

201132049A

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23年度 総括・分担研究報告書

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23年度 総括・分担研究報告書

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、また基本的な解析手法（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点などが特徴であり、複雑化の一途を辿る統計解析ソフトウェアとは異なる利用法を提供できるものと考えている。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、開発を進めてきた。最終年度に当たる平成 23 年度に実施した内容を、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) 新規サイトの 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. 分担研究報告	
1. 統計解析の理解促進に向けた Web サイトコンテンツの開発	----- 21
川下 理日人、岡本 晃典	
2. 統計解析の理解促進に資する Web サイト構築のためのプログラム群の開発	-- 29
高木 達也、岡本 晃典	
III. 研究成果の刊行に関する一覧表	----- 36
IV. 研究成果の刊行物・別刷	----- 37

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と比較して、高機能でもなく、また広範な手法を収載してもないが、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使用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実施した。その成果として、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。以下、平成 23 年度の実施内容を報告する。

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。これらの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う。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 においても医薬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目指す。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ている。

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」に分類され则认为している。

「調査」については前年度に、現行の MEPHAS におけるデザインも含めた Web サイト全体及び解析手法のプログラム群を見直し、現在の Web サイト構築技術に照らし合わせ、改善方針の検討を行った。平成 23 年度は現行の MEPHAS における解析手法や語句の説明等のページを中心に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作成を主に行う。昨年度はサーバー側のプログラム内でのエラー検出に関する修正とユーザビリティの向上につながるようなデザイン面も含めた Web ページの更新を行った。平成 23 年度は、試験公開に向けた準備である 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った。

「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数により、明確な評価が下され则认为している。従って、以降では「調査」と「開発」に絞って報告する。

実際の研究実施については分担を明確にするため、デザインも含めた Web サイトの修正、特に解析手法の説明や選択補助について検討し修正を実施する研究と、サーバー側のプログラム群などシステム的な面を検討し修正を行い、また MEPHAS に新たに追加する統計解析手法に関するプログラムを作成する研究の二つに分け、実施することにした。なお、それぞれの研究実施に関する方法については各分担研究報告書にて報告する。

(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 2. 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。

現行の MEPHAS のトップページを Figure 1 に、昨年度更新後の MEPHAS におけるトップページを Figure 2 にそれぞれ示した。なお、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。更新後の MEPHAS における基本的な Web ページデザインは、Figure 2 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。

平成 24 年 5 月末現在、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 サイトとして更新後の Web サイトを公開[3]している。平成 24 年度末には範囲を限定しない試験的な公開へ移行する予定で引き続き準備作業を進めている。

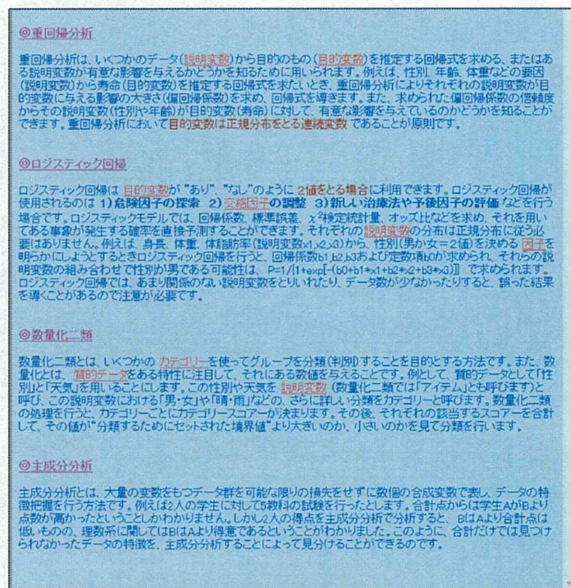


Figure 3. 現行の MEPHAS における『多変量解析など』についての説明ページ

Web サイト上ではフレームの中で表示される。

Figure 3 は、現行の MEPHAS において多変量解析などを説明した Web ページを示している。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。そこで、このような文章のみの説明については事例や図表を追加した内容へ更新を行うこととした。

