

201034074A

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 総括・分担研究報告書

研究代表者 岡本 晃典

平成23（2011）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 総括・分担研究報告書

研究代表者 岡本 晃典

平成23（2011）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、開発を進めている。平成 22 年度に初年度分として実施した内容を、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 ----- 20	
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 --- 30	
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ほど高機能でも広範な手法を収載してないが、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 サーバー側で稼働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、一定の開発成果を得た。

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図っている。しかし中には、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。

現在の 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

一側で受信し、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。現在の 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」に分類されると考えている。「調査」では、現在の 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、現在の 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う。

「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。現在の 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。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るので、まずは現在の MEPHAS に沿った手法を中心として、基本的なことや使用者を間違えた判断に導きかねないことを取り上げる。

「評価」に関しては、二年目以降に修正を進めている Web サイトを大学院生に使用してもらい、試験公開に向けた評価を行うことを検討している。従って平成 22 年度に関しては、現在の 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が保たれるかの評価を行う。

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に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 は現在の MEPHAS のトップページである。



Figure 1. 現在の MEPHAS トップページ

Figure 1 のトップページの右側に並んだボタンが各コンテンツへのメニューとなっている。2002年当時とはともかく、現在 Web 上で見られる他 Web サイトと比較してメニューや文章、図などが整理されて配置されているようには見受けられない。また、Figure 2 は現在の『手法の解説』のトップページを示している。

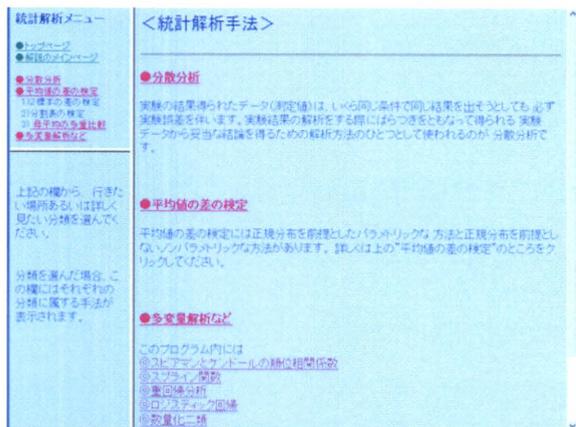


Figure 2. 現在の『手法の解説』ページ

『手法の解説』のページでは、Web ブラウザーの表示を3つに区切り、左側にメニューを配置し、一番大きな中央部に本文を配置している。先の Figure 1 に示したトップページと比較して、整理された印象を与え、この方が利用者にとってはより使い易いだろうと思われる。但し、Figure 2 で

使用されているフレームによる画面構成は近年では推奨されていない[3]。またメニューも Web ページの構造を反映したものとはあまりなっておらず、整理の必要性がある。本文についても読み易く表示されているとは思えないので、修正が必要と判断した。

Figure 3 は修正後の MEPHAS のトップページである。



Figure 3. 修正後の MEPHAS のトップページ

修正後の Web 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ります。以降、明記しませんが修正後の Web サイトに関する情報は全て同様です。

修正後の MEPHAS のトップページは上部に直下のコンテンツに関するメニューをまとめている。Figure 4 は『Explanation』のトップページであり、Figure 2 に該当するページである。



Figure 4. 修正後の MEPHAS における『Explanation』のトップページ

上部のメニューに関しては Figure 3 と同様に表示し、この『Explanation』のトップページ以下のコンテンツに関しては、左側にメニューを表示している。このメニューの配置については手法の選択補助である『Selection』に関しても同様にしており、一貫性を持たせることでユーザーを迷わせることが少ないようにしている。

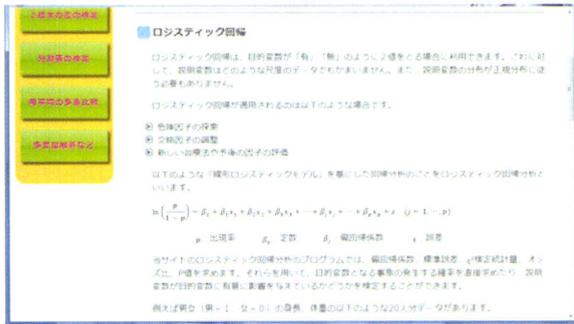


Figure 5. 新規サイトでの手法の説明のページ
多変量解析のロジスティック回帰の解説のページを取り上げている。

Figure 5 は新規サイトにおける手法の説明のページの一例、多変量解析の一つであるロジスティック回帰に関する説明を示している。場合によっては式表現も導入し、Figure 2 と比較してより見易くなったものと思われる。ただ、修正中の説明文章も含めて、用例や図表の追加が一部にとどまっているため、できる限り多く追加できるよう、引き続き説明の修正が必要と思われる。

