

SAS Enterprise Guide によるデータ解析入門

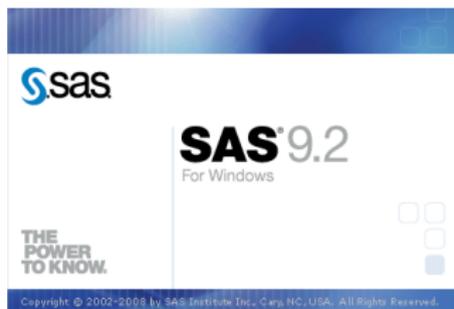
長島 健悟
城西大学 薬学部

Kengo NAGASHIMA
Laboratory of Biostatistics,
Department of Pharmaceutical Technochemistry,
Josai University

2011 年 11 月 15 日

本日の内容

- はじめに (SAS / SAS Enterprise Guide について)
- データ入力と読み込み
- 分割表データの解析
- 連続データの解析



SAS とは

SAS (Statistical Analysis System)

- 統計解析ソフトウェアの最高峰, 127ヶ国, 約 50,000 機関で利用され, 世界でも利用者数が多い統計解析パッケージの一つ
- 教育研究機関, 製薬企業, 医療機関での導入実績が多い
- 解析結果の信頼性は高いと認知されており, 論文執筆の上でも役立つ
- **ただし, 自由自在に解析するためにはプログラムを覚えなくてはいけない...**



SAS プログラムの例

```

188 *****/
189 proc datasets lib = work; delete temp;
190 ods exclude all;
191 proc genmod data = &data;
192 where TissueArea ^= ;
193 class group;
194 model &y. = group &z.
195 %do j = 1 %to &vnum;
196   &&vn&j.
197 %end;
198 / link = id dist = n type3;
199 ods output ModelFit = temp;
200 run;
201 data temp; set temp(obs=7 firstobs=7); keep value;
202 ods select all;
203 /*****/
204 data &out;
205 set %if "&i." ^= "1" %then &out; temp;
206 %end;
207 data &out; merge &out. allvarcomb;
208 proc datasets lib = work; delete allvarcomb Info temp;
209 run; quit;
210 proc rank data = &out. out = &out;
211 var value; ranks &out.rank;
212 proc sort; by &out.rank;
213 %mend modelselect2;
214
215 %modelselect2(Tb_Sp Ob_S_BS N_Ob_B_Pm ES_BS Oc_S_BS N_Oc_B_Pm, BV_TV, , bone2, aic1);
216 proc print data = aic1(obs=5); run;
217 %modelselect2(BV_TV Ob_S_BS N_Ob_B_Pm ES_BS Oc_S_BS N_Oc_B_Pm, Tb_Sp, , bone2, aic2);
218 proc print data = aic2(obs=5); run;
219 %modelselect2(BV_TV Tb_Sp N_Ob_B_Pm ES_BS Oc_S_BS N_Oc_B_Pm, Ob_S_BS, , bone2, aic3);

```

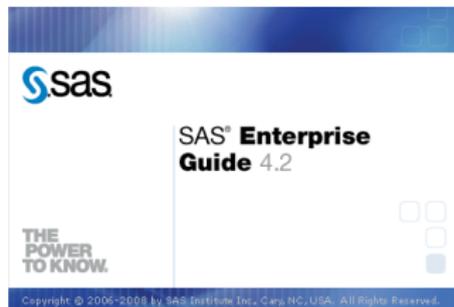
習得には非常に
時間がかかる・・・

SAS プログラムは大変すぎる...

- 統計解析を行う事が目的なのに、プログラム言語を覚えるために時間を浪費しすぎるのはもったいない
- 統計解析を行うために学ぶべきことは、データを読むための力を付ける事 (グラフ化) と、手法を適切に利用するための知識を付ける事 (統計的推定など) と、ソフトウェアパッケージを使った解析方法

SAS Enterprise Guide とは

- マウス操作だけで解析できる, Windows 用 SAS インターフェース
- プログラムを組む事なしに, SAS による信頼性の高い解析を実施できる
- データを読み込み, 手法, 変数, オプションをクリックして実行
- 実施した解析内容が記録され, 結果のレポートとともに保存できる

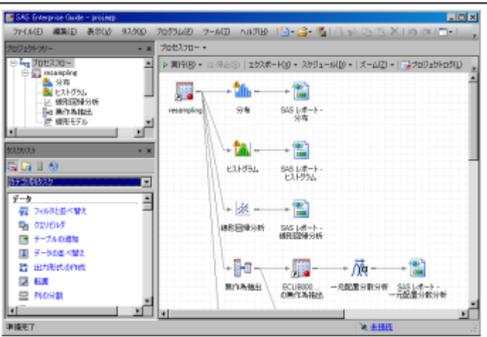
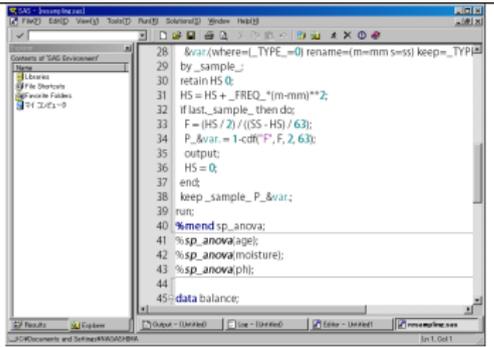


SAS と SAS Enterprise Guide の違い

- SAS: SAS プログラムを自分で記述する
- Enterprise Guide: 操作にしたがって, SAS プログラムを生成してくれる

ソフト名	SAS	Enterprise Guide
------	-----	------------------

画面



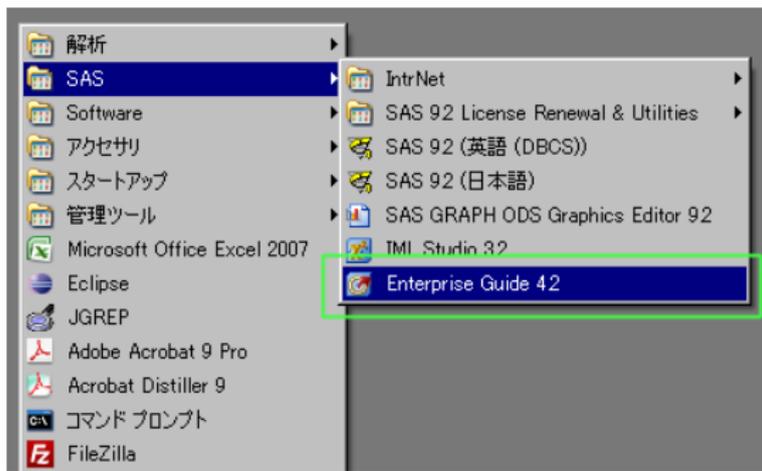
操作 対応範囲	難しいプログラム... 応用を含む解析	マウス操作のみ! 基本的な解析
------------	------------------------	--------------------

本演習の目標

SAS Enterprise Guideを用いた
データ解析手順を理解し実施できること

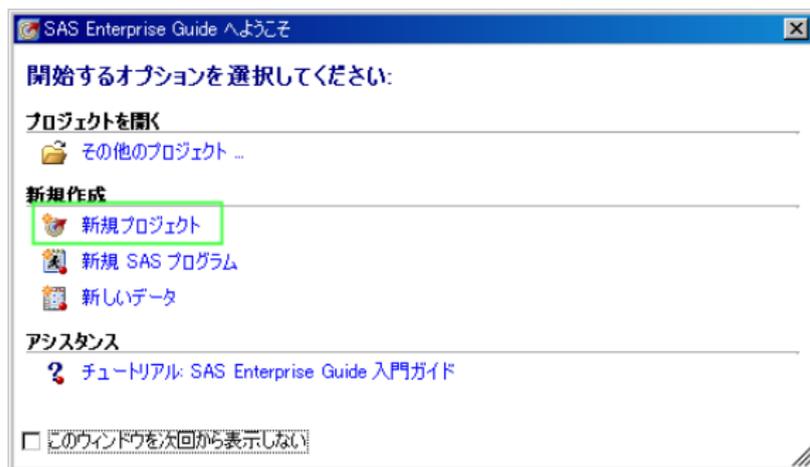
起動方法

- SAS Enterprise Guide が適切にインストールされた PC 上で
- スタートメニュー → SAS → Enterprise Guide 4.2 など



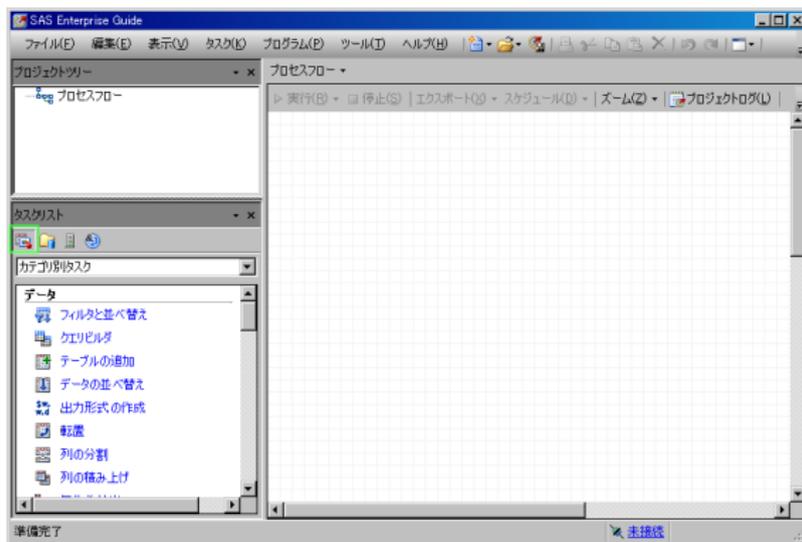
起動ダイアログ

- プロジェクトは、読み込むデータ、解析する方法とオプション、結果をひとまとめに記録したファイル、行った解析を全て記録しておく事ができる
- プロジェクトを開く: 最近使ったプロジェクトが表示される
- 新規作成: 新規にプロジェクト等を作成する
- アシスタンス: チュートリアルを表示する



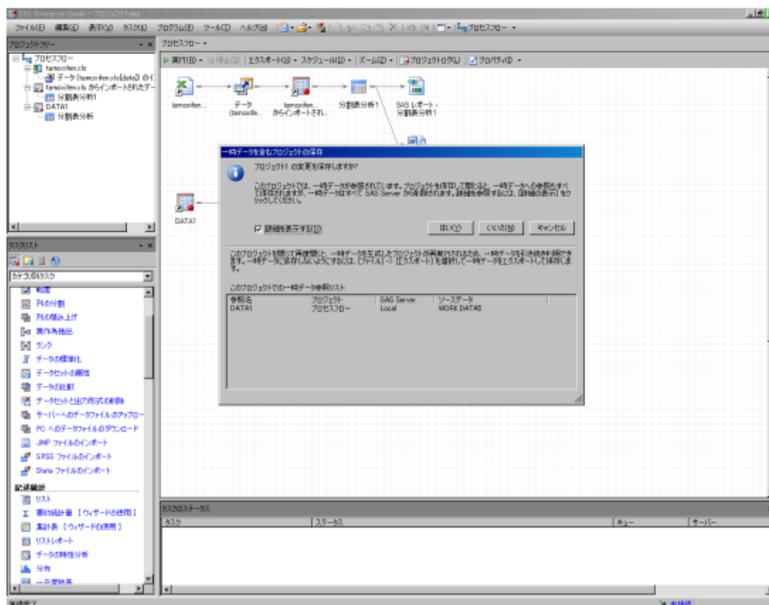
初期画面

- 左上: プロジェクト画面, 複数プロジェクトの切り替えなど
- 左下: リスト, いくつかの操作リストが表示できる, 一番左のタスクリストを表示しておくと便利
- 右: メイン画面, データや解析手法の設定, 結果の表示など



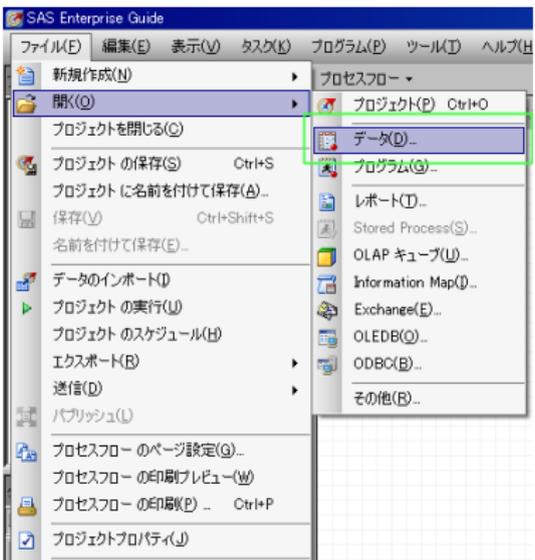
終了方法

- 右上の [×] をクリック
- 未保存のファイル等がある場合警告が表示される
 - 変更を保存して終了、保存せずに終了、またはキャンセル



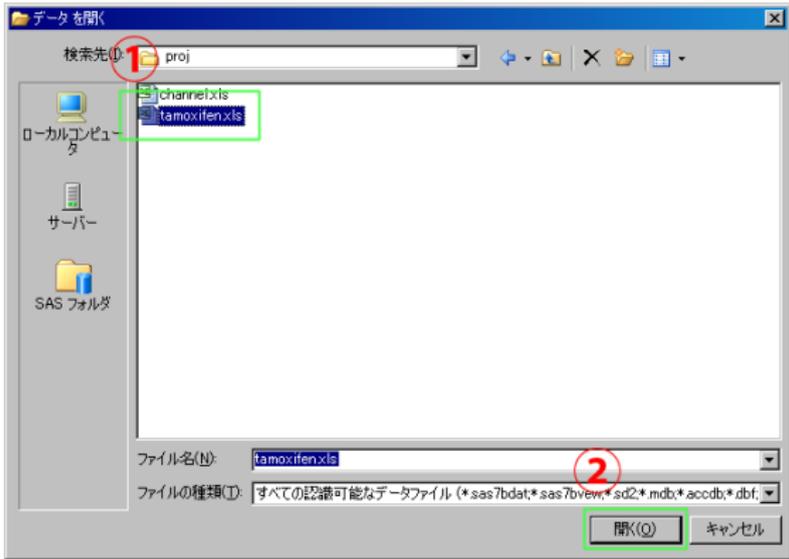
読み込みの操作

- [ファイル] → [開く] → [データ]