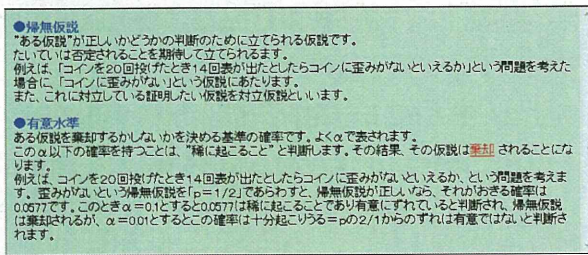


Figure 4. 現行の MEPHAS における語句に関する説明ページの一例

Figure 4 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。この説明ページは、現行の MEPHAS では別ウィンドウに表示される。語句の説明は理解促進の一助になると考え、更新後のよりよい活用法を検討した。



Figure 5. 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 5 は、Figure 3 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 5 に示すように、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、多変量解析ごとに独立した説明ページとして作成した。他の解析手法も同様に、現行の MEPHAS において事例や図表のある場合にはそれらを活用し、事例や図表が無い場合には可能な限りそれらを追加した説明内容へ更新を行った。

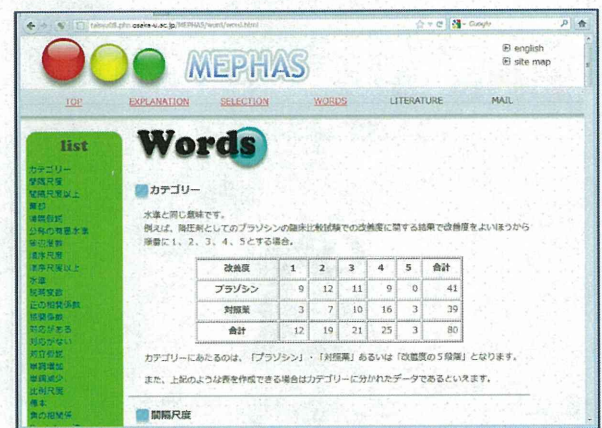


Figure 6. 更新後の MEPHAS における語句の説明一覧のページ

Figure 6 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た。今後は更新した Web ページの内容に合わせて、語句と説明の追加が必要である。

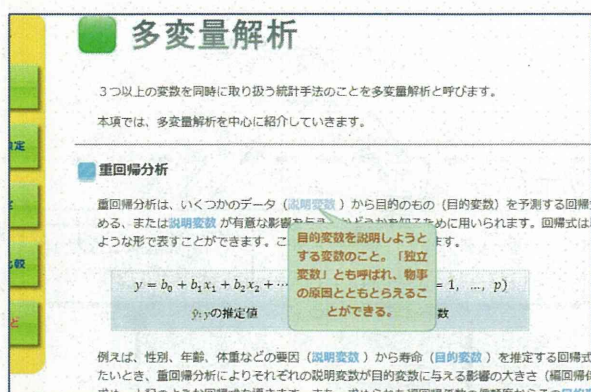


Figure 7. 更新後の MEPHAS における語句の概要説明のポップアップ表示

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 7 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 (CSS (Cascading Style Sheet)) によりスクリプトによらず表示可能であったため、試験公開時に再度変更した)。ポップアップにより概要が表示される語句は、今後も順次追加を進める予定である。



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 (両要因の水準数 5 の場合)

Figure 8 に示すように、現行の MEPHAS における二元配置分散分析では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。

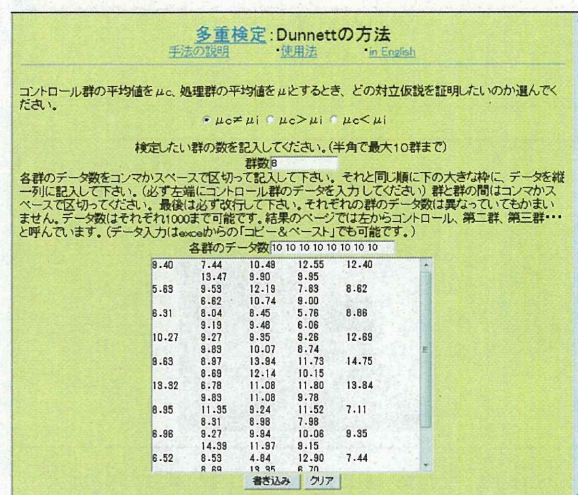


Figure 9. テキストボックス内へのデータ入力における表示のズレの一例

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 9 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り、入力が正しいかどうか使用者に不安感を与える。

