

201132049B

厚生労働科学研究費補助金

医薬品・医療機器等レギュラトリーサイエンス総合研究事業

医学分野で要する統計解析の理解促進に資する
Web アプリケーションの構築

平成22年度～23年度 総合研究報告書

研究代表者 岡本 晃典

平成24（2012）年 5月

厚生労働科学研究費補助金

医薬品・医療機器等レギュトリーサイエンス総合研究事業

医薬学分野で要する統計解析の理解促進に資する
Web アプリケーションの構築

平成22年度～23年度 総合研究報告書

研究代表者 岡本 晃典

平成24（2012）年 5月

はじめに

現在、統計解析を始めとするデータ解析は様々な分野や領域で利用されている。また、統計解析手法を計算可能なソフトウェアは無償、有償を問わず非常に多くのものが開発、利用されており、さらにそれらの大半は非常に高機能であり、計算結果に関する信頼性も高い。観察対象の特徴を捉えるための記述統計に始まり、母集団の一部として得られた標本を基に行う推測統計、そして有意差検定、さらには多変量解析や近年ではデータマイニング手法なども次々とソフトウェアの中に収載され、一つのソフトウェアにより非常に多様な解析を実施することが可能となっている。

研究代表者の所属する研究室においても、医薬学データ用統計解析プログラムとして Web サイト “ 医薬学データ用統計解析プログラム (MEPHAS)” を開発し、運用してきた。現在、年間のアクセス数は約 18,000 に上る。MEPHAS は先に記載した既存ソフトウェアのように広範囲、かつ多数の手法を利用できるわけではなく、利用可能な手法においてもその計算条件の設定や結果などは簡易なものに留まっているが、その一方で Web を通じて利用可能である点や入力項目が単純である点など、基本的な解析手法（例えば二群の平均値の差の検定や多重比較法など）が簡便に実行可能であることが特徴の一つであり、他の統計解析ソフトウェアとは異なる利用法を提供できる点になり得ると考えている。しかしながら、MEPHAS の利用者からの問い合わせにおいても、統計手法の適用としてあまり妥当でない手法を利用している例なども確認されており、データ解析の道具として利用する際にも一定以上の統計学に対する理解が必要であろう。そこで MEPHAS を統計初学者がより利用し易いサイトへとさらなる改善を図ることにより、利用者のニーズに答え、かつ統計に関する理解の裾野を広げることが期待できると考えたことが本研究の発端である。

医薬学分野など統計解析手法をデータ解析のツールとして利用する分野は多く、そういった分野における統計解析への入門ツールの一つとして MEPHAS が貢献していけることを期待し、研究・開発を平成 22 年度から 23 年度まで実施した。実施内容を以下報告する。

目 次

I. 総括研究報告	
医薬学分野で要する統計解析の理解促進に資する Web アプリケーションの構築	-- 1
岡本 晃典	
(資料) 更新後の MEPHAS の Web ページキャプチャー画像 (作成した Web サイトの一部のみ抜粋)	
II. 研究成果の刊行に関する一覧表	----- 43
III. 研究成果の刊行物・別刷	----- 44

厚生労働科学研究費補助金（医薬品・医療機器等レギュラトリーサイエンス総合研究事業）

（総合）研究報告書

医薬学分野で要する統計解析の理解促進に資する Web アプリケーションの構築

研究代表者 岡本 晃典 国立大学法人大阪大学大学院薬学研究科 助教

研究要旨

統計解析は、医薬学分野に限らず多くの分野においてデータ解析のための道具として活用されている。既存の統計解析ソフトウェアほど高機能でも広範な手法を収載してもいないが、MEPHAS は Web を通じて使用する統計解析プログラムパッケージであり、医薬学分野で必要となる基本的な解析手法を比較的簡便に利用できる点に特徴を有する。しかし、統計学的にあまり妥当でない解析を行おうとする例などもあることから、MEPHAS のユーザビリティをより向上し、かつ適切な情報を提供できるようコンテンツを追加あるいは修正できれば、利用者の統計解析への理解促進につながる事が期待できる。

そこで本研究では、MEPHAS のコンテンツを中心にデザインも含めた Web サイトの修正・追加を行う分担研究と、Web サーバー側で稼動している統計解析手法の計算を実行するプログラム群の修正・追加作成を行う分担研究に分けて平成 22 年度から 23 年度まで実施し、一定の開発成果を得た。それら成果を合わせて、更新後の MEPHAS を、現状アクセス可能な範囲を限定しているが試験的に公開した。この公開により計算エラーや記載間違い等を抽出し、それらを更新、修正した後、一般に広く公開する予定である。

研究分担者一覧

高木達也・国立大学法人大阪大学大学院薬学研究科・教授

川下理日人・国立大学法人大阪大学大学院薬学研究科・助教

A. 研究目的

医薬品、医療機器における有効性のみならず、安全性や品質などの調査や研究、開発の様々な場面において、推論の有効性を理論的に検証する統計学の有用性は高い。研究代表者の所属する研究室では以前から、医薬学データ用統計解析プログラムとして Web サイト “医薬学データ用統計解

析プログラム (MEPHAS)” [1] を開発、運用してきた。MEPHAS は医薬学分野において使用頻度の高い多重比較法を各種揃え、また手法の簡単な解説、手法の選択補助などを提供している。2002 年の公開以来、順調に利用者数を伸ばし、現在では年間のアクセス数は約 18,000 に上る。また利用者からのエラーの指摘や質問などに応じて現在もプログラム群の改善を行っている。しかし質問の中には、分散が 0 のデータにおいて平均値の差の検定を行うなど、統計的に妥当でない利用法をする例も見受けられる。そこで、MEPHAS を統計初学者がより利用し易いサイトへと更新を図ることは、科学的な正当性を根拠とする評価手法の整備につながり、また多くの利用者のニーズに答え

ること、科学的な論理に基づいた議論のツールである統計に関する理解の裾野を広げることも期待できることから研究の計画に至った。

現行の MEPHAS では Common Gate Interface (CGI) を通じて利用者が入力したデータをサーバー側で受信し、解析結果を提示している。これらデータの送受信の枠組みの更新も必要があれば検討するが、むしろユーザーインターフェースや結果の解説、解析したいデータに応じた適切な手法の選択補助や、場合によっては診断機能など、ユーザビリティの向上を図ることを中心に更新を行った。近年では統計解析ソフトウェアの種類も充実し、特に R[2] のように基本手法から最先端の手法までをカバーするソフトウェアもある。現行の MEPHAS においても医学分野でのニーズの高い多重比較法や多変量解析を備えるが、手法の充実よりむしろ統計初学者が本サイトを利用することにより、統計学の有用性への理解を深め、興味を覚えるような、統計学の導入となり得るサイトの構築を念頭に更新を実施した。各種統計解析ソフトウェアの高度化、専門化により、解析を行う統計の専門家と、統計の解析結果を必要とする利用者との間で起こりうる統計への理解、特に解析の結果の解釈における乖離が問題にならぬよう、統計学の共通理解を深めることは重要である。その統計学の理解を深めることに対し、統計学への導入となるようなサイト構築を目指す本研究課題は貢献できると考えた。

具体的には、MEPHAS 上において統計解析プログラムとして提供されている各手法の解説のための Web ページと手法選択補助のためのフローチャートの改訂を中心に修正、改訂を行うことにより MEPHAS 利用者のユーザビリティの向上を目的とした。また、特に手法の解説などは読み易く、かつ分かり易く情報を提供することも重要であり、それらは文字の大きさや 1 ページに表示される文字の量なども含めた Web サイトとして

のデザインに影響を受けるため、MEPHAS 全体のデザインを修正することも重要な実施内容の一つと考え、Web サイト全体のデザインやメニュー表示なども検討、修正を行い、ユーザビリティの向上を図ることを目的の一つとした。

