

設定

比較

と

多変量解析

3つ以上の変数を同時に扱う統計手法のことを多変量解析と呼びます。

本項では、多変量解析を中心に紹介していきます。

重回帰分析

重回帰分析は、いくつかのデータ（**説明変数**）から目的のもの（**目的変数**）を予測する回帰式を求める、または**説明変数**が有意な影響を与えているかどうかを知るために用いられます。回帰式は以下のような形で表すことができます。この式は、**目的変数を説明しようとする変数のこと。「独立変数」とも呼ばれ、物事の原因とともとらえることができる。**を説明します。

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

$\hat{y}$: yの推定値

$p = 1, \dots, p$

例えば、性別、年齢、体重などの要因（**説明変数**）から寿命（**目的変数**）を推定する回帰式を求めたいとき、重回帰分析によりそれぞれの説明変数が目的変数に与える影響の大きさ（偏回帰係数）を求め、上記のような回帰式を導きます。また、求められた偏回帰係数の信頼度からその**目的変数**の信頼度を評価します。

図 8. MEPHAS 新規サイトにおける語句の概要のポップアップ表示

[English](#)
[site map](#)



[TOP](#) [EXPLANATION](#) [SELECTION](#) [WORDS](#) [LITERATURE](#) [MAIL](#)

menu

分散分析

[一元配置分散分析](#)

[二元配置分散分析](#)

2標本の差の検定

[Studentのt検定](#)

[Welchのt検定](#)

[Wilcoxonの順位和検定](#)

[対応のあるt検定](#)

[Wilcoxonの符号付順位和検定](#)

分割表の検定

[Fisherの正確確率検定](#)

[カイ2乗検定\(2x2\)](#)

[カイ2乗検定\(2xm\)](#)

[カイ2乗検定\(mxn\)](#)

母平均の多重比較検定

[Tukeyの方法](#)

[Dunnnettの方法](#)

[Williamsの方法](#)

[Steel-Dwassの方法](#)

[Steelの方法](#)

[Shirley-Williamsの方法](#)

多変量解析など

[順位相関係数](#)

[重回帰分析](#)

[ロジスティック回帰分析](#)

[数量化二類](#)

[主成分分析](#)

二元配置分散分析

[手法の解説](#)・[使用法](#)・[in English](#)

各要因の水準数を選択してください。

要因A

要因B

各データを下のテキストボックスに入力してください。

- データは各要因のラベルをつけて入力してください。ラベルには半角数字を使用します。詳しくは使用法の解説を参考にしてください。
- データと各ラベルの区切りにはコンマかスペースを用いてください。
- 各データは改行で区別してください。

(c) Osaka University all rights reserved

図 9. MEPHAS 新規サイトにおける二次元配置分散分析のデータ入力ページ

	第1群			第2群	
	データ	順位		データ	順位
1	5.70	15.00	1	4.72	6.00
2	3.82	1.00	2	5.77	16.00
3	4.15	2.00	3	5.02	7.00
4	6.13	19.00	4	5.63	13.00
5	5.81	17.00	5	5.49	12.00
6	7.66	30.00	6	4.39	4.00
7	5.48	11.00	7	6.68	25.00
8	4.48	5.00	8	5.64	14.00
9	4.26	3.00	9	5.26	9.00
10	6.76	28.00	10	6.47	23.00
11	6.69	26.00	11	6.32	21.00
12	6.14	20.00	12	5.09	8.00
13	6.86	29.00	13	5.36	10.00
14	6.71	27.00	14	6.64	24.00
15	6.00	18.00	15	6.45	22.00
合計		251.00	合計		214.00
U1		94.00	U2		131.00