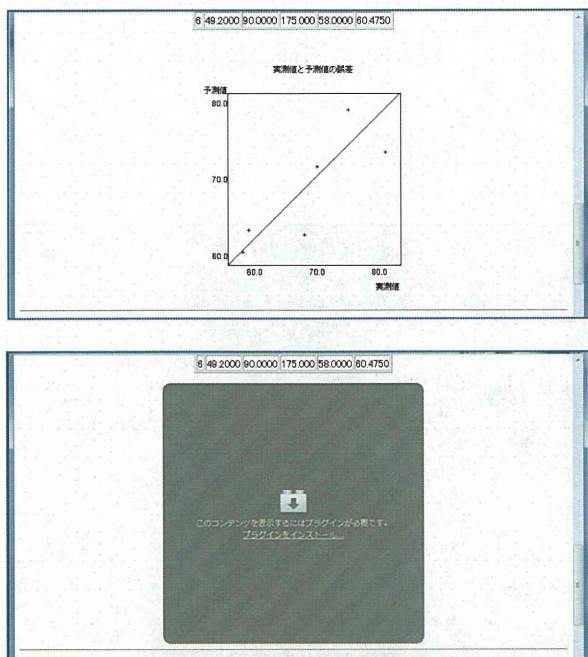


Figure 10. 現行のMEPHASにおける重回帰分析実施結果のグラフ表示（上図は表示可能な環境での結果、下図は表示不可能な環境での結果を示す）

多変量解析手法である重回帰分析や主成分分析では Figure 10 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 10 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 11 である。

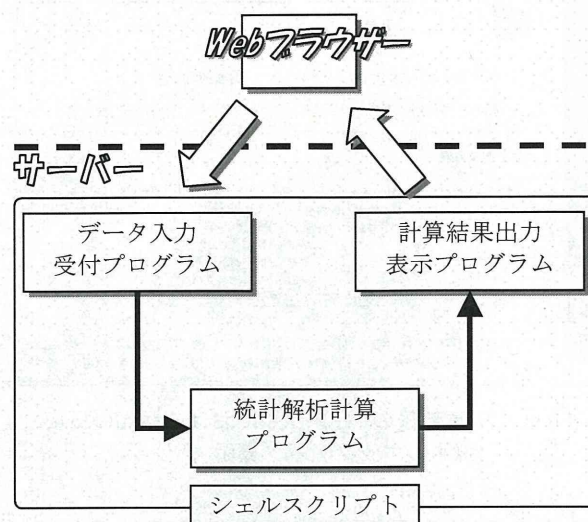


Figure 11. 現行のMEPHASにおいて動作しているプログラム群の概要図

Figure 11 に示される統計解析計算プログラムは現行の MEPHAS では Fortran 言語により作成されているが、今後のメンテナンスや更新を考慮し、近年データ解析分野における使用が多く情報も多く流通している R 言語[2]を利用した計算プログラムへと順次更新した。

Figure 12. 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 8 に示した多数のテキストボックスへの入力を、Figure 12 に示すように一つにまとめ、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。

第1群		第2群	
データ	順位	データ	順位
1	5.70	1	4.72
2	3.82	2	5.77
3	4.15	3	5.02
4	6.13	4	5.63
5	5.81	5	5.49
6	7.65	6	4.39
7	5.48	7	6.68
8	4.48	8	5.64
9	4.25	9	5.25
10	6.76	10	6.47
11	6.69	11	6.32
12	6.14	12	5.09
13	6.86	13	5.35
14	6.71	14	6.64
15	6.00	15	6.45
合計	251.00	合計	214.00
U1	94.00	U2	131.00

Figure 13. 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 9 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、Figure 13 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。

