

PROC MIANALYZEを用いた 多重代入法による結果の統合

石田和也・斎藤和宏
株式会社タクミインフォメーションテクノロジー

Combination of Results for Multiple Imputation Using PROC MIANALYZE

Kazuya Ishida, Kazuhiro Saito
Takumi Information Technology Inc.

要旨：

欠損値を含むデータの解析方法の1つに多重補完法 (Multiple Imputation) がある。SASではPROC MIで欠損値の補完をした後、結果を統合するためにPROC MIANALYZEを用いる。本発表では、PROC MIANALYZEについて中心にご紹介する。
なお、SASのバージョンは9.4 (SAS/STAT 12.3) を使用した。

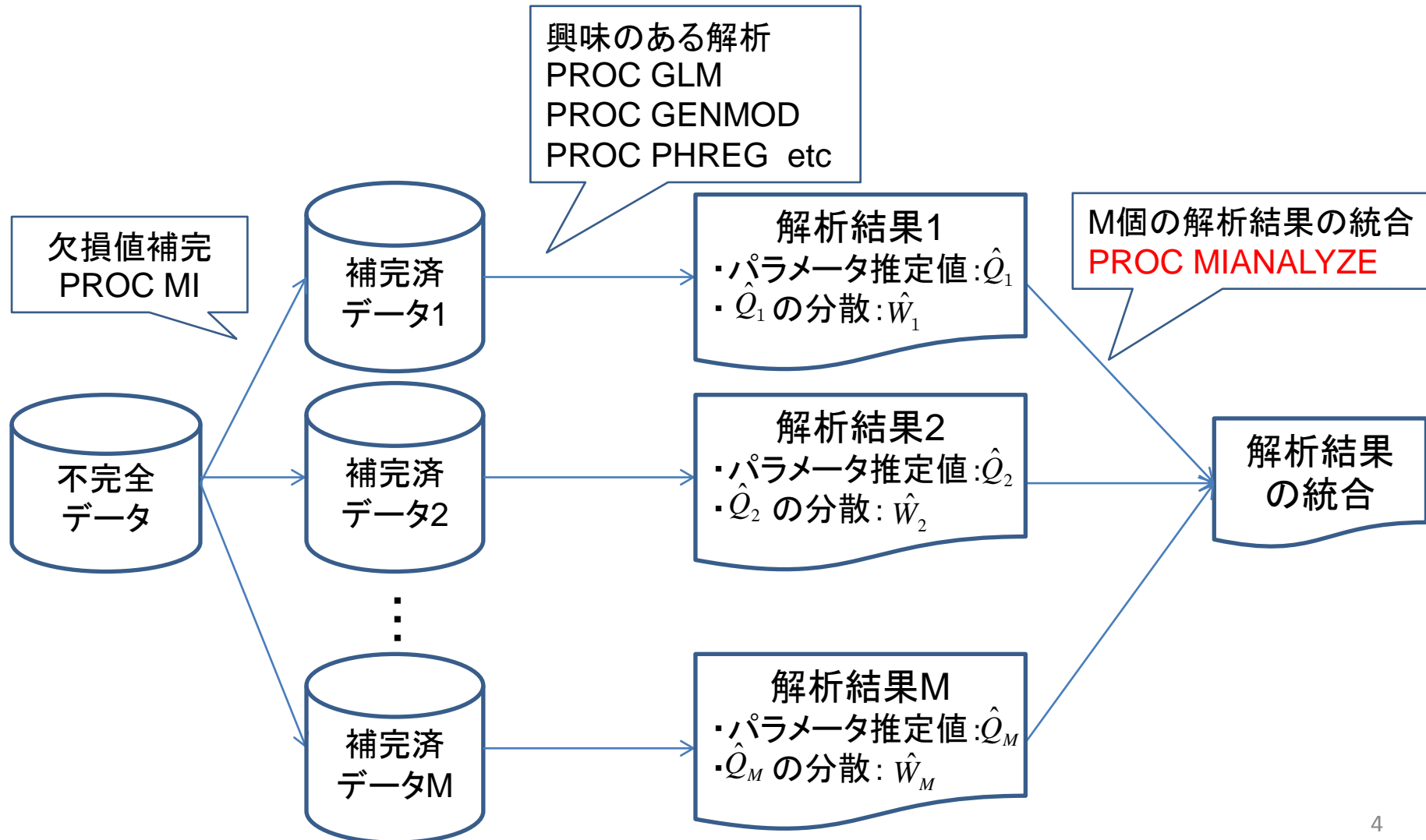
キーワード

PROC MI, PROC MIANALYZE, Multiple Imputation, TYPE3 Test, PROC SQL,

本日の発表構成

1. 多重代入法について
2. 利用事例
3. PROC MIによる欠損値補完
4. 補完後データセットの解析
5. PROC MIANALYZEによる結果の統合
6. 統合後の結果に対する主効果のTYPE3検定について