また、MEPHAS は Unix (もしくは Linux) 上において HTTP サーバープログラムである Apache [3] を用いて提供されており、CGI を利用した統計解析プログラムパッケージをその機能として提供している。サーバー側にて稼働しているプログラムは、シェルスクリプト、C 言語、Perl などの言語を用いて作成されている。それらプログラム群は Web サイトと同様、2000 年代の前半に作成されたものであるが、計算上の間違いや実行上のエラーが無い限りは修正の必要性は高くない。そこで、先の Web サイトの更新に伴い必要となる機能を実現するためのプログラム更新や新規プログラム作成を中心に検討し、既存プログラムについては今後のメンテナンス作業や更新作業の簡便化を念頭に更新の必要性を検討することも目的の一つとした。

B. 研究方法

本研究事業の実施内容については、三つのカテゴリーとして「調査」、「開発」、そして「評価」に分類し実施することとした。

まず「調査」では、現行の MEPHAS における CGI を利用したデータの送受信及び解析のプログラムの見直しもさることながら、旧来のフレーム等を多用した解説や利用目的に基づく解析手法の選択補助など、デザインも含めた Web サイト全体を、現行の Web サイト構築技術に照らし合わせ、スクリプト言語による動的なページ構築の検討も含め、改善方針の検討を行った。さらに、現行の MEPHAS のコンテンツの一つであり、統計を専門としない利用者が解析手法を選択する際の判断材料の一つとして重要であろう、各解析手法や語句の説明等のページについて、その記載内容

の読み易さ、理解し易さを向上するべく、改善すべき点の把握を行った。

「開発」では、調査において挙がってきた改善の検討内容に沿った Web ページ及びコンテンツ、サーバー側のプログラムの修正、更新や新規作成を主に実施した。現行の MEPHAS における問題点の一つに、解析がエラーをした際に使用者に表示されるエラーの原因を示すメッセージが不足している点が挙げられ、これについてはサーバー側のプログラム内でのエラー検出の箇所をより工夫することで修正を図ることとした。個々のデータに応じた統計解析手法の選択補助や解説などの追加は、膨大な量になることが想定され、また数値データのみのサーバー送受信に留まらない可能性があるため、まずは現行の MEPHAS で利用可能な手法に関する基本的な内容や使用者を間違った判断に導きかねない注意点が使用者へ伝わり易くなるよう、更新を行うこととした。また、「調査」により決められた改善の方向性に従い、ユーザビリティの向上につながるよう、デザインも含めた Web ページの更新を進めた。その後、Web ページ全体の更新作業と共に、解析手法等の説明に関して内容の変更や追加も含めた更新を行い、試験公開に向けた準備作業を実施した。

「評価」に関しては、試験公開までに更新前後での比較評価を大学院生等を対象に実施予定であったが、更新作業の一部遅延により、試験公開前の評価実施を見送り、試験公開を優先させた。今後、更新後の Web サイトが MEPHAS と同様、無制限にアクセス可能となった際に、アクセス数や使用ログなどにより、ある程度明確に評価されると考えている。なお、現行の MEPHAS には無い手法を追加する場合に既存の他の統計解析ソフトウェアの計算結果と比較することで作成し

たプログラムによる計算結果の妥当性の評価については適宜実行した。

(倫理面への配慮)

本研究事業での研究内容については電子データ、数値情報のみを取扱い、またそれら数値情報の入力、データの送信に際しても個人につながり得る情報の入力は全く必要ないため、倫理面に関して特別の配慮は必要ない。

C. 研究結果



Figure 1. 現行の MEPHAS トップページ

Figure 1 は現行の MEPHAS トップページを示している。図中に記載があるように、MEPHAS は 2002 年から公開しており、開発は 2000 年頃から 2002 年にかけて主に行われた。つまり Web サイトの構築方針や実際の作成方法なども 10 年以上前の考え方や技術などが使われている。2000 年当時のネットワーク回線では、ブロードバンドとして ADSL による 1Mbps (Mega bit per second; 1 秒間に 1M bit を送信できる回線が 1Mbps = 125 kbytes/s に該当する。ちなみに現在の光回線は Gbps、1Gbps = 125Mbytes/s が主流である) の回線が一般家庭用として普及し始めた年代である。つまり Web サイト利用者の通信回線は現在よりも遥かに通信回線の容量が少なく、従って Web ページを作成する際に、例えば多数の画像や動画、

Flash や JavaScript などのプログラムをページに配置すると表示に時間がかかるため控えるなど、通信の容量を多くしない工夫が尊ばれる時代でもあった。MEPHAS の作成においても同様に、画像などは少なく、また動的なデザインなどは取り入れず、できる限り利用者のストレスとならないことを、Web サイト構築時の主方針として作成した経緯がある。

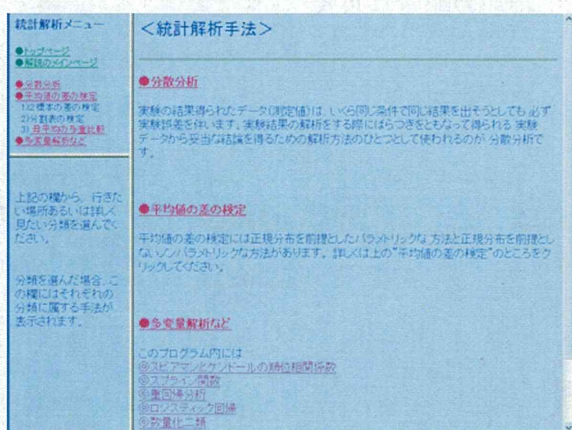


Figure 2. 現行の MEPHAS における『手法の解説』ページ

Figure 2 は Figure 1 のトップページの右側に表示されているメニューのボタンから『手法の解説』を選んだ際の移動先ページである。このページに関しても先のトップページと同様、可能な限り文字のみで情報を表現しようとしており、そのことは内容の分かり難さを助長していると思われる。また、現在の Web ページ作成においては検索エンジン最適化 (Search Engine Optimization: SEO) のためなどにより避けられる傾向のあるフレームによるページレイアウトが使用されている。勿論 MEPHAS は広告などのための Web サイトではないため、SEO などは関係無いが、World Wide Web Consortium (W3C) の HTML 4.01 や XHTML 1.0 の勧告[4,5] においてもフレームは厳密 (strict) な文書定義にも、移行上の (transitional) 文書定義にも属しておらず、フレームセット (Frameset) として別に文書定義が設けられており、実質的に推奨されていないものである。

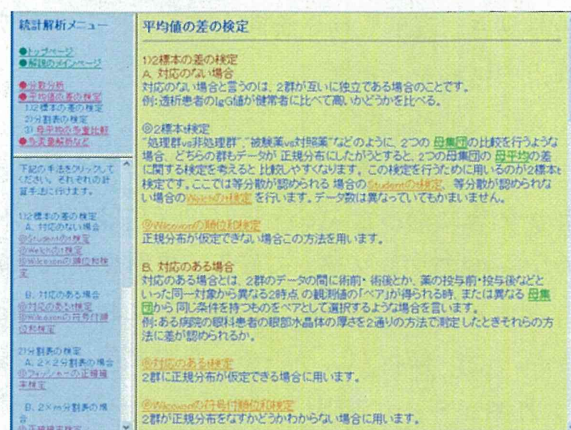


Figure 3. 『平均値の差の検定』の解説ページ

Figure 3 は Figure 2 の左フレームのメニューから『平均値の差の検定』を選択した際に表示されるページである。中央部フレームと左下部メニューのフレームが変化している。図に示されるように容量に負担をかけない方針もあり、図表も使用していない説明であり、また文章にしても決して読み易さや理解し易さに配慮したものとはなっていない。ほぼ Figure 3 と同様のため本報告書には記載しないが、他の解説ページも同様に図表などはほとんど使われていない。

