

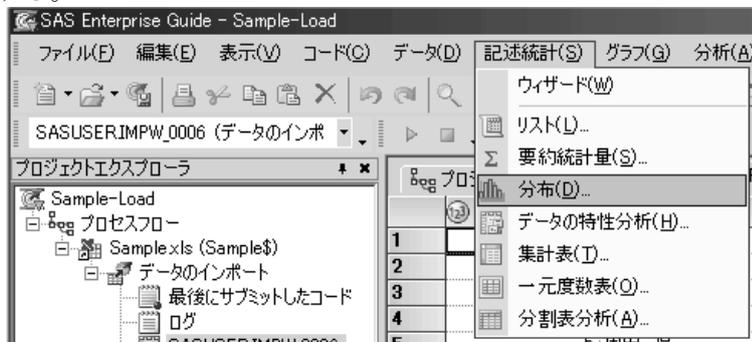
SAS (SAS Enterprise Guide) 使用方法
(2つのグループの1因子(1種類のデータ)の差を検定する)

目次

(1) 各グループ別のヒストグラムおよび、累積度数分布の確率プロットの表示	2
(2) t 検定 (t-test)を行う	5
(3) データを関数変換して分析を行う.....	12
(4) Mann-Whitney の U 検定 (Wilcoxon 順位和検定)を行う	15
(補足;..... 「中央値検定(メディアン検定)」.....)	20
(5) 記号の順位(順序)尺度データを数字に変換する.....	21
(6) χ^2 検定 (カイ二乗検定)を行う	25
(7) 対応のあるデータの差が正規分布しているかどうかを判定する.....	33
(8) 対応のある t 検定 (Paired t test)を行う.....	35
(9) Wilcoxon 符号順位和検定(Wilcoxon signed-ranks test)を行う	37

(1) 各グループ別のヒストグラムおよび、累積度数分布の確率プロットの表示

- 1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
- 2) 上部メニュー内の [記述統計] をクリックし、その下に表示されるメニュー内の [分布] をクリックする。



- 3) 新しいウィンドウが開くので、ヒストグラムや累積度数分布の確率プロットを表示する変数(正規分布しているかどうかを確認する変数)を「変数リスト」の下からのリストから選んでクリックする。そして、その右隣にある右向き矢印のボタンをクリックし、新たに表示されたメニューから「分析変数」をクリックする。



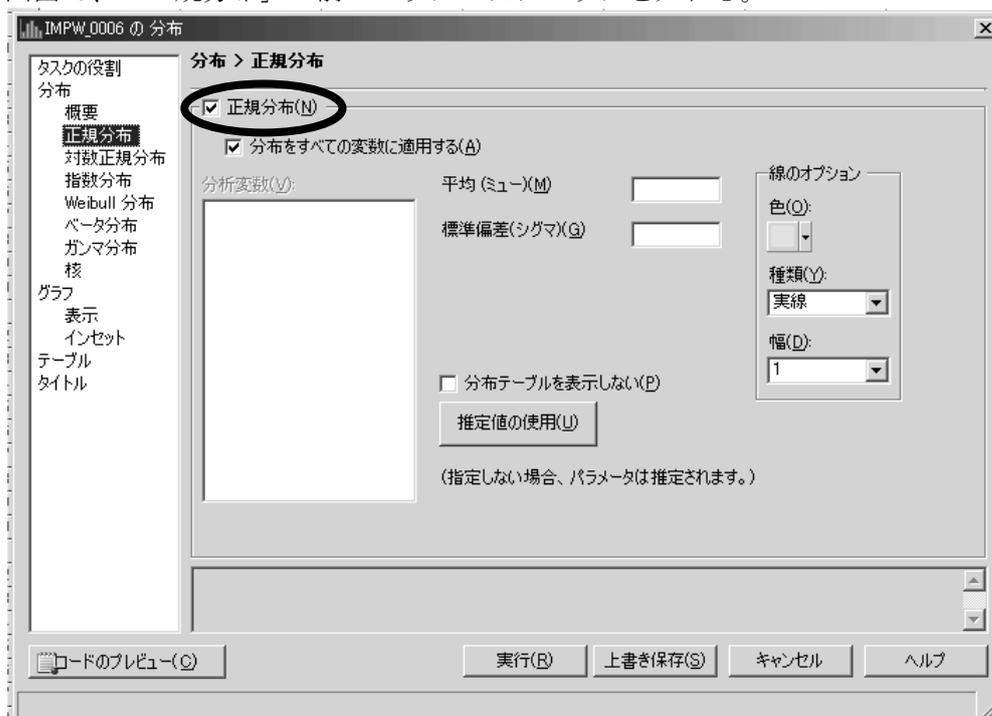
- 4) 右の「タスクの役割」の下「分析変数」の下に、選択した変数名が表示されていることを確認する。

- 5) グループ分けに使用する変数(その変数を使ってグループ分けする変数)を「変数リスト」の下からのリストから選んでクリックする。そして、その右隣りにある右向き矢印のボタンをクリックし、新たに表示されたメニューから「分類変数」をクリックする。

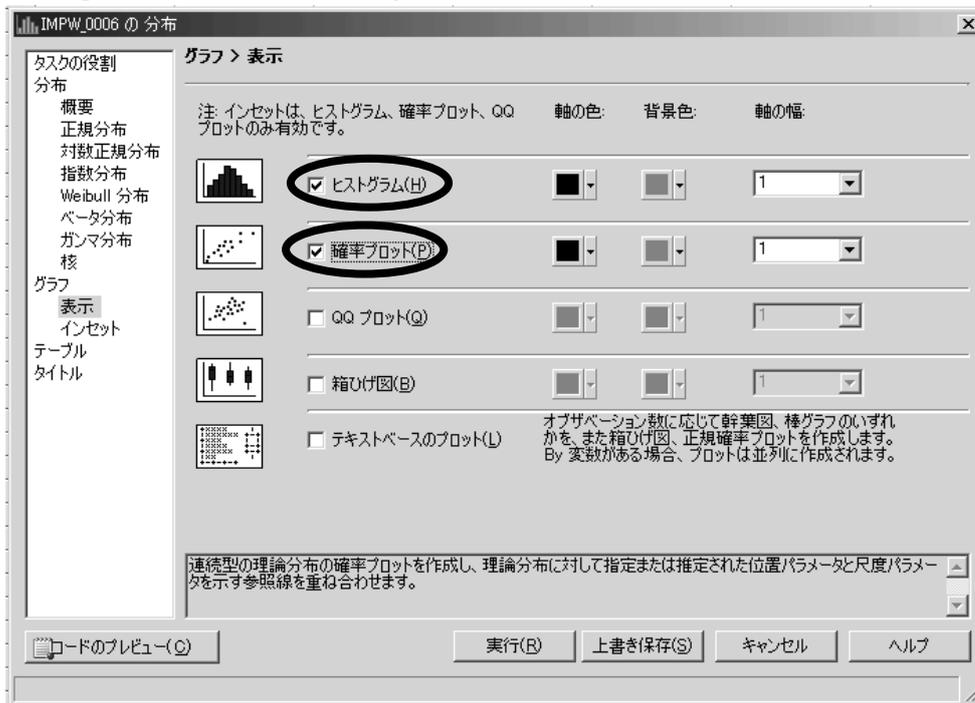


- 6) 右の「タスクの役割」の下「分類変数」の下に、選択した変数名が表示されていることを確認する。

- 7) 左端のメニューから「分布」の下「正規分布」をクリックし、新たに表示された右側の画面で、「正規分布」の前のボックスにチェックを入れる。



- 8) 左端のメニューから「グラフ」の下の「表示」をクリックし、新たに表示された右側の画面で、「ヒストグラム」と「確率プロット」の前のボックスにチェックを入れ、下にある「実行」ボタンをクリックする。

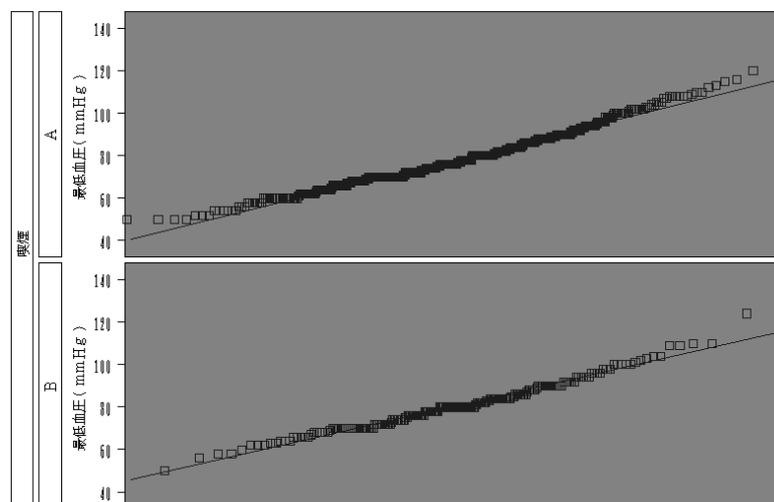


- 9) 画面に「分布」の分析結果が表示されるが、最下部にヒストグラムおよび、累積度数分布の確率プロットが、分類変数に設定した変数の値別に表示されている。

- 10) 最下部に下図のような分類変数に設定した変数の値別の累積度数分布の確率プロットがあるので、分析に使用するグループの確率プロット全てで、青い正方形のプロットが、左下から右上に伸びている黄色い直線上に乗っているかどうかを確認する。

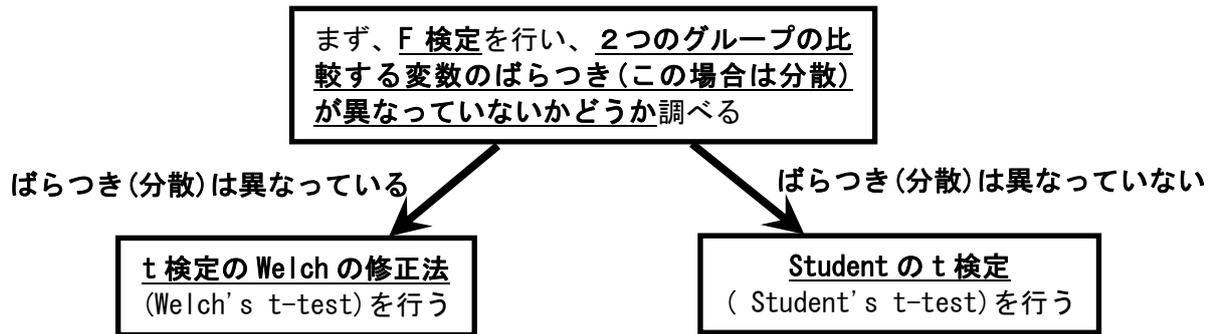
青いプロットが黄色い直線上にほぼ乗っている場合は、この変数(データ)は正規分布していると判断してよい。もし、青いプロットが黄色い直線から大きくずれている場合は、この変数は正規分布していないと判断できる。

分析: "最低血圧(mmHg)"n



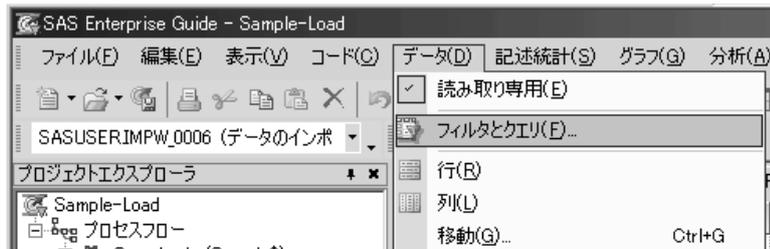
(2) t 検定 (t-test)を行う

t 検定は、必ず下記の流れで行うことを、まずしっかりと頭に入れておくこと



- 1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
- 2) t 検定を含めた比較検定を行う場合、その前に必ず、以下の操作で、グループ分けに用いる変数の値またはカテゴリーが、比較検定で用いる 2つの値またはカテゴリーであるデータのみ限定する必要がある。

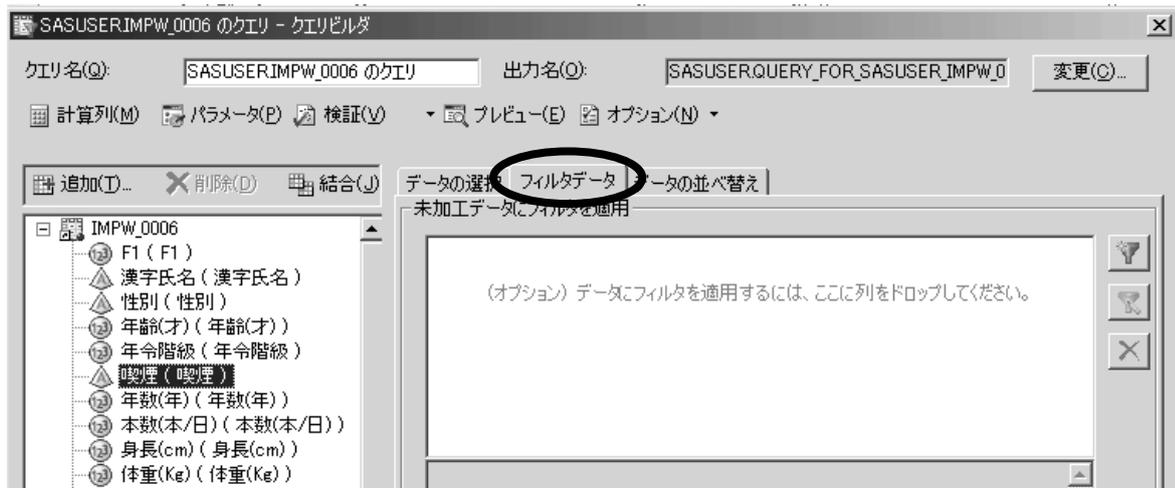
(a) 上部メニュー内の [データ]をクリックし、その下に表示されるメニュー内の [フィルタとクエリ]をクリックする。