図 10. MEPHAS 新規サイトにおける計算結果におけるデータ確認表示ページ

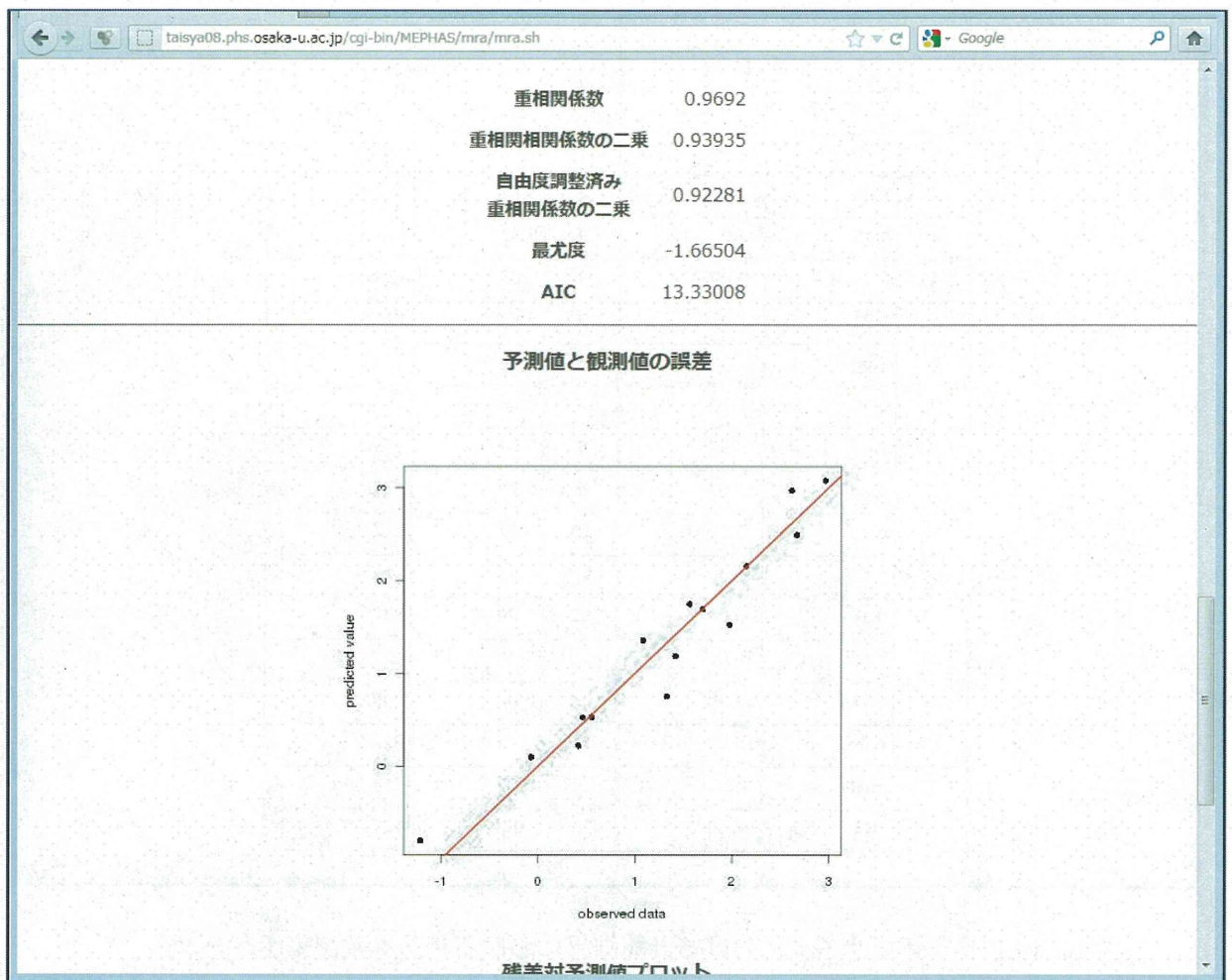


図 11. MEPHAS 新規サイトにおける重回帰分析におけるグラフ表示ページ

研究分担者 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

研究分担者 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

本事業において修正の基としている MEPHAS は 2002 年から公開されている Web サイトであり、本 Web サイト上において Web ブラウザーを通じて統計解析手法を利用可能である。Web サイトのコンテンツとしては統計解析手法の利用の他に、それら解析手法の簡単な説明と適切な手法の選択を補助するフローチャートなどがある。しかしそれらのコンテンツは、試験公開時点である 10 年以上前のデザインと技術で作成された Web ページを現在も用いており、現在の環境下で利用し易いとは言い難い。そこで本分担研究では、MEPHAS に設けられているコンテンツの内、特に統計解析手法の説明と選択補助についてより見易く、より使い易くなるよう修正することにより、それら説明と選択補助のコンテンツがより活用されるようになれば、MEPHAS 利用者の統計解析に対する理解の促進に寄与すると考え、修正、作成を行うことにした。以下、平成 23 年度に取り組んだ内容について報告する。

A. 研究目的

本研究において改訂を進めた Web サイト、MEPHAS[1] では統計解析プログラムを提供するにあたり、Common Gate Interface (CGI) を用いている。CGI を通じて使用者が解析したいデータをサーバー側へ送信し、サーバー側では統計計算を実施し、その結果を使用者の Web ブラウザーへと返す。但し、使用者がサーバー側へ提供するデータは数値のみで、その数値が何を示したのかという情報をサーバー側へ提供する必要はない。つまり、公表前の研究結果など他者へ公開することが躊躇われる内容のデータであったとしても、その数値のみから全容を知ることが不可能であれば、内容の漏洩による損失を気にすることなく Web を通じてデータを送受信可能である。一方で、解析を行うことで示したい内容を踏まえた上で、適切な解析手法をサーバー側で自動的に選択するわけではないため、解析手法の選択を行わな

ければならないのは使用者の側である。従って、既存の一般的な統計解析アプリケーションなども同様であるが、統計解析を専門としない使用者が利用する際においても、解析手法を適切に選択できるような説明や工夫の重要性は高い。そこで、昨年度は、MEPHAS で以前より手法選択を補助するための工夫の一つとして提供していたフローチャートを見直し、より分かり易く、利用し易いものへと更新した。また、Web サイト全体のデザインやメニュー表示などについても検討、修正を行い、MEPHAS 利用時のユーザビリティの向上を図った。

平成 23 年度は前年度に決定した Web サイトの修正方針に沿った更新内容を全てのページにて行い、試験公開（まずはアクセス可能な範囲を限定して公開する）に向けて準備を進めると共に、昨年度から引き続いて MEPHAS 利用時のユーザビリティの向上を目的に、各解析手法の説明に

関して事例や図表などを利用した内容への更新を実施した。

B. 研究方法

昨年度、ユーザビリティの向上を念頭に改善が必要と思われる点を現行の MEPHAS から洗い出し、以下の四点を重点的な修正が必要と判断し、修正の方向性を決定した上で、Web ページのデザイン更新内容を決めた。

- 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

今年度は、昨年度実施した 1)と 2)のデザインやメニュー表示の改訂を全てのページにおいて実施し、試験公開に向けた準備を中心に進めた。また上記の 3)に関し、昨年度は解説ページの読み易さの向上としてデザイン面の修正が中心になったため、今年度は事例や図表などの追加といった、内容面での更新を中心に実施した。また現行の MEPHAS には手法の解説ページとは別に、いくつかの用語についても簡単な解説が設けられており、それらについても読み易さ、利用し易さを向上するべく、修正を行った。

なお、MEPHAS の Web ページの中で、実際の統計解析手法を実施するためのデータ送信のページと結果表示を行うためのページに関しては、サーバー側のプログラム群との関連もあるため、CGI を介したデータ授受や統計解析の計算プログラム等の修正の検討を行う別の分担研究において実施することにした。

