

保健の科学 第36巻 1994 別刷

特集 ■ 健康調査と統計処理

調査データはHALBAUで

高木 廣文

はじめに

ある地域での食事調査をする、生活習慣を調べる、住民検診の資料を分析する、といった場合など、資料は集まったが、どうすれば自分の思ったような分析が行なえるのか困った経験はないだろうか。また、どうすればよいかは分かっているのだが、コンピュータや使用できる解析のためのソフトなどが身近にないなどの理由で、十分な統計学的な解析が行なえずに、腹立たしく感じたことはないだろうか。

20年ぐらい前までは、ちょっとした調査の集計でも、大型計算機でしかも分析用のプログラムを自作しなければならなかった。現在では、もはや大型計算機を使う必要はないし、通常の集計や解析ならば、プログラムも自作する必要はなくなった。当時のことを考えると、実に便利な世の中になったはずなのだが、その恩恵に浴していない人達も多いようである。実際にはどのようにすれば気軽に簡単に、データの集計や解析ができるのだろうか。分析のためには、パソコンは少なくとも必要であるが、集計や統計学的な分析のためには、HALBAU（はるぼう）さえあれば十分である。HALBAUといっても、初めてその名を聞く人も多いだろう。また、名前は知っているが、実際に何ができるのかはよく分かっていない人も多いものと思う。

ここでは、NEC-PC98シリーズかEPSON製の

同等のパソコンを用いて、簡単にデータ入力から集計・統計学的な解析を行なうためのパッケージ・ソフト HALBAU の簡単な紹介を行ないたい。

1 HALBAU とは何か

HALBAU は、“High-quality Analysis Libraries for Business and Academic Users” の略称であるが、一般に「はるぼう」と呼称されている。

HALBAU は、柳井・高木他（1986）による「多変量解析ハンドブック」に掲載された N88BASIC によるサブプログラムをもとに開発されたものである。当初は、柳井晴夫（大学入試センター、当時千葉大学）、市川雅教（東京外国語大学、当時東京大学大学院）、服部芳明（（株）東芝原子力事業部原子力技術研究所）、佐藤俊哉（統計数理研究所、当時東京大学大学院）、丸井英二（東京大学）および高木の6名（敬省略）が、同書の執筆とともにプログラムの開発にあたった。ただし、単に多変量解析の基本ルーチンの寄せ集めでは、実際の役に立たないので、高木がデータ入力・編集、基礎的な統計学の手法のためのプログラムなどを大幅に加えて、統計学パッケージとして統合した。同書の発行とともに、HALBAU も現代数学社より市販された。その後、佐伯圭一郎（文教大学、当時東京大学大学院）および中井里史（東京大学、当時東京大学大学院）の両氏にも改訂作業に参加して頂いた。現在は、主に高木が HALBAU に関する質問への回答、プログラムの訂正や方法の追

加などを行なっている。

HALBAU は、データ解析の考え方を基本とし、データ解説を行ないたいが行なえない環境にある研究者・学生・大学院生などに、簡単に手軽に高度な解析が行なえるソフトを提供することを試みたものである。

2 HALBAU の特徴

HALBAU は実際の調査データを解析できることを念頭に置いて開発されたものであり、以下のような特徴をもつ。

① データ入力・編集が容易でデータの変換も可能

統計学的な分析ができるだけでは、役に立たない。HALBAU はデータ入力のためのスクリーンエディタを含み、また HALBAU 形式以外のデータファイルとも、相互に交換できるような機能を持つ。

② 各種の指定が簡単

メニュー画面方式を採用し、分かりにくい指定にはヘルプ機能がついている。これにより、分析方法の指定、分析のための変数の指定などが簡単に行なえる。

③ 欠損値の処理が可能

調査データでは、対象者によっては無回答の項目があるなど、データに欠損のある場合が多い。このため、分析では欠損値の処理を自動的に行なう。また、特別な欠損値処理が必要な分析では、その処理方法を指定できる。

④ 各メニュー、指定画面、結果の表示は全て日本語

各種の指定を間違えないようにするためには、英語などではなく、日本語でメニューなどを表示することが絶対に必要である。また、分析結果の統計学用語も全て日本語で表示し、誤りが少なくなるようになっている。