Figure 6 は現在の MEPHAS において提供されている手法の選択を補助するフローチャートの一つである。

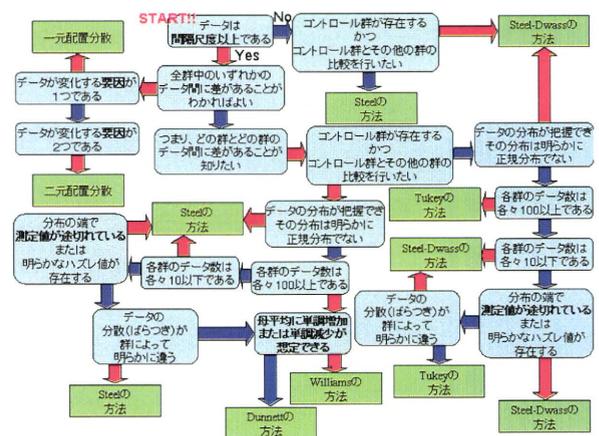


Figure 6. 現在の MEPHAS における『多群のデータに差があるかどうか 順序尺度以上の場合』のフローチャート

Figure 6 は、恐らく現在の MEPHAS において最も使用頻度の高い手法であろう多重比較法を含むフローチャートである。非常に入り組んでおり把握し辛く、また設問の中には選択肢が一つしかなく設問の体をなしていないものも含まれている。また Figure 6 に示した以外にも解析手法の選択補助のためのフローチャートが4つあり、設問と選択手法のチャート間での重複や MEPHAS に準備されていない手法へ行き着くなども修正の必要な点として考えられた。そこで、より見易く整理し、かつ可能な限り意義のある分岐を設定し、場合によっては現在の MEPHAS には準備されていない手法の追加を行うことで、解析手法の選択補助は改善が期待できると考えた。

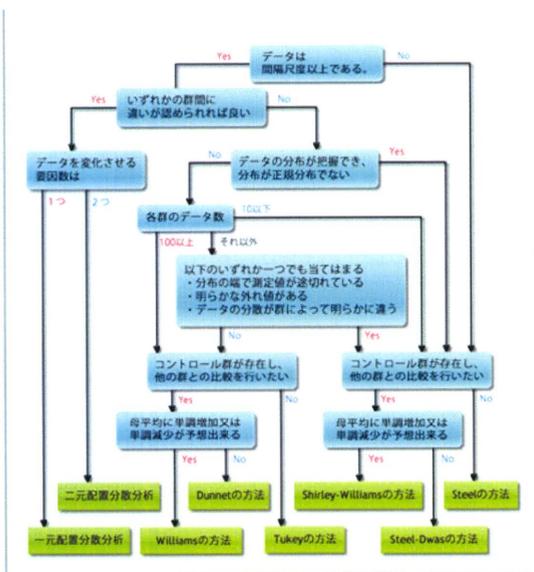


Figure 7. 新規サイトでの『多重比較・分散分析』のフローチャート

Figure 7 は Figure 6 のフローチャートを整理した、MEPHAS で新たに使用するフローチャートである。設問の配置を整理し、基本的の上から下へとチャートを辿る形式となっており、また設問自体も整理され、より分かり易くなっている。現在の MEPHAS では準備されていない、多群間に序列があり、かつコントロールとの対比較を行う場合のノンパラメトリックな多重比較法である Shirley-Williams の方法を新たにフローチャート中に追加し、また MEPHAS 上で実行できるようプログラムを作成した。Figure 8 は Shirley-Williams の方法を実行した際の計算結果を示している。なお、プログラムを新たに作成したため、複数のサンプルデータを用いて既存のソフトウェアによる Shirley-Williams の方法の計算結果と比較し、新たに作成したプログラムによる計算に間違いがないことを確認した。