読み込み対象ファイルの選択

- データを選択 → [開く] (今回は tamo.xls (Excel ファイル) を指定)



読み込み対象ファイルの確認

- [次へ] をクリック



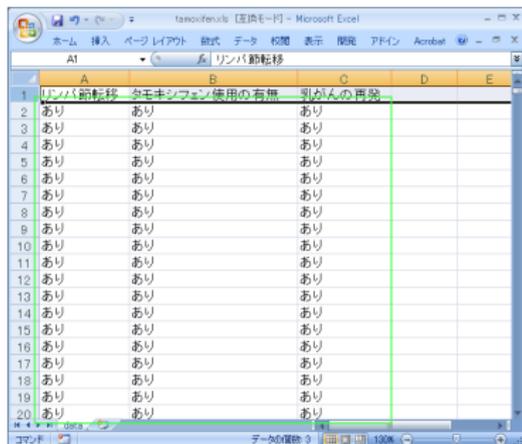
ワークシートのチェックと設定

- シート名の選択: [ワークシートを使用する] にチェックし、解析対象のシート名を選択する
- データの一行目に変数名の場合、チェックする
- [次へ] をクリック

The screenshot shows two windows side-by-side. On the left is Microsoft Excel with a spreadsheet containing data. On the right is the SAS Data Import Wizard dialog box. The dialog is titled '2 / 4 データソースの選択' (2 / 4 Select Data Source). It has a 'sas' logo in the top right. The 'Worksheet selection' section has a checked box for 'ワークシートを使用する(必)' (Use worksheet (required)) and a list box containing 'data'. A red circle '1' is around the 'data' selection. To the right, there is a checked box for '範囲の先頭行にフィールド名が挿入されている(必)' (First row of range contains field names (required)) and an unchecked box for 'SAS 命名規則に準拠するよう列の名称を変更する(可)' (Adjust column names to conform to SAS naming conventions (optional)). A red circle '2' is around the first checked box. At the bottom, there is a '次へ(右)' (Next (right)) button highlighted with a red circle '3'. Other buttons include '<戻る(左)' (Back (left)), '完了(D)' (Finish), 'キャンセル' (Cancel), and 'ヘルプ' (Help).

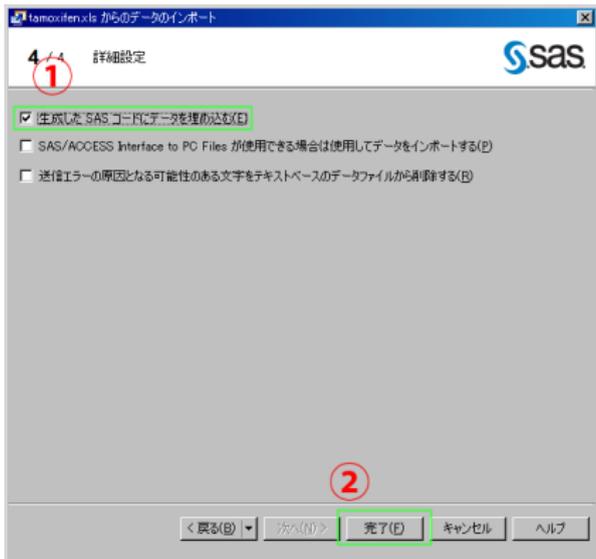
変数の確認と型変換

- 読み込まれた変数名の確認と、必要であれば文字型・数値型等を変更できる
- [次へ] をクリック



保存オプションと読み込み操作の完了

- [生成した SAS コードにデータを埋め込む] にチェック
 - 埋め込んでおくと再解析しやすい
- 完了をクリック



読み込み結果

- 正常に読み込みが完了すると、以下のような画面が表示されます
- これで、Excel データの読み込み作業が終了しました

	リンパ節転移	タモキシフェン使用の有無	乳がんの再発
1	あり	あり	あり
2	あり	あり	あり
3	あり	あり	あり
4	あり	あり	あり
5	あり	あり	あり
6	あり	あり	あり
7	あり	あり	あり
8	あり	あり	あり
9	あり	あり	あり
10	あり	あり	あり
11	あり	あり	あり
12	あり	あり	あり
13	あり	あり	あり
14	あり	あり	あり
15	あり	あり	あり
16	あり	あり	あり
17	あり	あり	あり

メイン画面に戻る

- データ (ファイル名.xls) のインポートと表示されている部分をクリック → [プロセスフロー]

SAS Enterprise Guide

ファイル(F) 編集(E) 表示(V) タスク(K) プログラム(P) ツール(T) ヘルプ(H) | 保存(S) 印刷(P) ヘルプ(H) | 終了(X)

プロジェクトツリー

- プロセスフロー
 - tamoxifen.xls
 - データ (tamoxifen.xls[data]) のインポート

データ (tamoxifen.xls[data]) のインポート -

Reg プロセスフロー

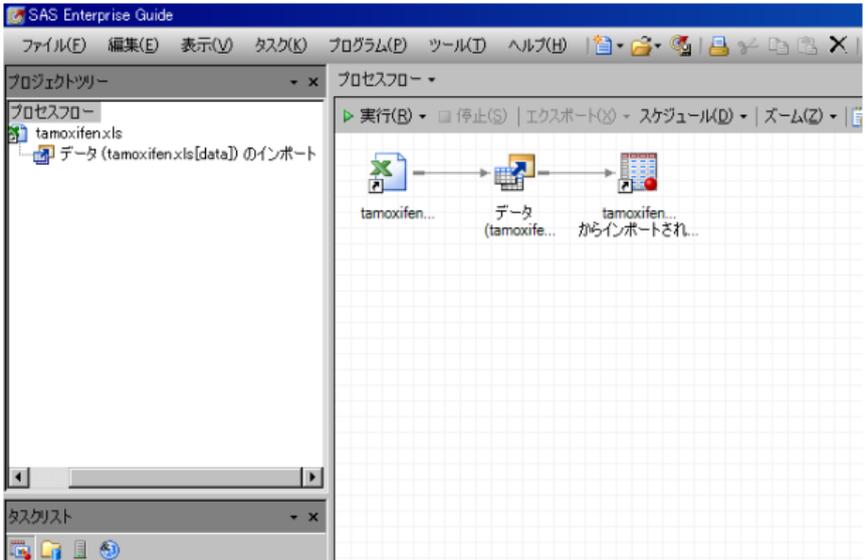
タスクの変更(Y) フィルタと並べ替え(L) クエリビルド(Q) データ(D)

	リンパ節転移	タモキシフェン使用の有無	乳がんの再発
1	あり	あり	あり
2	あり	あり	あり
3	あり	あり	あり
4	あり	あり	あり
5	あり	あり	あり
6	あり	あり	あり
7	あり	あり	あり
8	あり	あり	あり
9	あり	あり	あり
10	あり	あり	あり
11	あり	あり	あり
12	あり	あり	あり
13	あり	あり	あり
14	あり	あり	あり
15	あり	あり	あり
16	あり	あり	あり
17	あり	あり	あり

タスクリスト

メイン画面で結果を確認

- Excel ファイルのデータを読み込んで変換し、
- SAS データセット (SAS で扱うことのできるデータ形式) に変換されているか確認できます
- 下図一番右のアイコンが表示されていれば OK



先ほど読み込んだ分割表データについて

- 佐藤俊哉. 交絡 事実と反事実の比較. 岩波科学 2008 年 4 月号. より引用

1982 年から 1990 年にかけて日本の 9 施設で手術を受けた乳がん患者 4901 名について, 手術後にタモキシフェンという女性ホルモンによく似た乳がんの治療薬を使用したかどうかと, その後の乳がんの再発との関係を 1996 年まで追跡して調べた結果の一部である

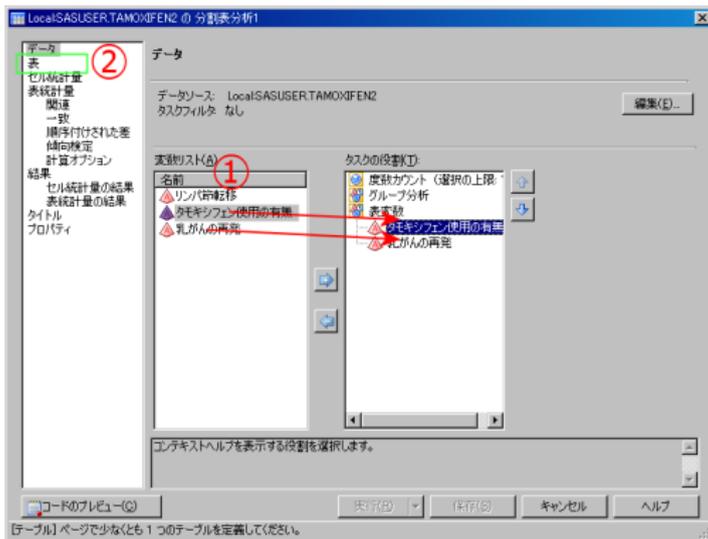
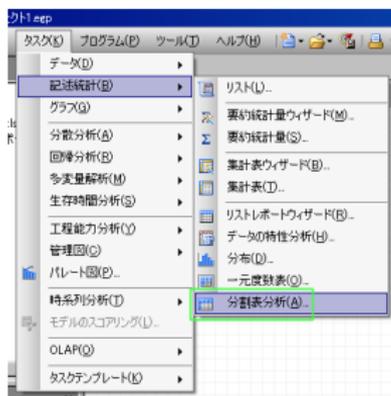
- 変数: タモキシフェン使用の有無, 乳がん再発の有無, リンパ節転移の有無

単変量解析

- 目的
タモキシフェン使用の有無によって、乳がんの再発割合に違いがあるかどうかを検討する事
- 目的に対応する解析内容
二元分割表の集計と、再発割合に違いがあるかどうかの仮説検定、再発割合の差の信頼区間によって評価を行う

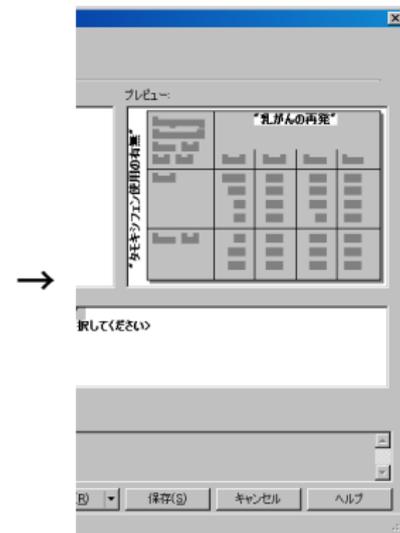
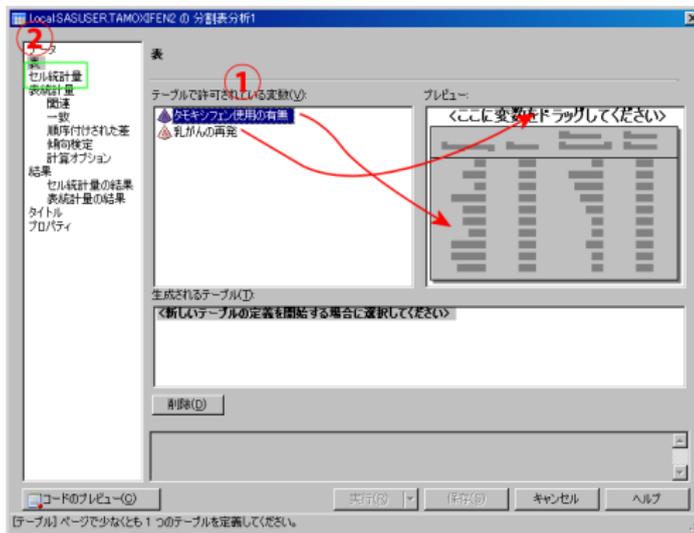
二元分割表の作成 (1)

- [タスク] → [記述統計] → [分割表分析]
- 変数 [タモキシフェン使用の有無] [乳がんの再発] を表変数に設定 (ドラッグ& ドロップ) → 表の設定



二元分割表の作成 (2)

