

# 統計解析ソフトウェア「SYSTAT」の利用法（その2）

## － SYSTAT による生存分析 －

医学部・地域医療科学教育研究センター 竹生政資  
takefu@cc.saga-u.ac.jp

### 1. はじめに

前回の広報では、医学サブセンターの教育用パソコン端末に導入された統計解析ソフトウェア「STSTAT J10.2」について、Microsoft Excel には含まれていないノンパラメトリック検定の例として Wilcoxon 符合付順位和検定と Wilcoxon 順位和検定（独立2群の場合として「Kruskal-Wallis」で代用）の二つの例を紹介した。

今回は、医学分野の研究においてよく利用されている生存分析について SYSTAT を用いた解析例を紹介する。また、SYSTAT で描画されたグラフの編集方法についても紹介する。生存分析プログラムは Excel の分析ツールや統計関数には含まれていないので、生存分析を行うためにはどうしても SYSTAT のような統計解析パッケージソフトウェアが必要となる。SYSTAT のほかにも SAS や SPSS や StatView などの統計解析パッケージソフトウェアも生存分析プログラムをもっている（これらは医学サブセンターのパソコンにはインストールされていない）。

### 2. SYSTAT による生存分析

生存分析の目的は、観察された生存データから生存関数を予測することと、生存データがいくつかのグループから構成されているときにそれぞれのグループ間で生存関数に有意差があるかどうかを検定することである。以下では、二つのグループの舌がん患者の生存データを例にとりて説明する。このデータは、「医学統計概論」、R.F.モールド著、日野寛三・田中真理子訳、p.165～167 の表 15.1 に掲載されているものである。

解析に用いるデータを表 1 に示す。このデータは、英国ハンマースミス病院で 1934-1938 年の間に口腔がんの放射線治療を受けた患者に対して、その後 1967 年に追跡調査したものである。A グループの 20 症例は前部 2/3 の舌がんで、B グループの 28 症例は後部 1/3 の舌がんである。A と B の二つのグループは、舌がんの解剖学的部位の違いによるものである。

表 1 の第 1 列は「治療グループ」を識別する記号で、今の例では A グループか、B グループかを表している。

表 1 の第 2 列は「数値化グループ」で、第 1 列の「治療グループ」を「数字」で表したものである。SYSTAT の生存分析プログラムでは、生存データのグループ別としては「数値」しか受け付けないので、解析に先立ってグループ別を数値化しておく必要がある。もっとも手っ取り早いのは、1, 2, 3, ... のように整数でもって数値化することである。なお、以下ではグループが 2 つの場合を例にとりて説明するが、SYSTAT は 3 つ以上のグ

グループに対しても同様に解析することができる。

表1の第3列は「生存期間(週)」で、今の場合、その患者が放射線治療を受けてから何週間生存し続けたかを示す。ここでは便宜上、生存期間の短いものから長いものの順にソートしてあるが、SYSTATに入力するデータには順序についての制限はない。したがって、表2のようにグループごとにまとめて入力してもよい(各グループ内で生存期間でソートする必要もない)。

治療グループ	数値化グループ	生存期間(週)	打ち切り情報
B	2	13	1
B	2	17	1
A	1	18	1
B	2	22	1
A	1	25	1
B	2	26	1
B	2	27	1
A	1	30	1
B	2	31	1
A	1	34	1
A	1	35	1
A	1	36	1
B	2	37	0
A	1	38	1
A	1	39	1
B	2	40	1
A	1	43	1
B	2	44	1
B	2	47	1
B	2	48	1
A	1	60	1
B	2	61	1
B	2	65	1
A	1	68	0
B	2	69	0
B	2	71	1
A	1	77	0
B	2	78	1
A	1	82	0
B	2	83	1
A	1	86	1
B	2	88	0
B	2	95	1
B	2	108	1
A	1	121	0
B	2	123	1
A	1	130	1
B	2	186	1
B	2	213	0
A	1	324	0
B	2	481	1
A	1	520	1
A	1	546	0
B	2	568	1
B	2	685	1
B	2	711	1
A	1	819	1
B	2	1105	1