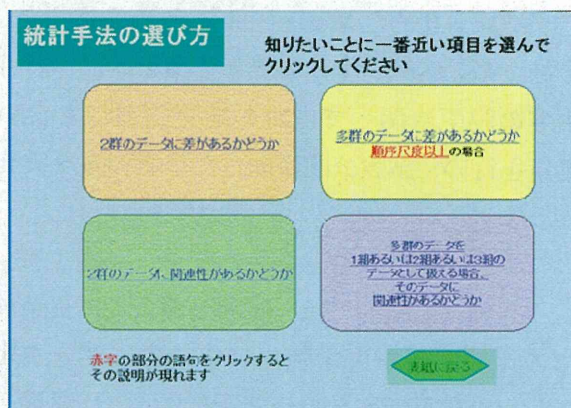


Figure 4. 『統計手法の選び方』のページ

Figure 4 は Figure 1 のトップページの右側に示しているメニューから『手法の選択』を選んだ際の移動先ページである。このページにおいても先の Figure 2 に示した『手法の解説』ページと同様、『表紙に戻る』として Web サイトのトップページへのリンクは準備されているものの、他ページへ

と移動するためのメニューはページ内に設けられていない。また先の『手法の解説』ページとは異なりフレームは使われていないが、その結果としてページごとにページデザインが変化し一貫性がないため、その意味でも Web サイトの利用者を迷わせる一因になり得ると考えられた。Figure 4 に示されたページの内容としては、ページ全体を図として表そうとしているものの、それにより内容を分かり易く伝えられているかには疑問が残る。Figure 4 に示した現在のページでは箇条書きと同様であり、改善の余地があるものと考えられた。

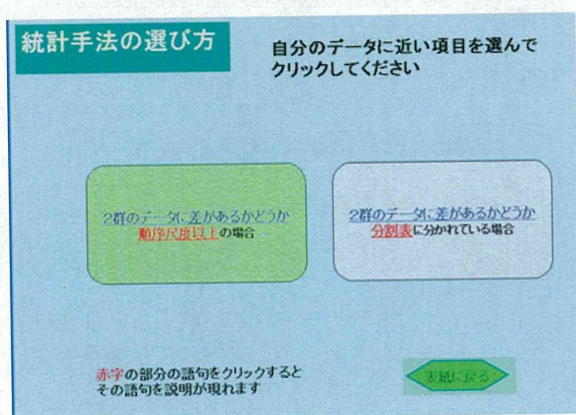


Figure 5. 『2群のデータに差があるかどうか』のページ

Figure 5 は Figure 4 に示したページの左上、『2群のデータに差があるかどうか』を選んだ際に移動するページである。Figure 4 から Figure 5 への移動を考えた場合、ページのデザインとしては一貫したものであるが、一方で、Figure 4 の『2群のデータに差があるかどうか』以外を選択した場合、例えば『多群のデータに差があるかどうか 順序尺度以上の場合』を選択した場合には Figure 6 に示すフローチャートが表示されるページへ移動する。

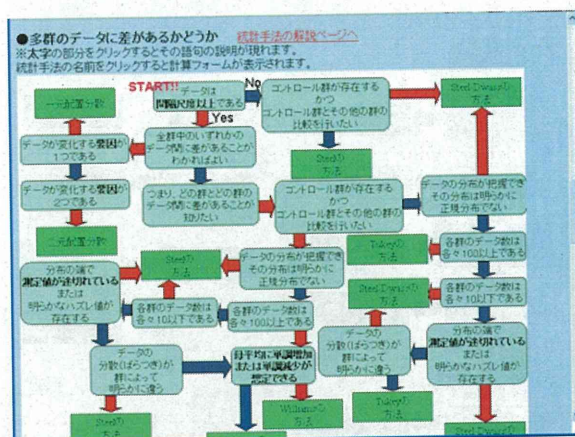


Figure 6. 手法選択のフローチャート掲載ページ中、『多群のデータに差があるかどうか』の箇所

Figure 4 に示される『統計手法の選び方』のトップページでは、ユーザーが現在解析を考えているデータセットの特徴に応じて4つのカテゴリーが示されている。上記のように Figure 4 からの移動先は Figure 5 に示されるページを挟む場合と挟まない場合があった上で、フローチャートへとリンクしている。それらフローチャートへ導くための設問は必ずしもその意図が分かり易い文章とはなっていない。また、単純な質問形式とはいえ、後に使用されるフローチャートとは異なる形式での選択肢の提示は一貫性が無く、やはり利用者を戸惑わせる要因の一つとなり得る点であり好ましくなく改善が必要と判断した。

Figure 7 から Figure 11 までは各フローチャートを示している。Figure 7 から順に、『2群のデータに差があるかどうか 順序尺度以上の場合』のフローチャート、『2群のデータに差があるかどうか 分割表の場合』のフローチャート、『多群のデータに差があるかどうか 順序尺度以上の場合』のフローチャート、『2群のデータに関連性があるかどうか』のフローチャート、そして『多群のデータを1組あるいは2組あるいは3組のデータとして扱える場合、そのデータに関連性があるかどうか』のフローチャート、計5つのフローチャートである。