- 変数 [タモキシフェン使用の有無] を表の行に, [乳がんの再発] を表の列に設定します
- 右図のようになっていれば OK です



二元分割表

- [行のパーセント], [セルの度数] にチェック → [実行]
- 二元分割表の結果が表示されます



分割表分析

結果

FREQ フロジヤ

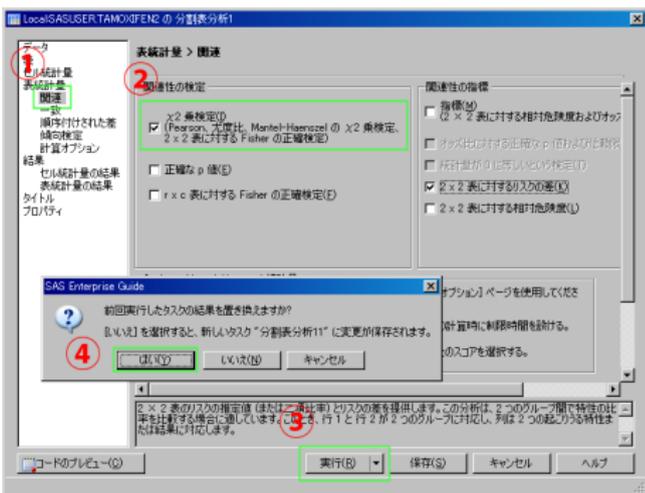
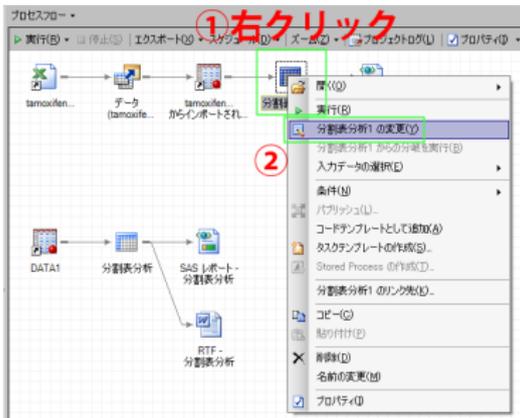
		乳がんの再発		合計
		あり	なし	
タモキシフェン使用の有無	あり	度数 464	2085	2549
	なし	度数 424	1928	2352
合計		度数 888	4013	4901

- 再発割合

タモキシフェン使用 $464/2549=18.2\%$, 非使用 $424/2352=18.0\%$

仮説検定

- プロセスフローに戻る → [分割表分析 1] を右クリック → [分割表分析 1 の変更]
- [表統計量] → [関連] → [関連性の検定] → [χ^2 検定] → [実行] → 結果の置き換え [はい]



仮説検定の結果

- 再発確率が異なるかどうかを、ピアソンのカイ二乗検定を用いて検定する ($\alpha = 0.05$ とする)

$$\begin{cases} H_0: & \text{タモキシフェン使用の有無で再発確率は変わらない} \\ H_1: & \text{タモキシフェン使用の有無で再発確率が異なる} \end{cases}$$

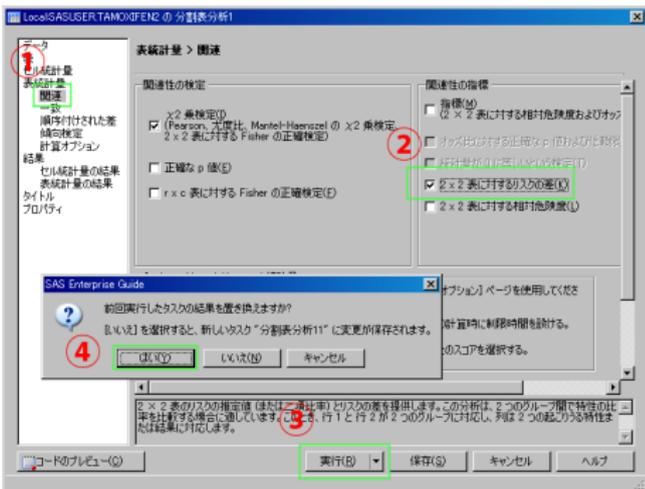
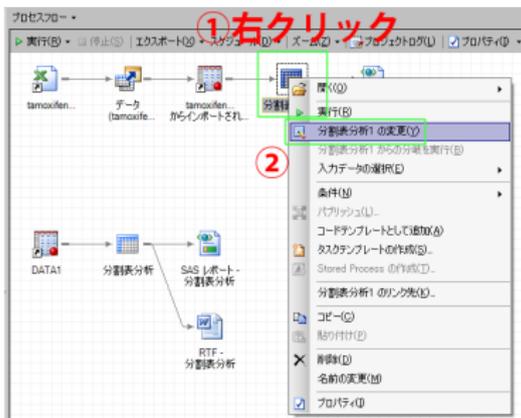
統計量	自由度	値	p 値
カイ 2 乗値	1	0.0255	0.8730
尤度比カイ 2 乗値	1	0.0255	0.8730
連続性補正カイ 2 乗値	1	0.0151	0.9023
Mantel-Haenszel のカイ 2 乗値	1	0.0255	0.8730
ファイ係数		0.0023	
一致係数		0.0023	
Cramer の V 統計量		0.0023	

Fisher の正確検定	
セル (1,1) 度数 (F)	464
左側 Pr <= F	0.5779
右側 Pr >= F	0.4513
表の確率 (P)	0.0292
両側 Pr <= P	0.8820

- P -value=0.873 であり, $\alpha = 0.05$ のもとで帰無仮説は棄却されない

リスク差の推定

- プロセスフローに戻る → [分割表分析 1] を右クリック → [分割表分析 1 の変更]
- [表統計量] → [関連] → [関連性の指標] → [2 × 2 表に対するリスクの差] → [実行] → 結果の置き換え [はい]



リスク差の推定結果

	リスク	漸近標準誤差	(漸近) 95% 信頼区間		(直接確率) 95% 信頼限界	
行1	0.1820	0.0076	0.1671	0.1970	0.1672	0.1976
行2	0.1803	0.0079	0.1647	0.1958	0.1649	0.1964
合計	0.1812	0.0055	0.1704	0.1920	0.1705	0.1923
差	0.0018	0.0110	-0.0198	0.0233		
行1 - 行2の差						

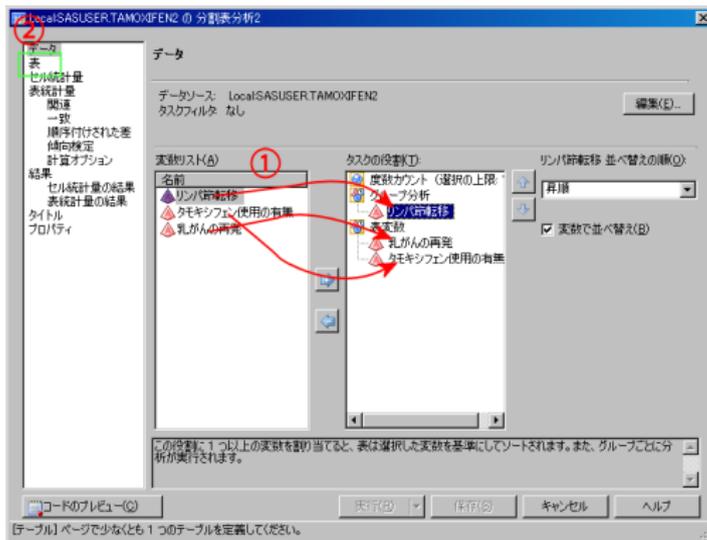
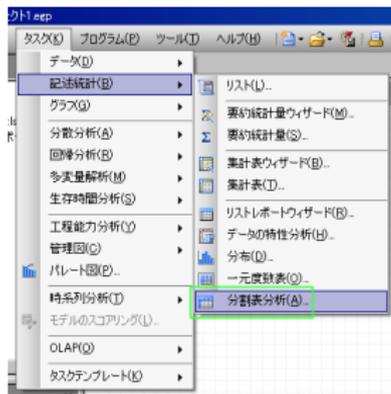
- リスク差=0.0018 であり, 95%信頼区間も 0 をまたぐため, 再発確率の差はそれほど大きくない

多変量解析

- 目的
リンパ節転移の有無を考慮した上で、タモキシフェン使用の有無によって、乳がんの再発確率に違いがあるかどうかを検討する事
- 目的に対応する解析内容
多元分割表の集計と、ロジスティック回帰モデルを用いた調整オッズ比の信頼区間を求めて評価を行う

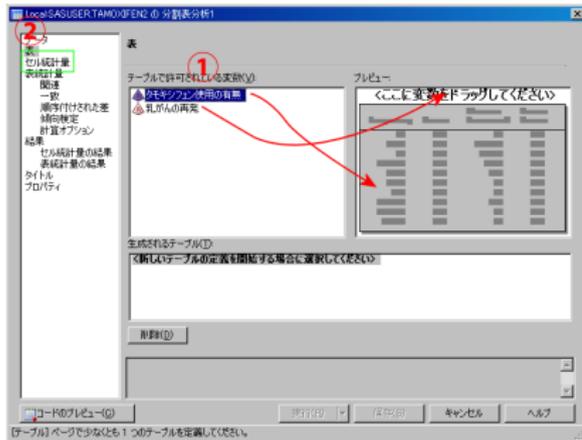
多元分割表の作成 (1)

- [タスク] → [記述統計] → [分割表分析]
- 変数 [タモキシフェン使用の有無] [乳がんの再発] を表変数に, [リンパ節転移] をグループ変数に設定 (ドラッグ&ドロップ) → [表] の設定



多元分割表の作成 (2)

- 変数 [タモキシフェン使用の有無] を表の行に, [乳がんの再発] を表の列に設定します
- [行のパーセント], [セルの度数] にチェック → [実行]



多元分割表

リンパ節転移'n=あり

		表：タモキシフェン使用の有無 * 乳がんの再発			
		乳がんの再発		合計	
		あり	なし		
タモキシフェン使用の有無	あり	度数	368	847	1215
		行のパーセント	30.29	69.71	
	なし	度数	253	507	760
		行のパーセント	33.29	66.71	
合計		度数	621	1354	1975

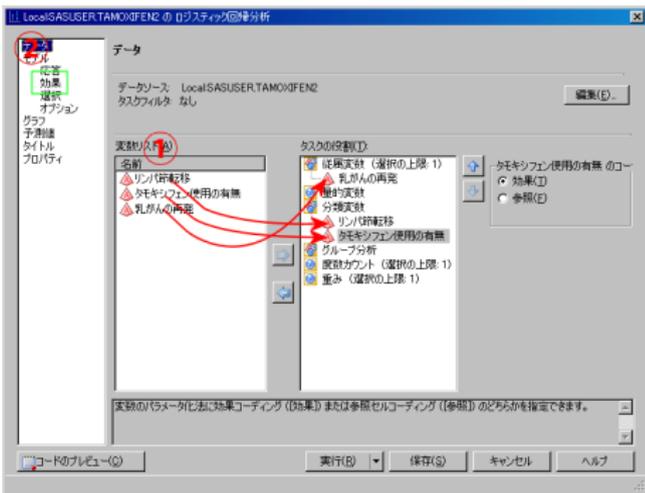
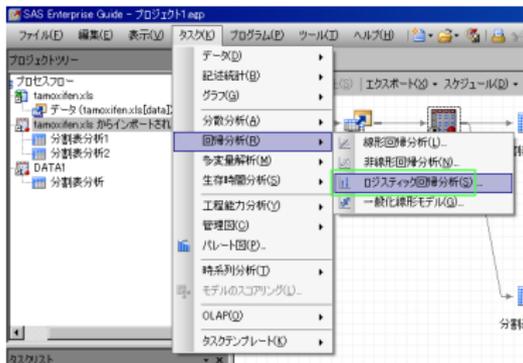
リンパ節転移'n=なし

		表：タモキシフェン使用の有無 * 乳がんの再発			
		乳がんの再発		合計	
		あり	なし		
タモキシフェン使用の有無	あり	度数	96	1238	1334
		行のパーセント	7.20	92.80	
	なし	度数	171	1421	1592
		行のパーセント	10.74	89.26	
合計		度数	267	2659	2926

- リンパ節転移ありの再発割合
タモキシフェン使用 $368/1215=30.3\%$, 非使用 $253/760=33.3\%$
- リンパ節転移なしの再発割合
タモキシフェン使用 $96/1334=7.2\%$, 非使用 $171/1592=10.7\%$

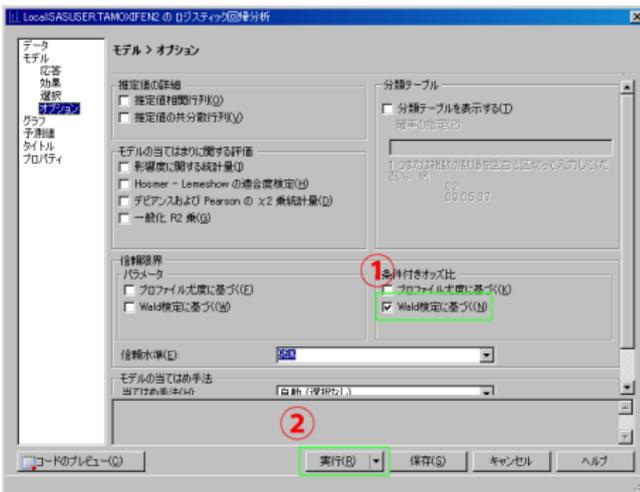
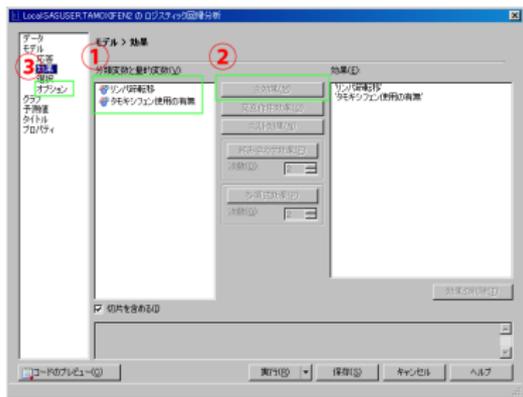
ロジスティック回帰モデルによる解析 (1)