（倫理面への配慮）

本研究の実施において取り扱うものは数値情報のみであり、それら数値情報も個人の特定につ

ながり得る情報は含まれないため、倫理面への配慮等は必要ない。

C. 研究結果



Figure 1. 現行の MEPHAS トップページ



Figure 2. 更新後のサイトのトップページ

更新後のサイト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サイト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 3. 更新後のサイトのコンテンツの一例

現行の MEPHAS のトップページを Figure 1 に、昨年度更新後の MEPHAS 新規サイトにおけるトップページを Figure 2 に、更新後のサイト内のコンテンツの一例（解析手法を選択する際のフローチャートのページ）を Figure 3 にそれぞれ示した。なお、更新後のサイト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サイト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。更新後の MEPHAS における基本的な Web ページデザインは、Figure 2 と 3 に示すように、MEPHAS 内の各コンテンツへのリンクを示すナビゲーションバーをページ上部に、Web ブラウザーに表示されている Web ページと同一のコンテンツに属するページへのリンクをページ左側に表示するデザインへと統一し、範囲を限定した試験公開に向けて各ページの更新を進めた。

範囲を限定した試験公開については平成 23 年度末を予定し進めてきたが、残念ながら 3 月には間に合わず、報告書提出期限の 5 月末に延期し、更新を進めた。平成 24 年 5 月末現在では、大阪大学内からのアクセスのみに限定し、現行の MEPHAS の稼働を阻害しないよう、別の Web サイトとして公開[2]している。この試験公開において、記載事項の誤りやプログラム更新に伴うエラーなどを可能な限り抽出、修正し、平成 24 年度

末には範囲を限定しない試験的な公開を実施する予定で引き続き作業を進めている。

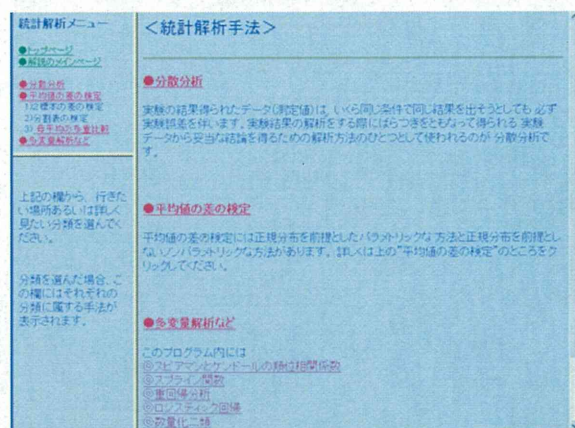


Figure 4. 現行の MEPHAS における『手法の解説』ページ



Figure 5. 更新後のサイトの『Explanation』のトップページ

内容は Figure 4 と同等である。

Figure 4 は、現行の MEPHAS における手法の解説ページの一例を示しており、Figure 5 は昨年度の更新後のページを示している。先述の通り、ナビゲーションバーを上部に、同一コンテンツページへのリンクを左側に備えている。

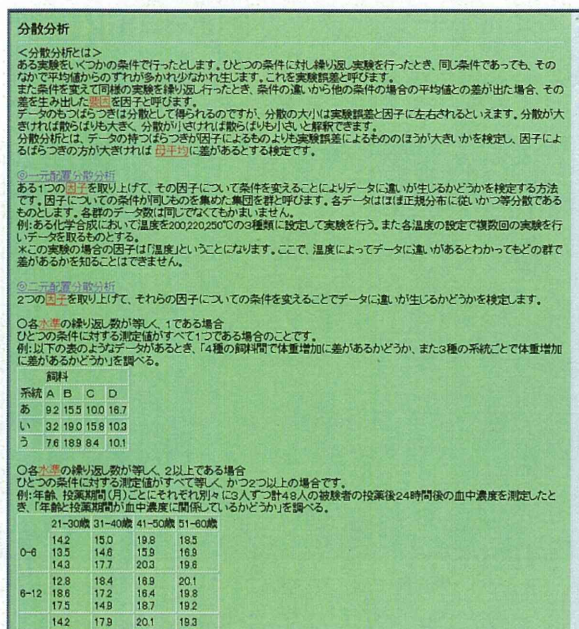


Figure 6. 現行の MEPHAS における『分散分析』についての説明ページ
Web サイト上ではフレームの中で表示される。

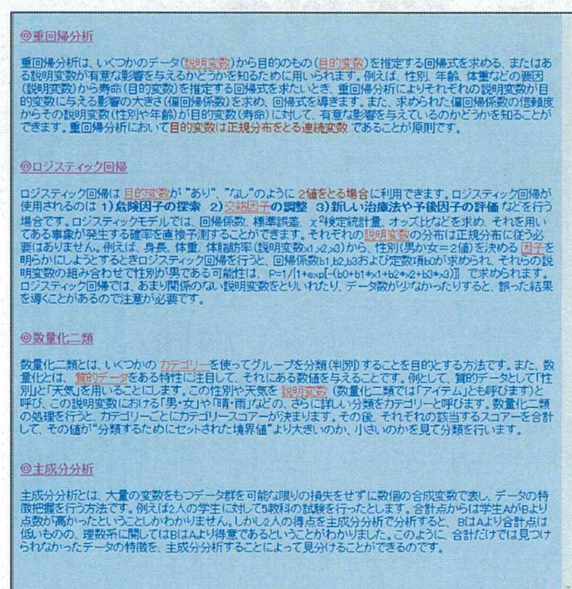


Figure 7. 現行の MEPHAS における『多変量解析など』についての説明ページ
Web サイト上ではフレームの中で表示される。

Figure 6は現行の MEPHAS における分散分析を、Figure 7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 7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と思われる。従って、Figure 6 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。