表1. 生存データ

治療グループ	数値化グループ	生存期間(週)	打ち切り情報
A	1	18	1
A	1	25	1
A	1	30	1
A	1	34	1
A	1	35	1
A	1	36	1
A	1	38	1
A	1	39	1
A	1	43	1
A	1	60	1
A	1	68	0
A	1	77	0
A	1	82	0
A	1	86	1
A	1	121	0
A	1	130	1
A	1	324	0
A	1	520	1
A	1	546	0
A	1	819	1
B	2	13	1
B	2	17	1
B	2	22	1
B	2	26	1
B	2	27	1
B	2	31	1
B	2	37	0
B	2	40	1
B	2	44	1
B	2	47	1
B	2	48	1
B	2	61	1
B	2	65	1
B	2	69	0
B	2	71	1
B	2	78	1
B	2	83	1
B	2	88	0
B	2	95	1
B	2	108	1
B	2	123	1
B	2	186	1
B	2	213	0
B	2	481	1
B	2	568	1
B	2	685	1
B	2	711	1
B	2	1105	1

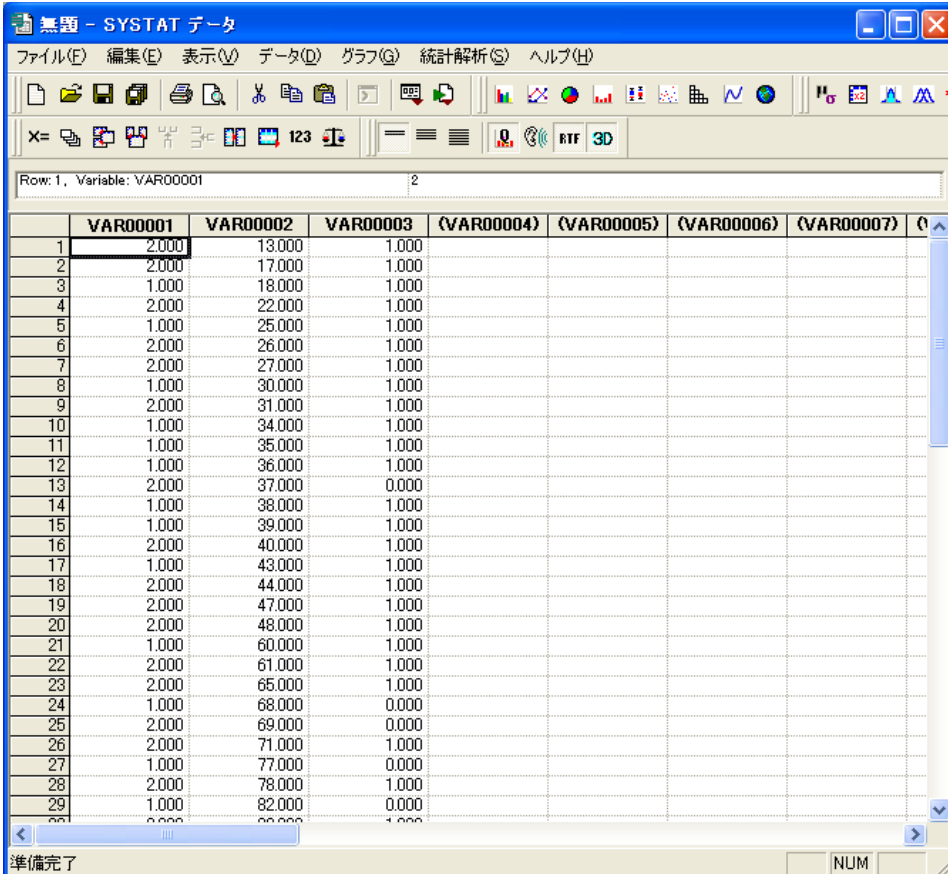
表2. 生存データ(別の形式)

表1の第4列は「打ち切り情報」で、「1」は「打ち切りのない」データであることを示し、今の例では患者が舌がんで死亡したことを意味する。一方、「0」は「打ち切った」データを表し、患者が追跡調査の時点まで生存していたか、あるいは患者が舌がんとは関係のない他の病気などで死亡したことを意味する。なお、「打ち切り情報」の表現は統計解析ソフトウ

ウェアによって異なるので注意を要する。例えば、StatView の場合、SYSTAT とは逆に、「打切りのない」データを「0」、「打切った」データを「1」として表すようになっている。