- [タスク] → [回帰分析] → [ロジスティック回帰分析]
- 変数 [乳がんの再発] を従属変数に, [タモキシフェン使用の有無] [リンパ節転移] を分類変数に設定 (ドラッグ& ドロップ) → [モデル] → [効果] の設定



ロジスティック回帰モデルによる解析 (2)

- 変数 [タモキシフェン使用の有無] [リンパ節転移] を選択し, [主効果] をクリックして効果を設定
- [モデル] → [オプション] の設定 → [条件付きオッズ比] → [Wald 検定に基づく] にチェック → [実行]



ロジスティック回帰モデルの解析結果 (1)

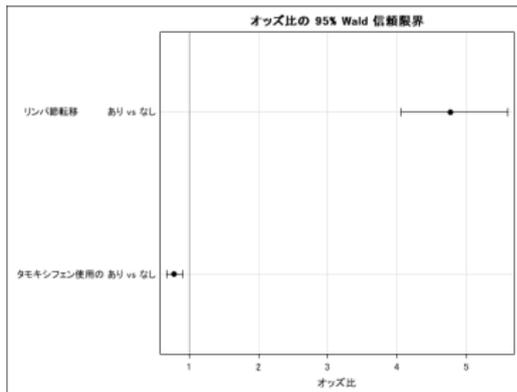
- ロジスティックモデル

$$\text{logit}(p_{\text{乳がんの再発}}) = \beta_{\text{切片}} + \beta_{\text{リンパ節転移の有無}} x_1 + \beta_{\text{タモキシフェン使用の有無}} x_2$$

- 上記のモデルを用いれば、リンパ節転移の有無で調整した、タモキシフェン使用の有無のオッズ比を求める事ができる

最尤推定値の分析						
パラメータ		自由度	推定値	標準誤差	Wald カイ2乗	Pr > ChiSq
Intercept		1	-2.1920	0.0718	932.4667	< .0001
リンパ節転移	あり	1	1.5617	0.0819	363.9424	< .0001
タモキシフェン使用の	あり	1	-0.2468	0.0789	9.7956	0.0017

オッズ比の推定値					
効果		点推定値	95% Wald 信頼限界		
リンパ節転移	あり vs なし	4.767	4.060	5.597	
タモキシフェン使用の	あり vs なし	0.781	0.669	0.912	



ロジスティック回帰モデルの解析結果 (2)

- タモキシフェン使用の効果についての推定値は -0.25 , $P\text{-value}=0.0017$ であり, $\alpha = 0.05$ のもとで帰無仮説は棄却された
- タモキシフェン使用の効果についての調整オッズ比は 0.781 , 95%信頼区間は $[0.669, 0.912]$ であった
- したがって, リンパ節転移を考慮するとタモキシフェン使用の有無で再発オッズが異なる

単変量・多変量解析の結果のまとめ

表 1. タモキシフェン使用と乳がんの再発

タモキシフェン	再発あり	再発なし	対象者数
使用	464 (18.2%)	2085	2549
非使用	424 (18.0%)	1928	2352
合計	888	4013	4901
再発割合の差 = 0.2% [-0.02, 0.02]			

表 2. リンパ節転移で層別したタモキシフェン使用と乳がんの再発

タモキシフェン	リンパ節転移あり			リンパ節転移なし		
	再発あり	再発なし	対象者数	再発あり	再発なし	対象者数
使用	368 (30.3%)	847	1215	96 (7.2%)	1238	1334
非使用	253 (33.3%)	507	760	171 (10.7%)	1421	1592
合計	621	1354	1975	267	2659	2926
再発割合の差 = -3%				再発割合の差 = -3.5%		

- 調整オッズ比=0.781 [0.669, 0.912]

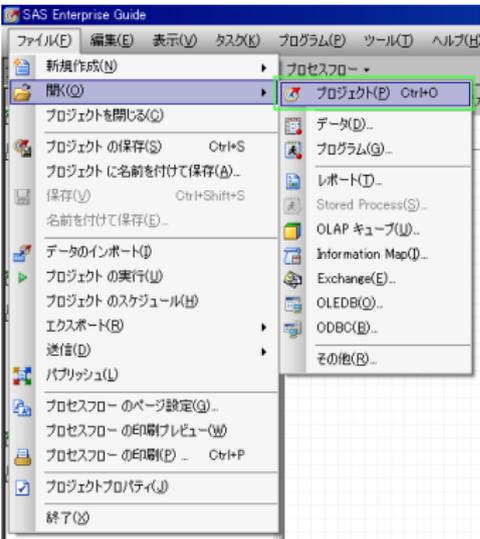
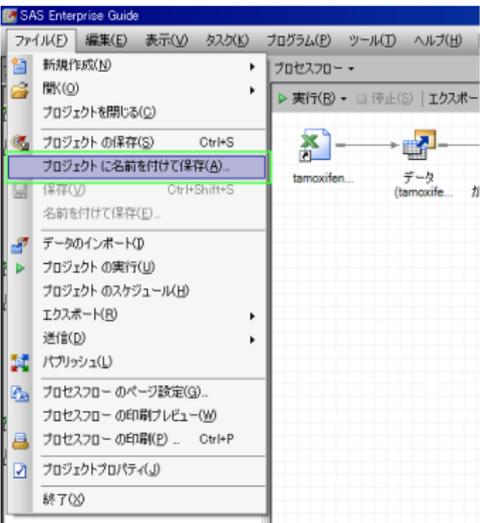
交絡因子

- 表 2 を横方向にみると...
 - タモキシフェン使用群, 1215:1334 でリンパ節転移あり (高リスク集団) が多い
 - タモキシフェン非使用群, 760:1592 でリンパ節転移あり (高リスク集団) が少ない
 - タモキシフェンに効果があるにも関わらず, 高リスクの人が多いので, 見かけ上効果が薄まって見えていた
- リンパ節転移の様な因子は交絡因子であり, 交絡因子を調整しない推定結果は正しくない
- 多変量解析を用いて, 交絡因子の状態を仮想的に同じとした場合の, タモキシフェン使用の再発予防効果をみるべき

タモキシフェンの乳がん再発予防効果はある

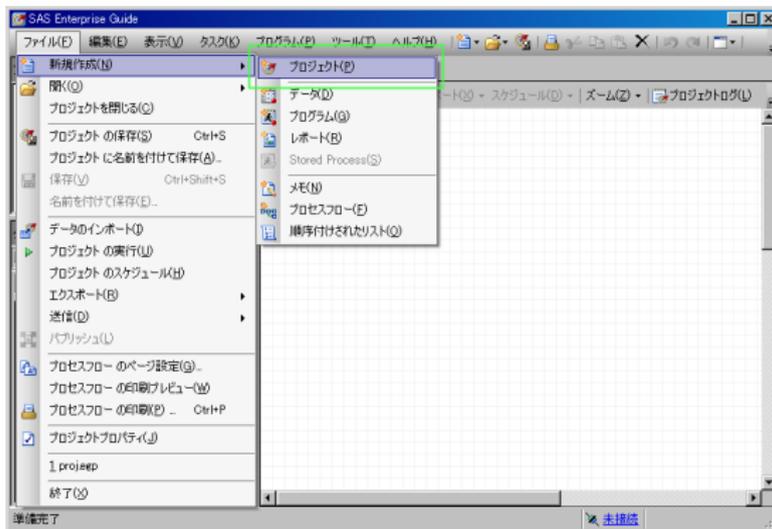
プロジェクトの保存 / プロジェクトを開く

- [ファイル] → [プロジェクトに名前を付けて保存] で今まで行った全ての解析をプロジェクトに保存します
- [ファイル] → [開く] → [プロジェクト] で保存しておいた解析を開くことができます



新規プロジェクトの追加

- [ファイル] → [新規作成] → [プロジェクト]

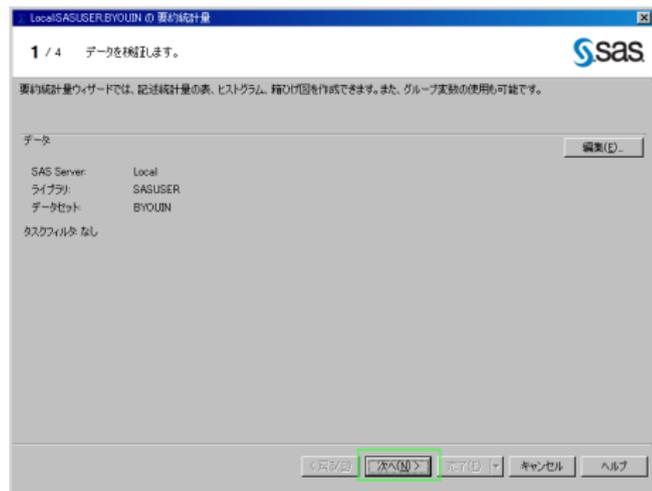
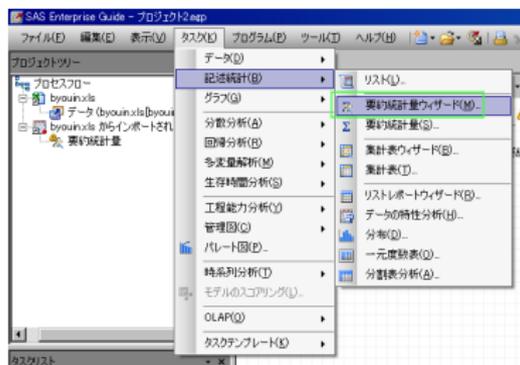


連続データの概要

- 2007 年の都道府県別の人口 10 万対病院数・歯科診療所数・一般診療所数^[4] と人口推計^[5]
- 高齢者の割合が多い県と少ない県で分けた [高齢者] という変数が含まれる
- 目的
高齢者の割合によって、人口 10 万対病院数・歯科診療所数・一般診療所数は異なるかどうかを検討する
- 目的に対応する解析内容
要約統計量, ヒストグラムや箱ひげ図の作成と, 2 群の母平均に対する仮説検定と信頼区間の計算
- 新しく作成したプロジェクトで, データ [byouin.xls] を読み込みます
- スライド 13 のデータ読み込みの手順を参照

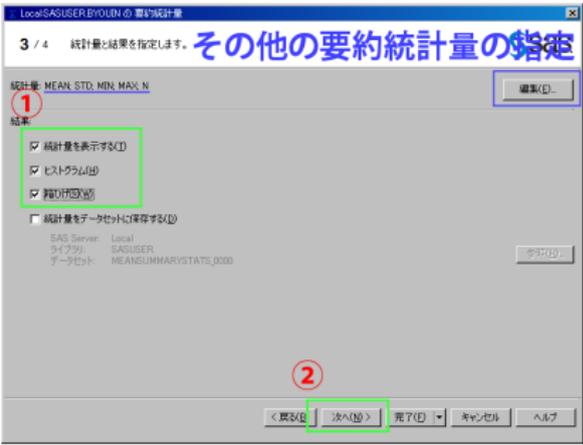
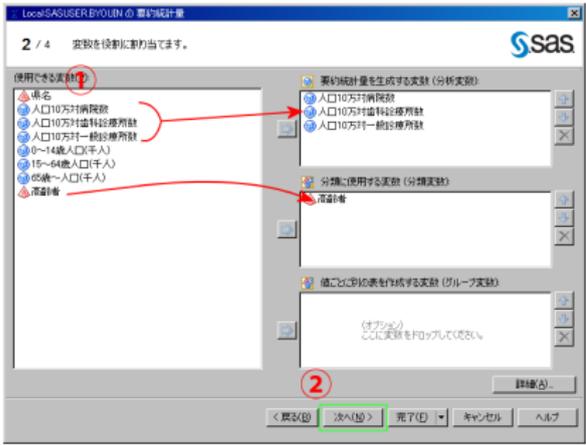
要約統計量の計算とヒストグラム・箱ひげ図の出力 (1)

- [タスク] → [記述統計] → [要約統計量ウィザード]
- [次へ]



要約統計量の計算とヒストグラム・箱ひげ図の出力 (2)

