パラメータを $\mathbf{b}'^{(-\lambda)}$ とすると、そのときの平均 2 乗誤差 $e(\eta)$ は以下で与えられる。

$$e(\eta) = \frac{1}{N}\sum_{\lambda=1}^{N}(\tilde{y}_\lambda - \sum_{i=1}^{p}\tilde{x}_{i\lambda}b_i'^{(-\lambda)})^2, \qquad \tilde{y}_\lambda = y_\lambda - \bar{y},\ \tilde{x}_{i\lambda} = x_{i\lambda} - \bar{x}$$

これは、以下のように書くこともできる[1]。

$$e(\eta) = \frac{1}{N}\,{}^t\mathbf{A}\mathbf{A}$$

ここに、

$$\mathbf{A} = diag(\mathbf{I} - \mathbf{H})^{-1}(\mathbf{I} - \mathbf{H})\mathbf{y},\quad \mathbf{H}(N \times N) = \mathbf{X}({}^t\mathbf{X}\mathbf{X} + \eta\mathbf{I})^{-1}\,{}^t\mathbf{X}$$

また、$diag(\mathbf{I} - \mathbf{H})^{-1}$ は対角要素が $(\mathbf{I} - (\mathbf{H})_{ii})^{-1}$ となる対角行列である。運用上はパラメータ η の値を変化させて、この $e(\eta)$ が最小になるようなパラメータ η を選ぶ。

もう少し安全性を考えて、以下の一般化交差確認検証法と呼ばれる方法から与えられる誤差 $e_{GCV}(\eta)$ を最小化する場合もある。

$$e_{GCV}(\eta) = \frac{1}{N}\,{}^t\mathbf{A}'\mathbf{A}'$$

ここに、$\mathbf{A}' = (\mathbf{I}-\mathbf{H})\mathbf{y}/[1-tr\mathbf{H}/N]$ である。我々のプログラムでは前者の判定法を利用している。

多重共線性がある場合、重回帰分析の予測は、そのデータに対してだけは良い精度を与えるが、他の新しいデータを用いた場合、予測の精度が著しく低下する。そのため１個抜き交差検証は必須である。

異常度について、リッジ回帰分析では以下で定義する。

$$a(y',\mathbf{x}') = \frac{(y'-b_0'-\mathbf{b}'\mathbf{x}')^2}{\sigma'^2}$$

ここで分散 σ'^2 については以下で推測する。

$$\sigma'^2 = \frac{D'}{N} = \frac{1}{N}[{}^t(\mathbf{y}-\mathbf{X}\mathbf{b}')(\mathbf{y}-\mathbf{X}\mathbf{b}') + \eta\,{}^t\mathbf{b}'\mathbf{b}']$$

我々のプログラムでは、重相関係数と寄与率について、残差分散を通常の重回帰分析通り以下で計算している。

$$\sigma^2 = \frac{1}{N}\,{}^t(\mathbf{y}-\mathbf{X}\mathbf{b}')(\mathbf{y}-\mathbf{X}\mathbf{b}')$$

そのため、出力される分散の値が他の分析（例えば以下の PLS 回帰）より大きいのに、重相関係数の値が大きく表示されるということもある。この定義が妥当かどうか今後考えることとする。

3）ＰＬＳ回帰分析

PLS 回帰分析ではまず、変数の線形結合を考える。

$$r_{i\lambda} = \sum_{j=1}^{p} u_{ij}\tilde{x}_{j\lambda} \qquad (i = 1, 2, \cdots, r\,;\, r < p)$$

この式を、行列記号を用いて書くと以下となる。

$$\mathbf{R} = \mathbf{X}\mathbf{U} \qquad \mathbf{U} = (\mathbf{u}_1\ \mathbf{u}_2 \cdots \mathbf{u}_r)$$

ここで、行列 $\mathbf{U}$ の各列ベクトルは直交し、順番に $\mathbf{X}\mathbf{u}_i$ と $\mathbf{y}$ との内積が最大化されるように選ばれる。詳細は後に示す。

この新しい変数を用いて、目的変数を以下のように予測する。

$$\tilde{y}_\lambda = \sum_{j=1}^{r} \beta_j r_{j\lambda} + \varepsilon_\lambda$$

即ち、

$$\mathbf{y} = \mathbf{R}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}\mathbf{U}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

最小２乗法を使い、以下の量を最小化するようにパラメータを決定する。

$$D'' = {}^t(\mathbf{y}-\mathbf{R}\boldsymbol{\beta})(\mathbf{y}-\mathbf{R}\boldsymbol{\beta})$$

その解は次のように与えられる。

$$\boldsymbol{\beta} = ({}^t\mathbf{RR})^{-1}\,{}^t\mathbf{Ry} = ({}^t\mathbf{U}\,{}^t\mathbf{XXU})^{-1}\,{}^t\mathbf{U}\,{}^t\mathbf{Xy}$$

これから、標準化偏回帰係数$\tilde{\mathbf{b}}$ は以下となる。

$$\tilde{\mathbf{b}} = \mathbf{U}\boldsymbol{\beta}$$

また、回帰係数は、以下で与えられる。

$$b_i'' = \tilde{b}_i s_y \big/ s_i\ ,\quad b_0'' = \bar{y} - \sum_{i=1}^{p} b_i'' \bar{x}_i$$

多重共線性の改善の程度については、変数を$\mathbf{U}$ 行列で変換した後のi 番目の説明変数を、他の説明変数で予測して重相関係数r_i を求め、以下の式で定義される VIF 指標を利用している。

$$VIF_i = 1\big/(1 - r_i^2)$$

異常度については以下で定義する。

$$a(y', \mathbf{x}') = \frac{(y' - b_0'' - \mathbf{b}''\mathbf{x}')^2}{\sigma^2}$$

ここで分散σ^2 については以下で推測する。

$$\sigma^2 = \frac{D}{N} = \frac{1}{N}\,{}^t(\mathbf{y} - \mathbf{Xb}'')(\mathbf{y} - \mathbf{Xb}'')$$

最後に行列$\mathbf{U} = (\mathbf{u}_1\ \mathbf{u}_2 \cdots \mathbf{u}_r)$の決定法について述べる。この行列の 1 列目$\mathbf{u}_1$ は$\mathbf{Xu}_1$ が最も$\mathbf{y}$ の方向に向くように、以下のように求める。

$$L_1 = {}^t\mathbf{yXu}_1 - \mu_1({}^t\mathbf{u}_1\mathbf{u}_1 - 1) \quad \rightarrow \qquad \text{最大化}$$

この解は以下で与えられる。

$$\mathbf{u}_1 = {}^t\mathbf{Xy} \big/ \left\| {}^t\mathbf{Xy} \right\|$$

次の$\mathbf{u}_2$ については、$\mathbf{X}$ から$\mathbf{u}_1$ 方向の成分を取り除き、以下のように求める。

$$L_2 = {}^t\mathbf{y}(\mathbf{X} - \mathbf{d}_1\,{}^t\mathbf{d}_1\mathbf{X})\mathbf{u}_2 - \mu_2({}^t\mathbf{u}_2\mathbf{u}_2 - 1) \quad \rightarrow \qquad \text{最大化}$$

ここに、$\mathbf{d}_1 = \mathbf{Xu}_1 \big/ \left\| \mathbf{Xu}_1 \right\|$ である。確かに$\mathbf{X} - \mathbf{d}_1\,{}^t\mathbf{d}_1\mathbf{X}$ は $\mathbf{u}_1$ 方向の成分を取り除いている。

$$(\mathbf{X} - \mathbf{d}_1\,{}^t\mathbf{d}_1\mathbf{X})\mathbf{u}_1 = \mathbf{Xu}_1 - \mathbf{Xu}_1\,{}^t(\mathbf{Xu}_1)\mathbf{Xu}_1 \big/ \left\| \mathbf{Xu}_1 \right\|^2 = \mathbf{0}$$

この解は以下で与えられる。

$$\mathbf{u}_2 = {}^t(\mathbf{X} - \mathbf{d}_1\,{}^t\mathbf{d}_1\mathbf{X})\mathbf{y} \big/ \left\| {}^t(\mathbf{X} - \mathbf{d}_1\,{}^t\mathbf{d}_1\mathbf{X})\mathbf{y} \right\|$$

このベクトル$\mathbf{u}_2$ は$\mathbf{u}_1$ と直交する。

$${}^t\mathbf{u}_1\mathbf{u}_2 \propto {}^t\mathbf{u}_1\,{}^t(\mathbf{X} - \mathbf{d}_1\,{}^t\mathbf{d}_1\mathbf{X})\mathbf{y} = {}^t(\mathbf{Xu}_1 - \mathbf{Xu}_1\,{}^t(\mathbf{Xu}_1)\mathbf{Xu}_1 \big/ \left\| \mathbf{Xu}_1 \right\|^2)\mathbf{y} = 0$$

これを続けると、k 番目の係数ベクトル$\mathbf{u}_k$ は以下のように求められることが分かる。

$$L_k = {}^t\mathbf{y}\left(\mathbf{X} - \sum_{i=1}^{k-1}\mathbf{d}_i\,{}^t\mathbf{d}_i\mathbf{X}\right)\mathbf{u}_k - \mu_k({}^t\mathbf{u}_k\mathbf{u}_k - 1) \quad \rightarrow \quad \text{最大化}$$

$$\mathbf{u}_k = {}^t\left(\mathbf{X} - \sum_{i=1}^{k-1}\mathbf{d}_i\,{}^t\mathbf{d}_i\mathbf{X}\right)\mathbf{y} \Big/ \left\| {}^t\left(\mathbf{X} - \sum_{i=1}^{k-1}\mathbf{d}_i\,{}^t\mathbf{d}_i\mathbf{X}\right)\mathbf{y} \right\|$$

どこまでの次元数を求めればよいかは、１つの方法として１個抜き交差検証法の重相関係数または同じことであるが、残差分散の大きさを元にして決めればよい。我々のプログラムではこの方法を用いている。

3.　プログラムの利用法

メニュー［分析－ＯＲ－品質管理－異常検知］を選択すると、図１のような分析実行メニューが表示される。

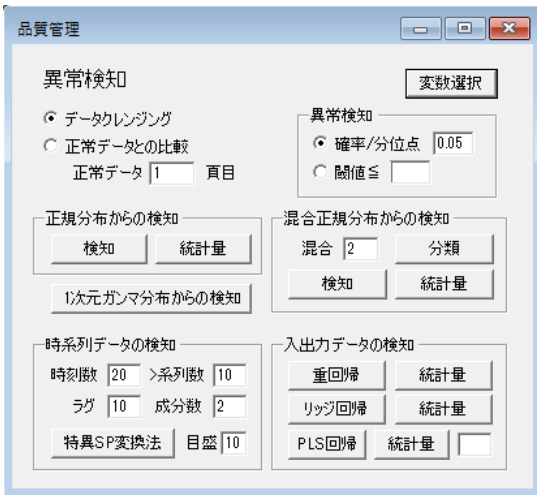


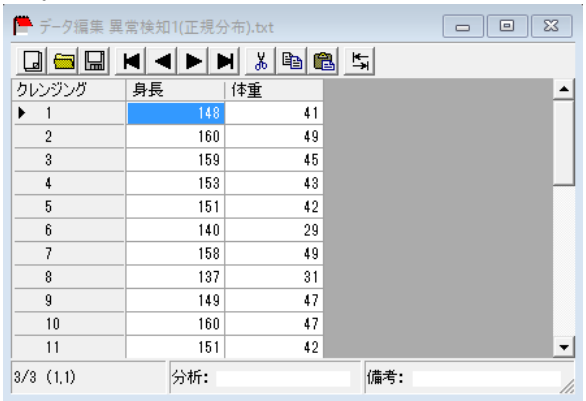
図１　分析実行メニュー

このプログラムでは、グリッドエディタに表示されているデータから、異常データを選別する「データクレンジング」機能と、別頁にある正常データを用いて、現在表示されているデータの異常データを選別する「正常データとの比較」機能がある。後者の場合は、正常データがどの頁にあるかを指定しなければならない。

分析には、(多変量)「正規分布からの検知」、(多変量)「混合正規分布からの検知」、「時系列データの検知」、「入出力データの検知」がある。最初の「正規分布からの検知」は１つの多変量正規分布からのデータのずれを検知するもので、マハラノビス距離を基にした手法である。これは分布が限定されているが、多くのデータでほぼ正規分布の仮定が成り立つと考えられるので、利用範囲は広い。「混合正規分布からの検知」は、多変量正規分布が仮定できない場合で、しかもどのような分布になっているか予想が困難な場合に適用が可能である。これは、分布を

複数の正規分布の重ね合わせとして考えるモデルで、いくつの正規分布の重ね合わせで考えると効果的かという判断も可能である。「時系列データの検知」では、最も実用的と思われる「特異 SP 変換法」(特異スペクトル変換法) が利用できる。「入出力データの検知」では、「重回帰」分析を基本として、多重共線性のある場合の手法で「リッジ回帰」分析と「PLS 回帰」分析が利用可能である。他に「1 次元ガンマ分布からの検知」があるが、これは変数が 1 つの場合にしか適用できないので、現実には利用しにくいかも知れない。

最初に 2 変数のデータを用いて、正規分布と混合正規分布からの検知について、プログラムの説明をする。図 2 に示すファイル「異常検知 1(正規分布).txt」の 3 頁目は、2 変数で異常値を含んだデータである。



クレンジング	身長	体重
1	148	41
2	160	49
3	159	45
4	153	43
5	151	42
6	140	29
7	158	49
8	137	31
9	149	47
10	160	47
11	151	42

図 2　異常値を含んだデータ

この中から異常値を検出するには、「データクレンジング」ラジオボタンを選び、「確率/分位点」ラジオボタンを選んで、異常値の確率値を指定する（この場合は確率値となる)。ここでは 5% に設定している。その後、正規分布からの異常検知の「検知」ボタンをクリックすると図 3 のような出力結果を得る。

	Maha2乗	異常度(f値)	確率	異常
15	2.011	0.908	0.415	0
16	2.851	1.288	0.292	0
17	0.658	0.297	0.745	0
18	1.026	0.463	0.634	0
19	0.565	0.255	0.776	0
20	1.789	0.808	0.456	0
21	14.438	6.520	0.005	1
22	1.718	0.776	0.470	0
23	3.303	1.492	0.242	0
24	0.310	0.140	0.870	0

図 3　データクレンジング検知結果

出力は、このデータから求めた多変量正規分布の平均からのマハラノビス距離の 2 乗、それを元にしたホテリング統計量（F 分布の f 値)、その検定確率、異常かどうかの判定である。判定は正常と異常でそれぞれ 0 または 1 で出力される。

次に、「統計量」ボタンをクリックすると、図４のように、平均や共分散等のパラメータ推定値等と共に、異常の判定に使われるホテリング統計量の閾値が出力される。利用者はこの値を参考にして、閾値としてホテリング統計量を用いてもよい。

平均・共分散・閾値

	身長	体重
平均	149.000	38.700
分散共分散		
身長	51.733	39.433
体重	39.433	40.343
閾値確率	0.050	
自由度	2.28	
閾値(Ht2-f値)	3.340	

図４　パラメータ推定値

次に「正常データとの比較」ラジオボタンをクリックし、同じファイルの１頁目を開く、正常データは２頁目に入っているものとして、「正常データ」テキストボックスの中に２を入力する。「検知」ボタンをクリックすると、図５のように２頁目の正常データから求められるパラメータを元にした、異常検知結果が出力される。

異常検知

	Maha2乗	異常度(f値)	確率	異常
1	2.295	1.036	0.368	0
2	0.484	0.219	0.805	0
3	10.957	4.948	0.014	1
4	0.368	0.166	0.848	0
5	2.384	1.077	0.354	0
6	0.645	0.292	0.749	0
7	2.763	1.248	0.303	0
8	2.048	0.925	0.408	0
9	2.412	1.089	0.350	0
10	6.417	2.898	0.072	0

図５　正常データとの比較検知結果

出力項目については図３と同様である。「統計量」ボタンをクリックした結果は、正常データを元にした結果であり、図４と同じ様式であるので省略する。

ファイル「異常検知 3(複合正規分布).txt [2]」を図６のように読み込み、「データクレンジング」ラジオボタンを選択し、データ処理の結果を見てみる。

データ編集 異常検知3(複合正規分布).txt

350/150	変数1	変数2
1	10.19	4.32
2	9.58	1.48
3	9.71	4.85
4	11.49	5.35
5	10.59	10.92
6	9.15	0.09
7	8.75	4.96
8	9.87	7.96
9	10.04	7.99
10	8.99	6.81

3/3 (1,1)　分析:　備考:

図６　非正規分布のデータ

このデータは、実際には 2 つの正規分布を合わせたものであるが、今の段階ではそれが分からないものとする。仮に「混合」テキストボックスを 2 とし、処理を進める。後にこの数字を変更して最も良いモデルを選択する。

「分類」ボタンをクリックすると、図 7 のように、2 つの群についてのデータの帰属度と分類結果が得られる。

混合正規分布分類

	帰属度1	帰属度2	分類
343	0.999	0.001	1
344	1.000	0.000	1
345	0.999	0.001	1
346	0.984	0.016	1
347	0.999	0.001	1
348	0.999	0.001	1
349	0.999	0.001	1
350	1.000	0.000	1
351	0.000	1.000	2
352	0.000	1.000	2

図 7　レコード毎の帰属度と分類結果

この分類結果をコピーして、元のデータに図 8 のように貼り付け、

データ編集 異常検知3(複合正規分布).txt

350/150	変数1	変数2	
1	10.19	4.32	1
2	9.58	1.48	1
3	9.71	4.85	1
4	11.49	5.35	1
5	10.59	10.92	1
6	9.15	0.09	1
7	8.75	4.96	1
8	9.87	7.96	1
9	10.04	7.99	1
10	8.99	6.81	1

3/3 (1,3)　分析:　備考:

図 8　分類項目の貼り付け

分析「相関と回帰分析」の「先頭列で群分け」ラジオボタンを選択して、図 9 のような散布図を描くことも可能である。

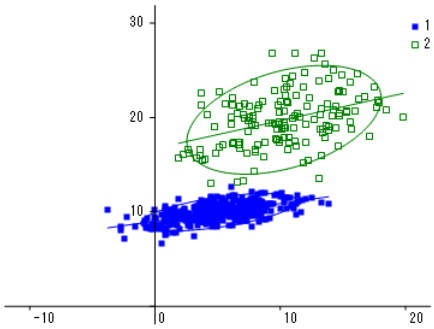


図 9　データの散布図

但し、このグラフには、グラフメニュー［設定－正規楕円半径］を用いて、2σの確率楕円を加えてある。

次に「統計量」ボタンをクリックすると、図 10 のように 2 つの群の平均と共分散、群の生起確率の推測値などが表示される。この中で一番下の BIC はモデル判定によく利用されるベイズ情報量基準と呼ばれるもので、この値が小さいほど良いモデルとして評価される。

混合正規分布パラメータ

	変数1	変数2
群1		
生起確率	0.697	
平均	9.973	5.091
共分散		
変数1	0.992	1.845
変数2	1.845	9.744
群2		
生起確率	0.303	
平均	19.738	10.267
共分散		
変数1	8.540	4.666
変数2	4.666	15.133
閾値(a値)	7.305	
対数尤度	-2392.768	
AIC	4809.535	
BIC	4860.110	

図 10　2 群の場合のパラメータ推定値

現在の 2 群の場合は BIC=4860.110 であるが、3 群にすると 4888.488 となり、2 群の方が良いモデルであると判断される。これは、2 群を故意に作ったモデルであるので、当然の結果である。もちろん AIC についても小さな値が良いモデルと判定される。