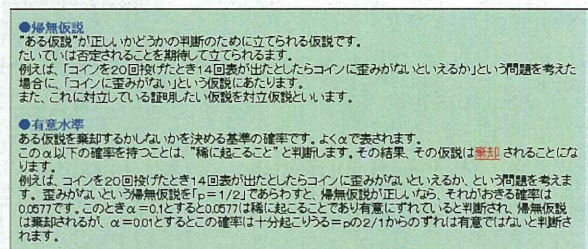


Figure 8. 現行の MEPHAS における語句に関する説明ページの一例

Figure 8 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。

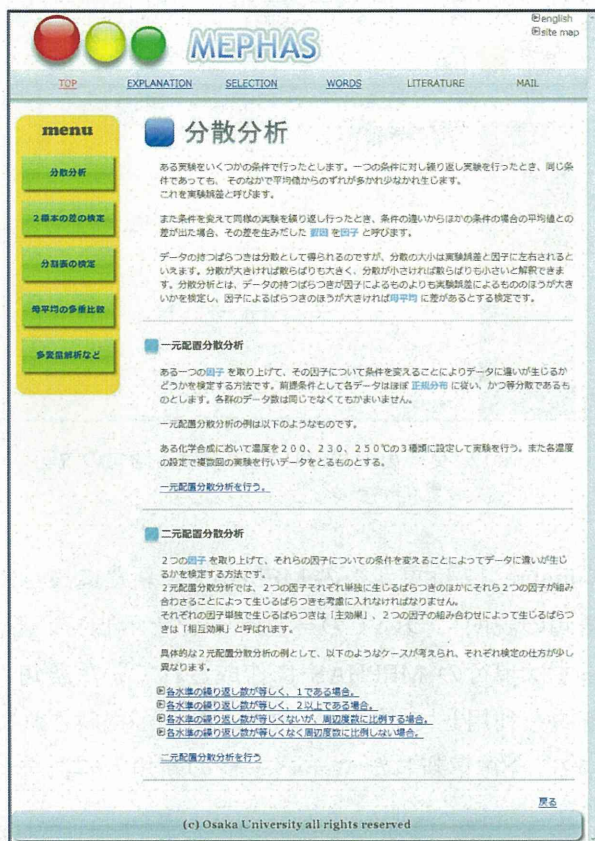


Figure 9. 更新後の MEPHAS における『分散分析』の説明ページ

Figure 9 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 3 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 6 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。

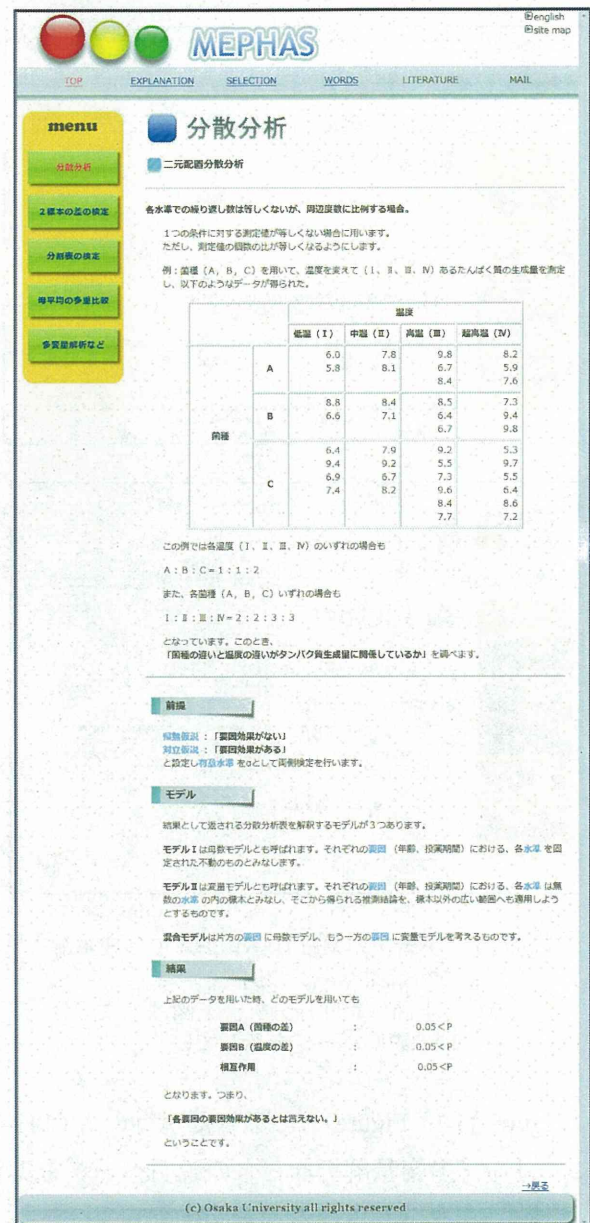


Figure 10. 更新後の MEPHAS における『二元配置分散分析』の説明ページの一例

Figure 10 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 11. 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 11 は Figure 7 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 7 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 12. 更新後のMEPHASにおける語句の説明一覧のページ