さて、表1（あるいは表2）のデータを実際に SYSTAT に入力して生存分析を行ってみよう。まず以下の手順によって生存データの入力を行う。

- (1) SYSTAT を起動して「ファイル」メニューから「新規作成」をクリックし、続いて「データ」をクリックする。
- (2) データウィンドウが開くので、これに表1（または表2）の第2列、第3列、第4列のデータを入力する（図1）。治療グループの区別は第2列の「数値化グループ」の情報によって行われるので、第1列の「治療グループ」は入力する必要はない。もし元データが Excel によって入力されている場合には、まずデータの入っている Excel ワークシートを表示させ、データ領域の範囲指定を行い、編集メニューから「コピー」を実行し、続いて SYSTAT のデータウィンドウを表示させ、「編集」メニューの「貼り付け」を実行することにより SYSTAT にデータをコピーすることができる。



	VAR00001	VAR00002	VAR00003	(VAR00004)	(VAR00005)	(VAR00006)	(VAR00007)
1	2.000	13.000	1.000				
2	2.000	17.000	1.000				
3	1.000	18.000	1.000				
4	2.000	22.000	1.000				
5	1.000	25.000	1.000				
6	2.000	26.000	1.000				
7	2.000	27.000	1.000				
8	1.000	30.000	1.000				
9	2.000	31.000	1.000				
10	1.000	34.000	1.000				
11	1.000	35.000	1.000				
12	1.000	36.000	1.000				
13	2.000	37.000	0.000				
14	1.000	38.000	1.000				
15	1.000	39.000	1.000				
16	2.000	40.000	1.000				
17	1.000	43.000	1.000				
18	2.000	44.000	1.000				
19	2.000	47.000	1.000				
20	2.000	48.000	1.000				
21	1.000	60.000	1.000				
22	2.000	61.000	1.000				
23	2.000	65.000	1.000				
24	1.000	68.000	0.000				
25	2.000	69.000	0.000				
26	2.000	71.000	1.000				
27	1.000	77.000	0.000				
28	2.000	78.000	1.000				
29	1.000	82.000	0.000				

図1. 生存データの入力

SYSTAT のデータウィンドウは、デフォルトでは数値の表示が小数点以下3桁までとな

っているので、例えば第1列の「治療グループ」は本来「2」と表示されるべきところが「2.000」のように表示される。同様に、第3列の「打切り情報」も「1」となるべきところが「1.000」のように表示される。これは単に表示の問題であり、このままでも生存分析には何ら支障はないが、必要であれば以下の手順によって表示方法を変えることができる。

- (1) SYSTAT のデータウィンドウの第1列の変数名「VAR00001」を右クリックするとサブメニューが表示されるので「変数のプロパティ」を選択する。
- (2) 図2のようなウィンドウが表示されるので、「数値表示オプション」の項目の「小数点位置」がデフォルトの「3」となっているので、ここを「0」に変更する。
- (3) 同様に、第3列の変数名「VAR00003」についても「小数点位置」を「0」に変更する。



図2. データの表示方法の変更

以上で生存分析に必要なデータ入力が終わったので、次に、実際に解析を実行するために、SYSTATのメインウィンドウ（「SYSTAT アウトプット オーガナイザ」ウィンドウ）の「統計解析」メニューから、「生存分析」を選択する。すると、図3のような変数設定ウィンドウが表示されるので、以下の手順によって変数設定を行う。

まず、左側のボックスの変数名「VAR00001」をクリックしてから、右下の「層別変数」ボックスの左側に表示されている「追加ボタン」をクリックする。すると、このボックスに「VAR00001」と表示される。これは SYSTAT のデータウィンドウに入力されている第1列（「VAR00001」）のデータを「層別変数」、つまり「グループ識別」データとして設定

