

CONTENTS

- 平均や比率に関する推定と検定をSASで行なう方法
- 10 Q&A
- 13 新刊マニュアルのお知らせ
- 14 SASトレーニングのお知らせ
- 15 最新リリース情報
- 15 SAS Technical News送付についてのご案内

特集

平均や比率に関する 推定と検定を SASで行なう方法



1. はじめに

手元に得られたデータをもとに分析を行なうとき、そのデータの概要をSASで探ることはこれまでのSAS Technical Newsでご紹介してきました。今号の特集では、SAS/STATソフトウェアのTTEST、およびFREQプロシジャを使用して、母平均や母比率に関する推定と検定を行なう方法をご紹介します。

2. 平均に対する推定と検定

分析を目的として標本データを観測しているとき、その値は大きく分けて量的な場合と質的な場合があります。たとえば、温度や金額、100点を満点としたテストの点数、血圧、身長などは、一般に量的な値と考えられます。量的なデータに対しては、その平均に関心を持つことが多々あります。このとき、母集団が正規分布に従うことを想定した上で、その平均について調べる、言い換えると平均に関する推定や検定を行なうことがあります。平均に関する検定は、いわゆるt分布に従う統計量を計算することから、t検定と総称されることがあります。また、ある場合によっては、t分布について初期の考察を行なったスチューデント(Student、これはウィリアム・ゴセットの論文投稿時のペンネーム)にちなんで、スチューデントのt検定と呼ばれるときもあります。

2.1 1標本の場合

次のような状況を考えてみましょう。A県の電気屋をランダムに20店舗選び、ある電化製品Xの価格を調べた結果が以下のようになったとします。

3180, 3599, 3280, 2980, 3500, 3099, 3200, 2980, 3380, 3780, 3199, 2979, 3680, 2780, 2950, 3180, 3200, 3100, 3780, 3200 (単位は円)

この商品の平均価格が3000円と異なるか、有意水準をα=0.10とした 1標本のt検定を行なってみましょう。ここでの帰無仮説H0と対立仮説H1は、 以下のように設定します。

H0: A県における電化製品Xの平均価格は3000円である H1: A県における電化製品Xの平均価格は3000円ではない

SASプログラムを扱った書籍には、この検定はBase/SASのUNIVARIATEプロシジャで対応していると記載されているものがありますが、現在ではSAS/STATのTTESTプロシジャでも実行できます。TTESTプロシジャで読み込むSASデータセットでは、分析で使用する変数が1つ用意されていれば十分です。ここでは、販売価格を表すPriceという変数ただ1つからなるデータセットSample1を作成し、TTESTプロシジャで読み込んで検定を行なうプログラムを書いてみましょう。

1標本の平均に関する検定のプログラム 1

```
DATA Sample1; /*サンプルデータの作成 */
INPUT Price @@; /* @@で1行から複数のデータを読み込む */
DATALINES;
3180 3599 3280 2980 3500 3099 3200 2980 3380
3780 3199 2979 3680 2780 2950 3180 3200 3100
3780 3200
;
RUN; /* TTESTプロシジャの実行 */
PROC TTEST DATA=Sample1 H0=3000 ALPHA=0.10; /* 分析する変数名をVARステートメントで指定する */
VAR Price;
RUN;
```

PROC TTESTステートメントでは、DATA=オプションで分析に使用する SASデータセット名を指定しています。これを省略すると、TTESTプロシジャ は直近に作成されたデータセットを読み込みます。昔の書籍では、DATA=オプションを省略しているプログラムが記載されていたりしますが、想定外の問題を避けるため、データセット名の指定は必ず行なっておきましょう。

このステートメントでは、他にも2つのオプションが与えられています。H0=オプションでは、帰無仮説で設定した数値(ここでは3000)を指定します。もしこのオプションを省略するとH0=0と指定されたものとして解析が行なわれるので、1標本の平均に対するt検定では、通常このオプションを指定する形になるでしょう。なおH0=オプションは、後でご紹介する「対応のあるt検定」や「2標本のt検定」でも使用できます。ALPHA=オプションは、信頼区間のための信頼水準を設定するものです。ここでは、0.10と指定したことから、1-0.10=0.90、すなわち母平均に対する90パーセント信頼区間が算出されます。

TTESTプロシジャの出力

TTEST プロシジャ

統計量

平均の平均の変数N 信頼限界の下限平均 信頼限界の上限Price203141.13251.33361.5

統計量

標準偏差の 標準偏差の 変数 信頼限界の下限 標準偏差 信頼限界の上限 標準誤差 Price 226.33 285.08 390.67 63.745

統計量

 変数
 最小値
 最大値

 Price
 2780
 3780

t 検定

変数 自由度 t 値 Pr > |t|
Price 19 3.94 0.0009

この出力では、「統計量」と「t検定」という2つのテーブルが現れています。 最初の「統計量」テーブルでは、標本から得られた統計量と、それらに基づく母集団パラメータの推定値が得られています。出力されている「平均」は、標本平均であり、かつ母平均の推定値でもあります。また、2つの「平均の信頼限界」は、信頼水準を90%とした母平均に対する区間推定の両端を表しています。2つ目の「t検定」テーブルでは、名前のとおり平均に関するt検定の結果が出力されています。t値は3.94であり、自由度19のt分布をもとに、0.0009というp値が計算されています。このp値は有意水準0.10と比べるとより小さいことから、この検定では帰無仮説H0は棄却され、A県における電化製品Xの価格は3000円ではないことが強く示唆されることになります。

実は、標本のおおきさ(N)、標本平均(MEAN)、および標本標準偏差(STD)という3つの要約統計量がわかっていれば、上記の検定を行なうことができます。これら3つの統計量を含むSASデータセットを用意しておくと、TTESTプロシジャはそれを認識し、検定の結果を出力することができます。なお、データセット中では統計量を表すキーワードを変数_STAT_、または_TYPE_で与えておく必要があります。

1標本の平均に関する検定のプログラム 2

```
/* 要約統計量からなるSASデータセットを作成 */
DATA Stat for Sample1:
  INPUT STAT $ Price;
DATALINES;
N
      20
MEAN
      3251.3
      285.08
STD
RUN;
                   /* TTESTプロシジャの構文は同じ */
PROC TTEST DATA=Stat_for_Sample1 H0=3000
ALPHA=0.10:
  VAR Price:
RIIN:
```