Figure 12 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。

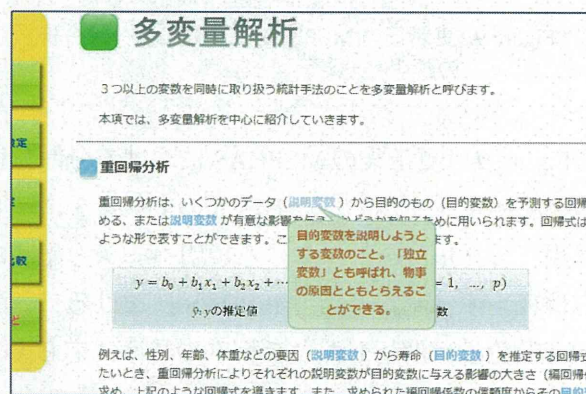


Figure 13. 更新後のMEPHASにおける語句の概要説明のポップアップ表示

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 13 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、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[3]、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。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、一覧と同数の現在 28 個であり、今後順次追加を進める予定である。

D. 考察

まず、試験公開に向けた作業として、昨年度実施したユーザビリティ向上のためのデザイン更新を MEPHAS の全てのページに適用した。試験公開は、アクセス可能な範囲を限定した上で実施期間の最後である平成 24 年 3 月末から開始の予定であったが、遅れてしまい報告書作成時（平成 24 年 5 月末）の公開となった。現在は大阪大学内からアクセスし利用することが可能である。この試験公開の期間中に解析実行時のエラーや Web ページの誤りなどを可能な限り低減させる予定である。なおアクセス可能な範囲を限定しない試験公開は平成 25 年 3 月末から開始する予定で、現在も作業を進めている。

次に MEPHAS において利用可能な各解析手法の説明ページについて、昨年度はデザインの更新のみに留まっていたため、平成 23 年度は事例や図表を追加するなど、内容の更新に努めた。更新に際しては、現行の MEPHAS において用いられている事例や図表を可能な限り利用することで省力化に努めると共に、それらの無い説明箇所については事例や図表等を取り入れることを検討し更新した。事例の紹介や図表による説明は解析手法自体の理解を促進すると共に、その手法により解析されるデータをイメージし易くなるため、昨年度のフローチャートの更新と同様、使用者の持つデータに対して適切な解析手法を選び易くする一助となることが期待できる。これまでの更新では解析手法を選択するためのフローチャートなど解析前の利用し易さの向上を中心に検討、実施してきたため、今後は MEPHAS による解析結果を利用した説明など、解析後の活用、つまり解析手法を利用して得られた結果を適切に用いる点についての説明に関しても充実を図る必要がある。

最後に、現行の MEPHAS において該当する語句のリンクにより別ウインドウへ表示することで提供されている語句の説明について、それら説明が利用し易くなるよう検討し更新を行った。語句の説明については、再確認としての使用頻度が高く、より簡便かつより即時的に内容を確認できることがユーザビリティの向上につながると判断し、ポップアップにより語句の説明を表示できるよう更新した。当初 JavaScript の利用を検討し導入したが試験公開時には CSS による表示に変更した。語句の説明についても解析手法の説明と同様、今後順次追加することで内容の充実を図る必要があるだろう。

昨年度の追加実施事項として挙げた英語版 Web サイトの構築も、先の更新に並行して進めている

(プログラムの計算結果も含め、文章の英訳中であり、まだ成果としては出ていないため結果からは割愛した)。

E. 結論

平成 23 年度に実施し、得られた成果としては以下の通りとなる。

- a. 昨年度決定した Web サイト修正内容に沿った Web ページ更新の作業実施と試験公開に向けた準備(若干遅れたが、予定通りアクセス可能な範囲を限定した試験公開まで完了)
- b. 手法の説明における事例や図表等の追加による更新
- c. MEPHAS 内の用語の説明法の改訂と用語説明の一覧ページの作成

引用文献

- [1]. MEPHAS, <http://www.gen-info.osaka-u.ac.jp/MEPHAS/>
- [2]. MEPHAS (試験公開版), <http://taisya08.phs.osaka-u.ac.jp/MEPHAS/> (但しアクセスは大阪大学内のみ可能)
- [3]. 情報処理振興事業協会セキュリティセンター, 「スクリプトウイルスに関する報告書」 2001 年 2 月, <http://www.ipa.go.jp/security/fy12/contents/virus/report/script-rep.PDF>, last accessed: 2012/05/17

F. 研究発表

1. 論文発表

特になし。

2. 学会発表

福井大介¹, 田雨時², 岡本晃典², 川下理日人^{2,3}, 後藤直久³, 安永照雄³, 高木達也^{2,3} (¹ 阪大薬, ² 阪大院薬, ³ 阪大微研), 医薬学データ用統計解析プログラム、MEPHAS の更新について(その 2),

30P2-am115, 日本薬学会 第 132 年会, 札幌, 2012 年 3 月

G. 知的財産権の出願・登録状況

1. 特許取得

特になし。

2. 実用新案登録

特になし。

3. その他

特になし。

研究分担者 高木 達也 国立大学法人大阪大学大学院薬学研究科 教授

研究分担者 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

本研究において開発の対象としている MEPHAS は Common Gate Interface (CGI) によりデータの授受を行う統計解析プログラムパッケージであり、それらを可能としているプログラム群はサーバー側にて稼働している。それら現在のプログラム群は 2002 年の試験公開から稼働しており、使用者からの指摘を中心にプログラムの見直しを適宜実施してきた。しかしながらそれら見直しの多くは、計算実行上のプログラミングミスや表示におけるミスなどであった。そこで本分担研究では MEPHAS における使用者の利便性の向上による MEPHAS 利用の促進、ひいては統計解析への興味を促進するよう、プログラム群を検討し直し、修正を行うことを目的に実施し、解析結果における入力データの確認やグラフ表示に関するプログラムの修正を行った。また、試験公開とそれ以降のメンテナンス作業などを行い易くするためのプログラム更新も実施した。以下、実施内容を報告する。