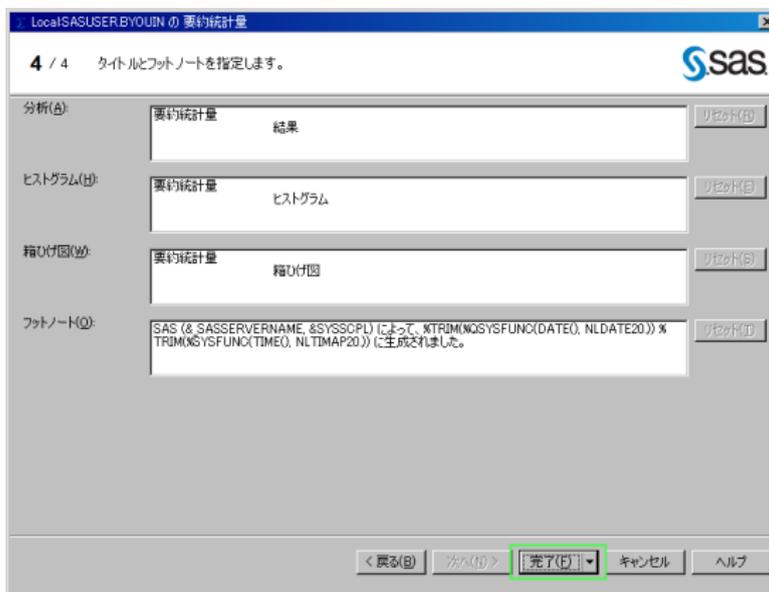
- 変数 [人口 10 万対病院・歯科・一般診療所数] を [分析変数] に, 変数 [高齢者] を [分類変数] にドラッグ&ドロップ → [次へ]
- [統計量を表示する], [ヒストグラム], [箱ひげ図] にチェック → [次へ]



- 青線部分に示された統計量以外を出力したい場合, 右上から設定できる

要約統計量の計算とヒストグラム・箱ひげ図の出力 (3)

- [完了]

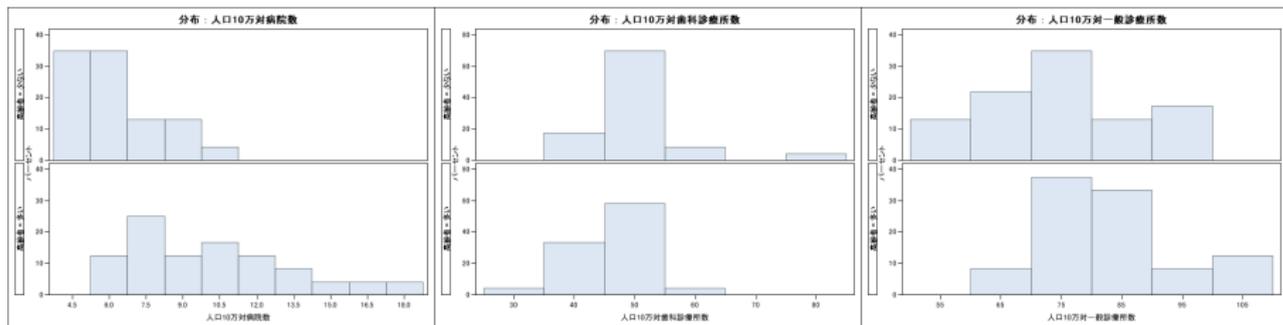


要約統計量

高齢者 オブザベーション数 変数			平均	標準偏差	最小値	最大値	N
少ない	23	人口10万対病院数	6.3130435	1.7535171	3.9000000	10.8000000	23
		人口10万対歯科診療所数	50.3956522	8.7888089	39.0000000	82.7000000	23
		人口10万対一般診療所数	74.8826087	12.9442740	55.4000000	99.1000000	23
多い	24	人口10万対病院数	10.0958333	3.2468686	5.7000000	17.6000000	24
		人口10万対歯科診療所数	45.6250000	4.8172922	33.5000000	55.1000000	24
		人口10万対一般診療所数	81.6666667	11.2479435	68.3000000	106.4000000	24

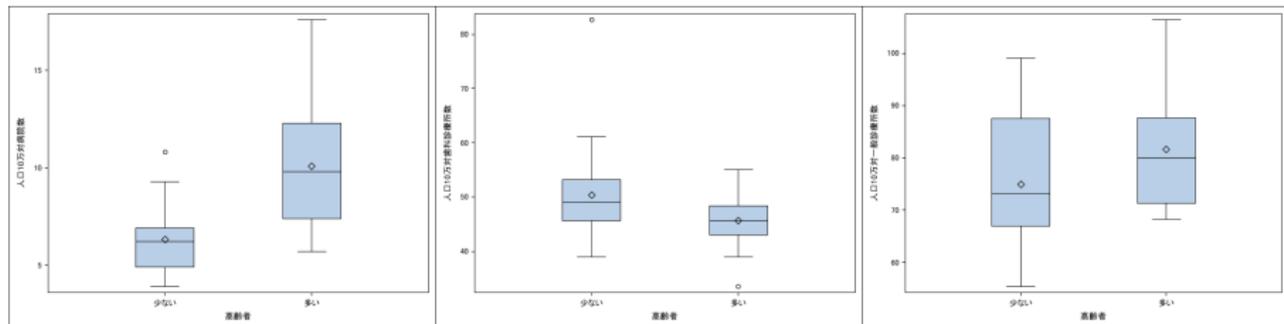
- 人口 10 万対病院数
 高齢者が少ない: 平均 6.3 施設, 標準偏差 1.75 施設
 高齢者が多い: 平均 10.1 施設, 標準偏差 3.24 施設
- 人口 10 万対歯科診療所数
 高齢者が少ない: 平均 50.4 施設, 標準偏差 8.79 施設
 高齢者が多い: 平均 45.6 施設, 標準偏差 4.81 施設
- 人口 10 万対一般診療所数
 高齢者が少ない: 平均 74.9 施設, 標準偏差 12.9 施設
 高齢者が多い: 平均 81.7 施設, 標準偏差 11.2 施設

ヒストグラム



- 人口 10 万対病院数
ピーク位置とばらつきが異なりそう → 対数変換して処理
- 人口 10 万対歯科診療所数
似た形状, 高齢者が少ない群に外れた値がある
- 人口 10 万対一般診療所数
似た形状, ピーク位置が異なりそう

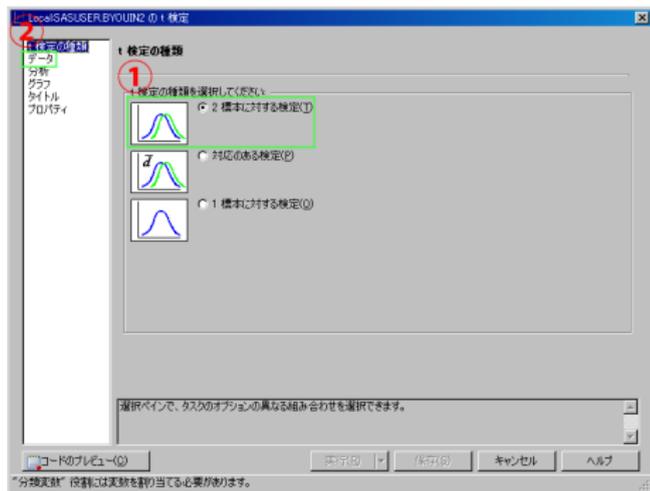
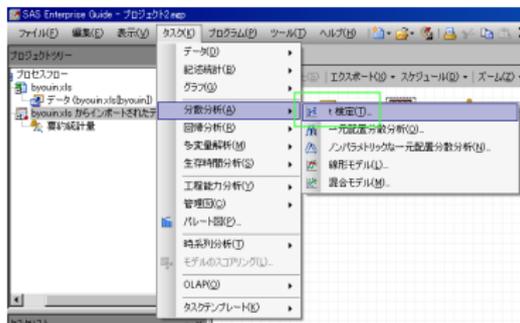
箱ひげ図



- 人口 10 万対病院数
ピーク位置とばらつきが異なりそう → 対数変換して処理
- 人口 10 万対歯科診療所数
似た形状, 高齢者が少ない群に外れた値がある
- 人口 10 万対一般診療所数
似た形状, ピーク位置が異なりそう

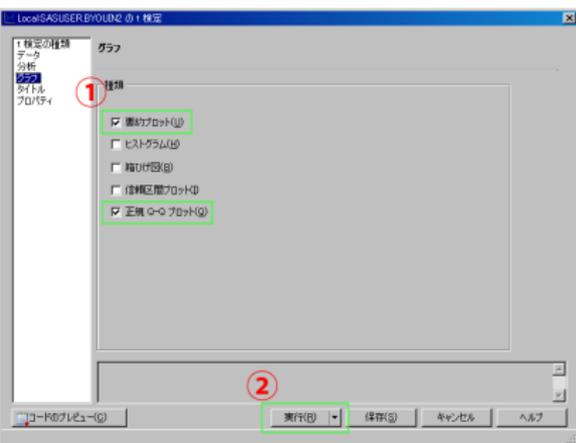
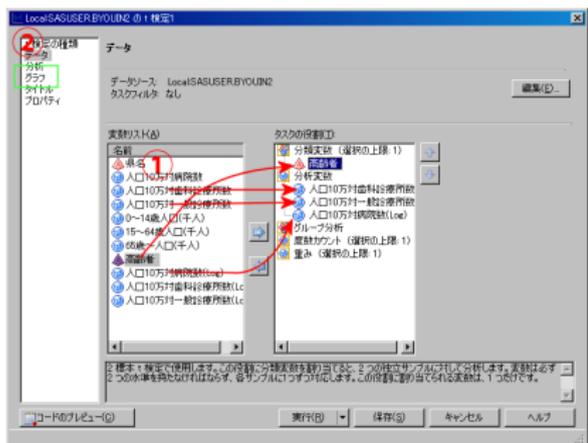
二標本 t 検定の実行 (1)

- [タスク] → [分散分析] → [t 検定]
- [2 標本に対する t 検定] → [データ] 設定画面へ



二標本 t 検定の実行 (2)

- 変数 [人口 10 万対歯科療所数・一般診療所数, 病院数 (Log)] を [分析変数] へ, 変数 [高齢者] を [分類変数] へドラッグ& ドロップ → [グラフ] の設定画面へ
- [要約プロット] および [正規 Q-Q プロット] にチェック → [実行]



二標本 t 検定の結果 (人口 10 万対病院数 (Log))

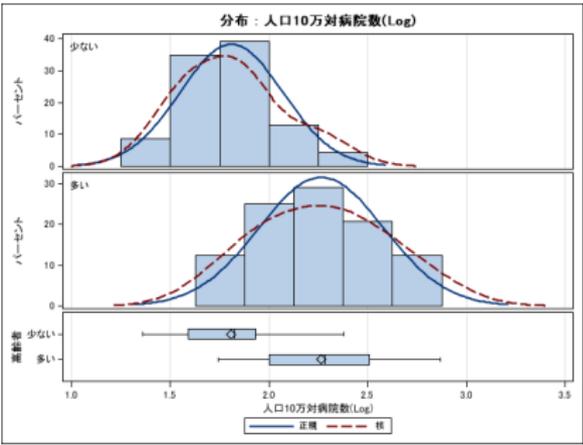
H_0 : 高齢者の割合によって、人口 10 万対病院数 (Log) に違いはない
 H_1 : 高齢者の割合によって、人口 10 万対病院数 (Log) は異なる

高齢者	N	平均	標準偏差	標準誤差	最小値	最大値
少ない	23	1.8089	0.2608	0.0544	1.3610	2.3795
多い	24	2.2640	0.3161	0.0645	1.7405	2.8679
Diff (1-2)		-0.4551	0.2904	0.0847		

高齢者	手法	平均	平均の		標準偏差の	
			95% 信頼限界	標準偏差	95% 信頼限界	
少ない		1.8089	1.6961	1.9217	0.2608	0.2017 0.3692
多い		2.2640	2.1305	2.3975	0.3161	0.2456 0.4434
Diff (1-2)	Pooled	-0.4551	-0.6258	-0.2845	0.2904	0.2409 0.3657
Diff (1-2)	Satterthwaite	-0.4551	-0.6252	-0.2851		

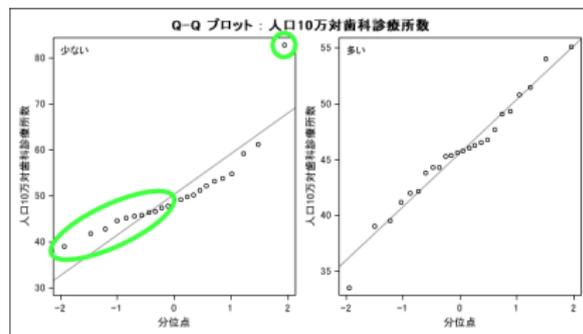
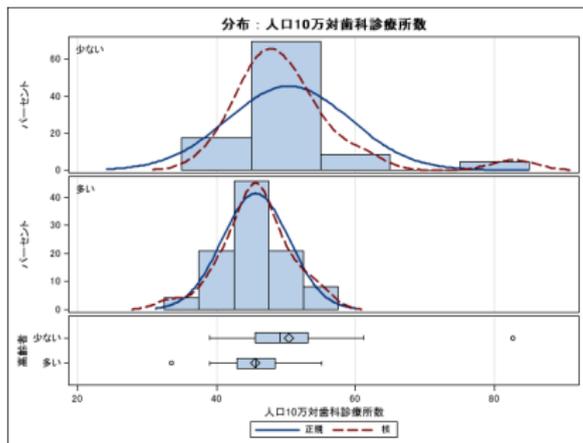
手法	分散	自由度	t 値	Pr > t
Pooled	Equal	45	-5.37	<.0001
Satterthwaite	Unequal	44.05	-5.39	<.0001