Figure 8. Shirley-Williams の方法による計算結果

Figure 9 は現在の MEPHAS において統計解析手法を実行する際に動作しているプログラム群を整理した概要図である。

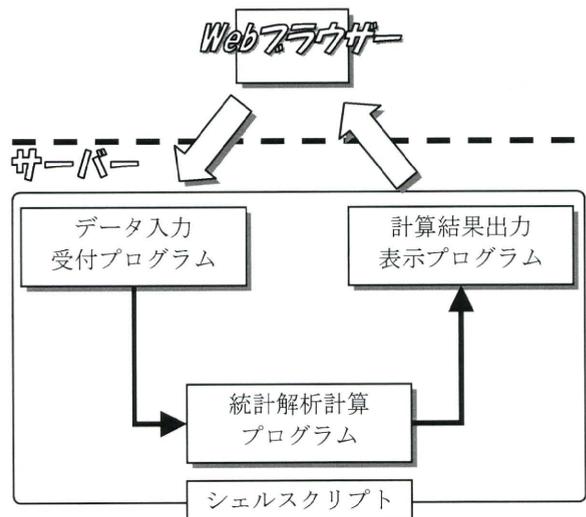


Figure 9. 現在の MEPHAS において動作しているプログラム群の概要図

Figure 10 は現在の MEPHAS において統計解析手法を使用した際に、データ入力受付プログラムにて何らかのデータ入力の間違ひを見つけた際に表示するエラーメッセージである。

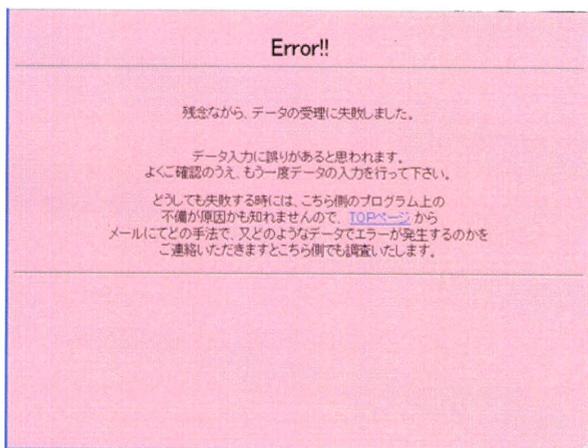


Figure 10. 現在のMEPHASにおけるデータ入力時の間違いに対するエラーメッセージの画面

使用者から入力されたデータに関して間違いと思われる箇所が検出された際に現在のMEPHASのデータ入力受付プログラムは、たとえ妥当な修正方法が考えられる場合（例えば全角数字を半角数字へ変換など）においてもエラーメッセージを表示させるようにしている。これは、万が一にも使用者の意図とは異なるデータを基にした解析結果を表示し、誤った計算結果に導くことを避けるために必要な措置であろう。一方で、こういった間違いであっても Figure 10 のエラーメッセージが表示されるため、使用者はデータ入力に間違いがあったとしても修正の方法に気付けない可能性がある。そこで、データ入力受付プログラムにおいて間違いとされた内容を使用者に伝えることによりデータ入力の間違いを訂正し易くするよう、プログラムの修正を行った。



Figure 11. プログラム修正後の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（データに空欄があった場合）

Figure 11 はプログラム修正後のMEPHASにおけるデータ入力時に、入力すべき箇所に空欄があったとデータ入力受付プログラムにおいて判断された際に表示されるエラーメッセージである。現状、Figure 11 のエラーメッセージを含めて、以下の6つの間違いがメッセージとして表示される。

- a. データの入力量が多すぎる場合
- b. 入力に空欄がある場合
- c. 入力が半角数字で行われていない場合
- d. 信頼区間など入力範囲があるもので範囲外の入力の場合
- e. 群数など整数値が期待されるもので整数値でない場合
- f. サンプルサイズと入力データが合致しない場合

D. 考察

現在、SAS[4] を始めとして S-PLUS[5] や SPSS[6] など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、医薬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 22 年度は、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、使用者の持つデータに合わせた適切な手法選択を補助するためのコンテンツを見直し、フローチャートの設問の改善も含めてフローチャートを整理、再構成し、修正を行った。また、解析手法の説明がより読み易く、内容が分かり易くなるよう、用例や図表、場合によっては数式なども導入し、説明の本文、及びページ構成などを修正した。さらに、Web サイト全体の構成やデザインが、公開された 2002 年当時から全く変わっておらず、メニュー表示なども利用者の使い勝手への配慮が足りないため、デザインも含め Web ページの修正も進めている。