(b) 下図のようなダイアログが開くので、左側の変数リストから、グループ分けに用いる変数と、比較する変数を、右側の「データの選択」の下のボックスにドラッグアンドドロップし、これらの変数名が表示されていることを確認する。



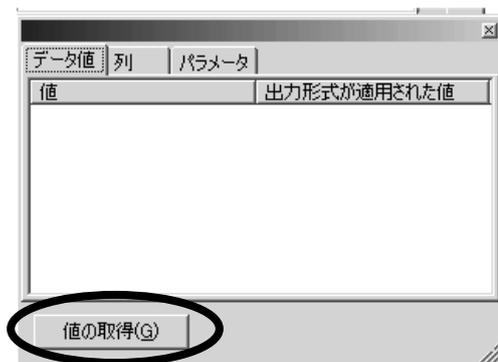
(c) 「データの選択」の右隣の「フィルタデータ」のタグをクリックし、その下に表示されたボックスに、グループ分けに用いる変数をドラッグアンドドロップする。



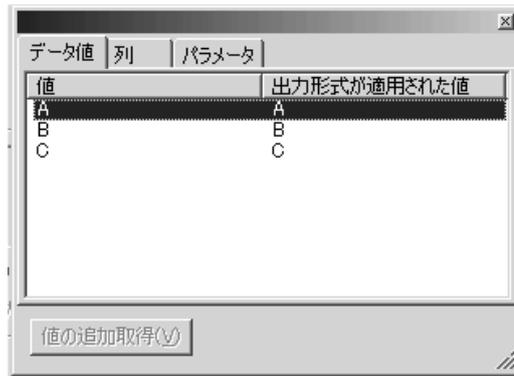
(d) 「フィルタの編集」というダイアログが開くので、「値 (V)」の行の右端の黒い逆三角形ボタンをクリックする。



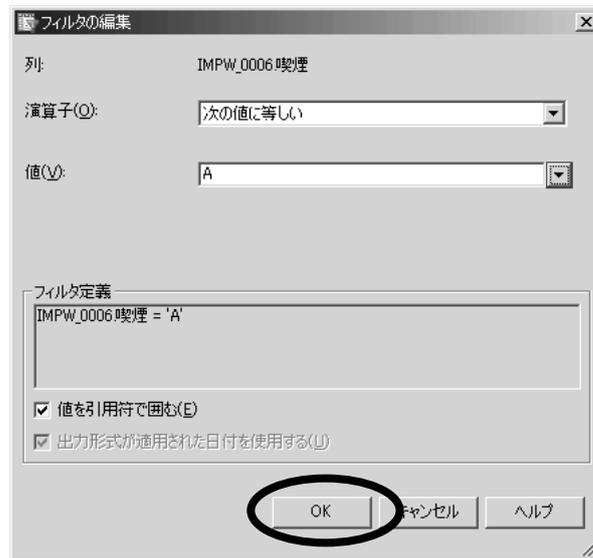
(e) 新しいダイアログが更に表示されるので、その下部にある「値の取得」ボタンをクリックする。



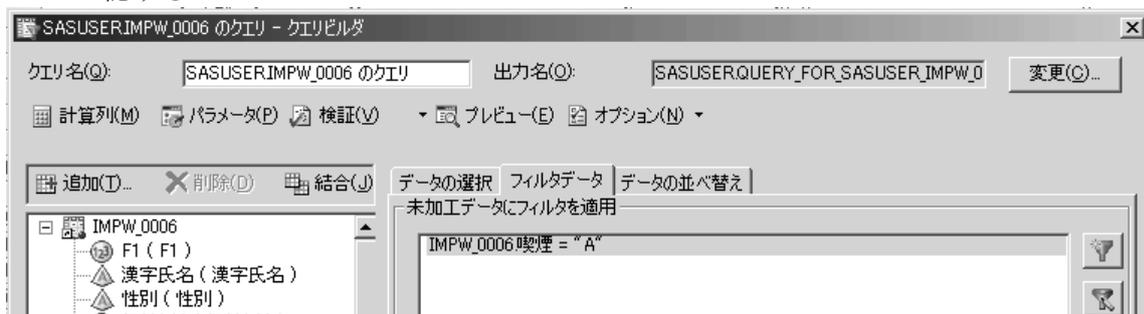
(f) グループ分けに用いる変数の値やカテゴリーのリストが表示されるので、まず、2つのグループの一方の値またはカテゴリーをクリックする。



(g) 「フィルタの編集」のダイアログに戻るので、「値」の右に、選択した値またはカテゴリーが表示されていることを確認し、下の「OK」ボタンをクリックする。

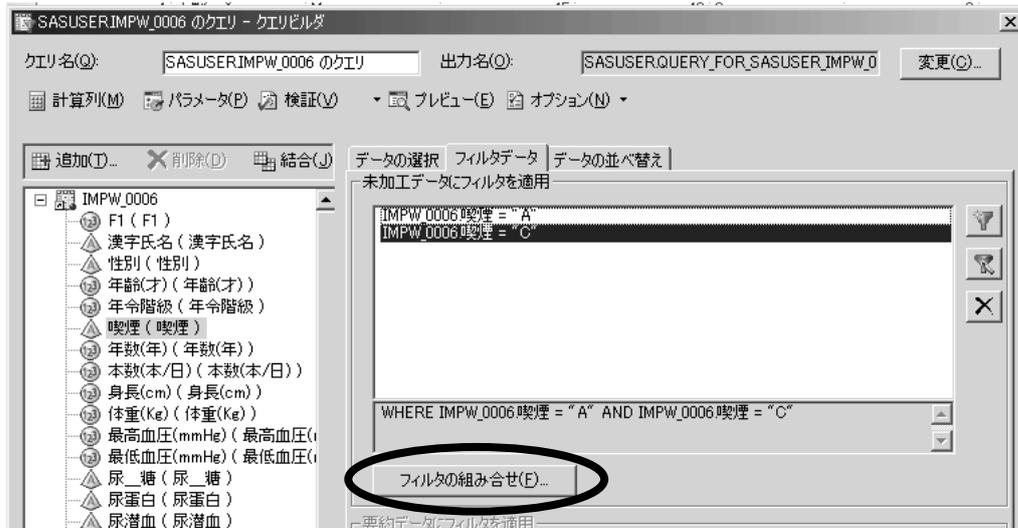


(h) 最初のダイアログに戻り、選択した値またはカテゴリーが表示されていることを確認する



(i) (c)～(h)の操作を繰り返し、2つのグループのもう一方の値またはカテゴリーを設定する。

(j) 2つのグループの値またはカテゴリが選択されていることを確認した後、その下の「フィルタの組み合わせ」ボタンをクリックする。



(k) 「フィルタの組み合わせ」というダイアログが新たに開くので、2つの漏斗の絵の間にある「AND ▼」をクリックし、その下に表示されたメニューから「ORに変更」をクリックする。

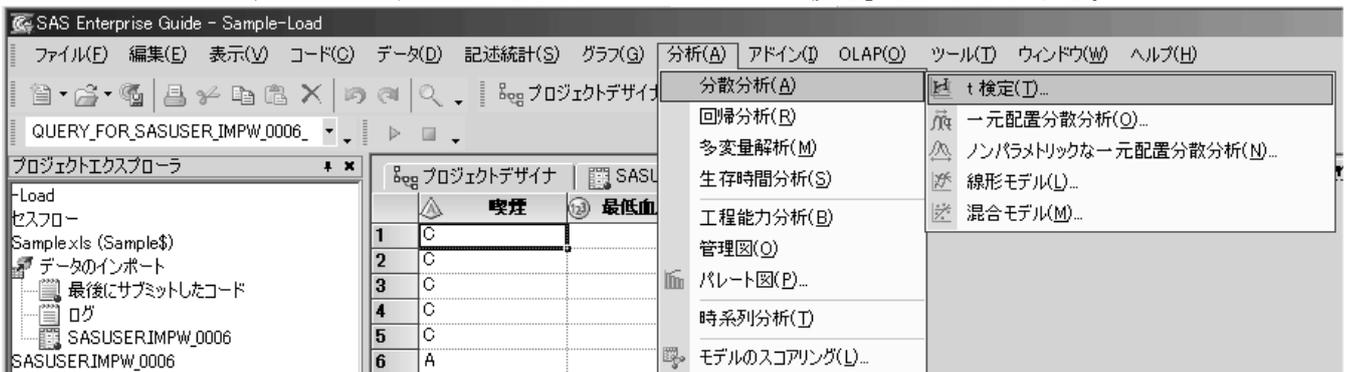


(l) 「AND ▼」が「OR ▼」に変わったことを確認した後、最下部の「OK」ボタンをクリックする。

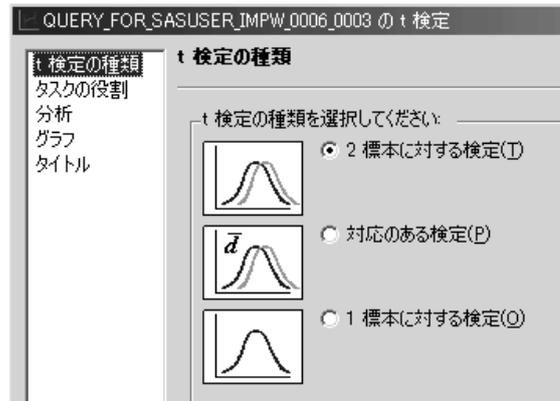
(m) 最初のダイアログに戻るので「実行」ボタンをクリックする。

(n) 画面上に、グループ分けに使用する変数と、比較する変数のみの表が、新たに表示されたことを確認する。

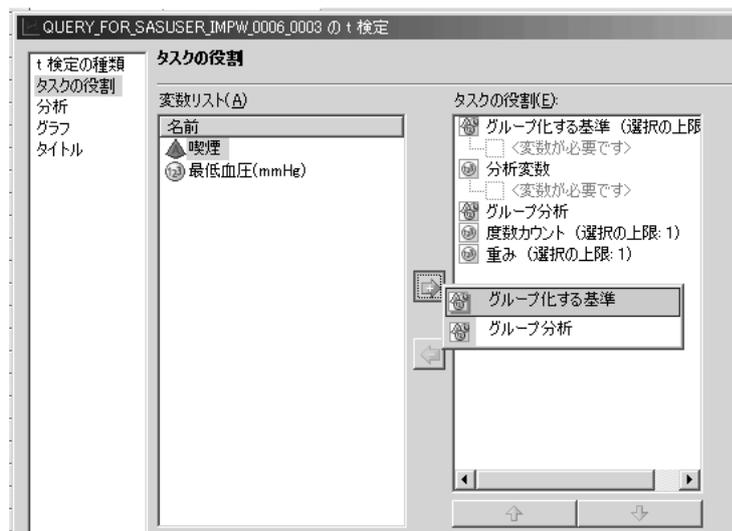
3) 上部メニュー内の [分析] をクリックし、その下に表示されるメニュー内の [分散分析] をクリックし、更にその右に表示されたメニュー内の 「t 検定」 をクリックする。



- 4) t 検定のダイアログが開くので、「t 検定の種類」で「2 標本に対する検定」の前にマークが入っていることを確認する。もし、マークが付いていない場合は、その前の丸をクリックしてマークを付ける。



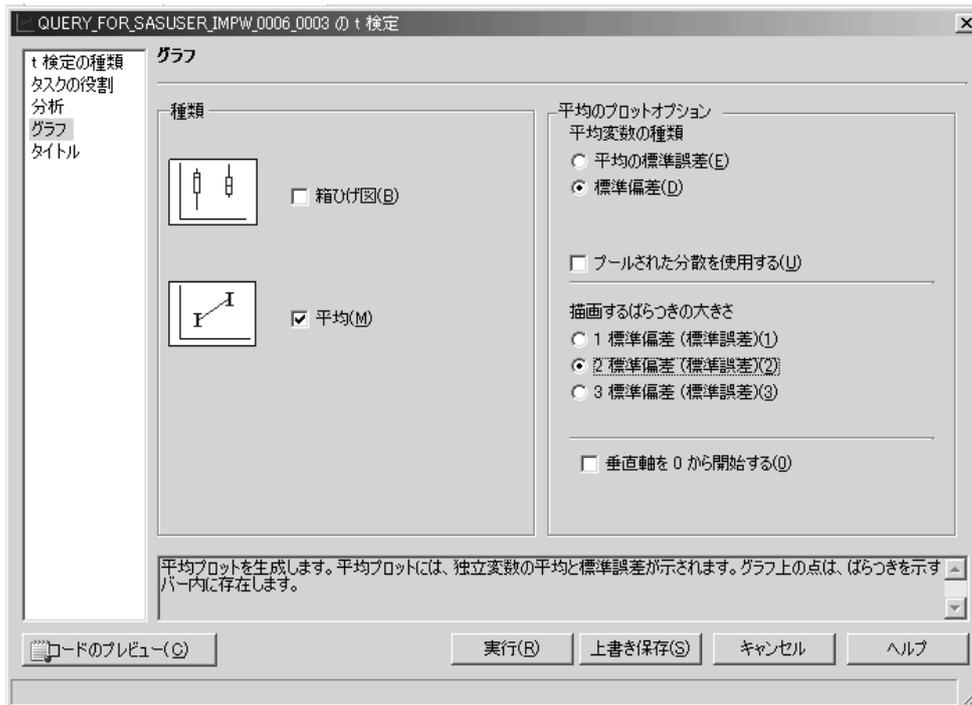
- 5) 左端のリストから「タスクの役割」をクリックし、その右に新たに表示された「変数リスト」からグループ分けに用いる変数をクリックした後、その右の右向き矢印のボタンをクリックし、新たに表示されたメニューから「グループ化する基準」をクリックし、「タスクの役割」の下の「グループ化する基準」の下に、その変数名が表示されたことを確認する。