1. 多重代入法について



1. 多重代入法について

◆ 補完後データによる解析の統合 (Rubin 1987)

➤ 多重代入法によるパラメータ Q の推定値 \bar{Q}
$$\bar{Q} = \frac{1}{M} \sum_{i=1}^M \hat{Q}_i$$

➤ \bar{Q} の分散 T
$$T = \bar{W} + \left(1 + \frac{1}{M}\right) B$$

➤ ここで $B = \frac{1}{M-1} \sum_{i=1}^M (\hat{Q}_i - \bar{Q})^2$ 補完データセット間の分散
(推定間のばらつき)

➤ $\bar{W} = \frac{1}{M} \sum_{i=1}^M \hat{W}_i$ 補完データセット内の分散

➤ 多重代入法によるパラメータ推定値の検定 ($H_0: Q = q_0$)

$$\frac{\bar{Q} - q_0}{\sqrt{T}} \sim t(\nu) \quad \nu = (M-1) \left[1 + \frac{\bar{W}}{\left(1 + \frac{1}{M}\right) B} \right]^2$$

M が大きい (= 補完データ
セット数が多い)
→ 自由度 ν が大きくなる
→ 検定統計量は漸近的に
正規分布にしたがう。

2. 利用事例

◆ 低体重出生児に対するリスクの解析

➤ 1986年マサチューセッツ州、スプリングフィールドにあるBaystate Medical Centerにおいて集められたデータの一部(データに欠損がない、完全データ)

変数名	ラベル	値	備考
LOW	出生体重が2.5kgを下回るか否か (2.5kgを下回った場合、低体重児とする)	0=低体重児ではない 1=低体重児	目的変数
AGE	母親の年齢(歳)		
LWT	最終月経期間における母親の体重(ポンド)		
RACE	母親の人種	1 = 白人, 2 = 黒人 3 = その他	
SMOKE	妊娠期間の喫煙の有無	0=なし, 1=あり	
PTD	過去の早産の有無	0=なし, 1=あり	
HT	高血圧症の有無	0=なし, 1=あり	
UI	子宮炎症の有無	0=なし, 1=あり	
FTV	妊娠後最初の3か月間に医師の診断を受けた回数	0=0回, 1=1回, 2+=2回以上	

2. 利用事例

◆ 完全データに対するPROC LOGISTICによる解析

```
proc logistic data = LBWI ;  
  class race(ref="white") smoke(ref="No") ptd(ref="No")  
        ht(ref="No")      ui(ref="No")      ftv(ref="0") / param=ref ;  
  model low(event="1") = age lwt race smoke ptd ht ui ;  
run ;
```

◆ 完全データに対するPROC GENMODによる解析

```
proc genmod data = LBWI descending ;  
  class race(ref="white") smoke(ref="No") ptd(ref="No")  
        ht(ref="No")      ui(ref="No")      ftv(ref="0") / param=ref ;  
  model low = age lwt race smoke ptd ht ui / link = logit d = binomial ;  
run ;
```


2. 利用事例

- ◆ 完全データに対するPROC LOGISTICによる解析結果
(PROC GENMODの結果は同じなので割愛)

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	0.6369	1.2303	0.2680	0.6047
age	1	-0.0377	0.0378	0.9968	0.3181
lwt	1	-0.0149	0.00704	4.4851	0.0342
race black	1	1.2127	0.5325	5.1870	0.0228
race other	1	0.8041	0.4484	3.2153	0.0730
smoke Yes	1	0.8464	0.4081	4.3020	0.0381
ptd Yes	1	1.2218	0.4630	6.9626	0.0083
ht Yes	1	1.8387	0.7033	6.8359	0.0089
ui Yes	1	0.7111	0.4631	2.3578	0.1247

2. 利用事例

◆ 欠損を発生させた際のPROC LOGISTICによる解析結果

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	0.5787	1.2983	0.1987	0.6558
age	1	-0.0349	0.0399	0.7656	0.3816
lwt	1	-0.0145	0.00752	3.7261	0.0536
race black	1	1.2646	0.5622	5.0600	0.0245
race other	1	0.7702	0.4579	2.8297	0.0925
smoke Yes	1	0.7892	0.4209	3.5149	0.0608
ptd Yes	1	1.0294	0.4992	4.2522	0.0392
ht Yes	1	2.0163	0.7435	7.3553	0.0067
ui Yes	1	0.8350	0.4857	2.9564	0.0855