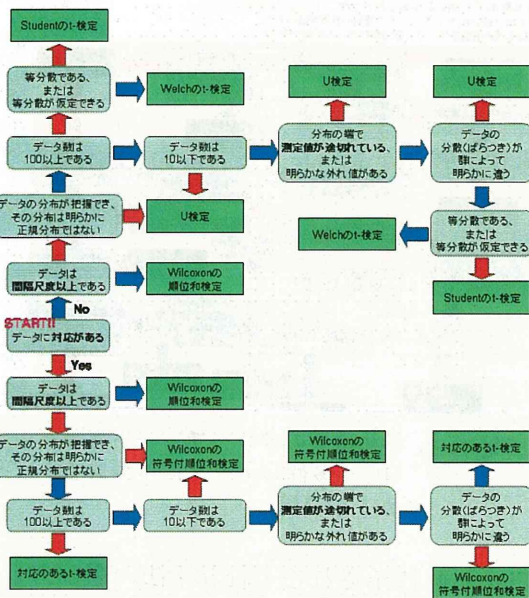


Figure 7. 『2群のデータに差があるかどうか 順序尺度以上の場合』のフローチャート

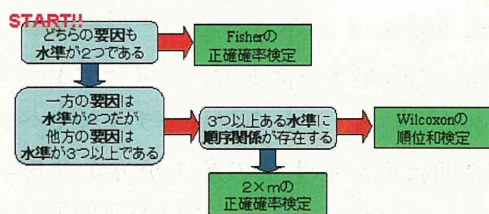


Figure 8. 『2群のデータに差があるかどうか 分割表の場合』のフローチャート

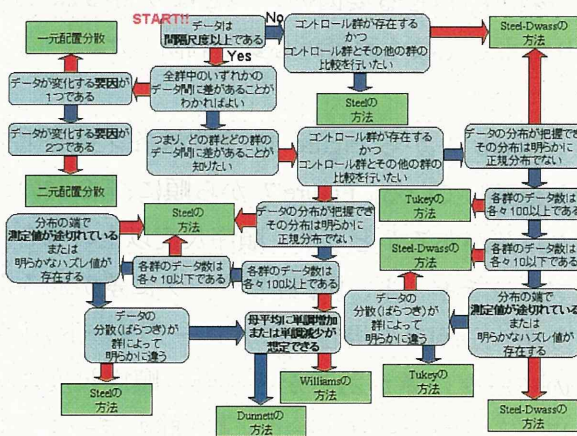


Figure 9. 『多群のデータに差があるかどうか 順序尺度以上の場合』のフローチャート

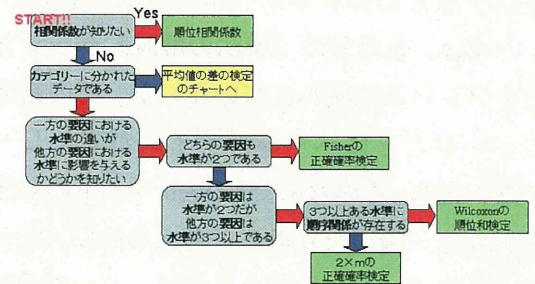


Figure 10. 『2群のデータに関連性があるかどうか』のフローチャート

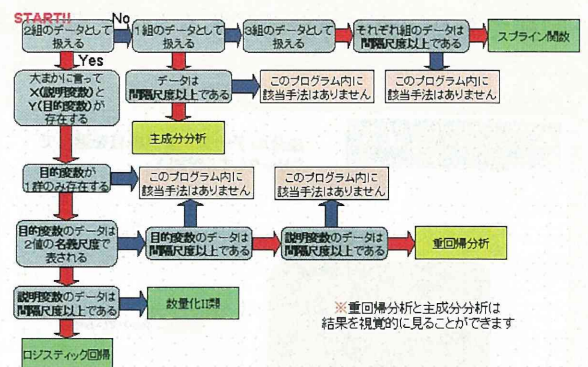


Figure 11. 『多群のデータを1組あるいは2組あるいは3組のデータとして扱う場合、そのデータに関連性があるかどうか』のフローチャート

Figure 10 のフローチャートは Figure 8 のフローチャートの中に含んでおり、さらにその直前には Figure 7 のフローチャートへ分岐する設問が設けられている。それらフローチャートの重複の解消も含め、手法選択のための手順を整理する必要があると考えられた。

また Figure 7 及び 9 のフローチャートは情報が詰め込まれ過ぎており、見づらいこともユーザビリティを悪くする要因の一つと考えられた。さらに Figure 11 のフローチャートでは『このプログラム内に該当手法はありません』との結果に行き着くことが多く、該当手法を MEPHAS に追加するか、あるいはフローチャート自体の構成を見直すことが必要であると考えられた。

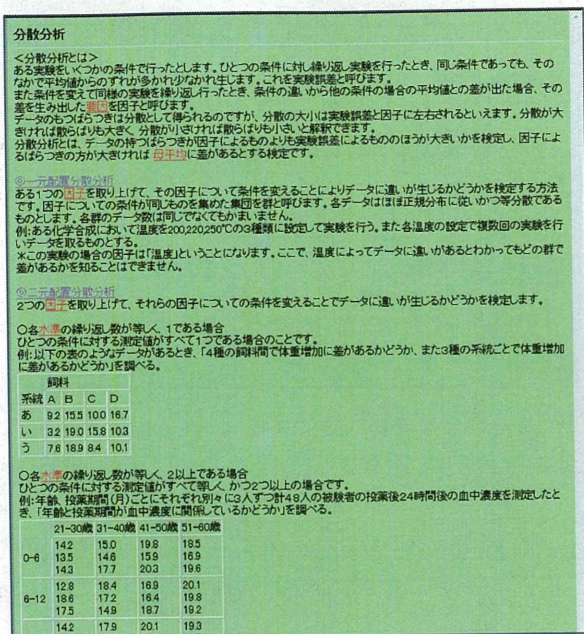


Figure 12. 現行の MEPHAS における『分散分析』についての説明ページ
Web サイト上ではフレームの中で表示される。

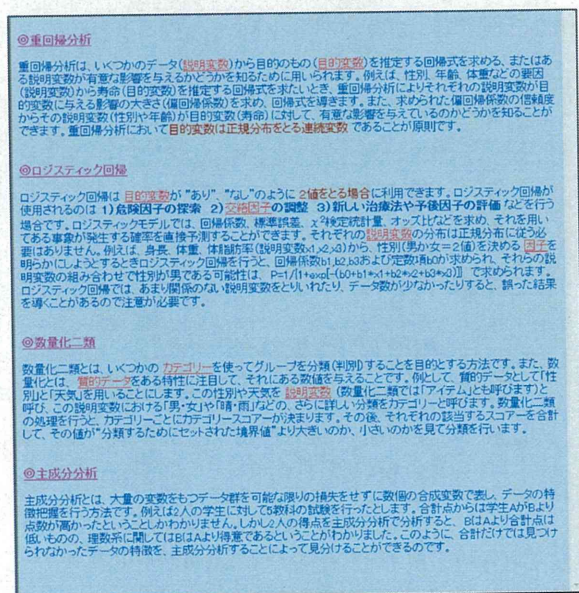


Figure 13. 現行の MEPHAS における『多変量解析など』についての説明ページ
Web サイト上ではフレームの中で表示される。

Figure 12 は現行の MEPHAS における分散分析を、Figure 13 は同じく多変量解析などを説明した Web ページを示している。後者の Figure 13 に示すような文章のみの説明では伝えられる情報に限りがあり、また事例なども含まれていないことか

ら、解析手法の説明の理解には適さないと思われる。従って、Figure 12 に示すような事例や図表が現行の MEPHAS においても使用されている場合にはできる限りそれらを活用することで作業の省力化を図りつつ、文章のみの説明については事例や図表を導入した内容へ更新を行うこととした。

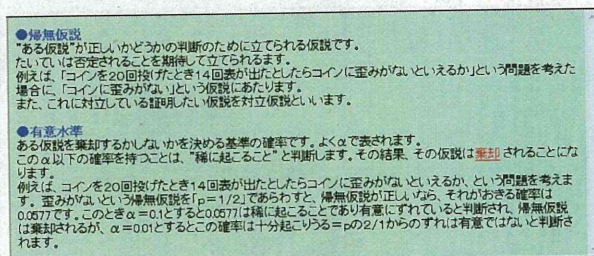


Figure 14. 現行の MEPHAS における語句に関する説明ページの一例

Figure 14 は、現行の MEPHAS 内に設けられた語句に関する説明ページの一例を示している。この語句に関する説明ページは、該当する語句に設定されたリンクをクリックした際に新規に別ウィンドウを起動し、そこに表示される。現行の MEPHAS においては手法の説明などを読む際に、語句の意味を確認したい時にクリックし内容を表示させることを考えて設けられている。別ウィンドウを起動させることもあり現行の MEPHAS では、幾つかの関連語句ごとに解説ページが作成されており、1つのページにまとめられてはいない。更新後の MEPHAS においても語句の説明は統計解析の理解促進の一助になると考え、よりよい利用法を検討した。

以上から、MEPHAS 作成時期が10年以上前であることも踏まえ、Webサイトのデザインも含め、以下の点を修正を行う項目として考えた。

- 1). ナビゲーション機能 (メニューの表示など) の充実
- 2). フレームの廃止
- 3). 統計手法の解説ページの読み易さの向上

- 4). 手法の選択に至るフローチャートの整理と修正
- 5). 統計手法の解説ページにおける解説内容への事例や図表の追加
- 6). 語句に関する説明の修正



Figure 15. 更新後の MEPHAS のトップページ

更新後の MEPHAS に関する画面表示などは全て開発中のものであり、最終公開時とは異なる可能性がある。以降、明記しないが更新後の MEPHAS に関する情報は全て同様に、最終公開時とは異なる可能性がある。

Figure 15 は更新後の MEPHAS のトップページを示している（以降、特に明記しないが、本報告書中の更新後の MEPHAS に関する情報は全て開発中のものであり、最終公開時とは異なる可能性がある）。フレームを用いずに各コンテンツへの移動メニューを上部にまとめたテンプレートを使用している。例えば『Explanation』からは現行の MEPHAS における『手法の説明』のコンテンツへと移動できる。Figure 16 はその移動先である『Explanation』のトップページである。



Figure 16. 更新後の MEPHAS の『Explanation』のトップページ

先の Figure 15 のトップページにおいて記載したように、各コンテンツへの移動メニューが上部にまとめられ、トップページと同様に構成されており、使用者にとっては現サイトよりも分かり易く、使い易いことが期待できる。また、この『Explanation』以下のコンテンツに関するメニューは左側にまとめられている。