- 6) 「変数リスト」から比較する変数をクリックした後、その右の右向き矢印のボタンをクリックし、新たに表示されたメニューから「分析変数」をクリックし、「タスクの役割」の下の「分析変数」の下に、その変数名が表示されたことを確認する。



- 7) 左端のリストから「グラフ」をクリックし、その右に新たに表示された「種類」で、「平均」の前のボックスにチェックを入れ、「平均のプロットオプション」では「標準偏差」にマークを付け、「描画するばらつきの大きさ」では「2標準偏差」の前にマークを付けた後、最下部の「実行」ボタンをクリックする。



- 8) 分析結果が表示されるので、まず、F 検定の結果を確認し、2つのグループ間で、比較する変数の分散(ばらつき)が異なっているかどうかを見る。

(注: 現在行っている分析は、2つのグループ間である変数の平均値に差があるかどうかを調べるものであるが、具体的な差の検定方法は、比較する変数の「ばらつき」が異なる場合と、異なる場合で、別々の検定方法を使う必要がある。このため、まず最初に、分散(ばらつき)に差があるかを調べる「F 検定」を行っている。しかし、本来の目的は、2つのグループ間である変数の平均値に差があるかどうかを調べることなので注意すること)

表示されている表の中から、下図のような「等分散性」と書かれた表を見る。この表は、2つのグループ間で比較する変数の分散(ばらつき)が異なっているかどうかを調べる F 検定の結果を表示したものであるが、その中から、「Pr > F」と書かれたところの下の数値を確認する。この値が、「F 検定の危険率」(2つのグループの分散に差が無い)ことが起こりうる確率の値)なので、この値が 5 % (0.05) 未満の場合は、「2つのグループの分散には差がある」とみなすことができる。この値が 5 % (0.05) 以上の場合は、2つのグループの分散に差があるかどうかの判定は「保留」されるが、この場合は便宜上「2つのグループの分散には差が無い」とみなす。

等分散性					
変数	手法	分子の自由度	分母の自由度	F 値	Pr > F
最低血圧(mmHg)	Folded F	697	280	1.116	0.5634

- 9) 8) で、もし、(便宜上)「2つのグループの分散(ばらつき)には差が無い」とみなすことができた場合、10)の処理を行う。

もし、「2つのグループの分散(ばらつき)には差がある」とみなすことができた場合は、11)の処理を行う。

10) 8)の「F 検定」の結果が、(便宜上)「2つのグループの分散(ばらつき)には差が無い」とみなすことができた場合、表示されている表の中から、下図のような「t 検定」と書かれた表を見る。その表中から、「分散」の下に「Equal」と書かれた行で、「Pr>|t|」の下の数値を確認する。この値が、「t 検定の危険率」(2つのグループの間で比較した変数の平均値に差が無い)ことが起こりうる確率の値)なので、この値が5% (0.05)未満の場合は、帰無仮説を棄却できる(否定できる)ため、5%の有意水準で「比較した変数の平均値は2つのグループの間で有意な差がある」とみなすことができる。

これに対し、この値が5% (0.05)以上の場合は、帰無仮説を棄却できない(否定できない)ため、2つのグループの間に差があるかどうかの判定は「保留」となり、「2つのグループの間で比較した変数の平均値に有意な差があるとはいえない」あるいは「2つのグループの間で比較した変数の平均値に有意な差は見られなかった」という結論になる。決して「2つのグループ間に差が無い」という結論は出せない。これは、危険率が「帰無仮説が正しいのに、間違っ棄却してしまう確率」であるためである。また現実的にも、サンプル数を増加させた場合や他の検定方法を使用した場合、有意差が検出される可能性があるためである。

t検定					
変数	手法	分散	自由度	t 値	Pr > t
最低血圧(mmHg)	Pooled	Equal	977	-2.08	0.0376
最低血圧(mmHg)	Satterthwaite	Unequal	531	-2.11	0.0354

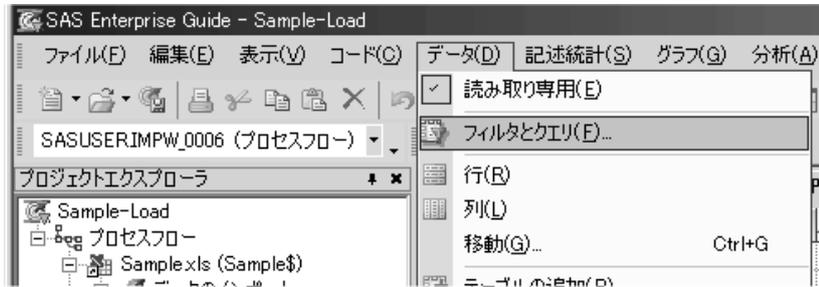
11) 8)の「F 検定」の結果が、「2つのグループの分散(ばらつき)には差がある」とみなすことができた場合、表示されている表の中から、下図のような「t 検定」と書かれた表を見る。その表中から、「分散」の下に「Unequal」と書かれた行で、「Pr>|t|」の下の数値を確認する。この値が、「t 検定の危険率」(2つのグループの間で比較した変数の平均値に差が無い)ことが起こりうる確率の値)なので、この値が5% (0.05)未満の場合は、帰無仮説を棄却できる(否定できる)ため、5%の有意水準で「比較した変数の平均値は2つのグループの間で有意な差がある」とみなすことができる。

これに対し、この値が5% (0.05)以上の場合は、帰無仮説を棄却できない(否定できない)ため、2つのグループの間に差があるかどうかの判定は「保留」となり、「2つのグループの間で比較した変数の平均値に有意な差があるとはいえない」あるいは「2つのグループの間で比較した変数の平均値に有意な差は見られなかった」という結論になる。決して「2つのグループ間に差が無い」という結論は出せない。これは、危険率が「帰無仮説が正しいのに、間違っ棄却してしまう確率」であるためである。また現実的にも、サンプル数を増加させた場合や他の検定方法を使用した場合、有意差が検出される可能性があるためである。

t検定					
変数	手法	分散	自由度	t 値	Pr > t
最低血圧(mmHg)	Pooled	Equal	977	-2.08	0.0376
最低血圧(mmHg)	Satterthwaite	Unequal	531	-2.11	0.0354

(3) データを関数変換して分析を行う

- 1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
- 2) 上部メニュー内の [データ] をクリックし、その下に表示されるメニュー内の [フィルタとクエリ] をクリックする。



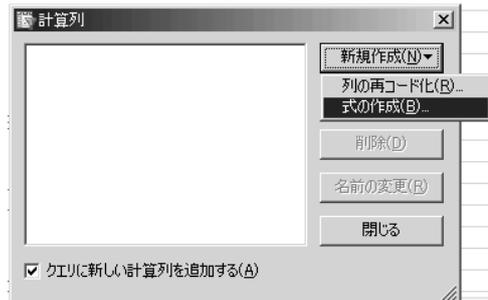
- 3) 新しいウィンドウが開くので、左側の変数のリストからグループ分けに用いる変数を選び、ダブルクリックし、右側の枠内にその変数が表示されていることを確認する。



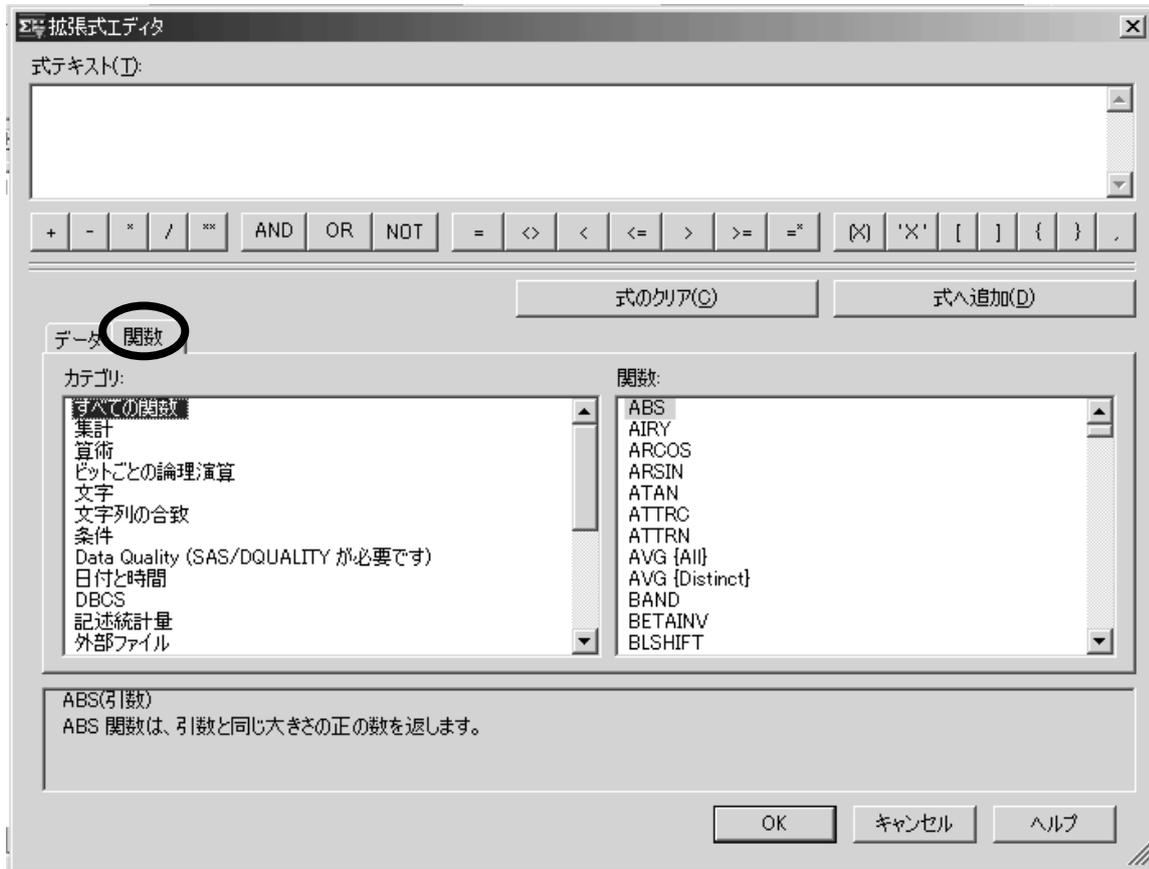
- 4) 左上にある「計算列」をクリックする。



- 5) 「計算列」というダイアログが開くので、右上の「新規作成」ボタンをクリックし、その下に新たに表示されたメニュー内の「式の作成」をクリックする。



- 6) 「拡張式エディタ」というダイアログが開くので、左中央の「関数」タグをクリックし、右下の「関数」の下から、変換に用いる関数を選び、ダブルクリックする。
(例; 対数変換を行う場合は、「LOG」の関数を選択する)



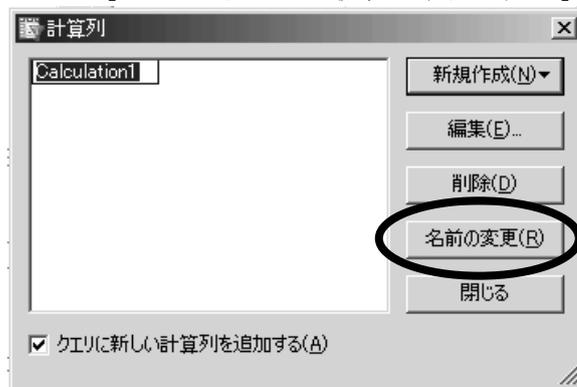
- 7) 「式テキスト」の下ボックスに選択した関数が表示されるので、その中の「<numValue>」をマウスで選択した後、「BackSpace」キーか「Delete」キーで削除する。



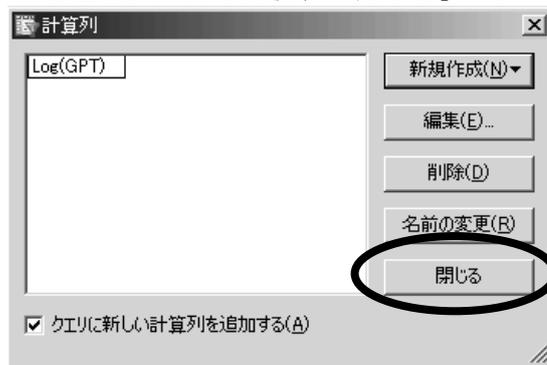
- 8) 左中央の「データ」タグをクリックし、「変数値」の下のリストから変換する変数を選んでダブルクリックし、「式テキスト」の下の関数の中にその変数が表示されたことを確認した後、最下部の「OK」ボタンをクリックする。



- 9) 「拡張式エディタ」が閉じて、「計算式」ダイアログに戻るので、左のボックス内に表示されている「Calculation1」をクリックした後、「名前の変更」ボタンをクリックする。



- 10) 変換後の変数の名前を適当に入力した後、「閉じる」ボタンをクリックする。



11) 新しい変数名が、グループ分けに用いる変数の下に表示されていることを確認した後、「実行」ボタンをクリックする。

12) グループ分けに用いる変数と関数変換後の比較する変数の表が表示されるので、これらを用いて、この資料の「(1) 各グループ別のヒストグラムおよび、累積度数分布の確率プロットの表示」の操作を行い、関数変換後の比較する変数が、2つのグループのどちらでも正規分布しているかどうかを判定する。

(注; なお、対応のある2つのデータの差を関数変換した場合は、配布資料「SAS (SAS Enterprise Guide) 使用方法 (1 因子(1 種類のデータ)の概要を見る)」の「(1) ヒストグラムおよび、累積度数分布の確率プロットの表示」の操作を行い、関数変換後の対応のある2つのデータの差の値が正規分布しているかどうかを判定する。

もし、正規分布していた場合は、13)の操作に進むが、もし、関数変換後も正規分布していない場合は、配布資料「統計学基礎(3) 2つのグループの1因子(1種類のデータ)の差を検定する」の(12)を参照する。