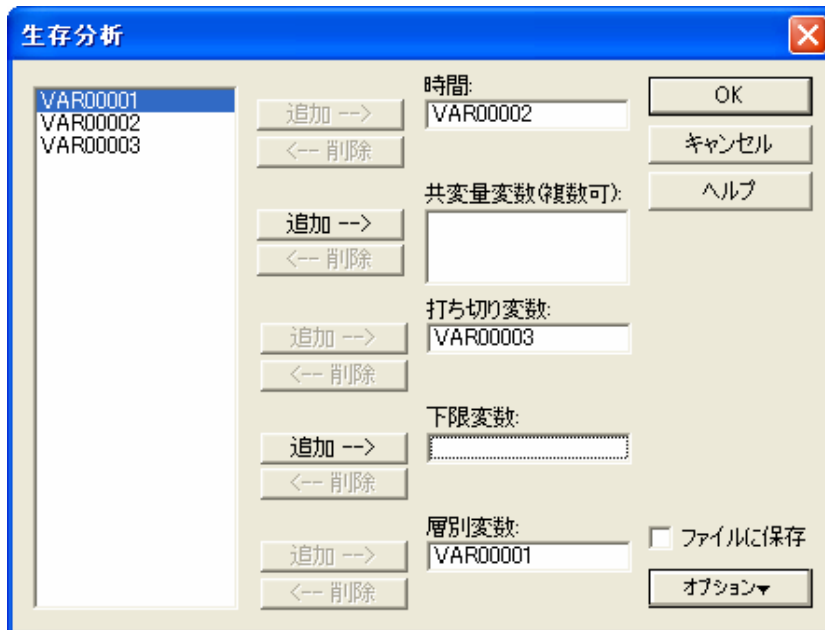


図 3. 生存分析の変数設定

することを意味する。

次に、左側のボックスの変数名「VAR00002」をクリックしてから、右上の「時間」ボックスの左側に表示されている「追加ボタン」をクリックする。すると、このボックスに「VAR00002」と表示される。これは SYSTAT のデータウィンドウに入力されている第 2 列（「VAR00002」）のデータを「時間」、つまり「生存時間」データとして設定することを意味する。

最後に、左側のボックスの変数名「VAR00003」をクリックしてから、右側の「打ち切り変数」ボックスの左側に表示されている「追加ボタン」をクリックする。すると、このボックスに「VAR00003」と表示される。これは SYSTAT のデータウィンドウに入力されている第 3 列（「VAR00003」）のデータを「打ち切り変数」データとして設定することを意味する。

以上の設定が終わったら、右上の「OK」ボタンをクリックする。すると図 4 のような結果が得られる。なお、SYSTAT のメインウィンドウはデフォルトでは下側に「コマンドペイン」ウィンドウが表示されるようになっているが、マウスを使って対話的に解析を行う場合にはこのウィンドウは必要ないので「表示」メニューから「コマンドペイン」を選択することによってこれを非表示にし、解析結果が表示される領域をできるだけ広げることができる（図 4）。

図 4 には、 Kaplan-Meier 法（あるいはプロダクト・リミット法とも言う）によって推定された二つのグループの生存関数が Survival Plot グラフとして重ねて描画される。またこのグラフの下には、この二つのグループの生存関数の間に有意差があるかどうかを

検定するための p-value (p 値) が表示される. p-value は MANTEL 法 (Logrank 法), BRESLOW-GEHAN 法, TARONE-WARE 法の 3 つの方法によって計算されたものが出力され, 今の場合, p-value はそれぞれ, 0.623, 0.691, 0.662 であり, いずれも有意水準 0.05 よりも大きいため, 「図 4 の二つ生存関数グラフの間に有意差はない」と結論することができる. つまり, 舌がんの放射線治療後の生存率は解剖学的部位の違い (前部 2/3 の舌がんか, あるいは後部 1/3 の舌がんかの違い) による有意差はない, という結論が得られる.