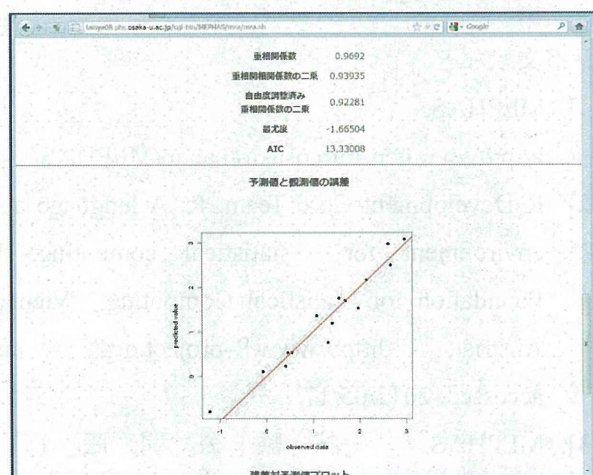


Figure 14. 更新後のMEPHASにおける重回帰分析実施結果のグラフ表示

また解析結果におけるグラフ表示については、Figure 10 に示した Web ブラウザーの環境による表示エラーを避けるため、解析を実行するごとにグラフ画像を作成し、Figure 14 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、Web サイトのコンテンツを中心に改訂、更新を行っている分担研究でのフローチャート

の更新を受けて、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。

データ				期待度数			
	I	II	合計		I	II	合計
1	22	15	37	1	18.89	18.11	37
2	40	25	65	2	33.18	31.82	65
3	15	18	33	3	16.85	16.15	33
4	10	15	25	4	12.76	12.24	25
5	10	20	30	5	15.32	14.68	30
合計	97	93	190	合計	97	93	190

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

Figure 15. カイ二乗検定による計算結果

Figure 15 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

現在、様々な統計解析ソフトウェアが販売されており、また S 言語のクローンとしてスタートした R は無償で非常に多くの手法（統計解析手法、データマイニング手法など）を利用可能[2]である。MEPHAS は手法の多様性や高機能化を目指すのではなく、医学分野において利用頻度の高い、基本的な統計解析手法を簡便に利用できるよう提供し、かつ統計初学者の学習に資するような Web サイトを目指していくことにより、それら統計解析ソフトウェアとの使い分けができるものと考えている。従って本研究は、ユーザビリティの向上や統計解析手法の理解促進につながるような改善を念頭に実施を進めることとした。

平成 23 年度も前年度と同様、実施内容を Web サイトの修正とサーバー側で稼働しているプログラムの修正とに大別し、開発研究を実施した。Web サイトの修正では、まずは試験公開に向けて昨年度決定したデザイン更新を全ページに適用する作業を進めた。また、解析手法の説明につい

ても、事例や図表の追加を含む更新を順次行った。語句の説明に関しても、語句説明の一覧ページを作成し、また Web ページ上の語句に簡単な説明を表示するポップアップを設定した。これにより、文章を読み進めている過程で語句の内容を再確認したい際に、より簡便に確認可能となった。

一方、サーバー側で稼働しているプログラム群の修正では、今後のメンテナンス等を考慮し、プログラミング言語の更新を行った。また、データ入力に際して発生し得る問題点や使用者の環境に応じて発生し得る結果表示の不具合について、プログラムの更新を行った。

これらの実施の後、更新後の Web サイトをアクセス可能な箇所を限定した上で、試験的に公開した。

今後は、アクセス可能な範囲を限定しない公開に向けて、更新後の Web サイトにおける不具合等を改善し洗練に努める。なおアクセス可能な範囲を限定しない公開については、平成 24 年度末を予定している。

また、解析手法や語句の説明などについては、今後も引き続き説明事項の追加や更新を行うことが必要と思われる。さらに解析手法を用いた後の工夫、例えば MEPHAS の解析結果を用いた手法の説明や事例を追加する、あるいはデータに応じて他の適切と思われる手法あるいは同様の解析を可能な手法を紹介するなど、解析結果を得た後の情報提供なども行えるようになれば、統計学への理解促進という点に関してより発展性があると考えている。

E. 結論

平成 23 年度も前年度と同様、MEPHAS のコンテンツや Web サイトのデザインを追加、修正することによりユーザビリティの向上を図る開発研究と、MEPHAS における統計解析手法の実行時に動作しているプログラム群の修正と解析手法

を追加するためのプログラムの作成を主に行う開発研究を実施し、一定の成果を得た。得られた成果は下記の通りである。

- 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
- d. 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
- e. 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
- f. 統計解析手法を追加するためのプログラム等の作成

引用文献など

- [1]. MEPHAS,
<http://www.gen-info.osaka-u.ac.jp/MEPHAS/>
- [2]. 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
- [3]. 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）

F. 健康危険情報

特になし。

（分担研究報告書には記入せずに、総括研究報告書にまとめて記入）

G. 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

H. 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。

(参考資料)