また、「検知」ボタンをクリックすると、図 11 のように異常度の値と判別結果が表示される。

混合正規分布検知

	異常度(a値)	異常
97	5.259	0
98	4.859	0
99	4.379	0
100	3.285	0
101	3.165	0
102	3.323	0
103	7.567	1
104	3.786	0
105	3.606	0
106	3.473	0

図 11　混合正規分布からの検知

判定には「確率/分位点」ラジオボックスの中の分位点を利用している。正規性を考えない場合は分位点で判定を与える。

1 次元ガンマ分布でも同様の結果の表示となるので、ここでは省略するが、ガンマ分布の場合は、2 つのパラメータが推測される。またこの場合、1 つのボタンで検知結果と統計量が両方表示される。

時系列データからの異常検知では、正常データを指定しないので、図 1 の分析メニューでは「データクレンジング」ラジオボタンを選んで実行する。時系列データは、図 12 のように与えられる。

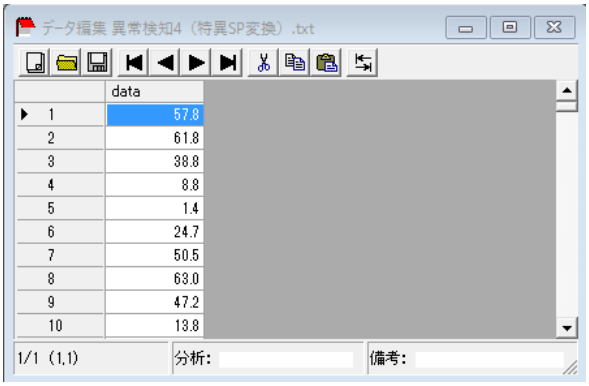


図 12　時系列データ（異常検知 4(特異 SP 変換).txt [2]）

異常検知の方法は簡単で、変数選択をして、「特異 SP 変換」ボタンをクリックする。その際、必要ならば「時刻数」、「系列数」、「ラグ」、「成分数」の値を変更する。これらの値はそれぞれ理論の説明の中の、w, k, L, m に相当する。「目盛」の値は、グラフの目盛間隔の値を与える。図 13 に実行結果を示す。

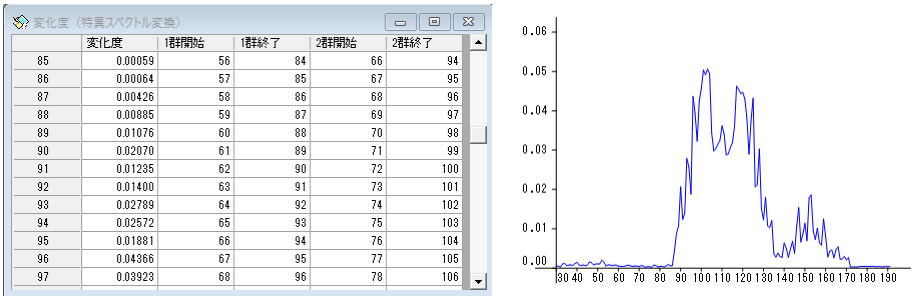


図 13　特異スペクトル変換による異常検知結果

図 13 の表の中で、左端が時刻であり、右 4 列は 2 つの群が利用したデータの範囲を示している。1 群は時刻の 1 つ前までのデータを利用し、2 群はそれから「ラグ」だけ遅れた時刻のデータを使っている。

入力と出力があるデータでは、重回帰分析が基本となる。通常の重回帰分析と同様に、目的変数、説明変数の順に変数を選んで、図 1 の分析メニューの「重回帰分析」ボタンをクリックすると図 14 のような実行結果が表示される。

異常度（重回帰分析）

	実測値	予測値	残差	異常度	判定
4	39	40.859	-1.859	0.160	0
5	81	84.237	-3.237	0.486	0
6	47	53.870	-6.870	2.192	0
7	92	91.659	0.341	0.005	0
8	75	76.076	-1.076	0.054	0
9	77	79.557	-2.557	0.304	0
10	64	61.139	2.861	0.380	0
11	76	82.910	-6.910	2.217	0
12	97	85.622	11.378	6.011	1
13	65	65.551	-0.551	0.014	0

図 14　重回帰分析実行結果（異常検知 6(入出力).txt 1 頁目）

また、隣の「統計量」ボタンをクリックすると、図 15 のような結果が表示される。

統計量（重回帰分析）

	偏回帰係数	標準化係数	VIF	残差分散	閾値	重相関R	寄与率R^2
x1	0.1490	0.0749	1.1796	21.5362	4.0705	0.943	0.889
x2	2.2233	0.0527	1.0554				
x3	2.7614	0.3522	3.9544				
x4	0.4314	0.5888	3.9670				
切片	6.8654	0.0000					

図 15　重回帰分析統計量

これは正常に重回帰分析が行われた結果である。閾値は異常度の閾値である。

データに多重共線性がある場合、例えば、図 16 は変数 x2 と x4 が殆ど同じで、1 番目のデータだけが、0.01 違っている例である。

データ編集 異常検知6(入出力).txt

完全多重共線性	y	x1	x2	x3	x4
1	83	67	3.0	7.4	3.01
2	90	71	3.7	8.0	3.7
3	80	57	3.9	6.5	3.9
4	39	43	2.8	1.8	2.8
5	81	63	3.6	6.1	3.6
6	47	51	3.7	2.7	3.7
7	92	72	4.1	7.9	4.1
8	75	62	3.8	4.6	3.8
9	77	69	3.6	5.8	3.6
10	64	59	3.6	4.2	3.6

4/5 (1,1)　分析:　備考:

図 16　強い多重共線性があるデータ（同 4 頁目）

このデータに対して重回帰分析の「統計量」ボタンをクリックすると、図 17 のような結果が得られる。

統計量（重回帰分析）

	偏回帰係数	標準化係数	VIF	残差分散	閾値	重相関R	寄与率R^2
x1	0.2251	0.1132	1.1862	38.0310	3.3607	0.8967	0.8041
x2	521.1502	12.3456	61886.5468				
x3	6.7407	0.8596	1.1435				
x4	-519.5957	-12.2978	61906.7203				
切片	18.5670	0.0000					

図 17　強い共線性があるデータの重回帰分析統計量

この結果ではほぼ同じ値である変数 x2 と x4 の偏回帰係数が極端に大きくなっている。これは予測において、x2 と x4 の値が少しずれると大きな差となって表れることを示しており、予測の頑健性において問題となる。

これを解決するための代表的な手法に、リッジ回帰分析と PLS 回帰分析がある。理論のところで示したように、リッジ回帰分析は多重共線性の元となる説明変数の共分散行列の対角成分に、ある定数（ここでは $\eta \div N$）を加える方法である。この定数 η は、1 個抜き交差検証の残差分散が最小となるように決める。

図 16 のデータに対して、「リッジ回帰分析」とその「統計量」を与えた結果を図 18 と図 19 に示す。

異常度（リッジ回帰分析）

	実測値	予測値	残差	異常度	判定
4	39	45.709	-6.709	1.005	0
5	81	78.913	2.087	0.097	0
6	47	54.497	-7.497	1.255	0
7	92	93.169	-1.169	0.030	0
8	75	69.391	5.609	0.702	0
9	77	78.488	-1.488	0.049	0
10	64	65.867	-1.867	0.078	0
11	76	78.259	-2.259	0.114	0
12	97	79.665	17.335	6.708	1
13	65	76.217	-11.217	2.809	0

図 18　リッジ回帰分析による異常検知

統計量（リッジ回帰分析）

	偏回帰係数	標準化係数	残差分散	閾値	重相関R	寄与率R^2
x1	0.2466	0.1240	44.7968	2.8086	0.8943	0.7998
x2	0.6154	0.0146			最良η	7.2000
x3	6.3469	0.8094				
x4	0.6093	0.0144				
切片	20.2521	0.0000				

図 19　リッジ回帰分析統計量

また、PLS 回帰分析は多重共線性に対して、独立成分として与える変数の数を減らす方法で対応する。どれだけの変数を減らすかは、リッジ回帰分析と同様に 1 個抜き交差検証の残差分散が最小となるように決める。図 20 と図 21 にメニューの指定で独立成分を 3 にした PLS 回帰分析を実行した結果を示す。

異常度（PLS回帰分析）

	実測値	予測値	残差	異常度	判定
4	39	44.704	-5.704	0.845	0
5	81	79.311	1.689	0.074	0
6	47	54.299	-7.299	1.383	0
7	92	94.278	-2.278	0.135	0
8	75	69.515	5.485	0.781	0
9	77	78.581	-1.581	0.065	0
10	64	65.789	-1.789	0.083	0
11	76	78.611	-2.611	0.177	0
12	97	80.817	16.183	6.801	1
13	65	75.948	-10.948	3.113	0

図 20　PLS 回帰分析による異常検知

統計量（PLS回帰分析）

	偏回帰係数	標準化係数	r-VIF	残差分散	閾値	重相関R	寄与率R^2
▶ x1	0.2120	0.1065	1.0988	38.5072	3.1128	0.8954	0.8017
x2	1.0505	0.0249	1.2990	交差検証R	0.8781	最良自由度	3
x3	6.6703	0.8507	1.2002				
x4	1.0563	0.0250					
切片	17.6834	0.0000					

図 21　PLS 回帰分析統計量

「最良自由度」は、目的変数の最適な自由度である。リッジ回帰分析、PLS 回帰分析とも重回帰分析で問題となった変数 x2 と x4 の偏回帰係数が小さくなっている。これにより、これらの変数のずれの影響は小さく抑えられる。

4.　おわりに

我々は、多変量正規分布に従うデータ、多変量正規分布しないデータ、時系列データ、入力と出力があるデータに関する異常検知のプログラムを作成した。多変量正規分布するデータについてはマハラノビスの距離を元にしたホテリング t^2 統計量を利用して異常検知を行っている。この方法では F 分布の確率値を用いて閾値を指定することができる。多変量正規分布しないデータについては多変量正規分布を確率的に重ね合わせてデータ分布を近似する。分布の重ね合わせには EM アルゴリズムという手法を用いている。

時系列データの異常検知については、時間をずらせて連続データを 2 本取り、その間の違いを定量化することによって異常を調べる特異スペクトル変換法という手法を用いている。入出力のあるデータについては、通常は重回帰分析を用いて予測値のずれの異常を調べるが、データの中に多重共線性が認められる場合も考えられる。その際には、リッジ回帰分析や、PLS 回帰分析を用いて多重共線性の影響を押さえるようにする。リッジ回帰分析は、多重共線性により正則性が問題となる行列の対角成分を変更し、正則性を保証するようにする手法であり、PLS 回帰分析は説明変数の自由度を減らして多重共線性を防ぐ手法である。

現在、これらの異常検知のプログラムは、データと独立に単独で与えられているが、本来はラインの中に組み込み、自動的に異常を検知できるようにして、力を発揮するものである。このプログラムの中から必要な部分が取り出され、実際にライン上で動作するように作り変えられれば大変興味深い。

参考文献

[1] 井出剛, 入門機械学習による異常検知, コロナ社, 2015.

[2] ホームページ http://www.heisei-u.ac.jp/ba/fukui/analysis.html 内のサンプルデータ Samples.zip 内のファイル

Multi-purpose Program for Social System Analysis 30
- Anomaly Detection -

Masayasu FUKUI[*1], Tomoyuki OYAMA[*2] and Nozomu ODA[*2]

[*1] *Department of Business Administration, Faculty of Business Administration, Fukuyama Heisei University*

[*2] *Nippon Pneumatic Manufacturing Co., Ltd.*

Abstract: We have been constructing a unified program on the social system analysis for the purpose of education. This time, we make programs on anomaly detection in the field of quality control. We explain methods of anomaly detection for multivariate normally distributed data, non-normally distributed data, time series data and input output data, and introduce operation of our programs.

Keywords: College Analysis, management science, quality control, anomaly detection, machine learning, EM algorithm

URL: http://www.heisei-u.ac.jp/ba/fukui/

福山平成大学経営学部紀要
第 13 号（2017），139－162 頁

社会システム分析のための統合化プログラム３１
－生存時間分析－

福井正康[*1]、呉暁娜[*2]

[*1] 福山平成大学経営学部

[*2] 福山平成大学大学院経営学研究科

要旨：我々は教育分野での利用を目的に社会システム分析に用いられる様々な手法を統合化したプログラム College Analysis を作成してきた。今回は、中途打ち切りを含むデータから死亡率や生存確率分布を予測する分析手法である生存時間分析のプログラムについて報告する。プログラムには生存時間分布表、Kaplan-Meier のグラフと予測曲線、Cox 比例ハザードモデル、最尤法による Weibull ハザードモデルと混合 Weibull ハザードモデルなどが含まれる。

キーワード：College Analysis、多変量解析、生存時間分析、混合 Weibull 分布、ハザードモデル、EM アルゴリズム

URL：http://www.heisei-u.ac.jp/ba/fukui/

1.　はじめに

生存時間分析は中途打ち切りを含むデータから死亡危険率や生存確率分布を予測する分析手法である。この分析は生物の生存時間だけでなく、機械の故障までの時間などにも利用できる。そのため、死亡という言葉は、あるイベントが発生するまでの時間とした方が的を射ているが、ここでは慣例的に使われてきた死亡や生存という言葉を使うことにする。

生存時間分析では生存数、期間死亡数、期間打ち切り数を元に、期間死亡率、期間生存率、累積生存確率（以後これを表す関数を生存関数と呼ぶ）、ハザードなどを計算し、その生存関数を視覚的にグラフで表す。また、その生存関数から累積死亡確率を求め、それに指数分布や Weibull 分布などの分布関数を当てはめ、モデルを作成する。また、対数ハザードに関して、死亡の原因と考えられる変数の影響を線形関数として導入し、その変数のハザードへの影響を議論する。さらに、分布を仮定して生存関数自体をこれらの変数で予測することも考える。このようなモデルによる分析を、ハザードモデルという。特に、分布によらず、変数の影響を考える有名なモデルには Cox 比例ハザードモデルがある。

人間の死亡や製品の故障のデータを、全時間を通して単一の指数分布や Weibull 分布で表すことは難しい。我々はパラメータの異なるこれらの分布を重ね合わせた混合指数分布や混合 Weibull 分布も考える。これはハザードモデルについても同様であり、混合 Weibull ハザードモデルも考えるようにした。

2. 生存時間分析の理論

2.1 生存時間分析の基礎

時刻$t=0$に$l(0)$個の個体があり、死亡により時刻tに個体数が$l(t)$個になっているものとする。時刻tからの単位時間の間に死亡する割合$p(t)=-dl(t)/dt$は、以下で与えられると仮定する。

$$-\frac{dl(t)}{dt}=\mu(t)l(t)$$

ここに$\mu(t)$を時刻tにおける死力という。

上式を時刻tと時刻$t+h$の間で定積分すると以下の関係を得る。

$$\log l(t+h)-\log l(t)=-\int_t^{t+h}\mu(\tau)d\tau=-\int_0^h\mu(t+\tau)d\tau$$

これより、

$$l(t+h)=l(t)\exp\left[-\int_0^h\mu(t+\tau)d\tau\right]$$

ここで、 $p(h;t)=\exp\left[-\int_0^h\mu(t+\tau)d\tau\right]$とおくと、$p(h;t)$は時刻$t$から$t+h$の間の期間生存率と呼ばれる。この期間生存率は以下のようにも書ける。

$$p(h;t)=\frac{l(t+h)}{l(t)}$$

同様にして、期間死亡率$q(h;t)$も以下のように与えられる。

$$q(h;t)=1-p(h;t)=\frac{l(t)-l(t+h)}{l(t)}=\frac{d(h;t)}{l(t)}$$

ここに$d(h;t)=l(t)-l(t+h)$は期間死亡数を表す。特に、$h=1$とした期間生存率、期間死亡率を単に時刻tでの生存率$p(t)$、死亡率$q(t)$という。

時刻t以降の生存時間の合計$T(t)$を個体数で割った量を平均余命$e(t)$という。

$$e(t)=\int_t^\infty l(\tau)d\tau\Big/l(t)=T(t)/l(t)$$

また、$t=0$での平均余命を平均寿命という。

死亡の発生までの時間を確率変数Tとする確率分布を考え、その密度関数を$f(t)$、分布関数を$F(t)$とすると、これらには以下の関係がある。

$$F(t)=P(0\le T\le t)=\int_0^t f(\tau)d\tau$$

分布関数$F(t)$は累積死亡関数である。これに対して、時刻tまで生きる確率を表す関数$S(t)$を生存関数といい、以下で表される。

$$S(t)=P(T>t)=1-F(t)=\int_t^\infty f(t)dt$$

時刻tにおける死亡発生危険率をハザード関数（故障率関数）$\lambda(t)$といい、以下のように定義される。

$$\lambda(t)=\frac{f(t)}{S(t)}=-\frac{d}{dt}\log S(t)$$

このハザード関数を積分した累積ハザード関数$\Lambda(t)$は以下のように与えられる。

$$\Lambda(t)=\int_0^t \lambda(\tau)d\tau=-\log S(t)$$

逆に生存関数は、以下のように表される。

$$S(t)=e^{-\Lambda(t)}$$

生存関数は$t\to\infty$で$S(t)\to 0$であるから、累積ハザード関数は$t\to\infty$で$\Lambda(t)\to\infty$でなければならない。

累積死亡分布には、指数分布や Weibull 分布が仮定される。指数分布の確率密度関数は以下で与えられる。

$$f(t)=\lambda e^{-\lambda t} \qquad (t\geq 0)$$

分布関数と生存関数はそれぞれ以下で与えられる。

$$F(t)=1-e^{-\lambda t},\quad S(t)=e^{-\lambda t} \qquad (t\geq 0)$$

確率変数の平均と分散はそれぞれ以下で与えられる。

$$E[T]=\frac{1}{\lambda},\quad V[T]=\frac{1}{\lambda^2}$$

ハザード関数は定数で与えられる。

$$\lambda(t)=\frac{f(t)}{S(t)}=\frac{\lambda e^{-\lambda t}}{e^{-\lambda t}}=\lambda$$

Weibull 分布の確率密度関数は以下で与えられる。

$$f(t)=(a/b)\left(t/b\right)^{a-1}\exp\left[-\left(t/b\right)^a\right] \qquad (t\geq 0)$$

分布関数と生存関数はそれぞれ以下で与えられる。

$$F(t)=1-\exp\left[-\left(t/b\right)^a\right],\quad S(t)=\exp\left[-\left(t/b\right)^a\right]$$

確率変数の平均と分散はそれぞれ以下で与えられる。

$$E[T]=b\,\Gamma\left(1+1/a\right),\quad V[T]=b^2\left[\Gamma(2+1/a)-\Gamma(1+1/a)\right]^2$$

ここに、$\Gamma(x)$はガンマ関数である。ハザード関数は以下で与えられる。

$$\lambda(t)=\frac{f(t)}{S(t)}=\frac{(a/b)\left(t/b\right)^{a-1}\exp\left[-\left(t/b\right)^a\right]}{\exp\left[-(t/b)^a\right]}=(a/b)\left(t/b\right)^{a-1}=at^{a-1}/b^a$$

実際のハザード関数は、初期段階で値が大きく、しばらく時間が経つと安定期に入り、最終的な段階でまた値が大きくなる。安定期では指数分布が使われ、初期段階では Weibull 分布がよく利用される。最終段階ではあまり当てはまりが良くないと言われることもあるが、我々は Weibull 分布を当てはめてみる。全体への当てはめの分布としては、後に述べる混合 Weibull 分布を考えてみることにする。