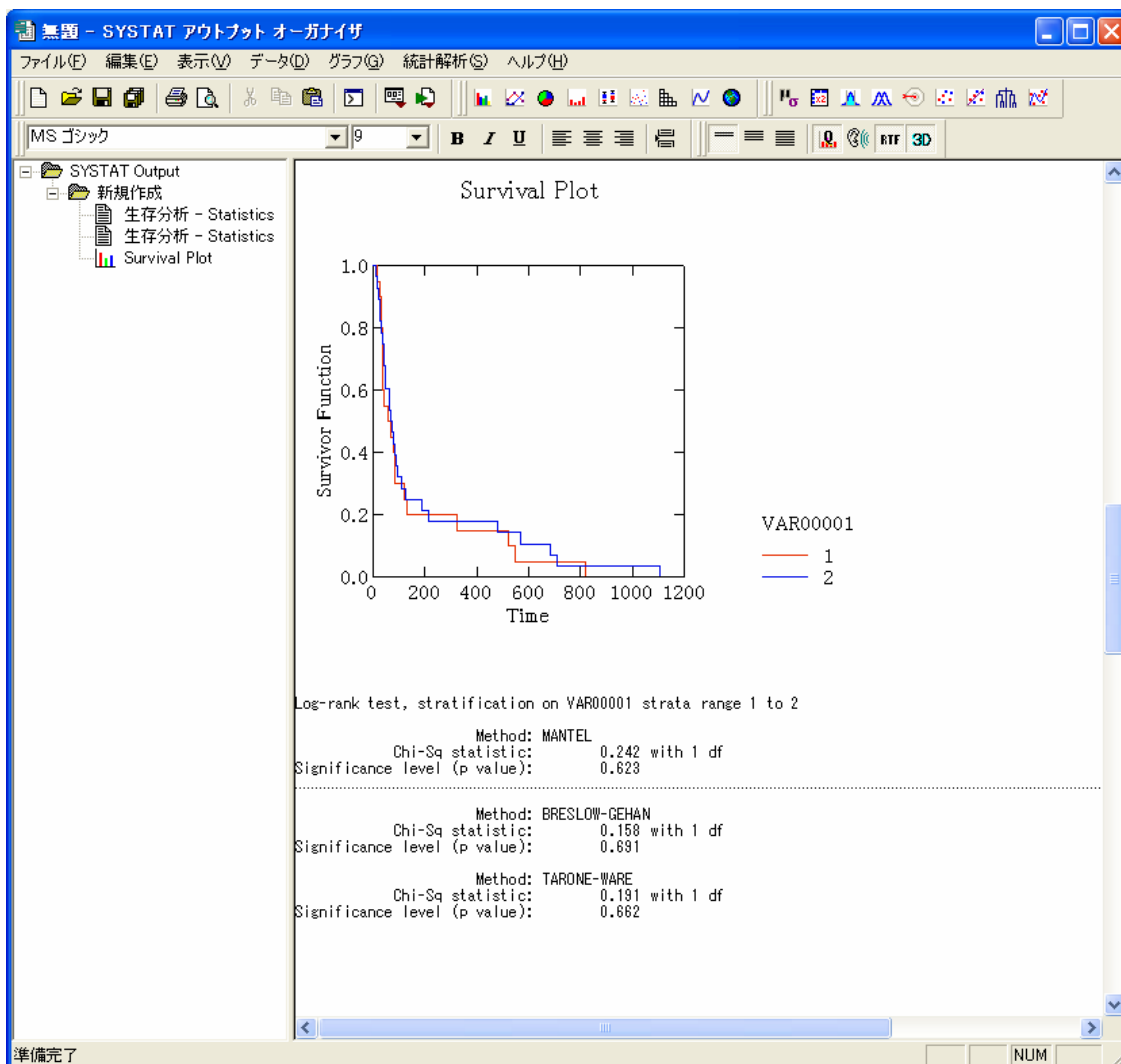


図 4. 生存分析結果 (Survival Plot と p-value)

図 4 の結果が表示されている SYSTAT のメインウィンドウの右側のスクロールバーを上側にスクロールさせると, 推定された生存関数の具体的な数値データを表示させることもできる (図 5). あるいは, SYSTAT のメインウィンドウの左側の枠の上側に表示されている「生存分析 - Statics」, 「生存分析 - Statics」, 「Survival Plot」, の部分をそれぞれマ