- 欠損を含むオブザベーションが解析から除外されているため…
 - 完全データでは有意であったLWT(最終月経期間における母親の体重)、SMOKE(喫煙有無)が有意ではない。
 - 全体的にパラメータの標準誤差が大きくなっている
- 解析の精度が下がっていることが分かる

3.PROC MIによる欠損値補完

◆ PROC MIを用いた欠損値補完のプログラム例

```
proc mi data = LBWI_miss out = LBWI_MI noprint
      seed = 123456 nimpute = 20 ;
class race ftv ht ;
fcs discrim(race / classeffects=include) logistic(ftv) logistic(ht) reg(lwt);
var race ftv ht lwt age ;
run ;
```

補完データセット数
Mの設定

判別関数による
多重補完(名義尺度)

ロジスティック回帰による
多重補完(順序尺度)

線形回帰による
多重補完(連続変数)

- SAS9.3より追加されたFCSステートメントにより、Fully Conditional Specificationによる補完が可能になった。(SAS9.3は評価版)
→RACEのような名義尺度についてもPROC MIによる多重補完が可能となった。

4. 補完後データセットによる解析

◆ PROC LOGISTICによる解析

```
proc logistic data = LBWI_MI outest=Log_Param covout ;  
by _imputation_ ;  
class race(ref="white") smoke(ref="No") ptd(ref="No")  
      ht(ref="No")      ui(ref="No")      ftv(ref="0") / param=ref ;  
model low(event="1") = age lwt race smoke ptd ht ui ;  
run ;
```

- 補完データセットごとの解析
- パラメータとその分散共分散行列のデータセット化

◆ PROC GENMODによる解析

```
ods output ParameterEstimates=Gen_Param COVB=Gen_Cov  
Parminfo=Gen_Info ;  
proc genmod data = LBWI_MI descending ;  
by _imputation_ ;  
class race(ref="white") smoke(ref="No") ptd(ref="No")  
      ht(ref="No")      ui(ref="No")      ftv(ref="0") / param=ref ;  
model low = age lwt race smoke ptd ht ui / link = logit d = binomial ;  
run ;
```

5. PROC MIANALYZEによる結果の統合

◆ PROC LOGISTICによる結果の統合

```
proc mianalyze data = Log_Param ;
  modeleffect Intercept age lwt raceblack raceother smokeYes ptdYes
  htYes uiYes ;
run ;
```

➤ MODELEFFECTステートメントにおいて、カテゴリ変数を、「変数名+水準」の形で指定する。

➤ OUTEST=オプションを使用することができるREGプロシジャ、PHREGプロシジャなどは、同じ指定方法となる。

パラメータと分散共分散行列が1つのデータセットにまとめられているので、PROC MIANALYZEの指定方法は他プロシジャと比べて容易

VIEWTABLE: Work.Log_param (Parameter Estimates and Covariance Matrix)

	Imputation Number	Link function	Type of Statistics	収束状態	Row Names for Parameter Estimates and Covariance Matrix	Intercept: low=0	母親
1	1	LOGIT	PARMS	0 Converged	low	0.736747536	****
2	1	LOGIT	COV	0 Converged	Intercept	1.5030304673	****
3	1	LOGIT	COV	0 Converged	age	-0.027282238	****
4	1	LOGIT	COV	0 Converged	lwt	-0.005628792	****
5	1	LOGIT	COV	0 Converged	raceblack	-0.041931557	****
6	1	LOGIT	COV	0 Converged	raceother	-0.215891363	****
7	1	LOGIT	COV	0 Converged	smokeYes	-0.124731095	****
8	1	LOGIT	COV	0 Converged	ptdYes	0.0205272132	****
9	1	LOGIT	COV	0 Converged	htYes	0.095609564	****
10	1	LOGIT	COV	0 Converged	uiYes	-0.081275363	****
11	2	LOGIT	PARMS	0 Converged	low	0.5615715911	****
12	2	LOGIT	COV	0 Converged	Intercept	1.5589405621	****
13	2	LOGIT	COV	0 Converged	---	0.00001175	****