A. 研究目的

現行の MEPHAS[1] 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 [2] 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提供している。サーバー側におい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、Fortran などのプログラミング言語を用いて作成されている。昨年度は Web サイトの更新に伴い必要となった新たな機能や解析手法のプログラムの作成を中心に実施した。

平成 23 年度も昨年度と同様 Web サイト更新に伴う新規プログラムの作成を行うと共に、試験公開後の維持管理を念頭にプログラムの更新を行うこととした。MEPHAS を使用する目的の大部分は統計解析を実施することであるため、解析を行った際に得られる結果の内容やその分かり易さはユーザビリティの大きな部分を占めている

と考え、特に解析の結果表示に関わる箇所を中心にプログラムの見直しを実施することを目的とした。さらに試験公開に向けて、またメンテナンスや今後のサイト更新の可能性も考慮し、解析手法の計算プログラムをより多用されているプログラミング言語へ置き替えることも検討することとした。

B. 研究方法

まずは現行の MEPHAS において解析の実施に関わるページである、データ入力画面と解析結果の表示画面において改善すべき点やユーザビリティの向上につながるであろう点を検討した。それらの改善を実現するための手段を検討し、実際の変更の実施の順に進めた。

プログラミング言語の置き換えについては、多用されておりかつ多彩な解析手法に対応可能であること、更新の頻度も多く今後の発展が十分に

期待されることなどから、利用ライセンス面も含めた検討の上、オープンソースの統計解析システムである R と R 言語[3]を利用することにし、順次プログラムの更新を進めた。

昨年度から引き続き、デザイン等も含めたユーザビリティ向上を目的とした Web サイトの見直しを検討する別分担研究における検討の結果に基づき、プログラム等に関して追加、あるいは修正が必要になった事項については適宜、対応を行った。

(倫理面への配慮)

本分担研究は、Web サイトとその統計解析手法の計算機能を提供するサーバーにおいて動作しているプログラム群の修正や、新規プログラムの作成を主に実施するものであり、そこで使用されるものは数値データのみであるため、倫理面での特別な配慮を必要とする研究ではない。

C. 研究結果

Figure 1. 現行の MEPHAS における Wilcoxon の順位和検定のデータ入力画面

Figure 1 は現行の MEPHAS における Wilcoxon の順位和検定のデータ入力画面を示している。一方、Figure 2 は昨年度実施した Web サイトのデザイン更新に伴い修正したデータ入力画面を示している。

Figure 2. 新規サイトにおけるデータ入力画面

Figure 1 に示すように、このページから移動できるリンク先はこの解析手法の使用法など一部のページに限られていたが、Figure 2 に示すように昨年度の更新時にトップページ等へのリンクを含むナビゲーションバーを各ページに配置するよう更新した。Figure 1 と 2 は「Wilcoxon の順位和検定」を例に示したが、他の多くの解析手法の入力画面についてもデザイン上の修正はあるものの、解析手法を実行する上で必要となる項目は同一であり、その点は変化していない。

Figure 3. 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 3 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 4 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 4. 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 4 に示すように、Figure 3 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 とするとき、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マンドスペ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) 群と群の間はコマンドスペ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 10 10 10

9.40	7.44	10.45	12.55	12.40
13.47	9.50	9.85		
5.89	9.59	12.19	7.89	8.82
6.82	10.74	9.00		
9.31	8.04	8.45	5.78	8.66
6.19	9.40	6.06		
10.27	9.27	9.35	9.26	12.69
9.83	10.07	8.74		
9.83	9.87	13.84	11.73	14.75
8.69	12.14	10.15		
13.32	6.78	11.08	11.80	13.84
9.83	11.08	9.78		
9.95	11.35	9.24	11.52	7.11
9.31	8.39	7.99		
9.96	9.27	9.84	10.08	9.35
14.39	11.97	8.15		
9.52	8.53	4.84	12.90	7.44
8.88	19.95	6.70		

書き込み クリア

Figure 5. 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 5 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。

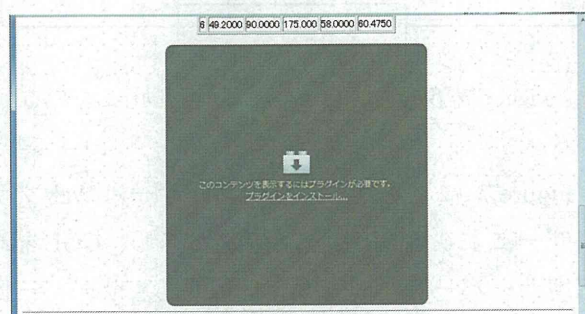
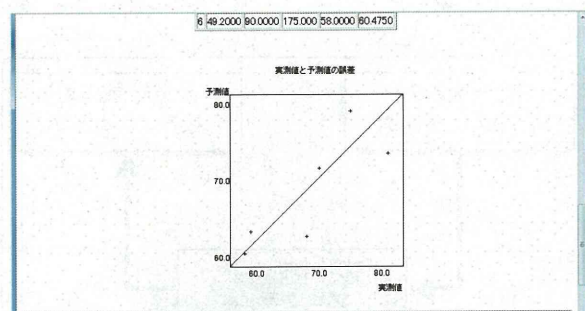


Figure 6. 現行の MEPHAS における重回帰分析実施結果のグラフ表示（上図は表示可能な環境での結果、下図は表示不可能な環境での結果を示す）

多変量解析手法である重回帰分析や主成分分析では Figure 6 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 6 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 7 である。