2.2 Kaplan-Meier 推定と log-rank 検定

観測対象 $\lambda = 1, \cdots, N$ に対して、生存時間を $t_\lambda = 0$ から $t_\lambda = T_\lambda$（打ち切りのないデータ）、$t_\lambda = 0$ から $t_\lambda = T_\lambda^+$（打ち切りのあるデータ、実際のデータでは 17+ 等と表記）とする。この終了時刻 T_λ を 0 から順番に並べた時刻を $t_0 = 0, t_1, \cdots, t_m$（同一のものもある）とし、$t_m$ ですべて死亡および打ち切りが確認されたものとする。これに対して、一定の時間間隔で時刻を取る方法もある。各時点での生存数を l_i、$t_i < t \le t_{i+1}$ の間に死亡した数を d_i、打ち切りになった数を w_i とする。これらを使って、死亡のリスクにさらされた数を $r_i = l_i - w_i/2$ とする。

死亡の期間発生率 q_i と期間生存率 p_i は以下で与えられる。

$$q_i = d_i / r_i, \quad p_i = 1 - q_i$$

生存関数 S_i、密度関数 f_i、ハザード関数 λ_i は以下のように計算される。

$$S_i = \prod_{k=0}^{i-1} p_k, \quad f_i = q_i S_i / (t_i - t_{i-1}), \quad \lambda_i = f_i / S_i = q_i / (t_i - t_{i-1})$$

このような生存関数の推定法を Kaplan-Meier の product-limit 推定法（以後 Kaplan-Meier 推定法と呼ぶ）という。生存関数 S_i のばらつきを表す標準誤差 $S.E.[S_i]$ は近似的に以下で与えられることが知られている。

$$S.E.[S_i] = S_{i-1}\sqrt{\sum_{k=1}^{i-1} \frac{d_k}{l_k(l_k - d_k)}} \qquad (i \ge 2)$$

指数分布や Weibull 分布の見極めは、累積ハザード関数に関する以下の関係を利用し、グラフが直線になるか否かで判断することができる。

指数分布　$-\log S(t) = \lambda t$

Weibull 分布　$\log(-\log S) = a \log(t/b) = a \log t - a \log b$

指数分布や Weibull 分布のパラメータの最小 2 乗推定は、以下の式によって与えられる。

指数分布　$S(t) = e^{-\lambda t}$

$$\lambda = -\sum_{i=0}^{m-1} t_i \log S_i \Big/ \sum_{i=0}^{m-1} t_i^2$$

Weibull 分布　$S(t)=\exp\left[-\left(t/b\right)^{a}\right]$

$t_i'=\log t_i$，　$S_i'=\log(-\log S_i)$　として、

$$a=\sum_{i=1}^{m-1}(t_i'-\bar{t}')(S_i'-\bar{S}')\Big/\sum_{i=1}^{m-1}(t_i'-\bar{t}')^2\ ,\quad b=\exp\left[-(\bar{S}'-a\bar{t}')/a\right]$$

分類数G の個体群について、生存時間データの差の検定を行うには以下の性質を用いる。第r 分類群のt_i 時点での期間死亡数をd_i^r 、生存数をl_i^r として

$$O_r=\sum_{i=0}^{m-1}d_i^r\ ,\quad E_r=\sum_{i=0}^{m-1}l_i^r(d_i/l_i)\ ,\quad \text{ここに、}l_i=\sum_{r=1}^{G}l_i^r\ ,\quad d_i=\sum_{r=1}^{G}d_i^r$$

を計算し、以下の近似的な関係を用いて群間の差を検定する。

$$\chi^2=\sum_{r=1}^{G}\frac{(O_r-E_r)^2}{E_r}\sim\chi^2_{G-1}$$

ここに、O_r は分類群r　の実測累積死亡数、E_r は分類群r　の予測累積死亡数である。この検定を Peto & Peto の log-rank 検定という。

2.3　パラメータの最尤推定

1）指数分布に基づく最尤推定

最初に通常の指数分布の最尤推定を考える。指数分布の確率密度関数と生存関数は以下で与えられる。

$$f(t)=\lambda\exp(-\lambda\,\mathrm{t})\qquad(t\geq 0)$$

$$S(t)=\exp(-\lambda\,\mathrm{t})\qquad(t\geq 0)$$

指数分布の最尤推定で、尤度$L(\lambda)$ は以下で与えられる。

$$L(\lambda)=\prod_{i=1}^{N}f(t_i)^{\delta_i}S(t_i)^{1-\delta_i}$$

ここで、打ち切りデータと非打ち切り（死亡）データをそれぞれ$\delta_i=0,1$ としている。

ハザード関数は以下で与えられる。

$$\lambda(t)=\frac{f(t)}{S(t)}=\lambda$$

対数尤度は以下となる。

$$\log L(\lambda)=\sum_{i=1}^{N}\left[\delta_i\log\lambda(t_i)+\log S(t_i)\right]=\sum_{i=1}^{N}\left[\delta_i\log\lambda-\lambda t_i\right]$$

対数尤度を微分してスコアベクトルに相当するものを作成するが、この場合はスカラーである。これを仮にスコアと呼ぶ。

$$\frac{\partial}{\partial\lambda}\log L=\sum_{i=1}^{N}\left[\delta_i/\lambda-t_i\right]=\frac{1}{\lambda}\sum_{i=1}^{N}\delta_i-\sum_{i=1}^{N}t_i=0$$

$$\lambda=\sum_{i=1}^{N}\delta_i\Big/\sum_{i=1}^{N}t_i$$

スコアをもう一度微分して、情報行列$\Im$に相当するものを作成する。この場合もスカラーである。

$$\Im=-\frac{\partial^2}{\partial\lambda^2}\log L=\frac{1}{\lambda^2}\sum_{i=1}^{N}\delta_i$$

この逆数は、推定値の分散を与える。

2) Weibull 分布に基づく最尤推定

最初に通常の Weibull 分布の最尤推定を考える。weibull 分布の確率密度関数と生存関数は以下で与えられる。

$$f(t)=(a/b)\left(t/b\right)^{a-1}\exp\left[-\left(t/b\right)^{a}\right] \qquad (t\geq 0)$$

$$S(t)=\exp\left[-\left(t/b\right)^{a}\right] \qquad (t\geq 0)$$

Weibull 分布の最尤推定で、尤度$L(a,b)$は以下で与えられる。

$$L(a,b)=\prod_{i=1}^{N}f(t_i)^{\delta_i}S(t_i)^{1-\delta_i}$$

ここで、打ち切りデータと非打ち切り（死亡）データをそれぞれ$\delta_i=0,1$としている。

ハザード関数は以下で与えられる。

$$\lambda(t)=\frac{f(t)}{S(t)}=\frac{(a/b)\left(t/b\right)^{a-1}\exp\left[-\left(t/b\right)^{a}\right]}{\exp\left[-(t/b)^{a}\right]}=(a/b)\left(t/b\right)^{a-1}=at^{a-1}b^{-a}$$

対数尤度は以下となる。

$$\begin{aligned}\log L(a,b)&=\sum_{i=1}^{N}\left[\delta_i\log\lambda(t_i)+\log S(t_i)\right]\\&=\sum_{i=1}^{N}\left[\delta_i\log\left(at_i^{a-1}b^{-a}\right)-t_i^{a}b^{-a}\right]=\sum_{i=1}^{N}\left[\delta_i\log\left(at_i^{a-1}e^{\beta}\right)-t_i^{a}e^{\beta}\right]\\&=\sum_{i=1}^{N}\left[\delta_i\left(\log a+(a-1)\log t_i+\beta\right)-t_i^{a}e^{\beta}\right]\end{aligned}$$

ここで、$b^{-a}=e^{\beta}$（$b=e^{-\beta/a}$，$e^{\beta}=b^{-a}\rightarrow\exp({}^{t}\mathbf{x\beta})$に相当）としている。

これを微分して、スコアベクトル$\mathbf{U}$と情報行列$\Im$をもとめると以下となる。

$$\boldsymbol{\beta}' = \begin{pmatrix} a \\ \beta \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} \partial \log L / \partial a \\ \partial \log L / \partial \beta \end{pmatrix}, \quad \mathfrak{I} = -\begin{pmatrix} \partial^2 \log L / \partial a^2 & \partial^2 \log L / \partial a \partial \beta \\ \partial^2 \log L / \partial a \partial \beta & \partial^2 \log L / \partial \beta^2 \end{pmatrix}$$

ここに、

$$\frac{\partial}{\partial a} \log L = \sum_{i=1}^{N} \left[\delta_i (1/a + \log t_i) - \log t_i \, t_i^a e^{\beta} \right]$$

$$\frac{\partial}{\partial \beta} \log L = \sum_{i=1}^{N} \left[\delta_i - t_i^a e^{\beta} \right]$$

$$\frac{\partial^2}{\partial a^2} \log L = -\sum_{i=1}^{N} \left[\delta_i / a^2 + (\log t_i)^2 t_i^a e^{\beta} \right]$$

$$\frac{\partial}{\partial a \partial \beta} \log L = -\sum_{i=1}^{N} \log t_i \, t_i^a e^{\beta}$$

$$\frac{\partial^2}{\partial \beta^2} \log L = -\sum_{i=1}^{N} t_i^a e^{\beta}$$

これらを用いてニュートン・ラフソン法でパラメータの推定を行う。

$$\boldsymbol{\beta}'^{(m+1)} = \boldsymbol{\beta}'^{(m)} + (\mathfrak{I}^{(m)})^{-1} \mathbf{U}^{(m)}$$

ここに右肩の添え字はニュートン・ラフソン法のループの段階を表している。

この情報行列の逆行列の対角成分はパラメータの分散を与える。

3）混合分布に基づく最尤推定

混合分布の最尤推定で、尤度 $L(\lambda)$ は以下で与えられる。

$$L(\lambda) = \prod_{i=1}^{N} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$$

K 種混合分布では、それぞれの密度関数を $f_k(t)$、生存関数を $S_k(t)$ として、全体の密度関数と生存関数は以下となる。ここに、π_k は分布の重ね合わせの確率である。

$$f(t) = \sum_{k=1}^{K} \pi_k f_k(t), \quad S(t) = \sum_{k=1}^{K} \pi_k S_k(t)$$

混合分布の最尤推定で、尤度 $L(\boldsymbol{\theta}, \boldsymbol{\pi})$ は以下で与えられる。

$$L(\boldsymbol{\theta}, \boldsymbol{\pi}) = \prod_{i=1}^{N} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} = \prod_{i=1}^{N} \left(\sum_{k=1}^{K} \pi_k f_k(t_i) \right)^{\delta_i} \left(\sum_{k=1}^{K} \pi_k S_k(t_i) \right)^{1-\delta_i}$$

ここで、打ち切りデータと非打ち切り（死亡）データをそれぞれ $\delta_i = 0,1$ としている。

対数尤度は以下となる。

$$
\begin{aligned}
\log L(\boldsymbol{\theta},\boldsymbol{\pi}) &= \sum_{i=1}^{N}\left[\delta_i \log\sum_{k=1}^{K}\pi_k f_k(t_i) + (1-\delta_i)\log\sum_{k=1}^{K}\pi_k S_k(t_i)\right] \\
&= \sum_{i=1}^{N}\left[\delta_i \log\sum_{k=1}^{K}q_k^{(i)}\frac{\pi_k f_k(t_i)}{q_k^{(i)}} + (1-\delta_i)\log\sum_{k=1}^{K}q_k^{(i)}\frac{\pi_k S_k(t_i)}{q_k^{(i)}}\right] \\
&\geq \sum_{i=1}^{N}\left[\sum_{k=1}^{K}q_k^{(i)}\delta_i \log\frac{\pi_k f_k(t_i)}{q_k^{(i)}} + \sum_{k=1}^{K}q_k^{(i)}(1-\delta_i)\log\frac{\pi_k S_k(t_i)}{q_k^{(i)}}\right] \\
&= \sum_{i=1}^{N}\sum_{k=1}^{K}q_k^{(i)}\left[\delta_i \log f_k(t_i) + (1-\delta_i)\log S_k(t_i) + \log\pi_k - \log q_k^{(i)}\right]
\end{aligned}
$$

上式の不等号は、$q_k^{(i)}$ の値によって、等号になることが知られている。

パラメータの推定には以下の手順①と②をパラメータ値が収束するまで繰り返す。このような 2 段階の推定法を EM アルゴリズムという。

①パラメータ $q_k^{(i)}, \pi_k$ の最適化

この $q_k^{(i)}$ について、$\sum_{k=1}^{K}q_k^{(i)} = 1$ の条件をつけて右辺を最大化するために、ラグランジュの未定定数法を用いる。

$$
\begin{aligned}
&\frac{\partial}{\partial q_k^{(i)}}\left[\log L(\boldsymbol{\theta},\boldsymbol{\pi}) - \sum_{i=1}^{N}\eta_i\left(\sum_{k=1}^{K}q_k^{(i)} - 1\right)\right] \\
&\quad = \delta_i \log\frac{\pi_k f_k(t_i)}{q_k^{(i)}} + (1-\delta_i)\log\frac{\pi_k S_k(t_i)}{q_k^{(i)}} + 1 - \eta_i \\
&\quad = \log\frac{\pi_k f_k(t_i)^{\delta_i} S_k(t_i)^{1-\delta_i}}{q_k^{(i)}} - 1 - \eta_i = 0
\end{aligned}
$$

これより、

$$
q_k^{(i)} = e^{-(1+\eta_i)}\pi_k f_k(t_i)^{\delta_i} S_k(t_i)^{1-\delta_i} = \frac{\pi_k f_k(t_i)^{\delta_i} S_k(t_i)^{1-\delta_i}}{\sum_{k=1}^{K}\pi_k f_k(t_i)^{\delta_i} S_k(t_i)^{1-\delta_i}}
$$

これを書き換えて、以下のようにすることもできる。

$$
q_k^{(i)} = \pi_k f_k(t_i) \Big/ \sum_{k=1}^{K}\pi_k f_k(t_i) \qquad \text{for } \delta_i = 1
$$

$$
q_k^{(i)} = \pi_k S_k(t_i) \Big/ \sum_{k=1}^{K}\pi_k S_k(t_i) \qquad \text{for } \delta_i = 0
$$

この $q_k^{(i)}$ を群 k への帰属度という。

次に、この尤度関数をパラメータ π_j で微分して 0 と置き、パラメータの推定を行うが、$\sum_{k=1}^{K}\pi_k = 1$ の条件をつけるために、ラグランジュの未定定数法を用いる。

$$\frac{\partial}{\partial \pi_j}\left[\log L(\boldsymbol{\theta},\boldsymbol{\pi}) - \eta\left(\sum_{k=1}^{K}\pi_k - 1\right)\right] = \sum_{i=1}^{N} q_j^{(i)} \Big/ \pi_j - \eta = 0$$

より、

$$\pi_k = \frac{1}{\eta}\sum_{i=1}^{N} q_k^{(i)}, \qquad \sum_{k=1}^{K}\pi_k = \frac{1}{\eta}\sum_{k=1}^{K}\sum_{i=1}^{N} q_k^{(i)} = \frac{1}{\eta}\sum_{i=1}^{N}\sum_{k=1}^{K} q_k^{(i)} = \frac{1}{\eta}\sum_{i=1}^{N} 1 = \frac{N}{\eta} = 1$$

となり、以下の関係を得る。

$$\pi_k = \frac{1}{N}\sum_{i=1}^{N} q_k^{(i)}$$

②パラメータ $\boldsymbol{\theta}$ の推定

パラメータ $\boldsymbol{\theta}$ の最尤法による推定では、$q_k^{(i)}, \pi_k$ は①の方法で求められた既知の定数として計算する。この部分の計算については具体的な関数形を用いて考える。

4）混合指数分布に基づく最尤推定

指数分布の確率密度関数と生存関数の以下の具体的な表式を代入すると

$$f_k(t) = \lambda_k \exp(-\lambda_k t), \qquad S_k(t) = \exp(-\lambda_k t)$$

対数尤度は以下のようになる。

$$\begin{aligned}\log L(\boldsymbol{\lambda},\boldsymbol{\pi}) &\geq \sum_{i=1}^{N}\sum_{k=1}^{K} q_k^{(i)}\left[\delta_i\left(\log\lambda_k - \lambda_k t_i\right) - (1-\delta_i)\lambda_k t_i + \log\pi_k - \log q_k^{(i)}\right] \\ &= \sum_{i=1}^{N}\sum_{k=1}^{K} q_k^{(i)}\left[\delta_i \log\lambda_k - \lambda_k t_i + \log\pi_k - \log q_k^{(i)}\right]\end{aligned}$$

これより、群 k への帰属度は以下となる。

$$q_k^{(i)} = \pi_k\lambda_k \exp(-\lambda_k t_i) \Big/ \sum_{k=1}^{K}\pi_k\lambda_k \exp(-\lambda_k t_i) \qquad \text{for } \delta_i = 1$$

$$q_k^{(i)} = \pi_k \exp(-\lambda_k t_i) \Big/ \sum_{k=1}^{K}\pi_k \exp(-\lambda_k t_i) \qquad \text{for } \delta_i = 0$$

対数尤度を微分して、スコアベクトルを求め、それを 0 とする。

$$\frac{\partial}{\partial\lambda_j}\log L = \sum_{i=1}^{N} q_j^{(i)}\left(\delta_i/\lambda_j - t_i\right) = \frac{1}{\lambda_j}\sum_{i=1}^{N} q_j^{(i)}\delta_j - \sum_{i=1}^{N} q_j^{(i)} t_j = 0$$

これより、

$$\lambda_j = \sum_{i=1}^{N} q_j^{(i)} \delta_j \Big/ \sum_{i=1}^{N} q_j^{(i)} t_j$$

スコアをもう一度微分して、情報行列 $\Im$ に相当するものを作成する。

$$\Im_{jk} = -\frac{\partial^2}{\partial\lambda_j \partial\lambda_k} \log L = \frac{\delta_{jk}}{\lambda_j^2} \sum_{i=1}^{N} q_j^{(i)} \delta_i$$

この逆行列の対角成分は、推定値の分散を与える。

5) 混合 Weibull 分布に基づく最尤推定

K 種混合 Weibull では、以下となる。

$$f(t) = \sum_{k=1}^{K} \pi_k f_k(t) = \sum_{k=1}^{K} \pi_k a_k t^{a_k - 1} b_k^{-a_k} \exp(-t^{a_k} b_k^{-a_k})$$

$$= \sum_{k=1}^{K} \pi_k a_k t^{a_k - 1} e^{\beta_k} \exp(-t^{a_k} e^{\beta_k})$$

$$S(t) = \sum_{k=1}^{K} \pi_k S_k(t) = \sum_{k=1}^{K} \pi_k \exp(-t^{a_k} b_k^{-a_k}) = \sum_{k=1}^{K} \pi_k \exp(-t^{a_k} e^{\beta_k})$$

混合 Weibull 分布の対数尤度は以下となる。

$$\log L(\mathbf{a}, \boldsymbol{\beta}, \boldsymbol{\pi}) \geq \sum_{i=1}^{N} \sum_{k=1}^{K} q_k^{(i)} \Big[\delta_i \left(\log a_k + (a_k - 1) \log t_i + \beta_k \right) - t_i^{a_k} e^{\beta_k} + \log \pi_k - \log q_k^{(i)} \Big]$$