5. PROC MIANALYZEによる結果の統合

◆ PROC MIANALYZEによる出力結果

Parameter Estimates					
Parameter	Estimate	Std Error	95% Confidence Limits		DF
intercept	0.701462	1.246553	-1.74198	3.14490	12163
age	-0.041679	0.037776	-0.11572	0.03236	144417
lwt	-0.014656	0.007191	-0.02875	-0.00056	9099.1
raceblack	1.247608	0.537399	0.19428	2.30094	27594
raceother	0.774883	0.451135	-0.10953	1.65930	5271.9
smokeYes	0.835670	0.405126	0.04159	1.62975	20539
ptdYes	1.210760	0.465283	0.29881	2.12271	89009
htYes	1.758649	0.743711	0.29954	3.21776	1203.7
uiYes	0.694328	0.466967	-0.22093	1.60958	75355

Parameter Estimates					
Parameter	Minimum	Maximum	Theta0	t for H0:	
				Parameter=Theta0	Pr > t
intercept	0.375152	1.181208	0	0.56	0.5736
age	-0.050861	-0.035429	0	-1.10	0.2699
lwt	-0.018433	-0.012926	0	-2.04	0.0416
raceblack	1.085831	1.374088	0	2.32	0.0203
raceother	0.489834	0.968404	0	1.72	0.0859
smokeYes	0.728863	0.961614	0	2.06	0.0391
ptdYes	1.074406	1.297235	0	2.60	0.0093
htYes	1.259269	2.086117	0	2.36	0.0182
uiYes	0.584834	0.772848	0	1.49	0.1370

5. PROC MIANALYZEによる結果の統合

◆ MODELEFFECTステートメントの自動化

➤ パラメータデータセットからマクロ変数を作成する

```
proc sql noprint ;
  select _NAME_
  into : modeleffect separated by ' '
  from Log_Param
  where _imputation_=1 and _TYPE_="COV" ;
quit ;
```

➤ SASログより

```
%put modeleffect = &modeleffect ;
modeleffect = Intercept age lwt raceblack raceother
smokeYes ptdYes htYes uiYes
```

➤ マクロ変数を用いると……

```
proc mianalyze data = Log_Param ;
  modeleffect &modeleffect ;
run ;
```

SELECT句で指定した変数をマクロ変数化する。その際……

- ・変数を横に並べる。
- ・変数の間はSEPARATED BYで指定したデリミタで区切る
(左記は半角空白で区切る場合)

VIEWTABLE: Work.Log_param (Parameter Estimates and Covariance Matrix)

	Imputation Number	Link function	Type of Statistics	収束状態	Row Names for Parameter Estimates and Covariance Matrix	Intercept: low=0	母親
1	1	LOGIT	PARMS	0 Converged	low	0.736747536	****
2	1	LOGIT	COV	0 Converged	Intercept	1.5030304673	****
3	1	LOGIT	COV	0 Converged	age	-0.027282238	****
4	1	LOGIT	COV	0 Converged	lwt	-0.005628792	****
5	1	LOGIT	COV	0 Converged	raceblack	-0.041931557	****
6	1	LOGIT	COV	0 Converged	raceother	-0.215891363	****
7	1	LOGIT	COV	0 Converged	smokeYes	-0.124731095	****
8	1	LOGIT	COV	0 Converged	ptdYes	0.0205272132	****
9	1	LOGIT	COV	0 Converged	htYes	0.095609564	****
10	1	LOGIT	COV	0 Converged	uiYes	-0.081275363	****
11	2	LOGIT	PARMS	0 Converged	low	0.5615715911	****
12	2	LOGIT	COV	0 Converged	Intercept	1.5589405621	****
13	2	LOGIT	COV	0 Converged	age	0.0000000000	****

5. PROC MIANALYZEによる結果の統合

◆ PROC GENMODによる結果の統合

```
proc mianalyze parms(classvar=level) = Gen_Param covb = Gen_COV
                parminfo = Gen_info ;
class race ;
modeleffect intercept age lwt race smoke ptd ht ui ;
run ;
```

➤ CLASSVAR=オプション

- FULL, LEVEL, CLASSVALの中から適切なものを選択。(デフォルトはFULL)
- PROC GENMODの場合、パラメータデータセットのカテゴリ水準の変数名の接頭辞がLEVELなので、LEVELと指定。

- CLASSステートメントでカテゴリ変数を指定するが、2水準のカテゴリ水準の場合は指定しなくてもよい。

VIEWTABLE: Work.Gen_param (パラメータ推定値の分析)