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、利用者からのデータ入力に間違い（例えば空欄や数字以外の文字の入力、サンプルサイズのずれなど）があり得るとプログラムにて判断された場合に利用者側で入力のやり直しを行い易くなるよう、間違いと判断した内容を含むエラーメッセージを利用者へと返すよう、修正した。また Web サイトの修正におけるフローチャートの修正に伴い、統計解析手法を追加するためにプログラム群を新たに作成し、多重比較法の一つである Shirley-Williams の方法を新たに追加した。

今後は、適切な手法の選択補助として使用しているフローチャートについて設問の妥当性などを再度見直し、適宜統計解析手法を追加することも必要であろう。また手法の解説としては、もう

少し用例等を追加することで内容の分かり易さの向上を図ることを引き続き検討する。解析手法の実行前に適切な手法を選択できるよう補助するだけでなく、何らかの手法により解析した際に、その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、手法の計算結果を出し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 22 年度は、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. Web サイト修正の方向性の決定とデザインの作成
- b. 手法の説明文章の修正
- c. 手法の選択のためのフローチャートの整理と改訂版の作成
- d. 入力データの誤りへの対応に向けたプログラムの修正
- e. 統計解析手法の追加の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]. The R Project for Statistical Computing,
<http://www.r-project.org/>
- [3]. W3C HTML 4.01 Specification,
<http://www.w3.org/TR/html401/struct/global.htm>

[4]. SAS,
<http://www.sas.com/offices/asiapacific/japan/platform/sas9/index.html>

[5]. S-PLUS, <http://www.msi.co.jp/splus/>

[6]. SPSS, <http://www.spss.co.jp/>

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,
29P-0115, 日本薬学会 第131年会, 静岡, 2011年3月
(東日本大震災のため開催中止となったが、組織委員会より要旨等も配布済みのため年会は成立したものとすとの通知あり)

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)

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)



図 1. MEPHAS 新規サイトのトップページ



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](#)

Explanation

分散分析

実験の結果得られたデータ（測定値）は、いくら同じ条件で同じ結果を出そうとしても必ず実験誤差を伴います。実験結果の解析をする際にばらつきを伴って得られる実験データから妥当な結論を得るための解析方法の一つとして使われるものが分散分析です。

[分散分析](#)

平均値の差の検定

平均値の差の検定には正規分布を前提としたパラメトリックな解析方法と正規分布を前提としないノンパラメトリックな方法があります。詳しくは上の“平均値の差の検定”のところをクリックしてください。平均値の差の検定はさらに以下のような種類に分けることができます。

[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

ある対象から得られた多数の変量を要約する手法を多変量解析と呼びます。

[スピアマンとケンドールの順位相関係数](#)[重回帰分析](#)[ロジスティック回帰分析](#)[数量化Ⅱ類](#)[主成分分析](#)

を行うことができます。各々の詳しい説明へは、上の“多変量解析など”のところか、それぞれの手法をクリックしてみてください。

[表紙へ戻る](#)

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



MEPHAS

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

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

menu

総論総論

2標本の差の検定

分割法の検定

母平均の乗乗比較

多変量解析など

2 標本の差の検定

平均値の差の検定のうち、特に2つの集団の間に差があるかないかを調べるために用いる方法です。さまざまな種類の検定方法があります。

- 「パラメトリック検定」であるか「ノンパラメトリック検定」であるか
- 「対応のあるデータ」か「対応のないデータ」か

でそれぞれの手法を分類しています。

「パラメトリック検定」か「ノンパラメトリック検定」か

パラメトリック検定

パラメトリック検定は母集団パラメーター（平均、分散、標準偏差など）について仮説を設ける検定方法です。一般に母集団の正規性や等分散性を仮定します。

「処理群と非処理群の間で差が生じているか」や、「被験薬を投与した集団と対照薬を投与した集団間で差が生じているか」を調べる場合などのように2つの集団の比較を行うような場合、どちらの群もデータが正規分布に従うとすると2つの集団の母集団の母平均の差に関する検定を考えると比較しやすくなります。この検定を行うために用いるのがt検定です。本サイトには以下の3種類の検定があります。