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）



図 1. 更新後の MEPHAS のトップページ

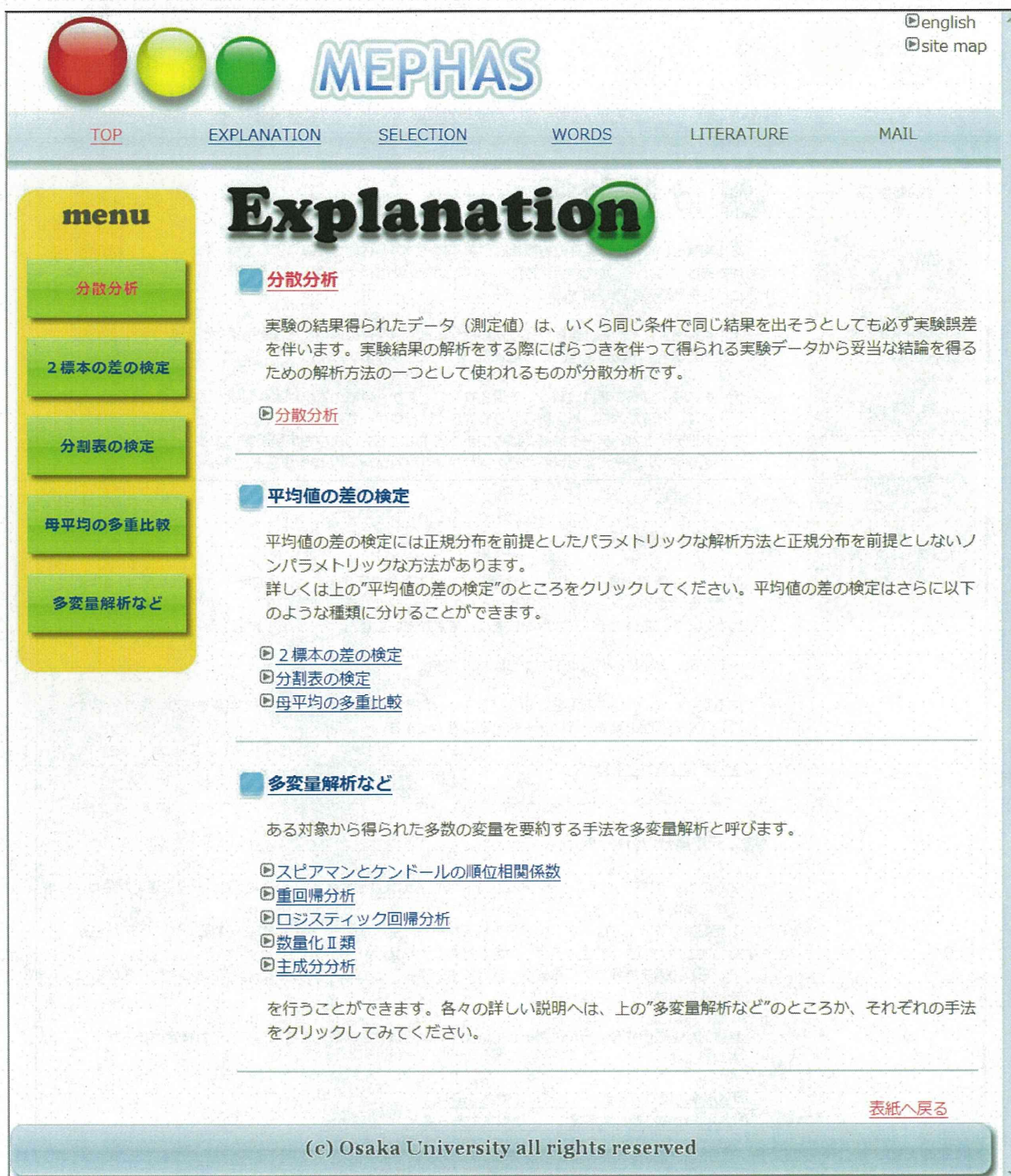


図 2. MEPHAS 新規サイトにおける手法の説明コンテンツのトップページ

[english](#)
[site map](#)