⑤ エラー処理／ストップキーの処理

分析中に誤ってキーを押したり、エラーが起きた場合などに、日本語による簡単なメッセージが

画面に表示され、それに対応できるようになっている。

⑥ デフォルト処理

分析を行なう場合、手法によっては細かい指定が必要となる。全てをその都度指定するのは煩雑である。分析用のファイル名などを含め、いくつかの指定では、単にリターンキーを押すだけで設定が行なえる。

⑦ 分析できるケース数が実用的

分析できる変数の個数は、BASIC の配列の大きさの制限があるが、扱えるケース数に、実用上の制限のないことが基本となっている。ただし、実際には、プログラム上の問題で、多くの方法で約3万2千例が上限となり、分析方法によっては、より小さなケース数が上限となっている手法もあるが、できる限り実データに対応できるようにした。

⑧ 分析結果のファイル保存

解析結果やグラフをプリンターに出力するだけでなく、ディスク上のファイルに保存することができる。ファイル保存した分析結果は、一太郎などの日本語ワープロやエディタで読み込むことができるので、数値の打ち間違えがなく、報告書や論文用に作表しなおすことが簡単にできる。また、グラフは HALBAU に専用の編集プログラムがあり、簡単な修正や編集が行なえる。

⑨ 新しい分析手法の組込みが簡単

将来のユーザーの要望に応じて分析手法を、新たに HALBAU に付け加える場合、複雑な変更を必要としないように考慮してある。

これまでも、ユーザーの要望によって、グラフのファイルへの保存機能や、信頼性係数の計算プログラムなどを追加した。現在でも、分析手法の要望は受け付けている（ただし、HALBAU に追加するかは、私の個人的な見解に左右される）。

3 HALBAU の基本構成

HALBAU は大きく分けると、3つの基本的な構成要素からできている。それらは、【1. データ

表1 データおよびファイルの操作

名 称	主 な 機 能 説 明
1. データファイルの作成と編集	分析用データ入力と編集, 2つのファイルのチェーンとマージ, HALBAU用ファイルの清書出力, ファイル消去など.
2. データファイルの変換	HALBAU用データファイルと固定長/MULTIPLANのSYLK形式/dBASE II/NAP形式(カンマ区切り)の各ファイルとの相互変換.
3. データ簡易変容とファイル分割	変数の重み付・べき乗線形和の計算, 変数の再カテゴリ化, 対数変換. ファイルの分割など. 特定の変数によるソート.
4. グラフの編集・作成	HALBAUでファイルに保存したグラフィックデータの編集, もしくは簡単な作図を行なう.
5. ドライブ・ファイル等の設定	HALBAUで使用する入力用のドライブ/ファイル, 出力用のドライブ/ファイル・プリンタ, 作業用ドライブの設定.
(データ変容)*	データの変容プログラム作成用エディタと, その実行.

*メニュー画面には表示されない.

およびファイルの操作】、【2. 基礎統計学パッケージ】、【3. 多変量解析パッケージ】である. この分類は便宜的なものであるが, 分析を指定するためのメニューなども, これに基づいており, 以下にそれぞれの内容について簡単に説明する.

(1) データおよびファイルの操作

質問紙などで収集した資料を, パソコンで処理できるようにフロッピーディスク上のファイルとして保存したり, 編集したりするためのものである. 表1に示したように, メニューとして5種類,