13) もし、関数変換後の比較する変数が、2つのグループのどちらでも正規分布している場合は、この資料の「(2) t 検定 (t-test)を行う」の操作を行い、2つのグループ間で比較する変数の関数変換後の平均値に差があるかどうかを判定する。

(4) Mann-Whitney の U 検定 (Wilcoxon 順位和検定)を行う

1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
2) Mann-Whitney の U 検定 (Wilcoxon 順位和検定)を含めた比較検定を行う場合、その前に必ず、以下の操作で、グループ分けに用いる変数の値またはカテゴリーが、比較検定で用いる2つの値またはカテゴリーであるデータのみ限定する必要がある。

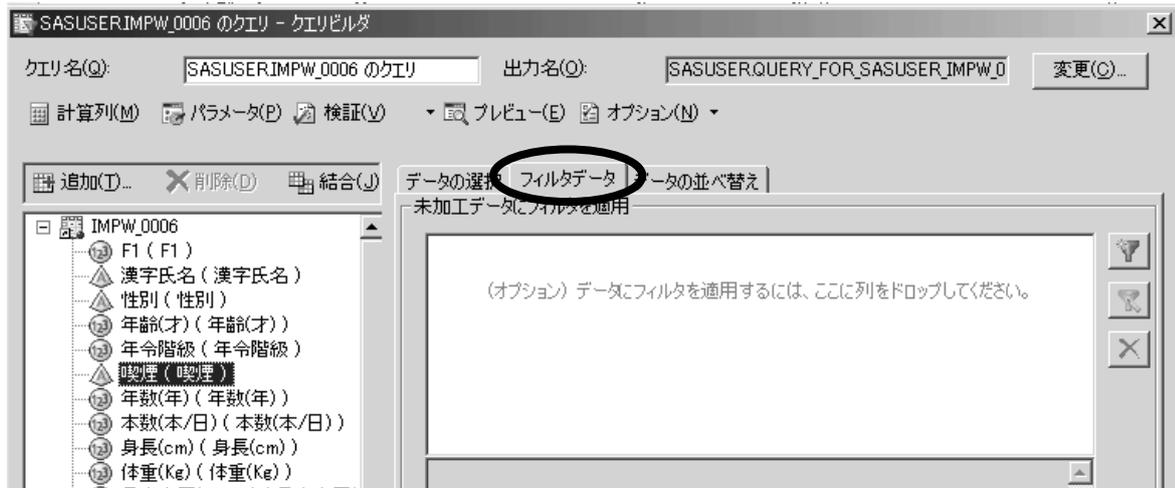
(a) 上部メニュー内の [データ]をクリックし、その下に表示されるメニュー内の [フィルタとクエリ]をクリックする。



(b) 下図のようなダイアログが開くので、左側の変数リストから、グループ分けに用いる変数と、比較する変数を、右側の「データの選択」の下のボックスにドラッグアンドドロップし、これらの変数名が表示されていることを確認する。



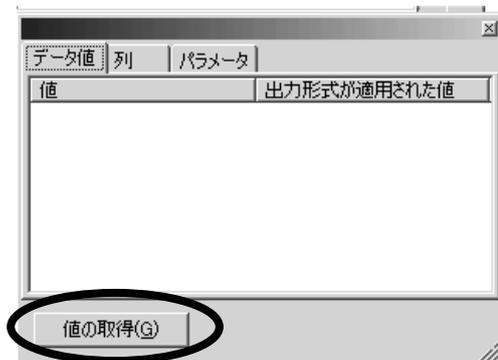
(c) 「データの選択」の右隣の「フィルタデータ」のタグをクリックし、その下に表示されたボックスに、グループ分けに用いる変数をドラッグアンドドロップする。



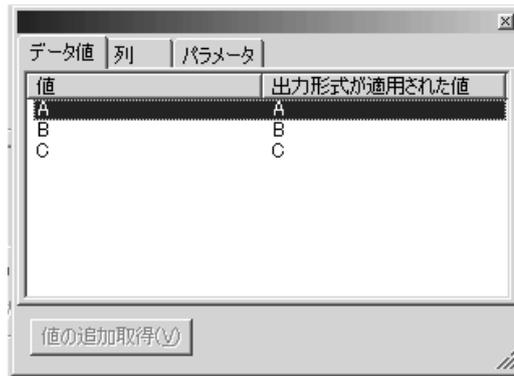
(d) 「フィルタの編集」というダイアログが開くので、「値(V)」の行の右端の黒い逆三角形ボタンをクリックする。



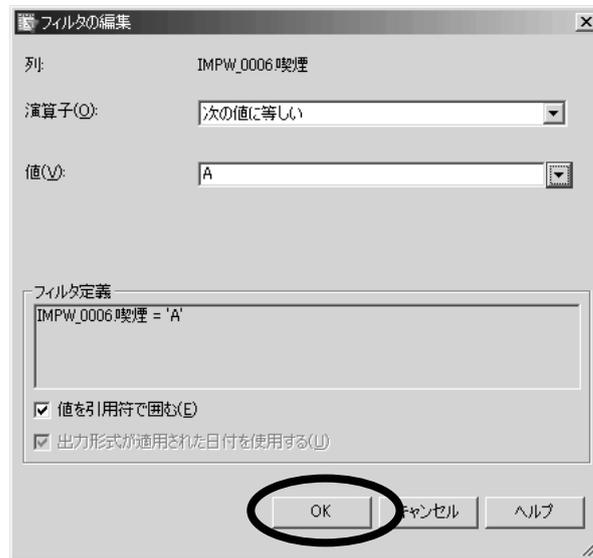
(e) 新しいダイアログが更に表示されるので、その下部にある「値の取得」ボタンをクリックする。



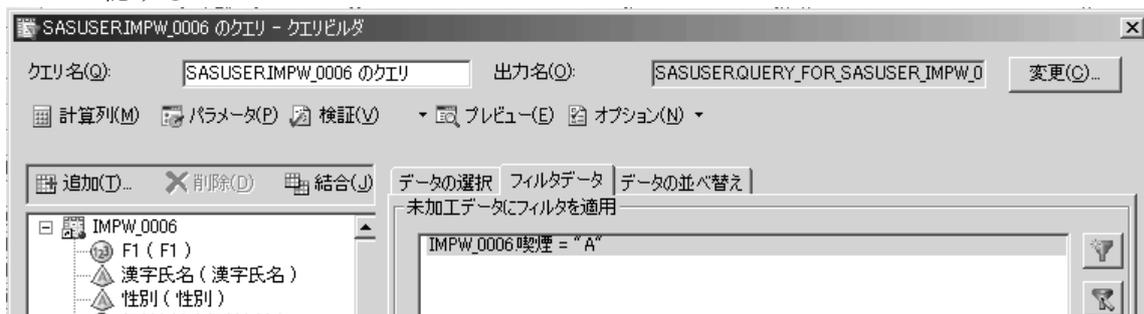
(f) グループ分けに用いる変数の値やカテゴリーのリストが表示されるので、まず、2つのグループの一方の値またはカテゴリーをクリックする。



(g) 「フィルタの編集」のダイアログに戻るので、「値」の右に、選択した値またはカテゴリーが表示されていることを確認し、下の「OK」ボタンをクリックする。

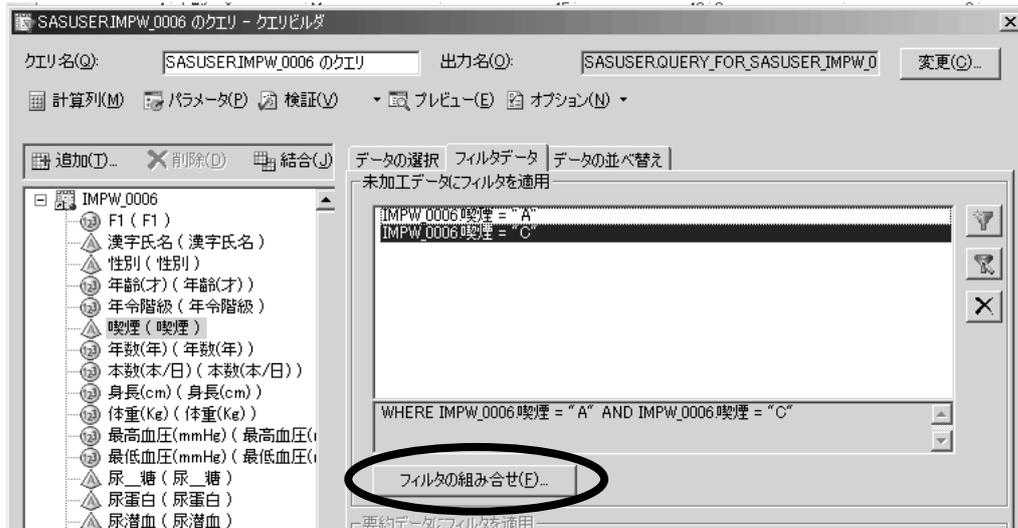


(h) 最初のダイアログに戻り、選択した値またはカテゴリーが表示されていることを確認する



(i) (c)～(h)の操作を繰り返し、2つのグループのもう一方の値またはカテゴリーを設定する。

(j) 2つのグループの値またはカテゴリが選択されていることを確認した後、その下の「フィルタの組み合わせ」ボタンをクリックする。



(k) 「フィルタの組み合わせ」というダイアログが新たに開くので、2つの漏斗の絵の間にある「AND ▼」をクリックし、その下に表示されたメニューから「ORに変更」をクリックする。

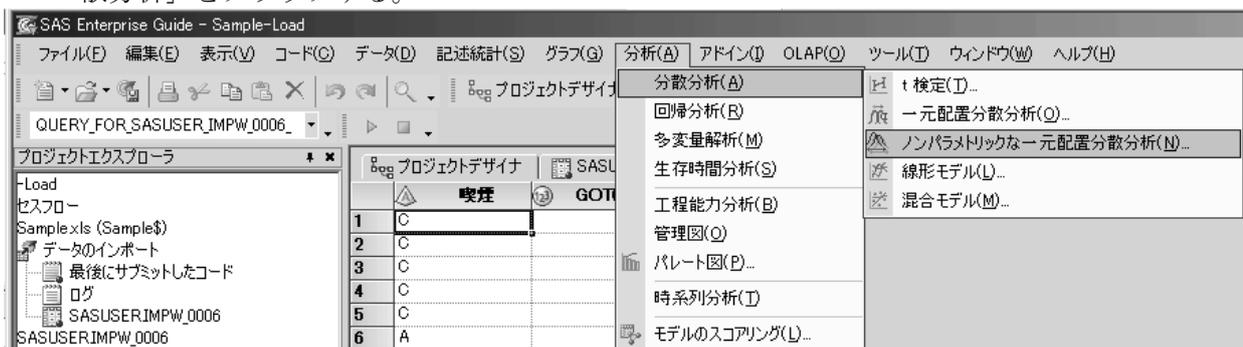


(l) 「AND ▼」が「OR ▼」に変わったことを確認した後、最下部の「OK」ボタンをクリックする。

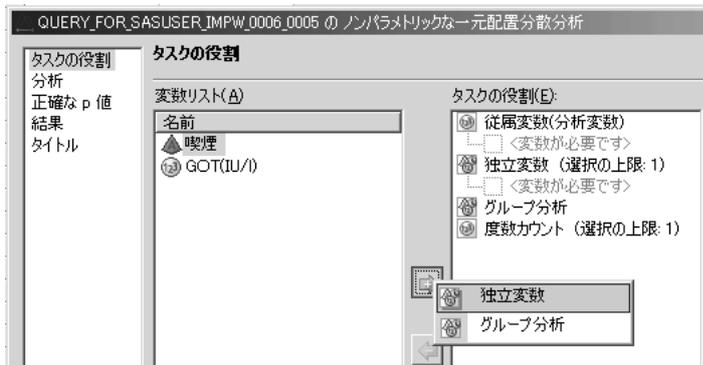
(m) 最初のダイアログに戻るので「実行」ボタンをクリックする。

(n) 画面上に、グループ分けに使用する変数と、比較する変数のみの表が、新たに表示されたことを確認する。

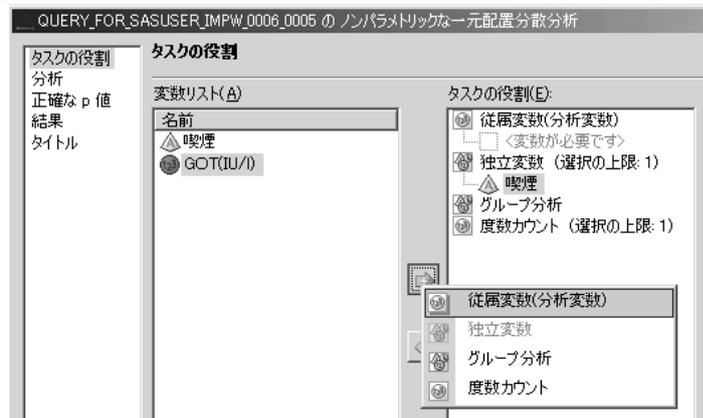
3) 上部メニュー内の [分析] をクリックし、その下に表示されるメニュー内の [分散分析] をクリックし、更にその右に表示されたメニュー内の「ノンパラメトリックな一元配置分散分析」をクリックする。