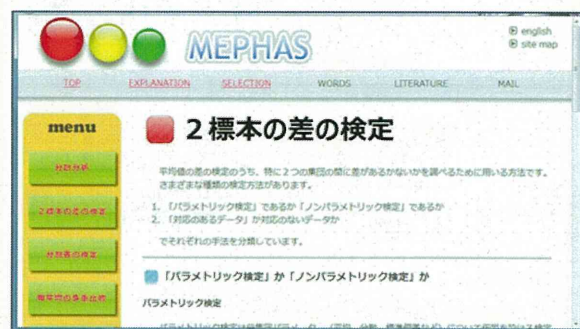


Figure 17. 更新後の MEPHAS の『2標本の差の検定』の説明ページ

Figure 17 は Figure 16 の『Explanation』のトップページ左側の『menu』から『2標本の差の検定』をクリックし移動したページである。外見上は、フレームを使っている現行の MEPHAS と同様に中央の本文の箇所のみが変化しているが、フレームを使わずに上部と左側のメニューの箇所を同一のまま中央の本文の箇所を変えた別 HTML ファイルへと移動している。また本文についても先の Figure 3 に示した現サイトの説明とほぼ同内容の説明箇所を Figure 17 に示しているが、背景

色の変更によってより見易くなっている他に、文章の題字や箇条書きなどを使用し、また行の間隔などもより広く取っており、評価に個人差はあるにしても、読み易くなっているものと期待できる。

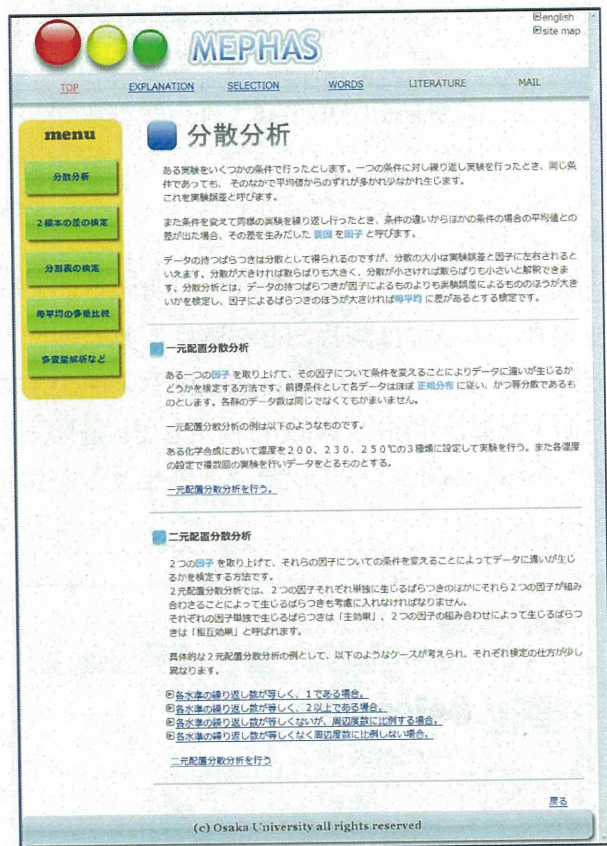


Figure 18. 更新後の MEPHAS における『分散分析』の説明ページ

Figure 18 は更新後の MEPHAS における分散分析の説明ページを示している。先の Figure 12 に示した現行の MEPHAS における説明ページと記載内容は変えずに利用した。MEPHAS における二元配置分散分析は解析したいデータの特徴に応じて4種類に分けられており、Figure 12 に示した現行の MEPHAS ではそれらの説明も一つのページにまとめられている。その説明ページでは1ページに記載されている情報量が多く、読み難いため、二元配置分散分析については各種類ごとに解説ページを分けて作成した。

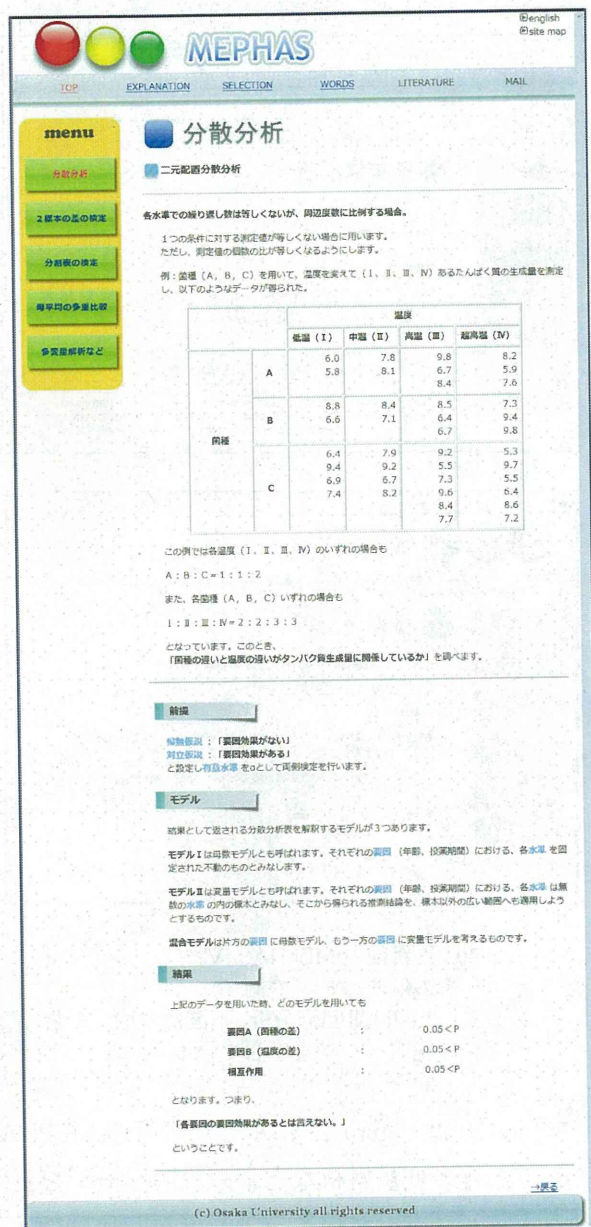


Figure 19. 更新後の MEPHAS における『二元配置分散分析』の説明ページの一例

Figure 19 は二元配置分散分析の説明ページの一例を示している。他の3種類の二元配置分散分析も同様に、解析手法に関する説明を追加しているが基本的には現行の MEPHAS における説明ページの事例や表を利用することで省力化を図りつつ、より読み易いページへの更新を行った。



Figure 20. 更新後の MEPHAS における多変量解析の解説ページの例
上図は重回帰分析、下図は主成分分析の説明ページを示している。

Figure 20 は Figure 13 に示した現行の MEPHAS における『多変量解析など』の説明ページを更新したページを示している。上図は重回帰分析、下図は主成分分析の説明ページをそれぞれ示している。Figure 13 に示すように、現行の MEPHAS では多変量解析に対する説明が十分ではなかったため、語句や解析結果の説明、事例や図表、数式なども加えて更新し、各多変量解析ごとに独立した説明ページとして作成した。

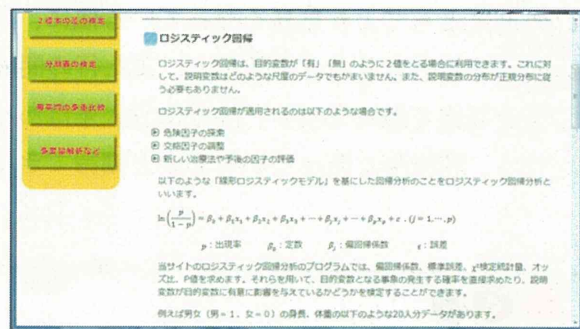


Figure 21. 更新後の MEPHAS での手法の説明のページ
多変量解析のロジスティック回帰の解説のページを取り上げている。

数式についても全てを数式で説明することは使用者によっては読むことを避ける要因の一つになり得ると思われるが、Figure 21 に示すように説明上必要な箇所については使用して、適切であり、かつ分かり易い説明になるように文章を作成した。