- Studentのt検定
- Welchのt検定
- 対応のあるt検定

ノンパラメトリック検定

ノンパラメトリック検定はパラメトリック検定に対して一切の仮説を設けない検定方法です。データ数が少ない場合にも適用でき、名義尺度や順序尺度のデータも使用できます。

ノンパラメトリック検定である順位和検定は、測定値を大きさの順番に置きなおしその順位の差を考える検定です。本サイトでは以下の2種類の検定を行うことができます。

- Wilcoxonの順位和検定（マン・ホイットニーのU検定）
- Wilcoxonの符号付き順位和検定

「対応のあるデータ」か「対応のないデータ」か

2標本の差の検定では、さらにデータに対応がある場合、対応がない場合で2つに分けることができます。

対応のある場合

対応のある場合とは、2つの集団が「術後と術前」や、「薬の投与前と投与後」のように同一対象から異なる2地点での観測値の「ペア」が得られる場合、または異なる母集団から同じ条件を持つものとして選択するような場合のことを言います。

本サイトでは以下の2種類の検定を行うことができます。

- 対応のあるt検定
- Wilcoxonの符号付順位和検定

[詳しくはこちら。](#)

対応のない場合

対応のない場合とは、上記の「対応のある場合」の条件を満たさないものです。2つの集団の母集団がたがいに独立である場合のことでもあります。

本サイトでは以下の3種類の検定を行うことができます。

- Studentのt検定
- Welchのt検定
- Wilcoxonの順位和検定

[詳しくはこちら。](#)

[戻る](#)

(c) Osaka University all rights reserved

図 3. MEPHAS 新規サイトにおける 2 標本の差の検定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

母平均の多量比較

多量分散分析

2 標本の差の検定

平均値の差の検定のうち、特に2つの集団の間に差があるかないかを調べるために用いる方法です。

対応のある場合

対応のある場合とは、2つの集団が「術後と術前」や、「薬の投与前と投与後」のように同一対象から異なる2地点での観測値の「ペア」が得られる場合、または異なる母集団から同じ条件を持つものとして選択するような場合のことを言います。

同一個体でなくても同じ条件下、同じ系統の場合もデータを「ペア」として扱うことができるためデータに対応があるといえます。データを「ペア」としているため、2群のデータ数は等しくなります。



このような対応のあるデータに用いるのが以下の検定です。

- 対応のある t 検定
- Wilcoxonの符号付き順位和検定

対応のある t 検定は、データに正規分布が仮定できる場合に用いる、パラメトリック検定です。

Wilcoxonの符号付き順位和検定は、2群が正規分布をなすかどうか分からない場合に用いるノンパラメトリック検定です。

データに正規分布が仮定できる場合

データに正規分布が仮定できる場合、対応のある t 検定を用います。

対応のある t 検定は、2つのデータの差を計算してその差を元にして検定を行います。Studentの t 検定やWelchの t 検定より優位差が見られやすくなります。

前提

帰無仮説：「母平均に差はない」

対立仮説：「母平均に差がある」

有意水準を α として両側検定を行います。

$P > \alpha$ の時、帰無仮説を採択し、「母代表値に差があるとは言えない」

$P \leq \alpha$ の時、帰無仮説を棄却し、「母代表値に差がある」

とします。

データに正規分布が仮定できない場合

データに正規分布が仮定できない場合、Wilcoxonの符号付き順位和検定を用います。

前提

帰無仮説：「母代表値に差はない」

対立仮説：「母代表値に差がある」

有意水準を α として両側検定を行います。

$P > \alpha$ の時、帰無仮説を採択し、「母代表値に差があるとは言えない」

$P \leq \alpha$ の時、帰無仮説を棄却し、「母代表値に差がある」

とします。

[戻る](#)

(c) Osaka University all rights reserved

図 4. MEPHAS 新規サイトにおける 2 標本の差の検定内、対応のある場合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

対話分析

入場券の価格決定

分類者の決定

海産物の価格決定

多変量解析とは

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。

本稿では、多変量解析を中心に紹介していきます。

ロジスティック回帰

ロジスティック回帰は、目的変数が「有」「無」のように2値をとる場合に利用できます。これに対して、説明変数はどのような尺度のデータでもかまいません。また、説明変数の分布が正統分布に従う必要もありません。

ロジスティック回帰が適用されるのは以下のような場合です。

- Ⓐ 色味因子の探索
- Ⓑ 交雑因子の調整
- Ⓒ 新しい治療法や手術の因子の評価

以下のような「線形ロジスティックモデル」を基にした回帰分析のことをロジスティック回帰分析といいます。

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \dots + \beta_nx_n + \epsilon \quad (U=1, \dots, p)$$

p: 出現率 β_0 : 定数 β_j : 偏回帰係数 ϵ : 誤差

当サイトのロジスティック回帰分析のプログラムでは、偏回帰係数、標準誤差、 χ^2 検定統計量、オッズ比、P値を求めます。それらを用いて、目的変数となる事象の発生する確率を直接求めたり、説明変数が目的変数に有意に影響を与えているかどうかを判定することができます。