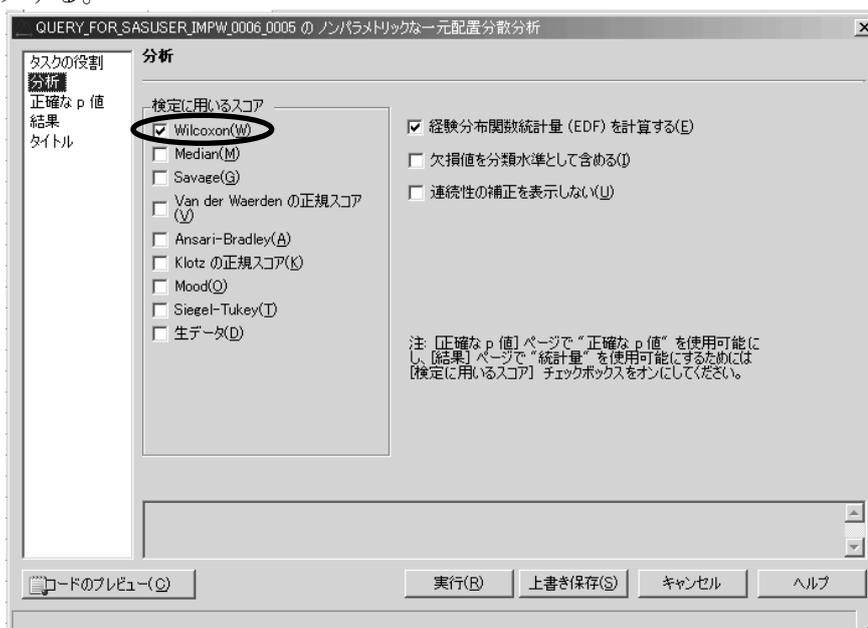
- 4) 新しいダイアログが開くので、「変数リスト」からグループ分けに用いる変数をクリックした後、その右の右向き矢印のボタンをクリックし、新たに表示されたメニューから「独立変数」をクリックし、「タスクの役割」の下の「独立変数」の下に、その変数が表示されたことを確認する。



- 5) 「変数リスト」から比較する変数をクリックした後、その右の右向き矢印のボタンをクリックし、新たに表示されたメニューから「従属変数」をクリックし、「タスクの役割」の下の「従属変数」の下に、その変数名が表示されたことを確認する。



- 6) 左端のリストから「分析」をクリックし、その右に新たに表示された「検定に用いるスコア」の下の「Wilcoxon」の前のボックスにチェックを入れ、最下部の「実行」ボタンをクリックする。



(補足; 「Wilcoxon」の下にある「Median」の前のボックスにチェックを入れると、「中央値検定(メディアン検定)」が行われる)

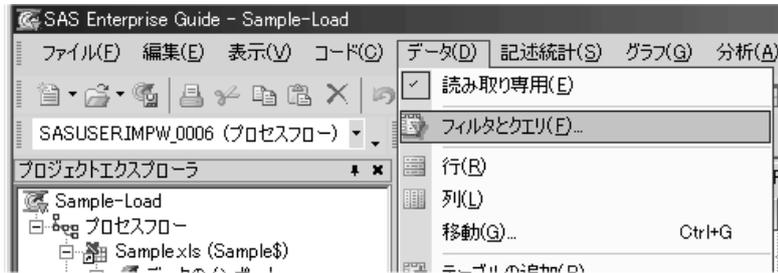
8) 分析結果が表示されるので、下図のような「Wilcoxon の順位和検定(2標本)」と書かれた表を見る。その表中から「正規近似」の3つ下にある「両側 $Pr > |Z|$ 」の右横の数値を確認する。この値が、「Mann-Whitney の U 検定 (Wilcoxon 順位和検定)の危険率」(2つのグループの間で比較した変数に差が無い) ことが起こりうる確率の値)なので、この値が **5% (0.05)未満**の場合は、帰無仮説を棄却できる(否定できる)ため、5%の有意水準で「**比較した変数は2つのグループの間で有意な差がある**」とみなすことができる。

これに対し、この値が **5% (0.05)以上**の場合は、帰無仮説を棄却できない(否定できない)ため、2つのグループの間に差があるかどうかの判定は「**保留**」となり、「**2つのグループの間で比較した変数に有意な差があるとはいえない**」あるいは「**2つのグループの間で比較した変数に有意な差は見られなかった**」という結論になる。決して「2つのグループ間で比較した変数に**差が無い**」という**結論は出せない**。これは、危険率が「帰無仮説が正しいのに、間違っ棄却してしまう確率」であるためである。また現実的にも、サンプル数を増加させた場合や他の検定方法を使用した場合、有意差が検出される可能性があるためである。

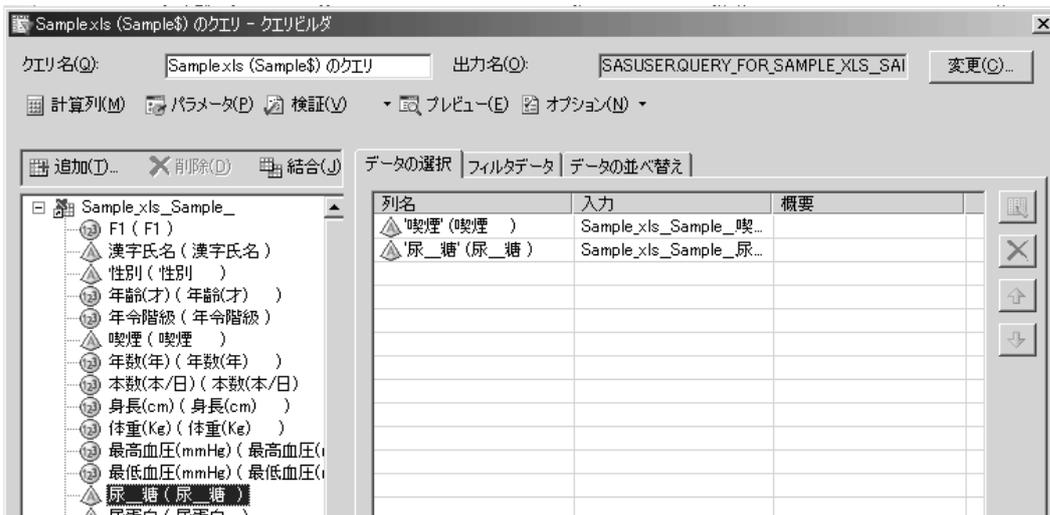
Wilcoxon の順位和検定(2標本)	
統計量	99725.0000
正規近似	
Z	-0.5769
片側 $Pr < Z$	0.2820
両側 $Pr > Z$	0.5640
t分布で近似	
片側 $Pr < Z$	0.2821
両側 $Pr > Z $	0.5642
Zには 0.5の連続性の補正が含まれています。	

(5) 記号の順位(順序)尺度データを数字に変換する

- 1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
- 2) 上部メニュー内の [データ] をクリックし、その下に表示されるメニュー内の [フィルタとクエリ] をクリックする。



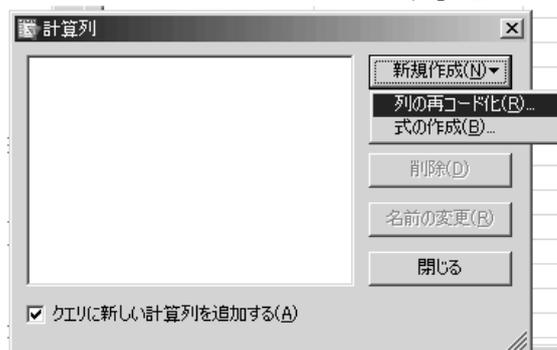
- 3) 新しいウィンドウが開くので、左側の変数のリストからグループ分けに用いる変数を選び、ダブルクリックし、更に、数字に変換する変数を選び、ダブルクリックし、右側の枠内にその2つの変数が表示されていることを確認する。



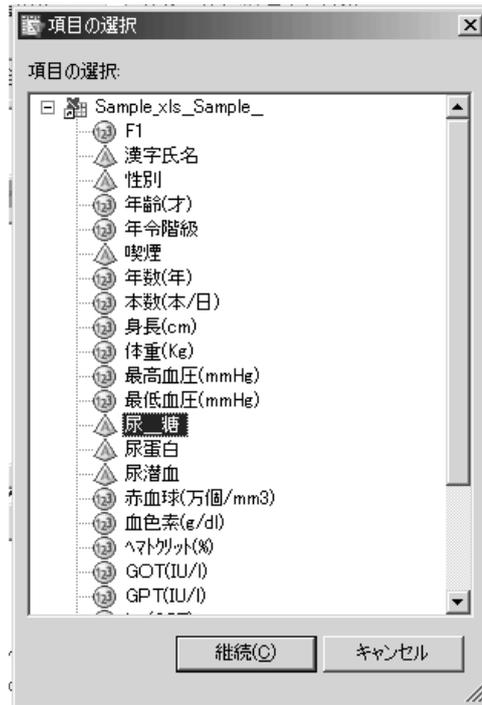
- 4) 左上にある「計算列」をクリックする。



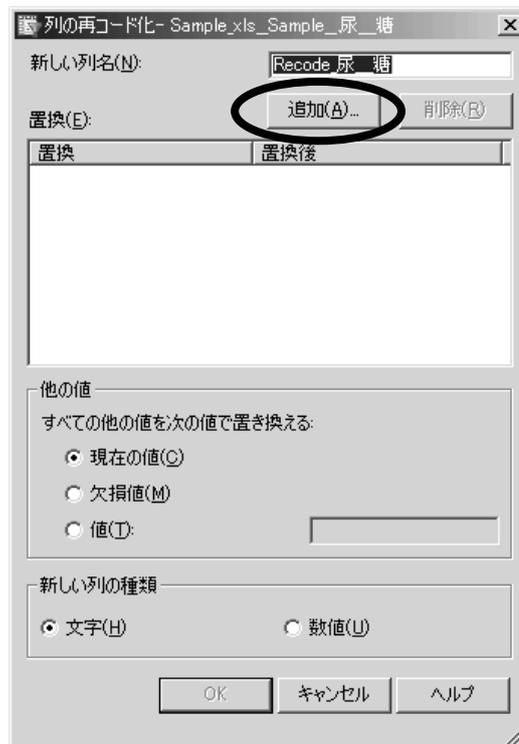
- 5) 「計算列」というダイアログが開くので、右上の「新規作成」ボタンをクリックし、その下に新たに表示されたメニュー内の「列の再コード化」をクリックする。



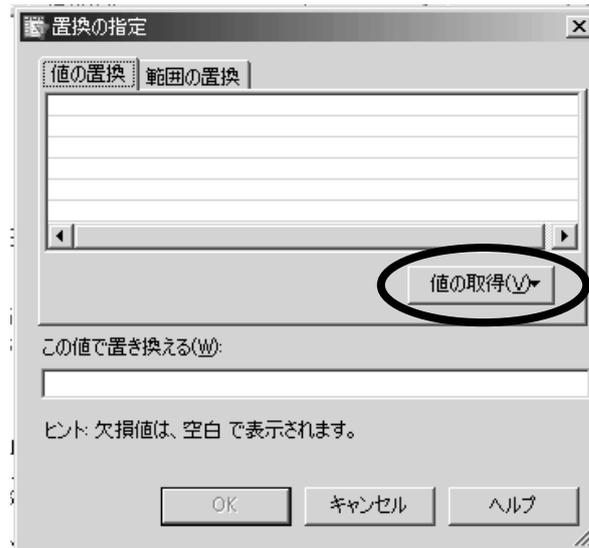
- 6) 「項目の選択」というダイアログが開くので、「項目の選択」の下のリストから数字に変換する変数をクリックし、最下部の「継続」ボタンをクリックする。



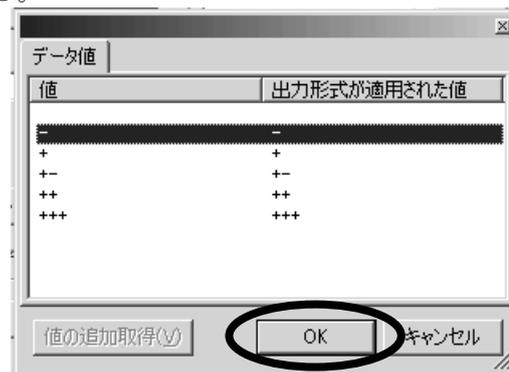
- 7) 「列の再コード化」というダイアログが開くので、もし、新しい変数名を付ける場合は、「新しい列名」の右横のボックス内の名前を変更する。次に、その下の「追加」ボタンをクリックする。



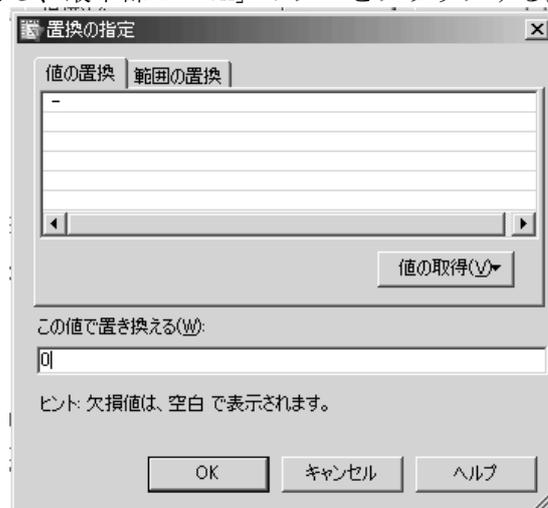
8) 「置換の指定」というダイアログが開くので、中央右にある、「値の取得」ボタンをクリックする。