表2 基礎統計学パッケージ

分析 方法	主 な 機 能 説 明
1. 基本統計量の計算	標本数, 平均値, (標本)分散, 標準偏差, 変動係数, 総和, 最小値, 最大値, 歪度, 尖度.
2. 項目別カテゴリの算出	質的変数の各カテゴリの度数と百分率.
3. 度数分布表とヒストグラム	適当な階級区分で度数, 累積度数, 相対度数, 累積相対度数, ヒストグラム, 折れ線グラフなどを作成.
4. 幹葉表示の作成	適当な階級区分で幹葉表示を作成.
5. クロス表の作成と検定	クロス集計と検定(カイ2乗検定, フィッシャーの正確な確率, mid-pによる正確な確率)など.
6. 四分表の併合	層別四分表の共通オッズ比, リスク比のマンテル-ハントウエル推定量などの算出.
7. 相関係数の計算と検定	ピアソンの積率相関係数/スピアマンの順位相関係数/ケンドールの順位相関係数などの算出と無相関の検定.
8. 相関図の作成	ピアソンの積率相関係数, 無相関の検定, 単回帰式など. (層別)相関図の作成, 領域図の作成など.
9. 3変数以上の同時相関図	3変数以上の相関係数行列のかわりに同時相関図を作成. 単相関係数, 単回帰式など.
10. 2群の母平均値の差の検定	独立2標本の場合, および対応のある場合について母平均値の差の検定を行なう.
11. 一元配置分散分析+ α	各群の平均値, 分散, 級間変動・級内変動・全変動およびその不偏分散, F検定. 各2群についての母平均値の差の検定.
12. 順位和検定など	ウィルコクソンの順位和検定(U検定), 対応のある場合には符号検定, ウィルコクソンの符号付順位和検定など.
13. Kruskal-Wallis+ α	クラスカル-ウォリス検定と対比較(ボンフェロニの方法, シュフェの方法).
14. 二元配置分散分析+ α^*	各要因の水準ごとの平均値, (不偏)標準偏差, 各要因変動・交互作用・不偏分散, F検定など.
15. 5数要約値と箱ヒゲ図	最小値, 第一四分位, 中央値, 第三四分位, 最大値, 四分位偏差, 箱ヒゲ図の作成など.
16. 生命表解析	カプラン-マイヤ法による生命表解析, マンテル-ハントウエル検定, ログランク検定, 一般化ウィルコクソン検定, 累積生存率の作図など.

*層別変数は指定できない.

その他に自分で簡単なプログラムを書き、それを実行するための〔データ変容〕のためのソフトから成っている。通常は、〔1〕を用いれば、データの入力、訂正、プリンタへのデータの一覧出力などが行なえる。実際には、データの入力を業者などに委託した場合などでは、固定長ファイルであったり、他の形式のファイルであったりする。そのようなデータファイルの変換も〔2〕で行なえる。また、いくつかの項目の合計を計算したり、データの対数値を計算したい場合などには〔3〕を用いればよい。また、HALBAUに含まれている分析手法で、グラフをファイルに保存することができるが、簡単なグラフの作成や編集は〔4〕を用いて行なえる。〔5〕は、分析のための各種の設定を行なうためのプログラムである。

(2) 基礎統計学パッケージ

基礎統計学パッケージには、表2で示したようなプログラムが含まれている。分布の特性値である平均値、分散、尖度、歪度などを求めたり〔1〕、度数分布表〔3〕や幹葉表示〔4〕、もしくは箱ヒ

ゲ図〔15〕などを簡単に作成できる。また、(単／順位)相関係数の計算や検定〔7〕、散布図〔8〕・〔9〕などが簡単に求められる。質的なデータでは、カテゴリごとの頻度を求めたり〔2〕、クロス集計〔5〕や疫学研究では不可欠なマンテル－ヘンツェルの方法による検定〔6〕などのルーチンもある。よく行なわれる2群の母平均値の差の検定〔10〕や分散分析〔11〕・〔12〕、さらに、ノンパラメトリックな手法を含む各種の検定方法〔12〕・〔13〕が組み込まれている。また生命表解析のためには Kaplan-Meier 法、一般化ウィルコクソン検定などがある〔16〕。

また、実際のデータ解析では、平均値を求める場合でも、性別や学年別などのように、特定の変数の値別に集計が必要なことも多い。そのような場合、「層別変数」として複数の変数を指定することで、一度に性別・学年別などの基本的な解析や集計が行なえる。