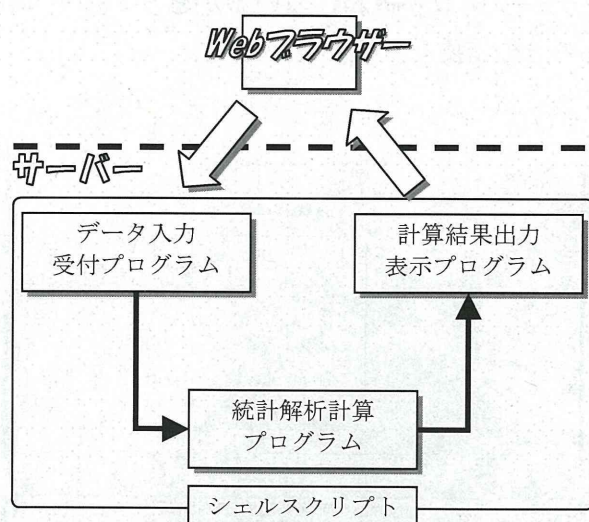


Figure 7. 現行の MEPHAS において動作しているプログラム群の概要図

Figure 7 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便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、更新を行っている分担研究と連携し、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。

因子A	因子B	結果
24.8	1	1
25.8	1	1
24.1	1	1
25.1	1	1
22.5	2	1
22.5	2	2
25.8	2	1
25.8	2	1
25.4	2	2
25.4	2	2
25.5	2	2
25.5	2	2

Figure 8. 更新後の MEPHAS における二元配置分散分析のデータ入力画面

二元配置分散分析におけるデータ入力画面の更新については、Figure 4 に示した多数のテキストボックスへの入力を、Figure 8 に示すように一

つにまとめることで簡便に済むよう、入力データに区分のためのラベルを含める形式へ変更した。またこのようなラベルを用いた入力形式は MEPHAS 中では新規であるため、入力のための説明事項を追加した。

第1群			第2群		
	データ	順位		データ	順位
1	5.70	15.00	1	4.72	6.00
2	3.82	1.00	2	5.77	16.00
3	4.15	2.00	3	5.02	7.00
4	6.13	19.00	4	5.63	13.00
5	5.81	17.00	5	5.49	12.00
6	7.66	30.00	6	4.39	4.00
7	5.48	11.00	7	6.68	25.00
8	4.48	5.00	8	5.64	14.00
9	4.26	3.00	9	5.26	9.00
10	6.76	28.00	10	6.47	23.00
11	6.69	26.00	11	6.32	21.00
12	6.14	20.00	12	5.09	8.00
13	6.86	29.00	13	5.36	10.00
14	6.71	27.00	14	6.64	24.00
15	6.00	18.00	15	6.45	22.00
合計	251.00		合計	214.00	
U1	94.00		U2	131.00	

Figure 9. 解析結果の出力ページにおける入力データの確認画面

Figure 5 に示した複数行、複数列からなる入力データにおける表示位置のズレに関しては、Excel や Google スプレッドシートのようなセル形式のデータ入力法を活用することも検討した。しかしながら、Web ページ上で動作するスクリプトは使用者の Web ブラウザの環境に依存して動作しない場合があることを考慮し、スクリプトの使用は機能に直接関わらない箇所では動作しない場合の代替が確保できる箇所のみに限定することが妥当と判断し、スクリプトが必要となるセル形式でのデータ入力は設けないことにした。そこで、入力画面の改善ではないが、Figure 9 に示すように解析結果の出力表示において、その統計解析計算プログラムで計算されたデータを確認できるよう、確認画面を作成した。

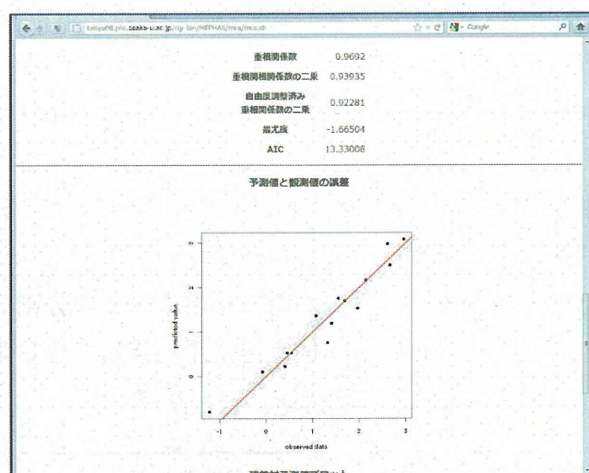


Figure 10. 更新後の MEPHAS における重回帰分析実施結果のグラフ表示

また解析結果におけるグラフ表示については、Figure 6 に示した Web ブラウザの環境による表示エラーを避けるため、該当する統計解析計算プログラム（現状は重回帰分析と主成分分析の計算プログラム）においてグラフ画像を作成し、Figure 10 に示すように Web ページ上に表示させるように変更した。またそれに合わせて、Figure 11 に示すようにデータ入力画面においてグラフ画像の有無について設定できるオプションを追加した。

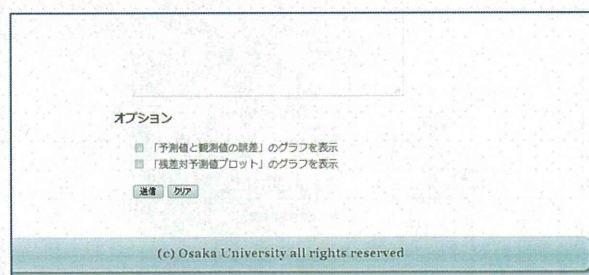


Figure 11. 更新後の MEPHAS における重回帰分析のグラフ表示に関するオプション設定

MEPHAS の Web サイトとしてのデザインやコンテンツの見直しを進めている別分担研究での検討を受けて、分割表の検定の一つであるカイ二乗検定を追加するために必要となるプログラム作成等を行った。現行の MEPHAS が作成された当初は Fisher の正確確率検定が 2×2 の分割表の