ウスでクリックすると、各グループの生存関数の数値リストや Survival Plot グラフを右側の枠に適切に表示させることができる。

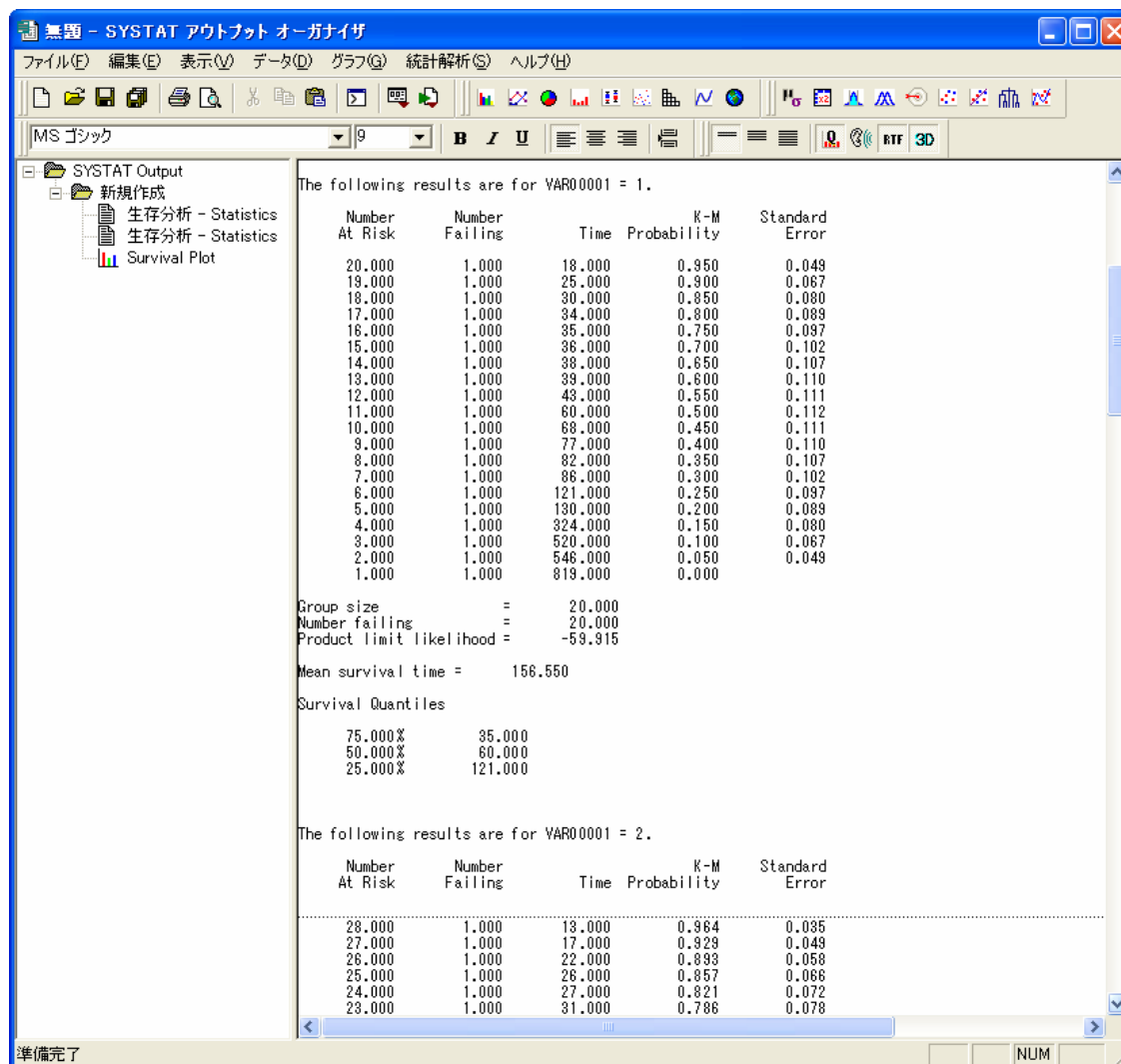


図5. 生存関数の数値リスト

以上が標準的な生存分析の方法であるが、最後に二つのことを補足しておこう。まず一つは生存分析のオプション設定についてである。図3の生存分析の変数設定ウィンドウの右下にある「オプション」ボタンをクリックするとサブメニューが表示されるので、ここで「テーブルとグラフ」を選択すると、図6のような生存分析のオプション設定ウィンドウが表示される。各設定項目の意味については、ウィンドウの右上から三番目の「ヘルプ」ボタンをクリックすると図7のようなオンラインヘルプが表示されるので、これを参考にして設定することができる。