出力結果を確認してみましょう。要約されたデータを分析に使用したため、 前のプログラムでは出力されていた最大値などは得られていませんが、本 質的に同じ結果となっていることがわかります。

Base SASのUNIVARIATEプロシジャを使用して同じ検定を行なうときには、 以下のようにプログラムを記述します。

UNIVARIATEプロシジャによるプログラム例

/* 要約統計量からなるSASデータセットを作成 */
PROC UNIVARIATE DATA=Sample1 MU0=3000;
VAR Price;
RUN;

前記のTTESTプロシジャとほぼ同様の指定となりますが、帰無仮説における数値を指定するオプションがMU0=となっています。また、TTESTプロシジャとは異なり、要約統計量を読み込んで検定を行なう機能はありません。

UNIVARIATEプロシジャによる出力例の一部

位置の検定 HO: MuO=3000				
検定		-統計量	p 値	Ī
Student の t 検定	t	3.942259	Pr > t	0.0009
符号検定	M	5	Pr >= M	0.0414
符号付順位検定	s	82	Pr >= S	0.0012

UNIVARIATEプロシジャの出力は多岐にわたりますが、「位置の検定」という項目にt検定の結果が現れます。出力される統計量の桁数が異なるものの、本質的に全く同じ結果が得られています。UNIVARIATEプロシジャにおける位置の検定では、他にもノンパラメトリックな手法である符号検定(signed test)と、(Wilcoxonの)符号付順位検定(signed rank test)の結果もあわせて出力されます。

なお、母平均=0の検定に限れば、他の様々なプロシジャ、たとえば MEANSプロシジャなどを用いてその結果を出力することができます。

2.2 対応のある2標本の場合

この節では、いわゆる「対応のある」t検定をSASで行なう方法を紹介します。ここで、「対応のある」データとは、何らかの意味で「対」として考えるべき数値のペアがあることを指します。たとえば、ある人にある画家の絵を見せる前と後のその画家が好きであるかを得点化しておくと、その2つの得点はペアとして考えることができます。もう少し緩やかに、同じような属性、たとえば年齢や性別が同じである2人をペアとして考え、異なった薬を投与し、反応の違いに差を見るといったことを考えることもあります。

対応のあるt検定での統計的な計算そのものは、前述の1標本のt検定と全く同じで、TTESTプロシジャを使用して実行できます。ただし、用意するSASデータセットの形状は異なり、1つのオブザベーションにおいて必ず対となる2つの観測値が存在しなくてはなりません。

次の例は、10人を対象としてある薬を使用した前と後で血圧の変化があるかを検討することを想定しています。有意水準α=0.05で検定をしてみましょう。SASデータセットSample2には2つの変数があり、変数Beforeは使用前、一方Afterは使用後の血圧を表しています。

対応のあるt検定のプログラム

DATA Sample2; /* サンプルデータの作成 */
INPUT Before After;
DATALINES;
122 128
142 146
162 161

133 129
155 155
130 128
136 140
145 150
137 140
134 138
;
RUN;
/* TTESTプロシジャによる対応のあるも検定 */
PROC TTEST DATA=Sample2;
/* PAIREDステートメントで2つの変数を指定 */
PAIRED After*Before;
RUN;

対応のあるt検定では、対となる2つの変数をPAIREDステートメントで指定します。このとき、アスタリスク "*" の前後にそれらの変数を与えますが、前の変数値から後の変数値を引いた結果を分析では利用します。

UNIVARIATEプロシジャによる出力例の一部

	TTEST プロ	コシジャ		
	統計量	<u>.</u>		
		均の	平均の	
差	N 信頼限界の	下限 平均	信頼限界の上限	
After - Before	10 -0.	519 1.9	4.3189	
	統計量			
	標準偏差の		標準偏差の	
差	信頼限界の下限	標準偏差	信頼限界の上限	
After - Before	2.3258	3.3813	6.173	
	統計量	ŧ		
差	標準誤差	最小值	最大值	
After - Before	1.0693	-4	6	
t 検定				
差	自由度	t値	Pr > t	
After - Before	9	1.78	0.1093	

1標本のt検定と同じ体裁で結果が出力されています。t検定におけるp値が0.05より大きいことから帰無仮説は棄却されず、この薬の摂取前と後では血圧に変化がある、とまでは言えない、という結論にいたります。

なお、もしUNIVARIATEプロシジャで対応のあるt検定を行なう場合には、 DATAステップであらかじめペアごとに差を計算しておいてから、 UNIVARIATEプロシジャを実行してください。

UNIVARIATEプロシジャを利用した対応のあるt検定の実行例

DATA Sample2_Diff;
SET Sample2;
Difference=Before-After; /* 差を計算する*/
RUN;
/* UNIVARIATEプロシジャ*/
PROC UNIVARIATE DATA=Sample2_Diff;
VAR Difference;
RUN;

2.3 独立した2標本の場合

2つの母集団間に統計学的に何らかの関連や対応が存在しないと思われる場合には、それらからそれぞれ標本をとり、群間に平均の差が存在するかを調べるいわゆる2標本のt検定は、同じくTTESTプロシジャで実現できます。

ある教育内容について、2つの異なった教え方で大学生に対して行なうことを考えてみましょう。1つはこれまでと同じ方法 (A) で、もう1つは新たに考案された方法 (B) です。この2つの教え方に違いがあるかを、学習後に100点を満点とした達成度テストを行ない、その結果から検討するものとします。そこで、教授方法 (A) を行なった後でのテスト結果の平均を μ A、また (B) に対する平均を μ Bとおきます。この平均差の検定における帰無仮説は2つの教授方法に差はない、対立仮説として2つの教授方法には差があると設定し、有意水準 α =0.05で検定を行ないましょう。

帰無仮説 H0: $\mu A = \mu B$ 対立仮説 H1: $\mu A \neq \mu B$

実際に得られたテスト結果は、以下のようになりました。

教授方法Aを行なった後のグループAのテスト結果 66, 64, 63, 74, 60, 58, 75, 48, 80, 75, 51, 48, 69, 45, 75