これより、群 k への帰属度は以下となる。

$$q_k^{(i)} = \frac{\pi_k a_k t^{a_k - 1} e^{\beta_k} \exp(-t^{a_k} e^{\beta_k})}{\sum_{k=1}^{K} \pi_k a_k t^{a_k - 1} e^{\beta_k} \exp(-t^{a_k} e^{\beta_k})} \qquad \text{for } \delta_i = 1$$

$$q_k^{(i)} = \frac{\pi_k \exp(-t^{a_k} e^{\beta_k})}{\sum_{k=1}^{K} \pi_k \exp(-t^{a_k} e^{\beta_k})} \qquad \text{for } \delta_i = 0$$

ここで、$b_k^{-a_k} = e^{\beta_k}$ （$b_k = e^{-\beta_k / a_k}$ に相当）としている。

$$\frac{\partial}{\partial a_j} \log L(\mathbf{a}, \boldsymbol{\beta}, \boldsymbol{\pi}) = \sum_{i=1}^{N} q_j^{(i)} \Big[\delta_i \left(1/a_j + \log t_i \right) - \log t_i \, t_i^{a_j} e^{\beta_j} \Big]$$

$$\frac{\partial}{\partial \beta_j} \log L(\mathbf{a}, \boldsymbol{\beta}, \boldsymbol{\pi}) = \sum_{i=1}^{N} q_j^{(i)} \Big[\delta_i - t_i^{a_j} e^{\beta_j} \Big]$$

$$\frac{\partial^2}{\partial a_j \partial a_k}\log L(\mathbf{a},\boldsymbol{\beta},\boldsymbol{\pi}) = -\delta_{jk}\sum_{i=1}^{N} q_j^{(i)}\left[\delta_i \big/ a_j^2 + (\log t_i)^2\, t_i^{a_j} e^{\beta_j}\right]$$

$$\frac{\partial^2}{\partial a_j \partial \beta_k}\log L(\mathbf{a},\boldsymbol{\beta},\boldsymbol{\pi}) = -\delta_{jk}\sum_{i=1}^{N} q_j^{(i)} \log t_i\, t_i^{a_j} e^{\beta_j}$$

$$\frac{\partial^2}{\partial \beta_j \partial \beta_k}\log L(\mathbf{a},\boldsymbol{\beta},\boldsymbol{\pi}) = -\delta_{jk}\sum_{i=1}^{N} q_j^{(i)} t_i^{a_j} e^{\beta_j}$$

2.4　比例ハザードモデル

比例ハザードモデルはハザート関数に対して、説明変数 $\mathbf{x} = {}^t(1, x_1, x_2, \cdots, x_p)$ とパラメータ $\boldsymbol{\beta} = {}^t(\beta_0, \beta_1, \beta_2, \cdots, \beta_p)$ を用いて、以下の仮定を行う。

$$\lambda(t \mid \mathbf{x},\boldsymbol{\beta}) = \lambda_0(t)\exp({}^t\mathbf{x}\boldsymbol{\beta}) \quad \text{ここに、} {}^t\mathbf{x}\boldsymbol{\beta} = \sum_{i=1}^{p} x_i\beta_i + \beta_0$$

Cox の比例ハザードモデルでは $\lambda_0(t)$ と定数項 β_0 について議論しないが、Weibull ハザードモデルでは以下のように時間に関して Weibull 分布のハザード関数を仮定する。

$$\lambda(t \mid \mathbf{x},\boldsymbol{\beta}) = (a/b)\left(t/b\right)^{a-1} = at^{a-1}b^{-a} = at^{a-1}\exp({}^t\mathbf{x}\boldsymbol{\beta})$$

1）Cox の比例ハザードモデル

Cox の比例ハザードモデルでは、尤度関数に対して近似的な部分尤度関数を考えて処理を行う。その対数尤度は以下で与えられる[3]。

$$\log L'(\boldsymbol{\beta}) = \sum_{i=0}^{m-1}\left[\sum_{j\in D_i} {}^t\mathbf{x}_j\boldsymbol{\beta} - d_i \log \sum_{j\in R_i}\exp({}^t\mathbf{x}_j\boldsymbol{\beta})\right]$$

ここに、$\boldsymbol{\beta}$ は定数項を除いた偏回帰係数ベクトル、D_i は $t_i < t \le t_{i+1}$ で亡くなった個体の集合、R_i は時刻 t_i で生存が確認されている個体の集合である。これを最大化するようにニュートン・ラフソン法を使って $\boldsymbol{\beta}$ を求める。ここでは $w_j = \exp({}^t\mathbf{x}_j\boldsymbol{\beta})$ として以下の値を示しておく。

$$\mathbf{U} \equiv \frac{\partial}{\partial\boldsymbol{\beta}}\log L'(\boldsymbol{\beta}) = \sum_{i=1}^{m-1}\left[\sum_{i\in D_i}\mathbf{x}_j - d_i\sum_{j\in R_i} w_j\mathbf{x}_j \Big/ \sum_{j\in R_i} w_j\right]$$

$$\mathfrak{I} \equiv -\frac{\partial^2}{\partial\boldsymbol{\beta}\partial{}^t\boldsymbol{\beta}}\log L'(\boldsymbol{\beta})$$

$$= \sum_{i=1}^{m-1} d_i\left[\sum_{j\in R_i} w_j\mathbf{x}_j\,{}^t\mathbf{x}_j \Big/ \sum_{j\in R_i} w_j - \sum_{j\in R_i} w_j\mathbf{x}_j \sum_{j\in R_i} w_j\,{}^t\mathbf{x}_j \Big/ \left(\sum_{j\in R_i} w_j\right)^2\right]$$

この $\mathbf{U}$ をスコアベクトル、$\mathfrak{I}$ を情報行列という。$\boldsymbol{\beta}$ の推定値は以下の計算を繰り返して求める。

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + (\mathfrak{I}^{(m)})^{-1}\mathbf{U}^{(m)}$$

ここに右肩の添え字はニュートン・ラフソン法のループの段階を表している。

2) Weibull ハザードモデル

Weibull ハザードモデルは、ハザード関数に対して以下の仮定を行う。

$$\lambda(t) = \frac{f(t)}{S(t)} = (a/b)\left(t/b\right)^{a-1} = at^{a-1}/b^{a} = at^{a-1}\exp({}^{t}\mathbf{x}\boldsymbol{\beta})$$

通常の Weibull 分布との関係は以下である。

$$b^{-a} = e^{\beta} \to \exp({}^{t}\mathbf{x}\boldsymbol{\beta}) \quad （\beta \to {}^{t}\mathbf{x}\boldsymbol{\beta} \equiv \sum_{i=1}^{p} x_i\beta_i + \beta_0）$$

これより、$b = \exp(-{}^{t}\mathbf{x}\boldsymbol{\beta}/a)$ であるから、$\mu \equiv E[T] = b\,\Gamma\left(1+1/a\right)$ より、

$$\eta \equiv {}^{t}\mathbf{x}\boldsymbol{\beta} = -a\log b = -a\log\left(\mu/\Gamma(1+1/a)\right)$$

となり、右辺が一般化線形モデルの連結関数となる。

この関係を用いて、密度関数と生存関数を求めると以下となる。

$$f(t) = at^{a-1}\exp({}^{t}\mathbf{x}\boldsymbol{\beta})\exp\left[-t^{a}\exp({}^{t}\mathbf{x}\boldsymbol{\beta})\right]$$

$$S(t) = \exp\left[-\left(t/b\right)^{a}\right] = \exp\left[-t^{a}b^{-a}\right] = \exp\left[-t^{a}\exp({}^{t}\mathbf{x}\boldsymbol{\beta})\right]$$

打ち切りデータと非打ち切り（死亡）データをそれぞれ $\delta_i = 0,1$ と区別し、尤度を求めると以下となる。添え字 i について、ここでは個体の番号として使っている。

$$L(\alpha,\boldsymbol{\beta}) = \prod_{i=1}^{N} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}$$

さらに、対数尤度は以下となる。

$$\begin{aligned}\log L(\alpha,\boldsymbol{\beta}) &= \sum_{i=1}^{N}\left[\delta_i\log\lambda(t_i) + \log S(t_i)\right] \\ &= \sum_{i=1}^{N}\left[\delta_i\log\left(at_i^{a-1}\exp({}^{t}\mathbf{x}_i\boldsymbol{\beta})\right) - t_i^{a}\exp({}^{t}\mathbf{x}_i\boldsymbol{\beta})\right] \\ &= \sum_{i=1}^{N}\left[\delta_i\left(\log a + (a-1)\log t_i + {}^{t}\mathbf{x}_i\boldsymbol{\beta}\right) - t_i^{a}\exp({}^{t}\mathbf{x}_i\boldsymbol{\beta})\right]\end{aligned}$$

対数尤度を微分してスコアベクトル $\mathbf{U}$ と情報行列 $\mathfrak{I}$ を求めると以下となる。

$$\boldsymbol{\beta}' = \begin{pmatrix} a \\ \boldsymbol{\beta} \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} \partial\log L/\partial a \\ \partial\log L/\partial\boldsymbol{\beta} \end{pmatrix},$$

$$\mathfrak{I} = -\begin{pmatrix} \partial^2 \log L / \partial a^2 & \partial^2 \log L / \partial a \partial {}^t\boldsymbol{\beta} \\ \partial^2 \log L / \partial a \partial \boldsymbol{\beta} & \partial^2 \log L / \partial \boldsymbol{\beta} \partial {}^t\boldsymbol{\beta} \end{pmatrix}$$

ここに、

$$\frac{\partial}{\partial a} \log L = \sum_{i=1}^{N} \left[\delta_i (1/a + \log t_i) - \log t_i t_i^a \exp({}^t\mathbf{x}_i \boldsymbol{\beta}) \right]$$

$$\frac{\partial}{\partial \boldsymbol{\beta}} \log L = \sum_{i=1}^{N} \left[\delta_i \mathbf{x}_i - \mathbf{x}_i t_i^a \exp({}^t\mathbf{x}_i \boldsymbol{\beta}) \right]$$

$$\frac{\partial^2}{\partial a^2} \log L = \sum_{i=1}^{N} \left[-\delta_i / a^2 - (\log t_i)^2 t_i^a \exp({}^t\mathbf{x}_i \boldsymbol{\beta}) \right]$$

$$\frac{\partial^2}{\partial a \partial \boldsymbol{\beta}} \log L = -\sum_{i=1}^{N} (\log t_i) \mathbf{x}_i t_i^a \exp({}^t\mathbf{x}_i \boldsymbol{\beta})$$

$$\frac{\partial^2}{\partial \boldsymbol{\beta} \partial {}^t\boldsymbol{\beta}} \log L = -\sum_{i=1}^{N} \mathbf{x}_i \, {}^t\mathbf{x}_i t_i^a \exp({}^t\mathbf{x}_i \boldsymbol{\beta})$$

これらを用いてニュートン・ラフソン法でパラメータの推定を行う。

$$\boldsymbol{\beta}'^{(m+1)} = \boldsymbol{\beta}'^{(m)} + (\mathfrak{I}^{(m)})^{-1} \mathbf{U}^{(m)}$$

ここに右肩の添え字はニュートン・ラフソン法のループの段階を表している。

求められたパラメータを使って、個人の予想寿命を以下のように求めることができる。

$$\mu \equiv E[T] = b\,\Gamma\left(1 + 1/a\right) = \exp(-{}^t\mathbf{x}\boldsymbol{\beta}/a)\Gamma\left(1 + 1/a\right)$$

この値を実際の寿命と比較することで相関係数等を求めることもできる。

3）混合 Weibull ハザードモデル

K 種混合 Weibull ハザードモデルでは以下を仮定する。

$$f(t) = \sum_{k=1}^{K} \pi_k f_k(t) = \sum_{k=1}^{K} \pi_k a_k t^{a_k - 1} \exp({}^t\mathbf{x}\boldsymbol{\beta} + \gamma_k) \exp(-t^{a_k} \exp({}^t\mathbf{x}\boldsymbol{\beta} + \gamma_k))$$

$$S(t) = \sum_{k=1}^{K} \pi_k S_k(t) = \sum_{k=1}^{K} \pi_k \exp(-t^{a_k} \exp({}^t\mathbf{x}\boldsymbol{\beta} + \gamma_k))$$

通常の Weibull 分と比較すると、ここでは以下を仮定している。

$$b_k^{-a_k} = e^{\beta_k} \to \exp({}^t\mathbf{x}\boldsymbol{\beta} + \gamma_k) \quad （\beta_k \to {}^t\mathbf{x}\boldsymbol{\beta} + \gamma_k \equiv \sum_{i=1}^{p} x_i \beta_i + \gamma_k）$$

これより、$b_k \to \exp\left[-({}^t\mathbf{x}\boldsymbol{\beta} + \gamma_k) / a_k \right]$ であるから、

$$\mu \equiv E[T] = \sum_{k=1}^{K} \pi_k b_k \Gamma\left(1 + 1/a_k\right)$$

となる。連結関数については、以下の関数の逆関数である。

$$\mu = \sum_{k=1}^{K} \pi_k \exp\left[-({}^t\mathbf{x}\boldsymbol{\beta} + \gamma_k)/a_k\right]\Gamma\left(1+1/a_k\right)$$
$$= \sum_{k=1}^{K} \pi_k \exp\left[-(\eta + \gamma_k)/a_k\right]\Gamma\left(1+1/a_k\right)$$

混合 Weibull 分布の対数尤度は以下となる。

$$\log L(\mathbf{a},\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\pi}) \geq \sum_{i=1}^{N}\sum_{k=1}^{K} q_k^{(i)}\Big[\delta_i\left(\log a_k + (a_k-1)\log t_i + {}^t\mathbf{x}_i\boldsymbol{\beta} + \gamma_k\right)$$
$$-t_i^{a_k}\exp({}^t\mathbf{x}_i\boldsymbol{\beta} + \gamma_k) + \log \pi_k - \log q_k^{(i)}\Big]$$

これより、群 k への帰属度は以下となる。

$$q_k^{(i)} = \frac{\pi_k a_k t_i^{a_k-1}\exp({}^t\mathbf{x}\boldsymbol{\beta}+\gamma_k)\exp(-t_i^{a_k}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_k))}{\sum_{k=1}^{K}\pi_k a_k t_i^{a_k-1}\exp({}^t\mathbf{x}\boldsymbol{\beta}+\gamma_k)\exp(-t_i^{a_k}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_k))} \qquad \text{for } \delta_i = 1$$

$$q_k^{(i)} = \frac{\pi_k \exp(-t_i^{a_k}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_k))}{\sum_{k=1}^{K}\pi_k \exp(-t_i^{a_k}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_k))} \qquad \text{for } \delta_i = 0$$

ここで、$b_k^{-a_k} \to \exp({}^t\mathbf{x}\boldsymbol{\beta}+\gamma_k)$ （$b_k \to \exp\left[-({}^t\mathbf{x}\boldsymbol{\beta}+\gamma_k)/a_k\right]$）としている。

対数尤度を微分してスコアベクトル $\mathbf{U}$ と情報行列 $\mathfrak{I}$ を求めると以下となる。

$$\boldsymbol{\beta}' = \begin{pmatrix}\mathbf{a}\\ \boldsymbol{\gamma}\\ \boldsymbol{\beta}\end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix}\partial\log L/\partial\mathbf{a}\\ \partial\log L/\partial\boldsymbol{\gamma}\\ \partial\log L/\partial\boldsymbol{\beta}\end{pmatrix},$$

$$\mathfrak{I} = -\begin{pmatrix}\partial^2\log L/\partial\mathbf{a}\partial^t\mathbf{a} & \partial^2\log L/\partial\mathbf{a}\partial^t\boldsymbol{\gamma} & \partial^2\log L/\partial\mathbf{a}\partial^t\boldsymbol{\beta}\\ \partial^2\log L/\partial\boldsymbol{\gamma}\partial^t\mathbf{a} & \partial^2\log L/\partial\boldsymbol{\gamma}\partial^t\boldsymbol{\gamma} & \partial^2\log L/\partial\boldsymbol{\gamma}\partial^t\boldsymbol{\beta}\\ \partial^2\log L/\partial\boldsymbol{\beta}\partial^t\mathbf{a} & \partial^2\log L/\partial\boldsymbol{\beta}\partial^t\boldsymbol{\gamma} & \partial^2\log L/\partial\boldsymbol{\beta}\partial^t\boldsymbol{\beta}\end{pmatrix}$$

ここに、

$$\frac{\partial}{\partial a_j}\log L = \sum_{i=1}^{N} q_j^{(i)}\left[\delta_i\left(1/a_j + \log t_i\right) - \log t_i\, t_i^{a_j}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_j)\right]$$

$$\frac{\partial}{\partial \gamma_j}\log L = \sum_{i=1}^{N} q_j^{(i)}\left[\delta_i - t_i^{a_j}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_j)\right]$$

$$\frac{\partial}{\partial \boldsymbol{\beta}}\log L = \sum_{i=1}^{N}\sum_{k=1}^{K} q_k^{(i)}\left[\delta_i\mathbf{x}_i - \mathbf{x}_i t_i^{a_k}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_j)\right]$$

$$\frac{\partial^2}{\partial a_j \partial a_k}\log L = -\delta_{jk}\sum_{i=1}^{N} q_j^{(i)}\left[\delta_i / a_j^2 + (\log t_i)^2 t_i^{a_j}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_j)\right]$$

$$\frac{\partial^2}{\partial a_j \partial \gamma_k}\log L = -\delta_{jk}\sum_{i=1}^{N} q_j^{(i)}\log t_i\, t_i^{a_j}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_j)$$

$$\frac{\partial^2}{\partial a_j \partial \boldsymbol{\beta}}\log L = -\sum_{i=1}^{N} q_j^{(i)}\log t_i\, \mathbf{x}_i t_i^{a_j}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_j)$$

$$\frac{\partial^2}{\partial \gamma_j \partial \gamma_k}\log L = -\delta_{jk}\sum_{i=1}^{N} q_j^{(i)} t_i^{a_j}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_j)$$

$$\frac{\partial^2}{\partial \gamma_j \partial \boldsymbol{\beta}}\log L = -\sum_{i=1}^{N} q_j^{(i)}\mathbf{x}_i t_i^{a_j}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_j)$$

$$\frac{\partial^2}{\partial \boldsymbol{\beta} \partial {}^t\boldsymbol{\beta}}\log L = -\sum_{i=1}^{N}\sum_{k=1}^{K} q_k^{(i)}\mathbf{x}_i\, {}^t\mathbf{x}_i t_i^{a_k}\exp({}^t\mathbf{x}_i\boldsymbol{\beta}+\gamma_k)$$

これらを用いてニュートン・ラフソン法でパラメータの推定を行う。

$$\boldsymbol{\beta}'^{(m+1)} = \boldsymbol{\beta}'^{(m)} + (\Im^{(m)})^{-1}\mathbf{U}^{(m)}$$

ここに右肩の添え字はニュートン・ラフソン法のループの段階を表している。

3.　プログラムの利用法

メニュー［分析－多変量解析他－生存時間分析］を選択すると、図１のような分析実行メニューが表示される。

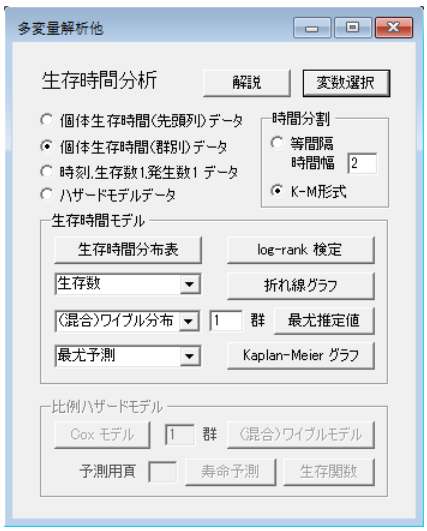
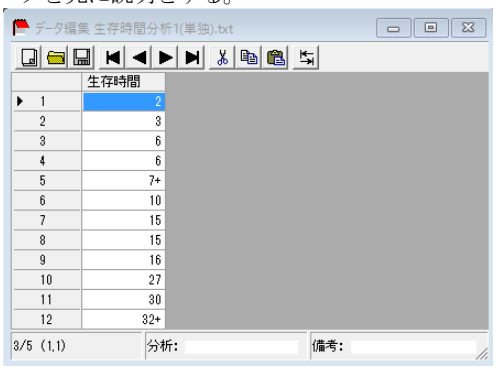