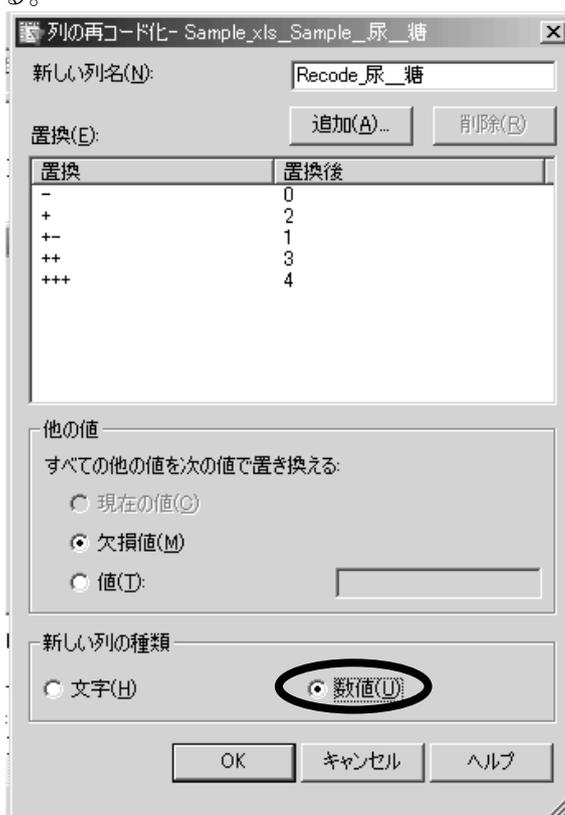
9) 新しいダイアログが開くので、データ値のリストからどれかの値をクリックし、「OK」ボタンをクリックする。



10) 「置換の指定」ダイアログに戻るので、「この値で置き換える」の下のボックスに変換する数字の値を入力し、最下部の「OK」ボタンをクリックする。



- 11) 7)～10)を繰り返して、再分類前の値と再分類後の値を全て設定する。
- 例) 例えば、尿糖の「-」「+-」「+」「++」「+++」を数字に変換する場合は、
 最初に9)で「-」を選んだ後、10)で「0」を入力、
 次に9)で「+-」を選んだ後、10)で「1」を入力、
 更に9)で「+」を選んだ後、10)で「2」を入力、
 その次に9)で「++」を選んだ後、10)で「3」を入力、
 最後に9)で「+++」を選んだ後、10)で「4」を入力する
- 12) 全ての設定が終了したら、「列の再コード化」ダイアログの下部にある「新しい列の種類」の所で、「数値」の前の丸をクリックしてマークを付けた後、最下部の「OK」ボタンをクリックする。



- 13) 「計算列」ダイアログに戻るので、最下部の「閉じる」ボタンをクリックする。
- 14) 最初のダイアログに戻るので、最下部の「実行」ボタンをクリックする。
- 15) グループ分けに用いる変数と、数字変換前のデータと数字変換後のデータの表が表示されるので、正しく数字変換されていることを確認する。
- 16) 上記の表を用いて、適切な統計手法で分析を行う。

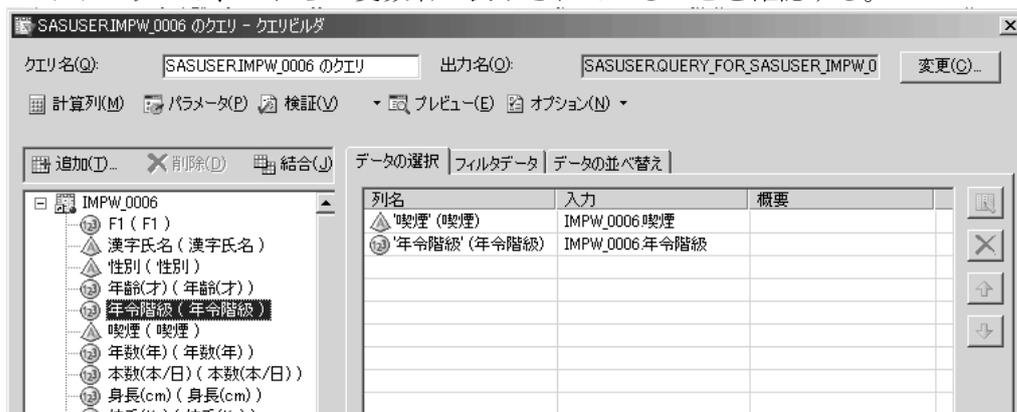
(6) χ^2 検定 (カイ二乗検定)を行う

- 1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
- 2) χ^2 検定 (カイ二乗検定)の場合、比較するグループが3つ以上でも特に問題は無いが、分析結果を判断する際に、グループが2つである方が分かりやすいため、以下の操作で、グループ分けに用いる変数の値またはカテゴリーを、2つの値またはカテゴリーであるデータのみ限定した方が良い(再分類によって、2つのグループにまとめ直す方法でも良い)。また、比較するデータ(変数)についても、分析後の判断を分かりやすくするためには、カテゴリーを限定するか、再分類することで、少なくすることが望ましい(グループ分けする変数も、比較する変数も、カテゴリーが2つである場合が、最も分かりやすいので、最良と言える)。

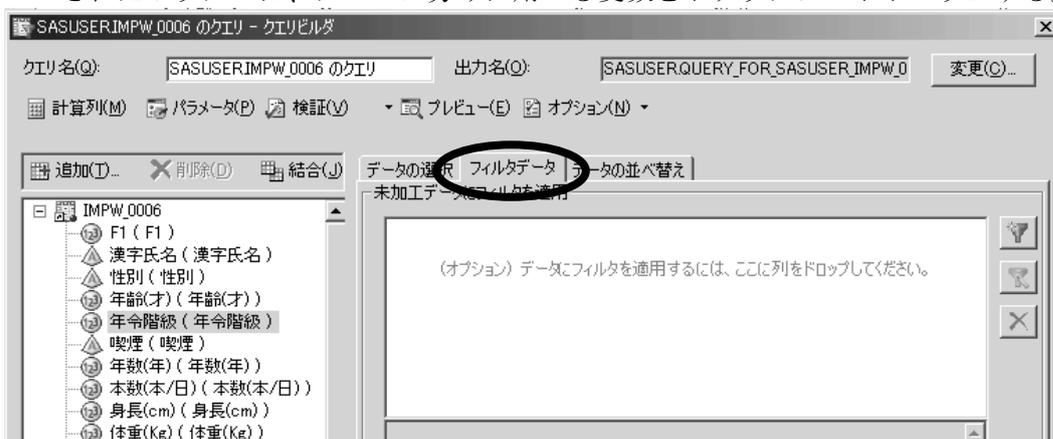
(a) 上部メニュー内の [データ]をクリックし、その下に表示されるメニュー内の [フィルタとクエリ]をクリックする。



(b) 下図のようなダイアログが開くので、左側の変数リストから、グループ分けに用いる変数と、比較する変数を、右側の「データの選択」の下のボックスにドラッグアンドドロップし、これらの変数名が表示されていることを確認する。



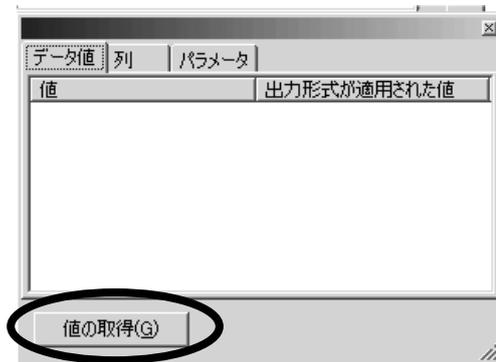
(c) 「データの選択」の右隣の「フィルタデータ」のタグをクリックし、その下に表示されたボックスに、グループ分けに用いる変数をドラッグアンドドロップする。



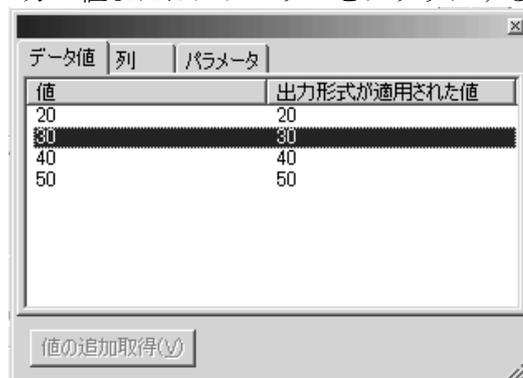
(d) 「フィルタの編集」というダイアログが開くので、「値(V)」の行の右端の黒い逆三角形ボタンをクリックする。



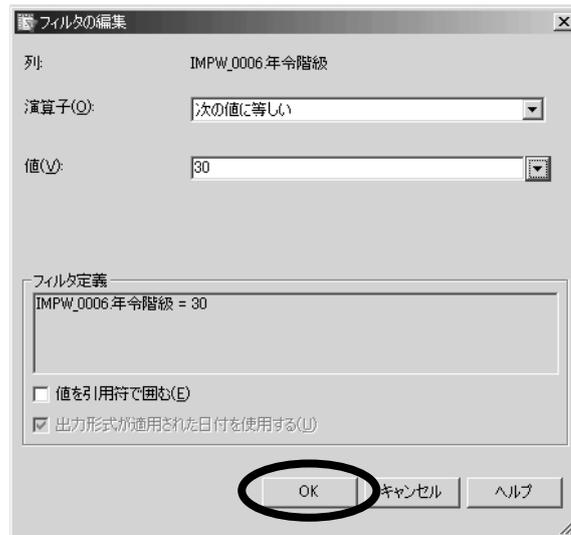
(e) 新しいダイアログが更に表示されるので、その下部にある「値の取得」ボタンをクリックする。



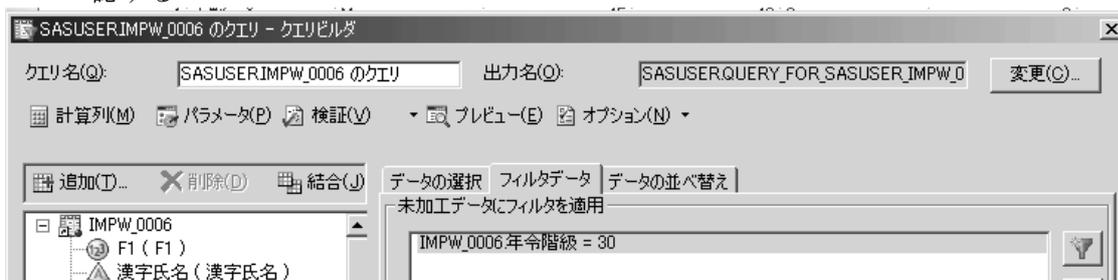
(f) グループ分けに用いる変数の値やカテゴリーのリストが表示されるので、まず、2つのグループの一方の値またはカテゴリーをクリックする。



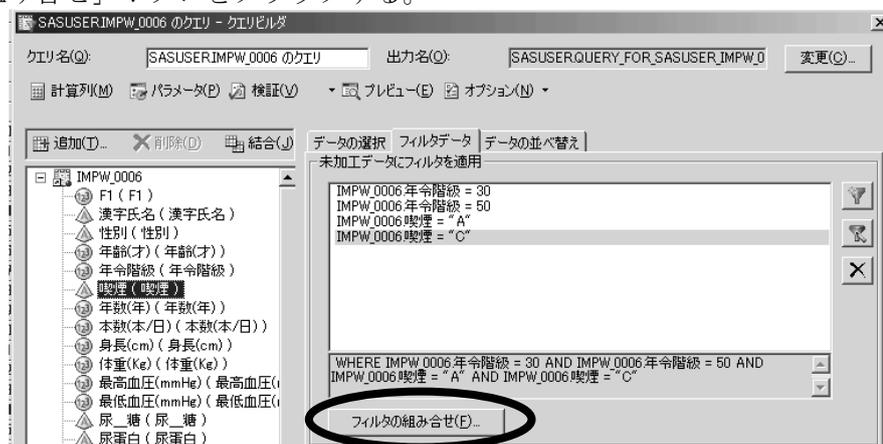
- (g) 「フィルタの編集」のダイアログに戻るので、「値」の右に、選択した値またはカテゴリーが表示されていることを確認し、下の「OK」ボタンをクリックする。



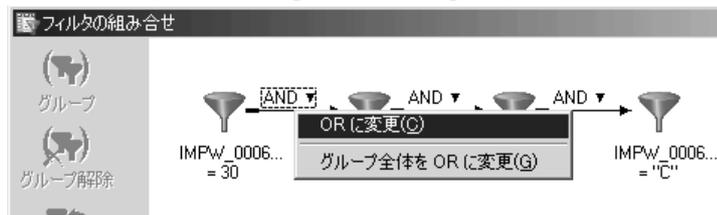
- (h) 最初のダイアログに戻り、選択した値またはカテゴリーが表示されていることを確認する



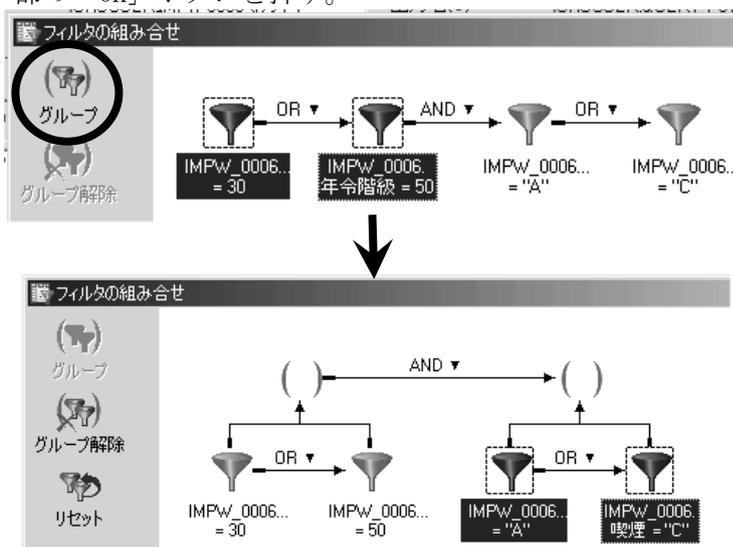
- (i) (c)～(h)の操作を繰り返し、2つのグループのもう一方の値またはカテゴリーを設定する。
 (j) もし、比較する変数もカテゴリーを限定する場合は、(c)～(h)の操作を繰り返し、分析に用いるカテゴリーを設定する。
 (k) 2つのグループの値またはカテゴリー、および、比較する変数も限定した場合は、それらの値やカテゴリーが選択されていることを確認した後、その下の「フィルタの組み合わせ」ボタンをクリックする。