教授方法Bを行なった後のグループBのテスト結果 84, 90, 67, 91, 72, 68, 86, 50, 61, 65, 96, 89, 83, 78, 71, 65

分析で使用するSASデータセットでは、測定したテストの点数と、2つの教授方法を指し示す変数を用意します。ここでは、それぞれScoreとMethodという名前としました。なお、ここではデータの形状を見るために、BOXPLOTプロシジャで各教授方法に対するスコアの分布を箱ひげ図で確認しています。

独立2標本の平均差の検定 1

DATA Sample3; /* SASデータセットSample3の作成 */
INPUT Method\$ Score @@;
DATALINES;
A 66 A 64 A 63 A 74 A 60 A 58 A 75 A 48 A 80 A 75
A 51 A 48 A 69 A 45 A 75
B 84 B 90 B 67 B 91 B 72 B 68 B 86 B 50 B 61 B 65 B 96
B 89 B 83 B 78 B 71 B 65
;
RUN;
/* BOXPLOTプロシジャで箱ひげ図 */
PROC BOXPLOT DATA=Sample3;

PLOT Score*Method / BOXSTYLE=SCHEMATIC; RUN;

/* TTESTプロシジャ */

PROC TTEST DATA=Sample3;

/* 群を識別する変数をCLASSステートメントで指定 */

CLASS Method;

VAR Score; /* 分析変数をVARステートメントで指定 */RUN:

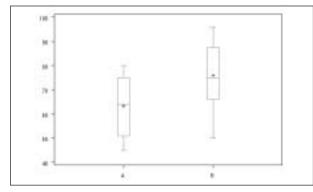


図1 BOXPLOTプロシジャによる箱ひげ図

+ で表示されている平均と、箱の中央部に存在する中央値を表す横線が 近く、また上下の箱の大きさがそれほど異ならないことから、各群の標本分 布はおおむね対称と考えられ、また上下に外れた点---いわゆる外れ値---も存在していないことがわかります。

TTESTプロシジャの結果は、この場合には「統計量」「t検定」「等分散性」という3つの項目に分かれています。順に見ていきましょう。

TTESTプロシジャの出力「統計量」テーブル

		¥	充計量			
				ম	平均の	
変数	Meth	od	N	信頼限界の)下限	平均
Score		A	15	57	.055	63.4
Score		В	16	69	.071	76
Score	Diff (1-	2)		-2	1.63	-12.6
		¥	充計量			
			平均の		平均の	
変数	Method	信 頓限 9	界の上限		-	標準偏差
Score	A		69.745		.3878	11.457
Score	В		82.929		.6051	13.003
Score	Diff(1-2)		-3.573		.7804	12.281
	(,			-		
		¥	充計量			
		標準	偏差の			
変数	Method	信頼限界	の上限	標準誤差	最小值	最大值
Score	A	1	8.068	2.9581	45	80
Score	В	2	0.124	3.2506	50	96
Score	Diff(1-2)	1	6.509	4.4136		

2つの群AとBそれぞれに対して、標本に対する様々な統計量が最初に出力されます。また、2標本間の差に関する統計量もあわせて算出されます。

TTESTプロシジャの出力 「t検定 | テーブル

t 検定 変数 手法 分散 自由度 t 値 Pr > |t|Score Pooled Equal 29 -2.85 0.0079 Satterthwaite 0.0077 Score Unequal 28.9 -2.87

2番目に現れる「t検定」のテーブルでは、名前のとおり独立2標本間の平均差に関するt検定の結果が出力されています。このとき、等分散性を仮定した場合のt検定が1行目に、また分散が異なることを前提としたいわゆるウエルチ(Welch)検定が2行目に出力されます。どちらの検定結果を見るべきかという問題は、次の「等分散性の検定」の結果をもとに判断することもあります。

TTESTプロシジャの出力「等分散性」テーブル

等分散性					
変数	手法	分子の自由度	分母の自由度	F 值	Pr > F
Score	Folded F	15	14	1.29	0.6412

「等分散性」の項では、2つの群における分散に違いがあると言えるかどうかを調べる検定を行なっています。この検定---TTESTプロシジャではFolded F検定と呼ばれています---におけるp値は0.6412と計算されていることから、等分散性を積極的に否定することはできません。したがって、通常はひとまず等分散であることを是認することになります。

なお、1標本のt検定と同じように、標本に関する要約統計量がわかっていれば平均差の検定も行なうことができます。群を表す変数を用意する必要がある以外は同じ構成のSASデータセットを用意する形になります。

独立2標本の平均差の検定 2

```
DATA Stat_for_Sample3;
  INPUT Method$ _STAT_$ Score;
DATALINES:
A N
        15
A MEAN 63.4
A STD
        11.457
ви
        16
B MEAN 76
       13,003
B STD
;
RUN:
                    /* TTESTプロシジャの構文は同じ */
PROC TTEST DATA=Stat_for_Sample3;
  CLASS Method;
  VAR Score;
RUN;
```

2.4 TTESTプロシジャで利用できる他のステートメント、オプション BYステートメント、WHEREステートメント、およびFORMATステートメント

BYステートメント、WHEREステートメント、およびFORMATステートメント などは、TTESTプロシジャでも利用できます。その他に用意されている構 文としては、以下のものがあります。

- ・FREQステートメント、WEIGHTステートメント それぞれ、度数、および重みを表す変数を指定します。
- ・COCHRANオプション (PROC TTESTステートメント)
 独立2標本に対するt検定では、母分散が異なると考えられるときにコクランの方法 (Cochran、コクラン・コックスの方法、コクラン・コックス検定とも呼ばれるときがあります)で自由度を調整した検定があります。近年では前述のWelch検定が使用されることが多いのですが、COCHRANオプションを指定することによってこの検定の結果が出力されます。

· CI=オプション (PROC TTESTステートメント)

TTESTプロシジャの出力において、先頭の「統計量」テーブルには母標準偏差に対する信頼限界が表示されています。このCI=オプションを利用すると、信頼限界の計算方法を変更することができます。このオプションを指定しなかった場合にはCI=EQUALという指定に基づいて計算されます。これは、ALPHA=オプションで与えた数値をpとおくと、自由度カイ2乗分布の両端において確率がp/2ずつとなるようにして得られた数値を信頼限界として出力します。一方、CI=UMPUと与えると、一様最強力不偏検定に基づいた信頼限界が算出されます。CI=NONEと指定すると、母標準偏差に対する信頼限界は出力されません。