図１　分析実行メニュー

この分析のデータ形式は大きく分けて 3 種類ある。 1 つは個体の生存時間を元にしたデータで、これは先頭列で分類される形式とすでに群別に並べられている形式に分けられる。これらの形式は基本統計のデータ形式に類似している。次に、すでに生存時間分布表に近い形式になっているデータである。これは、観測時刻、その時点での生存個体数、その時点より後で次の時点までに死亡する期間発生数が、すでに表の形式になっているデータである。変数としては、時刻，生存個体数 1，期間発生数 1，生存個体数 2，期間発生数 2，… のようになっており、生存個体数と期間発生数は複数組入力が可能である。但し、これは実際の処理では、あまり使われることがないと思われる。最後は、ハザードモデルデータで、重回帰分析などと同様の形式である。最初と最後の形式で、通常のデータと異なる部分は、観測の打ち切りデータが含まれる点である。打ち切りデータは、観測を打ち切られた時点の数値の後ろに+記号を付けて表す。観測が打ち切られた際の扱いは、生存時間分布表や Kaplan-Meier 推定においては、生存数から打ち切られたデータ数の半分を引いて、死亡リスクに晒されたデータ数として処理している[1]。

最初に図 2 の単独データを元に説明をする。



	生存時間
1	2
2	3
3	6
4	6
5	7+
6	10
7	15
8	15
9	16
10	27
11	30
12	32+

図 2　単独データ（生存時間分析 1(単独).txt 3 頁目）

ここに、データのファイル名はホームページ[4]上にあるサンプル内のファイル名である。このデータでは、2 個体が観測を打ち切られている。

「個体生存時間(群別)データ」ラジオボタンを選択し、変数選択を実行して、「生存時間分布表」ボタンをクリックすると図 3 のような結果が表示される。

生存時間分布表

	値<T	T<=値	間隔	生存数	期間発生数	打ち切り数	リスク数	期間発生率	期間生存率	生存関数	標準誤差	生存時間	密度関数	ハザード	累積ハザード
1	0.0	2.0	2.0	12	1	0	12.0	0.0833	0.9167	1.0000		2.0000	0.0417	0.0417	0.0000
2	2.0	3.0	1.0	11	1	0	11.0	0.0909	0.9091	0.9167	0.0870	0.9167	0.0833	0.0909	0.0870
3	3.0	6.0	3.0	10	2	0	10.0	0.2000	0.8000	0.8333	0.1183	2.5000	0.0556	0.0667	0.1823
4	6.0	7.0	1.0	8	0	1	7.5	0.0000	1.0000	0.6667	0.1701	0.6667	0.0000	0.0000	0.4055
5	7.0	10.0	3.0	7	1	0	7.0	0.1429	0.8571	0.6667	0.1361	2.0000	0.0317	0.0476	0.4055
6	10.0	15.0	5.0	6	2	0	6.0	0.3333	0.6667	0.5714	0.1706	2.8571	0.0381	0.0667	0.5596
7	15.0	16.0	1.0	4	1	0	4.0	0.2500	0.7500	0.3810	0.2204	0.3810	0.0952	0.2500	0.9651
8	16.0	27.0	11.0	3	1	0	3.0	0.3333	0.6667	0.2857	0.1835	3.1429	0.0087	0.0303	1.2528
9	27.0	30.0	3.0	2	1	0	2.0	0.5000	0.5000	0.1905	0.1804	0.5714	0.0317	0.1667	1.6592
10	30.0	32.0	2.0	1	0	1	0.5	0.0000	1.0000	0.0952	0.1806	0.1905	0.0000	0.0000	2.3514
11	32.0			0						0.0952					

図 3　生存時間分布表結果

図 3 では、様々な指標が区切られた時点毎に表示されている。ここで特に大切な指標は、「生存関数」と「ハザード」である。これらはそれぞれ、その時点まで生存している確率とその時点での死亡の危険率の意味を持つ。

図 3 の生存時間分布表の中で、生存数、生存関数、ハザード関数、累積ハザード関数については、コンボボックスで設定して、「折れ線グラフ」ボタンをクリックすると表示される。ここでは生存関数とハザード関数についてのグラフを図 4a と図 4b に示す。

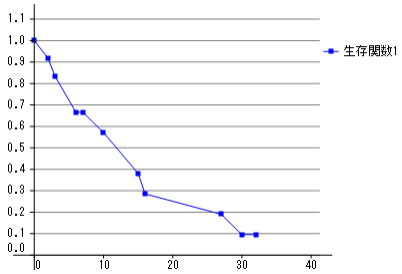


図 4a　生存関数

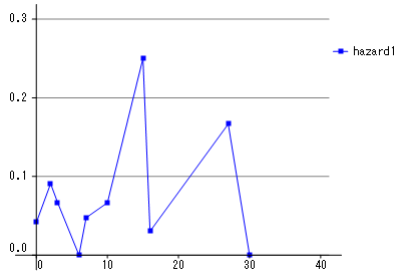


図 4b　ハザード関数

また、同じコンボボックスで「指数分布確認」または「Weibull 分布確認」を選択すると、図 5a と図 5b のような図が表示される。

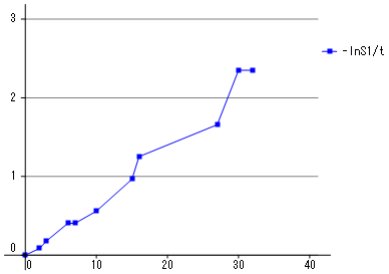


図 5a　生存関数確認

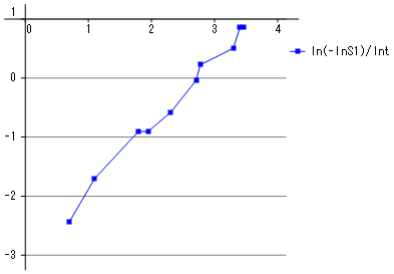


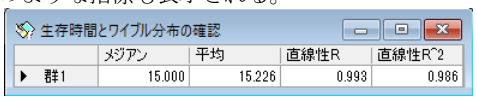
図 5b　ハザード関数確認

生存時間が指数分布または Weibull 分布に従うならば、それぞれの生存関数の時間依存性からこの点列は直線状に並ぶ。指数分布は Weibull 分布の特殊な場合であるので、指数分布が成り立つ場合は Weibull 分布も成り立つ。但し、Weibull 分布の場合の横軸は時間の対数である。

指数分布または Weibull 分布の確認の場合、「折れ線グラフ」をクリックすると、上図と共に分布の当てはまりの良さを示す、図 6a や図 6b のような指標も表示される。

	メジアン	平均	直線性R	直線性R^2
群1	15.000	15.226	0.992	0.985

図 6a　指数分布の指標



	メジアン	平均	直線性R	直線性R^2
群1	15.000	15.226	0.993	0.986

図 6b　Weibull 分布の指標

生存関数の Kaplan-Meier 推定のグラフは、「Kaplan-Meier グラフ」ボタンをクリックして表示される。その際、左のコンボボックスで指定して、指数分布または Weibull 分布の予想曲線を描くこともできる。予想曲線のないグラフと、Weibull 分布の予想曲線を付けて描いたグラフを図 7a と図 7b に示す。

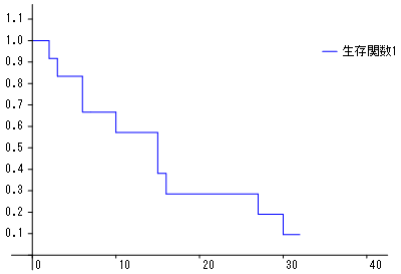


図 7a　Kaplan-Meier 生存関数グラフ

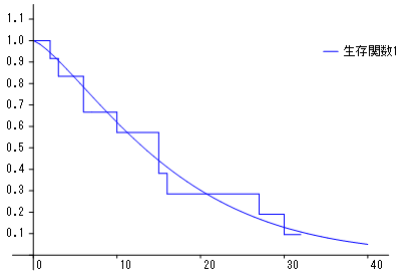


図 7b　予想曲線付き Kaplan-Meier グラフ

これらの予想曲線では最小 2 乗法によるものと最尤法によるものとが選択できる。上図は最尤法によるものである。

また、予想曲線は混合指数分布や混合 Weibull 分布についても表示することができる。その際は分布を選んだコンボボックスの右のテキストボックスで混合する数を指定する。図 8 に 2 種の混合 Weibull 分布による予想曲線を付けた Kaplan-Meier グラフを表示する。サンプルでは 2 つの時期に危険度が高くなっている。

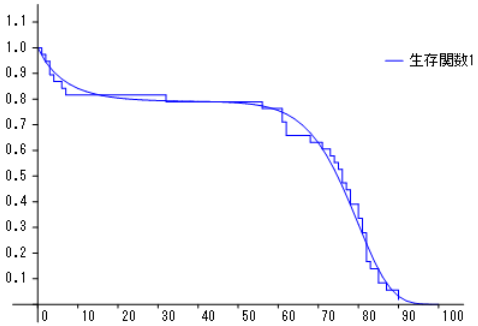


図 8　2 種混合分布による予測（生存時間分析 1(単独).txt 8 頁目）

このパラメータの値については、上と同じ設定で「最尤推定値」ボタンをクリックすると、図 9 のように表示される。

混合ワイブル分布推定結果

	推定値	標準偏差	5%下限	5%上限
▶ 生存時間	R	0.993	R^2	0.986
出現確率1	0.790			
a1	10.545	1.623	7.364	13.726
b1=exp(-β/a)	80.564			
β1	-46.283	7.171	-60.339	-32.227
出現確率2	0.210			
a2	0.937	0.233	0.480	1.394
b2=exp(-β/a)	6.964			
β2	-1.819	0.681	-3.154	-0.483

図 9　2 種混合 Weibull 予測（生存時間分析 1(単独).txt 8 頁目）

ここでは表示されていないが、混合がない場合には、右端に最小 2 乗推定による推定値も表示される。

複数群の生存時間分布表は、先頭列で群分けデータ（生存時間分析 2(2 群比較).txt）または群別データを元に図 10 のように縦に並べて表示される。

生存時間分布表

	値<T	T<=値	間隔	生存数	期間発生数	打ち切り数	リスク数	期間発生率	期間生存率	生存関数	標準誤差	生存時間	密度関数	ハザード	累積ハザード
1	0.0	1.0	1.0	12	0	0	12.0	0.0000	1.0000	1.0000		1.0000	0.0000	0.0000	0.0000
2	1.0	2.0	1.0	12	1	0	12.0	0.0833	0.9167	1.0000	0.0000	1.0000	0.0833	0.0833	0.0000
3	2.0	3.0	1.0	11	1	0	11.0	0.0909	0.9091	0.9167	0.0870	0.9167	0.0833	0.0909	0.0870
4	3.0	6.0	3.0	10	2	0	10.0	0.2000	0.8000	0.8333	0.1183	2.5000	0.0556	0.0667	0.1823
5	6.0	7.0	1.0	8	1	0	8.0	0.1250	0.8750	0.6667	0.1701	0.6667	0.0833	0.1250	0.4055
6	7.0	9.0	2.0	7	0	0	7.0	0.0000	1.0000	0.5833	0.1627	1.1667	0.0000	0.0000	0.5390
7	9.0	10.0	1.0	7	1	0	7.0	0.1429	0.8571	0.5833	0.1429	0.5833	0.0833	0.1429	0.5390
8	10.0	15.0	5.0	6	2	0	6.0	0.3333	0.6667	0.5000	0.1684	2.5000	0.0333	0.0667	0.6931
9	15.0	16.0	1.0	4	1	0	4.0	0.2500	0.7500	0.3333	0.2041	0.3333	0.0833	0.2500	1.0986
10	16.0	22.0	6.0	3	0	0	3.0	0.0000	1.0000	0.2500	0.1667	1.5000	0.0000	0.0000	1.3863
11	22.0	27.0	5.0	3	1	0	3.0	0.3333	0.6667	0.2500	0.1250	1.2500	0.0167	0.0667	1.3863
12	27.0	30.0	3.0	2	1	0	2.0	0.5000	0.5000	0.1667	0.1614	0.5000	0.0278	0.1667	1.7918
13	30.0	32.0	2.0	1	1	0	1.0	1.0000	0.0000	0.0000	0.1596	0.0000	0.0000	0.5000	0.0000
14	32.0			0						0.0000					
1	0.0	1.0	1.0	9	4	0	9.0	0.4444	0.5556	1.0000		1.0000	0.4444	0.4444	0.0000
2	1.0	2.0	1.0	5	1	0	5.0	0.2000	0.8000	0.5556	0.2981	0.5556	0.1111	0.2000	0.5878
3	2.0	3.0	1.0	4	2	0	4.0	0.5000	0.5000	0.4444	0.2070	0.4444	0.2222	0.5000	0.8109
4	3.0	6.0	3.0	2	0	0	2.0	0.0000	1.0000	0.2222	0.2772	0.6667	0.0000	0.0000	1.5041
5	6.0	7.0	1.0	2	0	0	2.0	0.0000	1.0000	0.2222	0.1386	0.2222	0.0000	0.0000	1.5041
6	7.0	9.0	2.0	2	1	0	2.0	0.5000	0.5000	0.2222	0.1386	0.4444	0.0556	0.2500	1.5041
7	9.0	10.0	1.0	1	0	0	1.0	0.0000	1.0000	0.1111	0.2095	0.1111	0.0000	0.0000	2.1972
8	10.0	15.0	5.0	1	0	0	1.0	0.0000	1.0000	0.1111	0.1049	0.5556	0.0000	0.0000	2.1972
9	15.0	16.0	1.0	1	0	0	1.0	0.0000	1.0000	0.1111	0.1049	0.1111	0.0000	0.0000	2.1972
10	16.0	22.0	6.0	1	1	0	1.0	1.0000	0.0000	0.0000	0.1049	0.0000	0.0000	0.1667	0.0000
11	22.0	27.0	5.0	0						0.0000					
12	27.0	30.0	3.0												
13	30.0	32.0	2.0												
14	32.0														

図 10　2 群の生存時間分布表

これ以外に、もっと群の違いを比較できる方法を考えて行きたい。

複数群の生存関数と Kaplan-Meire 生存関数グラフを図 11 と図 12 に示す。

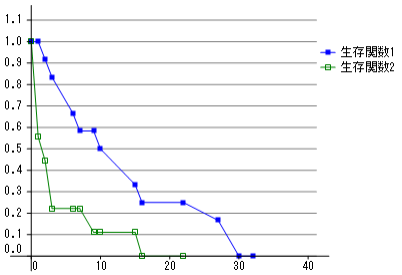


図 11　2 種類の生存関数グラフ

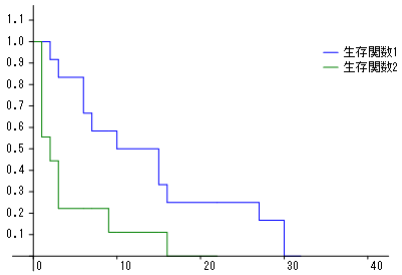


図 12　2 種類の Kaplan-Meier グラフ

複数群の生存関数間の差の log-rank 検定結果は、「log-rank 検定」ボタンをクリックすると図 13 のように表示される。

log-rank検定

log-rank検定結果

χ2値　5.0100
自由度　1
確率　0.0252

図 13　log-rank 検定結果

最後に、比例ハザードモデルの分析結果について示しておく。データは図 14 のような重回帰分析などと同じデータ形式である。

データ編集 生存時間分析3(ハザードモデル).txt

	寿命	身長	体重	要因
1	80	170	55	0
2	78	162	48	0
3	62	167	98	0
4	82	181	52	0
5	77	181	80	0
6	90+	157	44	0
7	75	160	67	0
8	80	172	73	0
9	68	173	85	0
10	85	164	73	0

2/4 (1,1)　分析:　備考:

図 14　比例ハザードモデルデータ（生存時間分析 3(ハザードモデル).txt）

ハザードモデルでは Cox 比例ハザードモデルと Weibull 比例ハザードモデルを組み込んでいる。ハザード関数について、2 つのモデルとも、前章でのべたように、以下の形を仮定する。

$$\lambda(t\,|\,\mathbf{x},\boldsymbol{\beta}) = \lambda_0(t)\exp({}^t\mathbf{x}\boldsymbol{\beta}) \quad \text{ここに、} {}^t\mathbf{x}\boldsymbol{\beta} = \sum_{i=1}^{p} x_i\beta_i + \beta_0$$

Cox 比例ハザードモデルは $\lambda_0(t)$ や β_0 の推定は行わないが、分布の形に依存しない利点がある。Weibull ハザードモデルでは、時間部分に Weibull 分布を仮定し、その 1 つのパラメータを説明変数で推定するという一般化線形モデルの形式を採用している。

$$\lambda(t\,|\,\mathbf{x},\boldsymbol{\beta}) = (a/b)\left(t/b\right)^{a-1} = at^{a-1}b^{-a} = at^{a-1}\exp({}^t\mathbf{x}\boldsymbol{\beta})$$

「Cox モデル」ボタンをクリックした結果を図 15 に、「Weibull モデル」ボタンをクリックした結果を図 16 に示す。

Cox比例ハザードモデル結果

	偏回帰係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
身長	-0.0038	0.0227	0.8669	-0.0483	0.0407	0.9962
体重	0.0294	0.0133	0.0272	0.0033	0.0555	1.0299
要因	2.4581	0.4826	0.0000	1.5123	3.4039	11.6825

図 15　Cox 比例ハザードモデル結果

ワイブル比例ハザードモデル結果

	偏回帰係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
身長	-0.0040	0.0216	0.8547	-0.0464	0.0385	0.9960
体重	0.0211	0.0119	0.0773	-0.0023	0.0445	1.0213
要因	2.0780	0.3916	0.0000	1.3104	2.8455	7.9884
切片	-21.4299	4.3297	0.0000	-29.9161	-12.9437	4.933E-10
a	4.7354	0.6323	0.0000	3.4961	5.9747	

図 16　Weibull 比例ハザードモデル

最後に Weibull 比例ハザードモデルが予想する生存時間の平均値と実際の観測値との比較を行ってみる。「予測用頁」テキストボックスを空欄のまま、「寿命予測」ボタンをクリックすると図 17a と図 17b の結果が示される。

Weibull寿命予測

	寿命	寿命予測	残差
29	81	76.199	4.801
30	83	74.542	8.458
31	23	47.878	−24.878
32	32	46.150	−14.150
33	43	45.474	−2.474
34	48	48.642	−0.642
35	34	44.575	−10.575
36	25	46.422	−21.422
37	40	47.904	−7.904
38	19	43.398	−24.398
39	32	45.818	−13.818
40	26	43.110	−17.110
R	0.869	R^2	0.755

図 17a　寿命予測図

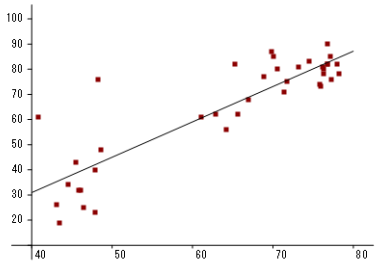


図 17b　実測/予測散布図

これには非打ち切り（死亡）データのみが用いられている。また、寿命予測の結果の最後に、予測値と実測値の相関係数の値とその 2 乗の値を表示している。寿命予測には、予測したいデータを別の頁に作っておき、「予測用頁」テキストボックスにその頁番号を入力して、「寿命予測」ボタンをクリックする方法もある。