	Imputation Number	パラメータ	Level1	自由度	推定値	標準誤差	95% Lower Confidence Limit	95% Upper Confidence Limit
1	1	Intercept		1	0.7367	1.2260	-1.6661	3.1396
2	1	age		1	-0.0354	0.0369	-0.1078	0.0369
3	1	lwt		1	-0.0155	0.0072	-0.0296	-0.0014
4	1	race	black	1	1.1509	0.5272	0.1175	2.1842
5	1	race	other	1	0.7556	0.4338	-0.0946	1.6059
6	1	smoke	Yes	1	0.8424	0.3943	0.0695	1.6153
7	1	ptd	Yes	1	1.1676	0.4586	0.2688	2.0664
8	1	ht	Yes	1	1.2593	0.6407	0.0035	2.5150
9	1	ui	Yes	1	0.6171	0.4622	-0.2889	1.5230
10	1	Scale		0	1.0000	0.0000	1.0000	1.0000
11	2	Intercept		1	0.5616	1.2486	-1.8856	3.0087
12	2	age		1	-0.0404	0.0375	-0.1139	0.0330
13	2	lwt		1	-0.0138	0.0072	-0.0278	0.0003
14	2	race	black	1	1.0858	0.5200	0.0666	2.1051
15	2	race	other	1	0.7718	0.4399	-0.0904	1.6340

6. 統合後の結果に対する主効果のTYPE3検定について

◆ Multivariate Inference (Rubin 1987, Schafer 1997)

➤ 多重代入法によるパラメータベクトル \mathbf{Q} の推定値 $\bar{\mathbf{Q}}$
$$\bar{\mathbf{Q}} = \frac{1}{M} \sum_{i=1}^M \hat{\mathbf{Q}}_i$$

Ex) \mathbf{Q} : 人種を表すパラメータベクトル

$$\mathbf{Q} = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} \text{黒人と白人の平均の差} \\ \text{その他と白人の平均の差} \end{bmatrix}$$

➤ $\bar{\mathbf{Q}}$ の分散 \mathbf{T}_0

$$\mathbf{T}_0 = \bar{\mathbf{W}} + \left(1 + \frac{1}{M}\right) \mathbf{B}$$

➤ ここで $\mathbf{B} = \frac{1}{M-1} \sum_{i=1}^M (\hat{\mathbf{Q}}_i - \bar{\mathbf{Q}})(\hat{\mathbf{Q}}_i - \bar{\mathbf{Q}})'$ 補完データセット間の分散共分散行列

➤ $\bar{\mathbf{W}} = \frac{1}{M} \sum_{i=1}^M \hat{\mathbf{W}}_i$ 補完データセット内の分散共分散行列

➤ 下記のような統計量 F_0 を考える

$$F_0 = (\mathbf{Q}_0 - \bar{\mathbf{Q}})' \mathbf{T}_0^{-1} (\mathbf{Q}_0 - \bar{\mathbf{Q}}) / p \sim F(p, \nu) \quad (H_0: \mathbf{Q} = \mathbf{Q}_0)$$

$$p: \text{Number of Level}, \nu = (M-1) \left(1 + \frac{1}{r}\right)^2$$

where $r = \left(1 + \frac{1}{M}\right) \text{trace}(\mathbf{B}\bar{\mathbf{W}}^{-1}) / p$

Average relative increase in variable

6. 統合後の結果に対する主効果のTYPE3検定について

◆ Multivariate Inference (Rubin 1987, Schafer 1997)

➤ F統計量 $F_0 = (\mathbf{Q}_0 - \bar{\mathbf{Q}})' \mathbf{T}_0^{-1} (\mathbf{Q}_0 - \bar{\mathbf{Q}}) / p \sim F(p, \nu)$ ($H_0 : \mathbf{Q} = \mathbf{Q}_0$)

➤ 問題点

➤ 補完データセット数が少ない(Mが小さい)ときに、補完データセット間の分散共分散行列 \mathbf{B} が不安定となる。特に $M \leq p$ のときには \mathbf{B} がフルランクとならない。

➤ 解決方法

➤ 補完データセット間、補完データセット内の分散共分散行列が **比例関係** にあることを仮定する。

➤ この仮定の下では $\bar{\mathbf{Q}}$ の分散 \mathbf{T} は $\mathbf{T} = \bar{\mathbf{W}} + r\bar{\mathbf{W}} = (1+r)\bar{\mathbf{W}}$

比例関係を仮定しない場合は...