例えば男女（男 = 1、女 = 0）の身長、体重の以下のような20人分データがあります。

身長	体重	性別	身長	体重	性別
181	67	1	175	57	1
160	74	1	157	64	0
140	55	0	163	74	0
152	50	0	157	64	0
142	58	0	175	66	1
154	62	1	173	79	1
145	63	0	172	83	1
168	60	1	175	57	1
177	67	1	178	58	1
166	65	1	166	83	0

身長、体重の2つを説明変数として、目的変数となる性別を予測しようとするとき、本サイトのロジスティック回帰分析を行うと以下のよう結果を得ることができます。

	偏回帰係数	標準誤差
定数項	-36.257	15.559
身長	0.2125	0.0975
体重	0.0305	0.1019

	wald統計量	オッズ比	P値
定数項	5.4304	0.1794 × 10 ⁻¹⁸	0.0198
身長	4.7529	1.2969	0.2925
体重	0.0898	1.0310	0.7645

偏回帰係数から、性別が男性である確率pは以下の式に回帰できます。

$$p = \frac{1}{1 + \exp(36.275 - 0.2125x_1 - 0.0305x_2)}$$

この確率の対数オッズ比を算出する式のような式になります。対数オッズ比はロジットとも呼ばれます。オッズとは、ある事象が起こる確率をpとしたとき、

$$\ln\left(\frac{p}{1-p}\right) = -36.275 + 0.2125x_1 - 0.0305x_2$$

この式は、身長と体重を説明変数とし、ロジットを目的変数とした直線回帰式に相当します。この式を一般化したモデルを厳密に示した「線形ロジスティックモデル」といいます。

また、下の表では各目的変数の評価を行っています。wald検定統計量は以下の式から求めることができます。

$$\text{wald 検定統計量} = \left(\frac{\text{偏回帰係数}}{\text{標準誤差}}\right)^2$$

wald検定統計量は、カイ二乗分布に従うことが知られており、この検定統計量を用いて各偏回帰係数が0でないかを検定することができます。有意水準 (α) を5%と定めるとき、今回の例では、身長の項目で、P < α となっていることから、性別を判別するにあたって、身長は有意な影響を与えているということがわかります。

ロジスティック回帰では、あまり関係のない説明変数を取り入れたり、データ数が少なかったりすると誤った結果を得ることがあるので注意が必要です。

[戻る](#)

(c) Osaka University all rights reserved

図 5. MEPHAS 新規サイトにおけるロジスティック回帰に関する説明コンテンツのページ



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](#)

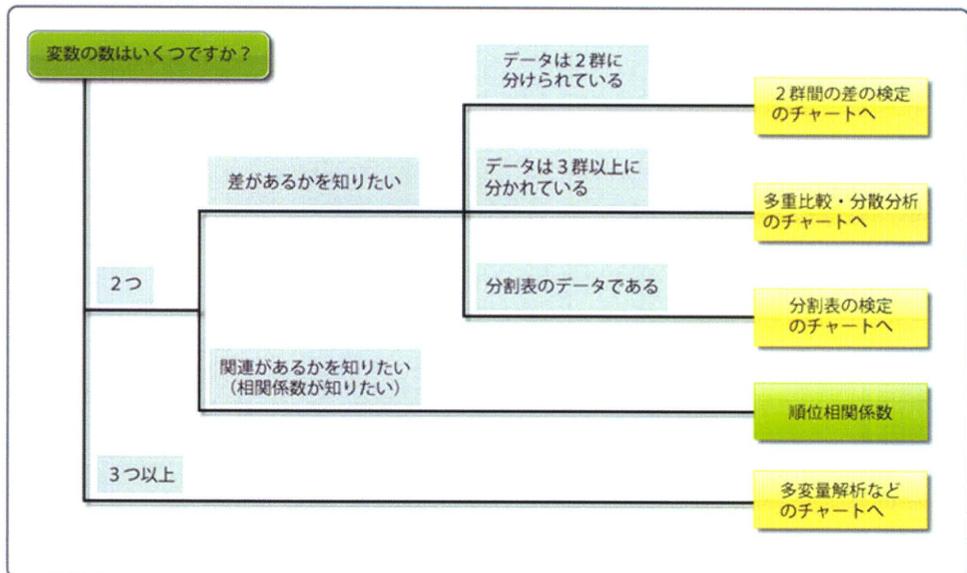
[多変量解析などのチャートへ](#)