このWeibullハザードモデルと混合Weibull分布のKaplan-Meier推定とを比較してみる。「予測用頁」テキストボックスを空欄のまま、「生存関数」ボタンをクリックし、各個体の生存関数を描画すると図 18 のようになる。また Weibull 分布を使った Kaplan-Meier 推定は図 19 のようになる。

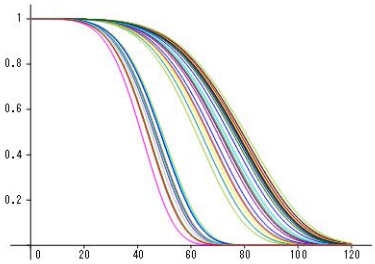


図 18　各個体の生存関数

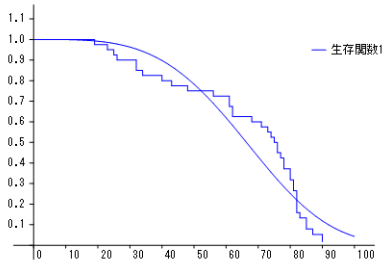


図 19　Weibull 分布による推定

Kaplan-Meier 推定はあまり適合しておらず、各個体の生存関数も当然のことであるが、寿命が長いか短いかしか分からず、あまり面白みがない。

次に、このモデルに例えば 2 種混合 Weibull ハザードモデルを適用してみよう。この場合、比例ハザードモデルの中の「群」テキストボックスに「2」を入れて、「(混合) Weibull モデル」ボタンをクリックする。図 20 に結果を示す。

混合ワイブル比例ハザードモデル結果

	偏回帰係数	標準偏差	両側確率	2.5%下限	2.5%上限	EXP(b)
▶ 身長	0.0274	0.0206	0.1830	−0.0129	0.0678	1.0278
体重	0.0526	0.0132	0.0001	0.0266	0.0785	1.0540
要因	5.2765	0.7407	0.0000	3.8248	6.7283	195.6922
出現確率1	0.2497					
a1	29.7025	5.6218		18.6837	40.7212	
γ1	−140.1803	25.9021		−190.9484	−89.4123	
出現確率2	0.7503					
a2	7.3947	1.0264		5.3829	9.4065	
γ2	−40.1365	6.2861		−52.4574	−27.8157	

図 20　混合 Weibull ハザードモデル（生存時間分析 3(ハザードモデル).txt 2 頁目）

このモデルによる実測・予測値と重相関係数 R の値、及びそのグラフを表示するには、「予測用頁」テキストボックスを空欄のまま、「寿命予測」ボタンをクリックする。結果は図 21 のようになる。

Weibull寿命予測

	寿命	寿命予測	残差
29	81	77.650	3.350
30	83	77.042	5.958
31	23	43.946	-20.946
32	32	44.003	-12.003
33	43	44.366	-1.366
34	48	46.858	1.142
35	34	39.913	-5.913
36	25	43.212	-18.212
37	40	45.561	-5.561
38	19	40.963	-21.963
39	32	43.342	-11.342
40	26	39.562	-13.562
R	0.855	R^2	0.731

図 21a　寿命予測図

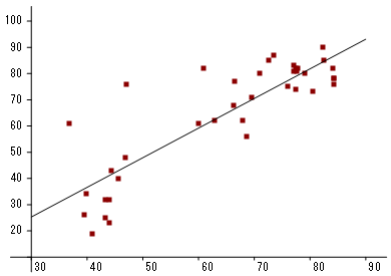


図 21b　実測/予測散布図

このモデルと 2 種混合 Weibull 分布の Kaplan-Meier 推定とを比較してみる。「予測用頁」テキストボックスを空欄のまま、「生存関数」ボタンをクリックし、各個体の生存関数を描画すると図 22 のようになる。また 2 種混合 Weibull 分布を使った Kaplan-Meier 推定は図 23 のようになる。

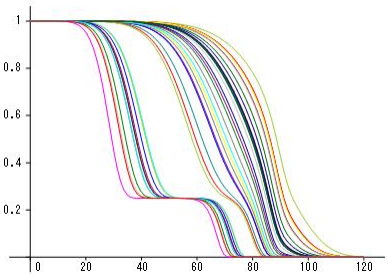


図 22　各個体の生存関数

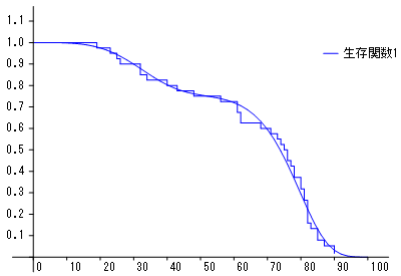


図 23　混合 Weibull 分布による推定

図 20 で生存関数が途中で折れ曲がっている個体は、寿命に及ぼす影響が大きいと思われる、データ中の「要因」因子が「1」の個体である。このようにして見ると、「要因」因子が「1」の個体と「0」の個体で生存関数が大きく異なっていることが分かる。

それでは、図 22 の個体の生存関数グラフとハザードモデルを考えない図 23 の混合 Weibull 分布モデルのグラフとの関係はどのようになっているのであろうか。図 23 のすべての曲線の時間ごとの平均を取ると図 24 のようになり、図 23 のハザードモデルを考えない形に非常に近くなることが分かる。

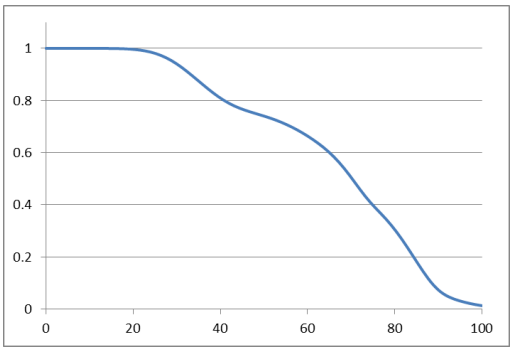


図 24　各個体の生存関数の平均

この結果はグラフのデータ出力機能を利用し、結果を Excel に貼り付けて平均を計算し描画したものである。

4.　おわりに

この論文では、Kaplan-Meier 推定法による生存分析表や生存関数グラフ、log-rank 検定、分布の最尤推定、Cox の比例ハザードモデル、Weibull ハザードモデルなど、生存時間分析法とそのプログラムについて解説して来た。特徴としては、混合 Weibull 分布をハザードモデルに適用し、使い易いプログラムにした部分である。これによって、寿命に影響を及ぼす変数がある場合、生存関数グラフが大きな変更を受ける可能性が見えてきた。視覚的なグラフにより危険性が示されることは、ハザードや平均寿命の数値だけによる表示に比べて注意を促しやすい。これがどの程度示されるのか、今後詳しく調べてみる必要がある。

このプログラムでは、生存時間分布表や生存関数の推定方法を、最もよく利用される Kaplan-Meier 推定法に限定している。また、それに当てはめる分布も、指数分布と Weibull 分布及び、それらの混合分布だけであり、他の分布は考えていない。今後これらの点については必要に応じて追加して行く予定である。また、生存関数、ハザード、推定されるパラメータの 95%信頼区間についても、すべてを取り上げていない。この問題についても今後の課題として残っている。改良すべき点を、段階を追って改良し、使い易いプログラムとして行きたい。

参考文献

[1] 打波守, Excel で学ぶ生存時間解析, オーム社, 2005.
[2] 柳井晴夫, 高木廣文編著, 多変量解析ハンドブック, 現代数学社, 1986.
[3] Annete J. Dobson, 田中豊他訳, 一般化線形モデル入門 原著第 2 版, 共立出版, 2008.
[4] ホームページ http://www.heisei-u.ac.jp/ba/fukui/analysis.html 内のサンプルデータ Samples.zip 内のファイル

Multi-purpose Program for Social System Analysis 31
- Survival Analysis -

Masayasu FUKUI[*1] and Xiaona WU[*2]

[*1] *Department of Business Administration, Faculty of Business Administration, Fukuyama Heisei University*

[*2] *Division of Business Information Study of Graduate School of Business Administration, Fukuyama Heisei University*

Abstract: We have been constructing a unified program on the social system analysis for the purpose of education. This time, we report about the program of survival analysis which is a method to predict mortality rate and survival probability distribution from sample including abort data. The program includes survival time table, Kaplan-Meier graph including prediction curve, Cox proportional hazard model, Weibull hazard model by maximum likelihood method and mixed Weibull hazard model.

Keywords: College Analysis, multivariate analysis, survival analysis, compound Weibull distribution, Cox proportional hazard model, Weibull hazard model, EM Algorithm

URL: http://www.heisei-u.ac.jp/ba/fukui/

福山平成大学経営学部紀要
第 13 号（2017），163－175 頁

フィリピン小学校英語教科書の分析
－１、２年教科書の量的分析－

渡辺 清美

福山平成大学経営学部経営学科

要旨：ESL環境にあるフィリピンの小学校１、２年の英語教科書をデータ化し、計量的分析を行った。総語数、異語数、リーダビリティ、特徴語、JACET8000のレベル、の分析を行った。比較対象データとして、中国の小学校英語教科書、日本の中高の英語教科書、そしてネイティブの英文データとしてネット上の英文ニュースのデータを使用して、分析した。分析の結果、フィリピンの教科書は、EFL環境の中国や日本の教科書と比べると語数や語の難易度において上回っているが、ネイティブの英文よりは低いレベルに統制されていること、また、内容的にも学習者の身近な題材を取り上げるなどの工夫がなされていることがわかった。

キーワード：量的研究、ESL環境、英語教科書研究

1. はじめに

フィリピンは東南アジア（ASEAN加盟国）では唯一の英語を公用語とする国である。英語学習の観点では、ESL (English as a Second Language)の環境であり、日本のようなEFL (English as a Foreign Language)の環境とは異なる。公用語は英語と標準語（タガログ語）であるが、１７０ほどの言語が母国語として使用されている多言語国家であり、子どもたちは小学校までは母国語の環境で育ち、正式に英語を学ぶのは小学校からである。フィリピンでは、英語が日常生活で使われている環境で育った子どもがいる一方、母国語を日常的に使用し、ほとんど英語に触れる機会がなく育った子どももいるという、様々な英語環境で育った子どもが小学校で英語を学ぶというのが実情である。

日本では、2020年度から英語の小学校３、４年での必修化、５、６年の教科化がなされる予定である。このような早期英語教育の実施を目前に控えて、上述のような環境にあるフィリピンの小学校英語テキストの研究を通して、日本の初等教育における英語学習への示唆を求めることが本研究の目的である。

1.1 フィリピンの小学校英語教育の概要

フィリピンの公用語はタガログ語を基礎にした「フィリピノ」と英語であるが、それ以外に１７０ほどの言語が母国語として使用されている。ルソン島ではタガログ語、ミンダナオ島ではビサヤ語、など島単位で異なる言語が使われている場合が多い。

フィリピンの教育省が作成しているPolicy Guidelinesによると、1年から１０年次[1]までの言語教育の目標は、母国語、フィリピノ、英語の３言語において、「読む」「聞く」「話す」「書く」の運用能力を身につけることとしている。この目標に従って、３年までは母国語で全ての教科が教えられ、４年以降は、公用語の英語とフィリピノとの共用で教科が教えられている。[2]

言語教育に関しては、母国語とフィリピノが教科として教えられており、母国語は1年前期・後期の１年間を通して、毎日５０分の時間で教えられている。ただし教科としての母国語学習は２年までである。フィリピノも１年の前期から毎日５０分の時間単位で６年生まで、７年生から１０年生までは週4時間、教えられている。英語に関しては、１年の後期から始まり、毎日３０分の時間単位で教えられている。１年では特に発話能力向上に力点を置いた授業がなされている。２年次からはフィリピノと同様に６年次まで毎日５０分の時間が取られ、７年から１０年までは週4時間が確保されている。[3]

1.2 英語教科書の内容

1.2.1 １年生

１年生のテキストは、4つのUnitに分けられており、Unitはそれぞれ５つの単元 (Lesson)によって構成されている。また、Unitの終わりにはUnit Testが置かれている。各単元は、最初に一つのテーマに基づいた会話文のセット（Reading Dialogue）が置かれ、後に学習目標の構文などが “REMEMBER” というタイトルで置かれている。次に “GRAMMAR POINT” というタイトルで、学習目標の文法が置かれ、 “Exercise” で文法の練習をするようになっている。次に、 “TUNE IN” というタイトルで、付属のCDにある音楽を聴き、歌う活動が続く。そして、さらに “SAY IT RIGHT” で、学習目標の構文を使った文章の発話練習をする。その後に、 “WRITE IT DOWN” で書く練習が行なわれる。次に “SPIN OFF” というタイトルで別なリーディング用の文章が置かれ、最後に “LESSON CHCEK UP” として、単元の復習を行うようになっている。練習問題などはテキスト自体に書き込むように十分なスペースが取られている。巻末にはGlossaryが置かれ、action word, rhyme, paragraph, punctuation mark などの文法用語や dialogue, fable, ownership, Tausug（ミンダナオに住む部族）などの１年生児童には馴染みの薄いと思われる語彙が記載されている。[4]

---

[1] “Guidelines on the Implementation of Grades 1 to 10”では、小学校ー中学校の区分ではなく、１～１０年の学年単位で記述されているため、本研究もこの学年単位の区分に従った。

[2] Ibid.,

[3] Ibid.,

[4] 教育省の Education Policy では、１年次は後期から英語が教えられるように指導されているが、テキストを見る限り、１年間分の内容量に相当すると思われる。

Unitごとの主な文法事項は以下の通りである。

Unit 1	be動詞、主格代名詞、現在動詞＋前置詞、命令文 現在進行形
Unit 2	所有格代名詞、三人称単数現在動詞、母音と子音について、 冠詞a, an　複数形名詞、this, that, these, those、未来形will 時間の表現　at noon, in the morning、形容詞の使い方
Unit 3	三人称単数現在動詞、動詞+ing、頻度を表す語every day, yearly 過去形 –ed、過去を表す語 last week, last month 助動詞　can, can't
Unit 4	形容詞の使い方、形を表す語　round, square, rectangular 形容詞の順序 数＋形容詞+名詞 同義語　synonyms　反義語　antonyms when, where, howを使った疑問文 命令文、依頼文、感嘆文

表１

いくつかの文法項目が重複して出ているが、ここから目標事項を反復して学習させる意図が見て取れる。

1.2.2 ２年生

２年生のテキストは、テキスト全体がFirst Quarter, Second Quarter, Third Quarter, Fourth Quarterの４つにまず区分されている。教育省の Education guidelines では前期、後期の２期制が前提であるが、テキストは４期制を前提として書かれている。各Quarterは二つのUnitに分かれ、一つのUnitは、二つないし三つの単元 (Lesson) によって構成されている。Quarter区分の最後にはTestが載せられている。

単元は、まずリーディング用の文章が置かれ、その後に目標文法を学ぶ "Grammar Point" が置かれている。その後に練習問題があり、"Don't Forget!" というタイトルの目標文法の復習箇所がある。その後に "Express Yourself" という主に自分の考えを書く活動をするものと、"Exchange Ideas" という自分の考えを発話する活動が置かれている。１年次テキストと同様に、練習問題やテストはテキストに書き込むように作られている。巻末にはGlossaryが置かれ、主に文法用語（adjectives, common nouns, interrogative pronouns, possessive pronouns等）が説明されている。

Quarterごとの文法事項は以下の通りである。

1st Quarter	名詞の使い方、名詞+be動詞、複数形 (-s, -es)、hasとhave、可算名詞と不可算名詞、wh疑問文 (who, what, where, when, why, How)

2nd Quarter	動詞の使い方、be動詞、過去形 形容詞の使い方、形容詞の順序（数+質+サイズ+形+色）
3rd Quarter	代名詞（単数、複数）、所有格代名詞（単数、複数）、 指示代名詞、疑問詞 (what, who, whose, which) 前置詞 (on, in at, under, above, across, behind)
4th Quarter	sentenceについて：平叙文、疑問文、命令文、感動文など 文章の書き方：段落の作り方、手紙の書き方

表２

表１の１年次の文法事項と比較すると明らかなように、動詞や形容詞など１年次で学んだと同じ文法事項が２年次にも学習目標文法として現れている（表２参照）。これは、文法事項の基本的な知識を理解し、次の段階でさらに深い内容を学ぶ、という段階的な学習方法をとっているためと思われる。また、認知的成長に合わせて内容を掘り下げて行くという合理的な学習方法がとられていると考えられる。

2. 先行研究及び研究方法

2.1 先行研究

本研究と同様の研究方法を用いて英語教科書の分析を行った研究では、渡辺・坂元 (2015)、渡辺・浅井・安部 (2016) や渡辺・浅井・赤瀬 (2016) がある。渡辺・坂元 (2015) と渡辺・浅井・安部 (2016) は、タイの英語教科書（小中学校）を日本の教科書と比較し、量的に分析した研究で、語彙量、語彙難度においてタイの教科書が数値を上回っていることが明らかになった。また、学年が上がるにつれて、総語数 (token)、異語数 (type)、や語の難易度も順次上がっていくのは、タイも日本も同じであるが、日本の教科書の特徴は、中学３年から高校１年にかけて、語数、異語数、また語の難易度がタイ（小学校から中学校）と比べて、大幅に数値が上回っていることが確認された。渡辺・浅井・赤瀬 (2016) では、中国の小学校用英語テキスト（６学年分）と日本の小学校（５、６年）から高校１年の教科書（６学年分）を量的に比較研究したものである。語彙量については、総語数、異語数、TTR (Type/Token Ratio) の分析では、総語数と異語数において中国のテキストが日本の教科書を上回っていることがわかった。また、JACET8000を基準とする語彙レベルの比較では、日本の中学までの教科書よりも中国の小学校テキストの平均レベルが上回っていることがわかった。また、BNC (British National Corpus) をもとにする特徴語の分析では、冠詞や代名詞の出現頻度が日本の教科書と比べて高いことがわかった。これは語彙の出現時期と頻度が教科書として統制されていることを示すものであることがわかった。

2.2 研究方法

本研究は、フィリピンの小学校の英語テキスト（１、２年）をデジタル化し、そのデータをデジタル言語分析ツールによって計量的分析を行った。

まず、スキャンした英語テキストをtextファイル化してコーパスを作成した。次に、そのデータを以下のデジタル言語ツールによって分析を行った。

ツール及び調査内容

1. AntConc	総語数、異語数、Type/Token Ratio 特徴語 対数尤度比 (loglikelihood ratio) 作者：Lawrence Anthony 早稲田大学 http://www.laurenceanthony.net/software/antconc/
2. Ozasa-Fukui Year Level	学習指導要領準拠の中学高校学年別英文リータビリティ測定 作者：小篠敏明（広島大学名誉教授）、福井正康（福山平成大学）
3. WordLevel Checker	JACET8000を基準とする単語難易度レベル分析 作者：染谷泰正　　関西大学 http://someya-net.com/wic/index_j.html

使用したフィリピンの小学校用英語教科書は以下の通りである。

小学校１年

Ribo, L.M., Jocson, J.V., Karaan, A.M., Aguinaldo, L.B., & Noche, C.I. (2010). ACROSS BORDERS THROUGH LANGUAGE, English Series 1, Second Edition, Manila, Philippines: Vibal Publishing House, Inc.