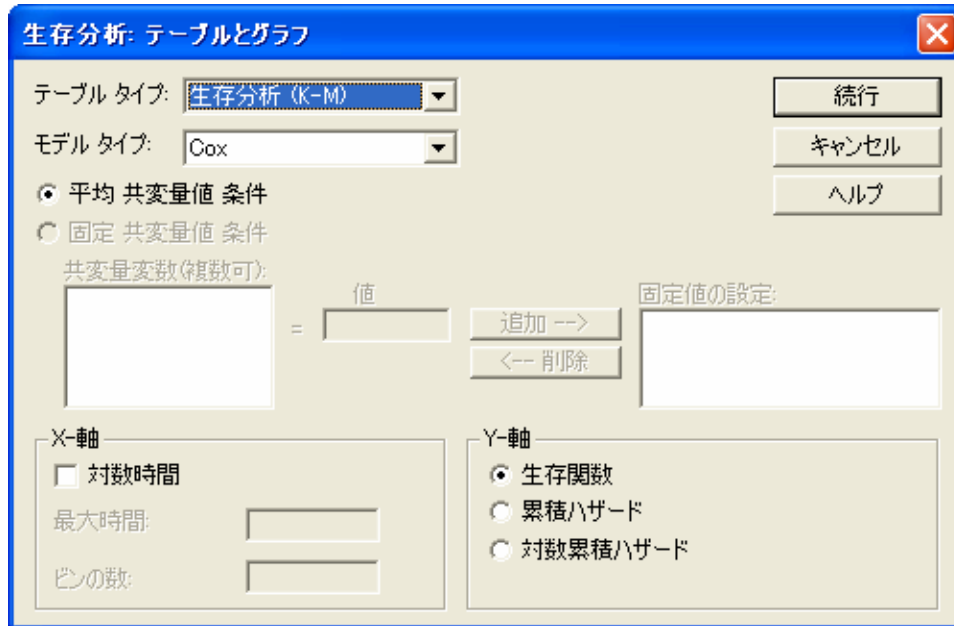


図 6. 生存分析のオプション設定

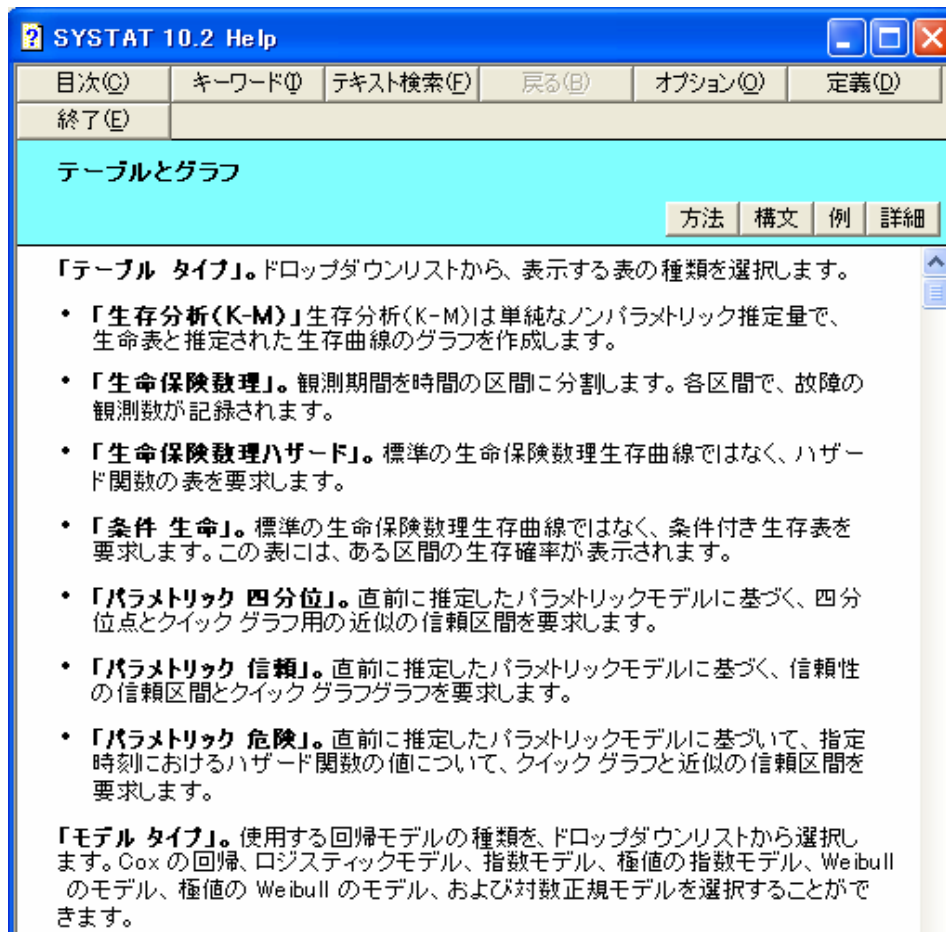


図 7. 生存分析のオプション設定のヘルプ画面 (一部)