- (l) 「フィルタの組み合わせ」というダイアログが新たに開くので、同じ変数の2つの漏斗の絵の間にある「AND ▼」をクリックし、その下に表示されたメニューから「ORに変更」をクリックし、「AND ▼」が「OR ▼」に変わったことを確認する。



- (m) 「Ctrl」(コントロール)キーを押しながら、同じ変数の漏斗をクリックして色を反転させた後、左端上の「グループ」をクリックし、同じ変数の漏斗を括弧でまとめた後、最下部の「OK」ボタンを押す。



- (n) 最初のダイアログに戻るので「実行」ボタンをクリックする。
 (o) 画面上に、グループ分けに使用する変数と、比較する変数のみの表が、新たに表示されたことを確認する。

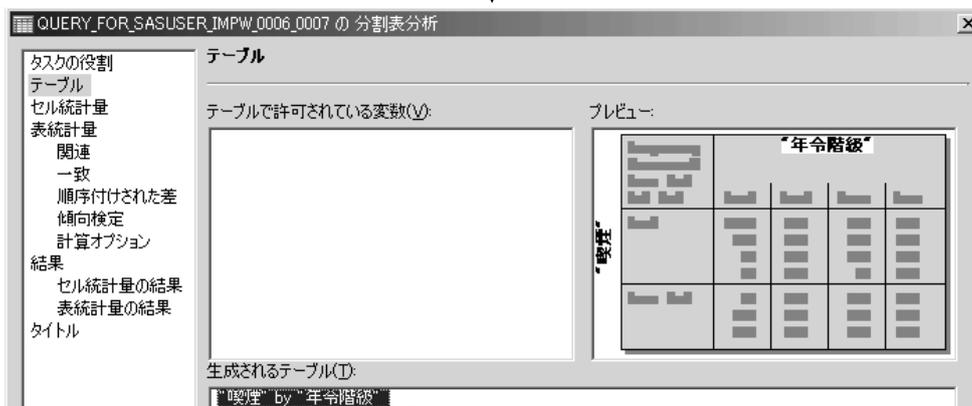
- 3) 上部メニュー内の [記述統計] をクリックし、その下に表示されるメニュー内の [分割表分析] をクリックする。



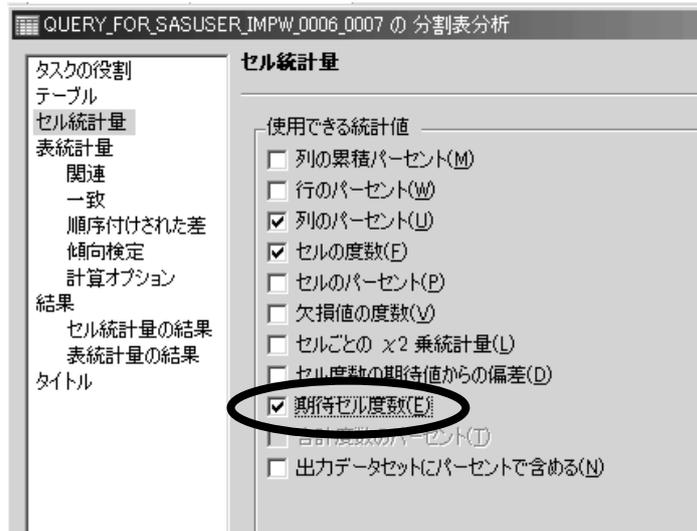
- 4) 分割表分析のダイアログが開くので、「変数リスト」からグループ分けに用いる変数と比較する変数をそれぞれクリックした後、それぞれで、その右の右向き矢印のボタンをクリックし、新たに表示されたメニューから「表変数」をクリックし、「タスクの役割」の下「表変数」の下に、グループ分けに用いる変数と比較する変数の両方が表示されたことを確認する。



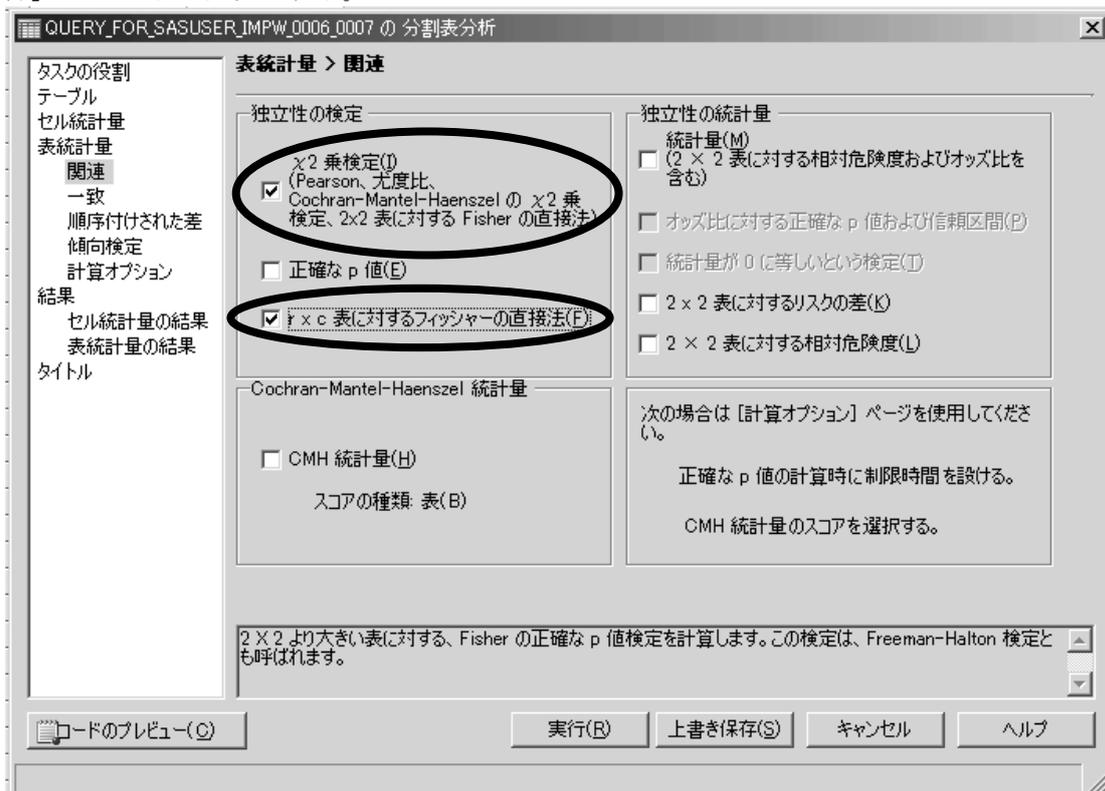
- 5) ダイアログの左端のリストから「テーブル」をクリックし、新たに右に表示された「テーブルで許可されている変数」の下リストから、最初にグループ分けに用いる変数をその右の「プレビュー」の下領域にドラッグアンドドロップし、続いて、比較する変数を同じ領域にドラッグアンドドロップする。その結果、「プレビュー」の下で、上にグループ分けに用いる変数、横に比較する変数が表示されていることを確認する。更に、その下の「生成されるテーブル」の下に「(比較する変数名) by (グループ分けに用いる変数名)」と記載されていることも確認する。



- 6) ダイアログの左端のリストから「セル統計量」をクリックし、新たに右に表示された「使用できる統計値」の下で、既にチェックの入っているものはそのままにして、「期待セル度数」の前のボックスにチェックを入れる。



- 7) ダイアログの左端のリストから「表統計量」の下の「関連」をクリックし、新たに右に表示された「独立性の検定」の下で、「 χ^2 乗検定 (Pearson、.....)」の前のボックスと、「 $r \times c$ 表に対するフィッシャーの直接法」の前のボックスにチェックを入れ、最下部の「実行」ボタンをクリックする。



8) 分析結果が表示されるので、まず、下図のような「FREQ プロシジャ」と書かれた下の表を見る。その表の各セルで、真ん中の「期待度数」を全て確認し、下記の条件に合うものがあれば、その指示に従う。

(a) 2×2表(分割表の行も列も2個のもの)で、どれかのセルの期待度数が5未満の場合、10)に進み、そこに記載された方法で結果を判定する。

(b) 2×2表以外の場合で、セルの期待度数が以下のどちらかの条件に当てはまる場合

(1) セルのうちで20%の期待度数が5未満

(2) セルの中に、1つでも期待度数が0のものがある

どちらかの条件に当てはまる場合は、11)に進み、そこに記載された方法で分析する。

上記の(a)(b)どちらの条件にも当てはまらない場合は、9)に進み、そこに記載された方法で結果を判定する。

FREQ プロシジャ

		表：喫煙 * 年令階級		
		年令階級(年令階級)		合計
喫煙(喫煙)	30	50		
	A	259 233.33 82.75	192 217.67 65.75	451
C	54 79.673 17.25	100 74.327 34.25	154	
合計	313	292	605	

度数
期待度数
列のパーセント

9) 8)の分割表のすぐ下にある下図のような表を見て、「カイ2乗値」の行の「p値」を確認し、以下の方法で判定する。

統計量	自由度	値	p値
カイ2乗値	1	22.9925	<.0001
尤度比カイ2乗値	1	23.2135	<.0001
連続性補正カイ2乗値	1	22.1056	<.0001
Mantel-Haenszelのカイ2乗値	1	22.9545	<.0001
ファイ係数		0.1949	
一致係数		0.1913	
CramerのV統計量		0.1949	

この値は「2つのグループ間で、比較したデータの割合に差が無い」ことが起こりうる確率の値である。従って、この値が**5%(0.05)未満**の場合は、帰無仮説を棄却できる(否定できる)ため、5%の有意水準で「比較したデータ(の割合)は2つのグループの間で有意な差がある」とみなすことができる。

この値が**5%(0.05)以上**の場合は、帰無仮説を棄却できない(否定できない)ため、2つのグループの間で比較したデータの割合に差があるかどうかの判定は「保留」となり、「2つのグループの間で比較したデータ(の割合)に有意な差があるとはいえない」あるいは「2つのグループの間で比較したデータ(の割合)に有意な差は見られなかった」という結論になる。決して「2つのグループ間で比較したデータ(の割合)に差が無い」と

いう結論は出せない。これは、危険率が「帰無仮説が正しいのに、間違っ
て棄却してしま
う確率」であるためである。また現実的にも、サンプル数を増加させた場合や他の検
定方法を使用した場合、有意差が検出される可能性があるためである。

10) 2×2表の分析で、4つのセルのうち1つでも期待度数が5未満のものがある場合は、Fisherの直接確率法(正確確率法, Fisher's exact probability test)を使用する必要がある。このため、9)のカイ2乗検定のすぐ下にある下図のような「Fisherの正確検定」の表の「両側 Pr<=P」右横の値を確認する。

この値は「2つのグループ間で、比較したデータの割合に差が無い」ことが起こりうる確率の値である。従って、この値が5%(0.05)未満の場合は、帰無仮説を棄却できる(否定できる)ため、5%の有意水準で「比較したデータ(の割合)は2つのグループの間で有意な差がある」とみなすことができる。

この値が5%(0.05)以上の場合は、帰無仮説を棄却できない(否定できない)ため、2つのグループの間で比較したデータの割合に差があるかどうかの判定は「保留」となり、「2つのグループの間で比較したデータ(の割合)に有意な差があるとはいえない」あるいは「2つのグループの間で比較したデータ(の割合)に有意な差は見られなかった」という結論になる。決して「2つのグループ間で比較したデータ(の割合)に差が無い」という結論は出せない。これは、危険率が「帰無仮説が正しいのに、間違っ
て棄却してしま
う確率」であるためである。また現実的にも、サンプル数を増加させた場合や他の検
定方法を使用した場合、有意差が検出される可能性があるためである。

Fisherの正確検定	
セル (1,1) 度数 (F)	259
左側 Pr <= F	1.0000
右側 Pr >= F	1.167E-06
表の確率 (P)	7.165E-07
両側 Pr <= P	1.720E-06

(注; 上記の表の危険率のように「1.720E-06」という数値は、「 1.720×10^{-6} 」を意味している。すなわち、「 $○○E-X$ 」という数値は、「 $○○ \times 10^{-x}$ 」のことである。

11) 2×2表以外の分析では、下記のような場合、 χ^2 検定(カイ二乗検定)はできない可能性がある。

(1) セルのうち20%の期待度数が5未満

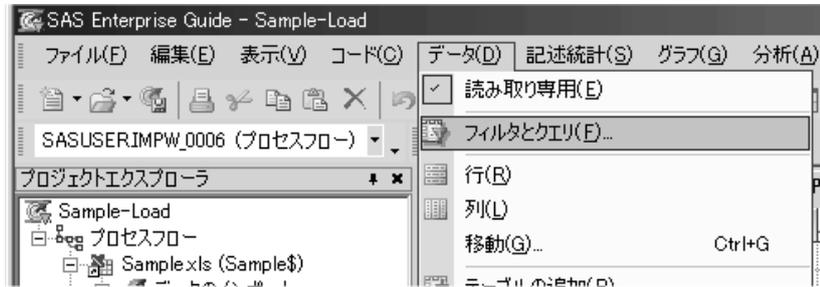
(2) セルの中に、1つでも期待度数が0のものがある

このような場合は、下記のような操作を行う。

- データの再分類を行い、期待度数の小さなセルを無いようにして、再度、 χ^2 検定(カイ二乗検定)を行う。
- データの再分類ができない場合は、 χ^2 検定(カイ二乗検定)の結果の解釈や採用するかどうかを慎重に検討(統計の専門家に相談するのも良い)し、場合によっては分析を中止する。

(7) 対応のあるデータの差が正規分布しているかどうかを判定する

- 1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
- 2) 上部メニュー内の [データ] をクリックし、その下に表示されるメニュー内の [フィルタとクエリ] をクリックする。