$$\mathbf{T}_0 = \bar{\mathbf{W}} + \left(1 + \frac{1}{M}\right) \mathbf{B}$$

➤ この \mathbf{T} を用いて、F統計量を構築しなおす。

➤ F統計量 $F = (\mathbf{Q}_0 - \bar{\mathbf{Q}})' \mathbf{T}^{-1} (\mathbf{Q}_0 - \bar{\mathbf{Q}}) / p \sim F(p, \nu_1)$ ($H_0 : \mathbf{Q} = \mathbf{Q}_0$)

◆ PROC MIANALYZEではこのロジックを採用している

$$\text{where } \nu_1 = \begin{cases} \frac{1}{2}(p+1)(m-1)\left(1+\frac{1}{r}\right)^2 & \text{if } t = p(m-1) \leq 4 \\ 4 + (t-4)\left[1 + \frac{1}{r}\left(1 - \frac{2}{t}\right)\right]^2 & \text{if } t = p(m-1) > 4 \end{cases}$$

6. 統合後の結果に対する主効果のTYPE3検定について

◆ TYPE3検定とは

- モデルにカテゴリ変数を用いた解析において、カテゴリ変数に含まれるいずれかの水準の間に有意な差があるかどうかを確認する検定。

- $H_0 : \mathbf{Q} = \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ $\left(\begin{array}{l} \mathbf{L} : \text{係数行列 (カテゴリ変数に対する対比行列)} \\ \boldsymbol{\beta} : \text{カテゴリ共変量のパラメータベクトル} \end{array} \right)$

- Type3検定のp値が0.05未満であれば、有意水準5%で当該カテゴリ変数のいずれかの水準の間に有意な差があると主張することができる。

- PROC LOGISTIC、PROC GENMODなどの統計解析プロシジャでは、CLASSステートメントで指定された共変量がモデルに含まれている場合はデフォルトでTYPE3検定の結果が出力されることが一般的であるが、PROC MIANALYZEではデフォルトでは出力されない。



- 係数行列 \mathbf{L} を理解した上で、自らTYPE3検定を行う(=プログラムを構築する)必要がある。
- TYPE3検定はPROC MIANALYZEのTESTステートメントで行うことができる。

6. 統合後の結果に対する主効果のTYPE3検定について

◆ 係数行列Lの確認

- 以下、カテゴリ変数をReference Codingで扱うこととする。
- Reference Codingでは係数行列Lは $(N-1)$ 次の正方行列(N :カテゴリ変数の水準数)
- 今回の数値例において、人種(白人、黒人、その他の3水準)のTYPE3検定を考える。
- 白人をReferenceとすると、係数行列Lは以下のようなになる。

$$\mathbf{L} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{array}{l} \text{1行目} \rightarrow \text{「黒人」と「白人」の水準の差、} \\ \text{2行目} \rightarrow \text{「その他」と「白人」の水準の差} \\ \text{をそれぞれ表す。} \end{array}$$

- 人種のカテゴリ変数に対するパラメータ推定値を β_1 (「黒人」と「白人」の水準の差)、 β_2 (「その他」と「白人」の水準の差)とする。

- 人種のカテゴリについてのTYPE3検定は下記のようなになる。

$$H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$$

$$\mathbf{L}\boldsymbol{\beta} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \mathbf{0}$$

$$\Rightarrow \beta_1 = 0 \quad \text{and} \quad \beta_2 = 0$$

$$\text{vs} \quad H_1 : \beta_1 \neq 0 \quad \text{or} \quad \beta_2 \neq 0$$

少なくとも1水準について、Reference水準との間に有意な差がある
(TYPE3検定:分散分散的な検定)

6. 統合後の結果に対する主効果のTYPE3検定について

◆ PROC MIANALYZEのTESTステートメント(PROC LOGISTICの場合)

```
proc mianalyze data = Log_Param ;
  modeleffect Intercept age lwt raceblack raceother smokeYes ptdYes
  htYes uiYes ;
  test raceblack = 0 , raceother = 0 / mult ;
run ;
```

- ・帰無仮説をカンマ区切りでコーディング
- ・MULTオプションで多変量に対する検定を実施

補完データセット間、補完データセット内の分散共分散行列が**比例関係**にあることを仮定している。

◆ PROC MIANALYZEによる出力結果(TESTステートメント)

Multivariate Inference
Assuming Proportionality of Between/Within Covariance Matrices

r	p	ν_1	F	
Avg Relative Increase in Variance	Num DF	Den DF	F for H0: Parameter=Theta0	Pr > F
0.046696	2	15412	3.07	0.0465