(3) 多変量解析パッケージ

基礎統計学パッケージには含まれていない、よ

表3 多変量解析パッケージ

分析方法	主な機能説明
1. 重回帰分析	偏回帰係数・標準偏回帰係数・F値・有意確率、重回帰係数・多重決定係数とF検定など。変数選択は、一括投入法、変数増加法・変数減少法、変数増減法、変数減増法、てこ比の算出、残差プロットの作成等。
2. 数量化理論1類	説明変数間のクロス表・カテゴリごとの基準変数の平均値と標準偏差、標準化数量、基準変数との偏相関係数、重相関係数など。
3. 主成分分析	主成分負荷量、固有値とその検定、寄与率、累積寄与率、主成分得点などの算出。主成分負荷量による変数の平面配置図の作成。
4. 正準相関分析	各変数の重みベクトル・構造ベクトル、各成分の固有値の検定・正準相関係数・冗長性など。
5. 数量化理論3類	各カテゴリおよびケースの重み数量、各成分の固有値など。各カテゴリの数量による平面配置図の作成。
6. 判別分析	2群および多群の判別、各変数の判別係数、パートレットのカイ2乗検定、ウィルコクスのAボックスのM検定などの算出。各群の判別得点の平均値、誤判別率など。
7. 数量化理論2類	2群および多群の判別。群ごとに説明変数間の各カテゴリのクロス表、カテゴリ数量、各群の判別得点の平均値、誤判別率など。
8. リスク関数などによる解析	多重ロジスティックモデル、多重ワイブルモデル、比例ハザードモデル、条件付きロジスティックモデルによる解析。各手法に関して各モデルに含まれる説明変数の係数、F検定、各変数のオッズ比と95%信頼区間の算出。生存解析については、期間別生存率・累積生存率など、累積生存率を作曲する。
9. 因子分析	主因子法、最尤法、共通性の初期値指定、反復推定指定可能。因子の回転は、直交回転・斜交回転を行なう。因子得点の算出、各変数については因子負荷量による平面配置図の作成など。
10. クラスター分析	項目分類、個体分類を階層的な手法により行ない、デンドログラムを作成する。
11. 尺度構成・信頼性係数	クロンバッハの α 信頼性係数、 θ 信頼性係数などを算出。各項目の第1主成分負荷量、他の変数によるSMCなどを算出する。

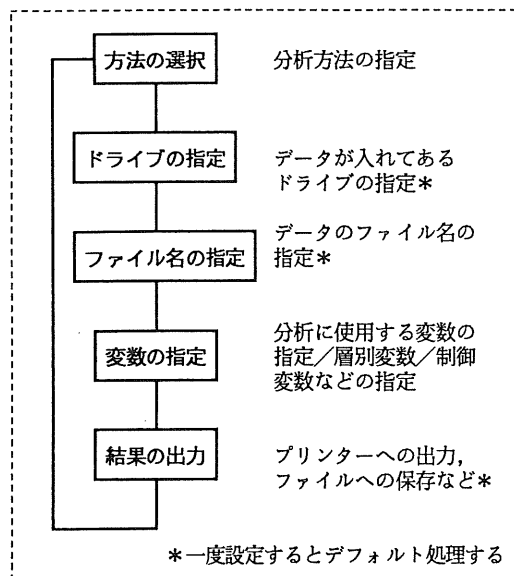


図1 HALBAUの基本的な操作

り複雑な事象間の関係进行分析するために作成されたもので、表3に示したように11種類の方法からなる。

各種の「予測」のために、重回帰分析[1]と数量化理論1類[2]がある。「識別問題」のため

には判別分析[6]と数量化理論2類[7]がある。また、変数間に存在する「内的構造分析」のためには、因子分析[9]と数量化理論3類[5]があり、「情報圧縮」のためには主成分分析[3]がある。さらに、変数群の関係を分析するためには正準相関分析[4]がある。「個体の分類」や「項目の分類」のためにはクラスター分析[10]がある。疾患の原因や死亡に対する特定の変数のリスクを解析するためには、多重ロジスティックモデルや比例ハザードモデルなどによる分析が簡単に行なえる[8]。さらに、QOLや生活習慣などを測るために、新たな質問項目から尺度構成を行なう場合などには、クロンバッハの α 信頼性係数などが、[11]により簡単に求められる。

4 HALBAU の操作と実行画面

HALBAUで分析を行なう場合の基本的な操作は図1で示したような流れになる。この図は、調査資料をコンピュータで読めるように加工した後の、分析を行なうときのものである。基本的な操作としては、①行ないたい分析方法を指定する、

HALBAU-4 メニュー	
1. データおよびファイルの操作	
2. 基礎統計学パッケージ	
3. 多変量解析パッケージ	
4. 終了	

基礎統計分析プログラム メニュー	
1. 基本統計量の計算	
2. 項目別カテゴリ度数の算出	
3. 度数分布表とヒストグラム	
4. 幹葉表示の作成	
5. クロス表の作成と検定	
6. 四分表の併合	
7. 相関係数の計算と検定	
8. 相関図の作成	
9. 3変数以上の同時相関図	
10. 2群の母平均値の差の検定	
11. 一元配置分散分析+ α	
12. 順位和検定など	
13. Kruskal-Wallis 検定+ α	
14. 二元配置分散分析: 母数モデル	
15. 5数要約値と箱ヒゲ図の作成	
16. 生命表解析	