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。



[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



chart menu

2群内の差の検定のチャートへ

分割表の検定のチャートへ

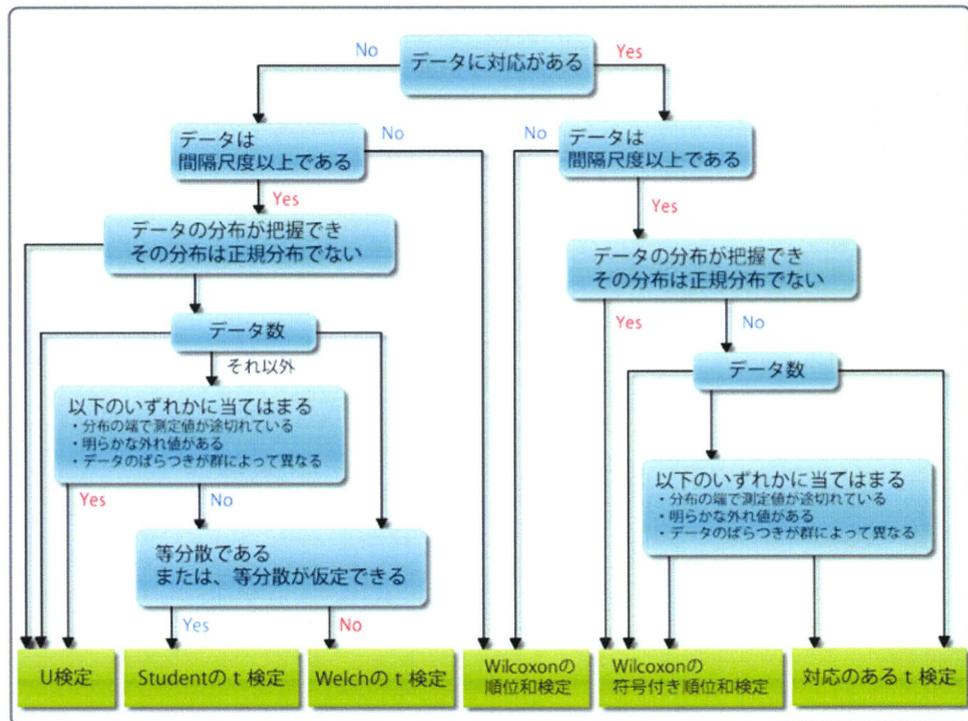
多重比較・多重比較のチャートへ

多変量解析などのチャートへ

2群間の差の検定

主に2群間に差があるかを検定します。

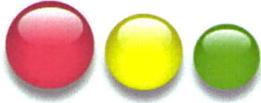
赤字の部分はクリックするとその語句の解説が表示されます。



[戻る](#)

(c) Osaka University all rights reserved

図 7. MEPHAS 新規サイトにおける手法の選択補助内、2群間の差の検定に関する
選択補助のフローチャートのページ



menu

Shirley-Williamsの方法

帰無仮説を $\mu_1 = \mu_2 = \dots = \mu_a$ として、証明したい対立仮説を選択してください。

- $\mu_1 > \mu_2 > \dots > \mu_3$
- $\mu_1 < \mu_2 < \dots < \mu_3$

有意水準 α を定めて下さい。(上側)

- $\alpha = 0.050$
- $\alpha = 0.025$
- $\alpha = 0.005$

検定したい群の数を入力してください。

群数

各群のデータ数をコンマかスペースで区切って入力してください。

各群のデータ数

データを縦一列に入力してください。

ただし以下の点に注意してください。

- ・ 群と群の間はコンマかスペースで区切ってください。
- ・ それぞれのデータ数は異なってもかまいません。
- ・ データ数は各群1000まで入力可能です。
- ・ コントロール群はデータの左端になるように入力してください。
- ・ 結果のページでデータの左から第二群、第三群…と呼んでいます。
- ・ データ入力はエクセルからの「コピー&ペースト」でも可能です。

図 8. MEPHAS 新規サイトにおける Shirley-Williams の方法のデータ入力ページ