6. 統合後の結果に対する主効果のTYPE3検定について

◆ TESTステートメントの自動化

➤ パラメータデータセットからマクロ変数を作成する

```
data MakeType3 ;
  set Log_param ;
  where _imputation_ = 1 and _TYPE_ = "COV" and index(_NAME_ , "race") > 0 ;
  Type3 = trim(_NAME_) || "=0" ;  %*TEST Statementのパーツの作成;
run ;
```

```
proc sql noprint ;
  select Type3
  into : Type3 separated by ", "
  from MakeType3 ;
quit ;
```

・TEST Statementの構築には、
カンマ区切りのプログラムが必要。
→SEPARATED BY で ", " と指定すればよい

➤ マクロ変数を用いると……

```
proc mianalyze data = Log_Param ;
  modeleffect &modeleffect ;
  test &Type3 / mult ;
run ;
```

VIEWTABLE: Work.Log_param (Parameter Estimates and Covariance Matrix)

	Imputation Number	Link function	Type of Statistics	収束状態	Row Names for Parameter Estimates and Covariance Matrix	Intercept: low=0	母親
1	1	LOGIT	PARMS	0 Converged	low	0.736747536	****
2	1	LOGIT	COV	0 Converged	Intercept	1.5030304673	****
3	1	LOGIT	COV	0 Converged	age	-0.027282238	****
4	1	LOGIT	COV	0 Converged	lwt	-0.005628792	****
5	1	LOGIT	COV	0 Converged	raceblack	-0.041931557	****
6	1	LOGIT	COV	0 Converged	raceother	-0.215891363	****
7	1	LOGIT	COV	0 Converged	smokeYes	-0.124731095	****
8	1	LOGIT	COV	0 Converged	ptdYes	0.0205272132	****
9	1	LOGIT	COV	0 Converged	htYes	0.095609564	****
10	1	LOGIT	COV	0 Converged	uiYes	-0.081275363	****
11	2	LOGIT	PARMS	0 Converged	low	0.5615715911	****
12	2	LOGIT	COV	0 Converged	Intercept	1.5589405621	****

6. 統合後の結果に対する主効果のTYPE3検定について

◆ PROC MIANALYZEのTESTステートメント(PROC GENMODの場合)

```
proc mianalyze parms(classvar=level) = Gen_Param covb = Gen_COV
                    parminfo = Gen_info ;

class race ;
modeleffect intercept age lwt race smoke ptd ht ui ;
test race = 0 ;

run ;
```

◆ SASログ

WARNING: The TEST statement can not be used when a CLASS statement is specified.

➤ TESTステートメントと、CLASSステートメントの併用ができない。

→パラメータデータセット(Gen_Param)とパラメータ詳細(Gen_info)データセットの加工が必要

Imputation Number	パラメータ	Level1	自由度	推定値	標準誤差	95% Lower Confidence Limit	95% Upper Confidence Limit
1	Intercept		1	0.7367	1.2260	-1.6661	3.1396
2	age		1	-0.0354	0.0369	-0.1078	0.0369
3	lwt		1	-0.0155	0.0072	-0.0296	-0.0014
4	age	black	1	1.1509	0.5272	0.1175	2.1842
5	age	other	1	0.7556	0.4338	-0.0946	1.6059
6	smoke	Yes	1	0.8424	0.3943	0.0695	1.6153
7	ptd	Yes	1	1.1676	0.4586	0.2809	2.0664
8	ht	Yes	1	1.2593	0.6407	0.0035	2.5150
9	ui	Yes	1	0.6171	0.4522	-0.2889	1.5230
10	Scale		0	1.0000	0.0000	1.0000	1.0000
11	Intercept		1	0.5616	1.2486	-1.8856	3.0087
12	age		1	-0.0404	0.0375	-0.1139	0.0330
13	lwt		1	-0.0138	0.0072	-0.0278	0.0003
14	race	black	1	1.0858	0.5200	0.0666	2.1051
15	race	other	1	0.7218	0.4399	-0.0904	1.6340

Imputation Number	パラメータ	効果	race	smoke	ptd	ht	ui
1	Prm1	Intercept					
2	Prm2	age					
3	Prm3	lwt					
4	Prm4	race	black				
5	Prm5	race	other				
6	Prm6	smoke		Yes			
7	Prm7	ptd			Yes		
8	Prm8	ht				Yes	
9	Prm9	ui					Yes
10	Prm1	Intercept					
11	Prm2	age					

6. 統合後の結果に対する主効果のTYPE3検定について

◆ PROC MIANALYZEのTESTステートメント(PROC GENMODの場合)