等分散性				
手法	分子の自由度	分母の自由度	F 値	Pr > F
Folded F	23	22	1.47	0.3717



- P -value < 0.0001 であり、帰無仮説は棄却される
 したがって、高齢者の割合によって、人口 10 万対病院数 (Log) は異なる
 と考えられる

二標本 t 検定の結果 (人口 10 万対歯科診療所数)



- Q-Q プロットから、はずれ値の影響で正規性の仮定が満たされない可能性が示唆される
よってノンパラメトリックな方法を用いて解析しなおすことにする

二標本 t 検定の結果 (人口 10 万対一般診療所数)

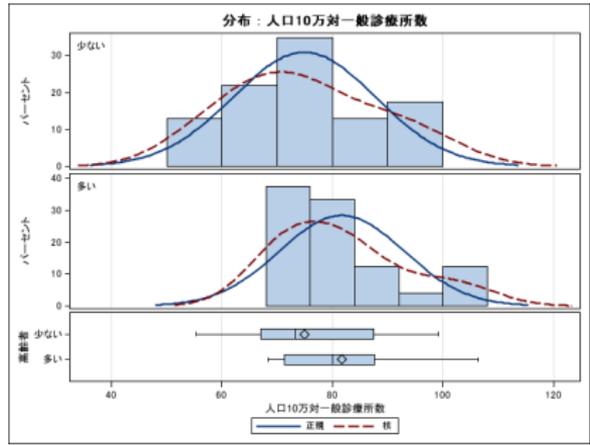
H_0 : 高齢者の割合によって、人口 10 万対一般診療所数に違いはない
 H_1 : 高齢者の割合によって、人口 10 万対一般診療所数は異なる

高齢者	N	平均	標準偏差	標準誤差	最小値	最大値
少ない	23	74.8826	12.9443	2.6991	55.4000	99.1000
多い	24	81.6667	11.2479	2.2960	68.3000	106.4
Diff (1-2)		-6.7841	12.1070	3.5328		

高齢者	手法	平均	95% 信頼限界	標準偏差	95% 信頼限界
少ない		74.8826	69.2851	80.4801	12.9443
多い		81.6667	76.9171	86.4163	11.2479
Diff (1-2)	Pooled	-6.7841	-13.8994	0.3313	12.1070
Diff (1-2)	Satterthwaite	-6.7841	-13.9276	0.3595	10.0420

手法	分散	自由度	t 値	Pr > t
Pooled	Equal	45	-1.92	0.0612
Satterthwaite	Unequal	43.548	-1.91	0.0621

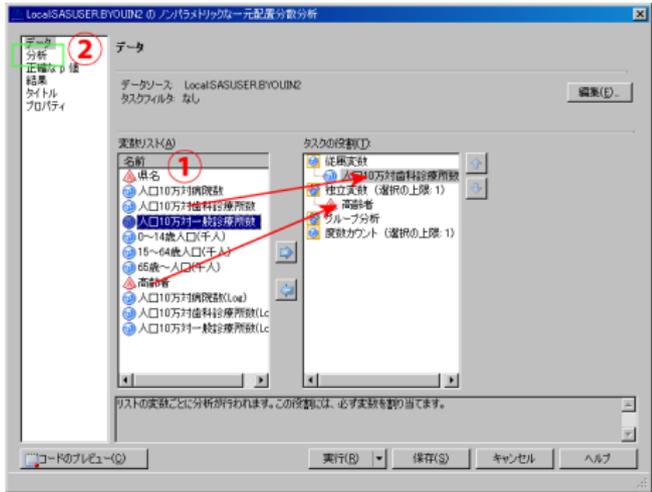
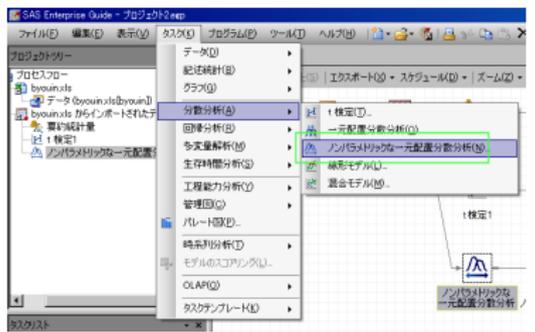
等分散性				
手法	分子の自由度	分母の自由度	F 値	Pr > F
Folded F	22	23	1.32	0.5083



- P -value=0.06 であり、帰無仮説は棄却されない
 したがって、高齢者の割合によって、人口 10 万対一般診療所数が異なるかどうかは分からない

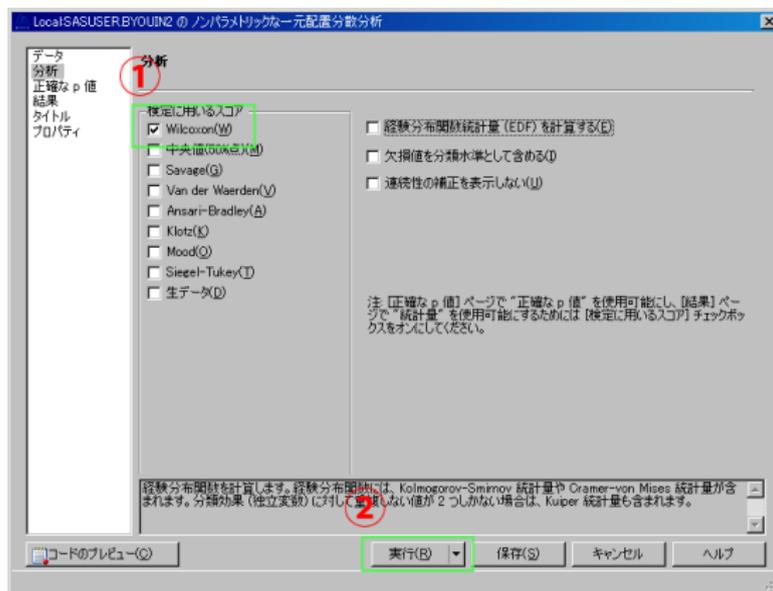
ノンパラメトリック検定の実行 (1)

- [タスク] → [分散分析] → [ノンパラメトリックな一元配置分散分析]
- 変数 [人口 10 万対歯科療所数] を [分析変数] へ, 変数 [高齢者] を [分類変数] へドラッグ& ドロップ → [分析] の設定画面へ



ノンパラメトリック検定の実行 (2)

- [検定に用いるスコア] → [Wilcoxon] のみにチェックが入るように変更 → [実行]



ノンパラメトリック検定の結果

- $$\begin{cases} H_0 : & \text{高齢者の割合によって, 人口 10 万対歯科診療所数に違いはない} \\ H_1 : & \text{高齢者の割合によって, 人口 10 万対歯科診療所数は異なる} \end{cases}$$

統計量	654.0000
正規近似	
Z	2.1604
片側 Pr > Z	0.0154
両側 Pr > Z	0.0307
t 分布で近似	
片側 Pr > Z	0.0180
両側 Pr > Z	0.0360
Z には 0.5 の連続性の補正が含まれています。	

カイ 2 乗	4.7133
自由度	1
Pr > Chi-Square	0.0299

- P -value=0.04 であり, 帰無仮説は棄却される
したがって, 高齢者の割合によって, 人口 10 万対歯科診療所数は異なると考えられる

参考文献

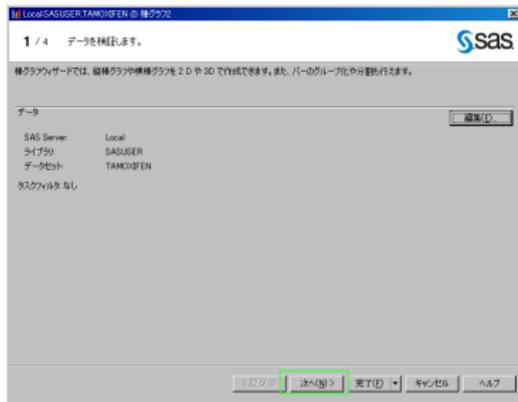
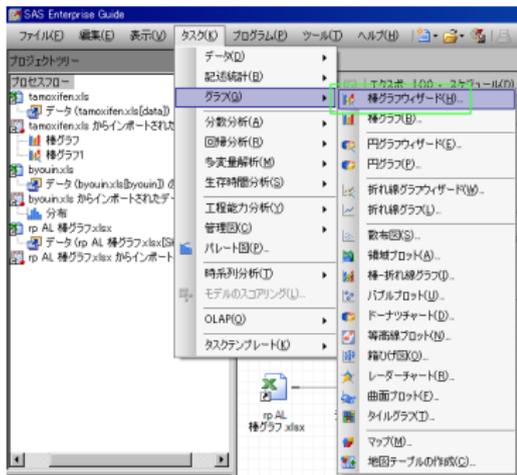
- [1] Der G, Everitt BS. *Basic statistics using SAS®Enterprise Guide®: a primer*. SAS Publishing, 2007.
- [2] 高柳良太 (著), SAS Institute Japan (監修). SAS による統計分析–SAS Enterprise Guide ユーザーズガイド. オーム社, 2008.
- [3] 佐藤俊哉. 交絡 事実と反事実の比較. 岩波科学 2008 年 4 月号.
- [4] 厚生労働省. 平成 19 年医療施設 (動態) 調査. 2007.
<http://www.mhlw.go.jp/toukei/list/79-1.html>
- [5] 総務省統計局. 人口推計 平成 19 年 10 月 1 日現在人口. 2007.
<http://www.stat.go.jp/data/jinsui/2007np/index.htm>

グラフの作成と読み方

- 離散変数 (性別, 重症度など)
 - 一変数: 棒グラフ, 層別: 積み上げ棒グラフ, 三次元棒グラフ
- 連続変数 (年齢, 臨床検査値など)
 - 一変数/層別: ヒストグラム, 箱ひげ図 (スライド 44), 二変数: 散布図
- 生存時間変数
 - 一変数/層別: Kaplan-Meier プロット

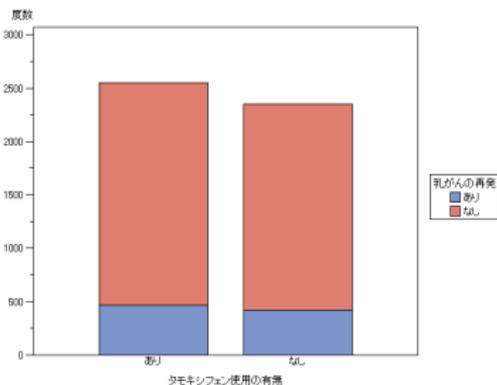
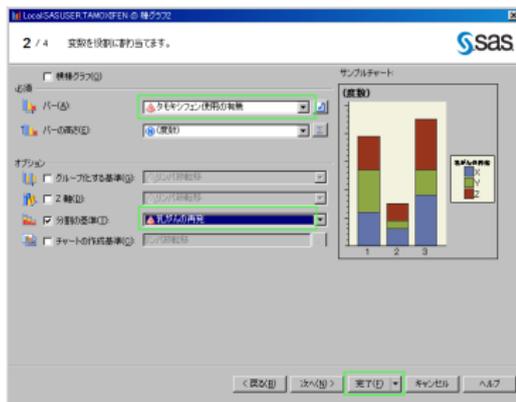
棒グラフ/積み上げ棒グラフ (1)

- [タスク] → [グラフ] → [棒グラフウィザード]
- データを選択し, [次へ]



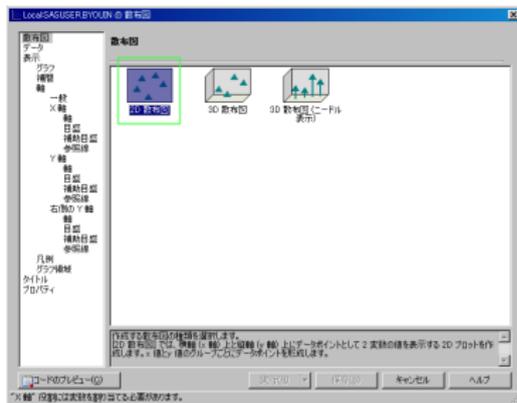
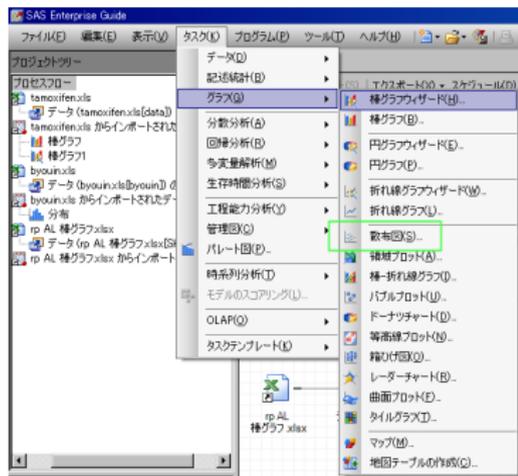
棒グラフ/積み上げ棒グラフ (2)

- [バー]: 横軸の変数
- [分割の基準]: 積み上げ棒グラフ
- [Z軸]: 層別因子が二つ以上の場合、三次元棒グラフを作成する
- [グループ化する基準]: 棒グラフを並べる
- 変数を選択し, [完了]