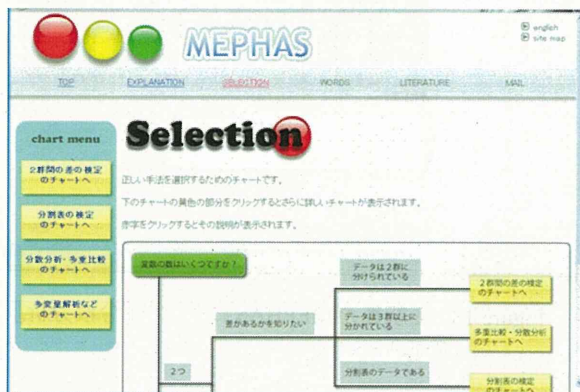


Figure 22. 更新後の MEPHAS での『Selection』のトップページ

Figure 22 は更新後の MEPHAS での『Selection』(現サイトでの『手法の選択』に該当する)のトップページである。このページにおいても先の『Explanation』のページと同様、上部に Web サイト全体のメニュー、左側に『Selection』以下のコンテンツのメニューが設けられている。また中央部の本文箇所には、現行の MEPHAS におけるカテゴリーによる分類ではなく、分類用のフローチャートを表示させている。Figure 23 はそのフローチャートを示している。

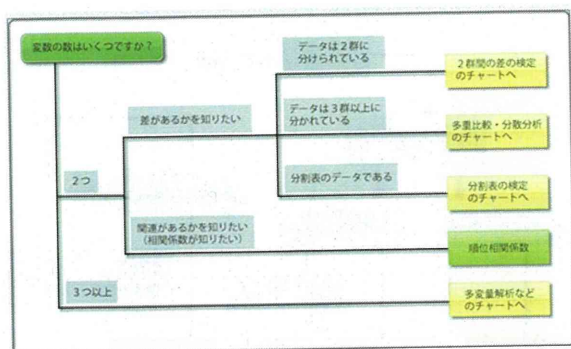


Figure 23. データの状況ごとのフローチャートを選択するためのフローチャート

Figure 23 のフローチャートは先の Figure 4, 5 に示された、使用者が解析したいデータの状況に応じた、各手法のフローチャートへの導入部分に該当する。右側の黄色のボタンから各手法のフローチャートへ、また黄緑色のボタンから解析手法のページ (順位相関係数) へと移動する。Figure 10 に示した『2群のデータに関連性があるかどうか』のフローチャートの最初の設問を Figure 23 に示されるようにフローチャート中に組み込むことにより、Figure 10 の『2群のデータに関連性があるかどうか』のフローチャートを削除し、Figure 8 の『2群のデータに差があるかどうか 分割表の場合』のフローチャート (Figure 23 では『分割表の検定のチャートへ』の移動先のフローチャートが該当する) へとまとめた。

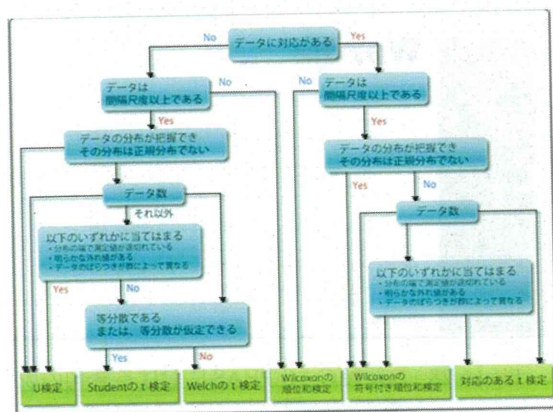


Figure 24. 更新後の MEPHAS での『2標本の差の検定』のフローチャート

Figure 24 は更新後の MEPHAS での『2標本の差の検定』のフローチャートを示している。このフローチャートは Figure 7 に示される『2群のデータに差があるかどうか 順序尺度以上の場合』のフローチャートに該当している。Figure 7 のフローチャートはできる限りブラウザー上でスクロールをせずに済むようまとめることも考えられており、その結果としてかなり入り組んでしまっている。従って、一見しただけでは開始点すら分かり難いほどであったが、Figure 24 のフローチャートでは上から下へとフローを流しており、一見して分かり易いフローチャートへと修正した。

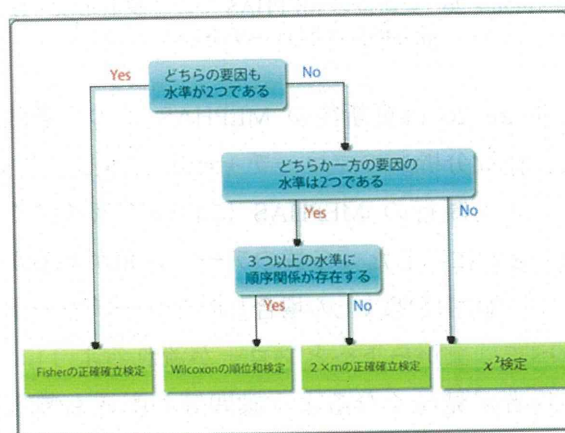


Figure 25. 更新後の MEPHAS での『分割表の検定』のフローチャート

Figure 25 は更新後の MEPHAS での『分割表の検定』のフローチャートを示している。これは先の Figure 8 に示される『2群のデータに差があるかどうか 分割表の場合』のフローチャートを修正したものである。フローチャートのデザインは先の Figure 24 に示したフローチャートと同様である。設問に関しては、Figure 8 の二つ目の設問に No の選択肢が無かったため、その選択肢を追加し、該当する手法としてカイ二乗検定の追加を行うこととした。

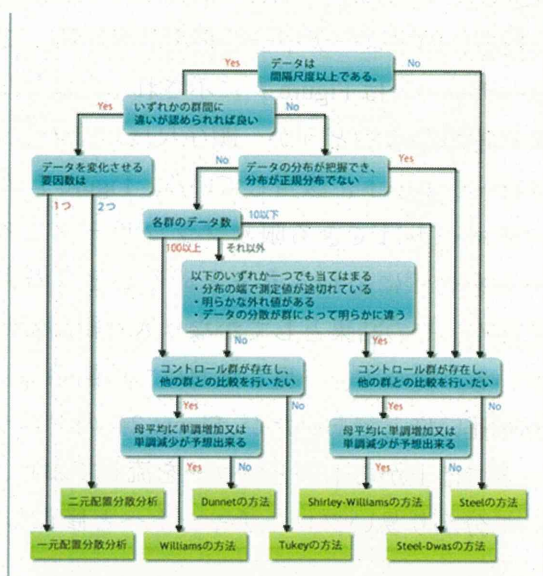


Figure 26 は更新後の MEPHAS での『多重比較・分散分析』のフローチャートを示している。こちらでも現行の MEPHAS において該当する、Figure 9 に示した『多群のデータに差があるかどうか 順序尺度以上の場合』のフローチャートに比べて、上から下へとフローが流れるようにし、使用者が見易くなるよう設問等の配置も整理した。また、Williams の方法（順序を設けることができる多群の場合に適用される、パラメトリックな多重比較法）に対するノンパラメトリックな手法である、Shirley-Williams の方法を新たに MEPHAS に追加することで、多重比較法のパラメトリックとノンパラメトリックな手法の対が構成され、フローチャートもよりまとまりが改善された。

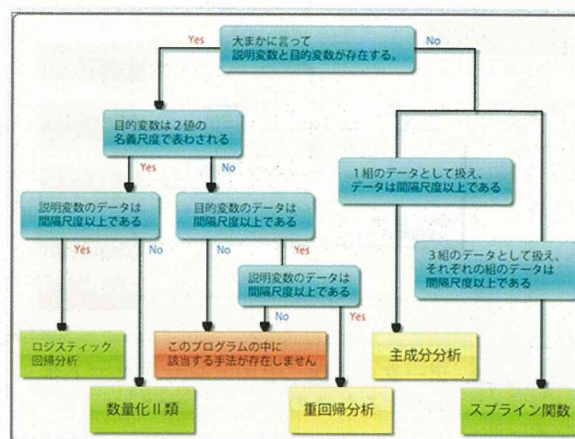


Figure 27 は更新後の MEPHAS での『多変量解析など』のフローチャートを示している。現サイトで該当する Figure 11 の『多群のデータを1組あるいは2組あるいは3組のデータとして扱う場合、そのデータに関連性があるかどうか』のフローチャートに比べ、見易さの面では向上したと考えているが、該当する手法が無いとの結果に行き着く点は必ずしも解決されていない。今後、手法を追加することにより対応し、もし手法の追加が困難である場合には適切な手法を提示することまでは行えるよう、修正が必要であろう。