➤ データセットの加工

```
data Gen_Param2 ;
  length Parameter $ 128 ;
  set Gen_Param ;
  if Parameter = "race" then Parameter = trim(Parameter) || trim(Level1) ;
run ;

data Gen_info2 ;
  length Effect $ 128 ;
  set Gen_info ;
  if race ^= "" then Effect = trim(Effect) || trim(race) ;
run ;
```

カテゴリ変数とその水準を結合して、
ユニークな変数値を作成する

	パラメータ	Imputation Number	Level1	自由度	推定値	標準誤差
1	Intercept	1		1	0.7367	1.2260
2	age	1		1	-0.0354	0.0369
3	lwt	1		1	-0.0155	0.0072
4	raceblack	1	black	1	1.1509	0.5272
5	raceother	1	other	1	0.7556	0.4338
6	smoke	1	Yes	1	0.8424	0.3943
7	ptd	1	Yes	1	1.1676	0.4586
8	ht	1	Yes	1	1.2593	0.6407
9	ui	1	Yes	1	0.6171	0.4622
10	Scale	1		0	1.0000	0.0000
11	Intercept	2		1	0.5616	1.2486
12	age	2		1	-0.0404	0.0375

	効果	Imputation Number	パラメータ	race	smoke	ptd
1	Intercept	1	Prm1			
2	age	1	Prm2			
3	lwt	1	Prm3			
4	raceblack	1	Prm4	black		
5	raceother	1	Prm5	other		
6	smoke	1	Prm6		Yes	
7	ptd	1	Prm7			Yes
8	ht	1	Prm8			
9	ui	1	Prm9			
10	Intercept	2	Prm1			
11	age	2	Prm2			
12	lwt	2	Prm3			

6. 統合後の結果に対する主効果のTYPE3検定について

◆ 修正後のプログラム (PROC GENMODの場合)

```
proc mianalyze parms = Gen_Param2 covb = Gen_COV  
                parminfo = Gen_info2 ;  
modeffect raceblack raceother ;  
test raceblack = 0 , raceother = 0 / mult ;  
run ;
```

◆ PROC MIANALYZEによる出力結果 (TESTステートメント)

Multivariate Inference
Assuming Proportionality of Between/Within Covariance Matrices

Avg Relative Increase in Variance	Num DF	Den DF	F for H0: Parameter=Theta0	Pr > F
0.046696	2	15412	3.07	0.0465

➤ PROC LOGISTICのときと同じ結果が得られる

まとめ

1. 多重代入法について

解析の精度が下がることを防ぐため、欠損値を含んだデータに対しては、多重代入法により、欠損値の補完を行った上で解析を行うことが有用である。SASでは、**PROC MIで多重代入法による欠損値補完**を、**PROC MIANALYZEで結果の統合**を行うことができる。

2. PROC MIANALYZEによる結果の統合

PROC MIANALYZEに結果の統合を行うためには、補完データセットごとの**解析結果のパラメータデータセットと分散共分散行列のデータセット**が必要となる。これらのデータセットの構造は、統計解析プロシジャにより異なるので、PROC MIANALYZEの指定方法が、使用した統計解析プロシジャにより異なる。

3. 統合後の結果に対する主効果のTYPE3検定について

モデルにカテゴリ変数を用いた解析では、カテゴリ変数に含まれるいずれかの水準の間に有意な差があるかどうかをTYPE3検定を用いることがある。PROC MIANALYZEでは、**TESTステートメントにより、TYPE3検定**を行うことができる。その際、**補完データセット間、補完データセット内の分散共分散行列が比例関係にあることを仮定している**ので、注意が必要である。

参考文献

- ◆MI Procedureによる多重代入 SAS ver 9.3における新機能の紹介(2013) / 多田圭佑
- ◆Multiple Imputation法によるネステッドコントロール研究、ケースコホート研究の解析(2012) / 野間久史・田中司郎・田中佐智子・和泉志津恵
- ◆ロジスティック回帰分析 SASを利用した統計解析の実際(1996) / 丹後俊郎・山岡和枝・高木春良 著
- ◆Applied Logistic Regression. Wiley Series in Probability and Statistics (1989) / Hosmer D.W. , Lemeshow S.
- ◆SAS for Linear Models (1996) / Ramon Littell , Walter W. Stroup, Rudolf Freund
- ◆Analysis of Incomplete Multivariate Data (1997) / J.L. Schafer
- ◆SAS/STAT 9.4 User's Guide

ご清聴ありがとうございました