2.5 SAS® Enterprise Guide® 4.1での実現方法

ここでご紹介した検定は、いずれもEnterprise Guide 4.1 (以下EG4.1) のタスクとして存在します。メニューから、[分析] > [分散分析] > [検定] と選択することで、前記の3つの検定をいずれも実現することが可能です。

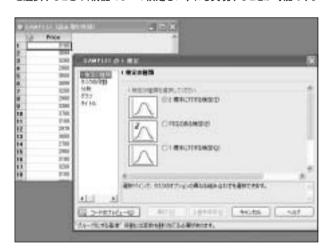


図2 t検定における画面

2.6 その他の話題

2006年に米国で開かれたSASユーザー会SUGI31では、次期リリース SAS 9.2におけるTTESTプロシジャの拡張点が開発担当者らによって紹介されています。

Like Wine, the TTEST Procedure Improves with Age http://www2.sas.com/proceedings/sugi31/208-31.pdf

その中では、SAS 9.2において以下のような機能が追加されるとのことです。

- · ODS Graphicsによるグラフ描画
- ・TOST (Two One-Sided Test) に基づく同等性の検定
- ・片側検定のp値と信頼区間の算出
- · AB/BAタイプの簡単なクロスオーバー計画に対する分析

上記以外にも、重要な拡張がいくつか計画されています。

3. 比率に関する推定と検定

本項以降では、「表と裏」、「YesとNo」など、観測する事象が高々2つまでである状況を考えます。このような場合には、それらの発生比率(割合)に関心を持つことがあります。1標本、または2標本の違いや、更にデータ内に対応が存在するか否かによって、一般に2項検定、マクネマー検定、またはピアソンのカイ2乗検定などを行なうことになります。この節で行なう検定では、いずれも有意水準を0.05として考えます。

3.1 1標本の場合

ここに一枚のコインがあるとします。このコインを投げ表か裏かを観測するとき、一般に表の出る確率は0.5と考えられますが、もしこのコインにゆがみがあるとその限りではありません。ここでは、帰無仮説H0と対立仮説H1をそれぞれ

H0: 表が出る確率は0.5である

H1: 表が出る確率は0.5ではない

と与え、コイン投げを20回行なっていわゆる2項検定を行ない、このコインにはゆがみがあると考えられるか、それともそれを示す明確な根拠はないかを調べてみましょう。分析用のSASデータセットでは、表をHead、裏をTailとし、変数Eventに格納しています。そのデータセットをFREQプロシジャで読み込み、分析を行ないます。

FREQプロシジャによる2項比率に関する分析プログラム 1

DATA Sample4; /* サンプルデータの作成 */
INPUT Event\$ @@;
DATALINES;
Head Tail Head Head Tail Tail Head Tail Tail
Tail Tail Head Head Tail Head Tail Head Tail
;
RUN;

PROC FREQ DATA=Sample4; /* FREQプロシジャ*/
TABLES Event /BINOMIAL(LEVEL="Head" P=0.5);
RUN;

TABLESステートメントでは、BINOMIALオプションを指定します。このプログラムでは、更にサブオプションとして括弧の中にLEVEL=とP=を与えています。LEVEL=サブオプションでは、2つの事象のうちで関心を持っている方の水準値を指定します。一方P=サブオプションは、帰無仮説における比率(割合)、ここでは0.5の数値を与えます。このプログラムを実行すると、以下のような出力が得られます。

FREQプロシジャの出力例

		FREQ プロシ	ジャ	
			累積	累積
Event	度数	パーセント	度数	パーセン
Head	8	40.00	8	40.00
Tail	12	60.00	20	100.00
	Event :	= Head の二項	頁分布の比率	
	 比率		0.4000	
	漸近標準	誤差	0.1095	
	95% 信頼	頁下限	0.1853	
	95% 信頼	頁上限	0.6147	
		正確な信頼限	!界	
	95% 1	言頼下限	0.1912	
	95% 1	言頼上限	0.6395	
	но: 母	比率 = 0.5(こ対する検定	
帰無仮詞	説が正しいもん	とでの漸近標準	誤差	0.1118
z			-0.8944	
片側 Pi	r < Z			0.1855
両側 P	r > z			0.3711
		標本サイズ =		

FREQプロシジャのアウトプットは、大きく3つのテーブルに分かれています。 1つ目に対しては項目名は特に与えられていませんが、いわゆる一元度数 表が出力されています。2つ目の「Event = Head の二項分布の比率」と いうテーブルには、母比率の推定値(標本における比率)、標準誤差、およ び正規分布に基づく漸近的な信頼限界、および正確分布に基づいた信 頼限界が出力されています。最後のテーブルには、項目名のとおり検定の 結果が表示されています。

正確分布に基づく2項検定の結果を得るためには、更にEXACTステートメントを追加してBINOMIALオプションを指定する必要があります。

EXACTステートメントの例

```
PROC FREQ DATA=Sample4;

TABLES Event /BINOMIAL(LEVEL="Head" P=0.5);

EXACT BINOMIAL; /* EXACTステートメントを追加 */
RUN;
```

出力結果の一部

HO: 母比率 = 0.5 に対する検定
帰無仮説が正しいもとでの漸近標準誤差 0.1118
z -0.8944
片側 Pr < Z 0.1855
両側 Pr > |z| 0.3711

正確検定
片側 Pr <= P 0.2517
両側 = 2 * 片側 0.5034

「正確検定」に関する出力が現れ、片側検定、および両側検定によるp値 の表示が追加されていることがわかります。

FREQプロシジャを利用しているときには、データがあらかじめ集計されていることがよくあります。このようなデータセットでは、水準、および度数を表す変数が対となって存在しており、FREQプロシジャはそのようなタイプのデータも受けつけます。他の分析プロシジャでは度数を表す変数はFREQステートメントで指定しますが、FREQプロシジャにはこのステートメントが存在せず、その代わりにWEIGHTステートメントを使用します。