もう一つの補足は、SYSTAT で描画された図形の編集方法についてである。例えば、図 4 の Survival Plot グラフの軸やレイアウトを編集するためには、図の上でダブルクリックすると、グラフ編集ウィンドウが表示される (図 8)。ここで「グラフ」メニューから「軸」や「レイアウト」を選択してグラフを編集することができる。あるいはツールバーの直線描画や四角形描画や文字描画のアイコンを選択してから、直接グラフの中に図形や文字を入力することもできる。しかし、ここでできる編集機能は限られており、むしろグラフを Microsoft PowerPoint に読み込んで編集するのが簡単である。そのためには、図 8 のグラフ編集ウィンドウの「編集」メニューから「グラフをコピー」をクリックしてから、PowerPoint を起動して新しいスライドを表示させ、「編集」メニューから「貼り付け」を実行する。すると SYSTAT のグラフが PowerPoint のスライド上にコピーされるので、引き続き PowerPoint の左下の「図形の調節」メニューから「グループ解除」を実行する。すると「これはインポートされた図で、グループではありません。これを Microsoft Office 描画オブジェクトに変換しますか？」という警告が表示されるので「OK」ボタンをクリックする。続けて、もう一度「グループ解除」を実行する (グラフが単純な場合「グループ解除」は一回で済むこともある)。これで SYSTAT のグラフを PowerPoint の描画オブジェクトに変換することができ、線や文字などを PowerPoint 上で自由に編集することができるようになる。

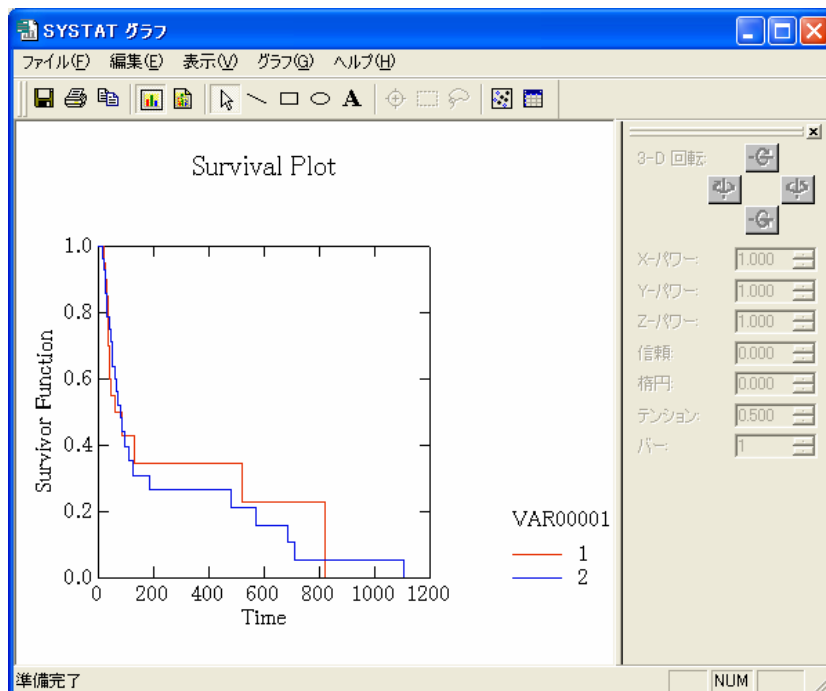


図 8. グラフ編集ウィンドウ

## 5. おわりに

本稿では、前回の広報記事に続いて、医学サブセンターの教育用パソコン端末に導入された統計解析ソフトウェア「STSTAT J10.2」の利用法として、放射線治療をほどこした二種類の舌がん患者の生存データを例にとり、具体的な生存分析の解析例を紹介した。

SYSTAT は、これまでに紹介した統計解析のほかにも、分散分析や回帰分析やクラスター分析などさまざまな統計解を行うことができる。また豊富なグラフィックス機能も持っている。さらに、統計解析だけでなく、フーリエ解析のような時系列解析の機能も装備している（ただし、データウィンドウ機能と「逆」変換機能はない）。特に、解析結果を自動的にグラフ表示する機能は Excel よりも洗練されており便利である。オンラインマニュアル等を参考にしてぜひ活用していただきたい。