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

menu

分散分析

2標本の差の検定

分割表の検定

母平均の多重比較

多変量解析など

分散分析

ある実験をいくつかの条件で行ったとします。一つの条件に対し繰り返し実験を行ったとき、同じ条件であっても、そのなかで平均値からのずれが多かれ少なかれ生じます。これを実験誤差と呼びます。

また条件を変えて同様の実験を繰り返し行ったとき、条件の違いからほかの条件の場合の平均値との差が出た場合、その差を生みだした **要因** を **因子** と呼びます。

データの持つばらつきは分散として得られるのですが、分散の大小は実験誤差と因子に左右されるといえます。分散が大きければ散らばりも大きく、分散が小さければ散らばりも小さいと解釈できます。分散分析とは、データの持つばらつきが因子によるものよりも実験誤差によるもののほうが大きいかを検定し、因子によるばらつきのほうが大きければ **母平均** に差があるとする検定です。

一元配置分散分析

ある一つの **因子** を取り上げて、その因子について条件を変えることによりデータに違いが生じるかどうかを検定する方法です。前提条件として各データはほぼ **正規分布** に従い、かつ等分散であるものとします。各群のデータ数は同じでなくてもかまいません。

一元配置分散分析の例は以下のようなものです。

ある化学合成において温度を200、230、250℃の3種類に設定して実験を行う。また各温度の設定で複数回の実験を行いデータをとるものとする。

[一元配置分散分析を行う。](#)

二元配置分散分析

2つの **因子** を取り上げて、それらの因子についての条件を変えることによってデータに違いが生じるかを検定する方法です。

二元配置分散分析では、2つの因子それぞれ単独に生じるばらつきのほかにそれら2つの因子が組み合わせることによって生じるばらつきも考慮に入れなければなりません。

それぞれの因子単独で生じるばらつきは「主効果」、2つの因子の組み合わせによって生じるばらつきは「相互効果」と呼ばれます。

具体的な2元配置分散分析の例として、以下のようなケースが考えられ、それぞれ検定の仕方が少し異なります。

- [各水準の繰り返し数が等しく、1である場合。](#)
- [各水準の繰り返し数が等しく、2以上である場合。](#)
- [各水準の繰り返し数が等しくないが、周辺度数に比例する場合。](#)
- [各水準の繰り返し数が等しくなく周辺度数に比例しない場合。](#)

[二元配置分散分析を行う](#)

戻る

(c) Osaka University all rights reserved

図 3. MEPHAS 新規サイトにおける分散分析の説明コンテンツのページ

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

menu

分散分析

2標本の差の検定

分割表の検定

母平均の多重比較

多変量解析など

分散分析

二元配置分散分析

各水準での繰り返し数は等しくないが、周辺度数に比例する場合。

1つの条件に対する測定値が等しくない場合に用います。
ただし、測定値の個数の比が等しくなるようにします。

例：菌種（A, B, C）を用いて、温度を変えて（Ⅰ、Ⅱ、Ⅲ、Ⅳ）あるたんぱく質の生成量を測定し、以下のようなデータが得られた。

		温度			
		低温（Ⅰ）	中温（Ⅱ）	高温（Ⅲ）	超高温（Ⅳ）
菌種	A	6.0	7.8	9.8	8.2
		5.8	8.1	6.7	5.9
	B	8.8	8.4	8.5	7.3
		6.6	7.1	6.4	9.4
	C	6.4	7.9	9.2	5.3
		9.4	9.2	5.5	9.7
	6.9	6.7	7.3	5.5	
	7.4	8.2	9.6	6.4	
			8.4	8.6	
			7.7	7.2	

この例では各温度（Ⅰ、Ⅱ、Ⅲ、Ⅳ）のいずれの場合も
 $A : B : C = 1 : 1 : 2$

また、各菌種（A, B, C）いずれの場合も
 $I : II : III : IV = 2 : 2 : 3 : 3$

となっています。このとき、
「菌種の違いと温度の違いがたんぱく質生成量に関係しているか」を調べます。

前提

無仮説：「要因効果がない」
対立仮説：「要因効果がある」
と設定し有意水準 α をとして両側検定を行います。

モデル

結果として返される分散分析表を解釈するモデルが3つあります。

モデルⅠは母数モデルとも呼ばれます。それぞれの要因（年齢、投薬期間）における、各水準を固定された不動のものとみなします。

モデルⅡは変量モデルとも呼ばれます。それぞれの要因（年齢、投薬期間）における、各水準は無数の水準の内の標本とみなし、そこから得られる推測結論を、標本以外の広い範囲へも適用しようとするものです。