移動: [→] [←] [↑] [↓]
確定: [↵] リターンキー

図2 HALBAUバージョン4のメニュー画面

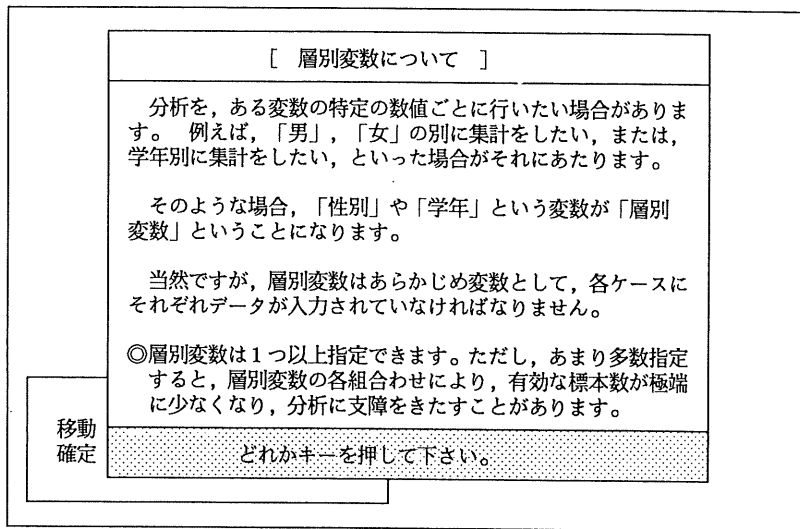


図3 層別変数についてのヘルプ画面

②データファイルの入っているドライブを指定する, ③そのファイル名を指定する, ④分析に使用する変数などを指定する, ⑤分析結果を画面で見たりファイルに保存したりする, ⑥次の分析に移る, という手順を示したものである。なお, ②と③の分析用のデータファイルの指定は, 一度指定すればリターンキーのみを押すだけで設定される。このように, HALBAU を用いれば, データファイルさえ作成してしまえば, その後は極めて簡単な操作で解析が行なえる。

実際に HALBAU を始動すると, 図2のような画面が表示される。メニュー画面で分析方法などを指定するためには, その番号をキー入力してもよいのだが, カーソルキーを使って指定することができる。

各種のメニュー画面を表示中に[HELP]キーを押すと, ヘルプ画面が表示され, 簡単な説明が得られる。図3は層別変数のヘルプ画面である。ヘルプ画面の表示の形式は, 他の場合(分析方法の指定, 変数の指定, 制御変数の指定, 欠損値の指定など)でも基本的には同一である。なお, 画面に表示されている内容で十分に操作が理解できると考えられる場面では, ヘルプ機能はつけられていない。そのような場合, 画面の右下側に, 落ち

着いて指示をよく読むようにメッセージが表示される。

どのような分析方法を用いたのであれ, 分析結果の出力を表示すると, 図4のような形式の画面表示になる。図4は基本統計量の出力画面を表示したものである。出力画面はカーソルキー[↑]・[↓], および[ROLL DOWN]・[ROLL UP]キーで, 画面表示をスクロール・逆スクロールさせ, 見たい結果を表示するようになっている。画面の最下段に表示されているように, 終了させたい場合には, [ESC]キーを押せばよい。その後, 分析結果を出力用のファイル, プリンタなどに出力するかをきいてくるので, 必要に応じてプリンタやファイルに出力を行なう。

図5は主成分分析での, 変数の主成分負荷量による平面配置図の1例である。

HALBAU のどの作図画面でも, 画面上の右側にグラフに関するメニューが表示される。図5では, 画面上のグラフの出力に関するメニューが表示されている。すでに述べたように, 保存したグラフファイルは, HALBAU のグラフ編集用プログラムで, 簡単な訂正や編集が行なえる。

[分析結果の画面出力]					
[基本統計量解析結果] File = life.hal 94/06/25					
層別変数名 ----- 層の範囲 -----					
1)性					
層 1	1.000	～	1.000		
層 2	2.000	～	2.000		
[全体]					
変数名	標本数	平均値	分散	標準偏差	変動係数
1)年齢	50	57.700	80.970	8.998	15.595
2)身長	49	155.912	44.373	6.661	4.272
3)体重	50	57.012	80.687	8.983	15.756
4)最高血圧	50	127.460	216.008	14.697	11.531
5)最低血圧	50	80.320	93.618	9.676	12.046
総 和		最小値	最大値	歪度	尖度
[↑]上 [↓]下 [ROLL DOWN]前画面 [ROLL UP]次画面 [ESC]終了					