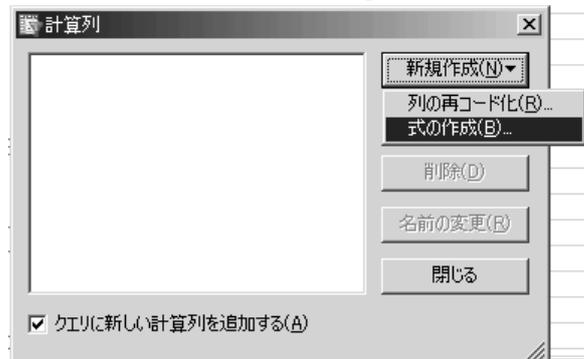
3) 新しいウィンドウが開くので、左側の変数のリストから対応のあるデータそれぞれをダブルクリックし、右側の枠内にその変数が表示されていることを確認する。



4) 左上にある「計算列」をクリックする。

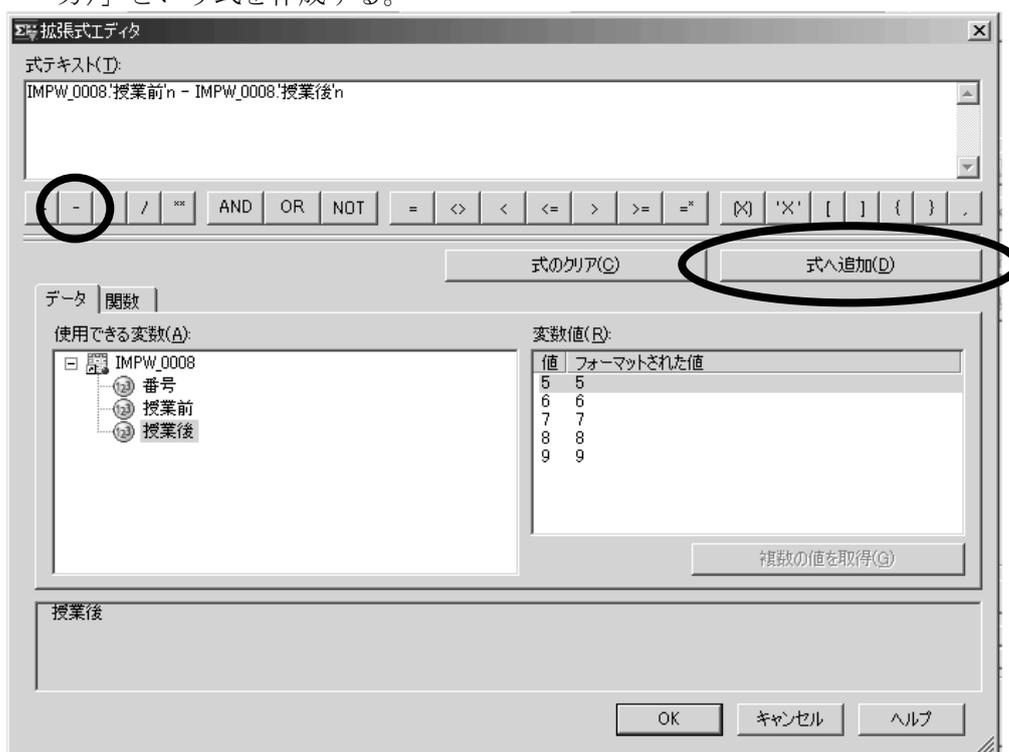


5) 「計算列」というダイアログが開くので、右上の「新規作成」ボタンをクリックし、その下に新たに表示されたメニュー内の「式の作成」をクリックする。

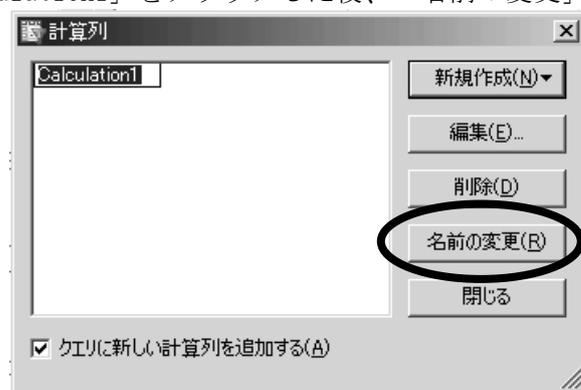


6) 「拡張式エディタ」というダイアログが開くので、左中央の「使用できる変数」の下のリストから、対応のあるデータの一方をクリックした後、右端の「式へ追加」ボタンをクリックして、上の「式テキスト」の下のボックス内にその変数の名前が表示されることを確認する。次に、その下のボタンの中で「-」（マイナス）ボタンをクリックし、「式テキスト」に表示されることを確認する。最後に、対応のあるデータのもう一方をクリックした後、右端の「式へ追加」ボタンをクリックして、上の「式テキスト」の下のボックス内にその変数の名前が表示されることを確認する。式が間違っていないことを確認した後、最下部の「OK」ボタンをクリックする。

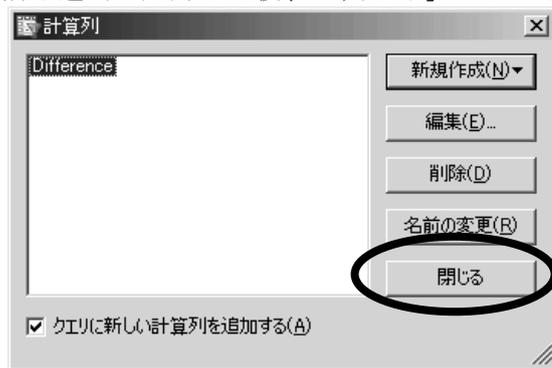
「式テキスト」には、「(対応のあるデータの一方) - (対応のあるデータのもう一方)」という式を作成する。



7) 「拡張式エディタ」が閉じて、「計算式」ダイアログに戻るので、左のボックス内に表示されている「Calculation1」をクリックした後、「名前の変更」ボタンをクリックする。



8) 変換後の変数の名前を適当に入力した後、「閉じる」ボタンをクリックする。



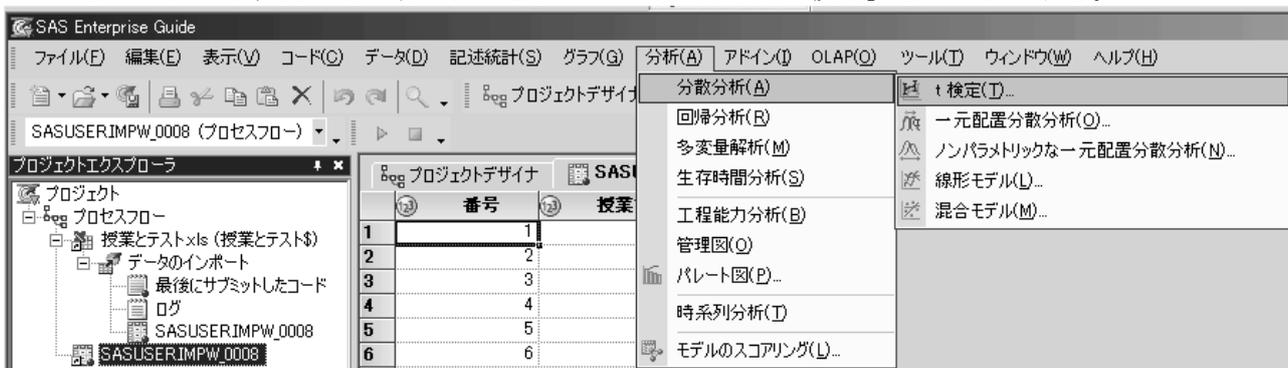
9) 新しい変数名が、グループ分けに用いる変数の下に表示されていることを確認した後、「実行」ボタンをクリックする。

10) 対応のある2つのデータとその差の表が表示されるので、この表を用いて、配布資料「SAS (SAS Enterprise Guide) 使用方法 (1 因子(1 種類のデータ)の概要を見る)」の「(1) ヒストグラムおよび、累積度数分布の確率プロットの表示」の操作を行い、対応のある2つのデータの差の値が正規分布しているかどうかを判定する。

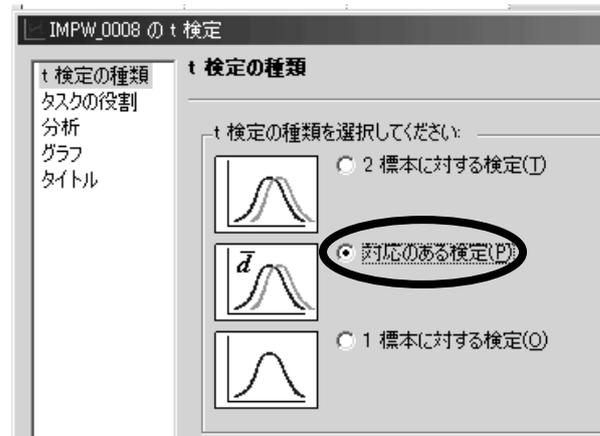
もし、正規分布していた場合は、この資料の「(8) 対応のある t 検定 (Paired t test) を行う」の操作を行い、2つのグループ間で比較する変数に差があるかどうかを判定する。もし、正規分布していない場合は、配布資料「統計学基礎(3) 2つのグループの1因子(1種類のデータ)の差を検定する」の(20)以降を参照する。

(8) 対応のある t 検定 (Paired t test) を行う

- 1) SAS Enterprise Guide を起動し、使用するデータの表が表示されている状態まで進める。
- 2) 上部メニュー内の [分析] をクリックし、その下に表示されるメニュー内の [分散分析] をクリックし、更にその右に表示されたメニュー内の「t 検定」をクリックする。



2) t 検定のダイアログが開くので、「t 検定の種類を選択してください」の下の「対応のある検定」の前の丸をクリックしてマークを付ける。



3) 左端のリストで「タスクの役割」をクリックし、右側に新たに表示された「変数リスト」から、対応のあるデータの一方をクリックした後、その右にある右向き矢印のボタンをクリックし、新たに表示されたメニューから「対応のある変数」をクリックする。更に、対応のあるデータのもう一方をクリックした後、その右にある右向き矢印のボタンをクリックし、新たに表示されたメニューから「対応のある変数」をクリックして、2つの対応のあるデータが、「タスクの役割」の下の「対応のある変数」の下に表示されていることを確認して、最後に最下部の「実行」ボタンをクリックする。



4) 分析結果が表示されるので、表示されている表の中から、下図のような「t 検定」と書かれた表を見る。その表中から、「Pr>|t|」の下の数値を確認する。この値が、「対応のある t 検定の危険率」（2つのグループの間で比較した変数に差が無い）ことが起こりうる確率の値)なので、この値が **5% (0.05) 未満** の場合は、帰無仮説を棄却できる(否定できる)ため、5%の有意水準で「**比較した変数は2つのグループの間に有意な差がある**」とみなすことができる。

これに対し、この値が **5% (0.05) 以上** の場合は、帰無仮説を棄却できない(否定できない)ため、2つのグループの間に差があるかどうかの判定は「**保留**」となり、「**2つのグループの間で比較した変数に有意な差があるとはいえない**」あるいは「**2つのグループの間で比較した変数に有意な差は見られなかった**」という結論になる。決して「2つのグループ間で比較した変数に**差が無い**」という**結論は出せない**。これは、危険率が「帰無仮説が正しいのに、間違っ棄却してしまう確率」であるためである。また現実的にも、サンプル数を増加させた場合や他の検定方法を使用した場合、有意差が検出される可能性があるためである。

t 検定			
差	自由度	t 値	Pr > t
授業前 - 授業後	9	-11.76	<.0001

(9) Wilcoxon 符号順位和検定(Wilcoxon signed-ranks test)を行う

1) この資料の「(7) 対応のあるデータの差が正規分布しているかどうかを判定する」を参照して、「対応のある2つのデータの差」の変数を作成し、その値が表示されている表を画面に出す。

2) 上部メニュー内の [記述統計] をクリックし、その下に表示されるメニュー内の [分布] をクリックする。



- 3) 新しいウィンドウが開くので、「対応のある2つのデータの差」を「変数リスト」の下
のリストから選んでクリックする。そして、その右隣りにある右向き矢印のボタンをクリッ
クし、新たに表示されたメニューから「分析変数」をクリックし、右の「タスクの役割」
の下「分析変数」の下に、選択した変数名が表示されていることを確認した後、最下部
の「実行」ボタンをクリックする。



- 4) 分析結果が表示されるので、表示されている表の中から、下図のような「位置の検定」
と書かれた表を見る。その表中から、「符号付順位検定」の行の「p 値」の下の数値を確認
する。この値が、「Wilcoxon 符号順位和検定の危険率」（2つのグループの間で比較した変
数に差が無い）ことが起こりうる確率の値)なので、この値が **5% (0.05) 未満** の場合は、帰
無仮説を棄却できる(否定できる)ため、5%の有意水準で「**比較した変数は2つのグルー
プの間で有意な差がある**」とみなすことができる。

これに対し、この値が **5% (0.05) 以上** の場合は、帰無仮説を棄却できない(否定できな
い)ため、2つのグループの間に差があるかどうかの判定は「**保留**」となり、「**2つのグ
ループの間で比較した変数に有意な差があるとはいえない**」あるいは「**2つのグループ
の間で比較した変数に有意な差は見られなかった**」という結論になる。決して「2つのグルー
プ間で比較した変数に**差が無い**」という結論は出せない。これは、危険率が「帰無仮説が
正しいのに、間違っ棄却してしまう確率」であるためである。また現実的にも、サンプ
ル数を増加させた場合や他の検定方法を使用した場合、有意差が検出される可能性がある
ためである。

位置の検定 H0: Mu0=0				
検定		統計量		p 値
Studentの t 検定	t	-11.7589	Pr > t	<.0001
符号検定	M	-5	Pr >= M	0.0020
符号付順位検定	S	-27.5	Pr >= S	0.0020