混合モデルは片方の要因に母数モデル、もう一方の要因に変量モデルを考えるものです。

結果

上記のデータを用いた時、どのモデルを用いても

要因A（菌種の差）	:	0.05 < P
要因B（温度の差）	:	0.05 < P
相互作用	:	0.05 < P

となります。つまり、
「各要因の要因効果があるとは言えない。」
ということです。

→戻る

(c) Osaka University all rights reserved

図 4. MEPHAS 新規サイトにおける二元配置分散分析に関する説明コンテンツのページ



[English](#)
[Site map](#)

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

menu

[分散分析](#)

[2 標本の差の検定](#)

[分割表の検定](#)

[母平均の多重比較](#)

[多変量解析など](#)

多変量解析

3 つ以上の変数を同時に取り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。この式を「重回帰式」と言います。

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_jx_j + \dots + b_px_p \quad (j = 1, \dots, p)$$

y : y の推定値 b_0 : 定数 b_j : 偏回帰係数

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（性別や年齢）が目的変数（寿命）に対して有意な影響を与えているかどうかを知ることができます。

重回帰分析において目的変数が正規分布をとる連続変数であることが原則です。

結果

当サイトのプログラムで重回帰分析を行うと以下の値が得られます。

t 値

求めた偏回帰係数や定数項が 0 でないかどうかを検定するために用いる検定統計量です。これをもとに t 検定を行い、「偏回帰係数（または定数項）は 0 である」という帰無仮説が棄却されるかどうかを見ます。帰無仮説が棄却された場合、その偏回帰係数は、0 でないと考えるので、目的変数（ y ）に何らかの影響を与えていることになります。また、帰無仮説を採用する場合、その偏回帰係数は 0 でない、つまり目的変数（ y ）に影響を与えていない可能性があるといえます。

決定係数

目的変数の実測値の分散は、求められた回帰式の予測値の分散と、残差の分散の和で表わされます。回帰方程式では、この残差の分散ができるだけ小さくなるように回帰式を定めています。そこで、回帰式の当てはまりの良さを表す指標として用いられるのが、決定係数です。決定係数は、0 以上 1 以下の値をとり、1 に近ければ近いほど予測精度は良いということです。一般に決定係数が 0.8 以上であれば、かなり良い精度で予測されているといえます。

重相関係数

実測される目的変数（ y ）の値と重回帰分析の結果として予測される目的変数の間の相関係数のことです。この値が大きければそれだけ回帰方程式から得られる予測値が、目的変数の実測値に近づくことになります。もしも重回帰分析による予測が完全であればこの値は 1 になります。

自由度調整済み重相関係数

説明変数を増やしていくと、重相関係数は単調に増加していき、説明変数（ x ）の個数が「データセット数 - 1」になると重相関係数は 1 になり、データセット数が独立変数の数に比べて小さい時には良い指標であるとは言えなくなってしまいます。そのために重相関係数をデータセットの数と自由度で修正します。この値が 1 になった時は予測した重回帰式には意味がないということになります。

回帰の分散分析表

ここでは、得られた回帰式の正当性を検定しています。重回帰分析の場合、結果の正当性は上記にある決定係数で判断されます。そこで、重回帰分析の正当性を評価するために、決定係数が十分に大きいといえるかを判断するために分散分析を行っています。ここで決定係数が大きいといえることが言えれば、回帰式は正当であるといえます。分散分析に関する詳しい解説は、こちらをご覧ください。

信頼限界

設定した有意水準のもとで母数（ここでは偏回帰係数のこと）が存在する区間の上限と下限のことです。有意水準を 5% とした時、95% の確率でこの区間に偏回帰係数が存在するということになります。

AIC

赤池の情報量基準。モデルの最大対数尤度と独立変数（説明変数）の数により解析的評価から求められたもの。この値が小さいほどそのモデルは予測が良いといえます。

信頼限界

設定した有意水準のもとで母数（ここでは偏回帰係数のこと）が存在する区間の上限と下限のことです。

戻る

(c) Osaka University all rights reserved

図 5. MEPHAS 新規サイトにおける重回帰分析に関する説明コンテンツのページ

- 14 -

[english](#)
[site map](#)