図4 結果の出力画面

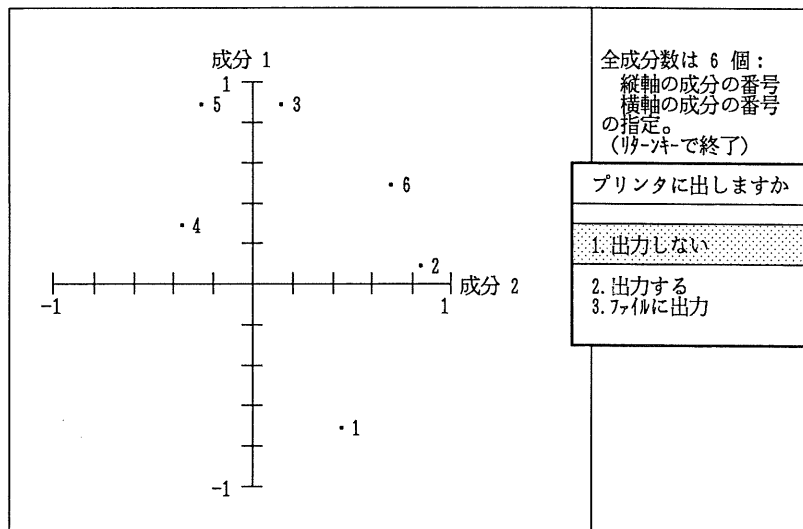


図5 主成分負荷量による変数の平面配置

5 HALBAU の問題点について

HALBAU のもつ問題点で最も強調すべきことは、統計学の誤用を助長するかもしれないという点にあるだろう。含まれる手法は極めて多数であ

り、あまりにも簡単に様々な分析ができるので、使用者が必ずしも最適な方法を用いてくれる保証はない。さらに、結果の解釈が正しく行なえるか疑問もあり、今後どのようにその対策をプログラムに組み込むかが当面の最大課題である。ただし、このような問題は、全ての統計学パッケージでの

問題でもある。この解決にはユーザー自身が、データ解析の各手法を正確に理解し、幅広い一般教養および専門的知識を身に付け、さらに直観力、知恵、知力などを育成する以外には方法がないのかもしれない。

参考文献

柳井晴夫, 高木廣文 編著 (1986) : 多変量解析
ハンドブック, 現代数学社

高木廣文, 佐伯圭一郎, 中井里史 (1989) : HAL-
BAU によるデータ解析入門, 現代数学社
高木廣文 (1994) : HALBAU-4 マニュアル, I,
基本編, 現代数学社
高木廣文 (1994) : HALBAU-4 マニュアル, II,
基礎統計学, 現代数学社
高木廣文 (1994) : HALBAU-4 マニュアル,
III, 多変量解析, 現代数学社

お知らせ

一般向けセミナー

癌のはなし

今回は、特に日本人に多い癌をとりあげ4回シリーズで、セルフチェックの仕方から診断、治療の最新情報、癌との付き合い方などについて、お話していただきます。

日 時：平成6年9月19日(月) 『乳癌』 PM 2:00～4:00
10月17日(月) 『肺癌』 PM 2:00～4:00
11月10日(木) 『消化器系の癌』 PM 2:00～4:00
12月9日(金) 『卵巣・子宮癌』 PM 2:00～4:00

申し込み・問い合わせ先

〒102 東京都千代田区平河町2-7-5 砂防会館5階
(財)ライフプランニングセンター TEL. 03(3265)1907

お知らせ

医療職向けセミナー

心電図基本コース

—これから心電図を学ぼうとする初心者の方のために—

日 時：平成6年10月17日(月) PM. 6:00～8:00
10月24日(月) PM. 6:00～8:00

講 師：久代登志男先生(日本大学医学部講師 駿河台日大病院循環器科)

申込方法 問い合わせ

〒102 東京都千代田区平河町2-7-5 砂防会館5階
(財)ライフプランニングセンター 心電図基本係 TEL. 03(3265)1907