小学校２年

Tanjutco, M.C. (2013). ENGLISH ENCOUNTERS: LANGUAGE 2. Manila, Philippines: REX Book Store.

AntConcを使用してデジタル化した教科書データのtoken（総語数）、type （異語数）とtype/token ratio （異語頻度）を調べた。次に、文科省の学習指導要領に準拠した英文リーダビリティ測定ツールであるOzasa-Fukui Year Levelを使用して、学年ごとの英文の平均難易度を調べた。さらに、AntConcの「特徴語」(keyness words)検索機能を使用して、特徴語を調べた。対象コーパスにはBritish National Corpus (BNC)と日本の教科書のコーパス（中学校１―３年）を使用した。次に、大学英語教育学会が作成した語彙表「JACET8000」を基にしたツールであるWordLevel Checkerを使用して、単語難易度を調べた。JACET8000は頻度順に８つのレベルに分けられており、そのレベルに従って調べた。

３．量的分析の結果と考察

３．１ 総語数、異語数、TTRの分析と考察

AntConcを使用して算出した総語数 (token)、異語数 (type)、TTR等の数値を算出した。

フィリピン小学校	token	type	TTR	R(Guiraud)値
1st year	28605	2612	9.13	17.81
2nd year	21498	2748	12.9	18.74
News corpus	28616	4992	17.4	29.51

表3

表３で明らかなように、総語数においては、１年次テキストが約２万８千語、２年次が約２万１千５百語と１年次が大きく上回っている。これは１年次テキストが読む、書く、聞く、話すという４つのスキルごとにそれぞれ練習問題やタスクを置いていることによるものと思われる。ただし、新語を表す異語数やその総語数における割合を表すTTRの数値では、２年次テキストが上回っており、文のレベルが上がっていることが伺える。

参考として本研究では、他のデータ（コーパス）との比較も行った。まず、ネイティブの英文と比較するために、インターネットの英文ニュース記事を集め、フィリピン英語テキストと同程度の語彙サイズのコーパスを即席で作り、その分析データと比較した。それがNews corpusのデータである。R(Ruiraud)値とはTTRの修正値である。総語数が大きく異なるデータ間で比較をする場合、両者のTTRの数値が必ずしも同等であるとは言えないため、修正する必要があるためである（石川, 2014）。ただし、News corpusは意図的にフィリピンテキストと同程度のサイズにしたので、R値への修正は参考として置いた。News corpusと比較すると２年次テキストのTTR値は５ポイントほど低く、英文のレベルが学習者に合わせて低く統制されていることがわかる。

次に、中国と日本の英語教科書との比較を行った。中国と日本のテキストデータは以下の通りである。

中国小学校	token	type	TTR	R 値（Guiraud 値）
１年	2056	249	12.11	5.49
２年	2364	327	13.83	6.73
３年	7753	810	10.45	9.20
４年	11186	1111	9.93	10.50
５年	16979	1538	9.06	11.80
６年	18581	1664	8.96	12.21

表４

日本	token	type	TTR	R 値（Guiraud 値）
小学校５年 (Hi, friends! 1)	329	94	28.57	5.18
小学校６年 (Hi, friends! 2)	487	164	33.68	7.43
中学校１年 (New Horizon 1)	2137	488	22.84	10.56
中学校２年 (New Horizon 2)	4340	844	19.45	12.81
中学校３年 (New Horizon 3)	4237	952	22.47	14.63

高校１年 (Prominence 1)	9407	1936	20.58	19.96
高校２年 (Prominence 2)	15494	2498	16.12	20.07
高校３年 (Prominence 3)	36238	4619	12.75	24.26

表５

まず、総語数であるが、フィリピンテキストが２万台の語数であるが、中国の小学校は６年次でも１万８千語程度で２万語はない。日本の英語教科書では、高校３年を除くと２万語以上あるテキストはなく、高校２年次でも１万５千語台である。新語を表す異語数でも、フィリピンは１年次が約２千６百語、２年次が２千７百語程度であるのに対し、中国のテキストは６年次でも１千６百語で､日本でも高校１年までは２千語に達していないが、これは、EFL環境のテキストである中国と日本のテキストとESL環境のフィリピンの違いによるものと考えられる。日常生活において英語が使用されている環境と教室内でのみ触れる機会のない環境の違いであろう。ESL環境では英語の習得をできるだけ早く行わなければならないという事情もあると思われる。

TTRについては、日本の数値がフィリピンや中国と比較してかなり高めの数値が算出されている。しかし、修正値であるR値では、高校テキストを除くとほぼ同じである。ネイティブの英文であるNews corpusのR値 (29.51)は、フィリピン、中国、日本よりもかなり高い数値であることがわかる。したがって、３カ国とも、テキストの語彙が学年（年齢）に合わせて統制されているということがわかる。

３．２　英文リーダビリティの測定と考察

Ozasa-Fukui Year Levelは、文科省の学習指導要領に準拠した英文リーダビリティツールである。これを使用して英文を調べるとその英文が中学から高校までの６年間でどの学年レベルにあたるかを算出できる。

表６は分析結果を示している。

Philippine text	OFYL
1st year	3.25
2nd year	3.78
News corpus	6.44

表６

１年次テキストの数値 (3.25)は、テキスト内の英文の平均レベルが日本の学習指導要領準拠の教科書の中学３年次前半の英文難易度に相当することを表している。２年次テキストの3.78は、１年次のよりも少し上の中学３年次後半に相当する。本研究のために作成したネイティブ英文のコーパスであるNews corpusは、6.44を示しているが、これは高校３年の半ば程度の難易度であることを示している。小学校レベルのテキストではあるが、日本

の中学３年次の難易度の英文を使用していることは、実践的な英語学習が必要なESL環境のテキストであることを示していると思われる。

３．３　特徴語の分析と考察

特徴語の分析は、二つのコーパス間における語彙の頻度を比較し、一方において顕著に現れる語を調べる。当該コーパスが対象コーパスとの比較において、どのような言語的・内容的性質を持っているかを調べることができる (石川, 2014)。本研究では、AntConcに付属する特徴語検索機能を使用して調べた。これは対数尤度比 (log-likelyhood ratio)という統計値を基準にして検索する機能である。Blake (2012)は、使用するツールによって算出される対数尤度比の数値が異なることが指摘している。本研究では比較参照した中国と日本のデータもすべてAntConcを使用している。また多くの場合、特徴語リストが膨大になるため、Cabrielatos & Marchi (2011) は上位１００語による比較が一般的としているため、本研究もこの慣例に従った。

今回は、対象コーパスとしてBritish National Corpusと独自に作成したNews corpusの２種類を選んで、特徴語の検索を行った。紙面の関係上、フィリピン教科書の特徴語がより明確に現れたNews corpusを参照対象コーパスとした場合の検索データで分析を行った。

まず小学校１年生のテキストをNews corpusを参照対象として特徴語検索を行った。上位１００語には、英語の文法関連の用語と学校・授業関連の語が多いことがわかった。

英語の文法関連の用語で上位１００語に含まれるのは以下の語である。

sentence(s), word(s), letter, action, describing, ing, telling, grammar

１年生であるため、いわゆる文法用語は出てこないが、“describing” や “telling” などは動詞の種類を示したものである。

学校・授業関連の用語で上位１００語に含まれるのは以下の語である。

read, write, exercise, say, listen(ing), lesson, blank, draw, question(s), grade, dialogue, underline, begin, answer(s), sing, correct, checkup, school, teacher, library

“read, write, say, listen, underline, begin, correct” などは、テキスト内の練習問題で使用される語である。以上の語群だけで上位１００語のおよそ３割を占めることがわかった。

次に、小学校２年のテキストをNews corpusを参照対象として特徴語検索を行った。上位１００語には小学校１年と同様に文法用語と授業や学校関連の用語が多く含まれていることがわかった。

英語の文法関連で上位１００語に含まれるのは以下の語である。

sentence(s), noun(s), will, verb(s), pronoun(s), adjective, paragraph, punctuation, grammar, spelling, plural, word, capitalization, letter, proper, singular

小学校１年のものと比べると、“noun, pronoun, adjective, capitalization, plural” など文法用語そのものが使われ始めていることがわかる。

授業や学校関連の語で上位１００語に含まれるのは以下の語である。

write(ing), exercise, ideas, correct, complete, class, reading, blank, book, express presentation, students, answer, circle, form, task, question(s), pupils, create, link, paper,

example(s), parentheses, activities, classroom, mistakes, grade, school, teacher, club, lesson, music, classmate, underline

文法関連と授業、学校関連の語だけで５０以上が確認され、上位１００語のうちの５割以上になることがわかった。この一つの理由は、テキストに練習問題などのタスク活動に関するものがかなり多く含まれていることにあると思われる。

3.4　JACET8000による語彙の難易度の分析

次に、大学英語教育学会が作成した語彙表「JACET8000」を基にして関西大学の染谷泰正教授が開発されたWordLevel Checkerを使用して、フィリピンテキストの語彙をレベルを調べた。JACET8000は日本人大学生を念頭に置いて開発されたものであり、英語母語話者のコーパスではなく、日本の英語教科書などをコーパス化したデータをベーシにして作成されている（相澤、他, 2015）。

JACET8000は、抽出した8000語を頻度別に８レベルに区別して、日本の英語学習者が目指す語彙量の目安となるようにしている。以下に、そのレベルを示す。

Level 1　中学の英語教科書に頻出する基本単語
Level 2　高校初級レベル。英字新聞の７５％をカバー。英検準２級に必要な語彙
Level 3　高校英語教科書レベル。センター試験のレベル。社会人の教養レベル
Level 4　大学受験、大学一般教養の初級相当
Level 5　難関大学受験、大学一般教養に相当
Level 6　英語を専門としない大学生やビジネスマンが目指すレベル。英検準一級に必要
Level 7　英語専攻の大学生やビジネスマンの到達目標レベル。英検１級、TOEICの95%の語彙をカバー
Level 8　日本人の英語学習者の一般的な単語学習の最終到達目標レベル

テキスト	総語数	平均レベル	Level 1		Level 2		Level 3		Level 4	
			語数	%	語数	%	語数	%	語数	%
小学１	28048	1.16	22789	81.25	1390	4.96	669	2.39	198	0.71
小学２	21201	1.28	16704	78.79	1462	6.90	627	2.96	211	1.00
News corpus	27184	1.21	18664	68.66	1733	6.38	647	2.38	841	4.09

テキスト	Level 5		Level 6		Level 7		Level 8		Unknown	
	語数	%	語数	%	語数	%	語数	%	語数	%
小学１	195	0.70	148	0.53	180	0.64	147	0.52	2332	8.31
小学２	220	1.04	299	1.41	83	0.39	167	0.79	1428	6.74
News corpus	306	1.13	223	0.82	167	0.61	189	0.70	4414	16.24

表７

表７はフィリピンの小学校１年と２年、そして参考資料として今回作成したNews corpusの分析データを示している。８つのレベルに含まれない（8000語に含まれない）語数を表す“Unknown”の数がそれぞれかなり多いのは、JACET8000が日本の教科書のデータに基づいており、それに含まれていない語数が多いことを示している。

この分析からわかることは、小学校低学年のテキストにもかかわらず Level８までの語が含まれているのがわかる。テキスト全体の平均レベルが、１年が1.16、２年が1.28とほとんど変わらないのと、ネイティブの英文であるNews corpusの平均も1.21とほぼ同じ数値を示している。渡辺、その他 (2016)の日本と同じEFL環境の中国の小学校英語テキストの調査では、JACET8000の平均レベルは小学校１年から６年までほぼ同一の1.1台であったが、小学校１、２年ではレベル７以上の語は一桁の語数であったことを参考にすると、フィリピンの小学校１、２学年テキストは、難易度の高い語を多く含んでいるということがわかる。

一方、Unknownの語数が多いことから、そこに含まれる語を調べてみた。まず、小学校１年テキスト内の“Unknown”に含まれる2332語の中で、使用頻度が１０回以上の語は４５あり、その中から人名や地名などの固有名詞、数字などを除いた語は以下の通りである。

checkup, classmate, astronaut

これらの語群からもわかる通り、Unknownに含まれる語には難度の高い語彙は使われていない。

小学校２年テキストの“Unknown”に含まれる1428語のうち、使用頻度が１０回以上の語は２７あり、その中から人名や地名などの固有名詞、数字などを除いた語は以下の通りである。

punctuation, capitalization, preposition, souvenir

ほとんどが文法用語であり、日本の中高教科書には文法用語はほとんど日本語が使われていることもあるため、JACET8000には含まれなかったということであろう。それ以外の難度の高い語は使用されていない。

News corpusの“Unknown”に含まれる4414語のうち、使用頻度が１０回以上の語は７１あり、固有名詞などの語を除いた主な語は以下の通りである。

Republicans, postseason

ほとんどが固有名詞であり、特に米国大統領選挙関連のものであった。難度が高い語のほとんどは使用頻度が１回のものであった。

したがって、“Unknown”に含まれる語は多いが、ほとんどは固有名詞などであり、難度の高い語は、少なくとも使用頻度の多い語群には含まれていないことがわかった。

結論

本研究では、ESL環境にあるフィリピンの小学校１、２年用の教科書をデジタル化し、デジタルツールを使用して計量的分析を行った。比較参考データとして英文ニュースのコ

ーパスと中国の小学校英語教科書と日本の中高の英語教科書のコーパスを使用した。

まず語数、異語数、TTRの分析では、総語数と異語数においてフィリピンの教科書は中国や日本（中高）の教科書よりも圧倒的に多かった。これはESL環境として早期に公用語である英語を学習させる必要から、授業時間を１年次から英語に多くの時間を割いているためと思われる。ただし、TTRの比較では、フィリピン教科書は中国や日本と大きくは変わらず、ネイティブの英文であるNews corpusよりはかなり低い数値であった。これは学習者の年齢や認知能力に対応するために語彙を統制している結果と思われる。

次に、日本の学習指導要領に準拠したリータビリティツールであるOzasa-Fukui Year Levelを使用した英文リーダビリティの分析では、１年教科書が日本の中学３年前半のレベルを示し、２年教科書は中学３年後半のレベルを示した。これもESL環境にあるとはいえ学習者の年齢を考慮してテキストが作成されていることを示すものである。

特徴語の分析では、News corpusを比較対象コーパスにして検索を行った。特徴語として検索された上位１００語の中の多くは小学校１年、２年教科書とも、文法関連の用語や授業や学校に関係する用語であった。これは予想されたことであるが、教科書の内容が学習者の身近なものを取り上げていることが量的分析においても明らかになったということである。

最後にJACET8000を基準とする語彙のレベル分析では、難易度が高い語彙が含まれるレベル８、９の語数がNews corpusと比べて少なかったが、中国や日本の教科書の数値と比べると比較的多いことがわかった。またUnknownに含まれる語数が多かったが、調べてみるとそのほとんどが文法用語であり、特に難度が高いと思われる語は含まれていなかった。したがって、語彙のレベルも低年齢の学習者のために意図的に統制されていることがわかった。

今回の研究から、ESL環境の教科書でも、学習者の年齢を意識して、わかりやすい英文で構成されていることが示されたと考える。今後は、本研究を継続して、小学校６年までの全学年の教科書をデータ化し、分析を行ってフィリピンの小学校英語教育の全体像を調べていきたいと考える。

参考文献

相澤一美・石川慎一郎・村田年　編 (2015). 「JACET8000英単語」東京：桐原書店.

石川慎一郎 (2014).『ベーシックコーパス言語学』東京：ひつじ書房.

文部科学省 (2008). 中学校学習指導要領解説　外国語編.

張世霞 (2016). 「中国における小学校の英語教育の現状」日本言語教育ICT学会2016研究大会にて口頭発表. 2016年9月10日－11日.

渡辺清美・浅井智雄・赤瀬正樹 (2016). 「中国小学校英語教科書の語彙の量的分析 ―日本の現行教科書との比較を中心にして―」日本言語教育ICT学会2016研究大会にて口頭発表. 2016年9月10日－11日.

渡辺清美・浅井智雄・安倍規子 (2016). 「タイ国小中学校英語教科書の語彙分析―日本の現行教科書との比較を中心として―」 『日本言語教育ICT学会研究紀要』第3号, 47-56.

渡辺清美・坂元真理子 (2015). 「タイと日本の英語教科書の分析―語彙を中心として―」『日本言語教育ICT学会研究紀要』, 2, 1-10.

Blake, J. (2012). Harnessing Keyness: Corpus-based Approach to ESP Material Development. pp. 1-7. Extended abstract in KOTESOL. International Conference 2012 handbook. Seoul Korea.

Gabrielatos, C., Marchi, A. (2011). Keyness Matching metrics to definitions. Presented at Corpus Linguistics in the South, University of Portsmouth, 5 November.

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. Applied Linguistics, 28(2), 241-265.

Reid, J. (1990). Responding to different topic types: A quantitative analysis from a contractive rhetoric perspective. In B. Kroll (Ed.) Second Language Writing: Research insights for the classroom (pp. 191-210). Cambridge University Press. cited by Kojima, M. in http://www.kojima-vlab.org/index.php?%E8%AA%9E%E5%BD%99%E3%81%AE%E8%B1%8A%E3%81%8B%E3%81%95 （アクセス日：2016年8月31日）

Guidelines on the Implementation of Grades 1 to 10, Department of Education, (2012)). Republic of the Philippines. Retrieved from http://www.deped.gov.ph/sites/default/files/order/2012/DO_s2012_31.pdf （アクセス日：2016年8月31日）

A Quantitative Study of Elementary School English Textbooks in Philippines -Focusing on Grade 1 and 2 textbooks-

Kiyomi WATANABE

Department of Business Administration, Faculty of Business Administration, Fukuyama Heisei University

Abstract: This paper studies the Elementary English school textbooks in the Philippines, the only ESL country in East Asia, for grades one and two by way of a series of quantitative analyses. The textbooks were first digitalized, then with the help of digital linguistic tools, following quantitative analyses were carried out; calculation of token, type and TTR, analysis of Readability levels according to Japanese school year (grade 7 through 12), creating keyword lists, and analysis of their vocabulary level according to JACET8000. The textbook data of Chinese elementary school and Japanese junior high and high school and an interim corpus data taken from News articles from the Internet were used as comparative data. The study finds that the vocabulary size of the textbooks from the Philippines is larger than both Chinese and Japanese textbooks, but the TTR results and the vocabulary level remains the same. The average Readability level of both textbooks are found to be in grade 8 of Japanese school year, the keywords in comparison with the News corpus indicate that the textbooks contain many words used for grammar, school, and classes. The vocabulary difficulty levels according to JACET8000 is about the same as those of Chinese and Japanese textbooks. Overall, the study concludes that although English textbooks from the Philippines contain a larger vocabulary size, their difficulty levels are controlled to suit elementary school pupils and that the topics focus on English learning and school activities.