みならず、 $2 \times m$ の分割表（2 個の水準と m 個の水準を持つ表）に対しても準備されていた（現在は入力データに依存して生じる過大な計算時間のため停止させている）ため、カイ二乗検定は省略されていた。今回は Figure 12 に示すように分割表の検定を選択するフローチャート中の 2×2 以外の分割表の計算において利用（ 2×2 の分割表のみ Fisher の正確確率検定を使用可能）できるよう、作成することにした。

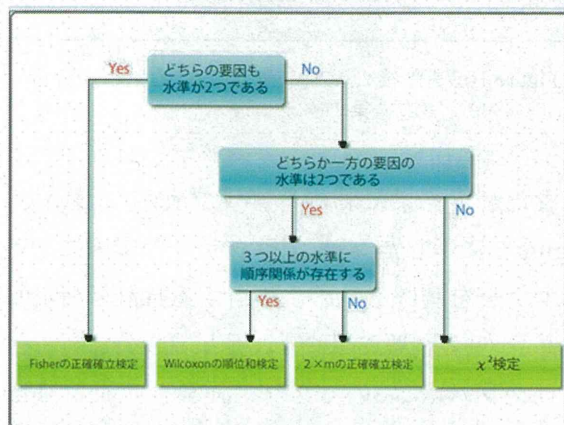


Figure 12. 更新後の MEPHAS における『分割表の検定』のフローチャート

結果

データ			
	I	II	合計
1	22	15	37
2	40	25	65
3	15	18	33
4	10	15	25
5	10	20	30
合計	97	93	190

期待度数			
	I	II	合計
1	18.89	18.11	37
2	33.18	31.82	65
3	16.85	16.15	33
4	12.76	12.24	25
5	15.32	14.68	30
合計	97	93	190

カイ二乗検定の結果
 χ^2 -squared = 9.3118, 危険率: 0.05376
 **: 1%有意, *: 5%有意

Figure 13. カイ二乗検定による計算結果

Figure 13 は作成したカイ二乗検定による計算結果を示している。

D. 考察

本分担研究では主に、MEPHAS の統計解析プログラムを実行するためにサーバー側で稼働しているプログラム群に関し、ユーザビリティの向

上につながる修正を中心にプログラム群の見直しと実際の修正を実施した。

現行の MEPHAS における各プログラムとデータ入力画面、解析結果表示画面を見直し、使用者の使い易さの向上につながるよう、修正点を検討した。その結果、以下の点を修正の対象とした。

- 1). データ入力画面における入力必要箇所の大量発生（例：二元配置分散分析）
- 2). データ入力画面における行列形式データの表示位置のズレとそれによる解析結果への不安（例：Dunnett の方法などの多重比較法）
- 3). 解析結果におけるグラフ表示の Web ブラウザ環境における表示エラー（例：重回帰分析）
- 4). 試験公開以降のメンテナンス等を考慮した統計解析計算プログラムの更新

1)から 3)のそれぞれについて、結果の箇所に例を示したように修正を行った。また 4)に関しては、今後の発展性等も念頭に、R 言語、R 実行環境への統計解析計算プログラムの更新を実施した。今後は、大阪大学内からのアクセスのみに限定した試験公開中にも修正の検討を進め、例えば使用者が選択した統計解析手法が適切でない場合はその理由を明示し適切な手法へと誘導する、適切であっても同様の解析手法がある場合にはそれを紹介するなどの機能を追加することで、統計解析の適切な利用、あるいは学習へとつながるような機能の追加を検討する。

また、別分担研究の実施項目である手法選択のフローチャートの修正に伴い、追加が必要となったカイ二乗検定を実施するためのプログラムの作成を行った。追加されたカイ二乗検定は Fisher の正確確率検定では過大な計算時間が発生してしまう 2×2 以外の分割表を対象とする手法として利用することを念頭に追加を行った。作成したプログラムは既存の統計解析ソフトウェアによ

る解析結果と比較し、違いが無いことを確認した。
今後も手法を選択する上で必要となった統計解析手法を中心に、手法の追加は引き続き検討する予定である。

E. 結論

平成 23 年度に実施し、得られた成果としては以下の通りとなる。

- a. ユーザビリティ向上を目指した入力、解析結果表示画面の更新とプログラムの修正
- b. メンテナンスや更新等を考慮した統計解析計算プログラムの更新
- c. 統計解析手法を追加するためのプログラム等の作成

G. 知的財産権の出願・登録状況

1. 特許取得

特になし。

2. 実用新案登録

特になし。

3. その他

特になし。

引用文献

[1]. MEPHAS,

<http://www.gen-info.osaka-u.ac.jp/MEPHAS/>

[2]. The Apache Software Foundation,

<http://www.apache.org/>, last accessed: 2012/05/17

[3]. R Development Core Team. R: A language and

environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria., <http://www.R-project.org/>, last accessed: 2012/05/17

F. 研究発表

1. 論文発表

特になし。

2. 学会発表

福井大介¹, 田雨時², 岡本晃典², 川下理日人^{2,3},
後藤直久³, 安永照雄³, 高木達也^{2,3} (¹ 阪大薬,
² 阪大院薬, ³ 阪大微研), 医薬学データ用統計解析プログラム、MEPHAS の更新について(その 2),
30P2-am115, 日本薬学会 第 132 年会, 札幌, 2012
年 3 月

研究成果の刊行に関する一覧表

学会発表

発表者氏名	発表タイトル名	発表学会名	開催場所	発表年月
福井大介，田雨時，岡本晃典，川下理日人，後藤直久，安永照雄，高木達也	医薬学データ用統計解析プログラム、MEPHASの更新について（その2），30P2-am115	日本薬学会第132年会	札幌	2012年3月