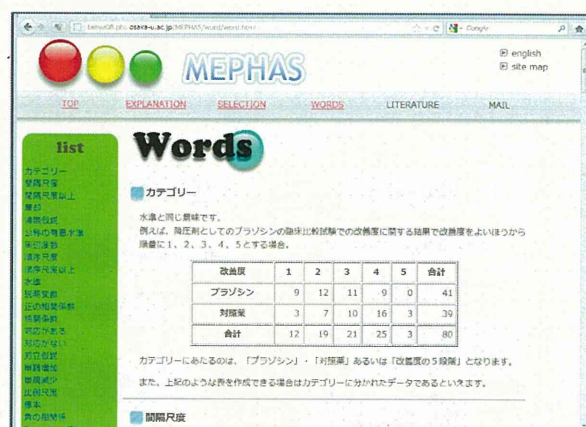


Figure 28 は更新後の MEPHAS に新たに設けた語句の説明についての一覧ページを示している。まずは現行の MEPHAS に作成されていた語句の解説を利用して作成し、現在 28 個が掲載されている。今後更新した Web ページの新規内容に合わせて、語句と説明の追加が必要である。

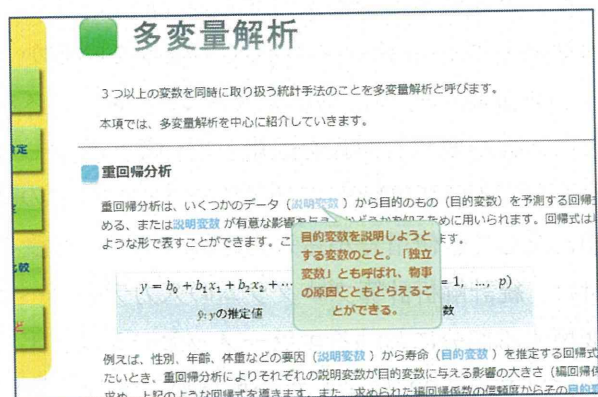


Figure 29. 語句の概要のポップアップ表示

また専門的な用語など語句の概要が手早く確認できることは内容の読み易さ、理解し易さの向上につながると考え、JavaScript を導入し、Figure 29 に示すように語句にカーソルを合わせると語句の簡単な内容を記載したポップアップが表示されるよう、更新した。

表示デザインの変更や簡単な数値処理などを Web 上で実現する手段の一つとして JavaScript を始めとする様々なスクリプト言語により作成されたプログラム（以下スクリプトと略する）が多くの Web サイトにおいて利用されている。一方で、Web ページを表示することにより動作するスクリプトはコンピュータウイルスの感染など攻撃行動に使用されることもあり[6]、スクリプトを動作しない設定において Web ブラウザーを利用しているユーザーもいる。従って、Web ページ上でスクリプトを利用する場合には、スクリプトが動作しない場合でも Web ページが利用できるような代替ページなどを Web サイト構築者が準備しておくことが重要である。そこで更新後の MEPHAS

においては、解析の利用や Web ページの閲覧など基本的な機能の利用には原則スクリプトを使用せず、基本的な機能以外に利用する際にもできる限り代替手段を準備することにした。但し、先の語句の説明に関しては、当初 JavaScript にて作成していたが、CSS (Cascading Style Sheet) で実現可能であったため、試験公開時に CSS による表示に変更した。なお、先のポップアップにより概要が表示される語句は、一覧と同様現在 28 個であり、今後も順次追加を進める予定である。

当初予定していた平成 24 年 3 月から若干遅延したが、平成 24 年 5 月にアクセス可能な範囲を限定した試験公開を実施した。この公開により判明するであろうエラーや間違いを修正した上で、平成 25 年度末に広く一般に公開する予定である。

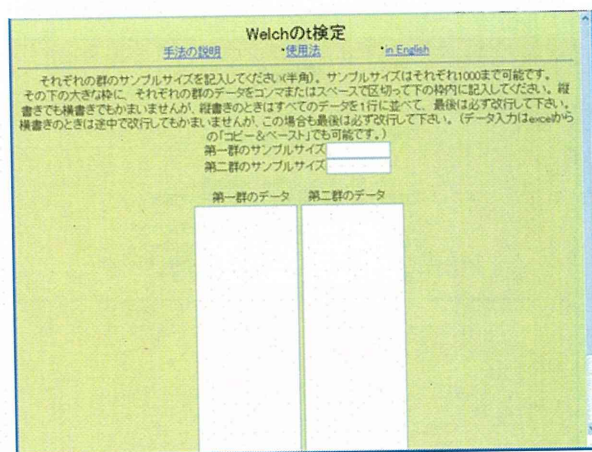


Figure 30. 現行の MEPHAS における Welch の t 検定のデータ入力画面

Figure 30 は現行の MEPHAS における Welch の t 検定のデータ入力画面を示している。上部に手法の説明や簡単な使用法のページ等へのリンクが準備されており、本文の箇所に入力項目の概説とサンプルサイズや実際の数値データを入力するためのボックス（CGI によるデータ送信のためのテキストボックスやテキストエリアなど）が用意されている。『手法の説明』は MEPHAS のコンテンツの一つである『手法の説明』の該当するべ

ージへのリンクとなっており、『使用法』のページは Figure 31 に示すように、先の Figure 30 に示された入力用のページにサンプルデータが入力され、少し説明が追加されている。

Figure 31. 現行の MEPHAS における Welch の t 検定の使用法のページ

Figure 32. 現行の MEPHAS におけるデータ入力時の間違いに対するエラーメッセージの画面

Figure 32 は現行の MEPHAS においてデータを入力し間違いがあった際に、そのまま計算はせず使用者（入力者）の側へ間違いがあったことを知らせるためのエラーメッセージのページを示している。MEPHAS に限らず、計算結果を返すアプリケーションにおいて注意すべきことのひとつとして、使用者が意図した入力と違うデータセットを基にした計算結果を返してしまうことが挙げ

られる。何らかの計算結果が返ってきた際に使用者がその結果を疑うことはまずあり得ないため、システム側で使用者の意図したデータ入力を受け取れているのかどうかをチェックすることが必要になる。従って MEPHAS においても、入力される数値データのみあればサンプルサイズや群数はその数値データから求めることが可能であり、Web サイトでの入力は必要ないが、使用者の意図しているデータの形式となっているかどうかの確認のためにも群数やサンプルサイズなどの入力を要する形式を採用している。しかしながら現行の MEPHAS ではエラーメッセージの種類に乏しく、使用者から入力されたデータを受け付け、統計解析計算プログラムの入力データへと変換する部分におけるエラーメッセージは、実質的には Figure 32 に示した一種類のみであった。

そこでまず、使用者の側がエラーメッセージを受け取った際に改善を図り易くするために、入力されたデータのチェックにおいて間違いを見つけた際のエラーメッセージのレポートリーを増やすようプログラムの修正を行うことにした。

Figure 33. 現行の MEPHAS における二元配置分散分析の要因数設定画面

Figure 33 は現行の MEPHAS における二元配置分散分析の両要因における水準数を設定する画面を、Figure 34 はその設定後に表示されるデータの入力画面を示している。

二元配置分散分析
各水準の繰り返し数が等しいが、周辺度数に比例する場合
データを半角数字で入力してください。
データ数は100個まで計算可能です。

要因B		水準 1	水準 2	水準 3	水準 4	水準 5
要因A	水準 1					
	水準 2					
	水準 3					
	水準 4					
	水準 5					