FREQプロシジャによる2項比率に関する分析プログラム2

```
DATA Sample5; /* サンプルデータの作成 */
INPUT Event$ Frequency;
DATALINES;
Head 8
Tail 12
;
RUN;

PROC FREQ DATA=Sample5;
  TABLES Event /BINOMIAL(LEVEL="Head" P=0.5);
  EXACT BINOMIAL;
  WEIGHT Frequency; /* WEIGHTステートメント */
RUN;
```

前記のFREQプロシジャと全く同じ出力が得られることを確認してください。

3.2 対応のある2標本の場合

母平均に関する検定を行なったときと同じように、母比率の検定においても対応のある2標本に対する検定と、対応関係を考えない独立なときの検定が考えられます。データの中で対応があると考えられるときにはマクネマー(McNemar)検定が、また独立していると考えられるときにはピアソン(Pearson)のカイ2乗検定が一般的に行なわれます。

マクネマー検定を行なうためには、FREQプロシジャにおいて分割表の定義をTABLESステートメントで与え、AGREEオプションをあわせて指定します。以下の例では、FORMATプロシジャを使用して出力形式(フォーマット)を作成しておき、分析用のSASデータセットに適用しています。FREQプロシジャに限らず、分析においては水準の並びが重要な意味合いを持つことが多々あります。変数の値として文字データをそのまま与えておいたときには、並び順が意図と異なって扱われてしまう恐れがあります。このように出

カ形式を準備し、その一方で水準の並び順はわかり易くしておくと、水準の並び順の誤りを防ぐことができます。

マクネマー検定の例

```
PROC FORMAT; /* FORMATプロシジャで出力形式を作成 */
  VALUE Approve 1="支持する" 2="支持しない";
RUN:
DATA Sample6;
  INPUT Last_Year This_Year Freq;
  FORMAT Last_Year This_Year Approve.;
DATALINES:
1 1 33
1 2 13
2 1 19
2 2 35
RUN:
PROC FREQ DATA=Sample6;
                             /* AGREEオプション */
  TABLES Last_Year*This_Year / AGREE;
  EXACT MCNEM;
  WEIGHT Freq:
RUN:
```

FREQプロシジャの出力の一部

```
      Last_Year * This_Year の統計量

      McNemar の検定

      統計量 (S) 1.1250

      自由度 1
      事近的な Pr > S 0.2888

      正確な Pr >= S 0.3771
```

なお、上記プログラムのようにEXACTステートメントを使用すると、ここでも正確な分布に基づく検定の結果も出力されます。いずれの検定方法においても、帰無仮説を棄却するまでにはいたらないという結果が得られています。

なお、FREQプロシジャにおけるマクネマー検定における検定統計量の計算式では、いわゆる連続修正項が含まれています。書籍やソフトウェアによっては、この修正項を含まない数値を統計量として算出しているものもあることにご留意ください。

AGREEオプションからは、いわゆるカッパ係数も計算されます。また、分割表が2×2よりも大きな場合には、ボウカー (Bowker) の対称性検定が行なわれます。また、分割表の性質によって、その他にも統計量が算出されます。

3.3 独立2標本の場合

分割表に対するピアソンのカイ2乗検定は、一般に行と列の関連性を調べるときに使用されますが、2×2表に対しては独立な2標本の比率に関する検定であると考えることができます。ここでは、ある集団からランダムに男女それぞれ30人ずつを選び出し、答えが「はい」「いいえ」のいずれかである質問をすることを想定します。その集計結果は以下のとおりでした。

	はい	いいえ	合計
女性	18	12	30
男性	14	16	30
合計	32	28	60

このデータを使用して、有意水準0.05でピアソンのカイ2乗検定を行なってみましょう。FREQプロシジャのTABLESステートメントでは、CHISQオプションを指定します。

ピアソンのカイ2乗検定のプログラム例

```
/* サンプルデータの作成 */
DATA Sample7;
  INPUT Sex$ Answer$ Cnt;
DATALINES;
Female Yes
              18
        _No
Female
              12
Male
        Yes
              14
        _No
Male
              16
RIIN:
PROC FREQ DATA=Sample7;
  TABLES Sex*Answer / CHISQ;
                             /* CHISQオプション */
  WEIGHT Cnt;
RIIN:
```

FREQプロシジャの出力の一部

Sex * Answer の統計量			
統計量	自由度	值	p値
カイ 2 乗値	1	1.0714	0.3006
尤度比カイ 2 乗値	1	1.0747	0.2999
連続性補正カイ 2 乗値	1	0.6027	0.4376
Mantel-Haenszel のカイ 2 乗値	1	1.0536	0.3047
ファイ係数		0.1336	
一致係数		0.1325	
Cramer の v 統計量		0.1336	

たくさんの統計量が出力されますが、一番上の行に出力される内容がここでの目的であるピアソンのカイ2乗検定の結果です。p値が大きな数値であることから、男女間に母比率の差があるとまでは言えないという結論になります。

3.4 EG4.1での実現方法

EG4.1ではメニューから「記述統計」>「一元度数表」と選択することによって、1標本の比率に関する推定と検定を行なうことができます。



図3 一元度数表における画面

同じく、メニューから「記述統計」>「分割表分析」と選択することにより、 マクネマー検定やピアソンのカイ2乗検定を行なうことができます。



図4 分割表分析における画面

ただし、ORDER=オプションなどの細かい設定に関しては、選択メニューとして用意されてないものもあります。

3.5 FREQプロシジャにおけるその他の機能

FREQプロシジャには、これまでにご紹介した以外にも数多くの解析機能が存在します。次のリストはその一部です。

- ・ピアソン (Pearson) の相関係数、スピアマン (Spearman) の相関係数、SomersのD、ラムダ統計量などの様々な連関統計量
- ・ヨンキー-タプストラ (Jonckheere-Terpstra) 検定 (*)
- ・コクラン-アーミテージ (Cochran-Armitage) 傾向検定
- ・オッズ比、リスク比、およびリスク差
- ・コクラン-マンテル-ヘンツェル (Cochran-Mantel-Haenszel) 検定 に関連する様々な統計量
- ・様々な統計量に対する信頼区間
- ・正確な分布に基づくp値の算出、またそのp値が現実的な時間で求めることが困難である場合のモンテ・カルロシミュレーション
- (*) JonckheereとTerpstraは、共に人名であり、日本語表記は人によって異なるようです。