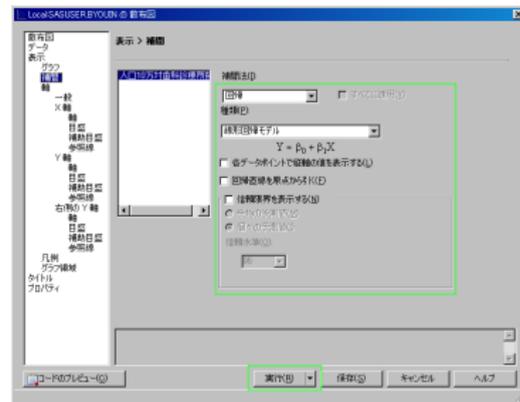
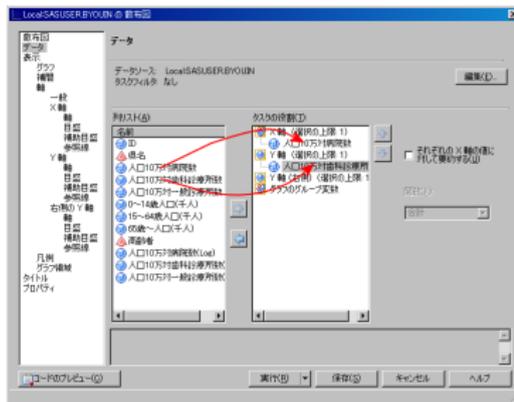
散布図 (1)

- [タスク] → [グラフ] → [散布図]
- [2D 散布図] をダブルクリック



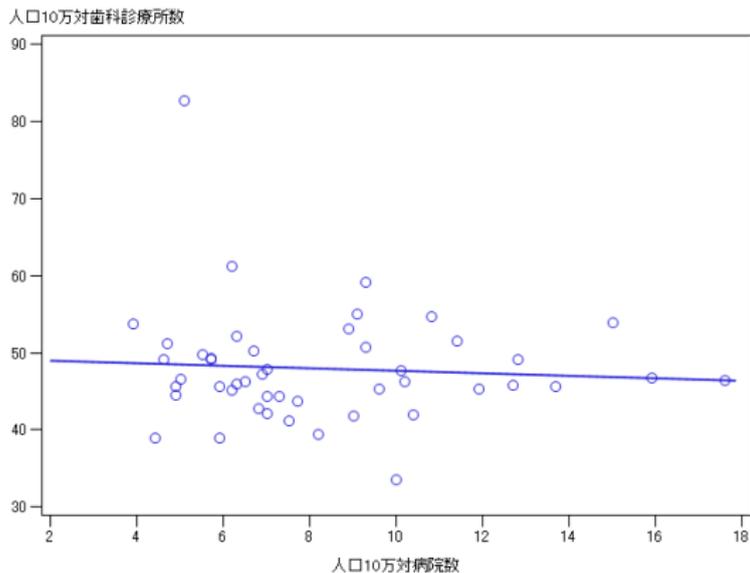
散布図 (2)

- x 軸, y 軸の変数をドラッグアンドドロップして設定する
- [補完] をクリックし, 補完法とオプションを設定し, [実行]



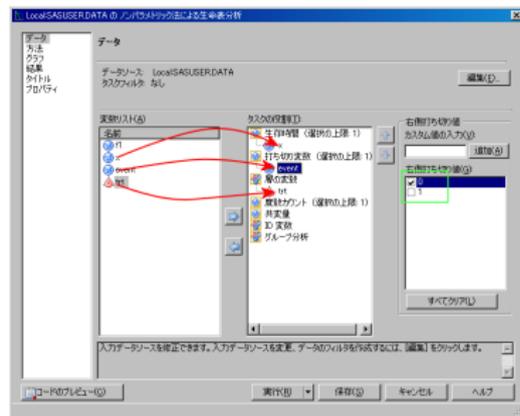
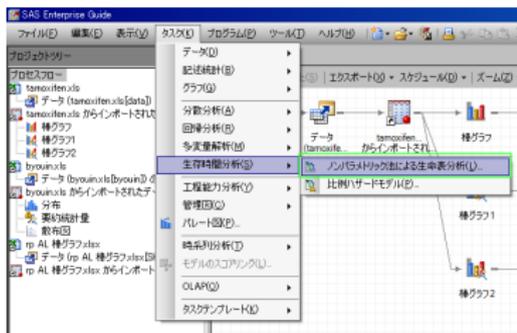
散布図 (3)

- 散布図からは、二変数の関連を読み取る事ができる
- 二変数間に直線的な関連, 二次関数的な関連 (減ってから増えたり), 周期的な関連, または何も関連が無いのかを視覚的に評価する



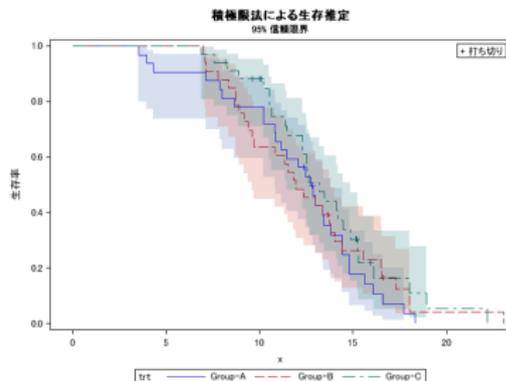
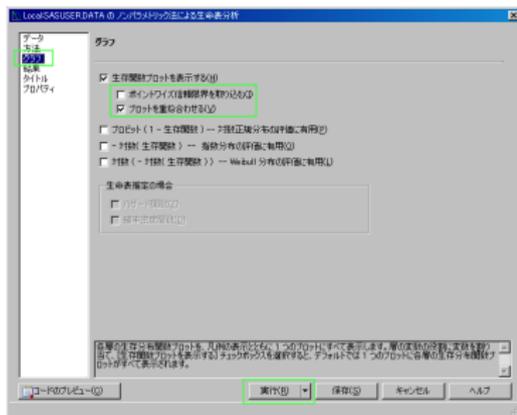
Kaplan-Meier プロット (1)

- [タスク] → [生存時間分析] → [ノンパラメトリック法による生命表分析]
- 生存時間変数, 打ち切り変数と打ち切りを表わす水準, 必要に応じて層別変数をドラッグアンドドロップして設定する



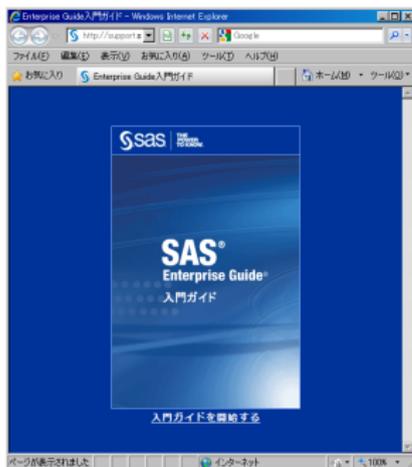
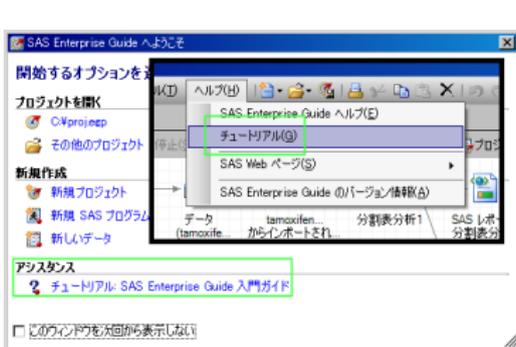
Kaplan-Meier プロット (2)

- [グラフ] をクリックし, [生存関数プロットを表示する] にチェックを入れる, 信頼区間を表示する場合は [ポイントワイズ信頼限界を取り込む] にチェックを入れる
- Log-rank 検定などの結果と Kaplan-Meier プロットが出力される



チュートリアルの表示

- 起動時に表示されるようこそ画面, または [ヘルプ] → [チュートリアル]



- 操作方法を忘れた時に読むと便利です

タスクリスト

- 左下: サーバリストからタスクリストに変更しておくと、解析手法一覧が表示される

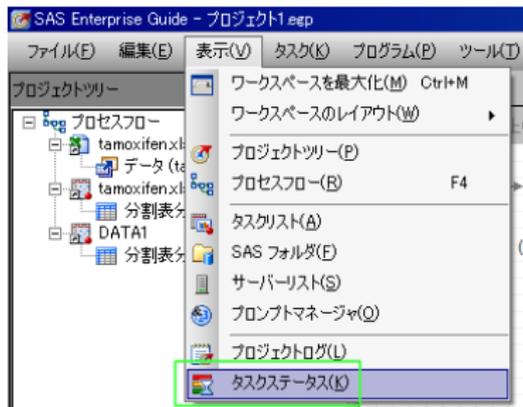
The screenshot displays the SAS Enterprise Guide interface. On the left, the 'タスクリスト' (Task List) pane is active, showing a list of tasks under the 'データ' (Data) category, including 'フィルタと並べ替え', 'クエリビルダ', 'テーブルの追加', 'データの並べ替え', '出力形式の作成', '転置', and '列の分割'. The main workspace shows a workflow diagram with the following steps:

- resampling** (represented by a calendar icon)
- 分布** (represented by a bar chart icon)
- ヒストグラム** (represented by a bar chart icon)
- 線形回帰分析** (represented by a scatter plot icon)
- 無作為抽出** (represented by a document icon)
- ECLIB000の無作為抽出** (represented by a document icon)
- 一元配置分散分析** (represented by a funnel icon)
- SASレポート-分布** (represented by a document icon)
- SASレポート-ヒストグラム** (represented by a document icon)
- SASレポート-線形回帰分析** (represented by a document icon)
- SASレポート-一元配置分散分析** (represented by a document icon)

Arrows indicate the flow from 'resampling' to each of the four analysis tasks, and from each analysis task to its corresponding SAS report. The 'ECLIB000の無作為抽出' task is connected to the '一元配置分散分析' task.

タスクステータスの表示

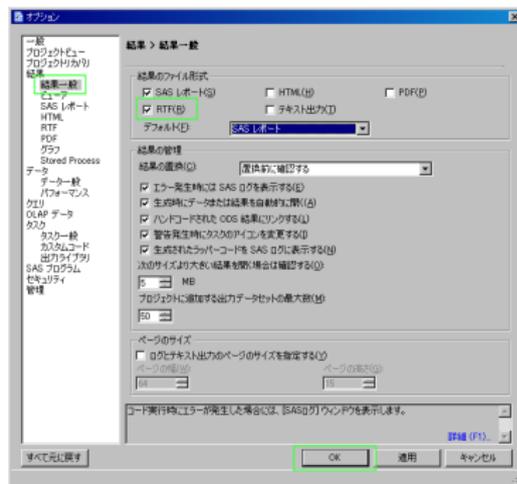
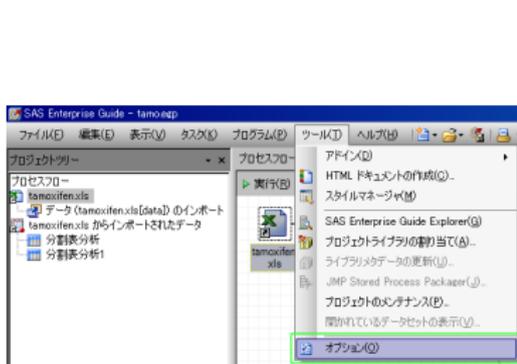
- 実行中・実行予定の処理内容を表示できる



タスクのステータス	
タスク	ステータス
データ (tamoxifen.xls[data]) のインポート	クレンジングした区切り記号付きテキストファイルを作成しています...
分割表分析1	[データ (tamoxifen.xls[data]) のインポート] の待機...
分割表分析	[分割表分析1, データ (tamoxifen.xls[data]) のインポート] の待機...