書き込み クリア

Figure 34. 現行のMEPHASにおける二元配置分散分析のデータ入力画面（両要因の水準数5の場合）

Figure 34 に示すように、Figure 33 の要因数の設定画面において設定した両要因の水準の数だけ、データ入力画面においてテキストボックスの入力フォームが作られる仕様となっており、水準の数によってはテキストボックスが非常に多くなり入力が煩雑になる。

多重検定: Dunnettの方法
手法の説明 使用法 in English

コントロール群の平均値を μ_c 、処置群の平均値を μ とすると、どの対立仮説を証明したいのか選んでください。
☐ $\mu_c \neq \mu$ ☐ $\mu_c > \mu$ ☐ $\mu_c < \mu$

検定したい群の数を記入してください。(半角で最大10群まで)

群数B

各群のデータ数をコンマカスベースで区切って記入して下さい。それと同じ順に下の大きな枠に、データを縦一列に記入して下さい。(必ず左端にコントロール群のデータを入力して下さい) 群と群の間はコンマカスベースで区切って下さい。最後は必ず改行して下さい。それぞれの群のデータ数は異なっても構いません。データ数はそれぞれ1000まで可能です。結果のページでは左からコントロール、第二群、第三群...と呼んでいます。(データ入力にはexcelからの「コピー&ペースト」でも可能です。)

各群のデータ数(10 10 10 10 10 10 10)

9.40	7.44	10.48	12.55	12.40
18.47	9.90	8.59		
5.89	9.53	12.19	7.89	8.62
	6.62	10.74	8.00	
6.31	8.04	8.45	5.78	8.86
	9.19	9.49	6.08	
10.27	9.27	9.95	9.26	12.89
	8.89	10.07	8.74	
9.69	8.87	13.84	11.73	14.75
	9.69	12.14	10.15	
13.32	6.78	11.08	11.80	13.84
	9.69	11.08	9.78	
8.95	11.35	9.24	11.52	7.11
	9.51	9.99	7.89	
8.96	9.27	9.84	10.06	9.95
	14.39	11.87	8.15	
6.52	8.53	4.84	12.90	7.44
	8.89	11.84	8.70	

書き込み クリア

Figure 35. テキストボックス内へのデータ入力における表示のズレ

また多変量解析や多重比較法など、データが複数行、複数列に渡るデータ形式の場合、Figure 35 に示すようにデータ入力画面ではテキストボックス内にデータを貼り付けるため、折り返しや表示位置のズレが起こることがある。この点はMEPHAS 使用者より、解析において意図した通り

にデータが取り扱われているかどうか不安があるとの指摘を受けていた。

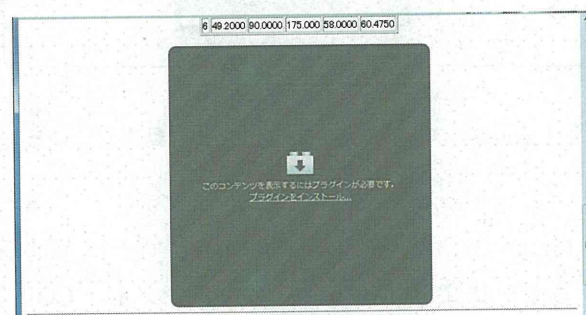
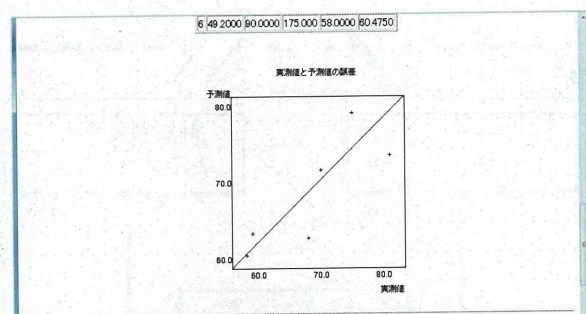


Figure 36. 現行のMEPHASにおける重回帰分析実施結果のグラフ表示（上図は表示可能な環境での結果、下図は表示不可能な環境での結果を示す）

多変量解析手法である重回帰分析や主成分分析では Figure 36 の上図に示すように計算結果の一部をグラフ表示により示しているが、その際には Java アプレットが使用されている。Figure 36 の下図に示すように使用者の Web ブラウザーの環境設定によっては動作しないこともあるため、可能であれば何らかの対応が必要であろうと判断された。

現行の MEPHAS における統計解析手法の計算を実施する際に動作しているプログラムについて整理したものが Figure 37 である。

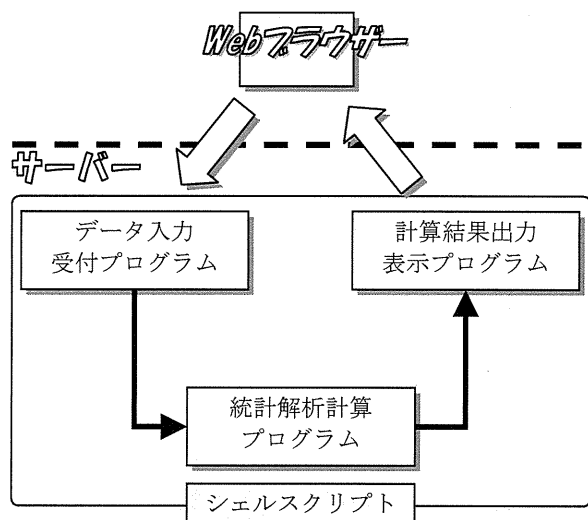


Figure 37. 現行のMEPHASにおいて動作しているプログラム群の概要図

Figure 37 に示されるようにユーザーは Web ブラウザーにより MEPHAS へアクセスし、CGI を通じてサーバー側へデータを送信する。サーバー側では3つのプログラムがシェルスクリプトにより1つにまとめられ、使用者からのデータを処理し、計算を実施し、計算結果を出力する。この中でデータ入力受付プログラムと計算結果出力表示プログラムはC言語もしくはPerlにより、統計解析計算プログラムはFortran（主にFortran77形式）により作成されている。

Fortran は科学技術計算分野で重用されてきたプログラミング言語であり、定評ある計算ライブラリーの豊富さなどから、現在においても科学技術計算分野で使用されている言語である。しかしC言語（C++を含む）やJavaなど、汎用的に使用されるプログラミング言語においても計算ライブラリーの提供が進んでいる現状においてはFortranの優先性は高くない。さらに近年ではオープンソースな統計解析システムであるRが、統計解析やデータマイニング分野における適用範囲の広さから研究における使用も増えている。

RはそのベースシステムがGNU GPL version 2に沿って提供されているオープンソースの統計

解析システムである。パッケージを追加することでベースシステムに解析手法を追加することが可能であり、世界中で様々な解析手法やデータ処理、グラフ表示などのパッケージが開発され、提供されている。またS言語のクローンであるR言語の実行環境でもあり、多様な解析手法を追加可能なパッケージ群と相まって、統計解析手法の開発、実行環境として非常に有用である。本研究課題のMEPHASでは統計初学者が簡単に利用でき統計を学ぶことに目的が置かれるため、多様な統計解析手法の提供は必須ではないが、R言語、R実行環境での解析計算ができるよう統計解析計算プログラムを置き換えることは、今後のメンテナンスや更新を考慮した際に、その言語の発展性の高さから有意義な選択であると考えられた。そこで、Fortranで作成されている統計解析計算プログラムを順次R言語を利用した計算プログラムへと更新した。

また、WebサイトのHTMLファイルやコンテンツを中心に改訂を行っている分担研究と連携し、多重比較法の一つであるShirley-Williamsの方法、また分割表の検定の一つであるカイ二乗検定をMEPHASで取り扱う解析手法の一つとして新たに追加できるよう、プログラムの作成を行った。