3.6 その他の話題

前述のSUGI31では、SAS 9.2におけるSAS/STATソフトウェア全般に対する拡張点を紹介したセッションがありました。

You Can't Stop Statistics

http://www2.sas.com/proceedings/sugi31/185-31.pdf

その中では、FREQプロシジャの追加機能として以下のものが挙げられています。

- ·Zelen検定
- ・2項比率に対する信頼区間の計算方法の追加
- · ODS Graphicsによるグラフ描画

これら3点以外にも、数多くの機能追加が予定されています。

4. おわりに

今号は、TTESTプロシジャやFREQプロシジャの基本的な使い方についてご紹介しました。今後は、他の分析プロシジャについてもとりあげてまいります。



Q&A

- 自動呼出しマクロの保存先を出力(表示)する
- ファイル参照名の存在を確認する方法
- RTFファイルの余白を制御する
- SASシステムビューアで日本語を表示する方法
- シミュレーション実行時の出力を抑制する
- CGM形式のグラフィックデバイスの指定
- 微分係数の算出方法





自動呼出しマクロを利用していますが、どのフォルダに保存さ れているマクロが呼び出されたかを確認する方法はありますか。



SAS9より追加されたMAUTOLOCDISPLAYマクロシステムオプションを利用することで、自動呼び出しマクロの起動時に、ログウィンドウへマクロのソース保存先を出力するこ

例:MAUTOLOCDISPLAYマクロオプションを指定したプログラム例

/* MAUTOLOCDISPLAY */

OPTIONS MAUTOLOCDISPLAY; %LET x=%TRIM(abc):

とが可能です。以下の例を参考にしてください。

例:出力結果例

- 1 OPTIONS MAUTOLOCDISPLAY;
- 2 %LET x=%TRIM(abc);

MAUTOLOCDISPLAY(TRIM): このマクロは、自動呼び出しファイル C:\Program Files\SAS\SAS 9.1\core\sasmacro\trim.sas からコンパイルされました。



FILENAMEステートメントを使用する際に、既存のファイル 参照名を上書きしないようにしたいのですが、ファイル参照 名が存在するかチェックすることは可能ですか。



FILEREF関数を使用することで、ファイル参照名が存在するかどうかの確認を行なうことが可能です。この関数の戻り値は以下のとおりです。

例:FILEREF関数の戻り値

- =0 ファイル参照名および参照先が存在する
- >0 ファイル参照名が割り当てられていない
- <0 ファイル参照名は割り当てられているが、参照先が存在しない

例:FILEREF関数を使用したマクロ例



ODS RTFステートメントにより作成するRTFファイルの余白はどのようにして制御することができますか。

A

TEMPLATEプロシジャにてODS RTFの際に用いられるテンプレートのLEFTMARGIN、RIGHTMARGIN、TOPMARGIN、BOTTOMMARGIN属性を変更することにより、作成される

RTFファイルの余白を指定することができます。デフォルトでは全て2.5インチに設定されています。

例:各MARGIN一覧

LEFTMARGIN: 文書の左側の余白を指定RIGHTMARGIN: 文書の右側の余白を指定TOPMARGIN: 文書の上の余白を指定BOTTOMMARGIN: 文書の下の余白を指定

例:TEMPLATEプロシジャ例

```
PROC TEMPLATE:
 DEFINE STYLE styles.test;
   PARENT=styles.rtf;
   STYLE BODY FROM DOCUMENT/
                             /* 左側の余白を指定 */
     LEFTMARGIN=3in
     RIGHTMARGIN=2.5in
                             /* 右側の余白を指定 */
                             /* 上部の余白を指定 */
     TOPMARGIN=2.5in
     BOTTOMMARGIN=2.5in;
                             /* 下部の余白を指定 */
 END:
RUN;
                     /* 上で作成したTEMPLATEの使用例 */
ODS RTF FILE="c:\temp\test.rtf" STYLE=styles.test;
```

※同様のサンプルプログラムが、以下のURLにも記載されています。 How can I set the margins for the ODS RTF destination? http://support.sas.com/faq/040/FAQ04001.html



SASシステムビューアで、日本語が含まれているデータセット を開くと文字化けします。どのようにすれば文字を正しく表示 できますか。



SASシステムビューアの標準のフォント設定を変更すること で、日本語のデータを表示できます。操作手順は以下のと おりです。

- 1. SASシステムビューアを起動します。
- 2. メニューから[表示] [オプション]を選択します。
- 3. [フォントの設定] タブに移動します。
- 4. 「プロパティ」 欄が 「Grid] および 「Text] の場合のフォントに、"MSゴ シック"などの日本語を表示できるフォントを指定します。
- 5. [OK] ボタンを選択します。



シミュレーションを実行していますが、多くのアウトプット、およ びログへのメッセージが出力されます。これらの出力をせず、必 要とする結果のみをデータセットに出力することはできますか。



ログ画面とアウトプット画面について各々に記載します。

ログ画面への出力に関して

マクロ関係以外WARNING、ERRORメッセージは抑制できませんが、 NOTEメッセージは、NONOTESオプションにて非表示とすることができます。

例:オプション指定例

OPTIONS NONOTES;

また 'OPTIONS NOTES:'の指定にて、デフォルトのNOTEメッセージ 表示に戻すことができます。

アウトプット画面への出力に関して

- ・プロシジャにおけるOUTEST=オプションなどを用い、データセットへ 結果を出力している場合には、NOPRINTオプションにてアウトプット 画面への出力を抑制することができます。
- ・ODS OUTPUTステートメントを用いてデータセットに出力している 場合には、アウトプット画面への出力を以下のように記述することによっ て抑制することができます。

例:プログラムでの指定例

ODS LISTING CLOSE: /* アウトプット画面への出力を抑制 */ < シミュレーションのプログラム >

ODS LISTING;

/* デフォルトに戻す */

また、結果ウィンドウへも多くの出力がされますので、以下のステートメント にて、抑制することが有用です。

例:オプション指定例

ODS NORESULTS:

'ODS RESULTS;'ステートメントにて、デフォルトに戻すことができます。

※ログ、アウトプットを外部ファイルに出力することによって、ログ画面、ア ウトプット画面における表示を抑制することもできます。この点に関しては、 以下のFAQをご参照ください。

http://www.sas.com/japan/service/technical/faq/list/body/ba254.html



グラフをCGM形式のファイルとして出力する際、 DEVICE=CGMMW6Cと指定していましたが、SAS9では、 'デバイスCGMMW6Cが無効です。'のメッセージが表示され ます。CGM形式にて出力するにはどのようにしたらよいですか。

'CGMMW6C'のデバイス名に関しては、SAS9以降ではデフォ ルトにて指定できるデバイス名のリストより外されており、今 後のリリースにて削除される予定となっています。このため、

CGM形式のファイルとして出力する場合には、デバイス名として'CGMOF97L'、 'CGMOF97P'に変更されることをおすすめします。その他、SAS9以降では、 'CGMOFML'、'CGMOFMP'のデバイス名が新たに追加されています。な お、デフォルトでは'CGMMW6C'のデバイス名を指定できませんが、この デバイスの情報は、SASHELP.DGDEVICEのカタログに含まれています。 このため、CATALOGプロシジャにてDEVICEドライバをコピーすることによっ て、'CGMMW6C'の指定を有効とさせることができます。次の例では、 GDEVICE0ライブラリとして既存のフォルダを指定し、SELECTステートメ ントでCGMMW6Cドライバだけをコピーしています。

例:CATALOGプロシジャによるCGMMW6Cドライバのコピー例

/* GDEVICEOのライブラリを作成 */

LIBNAME GDEVICEO 'D:\sasdev\gdev'; /* 任意の既存フォルダを指定 */ PROC CATALOG C=SASHELP.DGDEVICE;

/* CGMMW6Cの情報をコピー */

COPY OUT=GDEVICEO.DEVICES;

SELECT CGMMW6C / ET=DEV;

RUN:

QUIT;



任意の関数を与えたとき、その微分係数を求めることはでき ますか

SAS/ORソフトウェアのNLPプロシジャ、またはSAS/IMLのCALL NLPFDDを使用してください。t以下の例では、2*sin(x)*(cos(y)+1)という関数を考え、x=3.1415926、

y=1.5707963のときの偏微分係数を計算しています。NLPプロシジャでは、PHESSIANオプションを指定すると、2回微分を行なって得られる行列も計算することができます。また、このプロシジャには自動微分の機能が備わっており、基本的にプロシジャ内部で1次(及び2次)の導関数を解析的に求めてから微分係数を算出します。

例:NLPプロシジャの例

```
PROC NLP TECH=NONE PHESSIAN;

PARMS x=3.1415926, y=1.5707963;

f=2*SIN(x)*(COS(y)+1);

MAX f;

RUN;
```

一方、SAS/IMLではSTART/FINISHステートメントで目的関数を定めてから、CALL NLPFDDを使用します。

例:IMLプロシジャの例

```
PROC IML;
START func(x);
f=2*SIN(x[1])*(COS(x[2])+1);
RETURN(f);
FINISH;

x0={3.1415926 1.5707963};
CALL NLPFDD(f,g,h,"func",x0); /* fには関数の値が返る*/
/* gには偏微分係数からなるベクトル、hには
2階の偏微分係数からなる行列が返る*/
PRINT g h;
QUIT;
```



New Publications 新刊マニュアルのお知らせ

SAS[®] Learning Edition リリース 4.1発売開始

発売以来、多くの方にご利用いただいている自宅実習用パッケージ「SAS Learning Edition」の最新版リリース4.1が発売開始されました。
SAS Learning EditionはSAS9のBase機能であるBase SAS 9.1.3、
SAS Enterprise Guide 4.1、SAS/STAT、SAS/GRAPH、SAS/QC、
SAS/ETSを個人のPCにインストールし、付属の教本や豊富なサンプル素材、GUIのインターフェイスによる操作性の良さを活かして統計を学習されたい方に向けた、自宅実習用ソフトウェアパッケージです。

■価格

一般利用	52,500円(税込)
アカデミック利用	39,900円(税込)

※アカデミック利用でお申し込みの場合は、必ず学生証または教職員証のコピーが必要になります。

詳しくは、下記URLをご覧ください。

http://www.sas.com/japan/manual/le.html

■1月の新刊発売予定

Data Preparation for Analytics Using SAS[□]

http://support.sas.com/publishing/bbu/authorline/2006/11_

06/11_06_book.html

著 者: Gerhard Svolba

注文番号: 60502 税込価格: 10,920円

「Decision Trees for Business Intelligence and Data Mining: Using SAS[□] Enterprise Miner □

http://support.sas.com/publishing/bbu/authorline/2006/11_

06/11 06d.html

著 者: Barry de Ville

注文番号: 57587 税込価格: 8,295円

「Introduction to Data Mining Using SAS^a Enterprise Miner ⊓

http://support.sas.com/publishing/bbu/authorline/2006/11_ 06/11_06d.html

著 者: Patricia Cerrito

注文番号: 59829 税込価格: 17,220円

■「分析力のマネジメント」2007年1月発行

「Information Revolution」の邦訳版「分析力のマネジメント」が2007年 1月ダイヤモンド社より発行予定です。本書はSASのBI導入コンサルティング・ナレッジが凝縮された一冊です。企業内の情報活用を進化させるフレームワーク「情報進化モデル」の詳細と、自社のレベルを判定できるアセスメント方法を紹介しています。使い手の立場に立ち、技術のみならず価値観や体制のレベルアップも重視した内容です。本書に関するお問い合わせはマーケティング本部jpnwebmaster@sas.comまで。

SASマニュアル申込用紙、および最新のPublication Catalog (マニュアル案内パンフレット)は弊社ホームページ (http://www.sas.com/japan/manual/)にて公開しておりますので、あわせてご利用ください。

マニュアル販売係

- T E L 03-3533-3835
- F A X 03-3533-3781
- E-mail JPNBooksale@sas.com

SAS Training

SASトレーニングのお知らせ

■特別トレーニングコースの開催のご案内

● 「GENMODプロシジャによるGEE回帰モデル」コース(1日間)

日 程: 2007年2月16日(金)10:00~17:00(東京会場)