Key Words: quantitative analysis, ESL environment, English textbook analysis

研究業績一覧
（2016 年 1 月～12 月）

Ⅰ．学術論文（査読付き）・国際学会プロシーディング

１．朝日 亮太、「低費用航空会社の買収が競合航空会社に与える影響～サウスウエスト航空とエアトランのケース～」、『海運経済研究』第 50 号，pp. 41-49

本稿では、サウスウエストのエアトランの買収に焦点を当て、サウスウエスト、エアトランの運航した路線において、この買収がサウスウエストおよびライバル企業の運賃設定行動に対して与えた影響について、2009 年から 2014 年までの各年度の第 4 四半期の米国航空産業のデータを用いて分析している。需要関数と価格関数（疑似供給関数）の同時方程式の推定による分析の結果、買収後、サウスウエスト運航路線において、サウスウエストが運賃上昇行動をとる可能性が有ること、それに合わせライバルも運賃上昇行動をとることを示した。

２．市瀬 信子、「周京と杭州詩壇」、『中國中世文學研究』、第 67 号、pp39-61

清朝乾隆期の杭州では、盛んに詩会が開かれ、地方詩壇の隆盛期を迎えた。当時詩壇の領袖と言われたのが周京である。周京は若い頃、厲鶚らの先輩として杭州詩会の領袖となった後、無官のまま各地を遍歴し、乾隆初め頃に杭州に戻ると、再び詩会の領袖として杭州詩壇の隆盛期を築いた。乾隆時代の清代杭州詩壇は、「浙派」という、難解な文字や典故の工夫にこだわる詩風の一派として、文学史上に位置づけられてきた。しかし、領袖たる周京の詩風は浙派と大きく異なる。本論は、周京の活動と杭州詩壇の詩風を詳細に検証し、浙派の集団と目されてきた杭州詩壇が、より自由な詩風を誇る詩人集団であったことを明らかにした。

３．Eunji Seo “SOCO’s Impact on Individual Sales Performance: The mediating role of Cross-functional Integration”, *Journal of Japanese Management*,Vol. 1, No. 1, pp. 60-73.

The effects of sales orientation (SO) and customer orientation (CO) on individual performance and the mediating effects of cross-functional integration are examined in this research. The assumption, two types of selling motivation (SOCO) differentially influence individual performance, is modified by evaluating the mediating role of cross-functional integration based upon the rationale is that SO emphasizes self-interest whereas CO involves knowledge sharing and internal communication. Hypotheses are tested with structural equation modeling using survey data from 236 salespeople in a Japanese industrial company and it is found that SO directly influenced

individual performance but cross-functional integration completely mediated the relationship between CO and individual performance. Based on these results, the importance of SOCO both in order to increase upper performance is emphasized and a new perspective on the cross-functional integration of salespeople is provided.

４．徐 恩之、「国際マーケティング戦略における日本企業の標準化と国際経験に関する研究」、『商学研究』、第 10 号、pp. 33-48.

Despite extensive research conducted on assessing the effect of the marketing strategy on firm performance, there is no consensus in the literature about whether the marketing standardization affects firm's performance. This study examines the link between the marketing standardization strategy and firm's performance from the perspective of Japanese firms in international market and explores moderating effects of factors such as product category and international experience. Results showed that marketing standardization positively affects firm performance. Furthermore, international experience negatively affects positive link between marketing standardization and firm performance. This finding emphasizes the importance of strategic changes which levels of standardization depending on international experience of firm.

５．Eunji Seo and Youngjae Koh, "Balancing between exploratory and exploitative innovations: Moderating effects of interaction with international customer and corporate brand management in Japanese firms", *Proceedings of 2016 International Academic Business Conference San Francisco*, CA, pp.1-10

The aim of this paper is to address the following three issues using questionnaires that were randomly mailed to managers of international business in Japanese manufacturing companies. First, this study uses the data to test the relationship between exploratory and exploitative innovations, which are argued to have a trade-off. The analysis showed that the two innovations need not have a trade-off, but can be balanced as for the organizational learning. Second, moderating effects of interaction with customers, especially relational trust on international customers, and types of customer orientation (customer acquisition or retention) were examined. Lastly, the study focuses on corporate brand management as a moderating factor between the two types of innovation. Our findings suggest that an exploratory innovation stimulates an exploitative innovation, and this relationship is strengthened when firms have a high relational trust in international customers and pursue customer acquisition and corporate brand management.

６．奥野 友理子・元屋 愛子・高嶋 克義・徐 恩之、「医療従事者における自己犠牲志向と部門間連携が新機器受容に及ぼす影響」、『医療と社会』Vol. 26, No. 2、pp. 167-178

本研究の目的は、病院組織における新たな医療機器の導入に対する医療従事者の態度が、医療従事者の自己犠牲志向の程度と病院組織の部門間連携の状況によって影響されることを明らかにすることである。

本研究では，病院組織における新機器の導入に関する問題をイノベーション受容の視点から捉え，医療従事者の自己犠牲志向と部門間連携が新機器受容を高めることについての仮説を導出したうえで，医療従事者を対象とする質問票調査データに基づく仮説検証を実施した。その結果、医療従事者の自己犠牲志向が低い場合においては，部門間連携による新機器受容への正の影響が強まるものの，自己犠牲志向が高い場合には，その影響が弱くなることが確認され，医療従事者の自己犠牲志向の程度が部門間連携と医療従事者の新機器受容の正の関係を抑制する働きをすることが示された。この結果から、自己犠牲志向が高い場合には、医療従事者が他部門を巻き込むことへの配慮が高まり、新機器受容の意向を抑制することが推測された。このことに基づけば、病院組織の新機器導入を進めるためには、医療従業者の心理的要因を考慮する必要があり、部門間連携か自己犠牲志向の一方を高めるような戦略が有効であるという実践的インプリケーションを導くことができる。

７．Eunmi Kim, Insik Jeong and Eunji Seo, “A cross-national study on export strategies and performance of Korean and Japanese firms”, 韓国国際経営管理学会 proceedings, pp. 48-66.

This study examined the export strategies of Korean and Japanese firms and their performance in foreign markets. Hypotheses based on comparison of different impacts of cost leadership strategy, differentiation strategy, marketing standardization, and export diversification on export performance in Korean and Japanese exporters. Findings suggest that cost-based strategies enhance export performance in both Korean and Japanese firms. Adapting marketing promotion program to the specific needs of foreign markets also enhances export performance for Korean exporting firms. The relationship between geographical diversification and export performance is positively related in Korean and Japanese firms. This cross-national study could give several implication to researchers as well as exporting managers.

８．渡辺 清美・小篠 敏明、「新英文リーダビリティ指標開発における規準文の作成 Part 2」、『日本言語教育 ICT 学会研究紀要』vol. 3, March, 2016. (pp. 65-73)

学習指導要領に準拠した英文リーダビリティツールである Ozasa-Fukui Year Level を学習指導要領が改訂されたのに伴い、新教科書に基づいたバージョンを作成する過程において、リーダビリティを測定するにあたり重要な役目を果たしている規準文の選定における基準を見直した。1. 単元ごとの目標文法が含まれる文を選ぶ、2. 長い文章、イディオムなどは避ける、3. 本文に適当な文が見つからない場合は、欄外から選んでも良い、4. 選んだ基準

文のレベルのタグ付けは必ずしも単元にこだわらずに、下げることもできる、5. 学年上位では特に注意してレベルを決定する。

９．渡辺 清美・浅井 智雄・安部 規子、「タイ国小中学校の語彙分析」、『日本言語教育 ICT 学会研究紀要』vol. 3, March, 2016、pp. 47-64

タイと日本の英語教科書の語彙をデジタル言語分析ツールを利用し、計量的に分析した。その結果、語彙量においてタイのテキストがはるかに日本をしのいでいること、また、タイの小学校と中学校の間で大きな差があること、それがおそらく小学校では“Let’s Go”、中学校では “Access“という事なるテキストを使用しているためであることが量的分析においても明らかになった。

10．小篠 敏明・渡辺 清美・上西 幸治・坂元 真理子、「タイ国小中学校英語教科書のリーダビリティ分析」、『日本言語教育 ICT 学会研究紀要』vol. 3, March, 2016、pp. 1-13

英文リーダビリティ分析ツールである Ozasa-Fukui Year Level を使用して、タイの小中学校の英語教科書を分析した。結果は、日本の小学校の英語テキストと比べて、はるかに高く、日本の明治時代の英語テキストと匹敵するレベルの高さであった。

11．塩田 裕明・張 世霞・坂元 真理子・渡辺 清美、「日本・タイの英語教科書における題材内容の対照分析」、『日本言語教育 ICT 学会研究紀要』vol. 3, March, 2016、pp. 25-37

タイと日本の英語教科書を吉田(1997)が提唱する三つのスキーマを枠組みとした分析と Wmatrix3 を用いた意味領域の分析を行った。その結果、タイのテキスト日常生活の題材がより多く、個人と社会・文化的スキーマに分類されることがわかった。一方、日本のテキストはものの考え方や理解に関する題材が多く、半分以上が普遍的スキーマに分類されることがわかった。

12．Toshiaki Ozasa, Kiyomi Watanabe, Masayasu Fukui “Development of a Readability Index Attuned to the New English Course of Study of Japan (4) Development of Ozasa-Fukui Year Leve、Ver. 3.5nhnc1-6”, *International Conference on Social Science and Business, Conference Proceedings*, August 25-27, 2016, pp. 386-394

新学習指導要領の実施に伴って英文リーダビリティツールである Ozasa-Fukui Year Level の改定を行なった。２つの英語教科書セットをデジタル化し、文章の長さ、語の長さ、語彙の難易度、熟語の難易度を説明変数として、学年単位の難易度を目標変数とする

重回帰分析をまず行い、次に彫られた解析をもとに非線形分析を行なった。その結果、0.8912 の R2 値という高い数値を得ることができた。

13. Koji Uenishi, Kiyomi Watanabe, Mariko Sakamoto, Toshiaki Ozasa "Creating Criterion for Benchmark-Sentences for the Development of a New English Readability Index: Ozasa-Fukui Year Level, Ver. 3.5nhnc1-6", *International Conference on Social Science and Business, Conference Proceedings*, August 25-27, 2016, pp. 395-405

新学習指導要領に準拠するため Ozasa-Fukui Year Level を改定するに伴い、非線形分析を行う際に必要な基準文の選定基準を大幅に見直した。1. 基準文を厳選し、できるだけ少なくする、2. 選んだ文章の順序に齟齬がないかに注意する、3. 短縮、省略、固有名詞の多い文は避ける、4. 核となるテキスト（New Horizon）と扱う順序が異なる場合は、採用しない, 5. 学年上位（高校）では特に順序に気をつける、を選んだ。

Ⅱ. 学術論文（その他）

1. 朝日 亮太、「航空産業の規制緩和と変化」『季刊 Nextcom』,Vol.28 2016 Winter, pp.28-37

2. 朝日 亮太、「LCC ってなんですか？」『運輸と経済』第 76 巻, 第 4 号, pp96-100

3. 奥田 由紀恵、「情報教育初期段回におけるタッチタイピング習得法」、『経営研究（福山平成大学経営学部紀要）』、第 12 号、pp13-22

4. 尾崎 誠・福井 正康、「グラフィックスビューアを利用した 3 次元パラメータ関数の表示」、『経営研究（福山平成大学経営学部紀要）』、第 12 号、pp.23-30

5. 小玉 一樹、「パートタイマーの組織コミットメント」、『経営研究（福山平成大学経営学部紀要）』・福山平成大学経営学部紀要、第 12 号、pp.31-43

6. 佐藤 真司、「e ラーニング教材作成支援ツールの開発」、『経営研究（福山平成大学経営学部紀要）』、第 12 号、pp.45-60

7. 佐藤 幹、「地方自治体のマネジメント・コントロール技法等の有効性に関する実証分析のための予備的考察」、『保健医療経営大学紀要』、第 8 号 pp. 57-67

8. 福井 正康・尾崎 誠・朝日 亮太、「社会システム分析のための統合化プログ

ラム２５ － パス解析・多次元尺度構成法・数量化Ⅳ類 －」、『経営研究（福山平成大学経営学部紀要）』、第 12 号、pp.61-76

９．福井 正康・ 尾崎 誠、「社会システム分析のための統合化プログラム２６ － グラフと定積分の拡張・不等式グラフ －」、『経営研究（福山平成大学経営学部紀要）』、第 12 号、pp.77-94

10. 福井 正康・岩村 忠昭・尾崎 誠、「社会システム分析のための統合化プログラム２７ － 局所重回帰分析・パネル重回帰分析 －」、『経営研究（福山平成大学経営学部紀要）』、第 12 号、pp.95-111

11. 本田 良平、「T*he Sound and the Fury* と *Absalom, Absalom!* －燃え上がり消える炎としての人間像から伝播する波紋としての人間像への変化について―」、『経営研究(福山平成大学経営学部紀要)』、第 12 号、pp113-123

Ⅲ．学会報告（討論者を含む）

１．朝日 亮太、スペシャルセッション：国際交通研究の新たな地平―村上英樹教授の研究を踏まえて―「An Empirical Analysis of the Effect of Multimarket Contacts on US Air Carriers’ Pricing Behaviors」，日本海運経済学会第 50 回大会、10 月 22 日、神戸大学

２．市瀬 信子、「清代詩社と記録」、中国中世文学会平成二十八年度研究大会、10 月 29 日、広島大学東千田キャンパス

３．市瀬 信子、「陸游「秋暁倚闌」、「読陶詩」」、讀游會、11 月 12 日、日中友好協会兵庫県連合会

４．奥田 由紀恵、「学修システム『manaba』を導入して」、日本教育情報学会第 32 回年会、日本教育情報学会第 32 回年会論文集、pp246～247、8 月 21 日、福山大学宮地茂記念館

５．尾崎 誠・福井 正康、「HTML5 による幾何アニメーション」、日本教育情報学会第 32 回年会、日本教育情報学会第 32 回年会論文集、pp266～267、8 月 21 日、福山大学宮地茂記念館

６．徐 恩之、「国際マーケティング戦略における日本企業の標準化と国際経験に

関する研究」日本商学研究学会、6 月 18 日、大分大学

７．徐 恩之・鄭印植・金殷美、「日本企業におけるマーケティング標準化と製品戦略の関係に関する実証研究」、日本商業学会全国研究報告会、12 月 18 日、学習院大学

８．Eunji Seo and Youngjae Koh, “Balancing between exploratory and exploitative innovations: Moderating effects of interaction with international customer and corporate brand management in Japanese firms”, 2016 International Academic Business Conference San Francisco, CA, 2016. 8. 3, America San Francisco

９．Eunmi Kim, Insik Jeong and Eunji Seo , “A cross-national study on export strategies and performance of Korean and Japanese firms”, 韓国国際経営学会 2016 海外学術大会、6 月 28 日、中国 Nanjing

９．福井 正康・中村 雅子・永井 純子、「分析ソフト College Analysis の卒論指導での利用法」、第 48 回中国・四国学校保健学会プログラム・抄録集 pp. 17～18、6 月 26 日、福山平成大学

10．福井 正康・奥田 由紀恵・細川 光浩・尾崎 誠、「College Analysis の拡張と活用」、日本教育情報学会第 32 回年会論文集 pp. 168～171、8 月 21 日、福山大学宮地茂記念館

11．福井 正康・渡辺 清美・小篠 敏明、「英文難易度測定システム CheckRead の改訂」、日本教育情報学会第 32 回年会論文集 pp. 268～269、8 月 21 日、福山大学宮地茂記念館

12．福井 正康・小篠 敏明、「Ozasa-Fukui Year Level の決定法とリーダビリティ測定プログラム」、日本言語教育 ICT 学会・招待講演、9 月 11 日、拓殖大学国際教育会館

13．細川 光浩・福井 正康・渡辺 清美・小篠 敏明「Moodle を使った e-Learning の英語学習」、日本教育情報学会第 32 回年会、日本教育情報学会第 32 回年会論文集、pp. 162～165、8 月 21、福山大学宮地茂記念館

14．本田 良平、「“the darkness rushing past my breast” －*As I Lay Dying* における Dewey Dell の恐怖について―」、平成 28 年度中・四国アメリカ文学

会冬季大会、12 月 3 日、於　愛媛大学法文学部中会議室

15. 渡辺 清美、「フィリピンの小学校英語教科書の分析」日本言語教育 ICT 学会 2016 研究大会（東京大会）、9 月 10 日-11 日、拓殖大学国際教育会館

16. 渡辺 清美、「中国小学校英語教科書の語彙分析」日本言語教育 ICT 学会 2016 研究大会（東京大会）、9 月 10 日-11 日、拓殖大学国際教育会館

Ⅳ．その他

１．朝日 亮太「合併買収による米国航空産業の変化について～市場集中度の観点から～」『KANSAI 空港レビュー2016 年 5 月号』No.450, pp.24-27

２．福井 正康、「マルコフ連鎖モンテカルロ法による量子力学的視覚効果」、『経営研究（福山平成大学経営学部紀要）』、第 12 号、pp.125～133

＜経営学部著者紹介＞

渡辺　清美：福山平成大学経営学部経営学科教授
福井　正康：福山平成大学経営学部経営学科教授
小玉　一樹：福山平成大学経営学部経営学科教授
市瀬　信子：福山平成大学経営学部経営学科教授
佐藤　　幹：福山平成大学経営学部経営学科准教授
尾崎　　誠：福山平成大学経営学部経営学科准教授
奥田由紀恵：福山平成大学経営学部経営学科助手

＜編集委員＞

市瀬　信子：福山平成大学経営学部経営学科教授
奥田由紀恵：福山平成大学経営学部経営学科助手

経　営　研　究

第 13 号

平成 28 年 3 月 25 日　発行

発行所　：　福山平成大学経営学部

〒720-0001　広島県福山市御幸町上岩成正戸 117-1

電話 (084)972-5001, Fax (084)972-7771

＜経営学部著者紹介＞

渡辺　清美：福山平成大学経営学部経営学科教授
福井　正康：福山平成大学経営学部経営学科教授
小玉　一樹：福山平成大学経営学部経営学科教授
市瀬　信子：福山平成大学経営学部経営学科教授
佐藤　　幹：福山平成大学経営学部経営学科准教授
尾崎　　誠：福山平成大学経営学部経営学科准教授
奥田由紀恵：福山平成大学経営学部経営学科助手

＜編集委員＞

市瀬　信子：福山平成大学経営学部経営学科教授
奥田由紀恵：福山平成大学経営学部経営学科助手

経　営　研　究

第 13 号

平成 29 年 3 月 25 日　発行

発行所　：　福山平成大学経営学部

〒720-0001　広島県福山市御幸町上岩成正戸 117-1

電話 (084)972-5001, Fax (084)972-7771

ISSN1342-0860

Bulletin of
Faculty of Business Administration
Fukuyama Heisei University

No.13 March 2017

CONTENTS

<Articles>

The History of Zhou Jing
-With a Focus on “Zhou Mumen Muzhiming”-
……Nobuko ICHINOSE 1

Effect of Quick Touch-Typing Learning in the Information Literacy Beginning Stage
……Yukie OKUDA 21

Geometric Animation with HTML5 and Javascript ……Makoto OZAKI 33

Making Part-time workers as "Core Labor Force"
-Its Effect on Organizational Citizenship Behavior-
……Kazuki KODAMA 43

An Empirical Study of the Adoption of “Management Control Systems” in Hiroshima City Office
……Tsuyoshi SATO 59

Multi-purpose Program for Social System Analysis 28
-Meta-Analysis, Logistic Regression Analysis-
…… Masayasu FUKUI, Kazuki KODAMA and Makoto OZAKI 69

Multi-purpose Program for Social System Analysis 29
-Parameter Design, On Line Quality Engineering-
…… Masayasu FUKUI, Tomoyuki OYAMA and Nozomu ODA 99

Multi-purpose Program for Social System Analysis 30
-Anomaly Detection-
…… Masayasu FUKUI, Tomoyuki OYAMA and Nozomu ODA 121

Multi-purpose Program for Social System Analysis 31
-Survival Analysis-
……Masayasu FUKUI and Xiaona WU 139

A Quantitative Study of Elementary School English Textbooks in Philippines
-Focusing on Grade 1 and 2 textbooks-
……Kiyomi WATANABE 163

List of Research Achievement …… 176

FUKUYAMA HEISEI UNIVERSITY
117-1,Kamiiwanari,Miyuki-cho,Fukuyama,Hiroshima 720-0001,Japan