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

chart menu

2 群間の差の検定のチャートへ

分割表の検定のチャートへ

分散分析・多重比較のチャートへ

多変量解析などのチャートへ

Selection

正しい手法を選択するためのチャートです。

下のチャートの黄色の部分をクリックするとさらに詳しいチャートが表示されます。

赤字をクリックするとその説明が表示されます。

変数の数はいくつですか？

2つ

3つ以上

データは2群に分けられている

データは3群以上に分かれている

分割表のデータである

差があるかを知りたい

関連があるかを知りたい
(相関係数が知りたい)

2 群間の差の検定のチャートへ

多重比較・分散分析のチャートへ

分割表の検定のチャートへ

順位相関係数

多変量解析などのチャートへ

[戻る](#)

(c) Osaka University all rights reserved

図 6. MEPHAS 新規サイトにおける手法の選択補助コンテンツのページ

- 15 -

[english](#)
[site map](#)

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

list

[カテゴリ](#)
[間隔尺度](#)
[間隔尺度以上](#)
[棄却](#)
[帰無仮説](#)
[公称の有意水準](#)
[周辺度数](#)
[順序尺度](#)
[順序尺度以上](#)
[水準](#)
[説明変数](#)
[正の相関係数](#)
[相関係数](#)
[対応がある](#)
[対応がない](#)
[対立仮説](#)
[単調増加](#)
[単調減少](#)
[比例尺度](#)
[標本](#)
[負の相関係数](#)
[Bootstrap法](#)
[分割表](#)
[母集団](#)
[名義尺度](#)
[目的変数](#)
[有意水準](#)
[要因](#)

Words

■ カテゴリ

水準と同じ意味です。

例えば、降圧剤としてのプラソシンの臨床比較試験での改善度に関する結果で改善度をよいほうから順番に1、2、3、4、5とする場合。

改善度	1	2	3	4	5	合計
プラソシン	9	12	11	9	0	41
対照薬	3	7	10	16	3	39
合計	12	19	21	25	3	80

カテゴリにあたるのは、「プラソシン」・「対照薬」あるいは「改善度の5段階」となります。

また、上記のような表を作成できる場合はカテゴリに分かれたデータであるといえます。

■ 間隔尺度

数値の差のみに意味がある尺度。

「距離尺度」とも呼びます。順序尺度の性質も備えています。

例えば、温度が10℃から15℃になった時に、50%の温度上昇があったとは言いません。

温度が10℃から15℃になった時も、100℃から105℃になった時も共に5℃の上昇です。

5℃という数値には意味があります。

■ 間隔尺度以上

間隔尺度・比例尺度のいずれかである場合です。

■ 棄却

前提となる仮説を「捨てる・採択しない」ということです。

つまり、帰無仮説を棄却することとは、対立仮説を採択することになります。

■ 帰無仮説

“ある仮説”が正しいかどうかの判断のために立てられる仮説です。

たいていは否定されることを期待して立てられます。

例えば、「コインを20回投げたとき、14回大手が出たとしたらこのコインに歪みがないといえるか。」という問題を考えた場合に、「コインに歪みがない」という仮説に当たります。また、これに对立している証明したい仮説を対立仮説といいます。

■ 公称の有意水準

多重比較法において最初に決定した有意水準。第一種の過誤の確率がこれを超えないことが多重比較法として最低限必要な要件です。

■ 周辺度数

種差度数とは、以下の場合。

$(a1 \text{の周辺度数}) = (a1 \text{における} b1 \text{の度数}) + (a1 \text{における} b2 \text{の度数})$

ということです。

つまり、下記の白地の部分の度数（データの個数）を全て足したのがa1の周辺度数（今回は5）ということになります。

■ 順序尺度

大小関係にのみ意味がある尺度。

例えば、治療効果の判定において、悪化・不変・改善・有効をそれぞれ-1・0・1・2と数値に対応させたもの。

平均値は定義できませんが、中央値は定義できます。

■ 順序尺度以上

順序尺度・間隔尺度・比例尺度いずれかである場合です。

図 7. MEPHAS 新規サイトにおける語句の概要説明一覧のページの一部

- 16 -