出力形式の追加

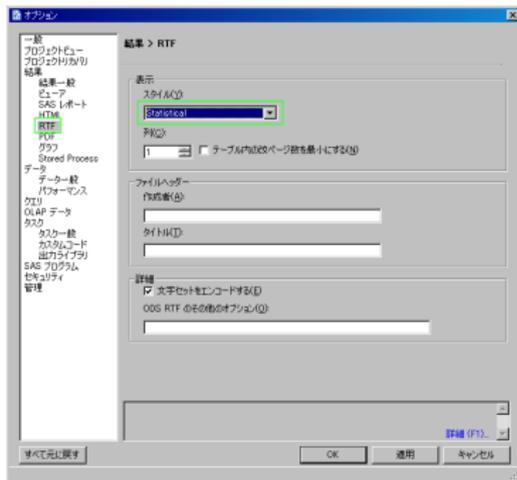
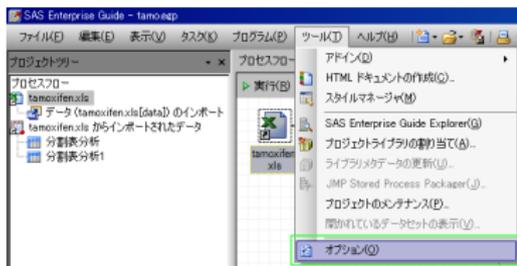
- [ツール] → [オプション]
- [結果一般] → [結果ファイルの形式] から [RTF] を探してチェック



- RTF 形式で出力すると、Word 等にコピー & ペーストして編集しやすい

出力デザインの変更

- [ツール] → [オプション]
- [RTF] → [スタイル]



出力デザインサンプル (一部)

デザインリスト: http://www.josai.ac.jp/~nagasima/contents/sas/odsstyle/ods_style.html

ods rtf style = BarrettsBlue
The FREQ Procedure

Table of x by y			
x(薬剤の投与)	y(治癒)		
Frequency			
Percent			
Row Pct			
Col Pct	あり	なし	Total
あり	63	87	150
	21.00	29.00	50.00
	42.00	58.00	
	60.00	44.62	
なし	42	108	150
	14.00	36.00	50.00
	28.00	72.00	
	40.00	55.38	
Total	105	195	300
	35.00	65.00	100.00

BarrettsBlue

ods rtf style = Listing
The FREQ Procedure

Table of x by y			
x(薬剤の投与)	y(治癒)		
Frequency			
Percent			
Row Pct			
Col Pct	あり	なし	Total
あり	63	87	150
	21.00	29.00	50.00
	42.00	58.00	
	60.00	44.62	
なし	42	108	150
	14.00	36.00	50.00
	28.00	72.00	
	40.00	55.38	
Total	105	195	300
	35.00	65.00	100.00

Listing

ods rtf style = Journal
The FREQ Procedure

Table of x by y

<i>x(薬剤の投与)</i>	<i>y(治癒)</i>		
<i>Frequency</i>			
<i>Percent</i>			
<i>Row Pct</i>			
<i>Col Pct</i>	<i>あり</i>	<i>なし</i>	<i>Total</i>
<i>あり</i>	63	87	150
	21.00	29.00	50.00
	42.00	58.00	
	60.00	44.62	
<i>なし</i>	42	108	150
	14.00	36.00	50.00
	28.00	72.00	
	40.00	55.38	
<i>Total</i>	105	195	300
	35.00	65.00	100.00

Journal

出力デザインのカスタマイズ (1)

- [ツール] → [スタイルマネージャ]
- カスタマイズしたいスタイルを選択 → [コピーの作成]

The image shows two screenshots from the SAS software interface. The main screenshot is the 'Style Manager' dialog box. It has two panes: 'Style List' on the left and 'Listing Preview' on the right. In the 'Style List' pane, a table lists various styles with their positions and UR values. The 'Listing' style is selected and highlighted with a green box. A red circle with the number '2' is placed over the 'Copy' button at the bottom of the dialog. To the left, a smaller screenshot shows the 'Tools' menu with 'Style Manager' selected and highlighted with a green box.

Style List Table:

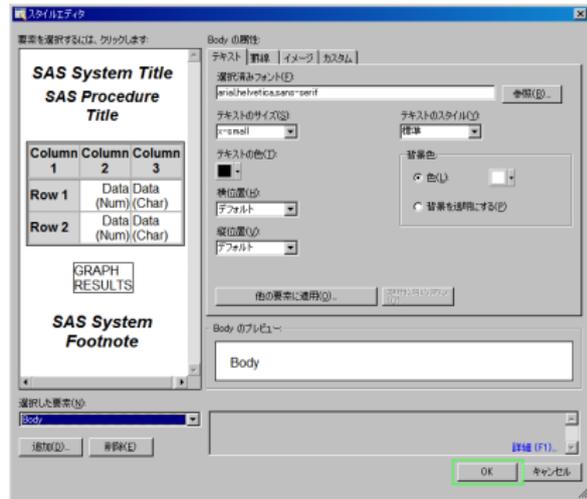
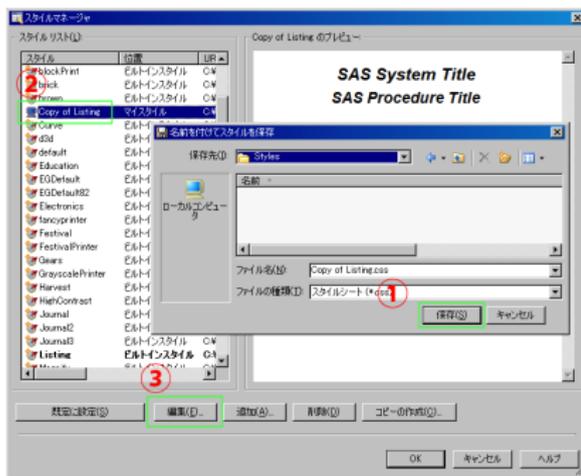
スタイル	位置	UR
Festival	ビルトインスタイル	0%
FestivalPrinter	ビルトインスタイル	0%
Gears	ビルトインスタイル	0%
GrayscalePrinter	ビルトインスタイル	0%
Harvest	ビルトインスタイル	0%
HighContrast	ビルトインスタイル	0%
Journal	ビルトインスタイル	0%
Journal2	ビルトインスタイル	0%
Journal3	ビルトインスタイル	0%
Listing	ビルトインスタイル	0%
Magnify	ビルトインスタイル	0%
Meadow	ビルトインスタイル	0%
MeadowPrinter	ビルトインスタイル	0%
minimal	ビルトインスタイル	0%
Money	ビルトインスタイル	0%
MonochromePrinter	ビルトインスタイル	0%
notfontdefault	ビルトインスタイル	0%
Normal	ビルトインスタイル	0%
NormalPrinter	ビルトインスタイル	0%
Ocean	ビルトインスタイル	0%
Plateau	ビルトインスタイル	0%
printer	ビルトインスタイル	0%
Printer	ビルトインスタイル	0%

Listing Preview Table:

Column 1	Column 2	Column 3
Row 1	Data (Num)	Data (Char)
Row 2	Data (Num)	Data (Char)

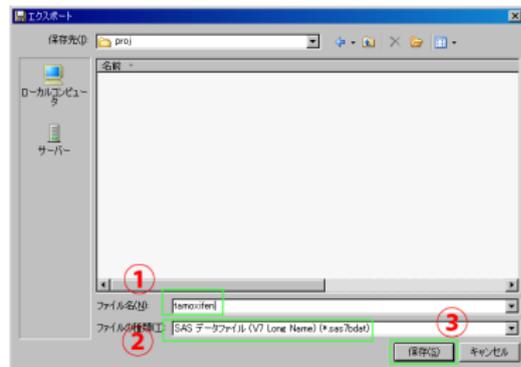
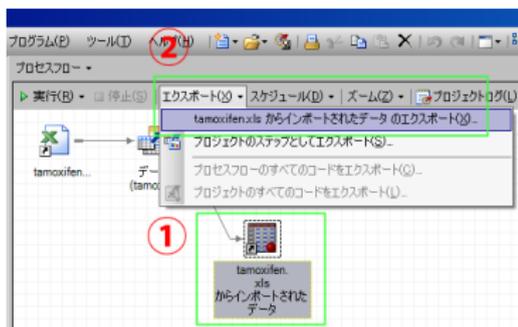
出力デザインのカスタマイズ (2)

- 作成したコピーに適切な名前を付けて [保存] → 作成したコピーを選択 → [編集]
- スタイルエディタで確認しながら編集 → [OK]



データのエクспорт

- SAS 形式のデータを選択 → [エクспорт] → [ファイル名] のエクспорт
- 保存先とファイル名および形式を指定 → [保存]



- 左の画像の下部のアイコンは SAS 形式のデータ
- SAS 形式以外も可

SAS / SAS Enterprise Guide インストールの注意点 I

- ① SAS のシステム必要要件 (Windows):
Windows XP Professional (Service Pack 2 を適用), Windows Vista Enterprise Edition / Business Edition / Ultimate Edition, Windows XP Professional x64 Edition, Windows Vista Enterprise x64 Edition / Business x64 Edition / Ultimate x64 Edition, Windows 7 Enterprise / Ultimate / Professional / Home Premium
Windows XP Home ではインストールも不可能
- ② SAS と SAS Enterprise Guide を別々にインストールすると SAS Enterprise Guide はうまく動作しませんので、同時にインストールして下さい。
- ③ ユーザアカウント名、PC 名、Workgroup 名が日本語の場合インストールできない可能性があるため、管理者権限を持ったユーザアカウント名、PC 名、Workgroup 名をアルファベットで新規作成してください。

SAS / SAS Enterprise Guide インストールの注意点 II

- ④ 以下のエラーメッセージが表示される場合 (Windows 7):
SAS Enterprise Guide にエラーが発生しました。作業中のファイルを保存して SAS Enterprise Guide を再起動してください。

エラー詳細：

System.IO.FileNotFoundException ファイルまたはアセンブリ
'Microsoft.Web.Services3, Version=3.0.0.0, Culture=neutral,
PublicKeyToken=31bf3856ad364e35'、またはその依存関係の 1 つが
読み込めませんでした。指定されたファイルが見つかりません。
→ Web Services Enhancements (WSE) 3.0 for Microsoft .NET をイン
ストールしてください。

<http://www.microsoft.com/download/en/details.aspx?id=14089>

SAS / SAS Enterprise Guide インストールの注意点 III

- ⑤ その他にもエラーが表示される場合、該当バージョンに対応するアップデートファイルをダウンロードしてインストールすることで問題を解決できる場合があります。

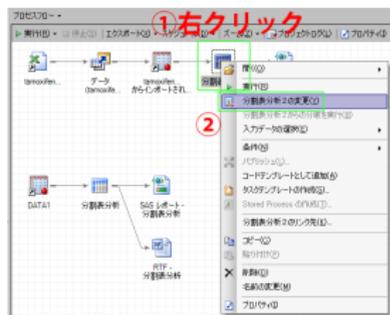
<http://ftp.sas.com/techsup/download/hotfix/HF2/A20.html>

<http://ftp.sas.com/techsup/download/hotfix/HF2/A32.html>

<http://ftp.sas.com/techsup/download/hotfix/HF2/A53.html>

Mantel-Haenszel 検定

- 目的
リンパ節転移の有無を考慮した上で、タモキシフェン使用の有無によって、乳がんの再発オッズに違いがあるかどうかを検討する事
- プロセスフローに戻る → [分割表分析 2] を右クリック → [分割表分析 2 の変更]
- [表統計量] → [関連] → [CMH 統計量] → [実行] → 結果の置き換え [はい]



仮説検定の結果

- リンパ節転移を考慮した上で再発オッズが異なるかどうかを、Mantel–Haenszel を用いて検定する ($\alpha = 0.05$ とする)

$$\left\{ \begin{array}{l} H_0 : \text{リンパ節転移のありの群でも, なしの群でもタモキシフェン使用} \\ \text{の有無で再発オッズは変わらない} \\ H_1 : \text{リンパ節転移のありの群でも, なしの群でもタモキシフェン使用} \\ \text{の有無で再発オッズは異なる} \end{array} \right.$$

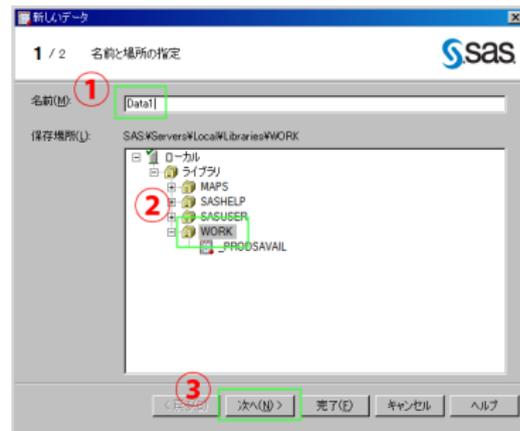
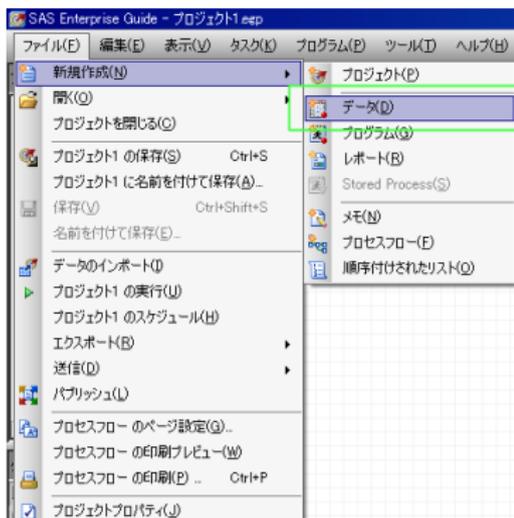
統計量	対立仮説	自由度	値	p 値
1 相関統計量		1	10.9946	0.0009
2 ANOVA 統計量		1	10.9946	0.0009
3 一般関連統計量		1	10.9946	0.0009

研究の種類	調整方法	値	95% 信頼限界
ケースコントロール研究 (オッズ比)	Mantel-Haenszel	0.6444	0.4962 0.8368
	ロジット	0.6444	0.4962 0.8368
コホート研究 (列 1 のリスク)	Mantel-Haenszel	0.6700	0.5275 0.8510
	ロジット	0.6700	0.5275 0.8510
コホート研究 (列 2 のリスク)	Mantel-Haenszel	1.0397	1.0164 1.0635
	ロジット	1.0397	1.0164 1.0635

- P -value=0.0009 であり, $\alpha = 0.05$ のもとで帰無仮説は棄却される
- リンパ節転移を考慮するとタモキシフェン使用の有無で再発オッズが異なる

分割表データの手入力 (1)

- [ファイル] → [新規作成] → [データ]
- ファイル名を指定 → 保存先として [WORK] をクリック → [次へ]



分割表データの手入力 (2)

- 列を三つにし、名前をそれぞれ [x] [y] [n] に設定 → [完了]
- 表を右図の様に入力し、左上のプロセスフローをダブルクリックして戻る