価 格: 57,750円(税込)/チケット捺印数 1 受講対象: SASによる統計解析に従事する方

前提知識: 「医薬向けカテゴリカルデータ解析2」コースを受講済みか、

同程度の知識のある方

学習内容: 本講座は、経時的なデザインあるいはクラスターデザインにより、測定値間に何らかの相関が生じるようなデータの解析に携わる方を対象としています。 たとえば、対象者の失明に関する 縦断研究や、世帯および個人の特性を調整した肝臓がんの家族集積性に関する横断研究をGEE(一般化推定方程式)によっ

てモデル化することができます。

コンテンツ: ■GEE回帰モデル

・2値結果変数のためのGEE回帰モデル

- ・順序結果変数のためのGEE回帰モデル
- ■上級者のためのGEE回帰
 - ・GEE回帰モデルのためのモデル診断
 - ·Alternating Logistic Regression Models

担当講師: 東京大学大学院 医学系研究科 生物統計学教室

■新規トレーニングコースの開催のご案内

● 「SAS Enterprise Guideによる高度なデータ加工」 コース(1日間)

日 程: 2007年2月9日(金)、3月7日(水) 10:00 ~ 17:00 (東京会場)

価格: 57,750円(税込)/チケット捺印数1

受講対象: SAS Enterprise Guideを使用して、ポイント&クリックでSAS の機能を活用したい方

前提知識: 「SAS Enterprise Guideによるデータ加工とレポート作成」 コースを受講済みか、同程度の知識のある方

コンテンツ: ■はじめに

・高度なクエリの紹介

・クエリビルダの基本操作

■各種機能の使用

- ・一般的な機能の使い方
- ・機能に関するSASヘルプ
- ■数値の処理
 - ・行毎に要約統計量を計算する方法
 - ・数値の精度を下げる方法
 - ・ 乱数の生成
- ■SAS日付値の処理
 - ・SAS日付値の作成
 - ・SAS日付値からの情報の抽出
 - ・日付間隔の計算
- ■データタイプの変換
 - ・文字値から数値への変換
 - ・数値から文字値への変換

■文字値の処理

- ・文字値の抽出
- ・文字値の結合
- 大文字小文字の修正
- ■条件をつけた値の作成
 - ・値の置換タブの使用
 - ・CASE式の使用
- ■クエリ内のデータ要約
- ・データのグループ化と要約
- 要約データのフィルタリング
- ■その他(自己学習)
 - テーブルの結合
 - DBMSテーブルの照会

● 「SAS Information Map Studioの活用」 コース(1日間)

日 程: 2007年3月27日(火) 10:00 ~ 17:00 (東京会場)

価格: 57,750円(税込)/チケット捺印数 1

受講対象: システム管理者、データ管理者、データ設計者、開発者前提知識: 下記2コースを受講済みか、同程度の知識のある方

・「SASプログラミング I |

·「SASによるSQL入門 |

コンテンツ: ■SASエンタープライズ・インテリジェンス・プラットフォーム入門

- ■SAS Information Map Studioの概要
- ■インフォメーションマップの拡張
- ■インフォメーションマップとフィルタ
- ■インフォメーションマップとOLAPキューブ
- ■インフォメーションマップとストアドプロセス
- ■インフォメーションマップでのデータ表示
- ■インフォメーションマップとXML

■価格改定のご案内

2007年1月より受講料の価格改定を実施いたします。詳細につきましては、 弊社Webサイトをご参照ください。

SAS Institute Japan株式会社では、今後も多岐にわたったトレーニング コースを追加していく予定です。コース内容・日程等の詳細は、順次弊社 Webサイトに公開しますので、以下のURLをご参照ください。

http://www.sas.com/japan/training/

その他、トレーニングに関する情報については、上記のURLをご参照いただくか、下記トレーニング担当までお問い合わせください。

トレーニング担当

- T E L 03-3533-3835
- F A X 03-3533-3781
- E-mail JPNTraining@sas.com

Latest Releases

最新リリース情報

■PCプラットフォーム

Windows版 SAS 9.1.3 9.1 TS1M3 64-bit Windows SAS 9.1.3 9.1 TS1M3 (Itanium) 版

■UNIXプラットフォーム

Tru64版	SAS 9.1.3	9.1 TS1M3
SunOS/Solaris版	SAS 9.1.3	9.1 TS1M3
HP-UX版	SAS 9.1.3	9.1 TS1M3
HP-UX (Itanium) 版	SAS 9.1.3	9.1 TS1M3
AIX版	SAS 9.1.3	9.1 TS1M3
Linux (Intel) 版	SAS 9.1.3	9.1 TS1M3
ABI+版	SAS 6.11	TS040

■ミニコンピュータプラットフォーム

 OpenVMS AXP版
 SAS 6.12
 TS020

 OpenVMS VAX版
 SAS 6.08
 TS407

■メインフレームプラットフォーム

IBM版(OS/390, z/OS)	SAS 9.1.3	9.1 TS1M3
富士通版 (F4, MSP)	SAS 6.09E	TS470
日立版 (VOS3)	SAS 6.09E	TS470
CMS版	SAS 6.08	TS410

Information SAS Technical News 送付についてのご案内

SAS Technical Newsは次の方を対象にお送りしています。

- ·SASコンサルタントとしてご登録の方
- · SAS Technical Newsの購読をお申し込みいただいている方

今後SAS Technical News購読が不要の方、配信先の変更等をご希望の方は、下記URLよりお手続きください。

配信停止

http://www.sas.com/japan/corporate/material.html

配信先変更手続き

http://www.sas.com/japan/sasj_privacy.html#inquiry



発行 SAS Institute Japan株式会社 テクニカルニュースに関するお問い合わせ先

テクニカルサポートグループ TEL:03-3533-3877 FAX:03-3533-3781



東京本社 〒104-0054 東京都中央区勝どき1-13-1 イヌイビル・カチドキ Tel 03 (3533) 6921 Fax 03 (3533) 6927 大阪支店 〒530-0004 大阪市北区堂島浜1-4-16 アクア堂島西館 12F Tel 06 (6345) 5700 Fax 06 (6345) 5655

SAS Institute Japan株式会社

www.sas.com/japan/